

Lab_11_Mariel: Labor Market Effects of the Mariel Boatlift

Introduction

In this lab we will be replicating a classical analysis originally done by Berkeley's own Prof. David Card on the effects of immigrants on labor markets. If we want to discover the causal effect of adding immigrants, then it's not enough to look at whether places with more immigrants have different rates of unemployment or wages. This is because immigrants choose their destination – and will all other things equal pick areas with strong labor markets. So, typically we will be unable to distinguish the causal effect of immigrants on labor markets from the reverse effect of labor markets on choice of place to immigrate to. Prof. Card's insight was that there are certain circumstances in which a wave of immigrants arrives in a way that has little or nothing to do with labor market conditions in the destination. The “Mariel boatlift” was such an event. Here is his description

The experiences of the Miami labor market in the aftermath of the Mariel Boatlift form one such [“natural”] experiment. From May to September 1980, some 125,000 Cuban immigrants arrived in Miami on a flotilla of privately chartered boats. Their arrival was the consequence of an unlikely sequence of events culminating in Castro's declaration on April 20, 1980, that Cubans wishing to emigrate to the United States were free to leave from the port of Mariel. Fifty percent of the Mariel immigrants settled permanently in Miami. The result was a 7% increase in the labor force of Miami and a 20% increase in the number of Cuban workers in Miami. (Card, 1990:245-6)

You should download a copy of the original paper and read at least pages 245-251 and pages 255-257. It is available at (davidcard.berkeley.edu/papers/mariel-impact.pdf)

Our goals in this lab are

1. To see if we can replicate Prof. Card's findings on the effects of this sudden influx of immigrants on the unemployment and wage rates of natives of different racial and ethnic groups. Even if we are not able to replicate the exact numbers of the original study, we can see if we can replicate the essential findings.
2. To see how a “control” group is useful even for natural experiments
3. To do a new analysis of the effects broken down by education.

The Current Population Survey “outgoing rotation groups” that we are using for the analysis is the largest sample available for this time period. Still, once we limit ourselves to Miami and the comparison cities, the sample sizes are still small.

Note: some of the “hints” in the auto-correct functions suggest running some additional R-code. You can do this by pasting the hint into the bottom of the last chunk of code and rerunning the chunk.

```
# Do not edit this chunk, but *do* press the green button to the answer key for the quiz info (the unre
tot = 0
answer.key = "eJytlk1v2zgQhu/9FYNgAdmAo03s+GML7CEJFuge0kNbo0iRosYS1xSp8MOK/31nKNmtqu7NF80kx0G8z7xDKn81s
library(quizify)
source.coded.txt(answer.key)
```

The data

We have combined the CPS files from 1979 to 1985 and put the relevant variables into a clean version of the file, which you can read-in below.

```
df <- read.csv("/data175/mariel_clean.csv")
esr <- df$esr # employment status
ftpt79 <- df$ftpt79 # full-time / part-time
ethrace <- df$ethrace ## combination of race and ethnicity
educ <- df$educ
year <- df$.id+1978 # years 1979, 1980, ..., 1985 for each record
weight <- df$weight # we are going to not use the weights (for simplicity)
earnwt <- df$earnwt # this is the weight one would use if you were using weights (we're not)
earnhr <- df$earnhr
earnhre<-df$earnhre # this is the wage variable we use
age <- df$age
smsarank <- df$smsarank # names of "standard metropolitan statistical areas" (e.g. Miami)
## some vectors for the categories we're interested in
ethrace.vec <- c("whites","blacks","cubans","hispanics") # good for ordering output
year.vec <- 1979:1985 # good for plotting
```

Q1.1 What categories are there in the “ethrace” variable?

A. White and Black B. White and Black and Cuban C. White, Black, Cuban, Hispanic, and Other D. Cuban, Hispanic, non-Hispanic?

```
## "Replace the NA with your answer (e.g., 'A' in quotes)"
answer1.1 = "C"
quiz.check(answer1.1)
```

```
## Your answer1.1 : C
## Correct.
```

Explanation: The original survey asked about race and Hispanic origin separately: we've combined the

Q1.2 What are the units of the earnhre variable?

A. 1980 dollars per hour B. 1980 cents per hour C. Nominal cents per hour (not adjusting for inflation) D. Nominal dollars per hour (not adjusting for inflation)

```
## "Replace the NA with your answer (e.g., 'A' in quotes)"
answer1.2 = "C"
quiz.check(answer1.2)
```

```
## Your answer1.2 : C
## Correct.
```

Explanation: Yes, these appear to be cents per hour. Card adjusts for inflation in his tables. We w

Q1.3 What cities make up the comparison group?

