

FIVETHIRTYEIGHT DATA

WHAT DO MEN THINK IT MEANS TO BE A MAN?

Homework #2

Report by:

Benita Diop

Baruch College - Zicklin School of Business

M.S in Statistics - Data Science

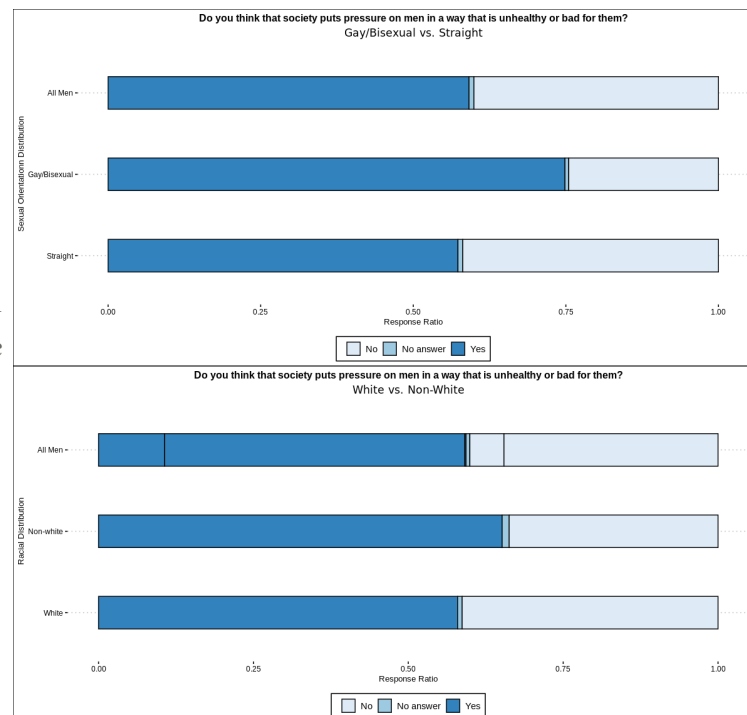
Course: Software Tools for Data Analysis

*** Please use your zoom (CTRL and +) to get a clear view of graphs

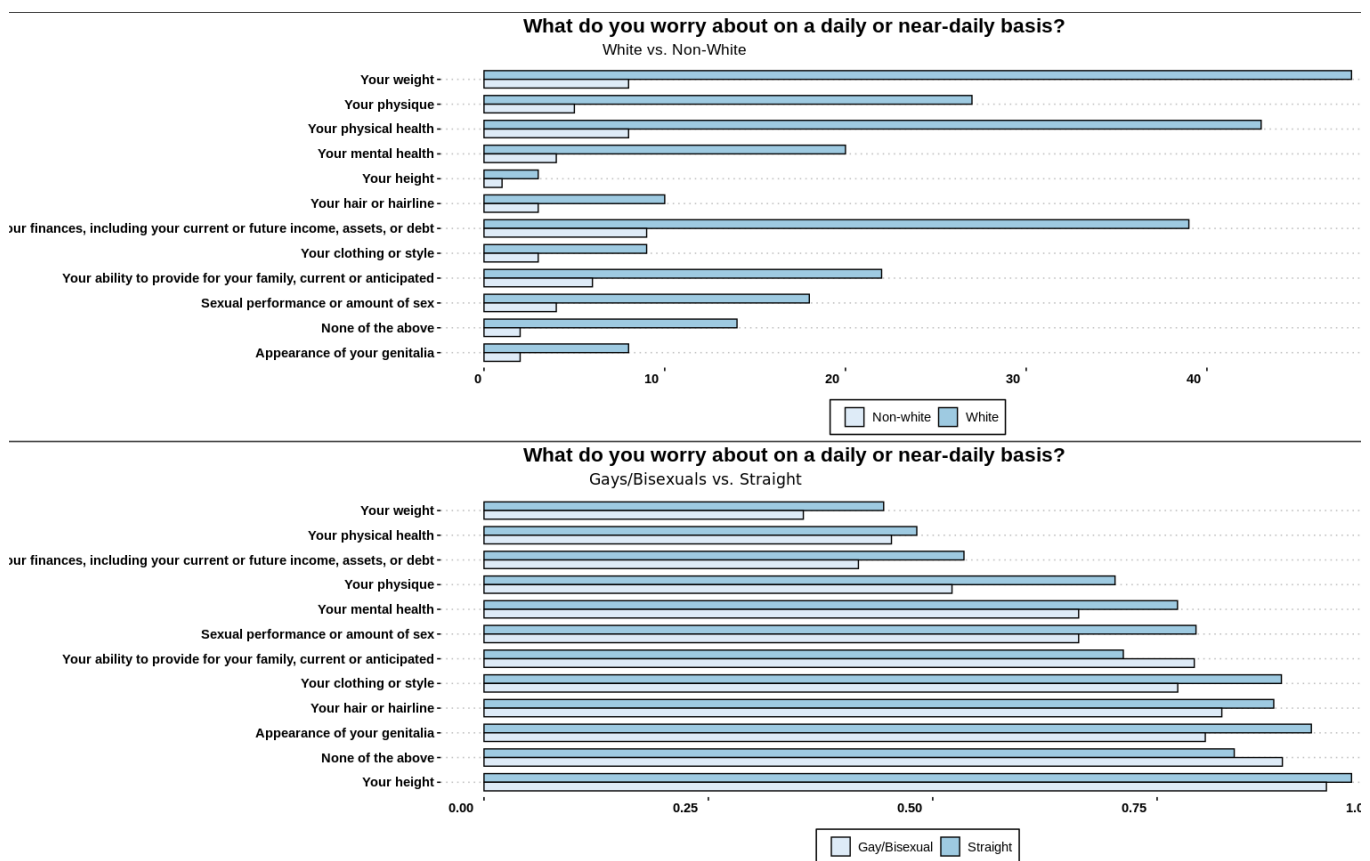
DISCUSSION:

