Data Exploration and Visualization

**Global Suicide Rates**

|  |  |
| --- | --- |
| Name: | **Ujwal Kavalipati** |

Table of Contents

[1.Introduction 3](#_Toc56376696)

[2.Design 3](#_Toc56376697)

[3.Implementation 5](#_Toc56376698)

[3.1 Libraries Used 5](#_Toc56376699)

[3.2 Implementation 5](#_Toc56376700)

[4. User Guide 6](#_Toc56376701)

[4.1 Prerequisites 6](#_Toc56376702)

[4.2 Crucial Note 7](#_Toc56376703)

[4.3 Installing 7](#_Toc56376704)

[4.4 Viewing and Exploring the Visualization 7](#_Toc56376705)

[5.Conclusion 11](#_Toc56376706)

[6.Biblography 11](#_Toc56376707)

[7. Appendix 12](#_Toc56376708)

# 1.Introduction

Data visualization is regularly promoted for its ability to reveal stories within data, yet these “data stories” differ in important ways from traditional forms of storytelling [1].  The main purpose of this narrative visualization is to explain our findings to the user. With the help of interactive plots, we can make the data speak to the user. Users can find it more engaging if we use narrative and interactive visualizations. As part of my exploration project, I have explored and visualized the global suicide rates dataset. The dataset is from Kaggle[2]. When trying to build the narrative visualization based on 5 design sheets, I had to use a shapefiles dataset to plot the boundaries of all continents. The shapefiles dataset is from ArcGIS Hub [3].

It takes so much courage to take one’s life. I selected the suicide dataset so that I can know about how every country’s suicide trends are changing with time and what age-groups are committing more suicides. The dataset does not have factors contributing to suicides, it only has demographic details about suicides in every country, gender, and generation types.

I want users to find out the following things from my narrative visualization.

* Find out which gender, generation type, and age-groups are committing suicides in each continent and compare the suicide trend in each continent.
* Find out the suicide rate in each country from 1985-2015 and compare between various countries.

This narrative visualization is intended for Government organizations and sociologists who are concerned about suicides. Knowing about suicides helps them to make better decisions and understand their people well.

# 2.Design

I have used five design sheet methodology to design my visualization. The FDS has 5 sheets total where there are three sheets for the initial design and one final design sheet which has the ultimate design for our visualization. Each sheet has advantages and disadvantages outlined in them which helped me in taking a good final design.

1. The first sheet is the **Brainstorming** sheet where the following things are considered.

* Consider the suicide dataset and observe the columns in it.
* Put down all my ideas I want to do. There are a couple of pie charts and line charts along with a world map. I wanted to give a slider bar for a selection of years as well.
* I filtered the ideas in my next step and removed redundant graphs or overlapping ideas.
* Next step, I have categorized the ideas and ordered them based on the world, continent, and country level.
* Then, I organized all my designs such that I tell the users the story I am willing to convey and make them understand the suicide statistics worldwide.

2. The second sheet is the **Initial design 1**:

* Here, a leaflet map is displayed showing all the continents of the world.
* Users can hover or select a continent. User can see the continent name and average suicides in that continent by hovering.
* The focus of the visualization would be the user able to see a pie chart of Gender suicide proportion, bar charts of age-groups and generation types vs suicides, and also the trend of suicides in that continent over the years 1985-2015.
* The advantage of this visualization is that it can help the users get the overall picture of various suicide statistics shown at a time through different graphs.
* Cons of this design sheet include that users can find that this information is not analyzed at a deeper level and look out for more options.

3. The third sheet is the **initial design 2**:

* This sheet focuses on conveying users about suicide statistics in the world.
* A line chart is shown initially which will display the suicide trend in the world as time progresses.
* Users can select categories like age-group, gender, generation type, and get a particular graph showing the selected statistics.
* The pros of this visualization include getting worldwide suicide statistics under different categories just by selecting options.
* The cons of this design include a lack of interaction in the graphs and over simplicity of graphs.

