

Information Visualization - Project

Marian Aldescu - MoSIG DS

January 13, 2021

Implement visualizations

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

1 Initial vs actual design

I will start by recalling the initial design of the visualization, which can be seen in Image 1. The D3 implementation does not have major differences in terms of appearance: I kept the same position for each one of the 6 small plots, and also the structure of bar chart. A change that I did, which is not necessarily prominent, but it worth mentioning, is the modification of the bars orientation in the second row from horizontal to vertical: I consider it is important that their variation in height to match the orientation of the above 2 graphs. Also, the bars for the CO₂ emission over houses will vary now in height and not in width, as planned. The main reason is the large number of houses, I think a variation on the vertical axis would be clearer, than on the horizontal.

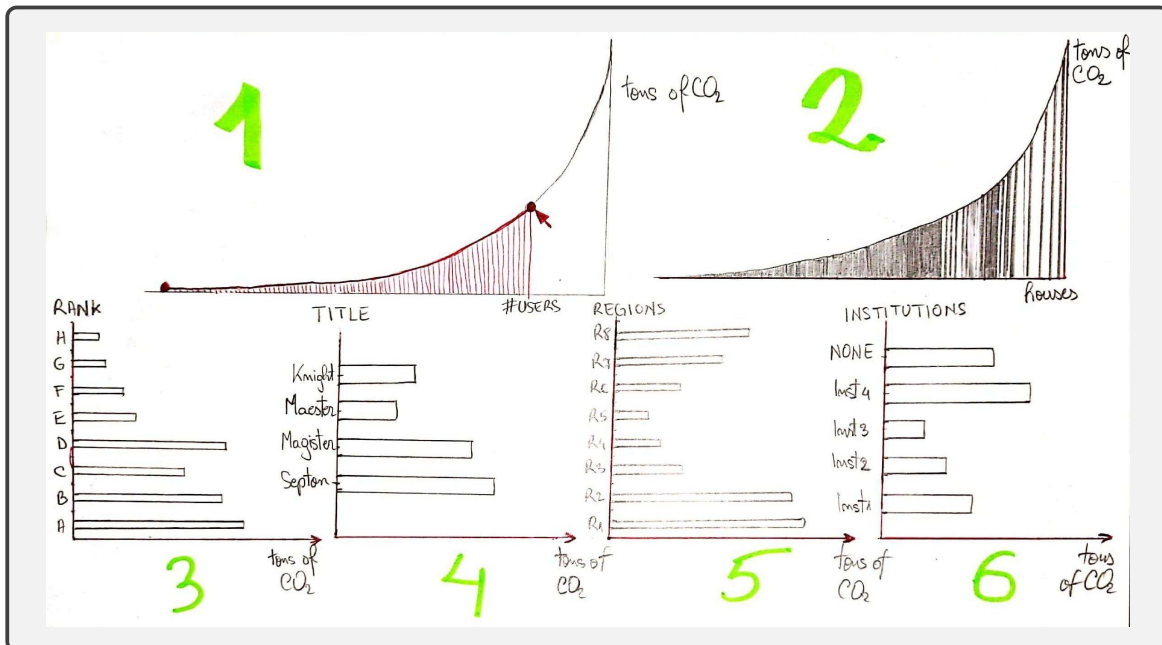


Figure 1: CO₂ - subsets statistics(initial design)

Interaction trade-off

Initially, I intended to use a slider on the users CDF, in order to allow the examination of smaller groups. For this step, I found 2 libraries that allow filtering the subsets, more specifically: Crossfilter[2] and DC[1]. I struggled to work with them, but due to the lack of time, I decided to make a compromise and choose a more "rigid" interaction, but that in the end could actually work. I split the first small plot in 6 equal vertical regions that are sensitive to mouse-over and click. By hovering the mouse above a region, this will change the color to differentiate it from the adjacent regions, and by clicking, the users of that region will be added to the other smaller plots. If a region is colored in blue, this is taken into consideration, if it is white, not.

In Figure 2 we can see an example in which all the dataset is analyzed.

