

Hmk4-Q3

```
setwd("/Users/Mavis1/Documents/ColumbiaUniv/Fall2018/EDAV/")
library(tidyverse)

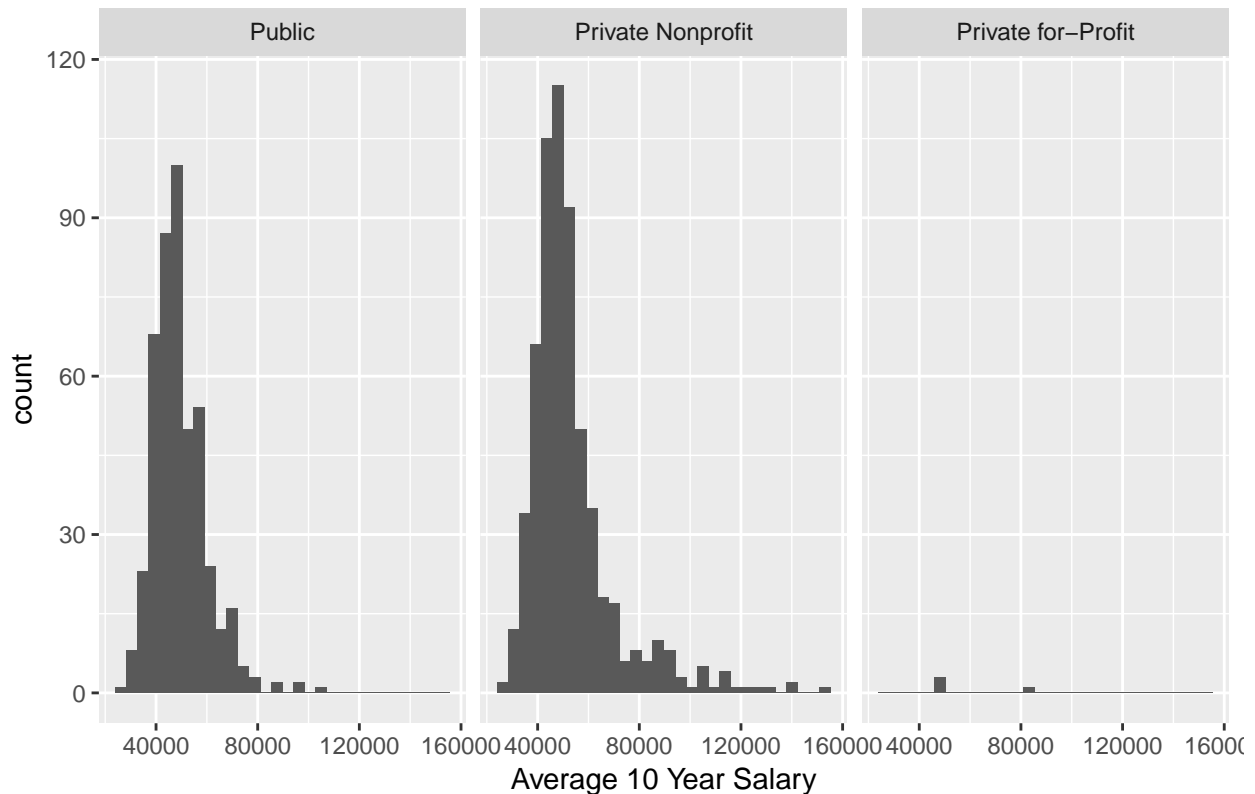
## -- Attaching packages ----- tidyverse 1.2
## v ggplot2 3.1.0      v purrr  0.2.5
## v tibble  1.4.2      v dplyr  0.7.7
## v tidyr   0.8.2      v stringr 1.3.1
## v readr   1.1.1      v forcats 0.3.0

## -- Conflicts ----- tidyverse_conflicts
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

college = read_rds("college.rds")
college[college=="NULL"] <- NA
