An Analysis of the Effects of COVID on the NYC Area and General Populus

STAT 405: Jana Alghamdi, Sharazad Ali, Yurie Han, Trey McCray, Andres Villada

April 11, 2023

Introduction and Data

Introduction

As of 2023, we live in a relatively post-pandemic era. COVID-19 affected us all in some way or another so we were intrigued to see what data trends reveal, on both a national and highly populated area, as the pandemic went on and what lessons we could learn from these trends to better prepare ourselves for future national health crises.

Primary Data Set

"COVID-19 Case Surveillance Public Use Data"

- Published by the U.S. Department of Health & Human Services
- Features 95M+ rows and 12 elements for all COVID-19 cases shared with CDC
- Elements include demographics, disease severity, and underlying medical conditions

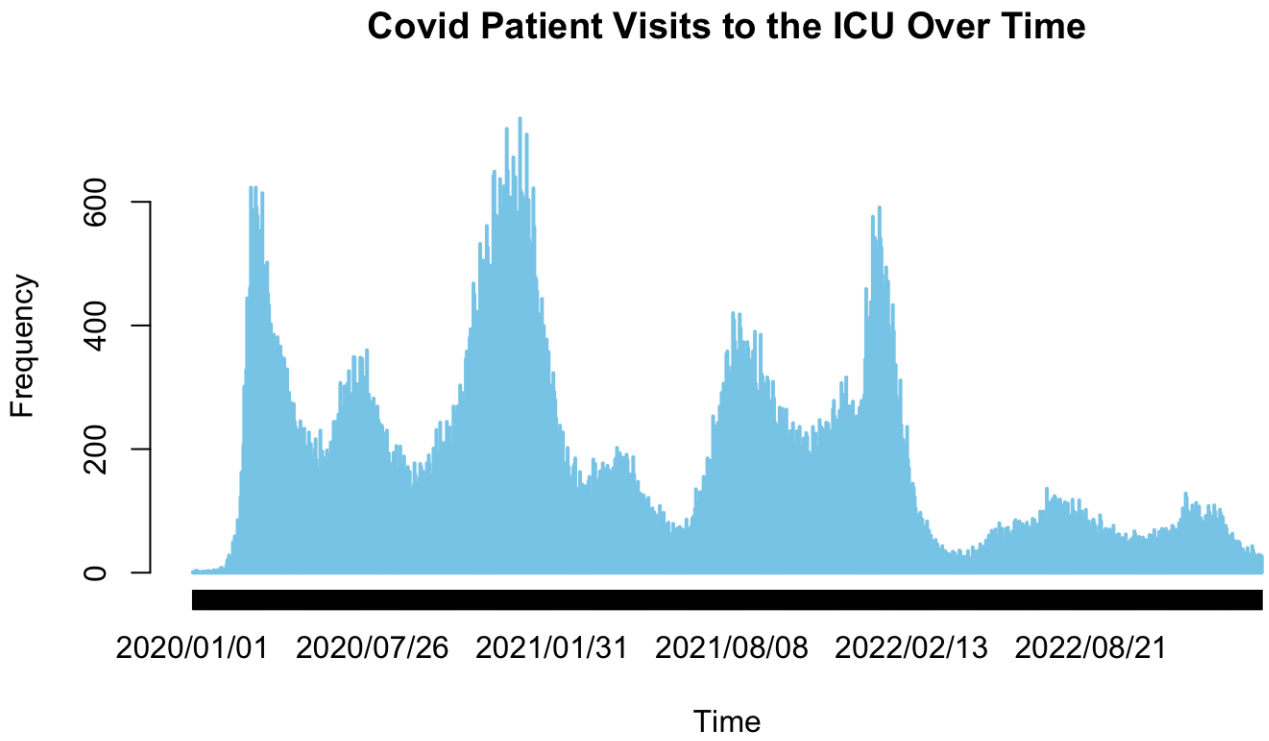
Secondary Data Sets

“Hospital Inpatient Discharges (SPARCS De-Identified): 2019” “Hospital Inpatient Discharges (SPARCS De-Identified): 2020” “Hospital Inpatient Discharges (SPARCS De-Identified): 2021”

