

data-science-project

Does an individual's track affect whether they are placed or not?

Tracks: - Commerce - Arts - Science - Sci&Tech - Comm&Mgmt - Others - Mkt&HR - Mkt&Fin

```
# Packages
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.2
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.6.2
```

```
##
## 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
```

```
campusRecruitment <- read.csv('campus-recruit.csv', check.names=FALSE)

# Total number of students
TotalNumStudents <- 21285

# Function for average
avg_num <- function(num1) {
  print(num1/TotalNumStudents)
}

# Total number of students who placed per High School track
# Commerce
placedCommerce <- filter(campusRecruitment, campusRecruitment$`HSC 12th Grade Track` ==
  "Commerce" & campusRecruitment$`Placement Status` == "Placed")
placedCommerceTotal <- nrow(placedCommerce)
placedCommerceTotal
```

```
## [1] 7821
```

```
# 37% of students who studied Commerce placed  
avgPlacedCommerce <- avg_num(placedCommerceTotal)
```

```
## [1] 0.3674419
```

```
# Arts  
placedArts <- filter(campusRecruitment, campusRecruitment$`HSC 12th Grade Track` == "Arts" & campusRecruitment$`Placement Status` == "Placed")  
placedArtsTotal <- nrow(placedArts)  
placedArtsTotal
```

```
## [1] 594
```

```
# 3% of students who studied Commerce placed  
avgPlacedArts <- avg_num(placedArtsTotal)
```

```
## [1] 0.02790698
```

```
# Science  
placedScience <- filter(campusRecruitment, campusRecruitment$`HSC 12th Grade Track` == "Science" & campusRecruitment$`Placement Status` == "Placed")  
placedScienceTotal <- nrow(placedScience)  
placedScienceTotal
```

```
## [1] 6237
```

```
# 29% of students who studied Commerce placed  
avgPlacedScience <- avg_num(placedScienceTotal)
```

```
## [1] 0.2930233
```

```
# SciTech  
placedScienceTech <- filter(campusRecruitment, campusRecruitment$`HS Track` == "Sci&Tech" & campusRecruitment$`Placement Status` == "Placed")  
placedScienceTechTotal <- nrow(placedScienceTech)  
placedScienceTechTotal
```

```
## [1] 4059
```

```
# 19% of students who studied Sci&Tech placed  
avgPlacedScienceTech <- avg_num(placedScienceTechTotal)
```

```
## [1] 0.1906977
```

```
# CommMgmt
placedCommMgmt <- filter(campusRecruitment, campusRecruitment$`HS Track` == "Comm&Mgmt"
  & campusRecruitment$`Placement Status` == "Placed")
placedCommMgmtTotal <- nrow(placedCommMgmt)
placedCommMgmtTotal
```

```
## [1] 10098
```

```
# 47% of students who studied Commerce placed
avgPlacedCommMgmt <- avg_num(placedCommMgmtTotal)
```

```
## [1] 0.4744186
```

```
# Others
placedOthers <- filter(campusRecruitment, campusRecruitment$`HS Track` == "Others" & cam
  pusRecruitment$`Placement Status` == "Placed")
placedOthersTotal <- nrow(placedOthers)
placedOthersTotal
```

```
## [1] 495
```

