Assignment 3

Lance Price

March 6, 2017

Project Description

Open records law in the state of Florida mandate that salaries paid to all state employees be a matter of public record. In particular, salaries for all employees of the State University System can be found at https://prod.flbog.net:4445/pls/apex/f?p=140:1:0:::::

Your task is to

1. Provide a comparison study of how New College stacks up relative to the other members of the SUS. This should include a comparison of salary data for faculty.

2. Provide a study of how administrative salaries compare to faculty salaries.

Required Submissions

Create an R Markdown document or iPython notebook which provides an analysis of the data. The report should:

1. Discuss the quality of the data and any data cleaning/munging you have done.

2. Carefully explain what calculations were done and why.

3. Include supporting graphics.

4. Be reproducible.

I will first download the data and store it in a data frame.

```
# Download data from url and store it in a data frame
# Something weird is happening with setwd()...it only works in console
# The same line that doesn't work when running the line from the chunk
#   works when run explicitly from console.
dir <- "C:/Users/lance/Desktop/classwork/Spring2017/dataVis/HW"
filePath <- paste(dir,"/FLSUSSalaries.csv",sep = "")

#    I don't need the following two lines any more...already downloaded
#    I am keeping them here anyway, so I know what I did
#url <- "https://prod.flbog.net:4445/pls/apex/f?p=140:30:0"
#download.file(url, destfile = filePath)



# When reading the file in, I changed empty fields and fields with nothing but a "." in it
#  to NA.  These fields have no data and I want it to register as such for functions like
```

```
#  complete.cases.
salaryData <- read.csv(filePath, na.strings = c('','.'))
```

Next, I will access the quality and completeness of the data. I can tell that if I look at all columns and rows together, then the completeness percentage will be exactly 0% because there is either data in the last column or the second to last column, but never both of them. I can check to see if this is true.

The following gives the number of rows with numbers in both columns 11 or 12.

```
sum(complete.cases(salaryData[,11:12]))
```

```
## [1] 0
```

As suspected there are no complete cases whwn considering both columns together.

The following gives me the number of rows that have a number in column 11 or 12.

```
sum(complete.cases(salaryData[,11]) | complete.cases(salaryData[,12]))
```

```
## [1] 87031
```

The is the same as the number of total rows.

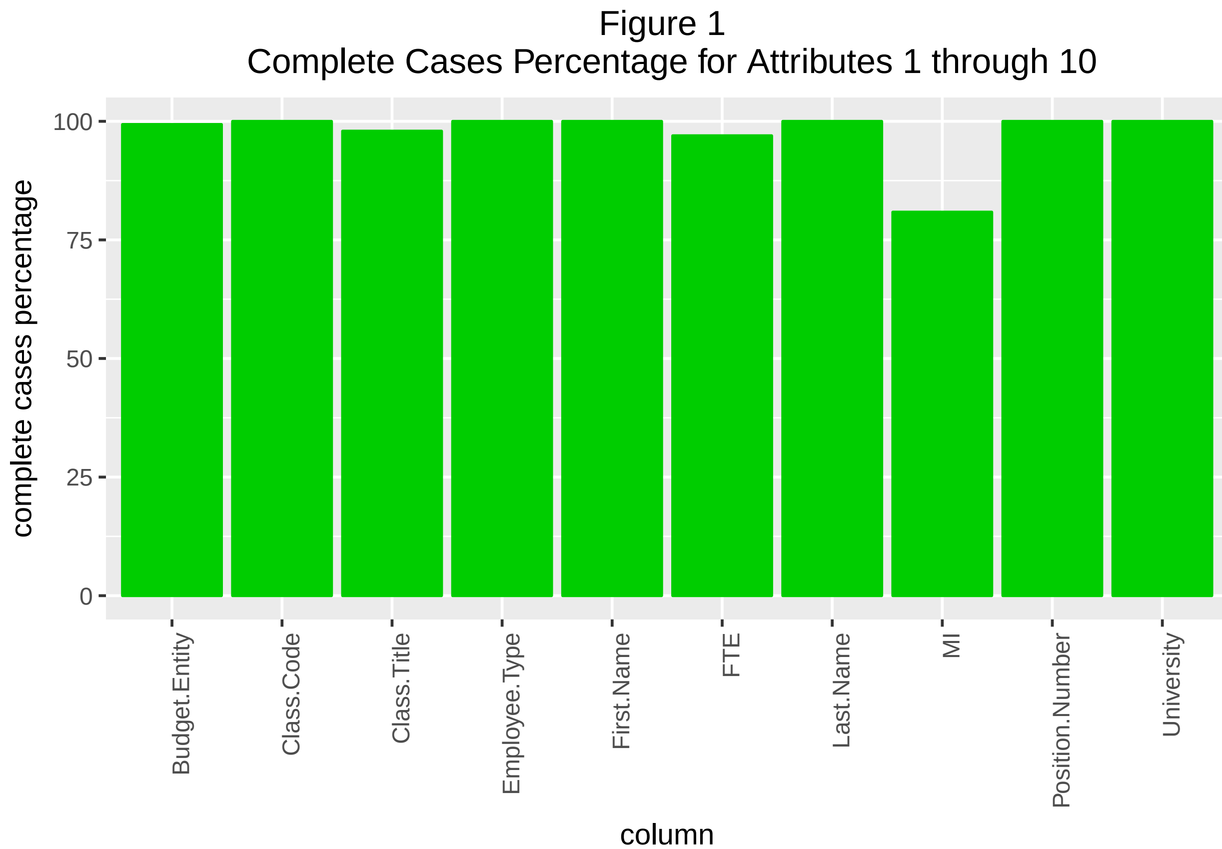
```
nrow(salaryData)
```

```
## [1] 87031
```

This information tells us that there is always a number in one of these columns but there is never a number in both columns for any particular row. Thus, I will say that they are both complete as a whole because there is always data on how much an individual is paid. The way they get paid can be different because some get paid by salary (column 11) and some get paid by term (column 12).

Now that I have looked at the completeness of these two columns, I will look at the other ones. I will make a bar graph that illustrates the percent complete for each of the other columns.

