COMP1378 Assignment_2

Manjiri Ramesh Pendharkar - 45933219

2023-03-31

Exercise 1

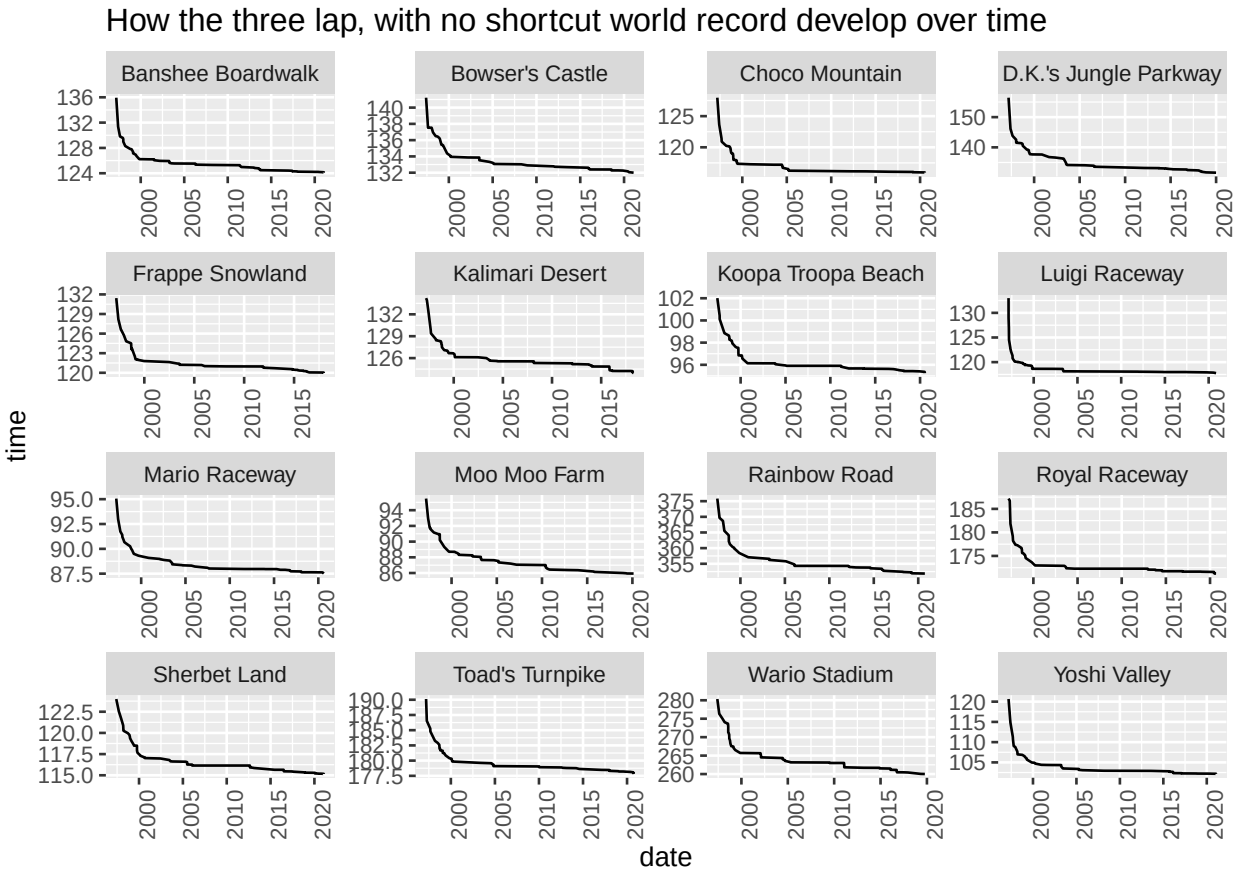
Data wrangling and visualisations

For this exercise , record.csv data is used, which was created by Benedikt Claus. It's about Mario Kart World Records and contains world records for the classic racing game on the Nintendo 64.

```
# importing the library to read the csv file
library(tidyverse)

# reading the file records.csv in tibble df1
df1 <- read_csv("records.csv")
```