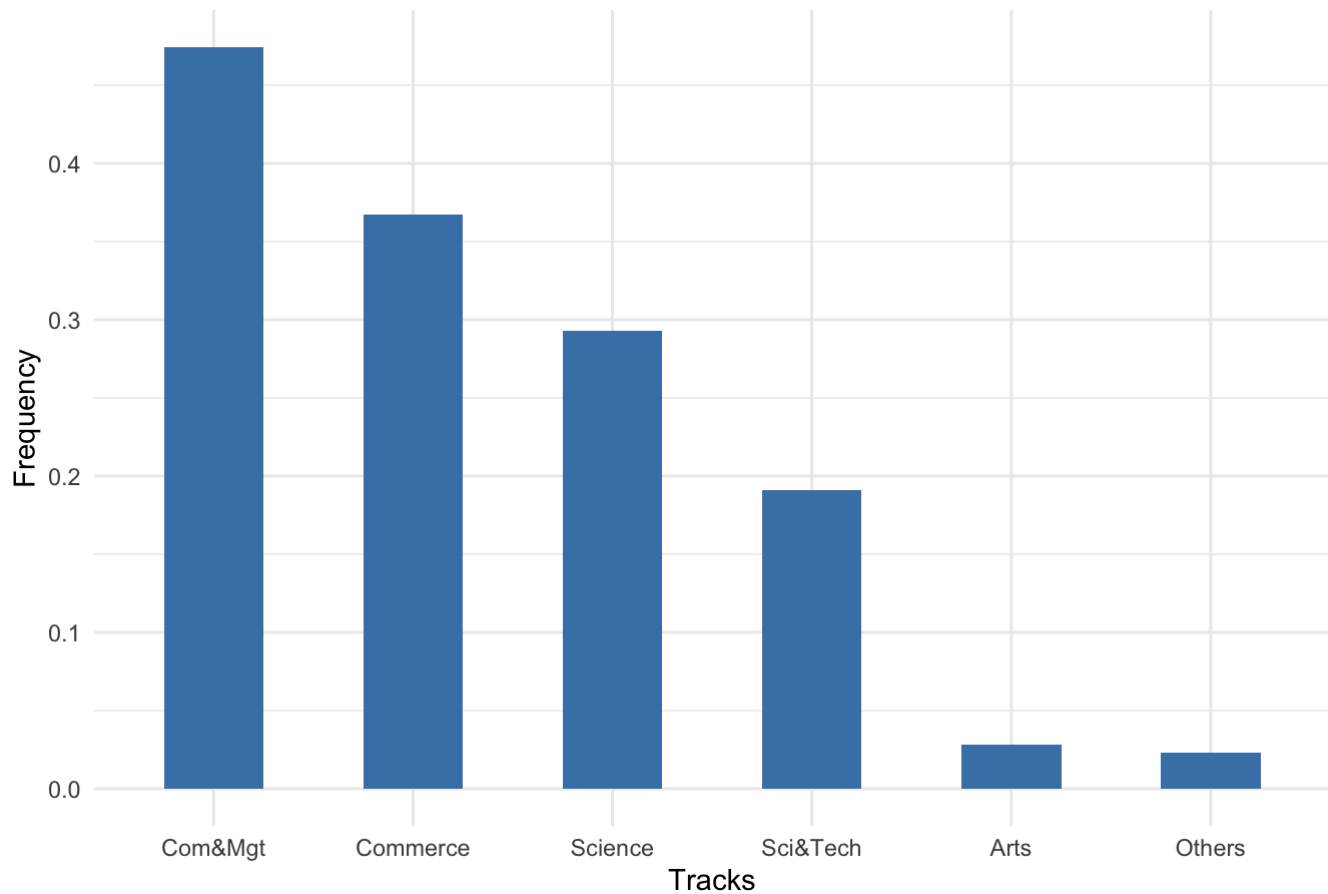
```
# 2% of students who studied Others placed
avgPlacedOthers <- avg_num(placedOthersTotal)
```

```
## [1] 0.02325581
```

```
# HS Track Bar Chart
avgPlacedPerTrack <- c(avgPlacedCommerce, avgPlacedArts, avgPlacedScience, avgPlacedScie
  nceTech, avgPlacedCommMgmt, avgPlacedOthers)
Tracks <- c("Commerce", "Arts", "Science", "Sci&Tech", "Com&Mgt", "Others")
dfTracks <- data.frame(Tracks, avgPlacedPerTrack)
dfTracks <- dfTracks[order(-dfTracks$avgPlacedPerTrack),]

ggplot(data=dfTracks, aes(x=reorder(Tracks, -avgPlacedPerTrack), y=avgPlacedPerTrack)) +
  geom_bar(stat="identity", fill="steelblue", width=0.5) +
  theme_minimal() +
  xlab("Tracks") +
  ylab("Frequency") +
  ggtitle("Average Placement by Studied High School Track")
```

Average Placement by Studied High School Track



```
# Total number of students who placed per College track
# Mkt&HR
placedMktHR <- filter(campusRecruitment, campusRecruitment$`College Track` == "Mkt&HR" &
  campusRecruitment$`Placement Status` == "Placed")
placedMktHRTotal <- nrow(placedMktHR)
placedMktHRTotal
```

```
## [1] 5247
```

```
# 25% of students who studied Commerce placed
avgPlacedMktHR <- avg_num(placedMktHRTotal)
```

```
## [1] 0.2465116
```

