Analysis of Staff Members in Jail Facilities

Michelle Seo and Lucy Wang

May 2020

1 Introduction

Numerous studies have been conducted to demonstrate relationships between different demographic characteristics and crime rates. The focus of these studies have always been on prisoners and drawing correlations between factors such as race or gender to an individual's likelihood of conviction. In our study, however, we want to focus instead on the condition of jails and their staff members. The main question being — do jail conditions differ depending on the average daily population (ADP) of the jail? Through this study, we hope to better understand as well as become more aware of jail conditions from the aspect of staff members in different facilities.

1.1 Questions

We first want to analyze whether there has been an increasing trend in the number of staff members in jail facilities. Has there been a greater number of correctional officers over the years? What about the number of other staff members? Have more staff members been employed in jail facilities recently?
Next, we examine if violence on staff members in jail facilities differ based on the average daily population of jails. Do jails with higher ADP have more cases of assaults than jails with lower ADP? Do certain types of staff members (e.g. correctional officer) have more cases of assaults or deaths by assaults?

We also investigate the gender ratio of staff members and the inmate-to-staff ratio over the years. Is there a significant difference between genders of staff members in jail facilites? Has there been an increasing gender gap in jail facilities over the years? Has there been more or fewer staff members employed in relation to the number of inmates over the years?

Through these questions, we ultimately reach a conclusion for our main question of whether jail conditions differ depending on the average daily population in jails.

2 Survey Description

2.1 Annual Survey of Jails

The Annual Survey of Jails is a survey series that samples jurisdictions from the Census of Jails and provides information on jails and their inmates. The survey consists of various attributes such as inmate count, inmate status, jail design capacities, number of assaults against staff etc. The survey data is for a wide range of audiences, federal/state agencies, local officers, researchers, and the public, to help address the issues facing jurisdictions today. The survey has been conducted yearly since 1982. Until 2014, the Census Bureau was in charge of collecting data, and starting 2015, RTI International has been collecting since. The Annual Survey of Jails provides web-based, mail, and fax surveys (from July to December) for its respondents. To minimize the nonresponse, the survey also conducted followup emails and calls to obtain more responses.

2.2 Survey Design

The Annual Survey of Jails is a stratified one-stage probability sampling method on jail jurisdictions. All jail jurisdictions are first classified into 10 strata based on two variables, average daily population (ADP) and the presence of juveniles in local jails. From the 10 strata, 8 of the 10 strata took a random sample of jails, while the other 2 selected all jurisdictions. Table 1 presents details on strata below. The sampling frame of the survey is the Census of Jails, which is a census on local jails administered every five to six years. In this report, we will be focusing on surveys from 2013 to 2018, and thus the project's sampling frame is the 2013 Census of Jails. The survey samples 882 out of 2871 jurisdictions, and receives 797 responses (with 11 ineligible jurisdictions).

A note to bear in mind is that the survey requires X marks next to numerical values if they are estimated values. Furthermore, facilities that are used solely as a purpose for temporary holding are not included in the survey, and facilities with multiple jurisdictions are contacted directly instead.

Table 1: Description of the 10 Strata*

	Presence of Juvenile	ADP	Jurisdictions	Number Sampled	Number Responded
1	Jurisdiction Certainties		253	257	237
1.1	California Jail Certainties		65	65	59
2	yes	264-499	95	35	33
3	yes	141-263	94	20	15
4	yes	69-140	72	8	8
5	yes	0-68	89	12	10
7	no	227-749	266	208	188
8	no	103-226	407	83	72
9	no	40-102	567	63	58
10	no	0-39	894	60	53
12	Regional Jail Uncertainties		69	71	64

*Referenced from Codebook Table 1

2.3 Description of Variables

For our project, we will be primarily observing variables that are closely related to staff variables. Table 2 delineates our variables of study.

Table 2: Variables

Variable	Description	Value
TOTALSTAFF	Number of total staff employed	The sum of correctional staff (CORRSTAFF) and other staff (OTHERSTAFF).
CORRSTAFF	Number of correctional officers in facilities (Correctional Officers refer to "deputies, monitors, and custody staff who spends more than 50 percent of their time with the incarcerated population")	The sum of female correctional officers (CORRSTAFFF) and male correctional officers (CORRSTAFFM).
OTHERSTAFF	Number of total other staff (administrators, clerical and maintenance, technical, and other)	The sum of female other staff (OTHERSTAFFF) and male other staff (OTHERSTAFFM).
TOTPOP	Total Population at year end (also denoted as S4QIV1C for 2013 and 2014 survey series)	Numerical value (ranges from 1 to 17060)
ASSAULTCORR	Number of assaults on correctional officers	Numerical value (ranges from 0 to 217)
ASSAULTTOTRSTAFF	Number of assaults on all other staff	Numerical value (ranges from 0 to 9)
DEATHSTAFFYN	Staff deaths due to assaults from inmates	Numerical value (ranges from 0 to 2)

