Team 8 Process Book

Team Name: Team 8
Project Name: Influenza between and within States
Team Members: Jenna Pollack, Garrett Mizell, Etsubdink Yergashewa, Faisal Rasheed Khan, Kate Martinez

---

Proposal + Charter (Phase 1)

Charter:

Technical Roles:

Code Monkey: Faisal Rasheed Khan
Technical Writing Manager: Garrett Mizell QM
Project Manager: Jenna Pollack QM
Resource Manager: Etsubdink Yergashewa

Scheduling:

Meeting evenings (6/7pm) on the weekdays through discord.

Conflict resolution strategies:

Contact the Professor. Jenna and Garrett will decide if it's worth contacting the professor since they both chose judge.

Collaboration tool preferences:

Jupyter Notebook for coding, will use discord for brainstorming and project management by utilizing Jira to monitor deadlines, upcoming tasks, and completed work.

Communication preferences and plans:

Over discord.

Proposal:

Domain and Project Motivation:

Big Idea: Analyzing Flu by State/City

The domain focuses on the flu outbreak over the different regions within the US. It aims to analyze and visualize the spread and impact of the flu across the different states of the US.

The project is driven by the intention of aiding users in addressing the spread of the flu by analyzing data and visually presenting patterns. By comprehending these patterns through visualizations, it aims to promote awareness and an understanding of where the flu is most prevalent.

Users:

People who are unsure about being vaccinated, people at risk for influenza, hospitals, healthcare workers

Main Tasks/Analysis Questions:

- The user will be able to locate the state/area that is most likely to have an outbreak of influenza based on lowest vaccine rate and highest influenza rate. {locate, state/area}
- The user will compare cases in areas with high vaccination rates to cases in areas with low vaccination rates. {compare, flu case counts}
- The user will notice if there is a correlation between influenza case rate and population density. {compare, correlation}
- The user will be able to find the seasonal patterns and peak periods of influenza within different regions or states. {identify, seasonal patterns and peak period}

Data:

First:

Data we will be using from data.world:
https://data.world/chhs/fc544658-35c5-4be0-af20-fc703bc57c13
The Original Source: https://www.cdph.ca.gov/Programs/CID/DCDC/Pages/Immunization/Influenza.aspx (at the bottom of the page)
Datatype: The dataset type is a table which has the data types are attributes and items
Semantic: year, data code (unique code for each week to assist with grouping by week), week of the data, region where samples are collected from, type of respiratory virus, amount of positive patients, amount of processed patients, and percentage of positive patients
Collected and processed: The Immunization Branch at the California Department of Public Health (CDPH) collects, compiles and analyzes information on influenza activity year-round in California and produces a weekly influenza surveillance report during October through May.
Limitations: This research contains both a region (Bay Area) and just the state without a region (CA), so we may need to import extra data with regions to map data correctly.
Transform: We expect a transformation of this data to be calculating the season based on which week of data is being used. (Week 8 in this list is the start of october. These seasons are listed by week and range from may to october).

Second:

The data using from data.world:
https://data.world/cityofchicago/8vvr-jv2g
The original source: https://dph.illinois.gov/data-statistics/syndromic-surveillance
Datatype: The dataset type is a table which has the data types are attributes and items
Semantic: peak flu season weeks (week 40 to week 20), start of the week of influenza being measured, end of week of influenza being measured for that week, zip code of influenza being measured, risk level of influenzal, record id of the visit, zip code location by coordinates.
Collected and processed: IDPH (Illinois Department of Public Health) collects data from the emergency department of 185 hospitals in Illinois.
Limitations: This Data set does have latitudes and longitudes to plot, but we will need the other state's longitude and latitude to make a comparison map graph.
Transform: We plan to combine the 'zip codes' and 'number of cases' columns to create a new column of 'cases per zip code'. The' longitude' and 'latitude' columns will be later aggregated to assist in plotting the 'cases per zip code' column.

Third:

The data using from data.world:
https://data.world/city-of-ny/2nwg-uqyg
The original source:

https://data.cityofnewyork.us/Health/Emergency-Department-Visits-and-Admissions-for-Inf/2nwg-uqyg/about_data
Datatype: The dataset type is a table which has the data types are attributes and items
Semantic: date the data was extracted, date of the emergency visit, zip code of collection, total ED visits, amount of influenza visits, amount of influenza admissions
Collected and processed: The data is collected by the Department of Health and Mental Hygiene (DOHMH) from Hospitals in New York City.
Limitations: Only has zip codes, will probably need to import country wide data with zip codes to map
Transform: We will need to combine the zip codes with the additional data such as county zip codes to make proper longitudes and latitudes.

Fourth:

The original source:
https://zipatlas.com/us/il/chicago/zip-code-comparison/highest-population-density.htm

Datatype: The dataset type is a table which has the data types are attributes and items
Semantic: Zip Code, People per Sq Mile, VS. State (not needed), VS. national (not needed)
Collected and processed: Don't know (likely the Gov census)
Limitations: If there are any missing zip codes (haven't verified yet)
Transform: We are going to use the Zip code and People per Sq. mile columns with the other Chicago dataset to plot the population density along with influenza cases.

Fifth:

The original source:
http://www.usa.com/rank/new-york-ny--population-density--zip-code-rank.htm?yr=9000&dis=&wist=&plow=&phigh=

Datatype: The dataset type is a table which has the data types are attributes and items
Semantic: Rank, Population density, Zip Code / total population (just need the zip code)
Collected and processed: Don't know (likely the Gov census)
Limitations: If there are any missing zip codes (haven't verified yet)
Transform: We are going to use the Zip code and People per Sq. mile columns with the other NYC dataset to plot the population density along with influenza cases.

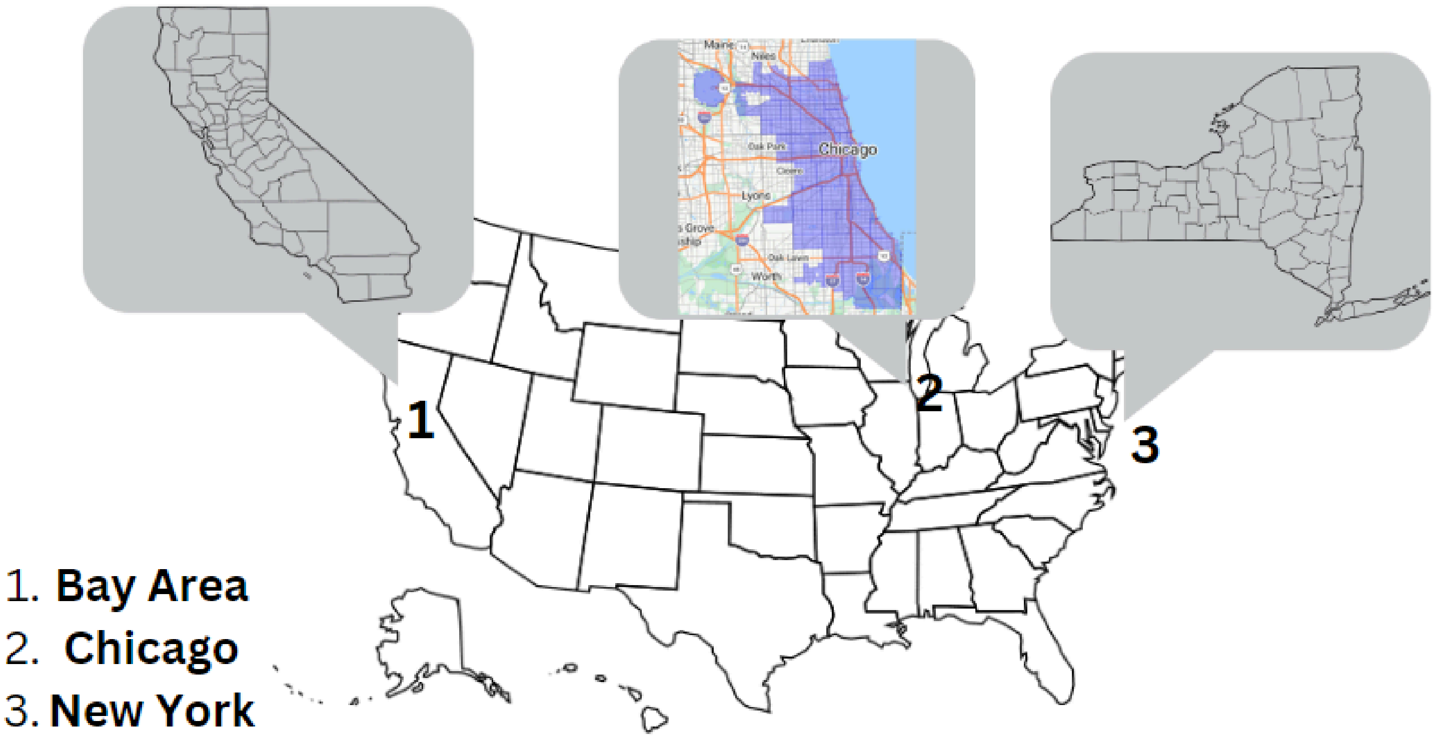
Project Contributions:

We will use Plotly and R. We would like to use heatmaps, geographic maps, and bubble maps to show regions by zip codes and coordinates of influenza. One dataset (CA) only lists general regions, so we will have to find a dataset with those same general regions. We will use some form of a bar chart to clearly see the influenza rate differences between each state without needing to look at a country map. Potential issues could be combining data with additional map data. Overlaying data onto a map could present a challenge since we have not done it before.

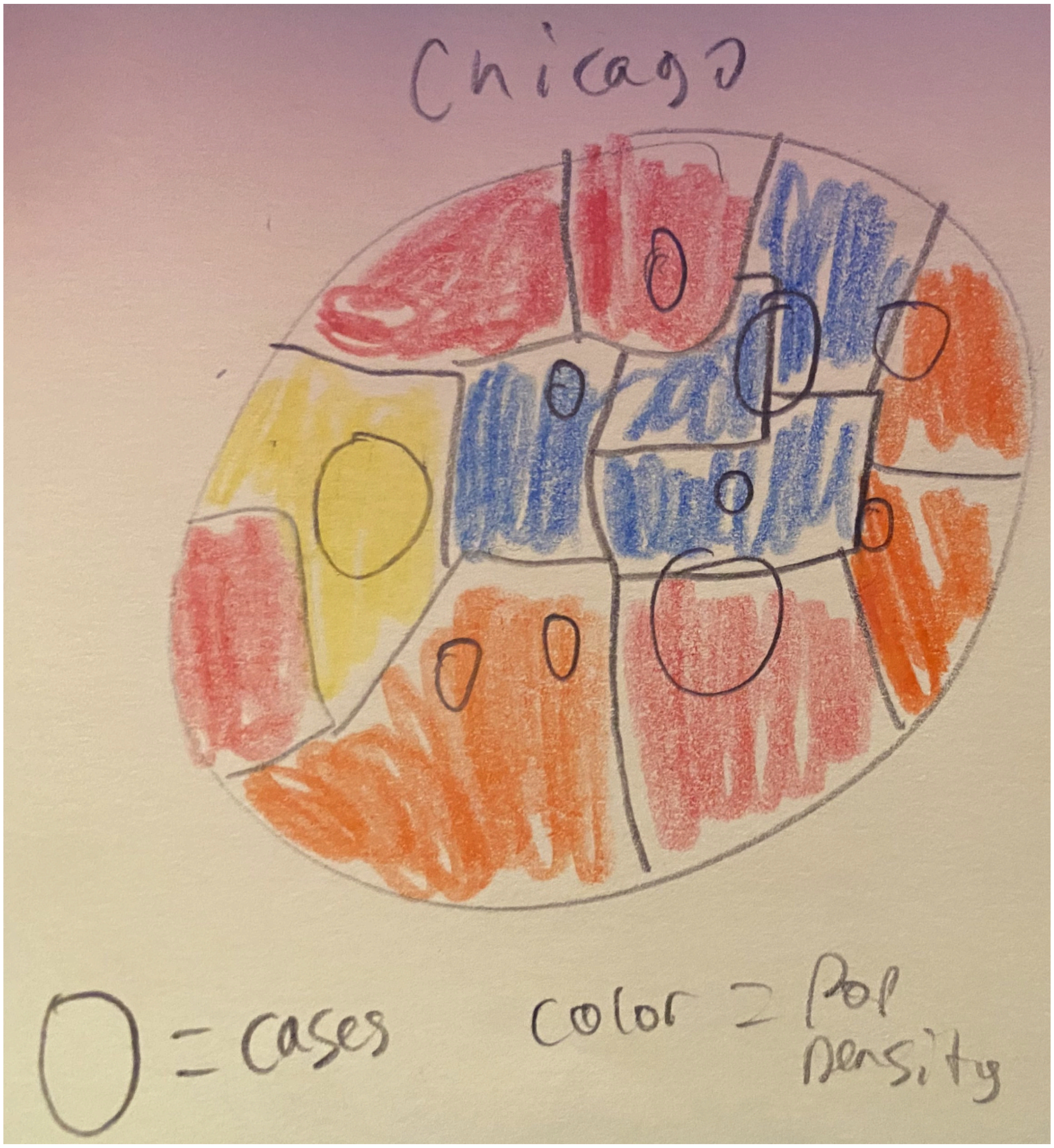
We plan to use interactivity to allow the user to hover over a zip code section to display the zip code and number of influenza cases. This will provide a numerical value to the sequential heatmap to give the user a clear understanding of the flu in their city. We do have some interactivity planned with our graphs, since we believe that being able to hover over a zip code that displays the specific case count would help us understand the data. We will present the user with graphs that contain a brief caption that will explain the purpose of the visualization. Our end product will have at least a graph for each city's state in our data (Chicago, New York, San Francisco) graph for each state and a comparison of all three states to give the user an overall picture of influenza cases (see Vis #1).

End Products:

Vis #1: A map of the U.S with a bubble map for each of the 3 cities based on the number of cases in the zip code (may have to group neighboring zip codes, will address this on creation.) (All datasets will be used for this vis)



Vis #2:A map of one or more of the cities with influenza. This map will be overlaid with a given state or city to show any correlation between the population density (heat map) using luminance (potentially different hues depending on the data we use) and number of influenza cases per zip code (bubble/dot/some kind of mark). (It will not just be a circle). (A singular dataset will be used for this vis, whichever city/state we decide to focus on).

