Exploring Hospital Fall Rates in 500 U.S. Cities

How hospitals can better predict and explain patient fall rates.

Why Explain and Predict Hospital Falls?

Patient Falls can increase:

- patient care costs
- length of stay
- liability

Centers for Medicare and Medicaid Services (CMS) no longer pays for health care costs associated with falls during hospitalization

How this project can help in the Healthcare Market.

With the constraints and demands facing hospitals I am seeking to:

- help hospital know how to use their time, money, and resources to reduce their patient fall rates.
- work with consulting healthcare companies to create products that can prevent hospital falls
- help a governmental agency categorize and monitor how hospitals are performing

From Fall Data to Modeling

To make a robust model, I wanted strong reliable data about hospital fall rates from around the country.

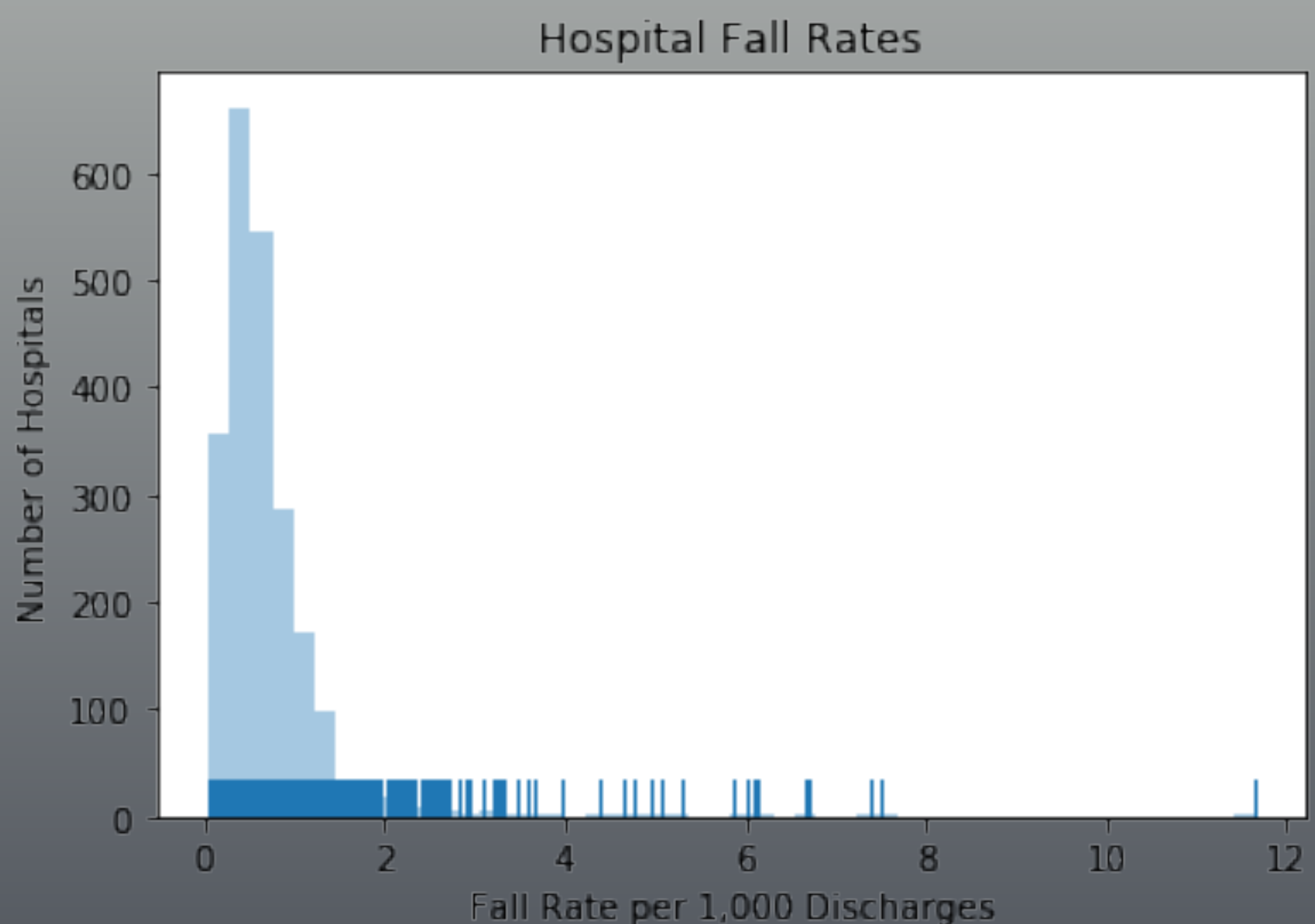
Since this dataset did not already exist, I decided to find and merge two separate fall dataset.

- The first came from hospital-data.com with its profiles of thousands of hospitals, medical clinics, nursing homes and home health centers

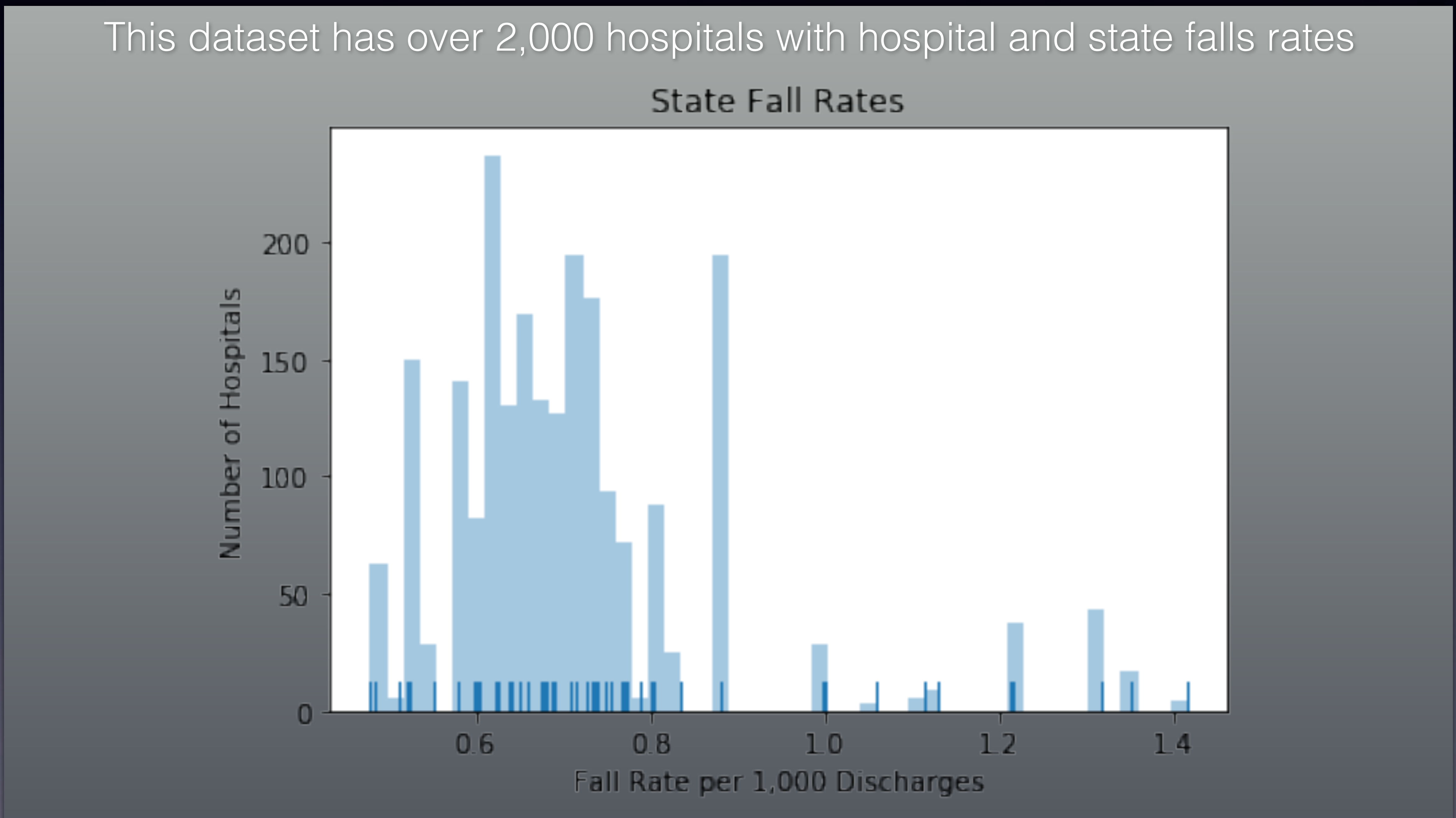
- The second was from the Centers for Medicare & Medicaid Services Hospital Level calculations for 8 Hospital Acquired Conditions (July 1, 2010 through June 30, 2012)

