Assignment 2, Social Science Inquiry II (SOSC13200-W22-3)

Friday 1/21/22 at 5pm

Packages

```
library(ggplot2)
```

Read in the data.

```
file <- "https://raw.githubusercontent.com/UChicago-pol-methods/SOSC13200-W22/main/data/card-krueger.cs
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

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

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

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

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

1.b

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

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

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

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

1.c

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

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

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

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

1.d

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

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

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

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

1.e

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

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

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

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

2.a

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

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

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

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

3.a

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

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

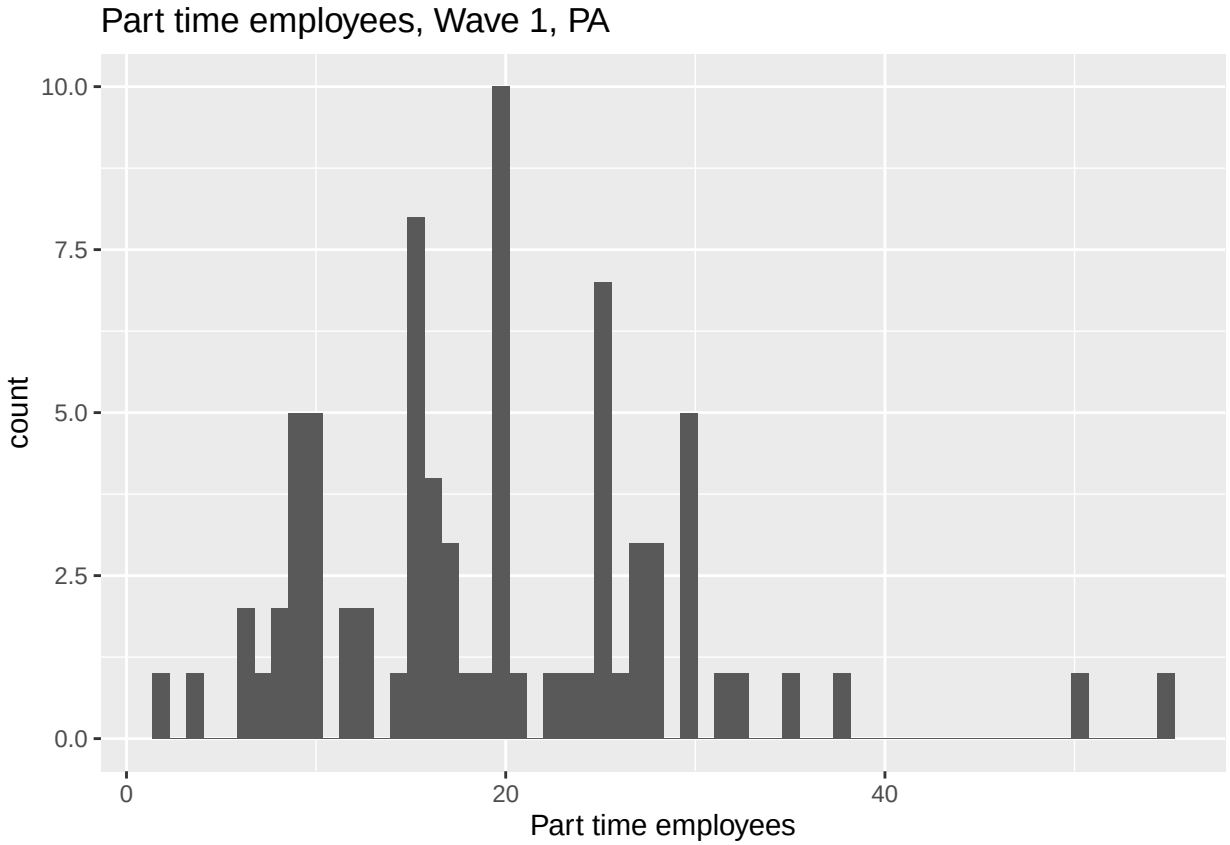
```
# PA
round(mean(dat$fte[which(dat$nj==0 & dat$d == 1)], na.rm = TRUE),1)
```

