Stats Modeling Project

Group 5: Xin Jin, Reid Ginoza, Heidi Lovejoy

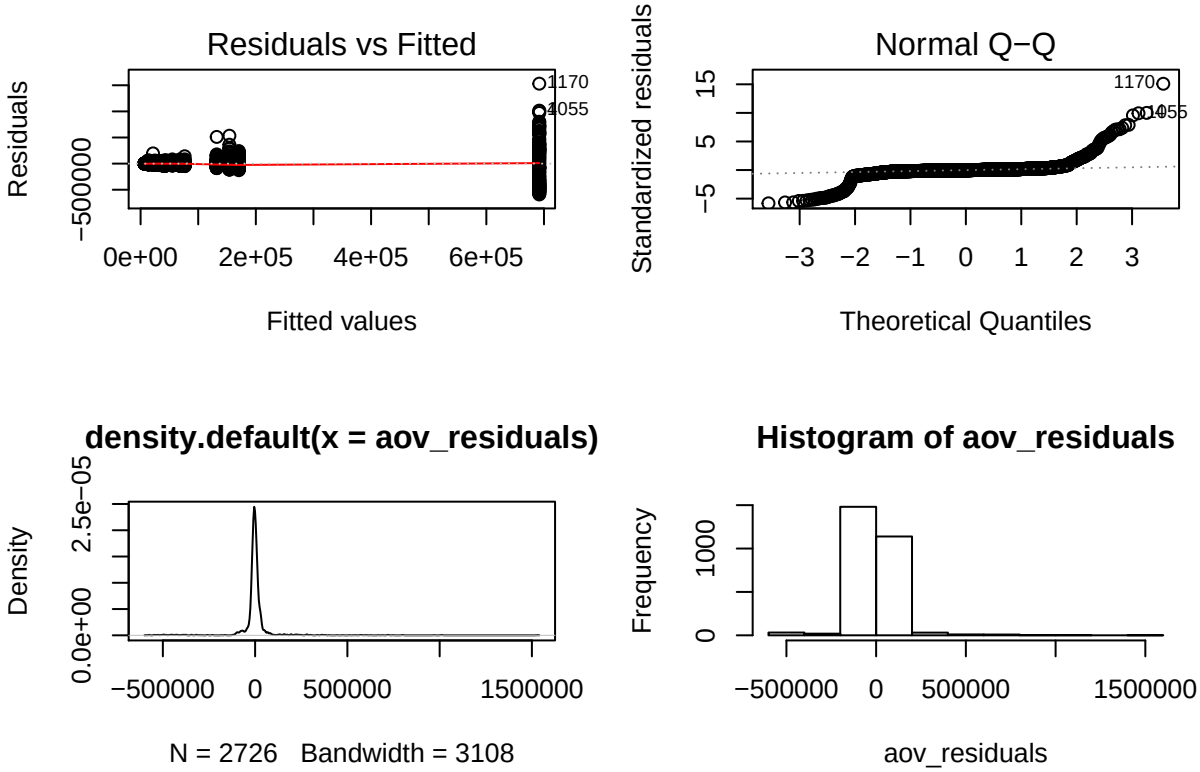
11/18/2019

```
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)
}
```

Almost SAS

```
classification.gender.interaction.results <- gender.separated_data %$%
  aov(Salary ~ Classification * Gender)
interaction.summary <- classification.gender.interaction.results %>% summary()
```

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



Notes for Kruskal-Wallis posthoc comparison

```
classification.kw.posthoc <- gender.separated_data %$% kruskalmc(Salary ~ Classification)
classification.kw.posthoc[[3]] %>% as_tibble(rownames = "Comparison") -> classification.pairwise.compari
```

