COM-480 Data Visualization Milestone 2

Laurent Brock Ben Kriesel Cindy Tang

05.05.2023

The aim of this report is to present our plan for visualizing the 'European Drug Development' Dataset for human medicine from the European Medicines Agency (EMA).

1 Sketches

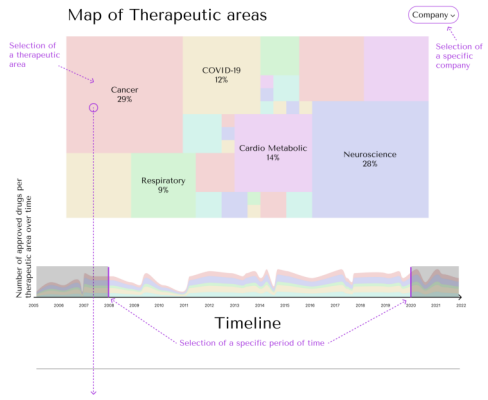
We plan to have a icicle diagram that users can interact with. The starting page will be an overview of the European medicine market by therapeutic area. The tiles will represent one therapeutic area and are scaled by how many drugs were approved for the market. By clicking on a tile the plot will change to another icicle plot showing additional information about that therapeutic area. Figure 1 shows a sketch of what the icicle diagram could look like. Users will be able to change the information the main icicle plot represents through two ways: A drop-down menu will give options for different information to be displayed and a year slider will show at which year the selected medicines were approved. Users could interact with the year slider to display information about a certain range.

2 Tools

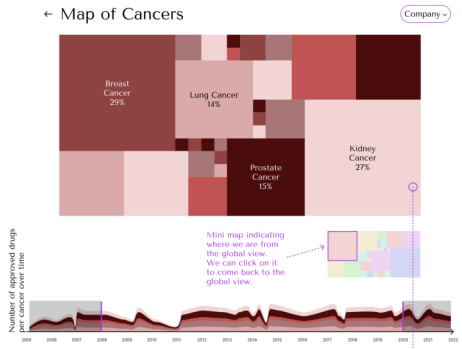
To create the visualisations, we will be using D3.js for the data visualisation aspects, and HTML/CSS/JavaScript for the user interface. Specifically, we will be using the D3.js layout API to create the icicle diagram. We will use concepts of the data visualization class specifically the ones plots, color perception, designing visualizations and tabular data. Preprocessing of the data will be done with Python. The dataset is frequently updated by the EMA. We want to be able to preprocess and visualize the data in such a way, that we can also visualize for future uploads.

3 Minimal Viable Product

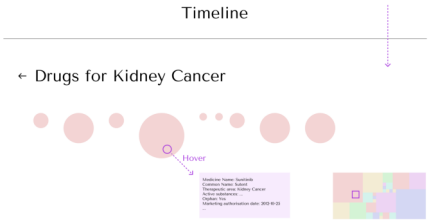
Our minimal viable product will be a basic icicle diagram that users can interact with by clicking and selecting from a drop-down menu on the top right of the graph a sorting criterion. The diagram will display labels for each node, and users will be able to hover over the nodes to display additional information.



(a) Initial visualisation (with timeline)



(b) Example 1: zoom in on cancer details on click



(c) Example 2: zoom into list of drugs for kidney cancer

Figure 1: Concept of our visualisation

4 Task breakdown and extra idea

In order to get to the final visualization we have to complete the following steps:

- Develop preprocessing script for the EMA Dataset
- Develop hierarchical tag system for icicle plot
- Implement icicle plot in D3.js
- Integrate plot in website
- Designing UX and color scheme

An idea to enhance the visualisation is to add an extra, second visualisation to our website. This visualization should display information about the molecular world, specifically which molecules used in drugs are similar to each other and which ones are not. To get the molecular structure we will scrape the SMILES notation for each medicine from PubChem, a global chemical medecine database and use the rdkit package in python to calculate similarity between different molecules. Using multiple dimension scaling we can achieve a 2D approximation to inform the viewer which molecules are physically similar and connect that graph to our main visualization.

5 Functional Prototype Review

At this point our website features the basic skeleton of the visualization/widgets.