Homework 1.2 - Case 1

Statistical Data Analysis - 29th of October, 2019

Group 5

Antoñanzas Martínez, Guillermo
Pueblas Núñez, Rodrigo
Martínez Capella, Ignacio
Burrell, David

Introduction:

For this assignment, we have chosen two examples of different types of simple graphs to analyse them and find their flaws.

First Simple Visualization:

Source: https://twitter.com/YouGov/status/838720989991223297

Figure 1: First sample graph

Here we can see a representation of the most liked pizza toppings in the UK using a "special" graph representation that looks like a pie chart.

The problem with this graph is very obvious from the beginning: it looks like a simple pie chart with the percentages of the values of the preferred pizza toppings, but in reality, it's a misrepresentation of the proportions, using the pizza just as visualization aid. The percentages are not adding to a 100%, which is the main point of a pie chart, and also in the bottom we can see that there are some observations that do not get represented with no logic behind why, because they have values similar to all the other ones that appear in the final graphic representation.

The easiest fix for this graph would be changing the kind of graph visualization from this to a bar plot. We have made a mock-up of how the graph would turn out using a similar style, with the Pizza types and toppings also displayed.

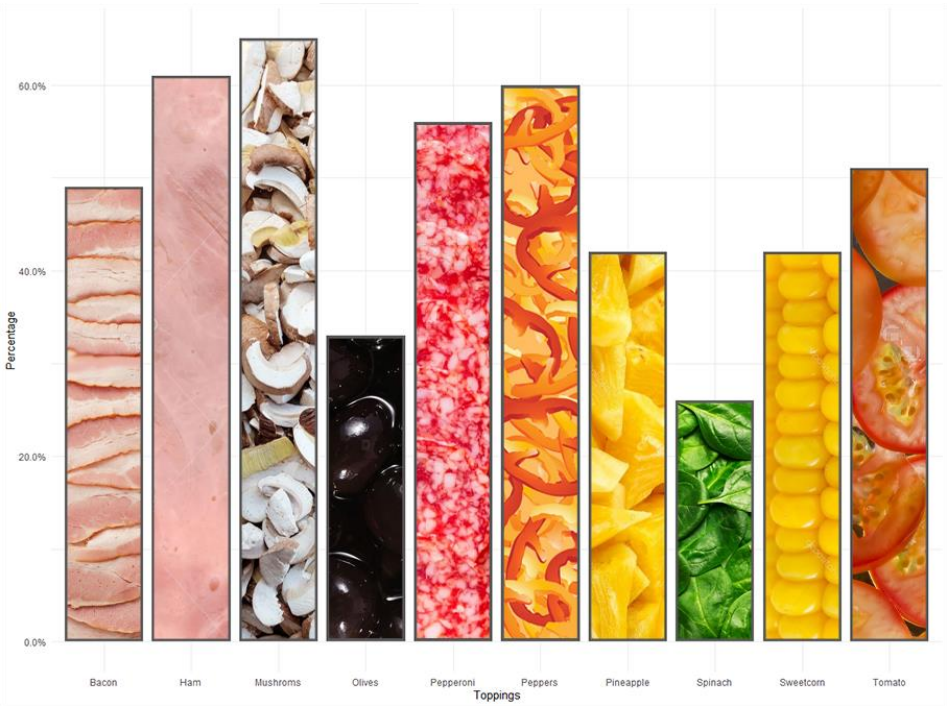

Figure 2: First sample graph fixed

Second Simple Visualization

Source: https://twitter.com/marianorajoy/status/946722575815794689?s=19

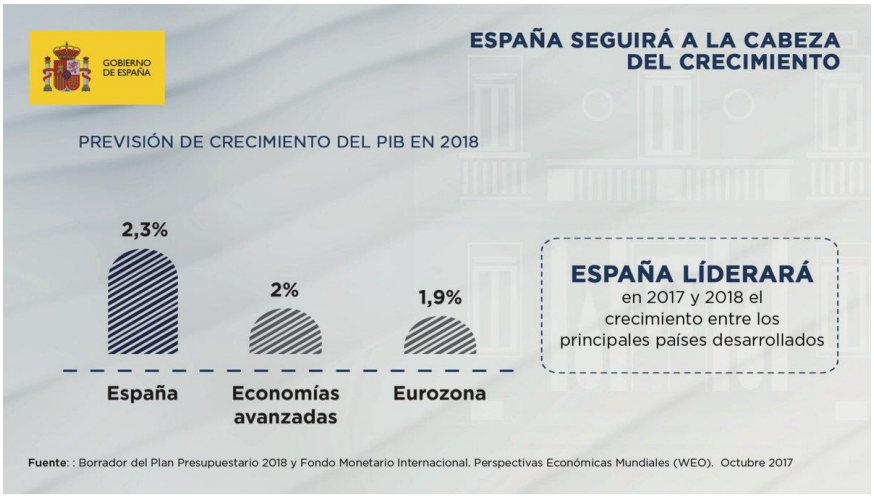

Figure 3: Second sample graph

The graph shown in Figure 3 is displays the GDP growth of Spain in comparison with the rest of Europe and other advanced economies.