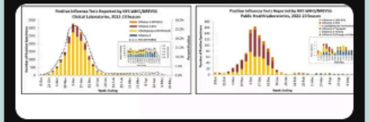


Reflection:

Looking at our padlet, we decided that we like the use of bar and line graphs to show comparison among the 3 states, however that is usually trivial. We decided to make overlapping line charts of each state and then combine them so that a user can compare all three charts together. We took inspiration from many heat graphs, so we agreed they would be useful in clearly displaying our influenza case density. We believe this would add clarity to showing the range of influenza cases. A potential visualization is with using a stacked bar chart, but the details still need to be completed. We thought using an overlay of the geographical shape of the states for which our 3 cities reside would give an accurate depiction and understanding of where influenza cases are located. We may allow the user to hover a zip code area for the font to enlarge for them. The idioms we are mainly thinking about are heat graphs, stacked bar charts, and juxtaposing multiple line graphs. We chose Influenza cases as our topic of study as, with the recent pandemic, understanding how viruses spread, specifically in high density cities, is vital for the population.

Screenshots of Padlet:

Etsub Yergashewa 7h



This graph illustrates the peak period of influenza activity. It visually presents the time frame when influenza reaches its highest levels, providing valuable insights into the seasonal trends of the virus.

Marks
line

Channels
hue

(3) 3

jennap3 7h
I like this idea, I just have a hard time understanding stacked bar charts. I bet we could do something more interesting.

Garrett Mizell 7h
This graph works well with our datasets that give the weeks. The same style and format can be used to represent the cases over time.

Faisal Rasheed Khan 7h
The visualization is good plot for the data, but it seems to increase

https%3A%2F%2Fwww.latimes.com%2Fcalifornia%2Fstory%...

Etsub Yergashewa 7h

Source:
https://www.latimes.com/california/story/2022-12-03/california-reporting-very-high-flu-activity

I like this map it demonstrates how California and New York were encountering notably high rates of influenza compared to other states.

Marks
area

Channels
color saturation, area, color Hue

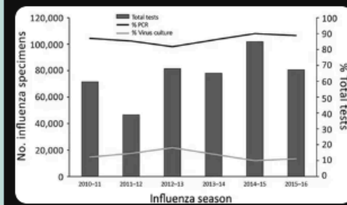
(3) 3

Garrett Mizell 7h
The style of heat map is perfect for what we want, just need to focus this in for the designated cities.

jennap3 7h
I really like how this graph demonstrates color use for showing

Faisal Rasheed Khan 9h

Influenza spread based on the specimens



Caption: The data shows the influenza specimens collected and tested. Based on the tests the influenza pattern is shown over the years. The bar plot represents the total tests.

This kind of plot helps the task we are doing by considering the states and based on the specimens tested we can see the outbreak over the period, and we can have a comparison plot in the same graph to visualize better.

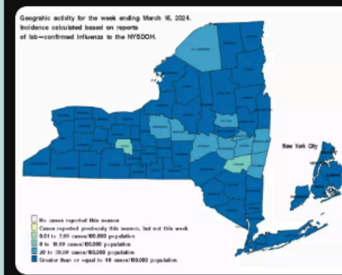
Source:
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6038762/

Marks
Lines

Channels
Identity Channels: None

Faisal Rasheed Khan 10h

New York Influenza Outbreak within the state



Caption: The New York region represents the weekly flu outbreak. The color saturation gives the information regarding the severeness of the outspread, where dark color represents the most affected area within the New York.

The plot helps in a good visualization to understand the outbreak of the influenza within the New York weekly state. The visualization answers the analysis question of the patterns/periods of the influenza as the data is represented weekly

Source:
https://www.health.ny.gov/diseases/communicable/influenza/surveilla

This graph works well with our datasets that give the weeks. The same style and format can be used to represent the cases over time.

Faisal Rasheed Khan 7h
The visualization is good plot for the data, but it seems to increase cognitive load for the user. Need to make it user friendly to visualize.

Add comment

Garrett Mizell 7h
The style of heat map is perfect for what we want, just need to focus this in for the designated cities.

jennap3 7h
I really like how this graph demonstrates color use for showing the worse and better parts of influenza.

Faisal Rasheed Khan 6h
The idea looks good, but need to figure out how we represent, may be juxtapose view because of the fact that we only require California state, New York state and Chicago city.

Add comment

articles/PMC6038762/

Marks
Lines

Channels
Identity Channels: None

Magnitude Channels: Horizontal spatial position, Vertical Spatial position

(3) 3

jennap3 7h
I think we definitely all agree a bar chart/line graph/combo is definitely an easy to understand path.

Garrett Mizell 6h
I think this is a good idea for the overall cases or we would have 3 separate lines in the line graphs for each city based on population density

Etsub Yergashewa 24m
I like this as Jenna said it's really easy to understand but adding color saturation would make it better

Add comment

question of the patterns/periods of the influenza as the data is represented weekly

Source:
https://www.health.ny.gov/diseases/communicable/influenza/surveillance/2023-2024/flu_report_current_week.pdf

Marks
Areas

Channels
Identity Channels: Color Hue
Magnitude Channels: Areas, Color Saturation

(3) 3

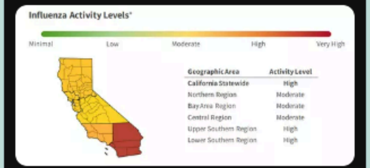
jennap3 7h
I really like this idea to show NY! the color saturation and areas make it a really nice visual.

Garrett Mizell 6h
The color saturation is amazing, this same idea but zoomed in to just NYC will work nicely

Etsub Yergashewa 2h
I like this map i also agree with Jenna and Garrett how the color saturation is important and make it easy to understand

Add comment

jennap3 22h



I like this graph because our CA data includes regions such as "Bay Area" and I like the idea of a heat graph.
https://www.cdph.ca.gov/Programs/CID/DCDC/CDPH%20Document%20Library/Immunization/Week2022-2347_FINALReport.pdf

Marks
Areas

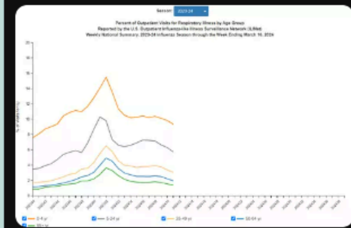
Channels
color hue, area

(3) Rate 3

Garrett Mizell 8h
This is perfect for the Bay area part of our dataset, I was worried we wouldn't be able to figure out what exactly and bay area was.

Faisal Rasheed Khan 7h
This is what we need as a visualization for the California map, but the coloring can be shown effectively with the color sequential and also with our parameters. Overall, this kind of visualization we

jennap3 22h



I like this map because it can easily show data comparison with CA, NY, and Chicago. It also uses weeks as data, and we have data in the form of weeks as well. Hoping for it to be good inspiration.
https://www.cdc.gov/flu/weekly/index.htm#ClinicalLaboratories

Marks
lines

Channels
tilt/angle, hue,

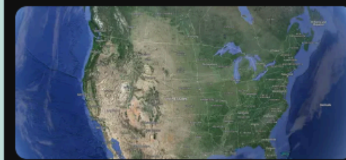
(3) Rate 3

Faisal Rasheed Khan 7h
The comparison would be good, but this visualization is not we are focused off. Because of the x-axis age as we dont compare the age

Garrett Mizell 6h
I like the style but i agree with Faisal as we can change age to another value if needed.

Garrett Mizell 1d

Map of the U.S.A with all 3 (Chicago, NYC, San Francisco) cities labeled.



This map goes perfect with visualization #1

https://www.google.com/maps/@41.7856216,-99.9757387,3419834m/data=!3m1!1e3!5m2!1e2!1e4?entry=ttu

We can add 3 small zoomed in maps of the 3 cities so all can be visible on one map. I took this straight from google maps, it shows a point for all 3 cities so they are easily identifiable. Using the juxtaposed views will give a better understanding of the cases in the different cities.

Marks
lines, points

Channels
Position, Texture, Orientation

Garrett Mizell 1d

Zip Code Map of Chicago

This map is perfect for plotting Vis #2, the population density and number of cases in Chicago
Source:
https://ofomaps.com/downloads/chicago-zip-code-map/

By overlaying a heat map onto the zip code regions we can display the density of influenza cases with population density. It should give a nice contrast.

Marks
lines, areas

Channels
Color, position, Potentially luminance

(4) 3

Faisal Rasheed Khan 7h
The zip codes for the visualization of Chicago are necessary as this

-2347_FINALReport.pdf

Marks
Areas

Channels
color hue, area

(3) Rate 3

Garrett Mizell 8h
This is perfect for the Bay area part of our dataset, I was worried we wouldn't be able to figure out what exactly and bay area was.

Faisal Rasheed Khan 7h
This is what we need as a visualization for the California map, but the coloring can be shown effectively with the color sequential and also with our parameters. Overall, this kind of visualization we
need with changes according to our setting.

Etsub Yergashewa 1h
i like this viz the color hue is really good as Garret said it's good for for bay area in our dataset

Add comment

Hoping for it to be good inspiration.
https://www.cdc.gov/flu/weekly/index.htm#ClinicalLaboratories

Marks
lines

Channels
tilt/angle, hue,

(3) Rate 3

Faisal Rasheed Khan 7h
The comparison would be good, but this visualization is not we are focused off. Because of the x-axis age as we dont compare the age

Garrett Mizell 6h
I like the style but i agree with Faisal as we can change age to another value if needed.

Etsub Yergashewa 1h
i also like the style it clear for and easy to understand

Add comment

1.7856216,-99.9757387,3419834m/data=!3m1!1e3!5m2!1e2!1e4?entry=ttu

We can add 3 small zoomed in maps of the 3 cities so all can be visible on one map. I took this straight from google maps, it shows a point for all 3 cities so they are easily identifiable. Using the juxtaposed views will give a better understanding of the cases in the different cities.

Marks
lines, points

Channels
Position, Texture, Orientation

(4) 2

Faisal Rasheed Khan 7h
The idea of the map is good and displaying using juxtapose is perfect. But the similar map presented above is confusing because of the cognitive load to look at these kind of maps.

jennap3 6h
I agree with Faisal, I bet we could find a less realistic map that still gets the point across.

Add comment

hicago-zip-code-map/

By overlaying a heat map onto the zip code regions we can display the density of influenza cases with population density. It should give a nice contrast.

Marks
lines, areas

Channels
Color, position, Potentially luminance

(4) 3

Faisal Rasheed Khan 7h
The zip codes for the visualization of Chicago are necessary as this
helps to plot the data within a state for the Chicago city region.

jennap3 6h
I like this idea a lot because it is easy to understand for the user. I think it will look a lot better when we use different color hues.

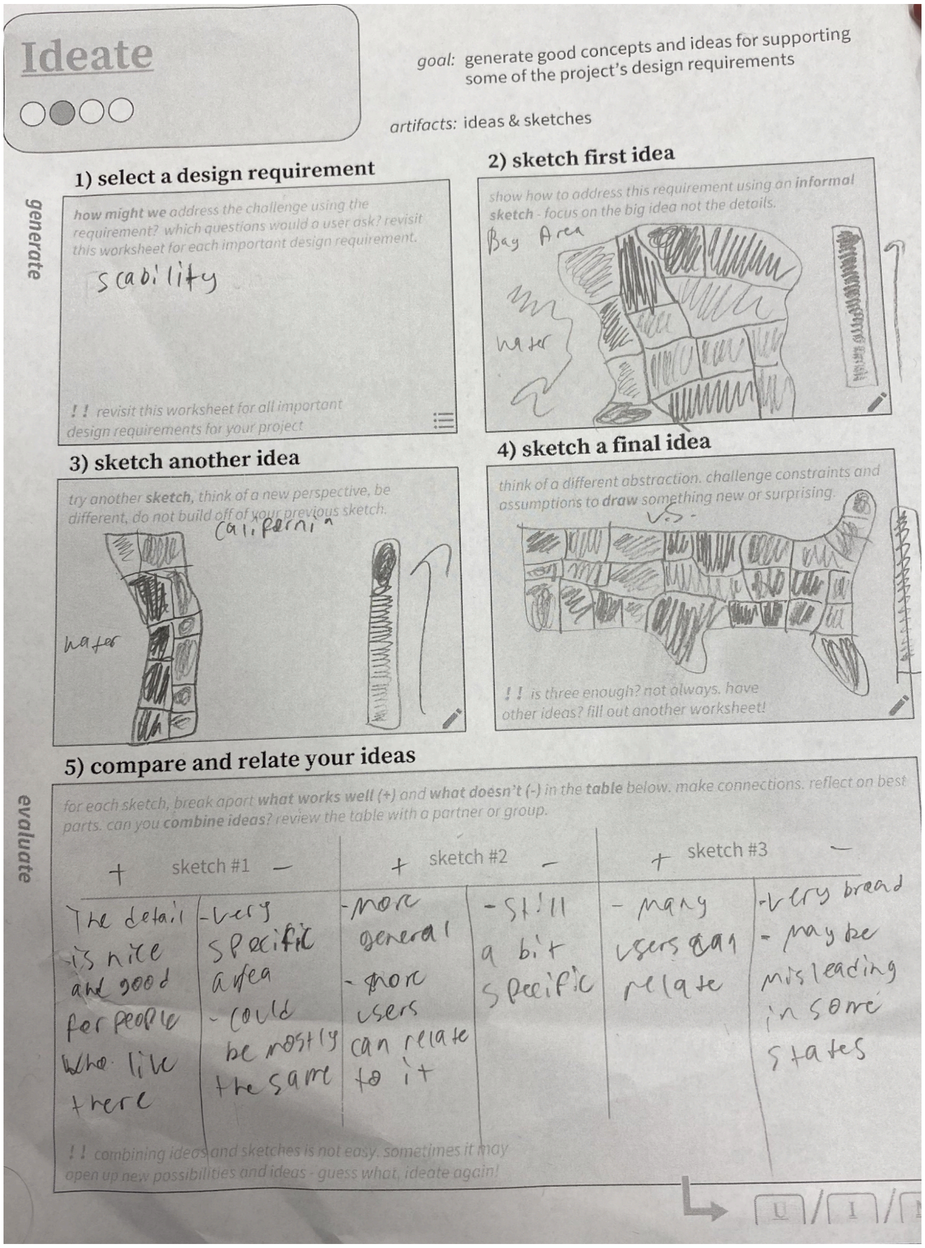
Etsub Yergashewa 1h
I agree with you jenna using different color hue will make this stand out

Add comment

Phase 3:

Task	Requirements
• The user will be able to locate the state/area that is most likely to have an outbreak of influenza based on the highest influenza rate.	• Will have a separate column to see influenza rate. • Zip code areas. • Colors to showcase the highest rates.
• Show users the rate of the state's flu for different years.	
• The user will notice if there is a correlation between influenza case rate and population density.	• Display a geographical map of one of the 3 cities sectioned off by zip code. • Use a heat map to plot the population density of each zip code. • Use a bubble map (or another mark) to represent the number of cases in that zip code.
• The user will be able to find the peak periods of influenza within different regions or states.	• Make a column of flu cases per peak flu month. • Plot that column per zip code.
• Show users the rates of different kinds of flu present. This allows users to see what kinds of flu are in which areas.	