When analyzing the survey responses through the lense of men that identify as “Gay/Bisexual” vs the alternative and men that Identify as “White” vs the alternative the analysis gave back insightful feedback. For the question; “Do you think that society puts pressure on men in a way that is unhealthy or bad for them?”, the original data showed that men within the age group of [18-34] responded with the most affinity to this sentiment vs their alternative; men [18-34] and men [65+]. When the survey was reproduce it showed that men who identify as gay and bisexual resonated with the sentiment more than their straight counterparts. Additionally, White men resonated with the question much less than their Non-White counterparts. Reviewing the data collectively, I believe that I can make the assumption that the subset of men age [18-34] who identify as gay/bisexual and who are not white will feel the most pressure from society. In terms of analyzing what men worry about, the reproduction of the data shows that all men worry about their weight a great deal, however, for Non-white men their physique it the greatest worry. In terms of sexual orientation, gay/bisexual men tend to have the same worries as straight men.

P L O T 1



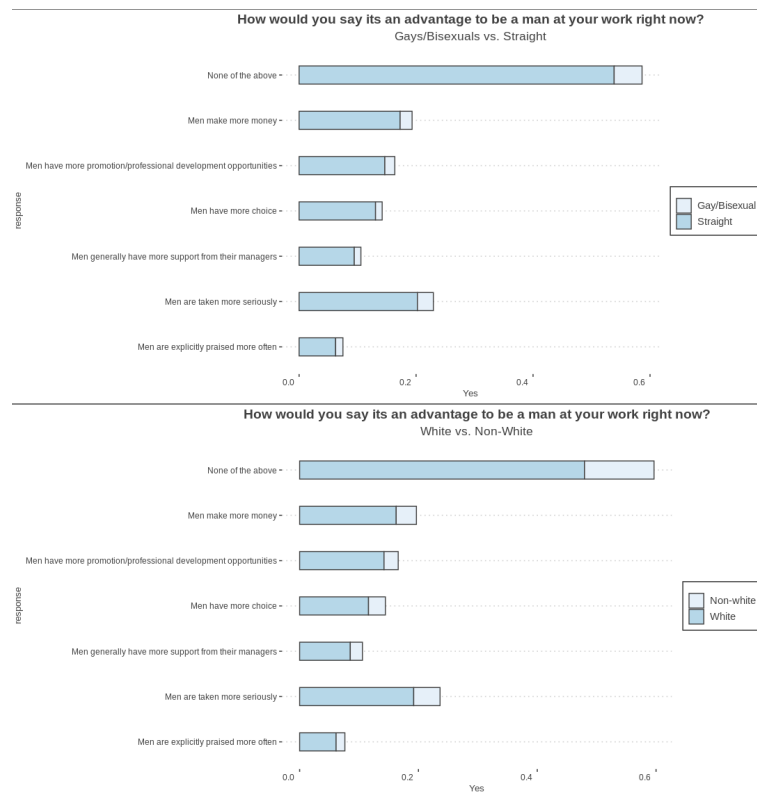
P L O T 2



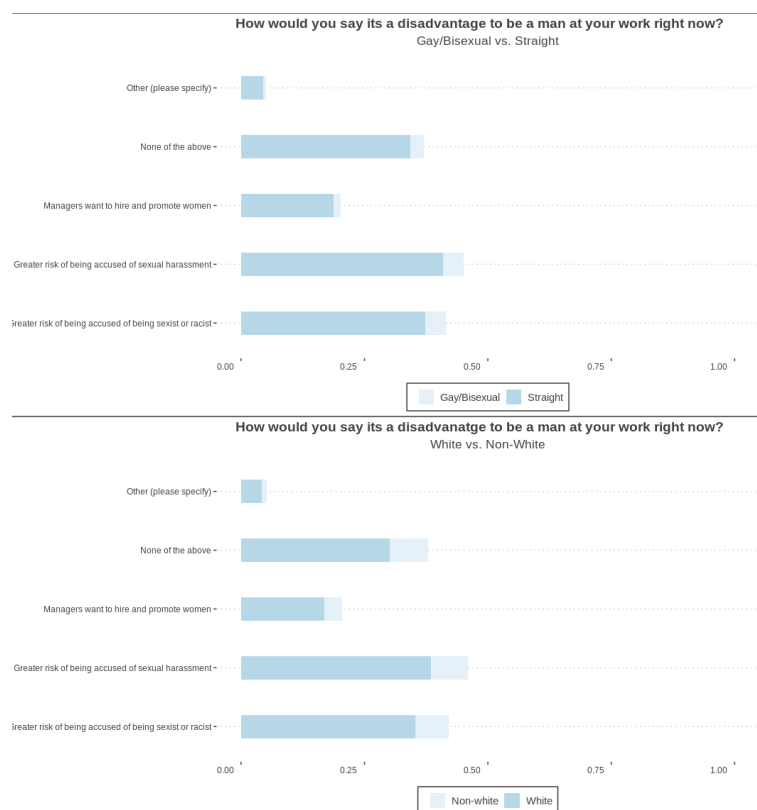
DISCUSSION:

From the original advantage vs. disadvantage plots on fivethirtyeight, most men did not feel that the options from the survey applies to them. I would categorize this particular survey question type as bias because it creates a lot of unknown in our analysis because many respondents do not resonate with the given option. However among the options that the respondents chose; when it comes to choosing what you consider to be an advantage to being a man; most gay/bisexual, straight, white and non-white males mutually agree that one of the advantages of being a man is that men are easily taken more seriously. Alternatively, when it comes to the disadvantages of being a man in the workplace most gay/bisexual, straight, white and non-white males mutually agree that men are at a greater risk of being accused of sexual harassment.

