|  |  |  |
| --- | --- | --- |
| Interactive exploratory data analysis on the web | 15/16 | |
| Yongyan Zheng (Carina) | |  |

Contents

[Abstract 2](#_Toc441526368)

[Introduction 2](#_Toc441526369)

[Motivations 2](#_Toc441526370)

[Preparations 3](#_Toc441526371)

[Methodology 4](#_Toc441526372)

[Primary Data Exploratory 4](#_Toc441526373)

[Functions Selections Progress 4](#_Toc441526374)

[Application Discussion 5](#_Toc441526375)

[Brief Result Interpretations 5](#_Toc441526376)

[Results 6](#_Toc441526377)

[Conclusion 7](#_Toc441526378)

[Future Development 8](#_Toc441526379)

[Appendices 9](#_Toc441526380)

[Data Classifications 9](#_Toc441526381)

[Glossary 11](#_Toc441526382)

[Full Code 11](#_Toc441526383)

[References 12](#_Toc441526384)

[Acknowledgement 12](#_Toc441526385)

# Abstract

[Here we are going to write the a brief summary of what this report and the project are for]

# Introduction

## Motivations

Visualization is the communication of information using graphical representations (Ward, Grinstein and Keim, 2010). Because human beings are visual beings, visualization provides a more efficient and completed method to obtain and understand information. In old times, the computer graphics were widely used. However, as the technology is getting more advanced day by day, people want to obtain information in a more interactive manner. Therefore, the idea of Interactive Communication comes out. Such concept tends to not only display the information but also enable users express with the visualizations, and users have the ability to choose how the information is represented.

In this project, we are building an interactive map of New Zealand along with the 2013 census data. Such map would show the means people travel to work and their weights (in percentages) in different territory authorities within New Zealand. We are aiming to be able to find the key variables that affect the weights of the travel means from the map. However, the most important purpose is to provide the possibility that users can visualize the changes of the travel means between different territory authorities, and at the same time, the map would enable them select and update the parameters in a more user-preferred manner. Potentially, users will be able to develop their own understandings towards this topic and potentially go further in their fields.

Data used in this research is from Statistics New Zealand, the major producer of official Statistical of New Zealand, which collects and analysis raw data, publishes useful information in general English to the public. It is efficient in helping individuals, community groups, businesses, and government agencies make better decisions.

The key definitions involved in this research are:

1. **Territory Authority**: It is either a city council or district council. There are 67 territorial authorities consisting of 12 city councils, 53 districts, Auckland Council, and Chatham Islands Council (Please refer to section 1 of the Appendices for the full list of the Territory Authorities involved in this research).

**Agenda:**

* + Excluding territorial authorities: Chatham Islands, Area Outside Territorial Authority.
  + The version used is 2011 v1.0.

1. **Population**:
   * On New Zealand level, people who are usually live in, and were present in, New Zealand on census night.
   * On Area level, people who usually live in that area and were present in New Zealand on census night.

For example, a person who usually lives in Christchurch city and is visiting Wellington city on census night will be included in the census usually resident population count of Christchurch city. (Statistics New Zealand, 2015)

1. **Main means of travel to work**: the method a person aged 15 years and over used to travel the longest distance to their place of employment on census day (for example, by bicycle, bus, walking, or jogging).

Responses indicate the main means of travel people chose on the census day and not necessarily indicate the responders’ usual mode of travel to work.

**Agenda:**

* + Excluding means: 77 Response unidentifiable, 99 Not stated.

1. **Rounding**:The data used in this research has been randomly rounded to protect confidentiality. Individual figures may not add up to totals, and values for the same data may vary in different tables.
2. **Extra Note**: some data is confidential; they are represented by “(c)”.

There are many different types of people who will use this map for specific purposes. The typical scenarios are:

1. To-graduate or newly-graduated high school students. These students are considering attending universities in other cities. Using the interactive map, students can find out if the cities match their expectations in regards to the main means of travel to work.
2. Biologists. They are designing experiments for any relevant research topics. Using the map, they will be able to make comparisons base on their understandings and findings, hence increase the accuracies and lower the risks of the experiments.
3. Politicians. Politicians have certain policies/plans that they want to submit to the authorities and hope the public will support them. Teams of the politicians then would use this map as their references/evidences to demonstrate the advantages of the new polices or plans.

In this research, the percentages of the travel means are calculated by applying the following the steps to all territory authorities:

1. Sum up the total number of people involved in the research by territory authorities (refer to "List of Territory Authorities", Data Classifications in Appendices).
2. Calculate the percentage of each travel mean within the same territory authority: Divide the number of people using such mean by the sum of people in the same territory authority.

For example, there are 168219 people from "Christchurch City" (territory authority) involved in the research. 5151 people chose "Public bus" (travel mean) on the survey conducted day. Therefore, the percentage of "Public bus" in "Christchurch City" is calculated as 5151/168219 = 0.03062080.

## Preparations

The components involved in this research are:

1. an Operating System combination of Ubuntu (Linux) and Windows.
2. RStudio, version 0.99.489 on Ubuntu.

