2013

Applying Data Visualisations to Open New Perspectives in Birdwatching

Frank Kendlin
Technological University Dublin, frank.kendlin@dit.ie

Follow this and additional works at: https://arrow.dit.ie/scschcomdis

Part of the Computer Engineering Commons

Recommended Citation

This Dissertation is brought to you for free and open access by the School of Computing at ARROW@TU Dublin. It has been accepted for inclusion in Dissertations by an authorized administrator of ARROW@TU Dublin. For more information, please contact yvonne.desmond@dit.ie, arrow.admin@dit.ie, brian.widdis@dit.ie.

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License
Applying Data Visualisations to Open New Perspectives in Birdwatching

Frank Kendlin

A dissertation submitted in partial fulfilment of the requirements of Dublin Institute of Technology for the degree of MSc in Computing (Data Analytics)

January 2013
I certify that this dissertation which I now submit for examination for the award of MSc in Computing (Data Analytics), is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

This dissertation was prepared according to the regulations for postgraduate study of the Dublin Institute of Technology and has not been submitted in whole or part for an award in any other Institute or University.

The work reported on in this dissertation conforms to the principles and requirements of the Institute’s guidelines for ethics in research.

Signed: _________________________________

Date: January 7 2013
ABSTRACT

Advances in data analytics for big data have affected many different domains and are providing new insights from these data for their respective communities. One such community is conservation science. A part of section of this community, ornithologists and bird conservationists has at its disposal millions of volunteers willing to contribute to a new ‘big’ data set. This can be achieved by means of new technology and the use of citizen science. This is already underway in North America through the excellent work of the Cornell Laboratory of Ornithology (CLO). This institute has harnessed the technology and the people and science is benefiting from the results. We would like to understand if a gap exists between the approach of the CLO and comparable groups in Ireland and the United Kingdom and if so, what issues need to be addressed to enable adoption of the new methodologies outside North America. This paper researches the effect of data visualisation techniques such as geo representation and tabular displays on motivating a community of interest to achieve a critical mass for successful delivery of a desirable new ‘big data’ set. Our work includes a review of data visualisation as part of data analytics, a review of how data visualisation has been used by conservation science in the past, and what the state of the art looks like today. We try to find key features in the techniques which are valuable for motivating the stakeholders. We also try to find the impact of the new insights provided from large spatio-temporal data sets on possible new beneficiaries.

Key words: Spatio-temporal data, data visualisation, ornithology, citizen science, geo-visualisation, bird watching, birding
I would like to express my sincere thanks to my supervisor for this dissertation, Brendan Tierney, for his encouragement and advice during the work. I would also like to thank Damian Gordon of DIT for his advice and support. I received considerable help and support from the staff of BirdWatch Ireland and, in particular the committee of the Tolka branch. I also thank the staff at the Cornell Laboratory for Ornithology and the British Trust for Ornithology. I thank my friends and family for their support and for putting up with me during the three months of work. I especially thank my sister Eilis who put in a marathon effort at short notice to help with the proofing. Any remaining errors are totally my own.
# TABLE OF CONTENTS

**Contents**

ABSTRACT ........................................................................................................................................ II

TABLE OF FIGURES ................................................................................................................ VIII

1 INTRODUCTION....................................................................................................................... 1

1.1 OVERVIEW .......................................................................................................................... 1

1.2 BACKGROUND ................................................................................................................... 2

1.2.1 Data Visualization and Wildlife conservation ...................................................... 2

1.2.2 Ornithology and Conservation ................................................................................. 3

1.2.3 Ornithological and Related Data ............................................................................. 3

1.3 RESEARCH PROBLEM .................................................................................................... 4

1.4 INTELLECTUAL CHALLENGE ......................................................................................... 5

1.5 RESEARCH OBJECTIVES ................................................................................................. 6

1.6 RESEARCH METHODOLOGY ............................................................................................ 7

1.6.1 Survey Questionnaire ............................................................................................... 8

1.6.2 Structured Interviews ............................................................................................... 8

1.7 SCOPE AND LIMITATIONS ............................................................................................... 9

1.8 ORGANISATION OF THE DISSERTATION ..................................................................... 11

2 DATA VISUALISATION .......................................................................................................... 13

2.1 INTRODUCTION ............................................................................................................... 13

2.2 HISTORICAL BACKGROUND ............................................................................................ 14

2.3 MODERN ERA OF DATA VISUALISATION ........................................................................ 15

2.4 DEFINITIONS ..................................................................................................................... 18

2.5 CATEGORISING DATA VISUALISATIONS ......................................................................... 20

2.5.1 Complexity .................................................................................................................. 20

2.5.2 Exploration Vs. Explanation ...................................................................................... 20

2.5.3 Infographics Vs. Data Visualisation ........................................................................... 21

2.6 TOOLS USED FOR DATA VISUALISATION ...................................................................... 21

2.7 PRACTICAL EXPLORATION OF TOOLS .......................................................................... 26

2.7.1 ‘R’ ................................................................................................................................. 26
## 5.3.4 Section 3 (4 questions about the respondents) ..................................... 64
## 5.3.5 Section 4 (5 questions on bird watching experience) ........................... 65
## 5.3.6 Section 5 (4 questions about experience with technology) ................... 66
## 5.3.7 Conclusion ............................................................................................ 67

## 5.4 THE INTERVIEWS: ............................................................................................. 67
### 5.4.1 Interview Section One ........................................................................... 69
### 5.4.2 Interview Section Two ........................................................................... 71
### 5.4.3 Conclusion ............................................................................................ 71

## 6 EVALUATION ................................................................................................... 72
### 6.1 INTRODUCTION ................................................................................................. 72
### 6.2 ANALYSIS OF THE QUESTIONNAIRE RESPONSES .................................................. 72
#### 6.2.1 Section One Analysis ............................................................................ 72
#### 6.2.2 Section Two Analysis ............................................................................ 73
#### 6.2.3 Section Three Analysis .......................................................................... 75
#### 6.2.4 Section Four Analysis ........................................................................... 76
#### 6.2.5 Section Five Analysis ............................................................................ 77
#### 6.2.6 Note of the tools used for the questionnaire ......................................... 79
### 6.3 ANALYSIS OF INTERVIEWS ............................................................................... 80
#### 6.3.1 Interview Part one ................................................................................ 80
#### 6.3.2 The Interview Part Two ........................................................................ 83
### 6.4 CONCLUSION OF EVALUATION ............................................................................. 84
#### 6.4.1 Conclusions from the survey ................................................................. 84
#### 6.4.2 Conclusions from the Interviews ............................................................ 85

## 7 CONCLUSION ................................................................................................... 86
### 7.1 INTRODUCTION ................................................................................................. 86
### 7.2 RESEARCH DEFINITION & RESEARCH OVERVIEW .................................. 86
### 7.3 CONTRIBUTIONS TO THE BODY OF KNOWLEDGE ........................................ 87
### 7.4 EXPERIMENTATION, EVALUATION AND LIMITATION .............................. 87
### 7.5 FUTURE WORK & RESEARCH ........................................................................... 88

## BIBLIOGRAPHY ........................................................................................................ I

## APPENDIX A ............................................................................................................. V
1. SECTION ONE ........................................................................................................... V
2. SECTION TWO ......................................................................................................... XV
TABLE OF FIGURES

FIGURE 2.1 MINARD’S *CARTE FIGURATIVE*, A GRAPHIC DEPICTION OF NAPOLEON’S *GRAND ARMÉES* DISASTROUS MARCH ON MOSCOW ................................................................. 14
FIGURE 2.2. THE BASIC BOX PLOT AND B THE BOX PLOT AS A DEVICE FOR COMPARING DISTRIBUTIONS ........................................................................................................ 15
FIGURE 2.3 SPARKLINES .......................................................................................... 17
FIGURE 2.4 THE ‘DESIGNER – READER – DATA TRINITY’ ........................................ 18
FIGURE 2.5 COMPLEX CHART EXAMPLE .................................................................. 19
FIGURE 2.6 A SIMPLE GRAPH PLOTTING DEATH ROW PRISONERS AND EXECUTIONS OVER 35 YEARS ................................................................................................ 20
FIGURE 2.7 THE RSTUDIO IDE AT COMMENCEMENT OF TRIAL ................................. 27
FIGURE 2.8 THE RSTUDIO IDE AFTER THE ‘READ.CSV’ FUNCTION ........................... 28
FIGURE 2.9 THE RSTUDIO IDE AFTER PLOTTING THE MAP OF FLORIDA .......................... 29
FIGURE 2.10 THE RSTUDIO IDE PLOTTING THE DISTRIBUTION OF THE NORTHERN CARDINAL IN FLORIDA 2011 .................................................................................. 30
FIGURE 3.1 EXTRACT FROM (FERREIRA ET AL., 2011, P.2377) ................................. 37
FIGURE 3.2 EXTRACT FROM (FERREIRA ET AL., 2011, P.2377) ................................. 38
FIGURE 3.3 EXTRACT FROM (FERREIRA ET AL., 2011, P.2379) ................................. 40
FIGURE 3.4 SMOOTHED POPULATION INDEX GRAPH AND ACCOMPANYING TEXT ............. 43
FIGURE 3.5 FLEDGLING NUMBERS DISTRIBUTION AND TREND ................................. 44
FIGURE 3.6 ABUNDANCE MAPS FOR BRITISH ISLES ................................................ 45
FIGURE 3.7 BREEDING CHANGE ATLAS BRITISH ISLES ............................................. 46
FIGURE 3.8 SIMPLE ATLAS REPRESENTATION OF THE CUCKOO DISTRIBUTION IN IRELAND ON THE BIRDWATCH IRELAND WEBSITE .......................................................... 48
FIGURE 3.9 MULTI DIMENSION VISUALISATION (SWEENEY ET AL., 2011) .................. 49
FIGURE 3.10 ANALYSIS OF BIRD FLIGHT PATTERNS TO AVOID BIRDSTRIKE ................. 50
FIGURE 4.1 THE AVIAN KNOWLEDGE NETWORK CYBERINFRASTRUCTURE .................. 55
FIGURE 5.1 SLIDE 1 ‘FIELD TRIP HOWTH’ ............................................................... 60
FIGURE 5.2 HOWTH FIELD TRIP (REDUCED IN SIZE FROM ORIGINAL) ................. 61
FIGURE 5.3 SECOND IMAGE: ‘THE BIRDWATCH LEAGUE TABLE’ ............................... 62
FIGURE 5.4 SECTION 2 ‘LEAGUE TABLE’ OF BIRDWATCHERS ................................. 63
FIGURE 5.5 DEMOGRAPHIC QUESTIONS ..................................................................... 64
FIGURE 5.6 BIRD WATCHING ACTIVITY ..................................................................... 65
FIGURE 5.7 SECTION 5. RESPONDENT’S USE OF TECHNOLOGY .................................. 66
FIGURE 5.8 THE INTERVIEW DOCUMENT SECTION ONE ........................................... 68
FIGURE 5.9 BIRDVIS VISUALISATION AND EXPLANATIONS ................................. 70
FIGURE 6.1 RESPONSES TO SECTION 1 ...................................................................... 73
FIGURE 6.2 ANALYSIS TO SECTION TWO RESPONSES ............................................ 74
FIGURE 6.3 ANALYSIS OF SECTION THREE – DEMOGRAPHICS (4 QUESTIONS) ............. 75
FIGURE 6.4 ANALYSIS OF SECTION FOUR BIRD WATCHING EXPERIENCE (5 QUESTIONS) ...... 77
FIGURE 6.5 ANALYSIS OF SECTION FIVE – USE OF TECHNOLOGY (4 QUESTIONS) ........ 78
1 INTRODUCTION

1.1 Overview

This dissertation is centred on the area of data visualisation and how this subject can influence and contribute in a practical way to a large community of interest, namely bird conservation, by providing motivating factors to mobilise ‘citizen science’ activities for generating ‘big data’ in Ireland and the United Kingdom as it has done elsewhere, and by providing inspiration to ornithologists and conservation groups to explore new possibilities enabled by this new data source.

Bird and wildlife conservation research relies heavily on knowledge relating to changes in geographic dispersion of species in time. There is a long tradition of collecting data on bird populations, both, by the natural science community, and the large community of volunteers and enthusiasts dating back to the early efforts of the Audubon Society in America and the Royal Society for the Protection of Birds in the British Isles. Recent years have seen an increase in the amount of data collected and available through new, cheaper remote sensing techniques and through the increased use of ‘citizen science’, harnessing the power of the internet and social media (B. L. Sullivan et al., 2009).

Our work will examine how this data is being used, particularly in the United States, to create new insights into bird populations and dispersion. We will also probe how data visualisation and related tools are being used to motivate a large scale ‘citizen science’ initiative and we will carry out a research experiment to find out what issues arise for such a project, in order for it to be initiated successfully in Ireland and the United Kingdom.

Our work will compare the use of data visualisation techniques in the United States, focussing on a case study at the Cornell Laboratory of Ornithology, with those in use in Ireland by BirdWatch Ireland and the British Trust for Ornithology. We will investigate two key facets: The use of visualisations of big data to provide insight for the science community and the use of visualisations to provide motivation and
feedback to the bird watching ‘citizen scientists’.

1.2 Background

1.2.1 Data Visualization and Wildlife conservation

The key function of Data visualization has been described as moving information from point A to point B. In exploratory visualization points A and B can be the dataset and the mind of the designer respectively, and, in explanatory visualization, points A and B can be the designer and the mind of the audience respectively (Steele & Iliinsky, 2011). Both these categories of visualization should be of interest to the wildlife conservation community. For example, exploratory visualization may be of use to the scientist looking for meaningful patterns, previously undetected, in the data. Explanatory visualization, on the other hand, may provide a novel and efficient way to explain patterns to other scientists and the wider community of enthusiasts. For data visualization, Fayyad stresses the role of data reduction and reducing complexity of data (U. M. Fayyad et al., 2002, p.5). Arguably, this function can also be brought to bear in the domain of wildlife conservation, particularly as a large amount of the data collected is geo-spatial, a form which readily lends itself to data visualization. Examples of this may be found in traditional literature, such as the species atlases found in many field guides, which show the concentration of bird populations in different regions. These maps are limited in that they provide static views at a given time and for a given level of granularity (Donald & Fuller, 1998a). Computer based visualizations have the potential to provide dynamic information such as population changes over time and can also offer some interactivity, for example, allowing the user to ‘drill down’ to higher levels of granularity. Recent developments, for example the novel visualization ‘Fat Fonts’, may provide new methods to visualize species’ population data. Overlaying species population visualizations with other data such as vegetation coverage or climate change may help to provide insight about the success or otherwise of a particular species in a given region. Other work is also being carried out in this domain, for example, visualizing flight patterns of birds to reduce bird strike incidents (Budgey, 2007) and visualization of phenological events (Marris, 2010)
1.2.2 Ornithology and Conservation

Ornithology is the zoological science concerned with the study of birds. It can trace its origins to the general expansion of biological science following the pioneering work of French writers Brisson and Buffon in the mid-18th century (Cheke, 2009). Paul Lawrence Farber describes the emergent scientific domain of Ornithology after Brisson and Buffon and ascribes, as the major factors in its development, the new abundance of data, the arrival of many new natural history museums, and the increases in quality and quantity of publications on birdlife (Davis, 1997). These early developments were based on increases in the available data from the area studies of naturalists and were fueled by European colonial activity.

1.2.3 Ornithological and Related Data

Data has always been at the heart of Ornithological science and, due to the high level of interest in birds in the general population, there has long been a valuable source of data arising from the activities of bird watching enthusiasts and nature lovers.

Recent years have seen a new surge in the amount of data available to the ornithological community. Some reasons for this are the availability of cheap remote sensing devices and the increasing importance of ‘citizen science’ initiatives, which harness internet features such as social media and the continuing enthusiasm of the public to engage in contributing their time and effort into worthwhile causes.

Studying conservation data is important for a variety of reasons. Preserving biodiversity is increasingly dependent on international policy intervention, according to Paul F Donald of the RSPB, but the lack of evidence based assessment of the effectiveness of this, puts such interventions in danger of abandonment (Donald et al., 2007). Donald has carried out research indicating a relationship between increasing agricultural intensity and a collapse of the farmyard bird population (Donald et al., 2001). The European Union has a considerably large budget allocated to farming in order to aid species preservation. The Rural Environment Protection Scheme (REPS) has paid out an average of €300 million annually in the 6 years from 2006 to 2011 in Ireland alone (Department of Agriculture, n.d.). Another important practical consideration is the understanding of the effect of agro-ecosystems on bird populations - notable cases of population explosions have caused considerable trouble to farmers in
By its nature, a large amount of wildlife data relates to the occurrence of species in their natural habitat (Ryder et al., 2010). Changes to numbers of these species, at various stages of their life cycle, can, and do occur naturally, due to many factors, such as season, food availability, predation and migration. Any significant change in species numbers is of interest to the conservation and scientific community. Monitoring these changes has always posed a significant challenge. Wildlife is mobile and generally doesn’t wish to be seen. The data can cover virtually any geospatial location, any time day or night, and the data required, can include observed behaviour such as courtship, mating, hunting and feeding etc., as well as the presence and number of a particular species.

There are comprehensive sources of data from the Avian Knowledge Network (AKN) in the United States (http://www.avianknowledge.net/content/datasets), also, in the U.K. from the British Trust for Ornithology (BTO) and the Royal Society for the Protection of Birds (RSPB), via the National Biodiversity Network (http://data.nbn.org.uk). Climate data is available from The National Climatic Data Centre (http://www.ncdc.noaa.gov) and the European Climate Assessment and Dataset (http://eca.knmi.nl). Land use data is available from the European Environmental Agency (http://www.eea.europa.eu/themes/landuse).

According to Fry, to extract meaning from data requires a combination of various and diverse fields such as statistics, data mining, graphic design and information visualization (Fry, 2008). We believe this applies as much to bird science, as any other field where large data sets have recently become available. We wish to probe news way in which this is happening.

### 1.3 Research problem

We wish to find out what new possibilities have opened up in the field of conservation and, in particular, ornithology, caused by recent developments in data analytics, particularly the growth of ‘citizen science’ and its potential to provide big data sets to the ornithological community. We are interested in a popular and new field in data
analytics, that of data visualisation. We wish to explore the understanding and usefulness of three particular forms of data visualisation. Firstly, we wish to look at the effectiveness of a very common representation of bio diversity in conservation science, geo-visualisation, for understanding of how it might impact on participants in the data collection effort. Are we adapting to this technology easily and is it useful? Secondly, we wish to discover, for the Irish bird watching community, if a simple tabular representation of a participant’s achievements and standings in their hobby, can significantly stir their interest and potentially have a motivating influence on their engagement as has been cited for their counterparts in the United States. Finally, we wish to find out about the attitude of the scientific community, specifically ornithological science, and those organisations involved in bird conservation in Ireland, towards new developments in data analytics which are being used elsewhere. We would like to test the level of awareness, among this community, of the revolution taking place in data science and the possibilities opening up in their field, as shown in the USA, due to the potential of ‘citizen science’ to provide new, big data sets. To do this we will examine their reaction to an existing data visualisation and solicit their opinion as to how useful they find it.

We will use, as a starting point, a significant development in recent years, namely BirdVis, a project of the Cornell Laboratory of Ornithology (CLO). The story surrounding the development of BirdVis is well documented and the issues relevant to our subject of data analytics, in particular data management and visualisation play a large role in this narrative (Chris Wood et al., 2011) (Caruana et al., 2006). We do not see comparable projects in the United Kingdom or Ireland and we would like to find out more about the reasons for this. We would like to find out, in the opinion of some key players in the Irish bird watching and ornithological community, what changes or initiatives would be needed to address the situation.

1.4 Intellectual challenge

The intellectual challenge in this dissertation has several components, some of which are internal to the work and some which are overarching.

