Assessing COVID-19 Infection Rates

Urban and Rural Low Income Counties

Team 126

Laura Davila | Kevin Brown | Kat Goodman | Phyllis Watts Morris
Cristhian Rodriguez | Andualem Teshome

Introduction

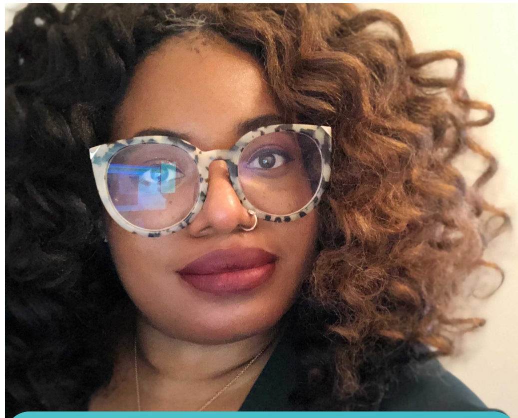
Meet the Team

Laura Davila

Team Lead
Data Scientist
Business Consultant

Cristhian Rodriguez

Data Scientist
Statistician
Python Guru



Kat Goodman

Data Scientist
Researcher
Dashboard Engineer

Introduction

Meet the Team

Andualem Teshome
Data Scientist
Researcher
Registered Nurse

Kevin Brown
Data Scientist
Dashboard Engineer

Phyllis Watts Morris
Data Scientist
Researcher
Writer/Tester

Introduction

As of August 7th, 2021, a total of 35,665,877 COVID-19 cases have been reported in the United States. Among those cases, more than 616,257 people have died as a result of the ongoing virus.

Growing evidence shows that there is a significant disparity in the infection rate among disadvantaged socio-economic communities. According to CDC, COVID-19 pandemic has brought social injustice and inequity to the forefront of public health. Ethnic minority groups and low income communities are known to be unequally affected by the pandemic.

Who is at greater risk should another pandemic occur?

Problem Statement

We examined the association of low income communities with COVID-19 infections among urban and rural US counties, hypothesizing a disproportionate difference in the cases of those who resided in counties classified as non-metro and a higher rate of poverty.

This study allows us to better understand how the United States can address and plan for the next global pandemic. Knowledge on the distribution of the virus over time is essential for planning purposes.