- Published by the New York State Department of Health
- Featured 2M+ rows containing patient information for the 2019-2021 years
- 33 Elements including discharge detail, diagnoses, treatments, services, and charges

Plots and Analysis

COVID Patient Visits to the ICU Over Time

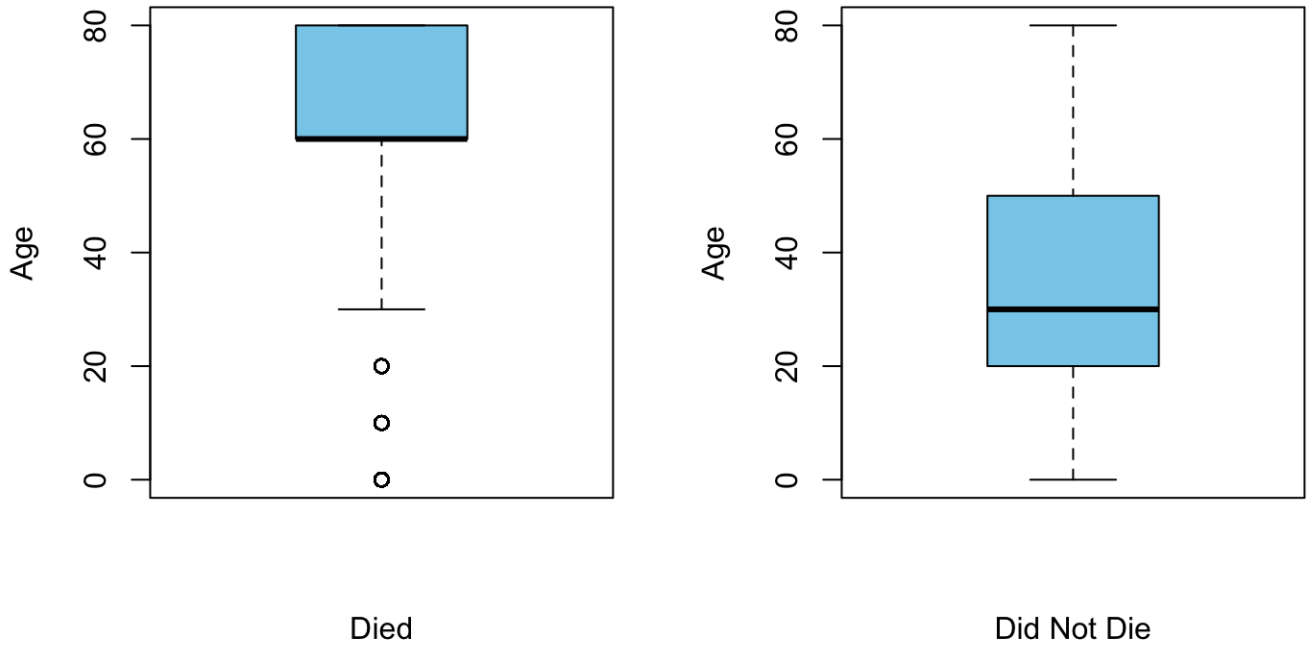


Question of Interest

What can we infer about the change in COVID ICU patient visits over time?

- Appearance of COVID first began with initial spike between Jan-Jun 2020. The frequency of patients reached a consistent peak of 550 daily ICU patients
- ICU patients decreased significantly after the initial peak between Jun-Nov 2020. Timeline coincides with the enforcement of lockdowns.
- Another spike in COVID cases in early 2021 likely due to a relaxation of lockdown procedures.
- Sharp decline and continuous decline starting on Feb 2022 likely due to vaccinations.
- It is inferred that changes in COVID patients is dependent on enforcement and reduction of lockdowns as well as vaccinations.

