Lab 2

Student Name

YYYY-MM-DD

Section 1

In Section 1, we will focus on analyzing PISA data we used in class. This dataset has math and reading scores from PISA assessment for 10 countries.

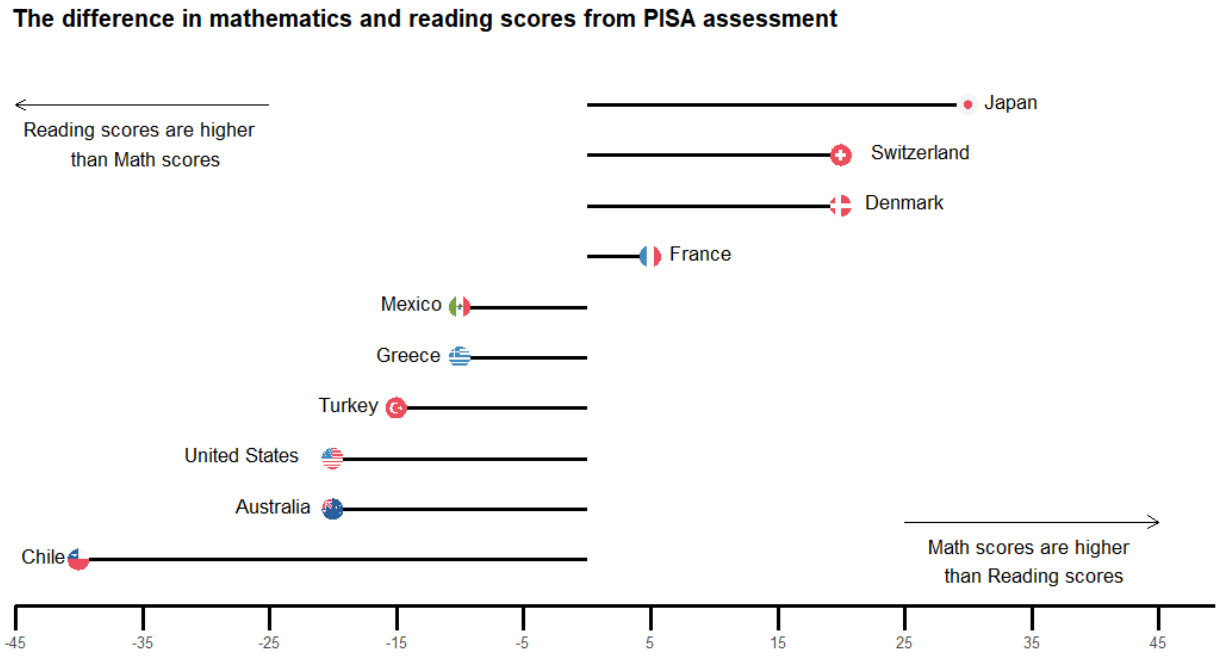
To begin working with this dataset, you are required to import and preprocess the data using the following code snippet provided below:

```
require(countrycode)

pisa      <- read.csv('pisa.csv')
pisa$iso  <- tolower(codelist[codelist$country.name.en %in% pisa$Country,]$iso2c)
pisa$diff <- pisa$Math - pisa$Reading

pisa
```

```
         Country Math Reading iso diff
1      Australia  480     500  au  -20
2          Chile  410     450  cl  -40
3        Denmark  520     500  dk   20
4         France  495     490  fr    5
5         Greece  440     450  gr  -10
6          Japan  535     505  jp   30
7         Mexico  410     420  mx  -10
8    Switzerland  510     490  ch   20
9         Turkey  435     450  tr  -15
10 United States  480     500  us  -20
```

Your task is to generate a figure that displays the difference between math and reading scores for 10 countries. The styling of the plot should closely resemble the example provided. While an exact match is not required, your plot should closely align with the given aesthetics. Please ensure your submission includes the complete R code necessary to reproduce the figure.

Hint: Please check the `ggflags` package and `geom_flag()` function to include the flags for each country as a symbol at the end of the line. https://github.com/jimjam-slam/ggflags