3 Exploratory Data Analysis

3.1 Data Cleaning

After downloading datasets from the NACJD (National Archive of Criminal Justice Data), we merged the data across different years (2013-2018) on the same columns – renaming any column names which were recorded differently in earlier years. We switched the file format from tsv to csv. We then appended a 'YEAR' column which indicated which year this data was from.

3.2 Nonresponse

In the Annual Survey of Jails, the survey deals with two kinds of nonresponse: unit nonresponse and item nonresponse. Initially, there are 916 responding units and 75 nonresponding units. The initial design weight, denoted as DESIGNWT, is calculated as follows:

$$w_h = 1/\pi_h = N_h/n_h$$

N_h = Number of jurisdictions in stratum h of census
n_h = Number of jurisdictions sampled from stratum h

The initial design weight is then adjusted for nonresponse, and this final design weight is denoted as FINALWT – calculated as follows:

$$w_h = \frac{1}{\pi_h * \phi_h}$$

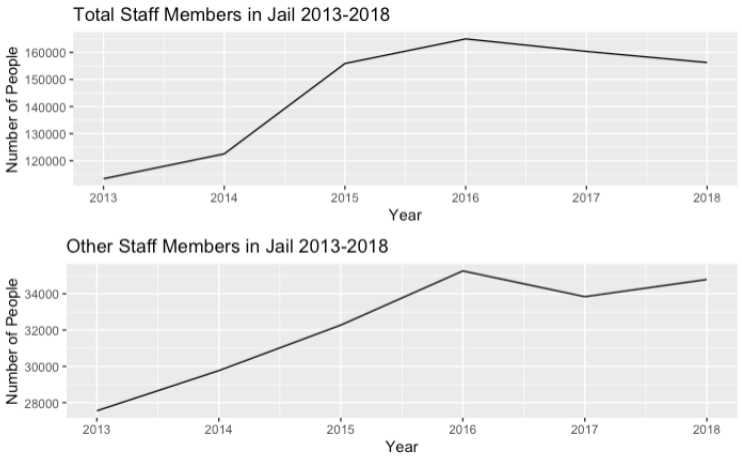
ϕ_h represents the sum of weights for respondents in stratum h over the sum of weights for sample in stratum h. The survey adjusts for nonresponse by assigning the design weight to the nonresponding jail jurisdictions based on their 2013 Census population and the sampling stratum. Within the same weighting class, nonresponse weights were adjusted to the responding jails weights. If the nonresponding jail jurisdictions are multi-reporting jail jurisdictions, the weights are calculated by applying a weighted sequential hot deck and cold deck. The final data consists of a total of 841 units in the jail jurisdictions.

The item response rate was slightly higher than the unit response rate, approximately around 95 to 100 percent. The missing values are decoded as -9 in the dataset. For nonresponse items, the survey used the SUDAAN software package to complete the missing values by employing weighted sequential hot deck and cold deck methods. The software package grouped the missing values into average daily population, and then randomly selected the missing values within their groups. The hot deck and cold deck method bases its data from census. The study further assumes that nonresponse is MAR (Missing at Random) and does not follow a specific model.

3.3 Exploratory Data Analysis

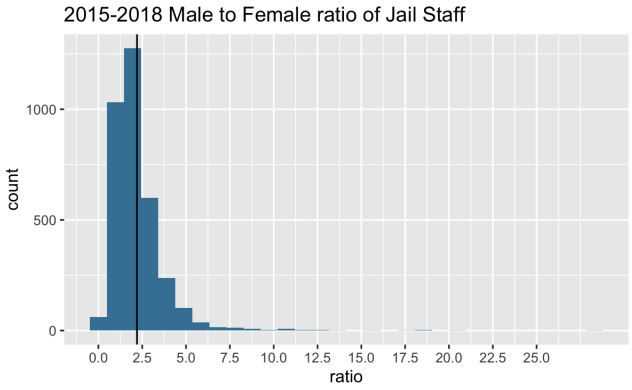
In order to understand the conditions of jail facilities for staff members, we first checked if there was a growing number of staff members in jail facilities. Figure 1 below depicts the overall trend of sampled jurisdictions through the five years. In 2013-2018, the number of staff members in jail facilities increased from years 2013 to 2016, and then decreased in 2017 and 2018. Similarly, the number of other staff members in jail facilities increased until 2016, and then decreased in 2017, with a slight increase in the last year.

