

# FIFA World Cup

Lauren Schmiedeler

```
library(tidyverse)

data <- readRDS("data.RData")

matches <- data[[1]] %>% mutate(country = factor(country), city = factor(city), stage = factor(stage), home_team = factor(home_team), away_team = factor(away_team), outcome = factor(outcome), win_conditions = factor(win_conditions), winning_team = factor(winning_team), losing_team = factor(losing_team), month = factor(month), dayofweek = factor(dayofweek))
matches$winning_team[matches$winning_team == "Portagul"] <- "Portugal"
summary(matches)
```

```

##      year      country      city      stage
## Min.   :1930   Germany:102   Mexico City   : 23   Round of 16   : 89
## 1st Qu.:1970   Brazil : 86   Montevideo    : 18   Group B       : 69
## Median :1990   Mexico : 84   Guadalajara   : 17   Group A       : 68
## Mean   :1987   France : 82   Johannesburg   : 15   Quarterfinals: 66
## 3rd Qu.:2006   Italy  : 69   Rio de Janeiro: 15   Group 1       : 59
## Max.   :2018   Russia : 64   Buenos Aires  : 12   Group C       : 57
##                (Other):413   (Other)       :800   (Other)       :492
##      home_team  away_team  home_score  away_score  outcome
## Brazil      : 84   Uruguay   : 37   Min.      : 0.000   Min.      :0.000   A:302
## Argentina   : 60   Italy     : 34   1st Qu.: 0.000   1st Qu.:0.000   D:169
## Italy        : 49   England   : 33   Median   : 1.000   Median   :1.000   H:429
## West Germany: 43   Spain     : 33   Mean     : 1.569   Mean     :1.262
## France       : 40   Mexico    : 31   3rd Qu.: 2.000   3rd Qu.:2.000
## England      : 36   Yugoslavia: 28   Max.     :10.000   Max.     :8.000
## (Other)      :588   (Other)   :704
##                win_conditions  winning_team  losing_team
## Italy won in AET                : 5   Brazil      : 76   Mexico     : 29
## Argentina won in AET            : 3   Argentina   : 47   Argentina: 24
## England won in AET              : 3   Italy       : 46   England    : 22
## Argentina won in penalties (4 - 3): 2   West Germany: 40   France     : 21
## Belgium won in AET              : 2   France      : 36   Spain      : 21
## (Other)                        : 47   (Other)     :486   (Other)    :614
## NA's                          :838   NA's        :169   NA's       :169
##      date      month      dayofweek
## Min.   :1930-07-13   Jul:150   Friday     : 92
## 1st Qu.:1970-06-14   Jun:727   Monday      : 82
## Median :1990-06-23   May: 23   Saturday    :152
## Mean   :1987-05-20                Sunday     :196
## 3rd Qu.:2006-06-19                Thursday   :111
## Max.   :2018-07-15                Tuesday    :119
##                Wednesday:148

```

```

cups <- data[[2]] %>% mutate(host = factor(host), winner = factor(winner), second = factor(second), third = factor(third), fourth = factor(fourth))
summary(cups)

```

```
##      year      host      winner      second      third
## Min.   :1930   Brazil   : 2   Brazil    :5   Argentina :3   Germany:3
## 1st Qu.:1958   France   : 2   Italy     :4   Netherlands :3   Brazil :2
## Median :1978   Germany  : 2   West Germany:3   West Germany :3   France :2
## Mean   :1977   Italy    : 2   Argentina  :2   Brazil       :2   Poland :2
## 3rd Qu.:1998   Mexico   : 2   France     :2   Czechoslovakia:2   Sweden :2
## Max.    :2018   Argentina: 1   Uruguay    :2   Hungary      :2   Austria:1
##          (Other) :10   (Other)    :3   (Other)      :6   (Other):9
##      fourth  goals_scored      teams      games
## Uruguay   : 3   Min.    : 70.0   Min.    :13.00   Min.    :17.00
## Brazil    : 2   1st Qu.: 89.0   1st Qu.:16.00   1st Qu.:32.00
## England   : 2   Median :126.0   Median :16.00   Median :38.00
## Yugoslavia: 2   Mean    :121.3   Mean    :21.76   Mean    :42.86
## Austria   : 1   3rd Qu.:146.0   3rd Qu.:32.00   3rd Qu.:64.00
## Belgium   : 1   Max.    :171.0   Max.    :32.00   Max.    :64.00
## (Other)   :10
##      attendance
## Min.    : 395000
## 1st Qu.: 943000
## Median :1774022
## Mean    :1898122
## 3rd Qu.:2724604
## Max.    :3568567
##
```

