Stats Modeling Project

Group 5: Xin Jin, Reid Ginoza, Heidi Lovejoy

11/18/2019

```
institution_data <- read_spss('InstLevel.sav') %>%
  select(institution_name, state_cd, ClassificationCode, Classification=classification_name, sector_cd,
  filter(grepl('4-year', sector_name))  %>% # only 4-year Schools
  filter(is.na(IL_PARTIC_COED_MEN) | IL_PARTIC_COED_MEN == 0 ) %>% # only schools with no male particip
  filter(is.na(IL_PARTIC_COED_WOMEN) | IL_PARTIC_COED_WOMEN == 0 ) %>% # only schools with no female pa
  select(-contains("COED")) %>% # ignore variables with the word "coed"
  filter(!is.na(HDCOACH_SALARY_MEN)) %>% # filter out NA
  filter(!is.na(HDCOACH_SALARY_WOMEN))  # filter out NA

ncaa.div.i_data <- institution_data %>%
  filter(grepl('NCAA Division I-', Classification)) # only NCAA division 1

men_data <- institution_data %>%
  select(-contains("WOMEN")) %>%
  rename(Salary = HDCOACH_SALARY_MEN, IL_PARTIC = IL_PARTIC_MEN)

women_data <- institution_data %>%
  select(-contains("_MEN")) %>%
  rename(Salary = HDCOACH_SALARY_WOMEN, IL_PARTIC = IL_PARTIC_WOMEN)

gender.separated_data <- bind_rows(
  men_data %>% add_column(Gender = "Men"),
  women_data %>% add_column(Gender = "Women")
)
```

Introduction

While it seems to be common knownledge that some of the highest paid public employees are college sports teams coaches, less is known about the salaries of college-level sports coaches in general. This project examined predictors of head coach salaries of sports teams at four-year colleges and universities in the United States. We used the institution-level data from the Equity in Athletics Survey, Year 2017-2018, from the U.S. Department of Education Office of Postsecondary Education (2018). Our analysis focused on the differences of head coach salaries between male and female teams, of classification, and of the number of student participants at an institution level.

The full data set listed completed surveys for 2079 institutions. The 578 two-year colleges were excluded, as well as the 120 four-year schools with coed teams and 18 other institutions with missing values for head coach salaries. Our working data set had 1363 four-year institutions without coed sports teams.

Hypotheses

We examined whether the classifications and the gender of the team had effects on the head coach salaries. To do that, we first investigated the variances of head coach salaries of male teams compared with that of female teams. We checked for interaction effects between classification and gender of team. We also looked at whether the number of participants institution-wide predicted the salary for the head coaches.

Methods

Data was analyzed using R version 3.6.1 (2019-07-05) (R Core Team 2019). Data is described using counts (percentage) for categorical variables and mean (standard deviation) for continuous variables.

