

INFORMATION VISUALIZATION

ENERGY GENERATION IN EUROPE

FINAL REPORT

JEROEN GREGOIRE 0092057 QUENTIN JULEMONT 0554389 MENGYAO SONG 0563486

Department of Applied Science and Engineering: Computer Science

Contents

1	Introduction	3
2	Target Users	3
3	Datasets 3.1 BP Statistical Review of World Energy June 2019 [1] 3.2 30 Years of European Solar Generation [2] 3.3 30 Years of European Wind Generation [3]	4 4 4 5
4	Pre-processing 4.1 Matrices to lists 4.2 Uniform structure 4.3 Overview of common data 4.4 Result in TSV format 4.5 Precalculated attributes	5 6 7 7 8 8
5	Visualization 5.1 Chorolepth Map 5.2 Stacked Area Graph 5.3 Aster Plot	9 9 11 12
6	Evaluation 6.1 The result of questionnaire 6.2 Improvement	13 13 14
7	Changes along the way	14
8	Conclusion & Potential Extension	14
9	Workload Division and Gantt Chart 9.1 Workload Division 9.2 Gantt Chart	15 15 17
10	Bibliography	18

1 Introduction

Energy has a great contribution to world's development, but it also brings pollution. Now people in growing numbers are beginning to be concerned about changing energy production. However, there is plenty of energy to be counted, like coal, solar, wind, geo biomass, etc. Faced with such numerous data, people have a hard time to read and understand them. Therefore, our project is focused on the visualization of Europe's production of energy from 1986 to 2015. By making different kinds of visualizations, we could do horizontal comparisons of the energy situation between European countries, as well as vertical comparisons of a country's energy composition in these years.

As we focus on renewable energy, we start by analyzing the renewable energies consumed in electricity generation. Next, we show the amount of each resource consumed each year in each country, which shows the changing structure of renewable energy sources. Also, we visualize a comparison between countries for the percentage of electricity generated by wind/solar and the total energy generation.

A working version of the visualization is published on the URL https://infovis.offgrid.expert/ The published version is optimized for a chrome browser on a windows desktop computer.

2 Target Users

Our target users are energy minded people with a desire to see renewable efforts and potential on a country level. The target users are interested in following the progress in changing energy generation. For each country in Europe, we want to show the change in generation over the last 20 years in relation with how the required energy is generated. More specifically, we want to visualize the change per country to renewable sources.

We created a single page dashboard, where the reactiveness and responsiveness have been the focus. We wanted to give the users a fast responding interface. The programming choices and preprocessed data are the key for this. We avoided calls to API's, or heavy calculations to ensure the user gets immediate feedback. The visualization is designed to be self-explaining. There are tooltips added to provide textual clarification as a backup to the intuitiveness.

3 Datasets

Based on the feedback, we combined the following three datasets.

3.1 BP Statistical Review of World Energy June 2019 [1]

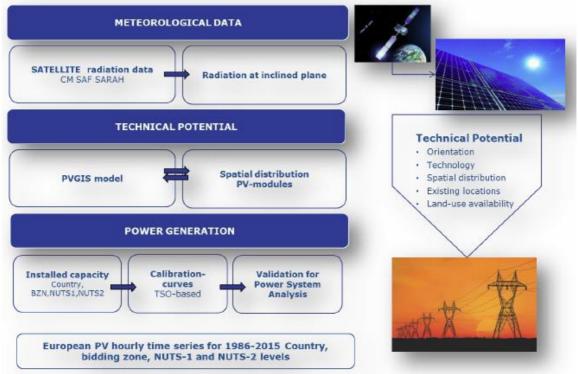
 $\underline{https://www.bp.com/content/dam/bp/businesssites/en/global/corporate/xlsx/energy-economics/statistical-review/bpstats-review-2019-all-data.xlsx}$

This dataset contains information about the consumption per country worldwide and how the energy is generated.