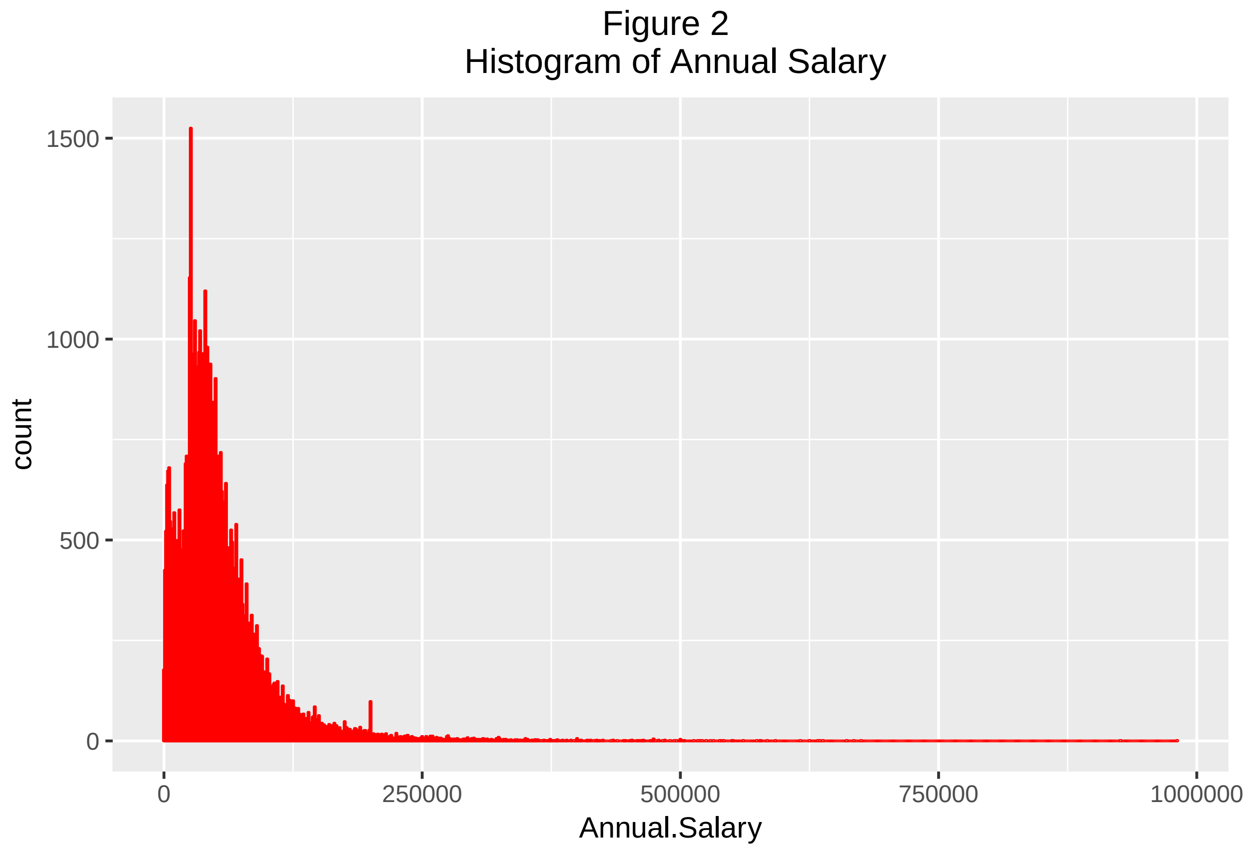
```
sum_complete <- function(x) {
  num_complete_cases <- sum(complete.cases(x))
  return(num_complete_cases)
}

matrix_df <- as.matrix(salaryData[1:10])
columns <- split(matrix_df, col(matrix_df))
complete_cases_percent <- unlist(lapply(columns, sum_complete))*100/nrow(salaryData)
complete_cases_percent <- cbind(complete_cases_percent,colnames(salaryData[,1:10]))
complete_percent_df <- as.data.frame(complete_cases_percent)
colnames(complete_percent_df)[2] <- "column"
f <- as.numeric(levels(complete_percent_df[,1]))[complete_percent_df[,1]]
complete_percent_df[,1] <- f
library(ggplot2)
ggplot(complete_percent_df, aes(column,complete_cases_percent)) +
  geom_bar(stat = "identity", color = 3, fill = 3) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  theme(plot.title = element_text(hjust = 0.5)) +
  ylab("complete cases percentage") +
  ggtitle(paste("Figure 1 \n","Complete Cases Percentage for Attributes 1 through 10"))
```



As we can see, the completeness percentage of each column is very high. The middle initial column is the lowest ($\approx 80\%$). I don't think the missing data for this attribute is very important when viewing the data set as a whole. I will not be using it in this report. All of the other attributes have a completeness percentage of ranging from 96.9677471 to 100. This is very good.

I will now look at some histograms of distributions of the numeric columns to identify any weird outliers and other funny business. The two numeric columns of interest are annual salary and OPS term amount.

```
ggplot(salaryData, aes(Annual.Salary)) +
  geom_histogram(binwidth = 1000, na.rm = TRUE, color = 2) +
  ggtitle(paste("Figure 2 \n","Histogram of Annual Salary")) +
  theme(plot.title = element_text(hjust = 0.5))
```



The long tail to the right startled me a bit because I had no idea that some people in the dataset make close to a million dollars. I decided to look at these lucky individuals - in particular anyone who makes over $600,000.

```
topSalRefinement <- which(salaryData[,11] >= 6e05)
topSals <- salaryData[topSalRefinement,1:11]
topSals <- topSals[order(topSals[,11]),c(1,2,4,5,10,11)]
library(knitr)
knitr::kable(topSals, caption = "Highest Salaries of Employees in Florida SUS", row.names = FALSE)
```