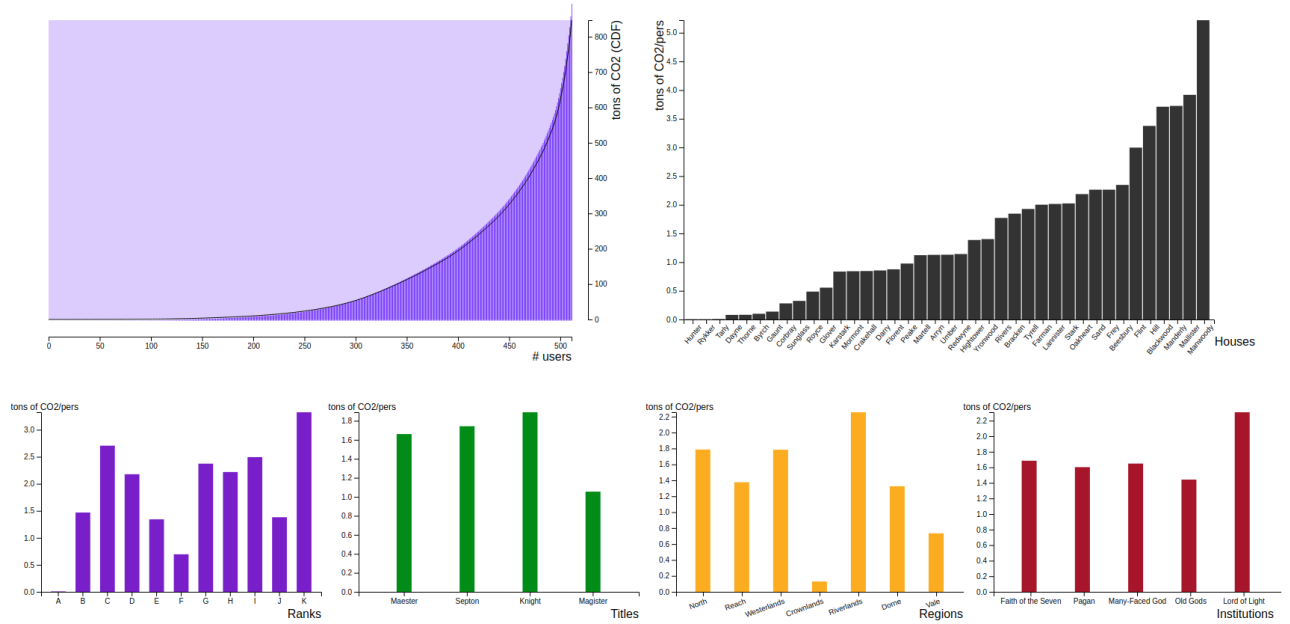


Figure 2: CO₂ - subsets statistics. All users selected

2 From overview to details

Before starting to discuss, note that the CO₂ emissions inside each house, rank title, region and institution are averaged by the number of users from that specific subset, because we do not want the size of the subsets to affect the perception. Therefore, the CO₂ level is expressed as an average of CO₂ tons/user.

When first looking at the visualization, we notice the almost exponential curve of the CDF chart. It is obvious that some users are polluting a lot more than others. Then, we notice that the houses CO₂ emissions fit almost the same distribution: the most pollutant houses can be seen in the right side (e.g. Manwoody, Mallister), on the other hand, the most "green" houses (e.g. Hunter, Rykker) seem to emit negligible quantities of CO₂.

Going to the second row of plots, we can easily see that some ranks produce very low CO₂ emissions(e.g. rank A), and they are very likely to live in the Crownlands region. Differences of the CO₂ emissions can be noticed also in the Titles(e.g. on average, users with the title of Magister pollute less) and Institutions(e.g. user from Lord of Light Institution usually pollute more than those from other institutions).

