**ANL501- DATA VISUALIZATIONS AND STORY TELLING - TMA 01**

**NAME:** SRUTHI MADHUSUDANAN

**SUBMISSION DATE:** 20/02/2023

Contents

[**1 a) Types of Tabular Data** 3](#_Toc127725458)

[**Cross-sectional data** 3](#_Toc127725459)

[**Time series data** 3](#_Toc127725460)

[**Longitudinal data** 4](#_Toc127725461)

[**1 b) Deriving subsets from tma\_accidents.csv for cross-sectional,time series and longitudinal data** 5](#_Toc127725462)

[**Cross sectional Data – Subset and visualization** 5](#_Toc127725463)

[**Subset** 5](#_Toc127725464)

[**Visualization** 5](#_Toc127725465)

[**Time series – Subset and visualization:** 6](#_Toc127725466)

[**Subset** 6](#_Toc127725467)

[**Visualization** 6](#_Toc127725468)

[**Longitudinal series -subset and visualization:** 7](#_Toc127725469)

[**Subset** 7](#_Toc127725470)

[**Visualization** 8](#_Toc127725471)

[**Report** 8](#_Toc127725472)

[**Analysis** 8](#_Toc127725473)

[**Conclusion** 9](#_Toc127725474)

[**1 c) Creating an animation for time series visualization** 10](#_Toc127725475)

[**2 a) Executive summary of the Pew Research Survey Spring 2021** 10](#_Toc127725476)

[**Introduction** 10](#_Toc127725477)

[**Dataset** 11](#_Toc127725478)

[**Conclusion** 11](#_Toc127725479)

[**2 b) Constructing data subset containing Q1 and Q2 and all variables related to Singapore.** 11](#_Toc127725480)

[**Dataset** 11](#_Toc127725481)

[**Coding procedure** 11](#_Toc127725482)

[**Code** 12](#_Toc127725483)

[**2 c) Attitudes in Singapore towards the current economic situation and future prospects for the younger generations** 13](#_Toc127725484)

[**Current economic situations and its relationships in Singapore** 13](#_Toc127725485)

[**Future outlook of younger generation for younger generation and its relationships in Singapore** 14](#_Toc127725486)

[**Report** 16](#_Toc127725487)

[**Introduction** 16](#_Toc127725488)

[**Analysis** 16](#_Toc127725489)

[**Influence of income on economic situations in New Zealand and Australia** 16](#_Toc127725490)

[**Conclusion** 18](#_Toc127725491)

[**Embedded Code** 18](#_Toc127725492)

[**Appendix** 19](#_Toc127725493)

[**R Code – Full** 19](#_Toc127725494)

[**Figures** 25](#_Toc127725495)

**ROAD ACCIDENTS OF SINGAPORE DATASET**

# **1 a) Types of Tabular Data**

## **Cross-sectional data**

Data is gathered by simultaneously observing many factors. Based on the variables being investigated, researchers select a group from a broad population or sample group, and a survey is then carried out. This provides you with an overview of the topic being studied. The field period is the period during which the survey is done. Time is not taken into consideration as a study variable in these investigations.

* Visualizing cross sectional data, helps us to look at numerous characteristics simultaneously thereby understand the prevailing conditions as it contains data for a particular period.
* Since the goal of a cross-sectional study is to confirm or refute a hypothesis or theory, we may easily arrive at a conclusion through the visualization of data.
* the data researchers collect from this process is immediately relevant so through visualization we will have an opportunity to create real-time updates within specific population groups.
* The data that researchers gather through this procedure is immediately applicable, giving us the chance to use visualization to produce real-time updates within population groupings.

**Example:** Consider the following scenario. Assume you want to determine the current blood pressure levels in a population, say for year 2022. From that population, 1000 people will be chosen at random. It is also known as a cross section of that population range. Their blood pressure will now be measured along with height, weight, and other health factors which gives you an overview of that population. This information will only show the current proportion of blood pressure levels. You can't determine whether the rate of blood pressure rise is low or high based on just one cross-sectional sample. It will, however, give you an idea of the situation.

## **Time series data**

Data that is recorded over time in a time series at regular intervals is known as a time series. Typically, this data tracks a single subject's observations throughout time. The observations could be recorded for any quantum of time (E.g., hour, day, month, year etc.)

* In time series data, the timing and order of the observations are critical since the values of the following observations could affect the value of the current observation.
* In time series graphs, the visual progression of counts or numerical values over time can be demonstrated.
* A time series data graph can be used to forecast future values of the variable(s) being monitored, which can be useful for planning and decision-making by examining prior trends and patterns in the data.
* Visualization helps us to understand the impact of events over time.

**Example:** Changes in stock prices are tracked. In finance, a time series monitors the movement of data points over time, such as the price of an asset, with data points being recorded at regular intervals. This can be monitored over a small period of time (such as a security's price at each hour throughout a business day) or a lengthy period of time (such as the price of a security at close on the final day of each month over a period of five years).

## **Longitudinal data**

Longitudinal data is provided by repeated observations of the same things at different points in time (often referred to as panel data).

* Visualizations help us to analyze variations in the target population's characteristics at both the group and individual levels.
* We could establish the correct sequence of events, identify changes over time, and provide insight into cause-and-effect relationships.

**Example:** Consider a study done to compare and contrast identical twins who were raised together against those who were not, and to identify any parallels or differences. The fact that every participant has identical twins remains a constant throughout the study's observation of many factors.

In this instance, researchers would wish to follow these subjects from infancy to adulthood in order to better understand how upbringing in various environments affects characteristics, routines, and personalities.

For a long period of time, researchers can observe both sets of twins as they navigate life without assistance. The idea that any genetic variations between the participants are the result of environmental circumstances, but only a careful study can draw those conclusions.