Counts of Divisions

```
institution_data %>%
  group_by(Classification) %>%
  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	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) %>%
  unique()
```

```
## # A tibble: 18 x 2
```

```
##    ClassificationCode Classification
##                 <int> <fct>
##  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,HDCOACH_SALARY_WOMEN) %>%
  # summarize(mean=mean(HDCOACH_SALARY_MEN)) # results in NA
  summary()
```

```
##                           Classification HDCOACH_SALARY_WOMEN
##  NAIA Division I                  :90     Min.   : 4077
##  Independent                      : 0     1st Qu.:22820
##  NAIA Division II                 : 0     Median :27126
##  NCAA Division I without football: 0     Mean   :28453
##  NCAA Division I-FBS              : 0     3rd Qu.:36206
##  NCAA Division I-FCS              : 0     Max.   :58666
##  (Other)                          : 0
```

```
two.way.anova.attempt <- institution_data %>%
  select(Classification, HDCOACH_SALARY_MEN, HDCOACH_SALARY_WOMEN) %>%
  gather("Gender", "Salary", -Classification)
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                    Gender  Salary
##    <fct>                             <chr>    <int>
##  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*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)

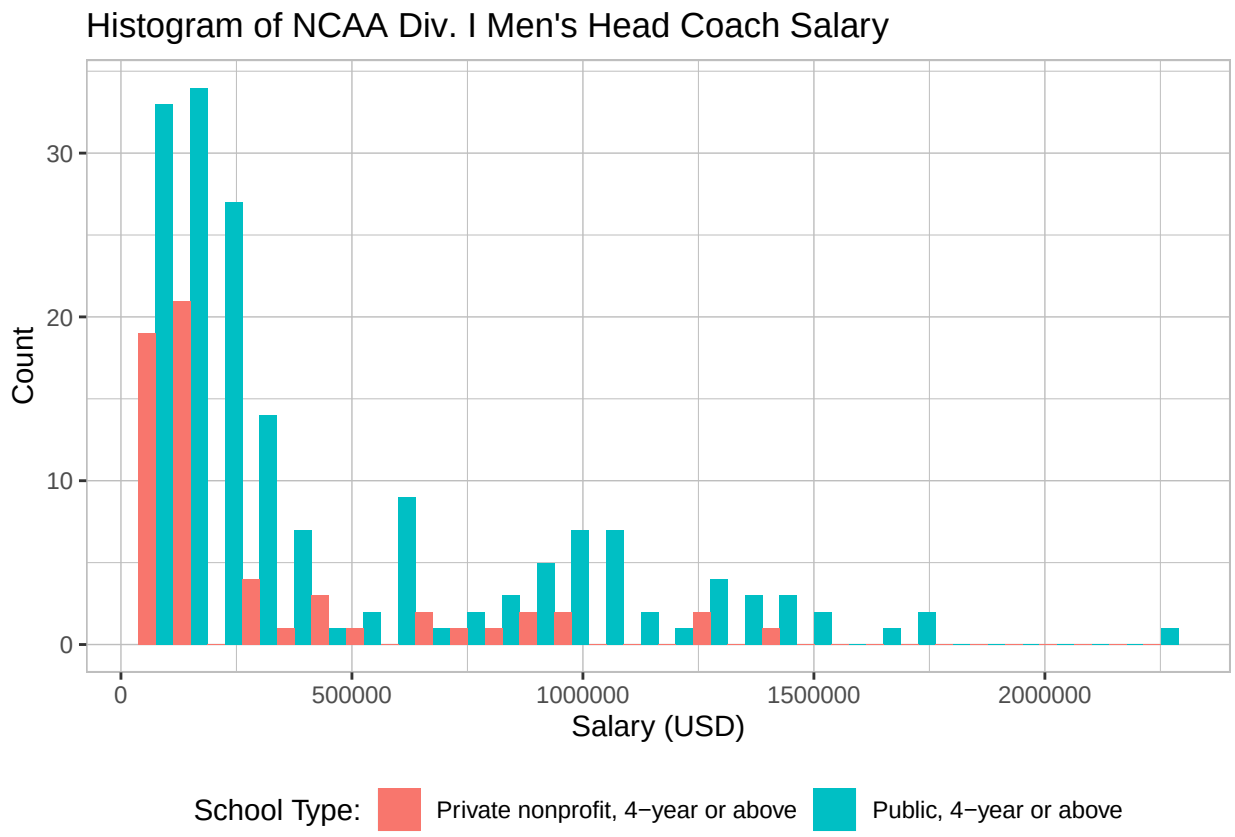
	Df	Sum Sq	Mean Sq	F value
Classification	17	3.413e+13	2.008e+12	193.6
Gender	1	2.428e+12	2.428e+12	234.1
Classification:Gender	17	1.402e+13	8.245e+11	79.5
Residuals	2690	2.79e+13	1.037e+10	NA

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

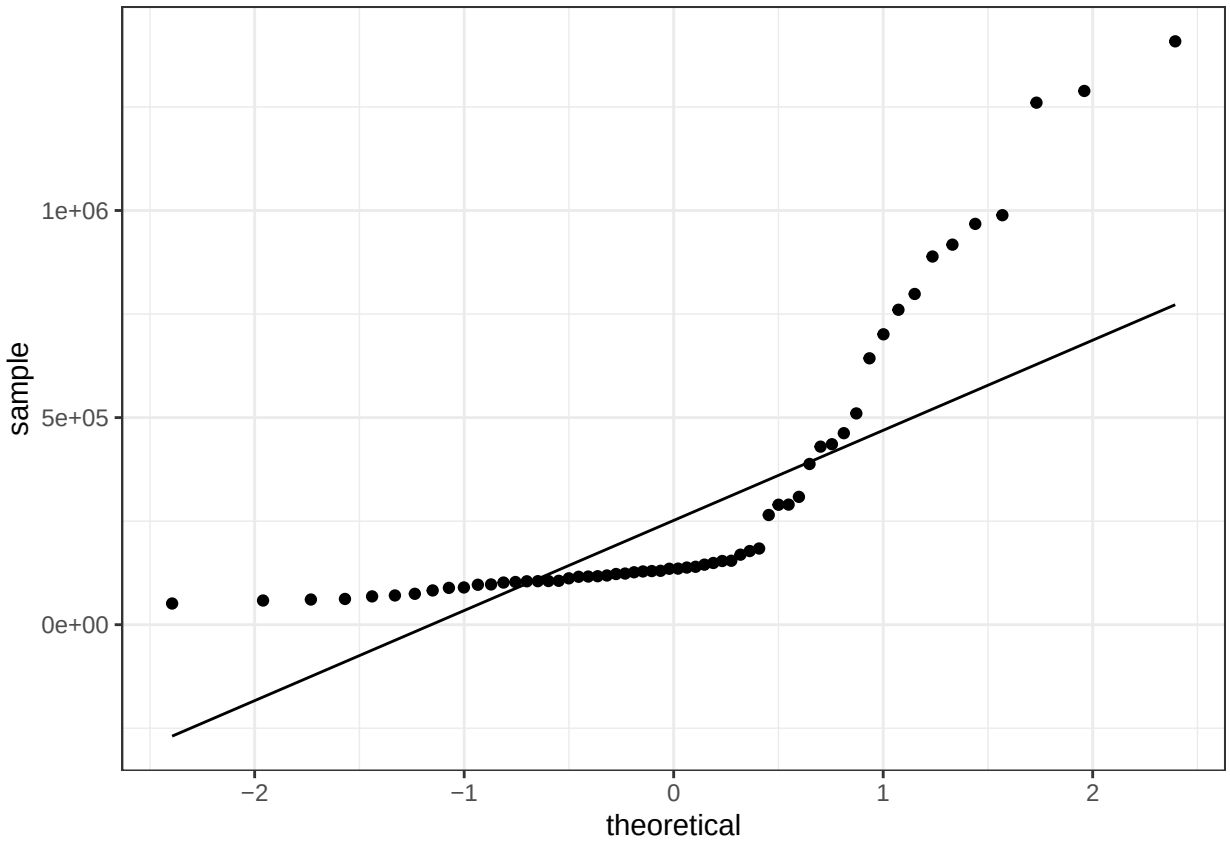
Results

Describing and Visualizing Head Coach data

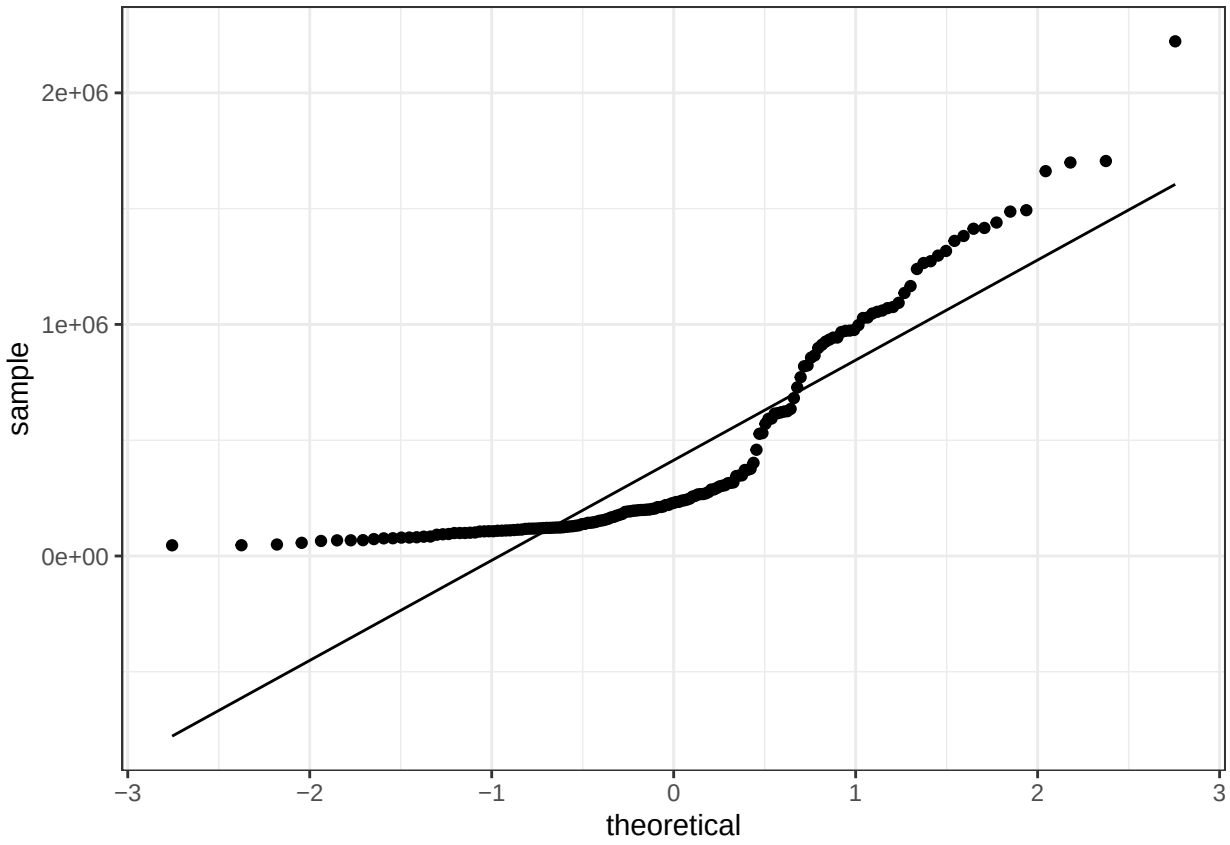
```
hd.coach.salary_data <- ncaa.div.i_data %>%
  select(sector_name, HDCOACH_SALARY_MEN)