Design Activity Sheets(Understand, Ideate)



Ideate

goal: generate good concepts and ideas for supporting some of the project's design requirements

artifacts: ideas & sketches

generate

1) select a design requirement

how might we address the challenge using the requirement? which questions would a user ask? revisit this worksheet for each important design requirement.

scability

! ! revisit this worksheet for all important design requirements for your project

2) sketch first idea

show how to address this requirement using an **informal sketch** - focus on the big idea not the details.

Bay Area

water

3) sketch another idea

try another **sketch**, think of a new perspective, be different, do not build off of your previous sketch.

California

water

4) sketch a final idea

think of a different abstraction, challenge constraints and assumptions to **draw** something new or surprising.

U.S.

! ! is three enough? not always. have other ideas? fill out another worksheet!

evaluate

5) compare and relate your ideas

for each sketch, break apart **what works well (+)** and **what doesn't (-)** in the **table** below. make connections. reflect on best parts. can you **combine ideas**? review the table with a partner or group.

sketch #1 +	sketch #1 −	sketch #2 +	sketch #2 −	sketch #3 +	sketch #3 −
The detail is nice and good for people who live there	-very specific area -could be mostly the same	-more general -more users can relate to it	-still a bit specific	-many users can relate	-very broad -may be misleading in some states

! ! combining ideas and sketches is not easy. sometimes it may open up new possibilities and ideas - guess what, ideate again!

U / I /

Ideate

goal: generate good concepts and ideas for supporting some of the project's design requirements

artifacts: ideas & sketches

generate

1) select a design requirement

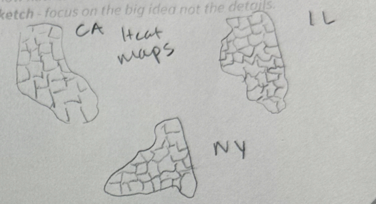
how might we address the challenge using the requirement? which questions would a user ask? revisit this worksheet for each important design requirement.

- easy to understand

!! revisit this worksheet for all important design requirements for your project

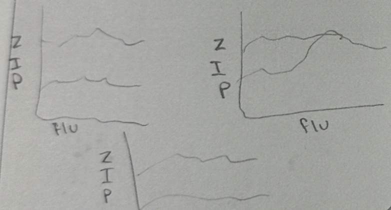
2) sketch first idea

show how to address this requirement using an **informal sketch** - focus on the big idea not the details.



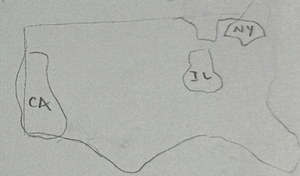
3) sketch another idea

try another **sketch**, think of a new perspective, be different, do not build off of your previous sketch.



4) sketch a final idea

think of a different abstraction, challenge constraints and assumptions to **draw** something new or surprising

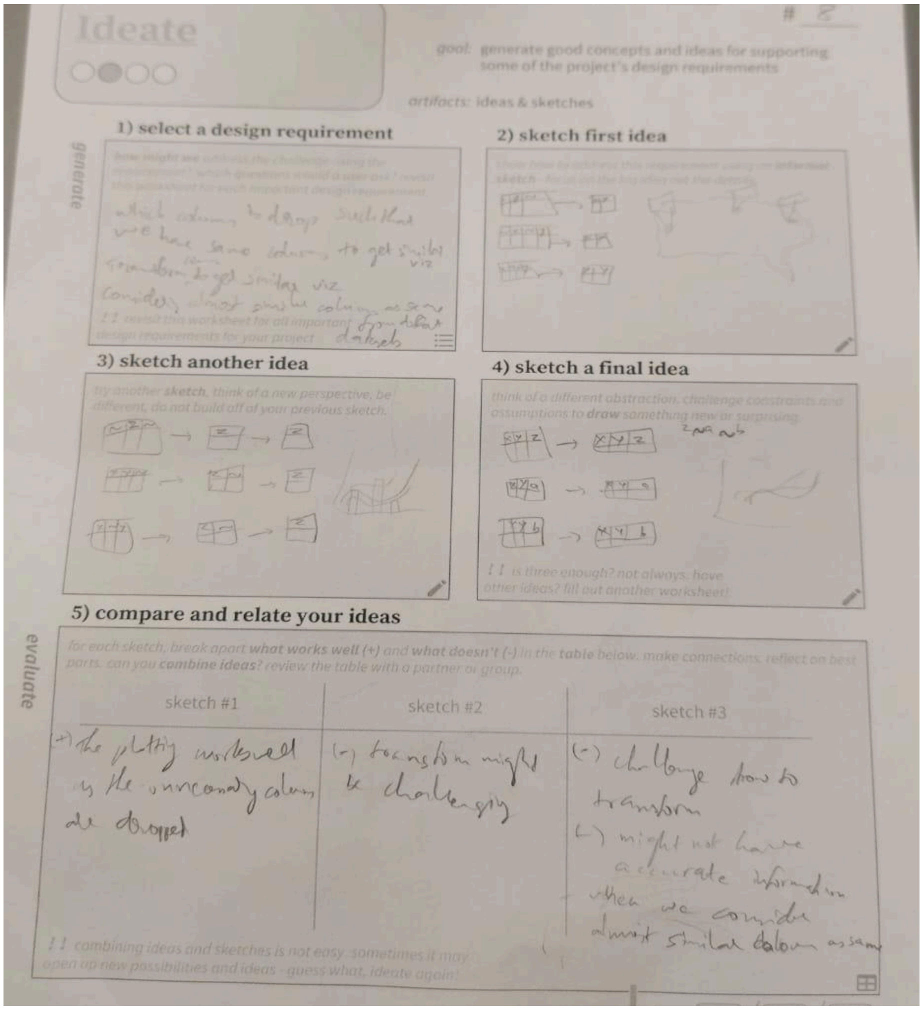


!! is three enough? not always, have other ideas? fill out another worksheet!

5) compare and relate your ideas

for each sketch, break apart **what works well (+)** and **what doesn't (-)** in the **table** below. make connections, reflect on best parts. can you **combine ideas?** review the table with a partner or group.

sketch #1	sketch #2	sketch #3
heat graphs help demonstrate comparison of flu cases by flu severity	having multiple line graphs to compare flu cases amoung states	might be too big to have a full world map



Color

The goal of using colormaps is to highlight where the flu is greatest. We plan on using diverging colormaps so the user will be able to see the slight increases of the flu. Thus, variance in influenza cases will affect the saturation which will play a great role in our graphs that measure ranges of the flu, but hues will be important in showing different categorical attributes such as seasons. Besides dataset columns, the other categorical data we have is the three states we are graphing together which are California, Chicago, and New York. The effect of changing the color binding to discrete is that it helps color the data making it easier to see trends or patterns instead of seeing the variance of different flu numbers. We will focus on showing one main color attribute. We looked online for several examples of different color schemes (see Fig 1.1) all of which were monochromatic, however that lacked visual appeal and creativity. An example is the gray dots which underlie the city add visual clutter and therefore will not be used in the final product. We decided on a more diverse color range for contrast and to allude to different influenza rates. Culturally, red is mostly used for danger, and with sickness, thus the worst of the flu will be shown as red, and the safer parts as green or blue (see fig 1.2). We are hoping to use color/heatmap for the task: “The user will be able to locate the state/area that is most likely to have an outbreak of influenza based on lowest vaccine rate and highest influenza rate.” Since discrete color does well with detecting patterns, we may use that color method to answer this task: “The user will be able to find the seasonal patterns and peak periods of influenza within different regions or states.” The seasons will be color coded as seen in figure 4.2. Being able to visualize the outbreaks on a map and compare the maps to each other is useful for understanding the areas of high rates of the flu. For fig 1.1, there are examples of color but also what the task may look like with different rates of the flu if you were to compare the 3 states. Fig 1.2 would show the heatmap for seasonal patterns.

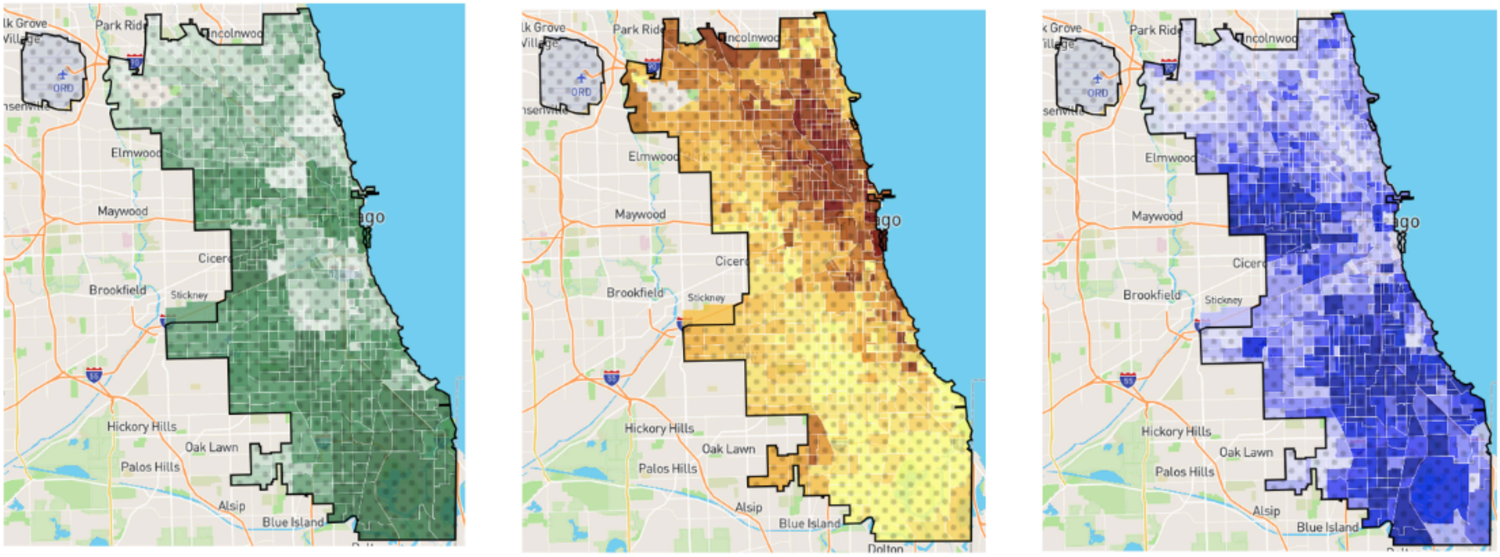


Fig 1.1 Shows examples found online of three different monochromatic color schemes that could be used. This is only meant to show different colors could be used and how hue may denote variance in influenza rates.

Fig 1.2 shows the final color schemes we decided to use. Mockup of heatgraph of chicago. Color encodes the amount of influenza with the color red denoting a high volume of influenza while blue denotes a low volume of cases.

Transformations

Transformation of the data is an important task for our visualization. As we are exploring different locations for our visualizations like Chicago, New York and California for Influenza spread, we needed to have similar data columns for comparison between the different locations. For example, the Chicago dataset has zip codes and latitude and longitude details to plot the geospatial visualization. The New York dataset on the other hand only has zip codes. To make it similar enough to the Chicago dataset for comparable plotting, we needed to transform our data based on the zip codes to get latitudinal and longitudinal information. A column called Zip Codes was added to the New York dataset utilizing a MapBox API which contains the relationship between zip codes longitude and latitude. A python function was made to do this for all values in the table. Similarly, for our California dataset, we need to find an appropriate way to transform its geospatial encoding (Bay Area, Central, Lower Southern, etc.). Additionally, we also need geojson codes to highlight the specific areas we are looking at in order to focus the user in on important locations such as Chicago, New York and California.

We are hoping to use transformations and color for the task: “The user will notice if there is a correlation between influenza case rate and population density,” Because this task has patterns and densities, using a heatmap will better convey our task to the user.This will be done by having a heat map of population density, and a dot on the map will encode the number of cases of influenza. Both population density and influenza rate will be encoded by a color scale. Fig 2.1 shows the first attempt at this which is a bit confusing with a lack of legend and too many circles encoding influenza rates. Fig 2.3 shows just a heatmap of the population density which is much more clear with a legend and less visual clutter. Influenza rate must still be added in the form of color coded dots to figure 2.3.

To get the population density, we needed to transform the 4th and 5th datasets but removed extra information. For example, we removed the rank and population in the 5 dataset (NYC population density) as it was not useful to our goal. We also removed the state VS. national column in the 4th dataset as it was irrelevant to our task.

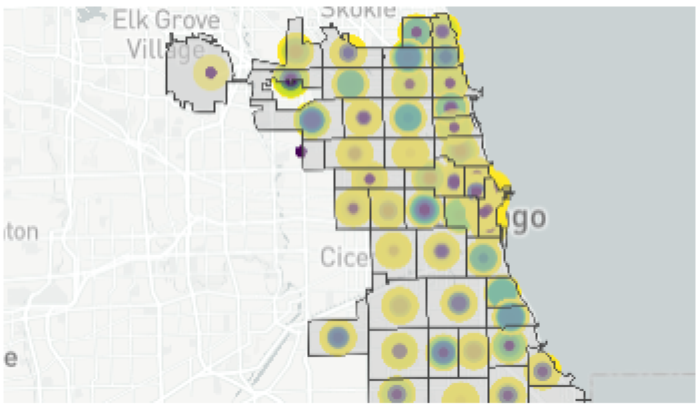


Fig 2.1 shows the first attempt at plotting of the Chicago dataset after the geojson zip codes transformations were completed.The size of the circle is not relevant, instead the color encodes the influenza rate. For example yellow denotes a high influenza rate while the dark purple represents a low influenza rate. The inclusion of circles with both color encodes change over time, however this proved difficult and caused visual clutter, so we decided to move in a different direction without encoding change over time.

