Gregory Matesi
Dr French
October 20, 2021

Project Proposal

Name: Greg Matesi

What data set do you want to analyze? Where did you obtain the data?
https://data-cdphe.opendata.arcgis.com/datasets/CDPHE::delayed-medical-care-in-adults-cdphe-community-level-estimates-census-tracts/about
Delayed Medical Care in Adults ($) - CDPHE Community Level Estimates (Census Tracts)
Obtained from the Colorado Department of Public Health and Environment (CDPHE)

This dataset has the percentage of adults who have delayed healthcare because of cost in the last 12 months for each census tract in Colorado. In also has the adult population for each census tract.

File formats:

1. cpg
2. dbf
3. prj
4. shp
5. shx
6. gal (made by me)
7. nb (made by me)

How up to date is it? The estimates are statistically modeled based on data from 2014-2017. The data file itself was updated in February 2019. However, I am not sure at this point if that update reflects current information for 2019 (i.e. they trained their model on 2014-2017 data and applied the model to produce estimates for 2019 data?).

How was it made? The estimates were based on statistical models trained on a number of healthcare related variables including age, race, gender, poverty, education, location and health conditions or risk behavior indicators.

"The estimates are based on statistical models and are not direct survey estimates. Using the best available data, CDPHE was able to model census tract estimates based on demographic data and background knowledge about the distribution of specific health conditions and risk behaviors."
https://data-cdphe.opendata.arcgis.com/datasets/CDPHE::delayed-medical-care-in-adults-cdphe-community-level-estimates-census-tracts/about

What kinds of auxiliary information are available for the data? How many variables does the data have? How many observations?
Observations are 1249 Census tracts for Colorado. I may narrow my analysis down to just Denver metro area (698 census tracts).

The data has the following variables for each census tract:

1. percent of adults that delayed medical care because of cost by census tract. 95 % confidence intervals.
2. County estimate with 95% confidence intervals.
3. State estimate with confidence intervals.
4. Population (adults 18+) for each census tract.

Gregory Matesi
Dr French
October 20, 2021

In addition to the above variable of interest, I could potentially merge similar census tract data for:

1. Percent of Adults with no medical checkout in the past year
2. Percent of Adults who are Heavy Drinking
3. Percent of Adults who have Health Insurance Coverage of any kind
4. Percent of Adults who currently Smoke Cigarettes
5. Percent of Adults who are Binge Drinking
6. Percent of Adults who experience more than 14 Physically Unhealthy days within the past 30 days
7. Percent of Adults who used Marijuana or Hashish 1 or more days out of the past 30 days
8. Percent of Adults ever diagnosed with Diabetes by a doctor, nurse, or other health professional

What statistical research question(s) do you hope to answer? What model(s)/methods do you initially plan to fit to answer the research question(s)? What will your response variable be?
I want to examine if there are certain areas of Colorado/Denver where more adults seem to be delaying medical care because of cost.

I initially believe that the Spatial Scan method or the Walter (1992) Modified Moran's I method will be good methods to apply to this analysis. I have attached an initial plot for the Spatial Scan method below.

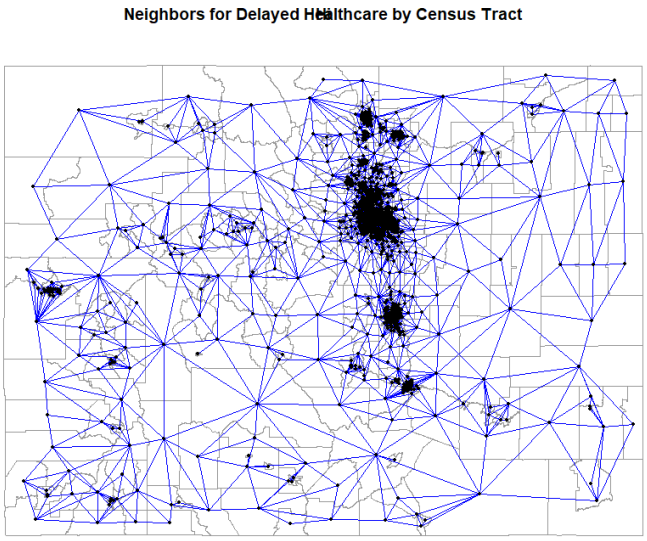
This analysis will potentially inform policy on creating public outreach and education programs targeting individuals in these areas with the goal of helping them find cost effective healthcare.

