Causal Inference Project Proposal

Overview of Data

All Data is from the NYC Open Data Website

```
######
# STEP 1
#load PS43 and PS 001 data
#######

library(tidyverse)
setwd("/cloud/project/scripts")
PS001_attendance <- read.csv("../data/PS001_attendance.csv")
PS001_demo <- read.csv("../data/PS001_demo.csv")
PS001_class_size <- read.csv("../data/PS001_class_size.csv")

PS043_attendance <- read.csv("../data/PS043_attendance.csv")
PS043_demo <- read.csv("../data/PS043_demo.csv")
PS043_class_size <- read.csv("../data/PS043_class_size.csv")
```

Just the Bronx

Let's just look at the Bronx to understand how these variables play out. There are 58 District 75 schools in NYC. I am still working on identifying which schools are D75, but for now, let's focus on PS043, which is definitely D75. X001 (also in the Bronx) does not have a D75 option within it. We want to compare the attendance of Sped students in X001 with attendance of Sped students in X043.

For now, I just compared PSX043 (D75) and PS001 (non D75)

```
## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 7 x 5
##   schoolyear DBN   black_per hispanic_per white_per
##        <int> <chr>     <dbl>        <dbl>     <dbl>
## 1   20052006 07X001     22.3         75.7       0.2
## 2   20062007 07X001     21.6         76.3       0.1
## 3   20072008 07X001     20.6         78         0.5
## 4   20082009 07X001     22.4         76         0.2
## 5   20092010 07X001     22.9         75.1       0.1
## 6   20102011 07X001     21.8         76.5       0.1
## 7   20112012 07X001     21.4         75.5       1.6

## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 7 x 5
##   schoolyear DBN   black_per hispanic_per white_per
##        <int> <chr>     <dbl>        <dbl>     <dbl>
## 1   20052006 07X043     33.7         62.6       0.1
## 2   20062007 07X043     28.8         69.9       0
```

```
## 3    20072008 07X043        26.9          70.9          0.2
## 4    20082009 07X043        25.9          72.9          0.2
## 5    20092010 07X043        27.8          70.9          0.4
## 6    20102011 07X043        21.6          75.9          1.9
## 7    20112012 07X043        23.9          74.2          1.2
```

```r
PS001_class_size %>%
  filter(PROGRAM.TYPE == "SPEC ED" | PROGRAM.TYPE == "GEN ED") %>%
  group_by(SCHOOL.CODE, GRADE, PROGRAM.TYPE) %>%
  summarise(AVERAGE.CLASS.SIZE)
```

```
## `summarise()` regrouping output by 'SCHOOL.CODE', 'GRADE' (override with `.groups` argument)
```

```
## # A tibble: 7 x 4
## # Groups:   SCHOOL.CODE, GRADE [7]
##   SCHOOL.CODE GRADE PROGRAM.TYPE AVERAGE.CLASS.SIZE
##   <chr>       <chr> <chr>                     <dbl>
## 1 X001        01    GEN ED                     27.3
## 2 X001        02    GEN ED                     21
## 3 X001        03    GEN ED                     17.5
## 4 X001        04    GEN ED                     25.8
## 5 X001        05    GEN ED                     19.5
## 6 X001        OK    GEN ED                     25.3
## 7 X001        OK-09 SPEC ED                    11.5
```

```r
PS043_class_size %>%
  filter(PROGRAM.TYPE == "SPEC ED" | PROGRAM.TYPE == "GEN ED") %>%
  group_by(SCHOOL.CODE, GRADE, PROGRAM.TYPE) %>%
  summarise(AVERAGE.CLASS.SIZE)
```

```
## `summarise()` regrouping output by 'SCHOOL.CODE', 'GRADE', 'PROGRAM.TYPE' (override with `.groups` a
```

```
## # A tibble: 8 x 4
## # Groups:   SCHOOL.CODE, GRADE, PROGRAM.TYPE [7]
##   SCHOOL.CODE GRADE PROGRAM.TYPE AVERAGE.CLASS.SIZE
##   <chr>       <chr> <chr>                     <dbl>
## 1 X043        01    GEN ED                     17.3
## 2 X043        02    GEN ED                     18
## 3 X043        03    GEN ED                     19.7
## 4 X043        04    GEN ED                     17.7
## 5 X043        05    GEN ED                     27
## 6 X043        OK    GEN ED                     19.3
## 7 X043        OK-09 SPEC ED                    11
## 8 X043        OK-09 SPEC ED                     8
```

```r
#attendance of only the SWD (studens with disabilities)
class(PS001_attendance$X..Days.Absent)
```

```
## [1] "character"
```

```r
PS001_attendance %>%
  filter(Demographic.Variable == "SWD", Grade==3) %>%
  filter(X..Attendance != 's') %>%
  mutate(days_absent = as.numeric(X..Attendance)) %>%
  group_by(Year) %>%
  mutate(avg_days_absent = mean(days_absent)) %>%
  select(DBN, Grade, Year, avg_days_absent)
```

```
## # A tibble: 6 x 4
## # Groups:   Year [6]
##   DBN     Grade Year    avg_days_absent
##   <chr>   <chr> <chr>             <dbl>
## 1 07X001 3      2013-14            88.9
## 2 07X001 3      2014-15            93.9
## 3 07X001 3      2015-16            92.1
## 4 07X001 3      2016-17            86.7
## 5 07X001 3      2017-18            90.7
## 6 07X001 3      2018-19            85.7
```

```
PS043_attendance %>%
  filter(Demographic.Variable == "SWD", Grade==3) %>%
  filter(X..Attendance != 's') %>%
  mutate(days_absent = as.numeric(X..Attendance)) %>%
  group_by(Year) %>%
  mutate(avg_days_absent = mean(days_absent)) %>%
  select(DBN, Grade, avg_days_absent)
```

```
## Adding missing grouping variables: `Year`
```

```
## # A tibble: 6 x 4
## # Groups:   Year [6]
##   Year    DBN    Grade avg_days_absent
##   <chr>   <chr>  <chr>           <dbl>
## 1 2013-14 07X043 3                90.5
## 2 2014-15 07X043 3                90.1
## 3 2015-16 07X043 3                94.2
## 4 2016-17 07X043 3                90.1
## 5 2017-18 07X043 3                88.1
## 6 2018-19 07X043 3                86
```

```
#attendance of only the SWD (studens with disabilities)
class(PS001_attendance$X..Days.Absent)
```

```
## [1] "character"
```

```
PS001_attendance %>%
  filter(Demographic.Variable == "SWD") %>%
  filter(X..Attendance != 's') %>%
  mutate(days_absent = as.numeric(X..Attendance)) %>%
  group_by(Grade) %>%
  summarise(avg_days_absent = mean(days_absent))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 7 x 2
##   Grade      avg_days_absent
##   <chr>                <dbl>
## 1 0K                    86.5
## 2 1                     89.0
## 3 2                     89.3
## 4 3                     89.7
## 5 4                     92.1
## 6 5                     91.5
## 7 All Grades            89.6
```

```r
PS043_attendance %>%
  filter(Demographic.Variable == "SWD") %>%
  filter(X..Attendance != 's') %>%
  mutate(days_absent = as.numeric(X..Attendance)) %>%
  group_by(Grade) %>%
  summarise(avg_days_absent = mean(days_absent))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 7 x 2
##   Grade      avg_days_absent
##   <chr>                <dbl>
## 1 0K                    88.8
## 2 1                     90.2
## 3 2                     90.0
## 4 3                     89.8
## 5 4                     89.9
## 6 5                     89.1
## 7 All Grades            89.4
```