

**Data Visualisation**

|  |  |
| --- | --- |
| **Assignment Deadline:** | **15/04/2024** |
| **Module Code:** | **CS5803** |
| **Title:** | **Global suicide visualisation** |
| **Module Leader:** | **Dr. Timothy Cribbin** |
| **Marker：** | **Dr. Timothy Cribbin** |
| **Student Name:** | **Meng Heng Chong** |
| **Student Number:** | **2374392** |
| **Degree Programme Title:** | **MSc in Artificial Intelligence** |
| **Academic Year:** | **2023/2024** |

CONTENT

1.0 Introduction

1.1 Data………………………………………………………………….…………….1

1.2 Persona and questions……………………………………………….…………….2

1.3 User requirement specification……………………………………………….…2-3

2.0 Design………………………………………………………………….……4

3.0 Implementation

3.1 Tableau implementation…………………………………………………………5-7

3.2 Power BI implementation………………………………………………………….7

3.3 Statistical modelling

3.3.1 Cluster analysis…………………………………………………….……8

3.3.2 Factor analysis……………………………………………………..….8-9

4.0 Walkthrough

4.1 Tableau walkthrough……………………………………………………….…10-11

4.2 Power BI walkthrough…………………………………………………………...12

5.0 Reflective Discussion

5.1 Difference in design and actual implementation…………………………...…13-14

5.2 Critical evaluation of visualisation tools…………………………………………14

5.3 Applied principle of good data visualisation………………………………….14-15

5.4 Limitations……………………………………………………………………15-16

5.5 Personal learning…………………………………………………………………16

6.0 Conclusion…………………………………………………………………16

7.0 Reference……………………………………………………….………….17

1.0 Introduction

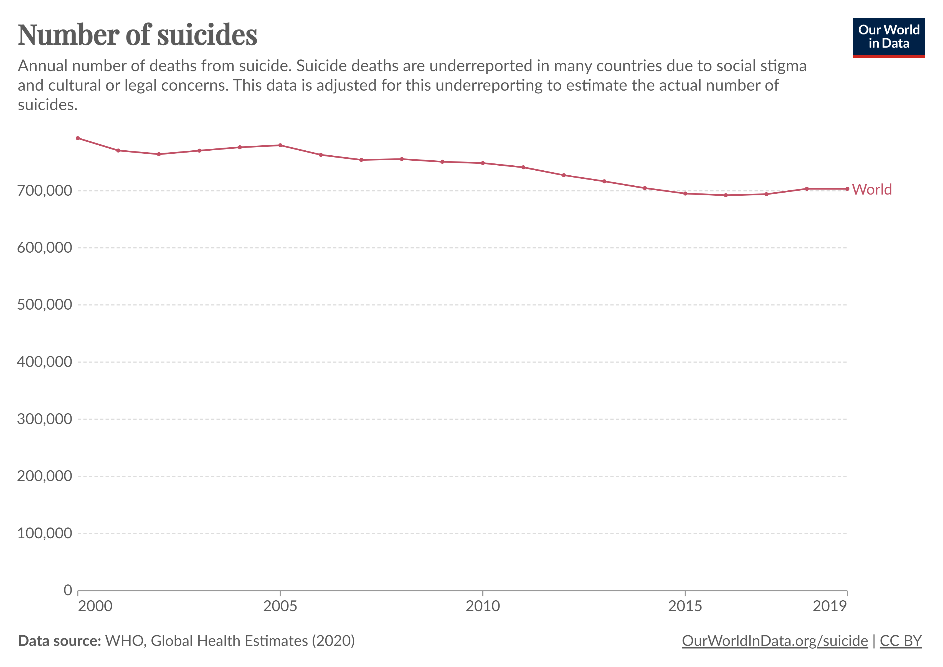


Figure 1.1 Global suicide cases from 2000-2019 (Dattani et al., 2024)

From 2000 to 2019, there are approximately 700k people around the world took their own lives every year and there is no sign of relief (Dattani et al., 2024). Such tragedies have raised serious public health issues around the world and raised attention by public health organisations such as WHO (WHO, 2023). To get insights on global suicides, global suicide data is analysed and visualised.

**1.1 Data**

‘Suicide Rates Overview 1985 to 2016’ was retrieved in 2018 from the Kaggle website (Rusty, 2018). The dataset consists of 12 variables and 27820 observations from 1985 to 2016 across 101 countries.

|  |  |  |
| --- | --- | --- |
| Name | Description | Domain |
| Country | Country in which the data is recorded | Nominal |
| Year | The year in which data is collected | Integer |
| Sex | Gender of the individual | Nominal |
| Age | The age group of suicide | Nominal |
| Suicide.no | Total number of suicide cases | Integer |
| Population | Population of the country | Integer |
| Country-year | A concatenation of the country and year columns | Nominal |
| Suicides/100k population | No. of suicides per 100k people in the population | Integer |
| HDI for year | Human development index for the year | Integer |
| gdp\_for\_year ($) | Annual gross domestic product (USD) | Integer |
| gdp\_per\_capita ($) | Annual gross domestic product per citizen (USD) | Integer |
| generation | Generation of suicide | Nominal |

Table 1.1 Data dictionary

**1.2 Persona and questions**

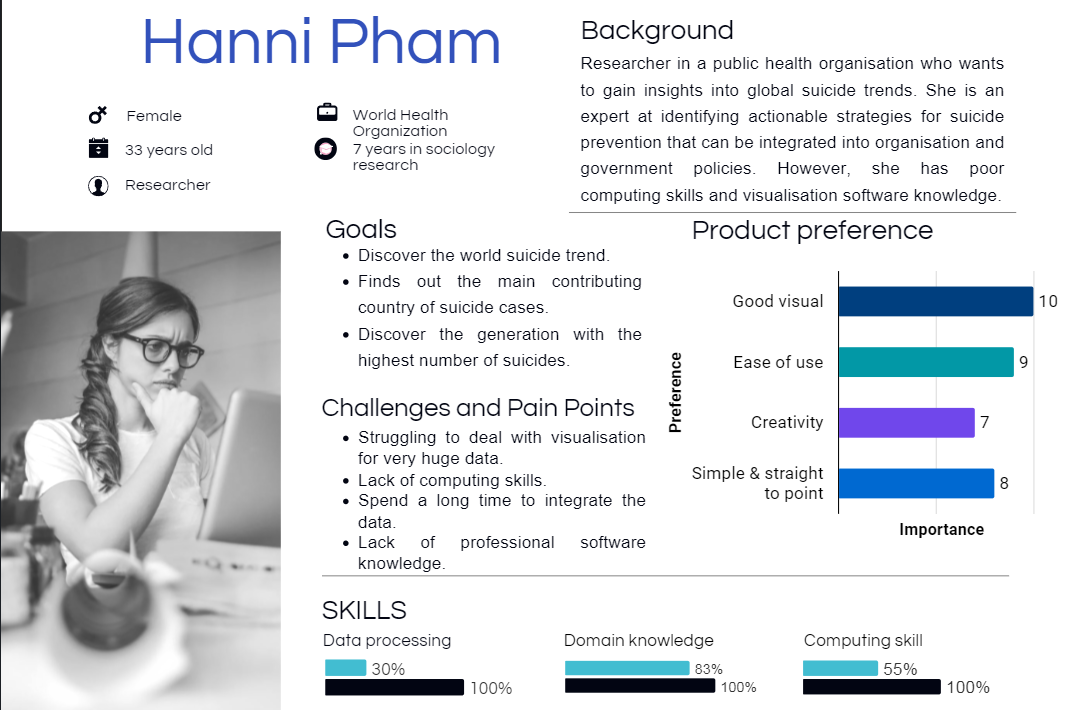


Figure 1.2 User Persona

From the user persona above, several research questions can be defined:

1. What are the top 3 countries with the highest number of suicides?

2. Which age group have the highest number of suicides and how many suicides in that age group?

3. How many male suicides in the 35-54 years age group in Japan during the year 2014?

**1.3 User requirement specification**

1**.** To answer Q1 the user needs to see the top 3 countries by number of suicides.

* Create a bar chart of countries by suicides with no. of suicides encoded into the length of the bar.
* Countries are arranged from high suicide to low suicide.

**2.** To answer Q2 the user needs to identify the number of suicides in the age group with the highest number of suicides.

* A bar chart of age groups vs no. of suicides with no. of suicides encoded into the length of the bar.

**3.** To answer Q3 the user needs to identify the number of male suicides in the 35-54 age group in Japan during the year 2014.

* Male suicides and female suicides are colour-encoded to identify the number of male suicides.
* Year and country filter to focus on a certain country in a certain year.
* Interactive filter to identify the number of suicides in specific countries, time ranges, genders and age groups.