For the general, overarching components, three challenges arise. Firstly, we wish to
produce a work which contributes to filling a gap in the knowledge relating to our
discipline, that of Data Analytics. This is a very broad area as is reflected by the
variety of topics covered in the syllabus, so we have a wide range of subjects to choose
from initially, and narrowing this down was one of our goals. Secondly, we wish to
relate our work to something practical and some activity carried out by a reasonably
large number of people to whom it might provide a small benefit. This would help
make the task more interesting for us and provide some motivation beyond satisfying
the academic imperative. Finally, we wish to put to use a lot of the subject matter
covered on the Master’s degree, in order that it would be consolidated into a coherent
whole that would represent a tangible body of new knowledge and skills which we
acquired over the duration of the course.

Internal to the dissertation we have at least three challenges which we hope to integrate
into a final body of work. Firstly, we try to find an efficient description of how data
science and emerging technologies are being used by a scientific community in the
USA, in cooperation with the large community of interest known as bird watchers. We
need to be able to use this description to communicate our research questions to two
distinct interest groups, namely, the hobbyists, who may provide a new valuable
source of ‘big data’, and the ornithologists, who may benefit from this new source of
data. Secondly, we need to formulate a research project which can provide new
information from which we can draw conclusions which may benefit these groups.
The methodology used needs to be efficient and low cost as we have a very limited
budget within which to operate and only one member of staff. Finally, we need to
distil our research findings and look for meaning in the data we collect thus finding the
best way to communicate these discoveries through the medium of the dissertation
paper.

1.5 Research objectives

We set out to achieve the following objectives through the dissertation and associated
research:

- We wish to learn more about the bird watching community in Ireland and how
  they have been contributing to citizen science projects in a way useful to
  ornithology and conservation. We wish to know to what level their reporting
activities are harnessed to make the best use of these large scale ‘citizen science’ techniques such as crowd sourcing.

- We would like to know how comfortable the bird watching community in Ireland are with the use of all the technologies and tools which could be used to deliver large data sets to the ornithological science community. These include internet, social networking, mobile applications such as smart phones and tablets as well as geo visualisations and geo positioning devices, or remote sensing technologies.

- We wish to know what attitude the bird watching community in Ireland have towards ‘competitive’ bird watching and reporting. Our initial research suggests that this is a highly motivational factor in the United States and results in higher participation rates. We would like to know if the same attitude prevails in Ireland.

- We wish to find out how aware the ornithological science community in Ireland are of developments in data analytical science and developments in technology which may affect their community. We would like to test for any evidence of active engagement with these technologies towards harnessing their potential. For example, do they have a strategy aimed at large scale collection of birdlife data as an input into newer, non-traditional analyses and visualisations?

1.6 Research methodology

Our research was aimed at two sides of a related community. On one side, the people who would most likely form a cohort of ‘citizen scientists’, from whom we needed to find indicators of attitudes and knowledge about their levels of skill, namely, the Irish bird watching community. On the other side, we wished to find out about levels of awareness of, and attitudes towards, new approaches in ornithological science, among the administrative and research practitioners in Irish bird science. For this reason we employed two different research methodologies, each aimed at the different sides of our research subjects as follows:
1.6.1 Survey Questionnaire

For the first group in our research, the birdwatchers, we needed as broad a representation as possible in order to provide us with reliable indicators of the attitudes of the overall population. We therefore chose to use a survey of members of the community based on a questionnaire which we developed. Given the busy time of year, and our lack of profile in the bird watching community and the limited time we would have to collect data, it was necessary to balance the target amount of information we wished to collect against the simplicity of the questionnaire and time demand on respondent’s behalf. Our survey, which was emailed to at least 150 people and was posted on the BirdWatch Ireland Facebook page, had 24 questions and could be answered in about 4 to 5 minutes. The questionnaire was accompanied by a separate ‘picture’ section, published on the Google apps server, and contained two visualisations on which some of the questions were based.

1.6.2 Structured Interviews

We wanted a different form of information from the second group in our research, the scientists and administrators of BirdWatch Ireland. This group was much smaller and was possibly more detailed in their particular subject knowledge, so we could afford to use open questions to gain as much nuanced and opinion based information as possible. With the small size of the group and to reduce the possibility of misinterpretation of questions due to the unfamiliar and novel subject matter, we decided to use face to face structured interviews to conduct this part of the research. To ensure a consistency of approach to the four interviewees and, in so far as was possible, to remove any bias from extraneous factors such as the interview location or the time of day, we asked the same questions of all the interviewees.

The interviews themselves consisted of two parts. The first part of the interview comprised 9 questions. The questions related to an extract from an article about new data analytics opportunities based on big data emanating from ‘citizen science’ in the field of ornithology. The article extract and the questions were emailed to the interviewees some days in advance of the meetings.

The second part of the interview was based on a visualisation from the BirdVis
(See below) which represented graphically a key functionality for exploring data made possible by large data sets and new methodologies for visualising such data. There were three open questions relating to this visualisation.

Although the nature of the interviews was thus structured, we did engage a conversational tone to the proceedings with some prompting and supplemental questions where it was felt necessary for clarification. A single person, the author, carried out all the interviews with one small exception. One of the participants was called away urgently at the start of her interview and offered to complete the process by writing her responses to the questions and emailing them. We do not believe that this affected the quality of the research in any significant way. The written responses were adequate. There is full transcript of these interviews in appendix A

1.7 Scope and limitations

The scope of this dissertation includes subjects from the syllabus of the Master’s degree in computer science (data analytics) as follows:

- **Data and database design for data analytics.** Part of the subject matter covers an analysis of data and database design, techniques of which are used to solve issues of data collection by diverse groups and integration of such data.
- **Data Mining and machine learning.** This is used to significant effect by one of our research subjects in the form of ‘bagged decision trees’ and the benefits flowing from it form a key driver in carrying out the research experiment.
- **Case Studies.** Our analysis of the case study of the ‘BirdVis’ project at The Cornell Laboratory of Ornithology during our initial research led to the formulation of our principle research question.
- **Problem solving and Communication.** This study is about problem solving in a community of interest using new developments in data analytics and ‘citizen science’. We have tried to practise our new lateral thinking skills in relation to the formulation and execution of our research.
- **Data management.** There were issues of data management raised in our research, particularly in the area of privacy of individuals submitting geo-location data. There are also issues of data ownership to be considered in the
borderless world of crowd sourcing.

- **Visualisation.** The role played by visualisation in both motivating the collection of data through ‘citizen science’ and, in the subsequent analysis of this data, provided us with a novel curiosity factor and enticed to follow the path we followed in our journey. We found this topic to be very interesting and embraced it in our research.

Apart from the scope of the data analytics subject matter, our dissertation involves an activity which is not only part of a science, but is a hobby for millions of people around the world. It is a hobby that is extremely well represented in Ireland, and the scientific aspect of it is highly important to humanity and life on our planet.

The project is principally concerned with how data analytics is being used more by one geographic section of a community (United States) and, not used so much by another section of the same community (Ireland/Europe). We are not developing new techniques to aid ornithologists, nor are we developing applications for use by the bird watching community. Rather we are trying to find out if there is a failure, or delay in exploiting an opportunity presented by new developments in data science which could benefit both the hobbyist and the scientist in Ireland as it seems to in the United States. We did, however, feel that it was important to cover in our research, as many areas as possible, which may have had a bearing on some of the success experienced by the American scientists. To this end we reviewed some of the literature, for example, in human interface design.

Our research, which is mostly qualitative, is limited to suggesting possible lacunae in the methods being used by our scientists. We hope to highlight areas where opportunities may exist for whoever may wish to take up the challenge of exploiting new methods. The precise level to which we can validate our findings is limited because, a validation model, suitable for this research, would be problematic with actual implementation. In this case, we hope to make our findings creditable through the exploration and presentation of the facts as we find them.
1.8 Organisation of the dissertation

Chapter 1 provides an introduction to our dissertation, a brief overview of the subject areas covered. We briefly outline our research problem, the methodology used, the design of the research and tools used to carry it out. We describe the scope and limitations of the dissertation and provide this organisation description.

Chapter 2 is the start of our literature review. It examines the subject of data visualisation. We look at its historical background, some of the recent and current ‘thought leaders’ in the field and describe their work. We give an analysis of different modalities and categorisations of data visualisations. We provide a review of some popular data visualisation tools and, in the final part of the chapter, we provide an example of how one of these tools was used to create a geographic data visualisation relevant to our work.

Chapter 3 continues our literature review and examines the current use of data visualisation specifically in the domain of ornithology, the scientific study of birds. We look at traditional data visualisation uses using examples from Ireland and the United Kingdom. We then feature a detailed examination of the BirdVis project in the Cornell laboratory of Ornithology. This case study, and how it compares with work carried out in Ireland and the United Kingdom, forms the basis of our experiment.

Chapter 4 provides an examination of data sets used in ornithology. Some of our exploration here feeds into our gap analysis between the modalities in use either side of the Atlantic. We see data science and management tackled formally in the USA but find a more traditional ‘silo’ approach in Ireland and the UK.

Chapter 5 describes our experiment from the rationale, through the design and the execution. We provide a description of the tools we used and why we used them.

Chapter 6 is our analyses and evaluation of our research findings. We summarise our findings and suggest possible implications and conclusions.’

Chapter 7 is our concluding chapter it restates our objectives in the light of having
completed the work. We summarise our work and outline some suggestions for future research.
2 DATA VISUALISATION

2.1 Introduction

This section is the start of a review of literature relative to our dissertation. Data visualisation, as a computer science subject, and as it pertains to this study, is a relatively recent emergence. It is instructive to look briefly at its historical background and how it has evolved recently in the era of ‘big data’. We will firstly, present a short historical perspective tracing the origins of visualisation by reference to key contributors in the field, through the early use of computerised data visualisation, and up to the recent and current thought leaders in the field. We will present some well-known examples of work from the field to illustrate important techniques and fundamental principles of data visualisation up to the current state of the art. We will offer some definitions of data visualisation and methods by which different visualisations may be categorised. This will provide a backdrop and framework, by which we can later assess the current use of data visualisation in the field of bird watching. We will then review a list of software tools currently used to create and develop data visualisations. This list will not be exhaustive, for, as with many software tools, one is limited only by ingenuity and imagination in terms of the wide range of possibilities available. We will finish this section with a brief summary of findings.
2.2 Historical Background

Visualisation has been applied to data at least as far back as the 18th Century arising from a variety of different motivations and by a variety of professions. The public works engineer Charles Joseph Minard (1781–1870) created informative visualisations, such as the one in figure 2.1, which depicts the fate of Napoleon’s army on its unfortunate round trip to Moscow. He also designed an early infographic depicting the relative size of regional sourcing of meat for the Paris market. We have also, the work of William Playfair (1759 – 1823), a Scottish engineer and economist, who, inspired by the earlier work of Joseph Priestly (1733 – 1804), is credited with the early development of bar charts, line graphs and pie charts. Noteworthy figures from the field of medicine in the 19th Century include; John Snow (1813 – 1858), who plotted data on a map of Soho, London in his solving of the source of a cholera outbreak in 1854. Snow is credited with founding the epidemiology branch of medicine. Also included is Florence Nightingale, ‘the Lady of the Lamp’, who, along with her famed contribution as a nurse in the Crimean war, also made a pioneering contribution to visualising statistics, by her design of the polar area diagram. In the

![Charles Joseph Minard (1781 - 1870)](image)

Figure 2.1 Minard’s *Carte Figurative*, a graphic depiction of Napoleon’s *Grand Armées* disastrous march on Moscow and subsequent rout by the Russians. The story is told with 6 variables – Size, Latitude, longitude,
20th Century Harry Beck (1903 – 1974) is credited as the designer of the iconic London Tube map, essentially an infographic which, by offering a stylised schema of the London Underground system, inspired by electronic circuit boards, greatly simplified its navigation.

2.3 Modern Era of Data Visualisation

The modern age of data visualisation started in the 1950’s, when graphic images were beginning to be produced by computers. A branch of computational science to produce images from data was recognised in 1987 in the Workshop on Visualization in Scientific Computing held in Washington D.C. in the USA. This workshop brought specialists in computer vision and graphics together with government representatives and scientists and produced a report, called Visualization in Scientific Computing, which marked a significant milestone in the emergent field. The report defines Visualisation, proposes a domain description and sets out recommendations for a funding model to support the development of the area for tool makers and tool users (Bruce H. McCormick et al., 1987). Computer visualisation was principally taken to mean scientific data visualisation - using visualisation techniques for simulation and modelling in scientific and engineering practises - up until the early 1990’s, when a new research area called ‘information visualisation’ emerged. This was to include analysis of data sets from a far wider range of disciplines such as business and finance,
as well as administration and digital media (Post et al., 2003, p.ix). Also, according to Post et al., by the late 1990’s the term data visualisation was gaining acceptance to include both scientific and data visualisation fields. Notable contributors in the modern era include, from the field of mathematics, John Tukey (1915 – 2000), who emphasised the importance of encouraging statisticians to use visual techniques as an aid in exploratory data analysis. As well as this significant contribution in the field of statistics, Tukey created the box plot (figure 2.2), a popular simple visualisation of data dispersion. The box plot provides an elegant and simple representation of a distribution which is robust and easy to compute. It has become one of the most widely used data visualisations and has had many adaptations and refinements (Mcgill et al., 1978). Tukey’s work at Bell Laboratory influenced the development of computer software packages like ‘S’ which in turn influenced ‘R’, which is widely used for statistical computations and data visualisation (Tukey was not a friend of the pie chart. He has been quoted, apocryphally, as stating ‘There is no data that can be displayed in a pie chart that cannot be displayed better in some other type of chart’. One thinks if he was to discuss this with Edward Tufte (see below) he might have been more specific and said ‘PowerPoint pie chart’.)

Edward Tufte (b. 1942) is somewhat of a polymath. He has a background as a statistician but also teaches political science and computer science, engages in philosophy and sculpture. Tufte worked for a time with John Tukey and, based on courses which he developed on statistical graphics, produced a pioneering work called The Visual Display of Quantitative Information. Tufte is meticulous about the display of data. He coined phrases such as the ‘data ink ratio’, ‘data density’ and ‘chartjunk’ to emphasise the importance of statistical noise elimination and to display trends or patterns as efficiently as possible. Tufte’s criticism of Microsoft PowerPoint is well known. He does not like the use of ‘out of the box’ graphics and claims that slideshows based on PowerPoint are often sub optimal as a means of communicating the message effectively (Tufte, 2003). In his ‘Beautiful Evidence’, by way of an in-depth critique of Josephs Minard’s Figurative map depicting Napoleon 1st’s ill-fated invasion of Russia (see figure 2.1) , Tufte elucidates on 6 principles of analytic design:

1. Show comparisons, contrasts and differences
2. Attempt to show causality, mechanism, structure and explanation
3. Show multivariate analysis
4. Integration of evidence - completely integrate words, numbers images and diagrams
5. Document sources and sponsors
6. Content counts most of all.

(Tufte, 2006, pp.123–139)

Tufte is credited with the invention of the spark line, which he describes as ‘small, intense, simple, word size graphics with typographic resolution’ (Tufte, 2006, p.8).

![Figure 2.3](image_url)

**Figure 2.3** The three sparklines on the top show the data points for 56 months of currency prices. The three on the bottom show these data ‘zoomed in’ to display the weekly data points for the last 12 months. For clarity red points mark the oldest and newest data points and blue points mark the highs and lows for the period. Source: (Tufte, 2006, p.8).

They are small scale trend lines designed to be embedded with text (Figure 2.3). Tufte is associated with the creation of the small multiple method of visualisation. This comprises a series of same size charts sharing one comparable axis as an aid to comparison of related data, or as an aid to visualisation of multivariate data. The credit for development of this technique is disputed in MacEachren who attribute it to Bertin in 1967 (MacEachren et al., 2003, p.2).
A significant figure in recent years is Hans Rosling (b. 1948). With his background in statistics and medicine and a heavy involvement in third world health, agriculture and economics, Rosling has had numerous awards for his communication and statistics expertise. His contribution to the field of data visualisation is evidenced by his co-creation of the ‘Gapfinder’ Foundation and its product. This company developed software for the interactive display of statistics. His TED talks and award winning presentation in the television documentary *The Joy of Stats* provide powerful examples of data visualisation in action. Rosling has had remarkable success in drawing attention to such things as the relationship between falling birth rates and rising affluence of countries such as China and India. His personal charisma and engaging delivery style renders the accompanying ‘commentary’ an essential part of the data story. It has been suggested also that the analyses Rosling performs throws up some very interesting insights but, this model relies, to a large extent, on the presenter itself for its attractiveness (Ridgway et al., 2012)

### 2.4 Definitions

Some definitions of data visualisation are as follows:

*‘The key function of data visualization has been described as moving information from point A to point B’* (Steele & Iliinsky, 2011, p.1). This message view harks back to the

![Diagram showing the 'designer – reader – data' schema from Steele and Iliinsky](image)
conceptual communication models such as the *Transmission Model* of Shannon and Weaver through the *SMCR and Transaction Models* of Berlo and Barlund respectively (Chandler, 1994). It also reminds us that data visualization is a method of communication. Steele and Illinsky propose using the communication model as a means to describe a categorization of visualization akin to that often used in statistical data analysis: ‘In exploratory visualization points A and B can be the dataset and the mind of the designer respectively, and, in explanatory visualization, points A and B can be the designer and the mind of the audience respectively’. Consideration of the audience and the originator of the message has a considerable bearing on the form of visualization produced but, so too has the data. Steele and Illinsky further develop this communication approach into a schema for positioning any visualization in the context of the data-reader-designer relationship combination see figure 2.4.

The rationale for data visualization as an emerging specialty seems very similar to the rationale for data mining, and that is, to a large extent, the result of the volume of data being produced in the late twentieth century and continuing to be produced in ever increasing volumes now in the early decades of the new century. This point has been made directly in relation to the field of data mining (Han & Kamber, 2001, p.1) (U. Fayyad et al., 1996). These masses of data are coming from a wide variety of sources in such diverse areas of activity as the space industry, the health industry and medical research, insurance and banking, retail sales, internet gaming, social media, and public utilities such as traffic control. Aided by the ubiquity of the internet and wireless communication, increasingly data is coming from a diverse range of ‘citizen...
science’ projects based on novel techniques such as crowd sourcing (Wiggins & Crowston, 2011) (R. Bonney et al., 2009).

2.5 Categorising Data Visualisations

2.5.1 Complexity

Noah Ilinsly and Julie Steele provide some useful methods for classification of data visualisations. Firstly, they suggest that visualisations may be assessed on their level of complexity. This complexity mainly relates to the number of dimensions involved in the visualisation, and not, for example, the size or cardinality of the data set. The more dimensions in a visualisation, the more encodings required and, therefore, the more challenging for the audience to interpret and the more difficult it is to tell the story of the data (Steele & Iliinsky, 2011, p.3). Noah Lliinsky, in an online lecture delivered to the staff at Linkedin, has suggested that it’s sometimes more practical to split the task and provide more than one visualisation, in order to avoid an overly complex visualisation (LinkedinTechTalks, 2012) (figure 2.5)

2.5.2 Exploration Vs. Explanation

Another dimension on which data visualisation may be categorised is, whether the purpose of the task at hand is exploration or explanation. Visualisations are useful as a method of trying to find patterns in data and therefore provide a useful tool in the data

![Figure 2.6 A simple graph plotting death row prisoners and executions over 35 years. Source: (The Guardian and Data Market, 2012)](image-url)
analysis phase (Steele & Iliinsky, 2011, p.7). Patterns could be interesting trends, correlations or outliers. For example nearest neighbour algorithms can find clusters in large data sets. Seeing these clusters may provide some clues as to a particular narrative in the data.

2.5.3 Infographics Vs. Data Visualisation

As we’ve seen, early data visualisations were driven by the need to communicate an idea or some information, derived from some data or statistic, to an audience for ease of comprehension. Two main aspects of this are the data, on one hand, and the picture or graphic representation, on the other.

Time spend browsing on the numerous websites devoted to data visualisation will reveal that a considerable amount of visualisations exist which display for example, patterns of traffic flow on the internet, airlines and vehicular journeys and patterns of activity on social networks and social news and bookmarking sites. These visualisations can show interesting trends and patterns and can have considerable aesthetic qualities. It is, however, sometimes difficult to ascertain if the visualisation is intended as a form of art, which some undoubtedly are, or if it has some practical or useful analytical purpose.

