State University System Salaries

Carlos Arias

In this project we explored the Florida State University System Employee Data. We looked at how New College compares to other school's in the state and investigated differences between Administrative and Faculty salaries.

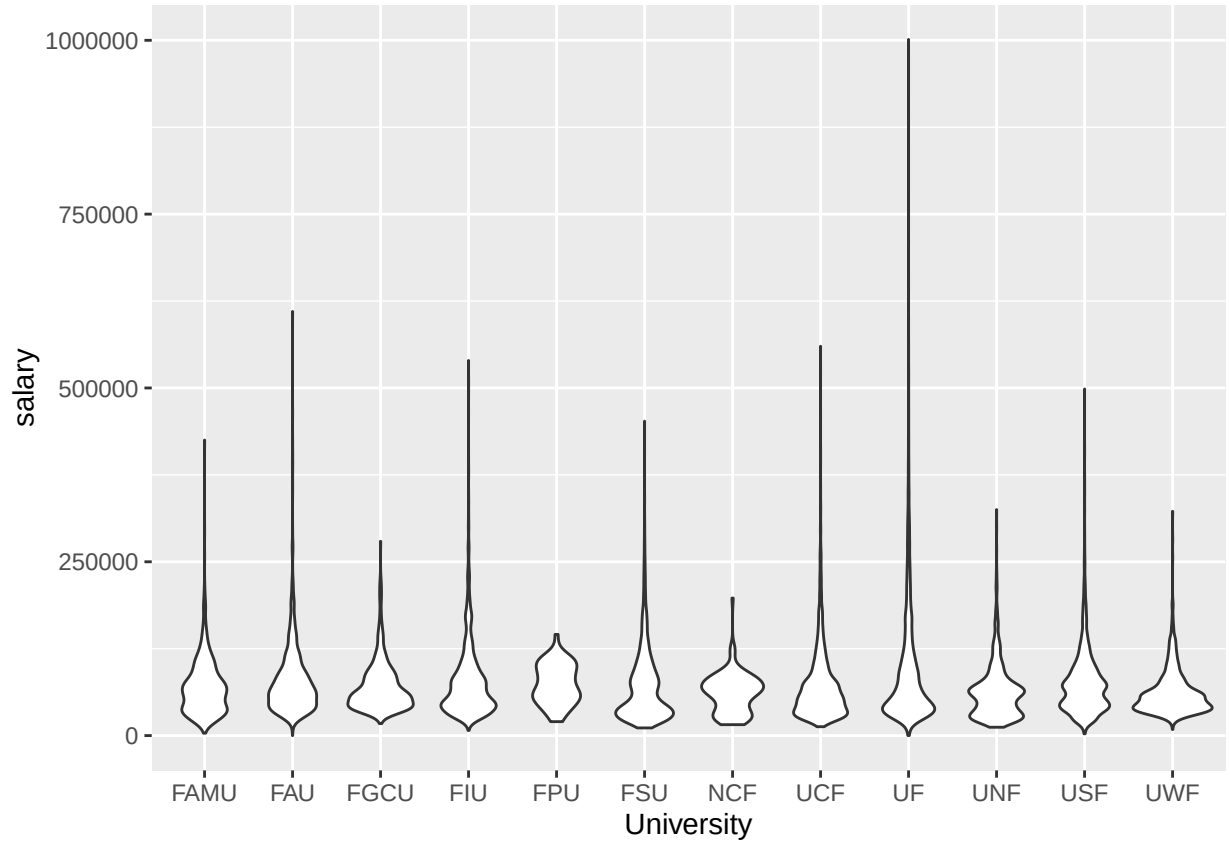
```
setwd('~/spring2016/viz/vizrepo/assignments/0221/')
library(data.table)
data <- fread('emp.csv')

# code type data was coded by hand. Only top 100 most common were used
code.types <- fread('code_types.csv')
```

