ECON 121 FA23 Problem Set 5

Robert Tso

Question 1

Verbal: list group members.

Robert Tso - A13829791

Tomas Lopez - A16798775

Stephanie Nguyen - A16215540

Akash Juwadi - A16372772

ChatGPT

Question 2

Code: Load packages and dataset, summarize data.

Verbal: Interpret summary statistics.

Conscription and crime rates seem constant across the birth years, although in 1961 there was an uptick in conscription and similarly a small uptick in crime. this matches the average conscription and crime rates across all years. I also included the draft numbers to check if the randomly assigned lottery was truly random and equally distributed, which it was, more so than conscription and crime.

```r
# The PDF will show the code you write here but not the output.
# Load packages and dataset here.

rm(list=ls())
library(tidyverse)
library(fixest)
library(plm)
library(ggplot2)


# load(url("https://github.com/tvogl/econ121/raw/main/data/crime_argentina.csv"))
crime_argentina <- read_csv("D:/Documents/Class/Econ 121/econ121/data/crime_argentina.csv")

view(crime_argentina)
```

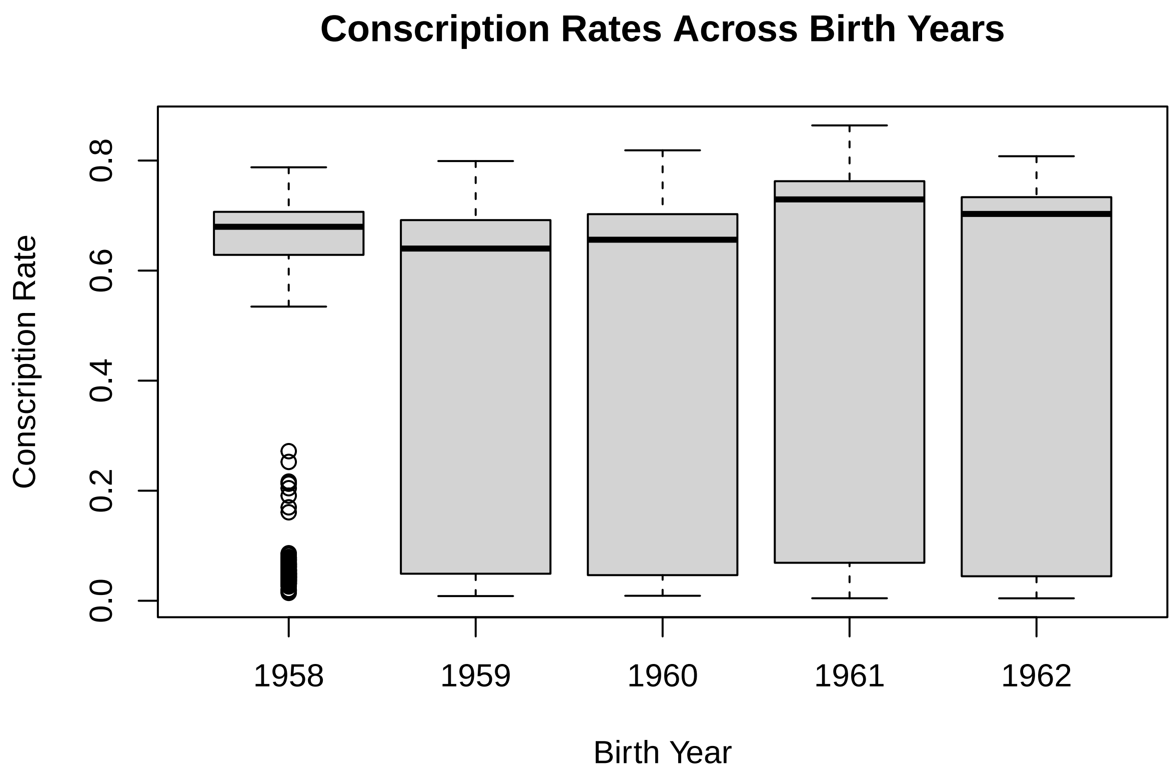
```r
# The PDF will show the code AND output here.
# Summarize the data here.
summary(crime_argentina)
```

```
##     birthyr      draftnumber      conscripted         crimerate
##  Min.   :1958   Min.   :   1.0   Min.   :0.004484   Min.   :0.01781
##  1st Qu.:1959   1st Qu.: 250.8   1st Qu.:0.059230   1st Qu.:0.05790
##  Median :1960   Median : 500.5   Median :0.677785   Median :0.06841
##  Mean   :1960   Mean   : 500.5   Mean   :0.503122   Mean   :0.06927
##  3rd Qu.:1961   3rd Qu.: 750.2   3rd Qu.:0.722437   3rd Qu.:0.08131
##  Max.   :1962   Max.   :1000.0   Max.   :0.863850   Max.   :0.14907
##    argentine        indigenous          naturalized
##  Min.   :0.9739   Min.   :0.0000000   Min.   :0.0000000
##  1st Qu.:0.9959   1st Qu.:0.0000000   1st Qu.:0.0000000
##  Median :1.0000   Median :0.0000000   Median :0.0000000
##  Mean   :0.9986   Mean   :0.0008938   Mean   :0.0005358
##  3rd Qu.:1.0000   3rd Qu.:0.0000000   3rd Qu.:0.0000000
##  Max.   :1.0000   Max.   :0.0144231   Max.   :0.0260870
```

