Visualizing the burns age-adjusted mortality in the world

Srikanth Gadicherla[1]
srikanth.gadicherla@aalto.fi, 546195.

The project assignment deals with visualizing the deaths due to burns/fire accidents in the year 2004 and checks the trend in the years 2002 and 2004 across all the countries in the world. The data age-adjusted mortality due to burns was visualized in the form of heat map, bar graph and area plot. The pros and cons of each method is discussed in detail.

I. THE DATA

The data was downloaded from `gapminder.org`. The data was collected and surveyed by *WHO* and *Global Burden of Disease*. The data was provided in a tabular form with the variables `Country`, `Y2002`, `Y2004`, where the second and third columns represent the age adjusted mortality rate in the respective years. The data has 193 countries and this assignment only considers the mortality rate in the year 2004. Sample data is presented in Table I.

TABLE I. Burns, age adjusted mortality per 100 000 standard population

Country	Y2002	Y2004
Afghanistan	5.460	5.120
Albania	0.271	0.485

Age adjusted rate is done for fairer comparisons between groups of different ages.

II. HYPOTHESIS

The aim of this endeavour is to prove that *"that the burns age-adjusted mortality is more in the developing countries than in developed countries and check the trend"*.

III. APPROACH TO VISUALIZE THE DATA

The first approach to prove the hypothesis is to visualize the data on the heat map and bar plot to get a sense of which countries have more mortality than others. To check the trend of mortality rate in the world the heat map on world map and area plot were drawn.

IV. MUNZNER'S MODEL

A. Domain

This visualization is aimed at social activist who can understand the current situation and take necessary steps for creating awareness among the people. This report could help in taking steps in increasing security of the concerned citizens of respective countries, so may also be targeted to all concerned politicians in the world.

B. Abstraction

This part of the Munzner's model answers the What and Why of the visualization. What data is being visualized and why the audience is viewing it?

1. Data Abstraction

The data at hand is a static tabular quantitative continuous data. One important aspect of the data is that it has geographical spatial location information in the form of country names.

2. Task Abstraction

Primary objective of visualizing this data is *to find the severity of the deaths due to burns and fire accidents.* Then later, also to check the trend in the deaths. First, we have to *produce distribution*, next we have to *compare trends*.

C. Description of Visual Idioms used

A visual idiom is a combination of marks and visual variables. Marks are basic units with which any visualization can be constructed. The visual variables are the properties of the marks which could be used to represent them. Below are given the visual idioms with their description.

