Information Visualization - Project

Marian Aldescu - MoSIG DS

November 26, 2020

Design visualizations

A. Global insights

1 What is the CO_2 distribution over time? In which period of the year are produced the biggest CO_2 emissions by Westeros missions?

Solution 1

To answer this question, we want to show how the CO_2 level evolves in time and to identify if there are periods with considerable high or low pollution. After taking a look on the dataset, I observed that the missions take place over a period of 8 years, so since the number is't not big, we can even compare how the CO_2 level changes from one year to another in the same visualization.

For this question, time is not an automatic property, because the entire dataset does not evolve with time, so I will use it as a quantity(**Q**), and reserve the **X** axis for it. On the **Y** axis I will map the level of CO_2 for each mission, which is also a quantity(**Q**).

On the X axis, I show the data for one year, because it would not be relevant to see all the emissions on the same axis, but overlapped or side by side. Therefore, we need to differentiate the CO_2 over years, and since the years are nominal(**N**) values, I will use the retinal property(**R**) color.

According to [2], the visual mapping table will look like this:

Name	D	F	D'	X	Y	Z	T	R	-	[]	CP
Time	Q			L							
CO_2	Q				L						
Year	N							C			

Table 1: CO_2 emissions evolution over time

In Figure 1, we can see a sketch of a plot using the previous visual mapping. I only drew the first 4/8 plots, normally another 4 would be overlaid. The user can choose to overlap all the plots in one by pressing **OVERLAP** button. This would help to better compare 2 or more years. Since in total, there are 8 years, the guideline for mapping of Bertin[1] is not respected, because he recommends to use maximum 7 colors in order for the eye to do a proper selection of the elements in the visualization, but this issue can be solved by choosing which year to keep in the plot, by checking the boxes in the right side. By pressing **EXPAND**, the visualization will be restored to the previous state.