Figure 1: Number of Staff Members in jail jurisdictions over the years 2013-2018



The data on assaults are not present in 2015-2018 data, and we will thus be studying assaults on staff members in jail facilities in 2013-2014. Throughout the two years, there has been a total of 2526 cases of assaults on correctional officers, 88 cases on assaults on other staff members, and 836 deaths on staff members by assaults reported in the Annual Survey of Jails. On average, each jail facility had 3.097 assaults on correctional officers, 0.137 assaults on other staff members, and 0.4819 deaths from assaults in facilities.

Figure 2: Gender Ratio in Staff Members



In EDA, we realized that staff data based on gender only exist for years 2015-2018, and so in our gender analysis we will only be working with 15-18 data. From Figure 2, it appears ratios are skewed to the right, centered around 2.2028 – there are about 2 male staff members per female staff.

Figure 3: Staff Distributions by Stratum

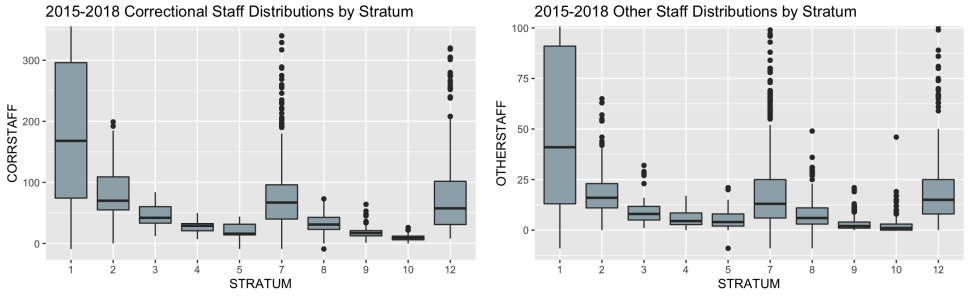


Figure 3 is a visualization of spread for number of correctional officers and other staff grouped by stratum. What we see is a a declining pattern, with stratum 1 having largest center and spread and then stratum 7 higher center than strata 8-10. This makes sense since if we look at the strata ranges in Table 1, we see that the first five as well as last five strata have declining ranges.

Figure 4: Inmate to Staff Ratio

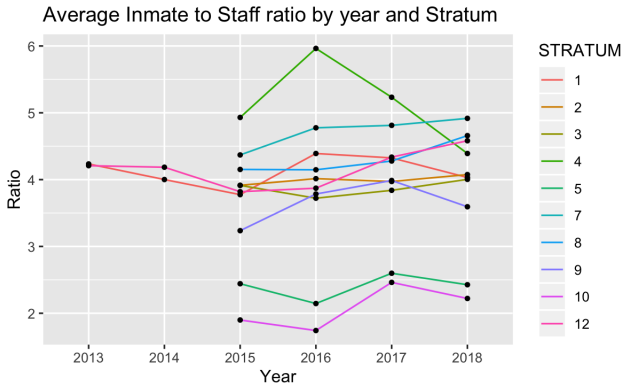


Figure 4 gives us a rough idea of where average ratios stand across years. From this plot, we can also see that only 2 strata have enough valid data points from 2013-2014 for inmate to staff ratios to be plotted.

4 Analysis and Findings

4.1 Total Number of Staff Members

We first wanted to analyze if the number of staff members in jail facilities has been increasing over time. As shown below in Table 3, the total number of staff members have been rising by approximately 100,000 members in five years. Despite the small decline in staff members in years 2017 and 2018, there has been an overall increasing trend in staff members.

Table 3: Total Staff Members in 2013-18

	2013	2014	2015	2016	2017	2018
Total	120274	131122	212489	226294	225690	221564
SE	13653	13688	12634	14067	14748	14500

4.2 Assaults on Staff Members

We then wanted to examine how jail conditions differ based on the average daily population of jails through the aspect of violence on staff members. Section 4.2.1 and 4.2.2 discusses our findings.

4.2.1

By taking a Chi-squared test of indepdence, we hoped to find an answer to the question: do certain types of staff members (ex. correctional officer) have more cases of assaults or deaths by assaults? When we conduct the test on two categorical variables, type of staff and presence of assault, we reach a p value less than 0.05. This result portrays that we can reject the null hypothesis that the two factors are independent, and conclude that there is evidence on the relationship between type of staff and presence of assault.

