DSC 385 - Project 1 Report

Name and EID

Juan Linares Velasquez EID: jgl2483

Setup

```
## Loading all the data files
## DO NOT CHANGE THIS CODE
championships <- read_tsv("WCA_export_championships.tsv.bz2")
competitions <- read_tsv("WCA_export_Competitions.tsv.bz2")
continents <- read_tsv("WCA_export_Continents.tsv.bz2")
countries <- read_tsv("WCA_export_Countries.tsv.bz2")
eligible_country_iso2s_for_championship <- read_tsv("WCA_export_eligible_country_iso2s_for_champi
events <- read_tsv("WCA_export_Events.tsv.bz2")
formats <- read_tsv("WCA_export_Formats.tsv.bz2")
persons <- read_tsv("WCA_export_Persons.tsv.bz2")
ranksaverage <- read_tsv("WCA_export_RanksAverage_333.tsv.bz2")
rankssingle <- read_tsv("WCA_export_RanksSingle_333.tsv.bz2")
results <- read_tsv("WCA_export_Results_333.tsv.bz2")
roundtypes <- read_tsv("WCA_export_RoundTypes.tsv.bz2")
scrambles <- read_tsv("WCA_export_Scrambles.tsv.bz2")
```

Active Speed Cubers

How many active (3x3x3) speedcubers are there registered with the WCA?

For this question an *active speeedcuber* is defined as any person registered in the WCA who has competed in at least two competitions in the years 2022–2024.

```
##
#Select only the ids of speedcubers that participated in competitions in 2022 and later
activeCompetitions=competitions |> filter(year >= 2022) |>select(id)
#From the general results keep only that ones who competed in 2022 and later
activeResults <- results |> filter(competitionId %in% activeCompetitions$id)
#Selext only the speedcubers who, on or after 2022, competed in 2 events or more
activeSpeedCubers = activeResults |> group_by(personId) |> summarize(n=n()) |> filter(n>=2)
#Show the results
nrow(activeSpeedCubers)
```

```
[1] 50439
```

There are 50439 active speedcubers.

World Records

This question has two parts:

   1. According to the data, who holds the world record single best solve? On what date was this record set?

```
bestTime= results |> arrange(best) |> filter(best>0)
bestTime1=bestTime[1,]
record1=bestTime1$personName
record1
```

```
[1] "Max Park"
```

Max Park holds the world record single best solve.

On what date was this record set?

```
#id1=bestTime1$competitionId
#competitionId1=toString(id1)
date1=competitions |> filter(id==bestTime1$competitionId) |> select(year,month,day)

#date1=competitions |> filter(id==competitionId1) |> select(year,month,day)
toString(date1)
```

```
[1] "2023, 6, 11"
```

The date of the record was June 11, 2023

   2. According to the data, who previously held the world record best single solve?

```
bestTime2=bestTime[2,]
bestTime2$personName
```

```
[1] "Luke Garrett"
```

Luke Garrett held the record previously.

On what date was this previous record set?

```
date2=competitions |> filter(id==bestTime2$competitionId) |> select(year,month,day)
toString(date2)
```

```
[1] "2023, 7, 22"
```

It was set on July 22, 2023

NOTE: For these questions, consider all speedcubers (not just active ones) and define "best" as the fastest time for a single solve (not for an average).

Regional Rankings

This question has two parts:

1. Amongst all speedcubers, who is the top ranked male speedcuber (for best single solve) in Australia?

```
malesAustralia = persons |> filter(countryId=="Australia"& gender=="m") |>select(id)
malesAustraliaResults=results |> filter(personId %in% unlist(malesAustralia))
sortedMaleAustralia= arrange(malesAustraliaResults,best) |> filter(best >0)
bestMaleAustralia=sortedMaleAustralia[1,] |> select(personName)
toString(bestMaleAustralia)
```

```
[1] "Jode Brewster"
```

Jode Brewster is the top ranked male speedcuber.