# **1 b) Deriving subsets from tma\_accidents.csv for cross-sectional,time series and longitudinal data**

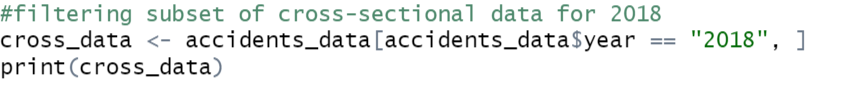
## **Cross sectional Data – Subset and visualization**

For cross sectional data, below are the data subset and visualization.

### **Subset**

**cross\_data** - Subset containing data for year 2018 across the following dimensions, year, accident\_classification, road\_user\_group, number\_of\_accidents and causes\_of\_accident.

We have examined the road user categories for the year 2018 in a cross section study to determine the causes and severity of accidents.



### **Visualization**

**A picture containing timeline

Description automatically generated**

**Fig 1 - Cross sectional analysis for 2018 by causes of accidents and user groups**

Chart, bar chart

Description automatically generated

**Fig 2 - Cross sectional analysis for 2018 by severity and user groups**

## **Time series – Subset and visualization:**

For time series analysis, below are the data subset and visualization.

### **Subset**

**time\_series** - Subset containing data for years across 2012-2018 across the following dimensions, year, accident\_classification, road\_user\_group ,number\_of\_accidents.

We have chosen the data range between the years of 2012 and 2018 for time series analysis. For this subset, we have examined the trends for drivers and pedestrians as well as the severity of the accidents from 2012 to 2018.

### **Visualization**

***Chart, line chart

Description automatically generated***

**Fig 3 - Time series analysis for number of accidents between 2012-2018 in SG**

***Chart, scatter chart

Description automatically generated***

**Fig 4 - Time series analysis for road user group and accident severity across 2012-2018**

## **Longitudinal series -subset and visualization:**

For longitudinal analysis, below are the data subset and visualization.

### **Subset**

**long\_data** - Subset containing data for years across 2012-2018 across the following dimensions, year, accident\_classification, road\_user\_group ,number\_of\_accidents.

Text

Description automatically generated

We have used the data range between the years of 2012 and 2018 for the longitudinal study. For this subset, we have examined the trend for road user groups and the severity of their accidents from 2012 to 2018.

### **Visualization**

A picture containing chart

Description automatically generated

**Fig 5 Longitudinal analysis for accidents across road user group and accident severity for 2012-2018**

## **Report**

Singapore is a city-state renowned for its excellent living standards, effective government, and cutting-edge technical infrastructure. Singapore is known for its safety and orderliness, but road accidents do take place here. However, the number of accidents has decreased in recent years.

### **Analysis**

These broad observations are based on the analysis above.

1. There has been a general decline in accidents in Singapore. As we can see, there has been a **33% decrease in the number of accidents** from 2012 in 2018.
2. Most accidents result in injuries rather than fatalities. On an average year, accidents result in **96% injury cases** and only **4% fatalities**.
3. The **number of fatal accidents is also falling** dramatically, **at over 50%** in 2018 compared to 2012. This is another encouraging development. This is highly encouraging because fatal accidents have the largest negative effects on people's life.

Let's now analyze the various road user categories in-depth using the time series & longitudinal visualizations. There are two categories of road users:

1. Drivers, Riders, and Cyclists
2. Pedestrians

From **Fig. 4 & 5,** it is clear that the drivers, riders, and cyclists group makes up the majority of the accident contributors (**93%)** and that this is to be expected. About **97%** of the incidents in this category result in injuries, most often as a result of carelessness or a gap in focus while driving. It's also important to note that the **injury trend (blue line) and fatality trend (red line) are both declining**, indicating that people are becoming more alert and conscious when driving. This means that there has been a **34% decline in accidents that resulted in injuries and a nearly 50% fall in accidents that resulted in fatalities**, which is a positive shift.

Thanks to Singapore's clearly laid out pedestrian traffic regulations, there are far fewer accidents involving pedestrians than they are with drivers, riders, and cyclists. The percent of fatal accidents differs significantly between the two categories, which is interesting. The **average number of fatal accidents involving pedestrians is 5%, which is greater than the other group's average of 3%.** Similar to other groups, pedestrians have had a downward trend in total injury and fatal accident trends, which is very encouraging.

Let's examine the causes of accidents in these categories using 2018 as our reference year and the cross-sectional data in **Fig 1**.

Three significant factors that contribute to accidents among drivers, riders, and cyclists are:

1. Failing to have a proper lookout.
2. Failing to have proper control.
3. Failing giving way to traffic with right of the way

It is obvious that this **group's carelessness and lack of focus** are the primary culprits of most of the incidents.

Although there are no clear warning signs for pedestrians, some of the main causes include.

1. Crossing when Red Man Light is displayed
2. Cross in between Vehicles with obstructed view
3. Other causes attributed to pedestrians.

In this group, most accidents happen **because people lack patience** and try to cross the road more quickly by disobeying the law.

### **Conclusion**

Overall, these **results point to a positive trend towards increasing awareness and diligence** among Singapore's road users, which has led to a decrease in fatalities and accidents over time. Nonetheless, there is always potential for development, particularly among motorists, bikers, and riders, to lower accidents brought on by negligence and lack of concentration. To decrease accidents brought on by disobedience, it is also essential to keep advocating for pedestrian safety and educating the general public on the value of following traffic regulations.

# **1 c) Creating an animation for time series visualization**

****

Graphical user interface, chart

Description automatically generated with medium confidence

**Fig 6 - Time series analysis gif for road user group and accident severity across 2012-2018**

**PEW RESEARCH SURVEY SPRING 2021**

# **2 a) Executive summary of the Pew Research Survey Spring 2021**

## **Introduction**

The impartial **Pew Research Center** educates people about the topics, attitudes, and trends shaping the globe through public opinion polling, demographic research, content analysis, and other data-driven social science research. This factual basis supports informed decision-making and enhances public discourse.