2.6 Tools Used For Data Visualisation

To give the reader some idea of how, and in what technical context data visualisations may be created, we provide a brief overview of the range of tools available. We will follow this synopsis with an overview of some the tools we have explored practically throughout this research. As a considerable amount of bird watching and conservation organisations are run as charities, we will focus attention mainly on open source and free software tools. There are, however, many commercial packages available and so we make mention of some. This is not intended as a comprehensive list. Rather, it is a sample from various different sources or genres. There are vast such amount of tools available to perform data visualisation. They vary greatly in terms of how they came into being, their primary audience, their focus in terms of end use, their technical design and ease of use relative to potential user goals.
Some data visualisation tools originated as add-ons to existing applications, for example, the graphing capabilities of spreadsheet and presentation software. These tools use a ‘chart typology’ as a menu of available visualisations. This has the benefit of simplicity for the user but is limited in terms of the variety of visualisations available (Heer & Shneiderman, 2012). One should bear in mind, of course, that a spreadsheet is, in itself, an important method of visualising data usually in tabular form. Other tools, such as the numerous ‘dashboard’ displays bundled as business intelligence modules within integrated enterprise resource management systems, would appear to have been market driven. In this case, there is a large selling point if the visualisation dashboard is already integrated with the larger business system. Companies like SAP and Oracle would be typical providers of this type of ‘dashboard’ offering. Indeed, SAP took ownership of Crystal Reports and now offers it, either as a standalone business intelligence system or, as a reporting and dashboard tool for SAP.

Given the nature of these commercial add-on feature visualisation tools, it’s logical to expect that the focus is on ease of use and integration with other features of the system. Part of this focus would be ‘plug and play’ readiness and ‘out of the box’ default encodings and parameters. These have received much comment and criticism from various commentators, as previously mentioned. They have been said to obscure information with needless and unsuitable legends, labels and styles. Their ease of use and attractive colours and encodings has the obvious temptation to lead analysts into superficial treatment of the information task at hand. They have even been accused of contributing to a notorious disaster (Tufte, 2003). However, considering the popularity and price of such packages, particularly in the face of free and arguably superior packages from other suppliers, one must acknowledge the fact that a substantial body of visualisations is delivered through these media. It may prove useful to adapt training courses on these packages in such a way as to optimise between their ease of use, on one hand, and delivering meaningful and efficient data visualisations on the other.

Alongside the visualisation tools available as integrated add-ons as discussed above, ‘best of breed’ data analytics and dashboarding packages are another option. Software like Pentaho, whose primary focus is data integration, and Spotfire, which has, as its
origins, a project in database exploration using dynamic visual querying in the Human Computer Interaction Lab at the University of Maryland, and whose focus is on data analytics (C. Ahlberg et al., 1992) (Christopher Ahlberg, 1996), offer analytic and visualisation packages which may be integrated to an existing separate ERP system or other database system(S) (Tibco spotfire, 2012). These packages can claim an advantage in that the primary focus of the developer is the analytic and visualisation functionality which enables them to offer deeper and broader features along with specialised consultancy. A recent such ‘pure play’ vendor in this arena is FusionCharts founded by Pallav Nadhani, a seventeen year old student from Kolkata, who turned his frustration at Excel’s limited charting functionality into a global product, used by multinationals such as Google and the Obama administration (The Economic Times, 2012). FusionCharts was originally developed in Flash and is now available in HTML5. Tableau also deserves a mention in this category as a leading player in this market. Tableau is also interesting in that its origins, an academic project at Stamford University and a proprietary visualisation language called VisQL, are in using visualisations to help write SQL statements. In this case, not only is the user able to visualise the data, but, in an all-encompassing fashion, the user employs visualisations of the data structures to explore and extract interesting information (Heer & Shneiderman, 2012) (Lazlo, 2007).

Packages which have their origins in the world of desktop publishing, web page authoring tools and graphical design, provide more visual functionality than spreadsheets, dashboards and data visualisation software, especially as their primary focus is the ‘look and feel’ of the visualisations. Examples of these include Adobe Illustrator, Indesign and Dreamweaver. Indeed some visualisation projects, developed on other platforms, are ‘finished’ using these packages because they provide far greater levels of visual refinement. The inevitable price of such refinement is a far greater need for specialist skills on behalf of the user. The markets for this type of tool appear to have a strong leaning toward people such as graphic designers and web designers, involved in visual design industries. This is evidenced by the content of the many training courses run by professional training companies for these packages. These applications are generally not amenable to be used ‘out of the box’ without a substantial amount of training and practise for the user.
Software packages designed specifically for statistics, data mining and mathematics have modules for data visualisation. For example ggplot2 is an optional package for R which is based on the work of Leland Wilkinson (ggplot2.org, 2012). Wilkinson’s opus, ‘The Grammar of Graphics’ itself lays out a detailed system proposing the development of a framework for creating visualisations in the form of a high level language. This approach involves programming knowledge but provides much more choice on the type of visualisation produced (Heer & Shneiderman, 2012) (Wilkinson, 2005). These packages have the ability to provide highly tailored images. They are, nonetheless, frequently output as quite basic plot images with the necessity to export to a ‘finishing’ product such as Adobe Illustrator as mentioned above, particularly if being incorporated into a media house style. Other examples of these packages include MATLAB and Wolfram Mathematica.

There are a variety of data visualisation tools available as libraries or functions in computer programming languages such as ‘R’ and ‘Processing’. These allow considerable flexibility with regard to how the data may be read and plotted. They come with the penalty of having to learn the language however. The outputs of ‘R’ can appear to be plain and functional. This ‘dry’ presentation may provide the type of uncluttered approach best suited to the task at hand. Alternatively, and particularly if there is an aesthetic requirement to match, for example, a marketing theme or ‘house style’, they are often amenable to be rendered subsequently in a product like Illustrator as previously mentioned.

There are other visualisation resources available on the World Wide Web which do not fit easily into the sorting described above. These often take their shape from the singular nature of the principle author. One example is ‘GapMinder’ the site created by Hans Rosling and son. Probably due to the passionate focus of its founder, GapMinder is available to use so long as you wish to create visualisations using the data provided on the GapMinder website. This data is themed mainly around the areas of world health and human development and the data providers include such organisations as OECD, World Bank, WHO and Lancet. GapMinder provides examples and resources for teachers to encourage children to develop a ‘fact based
Visual.ly is a web site providing visualisation resources mainly aimed at the infographic community. Through its ‘Create’ service visual.ly allows people to create infographics by customising templates provided on the site. The ‘Create’ service is currently aimed at visualising social media and internet usage. Visual.ly also provides consultancy and bespoke visualisations and acts as an exchange forum for practitioners (Betakit, 2012).

Google provide several tools providing data visualisation functionality and data visualisations. There are many such offerings and some examples are as follows: 

**Google Zeitgeist** provides an insight into popular Google searches and a quantitative and temporal analysis of when and to what level these searches were carried out by users. 

**Google Fusion Tables** allows users to upload spreadsheet or csv files to a Google server and perform joins and other query operations on them. The application can detect geographic co-ordinate fields in the table(s) and automatically plot the data to a map. 

**Google Chart Tools**, through the Google Visualisation API, exposes a range of chart types as JavaScript classes which enable customisation and display of charts on a webpage. These charts are rendered using HTML5 and SVG enabling cross platform browser capability including iPhone, iPad and Android mobile devices. The Google Visualisation API provides data table classes for staging table data and a data source protocol, implemented, for example, by Google Fusion Tables and Salesforce.com for database integration. A notable addition to Google Chart Tools is the ‘motion chart’. This resulted from the purchase, by Google, of the Hans Rosling ‘animated bubble chart’ concept in 2007 (Gesmann & De Castillo, 2012).

An interesting development linking some of the above technologies is the emergence of a new package for the ‘R’ programming language: ‘GoogleVis’. This provides an interface between ‘R’ and Google Chart Tools. The main feature is the ability to load ‘R’ data frames into the Google table class and, thereby eliminating the need to export data from ‘R’ into Google Fusion Tables or another database (Gesmann & De Castillo, 2012).
2.7 Practical exploration of tools

In this section we wish to explore in more detail some of the tools previously mentioned. We will describe the tool in question and give a brief demonstration of how it might be used to provide visualisations potentially useful to birdwatching.

2.7.1 ‘R’

R is a programming language aimed at statisticians and the data analytics community. It provides many classical statistics and graphical techniques such as linear and non-linear modelling and statistical tests. ‘R’ is based on the programming language ‘S’ which was developed by John Chambers and colleagues at Bell Laboratories, formerly AT&T. ‘R’ is an open source route to using ‘S’ and the source code is published under the GNU General Public Licence.

Two features of ‘R’ are cited as being particularly beneficial: Firstly, it is quite easy to produce plots which are of sufficient quality for publication, although, as stated earlier, these may be further rendered using a graphics package such as Adobe Illustrator. Secondly, R is highly extensible and, as well as the eight or so ‘packages’ delivered with the initial download, numerous additional ‘packages’ are available for download through the ‘Comprehensive R Archive Network’, aka ‘CRAN’. For example, the package ‘GGPLOT2’, created by Hadley Wickham, improves upon the standard R ‘plot’ package by implementing principles of ‘The Grammar of Graphics’ of Leland Wilkinson (Pedro, 2010). Another such package, Maptools, is described thus: ‘(a) Set of tools for manipulating and reading geographic data, in particular ESRI shape files. It includes binary access to GSHHS shoreline files. The package also provides interface wrappers for exchanging spatial objects with packages such as PBSmapping, spatstat, maps, RArcInfo, Stata tmap, WinBUGS, Mondrian, and others’ (Lewin-Koh et al., 2007). Also, a recently developed package, GoogleVis, as previously mentioned, provides for a convenient integration with Google Chart tools (Gesmann & De Castillo, 2012).

‘R’ can be linked to C and C++ programmes for computationally-intensive tasks where efficiency is critical and R objects can be manipulated directly by C programmes. ‘R’ stores its data as ‘data frames’, suited to most types of statistical analyses, and provides
tools to import to, and export from, these ‘data frames’. Although R is not suited to store very large datasets of more than a few hundred megabytes, it provides tools to integrate with most relational database management systems (R Project, 2012).

One of the most convenient aspects of using ‘R’ is the free and open source application ‘R Studio’. This provides an integrated development environment (IDE) for ‘R’ (see figure 2.7 for a sample screenshot of R Studio). Among its very useful features are:

- Syntax highlighting, code completion and smart indentation
• Source editor and ability to execute code directly
• Manage multiple working directories
• Code search and navigation
• Integrated help and documentation
• Data Viewer and workspace browser
• Searchable command history
• Plot history, zooming and pdf export
• Sweave and LaTeX to Pdf

Source: (RStudio, 2012)

RStudio which runs on Windows, Mac, Linux and on the web via a server makes the creation and modification of statistical programmes and graphics very easy from an administration perspective.

We trialled ‘R’ via RStudio using the 2011 data from the ‘focus species’ data included in the CLO eBird data set. The approach was to understand the steps required to plot the geographic distribution of the species as this is a fundamental and common representation of conservation data.
We first imported the csv file into Microsoft Access. This enabled us to easily choose a subset of the data for which to carry out some analysis. We decided to pick an abundant species, *Cardinalis cardinalis*, ‘The Northern Cardinal’. The first step was to query the ‘focus species’ file, which contains records for 38 species for the year 2011. Along with the geo-reference data and other variables such as the record ID, this file has 46 columns in total (Note: The ‘full species’ data set files have more than 1000 columns of data each due to the schema adopted by eBird which, in order to allow for ‘not present’ reporting, represents all possible North American species in each row of data. By using the ‘focus species’ data set we avoided the need to use a text parsing program to extract a species for our trial). The query selected reports which were labelled ‘Florida’ in the ‘State’ column. We also deselected all the species columns from the results set except the column, ‘*Cardinalis cardinalis*’. We exported the output of this query as a csv file called ‘cardflo11’. We now read this file into RStudio. Figure 2.8 shows the IDE having created a dataframe called ‘cardflo11’ using the ‘read.csv’ function to import the data from the cardflo11 csv file.

![Figure 2.8](image)

**Figure 2.8** The RStudio IDE after plotting the map of Florida:

A) The script extracting the map from the maptools database
B) The plot tab showing the map of Florida

from the results set except the column, ‘*Cardinalis cardinalis*’. We exported the output of this query as a csv file called ‘cardflo11’. We now read this file into RStudio. Figure 2.8 shows the IDE having created a dataframe called ‘cardflo11’ using the ‘read.csv’ function to import the data from the cardflo11 csv file.
The next step was to use the map function from the maptools package to plot a map of Florida from the maptools database. This plot can be seen in the ‘Plots’ tab in figure 2.9. It is a fairly minimalist representation based on vector graphics. There are other, ‘layered’ representations available, such as esri shape files and raster graphics, using extensions to the R project as previously mentioned. For simplicity and to expose the basic functionality in our exploration, we chose the map function and the state of Florida.

Finally, we plotted the locations of the bird sightings using the R ‘symbols’ function.

Figure 2.10 The RStudio IDE plotting the distribution of the Northern Cardinal in Florida 2011.

A) The symbols function and its arguments

This function uses, as its arguments, the ‘longitude’ and ‘latitude’ fields from the cardflo11 csv file which have been automatically created in the data frame by the ‘read.csv’ function. The ‘read.csv’ function, given no other arguments, assumes that column headers in a csv file will be used as the field names in the data frame and, therefore, can be referenced in the function. The ‘symbols’ function plots, in this case, circles, whose positions are read from the geo-coordinates in the cardflo11 dataframe, and whose sizes are relative to the count returned from the relative dataframe field.
Thus, with three lines of code, we can make a reasonable representation of the abundance of the Northern Cardinal species throughout the state of Florida in 2011. See figure 2.1 for the final result. It must be noted, of course, that this representation is a very basic visualisation of the occurrence of the Northern Cardinal in Florida throughout 2011. One can see indeed, that the species is quite abundant. However, we do not get any appreciation with this visualisation of the changes in the dispersion of the species throughout the year. To achieve this we would have to plot the locations, for example for each week (or day or month) of the year on a series of small multiples or on an animated visualisation. Both of these are achievable with R. It should also be noted that this visualisation is very far from a scientific prediction of species abundance such as that generated by BirdVis, or by traditional methodologies such as distance sampling or mark recapture. Such predictions are based on statistical methods from conservation and field ecology and machine learning techniques such as the ‘bagged decision trees’ used by BirdVis. Our focus here is much simpler and reflects a ‘wisdom of crowds’ approach. We are simply trying to reproduce, as clearly as possible, where birds have been sighted in the past, and in what numbers and at what times. Our primary audience is the bird watching community, and our goal is to provide that community with easy access to information on the reporting activity of others and to provide to that community, a place where the output of their hobby may be shared. We envisage a situation where the audience makes its own judgement as to what the likelihood is, on a given terrain or habitat, of observing a given species, or, where interesting observations are likely to be made based on the experience of others.

We found R relatively easy to learn, for this simple trial, through using examples and the hacking of existing code snippets from various tutorials. There is comprehensive documentation available through links in the RStudio application and through the CRAN website. We feel that ‘R’ is very much aimed at the statistical community on the evidence of the amount of statistical functionality available in the basic package and in the various extensions. Although people think of R as a statistics system, the developers would like to emphasise that R is an ‘environment within which statistical techniques are implemented’ (R Project, 2012). This distinction, in our view, is important as it emphasises the flexibility of ‘R’ and the wide range of possibilities it holds for implementing custom analyses as well as classical approaches. This may
have implications for novel explorations based on, for example, the newer, larger data sets becoming available through ‘citizen science’. It is very likely that conservation scientists and ornithologists will find a ready supply of classical statistical functions suited to their needs in the basic ‘R’ package, because such methods are commonly used by this community. However, they may also find extensions such as, for example, the ‘dsm’ package, which implements the ‘density surface modelling’ of line transect data, based on the work of Buckland suited to their needs for population prediction (David L. Miller et al., 2012), (Buckland et al., 2004). Also, they may find useful the ‘distance’ package, which provides for fitting detection functions to distance sampling data. Both these extension packages are examples of how different interested parties have contributed to the enhanced functionality of the ‘R’ package. If a required function or method is not already available in the CRAN archive, and if the effort is justified, it is possible, for those inclined, to create a new extension package to solve a particular need and to provide a useful artefact for the benefit of others.

Having made the above positive observations about the ‘R’ environment, particularly with regard to its accessibility through the RStudio IDE, we would also like to make some comments for those considering using it in a project for visualising data. We offered the view that it was reasonably easy to create some basic output based on sample code and by means of ‘hacking’ some tutorial code snippets. Whilst this may be true, we would remind the reader that ‘R’ is a programming language and, as such, would not be easy to learn to a professional level without considerable experience and knowledge of computer programming. As well as this, because of its origins and focus on the statistical community, it would be rather difficult to learn ‘R’, and perhaps pointless, without a very good grounding in quantitative statistical methods. Most of its functions implement statistical methods of varying levels of sophistication and complexity. The ‘R’ language is widely used and is well regarded by statisticians and throughout the academic world. It’s unlikely to be used as a complete solution for a comprehensive application such as BirdVis. Indeed, in the technical overview of BirdVis ‘R’ is given credit for the flexibility of its data handling, its statistical functionality and the fact that it is used by so many partners in the field. However, it is only a part of a larger system, which was developed in C++, which, along with OpenGL and QT was more suited to the development of the graphical user interface.
and widgets etc. This configuration makes good sense given the fact that ‘R’ is designed to able to be integrated with C++ programs using the ‘R’ extensions Rcpp and Rinside (Ferreira et al., 2011).

In summary, we recommend that careful consideration should be given to the use of ‘R’ in an application. Given that it’s open source and free, it offers a very good solution from cost of licence point of view. It is also widely used so we might assume a reasonable level of skilled personnel to be available and a responsive and skilled on-line community to offer advice. Any investment in learning the package should be worthwhile to someone who will be also involved in quantitative statistics. The combination of skills acquired and practised, in our opinion, should be in demand. Certainly the statistical knowledge would be easily transferable. The fact that ‘R’ integrates with other programs and is extensible offers the prospect of future proofing a skills investment. If the application requirement does not demand deep statistical functionality, or a wide variety of statistical techniques to be used, it might be considered overkill to use ‘R’ if a generic programming language would perform the task. In our case, where the principal requirement may be the geographic representation of data, we may find that a more suitable approach may be found in a product whose central design revolves around the display of geographical data.
3 DATA VISUALISATIONS IN ORNITHOLOGY

3.1 Introduction

In this section we wish to review the types of data visualisations currently being used in Ornithology and conservation science. The nature of data visualisation is such that many techniques may be used in many different contexts, depending on the imagination and skill of its creator, so this following list is certainly not exclusive and may, indeed be out dated in very a short time. However, by reviewing visualisations currently in use, we hope to rely on the ‘wisdom of crowds’ and find that the practitioners we review provide us with a good idea of the state of the art. A substantial element of this chapter is a review of the BirdVis project of The Cornell Laboratory of Ornithology (CLO). The reason for this focus on BirdVis is that, in our opinion, having reviewed their work, we find this project epitomises useful exploitation of ‘big data’ from the bird watching community, together with data mining and novel data visualisations in order to gain useful insight into bird populations and their distributions. Indeed, we think this work is so important, as a case study, and it contrasts so strongly with activity elsewhere, that it forms a substantial rationale for our research. We shall also examine other visualisations.