2.0 Design

|  |
| --- |
| Global suicides visualisation |
| 1. What are the top 3 countries with the highest number of suicides?  2. Which age group have the highest number of suicides and how many suicides in that age group?  3. How many male suicides in the 35-54 years age group in Japan during the year 2014? |
| Potential needs: Visualize the distribution of suicides in different countries  Q1) Rank suicides no. in different countries  Q3) Filters focus analysis on specific interval and country.    Q3) Set all chart as filter to form interactive dashboard  Potential needs: Visualise the suicide no. in each year.  Q2) See the age group with the highest suicides no.  Q3) See the no. of male suicides in 35-54 age group. |
| Result: Users can gain insights into the requirement of 3 research questions.  Q1: Top 3 countries with the highest suicides from the countries ranking chart.  Q2: Highest suicide age group from the chart of suicides in different age groups.  Q3: The number of male suicides in specific age groups, specific countries, and specific year intervals through brushing and linking.  The suicide trend charts and density map of world suicide distribution are built to give richer information in the dashboard. As the user is a sociology researcher, he/she might need to get insight into suicide trends and suicide distribution. |

3.0 Implementation

**3.1 Tableau Implementation**

To create a global suicide visualisation which can answer all research questions comprehensively, a total of 5 charts are implemented and combined to form a dashboard. Firstly, master.csv file is imported into Tableau.

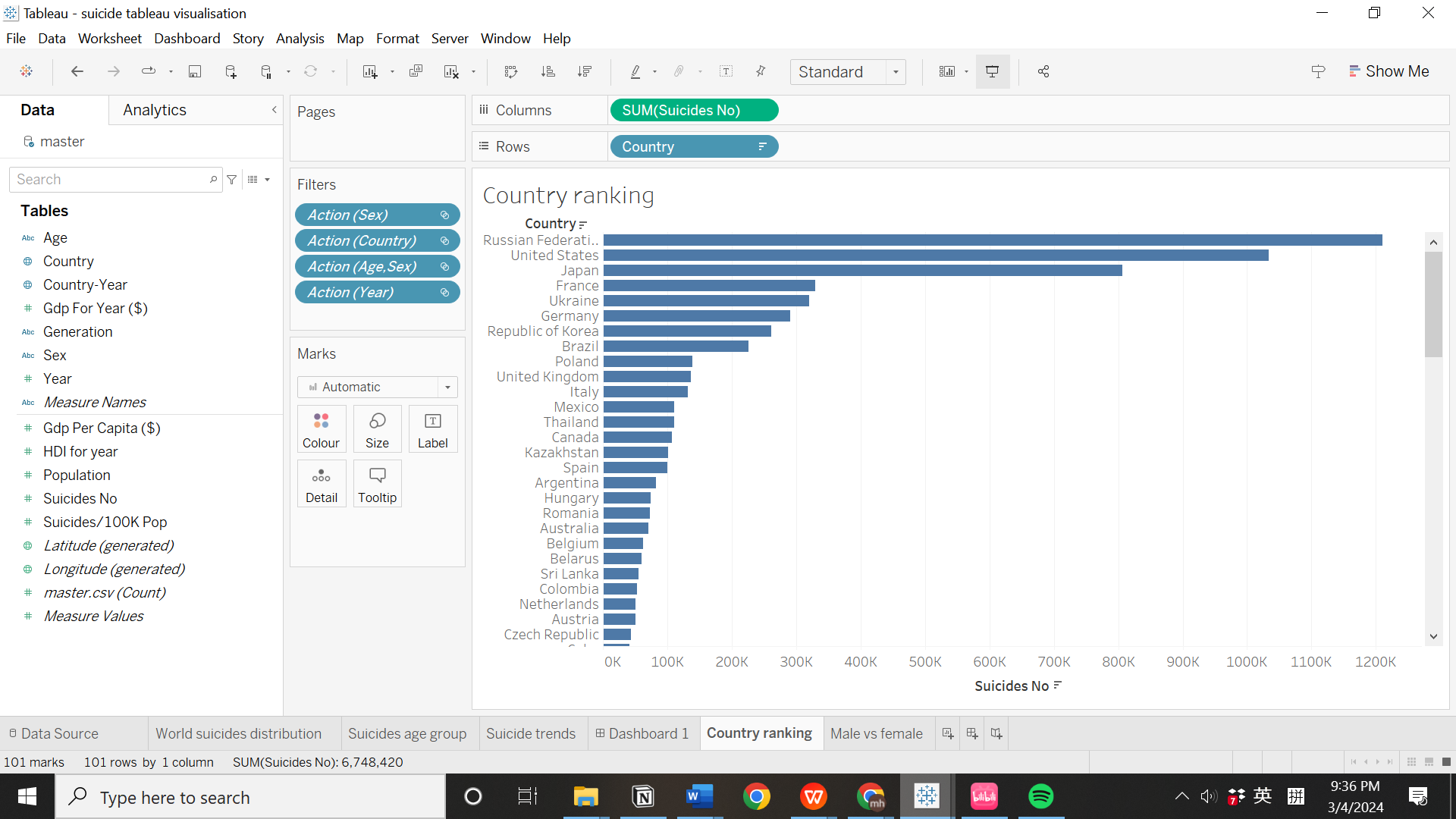


Figure 3.1 Bar plot showing country ranking in no. of suicides

A bar chart of ‘Country’ against ‘Suicides No’ is built to satisfy the user’s need to see the top 3 countries by suicide cases. Countries are arranged from high to low suicides through descending sorting. From the bar chart, users can easily find the top 3 countries.



Figure 3.2 Density map of world suicide distribution

While satisfying user needs to observe the top 3 suicide countries, a density map is also used to give a better insight into global suicide distribution. To create a density map, ‘Country’ and ‘Suicides No.’ are dragged to the worksheet, and then clicking the density map in the ‘Show Me’ panel.

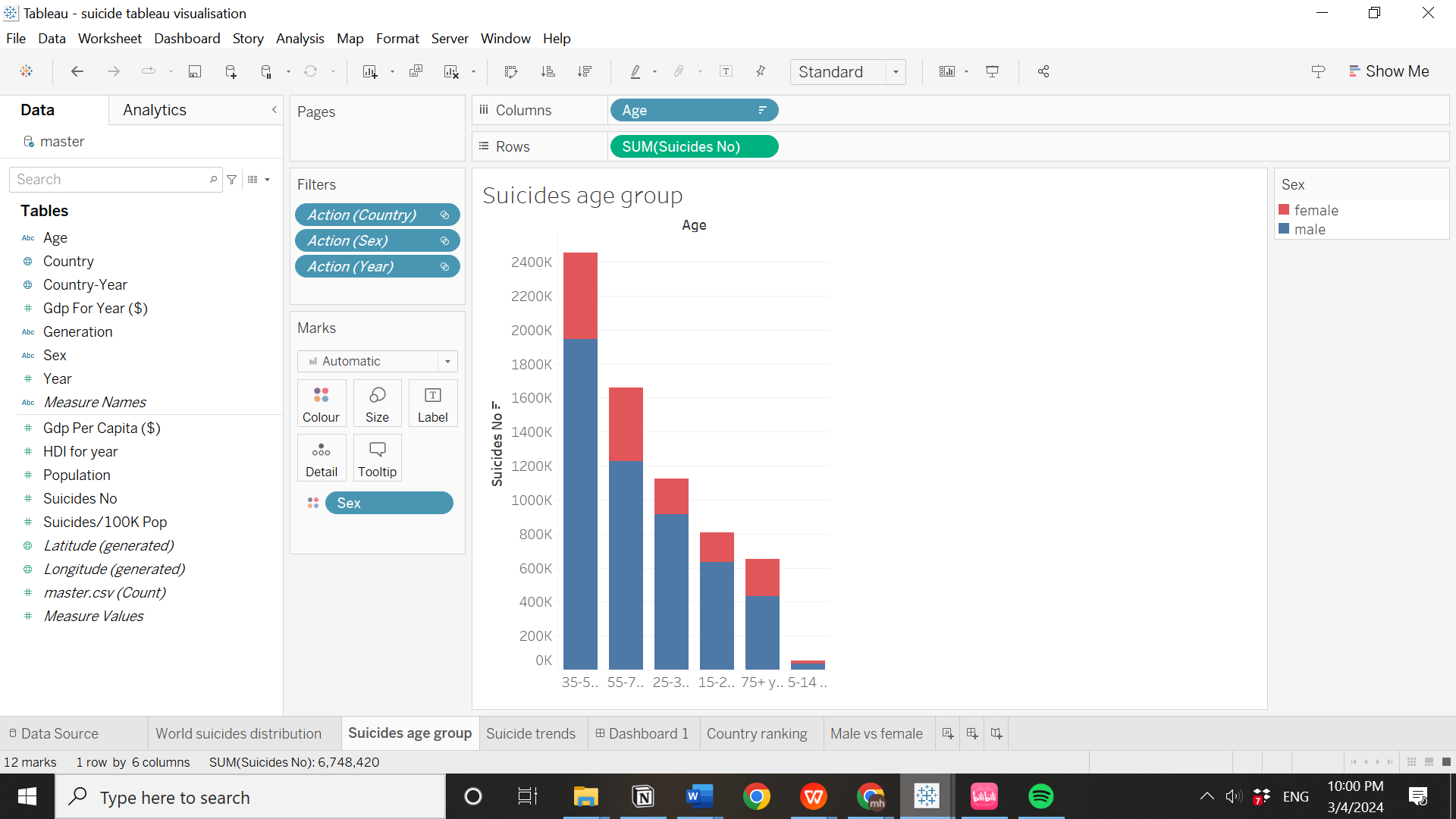


Figure 3.3 Completed stacked bar chart

A stacked bar chart of age groups against no. of suicides is built to satisfy the user’s need to see the age group with the highest number of suicides. The age group is arranged from high to low suicides through descending sorting. From the bar chart, users can easily find the highest suicide age group. To fulfil the requirement of Q3 to observe male suicide, the number of suicides in each age group is colour-encoded into males and females through colour mark card.

