COVID-19 Data Analysis and Visualization Application

Name – Kavya Gandhi (0778608)

Application URL - https://7v67yd-kavya-gandhi.shinyapps.io/covidapp/

Background

World economies, public health, and daily life are all affected by the COVID-19 pandemic. It can be overwhelming for researchers and decision-makers to keep up with the massive amount of data being generated and reported daily on COVID-19 cases, deaths, and recoveries. It is therefore important to have tools that can effectively analyze and visualize this data to identify patterns, trends, and insights. With a focus on time series and bar charts, this document describes a Shiny application that analyzes, summarizes, and visualizes COVID-19 data.

The analysis and visualization of data have been shown to improve the effectiveness of public health interventions during pandemics such as COVID-19 (Dong et al., 2020; Nair et al., 2020[2]; Schmitz et al., 2020). Based on Schmitz et al.'s (2020)[3] analysis, policymakers were able to identify regional and temporal differences in the spread of COVID-19 by visualizing the data. According to Dong et al. (2020)[1] , visualizing the spread of COVID-19 using spatial and temporal data allowed public health interventions to be targeted at high-risk areas.

Purpose of the Application

We designed this application to provide an easy-to-use tool for analyzing and visualizing COVID-19 data. With this dynamic application, users can load data from a CSV file, clean it, and create time series and bar charts. To explore data further, the application also provides summary statistics and a data table view.

Despite its simplicity, the application is designed to be easy to use even by people with little or no experience in data analysis or visualization. Each function is clearly labeled and explained with an intuitive user interface. As you load, clean, explore, and visualize data, the sidebar panel shows you all the options, while the main panel displays a table, summary statistics, and plots of the data. Furthermore, this application provides a clear understanding of the data without requiring deep analysis. Every day, large amounts of datasets are generated for covid-19 death cases, confirmed cases, and recovery cases. The application simplifies the task of data analysis, allowing effective measures to be taken to prevent this disease through effective analysis of the data.

The purpose of this application is to provide a simple and accessible tool for the analysis and visualization of COVID-19 data with the ultimate goal of supporting evidence-based decision-making and public health interventions.

Key Features

Several key features make the COVID-19 Data Analysis and Visualization application an effective tool for analyzing and visualizing COVID-19 data. These features include:

Data Loading options - A data loading option is available in the application that allows users to load data from a CSV file**.** With this feature, users can access the latest COVID-19 data immediately.

Data Cleaning Options - Users can remove missing values and other anomalies from data using the application's data cleaning feature. Analyzing and visualizing data with this feature ensures their accuracy and reliability.

Statistical Exploration - Statistical analysis of the loaded and cleaned data is available for users through the application**.** Users are able to identify trends and patterns in the data quickly using this feature.

Data Visualization Options - Users can visualize data in two ways: time-series plots and bar charts. A time-series plot depicts the change in data over time, while a bar chart shows the distribution of data across categories.

Problems Addressed by the application

For researchers and policymakers, the COVID-19 pandemic has presented many challenges, but thanks to COVID-19 data, they have been able to track and understand the virus' spread, analyze the effectiveness of control measures, and develop evidence-based recommendations for dealing with the disease. There are, however, a number of problems associated with analyzing and visualizing COVID-19 data.

The purpose of this application is to provide a convenient means of analyzing and visualizing COVID-19 data. As a result of the application, the following problems are addressed:

A challenge is navigating and analyzing COVID-19 data due to a large amount of data available. This problem is addressed by the COVID-19 Data Analysis and Visualization application, which provides users with an easy-to-use platform for loading, cleaning, and exploring COVID-19 data. It is also difficult to compare and analyze COVID-19 data effectively because the data quality varies significantly across countries and regions. This problem can be solved by using the COVID-19 Data Analysis and Visualization application, which provides users with tools for cleaning and analyzing COVID-19 data.

