University of Texas at Dallas

MIS 6380.004 DATA VISUALIZATION

GROUP PROJECT REPORT

GROUP 4

AMOL BHADANE
RADHIKA DUBEY
PRANAV MODY
HEMANGI PATIL
PRIYA VERMA
SUJAN NAIDU BHASKAR SURISETTY

Table of Contents

Project Description:

As New York State grappled with the COVID-19 pandemic, its healthcare workforce stood at the forefront, valiantly battling the virus. This project delves into the intricate patterns of COVID-19 infections, vaccinations, and hospitalizations across the state's diverse regions. By scrutinizing vaccination rates among healthcare workers and exploring potential disparities between metropolitan and non-metropolitan areas, the analysis aims to unravel the nuances that shaped the pandemic's trajectory. Additionally, it investigates the relationship between hospital density and vaccination uptake, shedding light on the factors influencing the adoption of this crucial line of defense.

Furthermore, the project expands its scope beyond New York State, exploring the nationwide impact of the pandemic. It examines the surge in internet usage during the crisis, potentially driven by the widespread adoption of remote work arrangements. Moreover, it delves into the intriguing hypothesis of gender-specific COVID-19 symptom manifestations, specifically concerning fever prevalence in May 2020. Through this comprehensive investigation, the study endeavors to uncover valuable insights that could inform future preparedness strategies and shape our understanding of this unprecedented public health challenge.

Objective:

This project aims to conduct a comprehensive analysis of COVID-19 patterns across New York State and the U.S., unveiling disparities in case distribution between metropolitan and non-metropolitan counties, especially for the 20-44 age group. It explores potential correlations between hospital density and vaccination uptake rates. Additionally, it examines the nationwide surge in internet usage during the pandemic, potentially linked to remote work adoption, and investigates hypothesis surrounding gender-specific COVID-19 symptom manifestations. Through data cleaning, standardization, and visualization, the goal is to present a visual model capturing the pandemic's impact on healthcare infrastructure, hospitalization rates, and vaccination trends in New York over time. The insights generated will inform future preparedness strategies, resource allocation, and public health policies, enhancing resilience against similar crises.

Data Description:

Dataset 1: https://health.data.ny.gov/Health/New-York-State-Statewide-Hospital-Staff-COVID-19-V/qfps-y8ta/about_data

This dataset includes information at the report date level by individual facilities on the total number of staff, how many staff are partially vaccinated, and how many are fully vaccinated. This information is only collected once a week from hospitals. The dataset 1 contains 17,900 rows, 14 columns.

Column Name	Description	Type
Report Date	The date the survey information was submitted	Date & Time
PFI	Unique Facility Identifier	Plain Text
Facility Name	Facility Name reporting the information	Plain Text
Hospital County	Facility County	Plain Text
Hospital Region	Facility Region	Plain Text
NY Forward Region	NY Forward Region where facility is located	Plain Text
Hospital Network	Network of the Facility	Plain Text
Total Employees	What is the total number of facility staff?	Number
Partially Vaccinated	Of the total number of facility staff, how many have received only one dose of a two dose MRNA COVID-19 vaccine series?	Number
Fully Vaccinated	Of the total number of facility staff, how many completed a COVID vaccine series?	Number
Staff Booster	The total number of facility staff that have received the additional or booster dose of the vaccine (Effective 1/4/2022)	Number
Total Staff Furloughed	Of the total number of facility staff, how may are currently furloughed due to a positive COVID 19 test or due to exposure?	Number
Direct Care Staff Furloughed	Of the total number of individuals employed by or used by your facility with "Hands On" patient/resident care responsibilities, how many are currently furloughed due to a positive COVID 19 test or due to exposure?	Number
Direct Care Employees	What is the total number of direct care staff employed by or used by your facility with "Hands On" patient/resident care responsibilities?	Number

Dataset 2: https://health.data.ny.gov/Health/New-York-State-Statewide-Hospital-Staff-COVID-19-V/qfps-y8ta/about_data

Annual Resident Population Estimates, Estimated Components of Resident Population Change, and Rates of the Components of Resident Population Change for States and Counties: April 1, 2020 to July 1, 2022. The dataset 2 contains 125 rows, 4 columns.

Dataset 3: https://health.data.ny.gov/Health/New-York-State-Statewide-COVID-19-Hospitalizations/jw46-jpb7/about_data

Description: This dataset includes information at the reporting facility level on patients hospitalized, admitted, discharged and fatalities. It also includes information on staffed beds. Patient information collected as part of the HERDS Hospital Survey are lab-confirmed COVID-19 positive. Hospitalized means patients admitted as inpatients in either inpatient or observation beds and does not include patients that were treated and released from an Emergency Department. The dataset 3 contains 240956 rows, 37 columns.

Columns in this Dataset:

Column Name	*Description*	*Type*
As of Date	The hospital reporting date through the Health Electronic Response Data System (HERDS) survey	Date & Time
Facility PFI	Facility PFI	Plain Text
Facility Name	Hospital Name	Plain Text
DOH Region	Hospital Regional DOH Office	Plain Text
Facility County	The NY county that the facility is located within	Plain Text
Facility Network	The network of the facility	Plain Text
NY Forward Region	NY Forward Region in which the facility is located	Plain Text
Patients Currently Hospitalized	How many confirmed positive COVID-19 patients does the facility have in either inpatient or observation beds at this time?	Number
Patients Admitted Due to COVID	How many patients with confirmed COVID were admitted due to COVID or complications of COVID?	Number
Patients Admitted Not Due to COVID	How many patients with confirmed COVID were admitted where COVID was not included as one of the reasons for admission?	Number
Patients Newly Admitted	How many confirmed, positive COVID-19 patients have been newly admitted since the last report?	Number
Patients Positive After Admission	How many of the positive COVID-19 patients were confirmed as positive AFTER admission AND since the last report?	Number
Patients Discharged	How many confirmed positive COVID-19 patients have been discharged from the facility since the last report?	Number
Patients Currently in ICU	How many confirmed, positive COVID-19 patients are there in the ICU at this time?	Number
Patients Currently ICU Intubated	This field is no longer updated due to changes in HERDS reporting requirements. Of the confirmed positive COVID-19 patients currently in the ICU, how many are intubated?	Number

