ECON 121 FA25 Problem Set 5

Solution

Question 2

Summary statistics appear below. The average conscription rate is 50 percent, and the average crime rate is 7 percent.

```r
summary(crime)
```

```
##     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
```

There are large variations in conscription rates by birth year, and somewhat smaller variations in crime rates by birth year.

```r
crime |>
  group_by(birthyr) |>
  summarize(conscripted_mean = mean(conscripted),
            crimerate_mean = mean(crimerate))
```

```
## # A tibble: 5 x 3
##   birthyr conscripted_mean crimerate_mean
##     <dbl>            <dbl>          <dbl>
## 1    1958            0.576         0.0690
## 2    1959            0.462         0.0680
## 3    1960            0.467         0.0690
## 4    1961            0.509         0.0698
## 5    1962            0.501         0.0706
```

To shed light on whether these birth year differences are statistically significant (this was not necessary for full credit), we can run regressions as follows:

```r
conscripted_model <- feols(conscripted ~ factor(birthyr), data = crime, vcov = 'hetero')
conscripted_model
```

```
## OLS estimation, Dep. Var.: conscripted
## Observations: 5,000
```

```
## Standard-errors: Heteroskedasticity-robust
##                      Estimate Std. Error  t value   Pr(>|t|)
## (Intercept)          0.576423   0.007637 75.47897  < 2.2e-16 ***
## factor(birthyr)1959 -0.114648   0.012059 -9.50755  < 2.2e-16 ***
## factor(birthyr)1960 -0.109412   0.012435 -8.79885  < 2.2e-16 ***
## factor(birthyr)1961 -0.066985   0.013029 -5.14119 2.8343e-07 ***
## factor(birthyr)1962 -0.075463   0.012723 -5.93133 3.2072e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.302072   Adj. R2: 0.017358
```

```r
fitstat(conscripted_model, "f")
```

```
## F-test: stat = 23.1, p < 2.2e-16, on 4 and 4,995 DoF.
```

```r
crimerate_model <- feols(conscripted ~ factor(birthyr), data = crime, vcov = 'hetero')
crimerate_model
```

```
## OLS estimation, Dep. Var.: conscripted
## Observations: 5,000
## Standard-errors: Heteroskedasticity-robust
##                      Estimate Std. Error  t value   Pr(>|t|)
## (Intercept)          0.576423   0.007637 75.47897  < 2.2e-16 ***
## factor(birthyr)1959 -0.114648   0.012059 -9.50755  < 2.2e-16 ***
## factor(birthyr)1960 -0.109412   0.012435 -8.79885  < 2.2e-16 ***
## factor(birthyr)1961 -0.066985   0.013029 -5.14119 2.8343e-07 ***
## factor(birthyr)1962 -0.075463   0.012723 -5.93133 3.2072e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.302072   Adj. R2: 0.017358
```

```r
fitstat(crimerate_model, "f")
```

```
## F-test: stat = 23.1, p < 2.2e-16, on 4 and 4,995 DoF.
```

The fitstat() output at the end tells us the p-value from the joint F test on all estimated coefficients. For both regressions, this p-value is less
than 0.05, indicating that the birth year variations in both variables are statistically significant.

I used fitstat() because the direct output from feols() does not include the joint F-statistic.

Question 3

OLS regressions of various crime rates on conscription rates.
Crime rates are higher in cells with higher conscription rates.

```r
feols(crimerate ~ conscripted + factor(birthyr) + indigenous + naturalized,
      data = crime,
      vcov = 'hetero')
```

```
## OLS estimation, Dep. Var.: crimerate
## Observations: 5,000
## Standard-errors: Heteroskedasticity-robust
##                      Estimate Std. Error   t value  Pr(>|t|)
## (Intercept)          0.067675   0.000766 88.371634 < 2.2e-16 ***
## conscripted          0.002265   0.000836  2.708184 0.0067883 **
## factor(birthyr)1959 -0.000731   0.000798 -0.914957 0.3602586
## factor(birthyr)1960  0.000271   0.000792  0.342535 0.7319626
```

```
## factor(birthyr)1961  0.000910   0.000802  1.134874 0.2564826
## factor(birthyr)1962  0.001615   0.000825  1.957338 0.0503634 .
## indigenous          -0.035845   0.127329 -0.281517 0.7783255
## naturalized          0.131270   0.171579  0.765071 0.4442652
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.017748   Adj. R2: 0.002647
```

