SUICIDE PREVENTION DASHBOARD

A PROJECT REPORT

Submitted by

Sudeepthi Rebbalapalli - 17BCE0955

CSE3020 – Data Visualization   (EPJ)
S l o t –   F1 / L55 + L56

Project Guide
Prof. Meenakshi P
Assistant Professor (Senior)
School of Computer Science and Engineering

Bachelor of Computer Science Engineering

IN

Vellore Institute of Technology

Winter Semester 2019-20

Table of Contents

1. INTRODUCTION

Every year there are hundreds of thousands of people take their own life and many more people who attempt suicide. Every suicide is a tragedy that affects families, communities and entire countries and has long-lasting effects on the people left behind. Suicide occurs throughout the lifespan and was the second leading cause of death among 15-29 year-olds globally in 2016. Suicide does not just occur in high-income countries, but is a global phenomenon in all regions of the world. In fact, over 79% of global suicides occurred in low- and middle-income countries in 2016. Suicide is a serious public health problem; however, suicides are preventable with timely, evidence-based and often low-cost interventions. For national responses to be effective, a comprehensive multisectoral suicide prevention strategy is needed. A suicide prevention dashboard will help in preventing suicides as it allows its users to identify major suicide trends. These trends will help families, police officers, teachers and others identify anyone around them that show these signs and stage an early intervention. In this project we make use of suicide data collected from various credible sources that includes information about the demographic data of people who have commited suicide over the past few years. After carefully analyzing the data and identifiying the semantics of the attributes, the data to be represented was selected and different idioms were proposed on how to best represent the data in a visual manner. The two most sutiable idioms were selected for the data choosen and two dashboards were created, one for each idiom. These dashboards were then validated by some test users who rated the them on the basis of whether the idioms are able to convey the intended message to the users and the level of clarity. Then, a comparison was done from this feedback and a conclusion was drawn about the effectiveness of the visual idioms for the suicide data and which visualization shows better results.

2. PROBLEM STATEMENT

Suicide is one of the leading causes of deaths around the world but suicide is also one of the few causes that can be prevented through timely intervention and assistance. To do this, familes and friends need to be able to identify who might be showing these depressive signs. This is not always obvious and by analyzing trends of suicide data we can conclude which individuals are at high risk. To do this, the idea of a suicide prevention dashboard is proposed. The dashboard can easily convey the risk factors and trends in people who commit suicide around around the world and can be understood by people who have do not have knowledge to perform data anaylsis. Here two dashboards of different visual idioms are proposed and implementation will be done using visualization tool Tableau. Validation will be done by test users and a comparision of results will be done. A conclusion will be achieved on how clearly the dashboards have managed to perform their duty based on the answers to these questions:

- What are the risk factors for suicide?

- What are the protective factors?

- What trends are obvious in the general case of suicides?

- Which countries are most vulnerable?

- During with time period was the suicide rate maximum?

- Is the trend of suicides increasing or decreasing?

- Which gender is more vulnerable?

- Which age groups are more vulnerable?

3. DATA COLLECTION

3.1 Method

The data being used in the creation of this dashboard is a secondary dataset whose data has been collected through government reports, statistical methods, surveys, and interviews. The dataset has 27805 cases of data. The data here choosen to be visualized is:

- Country of suicide
- Year of suicide
- Gender
- Age
- Population of country
- GDP per capita of country
- Generation