Data Prep

```
# only looking at salaried workers
all.data <- data[`Employee Type` == 'SALARIED' ]

all.data[, FTE := as.numeric(FTE)]
all.data[, `Annual Salary` := as.numeric(`Annual Salary`)]

# getting sums of fte and salary
all.data[, salary := sum(`Annual Salary`), by = list(`First Name`, `Last Name`, MI)]
all.data[, sum.fte := sum(FTE), by = list(`First Name`, `Last Name`, MI)]

# adding category
all.data <- merge(all.data, code.types, by.x = 'Class Title', by.y = 'code')

# aggregating
all.data <- unique(all.data[, list(University, `Last Name`, `First Name`, MI, sum.fte, salary, type)])

# calculating salary per fte
all.data[sum.fte != 0.00, fte.salary := salary / sum.fte]

all.data[, school := 'Other']
all.data[University == 'NCF', school := 'NCF']
```

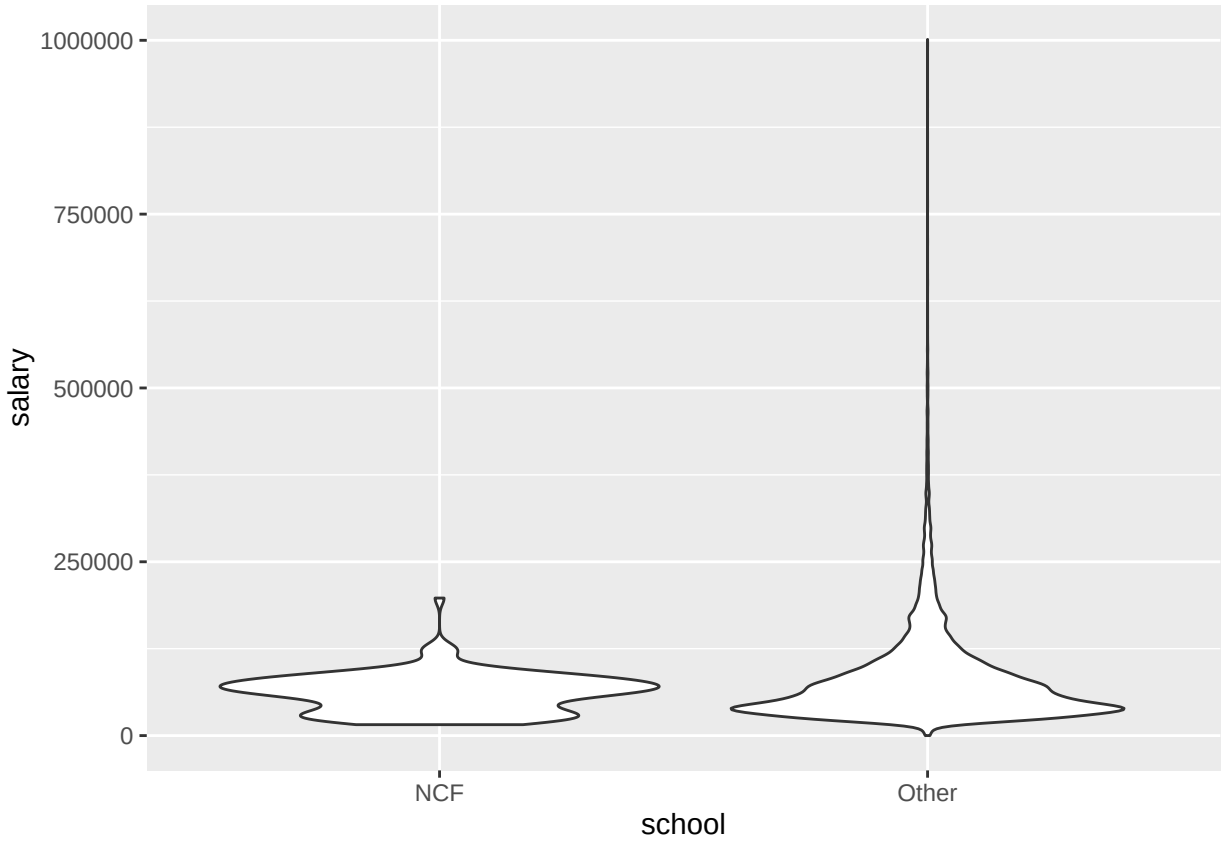
Most everyone has outliers

```
library(ggplot2)

plot <- ggplot(all.data, aes(University, salary))
plot + geom_violin()
```