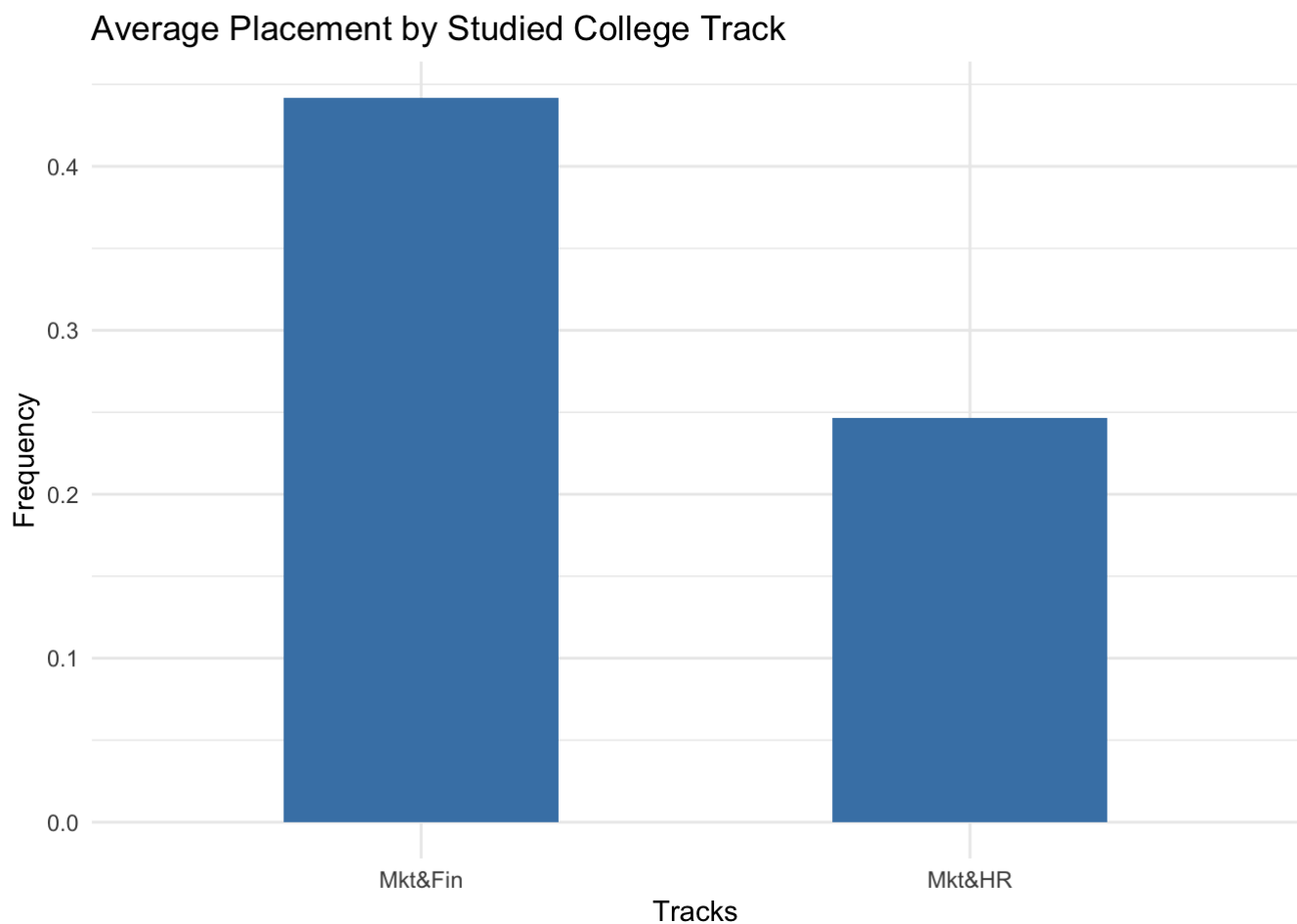
```
# Mkt&Fin
placedMktFin <- filter(campusRecruitment, campusRecruitment$`College Track` == "Mkt&Fin"
  & campusRecruitment$`Placement Status` == "Placed")
placedMktFinTotal <- nrow(placedMktFin)
placedMktFinTotal
```

```
## [1] 9405
```

```
# 44% of students who studied Commerce placed  
avgPlacedMktFin <- avg_num(placedMktFinTotal)
```

```
## [1] 0.4418605
```

```
# College Track Bar Chart  
avgPlacedPerTrack2 <- c(avgPlacedMktHR, avgPlacedMktFin)  
Tracks <- c("Mkt&HR", "Mkt&Fin")  
dfTracks2 <- data.frame(Tracks, avgPlacedPerTrack2)  
dfTracks2 <- dfTracks2[order(-dfTracks2$avgPlacedPerTrack2), ]  
  
ggplot(data=dfTracks2, aes(x=reorder(Tracks, -avgPlacedPerTrack2), y=avgPlacedPerTrack2))  
+  
  geom_bar(stat="identity", fill="steelblue", width=0.5) +  
  theme_minimal() +  
  xlab("Tracks") +  
  ylab("Frequency") +  
  ggtitle("Average Placement by Studied College Track")
```