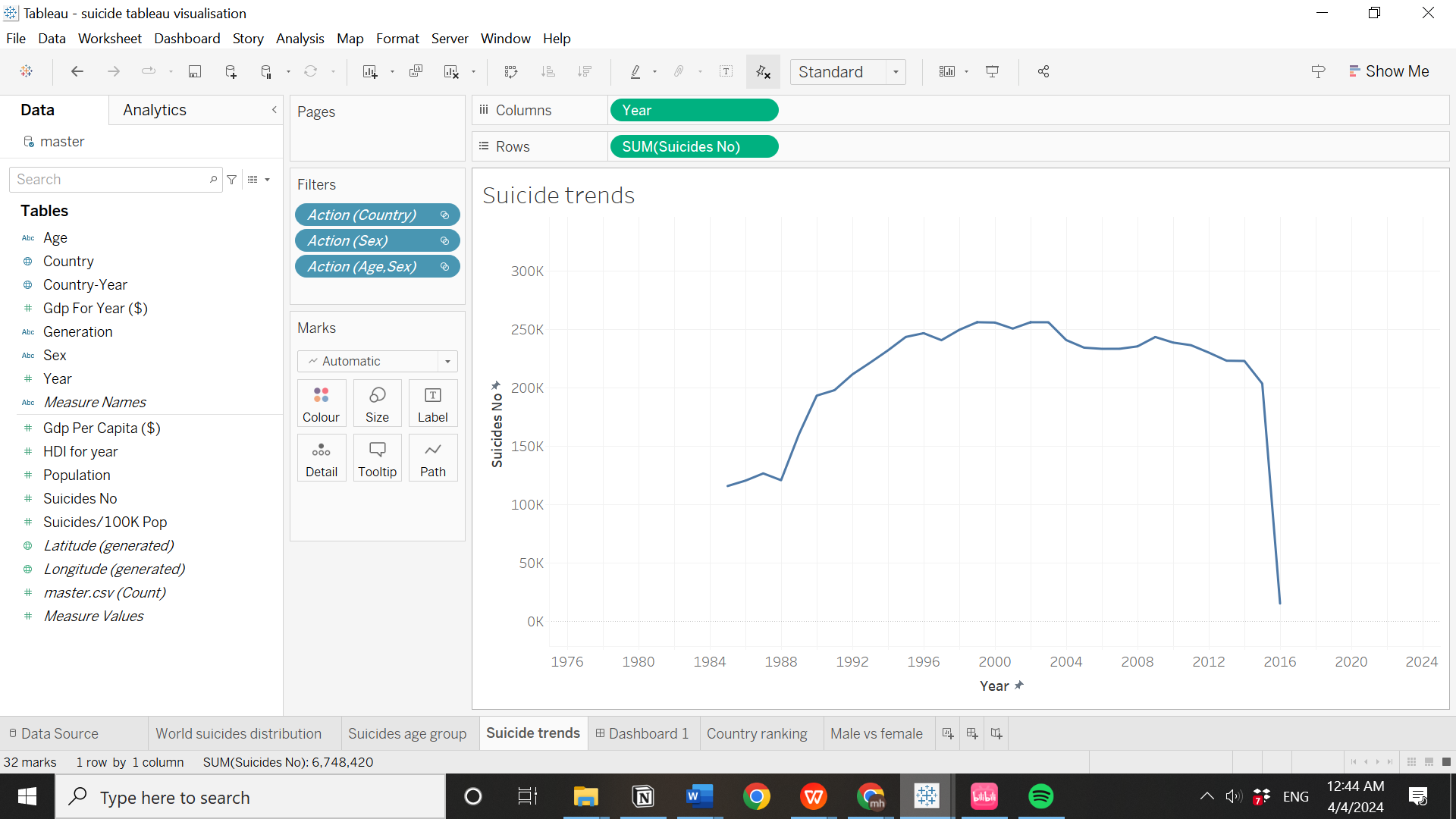


Figure 3.4 Completed line chart

To satisfy user needs to identify the number of suicides in specific intervals, we use a time series graph of ‘Suicides No’ against ‘Year’ to replace the function of the year filter while also showing the suicide trends to users.

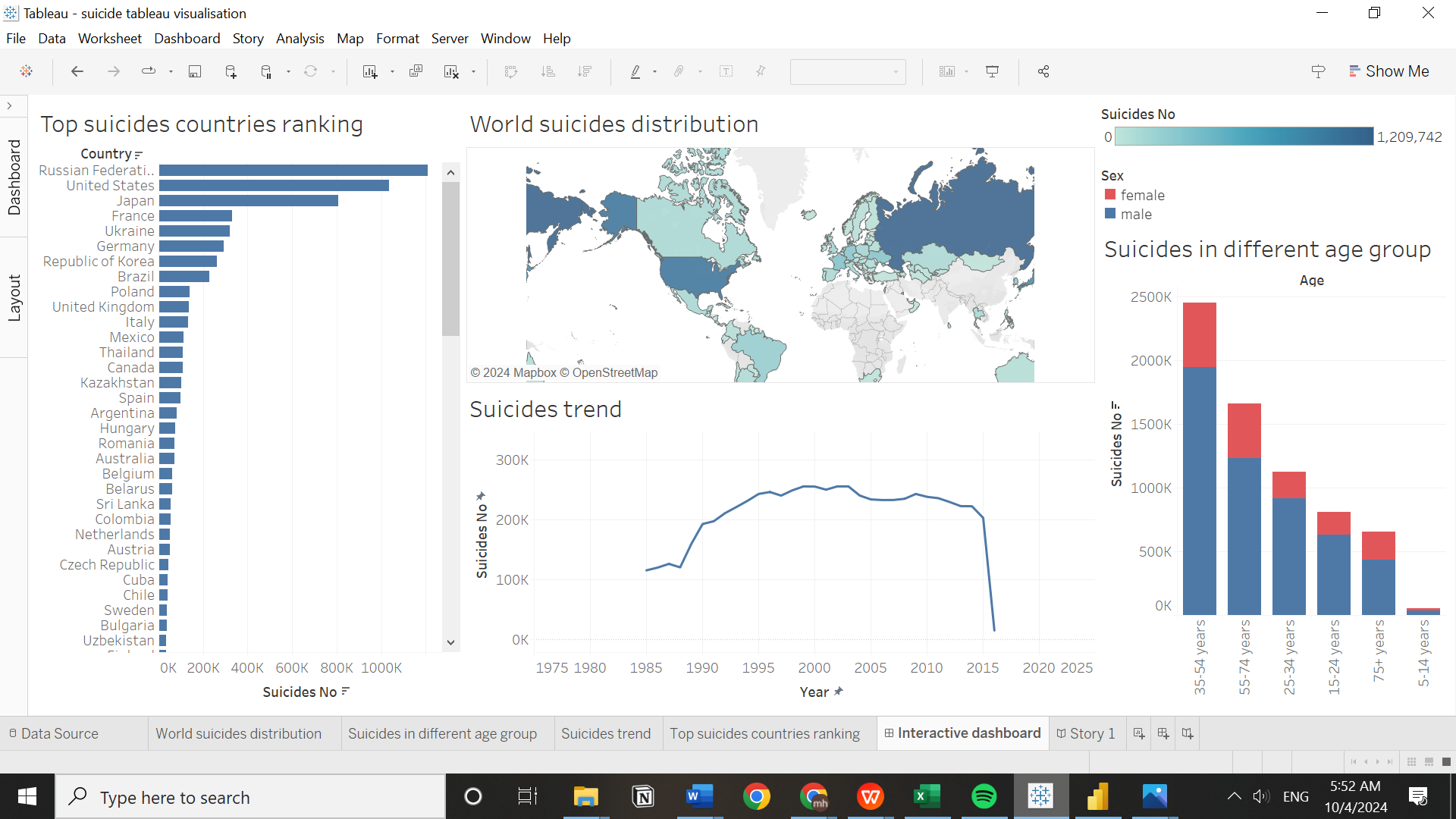


Figure 3.5 Completed Tableau dashboard

Finally, all charts are combined to form a dashboard. All charts are set to be used as filters to form an interactive dashboard. When users want to observe suicides in specific regions, groups or intervals, they can make selections in the chart, and other charts will be changed according to the selection.

