Gender bias in NYC employment

John Brandt, Yin Guo, Una Radojicic

8/31/2018

Introduction

The gender pay gap is an often studied issue in the United States where men are routinely paid more than women year over year. It is a difficult issue to study because it is explained varyingly by sexism, education access, tenure, job preferences, and unequal home duties. We were interested in the following two questions:

1. Where is the gender pay gap the worst in NYC?
2. Is the gender pay gap partially explained by an agency's male/female ratio?

Exploration/Analysis

```r
# Load and convert to RDS
library(tidyverse)
x <- readRDS("payrolldata.rds")
```

Fix date issues

```
library(lubridate)
x$Agency.Start.Date <- mdy(x$Agency.Start.Date)
x <- x[x$Agency.Start.Date < ymd("2017-01-01"),]
x <- x[x$Agency.Start.Date > ymd("1940-01-01"),]
```

Remove unreliable data

```
x$Pay.Basis <- gsub("^ ", "", x$Pay.Basis)
x$Title.Description <- gsub("\\s+$", "",
                            x$Title.Description)
x$Agency.Name <- gsub("\\s+$", "", x$Agency.Name)
# Calculate tenure length
x$tenure <- floor((as.numeric(
  ymd("2017-01-01") - x$Agency.Start.Date))/365)
```

Remove unreliable data

```
x <- x %>%
  filter(Fiscal.Year == 2017 &
           Regular.Gross.Paid >= 0 &
           Total.OT.Paid >= 0 &
           Total.Other.Pay >= 0 &
           Regular.Gross.Paid >= 0)
```

Bin together agencies by group

```
x$Agency.Name[grep("COMMUNITY BOARD|COMMUNITY BD",
                   x$Agency.Name)] <- "COMMUNITY BOARD"
x$Agency.Name[grep("COLLEGE",
                   x$Agency.Name)] <- "EDUCATION"
x$Agency.Name[grep("DISTRICT ATTORNEY",
                   x$Agency.Name)] <- "DISTRICT ATTORNEY"
x$Agency.Name[grep("PUBLIC ADMINISTRATOR",
                   x$Agency.Name)] <- "PUBLIC ADMINISTRATOR
x$Agency.Name[grep("BOROUGH PRESIDENT",
                   x$Agency.Name)] <- "BOROUGH PRESIDENT"
```

Estimate gender based on first name & census

```
gender <- gender::gender(x$First.Name)
gender.name <- gender %>%
  group_by(name, gender) %>%
  summarise(n=n())
x <- left_join(x, gender.name,
                by = c("First.Name" = "name"))
```

```
x$female <- NA
x$male <- NA
x$female[x$gender == "female"] <- 1
x$female[x$gender == "male"] <- 0
x$male[x$gender == "male"] <- 1
x$male[x$gender == "female"] <- 0
```

Wrangle

Create a dataframe of male and female salaary difference by grouped agency in terms of regular gross paid.

```
gender.group <- x %>%
  group_by(Agency.Name, gender) %>%
  mutate(salary = mean(Regular.Gross.Paid)) %>%
  ungroup() %>%
  group_by(Agency.Name) %>%
  summarize(female=sum(female, na.rm=T),
            male=sum(male, na.rm=T),
            salary = mean(salary)) %>%
  mutate(perc = female/(male+female)) %>%
  select(Agency.Name, perc)
```

Continued

```
x2 <- left_join(x, gender.group) %>%
  group_by(gender, Agency.Name) %>%
  summarise(salary=mean(Regular.Gross.Paid),
            percent=mean(perc))

x2 <- x2[!is.na(x2$gender),] %>%
  arrange(Agency.Name)
```