Table 4: Type of Staff v. Presence of Assault

	Correctional Officer	Other Staff
Assaults	2854	85808
No Assaults	150	27551

Pearson's Chi-squared test with Yates' continuity correction

data: chisquared table, X-squared = 600.58, df = 1, p-value < 2.2e-16

4.2.2

We then used another chi-squared test of independence to answer the question: do violence on staff members differ based on the average daily population of jails? As our p-value is again less than 0.05, we can conclude that there is an evidence relationship between the average daily population and the presence of assaults on staff members in facilities. Thus, violence on staff members are affected by the average daily population in jails.

Table 5: ADP v. Presence of Assault

	0-2000 (ADP)	2001-4000 (ADP)	4001- (ADP)
Assault	1470	878	598
No Assaults	60027	29406	20415

Pearson's Chi-squared test

data: chisquared table, X-squared = 26.216, df = 2, p-value = 2.029e-06

4.3 Male to Female Staff Ratio Analysis

The hypothesis of interest is shown below:

μ = Population mean male to female staff ratio

H_0: $\mu_{stratum1}$=$\mu_{stratum2}$=...=$\mu_{stratum12}$

H_1: At least one $\mu_{stratum1} \neq \mu_{stratum2}$

Data on the number of staff grouped by gender was only available for the years 2015-2018. Before conducting analysis on ratio by stratum, we first want to see if data across years are similar enough to merge for further analysis. To do so, we use ratio estimators to see the average Male/Female ratio by year as well as within year variance.

Table 6: Gender Ratio in Staff

	2015	2016	2017	2018
Ratio	1.893671	1.793516	1.745518	1.733858
SE	0.06548	0.04896	0.0487	0.05009

Figure 5

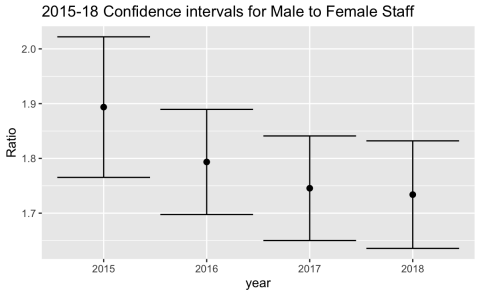


Table 7: ANOVA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
year	3	9	2.846	1.145	0.33
Residuals	3401	8456	2.486		

From Table 6 and Figure 5 above, we can see that there is not a huge difference between ratios between years. Concretely, fitting the ANOVA model with ratio year gives us a p-value of 0.33, which means there is no significant difference between ratio values between groups by year under $\alpha = 0.05$.

Table 8: ANOVA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Stratum	7	63	9.013	4.106	0.000173 ***
Residuals	1933	4243	2.195		

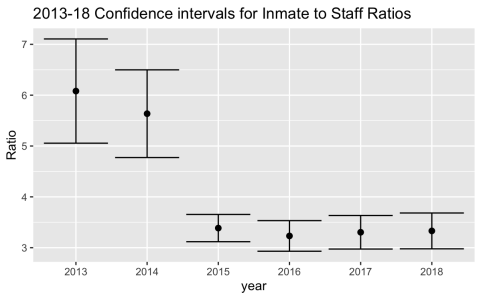
From ANOVA analysis of staff ratios grouped by stratum, there is significance pointing to the alternative hypothesis, which says at least one stratum has significantly different gender ratio of staff than another. In this analysis we dropped stratum 1 and 12 as our question of focus is differences in conditions for stratums by ADP.

4.4 Ratio of inmate to staff

Table 9: Inmate-to-Staff Ratio

	2013	2014	2015	2016	2017	2018
Ratio	6.0807	5.6357	3.3860	3.2316	3.3037	3.330
SE	0.5229	0.4399	0.1367	0.1535	0.1683	0.1797

Figure 6



From the statistics above and confidence intervals, it appears that years 2015-2018 have different inmate to staff ratios than 2013-2014, so it makes more sense to only merge 2015-2018 data for this portion of analysis.

In this portion, we test to see whether inmate to staff ratios remain the same across strata using a multiple pairwise t-test. This is important because in an ideal situation we'd want inmate to staff ratio to be the same everywhere – for safety reasons. Specifically, strata with a super high ADP should have the same ratio as a strata with super low ADP.

The assumption made in this test is that variances for different strata are not equal as different strata have different population sizes so naturally we cannot assume they're the same.