Additionally, many users may not possess the technical skills or knowledge required for effective analysis and visualization of COVID-19 data. In order to address this problem, COVID-19 Data Analysis and Visualization offers a user-friendly interface without requiring any technical knowledge or skills.

Thus, the application presented can help address problems across many sectors where data analysis is complex and challenging, and collecting data from multiple sources can lead to inaccurate and meaningless results. This application can be particularly useful in large-scale industries such as healthcare and finance.

For example, in the healthcare industry, doctors can collect data from patients regarding their health conditions and blood reports, and use this application to clean and analyze the data. Based on the analysis, doctors can predict patients' health conditions and make informed decisions about their treatment plans.

Additionally, individuals with limited knowledge of technology can also benefit from this application. They can use it to clean and remove missing data, as well as visualize the data in a way that is easy to understand. This can help individuals make better decisions based on the data they collect. Overall, the Shiny application has the potential to revolutionize data analysis and visualization across various sectors and industries.

Literature Review

Numerous studies have analyzed COVID-19 data to find insights and trends within data. During the early stages of the pandemic, Ghosal et al. (2020)[4] examined COVID-19 data from a number of countries, finding that cases and deaths increased rapidly. Time series analysis was also used by Abdollahpour-Alitappeh et al. (2021)[5] to examine the effects of different interventions on COVID-19 transmission. Distancing from others and wearing masks were effective in reducing virus spread, according to the study. The importance of data analysis and visualization in understanding the impact of COVID-19 is highlighted by these studies.

COVID-19 data analysis and visualization have been the subject of several research studies. As a result of these studies, it has become clear how crucial data analysis and visualization are to understand how the virus spreads, its impact on different populations, and the effectiveness of interventions to control the outbreak.

An analysis of COVID-19 data from different countries was conducted by Abdellaoui et al. (2020)[6] in order to identify the factors influencing the spread of the virus. To determine the correlation between population density, age distribution, and the spread of the virus, data analysis and visualization techniques were used.

Huang et al. (2020)[7] used data analysis and visualization to determine which factors contribute to the severity of COVID-19 in different populations. A study of COVID-19 data from several countries found significant influence on the severity of the disease from age, gender, and preexisting conditions.

The importance of data analysis and visualization in understanding COVID-19's spread and impact is highlighted in these studies. COVID-19 Data Analysis and Visualization provide users with an easy-to-use tool for analyzing and visualizing COVID-19 data.

Instructions for Operation

To use the COVID-19 Data Analysis and Visualization Application, follow these steps:

Platform-Independent Executable Link of application: https://7v67yd-kavya-gandhi.shinyapps.io/covidapp/

Step 1: Click on the above link to analyze the covid-19 data.
Step 2: To load your COVID-19 data, click on the "Choose CSV File" button in the sidebar panel, and select your CSV file. Note that the CSV file must contain a column with date values and columns for confirmed cases, deaths, and recoveries.
To make it easy, below are some links to the dataset to upload in the application:
  a)   https://www.kaggle.com/datasets/imdevskp/corona-virus-report
  b)   https://www.kaggle.com/datasets/meirnizri/covid19-dataset
  c)   https://www.kaggle.com/datasets/sudalairajkumar/novel-corona-virus-2019-dataset
Step 3: Click on the "Load Data" button to load your data into the application.
Step 4: To clean your data, click on the "Clean Data" button. This will remove any rows with missing or erroneous data. This cleaned data will be used for all subsequent analyses and visualizations.

Step 5: To view summary statistics of the cleaned data, click the "View Summary Statistics" button. The summary statistics will be displayed in the main panel of the application.

Step 6: To create a plot, choose a plot type (either "Time Series" or "Bar Chart") from the drop-down menu, choose an X variable and Y variable from the respective drop-down menus, and then click the "Create Plot" button. The plot will be displayed in the main panel of the application.

Step 7: To view the original loaded data, the data table will be displayed in the main panel of the application.

