Matthew Blake

11/5/2021

A5: Extension Plan

**Motivation Statement**

The COVID-19 pandemic has taken its toll on many Americans during these last two years. This toll is especially apparent when you examine the impact Covid-19 has had on employment and businesses, particularly small, local businesses. As we know, the pandemic had a disastrous impact on business hiring practices and operations, resulting in many businesses shutting down permanently. Financial security became a primary concern for most working Americans as unemployment skyrocketed during this period. However, not all industries were impacted in a similar fashion; some shut down due to the economic turmoil, while others continued to thrive in spite of COVID-19 outbreaks. Given this information, it would be interested to determine which groups of people have been impacted the most by this pandemic. With this query in mind, I’ll be extending the figure from A4 to reflect which demographics of residents were (and still are) impacted the most because of the Covid-19 pandemic. If possible, I’d also like to assess whether masks influenced the rate of Covid-19 contraction among these vulnerable groups; if most neighborhoods in the St. Louis County didn’t impose a mask mandate, then I’d like to compare the contraction rates among these groups to similar groups in a separate county where the mask mandate was imposed. Either way, such a visualization will not only provide insight as to whether masks are effective in preventing the spread of COVID-19 among these vulnerable groups.

**Research Questions & Hypotheses (Q: Question, H: Hypothesis)**

Q1: How did the COVID-19 pandemic affect the economy for various age groups at the local level in St. Louis County?

H1: Due to their lack of professional experience, recent college graduates and younger workers are expected to get hit the hardest by the pandemic compared to older, more experienced professionals.

Q2: How did the economic impact of the COVID-19 pandemic differ between rural and urban regions?

H2: Given that urban areas tend to be more populous than rural areas, they are expected to have higher rates of covid infections compared to rural areas, and thus, higher layoffs & business foreclosures.

Q3: How did the economic impact of the COVID-19 pandemic differ between various age groups?

H3: Due to their lack of professional experience, recent college graduates and younger workers are expected to get hit the hardest by the pandemic compared to older, more experienced professionals.

Q4: How did the COVID-19 affect the various racial demographics in this county?

H4: Given that no one particular race is more susceptible to contracting COVID-19, there should be no significant differences between the contraction rate between various racial groups in the county.

**Data that may be included in analyses (tentative to change)**

* Covid19\_deaths\_demographics: spreadsheet containing the Covid-19 deaths by sex, age, and race demographics in St. Louis County, Missouri.

Source: https://fred.stlouisfed.org

* Covid-19-open-data (aggregated.csv.gz): Covid-19 epidemiological database by the Google Cloud Platform containing data relating to demographics, economy, epidemiology, geography, health, hospitalizations, and government response. This is one data source that I can use to conduct my cross-demographical analyses.

Source: <https://github.com/GoogleCloudPlatform/covid-19-open-data>

License: https://github.com/GoogleCloudPlatform/covid-19-open-data/blob/main/LICENSE

* Raw US Confirmed Cases: contains the confirmed cases for states & counties throughout the U.S. from the Kaggle repository of John Hopkins University COVID-19 data.

Source: <https://www.kaggle.com/antgoldbloom/covid19-data-from-john-hopkins-university?select=RAW_us_confirmed_cases.csv>

* Masking\_mandates\_by\_county: U.S. state and territorial public mask mandates from April 10, 2020 through August 15, 2021 by county by day, provided by the Centers for Disease Control and Prevention (CDC).

Source: <https://data.cdc.gov/Policy-Surveillance/U-S-State-and-Territorial-Public-Mask-Mandates-Fro/62d6-pm5i>

* Covid-19\_US\_County-level\_Summaries: This data contains the county-level socioeconomic data for predictive modeling of epidemiological effects for the United States. I’ll be using this data as a compliment to the Covid-19 open data provided by the Google Cloud Platform; I intend use this data to create models and graphs that will serve to reinforce my arguments.