4. The fourth sheet is the **initial design 3**:

* This sheet is mainly designed so that the user can compare the suicide rates in different countries at a time with the choice of years.
* There is a year slider that can be used to select the years from 1985-2015. The years are kept as numerical so that users can better understand if the suicide rate increased mid-year or remained constant throughout.
* The country selection box can be used to select multiple countries.
* For each selection, the user gets the line chart showing each country’s suicide rates over the selected years.
* The advantage of this design is that it offers high flexibility to the user in selecting the years and also it can show the number of countries the user desires to analyze.
* The disadvantage of this design is that users can see only line charts and a deeper level of information is not provided.

5. The fifth sheet is the **Realization**:

* This is the final sheet which is considered after taking and weighing down the pros and cons of all previous design sheets.
* This design has 2 tabs mainly – blend of sheet 4 and sheet 2.
* It has world continents with categories on one tab and a line chart with a year slider and country input box.
* The dataset requires wrangling steps and data reformation.
* The time estimated to build the design is 15-20 hours.
* Suitable colors are selected based on discriminability and perceptual organization. Care is taken to ensure that proper color brewers are used and graphs are chosen to differentiate ordered and categorical data. Interactive graphs are put into the design so that the user is engaged.

I had planned to use complex charts but my intended audience may find it difficult to understand them. My main motive is to make my end users interact with different charts in my visualization and get a better picture of the suicide statistics around the globe.

I have not included world statistics in the final design as that was so simple and I want to give users options for comparison. I also considered grouped bar charts showing suicide proportions of gender but felt that the pie chart would be a better design.

Instead of showing 4 graphs at a time as in sheet 2, I thought it would be better if the user can select the category of preferred choice and get to see that graph in depth. The final sheet is designed so that users can understand the suicide statistics level by level. Continent level and country level. The data table is also shown in the countries tab so that users can get to see the suicide rate in that particular country. Continent tab helps users understand trends,age-groups, gender proportions, and also generation types.

# 3.Implementation

The first step in implementation was to convert my raw CSV file into a usable form. Wrangling steps were already made before for the data exploration.

I have decided to implement my design using **R shiny**. The challenging part of this visualization was to make the leaflet map interactive. I initially knew plotting a leaflet map with circle markers for which we need coordinates.

I have gathered coordinates for all continents and plotted them. That was not the type of graph I was expecting. I want the whole continent to be highlighted when the user hovers on the map. Then, I searched a lot and tried experimenting with shapefiles.

The shapefiles dataset I found helped me in plotting the boundaries. But I wanted to display the average suicides in that continent when the map has hovered. So I had to modify the shapefile data frame to include the suicide statistics and order the continents. Adding the interactivity also took quite a bit of time and research as I wanted to get the continent name when the map is clicked and based on that display the graphs below. I will further explain how I achieved this functionality.

## 3.1 Libraries Used

1. shiny – for interactive web applications
2. shinythemes - for various themes of the background
3. leaflet- for rendering the leaflet map (continents)
4. ggplot2- for plotting
5. dplyr – provides various functions for data frame manipulation
6. DT - DT provides an R interface to the JavaScript library DataTables.[4]
7. rgdal – R Geospatial Data Abstraction Library for dealing with shapefiles
8. plotly – for interactive and high-quality graphs
9. gganimate – for drawing animated charts
10. gifski – for combining images as a gif. gganimate uses this package.