This association would reflect a causal effect only if the lottery completely determined conscription – i.e., there was perfect compliance. But we know from the description that some individuals were exempted from military service, even if the lottery made them eligible. If people exempted from military service have different crime propensities than non-exempt people, then the OLS estimate is biased. For instance, clerics and individuals with dependents may be less prone to crime than non-exempt individuals, which would bias us toward finding a larger (more positive) association between conscription and crime.

Alternatively, it may be that patriotic individuals are more likely to sign up even if they are not eligible for the lottery, and are less likely to find ways to become exempt. In this case, conscription would be positively correlated with patriotism. If patriotic individuals are also less likely to commit crimes, then these patterns would bias us toward finding a smaller (more negative) association betetween conscription and crime.

Question 4

```r
crime <-
  crime |>
  mutate(eligible = if_else((draftnumber>=175 & birthyr==1958) |
                         (draftnumber>=320 & birthyr==1959) |
                         (draftnumber>=341 & birthyr==1960) |
                         (draftnumber>=350 & birthyr==1961) |
                         (draftnumber>=320 & birthyr==1962), 1, 0))
```

Question 5

To estimate the first-stage effect of eligibility on conscription, we regress conscription rates on eligibility, controlling for birth year dummies. The birth year dummies are required because the eligibility rule depended on birth year – elibility is randomly assigned within each birth year, but it varies systematically across birth years. Ethnic composition covariates are not necessary because the assignment of eligibility is unrelated to ethnicity.

```r
feols(conscripted ~ eligible + factor(birthyr),
      data = crime,
      vcov = 'hetero')
```

```
## OLS estimation, Dep. Var.: conscripted
## Observations: 5,000
## Standard-errors: Heteroskedasticity-robust
##                      Estimate Std. Error    t value   Pr(>|t|)
## (Intercept)          0.032299   0.001663  19.418726  < 2.2e-16 ***
## eligible             0.658746   0.001152 571.861723  < 2.2e-16 ***
## factor(birthyr)1959 -0.019130   0.002359  -8.109763 6.3272e-16 ***
## factor(birthyr)1960 -0.000060   0.001788  -0.033430 9.7333e-01
## factor(birthyr)1961  0.048295   0.001751  27.589448  < 2.2e-16 ***
## factor(birthyr)1962  0.020055   0.001862  10.770645  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.043263   Adj. R2: 0.97984
```

The results indicate that eligibility has a large, positive, and significant effect on conscription. Eligibility raises conscription rates by 66 percentage points, with a t-statistic of 572!

Question 6

Like the first stage, the reduced form regression must control for birth year dummies. This result reflects a causal effect, since eligibility is randomly assigned conditional on year of birth. However, it reflects the average effect of *eligibility*, not *conscription*.

```r
feols(crimerate ~ eligible + factor(birthyr),
      data = crime,
      vcov = 'hetero')
```

```
## OLS estimation, Dep. Var.: crimerate
## Observations: 5,000
## Standard-errors: Heteroskedasticity-robust
##                      Estimate Std. Error   t value  Pr(>|t|)
## (Intercept)          0.067524   0.000733 92.157027 < 2.2e-16 ***
## eligible             0.001759   0.000556  3.167214 0.0015483 **
## factor(birthyr)1959 -0.000747   0.000795 -0.939845 0.3473425
## factor(birthyr)1960  0.000320   0.000791  0.404771 0.6856635
## factor(birthyr)1961  0.001129   0.000803  1.406386 0.1596716
## factor(birthyr)1962  0.001853   0.000811  2.285724 0.0223121 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.017745   Adj. R2: 0.003432
```

We find that eligibility raises crime rates (by 0.2 %-points).