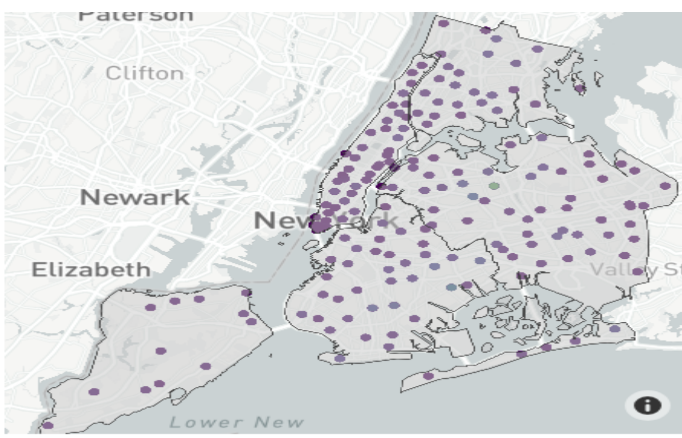


Fig 2.2 shows the plotting of the New York dataset after the latitude, longitude and geojson zip codes transformations were completed. This shows another way of encoding influenza rates as color coded dots, similarly to figure 2.1 with the small difference that they are solid colored dots. Once again, color encodes the number of influenza cases where yellow is high

and purple is low. One issue seen within this example is the values are too similar so finding the difference in rates is not clear.

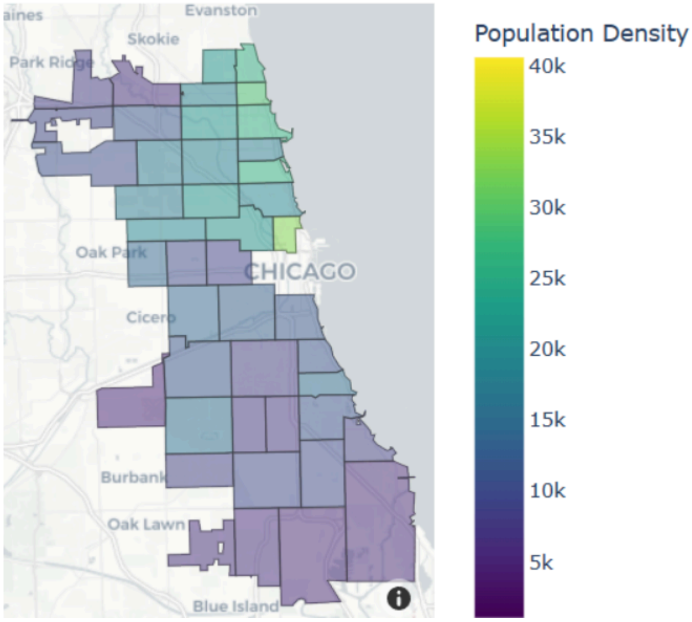


Fig 2.3 shows the heat map of population density for the city of Chicago by itself.

The chicago dataset used the mean ILI activity level but this is not achievable for the New york dataset. We instead will use the number of influenza-like visits to medical clinics. This will technically differ from from the chicago visualization but we feel the user will still

In the Chicago dataset, the Influenza details are given based on the spread of the flu, as quantified by the Influenza-Like Illness (ILI) activity level in our dataset. ILI is defined as when patients come in with an influenza-like illness even if it is not confirmed to be influenza. ILI risk level is defined as the current ILI percentage for the week compared to the mean ILI percentage during non-influenza months. This data is organized by zip code. This is different from the New York dataset where the influenza rates are simply determined by the total number of visits where patients displayed influenza-like symptoms. For New York and California, we have details on how many people got affected by the flu.. The new data we need is a geojson file (to overlay the state shape, we will need geojson files for it) of zip codes, latitude, longitude, zip codes to have the match between data semantics and the task Network semantics don't make sense for the dataset we are using, so we decided not to do so. Transforming the whole dataset from tabular to network would simply be too cluttered and shows no feasible benefits. There are also limitations in the dataset, namely a lack of useful columns. The effectiveness of the visualizations depends on the representation of the transformations, colors and marks & channels. As discussed in the colors section, colors are an effective way to communicate the differences in the amount of influenza cases, which is why we transformed the data to encode sequential color. Transformations allow for the derivation of data to create similarly comparable though not identical visuals as seen in figures 2.1 and 2.2.

Interaction

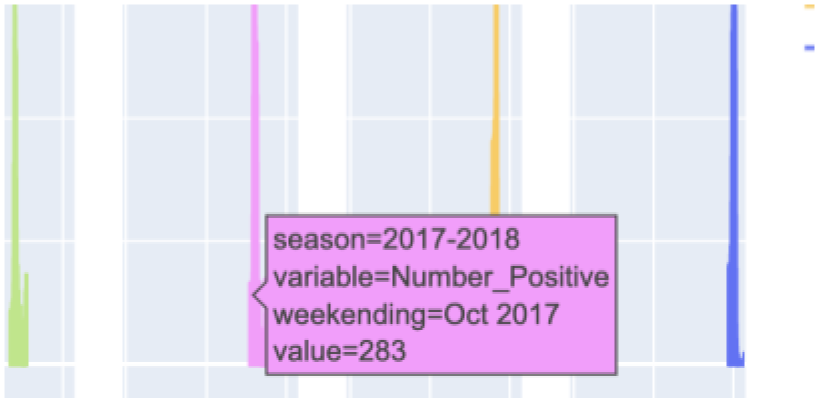


Fig 3.1 shows the interactivity the user can perform by hovering over the selected data with their cursor. See Figure 4.2 for deeper analysis.

Coordinating multiple views allows users to gain insights from different perspectives seamlessly. In our visualization, two main views employed are initial overview and a detailed on-demand view. The initial overview provides a broad perspective of trends over time in the zoomed out version which is fig 4.2, while the detailed view offers specific information based on detailed information to help the user understand. By showing different time periods or regions in Fig 4.2, interaction provides a dynamic view of the trends of influenza throughout different years or zip codes of states. When users hover over a data point on the graph, animation can be employed to smoothly reveal the tooltip with season, value, variable, and week ending details. This interactive graph adds a dynamic element to the interaction, making visually appealing for users by allowing users to see an uncluttered view of the trends in fig 4.2, but also being able to click on the bars of Fig 3.1,

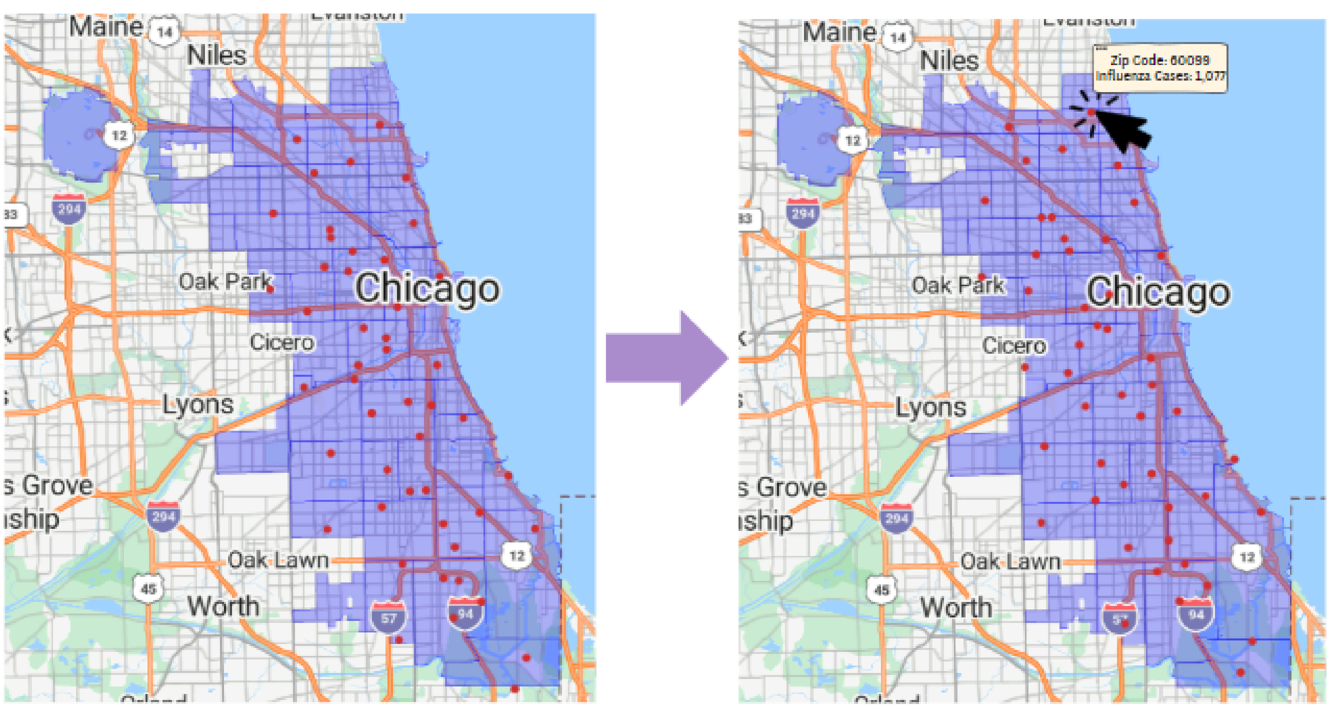


Fig 3.2 Shows a mock up of what the feature implemented in Fig 2.1 will look like on the map overlay of Chicago. Graph shows what interaction is possible for the Chicago map. The blue is just an outline of where the zip codes lie on the graph. The red dots encode different zip codes that allow for more zip codes to be hovered over and more information to be displayed.

Marks and Channels

After experimenting with various combinations of marks and channels to represent the number of specimens tested for each respiratory virus during the specified year, it's evident that scalability plays a crucial role in determining the effectiveness of the visualization. While a bar chart using position, color, and bars as visual channels and marks prove to be a straightforward and intuitive option, too much data can lead to clutter and reduced readability. However, other visual channels such as size, and shape can also be employed to encode additional attributes. For instance, the size of the circles could represent the number of specimens tested. Each of these visual channels encodes either sequential (fig 2.1) with saturation or categorical attributes (Fig4.1) with hue, providing flexibility in representing various aspects of the data. The use of multiple marks in the same vis, such as lines and points, can allow for the simultaneous representation of multiple attributes. Interference is occurring within the color channel and size channel due to the use of multiple similar shades of yellow and similar sizes in the bar chart making them hard to distinguish . This can make it difficult for users to distinguish between categories or data points effectively, undermining the clarity and interpretability of the visualization. In the upcoming phase, we will be working on the issue of color interference with more distinct colors in the bar chart to ensure maximum clarity for users.

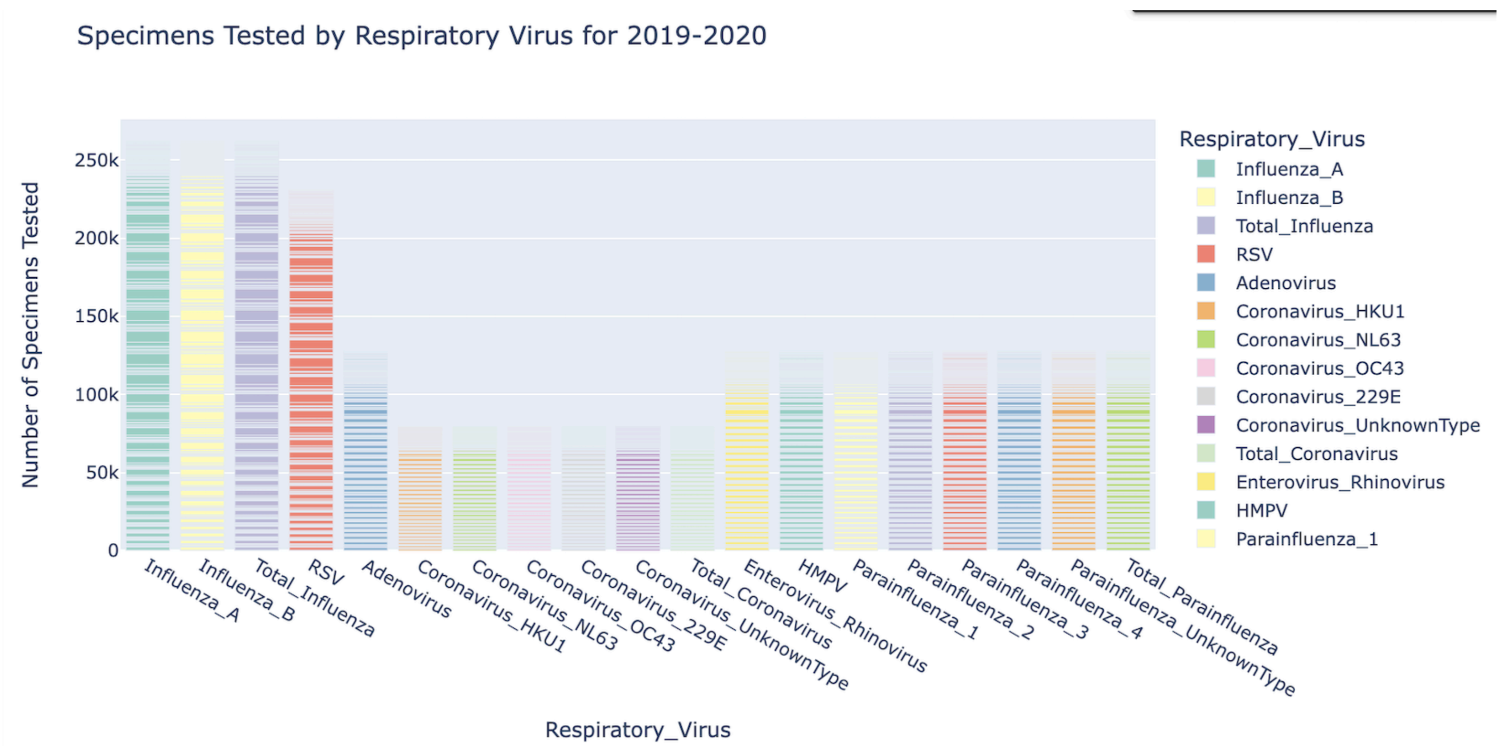


Fig 4.1 shows the numbers of specimens tested for each respiratory virus during 2019-2020.

