

Evaluable Exercise (Practice)

Recreate a plot

Data Visualization Course – Academic year 2021-22 – Professor: Clara Granell

Objectives.

In this project, the student will put into practice the concepts reviewed up to the moment in the practice lessons. The student is required to recreate two plots using R and ggplot2.

Methodology.

In this exercise, you need to recreate the two plots attached at the end of this document. Both plots have been generated using R and ggplot2, and the data used for this plot is the **mtcars** dataset, available in base R. You'll need to reproduce as accurately as possible the two graphs, using the tools we have learnt up to the moment in the practice lessons. Some details of the plot haven't been yet covered in the practice lessons, so you will need to research a bit.

About the plots.

The first plot is a bar plot of the miles per gallon fuel consumption for each of the cars of the dataset. In addition, the bars are colored according to the number of cylinders and the cars are ordered by fuel consumption.

In the second plot, we plot the z-score of the fuel consumption for each of the cars, using a bar plot. In statistics, the standard score is the number of standard deviations by which the value of a raw score (i.e., an observed value or data point) is above or below the mean value of what is being observed or measured. Raw scores above the mean have positive standard scores, while those below the mean have negative standard scores. It is calculated by subtracting the population mean from an individual raw score and then dividing the difference by the population standard deviation:

$$z = \frac{x - \mu}{\sigma}$$

where:

μ is the mean of the population.

σ is the standard deviation of the population.

This process of converting a raw score into a standard score is called standardizing or normalizing.

This plot also includes a color scale according to the values of the z-score, divided in two categories (low, high). Notice also the transparency effect¹.

¹ The transparency effect is there for practice purposes only. According to the theory, it induces an effect where the larger values are less transparent and therefore more noticeable, which is something unwanted.

Important remarks.

The plots have to be recreated using R and ggplot2 only. Some previous data analysis is required, and it should be done in R, but you are free to use the packages of your convenience for the data treatment. However, when it comes to building the plot, the only elements you can use have to be contained within ggplot2.

Materials.

The student needs to deliver their results using one of these two options:

- Traditional option: A zip file containing:
 - The ***.pdf**s of the two plots, as generated with ggplot2 and saved with ggsave.
 - The source code used to generate the plots, with comments that explain the steps followed.
- R-Notebook option: A zip file containing:
 - An R Notebook of the code used to generate the plots and the resulting plots. In this case, both the ***.rmd** and ***.html** files should be included in the zip file. The Notebook should include explanations on the steps followed. For more information about R Notebooks, please read this article: <https://bookdown.org/yihui/rmarkdown/notebook.html>

In both cases, the source code should compile and execute without errors. Don't forget to make sure all libraries you use are included at the beginning of the source code. In case you use a library that wasn't used in class previously, please use this code so that it will be installed if not present:

```
if(!require(somepackage)){  
  install.packages("somepackage")  
  library(somepackage)  
}
```

Grading.

The task will be graded according to the resemblance of the original plots compared to the delivered plots. In addition to this, the methodology used will also be evaluated, that is, how efficient or elegant the programming solution is².

The grade of this exercise contributes with a 30% towards the practice grade. Delivery of this exercise is not mandatory, but it is highly recommended, since it contributes to the Practice grade. Remember that passing Theory and Practice separately is required.

Delivery.

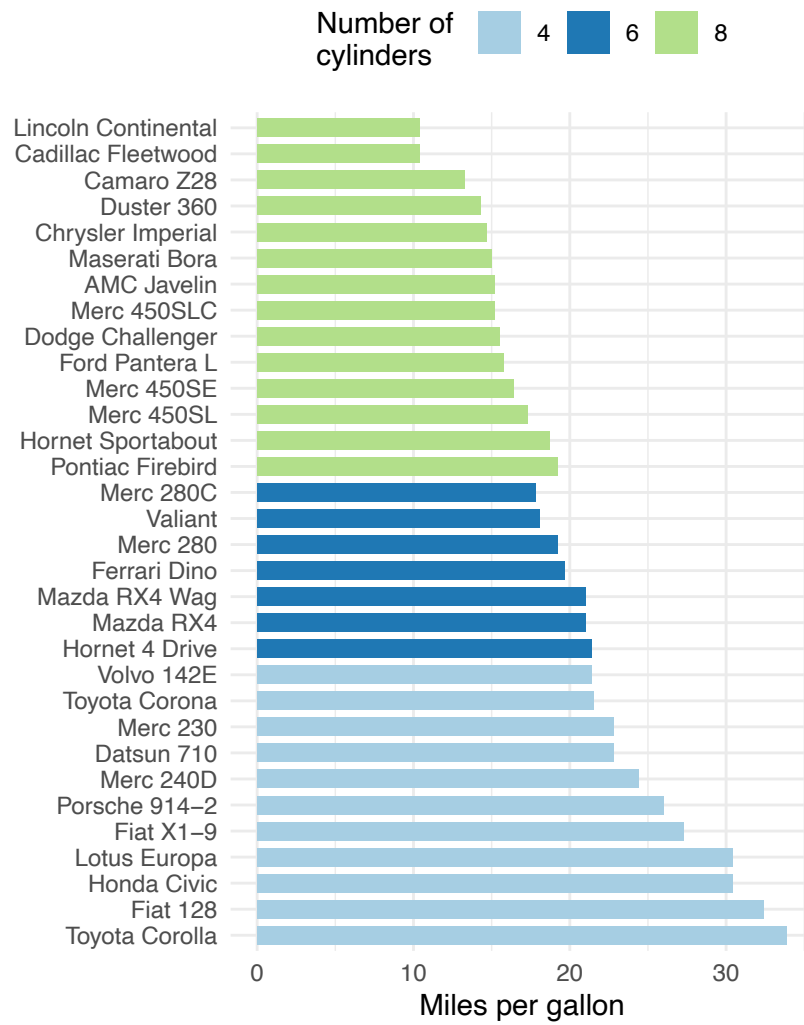
The exercise has to be delivered using the corresponding Moodle task before 19th of April 2022.

Plots to recreate.

(See next page).

² For example, if I want to draw a line plot that displays three lines of three different colors, I could make three calls to geom_line and assign a color manually to each one of the layers (bad solution), or I could find a way to use a single geom_line call to draw the three lines and use a color scale for colors (good solution).

Plot 1:



Plot 2:

