Estimate the Impact of Opioid Control Policies

Statistician Version

Team 2

Jingyi Wu, Zifan Peng, Shota Takeshima, Yu Gu

Nov 13, 2019

Motivation

- **Overview**

The increase in the use and abuse of opioids has caused tremendous prescription overdose deaths. As a result, the U.S. government has established new policies to control opioid over-prescription, thus reducing the death caused by drug overdose.

However, how much effect these policies created remains a question. In this project, we want to test whether these policies are really effective, that is, did these policies reduce opioid prescription and mortality from drug overdose significantly? Hence, we decided to make some statistical model to solve this problem.

- **Research Design**

Since we can't create a "parallel" world where there were no such policies, we need to test the effectiveness of policy based on real world data. Hence, we use data of county level, which is the administrative unit of U.S. To make our research complete and convincing, we do both pre-post analysis and difference-in-difference analysis, to test the change of opioid prescription and mortality from drug overdose from different perspectives.

Data Section

- **Population Data**

Because of the potential that areas with large population tends to have more drug shipments and higher drug-related mortality, we have to normalize the outcomes of interests by population at county level.

We use the Population Change for Counties in the United States and for Municipios in Puerto Rico: 2000 to 2010 (CPH-T-1) from the United States Census Bureau. The data can be found here. The data is on county level and variables are county name, state name, population in 2000 and 2010, change from 2010 to 2010 in magnitude and percentage.

Notice that the US only has a census every ten years and given that the population annual change is negligible when normalizing the deaths, we only use population in 2010 to normalize for the whole window (2004-2015). If one wants to get more precise results, the United States Census Bureau also provides predicted population annually on county level.

The normalization needs a connection on shipments/mortality data and population data. Therefore, the identifier should be multiple combining state and county, since there could be counties in different states having the same name.

- **Opioid Shipments Data**

1. Dataset Selection

The main dataset we used is maintained by the Drug Enforcement Administration that tracks the path of every single pain pill sold in the United States in every town and city. In order to analyze the effect of policy change in Florida, we conducted both pre-post analysis and difference-in-difference analysis. For the pre-post analysis, we used the shipment data for florida on county level, while for the difference-in-difference analysis, we picked five states, Louisiana, Mississippi, South Carolina, Alabama and Georgia, which are most proximate to Florida geographically.

Besides, in order to make the opioid shipments quantity of the three states in the same statistical level, we conducted normalization by population.

Therefore, the basic datasets are:

- Statewide opioid-shipment data for Florida from 2006 to 2012
- Statewide opioid-shipment data for Louisiana from 2006 to 2012
- Statewide opioid-shipment data for Mississippi from 2006 to 2012
- Statewide opioid-shipment data for South Carolina from 2006 to 2012
- Statewide opioid-shipment data for Alabama from 2006 to 2012
- Statewide opioid-shipment data for Georgia from 2006 to 2012
- Population dataset for all counties in 2010

  Download opioid-shipment data here

2. Data Engineering

- Variable selection
  - For the opioid shipment data, we chose six related variables:
    i. **TRANSACTION_DATE**: Date shipment occurred. Use this variable to represent time trend.
    ii. **BUYER_STATE**: State of entity receiving shipments from reporter
    iii. **BUYER_COUNTY**: County of entity receiving shipments from reporter
    iv. **CALC_BASE_WT_IN_GM**: DEA added field indicating the total active weight of the drug in the transaction, in grams.
    v. **MME_Conversion_Factor**: Morphine Milligram Equivalent, or how the specific drug compares to a morphine equivalent. Use (CALC_BASE_WT_IN_GM * MME_Conversion_Factor) to represent the total quantity of shipment.

vi. **BUYER_ZIP**: Zip code of entities receiving shipments from reporter. Use this variable to find the names of some counties who only have zip code but not the name in the data frame.

- For the population dataset, we chose three related variables:

i. **County**: county name. This is the key variable when merging population data with shipment data.
ii. **State**: state name. This is also the key variable when merging population data with shipment data on county level.
iii. **Poplulation_2010**: county-level population in 2010

- Read data

Read population data and opioid shipment data of Florida, Alabama, Georgia, Louisiana, Mississippi and South Carolina. The columns are shown below.

	BUYER_STATE	BUYER_ZIP	BUYER_COUNTY	TRANSACTION_DATE	CALC_BASE_WT_IN_GM	MME_Conversion_Factor
0	FL	33460	PALM BEACH	8182006	3.586	1.5
1	FL	33460	PALM BEACH	11292006	7.172	1.5
2	FL	33460	PALM BEACH	2062007	21.516	1.5
3	FL	33460	PALM BEACH	3012007	21.516	1.5
4	FL	33460	PALM BEACH	4162007	5.379	1.5

- Clean data

There are only some NA values in COUNTY Column. We used BUYER_ZIP of these counties to find out the county name and fill them up. All counties have population and shipment quantities after this step, so we didn't meet the situation where there are counties didn't report its shipment quantities.

- Transfer time format

The TRANSACTION_DATE is an 'int' type, for example, '08/18/2006' is written as '8182006'. In order to group by 'Year' and observe the time trend for the quantity of opioid shipment, we transferred the TRANSACTION_DATE to three columns: YEAR, MONTH and TRANSACTION_TIME. YEAR is a four-digit int type, MONTH is a two-digit int type and TRANSACTION_TIME is 'PeriodIndex' type.

	BUYER_STATE	BUYER_ZIP	BUYER_COUNTY	TRANSACTION_DATE	CALC_BASE_WT_IN_GM	MME_Conversion_Factor	TRANSACTION_TIME	YEAR	MONTH
0	FL	33460	PALM BEACH	8182006	3.586	1.5	2006-08-18	2006	8
1	FL	33460	PALM BEACH	11292006	7.172	1.5	2006-11-29	2006	11
2	FL	33460	PALM BEACH	2062007	21.516	1.5	2007-02-06	2007	2
3	FL	33460	PALM BEACH	3012007	21.516	1.5	2007-03-01	2007	3
4	FL	33460	PALM BEACH	4162007	5.379	1.5	2007-04-16	2007	4

- Add a column called "QUANTITY" and drop unrelated columns

CALC_BASE_WT_IN_GM represents the total active weight of the drug in the transaction, in grams and **MME_Conversion_Factor** is a factor representing how the specific drug compares to a morphine equivalent. Therefore, we use (CALC_BASE_WT_IN_GM * MME_Conversion_Factor) to get the morphine equivalent quantity of shipment and this variable is the shipment quantity that we want to observe.

	BUYER_STATE	BUYER_COUNTY	TRANSACTION_TIME	YEAR	MONTH	QUANTITY
0	FL	PALM BEACH	2006-08-18	2006	8	5.3790
1	FL	PALM BEACH	2006-11-29	2006	11	10.7580
2	FL	PALM BEACH	2007-02-06	2007	2	32.2740
3	FL	PALM BEACH	2007-03-01	2007	3	32.2740
4	FL	PALM BEACH	2007-04-16	2007	4	8.0685

- Merge Shipment data with population data

We can not simply compare quantity of opioid shipments of counties in different states since the population differs from each other. Therefore, we conduct normalization by using 'population'. In this way, the normalized opioid shipment quantity, which is called quantity per cap, can be used for comparison.

	BUYER_STATE	BUYER_COUNTY	TRANSACTION_TIME	YEAR	MONTH	QUANTITY	POP
0	FL	PALM BEACH	2006-08-18	2006	8	5.3790	1320134
1	FL	PALM BEACH	2006-11-29	2006	11	10.7580	1320134
2	FL	PALM BEACH	2007-02-06	2007	2	32.2740	1320134
3	FL	PALM BEACH	2007-03-01	2007	3	32.2740	1320134
4	FL	PALM BEACH	2007-04-16	2007	4	8.0685	1320134
5	FL	PALM BEACH	2007-04-27	2007	4	48.4110	1320134
6	FL	HERNANDO	2006-11-03	2006	11	3.0270	172778
7	FL	HERNANDO	2006-12-04	2006	12	3.0270	172778
8	FL	PINELLAS	2007-04-04	2007	4	2.6895	916542
9	FL	PINELLAS	2007-04-12	2007	4	5.3790	916542

- Group by state, county and year

Our unit of observation is quantity of opioid shipment per cap on county level. Therefore, we grouped the opioid data by state-county-year.

- Normalize shipment quantity by population & Add Pre-post variable

	BUYER_STATE	BUYER_COUNTY	YEAR	QUANTITY	POP
0	FL	ALACHUA	2006	82596.618688	247336
1	FL	ALACHUA	2007	95259.627977	247336
2	FL	ALACHUA	2008	114675.221922	247336
3	FL	ALACHUA	2009	141281.048185	247336
4	FL	ALACHUA	2010	150910.785876	247336
5	FL	ALACHUA	2011	145750.680899	247336
6	FL	ALACHUA	2012	127727.613022	247336
7	FL	BAKER	2006	11900.178638	27115
8	FL	BAKER	2007	13793.788183	27115
9	FL	BAKER	2008	15651.322154	27115

First, we use the formula below to conduct normalization

$$OpioidShipment\_PerCap\_County = \frac{Quantity\ of\ Opiod\ Shipment}{Population}$$

Then, in order to plot the pre-post graph, we add a column called POST, which is 1 if 'YEAR' >= 2010, 0 if 'YEAR' < 2010. This variable can tell us if a typical year is after or before the year when regulation went into effect.

The final dataframe is shown as below.

	BUYER_STATE	BUYER_COUNTY	YEAR	QUANTITY	POP	QUANTITY_PERCAP	POST
0	FL	ALACHUA	2006	54294.0	247336	0.219515	0
1	FL	ALACHUA	2007	56118.0	247336	0.226890	0
2	FL	ALACHUA	2008	64869.0	247336	0.262271	0
3	FL	ALACHUA	2009	73351.0	247336	0.296564	0
4	FL	ALACHUA	2010	74339.0	247336	0.300559	1
5	FL	ALACHUA	2011	75170.0	247336	0.303919	1
6	FL	ALACHUA	2012	64830.0	247336	0.262113	1
7	FL	BAKER	2006	8450.0	27115	0.311636	0
8	FL	BAKER	2007	11144.0	27115	0.410990	0
9	FL	BAKER	2008	11884.0	27115	0.438281	0

- ## Mortality Data

1. Dataset Selection

The national source data of mortality is from US Vital Statistics records. We use Nick's pre-downloaded data, which gives a summary of mortality for drug and non-drug related causes for every US county from 2003-2015. [The data can be found here.]

Note that the US Vital Statistics agency censors some data. If the number of people in a given category (i.e. one county / year / cause of death category) is less than 10, that data does not appear in the data. Similarly, zero-counts are also not reported. By summing deaths over full years, fewer counties end up near below this 10-death threshold, so the data is more complete. Even if there is still data not being reported, we don't give manipulation on them. This is because given that there is plenty enough counties which provide enough statistical power, a few omitted data impact negligible impact on plotting or regressions output.

The dataset is on county-year level and the variables are county(concatenates county name and state code), county code, year, year code, drug/alcohol induced cause, drug/alcohol induced cause code and deaths. Since we only want to investigate drug-related deaths, we filter off other causes and aggregate up all types of drug-related deaths.

However, in mortality data we only have state abbreviations while in population data we only have state names. Therefore, we have to look for a dictionary data that could actually be used as a bridge for mortality data and population data. We found this data through GitHub searching. The variables are state name, postal code (abbreviation) and FIPS code.

2. Data Engineering

The original dataset we use including the underlying cause of death from 2003 to 2015. The tuple of the dataset means the county-level data over the years. The tuple has columns values, such as year, state name, county name, number of deaths, and the cause of the deaths. First, we selected the tuples whose value of the cause of deaths is related to the drug poisonings and overdose.

Then the selected dataset is merged on another dataset that includes the population of each county. This is for the calculation of the death rate and compare these normalized death rates over the county.

- Variable selection
    - For the mortality data, we chose four related variables:
        1. **Year**: From 2003 to 2015. Use 'Year' from the source dataset.
        2. **State**: All states code from the source dataset. For our samples, there are three states as treatment groups: Florida, Texas, and Washington.
        3. **County**: All counties from the source dataset.
        4. **Deaths_County**: Total deaths caused by overdose for a specific county and a specific year. Use ' Deaths' from this dataset. Note that

there are 3 specific reasons for a drug overdose, but we just want to sum them up and regard it as a generic drug overdose.

- For the population dataset, we chose three related variables:
    i. **County**: county name
    ii. **State**: state name
    iii. **Poplulation_2010**: county-level population in 2010

- For the dictionary dataset, we chose three related variables:
    i. **name**: state name
    ii. **postal:** state abbreviation

- Clean data

Originally, the mortality data was divided into annual data files. We merged these files into one file and select tuples whose value of the mortality column is related to drug/alcohol-induced causes. Then, we converted 'Missing' (Categorial type) in **Deaths_County** into NaN values so that they can be operated in later plotting and regression.

- Merge mortality data with dictionary data and population data

We first merged the dictionary data with mortality data using 'State_Code' and then merged population data using 'State' and 'County' as keys.

There were some mismatched information after merging. We investigated the reason behind it and found that it was different naming and typos that led to this issue. There were different naming in state names in the dictionary data and typos in county names in the population data. We unified them and merged again to fix and get more accurate results.

- Normalize drug-related deaths by population

The core variable, 'Deaths_PerCap_County' is simply calculated like the below equation. This measurement will also be called 'mortality' at county level interchangeably in this report.