Patients Expired	How many confirmed positive COVID-19 patients have expired in the facility since the last report? Summary level reporting by the facility.	Number
Cumulative COVID-19 Discharges to Date	Cumulative Discharges	Number
Cumulative COVID-19 Fatalities to Date	The cumulative number of in-hospital fatalities to date. The reporting of cumulative in-hospital fatalities are from a patient-specific verified file reported by the hospital and may not match the summary level reporting of Patients Expired.	Number
Total Staffed Beds	Total Staffed Beds in Hospital. Data Replaced as of May 19, 2021 by Tot_Acute_Beds	Number
Total Staffed Beds Currently Available	Total Staffed Beds Currently Available in Hospital. Data Replaced as of May 19, 2021 by Tot_Acute_Occup	Number
Total Staffed ICU Beds	Total Staffed ICU Beds in Hospital. Data Replaced as of May 19, 2021 by Tot_ICU_New_Beds	Number
Total Staffed ICU Beds Currently Available	Total Staffed ICU Beds Currently Available in Hospital. Data Replaced as of May 19, 2021 by Tot_ICU_New_Occup	Number
Total Staffed Acute Care Beds	How many staffed acute care beds are currently at your hospital?	Number
Total Staffed Acute Care Beds Occupied	How many of those staffed acute care beds are currently occupied?	Number
Total Staffed ICU Beds 1	How many staffed ICU beds are currently at your hospital?	Number
Total Staffed ICU Beds Currently Occupied	How many of those staffed ICU beds are currently occupied?	Number
Total New Admissions Reported	Total New Admissions (Patients Newly Admitted + Patients Positive After Admission)	Number
Patients Age Less Than 1 Year	This field is no longer updated due to changes in HERDS reporting requirements. Currently hospitalized age category less than 1 year	Number
Patients Age 1 To 4 Years	This field is no longer updated due to changes in HERDS reporting requirements. Currently hospitalized age category 1 to 4 years	Number
Patients Age 5 to 19 Years	This field is no longer updated due to changes in HERDS reporting requirements. Currently hospitalized age category 5 to 19 years	Number
Patients Age 20 to 44 Years	This field is no longer updated due to changes in HERDS reporting requirements. Currently hospitalized age category 20 to 44 years	Number

Patients Age 45 to 54 Years	This field is no longer updated due to changes in HERDS reporting requirements. Currently hospitalized age category 45 to 54 years	Number
Patients Age 55 to 64 Years	This field is no longer updated due to changes in HERDS reporting requirements. Currently hospitalized age category 55 to 64 years	Number
Patients Age 65 to 74 Years	This field is no longer updated due to changes in HERDS reporting requirements. Currently hospitalized age category 65 to 74 years	Number
Patients Age 75 to 84 Years	This field is no longer updated due to changes in HERDS reporting requirements. Currently hospitalized age category 75 to 84 years	Number
Patients Age Greater Than 85 Years	This field is no longer updated due to changes in HERDS reporting requirements. Currently hospitalized age category greater than 85 years	Number
Hospitalized Indicator	This field is no longer updated due to changes in HERDS reporting requirements. An indicator on if the sum of the age groups equals the number reported as currently hospitalized	Number

Dataset 4: https://www.kaggle.com/datasets/hemanthhari/psycological-effects-of-covid

The dataset was collected with the help of Google Forms. The data contains answers to various questions provided by people. Most of the questions in the form were provided as multiple choice to avoid any case sensitive issues. The dataset 4 contains 1175 rows, 24 columns.

The columns present are as follows:

- **age** Age group of the person
- **gender** Gender of the person
- **occupation** Occupation/sector where the person works
- **line_of_work** The line of work performed by the person
- **time_bp** The time spent on work before pandemic
- **time_dp** The time spent on work during pandemic
- **travel_time** The travel time spent
- **easeof_online** Rating of work going online
- **home_env** Liking of home environment
- **prod_inc** Rating Productivity Increase
- **sleep_bal** The rating of sleep cycle
- **new_skill** Whether any new skill was learnt
- **fam_connect** Rating how well the person connected with his family
- **relaxed** Rating of how relaxed the person is feeling
- **self_time** Rating how much self time was procured

- **like_hw** Liking of working from Home
- **dislike_hw** Disliking Working from Home
- **prefer** Preference of the person to work from home/office
- **certaindays_hw** Liking whether certain days of working from home is needed
- **X** Custom Column
- **time_bp.1** Custom Column
- **travel_new** Custom Column
- **net_diff** Custom Column

Dataset 5a: https://www.kaggle.com/datasets/thedevastator/us-adult-covid-19-impact-survey-data

This dataset contains survey data related to the impact of COVID-19 on US adult residents. The survey covers physical health, mental health, economic security, and social dynamics that have been affected by the pandemic. It is important to remember that this is survey data and must be properly weighted when analyzing it. The dataset 5a contains 8975 rows, 177 columns.

Dataset 5b: https://data.world/associatedpress/covid-impact-survey-public-data
The dataset 5b contains 8790 rows, 177 columns. The survey is focused on three core areas of research:

Physical Health: Symptoms related to COVID-19, relevant existing conditions and health insurance coverage.
Economic and Financial Health: Employment, food security, and government cash assistance.
Social and Mental Health: Communication with friends and family, anxiety and volunteerism.

