BACKWARDS DESIGN STRATEGY

1.- Topic

Over the past two decades, the United States has seen a tremendous increase in the use and abuse of prescription opioids, leading not only to a huge rise in opioid addiction, but also a rise in prescription overdose deaths, and increasing deaths from non-prescription opioids like heroin and fentanyl as people who became addicted to opioids due to prescriptions turn to illegal markets to sustain their addiction.

In this project, we aim to analyze the impact of a set of policy changes implemented to reduce opioid abuse in various U.S. states on opioid drug prescriptions and drug overdose deaths.

2.- Project Questions

To analyze the effects of these policies, we want to answer the following questions:

I. *What is the impact of opioid drug prescription regulations on the volume of opioids prescribed in Florida and Washington?*
II. *What is the impact of opioid drug prescription regulations on drug overdose deaths in Florida, Washington, and Texas?*

3.- Project Hypothesis

Our hypotheses for these questions are as follows:

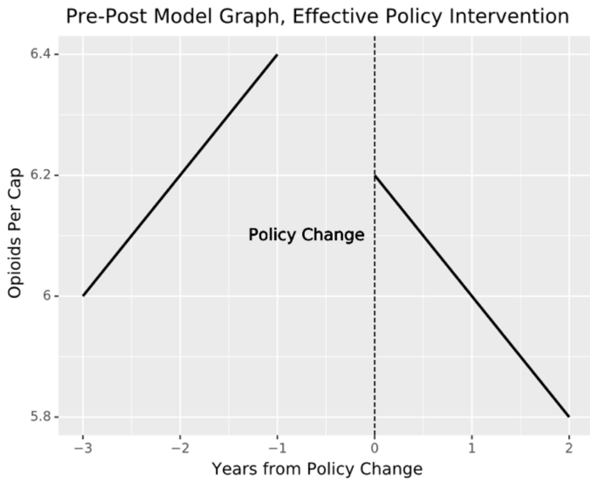
A. The opioid drug prescription regulations in Florida effective on February 2010, reduced both opioid shipments and overdose deaths.
B. The opioid drug prescription regulations in Washington effective on January 2, 2012, reduced both opioid shipments and overdose deaths.
C. The opioid drug prescription regulations in Texas effective on January 4, 2007, reduced overdose deaths.

4.- Model Results

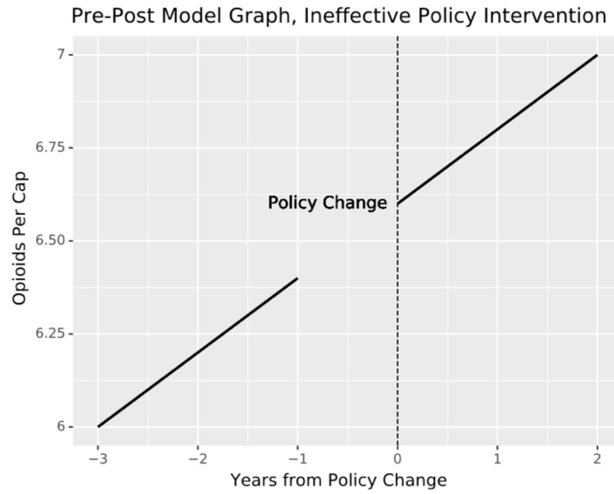
The response we are expecting in our graphs is as follows:

For the pre-post analysis:

Result if our hypothesis is **true** for each of the states: Florida, Washington, and Texas

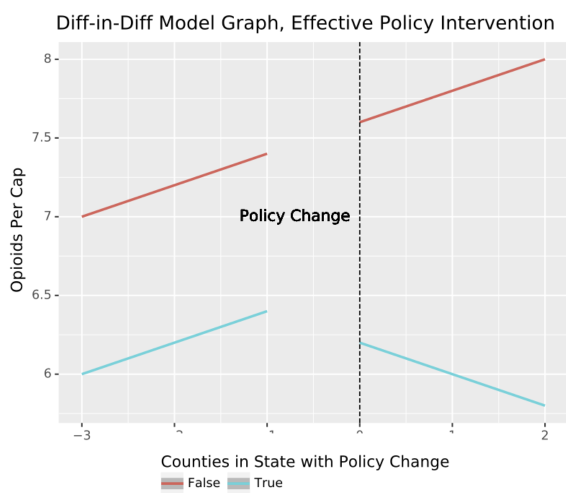


Result if our hypothesis is **false** for each of the states: Florida, Washington, and Texas