```
## [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.

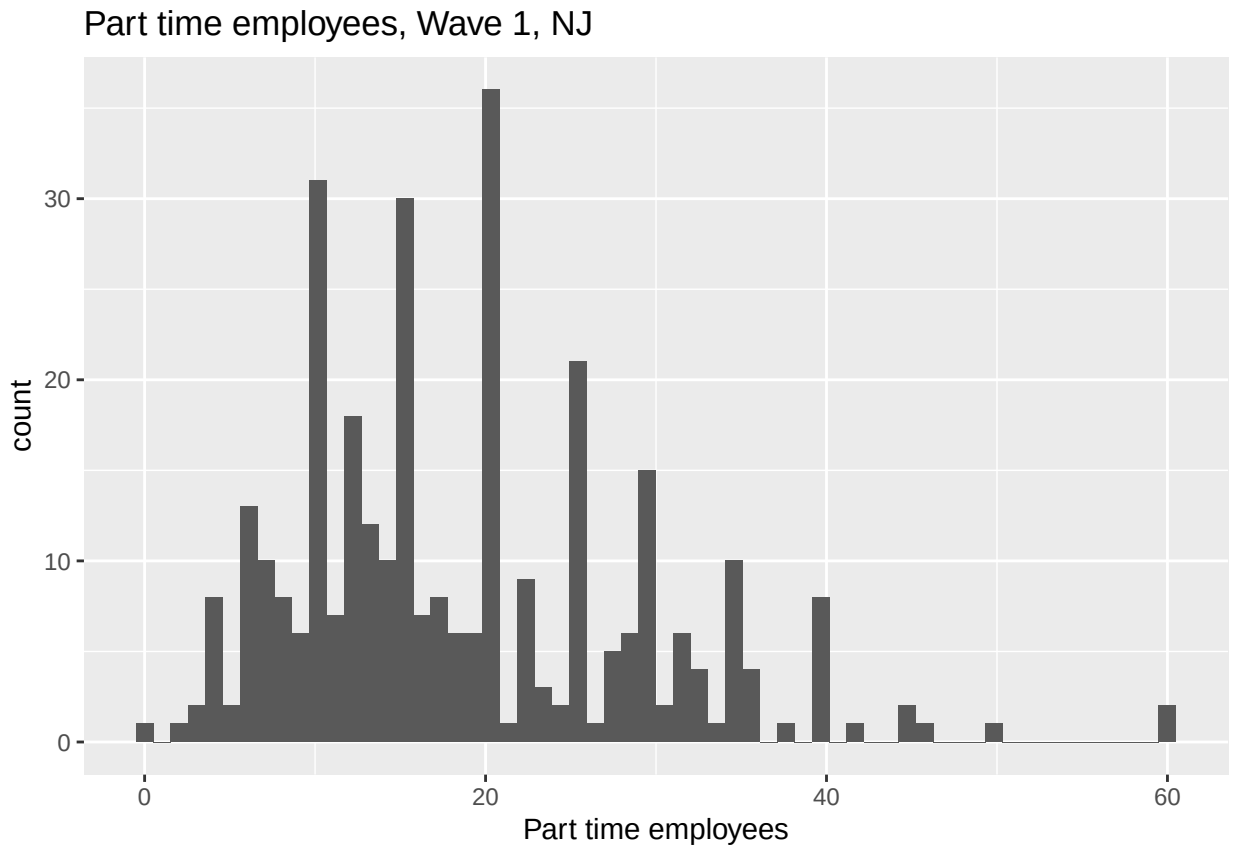
```
ggplot(dat[which(dat$d == 0 & dat$nj == 0),], aes(x = pt)) +
  geom_histogram(bins = max(dat$pt, na.rm = TRUE)) +
  xlab('Part time employees') +
  ggtitle('Part time employees, Wave 1, PA')
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```




```
ggplot(dat[which(dat$d == 0 & dat$nj == 1),], aes(x = pt)) +
  geom_histogram(bins = max(dat$pt, na.rm = TRUE)) +
  xlab('Part time employees') +
  ggtitle('Part time employees, Wave 1, NJ')
```