	A	B	C	D	E	F	G	H	I	J
1	country	year	sex	age	suicides_no	population	suicides/100k pop	country-year	gdp_per_capita ($)	generation
2	Albania	1987	male	15-24 years	21	312900	6.71	Albania1987	796	Generation X
3	Albania	1987	male	35-54 years	16	308000	5.19	Albania1987	796	Silent
4	Albania	1987	female	15-24 years	14	289700	4.83	Albania1987	796	Generation X
5	Albania	1987	male	75+ years	1	21800	4.59	Albania1987	796	G.I. Generation
6	Albania	1987	male	25-34 years	9	274300	3.28	Albania1987	796	Boomers
7	Albania	1987	female	75+ years	1	35600	2.81	Albania1987	796	G.I. Generation
8	Albania	1987	female	35-54 years	6	278800	2.15	Albania1987	796	Silent
9	Albania	1987	female	25-34 years	4	257200	1.56	Albania1987	796	Boomers
10	Albania	1987	male	55-74 years	1	137500	0.73	Albania1987	796	G.I. Generation
11	Albania	1987	female	5-14 years	0	311000	0	Albania1987	796	Generation X
12	Albania	1987	female	55-74 years	0	144600	0	Albania1987	796	G.I. Generation
13	Albania	1987	male	5-14 years	0	338200	0	Albania1987	796	Generation X
14	Albania	1988	female	75+ years	2	36400	5.49	Albania1988	769	G.I. Generation
15	Albania	1988	male	15-24 years	17	319200	5.33	Albania1988	769	Generation X
16	Albania	1988	male	75+ years	1	22300	4.48	Albania1988	769	G.I. Generation

3.2 Attribute Semantics

Here we have identified the attribute semantics of each of the attributes of the dataset we have choosen, so that we will be able to chose the right visual representation to represent the data.

Country of suicide – **Categorical, Spatial**

Year of suicide - **Nominal, Temporal**

Gender- **Categorical**

Age – **Diverging**

Population of country – **Ordered, Ratio**

GDP per capita of country - **Ordered, Ratio**

Generation - **Categorical, Temporal**

Now that we know what type of data we are dealing with, it is possible to think of the different ways each attribute can be represented.

3.3 Data Representation

Based on the semantics on the attributes, some the proposed visual representations for the data are:

1. Bar charts comparing suicides between the two genders over the years
2. Map of the world to visualize suicide rate over the world
3. Line graphs representing the trend of suicides over the years
4. Pie charts representing the age groups
5. Pie chart representing gender ratio
6. Bar charts representing suicide by country
7. World heat map to visualize suicide density
8. Histogram for frequency of suicides

We can choose the most suitable representations based on the target audience and design visual idioms for the dashboards.

3.4 Target Identification

The Suicide rate dashboard is used to track trend of suicides and is used to find signals

correlated to increased suicide rates among different cohorts globally, across the socio-economic spectrum. Some of the target users of this application would be:

• Police

• Hospitals/Therapists

• Newspapers/Journalists

• Schools/Parents

4. TASKS AND ACTIONS

We need to translate domain-specific situations into abstract visualization tasks in terms of tasks and actions.

- Present the data to enjoy and consume
- Record the outliers of each attribute
- Derive the dependencies of the attributes among each other
- Identify trends of each attribute
- Arrange the features in order of most to least important
- Filter the extremes to get the range
- Compare the features of the attributes
- Search for unseen trends and relationships
- Summarize the dependencies and trends

5. IMPLEMENTATION

Here we have used Tableau Desktop for Students to implement the visual idioms

5.1 Impmentation of Visual Idiom 1

The first dashboard implements the following visual idiom:

1. Map that shows the suicide density around the world

2. Bar Graph that shows suicides per gender per country

3. Pie Chart that shows the ratio of suicides per age group

4. Bar Graph that shows the relation between population and suicides in a country

5. Heat Map that shows the relation between suicide density and GDP of a country

6. Bar Graph that shows the suicides by the generation of birth

7. Line graph that shows number of suicides every year globally.

Suicide Rate DashBoard 1
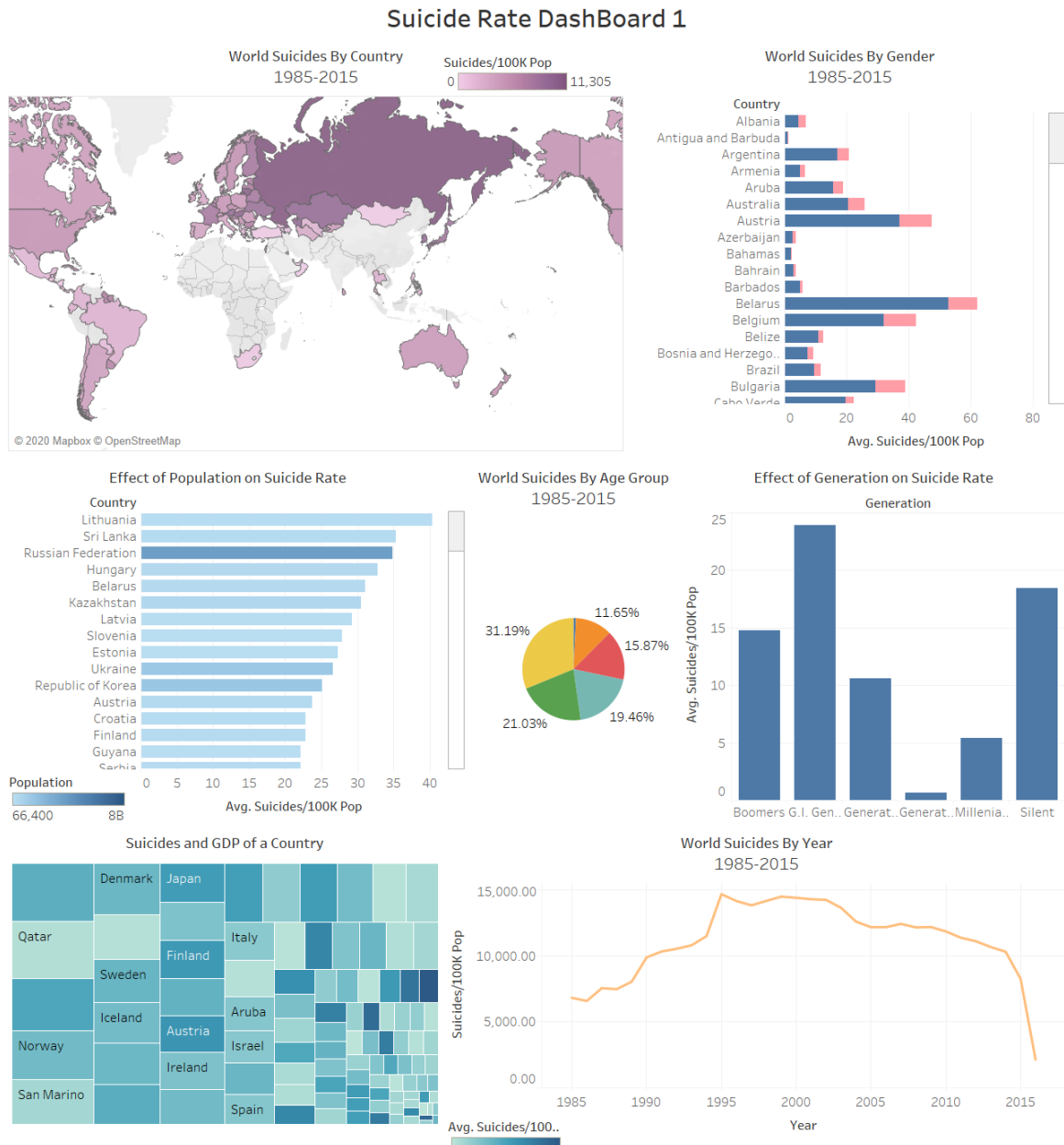
5.2 Implementaion of Visual Idiom 2

The second dashboard implements the following visual idiom:

1. Line graph that shows the suicide rate based on age
2. Bar Graph that shows rate of suicides per country
3. Pie Chart that shows the ratio of suicides per gender
4. Bubble Chart that shows the relation between suicide rate and GDP of a country
5. Heat Map that shows the relation between suicide density and population of a country
6. Area curve that shows the number of suicides every year globally
7. Line graph that shows effect of generation on suicides