3.2 BirdVis

3.2.1 Introduction

A substantial work has been produced by a collaborative effort between scientists at the Polytechnic Institute of New York University and the Cornell Lab of Ornithology at Cornell University. The result of this work is BirdVis, described as ‘an interactive visualisation tool that supports the analysis of spatio-temporal bird distribution models’ in an introductory paper describing its design and implementation (Ferreira et al., 2011, p.2375). The tool, which is claimed to be a fully functional application, is aimed at scientists in the Ornithological field, to enable them to explore, interactively and in their own way, bird distribution data over a variety of spatio-temporal
dimensions. The creators of BirdVis set out to provide new interactive functionality which they say is the first of its kind, and is a substantial improvement on the state of the art as of 2011. They make the assertion that the study of avian data provides a very important ‘window’ into other biotic processes and, as such, is an important tool in the increasingly important study of nature conservation. The following is a review of the functionality of the BirdVis tool, with particular regard to visualisations based on a paper published by the CLO team (Ferreira et al., 2011).

3.2.2 The data

The data used in BirdVis is typical of large data sets recently becoming available through ‘citizen science’ projects. These projects leverage the power of the internet and social media to recruit participants of varying skill levels, from professional to novice, for the purpose of carrying out large scale research at low cost (R. Bonney et al., 2009). In the case of the BirdVis project, the base data comes from eBird, the biggest such survey so far in the field of ornithology. These data, collected from numerous sources since 2002, are records of bird observations, standardised in terms of protocols followed by the participants while collecting the data (Chris Wood et al., 2011). The eBird data are processed by a statistical model called STEM (Spatio-Temporal Exploratory Model), designed by ornithologists and statisticians, to compute occurrence probability for each species for any given data point in time (based on a machine learning algorithm which scales relatively sparse observations to large space and temporal dimensions). STEM also calculates a relative weighting to one of 22 predictors for each data point. These include 16 different habitat classes (e.g. mixed forest, grassland etc.) within a 1.5 kilometre squared area centred on where the observation was made. The percentage cover of each of these classes is computed and forms the basis of a predictor importance measure. Other predictors in the data are, the length of time spent during the individual study, the length of the transect, the number of observers and the elevation of the observation point. The STEM data is further processed by BirdVis to provide an estimation of occurrence probability which is quick to compute for any given three dimensional space based on time, geographic area and predictor. In total BirdVis uses 622,000 observations from 107,000 unique locations from the eBird data set. This is modelled into a single daily occurrence map per week for each of the 52 weeks in the year in question. A single occurrence map
comprises the occurrence probability of each species in 130,769 geographically stratified random selection based on 30 Km pixels. There is a computation of the relative importance of habitat by means of estimating the improvement made by each habitat predictor to the splitting criterion of an ensemble of decision trees. The paper is unclear as to the exact details of this calculation but describes it intuitively as, the strength of association between each habitat and the predicted occurrence/absence of a given species.

3.2.3 The Goals of BirdVis

The Ferreira paper describing BirdVis explicitly sets out the goals of the project. The work was aimed at scientists (ornithologists, biologists and statisticians) involved in the generation, validation and interpretation of bird distribution models. Its aims were to help these scientist better understand the model predictions, tune the models and to communicate their results.

The scientists were themselves polled for their desiderata. These are summarised in the paper as follows:

- Identify and understand migration patterns. Besides visualizing bird movement, they need to explore the timing, direction, speed and duration of these movements.
- Validate the hypotheses that STEM predictor importance reflects actual species habitat preferences.
- Compare the behaviour of different species, both regarding migration patterns and habitat use.
- Perform statistical analyses to validate and improve the prediction model.

3.2.4 Challenges

The principle challenge faced by the BirdVis designers, as stated in the paper, was the multivariate aspect of the inquiries. They cite the following example: ‘Carrying out these tasks is challenging due to the multivariate aspect of the data: multiple dimensions (species, space and time) associated with multiple signals being studied
(occurrence probability, predictor importance). For example, to validate hypotheses regarding habitat associations for a species based on habitat predictor importance signals, it is necessary to compare 16 variables, each of which depends on space and time.

Visualisation is not new to the domain of bird conservation and the scientists involved make use of the state of the art which includes static plots, map views and animations generated with the statistical package ‘R’. The traditional form of map views are familiar as they appear in many bird guides and comprise of a geographic map (usually country or land mass specific to the book’s national market e.g. ‘Farm Birds of the British Isles’) per species with the occurrence of that species described by a coloured overlay on the map. There may be two versions of the map describing different seasonal distribution or a version describing the locations where breeding takes place.

According to BirdVis research, even to carry out investigation of a single species, may require a data modeller to generate several plots and maps. The ornithologist has to review these representations sequentially and try to retain the information to postulate the existence of patterns or relationships. This is very demanding and to most people represents a form of information overload.

In general, it is difficult to find a single colormap that is both perceptually good to display bird distribution and that represents well the range of the values in different times. Here, we show occurrence probability maps for Baltimore Oriole in May 25 and September 7 of 2009 using two configurations. The top row uses a continuous sequential colormap from probability 0 to 0.7. While this highlights important features in May 25, the details in September 7 are not clear. The bottom row uses the same colormap but with a range between 0 and 0.1, and dark red is used for values greater than 0.1. This reveals details in September 7, but it is not as effective for May 25.

Figure 3.1 Extract from (Ferreira et al., 2011, p.2377)
These challenges informed the design of the visualisations in the BirdVis application. The designers make the point that their tool enables the scientists choose a variety of parameters and to look at many dimensions at once.

### 3.2.5 BirdVis Solution

BirdVis employed a number of features to their tool set to overcome a variety of challenges in their initial aspirations. The first of these involves an interactive feature to control the parameters of the colormaps as follows: The generality of the visualizations were based on the concept of bird distribution maps. These are very familiar to those, both professionally engaged in ornithology and amateur ‘bird watchers’. In the case of the BirdVis target users, the scientists, these maps are in the form of distribution probability maps. The probability is expressed as a number $>0$, $< 1$. For some species/area/time combinations the range of actual values may differ. This makes it difficult to have a single colour map gradient. Figure 3.2 is an extract from the paper that describes the problem. BirdVis solve this by allowing the users decide the graduation of the colour map for different probability ranges. This is achieved by means of an *interactive colormap controller* which, as can be seen in figure 3.1, serves both as a controller and as a legend for the visualisation.

![Occurrence variation maps](image)

The second feature acts as an aid to visualising bird movement. It can sometimes be difficult to assess the movement patterns of species by looking at a series of colormaps...
displaying differences over time, so the BirdVis team created a set of complimentary occurrence probability differences. See figure 3.2 for their explanation of this in operation.

The third main feature of the BirdVis tool is, according to the authors, a key new technique which they are calling ‘Tag Cloud Lenses’. This innovative feature is a combination of the familiar tag clouds, which we find in many visualisations and, ‘magic lenses’, a visualisation technique introduced in 1993 by Bier et al. which uses a user defined semi-transparent window over a spatial area in order to select the region underneath for attention (Bier et al., 1993). In the BirdVis implementation these concepts are combined (for the first time as far as was known to the authors) and they have adopted the name ‘Tag Cloud Lenses’ for their innovation. The idea is that the user opens a ‘lens’ over a particular geographic area. Inside this lens the user will see the relative importance of the previously mentioned probability indicators, by means of a ‘word cloud’ in the area covered by the lens. The more relatively important the indicator is, the larger is its tag word. This suits the nature of the data, as bird distribution is not precisely mapped onto a particular habitat, but rather, shows ‘fuzzy preferences’

The lenses can be moved and/or resized. The tag cloud will dynamically change to represent relative importance of the indicator covered by the lens area. This is an excellent feature and allows, among other things, intuitive multi-dimensional analysis for easy comparison of:

- Same species for different geographic areas
- Same species for different temporal windows
- Different species for same geographic areas
A visual example of the ‘Tag Cloud Lens’ in one of its operating modes (multiple windows for different species) is provided in figure 3.3

3.2.6 Comments on the BirdVis Case Study

The BirdVis project provides a very good case study for how data visualization can be applied in a domain such as ornithology. It harnesses the new availability of the massive data set collected by eBird and the National Audubon Society and it uses both traditional and familiar methods of visual encoding (geo spatial maps, colormaps, line charts), as well as state of the art and novel techniques such as the colormap widget and the tag cloud lenses.

The project needed the cooperation and input from a number of disciplines including personnel with statistical, ornithological, computer programming and graphics expertise. This bears out the opinion of Fry, for example, that data visualisation requires a combination of many disciplines (Fry, 2008, p.5). This suggests to me that one of the key skills of the data visualisation practitioner might also include teamwork.

The project also seems to reflect, in so far as one can glean from the paper, Fry’s suggestion that visualisation is a process and that process starts with a question. In this case, the project team consulted in depth with the potential users to find out, in the first
instance, what their requirements were.

There was then a substantial treatment of the data, which involved its collection, its subsequent processing by the ornithologist statisticians using STEM, a data mining activity, and pre-processing by the visualisation team to generate probability estimates to enable the high speed interactive queries necessitated by the tag lens functionality.

We can assume that the data was filtered because the eBird data set contains more data than those used in the tool.

There are several representations of the data which would seem very suitable to the task at hand (the team reports that the tool is currently in use in the Cornell Laboratory for Ornithology). Some of these are quite basic, for example the bird probability trajectories at the foot of figure ?. It would seem that an effort was made to keep the representations simple and this is as it should be. There is a need to use only what is necessary to communicate a message. Anything else might be considered noise.

When we look at the visualisations, it’s clear that a lot of consideration was given to the quality of the display. In a situation where the target audience are skilled professionals who could be considered sufficiently motivated to make use of the tool, one might expect a lower quality ‘look and feel’ of the application. However, the introduction of the colormap widget and the ‘Tag Cloud Lens’ provides desired and meaningful functionality. This refinement is more than adequate justification for the visualisation techniques employed. Visualisation, as stated by Iliinsky and others, harnesses the immense bandwidth of the visual system to move a large amount of information to the brain very quickly and takes advantage of the brain’s ‘built in’ software to identify patterns and communicate relationships and meaning (Steele & Iliinsky, 2011, p.1).

The final part of the process, as defined by Fry, is to provide for interaction with the visualisation. This was achieved very well with the aforementioned colormap widget and Tag Cloud Lenses. It is also achieved by allowing the user look at multiple occurrence probability maps.
Fry does not explicitly recognise deployment and test as part of the process although he does discuss the iterative nature of same (Fry, 2008, p.9). In the case of BirdVis the paper describes two case studies where the scientists were able to confirm the tool’s ability to model a previously known hypothesis and, also the tool’s ability to help them identify new hypotheses.

Some limitations of the BirdVis tool are acknowledged by the team, namely, a known limitation of tag clouds for users to assign undue size to longer words. Also, the lens shape was limited to rectangles. A significant improvement would be some method to track trends and this was also acknowledged by the team by stating it as a likely future direction of research.

3.3 Irish United Kingdom Bird Organisation Visualisations

3.3.1 British Trust For Ornithology

On the BTO website in the ‘Trends’ link we can find examples of data visualisations which are simple and suit simple explanatory purposes. Figure 3.4 is one example in particular. There is no lack of merit in the fact of its simplicity. In his section on ‘principles’, Fry warns us to: ‘Avoid the All You Can Eat Buffet’. He cites the simple weather map and Beck’s iconic subway map as examples where noise is stripped out of the data and only the relevant information is displayed. The data visualisation in figure 3.4 simply shows the trend in the Mistle Thrush population in the UK between 1966 and 2010. There is an accompanying information icon which, when the mouse is held over it, tells us that it’s a smoothed index with an arbitrary base of 100 in 2009 and that the green lines represent 85% confidence intervals. It is a time series chart with two dimensions, time and population index. It’s very obviously telling us news about a decline in population of the Mistle Thrush. There is some accompanying explanatory information on the web page as follows in figure 3.4:
“Like those of Song Thrush and Blackbird, Mistle Thrush populations have declined significantly since the mid 1970s, especially on farmland. The species was recently moved from the green to the amber list because of population decline, and recent BBS data suggest that this decline is continuing. The Scottish BBS trend, in contrast to those elsewhere in the UK, is of strong increase since the late 1990s. There has been a minor increase in clutch size and perhaps in overall productivity; population decline is thus likely to have been driven by reduced annual survival (Siriwardena et al. 1998). Numbers have shown moderate decline across Europe since 1980 (PECBMS 2011a)”.

Figure 3.4 Smoothed population index graph and accompanying text. Source: (Baillie D.I. et al., 2012)

Given this additional commentary we might suggest that the visualisation could be made more complex and to carry more information. Compared with the richness of the BirdVis tool earlier we might, for example, expect some recognition of the fact that farmland populations are declining, or, perhaps some indication of a breakdown in the geographic figures, to see if there are any local outliers or reversal of the overall trend.

In the context of the website, the fact that this is a report aimed at the public in general, and the way website is structured, which allows the user to see this level of detail for a number of different threatened species (the Mistle Thrush is on ‘amber’ alert), we can see that this visualisation plays only a supporting role in an overall multi-faceted message. In this way, its resembles a weather forecast, where we only see a schema of the region with temperatures and wind direction indicators, accompanied by a narrative
to enhance and refine the visual message being displayed.

The chart in figure 3.5 is similar to that in figure 3.4. In this case, however, we find additional detail, in that the sample distributions are shown for each year. One might be justified in asking if there is any merit at all in this as the graduations on the y axis make it difficult to read precisely where the distributions are placed. One would think, in this case, a user might prefer a higher resolution graphic and/or a grid to support more accurate reading of the data. On the other hand, a distant examination of the chart does convey a sense of the year to year variability of the data and the variability of the samples in any given year. The visualisation has merit depending on what the user is looking for. These charts are explanatory and it’s clear that the user is expected to access the supporting information, provided by links on the website, for a fuller picture. A question may arise if newer techniques could be brought to bear to provide more information in the visualisation using the higher resolutions now available on most operating systems.

3.3.2 Bird atlases

Donald and Fuller give seven different reasons for collecting bird atlas data. These are:

- Education and recreation
• Conservation and research
• Documentation of changes to population and range
• Framework for survey design
• Assessing bird/environment associations
• Hypothesis generation for causes of rapid change
• Investigation of theoretical aspects of ecology (Donald & Fuller, 1998b, p.131).

They also suggest that, of these reasons, which are not comprehensive, the first three are primarily for descriptive purposes and the remaining four are primarily for analysis.

The BTO provides many visualisations in the form of Atlases. An example of these may be seen in figure 3.6. These are very simple geo-spatial representations of where the bird has been seen or has been observed breeding throughout the British Isles. There is a function on the website to display visualisations from two different survey periods side by side. This allows the user to see differences in the geographic distributions between the two periods.

The atlases simply describe the condition, ‘bird present yes or no’, and whether the

![Figure 3.6 Abundance maps for British Isles](image_url)
bird was known to be breeding for the geographic area. There is nothing to indicate the level of abundance of the species, either for the map as a whole, or for any particular region on the map. For abundance and population changes other maps are available. See figure 3.7

These visualisations have no interactive features. For example, the user cannot ‘zoom in’ to a higher resolution. The abundance map in the winter atlas in figure 3.17 shows three levels of abundance. This is noted, on the website, as a decision made after the surveys were completed (BTO, 2012). We may theorise that the data and the nature of the surveys play a large part in the level of complexity available to the data visualizer.

An interesting point is that, we are told on the website that the printed versions of these atlases contain a higher level of detail. The general level of image resolution on PC and mobile devices nowadays may well offer sufficient clarity to afford much more detail.

As far as any motivational slant on the BTO website’s visualisations is concerned, or, where they might stand in relation to Iliinsky and Steele’s ‘Designer-Reader-Data Trinity (Steele & Iliinsky, 2011, p.9) they appear as simple informative visualisations. There is no exaggeration or emphasis on bird population success or decline. Both are reported with equal dispassion. This has the advantage of removing noise from the data and making the visualisation more credible. It also lessens the possibility of any

Figure 3.7 Breeding Change Atlas British Isles
bias, for example, toward one species above another. The visualisations are simply trying to express the world as it is. On the other hand, it might be considered a good idea to add some ‘persuasive’ elements to the encoding. For example, the alert system employed by the BTO and The RSPB incorporates a colour coding for certain levels of population endangerment. This alert system is as follows. ‘Red listed’ species show the most decline and to qualify, among other things, the species population needs to show at least a 50% decline in the last 25 years or show at least a 50% contraction in its breeding range over the same period. ‘amber listed’ species have more protracted qualifying rules but basically have to show ‘moderate’ declines (25 to 49 %) over the 25 year period. Finally ‘green listed’ species are basically those which are not ‘Amber’ or ‘Red’ i.e. not in danger (RSPB, 2012). There are rigorous statistics to assess these declines (Baillie & Rehfisch, 2006). We might consider a visualisation which shows, in a simple way, bird species in or on a trajectory into those Alert zones.

3.4 BirdWatch Ireland

BirdWatch Ireland is the largest independent conservation organisation in Ireland. It is a registered charity and its mission is stated as follows:

*The primary objective of BirdWatch Ireland is the protection of wild birds and their habitats in Ireland through the efforts of our staff, members and volunteers alike. In order to fulfil this objective, we carry out extensive research and survey work, operate applied conservation projects and manage a network of reserves nationwide; we prepare and advocate policies; we recruit, retain and service a growing membership base; we promote the importance of this work through field education, dedicated media (such as Wings magazine, the BirdWatch Ireland website, our annual journal Irish Birds and other publications) and to the national media.*

(BirdWatch Ireland, 2012).

BirdWatch Ireland cooperates extensively with the British Trust for Ornithology and the online survey data is collected and shared with the BTO. The same geo visualisations are used for atlas information through the ‘bird atlas’ link on the on-line survey page of the BirdWatch Ireland website and volunteers are encouraged to use the
‘BirdTrack’ application, the same as for the British Volunteers. The point we are making here is that, with regard to data visualisation and citizen science, the activities in Ireland and the United Kingdom can be considered almost identical and can be contrasted as one unit against the work in the United States. BirdWatch Ireland provide simple distribution maps for Irish species and one example of these is shown in figure 3.8

**Figure 3.8** Simple atlas representation of the cuckoo distribution in Ireland on the Birdwatch Ireland website
3.5 Other Visualisations in Ornithology

Data visualisations in the domain of conservation and ornithology are not confined to atlas data and we found numerous other types of data visualisations in the literature. To give some idea of the kind of data being represented we provide an example from Sweeney et Al describing the mix of common species found in certain forest types (figure 3.9).

![Graph showing percentage composition of breeding bird assemblages in different forest types.](image)

**Figure 3.9 Multi dimension Visualisation (Sweeney et al., 2011)**

Gregory et al point to a lack of headline data visualisations in the realm of biodiversity such as those used in the field of economics and financial markets. They developed new indices for using birds as indicators of biodiversity in response to this shortcoming (Gregory et al., 2003).

A very interesting and novel combination of technologies relating to data visualisation of birds, and with potential lifesaving functionality is shown in figure 3.10. This was taken from an article on the MathWorks website extolling the virtues of the MATLAB
tool for performing data visualisations. Bird echoes were recorded in the vicinity of the Spurn bird observatory by the The Birdstrike Avoidance Team at the Central Science Laboratory, using radar technology specifically tuned for bird detection.

Figure 3.10 Analysis of bird flight patterns to avoid birdstrike.

Source: (Richard Budgey, 2007)

These echo signals are processed to produce shape files which can be represented on a map. The bird flight patterns are studied and reported to airport and military aviation centres to assist in avoiding ‘birdstrike’ incidents. (Richard Budgey, 2007)
3.6 Gap Analysis between US and Ireland/UK

We conclude this section by drawing the reader’s attention to the difference between the BirdVis project and the uses of geo visualisations in Ireland and the United Kingdom. We feel this represents a very distinct gap in terms of how data analytics and data visualisation are being exploited on either side of the Atlantic. We feel that the BirdVis project is using a new type of ‘big data’, collected by ‘citizen scientists’ over a huge geographical range. It is an example of a large scale process which seems to have been designed from the ground up to deliver value in terms of new insights and possible new discoveries in the data available. The British/Irish visualisations are based on traditional research methods and improvements may be limited to cosmetic factors such as improvements in resolution or encoding.

