WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

Bachelor Thesis

Visualization of Open Data

Hubert Baginski

Date of Birth: 04.06.1996
Student ID: 1551118

Subject Area: Information Business

Supervisor: Dr. Javier David Fernández García

Co-Supervisor: Dipl.-Ing., B.Sc. Sebastian Neumaier

Date of Submission: 23. January 2019

Department of Information Systems and Operations, Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria

DEPARTMENT FÜR INFORMATIONS-
VERARBEITUNG UND PROZESS-
MANAGEMENT DEPARTMENT
OF INFORMATION SYSTEMS AND
OPERATIONS

Contents

List of Figures

List of Tables

Abstract

This Bachelor Thesis analyses different methods in the field of Data Visualization and Visual Data Analysis. First, it gives an overview about what Open Data is and how it has developed historically. Second, different data visualization techniques and methods are introduced and their use cases, specifically in the context of Visual Data Analysis, elaborated. Consecutively, the objective of this thesis is to create a website that interactively visualizes Open Datasets, which are easy to interpret for users with little or no experience in neither the field of programming nor HTML. Finally, the concrete examples we have developed over the course of this thesis will be introduced and a conclusion will summarize the project and give future research ideas.

1 Introduction

In general Open Data can be understood as data that is publicly available for everyone with the permission to re-use and redistribute it freely, i.e everyone must be able to use, re-use, redistribute and combine it with different datasets. An explicit definition is provided by the Open Data Handbook[1]:

> "Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike"

Furthermore, in recent years Austria's government has put extensive effort into extending its Open Data portals such as data.gv.at[2] and opendataportal.at[3], remove barriers of Open Data and incoorparte best practises such as transparency and accountability[4]. Moreover, data mining has seen rapid development, especially in the context of big data, thus, the amount of raw data that is created by goverments, firms and private users has increased exponentially [Wu et al., 2014].

Consequently, one might say that we live in an era of information overload and the data that is collected is predominantly raw data, which only becomes useful once methods are applied to it and the corresponding insights are derived from it. Furthermore, humans are innately visual creatures, i.e. that they can easily detect patterns and interpret e.g. bar charts, and extract meaningful information through visualizations [Murray, 2017].

Thus, the usefulness of data visualizations is clear, it allows users to interpret complex datasets, given very little or no instructions, through simple visualization methods. Moreover, the complexity of the visualization can be as simple a bar chart, where the information of one single dataset is shown, or depict more complex visualization which are generated from more complex datasets, or through a combination of multiple datasets.

1.1 The Utilization of Open Data

The government, companies and private users upload and share their data through different websites in order to provide higher transparency, incentivize innovation, improve efficiency & effectiveness and generate new knowledge from combined data sources and patterns in large data volumes[5]. However, there is

[1] http://opendatahandbook.org/guide/en/what-is-open-data/

[2] https://www.data.gv.at/

[3] https://www.opendataportal.at/

[4] https://www.europeandataportal.eu/sites/default/files/country-factsheet_austria_2018.pdf

[5] http://opendatahandbook.org/guide/en/why-open-data/

no clearly defined default data format, thus, there is a wide range and variance in formats of open datasets. Consequently, one can access the publicly available data through an Open Data portal and be faced with data formats ranging from tabular or semi-structured data formats to, in some cases, images. Additionally, even if multiple files are accessible in the same format, their individual structure might severely vary, i.e. that the stored rent prices may not be numbers but plain text. Thus, an semi-automated or fully automated Open Data processing algorithm encounters multiple issues and must be supervised.

Summarizing, Open Data has no clearly defined standard data format, thus, it is increasingly difficult for inexperienced users to work with and interpret those datasets. Consequently, the need and demand for open data visualization, which enables intuitive data exploration, exists [Peña et al., 2014].

1.2 Extracting Information through Visualization for a wide Audience

The goal of this thesis is to create a website, that displays open datasets through interactive visualization methods, for users with little or no experience in neither programming nor HTML, and lets them explore the data. This will be partly achieved through the publication of the developed methods under a free license, making it available for everyone, honoring one of the key principals of Open Data [Health and Bizer, 2011] [opendatahandbook.org, 2018].

1.3 Research Question and Objectives

The research question this thesis is going to answer is:

> How can web-visualization methods help in supporting Open Data discovery by common citizens?

To answer this we will first look at already existing projects and literature regarding Open Data visualization, in order to determine the best set of visualization methods. Consequently, the plan is to create a website which will consist of multiple dashboards, wherein several datasets will be visualized in an unique dashboard. Additionally, one dashboard will be incorporated that will consist of multiple datasets in order to derive new information through their combination. Finally, the results will be introduced and elaborated, which will answer the research question, how can web-visualization methods help in supporting Open Data discovery by common citizens.

2 Background and State of the Art Visualization

Furthermore, an extension of certain key principals of Open Data must be added, in order to know which datasets are analogous to its definition and can be used and labeled as Open Data in this project. The most important principles of Open Data are summarized by the Open Data Handbook [opendatahandbook.org, 2018] as:

- *Availability and Accessibility*: the data must be available as a whole, preferably through the download over the internet. Additionally, the data must also be available in a convenient and modifiable format, thus, the possible data formats can range from tabular data formats such as CSV (comma separated value) files which currently is the most commonly used data format online[6], to semi-structured formats, such as JSON (JavaScript Object Notation) files or simple unstructured files, e.g. text files.

- *Re-use and Redistribution*: the data must be provided under an open license, which permits the re-use and redistribution of the data, as well as the integration and combination with other datasets

- *Universal Participation*: everyone must be able to access, use, re-use and redistribute the data, there must be no discrimination against the field of intended application or against persons or groups. E.g. datasets that are restricted by 'non-commercial' or 'only in education' licenses would prevent the application in commercial use, thus, they are prohibited.

Those attributes combined offer a definition for *Openness* of Open Data, which is essential since it guarantees *interoperability*, i.e. the ability of diverse systems and organizations to work together, to inter-operate. Consequently, interoperability is the ability to interoperate or intermix different datasets. With this definition in mind, if we were to access two different Open Data files in the same format, through the definition of openness and interoperability, we are able to ensure that the combination of the two files is, principally, possible.

2.1 History of Open Data

The term Open Data emerged for the first time in a paper from an American scientific agency back in 1995 [parisinnovationreview.com, 2019]. It engaged with the proper disclosure of geophysical and environmental data, and the authors

[6] Dave Tarrent, https://www.europeandataportal.eu/elearning/en/module9/#/id/co-01

heavily promoted "a complete and open exchange of scientific information between different countries" [parisinnovationreview.com, 2019], which was essential for both the analysis and the understanding of the observed global phenomena.

Consequently, Open Data was rooted deeply in the praxis of the scientific community before spreading into other branches. The researchers were the first who identified the potential benefit of both openness and sharing of data [parisinnovationreview.com, 2019]. Thus, the philosophy of Open Data has been long established, ever since the rise of the Internet and the World Wide Web the concept of publicly available information has been advancing, and this had a major impact of shaping Open Data as we know it today [Kitchin, 2014]. The goals of Open Data are similar to those of other open source projects such as: open-source-software, open content, open education and the open web [Lerner and Tirole, 2001]. It has gained immense popularity with the launch of open-data government initiatives such as data.gov[7] which was launched in 2000 and data.gov.uk[8] which was launched in 2010. Consequently, the Open Data development has been both adopted and endorsed by many countries and their respective governments. Europe's development and status as of 2018 can be seen in Figure 1, which shows the current state of Europe's Open Data maturity, classifying its countries into four categories based on the indicators policy, portals, impact and quality [europeandataportal.eu, 2019a]:

- beginner: < 30%
- follower: 31-59%
- fast-tracker: 60-79%
- trend-setter: > 80%

First, the indicator policy is determined through weighting attributes such as policy framework, national coordination, and licensing norms, and it is the furthest developed overall [europeandataportal.eu, 2019a]. Second, the indicator portals considers factors such as its features, usage, the data provision and the portals sustainability. Third, impact includes the strategic awareness, the political, social, environmental and economic impact. Finally, quality considers automation, data and metadata currency and the DCAT-AP compliance, which is the ability that guarantees cross-data portal and platform queries[9]. A more explicit and concise definition according to the European Data Portal is:

[7] https://www.usa.gov/history-of-website
[8] https://data.gov.uk/about
[9] https://joinup.ec.europa.eu/solution/dcat-application-profile-data-portals-europe

"A series of indicators have been selected to measure Open Data maturity across Europe. These indicators cover the level of development of national policies promoting Open Data, an assessment of the features made available on national data portals as well as the expected impact of Open Data" [europeandataportal.eu, 2019a].

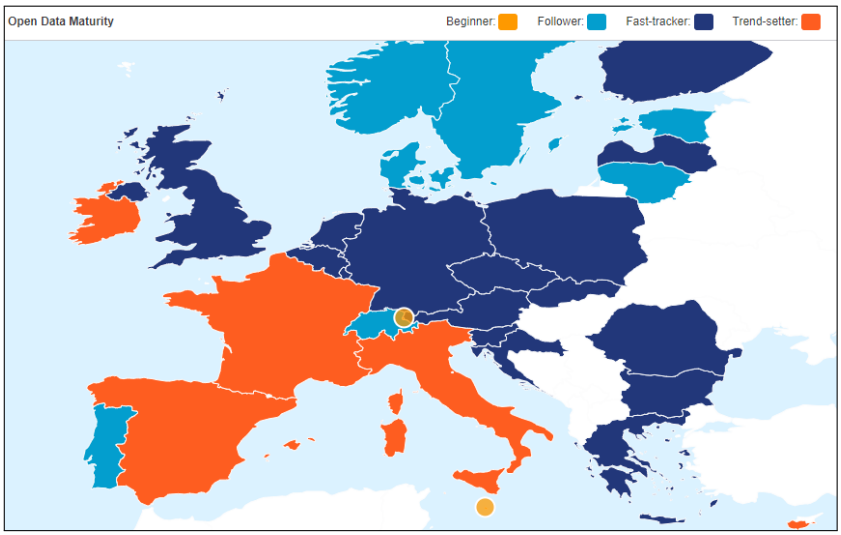


Figure 1: Country maturity map [europeandataportal.eu, 2019a] for Open Data 2018

Furthermore, the exact composition of a country's maturity consists of an evaluation of the priorly introduced attributes policy, portal, impact and quality and the corresponding score can be found in Figure 2.

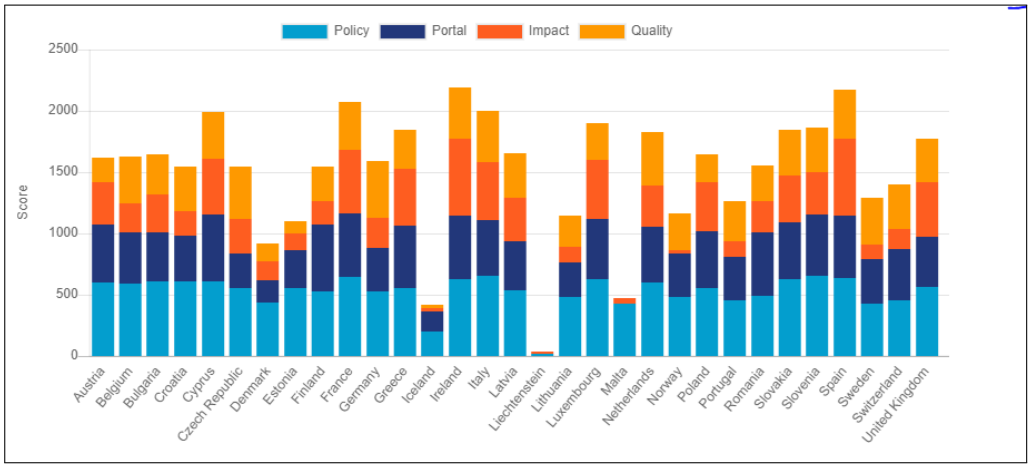


Figure 2: Country overview 2018 [europeandataportal.eu, 2019a]

Summarizing, Open Data has become more prevalent since the introduction of the Internet and Open Data government initiatives. Moreover, it is continuously developed and worked on in order to provide the best service possible as shown by the statistics [europeandataportal.eu, 2019a].

