Journal of Systemics, Cybernetics and Informatics (2024) 22(7), 6-15
https://doi.org/10.54808/JSCI.22.07.6

Transdisciplinary Applications of Data Visualization and Data Mining Techniques as Represented for Human Diseases

Richard S. SEGALL
Arkansas State University
Department of Information Systems & Business Analytics
Neil Griffin College of Business
State University, AR 72467 USA
E-mail: rsegall@astate.edu

ABSTRACT[1]

Data visualization and data mining are transdisciplinary tools for predictive and descriptive analytics. This presentation shows the abundance of data visualization tools currently available that are applicable for multi-disciplinary data. Some examples of visualization are presented as applied for multi-disciplines.

Results of applying the most commonly used data visualization tools of Tableau [1] and Power BI [2] are presented as preliminary outputs of a funded Seed Money Grant that applies data for transmissible diseases for humans and also plant pathology.

Tableau Software, LLC is an American interactive data visualization software company focused on business intelligence and can be used across all disciplines. Tableau products query relational databases, online analytical processing cubes, cloud databases, and spreadsheets to generate graph-type data visualizations.[1] Power BI (Business Intelligence) is a suite of business analytics tools developed by Microsoft that is widely used across various industries for business reporting, data analysis, and decision-making. [2]

Future directions of this research are discussed that include using other data mining and visualization methods such as image processing techniques for applications to data for plant pathology and human diseases.

[1]Acknowledgement is made to Venkata Jaipal Reddy Batthula, Assistant Professor of Computer Science, Southwestern College, Winfield, KS USA for editorial review of final submission.

Keywords: Data visualization, Tableau, Power BI, COVID-19, Flu, Pneumonia.

1. INTRODUCTION

According to IBM (2024), “data visualization is the representation of data through use of common graphics, such as charts, plots, infographics and even animations. These visual displays of information communicate complex data relationships and data-driven insights in a way that is easy to understand. Effective data visualization can be achieved if it uses correct data in an attractive design, that tells a story, and is exhibits shareability. [3]

There are many different types of data visualization software and tools. Simplilearn (2024) provides a detailed description of 23 data visualization tools from which a subset is selected as shown in Table 1 below with URL and brief characteristics.

Choosing the correct tool maybe challenging and Tableau is the most used and simplest to use. Plotly for Python is free and open-source software, licensed under the MIT license, and is premier platform for building, scaling, and deploying data apps. Enables Python users to create interactive web-based visualizations that can be displayed in Jupyter notebooks. free, open-source JavaScript library for data visualization, which supports eight chart types: bar, line, area, pie, bubble, radar, polar, and scatter.

Data mining is the process of extracting meaningful patterns and insights from large datasets using statistical, machine learning, and computational techniques. It involves the discovery of hidden patterns, correlations, and trends to inform decision-making and predict future outcomes.

Table 1: 9 Most Popular Data Visualization Tools

Software	Information	Web Address
Tableau	Offer free one-year Tableau licenses to students at accredited academic institutions.	http://www.tableau.com
Google Charts	Powerful, simple to use, and free.	https://developers.google.com/chart/interactive/docs/gallery
Plotly	Premier platform for building, scaling, and deploying data apps.	https://plotly.com/
Data Wrapper	20 interactive chart types & 3 interactive map types.	https://www.datawrapper.de/
Power BI	Part of Microsoft Power Platform interactive data visualization software product with a focus on business intelligence.	https://www.microsoft.com/en-us/power-platform/products/power-bi
Chart.js	Offers additional chart types like radar charts, polar area charts, and doughnut charts, which are not available in CHARTIST.	https://github.com/chartjs
Chartist.js	Provides more options than Chart.js for creating animated, interactive charts such as line charts, bar charts, pie charts, and scatter plots.	https://github.com/chartist-js/chartist
Sigma,js	Open-source JavaScript library aimed at visualizing graphs of thousands of nodes and edges using WebGL.	https://www.sigmajs.org/
Polymaps	Free JavaScript library for making dynamic, interactive maps in modern web browsers.	http://polymaps.org/

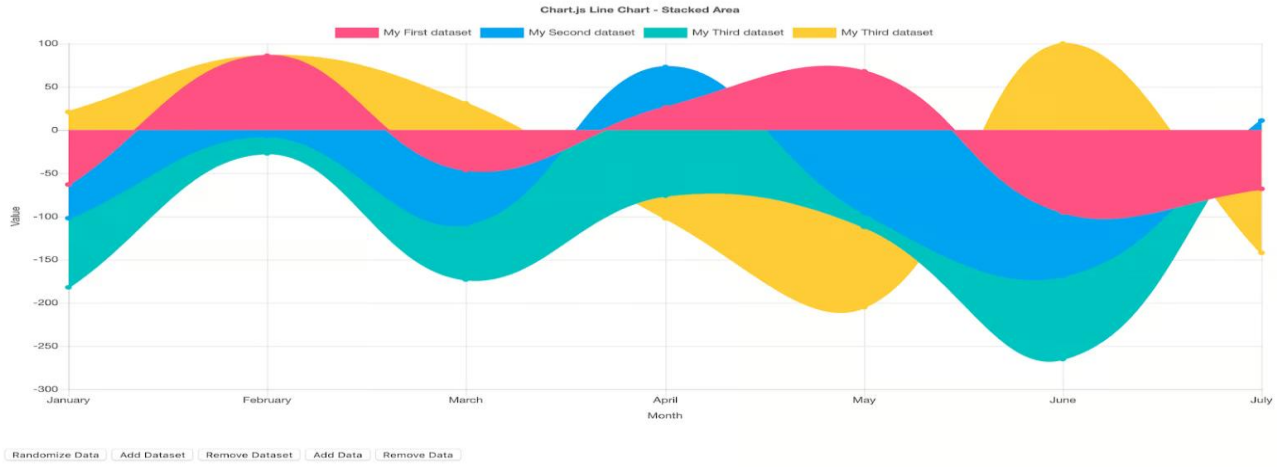

Figure 1: Example of Chart.js Line Chart with Stacked Area
Source: https://www.toptal.com/designers/data-visualization/data-visualization-tools