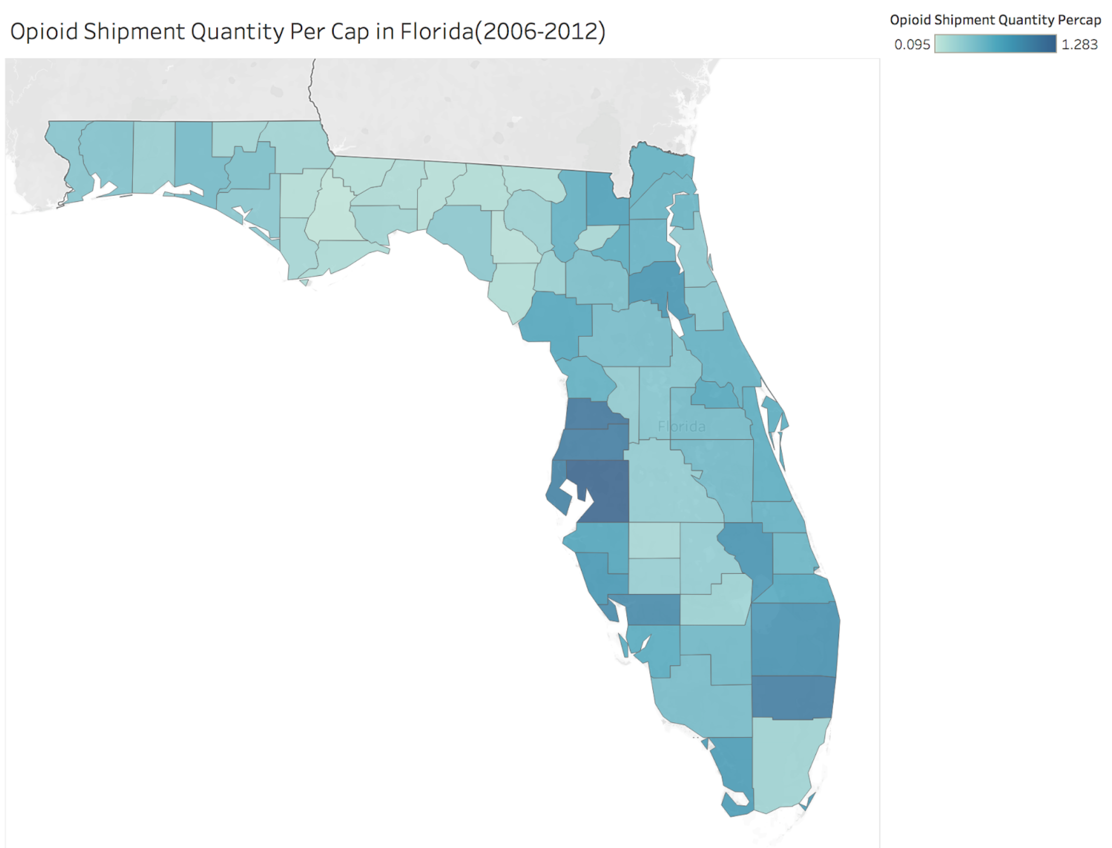
$$Deaths\_PerCap\_County = \frac{Deaths\_County}{Population\_2010}$$

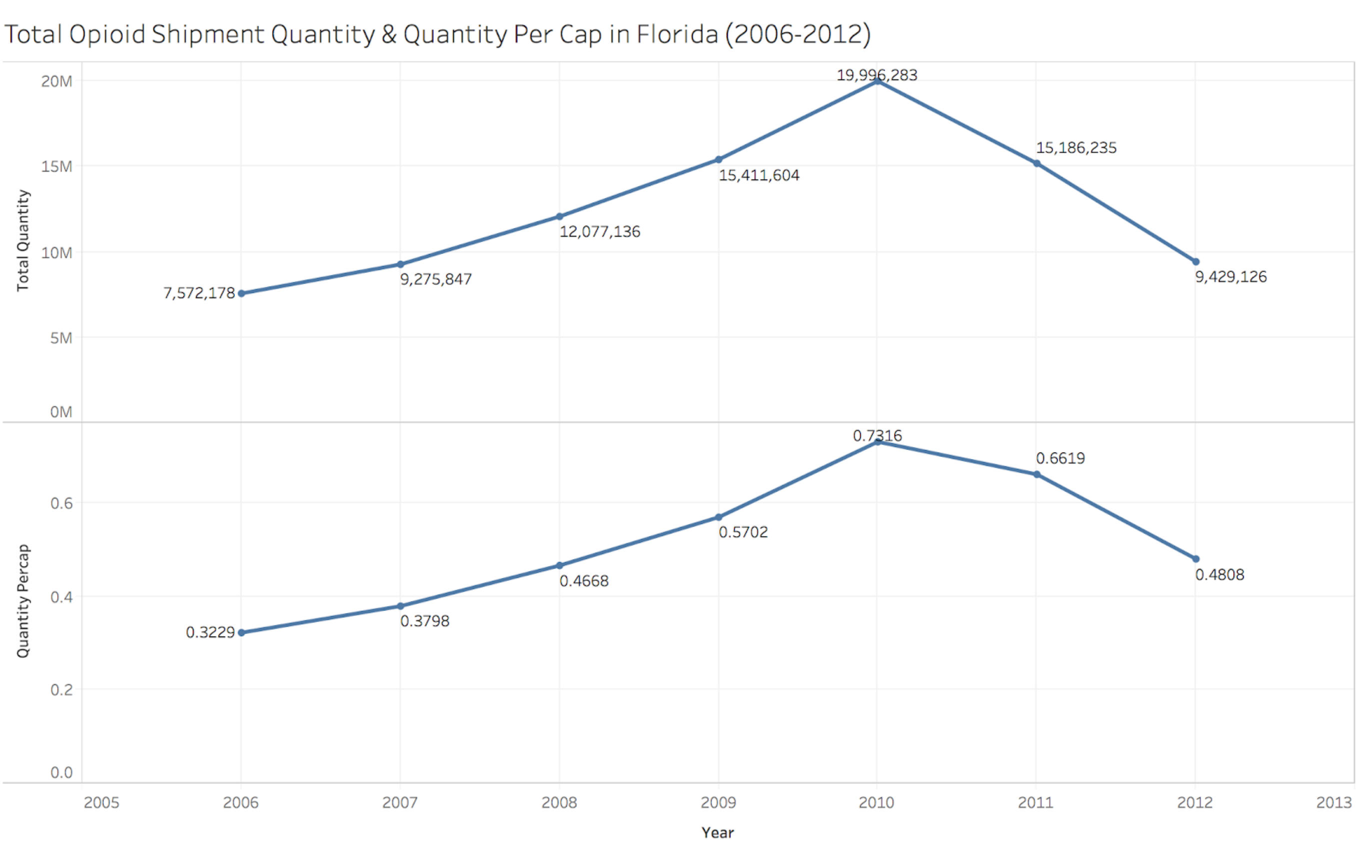
This is also what will be defined as our y-axis in the pre-post graphs and Difference-in-Difference graphs. Specifically, a result of 0.0001 means that there is 1 death caused by drugs out of ten thousand people in the county.

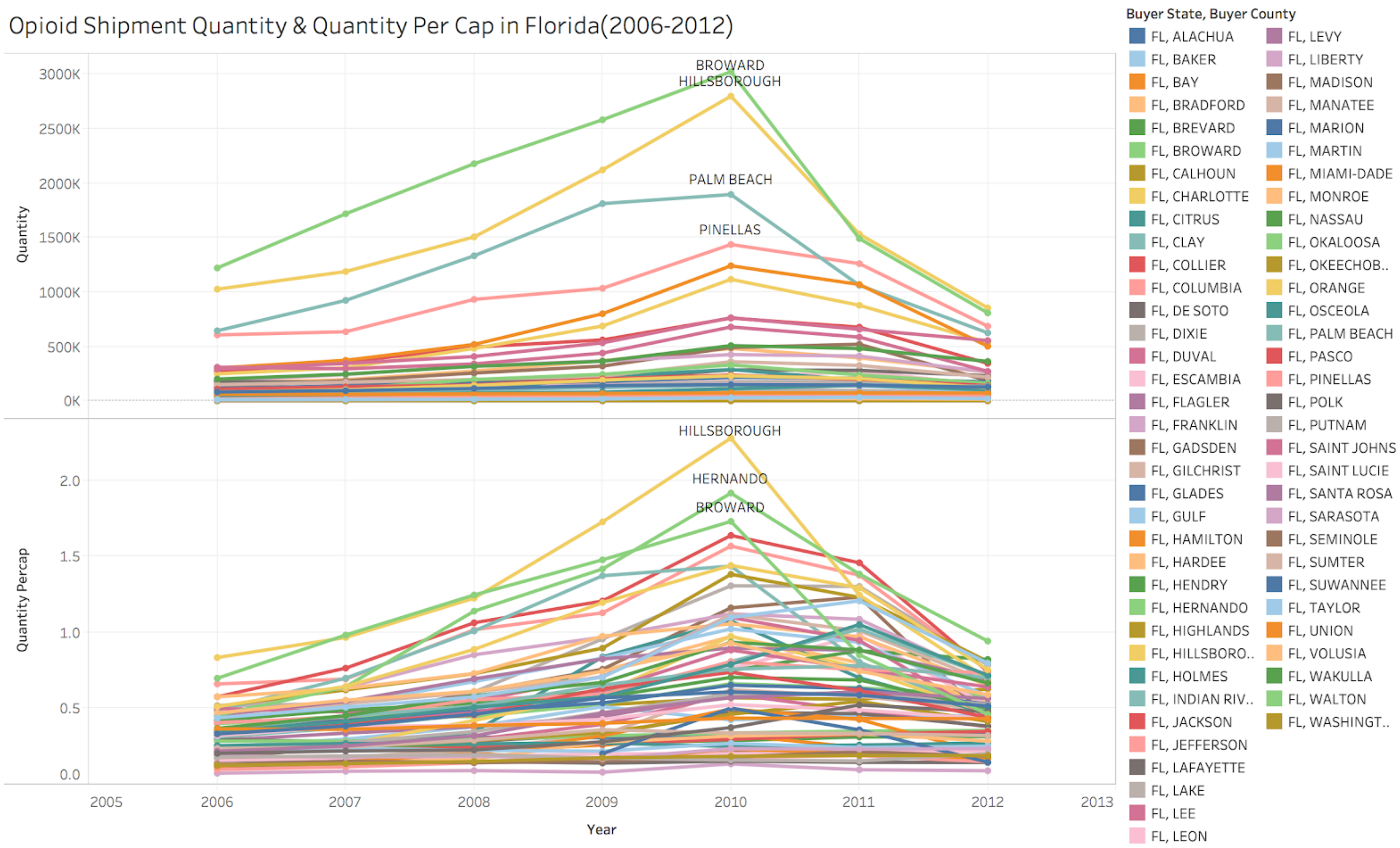
Opioid Shipment Analysis Section

- Statistics Overview

There are 67 counties in Florida with average quantity of opioid shipment 5579052 in total each year. The average quantity of opioid shipment per county from 2006 to 2012 is 83805.50, and the average quantity of opioid shipment per county per cap is 0.2676. Besides, Broward County has the largest quantity of shipment almost every year. We can also find that the total quantity of opioid shipment in Florida decreased after 2010, when the policy went into effect.

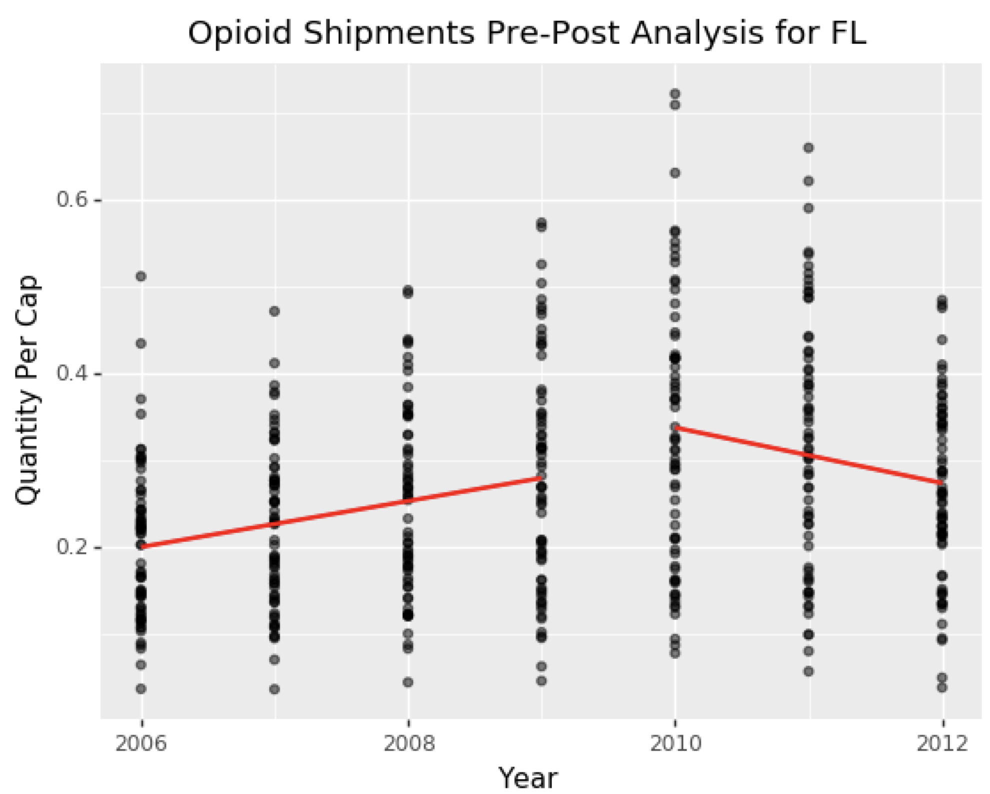






- Pre-Post Analysis

In order to compare how things were in Florida right before the policy went into effect, to Florida right after the policy went into effect, we conducted the pre-post analysis. The points denote the quantity of opioid shipment per cap in each county, while the regression line shows the trend of shipment quantity per cap before and after the policy change in the Florida, as is shown in the figure below.



From this pre-post plot, we can find that:

- Before the policy went into effect on Jan. 2010, the slope of the regression line is positive, which means that the quantity of opioid shipment increased, but after 2010, the slope is negative, which means the quantity of opioid shipment per cap decreased.
- The quantity of opioid shipment of Florida in 2010 is still larger than that in 2009, but consider that the implementation of a policy lags, this outcome is reasonable.

Therefore, we can conclude that potentially the policy change succeeded in reducing the quantity of opioid shipment. However, further analysis needed to be conducted to get more solid conclusion. In the next step, we conducted the difference-in-difference analysis.

- Difference-in-Difference Analysis

Even though the pre-post graphs already provide us with a good and concise perception about some sort of causal inference in this drug study, but it is sometimes has its own shortages, especially when there are a lot of factors driving the change.

For instance, if other states' trends also change from upward-sloping to downward-sloping, pre-post analysis is not sufficient to prove that the change was made by certain policy. Hence, we need to conduct a research to show if the change happens in this state is different from that in other states.

1. Sample Selection

- **Neighbor Level**

After we test the similarity of adjacent states, we use five adjacent states as our control group. Specifically, for Florida we use Louisiana, Mississippi, South Carolina, Alabama and Georgia.

2. Results and Analysis

- **Methods**

By using the Difference-in-Difference strategy, we are asking the question of whether the change we saw in treat state (the difference from pre-to-post) is larger than the change that occurred in other states or national-wide over the same period. In graphs, we plot the pre and post trends of policy change for treat state and its control group.

However, graphs of difference-in-difference aren't always valid and straightforward when it comes to different pre-trends. We will also estimate by a linear regression model that allows for non-parallel pre-trends. This will help us to check the consistency of our results in our two ways of analysis.