2.1.1 Austria's Open Data Development

This thesis will solely consider Austrian open datasets, thus, the development and current status of the country has to be inspected more closely. First, the European Data Portal will be queried in order to retrieve the specific information regarding Austria, the detailed statistics can bee seen in Figure 3.

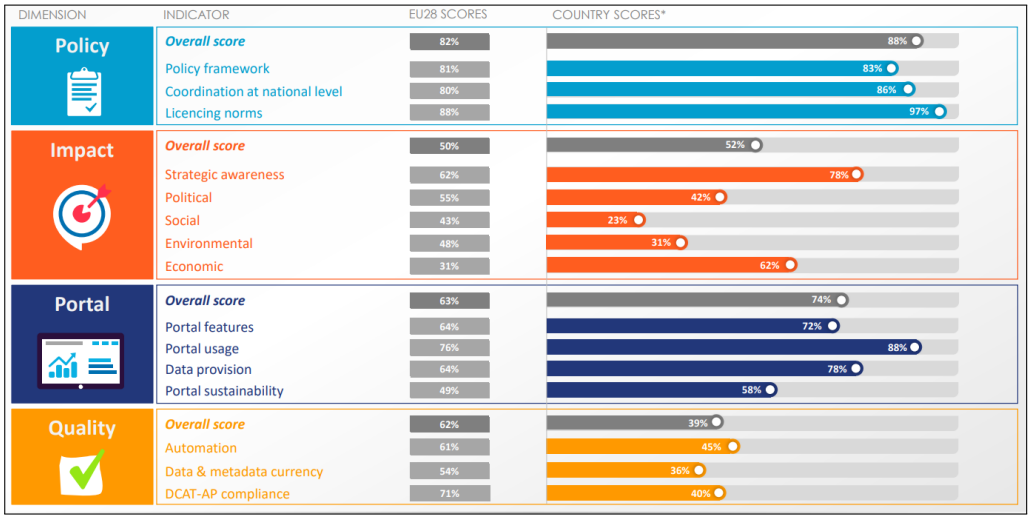


Figure 3: Austria's 2018 [europeandataportal.eu, 2019b] State-of-Play on Open Data

Derived from the statistics in Figure 3 we can see that the dimension which is developed the furthest is *policy*, with a score of 82% Austria is not only a trend-setter in this regard, but has become one within the last four years. However, it also shows that the *quality* of the provided data is lacking crucial features such as the possibility to automate, as well as the DCAT-AP compliance. Thus, Open Data which is provided by the Austrian government does only conform up to 40% with the European regulations, which makes it neither cross-portal queryable nor functional, i.e. that automated queries will not properly retrieve the relevant information from the dataset, since the required format has not been provided. Furthermore, to accurately determine the progress Austria has made in regards to Open Data within recent years one has to compare the development with the total number of published datasets per year on data.gv.at.

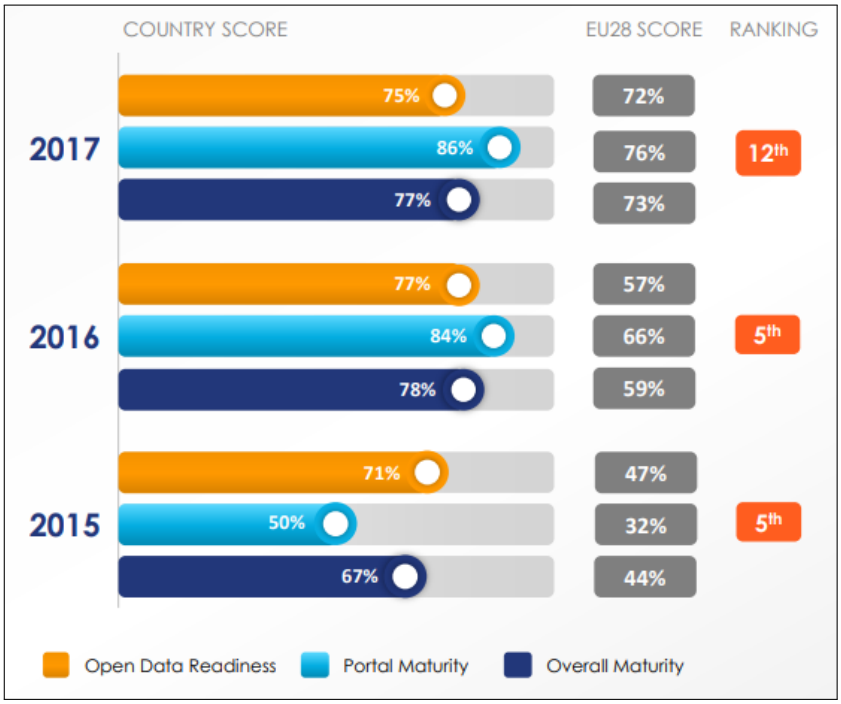


Figure 4: Austria's Open Data Progress [europeandataportal.eu, 2019b] 2015-2017

Figure 4 captures Austria's development in regards to Open Data and one can see that it has been ranked number five in 2015, and even though it has improved significantly since then, it was only ranked number twelve in 2017. This mean that the other EU countries have made more progress, thus, surpassing Austria. Furthermore, this development can also be seen by the by the increase of published datasets on data.gv.at:

Date	number of published datasets
December 2015	2308
January 2016	2316
February 2016	2341
March 2016	2443
April 2016	2478

Table 1: Increase of published datasets to data.gv.at [Beno, 2016]

Summarizing, Austria was one of the leading countries regrading Open Data

development, in all fours main aspects, in 2015 and 2016 where it placed fifth among all European countries. There is continuous improvement to both conform with EU regulations and data quality in general. The governmental Open Data portal for Austrian is data.gv.at[10] and was launched in 2011[11]. Ever since its launch it has seen an steadily increasing amount of published datasets which are utilized by the people, however, there is still much to improve especially in regards to standardized data formats and data structures, thus, barriers have to be removed and best practices incorporated.

2.2 Visualization for Data Analysis

The visualization method severely varies between the data one has and the information he wants to represent. Consequently, there are, generally speaking, four different visualization archetypes, they are:

- *simple charts*: very basic visualization of data in form of bar (see Figure 6), pie (see Figure 5) or line charts; timelines (see Figure 8) and time series (see Figure 7)
- *graphs*: visualization of data with a networked structure consisting of nodes and edges, depicts relation between nodes through incoming and outgoing edges (see Figure 9)
- *geomapping*: visualization of geographic data (geodata) such as traditional maps (see Figure 10, right)
- *three-dimensional data*: visualizes three dimensional data and allows the animation of it (see Figure 10, left)

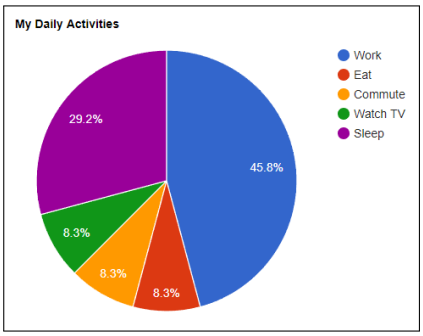


Figure 5: Simple pie chart[12]

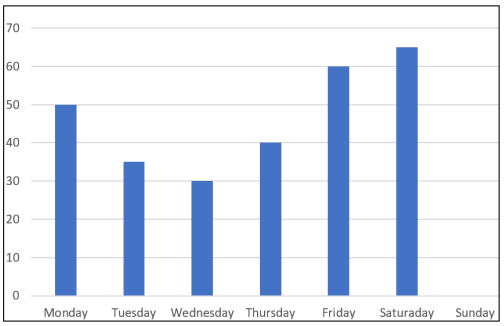


Figure 6: Simple bar chart

[10] https://www.data.gv.at/suche/
[11] https://www.data.gv.at/2016/06/03/5-jahre-open-government-data-in-oesterreich/
[12] https://developers.google.com/chart/interactive/docs/gallery/piechart

According to a study from Keim et al. [Keim et al., 2002] the benefits of bar charts, especially pixel bar charts, are[13]: i) they provide additional information on the data distributions of the dimensions, ii) they show patterns, correlations and trends between small subsets of the data, and iii) allow users to access detailed information about single customers. Consequently, even through a simple visualization information can be derived more quickly and more efficiently compared to only inspecting the raw data.

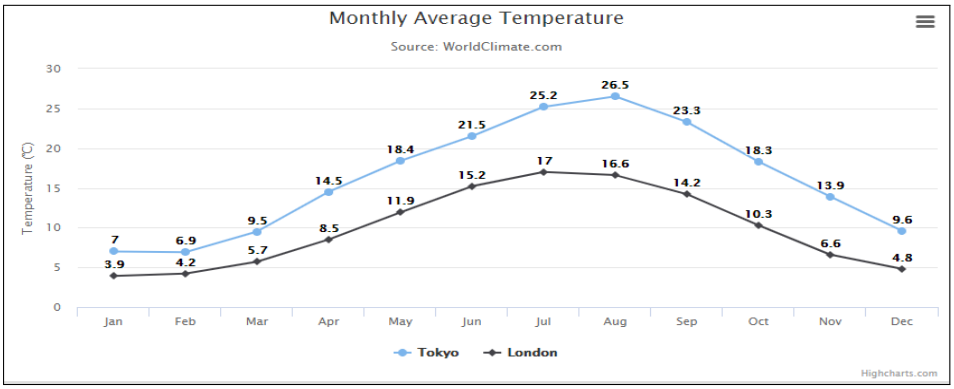


Figure 7: Timeseries example [14]

Traditionally time-series data have been displayed through statistical diagrams like Figure 7, which display the time on the horizontal axis and a single variable on the vertical axis [Monmonier, 1990]. However, time-series can also represent a logarithmic scale on the vertical axis to properly capture relative rates of change. If absolute change ought to be represented an arithmetical scale must be adopted, which will then be represented by the slopes. Consequently, there are various methods to represent time-series data and the apt method has to be chosen and applied to the dataset on a case-by-case basis [Monmonier, 1990].

[13] [Keim et al., 2002], p.33

[14] https://www.highcharts.com/demo/line-labels

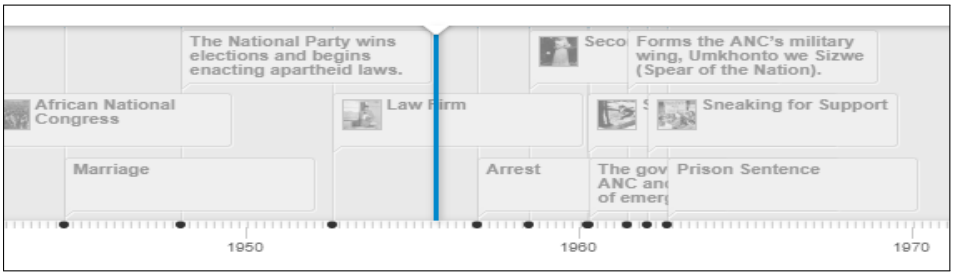


Figure 8: Timeline example [15]

"A timeline is a linear, graphical visualization of events over time"[16], it is used to create a graphical visualization from some record of events. Figure 8 displays a section from Nelson Mandela's life in form of a timeline, it depicts the events of this life over the evolution of time, representing it in an easy to understand design [Karam, 1994] [Plaisant et al., 2003].

