**MIS 6380.501 – DATA VISUALIZATION**

**GROUP 4 PROJECT REPORT**

**Florida State COVID Analysis**

**Group Members:**

1. Pragati Pramod Divekar

2. Pratik Devharkar

3. Yash Rakeshbhai Shiyani

4. Anagesh Ganta

5. Shivani Rege

6. Dhruv Patel

* **Executive Summary**

In this report, the impact of COVID-19 on Florida counties will explain the different effects of COVID-19 in a wide range of age groups. In other words, it provides the readers information of counties with highest spread of COVID-19 positive cases, to determine that COVID-19 is highly correlated with health conditions such as obesity and diabetes, which are major health issues in the United States today. Further, the visualizations have taken various filters applied to the eligible. With all of the different components of the data set, our team has created 5 hypotheses to visualize Florida counties COVID-19 effects.

The following sections go into further details about the hypotheses and their respective visualizations.

* **INDEX**
* **Data Description……………………………………**
* **DataSource……………………………….**
* **Data Cleaning………………………………………..**
* **Introduction………………………………………….**
* **Insights and Findings……………………………..**
* **Hypothesis 1………………….............**
* **Hypothesis 2……………………………..**
* **Hypothesis 3……………………………..**
* **Hypothesis 4……………………………..**
* **Hypothesis 5……………………………..**
* **Conclusion……………………………………………..**
* **Data Description**

Our dataset for this project is taken from Kaggle, publisher’s goal in this data is to provide a reproducible workflow to help data scientists and researchers study COVID19 in the United States. The data on Kaggle is based on the most recent figures reported by counties in USA. In this, The 3,142 counties of the United States span a diverse range of social, economic, health, and weather conditions data is given. County-level data on health, socioeconomics, and weather can help us address one of the primary tasks of the UNCOVER challenge, which is to identify which populations are at the greatest risk for COVID19.

* **Data Sources:**
* **Primary Data Source:** <https://www.kaggle.com/code/johnjdavisiv/us-counties-weather-health-hospitals-covid19-data/data?select=us_county_sociohealth_data.csv>
* **Secondary Data Source:**
* <https://data.world/deepspeak/sql-primer/workspace/file?filename=US+COVID+cases+by+County.csv>
* **Number of Records:** 17415 records
* **Data Cleansing Tool:** Python Pandas - It provides a rich set of functions to process various file formats from multiple data sources.
* **Visualization Tool:** Tableau- It’s a powerful Business Intelligence tool with more filtering features and less analysis time
* **Data Cleaning**

In our data cleaning part, we have used the following python code in which we have used Pandas and Numpy libraries of python.

* **Python code for data cleaning**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

#Reading the data from the CSV file and copying it into a data frame

covid\_health\_weather\_data\_df = pd.read\_csv("Cleaned\_data.csv")

#Print and check the Dataframe

covid\_health\_weather\_data\_df

#checking column names

print(covid\_health\_weather\_data\_df.columns)

#Displaying information of every column in dataset

covid\_health\_weather\_data\_df.info ()

#Displaying sum of empty rows in each column

covid\_health\_weather\_data\_df.isna(). sum()

covid\_health\_weather\_data\_df.drop(covid\_health\_weather\_data\_df.columns[[6,7,14,15,16,17,18,21,23,24,28,29,31,32,36,37,38,39,41,43,44,45,46,50,58,63,64,65,66,69,70,71,72,73,75,76,77,78,79,80,84,85,91,92,93,94,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109,110,111,112,132,133,142,148,158,159,160,161,162,163,164,165,166,167,168,169,170,171,172,173,174,175,176,177,178,179,180,181,182,183,184,185,186,187,188,189]], axis=1, inplace=True)

covid\_health\_weather\_data\_df

covid\_health\_weather\_data\_df.to\_csv('F:\sem2\DV\Project\Data\Cleaned\_Covid\_data.csv', sep='\t', encoding='utf-8’)

covid\_health\_weather\_data\_df = pd.read\_csv("Cleaned\_Covid\_data.csv")