From t-test results, 16 pairs of strata are significantly different under $\alpha = 0.05$. This means that for many of these stratum, the ratios are not equal. Specifically, stratum 4, 7, and 8 have the highest ratios – with corresponding ADP (average daily population) ranges of 69-140, 227-749, and 103-226. To summarize, this means jails with higher ADP are relatively understaffed compared to those with lower ADP, and thus conditions for these larger jails are somewhat worse. The table below shows several values from t-test ran.

Table 10: ADP v. Presence of Assault

Group 1	Group 2	P.adj
2	5	3.96e-07
2	7	2.00e-03
2	10	1.03e-11

5 Conclusion

In our project, we analyzed the data from the Annual Survey of Jails to examine whether jail conditions are different based on the average daily population (ADP) in facilities. As the sampling design is a stratified one stage probability sampling method on jail jurisdictions, we studied the data across 10 strata based on ADP. To understand how jail conditions differ, we researched on four main areas: number of staff members, existing violence and assaults on staff, gender ratio and inmate-to-staff ratio.

Based on the results on total number of staff, we see that there has been an overall growing number of staff members over the years. We also observe that assaults on both correctional staff and other staff members varies based on the ADP of jails.

From ratio estimates of male to female staff in 2015-2018, as well as ANOVA analysis, it's evident that the distribution of male staff vs female staff members are not the same. Further, a pairwise t-test of the

inmate-to-staff ratios show that across strata, inmate-to-staff ratios are not equivalent. Jails with higher populations tend to be relatively understaffed than jails with fewer average daily populations.

In conclusion, the jail conditions which we compared across strata based on ADP are not the same. This could be problematic in the future as larger jails tend to have worse conditions.

5.1 Limitations

Across strata, the amount of nonresponse differed, and although we do account for nonresponse with weights, some strata calculations are relying more heavily on a smaller number of points than others and that could have effect on our results.

6 Reference

United States Department of Justice. Office of Justice Programs. Bureau of Justice Statistics. Annual Survey of Jails, 2018. Inter-university Consortium for Political and Social Research [distributor], 2020-04-23. https://doi.org/10.3886/ICPSR37392.v1

