**Mid-Semester Project Outline**

Tasks before submitting this:

* Review
* Write a summary of the datasets we plan to use (Maybe not needed??)
* ~~Summarize how we plan to merge the datasets (summarize section 4/5)~~
* ~~Re-format this document~~

**Objective**

* Measure the effectiveness of opioid drug prescription regulations on the annual volume of opioids shipped to Florida and the annual drug overdose deaths in 3 states; Florida, Texas and Washington.
* The two approaches we are planning to utilize for comparing the results before enactment of policy to the results after enactment of policy are *Pre-Post Comparison* and *Difference-in-Difference.*

***Annual Drug Overdose Deaths***

* **Variables** in final dataset (*What variables will I need in this data?*)
  + - State: Extracted from County FIPS code of US Vital Statistics
    - Year: *‘Year’* column from US Vital Statistics
    - County: Extracted for US Vital Statistics
    - Overdose Deaths: Summed up the number of deaths caused by drug overdose for each State, Year and County combination.
    - Total Deaths: Total deaths (all causes)
    - Overdose Prop: Proportion of deaths caused by drug overdose.
    - Policy State: Indicate if the given county is in a state that experienced a policy change.
    - Post: Indicate whether we are in a period 12 after implementation of the policy change

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Year** | **State** | **County** | **Overdose Deaths** | **Total Deaths** | **Overdose Prop** | **Policy State** | **Post** |
|  |  |  |  |  |  |  |  |

* **Timeframe** (*What sample (what years, what counties, etc.) needs to be covered in this data?*)

|  |  |  |  |
| --- | --- | --- | --- |
| **State** | **Policy Enactment Year** | **Timeframe** | **Level** |
| Florida | 2010 | 2007-2013 | Annual |
| Texas | 2007 | 2004-2010 | Annual |
| Washington | 2012 | 2009-2015 | Annual |

* **Dataset Used** 
  + Drug Overdose Deaths Data 🡨US Vital Statistics
    - County Code: We will examine the data for each County FIPS code of the targeted states. The reason why we chose County FPIS as the identifier of each row of data is because the code contains information on the state and county.
    - Year: We will only examine years that are in our specified time window, which is differ by states.
    - Drug/Alcohol Induced Cause: We will only examine deaths resulted from drug poisoning (overdose).
    - Deaths: We will examine the number of deaths caused by drug overdose
* **Data Transformations**
  + - Group deaths by year, county, and death type, counts death type
    - Create 4th column of total deaths

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Year** | **State** | **County** | **Death Type** | **Overdose Deaths** | **Total Deaths** |
|  |  |  |  |  |  |

* + - Calculate overdose/total deaths

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Year** | **State** | **County** | **Death Type** | **Overdose Deaths** | **Total Deaths** | **Overdose Prop** |
|  |  |  |  |  |  |  |

* + - Average over state

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Year** | **State** | **Death Type** | **Overdose Deaths** | **Total Deaths** | **Overdose Prop** | **Policy** | **Post** |
|  |  |  |  |  |  |  |  |

* + - Create indicator for policy state or post regulation

***Annual Volume of Opioids Shipped***

* **Variables** in final dataset (*What variables will I need in this data?*)
  + - Buyer State: Extracted from shipment data
    - Buyer County: Extracted from shipment data
    - Year: Extract from shipment data (Extract from date column)
    - Total weight shipped: Summed up the shipment weights for each State, Year and County combination.
    - Total Deaths: Total deaths (all causes)
    - Population Estimates: Extracted from the census data
    - Policy State: Indicate if the given county is in a state that experienced a policy change.
    - Post: Indicate whether we are in a period 12 after implementation of the policy change

|  |  |  |  |
| --- | --- | --- | --- |
| **State** | **Policy Enactment Year** | **Timeframe** | **Level** |
| Florida | 2010 | 2008-2012 | Annual |
| Texas | 2007 | 2006-2008 | Monthly |
| Washington | 2011 | 2010-2012 | Monthly |

* **Timeframe** (*What sample (what years, what counties, etc.) needs to be covered in this data?*)
* **Datasets Used** 
  + Opioids Shipment Data 🡨Washington Post
  + 2000-2010 Population Estimates 🡨 US Census
  + 2010-2019 Population Estimates 🡨 US Census
* **Data Transformations**
  + - Convert date column to timestamp 🡪 extract year
    - Group shipments by year, state, county

|  |  |  |  |
| --- | --- | --- | --- |
| **Buyer State** | **Buyer County** | **Year** | **Total Weight Shipped** |
|  |  |  |  |

* + - Aggregate weight for target state/county
    - Population dataset: Aggregate population for state

|  |  |  |  |
| --- | --- | --- | --- |
| **State** | **County** | **Year** | **PopEst.** |
|  |  | **2004 PopEst.** |  |
|  |  | **…** |  |
|  |  | **2009 PopEst.** |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **State** | **County** | **2004**  **PopEst.** | **…** | **2009**  **PopEst.** |
|  |  |  |  |  |
|  |  |  |  |  |

Population Data

Stack

* + - Same with 2010-2019 dataset and append together
    - Merge PopDataset and Shipment on year, target state and county

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Year** | **Buyer State** | **Buyer County** | **Total Weight Shipped** | **PopEst.** |
|  |  |  |  |  |

* + - Calculate weight/pop

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Buyer State** | **Buyer County** | **Year** | **Weight Shipped** | **PopEst.** | **Weight per Cap** | **Policy** | **Post** |
|  |  |  |  |  |  |  |  |

* + - Average over state
    - Create indicator for policy state or post regulation

***Tasks***

* **Finding population dataset**
* 2010-2019: <https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-total.html>
* 2000-2010: <https://www.census.gov/data/tables/time-series/demo/popest/intercensal-2000-2010-counties.html>
* **Code for cleaning and transforming shipment dataset**
* Tasked to: Dean
* Reviewed by: Malcolm
* **Code for cleaning and transforming population dataset**
* Tasked to: Malcolm
* Reviewed by: Wilson
* **Code for cleaning and transforming deaths dataset**
* Tasked to: Wilson
* Reviewed by: Dean
* **Code for merging shipment and population datasets**
* Interpolating/extrapolating population values for individual years
* Tasked to: Dean
* Reviewed by: Malcolm
* **Code for plotting and saving shipment plots**
* Tasked to: Malcolm
* Reviewed by: Wilson
* **Code for plotting and saving death plots**
* Tasked to: Wilson
* Reviewed by: Dean
* **Writing introduction**
* Tasked to: Dean
* **Writing methods**
* Tasked to: Malcolm
* **Writing results/conclusion and limitations**
* Tasked to: Wilson

An outline of your project strategy is due on October 13th.

* *In writing this strategy, use a “backwards design” organizational scheme (we’ll discuss these much more in the future, but for now you get to explore yourself!):*

*Start by establishing what you want to achieve (i.e. the plots I’ve already specified).*

Pre-post analysis, and difference in difference analysis (Plots)

*Then ask: What dataset do I need to make these plots? Or more specifically:*

Datasets:

Pre and post data for each target state and at least 3 control states (for each target state):

Overdose deaths and opioid shipments

Data on the population and demographics for these counties

* **1. What variables will I need in this data?**

*Pre-post analysis: (For target state/county)*

Opioids shipments per cap vs year – Agg(Weight of opioid shipments/ population) vs year

Overdose deaths per cap vs year – Agg(Num overdose deaths/county deaths) vs year

*Difference in difference analysis:*

In addition to the data for the pre-post analysis

* For the control states/counties

Opioids shipments per cap vs year – Agg(Weight of opioid shipments/ population) vs year

Overdose deaths per cap vs year – Agg(Num overdose deaths/county deaths) vs year

* **2.** **What sample (what years, what counties, etc.) needs to be covered in this data?**

*Policy change years:*

Florida: 2010

Texas: 2007

Washington: 2011

*Opioid Prescriptions*

Florida: 2006-2009, 2011-2012

Texas: NA

Washington: 2006-2010, 2012-?

Overdose deaths

Florida: 2003-2009, 2011-2015

Texas: 2003-2006, 2008-2015

Washington: 2003-2010, 2012-15

*Counties*:

We will conduct our analysis only on counties that appear in the data set every year in our desired rage.

* **3. What should a single row of this data look like (i.e. what’s a unit of observation?)**

*Deaths*: State, Year, (Grouped Counties?), Overdose deaths, Overall deaths, Prop overdose deaths

*Shipment*: State, Year, Population, Total Weight Shipped, Weight/pop

* **4. Then step back and repeat the process: to get this dataset, what do I need to do?**
* **5. What do I need my input datasets to look like to get to this final analysis dataset?**
* Then again: to get to these intermediate datasets, what source datasets do I need? What variables do they need?

*Shipment*:

Group shipments by year and state

Aggregate weight for target state -> year and weight for target state

PopDataset: Aggregate population for state

Merge PopDataset and Shipment on year and target state

Calculate weight/pop

*Deaths*:

Group deaths by year, county, and death type, counts death type

Create 4th column of total deaths -> year, county, death type, dt counts, total death count

Calculate overdose/total deaths

* Then start trying to assign the tasks implied by that analysis to people. (OBVIOUSLY this will only be very approximate!! But it’s good to think about this before you start, then adapt as you move forward).

This should include a summary of the datasets you plan to use, a summary of how you plan to merge these datasets, who will be responsible for writing initial code for each step, and who will review each set of code.