2. Amongst all speedcubers, who is the top ranked female speedcuber (for best single solve time) in Europe?

NOTE: Europe is identified under the `continentId` column of the `countries` table.

```
europeCountries= countries |> filter(continentId=="_Europe") |> select(id)


femalesEurope = persons |> filter(countryId %in% unlist(europeCountries) & gender=="f") |>select(
femalesEuropeResults=results |> filter(personId %in% unlist(femalesEurope))
sortedFemaleEurope= arrange(femalesEuropeResults,best) |> filter(best >0)
bestFemaleEurope=sortedFemaleEurope[1,] |> select(personName)
toString(bestFemaleEurope)
```

```
[1] "Magdalena Pabisz"
```

Magdalena Pabisz is the top ranked female speedcuber in Europe.

Time Until Sub-5

Having a time below 5 seconds is considered an elite achievement and most speedcubers have to complete a large number of solves before they can obtain a sub-5 second solve.

1. For the current top 10 speedcubers in the world (as recorded in the `RanksSingle` table), on average, how many solves did they have to do before achieving a sub-5 second solve?

NOTE: Each round of a competition has 5 solves that should be considered separately when counting the number of solves.

```
top10ids = head(rankssingle[,1], 10)
times = list()
count = 0
#####
```

```r
colnames(competitions)[1] <- c("competitionId")
colnames(roundtypes)[1] <- c("roundTypeId")

resultsMod = results |> left_join(competitions, by="competitionId")
resultsMod$Date <- as.Date(with(resultsMod, paste(year, month, day, sep = "-")))
resultsMod = select(resultsMod, -name:-endDay)
resultsMod = resultsMod |> left_join(roundtypes, by="roundTypeId")
resultsMod = select(resultsMod, -cellName:-name)
resultsSorted <- resultsMod %>% arrange(Date, rank)


for (id in unlist(top10ids)){
  idResults = resultsSorted |> filter(personId==toString(id)) |> select(value1:value5)

  idResultsList=list()

  for (index in 1:nrow(idResults)){
    idResultsList = append(idResultsList,idResults[index,])
  }


  for (idResult in unlist(idResultsList)){
    if (idResult >0){
      count = count+1
    }

    if (idResult < 500 & idResult>0){
      break
    }

  }
  times=append(times,count)
  count=0
}
times=unlist(times)
average=mean(times)
average
```

```
[1] 503
```
It took 503 tries on average before having a sub 5-second solve.

2. For **one** of the top 10 speedcubers make a scatterplot of their individual single solve times vs. the date of the solve, with date on the x-axis and solve time on the y-axis.

```r
topIdResults = resultsSorted |> filter(personId==toString(top10ids[1,])) |> select(Date, value1:v

topIdResults_long <- topIdResults |> pivot_longer(cols = starts_with("value"), names_to = "Variab

ggplot(topIdResults_long, aes(x = Date, y = Value) )+
```

```
  geom_point(color = "black") +
  labs(title = "Scatterplot of Solve Times vs Date for speedcuber Max Park ", x = "Date", y = "Ti
```