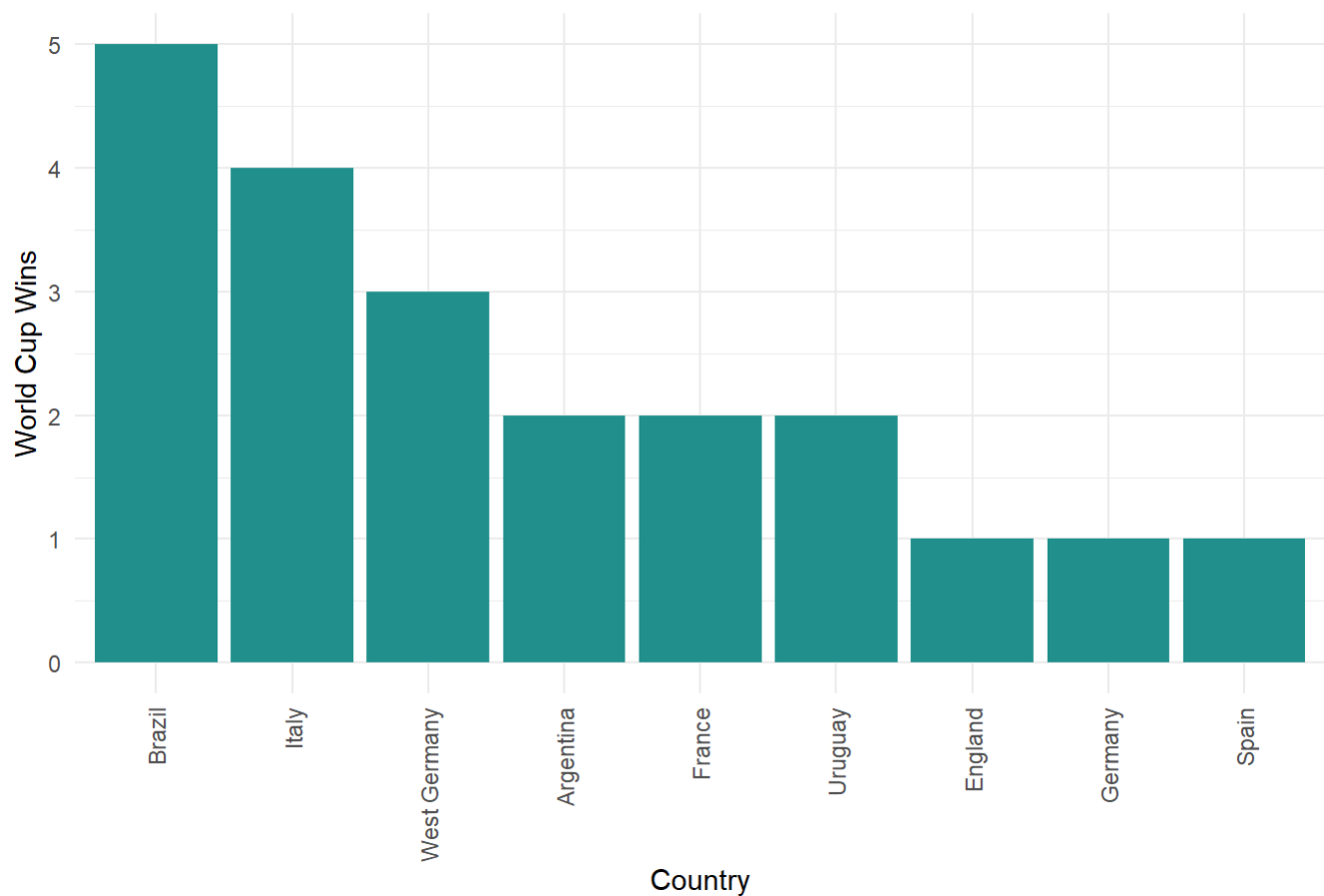
```
library(paletteer)
```

```
total_cup_wins <- cups %>% group_by(winner) %>%
  summarize(total_cup_wins = n()) %>%
  arrange(-total_cup_wins)
```

```
pal <- paletteer_c("viridis::viridis", 3)
```