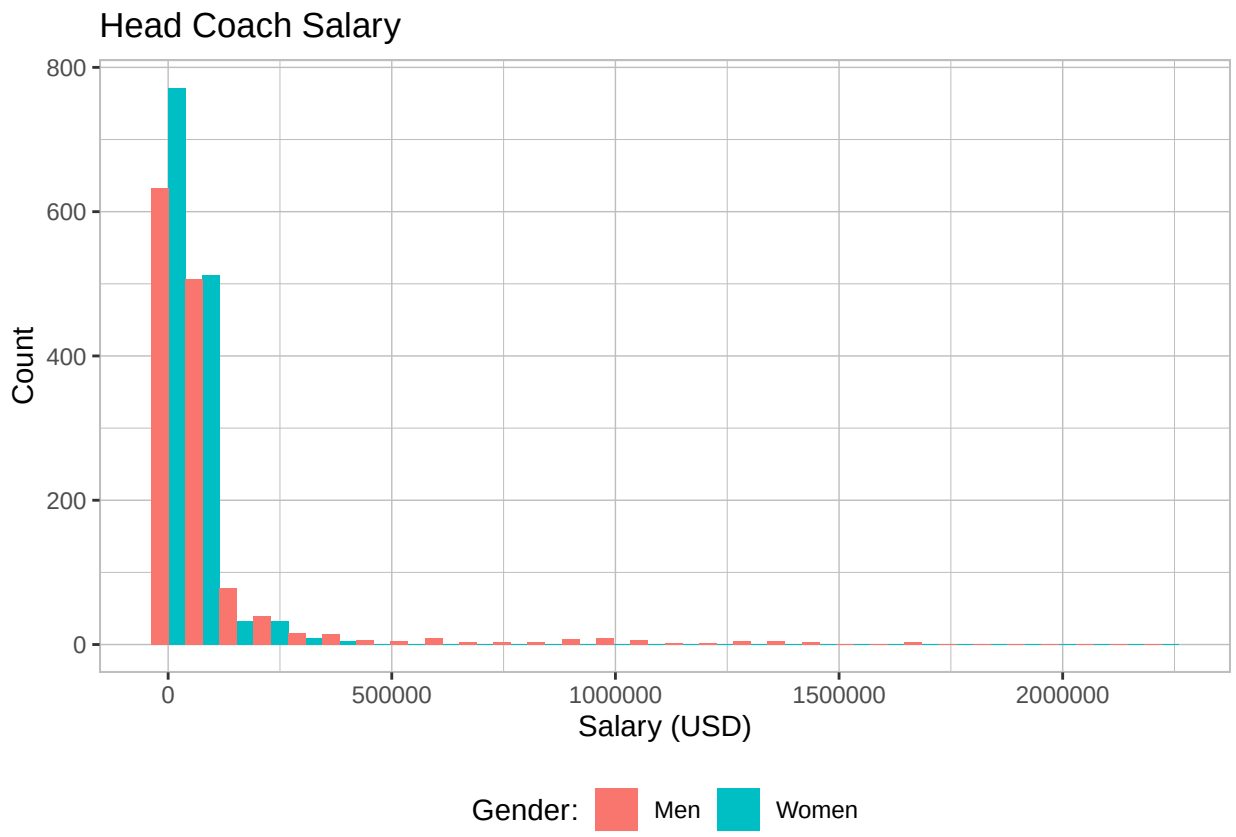
Results

Description of Data

Variable	Mean	SD
Head Coach–Men's Salary	107,801.45	230,718.91
Head Coach–Women's Salary	48,110.37	51,009.74
Head Coach–All Salary	77,955.91	169,698.1
Participation–Men	218.26	123.14
Participation–Women	169.02	94.32
Participation–Total	387.28	207.9

```
institution_data %>% group_by(Classification) %>% summarize(Count = n()) %>%
  mutate(Percentage = percent(Count / 1363)) %>% kable(booktabs = T)
```

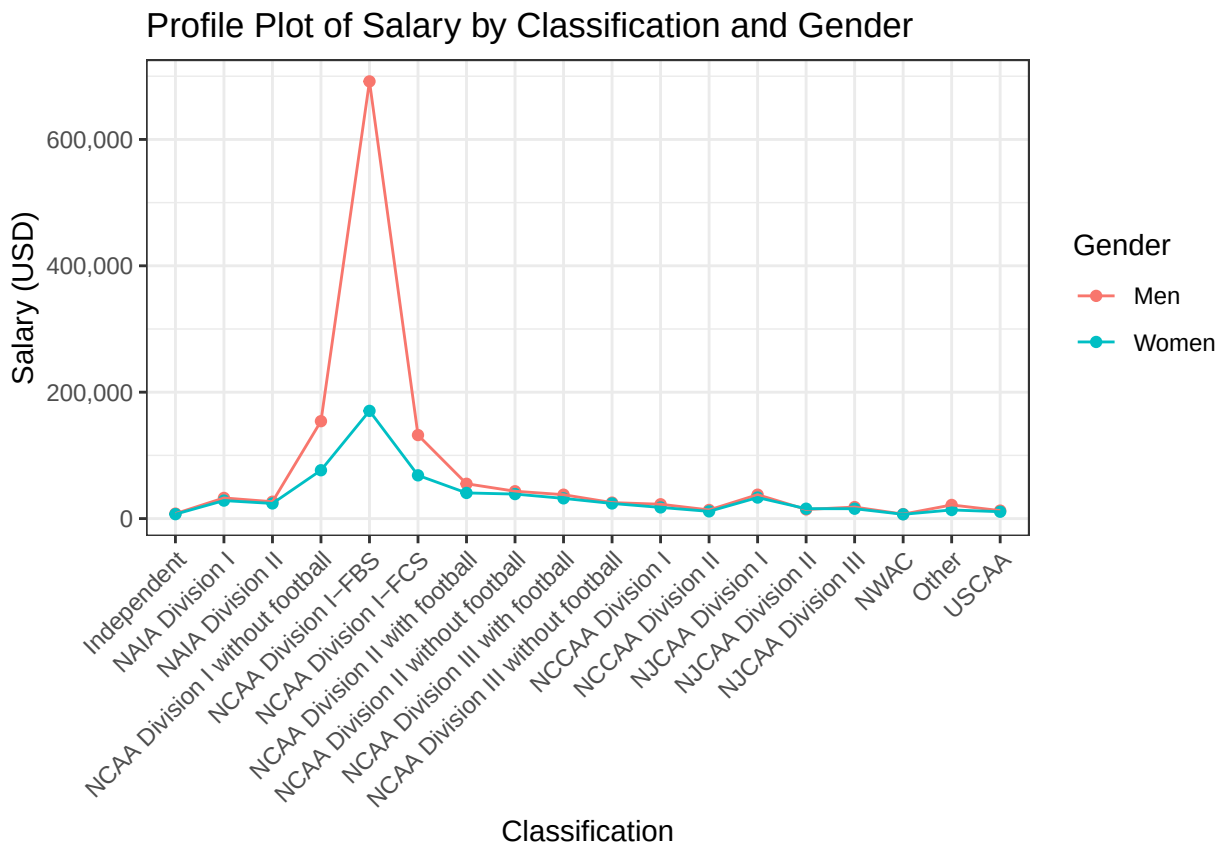
Classification	Count	Percentage
Independent	4	0.3%
NAIA Division I	90	6.6%
NAIA Division II	102	7.5%
NCAA Division I without football	95	7.0%
NCAA Division I-FBS	117	8.6%
NCAA Division I-FCS	114	8.4%
NCAA Division II with football	160	11.7%
NCAA Division II without football	138	10.1%
NCAA Division III with football	223	16.4%
NCAA Division III without football	162	11.9%
NCCAA Division I	9	0.7%
NCCAA Division II	25	1.8%
NJCAA Division I	35	2.6%
NJCAA Division II	9	0.7%
NJCAA Division III	4	0.3%
NWAC	9	0.7%
Other	33	2.4%
USCAA	34	2.5%



