SWINBURNE UNIVERSITY OF TECHNOLOGY

FACULTY OF SCIENCE ENGINEERING & TECHNOLOGY

COS30045–Data Visualization SEMESTER 1 2019

Project Process Book

DUE DATE: 3rd June 2019 11:59PM

STUDENT NAME: SANEE SALIM

STUDENT ID: 101887181

I hold a copy of this assignment that can be produced if the original is lost/damaged. To the best of my belief, no part of this assignment has been copied from any other student's work or from any other source except where due acknowledgement is made in the text. No part has been written for me/us by any other person, except where such collaboration has been authorised by the lecturer concerned. *Sanee* **31/05/2019**

Table of Contents

1 Introduction

According to Benno, Gang (2011), To date, an insufficient amount of research attention has focused on the elderly or seniors within Australia. Considering the increasing life expectancy in recent decades, it is essential that extensive researches are done on the social and financial background of senior Australians. This report aims to create an initial step towards that research process by producing some visualisations representing different aspects such as diversity, what they have been doing and other factors of Australian seniors' life. This report uses data from the Australian Institute of Health and Welfare that provides data to research on old age Australians. These researches can be reflected upon by Government officials, journalists, policy makers and people of the community who are willing to study about status of older Australians and services available to them. It could be used to develop facilities and policies that could lead to a better quality of life for Australian seniors in the upcoming future.

1.1 Background & Motivation:

According to the Australian Institute of Health and Welfare, in 2017, 15% of Australians were aged 65 and over. Considering the number of Australians categorised as old age Australians, many services such as aged care homes and social programs are arranged for these people. Besides the facilities provided to old aged Australians, there is still room for improvement in the facilities provided to Australian seniors. The goal of this project is to make meaningful visualisations that can be used by Government officials, Non-Governmental and other organisations that aim to benefit the seniors in our country. These organisations can use these visualisations and statistics and provide better facilities and services to improve the lifestyle of seniors in Australia. I have personally worked in, and Aged Care's hospitality service and have realised that undoubtedly there is a room for improvement in the facilities provided. It needs to be noticed that these services cannot be improved without substantial information. That is why I chose this project aiming to convey a message to the authorities about the measures that can be taken.

1.2 Aims & Objectives:

This study aims to study the current behaviours and past about the life of Australian seniors. The study would also research aspects such as work history, housing, marital status, social participation, technology usage for male and female seniors. The visualisations would aim to answer these questions

- Which work sector has had more influx of seniors previously?
- What is the rate of marriage amongst Australian seniors?
- Which counties contribute to the Older Australian's population?

The visualisations would answer these questions and many more questions about Australian Seniors. The visualisations would be interactive aiming to make genders, sectors and countries as channels. The visualisations would also include data from overseas and Aboriginal Australian seniors.

1.3 Project Schedule:

The below table provides a tentative schedule for the semester-long project, which was roughly followed to produce this report successfully.

Week	Task
5	Data Collection
6	Data Understanding
7	Data Preparation
8	Modelling
9	Evaluation
10	Deployment
11	Presentation
12	Reporting

Table 1: Project Schedule

1.3.1 Data Collection:

As Roger Sapsford, Victor Jupp (p.02) highlights that valid research can equally be carried out on data already collected by someone else. This approach was considered for this project and the data collected by the Australian Institute of Health & Welfare, that is a very renowned institute for researches in Australia. Many other sources were also considered, but this data had the most convincing aims and objectives which prompted me to choose this data.

1.3.2 Data Understanding:

This step was used to analyse if the data collected in the previous stage was valid and representative of the questions that are aimed to be answered by this project. It was concluded that it is indeed the perfect representation of the question. Although, the data is slightly old but it is still effective to push the idea of lack of focus on Australian seniors.

1.3.3 Data Preparation:

This part is also called the data processing. This involves cleaning, merging and formatting the data to convert it into readable CSVs. The data from the source was not representing the percentages. Instead it was representing the real values which can often be misleading. All the data was cleaned, and ratio-based values were produced using Microsoft Access. Unnecessary columns that were not used for the analysis were also removed. And lastly, the decimal figures were rounded off for better representation of the data in the visualisations.