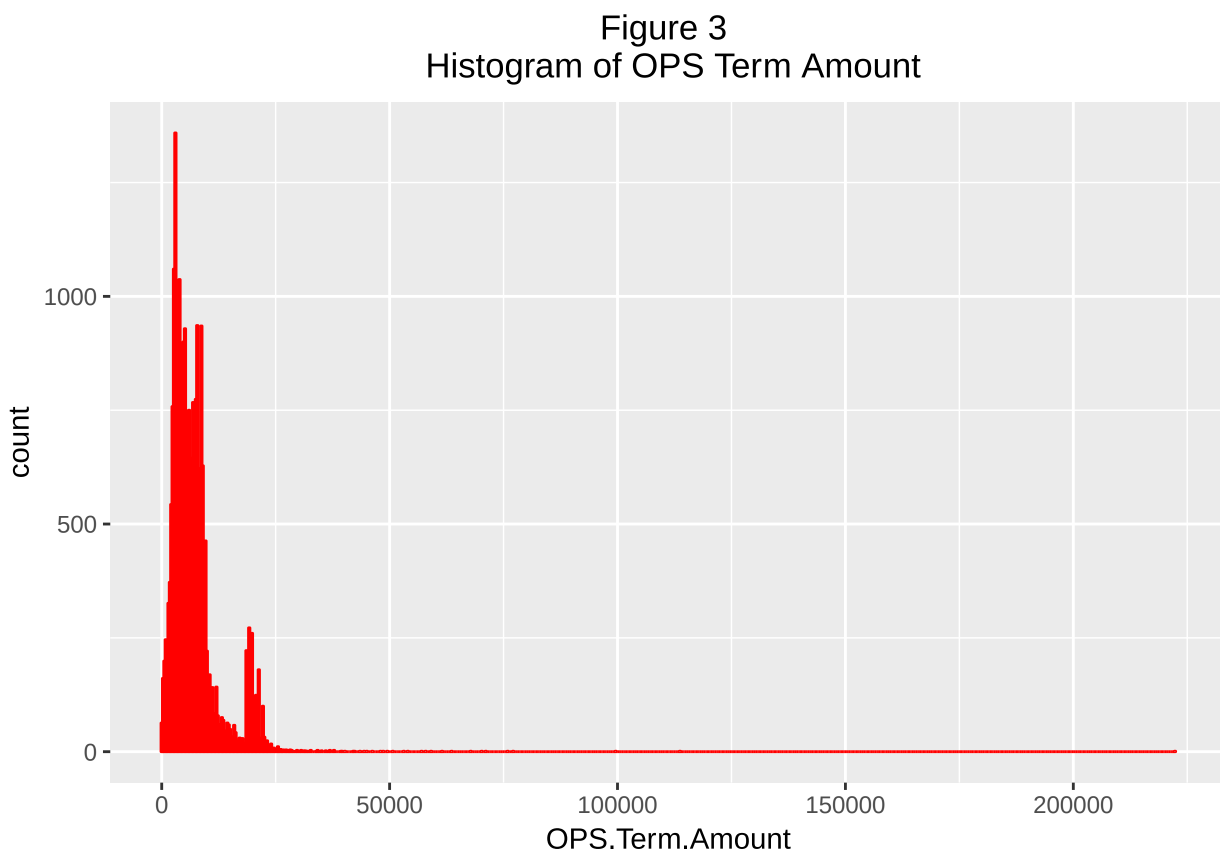
Table 1: Highest Salaries of Employees in Florida SUS

University	Budget.Entity	Last.Name	First.Name	Class.Title	Annual.Salary
UF	Contracts & Grants	GOOD	MICHAEL	PROFESSOR	616365
UF	Contracts & Grants	PIRRIS	JOHN	ASSISTANT PROFESSOR	625000
UF	Contracts & Grants	HOH	BRIAN	PROFESSOR	633281
UF	Contracts & Grants	TAVANAIEPOUR	DARYOUSH	ASSISTANT PROFESSOR	635125
UF	Contracts & Grants	SCARBOROUGH	MARK	PROFESSOR	638448
UF	Contracts & Grants	FUCHS	WESLEY	PROFESSOR	661340
UF	Contracts & Grants	ALDANA	PHILIPP	ASSOCIATE PROFESSOR	668272
UF	Contracts & Grants	DOUGHERTY	PAUL	PROFESSOR	675000
UF	Contracts & Grants	GUZICK	DAVID	PROFESSOR	926494
UF	Contracts & Grants	FRIEDMAN	WILLIAM	PROFESSOR	981000

All of these individuals are males who work at UF and are probably in charge of fancy, big-money labs. They all have traditionally masculine names, and they are all paid through Grants and Contracts budget entity. These high salaries make a some more sense now because they probably help to bring in a lot of money to UF and get paid through this incoming money. I must say that it is nice to see Professors, Assistant Professors, and Associate Professors as the highest paid individuals of UF. I have a sneaking suspicion that other schools do not have professors that are this high in the salary heirarchy. I also looked at the lowest salaried professors. There are 38 employees with a salary of less than or equal to $100. This looks suspicious and makes me doubt the quality of the data for this attribute. I don't know enough about the university system to judge whether these salaries are realistic or not. Even if this was a by term payment, I would still be highly sceptical about the quality of data for these employees. It sounds like slave labor or a recording error of some kind.

I will now look at a histogram of the OPS Term Amount attribute.

```
ggplot(salaryData, aes(OPS.Term.Amount)) +
  geom_histogram(binwidth = 300, na.rm = TRUE, color = 2) +
  theme(plot.title = element_text(hjust = 0.5)) +
  ggtitle(paste("Figure 3 \n","Histogram of OPS Term Amount"))
```