Fall Data from Hospital-Data.com

This dataset has over 2,000 hospitals with hospital and state falls rates

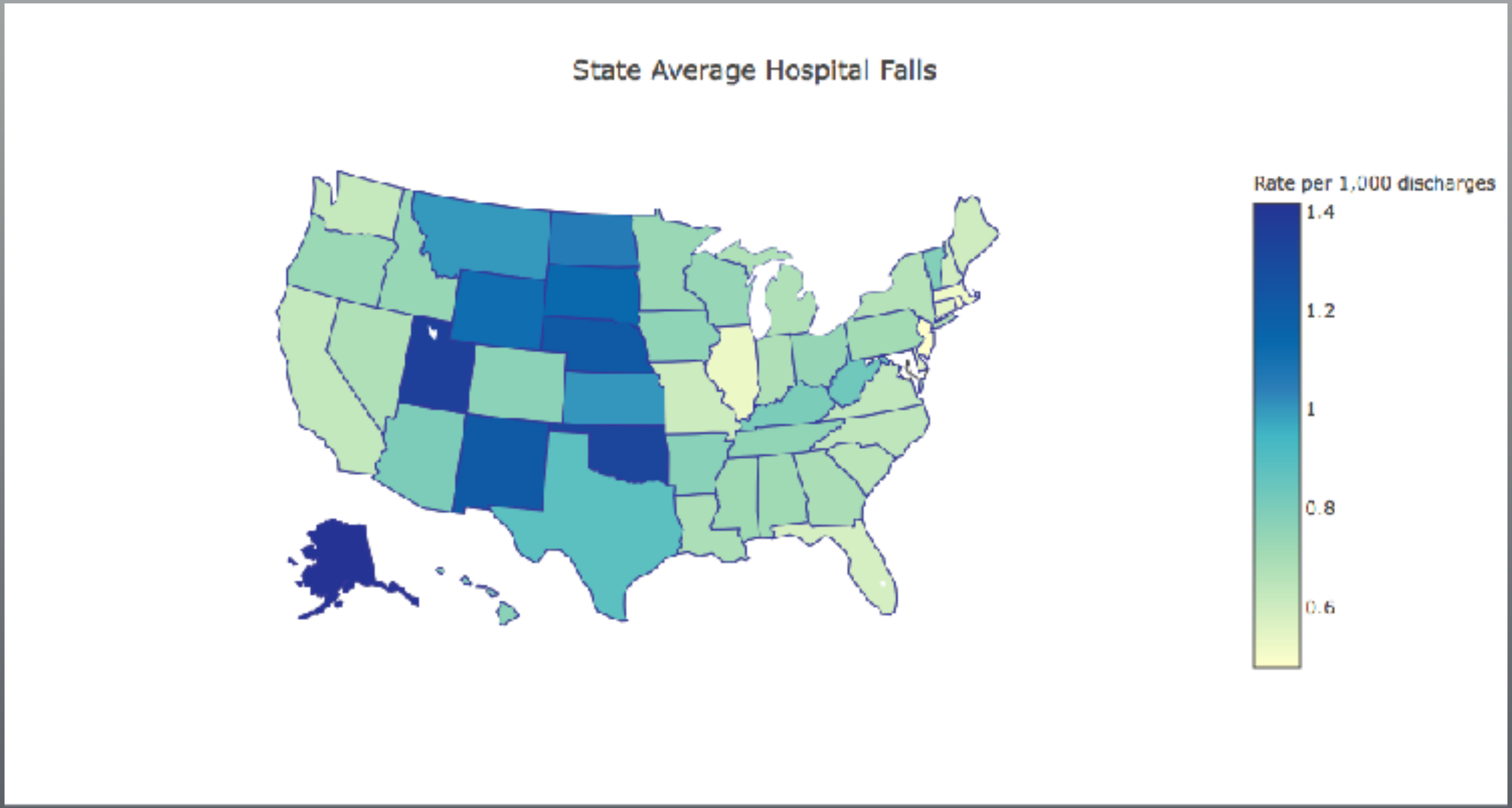


Fall Data from Hospital-Data.com



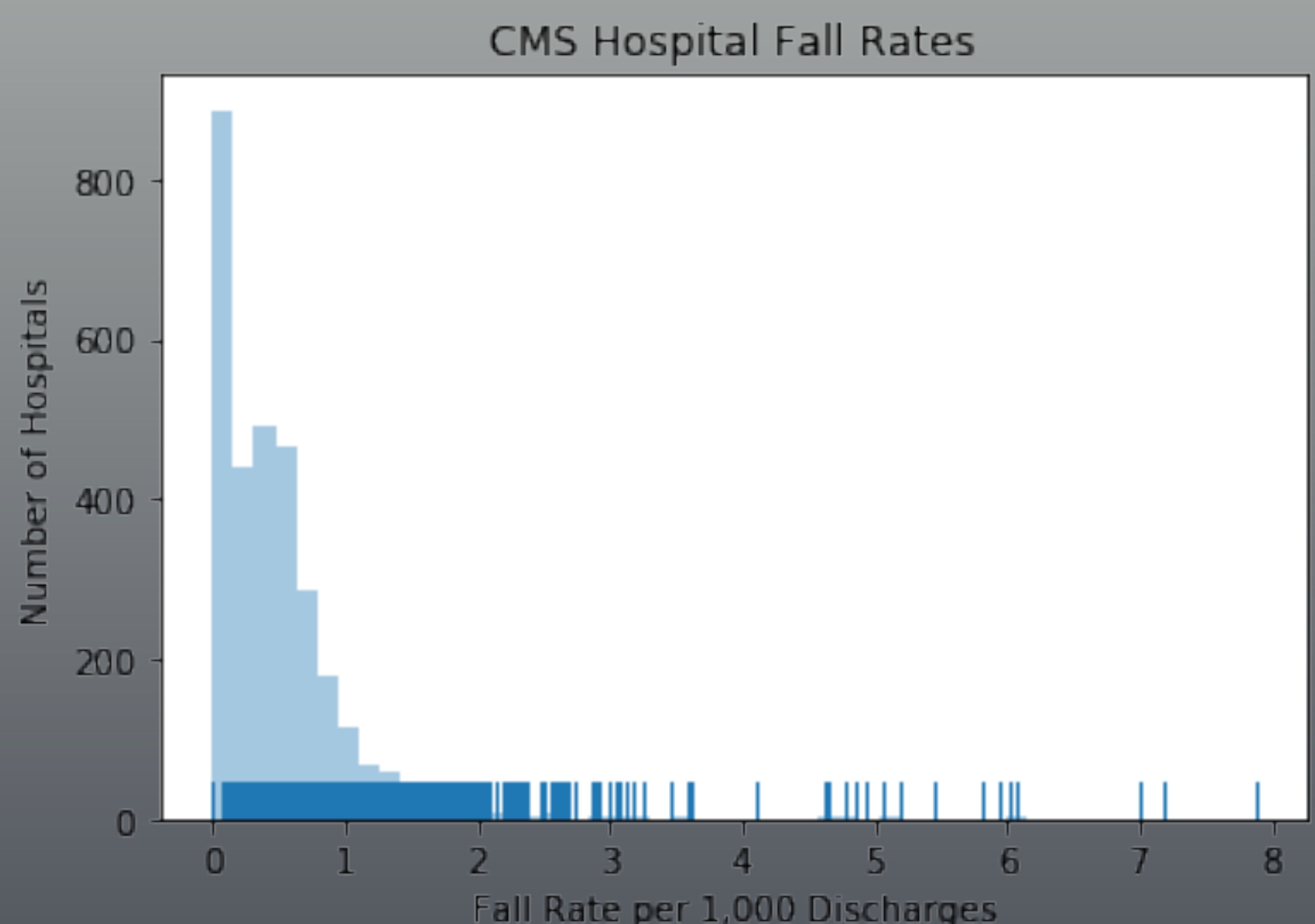
Fall Data from Hospital-Data.com

This dataset has over 2,000 hospitals with hospital and state falls rates



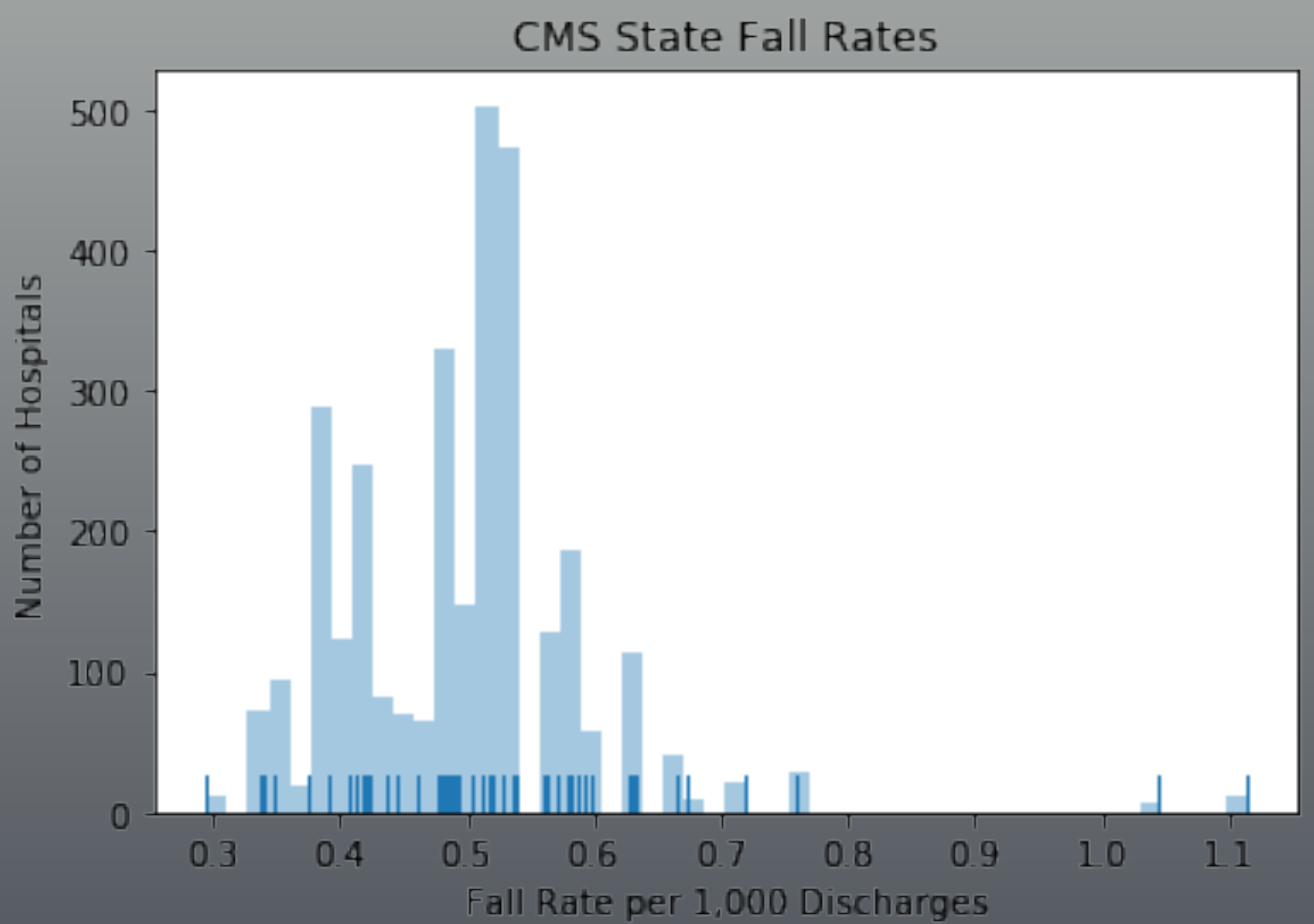
Fall Data from the Centers for Medicare & Medicaid Services

This dataset has over 3,000 hospitals with hospital and state falls rates



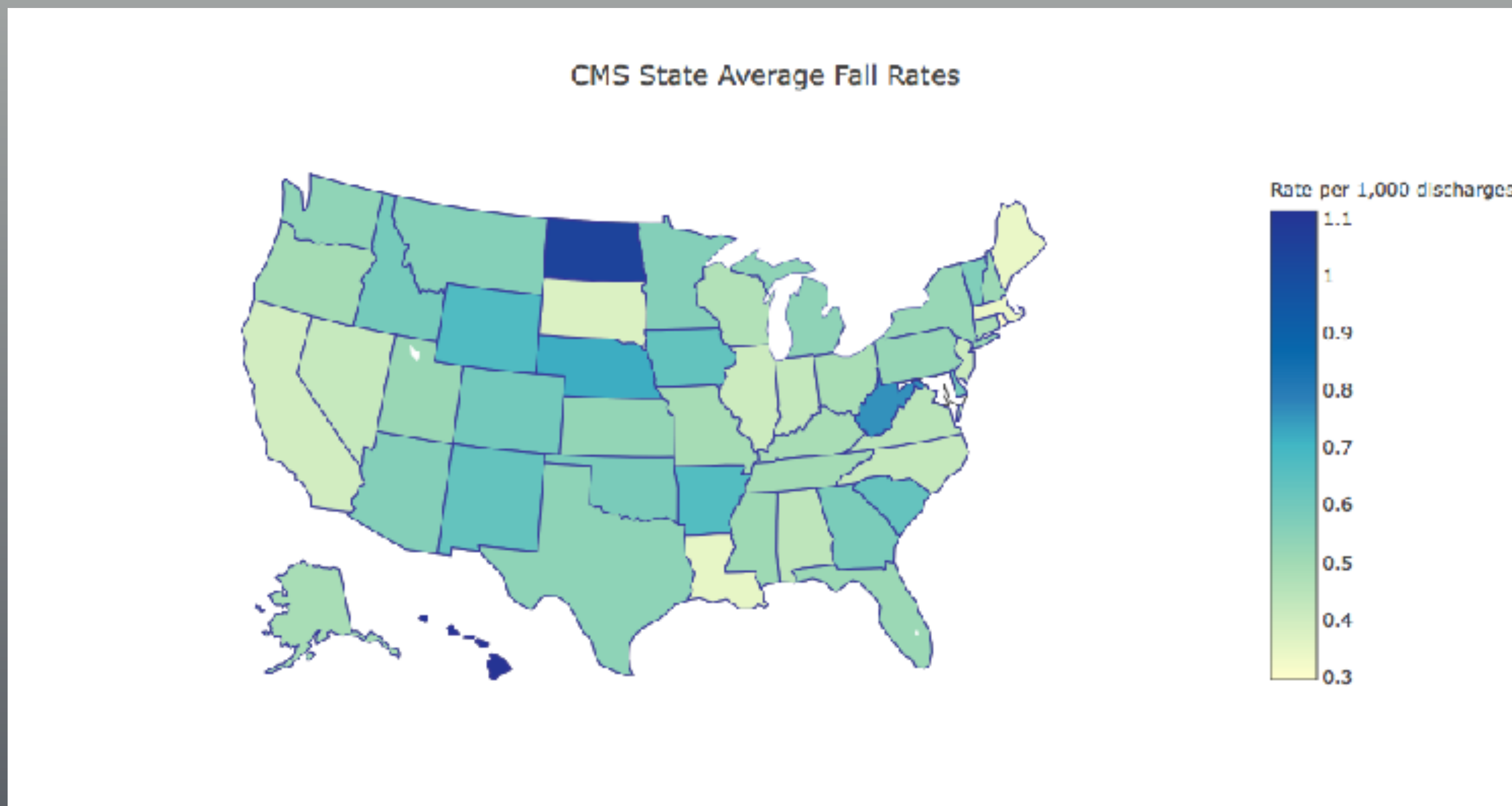
Fall Data from the Centers for Medicare & Medicaid Services

