OPIOID DATA SCIENCE CHALLENGE

ABHISHEK BABUJI & HARIKA PUTTI

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

- The primary objective is to transform the given JSON data into Zip Level and State Level CSVs which can be easily used for further Opioid Research. Also, give a brief description of the structure to help in understanding the CSV easily. The description should also include your inferences of the relationship between the JSON Data and the text data visible in the topographical visualization.
- The second objective is to do a trend analysis using different visualizations for the variation in the positivity rate.
- The third objective is to find a correlation between state positivity rate and employment statistics.

1 DESCRIPTION OF THE CSV FILES

Although there could be multiple csv files that could be rendered from the wrangled data, attached with the submission are 5 csv files which are most relevant and wholesome.

1. US.csv: This was obtained directly by wrangling quest-opi.json. The purpose of this csv is to help anyone understand how national rate varied. The reason why this is a separate csv file is to seperate the code US from the rest of the state codes while making visualizations with Tableau.

	Year	Mean Unemployment Rate	Country	Positivity Level	National Rate in %
0	2007	4.616667	US	2	0.32
1	2008	5.800000	US	2	0.34
2	2009	9.283333	US	2	0.39
3	2010	9.608333	US	2	0.34
4	2011	8.933333	US	2	0.36
5	2012	8.075000	US	2	0.37
6	2013	7.358333	US	2	0.39
7	2014	6.175000	US	2	0.39
8	2015	5.266667	US	2	0.41
9	2016	4.866667	US	2	0.40

Figure 1: Structure of US.csv

The DataFrame that was rendered into the csv file has 5 columns namely, Year, Mean Unemployment Rate, Country, Positivity Level and National Rate in % and 10 rows, one observation for each year from 2007 to 2016.

Year: represents the year in which the measurement was recorded ranging from 2007 to 2016.

Mean Unemployment Rate: represents the average of the unemployed percent of the civilian labor force [i.e., 100 times (unemployed/civilian labor force)] aggregated by by year. This was obtained from the Bureau of Labor Statistics.

Country: An indication that the reading is for the country United States, may come in handy when performing joins with other tables having state codes, to distinguish the country code from state code.

Positivity Level: One of the three measures used to represent the severity of opioid crisis. Also present in zipWise.csv, stateWise.csv and stateOpioidEmployment.csv

National Rate in %: The second of the three measures used to represent the severity of opioid crisis.

2. zipWise.csv: This was obtained directly by wrangling quest-opi.json. The purpose of this csv is to help anyone perform further transformations from the records obtained from its raw source. This file had information relating to each of the regions in the 50 states.

```
In [80]: print(zipWise.head())
   year  zipRange  positivityRate state
0  2016         6               0    PR
1  2016         7               0    PR
2  2016         8               0    PR
3  2016         9               0    PR
4  2016        10               2    MA

In [78]: print(zipWise.tail())
      year  zipRange  positivityRate state
8241  2013       995               3    AK
8242  2013       996               1    AK
8243  2013       997               2    AK
8244  2013       998               1    AK
8245  2013       999               3    AK

In [79]: print(zipWise.info())
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8246 entries, 0 to 8245
Data columns (total 4 columns):
year              8246 non-null int64
zipRange          8246 non-null int64
positivityRate    8246 non-null int64
state             8246 non-null object
dtypes: int64(3), object(1)
memory usage: 257.8+ KB
```

Figure 2: Structure of zipWise.csv

The `DataFrame` that was rendered into the `csv` file has 4 columns namely, `year`, `zipRange`, `positivityRate`, and `state`.

`year`: represents the year in which the measurement was recorded ranging from 2007 to 2016.

`zipRange`: represents the 3 digit zip code ranging from 006 to 999 is used to represent various regions belonging to a state.

`positivityRate`: is one of the 3 measures used to indicate the severity of opioid concentration based on the range of `stateRate`. The severity represented by `positivityLevel` is given below. The same column is present in `stateWise.csv` but renamed as `positivityLevel`. Both `positivityRate` and `positivityLevel` represent the same severity metric and can take a range between 0 to 4.

`state`: represents the set of uppercase abbreviations established by United States Postal Service to represent the 50 states in the United States.

3. `stateWise.csv`: This was also obtained directly by wrangling `quest-opi.json`. This file had cumulative information of each the 50 states instead and the positivity rate of US as a whole from the years 2006 to 2017.

```
print(stateWise.head())
```

```
   year state  positivityLevel  stateRate
0  2016    US                2   0.400000
1  2016    PR                0   0.109439
2  2016    MA                1   0.247107
3  2016    RI                1   0.147954
4  2016    NH                1   0.203089
```

```
print(stateWise.tail())
```

```
     year state  positivityLevel  stateRate
535  2013    HI                2   0.327250
536  2013    XX                0   0.065360
537  2013    OR                2   0.409043
538  2013    WA                2   0.336689
539  2013    AK                3   0.460482
```

```
print(stateWise.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 540 entries, 0 to 539
Data columns (total 4 columns):
year              540 non-null int64
state             540 non-null object
positivityLevel   540 non-null int64
stateRate         540 non-null float64
dtypes: float64(1), int64(2), object(1)
```

Figure 3: Structure of `stateWise.csv`

The `DataFrame` that was rendered into the `csv` file has 4 columns namely, `year`, `state`, `positivityLevel`, and `stateRate`.

`year`: represents the year in which the measurement was recorded ranging from 2007 to 2016.

`state`: represents the set of uppercase abbreviations established by United States Postal Service to represent the 50 states in the United States.

`stateRate`: is one of three measures to represent opioid severity. It indicares the opioid positivity percentage of the respective state in that respective year. The state positivity percentage ranges from 0% to 1.33% as you see in Figure X.

`positivityLevel`: is the second of the 3 measures used to indicate the severity of opioid concentration based on the range of `stateRate`. The severity represented by `positivityLevel` is given below.

positivityLevel	min	max
0	0.0000	0.1387
1	0.1445	0.2592
2	0.2610	0.4593
3	0.4605	0.8447
4	0.8624	1.3320

Figure 4: Severity represented by `positivityLevel` based on the range of State Positivity percentage

4. `stateOpioidEmployment.csv`: This `csv` is used to understand how the three measures to represent the severity of opioid crisis:
   a) `Opioid Severity`: Computed from `zipWise.csv`
   b) `Positivity Level`: Already present in `stateWise.csv`
   c) `State Rate in %`: Already present in `stateWise.csv`

   affected employment statistics such as Labor Force and Unemployment Rate in each state from the years 2006 to 2017.

```
In [98]: print(stateOpioidEmployment.head())
   Year  Opioid Severity  Positivity Level  State Rate in %  \
0  2007         2.611111                 3         0.512982
1  2008         2.666667                 3         0.590731
2  2009         3.176471                 3         0.767378
3  2010         2.944444                 3         0.778664
4  2011         3.444444                 4         0.934580

   Average Labor Force  Average # of employed subjects  \
0         2.175635e+06                    2.089058e+06
1         2.177421e+06                    2.053104e+06
2         2.163393e+06                    1.925772e+06
3         2.195945e+06                    1.964694e+06
4         2.202401e+06                    1.990570e+06

   Average # of unemployed subjects  Mean Unemployment Rate State
0                      86576.583333                   3.975000    AL
1                     124317.916667                   5.716667    AL
2                     237621.166667                  10.991667    AL
3                     231250.750000                  10.541667    AL
4                     211831.333333                   9.616667    AL
```

Figure 5: Structure of `stateOpioidEmployment.csv`

The `DataFrame` that was rendered into the `csv` file has 9 columns namely, `Year`, `Opioid Severity`, `Positivity Level`, `State Rate in %`, `Average Labor Force`, `Average # of employed subjects`, `Average # of unemployed subjects`, `Mean Unemployment Rate` and `State`.

`Year`: represents the year in which the measurement was recorded ranging from 2007 to 2016.

`Opioid Severity`: is the third metric which was generated from `zipWise.csv` to measure severity of opioid concentration. An example of how this new metric `Opioid Severity` was computed is given below.

In `zipWise.csv`, if you can recall, there were many regions represented by the 3 digit zip code within the same state which had different `positivityRate`. The `DataFrame` was grouped by `state` and then by `year` and then finally by `positivityRate`, to obtain an aggregation shown in

	state	year	positivityRate	Number of regions
0	AK	2007	1	3
1	AK	2007	2	2
2	AK	2008	0	2
3	AK	2008	1	1
4	AK	2008	2	2
5	AK	2009	0	1
6	AK	2009	2	2
7	AK	2009	3	2
8	AK	2010	0	2
9	AK	2010	1	1
10	AK	2010	3	2
11	AK	2011	0	1
12	AK	2011	1	2
13	AK	2011	2	2
14	AK	2012	0	3
15	AK	2012	1	1
16	AK	2012	3	1
17	AK	2013	1	2
18	AK	2013	2	1
19	AK	2013	3	2
20	AK	2014	1	1
21	AK	2014	2	3

Figure 6: `zipWise.csv` grouped by `state, year, positivityRate`

Let's take the year 2008 in the state `AK`

2	AK	2008	0	2
3	AK	2008	1	1
4	AK	2008	2	2

Figure 7: Grouped `zipWise.csv` in the year 2008, Alaska

The `Opioid Severity` for the state `AK` in the `Year` 2008 can be calculated as the weighted average:

$$\frac{(0\times 2)+(1\times 1)+(2\times 2)}{2+1+2} = 2.6111$$

`State Rate in %`: It is obtained from `stateWise DataFrame` by performing a join with it. It is one of the three measures used to represent severity of opioid concentration.

`Positivity Level`: It is obtained from `stateWise DataFrame` by performing a join with it. It is the second of the three measures used to represent severity of opioid concentration.

`Average Labor Force`: This is the average number of persons aggregated by state by year in the civilian non-institutional population ages 16 and older classified as either employed or unemployed as obtained from the Bureau of Labor Statistics [1].

`Average # of employed subjects`: This is the average of number of all persons aggregated by state by year who, during the reference week (the week including the 12th day of the month), (a) did any work as paid employees, worked in their own business or profession or on their own farm, or worked 15 hours or more as unpaid workers in an enterprise operated by a member of their family, or (b) were not working but who had jobs from which they were temporarily absent because of vacation, illness, bad weather, childcare problems, maternity or paternity leave, labor-management dispute, job training, or other family or personal reasons, whether or not they were paid for the time off or were seeking other jobs. Each employed person is counted only once, even if he or she holds more than one job. This was obtained from the Bureau of Labor Statistics [1].

`Average # of unemployed subjects`: This is the average of the number of all persons aggregated by state by year who had no employment during the reference week, were available for work, except for temporary illness, and had made specific efforts to find employment some time during the 4 week-period ending with the reference week. Persons who were waiting to be recalled to a job from which they had been laid off need not have been looking for work to be classified as unemployed. This was obtained from the Bureau of Labor Statistics [1].

`Mean Unemployment Rate`: This is the average of the unemployed percent of the civilian labor force [i.e., 100 times (unemployed/civilian labor force)] aggregated by state by year. This was obtained from the Bureau of Labor Statistics. [1].

`State`: represents the set of uppercase abbreviations established by United States Postal Service to represent the 50 states in the United States.

5. employmentCorrelations.csv: This csv has a numerical measure (Pearson Correlation Coefficient) to represent the correlation between the three severity metrics, Opioid Severity, Positivity Level and State Rate in % and the employment statistics such as Labor Force and Unemployment Rate in each state from the years 2006 to 2017.

```
In [99]: print(employmentCorrelations.head())
                Metric  Average Labor Force  Average # of employed subjects  \
0   Opioid Severity             0.032099                       -0.539349
1  Positivity Level            -0.095552                       -0.057370
2   State Rate in %            -0.030288                       -0.423044
0   Opioid Severity             0.151029                        0.125624
1  Positivity Level             0.796575                        0.693495

   Average # of unemployed subjects  Mean Unemployment Rate State
0                          0.505670                0.509066    AL
1                          0.025662                0.027276    AL
2                          0.380864                0.384192    AL
0                          0.103926                0.089673    AK
1                          0.482046                0.386715    AK
```

Figure 8: Structure of employmentCorrelations.csv

Metric: represents the severity metric that is being compared with which one of the employment statistic.

Average Labor Force: represents the Pearson Correlation Coefficient between this employment statisic and one of the severity metrics.

Average # of employed subjects: represents the Pearson Correlation Coefficient between this employment statisic and one of the severity metrics.

Average # of unemployed subjects: represents the Pearson Correlation Coefficient between this employment statisic and one of the severity metrics.

Mean Unemployment Rate: represents the Pearson Correlation Coefficient between this employment statisic and one of the severity metrics.

State: represents the state for which the Pearson Correlation Coefficient between a severity metric and an employment statistic is being calculated.

For example, in Figure X, the first row indicates that for the state AL (Alaska):
$\rho(\text{Opioid Severity}, \text{Average Labor Force}) = 0.0320$
$\rho(\text{Opioid Severity}, \text{Average \# of employed subjects}) = -0.5393$
$\rho(\text{Opioid Severity}, \text{Average \# of unemployed subjects}) = 0.5056$
$\rho(\text{Opioid Severity}, \text{Mean Unemployment Rate}) = 0.5090$

2 quest-topo.json

2.1 TopoJSON format

The are many formats designed for representing simple geographical features, along with their non-spatial attributes. The given file quest-topo.json is in an open standard format called as TopoJSON. TopoJSON is an EXTENSION of GeoJSON, that encodes geospatial topology and that typically provides smaller file sizes. It is mainly used to create interactive mapping with d3.js like you see in http://www.dtidrugmap.com/. Inside quest-topo.json contains fields like STUSPS, LSAD, ALAND, STATEFP, AWATER, GEOID, NAME, STATENS and AFFGEOID. These fields are standards that are used to create visualizations of geographical features.

2.2 Backtracking the source of quest-topo.json

Our guess is that They were created from cartographic boundary shape files from census.gov.

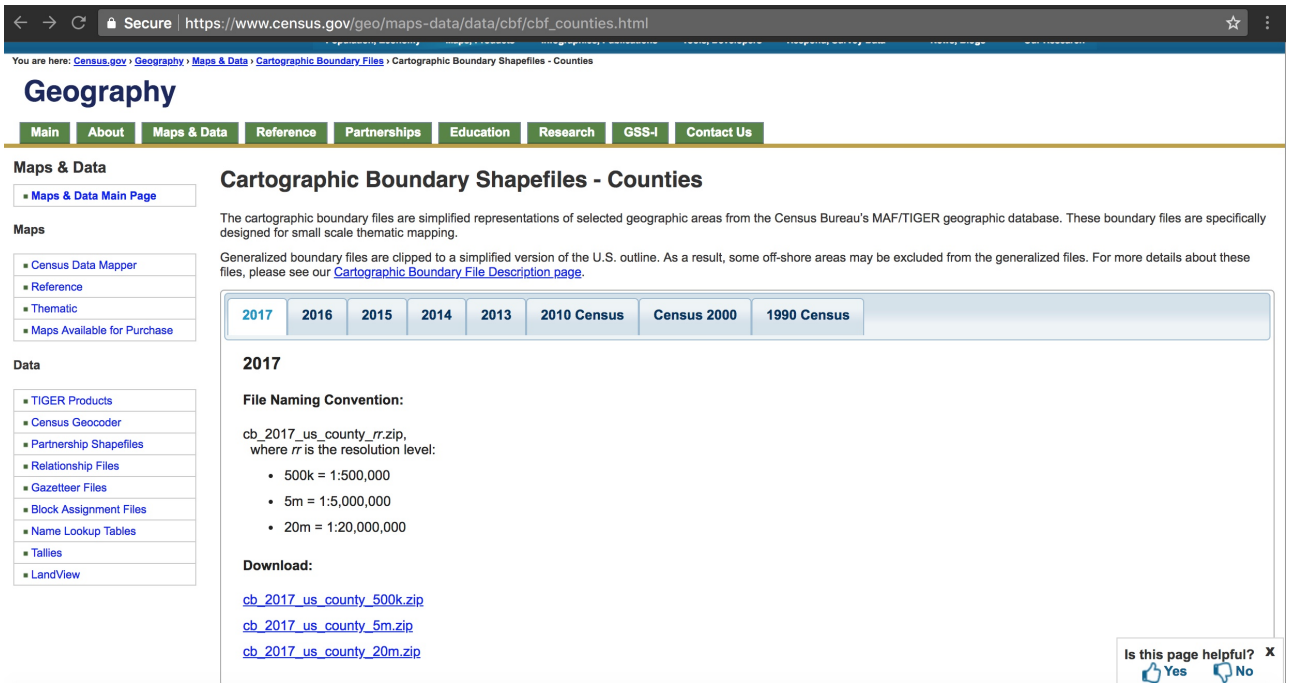

Figure 9: Potential Source of `quest-opi.json`

2.3 Example of a region recreated after converting TopoJSON to GeoJSON from `geojson.io`

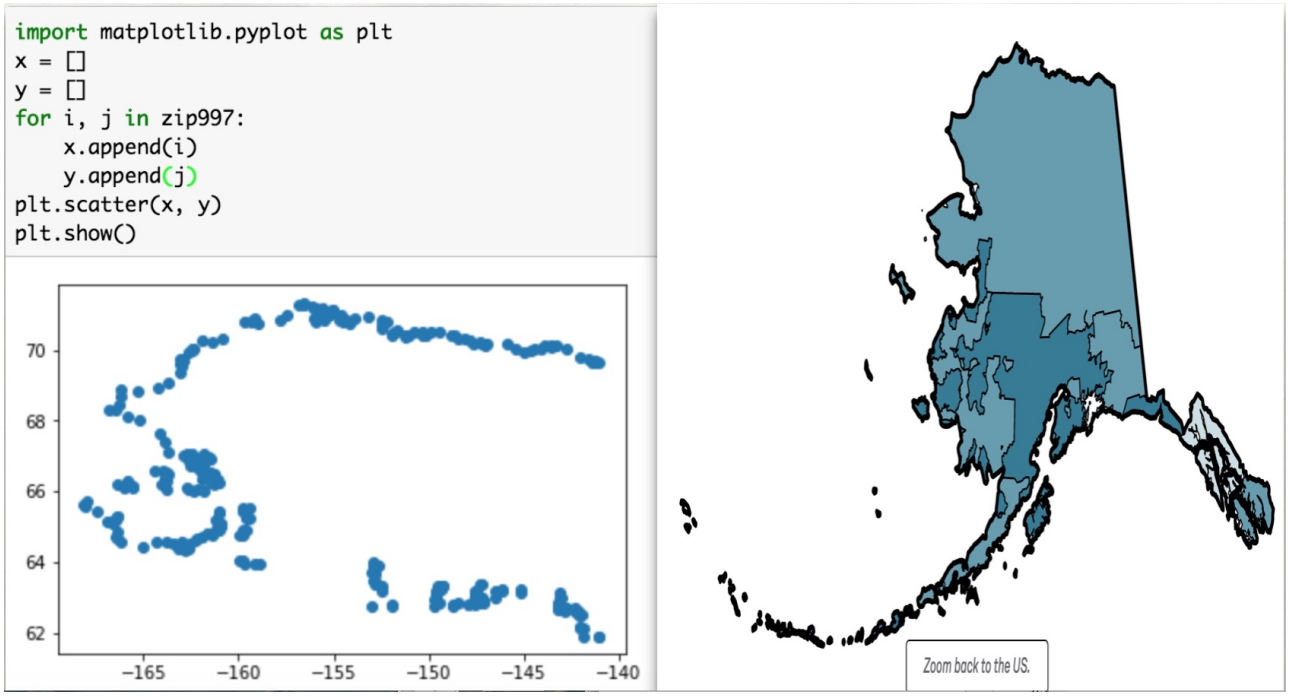

Figure 10: On the left, is a small portion of the map recreated from the given json file, On the right is the original portion of the map as seen on http://www.dtidrugmap.com/

3 VISUALIZATIONS

3.1 Variations in Severity of Opioid Crisis in States from 2007 to 2016

For a full set of visualizations in its most view-able form, visit this Tableau link. The Heat Maps were generated using the Latitude and Longitude values of each state using Tableau. Although numerous visualizations

can be drawn, the current section helps understand the concentration of opioid severity over the years using Heat Maps, Bar Graphs and Packed Bubble Chart.

3.1.1 *Over all years*

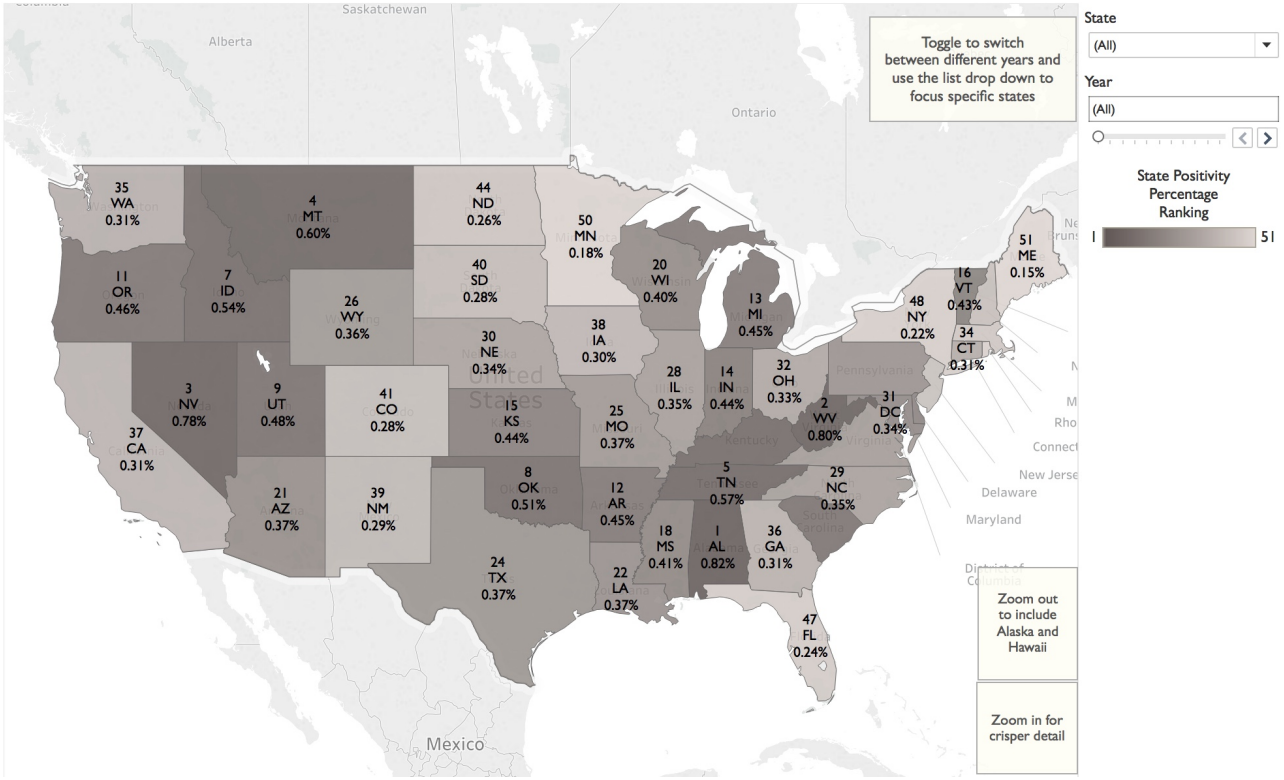

Figure 11: State Positivity % Ranking over all years combined: Heat Map

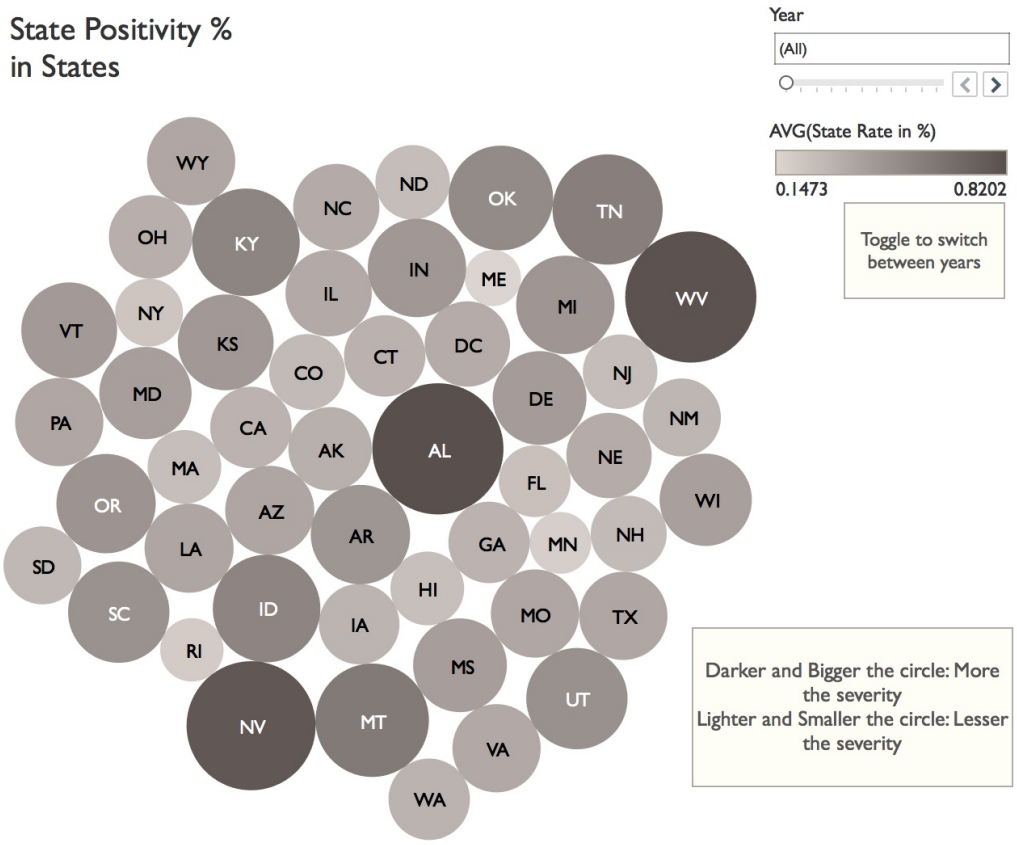

Figure 12: State Positivity % Ranking over all years combined: Packed Bubble Chart