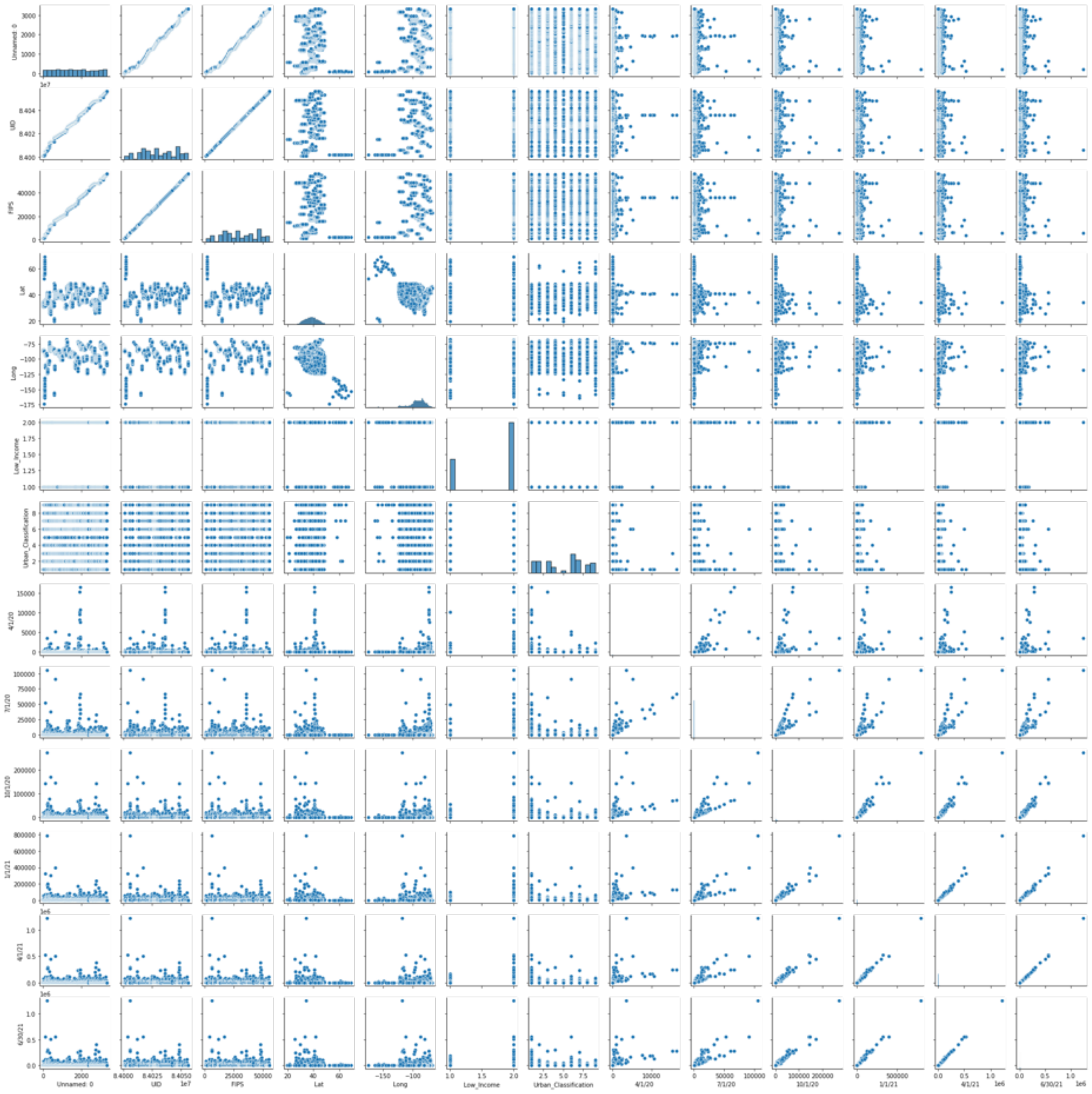
Data Cleaning & **Exploratory Data Analysis**

Data Cleaning

Exploratory Data Analysis

Datasets Used

1. John Hopkins University CSSE COVID-19 Dataset
2. NCHS Urban-Rural Classification Scheme for Counties
3. Median Household Income: 2015-2019 US Census
4. County Populations USA Facts



Data Cleaning
Exploratory Data Analysis

Datasets Used

1. John Hopkins University CSSE COVID-19 Dataset
2. NCHS Urban-Rural Classification Scheme for Counties
3. Median Household Income: 2015-2019 US Census
4. County Populations USA Facts

Disparity Cases Urban and Rural

Based on the classification of urban and rural counties, there was a notable disparity between the cases in urban counties vs rural counties. But over time, the difference between each category is less.

Low Income vs Not Low Income

Low income and not low income counties were also classified on a scale of 1 and 2, where 1 was low income and 2 was not low income, when compared to the median income per state. In the pair plot, we can also see the relationship and changes over time in counties with low income vs not low income counties.

Distribution of Cases

The pair plot also highlights how the case distribution changes over time per state. In the data set, each state is categorized by a FIPS code. In April 2020, we see outliers in different parts of the country when compared to the data in June 2021. We can see that over time, states assigned a lower FIPS code start to have more outliers

Population Ratio

In order to gather more insights, we found that we needed to add to our data the population for each county to compare how the cases increase over time per 50,000 people. By adding the rate, we can proportionally compare population groups of different sizes.

Hypothesis Testing

Analysis & Results

Summary Data & Hypothesis Testing

We defined our statistical significance as $p<0.05$. Our testing began with unpaired one & two tailed t – tests. To model the association between low income urban and rural COVID-19 infection rates, we used an analysis of variance (ANOVA) to analyze the effect of more than one categorical independent variables with our dependent variable — total cases per county.