This dataset has over 3,000 hospitals with hospital and state falls rates



Fall Data from the Centers for Medicare & Medicaid Services

This dataset has over 3,000 hospitals with hospital and state falls rates



Calculating Fall Meta-Analysis

Meta-analysis is:

- an approach to combine studies or samples when trying to answer a specific question.
- used because there is a chance that our sample or study is not representative of what is being studied.

Calculating Fall Meta-Analysis

If the studies or samples being combined do NOT use the same scale for measuring the outcome.

- Study scores are converted to what is called effect sizes
- One of the simplest and most common is called Cohen's d:
  - The difference in scores between two groups divided by the standard deviation.
  - d=(Mean1–Mean2) / SD

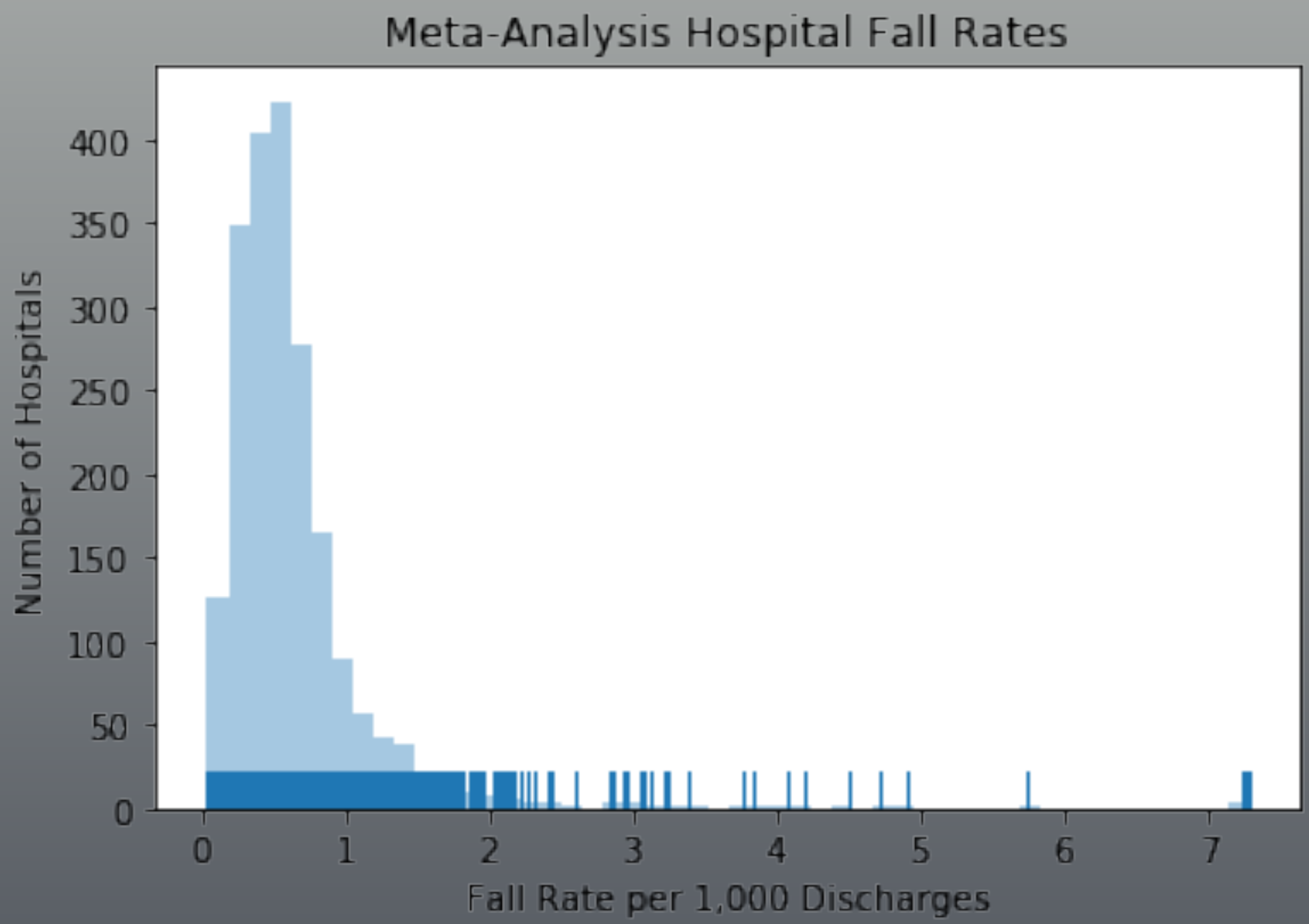
Calculating Fall Meta-Analysis

If the studies or samples being combined DO use the same scale for measuring the outcome.

- Then basic averaging between the samples is an option.
- For my capstone, both of my samples of hospital fall rates were measured on the same scale: rate per 1,000 discharges.
    - I decided to average the hospital and state fall rates for the same hospital from both datasets.

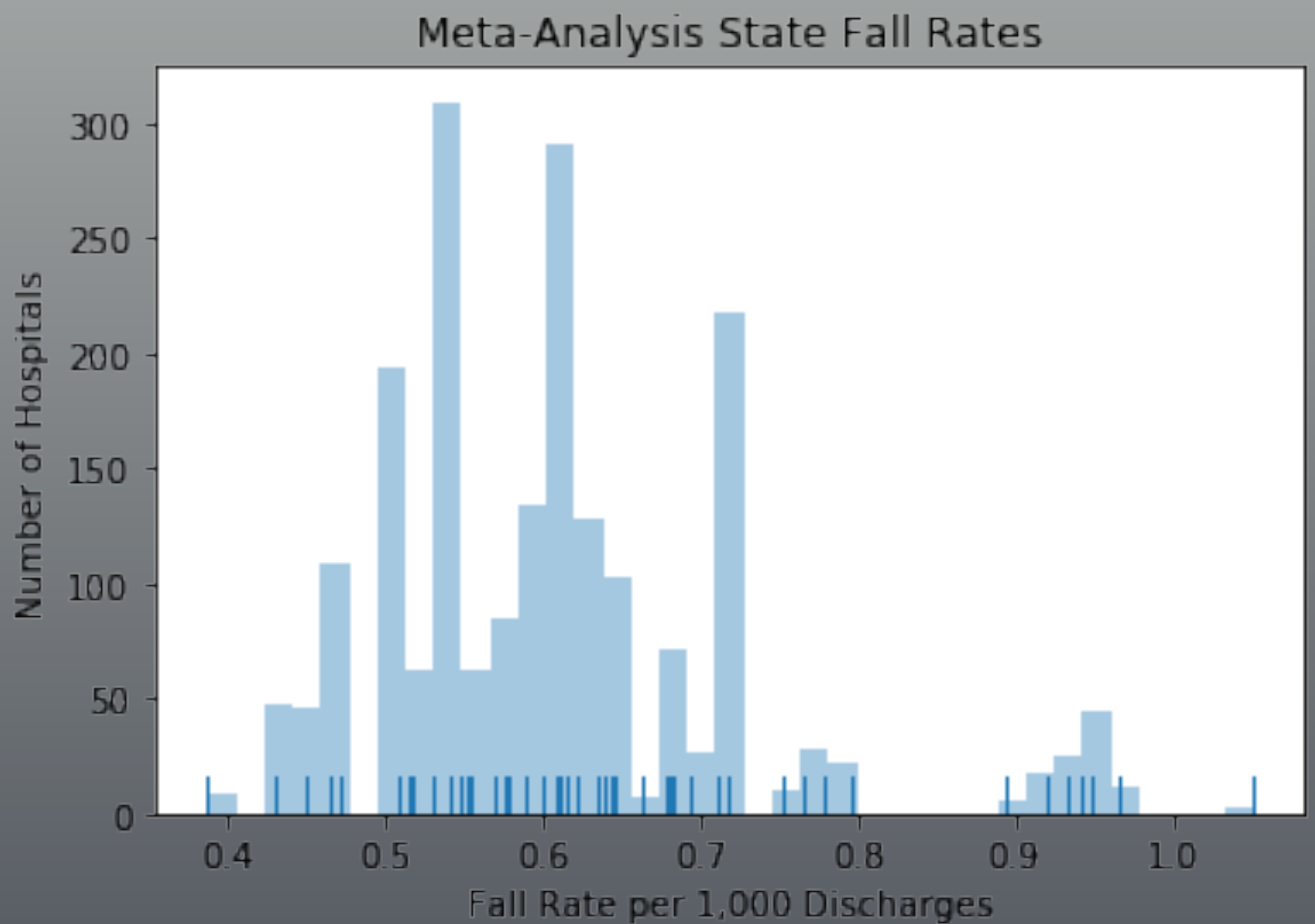
Merged Dataset with Meta-Analyses Fall Rates