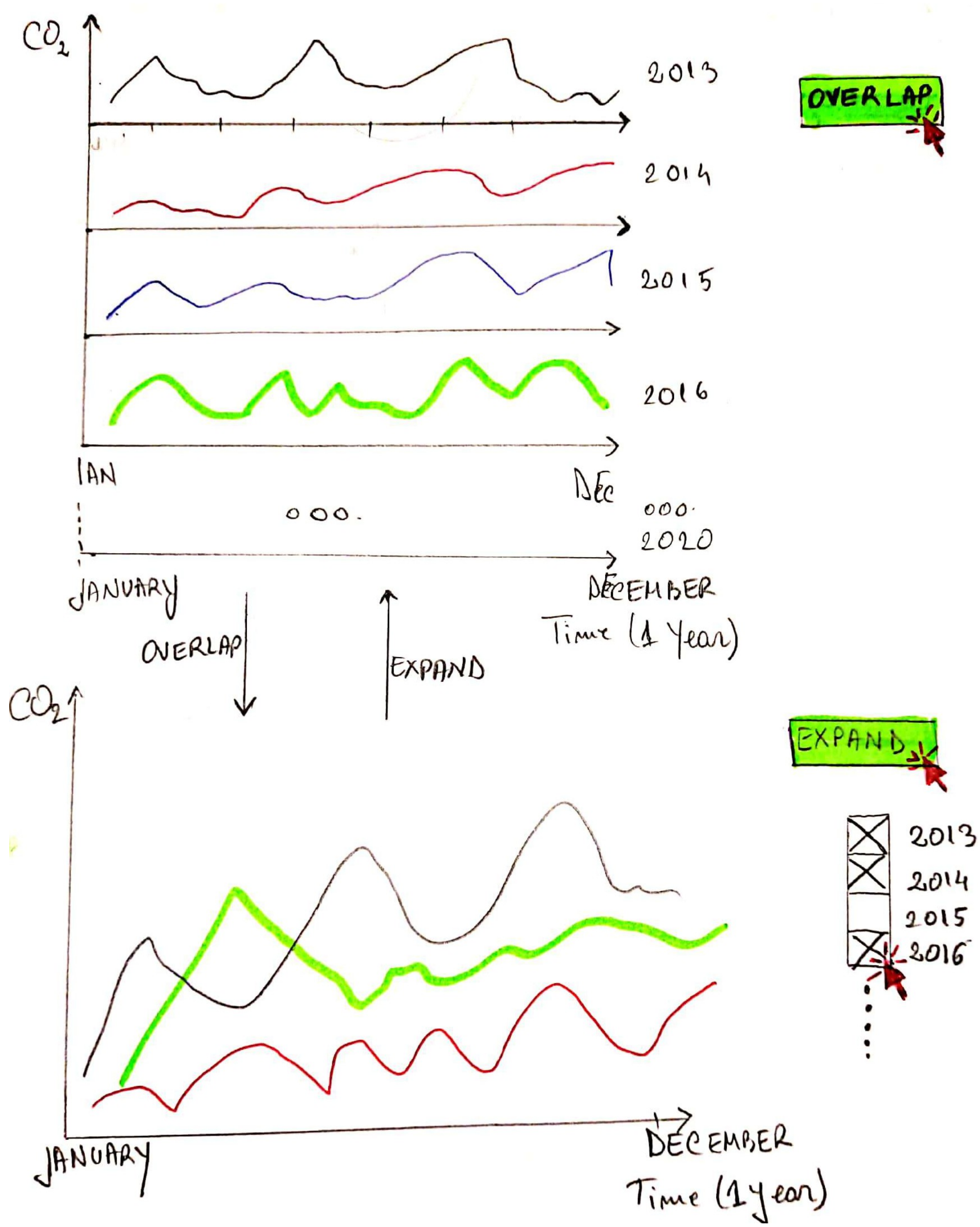

Figure 1: CO_2 emissions evolution over time

Solution 2

Because I chose to plot the CO_2 variation over one year, and a year can be imagined as a cyclic period of 365 days, we can transform the **X** axis in a circle, by transforming the n^{th} day from the 1^{st} of January in radians, with the following formula:

$$\theta = \frac{ArcLength * \pi}{L} \tag{1}$$

where L is the length of the circle given in days(365), and $ArcLength$ is the number of days since 1^{st} of January until the date of the mission. Both time and CO_2 are quantities(**Q**) and the level of CO_2 remains mapped on **Y** axis. In Table 2 we can see the visual mapping.

Name	D	F	D'	X	Y	Z	T	R	-	[]	CP
Time	Q	θ	Q	L							
CO_2	Q				L						
Year	N							C			

Table 2: CO_2 emissions evolution over time

A visualization would look like in Figure 2. In this case, the user can only choose which year to be displayed in the visualization and compare it with the average of all years. The list of check boxes in the right side is exclusive, so one choice will uncheck the previous one.