Mortality Distribution over Age Groups

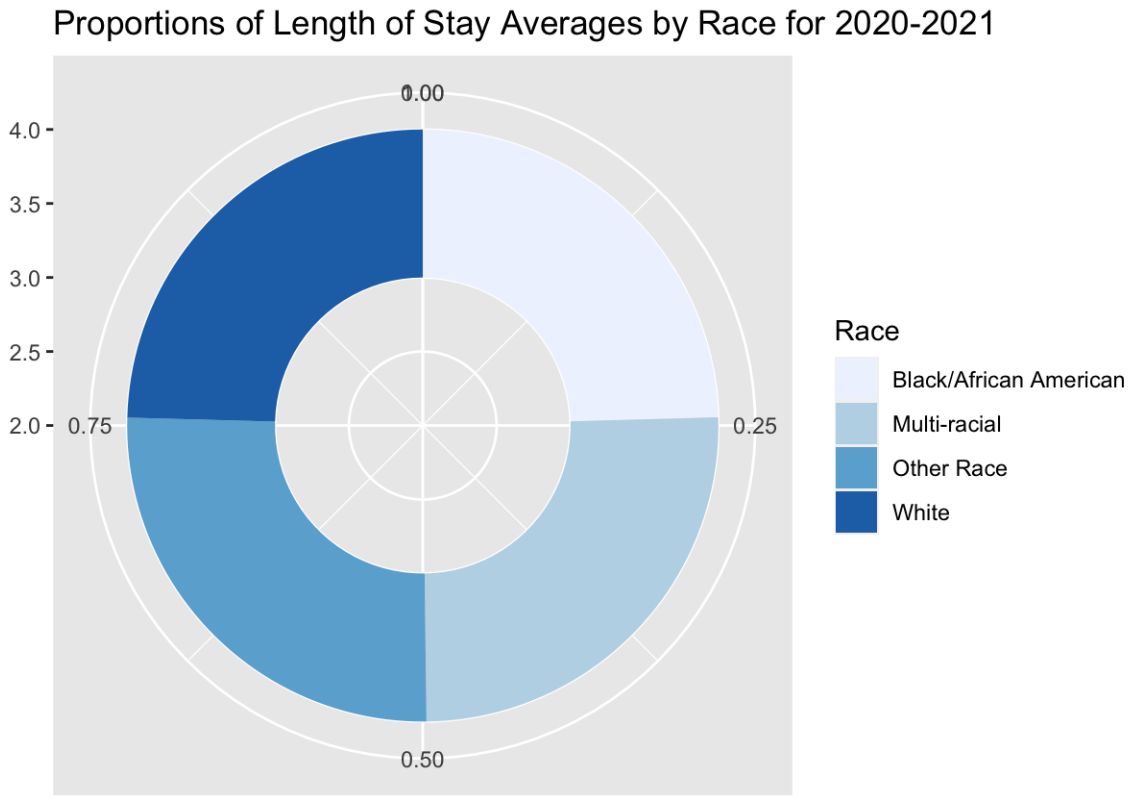


Question of Interest

Is there a difference in mortality risk for COVID between the age groups?

- Higher mortality is concentrated towards older ages (50% of the data lies from 60-80 years old)
- Lower mortality is concentrated towards younger ages
- Difference between contrasting age groups' mortality rates is likely due to health complications that arise with age
- Data for those who did not die is more spread out.

Proportions of Length of Stay Averages Grouped by Race (2020-2021)