United States Deptartment of Justice. Office of Justice Programs. Bureau of Justice Statistics. ANNUAL SURVEY OF JAILS: JAIL-LEVEL DATA, 2018 [37392-0001-Codebook-ICPSR]. Conducted by RTI, International. ICPSR37392. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

```
####Setup###################
options(warn = -1)
library(readr)
library(plyr)
library(ggplot2)
library(survey)
#########################

#####Data cleaning and writing files#####
cols13_14 <- c("NAME","STRATUM", "OTHERSTAFF", "TOTALSTAFF", "ASSAULTCORR",
"ASSAULTOTRSTAFFF", "DEATHSTAFFYN", "ADPF", "ADP", "DESIGNWT", "FINALWT")

cols15_18 <- c("NAME","STRATUM", "TOTALSTAFF", "CORRSTAFF", "CORRSTAFFF",
"OTHERSTAFF", "OTHERSTAFFF","OTHERSTAFFM","TOTPOP", "ADPF", "ADP", "DESIGNWT",
"FINALWT")

clean <- function(file, cols) {
  dat <-read_tsv(file)
  dict <- c("FACLNAME"="NAME","STRATUMNEW" = "STRATUM",
"FACILITY"="NAME","S4QIV1C"="TOTALSTAFF", "BJSW"="FINALWT",
"WEIGHT"="DESIGNWT")
  dat <- rename(dat, dict)
  dat <- dat[, cols]
  dat$year <- rep(substring(file, 6, 9), nrow(dat))
  return (dat)
}

files <- list.files(path="Data", pattern="*.tsv", full.names=TRUE, recursive=FALSE)
files13_14 <- files[1:2]
files15_18 <- files[3:6]

data13_14 <- data.frame()
for (file in files13_14){
  dat <- clean(file, cols13_14)
  data13_14 <- rbind(data13_14, dat)
}
write.csv(dataset,"dataset.csv")
data15_18 <- data.frame()
for (file in files15_18){
  dat <- clean(file, cols15_18)
  data15_18 <- rbind(data15_18, dat)
}
```

```
cols <- intersect(names(data13_14), names(data15_18))
data13_18 <- rbind(data13_14[,cols], data15_18[,cols])
write.csv(data13_14, "data13_14.csv")
write.csv(data15_18, "data15_18.csv")
write.csv(data13_18, "data13_18.csv")

#####################################

############Exploratory Data Analysis Q3-Q4#######
library(dplyr)
##filling in total_staff male/female calculation
##invalid data = 0 staff members
data15_18$TOTALSTAFFF <- data15_18$CORRSTAFFF+data15_18$OTHERSTAFFF
data15_18$TOTALSTAFFM <- data15_18$TOTALSTAFF - data15_18$TOTALSTAFFF
data15_18$STRATUM <- as.factor(data15_18$STRATUM)

summary(data15_18$TOTALSTAFF)
summary(data15_18$CORRSTAFFF)
summary(data15_18$OTHERSTAFFF)

indices <- which(data15_18$TOTALSTAFF==0)

data15_18$ratio <- data15_18$TOTALSTAFFM/data15_18$TOTALSTAFFF
data15_18$ratio[data15_18$ratio<=0 | is.infinite(data15_18$ratio)] <- NA

ratio <- data15_18$ratio[is.na(data15_18$ratio)==F]
ratio <- ratio[is.infinite(ratio)==F]

ggplot()+
geom_histogram(aes(ratio), bins = 30, fill="#4281a4")+
  ggtitle("2015-2018 Male to Female ratio of Jail Staff")+
  geom_vline(aes(xintercept=mean(ratio)))+
  scale_x_continuous(breaks = seq(0,25, 2.5))

data15_18 %>% group_by(STRATUM) %>% summarise(ct = n(), avg=mean(ratio, na.rm = T))

sts <- boxplot.stats(data15_18$CORRSTAFF)$stats
## Correctional Staff
ggplot(data15_18, aes(x=STRATUM, y=CORRSTAFF)) +
          geom_boxplot(aes(group=STRATUM),fill="#9cafb7")+
          scale_x_discrete(breaks = seq(1,12,1))+
          coord_cartesian(ylim = c(sts*1.05,sts/1.05))+
  ggtitle("2015-2018 Correctional Staff Distributions by Stratum")
```

```
## Other Staff
sts <- boxplot.stats(data15_18$OTHERSTAFF)$stats
##Gender comparison per type
ggplot(data15_18, aes(x=STRATUM, y=OTHERSTAFF)) +
        geom_boxplot(aes(group=STRATUM), fill="#9cafb7")+
        scale_x_discrete(breaks = seq(1,12,1))+
        coord_cartesian(ylim = c(sts*1.05,sts/1.05))+
  ggtitle("2015-2018 Other Staff Distributions by Stratum")

data13_18$STRATUM <- sapply(data13_18$STRATUM, round)
data13_18$year <- as.factor(data13_18$year)
data13_18$STRATUM <- as.factor(data13_18$STRATUM)

##inmate to staff ratio
data13_18$ratios2 <- data13_18$ADP/data13_18$TOTALSTAFF
data13_18$ratios2[is.infinite(data13_18$ratios2) | data13_18$ratios2 <=0] <- NA

grpRatio <- data13_18 %>% group_by(year, STRATUM) %>% summarise(avg = mean(ratios2,
na.rm=T))
grpRatio <- grpRatio[is.na(grpRatio$avg)==F,]

ggplot(grpRatio, aes(year, avg))+
  geom_line(aes(group = STRATUM, color=STRATUM))+