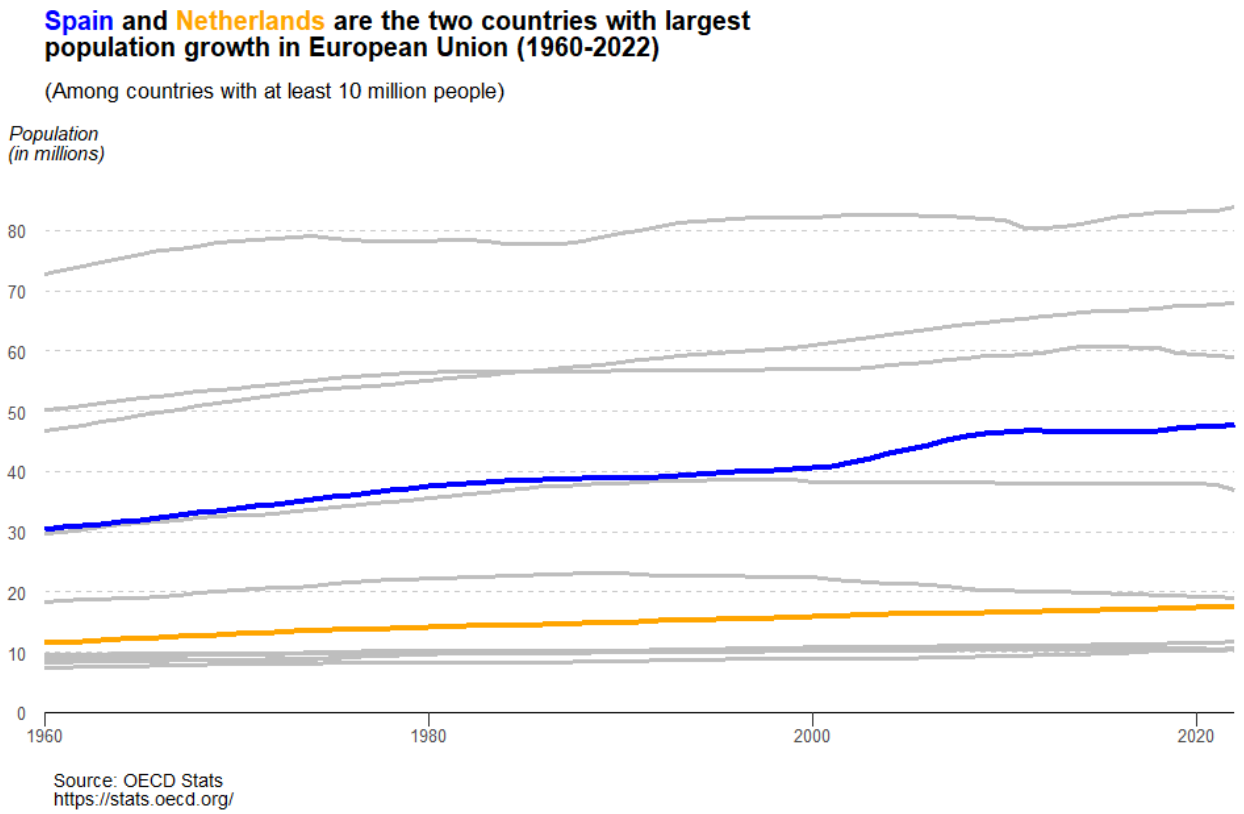
Section 2

In Section 2, our analysis will focus on world population. This dataset has the population for 266 countries from 1960 to 2022.

To begin, you should import the dataset using the following code snippet. This code will first filter the countries and include only European Union countries with at least 10 million people. It will also compute a % change in population from 1960 to 2022.

```
pop     <- read.csv('population.csv',fileEncoding="UTF-8-BOM")

EU <- c("Austria", "Belgium", "Bulgaria", "Croatia", "Cyprus",
        "Czechia", "Denmark", "Estonia", "Finland", "France",
        "Germany", "Greece", "Hungary", "Ireland", "Italy", "Latvia",
        "Lithuania", "Luxembourg", "Malta", "Netherlands", "Poland",
        "Portugal", "Romania", "Slovak Republic", "Slovenia", "Spain","Sweden")

pop <- pop[pop$Country.Name %in% EU,]
pop <- pop[which(pop$X2022>10000000),]
pop$per_change <- ((pop$X2022 - pop$X1960)/pop$X2022)*100
```

Your task is to generate a figure that displays the growth in population for these countries from 1960 to 2022, and highlight the two countries with the highest growth (Spain and Netherlands). The styling of the plot should closely resemble the example provided. While an exact match is not required, your plot should closely align with the given aesthetics. Please ensure your submission includes the complete R code necessary to reproduce the figure.



Section 3

In Section 3, our analysis will focus on the relationship between online hotel revenue and the number of travel agents over time using a connected scatterplot. To begin, you should import the dataset using the following code snippet.

```
hotel <- read.csv('hotel.csv',fileEncoding="UTF-8-BOM")
hotel$travel_agents <- hotel$travel_agents/1000

  # number of travel agents are in thousands
  # hotel revenue is in billion dollars

hotel
```

```
   year travel_agents hotel_revenue
1  2000       123.385         12.95
2  2001       110.583         19.95
3  2002       104.046         28.02
4  2003       103.501         40.12
5  2004        90.428         51.16
6  2005        88.521         64.10
7  2006        87.431         79.81
8  2007        85.252         89.79
9  2008        86.070         94.46
10 2009        76.809         90.00
11 2010        70.272         99.76
```

```
12 2011         67.276         116.11
13 2012         64.552         124.60
14 2013         64.280         143.49
15 2014         63.975         155.38
```

Your task is to generate a figure that displays the relationship between the number of travel agents and online hotel revenue over time (2000-2014). The styling of the plot should closely resemble the example provided. While an exact match is not required, your plot should closely align with the given aesthetics. Please ensure your submission includes the complete R code necessary to reproduce the figure.