3.2 30 Years of European Solar Generation [2]

https://www.kaggle.com/sohier/30-years-of-european-solar-generation/data

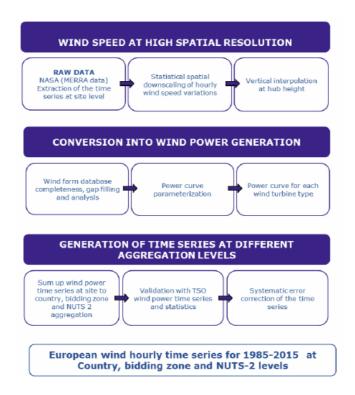
This dataset contains information about the weather conditions per hour since 1986 for European countries, expressed as a percentage of the installed capacity in 2015. The data is calculated based on information about installed capacity in the various countries, and the meteorological information of regions. Statistical techniques are used to predict values for gaps or missing data.



3.3 30 Years of European Wind Generation [3]

https://www.kaggle.com/sohier/30-years-of-european-wind-generation/data

The "30 Years of European Wind Generation"-dataset holds similar information than its solar counterpart. This dataset uses information from NASA. It uses wind in certain regions to calculate potential to wind generators installed in 2015 in the countries. It also uses statistical techniques to predict values for gaps or missing data.



4 Pre-processing

The datasets each hold information about countries in a certain year. In order to combine the sets, information needs to have the same identifiers. Also, the values to populate the visualizations require units that make sense, or are comparable. Each dataset was pre-processed to a generic format over the datasets.

4.1 Matrices to lists

The datasets were provided in csv and Microsoft Excel. Dataset "BP Statistical Review of World Energy June 2019" was a large file with matrices spread over several tabs. Each tab was holding information about a type of energy used, throughout a period of time in specific countries.

Our focus of the visualization is focused around renewable energy. The tabs holding other information were removed. Then the page layout was stripped away, and the non-European countries removed, leaving a matrix with country data per row, years as columns.

The 2 other datasets came in TSV format, with the countries of Europe in the header and a percentage per row. Each row represented an hour since 1986 (over 292000 rows), holding the % potential of the capacity installed in solar and wind generation. These 2 sets first required to be quantified with the actual installed capacity per country. This allowed calculating the potential energy consumption in Megawatts per hour per country. Then the information was aggregated to yearly per country, and changed to Terawatts per hour. Just like the first dataset, the result was a matrix with year per row and country as columns.

Sum of Value Colu	mn Labels 🔻					
Row Labels ▼ AT		BE	BG	СН	CZ	DE E
1986	4234468,636	36094398,79	8382876,409	6086002,835	18386974,69	1567043054
1987	3870411,91	32889962,28	7934168,349	5601991,684	16497703,88	1444464124
1988	4083087,505	32756613,41	8228403,137	5847535,978	17273105,45	1475744291
1989	4218851,689	36996870,07	8643978,821	6374657,507	17689012,99	1602979090
1990	4265193,361	36382866,69	8689994,311	6271753,569	18012946,2	1595907029
1991	4127512,642	36650636,21	7760488,544	6172848,554	17597906,53	1637725660
1992	4277536,143	35098278,62	8723739,95	5926375,688	18910979,92	1605774300
1993	4000690,719	33859355,85	8548274,913	5805183,429	17669633,67	1525752008
1994	4067853,833	33701658,04	8398250,063	5675239,496	17754691,17	1549313605
1995	3842877,04	35153007,2	8053457,32	5802513,487	16559282,19	1510946805
1996	3828076,71	34647268,87	7673649,865	5752262,152	16122197,33	1474015482
1997	4234618,795	36649429,44	8127931,542	6353811,768	18150310,94	1617644425
1998	3993038,777	31444228,92	8113218,965	6066471,408	16667017,79	1424659693
1999	3958069,778	35490315,45	8334015,637	5715305,054	17270449,42	1560927397
2000	4173484,626	32940323,69	8977097,437	5910858,115	17641564,22	1522168508
2001	4041371,406	33738421,64	8263815,646	5909833,576	16142674,64	1491745184
2002	4051832,796	34148659,6	7917200,361	5869896,659	17203798,88	1510055331

The first dataset has transposed values compared to the two other datasets. Transforming the data from a matrix to a list, solves this problem. The outcomes are three columns: country, year and value.