Suicide Rate Dashboard 2
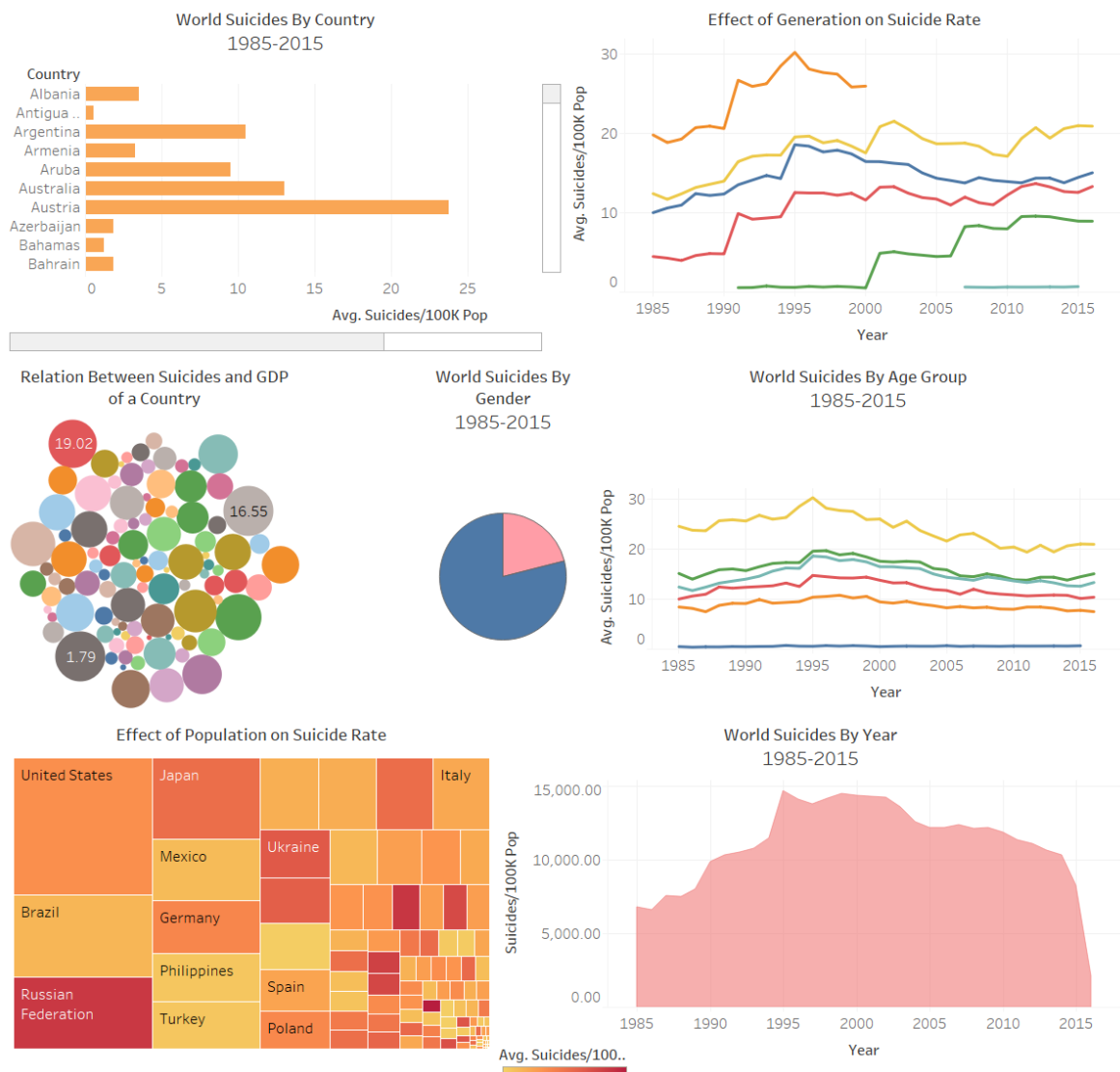
6. VALIDATION AND COMPARISION

By asking test users (some of the target audience) to use and rate their experience, we perform the four levels of validation on each of the dashboards, visual idiom 1 and visual idiom 2. We have used a rating system here where users can rate from between 1-10 with 1 being meaning unsuitable and 10 being the most suitable .