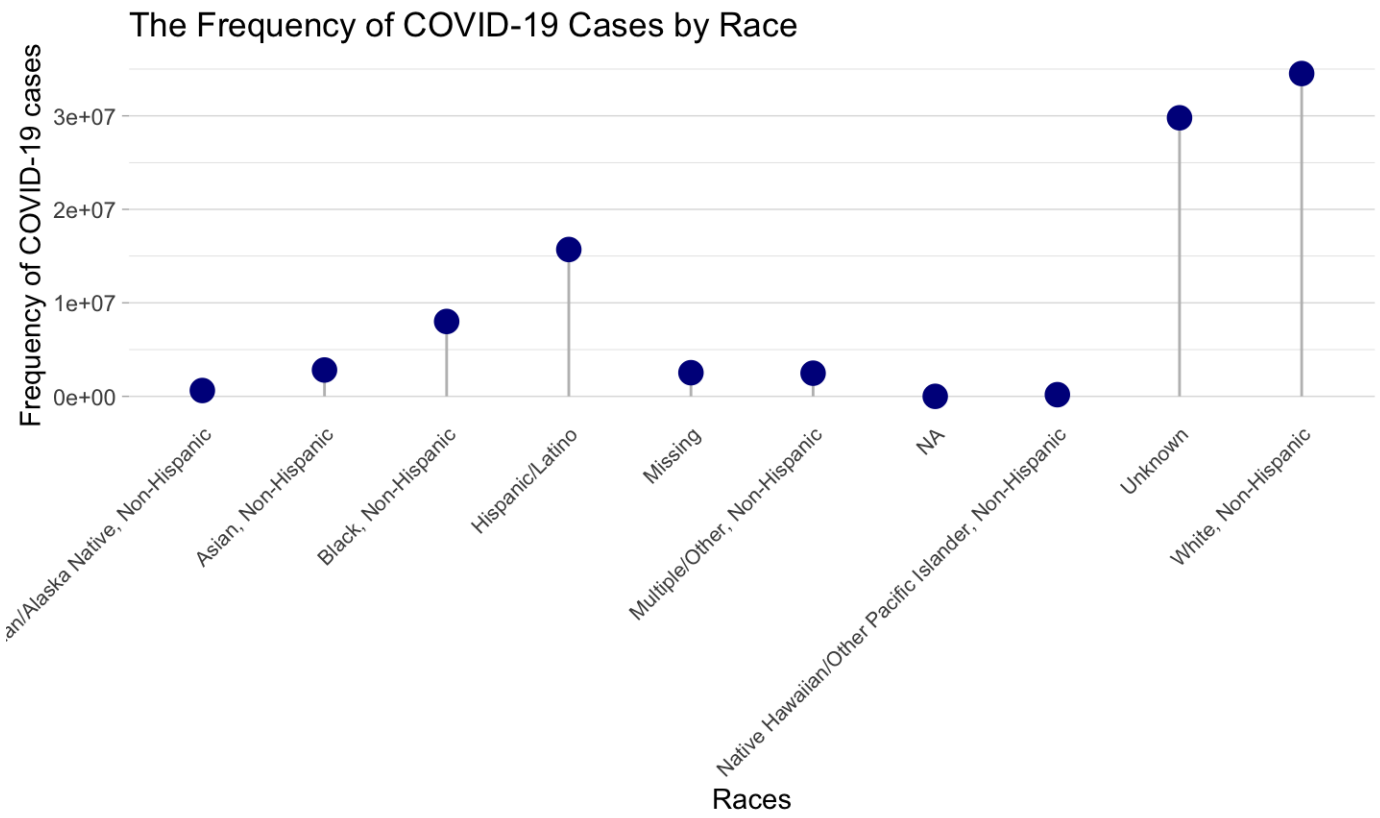


Question of Interest

Is there a difference in length of stay for COVID illness across races?

- It can be concluded that there is not a significant difference in length of stay for COVID-19 illness across races.
- Plot exhibits an even distribution of proportions between the races
- Maximum average length of stay by race was the Other Race category (8.893880 days)
- Minimum average length of stay was Black/African American (8.577218 days)
- Therefore, given the very small difference between the maximum and minimum average length of stay, there was no significant difference of Length of Stay by race for COVID patients.

