Exploratory Data Analysis

```
library(tidyverse)
```

```
## Registered S3 methods overwritten by 'ggplot2':
##   method         from
##   [.quosures     rlang
##   c.quosures     rlang
##   print.quosures rlang
```

```
## -- Attaching packages ---------------------------------------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.1.1     v purrr   0.3.2
## v tibble  2.1.1     v dplyr   0.8.1
## v tidyr   0.8.3     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(here)
```

```
## here() starts at /Users/denis/Dev/causal-forest-ohie
```

```
library(assertthat)
```

```
##
## Attaching package: 'assertthat'
```

```
## The following object is masked from 'package:tibble':
##
##     has_name
```

Assuming independence between respondents, the standard errors of ATE are given by:

$$se(\bar{y}_1 - \bar{y}_0) = \sqrt{\frac{1}{n_0} var(y_i|d_i = 0) + \frac{1}{n_1} var(y_i|d_i = 1)}$$

```
read_stata_dataset <- function(file_path) {
  file_path %>%
    here::here() %>%
    haven::read_dta(file_path) %>%
    haven::as_factor()
}

descriptive <- read_stata_dataset("data/OHIE_Public_Use_Files/OHIE_Data/oregonhie_descriptive_vars.dta")
state_programs <- read_stata_dataset("data/OHIE_Public_Use_Files/OHIE_Data/oregonhie_stateprograms_vars
survey12 <- read_stata_dataset("data/OHIE_Public_Use_Files/OHIE_Data/oregonhie_survey12m_vars.dta")
emergency <- read_stata_dataset("data/OHIE_Public_Use_Files/OHIE_Data/oregonhie_ed_vars.dta")
```

Average Treatment Effect of Being Picked at the Lottery

Heterogeneity here?

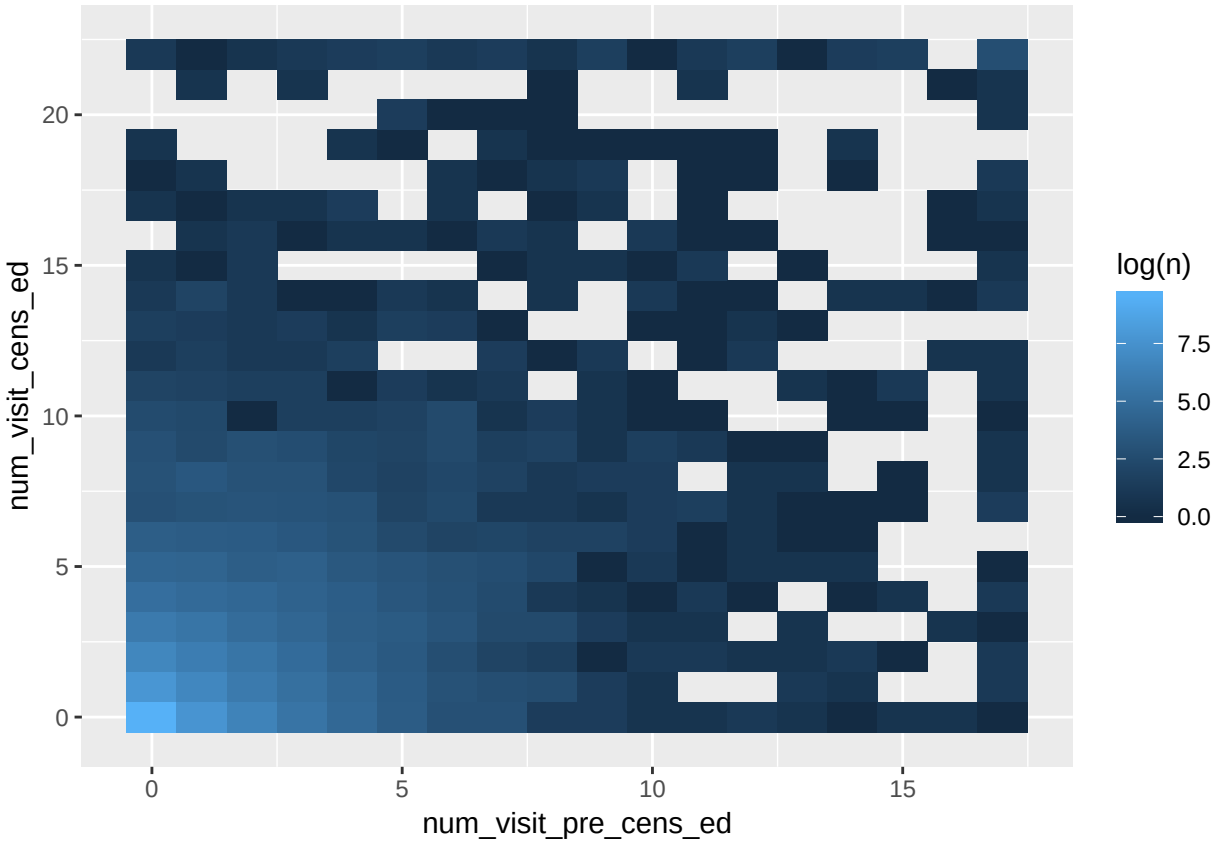
Average Treatment Effect of Medicaid enrollment

