Part 1. Graphic inquisition

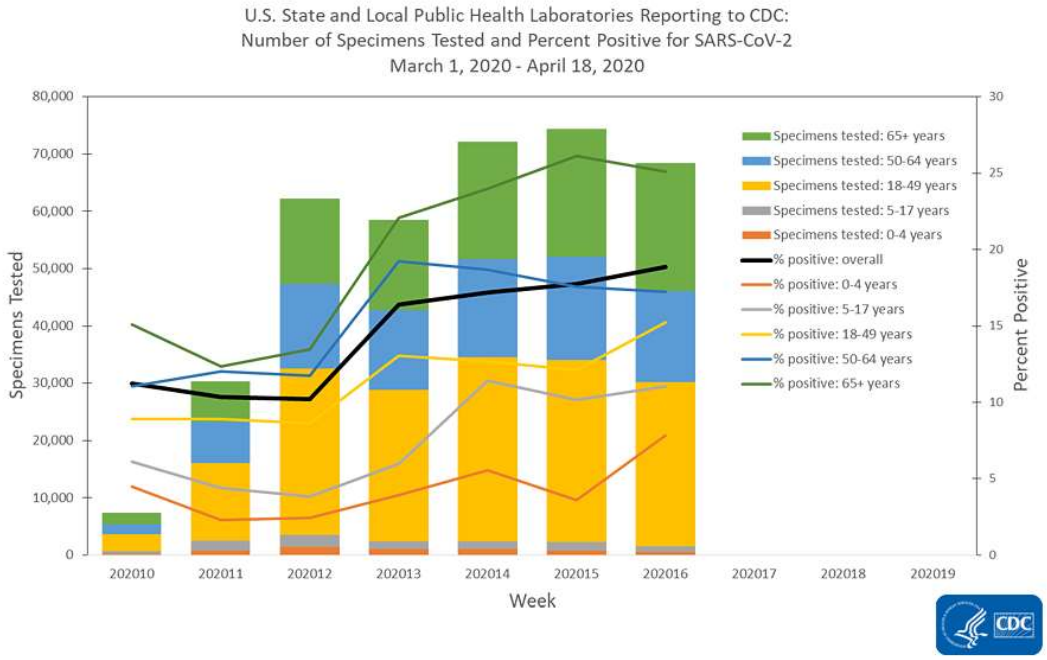

Figure 1: The graph by the Centers for Disease Control and Prevention (CDC) website created based on the weekly data provided by U.S. State and Public Health Laboratories during the surge of Covid 19 in March/April 2020.

Source: https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/04242020/public-health-lab.html

Last checked online- 18/10/2022 at 3:27.

The above CDC graphic was intended to show the number of specimens tested and percent of positive cases for SARS-CoV-2 in one single graphic by plotting both the variables in double Y- axis where X-axis depicted the age groups.

The graphic looks quite decent. Using stacked bars and simple lines makes it visually structured. Color is contextual and meaningful -five contrast colors represented 5 different age groups. Black line is explicit for readers to understand overall positive cases

The graphic, at the first sight looks a bit complex, but it is not super-hard to decode the necessary information. However, the uneven (jiggling) base line along X-axis makes the bar diagram heavy to understand the trends for the 65+ (green), 50-64 (blue), and 18-49 (yellow) age groups. Grid lines could be used to better anchor the reference points for effective scanning of bar compositions. Superimposition of bar and line graphs made it reader unfriendly.

Indeed, no chartjunk, distracting or redundant element was used, and data-ink ratio is fair enough to display the data on top of everything.

The graphic is trustworthy since the data is reliable and accessible on CDC website. However, on X-axis different age groups contain variable age differences, such as 4 (0-4), 12 (5-17), 31 (18-49) and so on, which might distort the size effect and confuse the readers.

Unclear and ambiguous labelling of year and week on X-axis makes the readers puzzled for a while. Even the last three points remained incomplete with no data shown but the legends lied upon. Double Y-axis were used to indicate the bar and line charts making it hard for viewers to immediately identify the values of individual data points. The coordination between the line chart and corresponding Y-axis (right) is lost in the space occupied by the legends.

Cramming together two quantitative variables into a single, overly-complex graph made it difficult for readability on its own with minimal effort. Rather, small multiples visualization concept introduced by Edward Tufte could be a good choice to make at least two vertically or horizontally faceted charts for better stand-alone readability.

