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.

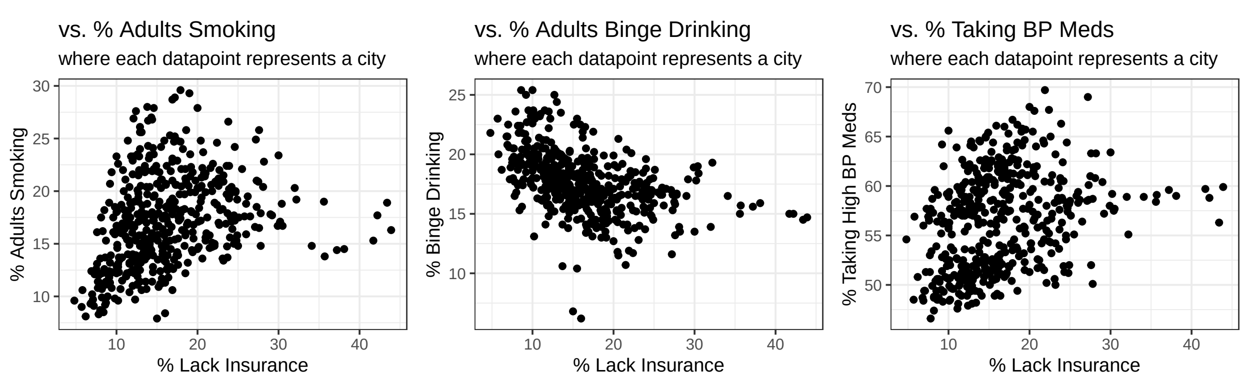
Exploratory Data Analysis

Research Question 1

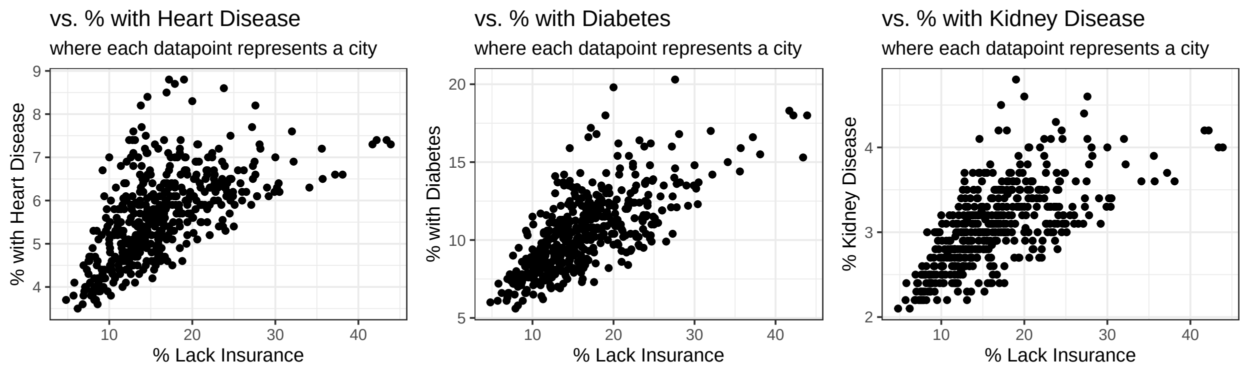
The following visuals are scatter plots between the insurance (the main explanatory variable we want to focus on) and a behavior or health outcome variable.

Figure 1: % Lacking Insurance vs. Behavior and Health Outcome Variables

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




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




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

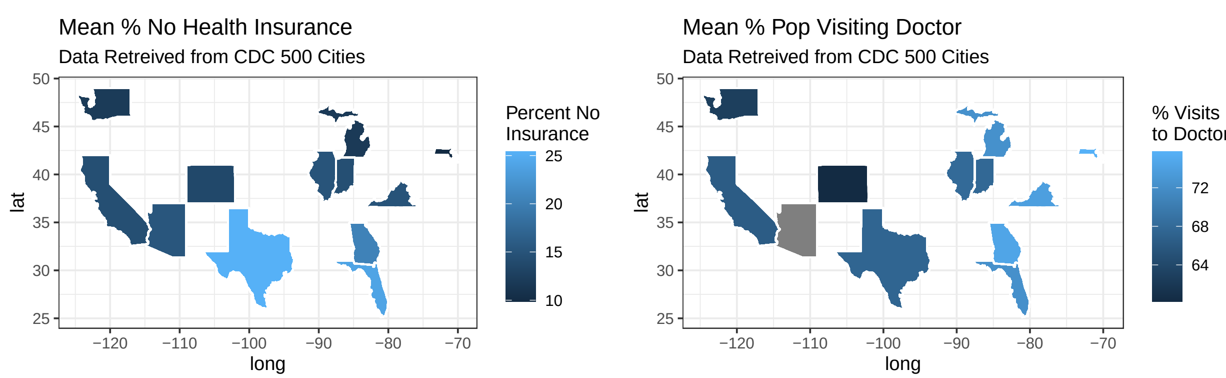
```
##
## attr(,"class")
## [1] "list"      "ggarrange"
```