Some distinctly fewer

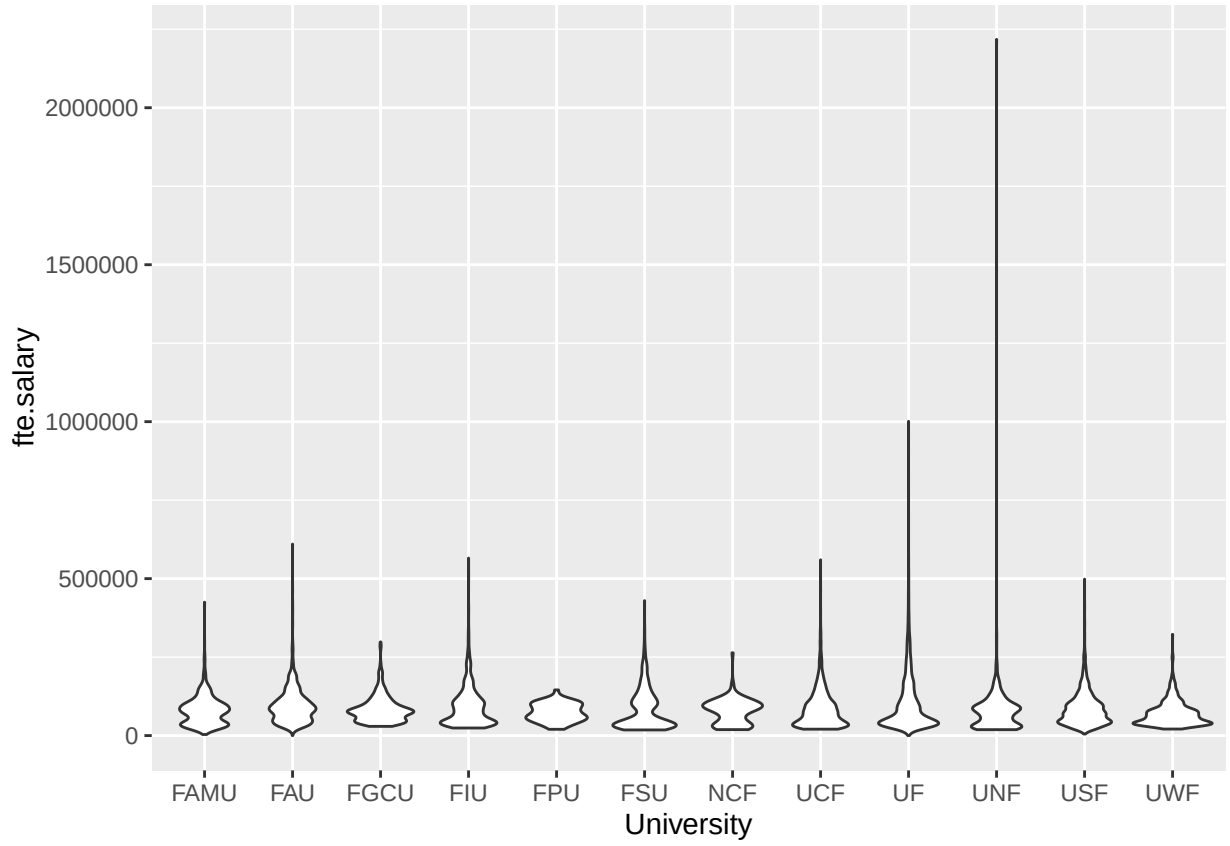
```
plot <- ggplot(all.data, aes(school, salary))
plot + geom_violin()
```