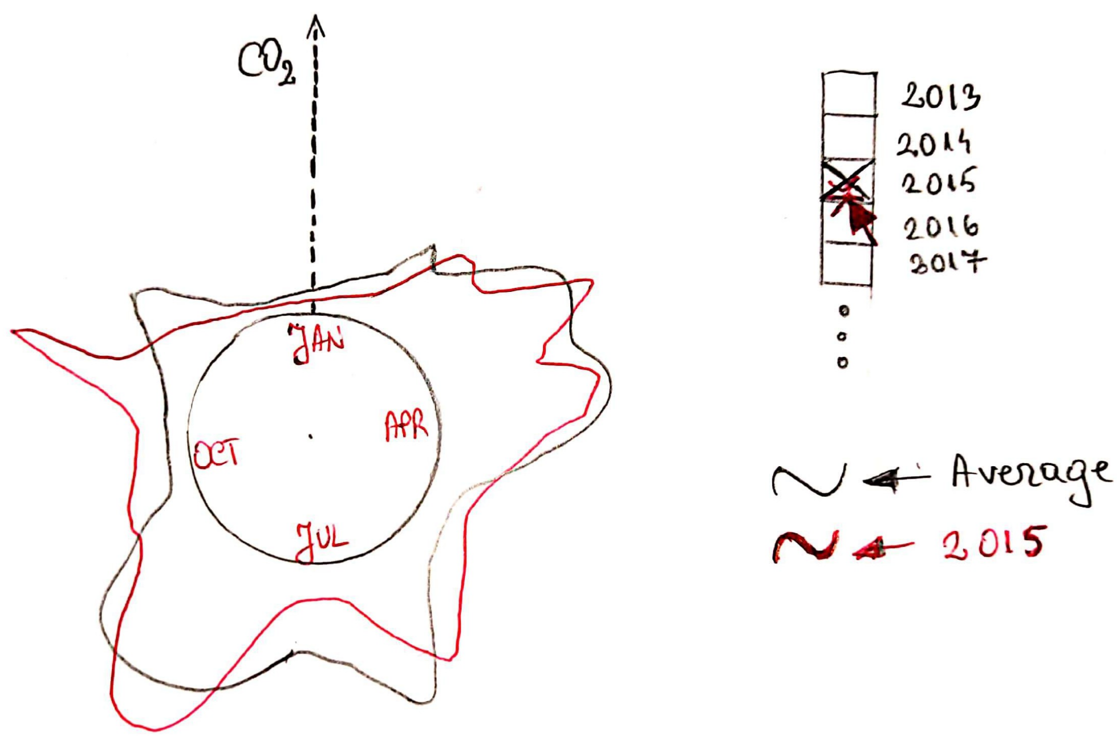

Figure 2: CO_2 emissions evolution over time

2 Where(countries, continents) do most of the users travel? (Distribution of trips to certain places)

Solution 1

A possible answer to this question would result by providing a list with the first n favorite destinations. To enhance user perception, these can be sorted by the total number of missions and each element of the list is accompanied by the name of the country. The number of missions is a quantity(**Q**), therefore we can map it on the **X** axis. Usually, names(of the countries) are nominal values, but I will sort them, thus they become ordinal(**O**) values. Also, each country can be colored according to its continent, which is a nominal value(**N**) and encoded as retinal property(**R**): color(**C**). For more information of the percentage of the number of missions we can add a Control Processing(**CP**) property, which will be mapped on the **X** axis. In Table 3 we can see the visual mapping.

Name	D	F	D'	X	Y	Z	T	R	-	[]	CP
Nb. of missions	Q		Q	L							
Country	N	fs	N		P						
Continent	N		N					C			
Percentage missions	Q		Q	P							tx

Table 3: Favorite missions destinations(Sol. 1)

A sketch is designed in Figure 3. What the user needs first to perceive from the image is the order of the countries. Then, for each bar, he will associate the quantity of missions with the country name. I did not encoded the country name as Control Processing(**CP**) property, since in this case it is more a label of the **Y** axis. Secondary information would be the grouping of the countries into continents by color of the bars. Notices that there are 7 continents, so Bertin rule on colors is respected. Depending on the length of the bars, we can use a logarithmic scale. This visualization is not intended to be interactive.