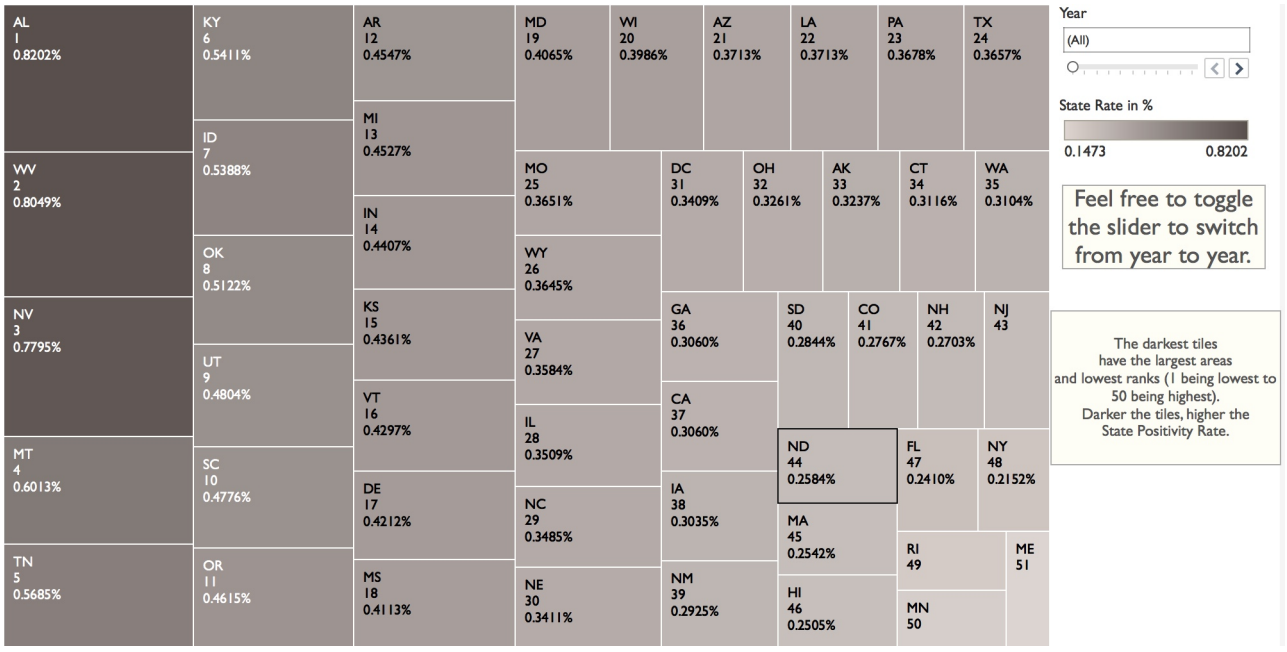

Figure 13: State Positivity % Ranking over all years combined: Tree Map

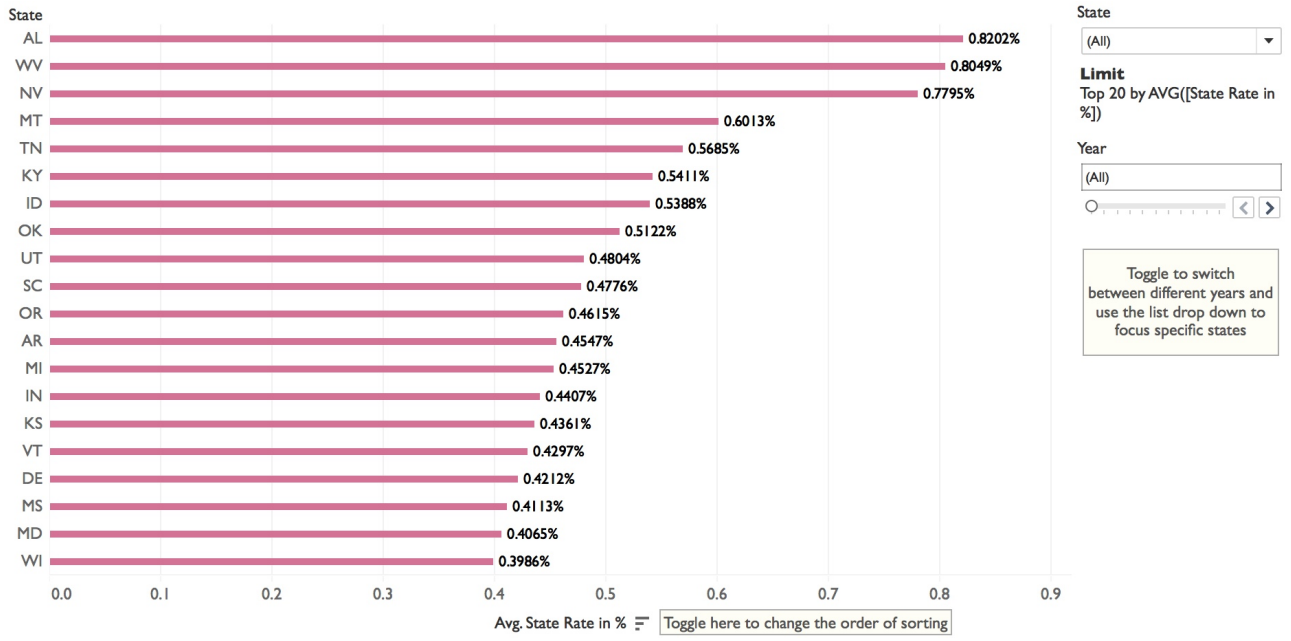

Figure 14: State Positivity % Ranking: Top 20 States with Most Severe Opioid Concentration - over all years

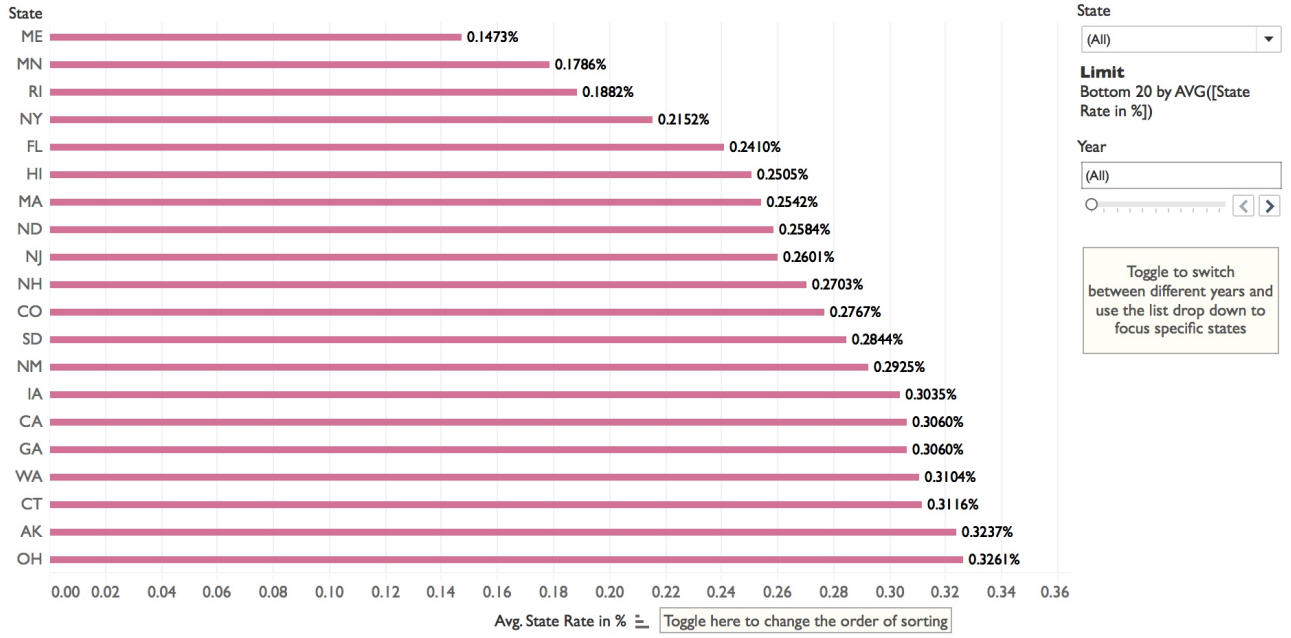

Figure 15: State Positivity % Ranking: Top 20 States with Least Severe Opioid Concentration - over all years

Over all the years, `AL` - Alabama, `WV` - West Virginia and `NV` - Nevada have been the states with the highest Opioid Severity. `ME` - Maine, `MN` - Minnesota and `RI` - Rhode Island have had the least Opioid Severity (using State Positivity % as the severity metric).

3.1.2 *2007*

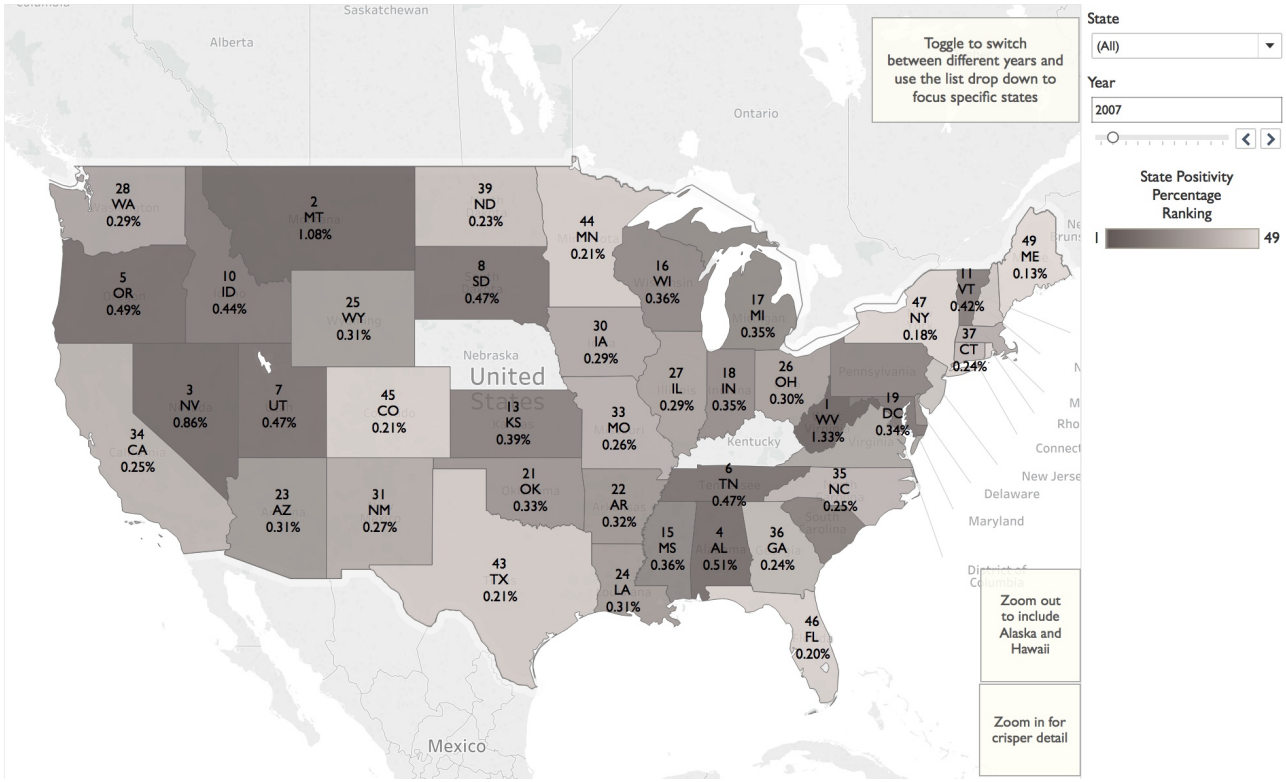

Figure 16: State Positivity % Ranking - 2007: Heat Map

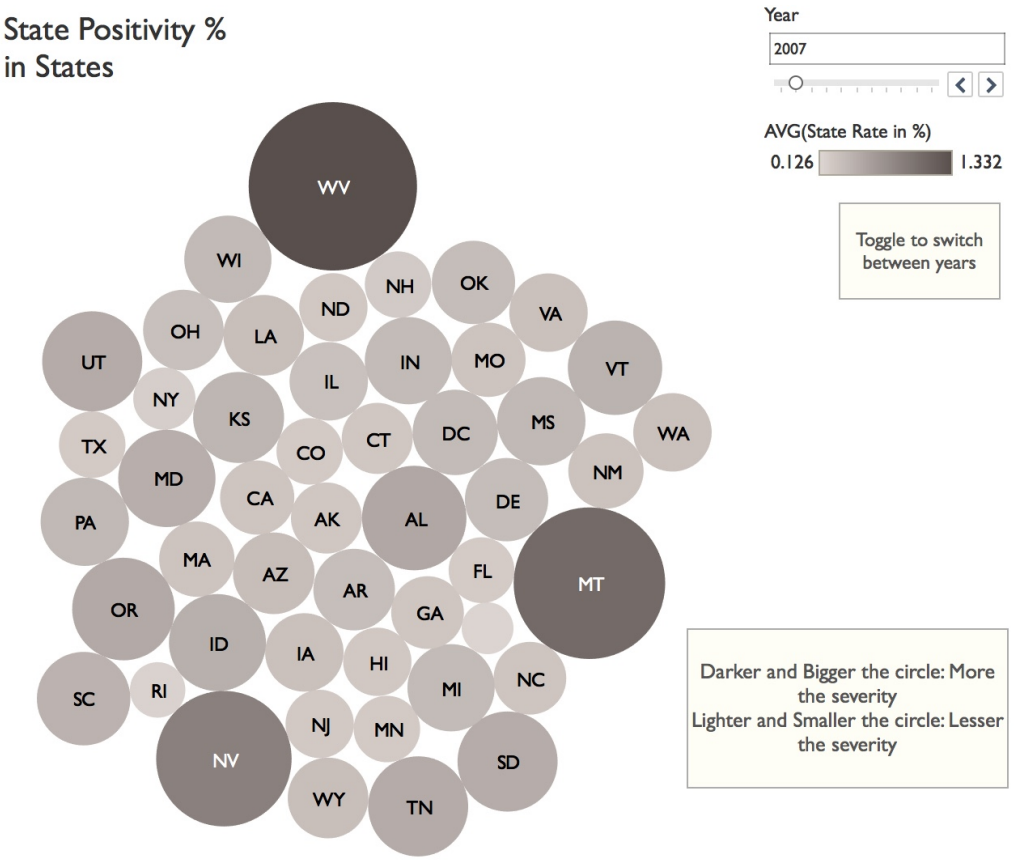

Figure 17: State Positivity % Ranking - 2007: Packed Bubble Chart

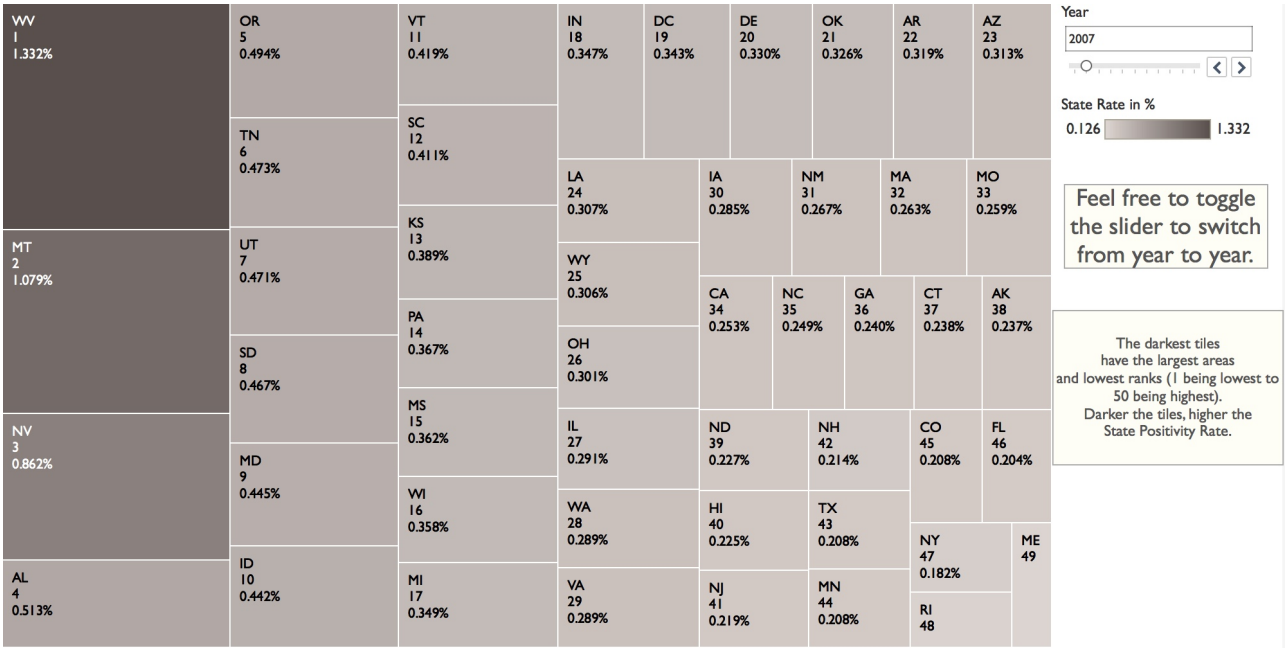

Figure 18: State Positivity % Ranking - 2007: Tree Map

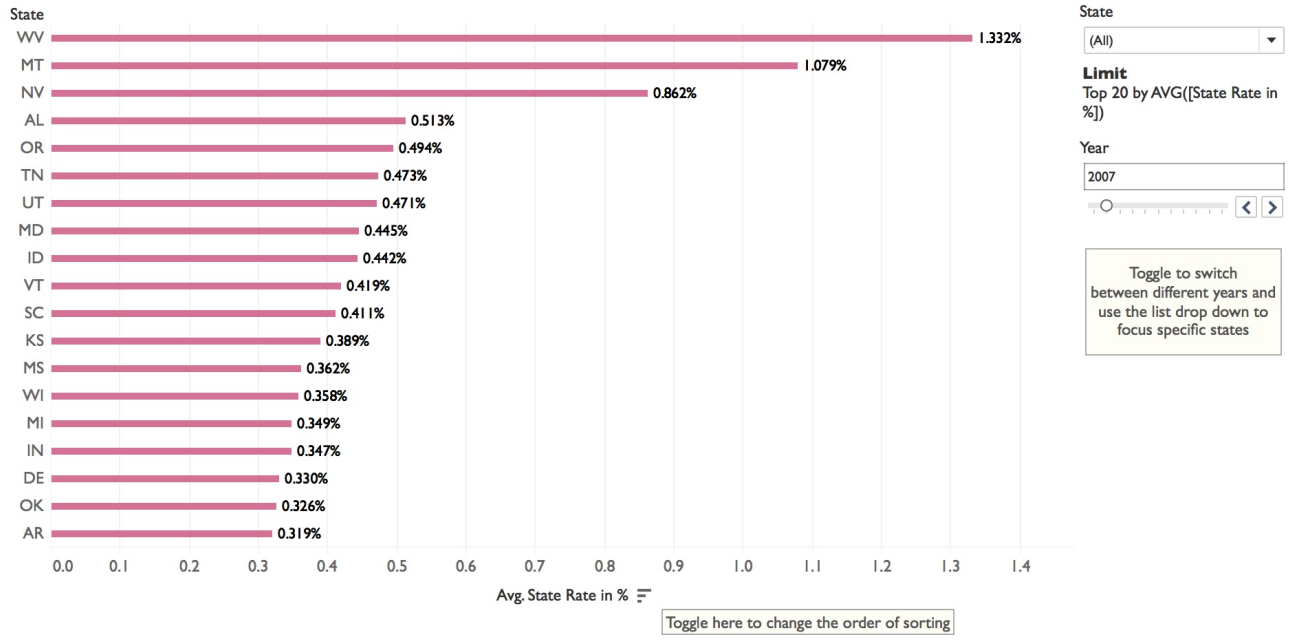

Figure 19: State Positivity % Ranking: Top 20 States with Most Severe Opioid Concentration - 2007

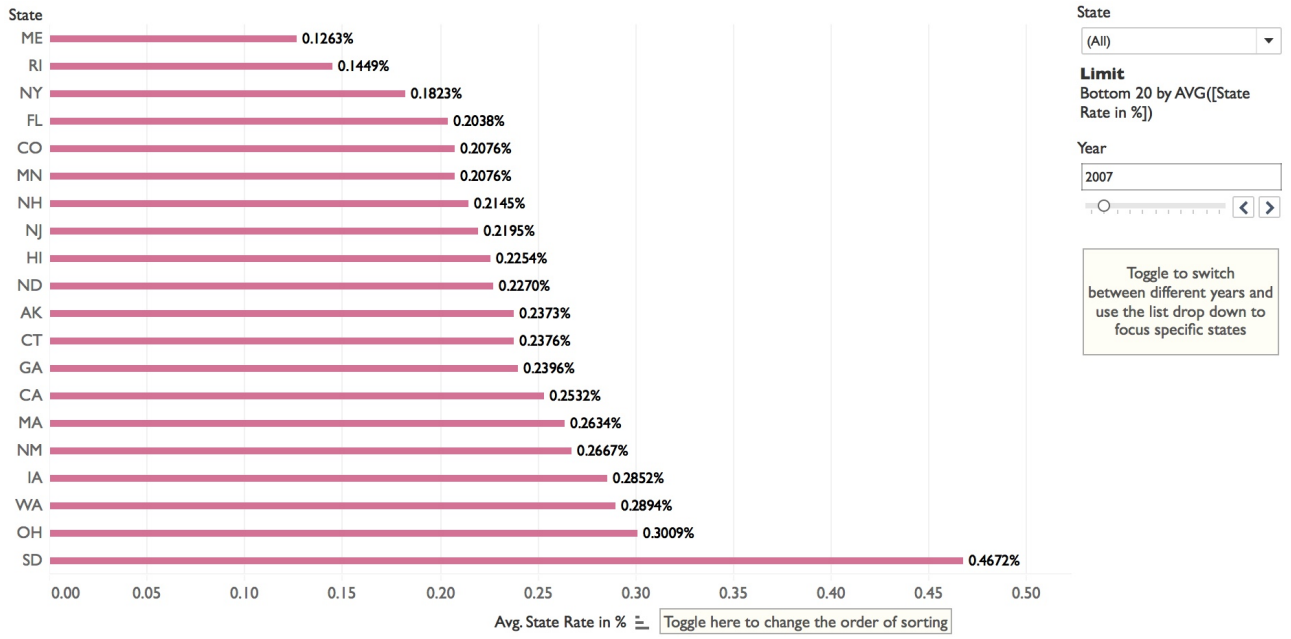

Figure 20: State Positivity % Ranking: Top 20 States with Least Severe Opioid Concentration - 2007

In the year 2007, `WV` - West Virginia, `MT` - Montana and `NV` - Nevada have been the states with the highest Opioid Severity. `ME` - Maine, `RI` - Rhode Island and `NY` - New York have had the least Opioid Severity (using State Positivity % as the severity metric).

3.1.3 *2008*

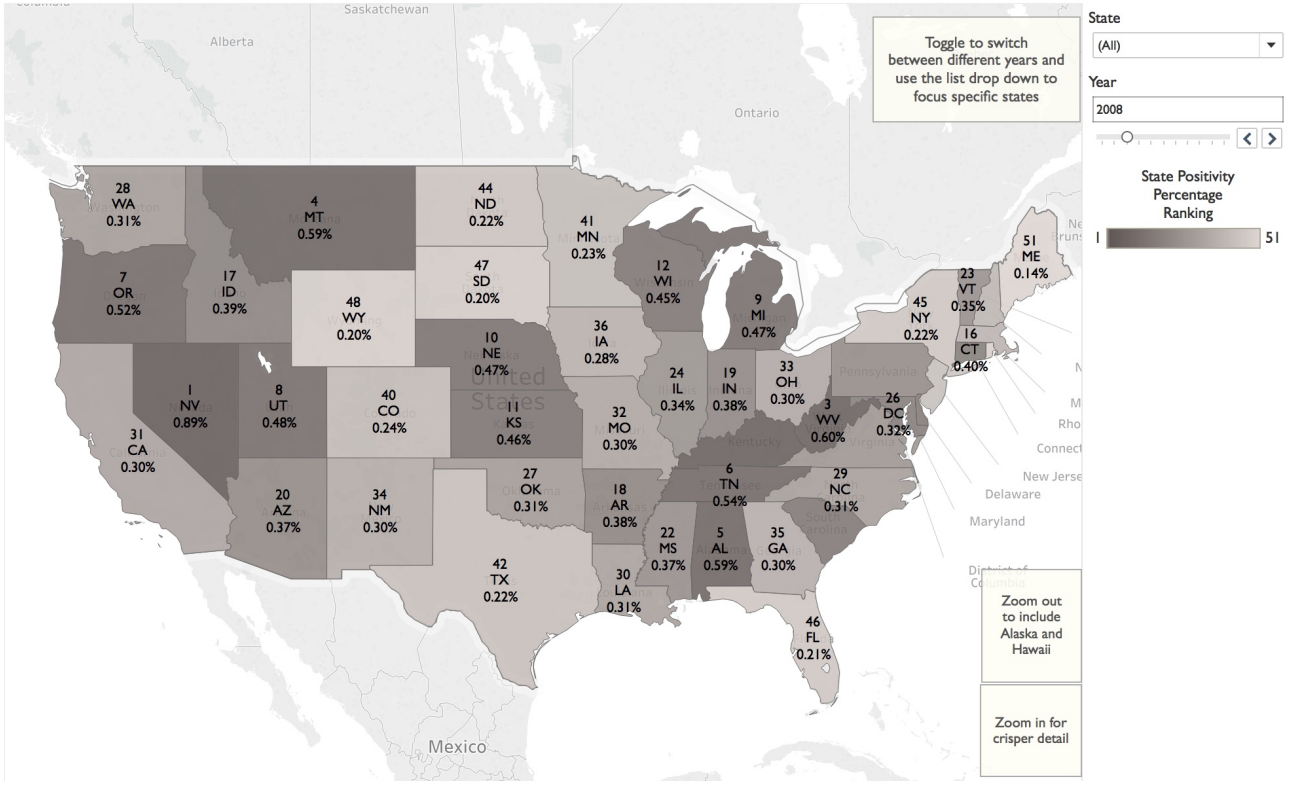

Figure 21: State Positivity % Ranking - 2008: Heat Map

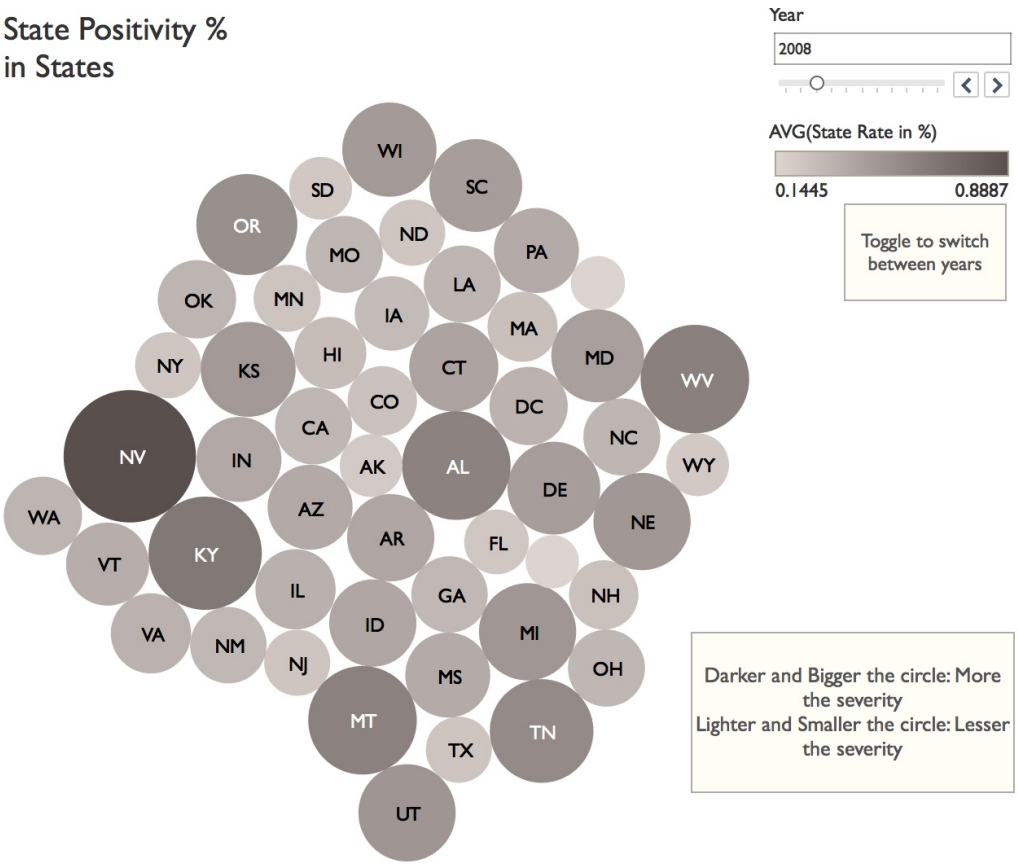

Figure 22: State Positivity % Ranking - 2008: Packed Bubble Chart

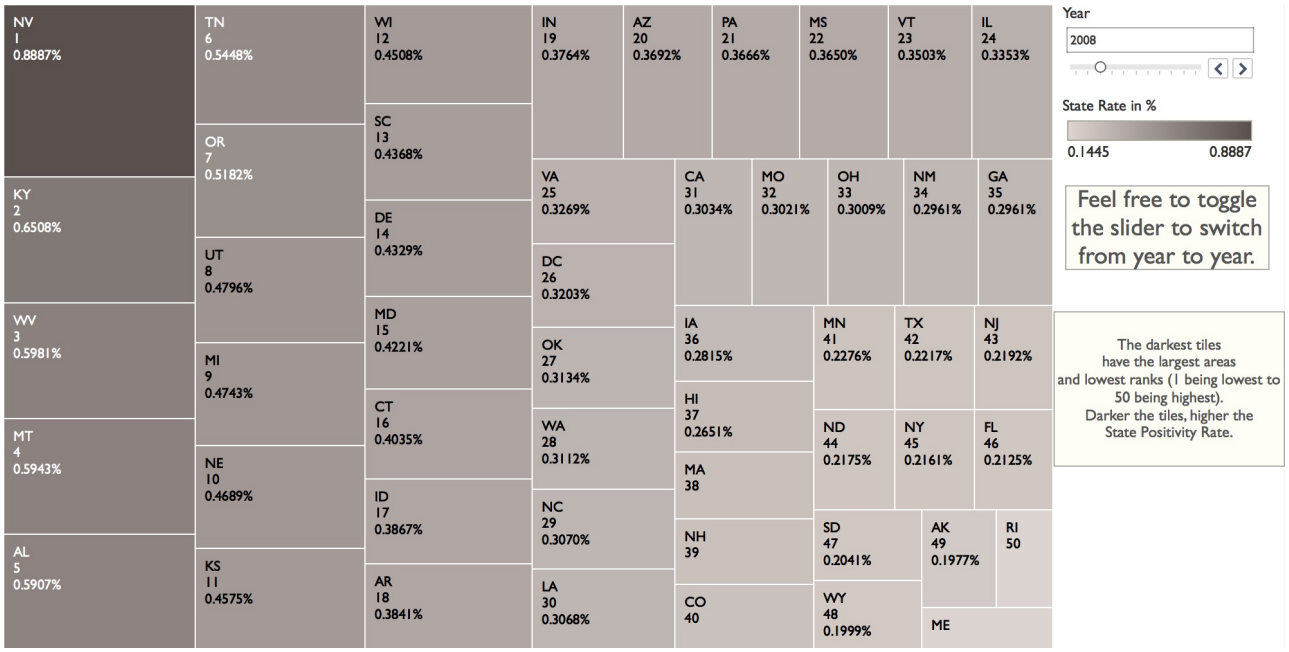

Figure 23: State Positivity % Ranking - 2008: Tree Map

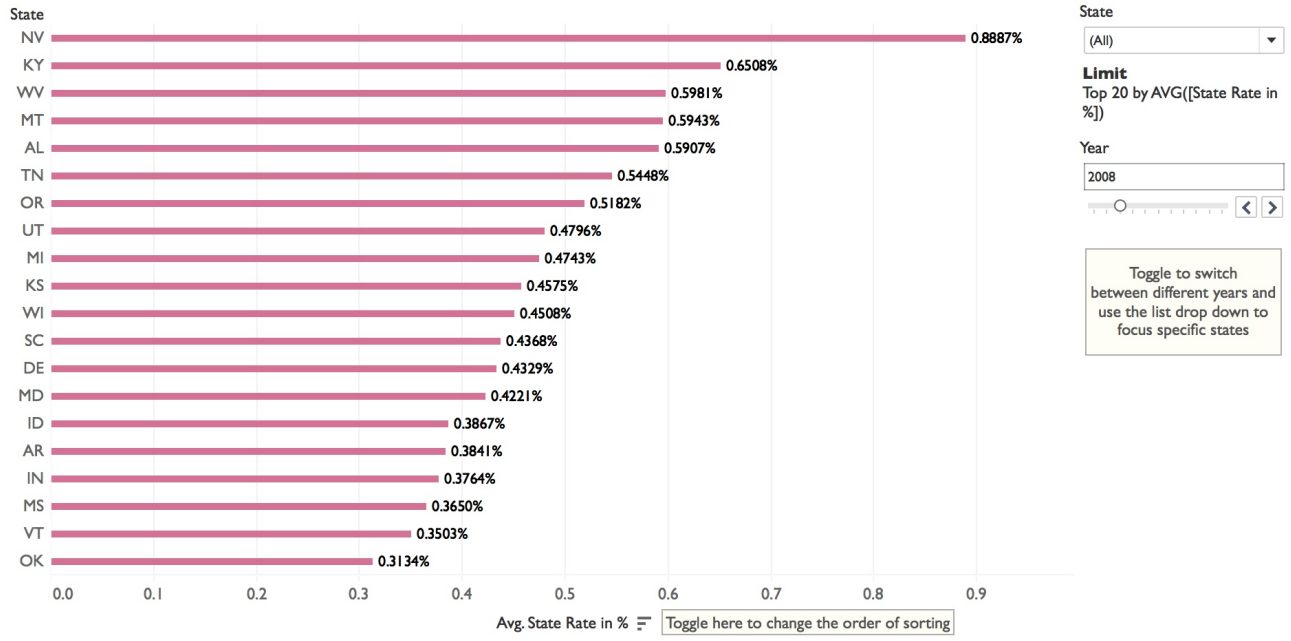

Figure 24: State Positivity % Ranking: Top 20 States with Most Severe Opioid Concentration - 2008