Because of the powerful pivot functions of Microsoft Excel, we used Microsoft Excel to generate the list from the matrix.

Row	▼ Column ▼	Value 💌	Columr ▼	Row	Column T	Value 💌	Columr ▼
1986	AT	4234469	4,234469	1986	AT	4234469	4,234469
1986	BE	36094399	36,0944	1986	BE	36094399	36,0944
1986	BG	8382876	8,382876	1986	BG	8382876	8,382876
1986	СН	6086003	6,086003	1986	СН	6086003	6,086003
1986	CZ	18386975	18,38697	1986	CZ	18386975	18,38697
1986	DE	1,57E+09	1567,043	1986	DE	1,57E+09	1567,043
1986	EL	49539258	49,53926	1986	EL	49539258	49,53926
1986	ES	1,68E+08	168,1526	1986	ES	1,68E+08	168,1526
1986	FI	49494,86	0,049495	1986	FI	49494,86	0,049495
1986	FR	1,65E+08	164,6636	1986	FR	1,65E+08	164,6636
1986	HU	260767,5	0,260768	1986	HU	260767,5	0,260768
1986	IE	1852,395	0,001852	1986	IE	1852,395	0,001852
1986	IT	5,36E+08	535,8119	1986	IT	5,36E+08	535,8119
1986	NL	18172504	18,1725	1986	NL	18172504	18,1725
1986	NO	58700,48	0,0587	1986	NO	58700,48	0,0587
1986	PL	1516178	1,516178	1986	PL	1516178	1,516178
1986	PT	3462893	3,462893	1986	PT	3462893	3,462893
1986	RO	12763861	12,76386	1986	RO	12763861	12,76386

4.2 Uniform structure

All the different fractions of datasets in the form of "country, year, value" were uploaded in a Google sheets file, with each dataset in another tab. The keys of the columns had ISO country codes for some, country names for others. Also, the units of the value varied. Then generate a page with all the data combined, each tab had a county column added. If a tab contained the ISO country code, the name was added and vise versa. The values were changed to the same unit (Terrawatts hour). This allowed the combination of all tabs, based on country and year.

4.3 Overview of common data

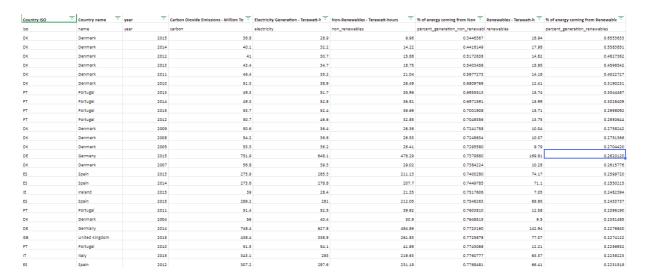
Not all the tabs contained information for the same countries nor for the same years. To make a decision on the time range to use, and which countries to withhold, an overview tab was created. This tab quickly showed which data per tab was available for the countries. A set of countries was taken based on sufficient information over the tabs.

The time range we set from 1986 to 2015. Mainly because the potential dataset contains information about this period, but also because it is the biggest common divisor between all the datasets. We selected the following countries: Austria, Belgium, Bulgaria, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Netherlands, Norway, Poland, Portugal, Romania, Spain, Sweden and the United Kingdom.



4.4 Result in TSV format

The final step in the pre-processing is to combine all the different tabs in one big set. This is done by index and match functions in google sheets, taking values only for the countries and years we specified. The complete set was downloaded in the TSV format. A tab character is better for separating the different values, especially when dealing with numbers. Future visualizations can be quickly added to this dataset, by adding a column with the corresponding formulas.