1. Heat map

The marks is the *area* with the visual variable being the *color* using the quantitative attribute. The area is decided by the geographical area of the country. The color palette used was R colors brewer's *'YlOrRd'*, this
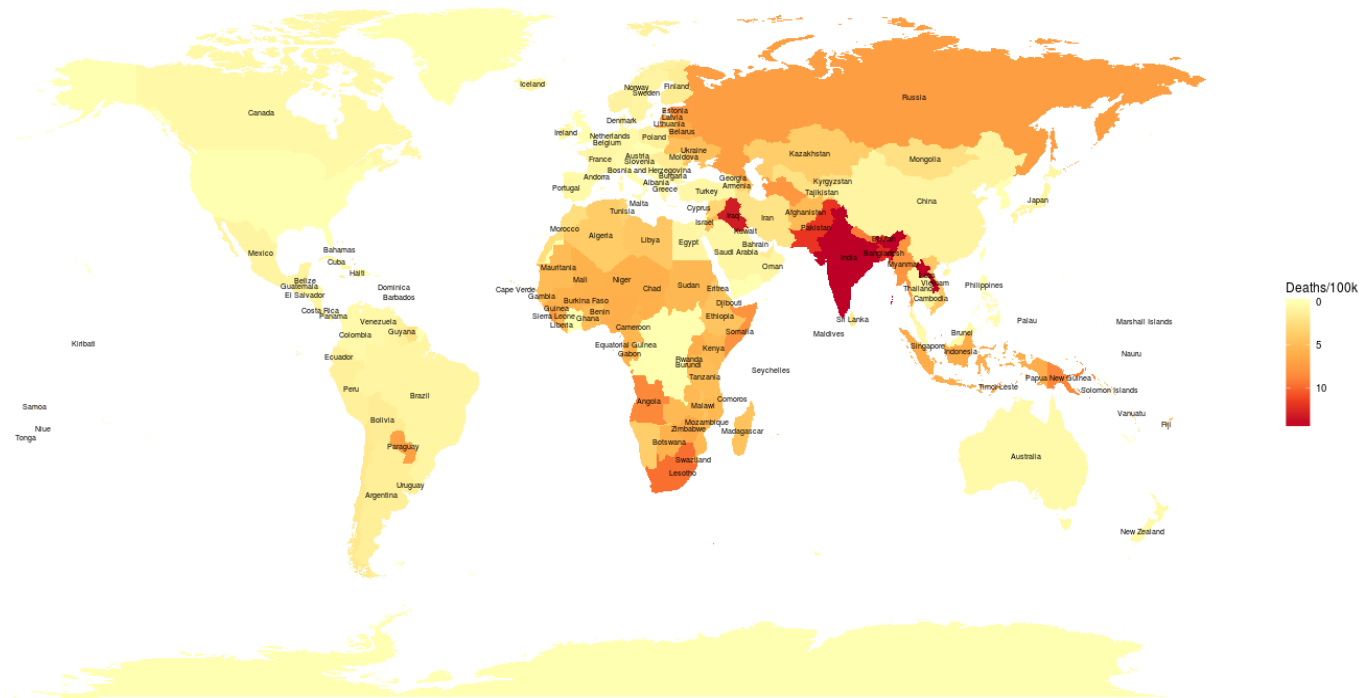
FIG. 1. Heatmap of burns age-adjusted mortality rate in 2004

palette has colour spectrum from yellow to red. The FIG. 1 shows the heat map of the 2004 year data.

For comparing trends, the palette used was 'RdGy' which goes from negative to positive values, useful for visualizing the change in the values over the years. FIG. 3 shows the heat map for the trend. The palette goes from black(high negative), white(neutral) to red(high positive).

2. Bar plot

The marks is the *line* here with the visual variable being the *length* of the quantitative variable. In the FIG. 2 where the bar plot is given, shows the data for the years 2002 and 2004.

The interactive version of the plot can be found at `https://goo.gl/VGXE3T`

3. Area plot

The marks in the area plot are the *lines* and the visual variable being the *length* of the quantitative variable. FIG. 4 shows the Area plot for the data 2002 and 2004.

The interactive version of the plot can be found at `https://goo.gl/YNcjxR`

V. TUFTE'S PRINCIPLES

The Tufte's principles were strictly followed and the data was showed without any distortion.

1. Heat map

In the heat map on the world map, the grid and bounding box were avoided and also the background was left white instead of filling any color maximizing the data ink. The labels were put to only those countries given in the data so as to reduce the non-data ink.

2. Bar plot and Area plot

The bar plot had also the grid and bounding box around the image being removed maximizing the data ink. The x-axis had the country names as the tick labels and y-axis had the mortality rate.

VI. GESTALT LAWS

The Gestalt laws used in the visualization are:

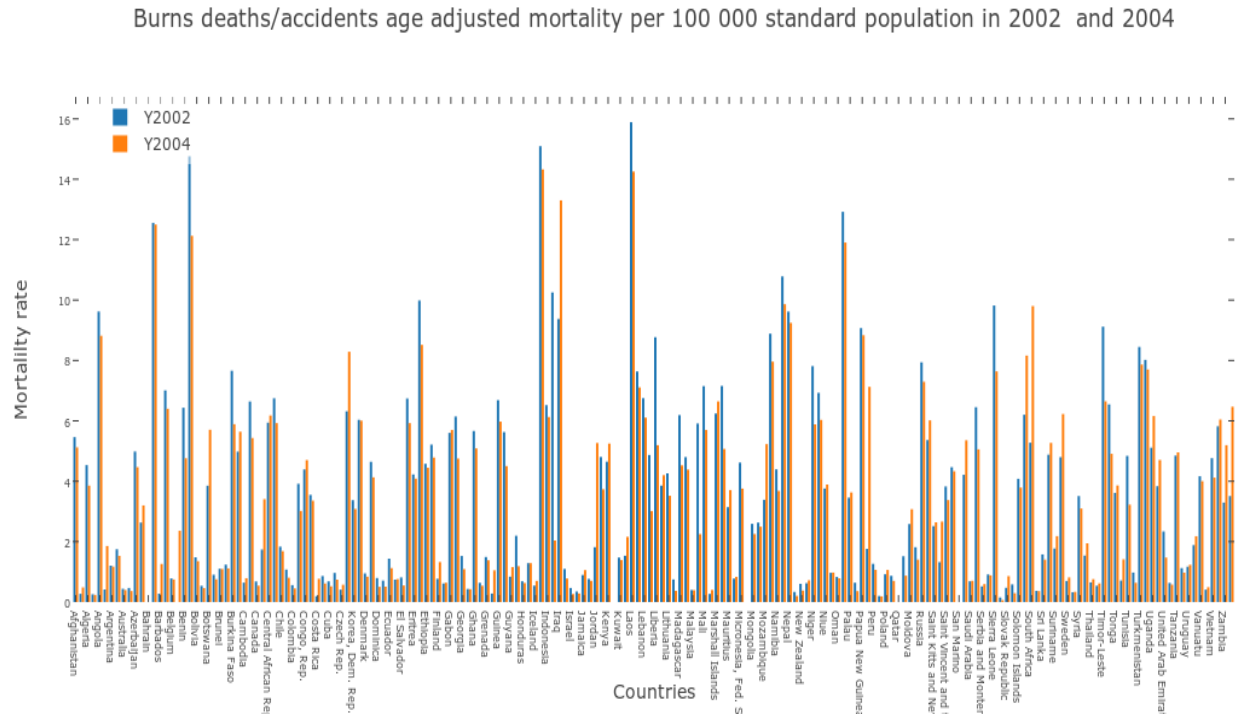*Proximity*: In the heat map, the countries which are

FIG. 2. Bar plot of burns age-adjusted mortality rate in 2002 and 2004

near belong to same continent and might show some relation in the analysis.

Similarity: In all the plots, the regions/lines with the same color are related to each other. In the heat map, the countries with the same country have the same death rate, while in the bar/area plot the lines with the same color are from the particular year, here, either from 2004 or 2002.

Connectedness: In the area plot, the trend for different countries for a particular year is connected with a line showing a relation.

Closure: In the heat map, the geographical nature of the figure gives all the countries the closed shape conforming to the closure property.

VII. METHODS USED FOR ANALYZING THE DATA

The programming software *R* was used for ploting the heat map on the world map. The packages mainly used for accomplishing the task was *ggplot2* and *ggmap*.

The bar plot and area plot were plotted online software *Plotly*.

VIII. RATIONALE/JUSTIFICATION FOR SELECTING THE VISUALIZATIONS

For visualizing the mortality rate is more in developing countries, it was evidently obvious to plot the mortality on world map to see which countries had more deaths due to burns/fire accidents.

For visualizing the trend in the years 2002 and 2004, it was thought appropriate to plot the heat map and area plot for the two years. With the scalability of area plot to hundreds of data points it was a sensible choice.

The heat map was plotted with the difference of mortality rates from 2004 to 2002. The area plot would have been appropriate for a time-series data, but here to compare difference in the values of the two years it was plotted. For each country the mortality rate value for that year was plotted and line was drawn for all the countries.

Similarly, it was done for the next year as well. If for a country, one of the colors is dominating the other, then we could say a change in trend either increase or decrease based on the color. This is not so good way for visualizing trend as the user would not get the idea behind the plot instead heat map should be considered.