1.3.4 Data Modelling:

The data was modelled, and extensive brainstorming was performed to choose the right visualisation for the data. All the marks, channels and optional and must have features were decided and designed using tableau which is a Business Intelligence tool. This gave a vision that had to be achieved for the visualisation project

1.3.5 Evaluation

This process involves evaluating the aspects of the visualisations to reach a conclusion as to whether the visualisations are or are not representative of the dataset. It is also kept in mind that the visualization answers the questions associated with the visualisation successfully as well.

1.3.6 Deployment

This part of the schedule aims at writing code that is required to show these visualisations in D3. First an HTML page was created with some basic styling using CSV. After that a JavaScript page was created which used the D3 library which was used to produce the visualizations on the webpage.

1.3.7 Presentation

This week was consumed by planning, preparing and executing the presentation of the project. Microsoft PowerPoint was used, and the project was presented in the specified tutorial as part of the submission guideline.

1.3.8 Reporting

Week twelve was used to form the report that highlights all the part of the project. This report also justifies the usage of the visualization, marks channels and the features of the website in which the visualization was deployed.

2 Data

2.1 Data Source & Format:

The data is provided by Australian Institute of Health and Welfare. The data is provided in Excel format and can be downloaded by the link provided. The data consist of 18 categories which include finance, employment, transport, social, technology usage, marital status and other factors that are related to Australian seniors. All these data sets are in the form of table. The data consist of data from 2006 and includes some future predictions as well. For this study I will only be using the data associated with years 2010 and 2011.

2.2 Data Type & Attributes:

Since data from three categories was used. The sections below would describe the associated variables with each of the category.

2.2.1 Work Sector

This was the first matter that was visualised. This visualisation aimed to answer the question "Which work sector has had more influx of seniors previously?" The dataset originally consisted of 6 fields those were:

1. Industry: Highlights the industry for which values are shown
2. 45-54: Number of 45-54 aged workers in the above industry
3. 55-59: Number of 55-59 aged workers in the above industry
4. 60-64: Number of 60-64 aged workers in the above industry
5. 65+: Number of 65+ aged workers in the above industry
6. 45+: Number of 45+ aged workers in the above industry

The above dataset was cleaned and comprised to four fields. The field 45-54 and 45+ was removed as it had an unequal interval compared to other fields.

2.2.2 Marriage

After the Work Sector was visualised. It was planned to visualise the marriage amongst Australian Seniors. This visualization aimed to answer the question "What is the rate of marriage amongst Australian seniors?" The dataset originally consisted of data from 2006 as well as 2010. But for my study I only considered 2010 as it has the more updated information. The fields are:

1. Married in a Registered Marriage: Number of people married in a registered marriage
2. Married in a de facto Marriage: Number of people married in a de facto marriage
3. Gender: Whether statistics for males or females are shown
4. Total Number: The total number of samples.

The above dataset was cleaned and comprised to two excel files, one for males and the other one for females. The percentages were calculated, and only married or unmarried status was considered as type of marriage was considered out of scope.

2.2.3 International

Lastly, the International distribution of the Australian Seniors was planned to be analysed. This visualisation aimed to answer the question "Which counties contribute to the Older Australian's population?". The data in this case had six fields which are described below:

1. Country: The country which is analysed.
2. 55-59: Number of 55-59 aged people from the above country
3. 60-64: Number of 60-64 aged people from the above country
4. 65-69: Number of 65-69 aged people from the above country
5. 70-74: Number of 70-74 aged people from the above country
6. 75+: Number of 75+ aged people from the above country

During data cleaning all the fields were converted to percentage-based values and rounded off to be easily readable and executed for visualization.

2.3 Data Processing:

Many of the categories have multiple tables and multiple tables have multiple attributes. All these tables would be merged and combined to form one table. Moreover, in some of the cases the data fields have percentage value, those values would be derived to form original numbers. The data processing will be implemented using Microsoft access or Microsoft excel.