6.1 Domain Validiation

 The major threat in this phase is that the problem is mischarectrized and the target users do not have this need. The users offered a rating on whether this dashboard is useful them and if it has answers to the possible general questions in they have their head about suicides. This was the same for both the dashboards as the domain situation for both of them was the same. **Rating:** 7

6.2 Abstract/Data Validiation

The major threat here is that the identified task abstraction and the designed data abstraction do not solve the characterized problems. The users were given the initial set of questions that the dashboard proposed to answer and rate how clearly the dashboard convyed the answer for each question. The below table shows the rating for both dashboards:

Problems	Visual Idiom 1	Visual Idiom 2
What are the risk factors for suicide?	6	7
What are the protective factors?	5	4
What trends are obvious in the general case of suicides?	8	7

| During with time period was the suicide rate maximum?	10	10
Is the trend of suicides increasing or decreasing?	9	10
Which countries are most vulnerable?	9	8
During with time period was the suicide rate maximum?	10	10
Which gender is more vulnerable?	8	8

6.3 Visual/Idiom Validiation

The major threat here is that the chosen idioms are not effective at communicating the desired abstraction to the users. Here the users rated whether each graph of each dashboard was actually effective at the conveying the data it was visually representing and how clearly they understood the visualization.

Visual Idiom/Dashboard 1:
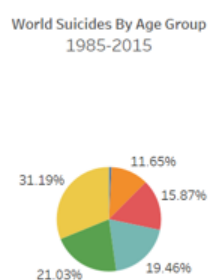
Graph	Rating
Map that shows the suicide density around the world	9
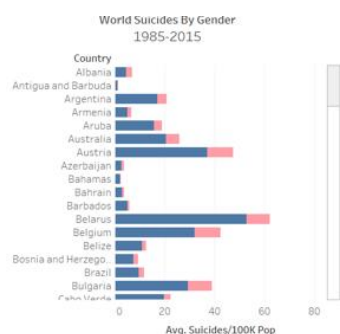
Bar Graph that shows suicides per gender per country World Suicides By Gender 1985-2015	**8**
Pie Chart that shows the ratio of suicides per age group World Suicides By Age Group 1985-2015	**9**
Bar Graph shows the relation between population and suicides in a country Effect of Population on Suicide Rate	**9**
Heat Mapshows the relation between suicide density and GDP of a country Suicides and GDP of a Country	**7**
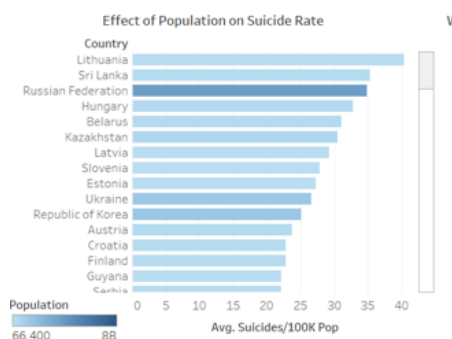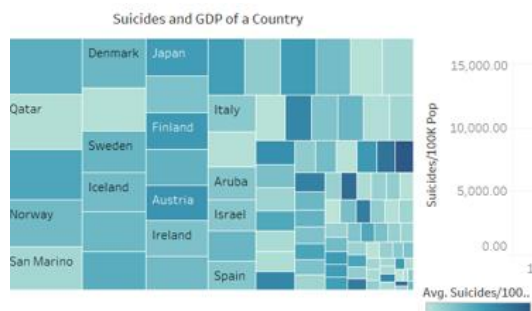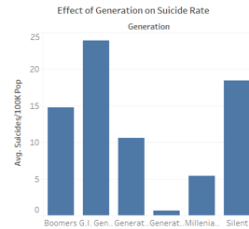
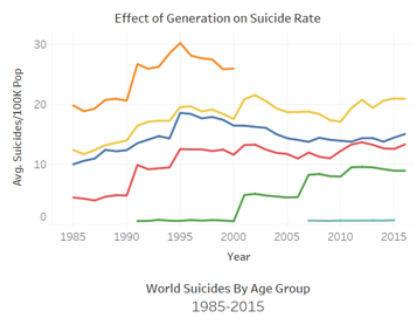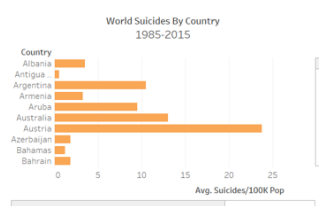| Graph	Rating
Bar Graph that shows the suicides by the generation of birth	8
Line graph that shows number of suicides every year globally.	9

