HI PROJECT ASSIGNMENT II

NAME: ESWAR RAO RAMINENI
MID: M15270716
DATASET: Cincinnati Health Care Centers

DATASET TYPE: Both Structured and Unstructured

DATA SOURCE: https://data.cincinnati-oh.gov/Thriving-Neighborhoods/Cincinnati-Health-Department-Health-Care-Centers/v8yh-wpss/data

FILE NAME: CincyHealthCenters.csv

For my research, I've chosen the "CincyHealthCenters.csv" dataset, which I've previously utilized in module 5. This dataset is structured, containing both textual and numerical values. It includes geographic details like addresses, cities, states, latitude, and longitude, along with statistical figures such as average lab fees and medical charges. These numeric values are crucial for conducting data analysis. Moreover, the dataset also features a "Service Provided" column containing unstructured data in the form of sentences. This presents an opportunity to employ text mining techniques for data analysis. To enhance the dataset, I've updated its original source with three additional columns: average medical fee, lab fee, and admission, utilizing Google resources. These added fields will facilitate a more detailed analysis to address the research questions at hand. Additionally, by extracting the "Service Provided" column data into a text file, I've created an unstructured dataset. The selection of the "CincyHealthCenters.csv" dataset allows me to explore both structured and unstructured data, enabling a comprehensive investigation for this research.

Data Variables:

VARIABLE NAME	DESCRIPTION	DATA TYPE
HEALTH_CENTER	Name of the healthcare center	Text/String
HEALTH_CENTER_CODE	Unique code for the healthcare center	Text/String
ADDRESS	Street address of the center	Text/String
STREET_NAME	Name of the street	Text/String
SUFX	Suffix for the street (if any)	Text/String
CITY	City where the center is located	Text/String
STATE	State where the center is situated	Text/String
ZIPCODE	Zip code of the location	Text/Numeric
PHONE	Contact phone number	Text/String
CARE_TYPE	Type of healthcare service provided	Text/String
MONDAY_OPEN to FRIDAY_CLOSE	Timings of the center for each day of the week	Time/Text
SPECIAL_HOURS_NOTES	Any special notes regarding operating hours	Text/String
SERVICES_PROVIDED	Description of services offered	Text/String
CENTER_DESCRIPTION	Description of the healthcare center	Text/String
INSURANCE_ACCEPTED	Accepted insurance types	Text/String
TITLE_X_SERVICES	Types of specialized services offered	Text/String
NEIGHBORHOOD	Neighborhood where the center is located	Text/String
LATITUDE	Geographic coordinate - Latitude	Numeric
LONGITUDE	Geographic coordinate - Longitude	Numeric
ADULT to SAFE_PLACE	Services available (Yes/No indicators)	Boolean/Text

HI PROJECT ASSIGNMENT II

NAME: ESWAR RAO RAMINENI
MID: M15270716

AVERAGE_MEDICAL_COST	Average medical cost for Health care hospital	Numeric
ADMISSION_FEE	Admission fee for Health care hospital	Numeric
AVERAGE_LAB_FEE	Average lab fee for Health care hospital	Numeric

RESEARCH QUESTIONS:

1. What types of healthcare services are predominantly available in Cincinnati?

2. Which neighborhoods have varying levels of nursing home availability and areas needing improvement?

3. How can good healthcare hospitals with lower admission and medical costs be identified?

4. What are the most popular services provided by healthcare facilities in Cincinnati?

5. Is there a statistically significant correlation between geographic location and average lab fees in Cincinnati?

6. How to identify the best budget hospital near to the patient?

7. Find the statistics of health care with average admission fee, lab fee and medical cost and find which is most expensive treatment?

PREPARATION OF DATA FOR ANALYSIS:

DATA UPDATE:

I enhanced the dataset for statistical validation by adding three new columns—average medical cost, admission fee, and average lab fee. These values were gathered from reliable sources found through Google. This updated data is particularly crucial for addressing research question 3 and will be utilized in both R and Tableau for analysis.

DATA CLEANING:

Within the dataset, there exists a column named 'services provided' that contains unstructured data. To tackle research question 4, I implemented three text mining cleaning techniques. These techniques involved removing stop words, punctuation, and excessive spaces in the text.

DATA CONVERSION:

After cleaning the unstructured data, I converted it to lower case. This step is essential for conducting statistical analysis in R, especially for determining word frequencies. Without converting the text to lowercase, eliminating duplicate words wouldn't be feasible. Subsequently, I calculated data frequencies, constructed a document-term matrix, and stored it in a separate file for further analysis.

HI PROJECT ASSIGNMENT II

NAME: ESWAR RAO RAMINENI
MID: M15270716

SLOUTIONS TO RESEARCH QUESTION USING R VISUALIZATIONS:

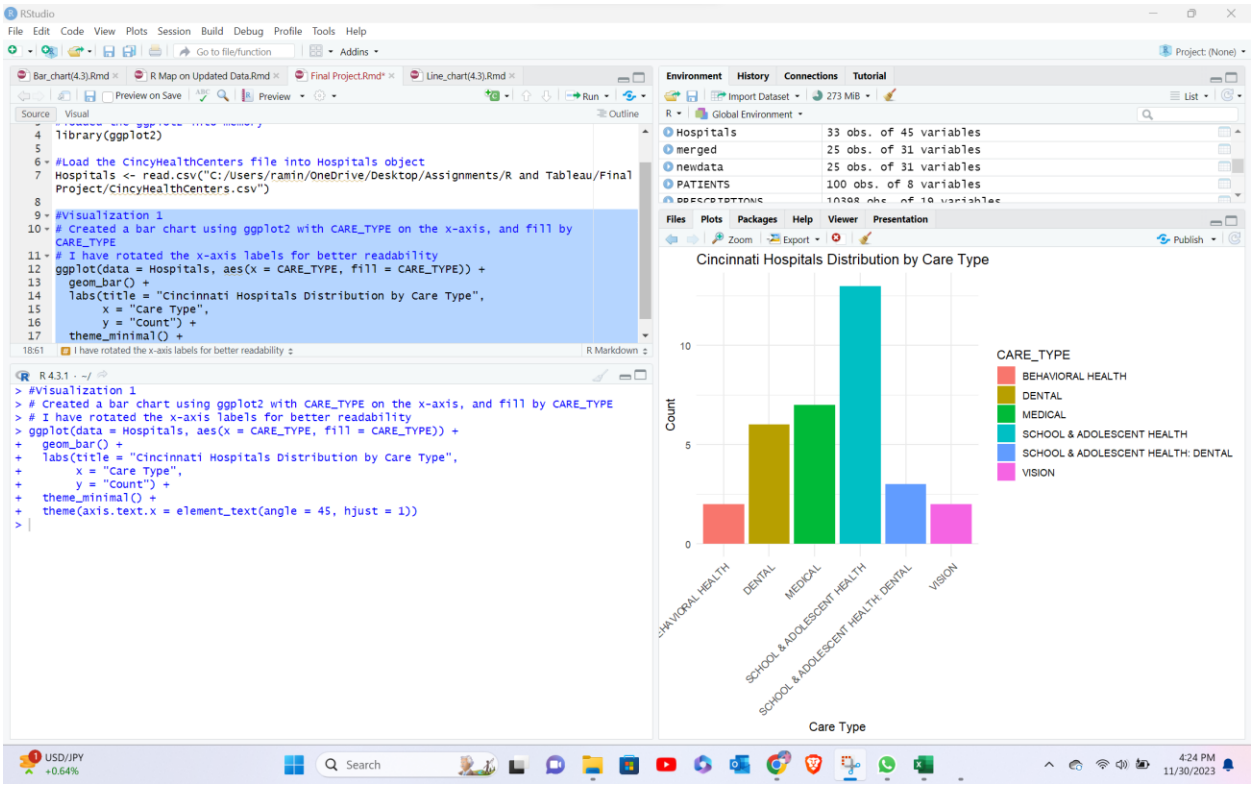
ROW	TYPE	R VISUALIZATIONS	SOLUTION
1.	Bar	Cincinnati Hospital Distribution by Care type	In R, using the ggplot package, I crafted a bar chart that offers a clear view of the Cincinnati Health Centers dataset. The goal was to address question one by tallying the occurrences of various healthcare types in Cincinnati. On the chart, the X-axis illustrates different care types with a simple appearance and color codes. Meanwhile, the Y-axis shows the count or frequency of these health cares. The observations reveal six distinct types of health care. Among these, 'School & Adolescent Health' stood out as the most frequent, while 'Behavioral and Vision Health' appeared least in number.
2.	Line	Frequency of Nursing Services by Neighborhood	In this R visualization, I created a line chart to show the presence of nursing services across various neighborhoods in Cincinnati. The chart helps answer research question 2. On the X-axis, you'll find the different neighborhoods of Cincinnati where various healthcare services are located. Meanwhile, the Y-axis indicates the frequency of nursing services. This chart has two lines: an orange line and a light green line. The orange line represents health care without nursing services, while the light green line represents health care that do have nursing services. What stands out in the visualization is that neighborhoods like HYDE PARK, LOWER PRICE HILL, and MADISONVILLE have a significant number of health cares. However, most of these health care in these areas do not offer nursing services. The graph distinctly shows the orange line being higher than the green line in these regions. This indicates that implementing or providing nursing services in these hospitals could significantly improve healthcare services for the people of Cincinnati. By enhancing nursing services in these areas, it could lead to better healthcare accessibility and quality for the community.
3.	Scatter Plot with regression line	Visualization for Geographic Location vs. Average Lab Fee	This visualization provides a solution for question 5. I conducted a scatter plot analysis representing the relationship between geographic location (longitude and latitude) and the average laboratory fees. The scatter plot had longitude on the x-axis and latitude on the y-axis, while the lab fees were indicated by various colors on the graph. Upon analysis, the intercept of the plot indicated that when the longitude is zero, the average predicted latitude is around 47.02. However, the p-value for the slope (0.274) suggested that the association between longitude and latitude wasn't statistically significant at the commonly accepted significance level of 0.05. The R-squared value, standing at 0.0384, signified that only about 3.84% of the variance in latitude could be explained by variations in longitude alone. This implies that the model based solely on longitude inadequately explains the latitude variability. Further, the adjusted R-squared value, considering the number of predictors in the model, was 0.007377, indicating an extremely weak fit of the model to the dataset. This suggests that the model has little explanatory power for latitude based on longitude. An analysis of residuals (Minimum, 1st Quartile, Median, 3rd Quartile, Maximum) helps in evaluating the model's goodness of fit by showing the differences between observed and predicted latitude values. The trend line on the plot passed through light, green-colored dots, indicating an average lab fee of 346. Consequently, we can infer that the most common average lab fee in Cincinnati is around 346, based on this trend line. External Factors that affect the lab fee, we observe that labs are spread across the city with different locations, sometimes lab equipment has issues and test subjects will be sent to another lab which is also increases the lab fee in hospitals.
4.	Map	Visualization for health care data on maps	To address research question 6, I utilized the leaflet package in R to create a visualization that displays various types of healthcare facilities across Cincinnati. The visualization employs yellow circles with black borders to represent these healthcare centers on the map. Each circle is interactive and displays a label when clicked, showing the type of hospital along with details about admission and medical costs. The use

