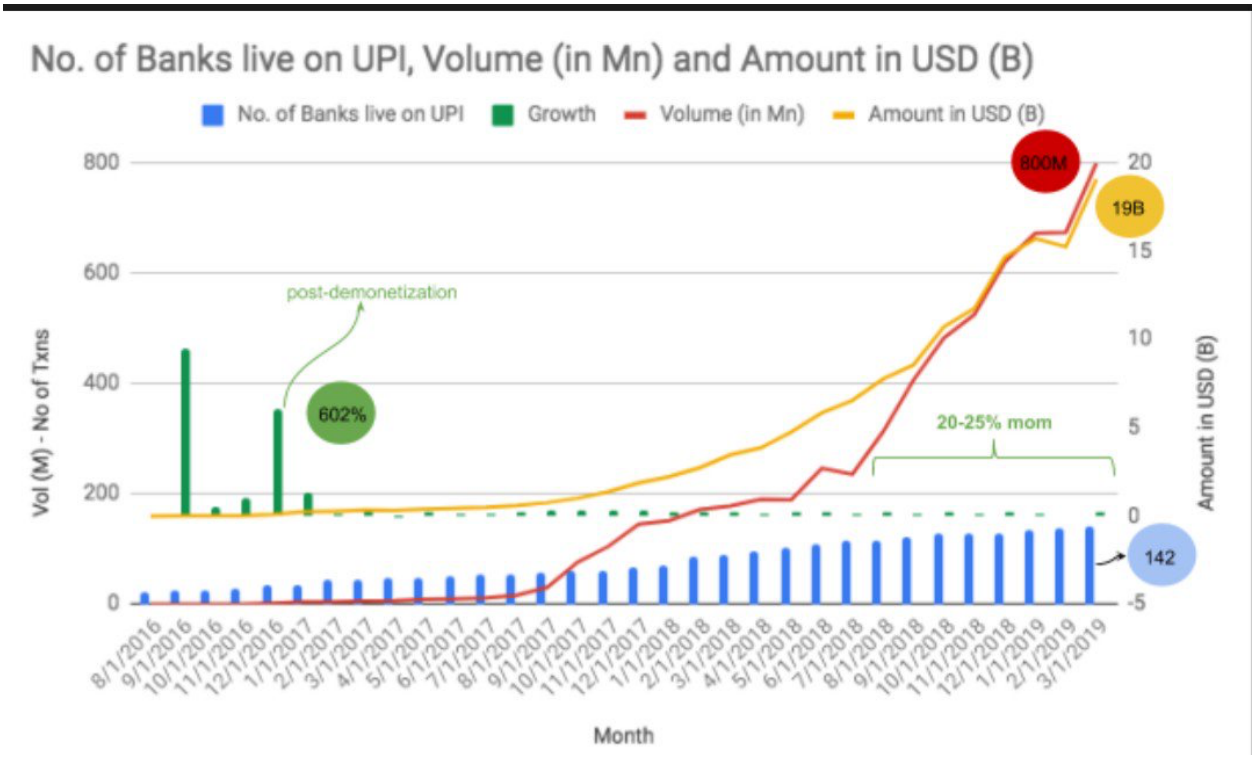
Component II: Data Visualization

Part 1. Graphic inquisition

Figure 1

Target Figure to Critique



Note. 2019, Retrieved from https://twitter.com/MohapatraHemant/status/1120212609045655552

1. **Gestalt principles & visual structure**

There are mixed bars and lines in the graphic making readers confusing about what they should focus on. Based on similarity principle, we would perceive the set of green-colored "bars" as a group, while it could also represent the 0-scale level measured by the right vertical axis (dotted line).

2. **Keep it simple: Decoding & Operations**

We can see the trend of lines in the graph, while the bars are not very sensible from both the trend and the values, especially we cannot gain any information about the growth. The grid line is not helpful in this case, as the number of scales in two vertical axes is not equal, and there is even negative scale on the right vertical axis.

3. **Less is more: Chartjunk & data-ink ratio**

The graphic design is too ambitious trying to display values with different magnitudes in one x-y coordinate axis. The bars are redundant in the graphic and should be removed. Instead, simple lines with clear scales and labels would tell the whole story.

4. **Graphical data integrity & lie factor**

As shown in the figure, the highest growth rate is depicted when both the number of transactions and amount of money were extremely low. The intersection point between two lines is also deceiving as we may think the two variables are comparable.