Source: <https://github.com/JieYingWu/COVID-19_US_County-level_Summaries>

**Unknowns & Dependencies**

* Many of the datasets I’ll be using span different timeframes during the pandemic, not the entirety of the pandemic itself. Therefore, I may experience some difficulties concerning consistency when examining timeseries or comparing the case rates between various demographics over time.
* As I can tell, the datasets that I will be using categorize Covid patients into various age groups, primarily ranging from 10-85+ years. I won’t be able to perform any trend analyses or group comparisons with patients who don’t fall into that 10-85 year-old range.
* The datasets that I’ve included in this plan categorizes patients into three racial categories: white, black, and other. Here I’m constrained as I expected that ‘other’ can be broken down into further subracial groups, such as Asian, Latino, Indian etc.
* While I can measure the fluctuation in GDP for various locations within St. Louis County via the Google Cloud Platform Covid-data, I don’t have sufficient data to infer how these fluctuations affect various age groups. Therefore, while comparing the economic impact of covid between rural and urban areas is possible, doing so between age groups is not.
* None of the datasets that I have listed contain data on mask usage between the various demographic groups, therefore I will be unable to infer whether Covid had less prevalence within a demographic due to mask use.

**Methodologies (in no particular order, tentative to change)**

* GDP growth comparisons between rural and urban zip codes – Using the Google Cloud Platform data, specifically the dataset that details the economic growth and production at the county level, I intend to create trendline plots detailing the change in GDP over the course of the pandemic in both the rural and urban zip codes within St. Louis County. To do this, I intend to subset the economic data by county for St. Louis County. Subsequently, I will randomly select at least 10 zip codes within my county, 5 rural and 5 urban, categorize them as ‘rural’ and ‘urban’, and plot the change in GDP growth/production overtime for urban and rural neighborhoods. While this won’t provide intuition as to the employment/unemployment rate in these areas, it certainly will illustrate the impact Covid-19 had on the economies of these local communities.
* One-way ANOVA between rural and urban zip codes on GDP growth – As a follow-up to the above comparison analysis, I’d like to provide information on whether the differences in GDP growth is significant between rural and urban areas. Using the same set of zip codes that I’ll use in the above analysis, I’ll perform a one-way ANOVA test to determine if the difference (if it exists) in GDP growth/production between the two areas is statistically significant. The results of the one-way ANOVA analysis will inform the reader whether residents living in rural areas were more or less impacted by the pandemic compared to residents living in urban areas.
* Confirmed Covid cases by age group over time – for this analysis, I intend to create a plot containing the timeseries for the various age groups from the datasets provided by the Google Cloud Platform and/or the JHU group. The processes for creating these timeseries is identical to what I will do for the GDP comparison above. This visualization will provide the reader with insight both as to difference in Covid prevalence among these groups and which group is more susceptible to contracting Covid.
* Confirmed Covid cases by zip code over time – I don’t intend to use every zip code in St. Louis County for this analysis, but I’d like to compare Covid-19 case rate timeseries between an equal number of rural and urban zip codes, similar to the comparison by the GDP growth in the above analysis. This will provide the reader with insight as to whether residents living in rural or urban areas are at greater risk of contracting Covid-19.
* Confirmed Covid cases by sex over time – Like the two timeseries above, I intend to compare Covid-19 case rate timeseries between the sexes in a similar fashion to the two comparisons above. I believe this will provide the reader with insight as to which sex is more at risk of contracting Covid-19 in this county.
* One-way ANOVA between various demographic groups on Covid case rate – These analyses will serve to compliment the above timeseries comparisons by showing a significant difference in the prevalence of Covid-19 among the various demographic groups I’ve included. The hope is that the results of these tests coincide with what is displayed on their respective timeseries comparisons.

**Projected Timeline**

Week of 11/15

* Gather all of the data in the forms of csv and excel files.
* Wrangle and curate the datasets in preparation for the subsequent analyses listed above. These clean datasets should only consist of relevant data for St. Louis County, Missouri.

Week of 11/22

* Build models and run analyses based on the clean data for St. Louis County
* Test the models and analyses for accuracy and statistical significance.
* Begin visualizing results of models & analyses

Week of 11/29

* Finishing creating visualizations from the prior week
* Create figures using the visualizations
* Put together a powerpoint presentation using the figures and data used for A6.
* Begin writing the research report for A7.

Week of 12/6

* Review and finish the presentation for A6
* Present research and findings for A6 to the class on 12/9
* Finish the research report for A7

Week of 12/13

* Review the research report for A7 prior to submission
* Submit research report for A7 on 12/14