The visualization exemplifies the Marks and Channels theme by offering an interactive exploration of the number of specimens tested for each respiratory virus throughout the designated year because it has a hover feature on each bar with additional information. This visualization provides a comprehensive view of the testing activity for various respiratory viruses over time, allowing users to identify trends, patterns, and fluctuations in testing volume for each virus throughout the specified year. Additionally, users can compare testing activity between different respiratory viruses to gain insights into their relative prevalence or impact during the specified time period. The graph answers the task: “Show users rates of outbreaks of different types of flu.” This allows users to see what kinds of flu that are represented in our graphs and to broaden their understanding of the kinds of flu that are existent for each of our graphs.

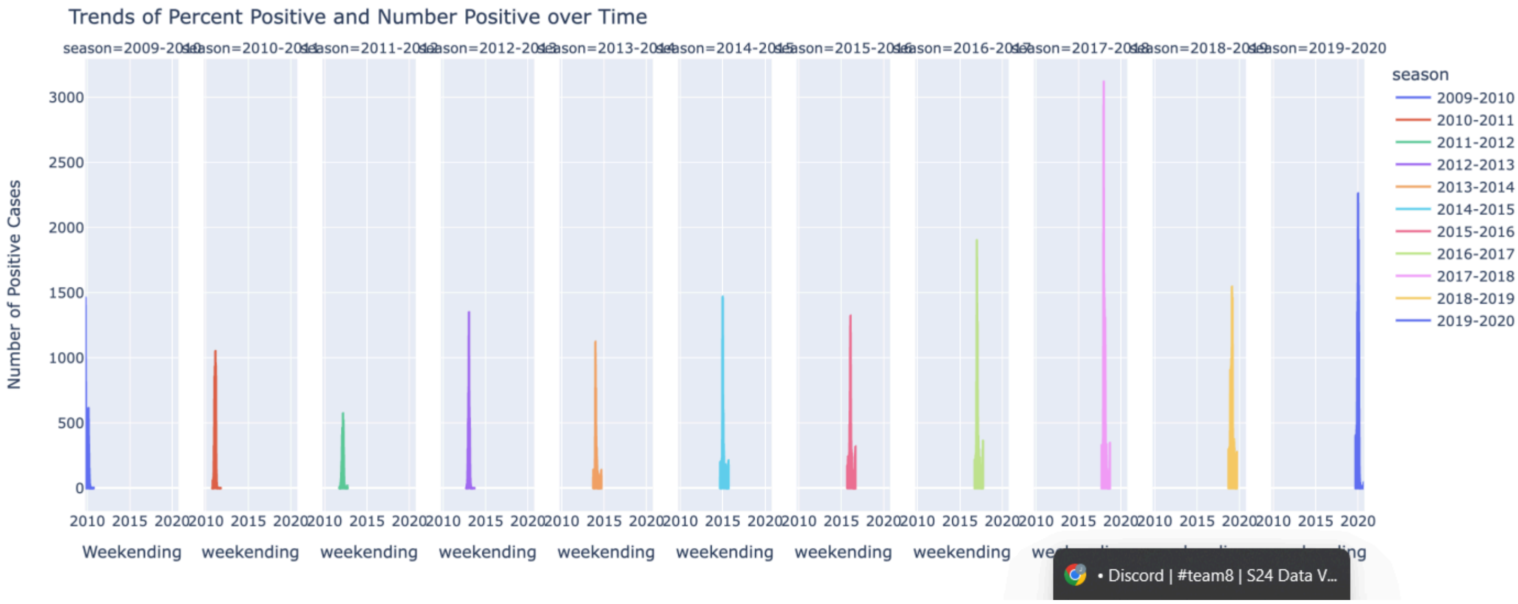


Fig 4.2 Dynamic visualization of influenza trends showcasing the number of positive cases over time, segmented by region and season

Centered around a theme of marks and channels, the visualization Fig 4.2 showcased offers users to explore trends in the percentage of positive influenza cases and the number of positive cases over time, segmented by different year. This visualization enables users to analyze how the influenza cases vary across different years, providing insights into potential temporal patterns in the spread of the virus.

For geographical graphs, such as Fig 2.1, area will be used as the main mark. Advantages of this could be an intuitive comparison between different areas' influenza rates. The main channel for this would be color hue which would denote the categorical attribute of influenza rate in fig 4.2 and for sequential data such as fig 1.2. For example one color, such as blue, would denote no influenza while another opposite color, such as red, would indicate the most influenza while colors in between such as purple would denote a value in the middle. Another mark would be one of the uniformed size dots representing each zip code area and with color saturation to denote the rate of the flu. This does not have an advantage over areas, but is already coded so we can guarantee that this is a potential mark to be used. We are planning on using color hues, and channels lines to answer this task: "The user will compare cases in areas with high vaccination rates to cases in areas with low vaccination rates" to easily visualize the differences. Because these tasks compare a high rate of influenza and a lower rate of influenza, having bars would be more useful and clear than a heat map or a state map.

Misleading Visualization

We are planning on having two map visualizations, one of Chicago and one of New York. Both will encode population density and influenza rate by color. Population density will color in the area of a certain zip code (fig 2.3) while influenza rate will color a dot within that zip code (2.1). One misleading part of these visuals is that there is a difference in the way "influenza rate" is defined in the Chicago dataset and the New York dataset due to the data available to us at this time. The Chicago dataset is defined by the amount of Influenza Like Illness (ILI). This means the amount of people who come into the hospital with an influenza-like illness in a certain week compared to the mean ILI percentage during non-influenza months. In contrast, the New York dataset does the same thing, but taking the mean ILI percentage year round, not during influenza only months. This can be misleading as Influenza Rate is being encoded and presented similarly while using a different metric to define it. We plan to mitigate misunderstandings through descriptive captions in the final product. If we had more time to complete this task, we would likely attempt to match the definitions of Influenza Rates to each other by changing the New York dataset to include the mean ILI percentage during non-influenza months. Unfortunately due to time constraints this might not be possible before the due date.

Beta Release

Alpha Release Feedback Summary

The feedback from Team 3 was useful in certain areas. Multiple comments referenced using a different color scale for the Geographical maps. This can be changed to a more distinct color scale (to be determined) which should make the visualization easier to understand. Another suggestion was to create a map of influenza cases throughout the entire U.S.. We felt this is too broad for our specific tasks and would be better represented using the 3 selected cities and overlaying them onto the map of the U.S so each city will have its own visualization and be overlaid onto the whole U.S. map so the cities can be juxtaposed. Another member of team 3 mentioned creating a visualization of future cases based on current data.

Professor had the following feedback on the visualizations: She mentioned using a diverging color scale for Figures 1.1, 1.2, and 2.3 to contrast the different regions more distinctly. She also mentioned for Figure 4.2, the scale needs to have a different aspect ratio. The axes also need to be formatted wider so that they are clearer. We used this feedback to create a different bar graph (figure 7.1) that allows the user to better understand the task. We also fixed many of the professor's comments in our process book.

Some tasks were updated, we removed seasonal from “The user will be able to find the peak periods of influenza within different regions or states.” as we no longer felt it was necessary to show. We also removed the task “The user will compare cases in areas with high vaccination rates to cases in areas with low vaccination rates” as we did not have a vaccination dataset. We replaced it with “Show users the rates of different kinds of flu present. This allows users to see what kinds of flu are in which areas.”. We felt this was an effective task to convey influenza information to the user. A new task was added, “Show users the rate of the state's flu for different years.”. Creating a graphical visualization of California had too many issues so we opted to use a bar graph (Figure 7.1) and line graph (Figure 7.2). We chose this task for our visualization to represent because we felt it was important for the user to understand the change in influenza rate over time as the other visualizations do not.

Final Form

Chicago

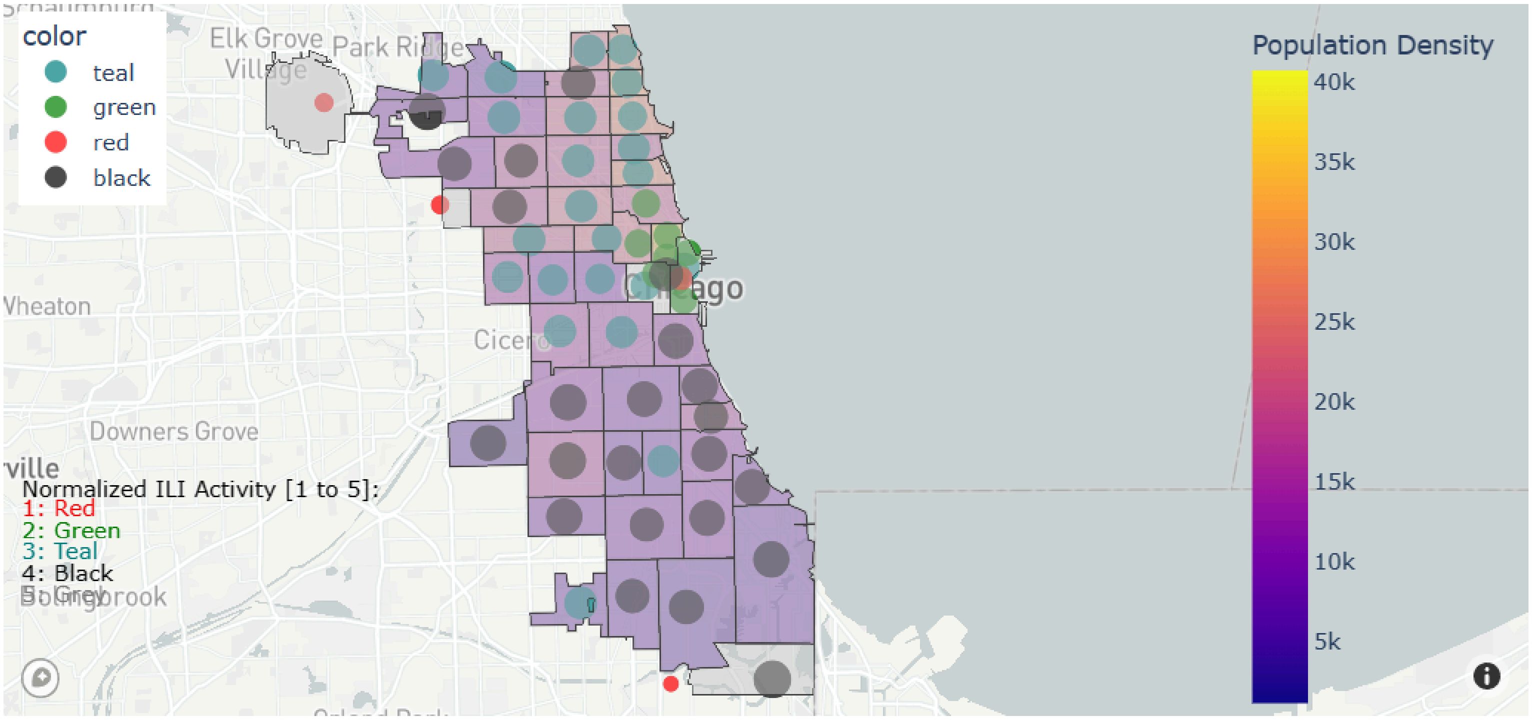


Fig 5.1 Represents a geographical choropleth map of the population density of the different zip code regions of Chicago with a secondary choropleth of circles representing the number of influenza cases. The size of the circle does not denote anything, and instead represents a specific zip code location. The Normalized ILI Activity represents the affected regions of the influenza where 5 being the highest (most cases) and 1 being the lowest (least cases). Using the color legend, Normalized ILI Activity can be filtered out.

We plan to have 2 state graphs, each measuring the influenza rate and population density. Population density will be shown via a choropleth . Figure 5.1 represents the improved version of Figures 2.3 and 2.1. We used a more distinct color scale and combined both features of the number of cases and population density. We plan to convert the current color scheme for the influenza dots to a different color scale but same style as the population density. This visualization also uses interactivity to allow the user to view the numerical value of the zip code and population density, as seen in figure 5.2. By clicking the color dot, it removes every flu case with that color level of flu.

The tasks in this graph will answer: "The user will notice if there is a correlation between influenza case rate and population density" by using the color scale and "The user will be able to locate the state/area that is most likely to have an outbreak of influenza based on influenza case rate and population density" based on the color scale and comparing it to the map.

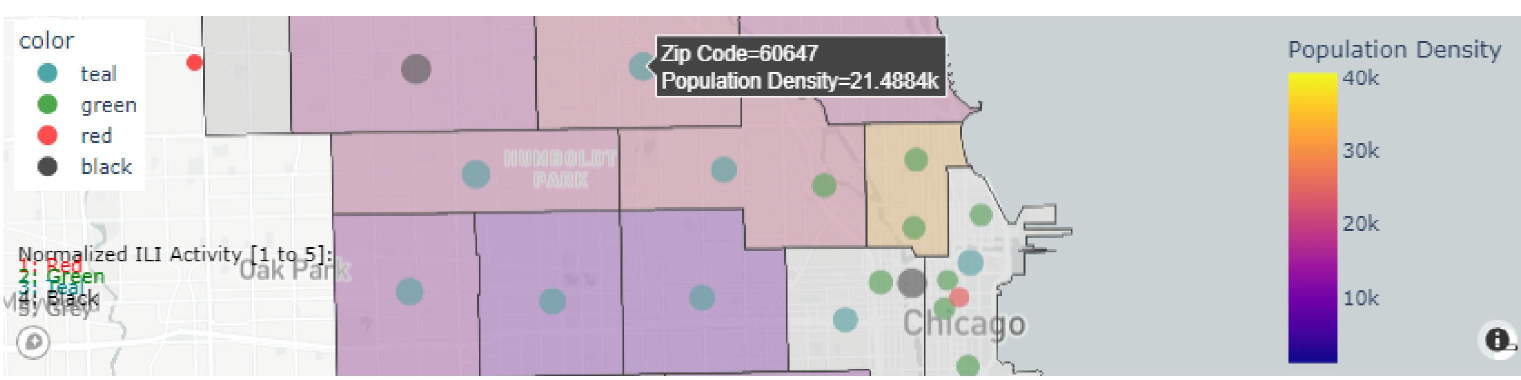


Figure 5.2 Represents a zoomed in version of Figure 2.3.2 showcasing the interactivity of the visualization.