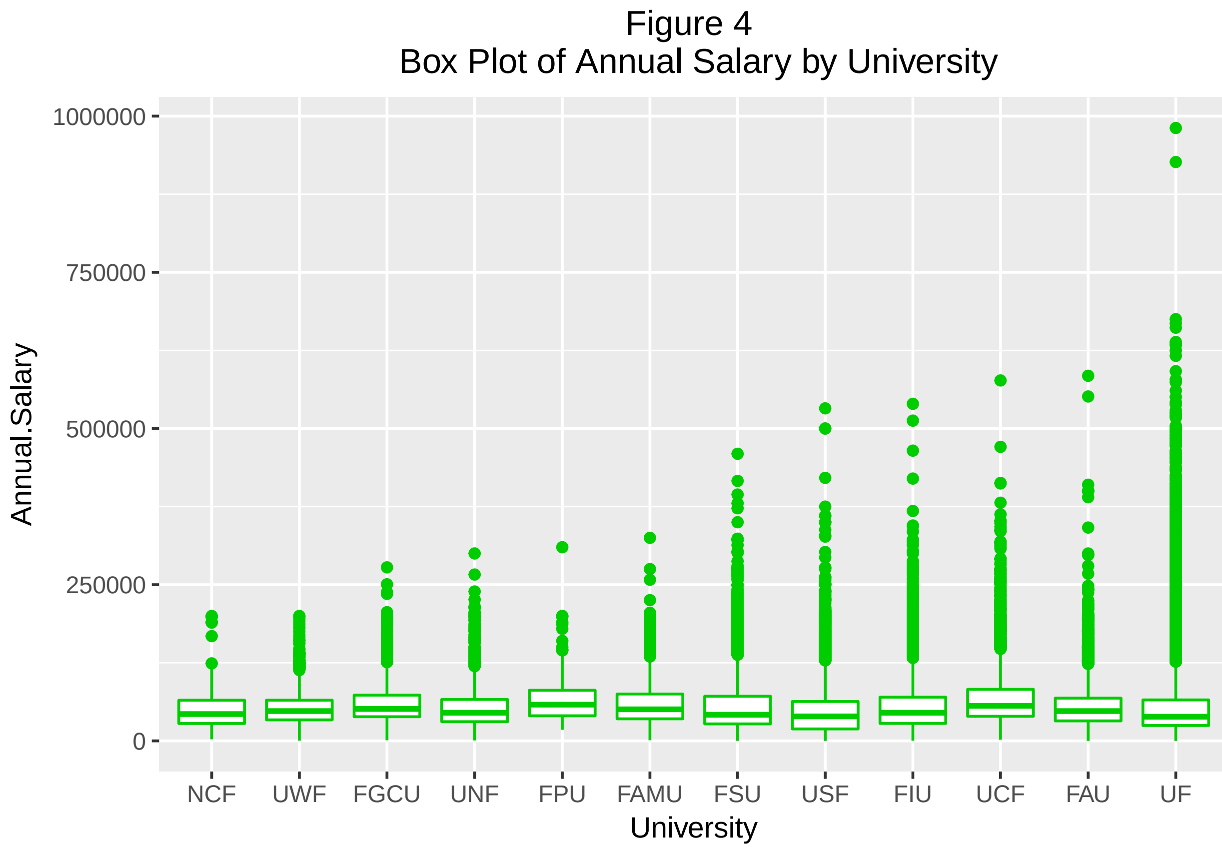


This attribute has a bimodal distribution. Much like the salary attribute, it has a very long right tail. The top ten ops term amounts are split half-way between the "Educational and General" and the "Grants and Contracts" budget entities, as opposed to the salary attribute in which all of the top ten salaries were accounted for by the Grants and Contracts budget entity. In addition, some women (4/10) snuck into the top ten ops term amounts (> $60,000 per term).

The top ops term amounts look realistic when taking into acount the top salaries. I am still sceptical about the extremely low salaries and ops term amounts. I can't quite wrap my head around these numbers. Maybe these people only worked a few hours for a given budget entity and position number. Some of these people are paid through multiple budget entities and have multiple positions. This would help to explain some of these low numbers.

I will now move on to some of the more in depth analysis. I will investigate how New College compares to the other members of the SUS. I will create a box plot that shows the distributions of salary for each university. I ordrred the schools on the maximum salary for each university.

```
schools <- unique(salaryData$University)

max_sal <- as.data.frame(rep(0,length(schools)))

rownames(max_sal) <- schools
colnames(max_sal) <- "max salary"

for (school in schools) {
  max_sal[school,1] <- max(salaryData$Annual.Salary[which(salaryData$University==school)],
                           na.rm = TRUE)
}

orderByMaxSal <- order(max_sal$'max salary')

for (i in seq(length(schools),1,-1)) {
  ith_school <- toString(schools[orderByMaxSal[i]])
  salaryData$University <- relevel(salaryData$University, ith_school)
}

ggplot(salaryData, aes(University, Annual.Salary)) +
  geom_boxplot(na.rm = TRUE, color = 3) +
  theme(plot.title = element_text(hjust = 0.5)) +
  ggtitle(paste("Figure 4 \n","Box Plot of Annual Salary by University"))
```



The first quartiles, medians, and third quartiles are similar across universities, but the maximum salaries change significantly across schools, with New College the lowest of all 12 colleges. Teachers are obviously not coming here for big pay checks. I would guess that they are attracted to the style of learning and the small student to teacher ratios. I also noticed that the outliers seem to be more dense for the other schools (i.e. there are more of them in general). Maybe this is due to the smaller student and teacher population at NCF, but I wanted to explore this observation in more detail. I first identified what a general outlier was for all schools put together because I didn't want to have a different outlier criteria for each school if I was going to use it to compare them.

```
minOutlierSals <- min(boxplot.stats(salaryData$Annual.Salary)$out)
```

The minimum salary for a general outlier of any school is $132974. I then used this statistic to see how many outliers there are for each school and normalize this by how many employees are in each school. This will give us a reasonable way to compare the schools density of outliers (or exceptionally paid employees with respect to the general number of employees at a given school). I ordered the following graph by this ratio of outliers to total employees.

```
numEmployees  <- rep(0,length(schools))
numOutliers <- rep(0,length(schools))

for (i in 1:length(schools)) {
  refineSchooli <- which(salaryData$University == schools[i])
  subsetSchooli <- salaryData[refineSchooli,1:11]
  refineSchooliOutliers <- which(subsetSchooli$Annual.Salary >=minOutlierSals)
  subsetSchooliOutlier <- subsetSchooli[refineSchooliOutliers, 1:11]
  numEmployees[i] <- nrow(subsetSchooli)
  numOutliers[i] <- nrow(subsetSchooliOutlier)
```

```
}

outliersPerEmp <- numOutliers / numEmployees
outliersPerEmp_df <- as.data.frame(outliersPerEmp)
rownames(outliersPerEmp_df) <- schools
outliersPerEmp_df <- cbind(outliersPerEmp_df, schools)
outliersPerEmp_df <- cbind(schools, outliersPerEmp_df)
outliersPerEmp_df <- outliersPerEmp_df[order(outliersPerEmp_df[, 2]),1:2]
colnames(outliersPerEmp_df) <- c("School", "Outliers Per Employee")

knitr::kable(outliersPerEmp_df, caption = "Ratio of Outliers per Employee", row.names = FALSE)
```