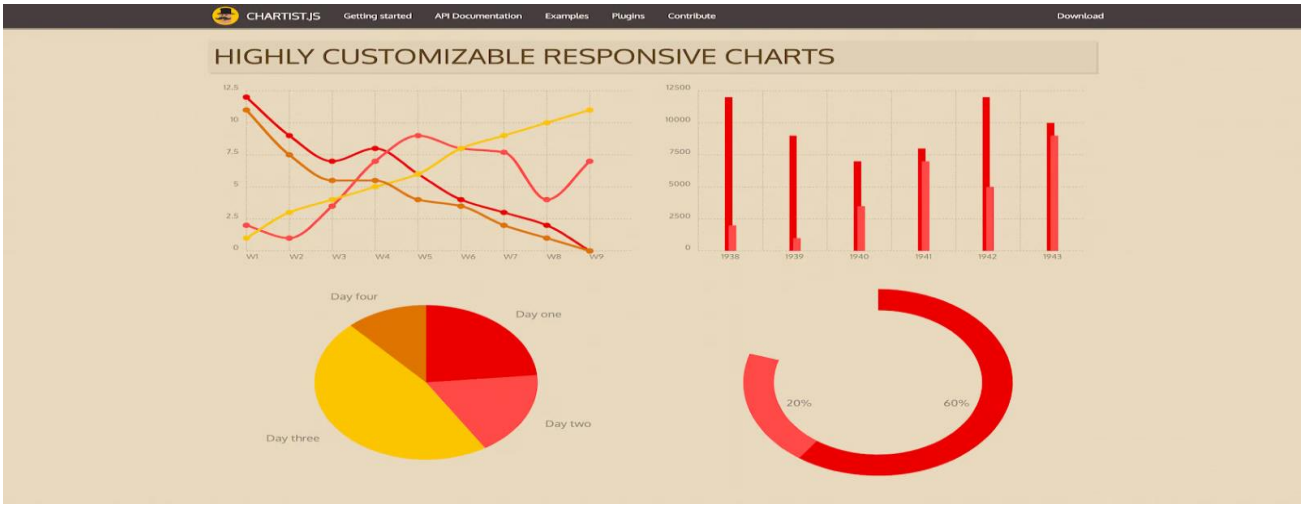

Figure 2: Example of Chartist.js Visualization Software
Source: https://www.toptal.com/designers/data-visualization/data-visualization-tools

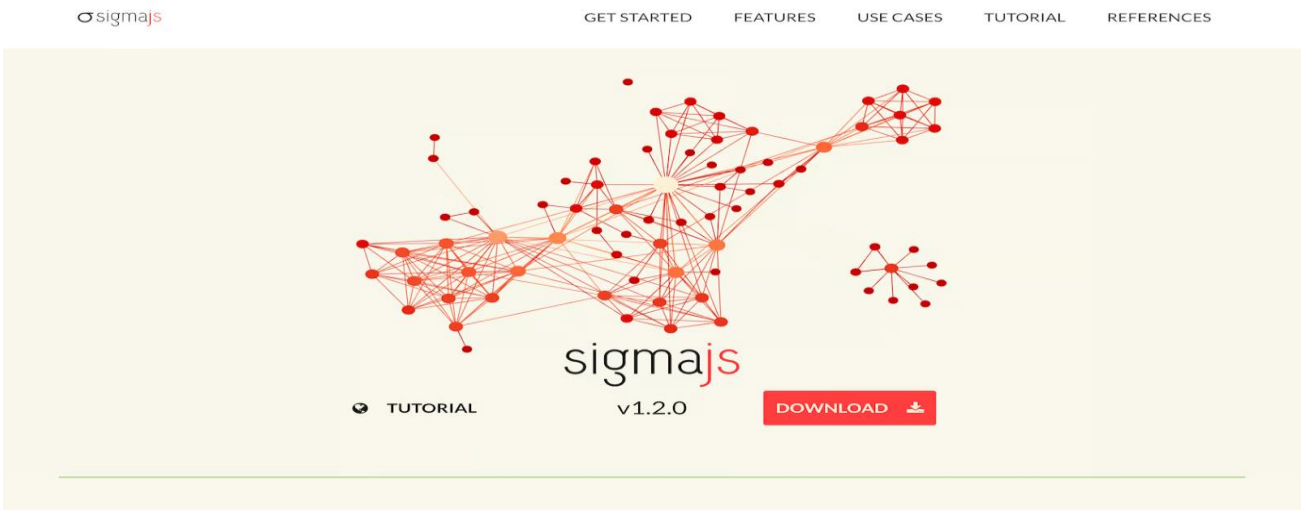

Figure 3: Example of Sigmajs visualization software to represent network of paths and nodes
Source: https://www.toptal.com/designers/data-visualization/data-visualization-tool

2. DATA VISUALIZATION FOR COVID-19

Figure 4 below is a tree diagram showing the number of COVID-19 recoveries by country. The United States and Brazil have the highest number of recovered cases, while the United States also has the highest number of deaths and confirmed cases. However, the number of recovered cases is also highest in the U.S., indicating that the response to the pandemic has been successful.