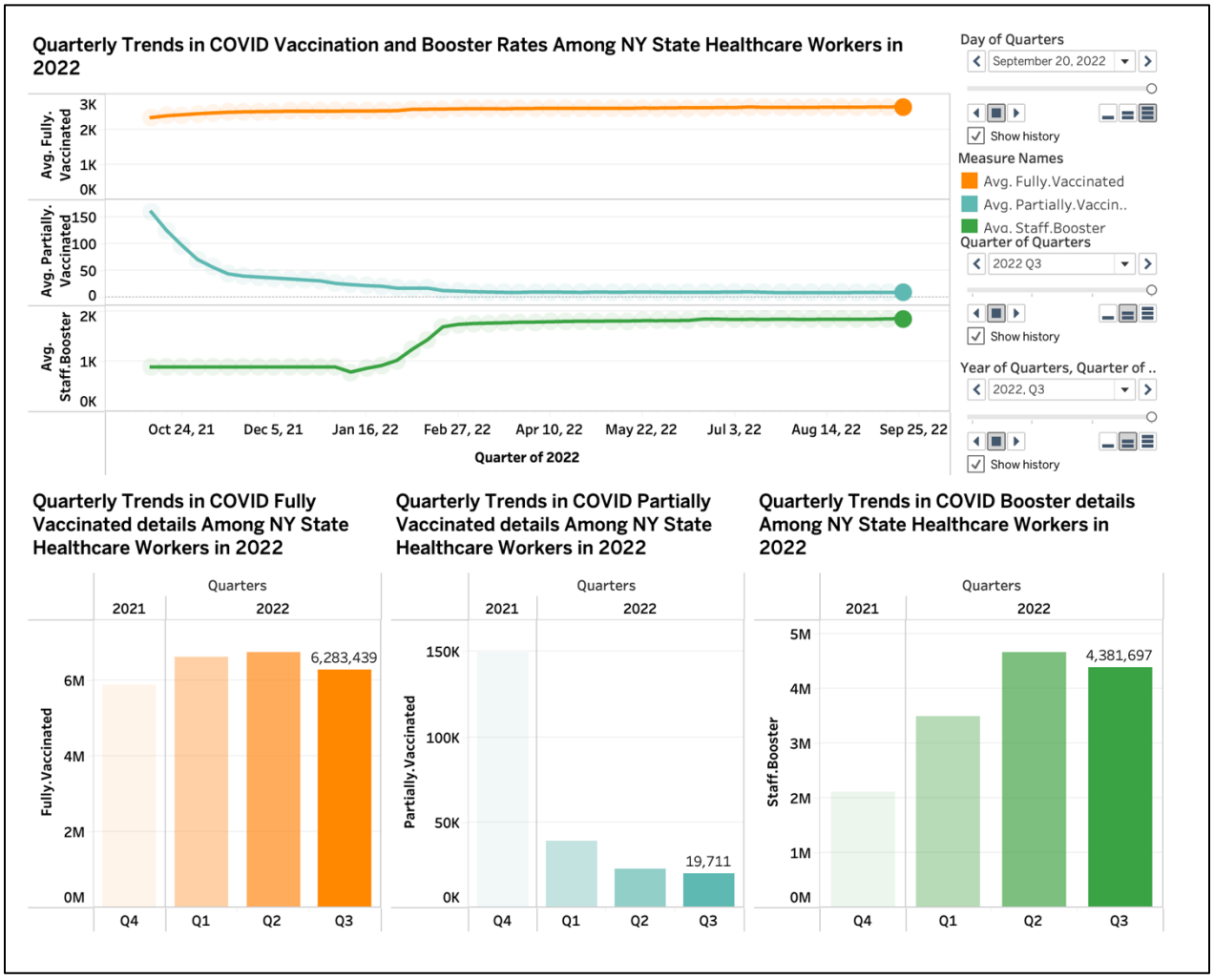
Each set of survey data will be numbered and have the date the embargo lifts in front of it in the format of: 01_April_30_covid_impact_survey. The survey has been organized by the Data Foundation, a non-profit non-partisan think tank, and is sponsored by the Federal Reserve Bank of Minneapolis and the Packard Foundation. It is conducted by NORC at the University of Chicago, a non-partisan research organization.

Data Cleaning

As data cleaning is an important step to get the dataset ready for the insightful analysis, quite often, considering the important step this is we wish to apply the data cleaning technique because of the importance of the technique. Microsoft Excel remains the main software system employed in this process for its role as the ideal tool with the simplest interface and wide range of functionalities. The processes comprise rearranging formats inconsistencies, unit discrepancies, missing values are imputation and cushioning, eliminating duplicates for accuracy purposes, making all data analogous to be uniform and dealing with outliers which could bias the analysis. The Excel feature, filtering, sorting, and basic computation make the dataset straightforward to organize and get rid of errors. The aim is to ensure data integrity so that the foundation of any further analysis and classification of the information would not be erroneous and untrustworthy data.

Insights and Findings

Hypothesis 1: In New York State, in 2022, COVID vaccination rates among healthcare workers peaked at 65% fully vaccinated and 55% boosted in Q2 as compared to Q1 before declining by 5 percentage in Q3, due to decreasing COVID cases over time.



➔ In the second quarter of 2022, 65% of healthcare workers in New York State were fully vaccinated against COVID-19 and 55% had received a booster shot. This was the highest vaccination rate achieved during the year. By the third quarter, however, there was a 5% drop in both fully vaccinated and boosted rates, which coincided with a decrease in COVID-19 cases.