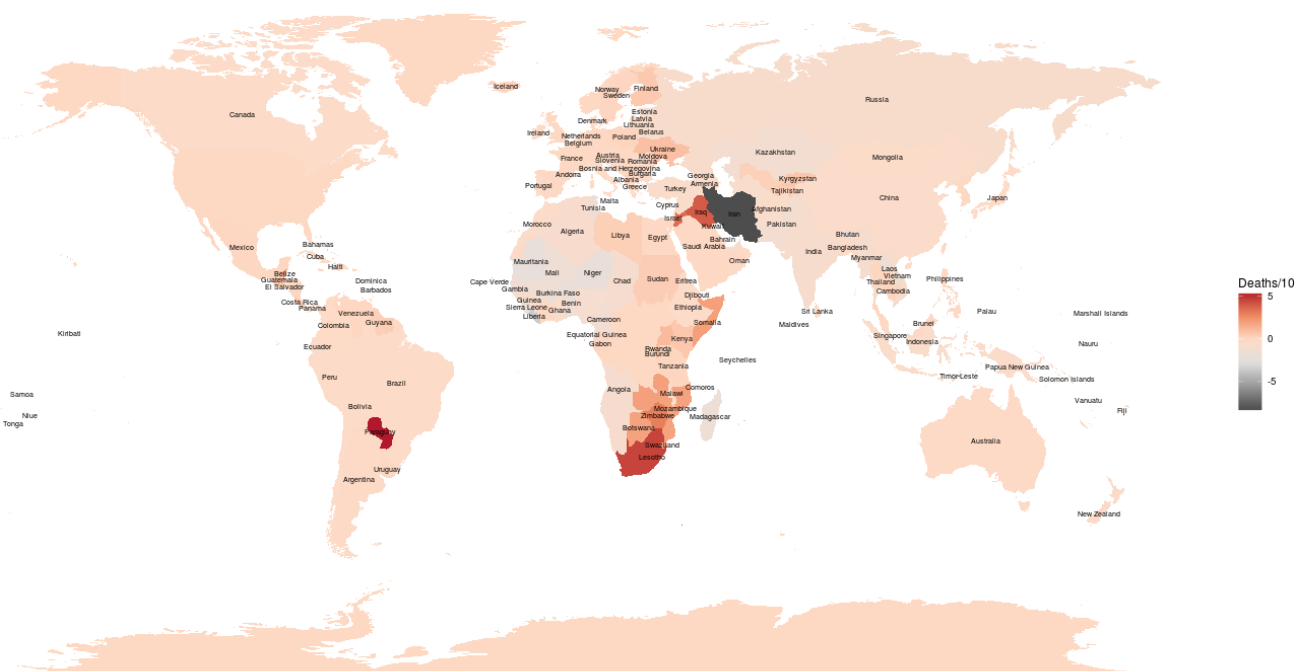Trend of burns deaths/accidents age adjusted mortality in 2004 and 2002