In the **spring of 2021**, Pew Research Center conducted a survey of the public to determine their opinions on a variety of crucial issues. Several nations, such as **Australia, Belgium, Canada, France, Germany, Greece, Italy, Netherlands, New Zealand, Singapore, South Korea, Spain, Sweden, Taiwan, and the United Kingdom**

## **Dataset**

The dataset for the Spring 2021 Pew Research Survey has **255 variables (Column) and 16254 observations (variables)**

* The answers provided by the responders to a particular inquiry are referred to as the "observation."
* Different questions that are asked in various countries make up the variables.

Say for example: consider the first question Q1.

***ASK ALL***

***Q1: Thinking about our economic situation, how would you describe the current economic situation in (survey public) – is it very good, somewhat good, somewhat bad or very bad?***

1. ***Very good***
2. ***Somewhat good***
3. ***Somewhat bad***
4. ***Very bad***
5. ***Don’t know (DO NOT READ)***
6. ***Refused (DO NOT READ)***

Here, respondents' attitudes on the state of the economy are surveyed, and they are asked to express their opinions based on the aforementioned possibilities. Here, not all respondents from all countries are asked every question. As seen by the **ASK ALL** indicator in the example above, Q1 is asked to all responders across all countries.

## **Conclusion**

To conclude Pew Research's spring 2021 dataset focuses on public opinion polling on a variety of topics, including the economy, children's future prospects, religion, identity, and income. It develops relationships between these variables and analyzes the effects of these variables on one another.

# **2 b) Constructing data subset containing Q1 and Q2 and all variables related to Singapore.**

## **Dataset**

Pew Research Center Global Attitudes Spring 2021 Survey Dataset

**Q1 relates to current economic situation.**

**Q2 relates to Future outlook of younger generation.**

## **Coding procedure**

**Primary Goal:** Create the subset comprising Q1, Q2, and all Singapore-related variables.

### **Code**

* Let’s start by obtaining the current working directory with **getwd ()** function.
* We use **setwd ()** function to set the working directory to the location where **pew\_dataset.sav** dataset is stored.
* **Foreign package** is loaded using **library ()** function to read spss file.
* **Tidyverse package** is installed and loaded using **install.packages ()**
* Dplyr and ggplot2 packages which are a part of the tidyverse package is loaded using library () function.
* **Read.spss ()** function from foreign package is used to read **pew\_dataset.sav** dataset and it is stored into a dataframe called df. **df contains 16254 observation and 255 variables**.
* Using pipe operator %>% lets assign df to sg\_df. Now sg\_df contains all the data which is contained in df .sg\_df contains **16254 observations and 13 variables**.
* Select () is used to select columns ID, ECON\_SIT(Q1), CHILDREN\_BETTEROFF2(Q2), region\_singapore,D\_RELIG\_SINGAPORE,D\_EDUC\_SINGAPORE,D\_INCOME\_SINGAPORE,D\_IDENTITY\_SINGAPORE,D\_INCOME2\_SINGAPORE,PARTYFAV\_SINGAPORE\_PAP,PARTYFAV\_SINGAPORE\_WP,D\_PTYID\_PROXIMITY\_SINGAPORE,public.sg\_df contains **16254 observations and 13 variables.**
* We use the **%in% operator to select only the rows with public = Singapore** and save them in pew\_singapore in order to subset the sg\_df data frame.
* **There will be 13 variables and 1124 observations in pew\_singapore and contains only respondents from Singapore.**
* Using print () pew\_singapore is printed which contains 1124 observations and 13 variables.

(Refer the Fig 7 Code snippet below)

Graphical user interface, text, application, email

Description automatically generated

**Fig 7 Code snippet for data preprocessing steps mentioned above.**

Alternatively, a respondents' affiliation with a certain country could be inferred through a thorough review of the dataset. Hence, when Singapore respondents are analyzed, the ID column's range is **6700013 to 6721593**. We can similarly use the in operator to select the range of rows from **ID 6700013 : 6721593**

Text

Description automatically generated

# **2 c) Attitudes in Singapore towards the current economic situation and future prospects for the younger generations**

For the analysis, we try to compare the responses for Q1 ( Economic Situation ) and Q2 ( Future prospects ) across different variables such as Religion , Education , Income and Identity.

## **Current economic situations and its relationships in Singapore**

Chart

Description automatically generated**Fig 8 - Current economic situation w.r.t to religion in Singapore**

Chart, bar chart

Description automatically generated

**Fig 9 - Current economic situation w.r.t to Education in Singapore**

Chart, histogram

Description automatically generated

**Fig 10 - Current economic situation w.r.t to Income in Singapore**

Chart, bar chart

Description automatically generated

**Fig 11 - Current economic situation w.r.t to Race in Singapore**

## **Future outlook of younger generation for younger generation and its relationships in Singapore**

A picture containing chart

Description automatically generated

**Fig 12 – Future outlook of younger generation w.r.t to education in Singapore**

Chart, histogram

Description automatically generated

**Fig 13 - Future outlook of younger generation w.r.t to income in Singapore**

Chart

Description automatically generated**Fig 14 - Future outlook of younger generation w.r.t to Religion in Singapore**

Chart, bar chart, waterfall chart

Description automatically generated

**Fig 15 - Future outlook of younger generation w.r.t to Race in Singapore**

## **Report**

### **Introduction**

The majority of responses on **Singapore's economic situation are positive, with somewhat good, somewhat bad, and very good coming in order of importance**. This demonstrates that the majority of respondents are content with the state of the economy, while the percentages for Modestly Bad and Very Good are nearly equal. Only a few respondents offered their opinions on Very Bad, but even those were quite few, suggesting Singapore's promising economic outlook.

### **Analysis**

#### **Current Economic situation – Singapore**

The next stage is to compare the influence of various variables in an effort to delve deeper. We can observe how several variables relate to the responses to the economic condition from Fig 8, 9, 10,11.

