Assignment1-Q1

2022-09-19

Question 2

a. Read the data into R using the read_csv function and save it to an R object named club_teams. Give the code you used to import the data into R and use the head() function to show the first six rows of data.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
club_teams = read_csv("spi_global_rankings.csv")
```

```
## Rows: 643 Columns: 7
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): name, league
## dbl (5): rank, prev_rank, off, def, spi
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
head(club_teams)
```

```
## # A tibble: 6 x 7
##    rank prev_rank name                league                     off   def   spi
##   <dbl>     <dbl> <chr>               <chr>                    <dbl> <dbl> <dbl>
## 1     1         1 Bayern Munich       German Bundesliga         3.35  0.45  92.9
## 2     2         2 Manchester City     Barclays Premier League   3.02  0.32  92.7
## 3     3         3 Paris Saint-Germain French Ligue 1            3.06  0.5   90.4
## 4     4         4 Liverpool           Barclays Premier League   2.8   0.47  89.0
## 5     5         6 Barcelona           Spanish Primera Division  2.66  0.46  87.9
## 6     6         5 Real Madrid         Spanish Primera Division  2.77  0.57  87.1
```

b. What class of object is club_teams?

```
class(club_teams)
```

```
## [1] "spec_tbl_df" "tbl_df"      "tbl"         "data.frame"
```

c. What is the mode of the column titled name?

```
mode(club_teams$name)
```

```
## [1] "character"
```

d. Add a column (named: diffxG) to the club_teams object containing the difference between the offensive expected goals (off) and the defensive expected goals (def).

```
club_teams["diffxG"] = club_teams["off"] - club_teams["def"]
head(club_teams)
```

```
## # A tibble: 6 x 8
##    rank prev_rank name                league              off   def   spi diffxG
##   <dbl>     <dbl> <chr>               <chr>             <dbl> <dbl> <dbl>  <dbl>
## 1     1         1 Bayern Munich       German Bundesliga  3.35  0.45  92.9   2.9
## 2     2         2 Manchester City     Barclays Premier~  3.02  0.32  92.7   2.7
## 3     3         3 Paris Saint-Germain French Ligue 1     3.06  0.5   90.4   2.56
## 4     4         4 Liverpool           Barclays Premier~  2.8   0.47  89.0   2.33
## 5     5         6 Barcelona           Spanish Primera ~  2.66  0.46  87.9   2.2
## 6     6         5 Real Madrid         Spanish Primera ~  2.77  0.57  87.1   2.2
```

e. Using slice and filter, write a line (or a few lines) of code that will return the first five rows from the league Italy Serie A and save it to an object called first_five_serieA. Give your code and print this object (using print) to show that you have done it correctly. Recall: the == operator can be used to test for whether the value on the left hand side in a vector is equal to the value no the right hand side. Hint: the correct order of slice and filter matters!

```
first_five_serieA = slice(filter(club_teams, league == "Italy Serie A"), 1:5)
print(first_five_serieA)
```

```
## # A tibble: 5 x 8
##    rank prev_rank name           league          off   def   spi diffxG
##   <dbl>     <dbl> <chr>          <chr>         <dbl> <dbl> <dbl>  <dbl>
## 1    12        14 Internazionale Italy Serie A  2.45  0.73  80.5   1.72
## 2    13        16 Napoli         Italy Serie A  2.36  0.67  80.5   1.69
## 3    17        17 AC Milan       Italy Serie A  2.18  0.64  79.0   1.54
## 4    33        29 Atalanta       Italy Serie A  2.04  0.8   73.0   1.24
## 5    35        33 AS Roma        Italy Serie A  2     0.8   72.5   1.2
```

f. Add a new column to club_teams called SPI_max_league_diff which contains the difference in SPI between the maximum SPI value **for that league* and the team in each row. In other words, the best team in each league should have an SPI_max_league_diff value of 0 and the second best team in each league should have a positive number (the difference between the league maximum and thir own SPI value). Hint: Think about the kid's height example from class. Give your code for adding the column and use head to show the first six rows of the updated object.

```
club_teams <- club_teams %>%
  group_by(league) %>%
  mutate(SPI_max_league_diff = max(spi)-spi) %>%
  ungroup()
head(club_teams)
```

```
## # A tibble: 6 x 9
##    rank prev_rank name                league      off   def   spi diffxG SPI_m~1
##   <dbl>     <dbl> <chr>               <chr>     <dbl> <dbl> <dbl>  <dbl>   <dbl>
## 1     1         1 Bayern Munich       German B~  3.35  0.45  92.9   2.9     0
## 2     2         2 Manchester City     Barclays~  3.02  0.32  92.7   2.7     0
## 3     3         3 Paris Saint-Germain French L~  3.06  0.5   90.4   2.56    0
```

```
## 4     4         4 Liverpool           Barclays~  2.8   0.47  89.0   2.33   3.68
## 5     5         6 Barcelona           Spanish ~  2.66  0.46  87.9   2.2    0
## 6     6         5 Real Madrid         Spanish ~  2.77  0.57  87.1   2.2    0.830
## # ... with abbreviated variable name 1: SPI_max_league_diff
```