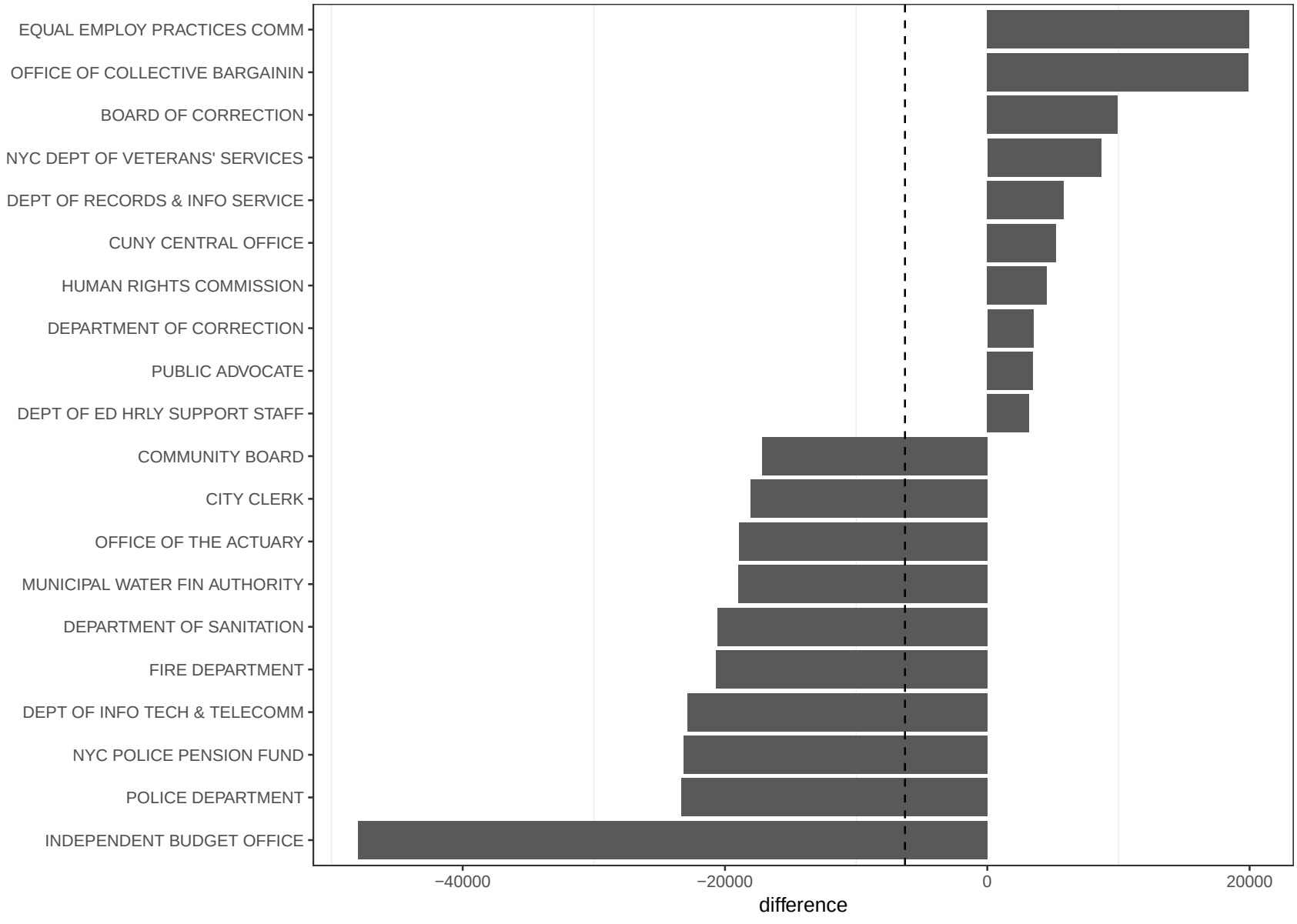
Calculate the actual pay difference in dollars

```
diff <- data.frame(Agency = x2$Agency.Name[x2$gender == "fe
                           difference = x2$salary[x2$gender
                             x2$salary[x2$gender == "male"])

top <- diff %>%
  top_n(10, difference)

bottom <- diff %>%
  arrange(difference)
bottom <- bottom[1:10,]
topdiffs <- rbind(top, bottom)
```

Code for Plot 1 - agency vs difference

```
p1 <- ggplot(data=topdiffs,
             aes(x=reorder(Agency, difference),
                                  y = difference))+
  geom_col()+
  coord_flip()+
  geom_hline(yintercept=mean(diff$difference),
             linetype="dashed")+
  theme_bw()+
  theme(panel.grid.major=element_blank())+
  xlab("")
```

