

UNIVERSITY OF MACEDONIA
MASTER OF SCIENCE PROGRAM
DEPARTMENT OF APPLIED INFORMATICS

DIGITAL DATA STORYTELLING OF PUBLIC DATA WITH USE OF PLOTLY'S
TOOLS
Master Thesis

by

Zacharias Samos

Thessaloniki, October 2019

DIGITAL DATA STORYTELLING OF PUBLIC DATA WITH USE OF PLOTLY'S
TOOLS

Zacharias Samos

BA Honours degree in Mechanical Engineering, AUTH, 2014

Master Thesis

submitted in partial fulfilment of the requirements of

MASTER OF SCIENCE PROGRAM
IN APPLIED INFORMATICS

Adviser
Georgia Koloniari

Approved by the three members committee at 30/10/2019

Georgia Koloniari

Georgios Evangelidis

Emmanuel Stiakakis

.....

.....

.....

Zacharias Samos

.....

Abstract

Data visualization has grown rapidly in terms of popularity in recent years. Communicating the findings from visualizing data to third parties is much more effective and is called digital data storytelling. In the context of this thesis, an overview of these disciplines was provided and great examples of data visualization were also presented. A data storytelling scenario has been developed that utilized public data to analyse and present the results of this analysis visually. The theme of the story is related to the sales of passenger cars and their characteristics registered in the last 8 years in Europe. Charts have been created with the help of Plotly Chart studio. This tool has the ability to create dynamic, interactive data visualizations in web browsers. The Plotly Chart Studio has proven to be a tool that offers many helpful capabilities to the user

Keywords: data visualization, data storytelling, Plotly Chart studio

Acknowledgements

I would first like to thank my thesis supervisor Assistant Professor Georgia Koloniari of the Department of Applied Informatics at University of Macedonia for her guidance and support during the selection of the subject, the design and elaboration of the present thesis.

I must also express my very profound gratitude to my family and to my friends for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Table of Content

1 Introduction	1
1.1 Problem statement – Significance	1
1.2 Scope	2
1.3 Contribution	3
1.4 Structure of the study	3
2 Methodology	5
2.1 Data visualization	5
2.1.1 Definitions	5
2.1.2 Why data visualisation is important?	5
2.1.3 Key figures in History of data visualization	7
2.1.4 Difference between static vs interactive data visualization	11
2.1.5 Exploratory vs explanatory analysis	11
2.1.6 Demonstration and comparison of data visualization tools	13
2.1.7 Design techniques	19
2.1.8 Great Examples of Data Visualization	21
2.1.9 Data visualization sites	25
2.2 Data storytelling	27
2.2.1 Definitions of Storytelling	27
2.2.2 The Biochemistry of Story	28
2.2.3 History of data storytelling	28
2.2.4 Trends in data storytelling	29
2.2.5 Best practices of data storytelling	29
2.2.6 Different types of story structures	30
3 Plotly Chart studio demonstration	34
3.1 Import Data	34
3.2 Interactivity tools	35
3.3 Scatter plots	36
3.4 Bubble charts	39
3.5 Line chart	40
3.6 Bar chart	42
3.7 Stacked bar chart	43

3.8 Heatmaps	46
3.9 Style the plots	47
4 Case study	52
4.1 Description	52
4.2 How the passenger car market evolved?	55
4.3 Average CO2 emissions in 2017	56
4.4 Historical CO2 emissions and adopted targets	57
4.5 Average parameters 2017	58
4.6 Heatmaps – Average parameters	59
4.7 Market share of new car sold in the EU per fuel	60
4.8 Market share evolution for top manufacturers	61
4.9 CO2 emissions per powertrain	62
4.10 Real world vs Euro standards limits NOx emissions per fuel type	63
4.11 New registrations of PHEVs	63
4.12 New registrations of BEVs	64
4.13 Technology shares	65
5 Conclusions	66
5.1 Research overview and findings	66
5.2 Limitations	67
5.3 Future research	68
6 References	69

List of Figures

Figure 1 Growth of data in the world [1], [2]	1
Figure 2 Bandwidth of our senses by Tor Norretranders.	6
Figure 3 The power of visuals [11]	7
Figure 4 Exports and imports to and from Denmark and Norway into England for the time period 1700-1780 [12]	8
Figure 5 Cholera cases in Soho (London) in 1854 [13]	9
Figure 6 Diagram of the causes of mortality in the army in the east.[14]	10
Figure 7 Age group of the Swedish population between the 18th and 19th centuries. [14]	10
Figure 8 Initial visual “Satisfaction loss by things gone wrong” - Exploratory analysis [15]	12
Figure 9 Steering wheel are top reasons for non-complete satisfaction. Explanatory analysis [15]	13
Figure 10 Collection of D3.js examples [17]	15
Figure 11 Sample chart from Chart.js [18]	16
Figure 12 Sample chart from Highcharts.[19]	17
Figure 13 Sample chart from Plottable.js [20]	18
Figure 14 Sample chart from Plotly.js [21]	19
Figure 15 How to lie with a bar chart.[22]	20
Figure 16 How to lie with a pie chart.[23]	21
Figure 17 Languages in the World	22
Figure 18 What's Really Warming the World?	23
Figure 19 Why Buses Bunch?	24
Figure 20 How People Like You Spend Their Time?	25
Figure 21 Progression of a story	31
Figure 22 Author-driven narratives	31
Figure 23 Viewer-driven narratives	32
Figure 24 The martini glass narrative structure	33
Figure 25 Plotly Chart studio GUI	34
Figure 26 Chart studio chart types	36
Figure 27 A scatterplot of random data with Plotly	37

Figure 28 Code for the scatterplot with random data	38
Figure 29 Add a trace in Plotly	38
Figure 30 An example of a bubble chart with Plotly	39
Figure 31 Code for the bubble chart	39
Figure 32 The previous bubble chart with a colorscale	40
Figure 33 Line chart showing three different modes	41
Figure 34 Code for the line chart	41
Figure 35 An example of bar chart with Plotly	42
Figure 36 Code for the bar chart	43
Figure 37 An example of stacked bar chart with Plotly	43
Figure 38 Option to switch from a single bar chart to a stacked bar chart	44
Figure 39 The anatomy of a boxplot	45
Figure 40 Set the properties of a histogram with Plotly	46
Figure 41 An example of heatmap chart with Plotly	46
Figure 42 Code for the heatmap	47
Figure 43 Available themes	48
Figure 44 Style tab	49
Figure 45 Annotation menu	50
Figure 46 An example of a curve fitting in a scatterplot with Plotly	51
Figure 47 How the passenger car market evolved?	55
Figure 48 How the passenger car market evolved? The cases of Greece, Portugal, Hungary and Sweden	56
Figure 49 Average CO2 emissions in 2017 from passenger cars in Europe	57
Figure 50 Historical CO2 emissions and adopted targets	58
Figure 51 Average mass, engine power and capacity in 2017	59
Figure 52 Average parameters through a heatmap	60
Figure 53 Market share of new car sold in the EU per fuel	61
Figure 54 Market share evolution for top manufacturers	62
Figure 55 CO2 emissions per powertrain	62
Figure 56 Real world vs Euro standards limits NOx emissions per fuel type	63
Figure 57 New registrations of PHEVs	64
Figure 58 New registrations of BEVs	65
Figure 59 Technology of passenger car shares in 2017	66

List of Tables

Table 1-1 A ranking of the top 10 JavaScript data visualization packages.	14
Table 2 EEA CO2 monitoring database	54

Abbreviations

API	Application program interface
BEV	Battery electric vehicle
CEO	Chief executive officer
CO2	Carbo dioxide
EDA	Exploratory data analysis
EEA	European environment agency
EPA	Environmental protection agency
EU28	European union 28 countries
FCEV	Fuel cell electric vehicle
GDP	Gross domestic products
GHG	Greenhouse gases
GUI	Graphical user interface
IQR	Interquartile range
NASA	National aeronautics and space administration
NEDC	New european driving cycle
Nox	Nitrogen oxides
PHEV	Plug-in hybrid electric vehicle
PM	Particulate matter
RSS	Really simple syndication
SQL	Structured query language
SVG	Scalable vector graphics
TDI	Turbocharged direct injection
URL	Uniform resource locator
US	United states
VW	Volkswagen
WLTP	World harmonized light-duty vehicles test procedure

1 Introduction

1.1 Problem statement – Significance

In recent years there has been an explosion of digital data production. This massive data growth is attributed to the fact that nowadays there are much more tools, devices and means available to people that are able to create and share information than even before. In fact, there is an exponential rise in ways to receive information by; social media, websites, e-mail, data from mobiles, RSS feeds etc. An overview of this incredible data growth rate has been presented in the IBM study 2017 which highlights that 90 % of the world's data have been generated only in the last two years [1]. This excessive amount of information can actually make it difficult for people to retrieve the required information from different sources and generally complicate data manipulation. However, in the workplace the information revolution is seen as a key to success and a way to gain competitive advantage to outperform the opponents.

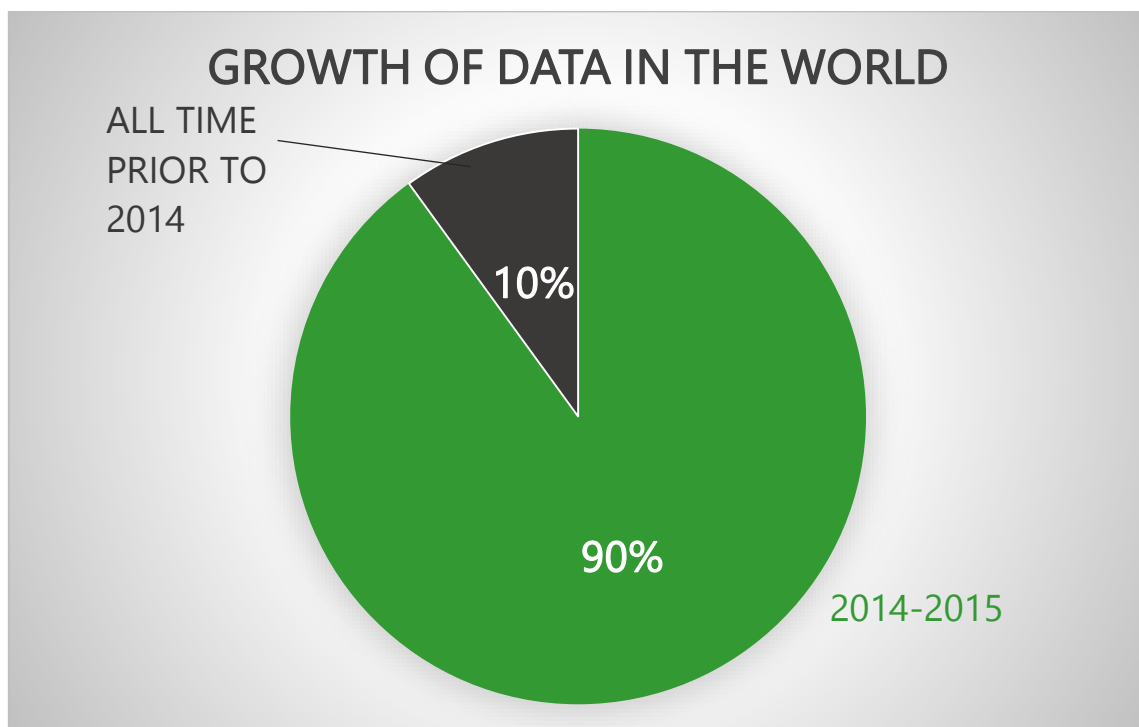


Figure 1 Growth of data in the world [1], [2]

Data contain valuable information that can be used to benefit businesses / organizations etc. Simply by exploring data (data exploration), managers / bi analysts can easily understand more about their businesses and to leverage this knowledge directly in order to make better decisions, follow better strategies and therefore increase their performance. This exploration involves visualizing the data.

Data visualization is a relatively new scientific field that has grown rapidly in terms of popularity in recent years. The term "data visualization" includes every effort to help people understand the importance of data by placing them in a visual context. Patterns, trends, and correlations that matter and may not be easily found in spreadsheets can be revealed and recognized through graphs. Moreover, visualizing the information allows the focus only in the important parts. It is commonly accepted that the human brain processes much better and easier visual data than language, words, numbers, tables etc [3].

However, visualizing data itself is not enough to convey a message. Simply visualizing data only express what is happening but it does not tell why this is happening. Communicating the findings from visualizing data to third parties (audiences, customers, etc.) with a compelling narrative that encourages them to engage in the story is much more effective and is called digital **data storytelling**. Data storytelling is considered to help significantly marketers / speakers to achieve their communication goals in many different ways.

Having said all the above, it is appropriate to examine these scientific areas more closely and, by implementing a technical example (case-study), to highlight why they are considered so important.

1.2 Scope

In the context of this thesis, a data storytelling dashboard has been created by utilizing public data sources. The data has been analysed in order to reveal interesting trends. The results of this analysis have been presented visually into a dashboard in a way that composes a story. The ultimate goal of this story is to engage readers/audience, to motivate them to interact with the graphs, and finally to understand and assimilate the main messages that the story wants to stress out. Graphs were created using the Plotly library of Javascript which has the ability to create interactive visualizations of data in web browsers.

The theme of the story created has to do with the sales of passenger cars recorded in the last 8 years in Europe. The dashboard has been designed in a manner to compose a story and gather the most interesting points of the data and present it to the public. The dashboard was based on the EEA's (European Environmental Agency) database which is published every year and contains the main characteristics of the cars sold in the European Member States (e.g. Engine displacement, Mass, CO2 emissions etc.).

1.3 Contribution

This study explores the scientific areas of data visualization and data storytelling. Through the literature review, these areas are analyzed and the abilities offered by them are presented. More specifically, the analysis includes definitions, explanation of the major features of these scientific fields, and presentation of some noteworthy examples.

The technical part of the thesis that involves the construction of a story through data is accompanied by a complete demonstration of the tool that was used to create the story. The tool that was used was the Chart studio of Plotly which is a Javascript's library. The main features of this library are investigated in the study and also this library is compared to other available tools for data visualization.

Initially the data was collected and gathered in a comprehensive database. The database was cleaned in order to facilitate the data analysis. Errors in the data were corrected as inaccurate or outlier values, some values stored as text were converted to numbers, etc. Then, the appropriate queries were created to begin analysing the data and finally the presentation of the data which tells a story about the passenger cars in Europe.

1.4 Structure of the study

In this section, the structure of the study is described. Specifically, chapter 2 provides information on the literature review conducted for collecting relevant material. Chapter 3 is the backbone of work, containing thoroughly analysis of the fields of interest. In this chapter, there are subchapters 3.1 and 3.2 where the fields of data visualization and data storytelling are explored respectively. Section 3 contains a presentation of the tool (Plotly chart studio) that was used to carry out the case study, which in turn is discussed in

Section 4. Chapter 5 includes the outline of the study, summarizing the conclusions of the work and talking about the bounds and limitations of the work, as well as the future extensions that can be studied. Finally, in chapter 6 there is the literature used in this study.

2 Methodology

2.1 Data visualization

2.1.1 Definitions

There are many ways to define data visualization. Some of the scientists in this field have given the following definitions:

Scott Murrey professor at the University of San Francisco defines data visualization as follows: *“Data visualization is a process of mapping values or information to visual content. The process of visualization can help us see the world in a new way, revealing unexpected patterns and trends in the otherwise hidden information around us.”*[4].

Cole Nussbaumer Knaflic, Founder & CEO of storytelling with data outlines that *“Data visualization is about turning numbers into pictures and stories which are relative easily understood by humans.”* [5]

According to Vitaly Friedman co-founder of the Smashing Magazine:

“The main goal of data visualization is its ability to visualize data, communicating information clearly and effectively. It doesn't mean that data visualization needs to look boring to be functional or extremely sophisticated to look beautiful. To convey ideas effectively, both aesthetic form and functionality need to go hand in hand, providing insights into a rather sparse and complex data set by communicating its key-aspects in a more intuitive way. Yet designers often tend to discard the balance between design and function, creating gorgeous data visualizations which fail to serve its main purpose — communicate information.” [6]

2.1.2 Why data visualisation is important?

It's often said that a picture is worth a thousand words and data visualization works the same way. A visual representation or a chart may contain a lot more information than text for a similar amount of space on a page/slide etc. Also, it has been proved that visuals are more efficient and effective than words when processed by human brain. Pictures are not just easier to perceive and process than words, but also easier to remember.

Danish Physicist Tor Norretranders created a visual to correlate human senses to computer terms and named it “Bandwidth of our senses”.