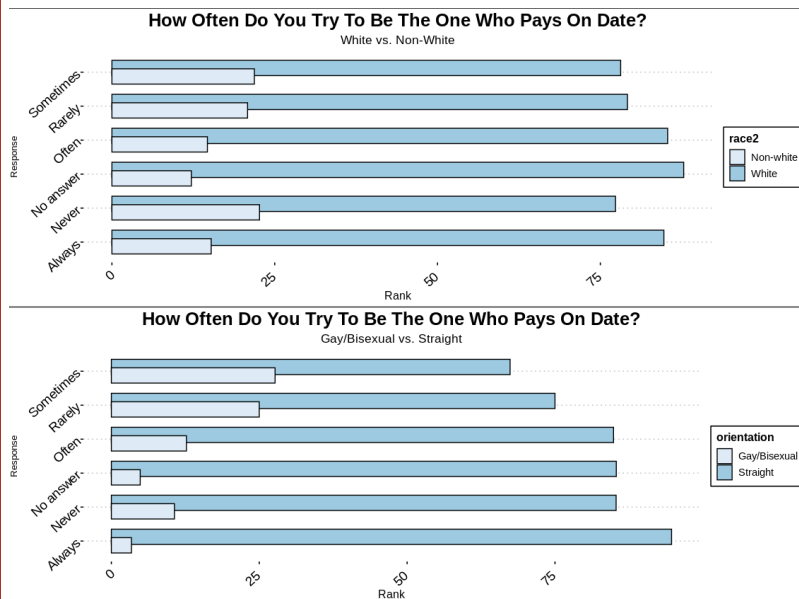
PLOT 3



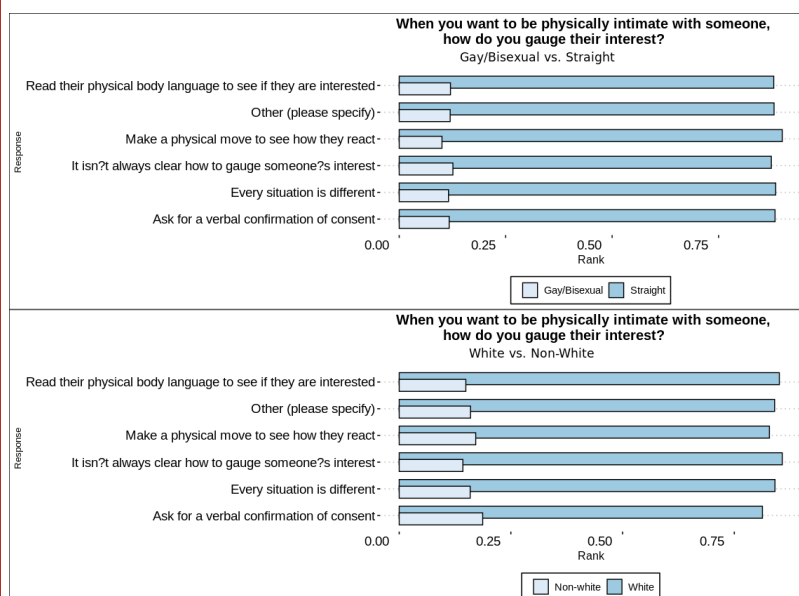
PLOT 4



P L O T 5

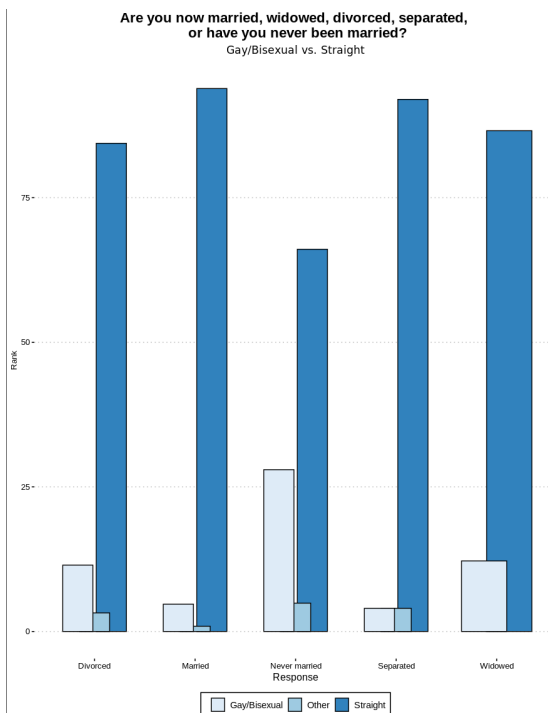


P L O T 6

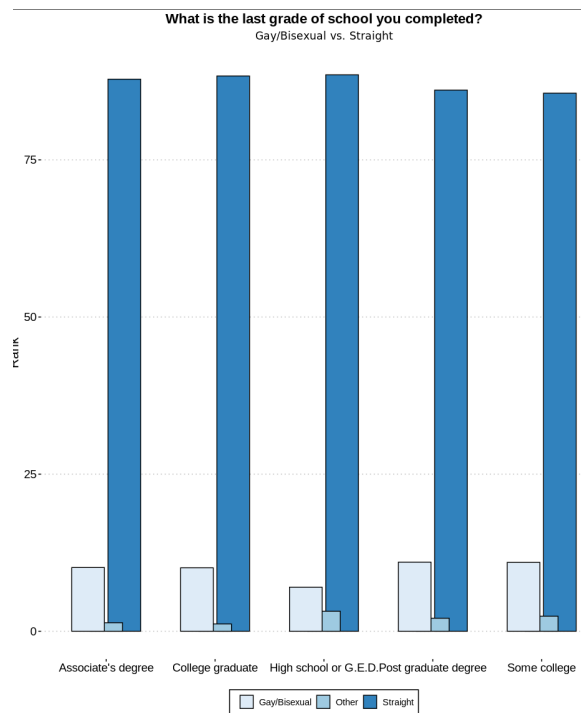


DISCUSSION:

In terms of dating and relationships for me, the original survey plot indicates that men try to pay for more dates as they age. When it comes to white vs non-white; white men tend to either always or often pay for dates. However, non-white men tend to never pay on dates. For gay/bisexual men vs. straight; straight men always try to be the one who pays on dates, while, gay/bisexual men tend to sometimes or rarely pay for dates. When it comes to physical intimacy and gauging an individual's interest; straight men and non-white men are more inclined to “make a move to see how the individual will react”, while gay/bisexual men and white men voted that its not always clear how to gauge an individuals interest. From this, I deduce that straight and non-white men tend to be more forward in romantic, intimate experiences.



PLOT 7



PLOT 8

DISCUSSION:

Two supplemental questions from the survey that I chose to analyze are 1) “Are you now married, widowed, divorced, separated, or have you never been married?” and 2) “What is the last grade of school you completed?”. I compared the survey results with men who identify as Gay/Bisexual, men who identify as straight and men who identify as other.

In terms of marriage, a roughly equal ratio of straight men have been married and separated. Gay/Bisexual men and “Other” men tend to mostly never marry or be divorced.

In terms of education, straight identifying respondent have equal ratio for men who have an associates degree, college graduate or High School or G.E.D. For Gay/ Bisexual men they are least likely to just have High School or G.E.D; many have pursued additional schooling.