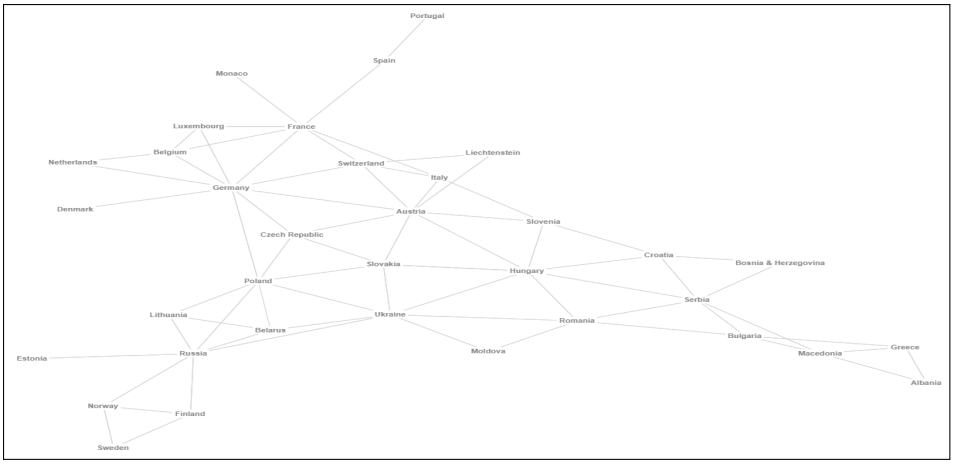


Figure 9: Graph example[17]

Graphs are widely used to represent relations between entities. There are numerous use cases such as models [Kalashnikov and Mehrotra, 2006], i.e.models which use graphs to determine the relations between one or more entities, business processes [Weske, 2012] where the graph determines the control flow of

[15] http://world.time.com/2013/12/05/nelson-mandelas-extraordinary-life-an-interactive-timeline/

[16] [Karam, 1994]p.125

[17] http://arborjs.org/atlas/

the process, or decision trees [Swain and Hauska, 1977] where graphs are used to represent the possible decisions one can make in the given model.

Figure 10: 3D Animation[18]and Geo-Mapping example

2.2.1 Visual Data Analysis

Generally, *visual analytics* can be be described as "the science of analytical reasoning facilitated by interactive visual interfaces" [Cook and Thomas, 2005], i.e. visual analytics is an iterative process that involves the collection of data, data pre-processing, knowledge presentation, interaction and decision making [Keim et al., 2008]. "The ultimate goal is to gain insight in the problem at hand which is described by vast amounts of scientific, forensic or business data from heterogeneous sources"[19].

Furthermore, visual analytics scope extends further than only visualization, it can be seen as an integral approach of combining visualization with both human factors and data analysis. Figure 11 depicts the full scope of visual analysis, however, this thesis will primarily consider *geospatial analysis*.

[18]https://threejs.org/examples/#webgl_decals/
[19] [Keim et al., 2008], p. 77

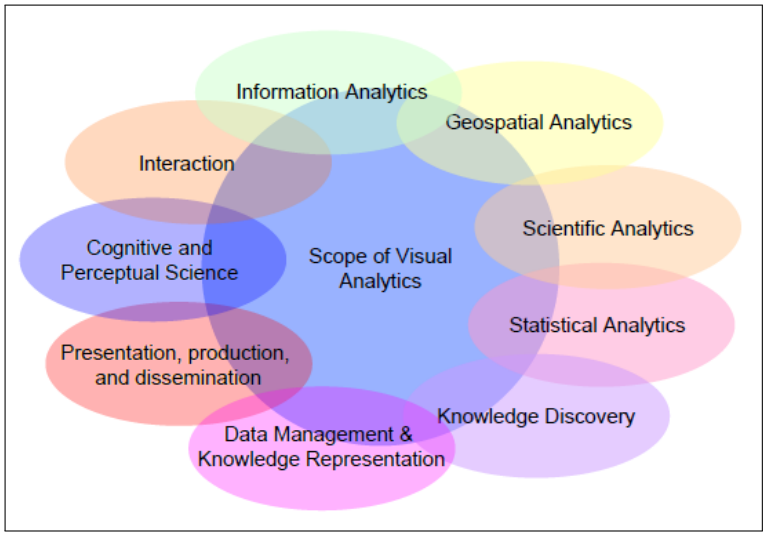


Figure 11: The Scope of Visual Data Analysis[20]

However, Visual Data Analysis poses major technical challenges such as rapidly increasing amounts of generated and collected data, and the thusly increasing growth of datasets, computational power and storage capacity which comes at great costs [Keim et al., 2006]. "Visual analytics combines strengths from information analytics, geospatial analytics, scientific analytics, statistical analytics, knowledge discovery, data management & knowledge representation, presentation, production & dissemination, cognition, perception, and interaction"[21]. Moreover, it is a goal oriented process to derive new insights into heterogeneous, contradictory and incomplete data through the aggregation of automatic analysis methods and human background knowledge [Keim et al., 2006].

Summarizing, there is a great variety of Visual Data Analysis methods, however, for the purpose and the context of this thesis the primary focus will be set on geospatial analysis through geovisualization. Consequently, creating interactive maps that display Open Data which can be discovery by the user intuitively, with the possibility to upload his own data, will be the primary objective. Lastly, simple charts, in form of time series will be implemented in order to accurately and most optimally represent two dimensional data with a focus on time, such as for example weather data.

[20] [Keim et al., 2008], p. 79
[21] [Keim et al., 2006], p.13

2.3 Geo-Mapping

"Human vision and domain expertise are powerful tools that (together with computational tools) make it possible to turn large heterogeneous data volumes into information and, subsequently, into knowledge (understanding derived from integrating information)" [MacEachren and Kraak, 2001]. In their paper, MacEachren and Kraak estimate that 80% of all digitally generated data includes geospatial referencing, i.e. they include either geographic coordinates, addresses or postal codes, and that the combination with cartography will return considerable benefits [MacEachren and Kraak, 2001].

Geo-Mapping is a technique to visualize geographic data onto a canvas, similarly to a map. Moreover, the availability of qualitative geospatial data has increased dramatically over the past decade [MacEachren and Kraak, 2001], which allows us to bind data from a dataset onto a map in form of points, lines, polylines, icons and heatmaps. Furthermore, the visualization of the data is possible through the utilization of different techniques, such as:

- *points, lines and polylines*: are used to represent at least one or many data points on a map, i.e. icons or geometric forms can be placed on top of the coordinates e.g. add a school icon for every school in the dataset; lines display a connection between two data points e.g. the path from one subway station to the next; polylines display the interconnection of multiple data points e.g. a countries border or an area.

- *choropleth*: a thematic map with areas that are shaded or patterned in proportion to the measurement of the statistical variable being displayed onto the map, such as population density, number of schools or per-capita income. This visualization method allows the display of data per geographic area [Dent et al., 1999]. Consequently, the granularity of the choropleth can be as low as e.g. population in billions of the seven continents, or as fine as e.g. number of schools per district in a city, and has to be evaluated and chosen individually in order to most appropriately convey the information one intends to.

- *heatmap*: A heatmap displays a map and merges geographically nearby data points into clusters, in order to represent the density of the distribution in an easily understandable format. The color red denotes the highest accumulation of data points in a given cluster, and as the clusters shrink, the color gradually fades into yellow, green and eventually blue.

2.3.1 Points

Datasets that contain points are the most common type of that that we have encountered. It is used to depict the exact coordinates of specific, mostly static, objects such as schools, traffic lights, parks or stores. Furthermore, it is the, compared to the other types, easiest to visualize, which can be achieved through a plane marker, a marker that depicts a cluster of points in close proximity to each other, a customized icon or a heatmap. Figure 12 depicts plane markers on the left, and custom icons on the right. Figure 13 captures clusters of datapoints, and depicts them in a gradually decreasing colour as the cluster shrinks.

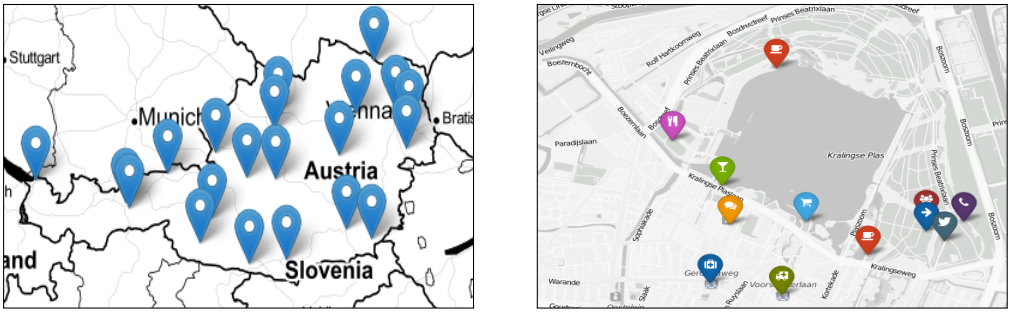


Figure 12: Customization options for point markers[22]

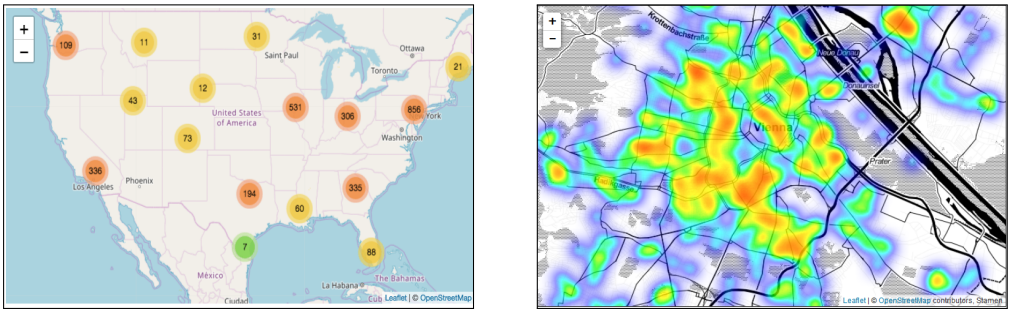


Figure 13: Point-cluster visualization[23]

[22]https://github.com/lvoogdt/Leaflet.awesome-markers
[23]https://www.datacamp.com/community/blog/course-maps-leaflet-r

2.3.2 Polylines

A polyline is a "line consisting of multiple segments, used to compose images on screen"[24]. It is widely used to depict distances between multiple points, such as time-varying data [Yagi et al., 2012], visualize orthogonal and force directed graphs [Tollis, 1996] or visualize multivariate correlations in parallely indexed points [Zhou and Weiskopf, 2018]. However, the main implementation of polylines in this thesis will be to depict the connection between datapoints, e.g. visualize the subway network of Vienna or the cities public sidewalks (see Figure 14).

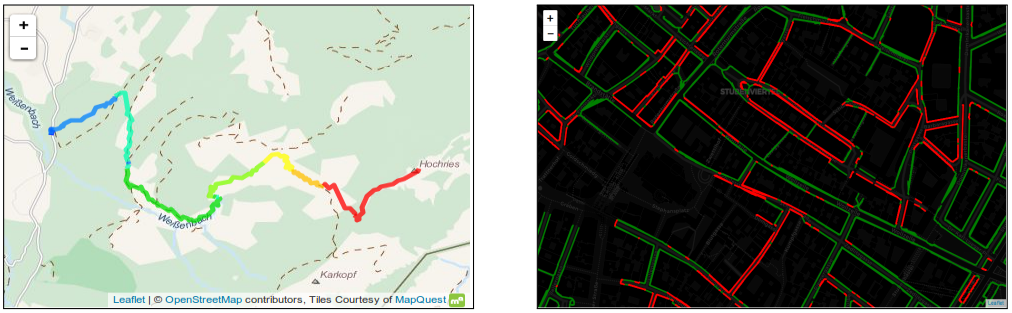


Figure 14: Polyline visualization[25]

2.3.3 Polygons

A polygon is a "figure, especially a closed plane figure, having three or more, usually straight, sides"[26], however, it is not limited to exclusively straight lines. Moreover, it is the most common type of visualization in regards to entities, specifically entity borders [Charaniya et al., 2010]. Polygons are commonly used to depict GeoJSON features, e.g. country borders or continents. The data set has a feature that contains multiple points (latitude-longitude pairs) that, if sequentially combined, form the shape of the object.