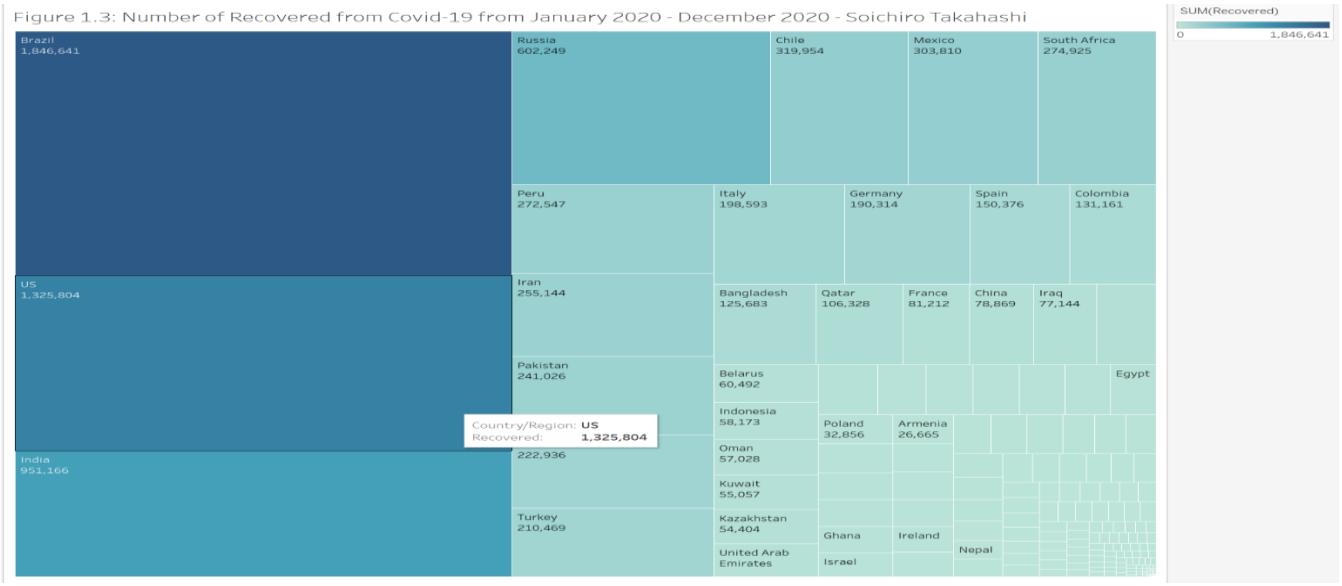

Figure 4: Tree Diagram showing the number of recovered people from COVID-19 by Countries [6]

Figure 5 is a heat map with vertical axis of vaccine manufacturer and horizontal axis for years 2020-2022. This heat map graph analyzes the use of various COVID-19 vaccinations. In the heatmap, changes in the number of vaccines utilized are represented by changes in color intensity. The darkest color, indicating the most utilized vaccine, is that of Pfizer's vaccine.

The highest usage was in the summer of 2021. Other vaccines such as Moderna, Johnson & Johnson, and Oxford also showed high usage. Chronologically, all of the companies' vaccines had low penetration rates when they were first offered, but in January 2021, the penetration rates of Pfizer and Moderna's vaccines increased significantly, followed by the other companies.

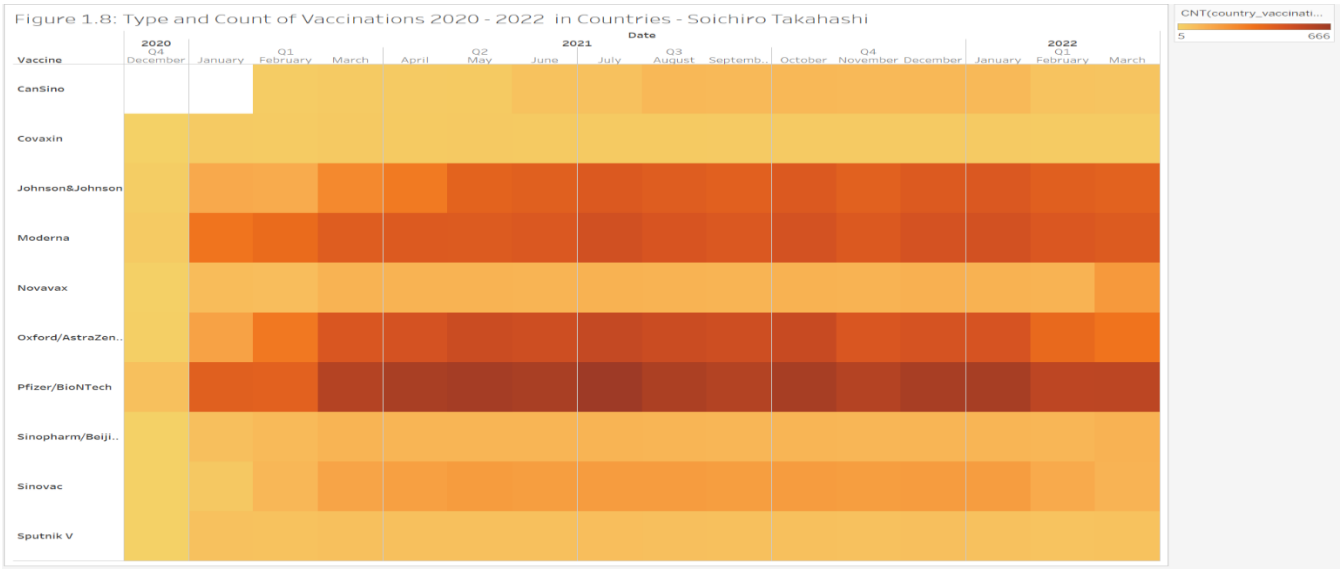

Figure 5: Heat map of types and counts of vaccinations by countries from December 2020 to March 2023 [6]

Figure 6 is a ribbon graph that displays stacked data for multiple categories (U.S. states) with time-series elements and a wave-like representation. The stacked data in this case is the total number of deaths due to COVID-19 from January 2020 to March 2023. The graph is arranged in the order of California, Texas, Florida, Maine, and New York, with the states with the highest values at the top.

The time series analysis reveals that there are still several states with no recorded COVID-19 deaths in 2020, and that these deaths will continue to be reported throughout the United States in 2021, 2022, and beyond.