`ohp_all_ever_firstn_30sep2009`: This variable takes a value of 1 if an individual was enrolled in any Medicaid program (including the lotteried program, OHP Standard) between the earliest notification date in the sample (10 March 2008) and 30 September 2009. In the analysis of the 12-month mail survey data in Finkelstein et al (2012), this variable was used as the definition of insurance coverage in estimating the effect of Medicaid. In the analysis in Taubman et al (201XX), this variable was used as the definition of insurance coverage in estimating the effect of Medicaid.

```
medicaid <- state_programs %>%
  select(person_id, ohp_all_ever_firstn_30sep2009)
```

```
emergency %>%
  count(num_visit_pre_cens_ed, num_visit_cens_ed) %>%
  filter() %>%
  ggplot(aes(num_visit_pre_cens_ed, num_visit_cens_ed, fill = log(n))) +
  geom_tile()
```

```
## Warning: Removed 13 rows containing missing values (geom_tile).
```



Local Average Treatment Effect per Previous Visit

```
per_n_visits <- emergency %>%
  left_join(descriptive) %>%
  mutate(n_visits_pre = case_when(
```

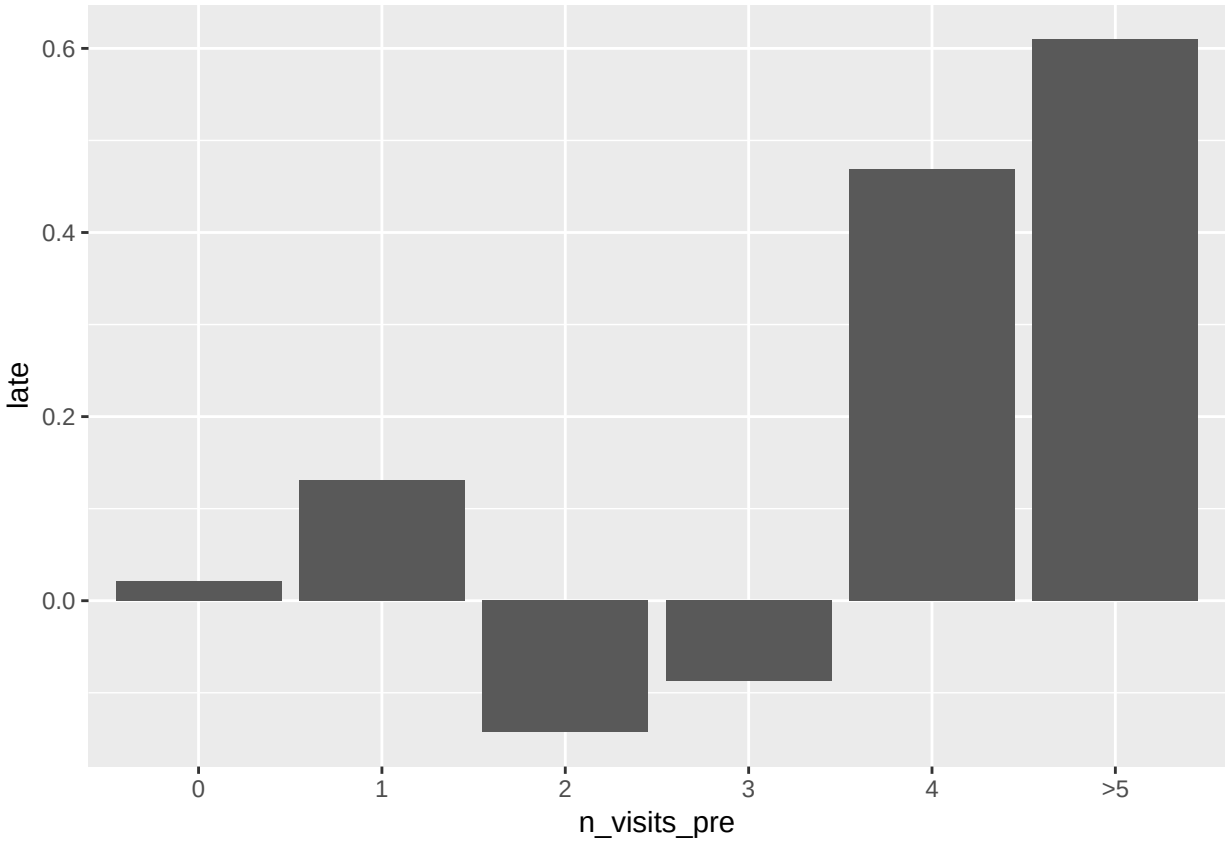
```
    num_visit_pre_cens_ed == 0 ~ "0",
    num_visit_pre_cens_ed == 1 ~ "1",
    num_visit_pre_cens_ed == 2 ~ "2",
    num_visit_pre_cens_ed == 3 ~ "3",
    num_visit_pre_cens_ed == 4 ~ "4",
    num_visit_pre_cens_ed >= 5 ~ ">5",
    is.na(num_visit_pre_cens_ed) ~ NA_character_,
    TRUE ~ "Else"
  )) %>%
  mutate(n_visits_pre = forcats::fct_relevel(n_visits_pre, "0", "1", "2", "3", "4", ">5"))
```

```
## Joining, by = "person_id"
```

```
# Make sure code logic is covering all cases
# assert_that(all(per_n_visits$n_visits_pre != "Else"))

# Now remove NAs
per_n_visits <- per_n_visits %>%
