Assignment 2, Social Science Inquiry II (SOSC13200-W23-2)

Tiffanie Huang

Friday 1/13/22 at 5pm

Packages

```
library(ggplot2)
```

Read in the data.

```
file <- "https://raw.githubusercontent.com/UChicago-pol-methods/SOSC13200-W23/main/data/card-krueger.csv
dat <- read.csv(file, as.is = TRUE)
```

1. Reproduce the reported means from table 2 of the Card and Krueger paper, for 1a-e, 2a, and 3a.

You do not need to reproduce the test of equality of means in the far right column, or the standard errors in parentheses.

```
#1
#a. Burger King
#round(nrow(dat[which(dat$d==1&dat$bk==1&dat$nj==1),])/nrow(dat[which(dat$d==1&dat$nj==1),])*100,1)
round(mean(dat$bk[which(dat$nj==1 & dat$d==0)])*100,1) #nj = 41.1%
```

```
## [1] 41.1
```

```
round(mean(dat$bk[which(dat$nj==0 & dat$d==0)])*100,1) #pa = 44.3%
```

```
## [1] 44.3
```

```
#b. KFC
round(mean(dat$kfc[which(dat$nj==1 & dat$d==0)])*100,1) #nj = 20.5%
```

```
## [1] 20.5
```

```
round(mean(dat$kfc[which(dat$nj==0 & dat$d==0)])*100,1) #pa = 15.2%
```

```
## [1] 15.2
```

```
#c. Roy Rogers
round(mean(dat$roys[which(dat$nj==1 & dat$d==0)])*100,1) #nj = 24.8%
```

```
## [1] 24.8
```

```
round(mean(dat$roys[which(dat$nj==0 & dat$d==0)])*100,1) #pa = 21.5%
```

```
## [1] 21.5
```

```
#d. Wendys
round(mean(dat$wendys[which(dat$nj==1 & dat$d==0)])*100,1) #nj = 13.6%
```

```
## [1] 13.6
```

```
round(mean(dat$wendys[which(dat$nj==0 & dat$d==0)])*100,1) #pa = 19.0%
```

```
## [1] 19
```

```
#e. Company-owned
round(mean(dat$co_owned[which(dat$nj==1 & dat$d==0)])*100,1) #nj = 34.1%
```

```
## [1] 34.1
```

```
round(mean(dat$co_owned[which(dat$nj==0 & dat$d==0)])*100,1) #pa = 35.4%
```

```
## [1] 35.4
```

```
#2a. Wave 1
round(mean(dat$fte[which(dat$nj==1 & dat$d==0)],na.rm=TRUE),1) #nj = 20.4%
```

```
## [1] 20.4
```

```
round(mean(dat$fte[which(dat$nj==0 & dat$d==0)],na.rm=TRUE),1) #pa = 23.3%
```

```
## [1] 23.3
```

```
#3a. Wave 2
round(mean(dat$fte[which(dat$nj==1 & dat$d==1)],na.rm=TRUE),1) #nj = 20.4%
```

```
## [1] 21
```

```
round(mean(dat$fte[which(dat$nj==0 & dat$d==1)],na.rm=TRUE),1) #pa = 21.2%
```

```
## [1] 21.2
```

2a. Make separate histograms showing the number of part time employees in each state, in the first wave only. Label your plots.

```
library(ggplot2)

mean(diff(sort(unique(dat$pt))))
```