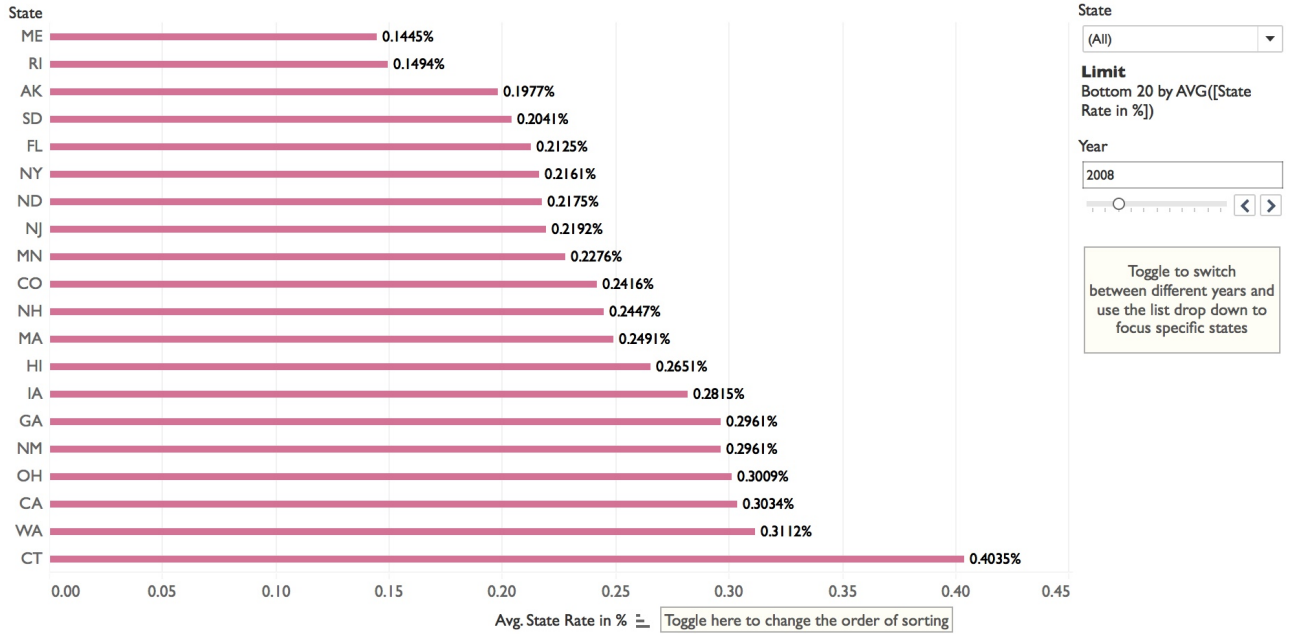

Figure 25: State Positivity % Ranking: Top 20 States with Least Severe Opioid Concentration - 2008

In the year 2008, `NV` - Nevada, `KY` - Kentucky and `WV` - West Virginia have been the states with the highest Opioid Severity. `ME` - Maine, `RI` - Rhode Island and `AK` - Alaska have had the least Opioid Severity (using State Positivity % as the severity metric).

3.1.4 *2009*

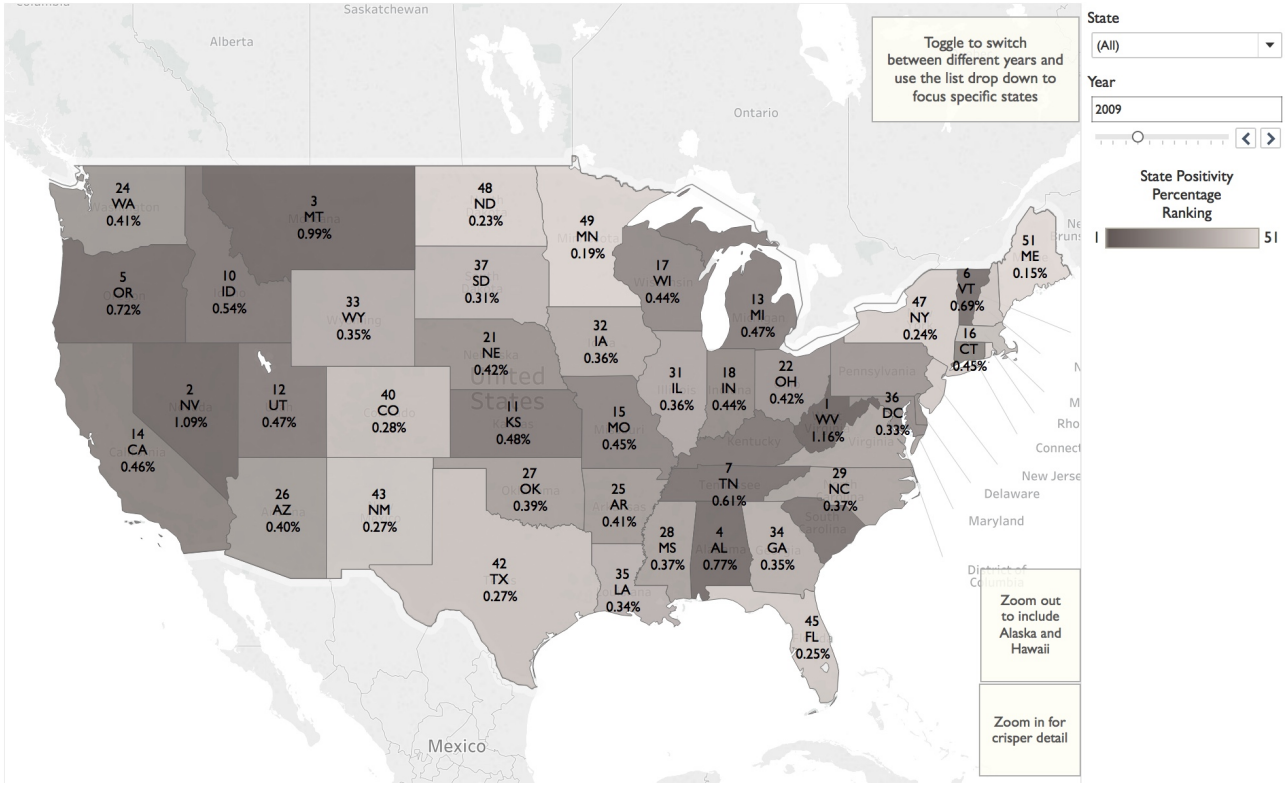

Figure 26: State Positivity % Ranking - 2009: Heat Map

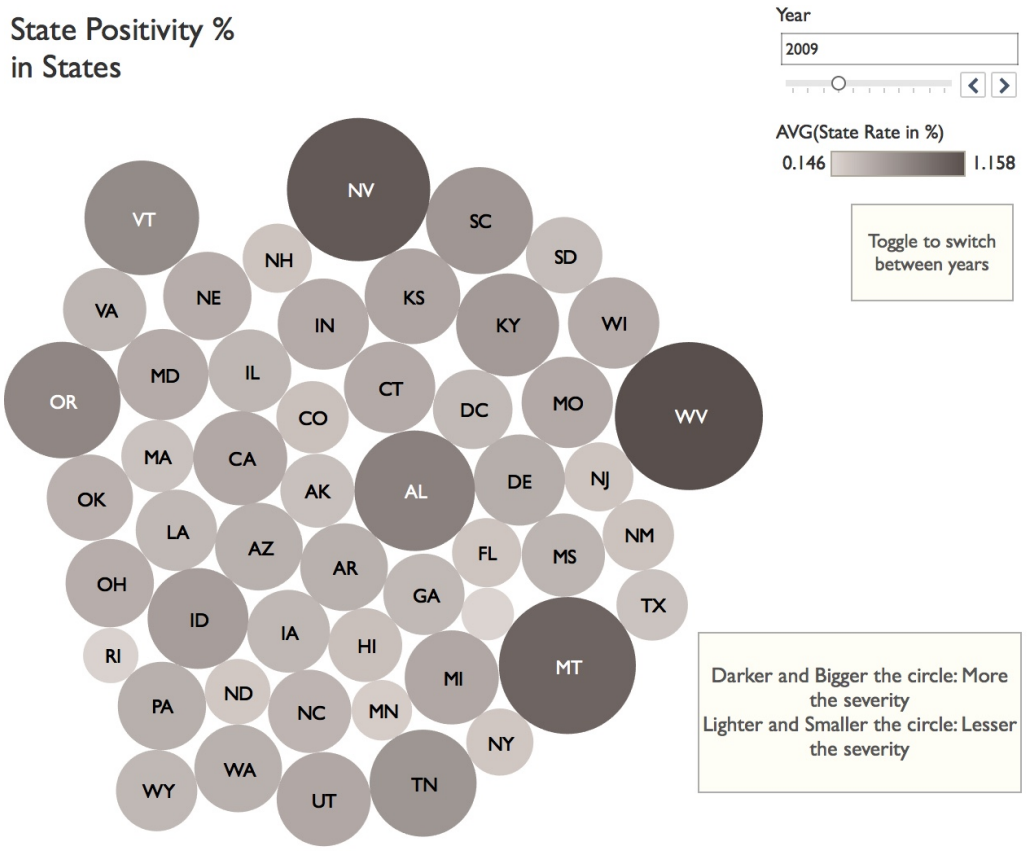

Figure 27: State Positivity % Ranking - 2009: Packed Bubble Chart

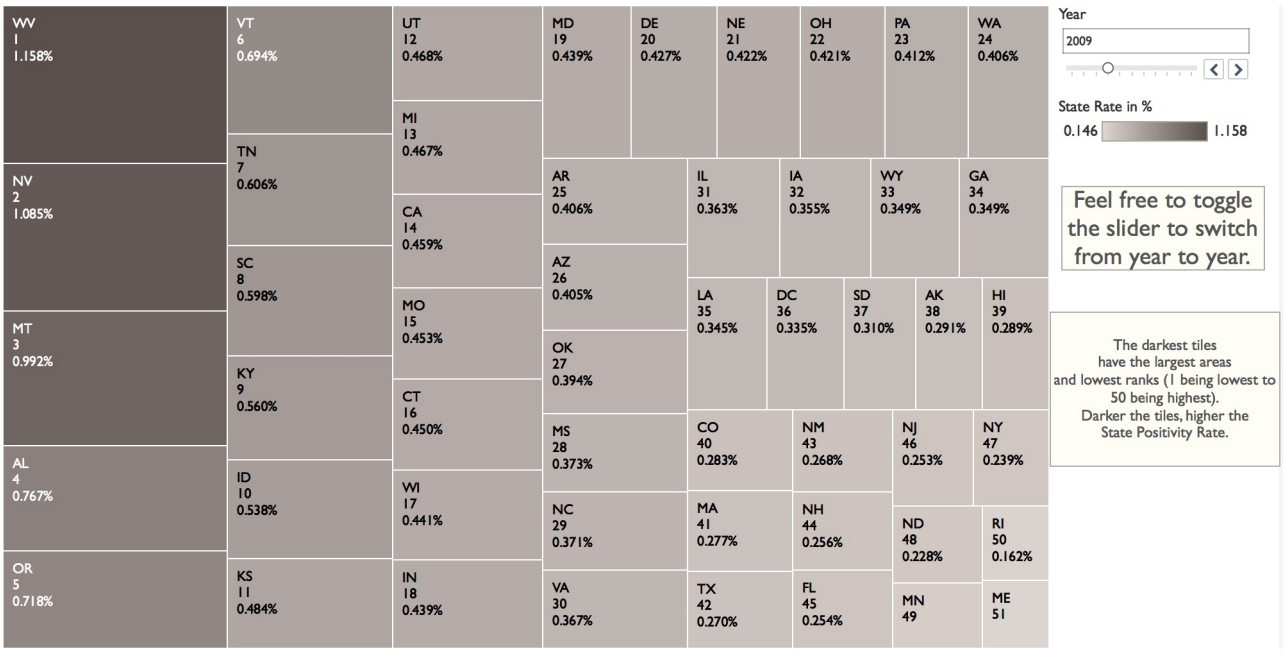

Figure 28: State Positivity % Ranking - 2009: Tree Map

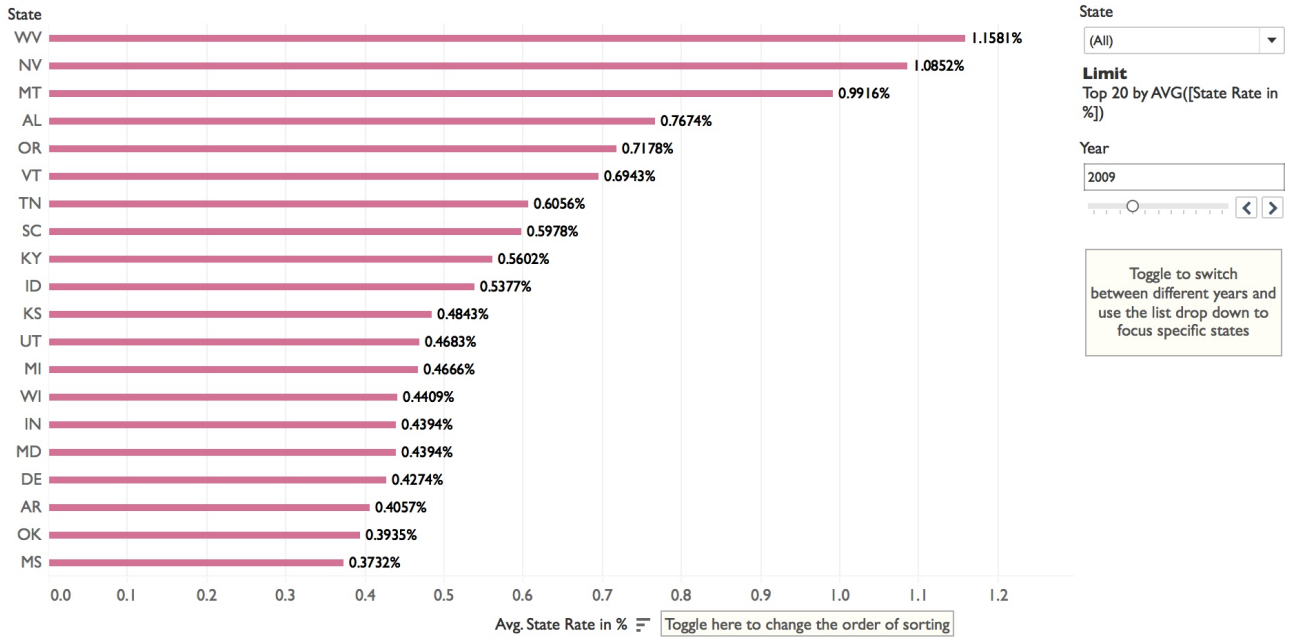

Figure 29: State Positivity % Ranking: Top 20 States with Most Severe Opioid Concentration - 2009

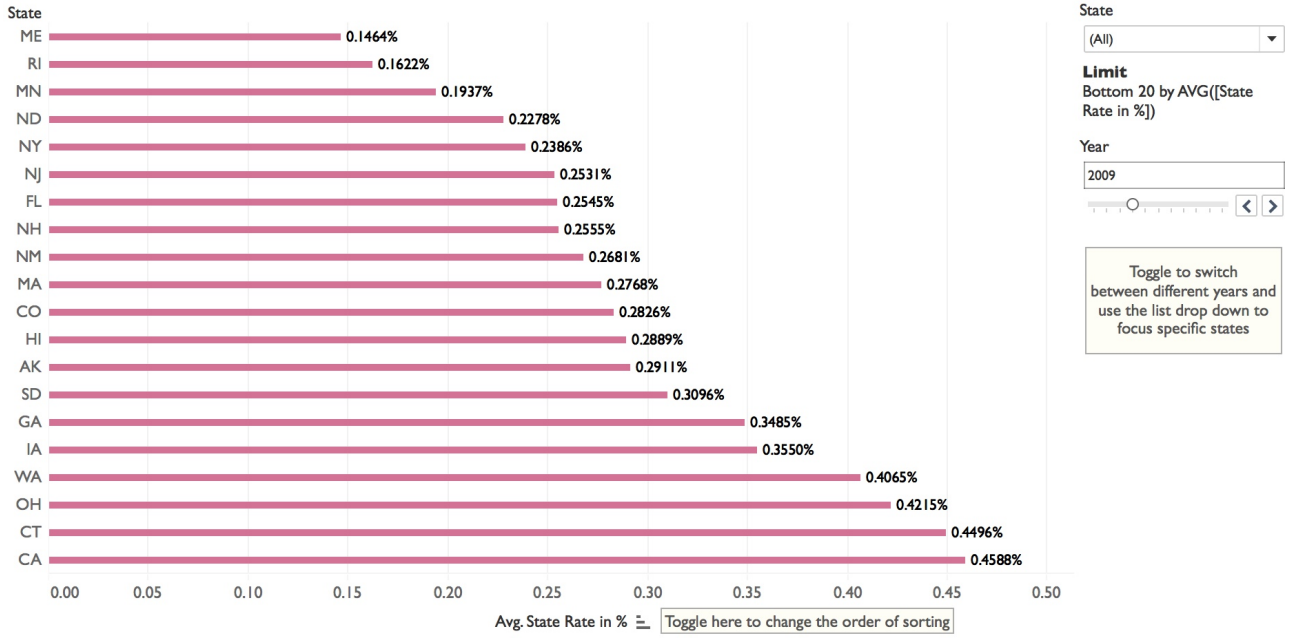

Figure 30: State Positivity % Ranking: Top 20 States with Least Severe Opioid Concentration - 2009

In the year 2009, `WV` - West Virginia, `NV` - Nevada and `MT` - Montana have been the states with the highest Opioid Severity. `ME` - Maine, `RI` - Rhode Island and `MN` - Minnesota have had the least Opioid Severity (using State Positivity % as the severity metric).