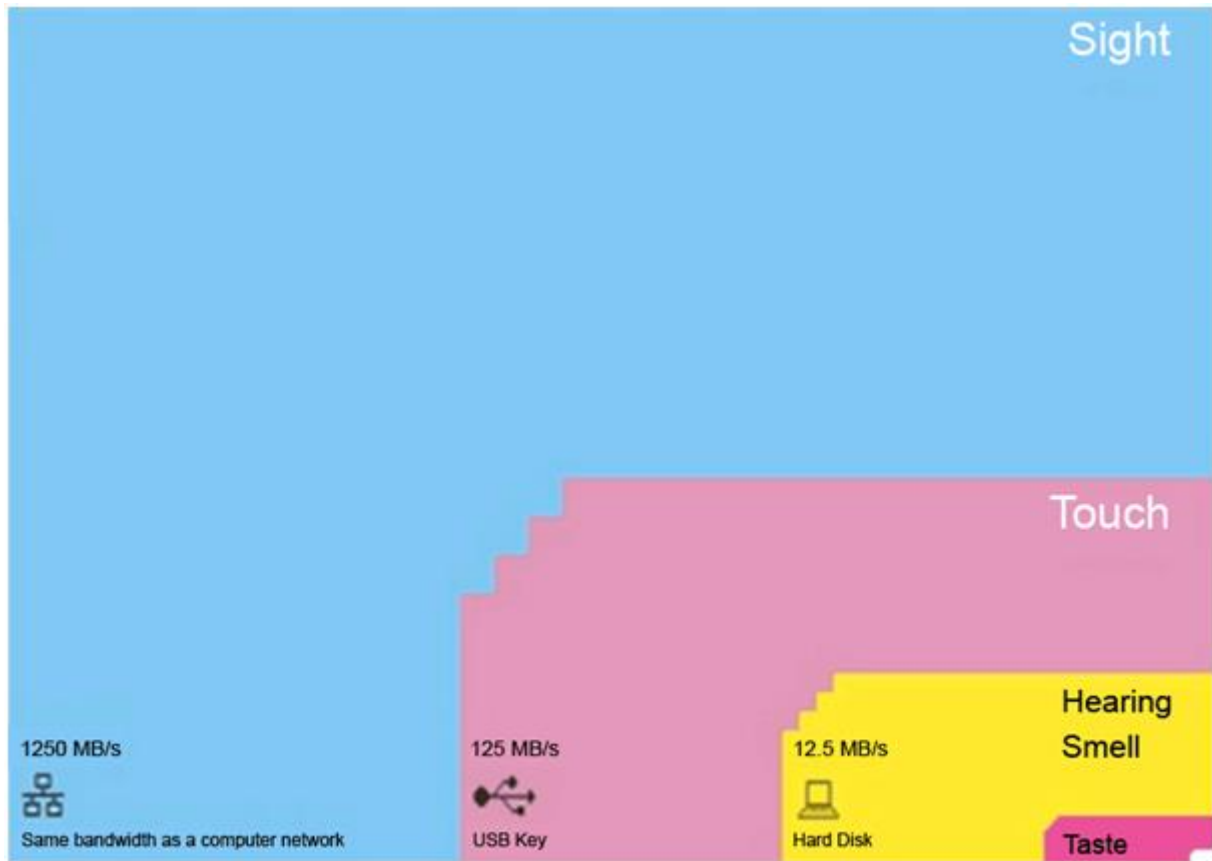


Figure 2 Bandwidth of our senses by Tor Norretranders.

Analytically, this visual compares the amount of information perceived by human senses per second. As seen from the above picture, vision has the dominant role among the senses, as the processing speed of vision is 10 times faster than the second strongest sense which is touch. There follows hearing and smell that perceive information 100 times slower than sight. Finally, it is worth noting that the small white spot on the right lower corner represents that humans are only aware of 0.7% of the information they experience.

A few other facts showing that visuals are more powerful than words:

- Visuals are processed 60,000 x quicker than text by human brain. [7]
- 90 % of information that comes to the brain is visual. [8]
- Individuals can recall the content of 2500 pictures with more than 90 % accuracy 72 hours later, looking at them for only 10 seconds. After a year, participants had 63 % recall of those same pictures. With traditional lecture format delivery, students only remember 10 % of the material 72 hours later. [9]

- Human eyes can enroll 36,000 visual messages per hour. [10]

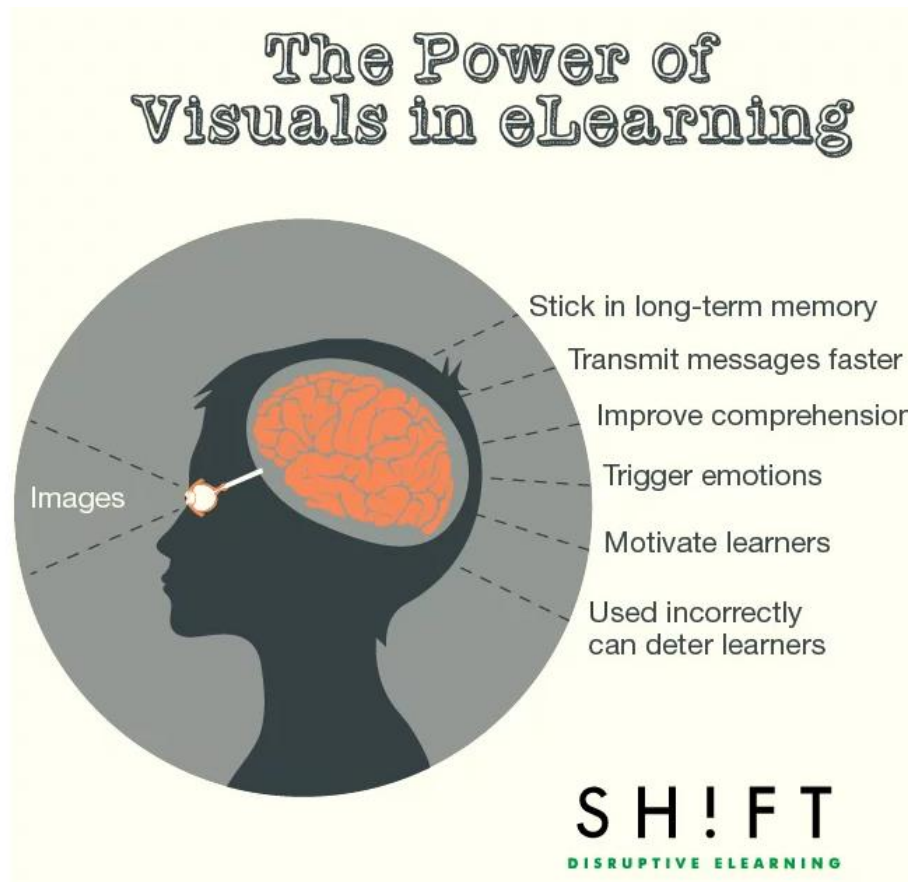


Figure 3 The power of visuals [11]

Any kind of images and visualizations have the ability to affect more their audience than written words and numbers. This is because, in general, visualizations are more closely tied to human memory. In addition, graphs or visuals repeat the message in a unique way, enhancing their impact on your audience.

However, in order to maximize the effectiveness, the correct chart or display for the situation should be utilized. Selecting the wrong data visualization might overwhelm or confuse the audience – which might lead to the opposite of the intended results.

2.1.3 Key figures in History of data visualization

It is commonly thought that data visualization is a recent development. However, data visualization has its origins in the very long past. This chapter describes some of the

most remarkable examples of data visualization of the past that shaped the current version of data visualization as we know it.

William Playfair (1759–1823)

A Scottish engineer considered as the father of graphical methods of statistics. He invented line, bar, area and pie charts that are commonly used today. One of his well-known graphs is depicted below where there is a comparison between exports from England with imports from Denmark and Norway into England for the time period 1700-1780.

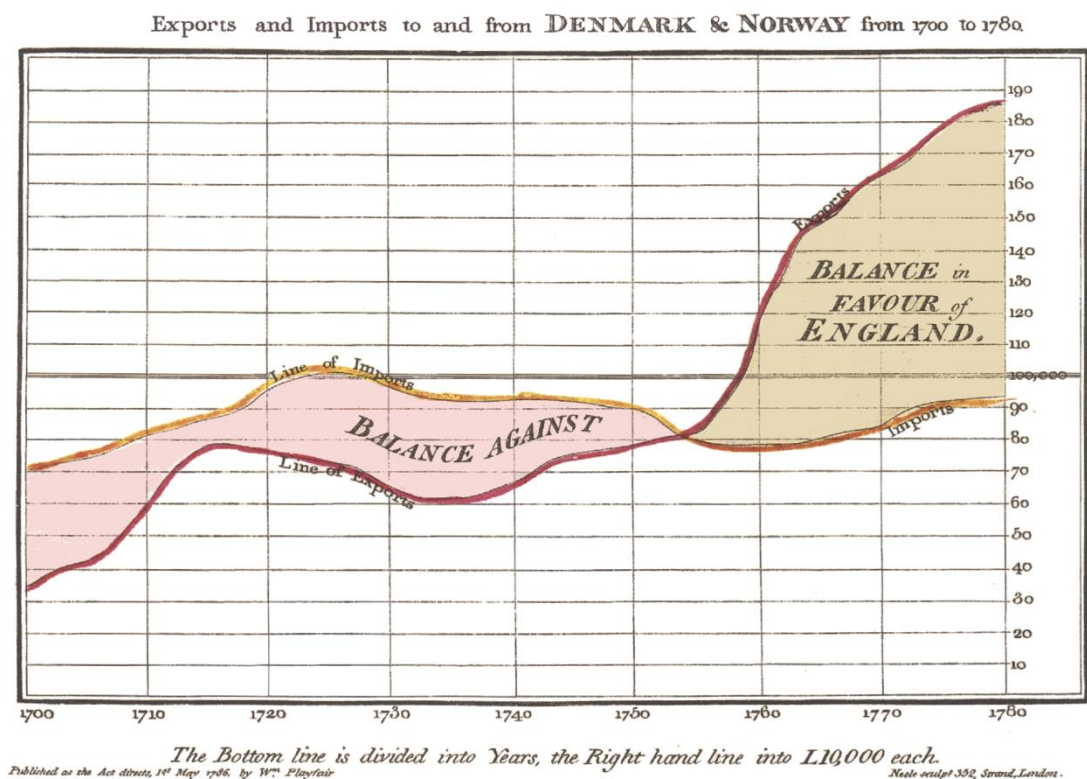


Figure 4 Exports and imports to and from Denmark and Norway into England for the time period 1700-1780 [12]

John Snow (1813–1858)

John Snow was a physician and is widely recognized as one of the founders of modern epidemiology. His best-known work is about the London's cholera outbreak in 1854. 127 people on or near Broad Street died. John Snow documented deaths around the community to determine what might be the cause of the epidemic. He created the following

map by recording the location of each death by a bar. The height or the length of these bars indicate that more deaths occurred in a particular location.

This map helped him to identify the source of the outbreak as the public water pump on Broad Street. Firstly, he discovered a meaningful anomaly in the results. No one of the workers employed in the close Broad Street brewery contracted cholera. As it turns out, the brewery gave an unlimited beer allowance to its workers, and since the fermentation process killed the cholera bacteria, none of the workers diseased.



Figure 5 Cholera cases in Soho (London) in 1854 [13]

Florence Nightingale (1820–1910)

During the Crimean War, Florence Nightingale was serving as a nurse. The terrible hygiene conditions that prevailed on the hospital facilities prompted her to carefully track the death rates in the hospitals and visualize the data. Nightingale is considered as a pioneer in the graphical representation of statistics. Its best-known contribution is a form of the pie chart presently known as the polar area chart. Her 'rose' graphs helped her battle for good sanitation and better medical conditions and eventually save lives.

2.1.4 Difference between static vs interactive data visualization

This section explores the difference that has a static visualization compared to a corresponding interactive one.

Static visualization can be considered as the most simple and straightforward way to present data. They combine images (accompanied with text) to present concepts in the form of graphs, plots and maps etc. Since, these infographics don't require user input, it should be carefully considered what data is and what is not being displayed because the viewer adjusts the view of final results. Interactive visualizations contain controls like zooming and filtering that give a lot more options. However, increasing the level of interactivity requires users to think and make their own decisions about what data to look at rather than presenting them precomposed views of the data. In many cases this increase level of interactivity and user autonomy is essential for the visualization to serve its purpose and meet the user's goals. Interactivity favors the presentation of the results from large datasets that contain complex relationships. It is easier to see the big picture but also to examine a lower level of detail. Ben Shneiderman, a data visualization pioneer has said "Overview first, zoom and filter and then details on demand".

Depending on the message or story the author has to say, he has to choose the appropriate visualization of his data (static or interactive). Some stories are relatively simple and straightforward and intended to present a single point. In this case, a static infographic is able to communicate the story in a straightforward but at the same time, compelling way.

2.1.5 Exploratory vs explanatory analysis

These terms can often seem very similar. **Exploratory** data analysis (EDA) is all the things that the author should do to obtain deep knowledge of the data. It contains all the questions that should be done that will help the author to determine all the interesting parts of the data. First step is the creation of some basic plots, scatter plots, histograms and other charts that will be very helpful in order to understand the data. In other words, exploratory analysis is a conversation between author and the data.

Explanatory analysis contains specific highlights of the data that are presented to the audience and is essentially the next step of the exploratory analysis. Correspondingly, we could say that explanatory analysis is a conversation between data and the audience.

An example that presents how the exploratory and explanatory analysis could differ in a specific visual. Suppose consumer-drivers provide feedback about how much they are disturbed by various things that go wrong while driving. The initial visual could be the following:

Satisfaction loss by things gone wrong

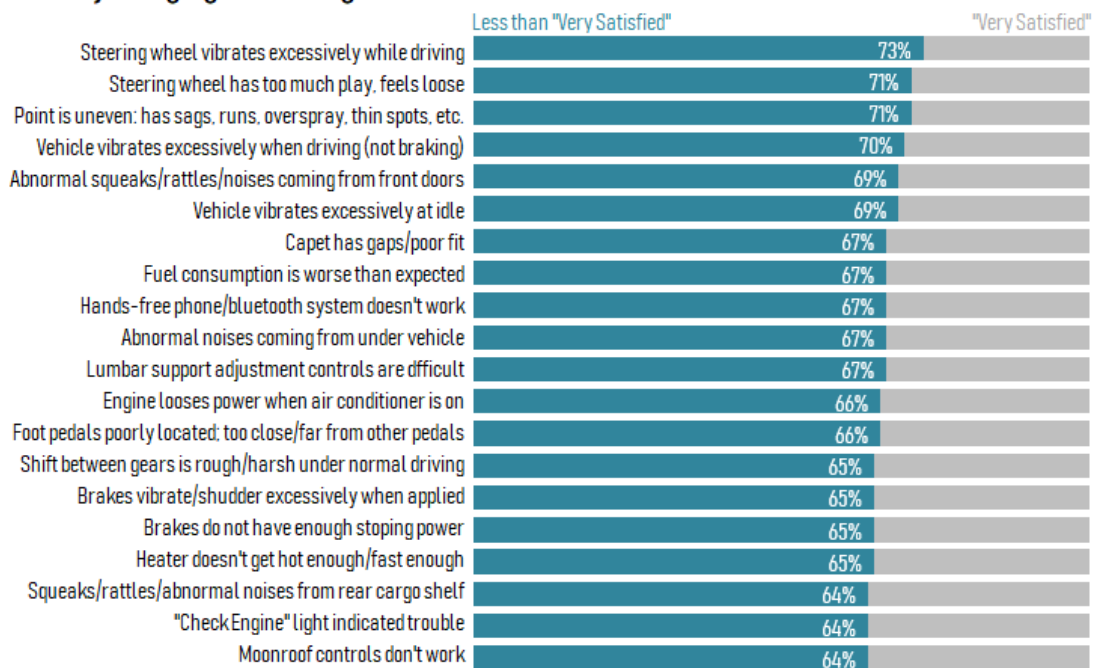


Figure 8 Initial visual “Satisfaction loss by things gone wrong” - Exploratory analysis

[15]

In the above diagram, the "less than very satisfied" responses (in deep blue) have been grouped, and presented in descending order. This process is very useful part of the exploratory analysis.

Once you've identified something or some things to focus on, in some cases it will make sense to create a different visual for the purpose of focusing on that thing or those things. Alternatively, the same visual can be modified for explanatory purposes by drawing attention to the points of interest, while preserving the other data for context:

When, things of the data that have particular interest for the analysis are identified, a new chart can be created. Alternatively, this chart can be modified or adjusted to highlight

those things. This is a typical example of explanatory analysis. In the following Figure 9, it is highlighted that the top 2 issues with the highest percentage of non-complete satisfaction are both steering wheel related.

The **top 2 issues** with the highest percentage of non-complete satisfaction are **both steering wheel** related.