Frequency of COVID-19 Patients Over Races



Frequency of COVID Patients Grouped by Race/Ethnicity

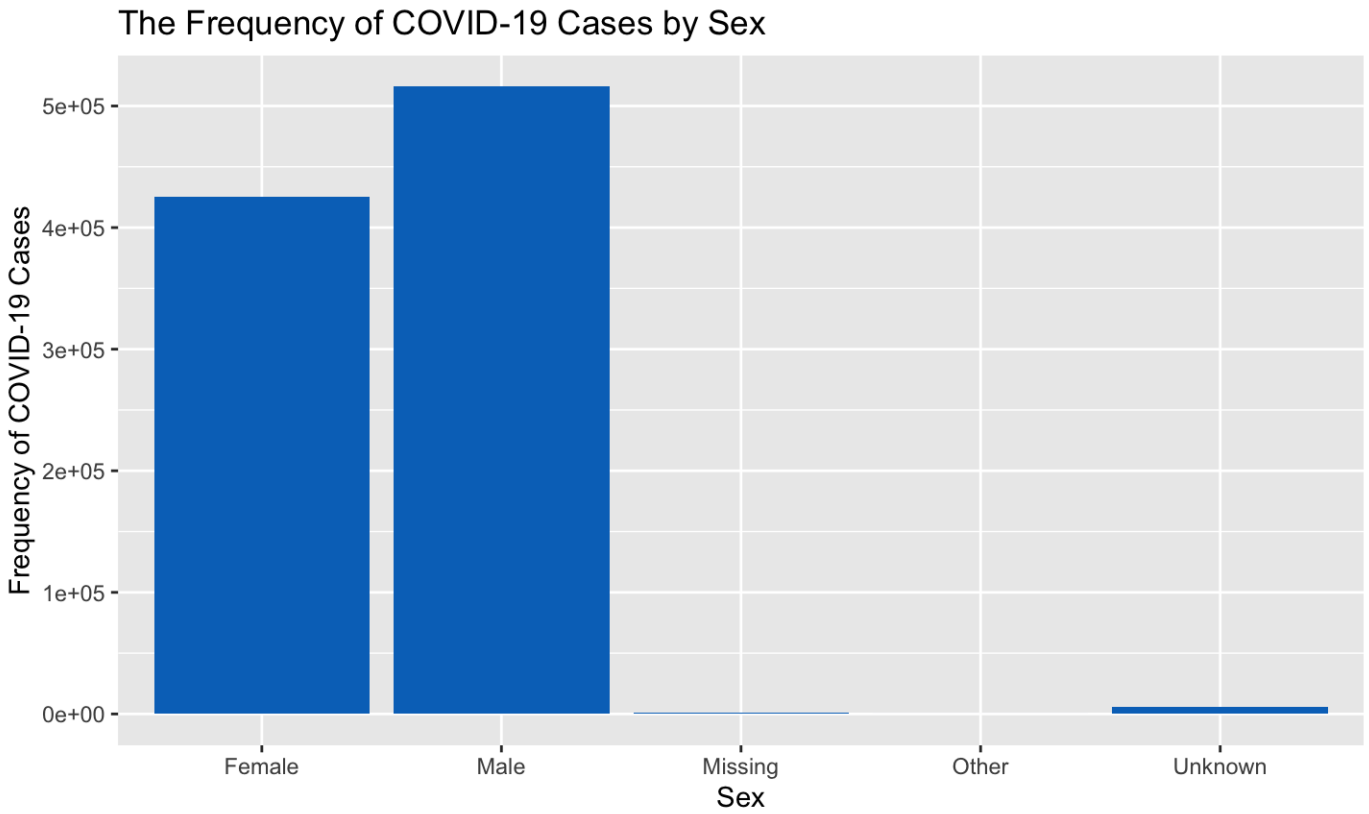
race_ethnicity_combined	freq
White, Non-Hispanic	34524397
Unknown	29790224
Hispanic/Latino	15703401
Black, Non-Hispanic	8007592
Asian, Non-Hispanic	2817740
Missing	2528751
Multiple/Other, Non-Hispanic	2475372
American Indian/Alaska Native, Non-Hispanic	623991
Native Hawaiian/Other Pacific Islander, Non-Hispanic	178012
NA	7

Question of Interest

How reflective is the frequency plot of COVID cases by race to US demographics?

- Maximum number of COVID cases occurred among White people
    - Supported by the fact that white people makes up most of the US (~60.1%).
- Hispanic people had the second highest number of cases
    - Hispanic people make up almost 18.5% of the US.
- Individuals of Native Hawaiian/Other Pacific Islander descent exhibited lowest number of COVID cases
    - Native Hawaiian/Other Pacific Islanders make up (~0.2%) of the US
- Thus, it is very important when looking at the frequency of COVID cases of the ethnicity/race of a certain population to consider what percentage each ethnicity/race makes up of the whole population at hand