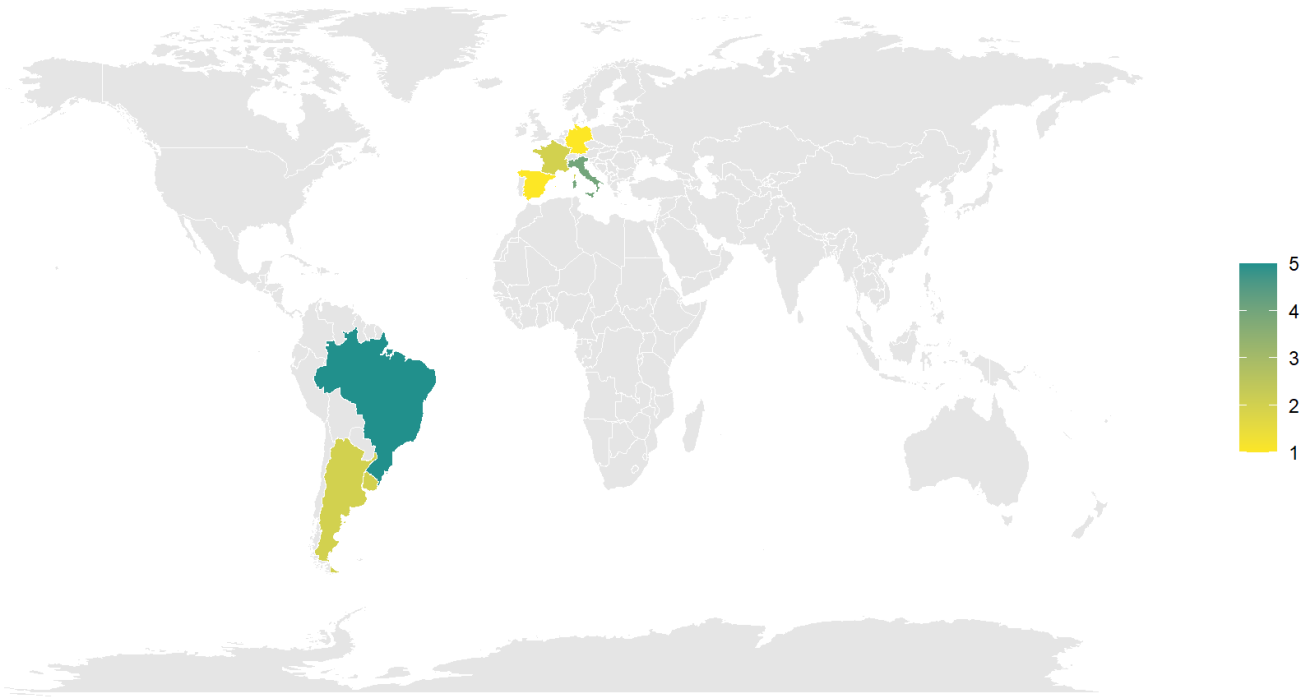
```
ggplot(total_cup_wins, aes(x = reorder(winner, -total_cup_wins), y = total_cup_wins)) +
  geom_bar(stat = "identity", fill = pal[2]) +
  theme_minimal() +
  labs(x = "Country", y = "World Cup Wins", title = "World Cup Wins by Country") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5))
```

World Cup Wins by Country



```
map_data("world") %>%
  merge(total_cup_wins, by.x = "region", by.y = "winner", all.x = T) %>%
  arrange(group, order) %>%
  ggplot(aes(x = long, y = lat, group = group, fill = total_cup_wins)) +
  geom_polygon(color = "white", size = 0.2) +
  scale_fill_gradient(low = pal[3], high = pal[2], na.value = "gray90") +
  theme_minimal() +
  theme(axis.text = element_blank(), axis.title = element_blank(), panel.grid = element_blank(),
        legend.title = element_blank())
```

