Part 3: Performance over years

### Importing relevant 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

library(gridExtra)

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

##   
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':  
##   
## combine

### Importing 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

### Removing unnecessary rows and columsn

df2 <- df %>% filter(Location=="IDN"|Location=="IRL"|Location=="AUT"|Location=="OAVG")  
df\_boys <- df2 %>% filter(Subject=="BOY") %>% select("Location","Time","Subject","Value") %>% arrange(Location)  
df\_girls <- df2 %>% filter(Subject=="GIRL") %>% select("Location","Time","Subject","Value")%>% arrange(Location)  
head(df\_boys)

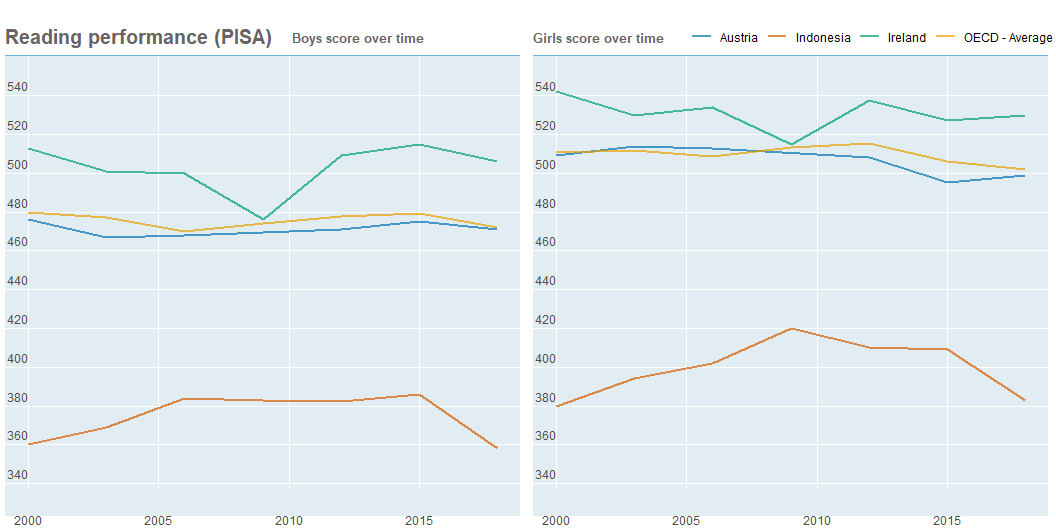
## Location Time Subject Value  
## 1 AUT 2000 BOY 476.000  
## 2 AUT 2003 BOY 467.000  
## 3 AUT 2006 BOY 468.000  
## 4 AUT 2012 BOY 471.093  
## 5 AUT 2015 BOY 475.000  
## 6 AUT 2018 BOY 471.000

head(df\_girls)

## Location Time Subject Value  
## 1 AUT 2000 GIRL 509.000  
## 2 AUT 2003 GIRL 514.000  
## 3 AUT 2006 GIRL 513.000  
## 4 AUT 2012 GIRL 508.021  
## 5 AUT 2015 GIRL 495.000  
## 6 AUT 2018 GIRL 499.000

### Plot

# plot 1 : trends of boys reading score over time for the specific countries  
gboys<-ggplot(df\_boys,aes(Time,Value,color=Location)) +   
 geom\_line(size = 1, alpha = 0.7) +  
 scale\_color\_manual(values = c("#0072b2", "#D55E00", "#009e73", "#E69F00"),name = NULL) +   
 annotate(geom = 'segment', y = Inf, yend = Inf, color = '#6eb4d5', x = -Inf, xend = Inf, size = 1)+ # creates a blue top margin  
 labs(title = "Reading performance (PISA)",subtitle = "Boys score over time") +  
 theme(  
 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(),  
 panel.background = element\_rect(fill = "#e2edf3"),  
 axis.line.x.bottom = element\_line("#e2edf3", size= 9.7), # Hack : a background colour line at bottom to hide lower gridlines   
 plot.title = element\_text(size=15,color = "#696363",vjust = -4,face="bold"),  
 plot.subtitle = element\_text(size=10,hjust = 0.75,vjust = 2,color = "#696363",face="bold"),  
 legend.position = "none", # removing legend for left visual  
 panel.grid.minor.x = element\_blank(),  
 axis.text.x = element\_text(vjust = -2)  
 ) +  
 scale\_y\_continuous(breaks = seq(340, 540, by = 20),minor\_breaks = seq(0, 20, 10),limits = c(341, 550))  
   
# plot 2 : trends of girls reading score over time for the specific countries  
ggirls<-ggplot(df\_girls,aes(Time,Value,color=Location)) +   
 geom\_line(size = 1, alpha = 0.7) +  
 scale\_color\_manual(values = c("#0072b2", "#D55E00", "#009e73", "#E69F00"),labels = c("Austria", "Indonesia","Ireland","OECD - Average")) +   
 annotate(geom = 'segment', y = Inf, yend = Inf, color = '#6eb4d5', x = -Inf, xend = Inf, size = 1)+ # creates a blue top margin  
 labs(title = " ",subtitle = "Girls score over time") +  
 theme(  
 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(),  
 panel.background = element\_rect(fill = "#e2edf3"),  
 axis.line.x.bottom = element\_line("#e2edf3", size= 9.7), # Hack : a background colour line at bottom to hide lower gridlines   
 plot.title = element\_text(size=15,color = "#696363",vjust = -4,face="bold"),  
 plot.subtitle = element\_text(size=10,hjust = 0,vjust = 2,color = "#696363",face="bold"),  
 legend.direction = "horizontal",  
 legend.key = element\_rect(fill = "white", color = NA),  
 legend.position = c(0.65, 1.043), # manually specifying legend location  
 legend.title = element\_blank(),  
 panel.grid.minor.x = element\_blank(),   
 axis.text.x = element\_text(vjust = -2)  
 ) +  
 scale\_y\_continuous(breaks = seq(340, 540, by = 20),minor\_breaks = seq(0, 20, 10),limits = c(341, 550))  
  
  
  
g<-grid.arrange(gboys, ggirls, nrow = 1)



ggsave(plot = g, filename = "Part3.png")