**3.2 PoweBI Implementation**

The implementation in Power BI is similar to Tableau. The only difference is that Power BI implementation is started by building charts directly on the dashboard while Tableau is started by building charts in different worksheets and combined to form a dashboard.

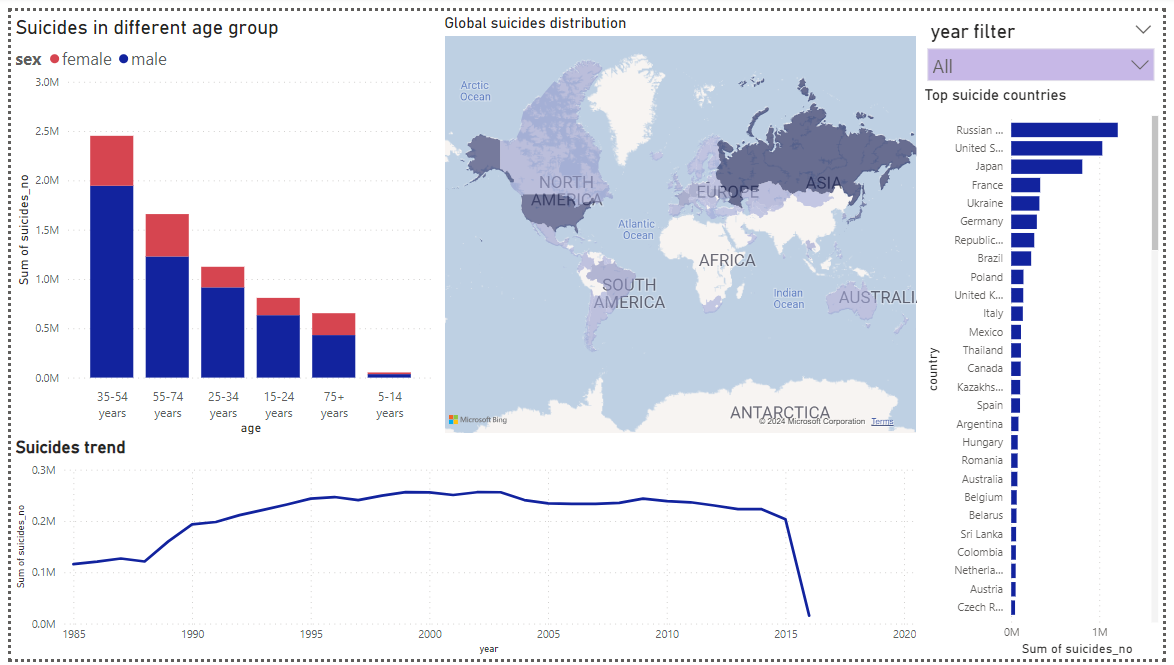


Figure 3.6 Completed Power BI dashboard

The completed Power BI dashboard is similar to the Tableau dashboard. There are differences in chart arrangements in the dashboard layout because Power BI allow more flexible chart arrangements in the dashboard compared to Tableau. There is an additional year filter at the top right corner because brushing and linking in Power BI only allows filtering selection in one chart at the same time. Unlike Tableau, which allows directly clicking on Japan from the map and the year 2014 from the suicide trend chart at the same time.

**3.3 Statistical modelling**

**3.3.1 Cluster analysis**

A cluster analysis is conducted in Tableau to explore whether the suicide no. can be classified into several groups according to population.

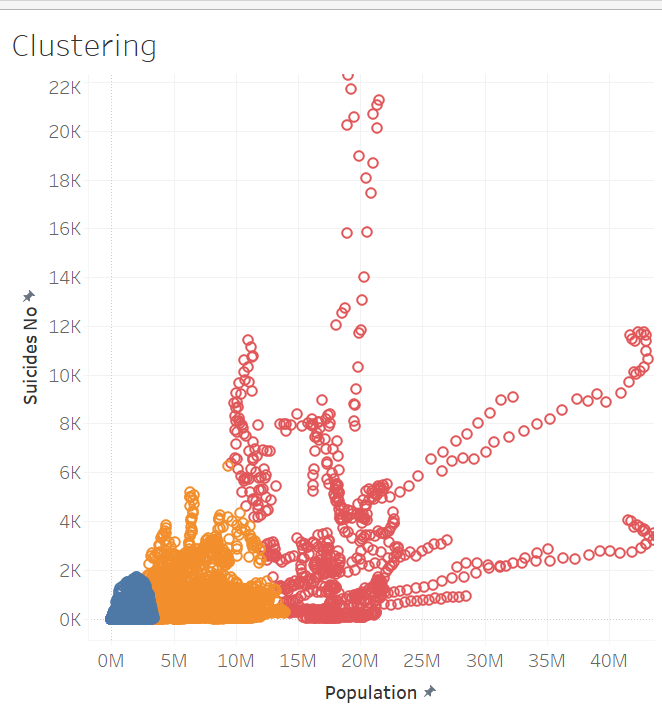


Figure 3.7 Clustering result

As shown in the figure, Tableau automatically classified suicide no. into 3 clusters and they can be interpreted as a cluster with a high suicide rate, a cluster with a medium suicide rate, and a cluster with a low suicide rate.

**3.3.2 Factor analysis**

A factor analysis is conducted in Tableau to explore underlying factors that contribute to the number of suicides.

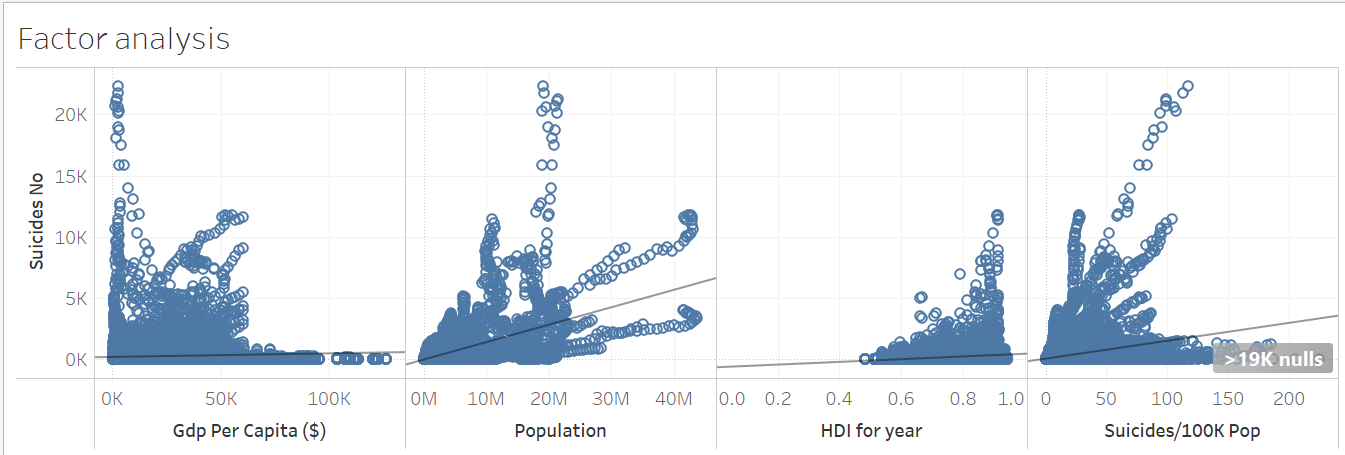


Figure 3.8 Correlation between ‘Suicides no.’ and numeric factors

A scatter plot with a trend line is used to explore the relationship between ‘Suicide.no’ against numeric factors. As a result, ‘Population’ and ‘Suicides/100K Pop’ show a strong positive relationship with ‘Suicides no’. ‘Gdp per Capita ($)’ and ‘HDI for year’ show a weak relationship with ‘Suicides no’.