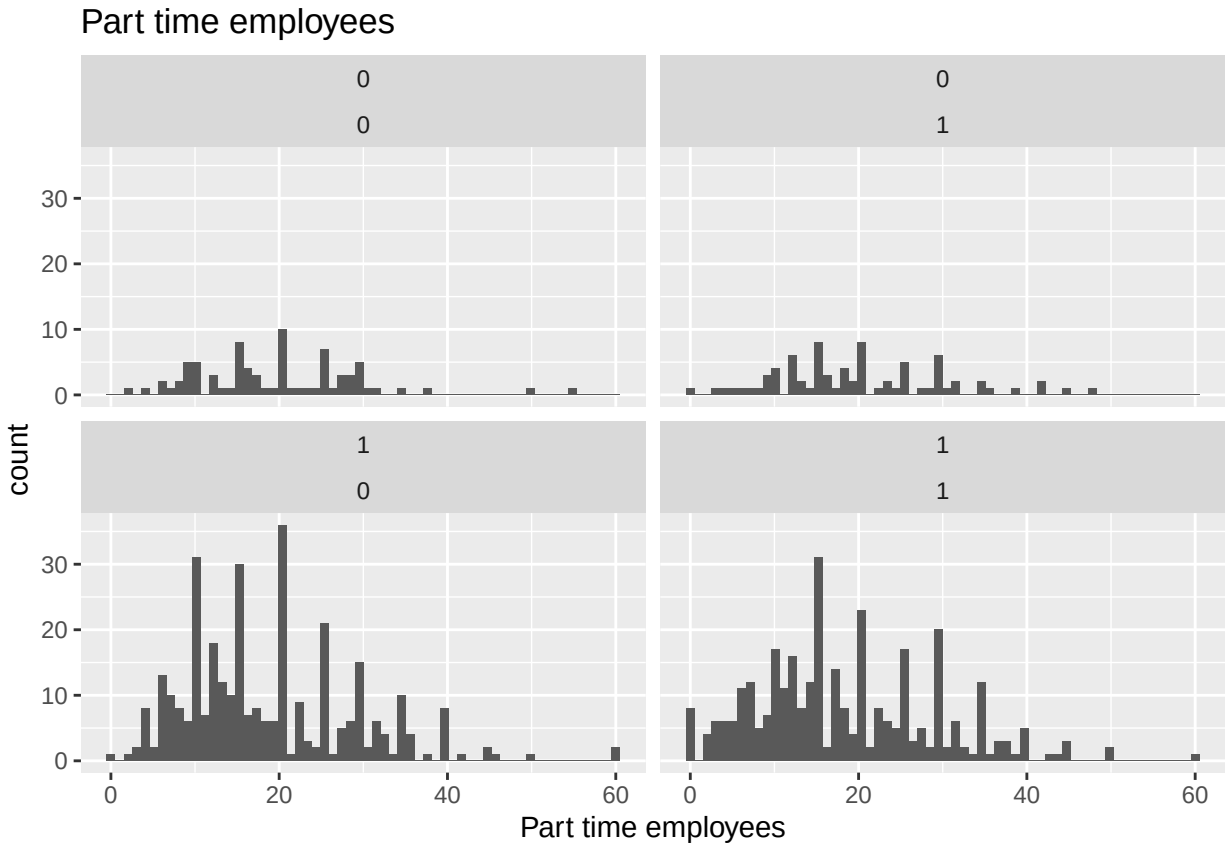
```
## Warning: Removed 3 rows containing non-finite values (stat_bin).
```



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

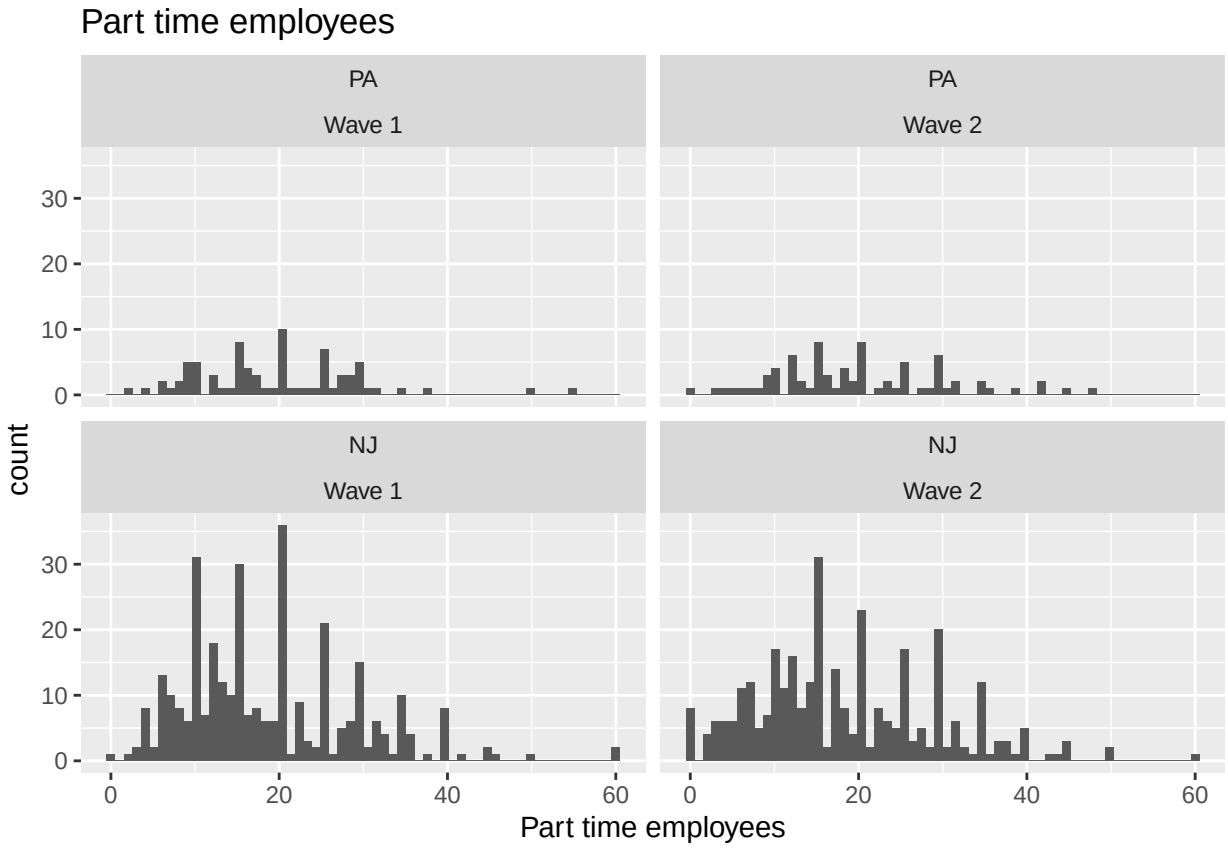
```
ggplot(dat, aes(x = pt)) +
  geom_histogram(bins = max(dat$pt, na.rm = TRUE)) +
  facet_wrap(vars(nj, d)) +
  xlab('Part time employees') +
  ggtitle('Part time employees')
```

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



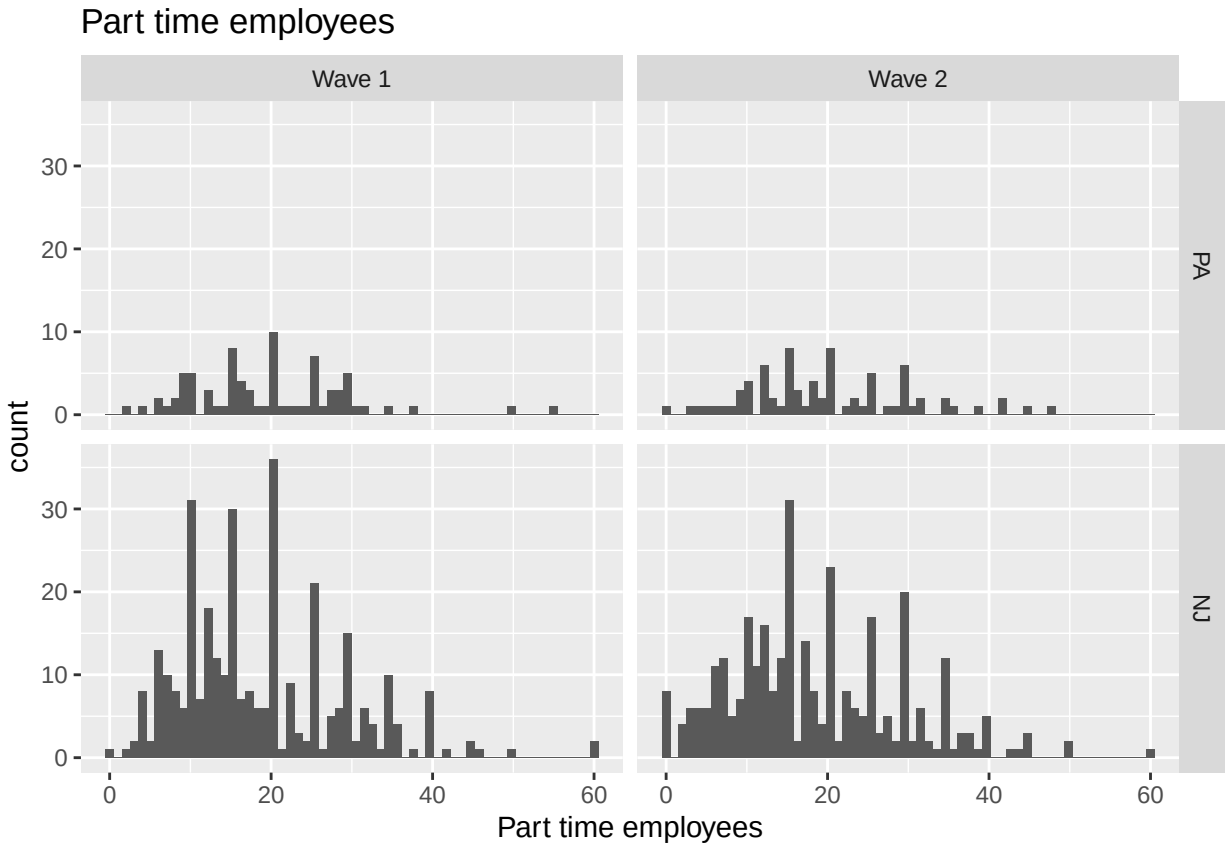