Question 7

We can compute the instrumental variables estimator by dividing the

reduced form coefficient by the first stage coefficient.

```r
0.0017595/0.6587458
```

```
## [1] 0.002670985
```

Question 8

We get exactly the same point estimates by running 2SLS.

```r
feols(crimerate ~ factor(birthyr) | conscripted ~ eligible,
      data = crime,
      vcov = 'hetero')
```

```
## TSLS estimation, Dep. Var.: crimerate, Endo.: conscripted, Instr.: eligible
## Second stage: Dep. Var.: crimerate
## Observations: 5,000
## Standard-errors: Heteroskedasticity-robust
##                      Estimate Std. Error   t value  Pr(>|t|)
## (Intercept)          0.067437   0.000751 89.851422 < 2.2e-16 ***
## fit_conscripted      0.002671   0.000843  3.166925 0.0015499 **
```

```
## factor(birthyr)1959 -0.000696   0.000798 -0.872989 0.3827113
## factor(birthyr)1960  0.000320   0.000791  0.404865 0.6855937
## factor(birthyr)1961  0.001000   0.000798  1.253678 0.2100177
## factor(birthyr)1962  0.001800   0.000809  2.224338 0.0261702 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.01775   Adj. R2: 0.002842
## F-test (1st stage), conscripted: stat = 238,468.2, p < 2.2e-16 , on 1 and 4,994 DoF.
##                     Wu-Hausman: stat =      11.6, p = 6.503e-4, on 1 and 4,993 DoF.
```

We find that conscription raises the crime rate by .27 percentage points, This magnitude is remarkably similar to the OLS result in Question 2. This similarity suggests that perhaps after all, the OLS results were not biased.

Question 9

1. The instrument is clearly relevant, with a first-stage t-statistic of over 500.
2. It also likely satisfies the independence assumption, since the lottery effectively randomizes eligibility.
3. Whether it satisfies the exclusion restriction is a little less clear. It seems likely that eligibility only affects crime through its effect on conscription, but alternative pathways are possible. For instance,suppose eligible people who were opposed to military service evaded conscription by hiding from the authorities, which exposed them to more crime.
4. The monotonicity condition is likely to be satisfied. It is very difficult to fathom why a person who would volunteer for the military without being drafted would refuse to serve if drafted.

Note: Instead of 2-4, you could have also discussed the instrument exogeneity assumption from the constant treatment effects setup.

Question 10

As noted above, the 2SLS result for crimerate suggests that conscription raises the probability of having a criminal record by 0.27 percentage points. Viewed through the lens of heterogeneous treatment effects, this estimate reflects the average treatment effect among people who were induced to serve by the draft lottery. It does not reflect the effects of conscription among people who would have volunteered anyway (always takers), nor the effects of conscription among exempted people (never takers).

If always takers exist in this context, then we should call the estimand a LATE, not a TOT. Recall the first stage regression:

```r
feols(conscripted ~ eligible + factor(birthyr),
      data = crime,
      vcov = 'hetero')
```

```
## OLS estimation, Dep. Var.: conscripted
## Observations: 5,000
## Standard-errors: Heteroskedasticity-robust
##                      Estimate Std. Error    t value   Pr(>|t|)
## (Intercept)          0.032299   0.001663  19.418726  < 2.2e-16 ***
## eligible             0.658746   0.001152 571.861723  < 2.2e-16 ***
## factor(birthyr)1959 -0.019130   0.002359  -8.109763 6.3272e-16 ***
## factor(birthyr)1960 -0.000060   0.001788  -0.033430 9.7333e-01
## factor(birthyr)1961  0.048295   0.001751  27.589448  < 2.2e-16 ***
## factor(birthyr)1962  0.020055   0.001862  10.770645  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## RMSE: 0.043263   Adj. R2: 0.97984
```