3.1.5 *2010*

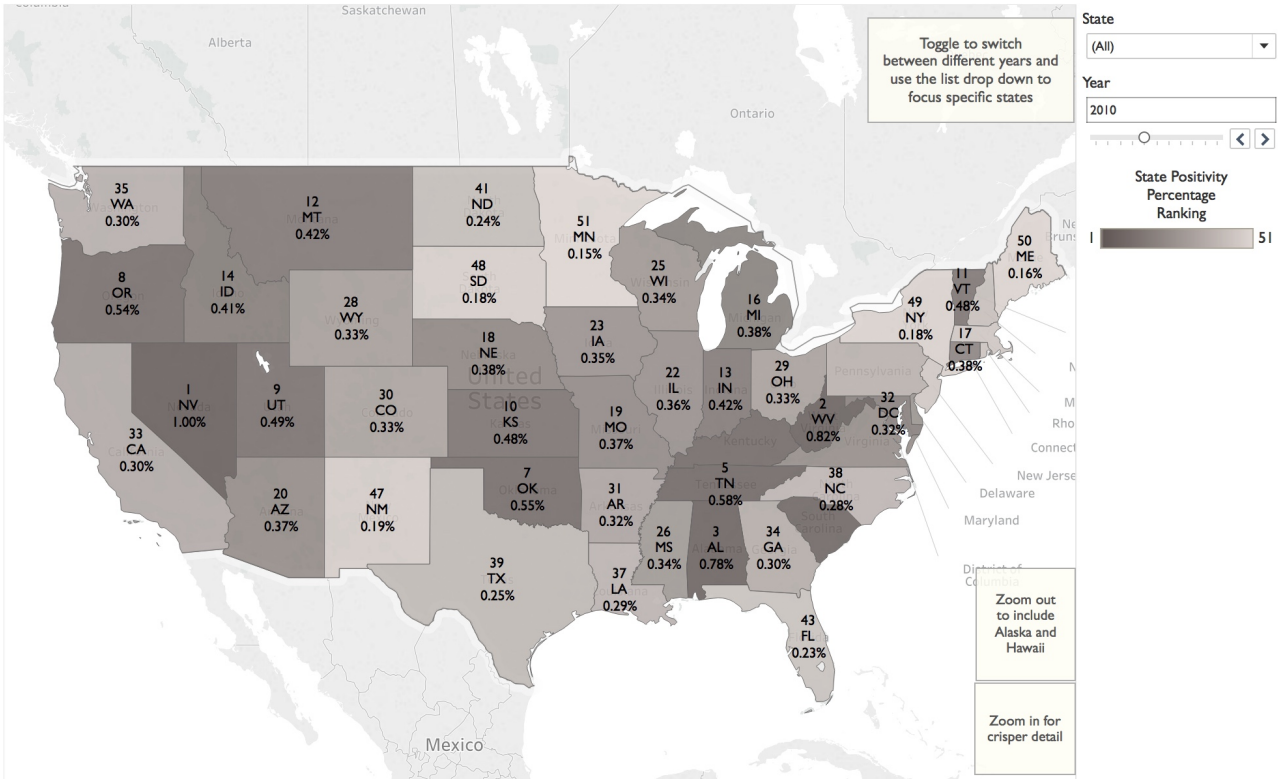

Figure 31: State Positivity % Ranking - 2010: Heat Map

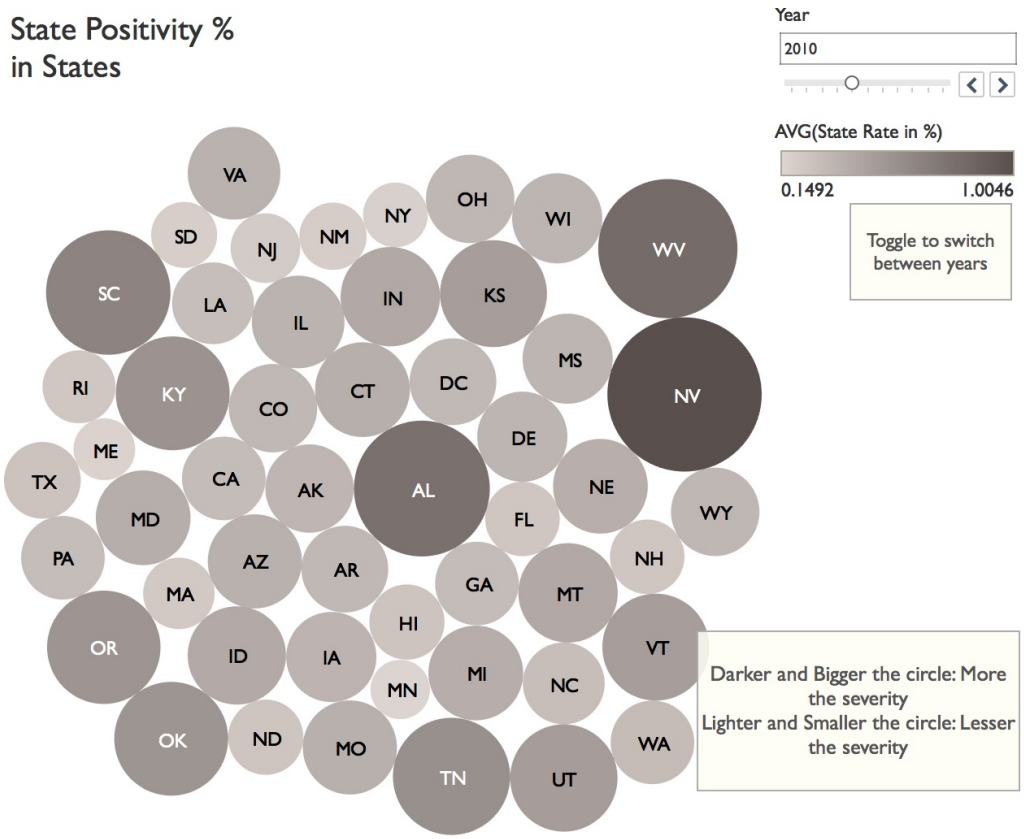

Figure 32: State Positivity % Ranking - 2010: Packed Bubble Chart

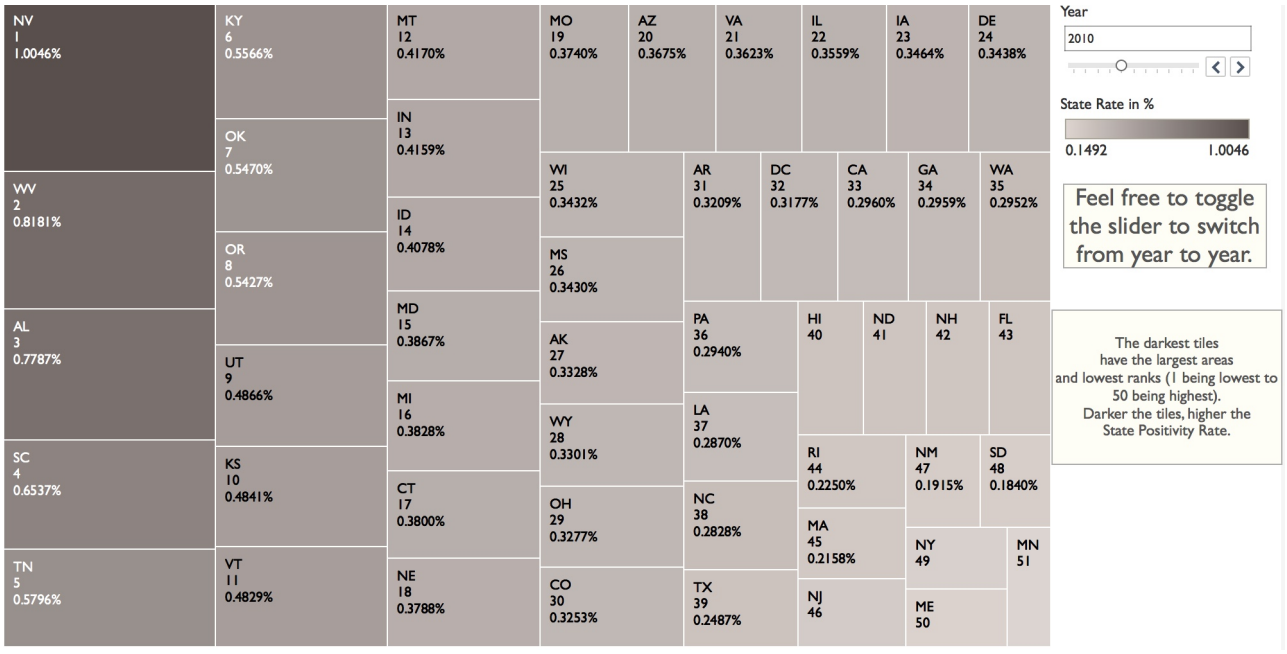

Figure 33: State Positivity % Ranking - 2010: Tree Map

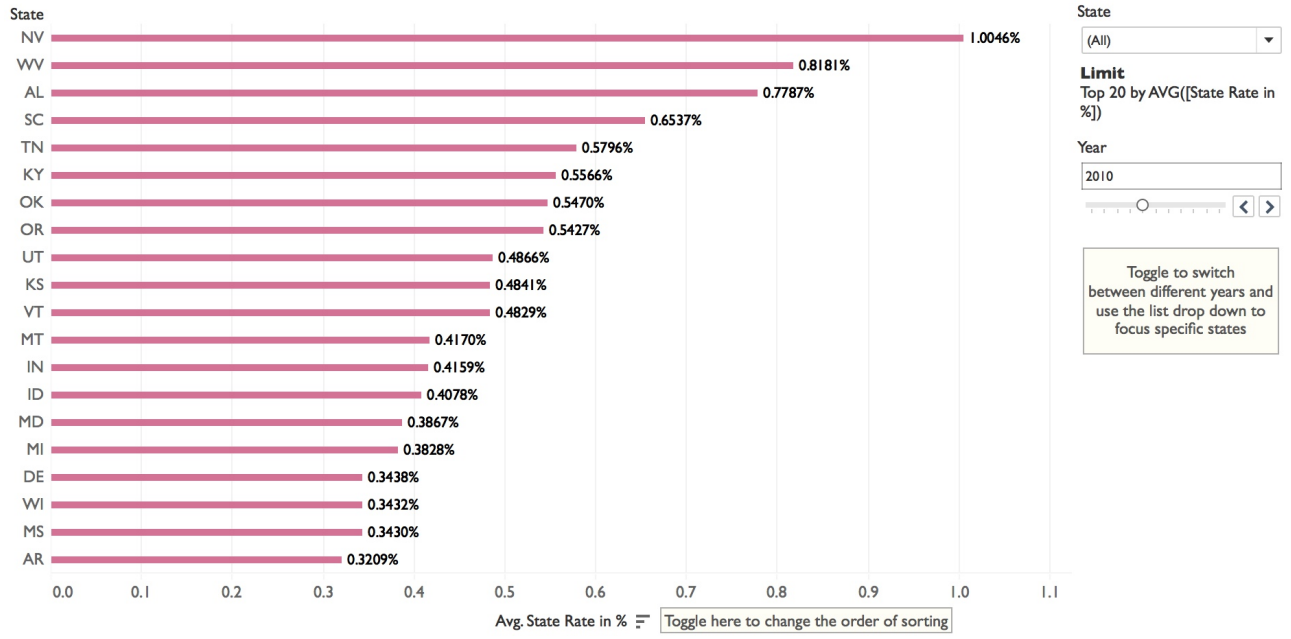

Figure 34: State Positivity % Ranking: Top 20 States with Most Severe Opioid Concentration - 2010

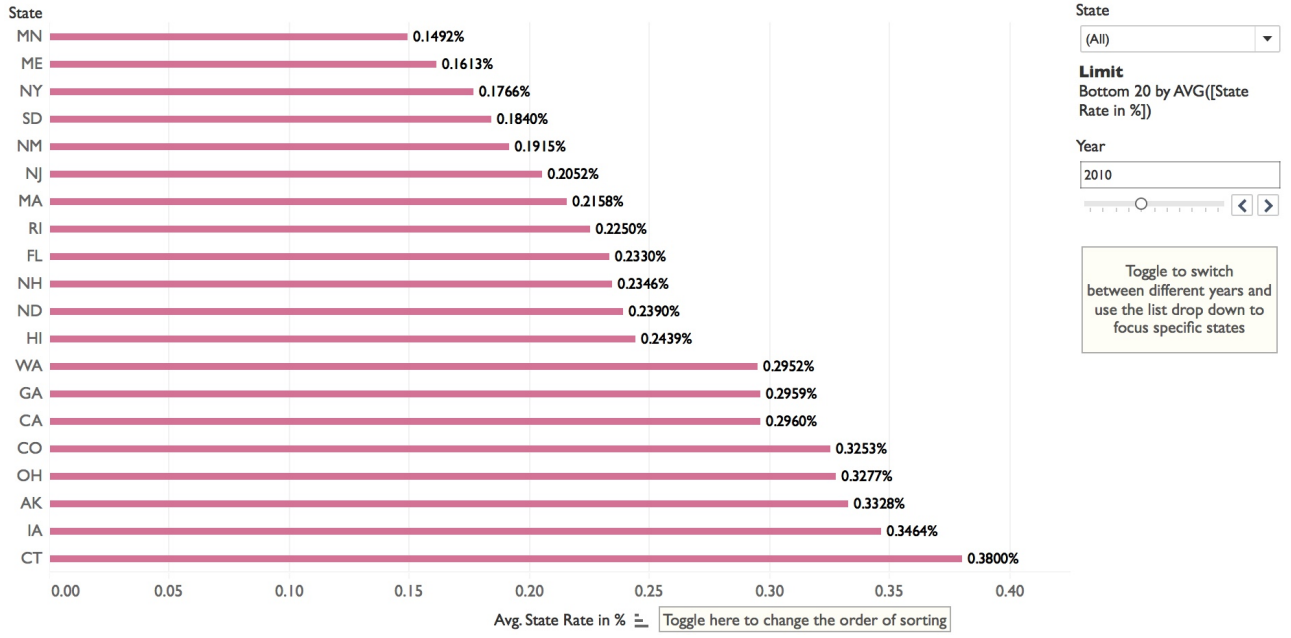

Figure 35: State Positivity % Ranking: Top 20 States with Least Severe Opioid Concentration - 2010

In the year 2010, `WV` - West Virginia, `NV` - Nevada and `AL` - Alabama have been the states with the highest Opioid Severity. `ME` - Maine, `NY` - New York and `MN` - Minnesota have had the least Opioid Severity (using State Positivity % as the severity metric).

3.1.6 *2011*

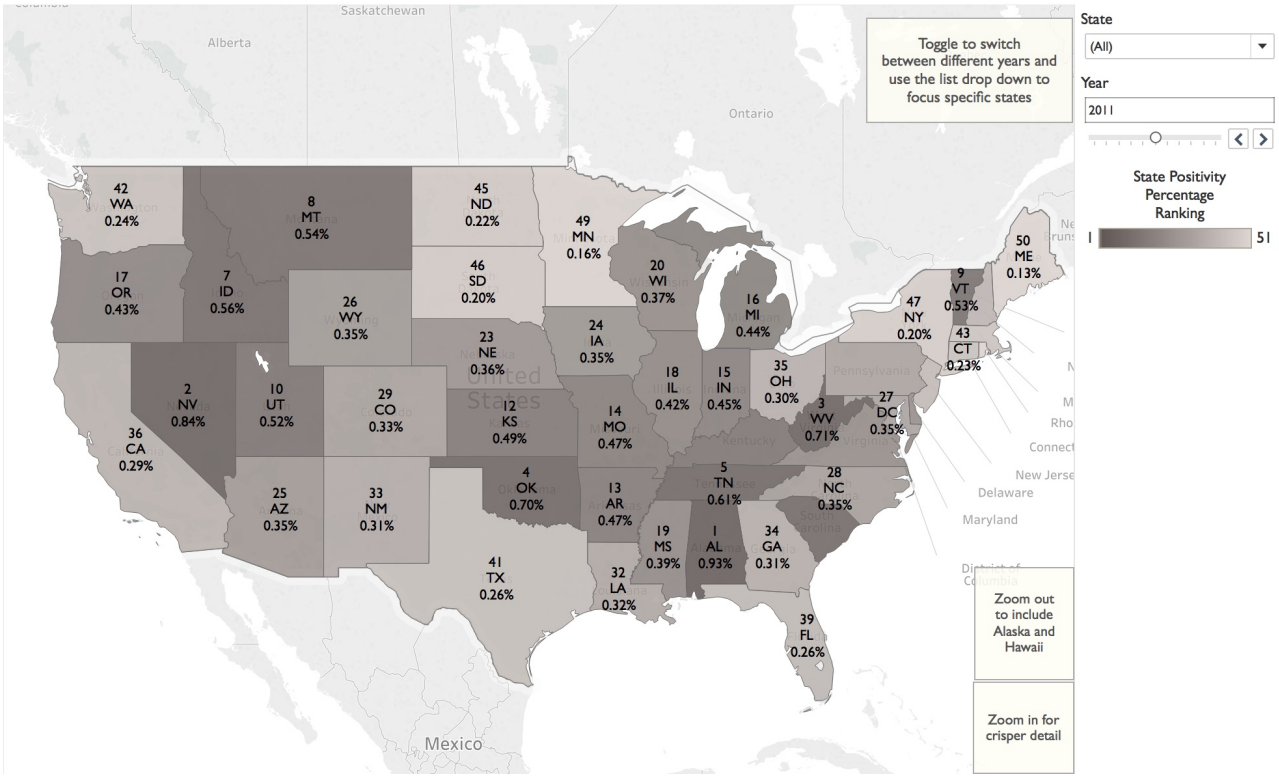

Figure 36: State Positivity % Ranking - 2011: Heat Map

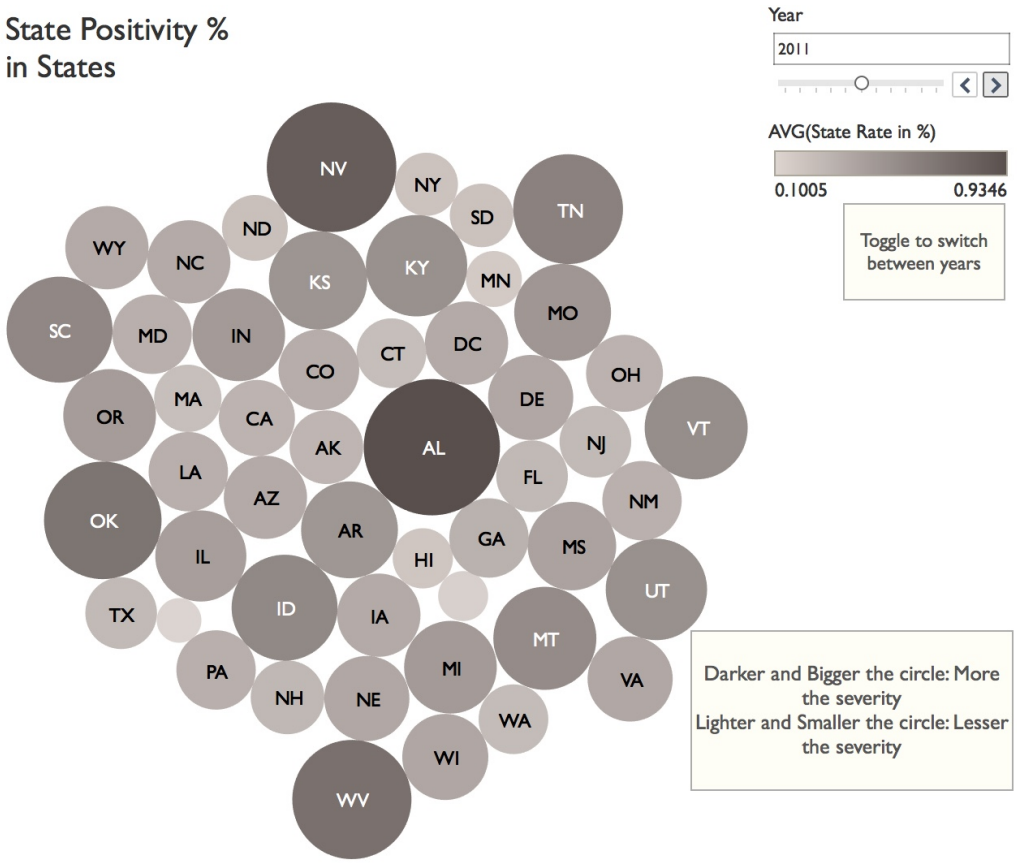

Figure 37: State Positivity % Ranking - 2011: Packed Bubble Chart

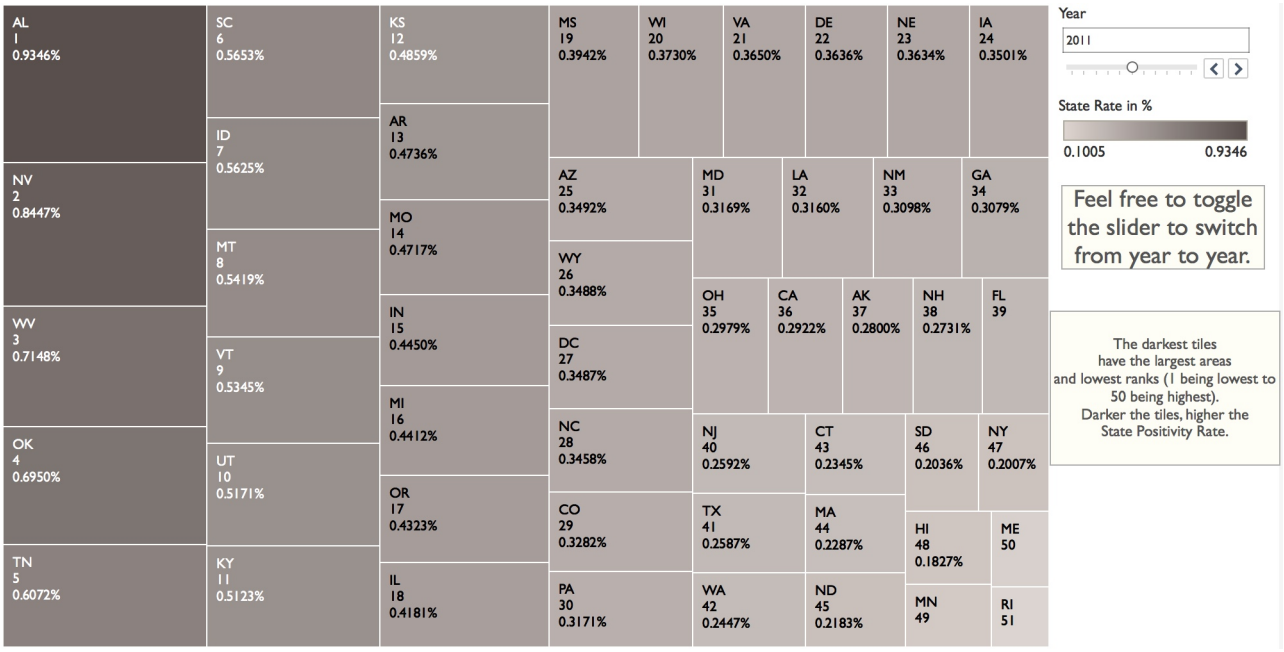

Figure 38: State Positivity % Ranking - 2011: Tree Map

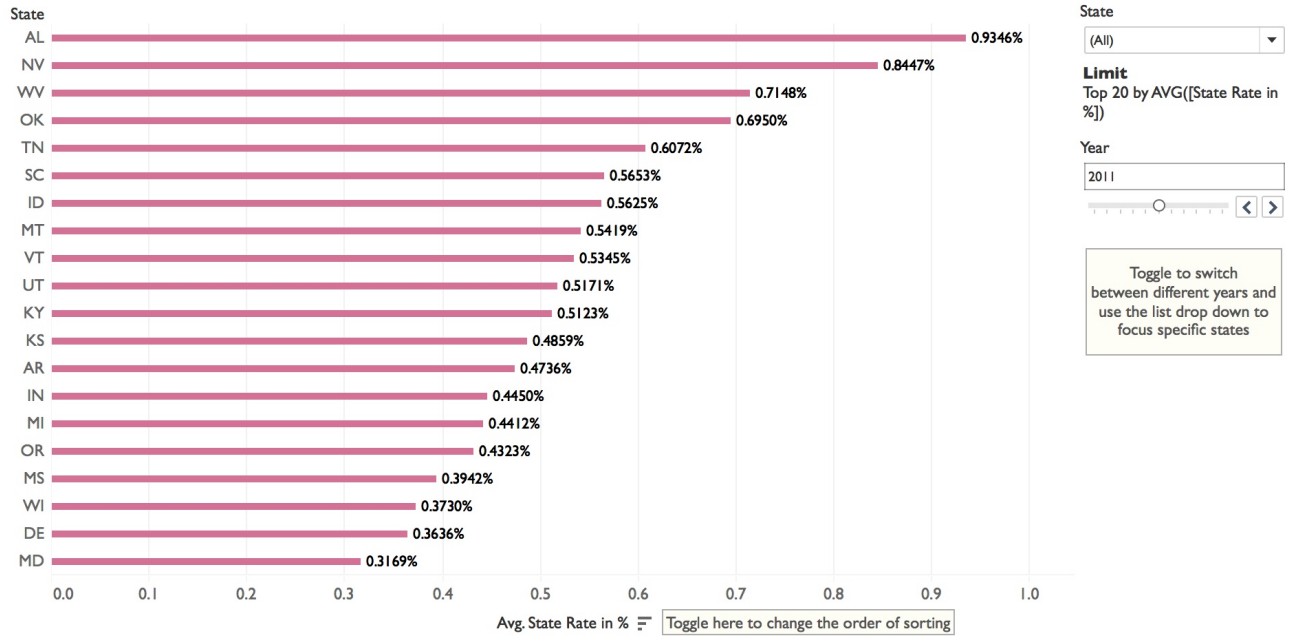

Figure 39: State Positivity % Ranking: Top 20 States with Most Severe Opioid Concentration - 2011

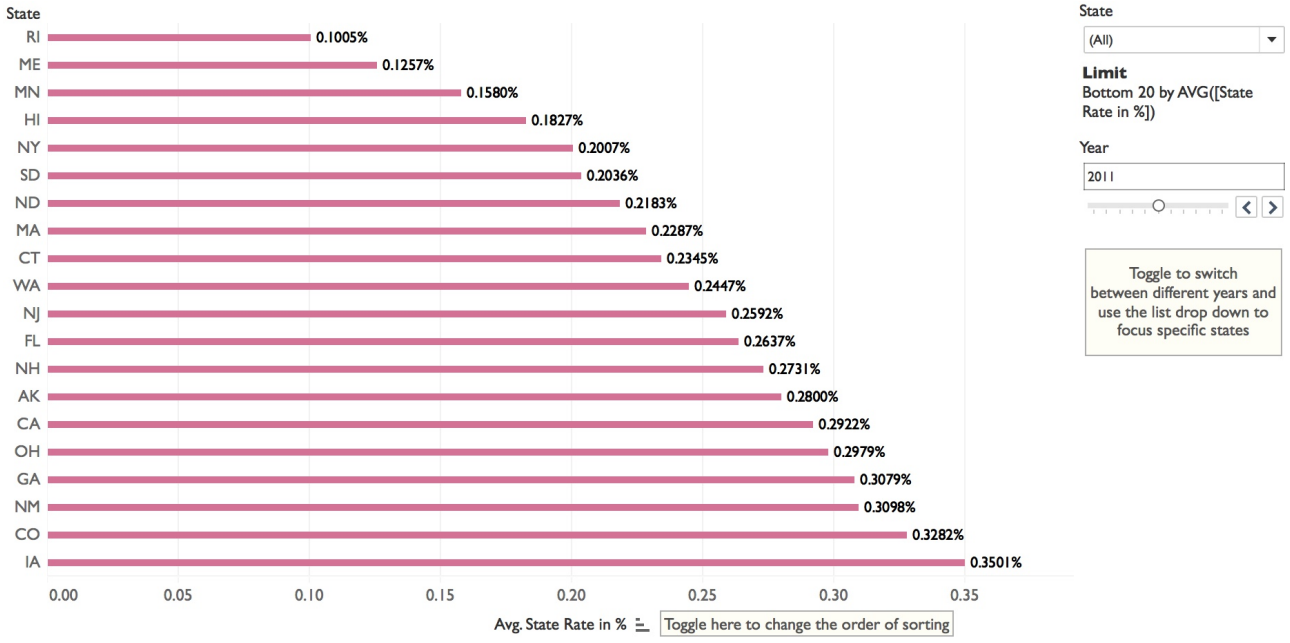

Figure 40: State Positivity % Ranking: Top 20 States with Least Severe Opioid Concentration - 2011

In the year 2011, `WV` - West Virginia, `NV` - Nevada and `AL` - Alabama have been the states with the highest Opioid Severity. `ME` - Maine, `RI` - Rhode Island and `MN` - Minnesota have had the least Opioid Severity (using State Positivity % as the severity metric).