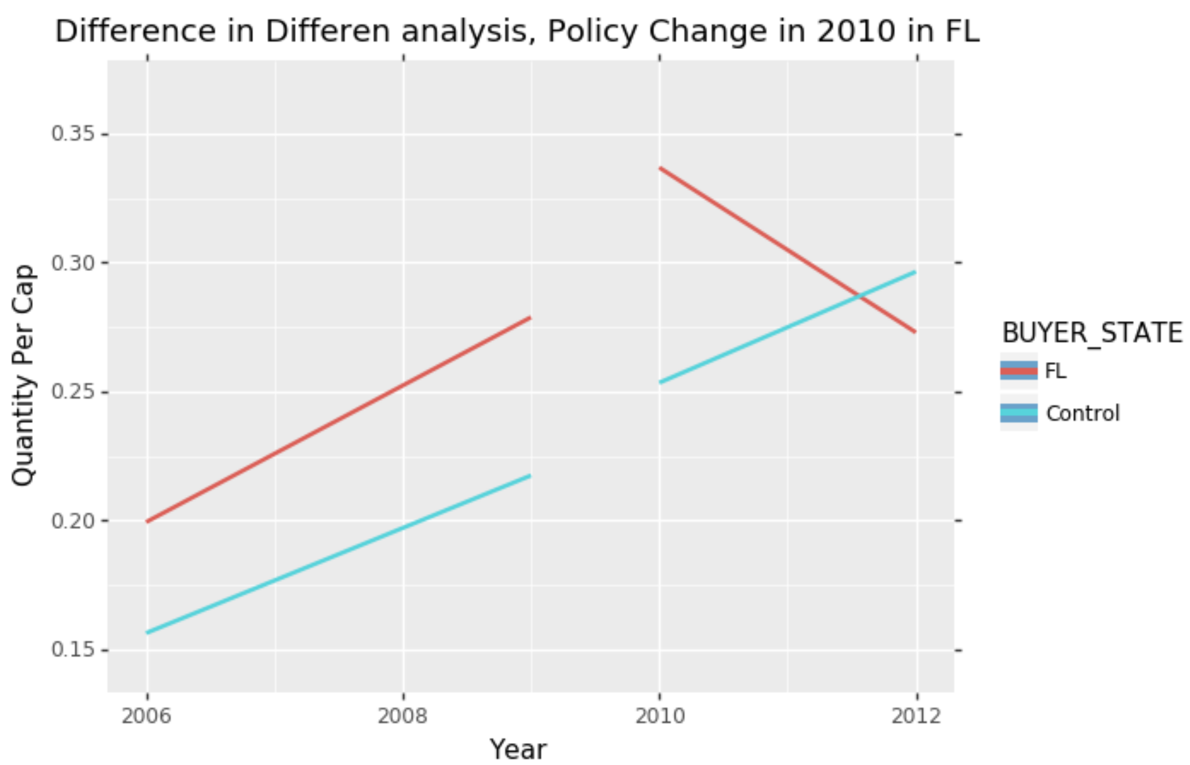
The specification would be:

$$Y_{c,t} = \alpha + \beta_1 Policy\_State_c + \beta_2 Year\_Adjust_t + \beta_3 Policy\_State_c * Year\_Adjust_t + \beta_4 Post_t + \beta_5 Post_t * Policy\_State_c + \beta_6 Post_t * Year\_Adjust_t + \beta_7 Post_t * Year\_Adjust_t * Policy\_State_c + \varepsilon_{c,t}$$

Where $Y_{c,t}$ is a county-year-level outcomes (shipments/drugs deaths) per capita, $Post_t$ is an indicator for whether we are in a period after implementation of the policy change, and $Policy\_State_c$ is an indicator for whether a given county is in a state that experienced a policy change. We adjust $Year_t$ to $Year\_Adjust_t$ so that it has a value of 0 in the year the policy goes into effect. In this way, β_5 is the difference-in-difference estimate for the change in intercepts, and β_7 is the difference-in-difference estimate for the change in slopes.

- **Plots and Regressions**

Firstly, since all other five states have similar levels and trends with Florida, we combine them together as a new control group. This combination gives our regression more statistical power.



From the graph we can see that before the policy change in 2010, Florida had similar upward sloping pre-trends with control group, and it's at a higher level. After the policy changed, while control group barely changed its trend, Florida has experienced a significant change in trend, from upward sloping to downward sloping. Meanwhile, its normalized shipment has even been lower than control group in 2012.

Test for Constraints

	coef	std err	t	P>|t|	[0.025	0.975]
c0	0.3711	0.028	13.449	0.000	0.317	0.425
c1	-0.1044	0.022	-4.729	0.000	-0.148	-0.061

The test above shows the results of testing if β_5 and β_7 equal 0. From the result, we can see that the P value for both estimators are 0, which indicates that we can reject that they are equal to zero. Hence, we could say from statistics perspective that the level and trend of Florida Opioid shipments have changed after the policy took effect.

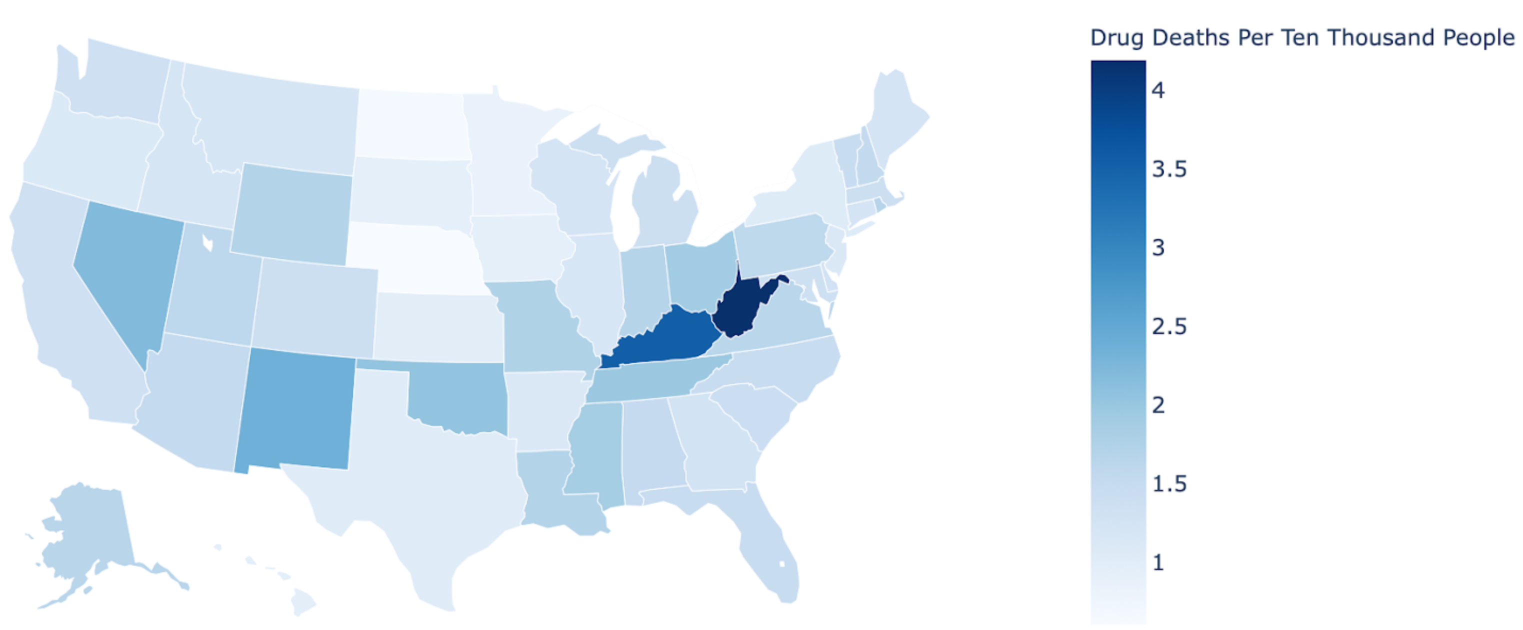
Mortality Analysis Section

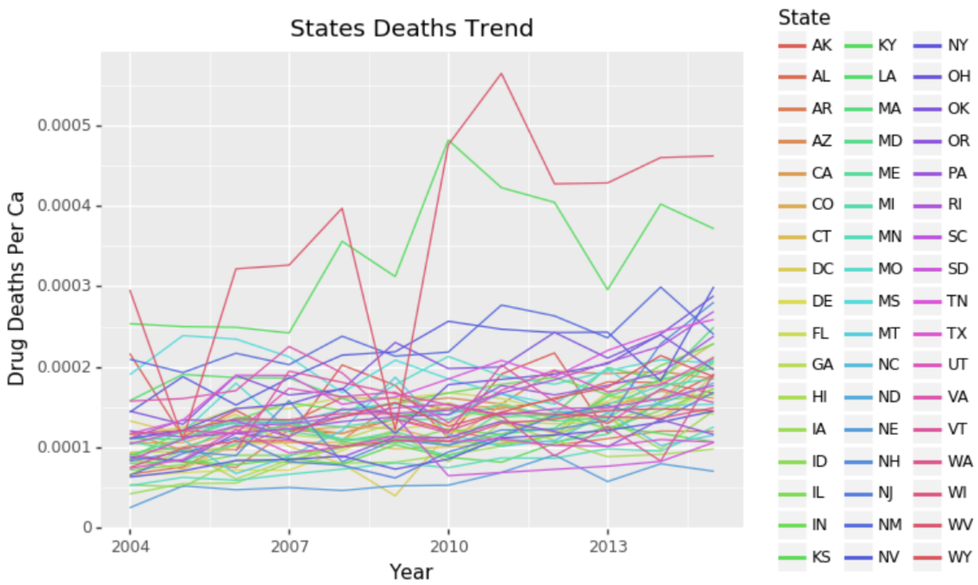
- Statistics Overview

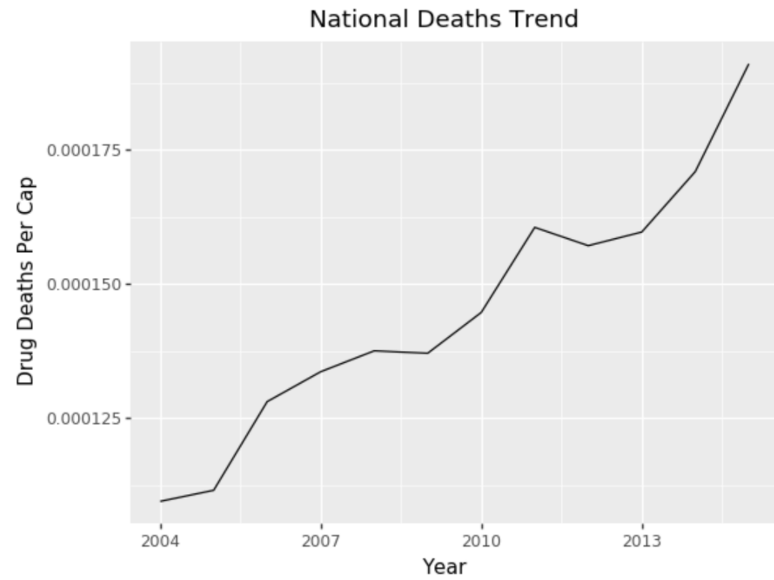
There are 767 counties in the intermediate dataset. The mean of deaths per county over the whole period is 36.32 with a standard deviation of 59.49. At the county level, the counties that have maximum/minimum mean of the number of deaths over 12 years are Los Angeles County and Aransas County(and other 37 counties), respectively. Likewise, at the state level, the maximum mean of deaths is of NV, 155.54 deaths. On the other hand, in ND, the number is 10.0 and minimal.

We draw a choropleth map to show the drug deaths rate (per 10,000 people) on state-level geometrically. This also helps us identify neighbors with similarity. A state-level and national-level trends of drug deaths per capita are also presented to give snapshots.