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()
```

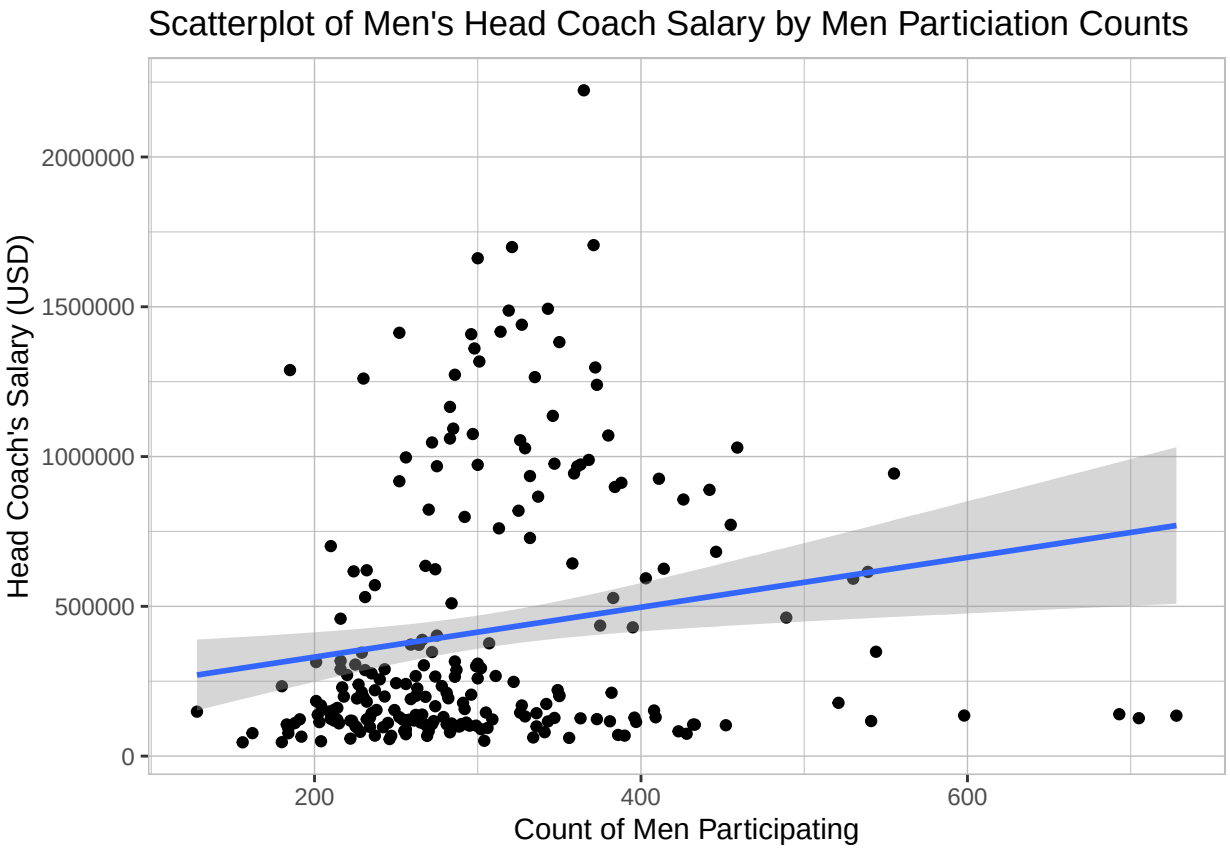



```
hd.coach.salary_data %>%
  filter(grepl('Public', sector_name)) %>%
  ggplot(aes(sample = HDCOACH_SALARY_MEN)) +
  stat_qq() + stat_qq_line() + theme_bw()
```



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")
  )
```



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/.