Table 2: Ratio of Outliers per Employee

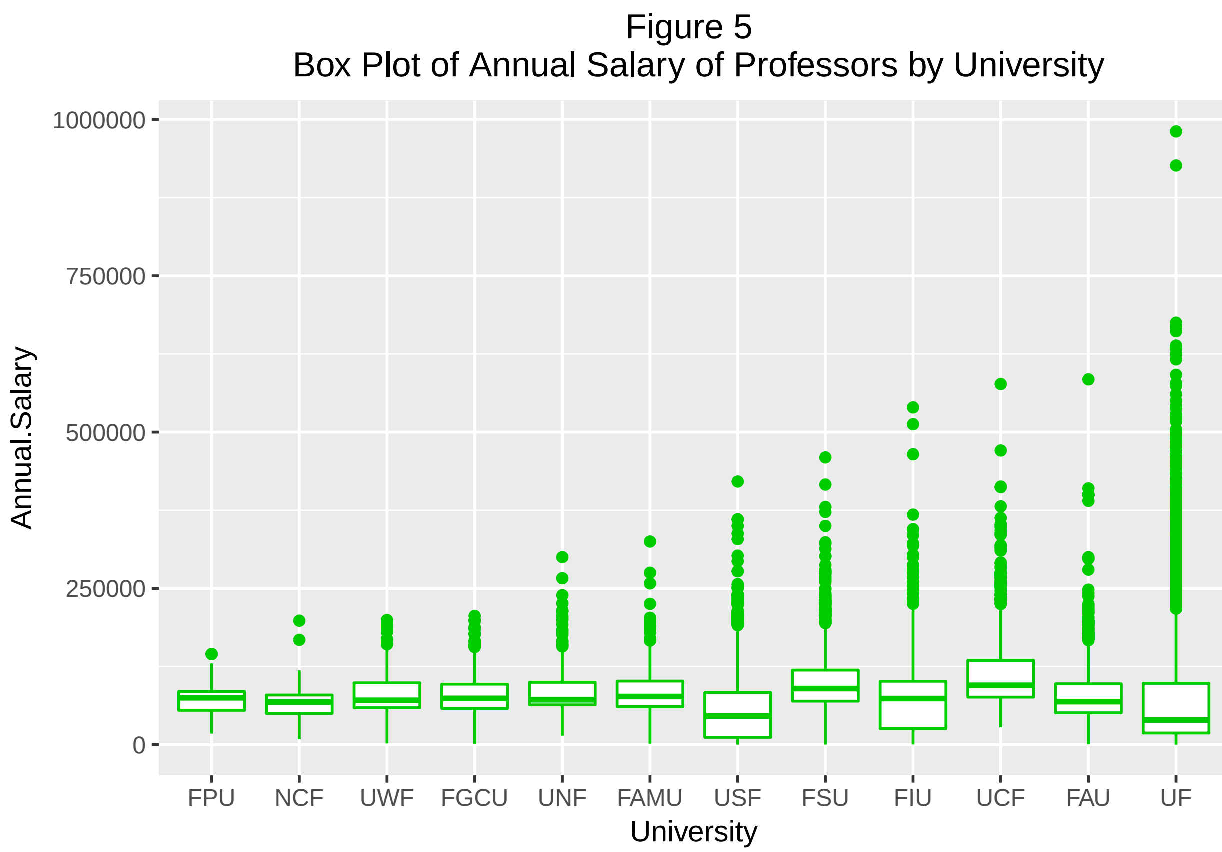
School	Outliers Per Employee
NCF	0.0139665
USF	0.0235948
UWF	0.0250553
UNF	0.0271084
FAU	0.0271352
FAMU	0.0294620
FGCU	0.0307150
FSU	0.0352710
FIU	0.0368606
UF	0.0539568
UCF	0.0573066
FPU	0.0573770

As suspected, New College is last according to this ranking criteria. This tells me that the exceptional employees (i.e. outliers) of New College are less likely to be rewarded with higher salaries than employees from other schools. There are other things at play here too, such as big money grants and contracts being rewarded to big labs and hospitals at these other institutions and the big name professors and medical doctors being rewarded for attracting such funds. It seems like you need to attract a lot of money to an institution for your salary to reflect your higher output and performance.

I will now subset the data set on professors. So the only employees that are accounted for in the next box plot are employees with "Professor" in their job title.

```
profRefine <- which(grepl("PROFESSOR", salaryData$Class.Title))
profSals <- salaryData[profRefine,1:11]

max_sal_prof <- as.data.frame(rep(0,length(schools)))

rownames(max_sal_prof) <- schools
colnames(max_sal_prof) <- "max salary"

for (school in schools) {
  max_sal_prof[school,1] <- max(profSals$Annual.Salary[which(profSals$University==school)],
                                na.rm = TRUE)
}

orderByMaxSalProf <- order(max_sal_prof$'max salary')
```

```r
for (i in seq(length(schools),1,-1)) {
  ith_school <- toString(schools[orderByMaxSalProf[i]])
  profSals$University <- relevel(profSals$University, ith_school)
}

ggplot(profSals, aes(University, Annual.Salary)) +
  geom_boxplot(na.rm = TRUE, color = 3) +
  theme(plot.title = element_text(hjust = 0.5)) +
  ggtitle(paste("Figure 5 \n","Box Plot of Annual Salary of Professors by University"))
```