US Drug Deaths Per Ten Thousand People by State, 2004-2015

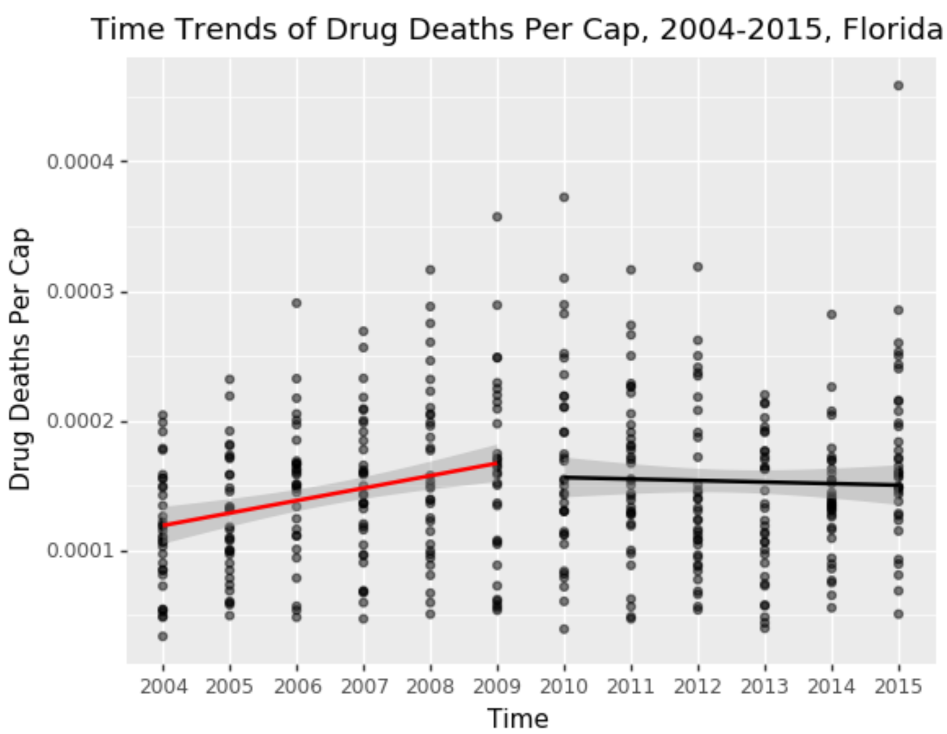




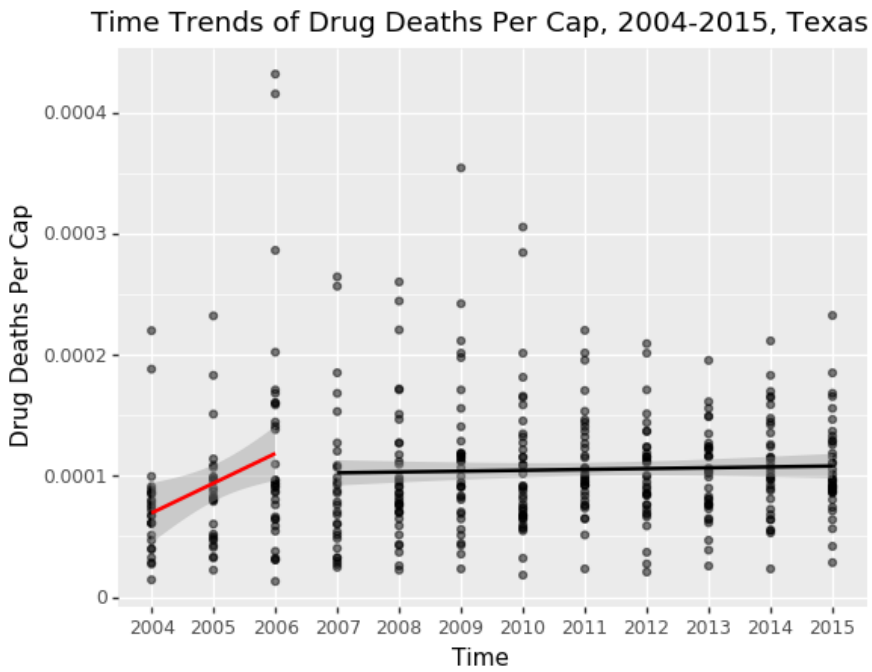


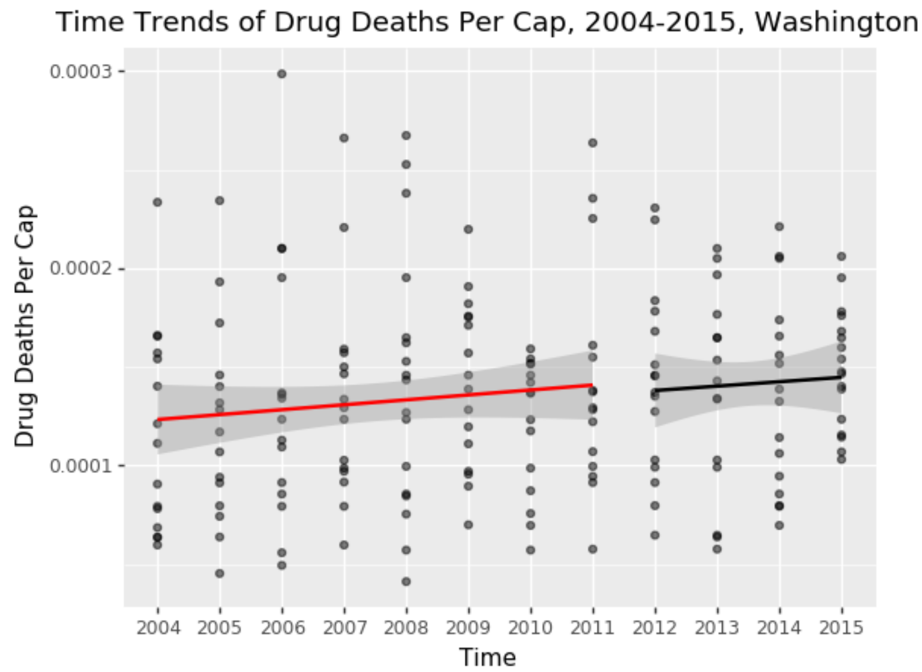
- Pre-Post Analysis

For the pre-post analysis, we create scatter plots of the drug mortality rates of counties in examined states, such as Florida, Texas, and Washington. The points denote the deaths rate in one of the counties over the years. The regression line shows the trends before and after the policy change in the state. For Florida, as the figure for Florida shows, the slope of the regression line after the policy change slightly decreases. This looks like potentially the policy change succeeded in reducing mortality if we only take advantage of this pre-post plot.



Likewise, in Texas, we can observe a similar trend to Florida. On the other hand, the slope of the regression of the death rate in Washington does not change compared to one from the regression before the policy change.





- Difference-in-Difference Analysis

Even though the pre-post graphs already provide us with a good and concise perception about some sort of causal inference in this drug study, but it is sometimes has its own shortages, especially when there are a lot of factors driving the change.

For example, that around 2010 (the same time Florida's policy went into effect) the US Customs service managed to dramatically reduce the importation of fentanyl into the United States. This would likely reduce the amount of overdose deaths throughout the United States, and so when we were looking at our pre-post graph from 2009 to 2011 for Florida, we would see a decline in overdose deaths and wrongly attribute that to Florida's policy change, which may be contributed largely by the importation reduction of fentanyl.

By using the Difference-in-Difference strategy, which serves as a counterfactual analysis helping us to answer the question of whether the change we saw in treat state (the difference from pre-to-post) is larger than the change that occurred in other states or national-wide over the same period.

In the following, we will first discuss our control selection and then present our plots. Then using the same regression specification as in the shipment part to check consistency and make comparison. Finally we will deliver some summary and analysis.

1. Control Selection

- **Neighbors**

For each treat state, we use 5 of its neighbor states as control group.

Treat State	Florida	Texas	Washington
Control Group Neighbors	Louisiana	New Mexico	Oregon
	Mississippi	Oklahoma	Idaho
	South Carolina	Arkansas	Montana
	Alabama	Louisiana	Nevada
	Georgia	Kansas	Wyoming

- **Similar Pre-trend**

For each treat state, we pick their control states by explicitly looking for the top five states with most similar pre-trends. We do regressions of pre-trends for each of all states and measure the similarity using intercept and slope. See more details in the appendix. We use this method as an independent validation test for our neighbor control, or as a supplement control method to it.

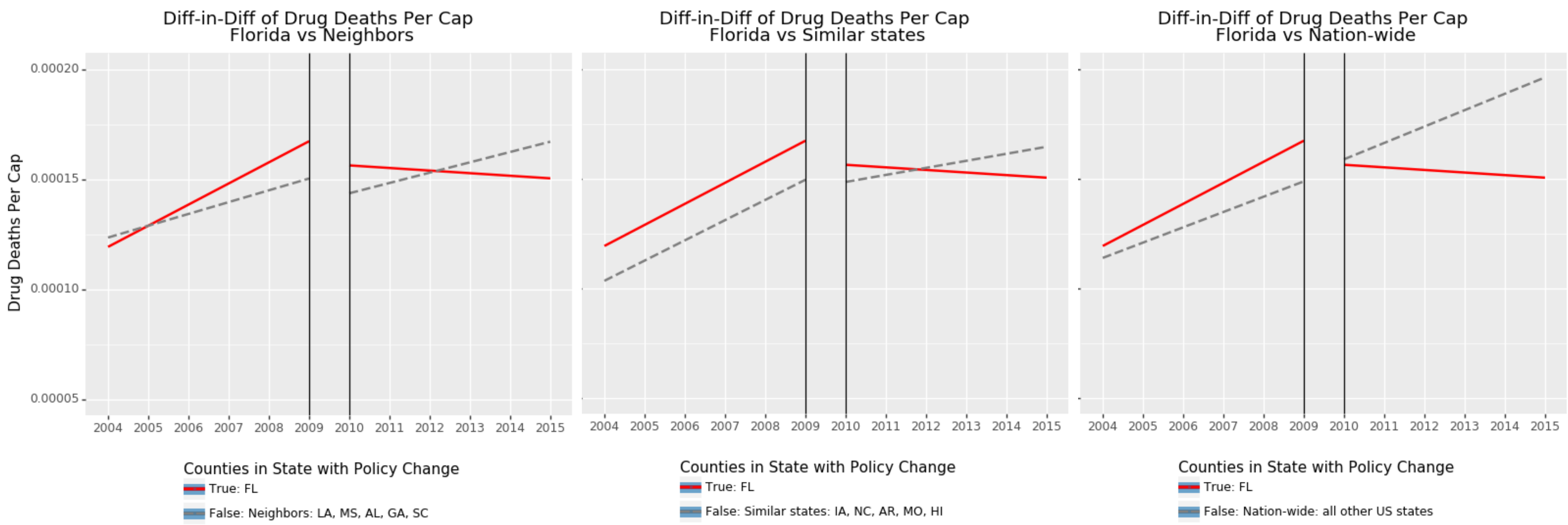
Treat State	Florida	Texas	Washington
Control Group Similar Pre-trend	Iowa	New Jersey	Illinois
	North Carolina	Massachusetts	Michigan
	Arkansas	Indiana	Connecticut
	Montana	Mississippi	Colorado
	Hawaii	Oklahoma	North Carolina

- **National Level**

While neighbor level gives us a more similarity to treat state, looking at national level gives us more statistical power. Given this trade-off, we analyze both to get less biased conclusions. For each treat state, the observations include all counties from all states (excluding Alaska).

2. Results and Analysis

- **Florida**



Firstly, we compare Florida to its neighbors. From the graph we can see that before the policy change in 2010, Florida had steeper upward-sloping pre-trends compared to its neighbors, and at a higher level. After the policy change, the trend of Florida significantly became downward-sloping and went to a lower level than its neighbor after 2012. Its neighbors' mortality were lower a little with an up-ward sloping at unchanged steepness. Generally, the graph shows that there was a difference between Florida and its neighbors/similar states for the change of mortality over the period. Specifically, Florida experienced a larger change with even a reversed post-trend and went below the level of its neighbor around 2012.

When we use states with similar pre-trends as control, the control group has little overlap with neighbor control. It seems that its neighbors are more similar in level but not slope. Nevertheless, we could not say much about the correlation of these two since this is affected by many factors. However, except for a parallel pre-trend, we got the same post-trend trajectory with neighbor control here.

Then, we compare Florida to nation-wide average (all other US states). From the graph we can see that Florida still had a higher level of drug mortality than national average during pre-period. After the policy change, the trend of Florida significantly became downward sloping and went less than national average right after 2010. The national trend of drug mortality seems continued after policy change as same as pre-period. Generally, the graph shows that there was a difference between Florida and national average for the change of mortality over the period. Specifically, Florida experienced a larger change with even a reversed post-trend and went below national level right after policy change.

Also, we examine the Difference-in-Difference combining the three ways of control. We can see that Florida's neighbors gave more respond to policy change than the national average. We may infer that policy change in Florida also affected its neighbors but only a little.

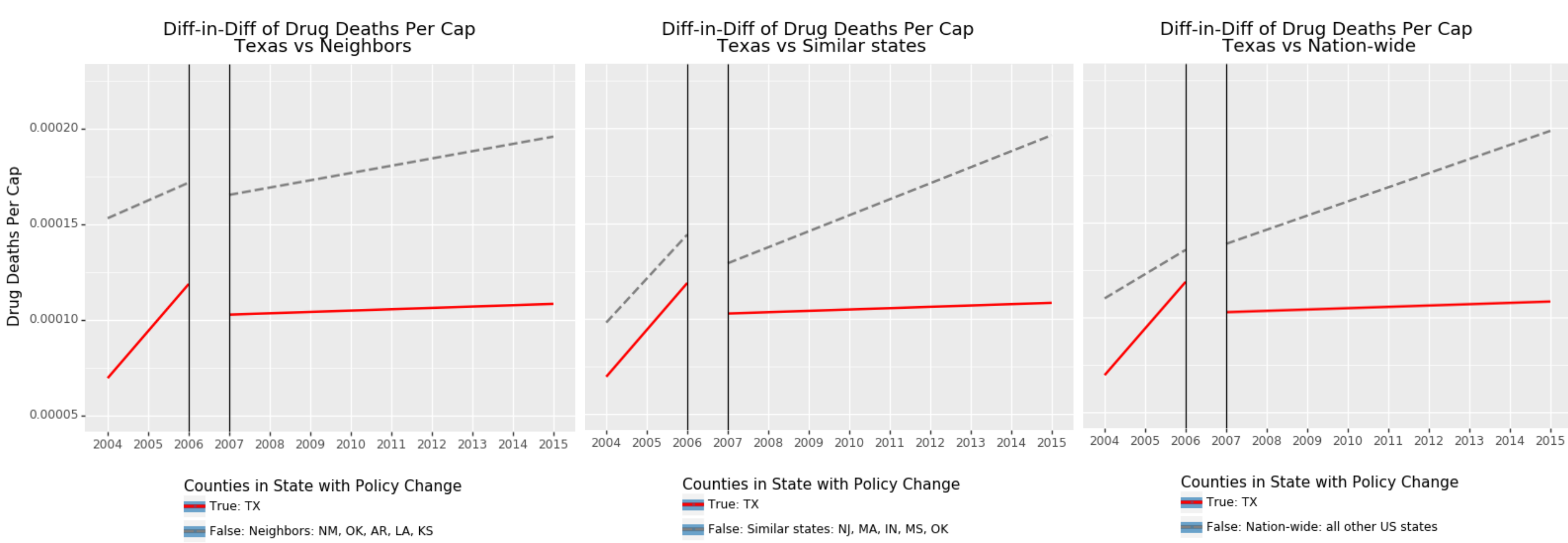
Now let's check plotting results with our regression output when using neighbors and nation-wide control.

	coef	std err	t	P>|t|	coef	std err	t	P>|t|
Intercept	0.0002	8.01e-06	19.439	0.000	0.0002	4.25e-06	36.629	0.000
Policy_State	2.118e-05	1.42e-05	1.497	0.135	2.118e-05	1.76e-05	1.201	0.230
Year_Adjust	5.367e-06	2.16e-06	2.484	0.013	6.983e-06	1.14e-06	6.111	0.000
Year_Adjust:Policy_State	4.236e-06	3.69e-06	1.149	0.251	2.621e-06	4.53e-06	0.579	0.563
Post	-1.209e-05	9.94e-06	-1.217	0.224	3.086e-06	5.21e-06	0.592	0.554
Post:Policy_State	-8.493e-06	1.76e-05	-0.482	0.630	-2.367e-05	2.2e-05	-1.078	0.281
Post:Year_Adjust	-6.745e-07	2.88e-06	-0.234	0.815	4.914e-07	1.5e-06	0.328	0.743
Post:Year_Adjust:Policy_State	-1.011e-05	5.03e-06	-2.010	0.045	-1.127e-05	6.24e-06	-1.808	0.071

Neighbors as Control   All Other States as Control

From the regression output we could also find evidence indicating a larger change in Florida when compared to its neighbors and nation-wide. When using neighbors as control, $\beta_5 = -8.493 * e^{-6} < 0$ and $\beta_7 = -1.011 * e^{-5} < 0$ indicates that both level and slope decreased relative to control group. The corresponding t statistics shows that the D-in-D of slope is significant at 95% confidence level. When using all other states, $\beta_5 = -2.367 * e^{-5} < 0$ and $\beta_7 = -1.127 * e^{-5} < 0$ indicates that both level and slope decreased. The corresponding t statistics shows that the D-in-D of slope is significant at 90% confidence level. These results are in line with our graphs that the difference of slope gap is larger than that of level gap in both control methods.

- **Texas**



Firstly, we compare Texas to its neighbor New Mexico, Oklahoma, Arkansas and Louisiana. From the graph we can see that before the policy change in 2007, Texas had similar upward-sloping pre-trends with its neighbors, though steeper, but at a significant lower level. After the policy change, the trend of Florida became nearly flat and the level decreased a little. Its neighbors' trend seems to continue as before roughly at the same level. Generally, the graph shows that there was a difference between Texas and its neighbors for the change of mortality over the period. Specifically, Texas experienced a larger change (mainly in the trend) than its neighbor.

Because similar states impose restrictions on both the level and slope while Texas's neighbors differs much on them, the overlap is not that much. It seems that the Difference-in-Difference is more distinct when we impose pre-trend restriction.

Then, we compare Texas to nation-wide average. From the graph we can see that Texas still had a lower level of drug mortality than national average during pre-period. Given a flatter trend of Texas and a continued trend of national average after policy change, the gap grew larger. Generally, the graph shows that there was a difference between Texas and national average for the change of mortality over the period. Specifically, Florida experienced a larger change.

Also, we examine the Difference-in-Difference combining the two ways of control. We can see that overall, Texas's neighbors have higher level of mortality than national average but gave more respond to policy change. We may infer that policy change in Florida also slightly affected its neighbors.

