

Assignment 2, Social Science Inquiry II (SOSC13200-W26-3)

Friday 1/16/26 at 11:59pm

Packages

```
library(ggplot2)
```

Read in the data.

```
file <- "https://raw.githubusercontent.com/UChicago-pol-methods/SOSC13200-W26/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. Use `data.frame()`. Format results in a table.

```
# Your code here.
```

2a. Make separate histograms showing the number of part time employees in each state, in the first wave only. Label the title, x-axis, and y-axis for your plots.

```
# Your code here.
```

2b. Using `facet_wrap()`, make the same figure for each state and both waves in the same plot. Label the title, x-axis, and y-axis for your plot.

```
# Your code here.
```

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. Again, make sure axes are labelled.

```
# Your code here.
```

4. Data audit + missingness map

Create a data audit table with one row per variable containing: variable name, type, % missing, number of unique values, and min/max (numeric only). For variables that have missingness, show how this varies between wave 1 and wave 2.

```
# Your code here.
```

5. Sample composition: chain and region (reshape required)

Subset to only wave 1 data. Create a “chain” variable from the indicator columns (bk, kfc, roys, wendys). Create a “region” variable from the location flags (centralj, southj, pa1, pa2).

Plot A: stacked bar chart of chain by state. Plot B: stacked bar chart of chain by region.

6. Short DGP memo (100-150 words)

Describe the unit of observation, sampling frame, and time structure. Explain how treatment status relates to the data. Note two measurement limitations you observed from your audit/plots.