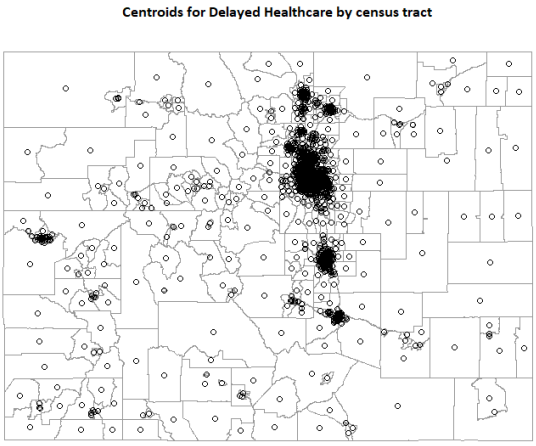
The response variable will be the number of adults who delayed health care because of cost "in the last 12 months" (during 2014-2017 or for February 2019?). I have successfully multiplied that percentage or adults who delayed healthcare by the adult population to obtain this response variable.

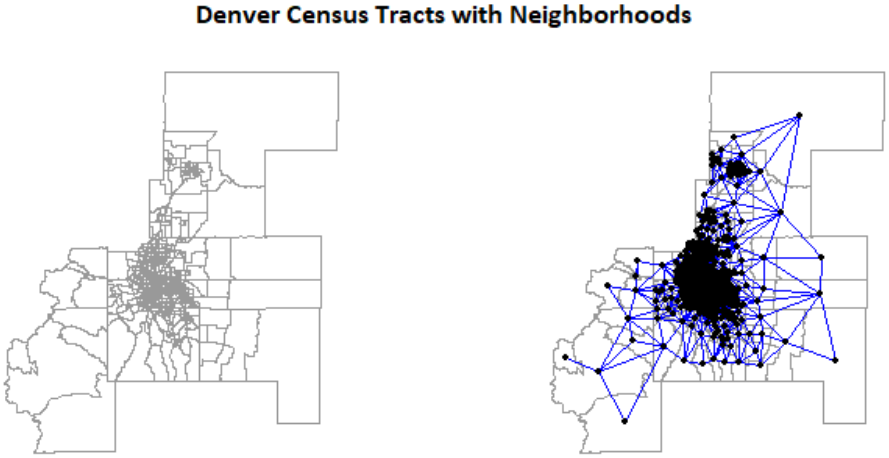
What is the title of your project (this should be fairly short and relate to your research question(s)
Intervention for Delayed Healthcare Due to Cost in Colorado/Denver

Provide an exploratory plot of your data.
Below is a plot of the neighbors for the census tract data and the centroids used. On the next page is a similar neighbors plot for the Denver metro area.

Gregory Matesi
Dr French
October 20, 2021







Summary Statistics

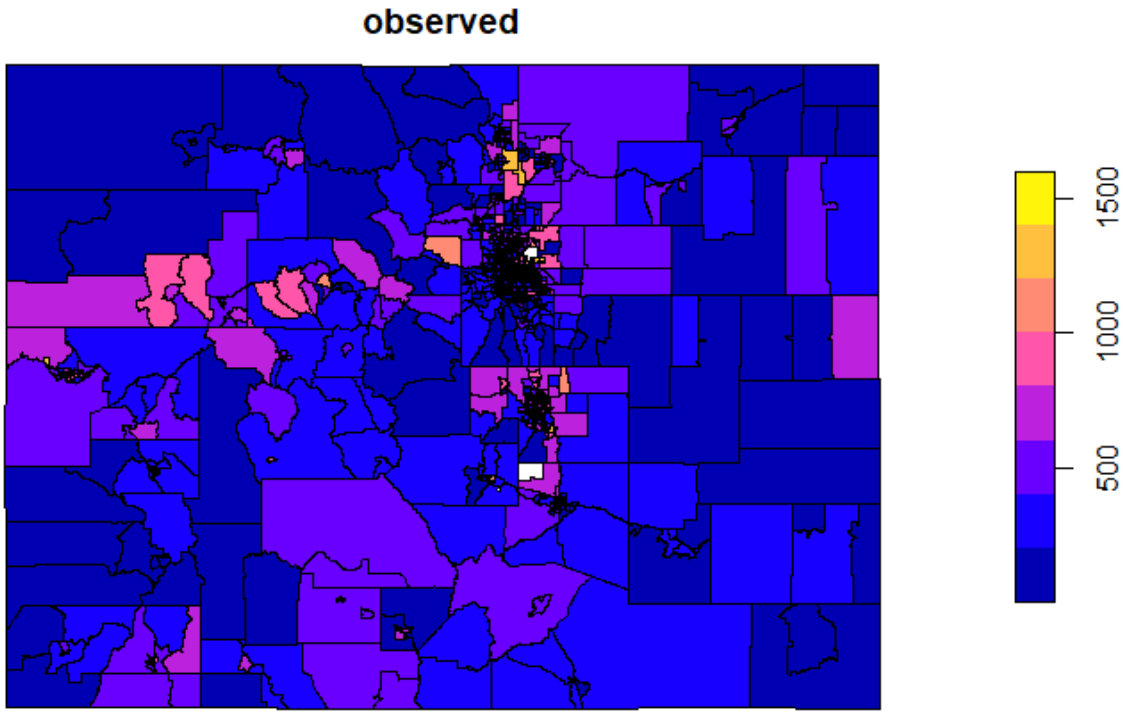
Below is a table of summary statistics for the dataset limited to census tracts in the ten Denver metro area counties.