covid\_health\_weather\_data\_df

* **Cleaned Data**

**Graphical user interface

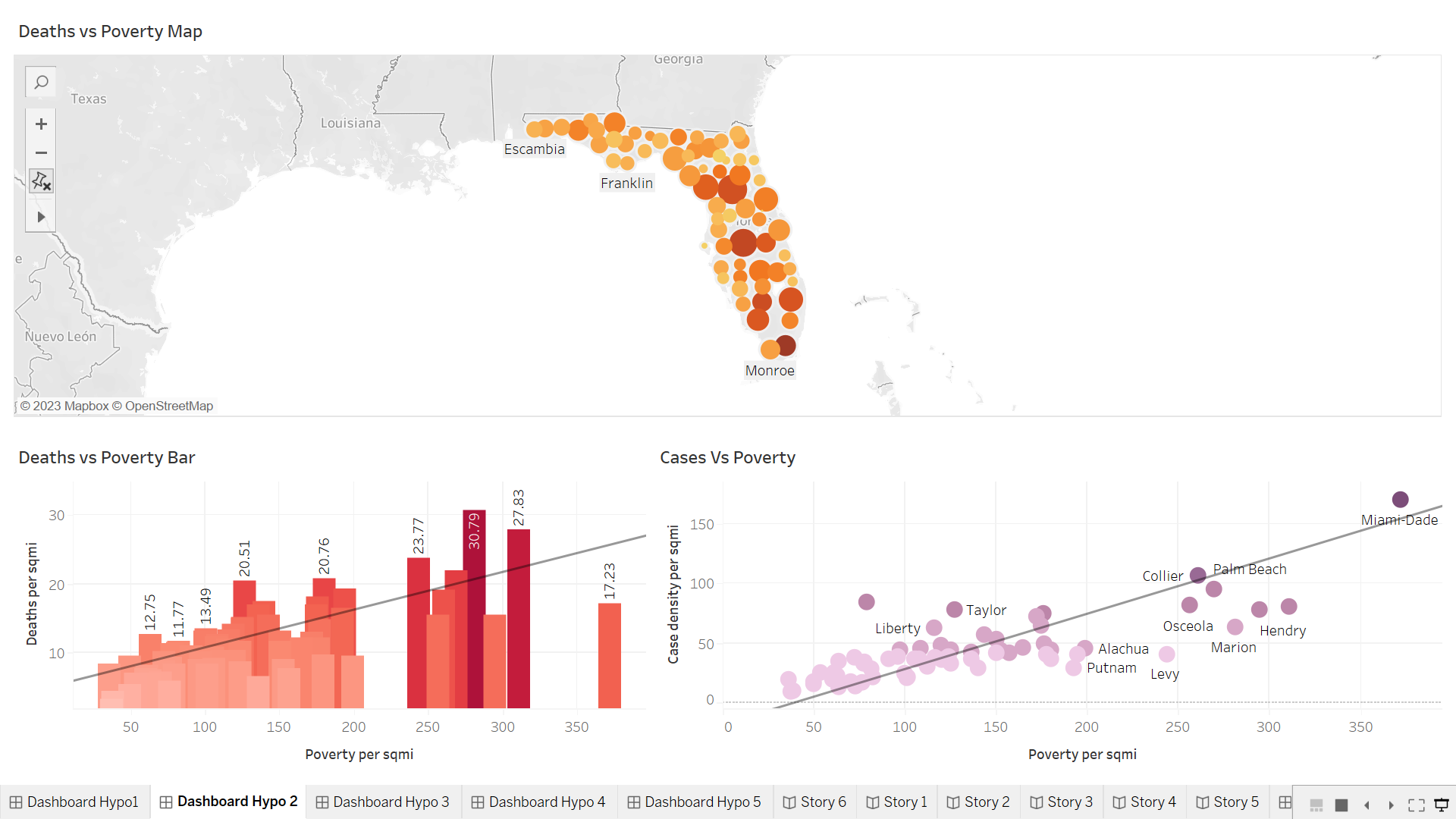
Description automatically generated**

* **General Introduction**

The COVID-19 pandemic is a global health emergency. It's also an unprecedented opportunity to use data science to help better understand the spread of this virus, as well as its potential impact on the world economy. In this project, we'll discuss how we used visualization tools to help track the spread of COVID-19 in Florida counties which span a diverse range of social, economic, health, and weather conditions.

