Final Report

Maya Ghanem and Isabelle Xiong

11/15/2021

Background and Significance

America, unlike other developed countries, does not have a universal healthcare program, meaning that not all individuals are granted free health insurance by the government. In America, Health insurance is purchased in the private marketplace or provided by the government only to certain groups, such as pregnant, low income, elderly, families with children.

The members of this project have personal experiences with United States health insurance. Having lived in Canada, where all citizens can apply for a public health insurance for free which grants them access to quality healthcare, Isabelle has never had to think of what it would be like without access to healthcare. In doing this research, Isabelle wanted to understand the effects of a lack of health insurance in the US, by finding the correlation between not having health insurance and the likelihood of contracting different diseases in 500 largest US cities. Maya worked at Westminster Free Clinic for four years, which offered healthcare services to the uninsured population in Ventura County, California. They witnessed how the obstacles our patients faced in accessing healthcare impacted their behaviors and healthcare outcomes.

Data Collection and Variables

The 500 cities dataset is provided by the Center for Disease Control and Prevention (CDC), Division of Population Health, Epidemiology and Surveillance Branch. Data from this dataset is sourced from 28,000 census tracts in "Census Bureau 2010 census population data, Behavioral Risk Factor Surveillance System (BRFSS) data (2017, 2016), and American Community Survey (ACS) 2013-2017, 2012-2016 estimates". All data was collected through the form of surveys. The meta dataset features data for a total of 5 "unhealthy behaviors", 13 "health outcomes'' and 9 "preventive services" related to 27 types of chronic disease in 500 largest cities in the US. Within the datasets for diabetes, coronary heart disease, mental health outcomes and health insurance, variables include city, state, state abbreviation, year, datasource, TractFIPS, CityFIPS, Geographic Level, data value, low confidence unit, high confidence unit, coordinate of geographical location of city.

The original dataset includes about 29,000 observations, with multiple measurements for each city based on the data source (census, BRFSS, ACS). We restricted our dataset to only include variables on percent of city with lack of insurance, visiting the doctor, taking high blood pressure medications, smoking, reporting binge drinking, not having physical activity, with heart disease, with diabetes, and with kidney disease. We took the mean of all the age-adjusted prevalence measurements within a city and filtered our dataset to these means. As a result, there was only one measurement per city. For ANOVA testing, we filtered our dataset to only include states with measurements for at least 10 cities and created logarithmic versions of all variables.

Research Questions

1) Do cities with a greater lack of healthcare access have poorer mental health and/or physical health outcomes?

2) Does healthcare access, mental health, and/or physical health outcomes vary by state?

Variables of Interest

1) Healthcare Access for Adults (18+): Percent of City Population that Lacks Insurance, Percent of City Population with visits to doctor for routine checkup within the past year, Percent of City Population who have high blood pressure and are taking medicine for high blood pressure control.

2) Geographic Distribution by State

3) Behavior for Adults (18+): Percent of city population currently smoking, percent of city population currently reporting binge drinking habits, percent of city population reporting No leisure-time physical activity

4) Health Outcomes for Adults (18+): Percent of city population with coronary heart disease, percent of population diagnosed with diabetes, percent of city population with kidney disease

For Research Question 1, Healthcare Access Variables (1) are the explanatory variables, whereas Behavior (3) and Health Outcomes (4) are the response variables. For Research Question 2, Geographic Distribution by State (2) is the explanatory variable, and all health indicators (1, 3, 4) are the response variables.
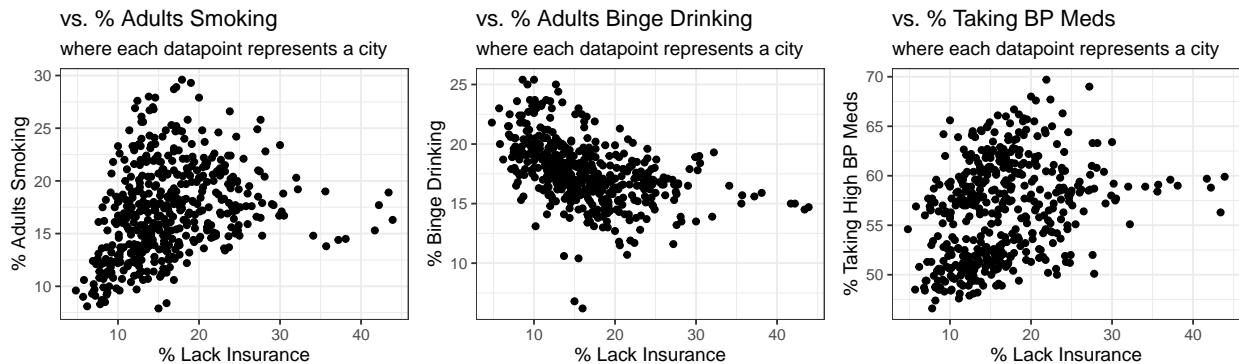
Exploratory Data Analysis

