Part 1

This graph depicts the performance of 15 year old students in term of literacy, mathematics and science. The graph shows the ranking of all the member countries of OECD.

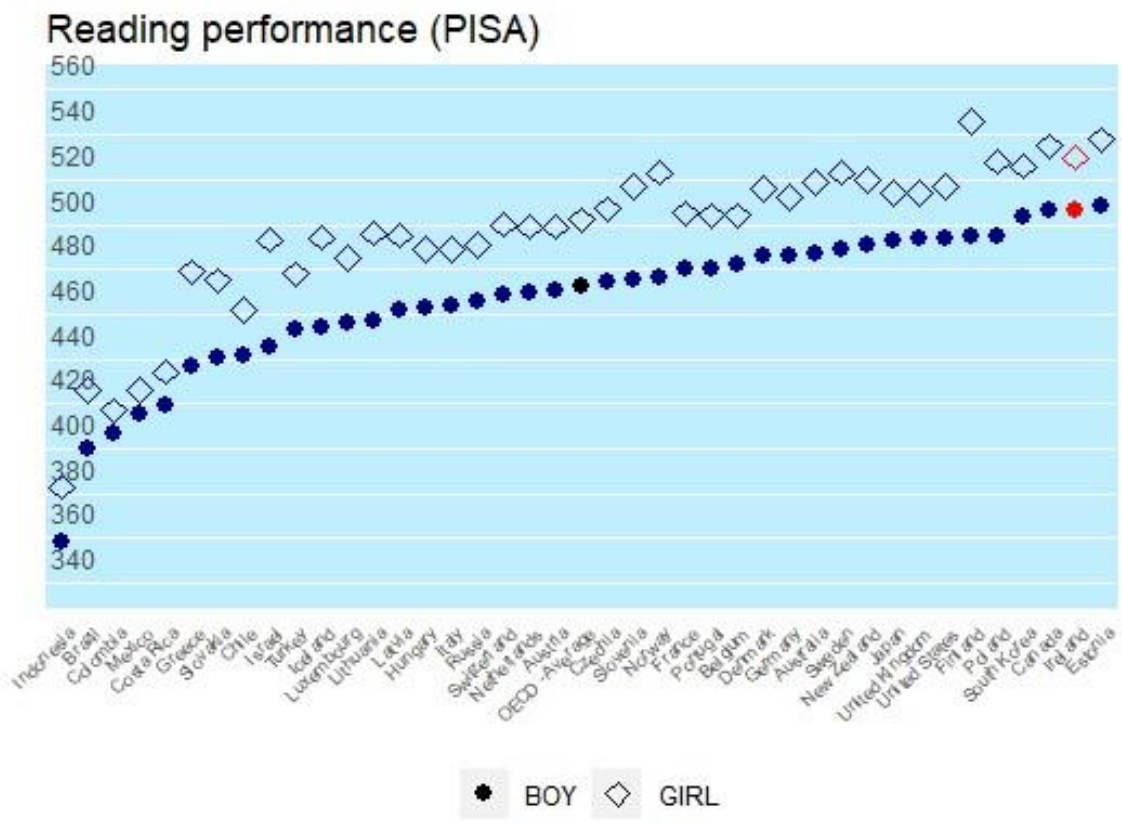
- The Boys are represented with a filled dot and girls with a empty diamond, in the graph IRELAND is shown in red color while the average of all OECD countries is shown in black.
- The axes are pretty straight forward, on the X axis is the list of countries that are the members of OECD while on the Y axis is the range of the average of performance of students in literacy, mathematics and science. The values on the Y axis start from 340 ranging to 560 with interval of 20.
- The plot has only horizontal grid lines with vertical grid lines ending towards the boys plotted value. This is very intuitive and even though there are a lot of countries it helps with identifying the performance for each country.
- The legend for this graph is pretty straight forward and it denotes only the shape for the boys and the girls.
- The background for the graph is light blue in color and matches the graph tone perfectly and complements the data, there isn't much in background it is plain.
- It was very interesting of the maker of graph to insert the average plotting exactly in the middle of the graph as it has a lot of focus, majorly due to its black color and position in the graph, also, the decision to make Ireland data red was also very interesting as it focuses the country.
- The point where the user has grid lines of different color connecting the boys and girls data points was very appealing and looks really good on the plot.
- Apart from that the position of the countries makes the graph look good in a flow and makes it very presentable.

