**14 September 2018**

1. Gathered a dataset with all the crimes reported in New York City between 2013 - 2015, directly from the US government (data.cityofnewyork.us)

(<https://drive.google.com/file/d/1938KG7qzdW06WUIPYFDpHxCeWKX_P2L3/view>)

1. Removed 7 columns:
   * latitude
   * longitude
   * latitude\_longitude
   * x-coordinates
   * y-coordinates
   * addresses
   * parks\_nvm (representing crimes that took place in a park)
2. Removed every row with missing data
3. Used R to filter the dataset and show only sex crimes (excluding sex crimes involving minors due to being highly triggering content). Also used grep to include and omit entries by a given string per row.
4. After all this filtering, the dataset went from over 100,000 rows and 24 columns to less than 5000 rows and 17 columns.