Question4

Ramzan Kamoto

2024-06-18

Question4

This readme is just to help organise my thoughts and execute my code for question 4. We start again by reading in our rds files using our read_rds function.

Olympics

We have data on the Olympics dating as far back as 1896 to 2012. With the upcoming olympic games, it is appropriate to take a look back and analyse some of the trends in terms of winning medals that we have seen in the past

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
list.files('/Users/ramzankamoto/Documents/Masters/DS_EXAM/23550716/Question4/code', full.names = T, recu

directory_path <- "/Users/ramzankamoto/Documents/Masters/DS_EXAM/23550716/Question4/data/olympics"

rds_data_list <- read_rds_files("/Users/ramzankamoto/Documents/Masters/DS_EXAM/23550716/Question4/data/
```

Data

Initially we wanted to analyse Indias performance in previous olympic games. However, in order to clearly see how many medals a country has we have to ensure that medals are counted by event and not by athlete. The data in its current state, inflates the performance of countries who participate in team sports like hockey.

We use a function to label team sports team and individual sports individual.

```
s_games <- label_team_sports(summer)
```

We can now do further analysis on individual awards given that there are relatively fewer Team sports (at least for the summer games) so the medal count should not be affected as much.

```
list(unique(summer$Country))
```

```
## [[1]]
##   [1] "HUN" "AUT" "GRE" "USA" "GER" "GBR" "FRA" "AUS" "DEN" "SUI" "ZZX" "NED"
##  [13] "BEL" "IND" "CAN" "BOH" "SWE" "NOR" "ESP" "ITA" "CUB" "ANZ" "RSA" "FIN"
##  [25] "RU1" "EST" "TCH" "NZL" "BRA" "JPN" "LUX" "ARG" "POL" "POR" "URU" "YUG"
##  [37] "ROU" "HAI" "EGY" "PHI" "IRL" "CHI" "LAT" "MEX" "TUR" "PAN" "JAM" "SRI"
##  [49] "KOR" "PUR" "PER" "IRI" "TRI" "URS" "VEN" "BUL" "LIB" "EUA" "ISL" "PAK"
##  [61] "BAH" "BWI" "TPE" "ETH" "MAR" "GHA" "IRQ" "SIN" "TUN" "KEN" "NGR" "GDR"
##  [73] "FRG" "UGA" "CMR" "MGL" "PRK" "COL" "NIG" "THA" "BER" "TAN" "GUY" "ZIM"
##  [85] "CHN" "CIV" "ZAM" "DOM" "ALG" "SYR" "SUR" "CRC" "INA" "SEN" "DJI" "AHO"
##  [97] "ISV" "EUN" "NAM" "QAT" "LTU" "MAS" "CRO" "ISR" "SLO" "IOP" "RUS" "UKR"
## [109] "ECU" "BDI" "MOZ" "CZE" "BLR" "TGA" "KAZ" "UZB" "SVK" "MDA" "GEO" "HKG"
## [121] "ARM" "AZE" "BAR" "KSA" "KGZ" "KUW" "VIE" "MKD" "SCG" "ERI" "PAR" "UAE"
## [133] "SRB" "SUD" "MRI" "TOG" "TJK" "AFG" NA    "BRN" "GUA" "GRN" "TTO" "BOT"
## [145] "MNE" "CYP" "SGP" "GAB"
```

```
summer_plotdata <- s_games %>% filter(Label == "Individual",
                                      Country == c("USA", "GER", "GBR","RUS", "JAM"))
```

Plot

Next we try to use this data to come up with a plot for Individual Medals per Country.

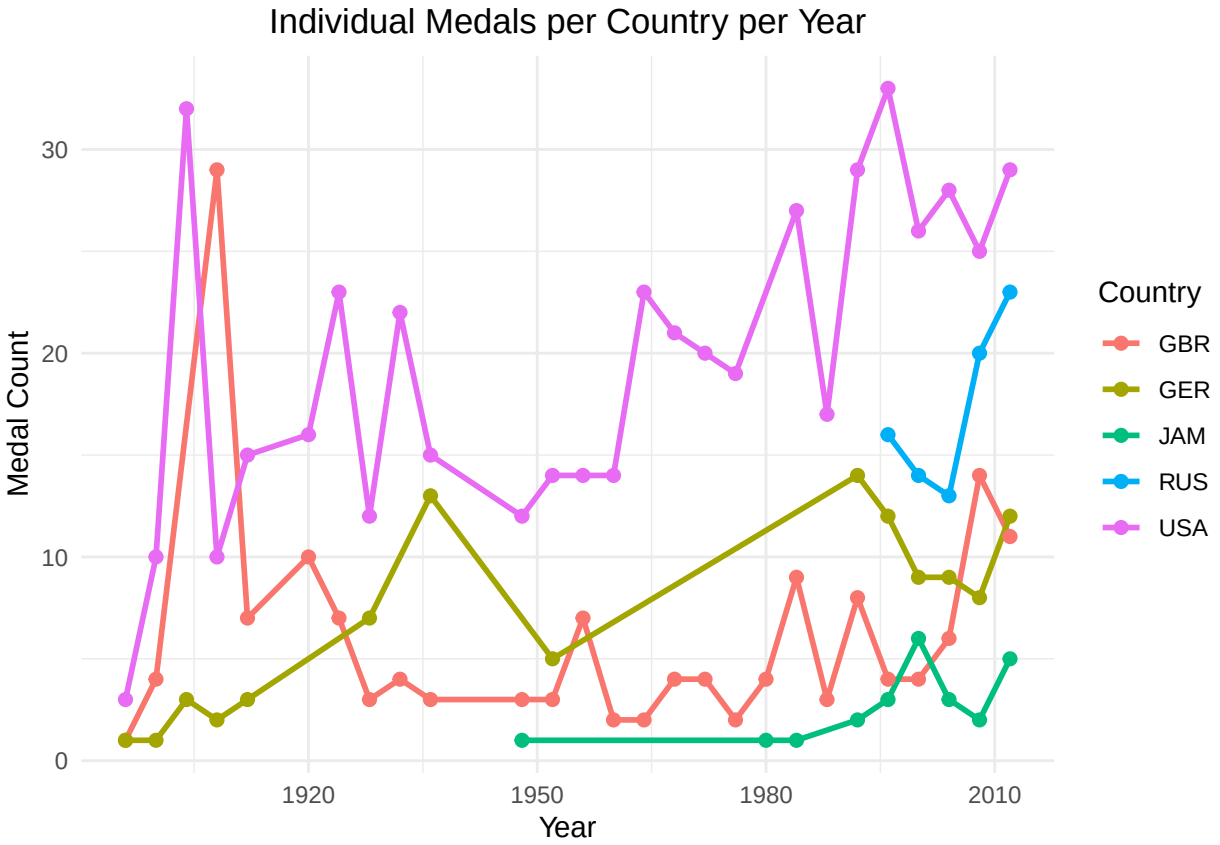
```
#first we aggregate the models so the data is usable
s_plot_data <- aggregate_medals(summer_plotdata)

#apply a function that adds up the medal count per year
aggregate_medals_summer <- aggregate_medal_counts(s_plot_data)

#Plot the line graph
summer_plot <- plot_medal_counts(aggregate_medals_summer)
```

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

```
print(summer_plot)
```



The plot shows some countries that typically do well at the olympic games. The US and Russia since independence are the only countries to consistently achieve more than 10 medals. Great Britain is initially very strong but drops below 10 individual models for pretty much the remainder of the period.