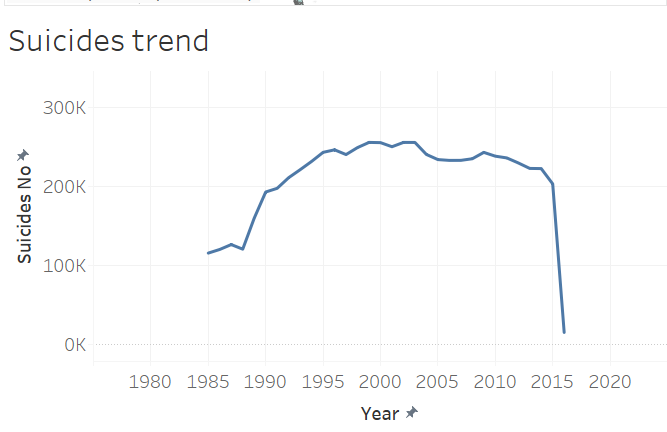


Figure 3.9 Suicide trend chart

From the time-series chart, we can infer that ‘Year’ shows a positive relationship with ‘Suicides.no’. The sudden drop at the end is due to the suicides record which was updated until early 2016.

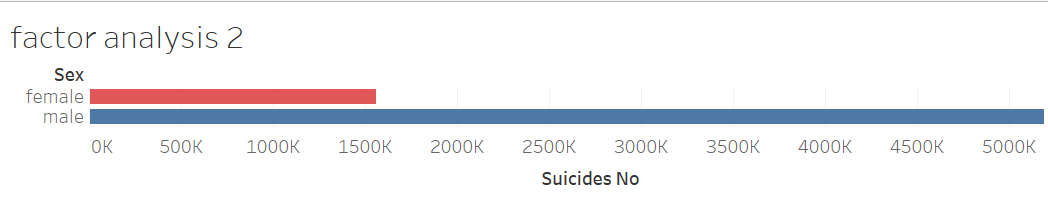


Figure 3.10 Comparison of Suicides.no between male and female

Categorical factors like ‘Countries’, ‘Sex’, and ‘Age’ also cause differences in ‘Suicides.no’. In the ‘Sex’ factor, male suicides are much greater than female suicides. In the ‘Countries’ factor, some countries like the Russian Federation and the United States have significantly high ‘Suicides.no’. In the ‘Age’ factor, certain age groups like 34-54 years have significantly high ‘Suicides.no’.

4.0 Walkthrough

**4.1 Tableau walkthrough**

1. What are the top 3 countries with the highest number of suicides?

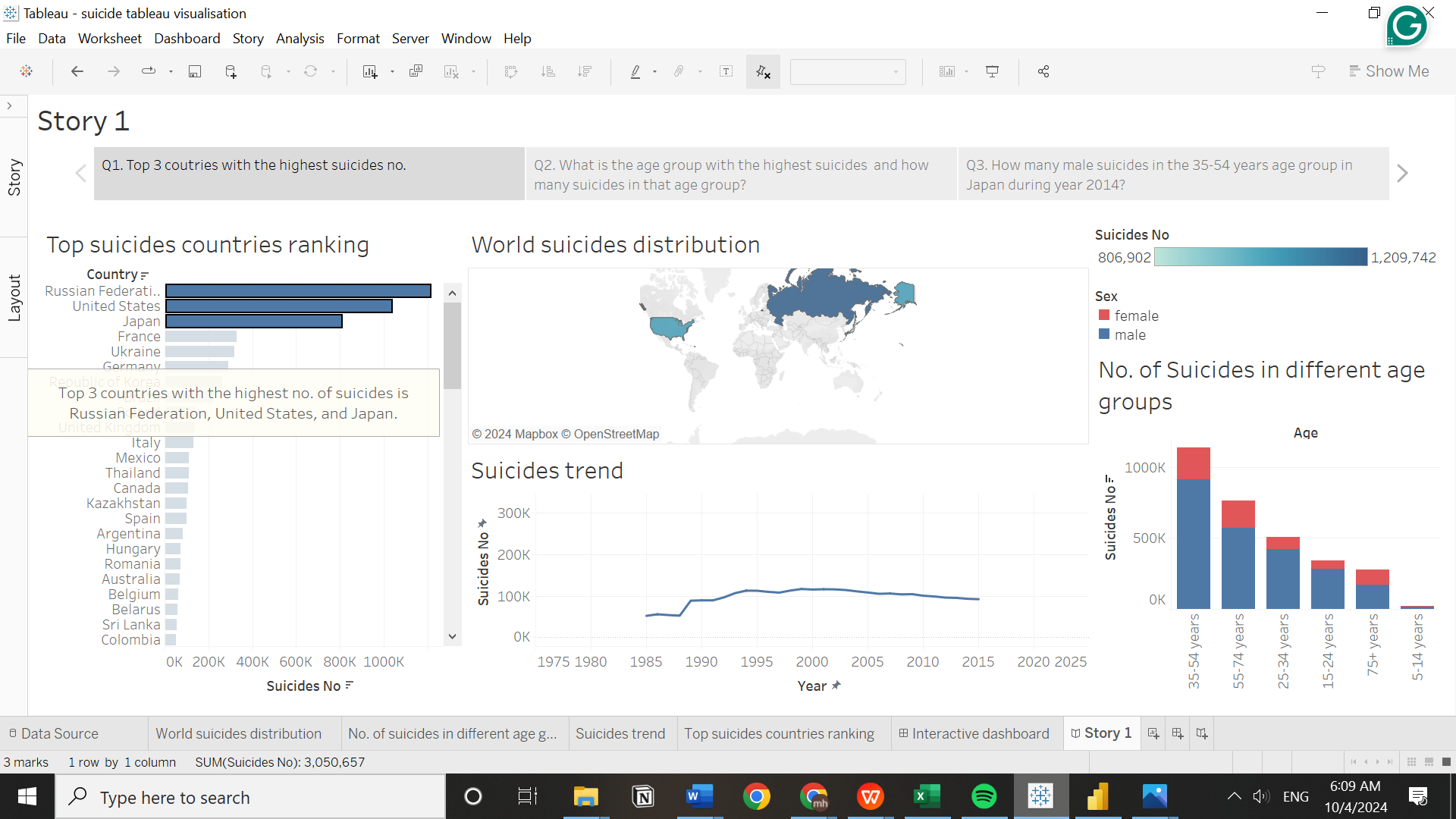


Figure 4.1 Tableau walkthrough in answering question 1

Question 1 could be answered by looking at the top suicide countries ranking chart. The top 3 countries with the highest suicide cases are the Russian Federation, the United States, and Japan.

2) Which age group have the highest number of suicides and how many suicides in that age group?