A. All U.S. cities except Miami B. All Florida cities except Miami C. Cities around the country that Card thought would be subject to the same macro-economic influences as Miami but that didn't receive many Cuban immigrants. D. Cities that also received a lot of Cuban immigrants.

```
## "Replace the NA with your answer (e.g., 'A' in quotes)"
answer1.3 = "C"
quiz.check(answer1.3)
```

```
## Your answer1.3 : C
## Correct.
```

```
## Explanation: This is the idea of a control group, which differs only in the variable of interest.
```

Attempt to replicate results for unemployment

We're going to begin by trying to replicate Card's results for unemployment in his Table 4. (We'll do wages in Table 3 later).

The goal here is 3-fold.

1. To compare over time to see if we can detect a change following the influx of Cuban migrants.
2. To compare across racial and ethnic groups to see if groups that we might assume are more likely to be in portions of the labor market influenced by the Cubans are influenced more by the influx of Cuban migrants.
3. To compare the changes in Miami with the comparison cities to try to isolate the effect of the migrant influx per se, separating it from the nationwide changes in the economy. (A large recession began in 1981, as a result of intentional monetary contraction, trying to slow inflation. See https://en.wikipedia.org/wiki/Early_1980s_recession#United_States)

Note: our replication will be slightly different than Card's original analysis for several reasons. First, we will not use the sampling weights. (We experimented with these but it did not make the replication much closer and so we set them aside for simpler programming – not recommended for research, but ok here for our educational purposes). Second, the selection of cases may not be the same. He removed some of the misreported wage data, but we don't. Third, while Card worked with CPI deflators, for simplicity we're going to use nominal wages (divided by 100) rather than deflated wages. (This should be ok, since we will be comparing Miami to the comparison cities, year by year.)

Programming background

We're going to use the R-function `tapply()`, which is similar to “pivot tables” in Excel. It will calculate values for each specific combinations of indices (e.g., whites in 1981) and return a table of the values for all combinations.

Here's an example of the logic, taking the mean of `x` by sex

```
x <- 1:4
sex = c("m", "m", "f", "f")
height = c("short", "tall", "short", "tall")
print(x)
```

```
## [1] 1 2 3 4
```

```
print(sex)
```

```
## [1] "m" "m" "f" "f"
```

```
print(height)
```

```
## [1] "short" "tall" "short" "tall"
```

```
average.by.sex.table <- tapply(X = x,
                               INDEX = list(sex),
                               FUN = mean)
print(average.by.sex.table)
```

```
##    f    m
## 3.5 1.5
```

Try to create the “average.by.height” using `tapply()`. You should get

short tall 2 3

(You can also check this by hand.)

In our case, we’re going to tabulate by year and ethrace, so we’ll “list” two elements when we specify the INDEX. In addition, we’re going to use the custom function `get.ue()` instead of `mean()`. The final difference is that we’re going to specify a specific subset of the data, ages 16-61 and Miami/not-Miami. For this we’ll use the TRUE/FALSE index “s”, and we’ll index our arguments using “[s]”. Note, we’ll reuse the same letter “s” for different subsets. So rerun each chunk in its entirety instead of line by line, so that you can be sure that the value R is using for “s” is correct.

```
get.ue <- function(esr)
{
  ## our function for calculating unemployment rate
  ## from esr variable
  sum(esr == "Unemployed-Looking") /
    sum(esr %in% c("Unemployed-Looking",
                  "Employed-At Work",
                  "Employed-Absent"))
}

# The unemployment rate for Miami
s <- smsarank == "Miami" & age %in% 16:61 #as the true/false index
ue.table.miami <- tapply(X = esr[s],
                        FUN = get.ue,
                        INDEX = list(ethrace[s], year[s]))
## express in percent and round to 1 digit
ue.table.miami.pretty <- round(100*ue.table.miami[ethrace.vec,], 1)
print("Unemployment Rates, as in Card's Table 4")

## [1] "Unemployment Rates, as in Card's Table 4"
print("Miami")

## [1] "Miami"
(ue.table.miami.pretty) # new trick: outer parentheses will print contents

##           1979 1980 1981 1982 1983 1984 1985
## whites      5.0  2.8  4.1  5.3  7.1  3.6  5.0
## blacks      8.7  5.1  9.3 16.3 19.2 14.5  7.5
## cubans       5.3  7.2 10.3 11.2 13.2  7.5  5.6
## hispanics    5.6  8.0  9.7  8.8  7.7 12.4  4.4

# The unemployment rate for comparison cities (not Miami)
s <- smsarank != "Miami" & age %in% 16:61
ue.table.notmiami <- tapply(X = esr[s],
                          FUN = get.ue,
                          INDEX = list(ethrace[s], year[s]))
## remove cubans in nonmiami because there are so few we don't want to be misled
ue.table.notmiami["cubans",] <- NA
print("Comparison Cities: Not Miami")

## [1] "Comparison Cities: Not Miami"
(ue.table.notmiami.pretty <- round(100*ue.table.notmiami[ethrace.vec,], 1))
```