NAME: ESWAR RAO RAMINENI
MID: M15270716

			of maps offers an optimal way to identify healthcare options that fit within an individual's budget and are conveniently located. By visually plotting these healthcare facilities on a map, individuals can easily identify the type of care available, assess associated costs, and find the nearest healthcare center. This visualization provides a user-friendly way to make informed decisions about healthcare options based on both affordability and proximity to one's location.
5.	Word Cloud	Word cloud for Services in Health Care	To address research question 4, I delved into the unstructured data found in the 'services provided' column within the Cincinnati Health Center dataset (CinyHealthcenter.csv file). Employing text mining techniques, I implemented several steps to refine the data for analysis. This involved the removal of stop words, punctuation marks, excessive spaces, and conversion of text to lowercase. The resultant refined dataset enabled a detailed analysis. I created a data frame that tallied the frequency of repeated words within the services provided. Through this analysis, it became apparent that 'nursing' and 'pediatrics' emerged as the most frequently occurring services, each having a frequency count of 18. From this data, a notable conclusion can be drawn nursing and pediatrics services were notably popular among the healthcare services provided in Cincinnati. The high frequency of these services suggests a significant demand for or emphasis on these healthcare offerings within the region.

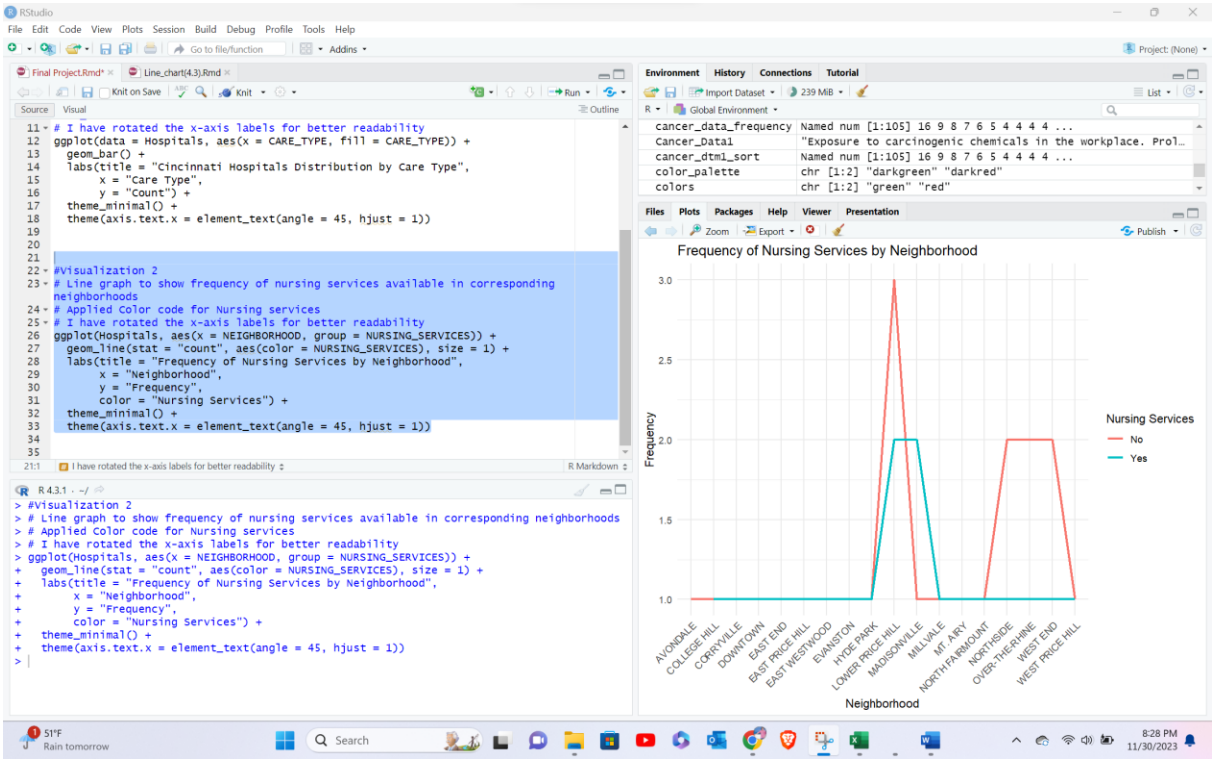
BELOW ARE THE R VISUALIZATIONS:

Visualization 1:

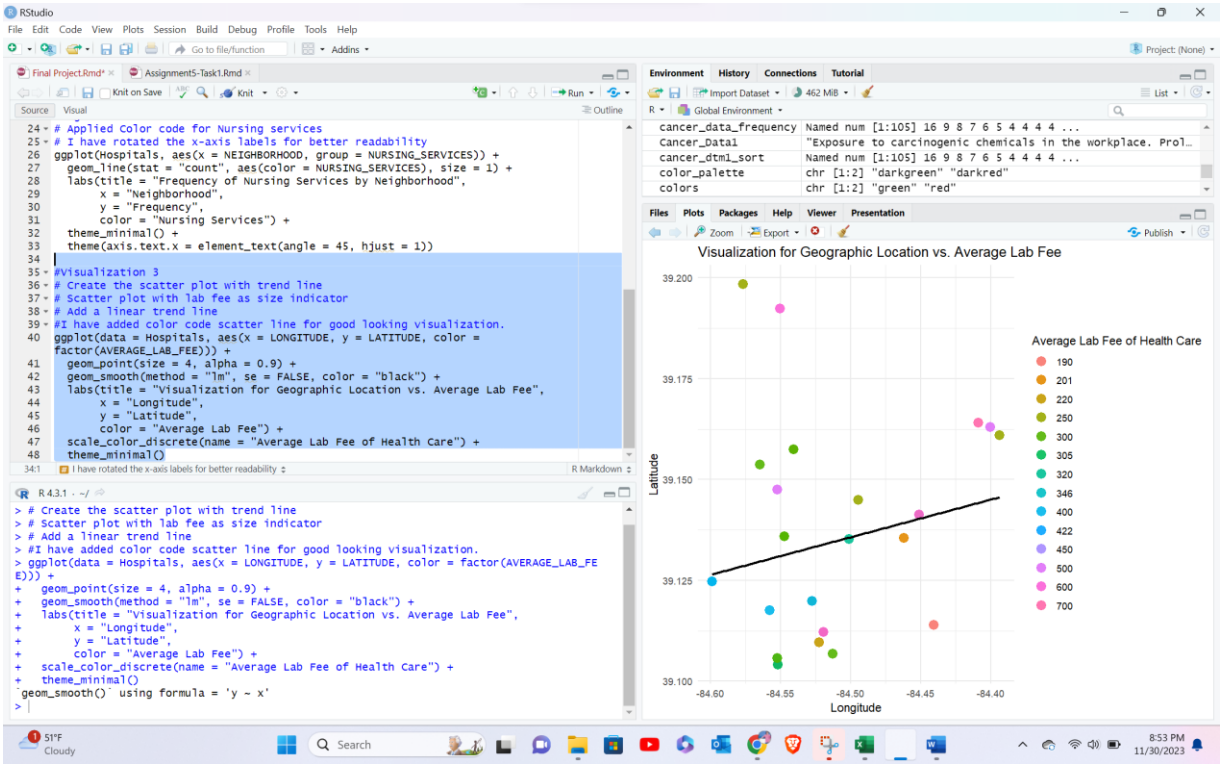


HI PROJECT ASSIGNMENT II

NAME: ESWAR RAO RAMINENI
MID: M15270716
Visualization 2:



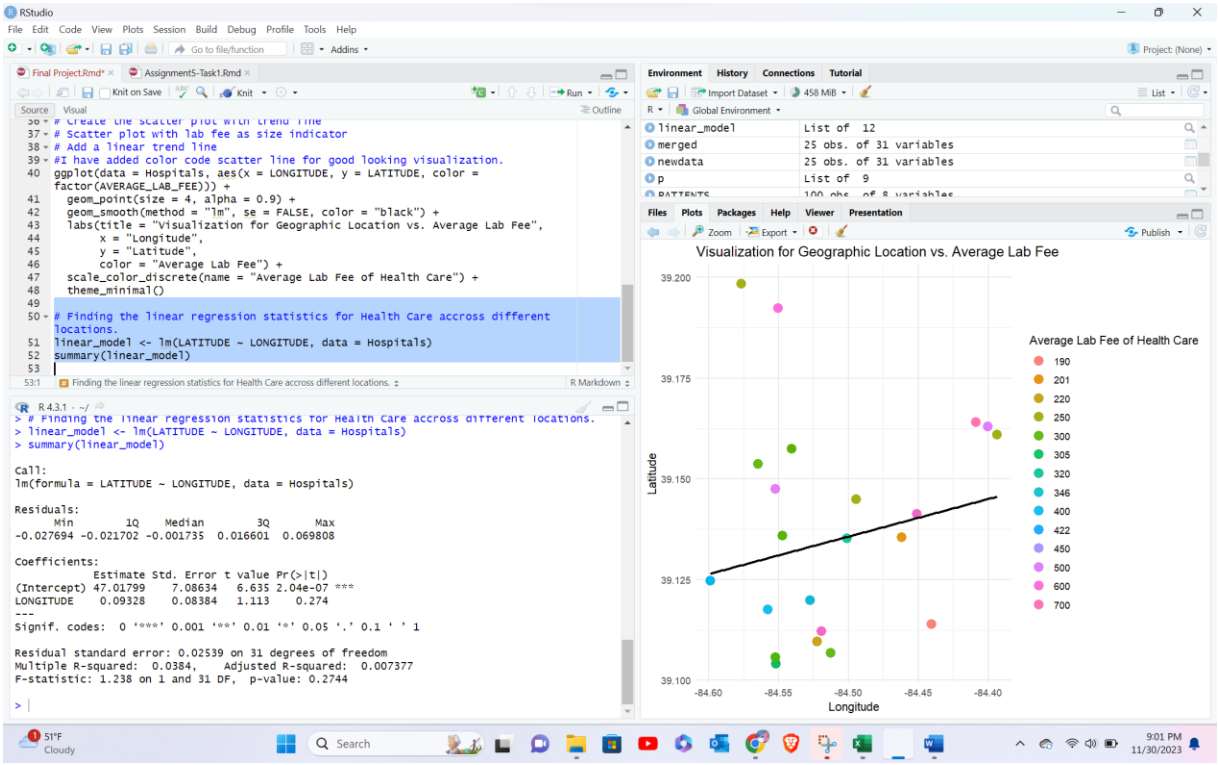
Visualization 3: Step 1: Creation of scatter plot with trend line



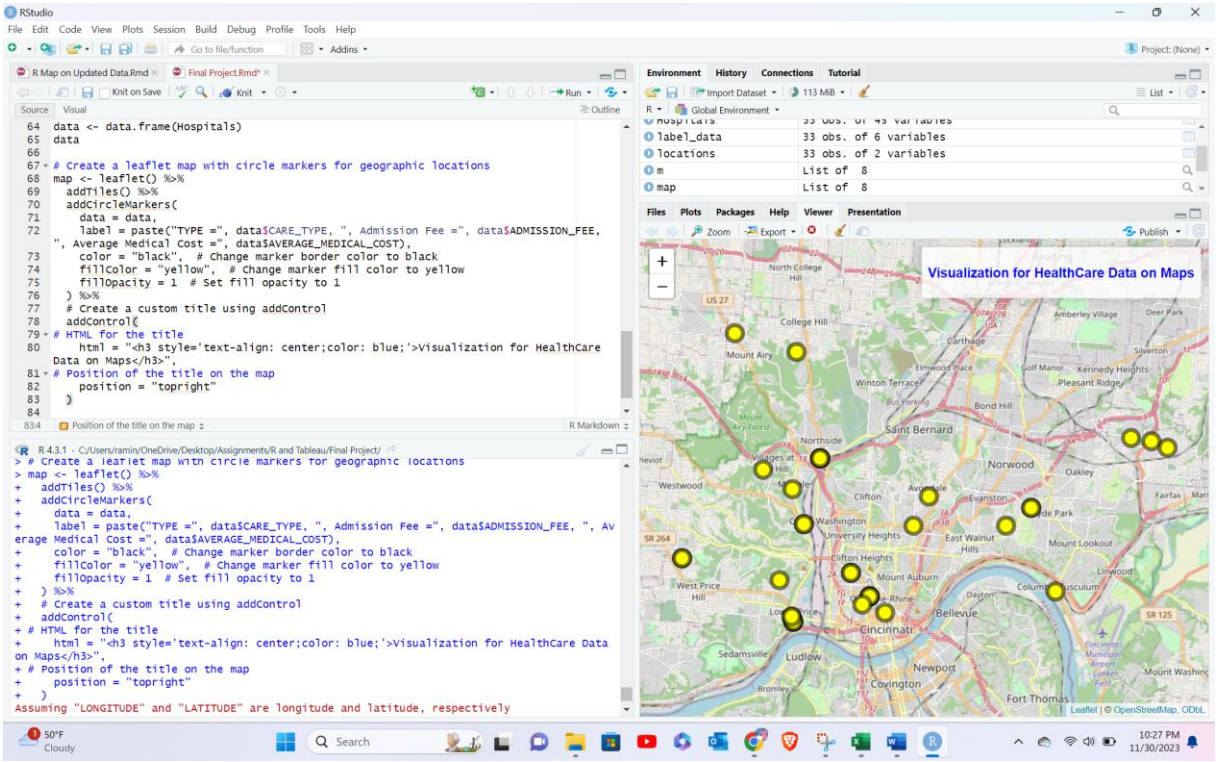
HI PROJECT ASSIGNMENT II

NAME: ESWAR RAO RAMINENI
MID: M15270716

Step 2: Finding regression line statistics.

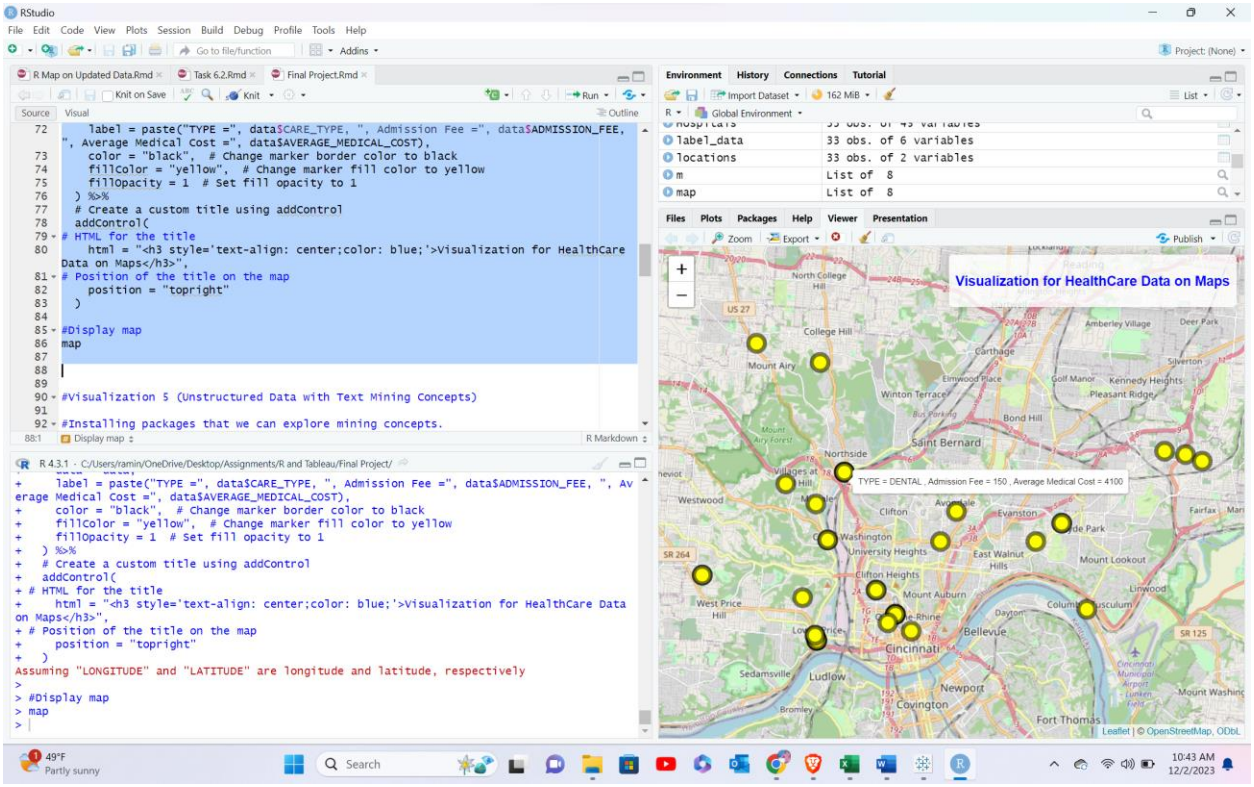


Visualization 4:

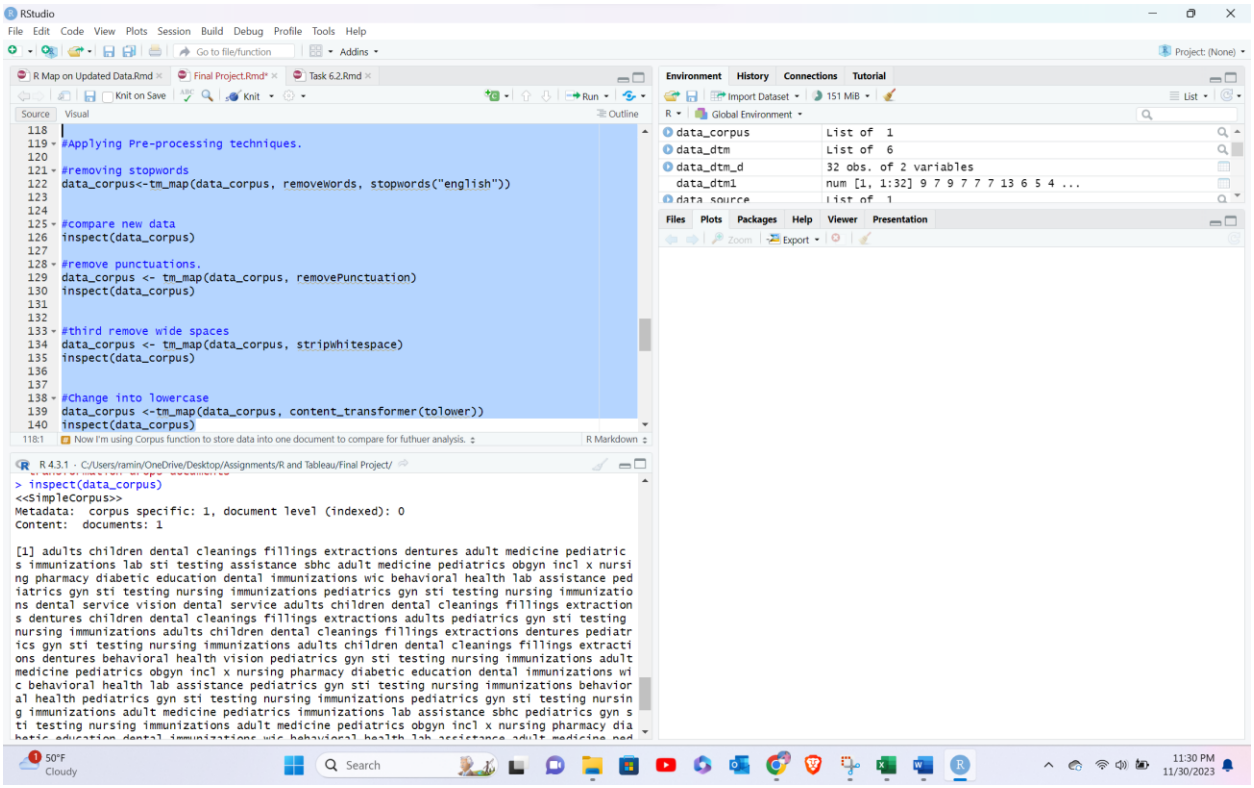


HI PROJECT ASSIGNMENT II

NAME: ESWAR RAO RAMINENI
MID: M15270716



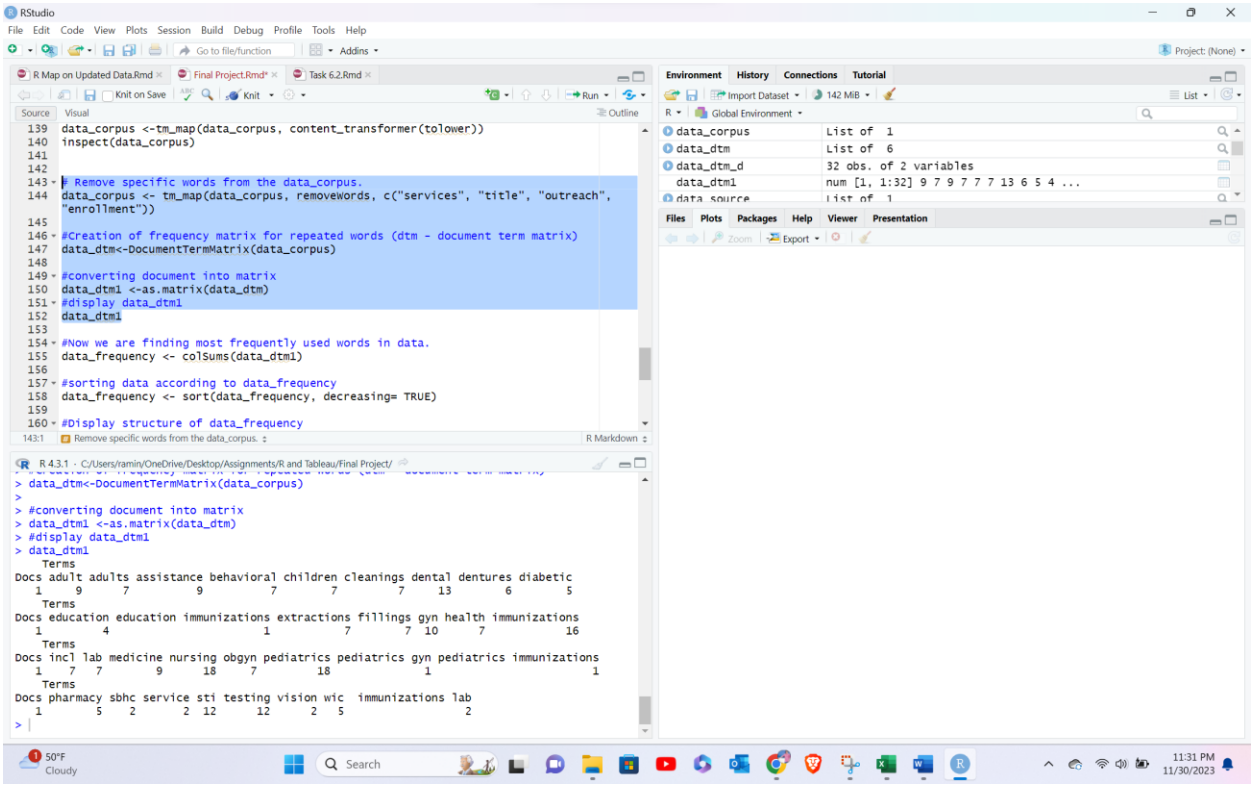
Visualization 5: Step 1: Text mining or data cleaning



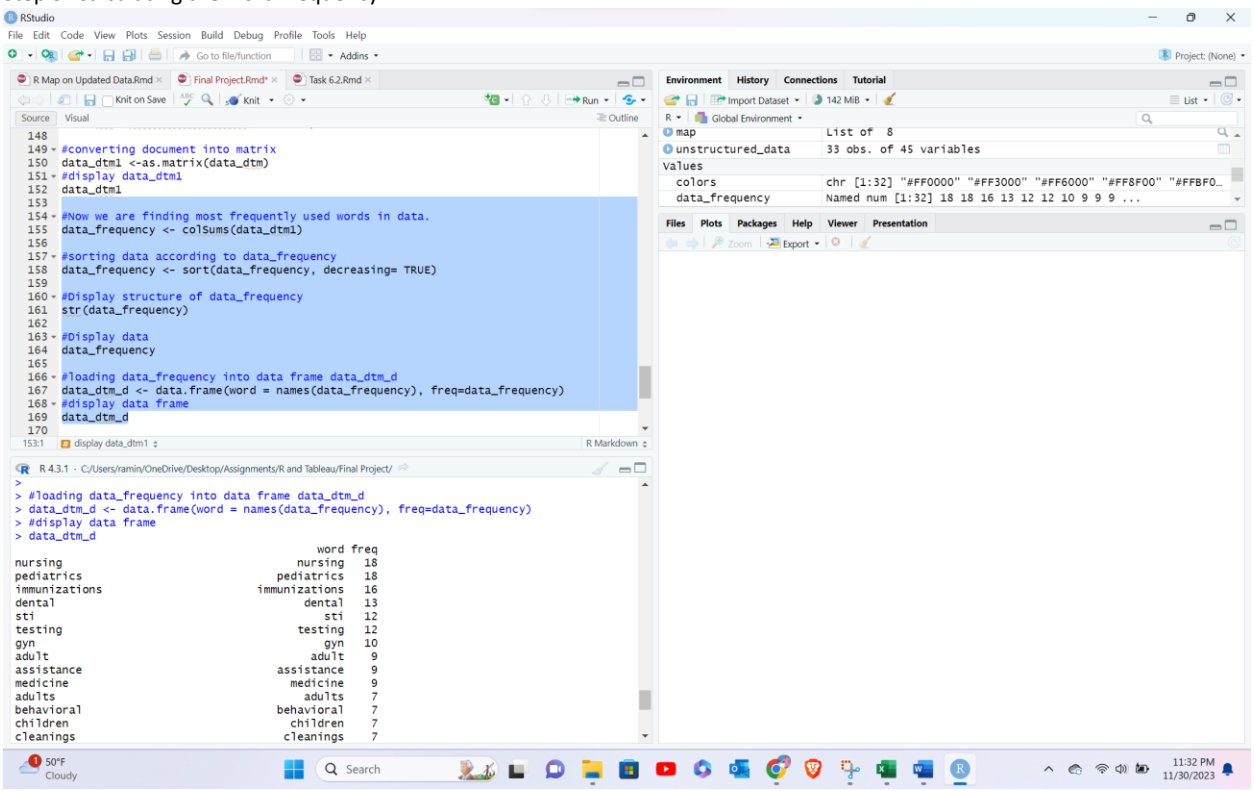
HI PROJECT ASSIGNMENT II

NAME: ESWAR RAO RAMINENI
MID: M15270716

Step 2: Converting document into matrix.

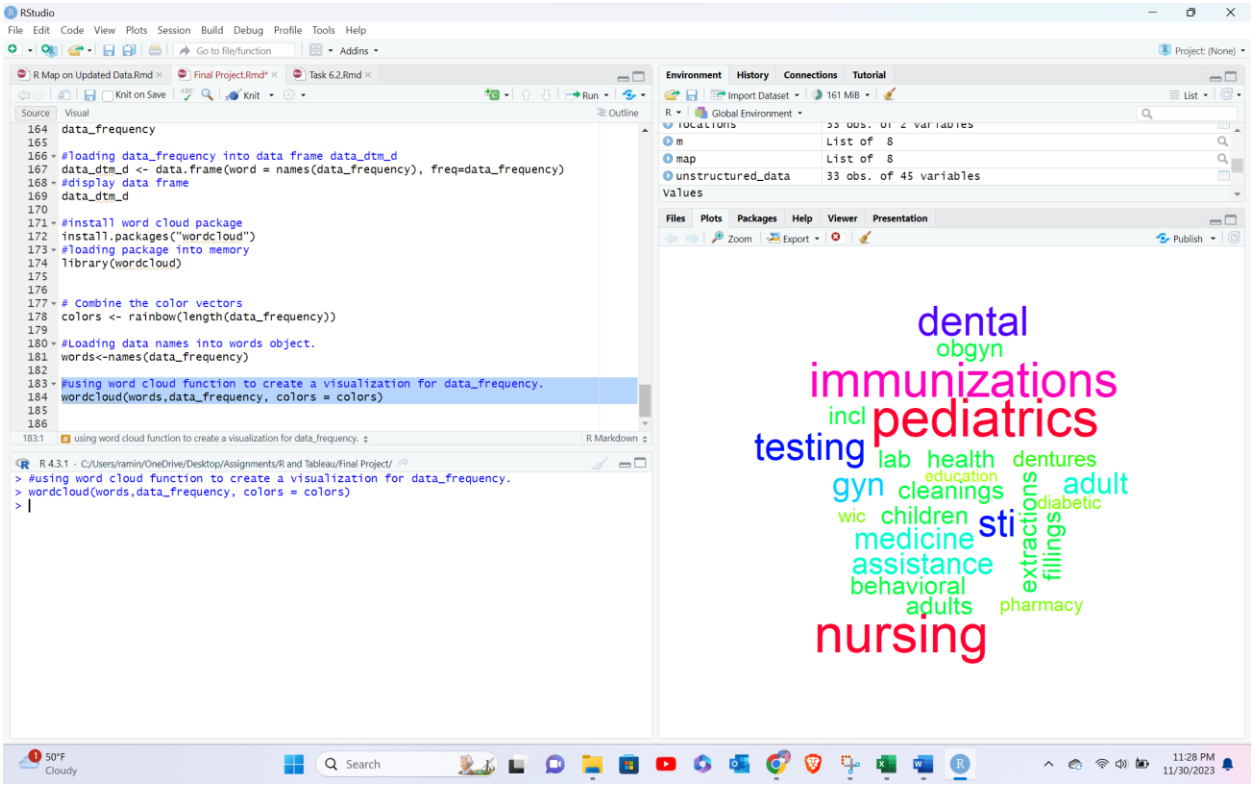


Step 3: Calculating the word frequency.