Satisfaction loss by things gone wrong

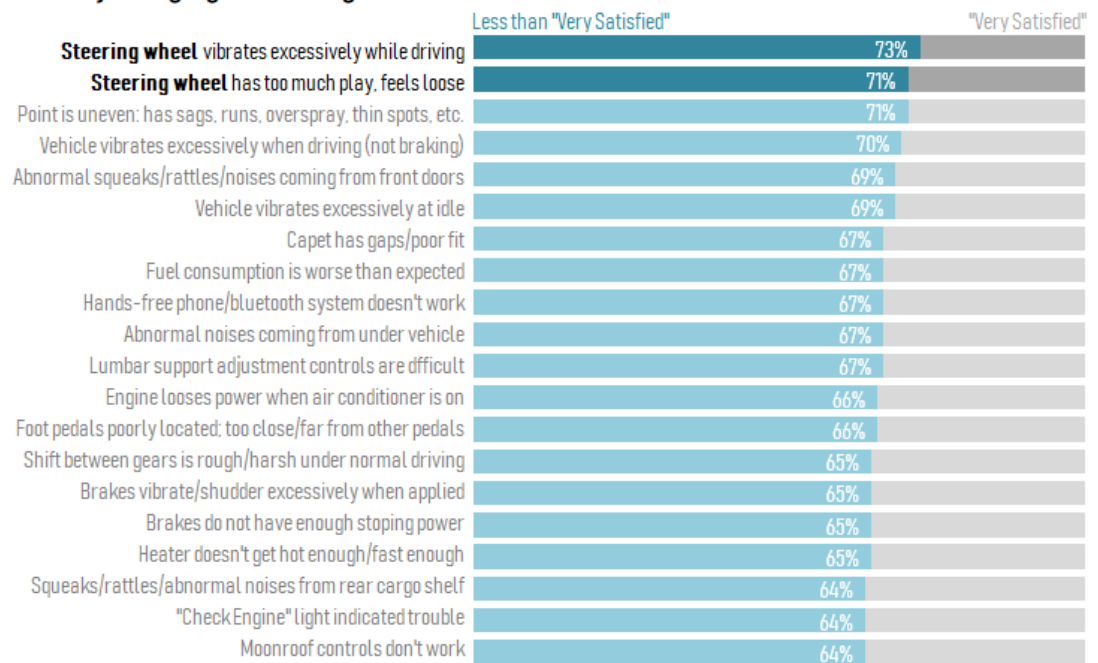


Figure 9 Steering wheel are top reasons for non-complete satisfaction. Explanatory analysis [15]

It should be noted that when focusing on specific data points, it is harder to see the whole picture. That is one of the main reasons why it's critical to do exploratory analysis before the explanatory phase.

2.1.6 Demonstration and comparison of data visualization tools

There are numerous tools for data visualizing, however, depending on the data to be used, the most appropriate tool should be selected. There are plenty of parameters to be considered when selecting the data visualization tool. The most important are listed below:

- The amount of data
- Format of the data

- What needs to be analyzed
- The data interconnections and interrelationships

JavaScript is the most simple and widespread technology for visualizing data on the web. Below are some of the most important JS charting libraries that can be used for complex or simple data visualization tasks. There are also many tools for visualizations that work locally in the computers. However, due to the increased number of these tools, they are not analysed in this study.

D3.js

D3.js is the most known JavaScript library for data visualization. It is open-source and through the very active community (20000 forks on Github) holds the dominant position among the charting libraries. This library contains a very helpful documentation and multiple examples which encourage users to utilize them by adjusting their data. According to “Data Incubator”, d3.js ranked 1st among 110 data visualization packages based on Github and Stack Overflow activity, as well as npmjs (javascript package manager) downloads. The Table 1 Table 1-1 A ranking of the top 10 JavaScript data visualization packages displays normalized scores where a value of 1 means one standard deviation above average (average = 0). [16]

Table 1-1 A ranking of the top 10 JavaScript data visualization packages.

Package	Rank	Overall	Github	Stack Overflow	Downloads
d3	1	16.34	7.5	4.84	-0.36
chart.js	2	4.93	3.27	0.45	-0.14
highcharts	3	3.96	0.45	2.68	-0.54
plottable	4	3.74	-0.23	2.88	-0.31
plotly.js	5	3.18	0.16	3.16	1.57
leaflet	6	2.92	1.46	0.78	1.48
britecharts	7	2.41	-0.22	-0.23	1.18
echarts	8	2.31	2.16	-0.21	1.36
c3	9	0.91	0.29	0.51	0.77

Chart.js is a very customizable JavaScript library for data visualisation particularly popular with web designers and developers. It offers an excellent basic solution for those who don't need lots of chart types and customization features but want their charts to look simple, clear and informative at first sight.

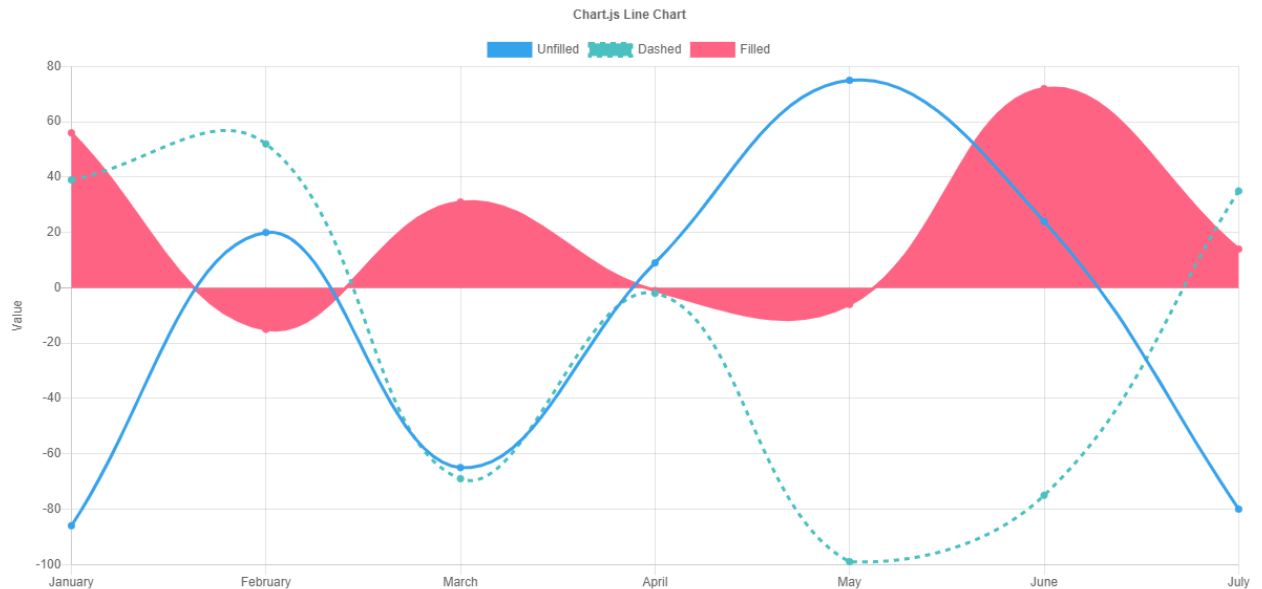


Figure 11 Sample chart from Chart.js [18]

Key features

- Eight chart types: line, area, bar, pie, radar, polar, bubble, and scatter.
- Customizable and responsive (when used on line)
- The use of plugins can extend functionality of Chart.js.

Highcharts

Highcharts is a popular JavaScript charting library based on SVG which supports plenty of chart types and also providing extensive customization options that are easy to use.

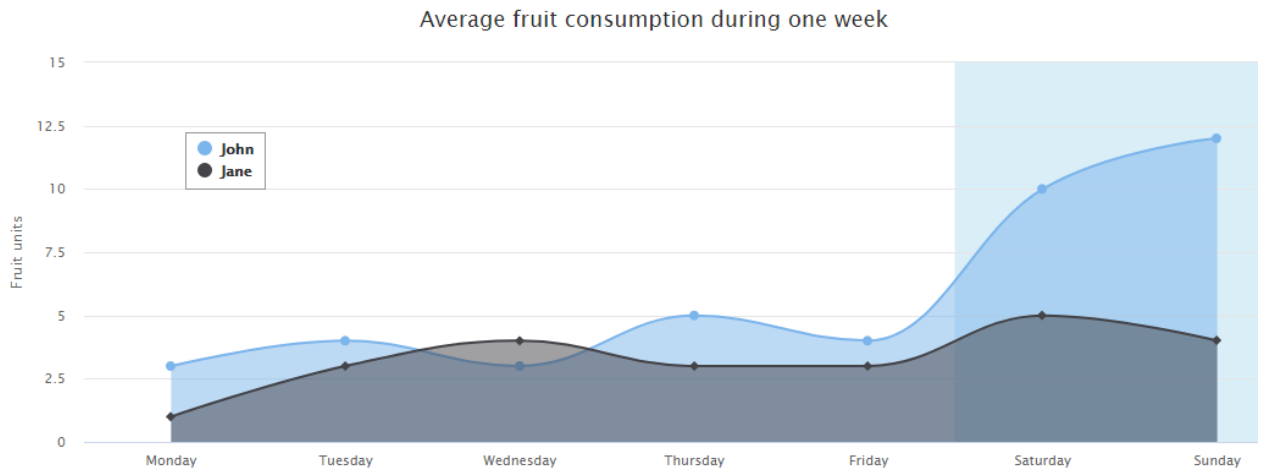


Figure 12 Sample chart from Highcharts.[19]

Key features

- Uses pure JavaScript, and data can be loaded externally.
- Robust documentation
- Can be used with React, Angular, Meteor, .NET, iOS, etc.

Plottable.js

Plottable is Built on top of D3 and gives web designers or developers a set of flexible, premade components that you can combine and rearrange to build charts.

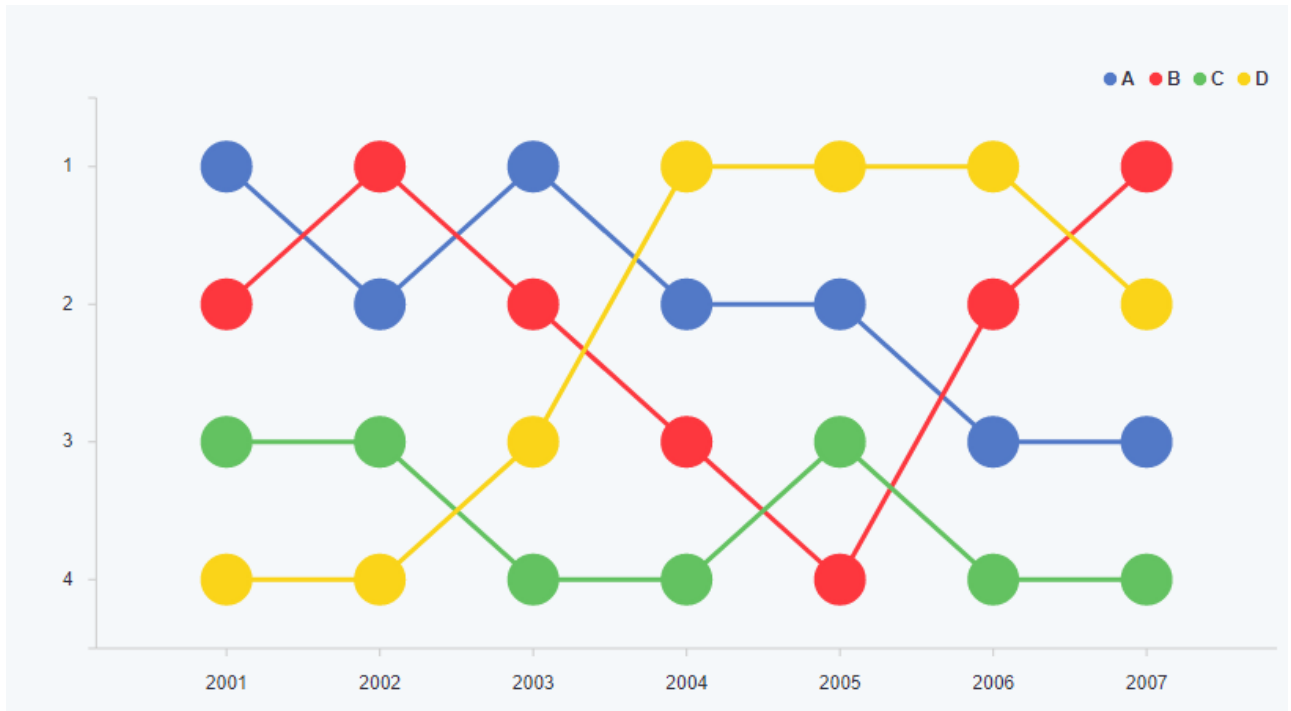


Figure 13 Sample chart from Plottable.js [20]

Key features

- Charts are easier to create and more reusable than charts built only with D3 because Plottable handles layout, sizing, and positioning
- The flexibility of D3 is still available.

Plotly.js

Plotly.js is a free and open-source high-level JavaScript library which built on D3.js and WebGL. It is used to create many different types of sophisticated and interactive charts. This library was used in the case study developed for the purpose of this thesis and will be further explored in a next chapter.

Plotly for Javascript and React

Create sophisticated, interactive charts in JavaScript for finance, engineering, and the sciences.

Plotly's JavaScript library is free, open-source, and does not require an online plot.ly account. Download today on GitHub.

GET STARTED

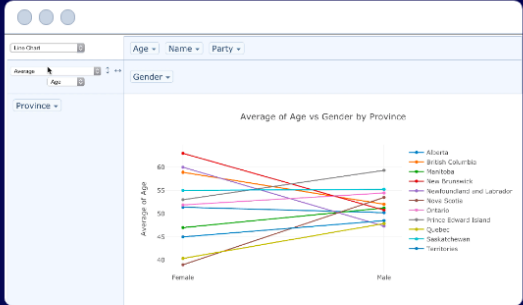


Figure 14 Sample chart from Plotly.js [21]

Key features

- A lot of chart types that can be embedded in websites or used to create dynamic presentations.
- A lot of different samples that can be used for inspiration in the web.
- Allows using Excel spreadsheets, or connect to your database.
- Extensive API documentation.
- Uses React.

2.1.7 Design techniques

The display of the data should be done as objectively as possible. In the past, there have been many examples where this data visualization rule has not been respected. In this section, through specific examples, the techniques that used to communicate messages that did not correspond to reality will be explored and in general how subjective data visualizations can mislead readers.

Bar charts subjective visualization example

A typical example of a strongly subjective visual is a chart depicted by Fox news back in 2012. In particular, the chart presented below wants to present to the audience what the change in the top tax rate will be if the Bush tax cuts expire, so they show the rate now and what'd it be in 2013. A big gap difference between the two bars and the increase in 2013 looks huge because of the lengths of the bars. This was caused by the fact that y axis doesn't start at zero, but at 34% which means the way this is plotted, the visual increase

between the two is 460%, versus the actual increase if we plot this you know, with the bars going all the way down to where they should be, the increase is something like 13%.