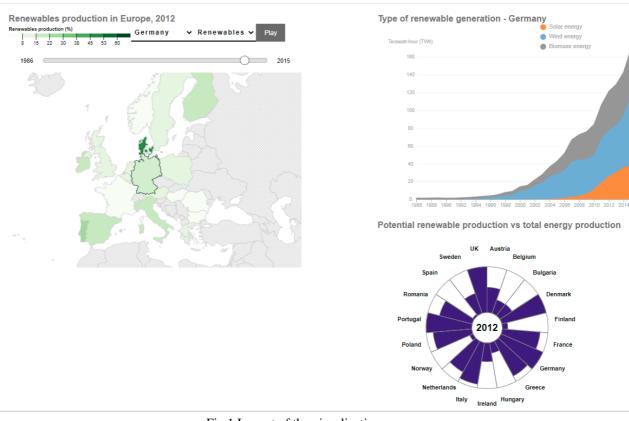
4.5 Precalculated attributes

The data from the pre-processed file is created to allow adding attributes easily. During the course of the project, we added multiple new attributes. These attributes are still available in the provided pre-processed data file. When we considered a new attribute, we always checked whether the attribute added meaningful new information to the visualization as a whole and whether the calculated data was usable for that purpose. In most cases the values where too low, or did not change that much over time.

By pre-calculating the attributes, we avoided doing calculation in real time.

5 Visualization

In the visualization, we use D3.js to make the maps and use node.js to display JavaScript. The layout of our visualizations is as follow:



30 years of Renewable Energy Production in Europe

Fig.1 Layout of the visualizations

5.1 Chorolepth Map

Up to now, we make four choropleth maps^{[4][5]} in one, each one having a different color scheme and being associated with one of the four energy production type that we have (renewables, solar, wind and biomass). The mouse changes into a pointer when it hovers a selectable country (only the European countries with data to show are clickable) and the borders of the hoverer country are highlighted. This clearly indicates to the user.

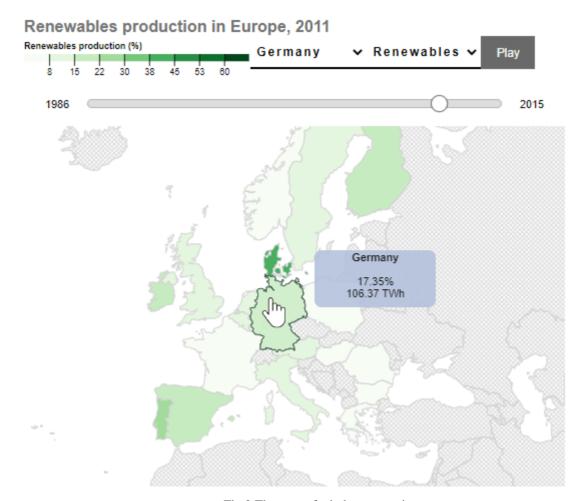


Fig.2 The map of wind consumption

1) Tooltips^[6]

On the map, we can move the mouse over a country and information about that country will appear in the tooltip. More specifically, the tooltip will show the country's name, the amount of energy and the percentage of electricity produced by the corresponding renewable resources and total electricity.

2) Date slider^{[7][8]}

A date slider is available above the map, which can adjust the year displayed. We can either click on the slider or use the 'Play' button, which will display a smooth transition from 1986 to 2015.

3) Threshold Scales^{[9][10]}

In order to make the maps clearer, we have defined a dynamic color scale for each map. Scales are set according to the maximum value of each diagram.

To identify the different visualizations, we have assigned different colors scales to the maps. Renewable energy is assigned to green; wind energy is assigned to blue; solar energy is assigned to orange, and geo-biomass is assigned to grey.