The same general pattern is going on with New College now second to last with this ranking criteria. In addition, the pattern of density (outliers per employee) seems to be the same as before. I can test this observation.

```r
numEmployees  <- rep(0,length(schools))
numOutliers <- rep(0,length(schools))

for (i in 1:length(schools)) {
  refineSchooli <- which(profSals$University == schools[i])
  subsetSchooli <- profSals[refineSchooli,1:11]
  refineSchooliOutliers <- which(subsetSchooli$Annual.Salary >= minOutlierSals)
  subsetSchooliOutlier <- subsetSchooli[refineSchooliOutliers, 1:11]
  numEmployees[i] <- nrow(subsetSchooli)
  numOutliers[i] <- nrow(subsetSchooliOutlier)
}
```

```r
outliersPerEmp <- numOutliers / numEmployees
outliersPerEmp_df <- as.data.frame(outliersPerEmp)
rownames(outliersPerEmp_df) <- schools
outliersPerEmp_df <- cbind(schools, outliersPerEmp_df)
outliersPerEmp_df <- outliersPerEmp_df[order(outliersPerEmp_df[, 2]),1:2]
colnames(outliersPerEmp_df) <- c("School", "Outliers Per Employee")

knitr::kable(outliersPerEmp_df, caption = "Ratio of Outliers per Professor", row.names = FALSE)
```

Table 3: Ratio of Outliers per Professor

School	Outliers Per Employee
NCF	0.0186916
FPU	0.0232558
FAMU	0.0675477
USF	0.0681385
UNF	0.0800000
UWF	0.0869565
FGCU	0.0887728
FAU	0.0920938
FIU	0.1301956
UF	0.1518906
FSU	0.1750159
UCF	0.2438193

Once again, NCF is last in this ranking system, so the behavior holds for professors and not just employees in general. This similarity makes sense because it seems like the pattern is driven by professors attracting huge amounts of money to these schools and being rewarded for these incoming money streams.

```r
refineNCFProfs <- which(profSals$University == "NCF" & profSals$Annual.Salary >= 1.25e5)
top2ProfSalsNCF <- profSals[refineNCFProfs, 1:11]
```

I then identified the two professors that are outliers for the NCF box plot because I though there might be something interesting about them. They are GORDON MICHALSON and STEPHEN MILES. They are an ex-president and a soon-to-be ex-provost. Nice gigs… This brings us to another interesting aspect of the data - the salaries of administrative positions and how they compare to faculty salaries. I am particularly interested to see if the same relationships between administrative and faculty salaries hold between institutions. I have a feeling that administrators at New College make a lot more than the faculty. I also have a feeling that the converse is true at universities such as UF. I must first figure out a reasonable way to split up the administrative employees from the others.

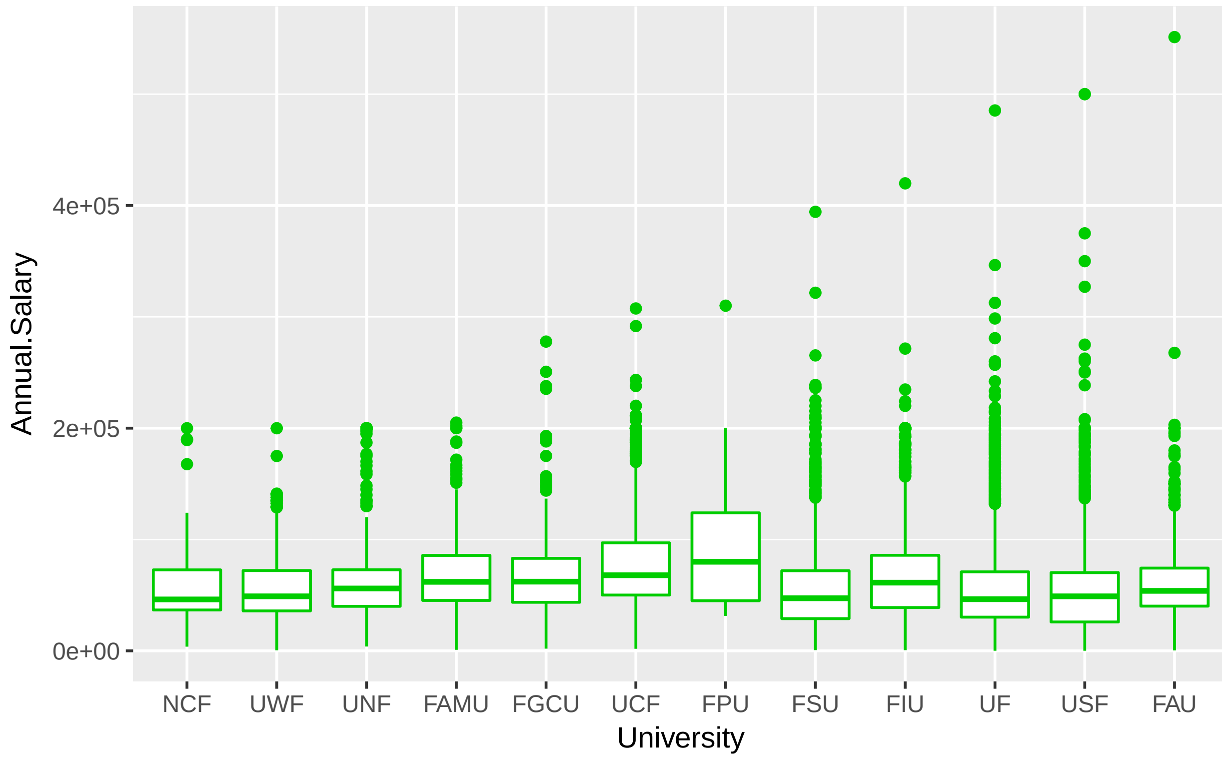
I first will make a list of keywords that would show up in an administrative position. This will not be an exhaustive list and I will be sure to miss some jobs while falsely labeling others. I am not sure if some jobs like "Administrative Assistants" are considered administrative jobs. My keyword list will include "ADMIN", "PRESIDENT", "VP", "PROVOST", "CIO", "CHIEF MARKETING OFFICER", "DIR" and "EXECUTIVE". I didn't realize how much I don't know about what an administrative position actually is. People use the word all the time, and I thought I knew what it meant. For example, is an athletic director an administrative position. It definitely sound like one, so I will include "ATHLETIC DIRECTOR", "ATHLETICS DIRECTOR" and "DIR, INTERCOLLEGIATE ATHLETICS" in my list. What about a head footbal coach? They are in charge of a lot of people, so I am tempted to include "HEAD FOOTBALL COACH", "HEAD BASKETBALL COACH", "HEAD ATHLETIC COACH" and "HEAD ATHL COACH" as well. I will do it, just to be sure I am including all possibilities. There are some other random things like "LIBRARIAN" that I will include because they are in charge of a lot of employees as well. I keep seeing "GENERAL COUNSEL" in the same general locations of administrative salaries for a given school, so that seems like they should also be included.

```
adminKeywords <- c("ADMIN", "PRESIDENT", "VP", "PROVOST", "CIO", "CHIEF MARKETING OFFICER",
                   "DIR", "EXECUTIVE", "ATHLETIC DIRECTOR", "ATHLETICS DIRECTOR",
                   "HEAD FOOTBALL COACH", "HEAD BASKETBALL COACH", "HEAD ATHLETIC COACH",
                   "HEAD ATHL COACH", "LIBRARIAN", "GENERAL COUNSEL")

refineAdmin <- which(grepl("ADMIN", salaryData$Class.Title))

for (key in adminKeywords) {
  refineAdmin <- c(refineAdmin, which(grepl(key, salaryData$Class.Title)))
}

adminSals <- salaryData[unique(refineAdmin),1:11]
```

