Stats Modeling Project

Group 5: Xin Jin, Reid Ginoza, Heidi Lovejoy

11/18/2019

```
institution_data <- read_spss('InstLevel.sav') %>%
  select(  # TODO will change later when we decide what variables we want
  -unitid, -addr1_txt, -addr2_txt, -city_txt, -zip_text, -sector_cd,
  -ClassificationCode, -ClassificationOther
  ) %>%
  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"
```

Introduction

%TODO Briefly give meaning to your project – what is the background and why is this project important? (20 points)

Hypotheses

%TODO Clearly state the hypotheses being tested. (20 points)

Methods

%TODO State and explain the methodology used to test the hypotheses. (20 points)

Results

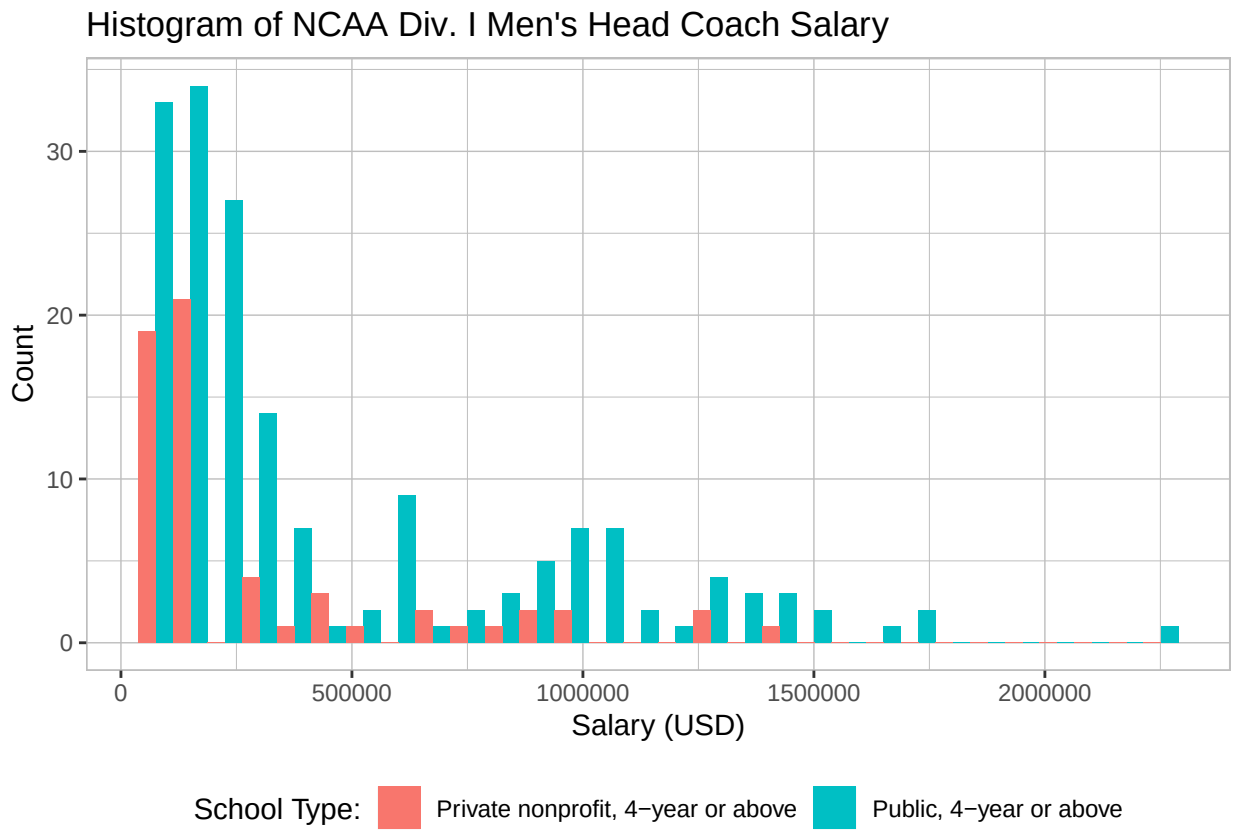
%TODO Describe the data (1st paragraph)

Visualizing Head Coach data

```
ncaa.div.i_hd.coach <- institution_data %>%
  filter(grepl('NCAA Division I-', classification_name)) %>%
  select(sector_name, HDCOACH_SALARY_MEN)

ncaa.div.i_hd.coach %>%
  ggplot(aes(x = HDCOACH_SALARY_MEN, fill = sector_name)) +
  geom_histogram(position = "dodge") +
  labs(title="Histogram of NCAA Div. I Men's Head Coach Salary", x = "Salary (USD)",
       y = "Count", fill = "School Type:") +
```

```
  theme(
    legend.position="bottom",
    panel.background = element_rect(fill = 'white', color = 'grey'),
    panel.grid.major = element_line(size = 0.25, linetype = 'solid',
                                    color = "grey"),
    panel.grid.minor = element_line(size = 0.1, linetype = 'solid',
                                    color = "grey")
  )
```