Salary adjusted for time (Salary / FTE) does not seem to show any major difference in distribution besides add more outliers. Note the dramaic change in y axis values

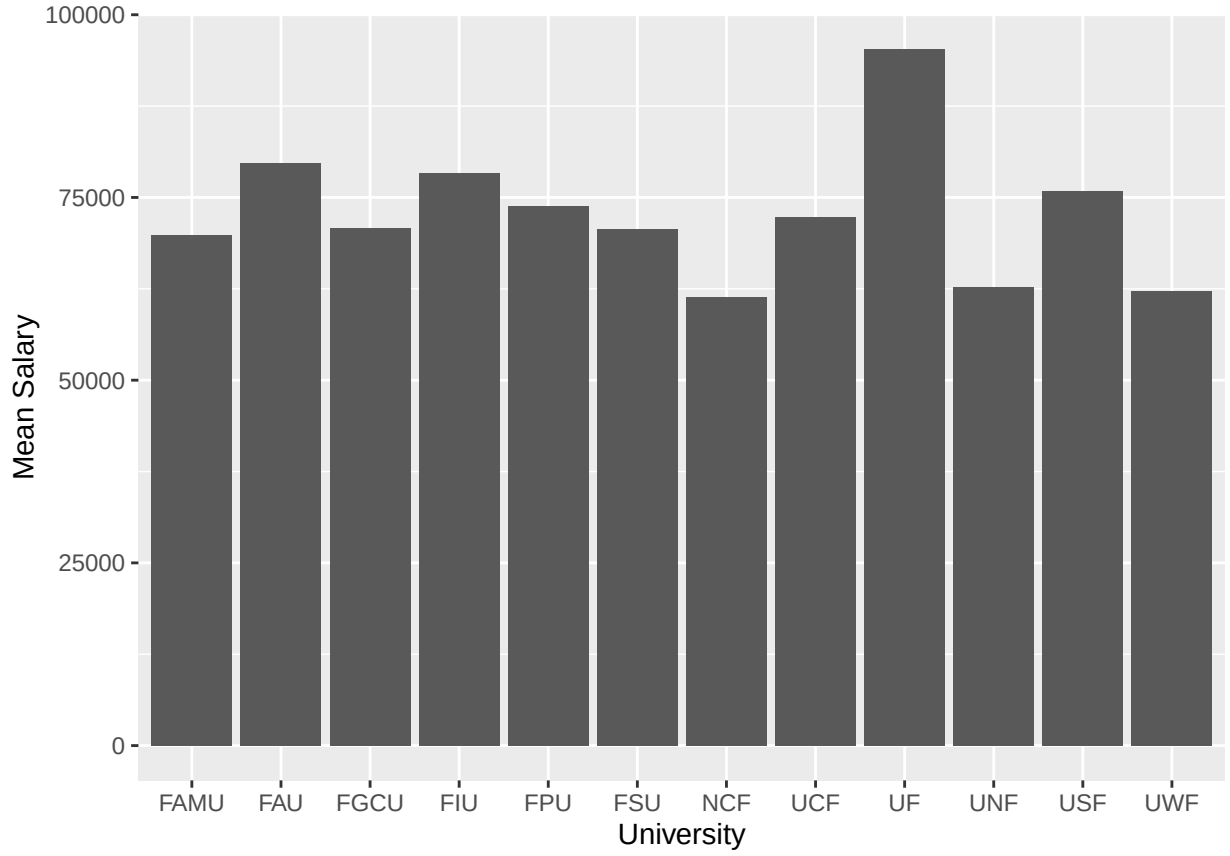
```
plot <- ggplot(all.data, aes(University, fte.salary))
plot + geom_violin()
```

```
## Warning: Removed 28 rows containing non-finite values (stat_ydensity).
```