Visual Idiom/Dashboard 2:

Graph	Rating
Line graph that shows the suicide rate based on age	7
Bar Graph that shows rate of suicides per country	9

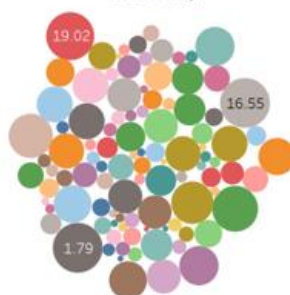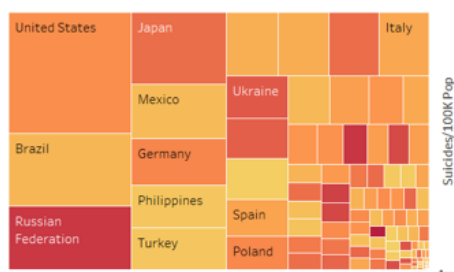|	
Pie Chart that shows the ratio of suicides per gender World Suicides By Gender 1985-2015	8
Bubble Chart shows the relation between suicide rate and GDP of a country Relation Between Suicides and GDP of a Country 19.02　16.55　1.79	10
Heat Map shows relation between suicide density and population of a country Effect of Population on Suicide Rate United States　Japan　Italy Mexico　Ukraine Brazil　Germany Philippines　Spain Russian Federation　Turkey　Poland Suicides/100K Pop	10
Area curve that shows the number of suicides every year globally World Suicides By Year 1985-2015 15,000.00 10,000.00 5,000.00 0.00 1985　1990　1995　2000　2005　2010　2015 Suicides/100　Year	10

Line graph that shows effect of generation on suicides	
World Suicides By Age Group 1985-2015	**8**
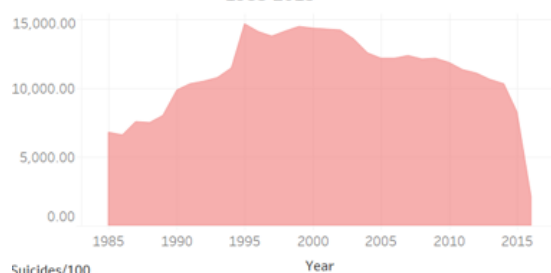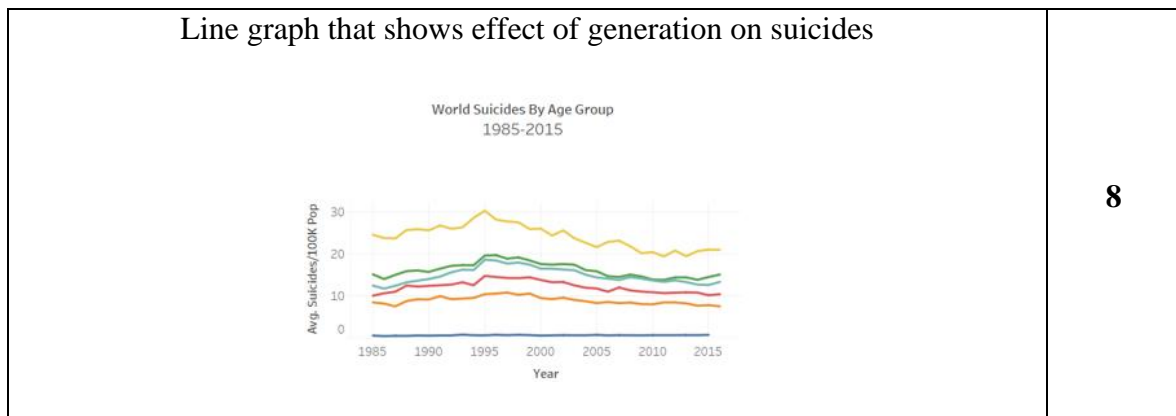
6.4 **Algorithm Validiation**

