Stats Modeling Project

Group 5: Xin Jin, Reid Ginoza, Heidi Lovejoy

11/18/2019

```
almost_sas <- function(aov.results){
  par(mfrow=c(2,2))
  plot(aov.results, which=1)
  plot(aov.results, which=2)
  aov_residuals <- residuals(aov.results)
  plot(density(aov_residuals))
  hist(aov_residuals)
}
```

```
{r import data} # 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, #   -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 participants in coed #
filter(is.na(IL_PARTIC_COED_WOMEN) | IL_PARTIC_COED_WOMEN ==
0 ) %>% # only schools with no female participants in coed #
select(-contains("COED")) %>% # ignore variables with the word
"coed" #   mutate(Grand.Total.Profits = GRND_TOTAL_REVENUE -
GRND_TOTAL_EXPENSE) %>% #   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_name)) # only
NCAA division 1 #
```

```
institution_data <- read_spss('InstLevel.sav') %>%
  select(institution_name, state_cd, ClassificationCode, classification_name, sector_cd, sector_name, HI
  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_name)) # only NCAA division 1
```

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

women_data <- institution_data %>%
  select(-contains("_MEN")) %>%
  rename(HD_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")
)
```

Counts of Divisions

```
institution_data %>%
  group_by(classification_name) %>%
  summarize('Mean Men Head Coach Salary'=mean(HDCOACH_SALARY_MEN), 'SD Men'=sd(HDCOACH_SALARY_MEN),
            'Mean Woen Head Coach Salary'=mean(HDCOACH_SALARY_WOMEN), 'SD Women'=sd(HDCOACH_SALARY_WOMEN
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover"),
                full_width = F,
                font_size = 14,
                position = "left")
```

classification_name	Mean Men Head Coach Salary	SD Men	Mean
Independent	7843.750	3368.628	
NAIA Division I	32668.956	11988.474	
NAIA Division II	26677.559	11896.811	
NCAA Division I without football	154140.947	113510.725	
NCAA Division I-FBS	691848.778	460008.876	
NCAA Division I-FCS	132124.781	73211.986	
NCAA Division II with football	55094.100	19053.739	
NCAA Division II without football	43269.167	21391.554	
NCAA Division III with football	37806.229	13236.055	
NCAA Division III without football	25276.660	15453.066	
NCCAA Division I	22683.000	14696.931	
NCCAA Division II	13652.840	11260.258	
NJCAA Division I	37834.543	22880.751	
NJCAA Division II	14425.000	8128.600	
NJCAA Division III	18207.750	13436.158	
NWAC	7047.556	1379.605	
Other	21572.939	40636.305	
USCAA	12626.029	14537.566	

```
institution_data %>%
  select(ClassificationCode, classification_name) %>%
  unique()
```

```
## # A tibble: 18 x 2
##    ClassificationCode classification_name
##                 <dbl> <chr>
##  1                  2 NCAA Division I-FCS
##  2                  5 NCAA Division II without football
##  3                  1 NCAA Division I-FBS
##  4                  6 NCAA Division III with football
##  5                  9 NAIA Division I
##  6                  4 NCAA Division II with football
##  7                 20 USCAA
##  8                  3 NCAA Division I without football
##  9                  8 Other
## 10                 12 NJCAA Division I
## 11                 15 NCCAA Division I
## 12                  7 NCAA Division III without football
## 13                 10 NAIA Division II
## 14                 16 NCCAA Division II
## 15                 13 NJCAA Division II
## 16                 14 NJCAA Division III
## 17                 18 Independent
## 18                 19 NWAC
```

```
institution_data %>%
  filter(ClassificationCode==9) %>%
  select(classification_name,HDCOACH_SALARY_WOMEN) %>%
  # summarize(mean=mean(HDCOACH_SALARY_MEN)) # results in NA
  summary()
```

```
##  classification_name HDCOACH_SALARY_WOMEN
##  Length:90           Min.   : 4077
##  Class :character    1st Qu.:22820
##  Mode  :character    Median :27126
##                      Mean   :28453
##                      3rd Qu.:36206
##                      Max.   :58666
```