# **Insights and findings**

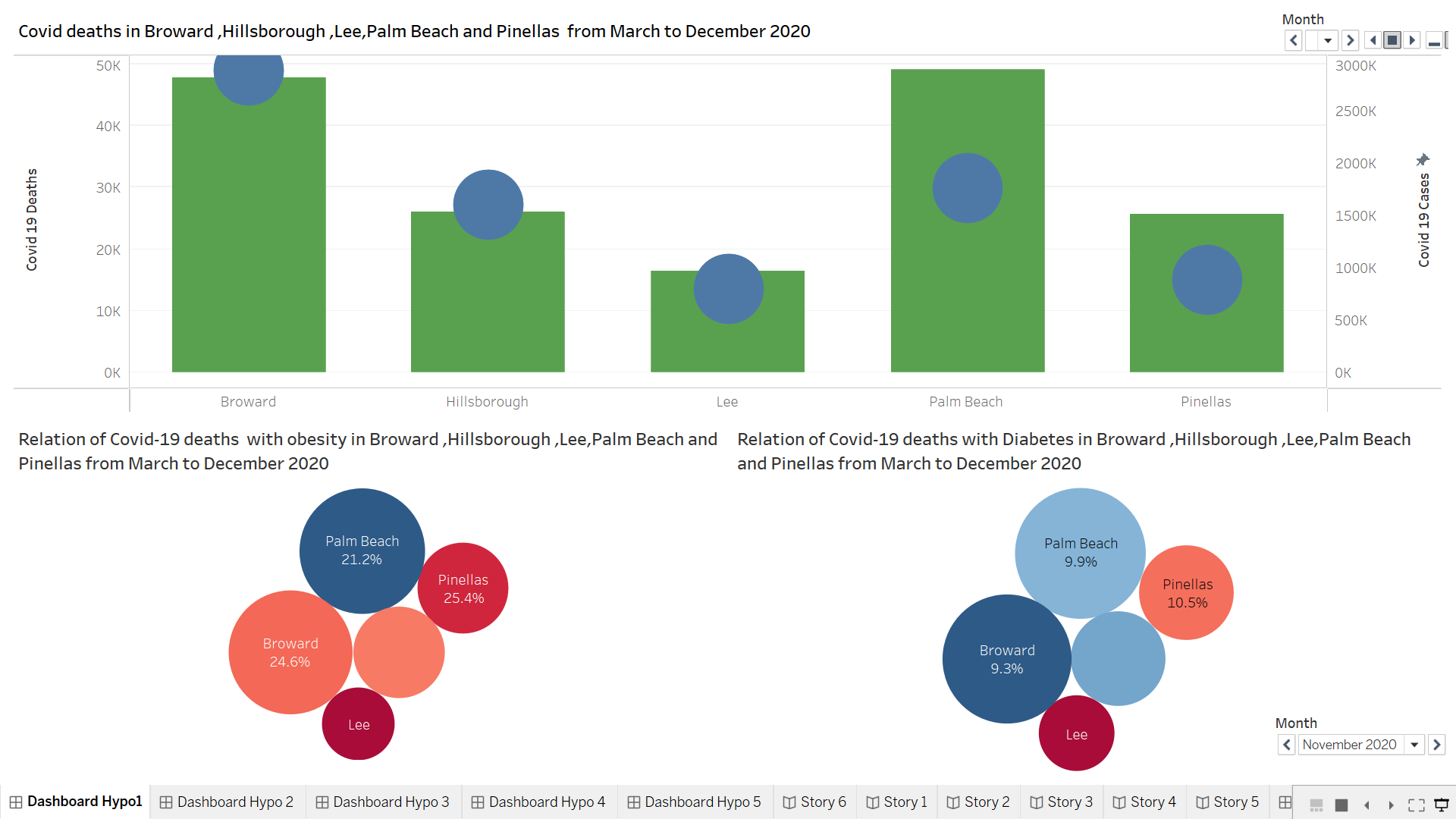
**HYPOTHESIS 1:** Whether Covid 19 deaths are more in counties with greater population below poverty line.



**Insight:**

The first hypothesis discusses the impact of poverty on no. covid cases. The first graph is a map view of the Florida counties where the intensity of the colour represents no. of people belonging to below poverty line and the size of the bubble depicts the no of deaths that happened due to covid 19. Hence, we can see that the darker colour bubbles have bigger size, it indicates that more people died in regions with high poverty. This can further be confirmed by the next two graphs too. The next graph plot consists of poverty and death cases. It shows that as the no. of people in poverty line increases, death cases increase too. The last one is a scatter plot which indicates Miami has the highest no. of poor people and the covid cases detected are high too. Hence the stated hypothesis is true. Several factors could contribute to this trend. People living in poverty have limited access to healthcare facilities, which makes them more vulnerable to the virus. They are less likely to get tested for COVID-19 and may not be able to afford treatment, leading to more severe cases and deaths. The lack of awareness and education is another factor that contributes to the trend. People living in poverty may not have access to information about the virus, its transmission, and preventive measures. This lack of awareness can lead to a higher risk of infection and can contribute to the spread of the virus. Furthermore, many people living in poverty work in jobs that do not allow them to work remotely or take time off if they fall ill. They may be forced to continue working even if they are infected with the virus, increasing the risk of transmission to others. It is essential to address these factors to mitigate the impact of the virus on vulnerable populations and ensure that everyone has access to healthcare facilities, education, and preventive measures. By doing so, we can reduce the number of COVID-19 cases and deaths in Florida County and other regions with a high percentage of the population below the poverty line.

