

## Project Strategy Outline

Team 2: Jennie Sun, Dapo Adegbile, Xiaohan Yang

The project is to estimate the impact of opioid control policies. In this analysis, we'll be analyzing three policy changes, Florida (February 2010), Texas (January 2007), and Washington (January 2012). For all three policy changes, we will perform a pre-post analysis and a difference-in-difference analysis. For the Florida case, we will analyze the effect of its policy change on both opioid shipments and overdose death. For Texas and Washington, we will only analyze overdose deaths for the time being.

## Variables Needed

In order to follow a "backward design" organization scheme, we will first figure out the datasets and variables we need in the data. Based on our questions of interest, a single row of our final data should look close to this:

Year	State	County	FIPS	County Pop/year	# of opioid prescribed/100 ppl	# of opioid prescribed/ county population	# of deaths 100,000 ppl	Deaths/county population
2019	FL	Alachua	12001	269,043	65	174,877	11	30

Therefore, the variables we will need for the analysis include: *Year, State, County, FIPS Codes, County Population for each year, Number of opioids prescribed, Number of opioids prescribed per cap, Number of deaths, Number of deaths per cap.*

One thing to note is that the opioids dataset contains different kinds of opioids that have different strengths, so this may be another potential variable to consider in the unit of observation. As we look more in-depth to the data, we may modify our framework above for better analysis.

## Datasets Needed & Sources

Currently, we are given two datasets:

- Mortality by Year by State by County from Drug/Alcohol Induced Cause
- Opioids shipments/prescribed by State by County

In order to construct a final dataset with the row entries similar to the one provided above, we will also need these two datasets below:

- County population by Year (US Census Bureau)
  - **County Population Totals: 2010-2019 :**  
[https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html#par\\_textimage\\_739801612](https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html#par_textimage_739801612)
  - **County Population Totals: 2000-2010:**  
<https://www.census.gov/data/tables/time-series/demo/popest/intercensal-2000-2010-counties.html>

- FIPS Codes by County (US Census Bureau)
  - **2013 FIPS Codes (State, County, MCD, and Incorporated Place):**  
<https://www.census.gov/geographies/reference-files/2013/demo/popest/2013-geocodes-state.html>

In order to quantify and compare the use of opioid drug pre and post policy, we will be looking at opioids per cap which describes the opioid prescription per capita. To calculate this variable, we need another dataset which contains the population data by county. Since there are likely many counties with the same names across different states, we also need FIPS codes to uniquely identify each county. Also, the Vital Statistics Mortality Data is ranged from the year 2003 to 2015, so we will need a dataset that covers this time range.

One of the dataset that satisfies our requirements above is the *Annual Estimates of the Resident Population for Counties* data ranging from 2000 to 2019 from the US Census Bureau. It contains the annual population estimate, generated from the 2000 and 2010 census, for each county in the US.

## Data Merging

- **Merge FIPS code data into county population data (geography data)**
  - In order to get our data into one manageable dataset, we plan on finding a dataset with Counties and their corresponding FIPS codes. This is necessary in guaranteeing we're referencing the correct county and state because it's not uncommon for two distinct counties to have the same name. We plan to merge the FIPS code data into the county data on the county name column. We will refer to this dataset as geography data.
- **Merge Geography to Mortality data**
  - It's necessary to merge the Mortality data (deaths by overdose per county) into our geography data, on the "County" column so we can create a new variable, "Deaths per Capita", where we divide the number of drug overdose deaths by county population number.
- **Merge Geography to Shipment data**
  - We need to merge geography data to shipment data to be able to uniquely identify counties on a consistent basis. We plan on merging these on the "county" name in geography data and "buyer\_county" column in shipment data.

## Task Delegation (Temporary)

- Data Cleaning -- Jennie
- Data Merging -- Dapo
- Data Wrangling and Analysis -- Together
- Plotting -- Xiaohan
- Reports Writing -- Together