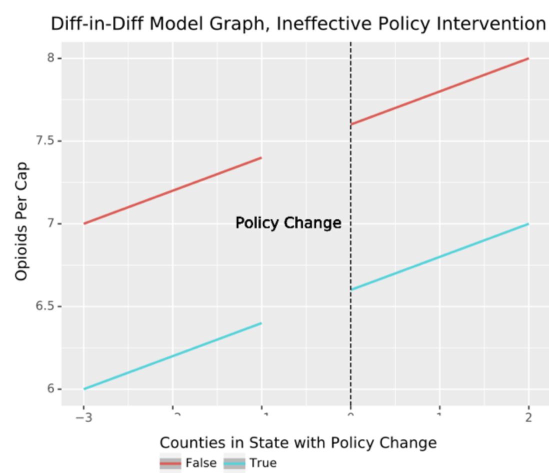


For the difference-in-difference analysis:

Result if our hypothesis is **true** for each of the states: Florida, Washington, and Texas with their respective control states.



Result if our hypothesis is **false** for each of the states: Florida, Washington, and Texas with their respective control states.



5.- Final Variables Required

Variables	Description
Year	2006-2015
County Name	Every single County in the target states (Florida, Washington, Texas) and Control States.
State Name	Target States (Florida, Washington, Texas) and Control States.
County Code	2-digit State FIPS code.
State Code	3-digit State FIPS code.
County-Level Overdose Death Rate:	County-level total overdose death rate per county population in target and control states for the year range 2006 to 2015.
County-Level Opioids Per Capita:	County Level Total Dosage Unit & Calculation of base weight in grams /County Population for Target and Control States for the year range 2006 to 2015.

The unit of observation is county by year, meaning each row in the dataset would represent a particular county within a specific year. Each row would have a unique combination of the year and county (or county code), with corresponding data for that time and place regarding overdose death rates and opioids per capita.

6.- Data Sources

First, let's look at a summary of the datasets we will be working with:

- **Vital Statistics Mortality dataset**

The Vital Statistics Mortality Dataset contains mortality data from the United States of America from 2003 - 2015 with different variables such as the information about the county, county code in each particular year, causes of the mortality and its representative code. Exploring the dataset, it was also observed that causes of mortality were grouped into 7 main categories namely.

- All other non-drug and non-alcohol causes.
- All other alcohol-induced causes
- Unintentional drug poisonings (overdose)
- Suicidal drug poisonings (overdose)
- Undetermined drug poisonings (overdose)
- All other drug-induced causes.
- Alcohol poisoning and overdose.

Alcohol poisoning isn't in 2006 and it was also observed that in 2015 there were 5 recorded cases of Homicide caused by drug poisoning overdose.

- **Opioid Prescriptions dataset**

This large dataset is about opioid drug shipments in the US from 2006 to 2019. The dataset has 33 variables in total. 10 columns pertain to the reporting entity, which is the seller that reported to the DEA. These columns provide identifiers such as DEA numbers, names, and zip codes. Another set of 10 variables concerns the buyer, indicating the destination of the shipments. The key variables for our analysis are buyer county and buyer state, as our dataset is organized at the county level, allowing for a singular observation per county. Additionally, the transaction date variable is crucial for assessing temporal patterns in drug shipments and the impact of policy changes over time. We will utilize Python's datetime capabilities to compile the data into annual aggregates. Variables like dosage unit and the total active weight of the drug in the transaction, in grams are also significant, as they will enable us to quantify the influence of policies on drug distribution volumes.

Just to summarize, here are the important variables for our analysis, along with their definition:

1. “BUYER_COUNTY”: County of entity receiving shipments from the reporter.
2. “BUYER_STATE”: State of entity receiving shipments from reporter.
3. “TRANSACTION_DATE”: Date shipment occurred.
4. “DOSAGE_UNIT: DEA calculated field indicating the number of pills, patches, or lozenges, among others, shipped as part of the transaction.
5. “CALC_BASE_WT_IN_GM”: DEA added a field indicating the total active weight of the drug in the transaction, in grams.

- **County Population dataset**

We selected county population data as our third dataset based on the project guidelines, which note that U.S. county populations can range significantly, from as few as 10,000 to as many as 1 million residents, and that county boundaries are often drawn arbitrarily. To ensure our analysis of the two outcome variables is normalized and comparative, we require precise population figures for each county across the relevant years.

We sourced this data from the National Historical Geographic Information System (NHGIS), which provides comprehensive population statistics for U.S. counties from 2006 to 2015. The NHGIS allows us to locate data across different geographic divisions, facilitating our research.

For methodology, we navigated the NHGIS Data Explorer to gather the necessary datasets year by year. Where we encountered missing values for certain counties in some years, we employed moving averages and linear extrapolation techniques to achieve a complete dataset for our analysis.