```
two.way.anova.attempt <- institution_data %>%
  select(classification_name, HDCOACH_SALARY_MEN, HDCOACH_SALARY_WOMEN) %>%
  gather("Gender", "Salary", -classification_name)
```

```
## Warning: attributes are not identical across measure variables;
## they will be dropped
```

```
two.way.anova.attempt$Gender %<>% recode(HDCOACH_SALARY_MEN="Men", HDCOACH_SALARY_WOMEN="Women")
two.way.anova.attempt
```

```
## # A tibble: 2,726 x 3
##    classification_name               Gender  Salary
##    <chr>                             <chr>    <dbl>
##  1 NCAA Division I-FCS               Men      56920
##  2 NCAA Division II without football Men      73267
##  3 NCAA Division I-FCS               Men      84018
##  4 NCAA Division I-FBS               Men    1705741
##  5 NCAA Division II without football Men      37051
##  6 NCAA Division I-FBS               Men    1360838
##  7 NCAA Division III with football   Men      44759
##  8 NAIA Division I                   Men      59471
##  9 NCAA Division III with football   Men      64314
## 10 NCAA Division II with football    Men      56806
## # ... with 2,716 more rows
```

```
adv_aov.results <- aov(Salary ~ classification_name*Gender, data = two.way.anova.attempt)
adv_aov.table <- adv_aov.results %>% summary()
adv_aov.table %>% pander(style = 'rmarkdown')
```

Table 1: Analysis of Variance Model (continued below)

	Df	Sum Sq	Mean Sq	F value
classification_name	17	3.413e+13	2.008e+12	193.6
Gender	1	2.428e+12	2.428e+12	234.1
classification_name:Gender	17	1.402e+13	8.245e+11	79.5
Residuals	2690	2.79e+13	1.037e+10	NA

	Pr(>F)
classification_name	0
Gender	9.708e-51
classification_name:Gender	9.674e-223
Residuals	NA

Introduction