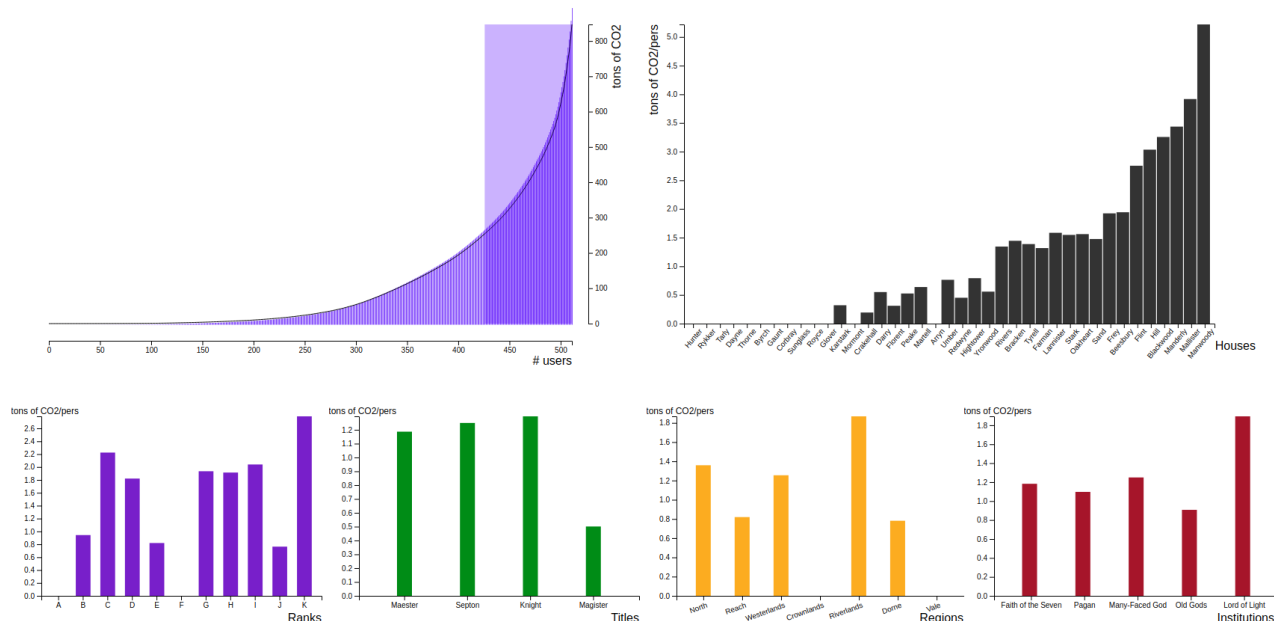


Figure 3: CO₂ - subsets statistics. Last 1/6 of users selected

Now we want to get even more detailed information about subgroups of users that emit similar amounts of CO₂. For example, in Figure 3, I selected the last 1/6 of users: many of them live in Manwoody house or have the title K. On the other hand, we do not encounter them in Vale or Crownlands regions and do not have the titles A and F.

Let's take a look to the first 1/6 of users selected(Figure 4). The intuitions that many of them are part of the houses that pollute the least(from Figure 2) or that they normally have the title A are now confirmed. It is interesting that they also are part of the Lord of Light Institution, the one in which we find the 1/6 of users that pollute the most.

In conclusion, I think this visualization gives a good intuition about the CO₂ distribution over the subsets of users. Even though my way of "brushing" the data is not as versatile as a normal slider, it still offers interesting details about the predominant characteristics of some representative subgroups of users.

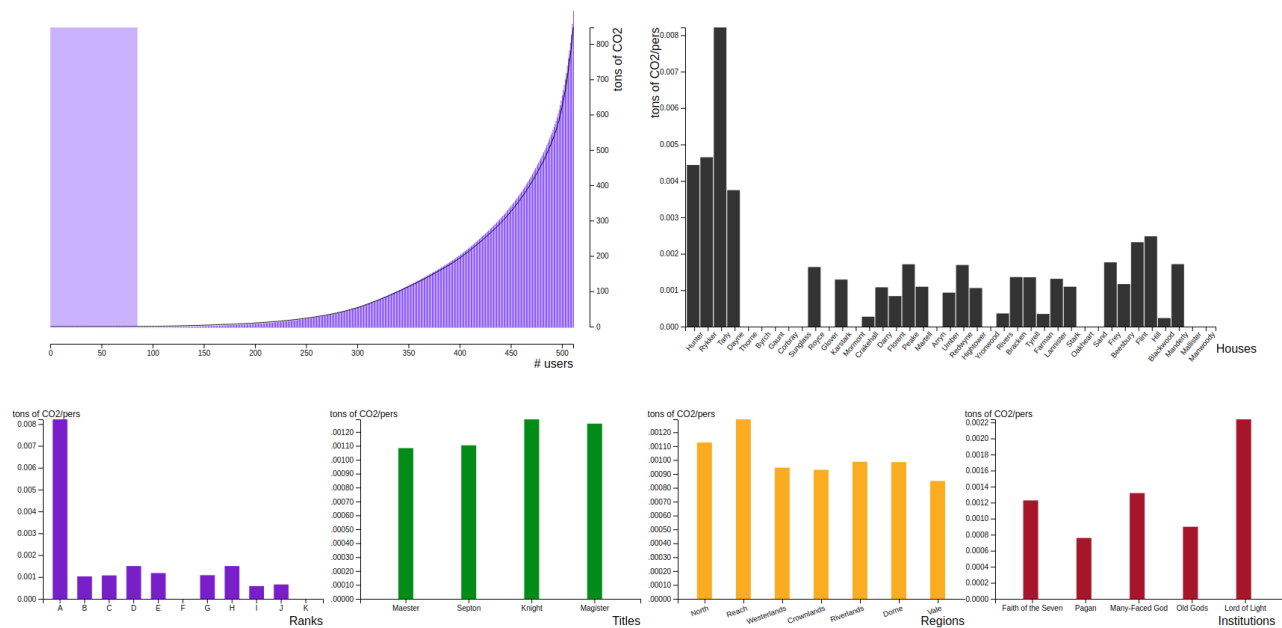


Figure 4: CO₂ - subsets statistics. First 1/6 of users selected

References

- [1] Dimensional charting. <https://dc-js.github.io/dc.js/>.
- [2] Mike Bostock. Crossfilter. <https://square.github.io/crossfilter/>.