Visualization of EU Funding Programmes
186.834 Praktikum aus Visual Computing
WS 2016/17

Daniel Steinböck

January 28, 2017

Abstract

To fund research and technological development, not only in Europe but all over the world, the European Union created so-called Framework Programmes. The data of these programmes, containing information about projects, corresponding topics, funding sums, funding periods and recipient countries, is publicly available, but hard to analyze without visual support. Therefore, a multiple coordinated view approach is developed in course of this project. The different visualization techniques used, like bar charts, treemap, choropleth map and line graph, make it possible to filter, analyze and further explore the available data through brushing and linking. The project was developed in close collaboration with the end users of the *Centre for Social Innovation* and received an overall positive feedback from them.

1 Introduction

Since 1984 the European Union has created eight funding programmes called *Framework Programmes for Research and Technological Development*. The latest one, *Horizon 2020*, runs from 2014 to 2020. The goal of these framework programmes and especially of *Horizon 2020 (H2020)* is to promote research, technological development and innovation [2]. The datasets that hold the information of around 9000 funded projects of different EU member states and third countries is published at the *European Union Open Data Portal* [3]. The supplemental data that is linked to these projects, is further explained in Section 3.1. To explore these datasets, appropriate visualization is advantageous. In cooperation with the *Centre for Social Innovation (ZSI)*, three main questions of interest have been formulated:

- What is the geographical distribution of certain categories (e.g. topics) of funds allocated to entities (e.g. countries) in certain programmes (e.g. H2020)?

- What is the role of certain recipient entities (e.g. third countries) compared to others (e.g. EU Member States)?

- How is the distribution of funds per entity and category progressing over time?

The goal was to develop an interactive web-based visualization that makes it possible to answer these questions. For that purpose the JavaScript library *D3.js* [5] was used, since it offers the possibility to easily create a wide range of different diagram types in web browsers by manipulating the *Document Object Model* (DOM) and is thus very flexible.

2   Related Work

The main technique used in the implementation is *Multiple Coordinated Views* [8]. Generally, if data is represented in multiple windows, it is referred to as *Multiple Views*. If operations on these views are also coordinated, it is then referred to as *Multiple Coordinated Views* (MCV). This technique makes it possible for the user to easily compare data from different representations. MCV has constantly developed over the last 15 years, producing variable exploration, interaction and manipulation techniques [8].

The main concept consists of letting the user interact with the data and offer support in finding information in different datasets of multiple components and types. It should be feasible to discover anomalies, detect certain differences or similarities between datasets, respectively or understand trends. The more attributes need to be visualized, the harder it gets to display these in static visualizations. Thus the used visualization has to be highly interactive. Through the interaction with the data, the user can not only formulate a problem, but also solve it at the same time [10]. Consequently, the user is gaining insight through the interaction with the data.

Roberts [8] distinguishes between two different types of interaction - the indirect and the direct manipulation:

- **Indirect** manipulation lets the user interact with sliders, menus and buttons to filter the data and change how the information is displayed.

- **Direct** manipulation allows the user to select respectively filter elements from the visualization itself. They key principle here, is the technique of *brushing & linking*, where elements are selected and highlighted in one view and the same information gets highlighted in any other linked view.

As presentation design guideline, *Baldonado et al.* [11] suggest using perceptual cues to draw attention to a certain view at a certain time and to use them also to make relationships between different views more visible to the user.

When combining different views, there is a wide range of visualization techniques available to choose from. *Andrienko & Andrienko* [4] pointed out, that it can be difficult to choose between the suitable combination of visualization methods when developing a MCV system. Further, the process of designing involves considerable expense and can not be easily automated.

The other concept that is used in this implementation is the *Dashboard. Few* [7] defines it as "a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance". In combination with the MCV approach, this results in a single information screen, where all essential information is displayed and can be manipulated and explored. Thereby it is important to use appropriate visual representations to support information understanding [6].

3 Implementation

The tool was implemented in HTML and JavaScript using the D3.js library. The following sections cover the structure of data and the used types of visualizations.

3.1 Structure of Data

To understand the chosen visualizations and their interactions, it is necessary to get an idea of how the underlying data is structured.

Programmes: *Horizon 2020* consists of eight main research areas and programmes, respectively. Each of them is further subdivided in subprogrammes, whereupon these subprogrammes can also be subdivided themselves and so on. There are around 270 subprogrammes.

Topics: Topics are assigned to certain subprogrammes. In the original dataset one topic could be assigned to multiple subprogrammes. Due to the lack of percentage contributing to each subprogramme it would not be possible to estimate the precise summed up funding sum for each programme. Hence, the domain expert of ZSI offered a relationship table where each of the 130 topics is assigned to exactly one subprogramme.