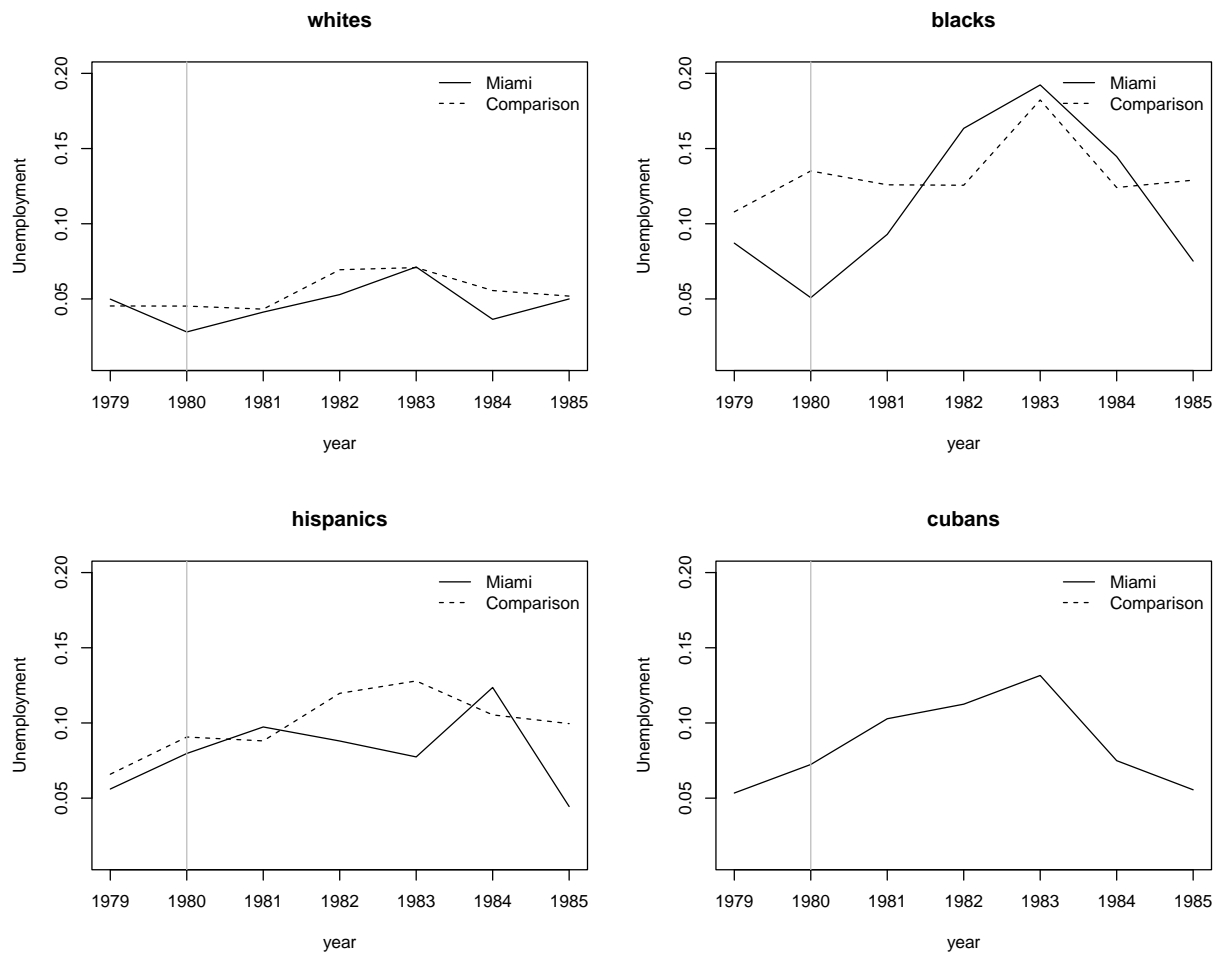
```
##           1979 1980 1981 1982 1983 1984 1985
## whites    4.5  4.5  4.3  6.9  7.1  5.6  5.2
## blacks   10.8 13.5 12.6 12.6 18.2 12.4 12.9
## cubans      NA   NA   NA   NA   NA   NA   NA
## hispanics  6.6  9.1  8.8 12.0 12.8 10.5 10.0
```