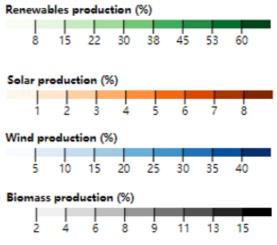


Fig.3 Threshold Scales of four energies

5) Selecting Country

A country can be selected in 2 ways:

- Clicking on a specific country on the map
- Selecting a country from the country dropdown menu

When a country is selected, the borders on the map will be stroked with a dark color. The same effect is added as hovering over a country with the mouse. The highlighting of the borders is an extra visual feedback of selecting a country. The selected country name is also shown in the titles of the graphs using the country as a selector, and on the country dropdown as a title.

6) Brushing with Stacked Area Graph and Aster plots

Clicking on a country in Choropleth map can brush the Stacked Area Graph accordingly. This part will be detailed in 5.2. Similarly, operations on the date slider will brush the Aster plot.

5.2 Stacked Area Graph

To analyze the renewable energy structure changes of each country, we can visualize the percentage of energy for each country in a Stacked Area Graph^[11-14]. In addition to the map clicking function, another select box/dropdown is available and allow the user to switch between countries. These two features will have the effect of changing the visualized country through a smooth transition.^[16] The tooltips show the year and the exact amount of the energy generation. Apart from tooltips, to make it more readable, we add horizontal lines in the middle, axes and legend in the graphs.^[15] With the distinct color and tooltips, it is intuitive for users to observe the generation.

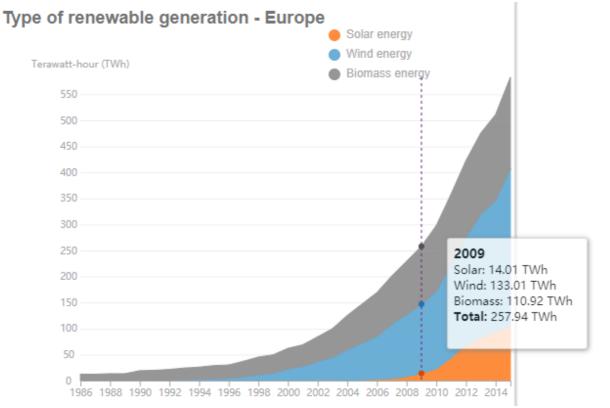
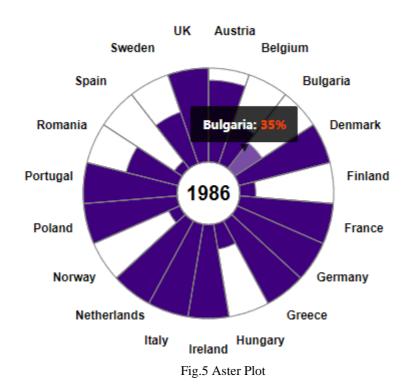


Fig.4 Stacked Area Graph

5.3 Aster Plot

Potential renewable production vs total energy production



With the third visualization, an Aster Plot^[17], we depict the percentage of the potential energy generated by wind/solar compared with the total energy generation for each country. The date slider used with the choropleth map can also be controlled to change the date of the aster plot. With 19 countries to be shown in one plot, legends may be confusing, so instead of using legends, we choose to put the name of the country next to each sector^[18]. Note that the chosen color used for this plot is meaningless. The color was chosen to not be mistaken for a color in one of the other graphs. The Aster Plot shows the potential of the total energy production in a country in a specific year, that could have been produced by renewable sources, if the same infrastructure as in 2015 was available in the specific year.

6 Evaluation

We conducted a questionnaire, which contains 11 rating questions and one optional comment, the full score is five for each question respectively, recruiting 30 participants in wenjuanxing. The link of the questionnaire is: https://www.wjx.cn/jq/77671056.aspx

6.1 The result of questionnaire

The rating questions are divided into 4 types: 1) Memorability measures the impression users can get from the visualizations. 2) Efficiency measures how much time the visualizations take the users to understand. 3) Subjective Satisfaction rates the user's subjective feelings to the visualizations. 4) The brushing means the degree of the two images are related to each other. The results of our questionnaire is shown as follow:

Choropleth Map				
Memorability	3.47/5			
Efficiency	3.83/5			
Subjective Satisfaction	3.7/5			
The Stacked Area Graph				
Memorability	4/5			
Efficiency	3.77/5			
Subjective Satisfaction	3.7/5			
The brushing of the Area graph and the Map	3.87/5			
The Aster Plot				
Memorability	3.8/5			
Efficiency	3.7/5			
Subjective Satisfaction	3.87/5			
The brushing of the Area graph and the Map	3.9/5			

6.2 Improvement

There are some comments for the map which suggest improving the color contrast of the map or use darker background color. According to the feedback, we need to set the threshold scales more finely to support richer color changes. According to the feedback, we increased the saturation of the lowest level of threshold scales so that it is more easily to distinguish between the colored country and the background. For the aster plot, in the first version we assigned different color to each country, which is meaningless. In the latest version, we gave them the same color so it will not make any distraction.

7 Changes along the way

We tried to make graphs since we are holding data about the actual generation of solar and wind energy in the various countries per year as well as the information about the potential generation for the same countries. The correlation between these 2 sets over time should hold information about the various countries' advancement in renewable energy.

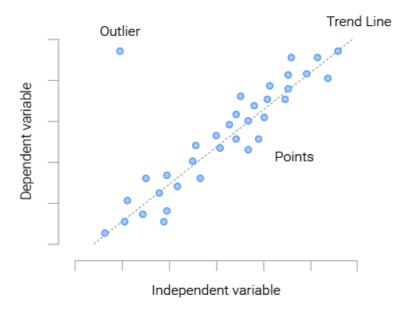


Fig.6 Scatter graphs

However, after being visualized, the scatter graphs are meaningless to give any useful information. Therefore, we had to look for more appropriate graphs. The aster plots, which can show the percentage intuitively as well as the comparison between countries, is selected finally.

8 Conclusion & Potential Extension

In this report, we present the visualization of Europe's energy production. By now, we have made the following type of visualization: 1) Choropleth maps 2) Stacked Area Graphs 3) Aster plot, which are up to our expectations. By these visualizations, it is feasible for supporting the users to follow the progress in changing energy generation.

Adding to our own evaluations, we also discussed our project with several participants. We are excited to further explore potential extension. First, in the current version the layout is fixed. It will have a better appearance if the responsive design is implemented which can respond to the user's behaviour and environment based on screen size, such as mobile phone as well as multi browsers. Secondly, the current project provides a limited description of the visualizations, therefore, we propose combining our visualizations with AI to get real-time analyzation makes it possible for users to view the visualization and at the same time can get analyzation in the future. As for the datasets, now our data is from 1989 to 2015, it will be better if we could find a reliable data source that the data could be updated automatically.

9 Workload Division and Gantt Chart

9.1 Workload Division

The workload division of our project is as follow:

Jeroen Grégoire: Select datasets, analyse demand, pre-process data, publish code online in dashboard style.

Quentin Julemont: Do the visualizations of Choropleth map, Stacked area graphs and Aster plots.

Mengyao Song: Make slides, write report, do the evaluation.