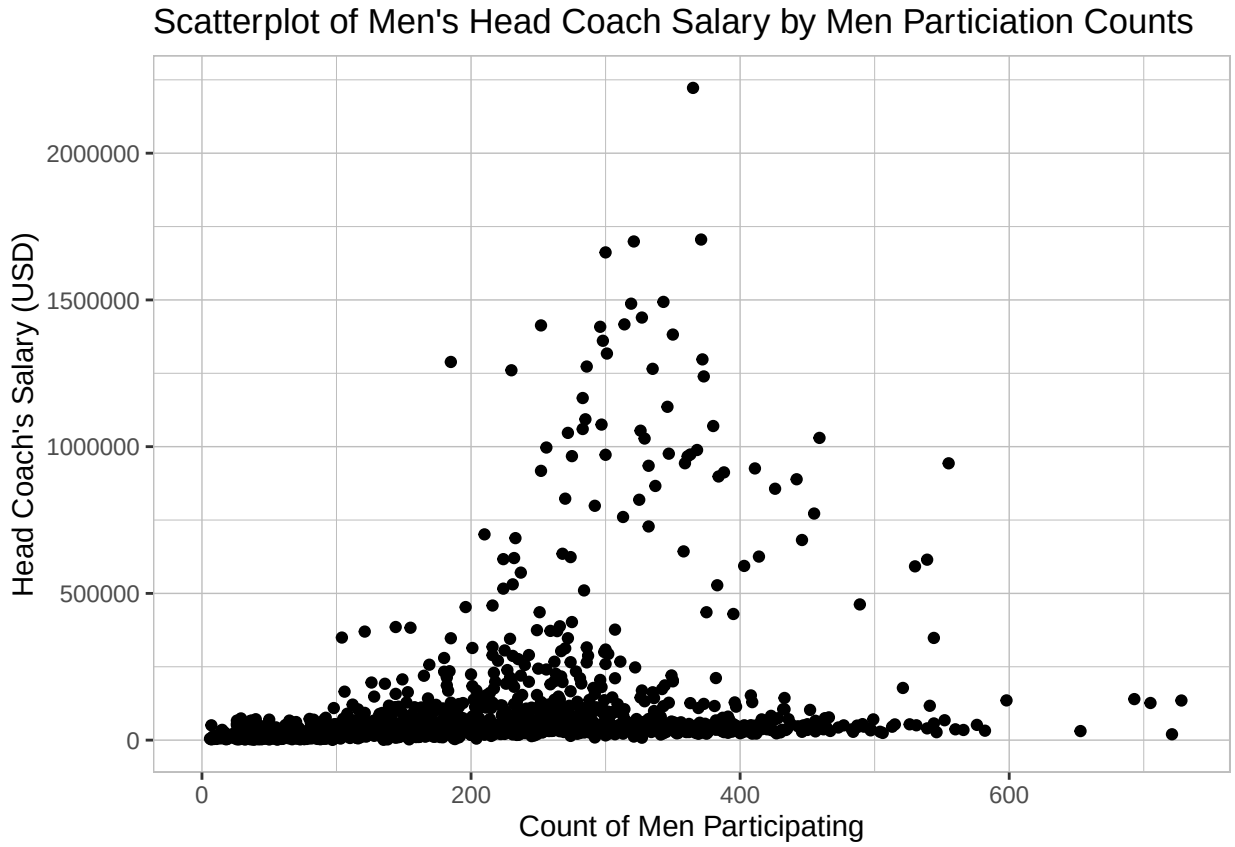
Visualizing Number of Athletes and Head Coach Salaries

```
participation.data <- institution_data %>%
  select(IL_PARTIC_MEN, HDCOACH_SALARY_MEN)

participation.data %>% ggplot(aes(x=IL_PARTIC_MEN, y=HDCOACH_SALARY_MEN)) +
  geom_point() +
  labs(title="Scatterplot of Men's Head Coach Salary by Men Particiation Counts", x = "Count of Men Par
       y = "Head Coach's Salary (USD)") +
  theme(
    panel.background = element_rect(fill = 'white', color = 'grey'),
    panel.grid.major = element_line(size = 0.25, linetype = 'solid',
                                    color = "grey"),
    panel.grid.minor = element_line(size = 0.1, linetype = 'solid',
```

```
                                color = "grey")
 )
```

```
## Warning: Removed 11 rows containing missing values (geom_point).
```



%TODO And analysis results (order of results should correspond to hypotheses listed). (20 points)

Conclusion

%TODO Restate the results in terms of the larger picture; state limitations and opportunities forfuture research (20 points)

Session Info

```
sessionInfo()
```

```
## R version 3.6.1 (2019-07-05)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS High Sierra 10.13.6
##
## Matrix products: default
```

```
## BLAS:   /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] haven_2.1.1      magrittr_1.5     kableExtra_1.1.0 knitr_1.24
##  [5] forcats_0.4.0    stringr_1.4.0    dplyr_0.8.3      purrr_0.3.2
##  [9] readr_1.3.1      tidyr_0.8.3      tibble_2.1.3     ggplot2_3.2.1
## [13] tidyverse_1.2.1
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.2        cellranger_1.1.0  pillar_1.4.2
##  [4] compiler_3.6.1    tools_3.6.1       zeallot_0.1.0
##  [7] digest_0.6.20     viridisLite_0.3.0 lubridate_1.7.4
## [10] jsonlite_1.6      evaluate_0.14     nlme_3.1-140
## [13] gtable_0.3.0      lattice_0.20-38   pkgconfig_2.0.2
## [16] rlang_0.4.0       cli_1.1.0         rstudioapi_0.10
## [19] yaml_2.2.0        xfun_0.9          withr_2.1.2
## [22] xml2_1.2.2        httr_1.4.1        vctrs_0.2.0
## [25] generics_0.0.2    hms_0.5.1         webshot_0.5.1
## [28] grid_3.6.1        tidyselect_0.2.5  glue_1.3.1
## [31] R6_2.4.0          readxl_1.3.1      rmarkdown_1.15
## [34] modelr_0.1.5      backports_1.1.4   scales_1.0.0
## [37] htmltools_0.4.0   rvest_0.3.4       assertthat_0.2.1
## [40] colorspace_1.4-1  labeling_0.3      stringi_1.4.3
## [43] lazyeval_0.2.2    munsell_0.5.0     broom_0.5.2
## [46] crayon_1.3.4
```

References

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