This final dataset has over 2,000 hospitals from both dataset



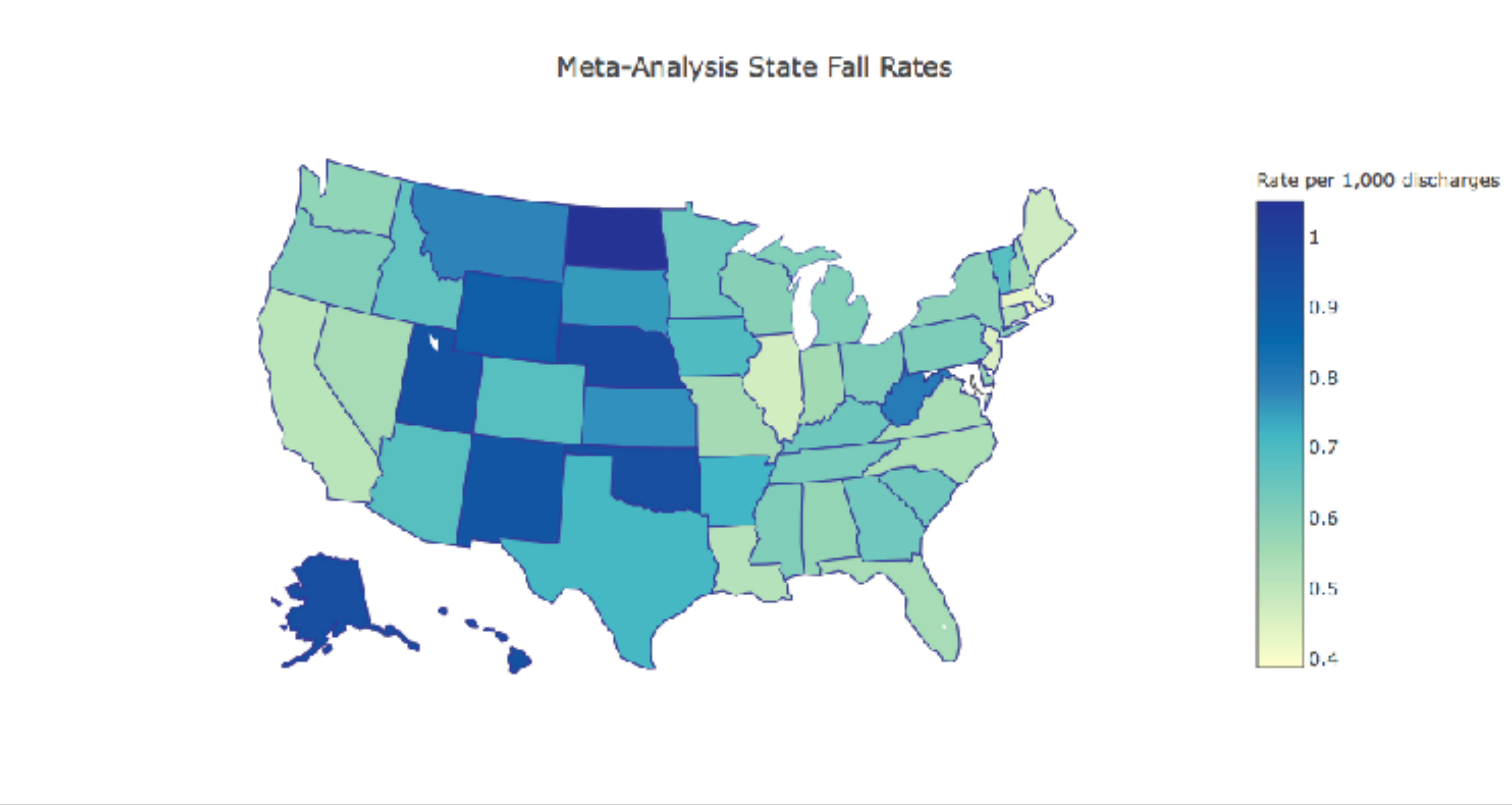
Merged Dataset with Meta-Analyses Fall Rates

This final dataset has over 2,000 hospitals from both dataset



Merged Dataset with Meta-Analyses Fall Rates

This final dataset has over 2,000 hospitals from both dataset



Exploring and Adding Features

- Hospital Feature
  - Trauma Centers
- State Feature
  - Healthcare Employee Injury Rates
- City Feature
  - 500 largest cities in the United States with estimates for chronic disease, health outcomes, and preventive services.

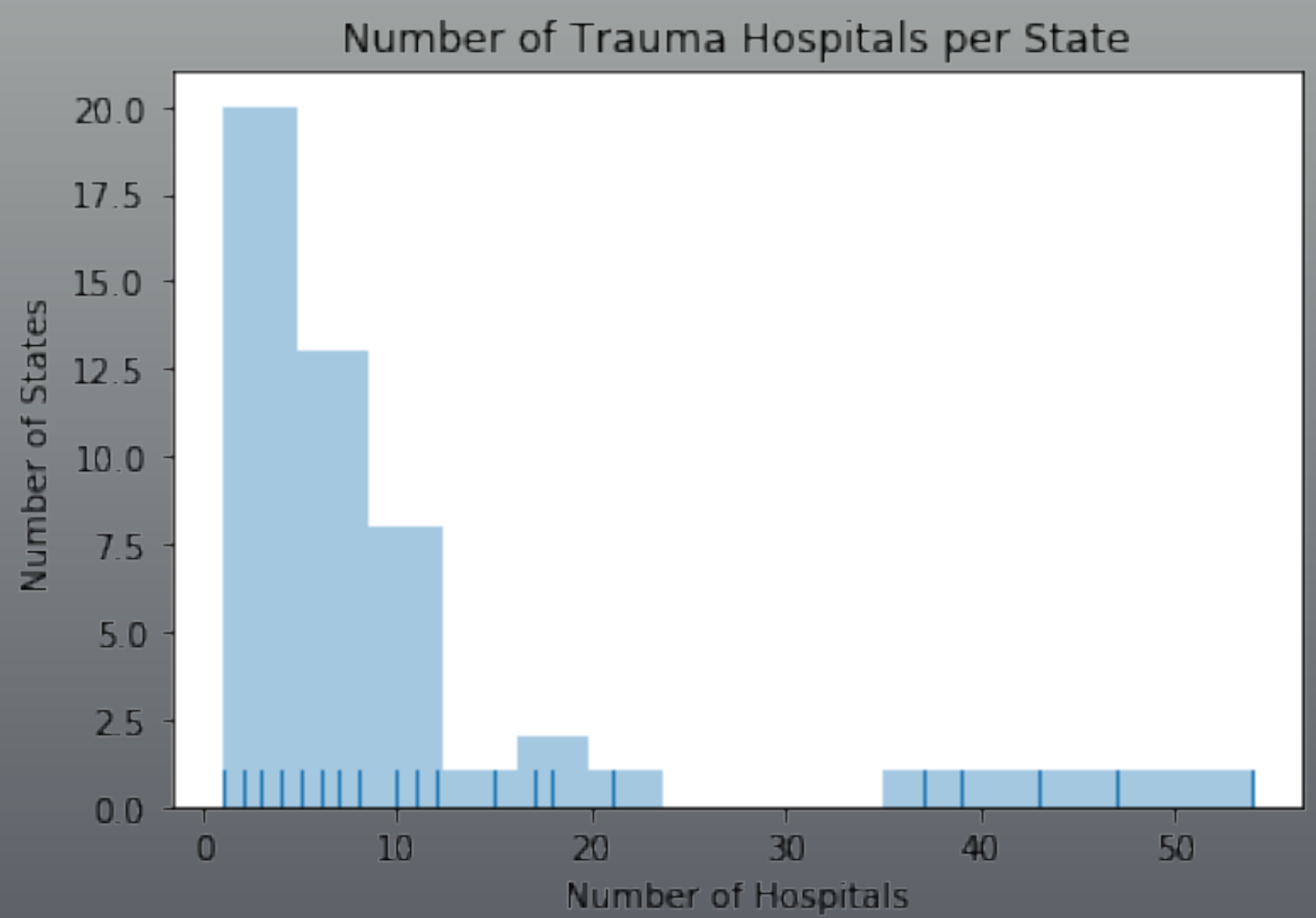
Hospital Feature - Trauma Centers

- A trauma center is a hospital equipped and staffed to provide care for patients suffering from major traumatic injuries such as falls, motor vehicle collisions, or gunshot wounds.

- Trauma centers are separated into adult and pediatric and are then ranked from level 1 being the highest and level 3 the lowest.

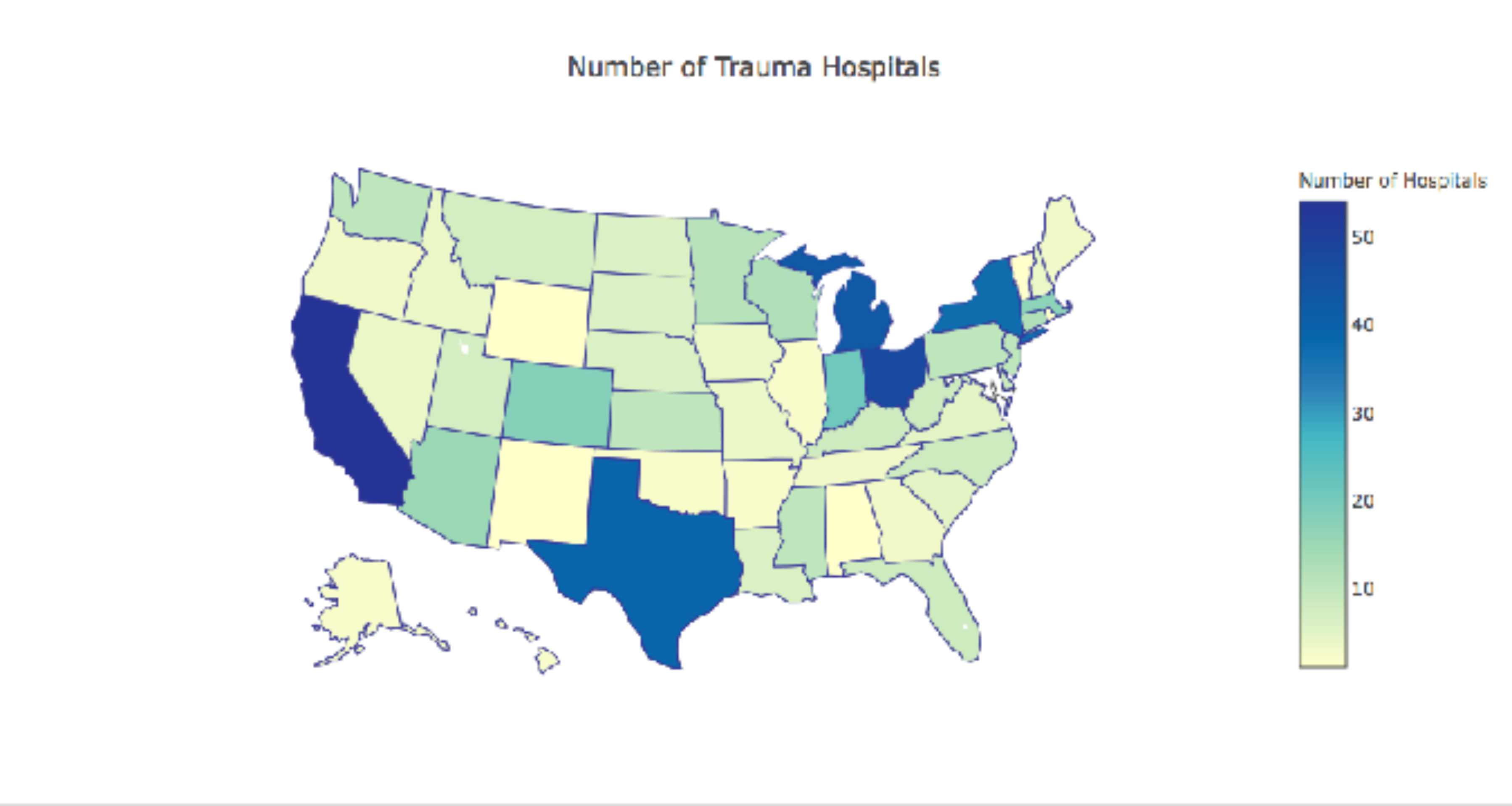
Hospital Feature - Trauma Centers

There are close to 500 hospitals that are trauma centers in the U.S.