The majority of respondents are Buddhist, Muslim, Chinese, and Malay, which represents Singapore's native population. As would be predicted, these groups tend to have a largely optimistic attitude on the country's economic status (Figs. 8 and 11).

On the other side, when we look at the Education (Fig. 9) and Income (Fig. 10) charts, we can observe that the majority of students in universities, polytechnics, junior colleges, and early to mid-career professionals with monthly salaries under 11K SGD provided positive responses.

The responses from senior executives, who typically make more than 17K SGD per month, should be highlighted in Fig. 10 Chart. It's interesting to note that high-earning professionals' opinions of the state of the economy are evenly divided among Slightly Good, Somewhat Bad, Very Good, and Very Bad. This might be because top executives don't inclined to be unduly optimistic because they have a solid understanding of the actual economics.

#### **Future outlook for younger generation – Singapore**

The only two response options in Q2 (Future Prospects), in contrast to Q1 (Economic Situation), are Better off and Worse off, as the other two have little responses. As seen in Fig.12-15, it's interesting to note how closely the responses for future prospects are split between the two factors, with Better off coming out on top. This demonstrates that future prospects are generally seen as being positive, albeit marginally.

The same respondent categories, such as Native population, students, and early to mid-career professionals, are the ones who have the most influence on the survey with numerous responses for Q2 when we dig deeper across multiple variables, much like Q1 had.

### **Influence of income on economic situations in New Zealand and Australia**

One of the most direct and interesting influencers was income on economic situation as we discussed above.

As next step, we will analyze the impact of income on responses across two different countries New Zealand and Australia, which has very similar economics as Singapore.

#### **Economic Situation – ANZ vs Singapore**

Chart, bar chart

Description automatically generated

**Fig 16 - Current economic situation w.r.t to Income in New Zealand**

Chart, bar chart

Description automatically generated

**Fig 17 - Current economic situation w.r.t to Income in Australia**

Fig 16 and 17 show that, in contrast to Singapore, ANZ, particularly New Zealand, has significantly more responses for Somewhat Good, followed by Somewhat Bad and Very Good.

Much like Singapore, this group is heavily influenced by early to mid-career professionals who earn less than $85,000 annually. This demonstrates that the majority of respondents believe that the economy is doing well overall.

The responses from senior executives who earn more than $180,000 per year, however, represent a significant change. In contrast to Singapore, senior executives' outlooks are generally favorable in ANZ.

This could be as a result of the fact that the impact of COVID on the economy is still more pronounced in Singapore than in ANZ. ANZ was among the first nations to open its borders and economy. Whereas in Singapore, border closures took place for a significantly longer period of time and had a significant negative impact on the economy, as seen by the sharp increase in rental and commodities costs as well as the hiring of foreign expertise. Whereas ANZ with opened borders was able to dilute the impact little earlier and these countries are planning on large scale migration programs to source high qualified foreign talents .

Senior executives think long term in terms of the economy, in contrast to early to mid-career professionals who think short term. As a result, we could see the difference in their responses across Singapore and ANZ, as opposed to early to mid-career professionals, whose responses were similar across Singapore and ANZ.

### **Conclusion**

Overall, the majority of Singaporean and ANZ respondents are optimistic about the economy's present state and future prospects. Senior executives in Singapore and ANZ, however, have a more nuanced perspective of the economy as a result of their knowledge and expertise in actual economics. For futuristic, it is critical for policymakers and business executives to comprehend the perspectives of people in various income categories and professions.

# **Embedded Code**

****

**Appendix**

## **R Code – Full**

At times embedded R object doesn’t open in MacBook. Hence attaching the full code for reference

