

Lab 3: Data Wrangling on Soccer Tournament Data

July 9, 2020

Soccer tournament data wrangling

Read the dataset of football games.

```
d <- read_csv("data/results.csv")
```

```
## Parsed with column specification:
## cols(
##   date = col_date(format = ""),
##   home_team = col_character(),
##   away_team = col_character(),
##   home_score = col_double(),
##   away_score = col_double(),
##   tournament = col_character(),
##   city = col_character(),
##   country = col_character(),
##   neutral = col_logical()
## )
```

1. Select variables `date`, `home_team` and `away_team`.
2. Subset games with **Brazil** as the home team.
3. Choose the games that Brazil won as the home team, and select variables `date`, `away_team` and `tournament`.
4. Add the difference of goals, and an indicator variable called `goleada` for when the difference of goals is large, and select what we did only for Brazil. **Hint: use `ifelse`**.
5. What was the largest difference in goals within these games?
6. The top 5 `goleadas`?
7. Summary on goals scored by home teams, such as `mean` of `home_score` and `away_score`, `std`, using `group_by` and `summarise`
8. Proportion of victories of **Brazil** on different tournaments against each opponent, for instance, **Argentina**.