## 3.2 Implementation

* For implementing the design, I have chosen R shiny over D3 because I am more confident in building applications with shiny. With the time I had in hand, I thought R would be a better choice for me to build the application according to the five design sheet methodology.
* I strongly believe that any application should be user friendly and serve its purpose. So, I decided that my design would be simple and easy to use.
* I thought to give an introduction to my application in the first tab. This contains information about how to use the application and what to expect from it. I have placed an animated line chart here which tells about the suicide trend worldwide.
* The trend is a decreasing one. I felt that users using my application should not start with a negative mood and this graph can cheer them up as it clearly shows global suicide rates are decreasing.
* I have used gganimate with gifski library to build the animated chart. The only difference between a static graph in ggplot and an animated line chart (gganimate) is that we have **transition\_reveal()** function in gganimate which will do all the magic. The year column is passed in it and thus the chart becomes animated.
* The second tab is the **Continent statistics** tab which has a leaflet map and 4 types of graphs.
* The critical part of implementing the application is making the world map interactive. As I told before, I have used the shapefiles dataset. The dataset is read through **readOGR()** function of the **rgdal** library.
* A continent Tibble is created and it had an average suicide rate in each continent. Next thing was to include this information in the shapefile dataset. For that wrangling is required to add the missing continents in the Tibble.
* After a few manipulations, the shapefile is ready to be plotted. A leaflet map is drawn with highlight options and customized labels. It is a choropleth map and depicts the color based on the suicide rate.
* **layerId=~CONTINENT** is used in the leaflet map to make it return the continent name when clicked. The shape click event by default returns an id, latitude, and longitude of the position where the mouse is clicked. giving the layerId as ‘continent’ is a breakthrough as it helped in the further steps.
* Based on the continent clicked, I have filtered the **suicide** data frame and plotted the graphs.
* I have used if-else statements to check which radio button is selected and displayed the corresponding graph through **plotly()**.
* Proper labels are given when the user hovers on the graphs.
* The next tab is **Countries** where line chart is displayed.
* For implementing this functionality, I have used **sliderInput** for selecting the years and **selectizeInput** for selecting the countries. selectizeInput acts as both a search box and a select box with options of multiple selections.
* Based on the user inputs, a line chart is plotted by **ggplot.** I have also displayed Data Table below the selection boxes so that if some official wants to know the suicide rate in any particular year, he/she can just see it from there.
* All the countries selected are ordered alphabetically in the table.

# 4. User Guide

## 4.1 Prerequisites

1. R version: 3.6.3 and above
2. shiny version:1.4.0.0 and above
3. shiny themes
4. leaflet
5. ggplot2
6. dplyr
7. DT
8. rgdal
9. plotly
10. gganimate
11. gifski

## 4.2 Crucial Note

To render the image and animated line chart in the **Introduction** panel, it is crucial to place the wrangled file **sucide.csv** : It has 27493 records and 11 attributes in the same folder as the R shiny app.

The following folders:

**“World\_Continents”** – containing the shapefiles starting with “4a7d27e1-84a3-4d6a-b4c2-6b6919f3cf4b202034-1-2zg7ul.ht5ut” name.

**“www”** – containing the image **“depression.jpg”** and animated line chart **“outfile.gif”.**

should be in the same folder as the R shiny app.

## 4.3 Installing

* The first and foremost thing is to ensure that R is installed.
* The next step is to install all the necessary packages with the following command.



* install. packages(“package name”) should be used to install the required packages for the application to run.
* A message should be displayed that the package is installed successfully.
* You can press ‘??package name’ to read the official R documentation for that package.
* You can use ‘require(package name)’ to check if a package is installed. The statement should not give any error.

## 4.4 Viewing and Exploring the Visualization

* If you are using R studio, you can click on the **Run App** button on the top right corner to start the application.
* Global Suicide Rates shiny app is divided into three panels:

a. Introduction

b. Continent Statistics

c. Countries

* **Global Suicides** is the dashboard name.
* Initially, the panel with the **Introduction** is displayed with a suicide image[5] and an animated line chart. The image is copyright free. Users can click on **Continent Statistics** or **Countries** to toggle between the tabs.

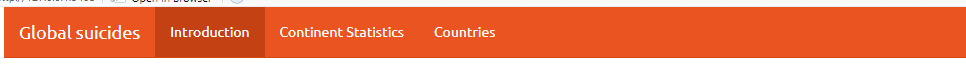


Figure Panels of the dashboard

* The first panel displays some information about Global Suicide Rates and what users can expect from the application.