When we requested data from the British Trust for Ornithology and from BirdWatch Ireland, both bodies were extremely kind and helpful. However, they could only supply us with data subsets based on specific species and in formats which were difficult to use. This, we think, is indicative of a data management policy which is not designed around sharing data. We had been told, for example, by one of the staff at BirdWatch Ireland that the data were stored in ad-hoc databases and it was not a simple matter to retrieve it. We wish to state that this is, in no way, a negative criticism of the organisation. Their data policy may be suitable for the current requirement. When we requested data from the Cornell Laboratory of Ornithology from which BirdVis was developed, we were directed to the Avian Knowledge Network. This allows access to a wide range of data in a standardised format.

So, although BirdVis is a very impressive data visualisation application, we think it is only made possible by the breadth of the data management and integration effort. The BTO and Birdwatch Ireland are excellent organisations and are very productive. However, we are of the opinion that data policy would be a very important area for discussion if anything like the facility at Cornell was to be considered. Once the data is in place, the production of new visualisations could be made open to a large number of professionals and volunteers.
4 DATA SETS

4.1 Introduction
The collection of large Ornithology data can be traced back to the origins of organisations engaged in bird conservation. The main sources of data used in ornithology come from focused studies carried out by ornithologists alone or in tandem with volunteer researchers. Collecting such data is highly labour intensive and, without the aid of these volunteers, could be a prohibitively expensive operation (Bibby, 2003). This necessitates very careful planning of the collecting effort and the site in order that the maximum benefit is accrued for the resources spent on the study (Dunn & Canadian Wildlife Service, 2006). The data collected are specific to the site and usually specific to the particular study being carried out. This may have an effect, whereby the data is stored in an ad-hoc fashion depending on the preferences of the scientists involved. Here, we first briefly examine categories of ornithological data and then we name some of the main sources of this data. We limit this to Ireland, the U.K and the U.S.A. but we mention one pan-European organisation of note.

4.2 Categories of Ornithological data
Bibby, in his analyses of the subjects covered in the Journal Bird Study over a 50 year period, provides some useful hints for categories by which Ornithological data are collected. He describes contributions under following broad areas:

- Numbers, Distributions and Trends
- Population Dynamics
- Migration and Movement
- Behaviour
- Ecology
(Bibby, 2003, pp.194–210)

Our main interest for this dissertation is the first category, ‘numbers, distributions and trends’. It is worth noting that this subject and its data are largely affected by the other
areas in the list. For example, a change in agricultural activity in a particular area may have an effect on a given bird species population by reducing or increasing its numbers. Observing and analysing such changes, may provide vital information to conservationists and those concerned with environmental change. In an interesting section on methods Bibby describes the increasing acceptance of the importance of statistical methods to derive inferences from collected data (Bibby, 2003, p.201).

The last 50 years have seen a big improvement in knowledge of numbers, distributions and trends, moving from qualitative for most species to quantitative data for all (Bibby 2003, p.196). Atlas surveys provide the majority of the data in the data sets. Bird data sets tend to be large. As far back as 1952, in the UK for example, a survey of 30 species at a resolution of 50 Km squares yielded 27,900 records. These were actually transferred to punch cards, in duplicate for error checking (Bibby, 2003, p.196).

4.3 Sources of Data

4.3.1 Ireland and UK

Organisations in Ireland and the U.K. cooperate extensively for the purpose of collecting bird data. The following is a selection of the main surveys:

- Common Bird Census (CBC). This was a long running mapping survey which ran from 1961 to 2000 and has been superseded by the Breeding Bird Survey.
- Breeding Bird Survey (BBS). This was launched in 1994 in order to overcome some of the weakness of the CBC.
- Countryside Bird Survey (CBS). This is similar to the BBS and takes place in Ireland.
- Waterways Bird Survey (WBS). This was carried out from between 1974 to 2007. It is similar to the CBC but for riparian (rivers and river banks) species.
- Waterways Breeding Bird Survey (WBBS). This replaced the WBS in a similar way to BBS replacing CBC.
- Wetlands Birds Survey (WeBS). This monitors the populations of non-
breeding water birds in the UK.

- **Irish Wetlands (i-WeBS).** Same as above but for Ireland.
- **Bird Atlases (1976, 1986, 1994).** These are compiled using the input of many thousands of volunteers over periods of 10 to 15 years and contain the best estimates of distributions of all species in Ireland and the UK. They are the nearest approximation to a high resolution geo spatial data set.
- **Rare Bird Breeding Panel (RBBP).** This collects data on rare breeding birds
- **Sea Bird Surveys.** These periodically survey all seabird breeding sites.

As well as the above there are numerous periodic surveys of individual species throughout Ireland and the UK. Sources: BTO.org and BirdWatchIreland.ie.

4.3.2 **U.S.A.**

- **Christmas Bird Count.** This is reputed to be the oldest running citizen science project in the world. It began on Christmas day 1900 and it takes place every year, from mid-December to early January all over the U.S.A. and involves tens of thousands of volunteers. Source; Audubon.org
- **Breeding Bird Survey.** The BBS is a cooperative program for gathering and analysing quantitative information on populations of breeding birds in North America north of Mexico. The data is entered on-line and is made available to the public on-line (D. Ziolkowski, Jr et al., 2010)
- **Avian Knowledge Network.** This represents a collaboration between academic government and non-government organisations with the purpose of providing access to bird observation data from all participating data providers. See the section below ‘standardising of ornithological data’ for further discussion.

4.3.3 **Continental Europe**

There are various organisations in the countries of the European Union and Continental Europe which collect data on bird species. The UK, Ireland, Sweden and Holland are regarded as leading countries for birdwatching (See interview transcripts in appendix A). We have discussed Ireland and the UK and other individual European countries are beyond the scope of this dissertation. However, we briefly mention the
EBCC below simply to provide an example of a trans-European organisation which can be compared and contrasted with the pan American organisations.

- The European Bird Census Council (EBCC). This Dutch organisation has among its goals to
  - Encourage communication and collaboration between organisations, institutions and individuals interested in monitoring bird populations and their distribution
  - Encourage monitoring of bird populations and their distributions aimed at better conservation and management of bird populations.

### 4.4 Standardisation of ornithological data

The Avian Knowledge Network (AVN) in the USA, in our view leads the way with
regard to data management of ornithological data. It provides a data federation facility to integrate bird observation data from a wide range of sources. At the centre of this federation, it utilises a standard called ‘Bird Monitoring Data Exchange Schema’ (BMDE). This is an extension of the Darwin Core, a body of standards for exchanging data on biological organisms, their habitats and their spatio-temporal occurrence. Figure 4.1 provides a schema of the AVN Cyber infrastructure (M. Iliff et al., 2008).

4.5 eBird Data Set

By far the largest data set contributing to the AVN, with over 63 million observations is eBird and, to a large extent, the enthusiastic and pioneering approach taken by the people behind eBird (Audubon society and Cornell Laboratory of Ornithology) has inspired this research. We were impressed at how they were using ‘citizen science’ and data science to develop new methods of experimentation and knowledge discovery in ornithology. In particular, the philosophy and techniques employed to engage the bird watching community seems to have provided spectacular results in this regard. (Chris Wood et al., 2011). This gave rise to our curiosity about similar approaches on this side of the Atlantic. The eBird data set is used extensively in the BirdVis Project discussed in chapter 3.

4.6 Conclusion

The range and diversity of organisations involved in bird conservation is enormous. This is beneficial, in that there is increasing need for scientists to monitor biodiversity and to research into causes of change. There is an opportunity to increase the scope of this research through the increased use of data science, particularly through the use of large spatio-temporal data sets. We think eBird, as a data collection methodology and the AVN, as a data federation, combine to be an excellent model for this new research opportunity. We would like to explore some aspects of the success of eBird, and, with regard to Ireland and the UK to see what elements might prove helpful to motivating a similar model here.
5 EXPERIMENT

5.1 Introduction

In this chapter we describe our experiment. In the first section we elucidate the rationale of the experiment and summarise the main questions we wished to ask. We then describe the various methodologies we used to ask our questions. We explain the approach taken and the tools we used for each approach. We then describe the actual experiment and any difficulties or limitations we incurred.

5.2 Rationale

When we carried out the literature review into data visualisation in the domain of bird watching we found that the novelty of data visualisation as part of data analytics was certainly beginning to be represented in new publications and papers as one would expect in line with a general rise in interest in big data and visualisation. We also found a variety of data representations, a large part of which were rooted in the traditional statistics which are used in field ecology. In this domain regression plots and box plots are commonly found as the results of focused studies of selected species in particular localities. With regard to more modern visualisations and those related to data analytics we found that most seem to be coming from the United States. We found that ‘citizen science’ projects seem to be driving an increase in the availability of widespread presence and absence species data.

We found that, the data in the UK and Ireland seemed to be confined to ‘silos’ contained in the various organisations with responsibility for bird science and conservation. In this regard we found that organisations had data which were collected for specific studies and were not amenable particularly for sharing. In particular we found a difference between the enthusiasm and recognition in the Cornell Laboratory for Ornithology for large geographically dispersed data sets being accumulated through the use of data standards and federation and through the use of ‘people as sensors’. We were particularly drawn to an observation by the Cornell team that, when metrics were fed back to the participants in the form of ‘rankings’ among bird
watchers, the participation of the bird watching community increased dramatically (B. L. Sullivan et al., 2009). We wished to see if there was any evidence if this would be the case in Ireland.

We also noticed that, on the application for reporting bird sightings a considerable amount of representation is made by means of geo-located data. This form of representation is becoming very prevalent in tandem with the popularity and massive take up in the use of smart phone and 3G and 4G communications technology and social media. We wished to find out about the level of awareness of this technology among bird watchers - if they easily understand this common representation of bird sight reporting and how they felt about certain issues regarding its potential to increase their enjoyment and knowledge of their hobby. We also wished to see if there were any subgroups among the population who displayed more or less interest in the concepts.

As well as finding out more about the reactions of bird watchers and their potential to be providers of data as ‘human sensors’, we wished also to learn more about how the conservation organisation personnel felt about the potential for using larger data sets coming from the bird watching community. We wished to know did they think such data would be useful or did they have any suggestions as to how the data might be used in the future. We also wished to find out what problems might be encountered by collecting such data.

The Cornell Laboratory for Ornithology people have been enthusiastic about the potential for ‘citizen science’ projects to provide them with new tools and data with which to study bird dispersion and abundance as well as predictors such as habitat. They cite as an example the use of ‘tag cloud lenses’ as a useful visualisation technique, both for the testing of hypotheses and for the formation of new hypotheses. We wished to gauge the reaction of their Irish counterparts to these tools. We wished to see if they easily understood the nature of the tools and if they thought that these tools would be useful in Ireland and the British Isles.

To summarise, our research is in two broad areas which are divided between the people who would be providers of data through their hobby, and those who may be
able to make use of the new type of data available. Our hypotheses are that:

a) People would like to get involved in ‘citizen science’ by providing data from their field trips and that measuring themselves against others, in a similar way to the eBird project, would encourage this involvement.

b) The people involved on a professional level as conservationists and ornithologists would respond enthusiastically to the prospect of having new and different amounts of data on which to perform analyses.

The next section explains how we designed two experiments to help provide answers to those two broad questions.

5.3 The Questionnaire

5.3.1 Overview

We decided to approach hypothesis a) by means of a questionnaire. This was because we were fairly confident of getting responses from at least fifty randomly selected individuals who had an interest in bird watching. This was on the basis of the very kind assistance of members of the committee of the Tolka branch of BirdWatch Ireland and by offers of help from the head of BirdWatch Ireland. Although, we would not have access to the full membership of BirdWatch Ireland, which would have given us some potentially interesting quantitative insights, we received email addresses and permission to contact the addressees through the ‘networking’ efforts of the people above. Ireland’s National Education and Research Network (HEAnet) run a mailing list service for member institutions. There is a network of bird watchers who communicate through a mailing list on the HEAnet list server. We were advised of it being a reliable source of bird watchers and we used the mailing list service to solicit responses to our survey. This tactic proved successful and we managed to get 56 responses from people that, we have a high level of confidence, are interested in bird watching as a hobby.

We wished to achieve a balance in our questionnaire between the amount of time it would take the respondents to complete it, the complexity of the questions and quality of the information we would receive from it. This was achieved by testing the
questions using a select few people who would not be solicited later as part of the survey. We also solicited opinions from one or two people who ran surveys previously. December is a very busy time for most people and we did not wish to unduly impinge on their time. We did not wish to have people return incomplete questionnaires so we made most of the questions mandatory. We tried to limit the time it would take to complete a full response to five minutes and we tested this using our select group.

The questionnaire was composed of five main sections. The following is a description of these sections, the questions we asked and the reasons we asked them:

5.3.2 Section One (4 Questions)

The first section depended on the respondents viewing a geographical representation of

![Figure 5.1 Slide 1 'Field Trip Howth'](image)

A) Authentic caption B) Blue line indicating track, yellow points indicating sightings C) Clue for yellow point meaning. Note: the letters are not part of the visualisation
a fictitious bird watching session recorded on a ‘Google Earth’ map visualisation and displayed as a presentation on-line. The picture is replicated in figure 5.1 and is typical of the ‘trips’ and ‘places’ functionality available on Google Earth. The user is asked to view the geo-visualisation as displayed in figure 5.1 and is then asked to respond to a series of questions on section one of the on-line questionnaire. These questions are displayed in figure 5.2 (resized to a smaller size here to show all questions). The first question tests the respondent’s knowledge of the Google Earth interface and thus tries to establish the level of familiarity with this type of display. This would serve as an indicator of the ease of adoption of this type of visualisation if used, for example in an application aimed at the Irish bird watching community. The next question is a simple attempt to test the respondent’s ease of comprehension of the content and message contained within the picture. This could serve as an indicator, for example, of how much information needs to be displayed at a given map resolution. The third question solicits the respondents directly for their opinion of the visualisation. We wished to know if they were positively or negatively disposed towards the image. A strong opinion either way would warrant further investigation and or consideration of alternative styles of geo visualisations. Finally, we asked the

Figure 5.2 Howth Field Trip (Reduced in size from original)
respondents if they have already used this style of geo-visualisation for bird watching or otherwise. Similar to question one, this would be an indication of the ease with which such a representation would be adopted.

5.3.3 Section 2 (7 Questions)

For the second section of our questionnaire the respondents were asked to review a second image and answer 7 questions about the visualisation. This image was an approximate reproduction of a key element of the eBird website, which has a substantial level of interest in the United States as previously discussed. The image displayed is a visualisation in the form of a table displaying the standing of a fictitious

![Birdwatch League Table](image)

**Figure 5.3** Second image: ‘The Birdwatch League Table’

This is a mock up web page/app page displaying statistics and standings for a fictitious member ‘Pat Murphy’.

bird watcher and some key statistics regarding his bird sighting activity. Some slight changes were made for an Irish audience. Figure 5.3 shows the on-line slide containing the picture. The questions we asked were as follows and can be seen in figure 5.4. The first three questions test the respondent's ability to easily read the information displayed in the table. If the respondents score well here, we would regard that as an indication of the efficiency of the table for delivering the information and
would reinforce the advantages of this type of visualisation. The fourth question, ‘What is the best measurement for a bird watchers league?’ is to ascertain which aspect of bird watching the community itself finds most important. The next question asks the respondents directly to rate the ‘league table’. We were interested if this rating would differ to that in the United States.

We then asked if the respondents would like to have their records held on-line. The people at The Cornell Laboratory of Ornithology expressed the opinion that the provision of such a service free of charge is a reward which encourages uptake of the reporting activity of the birder watchers. We thought that we could get an indication of whether this would hold true for our Irish based respondents. Finally, we asked the respondents about their level of interest in participating in a ‘league table’. The contention from the people at Cornell is that this was a strong factor in increasing participation in the ‘citizen science’ project. This question may seem like a repetition of the fifth question which asks for a rating of the league table. However, it is possible that the respondents may wish to participate in a competitive league table but they dislike our representation of it. This would be an indicator that further research was needed to establish what other encodings or styles would be more attractive for the

Figure 5.4 Section 2 ‘League Table’ (rendered here in a split screen format to display all 7 questions.)
5.3.4 Section 3 (4 questions about the respondents)

The third section of the questionnaire was to gather some demographic information for possible analysis by age, gender, handedness and interest in sketching. These questions are displayed in figure 5.5.

We simply wished to gather an age profile of the respondents and their gender. This may show up differences in preferences between different age groups and genders. Both age and gender have been studied in relation to cognitive abilities in geovisualisation (Slocum et al., 2001). We also thought there may be some interesting differences in the attitude toward competitive analysis in our ‘league table’. The question relating to whether the respondents were right or left handed was a matter of curiosity related to Human Computer Interaction (HCI) literature. Handedness has been studied in the field of HCI and has been the subject of some disagreement and controversy (Hibbard, 2004). We wished to take the opportunity to see if there were different reactions amongst the respondents on the basis of their handedness. This may provide some interesting insights for persons intending to develop computer or smart users.
phone applications relating to our topic. We asked respondents if they liked sketching as a substitute or proxy for possible enhanced spatial abilities.

5.3.5 Section 4 (5 questions on bird watching experience)

The penultimate section was aimed at establishing the level of bird watching experience among the respondents and the nature of any reporting they have been involved in. Figure 5.6 show the questions from this section. The questions in this section all reveal information about the experience and level of involvement of the respondents in field reporting. Each answer choice represents an increased level of engagement of the respondent. For example, for the question ‘Please state the length of time you have been involved in bird watching’, the respondent is allowed three answers in order of increased experience: 1 to 3 years, 4 to 10 years and over 10 years. The remaining questions are similarly provided with progressive answer choices. We are interested in participation in formal surveys so it would be interesting to see what the level is of this activity. The question on foreign bird watching is meant as an indicator of interest although we have been since made aware that there are highly
experienced bird watchers who rarely go abroad.

5.3.6 Section 5 (4 questions about experience with technology)

The final section attempts to establish the level of experience and awareness of relevant technology among the respondents. We would like to know a little about the use, among the respondents, of internet technologies. This could have implications for recommendations we may make about the need, for example, for simplicity of design, or for the need and degree of promotion effort to encourage adoption of new technology. We were interested to see how the level of familiarity that the Irish bird watching community has with technologies for bird watching, compares with levels of

![Figure 5.7 Questionnaire section 5. Respondent’s Use of technology](image-url)

those the United States.

The first question in this section establishes the level of familiarity among the
respondents with various internet technologies. We would like to see if the bird watching community displays any significant traits in this regard. We would expect to see a normal pattern here for email and flight reservation, in that we think most people in Ireland use these applications regularly. We think that a high usage of social media like Facebook or Linkedin might indicate a potential ease of adoption of a bird watching ‘report sharing’ or ‘experience sharing’ type of application. We also think a usage of SATNAV technology would indicate a level of comfort for report birds using geo-location functionality.

The second question relates to the type of devices in use. We are mainly trying to establish the level of use between pc and mobile devices. We think mobile applications have significant potential as an aid to field reporting.

The fourth question asks about Apple products. This question has two purposes: On one hand this question tries to ascertain if there is a strong leaning towards the iPhone platform, on the other hand, we were being slightly mischievous, in that, we suspect people who would express an explicit preference for Apple products, might have a design or visual leaning.

5.3.7 Conclusion

Our survey was an attempt to gain an understanding of the level of comfort the bird watching community had with the types of technology being used to collect data in the United States. We wished to examine their cognition of geo-visualisations and if they would find such technology attractive. We also wished to test their comprehension and reaction to a motivating concept, which we found in our research, to be effective in the United States. This was in the form of a tabular visualisation of the bird watchers track record and their ‘standing’ in a ‘league table’ against other bird watchers.

5.4 The Interviews:

For testing hypotheses b) we decided to use a structured interview approach and we asked four prominent members of the bird watch scientific community if they would give their time for a structured interview. We gratefully accepted their kind
affirmative answers. A brief profile of the four participants is provided as follows:

Niall Hatch.

• Development Officer with BirdWatch Ireland
• Avid birdwatcher since the age of 2;
• Has travelled extensively on 6 continents in search of birds and has contributed time to many conservation and survey projects across the world.
• Contributes records to I-WeBS, CBS, GBS, BirdAtlas, BirdTrack, eBird, Observado, World Birds and other online ornithological surveys and data collection portals.