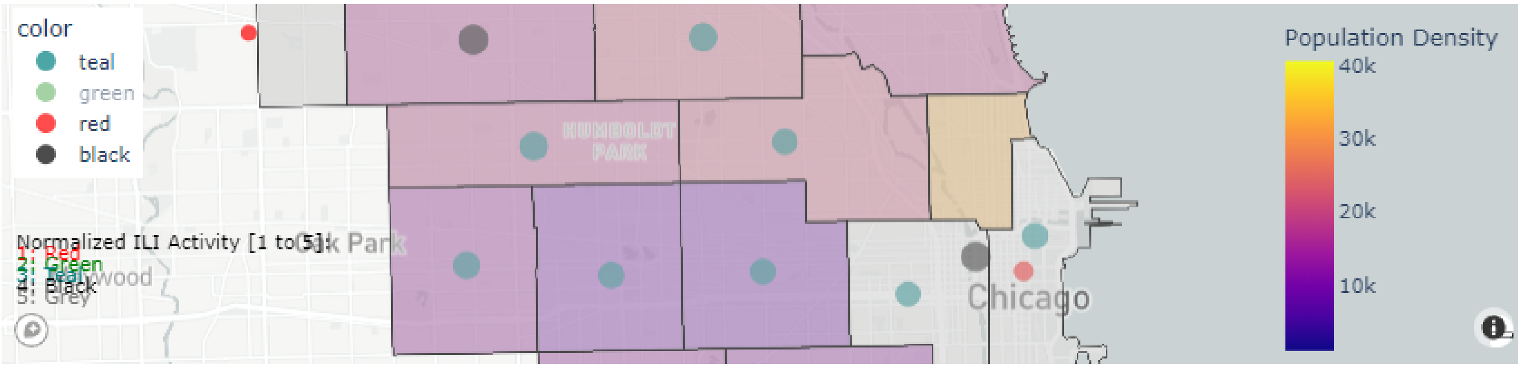


Figure 5.3 Represents the interactivity of the color legend by turning the green (level 2) attribute off, removing those marks from the visualization..

New York

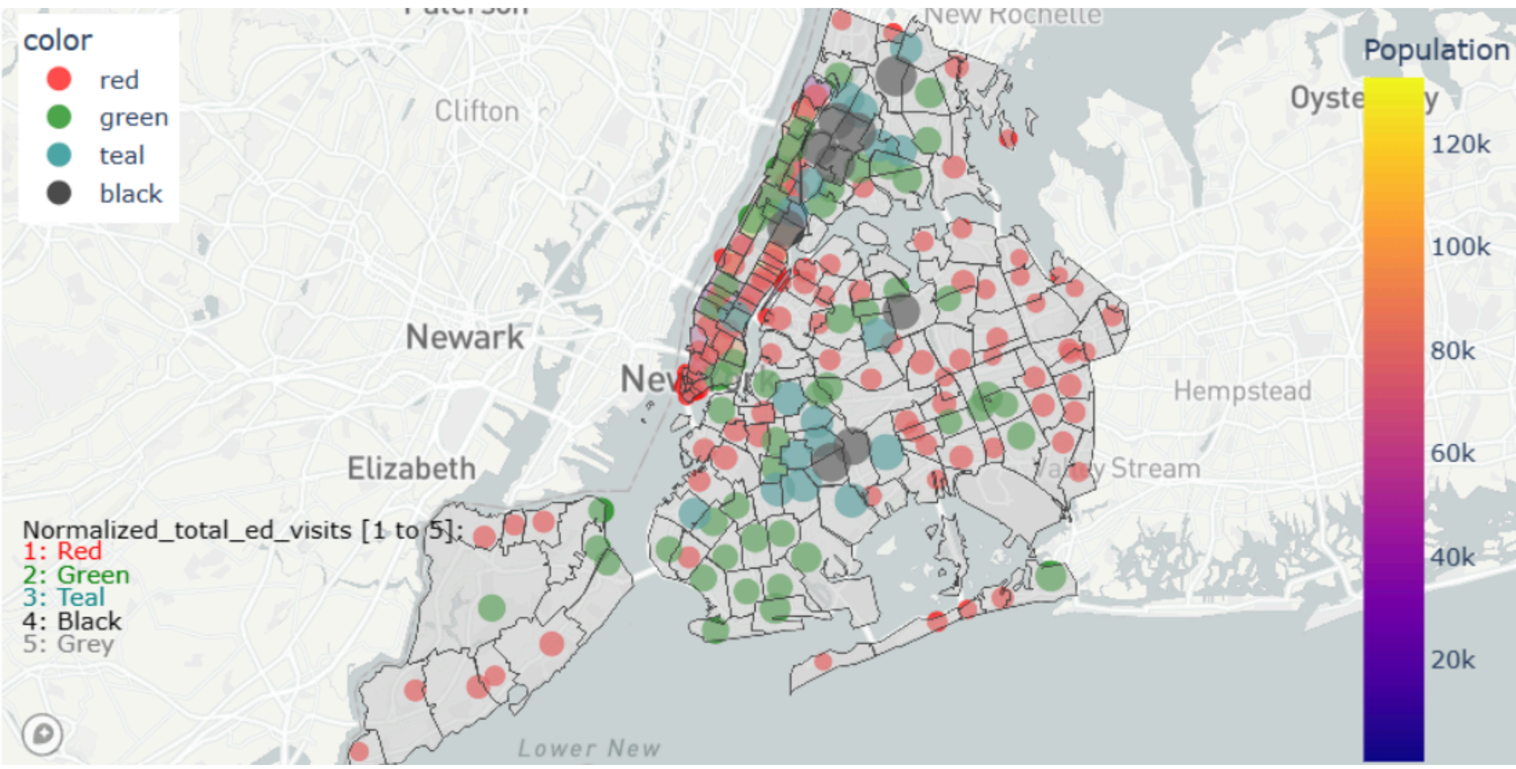


Fig 6.1 Represents a geographical heatmap of the population density of the different zip code regions of NYC with a secondary choropleth of circles representing the number of influenza cases. The Normalized ILI Activity represents the affected regions of the influenza where 5 being the highest and 1 being the lowest. With the color legend Normalized ILI Activity can be filtered out.

Fig 6.1, population density is not fully visible because our geojson did not have zip codes for every part of New York and the areas that are colored in for population density are too difficult to see zoomed out. Fig 6.2 is a zoomed in version to show that we don't have an entire NY zip code geojson but the part that we do have is how all of NY will look once we find a new geojson to use. Overall, NY Fig 6.1 will look the same to Chicago Fig 5.1.

The tasks this graph will answer: "The user will notice if there is a correlation between influenza cases recorded by hospitals and population density" by seeing if the colors of both are similar which will be implemented by giving both a different color wheel , and "the user will be able to locate the state/area that is most likely to have an outbreak of influenza based on influenza case rate and population density" as the user will be able to use the color scale to determine area with the most influenza cases.

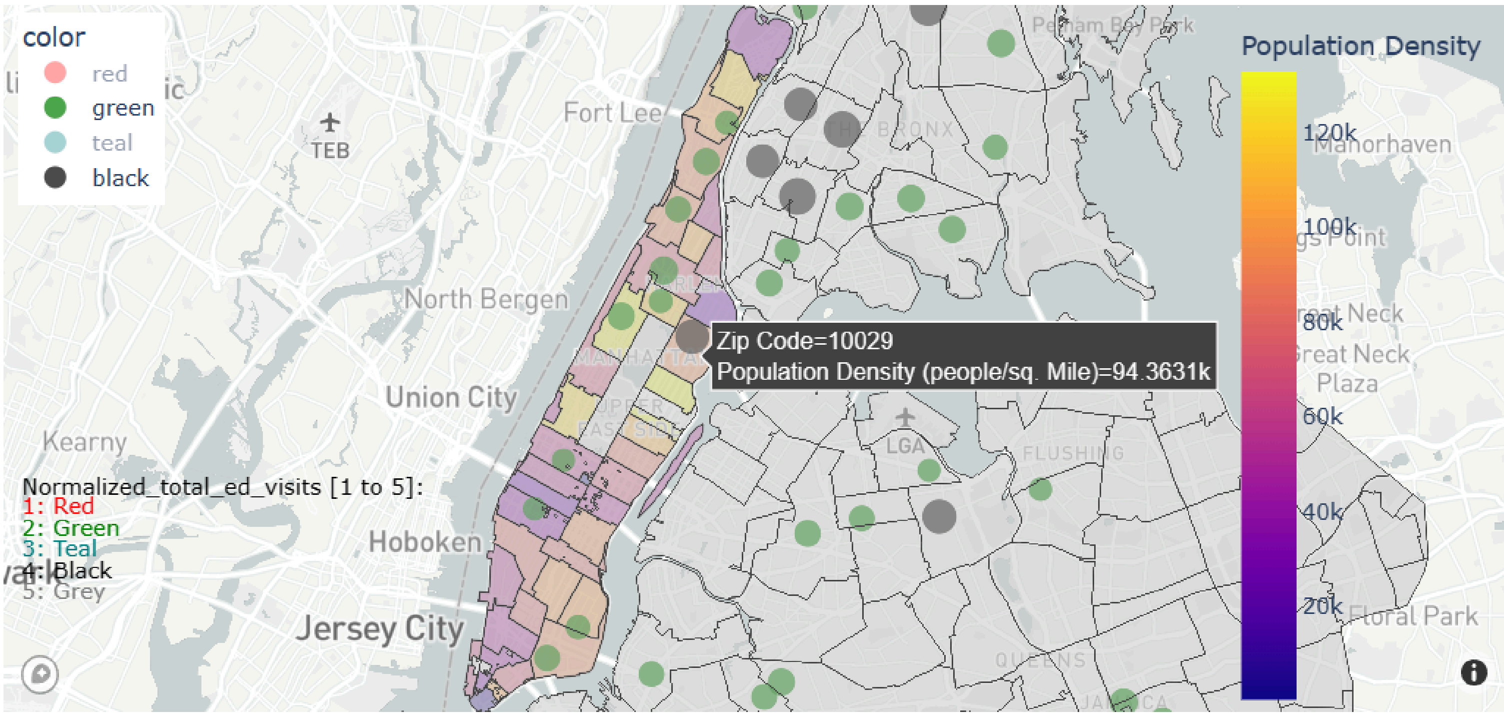


Fig 6.2 showcases a zoomed in version of Fig 6.1 highlighting the population density and zip codes.

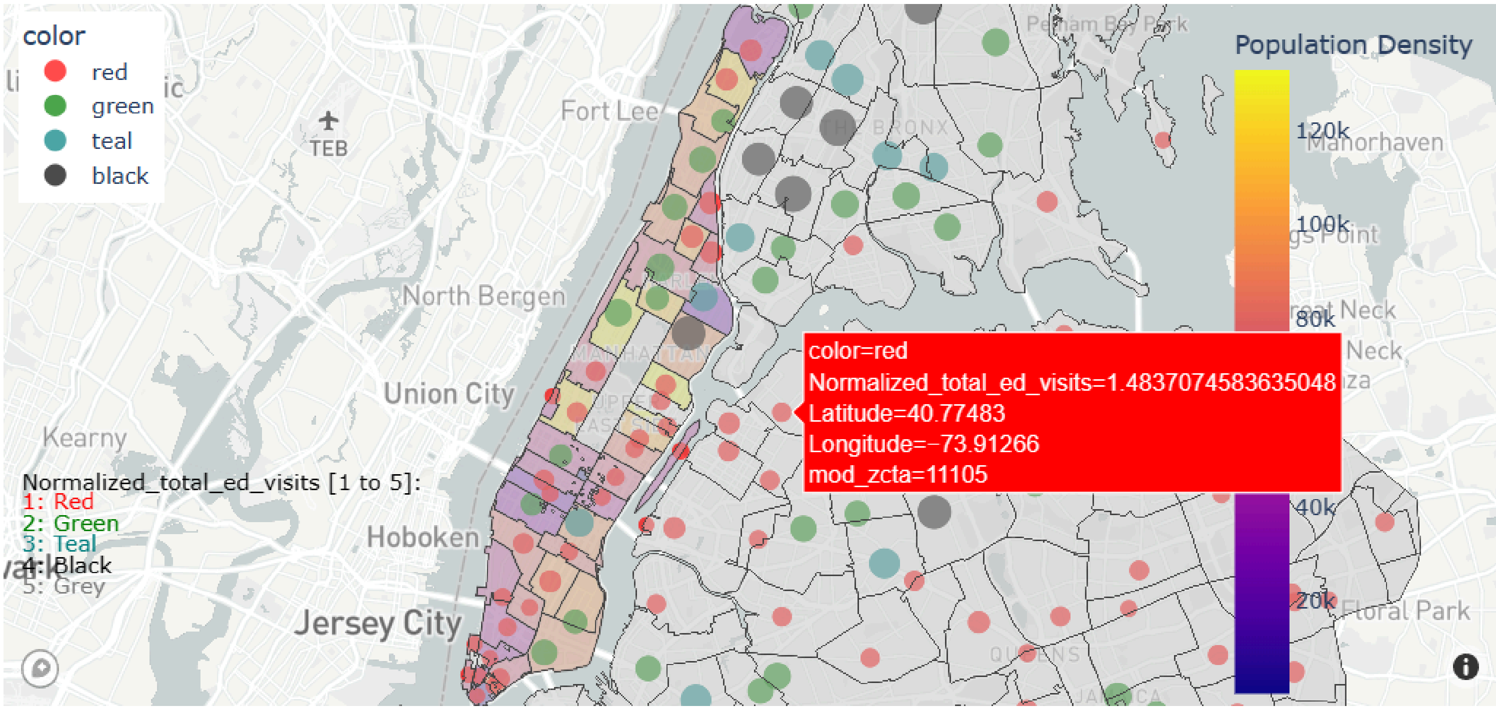


Fig 6.3 showcases a zoomed in version of Fig 6.1 highlighting the normalized total ed visits and the geographical coordinates.