Now let's check with our regression output.

	coef	std err	t	P>|t|
Intercept	0.0002	2.12e-05	8.525	0.000
Policy_State	-3.771e-05	3.14e-05	-1.202	0.230
Year_Adjust	9.304e-06	1.02e-05	0.916	0.360
Year_Adjust:Policy_State	1.529e-05	1.5e-05	1.018	0.309
Post	-1.57e-05	2.26e-05	-0.696	0.487
Post:Policy_State	-2.492e-05	3.35e-05	-0.743	0.457
Post:Year_Adjust	-5.498e-06	1.03e-05	-0.535	0.593
Post:Year_Adjust:Policy_State	-1.84e-05	1.52e-05	-1.209	0.227

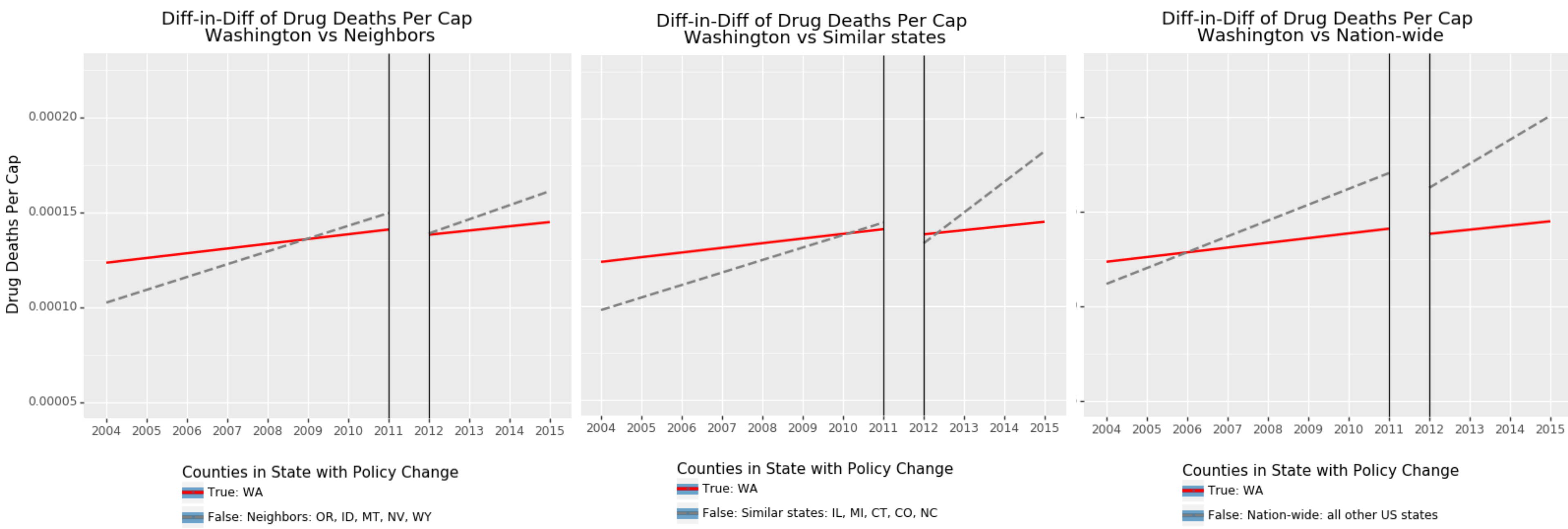
Neighbors as Control

	coef	std err	t	P>|t|
Intercept	0.0001	7.33e-06	20.222	0.000
Policy_State	-5.017e-06	2.96e-05	-0.170	0.865
Year_Adjust	1.278e-05	3.46e-06	3.691	0.000
Year_Adjust:Policy_State	1.182e-05	1.42e-05	0.835	0.404
Post	-9.609e-06	7.8e-06	-1.231	0.218
Post:Policy_State	-3.101e-05	3.17e-05	-0.978	0.328
Post:Year_Adjust	-5.336e-06	3.5e-06	-1.523	0.128
Post:Year_Adjust:Policy_State	-1.856e-05	1.44e-05	-1.293	0.196

All Other States as Control

From the regression output we could also find evidence indicating a mild larger change in Texas when compared to its neighbors and nation-wide. When using neighbors as control, β_5 = -2.492 * e^{-5} < 0 and β_7 = -1.84 * e^{-5} < 0 indicates that both level and slope decreased relative to control group. However, the corresponding t statistics shows that the D-in-D of level and slope are neither significant at conventional confidence level. When using all other states, β_5 = -3.101 * e^{-5} < 0 and β_7 = -1.856 * e^{-5} < 0 indicates that both level and slope decreased relative to control group. However, the corresponding t statistics shows that the D-in-D of level and slope are neither significant at conventional confidence level. These results are in line with our graphs that, though not significant, the slope and level gaps are larger in both control methods.

- **Washington**



Firstly, we compare Washington to its neighbor Oregon and Idaho. From the graph we can see that before the policy change in 2012, Washington had similar upward-sloping pre-trends and level with its neighbors, though slightly flatter. After the policy change, there is nearly no change in the trend and level of mortality in Washington. Its neighbors' trend seems to continue the same as before but followed by a little lower level right after policy change. Generally, there is little change in Washington and its neighbors in terms of mortality over the period.

It seems that we can hardly find a parallel pre-trend as the second graph shows. Given Washington experience little effect, it seems to act like a control, making its neighbors and other states behave like treat group.

Then, we compare Washington to nation-wide average. From the graph we can see that national average of mortality level grew over that of Washington after 2006. Since national average of mortality increased to grow after 2006, the gap then became larger as time moving forwards and even bigger after the policy change. Generally, the graph shows that Washington were not responsive to policy change while national average mortality became even more negative.

Also, we examine the Difference-in-Difference combining the two ways of control. We can see that overall,national average have higher level of mortality than Washington’s neighbors but they neither gave any responds to policy change.

Now let's check with our regression output.

	coef	std err	t	P>|t|
Intercept	0.0002	1.12e-05	13.967	0.000
Policy_State	-1.294e-05	1.63e-05	-0.794	0.428
Year_Adjust	6.739e-06	2.45e-06	2.748	0.006
Year_Adjust:Policy_State	-4.242e-06	3.4e-06	-1.247	0.213
Post	-1.751e-05	1.58e-05	-1.105	0.270
Post:Policy_State	1.227e-05	2.34e-05	0.524	0.600
Post:Year_Adjust	6.979e-07	6.2e-06	0.112	0.911
Post:Year_Adjust:Policy_State	-9.708e-07	9.33e-06	-0.104	0.917

Neighbors as Control

coef	std err	t	P>|t|
0.0002	3.33e-06	53.679	0.000
-3.526e-05	2.13e-05	-1.652	0.099
8.365e-06	6.99e-07	11.974	0.000
-5.868e-06	4.26e-06	-1.378	0.168
-1.612e-05	4.75e-06	-3.393	0.001
1.088e-05	3.1e-05	0.351	0.726
4.32e-06	1.9e-06	2.278	0.023
-4.593e-06	1.25e-05	-0.366	0.714

All Other States as Control

From the regression output we could also find evidence indicating a negligible change in Washington when compared to its neighbors and nation-wide. When using neighbors as control, β_5 =1.227 * e^{-5} > 0 and β_7 = -9.708 * e^{-7} < 0 indicates that the level increased and slope decreased relative to control group. However, the corresponding t statistics shows that the D-in-D of level and slope are neither significant at conventional confidence level. When using all other states, β_5 = 1.088* e^{-5} < 0 and β_7 = -6.593 * e^{-5} < 0 indicates that the level increased and slope decreased relative to control group. However, the corresponding t statistics shows that the D-in-D of level and slope are neither significant at conventional confidence level. These results are in line with our graphs that, the change of slope and level gaps are not significant and there was no obvious respond in Washington.

Conclusion

The core question of our research is that what is the effect of opioid drug prescription regulations on (1) the volume of opioids prescribed, and (2) drug overdose deaths. By using Difference-in-Difference strategy, we could examine whether a policy change on regulation exerted an expected impact on the outcomes of interest.

Our results indicates that for Florida, the treated sample state with a high level of overdose deaths overall, showed obvious responds both to volume of prescription and overdose deaths. They decreased in level and presented a downward-sloping trend after policy change in 2010. The trajectory of prescription and overdose deaths over the period are quite similar, so we can say that they are of high correlation. However, the causal inference on whether the decrease in prescription is the reason of decreased overdose deaths cannot be determined and interpreted merely by this correlation.

Also, for Texas, the treated sample state with a low level of overdose deaths overall, D-in-D analysis also provides evidence of the policy change's effect on overdose deaths. There is a decrease in the level and a flattened trend after policy change in 2007.

However, Washington seems to experience negligible change to the policy. Even its neighbor seems to be more responsive to the policy change. Also, it looks like that it's hard to find a states that have parallel pre-trend with Washington.

Since D-in-D strategy requires a conventional assumption of parallel pre-trends, we should be careful when interpreting the results. This brings us to the problem of better way to select control groups. Even if we could find the trajectory of mortality or shipments most like treat state, it actually were not for the policy change, since there are many other factors like demographics and governance driving the trajectory. Parallel trends only capture one dimension of similarity, and a not very theoretically-motivated one. Nevertheless, it can still be used as an independent validation test. In addition, there is a trade-off of sample selection. More states means more statistical power, but with more states we may end up with states that are less like the treat state. Therefore, we do both, with all other states to generate more statistical power, and neighbors that keep more similarity.

Apart from sample selection, we also notice that each treat sample has different pre and post window length. This will make it hard to do comparison across different treat states. In other words, shorter window will have less statistical power and make the regression fit not representative enough. For example, in analysis for Texas, while there is 9 years in post-period, there is only 3 years in pre-period which may contribute to an overestimate of steepness of the fitted line and will affect our comparison to the post-period and related interpretation. Therefore, for better analysis, we should take consideration of a more balanced pre-post windows if possible.

Nevertheless, Difference-in-Difference still provides us with a good estimate of the impact of opioid control policies. The next step may be digging further into the question of why the policy seems to be more effective in some states than others.

Appendix

a. Similar Pre-trend States Selection of FL for Opioid Shipment

	level	slope
AL	-8.776954	0.004396
GA	-13.730238	0.006866
MS	-18.825820	0.009414
SC	18.658328	-0.009285

We run regression for each state before the policy took effect, and calculate level & slope difference to select states most similar with FL. The result shows that AL is the most similar state, with GA slightly less alike. Hence, we chose these two states as our control group.

b. D-in-D regression of FL for Opioid Shipment

FL: Using AL and GA as Control

Test for Constraints

	coef	std err	t	P>|t|	[0.025	0.975]
c0	0.3711	0.028	13.449	0.000	0.317	0.425
c1	-0.1044	0.022	-4.729	0.000	-0.148	-0.061

In this test, c0 represents β_5 which is level change, and c1 represents β_7 which is time trend change.From the regression output,we find P value for both of them are zero, which shows we can reject there is no change in level and trend. Generally, it indicates that while the shipment per capita increased for the first year since the policy took effect, it has continued to decrease since then.

c. Similar Pre-trend States Selection for Mortality

State	Slope	Level
FL	0.000000e+00	0.000000
IA	9.924763e-07	0.001914
NC	1.063865e-06	0.002124
AR	1.083127e-06	0.002213
MO	1.164851e-06	0.002340
HI	1.172019e-06	0.002296

State	Slope	Level
TX	0.000000e+00	0.000000
NJ	5.789646e-07	0.001157
MA	7.472885e-07	0.001517
IN	1.413288e-06	0.002813
MS	2.013388e-06	0.004164
OK	2.021262e-06	0.004113

State	Slope	Level
WA	0.000000e+00	0.000000
IL	3.508046e-08	0.000114
MI	3.813047e-07	0.000731
CT	3.986060e-07	0.000836
CO	6.393788e-07	0.001294
NC	7.591676e-07	0.001526

(1) FL

For the difference-in-difference analysis, we pick up five neighbor states as the controlled states. Also, we use the national average death ratio by years to compare to treated states. On the other hand, as we did in the opioid shipment analysis, we run the regression for each state before the policy took effect, and calculate the level and slope difference to select states most similar to FL. The figure above shows the difference between the slope of the regression for each state and one for Florida. As the figure shows, we could say that IA, NC, AR, MO, and HI had similar trends to Florida.

(2) TX

The figure right shows the difference between the slope of the regression for each state and one for Florida. As the figure shows, we could say that NJ, MA, IN, MS, and OK had similar trends to Florida.

(3) WA

The figure right shows the difference between the slope of the regression for each state and one for Florida. As the figure shows, we could say that IL, MI, CT, CO, and NC had similar trends to Florida.

Estimate the Impact of Opioid Control Policies

Policy Maker Version

Team 2

Jingyi Wu, Zifan Peng, Shota Takeshima, Yu Gu

Nov 13, 2019

Motivation for the Project

The increase in the use and abuse of opioids has caused tremendous prescription overdose deaths. As a result, the U.S. government has established new policies to control opioid over-prescription, thus reducing the death caused by drug overdose.

However, how much effect these policies created remains a question. In this project, we want to test whether these policies are really effective, that is, did these policies reduce opioid prescription and mortality from drug overdose significantly? Hence, we decided to make some statistical model to solve this problem.

Overview of the Data Being Used

Population Data

Because of the potential that areas with large population tends to have more drug shipments and higher drug-related mortality, we have to normalize the outcomes of interests by population at county level.

We use the Population Change for Counties in the United States and for Municipios in Puerto Rico: 2000 to 2010 (CPH-T-1) from the United States Census Bureau. The data can be found here.

Notice that the US only has a census every ten years and given that the population annual change is negligible when normalizing the deaths, we only use population in 2010 to normalize for the whole window (2004-2015). If one wants to get more precise results, the United States Census Bureau also provides predicted population annually on county level.

Opioid Shipments Data

The main dataset we used is maintained by the **Drug Enforcement Administration** that tracks the path of every single pain pill sold in the United States in every town and city.

In order to analyze the effect of policy change in **Florida**, we used the **shipment data for florida on county level**. To make deeper analysis, we picked neighbor states, **Alabama** and **Georgia**, to compare with. Besides, in order to eliminate the differences caused by population, we normalized the quantity of opioid shipments by **population**. Therefore, the basic four datasets are:

- Statewide opioid-shipment data for Florida from 2006 to 2012
- Statewide opioid-shipment data for Alabama from 2006 to 2012
- Statewide opioid-shipment data for Georgia from 2006 to 2012
- Population dataset for all counties in 2010

[Download opioid-shipment data here](#)
[Download population data here](#)

Mortality Data

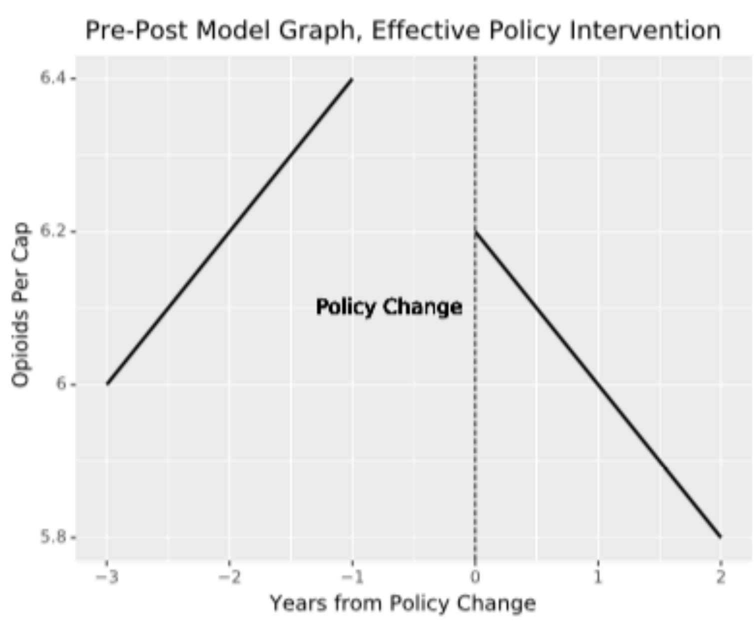
The national source data of mortality is from **US Vital Statistics records**. We use Nick's pre-downloaded data, which gives a summary of mortality for drug and non-drug related causes for every US county from 2003-2015. Still, we will normalize them by population. The core measurement of mortality we are investigating in this report is deaths per capita in a county. This measurement will also be sometimes called 'mortality' at county level and will be used interchangeably in this report.