[24] https://en.oxforddictionaries.com/definition/polyline

[25] https://stackoverflow.com/questions/53033031/cartography-vizualizing-speed-of-movement-with-color-scale-on-map-in-python?rq=1

[26] https://www.dictionary.com/browse/polygon

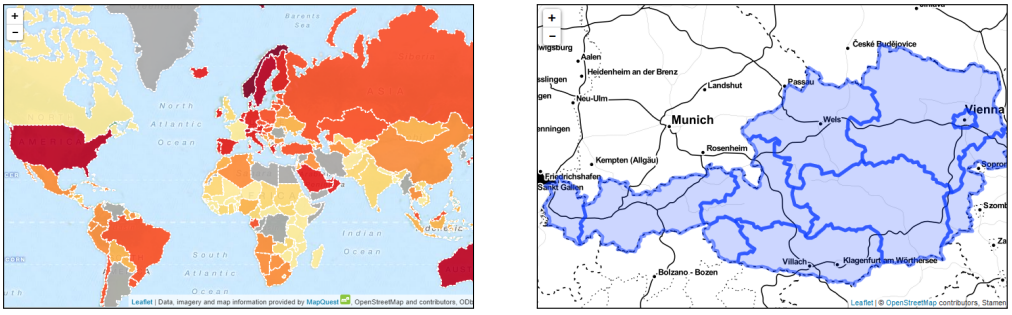


Figure 15: Polygon visualization[27]

Summarizing, geo-mapping allows for a big variety of different visualization methods which enable Visual Data Analysis. Each method has unique benefits and drawbacks, and its implementation has to be evaluated and tested, on a case-by-case basis for any dataset, in order to represent the data correctly and convey meaningful information.

2.4 Related Work: Visualization for Open Data

This thesis and the correlating project has been strongly inspired by the BarcelonaNow project, which was successfully launched in 2018 [Marras et al., 2018] (see Figure 16). However, this project differs in multiple key aspects, such as: this project does not offer a framework for data exploration, however, it combines a pair of front-end and back-end architecture in order to visualize the collected data, which is stored and accessed locally during development or through a server during production. Furthermore, it strongly focuses on static data, i.e. that for any given datasets their format is final and their visualizations can not be changed by the user. Moreover, a dashboard was implemented which represents an example of data combination, i.e. multiple datasets have been combined in order to generate new insights, which were not available earlier.

[27] https://blogs.sap.com/2014/10/06/sapui5-choropleth-map-control/

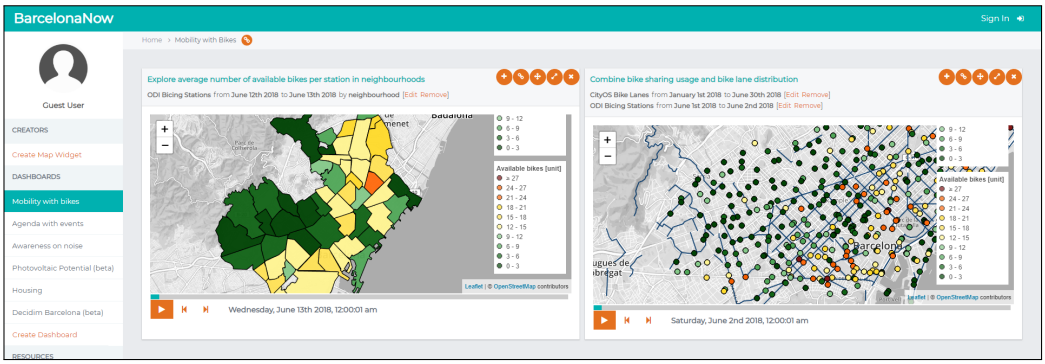


Figure 16: BarcelonaNow - Mobility with bikes dashboard[28]

Second, a framework was developed by Heil and Neumaier which allows for the automated visualization of Open Data [Heil and Neumaier, 2018]. It crawls the datasets from the two Austrian Open Data portals, data.gv.at and opendataportal.at, and classifies the CSV's columns and the metadata descriptions using the geo-entities they have accumulated in their specifically created knowledge graph. Consequently, an elastic search is executed, which contains all the cell values of the table, the potential geo-labels, the metadata of the CSV, and any additionally extracted geo-labels from the metadata. Finally, it is combined with the web application, which one can find at reboting.com, and visualized on the website.

Lastly, there are numerous community based applications, which were created and uploaded by private users on data.dv.at[29], which visualize at least one or multiple datasets. Those applications range from relatively simple only visualizing one dataset, e.g. Vienna's public swimming places and pools, to rather complex with live updates of public transportation, calculating the shortest path of transit and updating the user of the fastest way of reaching the target destination through push-notifications. Those applications show the innovation potential that the provision of governmental Open Data enables.

[28] http://bcnnow.decodeproject.eu/
[29] https://www.data.gv.at/anwendungen/

3 Exploring different Visualizations for Open Data

The intention of this thesis is to create a website that displays Open Data through different visualization methods in order to enable Visual Data Analysis and convey the information in a visualized manner to an inexperienced user. The site should be easily understood as well as its contents be both interactive and intuitive. In order to determine which visualization method fits the best for each individual dataset, we had to find some datasets and determine their content type, i.e. whether the data contains points, polylines or polygons. This influences the visualization style severely since points have exactly one latitude-longitude pair for each point, where as both polylines and polygons contain multiple points in the same row.

The applied visualizations will primarily be Geo-mappings, since most of the datasets are accessed in a GeoJSON format, i.e. maps with different style elements and layers plotted over them, conveying multiple unique perspectives to the chosen dataset. Furthermore, through personally testing already existing projects and their visualization methods, especially the BarcelonaNow project, the visualization styles we chose and that will be used are heatmaps, choropleth maps and regular markers, whether they are represented through polygons, circles or icons.

Thus, a workflow model has been developed which allows the discovery of open datasets and their visualization, which will support the development of an Open Data dashboard. One dashboard will be used to depict the information of a dataset through the pair of most meaningful visualization methods and styles. Furthermore, the website can be extended by multiple dashboards, however, for this thesis we focus on the development of four dashboards and their respective visualizations.

3.1 Workflow

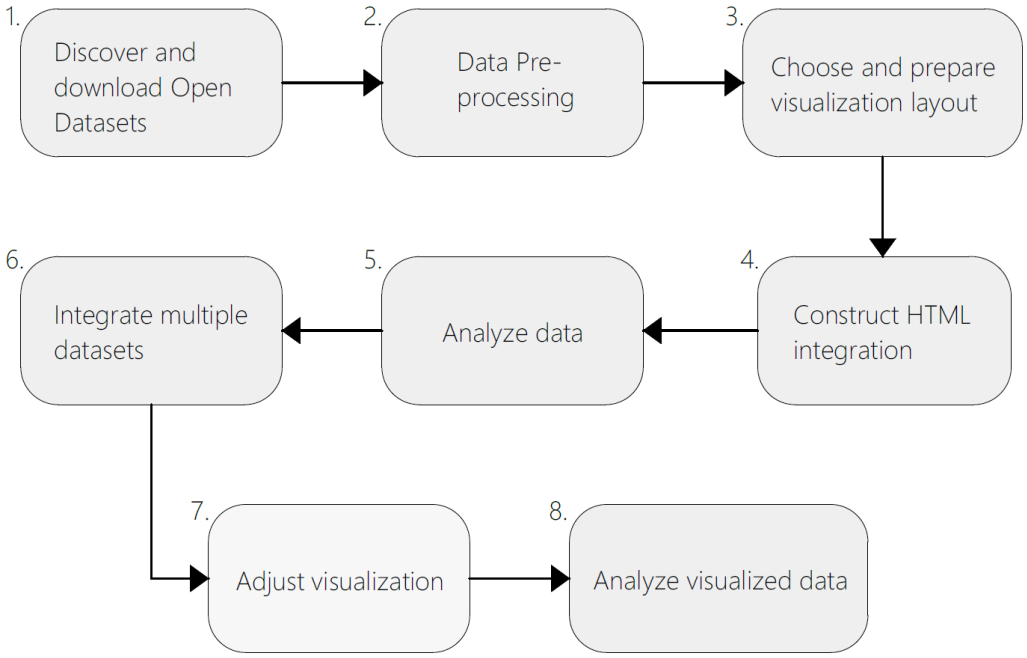


Figure 17: Visualization Workflow

The workflow which was executed in order to develop a dashboard can be seen in Figure 17. It consists of eight tasks which are executed sequentially:

First, one has to browse Austrian Open Data Portals, either data.gv.at or opendataportal.at, in order to discover potentially interesting datasets. It is important that the correct data format is available, i.e. either a GeoJSON or a CSV file since those are the main focus of this thesis. Additionally, a brief inspection of the dataset will determine the actual content type and it will be evaluated whether it is a suitable candidate for the visualization process or not. Some criteria which may disqualify a dataset are: it contains no geospatial information at all, i.e no address, coordinates etc., it does contain invalid coordinates, i.e. the latitude-longitude combination does not return the expected location, and the size of the dataset. If the file size is too large, a quick visualization will not be possible, similarly, if the dataset is too small, there is not enough data in order to produce significant results.

Second, the datasets are inspected more closely which is necessary in order to determine which adaptations must be made to the dataset. Due to the fact that the JavaScript libraries, which are used to visualize the data, require a

specific input format, the data has to be processed accordingly. An example of pre-processing is the modification of a downloaded CSV file. Austria's default syntax for CSV files is distinct from the default, we use a semicolon (';') instead of a comma (',') as a separator, and commas (',') to display decimals rather than dots ('.'), e.g. 15,4 must be 15.4. This leads to issues since the JavaScript libraries only allow true CSV files, thus, the changes must be made in order to conform with the requirements. The adapted datasets are saved and stored locally for development.

Third, a set of suitable visualization methods is created. Due to the fact that different data types, i.e. points, lines and polygons, require different visualization techniques, this list is crucial. A list for points could contain e.g. marker, heatmap, choropleth, whereas polygons would only contain a choropleth map. Consecutively, the maps are loaded into the HTML in order to select their positions and sizes, however, they do not contain any data at this point.

Fourth, the pre-processed data, which is currently stored locally, is loaded into the HTML file in order to access it and work with it directly in the browser through JavaScript. This step is crucial as it constructs the HTML integration, specifically, through the method 'fetch' we are able to load the data and work with it in the Firefox browser. The code snippet below shows an example which allows us to load in a local file and access & modify the entire dataset through regular JavaScript notation. Last, the maps are populated with the predefined visualization styles.

```
// create variable that stores data (coordinates)
var addressPoints;
//define file path
fetch("file:///C:/Users/Hubert/bach_thesis/school_data.json")
.then(res => res.json())
.then (data => addressPoints = data)
// create heatmap with data
.then(() => L.heatLayer(addressPoints,
{radius: 12,blur:20,maxZoom:11}).addTo(map))
```

Fifth, the first visualization of the data can now be explored through Visual Data Analysis, however, it only contains the information of one dataset. The goal is to determine whether the displayed data conveys the intended information, and if it is sufficient or not. If the displayed information is satisfactory the workflow ends here.