California

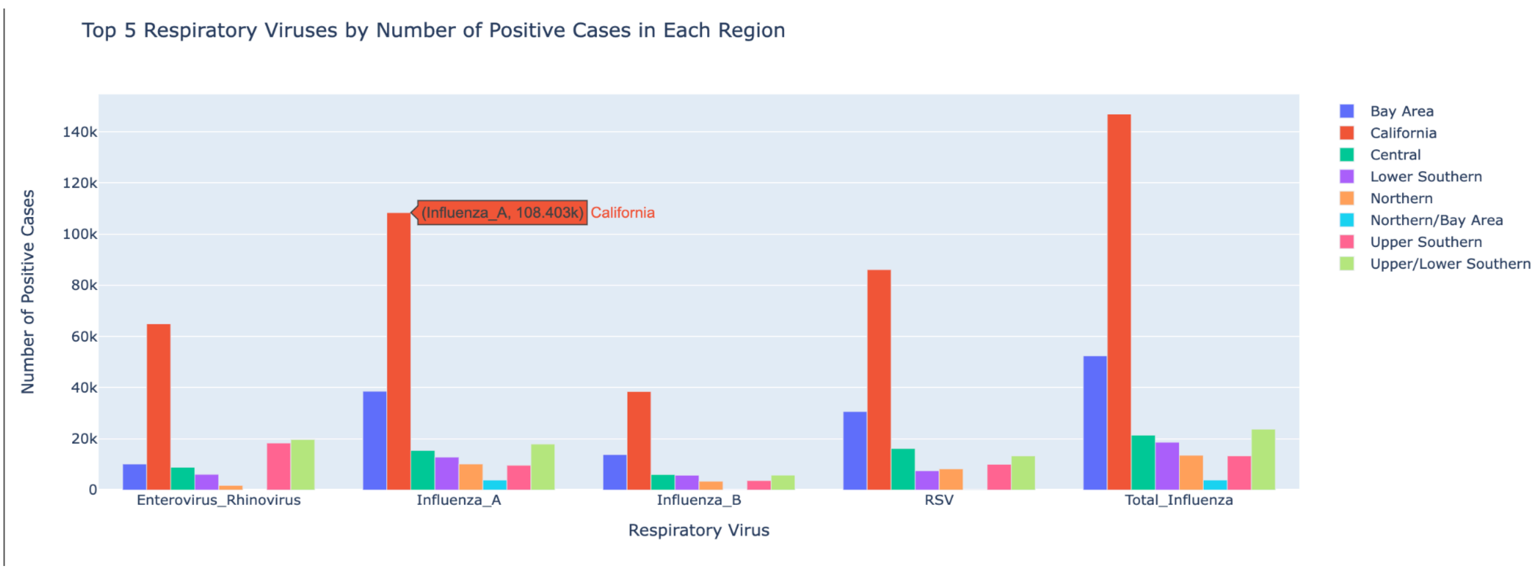


Fig 7.1 represents the total number of cases for the different influenza-like illnesses in the different regions of California over the years 2009 - 2020.

The original bar chart contained discontinuous bars that were not necessary which were removed. Interference between color and size was a problem so distinct colors are used.

This will answer the task: "show users different kinds of flu that are present" because the user will be able to see how severe each kind of flu is. "The user will be able to find the peak periods of influenza within different regions or states" because they will be able to locate their region's color and see the number of cases their region has and how it compares to the other regions.

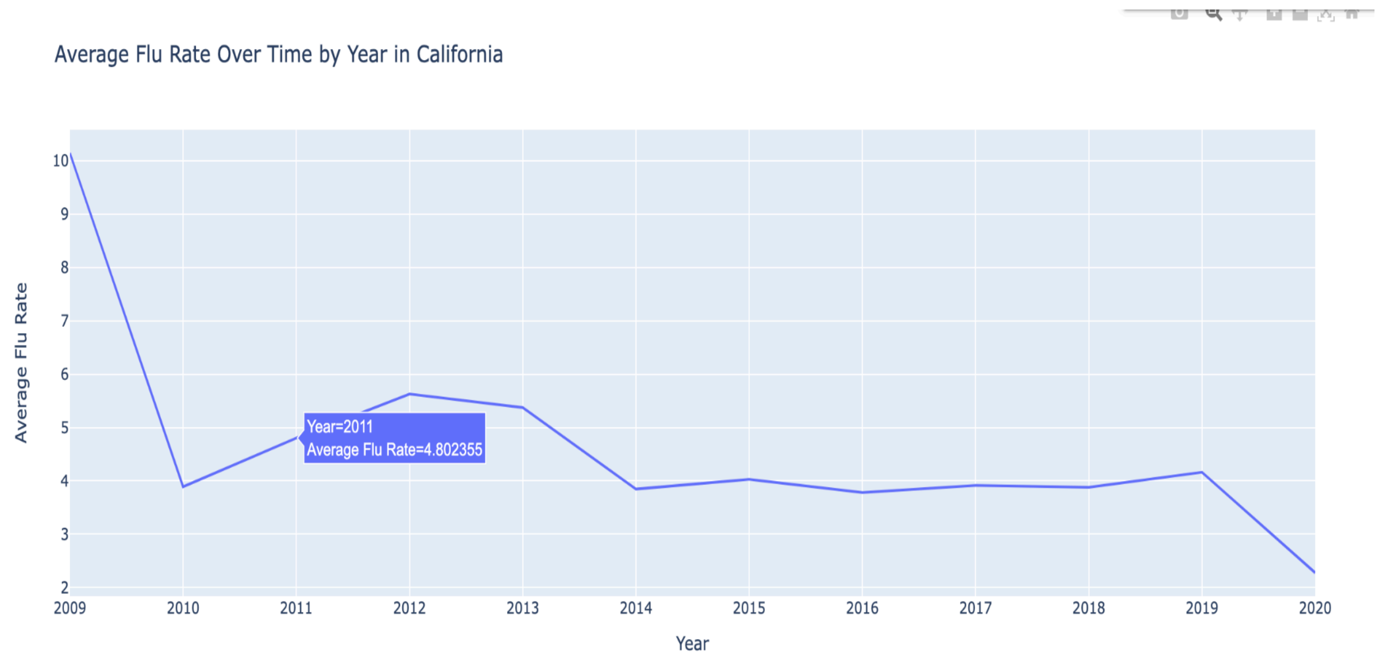


Fig 7.2 represents the average number of cases of influenza-like illnesses in California over the years 2009-2020.

This data showcases every flu-like illness reported to the hospitals in different regions of California. The regions used are Bay Area, Central, Lower Southern, Northern, Northern/Bay Area, Upper Southern, Upper/Lower Southern and California as a whole. This helps users visualize which of the top 5 flu-like illnesses are the most prevalent in California. The graph is also interactive allowing the user to hover over a bar to view more details. The original bar chart for the year progression of flu was originally thin and hard to see, so we got rid of a bar chart and opted for a line graph.

This will answer the tasks: "Show users the rate of the state's flu for different years,"

Juxtaposed Visualizations

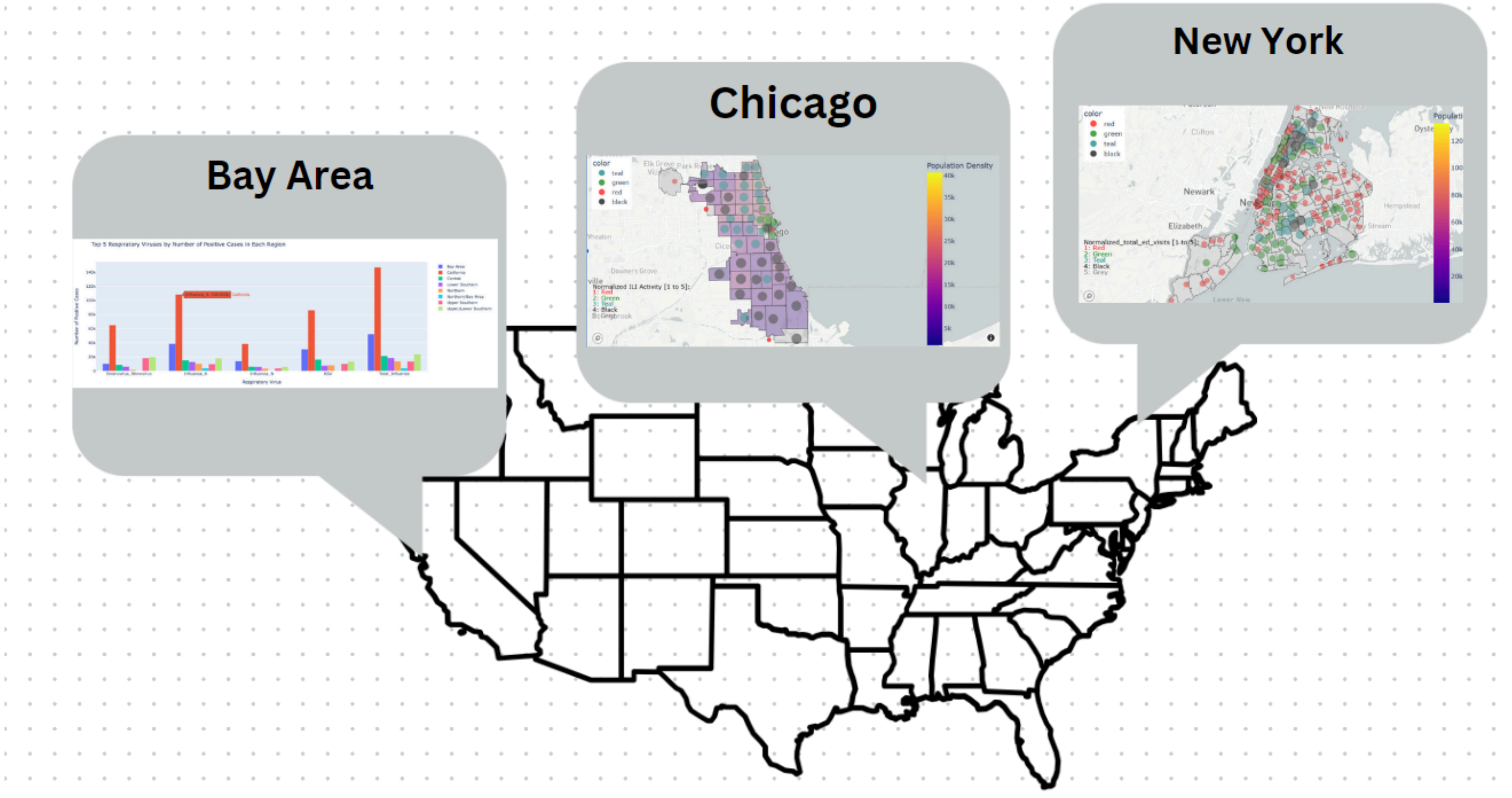


Fig 8.1 shows the juxtaposed view of Figures 7.1, Figure 5.1, Figure 6.1. It illustrates how influenza affects these 3 cities.

This mock up graph will be the whole USA map with a highlighted preview of the cities we focus on including California, New York and Chicago. The user will either click on a state or hover to view the individual graphs we have. This allows for a broad overview of the sum of our data and tells the user exactly what they can expect from our dataset.

This will answer the tasks: “The user will be able to locate the state/area that is most likely to have an outbreak of influenza based on the highest influenza rate” because they will be able to compare colors for 2 states and numbers for all 3 states.

Extra Credit

Prompt: Dive deep into human visual perception and psychophysics as it relates to your project.

For our project we utilized the way the human brain perceives visual information to communicate our tasks effectively. This was done through the utilization of several different visual tools. We showed data on a geographical map in order to communicate location more effectively. In order to compare two different attributes of data we overlaid population density and influenza cases using a heatmap and color scale respectively. By making a heatmap to convey our information, color encoding the quantity of cases, allows for easier pattern recognition as well as identifying any geographical patterns present. Through the use of color hue, spatial positioning, and juxtaposed views, we were able to effectively utilize the brain's order of perception. Spatial region is the most effective categorical attribute followed by Color hue. Both of these channels are heavily utilized in Figures 2.3.2, 2.2.2, and 5.1. Spatial region allows the user to quickly grasp the location of the visualization. Color hue was used to represent and contrast the population density with influenza-like illnesses. By using the top 2 most effective channels, we can quickly begin utilizing the brain's visual capabilities. Juxtaposing our visualization took advantage of the “Eyes Beat Memory” Rule of Thumb allowing the user to more quickly and effectively compare the influenza statistics between areas.

Final Release

Chicago:

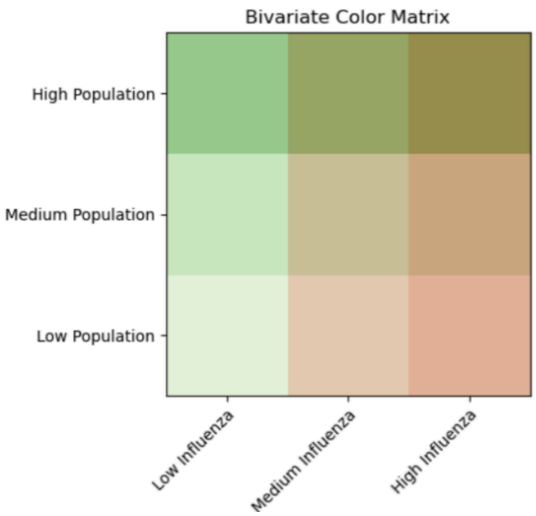


Fig 9.1: shows the bivariate legend for Fig 9.2 and Fig 9.3, where the green color encodes the population and the pink color encodes the Influenza. The dark color encodes the highest and the light color encodes the lowest for both the population and Influenza. The 3x3 bivariate matrix represents the combination of both Population and Influenza for low, high and medium values.

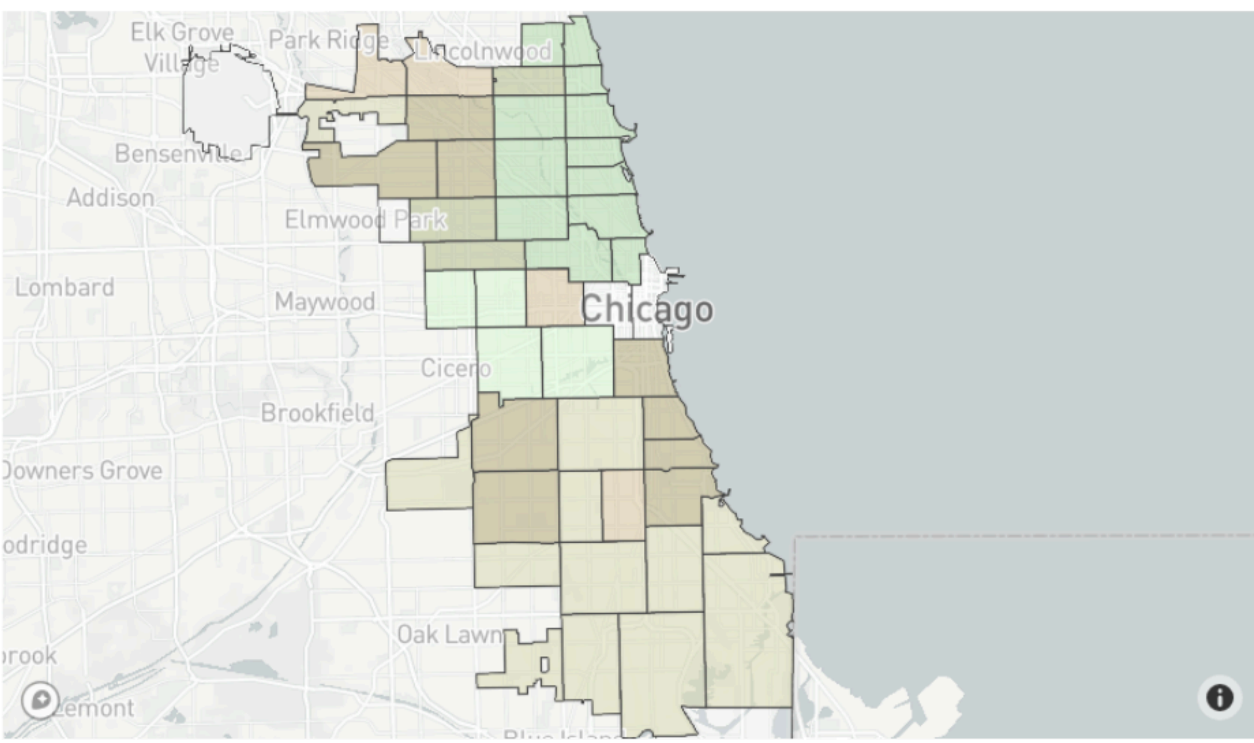


Fig 9.2: Shows the bivariate choropleth for Chicago of influenza rate and population density, where green represents high population and low influenza in that area and pink color represents High influenza among the low population region. The combination of colors are represented in the bivariate matrix to encode the details of the population and influenza spread there.