# Methodology

This section in separated into three sub-sections. The first two sections explain the primary setup of the methodology, followed by the process of applying the target data and the selected functions. At the end, it generates the outputs and illustrates the meanings behind the results at the first look.

## Primary Data Exploratory

## Functions Selections Progress

The functions selections progress is separated into two parts: background foundation (FD) and user interface (UI). The background foundation provides and pre-process the data required for the interactivities with the users. And as its name, the users are not able to view the processes. The user interface is the platform that users see, use and obtain information from. The background foundation also receives the indications made by users on user interface, while the user interface sends and displays the generated result.

1. Data table

The source data is being stored in the format of Comma Separated Values (csv). Storing and reading data in csv file have the following advantages:

* 1. It can be opened by Notepad (Windows and Linux) and Microsoft Office Excel.
  2. The data fields are separated by comma (",") which makes the data and plain reading more accurate and clear.

1. Interval selection
2. Colour definition
3. New Zealand map (of single travel mean)
4. Dynamic Scatterplot for two travel means
5. New Zealand map (of two travel means)
6. Shiny

Shiny is an open source R package. Developers can build web applications by using R along with Shiny (web framework). It can also be used with HTML, CSS, and JavaScript for better styling and enhanced functionalities. More information can be found at <http://shiny.rstudio.com/>.

## Application Discussion

[Here I’m going to discuss the difficulties I have met during the development]

## Brief Result Interpretations

# Results

# Conclusion

Summary:

1. Conclude the findings from previous sections
2. Discussion on the conclusions

# Future Development

Given the short time frame of the research project, some ideas cannot be performed. However, they shall be descripted and discussed in this section.

1. Interactive Linear Regression Plot.

Summary:

1. What I haven’t done/gone through because of I’m short of time
2. What I would do if I got more time
3. Suggestions and thoughts to data visualization and interactive map

# Appendices

## Data Classifications

1. List of Territory Authorities:
   * 001 Far North District
   * 002 Whangarei District
   * 003 Kaipara District
   * 011 Thames-Coromandel District
   * 012 Hauraki District
   * 013 Waikato District
   * 015 Matamata-Piako District
   * 016 Hamilton City
   * 017 Waipa District
   * 018 Otorohanga District
   * 019 South Waikato District
   * 020 Waitomo District
   * 021 Taupo District
   * 022 Western Bay of Plenty District
   * 023 Tauranga City
   * 024 Rotorua District
   * 025 Whakatane District
   * 026 Kawerau District
   * 027 Opotiki District
   * 028 Gisborne District
   * 029 Wairoa District
   * 030 Hastings District
   * 031 Napier City
   * 032 Central Hawke's Bay District
   * 033 New Plymouth District
   * 034 Stratford District
   * 035 South Taranaki District
   * 036 Ruapehu District
   * 037 Wanganui District
   * 038 Rangitikei District
   * 039 Manawatu District
   * 040 Palmerston North City
   * 041 Tararua District
   * 042 Horowhenua District
   * 043 Kapiti Coast District
   * 044 Porirua City
   * 045 Upper Hutt City
   * 046 Lower Hutt City
   * 047 Wellington City
   * 048 Masterton District
   * 049 Carterton District
   * 050 South Wairarapa District
   * 051 Tasman District
   * 052 Nelson City
   * 053 Marlborough District
   * 054 Kaikoura District
   * 055 Buller District
   * 056 Grey District
   * 057 Westland District
   * 058 Hurunui District
   * 059 Waimakariri District
   * 060 Christchurch City
   * 062 Selwyn District
   * 063 Ashburton District
   * 064 Timaru District
   * 065 Mackenzie District
   * 066 Waimate District
   * 068 Waitaki District
   * 069 Central Otago District
   * 070 Queenstown-Lakes District
   * 071 Dunedin City
   * 072 Clutha District
   * 073 Southland District
   * 074 Gore District
   * 075 Invercargill City
   * 076 Auckland
2. Main means of travel to work:
   * 01 Worked at home
   * 02 Did not go to work today
   * 03 Drove a private car, truck or van
   * 04 Drove a company car, truck or van
   * 05 Passenger in a car, truck, van or company bus
   * 06 Public bus
   * 07 Train
   * 08 Motor cycle or power cycle
   * 09 Bicycle
   * 10 Walked or jogged
   * 15 Other (includes taxi, ferry, helicopter, aeroplane)

## Glossary

Summary:

1. List of the codes or special/professional nouns used
2. Extra helpful definitions

## Full Code

# References

1. Ward, Grinstein and Keim, 2010
2. Statistics NZ website

Summary:

1. Purely references corresponding to the references from all previous sections
2. Statistics New Zealand (2015) <http://stats.govt.nz/> (checked on 22nd December 2015)

# Acknowledgement

[section for supervisor, funding/relevant organization, helping staff etc.]

**LOG: LAST UPDATED ON 25/1/16 19.15**