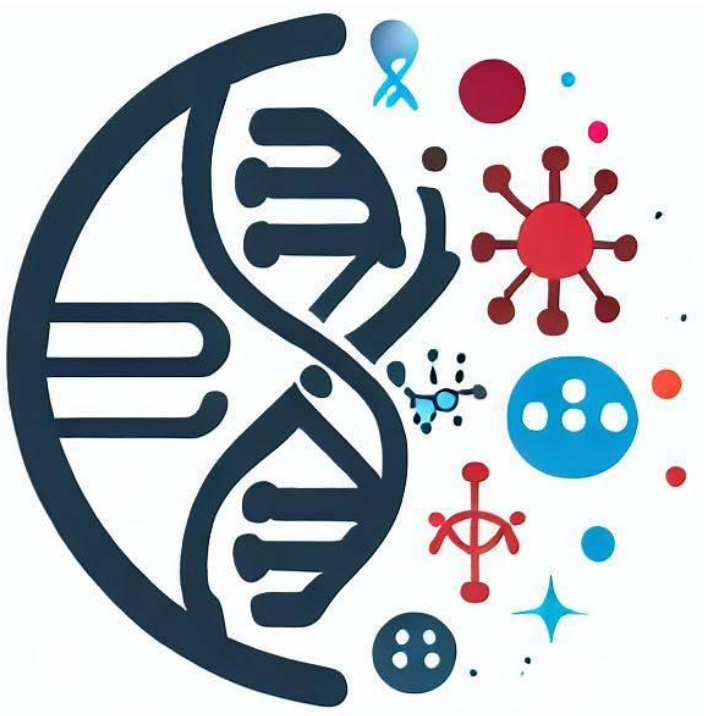
Project Report

Comorbidity and Genetic Factors and their Impacts on Patients with COVID-19



Final Version – 05/04/2023

Team Members:

Riese Atianzar
Rohan Chatterjee
Francisco Contreras
Emily Gonzalez
Luis Gonzalez
Ting Fung Ha
Carlos Hernandez
Juan Hernandez
Jimuel Julaton
Chen-Ching Lin

Faculty Advisor:

Dr. Navid Amini

Project Liaisons:

Haley Kirk

Table of Content

1. Introduction:

1.1 Background:

Coronavirus disease 2019 (COVID-19) has had a profound impact on the world, affecting health, economies, and daily life. This highly infectious respiratory illness, caused by the SARS-CoV-2 virus, spreads through infected persons who exhale droplets and small particles containing the virus. These particles can be inhaled by others or deposited on their eyes, noses, or mouth, and in some circumstances can contaminate surfaces they touch[1]. In severe cases, COVID-19 can even lead to death, particularly in individuals with underlying health conditions or those who have weak immune systems. As of March 15, 2023, there have been more than seven hundred and sixty million confirmed cases of COVID-19 resulting in more than almost 7 million confirmed deaths globally. Additionally, COVID-19 has led to a decrease in the United States' GDP of 32.9% within the first year of the pandemic resulting in the largest economic downturn since the Great Recession [2], [3] resulting in long term ramifications for the economy, job market, people's daily lives and mental health.

Given the devastating impact of the virus, COVID-19 has become one of the areas of study in the world. However, despite the extensive research that has been conducted on the virus, there are several overlooked aspects related to the disease that have been ignored by a large section of the scientific community in favor of more known informative and logical areas related to prevention, vaccination, and cure. However, to have a complete understanding of this virus it is important to all possible areas related to the disease.

This project seeks to address this gap in research and explore some of the neglected areas of COVID-19. This project focuses on three core branches that analyze the spread and mortality of COVID-19. The three branches can be categorized into two parts: visualization of already available data and the use of Machine Learning to identify new data. By employing these two approaches, this project seeks to address the neglected areas of COVID-19 and provide insights into the spread of the disease and mortality.

The first branch of the project is a creation of an online Tableau Public portal. This portal will provide easy access to a large number of visualizations that display specific aspects related to COVID-19, including virus cases, vaccinations, mortality, and their relation to other factors. The Tableau software is being used to design, experiment and iterate through with different visualizations and elements, making it a straightforward tool to use. However, there are significant drawbacks to this software

when it comes to efficiency, data usage, interactivity, accessibility, and customization of the portal itself.

Therefore, the second branch of the project is a JavaScript Data Visualization Application which circumvents the limitations of the Tableau portal. This application lets the user easily navigate through and interact with remade versions of visualizations from the tableau portal but now with live data. The application also is also faster and easier to use as it is being made from ground up for the purposes of visualizing COVID-19 data. This application can easily be accessed by anyone online and be available for everyone to download.