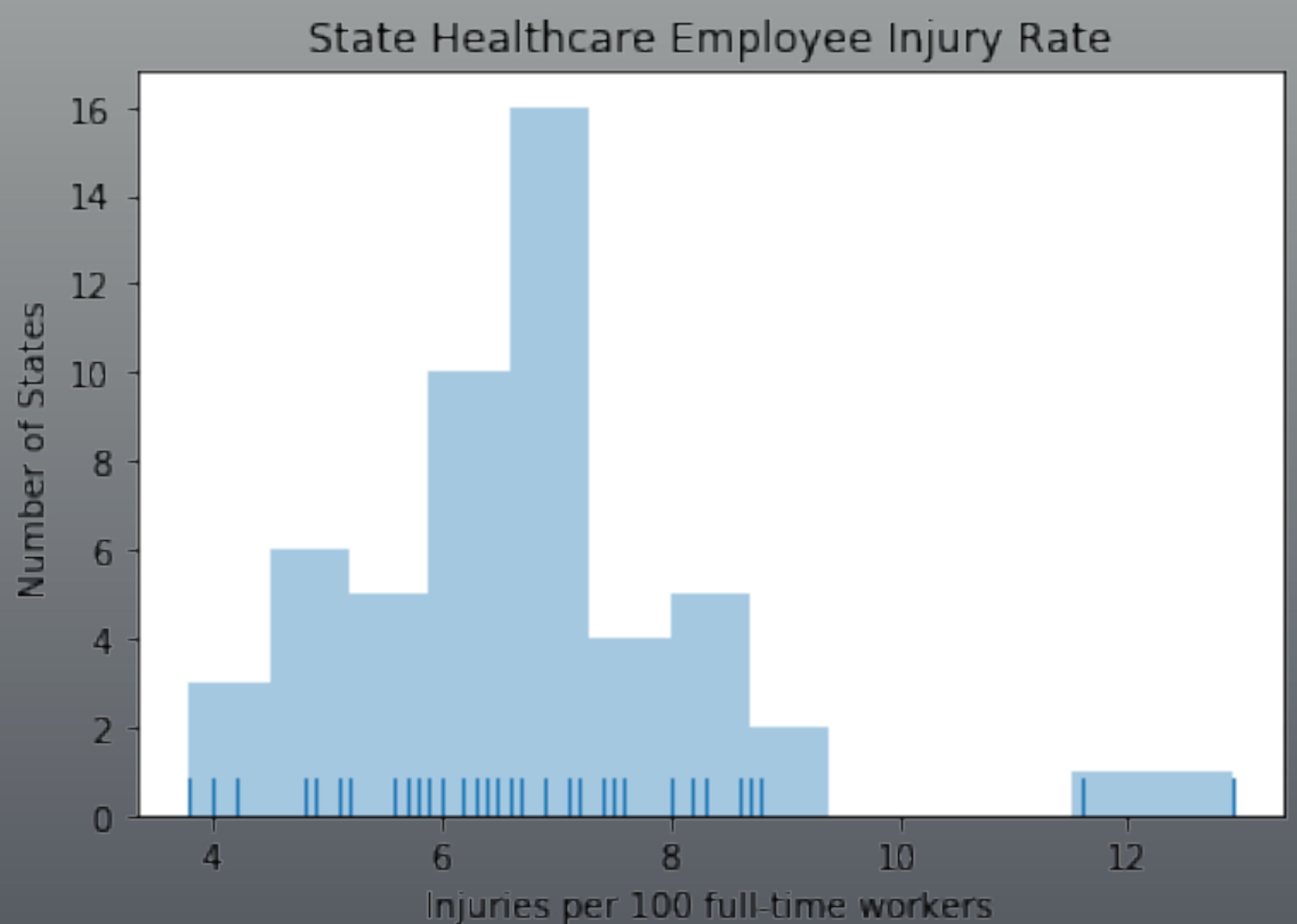
Hospital Feature - Trauma Centers

There are close to 500 hospitals that are trauma centers in the U.S.



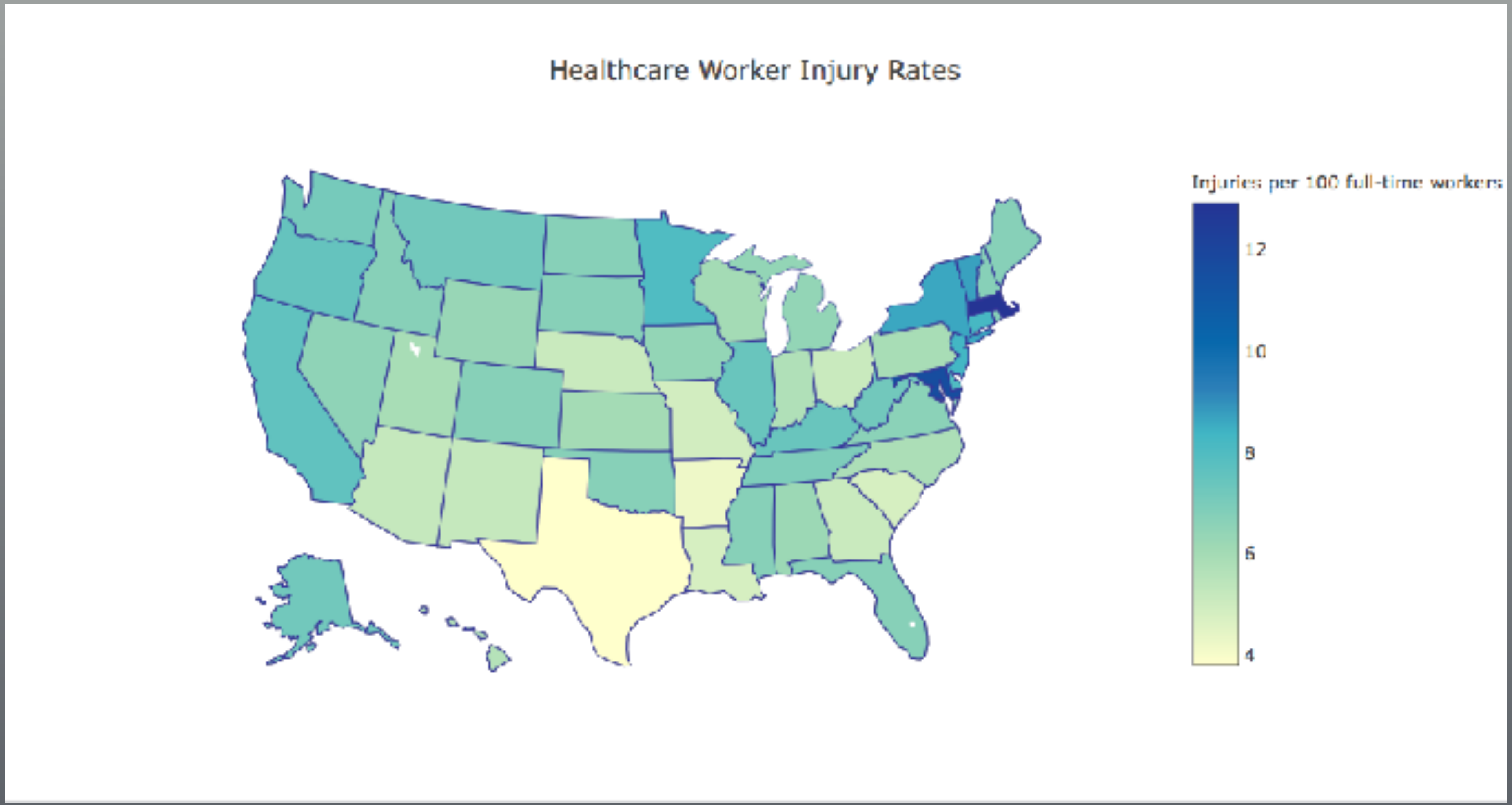
State Feature - Healthcare Employee Injury Rates

The Bureau of Labor Statistics State Occupational Injuries, Illnesses, and Fatalities



State Feature - Healthcare Employee Injury Rates

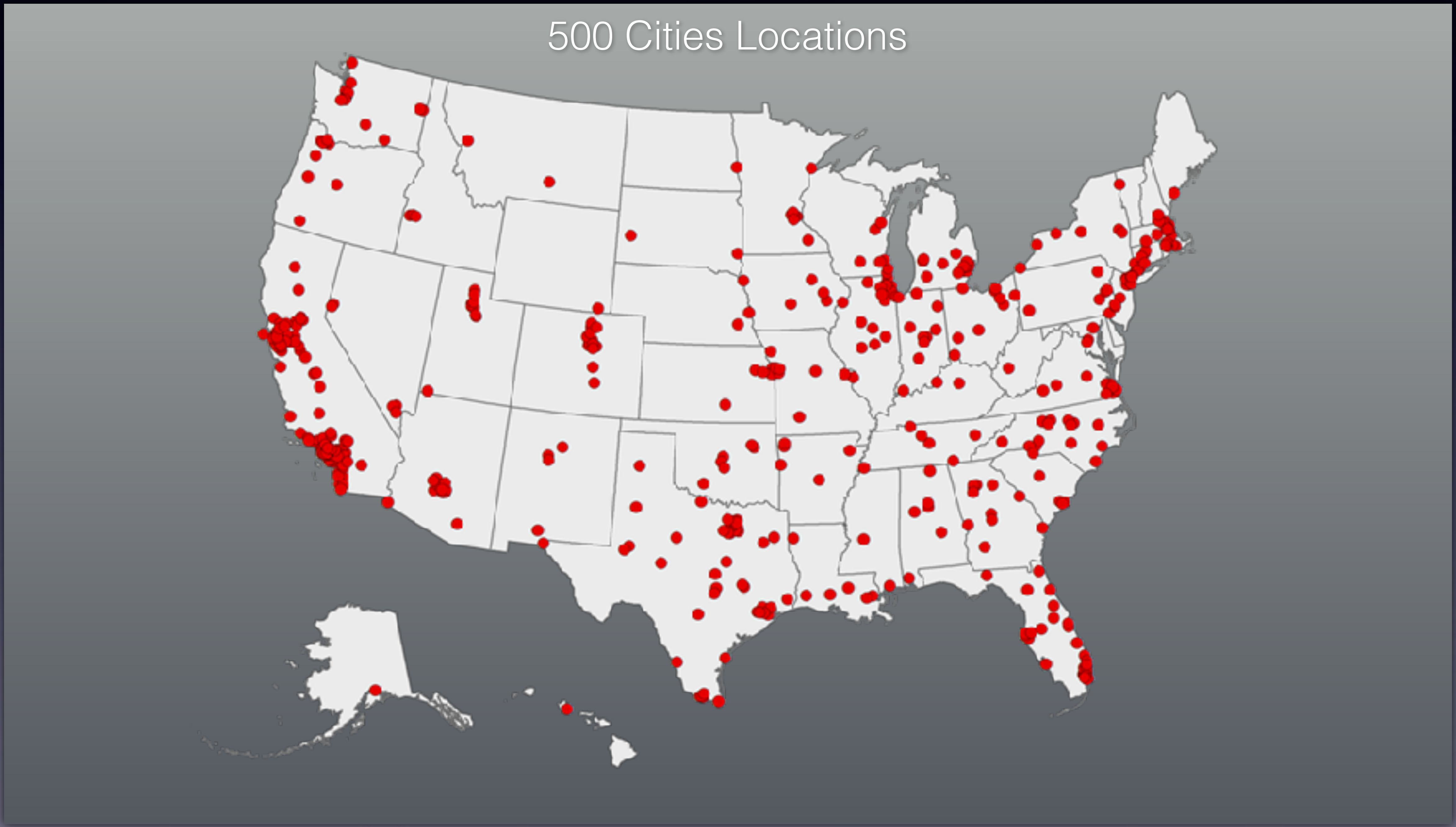
The Bureau of Labor Statistics State Occupational Injuries, Illnesses, and Fatalities