Part 2. Graphic design

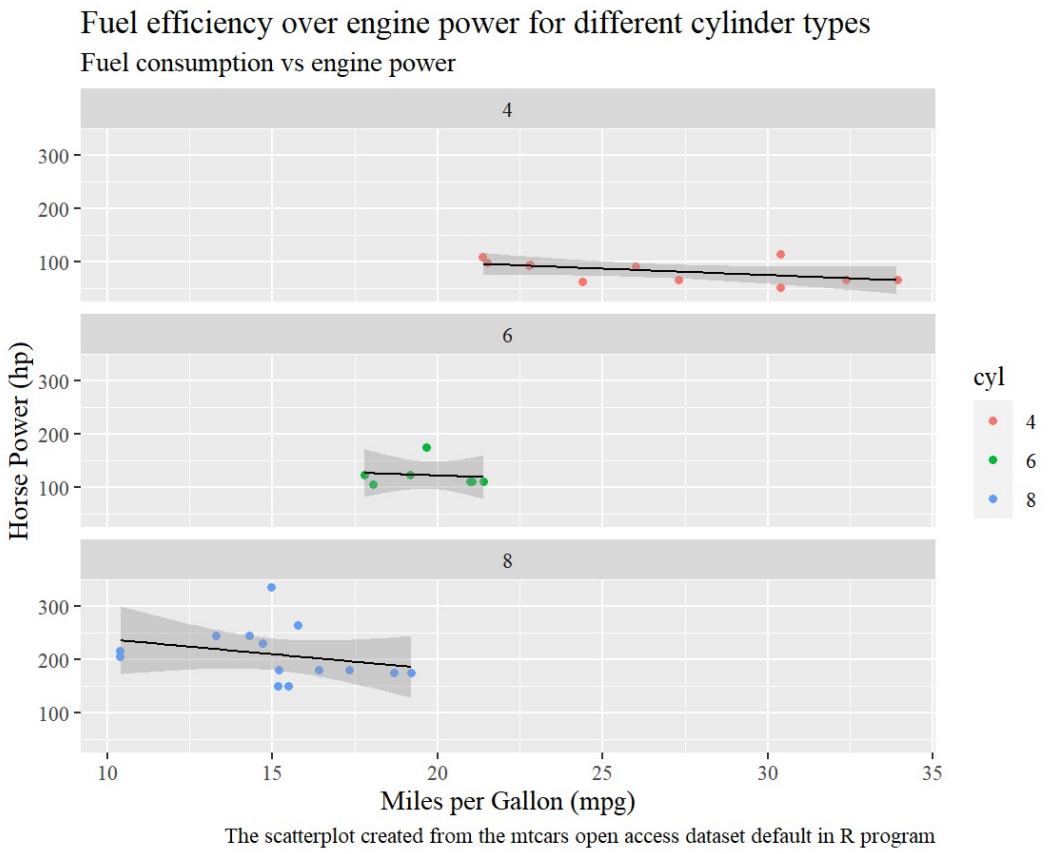

Figure 2: How fuel efficiency varies with the engine output for mtcars with three different types of cylinders.

The mtcars dataset contains measurements on 11 different attributes for 32 different cars. A car engine with higher cylinder numbers would be more powerful and ideally would consume more fuel. Hence, the plot was intended to see how fuel efficiency (mpg) varies with the engines' output (hp) in case of different cylinder types (cyl).

For the graphic, ggplot2 was chosen because of its advantages options over the basic graphics, such as adding multiple layers, themes, facets etc. making the visualization structural, effusive and more engaging. Two discrete (hp & cyl) and 1 continuous (mpg) variables can better be represented by scatterplot where "cyl" was converted to categorical for differentiating facets. The "geom_ point()" could sufficiently show the relationship between two aesthetically mapped variables ("mpg", & "hp"), however. "geom_jitter()" is better to show even small amount of random variation to the location of each point, and to avoid overplotting caused by discreteness in this smaller datasets. The layer "geom_smooth()" was used to aid seeing patterns in the presence of potential overplotting and method "lm" was used to fit linear models. Labels was helpful to make the plot interactive. The panel facet_wrap() was well-enough to form the matrix by single variable "cyl".

Three facets show the relationship between fuel consumption and engine power for 3 types of cylinder cars. In case of 4-cylinder, fuel efficiency reduces from 34 to 22 mpg where engine power increases from 52 to 120 hp indicating a negative linear relationship. Negative relation though weaker is observed between mpg and hp for 8-cylinder cars. In contrast, the 6-cylinder cars are prone to have less impact on fuel efficiency (fluctuating around 20 mpg) while engine power increases from 105 to 175 hp. Overall cylinder number has inverse proportionality with the fuel efficiency meaning high cylinder number car has low fuel efficiency and vice-versa. The graphic showing some data points remain outside the geometric smooth linear model indicating discreteness – that might be due to impact of other constant and latent variables (such as disp, wt, gear.).