Projects: Projects are assigned to certain topics, where each project has a certain start and end date, funding sum and different countries participating in it. As mentioned in Section 1, there are around 9000 projects.

Countries: Countries are assigned to projects, where each country is granted a different share in the project's funding sum.

Funding period: The funding period of each project is specified by its start and end date.

Funding Sum: The funding sum is the main attribute. All other entities mentioned above rely on it.

So here the funding sum is the dependent variable, all others are independent variables. This is important to know for the following section, where the different types of visualizations are discussed.

3.2 Visualization

Since the key questions from Section 1 are hard to answer with a single diagram, a multiple coordinated view approach was chosen. With this kind of visualization it is not only possible to get a broad overview of the data and answer these questions, but also to be able to interactively explore the data and gain more insight to relations between different entities.

The idea was to present all needed diagrams in one single dashboard without the necessity to change between different views. This makes it possible to easily brush data in the different views and immediately see the filtered result in all other. Additionally, a filter bar is placed on top to keep track of the current selections. The overall design follows *Shneiderman's* Visual Information-Seeking Mantra "Overview first, zoom and filter, then details-on-demand" [9].

Descriptions and considerations of the usage of the different views are presented below and are visualized in Figure 1 and Figure 2:

Visualizations:

1. **Programmes:**
Since the data is hierarchically structured, a treemap was used because it offers a compact way of visualization and lets the user easily navigate through the different levels of the hierarchy. To enable faster navigation, there is also the possibility to search for a certain programme or topic in a list.

2. **Topics:**
For reason of comparison and due to the partially huge number of topics, their funding sums are additionally visualized in a separate bar chart with the possibilities to sort the topic names alphabetically and change the scale between linear and logarithmic.

3. **Countries:**
For the main focus on how countries contribute to certain topics or programmes and how they cooperate with each other, there are two separate views. One of them shows participating countries in a bivariate choropleth map, where one color is used for EU member states and another color for third countries. This visualization makes it possible to show the geographical distribution of topics and programmes. Since it is difficult to compare the exact funding sums of different countries with each other using color, a bar chart is used in addition to the choropleth map. The country bars are colored in the same way as in the choropleth map and can be manipulated like the bars in the topic's bar chart.

4. **Progression over time:**
To see the progression of funding sums over time, a line chart with a focus & context approach was implemented. The context shows the total chronological progression of funding sums for the current selections and offers the possibility to apply a filter for a specific time span. The focus shows the filtered time domain and provides the option to zoom and pan within the domain.

5. **Distribution of project funding sums:**
An additional view, which can be switched with the line chart, shows a histogram where the funding sums of individual projects are binned (see Figure 2). This makes it possible to follow the funding distribution and also to filter it by selecting a certain range.

Interaction techniques:

6. **Filtering:**
For filtering the data, the JavaScript library *Crossfilter* [1] was used, due to the fact that it is fast for filtering on multiple dimensions and thus ideal for the application in a coordinated views setup. Whenever a selection is made in one of the views, the filter bar gets updated and offers the possibility to reset a certain selection or reset all of them.