The first two branches of the project center around the analysis of easily available COVID-19 data and is a part of the Visualization component however the third branch takes a deeper look into covid by analyzing patients and comorbidities related to COVID-19 severity and mortality. More specifically this project aims to identify possible associations between COVID-19 severity and blood type which is a neglected factor in research. While there has been some research into the relationship between blood type and COVID-19, this area of study has not been thoroughly explored in favor of a more concrete area of study. By examining private clinical data to identify comorbidities that may have an effect on COVID-19 patients related to their blood type, this branch hopes to shed light on the potential role of blood type in determining the outcome of COVID-19 patients.

Blood type has been linked to a number of health outcomes, including susceptibility to certain diseases and response to treatment. There have also been multiple studies that have shown blood type B and AB are more likely to test positive for COVID-19. Some studies also suggest that blood type may influence infection through the inhibitory effects of anti-A antibodies in individuals with blood type O or B.

However, its association with COVID-19 severity and mortality still remains unclear. This project will analyze data from COVID-19 patients and compare their blood types to the severity of their illness, taking into account other factors that may contribute to COVID-19 severity. By identifying any potential associations between blood type and COVID-19 severity, this project could contribute to our understanding of the disease and potentially inform future treatment strategies.

Overall, this project represents an important step in addressing the overlooked aspects of COVID-19 research and identifying potential factors that may contribute to the severity of the disease. The visualizations will help provide new information that will be simple to understand, the application will allow easy access, navigation of the visualizations while also keeping them up to date, finally the machine learning analysis

will help provide a answer to see if there is any associations between blood type the effects of COVID-19 in relation to severity and mortality.

1.2 Design Principles:

Our goal for this project is to gain a better understanding of COVID-19 by studying aspects that are overlooked or unconsidered by a large part of the scientific community in favor of more important areas of study. The application section of the project is built with JavaScript and a number of JavaScript libraries. With the help of these libraries, our application is able to present all the critical data in a simplistic, yet aesthetic manner. Datasets will be retrieved from various trustworthy and credible source providers, where the JavaScript libraries will convert a dataset into visual representations for the application users to interact and analyze. After the conversion, the user will have the opportunity to interact and customize their individual viewing options based on their own desires. This section is part of the visualization component of our project while the other section uses machine learning to study comorbidity and genetic factors and their impacts on patients with COVID-19. In the end, Boruta, a feature selection algorithm and multiple machine learning models are used to determine if there are any factors associated with COVID-19 mortality and severity.

1.3 Design Benefits:

By following the design principle above, our team was able to gain a number of benefits including reusability, compatibility and ease of access throughout the pages of the JavaScript data visualization application. With the support of a selection of JavaScript libraries, new pages can use the layout and design from the previous page, rather than implementing a new structure. On the other hand, the compatibility between each library allows all the implementing functions to work properly. This also allows users to visit our application with other different devices besides computers. In the end, the implementation of a navigation bar can help the user gain access to some of our frequently used functions quickly.

1.4 Achievements:

The team was able to create visualizations that give us new insights to COVID-19 through the use of Tableau desktop and then transferred to Tableau online

for our public portal. These visualizations were then transferred to the JavaScript data visualization application where the datasets used can be shown through visualizations in real time. The application contains filters for the visualizations to alter them such as changing the view to only see the data from a certain date range. The application was optimized for a smooth viewing experience and allows for the visualizations to load efficiently.

The machine learning section of the project combines with a feature selection algorithm, known as Boruta algorithm. After feeding the algorithm with confidential medical data and training it, the algorithm is able to generate results that may act as a key factor in COVID-19 prognosis.

2. Related Technologies:

2.1 Existing Solutions:

An application created by a group of doctorate students from John Hopkins University shares many similarities with our project. However, the application from John Hopkins University works more like a tracker for COVID-19 cases, where our application offers an interactive and customized window for users to see and manipulate the data base on user's desires.

2.2 Reused Products:

Our team was able to reuse a great number of our code based on previously implemented contents. We consistently use the D3 library for creating a visualization, along with its interactions. Secondly, because the team reused the style and design from Material UI, the layout of most pages are almost identical. In the end, no matter which page the user accesses, the navigation bar will always remain the same. For that reason, the component of the navigation bar remains unchanged.