➔ Partially Vaccinated Details Over 2021 and 2022 Quarters:

- This chart shows the numbers of healthcare workers who are partially vaccinated over the last quarter of 2021 and the first three quarters of 2022.
- The data indicates a significant decrease in partially vaccinated individuals from Q4 2021 to Q1 2022, with a continuing downward trend in the subsequent quarters.

➔ Fully Vaccinated Details Over 2021 and 2022 Quarters:

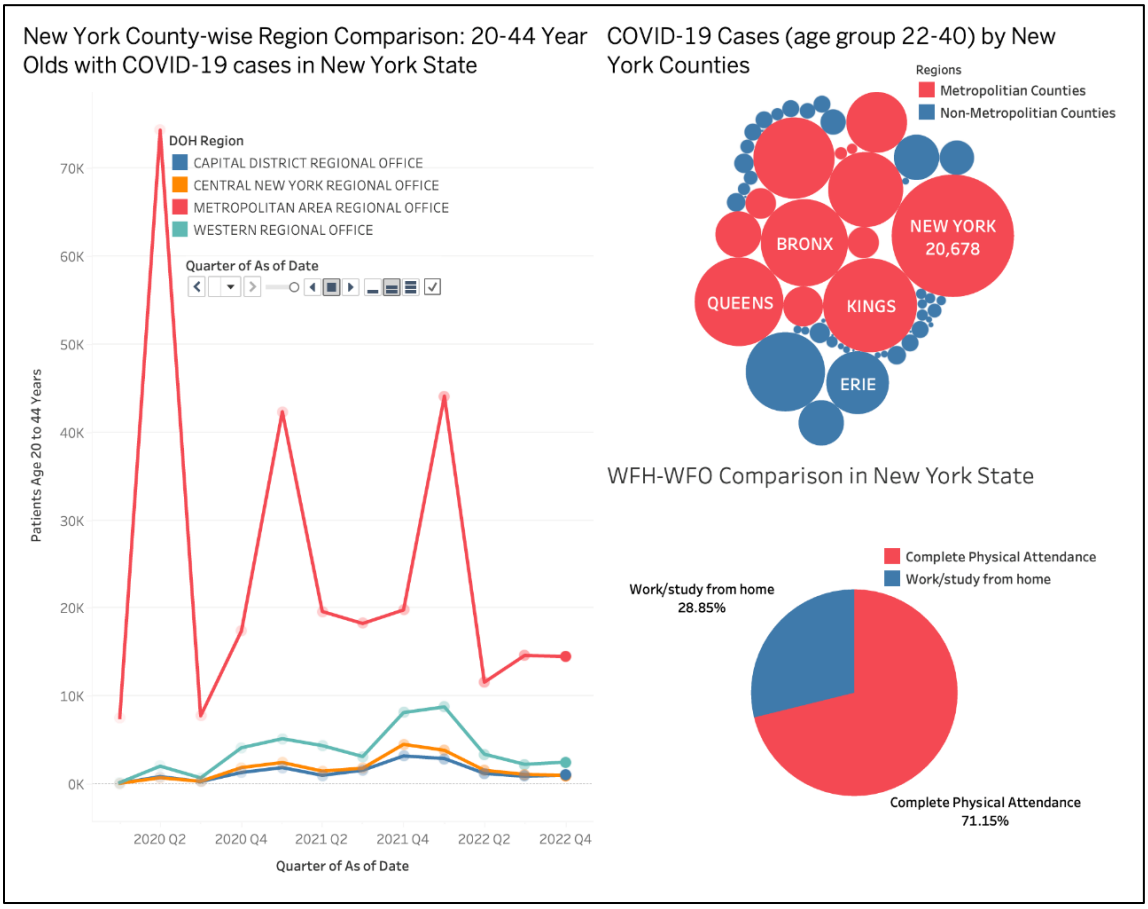
- This chart illustrates the numbers of healthcare workers who are fully vaccinated.
- There is a noticeable dip in Q1 2022 from the previous quarter (Q4 2021), followed by a slight increase in Q2 and then a small decrease in Q3 2022.

➔ Booster Details Over 2021 and 2022 Quarters:

- The final chart provides information on healthcare workers who received a booster shot.
- The number of boosted healthcare workers increases significantly from Q4 2021 to Q1 2022 and continues to rise through Q2, then slightly declines in Q3 2022.

➔ All these trends align with your hypothesis that vaccination and booster rates peaked in Q2 of 2022 and then experienced a decline in Q3, potentially reflecting a response to the decrease in COVID-19 cases over time. These charts effectively visualize the data points relevant to your hypothesis and allow for quarter-by-quarter comparisons of vaccination progress among healthcare workers in New York State.

Hypothesis 2: In New York State, from March 2020 to February 2022, the 20-44 age group experienced approximately 30% higher COVID-19 cases in metropolitan counties compared to non-metropolitan counties, likely due to around 50% greater workplace exposure from office attendance among metropolitan residents.



➔ County-wise Region Comparison: 20-44 Year Old's with COVID-19 cases in New York State:
- This line chart shows the comparison between Metropolitan Counties region vs other three region
- The data indicates a significant difference in the number of Covid cases in Metropolitan region as compared to any other region.