Olivia Crowe
• Senior Conservation Officer BirdWatch Ireland.
• More than 14 years’ experience in bird monitoring projects.
• Currently project manages surveys including the Irish Wetland Bird Survey, Countryside Bird Survey and Garden Bird Survey.

John Fox
• Committee member Birdwatch Ireland Tolka Branch.
• 8 years studying and photographing birds.
• Architect, currently studying field ecology.
• Contributes records to IWEBS CBS and GBS

Dermot McCabe
• Chairman BirdWatch Ireland Tolka Branch.
• 30 years bird watching travelling extensively in Africa to pursue his hobby.

We elucidate here on the questions we asked and why we asked them. In the evaluation chapter, we will discuss the responses we received. A full transcript of the interviews is in appendix a. The following is a list of the questions we asked at the interviews. There were two broad sections:

5.4.1 Interview Section One

Section one comprised an extract from an article about new developments in citizen science in ornithology which the interviewees were asked to review. There followed 9 questions prompted by the article’s content. Figure 5.8 shows the interview document, containing the extract and the question, which was emailed in advance to the interviewees. We will examine each of the questions in turn and the responses given in the evaluation in chapter 6. The broad thrust of the questions was toward finding out what level of awareness the interviewees had of the potential for new analysis based on big data from citizen science. Also, what kind of issues, both positive and
negative, could they foresee with such data. The intention of showing the extract was to provoke interest and comment on the subject of our research, namely, how will the arrival of new data and visualisation techniques impact on the bird watching community, for both hobbyist and scientist?

Top: plot of predictor importance (PI) for 16 habitat variables in the year of 2009 for the species Grasshopper Sparrow. Bottom: tag clouds based on PI values to highlight which habitats are more relevant in three different dates. Tag clouds provide a more effective representation, helping scientists more quickly understand the relative importance of variables in a specific date. The actual tag cloud conveys the information directly and does not require the user to carefully inspect the individual plot curves and the legend. The sequence of tag clouds also clearly shows an important ecological fact: the habitat association for the Grasshopper Sparrow changes throughout the year, from Shrub scrub to Cultivated Crops and finally, to Grassland.

Source: http://vgc.poly.edu/projects/birdvis/resources/birdvis.pdf

**Figure 5.9 BirdVis visualisation and explanations**
5.4.2 Interview Section Two

For the second part of the interview we asked the interviewees to review a visualisation from a paper about the Cornell Laboratory’s BirdVis project. The visualisation is accompanied by a brief explanation from the same paper shown in figure 5.9.

After reviewing the visualisation we asked the interviewees three questions as follows:

Q1) What do you think of the way the data is visualised in the image? Is it useful? Is it easy to understand?

Q2) What data could be used from Ireland, UK or Europe to produce similar visualisations. Do we have such data?

Q3) What steps with data would have to be taken to achieve such visualisations in Ireland, UK or Europe?

The purpose of the questions in section two was to, firstly, engage the interest of the interviewees and to see if they were aware of such developments in their field; Also, to see if they thought this kind of analysis would be useful to them and, finally, to see what steps they would be needed to introduce similar techniques on this side of the Atlantic.

5.4.3 Conclusion

The interview part of our experiment was aimed at the scientific side of the bird watching community. They would be the people who would have the use of potentially massive new data sets if the tools were put in place and if the hobbyists reacted positively to their availability. In the first section, we wanted to see if the professionals were aware of this new resource above and beyond the traditional data collection efforts in conservation. We also wanted to see what thoughts they had on the impact such data collection effort would have on their community in terms of, for example, administration, data skills and governance issues.

The second section was show a particular scientific use of the newer large type of data set. We wished to find out if the scientists found the analysis useful and how they thought such data might become available in Ireland, UK or Europe.
6 EVALUATION

6.1 Introduction

6.2 Analysis of the Questionnaire Responses

The Google forms application automatically generates charts to analyse the responses to the questionnaire. These charts provide a very convenient exploratory overview of the responses. There follows a brief analysis of the questionnaire. Note: this analysis covers 43 responses. 13 responses are contained in three separate spreadsheets and are not included here. We don’t think this will significantly affect our discussion here.

6.2.1 Section One Analysis

Figure 6.1 shows the analysis of section one of the questionnaire. We can see from the responses to the first two questions that the respondents are familiar with Google maps and the representation of the field trip. 88% answered question 1 correctly and 77% got question 2 correct. There was quite a positive response to the idea of using a mapping application with a total of 49% expressing positive sentiments. 14% did say they prefer a manual system. We might have offered an option of recording manually with transfer to the map later at home or perhaps, the respondents were unaware how easy the recording could be with a smart phone. Finally, a high percentage of the respondents use computer based maps. We would expect this from people with an outdoor hobby and it was gratifying to see this confirmed. Section 1 shows a positive overall attitude toward our geo-visualisation and augurs well for such representation.
Figure 6.1 Responses to section 1

6.2.2 Section Two Analysis

Figure 6.2 shows the analysis for section two of the questionnaire. 91% got the first question correct and 67% the second. Familiarity would probably improve on the 33% who answered ‘1460’ or perhaps a redesign of the table. The third question posed very little problems with 93% correct. The fourth question is interesting in the spread of responses. The largest cohort (35%) thought that ‘unique species observed’ was the best measure, but ‘number of records submitted’ and ‘bird identification’ both returned similar results. This is not a very definite response and would certainly warrant further research. Interestingly, one respondent went to the trouble of contacting me by email to insist that ‘number of unique species observed’ was what most bird watchers cared about. Perhaps the respondents were trying to give a ‘desired’ answer or didn’t
understand the question. The response to the ‘league table’ rating question was inconclusive. 28% rated it a at ‘0’ and, although 33% rated it somewhat positively at ‘3’, the overall sentiment was lukewarm. This contrasts a little with the endorsement of the people at the Cornell Laboratory. Interestingly, the last question in this section indicates a much more negative attitude toward the ‘league table’. This shows a disharmony in the two answers and would warrant more investigation. Perhaps the users liked the representation of our table but are not interested in participating in a
competitive way. Perhaps this appeals more to American culture? The sixth question shows a positive sentiment to being able to keep record on-line although 37% state they are not interested. This might improve if people started sharing their experiences.

6.2.3 Section Three Analysis

The section three analysis shows the demographic breakdown and this is displayed in figure 6.3 below. The age profile looks to be normally distributed with 46 to 55 being

![Figure 6.3 Analysis of section three – demographics (4 questions)](https://docs.google.com/spreadsheet/ccc?key=0AklTniXb5nqMTw0DRisam5cHFZIUxIYk1Za12cFUES&gridId=0#chart)
the most well represented age group at 37%. The vast majority of bird watchers seem to be men. Perhaps we can check how this sample compares with a larger database if we ask BirdWatch Ireland to give us their breakdown. Overall, bird watching seems to appeal mostly to middle aged men. With regard to ‘handedness’, the figure of 86% right handed is not controversial and lies within expectations (M.K. Holder, 1997). The question on sketching is interesting but may require further analysis. The respondents may have understood the question to mean sketch birds which, one might expect, is popular among the bird watching community.

6.2.4 Section Four Analysis

Figure 6.4 shows the analysis for section four of our questionnaire – bird watching experience. The majority (81%) have more than 10 years’ experience. This may have implications for our results because, newer and, as we have seen, younger people are underrepresented. If there was engagement using computer applications, there may be a change in the demographic profile. Question two reaffirms the level of interest with 74% going on regular bird watch activities.

Question three shows that the bird watchers regularly take part in surveys which would indicate that they are already predisposed to ‘citizen science’. Although some people 19% didn’t understand the term ‘complete list’ (we should have been clearer), 46% showed they had submitted them and would possible be good candidates for less structured reporting (in other words reporting not as part of an organised survey but simply while in the field). Finally, 65% of the respondents travel abroad to take part in their hobby. This is potentially quite important for projects in Europe where reporting applications may have to be available in many languages to support the kind of widespread data collection which happens in the United States. This adds a potential level of interest with regard to sharing data across borders and multinational organisations.
6.2.5 Section Five Analysis

Finally, we look at the analysis of section five, the questions pertaining to the use of technology. The results are shown in figure 6.5. We can see from the analysis that the respondents are fairly *au fait* with technology. The first question shows a high usage
of common internet technologies and social media. This should mean they would easily adapt to a reporting and sharing application if designed in accordance with these technologies. Question two shows a high percentage using smart phones and Satnav.
This may indicate that the respondents would understand the idea of being able to carry out remote sensing for logging bird sightings, which would lend itself to increased reporting. 21% of respondents say they have a ‘strong preference for Apple products’. This might have implications for application design in that these people may not be prepared to use an android device if the application wasn’t available on iOS. We don’t know if this is a commonly held bias but it certain seems unambiguous among the Irish bird watching community. Apple has a loyal fan base.

6.2.6 Note of the tools used for the questionnaire

To carry out our questionnaire we used Google Forms. This cloud based application is available free as part of Google Drive (formerly Google docs), which now enables ubiquitous file management similar to Dropbox. The form is opened as a new form and has a drop down menu to choose from various question types. Multiple choice, ‘choose from a list’, Lickert scales, grids and freeform text are all supported as question types. There are questionnaire templates to suit certain types of projects and features such as flow control to skip to certain questions based on answers to earlier ones. Any question can be made mandatory to answer and the form will gently remind the respondent of this. In this way, we could develop our questionnaire very easily. There is an option to email the form, in which case we emailed the list server of the Irish Bird Network (IBN). One can also embed the form as a link on a website, in which case, we placed it on the Facebook page of BirdWatch Ireland. When the user opens the email or clicks on the website link, the questionnaire appears and the user can easily navigate through it. Both, emailing and embedding worked for us as strategies, and we found the process easy from a usability point of view.

When the respondent completes the form, a spreadsheet is automatically populated in the ‘Google Drive’ folder with the responses. There is automatic, built in descriptive analysis which can be viewed by the owner at any time. There are some drawbacks. Here’s a brief list of the one’s we found:

- A Google form can only be customised by using the templates provided.
- One cannot put an image on a form.
- The process can only collect respondent’s user name if the survey is limited to
a Google Apps account.

- One cannot prevent a user responding multiple times to a survey
- The automatic descriptive analysis is always generated for an entire form and one cannot limit it to particular groups of respondents, nor, can two separate forms be merged for a single analysis.

We found solutions to most of these shortcomings. For example, we were able to embed a link to our picture, a Google Presentation, on the forms ‘welcome text’. We trusted the community to offer only one response. Our survey was short, simple and aimed at a particular interest group so we did not feel the need to dress it up with a template. We did not wish to collect user names as our survey was anonymous. We had 43 respondents on one particular form and we felt that that would sufficiently represent the group to the level we required.

Overall we found the Google forms application has good support through user groups and on-line support with training documents and You Tube videos. We were very happy to use it and would recommend it for simple surveys on a low budget.

6.3 Analysis of Interviews

6.3.1 Interview Part one

Q1) What do you think of the usefulness of ‘using people as sensors’ in order to collect scientific data in the manner now known as ‘crowd sourcing’ or ‘citizen science’?

Responses: The four interviewees were unanimous in their opinion that people are very useful, indeed crucial to data collection. There was a danger they were stating the obvious but we are happy from the conversation they understood this to mean large scale data collection effort and not traditional organised surveys.

Q2) What problems do you envisage in data collected by enthusiasts and amateurs?

Responses: All expressed concern that data quality is a potential problem. Traditionally, survey participants are trained and carry out the work under supervision. There is a hierarchy of experience which is maintained through networking and
personal contact. There are various initiatives such as the garden bird survey which introduces children to the ‘citizen science’ practice. Some data collection efforts are more sensitive to accuracy than others. For example, a survey of a rare species would have to be carried out by experienced personnel. Large scale reporting may not be so sensitive.

Q3) How do you think such problems, if any, may be minimised?

Responses: There were slightly mixed opinions on this. Three of the participants expressed interest in and enthusiasm for more use of technology to help with data quality. All agreed that training was still very important.

Q4) What impact do you think the widespread availability of geo-position aware smart phones will have on the collection of useful data for ornithology?

Responses: Using the ever increasing quality of smart phone photographic images and sound recordings, as well as database and workflow technology to identify outliers and route these outliers to experts for vetting, were cited as potentially very useful possibilities. There was also recognition that, flowing on from this, artificial intelligence methods may be useful to recognise images and sounds automatically. There was agreement that smart phones would improve data quality. There was a small concern that the geo location may skew data to the position of the bird watcher and not the bird (This was interesting as most of the bird watchers understood the difference during the questionnaire when asked about the ‘yellow points’)

Q5) Which data governance/privacy problems can you think of using volunteer field records at such potential level of detail as dates times and locations of field trips?

Responses: There was a sharp difference between one of the interviewees who works professionally for BirdWatch Ireland and two who were on the Tolka branch committee. The professional, who is a qualified solicitor, assured us that data privacy issues were very important and members were allowed to maintain anonymity on surveys etc. This view may be influenced by his position as an administrator. This view was echoed by the other member of staff who we interviewed. The other two interviewees were not particularly concerned and, indeed, reported that most bird watchers liked to have their name associated with their work (This was slightly at odds
with the survey results which showed a lukewarm reception to the ‘league table’ idea).

**Q6) People in Ireland can use ‘Birdtrack’ or other apps like ‘eBird’ to submit data. This may lead to smaller data sets residing in different organisations. How do you think the data collected in citizen science efforts should be organised and shared?**

**Responses:** Three interviewees expressed a concern that there is a danger of duplication of effort and some loss of data to individual organisations. The suggested solution went along the line of high level agreement in organisations sharing their information. There seemed to be an attitude that this should be a reasonably easy objective to achieve and that good communication exists between BirdWatch Ireland, the BTO and The Cornell Laboratory. There was mention of a fractious situation in the United States where, a lot of people are members of small local clubs, in comparison to Europe, where most enthusiasts are members of national organisations.

**Q7) How would you compare the collection and organising of a database such as the Avian Knowledge Network’s ‘eBird’ in the United States, and anything available in Ireland, UK and Europe?**

**Responses:** This question was difficult for the interviewees as they were not so au fait with the eBird technology. One was very knowledgeable and described the Dutch system ‘Observado’. He offered some criticism of eBird in that it does not allow for certain types of data. However, he recognised that eBird and the traditional methods serve different purposes and complement each other.

**Q8) ‘All too often, scientists must become computer scientists and statisticians in addition to their chosen discipline. They need a ‘tool layer’ to support the information life cycle from initial research design through instrumentation, data capture, data management, analysis, publication, and curation’.**


**How do you think will the availability of massive geo-temporal data sets affect the skills needs of ornithologists and conservation science staff?**

**Responses:** All agree that ornithologists and conservation scientists needed to keep up with technology to make the most from the data. There is a significant cost to training
staff and this was expressed as a concern. The sensible view was expressed that any applications should be made as simple as possible to encourage uptake by the users.

**Q9** What interesting developments can you envisage in the future given the pace and uptake of database and mobile technology?

**Responses:** This question got an enthusiastic reaction. Mention was made of the benefits of increased accuracy of the location data. This could be automatically mapped accurately to land cover and land use data. The knowledge accrued from new small radio devices was mentioned and the fact that information from studying bird location data was very important during a recent outbreak of avian flu. The disease was tracked to the rail networks indicating that the spread was by means of commercial trading of domestic birds and not by wild birds.

6.3.2 The Interview Part Two

This part of the interview was based on the interviewee’s responses to the image displayed in figure ??. There were three questions:

**Q1** What do you think of the way the data is visualised in the image? Is it useful? Is it easy to understand?

All found the representation easy to understand and agreed it was aimed at a scientific audience. All seemed impressed with the visualisation. All found it useful and there was a general feeling that it would be nice to be able to see such representations for local species. The point was made that the land cover data would be more useful if it was more granular.

**Q2** What data could be used from Ireland, UK or Europe to produce similar visualisations. Do we have such data?

There was a kind of hesitancy about this question. The general feeling was that the habitat data was not sufficiently integrated or perhaps not available. One person declared himself to be insufficiently knowledgeable to offer an opinion.

**Q3** What steps with data would have to be taken to achieve such visualisations in Ireland, Uk or Europe?
One person again declared himself to be insufficiently knowledgeable to offer an opinion. The others ventured that the main issue was the refinement and availability of integrated habitat data.

### 6.4 Conclusion of evaluation

#### 6.4.1 Conclusions from the survey

Our research goal is to find out if data visualisation techniques such as geo-visualisation and clean tabular displays could enable large scale citizen science data capture activity in the bird watching community. If we remember from 6.1 above, our hypothesis was that: ‘People would like to get involved in citizen science by providing data from their field trips and that measuring themselves against others, in a similar way to the eBird project, would encourage this involvement’. We discovered something about the issues which might encourage or prevent this as follows:

- Familiarity with and comprehension of geo-visualisation as a means of communicating bird watching activity. We chose a variation of a commonly used representation, that of Google Earth, and found, that the respondents were familiar with and would have little difficulty adapting to this form of visualisation as a communication tool.

- Existing willingness to spend time bird watching and report sightings both in Ireland and abroad. We found that, indeed Irish bird watchers are dedicated to their hobby and this includes participating in studies. We have been told, anecdotally, that this is even more so in the United Kingdom. This would point to a positive outlook for using crowd sourcing techniques to harness this enthusiasm.

- Usage of current technologies. We found our respondents had access to the latest technologies and were using smart phones, tablets and Satnavs. There would be few barriers to participation with regard to tools, so long as any data capture applications were reasonably well designed. Indeed some of the respondents are already using applications for the purpose of recording field activities.
The testing of the ‘league table’ ranking and score keeping approach was inconclusive in our research. It has been much vaunted by the designers of the eBird application as a factor in its great popularity but the Irish bird watchers did not show themselves to be overly enthusiastic about it. This would have to be explored more in light of the contrasting experiences. The ambiguous results may have been caused by the design of the visualisation, or by the direct nature of the question. Perhaps people did not wish to admit to having a competitive nature.

6.4.2 Conclusions from the Interviews.

The general tone from the interviews was positive throughout and the four interviewees were concordant in most of their views. For the first section of the interviews, they all expressed enthusiasm for using new technology and the idea of collecting and sharing big data using ‘citizen science’. As one would expect, the interviewees were far more interested in their core topic than in technology per se. They were self-effacing in that they professed only lay knowledge with regard to matters such as database technology, although, they did seem to us to be well informed. An example of this was the awareness of certain possibilities for bias caused by poor vetting of geo positioning data. One of the group members was very knowledgeable about the need for proper data privacy policies. There was something of a presumption that the technology would happen inevitably and they would adapt to it. This ‘passive’ view possibly demonstrates a certain inevitability of a slower uptake of new data methodologies. Cost was mentioned as a concern and, it is possible that the Cornell Laboratory has better funding. It is probably natural, in a not for profit organisation, that initiatives, such as the BirdVis project, would need a visionary and/or a substantial sponsorship to provide the impetus to get moving. Perhaps a European wide project could provide the resources.
7 CONCLUSION

7.1 Introduction

This dissertation examined the emerging use of data analytics, particularly data visualisation, as it is being brought to bear on a community of interest, bird watchers and their counterparts in the conservation science community, ornithologists. We examined data visualisation as data science discipline and how it is currently used by bird watchers and ornithologists. We contrasted the different approaches in two parts of the world, and, by doing so we hoped to expose the benefits of this new approach to the parties involved. We carried out an experiment to research some of the techniques used. Our goal was to discover the effect of these techniques and try to find reasons to encourage faster adoption by the community of these new developments.

7.2 Research Definition & Research Overview

Our research wished to explore a gap we discovered in the use of data visualisation between groups in North America and similar groups in Ireland/United Kingdom. Our research was aimed at two separate potential beneficiaries of new uses of data visualisation, the Irish bird watching community and people involved at a high level in bird conservation science in Ireland. We constructed two hypotheses:

a) People would like to get involved in ‘citizen science’ by providing data from their field trips and that measuring themselves against others, in a similar way to the eBird project, would encourage this involvement.

b) The people involved on a professional level as conservationists and ornithologists would respond enthusiastically to the prospect of having new and different amounts of data on which to perform analyses.