Frequency of COVID-19 Cases by Sex

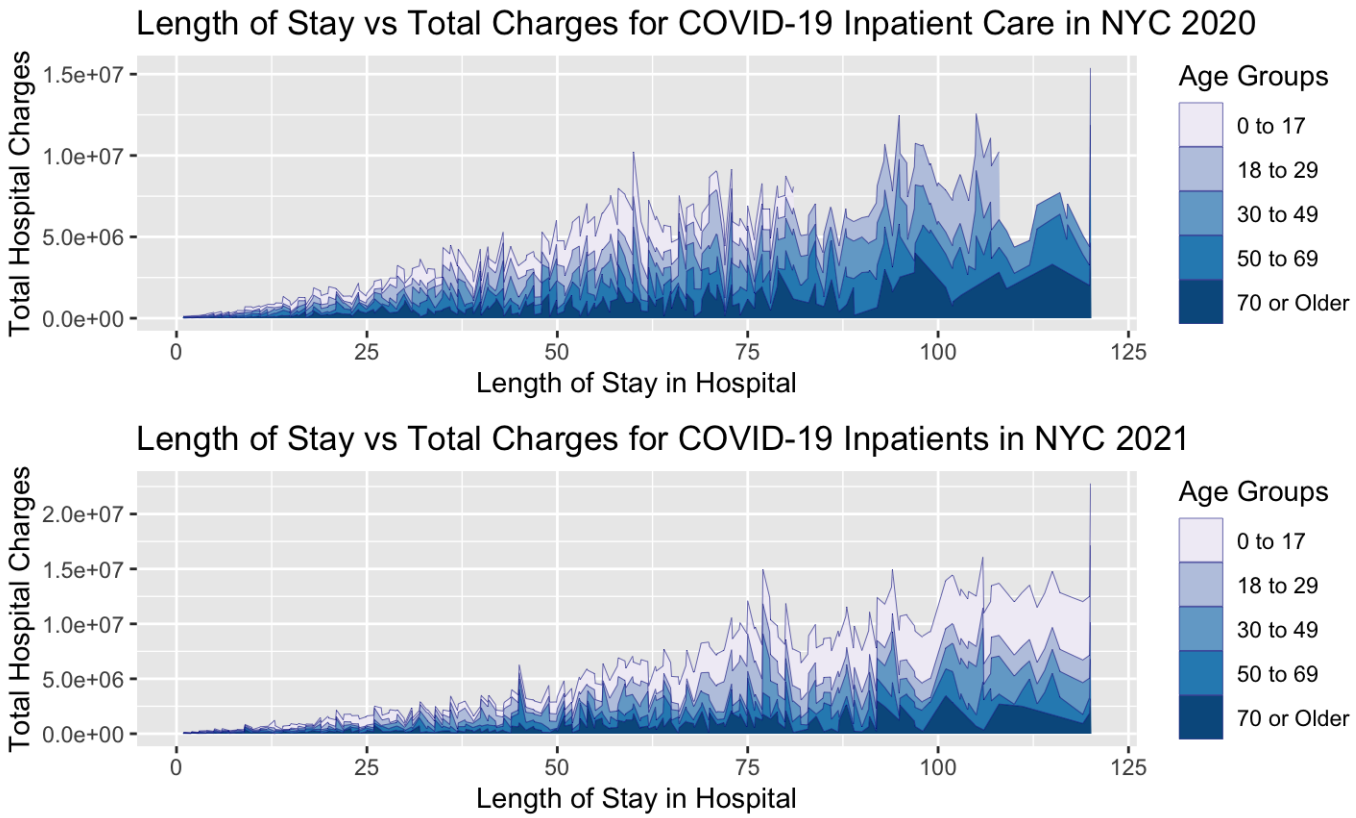


Question of Interest

Is there statistical evidence for sex differences in death caused by COVID?

- According to the graph, males ranked the highest when it came to the reported COVID cases, followed by females However, the difference between males and females is minimal
    - Barely a difference between the sexes and the reported COVID cases through the years 2020-2023.
    - The slight difference between them may be contributed to more males reported COVID cases more than the females did.

Length of Stay over Total Charges for COVID Patients (2020-2021)



Correlation Between Length of Stay and Total Charges (Grouped by Age Group)

Age.Group	correlation
0 to 17	0.8652675
18 to 29	0.8626409
30 to 49	0.8193177
50 to 69	0.8165139
70 or Older	0.7485752

Question of Interest

What is the correlation between length of stay and total charges within each age group?