3. System Architectures:

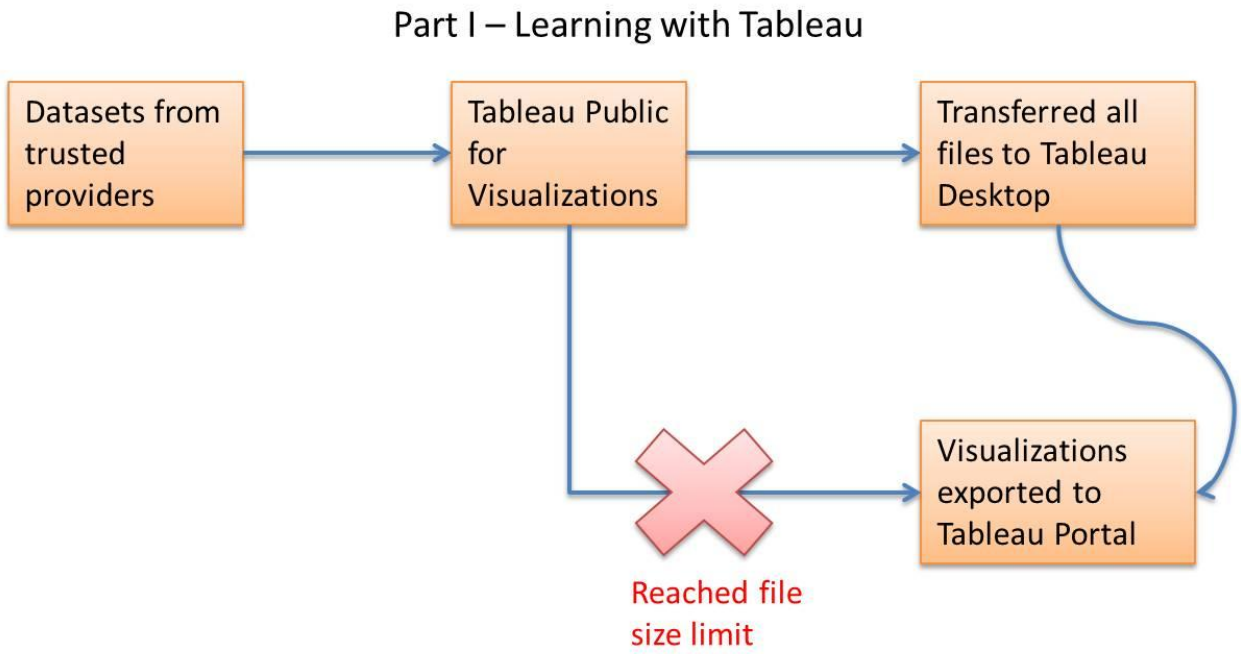
3.1 Overview:

3.1.1 Tableau Portal

Data Source Provider: This section is the beginning of our project. Online datasets from credible source providers will be taken as input and fed into the application perspective, which is covered in the next section.

JavaScript Designs and Layouts: The application will provide a simplistic and aesthetic layout to host all the information and all the visualizations generated from online datasets. Each dataset and its relevant visualizations will be grouped into an organized fashion.

Dataset Visualization: All graphical representations will be processed and converted into an interactive form. These visualizations will be hosted into the design and layouts defined in the previous section.