9.2 Gantt Chart

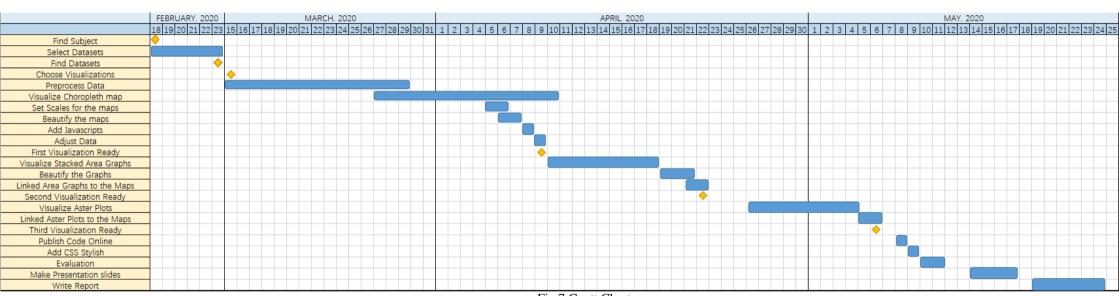


Fig.7 Gantt Chart



10 Bibliography

- [1] bp.p.l.c. (2019) *Energy Economics*, [web site] http://www.bp.com/en/global/corporate/energy-economics.html, retrieved on Febury 23, 2020 [2] Sohier Dane. (2017) *30 Years of European Solar Generation/Kaggle*, [web site] https://www.kaggle.com/sohier/30-years-of-european-solar-generation/data, retrieved on Febury 23, 2020
- [3] Sohier Dane. (2017) *30 Years of European Wind Generation/Kaggle*, [web site] https://www.kaggle.com/sohier/30-years-of-european-wind-generation/data, retrieved on Febury 23, 2020
- [4] Mariella CC. (2015) *Europe map step 1*, [web site] https://bl.ocks.org/MariellaCC/0055298b94fcf2c16940,
- [5] Sam. (2014) *mapsam*, [web site] http://mapsam.com/posts/d3-background-hash, retrieved on March 15, 2020
- [6] Chris Viau. (2019) *Simple D3 tooltip*, [web site] http://bl.ocks.org/biovisualize/1016860, retrieved on March 15,2020
- [7] Jane Pong. (2020) *Date slider with play/pause button*, [web site] https://bl.ocks.org/officeofjane/47d2b0bfeecfcb41d2212d06d095c763, retrieved on March 25, 2020
- [8] Bertspaan. (2015) *D3.js animated U.S. states border map*, [web site] https://github.com/maptime-ams/animated-borders-
- <u>d3js?utm_content=bufferb12b2&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer</u>, retrieved on March 25, 2020
- [9] Jason Heppler. (2017) *Threshold scale*, [web site] https://bl.ocks.org/hepplerj/f2f4e5f4a9321428b11614674c741177, retrieved on March 18, 2020
- [10] Mike Bostock. (2019) *Threshold Choropleth*, [web site] https://observablehq.com/@d3/threshold-choropleth, retrieved on March 18, 2020
- [11] Yan Holtz. (2018) *Basic stackedarea in d3.js from wide input format*, [web site] https://www.d3-graph-gallery.com/graph/stackedarea_wideinput.html, retrieved on April 12, 2020
- [12] Fabio Mainardi. (2015) *Stacked area chart with multiple tooltips*, [web site] https://bl.ocks.org/fabiomainardi/3976176cb36e718a608f, retrieved on April 12, 2020
- [13] Yong Lee. (2019) *Line Chart with Tooltip 1*, [web site] https://bl.ocks.org/Qizly/8f6ba236b79d9bb03a80, retrieved on April 12, 2020
- [14] Alan Dunning. (2018) *Line Chart with Circle Tooltip D3 V4*, [web site] https://bl.ocks.org/alandunning/cfb7dcd7951826b9eacd54f0647f48d3, retrieved on April 25, 2020
- [15] Dianaow. (2019) *D3 V4 Multiple-line chart: Hover tooltip*, [web site] https://bl.ocks.org/dianaow/0da76b59a7dffe24abcfa55d5b9e163e, retrieved on April 28, 2020 [16] Amber Thomas. (2019) *d3v4 Multi-Line Chart with Dropdown Menu and Nesting*, [web site] https://bl.ocks.org/ProQuestionAsker/8382f70af7f4a7355827c6dc4ee8817d, retrieved on April 28, 2020

[17] Ben Best. (2020) *Aster Plot in D3*, [web site] http://bl.ocks.org/bbest/2de0e25d4840c68f2db1, retrieved on May 4th, 2020 [18] Frank Guerino. (2015) *Simple D3 Pie Chart with Magnitudes in Arcs and Legends Outside and Along Arcs*, [web site] http://bl.ocks.org/Guerino1/2295263, retrieved on May 4th, 2020