This project examines NCAA Division I athletics. We used the data from the Equity in Athletics Survey, Year 2017-2018, from the U.S. Department of Education Office of Postsecondary Education (2018).

Hypotheses

Divisions and Profits

Head Coaches of Men's Teams and School Type

H_0:
H_1:

Head Coaches of Men's Teams and Participation

H_0:
H_1:

Difference in Classifications

There are 18 different classifications, including NCAA Division I-FCS, NCAA II without Football, and so forth. This question looks at the means of the profits for each classification. H_0: $\mu_1 = \mu_2 = \cdots = \mu_{18}$
H_1: At least one is different

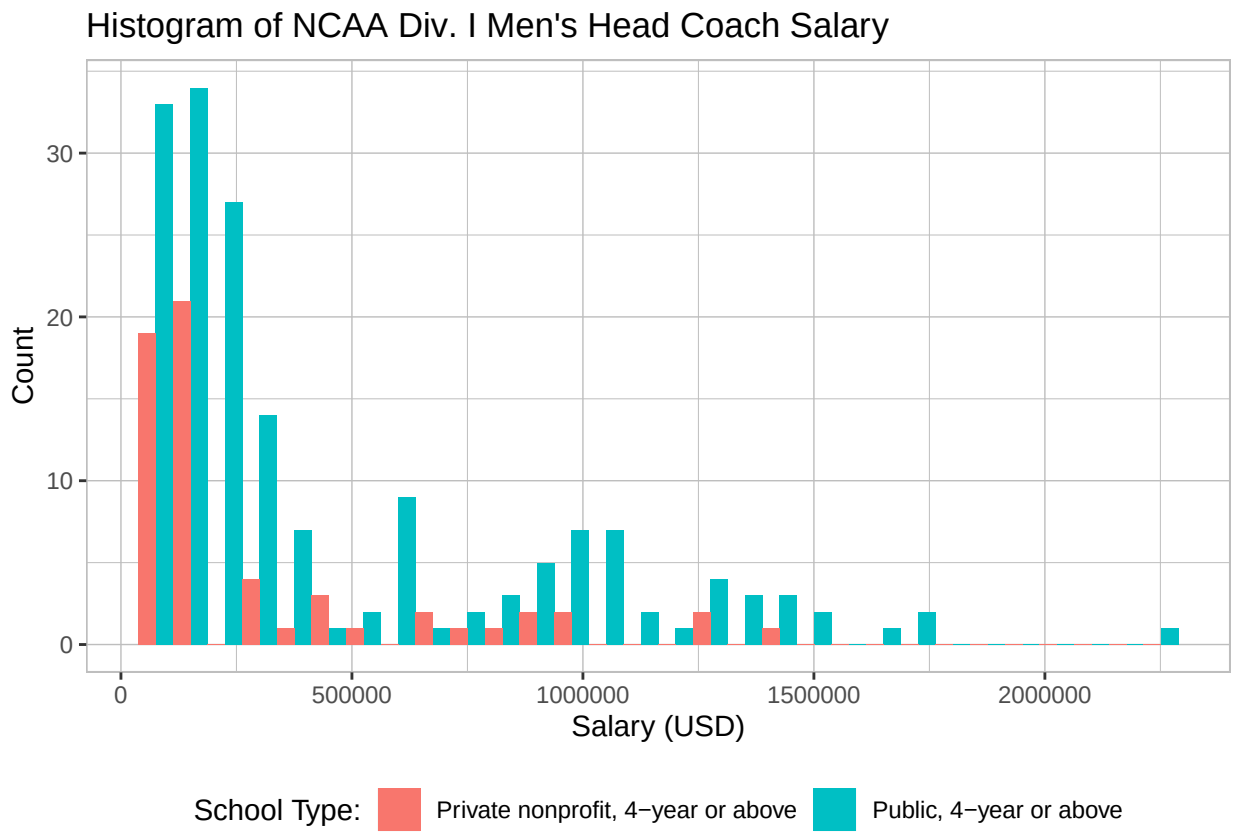
Difference in Classfications (Nonparametric)

Methods

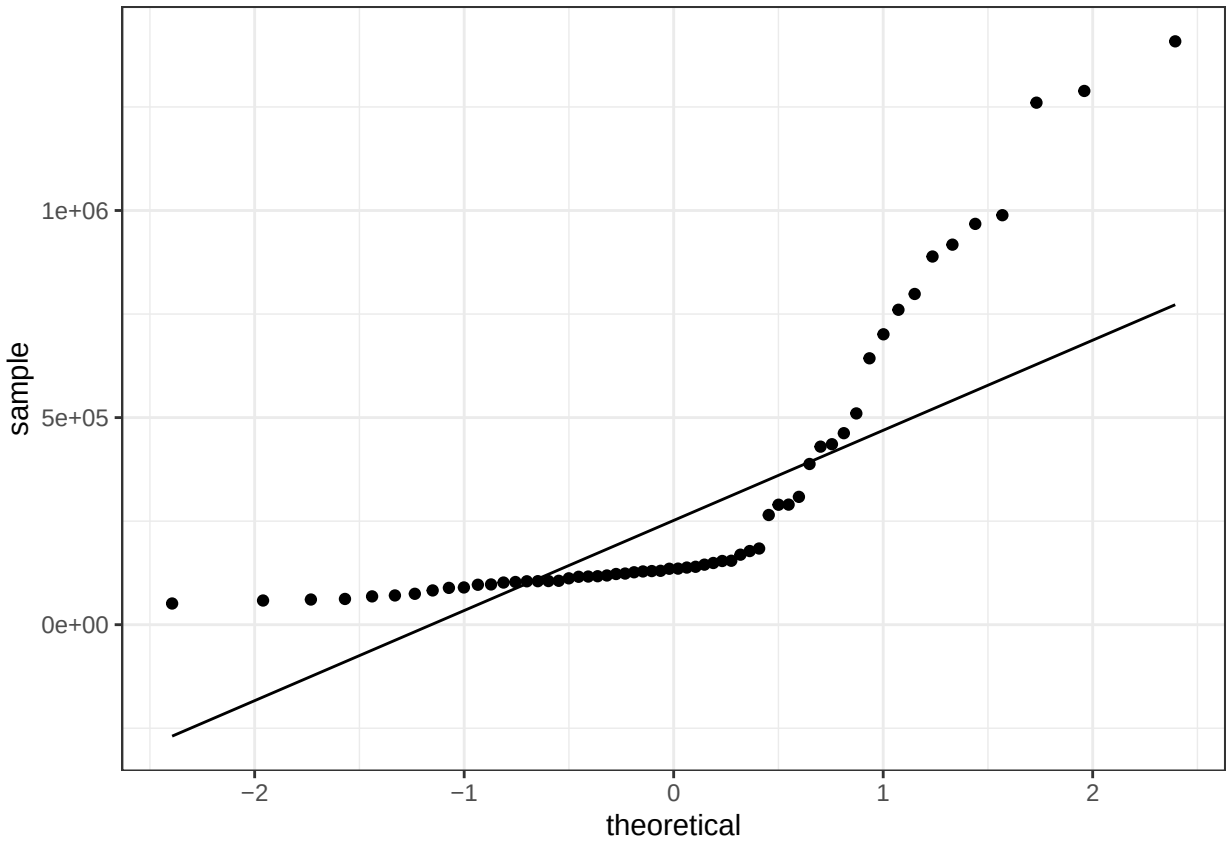
Results

Describing and Visualizing Head Coach data

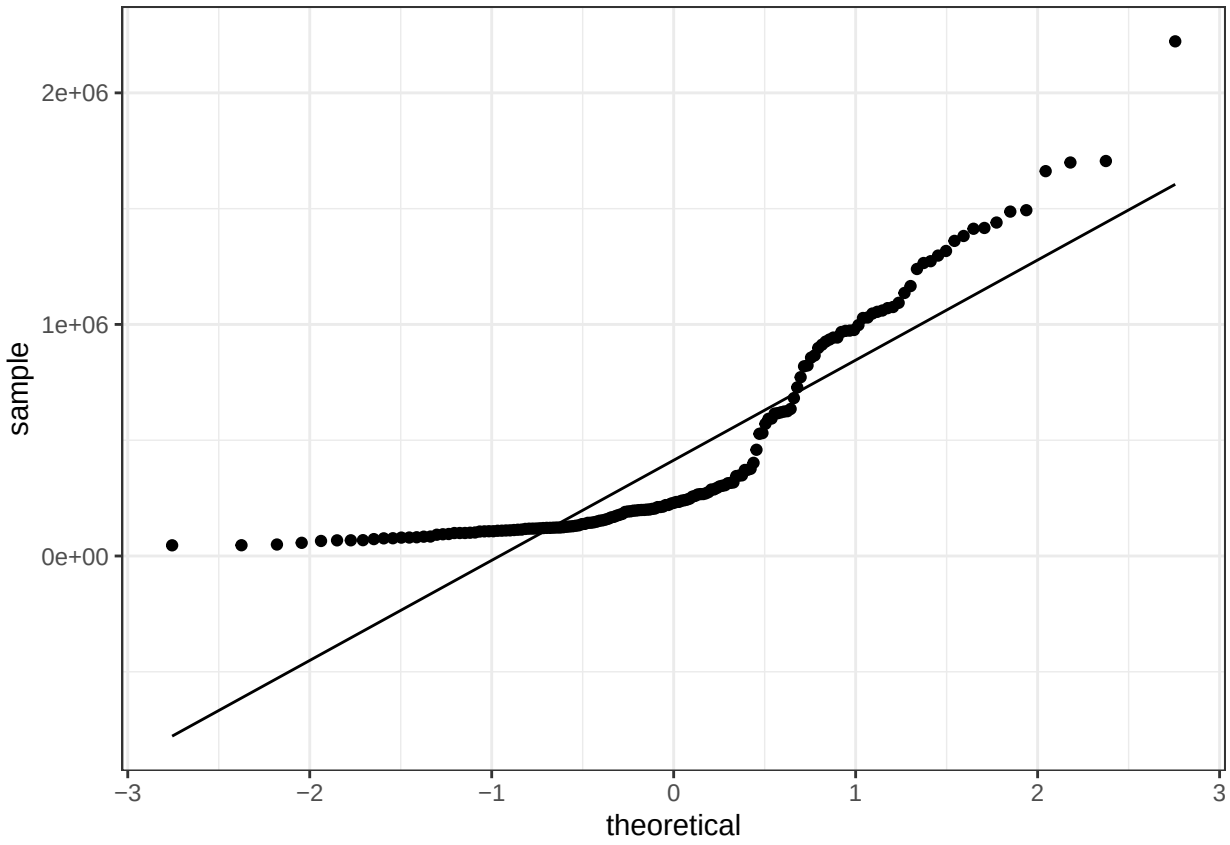
```
hd.coach.salary_data <- ncaa.div.i_data %>%
  select(sector_name, HDCOACH_SALARY_MEN)