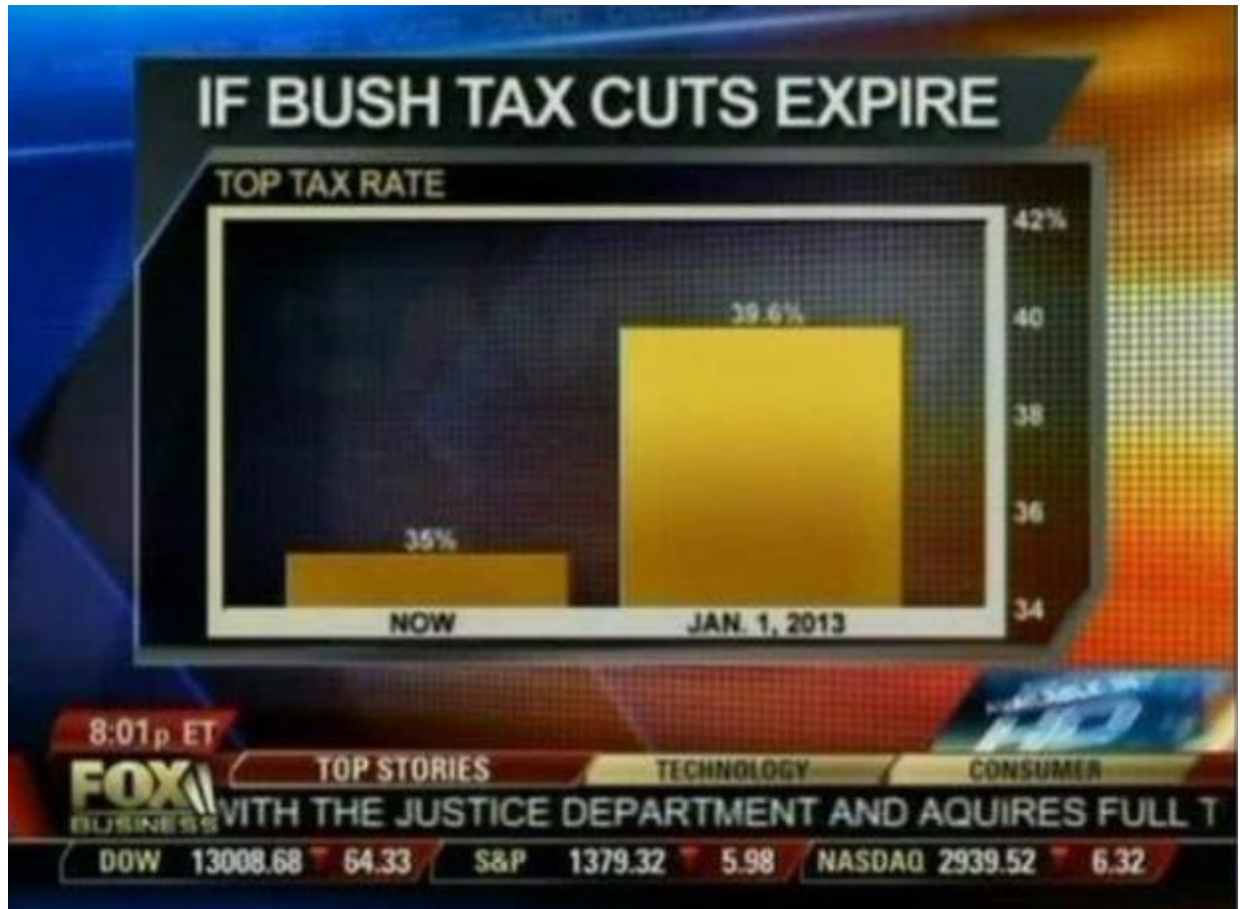


Figure 15 How to lie with a bar chart.[22]

Pie charts subjective visualization example

Another example of a data visualization, that tells a subjective story, rather than an objective one has been presented by Steve Jobs which show the US smartphone market share in a 3D pie chart. This chart can become very confusing for the audience. The 19.5% represented by the green part of the pie (Apple) looks to the eye larger than the 21.2% represented by the blue part.

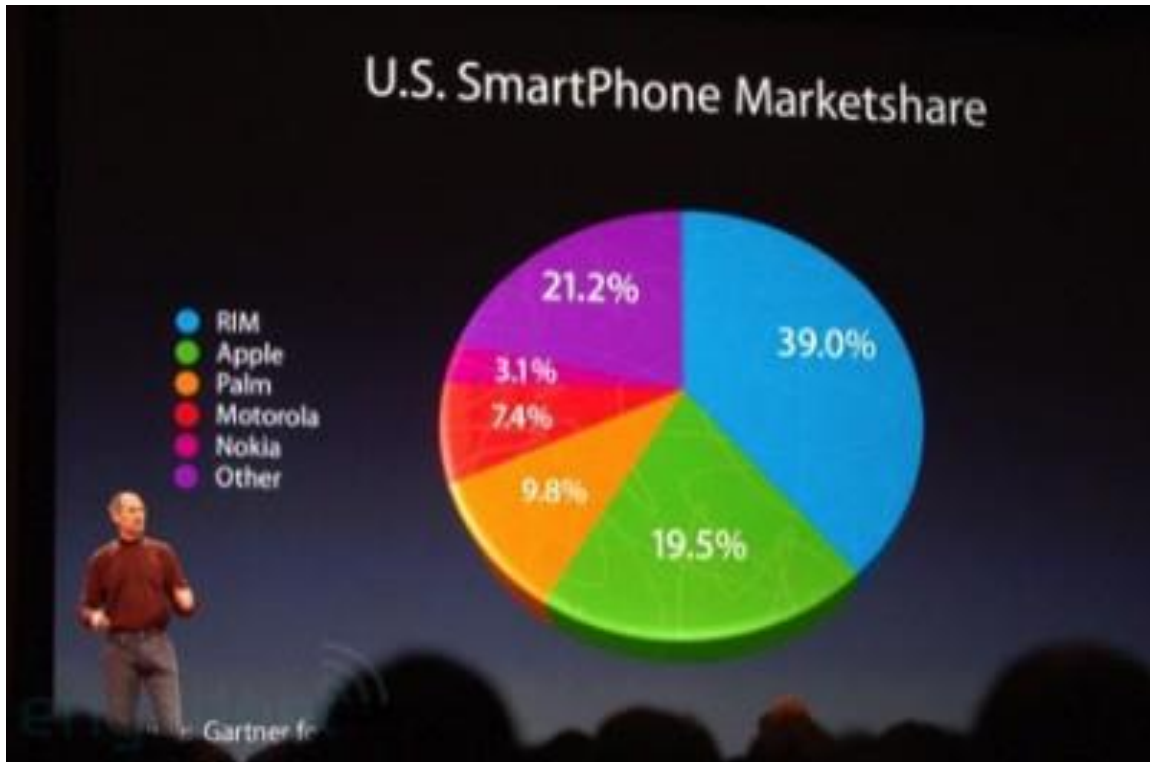


Figure 16 How to lie with a pie chart.[23]

2.1.8 Great Examples of Data Visualization

This section investigates some of the most interesting data visualization applications, which stand out for both their aesthetics but also for the direct and easy way in which they transmit complex information to the reader.

Languages in the World

In this interactive visual, the languages spoken in the world are presented. The user is encouraged to explore topics such as common languages families, discover which languages are most spoken, and see what languages are spoken round the world. This is an excellent visual narrative, where a complex subject is broken down into pieces which the reader can easily understand [24]

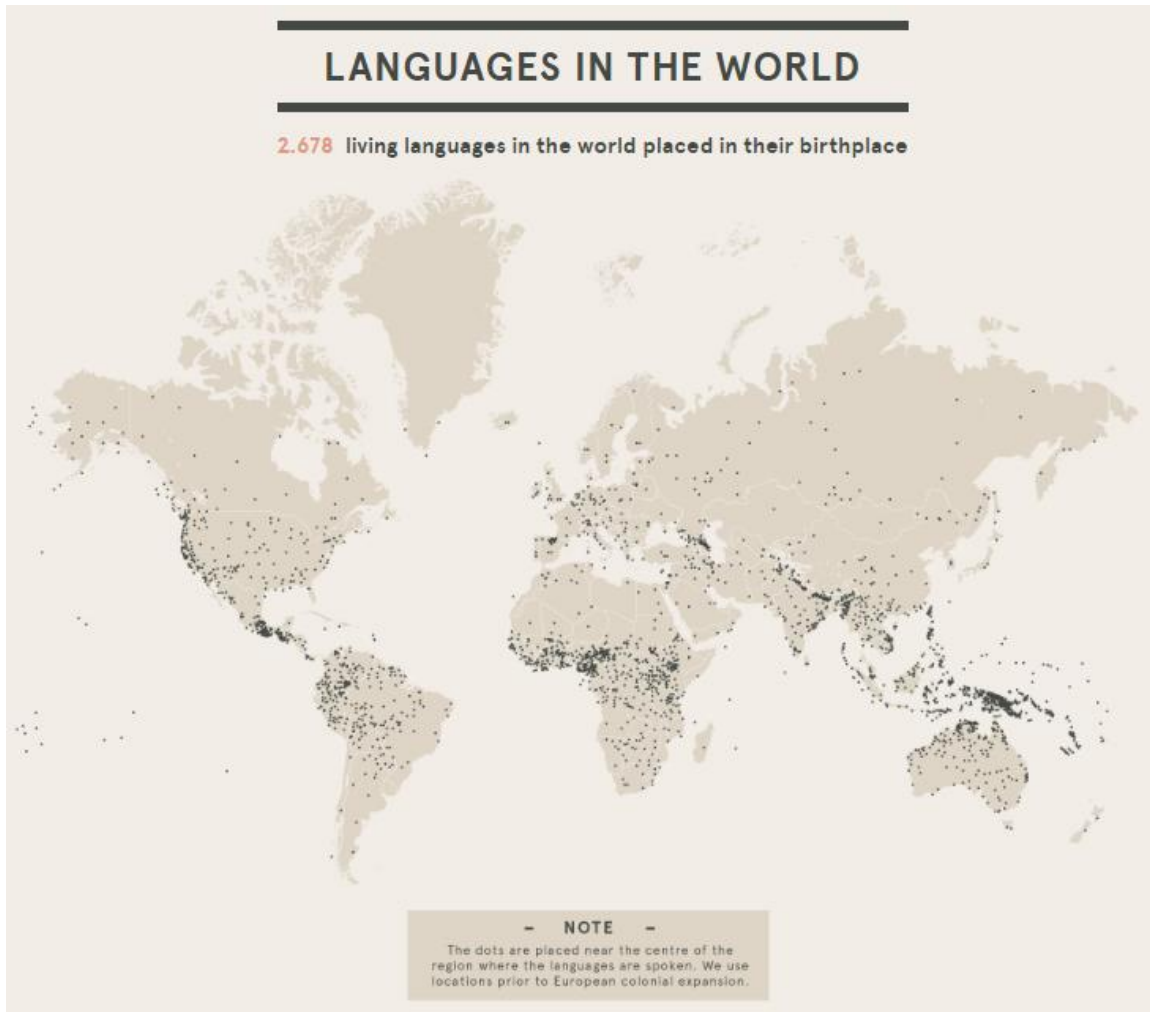


Figure 17 Languages in the World

What's Really Warming the World?

Bloomberg Business created a visualization which tries to tell a story to the readers. Its interactive part used to give life to the story and raise the attention on the readers [25].

What's Really Warming the World?

By Eric Roston and Blacki Migliozi | June 24, 2015

Skeptics of manmade climate change offer various natural causes to explain why the Earth has warmed 1.4 degrees Fahrenheit since 1880. But can these account for the planet's rising temperature? Scroll down to see how much different factors, both natural and industrial, contribute to global warming, based on findings from NASA's Goddard Institute for Space Studies.

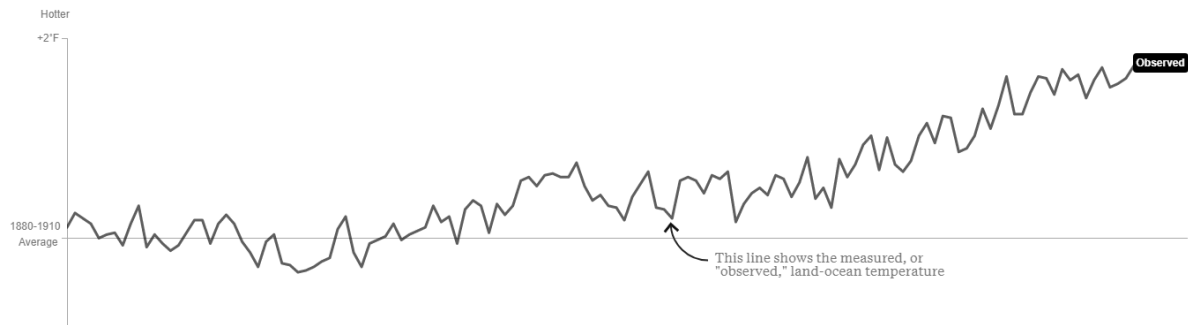


Figure 18 What's Really Warming the World?

The meaning of imaging is to overturn the theories that claim that natural causes cause global warming. The first thing you will see is that the temperature, although it exhibits some fluctuations, has in general an upward trend since 1880 until today. The earth has warmed 1.4 Fahrenheit since 1880. NASA made a research to see how much different factors, both natural and industrial, contribute to global warming. As the user scrolls down, the visualization progressively reveals how much different factors contribute to global warming in comparison to what's been observed. The research highlights a strong conclusion, that also is communicated directly to the user, that greenhouse gases play a major role in climate change.

Why Buses Bunch?

This is an excellent example of a complex dataset which is significantly simplified and feels like a game for the user. In this visualization, it is presented how "bus bunching" happens, i.e. when a bus gets delayed and later causes multiple buses to arrive at a single stop at the same time.

This story it would be extremely difficult to described with numbers but it is easier for the human eye to see how a delay of a bus (user can interactively delay the bus by clicking and holding a button) causes the buses to bunch together after a time [26]

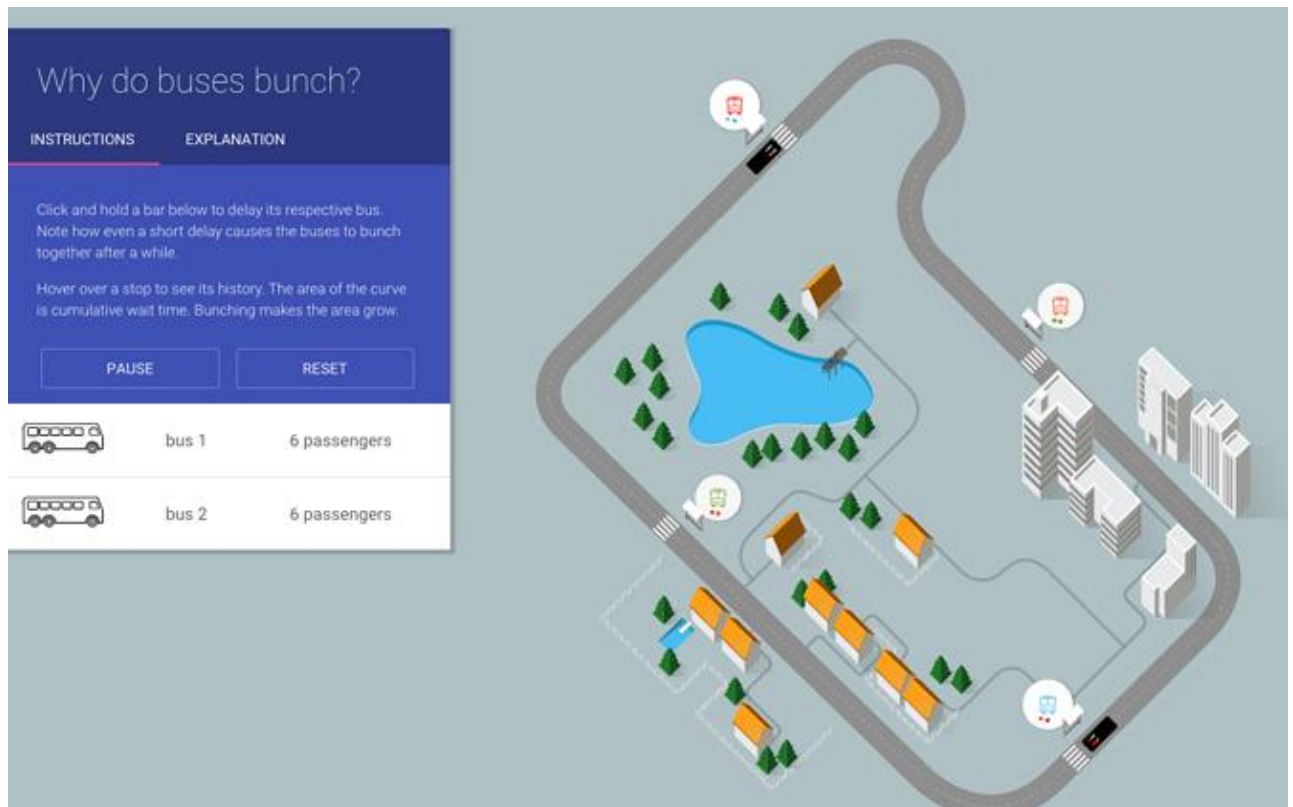


Figure 19 Why Buses Bunch?

How People Like You Spend Their Time?

There is an on-going research conducted by The American Time Use Survey that asks people what they did during the previous day. The user can choose gender, age, if he is employed or not and if he wants to see results for weekday or weekend. The results are presented in a smart way based on the choices the user is asked to make. The grey lines show the hours spent in one person's day. The orange line shows the median hours for each activity.

For household work and leisure, the number of hours spent significantly varies, which is why there are a lot of crossing lines. Noteworthy is the fact that both sleeping and work, have a median value of 8 hours which is rather expected [27].