New York:

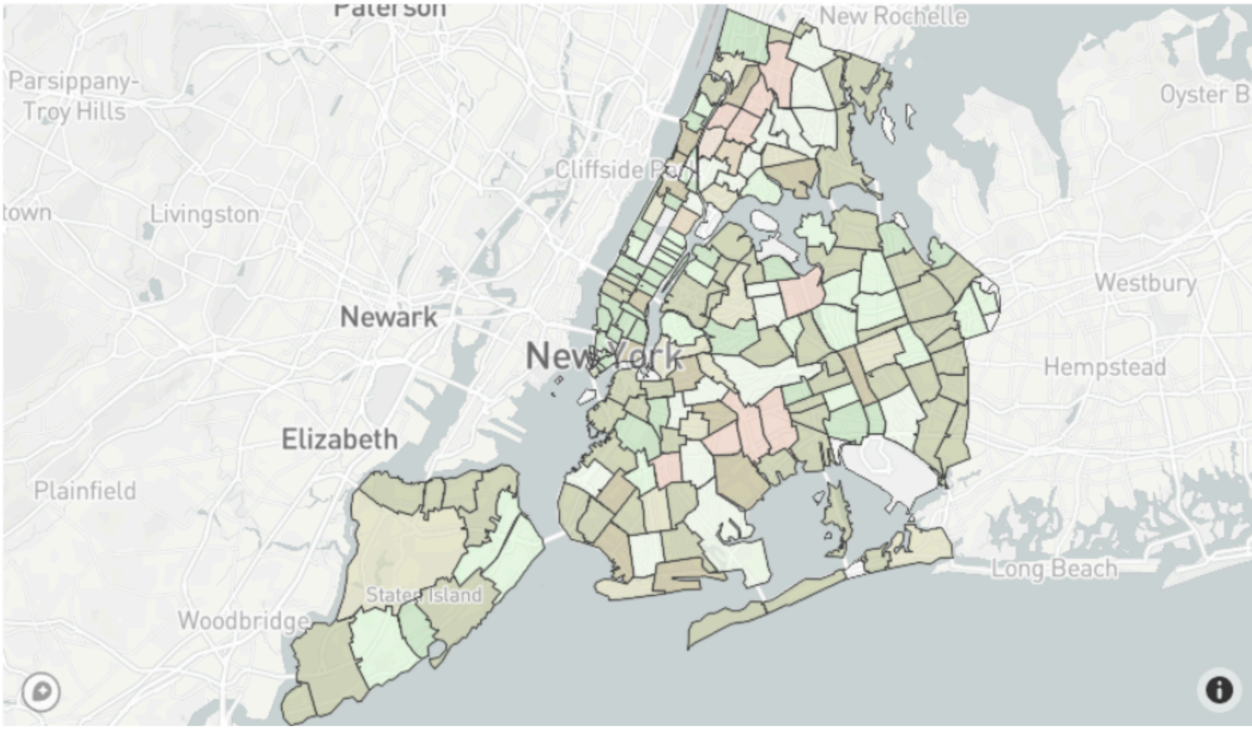


Fig 9.3: Shows the bivariate choropleth for New York of influenza rate and population density, where green represents high population and low influenza in that area and pink color represents High influenza among the low population region. The combination of colors are represented in the bivariate matrix to encode the details of the population and influenza spread there.

<u>California:</u>

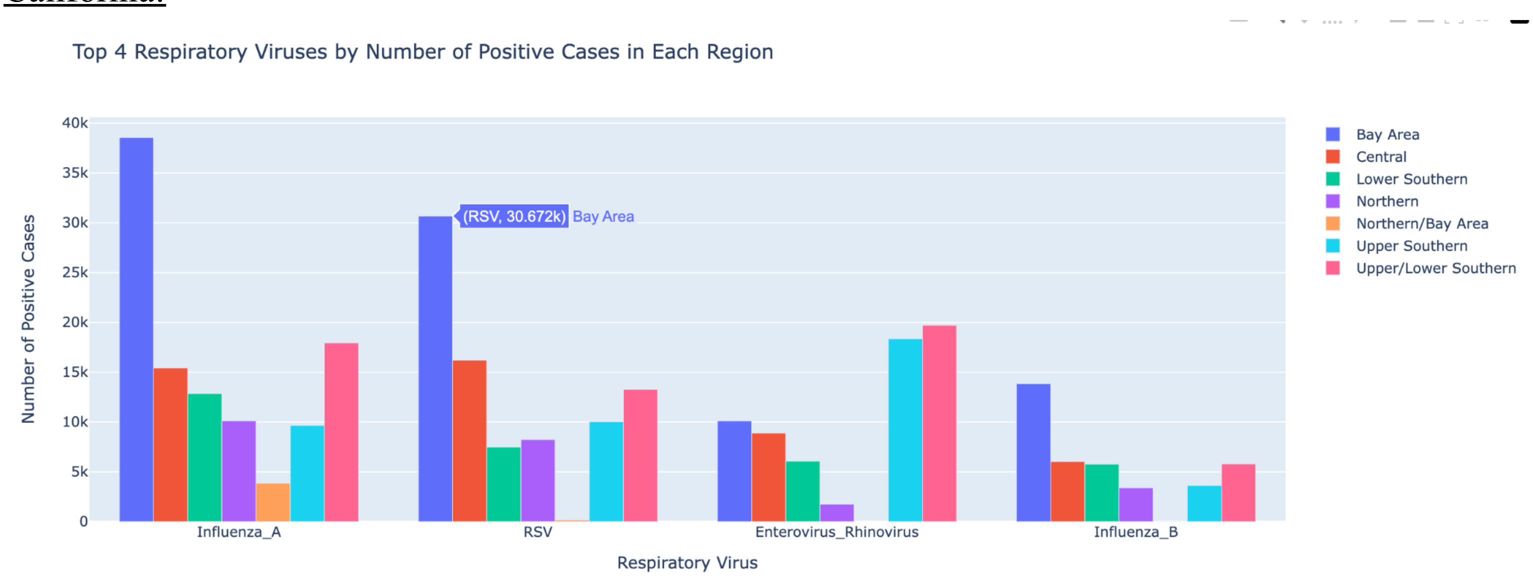


Fig 10.1 This visualization showcases the top four respiratory viruses by the number of positive cases in each region of California. Each region is represented by a distinct color, with bars indicating the number of positive cases for each respiratory virus. Hovering over a bar provides additional information, including the number of positive cases and the corresponding region. This visualization facilitates a comparative analysis of respiratory virus prevalence across different regions of california.

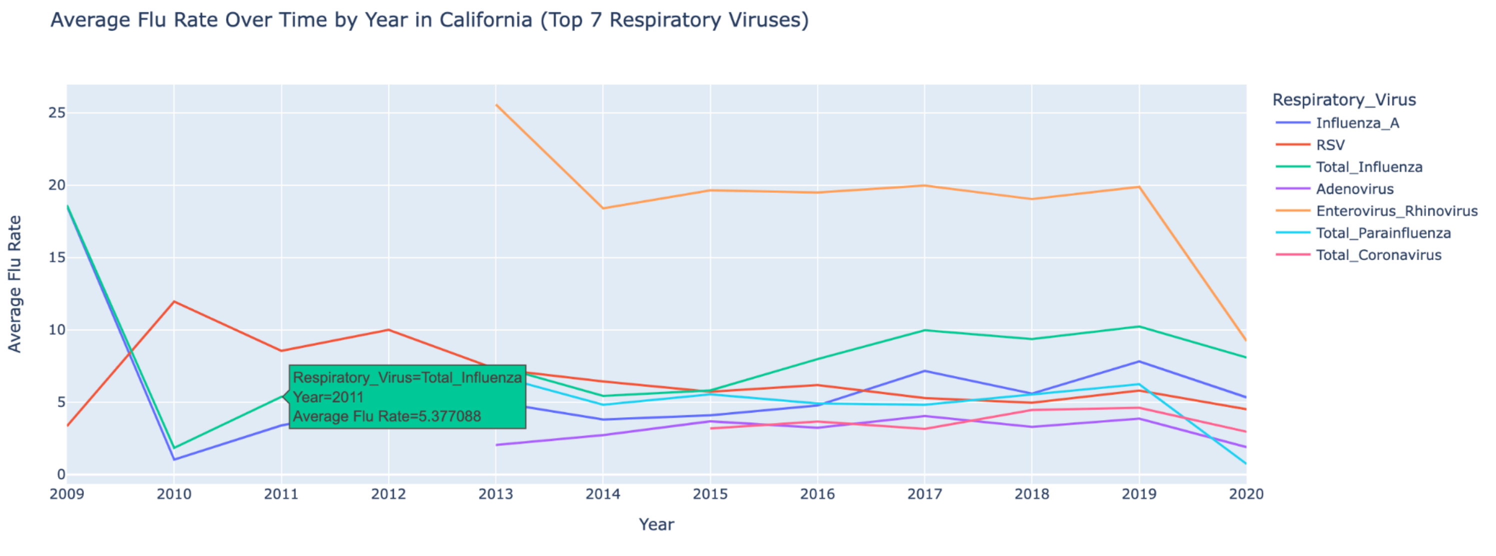


Fig 10.2 This visualization illustrates the average flu rate over time in California, focusing on the top 7 dominant respiratory viruses. The top 7 respiratory viruses are determined based on their mean flu rate. Each line in the visualization represents a different respiratory virus where the x-axis represents year, illustrating the change over time. This visualization allows viewers to understand the temporal patterns of flu rates for the most prevalent respiratory viruses in California.

<u>A short summary of what the user will experience and the user tasks that can be accomplished:</u>

The user will be able to experience viewing the flu rates in their city. This is calculated by flu rate and population density. The user will be made aware of this with captions under the graph. The user will be able to see which part of Chicago or New York by zip codes is the worst or best. For California, the user will be able to compare data in the forms of a line graph and a bar graph. The bar graph will show which case of flu-like disease is not only present in California but also which is the most common flu-like disease. The line graph will show the user the top 7 most prevalent flu-like diseases from 2009-2020. The user will be able to click on the lines for more information.

The tasks the user will be able to accomplish are: “The user will notice if there is a correlation recorded by hospitals and population density,” and “The user will be able to locate the state/area that is most likely to have an outbreak of influenza based on the highest influenza rate” through the choropleth maps, “Show users the rate of the state's flu for different years.” by the line graph, “The user will be able to find the peak periods of influenza within different regions or states,” and “Show users the rates of different kinds of flu present” by the bar graph.

References

Related work by WHO:

https://www.who.int/teams/global-influenza-programme/surveillance-and-monitoring/influenza-surveillance-outputs

T work by CDC: https://www.cdc.gov/flu/spotlights/2023-2024/22-23-summary-technical-report.htm

"Counties in California - GeoJSON - WIFIRE Commons Data Catalog." *Wifire-Data.sdsc.edu*, wifire-data.sdsc.edu/nl/dataset/counties-in-california/resource/1dba9680-5ed6-4061-b070-165d394b0508?inner_span=True. Accessed 8 Apr. 2024.

Emergency Department Visits and Admissions for Influenza-... - dataset by city-of-ny. (n.d.). Data.world. Retrieved March 8, 2024, from https://data.world/city-of-ny/2nwg-uqyg

Influenza Surveillance - dataset by chhs. (n.d.). Data.world. Retrieved March 8, 2024, from https://data.world/chhs/fc544658-35c5-4be0-af20-fc703bc57c13

Influenza Risk Level by ZIP Code - dataset by cityofchicago. (n.d.). Data.world. Retrieved March 8, 2024, from https://data.world/cityofchicago/8vvr-jv2g

Location Inc, "Chicago, IL Real Estate & Demographic Data," NeighborhoodScout, Retrieved Apr. 5, 2024, from https://www.neighborhoodscout.com/il/chicago

Mapbox. "Mapbox." *Mapbox*, Mapbox, 2017, www.mapbox.com/.

PUI2015_EC/mam1612_EC/nyc-zip-code-tabulation-areas-polygons.geojson at master · fedhere/PUI2015_EC Fedhere https://github.com/fedhere/PUI2015_EC/blob/master/mam1612_EC/nyc-zip-code-tabulation-areas-polygons.geojson

"Nyc-Maps/Boroughs.geojson at Master · Dwillis/Nyc-Maps." *GitHub*, github.com/dwillis/nyc-maps/blob/master/boroughs.geojson. Accessed 8 Apr. 2024.

World Media Group, LLC. (n.d.). *New York, NY population density zip code rank*. USA.com. http://www.usa.com/rank/new-york-ny--population-density--zip-code-rank.htm?yr=9000&dis=&wist=&plow=&phigh=

M, M. (n.d.). *PUI2015_EC/MAM1612_EC/NYC-Zip-code-tabulation-areas-polygons.geojson at master · fedhere/PUI2015_EC*. GitHub. https://github.com/fedhere/PUI2015_EC/blob/master/mam1612_EC/nyc-zip-code-tabulation-areas-polygons.geojson