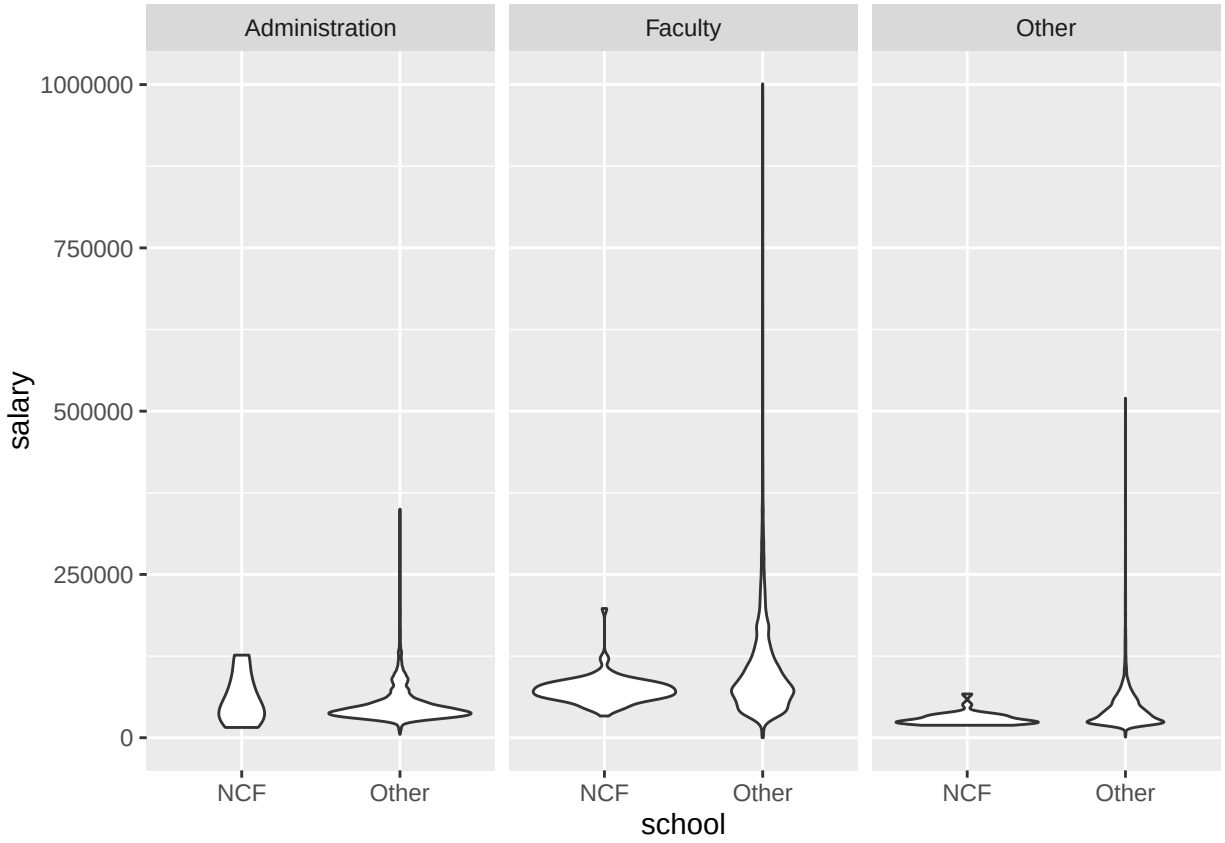
Means show some tendencies but not much difference

```
by.school <- all.data[, mean(salary), by = University]
plot <- ggplot(by.school, aes(University, V1))
plot + geom_bar(stat = 'identity') + ylab('Mean Salary')
```



These outliers seem to be mostly Faculty

```
plot <- ggplot(all.data, aes(school, salary))
plot + geom_violin() + facet_wrap(~ type)
```


Administration
Faculty
Other
1000000
750000
500000
250000
0
salary
NCF
Other
NCF
Other
NCF
Other
school