hd.coach.salary_data %>%
  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")
  )
```




```
hd.coach.salary_data %>%
  filter(grepl('Private', sector_name)) %>%
  ggplot(aes(sample = HDCOACH_SALARY_MEN)) +
  stat_qq() + stat_qq_line() + theme_bw()
```

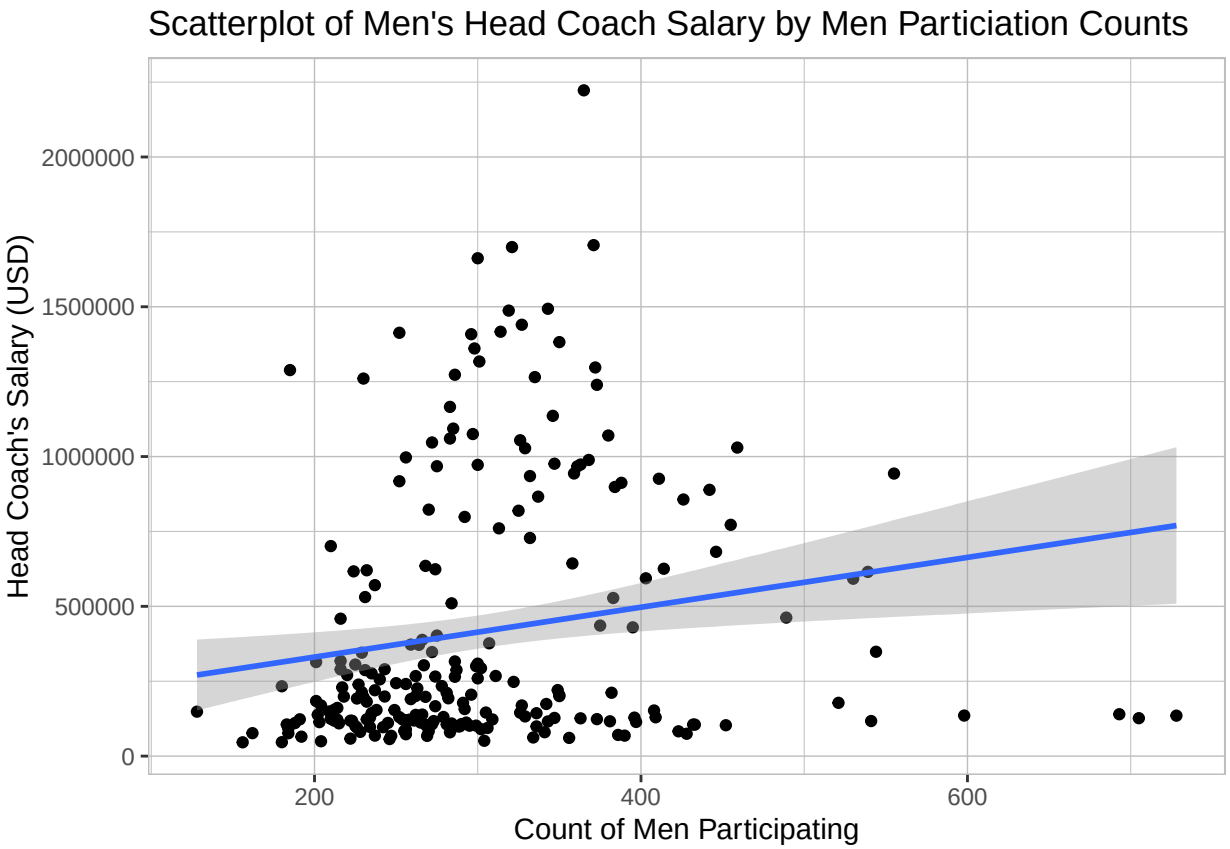