3.1.7 *2012*

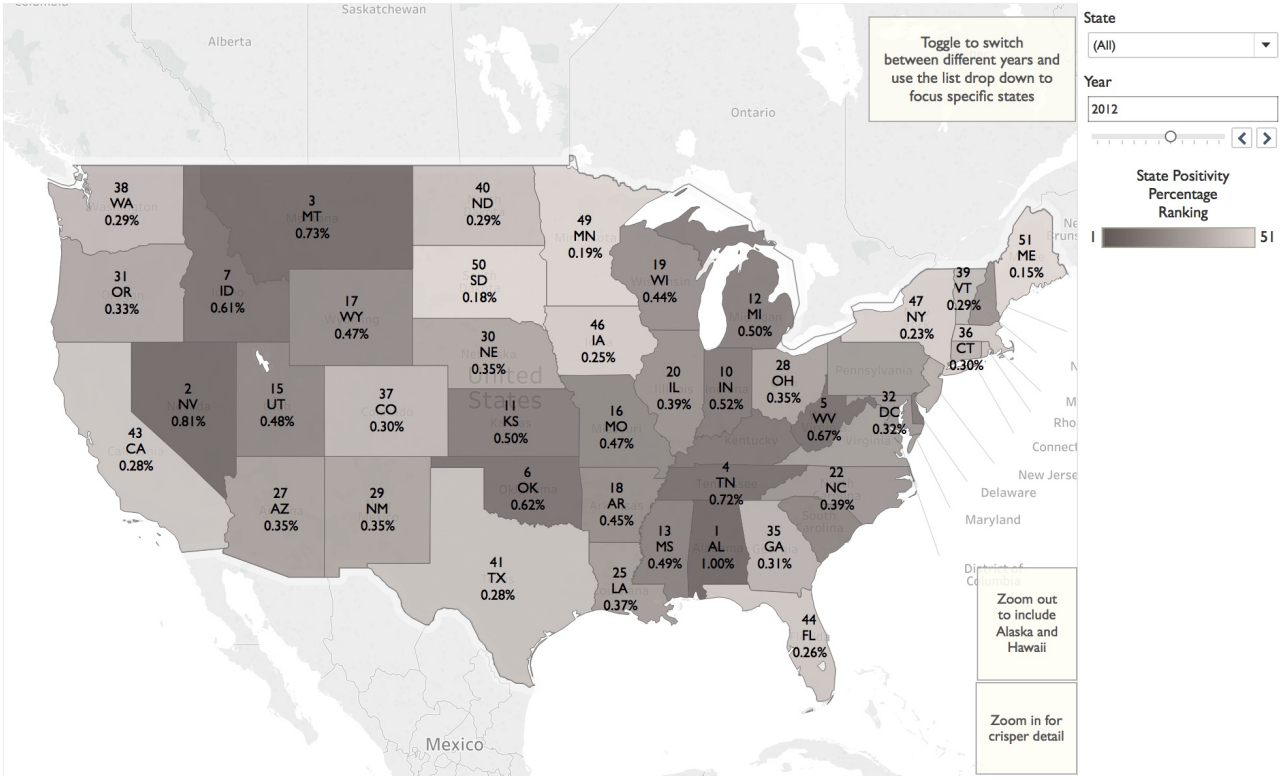

Figure 41: State Positivity % Ranking - 2012: Heat Map

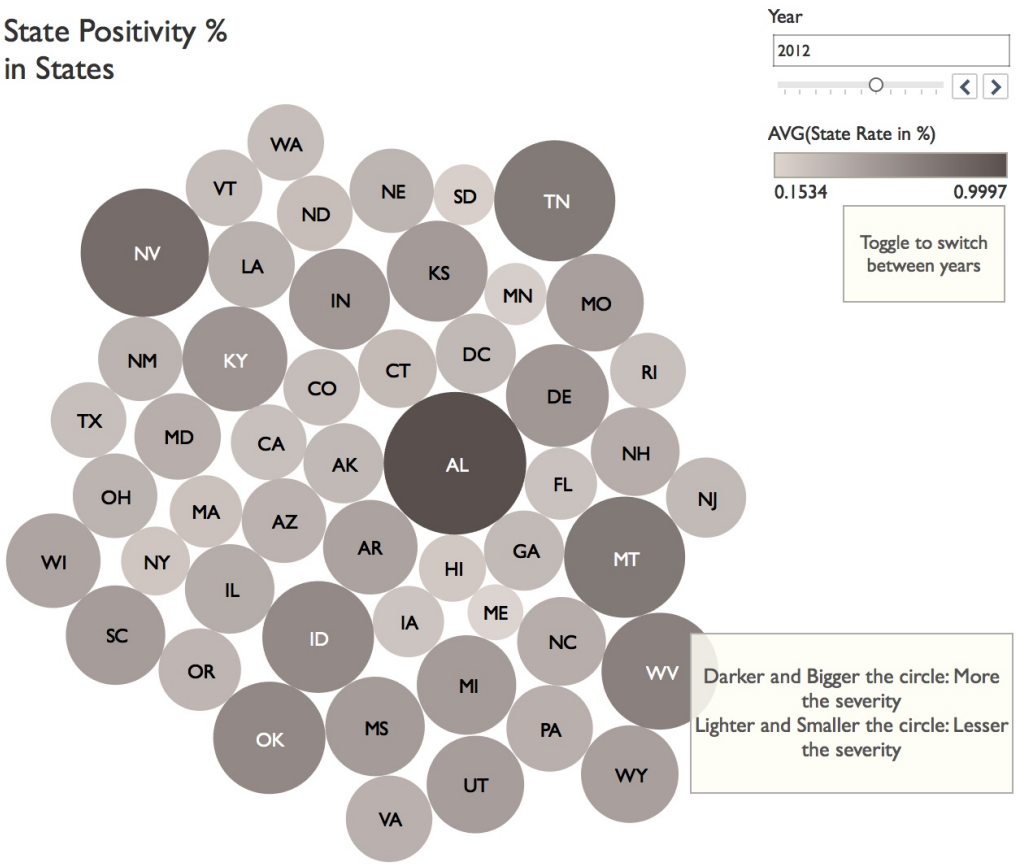

Figure 42: State Positivity % Ranking - 2012: Packed Bubble Chart

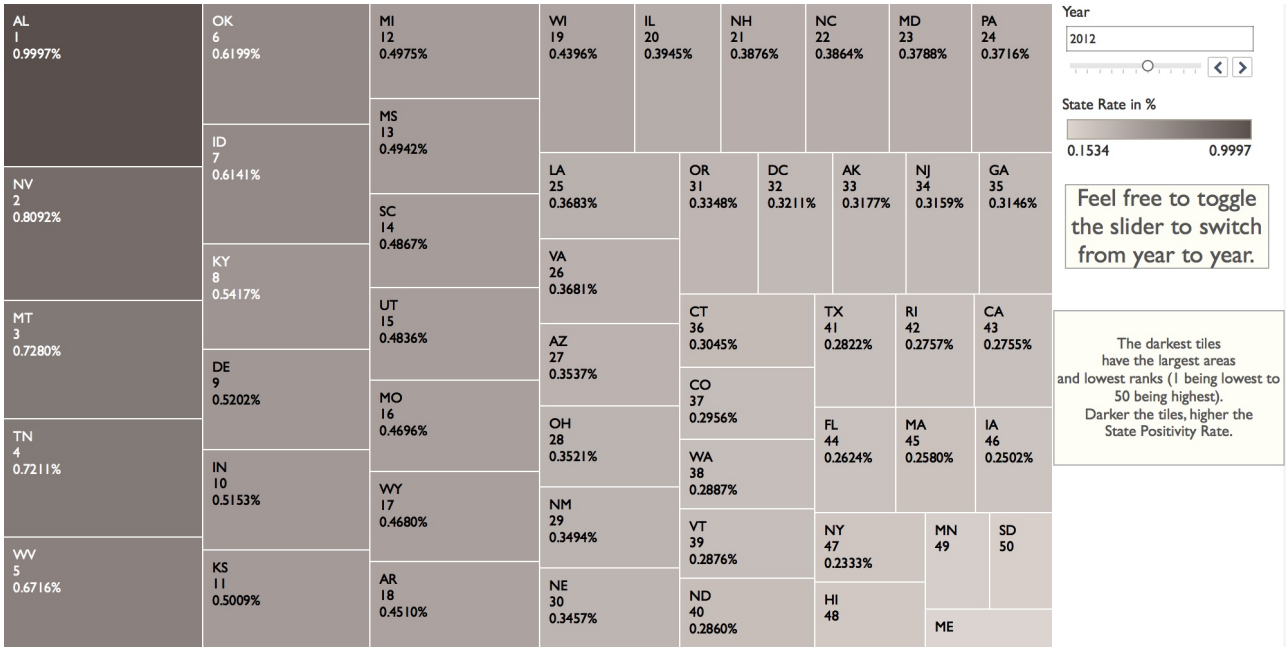

Figure 43: State Positivity % Ranking - 2012: Tree Map

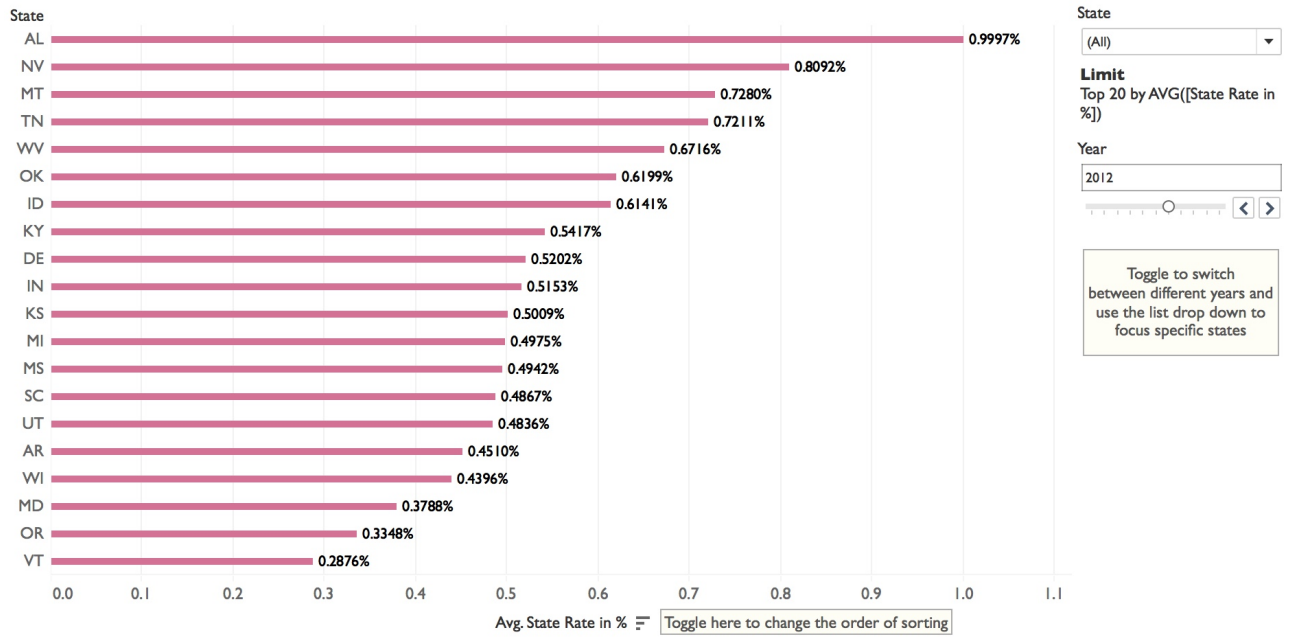

Figure 44: State Positivity % Ranking: Top 20 States with Most Severe Opioid Concentration - 2012

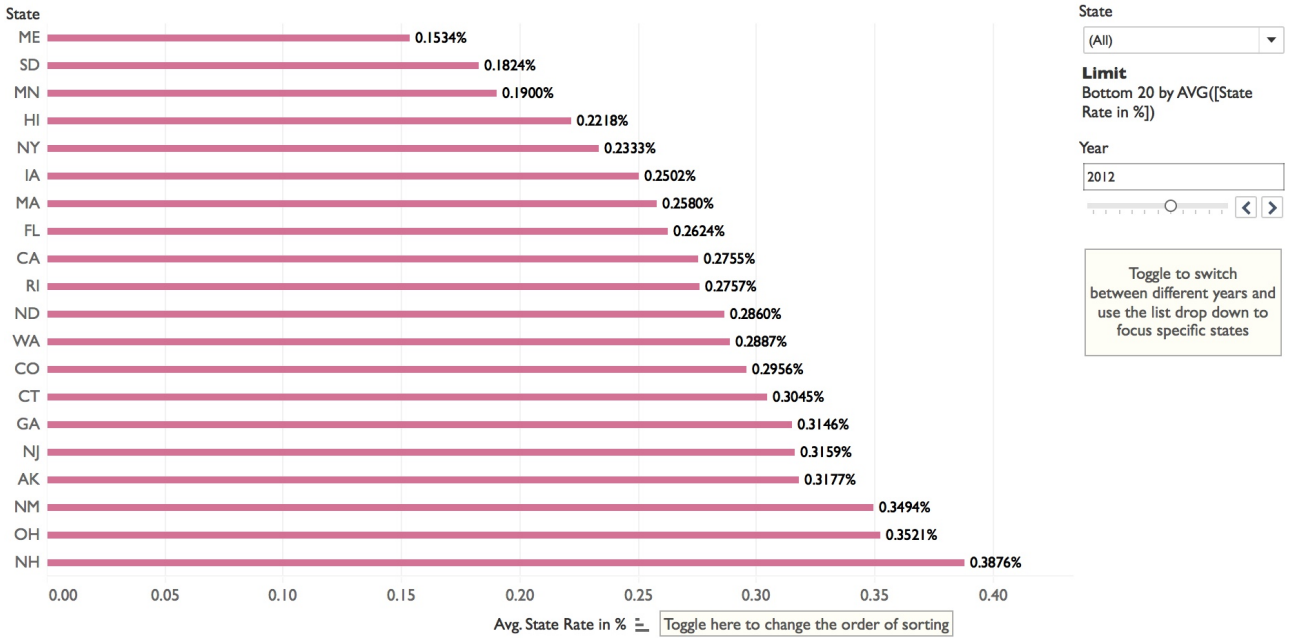

Figure 45: State Positivity % Ranking: Top 20 States with Least Severe Opioid Concentration - 2012

In the year 2012, `AL` - Alabama, `NV` - Nevada and `MT` - Montana have been the states with the highest Opioid Severity. `ME` - Maine, `SD` - South Dakota and `MN` - Minnesota have had the least Opioid Severity (using State Positivity % as the severity metric).

3.1.8 *2013*

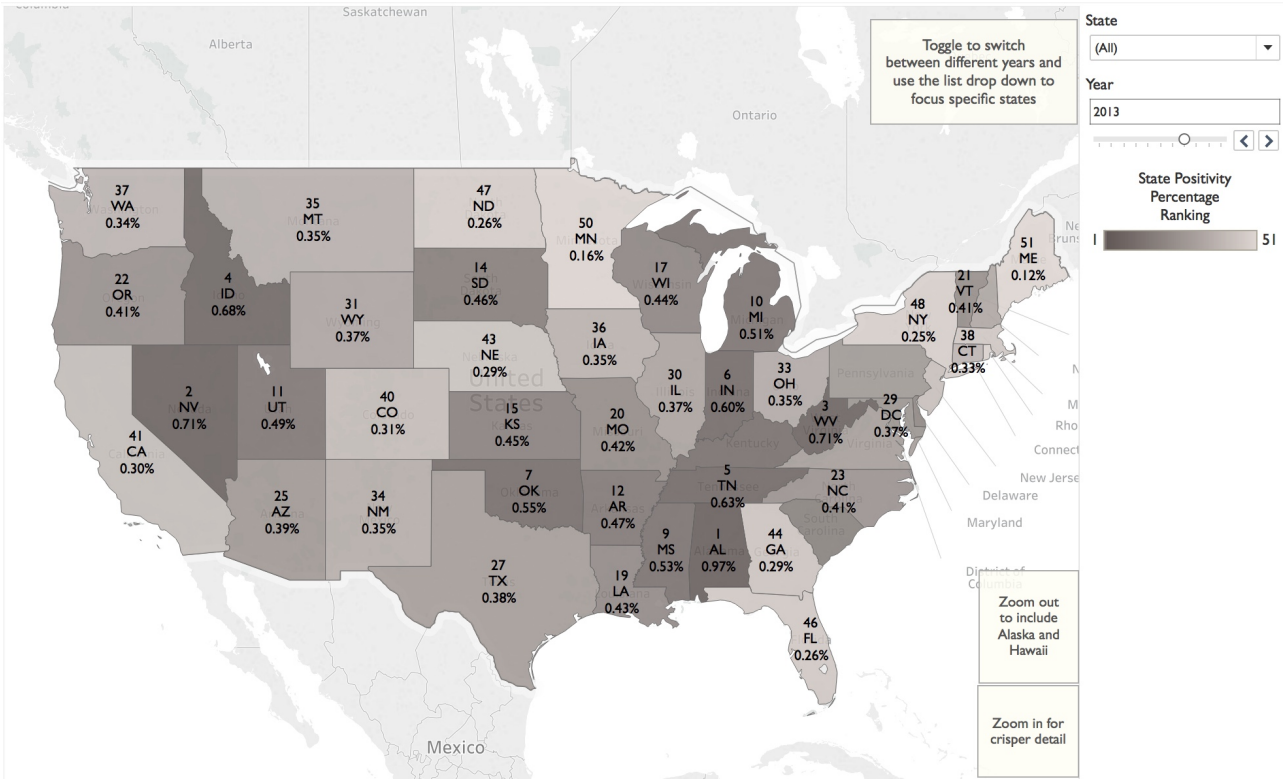

Figure 46: State Positivity % Ranking - 2013: Heat Map

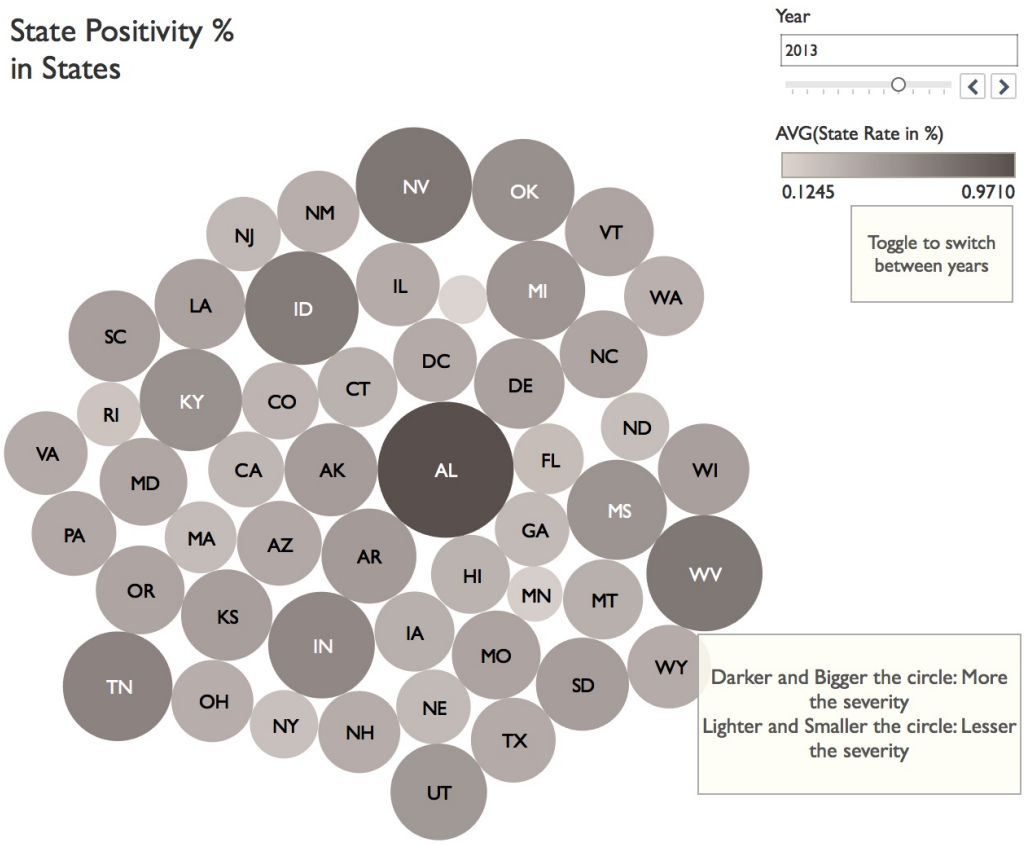

Figure 47: State Positivity % Ranking - 2013: Packed Bubble Chart

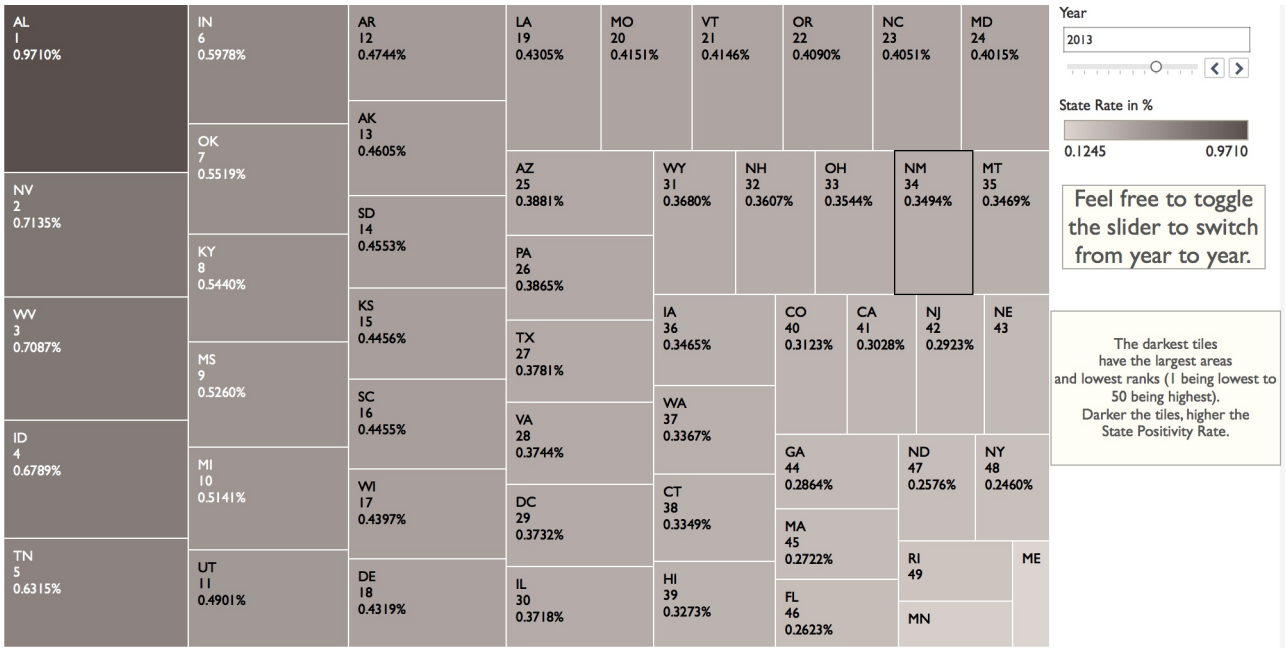

Figure 48: State Positivity % Ranking - 2013: Tree Map

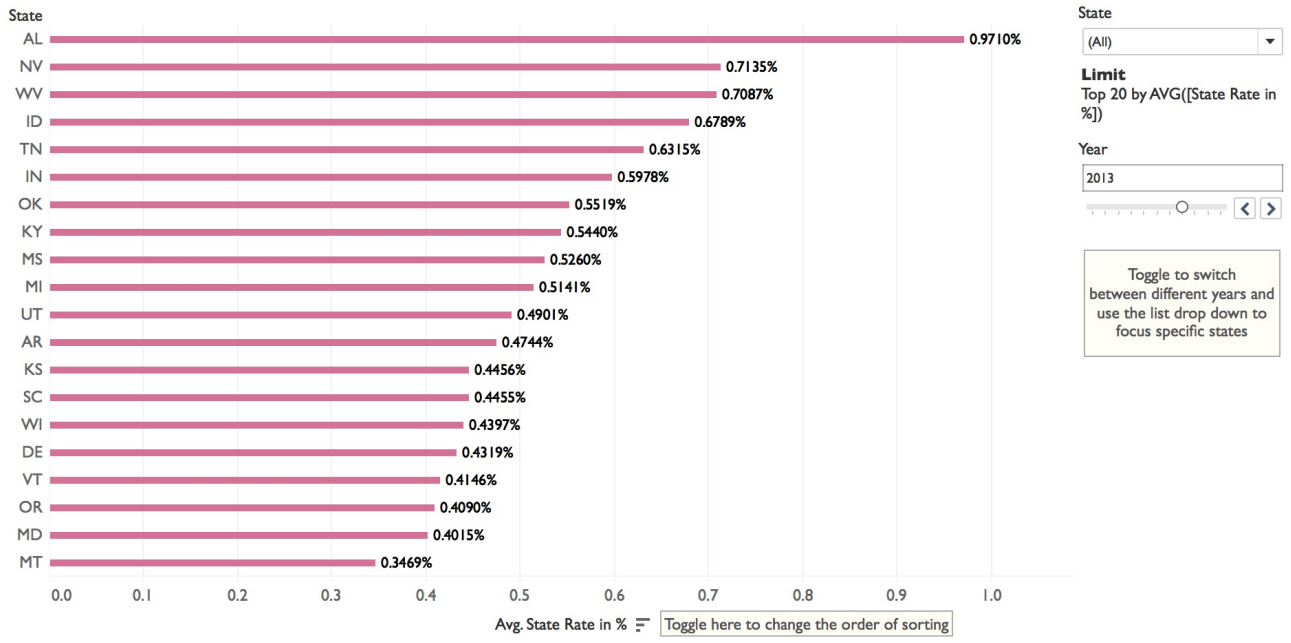

Figure 49: State Positivity % Ranking: Top 20 States with Most Severe Opioid Concentration - 2013

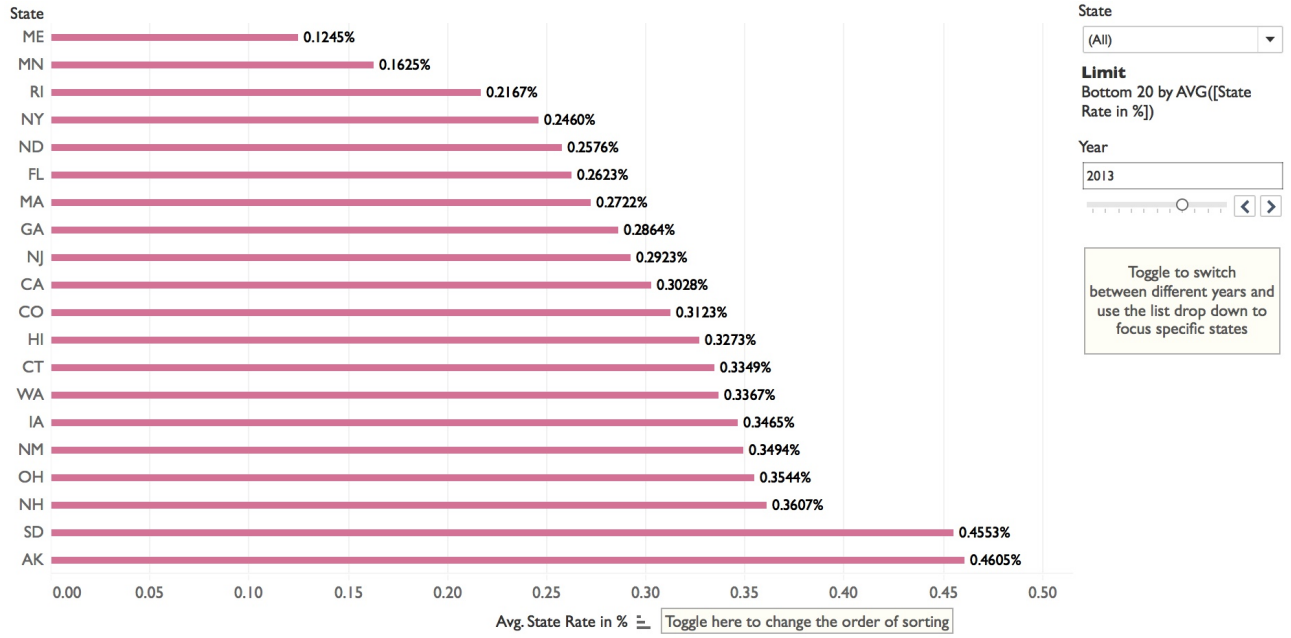

Figure 50: State Positivity % Ranking: Top 20 States with Least Severe Opioid Concentration - 2013

In the year 2013, `AL` - Alabama, `NV` - Nevada and `WV` - West Virginia have been the states with the highest Opioid Severity. `ME` - Maine, `RI` - Rhode Island and `MN` - Minnesota have had the least Opioid Severity (using State Positivity % as the severity metric).