Sixth, if the displayed information, which was produced in *5. Analyze data*, is not yet satisfactory, i.e. more insight can potentially be generated if multiple datasets are added and their data combined. Thus, the additionally chosen

datasets, which were pre-processed accordingly, are integrated into the HTML file.

Seventh, the visualization has to be adjusted since the goal was to display information which was generated through the combination of two, or more, datasets. Once the visualization has been adapted accordingly some style elements are incorporated into the layout. Elements such as a legend, which gives detailed descriptions of both the color scheme and the total number of elements in a choropleth map and a function that displays more detailed information such as the elements name, district, address and value were added.

Last, the final state of the website has to be analyzed, i.e. does the site in its current state represent the goals of this thesis. This means, does the visualized data enable Visual Data Analysis for the common citizen and does the layout provide sufficient information for the independent start of the discovery process. If it does not fulfill those requirements, the necessary adaptation have to be made, whether it is in regards to the visualization method or the visualized data. If the dashboard classifies as satisfactory the workflow finishes and the dashboard is in its final state, which will be seen by the user. The workflow can be started once more, starting from step *1. Discover and download Open Datasets*, if the goal is to create an additional dashboard.

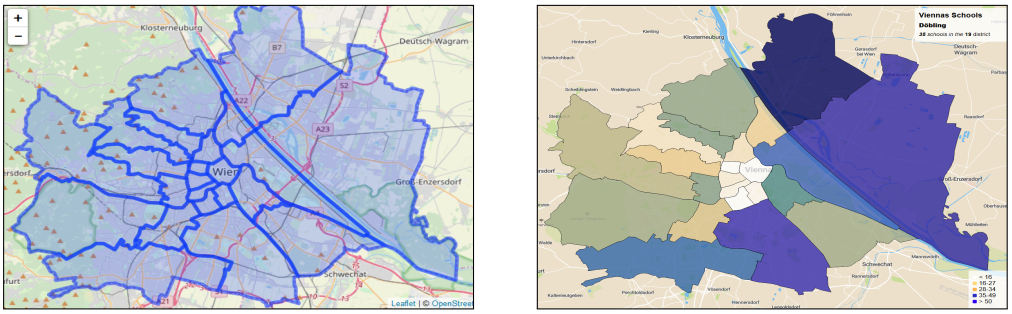


Figure 18: Choropleth maps produced during steps 3 and 7 of workflow

The left image of Figure 18 shows the initial state of step three, *choose and prepare visualization layout*, which depicts Vienna and its borders & districts. The right image of Figure 18, on the other hand, depicts the choropleth map which is weighted by the number of schools per district, and is the result of step seven, *adjust visualization*. Specifically, three style elements were added on top of the choropleth map which provide more clarity and information:

- *infobox*: the box shows a district's name and number of schools when the user hovers over it

- *legend*: a legend was added in order to clarify the color scheme and show the number of schools and how they are classified
- *map layer*: another default map layer, i.e. background layer, was chosen to make the data stand out more clearly

4 Practical Visualizations of Open Data

Table 2 summarizes all of the developed dashboards, provides the sources of the used datasets, and shortly describes them. The sequentially introduced and elaborated dashboards follow the workflow graph as seen in Figure 17.

Dashboard	Data	Data Format	Description
Schools	School Data	GeoJSON [data.gv.at, 2018a]	Each School has a set of attributes: name, address, district and coordinates
WLAN	WLAN Data	GeoJSON [data.gv.at, 2018b]	Each WLAN Access Point has a set of attributes: address, district, coordinates
Weather	Weather Data	CSV [data.gv.at, 2018c]	contains hourly updated weather data from 21 stations, such as: temperature, wind speed and humidity, etc.
Rent	Public Transportation Stations	GeoJSON [data.gv.at, 2018d]	Full list of public transportation Stations in Vienna
	Public Transportation Lines	GeoJSON [data.gv.at, 2018e]	Full list of public transportation connection
	Vienna's m2 Rent Prices per District	PDF [immopreise.at, 2018]	Contains average rent prices per district
	Vienna's Rent Price close to Subway Stations	PNG [immoscout.at, 2018]	Containts average rent price near subway Stations in Vienna
Common	Vienna District Borders	GeoJSON [data.gv.at, 2018f]	Contains the district borders of Vienna
	Austria State Borders	GeoJSON [opendataportal.at, 2018]	Contains state borders of Austria

Table 2: Sources and descriptions of used datasets

The produced code will be uploaded to the following GitHub repository: https://hubertbaginski.github.io/open_data_vis

This thesis project will produce a website through which all of the visualized datasets can be explored by the user. Sequentially, a sidebar will be added which will both display the name of the dataset and redirect to it. Thus, one

dashboard will be added to the sidebar for each dataset and a unique webpage will be created as well, which will allow the user to choose one of the datasets he wants to explore more closely. The structure of the website will be as follows, it will be separated into four equidistant sections, splitting the page in half, both vertically and horizontally. One section will contain exactly one map window with the width=700px and the height=450px, this will guarantee that on a regular monitor with a resolution of 1920x1080px two maps can be seen side by side simultaneously, as well as the sidebar containing the available dashboards. Furthermore, there will be a short description added to each map briefly introducing the visualized information as well as the visualization method.

Furthermore, there will be a dashboard that will display hourly collected streaming data, however, it will be downloaded through a script and handled as a static data example, with the option to explore the daily data for the time period the script was gathering the data. Subsequently, a function will be added that semi-automatically downloads the current weather data, i.e. for the current full hour, and dynamically displays it on the dashboard. However, since the remaining data for the rest of the day has not been stored, it will only be displayed as a single set of datapoints for the hour of the instantiation. This dataset, however, significantly differs form the remaining datasets. First, it contains data collected from Austria‘s 21 weather stations, thus, the priorly introduced visualization methods will not suffice. Consequently, a two dimensional approach will be taken and the according visualization method chosen, i.e. a combination of markers on the map displaying the exact location of the stations as well as the data which will be represented on a two dimensional coordinate system, with the horizontal axis displaying the time in hours for a single day, and the vertical axis containing the specific data e.g. temperature, humidity and wind speed.

4.1 Dashboard - Schools

Setup. The schools dashboard is fundamentally built on one primary dataset, it is the schools dataset [data.gv.at, 2018a] and contains the locations of all schools, both of public and private, in Vienna. Furthermore, it was accessed in the GeoJSON format and the only attribute which is used in this dashboard is the location of the school, since the visualizations we want to implement are a heatmap and a choropleth map. However, the GeoJSON has to be extended by another attribute, the number of schools per district, in order for the choropleth visualization. This was done manually, i.e. the dataset was imported as a CSV file into a database and queried looking for number of schools per district.

```
SELECT * from schools WHERE name LIKE 'Innere Stadt';
```

Querying for the district name removes possible redundancies which might occur if the search by district number is not done properly, and districts such as 1, 10-19 and 21 might be included in the search for the first district. Consequently, the final amount of returned rows was appended to the corresponding district in the GeoJSON file. Lastly, the JavaScript package which renders the heatmap requires an array of arrays containing all of the latitude-longitude pairs of the dataset. However, the contained coordinates in the file were longitude-latitude, which had to be ordered properly. This was done by running the necessary Python code and providing the processed data, with an option to download the file, through running a Flask server. Lastly, it is combined with the GeoJSON file [data.gv.at, 2018f] containing the borders of Vienna and its districts, this is crucial as it is the foundation of the choropleth map, which is sequentially weigthed by the number of schools per district.

The homepage of this project, i.e. the site that is presented to the user when he first visits the website, is the schools dashboard. Both of the top sections have been populated with a map each, the left map displays a choropleth map of Vienna and is weighted by the number of schools for every district. Furthermore, a legend has been added to right-hand side of the map which further elaborates the colour scheme and the total number of schools per district. Lastly, a function has been added that displays detailed information regarding the selected district, such as the district number & name and the total number of schools. The second map displays a heatmap, which means that geographically nearby schools (points) are clustered together and displayed in a more intense color. The map is interactive and recomputes the clusters based on the current zoom, displaying the highest accumulation of nearby schools as orange and the loosest clusters, i.e. single schools without any nearby schools, as blue, the colour grading starts from orange and gradually fades into blue.

Visual Data Analysis. The user can now investigate where individual schools or clusters of schools are through the heatmap. Furthermore, he has the option to find the highest density of schools and confirm or refute a theory such as, the biggest sized district must have the most schools. Additionally, he has the option, through the choropleth map, to inspect the individual districts and check how many schools there are in total. Both visualizations can be seen in Figure 19 and more detailed in Appendix A.1. Finally, one could explore how the schools are distributed and if their distribution is related to a districts size, population or price per square meter in regards to rent in order to determine the 'best' district for school coverage.

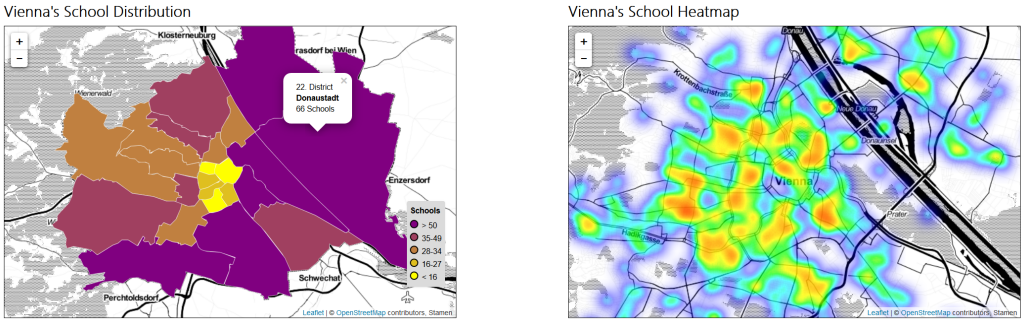


Figure 19: Schools Dashboard

4.2 Dashboard - WLAN

Setup. The WLAN dashboard is, similarly to the schools dashboard, built on one GeoJSON dataset [data.gv.at, 2018b] containing the locations of all public WLAN access points of Vienna. Furthermore, it contains points, which enables the visualization of both the heatmap and the choropleth map. However, to differentiate it from the school dashboard, the locations of the access points have been marked with a circle with a radius of 100 meters. The choropleth map requires the total number of access points per district, which was summed and appended to the file manually. The heatmap requires the array of arrays containing the latitude-longitude pairs, which can be downloaded after running the Python code through the Flask server.

Second, through the sidebar an easy navigation between the different dashboards, which contain different datasets, is possible. The next dashboard, named WLAN, contains all of Vienna‘s public WLAN access points, here three maps have been added to the webpage. Firstly, a heatmap has been added which functions similarly to the priorly introduced school heatmap, with the only difference being that instead of schools, public WLAN access points have been plotted onto the map and the heatmap has been created. Secondly, all of the WLAN access points have been added to the second map, with a circle displaying their coverage, which is estimated for externally installed access points, i.e. not inside a building, of approximately 100 meters. Thus, the user can see if the area he is interested in is covered by free internet. Lastly, a choropleth map has been added and the colours are weighted by the total number of WLAN access points per district. Furthermore, detailed information can be requested by the user, i.e. the district number and name as well as the total number of public access points in the district will be displayed once requested.

Visual Data Analysis. The user has the option to discover where the most public WLAN access points are and check if the area he is interested in is actually covered by free internet. Additionally, he can check which district has the most access points and where the highest accumulation of access points are. Consequently, he has the ability to deduct how the city of Vienna is installing access points and discover a clear pattern, such as that the most frequented areas of the city, whether by tourists or residents, have the best coverage. Figure 20 shows the dashboard and Appendix A.2 shows it in more detail. One can test the theory, that the highest distribution of access points is in areas such as public parks, highly frequented shopping streets and central train & subway stations as it would provide the best value that is achievable.