City Features - Estimates for chronic disease, health outcomes, and preventive services

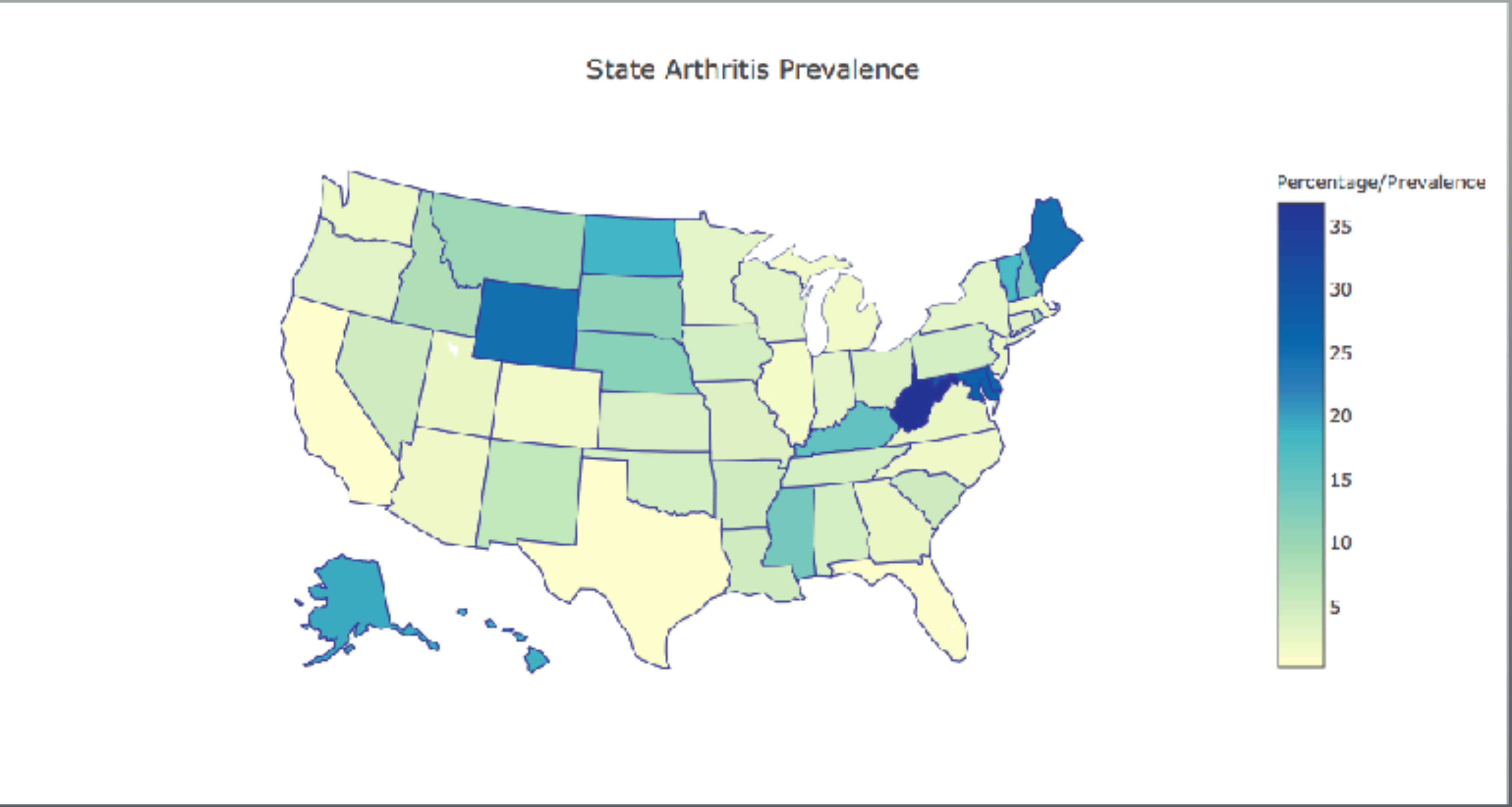
- These city features came from the Center of Disease Control's 500 Cities Project.
- This project provides small area estimates for chronic disease risk factors, health outcomes, and clinical preventive service use for the largest 500 cities in the United States.

City Features - Estimates for chronic disease, health outcomes, and preventive services

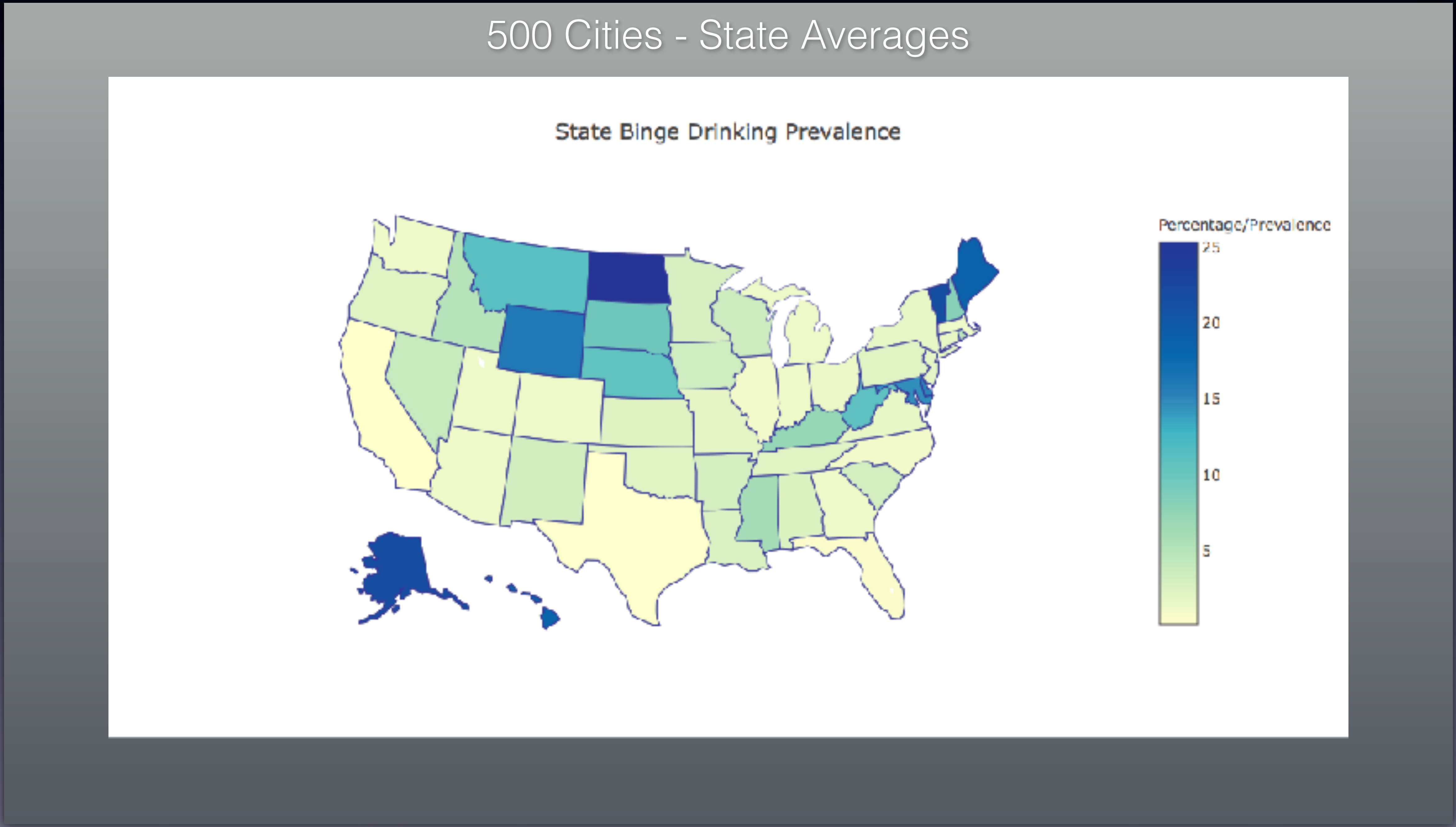


City Features - Estimates for chronic disease, health outcomes, and preventive services

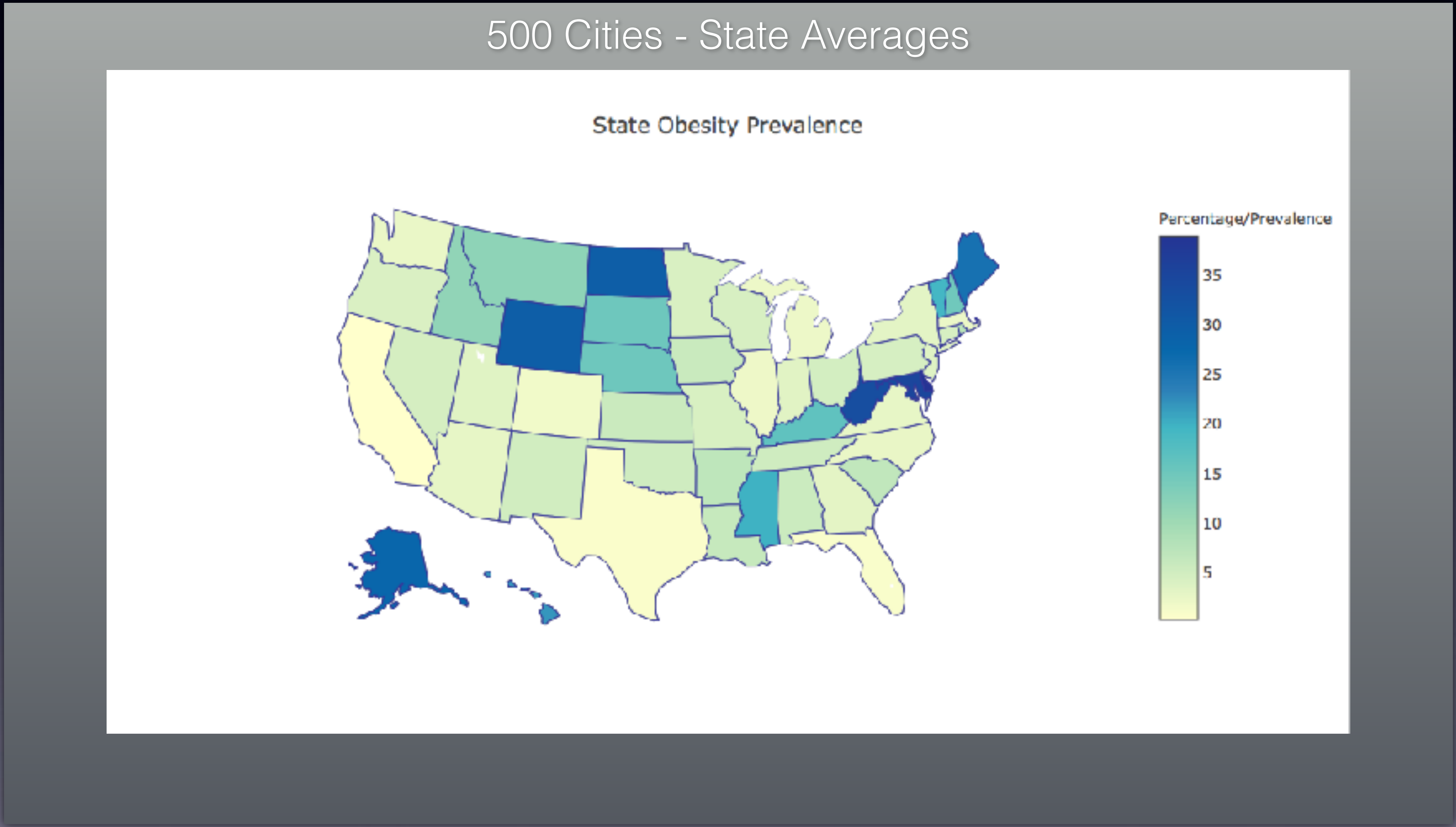
500 Cities - State Averages



City Features - Estimates for chronic disease, health outcomes, and preventive services