Now plot results The following code makes a function that takes in the subgroup we want to make a plot for as an input and output the plot.

```
ue.plot.fun <- function(my.ethrace)
{
  ## based on:
  ##   plot(year.vec, ue.table.miami["whites",], type = "l",
  ##   ylim = my.ylim)
  ## lines(year.vec, ue.table.notmiami["whites",], lty = 2)
  my.ylim = c(0.01, .2)
  plot(year.vec, ue.table.miami[my.ethrace,], type = "l",
        ylim = my.ylim,
        ylab = "Unemployment",
        xlab = "year")
  lines(year.vec, ue.table.notmiami[my.ethrace,], lty = 2)
  title(my.ethrace)
  abline(v = 1980, col = "grey")
  legend("topright",
        lty = c(1,2),
        legend = c("Miami", "Comparison"),
        bty = "n")
}
```

Don't forget to keep the quotation marks for inputs.

```
par(mfrow = c(2,2))
ue.plot.fun("whites")
ue.plot.fun("blacks")
ue.plot.fun("hispanics")
ue.plot.fun("cubans")
```



Question 1.4 Why do the “Cubans” have no comparison group?

- A. Because there’s a mistake in the code B. Because there are not many Cubans in the comparison cities C. Because there are many Cubans in the comparison cities and it would be confusing to include them.

```
## "Replace the NA with your answer (e.g., 'A' in quotes)"
answer1.4 = "B"
quiz.check(answer1.4)
```

```
## Your answer1.4 : B
```

```
## Correct.
```

```
## Explanation: The absence of a Cuban influx is what makes comparison cities a useful control group.
```

Question 1.5 Unemployment after the Mariel boatlift goes up for all groups. Why does Card argue that there is “There is no evidence that the Mariel influx adversely affected the unemployment rate of either whites or blacks.” (p. 250)

- A. Because our replication gives different numbers than Card’s original analysis
B. Because the increases in unemployment were also seen in cities that didn’t have the sudden Cuban migration.

```
## "Replace the NA with your answer (e.g., 'A' in quotes)"
answer1.5 = "B"
quiz.check(answer1.5)
```

```
## Your answer1.5 : B
```

```
## Correct.
```

```
## Explanation: Although our numbers are not exactly the same, the trend in Miami is not worse than in
```

Question 1.6 How much attention should we pay to the ups and downs in these graphs? Are these chance fluctuations from the sample survey (“noise”), or are they important information that we should pay attention to (“signal”)?

A. They are signal B. They are noise C. We can’t tell just by looking, but one could in theory (and with the help of a statistics course) quantify the magnitude of fluctuations that we would expect from random sampling.

```
## "Replace the NA with your answer (e.g., 'A' in quotes)"
```

```
answer1.6 = "C"
```

```
quiz.check(answer1.6)
```

```
## Your answer1.6 : C
```

```
## Correct.
```

```
## Explanation: This could be done, for example, by a technique called the 'bootstrap'.
```

Question 1.7 Is there any evidence that unemployment rates for Blacks rose because of the migrant influx?

A. Maybe, since there is such a large spike. B. Maybe, but since we would expect Hispanics to be closer substitutes, the fact that we don’t see the rise for Hispanics should make us skeptical. C. Not much because sample sizes are small. D. All of the above.

```
## "Replace the NA with your answer (e.g., 'A' in quotes)"
```

```
answer1.7 = "C"
```

```
quiz.check(answer1.7)
```

```
## Your answer1.7 : C
```

```
## Correct.
```

```
## Explanation: The increase for Blacks is bigger in Miami, but the general trend is not very different
```

Wages

Now we will try to replicate Card’s findings that the Mariel boatlift also had little or no effect on wages of natives.

For simplicity we will not deflate the wages but instead consider the nominal wages. (Note: we’re also doing this because even with the published deflators, we were unable to very closely replicate Card’s Table 3.)

```
log.w <- log(earnhre/100) # we divide by 100 to get roughly the same units as Card's deflated log wages
```

```
## Miami
```

```
s <- age %in% 16:61 & smsarank == "Miami" & !is.na(earnhre) &  
  ftpt79 %in% c("Employed full-time")
```

```
wage.table.miami <- tapply(X = log.w [s],  
                           INDEX = list(ethrace[s], year[s]),  
                           FUN = mean)
```

```
(round(wage.table.miami[ethrace.vec,], 2))
```