It was planned that one excel file will be created with all the data but after reading files using CSV in our labs it was realized that it will be better to create multiple csv files so that it is easy to read the data. Four excel files were created, the contents of each of the files is given below

1. Percentages of age groups in each sector in 2011
2. Percentage of age groups from other countries in 2010
3. Marital status of females from age groups in 2010
4. Marital status of males from age groups in 2010

Besides conversion and cleaning data, new calculated fields were created to calculate percentages by converting numbers into percentages wherever required. All the figures were rounded off to easily visualize using d3.

3 Requirements

3.1 Must-Have Features:

3.1.1 Work Sector:

It was necessary that a scattered plot was made and successfully differentiates the age groups based on colors.

3.1.2 Marriage:

A grouped bar chart that represents males and females with different colors and shows the marriage rate for each aged group

3.1.3 International:

A map based visualisation showing the distribution of the Australian adults from countries based on their geographical location

3.2 Optional Features:

The optional features remain same for all the visualizations.

- Details on demand by using browser tooltips so that the values are showed on hovering.
- Legend highlighting the colors used for each group.
- A title of the visualisation.
- Filtering ability between age groups.
- Storing data in a SQL Database and accessing them securely.

4 Chart Justification

In the below section all the justification for choosing the chart is provided:

4.1 Work Sector (Stacked Bar Chart):

A stacked bar chart is a great way to show the parts of multiple totals. The chart successfully demonstrates the total proportion of each of the industry ad also tells the proportion for each of the three aged groups. A grouped bar hart could've also been used but that won't successfully demonstrate the total proportion in each of the industries. (Richard, Naomi 2014)

4.2 Marriage:

A grouped bar chart is good for comparing between elements in each category, and in our case, we aim to compare amongst males and females for each age group category. An alternative could've been two separate bar graphs but that disables the viewer to compare it and reach conclusion instantaneously. (Richard, Sandra 2010)

4.3 International:

Map based visualisations are used to show geographical location and corresponding information to the viewer. A map based visualisation is the best pick for this case as we have multiple countries that contributes to senior population in Australia. A bar graph could've also been used but it will have many bars which would be visually undisciplined.

5 Visualization Design:

This section describes the design used for each of the visualizations below All the visualizations are placed in the Appendix.

5.1 Work Sector:

This visualization shows a stacked bar chart to represent proportions of Australian seniors in each of the industries. Industries are nominal values, and the percentage is a ratio-based value. Length is used as a mark to represent the proportion for each industry while multi-color hue is used to represent different categories of the aged groups. The colors for the aged groups have been selected using the color scheme category10 function in D3 Library. The visualization shows that 33 % of the workers working in the Agriculture industry are amongst the old age Australians. On the other hand, the accommodation sector is the least populated by old age Australians.

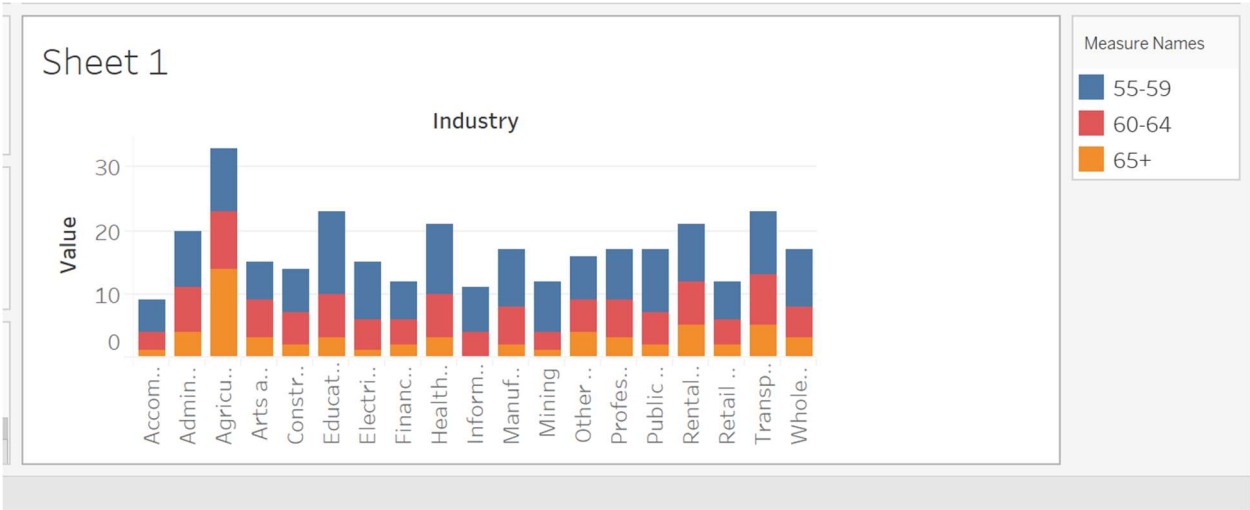
5.2 Marriage:

The visualization shows a grouped bar chart to represent proportions of Australian seniors both males and females that are married and unmarried. Gender and age group are nominal values while the percentage of marriage is a ratio-based value. Length is used as a mark to represent the proportion for each gender and each age group. Single color hue was used as a channel to differentiate between each gender.

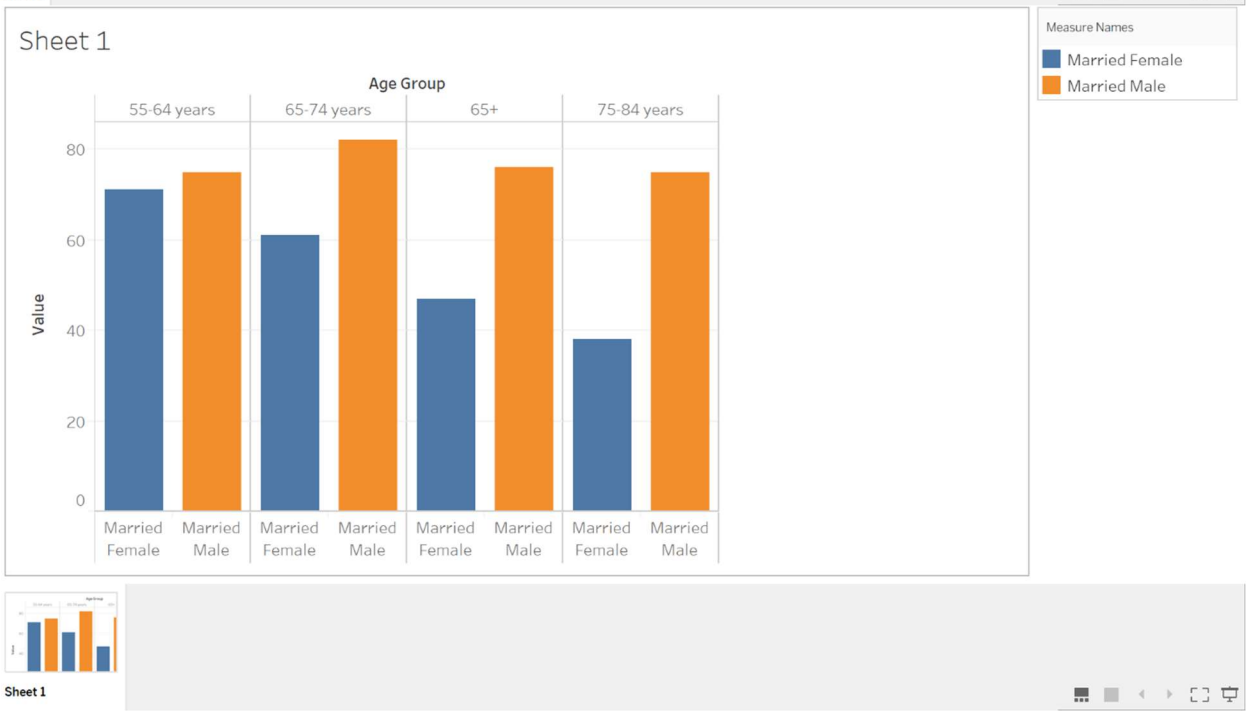
5.3 International:

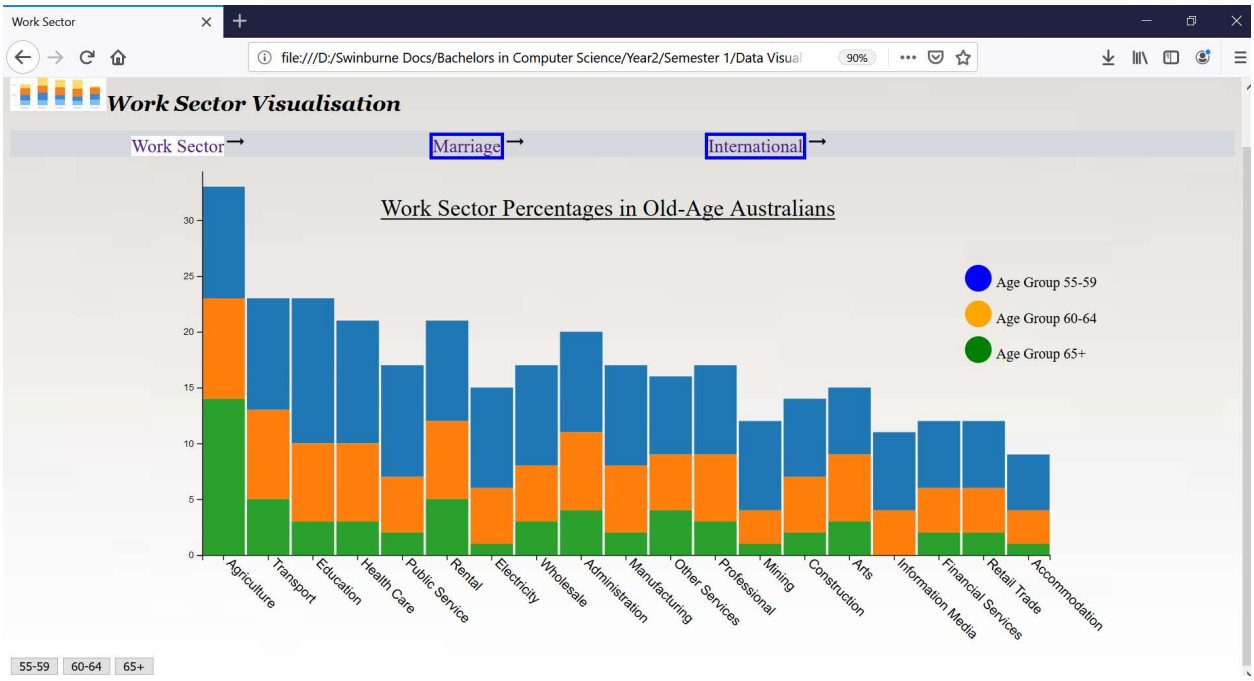
The visualisation showed a map-based display in which a point was used as a mark and area was used as a channel to display the amount of people coming into Australia from other countries. (Ann, Cees 2011)