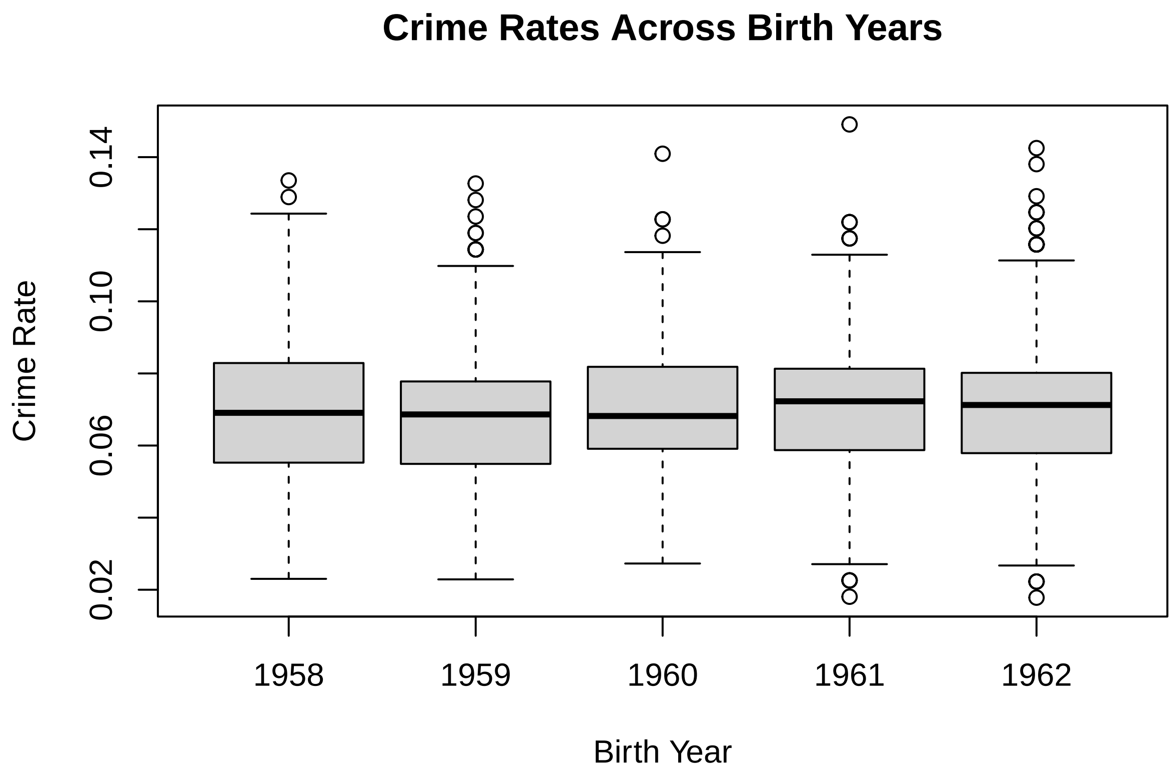
```r
# Check for missing values in the entire dataset
any_missing <- any(is.na(crime_argentina))

# Print the result
if (any_missing) {
  cat("There are missing values in the dataset.\n")
} else {
  cat("There are no missing values in the dataset.\n")
}
```

```
## There are no missing values in the dataset.
```