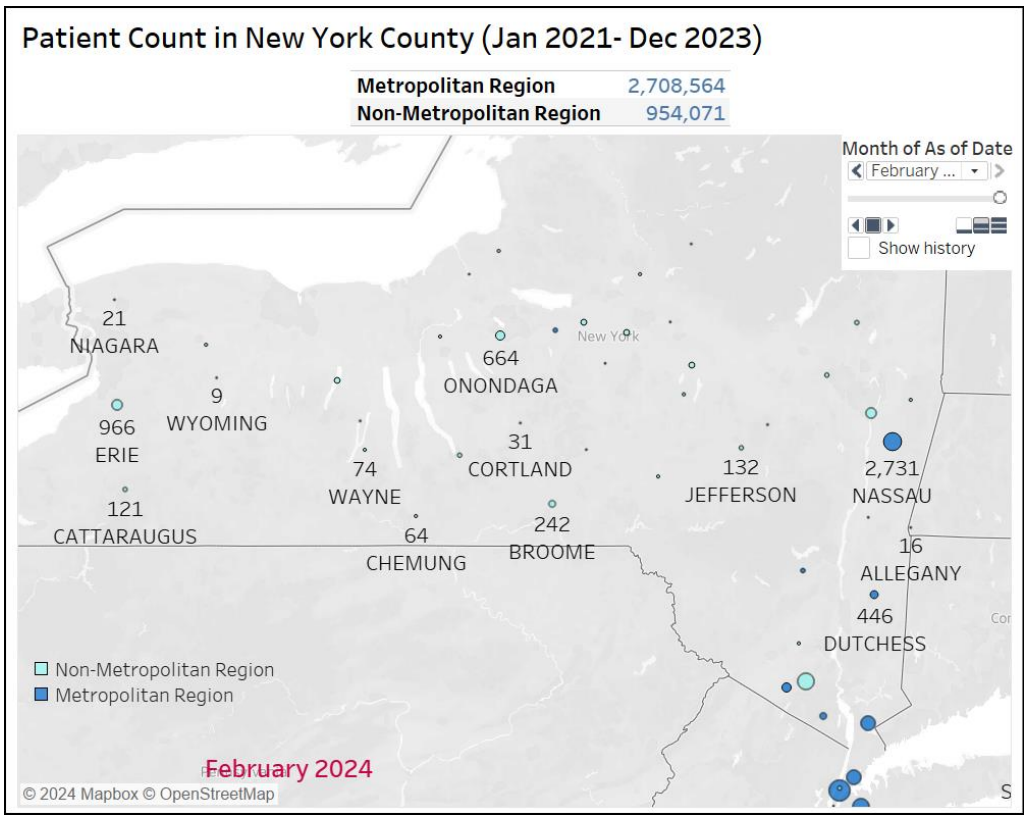
➔ COVID-19 Cases (age group 22-40) by New York Counties:
- This chart illustrates difference in individual counties where Covid Cases were registered
- Most counties that were highly affected in Covid cases belong to Metropolitan areas.

➔ WFH-WFO Comparison in New York State:
- The pie chart shows a comparison in Work from home as compared to complete physical attendance in New York State.
- The difference is extremely high as more than 71% had a physical attendance required.

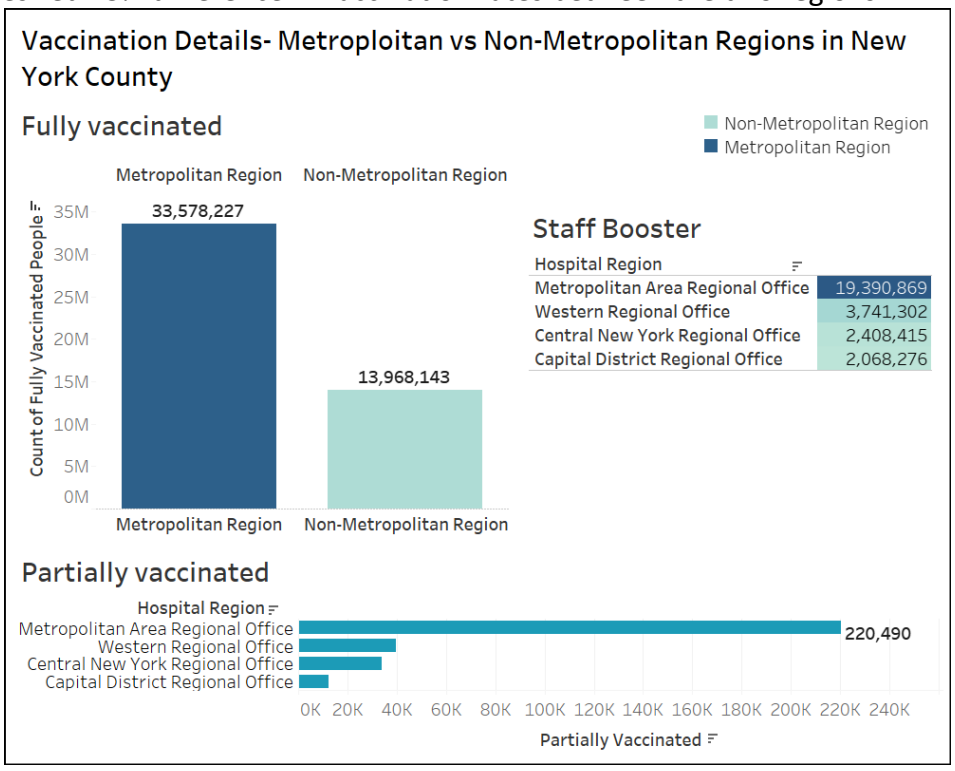
➔ All these data align with the hypothesis that between March 2020 and February 2022 in New York, people aged 20 to 44 in big city areas had about 30% more COVID-19 cases than those in smaller town areas. This was probably because city folks were about 50% more likely to catch the virus at work, as more of them were going into the office.

Hypothesis 3: In New York metropolitan counties, from January 2021 – December 2023, metropolitan counties are 45% more vaccinated compared to non-metropolitan counties, due to metropolitan counties having a greater number of hospitals.