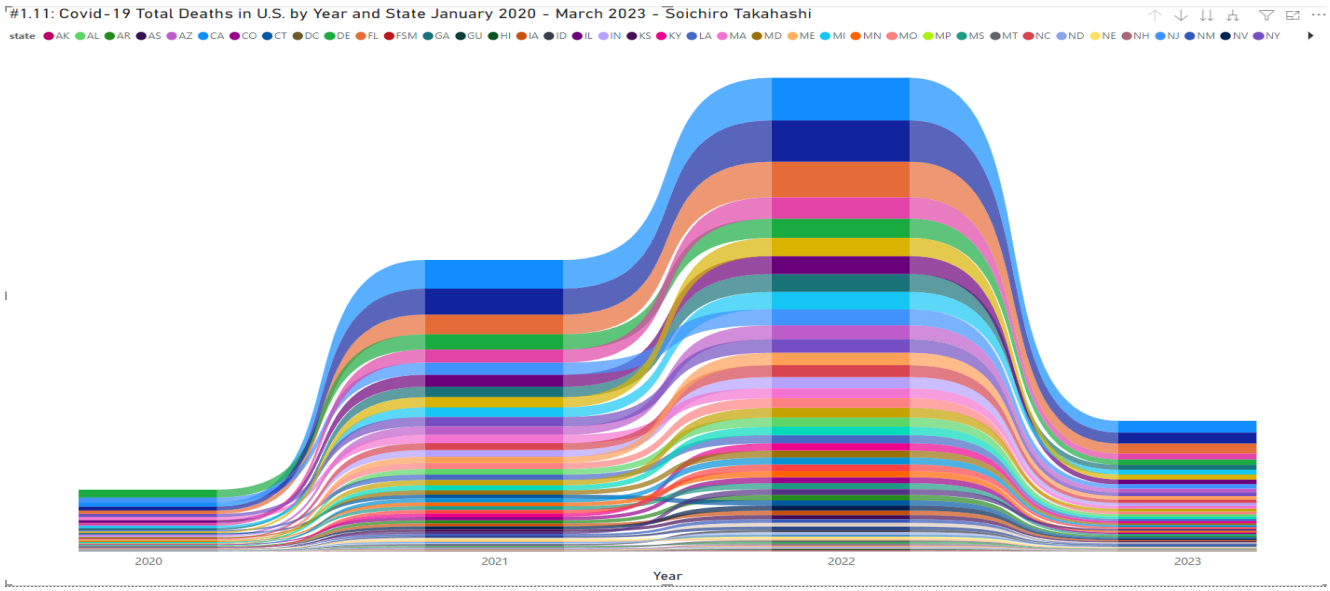

Figure 6:Ribbon graph showing COVID-19 Total Deaths in U.S. by Year and States for January 2020 -March 2023[6]

Figure 7 is called doughnut chart displaying the percentage of COVID-19 cases in each U.S. state from January 2020 to March 2023. California has the largest number of cases at 915M, accounting for 13.51% of the total, followed by Texas at 673M (9.94% of the total) and Florida at 583M (8.61% of the total). Other states have smaller percentages and there are no significant differences among the states.

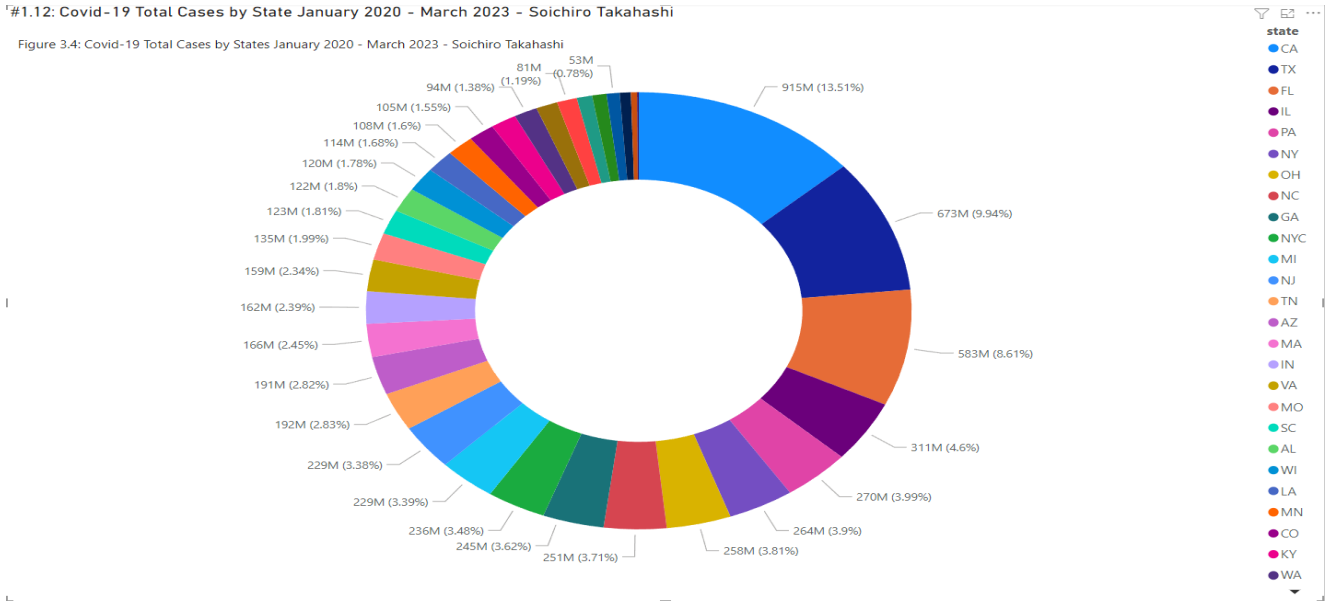

Figure 7: Doughnut Chart showing COVID-19 Total Cases by States for January 2020 - March 2023 [6]

3. DATA VISUALIZATION FOR PNEUMONIA

Figure 8 shows us the percentage of Children with Symptoms of Pneumonia, who went to receive health care from a health care provider during the years of 1990 – 2018. The dataset gathers data from over one-hundred countries.

When you hover your mouse over circles in this visualization created, Tableau will show the percentage of children under five who got taken to a health care provider. When we hover over the squares, it will tell us about the year that child got taken to a healthcare provider. Figure 8 shows us when we hover over each dot with the mouse it tells information about the percentage of children under five that was taken to a health care provider if they had symptoms of pneumonia.

When you hover over the squares in this Tableau visualization created, Tableau will show the year that child was taken. This gives us a lot of information about if a child received health care or not and helps to explain why some countries have higher death rates caused by pneumonia. It also tells us that children who received health care are less likely to die.