Research Question 2

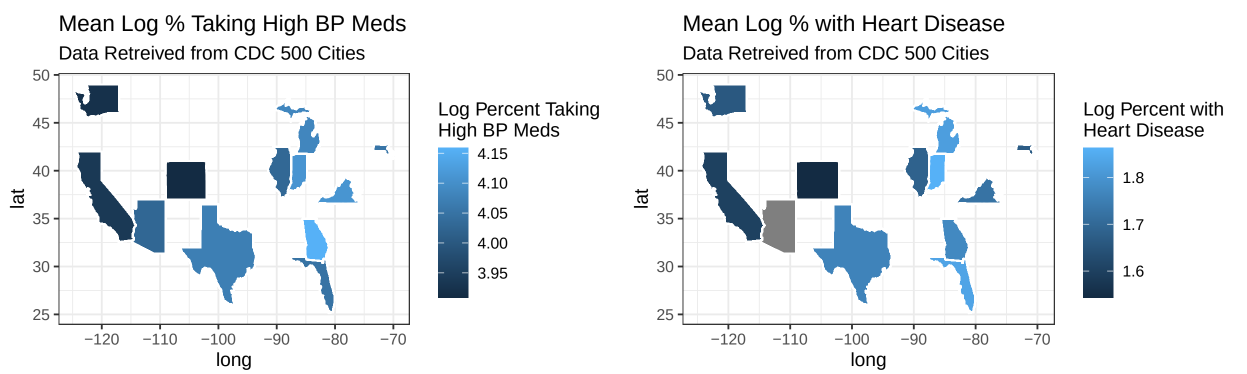
The maps below visualize the distribution healthcare access, behavior, and outcome variables by state. Only states that have at least ten observations are included, and only four variables that fit the assumptions for ANOVA testing.

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

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




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




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

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

Analytic Methods

Research Question 2

We conducted ANOVA testing to determine if there was variance within and between states for the healthcare access, behavior, and health outcome variables. After first checking for ANOVA test assumptions, we determine that state distributions of % lack of insurance, % visiting the doctor, log % with heart disease, and log % taking high BP medications had normal distributions within groups and homoscedastic variance. Since it is highly unlikely that the samples are independent, we employed a Bonferroni correction with a subsequent significance level of $p = 0.000641025641$.

Results

ANOVA tests provided evidence that there is significant variances between and within states for % lack of insurance, % visiting the doctor, log % with heart disease, and log % taking high BP medications. The step down tests demonstrated that there are 19 significant pairs for % lack of insurance, 47 significant pairs for % visiting doctor, 39 significant pairs for log % taking high BP meds, and 7 significant pairs for log % with heart disease. Florida and Texas had the most significant pairs (9 each) for the insurance step down tests, Colorado and Washington had the most significant pairs (10 each) for the visits to doctor step down tests, California, Colorado, Georgia, and Washington (9 each) had the most significant pairs for the log medicine high BP step down tests, and Colorado had the most significant pairs (3) for the log % heart disease step down tests. The only step-down test that had states with no significant pairs was for log % heart disease.

Discussion

Research Question 2

The ANOVA testing demonstrated that there is variance within and across states for lack of insurance, visits to doctor, heart disease, and taking High BP medication rates. Based on their higher F-squared values and greater number of significant pairs, visits to doctor and High BP Medication rates seemed to have more variance than lack of insurance and heart disease.

These results are significant in supporting Zuckerman's article on how there are differences in insurance coverage across states, with California, Texas, and Colorado having disproportionately high uninsurance rates (Zuckerman 1999, 8). These variances for health insurance correspond with state based variances for doctors' visits, heart disease prevalence, and people taking high blood pressure medications. Although the ANOVA testing does not depict which states have disproportionately high or low insurance rates, the fact that California, Texas, and Colorado commonly appear in significant pairs for the step-down tests supports Zuckerman's results.

A Bonferroni correction is implemented on the ANOVA testing because it is unlikely that the state distributions are independent from each other, but the research is limited in identifying what exactly causes a lack of independence. Identifying this cause could allow one to correct the lack of independence more accurately.

Further research should investigate the socioeconomic and policy factors that contribute to differences between and within states. Are the differences due to differing statewide healthcare policies or the result of inherent economic inequality between states?

References

https://chronicdata.cdc.gov/500-Cities-Places/500-Cities-Local-Data-for-Better-Health-2019-relea/6vp6-wxuq

https://www.urban.org/sites/default/files/publication/66251/309311-Snapshots-of-America-s-Families.PDF

https://www.annualreviews.org/doi/abs/10.1146/annurev.publhealth.28.021406.144042

Appendix

Overall ANOVA Tests

State Distribution of People Lacking Insurance

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## StateDesc    11   7449   677.2    25.2 <2e-16 ***
## Residuals   297   7982    26.9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

State Distribution of People Visiting Doctor

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## StateDesc    11   3780   343.6   90.03 <2e-16 ***
## Residuals   296   1130     3.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 1 observation deleted due to missingness
```

State Distribution of People Taking High BP Meds

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## StateDesc    11  1.627 0.14794   113.3 <2e-16 ***
## Residuals   297  0.388 0.00131
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

State Distribution of People with Heart Disease

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## StateDesc    11  2.834 0.25760   10.33 5.44e-16 ***
## Residuals   296  7.384 0.02495
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 1 observation deleted due to missingness
```