

Data Visualization Assignment 2

Name : Varun Ramesh
Student ID : s3793675

RPubs Link:

<https://rpubs.com/varunr/BeliefInVaccinesByCountry>

DataSource Link:

<https://wellcome.ac.uk/sites/default/files/wellcome-global-monitor-2018.pdf> (Wellcome Global Monitor page 114,116 for data table)

Original DataVisualization Link:

<https://wellcome.ac.uk/reports/wellcome-global-monitor/2018#&gid=1&pid=9>

R-MarkDown:

title: "Assignment 2"

author: "Varun Ramesh, S3793675"

subtitle: Deconstruct, Reconstruct Web Report

output:

pdf_document: default

html_document: default

urlcolor: blue

```
```${r setup, include=FALSE}
```

```
knitr::opts_chunk$set(warning = FALSE, message = FALSE)
```

```
```
```

{.tabset}

Click the **Original**, **Code** and **Reconstruction** tabs to read about the issues and how they were fixed.

Original

<center>

</center>

<center>*Source: Wellcome Global Monitor 2018 (link to original visualization)
*</center>

Introduction

Wellcome Trust is a research-charity based foundation in London. The Wellcome Global Monitor has findings about global attitude to science and health with help of a survey. It has minds of different people from different countries and their answers to how science and health can benefit the society. The survey which is picked in the report is related to how the world perceive vaccines for diseases.

Objective

Explain the objective of the original data visualisation and the targetted audience.

The visualisation chosen had the following three main issues:

* Perceptual Issue; We can see that the graph has no Y-axis present and it can be very hard to judge what the graph is about. There is no point to distribute the points in both two axes if there is data being displayed only in one axis. This creates a confusion to understand what the Y-axis denotes.

* Deceptive Issue: As shown from the original graph, Asia is on the top, followed by Africa, North America, Europe and then finally the Soviet Union. This can give out a false information saying Asia is doing better than all other continents. The data is so inconsistent that it depicts only some part of Asia is doing good while some parts of North America is doing bad. Although many dots are plotted, we cannot have direct comparison between two countries. The graph could have been acceptable if it was an interactive one with a hover-over feature.

* Data Integrity: Looking at the Asia Section, we can see New Zealand has also been included in Asia. This is a typo and New Zealand should never have been included in Asia.

****Reference****

* Wellcome Globe Monitor 2018. Retrieved May 5, 2020, from an UK based global charity program and this survey was taken from a journal which talks about how people around the world feel about science and major health challenges: <https://wellcome.ac.uk/sites/default/files/wellcome-global-monitor-2018.pdf> Wellcome Global Monitor (pg 114,116 for data table)

Code

The following code was used to fix the issues identified in the original. A new dataFrame had to be created with already present data existing. Countries from North America had to be excluded because there were very few countries present and would give out a convoluted result. As we can see, only the country names were given and they had been wrongly assigned to different regions. Creating a new dataFrame with a new column with regions included so we can make the visualization more sensible. We can add region names to the Y-axis to avoid the deception issue

```
`{r}  
library(ggplot2)  
library(tidyr)  
library(dplyr)
```

```
vaccines <- data.frame(Countries =
c("Bangladesh","Egypt","Ethopia","Liberia","Tanzania","India","Afghanistan","Rwanda","Thailand","
Sierra
Leone","Uzbekistan","Venezuela","Jordan","Laos","Malawi","Mozambique","Nepal","Nicaragua","M
alaysia","Myanmar","Nigeria","Palestinian","Tajikistan",
"France","Gabon","Togo","Russia","Switzerland","Armenia","Austria","Belgium","Iceland","Burkina
Faso"),
```

```
Perc =
c(97,97,96,96,96,95,94,94,93,93,93,93,92,92,92,92,92,92,91,91,91,91,91,33,26,25,24,22,21,21,21,21
,20),
```

```
Region = c('Asia','Gulf','Africa','Africa','Africa','Asia','Asia','Africa','Asia','Africa','Former Soviet
Union','South America','Gulf','Asia','Africa','Africa','Asia','South
America','Asia','Asia','Africa','Gulf','Gulf','Europe','Africa','Africa','Former Soviet
Union','Europe','Gulf','Europe','Europe','Europe','Africa'))
```

...

If we look at the original graph, we can see that there is a line in the graph which deotes median. This makes the graph very cluttered. To make it much more simpler, a new graph can be added which has the means of each of these regions. This becomes more informative and efficient. Firstly, the means of each of the regions are computed. Then we make a second dataFrame with only the means and regions.

```
``{r}
```

```
region_mean = c(mean(filter(vaccines, Region == "Africa")$Perc),mean(filter(vaccines, Region ==
"Asia")$Perc),mean(filter(vaccines, Region == "Europe")$Perc),mean(filter(vaccines, Region ==
```

```
"Former Soviet Union")$Perc),mean(filter(vaccines, Region == "Gulf")$Perc),mean(filter(vaccines,
Region == "South America")$Perc))
```

```
means_df <- data.frame(Means = c(74.6, 93.1, 23.6 ,58.5 ,78.4 ,92.5),
                        Regions = c('Africa','Asia','Europe','Former Soviet Union','Gulf','South America'))
```

```
```
```

```
Graph
```

```
```{r}
```

```
#Graph for vaccines %
```

```
vaccine_plot <- ggplot(data = vaccines, mapping = aes(x = Countries, y = Perc)) + theme_classic() +
geom_bar(stat = "identity", aes(fill = Region), position = "dodge") + coord_flip() + facet_grid(rows =
vars(Region), scales = "free_y", space = "free_y", switch = "y") + theme(strip.placement = "outside",
strip.background = element_blank(),
panel.border = element_blank(),
strip.text.y = element_text(angle = 180)) + guides(fill = FALSE) + xlab("") + ylab("% of people who
believe vaccines are safe") + labs( title = "Vaccine awareness by country and region", caption = "Data
Source: Wellcome Globe Monitor 2018")
```

```
#Graph with means
```

```
mean_plot <- ggplot(data=means_df, aes(x=Regions, y=Means)) + geom_bar(stat="identity",
width=0.5, fill="steelblue") + geom_text(aes(label=Means), vjust=1.6, color="white", size=3.5) +
theme_minimal() + labs( title = "Mean of Region-Based vaccine awareness", caption = "Data Source:
Wellcome Globe Monitor 2018")
```

```
#ggplot(vaccines, aes(x = Perc, y = Region)) + geom_line(size = 1.5, color="black", group = 1)
```

```

#### **\*\*Data Reference\*\***

\* Wellcome Globe Monitor 2018. Retrieved May 5, 2020, from an UK based global charity program and this survey was taken from a journal which talks about how people around the world feel about science and major health challenges: <https://wellcome.ac.uk/sites/default/files/wellcome-global-monitor-2018.pdf>>Wellcome Global Monitor (pg 114,116 for data table) </a>

#### **#### Reconstruction**

The following plot fixes the main issues in the original.

```
```{r fig.align="center", echo = FALSE, fig.height=10,fig.width=11}
```

vaccine_plot

'''

```{r fig.align="center", echo = FALSE, fig.height=5,fig.width=8}

mean\_plot

'''