



Universidad de San Andrés

COMPUTATIONAL TOOLS FOR RESEARCH

PROFESSOR: PHD AMELIA GIBBONS

Work N°5

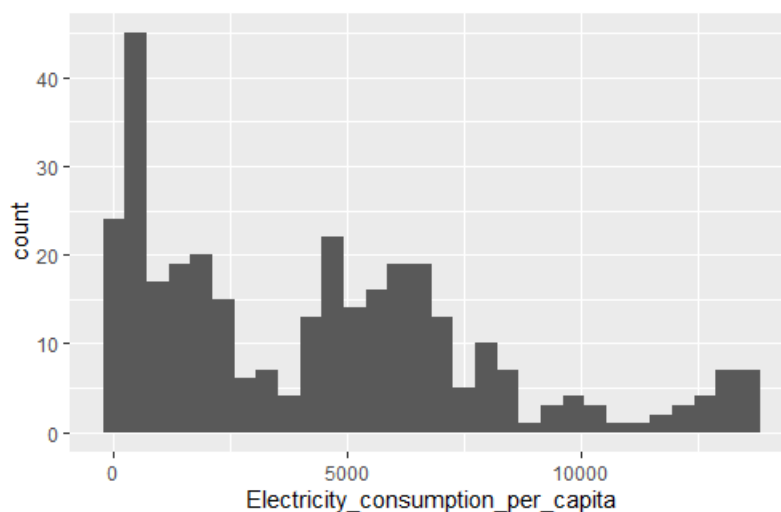
PERUCHIN TOMÁS
FRANCO RIOTTINI DEPETRIS

August 1, 2022

1 Plots corrected to verify Schwabish principles.

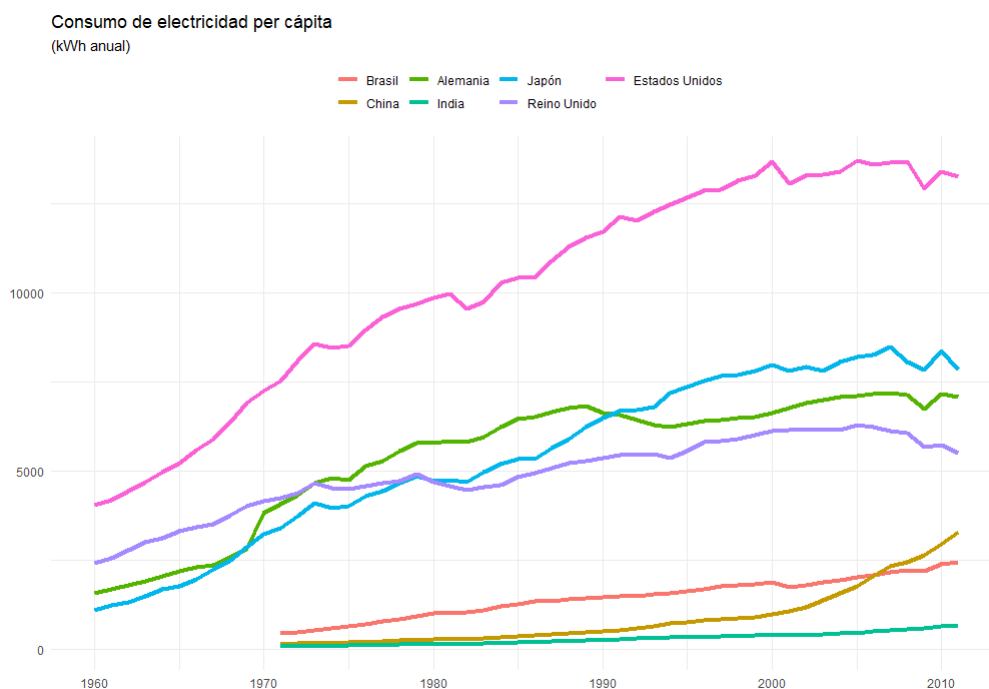
Here we will show the plots of the first video, the originals and the modified. We follow the Schwabish principles to improve the plots.