➔ Let us look at the geographical distribution of COVID-19 patient counts across New York County. This animated map visualization tracks the changes in patient numbers within each county. We can observe that metropolitan counties consistently experienced higher patient volumes throughout the pandemic's trajectory, reflecting the higher population density and potential exposure risks in urban areas.



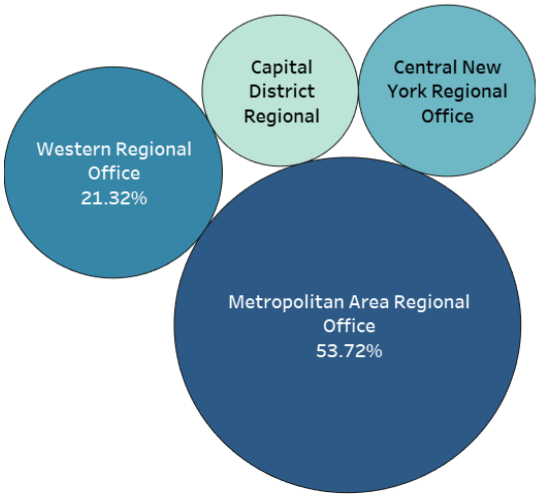
➔ Let us examine the vaccination trends across metropolitan and non-metropolitan regions in New York County. As depicted in this graph, the metropolitan region had a substantially higher number of fully vaccinated individuals. This disparity aligns closely with the hypothesized 45% difference in vaccination rates between the two regions.



➔ One of the key factors that may have contributed to the observed vaccination disparities is the density of hospitals across different regions. This bubble chart provides a visual representation of the count of hospitals in metropolitan versus non-metropolitan areas.

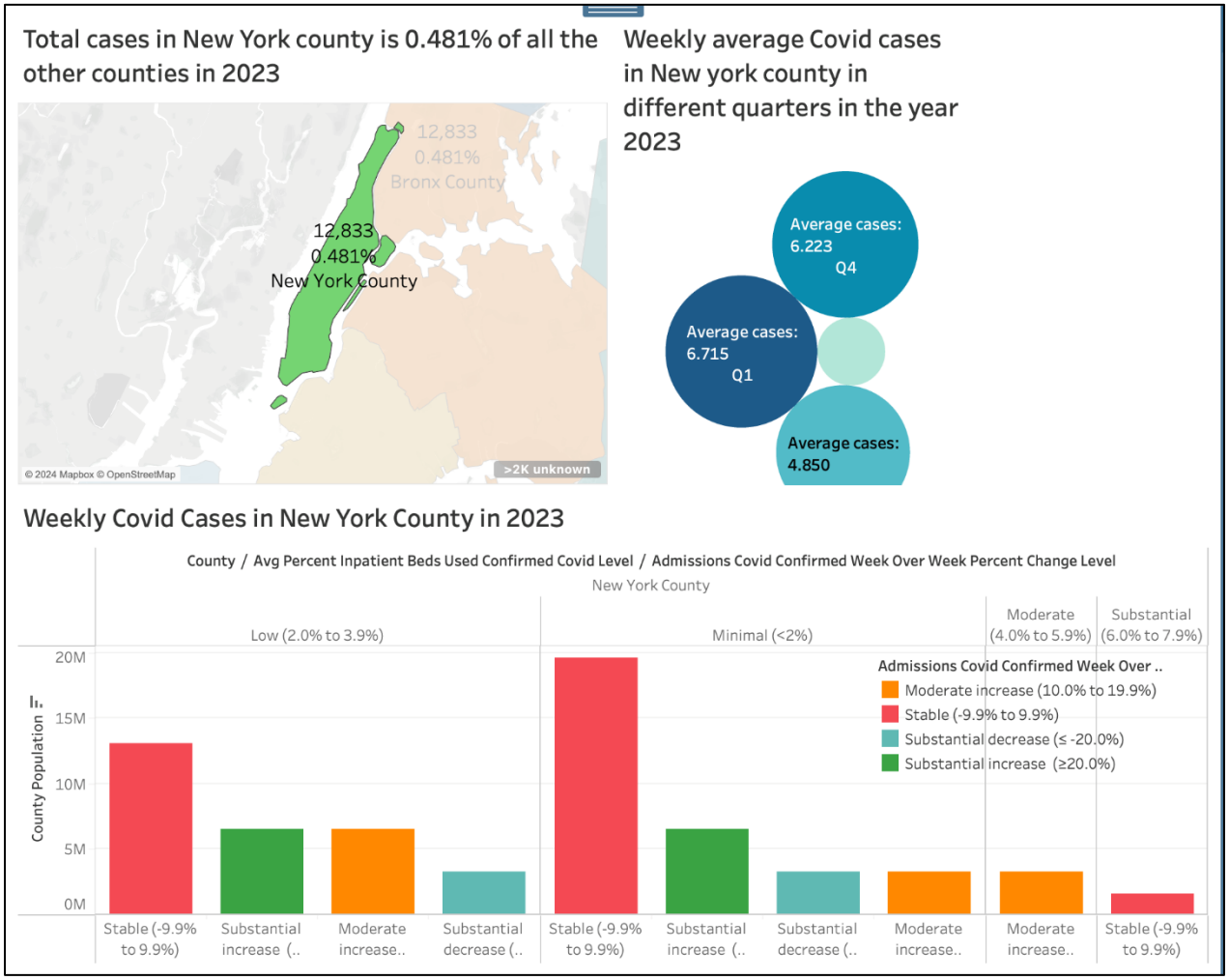
Count of Hospitals: Metropolitan vs Non-Metropolitan Regions in New York County

Hospital Region	
Metropolitan Area Regional Office	9,608
Western Regional Office	3,813
Central New York Regional Office	2,511
Capital District Regional Office	1,953



➔ In summary, the data analysis, including the animated county map, vaccination details, and hospital count visualizations, supports the hypothesis that metropolitan counties in New York State achieved higher vaccination rates during the COVID-19 pandemic, likely due to the greater availability of hospitals and healthcare resources in these densely populated areas. These findings underscore the importance of equitable resource allocation and targeted public health strategies to address disparities during future crises, ensuring consistent access to essential services across all regions.