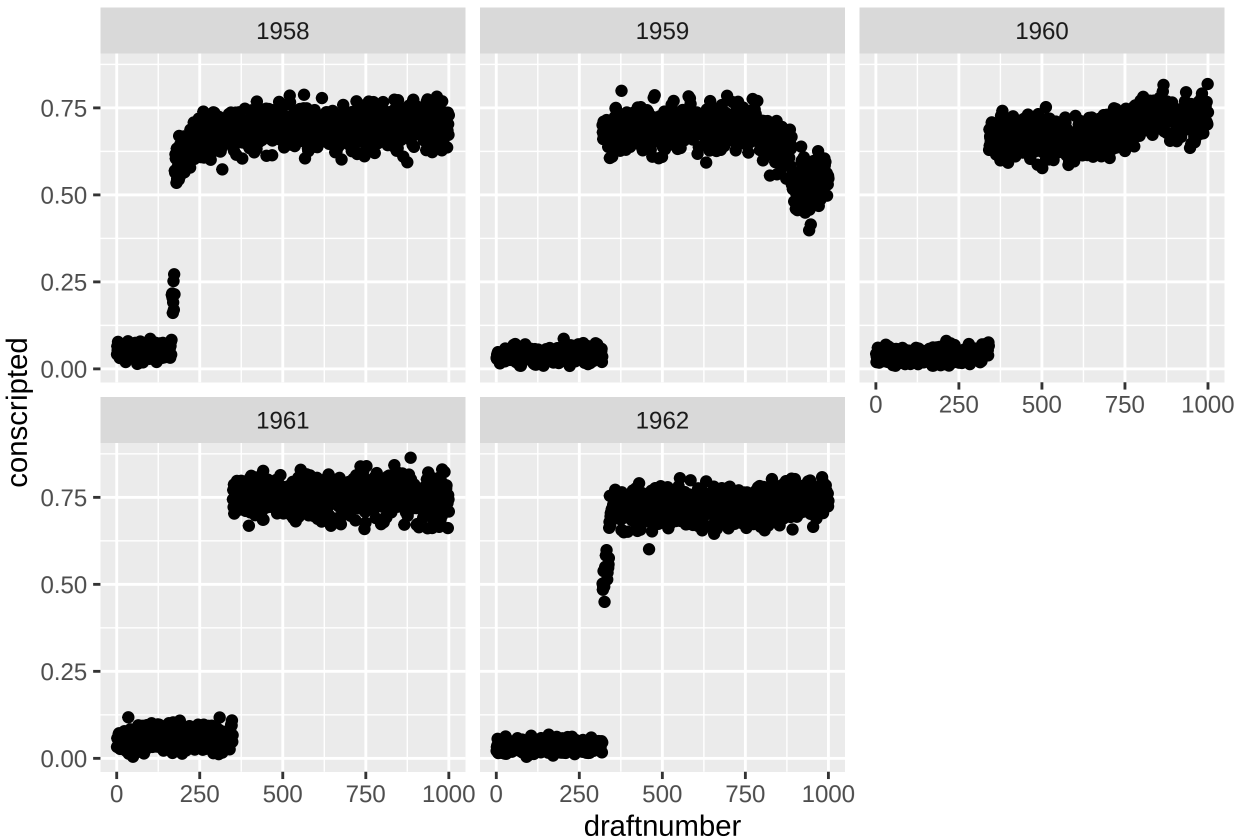
Because the intercept is positive, we can conclude that some non-eligible people did sign up for the military. So there are always takers, and we should call it a LATE.

Question 11

a.

The scatterplot represents the effect of crossing the cutoff on conscription. It shows an opportunity to use the cutoff rule in a fuzzy RD design. It is "fuzzy" because conscription rates are positive below the cutoff and less than 1 above the cutoff, so the change in treatment probability is smaller than 1.

```
ggplot(data = crime, aes(draftnumber, conscripted)) +
  geom_point() +
  facet_wrap(~ birthyr)
```



b.