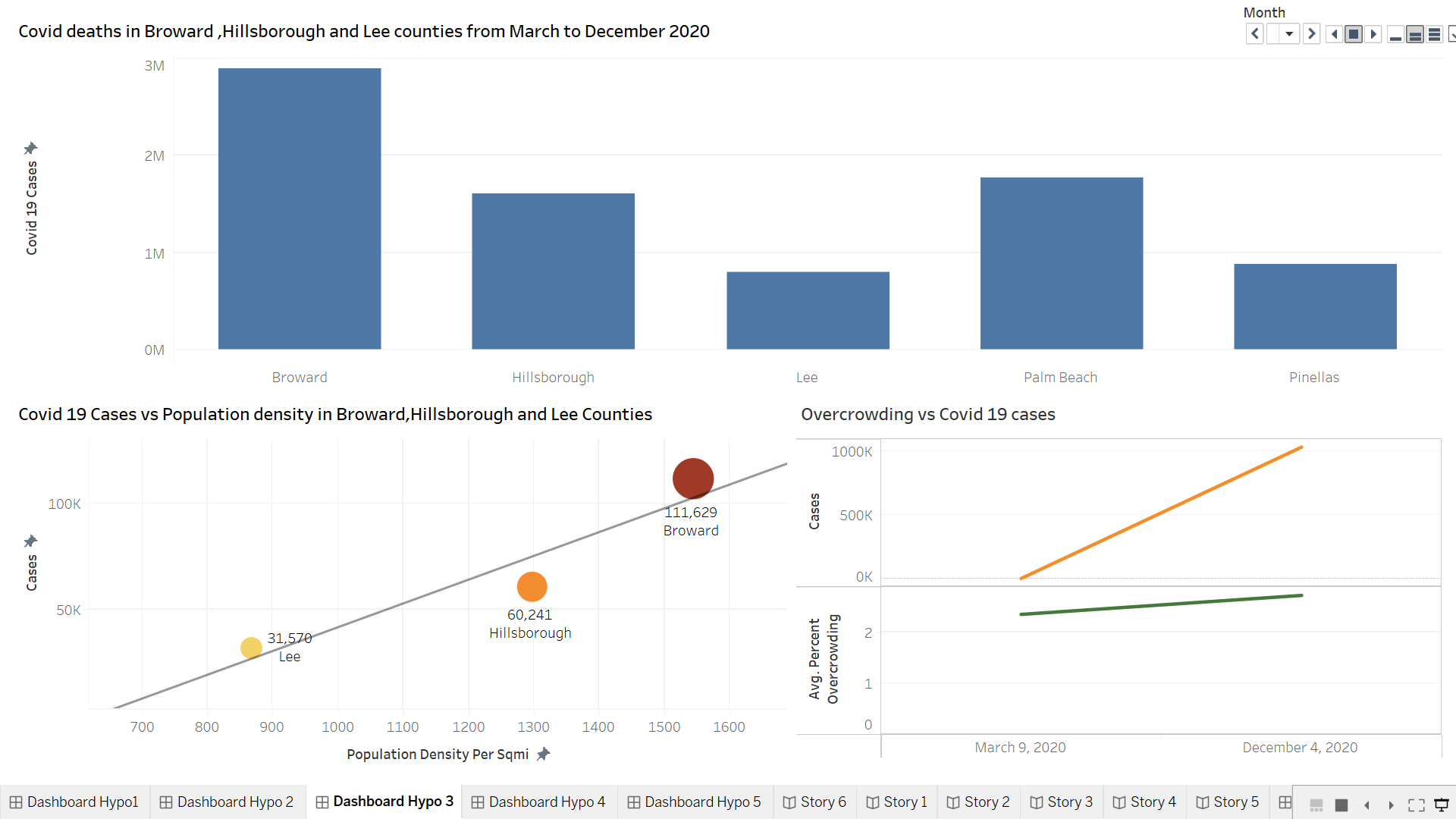
**HYPOTHESIS 2:** Whether medical conditions such as diabetes and obesity have an impact on increasing Covid 19 deaths in various counties.