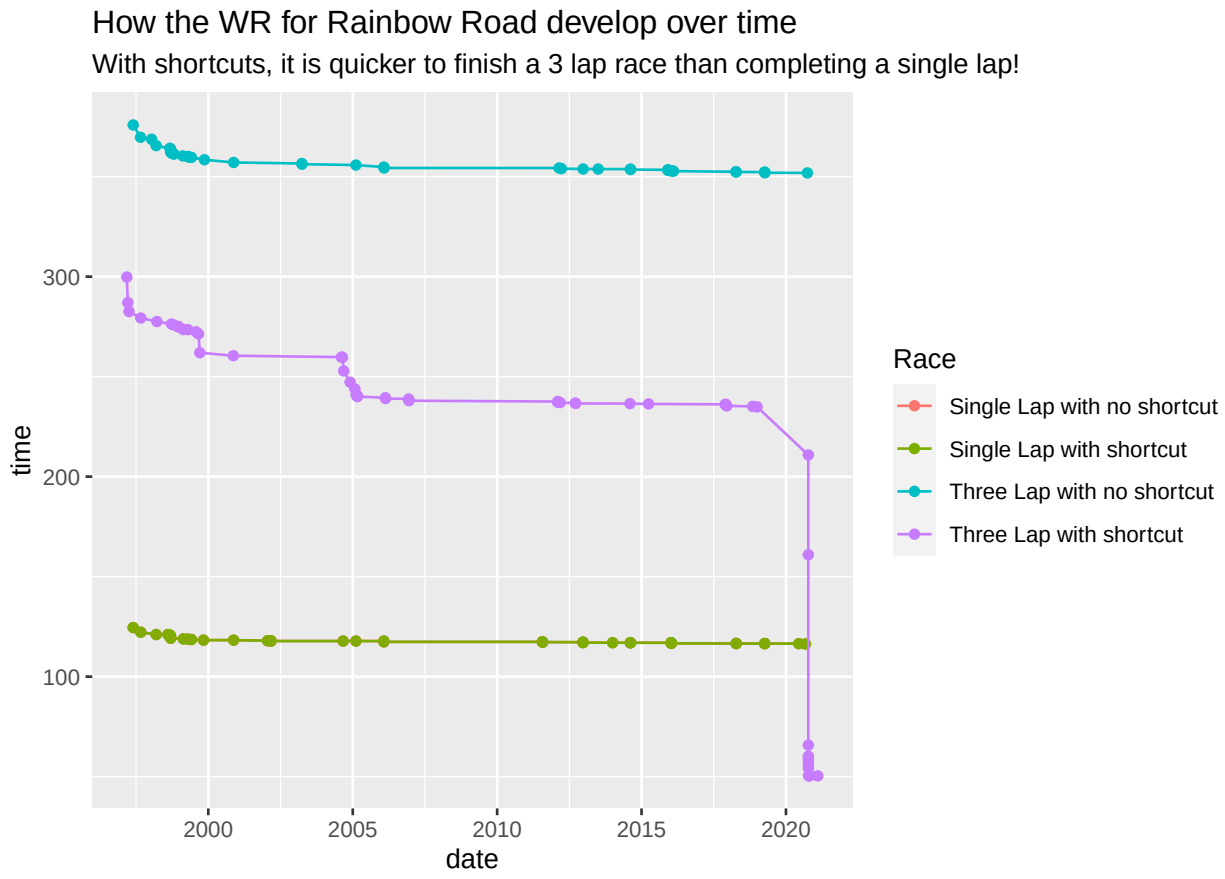
Plot 1

```
#filtering the dataset to include records of three lap with no shortcuts race and grouping by track
q1 <- df1 %>%
      filter(type == "Three Lap", shortcut == "No")%>%
      group_by(track) %>%
      select(date, time, track)

#plotting graph of three lap with no shortcuts records
plot1 <- ggplot(data=q1, aes(x = date, y = time)) +
         geom_line() +
         facet_wrap(~track, scales = "free")

#adding title and adjusting axis
plot1 + ggtitle("How the three lap, with no shortcut world record develop over time") +
        theme(axis.text.x = element_text(angle = 90))
```



Plot 2

```
#filtering the dataset to include records of rainbow road track grouped by type and shortcut
q2 <- df1 %>%
      filter(track == "Rainbow Road")%>%
      group_by(type, shortcut) %>%
      select(date, time, type, shortcut)

#plotting graph to compare records by type and shortcut
plot2 <- ggplot(data= q2 , aes(x = date, y = time,colour=interaction(type, shortcut))) +
         geom_point()+
         geom_line()

#adding legend name, labels, graph title and subtitle
plot2 + scale_colour_manual(name = "Race",
        breaks=c("Single Lap.No","Single Lap.Yes","Three Lap.No",'Three Lap.Yes'),
        labels=c('Single Lap with no shortcut','Single Lap with shortcut',
                 'Three Lap with no shortcut', 'Three Lap with shortcut'),
        values=c("#f8766d","#7cae00","#00bfc4","#c77cff")) +
        ggtitle(label = "How the WR for Rainbow Road develop over time",
        subtitle = "With shortcuts, it is quicker to finish a 3 lap race than completing a single lap!")
```