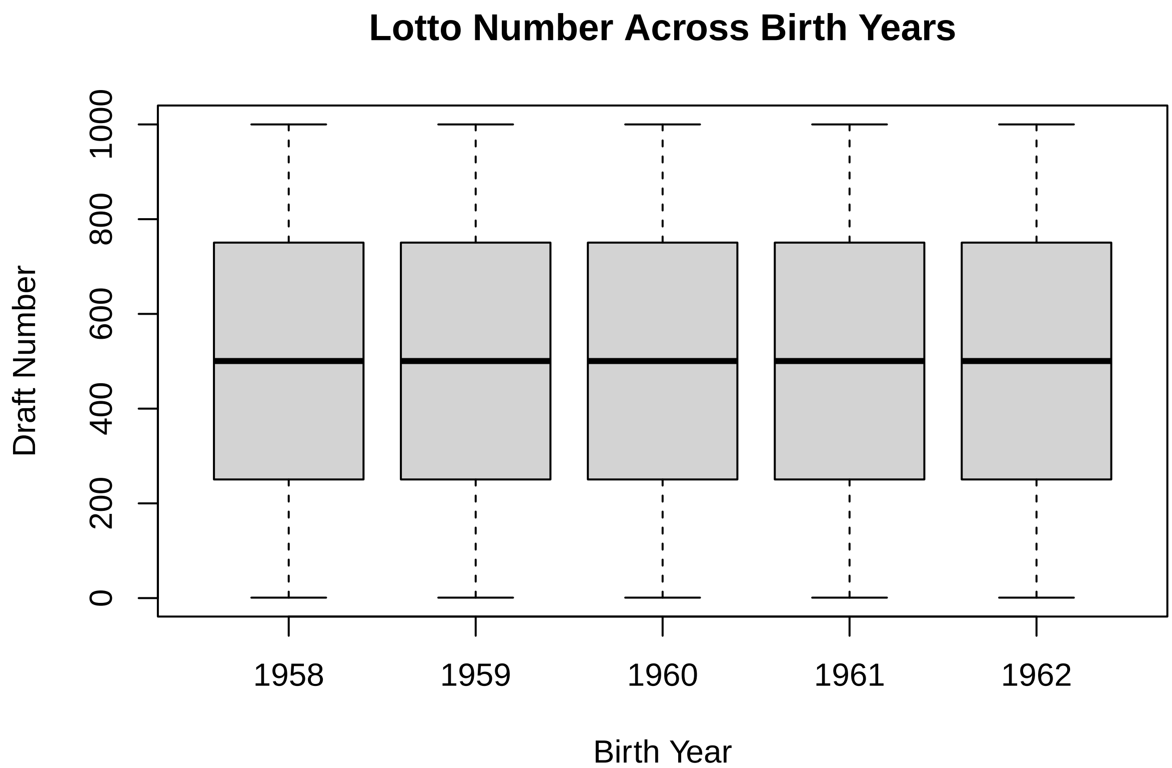
```r
# Explore differences in conscription rates across birth years
boxplot(conscripted ~ birthyr, data = crime_argentina, main = "Conscription Rates Across Birth Years", :
```




```r
# Explore differences in crime rates across birth years
boxplot(crimerate ~ birthyr, data = crime_argentina, main = "Crime Rates Across Birth Years", xlab = "Bi
```




```
# Explore differences in lottery numbers across birth years
boxplot(draftnumber ~ birthyr, data = crime_argentina, main = "Lotto Number Across Birth Years", xlab =
```


Lotto Number Across Birth Years
Draft Number
0
200
400
600
800
1000
1958
1959
1960
1961
1962
Birth Year

Question 3

Code: Regression.

Verbal: Interpret.

The statistically significant coefficient for conscription's effect on crime rate is 0.002420 , implying a positive correlation. There are likely biases as we see that birth year also has a statistical significant coefficient of 0.000478, meaning that the relative year of conscription based on military demand were possibly pulling more or less young adults off the streets away from committing crime. Conversely, there may be a negative career impact from conscription on young adults who exit the service and they turn to crime.

```r
# All question 3 code here.


# Run OLS regression
ols_crime_conscript <- feols(crimerate ~ conscripted + birthyr +
                                          argentine + indigenous + naturalized,
                            data = crime_argentina,
                            vcov = 'hetero')


# Summarize the model
summary(ols_crime_conscript)
```

```
## OLS estimation, Dep. Var.: crimerate
## Observations: 5,000
## Standard-errors: Heteroskedasticity-robust
##              Estimate Std. Error   t value  Pr(>|t|)
## (Intercept) -1.463170   0.973816 -1.502511 0.1330284
## conscripted  0.002420   0.000827  2.925830 0.0034509 **
## birthyr      0.000478   0.000185  2.591070 0.0095956 **
## argentine    0.593762   0.919131  0.646003 0.5183069
## indigenous   0.566639   0.921980  0.614589 0.5388541
## naturalized  0.757181   0.943691  0.802361 0.4223822
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.017752   Adj. R2: 0.002616
```