**############## Cross sectional- subset and visualization#############**

**#current directory**

**getwd()**

**#setting the working directory**

**setwd("C:/Users/user/Desktop/anl 501/TMA")**

**library(dplyr)**

**library(ggplot2)**

**library(graphics)**

**#read CSV file**

**df <- read.csv("TMA\_accidents.csv")**

**accidents\_data <- df[,c("year","accident\_classification","road\_user\_group",**

**"number\_of\_accidents","causes\_of\_accident")]**

**#filtering subset of cross-sectional data for 2018**

**cross\_data <- accidents\_data[accidents\_data$year == "2018", ]**

**print(cross\_data)**

**#omit NA**

**cross\_section <- na.omit(cross\_data)**

**print(cross\_section)**

**#visualize through barplot**

**p <- ggplot(cross\_section, aes(x = accident\_classification,**

**y = number\_of\_accidents,**

**fill = road\_user\_group))**

**p + geom\_bar(position = "dodge", stat = "identity",group=1 )+**

**scale\_y\_continuous(limits = c(0, 2000), breaks = seq(0, 2000, 250)) +**

**theme(legend.position = "right",legend.box = "vertical") +**

**labs(x="Classification of Accidents",y= "Number of Accidents",**

**title = "Road accidents in singapore in 2018",**

**subtitle = "Breakdown by Severity of accidents and User Group",**

**caption="Source:TMA\_Accidents" )+ theme(plot.title = element\_text(size=14),**

**plot.subtitle = element\_text(size = 12)) +**

**scale\_fill\_discrete(name = "Road Users")**

**##################Time series - subset and visualization##############**

**#current working directory**

**getwd()**

**#setting the working directory**

**setwd("C:/Users/user/Desktop/anl 501/TMA")**

**getwd()**

**#load packages**

**install.packages("gganimate")**

**install.packages("gifski")**

**library(dplyr)**

**library(ggplot2)**

**library(graphics)**

**library(gganimate)**

**library(gifski)**

**#install.packages("transformr")**

**#read the TMA\_accidents dataset**

**accidents <- read.csv("TMA\_accidents.csv")**

**accidents <- na.omit(accidents)**

**#select year, accident classification,number of accidents column**

**accidents\_1 <- accidents[,c("year","accident\_classification",**

**"road\_user\_group","number\_of\_accidents")]**

**print(accidents\_1)**

**# group the data by year and number of accidents**

**time\_series <- accidents\_1 %>%**

**group\_by(year, accident\_classification, road\_user\_group) %>%**

**summarize(total\_accidents = sum(number\_of\_accidents))**

**print(time\_series)**

**# visualise by line chart**

**p<-ggplot(time\_series, aes(x = year, y = total\_accidents,**

**color = accident\_classification,**

**linetype = road\_user\_group)) +geom\_line() +**

**scale\_y\_continuous(limits = c(0, 7000), breaks = seq(0, 7000, 1000))+**

**geom\_text(aes(label = total\_accidents), hjust = -0.1, size = 3, nudge\_y = 125) +**

**labs(x = "Year", y = "Accidents",**

**color = "Severity of Accidents ",**

**linetype = "Road User Category",**

**title="Number of Accidents 2012-2018",**

**subtitle="Breakdown by Severity of accidents and User Group",**

**caption="Source:TMA\_accidents") +**

**theme(legend.position = "right",**

**legend.box = "vertical",**

**plot.title = element\_text(hjust = 0))+**

**theme(plot.title = element\_text(size=14),**

**plot.subtitle = element\_text(size = 12)) +**

**scale\_fill\_discrete(name = "Road Users")**

**p+facet\_wrap(~road\_user\_group )+geom\_point()**

**#################Longitudinal data - subset and visualization############**

**#current working directory**

**getwd()**

**#setting the working directory**

**setwd("C:/Users/user/Desktop/anl 501/TMA")**

**#load packages**

**library(dplyr)**

**library(ggplot2)**

**library(graphics)**

**#read the TMA\_accidents dataset**

**df <- read.csv("TMA\_accidents.csv")**

**accidents <- na.omit(df)**

**#select year, accident classification,number of accidents, road user group**

**accidents\_1 <- accidents[,c("year","accident\_classification",**

**"road\_user\_group","number\_of\_accidents")]**

**print(accidents\_1)**

**# group the data by year and number of accidents**

**long\_data <- accidents\_1 %>%**

**group\_by(year, accident\_classification, road\_user\_group) %>%**

**summarize(total\_accidents = sum(number\_of\_accidents))**

**print(long\_data)**

**#visualize by line chart**

**ggplot(long\_data, aes(x = year, y = total\_accidents,**

**color = accident\_classification, linetype = road\_user\_group)) +**

**geom\_line() + scale\_y\_continuous(limits = c(0, 7500), breaks = seq(0, 7500,1000))+**

**geom\_text(aes(label = total\_accidents), hjust = -0.1, size = 3, nudge\_y = 100) +**

**labs(x = "Year", y = "Number of Accidents", color = "Severity",**

**linetype = "Road User Category", title="Number of Accidents 2012-2018",**

**caption = " Source: TMA\_accidents",**

**subtitle="Breakdown by User groups and severity ") +**

**theme(legend.position = "right",**

**legend.box = "vertical")+theme(plot.title = element\_text(size=14),**

**plot.subtitle = element\_text(size = 12))**

**##################### Time series - gif ##############################**

**#current working directory**

**getwd()**

**#setting the working directory**

**setwd("C:/Users/user/Desktop/anl 501/TMA")**

**getwd()**

**#load packages**

**install.packages("gganimate")**

**install.packages("gifski")**

**library(dplyr)**

**library(ggplot2)**

**library(graphics)**

**library(gganimate)**

**library(gifski)**

**#install.packages("transformr")**

**#read the TMA\_accidents dataset**

**accidents <- read.csv("TMA\_accidents.csv")**

**accidents <- na.omit(accidents)**

**#select year, accident classification,number