Tools used for analysis: Python Libraries: Pandas, Seaborn and Matplotlib, Jupyter notebook and Excel.

Summary Data

3,143
Total Counties

94%
Counties Analyzed

918
Low Income Counties

33,057,752
Total Cases June 2021

1706
Rural Counties*

**NCHS Urban-Rural Classification was not available for 192 counties*

Hypothesis Testing

Analysis & Results

WE REJECTED OUR NULL HYPOTHESIS

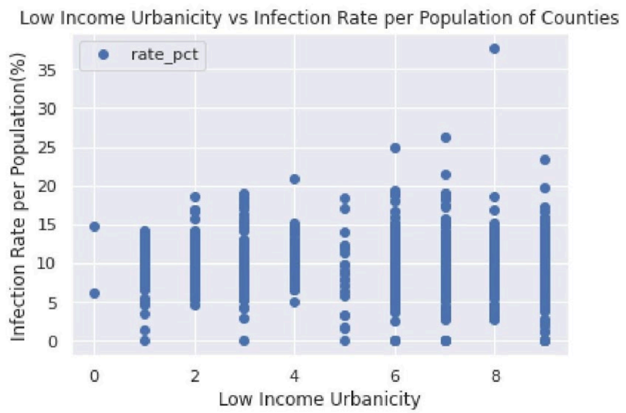
H_o= There is no statistical significant difference in COVID-19 infection rates in urban and rural counties, among low income groups.

H_a = There is a statistical significant difference in COVID-19 infection rates in urban and rural counties, among low income groups.

0.8 Urbanicity vs Rate(%) (Low Income Only)

```
[8]: subset_df.plot(x = "Urban_Class", y = "rate_pct", style = 'o')
     plt.title ('Low Income Urbanicity vs Infection Rate per Population of Counties')
     plt.xlabel ('Low Income Urbanicity')
     plt.ylabel("Infection Rate per Population(%)")
     plt.show
```

```
[8]: <function matplotlib.pyplot.show(close=None, block=None)>
```



When considering only low income, there are higher infection rates in rural areas than there are in urban areas

```
[10]: pingouin.ttest(subset_df["Urban_Class"],  subset_df["rate_pct"])
```

```
[10]:              T          dof       tail          p-val           CI95%  \
      T-test -33.073698  2033.713614  two-sided  2.544139e-192  [-4.48, -3.98]

              cohen-d        BF10  power
      T-test  1.408246  2.241e+190    1.0
```

Summary: Since the p-value of 2.54e-192 is less than 95% confidence level we are rejecting null hypothesis!

Visualizations: **Interactive Dashboard**

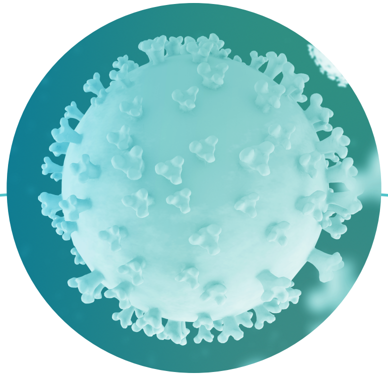
Summary & Conclusion

Discussion & Recommendations

While the infection rate was more experienced by poorer and more rural areas as of June 30th, 2021, disparities in COVID-19 infections exist beyond those explained by differences of income and urban classifications.

Insufficient access to disaggregated data and lack of time prevented us from further exploring and analyzing other driving factors in the infection rates. We believe that in addition to analyzing income and urban classification, future studies should leverage analyzing additional casual factors such as age distribution, racial/ethnic data, individual-level health data, mobility, and other social drivers.

Thank You!



correlation one
DATA SCIENCE FOR ALL™

Acknowledgements:
Correlation One, DS4A Empowerment, Casey Kongpanickul & Serena Zhou– TA, DS4A Empowerment, Carlton Triolo-Sapp – Mentor, Northwestern Mutual,, John Hopkins University , Centers for Disease Control , (other owner of 3rd dataset), Barbara Mullen, Luis Valenzuela (Former Team Member) Team Members 126