Alexis Montavon
Yu Yamashita
Boris Flückiger

Data visualisation project booklet

Week 9 : Data Retrieval

After submitting our proposal, we started thinking of a way to get the data we needed. We decided to scrap the “ratebeer” website in order to get the top 50 beers by country. We could fastly accumulate a total of 4400 beers with their respective features as : type, rating, amount of alcohol, calories...

Week 10 : First Map, Data Cleaning, Beer wheel

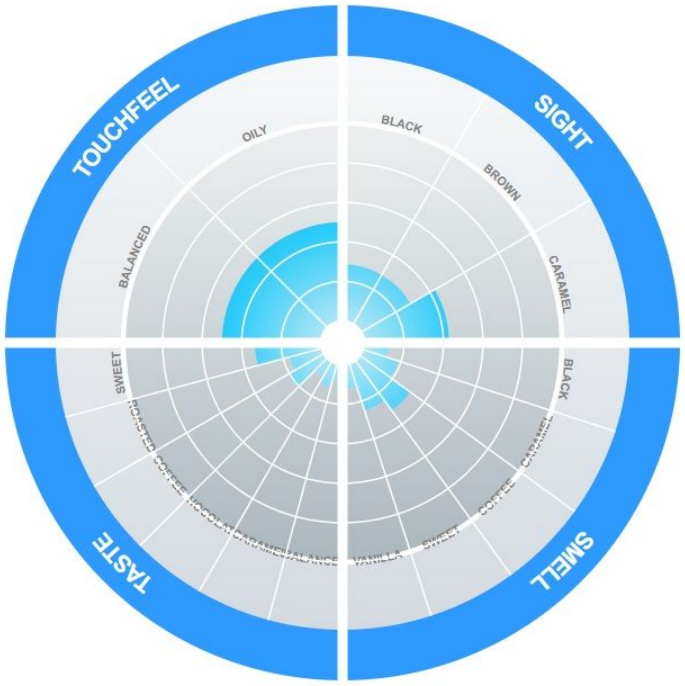
At the beginning of this week we could get a first working visualisation of a map using D3. Up to there we tried to improve it in order to have a more interactive map changing parameters next to the visualization. We could thus represent a choropleth map according to these parameters : “rating score mean”, “alcohol percentage mean” and “calorie mean”. Moreover, we succeed in representing each beer location by a clickable point from which we could get informations about the beer and the country zooming on it (as well as the top five beers). These informations are then displayed next to the map.

In parallel, we focused on having a clean and easy to use dataset. Moreover, we tried to find a interesting way to compare the different beers. We thus decided to scrap the comments used to describe each beer containing informations about the taste, the smell, the sight and feel.

Alexis Montavon
Yu Yamashita
Boris Flückiger

Data visualisation project booklet

From there we found relevant to represent these informations under the form of a wheel partitioned in this way and displayed next to the map (when clicking on a beer from the top beer list of the country.)



To do that, we agglomerated each comments registred on ratebeer.com for each beer. Using a “countvectorizer” from sklearn, we managed to represent each beer with a vector defined on a “prebuilt” vocabulary used to describe the beers. We got this vocabulary from the web site :
https://www.winning-homebrew.com/beer-flavor-descriptors.html
We also added a few more words as a the list of existing fruits for example, to complete this vocabulary. Then our vectorizer, automatically added for each entry (word of our vocabulary) the number of time a given word appeared in the comments of the beer.
From this vector we decided to map each word to a more general category (the one from the wheel) in order to display it as above.

Next objectives :

The next objective for this project, will be to be able to draw edges on the map linking beers by their similarities. This similarity will be computed in function of their type and mostly based on their vector representation from which we will be able to compute a cosine similarity for example.

Alexis Montavon
Yu Yamashita
Boris Flückiger

Data visualisation project booklet

We will also try to improve the css rendering so that the visualisation become more intuitive