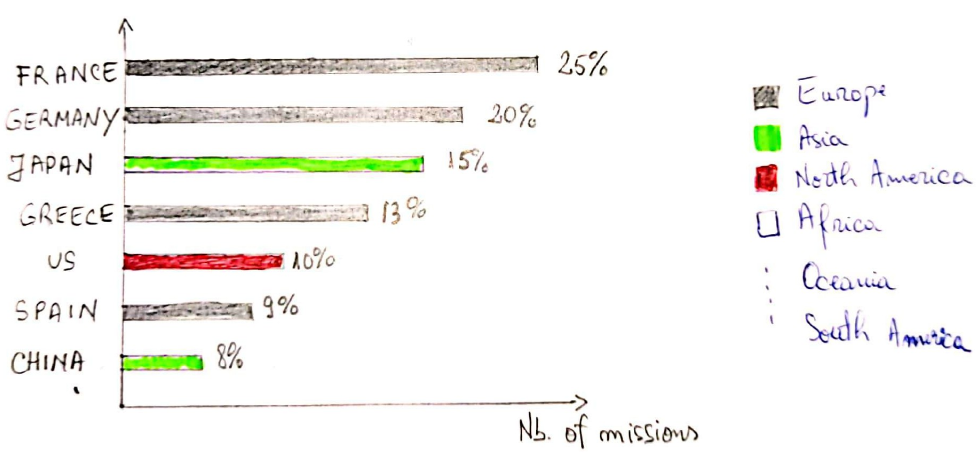

Figure 3: Favorite missions destinations(Sol. 1)

Solution 2
Keeping in mind that we want to show the favorite destinations, we can put a focus on the continents by mapping the number of missions on each continent in a pie-chart. Each slice will have an opening angle(θ) which is directly proportional to the length of the arc represented by the number of missions. For this we can use the transformation in radians: Eq.(1), hence the slice for each continent will be a surface(**S**):

Q→**X**:S, **Y**:S

All the resulted slices will be sorted by the length of the arc(number of missions).

For each country, we calculate the number of missions(**Q**) and map it as a line on **Y** axis.

Q→**Y**:L

Countries are nominal(**N**) values, represented as points on the **X** axis and sorted inside a continent by the number of missions. Each country is then mapped on the arc which corresponds to the continent to which the country belongs, by using Eq (1).

Continents are differentiated by the retinal property(**R**): color and all the countries inside a continent with the same color.

In Table 4 we can see the visual mapping.

Name	D	F	D'	X	Y	Z	T	R	-	[]	CP
Nb. of missions(continent)	Q		Q	S	S						
Nb. of missions(country)	Q		Q		L						
Country	N	fs, θ	N	P							
Continent	N	fs, θ	N	P				C			

Table 4: Favorite missions destinations(Sol. 2)

In Figure 4 we see the visual representation, which it is not interactive.

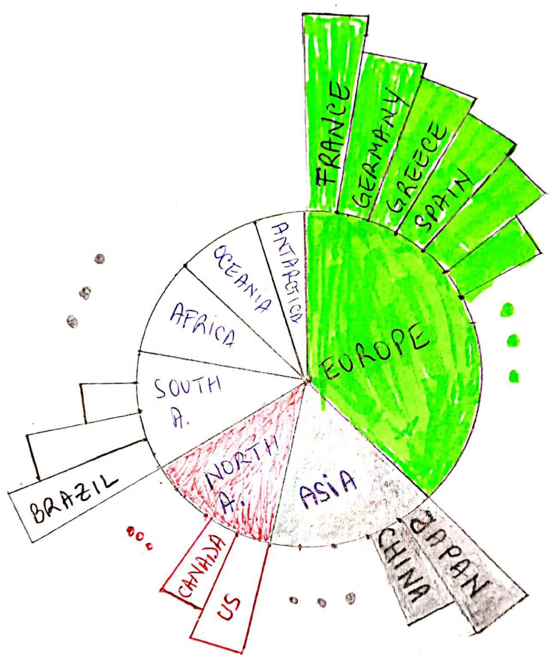

Figure 4: Favorite missions destinations(Sol. 2)

Solution 3
Another option would be to visualize the number of missions on the world map.

We know the destination country of each mission, therefore we can find out its positional coordinates(long, lat)(we ca use an external dataset which maps countries to their approximately geographical central point).

QXlon→**X**:P
QYlat→**Y**:P

The number of missions can be calculated for a certain time interval, hence a **slider** can be added to the visualization, to choose the extremities of the interval. We use the retinal property(**R**) size(**S**) to encode the number of missions in the interval. Table 5 shows the visual mapping just described.

Name	D	F	D'	X	Y	Z	T	R	-	[]	CP
Long.	QX lon		QX lon	P							
Lat	QY lat		QY lat		P						
Nb. of missions(country)	Q	sl	Q					S			

Table 5: Visual mapping: favorite missions destinations(Sol. 3)

In Figure 5 we can see how a visualization would look like for the previous visual mapping. The initial map shows all the continents with a red circle on each country, with the diameter directly proportional to the number of missions to that destination. Using the mouse cursor, a rectangle region can be selected to be maximized. Selecting the European continent, the result it's shown in Figure 6.