3.1.9 *2014*

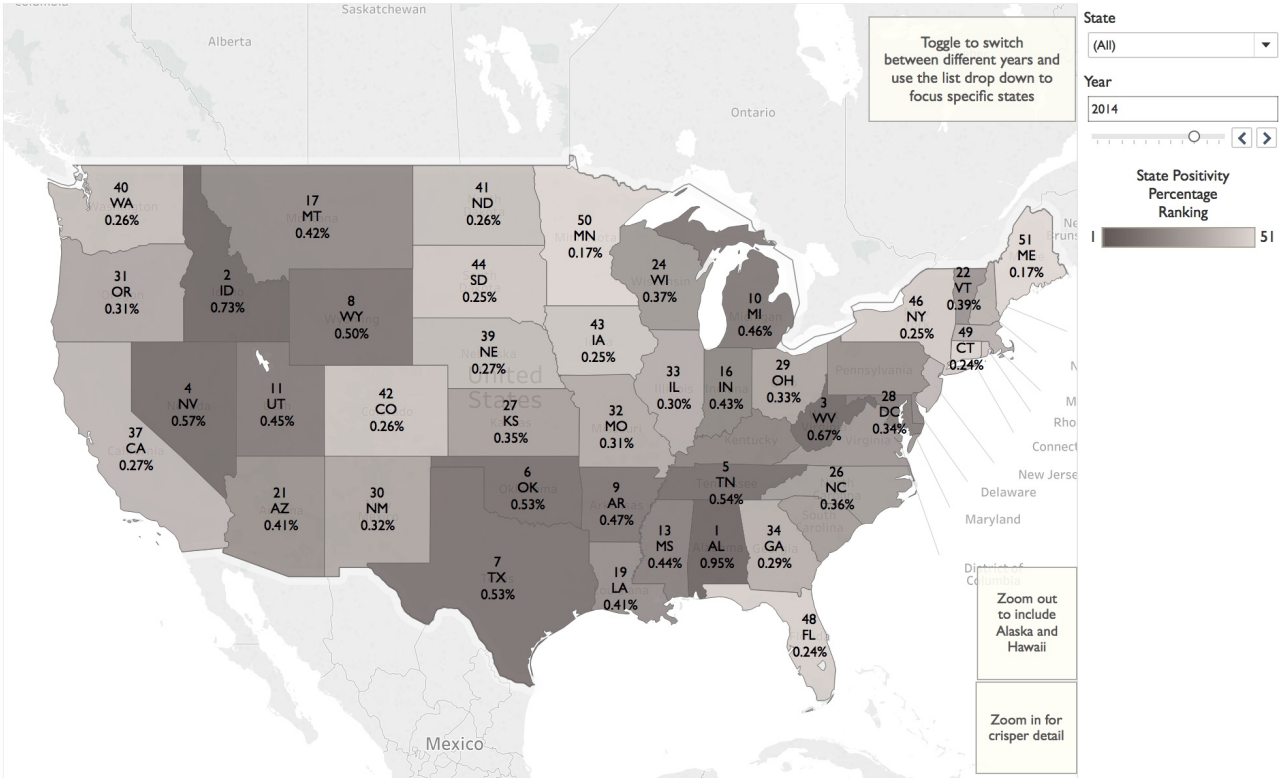

Figure 51: State Positivity % Ranking - 2014: Heat Map

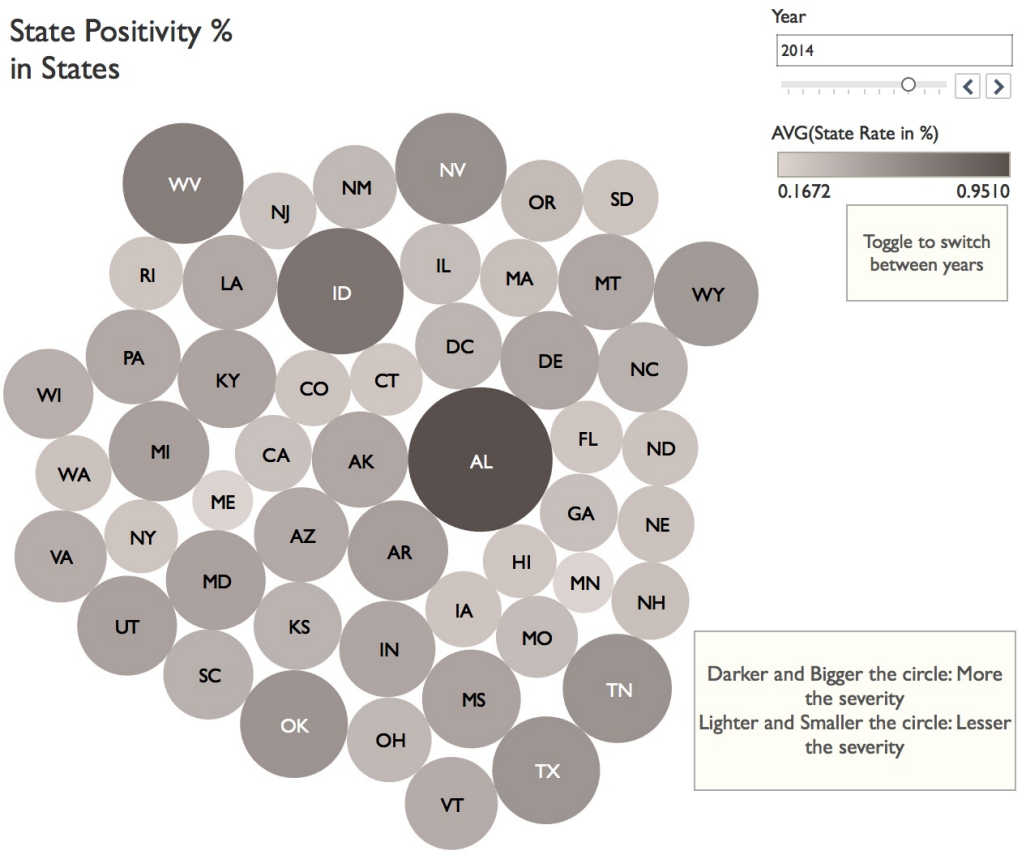

Figure 52: State Positivity % Ranking - 2014: Packed Bubble Chart

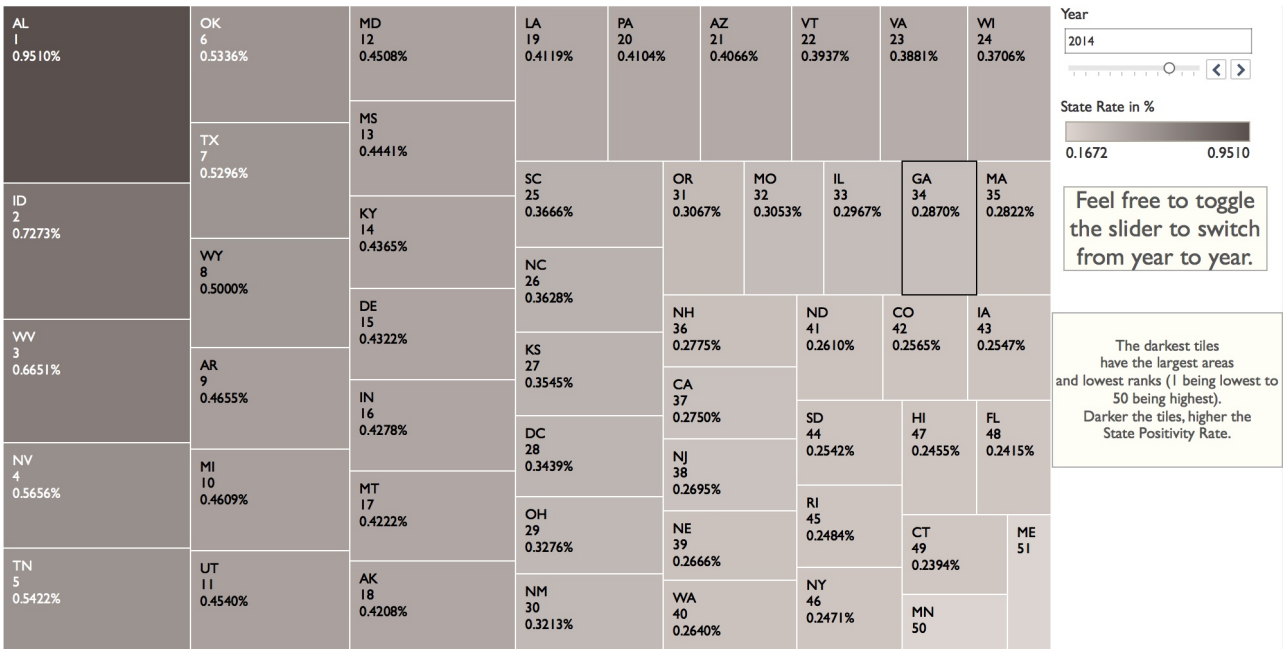

Figure 53: State Positivity % Ranking - 2014: Tree Map

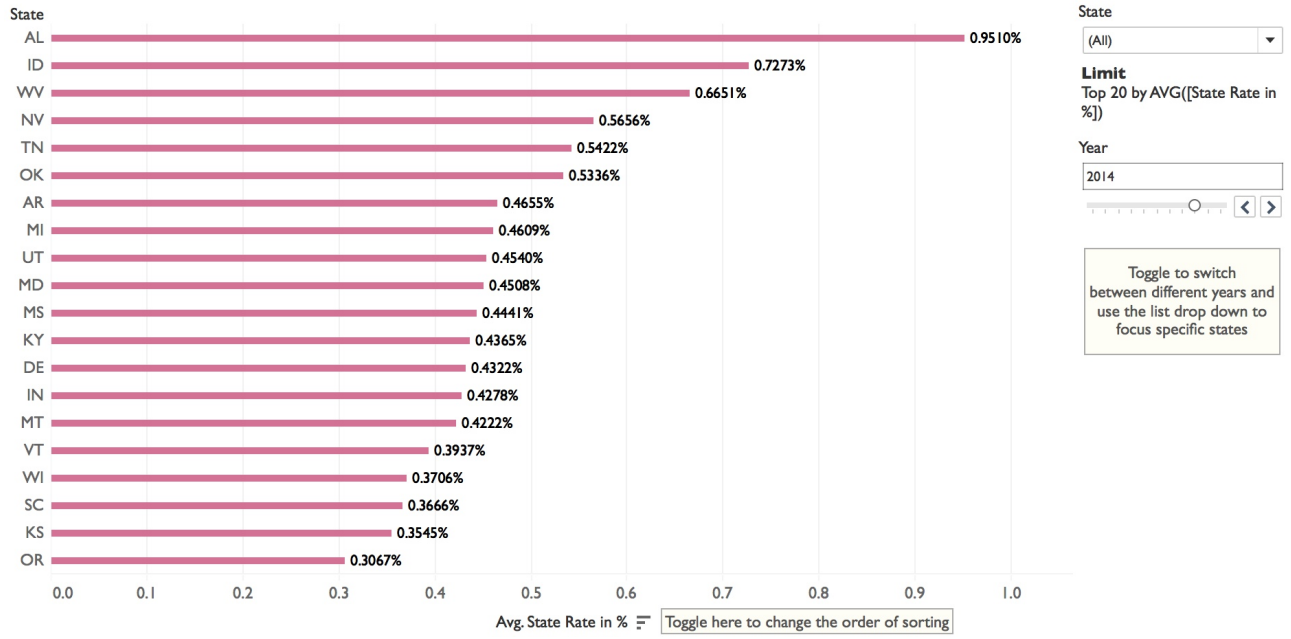

Figure 54: State Positivity % Ranking: Top 20 States with Most Severe Opioid Concentration - 2014

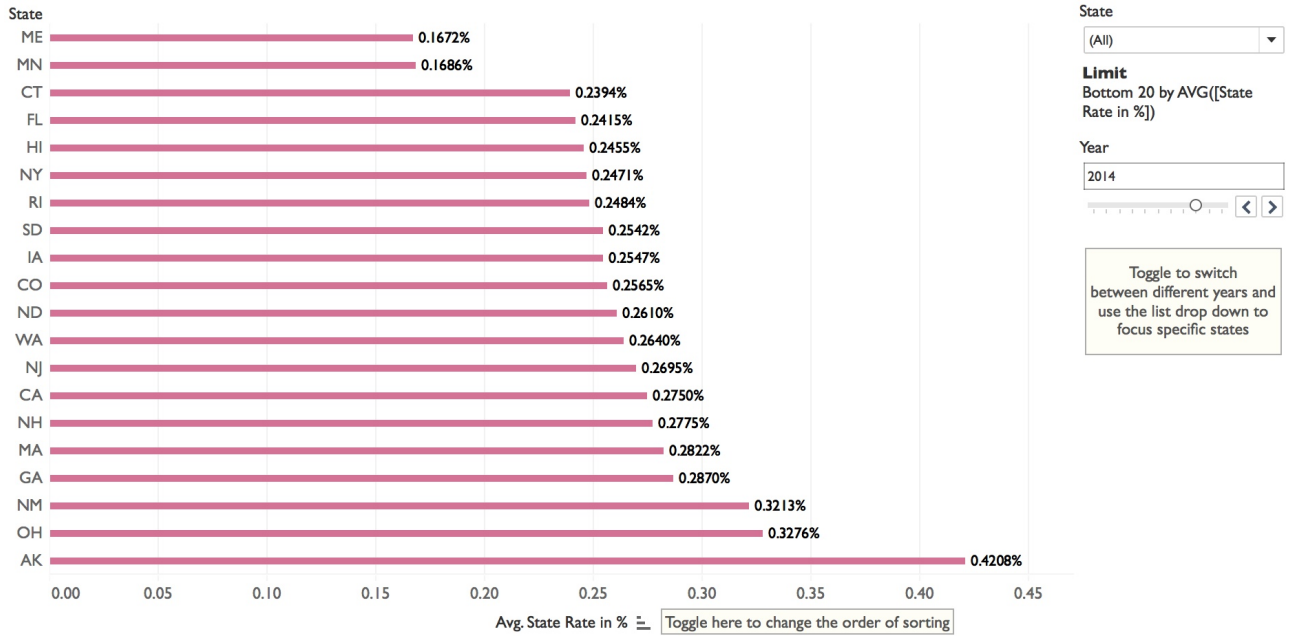

Figure 55: State Positivity % Ranking: Top 20 States with Least Severe Opioid Concentration - 2014

In the year 2014, `AL` - Alabama, `ID` - Idaho and `WV` - West Virginia have been the states with the highest Opioid Severity. `ME` - Maine, `CT` - Connecticut and `MN` - Minnesota have had the least Opioid Severity (using State Positivity % as the severity metric).

3.1.10 *2015*

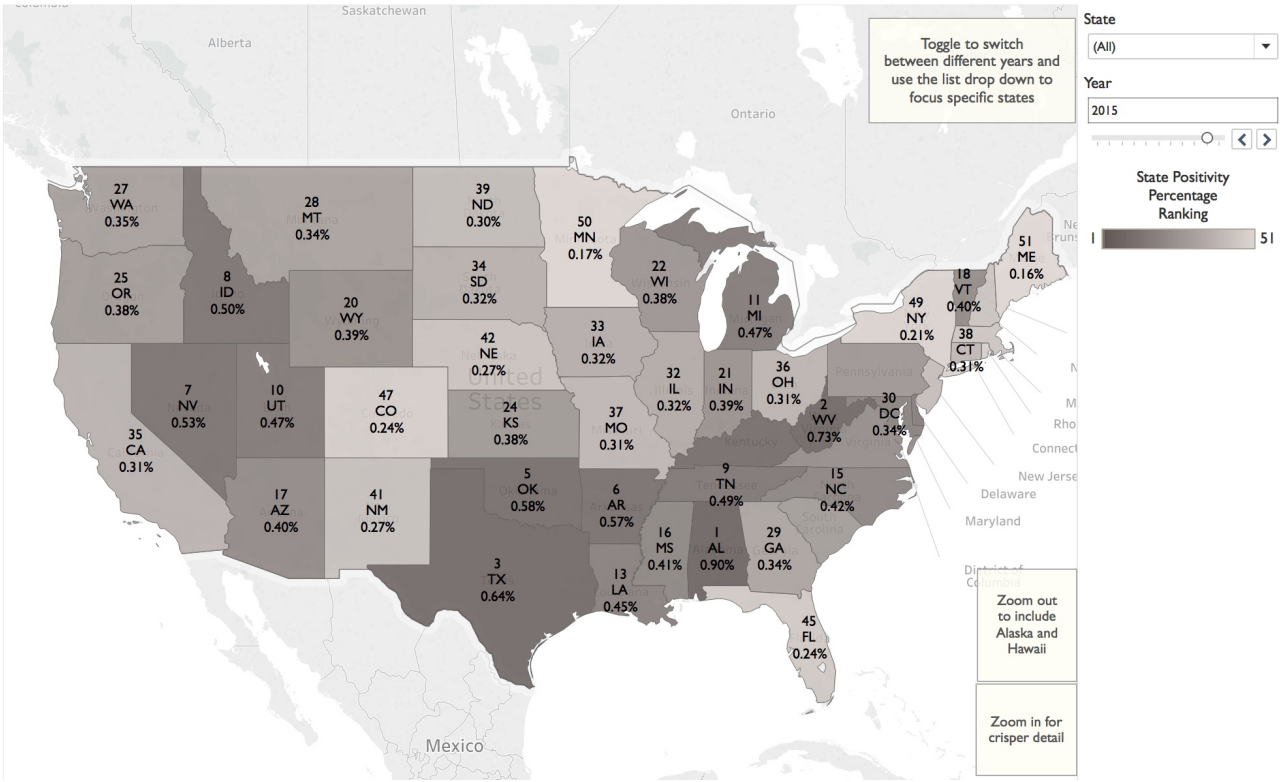

Figure 56: State Positivity % Ranking - 2015: Heat Map

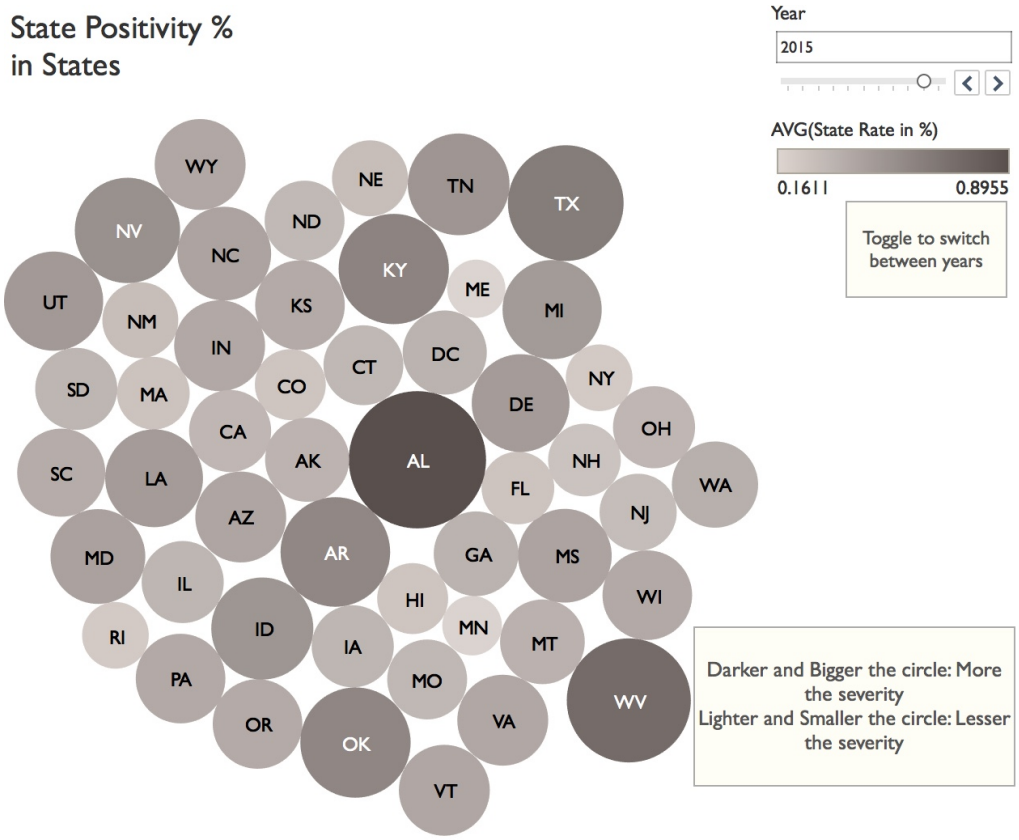

Figure 57: State Positivity % Ranking - 2015: Packed Bubble Chart

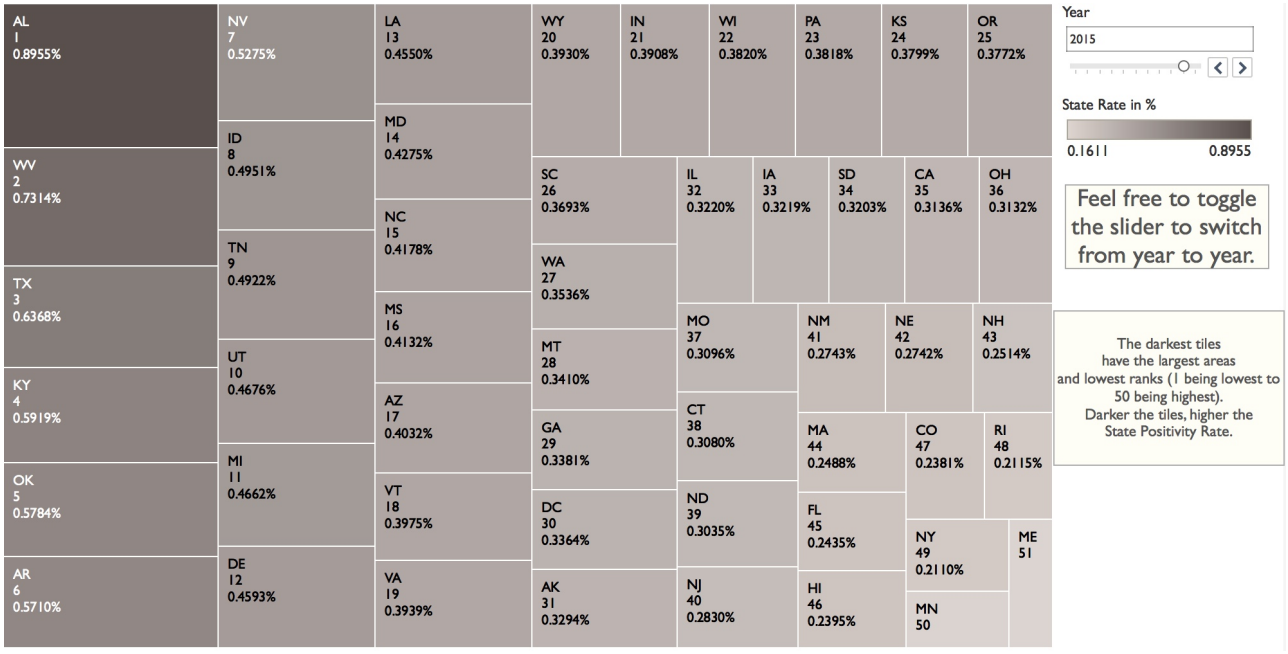

Figure 58: State Positivity % Ranking - 2015: Tree Map

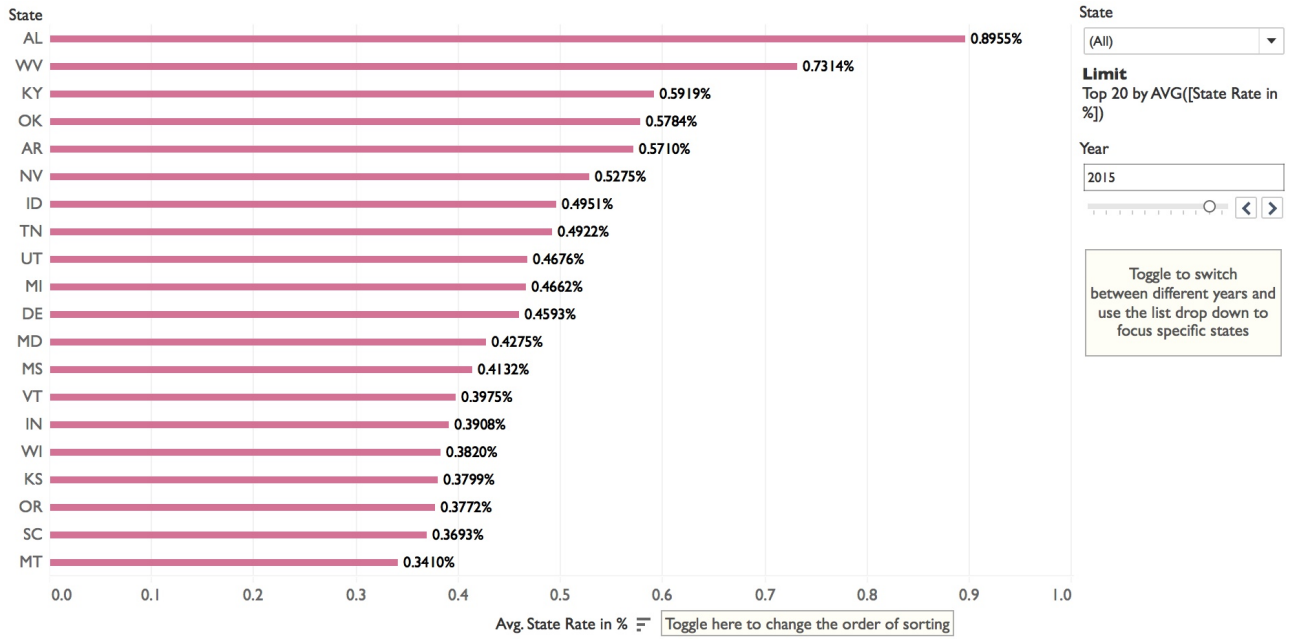

Figure 59: State Positivity % Ranking: Top 20 States with Most Severe Opioid Concentration - 2015

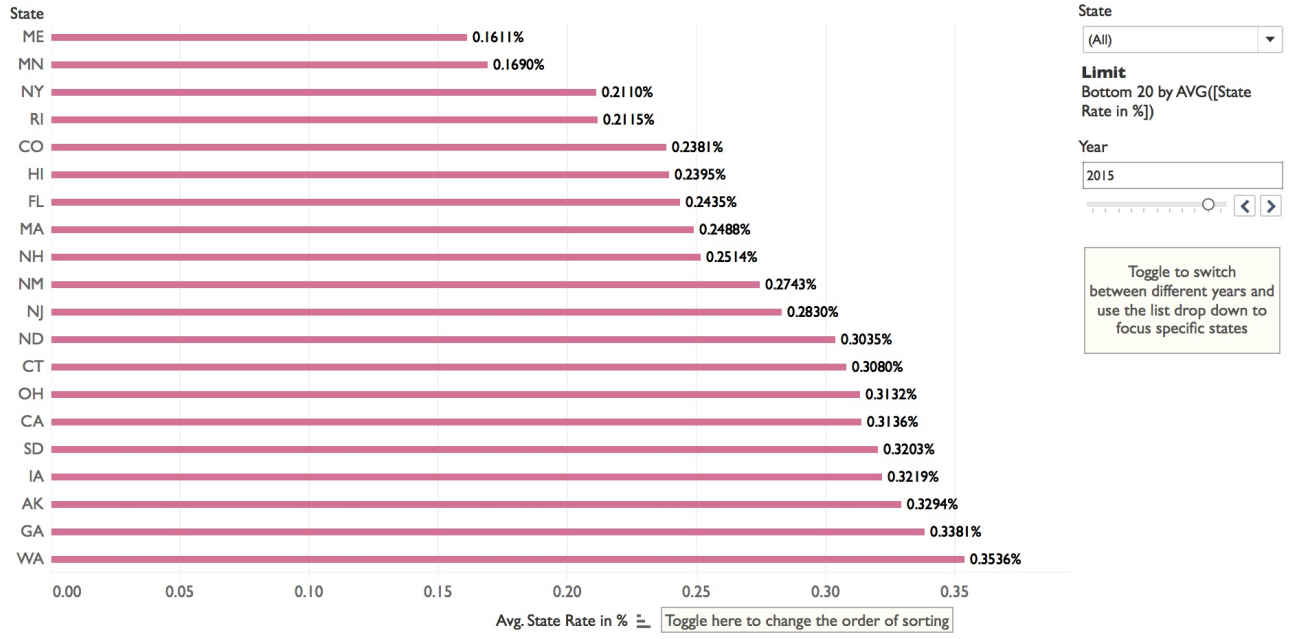

Figure 60: State Positivity % Ranking: Top 20 States with Least Severe Opioid Concentration - 2015

In the year 2015, `AL` - Alabama, `TX` - Texas and `WV` - West Virginia have been the states with the highest Opioid Severity. `ME` - Maine, `NY` - New York and `MN` - Minnesota have had the least Opioid Severity (using State Positivity % as the severity metric).