Plot Analysis:

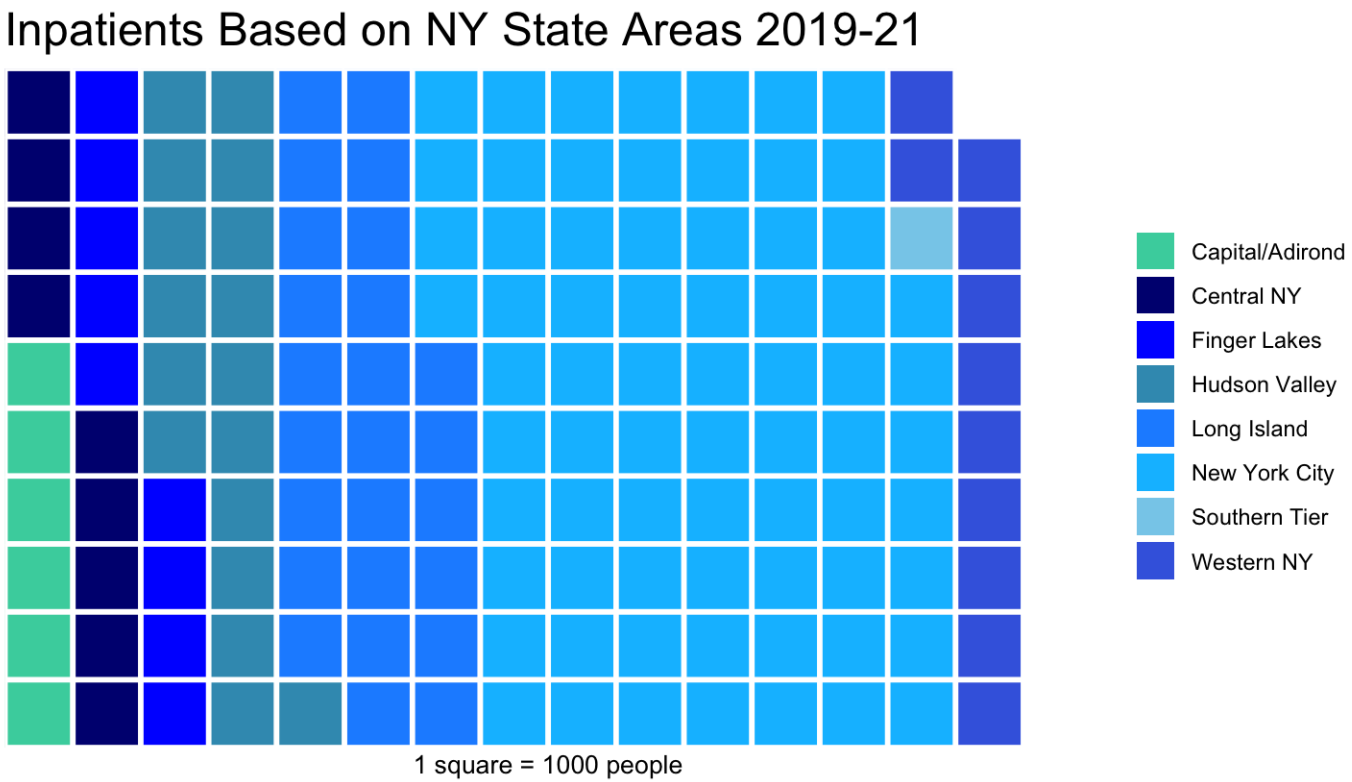
- 2020 and 2021 plots are similar in that both exhibit:
  - A linear relationship between patient length of stay and total charges
  - An inverse relationship between age group and total charges
  - Highest charges occurred for the 18 to 29 year old age group
    - 2020: Highest charge for approximately 105 days
    - 2021: Highest charge for approximately 80 days

- As each age group gets older, each group reaches their highest charge at a higher length of stay.
    - Younger people tend to have less access to health insurance or life savings.
    - Less likely for younger people to have a prolonged hospital visit due to less severity in their illness in comparison to older age groups.

Table Analysis:

- Across all age groups, the correlation proved to be very strong, as all of the age groups had a correlation in the ~.80 range, while the 70 or older age group had a correlation of .75.
- Exhibits strong evidence that there is a direct relationship between length of stay and total charges: as the length of stay increases, the total charge for each patient also increases across all age groups.