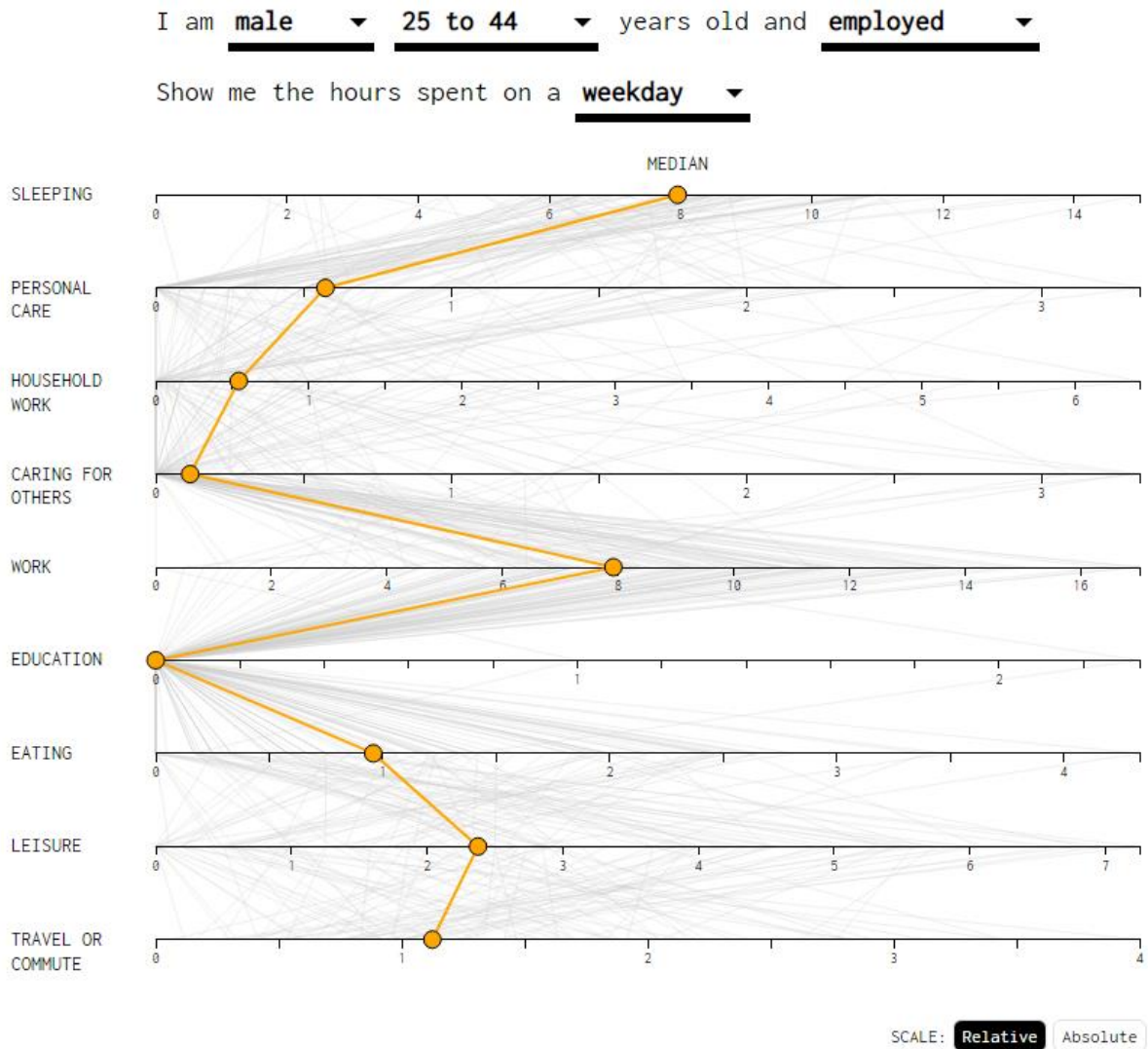


Figure 20 How People Like You Spend Their Time?

2.1.9 Data visualization sites

The Pudding

The Pudding is known as a blog that creates visual essays to enlighten ideas. Complex topics are explored by investigating on wielding prototype datasets, primary research and interactivity. Visual essays are an emerging form of journalism.

The theme of the blog covers various topics such as sports, culture, entertainment, music and social issues. A very nice and fun approach explains issues and analyses situations that have not been analysed so far by conventional forms of journalism.

The journalist-engineers of The Pudding also document their own tools and methods in a growing series of How To articles for everyone to learn from [28].

URL: <https://pudding.cool/>

Information is Beautiful

Information is Beautiful is a blog created by David McCandless to present bespoke projects with a bright and minimalistic way.

The visualizations include static and interactive graphics that analyzing data from news and current events. Most of the visuals, like the site itself, are clean and minimalistic, featuring simple shapes and colors that depict large data sets in digestible chunks. In addition, this blog publishes its data sources [29].

URL: <https://informationisbeautiful.net/>

Flowing Data

Flowing Data was created by the statistician Nathan Yau. It possesses a key role among other data visualization site. FlowingData looks at how statisticians, designers, data scientists, and others use analysis, visualization, and exploration to understand data. It is updated almost daily with interesting graphics and visualizations for a wide range of topics.

Flowing Data also features tutorials and courses on how to make helpful visuals and how to exploit the tools to do so [30].

URL: <https://flowingdata.com/>

Tableau Blog

Tableau blog contains a wide variety of examples, ideas, tips & tricks, community highlights, and stories about the social impact of data. In the visualization gallery there are stunning examples from across the web created with Tableau Public [31].

URL: <https://www.tableau.com/about/blog>

FiveThirtyEight

FiveThirtyEight sometimes rendered as '538', is a data journalism website that makes visualizations, interactive visual models. It is worth noting that it contains interactive political visualizations, which covers things like modeling the US election data

and aggregating presidential approval ratings. The site also deals with economics, sports, cultural, and science / health [32].

URL: <https://fivethirtyeight.com/>

2.2 Data storytelling

2.2.1 Definitions of Storytelling

Data storytelling is a term that is becoming more and more popular in the field of analytics. From 2012, the interest in Data Storytelling has grown approximately 6 times as measured by Google Trends [33].

Data storytelling can be defined as the methodology for communicating information, tailored to a specific audience, with an exciting narrative. It is the last step of the data analysis and undoubtedly the most important aspect.

According to Caine [34], storytelling is the primary way that people access, express, and retain information and knowledge. They claim that brain research confirms the theory that information is naturally organized in our minds according to story form. Richter and Koppett [35] share a similar opinion stating that stories form the foundation of how people communicate and that we remember and integrate new information by placing them into a story format.

Data storytelling combines 3 key fields of knowledge: Data science, Visualizations and Narrative.

Narrative uses language in a format tailored to our special needs, enhancing our full understanding of new information. A narrative is the best way to convey insights, with visualizations and data providing significant evidence.

Stories have been used throughout history to draw attention, convey ideas, trigger the imagination. Data can be convincing, but stories are much more. Stories have strong impact to the audience that is not feasible to achieve with data alone. They are memorable, they are often passed on and repeated from person to person, they encourage personal involvement, and they stimulate people to take action.

As we have seen above, simply collecting the data, and presenting it in dashboards and spreadsheets only answers questions about **what** is happening, but it does not give the reasons for **why** this is happening.

Tables, dashboards, spreadsheets and BI tools in general, have a number of limitations:

- There is need for human intervention which slows the data analysis process and communications within organisations.
- These tools only show data as numbers and charts. However, they lack the narrative which should be used for effectively communicating information and insights.

2.2.2 The Biochemistry of Story

There is a bio-chemical reason that people like stories [36]. Storytelling has actually a chemical effect on human brain that activates it to perceive, digest and store information. By hearing a good story, people think themselves as the central character in that story. The scientific term for this is “neural coupling”. There is also an emotional bond build with the storyteller that scientists call mirroring. Compared to the raw data, stories are much more interesting and wake up several parts of the brain, which is mentioned to as cortex activity. Finally, all this activity leads to the release of dopamine, which in turn causes euphoria.

2.2.3 History of data storytelling

Storytelling and its role(s) in society go before written human history as verbal storytelling was the only tool existed to retain and share cultural customs, beliefs and heritage. Story verified to be the oldest and the most substantiated means for people to recall and story information [37]. Therefore, storytelling proved to be a very effective way of coding knowledge in oral cultures because it made them more memorable and easily passed on to others. Actually, Abrahamsen claims that storytelling helped the civilizations to survive because they ensured the continuity of life experiences and practices to following generations.

The first record of storytelling is placed back in 4000 B.C. as stated by Ruth Sawyer in her book with title “Tales of the Magicians” (as cited in Abrahamsen, 1998). Storytellers were highly valued in early Egyptian and medieval times and were also used by nomadic tribes such as the Gypsies. Stories are evident in the Roman culture as well. Even more,

storytellers in ancient times were regularly highly esteemed among the community as they facilitated people to understand complex cultural or religious issues, but also provided simple entertainment (Parkin, 2004; "What use is Storytelling", 2005). As expressed by Abrahamson, "it appears through a review of literature that the story is one of the most important inventions of humankind" (para. 3)

2.2.4 Trends in data storytelling

Even though storytelling is considered a very old art and discipline, continues nowadays as a key means of communication and entertainment. Consequently, storytelling continues to have a strong presence in today's society (Abrahamson, 1998). It is found within cultures, religions, families, businesses and in educational institutions. In current day civilization, stories are existing in many forms; stories reach out from tv, radio, email, magazines, newspapers and books to refer a few of them.

Stories are usually the way in which people communicate to each other to recap the day's events to friends or family, as advice is offered to a colleague, or as children are taught meaningful lessons for their own lives and interactions with society. Moreover, educational researchers have strong signs of the renaissance of storytelling as a teaching pedagogy and learning tool in modern educational organizations and business organizations (Parkin, 2004; "What use is Storytelling", 2005). There is as well a movement toward oral storytelling in the entertainment industry as the popularity and number of storytelling increases on both national and international levels.

2.2.5 Best practices of data storytelling

Above we have examined the importance of data storytelling, but in order to achieve its goals, the storyteller should follow a few rules. Below are presented the best practices to be used in storytelling [38].

- **Focus on the story.** The target of storytelling is to draw listeners' attention with a narrative that is reinforced by visualizations. Stories should have a beginning, middle, and end. Even though the details will vary, a usual story flow is to define

the current situation, provide estimates should it continue, describe the options, make a recommendation, and explain the expected result from implementing the proposal.

- **Engage the audience.** Audience will ask themselves what the story means to them. Storyteller should put himself in their place when making the story. Also, he should include specific, personalized examples that people will easily recall. For instance, politicians frequently name and describe an individual to prove a point.
- **Explanation of the story.** Storyteller should analyse the situation in a manner that is meaningful to the listener and also to support this description with data and let listeners know its source, accuracy, and relevance. It is good to give the impression that storyteller is confident and understands the business and issues.
- **Define the problem or conflict.** Visualizations are critical because they can identify a problem or issue or prove a point. Storyteller should be strategic with the visualizations and keep them simple, organised, and tasteful. It is essential to highlight the most important parts. The more interesting the visualization, the more time and attention the audience will give it.
- **Solution suggestions and future perspectives.** Presentations and storytelling should do more than identify a problem. They should suggest a resolution or at least lead to a conversation. Data can be used to make predictions, outline alternatives, identify a solution, create a call to action, and show likely results.

2.2.6 Different types of story structures

It is commonly known that good stories have a predictable arc. They typically start out with an introduction, followed by a rising action, a climax, and finish with a final resolution (Figure 21). This is the way storytelling has been done for years and works well in a variety of formats such as oral and written stories in TVs and movies. The explosion of new mediums, namely computers and the Internet, have given rise to new narrative structures that can often be much more nuanced and complex [39].

The most usual type of narratives are the Author-driven narratives, Viewer-driven narratives and a mix of them called the martini glass narratives. These narratives are described in the following paragraphs.

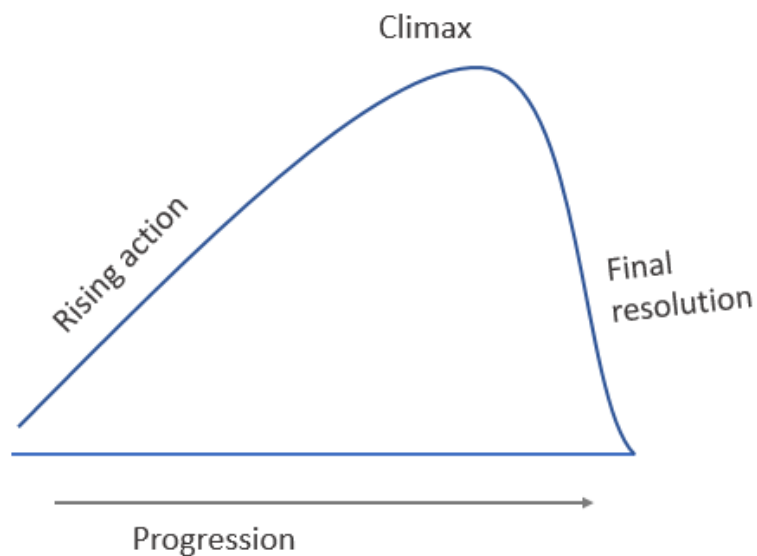


Figure 21 Progression of a story

Author-driven narratives typically have a well-defined start and endpoint and proceed in a linear flow that's directed by the author of the visualization.



Figure 22 Author-driven narratives

Viewer-driven narratives on the other hand, while still having a well-defined start point, allow a freedom of choice for the viewer to choose the direction in which the narrative progresses. In this way while each user starts in the same place there are various endings that they could end up depending on how they look at the data or the choices they made along the way. This type of narrative is very powerful because it allows the user to choose their own narrative.

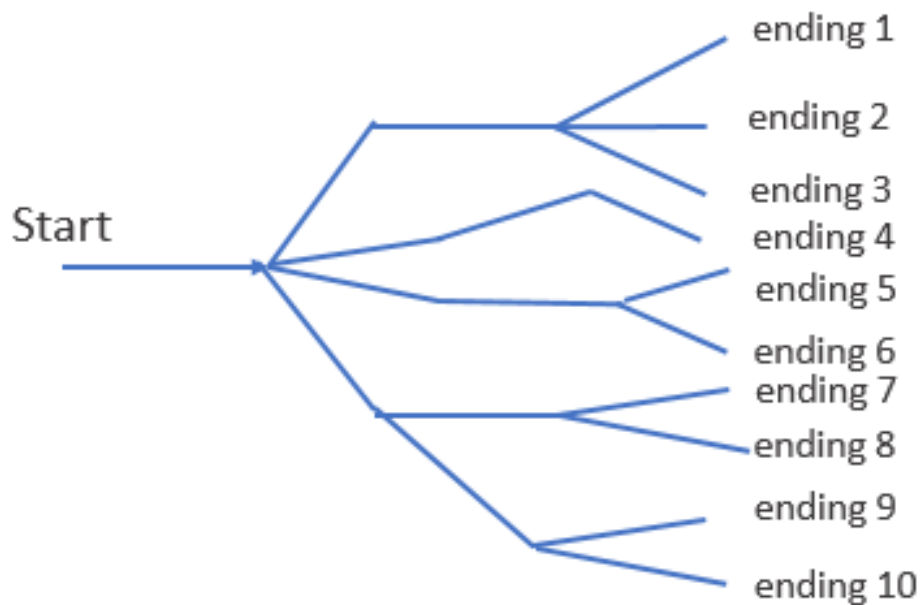


Figure 23 Viewer-driven narratives

It is like the reader choose his own adventure book.

We often see stories fall within a spectrum, between author-driven and viewer-driven. Visualizations that have few exploratory elements, are often called author-driven. And are useful for narratives that have strong ordering, heavy messaging, or have a need for clarity and speed in the narrative.

Viewer-driven narratives on the other hand, take a very different approach. Rather than dictating a specific narrative, it allows the viewer to freely interact with the data. Allowing you to ask questions, explore, and tell their own data story.

There's a third narrative structure, common in more complex data visualizations called **the martini glass**. It's called the martini glass because of its unique shape. Think of a martini glass set it on its side. It has a base, a long narrow stem, with a wide mouth. Viewers start at the base and work through the neck of the martini glass.

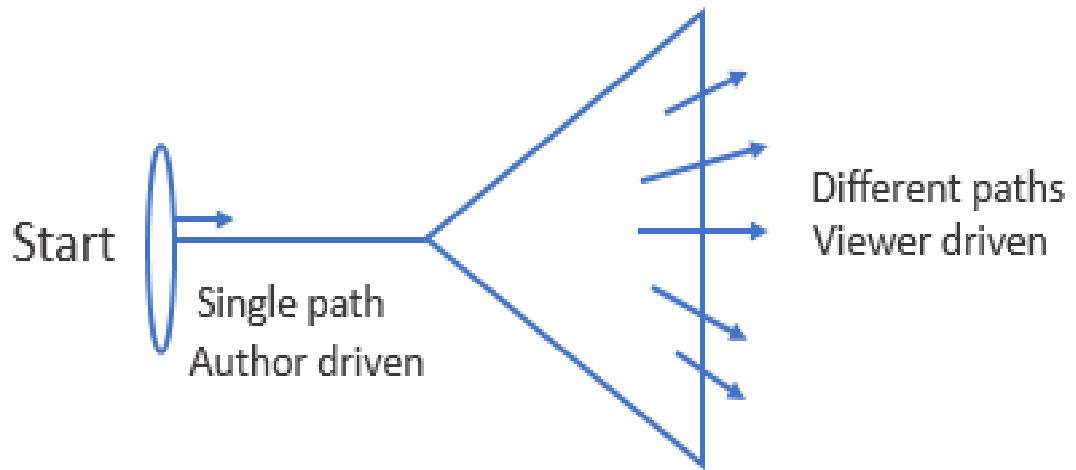


Figure 24 The martini glass narrative structure

Typically, through a single path that is author-driven. When they reach the mouth, the visualization opens up and allows the viewer to freely explore different paths within the data. And to tell their own narrative. This way, it combines with the author-driven and viewer-driven narrative structures in a single cohesive whole.