Question 4

Code: Generate variable.

```r
# All question 4 code here

# Define birth years and corresponding cutoffs
years <- c(1958, 1959, 1960, 1961, 1962)
cutoffs <- c(175, 320, 341, 350, 320)

# Code a variable indicating eligibility (1 if eligible, 0 if not)
crime_argentina$eligible <- with(crime_argentina,
                  ifelse(birthyr == years[1] & draftnumber >= cutoffs[1], 1,
                  ifelse(birthyr == years[2] & draftnumber >= cutoffs[2], 1,
                  ifelse(birthyr == years[3] & draftnumber >= cutoffs[3], 1,
                  ifelse(birthyr == years[4] & draftnumber >= cutoffs[4], 1,
                  ifelse(birthyr == years[5] & draftnumber >= cutoffs[5], 1,
                         0))))))

# Check the new variable
view(crime_argentina)
```

Question 5

Code: Regression.

Verbal: Interpret.

We control for birth year because the age of adulthood or parenthood has an impact on whether or not a man is enlisted when his number is drawn for conscription, younger ages are more likely to be enlisted since they are less likely to be fathers or clerics at the time of the drawing. We control for ethnic composition because there may be a racial systemic factor in conscription. Upon regressing we see that the 2 statistically significant coefficients are eligible at 0.659805 and birth year at 0.010910. Eligibility should have a direct correlation on conscription since it determines if that person qualifies for service. Higher birth year means younger conscripts so they are more likely to be enlisted as well.

```r
# All question 5 code here

#first stage regression
ols_conscript_eligibility <- feols(conscripted ~ eligible + birthyr +
                                              argentine + indigenous +
                                              naturalized,
                                   data = crime_argentina,
                                   vcov = 'hetero')

summary(ols_conscript_eligibility)
```

```
## OLS estimation, Dep. Var.: conscripted
## Observations: 5,000
## Standard-errors: Heteroskedasticity-robust
##               Estimate Std. Error   t value   Pr(>|t|)
## (Intercept) -18.691883   2.570654  -7.27126 4.1140e-13 ***
## eligible      0.659805   0.001149 574.25224  < 2.2e-16 ***
## birthyr       0.010910   0.000465  23.47866  < 2.2e-16 ***
## argentine    -2.648813   2.428956  -1.09051 2.7554e-01
## indigenous   -3.162557   2.450683  -1.29048 1.9694e-01
## naturalized  -3.225953   2.446753  -1.31846 1.8741e-01
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.04649   Adj. R2: 0.97672
```

Question 6

Code: Regression.

Verbal: Interpret.

When regressing on the IV to find the reduced form we see that eligible has a statistically significant coefficient of 0.001839 compared to regressing crimerate against conscription for 0.002420. These are both in the same orders of magnitude and seeing that the coefficient of eligible on conscripted was 0.659805, which is closer to 1.00 than not, we imagine how eligible can be a strong IV for conscription.

```r
# All question 6 code here

# Run reduced form regression
reduced_form_model <- feols(crimerate ~ eligible + birthyr +
                                        argentine + indigenous + naturalized,
                            data = crime_argentina,
                            vcov = 'hetero')

# Summarize the model
summary(reduced_form_model)
```

```
## OLS estimation, Dep. Var.: crimerate
## Observations: 5,000
## Standard-errors: Heteroskedasticity-robust
##              Estimate Std. Error   t value  Pr(>|t|)
## (Intercept) -1.522633   0.974566 -1.562371 0.1182641
## eligible     0.001839   0.000552  3.333452 0.0008640 ***
## birthyr      0.000513   0.000185  2.765164 0.0057103 **
## argentine    0.585824   0.919388  0.637188 0.5240314
## indigenous   0.557337   0.922264  0.604314 0.5456625
## naturalized  0.747081   0.943790  0.791575 0.4286463
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.017747   Adj. R2: 0.003123
```

Question 7

Code: Calculation.

Verbal: Interpret.

```r
# All question 7 code here

# First Stage Estimate
first_stage_estimate <- coef(ols_conscript_eligibility)['eligible']

# Reduced Form Estimate
reduced_form_estimate <- coef(reduced_form_model)['eligible']

# Instrumental Variables (IV) Estimate
iv_estimate <- reduced_form_estimate / first_stage_estimate

# Print the IV Estimate
cat("Instrumental Variables (IV) Estimate:", iv_estimate, "\n")
```

```
## Instrumental Variables (IV) Estimate: 0.002787198
```

Question 8

Code: Regression.