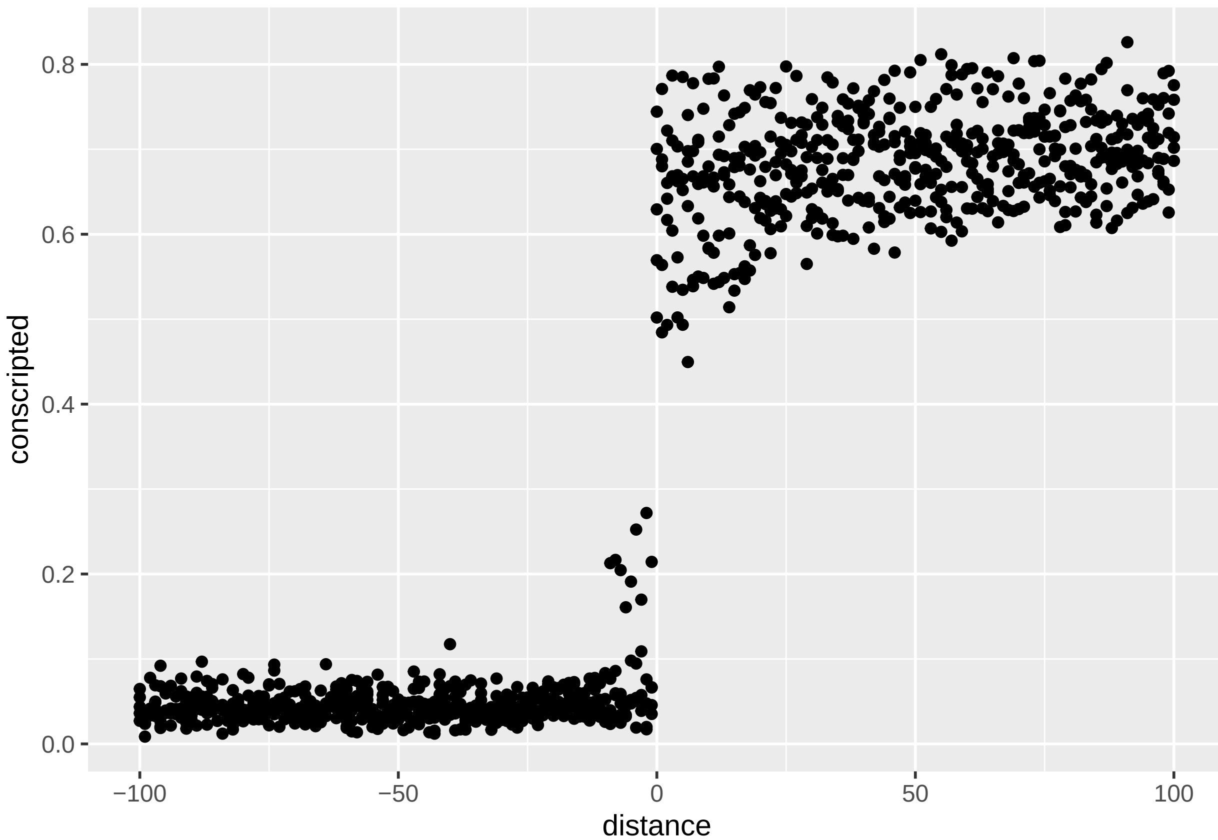
```
# generate distance to cutoff, filter to window from -100 to +100
crime_local <-
  crime |>
  mutate(distance = (draftnumber-175) * if_else(birthyr == 1958,1,0) +
                    (draftnumber-320) * if_else(birthyr == 1959,1,0) +
                    (draftnumber-341) * if_else(birthyr == 1960,1,0) +
                    (draftnumber-350) * if_else(birthyr == 1961,1,0) +
```

```
                    (draftnumber-320) * if_else(birthyr == 1962,1,0)) |>
  filter(distance<=100 & distance>=-100)
```

c.