I am seeing some issues with the analysis. There are many employees who are getting multiple salaries, and counting each one separately will lead to some very skewed results. Most of the analysis that was performed should probably be redone to account for this. I will do this at a later point in time. It will be included in the munging section.

```
max_sal_admin <- as.data.frame(rep(0,length(schools)))

rownames(max_sal_admin) <- schools
colnames(max_sal_admin) <- "max salary"

for (school in schools) {
  max_sal_admin[school,1] <- max(adminSals$Annual.Salary[which(adminSals$University==school)],
                           na.rm = TRUE)
}

orderByMaxSalAdmin <- order(max_sal_admin$'max salary')

for (i in seq(length(schools),1,-1)) {
  ith_school <- toString(schools[orderByMaxSalAdmin[i]])
  adminSals$University <- relevel(adminSals$University, ith_school)
}

ggplot(adminSals, aes(University, Annual.Salary)) +
  geom_boxplot(na.rm = TRUE, color = 3) +
  theme(plot.title = element_text(hjust = 0.5)) +
  ggtitle(paste("Figure 6 \n","Box Plot of Annual Salary of Administrative Positions by University"))
```

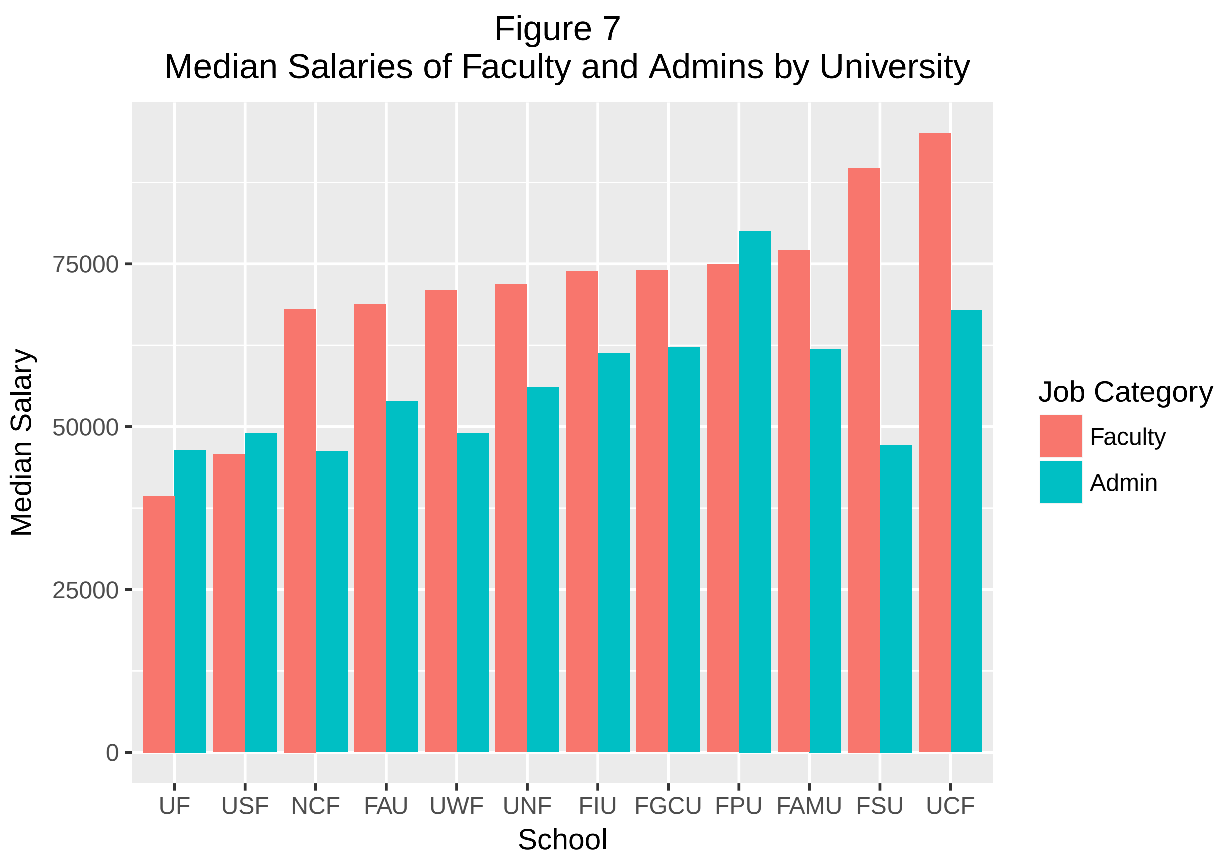
Figure 6
Box Plot of Annual Salary of Administrative Positions by University