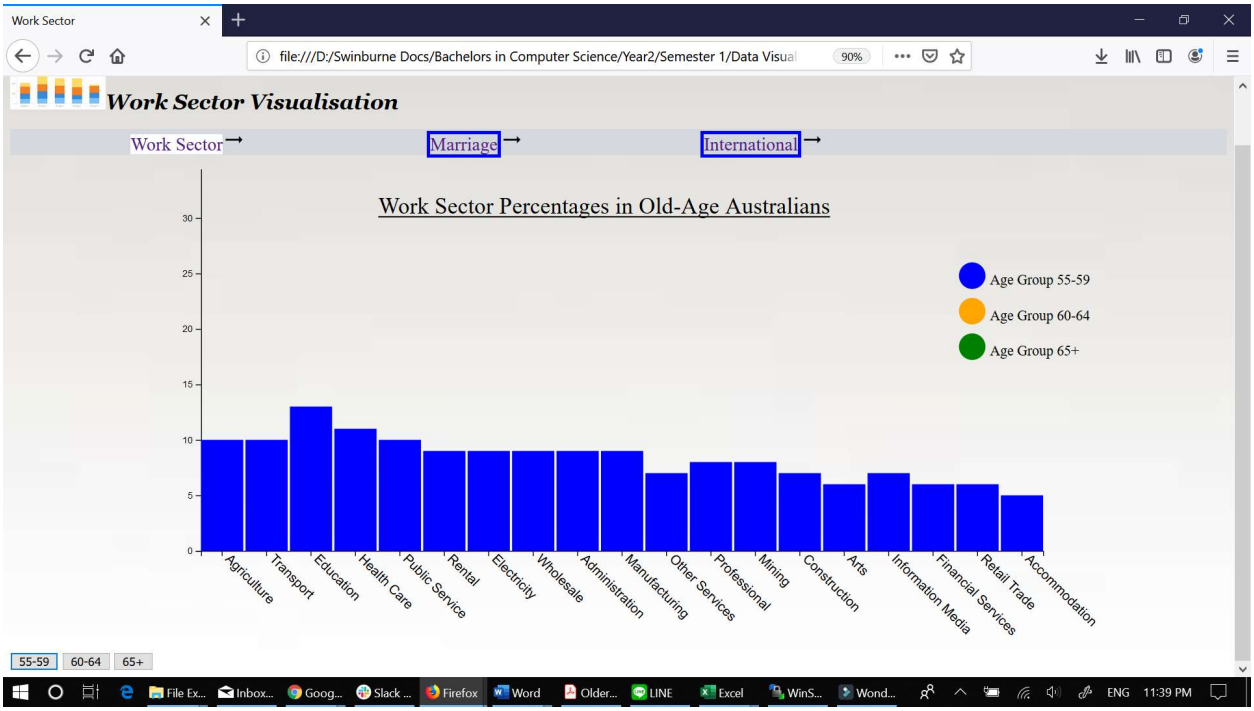
6 Appendix

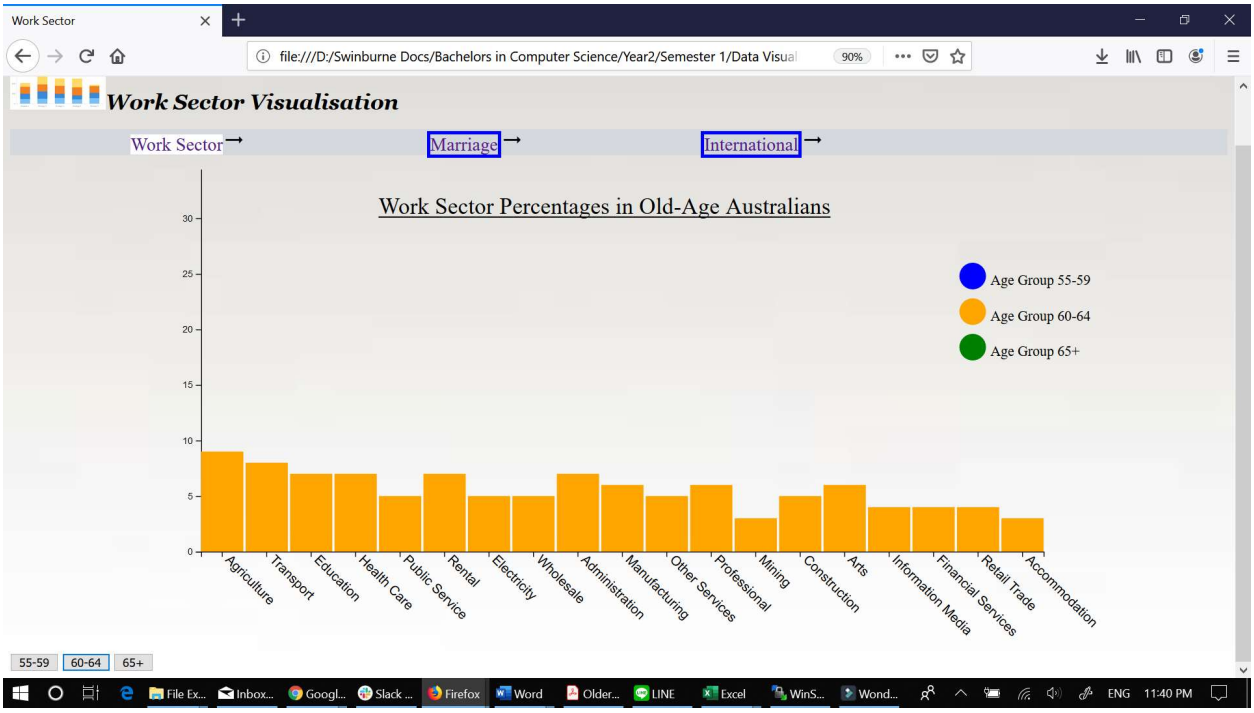

Sheet 1
Industry
Value
30
20
10
0
Accom..
Admin..
Agricu..
Arts a..
Constr..
Educat..
Electri..
Financ..
Health..
Inform..
Manuf..
Mining
Other ..
Profes..
Public ..
Rental..
Retail ..
Transp..
Whole..
Measure Names
55-59
60-64
65+