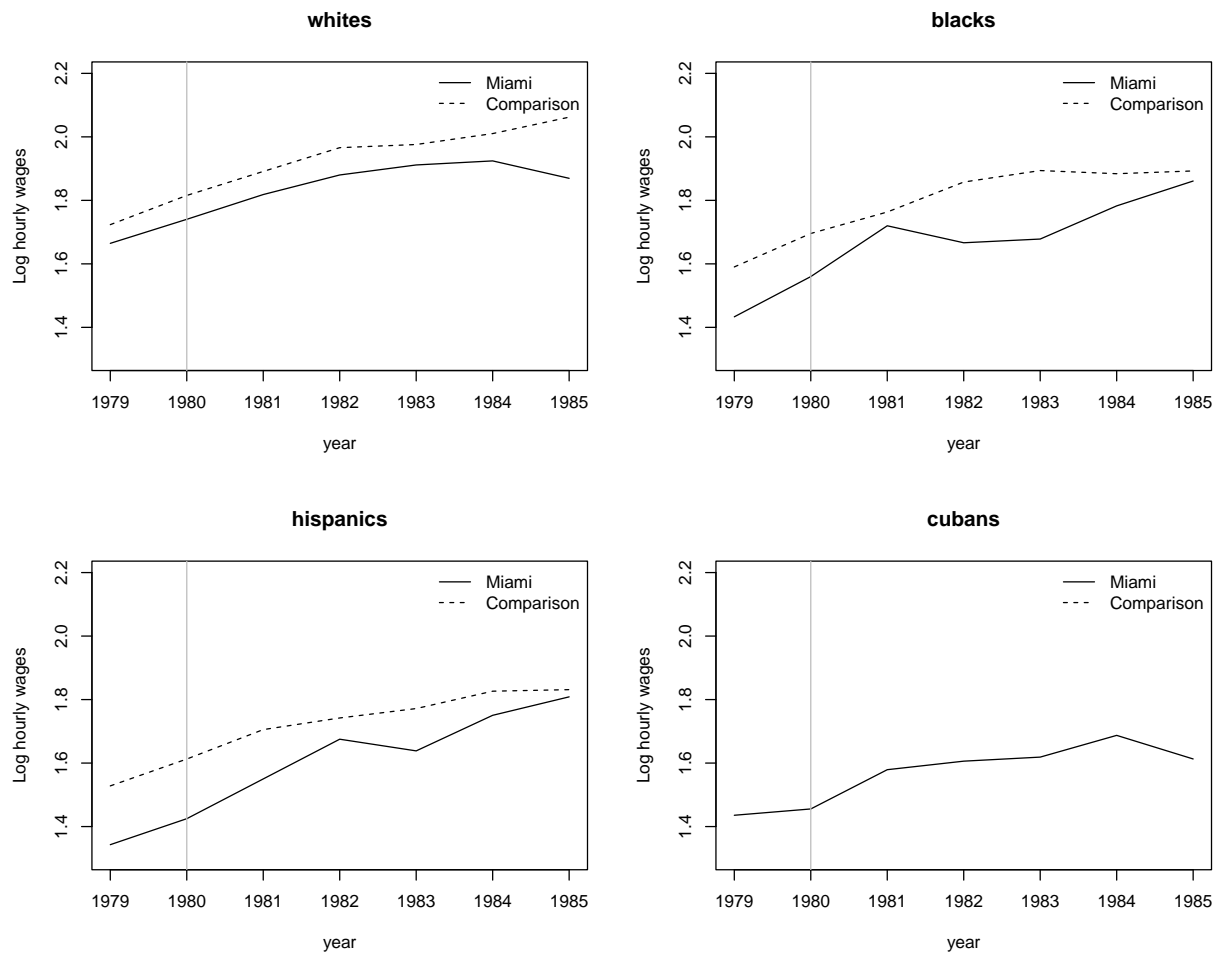
```
##           1979 1980 1981 1982 1983 1984 1985  
## whites    1.66 1.74 1.82 1.88 1.91 1.92 1.87  
## blacks    1.43 1.56 1.72 1.67 1.68 1.78 1.86  
## cubans    1.44 1.46 1.58 1.61 1.62 1.69 1.61  
## hispanics 1.34 1.42 1.55 1.68 1.64 1.75 1.81
```

```
## Not Miami
s <- age %in% 16:61 & smsarank != "Miami" & !is.na(earnhre) &
  ftpt79 %in% c("Employed full-time")
wage.table.notmiami <- tapply(X = log.w[s],
                              INDEX = list(ethrace[s], year[s]),
                              FUN = mean)
wage.table.notmiami["cubans",] <- NA
(round(wage.table.notmiami[ethrace.vec,], 2))
```

```
##          1979 1980 1981 1982 1983 1984 1985
## whites    1.72 1.82 1.89 1.97 1.98 2.01 2.06
## blacks    1.59 1.70 1.76 1.86 1.89 1.88 1.89
## cubans      NA  NA  NA  NA  NA  NA  NA
## hispanics 1.53 1.61 1.71 1.74 1.77 1.83 1.83
```