About Application

In the main panel, the app displays the data table, summary statistics, and data visualization. The app uses the ggplot2, dplyr, tidyr, and readr packages to perform data analysis and visualization. The server function defines eventReactive functions for loading and cleaning data, creating select inputs for x and y variables, generating plots based on the selected plot type and variables, and creating tables for summary statistics and data. It utilizes various functions such as read.csv, na.omit, eventReactive, renderUI, renderPlot, renderPrint, and renderDataTable.

The application allows users to interactively explore and visualize COVID-19 data, making it easier to gain insights and understand the trends and patterns of the pandemic. This Shiny web application allows users to load, clean, explore and visualize COVID-19 data. It consists of a user interface and a server function and utilizes various packages such as ggplot2, dplyr, tidyr, and readr.

Prototype

The user interface includes a title panel, sidebar panel, and main panel. Sidebar Panel is shown below:
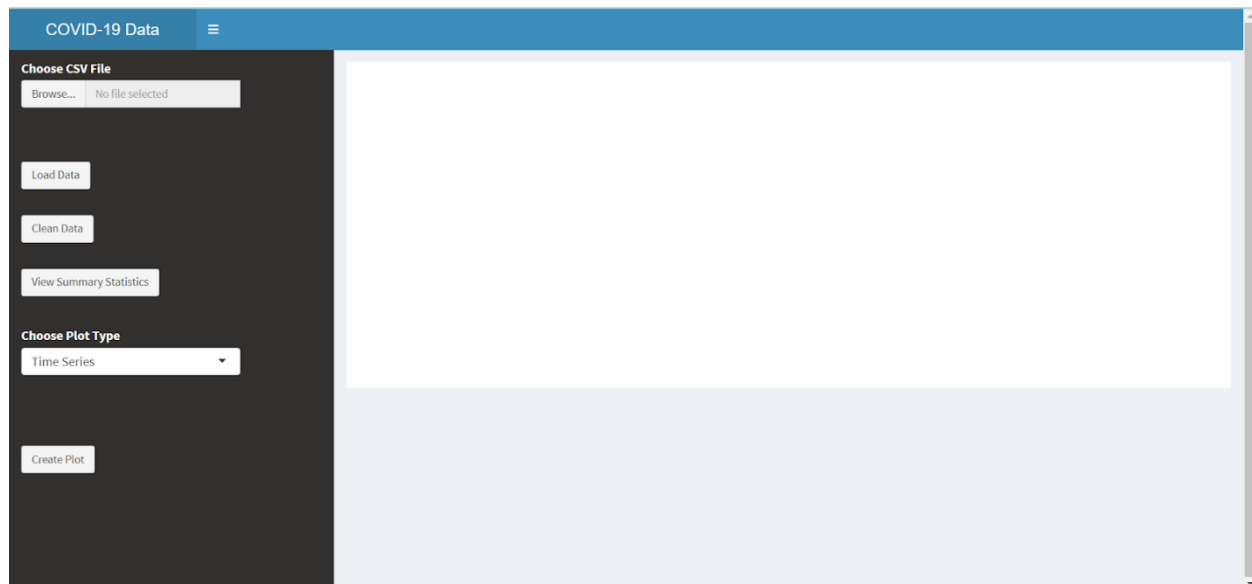
The sidebar panel contains options for loading and cleaning data, exploring summary statistics, selecting a plot type, choosing x and y variables for the plot, and creating the plot. Also in the sidebar panel, users can choose to load a CSV file, clean the loaded data, view summary statistics,

and create data visualizations. The sidebar panel also includes inputs for choosing plot type, X variable, and Y variable.

 The main panel shows the data table, summary, and charts as shown below:

The "Load Data" option allows users to choose a CSV file containing COVID-19 data, which is read into the application using the read.csv function. The data is stored as a reactive object using the eventReactive function, which means that it is only loaded once the user clicks the "Load Data" button. The "Data Table" option displays the loaded data in a table using the renderDataTable function. The table is only displayed once the user clicks the "Load Data" button.