Verbal: Interpret. fit_conscripted is 0.002787 and the OLS conscripted is 0.002420 which are similar but different values. This makes sense as we are accounting for the IV eligibility in the 2SLS regression, while in OLS we have a much less accurate coefficient for conscription.

```r
# All question 8 code here

tsls_crimerate <- feols(crimerate ~ birthyr + argentine + indigenous +
                                   naturalized| conscripted ~ eligible,
                        data = crime_argentina)

summary(tsls_crimerate, stage= 1)
```

```
## TSLS estimation, Dep. Var.: conscripted, Endo.: conscripted, Instr.: eligible
## First stage: Dep. Var.: conscripted
## Observations: 5,000
## Standard-errors: IID
##               Estimate Std. Error   t value   Pr(>|t|)
## (Intercept) -18.691883   2.719671  -6.87285 7.0650e-12 ***
## eligible      0.659805   0.001442 457.41628  < 2.2e-16 ***
## birthyr       0.010910   0.000481  22.67033  < 2.2e-16 ***
## argentine    -2.648813   2.581248  -1.02618 3.0486e-01
## indigenous   -3.162557   2.601585  -1.21563 2.2418e-01
## naturalized  -3.225953   2.627641  -1.22770 2.1962e-01
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.04649   Adj. R2: 0.97672
## F-test (1st stage): stat = 209,229.6, p < 2.2e-16, on 1 and 4,994 DoF.
```

```r
summary(tsls_crimerate, stage= 2)
```

```
## TSLS estimation, Dep. Var.: crimerate, Endo.: conscripted, Instr.: eligible
## Second stage: Dep. Var.: crimerate
## Observations: 5,000
## Standard-errors: IID
##                  Estimate Std. Error   t value   Pr(>|t|)
## (Intercept)     -1.470535   1.038134 -1.416518 0.15668621
## fit_conscripted  0.002787   0.000835  3.338784 0.00084762 ***
## birthyr          0.000482   0.000183  2.633829 0.00846883 **
## argentine        0.593206   0.985641  0.601849 0.54730231
## indigenous       0.566152   0.993405  0.569910 0.56876423
## naturalized      0.756072   1.003351  0.753547 0.45115677
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.017752   Adj. R2: 0.002576
## F-test (1st stage), conscripted: stat = 209,229.6   , p < 2.2e-16 , on 1 and 4,994 DoF.
##                      Wu-Hausman: stat =       8.3059, p = 0.003969, on 1 and 4,993 DoF.
```

Question 9

Verbal: Assessment.

Yes, eligibility meets the requirements of Relevance and Exogeneity. It is relevant because eligibility is based off of the draft number, which determines who gets drafted. The lottery assignment of the draft number as demonstrated in the summarized data, showed that it was equally distributed across all birth years. This randomness is ideal for an IV because we can eliminate a selection bias from the initial drawing of the numbers.

Question 10

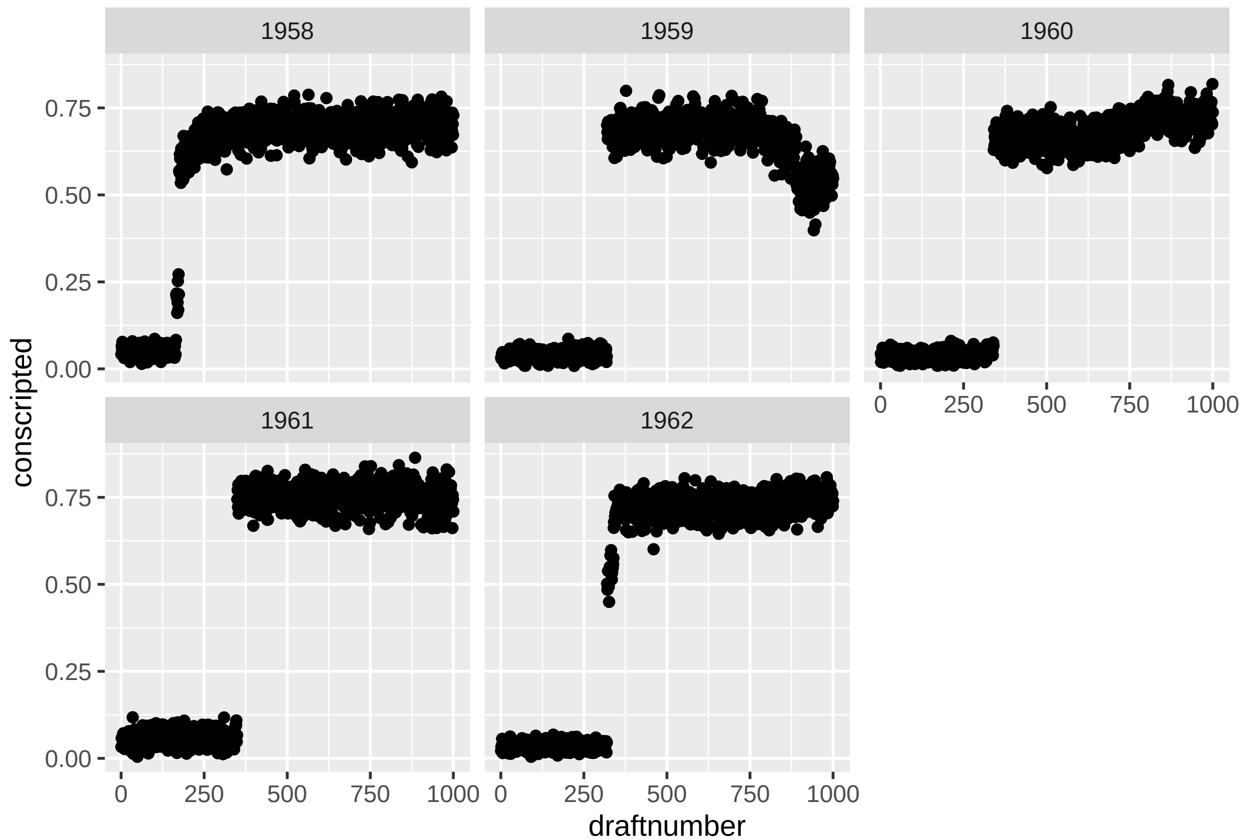
Verbal: Interpretation.