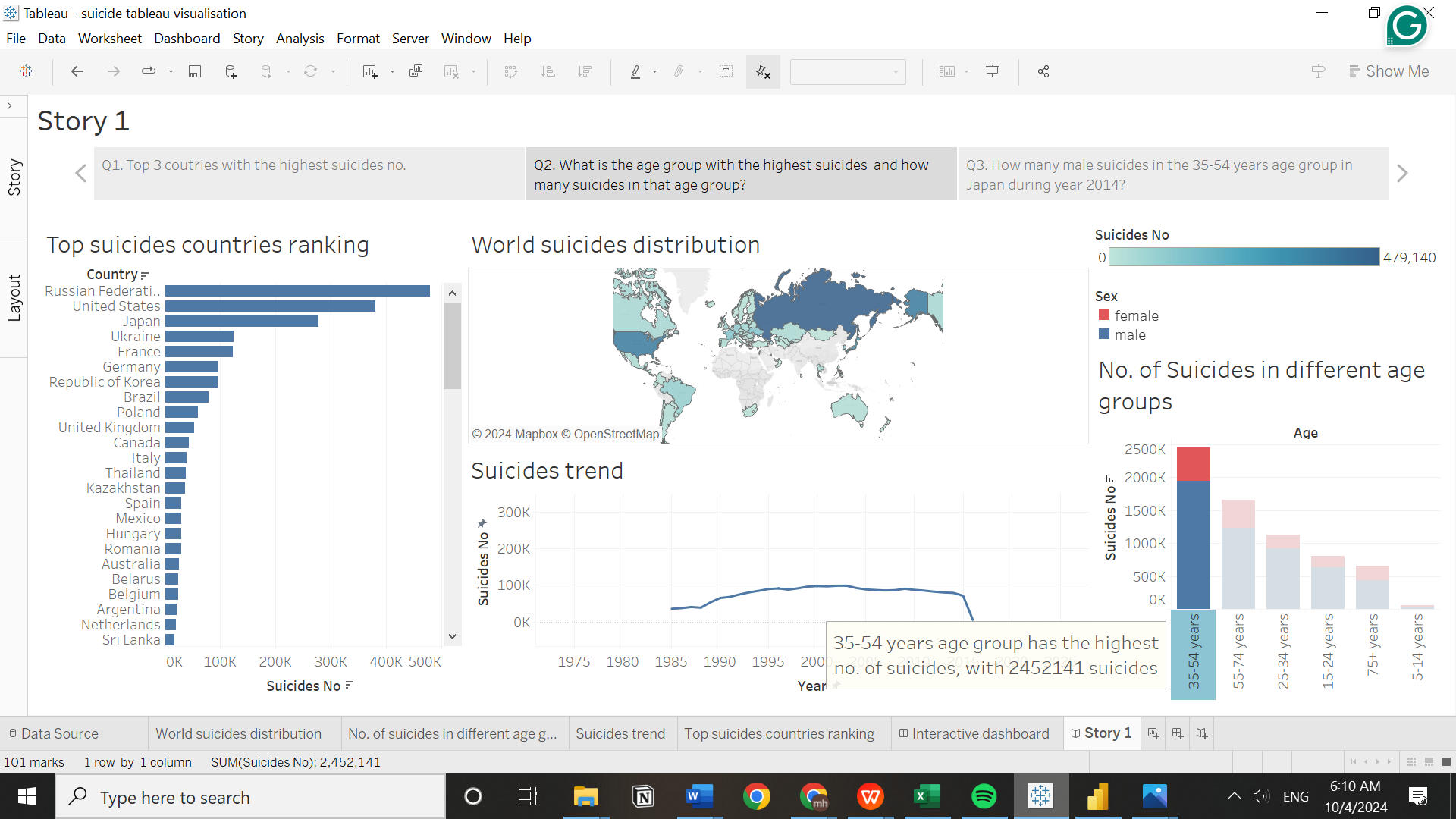


Figure 4.2 Tableau walkthrough to answer question 2

Question 2 could be answered by looking at the suicides age group chart. The age group with the highest suicides is 35-54 years, with a total of 2452141 suicides.

3. How many male suicides in the 35-54 years age group in Japan during year 2014?

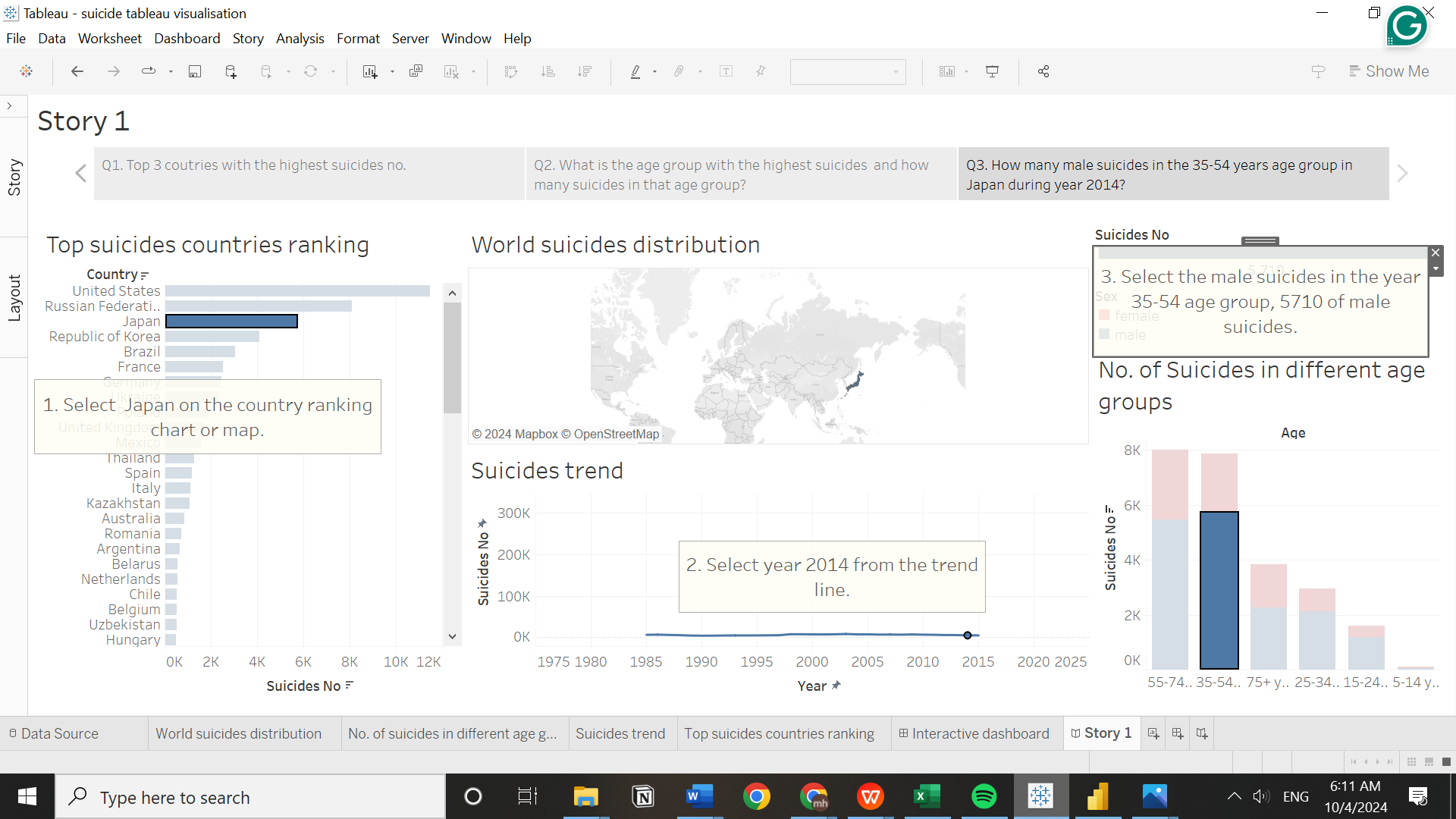


Figure 4.3 Tableau walkthrough to answer question 3

To answer question 3, the interactive dashboard is needed. From the suicide trending chart, the year 2014 is selected from the suicide trend line. Japan can be either selected from the density map or country ranking chart. From the age group chart, there are 5710 male suicides in the 35-54 years age group.

**4.2 Power BI Walkthrough**

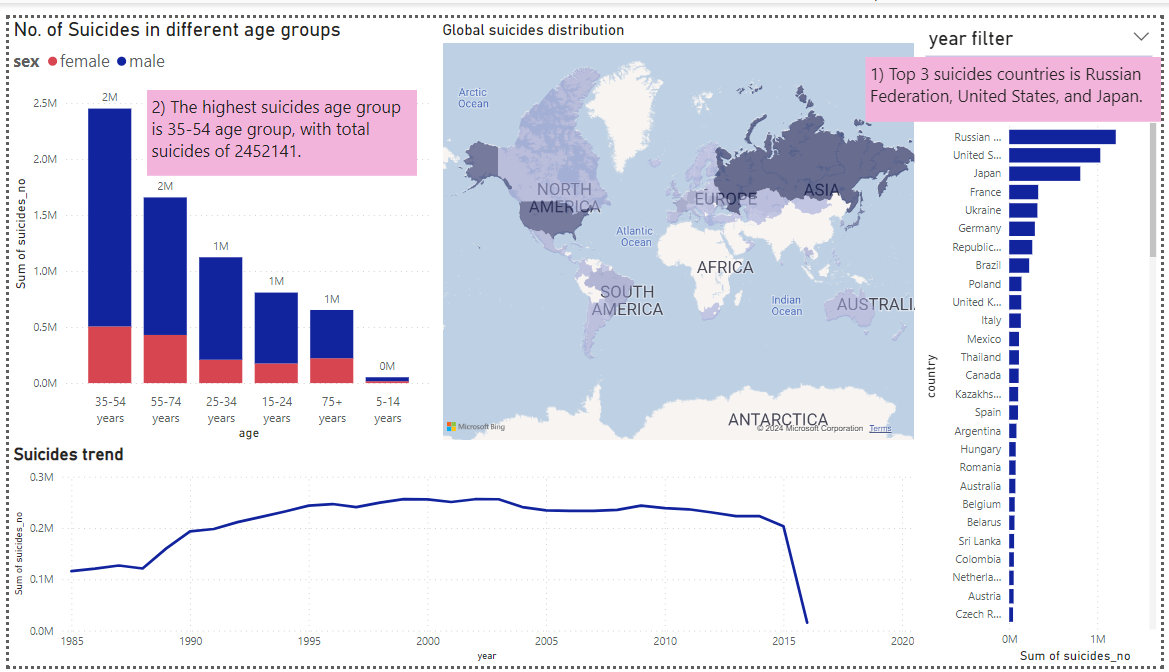


Figure 4.4 Power BI question 1 & 2 walkthrough

1) What are the top 3 countries with the highest number of suicides?

* Looking at the top suicide countries ranking, the top 3 suicide countries are the Russian Federation, the United States, and Japan.

2) Which age group have the highest number of suicides and how many suicides in that age group?

* Look at the suicides in different age groups, the highest suicides age group is 35-54 years, with total suicides of 2452141.