Part 2

- For the recreation of the plot I pretty much followed on the similar lines of the given plot.
- The plots is depicted in legends for boys and girls where boys is filled dot while girls is empty diamond, the y label Ireland and oecd average have colors red and black respectively while other are in darkblue.

- The X axis has list of countries that are members of OECD and the Y axis is plotted inside the graph which denotes the read performance in the year 2018. The axes don't show the title.
- The grid line are without tick and only major horizontal grid lines are plotted and vertical lines are kept blank. The Y axis has a continuous scale ranging from 340 to 560 with an interval of 20.
- the legend shows the shape for the boys and girls data and is place at the bottom of the chart an has no title.
- The background has a light blue color and all the margin lines are kept blank.
- The parts of the graph which I was not able to recreate were -
    - The space in the margin of left of the graph where the Y axis label is shown, It is separate in graph but I was not able to plot it that way.
    - The legend is towards left side in bottom but I was not able to place it in bottom left position while attempting to do that it overlapped the graph so I left I left it as bottom only.
    - The OECD- Average and Ireland X axis label are colored, I wasn't able to do that neither was I able to fill the shape for these label's but I did color the geom-points based on those.
    - The original plot has vertical grid lines starting from x label till geom point but I wasn't able to plot that.

Plot 2



Code - Part 2

```
library(ggplot2)
library (dplyr)
library(countrycode)
library(forcats)

# Data Preprocessing

oecdpisa =  read.csv(file = 'OECD_PISA.csv')
new_csv <- subset(oecdpisa, TIME==2018 )
new_csv <- select(new_csv,ï..LOCATION,SUBJECT,Value)
names(new_csv)[names(new_csv)== 'ï..LOCATION'] <- 'Location'
names(new_csv)[names(new_csv)== 'SUBJECT'] <- 'Gender'
new_csv <- new_csv[!(new_csv$Gender=='TOT'),]
new_csv$Location <- countrycode(new_csv$Location, origin =
"iso3c", destination = "country.name")
new_csv$Location[73] <- 'OECD -Average'
new_csv$Location[74] <- 'OECD -Average'

# sort based on gender and relative values

new_csv <- new_csv[with(new_csv, order(Gender,Value)),]
```

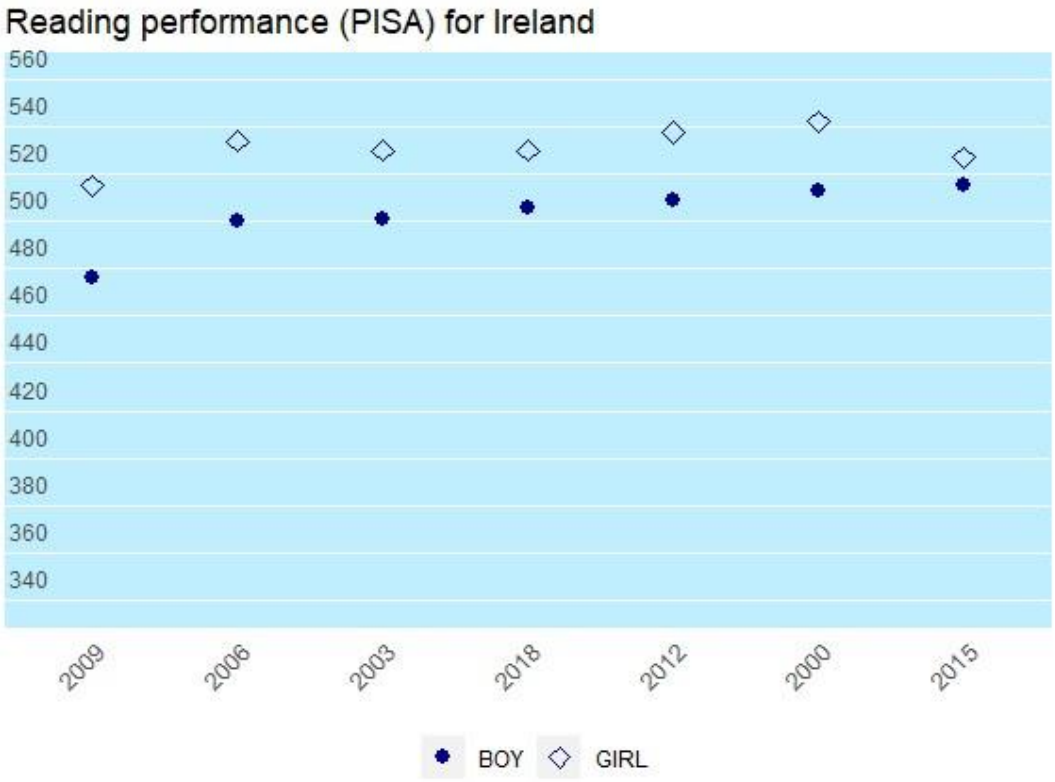
```
# convert location to factor to keep the order same in
plotting

new_csv$Location <- factor(new_csv$Location, levels =
unique(new_csv$Location), ordered =TRUE)

#plotting
new_csv$highlight <-
ifelse(new_csv$Location=='Ireland',"highlight","normal")
mycolors <-c("highlight" = "red", "normal" = "darkblue","gl" =
"black")
new_csv$highlight[21] <- 'gl'
new_csv$highlight[62] <- 'gl'

gg <-ggplot(new_csv, (aes(x= Location, y=Value, group =
Gender,ylab=''))) +
    theme(panel.background = element_rect(fill =
"lightblue1",colour = "lightblue1"),axis.text.x =
element_text(angle = 45, size = 6,vjust = 1, hjust =
1),axis.ticks.y = element_blank(),
axis.ticks.x = element_blank(),axis.title.x =
element_blank(),axis.title.y = element_blank(),
panel.grid.minor = element_blank(),panel.grid.major.x =
element_blank(),
axis.text.y = element_text(vjust = -0.7,margin = margin(l =
100, r = -20)),legend.position = 'bottom', legend.title
=element_blank())+
  scale_y_continuous(limits = c(340,560),breaks =
seq(340,560,by = 20))+
  geom_point(aes(shape = Gender, size=Gender, colour =
highlight ), size=2.5)+
  scale_shape_manual(values=c(19,5),name = "Gender")+
  guides(color = FALSE)+
  scale_color_manual("Status", values = mycolors)+
   ggtitle("Reading performance (PISA)")

new_csv
gg
```