3.1.11 *2016*

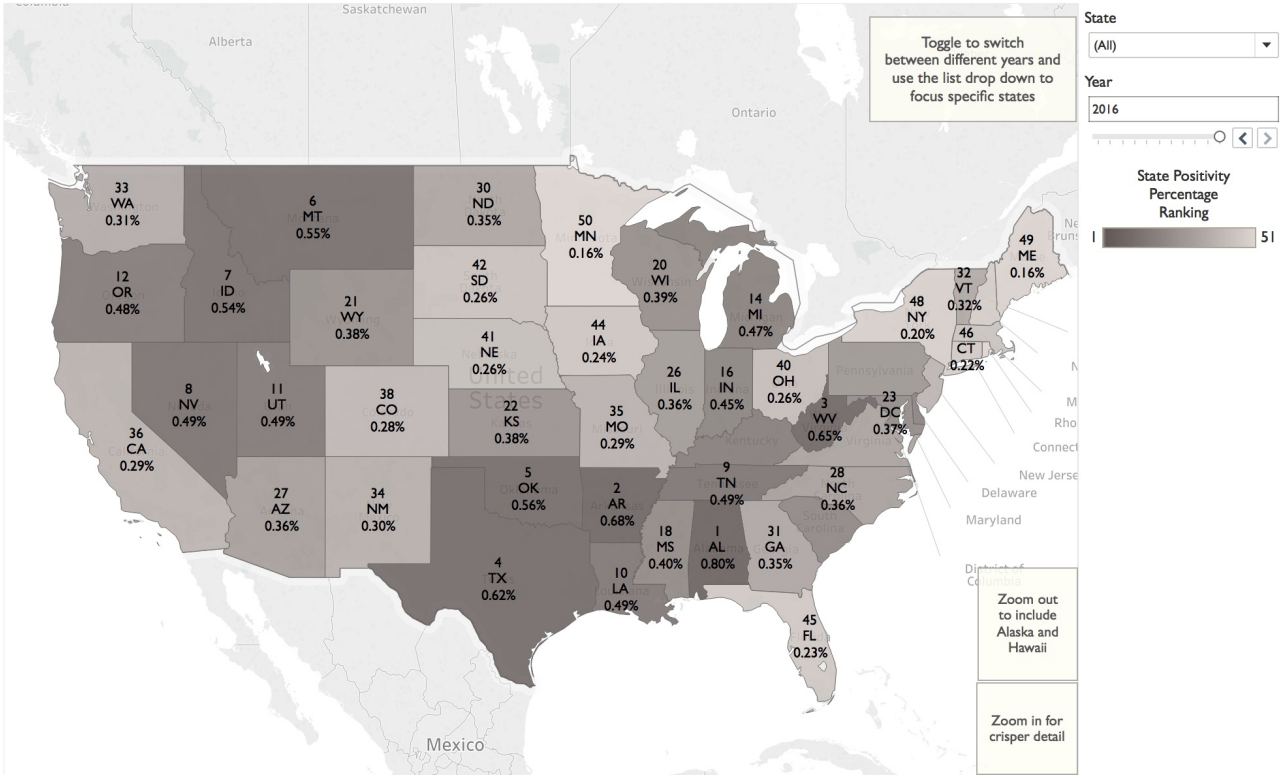

Figure 61: State Positivity % Ranking - 2016: Heat Map

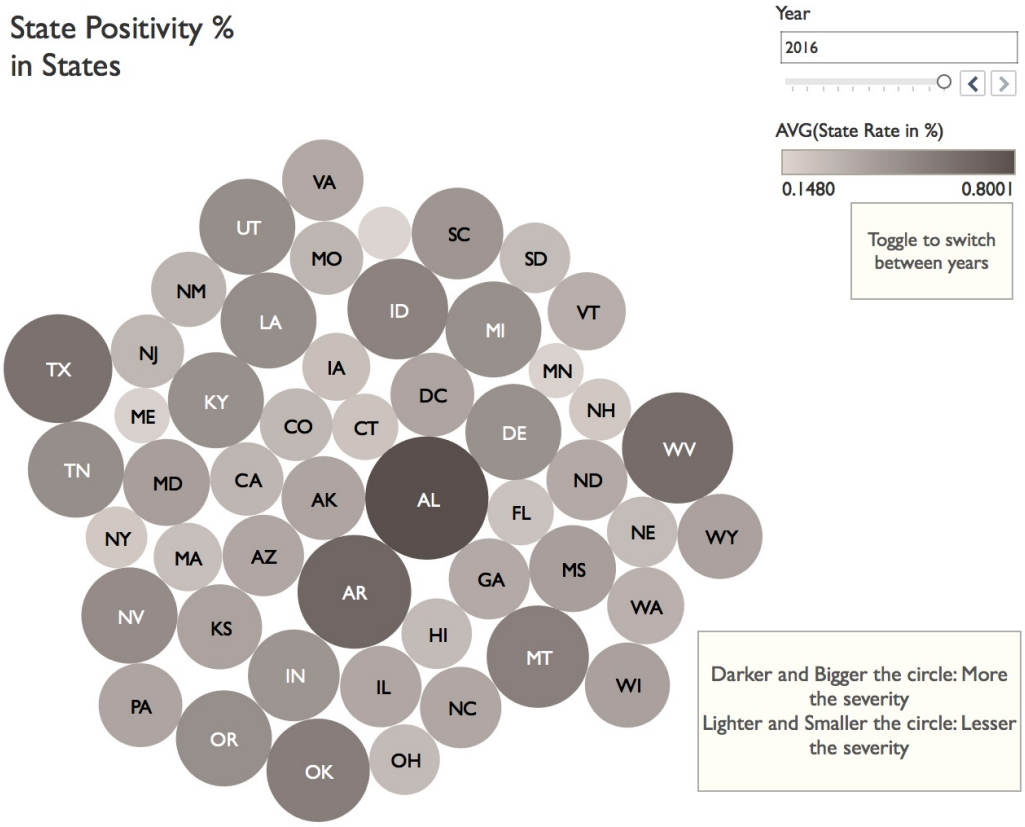

Figure 62: State Positivity % Ranking - 2016: Packed Bubble Chart

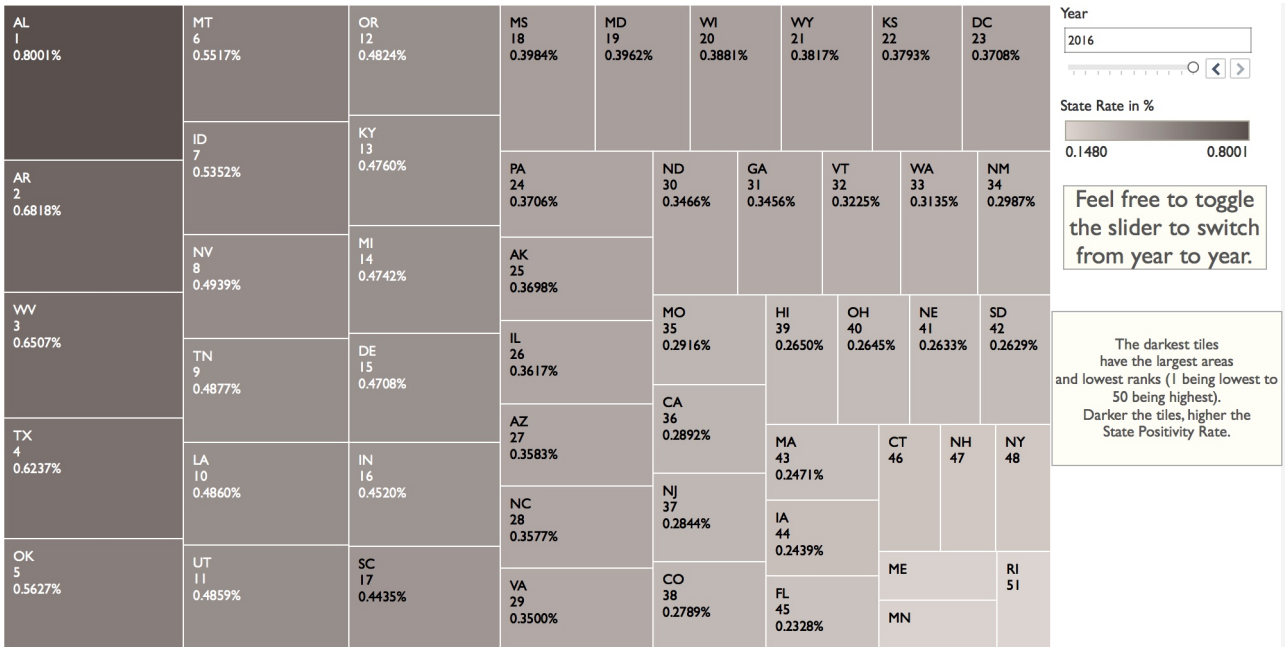

Figure 63: State Positivity % Ranking - 2016: Tree Map

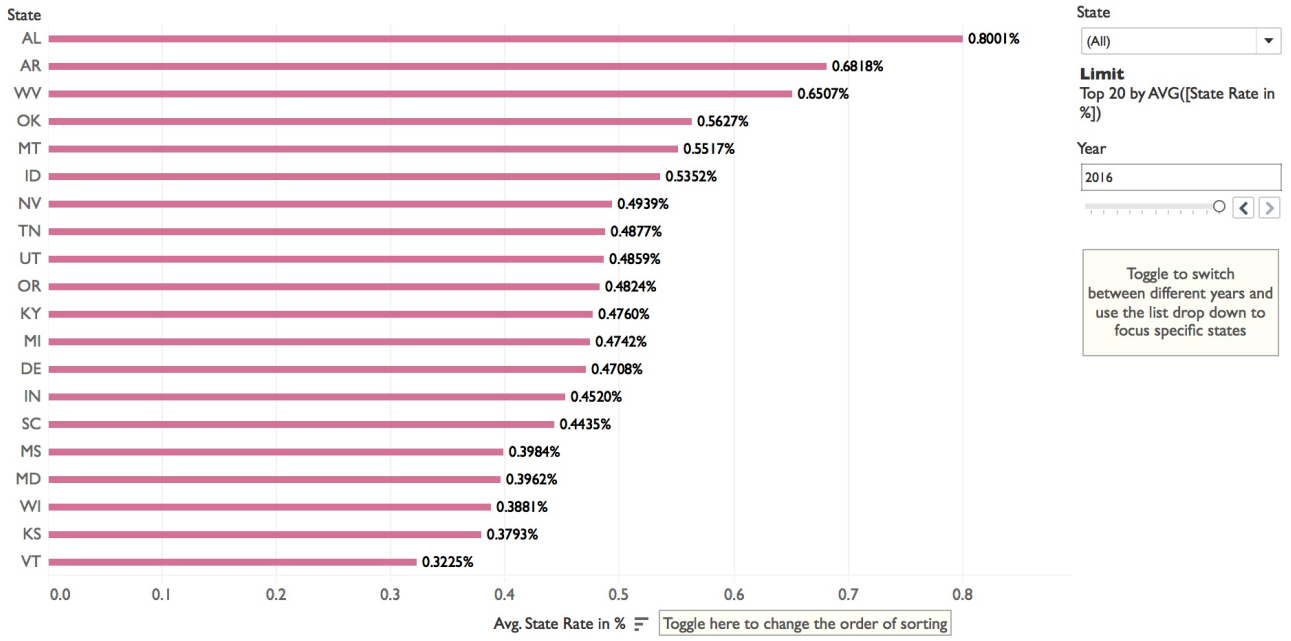

Figure 64: State Positivity % Ranking: Top 20 States with Most Severe Opioid Concentration - 2016

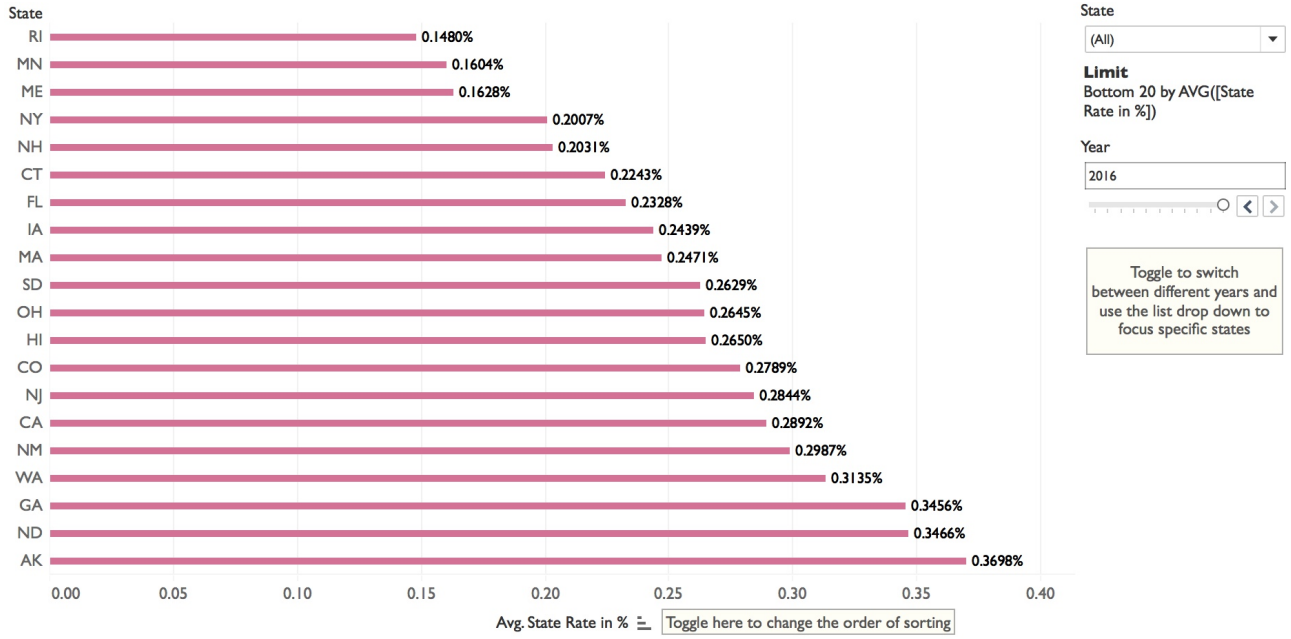

Figure 65: State Positivity % Ranking: Top 20 States with Least Severe Opioid Concentration - 2016

In the year 2016, `AL` - Alabama, `AR` - Arkansas and `WV` - West Virginia have been the states with the highest Opioid Severity. `ME` - Maine, `RI` - Rhode Island and `MN` - Minnesota have had the least Opioid Severity (using State Positivity % as the severity metric).

3.2 Variance of State Positivity Rate: Sporadic Fluctuations

The goal of this visualization is to help understand which of the states have had the maximum variance and minimum variance in their state positivity rate. We would expect that while some states may have been consistently worse and some states may have been consistently good, there are some states where the state positivity rate has been very varied. Variance will help us understand if the positivity rate of states have been all over the place! This may uncover which states underwent some major changes that would have contributed to the sporadic fluctuations.

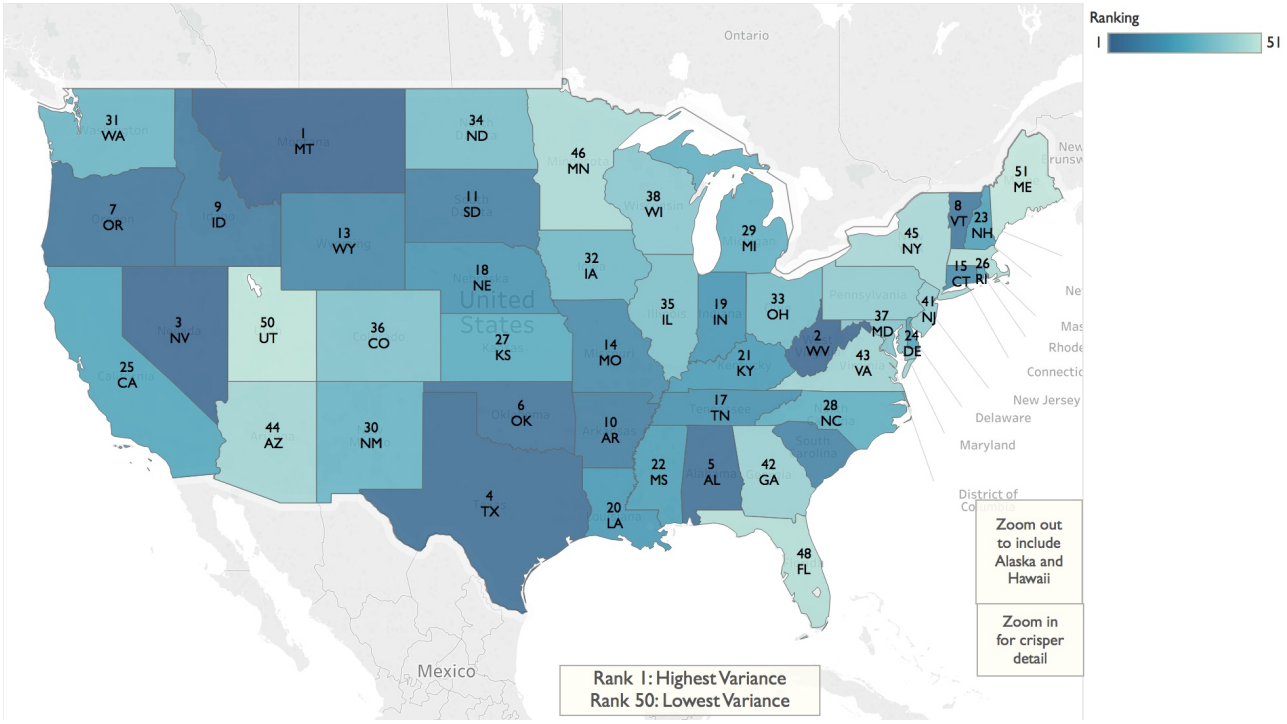

Figure 66: Variance of State Positivity Rate in States from 2007 to 2016: Heat Map

On closer examination, it turns out that, those states that have been consistently worse like `MT` (Montana), `MT` (West Virginia) and `NV` (Nevada), actually have had the highest variation within their state positivity rate. You would expect that these states which bag the medals for constantly have a high positivity rate for many years would have be centered around its own average! It turns out that although they have a high positivity rate, the positivity rate of these states are very spread apart from their mean and fluctuated a lot compared to other states! The states with the lowest positivity rate like `MN` (Maine) and `NY` (New York) actually have their positivity rates centered around their mean (lowest variances). But when you look at the table for some reason, `TX` - Texas is rank 4! This brings us to our next visualization.

3.3 Variations in Severity of Opioid Crisis because of being in close proximity to Mexico

Depending on the source of drug cartels and manufacturing plants and its proximity from them, the opioid concentration could vary! Many news articles like this one from Washington Times explains how drugs coming from the southern border (Mexico) is not helping the United States at all.

The metric used to judge severity of opioid crisis is the `State Positivity Rate`.

The following Heat Maps below help understand how the ranking of opioid severity of states near the border has been fluctuating over the years.

3.3.1 *Over all years*

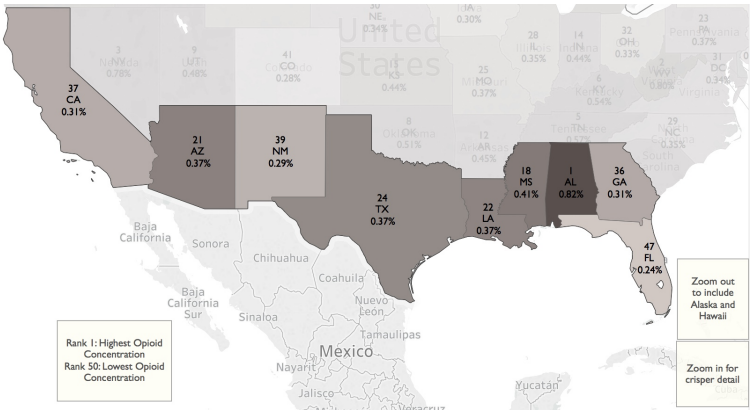

Figure 67: Rank of State Positivity Rate in states closest to the Mexican Border - Over all years

3.3.2 *2007*

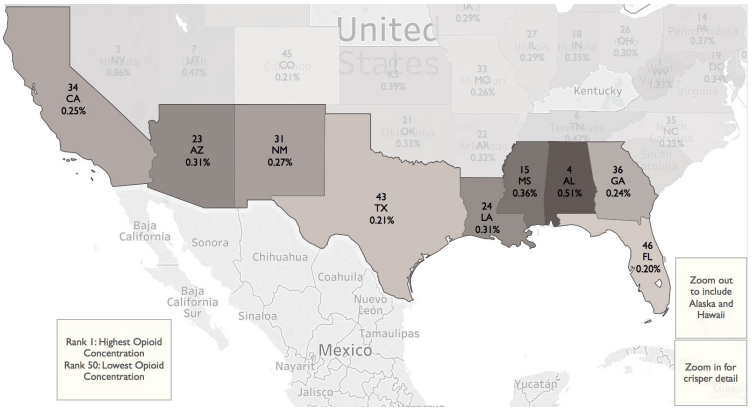

Figure 68: Rank of State Positivity Rate in states closest to the Mexican Border - 2007

3.3.3 *2008*

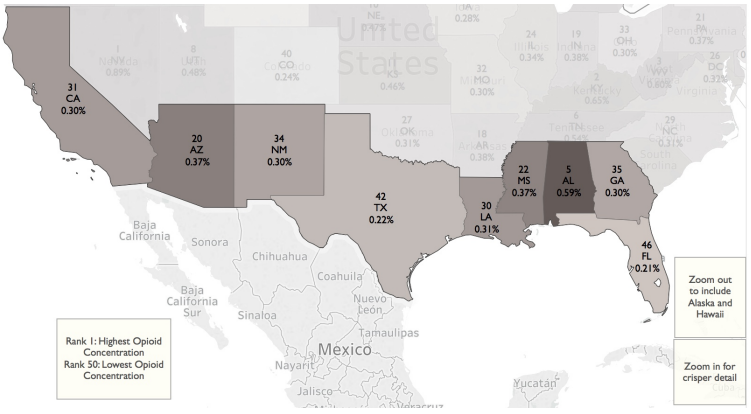

Figure 69: Rank of State Positivity Rate in states closest to the Mexican Border - 2008

3.3.4 *2009*

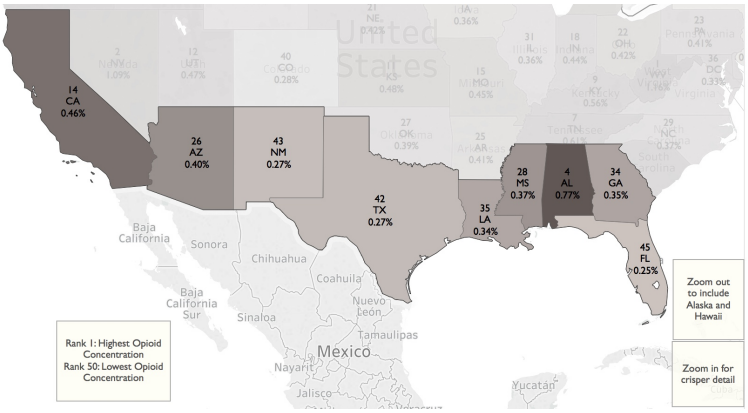

Figure 70: Rank of State Positivity Rate in states closest to the Mexican Border - 2009

3.3.5 *2010*

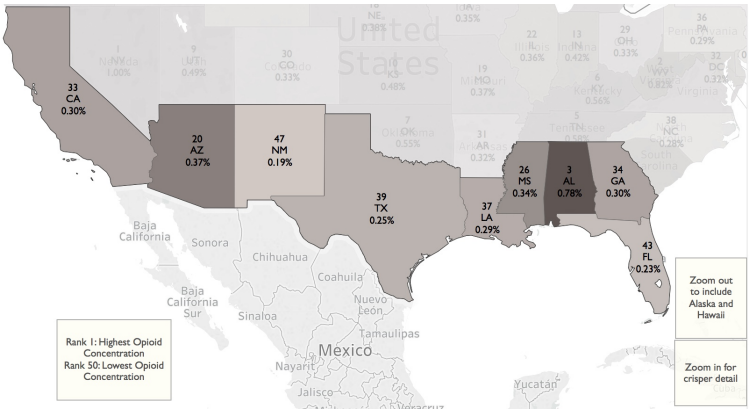

Figure 71: Rank of State Positivity Rate in states closest to the Mexican Border - 2010

3.3.6 *2011*

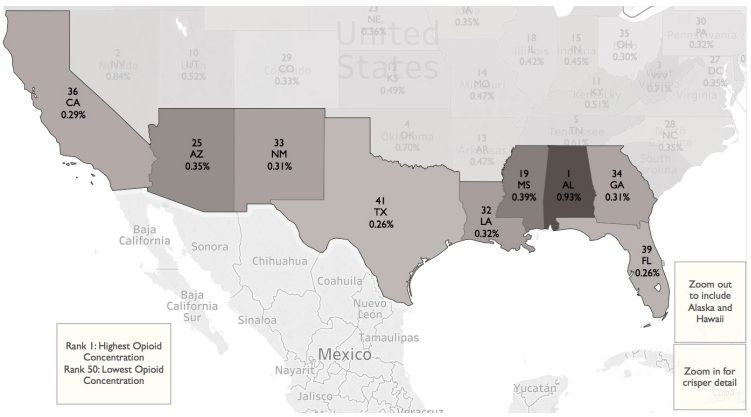

Figure 72: Rank of State Positivity Rate in states closest to the Mexican Border - 2011

3.3.7 *2012*

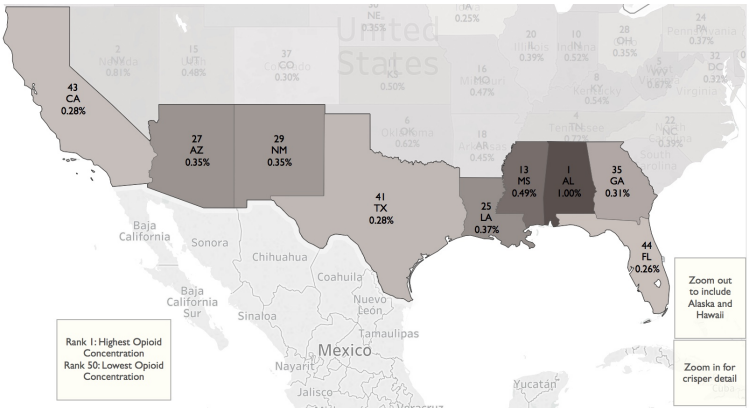

Figure 73: Rank of State Positivity Rate in states closest to the Mexican Border - 2012

3.3.8 *2013*

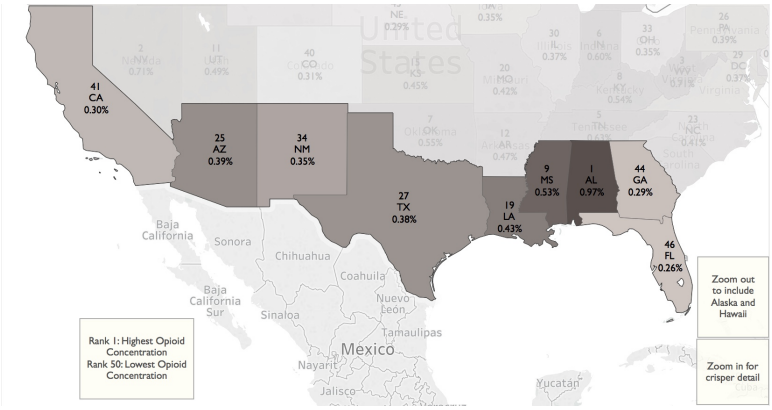

Figure 74: Rank of State Positivity Rate in states closest to the Mexican Border - 2013

3.3.9 *2014*

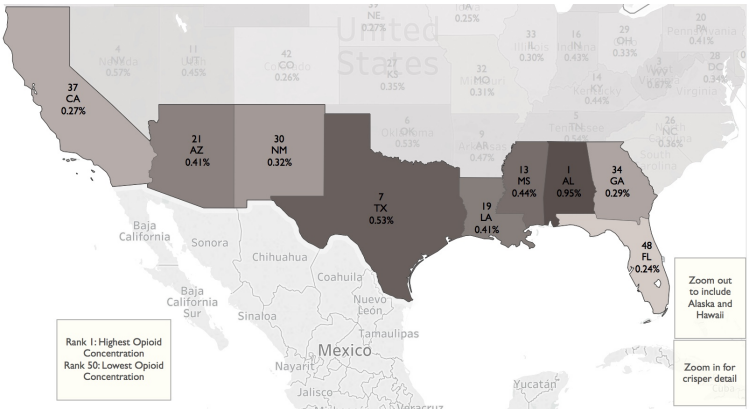

Figure 75: Rank of State Positivity Rate in states closest to the Mexican Border - 2014

3.3.10 *2015*

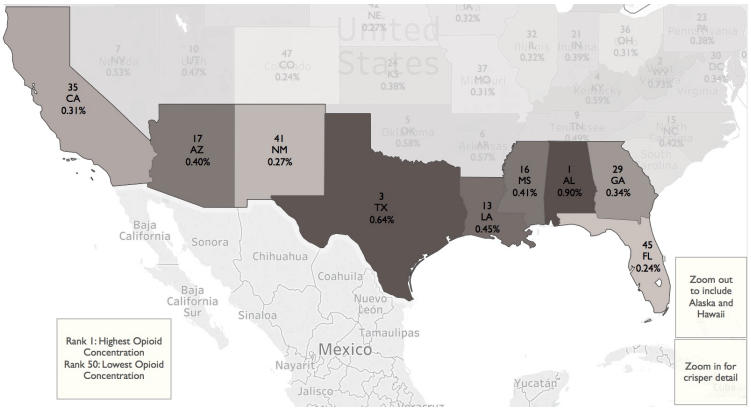

Figure 76: Rank of State Positivity Rate in states closest to the Mexican Border - 2015

3.3.11 *2016*

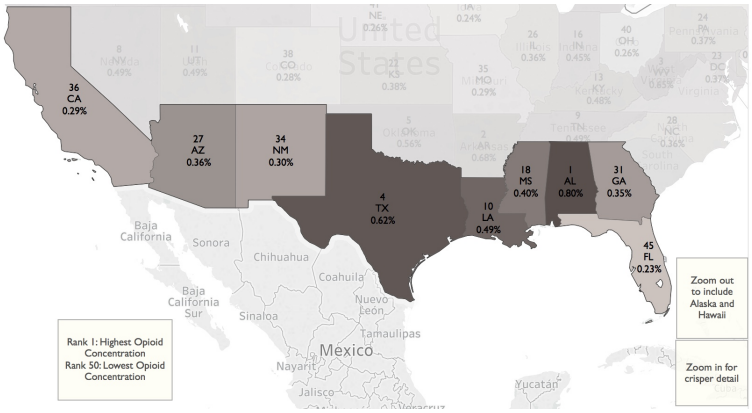

Figure 77: Rank of State Positivity Rate in states closest to the Mexican Border - 2016

3.4 Texas, the state closest to the Mexican Border gets progressively worse from year 2012

Table 1: State Positivity Rate Rank of Texas

Year	Rank (State Positivity Rate)
2007	43
2008	42
2009	42
2010	39
2011	41
2012	31
2013	27
2014	7
2015	3
2016	4

When we used the metric we extracted from the data called `Opioid Severity` (Weighted Average Metric), we get a similar picture.

Table 2: Opioid Severity Rank (Weighted Average Metric)

Year	Rank (Weighted Average Metric)
2007	42
2008	39
2009	43
2010	39
2011	41
2012	38
2013	21
2014	4
2015	3
2016	1

When we use the Weighted Average Metric, from the year 2012, Texas goes from being Rank 38 to Rank 1 in 2016. That just goes to show how being close to the source of manufacture may have an impact on how much of those drugs are being distributed.

4 CORRELATIONS BETWEEN OPIOID SEVERITY AND EMPLOYMENT STATISTICS

The two opioid severity metrics we will be considering are:

1. `State Rate in %`
2. `Size of Labor Force`

All employment related statistics which are present in `stateOpioidEmployment.csv` were aggregated by state by year using the original data obtained from Bureau of Labor Statistics.

4.1 Pearson Correlation Coefficient

Pearson Correlation Coefficient is one of the ways to quantify the existence of a linear relationship between two variables.

