Part 2 : Replicating the Visual

### Importing relavnt libraries

library(readr)

## Warning: package 'readr' was built under R version 4.0.4

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.0.4

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(tidyr)  
library(lubridate)

##   
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':  
##   
## date, intersect, setdiff, union

library(scales)

##   
## Attaching package: 'scales'

## The following object is masked from 'package:readr':  
##   
## col\_factor

library(kableExtra)

## Warning: package 'kableExtra' was built under R version 4.0.4

##   
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':  
##   
## group\_rows

library(countrycode)

## Warning: package 'countrycode' was built under R version 4.0.5

library(grid)  
library(repr)

## Warning: package 'repr' was built under R version 4.0.5

### Reading the original dataset

df <- read.csv("OECD\_PISA.csv") %>% select("ï..LOCATION","SUBJECT","TIME","Value")  
colnames(df)<-c("Location","Subject","Time","Value")  
head(df)

## Location Subject Time Value  
## 1 AUS BOY 2000 513.00  
## 2 AUS BOY 2003 506.00  
## 3 AUS BOY 2006 495.00  
## 4 AUS BOY 2009 496.00  
## 5 AUS BOY 2012 495.09  
## 6 AUS BOY 2015 487.00

### Filtering the relevant rows for this part

df2 <- df %>% filter(Time==2018) %>% filter(Subject=="BOY"|Subject=="GIRL")  
head(df2)

## Location Subject Time Value  
## 1 AUS BOY 2018 487  
## 2 AUS GIRL 2018 519  
## 3 AUT BOY 2018 471  
## 4 AUT GIRL 2018 499  
## 5 BEL BOY 2018 482  
## 6 BEL GIRL 2018 504

### Dividing dataset into boys and girls to find ordering of data

df\_boys <- df2 %>% filter(Subject=="BOY") %>% select("Location","Subject","Value") %>% arrange(Value)  
df\_girls <- df2 %>% filter(Subject=="GIRL") %>% select("Location","Subject","Value")%>% arrange(Value)  
head(df\_boys)

## Location Subject Value  
## 1 IDN BOY 358  
## 2 BRA BOY 400  
## 3 COL BOY 407  
## 4 MEX BOY 415  
## 5 CRI BOY 419  
## 6 GRC BOY 437

head(df\_girls)

## Location Subject Value  
## 1 IDN GIRL 383  
## 2 COL GIRL 417  
## 3 MEX GIRL 426  
## 4 BRA GIRL 426  
## 5 CRI GIRL 434  
## 6 CHL GIRL 462

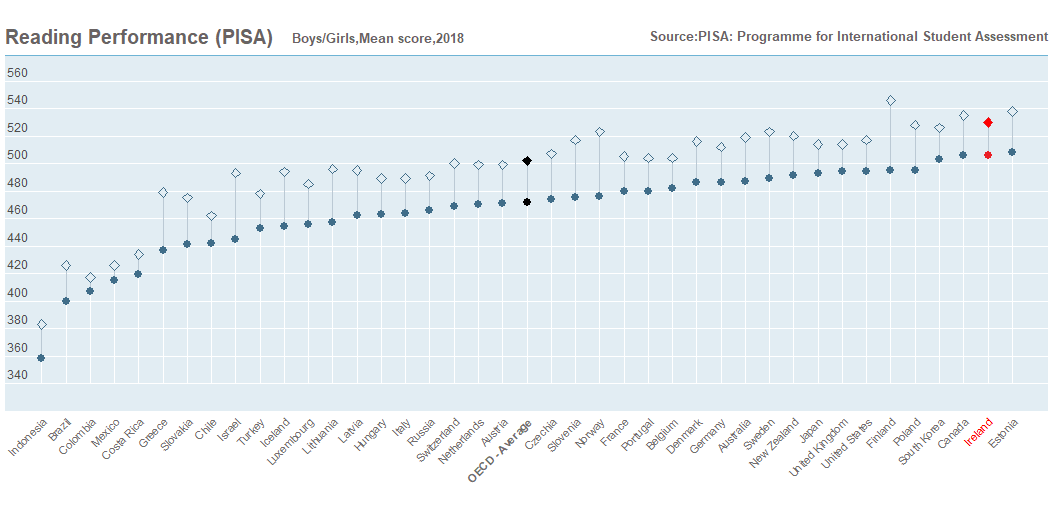
# The plot orders in the flow of boys dataset

### Finding the ordered country codes and country names

# the level of boys will be used in visual as it creates an increasing order  
countrycodes <- df\_boys$Location  
countryfullnames <- countrycode(df\_boys$Location,origin = 'iso3c', destination = 'country.name')  
# warning: Some values were not matched unambiguously: OAVG  
countryfullnames[is.na(countryfullnames)]='OECD - Average'  
countryfullnames