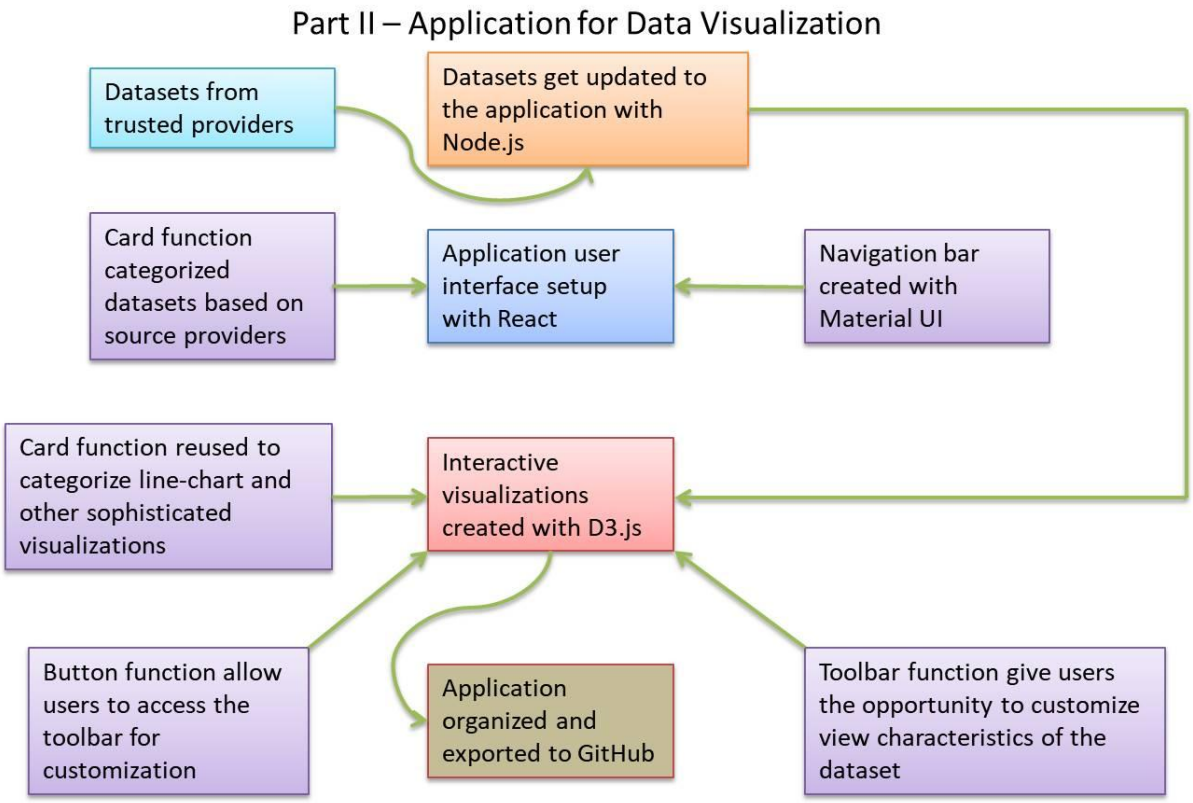


3.1.2 Application

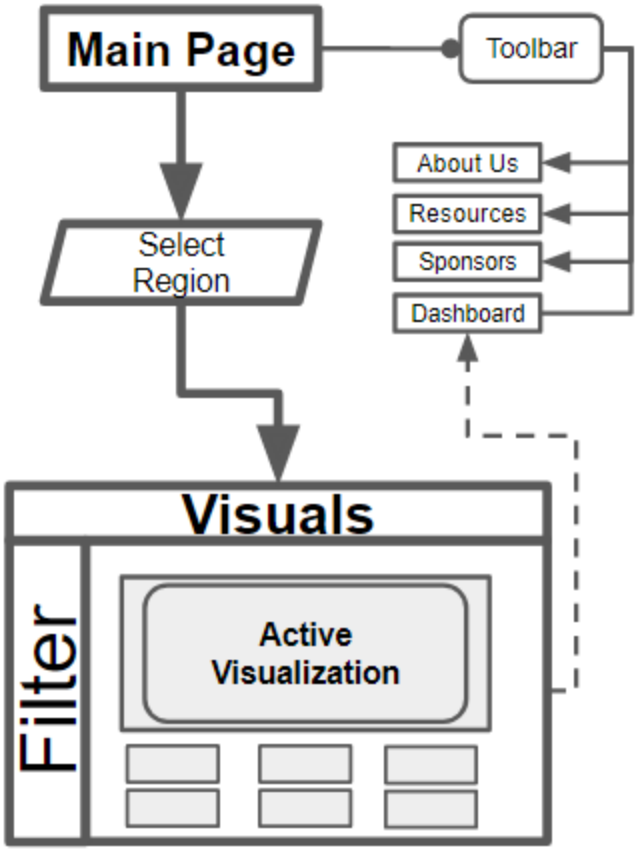
Data Source Provider: This section is the beginning of our project. Online datasets from credible source providers will be taken as input and fed into the next section, the application perspective.

JavaScript Designs and Layouts: The application will provide a simplistic and aesthetic layout to host all the information and all the visualizations generated from online datasets. Each dataset and its relevant visualizations will be grouped into an organized fashion.

Dataset Visualization: All graphical representations will be processed and converted into an interactive form. These visualizations will be hosted into the design and layouts defined in the previous section.



3.2 Data Flow:



Main page: The main page contains the toolbar in the top for navigation purposes, while giving the user options to select datasets.

Toolbar: The toolbar, or navigation bar, will act as a shortcut with four other subcategories, allowing the user to quickly navigate to one of the pages. The toolbar will remain in the top for all the other pages.

Select Region: This section gives users the option to select a dataset that focuses domestically or internationally.

Visuals: This section will house a number of subsections like active visualization and filters.

Active Visualization: The visuals section will display a quick visualization which summarizes the selected dataset.

Filter: The filters will provide the users with predefined options to change the viewing windows and options.