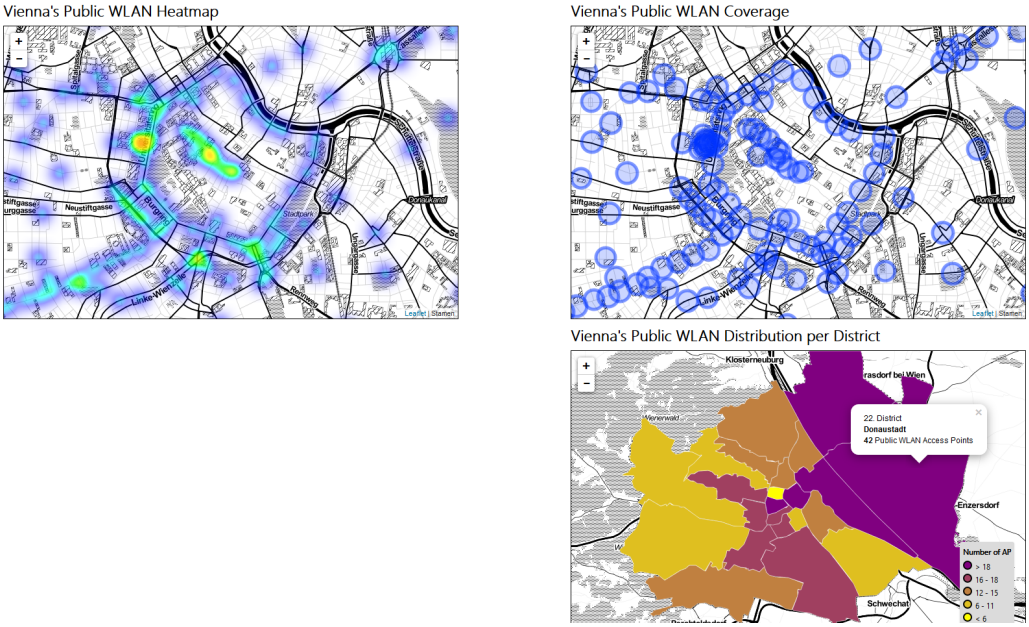


Figure 20: WLAN Dashboard

4.3 Dashboard - Weather

Setup. The third dashboard was initially planned to display a streaming data example which was added under the 'Weather' dashboard. For the duration of five months a script has been downloading the hourly updated weather dataset [data.gv.at, 2018c] provided by the Central Institute for Meteorology and Geodynamics (ZAMG). Subsequently, the individual CSV files have been merged and processed to conform with the necessary format for its visualization. Step 2, *Data Pre-processing*, was especially important for this dataset, as it did not

contain the default CSV syntax, but it was a commonly encountered 'Semicolon-separated-value' file, which had to be changed into a regular CSV in order for the visualizations to compile correctly. Furthermore, it was the first dataset that contained a time component, thus, a different visualization style had to be theorized and developed in step 3, *choose and prepare visualization layout*, of the workflow. Moreover, the dataset contained a big variety of different attributes, however, we chose to solely work with the temperature, dew point, average wind speed, peak wind speed, relative humidity and relative sunshine.

Thus, we decided to add a map with Austria's states and the 21 weather stations which are displayed through one regular marker each, with the option to display the station name and height upon user request. Next, two two-dimensional graphs have been added, the horizontal axis displays the time of one day, in hours, and the vertical axis displays the values, such as: temperature, dew point, wind speed and humidity. Furthermore, the data is queryable, i.e. the user can pick a date which is then queried by the browser and finally returned to the user without refreshing the page. However, it will only return the daily data for one stations in order to minimize clutter and keep the visualized data as simple and understandable as possible. Thus, the user has the option to choose a station, and once selected the corresponding station data for the day is queried and displayed. Lastly, a function was added to display the current live weather data. During development, a server must be running which executes the necessary data processing code. Subsequently, the current weather data can be downloaded, and processed into the necessary format, through the download button. Next, the user has to manually move the file from the download folder to the predefined folder in order for the visualization to work. Finally, the 'Get Live Data' button will compute the live data and display it on the webpage.

Visual Data Analysis. The user has an overview of the country and all of its weather stations, he can also see the stations name and height. Additionally, he has the option to inspect the development of the attributes over the course of a day and identify extreme values easily. He can also select a specific station he is especially interested in, as well as looking up the daily data of a specific date. The data can be seen in Figure 21 and more detailed in Appendix A.3. One can test the theory that the higher the station height is, the more extreme temperatures and wind speeds are experienced by the station.

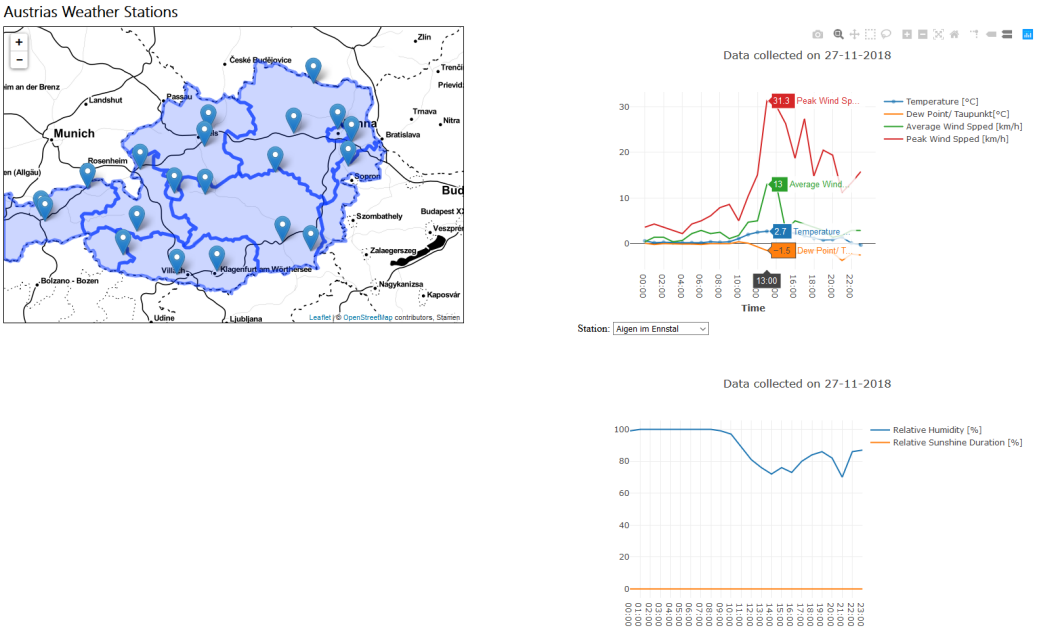


Figure 21: Weather Dashboard

Summarizing, the weather data can be queried by day and station. Furthermore, the user has the possibility to download the current live data and visualize it on the site, without leaving the current session, i.e. the site does not reload or redirect, which provides increased user friendliness.

4.4 Dashboard - Rent

Setup. This dashboard combines multiple datasets in order to generate new insights. It combines two open datasets, one that contains all of Vienna's public transportation lines and stations [data.gv.at, 2018e] in a GeoJSON format, and Vienna's borders and district borders [data.gv.at, 2018f] which is also accessed in a GeoJSON format. Furthermore, it combines two public datasets which contain the average rent prices per district [immopreise.at, 2018] in a PDF format, and the average rent price for a 73 m2 apartment in close proximity to subway stations [immoscout.at, 2018] in a PNG format. Since the public data sets are not open, additional data processing has to be done, concretely, the average rent prices per district have to be appended to the cities GeoJSON file as a new attribute named 'price', and the prices close to subway station have to be appended to the corresponding stations, again under the attribute 'price'. However, we are only interested in the subway lines, thus, the file is adapted and only the subway stations remain. The remaining attributes are the subway

line, the coordinates, station name and the price per station.

The final dashboard has been added, it displays Vienna's public subway lines and the rent prices in close proximity to the stations. This visualization is an example of data combination, since it combines the data of three datasets, i) the subway lines and points, ii) the average total rent price for a 73 square meter apartment in close proximity to the subway stations, and iii) the average rent price per square meter for each district. Thus, two maps have been added, the first one displays all of the subway lines, i.e U1, U2, U3, U4 and U6, in their respective native colours, and a circle marker that is coloured according to the rent price, ranging from 'cheap' to 'average' to 'expensive'. Furthermore, a function had been added to toggle the colours of the subway lines since it causes a lot of clutter and diminishes the readability of the map. Hence, the user can choose if he is interested in the specific subway lines or the 'best option' for renting an apartment that meets his criteria the best. The second map displays a choropleth map of the average square meter rent prices for each district. It is coloured according to the average rent price, and detailed information can be requested, it will display information such as district number & name and the average rent price per square meter of the district. Furthermore, a legend has been added to further elaborate the colour scheme of the map and provide more clarity to the visualization.

Visual Data Analysis Through the dashboard the theory can be tested that the closer to the city center one lives, the more expensive the rent of the apartment is. We can confirm this by checking the visualization in Figure 22, the most expensive district is the city's centre, and the further apart the apartment is from the center the cheaper it becomes. However, there is one major exception, it is the 22. district which has a significantly higher average price per square meter than it should have, based on the theory which was introduced earlier. Additionally, if one wants to find a subway station which is close to the center but in comparison rather 'cheap', there are four candidates which are only a few stations away, they are marked as green circles. Finally, there is a function which shows the average monthly rent of an apartment near subway stations, it depicts detailed information such as station name and price.

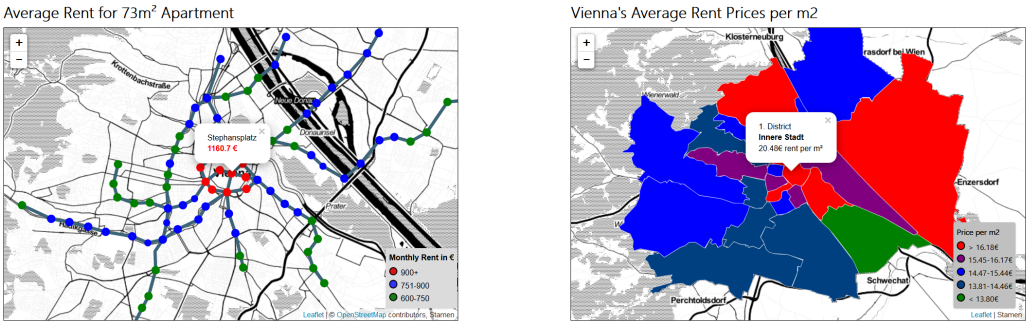


Figure 22: Rent Dashboard

Figure 22 shows the rent dashboard and depicts the most expensive areas as red, and the cheapest as green. Appendix A.4 shows this in more detail.

5 Conclusions and Future Work

Due to the steady growth and development of Open Data the amounts of publicly available datasets is increasing continuously. However, many people can not efficiently use or work with the provided raw data, thus, visual analysis becomes a crucial tool for a better understanding of that data. Once the data is represented in a easily comprehensible visualization, data discovery and exploration for the common citizen are enabled. In this thesis we developed multiple dashboards that visualize open static datasets and provide a tool for a wide audience to discover this data. Furthermore, one dashboard was added where multiple datasets were combined to generate new information, which was not available in this format before. First, we proposed a workflow, which if followed enables the visualization of Open Data, however, there are still many manual tasks which have to be executed in order for the dashboards to be properly visualized. There are some papers that investigated both recommendations for personalized visualizations such as VizRec [Mutlu et al., 2016] and automated visualization of Open Data such as reboting.com [Heil and Neumaier, 2018], however, this exceeds the scope of this thesis.