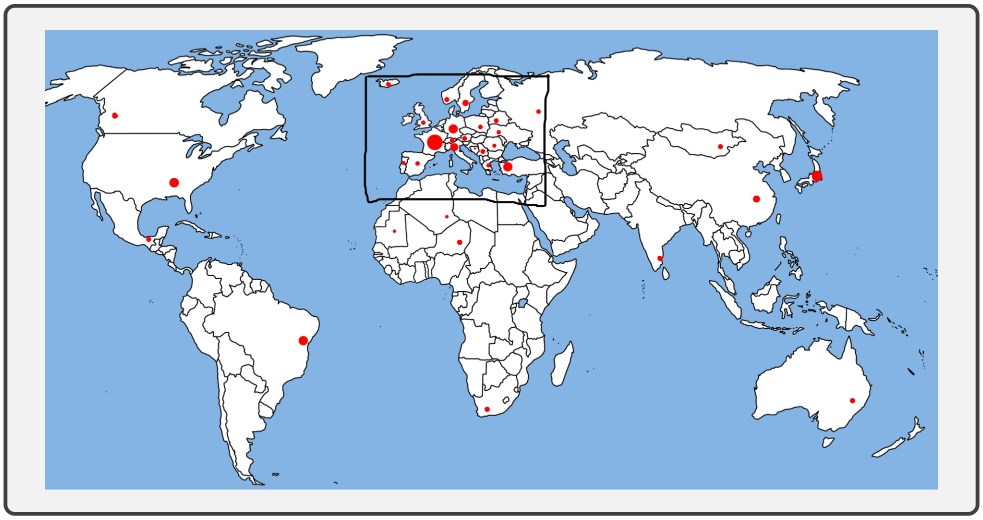

Figure 5: Favorite missions destinations(Sol. 3) - Fictive data

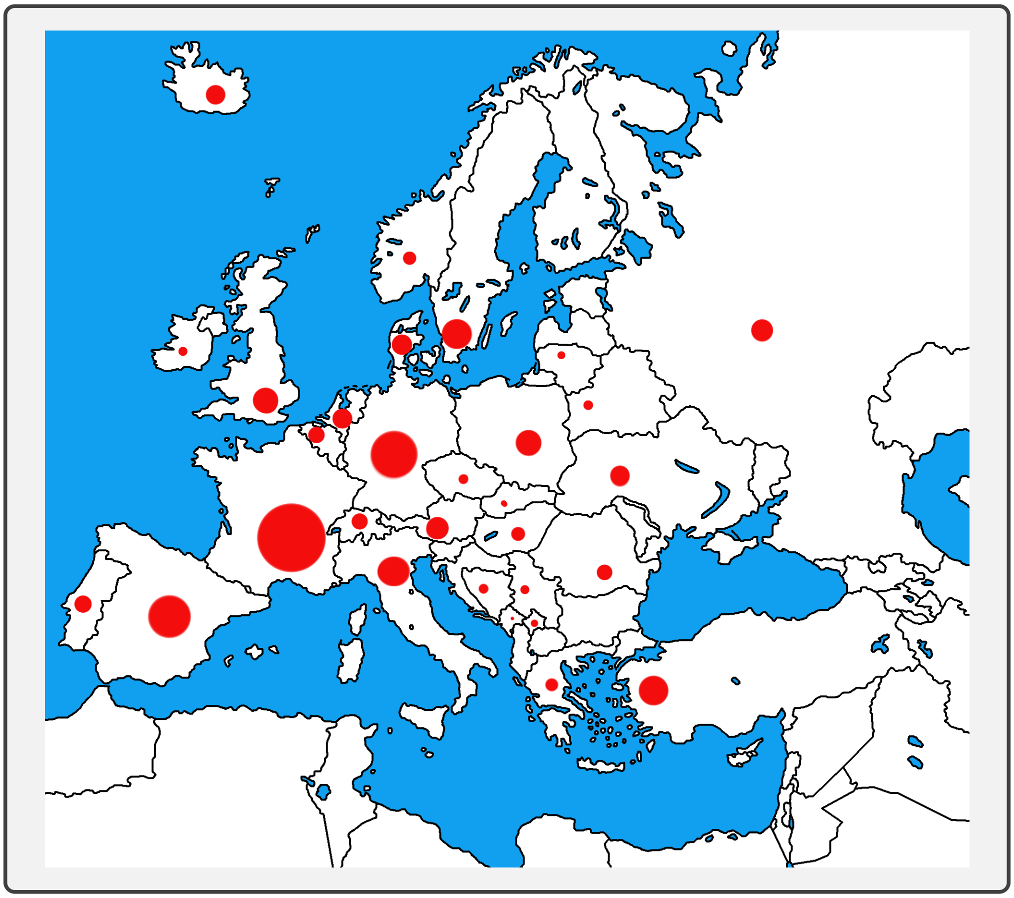

Figure 6: Favorite missions destinations(Sol. 3) - Fictive data

The visualization has a slider attached that is represented in Figure 7. For the plot, it will be counted only the missions that started in the desired interval. The user can change the extremities of the interval by moving the red triangles, or can slide it, in this way providing a continuity effect of the evolution of the number of missions in the range.

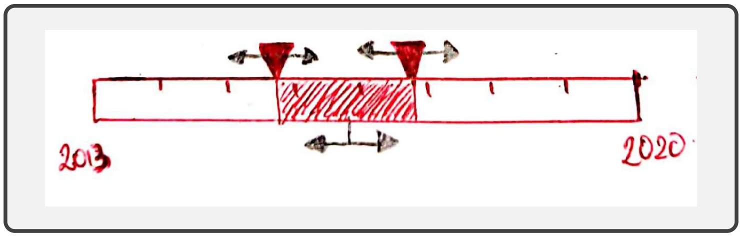

Figure 7: Date glider

B. Focusing on subsets

3 Which person, house, region, rank, title, institution pollutes the most?

Starting from the sample visualization provided in the project, of the CO_2 distribution over people, it is interesting to observe if any of the subsets mentioned in the question is predisposed to higher/lower levels of CO_2 emissions for any of its components.

An example would be to see if people who produce small amounts of CO_2 emissions have lower ranks, or maybe they are part of a certain institutions, or they have a prestigious title.

With this in mind, and considering that a person may be considered a subset of dimension 1, we can design a visualization that shows statistics on all subsets of our dataset and possible correlations between them.

A key point in selecting people who produce large/small amounts of CO_2, is to add a slider in the provided visualization, to take into consideration only a range of users. On the **X** axis, users are mapped as quantity(**Q**), and on the **Y** axis, the volume of CO_2 is mapped also as quantity(**Q**). We want to have the users sorted by the quantity of CO_2 produced, so we apply a sort function, then a slider in order to select the users from a certain interval. In Table 6 we have the visual mapping.

We can interact with the visualization by moving the interval extremities with the cursor. The surface in the interval can be represented with a color different than the one outside the interval.

Name	D	F	D'	X	Y	Z	T	R	-	[]	CP
Users	Q	s,sl	O	S							
CO_2	Q		Q		S						

Table 6: Visual mapping: CO_2 over users(vis.1 in Figure 8)

For the: distribution of CO_2 over ranks, titles, regions and institutions, the visualizations are similar. On the **X** axis, we map the emissions of CO_2 as quantity(**Q**) and on the **Y** axis we map the categories mentioned before as points(the order is not taken into account). I will provide the visual mapping only for the rank distribution in Table 7, because the others(vis. 4, 5, 6 from Figure 8) are almost identical. It is important to mention that for each title, rank, etc., we compute the average of the CO_2 produced by the users inside each group.

