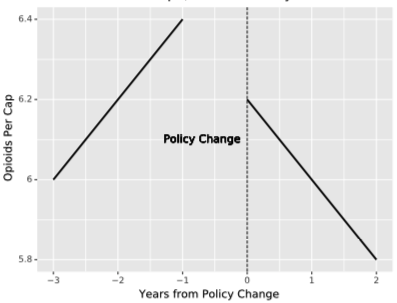
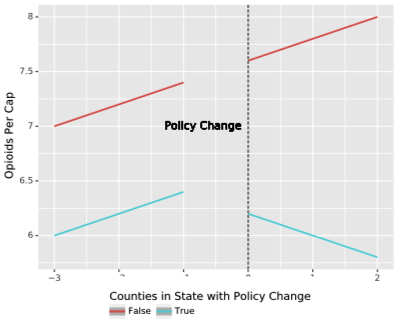
**Goal**

The goal of the project is to conduct a pre-post analysis and a difference-in-difference analysis to estimate the impact of opioid control policies in 3 different states.

A sample final output of this analysis is as follows:

Pre-Post Analysis

Difference in Difference Analysis Analysis

**Measure of Impact**

The impact of policy change will be measured in terms of the following:

1. The volume of drugs prescribed: The details on basic considerations are mentioned in the below table:

|  |  |  |  |
| --- | --- | --- | --- |
| State | Policy change year | Time period of analysis | Yearly/Monthly level |
| Florida | 2010 | 2008 - 2012 | Yearly |
| Texas | 2007 | 2006 - 2008 | Monthly |
| Washington | 2011 | 2010 - 2012 | Monthly |

1. Drug overdose deaths: The details on basic considerations are mentioned in the below table

|  |  |  |  |
| --- | --- | --- | --- |
| State | Policy change year | Time period of analysis | Yearly/Monthly level |
| Florida | 2010 | 2006 - 2014 | Yearly |
| Texas | 2007 | 2003 - 2011 | Yearly |
| Washington | 2011 | 2007 - 2015 | Yearly |

**Part 1 – Analysis on Volume of Drugs Prescribed:**

In order to be able to make plots like how it is shown under ‘Goal’ in the previous page, we would be fitting the data into a linear model. For example,



There will be 3 final datasets for each state under consideration for the 3 states of study, which would have the following columns:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **FIPS Code** | **County** | **State** | **Month** | **Year** | **Opioids Per Capita** | **Period (Pre/Post)** | **Policy State?** |
|  |  |  |  |  |  |  |  |

To get to the final dataset, we will be using various columns from the following datasets:

1. Dataset on Reporting to the US Drug Enforcement Agency (DEA):

<https://www.washingtonpost.com/national/2019/07/18/how-download-use-dea-pain-pills-database/?arc404=true>

* We will use the Washington Post version of the dataset. This dataset includes data on only two opioids – Hydrocodone & Oxycodone. These opioids however account for more than 75% of opioids prescribed.
* The COUNTY and STATE columns in the goal dataset would be gotten from the BUYER\_COUNTY and BUYER\_STATE columns.
* The MONTH and YEAR can be obtained from TRANSACTION\_DATE. Month level analysis will only be done for Washington and Texas due to lack of sufficient data post policy change. For Florida, a yearly analysis will be done.
* The OPIOD PER CAPITA can be calculated by dividing the total volume of opioid bought inside of a county divided by the population of the county. To calculate the total volume of opioid bought in a county, we multiply the “CALC\_BASE\_WT\_IN\_GM” with the “MME\_Conversion\_Factor” and sum up the value at county level.
* PERIOD: This column will be calculated based on pre and the post periods. 0 for pre and 1 for post.
* POLICY STATE: This will tell us if the state has and opioid policy enforced. This will be based on the “BUYER\_STATE” column in the data.

1. Population Data from the US Census – POPULATION:

<https://data.ers.usda.gov/reports.aspx?ID=17827&AspxAutoDetectCookieSupport=1>

* This data frame comprises of population data as collected by the US census at a county level.
* The population for the year 2010 will be the most relevant to this analysis as it is the closest to the years under consideration.
* We choose to use the population for 2010 for all the years under consideration as using rate of change of population assumes a linear rate of increase and may not be ideal for all the counties. Moreover, we do not have the rate of change of population for the years prior to 2010. This makes using 2010 population data throughout more reasonable.

1. Zip code to FIPS code data – ZIP-FIPS.

<https://www.zipinfo.com/products/cz/cz.htm>

* This dataset will primarily be used to relate the county names to the FIPS code via the zip code. We primarily do this to deal with the chance that counties may have identical names. The FIPS code will help us uniquely identify the county.
* The columns included in this data frame are County name, zip code, zip code type, County FIPS code, city name, state code, address record count.
* The FIPS Code on the goal dataset can be obtained by merging the BUYER\_ZIP column in the DEA TRANSACTION data frame with the ZIP CODE in the ZIP-FIPS data frame. We now have a unique FIPS code for each county.

**Part 2- Analysis on Drug Overdose Deaths:**

In order to be able to make plots like how it is shown under ‘Goal’ in the previous page, we would be fitting the data into a linear model. For example,



There will be 3 final datasets for each state under consideration for the 3 states of study, which would have the following columns:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **FIPS Code** | **County** | **State** | **Year** | **Drug overdose deaths per capita** | **Period (Pre/Post)** | **Policy State?** |
|  |  |  |  |  |  |  |

1. US Vital Statistics records: for data on drug overdoses <https://www.dropbox.com/s/kad4dwebr88l3ud/US_VitalStatistics.zip?dl=0>

* FIPS Code: We will use ‘county code’ column. A preceding ‘0’ may have to be added to it before we can use it.
* State: We will extract state name from ‘County’ column.
* Year: The year column is already present in the dataset.
* Overdose Death Per Cap: Divide Drug/Alcohol Induced Cause Deaths by Population of county got from the other dataset. We will be filtering for only deaths related to overdose deaths.
* PERIOD: This column will be calculated based on pre and the post periods. 0 for pre and 1 for post.
* POLICY STATE: This will tell us if the state has and opioid policy enforced. This will be based on the “BUYER\_STATE” column in the data.

1. Population data by county: for population and county FIPS codes

<https://data.ers.usda.gov/reports.aspx?ID=17827>

* This data frame comprises of population data as collected by the US census at a county level.
* The population for the year 2010 will be the most relevant to this analysis as it is the closest to the years under consideration.

**Note:** Identification of a non-policy state for comparison:

For the difference-in-difference analysis, we plan to identify a state that has a similar trend compared to the state under study in the pre-period of the study. For this, we will obtain data for each state at a county level and plot it to look for similarities in the trend.

In case we find more than one state, we will research online to figure out other related characteristics to pick a state. In case we are unable to make a distinction, we may compare the difference-in-difference with more than one state.

**Action plan**

|  |  |  |
| --- | --- | --- |
|  | **Responsible** | **Reviewer** |
| Coding for Drug overdose data | Roderick | Abhiraj |
| Coding for Drug prescriptions | Vishaal | Abhiraj |
| Identifying non-policy state | Abhiraj | Vishaal |
| Plotting and insights | Roderick | Vishaal |
| Reports | All 3 | All 3 |