  geom_point(size = 1)+
  ggtitle("Average Inmate to Staff ratio by year and Stratum")+
  xlab("Year")+
  ylab("Ratio")

data15_18$ratios3 <- data15_18$ADP/data15_18$CORRSTAFF
data15_18$ratios3[is.infinite(data15_18$ratios3) | data15_18$ratios3 <=0] <- NA

grpRatio <- data15_18 %>% group_by(year, STRATUM) %>% summarise(avg = mean(ratios3,
na.rm=T))
grpRatio <- grpRatio[is.na(grpRatio$avg)==F,]
grpRatio
#####################################

#######Functions for estimating ratios##############
ratio_fn <- function(dat) {
  ##Returns Ratio and SE
strat_design<-svydesign(ids = ~1, strata = ~STRATUM, weights=~FINALWT, data = dat)
r <- svyratio(~ADP, ~TOTALSTAFF, strat_design, na.rm=T)
```

```
c(unname(r$ratio)[[1]], sqrt(unname(r$var)[[1]]))
}

yrs <- as.character(2013:2018)

estimates <- data.frame()

for (i in 1:6) {
  yr <- yrs[i]
  dat <- data13_18[data13_18$year==yr,]
  estimates <- rbind(estimates, ratio_fn(dat))
}

colnames(estimates) <- c("Ratio", "SE")
estimates$year <- yrs

ggplot(estimates, aes(x = year, y = Ratio)) +
  geom_point(size = 2) +
  geom_errorbar(aes(ymax = Ratio+1.96*SE, ymin = Ratio-1.96*SE))+
  ggtitle("2013-18 Confidence intervals for Inmate to Staff Ratios")

estimates <- estimates[c("year","Ratio","SE")]
yrs <- as.character(2015:2018)
##Female to Male
estimates <- data.frame()

ratio_fn2 <- function(dat) {
  ##Returns Ratio and SE
strat_design<-svydesign(ids = ~1, strata = ~STRATUM, weights=~FINALWT, data = dat)
r <- svyratio(~TOTALSTAFFM, ~TOTALSTAFFF, strat_design, na.rm=T)
c(unname(r$ratio)[[1]], sqrt(unname(r$var)[[1]]))
}

for (i in 1:4) {
  yr <- yrs[i]
  dat <- data15_18[data15_18$year==yr,]
  estimates <- rbind(estimates, ratio_fn2(dat))
}

colnames(estimates) <- c("Ratio", "SE")
estimates$year <- yrs

#####################################
```

```
############Plot of CI Interval + ANOVA ##############

ggplot(estimates, aes(x = year, y = Ratio)) +
  geom_point(size = 2) +
  geom_errorbar(aes(ymax = Ratio+1.96*SE, ymin = Ratio-1.96*SE))+
  ggtitle("2015-18 Confidence intervals for Male to Female Staff")

estimates <- estimates[c("year","Ratio","SE")]
data15_18$year <- as.factor(data15_18$year)
summary(aov(ratio~year, data15_18))

##############################################

############ANOVA Testing###################

library(tidyverse)
library(ggpubr)
library(rstatix)

#Inmate to staff
dat <- data15_18[data15_18$STRATUM != 1 & data15_18$STRATUM != 2,]
summary(aov(ratio~STRATUM, dat))
inm_staff<-data13_18[data13_18$year != "2013" & data13_18$year != "2014",]
##########################################

############t-test#########################
t_test <- inm_staff %>%
pairwise_t_test(ratios2~STRATUM, pool.sd = F)
t_test <- t_test[t_test$p.adj < 0.05,]
t_test <- t_test[t_test$group1!=1 & t_test$group2 !=12,c("group1", "group2","p.adj")]
head(t_test)
```
```
##########################################

########################################
library(ggplot2)
library(dplyr)
library(gridExtra)
library(survey)
########################################
################Loading Data#############
```

```
fifteen_eighteen_data <- read.csv("data15_18.csv")
thirteen_fourteen_data <- read.csv("data13_14.csv")
total_data <- read.csv("data13_18.csv")

################Exploratory Analysis for Q1-2#################
staff_data <- total_data %>% select(year, TOTALSTAFF, OTHERSTAFF) %>% group_by(year) %>%
summarise_all(funs(sum))
staff_data
#reference https://stackoverflow.com/questions/1249548/side-by-side-plots-with-ggplot2
a <- ggplot(data=staff_data)+geom_line(aes(x = year, y = TOTALSTAFF))+xlab("Year")+ylab("Number
of People")+ggtitle("Total Staff Members in Jail 2013-2018")
b <- ggplot(data=staff_data)+geom_line(aes(x = year, y = OTHERSTAFF))+xlab("Year")+ylab("Number
of People")+ggtitle("Other Staff Members in Jail 2013-2018")
grid.arrange(a, b)
assault_data <- thirteen_fourteen_data %>% select(ASSAULTCORR, ASSAULTOTRSTAFFF,
DEATHSTAFFYN)
assault_data$DEATHSTAFFYN[assault_data$DEATHSTAFFYN==-9] <- 0