2SLS is measuring the sub-population that have been conscripted and the Average treatment effect (ATE) on crime rate. By using the eligibility IV we are now measuring the local average treatment effect since conscription status changes based off of eligibility. Treatment-on-the-treated does not apply since there were a handful of people who were conscripted while they were ineligible by the lottery cut-off, meaning those who were supposed to be untreated by measure of eligibility, still got treatment by conscription.

Question 11

a. Code.

```
ggplot(crime_argentina, aes(draftnumber, conscripted)) +
geom_point() +
facet_wrap(~birthyr)
```



b. code

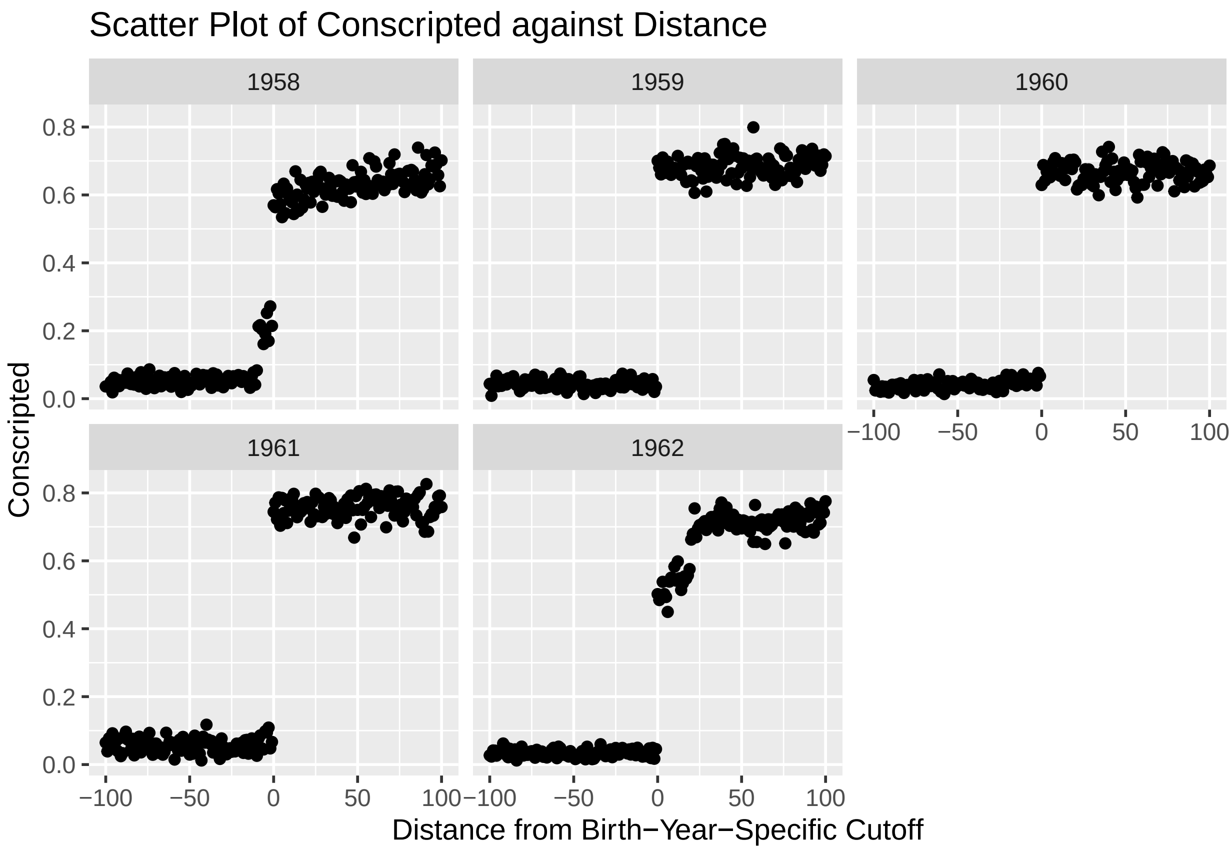
```
crime_argentina$distance <- with(crime_argentina,
                  ifelse(birthyr == years[1], draftnumber - cutoffs[1],
                  ifelse(birthyr == years[2], draftnumber - cutoffs[2],
                  ifelse(birthyr == years[3], draftnumber - cutoffs[3],
                  ifelse(birthyr == years[4], draftnumber - cutoffs[4],
                  ifelse(birthyr == years[5], draftnumber - cutoffs[5],
                         0))))))

crime_argentina_distance <- crime_argentina %>%
  filter(distance >= -100, distance <= 100)
```

```r
view(crime_argentina_distance)
```

