**Project: Cases of miscarriage of justice. How “blind” is the American justice system?**

**what is the subject of the project?**

The subject of my project, as discussed today, is “Cases of miscarriage of justice in US. As I was struggling to find data, I turned into Wikipedia’s reference section. In there, I think I found what I was looking for.

**what is your main research question?**

That is data about any case of wrongful conviction in the US. The answer was lying basically under two sources (see below). The main research question is “How blind the justice system of United States can be?”

**what is/are your data source(s)? please provide links**

My data sources are the following two (to begin with):

1. [The national Registry of Exonerations](https://www.law.umich.edu/special/exoneration/Pages/about.aspx)
2. [Innocence Project](https://www.innocenceproject.org/all-cases/)
3. <http://www.law.northwestern.edu/>

**how will you transform the data into your own data set: scraping, regex, etc?**

Mostly through scraping!

**how can this data set be geocoded?**

Through shapes? (I guess)

**on the map, what level of study would be displayed: Country, State, City,neighborhood?**

State, County

**what information would be displayed when you click on/rollover a country (city, etc)?**

It depends on the categories/groups we will create. For example name, race, crime, sentence…

rest

<https://www.ny.gov/>

<https://library.csi.cuny.edu/c.php?g=452337&p=3089682#s-lg-box-9507264>

<https://www.criminaljustice.ny.gov/crimnet/ojsa/stats.htm>

<https://data.ny.gov/Public-Safety/Jail-Population-By-County-Beginning-1997/nymx-kgkn>

**Project presentation**

**What you should present CONCEPTUALLY: process, outcomes, horizon**

**Process:**

**-Why did you choose the project you chose?**

The motivation for my project was a Netflix series called “When they see us”. Do yourself a favor and see that please. It’s about one of the most famous cases of miscarriage of justice in the United States. True story.

**-What did you hope to learn?**

At first I wanted to learn the number of cases of miscarriage of justice, how many wrongful convictions occurred, where and when during the recent years in the United states.

**-How did you obtain your data?**

For this reason I started my search to Wikipedia, in order to find maybe lists or catalogues, if any, with such cases. And of course there where many. But the thing is that I went to Wikipedia not to find and scrape Wikipedia’s pages (cause that is very bad practice) but I went there just to see what was inside the reference section in order to search in the original sources of information.

So there it was down there along with many other stuff, my two main data sources.

1. The national Registry of exonerations
2. And the innocence project

Basically the main dataset came from the national registry of exonerations which contain 2.468 names with cases from 1991 till today. They also had cases prior of 1991 but I decided not to deal with them at this stage. Fortunately for me my secondary source wasn’t very useful because many of the cases displayed in their website were already existed in the dataset.

**-What were the central challenges of transforming and aggregating your data?**

Acquiring the data wasn’t hard at all. My dataset was there and ready to process it. The challenge came when I had to use API’s which I don’t like them at all. I used two API’s, the NEWS API and the New York Times API. The reason why, was John’s advice on how to approach this delicate issue of wrongfully convicted individuals. That was, “we can’t just display those people in a map as dots. Let’s search for the news coverage for each of these cases, if any. In other words, let’s see how many news outlets, news sources covered these cases and see their impact. So I created a list with all the names from my dataset and made the requests in the API’s. The results were too many and too messy. I struggled to make my requests more specific but I still couldn’t get there. So I had to narrow down my sample and I sorted by the total years of conviction and got out a top\_100 names. Then I went back and used only the NEWS API (because it was the strongest) and I got back fewer results. However it was still messy with articles that although included the name that I wanted inside them, but this person was just a person who had the same name with the one that I wanted to find. So I said ok, I will create my dataframe and then I’ll try to clean it up.

I tried to do that but unfortunately there was not enough time left and I didn’t. But I will the following days.

**-The Outcome**

Now regarding the map, I will, in the future I hope, do a better version of this map, displaying interactively the:

1. Name of the person
2. gender
3. Years of conviction
4. Total news coverage