The "Clean Data" option allows users to clean the data by removing any rows containing missing values. The cleaned data is stored as a reactive object using the eventReactive function, which means that it is only cleaned once the user clicks the "Clean Data" button.

By just clicking on View Summary, it shows a statistical summary of the data shown below:

Province.State	Country.Region	Lat	Long	Date	Confirmed	Deaths	Recovered	Active	WHO.Region
	Bangladesh	23.68500	90.356300	2020-01-22	0	0	0	0	South-East Asia
	Barbados	13.19390	-59.543200	2020-01-22	0	0	0	0	Americas
	Belarus	53.70980	27.953400	2020-01-22	0	0	0	0	Europe
	Belgium	50.83330	4.469936	2020-01-22	0	0	0	0	Europe
	Benin	9.30770	2.315800	2020-01-22	0	0	0	0	Africa

Showing 1 to 25 of 49,068 entries   Previous 1 2 3 4 5 ... 1963 Next

```
 Province.State    Country.Region        Lat             Long
Length:49068      Length:49068      Min.   :-51.796   Min.   :-135.00
Class :character  Class :character  1st Qu.:  7.873   1st Qu.: -15.31
Mode  :character  Mode  :character  Median : 23.634   Median :  21.75
                                    Mean   : 21.434   Mean   :  23.53
                                    3rd Qu.: 41.204   3rd Qu.:  80.77
                                    Max.   : 71.707   Max.   : 178.06
     Date            Confirmed          Deaths          Recovered
Length:49068      Min.   :      0   Min.   :     0.0   Min.   :      0
Class :character  1st Qu.:      4   1st Qu.:     0.0   1st Qu.:      0
Mode  :character  Median :    168   Median :     2.0   Median :     29
                  Mean   :  16885   Mean   :   884.2   Mean   :   7916
                  3rd Qu.:   1518   3rd Qu.:    30.0   3rd Qu.:    666
                  Max.   :4290259   Max.   :148011.0   Max.   :1846641
     Active          WHO.Region
Min.   :   -14    Length:49068
1st Qu.:     0    Class :character
Median :    26    Mode  :character
Mean   :  8085
3rd Qu.:   606
Max.   :2816444
```

The "View Summary Statistics" option allows users to view summary statistics for the cleaned data using the summary function. The summary statistics are displayed in a verbatim text output using the renderPrint function.

Below is the visualization of the data in the Time series and Bar chart:

Visualization in Time Series

```
Mode  :character  Mode  :character  Median : 23.634   Median :  21.75
                                    Mean   : 21.434   Mean   :  23.53
                                    3rd Qu.: 41.204   3rd Qu.:  80.77
                                    Max.   : 71.707   Max.   : 178.06
     Date            Confirmed          Deaths          Recovered
Length:49068      Min.   :      0   Min.   :     0.0   Min.   :      0
Class :character  1st Qu.:      4   1st Qu.:     0.0   1st Qu.:      0
Mode  :character  Median :    168   Median :     2.0   Median :     29
                  Mean   :  16885   Mean   :   884.2   Mean   :   7916
                  3rd Qu.:   1518   3rd Qu.:    30.0   3rd Qu.:    666
                  Max.   :4290259   Max.   :148011.0   Max.   :1846641
     Active          WHO.Region
Min.   :   -14    Length:49068
1st Qu.:     0    Class :character
Median :    26    Mode  :character
Mean   :  8085
3rd Qu.:   606
Max.   :2816444
```
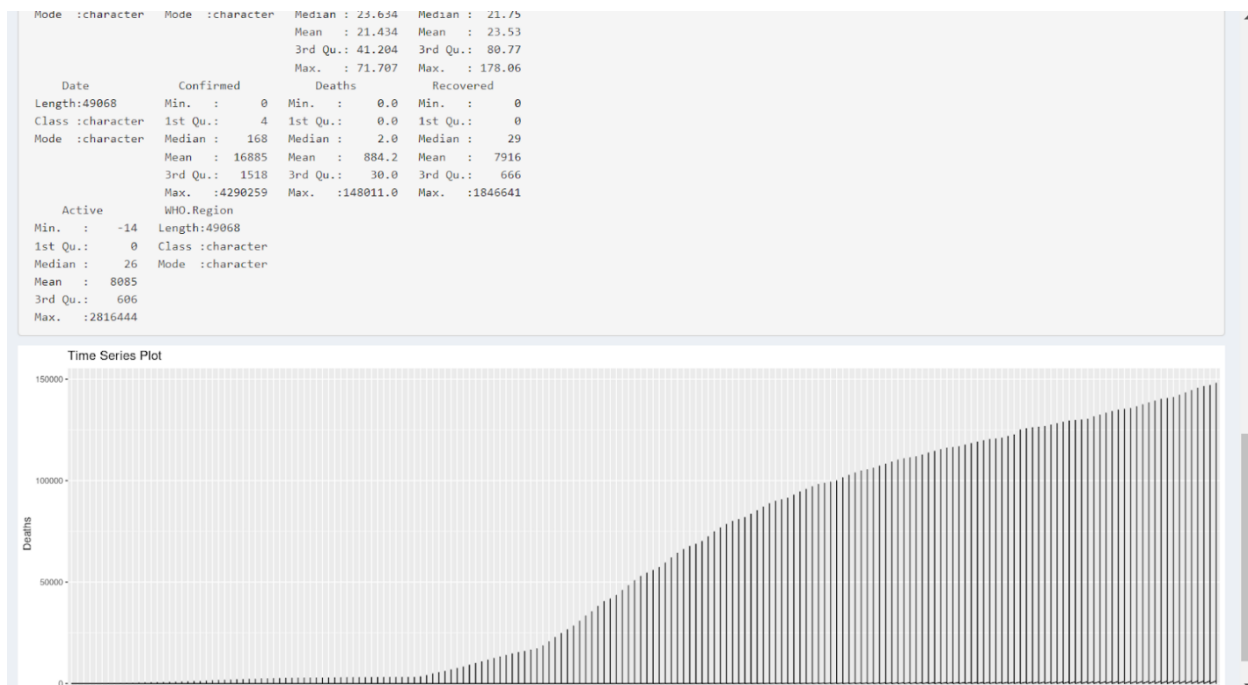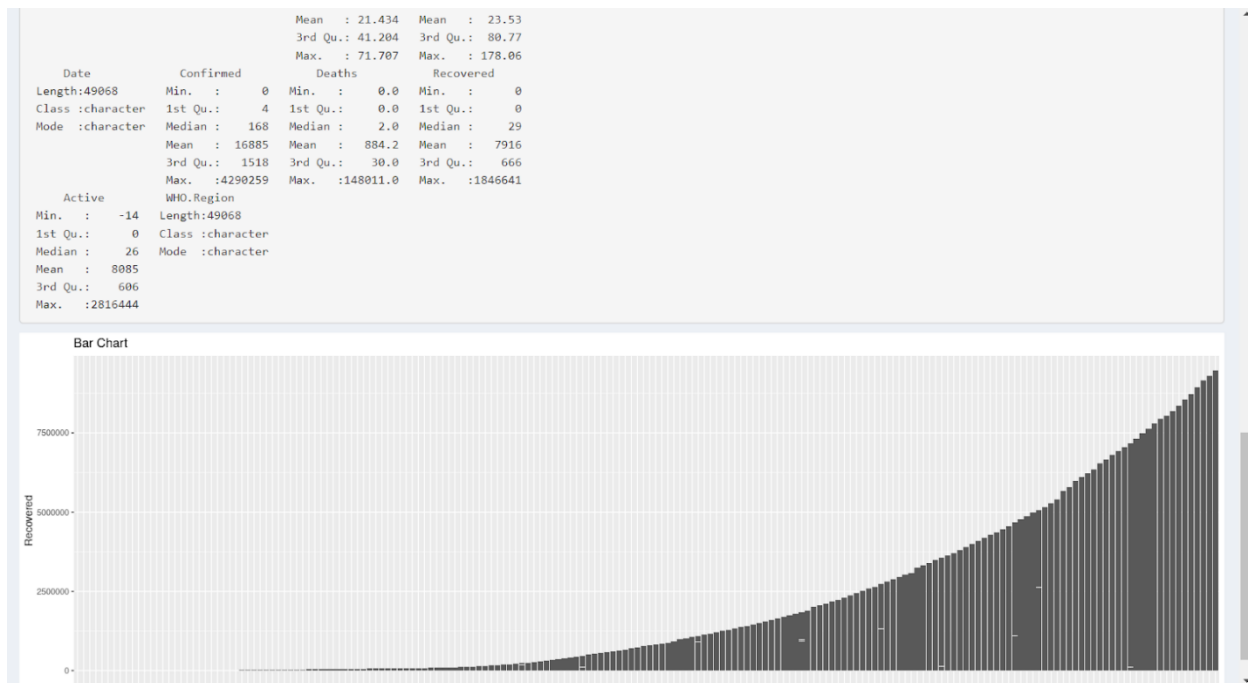