```
hd.coach.salary_data %>%
  filter(grepl('Public', sector_name)) %>%
  ggplot(aes(sample = HDCOACH_SALARY_MEN)) +
  stat_qq() + stat_qq_line() + theme_bw()
```



Describing and Visualizing Participation of Men and Head Coach Salaries

```
men.participation.data <- ncaa.div.i_data %>%
  select(IL_PARTIC_MEN, HDCOACH_SALARY_MEN)

men.participation.data %>% ggplot(aes(x=IL_PARTIC_MEN, y=HDCOACH_SALARY_MEN)) +
  geom_point() +
  geom_smooth(method = lm) +
  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")
  )
```



Results of Head Coach data

Results of Participtionof Men and Head Coach Salaries

Conclusion

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  pander_0.6.3     pgirmess_1.6.9
##
## loaded via a namespace (and not attached):
##  [1] httr_1.4.1         viridisLite_0.3.0  jsonlite_1.6
##  [4] splines_3.6.1      modelr_0.1.5       gtools_3.8.1
##  [7] assertthat_0.2.1   expm_0.999-4       sp_1.3-1
## [10] cellranger_1.1.0   yaml_2.2.0         LearnBayes_2.15.1
## [13] pillar_1.4.2       backports_1.1.4    lattice_0.20-38
## [16] glue_1.3.1         digest_0.6.20      rvest_0.3.4
## [19] colorspace_1.4-1   htmltools_0.4.0    Matrix_1.2-17
## [22] pkgconfig_2.0.2    broom_0.5.2        gmodels_2.18.1
## [25] webshot_0.5.1      scales_1.0.0       gdata_2.18.0
## [28] generics_0.0.2     withr_2.1.2        lazyeval_0.2.2
## [31] cli_1.1.0          crayon_1.3.4       readxl_1.3.1
## [34] deldir_0.1-23      maptools_0.9-8     evaluate_0.14
## [37] fansi_0.4.0        nlme_3.1-140       MASS_7.3-51.4
## [40] xml2_1.2.2         foreign_0.8-71     class_7.3-15
## [43] tools_3.6.1        hms_0.5.1          munsell_0.5.0
## [46] compiler_3.6.1     e1071_1.7-2        rlang_0.4.0
## [49] classInt_0.4-2     units_0.6-5        grid_3.6.1
## [52] rstudioapi_0.10    labeling_0.3       rmarkdown_1.15
## [55] boot_1.3-22        gtable_0.3.0       DBI_1.0.0
## [58] R6_2.4.0           splancs_2.01-40    lubridate_1.7.4
## [61] rgdal_1.4-6        utf8_1.1.4         rgeos_0.5-2
## [64] zeallot_0.1.0      spdep_1.1-3        KernSmooth_2.23-15
## [67] stringi_1.4.3      Rcpp_1.0.2         vctrs_0.2.0
## [70] sf_0.8-0           spData_0.3.2       tidyselect_0.2.5
## [73] xfun_0.9           coda_0.19-3
```

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/.