of accidents column**

**accidents\_1 <- accidents[,c("year","accident\_classification",**

**"road\_user\_group","number\_of\_accidents")]**

**print(accidents\_1)**

**# group the data by year and number of accidents**

**time\_series\_anim <- accidents\_1 %>%**

**group\_by(year, accident\_classification, road\_user\_group) %>%**

**summarize(total\_accidents = sum(number\_of\_accidents))**

**print(time\_series\_anim)**

**# visualise by line chart**

**p<-ggplot(time\_series\_anim, aes(x = year,**

**y = total\_accidents,**

**color = accident\_classification,**

**linetype = road\_user\_group)) +**

**geom\_line() + scale\_y\_continuous(limits = c(0, 7000), breaks = seq(0, 7000, 1000))+**

**geom\_text(aes(label = total\_accidents), hjust = -0.1, size = 3, nudge\_y = 125) +**

**labs(x = "Year",**

**y = "Accidents",**

**color = "Severity of Accidents ",**

**linetype = "Road User Category",**

**title="Number of Accidents 2012-2018 in Singapore",**

**subtitle="Breakdown by Severity of accidents and User Group",**

**caption=" Source: TMA\_accidents") +**

**theme(legend.position = "right",**

**legend.box = "vertical",**

**plot.title = element\_text(hjust = 0))+**

**theme(plot.title = element\_text(size=14), plot.subtitle = element\_text(size = 12))**

**p+facet\_wrap(~road\_user\_group )+geom\_point()+transition\_reveal(year)**

**###########################Subset of Q1,Q2 and all variables related to singapore############**

**#Constructing the subset containing Q1 , Q2 and all variables related to Singapore**

**#current working directory**

**getwd()**

**#setting the working directory**

**setwd("C:/Users/user/Desktop/anl 501/TMA/DATASET")**

**#loading the foreign package**

**library(foreign)**

**#install tidyverse**

**install.packages("tidyverse")**

**library(dplyr)**

**library(ggplot2)**

**#read spss data file**

**df<- read.spss('C:/Users/user/Desktop/anl 501/TMA/DATASET/pew\_dataset.sav',**

**use.value.labels =TRUE,**

**to.data.frame = T)**

**sg\_df <- df %>%**

**# select Q1 , Q2 and all variables related to Singapore**

**select(ID, ECON\_SIT, CHILDREN\_BETTEROFF2,region\_singapore,**

**D\_RELIG\_SINGAPORE,D\_EDUC\_SINGAPORE, D\_INCOME\_SINGAPORE,**

**D\_IDENTITY\_SINGAPORE,D\_INCOME2\_SINGAPORE,PARTYFAV\_SINGAPORE\_PAP,**

**PARTYFAV\_SINGAPORE\_WP,D\_PTYID\_PROXIMITY\_SINGAPORE,public)**

**#pull the rows related to Singapore**

**pew\_singapore <- sg\_df[sg\_df$public %in% c(“Singapore”),]**

**#print the subset**

**print(pew\_singapore)**

**#################### Economic situation and its relationship to Religion###########**

**#current working directory**

**getwd()**

**# set working directory**

**setwd("C:/Users/user/Desktop/anl 501/TMA/DATASET")**

**#load packages**

**library(foreign)**

**install.packages("tidyverse")**

**library(dplyr)**

**library(ggplot2)**

**#read spss file**

**df<- read.spss('C:/Users/user/Desktop/anl 501/TMA/DATASET/pew\_dataset.sav',**

**use.value.labels =TRUE, to.data.frame = T)**

**sg\_df <- df %>%**

**#select the required columns**

**select(ID, ECON\_SIT, CHILDREN\_BETTEROFF2,region\_singapore,public,D\_RELIG\_SINGAPORE,**

**D\_EDUC\_SINGAPORE,D\_INCOME\_SINGAPORE,D\_IDENTITY\_SINGAPORE)**

**#filter the datas of singapore**

**ecorelig\_sg <- sg\_df[sg\_df$public %in% c("Singapore"),]**

**#barplot visualization**

**ggplot(ecorelig\_sg,aes(x= ECON\_SIT,fill= D\_RELIG\_SINGAPORE ))+**

**geom\_bar(group = "D\_RELIG\_SINGAPORE",position="dodge")+**

**labs(x="Economic situation ",y="Number of Respondent",**

**title = "Current Economic situation ",**

**subtitle = "Country: Singapore",**

**fill = "Religion",**

**caption ="source: Pew Research spring survey 2021")+**

**theme(plot.title = element\_text(size = 16), plot.subtitle = element\_text(size = 14))**

**#################### Economic situation and its relationship to Identity ###########**

**#current working directory**

**getwd()**

**# set working directory**

**setwd("C:/Users/user/Desktop/anl 501/TMA/DATASET")**

**#load packages**

**library(foreign)**

**install.packages("tidyverse")**

**library(dplyr)**

**library(ggplot2)**

**#read spss file**

**df<- read.spss('C:/Users/user/Desktop/anl 501/TMA/DATASET/pew\_dataset.sav',**

**use.value.labels =TRUE, to.data.frame = T)**

**sg\_df <- df %>%**

**#select the required columns**

**select(ID, ECON\_SIT, CHILDREN\_BETTEROFF2,region\_singapore,public,D\_RELIG\_SINGAPORE,**

**D\_EDUC\_SINGAPORE,D\_INCOME\_SINGAPORE,D\_IDENTITY\_SINGAPORE)**

**#filter the datas of singapore**

**e\_id\_sg <- sg\_df[sg\_df$public %in% c("Singapore"),]**

**#barplot visualization**

**ggplot(e\_id\_sg,aes(x= ECON\_SIT,fill= D\_IDENTITY\_SINGAPORE ))+**

**geom\_bar(group = "D\_IDENTITY\_SINGAPORE",position="dodge")+**

**labs(x="Economic situation ",y="Number of Respondent",**

**title = "Current Economic situation ",**

**subtitle = "Country: Singapore",**

**fill = "Identity",**

**caption ="source: Pew Research spring survey 2021")+**

**theme(plot.title = element\_text(size = 16), plot.subtitle = element\_text(size = 14))**

**#################### Economic situation and its relationship to Income ###########**

**#current working directory**

**getwd()**

**# set working directory**

**setwd("C:/Users/user/Desktop/anl 501/TMA/DATASET")**

**#load packages**

**library(foreign)**

**install.packages("tidyverse")**

**library(dplyr)**

**library(ggplot2)**

**#read spss file**

**df<- read.spss('C:/Users/user/Desktop/anl 501/TMA/DATASET/pew\_dataset.sav',**

**use.value.labels =TRUE, to.data.frame = T)**

**sg\_df <- df %>%**

**#select the required columns**