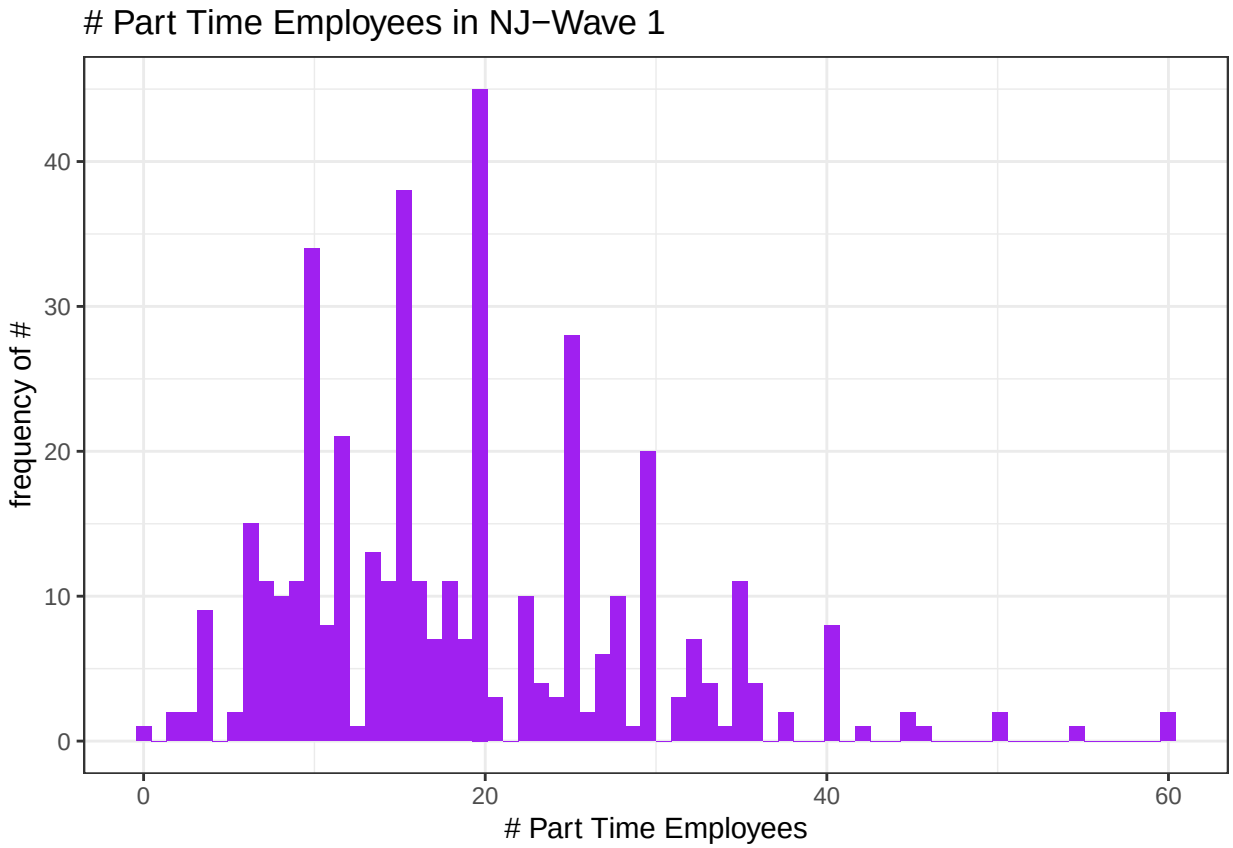
```
## [1] 0.8823529
```

```
diff(range(dat$pt, na.rm=TRUE))/0.8823529
```

```
## [1] 68
```