Name	D	F	D'	X	Y	Z	T	R	-	[]	CP
CO_2	Q		Q	L							
Rank	Q		Q		P						

Table 7: CO_2 distribution over ranks(vis.2 in Figure 8)

Visualization 2 from Figure 8 could provide some interesting insight about the emissions of CO_2 inside a house. It is like vis. 1, but instead of having the users on the **X** axis, we map the houses, and on the **Y** axis, the average CO_2 quantity produced inside a house, then all the values are sorted, resulting almost an exponential curve. We consider that all the houses are "vertical cylinders" of equal width. By moving the slider extremities in the vis. 1, we keep inside a house just a number of users equal with those that are in the slider range interval. Therefore, some regions in the houses distribution will become whiter in color, if we consider that inside a house, we eliminate randomly the users from a house, its color going from black to gray (if the initial color is black). In Figure 8, vis.2, we can see such an effect.

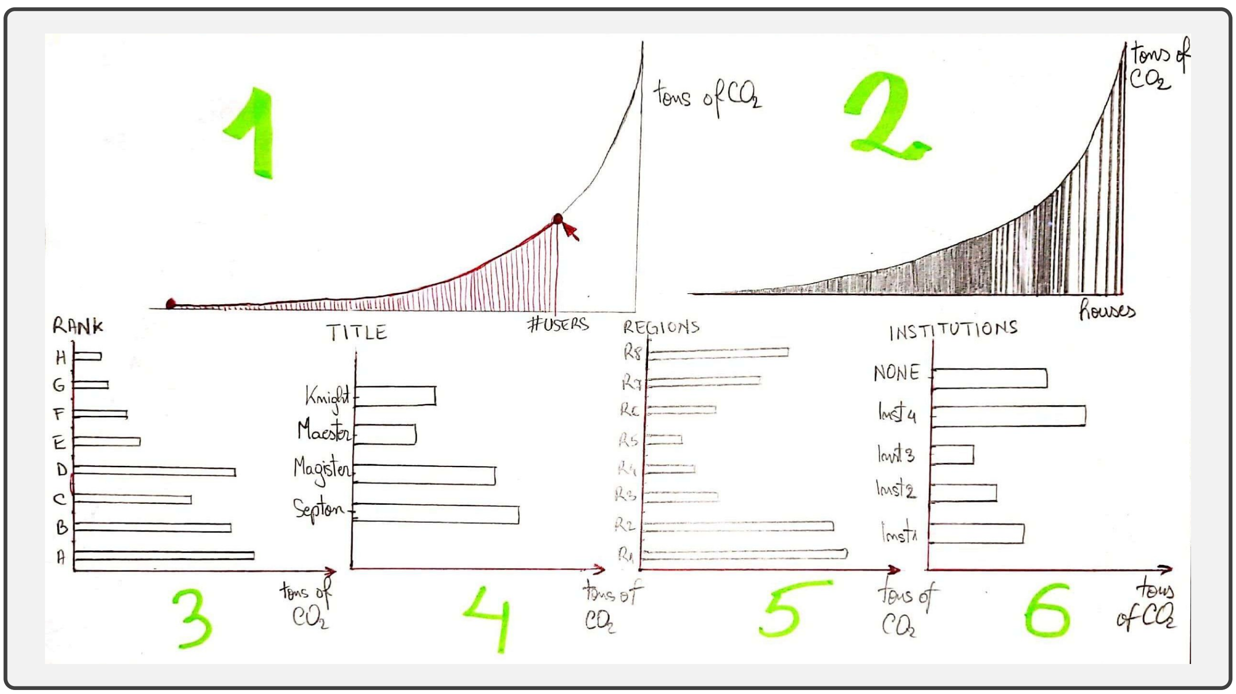

Figure 8: CO_2 - subsets statistics

References

[1] J Bertin. Semiology of graphics: Diagrams, networks, maps. 1967.

[2] Stuart Card and Jock Mackinlay. The structure of the information visualization design space. pages 92–99, 11 1997.