FIG. 3. Heat map for showing the trend in years 2002 and 2004 in the countries

Burns deaths/accidents age adjusted mortality per 100 000 standard population in 2004 and 2002
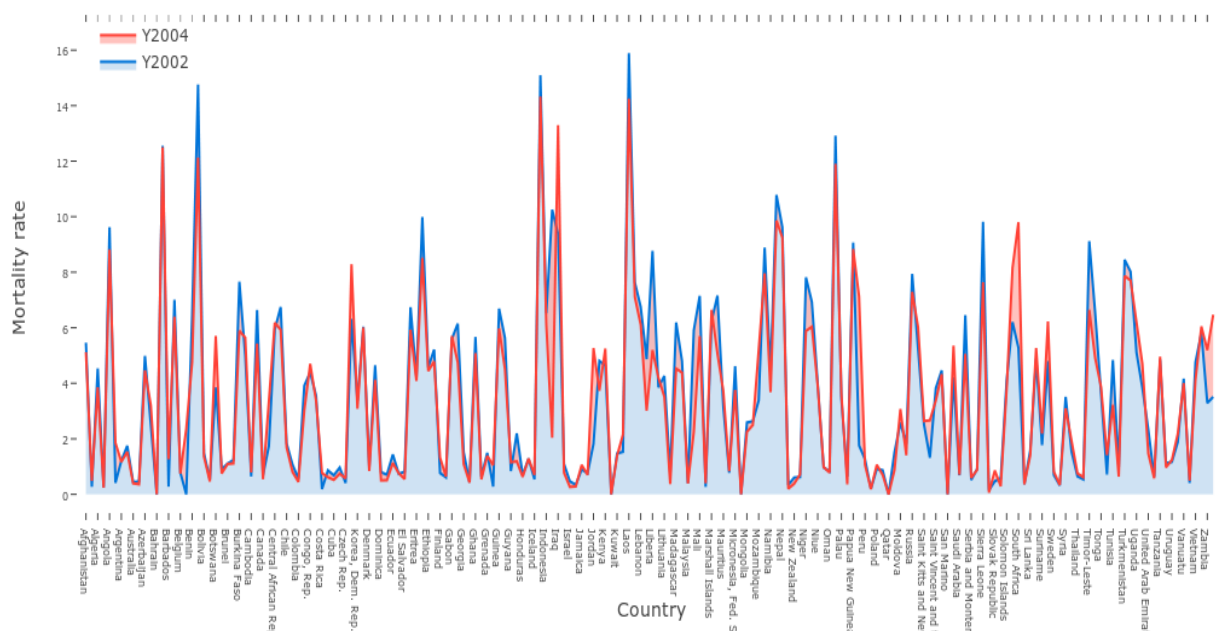
FIG. 4. Area plot of burns age-adjusted mortality rate in 2002 and 2004

IX. COMPARISON OF THE PLOTS

From the study of pre-attentive features, it could be seen that in heat map that darkest color has greatest degree of pop out than others, so visualizing any interesting phenomenon would be easy. But going by the Tufte's laws, it would be optimal to plot the bar plot for visualizing the mortality rates.

For comparing the trends, the heat map of the difference of the values of 2004 with 2002 and area plot were plotted. From the discussion in the previous section, the heat map is a better option. But these method would not suffice if more years are added to the data, then one should plot conventional 2D matrix-type heat map for visualizing the trend.

X. DRAWBACKS OF THE PLOTS

The drawbacks of the plots drawn are that in case of heat map on world map plotting of mortality rates for multiple years is not possible, hence the bar plot was used for plotting the mortality rate for the years. Another drawback of the heat map on the map is that the size of the area of the country takes precedence over the actual information.

The bar plot comes with its set of problems, the plot would get cluttered and it would be difficult to accommodate the labels on the axes, so the interactive version is also needed where it could be zoomed to check the details for an individual country. Hovering over the plot would give additional information. It should be noted that not all countries were labelled on the x-axis in the static plot provided for bar and area plot.

XI. RESULTS AND DISCUSSIONS

For the year 2004, from FIG. 1, it could be seen that deaths are more in the countries like India, Iraq, Vietnam, Pakistan and Laos, which are developing countries. This could be even confirmed from the bar and area plot.

So proving the hypothesis, it could be said that mortality rate is higher in the developing countries than in developed countries. It could be also said it's almost non-existent in developed countries.

For the trend, from FIG 3 it could be said that Iran saw a considerable decrease in the deaths while the countries like Iraq, Paraguay, South Africa and Jordan saw an increase.

XII. REFERENCES

[1] Luana Micallef,(2016) Lecture 6, T-61-5010, Information Visualization, *Aalto University.*

[2] Luana Micallef,(2016) Lecture 7, T-61-5010, Information Visualization, *Aalto University.*

[3] Nikolaj Tatti,(2016) Lecture 2, T-61-5010, Information Visualization, *Aalto University.*

[4] Nikolaj Tatti,(2016) Lecture 3, T-61-5010, Information Visualization, *Aalto University.*

[5] Munzner.T.(2015) Visualization Analysis and Design, *CRC press,* 157-159.

[6] Jeffrey Heer, Michael Bostock, and Vasdim Ogievetsky (2010) A Tour Through the Visuallization Zoo, *Communications of the ACM 53(6):,* 59-67.