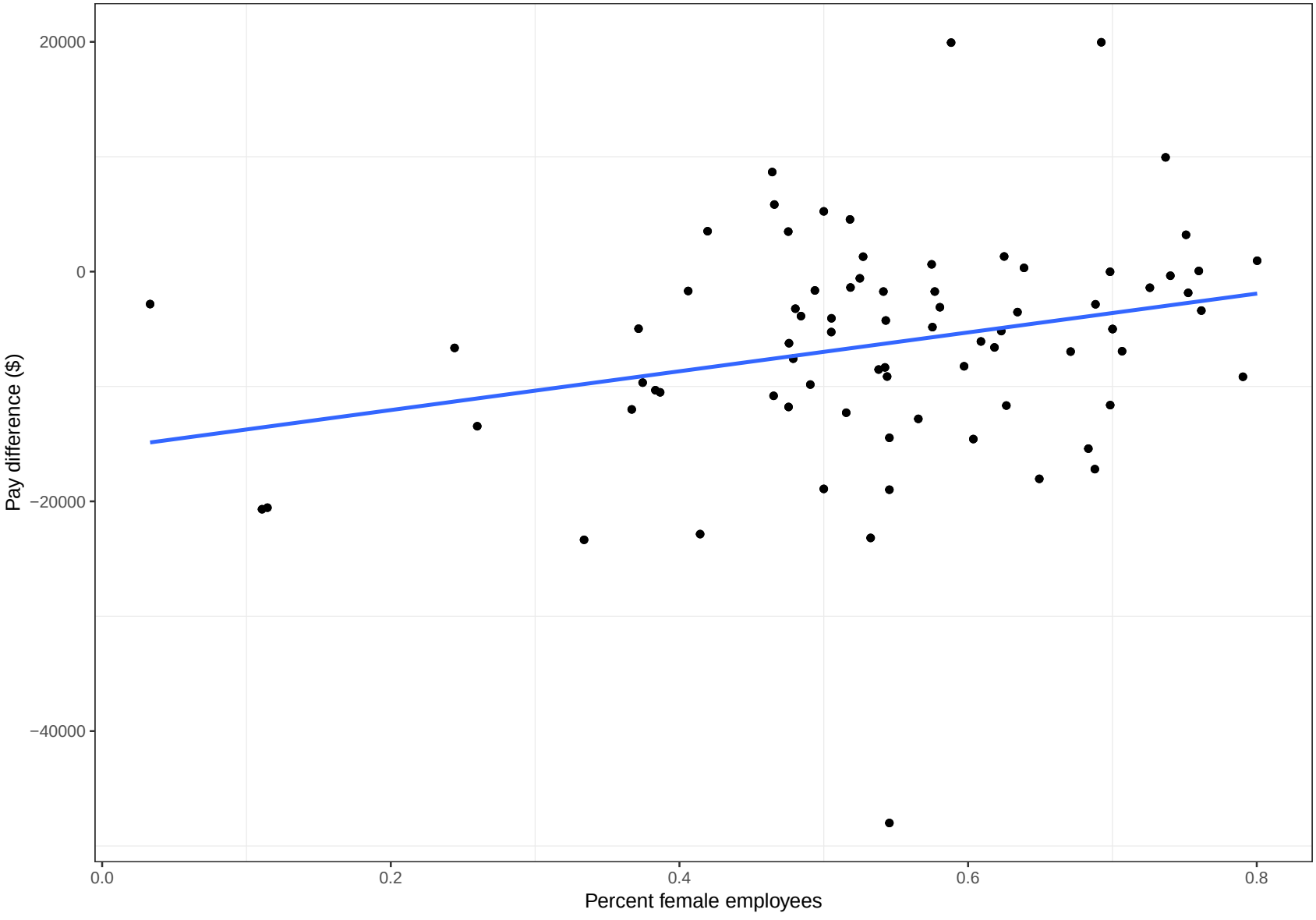
Figure 1


EQUAL EMPLOY PRACTICES COMM
OFFICE OF COLLECTIVE BARGAININ
BOARD OF CORRECTION
NYC DEPT OF VETERANS' SERVICES
DEPT OF RECORDS & INFO SERVICE
CUNY CENTRAL OFFICE
HUMAN RIGHTS COMMISSION
DEPARTMENT OF CORRECTION
PUBLIC ADVOCATE
DEPT OF ED HRLY SUPPORT STAFF
COMMUNITY BOARD
CITY CLERK
OFFICE OF THE ACTUARY
MUNICIPAL WATER FIN AUTHORITY
DEPARTMENT OF SANITATION
FIRE DEPARTMENT
DEPT OF INFO TECH & TELECOMM
NYC POLICE PENSION FUND
POLICE DEPARTMENT
INDEPENDENT BUDGET OFFICE
−40000
−20000
0
20000
difference

Code for plot 2 - difference vs. male/female ratio

```
x2 <- left_join(x2, diff,
                by = c("Agency.Name" = "Agency"))
p2 <- ggplot(data=x2, aes(y=difference, x=percent))+
  geom_point()+
  geom_smooth(method="lm", se=F)+
  theme_bw()+
  theme(panel.grid.major=element_blank())+
  xlab("Percent female employees")+
  ylab("Pay difference ($)")
```

Figure 2


20000
0
−20000
−40000
Pay difference ($)
0.0
0.2
0.4
0.6
0.8
Percent female employees