Scatterplot of Solve Times vs Date for speedcuber Max Park
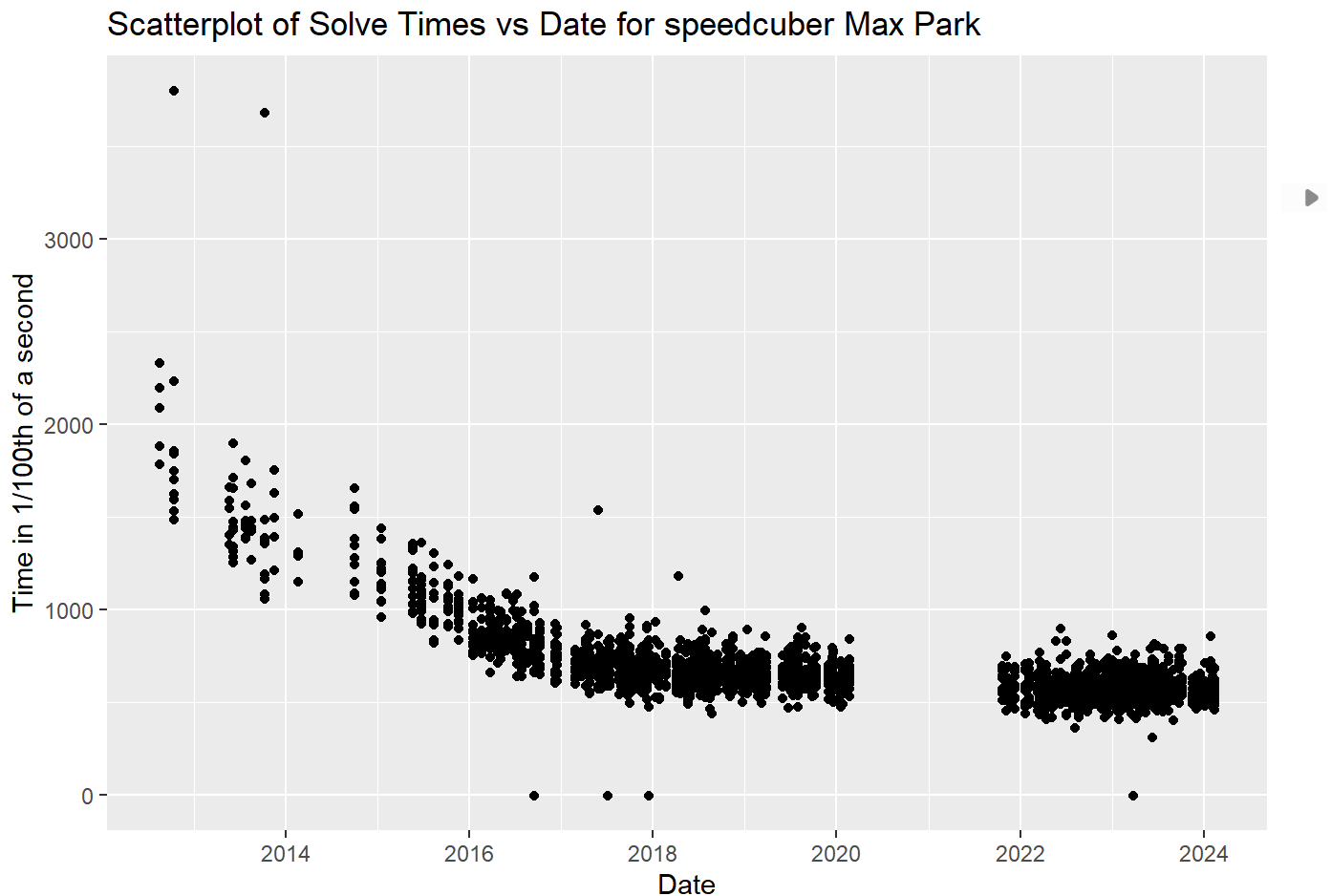
Up-and-Coming Speed Cubers

Which speed cubers **not** in the top 10,000 (worldwide for single best time) should we keep an eye on for the near future?

The idea here is to identify "up-and-coming" speedcubers who are not yet achieving elite times. Come up with a list of **five** speedcubers (provide their names and WCA IDs) that you have identified as "up-and-coming". There is no one way to answer this question and the goal is to provide an analysis of the data that justifies the selection of your five names.

```
notTop10000 = rankssingle[-(1:10000),] |> select(personId)
resultsNotTop10000 = results |> filter(personId %in% notTop10000$personId)
resultsNotTop10000Sorted=resultsNotTop10000 |> arrange(average) |> filter(average >0)

upAndComing = head(resultsNotTop10000Sorted, 5) |> select(personName, personId)
upAndComing
```

```
# A tibble: 5 × 2
  personName          personId
```

```
  <chr>              <chr>
1 Leemar Bonite        2018BONI01
2 Junlin Zhou (周俊霖)  2019ZHOU50
3 Adam Quinn           2015QUIN07
4 Muhammad Fahri Rahman 2016RAHM09
5 Trịnh Quang Dũng     2019DUNG01
```

The up-and-coming speedcubers (selected by average solve time)are:

Name WCA ID Leemar Bonite 2018BONI01

Junlin Zhou (周俊霖) 2019ZHOU50

Adam Quinn 2015QUIN07

Muhammad Fahri Rahman 2016RAHM09

Trịnh Quang Dũng 2019DUNG01