7. **Details-on-Demand:**
The final feature of the dashboard are details on demand, which are displayed in every view if the cursor is hovered over a certain element (country, bar, tile, line).
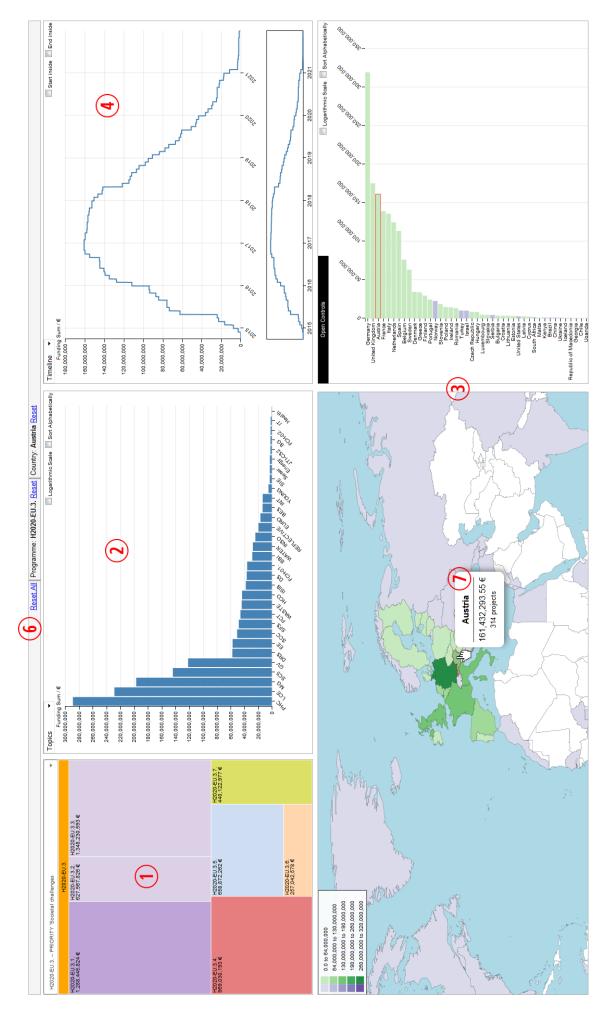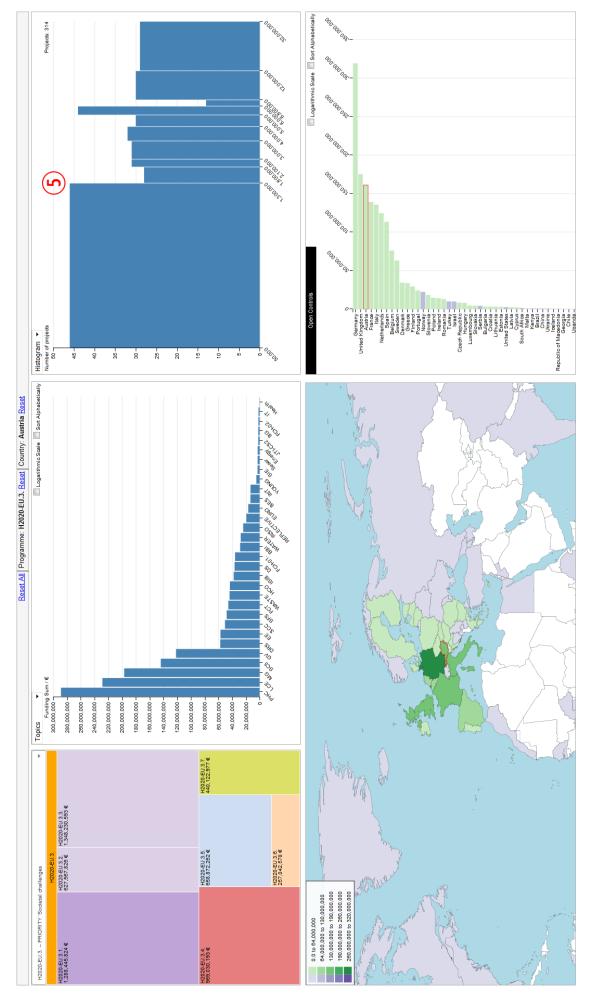
Figure 1: Implemented MCV design.

Figure 2: Overview where progression over time is switched with histogram of project funding sums.

References

[1] Crossfilter. Accessed: 01 September 2016. URL: `https://square.github.io/crossfilter/`.

[2] Overview of EU Funds for Research and Innovation. Accessed: 01 September 2016. URL: `http://www.europarl.europa.eu/RegData/etudes/BRIE/2015/568327/EPRS_BRI(2015)568327_EN.pdf`.

[3] CORDIS - EU Research Projects Under Horizon 2020, 09 2016. Accessed: 01 September 2016. URL: `https://data.europa.eu/euodp/en/data/dataset/cordisH2020projects`.

[4] G. Andrienko and N. Andrienko. Coordinated multiple views: A critical view. In *Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV 2007)*, pages 72–74, July 2007.

[5] M. Bostock, V. Ogievetsky, and J. Heer. D3 Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, Dec 2011.

[6] M. Elias and A. Bezerianos. *Exploration Views: Understanding Dashboard Creation and Customization for Visualization Novices*, pages 274–291. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[7] S. Few. *Information Dashboard Design: The Effective Visual Communication of Data.* O'Reilly Series. O'Reilly Media, Incorporated, 2006.

[8] J. C. Roberts. State of the art: Coordinated multiple views in exploratory visualization. In *Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV '07. Fifth International Conference on*, pages 61–71, July 2007.

[9] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages*, VL '96, pages 336–343. IEEE Computer Society, 1996.

[10] R. Spence. *Information visualization*, volume 1. Springer, 2001.

[11] M. Q. Wang Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for using multiple views in information visualization. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '00, pages 110–119. ACM, 2000.