Plot results

```
wage.plot.fun <- function(my.ethrace)
{
  my.ylim = c(1.3, 2.2)
  plot(year.vec, wage.table.miami[my.ethrace,], type = "l",
        ylim = my.ylim,
        ylab = "Log hourly wages",
        xlab = "year")
  lines(year.vec, wage.table.notmiami[my.ethrace,], lty = 2)
  title(my.ethrace)
  abline(v = 1980, col = "grey")
  legend("topright",
        lty = c(1,2),
        legend = c("Miami", "Comparison"),
        bty = "n")
}
par(mfrow = c(2,2))
wage.plot.fun("whites")
wage.plot.fun("blacks")
wage.plot.fun("hispanics")
wage.plot.fun("cubans")
```

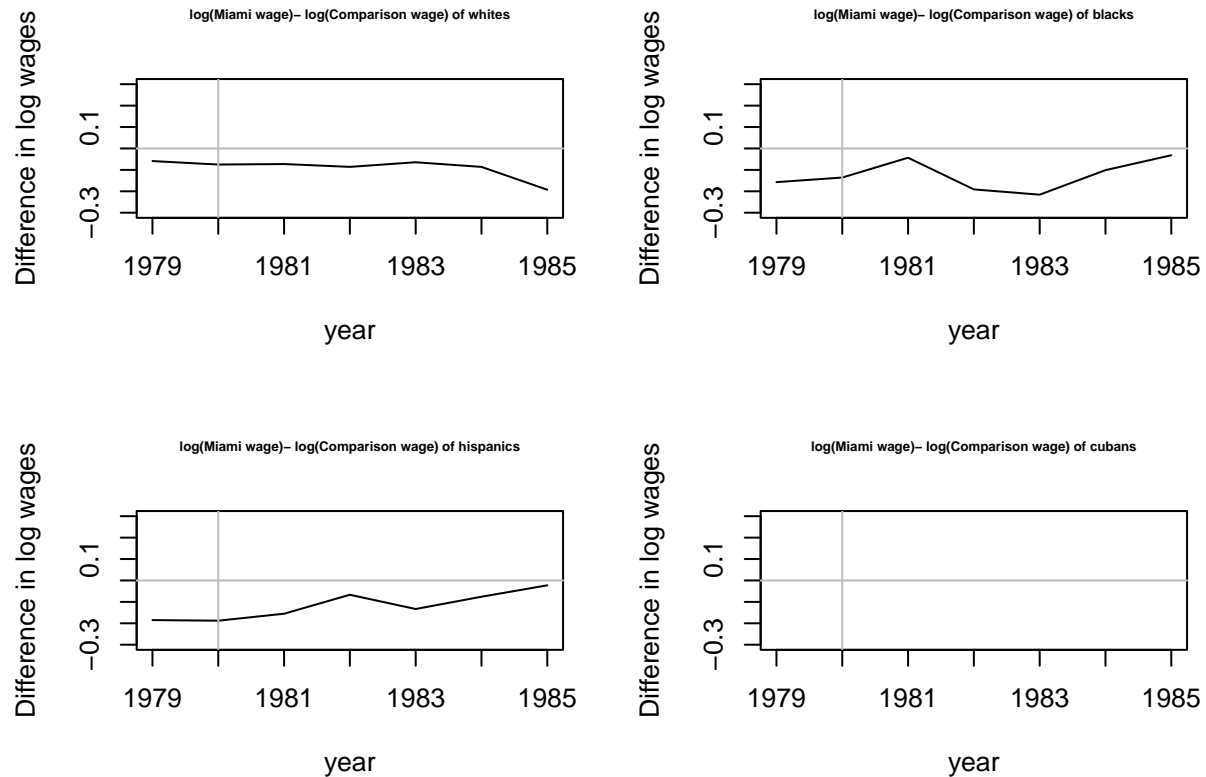



Our numbers differ from Card's Table 4 because we are not accounting for inflation. In order to make inferences about the effect of the boatlift on wages easier, let's plot the differences between Miami and the Comparison Cities.

```
diff.wage.plot.fun <- function(my.ethrace)
{
  my.ylim = c(-.3, .3)
  plot(year.vec,
    (wage.table.miami[my.ethrace,] - wage.table.notmiami[my.ethrace,]),
    type = "l",
    ylim = my.ylim,
    ylab = "Difference in log wages",
    xlab = "year")
  title(paste("log(Miami wage)- log(Comparison wage) of", my.ethrace),
    cex.main = .5)
  abline(v = 1980, col = "grey")
  abline(h = 0, col = "grey")
}

par(mfrow = c(2,2))
diff.wage.plot.fun("whites")
diff.wage.plot.fun("blacks")
```

```
diff.wage.plot.fun("hispanics")
diff.wage.plot.fun("cubans")
```