Figure 1: Original first plot



Source: Own elaboration based on Video 1 of Lesson 2.

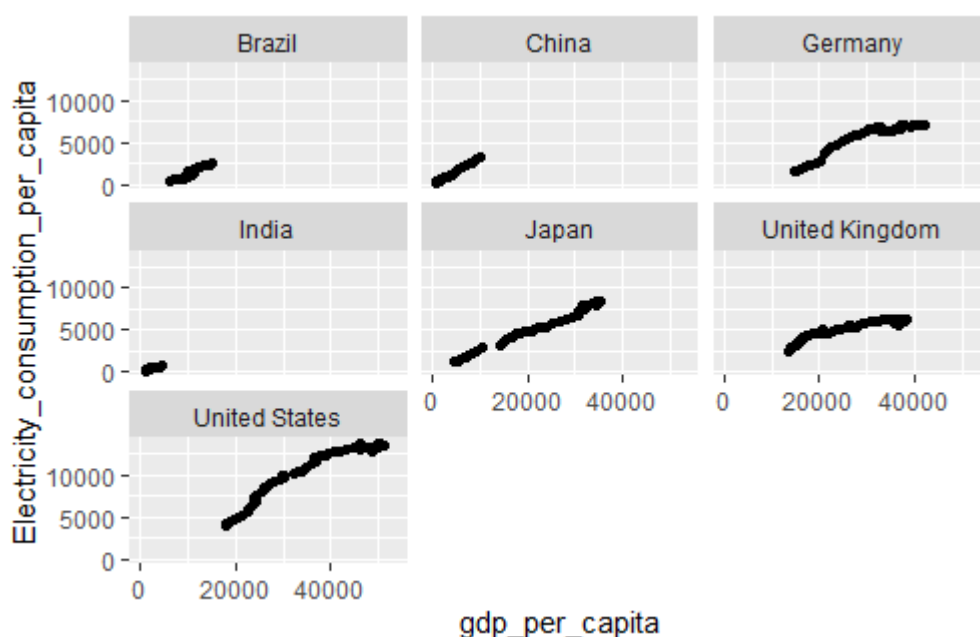
Figure 2: Modified first plot



Source: Own elaboration based on Video 1 of Lesson 2.

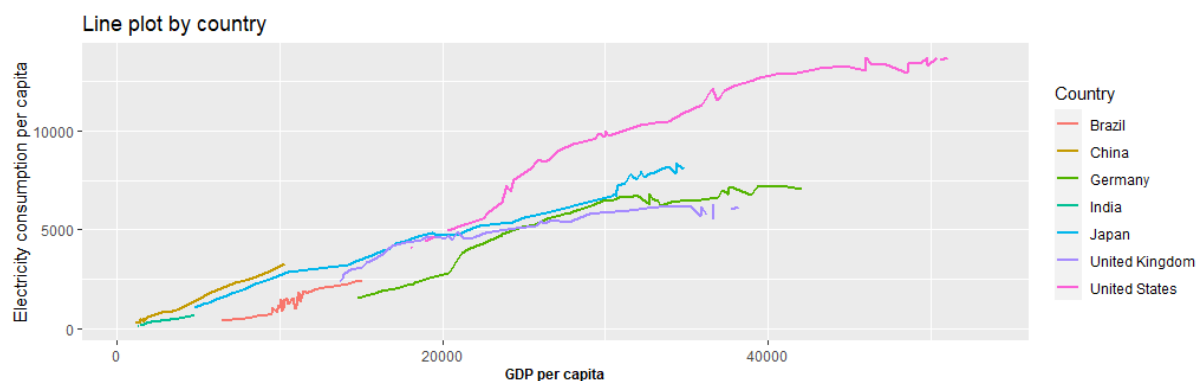
In the original graph (Figure 1), a histogram was being made with data from 7 countries and many years from each country. This graph does not convey a clear idea. For this reason, we select the countries from the year in which they begin to have data for this variable, and we graph it by year and country, conveying an idea of the evolution of the variable. In addition, we add a title, specify the value unit and include a legend.

