

## Data Visualisation - Assignment 2

### Deconstruct, Reconstruct Web Report

Vishal Patnaik Damodarapatruni - s3811521

02/05/2021

RPubs link: <https://rpubs.com/VISHAL34PATNAIK/764369>

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

#### Original

*Source:* The United States of Debt. (2021). Retrieved 27 April 2021, from <https://howmuch.net/articles/the-us-debt-map-2020>

#### Objective

The main objective of the visualization is to understand the debts of United States. Whether each state is in debt or not. If so, how much debt they are in and which states already went into a poor state due to huge debts based on Liabilities and assets of the states. It mainly focused on the data of year 2020. As we all know how 2020 has been and it is not a kind one for budgets. Whether the state is spending more than the previous years or is minimizing costs to lower the budget.

#### Targeted Audience:

- Financial institutions
- Government officials
- Economists
- Data analysts
- People all over the world.

#### The visualization chosen had the following three main issues:

- The values are rounded to one decimal which results in the loss of exact data in millions. For example: 24,467,342,000 is rounded as 24.5 B and 24,447,387,000 is rounded to 24.4 B. In the first case there is an addition of nearly 33 million and in the second case there is a deficit of 47 million approx. This difference in million has a serious issue in analyzing or understanding the data by most of the audience.

- On observing the map in the first sight we will assume that higher the size of the circle the higher will be the debt of the state and there by color intensity will be high in red color. Similarly, if the debt is low the circle will be small with color variations in green. But here we should observe more so that we will understand that the size is based on Liabilities and color is based on Debts. It will be difficult to understand in the first case in most of the cases to most of the people.
- The visualization has the data of 50 different states and each state has different Liabilities and Debt ratios. Placing these 50 states on a single visualization makes it clumsy and difficult to observe the data of some states. For example, the top left corner of the map has the bubble overlaps. This will not be a visual for most of the people as different sized bubbles in different colors overlapping at a single point due to smaller size and larger liabilities and debts of the state. This adds complexity in understanding the visualization. It would be better if some of the countries having lower debts are trimmed as we need the states with most debts.

## Reference

The United States of Debt. (2021). Retrieved 27 April 2021, from <https://howmuch.net/articles/the-us-debt-map-2020>

## Code

The following code was used to fix the issues identified in the original.

```
library(dplyr)
library(tidyr)
library(rvest)
library(stringr)
library(ggpubr)
library(egg)
library(ggplot2)

V_USDEBTS <- as.data.frame(read_html("https://docs.google.com/spreadsheets/d/
e/2PACX-1vQwS4OadAKoTRIL8agGqzs1D_Q-V1y9pcGYAt8pM4XvaII_nIg-MPwApk-eoYINEHxIg
aVorfKazBMU/pubhtml/sheet?headers=false&gid=0") %>% html_table(fill=TRUE))

V_USDEBTS <- V_USDEBTS[-1, -c(1, 4, 7:9)]

V_USDEBTS <- V_USDEBTS %>%
  rename("Rank" = "Var.2", "State" = "Var.3", "Total Liabilities" = "Var.5",
"Debt Ratio" = "Var.6")

V_USDEBTS$Rank <- as.integer(V_USDEBTS$Rank)

V_USDEBTS$`Total Liabilities` <- str_remove_all(V_USDEBTS$`Total Liabilities`
, "[$,]")
V_USDEBTS$`Total Liabilities` <- as.double(V_USDEBTS$`Total Liabilities`)
V_USDEBTS$`Total Liabilities` <- V_USDEBTS$`Total Liabilities`/100000000

V_USDEBTS$`Debt Ratio` <- str_remove_all(V_USDEBTS$`Debt Ratio`, "[%]")
V_USDEBTS$`Debt Ratio` <- as.integer(V_USDEBTS$`Debt Ratio`)
```

```

V_USDEBTS <- V_USDEBTS[order(V_USDEBTS$Rank), ]

V_USDEBTS <- V_USDEBTS %>% mutate(`DebtLevel` = ifelse(`Debt Ratio` >= 400 &
`Debt Ratio` <= 500, "400% - 500%",
                                     ifelse(`Debt Ratio`
>= 300 & `Debt Ratio` <= 400, "300% - 400%",
                                     ifelse(`Debt
Ratio` >= 200 & `Debt Ratio` <= 300, "200% - 300%",
                                     ifelse
(`Debt Ratio` >= 100 & `Debt Ratio` <= 200, "100% - 200%",
ifelse(`Debt Ratio` >= 75 & `Debt Ratio` <= 100, "75% - 100%",
ifelse(`Debt Ratio` >= 50 & `Debt Ratio` <= 75, "50% - 75%",
ifelse(`Debt Ratio` >= 25 & `Debt Ratio` <= 50, "25% - 50%", "0% - 25%"
)
)
)
)
)
)
)
)
)
)
)

V_USDEBTS$DebtLevel <- factor(V_USDEBTS$DebtLevel,
                             levels = c("400% - 500%", "300% - 400%", "200% - 300
%", "100% - 200%", "75% - 100%", "50% - 75%", "25% - 50%", "0% - 25%"), order
ed = TRUE)

#Taking top 20 for better visualization.
V_USDEBTS1 <- V_USDEBTS[1 : 20, ]

V_Plot1 <- ggplot(data = V_USDEBTS1, aes(x = `State`, y = `Total Liabilities`
)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  theme_minimal() +
  expand_limits(y = c(0, 500)) +
  geom_text(aes(label = paste("$", `Total Liabilities`, " B", sep="")), nudge
_y = -2, nudge_x = .05, hjust = -0.075) +
  labs(title = "Total Liabilities (Billion doll
ars) and")

