Waiting to Abort

This is a fun side project that aims to answer the following question

Question: Do pregnant 17-year-old teenagers wait until turning 18 to get an abortion?

Why this question?

We know that *18 year olds enjoy more privileges compared to their slightly younger 17-year-old peers*. 18 year olds are considered to be adults and can make a lot of decisions by themselves including getting an abortion. In a lot of places, while 17 year olds need to get their parents' consent to get an abortion, or at least inform them, 18 year olds do not need to.

However, note that *the line of being treated as an adult or as a minor can be very thin* for some people: a month, a week, or even a day if tomorrow is the birthday. So, should the soon-to-be-adult minor decide to get an abortion, depending on how early/late her pregnancy is, she *can* wait to get an abortion and avoid talking to her parents. So, do (at least) some of them wait?

Looking at this question another way, we actually try to answer *whether involving parents in abortion is a burden* to minors. If it is not a burden, then we should not see any waiting: minors tell their parents and get an abortion as they need. If it is a burden, waiting is one piece of evidence of that burden.

If we want to understand the impact of abortion law, then understanding how the law may affect different groups, not the least minors, is crucial.

I don't care about abortion. Is this project still interesting to me?

Yes. In answering this question, this project shows an estimation method that is relevant in many settings. Removing "abortion" out of the question, the method can answer questions of the type **"do people wait until time T to do X?"** where X is an interesting activity and T is a time threshold for that X. For example,

- Do people wait until the sales period to buy stuff? (X = buy, T = start of sales)
- Do people wait until the New Year to start a routine? (X = start a routine, T = New Year day)

The question can also be flipped **"do people rush to do X before time T?"**. This means that instead of looking at decisions to delay, we can look at decisions to avoid delay beyond a deadline. For example,

- Do people work more before the last day of the month? (X = work, T = last day of the month)
- Do people settle outstanding financial obligations before the end of the (tax) year? (X = financial activity, T = end of (tax) year)

And the threshold is not necessarily just time, it can also be space. We can modify the questions to **"do people cross a border B to do X?"**. For example,

- Do people cross state boundary to pump cheaper gase? (X = pump gas, B = state boundary)
- Do people cross zone boundary to board train stations with cheaper price? (X = board train, B = zone boundary)

At heart of all of the above questions is the question whether at the margin, doing X is costly. Besides, if the underlying question has a sense of the additional cost of (not) waiting (for example, the sale period has a specific cost reduction), then comparing the number of people who wait and the increase/decrease in cost gives us a measure of cost sensitivity. In this project, we cannot measure the burden of talking to your parents about abortion i dollar terms, so we cannot say anything about cost sensitivity. We provide only a yes/no answer to whether the cost is positive.

So what's the method?

I illustrate how the method works by showing how it is done in this project. I will then summarize the main points of the method below.

First, let me explain a little bit about the data. Due to the sensitive nature of this topic, I will not disclose the source of the data. I will say only the relevant features of the data for this project. It covers the universe of abortion in a territory for a few years. One thing about abortion that you may be worried about is that people may travel to get an abortion. For example, in the US, people can go to a different state where it is easier to get an abortion. We can never completely rule out this channel but for this data, this is not a major concern because of various institutional details.

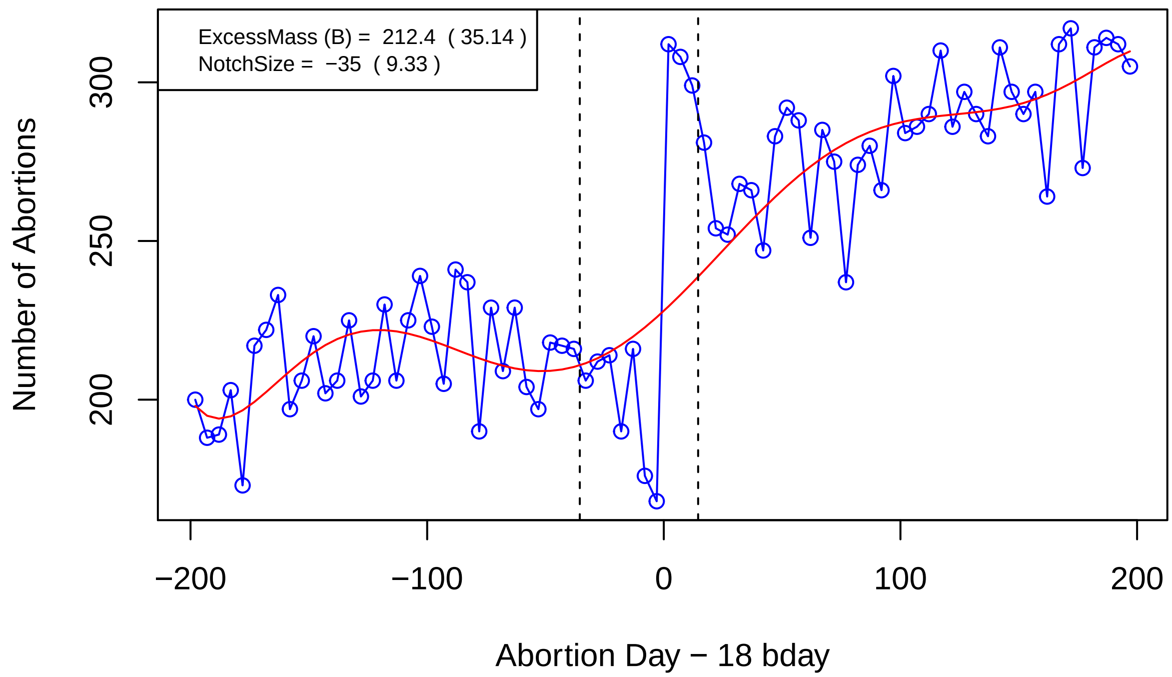
This data includes the exact birthday of the women seeking the abortion and the exact day of abortion. So it is straightforward to know whether the abortions were performed before or after the 18th birthday and how far/close away that abortion from the 18th birthday.