Homework #2

C O D E

```
rnorm_round <-
  function(n, mean, sd) {
    return(round
      (rnorm
        (n, mean = mean, sd=sd)
      )
    )
  }
```

```
Team <- c("Jersey Lions", "Westchester Cats", "Long Island Tigers", "Staten Island Dogs", "The Bronx Foxes", "Queens Bears", "Manhattan Ducks", "Brooklyn Rats")
A <- c(7, 6, 7, 8, 9, 6, 7, 8)
D <- c(7,8,6,6,6,9,6,7)
H <- c(3,3,4,3,3,5,3,4)
R <- c(0.5,0.25, 0.3, 0.25, 0.2, 0.35, 0.2, 0.25)
Basketball_League <- data.frame(Team, A, D, H, R) %>% arrange(Team) %>% column_to_rownames("Team")
Basketball_League
```

```
##           A D H   R
## Brooklyn Rats      8 7 4 0.25
## Jersey Lions       7 7 3 0.50
## Long Island Tigers 7 6 4 0.30
## Manhattan Ducks    7 6 3 0.20
## Queens Bears       6 9 5 0.35
## Staten Island Dogs 8 6 3 0.25
## The Bronx Foxes    9 6 3 0.20
## Westchester Cats   6 8 3 0.25
```

```
basketball_game_sim = function(x, y) {
  x <- Basketball_League[x, ]
  y <- Basketball_League[y, ]
  mean = (0.6*x$A + 0.4*x$D) - (0.4*y$A + 0.6*y$D) + x$H
  sd <- 1/x$R + 1/y$R
  spread <- 0
  while(spread == 0) {
    spread <- rnorm_round(1, mean = mean, sd = sd)
  }
  return(spread)
}
basketball_game_sim("Jersey Lions", "Queens Bears")
```

```
## [1] 7
```

```
season <- permutations(8, 2, row.names(Basketball_League), repeats.allowed = F) %>%
  rbind(., .) %>%
  data.table()
colnames(season) <- c("x", "y")
season
```

```
##           x           y
## 1: Brooklyn Rats Jersey Lions
## 2: Brooklyn Rats Long Island Tigers
## 3: Brooklyn Rats Manhattan Ducks
## 4: Brooklyn Rats Queens Bears
## 5: Brooklyn Rats Staten Island Dogs
## ---
## 108: Westchester Cats Long Island Tigers
## 109: Westchester Cats Manhattan Ducks
## 110: Westchester Cats Queens Bears
## 111: Westchester Cats Staten Island Dogs
## 112: Westchester Cats The Bronx Foxes
```

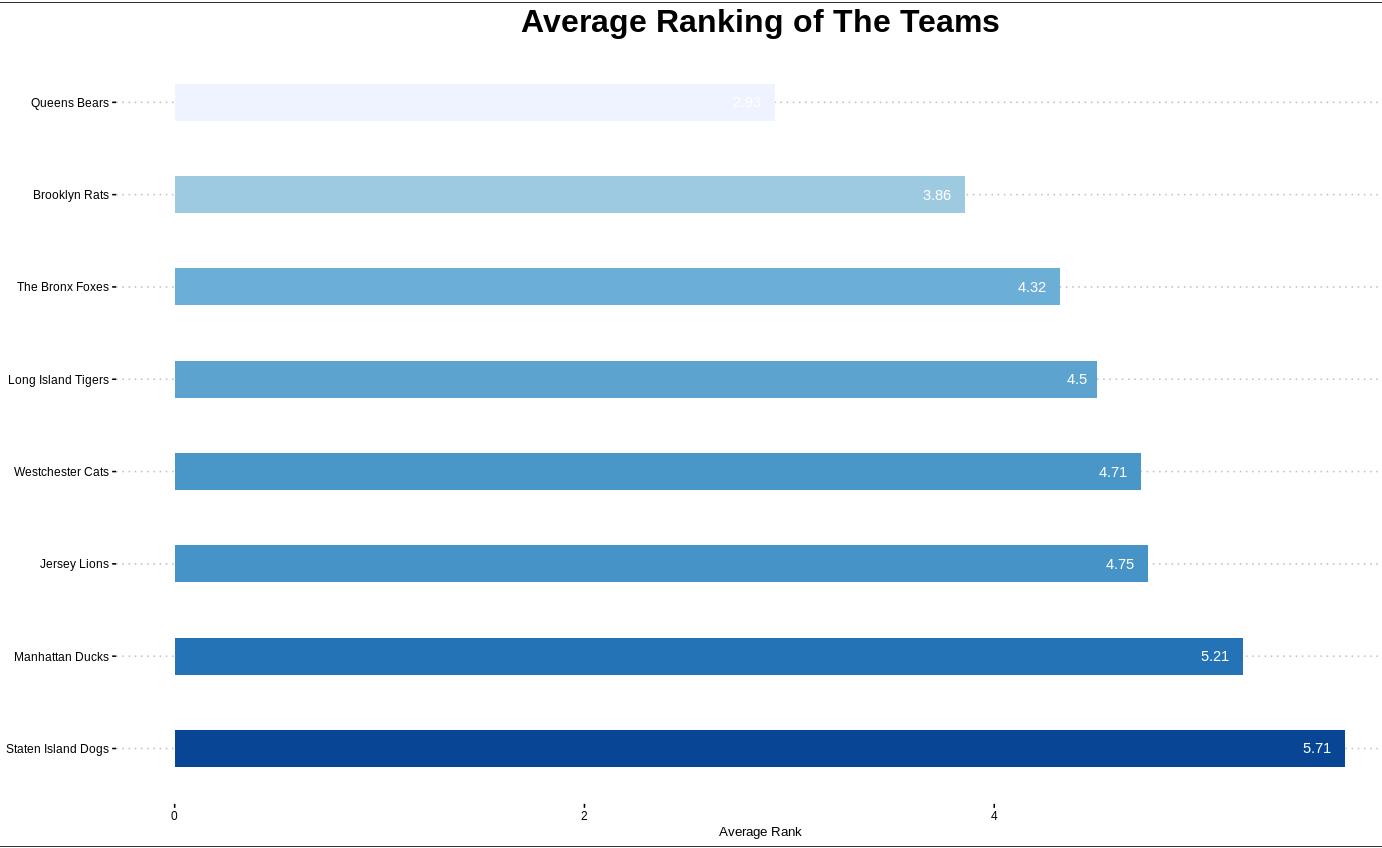
DISCUSSION :