Research Question 1

The following visuals are scatter plots between the insurance (the main explanatory variable we want to focus on) and a behavior or health outcome variable.
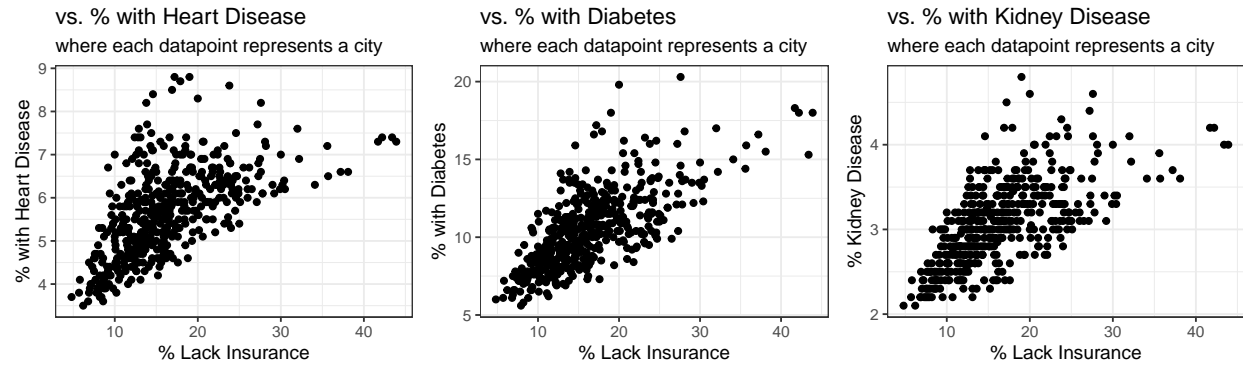
Figure 1: % Lacking Insurance vs. Behavior and Health Outcome Variables

`## $`1``
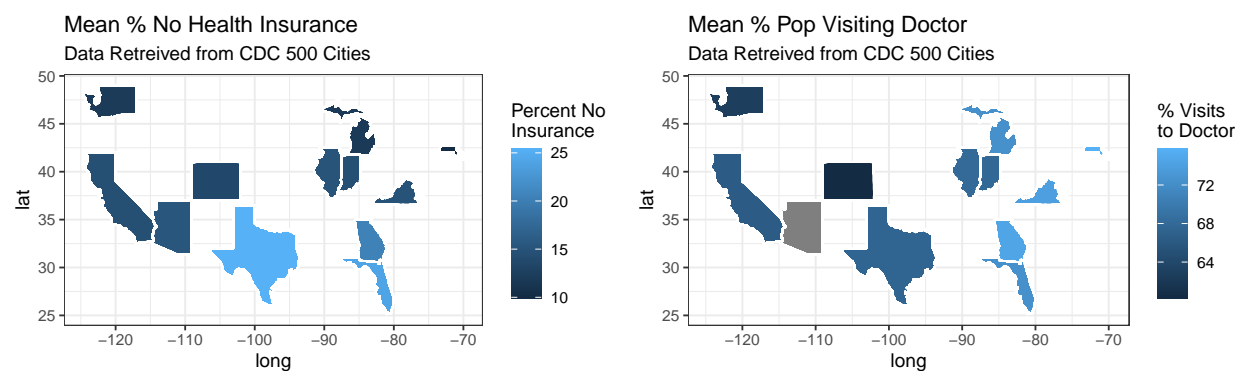


```
##
## $`2`
```

```
##
## $`3`
```

```
##
## attr(,"class")
## [1] "list"        "ggarrange"
```
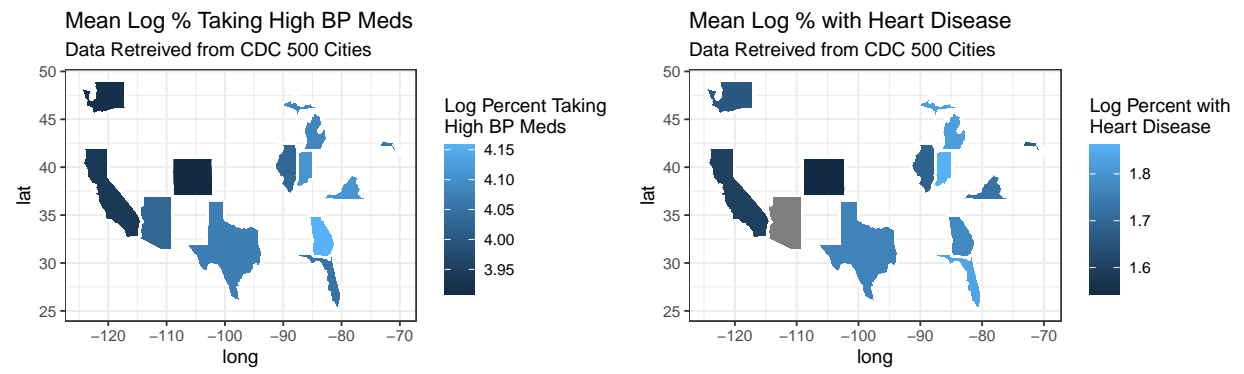
Research Question 2

The maps below visualize the distribution healthcare access, behavior, and outcome variables by state. Only states that have at least ten observations are included, and only four variables that fit the assumptions for ANOVA testing.

Figure 2: State Distribution of Healhcare Access, Behavior, and Health Outcome Variables

```
## $`1`
```



```
##
## $`2`
```

```
##
## $`3`
```

```
##
## attr(,"class")
## [1] "list"          "ggarrange"
```