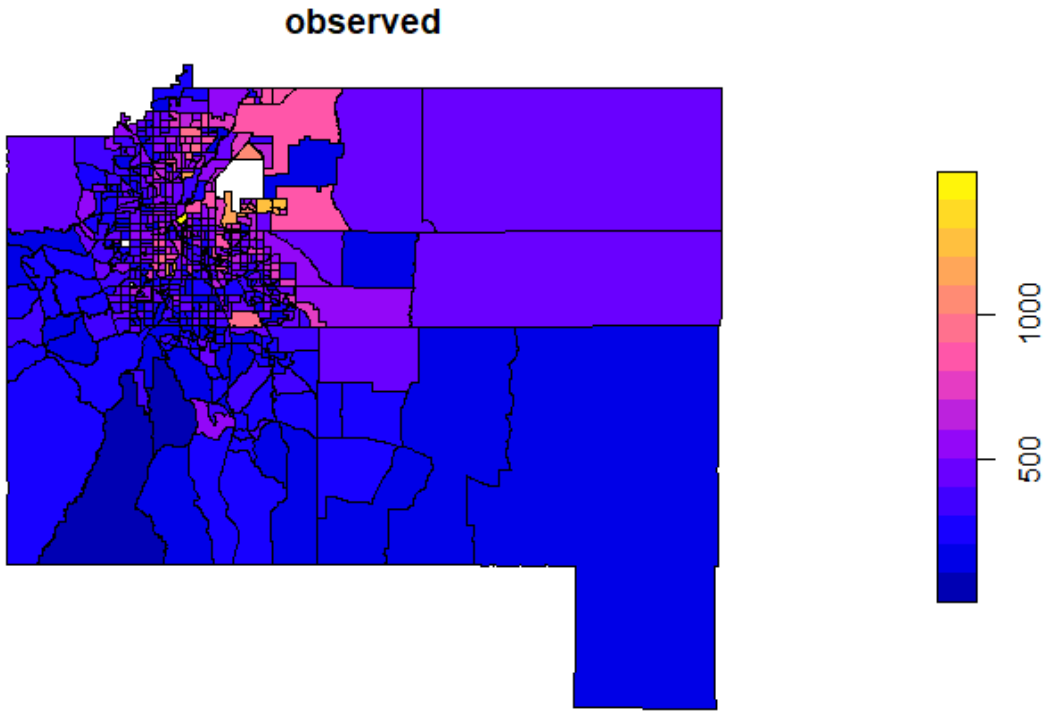
Variable	Adult population	Percet adults who delayed healthcare	Number of adults who delayed healthcare
Min	249	3.80 %	31.12
Mean	3397	12.33 %	421.47
Median	3257	12.40 %	388.23
Max	9803	22.20 %	1421.43

7 census tracts were removed from the original data because they reported NA for the percent of adults who delayed healthcare due to cost. This was because those census tracts had less than 50 individuals estimated to have delayed healthcare. 6 of these census tracts are in the Denver metro area and have been left out of these summary statistics.

Gregory Matesi
Dr French
October 20, 2021

I have included choropleth maps below of the observed response variable for Colorado and for Denver counties excluding Weld, Park, Gilpin, and Clear Creek counties.





Gregory Matesi
Dr French
October 20, 2021

And finally, I have initial results implementing the spatial scan method with default function settings. More analysis is needed to find out how to best use this method.