PSY6422 Module Project

Registration Number: 210155394

Data Origins

The World Health Organisation (WHO) recommends several vaccinations for one-year-olds around the world. It is important to understand how vaccinations can prevent disease, since they have greatly reduced the prevalence of diseases everywhere in the world. My visualisations in in this project will show the progression of how many one-year-olds are vaccinated in the United Kingdom (UK) from the years 1980-2019 and the worldwide percentage of twelve vaccines in one-year-olds in the year 2019.

The data I have used in this project was retrieved from https://ourworldindata.org/vaccination, where I downloaded the full data set on the number of one-year-olds vaccinated globally. The raw data set included 7708 rows and 15 columns.

The variables I will be looking at in this project will be the twelve different vaccinations. For this project, I focused on United Kingdom data and World coverage data.

```
library(tidyverse) #Attaching 'tidyverse' package
```

```
# Source of Data: World Health Organisation (WHO)/ UNICEFs Estimates of National Immunisation Coverage.
Full_Dataset <- read.csv("~/Desktop/R/Module_Project/Data/Full_Dataset.csv")
Full_Dataset_tb <- as_tibble(Full_Dataset)
head(Full_Dataset_tb, 5) #Showing the first five rows of the raw data
```

```
## # A tibble: 5 × 256
##   Entity      Code   Year BCG,...of.one.year.… Hep3,...of.one… Hib3,...of.one.…
##   <chr>       <chr> <int>                <int>           <int>           <int>
## 1 Afghanistan AFG    1982                   10              NA              NA
## 2 Afghanistan AFG    1983                   10              NA              NA
## 3 Afghanistan AFG    1984                   11              NA              NA
## 4 Afghanistan AFG    1985                   17              NA              NA
## 5 Afghanistan AFG    1986                   18              NA              NA
## # … with 250 more variables: IPV1....of.one.year.olds.immunized. <int>,
## #   MCV1....of.one.year.olds.immunized. <int>,
## #   PCV3....of.one.year.olds.immunized. <int>,
## #   Pol3....of.one.year.olds.immunized. <int>,
## #   RCV1....of.one.year.olds.immunized. <int>,
## #   RotaC....of.one.year.olds.immunized. <int>,
## #   YFV....of.one.year.olds.immunized. <int>, …
```

Research Questions

My visualisations will address the following questions:

- **In 2019, what percentage of vaccines was vaccinated worldwide?**
- **How has the percentage of vaccinated one-year-olds changed from the years 1980-2019 in the UK?**

Codebook

The table below outlines each variable that the data set that will be used in the visualisation.

Variable	Variable Description
BCG, Hepb3, Hib3, IPV1, MCV1, PCV3, Pol3, RCV1, RotaC, YFV, DTP3, MCV2	Tuberculosis, Hepatitis B, H.influenzae Type B, Inactivated Polio Vaccine, Measles First Dose, Pneumococcal Vaccine, Polio, Rubella, Rotavirus, Yellow Fever, Diptheria/Tetanus/Pertussis, Measles
Full_Dataset	The full data set (raw)
Full_Dataset_tb	The full data set (raw) as a tibble
Full	The full data set (processed)
World	The data set used for Figure 1 (processed)
UK	The data set used for Figure 2 (processed)
p	Plot 1 (Visulisation/Figure 1)
pl	Plot 2 (Visualisation/Figure 2)