Once again, NCF has the lowest maximum salary of all of the Universities - this time for administrative positions. For a quick comparison of the salaries of professors and administrators, I looked at the median salaries for faculty and admins by university.

```
medProfAdminSals <- rep(0,2*length(schools))

for (i in 1:length(schools)) {
  profSalsSchooli <- profSals[which(profSals$University == schools[i]),11]
  adminSalsSchooli <- adminSals[which(adminSals$University == schools[i]),11]
  medProfAdminSals[i] <- median(profSalsSchooli, na.rm = TRUE)
  medProfAdminSals[i+12] <- median(adminSalsSchooli, na.rm = TRUE)
}

medProfAdminSals_df <- as.data.frame(medProfAdminSals)
schools_twice <- as.data.frame(schools)
schools_twice <- rep(schools_twice,2)
library(reshape2)
schools_twice <- melt(schools_twice)[,1]
prof_admin <- c(rep(1,12),rep(2,12))
medProfAdminSals_df <- cbind(schools_twice, medProfAdminSals_df, prof_admin)
colnames(medProfAdminSals_df) <- c("School", "MedianSalary", "ProfAdmin")

medSalsProf <- medProfAdminSals_df[which(medProfAdminSals_df[,3]==1),1:2]
orderByMedSal <- order(medSalsProf$'MedianSalary')

for (i in seq(length(schools),1,-1)) {
  ith_school <- toString(schools[orderByMedSal[i]])
```

```
  medProfAdminSals_df$School <- relevel(medProfAdminSals_df$School, ith_school)
}

ggplot(medProfAdminSals_df, aes(x = School, y = MedianSalary, fill=factor(ProfAdmin))) +
  geom_bar(stat="identity", position="dodge") +
  scale_fill_discrete(name="Job Category",
                      breaks=c(1, 2),
                      labels=c("Faculty", "Admin")) +
  xlab("School")+ylab("Median Salary") +
  ggtitle(paste("Figure 7 \n", "Median Salaries of Faculty and Admins by University")) +
  theme(plot.title = element_text(hjust = 0.5))
```



In general, the median salary for administrators is lower than the median salary for faculty (it is higher for UF, USF, and FPU). I then looked at the mean salaries for faculty and admins by university.

```
meanProfAdminSals <- rep(0,2*length(schools))

for (i in 1:length(schools)) {
  profSalsSchooli <- profSals[which(profSals$University == schools[i]),11]
  adminSalsSchooli <- adminSals[which(adminSals$University == schools[i]),11]
  meanProfAdminSals[i] <- mean(profSalsSchooli, na.rm = TRUE)
  meanProfAdminSals[i+12] <- mean(adminSalsSchooli, na.rm = TRUE)
}

meanProfAdminSals_df <- as.data.frame(meanProfAdminSals)
meanProfAdminSals_df <- cbind(schools_twice, meanProfAdminSals_df, prof_admin)
```

```r
colnames(meanProfAdminSals_df) <- c("School", "MeanSalary", "ProfAdmin")

meanSalsProf <- meanProfAdminSals_df[which(meanProfAdminSals_df[,3]==1),1:2]

orderByMeanSal <- order(meanSalsProf$'MeanSalary')

for (i in seq(length(schools),1,-1)) {
  ith_school <- toString(schools[orderByMeanSal[i]])
  meanProfAdminSals_df$School <- relevel(meanProfAdminSals_df$School, ith_school)
}

ggplot(meanProfAdminSals_df, aes(x = School, y = MeanSalary, fill=factor(ProfAdmin))) +
  geom_bar(stat="identity", position="dodge") +
  scale_fill_discrete(name="Job Category",
                      breaks=c("1", "2"),
                      labels=c("Faculty", "Admin")) +
  xlab("School")+ylab("Mean Salary") +
  ggtitle(paste("Figure 8 \n", "Mean Salaries of Faculty and Admins by University")) +
  theme(plot.title = element_text(hjust = 0.5))
```

All schools but one (FPU) have a decrease in mean salary from faculty to admins.

To test the significance of these differences in mean, I ran a t-test between the distributions of professor salaries and administrator salaries by university. The following table shows the results.

```r
tTestProfsAdminsSals<- rep(0,length(schools))
pValProfsAdminsSals<- rep(0,length(schools))
```

```r
for (i in 1:length(schools)) {
  profSalsSchooli <- profSals[which(profSals$University == schools[i]),11]
  adminSalsSchooli <- adminSals[which(adminSals$University == schools[i]),11]
  tTestProfsAdminsSals[i] <- t.test(profSalsSchooli, adminSalsSchooli)$statistic
  pValProfsAdminsSals[i] <- t.test(profSalsSchooli, adminSalsSchooli)$p.value
}

tTest_df <- as.data.frame(tTestProfsAdminsSals)
pVal_df <- as.data.frame(pValProfsAdminsSals)
tTestpVal_df <- cbind(schools, tTest_df, pVal_df)
tTestpVal_df <- tTestpVal_df[order(tTestpVal_df[, 2]),1:3]
colnames(tTestpVal_df) <- c("School", "t Statistic for Admin and Professor Salaries",
                            "p-value of t Statistic")

knitr::kable(tTestpVal_df,
             caption = "t Statistics and p-values for Comparison of Professor and Admin. Salaries",
             row.names = FALSE)
```

Table 4: t Statistics and p-values for Comparison of Professor and Admin. Salaries

School	t Statistic for Admin and Professor Salaries	p-value of t Statistic
FPU	-2.631109	0.0099163
USF	1.279919	0.2006357
NCF	2.141354	0.0336428
FGCU	3.388358	0.0007525
FAMU	4.451915	0.0000097
FIU	4.603911	0.0000043
FAU	7.408634	0.0000000
UNF	8.588179	0.0000000
UWF	9.613635	0.0000000
UF	14.239531	0.0000000
UCF	15.623567	0.0000000
FSU	22.756061	0.0000000

With an alpha risk level of 0.05, all of the comparisons (except for USF) allow me to reject the null hypothesis that the means of the distributions are the same. In the case of FPU, the admin salaries trended up from the professor salaries. For the other 10 schools (i.e. the 12 total schools minus FPU and USF), the salaries tended to go down from professors to administrators. I have a feeling that if I remove administrative assistants, these general trends might be less apparent or reversed in some cases. In general, the scale of Figure 5 (i.e. professor salaries) is larger than the scale of Figure 6 (the admin. salaries), so this general tend is not that surprising. The admins don't get the crazy high paychecks of the hot shot MD-PhD professors that attract the big money grants.