3 Plotly Chart studio demonstration

The company that develops Plotly (also known as Plot.ly) JavaScript language is also named Plotly develops online data analytics and visualization tools [21]. This library uses D3.js behind the scenes. One of the main products of this company is the **Chart Studio**. Chart Studio is a graphical user interface (GUI) for importing and analyzing data into a grid and utilizing various statistical tools. In other words, it is a sophisticated editor for creating D3.js and WebGL charts faster and more efficient without coding experience to be required. There is a “free” pricing plan for the Chart Studio named Community but with this plan the user can not private save his charts and dashboards, instead the charts remain public in the Chart studio gallery, in order to inspire other users. Annual pricing starts from 99 USD if the user wants to enable private saving of his charts.

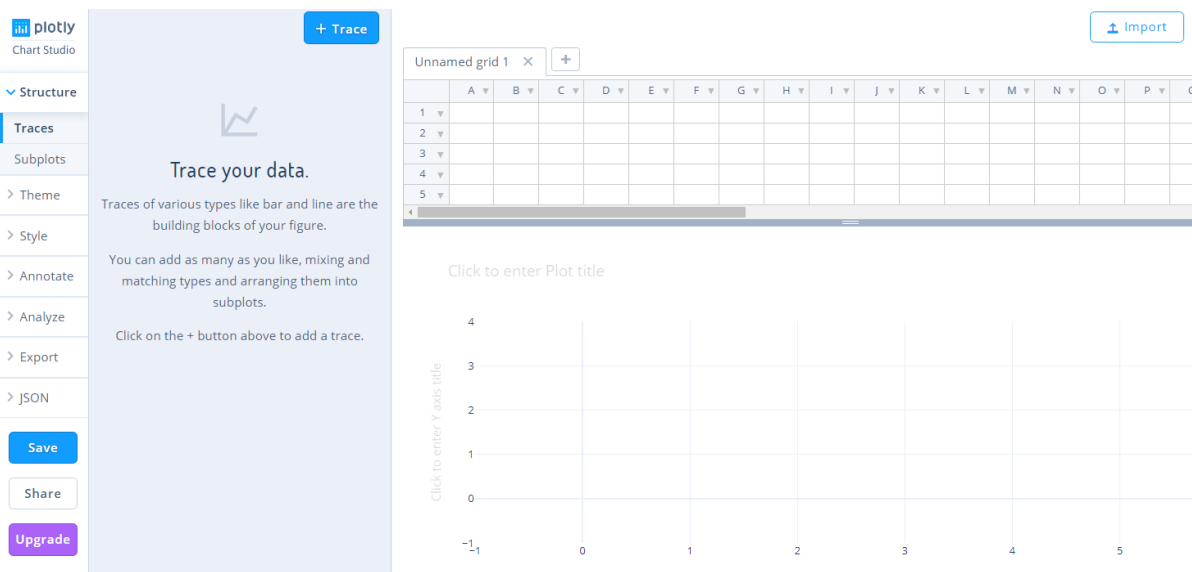


Figure 25 Plotly Chart studio GUI

Below, there is a description of how Chart studio works and what it can offer to the user who will use it.

3.1 Import Data

Importing data to the grid is the first step in creating a Chart Studio graph. There are several options for the user of how this could be done.

- Type the data directly into the grid, or copy it from a spreadsheet and paste it in.
- Import a Data File (.xls, .xlsx, or .csv). Drag and drop can be used.
- Import data by entering the URL of an online dataset.
- Connect to an SQL database

3.2 Interactivity tools



There are several interactivity tools that allows user to bring graphing and data analysis to the next level with zoom, pan, and hover. These tools are at the right upper side of the graph.

The 4 first icons are for the drag mode that is selected.

If the plot's drag mode is set to 'Zoom', the user can click and drag on the plot to zoom-in and double-click to zoom-out completely, i.e., autoscale both the axes.

If the plot's drag mode is set to 'Pan', click and drag on the plot to pan and double-click to reset the pan.

Box select and Lasso select icons are used to filter the data by simply selecting an area of the chart. If there are other charts that based on the same data, they will be filtered as well, so the user will see only the data corresponds to the previous selection.

Hover Options: Only one of the following two buttons is selected each time. The first button 'Show closest data on hover' will display the data for just the one point under the cursor while the button 'Compare data on hover' will show you the data for all points with the same x-value.



In addition, if the user enables the following button, when the cursor is over data point, spike lines will be appear showing the intersection with the axes.



The chart studio offers a wide range of graphs to the user. These are categorized according to their characteristics in **Simple, Distributions, 3D, Maps, Finance and Specilized** charts.

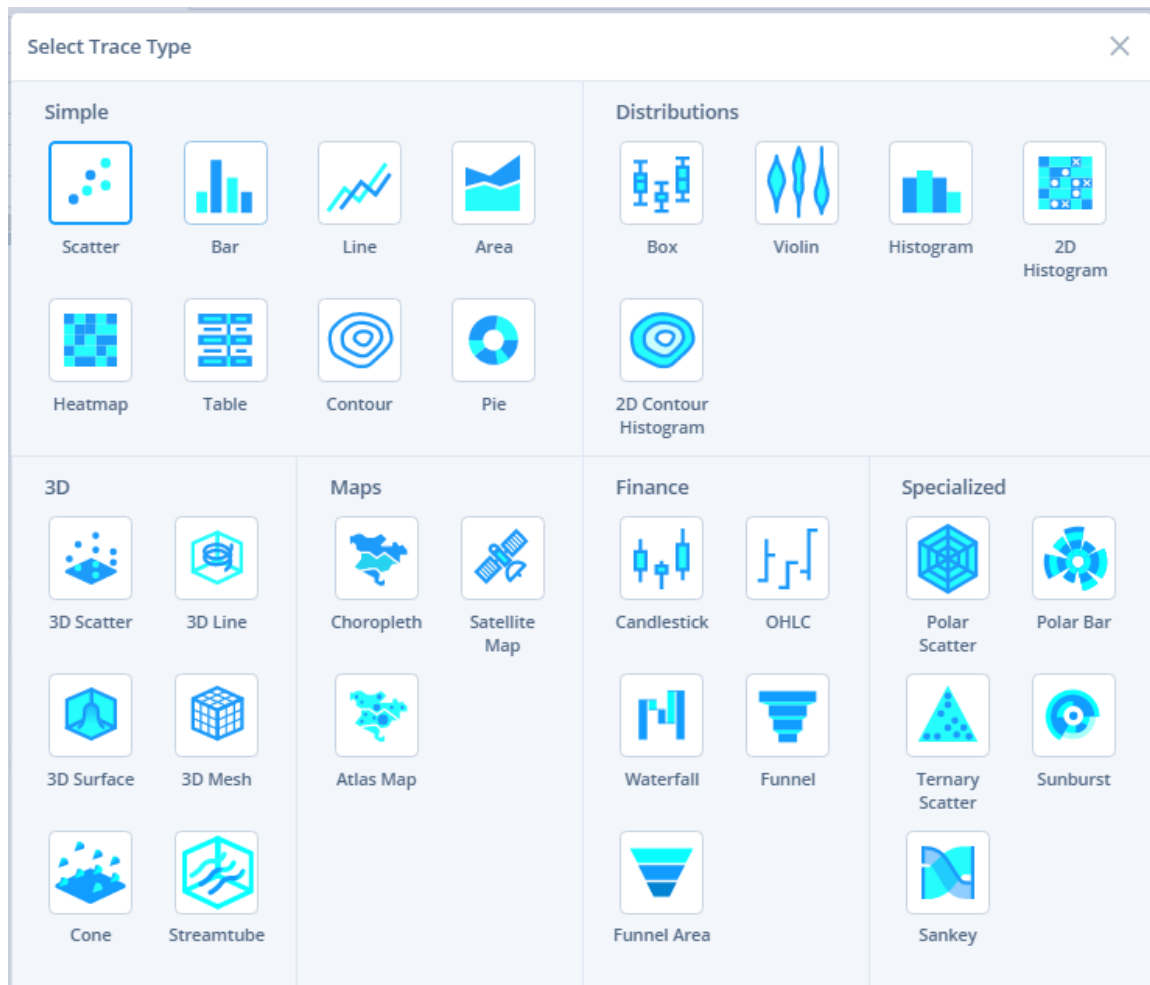


Figure 26 Chart studio chart types

Below, some of the above diagrams are described to illustrate how Chart studio works.

3.3 Scatter plots

Scatter plots allow the comparison of two variables for a set of data. Here we have an example scatterplot where one variable is on the y axis and another variable is on the x axis and where those two variables end up meeting there is a little point. The data for this chart came from random numbers. Depending on the trend that the scatter points may have, we can actually interpret some sort of correlation. In our particular case here for the following image, it doesn't seem like these two variables have any sort of correlation (as we said the data are random numbers) but if the points happen to go up on the y axis as the value of x got larger, then there will be a kind of correlation. A positive correlation would take place if as one variable goes up the other variable also tends to increase.

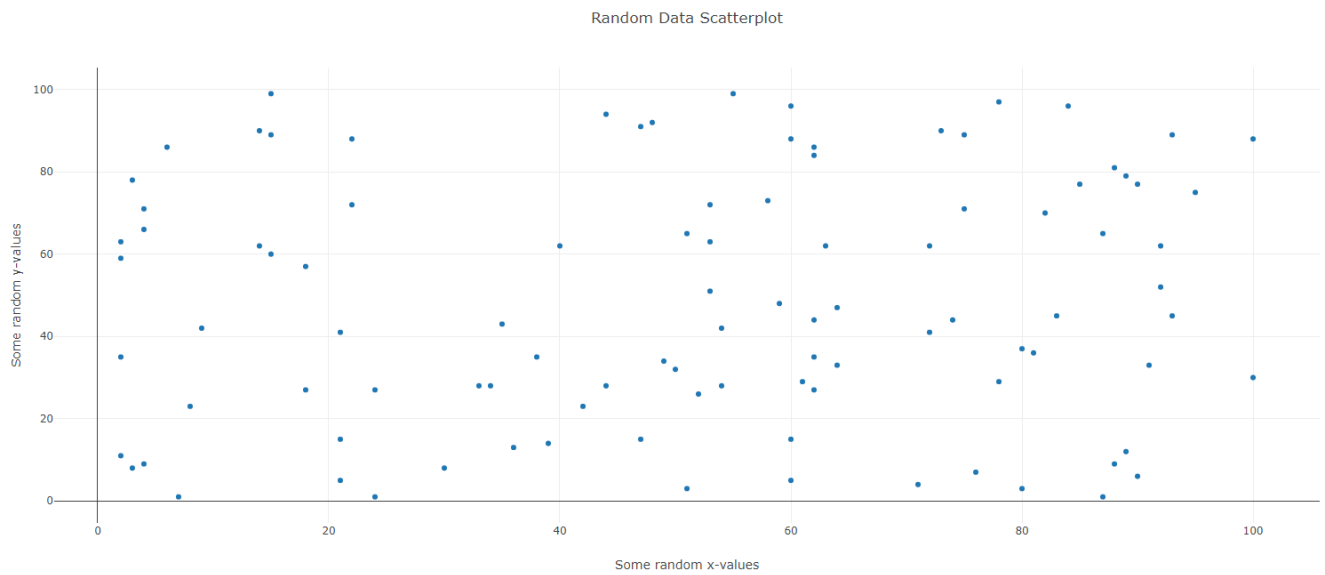


Figure 27 A scatterplot of random data with Plotly

```

6 import plotly.offline as pyo
7 import plotly.graph_objs as go
8 import numpy as np
9
10 np.random.seed(42)
11 random_x = np.random.randint(1,101,100)
12 random_y = np.random.randint(1,101,100)
13
14 data = [go.Scatter(
15     x = random_x,
16     y = random_y,
17     mode = 'markers',
18 )]
19 layout = go.Layout(
20     title = 'Random Data Scatterplot', |
21     xaxis = dict(title = 'Some random x-values'),
22     yaxis = dict(title = 'Some random y-values'),
23     hovermode = 'closest'
24 )
25 fig = go.Figure(data=data, layout=layout)
26 pyo.plot(fig, filename='scatter2.html')
27

```

Figure 28 Code for the scatterplot with random data

In Chart studio, the user has to add a trace by clicking the blue button (Figure 29) and select the “Type” to be “Scatter”. Also, the user has to define the data that will be used by each variable (e.g. x variable data will come from column A of the grid and y variable data will come from column B of the grid)

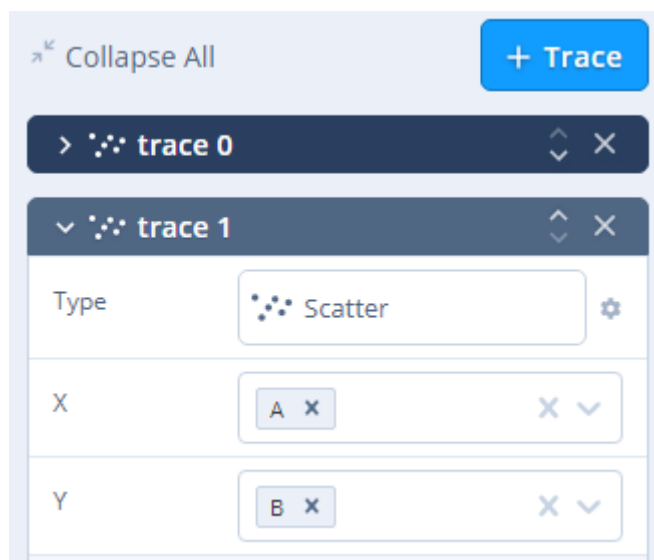


Figure 29 Add a trace in Plotly

3.4 Bubble charts

Bubble charts are actually very similar to scatterplot except there is a 3rd variable's information through the size of the actual markers. In the typical scatterplot as we have already seen, all the markers have the same size. In order to create a bubble chart, the user should change the size of the markers from the Style menu (The side menu will be further analyzed in a separate section). There are two modes: 'Constant' or 'Variable'. To set a varying size for the markers based on a dimension of data, the user should click on 'Variable' and select the desired column name from the dropdown menu.