Comparing the two systems, I believe the initial system offers the teams an opportunity to compete in more games as opposed to the matchup tournament. If I was to recommend a system to the league organizers, I would recommend the game where each team plays 28 games that way you can get a better scope of the abilities of the given team. In terms of fairness, I believe that the tournament is the least fair because there is no randomness to the pairing of the teams.

Since, the highest ranked team (top-dog) is always set to compete with the least-ranked team (underdog), this immediately gives the highest ranked team an advantage.

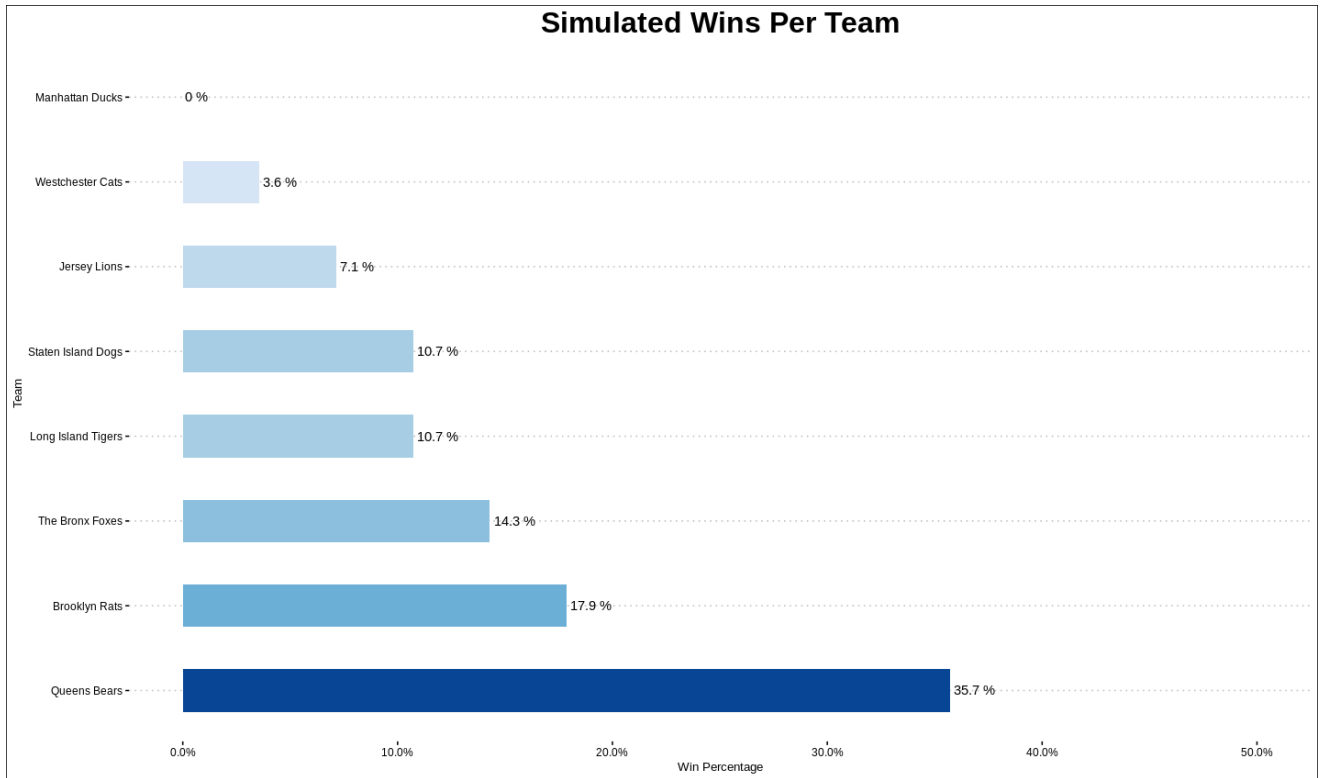
Additionally, in the matchup, the highest ranked team plays at home during the first game and plays at home again during an event of a tie. I believe this gives the highest rank team home advantage. Clearly, the matchup favors the highest ranked team and its unfair.

