**Does income impact COVID rates?**

Ho Income has no impact on COVID rates and cases

Ha The p-value of 0.13 demonstrates we cannot reject the null hypothesis

Understanding if income levels of the population impact the COVID rates were limited by accounting levels of the data categories down to differing geographic increments.

COVID data is reported for the United States down to the county level.

Census data for median income is reported to zip code level.

A sample dataset of cases in Florida was identified that reported cases down to the zip code. Florida was chosen as it had a large data density of 680,000 cases reported. The initial analysis was conducted using symmetrical bins based upon categorical distributions of median income in $10,000 increments. This distorted the distribution of data. A calculation was used to determine data quartiles. Then the the dataset was divided into appropriate bin increments.

The analysis demonstrated a direct correlation between median income and rate visually on the box plots. As income increased in each boxplot, rate increased.

More statistical tests were conducted to quantify correlation. The higher the income, the higher the rate represented by y=0.09x + 18,544. The Pearson correlation coefficient of 0.05 indicated a wide variation around the line of the best fit. With a large distribution of COVID rates between $20,000 and $40,000 median income range.

**DATA LIMITATIONS**

**Reporting over Time**The virus began to spread through the United States over time. The world was not ready for an event such as this and with it being a novel virus, tests did not exist. Once it was declared a pandemic, the reporting of the data and testing began to accelerate. Therefore, appropriate times were selected to eliminate sparse data density due to limited reporting versus rates.

**Differing Geographies**  
This analysis was limited by the level of accounting for the variables. For instance, COVID data is reported for the United States only down to the county level, whereas the Census data for median income is reported to zip code level. FIPS and zip code tables, and sample datasets were used to align datasets.

**Different Time Elements**  
The COVID data was reported daily for all counties in the United States and the Census Data is reported annually. The census data used was legacy data for 2018. The mobility data was a precursory data point for COVID rates, so adjusting the time periods for comparative analysis was needed. It was moved two weeks forward to align with rates reporting due to COVID incubation period.

**Causal Elements**The granularity of the data prevented an in-depth investigation of real cause. Exposure would be a large contributing factor but it is difficult to quantify. Understanding correlation is not necessarily causation, other impacts could have critical impacts to the data. For example, the analysis presented questions regarding the data distribution, exposure, population density and other causal elements.

**Assumptions**

There were multiple assumptions made in calculating the analysis.

1. The data was normally distributed.
2. The data was independent.
3. The data was homogenous.
4. The standard deviations were assumed equal.