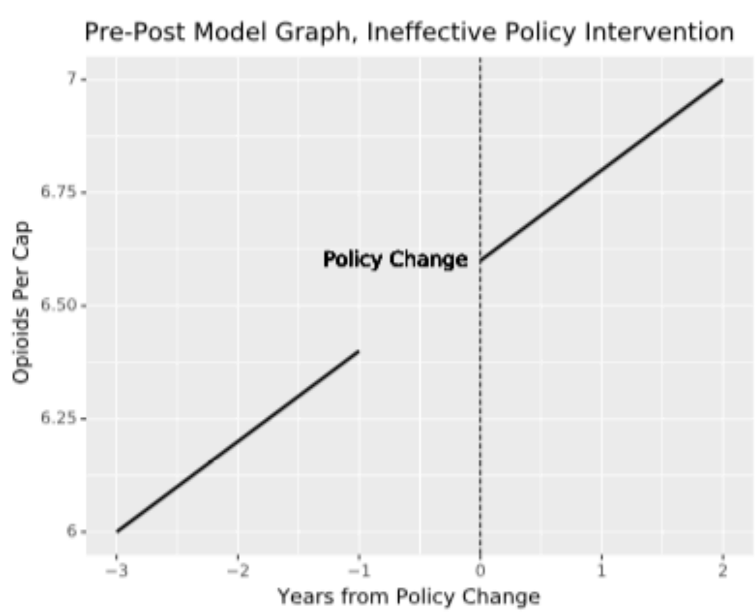
However, in mortality data we only have state abbreviations while in population data we only have state names. Therefore, we use a dictionary data that could actually be used as a bridge for mortality data and population data. [We found this data through GitHub searching.](#) The data contains state name, postal code (abbreviation) and FIPS code.

Overview of Analysis Methods Being Used

Pre-post analysis

It aims to compare how things were in the state right before and after the policy went into effect. If the policy takes effect, the trend before will change while it won't change too much if the policy does not go into effect, which matches the situation in the first figure and the second figure.



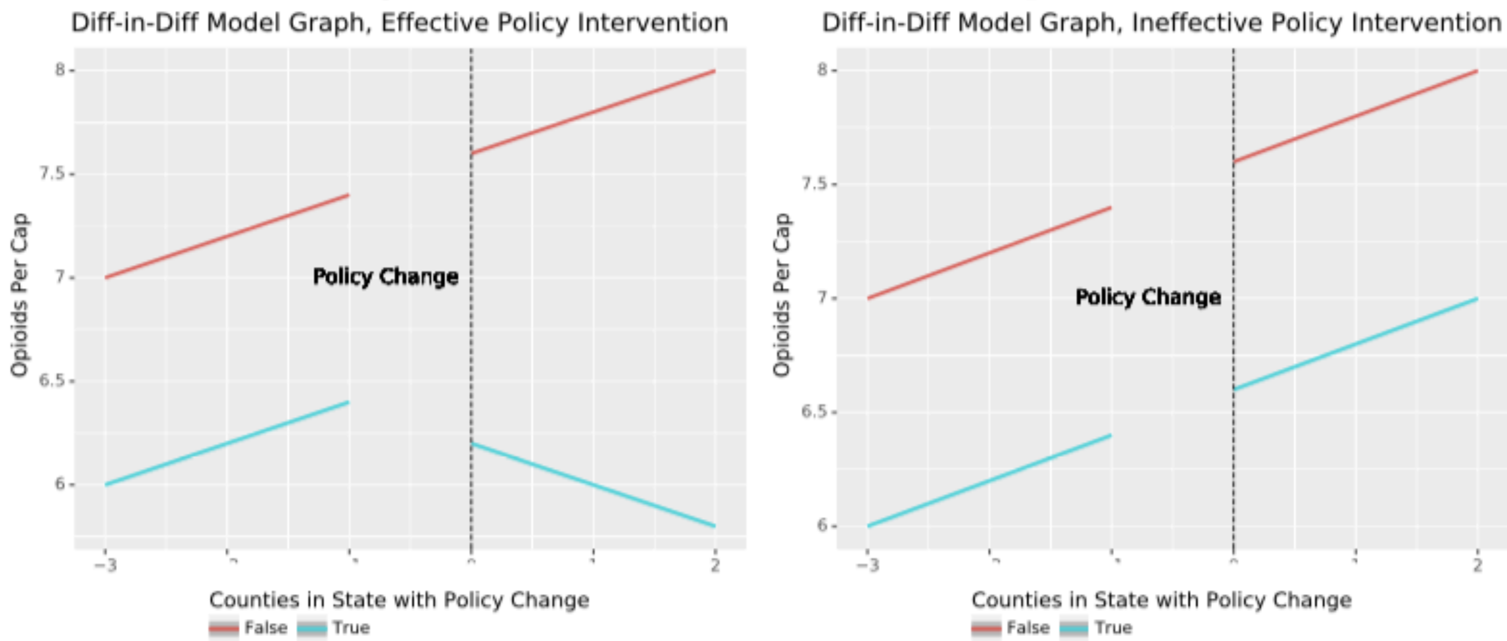


Difference-in-difference Analysis

Even though the pre-post graphs already provide us with a good and concise perception about some sort of causal inference in this drug study, but it is sometimes has its own shortages, especially when there are a lot of factors driving the change.

For instance, if other states' trends also change from upward-sloping to downward-sloping, pre-post analysis is not sufficient to prove that the change was made by certain policy. Hence, we need to conduct a research to show if the change happens in this state is different from that in other states. Please refer to the figure following.

Figure 3: Difference-in-Difference Example Plots

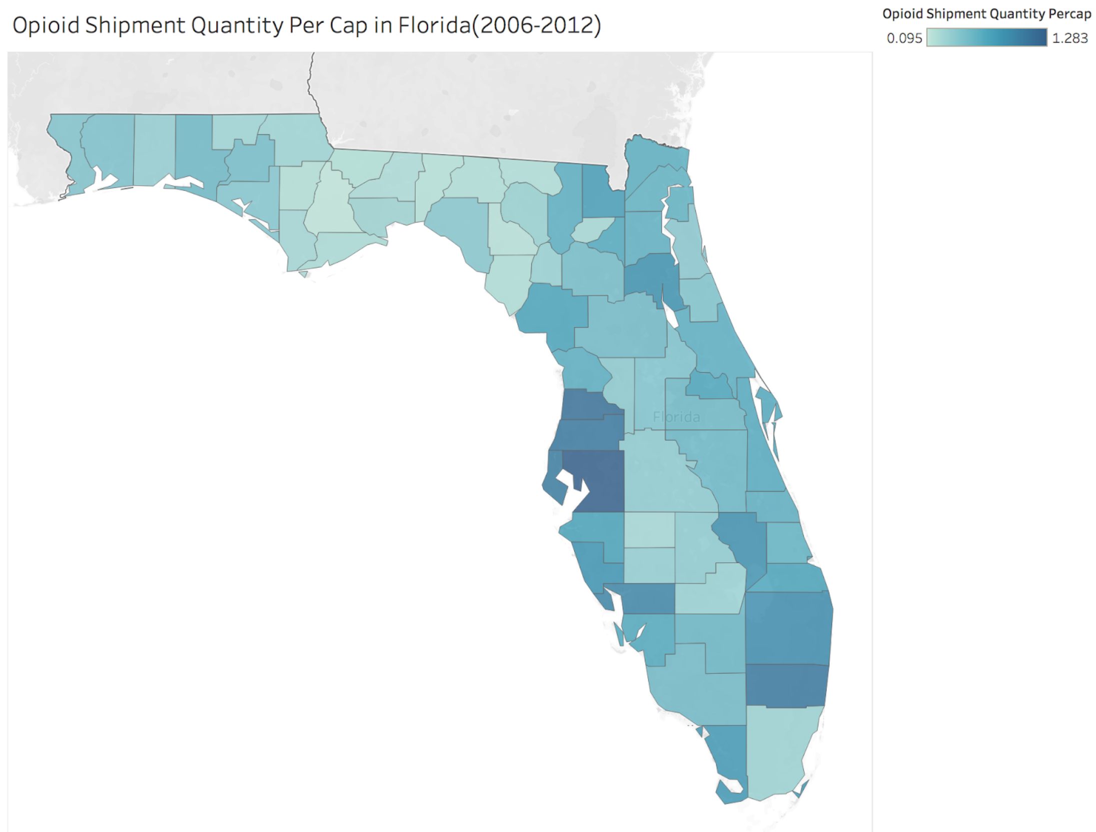


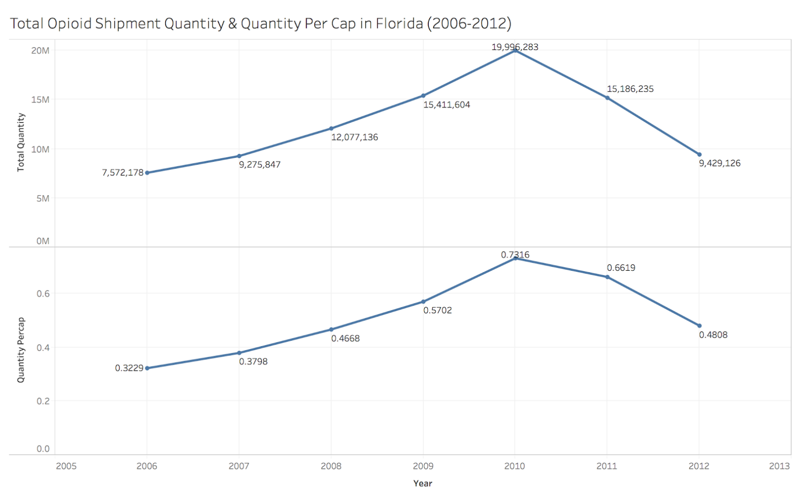
Analysis

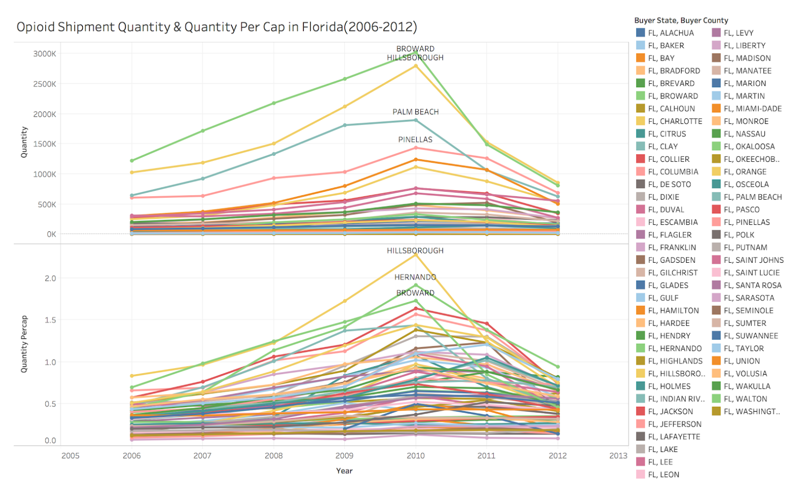
Opioid Shipment Analysis Section

- **Statistics Overview**

There are 67 counties in Florida with average quantity of opioid shipment 5,579,052 in total each year. The average quantity of opioid shipment per county from 2006 to 2012 is 83,805, and the average quantity of opioid shipment per county per cap is 0.27. Besides, Broward County has the largest quantity of shipment almost every year. We can also find that the total quantity of opioid shipment in Florida decreased after 2010, when the policy went into effect.





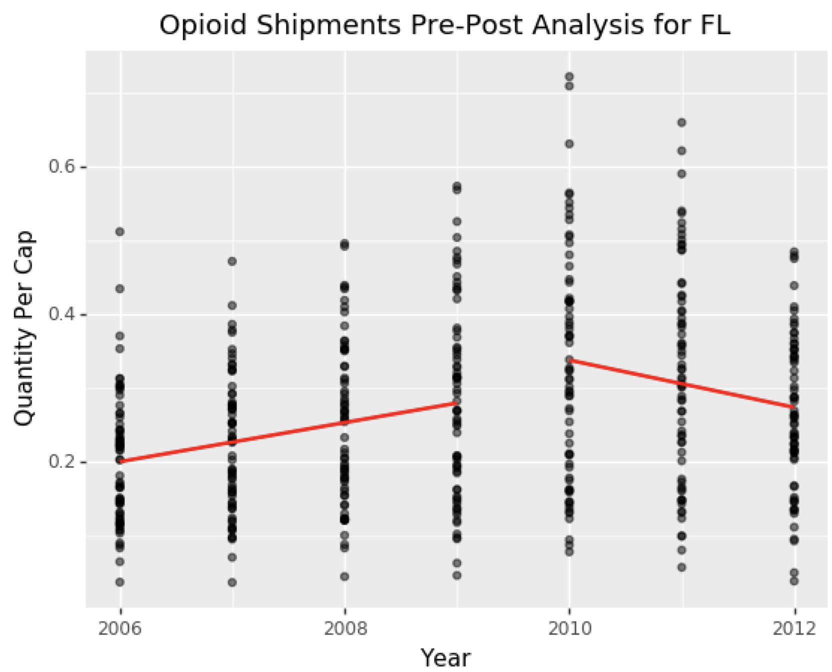


- Pre-Post Analysis

In order to compare how things were in Florida right before the policy went into effect, to Florida right after the policy went into effect, we conducted the pre-post analysis. The analysis aims to compare whether there's change before and after the policy changes. From this pre-post plot, we can find that:

- Before the policy went into effect on Jan. 2010, the slope of the regression line is positive, which means that the quantity of opioid shipment increased, but after 2010, the slope is negative, which means the quantity of opioid shipment per cap decreased.
- The quantity of opioid shipment of Florida in 2010 is still larger than that in 2009, but consider that the implementation of a policy lags, this outcome is reasonable.