3) How many male suicides in the 35-54 age group in Japan during year 2014?

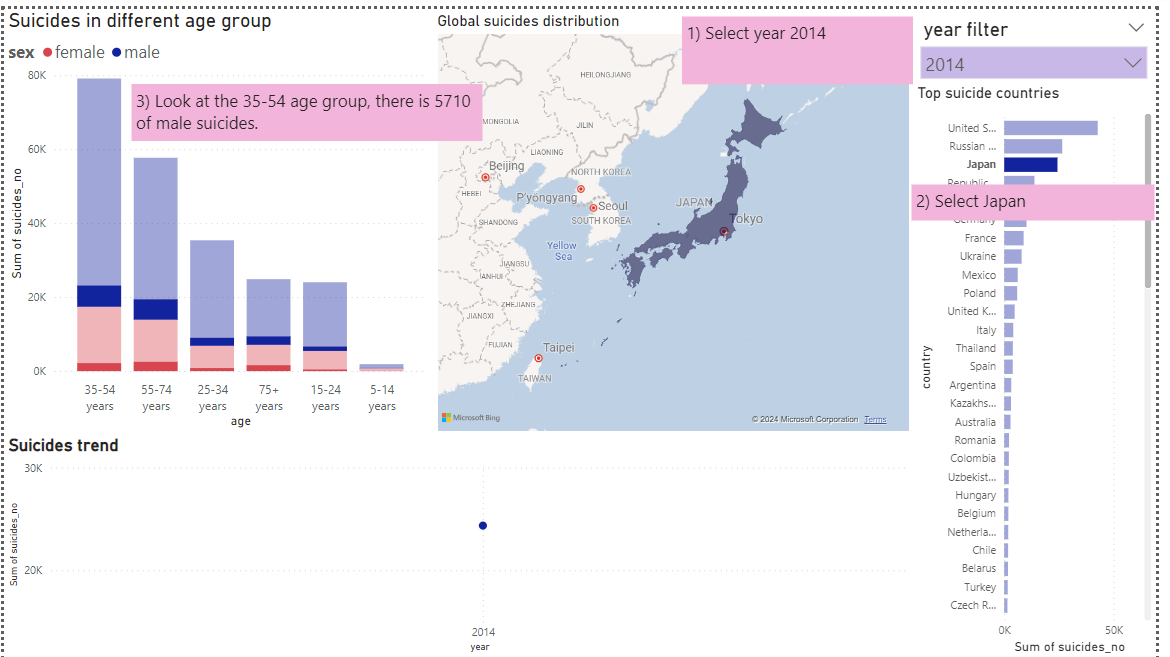


Figure 4.5 Power BI question 3 walkthrough

To answer question 3, select Japan from the country chart or map, and the year 2014 from the filter, then look at the 35-54 age group, there are 5710 male suicides.

5.0 Reflective Discussion

**5.1 Difference in design and actual implementation**

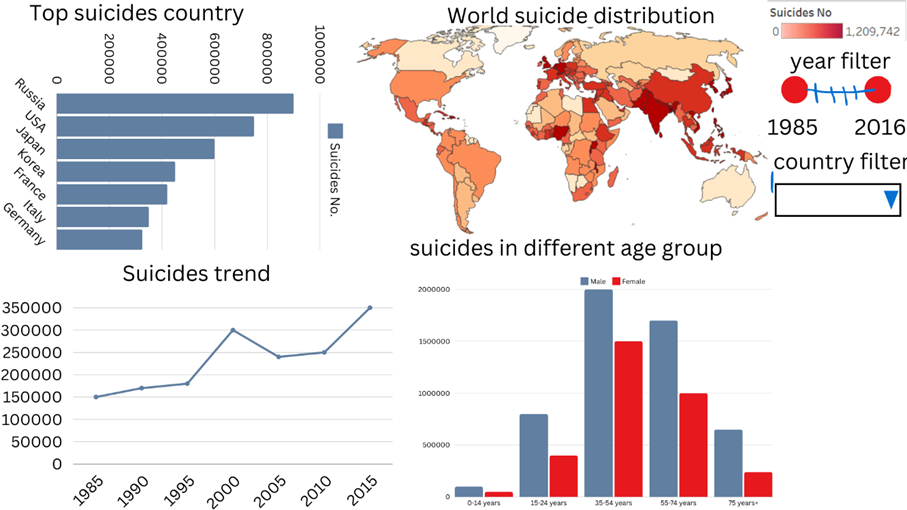
****

Figure 5.1 Paper landscape design

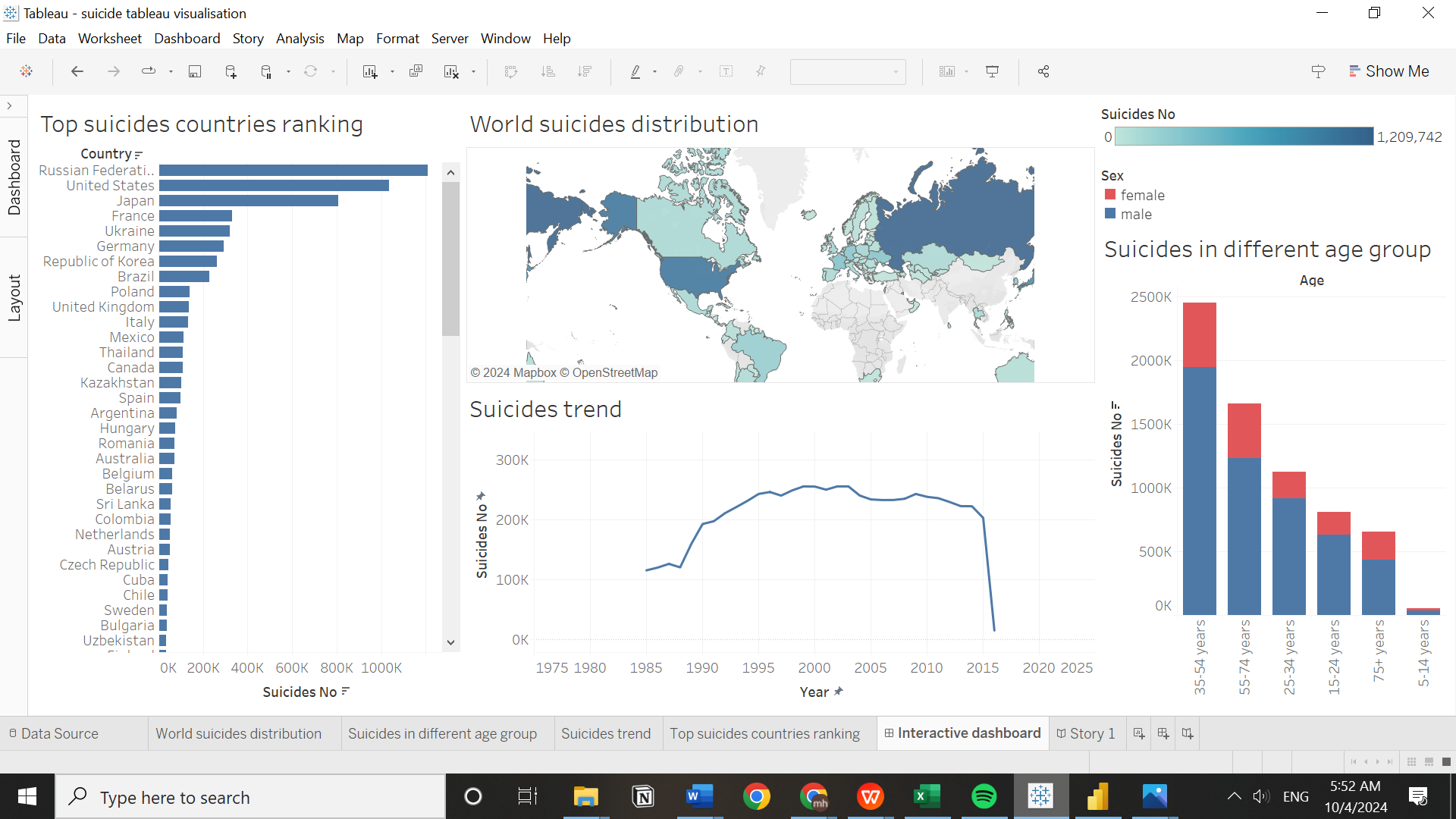


Figure 5.2 Actual Tableau implementation

There are differences between the paper landscape design and the actual tableau design. Firstly, the chart arrangement in the dashboard is changed to maximize the utilization of dashboard space and optimize the chart display. For instance, the country's ranking chart should be deep enough to better visualise the ranking of 101 different countries, so users don’t need extensive scrolling to see all countries. Another example is that the time-series chart of suicide trends should be long enough so users can click on each year from the trend line easily.