Draw the first-stage scatterplot pooling all birth years:

```
ggplot(data = crime_local, aes(distance, conscripted)) +
  geom_point()
```

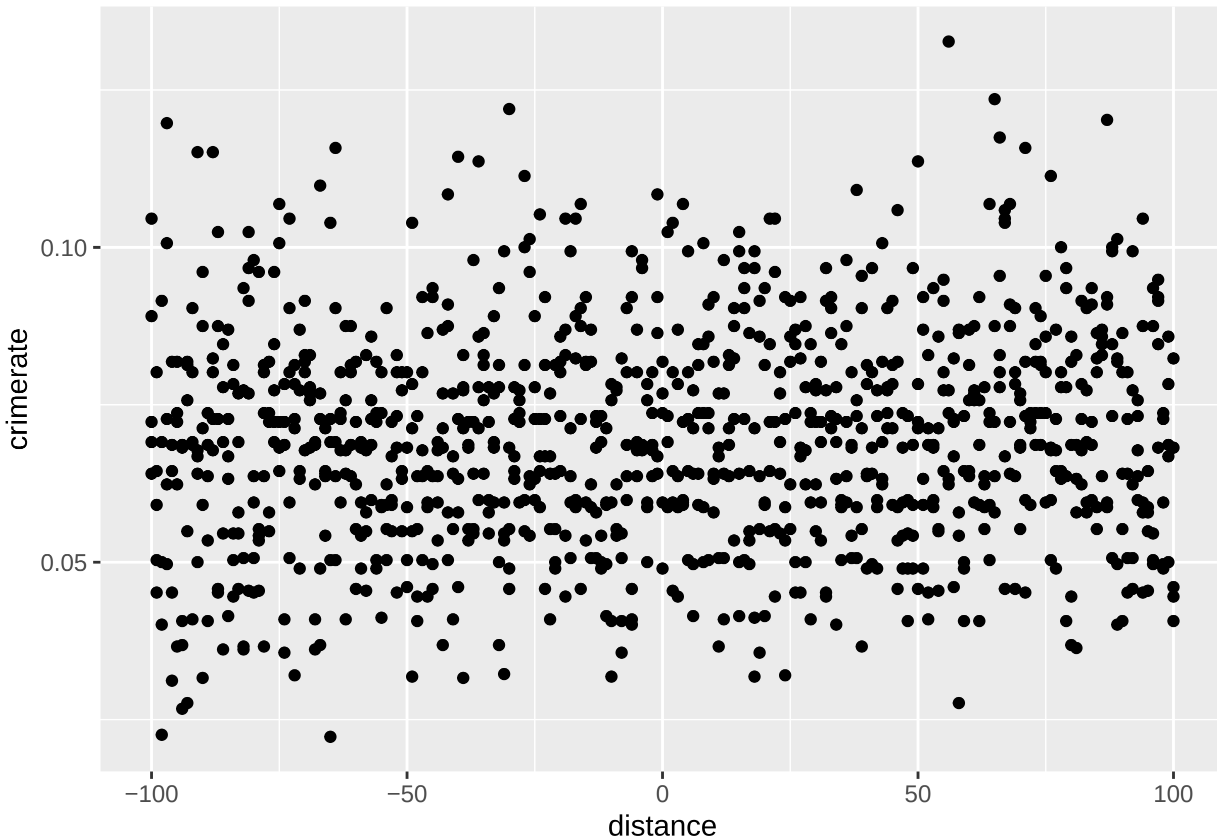


The discontinuity is clear. Crossing the threshold has an effect on conscription.

d.

Draw the reduced-form scatterplot pooling all birth years:

```
ggplot(data = crime_local, aes(distance, crimerate)) +
  geom_point()
```



Here, a discontinuity is very hard to see here. Crossing the threshold does not seem to have an effect on conscription.

e.

Estimate a 2sls regression using local linear regression with a bandwidth of 100:

```
crime_local <- crime_local |> mutate(distXelig = distance*eligible)
fuzzy_rd <-
  feols(crimerate ~ distance + distXelig | conscripted ~ eligible,
        data = crime_local,
        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.
```

We can see from the first stage that crossing the cutoff raises conscription by almost 60 percentage points. But there is no significant discontinuity in crime rates. These findings confirm the patterns in the scatter plots.

f.

Why this disappointing result? A big problem is that by focusing right around the cutoff, we have increased the variance (std. errors) of our estimator a lot. The 2sls result earlier in the problem set had an SE of 0.0008. Here we have a standard error of 0.0037, much larger! We cannot reject an effect of 0.0027, which is what we found earlier.