Data Preparation and Processing

To prepare the data, there needs to be a set up of a working directory where all figures will be saved. The raw data set needs to be cleaned by removing columns and rows which will not be used in the data processing and setting data frames that will be used in the data visualisation.

```
#Loading the data
library(tidyverse)
setwd("~/Desktop/R/Module_Project/Figures") #Setting a working directory where the figures will be saved
Full_Dataset <- read.csv("~/Desktop/R/Module_Project/Data/Full_Dataset.csv")
Full_Dataset_tb <- as_tibble(Full_Dataset)

#Cleaning the data
#Deleting unwanted rows and columns
Full <- Full_Dataset_tb[-c(117243, 7284:7587, 7589:7706),] #Deleting rows from the unwanted Entity list
Full <- subset (Full, select = -Code) #Deleting the unwanted 'Code' column

#Renaming columns
library(dplyr)
colnames(Full) <- c("Country", "Year", "BCG", "Hepb3", "Hib3", "IPV1", "MCV1",
                    "PCV3", "Pol3", "RCV1", "RotaC", "YFV", "DTP3", "MCV2") #Renaming the columns so they are eas
ily recognised and take up less space
```

```
#Processing the data for Figure 1
World <- Full[-c(1440, 42),] # Removing the unwanted UK rows
World <- subset (World, select = -Year) # Removing the Year column
World <- subset (World, select = -Country) # removing the Country column

World <- subset (World, select = -c(13:253)) # Deleting unwanted columns

World <- t(World) # Changing the data set from wide to long data
colnames(World)<-c("Percentage") # Renaming a column

World <- as.data.frame(World) #Changing 'World' from a matrix to a data frame
```

```
head(World) #Showing the first 5 rows of the processed data
```

```
##      Percentage
## BCG          88
## Hepb3        85
## Hib3         72
## IPV1         82
## MCV1         85
## PCV3         48
```

```
#Processing the data for Figure 2
UK <- Full[-c(41),] # Removing the unwanted World row
UK <- subset (UK, select = -c(15:253))
UK <- as.data.frame(UK)
UK[UK==0]<- NA # All the NA values changed to '0' for Figure 2
```

```
head(UK, 2) #Showing the first two rows of processed data
```

```
##         Country Year BCG Hepb3 Hib3 IPV1 MCV1 PCV3 Pol3 RCV1 RotaC YFV DTP3
## 1 United Kingdom 2019  NA    93   97   91   91   93   91   90   NA  93
## 2 United Kingdom 1993  NA    NA   89   NA   91   NA   95   91   NA  NA   93
##   MCV2 NA NA
## 1   87 NA NA
## 2   NA NA NA
```

Data Visualisation

[Figure 1:]

This visualisation will show the percentage of one-year-olds who have been vaccinated worldwide in 2019. From this I can see the coverage of many vaccines is still quite low. I chose a bar chart to look at this data because it nicely shows all vaccinations alongside each other and easily shows the percentages along the x-axis and as a bar label.

```
library(ggplot2) # Loading the ggplot library
```

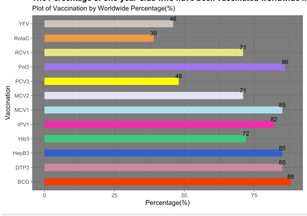
```
# geom_plot and geom_bar create a bar graph, where 'World' is the data set, with the row names on the x-axis and
percentage on the y-axis.
p <- ggplot(World, aes(rownames(World),Percentage)) +
       geom_bar(stat="identity", width=0.5,
                fill=c('orangered2', 'royalblue1', 'seagreen1','maroon2',
                       'lightblue2', 'yellow1', 'mediumpurple2', 'khaki2',
                       'tan2', 'seashell3', 'pink3', 'lavender')) + # Changing the colour of the bars.

# geom_text and ggtitle, allow the adjustment of titles.
  geom_text(aes(label=Percentage), vjust=-0.7, size=3.5,)+
  ggtitle("The Percentage of one-year-olds who have been vaccinated worldwide in 2019",
          subtitle="Plot of Vaccination by Worldwide Percentage(%)")+

# Now to change the x- and y-axis labels.
# Change the style of the title and change the theme to 'dark'.
  labs(y="Percentage(%)", x="Vaccination")+ theme_dark() + scale_fill_manual()+ theme(plot.title = element_text(f
ace="bold.italic"))+

# Changing the coordinates, so x-axis is now vertical and y-axis is horizontal.
  coord_flip()

p
```




```
ggsave("myplot1.png", plot=p) # Saving the figure as a png file
```

```
## Saving 7 x 5 in image
```

[Figure 2:]

This visualisation will show the percentages of one-year-olds in the United Kingdom who have been vaccinated from 1980 - 2019. You can see all vaccinations have increased in the amount they have been given to one-year-olds over this period of time.