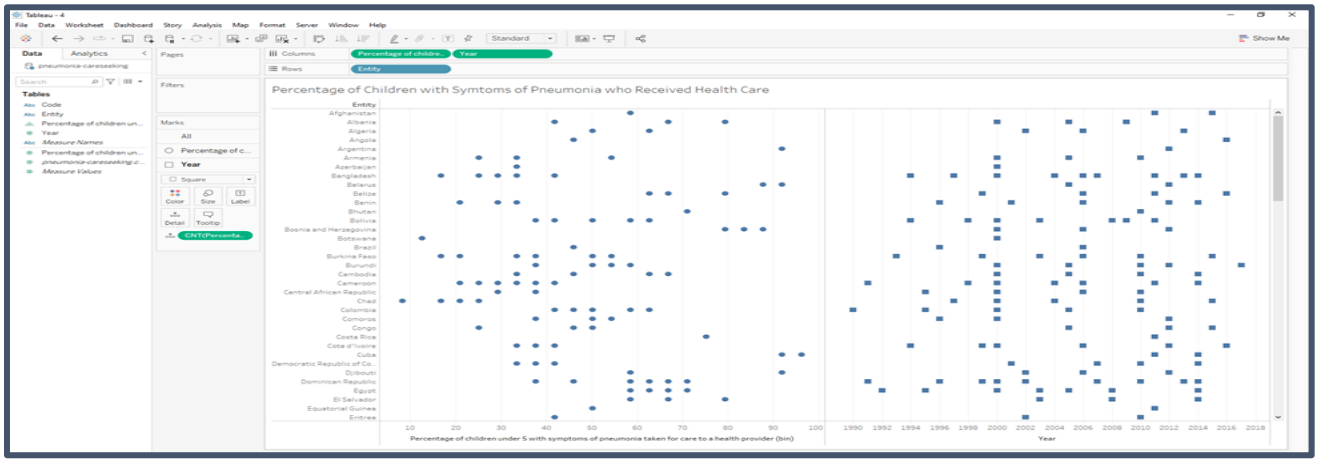

Figure 8: Percentage of Children with Symptoms of Pneumonia who Received Health Care (1990-2018) [7]

Figure 9 shows deaths caused by pneumonia in different five different age groups from the years 1990 to 2019. The age groups used were 5 and under, 6 – 14, 15 – 49, 50 – 69, and 70 or older. Figure 9 gathers data about five different age groups and how many died due to pneumonia. The conclusion we can draw from this data is children under 5 and people belonging to the age groups between 50 and 70 plus, are more likely to die if they contracted pneumonia.

Figure 9: Deaths Caused by Pneumonia in Different Age Groups (1990-2019) [7]

Figure 10 shows Circle Graphs for information about thinness of patients aged anywhere between 1 to 19 years old from the years of 2000 to 2015. When you hover your mouse over each circle, it will show the year that circle represents and how many children between the ages of 1 to 19 have experienced weight loss due to pneumonia. The conclusions we can draw from this Figure 10 is how many children each between the ages of 1 to 19 years old had problems with thinness while having pneumonia.

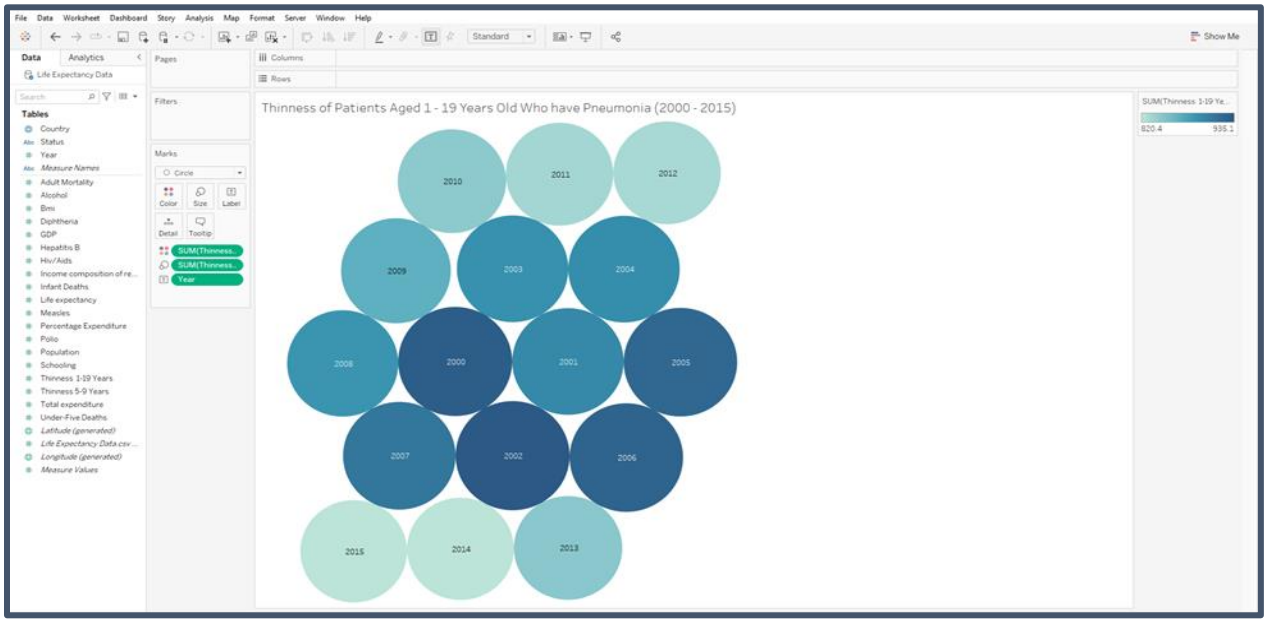

Figure 10: Thinness of Patients Aged 1 - 19 Years Old Who Had Pneumonia (Tableau Visual) [7]