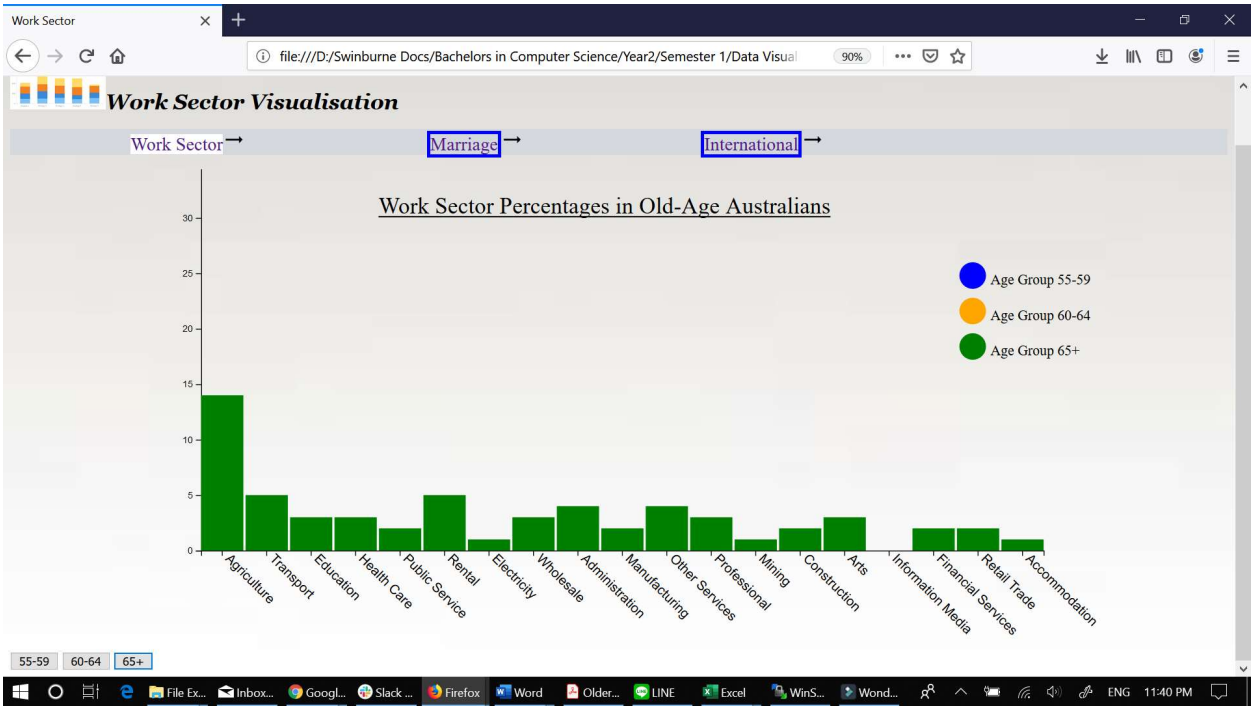

Sheet 1
Belarus
Ukraine
Kazakhstan
Mongolia
France
Spain
Turkey
Iraq
Iran
China
Japan
Algeria
Libya
Saudi Arabia
Pakistan
India
Sudan
Yemen
Democratic Republic of Congo
© OpenStreetMap contributors
3 unknown
Sheet 1


Sheet 1
Age Group
55-64 years
65-74 years
65+
75-84 years
Value
80
60
40
20
0
Married Female
Married Male
Measure Names
Married Female
Married Male
Sheet 1


Work Sector
file:///D:/Swinburne Docs/Bachelors in Computer Science/Year2/Semester 1/Data Visual
90%
Work Sector Visualisation
Work Sector
Marriage
International
Work Sector Percentages in Old-Age Australians
Age Group 55-59
Age Group 60-64
Age Group 65+
0
5
10
15
20
25
30
Agriculture
Transport
Education
Health Care
Public Service
Rental
Electricity
Wholesale
Administration
Manufacturing
Other Services
Professional
Mining
Construction
Arts
Information Media
Financial Services
Retail Trade
Accommodation
55-59
60-64
65+


Work Sector
file:///D:/Swinburne Docs/Bachelors in Computer Science/Year2/Semester 1/Data Visual
90%
Work Sector Visualisation
Work Sector
Marriage
International
Work Sector Percentages in Old-Age Australians
Age Group 55-59
Age Group 60-64
Age Group 65+
0
5
10
15
20
25
30
Agriculture
Transport
Education
Health Care
Public Service
Rental
Electricity
Wholesale
Administration
Manufacturing
Other Services
Professional
Mining
Construction
Arts
Information Media
Financial Services