3.3 Implementation:

The entire project was broken down into four groups, obtaining datasets, implementing designs and layout for the application, creating visualizations, and machine learning.

3.3.1 Obtaining Datasets

In order to retrieve quality and reliable data, the providers must have a certain level of credibility. At the same time, the provider should update its dataset regularly.

3.3.2 Design and Layout Implementation

The designs and layouts were designed with JavaScript and two of its libraries. This implementation can be found on many other contemporary webpages on the Internet. These work will set the foundation for our application as a whole.

3.3.3 Creating Visualizations

The datasets obtained from the previous section will be handled here, by a JavaScript library known as D3. This library will convert the dataset into interactive and dynamic visualizations. Afterwards, some functions like filter options and saving screenshots to the dashboard were added by the team to enhance quality and variations.

3.3.5 Clinical Analysis with Machine Learning

This is the second portion of the project, where a machine is being trained with a feature selection algorithm known as Boruta. After numbers of training iterations, the machine will generate some categories that the machine considered important for COVID-19 prognosis.

The second half of the project involves training machine models with a feature selection algorithm known as Boruta. After numbers of training iterations, the machine will determine if blood type is a determining factor in severity and mortality of COVID-19.

4. Conclusions:

4.1 Results

In the course of a school year, our team researched, designed and implemented an online application that accepts datasets as input, and returns a quantity of interactive visualizations expressed in different perspectives as output. Additionally, the team created other side functions like filtering options and saving the visualizations to the dashboard for later.

The application is not just limited for PC users, but rather for a wide range of mobile users across some of the popular platforms today.

4.2 Future Works:

- Convert the application into tracker with visualizations not just for COVID-19, but for other infections
- Implement more filter options and visualizations for datasets
- Health watch system with reminders, local and international health informations

5. References:

1. Our World in Data, "COVID-19 Data," https://github.com/owid/covid-19-data/blob/master/public/data/README.md, accessed March 8, 2023.
2. World Bank, "Urban population (% of total population)," https://data.worldbank.org/indicator/SP.URB.TOTL, accessed March 8, 2023.
3. Fraser Institute, "Economic Freedom of the World 2023 Annual Report," https://www.fraserinstitute.org/resource-file?nid=14426&fid=17063, accessed March 8, 2023.
4. Our World in Data, "Share of individuals using the internet," https://ourworldindata.org/grapher/share-of-individuals-using-the-internet, accessed March 8, 2023.
5. Helliwell, J. F., Layard, R., & Sachs, J. D. (2022). World Happiness Report 2022. New York: Sustainable Development Solutions Network.
6. World Bank, "GDP (current US$)," https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?end=2021&start=2021, accessed March 8, 2023.
7. “How Coronavirus Spreads,” Cent. Dis. Control Prev. CDC, Aug. 2022, [Online]. Available:https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covid-spreads.html.
8. “How has COVID-19 impacted the US economy?,” World Economic Forum, Jul. 31, 2020. [Online]. Available: https://www.weforum.org/agenda/2020/07/covid-19-coronavirus-usa-united-states-economy-gdp-decline/
9. COVID-19: U.S. Economic Effects. in Congress. 2020. [Online]. Available: https://crsreports.congress.gov/product/pdf/IN/IN11388
10. S. L. Franconeri, L. M. Padilla, P. Shah, J. M. Zacks, and J. Hullman, “The Science of Visual Data Communication: What Works,” Psychol. Sci. Public Interest, vol. 22, no. 3, pp. 110–161, Dec. 2021, doi: 10.1177/15291006211051956.
11. J. J. Van Wijk and H. Van de Wetering, “Cushion treemaps: visualization of hierarchical information,” in Proceedings 1999 IEEE Symposium on Information Visualization (InfoVis’99), San Francisco, CA, USA: IEEE Comput. Soc, 1999, pp. 73-78,. doi: 10.1109/INFVIS.1999.801860.
12. C. A. Latz et al., “Blood type and outcomes in patients with COVID-19,” Ann. Hematol., vol. 99, no. 9, pp. 2113–2118, Sep. 2020, doi: 10.1007/s00277-020-04169-1.
13. Y. Zhang, R. Garner, S. Salehi, M. La Rocca, and D. Duncan, “Association between ABO blood types and coronavirus disease 2019 (COVID-19), genetic associations, and underlying molecular mechanisms: a literature review of 23 studies,” Ann. Hematol., vol. 100, no. 5, pp. 1123–1132, May 2021, doi: 10.1007/s00277-021-04489-w.
14. Rundel, C. (2013, April 15). Lecture 20 - Logistic Regression. Duke Statistical Science. Retrieved Available