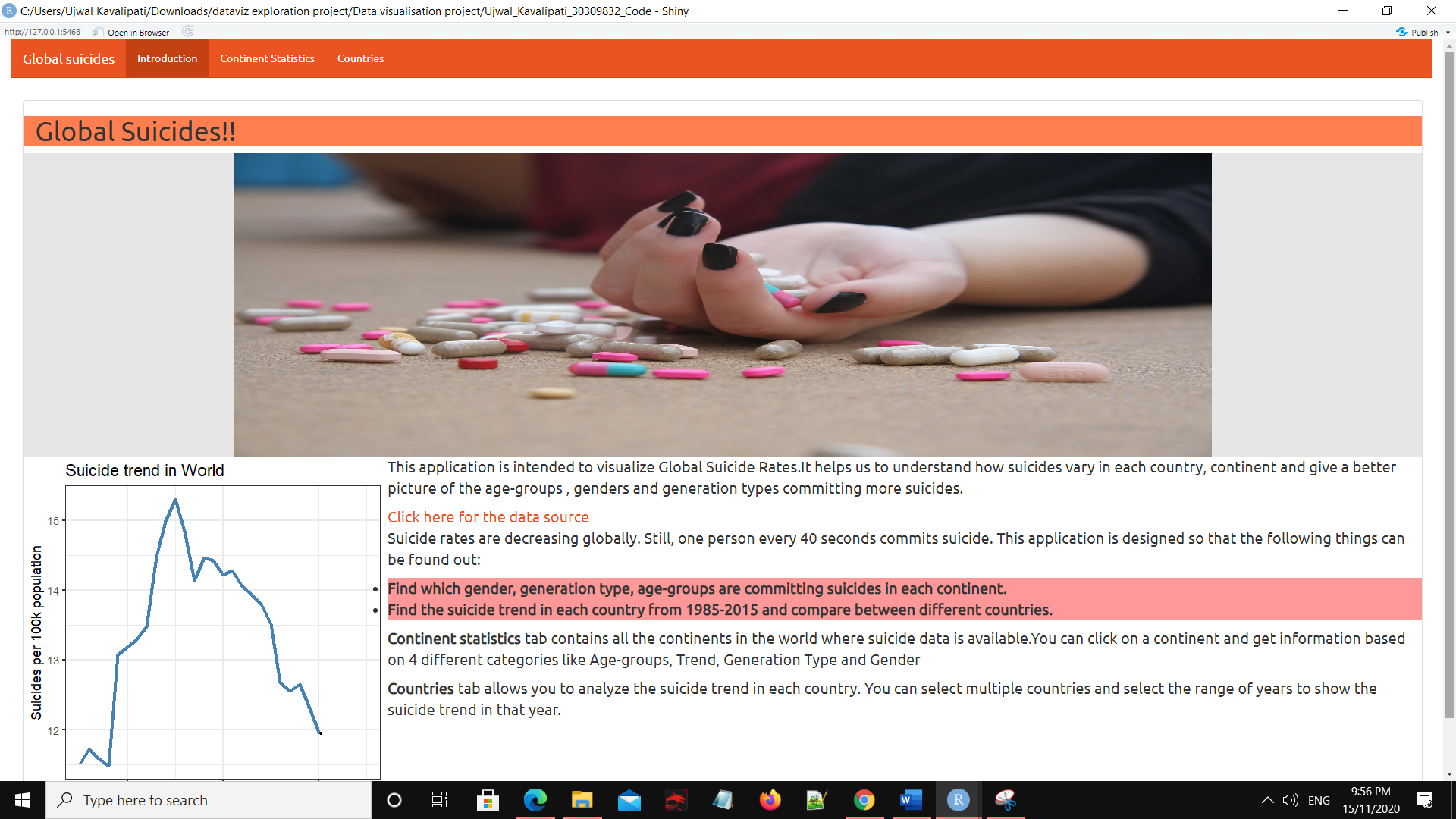


Figure Introduction panel

* If the user clicks on the **Continent Statistics** tab, a world map is displayed first on the screen.
* Sometimes, it might take a couple of seconds to display the map. But the options below the map and informational text will be displayed soon.