HI PROJECT ASSIGNMENT II

NAME: ESWAR RAO RAMINENI
MID: M15270716
Step 4: visualize the word cloud.



HI PROJECT ASSIGNMENT II

NAME: ESWAR RAO RAMINENI
MID: M15270716

SLOUTIONS TO RESEARCH QUESTION USING TABLEAU VISUALIZATIONS:

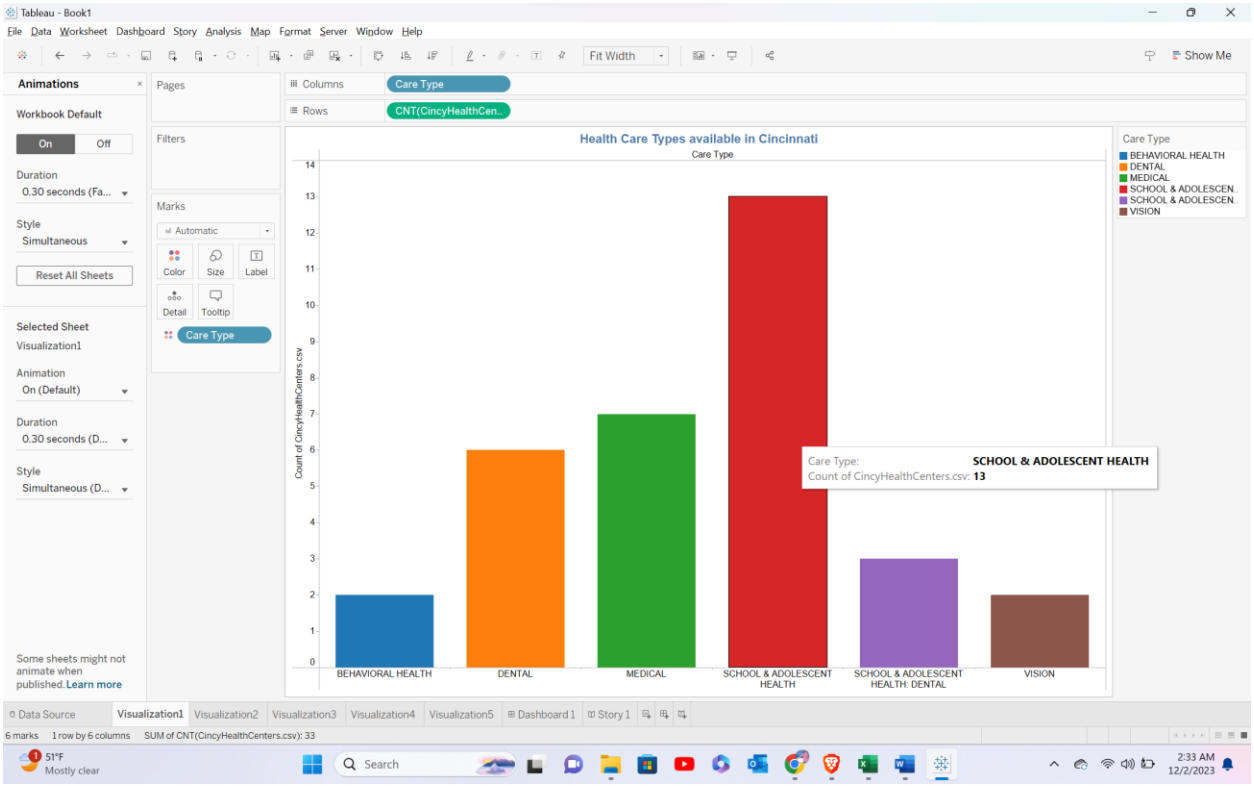
ROW	TYPE	TABLEAU VISUALIZATIONS	SOLUTION
1.	Bar	Health Care Types available in Cincinnati	In Tableau, I crafted a bar chart that offers a clear view of the Cincinnati Health Centers dataset. The goal was to address question one by tallying the occurrences of various healthcare types in Cincinnati. On the chart, the X-axis illustrates different care types with a simple appearance and color codes. Meanwhile, the Y-axis shows the count or frequency of these health cares. The observations reveal six distinct types of health care. Among these, 'School & Adolescent Health' stood out as the most frequent, while 'Behavioral and Vision Health' appeared least in number. And these results are same as R results.
2.	Map	Map for Health Centers in Cincinnati	In response to research question 6, I employed Tableau to generate a map visualization that addresses the inquiry effectively. Using this visualization, I plotted circle dots of distinct colors on the map, with each color symbolizing a different type of health care available in the area. I ensured that the labels associated with these circle dots contained essential information such as the care type, health center name, corresponding zip code, admission fee, and average medical cost. This comprehensive data representation aids in identifying hospitals or health centers that are both nearby and cost-effective for patients. The interactive nature of this map visualization is highly beneficial. Users can easily navigate through the map, utilizing features like zooming in or out to focus on specific areas of interest. Furthermore, the ability to pinpoint precise locations on the map enhances its usability, allowing individuals to locate and assess healthcare options based on proximity and affordability with convenience and precision.
3.	Bubble	Bubble Visualization for availability of Nursing services	To address research question 2, I utilized a bubble visualization technique to depict different types of health care services across various neighborhoods. In this visualization, bubbles are represented in two distinct colors: orange and blue. The orange-colored bubbles denote healthcare facilities that offer nursing services, while the blue-colored bubbles represent those without nursing services. Moreover, I included essential details such as the care type, health center name, neighborhood, nursing service availability, and average medical cost in the labels associated with each bubble. This enriched information helps patients seeking specific nursing service details of hospitals within their neighborhood. Upon observing this visualization, it becomes evident that the blue-colored regions signify areas where nursing services are currently unavailable within healthcare facilities. Enhancing nursing services in these blue regions is crucial as it would significantly increase access and availability of such services to patients. Improving nursing services in these areas could bridge the gap, ensuring better healthcare accessibility and options for patients within those neighborhoods.
4.	Table	Tabular Data for Health Care in Cincinnati	To address research question 6, this table data serves as a valuable resource for identifying budget-friendly hospitals. The structured format of this table includes several columns such as care type, behavior, health center name, admission fee, average lab fee, average medical fee, and admission fee. Analyzing this table data provides insights into hospitals or healthcare centers that offer services at affordable rates. By examining the admission fees, average lab fees, and average medical fees across different care types and health centers, individuals can make informed decisions to identify facilities that align with their budget constraints. This structured data serves as a practical reference point for individuals seeking cost-effective healthcare options. It assists in comparing various hospitals based on their fees and services offered, thereby aiding in the identification of budget-friendly healthcare options that cater to specific needs and financial considerations.
5.	Side Bar	Health Care vs Average Cost of Expenditures	To address research question 7, I utilized Tableau to create a side-by-side bar chart that compares various types of healthcare treatments based on their associated average admission fees, lab fees, and medical costs. In this visualization, the different types of health care treatments are grouped together, and each group is represented by three distinct color bars. The blue bars indicate the average admission fees, the purple bars represent the average lab fees, and the green bars signify the average medical costs

NAME: ESWAR RAO RAMINENI
MID: M15270716

for each corresponding healthcare sector. Moreover, the visualization includes a label pop-up feature, enabling users to view specific values by hovering the mouse cursor over the bars. This interactive functionality enhances the chart's usability by providing detailed information about the average costs associated with different healthcare treatments. By examining this side-by-side comparison, individuals can easily identify which healthcare treatment tends to have higher average costs. This visual representation aids in determining the most expensive treatment by considering and comparing the average admission fees, lab fees, and medical costs across various healthcare sectors.