I chose an animated line graph to show this data to allow a visualisation of all vaccinations at the same time for a comparison but also animation is more appealing.

```
# First, I need to install the appropriate packages to allow an animated figure to be produced.
install.packages("gganimate")
install.packages("gifski")
```

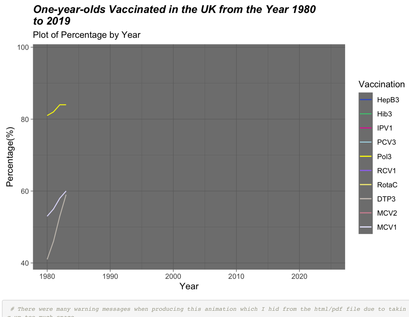
```
library(ggplot2)
library(gganimate)
```

```
# Producing a line graph for each vaccination, discarding the vaccinations with no data
pl <- ggplot() + geom_line(data=UK, mapping=aes(x=Year, y=Hepb3 ,color="Hepb3"))+
  geom_line(data=UK, mapping=aes(x=Year, y=Hib3, colour="Hib3"))+
  geom_line(data=UK, mapping=aes(x=Year, y=IPV1, colour="IPV1"))+
    geom_line(data=UK, mapping=aes(x=Year, y=PCV3, colour="PCV3"))+
  geom_line(data=UK, mapping=aes(x=Year, y=Pol3, colour="Pol3"))+
  geom_line(data=UK, mapping=aes(x=Year, y=RCV1, colour="RCV1"))+
  geom_line(data=UK, mapping=aes(x=Year, y=RotaC, colour="RotaC"))+
  geom_line(data=UK, mapping=aes(x=Year, y=DTP3, colour="DTP3"))+
  geom_line(data=UK, mapping=aes(x=Year, y=MCV2, colour="MCV2"))+
    geom_line(data=UK, mapping=aes(x=Year, y=MCV1, colour="MCV1"))+

labs(x="Year", y="Percentage(%)", title="One-year-olds Vaccinated in the UK from the Year 1980
     to 2019", subtitle=
       "Plot of Percentage by Year")+
  theme(plot.title = element_text(face='bold.italic'))+theme_dark()+ #Renaming titles and changing the theme of t
he graph
transition_reveal(Year)+ xlim(1980, 2025)+ # Changing the x-axis and including a transition in the graph

  scale_color_manual(name = "Vaccination", values = c("Hepb3" = "royalblue1",
                                                      "Hib3" = "seagreen1",
                                                      "IPV1" = "maroon2",
                                                      "PCV3" = "lightblue2",
                                                      "Pol3" = "yellow1",
                                                      "RCV1" = "mediumpurple2",
                                                      "RotaC" = "khaki2",
                                                      "DTP3" = "seashell3",
                                                      "MCV2" = "pink3",
                                                      "MCV1" = "lavender"))+
  theme(plot.title = element_text(face='bold.italic'))+ ease_aes('cubic-in-out') # Changing the colour of each l
ine and producing a legend

  pl
```




```
# There were many warning messages when producing this animation which I hid from the html/pdf file due to takin
g up too much space
```

Summary

I enjoyed working on this project and producing visualisations for the data set on worldwide vaccinations. I liked the challenge of finding advanced code for the visualisations.

Figure 1 shows the percentage of vaccinations worldwide in 2019 averaged at 71.5%. It would be interesting to find out if specific countries dramatically effected the percentages of certain vaccines and to compare the factors.

Figure 2 sows that percentage of all vaccinations have increased over time. Some vaccinations such as BCG were not included in this graph due to the vaccination not being needed anymore (e.g. tuberculosis rates being very low in the general population).

Some limitations in my project would be in my Figure 2 line graph, with the data having 'NA' and overlapping values. I would like the graph to have individual points for each value and for the lines not to be overlapping. If I were to produce this graph again, the animated line graph may not be the most suitable to display the data.