We created some sample data visualisations based on those which had been successful in the BirdVis project in the Cornell Laboratory of Ornithology and tested the reaction of birdwatchers in Ireland by means of a survey questionnaire. We also carried out
structured interviews among members of the bird conservation science community to test their awareness of the new science of big data and its potential use to their domain.

7.3 Contributions to the Body of Knowledge

We think we have made contributions to the body of knowledge of data analytics in the following ways:

- We have drawn attention to the potential value of data analytics, in particular data visualisation, for providing new insights from data for use by the conservation science community and hobbyists.
- We have highlighted areas of importance in data management for ‘citizen science’ initiatives in the domain of bird watching and ornithology.
- We have emphasised the usefulness of new data visualisation tools, particularly Google Earth geo spatial representations for possible motivation of certain communities of interest.

7.4 Experimentation, Evaluation and Limitation

Our experiment set out to discover, from two different related interest groups, what issues need attention, if new developments in ‘big data’ enabled by ‘citizen science’ and new tools in information and communications technology, are to be exploited to deliver new insights into bird conservation. On one hand, we wished to find out if having a place for volunteers to store their field reports, by means of recording field trip sightings using geo location, and being able to visualise their competitive standings against peers via a ‘league table’ would provide noticeable motivation factors with potential to increase level of participation in ‘citizen science’.

We created some sample data visualisations based on those which had been successful in the BirdVis project in the Cornell Laboratory of Ornithology and tested the reaction of birdwatchers in Ireland by means of a survey questionnaire. We also carried out structured interviews among members of the bird conservation science community to test their awareness of the new science of big data and its potential use to their domain.
Our analysis of the experiments and subsequent evaluation of the results showed high levels of enthusiasm for contributing to citizen science on behalf of the hobbyists and high levels of comfort with the relevant technologies. There was conflicting evidence with regard to the motivating effect of a ‘league table’. The survey returned a negative sentiment but anecdotally, we were given the impression that Irish birdwatchers liked recognition of their work.

Those representing possible consumers of new ‘big data’ emanating from using ‘citizen science’, the conservation organisation, reacted positively towards the new ideas presented. They expressed an understanding and appreciation of the value of the new techniques developed in the Cornell Laboratory. There was however a ‘passive’ response to questions about how similar initiatives might take place in Ireland or United Kingdom.

This research was looking at ‘soft areas’ such as sentiment and opinions. The conclusions drawn are, therefore only indicators of possible realities. Ireland is very different to the United States in terms of infrastructure, geography and resources. This limits conclusions made in comparative studies of the two. We did not research the British community but we assumed a similar approach because of the close ties between them and the Irish. This may not be the case.

7.5 Future Work & Research

We would like to see further research into cooperative ventures for data federation possibly at a European level. We feel that availability of standardised shared data sets will greatly increase the possibilities for using ‘big data’ analytics in the domain of ornithology in Europe. A comprehensive survey of the existing ornithological data sets would be a very good starting point for this research direction.

The use of a citizen science approach has enabled the creation of large temporal geospatial data sets in the USA. So far, the analysis has been concentrated on the geographic dispersion of species over time and the factors or co-variants influencing these dispersions. We would like to see research in Europe, using a similar approach,
in particular with the integration of land use data sets. A survey of these land use data sets would be a good starting point. This research may lead to coordination of conservation efforts across European (and wider) borders.

We would also suggest further research into motivation factors affecting bird watchers in providing good quality data as citizen scientists. We looked at a competitive ‘league table’ as a possible motivator. The reaction was inconclusive. Further research using different table layouts and skills criteria may indicate different motivating factors. We would also suggest the possible use of a gamification approach. Such an approach could produce, for example, useful features to use as a training aid to improve reporting skills of the participants, or to encourage and educate younger bird watchers. We would also like to see this research widened out to a pan-European level. There may be considerable differences in the motivational triggers between different cultures and/or nationalities.

We would like to see further research focusing on subgroups of the birdwatcher population. For example, are there differences between the effort and/or skills of people with different genders, age groups or nationalities? Knowledge of these differences may lead to group specific refinements in data collection applications for these subgroups. Refinements to these applications could have the beneficial effects of reducing bias in the data collection effort. This research may take the form of surveys coordinated between the various national birdwatcher organisations.
BIBLIOGRAPHY


Fayyad U, Piatetsky-Shapiro G and Smyth P (1996) From Data Mining to Knowledge Discovery in Databases. AI Magazine. 17 (3).


Gesmann M and De Castillo D (2012) Using the Google Chart Tools with R: googleVis-0.3. 3 Package Vignette. Available at: ftp://cran.r-project.org/pub/R/web/packages/googleVis/vignettes/googleVis.pdf (accessed 27/12/12).


Post FH, Nielson GM and Bonneau GP (2003) *Data visualization: the state of the art*. Springer. Available at: http://books.google.com/books?hl=en&lr=&id=WAZYsfEMi4kC&oi=fnd&pg=PR9&dq=Visualization+the+state+of+the+art&ots=gBeRUt5F5v&sig=WIYh8ugu1ye0nCkD UWjiKYjCqNA (accessed 31/10/12).


APPENDIX A
The following is a copy of the interview introduction, the twelve interview questions and a transcript of the responses from the four interviewees. Each question printed once and is followed by each of the interviewee`s responses. Occasion prompting by the interviewer is in parentheses and is commenced with `Frank:´

Hi,

As part of my master’s dissertation (Data Analytics) I am conducting research into data visualisation of ‘citizen science’ data in the field of bird conservation and ornithology. I hope to conduct some interviews with personnel engaged in this field. The idea is for me to collect some qualitative data (your opinions) to add to my research. The interview should take no more than 15 minutes and comprises two sections below. Section one is about the recent potential to collect large amounts of data using citizen science techniques. There is a small extract from a paper followed by 9 questions. Section 2 has a data visualisation from the Cornell Laboratory of Ornithology followed by three questions. I would be very grateful for your help.

1. Section one
The following is an extract from the introduction to a paper from scientists at Cornell University (full text available here: http://www.cs.cmu.edu/~daria/papers/BirdMining.pdf)

`Ecology is fundamentally the science of understanding the distribution and abundance of organisms. Ecologists interested in efficient environmental manipulation for conservation and management of wild birds have two general needs: (1) to be able to accurately predict where a species is and is not found; and (2) to understand the causes of presence and absence of a species. Within ecology, the conventional paradigm for analysing data and gaining insights has been the formulation and testing of a small set of statistical models that are assumed, based on expert opinion, to be the most likely descriptions of the biological processes at work. This conventional paradigm is now becoming unworkable, overwhelmed by increasingly available large ornithological data sets with many potentially important features (e.g., geographic data sets based on satellite imagery). One example of this is the Avian Knowledge Network (AKN, http://avianknowledge.net), a group of university, governmental, and non-governmental ornithological organizations that are combining their existing databases of bird distribution information. Currently, over 25 million bird observation records exist in the AKN’s data warehouse, each record associated with data on over 200 environmental features, not even counting the additional geographic data. This volume of data requires new scalable analytical tools that provide ecologists with initial insights (hypotheses) to be subsequently examined in greater detail. For this, ecologists need to identify features that are strongly associated with interesting patterns of species’ occurrence and visualize their effects.’

The above extract points to the vast increase in the size of data sets potentially available to the ornithological community. The authors go on to propose a set of techniques to ‘mine’ the data for scientific knowledge. I am primarily interested in the availability of the data and how it is collected. The Cornell Laboratory of Ornithology through the Avian Knowledge Network have made great strides in recent years to consolidate a variety of sources of bird observations, particularly those sources using ‘citizen science’ initiatives, whereby the bird watching public post their observations via a website or a smart phone. These data are collected and stored in a formal database, eBird, which is available to the public and scientists alike for use in further understanding large scale spatial and temporal patterns and effects in bird population and
dispersion. I have some questions regarding the development and use of such data in Ireland and Europe. Please allow me some leeway in that I’m not an ornithologist. My background is in data analytics (collecting, storing and analysing data), and as a result my questions may seem naïve or misplaced.

**Q1) What do you think of the usefulness of ‘using people as sensors’ in order to collect scientific data in the manner now known as ‘crowd sourcing’ or ‘citizen science’?**

**Respondent 1:** Erm I think at the moment it’s the only way to gather data on birds and birdwatching is by using people to record what they have seen. So I understand that there is new technology afoot to perhaps you – as recognition, that sort of thing. Recognition of the sounds of birds, but I think that’s in its infancy and for the moment we just have to rely on ………………

**Respondent 2:** I think it’s remarkably useful particularly in the sphere of Ornithology/Bird Conservation/Birdwatching because I think Ornithology is one of the areas where amateurs, interested amateurs can have a bigger role to play than professionally trained scientists because there are just so many more of them who are getting out there with very high levels of skills and experience. XXXXXXXXXXXX into that as best we can, and of course, modern technology advancements and online applications and so on make that easier than before and more workable. So something that’s actually been going on here in Birdwatch Ireland for many decades, erm you know in the past people have to send us observations and data forms in the post and now it’s a lot more automated. But having said that, there are still people who use the old methods and we welcome whatever way we can receive the data. But I think it’s vitally important that it has been recognised as being crucial for the future of the whole science.

**Olivia Crow:** Most of our national surveys rely on the inputs of interested volunteers – they’re inputs are incredibly useful.

**Respondent 4:** This has absolutely enormous potential because it increases the amount of data that’s going to be available once it’s processed properly and properly vetted and you can assume that there’s somebody looking after the raw data and being judicious about the kind of stuff...

(Frank: And I know you go out yourself regularly with groups so there plenty of people interested in it).

... Loads and loads... the more there are, the better the quality of the data because perhaps if you were to limit it to a few very enthusiastic amateurs the first year or so that people get involved they’re inclined to see rarities at the bottom of every tree. So once the processors of the data realise that there’s phoenixes in the phoenix park they ought to treat it with a little judicious eh eh but other than that we’re most enthusiastic about this wonderful eh the more data the better.

**Q2) What problems do you envisage in data collected by enthusiasts and amateurs?**

**Respondent 1:** Why is it the obvious one. That there are different standards in the birdwatching world. Some people are well able to recognise hundreds of species of birds then there are others that can only manage a few. But they’re all quite keen to submit records and quite often you’ll get dodgy information because people aren’t that skilled. You also have the situation where people are hoping to see something and because they’re hoping to see it they quite often will try and turn something that looks closely like what they are hoping to see into what they see in the report. So you get problems in that regard.
People hoping to see something and then reporting it on the basis of a glimpse or a brief sighting

**Respondent 2:** Erm. There are always certain problems that are inherent in this. It does take a bit of time and a bit of resources on our part to try and make sure that we’re on top of it. Certainly, er, when you’re dealing with volunteers and you’re dealing with amateurs, first of all it can be difficult to assess levels of skill, and the accuracy of some of the data. Erm in some surveys that’s extremely important and in some others it isn’t so much. Sometimes, for example, if we’re looking at population transits of certain birds, particularly on a basic level of a garden bird survey which is the biggest citizen science project going on in Ireland at the moment. With thousands of people taking part in winter. If people are consistently making the same mistakes it actually doesn’t affect our trends at all because they’re making the same mistakes all the time with the birds. Whereas, other more kind of complex surveys like the Bird Atlas, which we’ve just concluded from 2007-2011 accuracy is much more important there and so therefore for those we have to spend a lot more time verifying the records and querying them.

**Respondent 3:** Data quality and significant efforts are placed into ensuring that our survey participants get the necessary training for the surveys that they are asked to participate in. We have dedicated project coordinators that spend probably more than 70% of their time working exclusively with the volunteer network ensuring that they are clear about the tasks and that they return their data promptly.

**Respondent 4:** Clearly when people are observing trends with commoner species that are easy to identify and there in more or less the areas you expect it’s going to be really useful because you can be pretty well sure that it’s going to be accurate and on the occasions when something odd is reported, the fact that it is reported is going to allow people to check up on it. So problems would be maybe over enthusiastic sightings of things that are unexpected and the trouble is this could happen but more often than not...

Q3) How do you think such problems, if any, may be minimised?

**Respondent 1:** Erm well somebody is going to have to vet all the data that’s sent in. or something whether that be a computer system that – using some kind of advanced database or whether it’s individuals going through records and looking at them and saying … well that’s highly unlikely for that area… maybe contact the person who’s submitted the record and talk to them to see if they can verify how good a sighting was it, how skilled is that particular person etc. etc.

(Frank: You’ve submitted records for Sharrock and you said that they were pretty strict about coming).

For the bird atlas, yes

Erm I wasn’t queried about any particular data that I sent it, but I would regard myself as a reasonably competent birdwatcher whereas other people who have submitted things that were a little bit spurious – they were queried about the actual records in the particular location

**Respondent 2:** Erm so the work that we’ve been doing in conjunction with the BTO in the UK, they’ve developed some computer tools to try and filter some of these records or to make queries where, for example, an unusual species is seen or unusually high number of a particular scarce species is seen, or the time of the year seems wrong. Those are queries and when submitted electronically they are automatically sent back to the observers to verify
that’s correct or not. So that’s …. People still can make mistakes. Something we need to keep an eye on. So it’s not a simple matter of just putting surveys out there and let them take care of themselves. It still needs a lot of supervision.

Well on that, certainly what we’ve been trying to do, and it’s difficult with limited resources, but we’ve a good branch network across Ireland. We’ve thirty local branches that help with this. We run training courses for people. We also try to get people involved in a mentoring system so that they will go out and do some of the bird counts. For example – Irish Wetland Bird Survey takes place each winters – IWEBS we call it for short – it’s counting key water birds, waders species, ducks, swans, geese birds like that as well – at key internationally and nationally important wetland sites. Maybe run by teams. What we try to do we try to have people almost sort of an apprentice as part of that where they would go out with an established count team. People that we know and have been trained up and we know that their skills are absolutely top notch and they can learn the ropes from those people. Starting off with some easier species and then graduating on to the more difficult and more cryptic species, and over time then we can have the confidence that these people are able to do it correctly. So that’s one way that we’re doing that.

Also, another thing that we do is we try and spread the word about birds as much as we can to get as many people involved in the simple surveys as we can. So the garden bird survey I mentioned a few moments ago, that’s a very popular survey. Partially because a lot people, who have a garden they can watch out, or not they can do it in a local park or even on a window box. Partly because it’s so simple. Most people can manage to do it. If they don’t know some of the species it doesn’t matter they leave it out. So they get sort of trained in that methodology and that idea. And then the idea is that the data we get from that is extremely useful. We build up a great data set of garden bird information that’s second to none. It’s the only source of this in Ireland. But, the idea a significant proportion of these people will also after a year or two be confident enough to graduate on to the more involved, more complicated, more technical surveys. And that seems to be paying dividends.

And we’ve also been trying with young children. We have a citizens science project called Spring Alive which we run with our partners in Birdlife International all across Europe and Africa, and that’s a great way from the age of seven or eight recording mind-sets so that they can then graduate further, and perhaps go on to do things like the Young Scientists exhibition and so on later on.

So it is very people centric. It’s very oriented towards people and training them etc. And it’s the only way we can achieve that level of coverage because there’s only a limited number of professional Ornithologists in Ireland and even worldwide terms the number of committed birdwatchers far exceeds that. Erm so it makes sense to tap into that resource.

Respondent 3: Training – as specified above

Respondent 4: A lot of these smart phones have pretty powerful, probably not professional but powerful enough photography that can be linked to the sightings. This is a wonderful thing I mean the might change even the serious scientific recording and the Irish rare bird committee who vet records are now much more inclined to make judgements if photography is available, if photographic records are available and in many cases where identification is difficult they’re beginning to insist on this because there’s so much.. photographic opportunities have increased because there are so many types of instruments you can carry like digital cameras, phones. It’s changed
Q4) What impact do you think the widespread availability of geo-position aware smart phones will have on the collection of useful data for ornithology?

Respondent 1: I think it’s ultimately probably where it’s going to go in that everybody is carrying a smart phone these days or very few people don’t have them erm and once geo location becomes sort of second nature in these things and it’s not something that you’re paying for, which inevitably it will probably become free – you will always know exactly where they are and they will start to use their smart phones to record where they’ve seen a particular species because if it’s not just for their own records it may be submitted to some larger database.

(Frank: Do you think it would improve the quality of the data? Or just the quantity).

Well it will improve the quality of the data in that people will be able to pinpoint exactly where it was seen. Erm whether it increases the quantity I’m not sure. If it’s a very easy thing to do, then yes, it probably will.

Respondent 2:

Erm I think that there’s certainly a great number of new advances and advantages we can get from this and we have been already. With Birdtrackers for example we already have an App available for an Android system and the IPhone system is due out later this months I believe and hopefully, people will be using that now early in the new year. Erm it makes it more immediate. I think for a lot of birdwatchers it’s easier for them to record things where they actually see them. Mostly the geopositioning software GPS and those – that works really well enough to bother finding out the grid references – much easier to pinpoint where they are when you see the bird. What that records for us, is not necessarily where the bird is, but where the person was standing when they saw the bird. So if something like a large bird of prey – a kite, an eagle or a buzzard – that could be a couple of kilometres away and the habitat that that bird is using could be completely different from where the person is standing.

(Frank: So that might be a problem....)

So we still need to know. I think we still need to know ... there’s still a bit of judgement on the part of the observer to be able to record that kind of thing. It’s not just like ..... we have skewed data. I think, for example, we’re here in north County Wicklow a lot of bird watching goes on along the Wicklow coast, we have some reserves. But you can see into the Wicklow Mountains where with binoculars or a telescope you could quite easily see a bird like a Buzzard four or five kilometres away. So if we were just going on GPS data on a phone or another device it would seem like all these birds are concentrated along a very narrow walkway next to the railway track by the beach and that wouldn’t be realistic. It’s about trying to make sure that erm you know that we are also able to get an understanding of the kind of the habitat at least that the bird is using so that we can better position that.

Having said that I think the idea of people being able to update the records and keep these lists as they walk through the field in real time, I think that’s really useful because it’s much less effort from them. If they have to go back home. What used to happen – and still does with many people – they have to go home where they set up a paper form or type into the computer. That’s an extra hour of their time. It’s also – there could be more errors in there, because they may misremember things or maybe not recording some of the more common birds.
One of the big stumbling blocks we have – a lot of birdwatchers in Ireland often think we’re only interested in the rare birds. Whereas in fact, you can actually sometimes learn a lot more about trends in the really common birds because they are so much more numerous. So if we changes in the population of – let’s say common birds like starlings or robins or blackbirds – that can actually be a lot more significant than seeing changes in the population of very scarce species. Because you have much bigger.

Respondent 3: Will be useful in time, when the majority of phone users are on smart phones. The ability to record directly onto a digital interface will be very useful for our surveys

Q5) Which data governance/privacy problems can you think of using volunteer field records at such potential level of detail as dates times and locations of field trips?

Respondent 1: I don’t think it’s a major problem in that the birdwatching world in Ireland anyway is pretty small. But the issue that could arise is that if somebody has decided to perhaps take a leave from work, has thrown a sickie shall we say, and they have gone off to see a rare bird that has turned up, they’re less likely to record that to a website that is publicly viewed by everybody because if his work colleagues knew that perhaps this guy was an avid birdwatcher, they could check the website to see if he had reported anything that day. So I think that’s the only issue that I can see. I don’t think … most birdwatcher are quite happy to put up their records and quite happy to have their names alongside them.

(Frank: What about sensitive birds that you wouldn’t want......?)

Well there are issues about breeding birds particularly – particularly in Ireland where there are a few species that are very rare breeding birds and those kinds of records are generally hushed up and they’re not recorded. But that situation ... the records get out occasionally, but I don’t think that’s going to change hugely with using advanced technology to record the locations. I think it’s the same thing now. There are people who are a bit casual about it and will record it when they probably shouldn’t and then there are more people who just – you know – know that’s it not a good idea to record a rare breeding bird.