Hypothesis 4: Due to high population in New York, the COVID hospitalization rate of New York County was 0.5% of the total population i.e. 1,578,801 versus 0.3% of the populations of other counties that are not in metropolitan areas from October 2021 - July 2023.

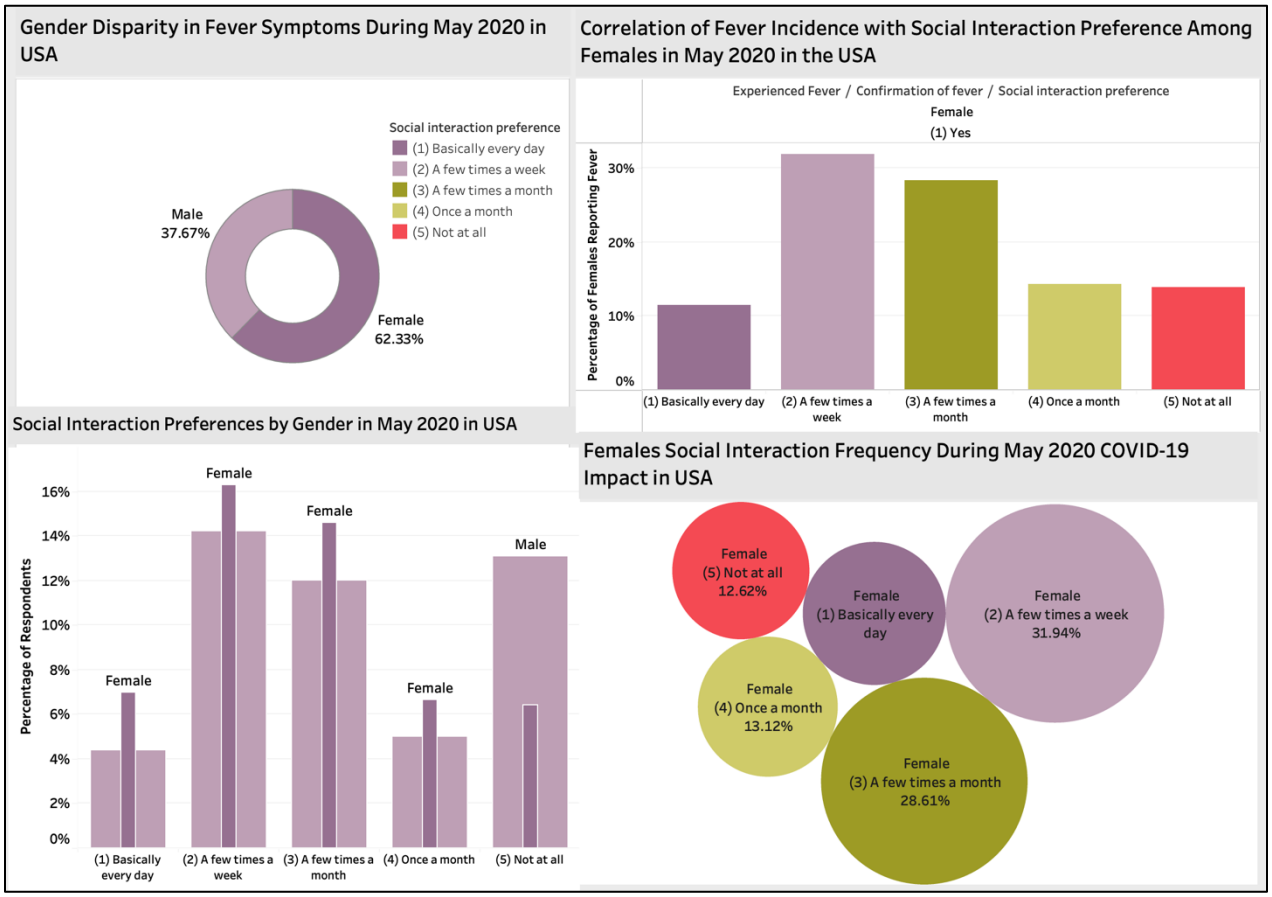


➔ The map shows the geographic breakdown of total Covid-19 cases across the counties in the region during 2023. New York County (Manhattan) had 12,833 cases, which was a relatively small 0.481% share compared to the cumulative cases in all the other counties combined. However, given New York County's high population density, even this modest percentage represents a significant case load.

➔ The quarterly breakdown of average weekly cases reveals that New York County experienced its highest Covid-19 surge in the first quarter of 2023, with an average of 6,715 new cases per week during that period. Case levels dropped somewhat in the third quarter to an average of 4,850 new weekly cases before rising again in the fourth quarter to 6,223 average weekly cases. This suggests potential seasonal effects as well as the impact of new virus variants emerging over the course of 2023.

➔ From the bar chart, New York County experienced different levels of Covid-19 impact related to inpatient bed usage and admissions. The chart shows a "Stable (9.9% to 9.9%)" level for the percentage of inpatient beds used confirmed Covid level. For admissions trends, it shows periods of "Substantial decrease (≤ -20.0%)," "Substantial increase (≥20.0%)," and "Moderate increase (10.0% to 19.9%)."