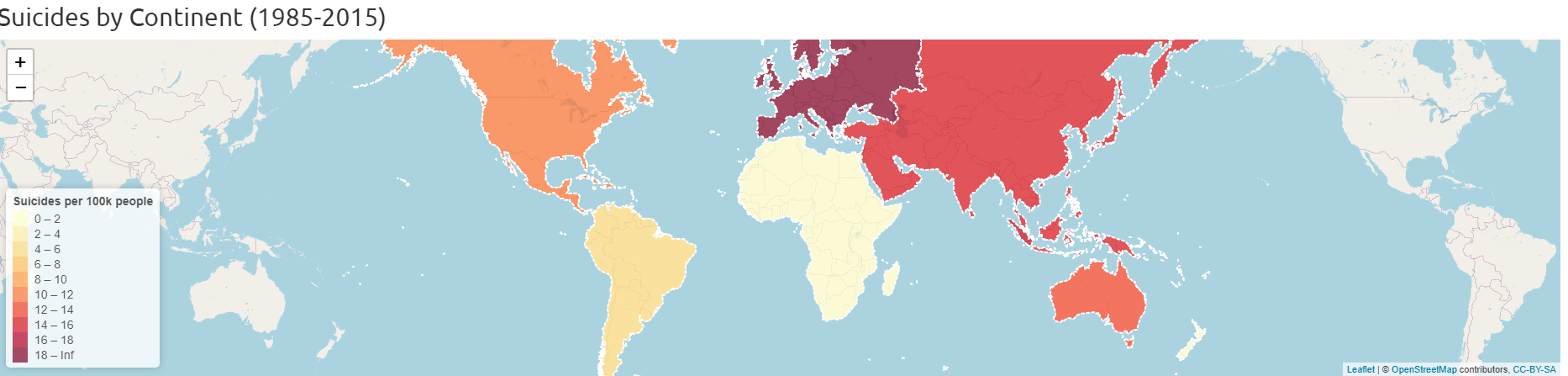


Figure Leaflet map showing all continents

* Users can hover on the map to get information about the continent name and suicides per 100k population in that continent.
* For example, if the user hovers on Australia, the following is displayed.



Figure Hovered Australia

* Users can select the radio buttons given below the map to explore the options like Age-groups, Trend, Generation Type, and Gender. User can only click on the options after he/she clicks on a continent. Without clicking on the continent, if the user just changes the options, no plot would be displayed.

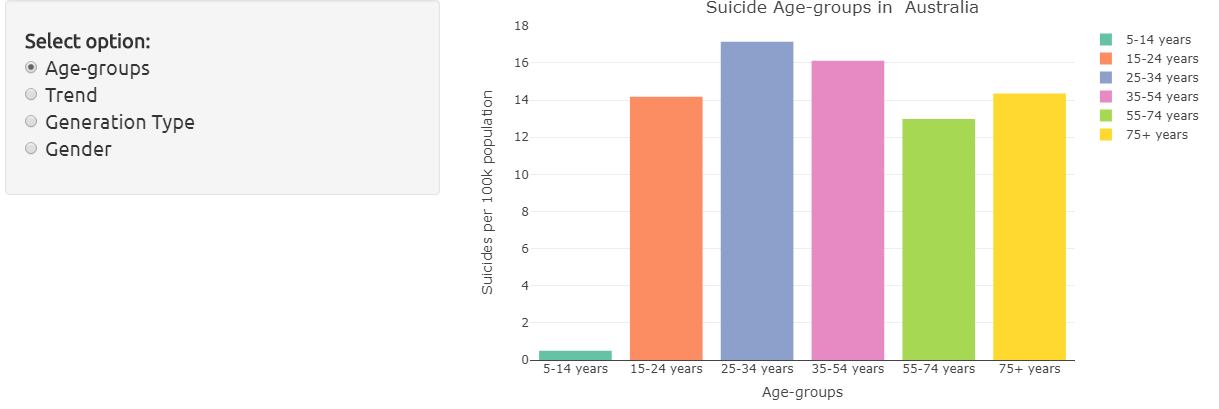


Figure Showing the options and Age-group Vs Suicide graph in Australia

* For example, the below graph is showed when the user clicks on Europe and selects the Gender option.