summary(assault_data)
assault_data <- thirteen_fourteen_data %>% select(ASSAULTCORR, ASSAULTOTRSTAFFF,
DEATHSTAFFYN)
assault_data$DEATHSTAFFYN[assault_data$DEATHSTAFFYN==-9] <- 0
assault_data <- assault_data %>% select(ASSAULTCORR, ASSAULTOTRSTAFFF,
DEATHSTAFFYN) %>% summarise_all(sum)
assault_count <- data.frame("Type"=c("Assault on Corr Officers", "Assault on Other", "Deaths by
Assaults"), "Count"=c(2526, 88, 836))
assault_count
ggplot(data=assault_count)+geom_col(aes(x=Type, y=Count))+xlab("Assault Type")+ylab("Number of
Cases")+ggtitle("Assaults and Deaths of Staff Members in Jail Facilities in 2013-14")
################

#Individual Years
thirteen_data <- total_data %>% filter(year==2013)
fourteen_data <- total_data %>% filter(year==2014)
fifteen_data <- total_data %>% filter(year==2015)
sixteen_data <- total_data %>% filter(year==2016)
seventeen_data <- total_data %>% filter(year==2017)
eighteen_data <- total_data %>% filter(year==2018)

#Total staff values using survey
survey_design13 <- svydesign(data=thirteen_data, ids=~1, weights=thirteen_data$FINALWT,
strata=thirteen_data$STRATUM)
svytotal(thirteen_data$TOTALSTAFF,design=survey_design13)
```

```
survey_design14 <- svydesign(data=fourteen_data, ids=~1, weights=fourteen_data$FINALWT,
strata=fourteen_data$STRATUM)
svytotal(fourteen_data$TOTALSTAFF,design=survey_design14)
survey_design15 <- svydesign(data=fifteen_data, ids=~1, weights=fifteen_data$FINALWT,
strata=fifteen_data$STRATUM)
svytotal(fifteen_data$TOTALSTAFF,design=survey_design15)
survey_design16 <- svydesign(data=sixteen_data, ids=~1, weights=sixteen_data$FINALWT,
strata=sixteen_data$STRATUM)
svytotal(sixteen_data$TOTALSTAFF,design=survey_design16)
survey_design17 <- svydesign(data=seventeen_data, ids=~1, weights=seventeen_data$FINALWT,
strata=seventeen_data$STRATUM)
svytotal(seventeen_data$TOTALSTAFF,design=survey_design17)
survey_design18 <- svydesign(data=eighteen_data, ids=~1, weights=eighteen_data$FINALWT,
strata=eighteen_data$STRATUM)
svytotal(eighteen_data$TOTALSTAFF,design=survey_design18)

#Assault Analysis for years 13-14
thirteendata <- thirteen_fourteen_data %>% filter(year==2013)
fourteendata <- thirteen_fourteen_data %>% filter(year==2014)
thirteensurvey <- svydesign(data=thirteendata, ids=~1, weights=thirteendata$FINALWT,
strata=thirteendata$STRATUM)
svytotal(~ASSAULTCORR, design=thirteensurvey)
svytotal(~ASSAULTOTRSTAFFF, design=thirteensurvey)
fourteensurvey <- svydesign(data=fourteendata, ids=~1, weights=fourteendata$FINALWT,
strata=fourteendata$STRATUM)
svytotal(~ASSAULTCORR, design=fourteensurvey)
svytotal(~ASSAULTOTRSTAFFF, design=fourteensurvey)

#################Chisq Test of Independence################
sum(thirteendata$TOTALSTAFF-thirteendata$OTHERSTAFF)
sum(thirteendata$OTHERSTAFF)
sum(thirteendata$ASSAULTCORR)
sum(thirteendata$ASSAULTOTRSTAFFF)
correctional_staff <- c(2854, 85808)
other_staff <- c(150, 27551)
chisquared_table <- data.frame(rbind(correctional_staff, other_staff))
names(chisquared_table) <- c("assault", "no_assault")
chisquared_table
chisq.test(chisquared_table)

#################Chisq Test of Independence################
emptyvector <- c()
emptyvector[thirteendata$ADP < 2000] <- "one"
```

```
emptyvector[thirteendata$ADP > 2001 & thirteendata$ADP < 4001] <- "two"
emptyvector[thirteendata$ADP > 4001] <- "three"
thirteendata$category <- emptyvector
gettingvals1 <- thirteendata %>% select(category, ASSAULTCORR, ASSAULTOTRSTAFFF) %>%
group_by(category) %>% summarise(sum=sum(ASSAULTCORR),
sum2=sum(ASSAULTOTRSTAFFF))
thirteendata$noassault <- thirteendata$TOTALSTAFF - thirteendata$ASSAULTCORR -
thirteendata$ASSAULTOTRSTAFFF
gettingvals2 <- thirteendata %>% select(noassault, category) %>% group_by(category) %>%
summarise(sum=sum(noassault))
assault <- c(1470, 878, 598)
no_assault <- c(60027, 29406, 20415)
chisquared_table <- data.frame(rbind(assault, no_assault))
names(chisquared_table) <- c("one", "two", "three")
chisquared_table
chisq.test(chisquared_table)
################################
```