**Insight:**

It is believed that medical conditions such as diabetes and obesity weaken the immune system, making diabetic and obese individuals more vulnerable to infections and hence increasing the death rate. We have displayed the COVID-19 deaths in Broward, Hillsborough, Lee, Palm Beach, and Pinellas counties from March 2020 to December 2020. The green bars represent deaths, and the blue circles show COVID-19 cases. We have observed that Broward and Palm Beach counties have the highest number of COVID-19 deaths, while Lee County has the least death count compared to the others. In the Covid deaths Vs Obesity bubble plot, we have examined the relationship between COVID-19 deaths and obesity in the respective counties. The size of the circle represents the number of deaths, while the colour represents the percentage of obese people. As the intensity of the red colour increases, the percentage of obese people also increases, and vice versa. Interestingly, we can observe that Lee County has the highest percentage of obese people, but it has the least number of COVID-19 deaths. This suggests that obesity does not significantly contribute to COVID-19 deaths. In the Covid deaths Vs Diabetes plot, we have plotted deaths with the percentage of diabetic people. We can clearly see that Palm Beach and Broward counties have the highest number of deaths. However, the blue colour indicates that they have the least percentage of diabetic people. Therefore, from our data, we can conclude that diabetes and obesity do not significantly contribute to increasing COVID-19 death cases in Florida counties.

**HYPOTHESIS 3:** Whether Population density is the reason behind spread of Covid 19 in Broward County.



**Insight:**

In this visualization, we have compared different counties from the state of Florida. For that, we analysed their population density with respect to covid cases and decided whether overcrowding affected the covid cases. From first graph we can get the information about counties cases from month of march to December from 2020. In second graph, the line describes the best that as we increase population density per sqmi. The covid cases are significantly increasing due to ease of spread. From last graph, the simple line graph is making it lot easier to say that from march to December of 2020 as the average percent of overcrowding going up, the cases are upscaling. So, from overall analysis it can be said that our hypothesis stated it right that Population density is the reason behind spread of covid 19 in Broward County.

**HYPOTHESIS 4:** **Whether counties with a greater percentage of Asian residents experienced a lower percentage of covid deaths**.

Scatter chart

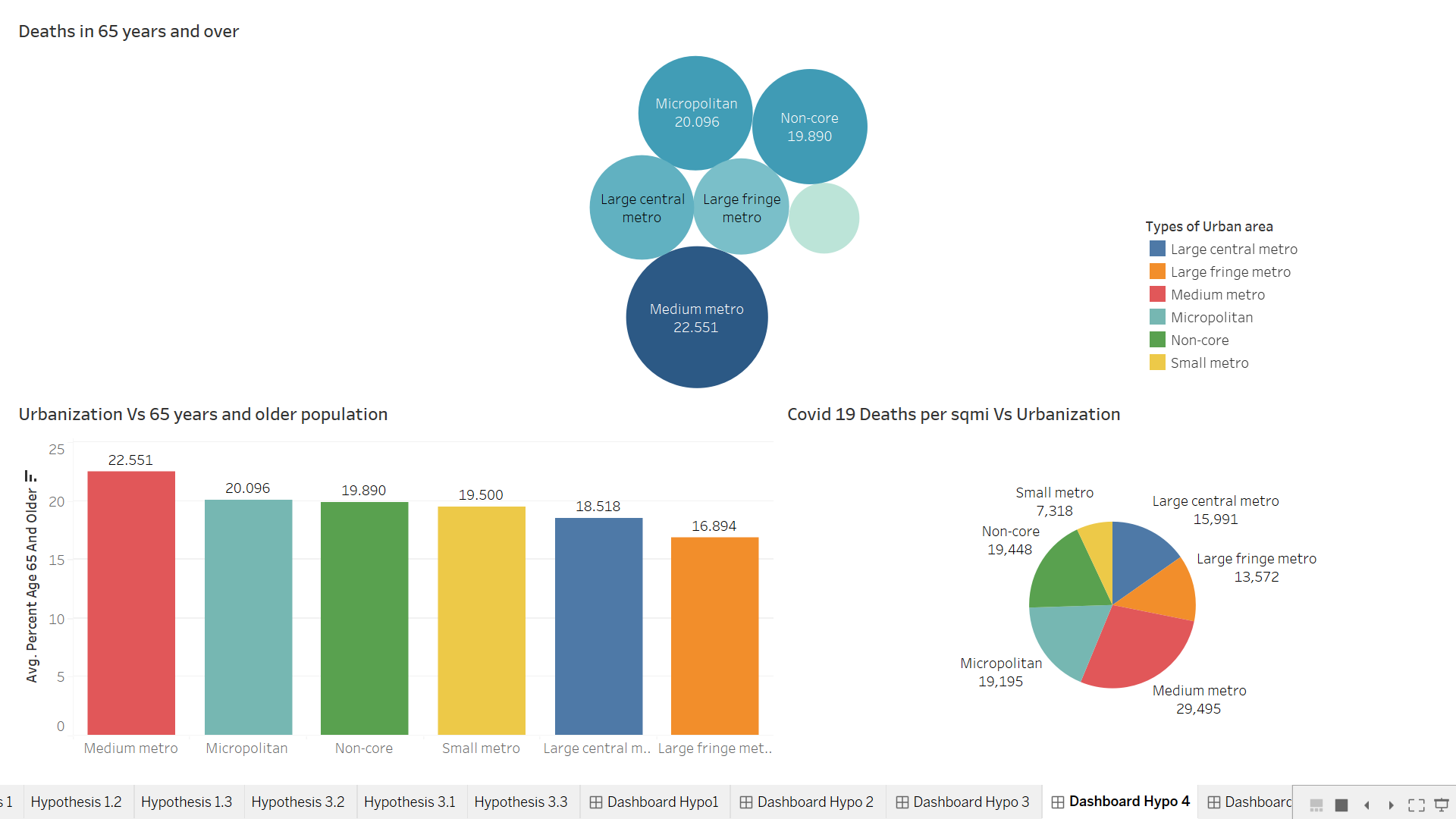
Description automatically generated with low confidence