Exercise 2

Data wrangling and visualisations

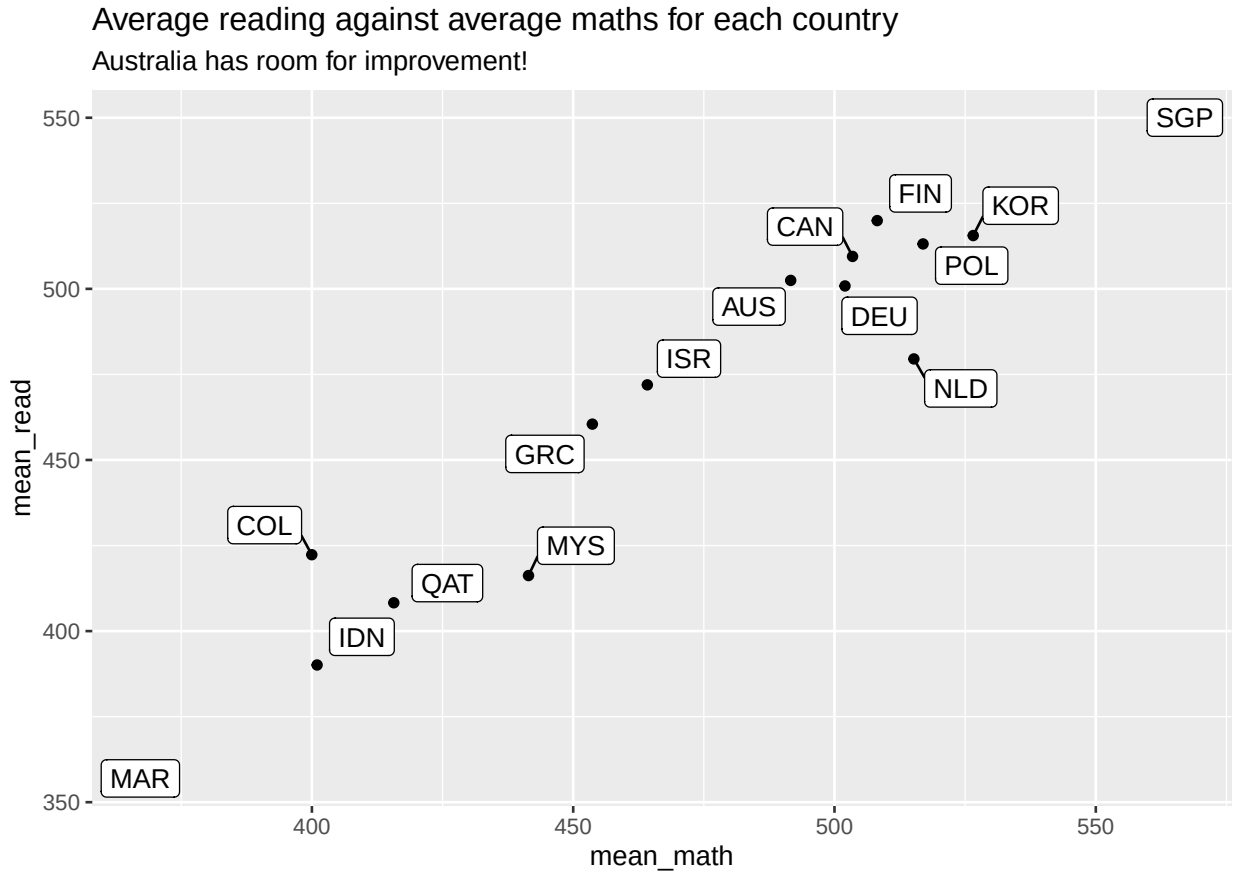
For this exercise , pisa_2018.csv data is used. It's about the Programme for International Student Assessment (PISA), an international assessment measuring student performance in reading, mathematical and scientific literacy.

```
library(tidyverse)
library(ggrepel)

# reading the file pisa_2018.csv in tibble df2
df2 <- read_csv("pisa_2018.csv")
```

Plot 1

```
#calculating average math and reading scores grouped by country
q1 <- df2 %>%
      group_by(country) %>%
      summarise(mean_math = mean(math), mean_read = mean(read))
```

```
#plotting average math and reading scores grouped by country
plot1 <- ggplot(data=q1, aes(x = mean_math, y = mean_read, label=country)) +
         geom_point() +
         geom_label_repel(aes(label = country))

#adding title and subtitle
plot1 + ggtitle(label = "Average reading against average maths for each country",
        subtitle = "Australia has room for improvement!")
```



Plot 2

```
#dataset grouped by gender and country
a <- df2 %>%
  group_by(country, gender) %>%
  summarise(mean_math = mean(math), mean_read = mean(read), mean_sci = mean(science))

#making the dataset wider taking names from gender and values from mean of read, math, science
b <- a %>%
   pivot_wider(names_from = gender, values_from =c(mean_math,mean_read,mean_sci) )

#calculating the gender mean difference of each subject
```

```r
c <- b %>%
     mutate(diff_math= mean_math_female - mean_math_male,
     diff_read = mean_read_female - mean_read_male,
     diff_sci = mean_sci_female - mean_sci_male)

#making the dataset longer so that all the subject differences are in one column and on each row
d <- c %>%
     pivot_longer(cols = starts_with("diff"),names_to = "mean_subjects", values_to = 'diff' )

#creating difference column where if difference is greater than 0 then true otherwise false
e <- d %>%
     mutate( difference  = if_else(diff > 0, TRUE, FALSE)) %>%
     select(country, mean_subjects, diff, difference)

#plotting the graph with geom vline to have vertical line in facet wrap
plot2 <- ggplot(data=e, aes(x=diff, y=country, colour=difference)) +
      geom_point() +
      geom_vline(xintercept = 0) +
      facet_wrap(~mean_subjects)

plot2 + scale_colour_discrete(name = "diff > 0") +
        ggtitle(label = "Average gender difference (diff = female minus male) per Country",
                subtitle = "Gender gap in reading is universal, but the math gap is not.")
```