Figure 3: Original second plot



Source: Own elaboration based on Video 1 of Lesson 2.

Figure 4: Modified second plot

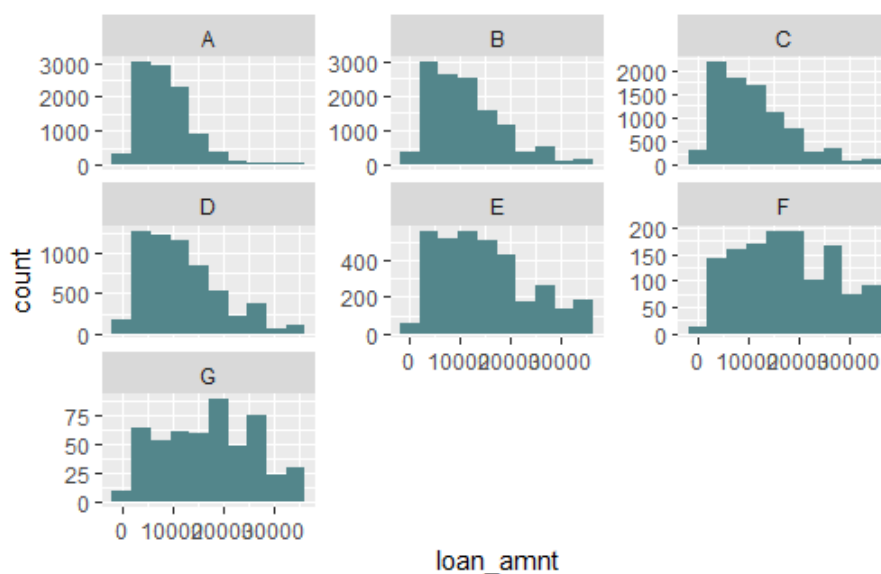


Source: Own elaboration based on Video 1 of Lesson 2.

In the second place, the original figure 3, the plot is confusing because the values of the axes affect the distribution of points for the case of India, Brazil and China. To solve this, we choose a line graph 4 that shows the same variables for both axes but for each country with a different colour, in order to distinguish the relationship between electricity consumption per capita and GDP per capita.

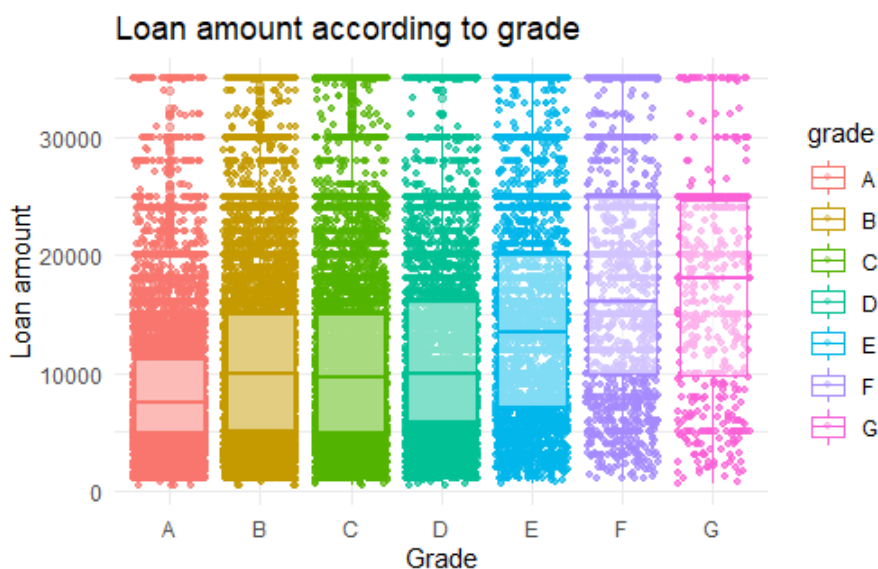
Finally, we modify the graph present in Figure 5. Following the Schwabish principles we create the box plot 6 that shows each credit debt classification in X-axis and loan amount in Y-axis. This makes clear the differences between the dispersion of loan amount for the different categories.