Besides that, the side-by-side bar chart of suicides in different age groups in design is changed to a stacked bar chart for better visualisation of total suicides in each age group.

Furthermore, unnecessary year filters and country filters are removed since their functions can be replaced by brushing and linking between charts. After setting up the interactive dashboard, specific countries can be assessed by clicking country on the map or countries ranking chart and specific years can be accessed by clicking year on the suicides trend line.

**5.2 Critical evaluation of visualisation tools**

As aforementioned, the global suicide visualisation is implemented on Tableau and Microsoft Power BI. Both visualisation tools are evaluated and compared in terms of flexibility, ease of implementation, and user experience.

In terms of flexibility, Power BI has higher dashboard layout flexibility than Tableau. Power BI doesn’t restrict the dashboard layout; any chart arrangements are allowed and this allows maximise utilization of the dashboard layout. Tableau has limited dashboard layout options; chart arrangements are sometimes limited. Therefore, the chart arrangement of global suicide visualisation in Powe BI implementation is different from Tableau implementation.

In terms of ease of implementation, Power BI implementation is more straightforward than that of Tableau. Power BI implementation starts by building every chart directly on the dashboard while Tableau implementation starts by creating charts in different worksheets and combining them on the dashboard. However, building charts is easier in Tableau due to the ‘Show me’ panel, Power BI have many complex functions which is not friendly to new users. Also, Tableau don’t have a density map option therefore customised setting of colour depth is required to generate the density map of global suicide distribution.

In terms of user experience, Power BI’s density map is easier to control compared to Tableau. For instance, Tableau doesn’t allow dragging operations in the density map of global suicides visualisation to observe interested regions. However, the dashboard theme in Tableau is more visually comfortable than Power BI.

Tableau and Microsoft Power BI have their unique advantages. In my personal opinion, Tableau is better.

**5.3 Applied principle of good data visualisation**

According to Tufte (2001), effective data visualisation should maximize the data-ink ratio, prioritize the data representation itself, focus on comparison, serve a clear purpose, and avoid distorting the data. Tufte (2001) explained maximization of the data-ink ratio is important to avoid distraction in data representation. For evidence, any visual encoding that didn’t convey information was removed, only colour encoding that conveyed the information of suicide density map and sex remained. Besides that, Tuft (2001) states that each visualisation should show a clear purpose to address the audience’s question. For example, the design of global suicide visualisation is guided by user requirements and successfully addresses every research question. Furthermore, Tufte (2001) also implies that data visualisation should focus on the comparison of data points to provide a key understanding to audiences. For instance, this work has focused on the comparison of suicides in different countries, sex, age group, and time series. As an addition, Tufte (2001) also prioritizes the data representation itself over decorative elements and avoids distorting the data. For instance, this work keeps the element as simple as possible while providing sufficient information to answer every research question correctly. Another good data visualization principle includes the use of suitable charts for different type of comparison (Few, 2004). For instance, this work uses a bar chart for value comparison between unordered groups like country, age group etc. and uses a line chart to perceive time-series suicide trends.

UCD (User-centered design) theory can significantly improve the user experience of data visualization projects through a systematic focus on the needs, context, and behaviour of end-users. The project should be started by user research to identify user goals, use cases, and how users can interact with the dashboard comfortably (Cooper, 2007). For example, identifying user personas and scenarios that represent the target user at the initial stage of the visualization project.

The HCI (Human-computer interaction) principle enhances the prototype design by guiding each decision in prototype implementation around usability, consistency, and user feedback. Benyon (2014) explained that usability focused on making prototypes easy to use, learn, and memorise. It guides the design to have minimal actions to reach customer needs. Consistency in design helps users to apply previous interactive experiences to a new system, making it easy to use (Benyon, 2014). The evaluation of user experience can be done through user participation, observing user interaction with the visualization projects and feedback collection from them (Wickramasinghe, 2014). Benyon (2014) emphasises the importance of user feedback in prototype design as they are clear responses from the user to improve the usability and user experience.

**5.4 Limitations**

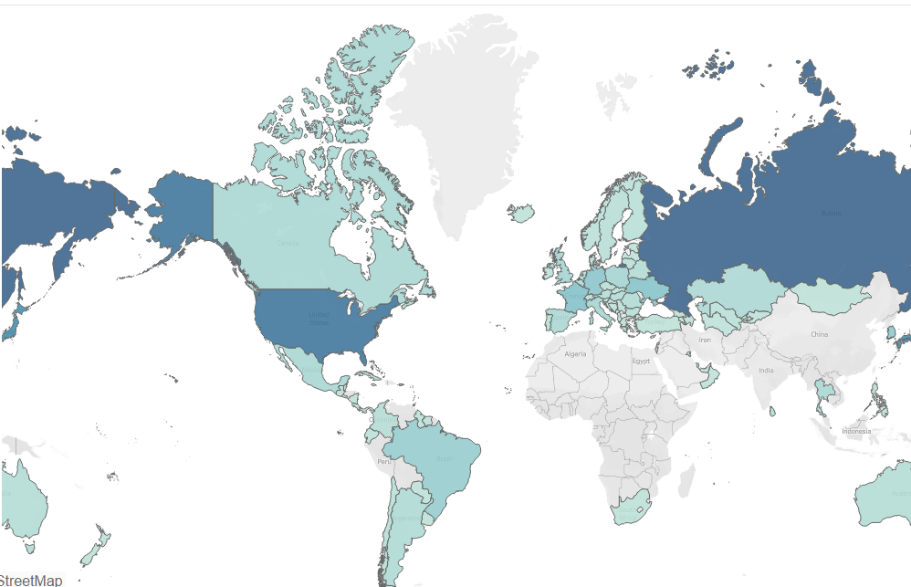


Figure 5.1 Density map of world suicide distribution.

The limitation of this work is regarding the constraint of the dataset. As shown in the density map, ‘Suicide Rates Overview 1985 to 2016’ didn’t cover most of the countries in East Asia and Africa. As a consequence, the user is unable to analyse suicides in those countries. Besides that, the dataset is quite outdated as it only covered suicide records until 2016. The user might not be able to analyse recent suicide trends.

**5.5 Personal learning**

This project shows the power of data visualisation in discovering meaningful patterns from complex data. It allowed me to turn theoretical knowledge from the lecture into practical context, significantly enhancing my data visualisation skills. Besides that, navigating through Tableau and Power BI provided me with invaluable hands-on experience in data visualisation software and this is very helpful for my future career in the data science field. Moving forward, I aim to deepen my hands-on experience and skills by trying more complex data visualisation projects like marketing visualisation, financial visualisation, disease visualisation etc, so I am readily prepared when a real-world project comes.

6.0 Conclusion

To satisfy user requirements in analysing global suicides, this project aims to design and implement a visualisation dashboard for generating value and insights from global suicide data. This project emphasized the importance of addressing users' needs by setting 3 research questions and user-specific requirements which guide the design process. To ensure the final implementation matches user requirements, a paper landscape design is drafted. Throughout the Tableau and Powe BI implementation process, dashboard design is guided by paper landscape design, data visualization principles learned from the lecture, user-centred design and human-computer interaction theory. With the use of interactive visualisation, both Tableau and Power BI dashboards successfully satisfy users' requirements and address three research questions within 3 clicks during the walkthrough process.

7.0 Reference

1) Lee, L., Roser, M. and Ortiz-Ospina, E. (2019). *Suicide*. [online] Our World in Data. Available at: https://ourworldindata.org/suicide.

2) Rusty (2018). *Suicide Rates Overview 1985 to 2016*. [online] www.kaggle.com. Available at: https://www.kaggle.com/datasets/russellyates88/suicide-rates-overview-1985-to-2016.

3) *WHO (2023) Suicide*, *World Health Organization*. Available at: https://www.who.int/news-room/fact-sheets/detail/suicide.

4) Tufte, E. R. (2001). "The Visual Display of Quantitative Information". Graphics Press, Cheshire, CT.

5) Benyon, D. (2014). *Designing interactive systems: a comprehensive guide to HCI and interaction design*. Harlow: Pearson.

6) Cooper, A., Reimann, R. and Cronin, D. (2007). *About Face 3*. John Wiley & Sons.

7) Wickramasinghe, B. (2020). *Human-Computer Interaction — Principles, Evaluation and Universal Design Principle*. [online] Medium. Available at: [https://bimalics.medium.com/human-computer-interaction-principles-evaluation-and-universal-design- principle3687123b5b2a#:~:text=The%20seven%20principles%20are%20equitable](https://bimalics.medium.com/human-computer-interaction-principles-evaluation-and-universal-design-%20%20principle3687123b5b2a#:~:text=The%20seven%20principles%20are%20equitable).

8) Few, S. (2004) Show Me the Numbers: Designing Tables and Graphs to Enlighten. Analytics Press. (Ch.5)