(Frank: Maybe the rare ones could be filtered out of the data base anyway a lot of the time...).

If there is somebody with that kind of – I suppose approach – in the data monitoring website or whatever it is, then yes, that information could easily be filtered out.

Respondent 2: I think there’s definitely an issue there and it’s something that people are becoming more aware of. I think that the privacy protection and data protection legislation.

(Frank: You’re a trained solicitor...)

I am. So I do know a bit about this area. Certainly I think that people in Ireland in the past have been a little bit lax about this kind of thing. I think it’s right to be prudent on this. Take care of the personal details. Of course in Birdwatch Ireland we are very mindful of our responsibilities under data protection acts and that’s something we are totally compliant with. So we do not share personal data or make it publicly available without prior consent. In most cases there will be no need for that though for that consent to be given. There’s no need for the general public to know the names of these observers or data information. And then obviously when we produce papers based on these results it’s done not an individually identifiable basis. It’s done on sample size. So from that point of view we do keep that protection. Obviously we ourselves need to know who these observers are but it’s kept completely confidential and we are completely compliant there with our obligations.
Respondent 3: We provide survey volunteers with the opportunity to have their records held in confidence. Not really sure about the question.

Respondent 4: This is actually eh there are two answers if you like. One is that there’s already been an many years a serious of pagers and phone numbers that collect data and put it out as live as they can about rarities because there are people who are interested in chasing up rarities, either twitching or just interested in seeing rarities for their own sake and that already exists eh bird that could be disturbed by attention then you have to be judicious about it plus of course there’s also the case of when things turn up on private property or in areas that could cause ill feeling if you disturb people that has to be considered but generally speaking that’s grand. When it comes to rare breeding birds that’s completely different and it’s considered proper form to conceal close information.

(Frank: An the volunteers themselves, once they can have a user name – like they can call themselves ‘batman’ or something if they wanted to, they’re going to be happy enough that their own privacy in terms of where they’re going is not....)

Most people curiously enough with most people the reverse is true. They like to be associated with a finding so people.. I don’t know of anyone who is anonymous or conceals their identity when they’re reporting on something interesting because you’ll see for example if you look at the annual bird report for each year that comes out in Irish Birds people are always anxious to have their associated with an observation. As you say.. before of course.. a more comical issue if you like - a guy being photographed in the newspaper with an eagle when he’s supposed to be at his granny’s funeral... that kind of thing is always...but it’s not a big problem.

Q6) People in Ireland can use ‘Birdtrack’ or other apps like ‘eBird’ to submit data. This may lead to smaller data sets residing in different organisations. How do you think the data collected in citizen science efforts should be organised and shared?

Respondent 1: Well I think if it’s available to everybody then more people are going to use it. You know if you’re submitting records to something like Birdtrack and you can’t access the data yourself then you’re not going to bother doing it anymore. Everybody wants to see their record up there. And if it’s unavailable then I think people just won’t use it. Unless they know it’s going to something that will be produced in the future, such as a book, or a database that would be – maybe not an interactive one – but one that kind of has records over the last five years and the data will appear, then they’ll be happy to submit their stuff to it. But if it’s not available to the people reporting their sightings then it probably won’t happen.

(Frank: Is there a danger that the effort could be divided if some people report to EBird and some people report to Birdtrack that either data set is just not sufficient then?)

That is a problem, yeah. If half the population favour one site and half favour another site. I mean you do get situations like this going on in Ireland at the moment where people report to one site, and other people report to another site and some people report to both of them knowing that they’re not sharing the information. So, in an ideal world – if there are two rival organisations – that they would put their heads together to share the data because it would be a much more complete record of – you know – what’s out there and what’s about

Respondent 2: EBird is a remarkable tool. I think its tenth anniversary now. So it’s relatively new and it coincides with the big erm sort of increase and that in availability of this new technology that we have. I think that a lot of the systems that are in place in Europe, and in
other parts of North America as well, there are other systems there as well. They’re sort of trying to continue on previous methodologies back from the paper and pen era and to try and make sure that the data is comparable between those periods. I think that’s very important as well. I think EBird is a very useful erm – new tool taking advantage of this new technology and also willing to offer more immediate rewards to the people who contribute that data as well. I mean it’s a great use as a tool for a lot of birdwatchers. I know EBird has an app for example – if you want, there’s a certain bird that you really want to see and you haven’t seen it before and you happen to be travelling to a new part of the US or Canada or Mexico or wherever and you want to know where that bird was last seen, when was the last report on EBird, how close is it to me now, what’s the nearest road. That’s really useful in that you can trace the rarity that way. Certainly in Europe, particularly in Western Europe there’s a long tradition of birdwatching. Particularly like the UK, the Netherlands, Scandinavian countries where there are lots of birdwatchers, a big concentration. They would have systems over the years where, you know, they were using the web technology, pager systems, mobile phones etc. to spread that word faster than it would be than in some large area like the United States. So I think that’s certainly very helpful.

There’s another system in Europe, it’s done through an organisation in the Netherlands. It’s called observado, observado.org is the website for that. It’s sort of, it’s not quite as slick as EBird but it does the same sort of thing. You can pinpoint on a map exactly where the bird is. It’s more sort of adhoc.

What I find with things like, all the surveys we’ve been doing on our Bird Atlas and a lot of our scientific surveys that we do here, they are very precise methodologies where the information we’re looking for it still isn’t actually captured on Ebird. So we’re doing a breeding bird survey for example as part of our Bird Atlas. We want to be able to answer certain questions about the probability of the likelihood of bird breeding. Other observations about behaviour and so on. Which Ebird doesn’t necessarily capture.

(Frank: So it’s clear that the Citizen Science style collection does not replace the traditional methods. But might it add something new?)

I think definitely. They complement each other very well. I think there is quite a degree of overlap there. Erm, it’s – you know – certainly I think the idea behind Ebird of submitting – the idea obviously is that they want people to submit a complete checklist for a particular trip they went on. I think that’s a very good thing to do. I think there’s – and a lot of possibilities there – it shows a lot about the birds that are absent from habitats as well as present. A lot depends on observer effort and observer experience. Erm a lot depends the weather. I know they think about that and they build that into their monitoring.

(Frank: You mentioned various different apps that are available in Europe, but is it a danger that the data is not getting into a central shared repository?)

ERm. yes that is a concern. Having said that erm ourselves and the BTO, through the BTO who organise the Birdtrack and so on and help in some of the other surveys. There’s a lot done between our people and Cornell. So the idea is that information will be shared. Erm it obviously on duplication of effort, People aren’t going to submit to two different ones. But certainly there’s other things .... The same issue in North America with lots of local bird atlases, lots of local bird clubs. We find in Europe that people tend to be members of the national organisations. Particularly the Birdlife International part in each country. So in the Republic of Ireland that’s ourselves, Birdwatch Ireland. In the UK it’s the RSPB. It’s a massive organisation with 1.2 million members, a lot of people. Erm again ....
Respondent 3: Greater communication/ liaisons at higher levels – between the bigger schemes

Respondent 4: If they don’t share it seems idiotic because by and large I don’t know of what interest the American collectors of data in the general run of stuff here would be except for our winter visitors from the New World which of course would be of interest to them or on the occasional turn up of vagrants but eh it seems… not being a person who knows mush about IT but I mean it seems to me to be so easy to amalgamate all of this data and so useful because there’s a certain zoological overlap between the two areas that has enough of a whole to have the two sets of data for mutual interest

Q7) How would you compare the collection and organising of a database such as the Avian Knowledge Network’s ‘eBird’ in the United States, and anything available in Ireland, UK and Europe?

Respondent 3: Not very familiar with the American schemes – we’re probably a little behind over here in Europe, and our BirdTrack scheme continues to evolve. Better to get feedback from users of a number of these schemes for this question

Q8) ‘All too often, scientists must become computer scientists and statisticians in addition to their chosen discipline. They need a ‘tool layer’ to support the information life cycle from initial research design through instrumentation, data capture, data management, analysis, publication, and curation. -Borgman,Wallis,Enyedy, IJDL paper final revisions, Nov 25, 2006, Page 5 of 20

How do you think will the availability of massive geo-temporal data sets affect the skills needs of ornithologists and conservation science staff?

Respondent 1: Well ornithologists currently I think do have to know how to use the data manipulation software and various kinds of formulae would be used to analyse this data and it is becoming more and more important to learn a lot of the terms. Because the amount of data that’s being gathered now is vast and it’s just going to get more and more as time goes on. So it is something that has to – that all Ornithologists that are working in science – we’re not talking about casual birdwatchers – but you know science based Ornithologists are going to have to get grips with…. And erm, I think the important thing for whoever is developing this kind of data is that this stuff is easy for Ornithologists who aren’t particularly interested in how it works make a report. So the idea is that it’s easy to work, you don’t have to necessarily understand you know, how it works. So make it fairly user friendly.

Respondent 2: It’s certainly a major part of it and something that my scientific colleagues here in Birdwatch Ireland have to become quite adept with. Erm this does actually predate the modern sort of immediacy of this – I don’t know the GPS technology and so on and online submission. Erm this familiarity with mapping and mapping techniques has always been absolutely vital. It’s a crucial role. We actually have a dedicated person here in Birdwatch Ireland that does that. I know the BTO have many people, a whole team that’s involved with this. Yes it’s crucial. It’s no longer enough for bird researchers to just have knowledge about the birds and the wildlife and so on. And survey methodology, they need to know about computer mapping techniques….

And data storage and data, technical data issues as well. Becoming computer scientists almost or having to have some specialist …..

That takes quite an investment on our part. And a charity like Birdwatch Ireland it’s always a struggle to try and meet those needs. We do at the moment but as the data grows you need more storage space, you need more computer power to analyse this, you need more people

XIII
as well to make sense of this data. It becomes a challenge, absolutely, but it’s a challenge that we have to meet because it’s the way it’s going to work in the future.

**Respondent 3:** In that boat – tend to learn the necessary skills as and when needed. Some folks are not very good with computers and unfortunately if you fall behind in the technology side, then over time become more and more inefficient (relative to the others becoming more efficient)

**Respondent 4:** Clearly to draw meaningful conclusions from raw data you need to know how to work it.. how to make your conclusions and all that kind of thing.. certainly the people who are going to process the data and eh come to conclusions with it are going to need statistical skills but eh for the people submitting it you don’t need to organise it too much you to …distribution curves or anything like that you know just send in the stuff.

**Q9)** What interesting developments can you envisage in the future given the pace and uptake of database and mobile technology?

**Respondent 1:** I think with the , well I think yeah, as people are now carrying more and more smartphones and that kind of thing they’re going to use the technology more and there’s going to be much more data out there and hopefully it will become more accessible to everybody as well. That’s where I see it going.

**Respondent 3:** Mapping/ ability to locate a record is badly needed – and the ability to build this into websites/ online surveys etc.

**Respondent 4:** Well em data and mobile.. I don’t know if ..I don’t know if this is quite what you had in mind but the most interesting new development in recording new bird data is the change over from ringing to radio tracking. Of course the old way of finding information about bird movements was to trap birds or catch them in the nest to put rings on their legs and in the case of large birds you can possible spot ring with binoculars or a telescope but in the case of tiny little birds you have to retrap them or find them dead to get any results and a ringer might put thousands of rings out in the hope of getting one or two back, but now for example in Britain, they just trapped six male cuckoos two years ago and put radios on them and followed them day by day on their migration back to Africa into wherever they went in Africa and back again to Britain and em with six individual birds they had more information about cuckoos within a season than all the ringers had ever got.
2. Section Two

Top: plot of predictor importance (PI) for 16 habitat variables in the year of 2009 for the species Grasshopper Sparrow.
Bottom: tag clouds based on PI values to highlight which habitats are more relevant in three different dates. Tag clouds provide a more effective representation, helping scientists more quickly understand the relative importance of variables in a specific date. The actual tag cloud conveys the information directly and does not require the user to carefully inspect the individual plot curves and the legend. The sequence of tag clouds also clearly shows an important ecological fact: the habitat association for the Grasshopper Sparrow changes throughout the year, from Shrubscrub to Cultivated Crops and finally, to Grassland.

Source: http://vgc.poly.edu/projects/birdvis/resources/birdvis.pdf

Q1) What do you think of the way the data is visualised in the image? Is it useful? Is it easy to understand?

Respondent 1: Yes. Very obviously looking at it the first time, what it was about and I would be very happy to use data like that – you know – if I was particularly interested in the type of information I can provide. So I think it is quite easy to understand and looks well.

Respondent 2: Erm I think the way I look at it is that it’s useful to certain people, easy to understand to certain people and then for other audiences not at all. Erm looking at the way it’s portraying the image here erm the graphs and so on, there is quite a lot of data there, it’s quite complicated. I see the way it’s working. I’m uses to looking at graphs so I can see exactly where it’s coming from. I think that the colour coding can be a little bit confusing because there are so many colours there, it takes a bit of analysis that way. Having said that I think that for a lot of people to take part in our surveys it would be kind of baffling. It looks a little bit remote from being out in the field actually birdwatching.
(Frank: So it’s more aimed towards the scientists .......)  

Yes I think it is. Exactly.......It’s also I mean this is - one species – the grasshopper sparrow which is a North American species erm quite – a very widespread bird with wide habitat preferences. In terms of the actual criteria they are using now in a European context, particularly an Irish context it would be quite different. I notice there for example, cultivated crops is obviously very important. For us in Ireland that wouldn’t be good enough we need to know what kind of crop it is because it’s done on a much – a lot of our farmland birds in Ireland are declining dramatically and some crops are better than others. Erm so that’s – but that’s obviously specific to each species – I think as a general tool it is very useful. Erm but I think it does take a bit of training, a bit of experience to be able to use it to its maximum potential.

Respondent 3: This is all very useful – I currently cannot see any use for it in my line of work – that may change

Respondent 4: Yeah, I’d have to say, compared with the old fashioned graphs it’s both useful and very easy to understand.

Q2) What data could be used from Ireland, UK or Europe to produce similar visualisations. Do we have such data?

Respondent 1: Well we would have to – I think – log the data against – against habitat type in this country. I don’t know if anybody is doing that at the moment but erm when anybody records sighting of a bird or whatever it would need to be linked to the type of habitat that it was found in. And that’s something that I don’t think we have the facility to do at the moment but I think the data is there, it’s just a matter of putting it together.

Respondent 2: We do and particularly in the areas in Ireland – particularly in the areas of wetland birds, birds using mud flats, coast estuaries, beaches and so on. We are already recording that kind of data – to a very detailed degree. Erm the different kind of literal habitats that they are using. Erm, so we do have that, we do have a map of different species and different distributions. Also with our countryside birds survey which is one of our most technical surveys where we have 300 randomly selected one kilometre squares across Ireland which are surveyed twice during each breeding season. That erm, for those – the people who record very, very detailed habitats data and the changes in that between each visit year on year. Erm, and, it’s difficult in Ireland to find enough people with the skills and time to be able to do that. That’s why we have to limit it to 300 but from ratio able to build up a very good picture of patterns and habitat use. So this habitat use has been also a very important part of the work we do. It’s worth erm mentioning of course that Birdwatch Ireland’s whole reason for existence is as a conservation organisation. So the data that we are gathering, the ultimate aim is the conservation of birds and their habitats and by extension of course, the whole suite of flora and fauna out there because we can’t just protect birds in isolation. So to do that we have to make sure that our habitat data is as accurate as possible.

Also Ireland is home to some really important threatened habitats on a European scale and that’s a very important part of what we do. Erm so a lot of the media recently have been about raised bogs in Ireland. It’s a really unique and important habitat that we have here. And it’s very controversial. We understand each side of the debate. We understand the need for local communities and traditions to .........
Well the thing is. The very important part of this is that we engage – we do – with people like Liam Flanagan and other interest groups because it’s not good to be reactionary and just to say no! no! to this. We have to be able to make our argument, to make our case. To say why we think this is important. Why a given area perhaps isn’t so important and the only way we can do that, is not through well-meaning passionate but ultimately ineffective statements. It’s scientific evidence. That’s what we need and that’s what we’ve been over for the last four decades. That’s what Birdwatch Ireland has built on. So we have quite a big database on this in Ireland. Not perhaps as user friendly at the moment or analysed to the degree that the Cornell data is in the graph that we just been looking at. But you know, the data we would be able to produce something like that is the need came if we had the resources to do something that habitat data is absolutely vital. And in fact without that a lot of the bird site data is absolutely meaningless. And that’s one of the things with – one of the issues I see sometimes with submitting complete trip lists and so on. These are a great way to do it and capture it but of course when a person is doing it on a trip walking along they might say, ok well I saw – let’s say five chaffinches here, three chaffinches there, eight chaffinches there – a very common bird the chaffinch. But they could be in completely different habitat types on the walk. It’s just been recorded as something like twenty or thirty chaffinches. That’s not telling us too much about the habitat use. The Geodata might if it was combined electronically – would pick it up automatically if it was accurate enough in the future that might be ……

I think the thing about the new technology advances recently is that it allows us to capture all sorts of data that we may up until now not even have realised that there was a use for, but it’s there to be used down the line. So I think the idea is we should capture as much data as we possibly can, every aspect, because you never know when it could come in handy, and when we need it we can bring it out. A good example – but actually very handy for us as well – was erm, you may remember a few years ago all the fuss about bird flu, it was all over the media. We actually had our profile raised quite a lot about this because the media were talking to us a lot and from that we were actually able to be quite reassuring. We had a lot of data on the movements of migratory water birds. And for instance, a lot of hysteria in the media about how birds were going to kill everybody and how there’s going to be no listen the birds that migrate to Ireland are coming from here, and here and here, completely separate from the from the important point of view of infection. Not only that we were able to show through the data from our Birdlife partners that actually the pathway of these diseases was not following migratory bird pathways at all it was following Europe’s rail and road networks. It was transportation of poultry that was causing the infections, not migratory wild birds. And they were in fact victims of this disease rather than the vectors and that - we were able to be very reassuring about this and were actually able to come out and categorically say in the media, no there’s no need to panic on this we know what’s happening here and we’re not just saying this because we love birds we’re saying it because we have scientific evidence that backs us 100 per cent.

Respondent 3: This is all very useful – I currently cannot see any use for it in my line of work – that may change

Q3) What steps with data would have to be taken to achieve such visualisations in Ireland, UK or Europe?
Respondent 1: Yeah. Well the other thing I would have said about that is we possibly need to refine the habitat mapping for this country a little more. Get the habitat survey better. Maybe have it done on a more regular basis because habitat is constantly changing. So it’s just to kind of get that sort of up to speed so it can be linked to the records or stuff being recorded.

Respondent 2: I think it will take a bit of work to get the data into that kind of form. I think it can be done. It may take erm some more habitat surveying in certain areas. Because certainly we wouldn’t have habitat data for every single record that we have through something like a Bird Atlas or Birdtrack. Certainly the size of our survey for things like the Irish Wetland Bird Survey, The Countryside Bird Survey we do have a good idea of that. Even with our garden bird survey we have data on the kind of plants people put in the garden and also particularly for that point of view the kind of food people put out in the garden. Because that’s habitat too. That’s an artificial habitat people are creating. Because the birds are there for that reason so it’s every bit as valid a habitat as any other. I mean something like a cultivated field is every bit as artificial as that, it’s man made. Erm so there’s a lot we can do there. But I think it’s going to become more and more important part of the survey work that we do that we’re asking people to do habitat type surveys while they’re also gathering the bird data, before or after gathering the bird data. The nice thing of course about habitat being it doesn’t change as rapidly as the presence of a bird. So what we would recommend on the Countryside Birds Survey for example is count the birds first and when you’ve done your hour long survey then go back and plot the habitat types because the crops won’t have flown away, or the trees won’t have fallen down at that stage so you have more time to do that. I think that’s a very important part of it. And the idea is then also to get our disparate databases from each survey interconnecting more and to have one sort of central repository for this. But there’s a lot of work involved in that and as a charity ourselves relying on the support of our members to fund our work it’s difficult to get everything done that we’d like.

Respondent 3: This is all very useful – I currently cannot see any use for it in my line of work – that may change