c. code and verbal

Yes, the jump from the left and right of Zero distance is pronounced enough to show that conscription rates increase.

```r
ggplot(crime_argentina_distance, aes(distance, conscripted)) +
  geom_point() +
  labs(title = "Scatter Plot of Conscripted against Distance",
       x = "Distance from Birth-Year-Specific Cutoff",
       y = "Conscripted") +
  facet_wrap(~birthyr)
```

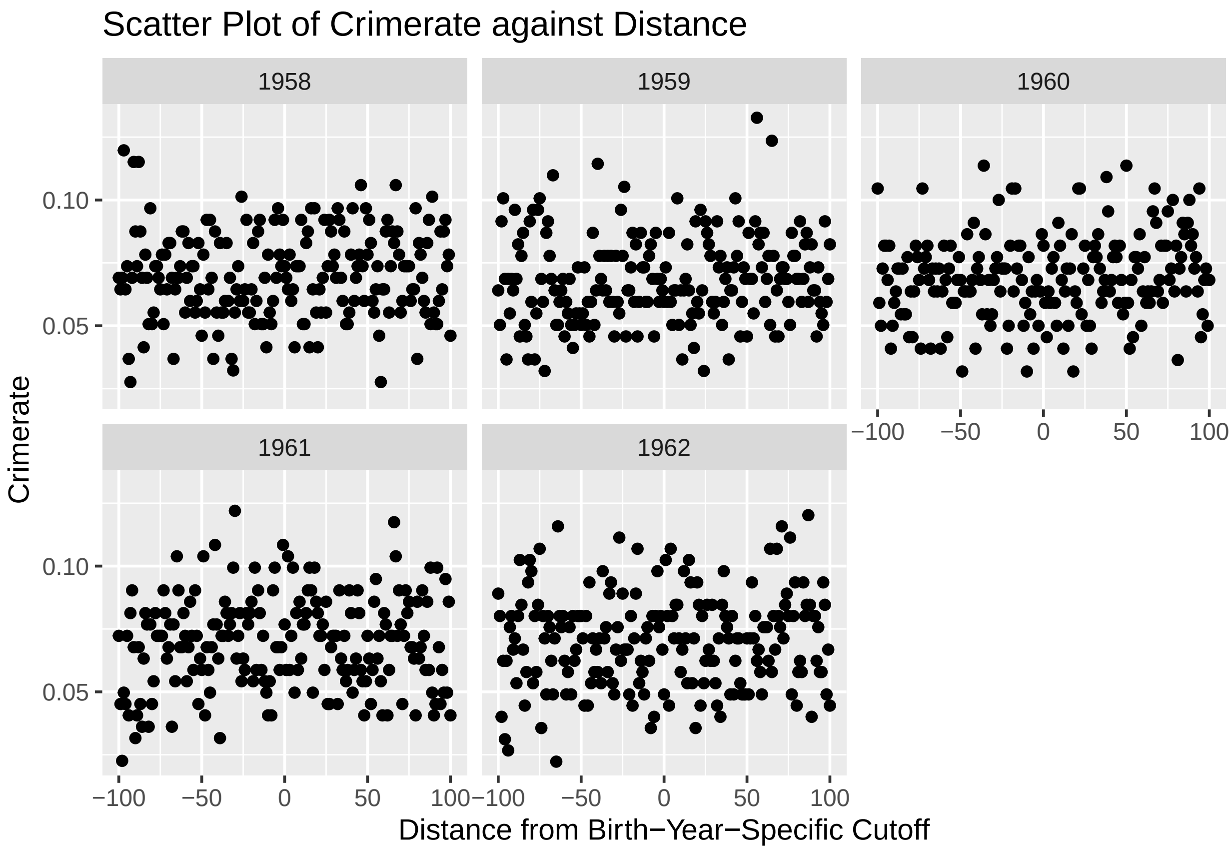


d. code and verbal

There does not seem to be a rise of crime after the cutoff, the plots look equally distributed between 0.05 and 0.10, with some outliers that are inconsistent across the 5 years.

```r
ggplot(crime_argentina_distance, aes(distance, crimerate)) +
  geom_point() +
```

```
  labs(title = "Scatter Plot of Crimerate against Distance",
       x = "Distance from Birth-Year-Specific Cutoff",
       y = "Crimerate") +
  facet_wrap(~birthyr)
```



e. code and verbal

The 2SLS regression includes the interaction term between distance and eligibility to account for negative slopes on either side of the cutoff. In the first stage, crime rate is regressed against distance, eligibility and the interaction distance x eligibility. In the second stage, crime rate is regressed against conscripted, distance, and the interaction of eligibility x interaction. We can see in the second stage that only the intercept is statistically significant, so there is unlikely a relationship between conscription and crimerate.

```
crime_argentina_distance <- crime_argentina_distance %>%
  mutate(distXelig = distance*eligible)

fuzzy_rd <- feols(crimerate ~ distance + distXelig | conscripted ~ eligible,
                  data = crime_argentina_distance, vcov = "hetero")

summary(fuzzy_rd, stage = 1)
```

```
## TSLS estimation, Dep. Var.: conscripted, Endo.: conscripted, Instr.: eligible
## First stage: Dep. Var.: conscripted
```

```