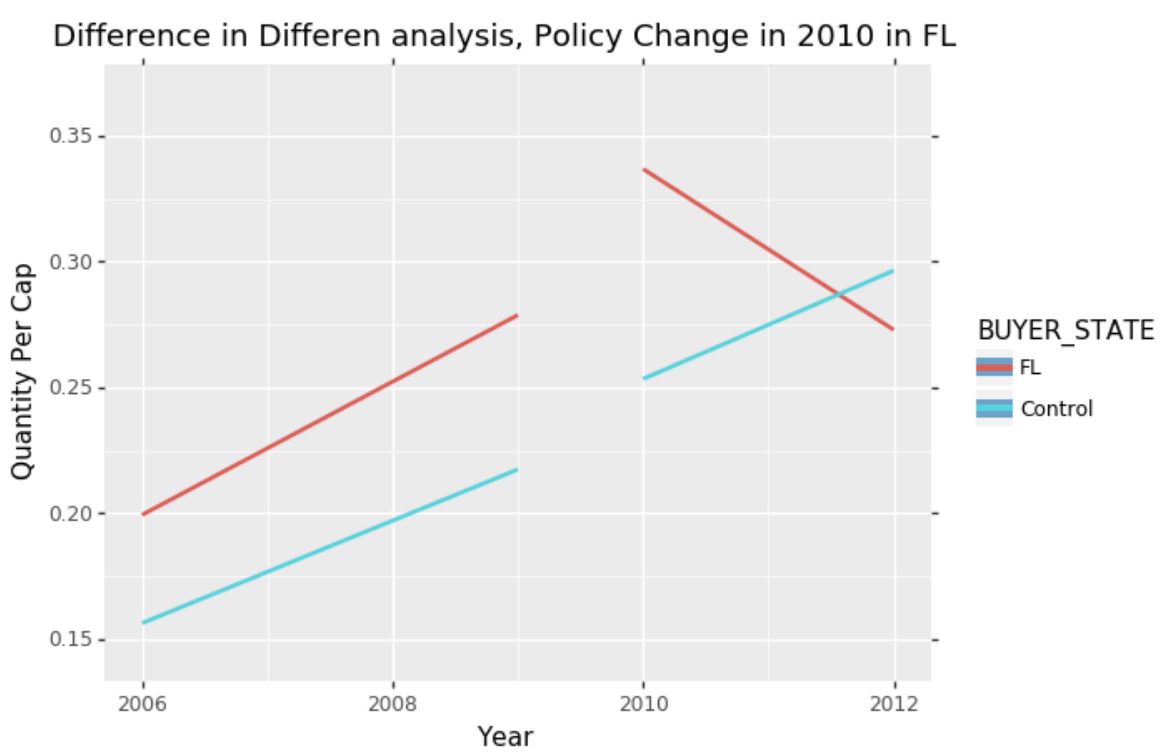
Therefore, we can conclude that potentially the policy change succeeded in reducing the quantity of opioid shipment. However, as we stated before, this change can also happen in other states. Hence, further analysis needed to be conducted to get more solid conclusion. In the next step, we conducted the difference-in-difference analysis.



- **Difference-in-Difference Analysis**
    - **Sample Selection**
        - **Neighbor Level**

    After we test the similarity of adjacent states, we use five adjacent states as our control group(Generally, more data, more solid the results). Specifically, for Florida, we use all counties from Alabama, Mississippi, South Carolina, Georgia, Louisiana.

    - **Results and Analysis**



From the graph we can see:

- Before the policy change in 2010, Florida had similar upward sloping pre-trends with five other states (two lines are almost parallel), and it's at a higher level(Florida is above other states).
- After the policy changed, while control group barely changed its trend, Florida has experienced a significant change in trend(red line goes downwards), from upward sloping to downward sloping.

Besides, we also use other statistical methods to test if Florida really experienced a decreasing trend in Opioid shipments, and the answer is: Yes!

Mortality Analysis Section

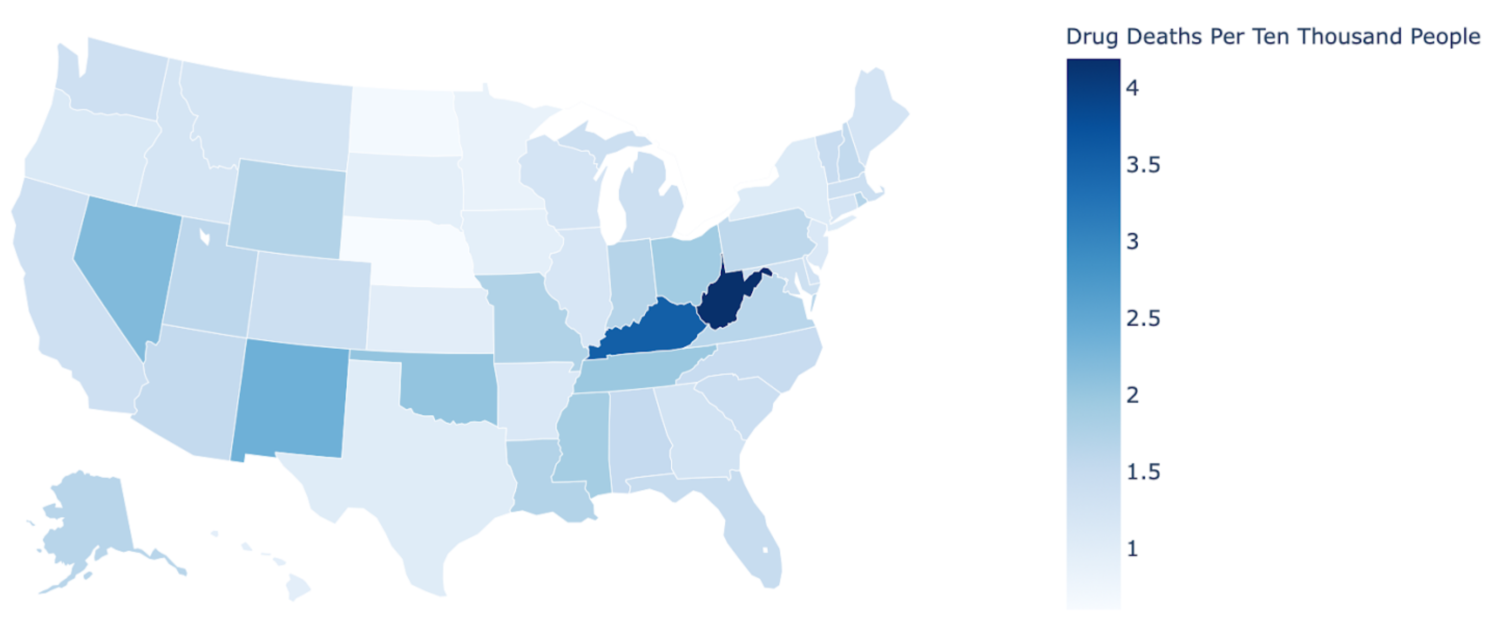
- **Statistics Overview**

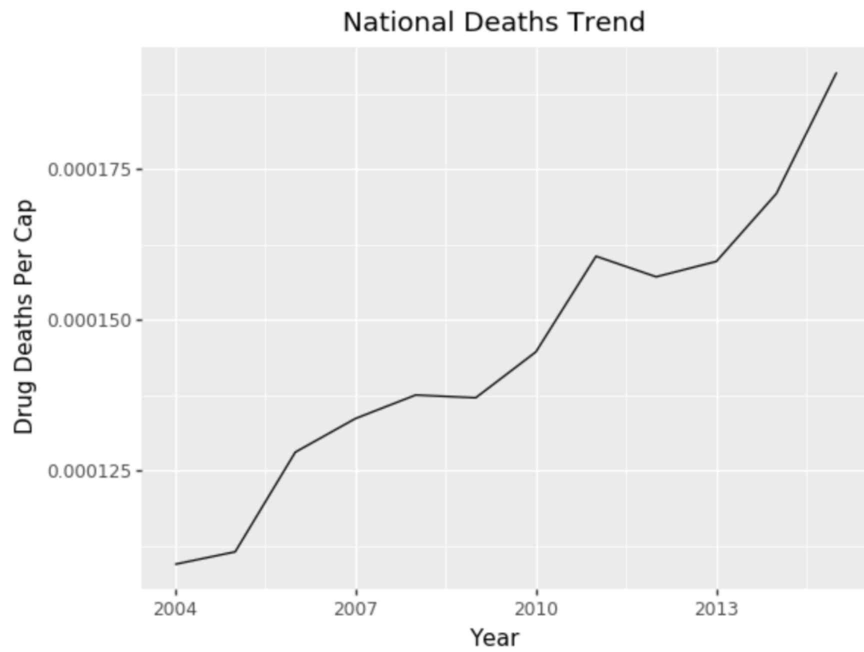
There are 767 counties in total being examined. The mean of deaths per county over the whole period is 36.32. Los Angeles County has a maximum at mean level while Aransas County and other 37 counties get a minimum. Likewise, at the state level, NV has a maximum mean of deaths at 155.54. On the other hand, in ND have a minimal at 10.0. We show the state-level and national trends related to drug poisonings at the plots below.

Specifically, we provide a choropleth map to show the drug deaths per 10,000 people on state-level geometrically. From it we can see that there is a cluster effect, which also helps us identify neighbors with similarity.

A state-level and national-level trends of drug deaths per capita are also presented to give snapshots. We can see that deaths rate trend are still growing on national-level.

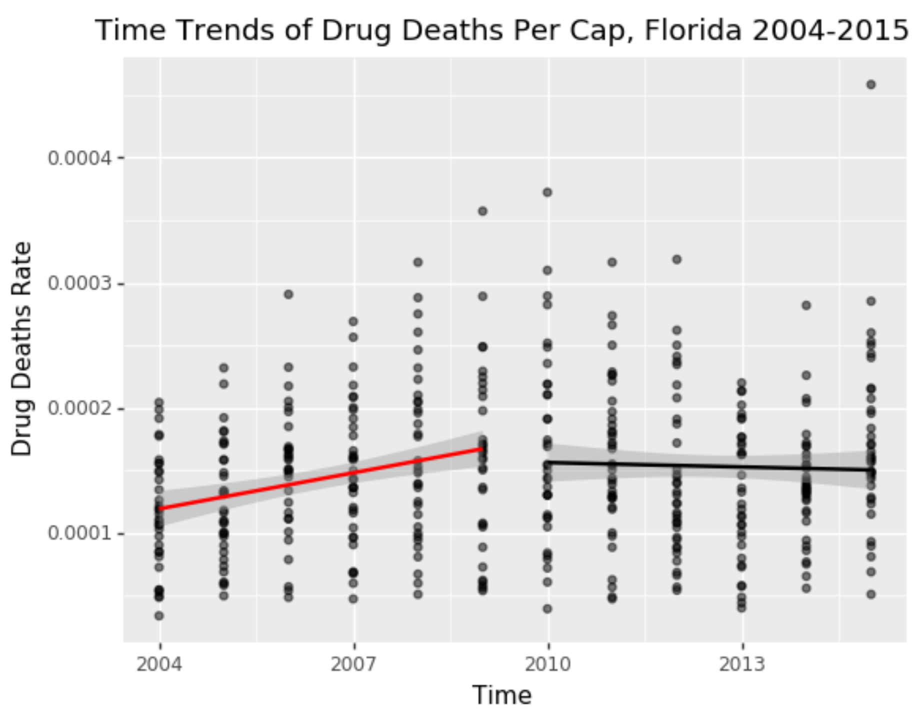
US Drug Deaths Per Ten Thousand People by State, 2004-2015



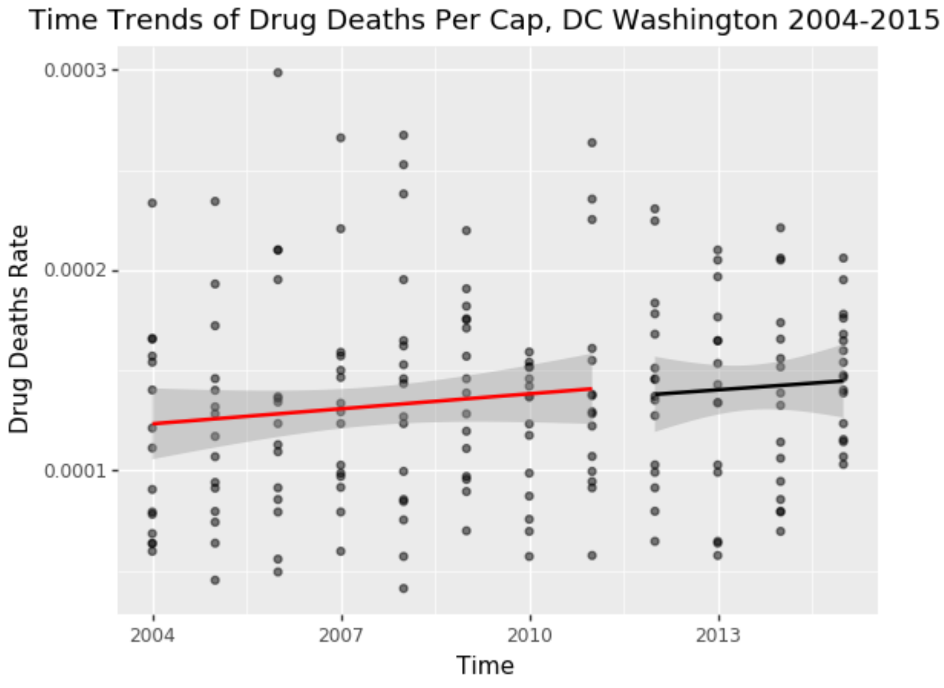


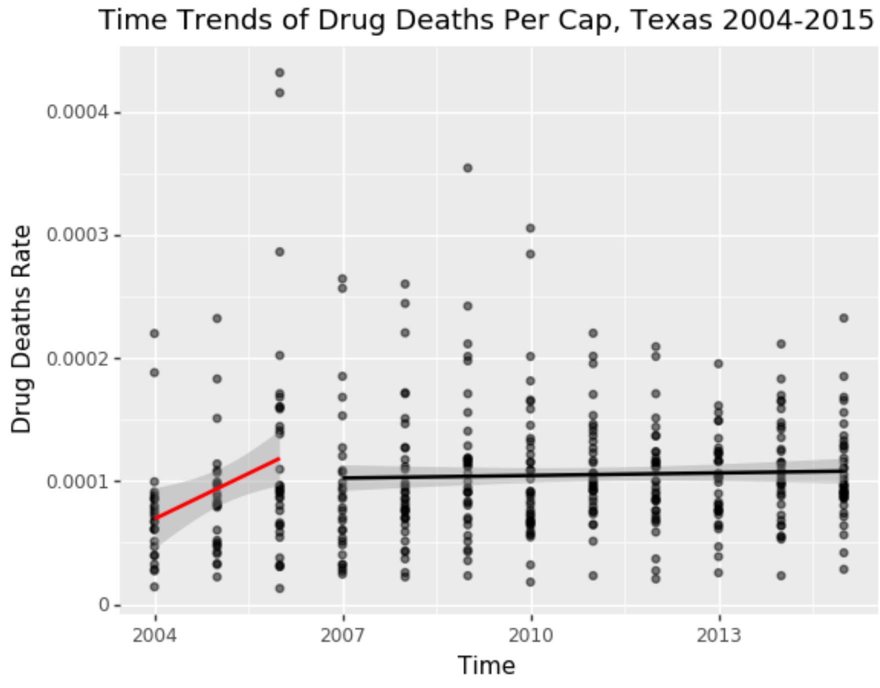
- Pre-Post Analysis

For the pre-post analysis, we create scatter plots of the drug mortality rates of counties in examined states, such as Florida, Texas, and Washington. The points denote the deaths rate in one of the counties over the years. The line shows the trends before and after the policy change in the state. For Florida, as the figure for Florida shows, the slope of the line after the policy change slightly decreases. This looks like potentially the policy change succeeded in reducing mortality if we only take advantage of this pre-post plot.