What is the average placed status by gender?

```
# Total number of students by gender who placed (not proportioned)
```

```
# Females
```

```
placedFemales <- filter(campusRecruitment, campusRecruitment$`Gender` == "F" & campusRecruitment$`Placement Status` == "Placed")
placedFemalesTotal <- nrow(placedFemales)
placedFemalesTotal
```

```
## [1] 4752
```

```
# 22% of female students who placed out of total students
```

```
avgPlacedFemales <- avg_num(placedFemalesTotal)
```

```
## [1] 0.2232558
```

```
# Males
```

```
placedMales <- filter(campusRecruitment, campusRecruitment$`Gender` == "M" & campusRecruitment$`Placement Status` == "Placed")
placedMalesTotal <- nrow(placedMales)
placedMalesTotal
```

```
## [1] 9900
```

```
# 47% of male students who placed out of total of students
```

```
avgPlacedMales <- avg_num(placedMalesTotal)
```

```
## [1] 0.4651163
```

```
# Total number of students by gender who placed (proportioned)
```

```
# Total number of female students 7,524
```

```
totalFemaleStudents <- filter(campusRecruitment, campusRecruitment$`Gender` == "F")
totalFemales <- nrow(totalFemaleStudents)
totalFemales
```

```
## [1] 7524
```

```
# Total number of male students 13,761
```

```
totalMaleStudents <- filter(campusRecruitment, campusRecruitment$`Gender` == "M")
totalMales <- nrow(totalMaleStudents)
totalMales
```

```
## [1] 13761
```

```
# 63% of females placed out of the total female students
avgFemalesPlacedOfFemales <- placedFemalesTotal/totalFemales
avgFemalesPlacedOfFemales
```

```
## [1] 0.6315789
```