```

```
V_Plot2 <- ggplot(data = V_USDEBTS1, aes(x = `State`, y = `Total Liabilities`
)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  theme_minimal() +
  expand_limits(y = c(0, 500)) +
  theme_minimal() +
  geom_text(aes(label = paste("$", `Total Liabilities`, " B", sep="")), nudge
_y = -2, nudge_x = .05, hjust = -0.075)
```

```
V_Plot3 <- ggplot(data = V_USDEBTS1, aes(x = `State`, y = `Total Liabilities`
)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  theme_minimal() +
  geom_text(aes(label = paste("$", `Total Liabilities`, " B", sep="")), nudge
_y = -2, nudge_x = .05, hjust = -0.075)
```

```
V_Plot4 <- ggplot(data = V_USDEBTS1, aes(x = `State`, y = `Debt Ratio`, fill
= `DebtLevel`)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  expand_limits(y = c(0, 500)) +
  theme_minimal() +
  labs(title = "Dept ratio (%) by State", y = "Debt Ratio") +
  geom_text(aes(label = paste(`Debt Ratio`, "%", sep="")), nudge_y = -2, nudg
e_x = .05, hjust = -0.075) +
  scale_fill_brewer(palette = "Greens", direction = -1)
```

```
V_Plot5 <- ggplot(data = V_USDEBTS1, aes(x = `State`, y = `Debt Ratio`, fill
= `DebtLevel`)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  expand_limits(y = c(0, 500)) +
  theme_minimal() +
  labs(title = "Dept ratio (%) by State", y = "Debt Ratio") +
  geom_text(aes(label = paste(`Debt Ratio`, "%", sep="")), nudge_y = -2, nudg
e_x = .05, hjust = -0.075) +
  scale_fill_brewer(palette = "Greens", direction = -1)
```

```
V_Plot6 <- ggplot(data = V_USDEBTS1, aes(x = `State`, y = `Debt Ratio`, fill
= `DebtLevel`)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  expand_limits(y = c(0, 500)) +
```

```
theme_minimal() +
  geom_text(aes(label = paste(`Debt Ratio`, "%", sep=""), nudge_y = -2, nudg
e_x = .05, hjust = -0.075) +
  labs(title = "Dept ratio (%) by State", y = "Debt Ratio") +
  scale_fill_brewer(palette = "Greens", direction = -1)
```

## Data Reference

- The United States of Debt. (2021). Retrieved 27 April 2021, from <https://howmuch.net/sources/the-us-debt-map-2020>
- DePietro, A. (2021). States With The Most And Least Debt In 2020. Retrieved 27 April 2021, from <https://www.forbes.com/sites/andrewdepietro/2020/11/23/states-with-the-most-and-least-debt-in-2020/?sh=6f4c520e78a3>

## Reconstruction

The plot which fixes the three main issues in the original visualization is shown below,