  filter(!is.na(n_visits_pre))

per_n_visits <- per_n_visits %>%
  group_by(n_visits_pre,
           treatment) %>%
  summarise(avg_visit = mean(num_visit_cens_ed, na.rm = TRUE),
            n = n()) %>%
  group_by(n_visits_pre) %>%
  mutate(total_n = sum(n),
         pp = n / total_n) %>%
  mutate(pp_treated = pp[treatment == "Selected"])

per_n_visits
```

```
## # A tibble: 12 x 7
## # Groups:   n_visits_pre [6]
##    n_visits_pre treatment    avg_visit     n total_n    pp pp_treated
##    <fct>        <fct>            <dbl> <int>   <int> <dbl>      <dbl>
##  1 0            Not selected     0.418 10213   16930 0.603      0.397
##  2 0            Selected         0.439  6717   16930 0.397      0.397
##  3 1            Not selected     1.11   2380    3881 0.613      0.387
##  4 1            Selected         1.25   1501    3881 0.387      0.387
##  5 2            Not selected     1.86   1001    1594 0.628      0.372
##  6 2            Selected         1.71    593    1594 0.372      0.372
##  7 3            Not selected     2.65    523     839 0.623      0.377
##  8 3            Selected         2.56    316     839 0.377      0.377
##  9 4            Not selected     3.12    282     445 0.634      0.366
## 10 4            Selected         3.59    163     445 0.366      0.366
## 11 >5           Not selected     6.44    613     945 0.649      0.351
## 12 >5           Selected         7.05    332     945 0.351      0.351
```

```
per_n_visits %>%
  select(-n, -pp) %>%
  spread(key = treatment, value = avg_visit) %>%
  mutate(late = Selected - `Not selected`) %>%
  ggplot(aes(n_visits_pre, late)) +
  geom_col()
```




```
knitr::knit_exit()
```