3\_visualization.r

abhia

2019-11-18

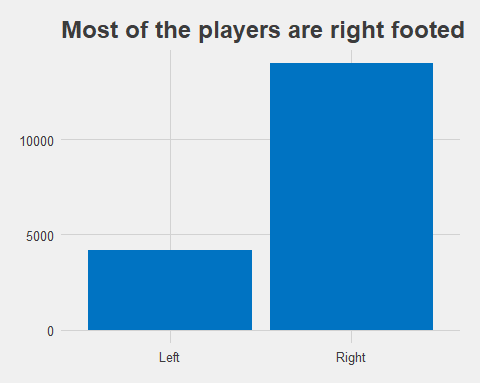
fifa = read.csv("1\_cleaned\_data.csv")  
library("ggplot2")  
library("ggthemes")  
library("magrittr")  
library("dplyr")

##   
## Attaching package: 'dplyr'

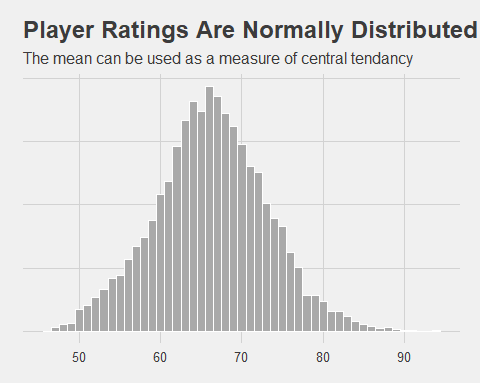
## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

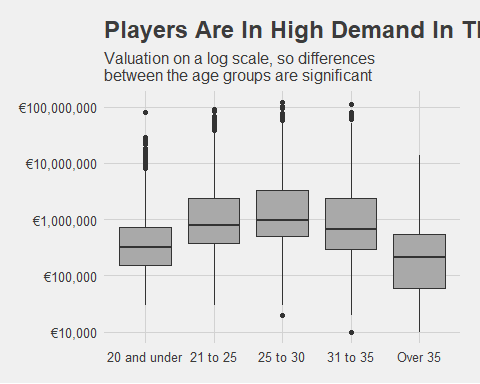
library("scales")  
options(scipen = 999)  
  
#FREQUENCY PLOT OF FOOT PREFERRENCE  
print(ggplot(fifa, aes(Preferred.Foot)) +geom\_bar(fill = "#0073C2FF") + ggtitle("Most of the players are right footed") + theme\_fivethirtyeight())



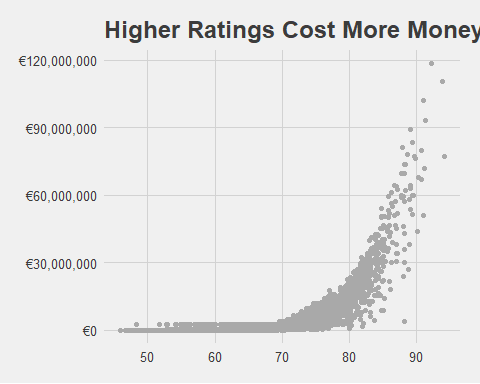
#HISTOGRAM OF PLAYER RATING  
print(fifa %>%  
 ggplot(aes(x= Overall)) +  
 geom\_histogram(color = "white", fill = "darkgrey", binwidth = 1) +  
 ggtitle("Player Ratings Are Normally Distributed", subtitle = "The mean can be used as a measure of central tendancy") +  
 theme\_fivethirtyeight() +  
 theme(axis.text.y = element\_blank()))



#BOXPLOT OF AGE VS VALUE  
fifa <- fifa %>%  
 mutate(AgeGroup = ifelse(Age <= 20, "20 and under", ifelse(Age > 20 & Age <=25, "21 to 25", ifelse(Age > 25 & Age <= 30, "25 to 30", ifelse(Age > 30 & Age <= 35, "31 to 35", "Over 35")))))  
print(fifa %>%  
 ggplot(aes(x= AgeGroup, y= Value)) +  
 geom\_boxplot(fill = "darkgrey") +  
 scale\_y\_log10(labels = dollar\_format(prefix = "€")) +  
 ggtitle("Players Are In High Demand In Their Mid-20s", subtitle = "Valuation on a log scale, so differences \nbetween the age groups are significant") +  
 theme\_fivethirtyeight())



#SCATTERPLOT OF VALUE VS OVERALL RATING  
print(fifa %>%  
 ggplot(aes(x= Overall, y= Value)) +  
 geom\_point(position = "jitter", color = "darkgrey") +  
 ggtitle("Higher Ratings Cost More Money") +  
 scale\_y\_continuous(labels = dollar\_format(prefix = "€")) +  
 theme\_fivethirtyeight())



# FREQUENCY PLOT OF TEAM VS TOTAL WAGE  
print(fifa %>%   
 group\_by(Club) %>%   
 summarise(TotalWages = sum(Wage, na.rm = T),) %>%  
 arrange(desc(TotalWages)) %>% head(n= 20) %>%  
 ggplot(aes(x= reorder(Club, TotalWages), y= TotalWages)) +  
 geom\_col(colour = "black") +  
 scale\_y\_continuous(labels = dollar\_format(prefix = "€")) +  
 coord\_flip() +  
 ggtitle("The 20 highest wage bills in FIFA19 and how much one rating point costs in wages") +  
 theme\_fivethirtyeight() +  
 theme(legend.position = "none"))