## Observations: 1,005
## Standard-errors: Heteroskedasticity-robust
##             Estimate Std. Error  t value   Pr(>|t|)
## (Intercept) 0.058729   0.003923 14.97162  < 2.2e-16 ***
## eligible    0.594232   0.007541 78.80529  < 2.2e-16 ***
## distance    0.000211   0.000059  3.58108 0.00035862 ***
## distXelig   0.000406   0.000116  3.49586 0.00049325 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.046265   Adj. R2: 0.979227
## F-test (1st stage): stat = 10,317.7, p < 2.2e-16, on 1 and 1,001 DoF.
```

```r
summary(fuzzy_rd, stage = 2)
```

```
## TSLS estimation, Dep. Var.: crimerate, Endo.: conscripted, Instr.: eligible
## Second stage: Dep. Var.: crimerate
## Observations: 1,005
## Standard-errors: Heteroskedasticity-robust
##                     Estimate Std. Error   t value  Pr(>|t|)
## (Intercept)      0.06962858   0.001754 39.689883 < 2.2e-16 ***
## fit_conscripted  0.00031324   0.003702  0.084610   0.93259
## distance         0.00002882   0.000030  0.971110   0.33173
## distXelig       -0.00000960   0.000039 -0.243553   0.80763
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.017601   Adj. R2: 0.004089
## F-test (1st stage), conscripted: stat = 10,317.7     , p < 2.2e-16 , on 1 and 1,001 DoF.
##                      Wu-Hausman: stat =      0.697271, p = 0.403901, on 1 and 1,000 DoF.
```

f. verbal

Because we excluded distance beyond the range of 100 and -100 we are focused on the middle 200 lottery numbers drawn in the data set. This is akin to zooming into the scatter plot or dataset. The differences between one drafted number to the next may not vary much on a small scale and can indicate low statistical power, ie. we may have dropped too many data points. There may still be a correlation when we reintroduce or widen the range, or maybe include more birth years beyond 1958-1962. In regards to regression we have also excluded control variables of ethnicity which may account for this discrepancy if we introduce interaction terms between "argentine", "ingenous", "naturalized" and distance/eligibility.