This is probably the worst case discussed in this assignment by our group. Without paying attention, one can interpret this as a good graph: it seems simple and very clear. However, when you look at the gap between each group and the corresponding percentage, the representation is plain wrong. This is not that bad, you could say. But the serious problem is that it has been modified with misleading intentions with political purposes. It is inexcusable that a public government spreads an intentional misleading visualization as this one.

There are multiple problems with this graph:

1. The range of the plot has been cut to accentuate the differences between bars. This is not intrinsically wrong, but in this case, it is used with this intention.
2. The value of the left bar does not correspond to the height of the bar.
3. The rounded ends of the bars make more difficult the comparison between them.
4. What are “Economías avanzadas” (advanced economies)?

There is even a misspelling in the word “Liderará”! (which has an extra accent mark)

We can see an approximate correction of the barplot in figure 4.

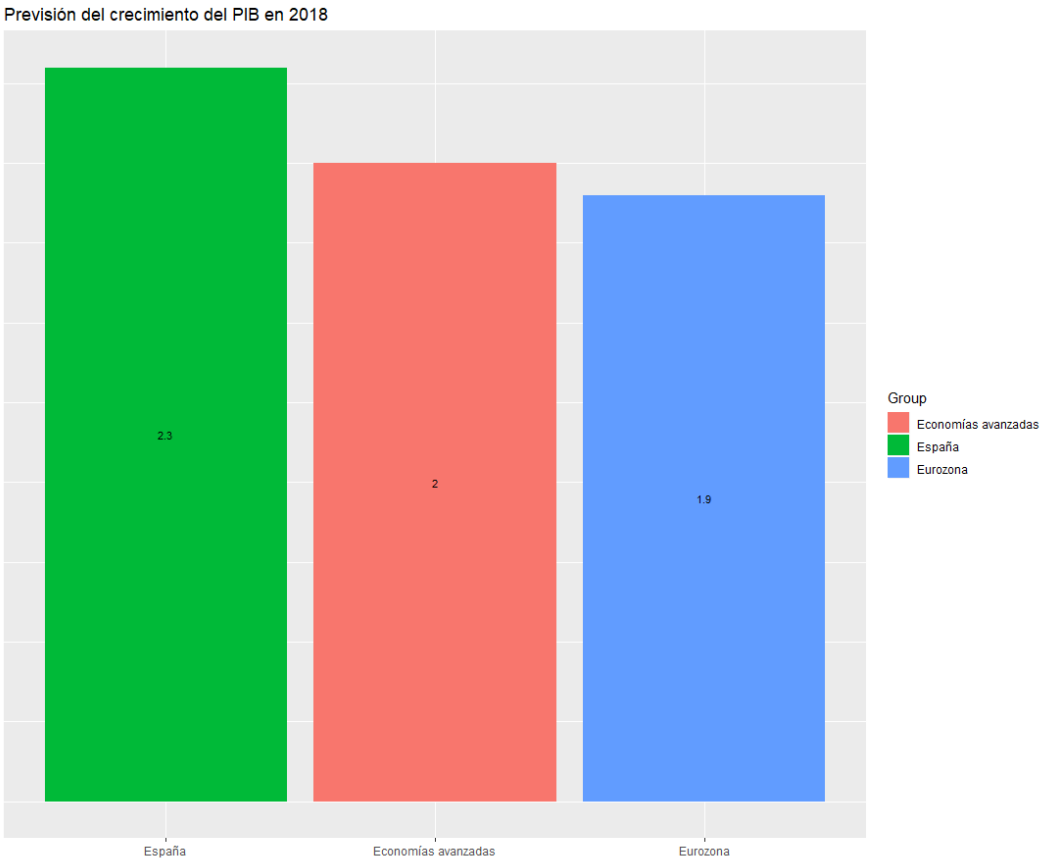

Figure 4: Second sample graph fixed

Appendix:

First Simple Visualization:

```
library(ggplot2)
library(scales)

 values = c(0.42,0.6,0.56,0.33,0.51,0.65,0.61,0.42,0.26,0.49)
 tags =
c("Sweetcorn","Peppers","Pepperoni","Olives","Tomato","Mushrooms","H
am","Pineapple","Spinach","Bacon")

 dataf <- data.frame(values, tags)
 ggplot(dataf) +
   geom_bar(aes(x = tags, y = values), stat = "identity") +
   theme_minimal() +
   xlab("Toppings") +
   ylab("Percentage") +
   scale_y_continuous(labels = percent)
```

Second Simple Visualization:

```
library(ggplot2)

x <- data.frame("Group" = c("España", "Economías avanzadas",
"Eurozona"), "Crecimiento" = c(2.3,2.0,1.9))

ggplot(data=x, aes(x=reorder(Group, -Crecimiento), y=Crecimiento,
fill=Group, label = Crecimiento)) +
  geom_bar(stat="identity") +
  xlab("") +
  ggtitle("Previsión del crecimiento del PIB en 2018") +
theme(axis.title.y=element_blank(),axis.text.y=element_blank(),
axis.ticks.y=element_blank()) +
  geom_text(size = 3, position = position_stack(vjust = 0.5)
```