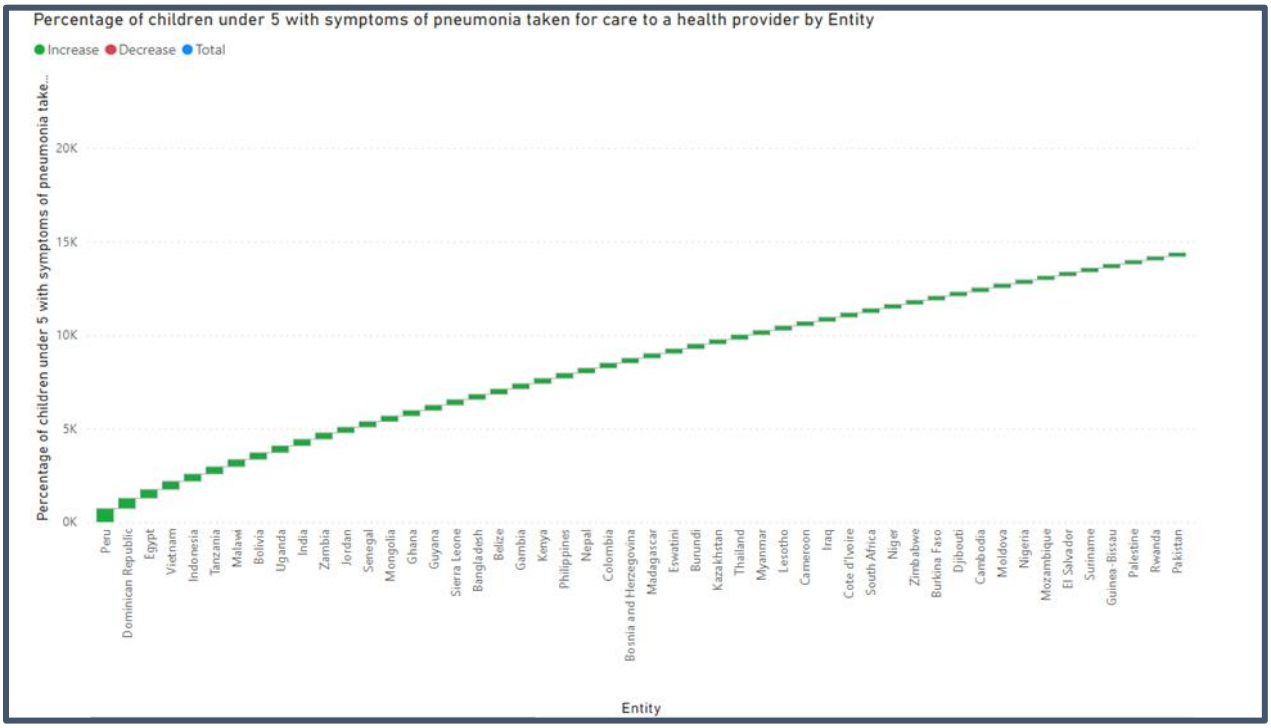

Figure 11: Percentage of Children Under 5 with Symptoms of Pneumonia Who Received Health Care by Country [7] (Power BI Visual)

4. DATA VISUALIZATION FOR FLU

In Figure 12 below for "Flu Vaccine Supply by State 2021" a geographical chart is shown to fully display each state in the U.S. and its supply level of the flu vaccine. Not only can the supply level show individual states concern with the flu, but it can also show how many flu vaccines are being given.

Highly populated states like California and Florida have a high number of vaccine supply, while some also high populated states like Oregon/Washington have negative supply. This could either mean that those two states are using a lot of vaccines, don't carry as many, or are having a hard time supplying their state with vaccines.

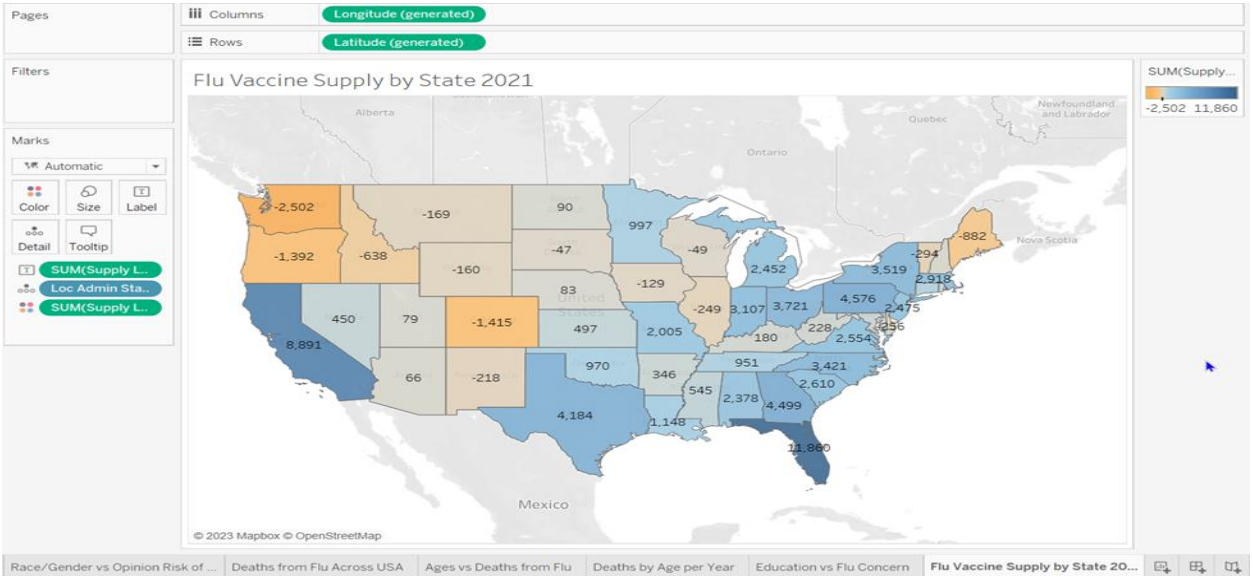

Figure 12: Flu Vaccine Supply for each State of USA in 2021 [8]

Figure 13 below shows Sum of All Deaths by City shows the number of deaths, larger or smaller circles, in different populated cities across the globe. As we can see from this Figure 13, the east coast in the United States is heavily affected by the flu virus and see a multitude of deaths occurring around the east coast of the United States.