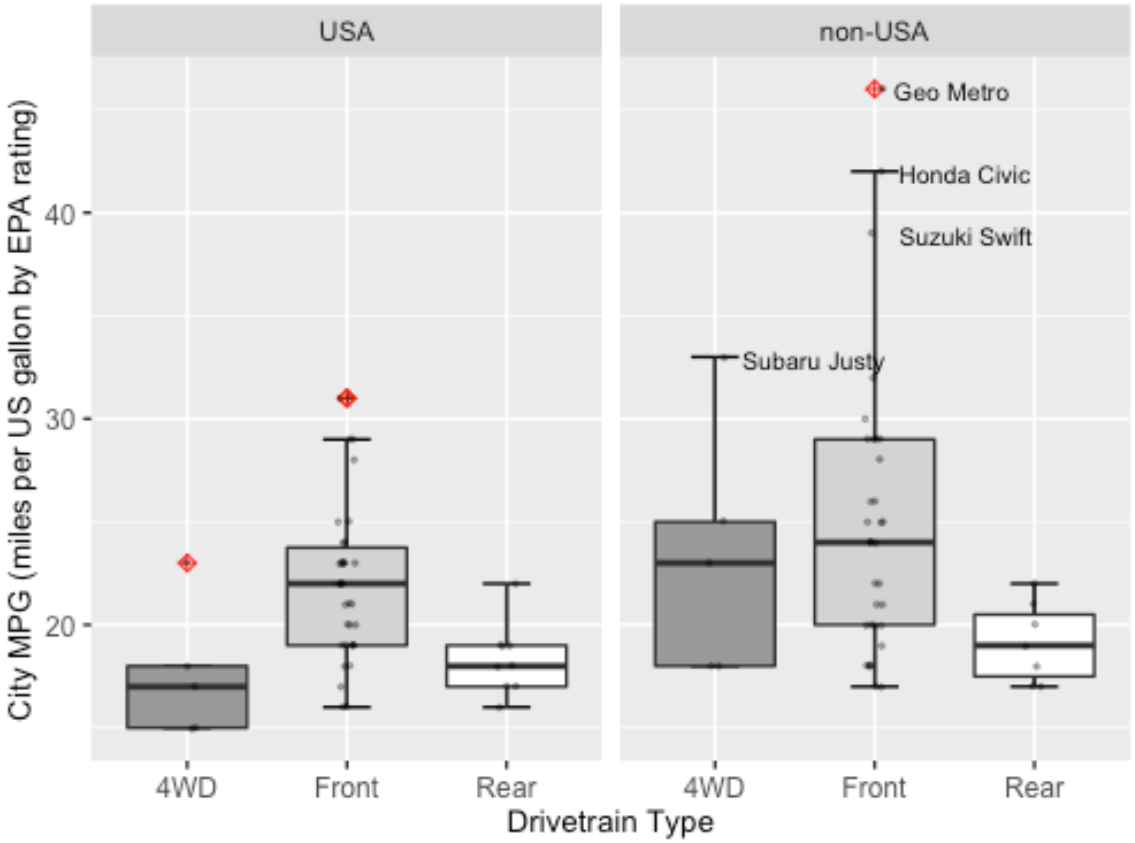
5. **Annotation & stand-alone readability**

The graphic is designed with title, legend, and axis labels. However, with the horizontal axis labeled month, the actual scale is disordered in the format of date, and it is basically not readable. Other marks like "20-25% mom" do not make too much sense. Too much abbreviation involved in the annotation increases difficulties to read.

Part 2. Graphic design

Figure 2

Comparison of MPG of Different Drivetrain Types for USA and Non-USA Cars



Note. The figure is represented by using the *Cars93* dataset in the MASS package in R. In the dataset, 93 cars were selected randomly from passenger car models in 1993 sold in the USA. The two panels demonstrate cars with different drivetrain types which originally from America or not. Four outliers stand out in the category of non-USA cars by using the level of 1.5 times IQR (interquartile range).

Intention and Motivation

The graphic intends to show which drivetrain type of cars is more fuel-efficient among 93 car models made in/outside America. There are three variables represented in the figure, including a continuous variable MPG.city and two categorical variables, drive train type (front wheel, rear wheel, and 4WD) and origins (USA and non-USA). As the graphic illustrated, cars with front wheel drivetrain are more fuel-efficient than those with the other two types. In addition, non-American cars can be driven slightly further per gallon of gasoline compared to American cars for all three types of drivetrain. Four outliers are found in the right panel by extending the greater values of MPG.city to a max of 1.5 times the interquartile range, which help us identify the most fuel-efficient cars in the sample.

The reason why I choose boxplots to represent the data is that we can compare MPGs corresponding to each type of drivetrain independently. The other advantage of using boxplots is to notify outliers (the most fuel-efficient ones or the least ones) easily. Specifically, the same color is applied for each drivetrain type in two facets helps us to sense the differences intuitively. The grid line in the background together with consistent scales assists us read values of five numbers plotted in each boxplot. Jitter is added to show the distribution of observations in each type of drivetrain. In addition, the graph is designed with annotation, including one descriptive title and labels for two axes.