Retail Trade
Accommodation
55-59
60-64
65+
ENG
11:39 PM





ScreenCapture_2019-6-3 23.30.23.mp4

7 References

7.1 Data Source

https://www.aihw.gov.au/reports/older-people/older-australia-fourth-edition/data

7.2 Repository Source

https://github.com/sreimoo/Data-Visualisation

7.3 Referred Material

- BennoEngels, 2011, Social exclusion, location and transport disadvantage amongst non-driving seniors in a Melbourne municipality, Australia Journal of Transport Geography Volume 19, Issue 4, July 2011, Pages 984-996. Available From: <https://reader.elsevier.com/reader/sd/pii/S0966692311000408?token=8B9B3F290FD0B304FD42DB2FBB276155FD7B7D60DC099C75E6C2D1A9857303B4928226EE36B85789DDC9DDBEC4BFCFAF>

- Roger Sapsford, Victor Jupp, 1996, Data Collection and Analysis. Available from: <https://books.google.com.au/books?id=euLIa2CNwtEC&dq=data+collection&lr=&source=gbs_navlinks_s>

- Richard M. Heiberger, Naomi B. Robbins, 2014, Design of Diverging Stacked Bar Charts for Likert Scales and Other Applications, Journal of Statistical Software March 2014, Volume 57, Issue 5.

- Richard Burns, Sandra Carberry, Stephanie Elzer, 2010, Visual and Spatial Factors in a Bayesian Reasoning Framework for the Recognition of Intended Messages in Grouped Bar Charts, Association for the Advancement of Artificial Intelligence

- Ann Van, Cees Van, 2011, On the argumentative work of map-based visualization, Landscape and Urban Planning