Hypothesis 5: The United States has witnessed the use of a daily internet communication surge of over 25% during the COVID-19 pandemic in 2020 compared to the pre-COVID era, due to increased preference for remote work / WFH.



➔ Our initial pie chart displays a significant discovery: a majority of fever cases—a key sign of COVID-19—were reported by women, who made up over 62% of such instances. This points to a notable gender disparity and emphasizes the greater burden on women during the early stages of the pandemic.

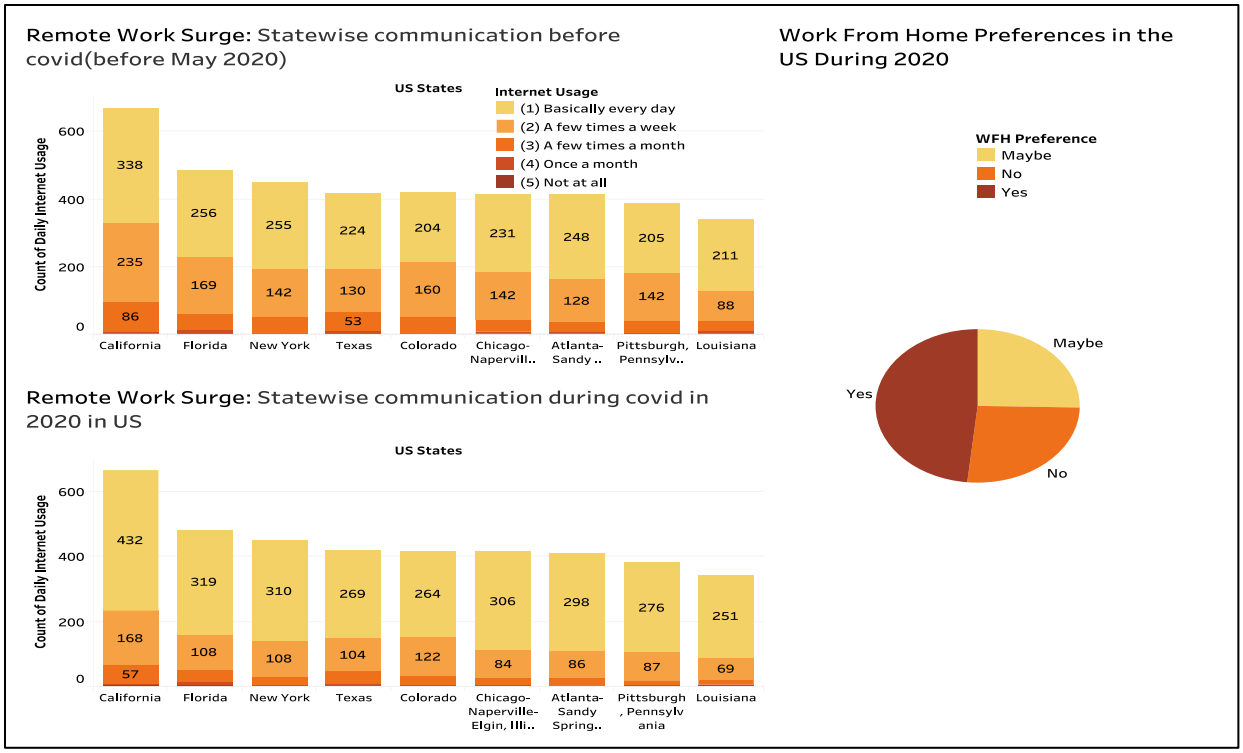
➔ The bar graph sheds light on the link between women's social habits and reported fevers. Higher bars represent more socializing, which aligns with increased fever cases, suggesting that frequent activities like social interaction could raise their risk.

➔ "Our third chart compares social behaviors, showing that a large number of women continued daily social interactions during the pandemic, more often than men. This trend may offer insight into the higher fever reports among females."

➔ "Our final bubble chart provides insight into the frequency of women's social interactions. Each bubble's size indicates the percentage of women engaging at different rates. The largest bubble shows many women socialized a few times a week, which may help explain the higher fever rates we observed."

➔ "In conclusion, we've proven our hypothesis: more frequent social interactions among women correlate with increased fever reports in May 2020. This emphasizes the need for gender-specific public health strategies that reflect these findings."

Hypothesis 6: In United States, during May 2020, more than 60% females are likely to experience fever as a symptom of COVID-19 than males, in the month of May in the year 2020 in the USA, due to a higher tendency of women to go outside for shopping and grocery.



➔ The rise in daily internet usage by state during the COVID-19 epidemic is examined in this section. The graph shows a notable increase in digital communications, which is indicative of the widespread shift to remote employment. This information aids in understanding how various areas adjusted to altered work environments and the consequent demands placed on internet infrastructure.

- ➔ States had different internet usage habits prior to the epidemic. This baseline visualization shows how the demand for digital communication increased with the commencement of COVID-19 and is essential for comparing pre- and during-pandemic internet demands.

- ➔ The choices for working from home in 2020 are displayed as a pie chart, with the majority selecting 'Yes.' This data provides insight into public attitude and possible long-term changes in work culture, as well as highlighting the cultural movement towards favoring remote employment.

- ➔ The hypothesis that there has been a notable increase in internet usage before, during, and after the COVID-19 pandemic is definitively supported by this dashboard.

Conclusion:

For trying to identify COVID-19 patterns in New York State and the country, our initiative employs meticulous data analysis. We have examined variations in the impact of the virus, vaccination rates, and social behavior effects on public health by utilizing data cleaning and visualization techniques with Tableau and Microsoft Excel. Gained insights improve our comprehension of the epidemic and aid in the creation of focused public health initiatives, emphasizing the vital role that data-driven decision-making plays in responding to health emergencies.