## [1] "Indonesia" "Brazil" "Colombia" "Mexico"   
## [5] "Costa Rica" "Greece" "Slovakia" "Chile"   
## [9] "Israel" "Turkey" "Iceland" "Luxembourg"   
## [13] "Lithuania" "Latvia" "Hungary" "Italy"   
## [17] "Russia" "Switzerland" "Netherlands" "Austria"   
## [21] "OECD - Average" "Czechia" "Slovenia" "Norway"   
## [25] "France" "Portugal" "Belgium" "Denmark"   
## [29] "Germany" "Australia" "Sweden" "New Zealand"   
## [33] "Japan" "United Kingdom" "United States" "Finland"   
## [37] "Poland" "South Korea" "Canada" "Ireland"   
## [41] "Estonia"

# creating a separate dataframe for this exercise  
df3 <-df2  
df3$Color <- ifelse(df3$Location == "IRL"|df3$Location == "OAVG",ifelse(df3$Location == "IRL","red","black"),"blue")  
df3$shape <-ifelse(df3$Subject=="BOY","A",ifelse(df3$Location == "IRL"|df3$Location == "OAVG","B","C"))  
df3$fill <- ifelse(df3$Subject=="BOY","A",ifelse(df3$Location == "IRL","B",ifelse(df3$Location == "OAVG","C","D")))  
  
# Define plot title,subtile and caption  
plot.title = "Reading Performance (PISA)"  
plot.subtitle = "Boys/Girls,Mean score,2018"  
plot.caption = "Source:PISA: Programme for International Student Assessment"  
  
g <-ggplot(df3,aes(x=factor(Location,levels = countrycodes), y=Value,colour=Color,shape=shape,fill=fill)) + #countrycodes represent country code orders of boys dataset  
 geom\_segment(aes(xend = Location), yend = 0, colour="white", size=0.25) + # create a white line till boys coordinate  
 geom\_line(aes(group = Location),colour = "#bbc9d4", size=0.5,aplha=0.7) + #creates a coloured line from boys to girls coordinate  
 geom\_point(size=2,stroke = 1) +  
 annotate(geom = 'segment', y = Inf, yend = Inf, color = '#6eb4d5', x = -Inf, xend = Inf, size = 1)+ # creates a blue top margin  
 labs(title = plot.title,subtitle = plot.subtitle,caption = plot.caption)+  
 theme(  
 axis.text.x = element\_text(angle = 45, vjust = 0.8, hjust = 1,color = ifelse(countryfullnames=="Ireland","Red","#716d6d"),face = ifelse(countryfullnames=="OECD - Average","bold","plain"),size = 8), # specific colours for special cases like Ireland and OECD Average  
 axis.text.y = element\_text(vjust = -0.5,margin = margin(r = -20)), # The text of y axis exists above the gridlines  
 axis.title.x = element\_blank(),  
 axis.title.y = element\_blank(),  
 axis.ticks.y = element\_blank(),  
 axis.ticks.x = element\_blank(),  
 axis.line.x.bottom = element\_line("#e2edf3", size= 9.7), # Hack : a background colour line at bottom to hide lower gridlines   
 panel.background = element\_rect(fill = "#e2edf3"),  
 panel.grid.major.x = element\_blank(),   
 #legend.position = c(0.077, 1),  
 legend.position = "none",  
 legend.title = element\_blank(),  
 legend.direction = "horizontal",  
 legend.key = element\_rect(fill = "white", color = NA),  
 plot.title = element\_text(size=15,color = "#696363",vjust = -4,face="bold"),  
 plot.subtitle = element\_text(size=10,hjust = 0.33,vjust = 2,color = "#696363",face="bold"),  
 plot.caption = element\_text(size=10,hjust = 1,vjust = 151,color = "#696363",face="bold"),  
 plot.margin = unit(c(5.5,5.5,26,5.5), "points") # Default margins :theme\_get()$plot.margin -> 5.5points 5.5points 5.5points 5.5points  
 ) +  
 scale\_colour\_manual(values = c("#000000","#406d89","#ea1f25")) +  
 scale\_shape\_manual(values = c(19, 23,23)) +   
 scale\_fill\_manual(values = c("yellow","red","black","#e2edf3"))+  
 scale\_y\_continuous(breaks = seq(340, 560, by = 20),minor\_breaks = seq(0, 20, 10),limits = c(341, 567)) +   
 scale\_x\_discrete(labels= countryfullnames,expand = expansion(add = 1.5))  
g



ggsave(plot = g, filename = "Part2.png")  
  
# References:  
# https://www.datanovia.com/en/blog/ggplot-legend-title-position-and-labels/  
# https://stackoverflow.com/questions/48214915/how-to-increase-size-of-ggplot-squeezed-horizontal-bar-chart  
# https://www.listendata.com/2017/03/if-else-in-r.html  
# https://viz-ggplot2.rsquaredacademy.com/labels.html  
# https://stackoverflow.com/questions/55406829/ggplot-put-axis-text-inside-plot  
# https://stackoverflow.com/questions/56097381/adding-some-space-between-the-x-axis-and-the-bars-in-ggplot/56097971  
# https://stackoverflow.com/questions/46256851/how-to-add-line-at-top-panel-border-of-ggplot2  
# https://stackoverflow.com/questions/10836843/ggplot2-plot-area-margins