The below figure summarizes everything about the method and the data relevant to this question. Because I am not making the data available publicly, you will not be able to reproduce this figure. So the next best thing I can do is to show you the code. All functions are available in the folder.

```
#all functions
source("notchReg.R")
source("notchEst.R")
source("notch.R")

data <- df[["diff18"]]

ans <- notch(data,notchPoint = -0.5, binw = 5,
             regFrom = -200.5, regTo = 199.5, excludeFrom = -50.5, excludeTo = 14.5,
             polySize = 7, optimalMiss = TRUE, graph = 1, nboots = 100,
             xL = "Abortion Day - 18 bday", yL="Number of Abortions" )
```



Let's go through each feature of figure

- **The x-axis** shows the difference between the 18th birthday and the day the abortion is performed. 0 means that the woman aborts on her 18th birthday. -1 means that the woman aborts the day before her 18th birthday. 1 means that the woman aborts the day after her 18th birthday. The axis runs from -200 to 200, meaning more than 6 months before and after the 18th birthday.
- **The y-axis** marks the number of abortions.
- The blue line is the "histogram" of the number of abortions for the corresponding difference between the 18th birthday and the abortion day. The word histogram is in quotes because the counts are binned in periods of 5 days to smooth out the noises in the data. Note also that this histograms are drawn in lines instead of the usual bars to make other features of the figure more readable. From this histogram, we see more than 300 women abort on the 5 days after their 18th birthdays.
- The red line is the "counterfactual" number of abortions. This is the crux of the method.
  - To measure whether/how much minors wait to abort, we need to have benchmark. A good benchmark is to imagine a world in which there is *no reason for them to wait*, i.e., the law treats 17 year olds the same as 18 year olds. This world is called *the "counterfactual" world.*
  - Now, We need to estimate how 17 and 18 year olds would have behaved in the "counterfactual" world, i.e., the "counterfactual" number of abortions. We do this by fitting a model from regions where we do not expect any waiting to the region we expect to the region with waiting. We talk about how to delineate the regions later but it helps to first understand this conceptually with an example.
  - Example: Imagine that you are 17 year olds and want to get an abortion. Your 18th birthday is 3 months away. In all likelihood, you will not wait, not the least because the law may not allow you to abort when the pregnancy is more developed. At 3 months away, there is virtually no benefit to

waiting for you even in the current world. So I use your decision to predict what someone like you would have done 1 week from her 18th birthday *in the counterfactual world*, or in a world where there is no benefit for her to wait either.

  - So the red line is the fitted model from using data far from the 18th birthday and predict data near the 18th birthday. You can see the line follows rather closely the data points far from 0 but not around the points close to 0.

- **The dotted lines** delineate what is considered far from 0 and what is considered close to 0.
  - **The dotted lines** were estimated using an intuitive assumption: the number of "missing abortions" before the 18th birthday should equal the number of "excess abortions" after the 18th birthday.
  - In terms of estimation, we start with a guess, estimate the red line, check if the assumption is satisfied. If it is satisfied, great, we are done. If it is not, we adjust the lines until we get the assumption right to an acceptable error level. If we have too many "missing abortions", we either increase the boundary on the right of 0 or decrease the boundary on the left of 0.
- Both **the dotted line** and the red line are estimated over and over again by resampling the data.
- After getting the red line and **the dotted lines**, we find the difference between the actual number of abortions (the blue line) and the counterfactual number (the red line) in the region between 0 and the right dotted line to find the "excess mass", meaning the number of excess abortions that come from waiting. The number here is 212 abortions. We also report the "notch size", or the position of the left dotted line, to indicate how far away from their 18th birthday are women willing to wait. The answer is 35 days before, or slightly more than a month.
- The red line and **the dotted lines** are estimated again and again using resampling so that we can have a sense of the noises in the estimates. The standard errors from the re-estimates are in parentheses.

This method is called bunching estimation. You can see why it is called "bunching". It is often used in the tax literature in economics because tax rates often jump at some thresholds. You can imagine that people or businesses will adjust their tax returns to stay below a certain threshold to enjoy a better tax rate, especially if they are around the threshold. The reference of this method is: Kleven and Wassem (2013). I adapted this method to this particular data for the project.

For a quick summary of the method, here are the steps:

1. Plot the histogram to give a sense of the waiting/cross behavior at the threshold
2. Estimate the counterfactual behavior where the counterfactual world is one which does not have the threshold

   2a. Guess the boundaries that delineate the regions where we expect no waiting/crossing (~control region) and and where we expect waiting/crossing (~treated region). Fit a model from the control region to predict the behavior in the treated region.

   2b. Check if the "missing mass" and the "excess mass" balance. If yes, done. If no, adjust the boundaries and re-estimate the counterfactual.
3. Calculate the excess mass and the notch size
4. Optional: find the cost sensitivity by comparing the excess mass with the increase/decrease in cost across the threshold
5. Redo steps 2 to 4 with resampled data to find the standard errors of the estimates.