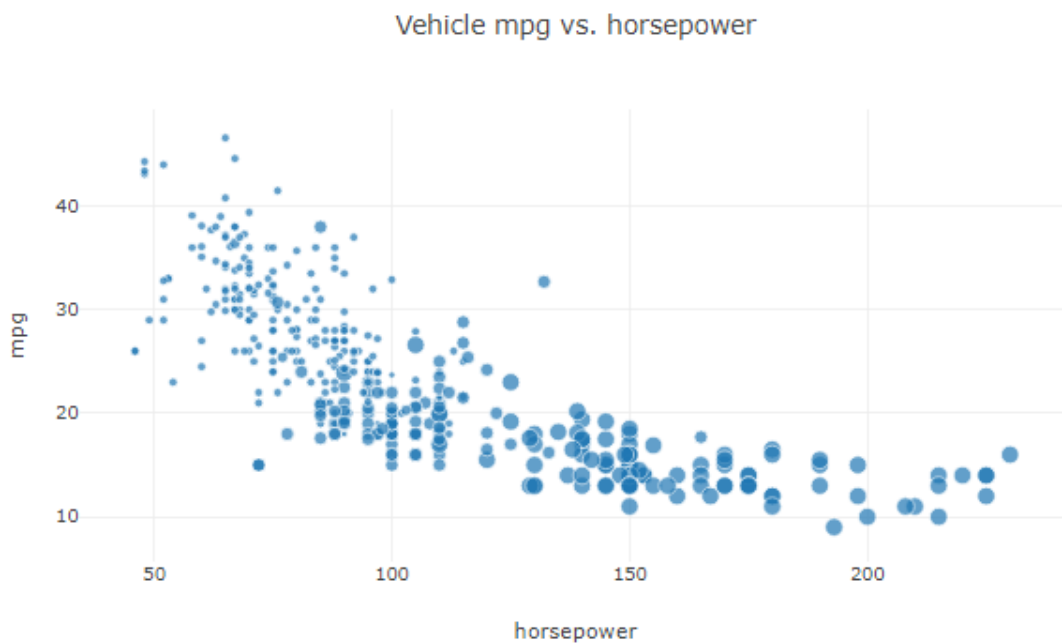


Figure 30 An example of a bubble chart with Plotly

```
17 ▼ data = [go.Scatter(  
18     x=df['horsepower'],  
19     y=df['mpg'],  
20     text=df['text2'],  
21     mode='markers',  
22     marker=dict(size=1.5*df['cylinders'])  
23 )]  
24 ▼ layout = go.Layout(  
25     title='Vehicle mpg vs. horsepower',  
26     hovermode='closest'  
27 )  
28 fig = go.Figure(data=data, layout=layout)  
29 pyo.plot(fig, filename='bubble2.html')
```

Figure 31 Code for the bubble chart

On top of these, user can also continue to add a variable information by coloring points based off a category. In the following example, the previous bubble chart contain another one variable information (the GDP per capita)

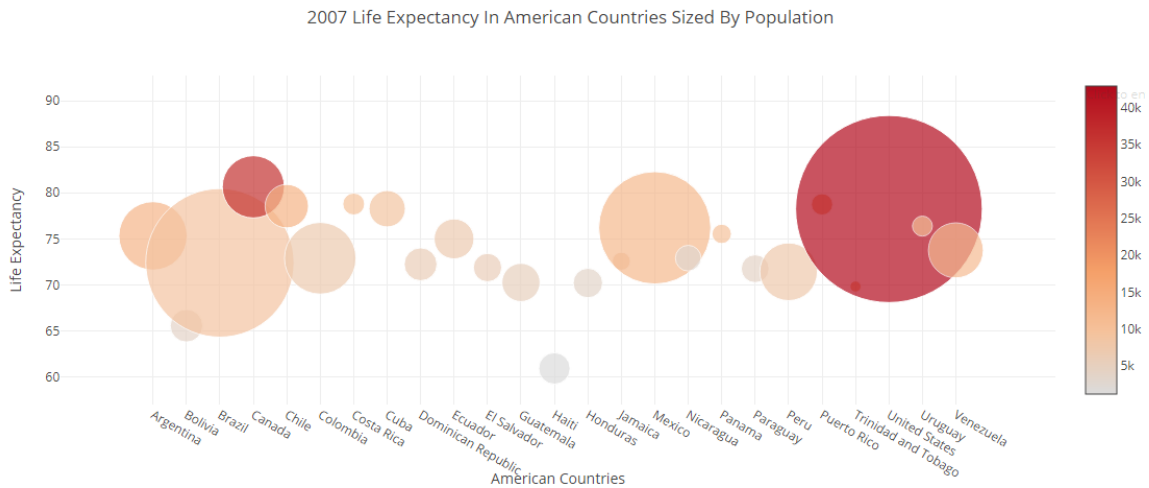


Figure 32 The previous bubble chart with a colorscale

3.5 Line chart

Line charts display a series of data points called markers connected by line segments. So, this actually means that a line chart is essentially the same thing as a scatterplot except this time there are line segments that connect each of the points. One of the main differences however between a typical line chart versus just taking a scatterplot and connecting all those points together is that the x value on the x axis is going to be sorted or ordered in some way. Therefore, it is really useful to use a line chart to visualize a trend over an interval of time which is known as a time series. Below there is an example of a simple line chart showing the flow of the river Nile across the years. Actually, there are some sorted points that have all been connected with lines.

Again, the process to create a line chart in Chart studio does not differ a lot from the scatterplot. In the above Figure 29, instead of selecting “Scatter” in the “Type” field, the user has to select “Line”.

Line chart showing three different modes

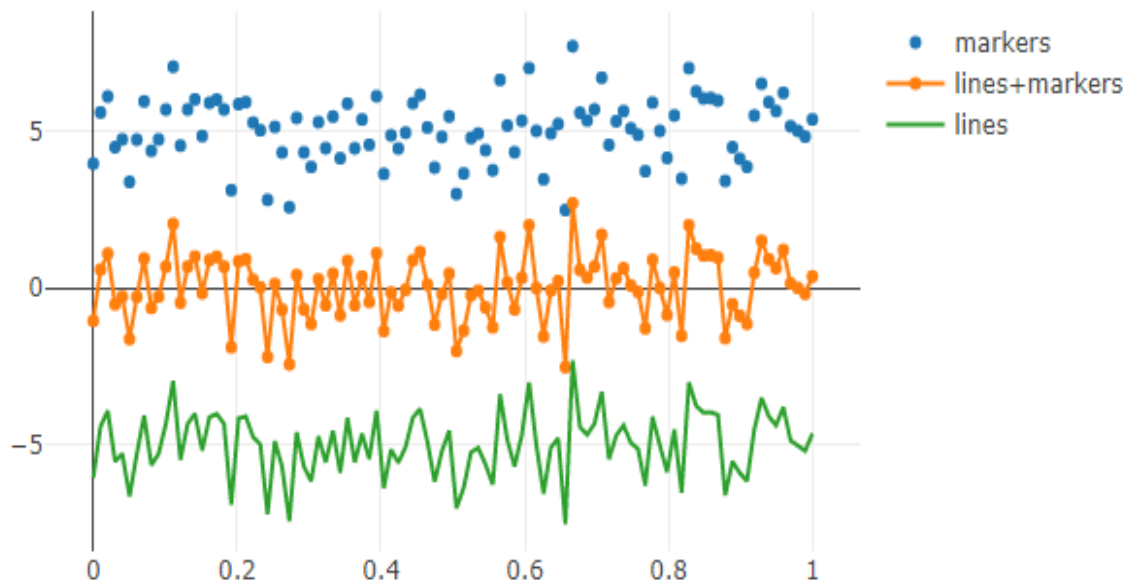


Figure 33 Line chart showing three different modes

```
9 np.random.seed(56)
10 x_values = np.linspace(0, 1, 100)
11 y_values = np.random.randn(100) # 100 random values
12
13 # create traces
14 trace0 = go.Scatter(
15     x = x_values,
16     y = y_values+5,
17     mode = 'markers',
18     name = 'markers'
19 )
20 trace1 = go.Scatter(
21     x = x_values,
22     y = y_values,
23     mode = 'lines+markers',
24     name = 'lines+markers'
25 )
26 trace2 = go.Scatter(
27     x = x_values,
28     y = y_values-5,
29     mode = 'lines',
30     name = 'lines'
31 )
32 data = [trace0, trace1, trace2] |
33 layout = go.Layout(
34     title = 'Line chart showing three different modes'
35 )
36 fig = go.Figure(data=data,layout=layout)
37 pyo.plot(fig, filename='line1.html')
```

Figure 34 Code for the line chart

3.6 Bar chart

A bar chart presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent. At this point we have to explain what is categorical data and what is continuous data.

In general variables and data either represent measurements on some continuous scale or they represent information about some categorical or discrete characteristics. For example, the weights, heights and age of respondents in a survey would represent **continuous** variables because each of those features is on a continuous scale. However, things like a person's gender, occupation or marital status are going to be **categorical or discrete** variables because they're categories and they're not really on a continuous scale.

By using bar charts, categorical data can be visualized. Typically, the x axis is the categories and the y axis represents count that is the number of occurrences in each category.

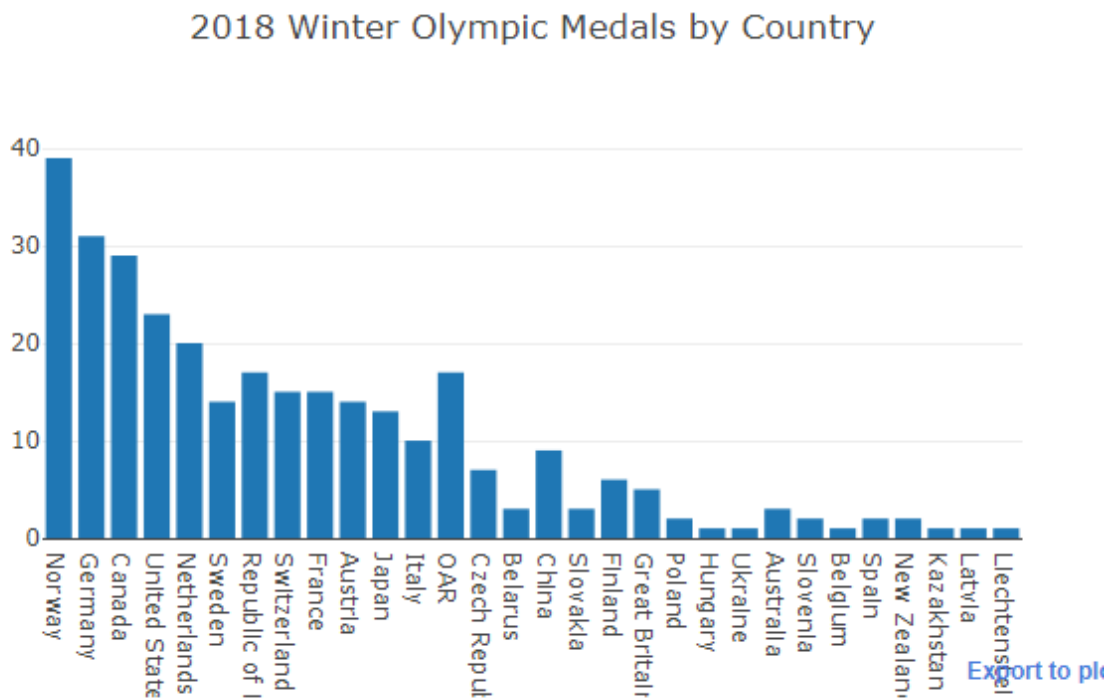


Figure 35 An example of bar chart with Plotly

```

9 df = pd.read_csv('../data/2018WinterOlympics.csv')
10
11 data = [go.Bar(
12     x=df['NOC'], |
13     y=df['Total']
14 )]
15 layout = go.Layout(
16     title='2018 Winter Olympic Medals by Country'
17 )
18 fig = go.Figure(data=data, layout=layout)
19 pyo.plot(fig, filename='bar1.html')

```

Figure 36 Code for the bar chart

In the above Figure 35 there is a basic bar chart of number of animals in a specific area. Notice that there's actually spaces in between these bars. The reason they're spaces in between them is because each of these bars is going to represent a different category and in this case each category is a different animal.

3.7 Stacked bar chart

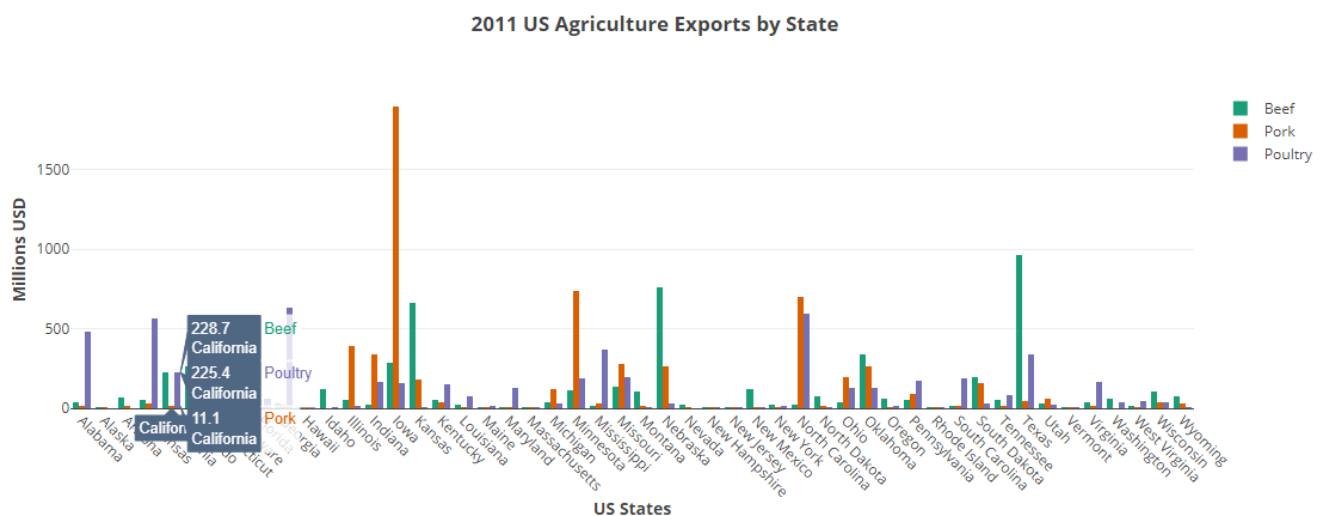


Figure 37 An example of stacked bar chart with Plotly

The above chart was created by adding more than one traces in a single bar chart. Chart studio provides the option to simply switch to a stacked bar chart. To make this bar chart a stacked bar chart, user has to go to the property 'Bar mode' under 'Bar Size and Spacing' and select 'Strict Sum Stacked' as the bar mode from the dropdown list.

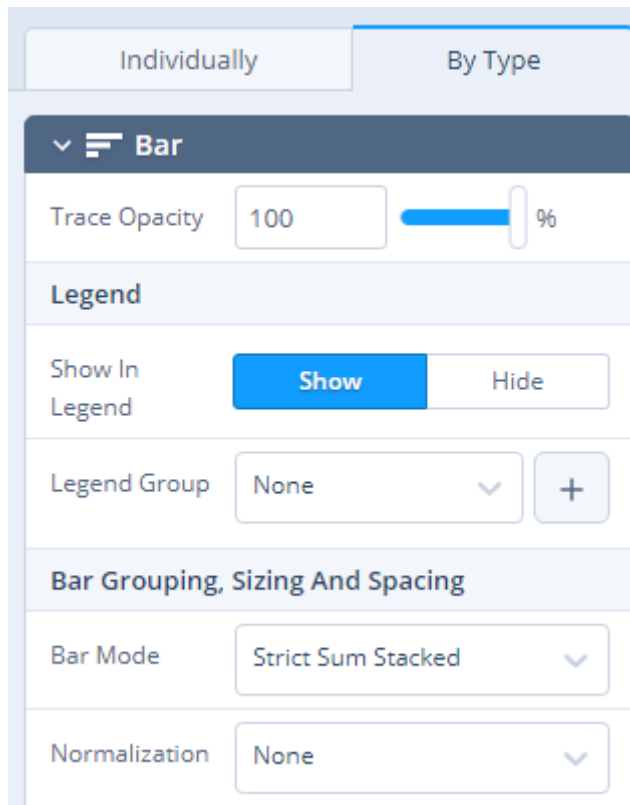


Figure 38 Option to switch from a single bar chart to a stacked bar chart

Box plots

Box plots visualize the variation of a feature by depicting the continuous numerical data through core tiles. Then the data can be separated based on a categorical feature to compare the continuous feature based off the category. Boxplots are a standardized way of displaying the distribution of data based on a five-number summary (“minimum”, first quartile (Q1), median, third quartile (Q3), and “maximum”). **Median** is the middle value of the data, **first quartile (Q1)** is the middle number between the smallest number (not to be confused with “minimum”) and the median of the dataset, **third quartile (Q3)** is the middle value between the median and the highest value (not to be confused with “maximum”) of the dataset., **interquartile range (IQR)** is 25th to the 75th percentile., **“maximum”**: $Q3 + 1.5 \cdot IQR$, **“minimum”**: $Q1 - 1.5 \cdot IQR$. A small IQR indicates that most values fall near the median value.

It is easy to create a boxplot in Plotly’s Chart studio. The user has to add the distribution data in the grid and to create a trace base on this data by simply selecting ‘Boxplot’ in the Type menu.

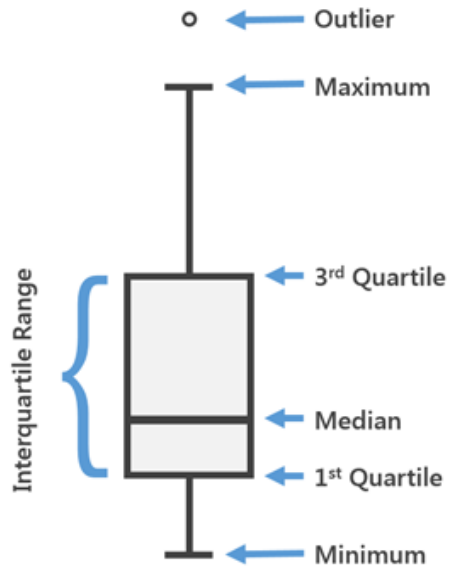


Figure 39 The anatomy of a boxplot

Histogram

Histogram plots are used to better understand how frequently or infrequently certain values occur in a given set of data. In order to construct a histogram, the user should divide that entire range of values of a continuous feature into a series of intervals and these series of intervals are known as bins. Each bin contains the number of occurrences. Notice that unlike a bar chart, there are no "gaps" between the bars. This is because a histogram represents a continuous dataset.