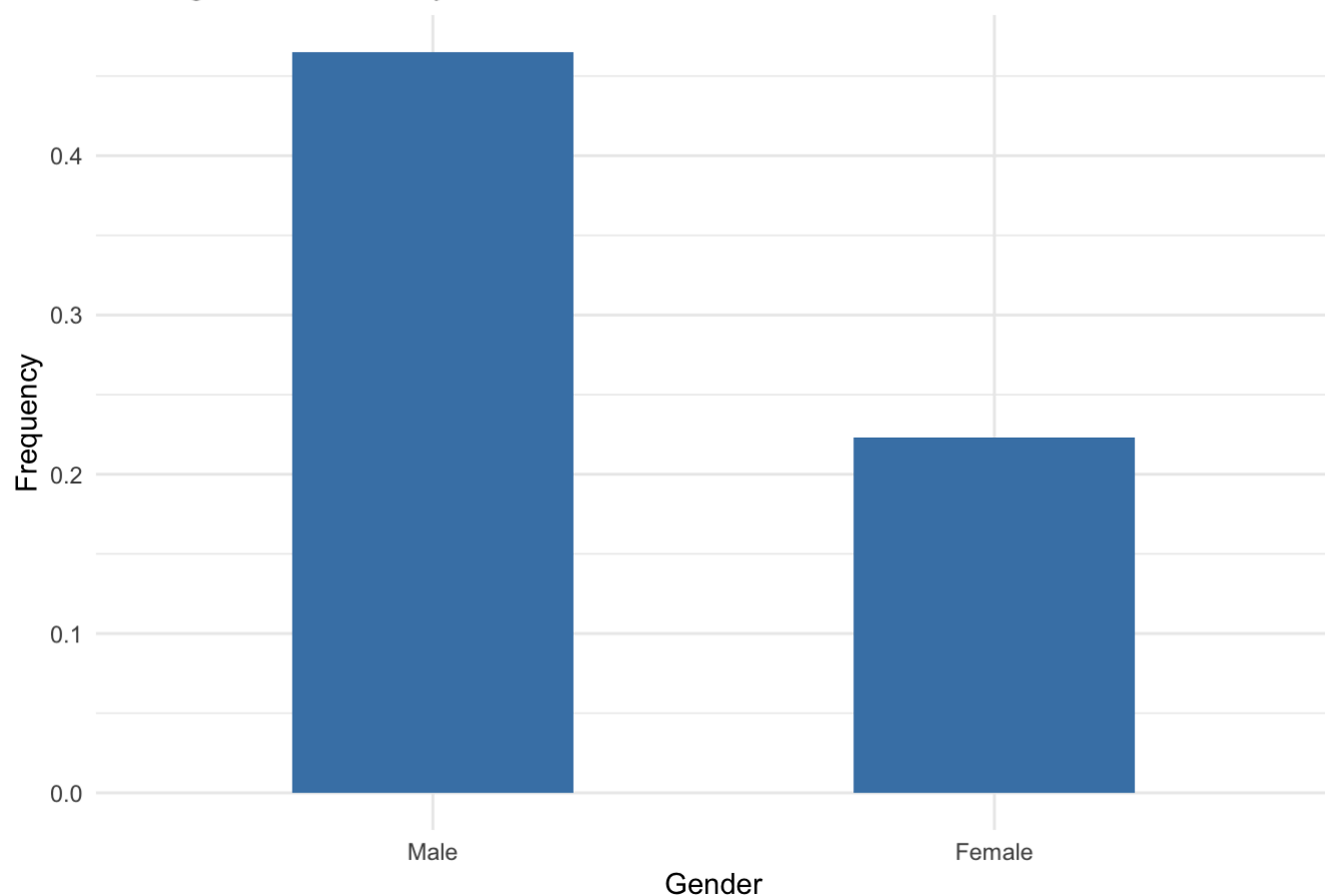
```
# 72% of males placed out of the total male students
avgMalesPlacedOfMales <- placedMalesTotal/totalMales
avgMalesPlacedOfMales
```

```
## [1] 0.7194245
```

```
# Bar Chart (Total Female vs Male)
avgPlacedPerGender <- c(avgPlacedFemales, avgPlacedMales)
Gender <- c("Female", "Male")
dfGender <- data.frame(Gender, avgPlacedPerGender)
dfGender <- dfGender[order(-dfGender$avgPlacedPerGender),]

ggplot(data=dfGender, aes(x=reorder(Gender, -avgPlacedPerGender), y=avgPlacedPerGender))
+
  geom_bar(stat="identity", fill="steelblue", width=0.5) +
  theme_minimal() +
  xlab("Gender") +
  ylab("Frequency") +
  ggtitle("Average Placement by Total Gender")
```