Figure 5: Original third plot



Source: Own elaboration based on Video 1 of Lesson 2.

Figure 6: Modified third plot



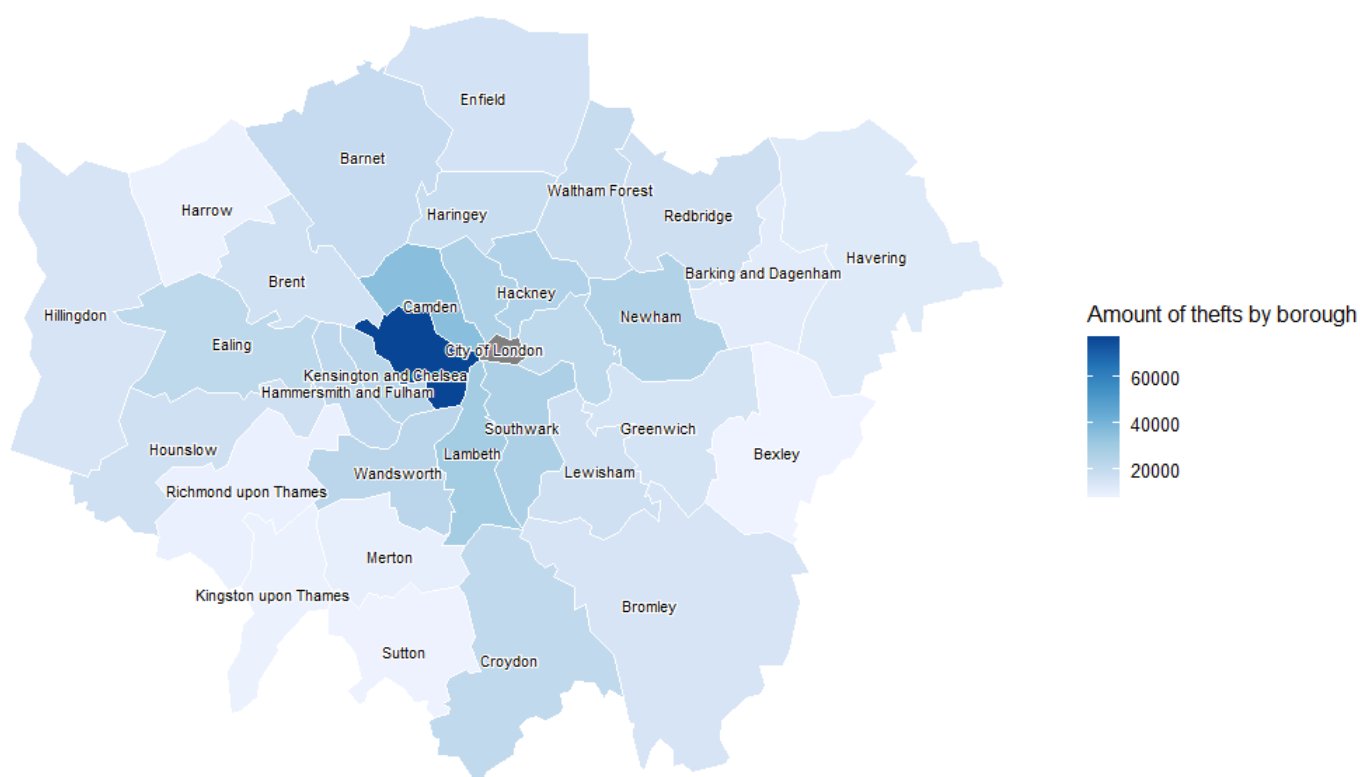
Source: Own elaboration based on Video 1 of Lesson 2.

2 Second part

Map with ggplot2

The following map is made with the ggplot2 package in R. We see the number of robberies by neighbourhood in London. From it, we see that the neighbourhoods with the highest number of robberies are those that are closest to the City of London, for which unfortunately we do not have data available.