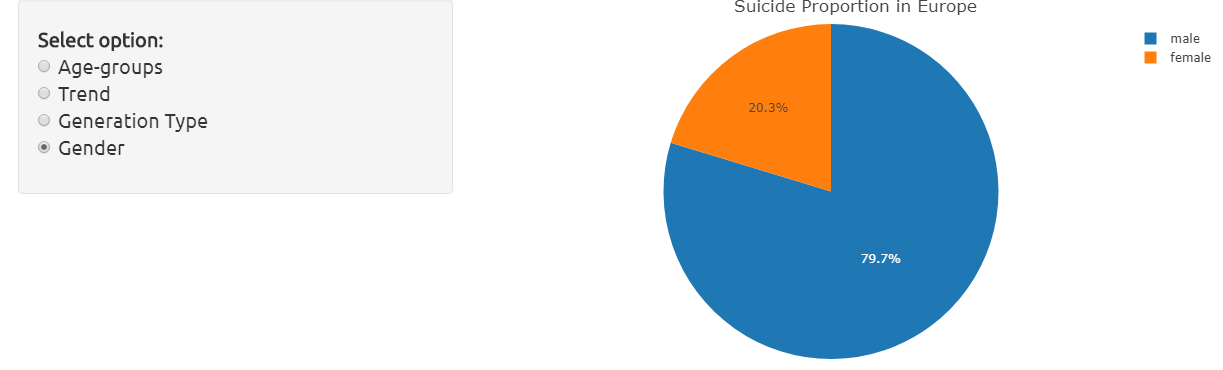


Figure Showing Suicide proportion in Europe

* If users click on the **Countries** tab, by default a line chart will be displayed showing **Albania’s** suicide trend between the years 1985-2000. Users can select the years from the slider input and countries from the input selection box. The box can also work as a search box for the user to type the country’s name.