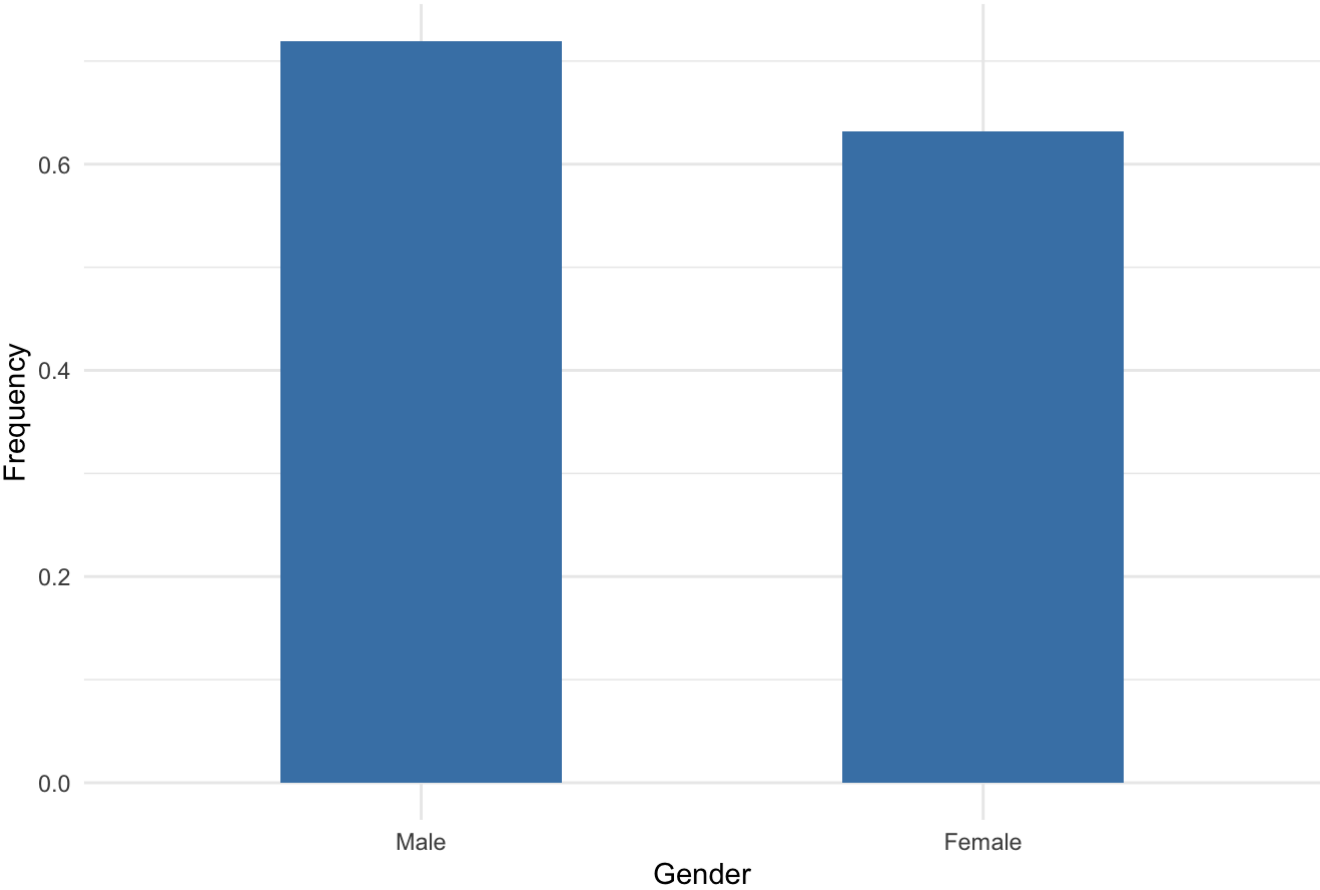
Average Placement by Total Gender



```
# Bar Chart (Avg Female vs Male)
avgPlacedPerGender2 <- c(avgFemalesPlacedOfFemales, avgMalesPlacedOfMales)
Gender <- c("Female", "Male")
dfGender2 <- data.frame(Gender, avgPlacedPerGender2)
dfGender2 <- dfGender2[order(-dfGender2$avgPlacedPerGender2), ]

ggplot(data=dfGender2, aes(x=reorder(Gender, -avgPlacedPerGender2), y=avgPlacedPerGender2)) +
  geom_bar(stat="identity", fill="steelblue", width=0.5) +
  theme_minimal() +
  xlab("Gender") +
  ylab("Frequency") +
  ggtitle("Average Placement by Total Within Each Gender")
```


Average Placement by Total Within Each Gender



When taking the average of students who placed by Gender using the total number of students, there appears to be a large margin between females and males who placed, however when you compare using the total within each gender you can see that the placement is comparable between females and males who placed. This is because there are more male students than female students so the numbers are skewed unless you take the average out of the total number of students within each gender: female (7,524) and male (13,761) versus out of the total number of students (21,285)