The major threat here is that the algorithm performance is suboptimal. Here the users ranked the speed and ease of use of the application. Since both the dasboards were constructed in Tableau which is a powerful visualization tool and very easy to use and user-friendly they both saw the same excellent results for algorithm validation. **Rating: 10**

6.5 Comparison

Based on the validations done by the users of the dashboards, we can now compare them to draw a conclusion on which is better

Validation Type	Visual Idiom 1 Average Rating	Visual Idiom 2 Average Rating
Domain	7	7
Abstract/Data	8.125	7.75
Visual/Idiom	8.412	8.857
Algorithm	10	10
Final Average of Validation	**8.384**	**8.401**

From these results obtained from comparing the validations of Visual Idiom 1 and visual Idiom 2, we can draw certain conclusions. The first visual idiom performs much better at achieving the purpose of the dashboard, i.e. it conveys to the user all the answers to questions that the dashboard proposed to answer. It also does significantly choosing the visual

representations to clearly convey the data. This means the graphs chosen were able to convey the right information but could have been more aesthetically pleasing. The second visual idiom performs average in answering the right questions and fulfilling the purpose of the dashboard. But the visualizations chosen in this idiom were able to reach the users better both aesthetically and information-wise. This means the graphs used conveyed the data well but they were not the best for the purpose. But from the overall rating we can say that the visual idiom 2 slightly outperforms the visual idiom 1.

7. CONCLUSION

The objective of this project was to create a suicide prevention dashboard that can help familes and friends identify high risk patients and prevent suicides by early intervention. The data collected for visiualization contained various trends among people who have commited suicide and their demographic data. The idea was to construct a dashboard that can easily convey the risk factors and trends in people who commit suicide around around the world and can be understood by people who have do not have knowledge to perform data anaylsis. Based on the analyzed data the two best visual idioms were proposed and task abstraction was done. These idioms were implemented using the visualization tool Tableau and two dashboards were created. These dashboards were then used by a portion of the target audience who rated them in various categories like usefulness, effectiveness, visually pleasing, performance etc. and these results were used in the validation. The validations of both the dashboards were compared and the conclusion was that while the first dashboard was able to serve the purpose better and conveyed information that it was meant to, the users preferred the second dashboard. Even though the second dashboard was not able to fully convey the information, it was simpler and more pleasing to eye. Therefore in the validation, the second visual idiom did slightly better. Finally, from these results we can conclude that the ideal visual idiom would be one where we can combine the best features of both of the above proposed idioms to make a single one which is simple, visually pleasing and accurate.

REFERENCES

[1] United Nations Development Program. (2018). Human development index (HDI). Retrieved from http://hdr.undp.org/en/indicators/137506

[2] World Bank. (2018). World development indicators: GDP (current US$) by country:1985 to 2016. Retrieved from http://databank.worldbank.org/data/source/world-development-indicators# [Szamil]. (2017).

[3] Suicide in the Twenty-First Century [dataset]. Retrieved from https://www.kaggle.com/szamil/suicide-in-the-twenty-first-century/notebook

[4] World Health Organization. (2018). Suicide prevention. Retrieved from http://www.who.int/mental_health/suicide-prevention/en/

[5] https://www.tutorialspoint.com/tableau/index.htm

[6] https://www.sisense.com/blog/10-useful-ways-visualize-data-examples/