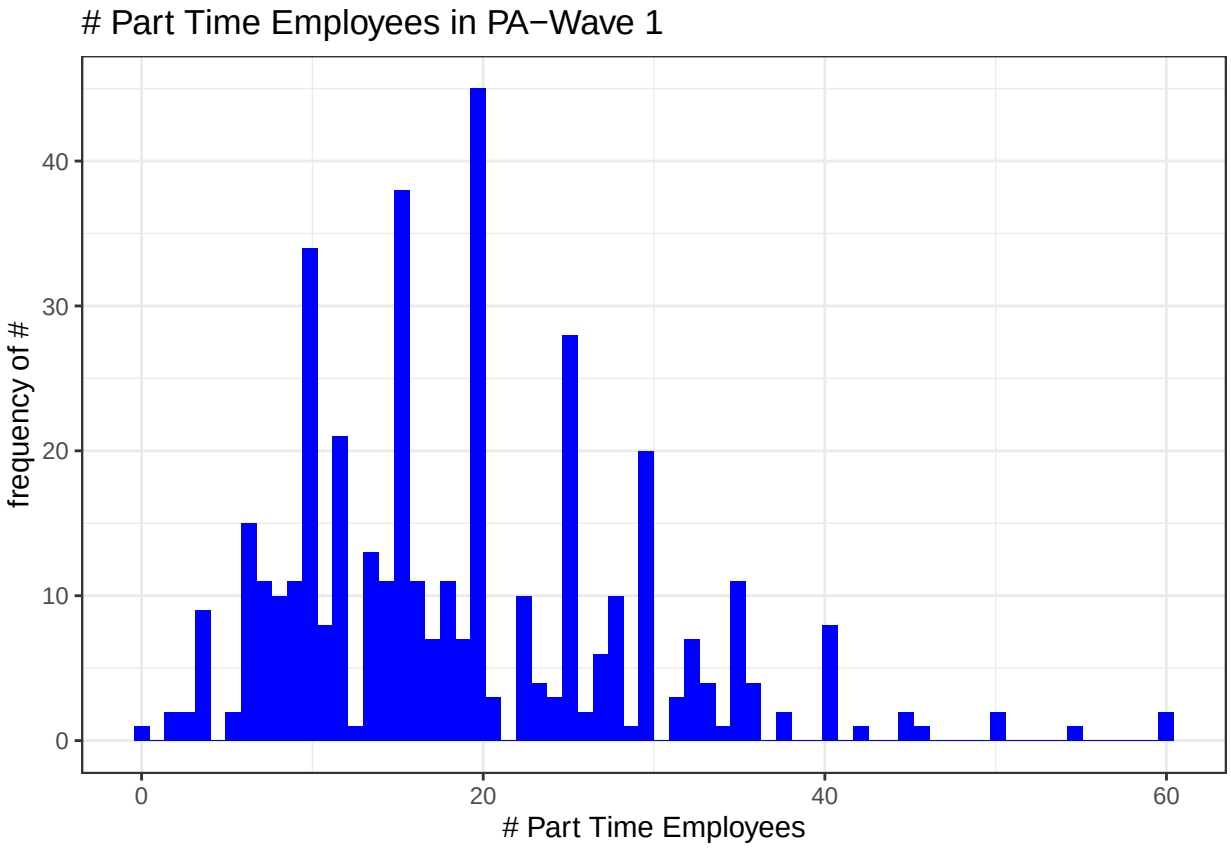
```
#New Jersey PTE - Wave 1
ggplot(dat[which(dat$nj==1) & dat$d==0,],aes(x=pt)) +
  geom_histogram(bins=68, fill="purple", na.rm=TRUE) +
  xlab("# Part Time Employees") +
  ylab("frequency of #") +
  theme_bw() +
  ggtitle("# Part Time Employees in NJ-Wave 1")
```

```
## Warning in which(dat$nj == 1) & dat$d == 0: longer object length is not a
## multiple of shorter object length
```



```
#Pennsylvania PTE - Wave 1
ggplot(dat[which(dat$nj==0) & dat$d==0,],aes(x=pt)) +
  geom_histogram(bins=68, fill="blue", na.rm=TRUE) +
  xlab("# Part Time Employees") +
  ylab("frequency of #") +
  theme_bw() +
  ggtitle("# Part Time Employees in PA-Wave 1")
```

```
## Warning in which(dat$nj == 0) & dat$d == 0: longer object length is not a
## multiple of shorter object length
```



2b. Using `facet_wrap()`, make the same figure for each state and both waves in the same plot.

```
#encode a factor: factor(vector name, labels=c("label names"))
```

```
#New Jersey and Pennsylvania PTE - Waves 1 and 2
dat$State <- factor(dat$nj,labels=c("PA","NJ"))
dat$Waves <- factor(dat$d+1,labels=c("Wave 1","Wave 2"))
state.lab <- c(`0`="PA",`1`="NJ")
wave <- c(`0`="Wave 1",`1`="Wave 2")
```

```
ggplot(dat,aes(x=pt,fill=State)) +
  scale_fill_manual(values=c("turquoise4", "rosybrown1")) +
  geom_histogram(bins=68) +
  facet_wrap(vars(nj,d),
             labeller=labeller(nj=state.lab,d=wave)) +
  xlab("# Part Time Employees") +
  ylab("frequency of #") +
  theme_bw() +
  ggtitle("# Part Time Employees in PA-Waves 1+2")
```

```
## Warning: Removed 14 rows containing non-finite values (`stat_bin()`).
```

3. Using `geom_boxplot()`, create a box and whiskers plot of the distribution of full time employees.

Include wave as a secondary aesthetic, and state as color, so that you should have two paired plots for each wave.

```
ggplot(dat, #data
       aes(x=ft,y=Waves,fill=State)) +
  geom_boxplot(na.rm=TRUE) +
  scale_fill_manual(values=c("turquoise4", "rosybrown1")) +
  xlab("# Full Time Employees") +
```

```
  ylab("Wave") +
  theme_bw() +
  ggtitle("Full Time Employees for Waves 1 and 2")
```