d3.js and its possibilities in infographics

Creating a diagram showcase using ukrainian refugee data of the ongoing armed conflict

Luis Rothenhäusler

20202459

Technische Hochschule
Brandenburg
University of
Applied Sciences
Fachbereich
Informatik und Medien

Bachelorarbeit - Exposé

Fachbereich Informatik
und Medien
Technische Hochschule Brandenburg

Betreuer: Prof. Julia Schnitzer
2. Betreuer: Hier könnte Ihr Name stehen

Brandenburg, den 14.05.2022

Bearbeitungszeit: dd.mm.yyyy - dd.mm.yyyy

Table of contents

1 Introduction 2

2 Basics 2

3 Goals 2

4 How 2

5 Data 3

6 Diagrams 3

7 Expectations 3

8 Preliminary Structure 4

1 Introduction

The postmodern world produces huge amounts of data every second. Analyzing this data leads to better-informed decision-making in every sector. Yet the wast amounts of created data is often hard to comprehend with the human mind. Infographics is about finding ways to represent this data in visually appealing, yet easily understandable visual representations. Doing this quickly and always up to date can be crucial. There are many tools available to help with the creating of infographics. Some of these tools are software based, others require programming. This thesis will be a deep dive into the possibilities of one of these tools, the 'd3.js'(D3) library for JavaScript. To show its capabilities we will use D3 to create a showcase containing several different graphics.

The following sections describe what the goal of this thesis is, as well as why they should be pursued and how they will be achieved.

2 Basics

D3 stands for "Data-Driven Documents" and "[...] is a JavaScript library for manipulating documents based on data. D3 helps you bring data to life using HTML, SVG, and CSS." [1] Since the first notation of the development of D3 in 2011, the library has seen continuos development. During this time many different resources and usage examples have been created for D3. Yet a lot of these openly available examples are created in online environments with D3 preloaded, are lacking documentation, use different styles of JavaScript or do not react to changing data.

3 Goals

The main result of this thesis will be the creation of a showcase of diagrams created in D3. And there are several questions that will be answered in the process. What are the strengths and weaknesses of D3? How easy is D3 to use? How hard is it to initially get into? When is D3 a viable option for creating graphics?

Furthermore there are certain requirements the showcase should fulfill. Each diagram should be well documented and explained, be easily adaptable to ones personal use and use some aspect of D3 which is not used by the other diagrams. They should also be easily comparable to each other, to allow fast comprehension and understanding of the differences in the creation process. Furthermore all diagrams should be able to react to data changes, so they can continuously update as the data changes.

4 How

To achieve our goals, we will set certain limitations. Besides using a consistent code style and proper documentation, each diagram should be fully independent of the other graphics. This will allow easier adaptation and an easier comparison in effort between the

graphics. To further simplify the comparison, we will use a limited number of datasets. Furthermore the diagrams should be implemented using only JavaScript, HTML, CSS and the D3 library, and not rely on a specific framework. This makes them lightweight and easier to adapt them to all kinds of web based projects. To mimic changing data, the showcase has an option to manipulate the used data-sets.

To make sure we actually make proper use of the D3 library, and be able to evaluate the strengths and possibilities of D3, we will make use of as many different sections of the API[2] as possible.

5 Data

Data is produced everywhere. Most of it is created in a commercial setting. But there are many public data sources available as well. Some is open source or available through api requests. Yet most of the available data on current events is highly individual and not machine readable. This thesis only uses data which is publicly available and machine readable. As this thesis focuses on the technical implementation and to show the possibilities of D3, we do neither need nor claim the data to be correct or unbiased.

Data can be categorized into two main categories. Qualitative and quantitative data. Each category has two subtypes. Qualitative dta can be either nominal and ordinal. Whilst nominal data does not have a natural order, ordinal data does. Quantitative data can be either discreet or continuos. Different types of data can be shown differently in diagrams. In our implementations we will use data of all four kinds. Our first dataset contains the refugees in absolute people(discreet) per country(nominal). we can also express this in refugees in percent(continuos) per country. The second dataset is about the total refugee count(discreet) over time(ordinal).

All data comes from the UNHCR, the UN Refugee Agency[3].

6 Diagrams

When selecting the type of diagram, it is important to pay attention to what diagram can represent which types of data well. Usually it is also worth paying attention to how accurately the presented data is perceived. This depends on the chosen shapes, colors and diagram types[4]. It can also differ depending on the types of data which are presented[5]. As we are trying to create a showcase for the possibilities of D3, we will not need to consider the best didactic choices. Yet the chosen combinations of data and diagrams should still be reasonable and realistically usable, while using a wide range of D3's possibilities.

7 Expectations

I think the possibilities of creating the showcase are quite unlimited. It will be an interesting journey going through and refining the different aspects. But I do think the

benefits of having such a showcase can be quite strong. It can act as a baseline and possible starting point for others in understanding and working with D3.

Furthermore I think it will be interesting to see in which cases using D3 can be reasonably justified. As the visualization tools of other software, i.e. Excel, are quite strong, I assume that there are many cases where it can be easier and faster to not use D3.

8 Preliminary Structure

This is the preliminary structure of the actual bachelors thesis:

- German Abstract
- Introduction
- Basics
  - Infographics
  - D3.js
  - Data
  - Diagrams
- Implementation
  - Diagrams
  - Showcase
- Conclusion
- Sources
- Appendix

References

[1] M. Bostock, "Data-driven documents," accessed:31.03.2022. [Online]. Available: https://d3js.org/

[2] ——, "D3/api.md at main · d3/d3," Apr 2022. [Online]. Available: https://github.com/d3/d3/blob/main/API.md

[3] UNHCR, "Operational data portal," accessed:12.05.2022. [Online]. Available: https://data2.unhcr.org/en/situations/ukraine

[4] J. Heer and M. Bostock, "Crowdsourcing graphical perception: using mechanical turk to assess visualization design," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2010, pp. 203–212.

[5] J. Mackinlay, "Automating the design of graphical presentations of relational information," *Acm Transactions On Graphics (Tog)*, vol. 5, no. 2, pp. 110–141, 1986.