Inpatients Based on NY State Areas (2019-2021)

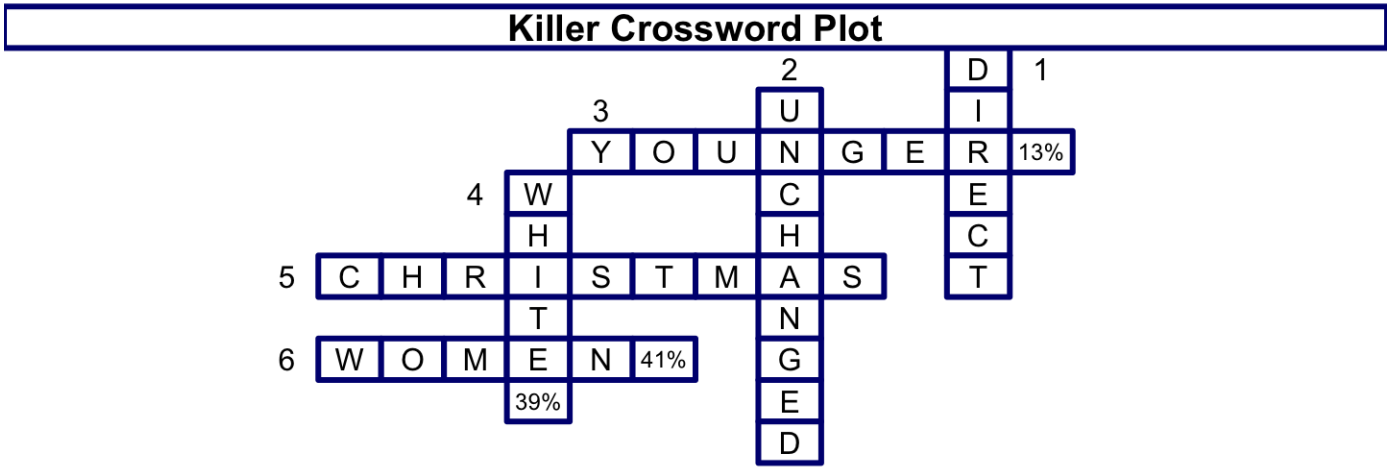


Question of Interest

Is COVID more likely to transmit in some areas than others?

- New York City area had the highest inpatients
- Less inpatients from less populated areas
- Positive correlation between densely populated areas with the number of inpatients admitted
- Proportion of people in each area remained relatively unchanged

Killer Plot



Across

3. Age Group with Highest Hospital Charges.
5. What holiday did patients visit the ICU the most?
6. Which Sex had the highest frequency of covid?

Down

1. Relationship between length of stay and charges incurred.
2. The trend in migration in NY remained --------?
4. Which race had the highest number of covid cases?

Killer Plot Interpretation

- Crossword Puzzle
    - Represents the outcomes we found from the overall data visualizations we have done on the COVID datasets.
        - Clues present for some of the big takeaways for each outcome
        - Word in puzzle corresponds to its clue

Conclusion and Future Work

Conclusion

Findings:

- After looking at the hospital trends overall, we noticed that the trends in ICU patients over time are indicative of the chronological order of the pandemic
    - For example, during lockdown rulings there seemed to be less cases, while during peaks of the pandemic, ICU patient numbers rose (especially in more populated areas).
- The general demographic breakdown of our results are as follows
    - Across all races, White individuals had the highest frequency of cases and men exhibited the highest rate of cases compared to other sexes.
    - Across all races, there seemed to be an even length of hospital pay, which substantiates proof that race does not necessarily impact the severity of COVID-19 symptoms.

- Younger age groups incurred higher charges for the same length of hospital stay
    - Lack of health insurance and savings

Future Work

- Consider looking further into COVID-19's effects on race
    - Calculate the rate of illness in each race to New York's overall racial demographics
        - By observing this statistic, we would be better able to answer the question: Does COVID-19 impact marginalized groups at a higher rate?
        - Through the study and analysis of this data, we can better understand how infectious diseases spread throughout a population as well as its effect on our society.
        - By detecting the general trends that may occur in transmission, we will be better-equipped with predicting who is more vulnerable and susceptible to the illness and mitigating necessary measures to protect these people.