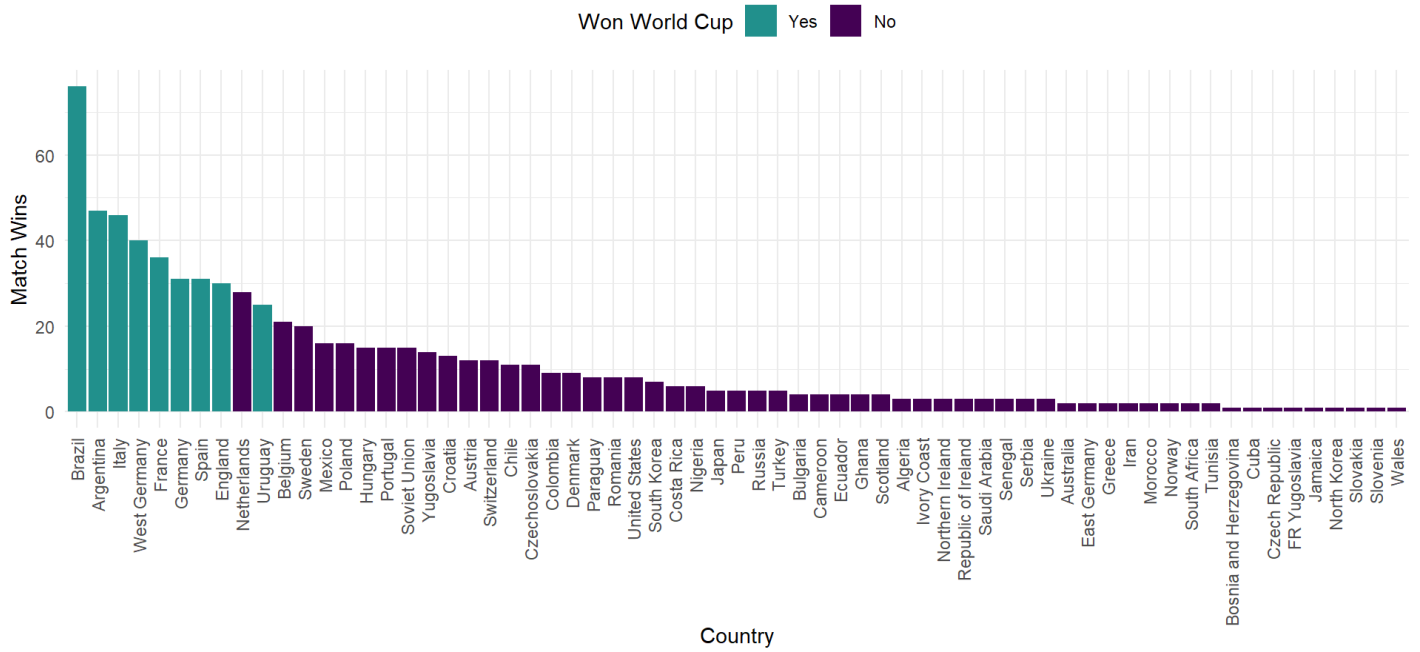
```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
```



```
total_match_wins <- matches %>% group_by(winning_team) %>%
  summarize(total_match_wins = n()) %>%
  na.omit() %>%
  arrange(-total_match_wins) %>%
  mutate(won_world_cup = ifelse(winning_team %in% total_cup_wins$winner, "Yes", "No"))
total_match_wins$won_world_cup <- factor(total_match_wins$won_world_cup, levels = c("Yes", "No"))

ggplot(total_match_wins, aes(x = reorder(winning_team, -total_match_wins), y = total_match_wins,
  fill = won_world_cup)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(x = "Country", y = "Match Wins", title = "Match Wins by Country", fill = "Won World Cup")
+
  scale_fill_manual(values = c(pal[2], pal[1])) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5),
    legend.position = "top")
```

## Match Wins by Country



```
map_data("world") %>%
  merge(total_match_wins, by.x = "region", by.y = "winning_team", all.x = T) %>%
  arrange(group, order) %>%
  ggplot(aes(x = long, y = lat, group = group, fill = total_match_wins)) +
  geom_polygon(color = "white", size = 0.2) +
  scale_fill_gradient(low = pal[2], high = pal[1], na.value = "gray90") +
  theme_minimal() +
  theme(axis.text = element_blank(), axis.title = element_blank(), panel.grid = element_blank(),
        legend.title = element_blank())
```