Results

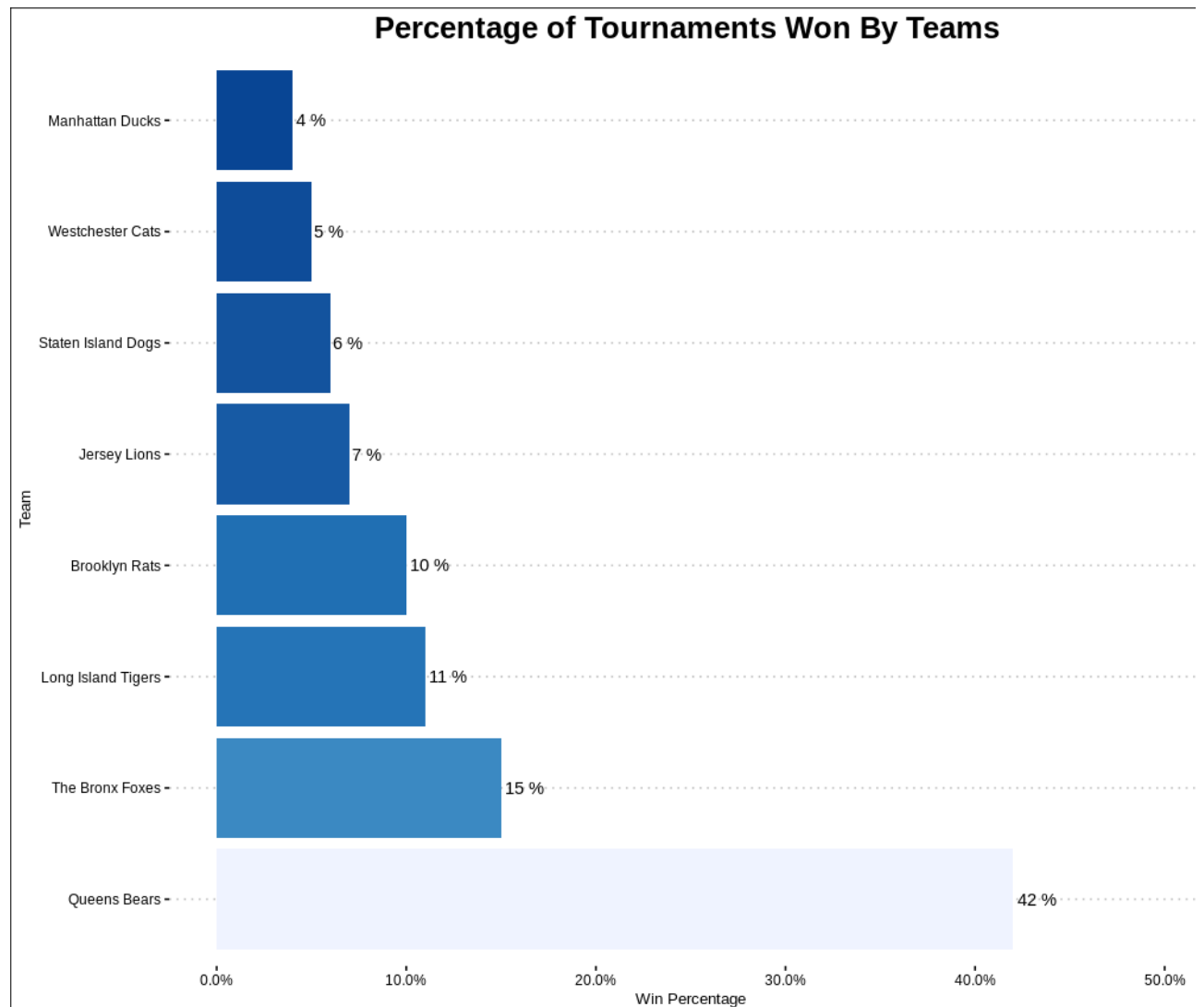


Results

Simulated Wins Per Team



Percentage of Tournaments Won By Teams



RESTAURANT

Homework #2

CODE

Invest in the marketing campaign !