college = na.omit(college)
recode <- c('1'='Public','2'='Private Nonprofit','3'='Private for-Profit')
college$ownership <- recode[college$ownership]
college$ownership<-fct_relevel(college$ownership,'Public','Private Nonprofit','Private for-Profit')

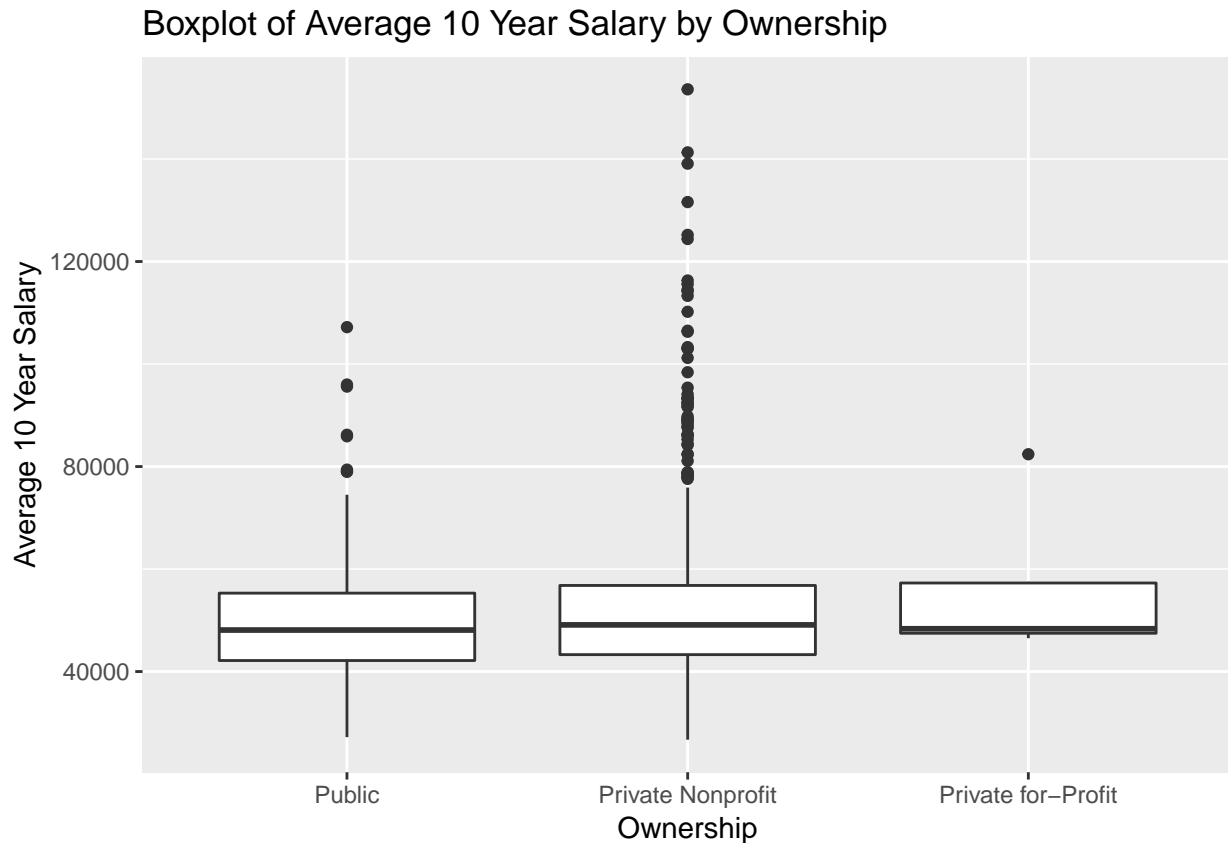
ggplot(college, aes(x=avg_10yr_salary))+
  geom_histogram(bins = 30)+
  facet_grid(.~ownership)+
  labs(x="Average 10 Year Salary")+
  ggtitle("Histogram of Average 10 Year Salary by Ownership")
```