**select(ID, ECON\_SIT, CHILDREN\_BETTEROFF2,region\_singapore,public,D\_RELIG\_SINGAPORE,**

**D\_EDUC\_SINGAPORE,D\_INCOME\_SINGAPORE,D\_IDENTITY\_SINGAPORE)**

**#filter the datas of singapore**

**e\_income\_sg <- sg\_df[sg\_df$public %in% c("Singapore"),]**

**#barplot visualization**

**ggplot(e\_income\_sg,aes(x= ECON\_SIT,fill= D\_INCOME\_SINGAPORE ))+**

**geom\_bar(group = "D\_INCOME\_SINGAPORE",position="dodge")+**

**labs(x="Economic situation ",y="Number of Respondent",**

**title = "Current Economic situation ",**

**subtitle = "Country: Singapore",**

**fill = "Income",**

**caption ="source: Pew Research spring survey 2021")+**

**theme(plot.title = element\_text(size = 16), plot.subtitle = element\_text(size = 14))**

**#################### Economic situation and its relationship to Education ###########**

**#current working directory**

**getwd()**

**# set working directory**

**setwd("C:/Users/user/Desktop/anl 501/TMA/DATASET")**

**#load packages**

**library(foreign)**

**install.packages("tidyverse")**

**library(dplyr)**

**library(ggplot2)**

**#read spss file**

**df<- read.spss('C:/Users/user/Desktop/anl 501/TMA/DATASET/pew\_dataset.sav',**

**use.value.labels =TRUE, to.data.frame = T)**

**sg\_df <- df %>%**

**#select the required columns**

**select(ID, ECON\_SIT, CHILDREN\_BETTEROFF2,region\_singapore,public,D\_RELIG\_SINGAPORE,**

**D\_EDUC\_SINGAPORE,D\_INCOME\_SINGAPORE,D\_IDENTITY\_SINGAPORE)**

**#filter the datas of singapore**

**e\_edu\_sg <- sg\_df[sg\_df$public %in% c("Singapore"),]**

**#barplot visualization**

**ggplot(e\_edu\_sg,aes(x= ECON\_SIT,fill= D\_EDUC\_SINGAPORE ))+**

**geom\_bar(group = "D\_EDUC\_SINGAPORE",position="dodge")+**

**labs(x="Economic situation ",y="Number of Respondent",**

**title = "Current Economic situation ",**

**subtitle = "Country: Singapore",**

**fill = "Education",**

**caption ="source: Pew Research spring survey 2021")+**

**theme(plot.title = element\_text(size = 16), plot.subtitle = element\_text(size = 14))**

**#################### Future outlook of younger generation and its relationship to Income ###########**

**#current working directory**

**getwd()**

**# set working directory**

**setwd("C:/Users/user/Desktop/anl 501/TMA/DATASET")**

**#load packages**

**library(foreign)**

**install.packages("tidyverse")**

**library(dplyr)**

**library(ggplot2)**

**#read spss file**

**df<- read.spss('C:/Users/user/Desktop/anl 501/TMA/DATASET/pew\_dataset.sav',**

**use.value.labels =TRUE, to.data.frame = T)**

**sg\_df <- df %>%**

**#select the required columns**

**select(ID, ECON\_SIT, CHILDREN\_BETTEROFF2,region\_singapore,public,D\_RELIG\_SINGAPORE,**

**D\_EDUC\_SINGAPORE,D\_INCOME\_SINGAPORE,D\_IDENTITY\_SINGAPORE)**

**#filter the datas of singapore**

**child\_income\_sg <- sg\_df[sg\_df$public %in% c("Singapore"),]**

**#barplot visualization**

**ggplot(child\_income\_sg,aes(x=CHILDREN\_BETTEROFF2 ,fill= D\_INCOME\_SINGAPORE ))+**

**geom\_bar(group = "D\_INCOME\_SINGAPORE",position="dodge")+**

**labs(x="Future outlook for children ",y="Number of Respondent",**

**title = "Future outlook for children",**

**subtitle = "Country: Singapore",**

**fill = "Income",**

**caption ="source: Pew Research spring survey 2021")+**

**theme(plot.title = element\_text(size = 16), plot.subtitle = element\_text(size = 14))**

**############### Future outlook of younger generation and its relationship to Identity ###########**

**#current working directory**

**getwd()**

**# set working directory**

**setwd("C:/Users/user/Desktop/anl 501/TMA/DATASET")**

**#load packages**

**library(foreign)**

**install.packages("tidyverse")**

**library(dplyr)**

**library(ggplot2)**

**#read spss file**

**df<- read.spss('C:/Users/user/Desktop/anl 501/TMA/DATASET/pew\_dataset.sav',**

**use.value.labels =TRUE, to.data.frame = T)**

**sg\_df <- df %>%**

**#select the required columns**

**select(ID, ECON\_SIT, CHILDREN\_BETTEROFF2,region\_singapore,public,D\_RELIG\_SINGAPORE,**

**D\_EDUC\_SINGAPORE,D\_INCOME\_SINGAPORE,D\_IDENTITY\_SINGAPORE)**

**#filter the datas of singapore**

**child\_id\_sg <- sg\_df[sg\_df$public %in% c("Singapore"),]**

**#barplot visualization**

**ggplot(child\_id\_sg,aes(x=CHILDREN\_BETTEROFF2 ,fill= D\_IDENTITY\_SINGAPORE ))+**

**geom\_bar(group = "D\_IDENTITY\_SINGAPORE",position="dodge")+**

**labs(x="Future outlook for children ",y="Number of Respondent",**

**title = "Future outlook for children",**

**subtitle = "Country: Singapore",**

**fill = "Identity",**

**caption ="source: Pew Research spring survey 2021")+**

**theme(plot.title = element\_text(size = 16), plot.subtitle = element\_text(size = 14))**

**############### Future outlook of younger generation and its relationship to Religion ###########**

**#current working directory**

**getwd()**

**# set working directory**

**setwd("C:/Users/user/Desktop/anl 501/TMA/DATASET")**

**#load packages**

**library(foreign)**

**install.packages("tidyverse")**

**library(dplyr)**

**library(ggplot2)**

**#read spss file**

**df<- read.spss('C:/Users/user/Desktop/anl 501/TMA/DATASET/pew\_dataset.sav',**

**use.value.labels =TRUE, to.data.frame = T)**

**sg\_df <- df %>%**

**#select the required columns**

**select(ID, ECON\_SIT, CHILDREN\_BETTEROFF2,region\_singapore,public,D\_RELIG\_SINGAPORE,**

**D\_EDUC\_SINGAPORE,D\_INCOME\_SINGAPORE,D\_IDENTITY\_SINGAPORE)**

**#filter the datas of singapore**

**child\_relig\_sg <- sg\_df[sg\_df$public %in% c("Singapore"),]**

**#barplot visualization**

**ggplot(child\_relig\_sg,aes(x=CHILDREN\_BETTEROFF2 ,fill= D\_RELIG\_SINGAPORE ))+**

**geom\_bar(group = "D\_RELIG\_SINGAPORE",position="dodge")+**

**labs(x="Future outlook for children ",y="Number of Respondent",**

**title = "Future outlook for children",**

**subtitle = "Country: Singapore",**

**fill = "Religion",**

**caption ="source: Pew Research spring survey 2021")+**

**theme(plot.title = element\_text(size = 16), plot.subtitle = element\_text(size = 14))**

**################Future outlook of younger generation and its relationship to Education ###########**

**#current working directory**

**getwd()**

**# set working directory**

**setwd("C:/Users/user/Desktop/anl 501/TMA/DATASET")**

**#load packages**

**library(foreign)**

**install.packages("tidyverse")**

**library(dplyr)**

**library(ggplot2)**

**#read spss file**

**df<- read.spss('C:/Users/user/Desktop/anl 501/TMA/DATASET/pew\_dataset.sav',**

**use.value.labels =TRUE, to.data.frame = T)**

**sg\_df <- df %>%**

**#select the required columns**

**select(ID, ECON\_SIT, CHILDREN\_BETTEROFF2,region\_singapore,public,D\_RELIG\_SINGAPORE,**

**D\_EDUC\_SINGAPORE,D\_INCOME\_SINGAPORE,D\_IDENTITY\_SINGAPORE)**

**#filter the datas of singapore**

**child\_edu\_sg <- sg\_df[sg\_df$public %in% c("Singapore"),]**

**#barplot visualization**

**ggplot(child\_edu\_sg,aes(x=CHILDREN\_BETTEROFF2 ,fill= D\_EDUC\_SINGAPORE ))+**

**geom\_bar(group = "D\_EDUC\_SINGAPORE",position="dodge")+**

**labs(x="Future outlook for children ",y="Number of Respondent",**

**title = "Future outlook for children",**

**subtitle = "Country: Singapore",**

**fill = "EDUCATION",**

**caption ="source: Pew Research spring survey 2021")+**

**theme(plot.title = element\_text(size = 16), plot.subtitle = element\_text(size = 14))**

**#################### Economic situation in Australia and its relationship to Income ###########**

**#current working directory**

**getwd()**

**# set working directory**

**setwd("C:/Users/user/Desktop/anl 501/TMA/DATASET")**

**#load packages**

**library(foreign)**

**install.packages("tidyverse")**

**library(dplyr)**

**library(ggplot2)**

**#read spss file**

**df<- read.spss('C:/Users/user/Desktop/anl 501/TMA/DATASET/pew\_dataset.sav',**

**use.value.labels =TRUE, to.data.frame = T)**

**aus\_df <- df %>%**

**#select the required columns**

**select(ID, ECON\_SIT,public,D\_INCOME\_AUSTRALIA)**

**#filter the datas of Australia**

**eco\_aus <- aus\_df[aus\_df$public %in% c("Australia"),]**

**#barplot visualization**

**ggplot(eco\_aus,aes(x=ECON\_SIT ,fill= ,D\_INCOME\_AUSTRALIA ))+**

**geom\_bar(group = ",D\_INCOME\_AUSTRALIA",position="dodge")+**

**labs(x="Economic situation ",y="Number of Respondent",**

**title = "Economic situation",**

**subtitle = "Country: Australia",**

**fill = "Income",**

**caption ="source: Pew Research spring survey 2021")+**

**theme(plot.title = element\_text(size = 16), plot.subtitle = element\_text(size = 14))**

**#################### Economic situation in New zealand and its relationship to Income ###########**

**#current working directory**

**getwd()**

**# set working directory**

**setwd("C:/Users/user/Desktop/anl 501/TMA/DATASET")**

**#load packages**

**library(foreign)**

**install.packages("tidyverse")**

**library(dplyr)**

**library(ggplot2)**

**#read spss file**

**df<- read.spss('C:/Users/user/Desktop/anl 501/TMA/DATASET/pew\_dataset.sav',**

**use.value.labels =TRUE, to.data.frame = T)**

**nz\_df <- df %>%**

**#select the required columns**

**select(ID, ECON\_SIT,public,D\_INCOME\_NEWZEALAND)**

**#filter the datas of Australia**

**eco\_nz <- nz\_df[nz\_df$public %in% c("Newzealand"),]**

**#barplot visualization**

**ggplot(eco\_nz,aes(x=ECON\_SIT ,fill= ,D\_INCOME\_NEWZEALAND ))+**

**geom\_bar(group = ",D\_INCOME\_NEWZEALAND",position="dodge")+**

**labs(x="Economic situation ",y="Number of Respondent",**

**title = "Economic situation",**

**subtitle = "Country: Newzealand",**

**fill = "Income",**

**caption ="source: Pew Research spring survey 2021")+**

**theme(plot.title = element\_text(size = 16), plot.subtitle = element\_text(size = 14))**

**###################################################################################################**

## **Figures**

**Fig 1 - Cross sectional analysis for 2018 by causes of accidents and user groups.**

**Fig 2 - Cross sectional analysis for 2018 by severity and user groups.**

**Fig 3 - Time series analysis for number of accidents between 2012-2018 in SG.**

**Fig 4 - Time series analysis for road user group and accident severity across 2012-2018.**

**Fig 5 -Longitudinal analysis for accidents across road user group and accident severity for 2012-2018.**

**Fig 6 -Time series analysis gif for road user group and accident severity across 2012-2018.**

**Fig 7 -Code snippet for data preprocessing steps mentioned above.**

**Fig 8 - Current economic situation w.r.t to religion in Singapore**

**Fig 9 - Current economic situation w.r.t to Education in Singapore**

**Fig 10 - Current economic situation w.r.t to Income in Singapore**

**Fig 11 - Current economic situation w.r.t to Race in Singapore**

**Fig 12 – Future outlook of younger generation w.r.t to education in Singapore**

**Fig 13 - Future outlook of younger generation w.r.t to income in Singapore.**

**Fig 14 - Future outlook of younger generation w.r.t to Religion in Singapore.**

**Fig 15 - Future outlook of younger generation w.r.t to Race in Singapore**

**Fig 16 - Current economic situation w.r.t to Income in New Zealand.**

**Fig 17 - Current economic situation w.r.t to Income in Australia.**

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*END\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***