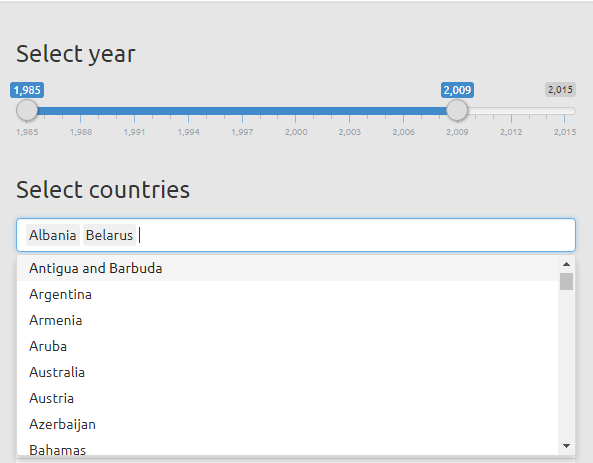


Figure Showing slider bar and the country select box

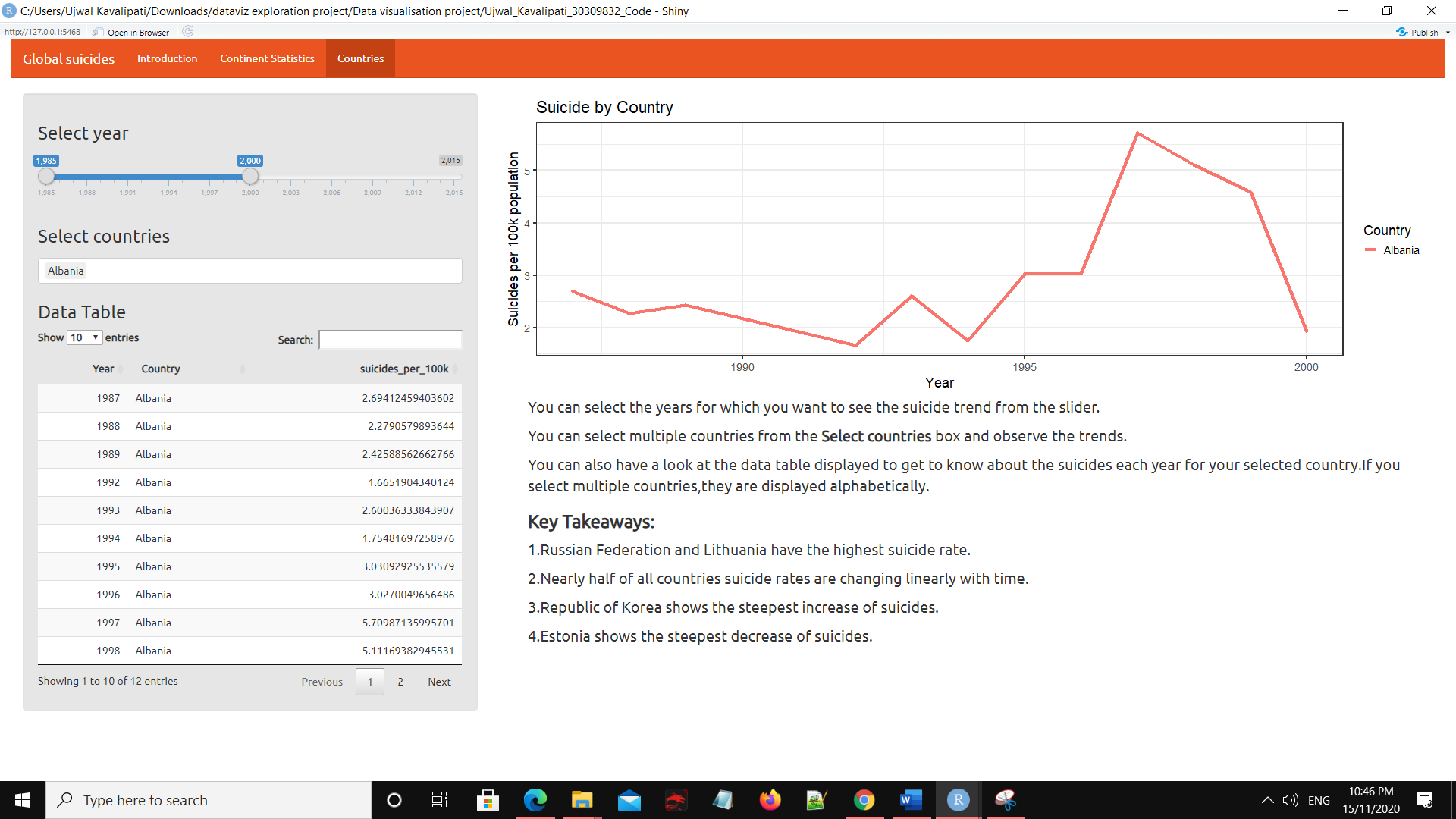


Figure Showing default page on clicking of **Countries** tab

* Users can select the required years and countries to observe and compare the suicide trends. If the user selects the years 1999-2015 and countries Canada, Guyana, and Lithuania, the following plot is displayed.

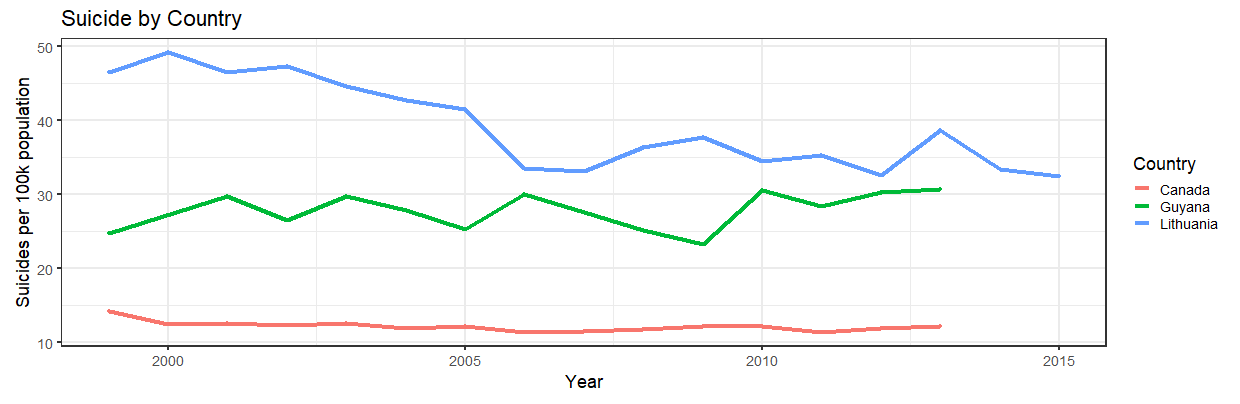


Figure Showing suicide trend in Canada, Guyana, and Lithuania

* Along with the plot, a Data Table is also displayed which displays the Year, Country, and suicides\_per\_100k population details. This information can help the user to get the suicide rate for the selected years and countries sorted alphabetically.

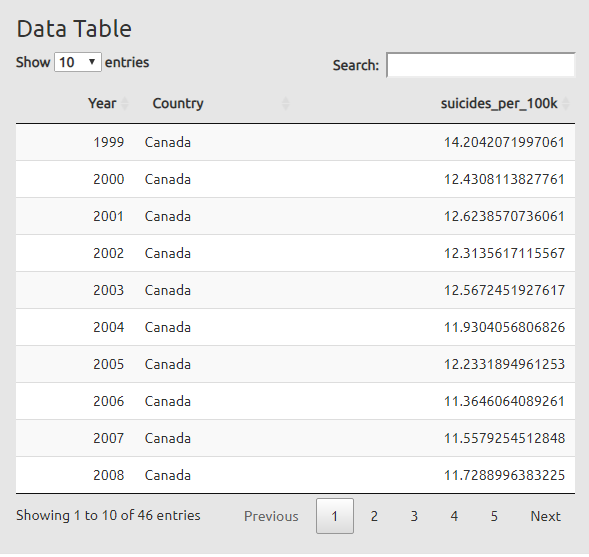


Figure Showing data table for Canada

# 5.Conclusion

From the visualizations, it is observed that Suicide rates are decreasing globally. Also

* South America and North America have increasing suicide trends.
* Men’s suicide rate is higher than for women.
* GI Generation committed the highest number of suicides.
* Russian Federation and Lithuania have the highest suicide rate.

Apart from the findings from Global Suicide Rates, I have learned how to built robust shiny applications and use various libraries in R. I also learned about shapefiles and rgdal library. I am grateful to use my skills and achieve my desired functionality. It is very satisfying that a web application is developed from scratch and interactive visualizations are used in it.

If I want to do this again, I would try using D3 to build the application. D3 offers various controls for the user interface and beautiful interactive graphs. I would also try to do some analysis with the gdp\_for\_year and gdp\_per\_capita columns to know if there is any relation between a country’s economic status and suicide rates. Thus, the global suicide rates dataset helped me to know about the suicides globally in each country and continent and also understand the suicides for gender, age, and generation type.

# 6.Biblography

[1] <https://ieeexplore.ieee.org/document/5613452>

[2] <https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016>

[3] <https://hub.arcgis.com/datasets/esri::world-continents/data?geometry=-174.023%2C-84.708%2C174.023%2C57.126&selectedAttribute=CONTINENT>

[4] <https://rstudio.github.io/DT/>

[5] <https://pixabay.com/photos/depression-mental-health-sadness-824998/>

# 7. Appendix

A picture containing text, whiteboard

Description automatically generated

Figure 11:Sheet 1- Brainstorming

A close up of text on a whiteboard

Description automatically generated

Figure 12: Sheet 2- Initial Design 1

A close up of text on a whiteboard

Description automatically generated

Figure 13: Sheet 3- Initial Design 2

Text, letter

Description automatically generated

Figure 14: Sheet 4- Initial Design 3

Diagram

Description automatically generated

Figure 15 : Realization