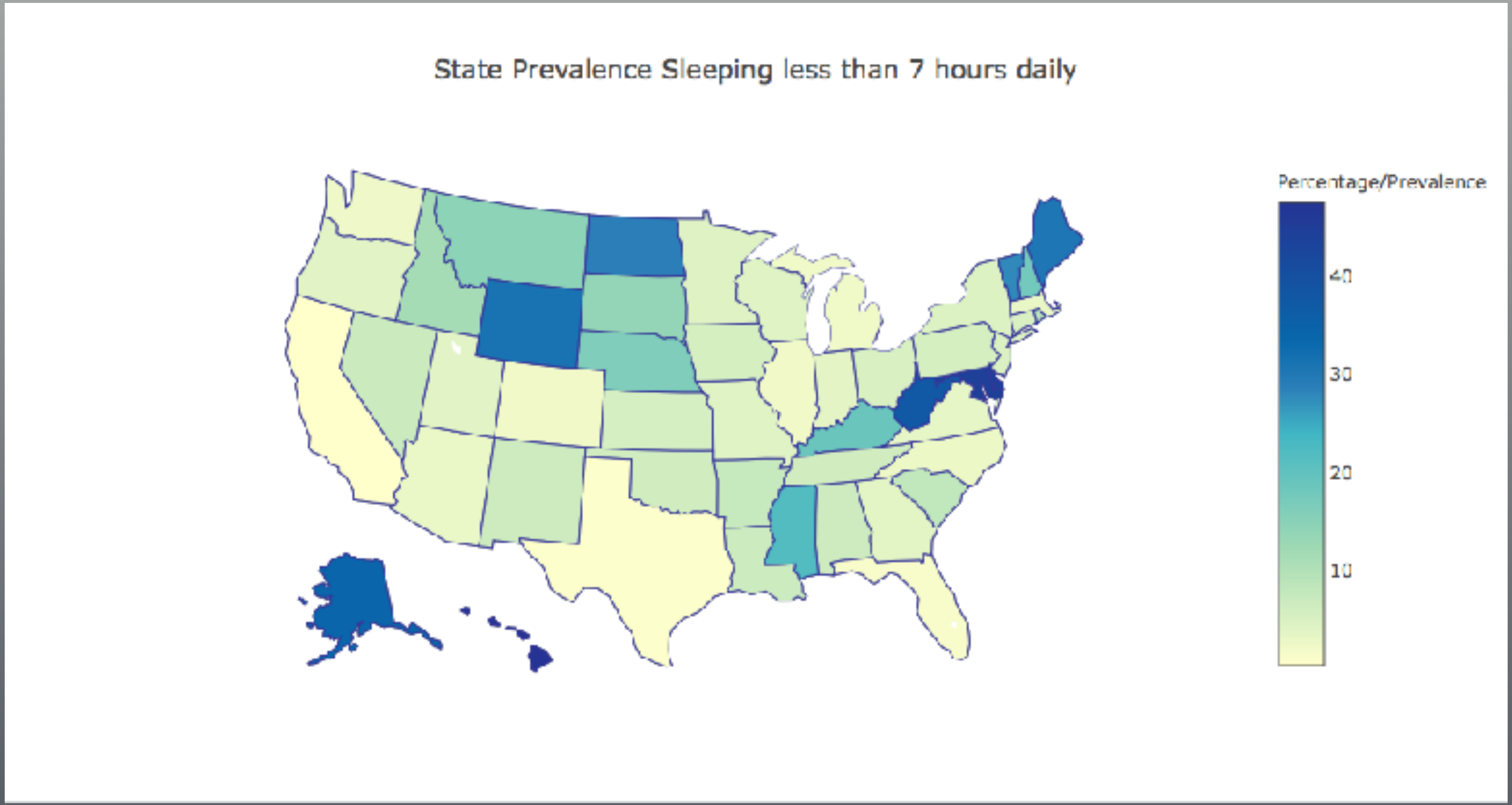


City Features - Estimates for chronic disease, health outcomes, and preventive services



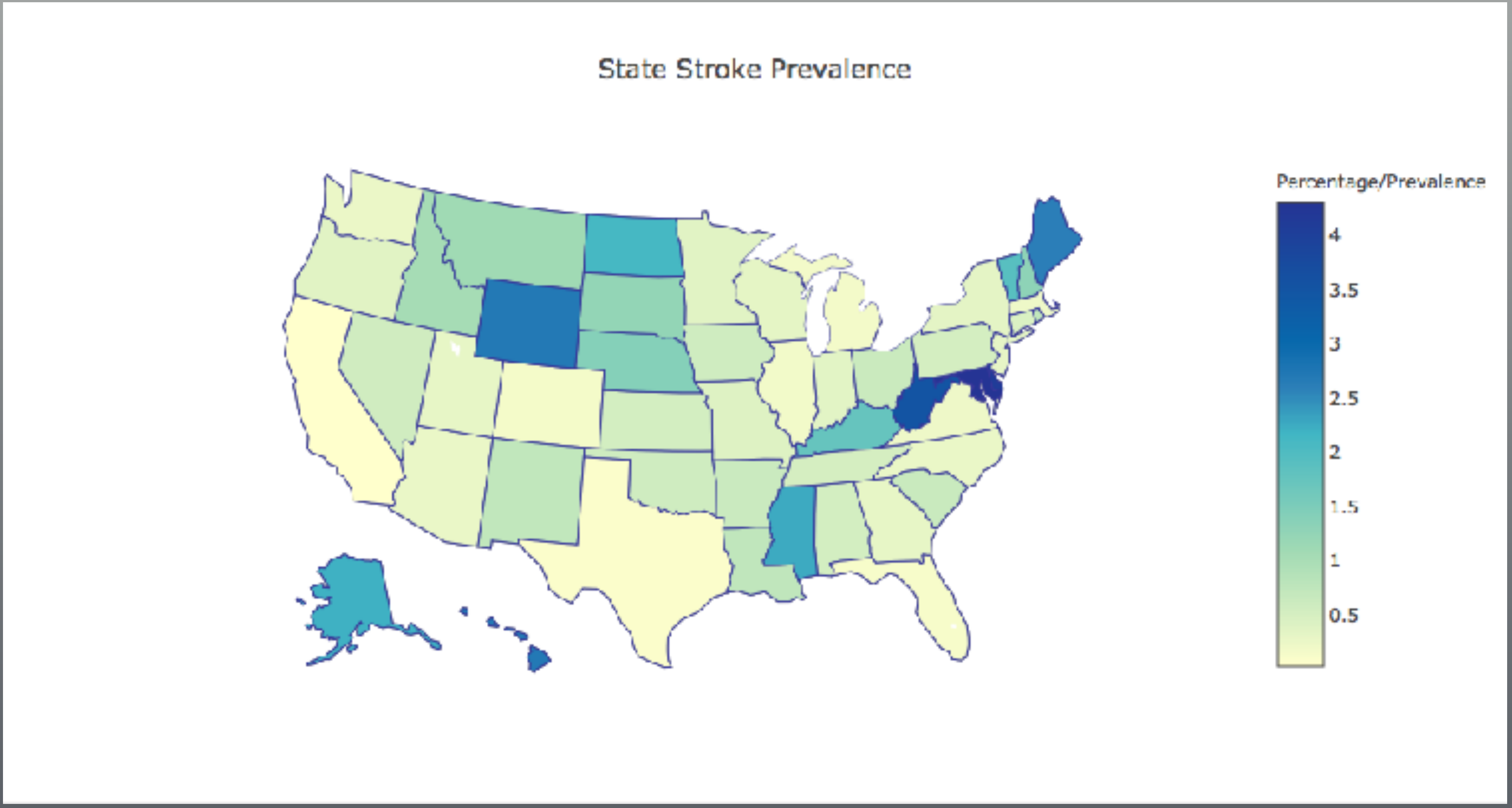
City Features - Estimates for chronic disease, health outcomes, and preventive services

500 Cities - State Averages



City Features - Estimates for chronic disease, health outcomes, and preventive services

500 Cities - State Averages



Finalized Dataset and Limiting Features

The meta-analysis fall dataset has a shape of (2069, 5).

When I merged it with each of my respective features:

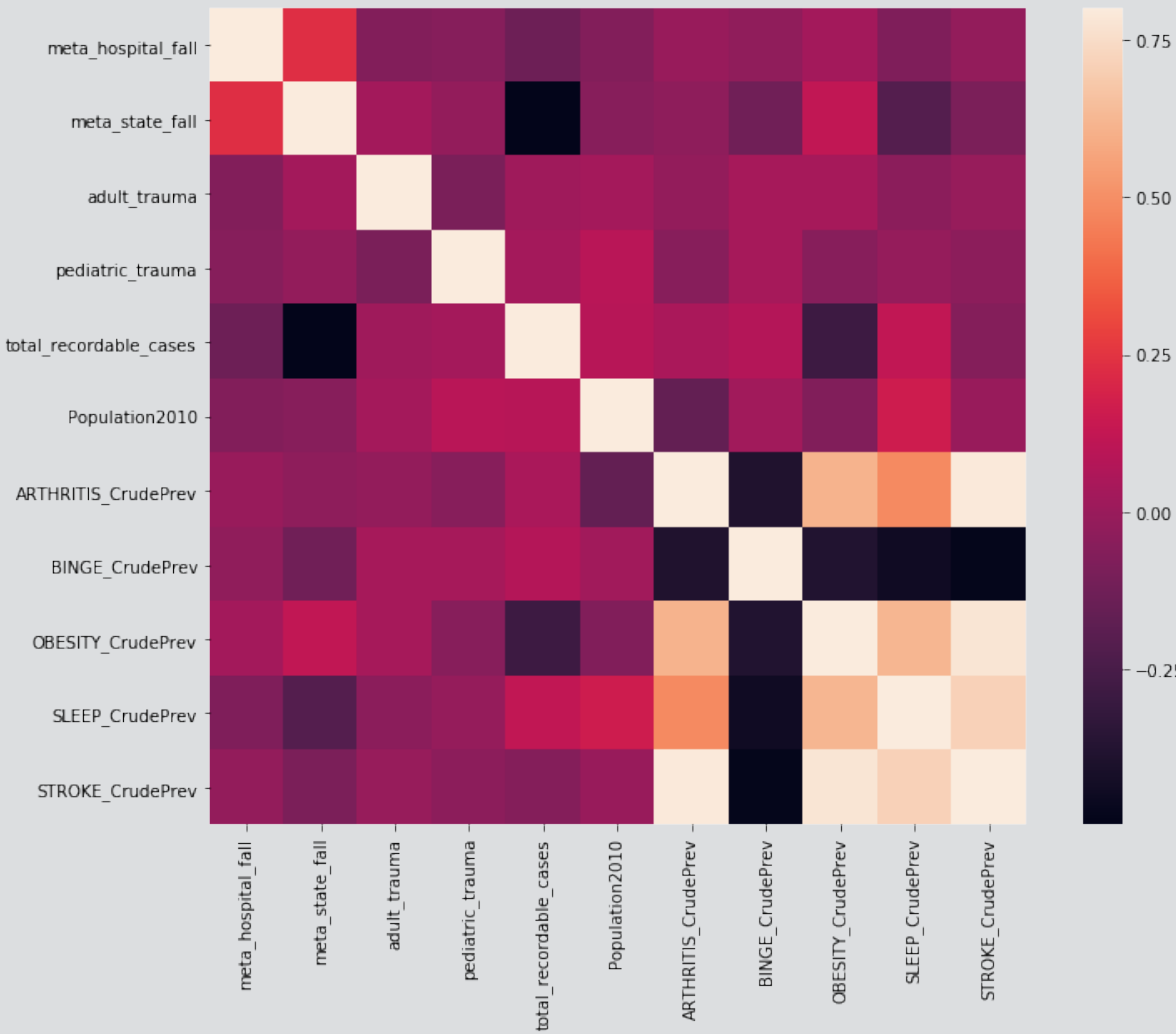
- Trauma Centers - shape: (2069, 7)
- Healthcare Employee Injury Rates - shape: (2069, 8)
- City Health Estimates - shape: (796, 14)

Explanatory Model

- My goal is to create an explanatory model that will help hospital understand what features are contributing to their fall rates.

- I will start by exploring the correlation between features and my dependent outcome, hospital falls.

Feature Correlation with Hospital Fall Rate



Explanatory Regression Models

- Regression models are particularly useful for explaining the relationship between the features and outcome.
- I decided to use both random forest regression and linear regression models.

Explanatory Regression Models

Random Forest Regression

- Training Score: 0.25
- Each Cross Validated R2 score:
  - [-0.00782244 -0.33508102 -1.01297976 -0.02362532 0.00242513]

- Overall Random Forest Regression R2: -0.28 (+/- 0.78)

- Feature Importance:
  - [('meta_state_fall', 0.16),
  - ('Population2010', 0.14),
  - ('STROKE_CrudePrev', 0.14),
  - ('SLEEP_CrudePrev', 0.11),
  - ('BINGE_CrudePrev', 0.1),
  - ('OBESITY_CrudePrev', 0.09),
  - ('adult_trauma', 0.08),
  - ('ARTHRITIS_CrudePrev', 0.06),
  - 'state_NM', 0.03),
  - ('total_recordable_cases', 0.02),
  - ('pediatric_trauma', 0.01)]

Explanatory Regression Models

Random Forest Regression

- This Random Forest Regression Model performed very poorly. It was affected by over-fitting with a training R-squared score of 0.25 but a cross-validated score of -0.28.
- I believe that this is likely from poor correlation of features with my outcome and non-normative distribution of hospital fall rates.

Explanatory Regression Models

Linear Regression

- Training Score: 0.10
- Each Cross Validated R2 score:
  - [-0.00222326 -0.23515336 -0.19107563 0.01601725 -0.03245238]
- Overall Linear Regression R2: -0.09 (+/- 0.21)

Explanatory Regression Models

Linear Regression

- The linear regression model also performed poorly but it did have less over-fitting that the Random Forest Regression model.
- Similarly, I believe that this is from the poor correlation of features with my outcome and non-normative distribution of hospital fall rates.

Explanatory Classification Models

- Classification models usually give less insight than regression models but can still provide information about features and the outcome.
- Since my regression models performed poorly, I am going to move onto classification models.
- I will be using random forest classification model and gradient boosted classifier model.

Explanatory Classification Models

- In order to create a classification model. I am going separate the hospitals into high and low fall rates and then use a classification model.
- I will change my outcome to a binary feature for high and low fall rates.
- I will set the threshold at 0.6 because the mean for hospital falls is 0.59 with the 50th percentile at 0.48.

Explanatory Classification Models

Random Forest Classification

- Training Score: 0.79
- Each Cross Validated Accuracy:
  - [0.675 0.60625 0.43125 0.5 0.47468354]
- Overall Random Forest Classification Accuracy: 0.54 (+/- 0.18)

Report:	precision	recall	f1-score	support
0	0.78	0.98	0.87	543
1	0.90	0.42	0.57	253
avg / total	0.82	0.80	0.77	796

Confusion Matrix:

([[531, 12],
[148, 105]])

Explanatory Classification Models

Random Forest Classification

- Feature Importance:
  - [('Population2010', 0.13),
  - ('ARTHRITIS_CrudePrev', 0.11),
  - ('OBESITY_CrudePrev', 0.11),
  - ('SLEEP_CrudePrev', 0.11),
  - ('BINGE_CrudePrev', 0.1),
  - ('meta_state_fall', 0.08),
  - ('STROKE_CrudePrev', 0.08),
  - ('adult_trauma', 0.06),
  - ('total_recordable_cases', 0.05),
  - ('pediatric_trauma', 0.01),
  - ('state_FL', 0.01),
  - ('state_IN', 0.01),
  - ('state_MI', 0.01),
  - ('state_NE', 0.01),
  - ('state_OK', 0.01),
  - ('state_TN', 0.01),
  - ('state_TX', 0.01),
  - ('state_UT', 0.01),
  - ('state_AK', 0.0),
  - ('state_AL', 0.0)]

Explanatory Classification Models

Random Forest Classification

- My Random Forest Classification model performed much better than the regression model
- It still has some over-fitting
- The accuracy score is good as well as the precision and recall scores.
- From my feature importance, I can see that population, arthritis, and obesity are the three most important features in this model.

Explanatory Classification Models

Gradient Boosted Classification

- Training Score: 0.85
- Each Cross Validated Accuracy:
  - [0.66875   0.625   0.40625   0.41139241 0.5   ]
- Overall Gradient Boosted Classifier Accuracy: 0.52 (+/- 0.22)

Report:	precision	recall	f1-score	support
0	0.86	0.93	0.90	543
1	0.82	0.68	0.75	253
avg / total	0.85	0.85	0.85	796

Confusion Matrix:

([[506, 37],
  [80, 173]])

Explanatory Classification Models

Gradient Boosted Classification

- Feature Importance
  - [('Population2010', 0.16),
  - ('adult_trauma', 0.12),
  - ('ARTHRITIS_CrudePrev', 0.12),
  - ('SLEEP_CrudePrev', 0.12),
  - ('OBESITY_CrudePrev', 0.11),
  - ('BINGE_CrudePrev', 0.07),
  - ('STROKE_CrudePrev', 0.07),
  - ('meta_state_fall', 0.06),
  - ('total_recordable_cases', 0.04),
  - ('pediatric_trauma', 0.01),
  - ('state_AZ', 0.01),
  - ('state_CA', 0.01),
  - ('state_CT', 0.01),
  - ('state_IN', 0.01),
  - ('state_NY', 0.01),
  - ('state_OH', 0.01),
  - ('state_RI', 0.01),

Explanatory Classification Models

Gradient Boosted Classification

- This Gradient Boosted Classification model performed best out of all the models.
- It continued to have over-fitting
- But the accuracy was consistent and it had the best precision and recall scores.
- From my feature importance, I can see that population, adult trauma, and arthritis are the three most important features in this model.

Conclusion

Gathering Data

- This Capstone Project has allowed me the opportunity to explore some of the features that may explain hospital fall rates.

- One of the challenges was that I did not have a single dataset with all my features, but rather I had to find separate datasets with the features that I believe would help in my modeling.

- Most of the features that I collected were on the city and state level, so if I had more time and resources available I would gather additional data on the hospital level.

- I would also search for features that had a higher correlation with hospital fall rates since the correlation matrix and feature importance show the overall low level of correlation with my dependent outcome.

Conclusion

Data Distribution

- When exploring hospital fall distribution, it is easily seen that most of the fall rates are close to zero.
- One explanation for this might be that some of the 'hospitals' are actually surgery centers where people leave the same day after having surgery.
- Surgery centers are commonly known to have low fall rates because people are not staying there or walking alone in the facility.
- If I had more time, I would explore the hospitals with low fall rates and attempt to removed surgery centers from the dataset because they are altering the model.

Conclusion

Meta-Analysis

- This capstone project has allowed me to learn about and use meta-analysis for the first time.
- I can see how meta-analysis is very useful for strengthening a model and avoiding a single dataset's bias by combining their information.
- In my case both of my datasets recorded fall rates at the same level, rate per 1,000 discharges. If the datasets had different measurement rates then I likely would have had to use effect size for my comparison.

Conclusion

Regression Models

- For this project I believe that the regression models highlighted primarily the weakness of my dataset.

- The random forest and linear regression models experienced over-fitting and their R-squared scores were negative. This was likely caused by the poor correlation of features with my outcome and non-normative distribution of hospital fall rates.

Conclusion

Classification Models

- The random forest and gradient boosted classification models did much better overall.
- The accuracy for both are roughly the same 0.54 and 0.52.
- This shows that these models don't perform great but the aspect that I am happiest about is their precision and recall.
- Precision is out of the total number true and false positives, how many were true positive. In both classification models, their precision was good at 0.82 and 0.85.
- Recall is out of the total number true positives and false negatives, how many were true positives. In my models, the gradient boosted classifier did better with a recall of 0.85 compared to 0.80 for the random forest classifier.

Conclusion

Future Application and Research

- I believe that analyzing the trends and modeling the distribution of hospital fall rates could greatly benefit hospitals and healthcare organizations.
- Hospitals have major ethical and financial incentives to understanding and reduce the rates and severity of falls.
- Creating explanatory models like this one could be used for other challenges that the healthcare industry is also trying to understand or reduce.
- I hope to continue seeing data science and healthcare joining together to tackle some of the world's most challenging health situations.

References

[1] Falls among Adult Patients Hospitalized in the United States: Prevalence and Trends https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3572247/

[2] Profiles of thousands of hospitals, medical clinics, nursing homes and home health centers. http://www.hospital-data.com/

[3] Hospital Level calculations for 8 Hospital Acquired Conditions (July 1, 2010 through June 30, 2012) https://data.cms.gov/Medicare-Inpatient/Selected-Hospital-Level-HAC-Rates/b5av-3pcr

[4] American College of Surgeons https://www.facs.org/search/trauma-centers?country=United%20States&n=250

[5] CDC estimates for chronic disease risk factors, health outcomes, and clinical preventive service use for the largest 500 cities in the United States. https://chronicdata.cdc.gov/500-Cities/500-Cities-City-level-Data-GIS-Friendly-Format-201/k56w-7tny

[6] Bureau of Labor Statistics State Occupational Injuries, Illnesses, and Fatalities https://www.bls.gov/iif/oshstate.htm#CA