```
# (To have the plots correctly labeled)
ggplot(dat, aes(x = pt)) +
  geom_histogram(bins = max(dat$pt, na.rm = TRUE)) +
  facet_wrap(vars(nj,d),
             labeller = labeller(nj = c(`0` = 'PA', `1` = 'NJ'),
                                 d = c(`0` = 'Wave 1', `1` = 'Wave 2'))) +
  xlab('Part time employees') +
  ggtitle('Part time employees')
```

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




```
# OR
ggplot(dat, aes(x = pt)) +
  geom_histogram(bins = max(dat$pt, na.rm = TRUE)) +
  facet_grid(rows = vars(nj), cols = vars(d),
             labeller = labeller(.rows = c(`0` = 'PA', `1` = 'NJ'),
                                 .cols = c(`0` = 'Wave 1', `1` = 'Wave 2'))) +
  xlab('Part time employees') +
  ggtitle('Part time employees')
```

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

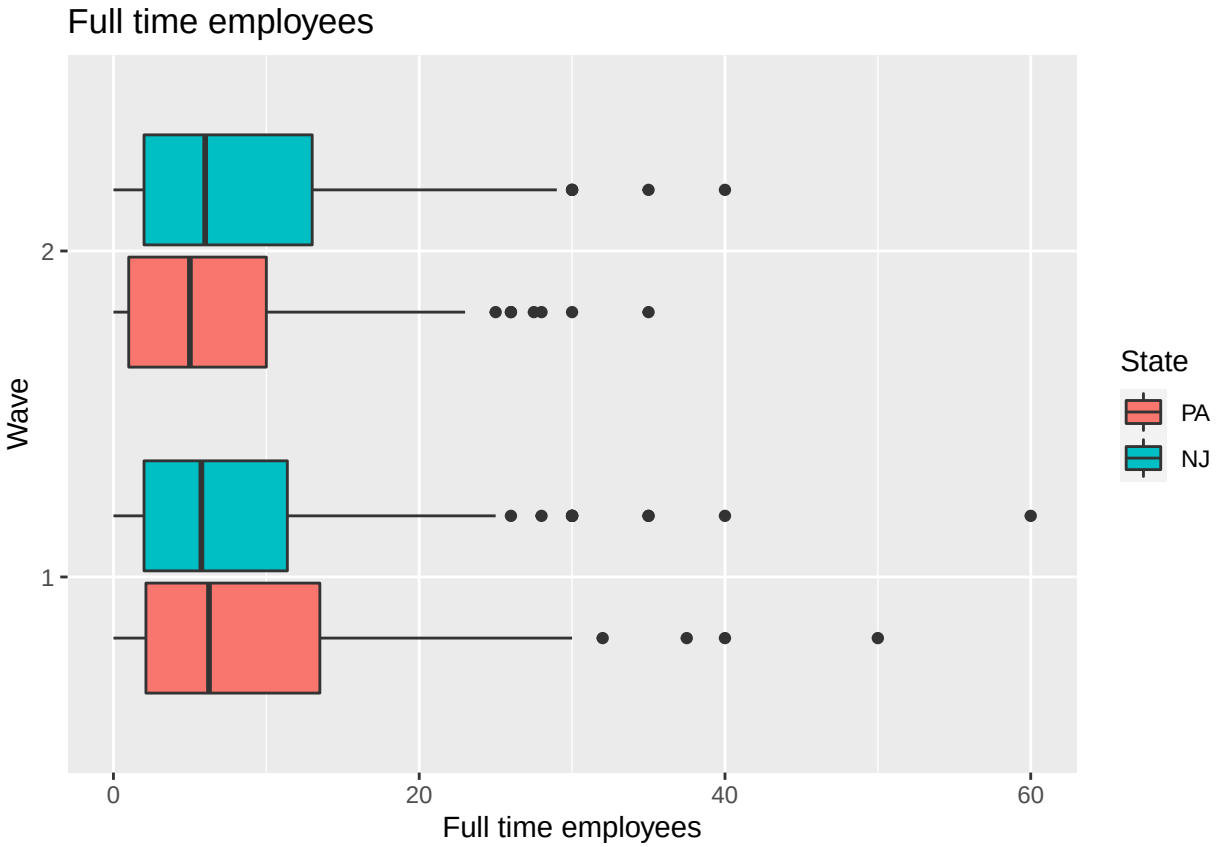


3. Using `geom_boxplot()`, create a box and whiskers plot of the distribution of full time employees.

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

```
ggplot(dat, aes(x=ft, y=as.factor(d+1), fill=as.factor(nj))) +
  geom_boxplot() +
  scale_fill_discrete(name = 'State', labels = c('PA', 'NJ')) +
  xlab('Full time employees') +
  ylab('Wave') +
  ggtitle('Full time employees')
```

```
## Warning: Removed 18 rows containing non-finite values (stat_boxplot).
```


Full time employees
Wave
2
1
0
20
40
60
Full time employees
State
PA
NJ