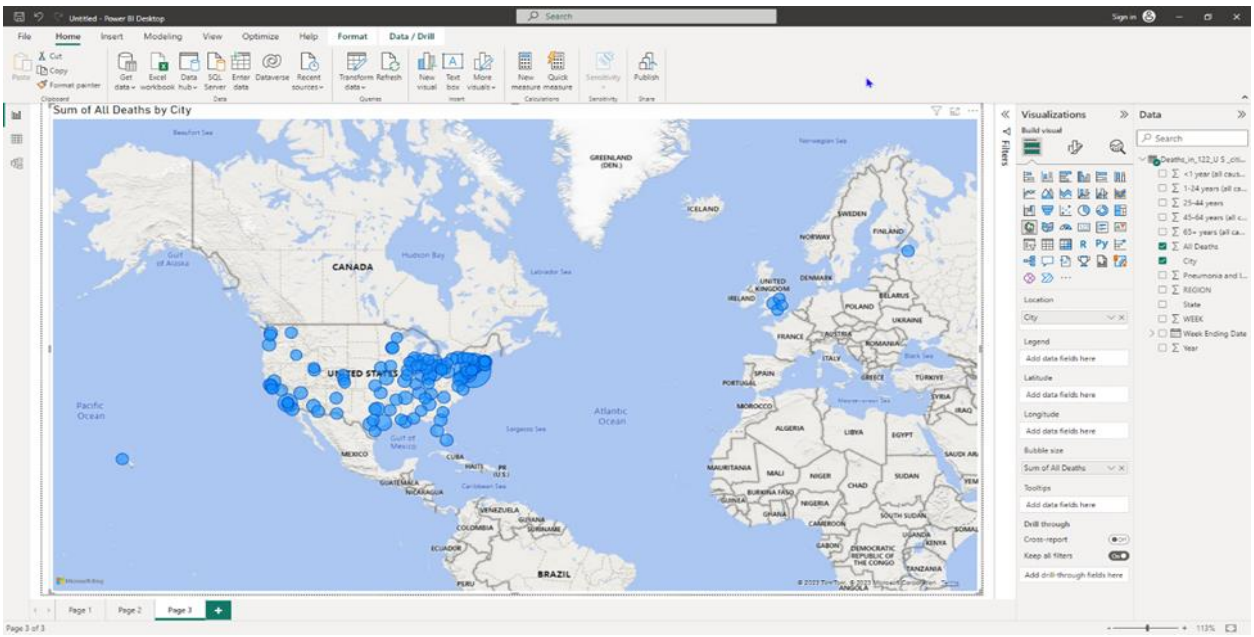

Figure 13: Sum of All Deaths due to Flu by City [8]

From area chart shown in Figure 14, we can see the flu knowledge, concern, and health insurance that different age groups have. By combining all of this, we can tell which group of individuals have a concern with the virus and whether they are invested in health insurance or not.

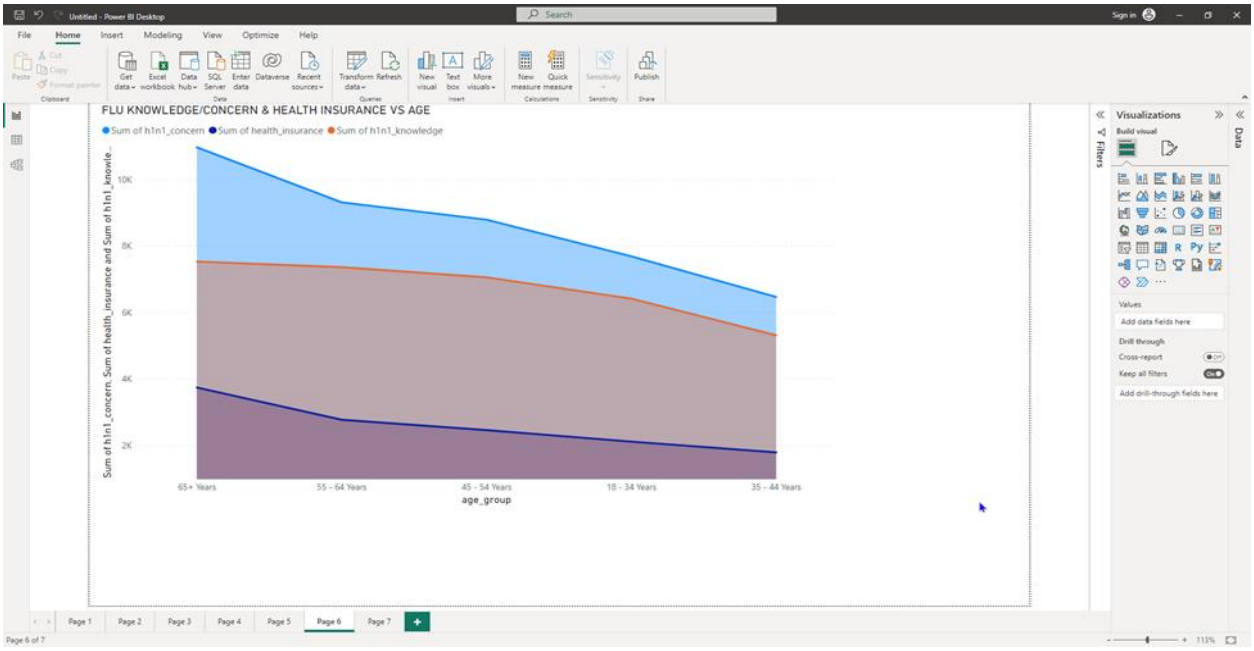

Figure 14: Area Chart showing Flu Knowledge/Concern & Health Insurance vs. Age [8]

From Figure 15 below, we can compare different levels of education/employment and those individuals' opinions of effectiveness that the vaccine for the flu has. From Figure 15 we can see that employed/highly educated individuals tend to believe that the vaccine is effective and vice versa.