Histogram of Average 10 Year Salary by Ownership



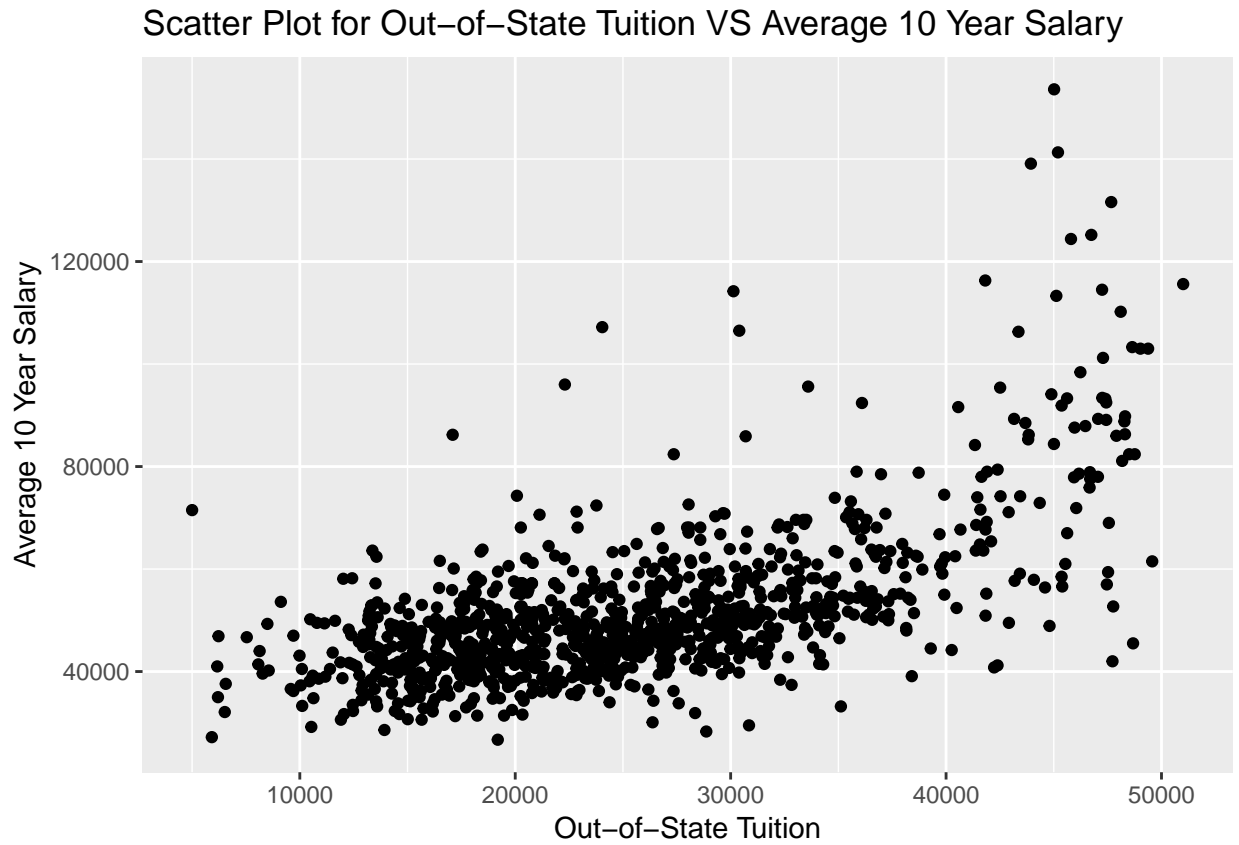
The average 10 year salary variable represents the mean earnings of students working and not enrolled at school 10 years after entry. Histogram above shows that the average 10 year salary follows a distribution skewed right. By faceting average 10 year salary into different ownership group, we can see that overall the distribution is similar for public and private nonprofit schools, and there is too few data for private for-profit schools. Comparing to public schools, the private nonprofit schools have the distribution more skewed right. This might indicate students from private nonprofit schools can have higher chance to get high salary, so this can be an interesting question to explore.

```
ggplot(college, aes(x=reorder(ownership, avg_10yr_salary), y=avg_10yr_salary)) +
  geom_boxplot() +
  labs(y="Average 10 Year Salary", x="Ownership") +
  ggtitle("Boxplot of Average 10 Year Salary by Ownership")
```



The boxplot for average 10 year salary gives us a better understanding of the median, as well as outliers of the data. We can see that the median for public schools is lower than private nonprofit ones. Also, the middle fifty percent of the data for public schools have lower average 10 year salary than private nonprofit schools. Moreover, there are more outliers for private nonprofit schools, and the outliers have higher salary than those for public schools. This also indicates students from private nonprofit schools might have higher average 10 year salary than those from public schools.

```
ggplot(college, aes(x=tuition_out,y=avg_10yr_salary))+
  geom_point()+
  labs(x="Out-of-State Tuition",y="Average 10 Year Salary")+
  ggtitle("Scatter Plot for Out-of-State Tuition VS Average 10 Year Salary")
```



We can see from the scatter plot that there is some positive correlation between out-of-state tuition and average 10 year salary. The more expensive the out-of-state tuition is, the higher the average 10 year salary is.