Question 1.8 If wages were hurt by the influx of migrants, we would expect this graph to show

- A. A decrease after 1980, as Miami wages went down relative to other cities
- B. Values below 0 for all periods, because Miami would always have lower wages
- C. An uptick after 1980 because we are working with logarithms.

```
## "Replace the NA with your answer (e.g., 'A' in quotes)"
answer1.8 = "A"
quiz.check(answer1.8)
```

```
## Your answer1.8 : A
## Correct.
## Explanation: Yes, but we don't see any decrease, so it appears wages were unaffected.
```

So it seems that indeed our analysis is consistent with Card's conclusion that "the Mariel immigration had virtually no effect on wages or unemployment outcomes of non-Cuban workers in the Miami labor market" (p. 255).

Education

We would expect any negative effect of the influx of immigrants to be strongest on the group that they most resemble. Because most of the Cuban immigrants in the boatlift were unskilled, we would expect the strongest

effect on natives with the least education, with perhaps the clearest comparison group being Hispanics with the least education.

Card used a different approach, looking at the effects for low-skilled workers by predicting wages based on education and years of experience. Here we do something a bit simpler, using education only. This portion will provide the material you need to answer the next graded questions.

Programming background (continued)

We can use the above approach with `tapply()`, adding an additional index, “educ”. The result will then be an array of 3 tables, one for each level of education. To index the array, we need to specify 3 dimensions $A[i,j,k]$.

Unemployment by year and education for Miami

```
s <- age %in% 16:61 & smsarank == "Miami"
ue.educ.table.miami <- tapply(X = esr[s],
                             FUN = get.ue,
                             INDEX = list(educ[s], year[s]))
print("Unemployment in Miami")
```

```
## [1] "Unemployment in Miami"
```

```
(ue.educ.table.miami.pretty <- round(100*ue.educ.table.miami,1))
```

```
##      1979 1980 1981 1982 1983 1984 1985
## BA      3.6  3.8  3.2  3.1  2.6  2.4  2.0
## HS      5.0  3.8  6.9  8.4 10.8  8.0  6.2
## lessHS  9.5  8.5 13.0 16.4 17.9 12.5  7.8
```

Graded Question 1. What happens to the unemployment rates of those with a college education (BA) between 1980 and 1982, when the effects of the Mariel boatlift should have been felt? What happens to those with the least education? (“lessHS”). Is this consistent with a large effect of immigration on the least educated? (Three sentences are fine here).

Unemployment rate decreased from 3.8 to 3.1 from 1980 to 1982. With least education, increases from 8.5 to 16.4. The increase in this case is dramatic. This is consistent with a large effect of immigration on the least educated since most incoming immigrants were not highly educated, having an effect on the natives as well.

Graded Question 2. Modify the following chunk of code and calculate the same table for the comparison cities. Does this change your answer about the effect of the boatlift? (< 50 words).

```
## not miami
## Hint: you need to change the code defining "s" below so that it specifies the comparison cities. Exam
s <- age %in% 16:61 & smsarank != "Miami"
ue.educ.table.notmiami <- tapply(X = esr[s],
                                FUN = get.ue,
                                INDEX = list(educ[s], year[s]))
print("Unemployment in Comparison Cities, not Miami")
```

```
## [1] "Unemployment in Comparison Cities, not Miami"
```

```
(ue.educ.table.notmiami.pretty <- round(100*ue.educ.table.notmiami,1))
```

```
##      1979 1980 1981 1982 1983 1984 1985
## BA      3.6  3.1  2.7  4.4  4.2  4.2  3.5
## HS      5.0  5.9  5.5  8.0  8.7  7.0  6.6
## lessHS  9.2 11.4 11.5 15.2 17.9 13.1 13.3
```

Yes, it changes my answer about the boatlift. Now, there is still an increase in least education on comparison cities for least educated people. This means that immigration is not the sole effect of the unemployment.