Time Series Plot

Visualization in Bar Chart

The "Choose Plot Type" option allows users to select between a time series plot and a bar chart. The available choices are displayed using the selectInput function. The "Choose X Variable" and "Choose Y Variable" options allow users to select variables to be plotted on the x-axis and y-axis, respectively. The available choices are displayed dynamically based on the loaded data using the renderUI function. The "Create Plot" option allows users to create the selected plot type using the ggplot2 package. The plot is displayed using the renderPlot function.

Overall, this prototype provides a basic framework for a COVID-19 data analysis and visualization web application using R Shiny. With further development, additional features could be added, such as filtering the data, creating interactive visualizations, and incorporating machine learning models for predicting COVID-19 outcomes.

This application provides a simple yet powerful tool for analyzing and visualizing COVID-19 data. Overall, the COVID-19 Data Analysis and Visualization application is a useful tool for analyzing and visualizing COVID-19 data in real time. The application's key features, including data loading options, data cleaning options, data exploration options, and data visualization options, make it a valuable resource for policymakers, health officials, and researchers. However, the application also has some limitations, such as limited data sources, limited analysis options, and limited customization options.

References

[1] Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real-time. The Lancet Infectious Diseases, 20(5), 533-534.

[2] Nair, R., Jayakumar, V., & Manickam, S. (2020). A Shiny App for Interactive Analysis and Visualization of COVID-19 Pandemic Data. Journal of Medical Systems, 44(8), 141.

[3] Schmitz, A., Salzberger, B., & Glück, T. (2020). COVID-19 Dashboard: Real-time monitoring, visualizing, and analyzing of worldwide COVID-19 data. Journal of Medical Internet Research, 22(8), e20997.

[4] Ghosal, S., Sinha, B., Majumder, M., & Misra, A. (2020). Estimation of effects of nationwide lockdown for containing coronavirus infection on worsening of poverty, health, and hunger indices in India: a panel data analysis. Social Science & Medicine, 113660.

[5] Abdollahpour-Alitappeh, M., Mirzaei, R., Ghorbani, J., & Tavakoli-Yaraki, M. (2021). COVID-19 and Interventions to Counteract Associated Risk Factors: A Review. Archives of Medical Research, 52(4), 365-374.

[6] Abdellaoui, B., El Khalfi, B., & Lahcen, A. A. (2020). The impact of population density, age distribution, and interventions on COVID-19 spread: a comparative study on 25 countries. Environmental Science and Pollution Research, 27(29), 36547-36556.

[7] Huang, Y., Li, L., Gan, Y., Wang, W., Liu, J., Huang, Y., & Hu, B. (2020). Factors influencing the severity of COVID-19 in China: A meta-analysis. Journal of Medical Virology, 92(10), 2188-2196.