Part 3

- To plot the data according to all the years till date I chose the country Ireland.
- The plot has similar legend type i.e. filled dot for boys and empty diamond for girls.
- The x axis has a Date type which is the years of the performance and Y axis has values of the performance in the respective years.

- There are only horizontal grid lines and vertical lines are kept blank.
- The date is converted to factor and is plotted in ascending order of the value.
- Everything else is pretty much the same compared to other plots.

Plot



Code - Part 3

```
#Part 3
library(ggplot2)
library (dplyr)
library(countrycode)
library(forcats)

# Data Preprocessing
oecdpisa1 =  read.csv(file = 'OECD_PISA.csv')
new_csv1 <- subset(oecdpisa1, ï..LOCATION =='IRL')
new_csv1 <- select(new_csv1,SUBJECT,Value,TIME)
names(new_csv1)[names(new_csv1)== 'ï..LOCATION'] <- 'Location'
names(new_csv1)[names(new_csv1)== 'SUBJECT'] <- 'Gender'
new_csv1 <- new_csv1[!(new_csv1$Gender=='TOT'),]
new_csv1 <- new_csv1[with(new_csv1, order(Gender,Value)),]

new_csv1$TIME <-factor(new_csv1$TIME,levels =
unique(new_csv1$TIME), ordered = T)
```

```
gg1 <- ggplot(new_csv1, aes(x = TIME, y = Value, group =
Gender)) +  theme(panel.background = element_rect(fill =
"lightblue1",colour = "lightblue1"),axis.text.x =
element_text(angle = 45, size = 10,vjust = 1, hjust =
1),axis.ticks.y = element_blank(),axis.ticks.x =
element_blank(),axis.title.x = element_blank(),axis.title.y =
element_blank(), panel.grid.minor =
element_blank(),panel.grid.major.x =
element_blank(),axis.text.y = element_text(vjust = -0.7,margin
= margin(l = 100, r = -20)),legend.position = 'bottom',
legend.title =element_blank())+
  scale_y_continuous(limits = c(340,560),breaks =
seq(340,560,by = 20))+
  geom_point(aes(shape = Gender, size=Gender), size=2.5, color
= "darkblue")+
  scale_shape_manual(values=c(19,5),name = "Gender")+
  guides(color = FALSE)+
   ggtitle("Reading performance (PISA) for Ireland")

new_csv1
```