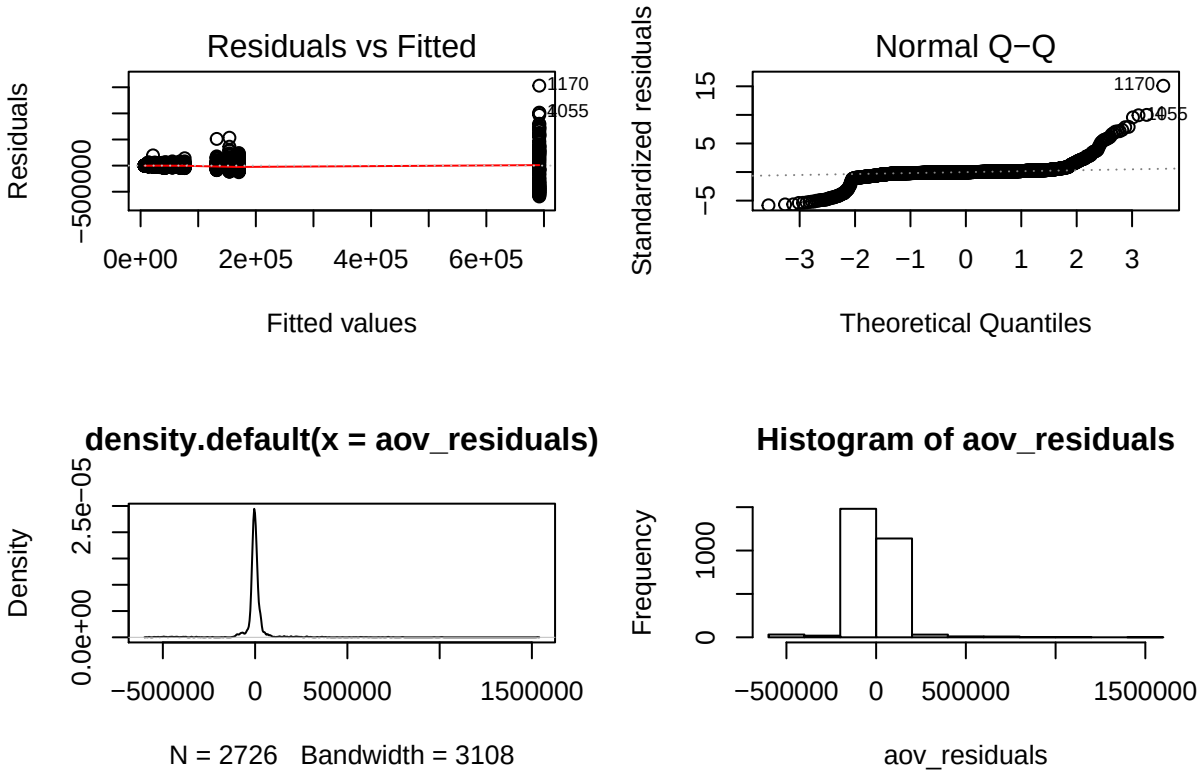
Analysis Results

```
classification.gender.interaction.results <- gender.separated_data %$% aov(Salary ~ Classification * Ge
```

```
gender.separated_data %>% select(Salary, Classification, Gender) %>%
  ggplot() + aes(x = Classification, color = Gender, group = Gender, y = Salary) +
  stat_summary(fun.y = mean, geom = "point") + stat_summary(fun.y = mean, geom = "line") +
  labs(
    title = "Profile Plot of Salary by Classification and Gender",
    x = "Classification",
    y = "Salary (USD)"
    ) +
  theme_bw() + theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1)) + scale_y_continuous(la
```




```
classification.gender.interaction.results %>% almost_sas()
```



Conclusion

Our data set contained only the averages of head coach salaries of a given institution, so we were unable to capture any variance within an institution. Additionally, this study only used the surveys from one year and future studies should investigate the salaries over multiple years.

References

Office of Postsecondary Education. 2018. "Equity in Athletics Data Analysis." U.S. Department of Education. https://ope.ed.gov/athletics/#/datafile/list.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.