Invest in the marketing campaign because the 85% return on investment is worth the investment.

```
# Defining Variables
SALESnorm_WD_mean = 2000
SALESnorm_WD_sd = 500

SALESnorm_W3_mean = 3000
SALESnorm_W3_sd = 700

rent = 2500
laborCost = 4500
foodCost = 4500
other = 2500
TotalExpenses = (rent + laborCost + foodCost + other)
MarketingCost = 10000
# probability of a successful marketing campaign
p1 <- 0.8
p0 <- 0.2
```

```
# Adjusted Daily Means - Marketing Accounted For
new_WD_mean = (SALESnorm_WD_mean + .15*SALESnorm_WD_mean)
new_n3_mean = (SALESnorm_W3_mean + .15*SALESnorm_W3_mean)

# Weekly Profits - with and without marketing
NoMarketingMean <- 5 *SALESnorm_WD_mean + 2 *SALESnorm_W3_mean - TotalExpenses
week_Tmean <- 5 *new_WD_mean + 2 *new_n3_mean - TotalExpenses

# Yearly Profits
AnnualProfit <- 52*(5*SALESnorm_WD_mean + 2*SALESnorm_W3_mean - TotalExpenses)

# Forecasted Profits based of marketing success or failure
Failure <- 52 * NoMarketingMean - MarketingCost
Success<- 52 * week_Tmean - MarketingCost
Weighted_Year_Profit <- p1 * Success + p0 * Failure
```

```
nsims = 1e3
randomweek = numeric(nsims)
for (i in 1:nsims) {
  sim_weekday = sample(x = nsims , size = 5) %>% rnorm(mean = SALESnorm_WD_mean, sd = SALESnorm_WD_sd)
  sim_weekend = sample(x= nsims, size = 2) %>% rnorm(mean = SALESnorm_W3_mean, sd= SALESnorm_W3_sd)
  week <- sum(sim_weekend) + sum(sim_weekday) - TotalExpenses
  randomweek[i] <- week
}

print(map_lgl(randomweek, function(x) ifelse(x > 0, T, F)) %>% mean())
```

```
## [1] 0.927
```

```
nsims = 5000
randomMonth = numeric(nsims)
for ( i in 1:nsims) {
  sim_weekday = sample(x = nsims , size = 22) %>% rnorm(mean = SALESnorm_WD_mean, sd = SALESnorm_WD_sd)
  sim_weekend = sample(x= nsims, size = 9) %>% rnorm(mean = SALESnorm_W3_mean, sd= SALESnorm_W3_sd)
  December <- sum(sim_weekend) + sum(sim_weekday) - (TotalExpenses * 32/7)
  randomMonth[i] <- December
}

map_lgl(randomMonth, function(x) ifelse(x> 0, F, T)) %>% mean()
```

```
## [1] 0.0128
```

```

nsims = 5000
randomYear = numeric(nsims)
for ( i in 1:nsims) {
  sim_weekday = sample(x = nsims , size = (5*52)) %>% rnorm(mean = SALESnorm_WD_mean, sd = SALESnorm_WD_sd)
  sim_weekend = sample(x= nsims, size = (2*52)) %>% rnorm(mean = SALESnorm_W3_mean, sd= SALESnorm_W3_sd)
  Year <- sum(sim_weekend) + sum(sim_weekday) - (TotalExpenses * 52)
  randomYear[i] <- Year
}

print("Simulated Annual Profit:")

```

```
## [1] "Simulated Annual Profit:"
```

```
mean(randomYear)
```

```
## [1] 103973.3
```

```
print("Calculated Annual Profit:")
```

```
## [1] "Calculated Annual Profit:"
```

```
AnnualProfit
```

```
## [1] 104000
```

```

n = 1e4
ProjectProfit <- c()
for (i in rbinom(n, 1, 0.8)) {
  if(i == 1){
    ProjectProfit <- c(ProjectProfit, Success)
  }
  else{
    ProjectProfit<- c(ProjectProfit, Failure)
  }
}

print("Simulated Annual Profit With Marketing Campagin")

```

```
## [1] "Simulated Annual Profit With Marketing Campagin"
```

```
mean(ProjectProfit)
```

```
## [1] 194626.2
```

```
print("Simulated Annual Profit Without Marketing")
```

```
## [1] "Simulated Annual Profit Without Marketing"
```

```
mean(randomYear)
```

```
## [1] 103973.3
```

```
print(" % Profit Increase:")
```

```
## [1] " % Profit Increase:"
```

```
round(100*((mean(ProjectProfit) - mean(randomYear))/mean(randomYear)))
```

```
## [1] 87
```