This project was developed in a development environment, however, the ultimate goal is to launch it on a server for production. Moreover, for this project the main limitations are that some data files have to be moved manually in order negate Cross-Origin Request Errors, those file transfers ought to be automated in the future. The projects launch on a server might resolve the Cross-Origin request issues, which are currently forcing the manual transfer of files.

The agenda to further improve this project is to implement a function that allows the user to pick an open dataset from an Open Data Portal, such as data.gv.at, and upload it to the website. Next, it is analyzed and dynamically visualized on the website. The analysis would determine the content type and choose a set of appropriate visualization methods, The user could then pick some layout he finds especially appealing and submits the visualizations methods he wants to see. Consequently, the website processes, computes and visualized his Open Dataset, and enables the Visual Data Analysis for his dataset. The dataset and visualizations are stored on a server under a new dashboard, and are accessible by everyone visiting the site. Finally, the extension to all browsers and its compatibility with the code should be a main priority, since the current focus was on the compatibility with Firefox.

References

[Beno, 2016] Beno, M. (2016). Open data hopes and fears, determining the barriers of open data. https://aic.ai.wu.ac.at/~polleres/supervised_theses/Martin_Beno_BSc_2016.pdf.

[Charaniya et al., 2010] Charaniya, A., Rohlf, J., and Jones, M. T. (2010). Streaming and interactive visualization of filled polygon data in a geographic information system. US Patent 7,746,343.

[Cook and Thomas, 2005] Cook, K. A. and Thomas, J. J. (2005). Illuminating the path: The research and development agenda for visual analytics. Technical report, Pacific Northwest National Lab.(PNNL), Richland, WA (United States).

[data.gv.at, 2018a] data.gv.at (2018 (accessed December 20, 2018)a). https://www.data.gv.at/katalog/dataset/c1ba372b-dba2-4bce-b72e-b5c832eaaf44.

[data.gv.at, 2018b] data.gv.at (2018 (accessed December 20, 2018)b). https://www.data.gv.at/katalog/dataset/354f9aa0-2f0a-4cdf-9b60-d023da93cc76.

[data.gv.at, 2018c] data.gv.at (2018 (accessed December 20, 2018)c). https://www.data.gv.at/katalog/dataset/9b40a0af-a6fe-47ff-9624-2ea8f40c746f.

[data.gv.at, 2018d] data.gv.at (2018 (accessed December 20, 2018)d). https://www.data.gv.at/katalog/dataset/f1f6f15d-2faa-4b62-b78b-80599dd1c66e.

[data.gv.at, 2018e] data.gv.at (2018 (accessed December 20, 2018)e). https://www.data.gv.at/katalog/dataset/36a8b9e9-909e-4605-a7ba-686ee3e1b8bf.

[data.gv.at, 2018f] data.gv.at (2018 (accessed December 20, 2018)f). https://www.data.gv.at/katalog/dataset/stadt-wien_bezirksgrenzenwien.

[Dent et al., 1999] Dent, B. D., Torguson, J. S., and Hodler, T. W. (1999). Cartography: Thematic map design. *Boston: WCB/MCGraw-Hill*, 5.

[europeandataportal.eu, 2019a] europeandataportal.eu (2019 (accessed January 1, 2019)a). *Open Data in Europe*. https://www.europeandataportal.eu/en/dashboard#tab-overview.

[europeandataportal.eu, 2019b] europeandataportal.eu (2019 (accessed January 1, 2019)b). *State-of-Play on Open Data - 2018*. https://www.europeandataportal.eu/sites/default/files/country-factsheet_austria_2018.pdf.

[Health and Bizer, 2011] Health, T. and Bizer, C. (2011). Linked data: Evolving the web into a global data space. *Synthesis Lectures on the Semantic Web: Theory and Technology*, 1(1):1–136.

[Heil and Neumaier, 2018] Heil, E. and Neumaier, S. (2018). reboting.comm towards geo-search and visualization of austrian open data.

[immopreise.at, 2018] immopreise.at (2018 (accessed December 20, 2018)). https://www.immopreise.at/Wien/Wohnung/Miete.

[immoscout.at, 2018] immoscout.at (2018 (accessed December 20, 2018)). https://www.immobilienscout24.at/ratgeber/wohnung-suchen/mietpreise-wien/_jcr_content/par/textimage_858270447/image.adapt.700.medium.png/1481729570326.png.

[Kalashnikov and Mehrotra, 2006] Kalashnikov, D. V. and Mehrotra, S. (2006). Domain-independent data cleaning via analysis of entity-relationship graph. *ACM Transactions on Database Systems (TODS)*, 31(2):716–767.

[Karam, 1994] Karam, G. M. (1994). Visualization using timelines. In *Proceedings of the 1994 ACM SIGSOFT international symposium on Software testing and analysis*, pages 125–137. ACM.

[Keim et al., 2002] Keim, D. A., Hao, M. C., Dayal, U., and Hsu, M. (2002). Pixel bar charts: a visualization technique for very large multi-attribute data sets. *Information Visualization*, 1(1):20–34.

[Keim et al., 2008] Keim, D. A., Mansmann, F., Schneidewind, J., Thomas, J., and Ziegler, H. (2008). Visual analytics: Scope and challenges. In *Visual data mining*, pages 76–90. Springer.

[Keim et al., 2006] Keim, D. A., Mansmann, F., Schneidewind, J., and Ziegler, H. (2006). Challenges in visual data analysis. In *Information Visualization, 2006. IV 2006. Tenth International Conference on*, pages 9–16. IEEE.

[Kitchin, 2014] Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.

[Lerner and Tirole, 2001] Lerner, J. and Tirole, J. (2001). The open source movement: Key research questions. *European economic review*, 45(4-6):819–826.

[MacEachren and Kraak, 2001] MacEachren, A. M. and Kraak, M.-J. (2001). Research challenges in geovisualization. *Cartography and geographic information science*, 28(1):3–12.

[Marras et al., 2018] Marras, M., Matteo, M., Boratto, L., Fenu, G., and Laniado, D. (April 23–27, 2018). Barcelonanow: Empowering citizens with interactive dashboards for urban data exploration. *In WWW '18 Companion: The 2018 Web Conference Companion, Lyon, France. ACM, New York, NY, USA 5 Pages. https://doi.org/10.1145/3184558.3186983.*

[Monmonier, 1990] Monmonier, M. (1990). Strategies for the visualization of geographic time-series data. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 27(1):30–45.

[Murray, 2017] Murray, S. (2017). Interactive data visualization for the web: An introduction to designing with. *O'Reilly Media, Inc.*

[Mutlu et al., 2016] Mutlu, B., Veas, E., and Trattner, C. (2016). Vizrec: Recommending personalized visualizations. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 6(4):31.

[opendatahandbook.org, 2018] opendatahandbook.org (2018 (accessed December 28, 2018)). *What is Open Data?* `http://opendatahandbook.org/guide/en/what-is-open-data/`.

[parisinnovationreview.com, 2019] parisinnovationreview.com (2018 (accessed January 19, 2019)). *A brief history of Open Data.* `http://parisinnovationreview.com/articles-en/a-brief-history-of-open-data`.

[Peña et al., 2014] Peña, O., Aguilera, U., and López-de Ipiña, D. (2014). Linked open data visualization revisited: a survey. *Semantic Web Journal.*

[Plaisant et al., 2003] Plaisant, C., Mushlin, R., Snyder, A., Li, J., Heller, D., and Shneiderman, B. (2003). Lifelines: using visualization to enhance navigation and analysis of patient records. In *The Craft of Information Visualization*, pages 308–312. Elsevier.

[Swain and Hauska, 1977] Swain, P. H. and Hauska, H. (1977). The decision tree classifier: Design and potential. *IEEE Transactions on Geoscience Electronics*, 15(3):142–147.

[Tollis, 1996] Tollis, I. G. (1996). Graph drawing and information visualization. *ACM Computing Surveys (CSUR)*, 28(4es):19.

[Weske, 2012] Weske, M. (2012). Business process management architectures. In *Business Process Management*, pages 333–371. Springer.

[Wu et al., 2014] Wu, X., Zhu, X., Wu, G.-Q., and Ding, W. (2014). Data mining with big data. *IEEE transactions on knowledge and data engineering*, 26(1):97–107.

[Yagi et al., 2012] Yagi, S., Uchida, Y., and Itoh, T. (2012). A polyline-based visualization technique for tagged time-varying data. In *Information Visualisation (IV), 2012 16th International Conference on*, pages 106–111. IEEE.

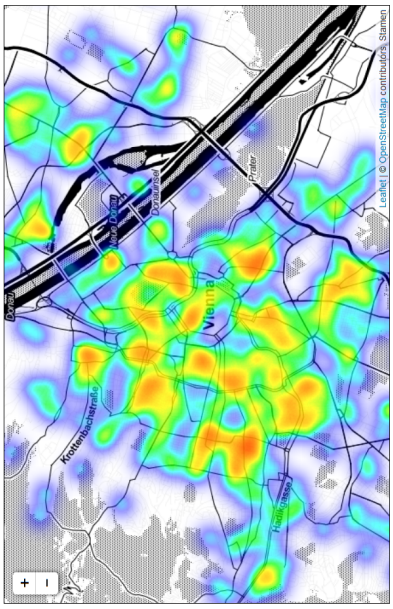
[Zhou and Weiskopf, 2018] Zhou, L. and Weiskopf, D. (2018). Indexed-points parallel coordinates visualization of multivariate correlations. *IEEE transactions on visualization and computer graphics*, 24(6):1997–2010.