Likewise, in Texas,we can observe a similar trend to Florida. On the other hand, the trend of the death rate in Washington does not change compared to one from pre-period.





1. Difference-in-Difference Analysis

By using the Difference-in-Difference strategy, we are asking the question of whether the change we saw in treat state (the difference from pre-to-post) is larger than the change that occurred in other states or national-wide over the same period. In graphs, we plot the pre and post trends of policy change for the three states being examined comparing to its control group.

1. Two-fold Control Selection

- **Neighbor Level**

For each treat state, we use 5 of its neighbor states as control group.

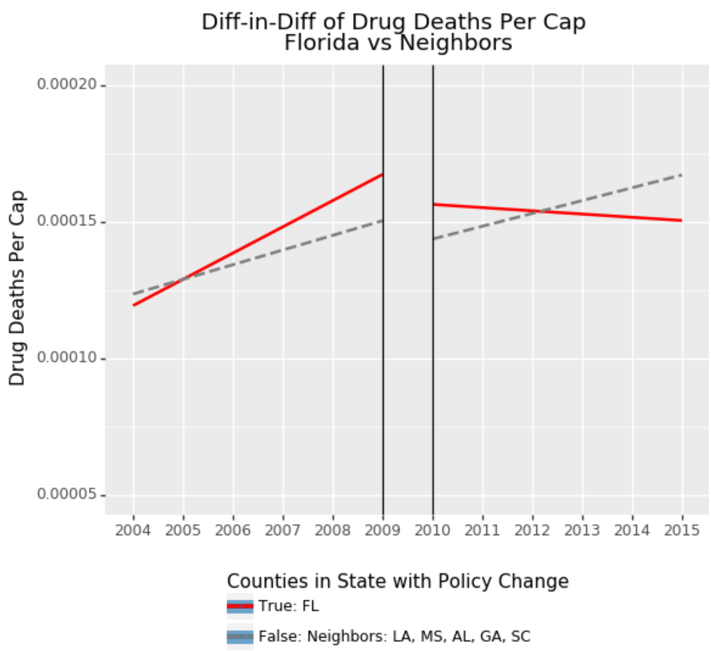
Treat State	**Florida**	**Texas**	**Washington**
Neighbors as control	Louisiana	New Mexico	Oregon
	Mississippi	Oklahoma	Idaho
	South Carolina	Arkansas	Montana
	Alabama	Louisiana	Nevada
	Georgia	Kansas	Wyoming

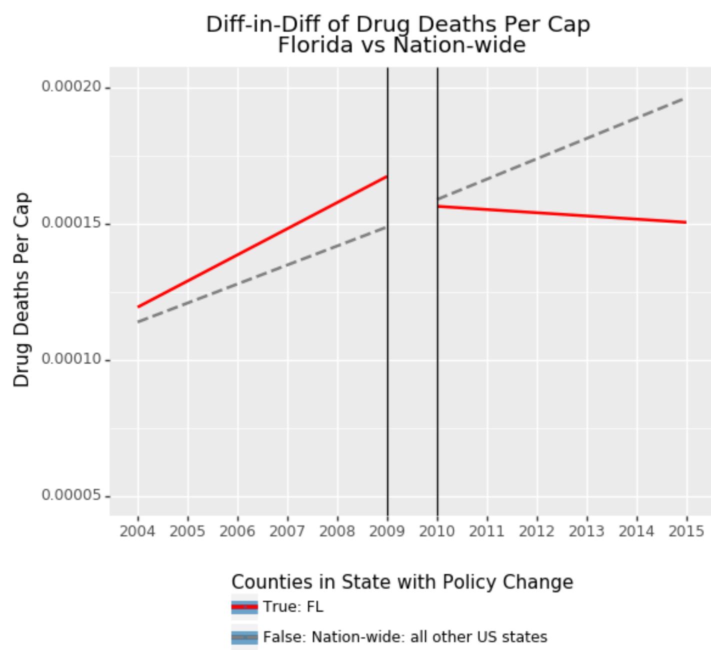
- **National Level**

While neighbor level gives us a more similarity to treat state, looking at national level gives us more statistical power. Given this trade-off, we analyze both to get less biased conclusions. Also, by comparing to the national average level, we could get where the state being examined stands in the pool of all states. For each state being examined, the observations include all counties from all states (excluding Alaska).

2. Results and Analysis

- **Florida**



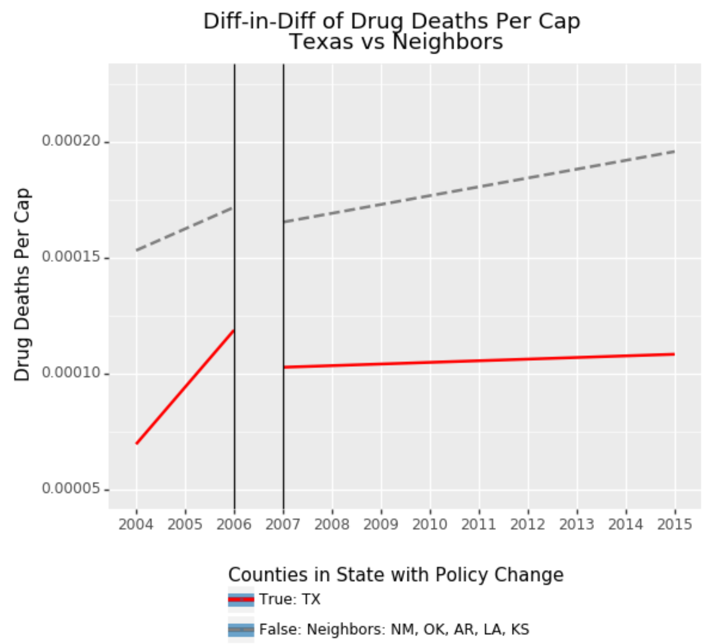


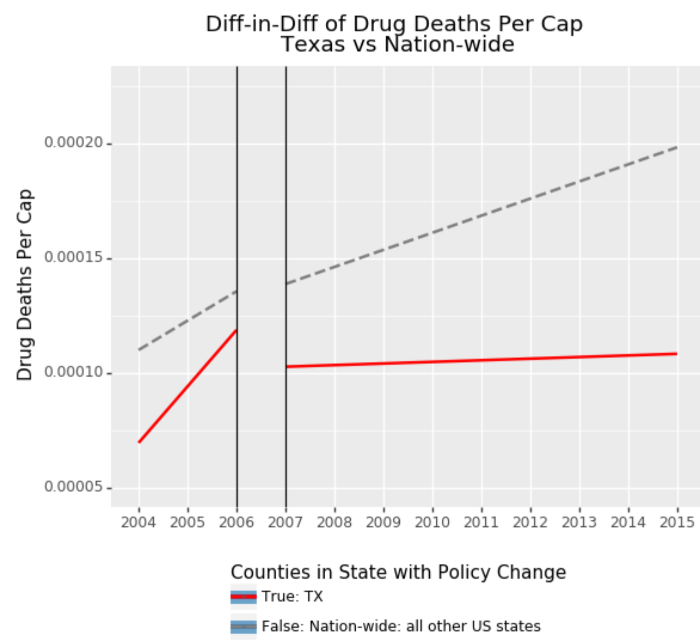
Firstly, we compare Florida to its neighbors. From the graph we can see that before the policy change in 2010, Florida had steeper upward-sloping pre-trends compared to its neighbors, and at a higher level. After the policy change, the trend of Florida significantly became downward-sloping and went to a lower level than its neighbor after 2012. Its neighbors' mortality were lower a little with an up-ward sloping at unchanged steepness. Generally, the graph shows that there was a difference between Florida and its neighbors/similar states for the change of mortality over the period. Specifically, Florida did experience a larger change with even a reversed post-trend and went below the level of its neighbor around 2012.

Then, we compare Florida to nation-wide average. From the graph we can see that Florida still had a higher level of drug mortality than national average during pre-period. After the policy change, the trend of Florida significantly became downward sloping and went less than national average right after 2010. The national trend of drug mortality seems continued after policy change as same as pre-period. Generally, the graph shows that there was a difference between Florida and national average for the change of mortality over the period. Specifically, Florida experienced a larger change with even a reversed post-trend and went below national level right after policy change.

Given the analysis above, we could conclude that the policy change in 2010 did reduce drug mortality in Florida and made it went below the national average.

- **Texas**



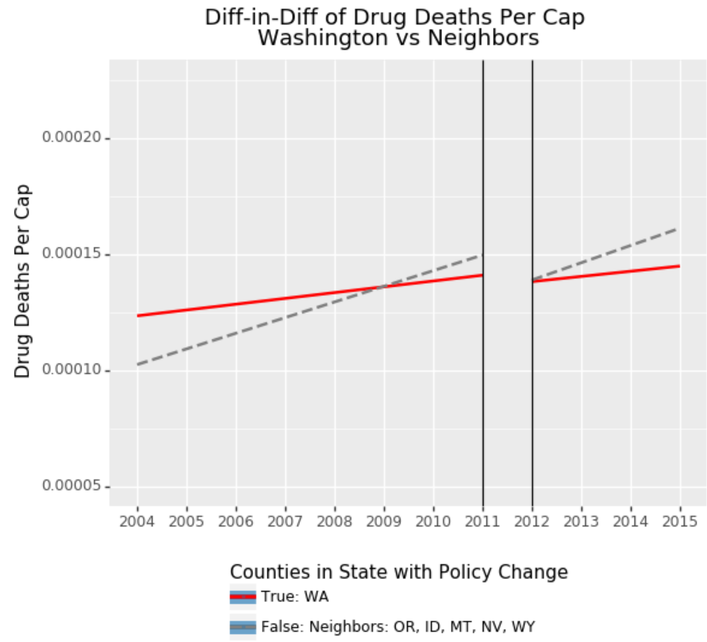


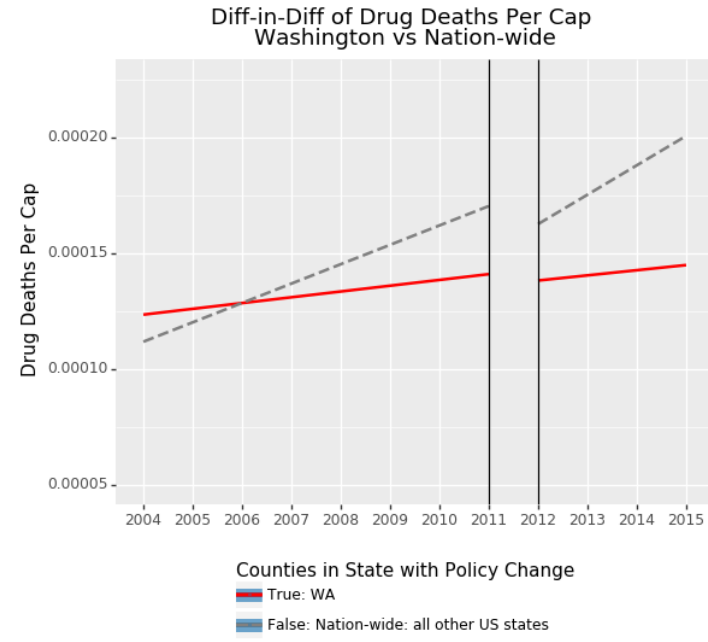
Firstly, we compare Texas to its neighbors. From the graph we can see that before the policy change in 2007, Texas had similar upward-sloping pre-trends with its neighbors, though steeper, but at a significant lower level. After the policy change, the trend of Florida became nearly flat and the level decreased a little. Its neighbors' trend seems to continue as before roughly at the same level. Generally, the graph shows that there was a difference between Texas and its neighbors for the change of mortality over the period. Specifically, Texas experienced a larger change (mainly in the trend) than its neighbor.

Then, we compare Texas to nation-wide average. From the graph we can see that Texas still had a lower level of drug mortality than national average during pre-period. Given a flatter trend of Texas and a continued trend of national average after policy change, the gap grew larger. Generally, the graph shows that there was a difference between Texas and national average for the change of mortality over the period. Specifically, Florida experienced a larger change.

Given the analysis above, we could conclude that even though Texas is of low drug mortality, the policy change in 2007 still reduced drug mortality and kept down its growing trend in the first place.

- **Washington**





Firstly, we compare Washington to its neighbors. From the graph we can see that before the policy change in 2012, Washington had similar upward-sloping pre-trends and level with its neighbors, though slightly flatter. After the policy change, there is nearly no change in the trend and level of mortality in Washington. Its neighbors' trend seems to continue the same as before but followed by a little lower level right after policy change. Generally, there is little change in Washington and its neighbors in terms of mortality over the period.

Then, we compare Washington to nation-wide average. From the graph we can see that national average of mortality level grew over that of Washington after 2006. Since national average of mortality increased to grow after 2006, the gap then became larger as time moving forwards and even bigger after the policy change. Generally, the graph shows that Washington were not responsive to policy change while national average mortality became even worse.

Given the analysis above, we could conclude that Washington gave little respond to its policy change in 2012 and the trend seemed to be stable and flat over the whole period. The growing national level made Washington beame below average in the end.

Conclusion

The core question of our research is that what is the effect of opioid drug prescription regulations on (1) the volume of opioids prescribed, and (2) drug overdose deaths. By using Difference-in-Difference strategy, we could examine whether a policy change on regulation exerted an expected impact on the outcomes of interest.

Our results indicates that for Florida, the treated sample state with a high level of overdose deaths overall, showed obvious responds both to volume of prescription and overdose deaths. They decreased in level and presented a downward-sloping trend after policy change in 2010. The trajectory of prescription and overdose deaths over the period are quite similar, so we can say that they are of high correlation. However, the causal inference on whether the decrease in prescription is the reason of decreased overdose deaths cannot be determined and interpreted merely by this correlation.

Also, for Texas, the treated sample state with a low level of overdose deaths overall, D-in-D analysis also provides evidence of the policy change’s effect on overdose deaths. There is a decrease in the level and a flattened trend after policy change in 2007.

However, Washington seems to experience negligible change to the policy. Among all 3 treat states, WA is the only one that has non-parallel pre-trend as national average. Since D-in-D strategy requires a conventional assumption of parallel pre-trends, we should be careful when interpreting the results.

Difference-in-Difference can provide policy makers with a good estimate of the impact of opioid control policies and support relevant policy evaluation. In general, the three policy change being investigated did have effects, big or small, on the outcomes of our interest, especially for Florida. The next step may be digging further into the question of why the policy seems to be more effective in Florida.