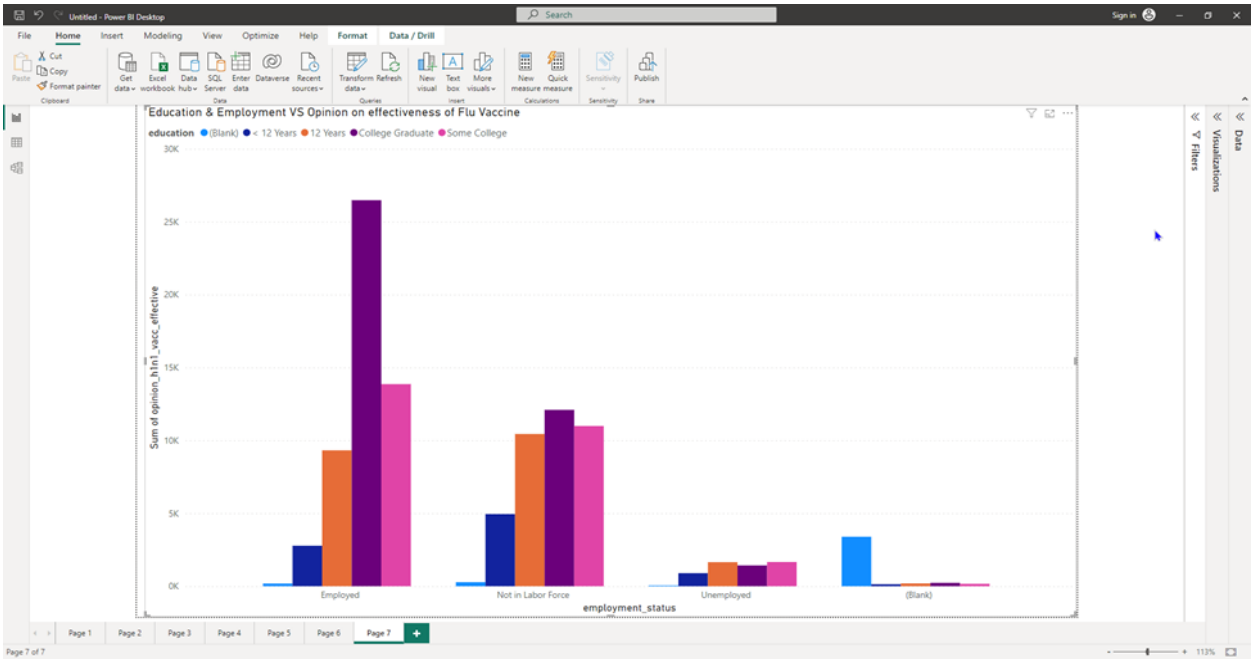

Figure 15: Education & Employment vs Opinion of Vaccine Effectiveness [8]

5. CONCLUSIONS

5.1 COVID-19 [6]

1. The United States had the highest number of deaths from COVID-19 in the early stages of the pandemic compared to other countries, but also had the highest number of recoveries

2. Vaccines played an important role in reducing the severity of COVID-19, as shown by the marked decrease in deaths that coincided with increased vaccination.

3. A comprehensive analysis using different types of graphs (scatter plots, ribbon graphs, donut graphs) and data from 2020 to 2023 provided a more detailed understanding of the impact of COVID-19 in the US and worldwide.

5.2 Pneumonia [7]

1. The biggest factors of pneumonia survival rate are that a nation's GDP, wealth, and the status of whether a nation is developing or developed is what affects the survival rate of people worldwide.

2. Overall is that people who smoked versus those just exposed to secondhand smoke had a higher rate of death from pneumonia than the other group of people did.

3. Those who received vaccination from pneumonia versus those who did not have the vaccine, had a much higher survival rate than the other group who were not fully vaccinated.

5.3 Flu [8]

1. The Flu is massively present in the United States, specifically within highly populated environments such as the east coast, New York, and other areas as well.

2. The virus seems to be more present with individuals who have not been vaccinated, which from the data, are more commonly the individuals with less education.

3. Fatalities with this disease, statistically, are more evident in the elderly and newborns. It is crucial that these two age groups do not receive this virus, or there is a heavy potential in fatal symptoms.

6. ACKNOWLEDGEMENTS

(1.) Dr. Richard Segall needs to acknowledge the support for continuation of this research by a Seed Money Grant awarded by the Arkansas Biosciences Institute (ABI) and (2.) Soichiro Takahashi needs to acknowledge support for continuation of this research by a subsequent 2023 Summer Internship awarded by Arkansas Biosciences Institute (ABI).

The division of labor also needs to be acknowledged in this research in that Soichiro Takahashi performed the research on COVID-19, Tucker Pierce on Pneumonia, and Joshua Curry for Flu that were also performed as part of course term projects for ISBA 3423 Data Visualization for Business in Neil Griffin College of Business at Arkansas State University under supervision of Professor Richard S. Segall.

7. REFERENCES

[1] **Tableau Software.**
https://en.wikipedia.org/wiki/Tableau_Software

[2] **Microsoft Power BI.**
https://en.wikipedia.org/wiki/Microsoft_Power_BI

[3] IBM (2024). **What is Data Visualization?** Retrieved May 13, 2024 from https://www.ibm.com/topics/data-visualization

[4] Simplilearn (2024). **23 Best Data Visualization Tools for 2024.** Retrieved May 13, 2024 from https://www.simplilearn.com/data-visualization-tools-article

[5] **Data Visualization**. Retrieved May 13, 2024 from https://static.javatpoint.com/tutorial/tableau/images/data-visualization.png

[6] Takahashi, S. (2023). **Semester Project: Data Visualization for COVID-19.** ISBA 3423 Data Visualization for Business, Neil Griffin College of Business, Arkansas State University, USA.

[7] Pierce, T. (2023). **Semester Project: Data Visualization for Pneumonia**. ISBA 3423 Data Visualization for Business, Neil Griffin College of Business, Arkansas State University, USA.

[8] Curry, J. (2023). **Semester Project: Data Visualization for Flu.** ISBA 3423 Data Visualization for Business, Neil Griffin College of Business, Arkansas State University, USA.