Appendices

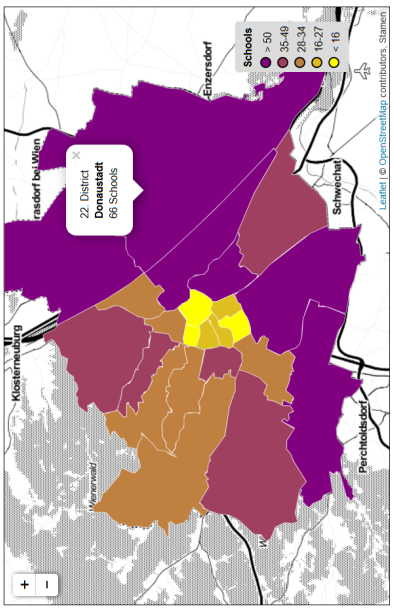
A Dashboards

A.1 Dashboard - Schools

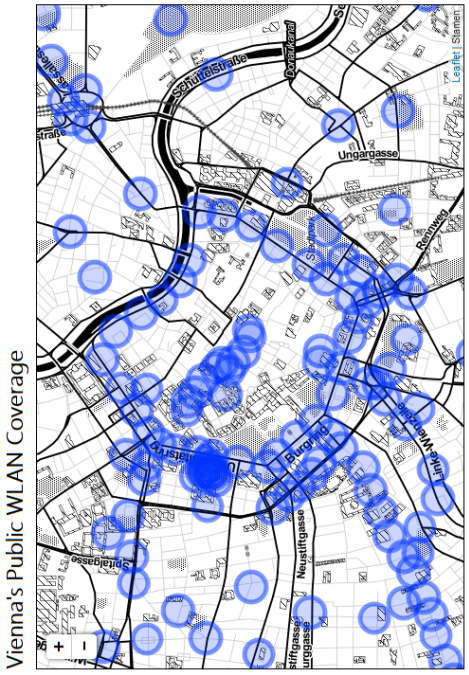
Vienna's School Heatmap

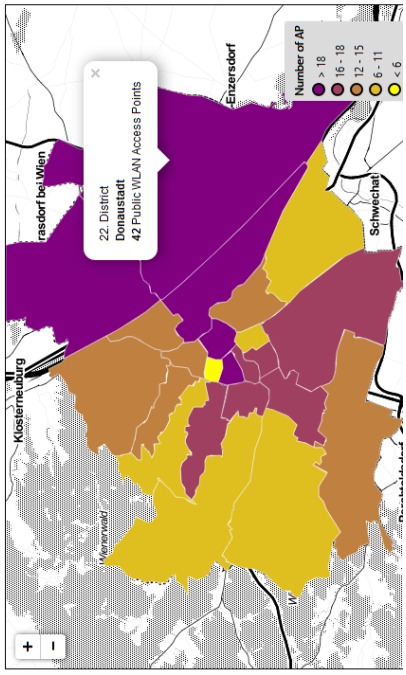


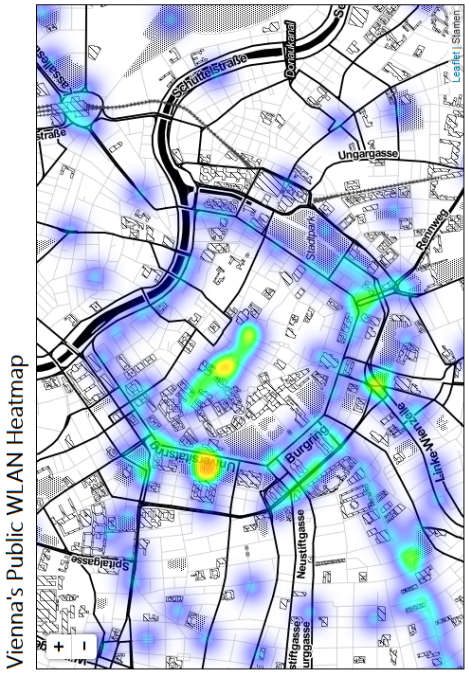
Vienna's School Distribution



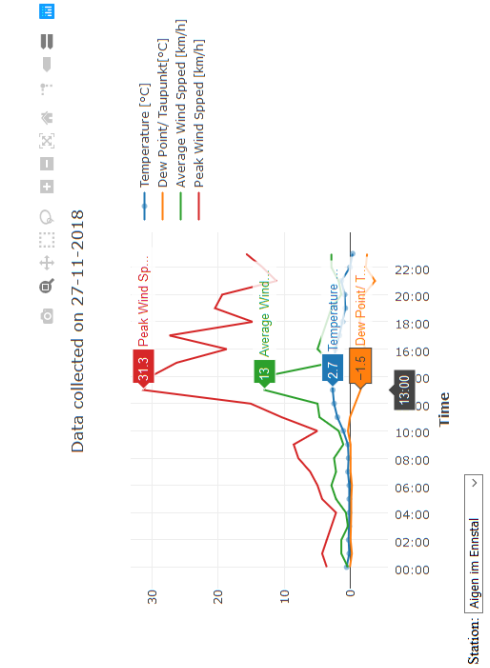
A.2 Dashboard - WLAN

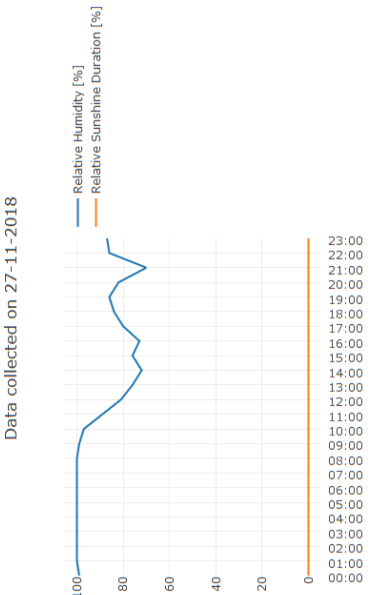


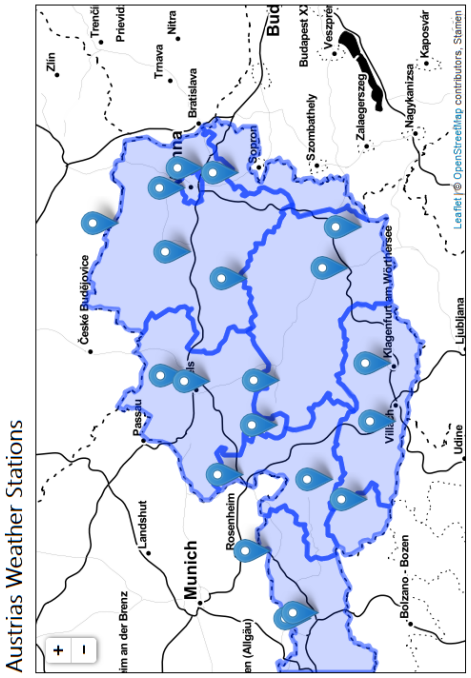




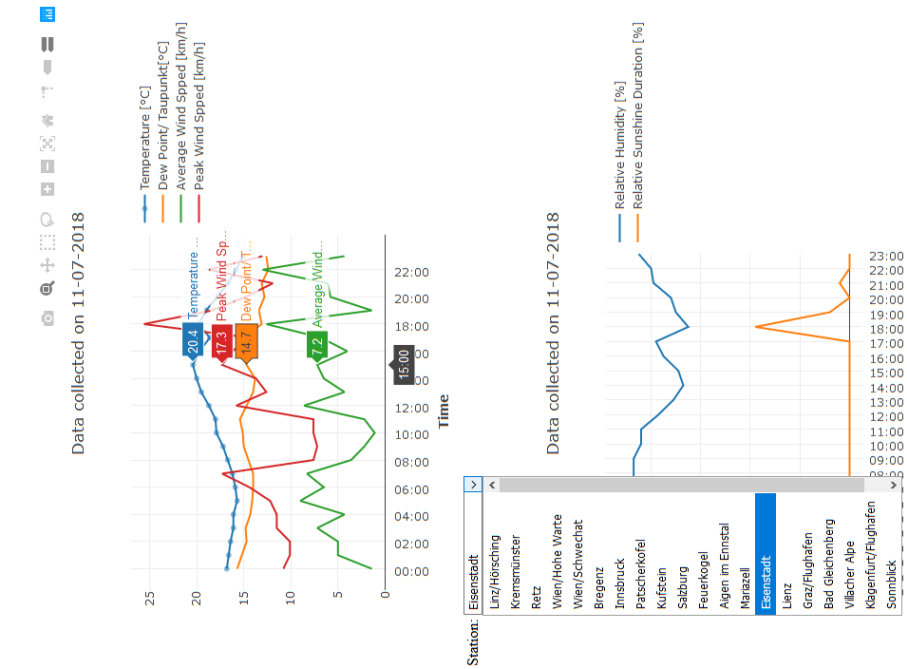
A.3 Dashboard - Weather: Overview

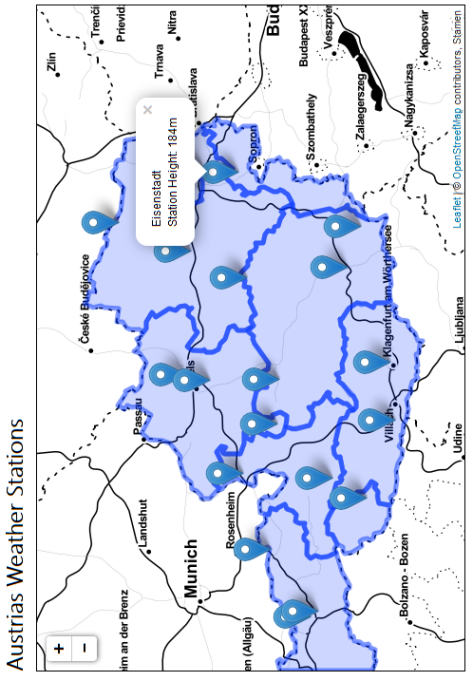




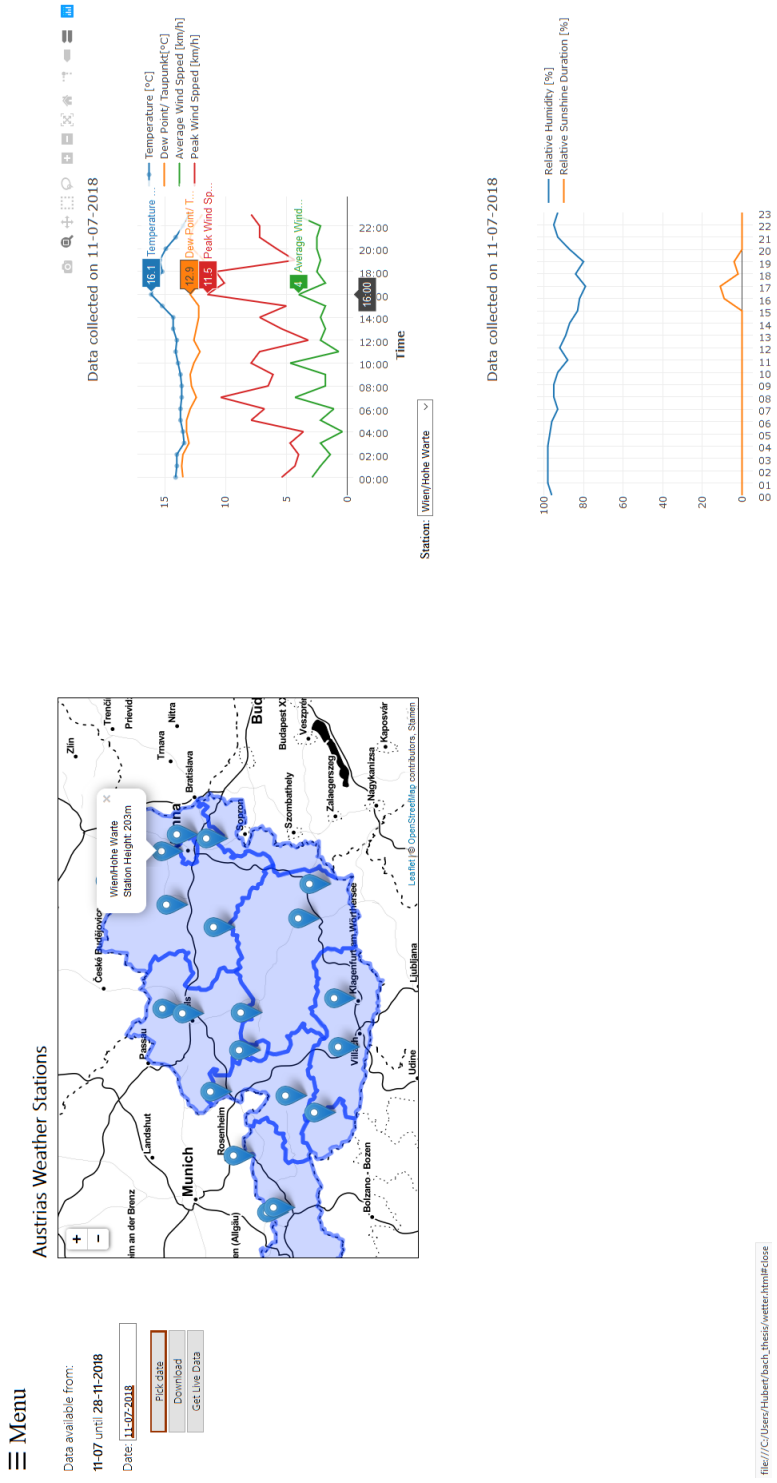


A.3.1 Dashboard - Weather: Pick Station





A.3.2 Dashboard - Weather: Pick Date



A.4 Dashboard - Rent