The variables included in our population dataset are as follows:

1. 'GISJOIN': This is an internal nhgis identifier variable.
2. 'STATE_CODE': Code for the US State.
3. 'COUNTY_CODE': Code for the US County.
4. 'COUNTY': Name of the County in the US.
5. 'STATE': Name of the State.
6. ‘POPULATION’: '2006', '2007', '2008', '2009', '2010', '2011', '2012', '2013', '2014', '2015'.

Plan to merge:

Our analysis utilizes three datasets, each with overlapping variables that facilitate the creation of a consolidated dataset. The vital statistics mortality dataset, spanning 2003-2015, and the opioid prescription dataset, covering 2006-2019, share common identifiers such as county and state names, allowing for a merged dataset from 2006 to 2015 when both datasets overlap.

This combined dataset provides a detailed look at drug-related fatalities and opioid distribution on a county level for all U.S. states, except for Alaska, during the specified period. We have enhanced this dataset further with a third dataset — the population records — which includes comprehensive demographic details like the two-digit state code, county code, county and state names, and annual population counts for each county from 2006 to 2015.

By integrating the mortality and prescription datasets with population data using the county code as a key, we are equipped to analyze critical metrics: the total number of county-level overdose deaths, the county-level drug-induced death rates, the total volume of opioids distributed in the country-level, and the per capita opioid distribution within each county throughout the years 2006 to 2015.

It is important to note that our analysis excludes Alaska due to its unique administrative divisions of boroughs, which differ from the conventional county system and could potentially skew the comparative analysis across other states in our dataset.

Merge process:

Our plan is to merge the Source Dataset by County, County Code, State, and State Code.

1. Merge mortality with opioids: Merge by state and county.

2. Merge with population: Merge by county code.

3. After the merge, we calculate the county-level overdose death rate and county-level opioids per capita, which gets us to the intermediate dataset.

4. We then select the most important variables from the intermediate dataset to our final dataset, ready for pre-post and diff-in-diff analysis.

Source Dataset

1. Vital Mortality dataset
   - County
   - State
   - County Code
   - Year
   - Drug/Alcohol Induced cause
   - Deaths

2. Opioids Prescription dataset
   - Buyer County
   - Buyer State
   - Transaction Date
   - Dosage Unit
   - Calculation of base weight in grams

3. Population dataset
   - County
   - County Code
   - State
   - State Code
   - Population for each year from 2006 - 2015

Intermediate Dataset

- Year: 2006-2015
- County: Every County in every state (except Alaska) in the US
- County Code
- State
- State Code
- County Population
- County-Level Total Overdose Death
- County-Level Total Dosage Unit & Calculation of base weight in grams (as a measure of county-level opioids prescription)
- County-Level Overdose Death Rate: County-Level Total Overdose Death Number/County Population
- County-Level Opioids Per Capita: County Level Total Dosage Unit & Calculation of base weight in grams /County Population

Intermediate Dataset

- Year: 2006-2015
- County: Every County in every state (except Alaska) in the US
- County Code
- State
- State Code
- County Population
- County-Level Total Overdose Death
- County-Level Total Dosage Unit & Calculation of base weight in grams (as a measure of county-level opioids prescription)
- County-Level Overdose Death Rate: County-Level Total Overdose Death Number/County Population
- County-Level Opioids Per Capita: County Level Total Dosage Unit & Calculation of base weight in grams /County Population

Final Dataset

- Year: 2006 - 2015
- County Name: Every Single County in the Target States (Florida, Washington, Texas) and Control States.
- State Name: Target States (Florida, Washington, Texas) and Control States.
- County Code
- State Code
- County-Level Overdose Death Rate
- County-Level Opioids Per Capita

7.- Division of Labor

- **Initial Code:**

Faraz: Data cleaning and merging.
Shaila & Keon: Looking for control states.
Tina: Performing pre-post and diff-in-diff analysis.

- **Code Review:**

Shaila & Keon: Data cleaning and merging.
Tina: Looking for control states.
Faraz: Performing pre-post and diff-in-diff analysis.

Data Sources:

Centers for Disease Control and Prevention. (Year). U.S. Vital Statistics [Data set]. Dropbox. https://www.dropbox.com/s/kad4dwebr88l3ud

The Washington Post. (2019, July 18). How to download and use the DEA pain pills database. https://www.washingtonpost.com/national/2019/07/18/how-download-use-dea-pain-pills-database/?arc404=true

National Historical Geographic Information System. (n.d.). Retrieved November 5, 2023, from https://www.nhgis.org/