**Insight:**

Our fourth hypothesis examines whether a higher percentage of Asian residents in Florida counties is associated with a lower percentage of COVID deaths. The rationale behind this hypothesis is that counties with a greater Asian population may have a higher median household income, providing them with greater access to healthcare and exercise opportunities.

As illustrated in our first graph, there is a clear negative correlation between the percentage of Asian residents and the percentage of COVID deaths in Florida counties. The graph below the first one reveals a positive relationship between the percentage of Asian residents and median household income, supporting our hypothesis. In the graph on the top right, we show that counties with a higher percentage of Asian residents generally have better access to healthcare facilities, as indicated by the blue bubbles. Conversely, counties with a lower percentage of Asian residents have inadequate healthcare facilities, as indicated by the red bubbles. The size of the bubbles corresponds to the percentage of Asian residents in each county. Similarly, our fourth and final graph demonstrates that counties with a higher percentage of Asian residents tend to have more exercise opportunities, such as gyms, as indicated by the red bubbles. Counties with a lower percentage of Asian residents have less access to exercise opportunities, as indicated by the blue bubbles. Once again, the size of the bubbles reflects the percentage of Asian residents in each county. In conclusion, our findings support our hypothesis that counties with a higher percentage of Asian residents in Florida generally experience a lower percentage of COVID deaths.

**HYPOTHESIS 5:** **Whether Large metros experienced higher deaths per sqmi because of overcrowding.**



**Insight:**

These are 6 metro types :  
 Large central metro  
 Large fringe metro  
 Medium metro  
 Micropolitan  
 Non-core  
 Small metro  
   
and according to our hypothesis whether large central metro and large fringe metro which are part of large metros with overcrowded population experiencing higher deaths comparing other metro areas.

Analysis of data from all the metro types showed interesting patterns. Large central metros and large fringe metros had lower COVID-19 deaths compared to medium metros. Further analysis revealed that the population of the 65+ age group is a significant factor influencing COVID-19 deaths.

Medium metros had higher populations of 65+ age people compared to large central and large fringe metros. In conclusion, our analysis suggests that the relationship between overcrowding and COVID-19 deaths in metropolitan areas is complex and influenced by various factors.

Overcrowding alone may not be the sole determinant of COVID-19 deaths, and other factors such as age demographics need to be considered.  
Thus, stated hypothesis proved wrong.

* **Conclusion**
* Conditions like diabetes and obesity have no such impact on the increase in Covid-19 deaths in various counties.
* If the population below the poverty line is greater in a county, then Covid-19 deaths are more in that county.
* High population density has led to more no of covid cases in Broward County,
* Florida counties, which experienced overcrowding, had higher number of deaths per sqmi in those respective areas.
* Counties with greater percentage of Asian residents have lower percentage of COVID deaths.