Figure 7: Amount of thefts by borough of London



Source: Own elaboration with <https://github.com/Robinlovelace/Creating-maps-in-R>

Map with tmap

In this case, the following map shows the number of robberies by neighborhood in London, but using the tmap package in R. Although we still do not have data for the City of London, we can see how the clearest category means less quantity of thefts. Now this scale is no longer a continuum but there are demarcations every 20 thousand thefts.

Figure 8: Amount of thefts by borough of London

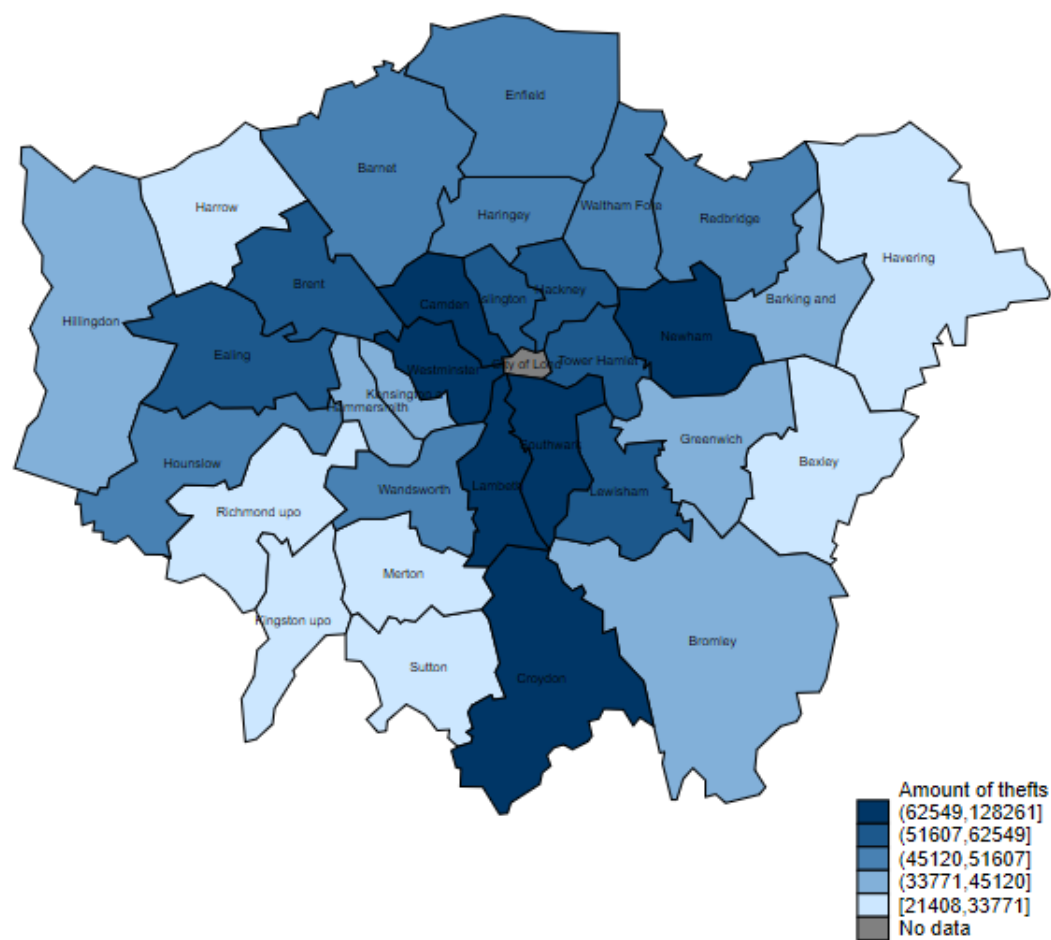


Source: Own elaboration with <https://github.com/Robinlovelace/Creating-maps-in-R>

Map with STATA and spmap

We present again the map of thefts in London, now made with the STATA spmap command. Once again, the neighborhoods on the outskirts of the city center are the ones with the lowest number of robberies.

Figure 9: Amount of thefts by borough of London



Source: Own elaboration with <https://github.com/Robinlovelace/Creating-maps-in-R>