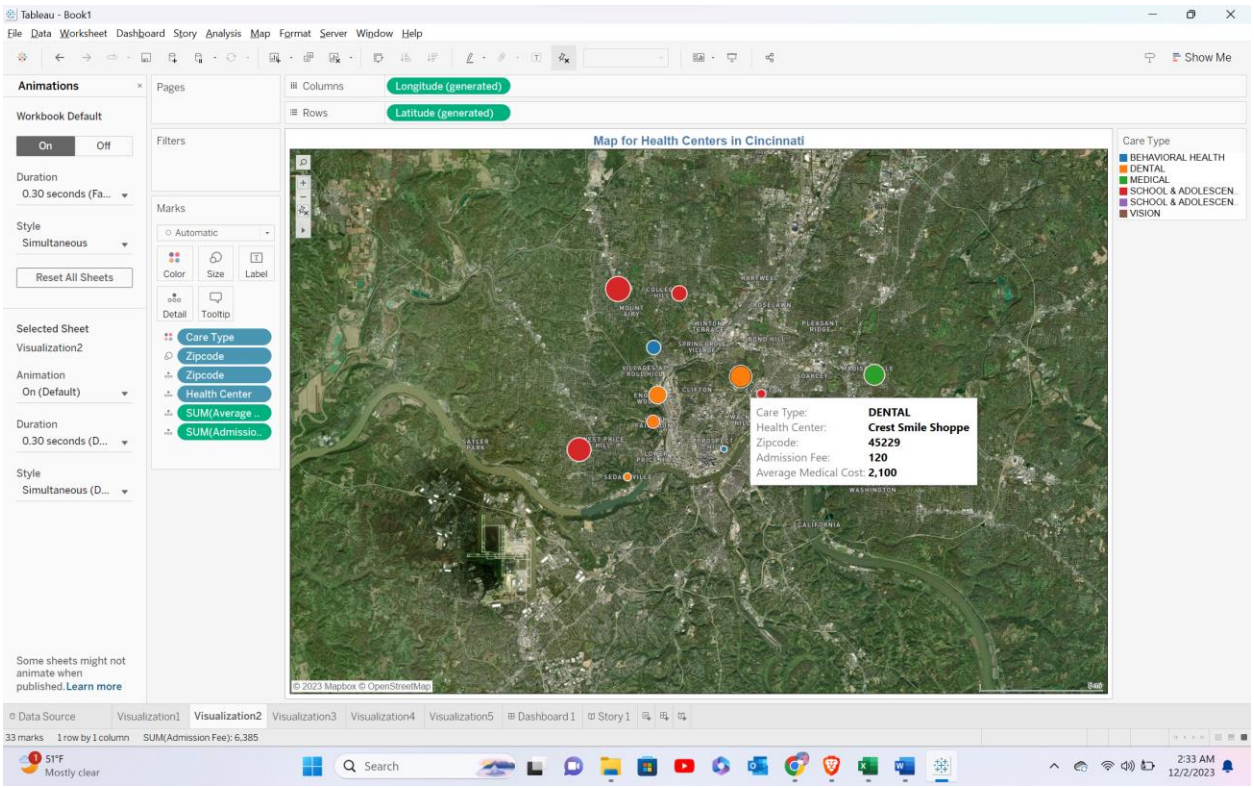
TABLEAU VISUALIZATIONS:

Visualization 1:

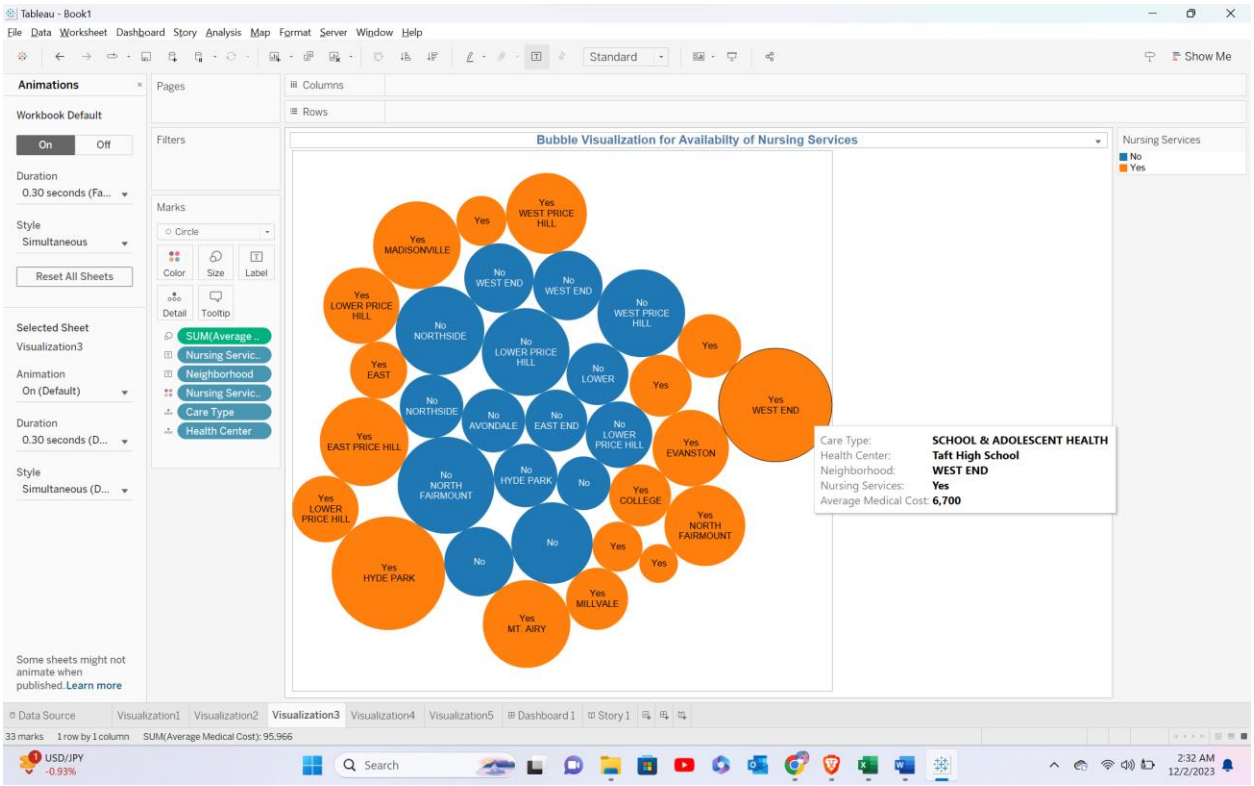


HI PROJECT ASSIGNMENT II

NAME: ESWAR RAO RAMINENI
MID: M15270716
Visualization 2:

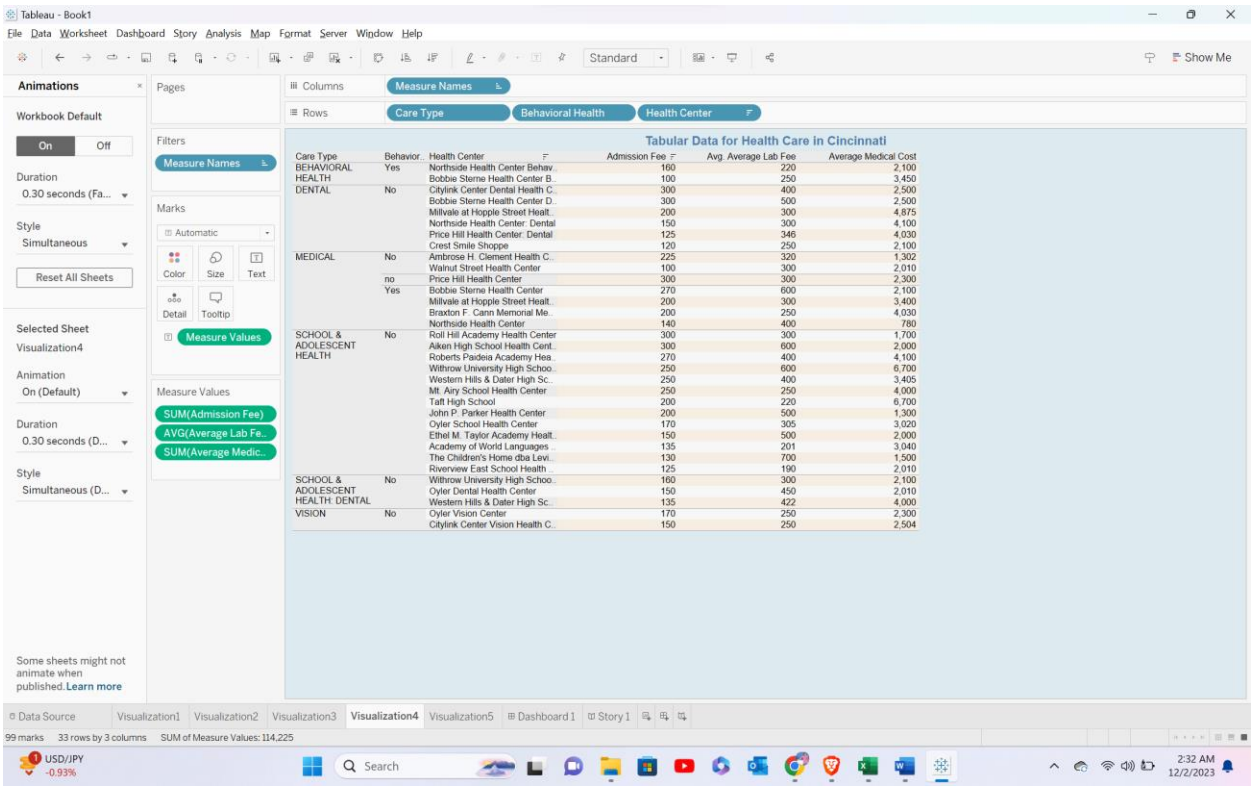


Visualization 3:

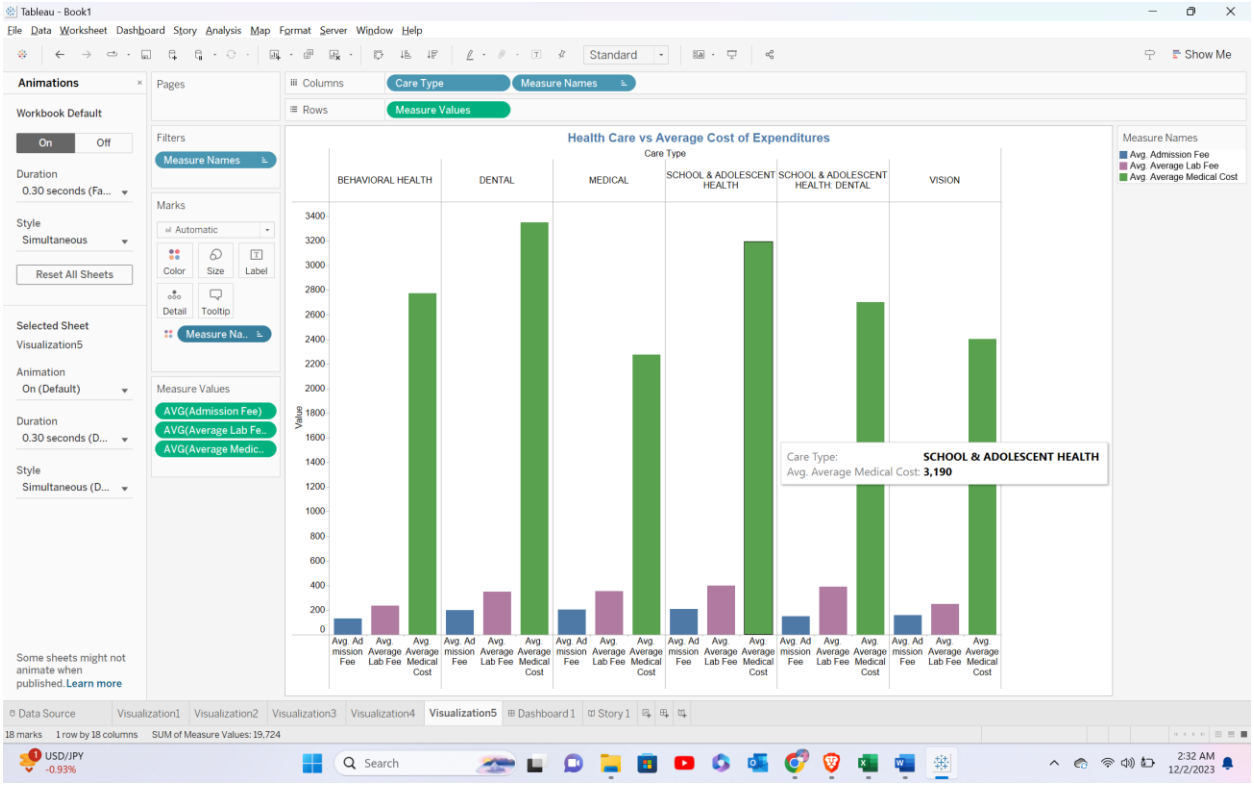


HI PROJECT ASSIGNMENT II

NAME: ESWAR RAO RAMINENI
MID: M15270716
Visualization 4:



Visualization 5:



HI PROJECT ASSIGNMENT II

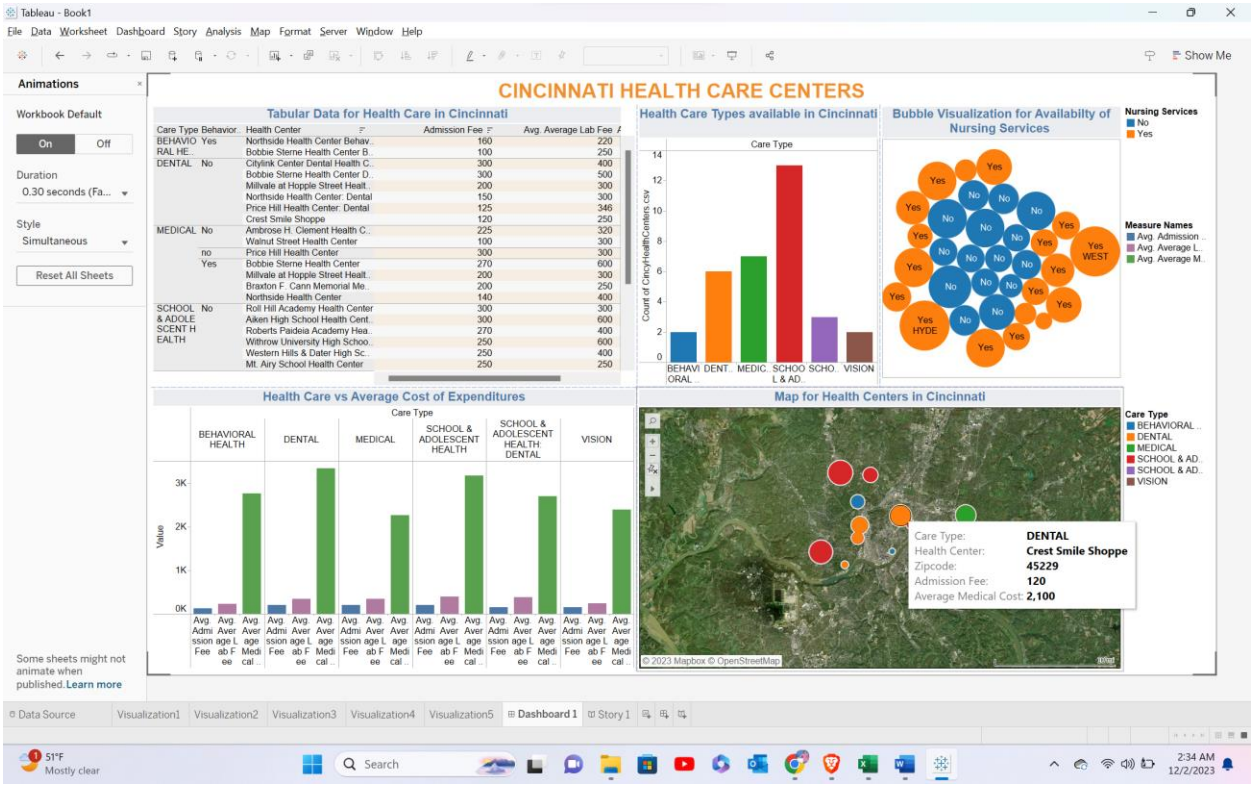
NAME: ESWAR RAO RAMINENI
MID: M15270716

DASHBOARD CREATION:

Tableau offers a cool thing called a dashboard. It's like a place where you can put all your pictures and charts together. For my data about "CINCINNATI HEALTH CARE CENTERS," I made this special dashboard. It has five different types of pictures and charts.

First, there's a table showing information about health care in Cincinnati. Then, there's another picture showing the different types of health care available in Cincinnati. Next, there's a bubble picture that shows where nursing services are available. Also, there's a chart comparing different health care and how much they usually cost.

Finally, there's a map that shows where all the health centers are in Cincinnati. I made sure the dashboard looks cool by using a modern theme. This makes it easier for patients or anyone using it to click around and see the information they need. The dashboard is like a neat collection of pictures that help people understand health care in Cincinnati better.



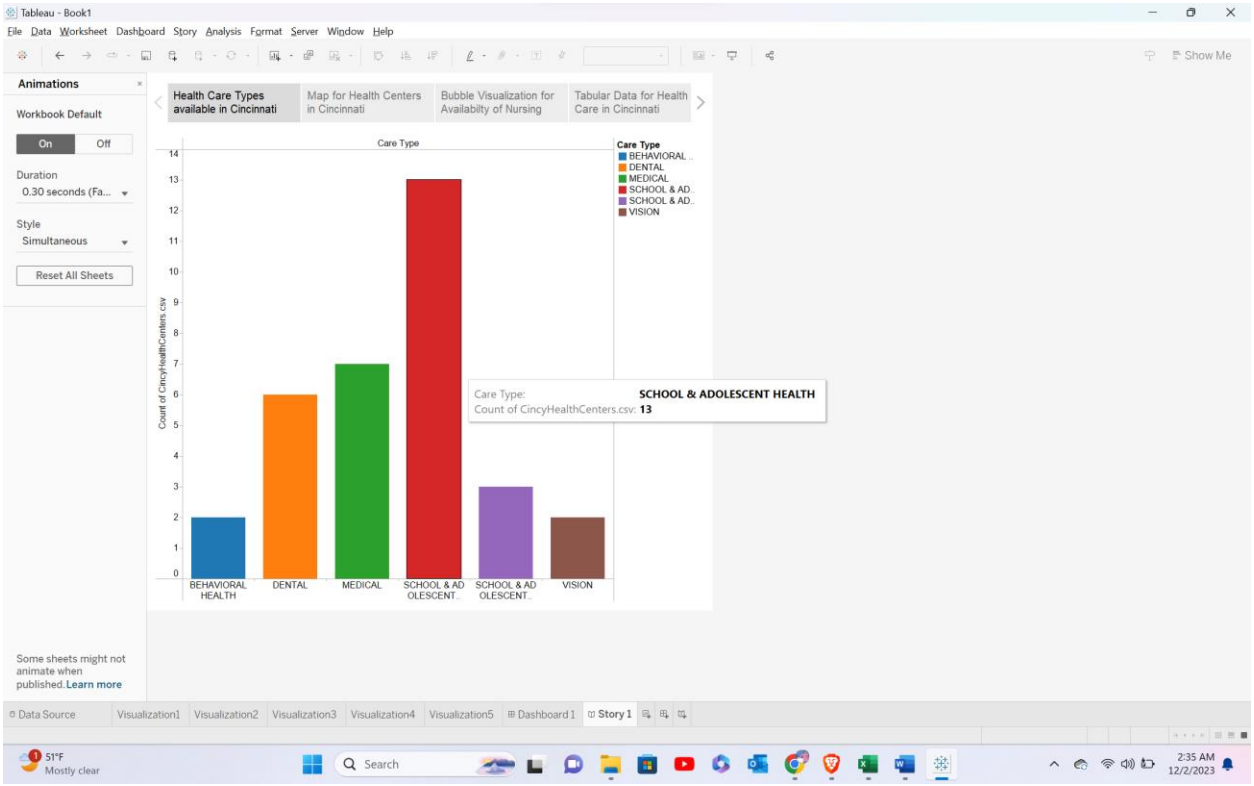
HI PROJECT ASSIGNMENT II

NAME: ESWAR RAO RAMINENI
MID: M15270716

USER STORY:

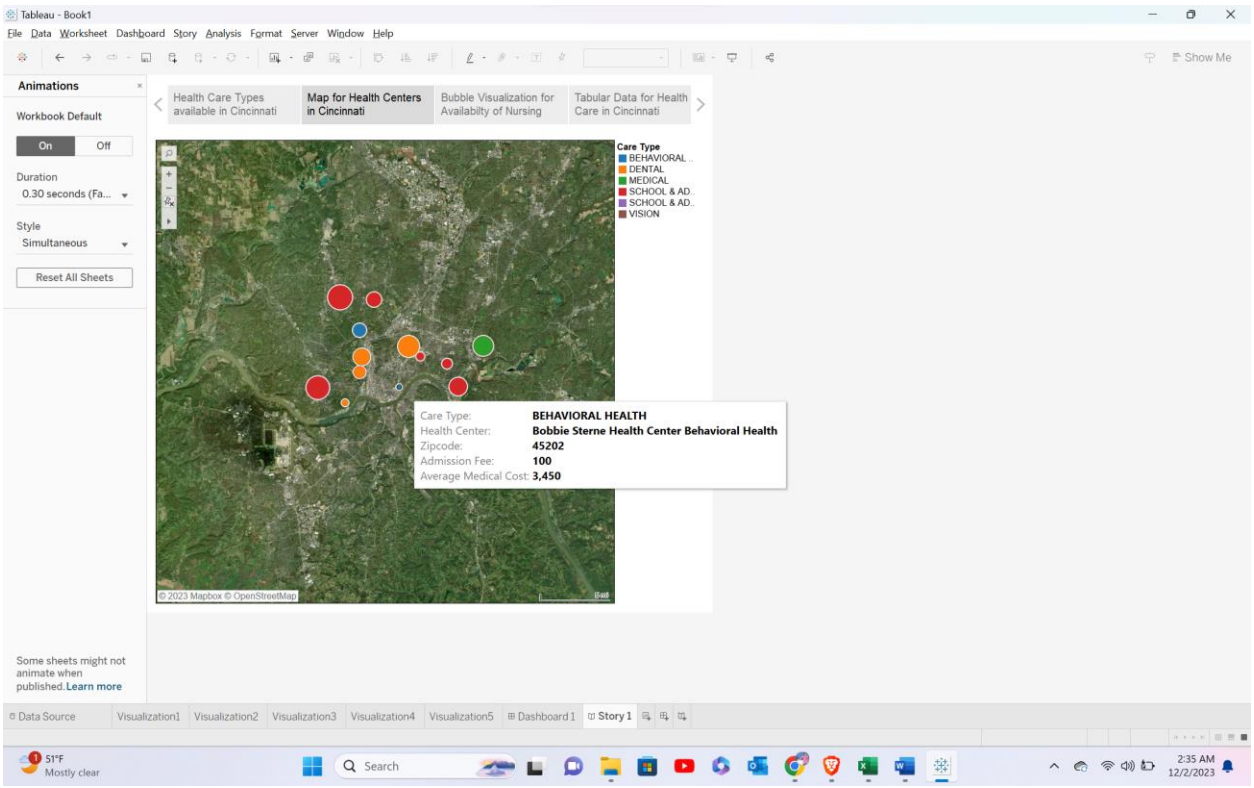
In tableau we have an option to create a story for visualizations. This story contains series of status followed by one-by-one visualization, this stories greatly helps to understand and presentation to users. It also has good options such as animations adding text or description to stories.

Status 1: Health Care Types available in Cincinnati

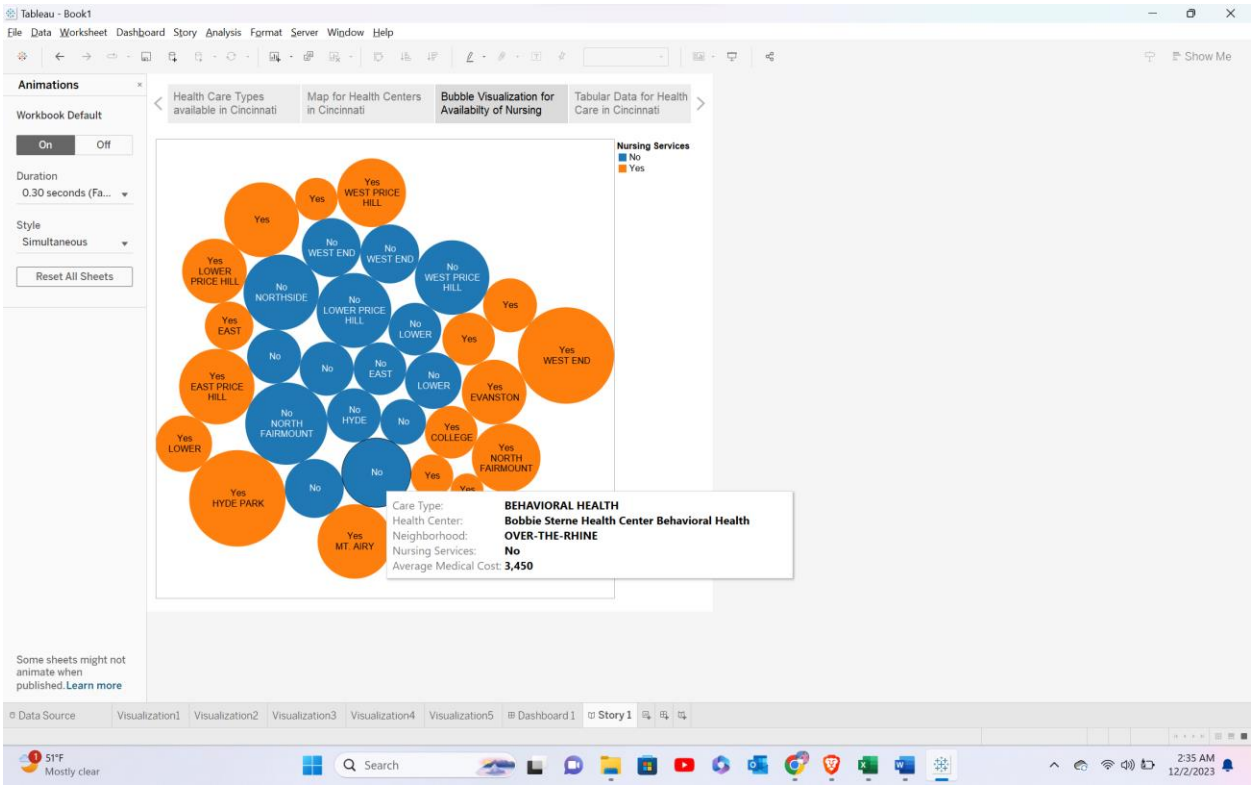


HI PROJECT ASSIGNMENT II

NAME: ESWAR RAO RAMINENI
MID: M15270716
Status 2: Map for Health Centers in Cincinnati

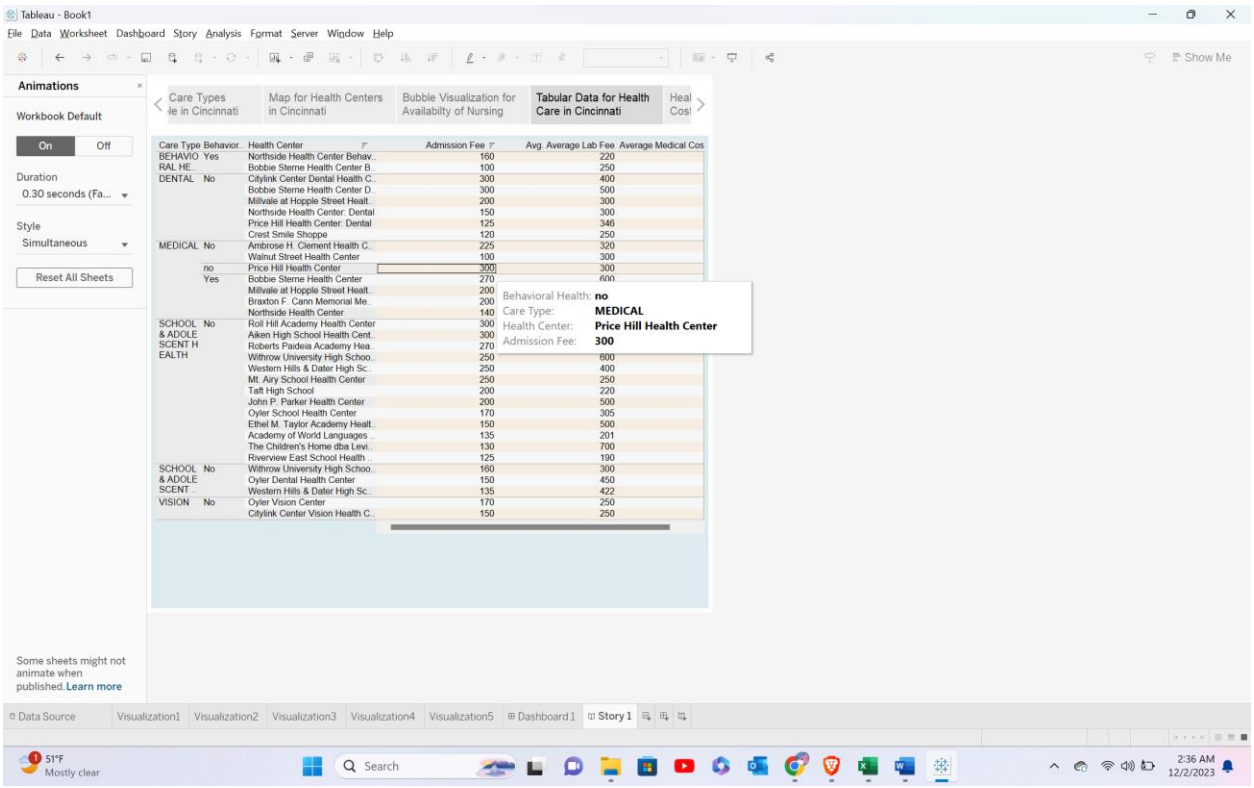


Status 3: Bubble Visualizations for Availability of Nursing services



HI PROJECT ASSIGNMENT II

NAME: ESWAR RAO RAMINENI
MID: M15270716
Status 4: Tabular Data for Health Care in Cincinnati.



Status 5: Health Care vs Average Cost of Expenditures.