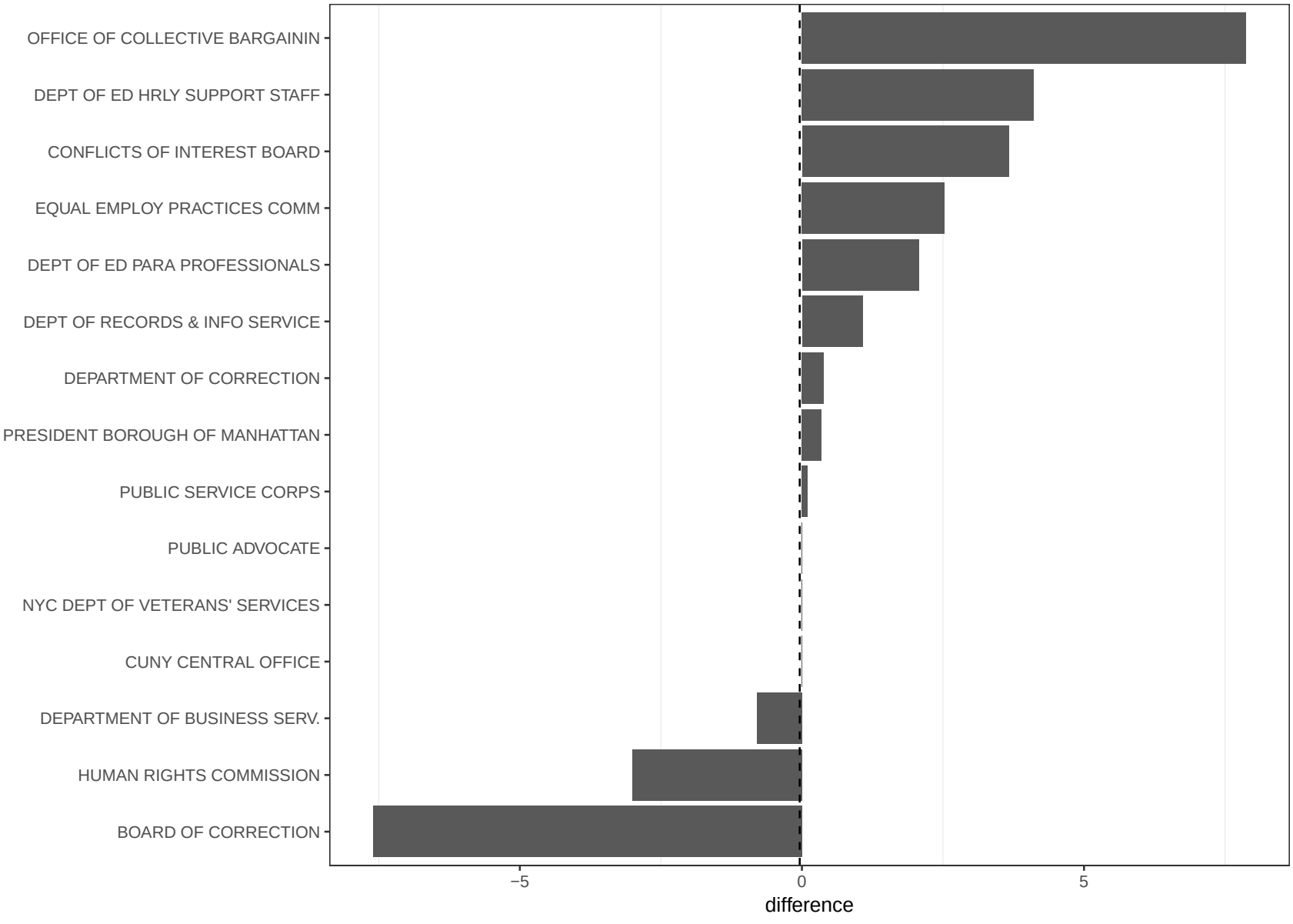
Tenure

```
tenure.group <- x %>%
  group_by(Agency.Name, gender) %>%
  mutate(tenure2 = mean(tenure)) %>%
  ungroup() %>%
  group_by(Agency.Name) %>%
  summarize(female=sum(female, na.rm=T),
            male=sum(male, na.rm=T),
            tenure = mean(tenure)) %>%
  mutate(perc = female/(male+female)) %>%
  select(Agency.Name, perc)
```

Top 15 jobs in which women are paid better than men

```
p11 <- ggplot(data=tenuretop,
              aes(x=reorder(Agency, difference),
                                  y = difference))+
 geom_col()+
 coord_flip()+
 geom_hline(yintercept=mean(tenure.diff$difference),
            linetype="dashed")+
 theme_bw()+
 theme(panel.grid.major=element_blank())+
 xlab("")
```

Top 15 jobs in which women are paid better than men



Top 15 jobs in which women are paid less than men

```
p12 <- ggplot(data=tenurebottom,
              aes(x=reorder(Agency, difference),
                                    y = difference))+
 geom_col()+
 coord_flip()+
 geom_hline(yintercept=mean(tenure.diff$difference),
            linetype="dashed")+
 theme_bw()+
 theme(panel.grid.major=element_blank())+
 xlab("")
```

Top 15 jobs in which women are paid less than men