- Pearson Correlation Coefficient > 0.5 - Moderate to Strong Positive Linear Relationship
- 0.5 < Pearson Correlation Coefficient < -0.5 - Close to no Linear Relationship
- Pearson Correlation Coefficient < -0.5 - Moderate to Strong Negative Linear Relationship

Also,

- Strong Positive Linear Relationship - Both variables increase together
- Strong Negative Linear Relationship - As one variable increases, the other variable decreases

4.2 Correlation between State Positivity Rate and Average Unemployment Rate

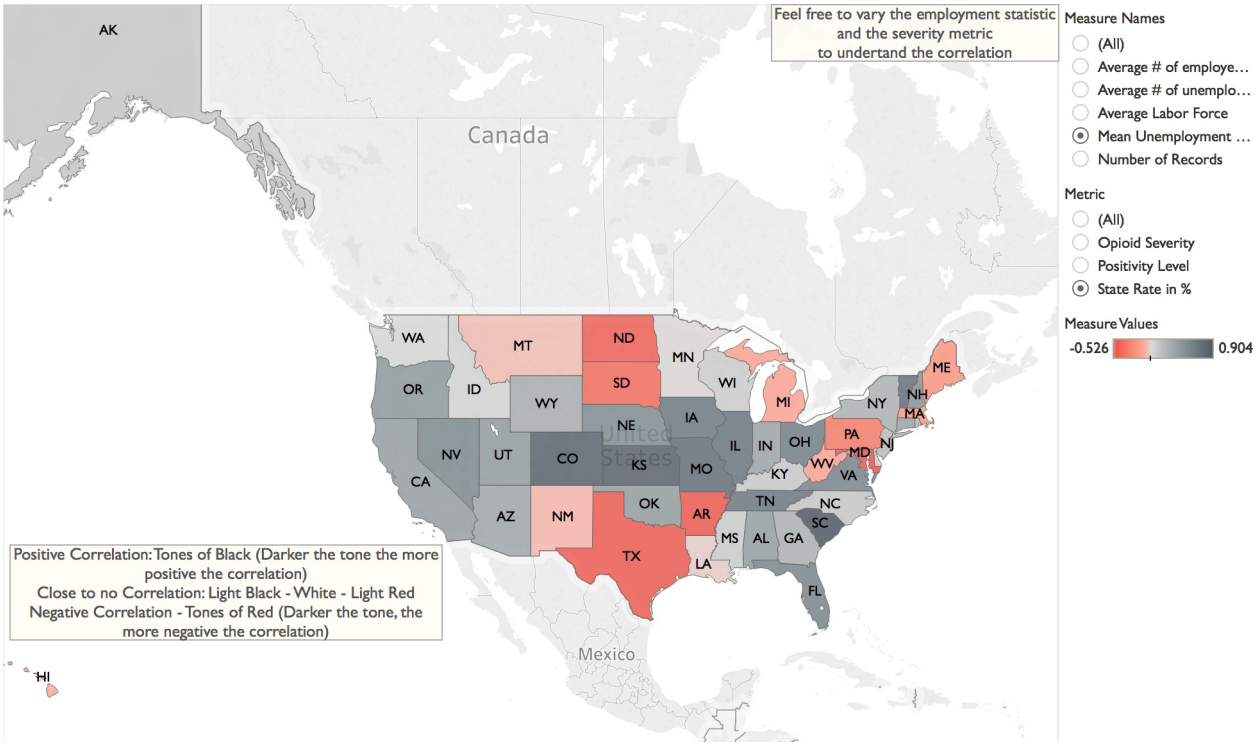

Figure 78: Correlation between State Positivity Rate and Average Unemployment Rate: Heat Map

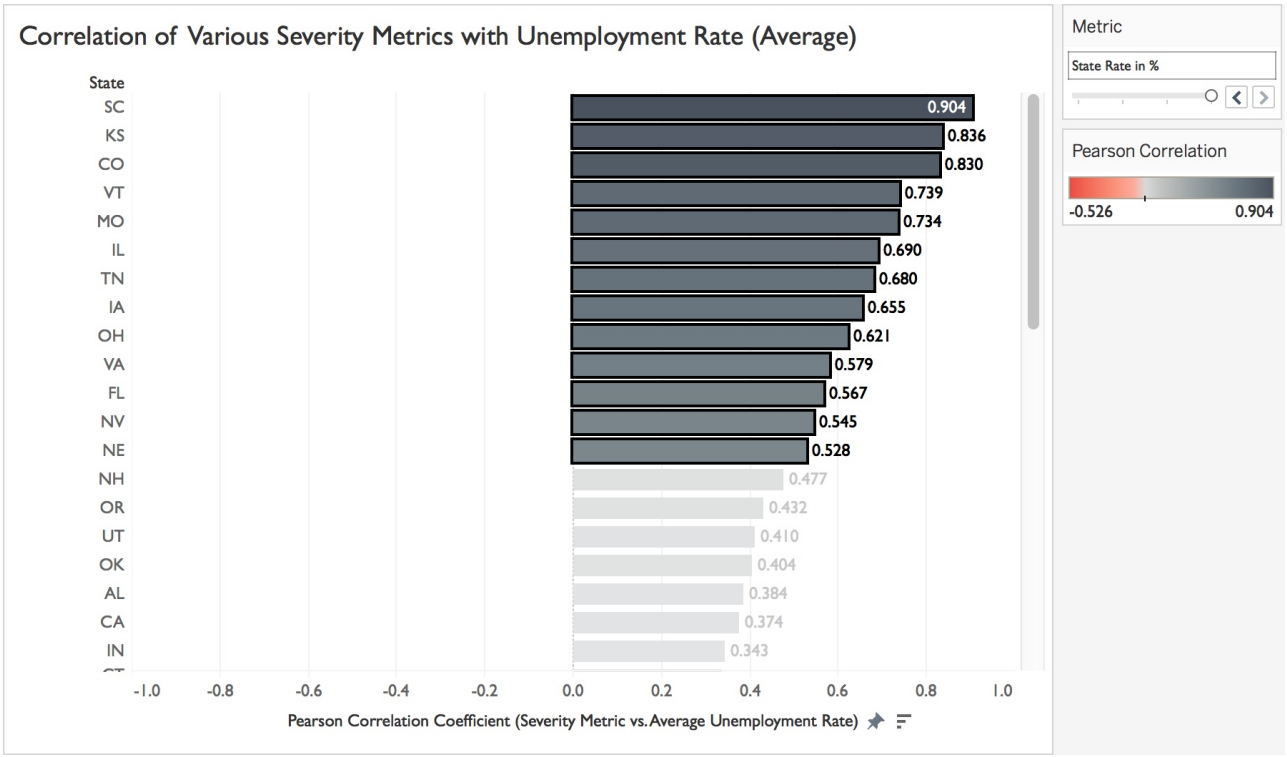

Figure 79: Positive Correlation between State Positivity Rate and Average Unemployment Rate

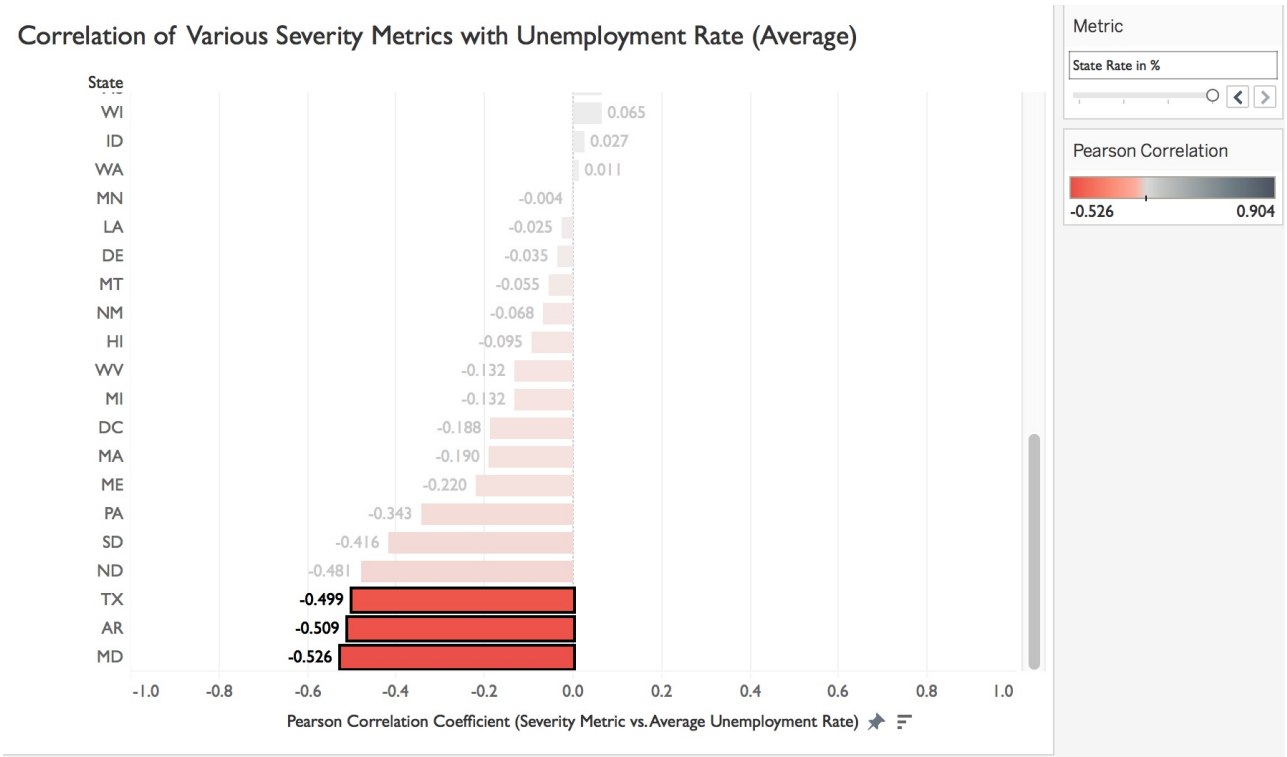

Figure 80: Negative Correlation between State Positivity Rate and Average Unemployment Rate

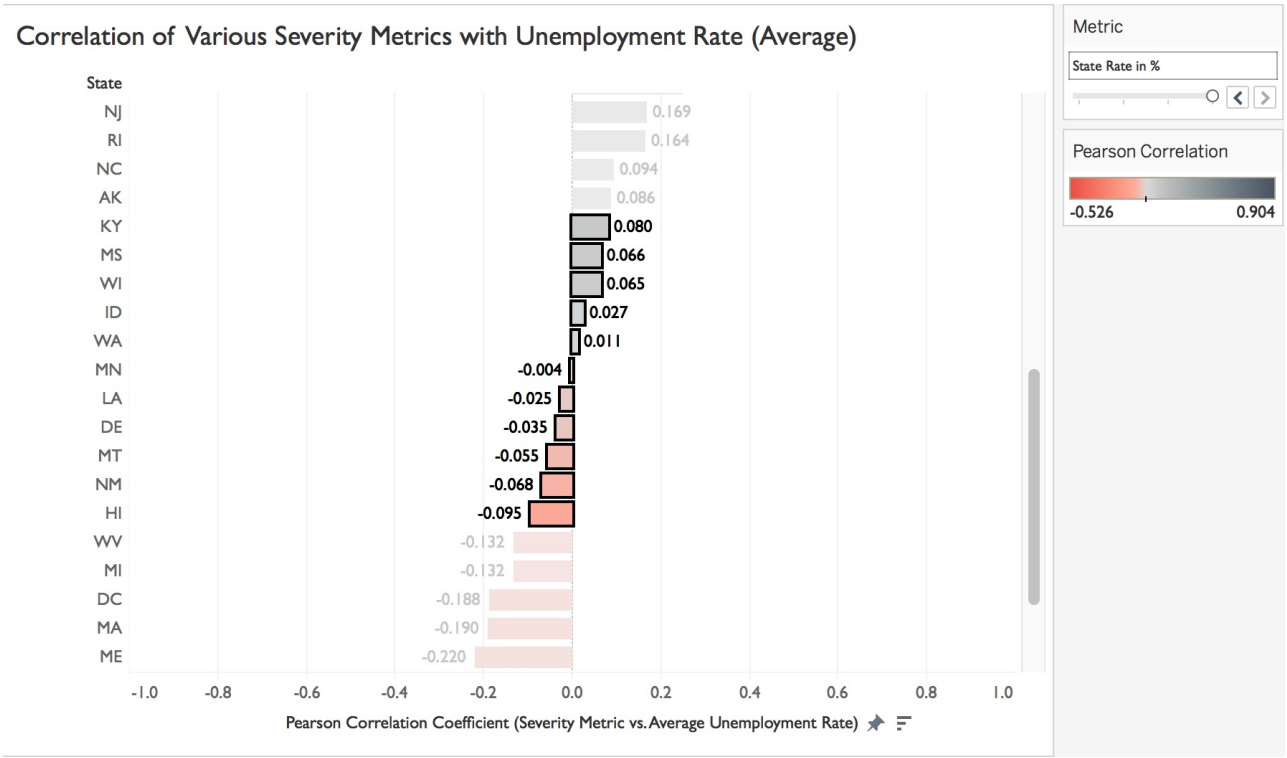

Figure 81: Close to no Correlation between State Positivity Rate and Average Unemployment Rate

4.3 Correlation between State Positivity Rate and Average Labor Force

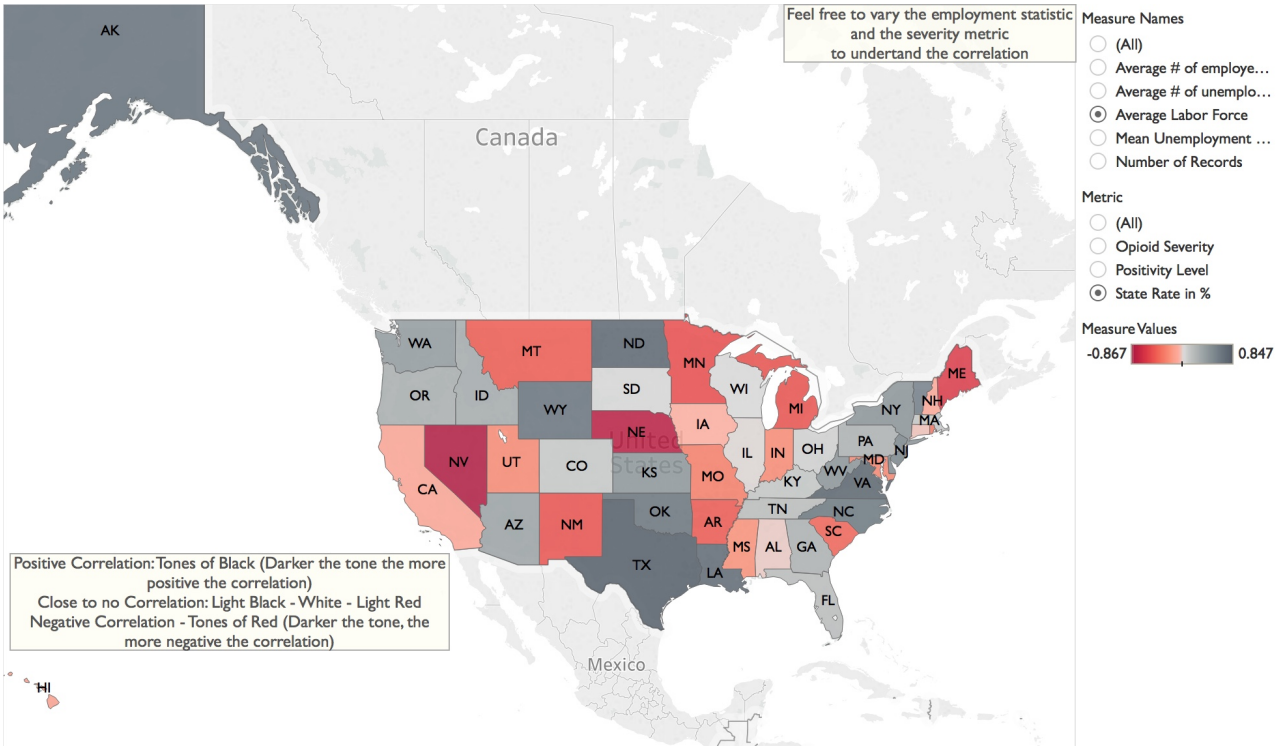

Figure 82: Correlation between State Positivity Rate and Average Labor Force: Heat Map

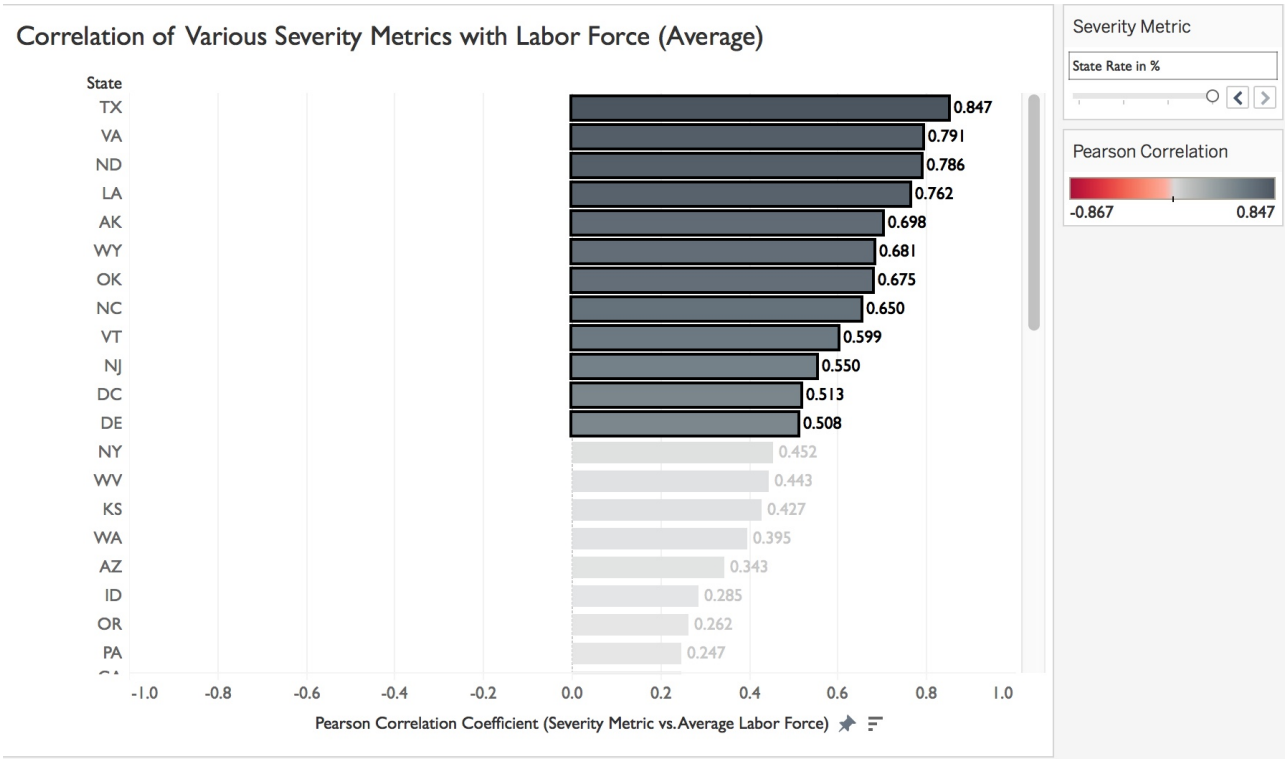

Figure 83: Positive Correlation between State Positivity Rate and Average Labor Force

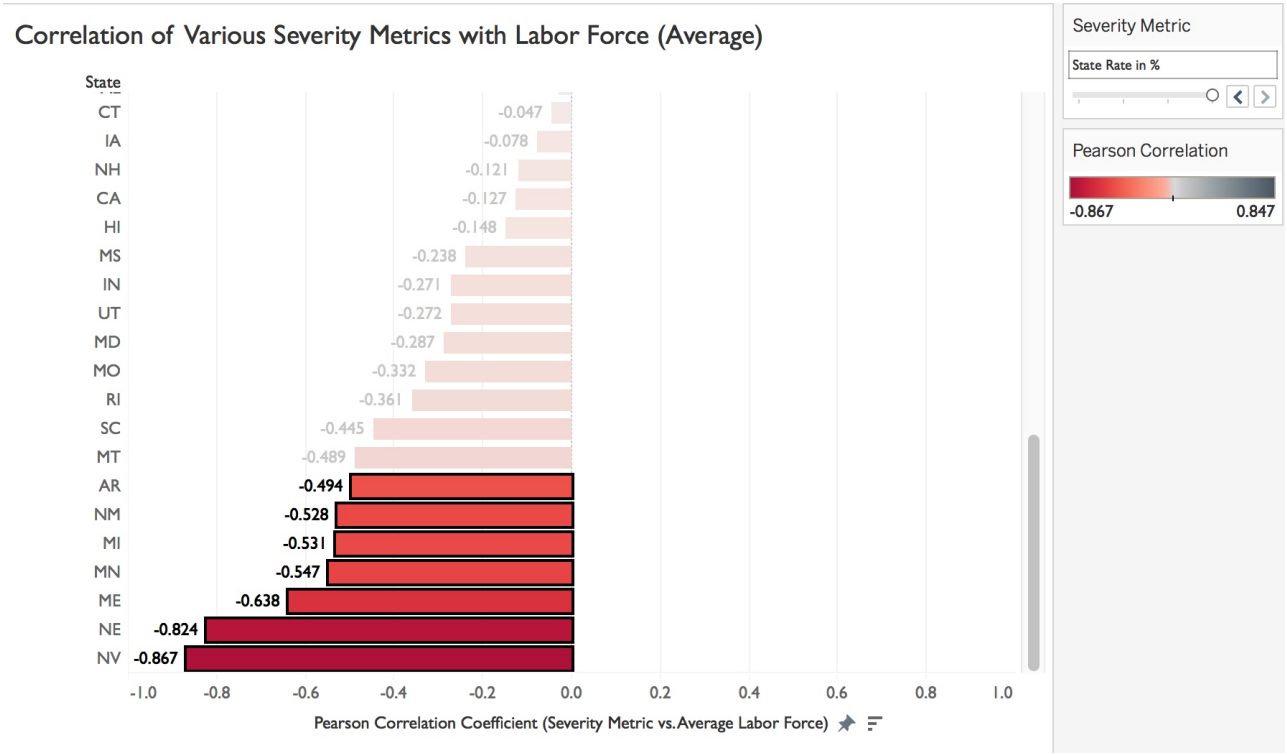

Figure 84: Negative Correlation between State Positivity Rate and Average Labor Force

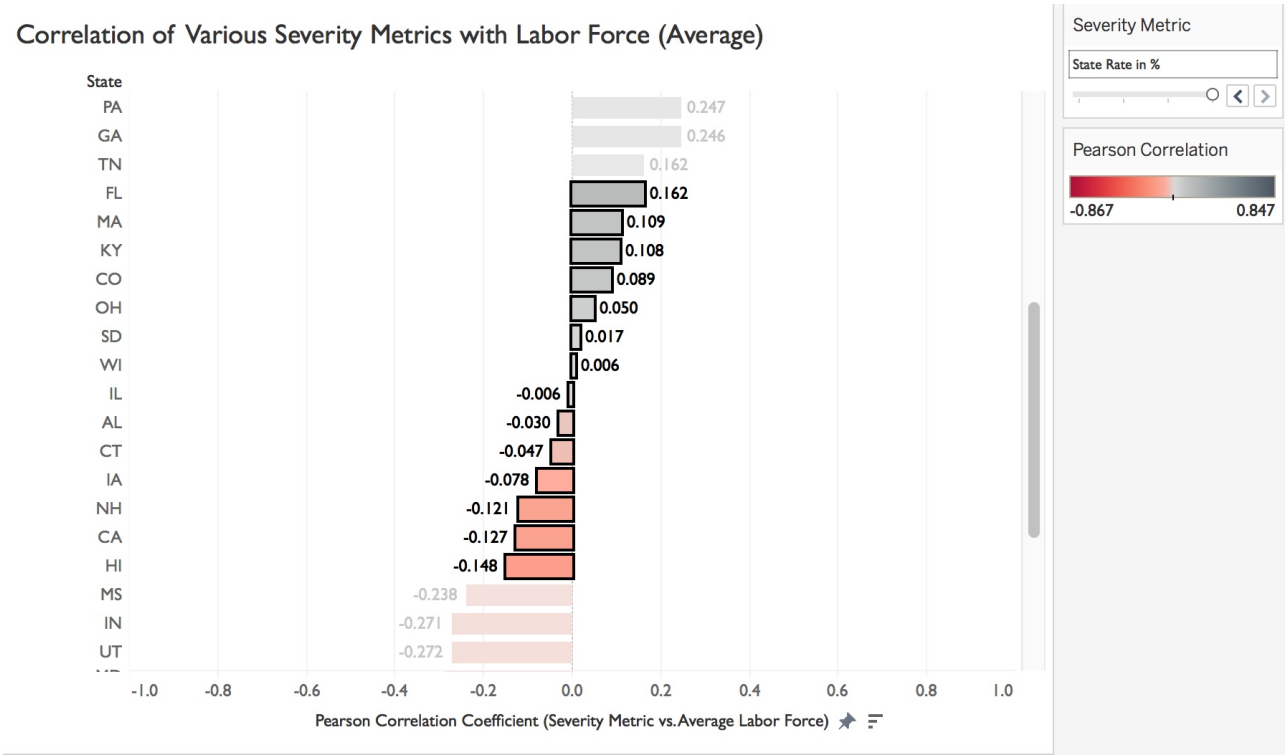

Figure 85: Close to no Correlation between State Positivity Rate and Average Labor Force

4.4 Summary

4.4.1 *Correlation between State Positivity Rate and Average Unemployment Rate*

STATES THAT ARE POSITIVELY CORRELATED WITH UNEMPLOYMENT RATE	
STATE	PEARSON CORRELATION COEFFICIENT
South Carolina	0.904
Kansas	0.836
Colorado	0.83
Vermont	0.739
Missouri	0.734
Illinois	0.69
Tennessee	0.68
Iowa	0.655
Ohio	0.621
Virginia	0.579
Florida	0.567
Nevada	0.545
Nebraska	0.528

Figure 86: States having Positive Correlation between State Positivity Rate and Average Unemployment Rate

STATES THAT ARE NEGATIVELY CORRELATED WITH UNEMPLOYMENT RATE	
STATE	PEARSON CORRELATION COEFFICIENT
South Dakota	-0.416
North Dakota	-0.481
Texas	-0.499
Arkansas	-0.509
Maryland	-0.526

Figure 87: States having Negative Correlation between State Positivity Rate and Average Unemployment Rate

STATES THAT HAVE CLOSE TO NO CORRELATION WITH UNEMPLOYMENT RATE	
STATE	PEARSON CORRELATION COEFFICIENT
Kentucky	0.08
Mississippi	0.066
Wisconsin	0.065
Idaho	0.027
Washington	0.011
Minnesota	-0.004
Louisiana	-0.025
Delaware	-0.035
Montana	-0.055
New Mexico	-0.068
Hawaii	-0.095

Figure 88: States having close to No Correlation between State Positivity Rate and Average Unemployment Rate

4.4.2 *Correlation between State Positivity Rate and Average Labor Force*

STATES THAT ARE POSITIVELY CORRELATED WITH SIZE OF LABOR FORCE	
STATE	PEARSON CORRELATION COEFFICIENT
Texas	0.847
Virginia	0.791
North Dakota	0.786
Louisiana	0.762
Alaska	0.698
Wyoming	0.681
Oklahoma	0.675
North Carolina	0.65
Vermont	0.599
New Jersey	0.55
District of Columbia	0.513
Delaware	0.508

Figure 89: States having Positive Correlation between State Positivity Rate and Average Labor Force

STATES THAT ARE NEGATIVELY CORRELATED WITH SIZE OF LABOR FORCE	
STATE	PEARSON CORRELATION COEFFICIENT
Nevada	-0.867
Nebraska	-0.824
Maine	-0.638
Minnesota	-0.547
Michigan	-0.531
New Mexico	-0.528
Arkansas	-0.494

Figure 90: States having Negative Correlation between State Positivity Rate and Average Labor Force

STATES THAT HAVE CLOSE TO NO CORRELATION WITH SIZE OF LABOR FORCE	
STATE	PEARSON CORRELATION COEFFICIENT
Florida	0.162
Massachusetts	0.109
Kentucky	0.108
Colorado	0.089
Ohio	0.05
South Dakota	0.017
Wisconsin	0.006
Illinois	-0.006
Alabama	-0.03
Connecticut	-0.047
Iowa	-0.078
New Hampshire	-0.121
California	-0.127
Hawaii	-0.148

Figure 91: States having close to No Correlation between State Positivity Rate and Average Labor Force

---

5 LINKS

Please visit the following links for a full set of visualizations.

- Opioid State Positivity Visualizations
- Correlations between Employment Statistics and State Positivity Rate

REFERENCES

[1] Bureau of Labor Statistics. URL: https://www.bls.gov/lau/laufaq.htm#Q03.