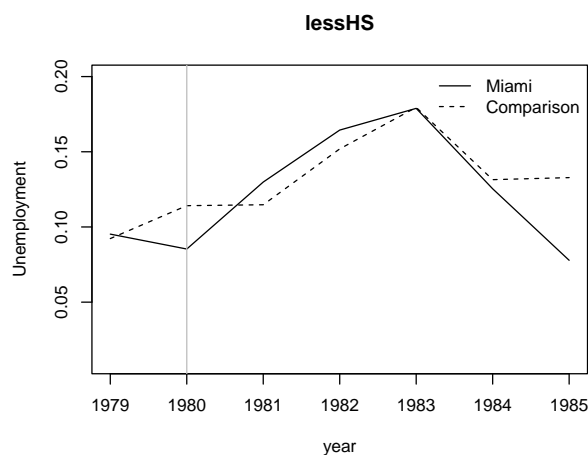
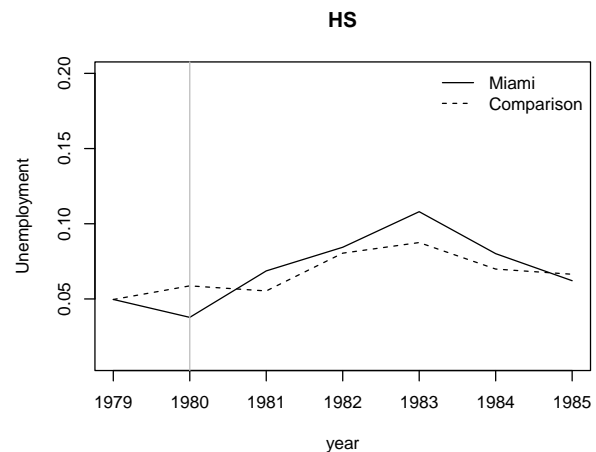
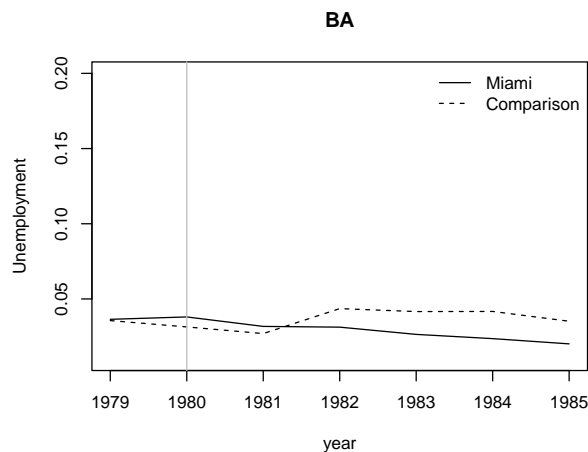
Graded Question 3. Plot the results of your educational comparison. Copy your image as the answer.

(No additional coding is required. But if you want you can add your name using the `text()` function like we did in previous labs.)

Here is code that should do this for you, as long as you have created `ue.educ.table.miami` and `ue.educ.table.notmiami` above:

```
ue.plot.fun <- function(my.index, miami.table, notmiami.table)
{
  ## based on:
  ##   plot(year.vec, ue.table.miami["whites",], type = "l",
  ##   ylim = my.ylim)
  ## lines(year.vec, ue.table.notmiami["whites",], lty = 2)
  my.ylim = c(0.01, .2)
  plot(year.vec, miami.table[my.index,], type = "l",
        ylim = my.ylim,
        ylab = "Unemployment",
        xlab = "year")
  lines(year.vec, notmiami.table[my.index,], lty = 2)
  title(my.index)
  abline(v = 1980, col = "grey")
  legend("topright",
        lty = c(1,2),
        legend = c("Miami", "Comparison"),
        bty = "n")
}

par(mfrow = c(2,2))
ue.plot.fun(my.index = "BA", miami.table = ue.educ.table.miami,
            notmiami.table = ue.educ.table.notmiami)
ue.plot.fun(my.index = "HS", miami.table = ue.educ.table.miami,
            notmiami.table = ue.educ.table.notmiami)
ue.plot.fun(my.index = "lessHS", miami.table = ue.educ.table.miami,
            notmiami.table = ue.educ.table.notmiami)
```



Graded question 4. Like Card's study, many empirical works find very small or no impact of immigration on local workers' wages and employment. Several studies even found positive impact of skilled immigration on wages and employment. What are 2 possible reasons that having immigrants would benefit the native-born workers? (< 50 words)

One reason is that immigrants may bring in new technology or innovation, which will help natives become more productive and efficient. There is also an increase supply of labor as more immigrants come, thus more opportunity and economy expansion.

Congratulations: You have finished the lab 11!