The same procedure with the above charts can be followed to create a histogram in the Chart studio. It is very easy to modify the properties of the traces such as histogram normalization, binning, and bar size and spacing.

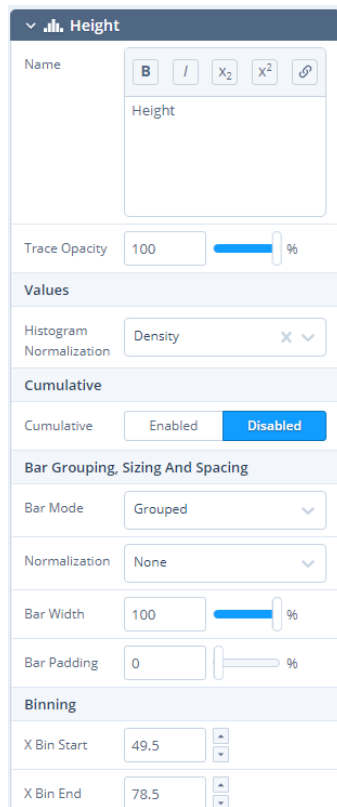


Figure 40 Set the properties of a histogram with Plotly

3.8 Heatmaps

A graphical representation of data where the individual values contained in a matrix are represented as colours.

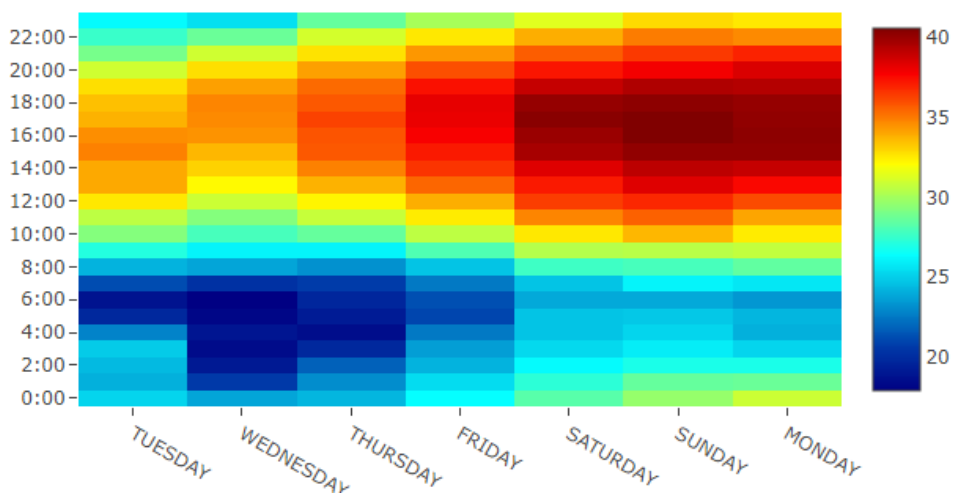


Figure 41 An example of heatmap chart with Plotly

In order to create a heatmap in Chart studio, user has to select 'X' and 'Y', and fill out the 'Z' values from their respective dropdown menus.

```
7
8 df = pd.read_csv('../data/2010Thessaloniki.csv')
9
10 ▼ data = [go.Heatmap(
11     x=df['DAY'],
12     y=df['LST_TIME'],
13     z=df['T_HR_AVG'],
14     colorscale='Jet'
15 )]
16
17 ▼ layout = go.Layout(
18     title='Hourly Temperatures, June 1-7, 2010 in<br>\
19     Thessaloniki'
20 )
21 fig = go.Figure(data=data, layout=layout)
22 pyo.plot(fig, filename='Thessaloniki.html')
23
```

Figure 42 Code for the heatmap

3.9 Style the plots

Once the user created the plots, there are a lot of features in the Chart studio to make the plot look just the way he wants.

3.9.1.1 Themes

There are 6 built-in themes available for the user to select in the free version. The user can only create its theme only in the Pro version, which is also contain a lot of other themes in the Pro version.

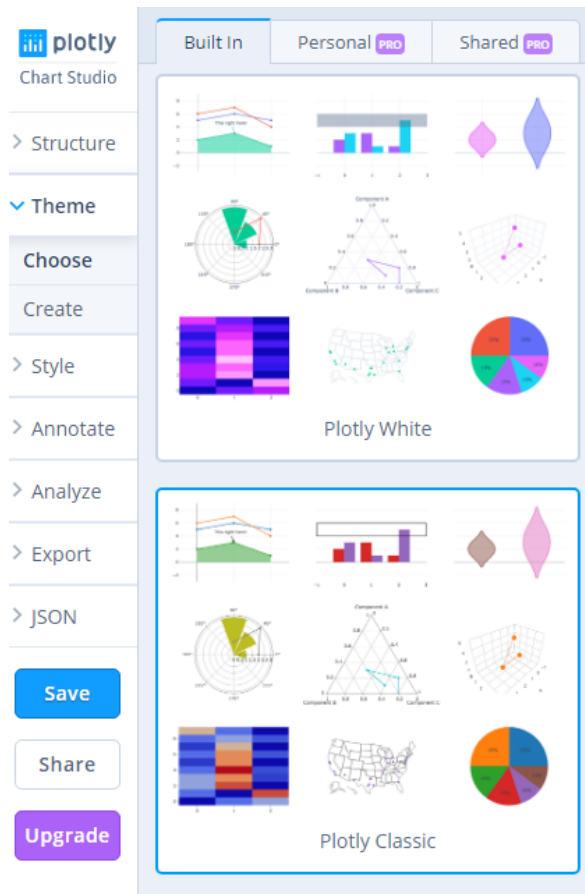


Figure 43 Available themes

3.9.1.2 Style

Under the Style tab, there are some options for the user which depend on the chart type that is created as each chart type is unique.

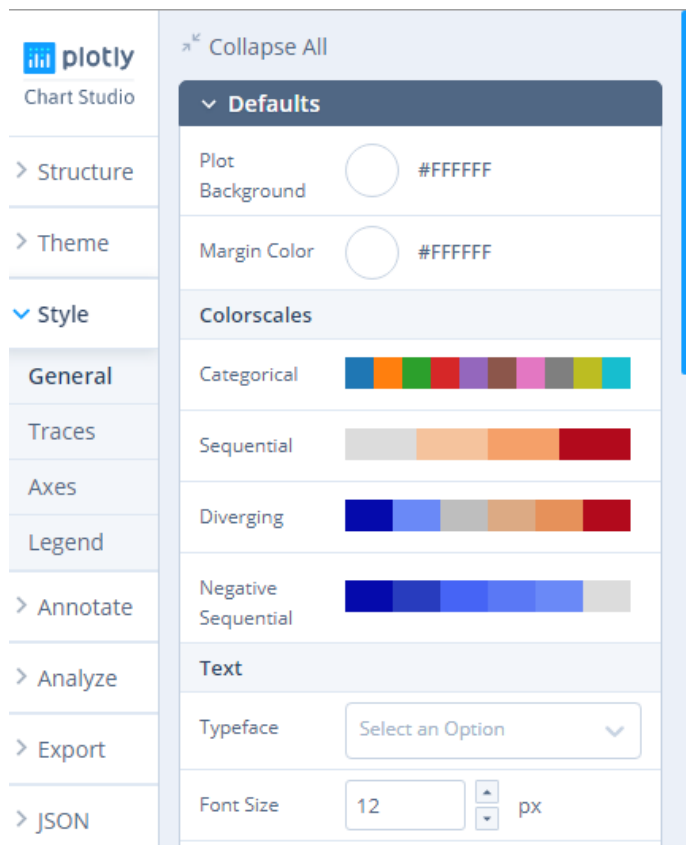


Figure 44 Style tab

The 'General tab' is a section where the user can edit general attributes such as the background colour of the plot and other characteristic of the charts such as margins and sizes etc.

In the 'Traces' tab, the user can modify style attributes of the chart's values or data. For simple charts such as line and scatter, user can alter the color, thickness and type of the lines, play with the color, diameter and symbol of the points.

In case of bar charts or histograms, here the user can alter the size and spacing between each bar, in addition to the bin size (the width of each bin on the number line) of the histogram.

Also the user can edit here the color scale that some chart types contains (e.g. scatter plots, heatmaps, choropleth maps, and 3D charts). The heatmap has a 'smoothing' choice, which will create a continuous heatmap instead of color blocks.

The 'Axes' section has options that relate to the values and labels of the axes on the plot. Here the user can modify the range of the axes by setting minimum and maximum values,

set the label format of the axis, change the status of showing tick marks, and show the spike lines when hovering on data points.

The 'Legend' section has option related to the text, font color and font size of the labels of the legend. User can also make a border around your legend, fine-tune the border width, play around with the color of its boundaries and its background. The positioning of the legend can be changed as well, along with the orientation. Finally, there are two options for the trace order in the legend; showing your traces as is (automatic), or reversing the order.

3.9.1.3 Annotate

As we have already seen, Chart Studio has many features to improve the graphs, including addition of a logo, text and images to the plot.

The section 'Annotate' contains the sub-tabs 'Text', 'Shapes' and 'Images'. The user should select blue button "+Annotation" and simply put the text in the "Note text" section.

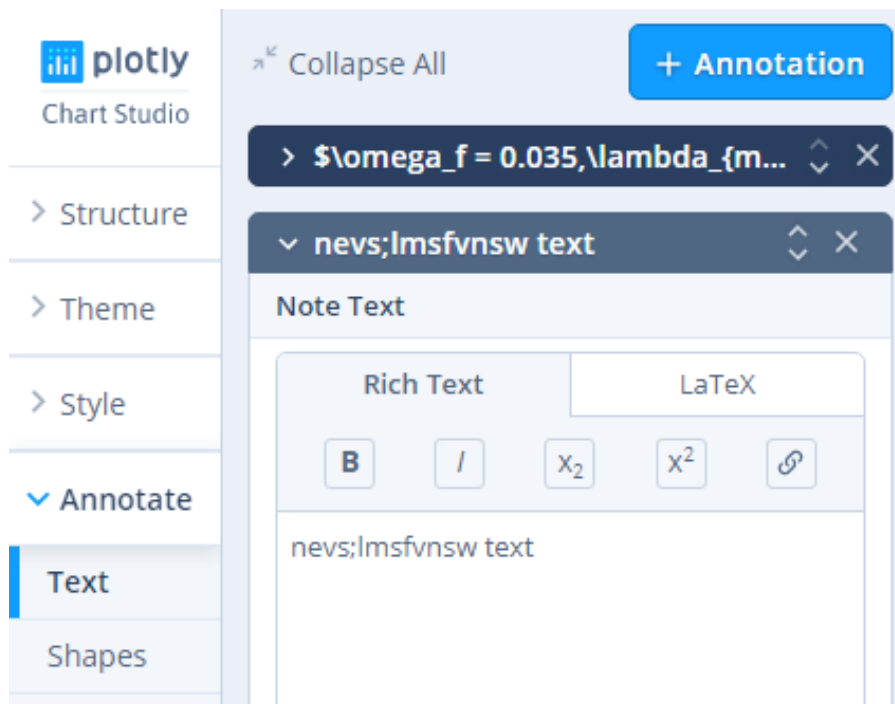


Figure 45 Annotation menu

Similarly, user can add a line, a rectangle or an ellipse, and to upload an image (e.g. a logo, or watermark).

3.9.1.4 Other features

There are also some statistical tools in the Chart studio, such as the ‘Curve fitting’. A trend line or linear regression, is a straight line on a graph that best represents the data on a plot. This line passes through some of the points, all of the points, or none of the points. It can be used to make forecasts or to reveal trends in data.

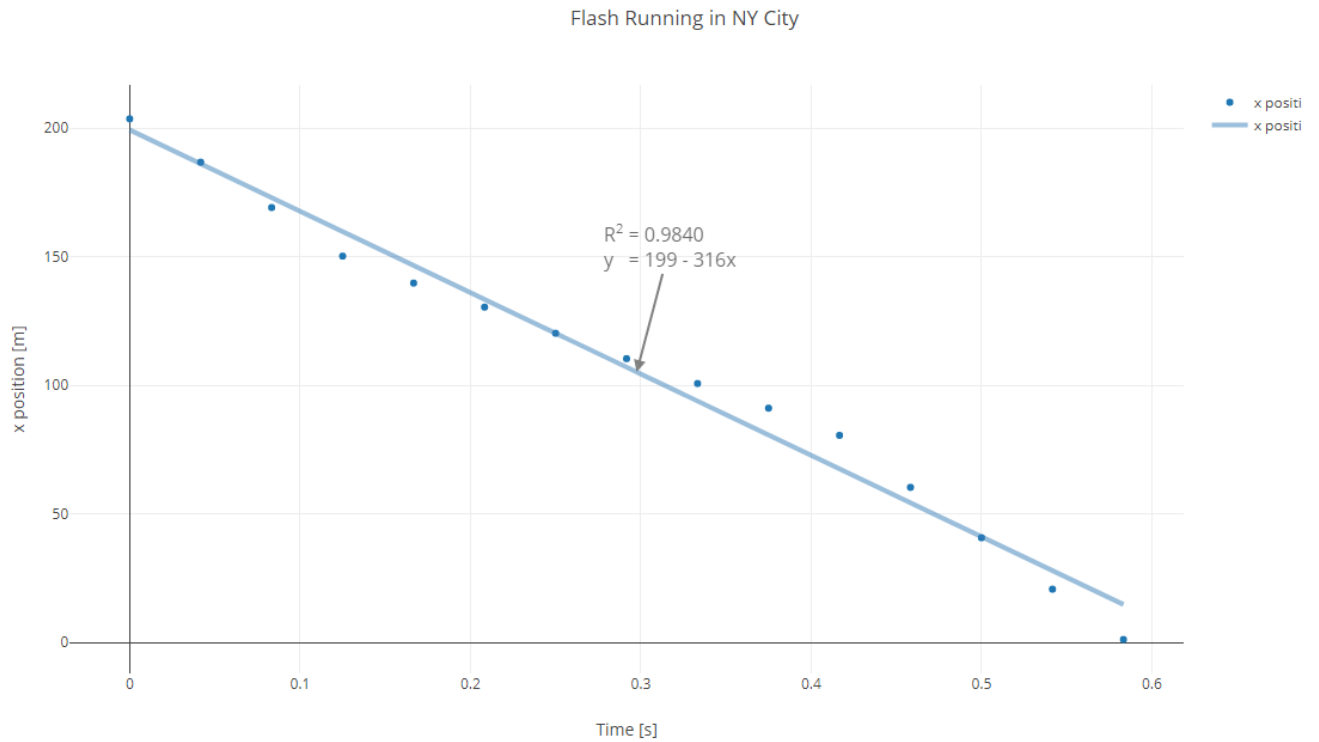


Figure 46 An example of a curve fitting in a scatterplot with Plotly

4 Case study

4.1 Description

In the context of this thesis, a case study of data storytelling was developed in which open data has been analysed and the results of this analysis are presented visually. Charts have been created with Plotly's Chart studio, a web application that has the potential to create dynamic, interactive visualizations of data.

The theme of the story is related to the sales of passenger cars and their characteristics registered in the last 8 years in Europe. The data came from the EEA's (European Environmental Agency) database, which is published every year and includes the main characteristics of cars sold in the European Member States (e.g. Engine displacement, Mass, CO₂ emissions etc.).

The Regulation (EC) No 443/2009 requires Member States to record information for each new passenger car registered in its territory. Every year, each Member State shall submit to the Commission all the information related to their new registrations. In particular, the following details are required for each new passenger car registered: manufacturer name, type approval number, type, variant, version, make and commercial name, specific emissions of CO₂ (NEDC and WLTP protocols), masses of the vehicle, wheel base, track width, engine capacity and power, fuel type and mode, eco-innovations and electricity consumption. Data for EU-28 are reported in the main database. All the parameters that are reported by the Member States are show in the Table 2 [40].

This database has been created for the first time in 2010. Each year the following process is taking place:

Member States shall record and transmit this information to the Commission by 28 February of each year for each new passenger car and van registered in their territory. Data shall be submitted to the Central Data Repository managed by the EEA

Many reliability checks were carried out by the EEA in order to assess the consistency and performance of the dataset. On the basis of the checks and feedback from the Member States, the EEA finalizes and publishes a preliminary database. At the same time, notification letters are sent to manufacturers informing them of their provisional CO₂ performance.

Manufacturers may inform the Commission of any errors in the data within three months of confirmation of the provisional estimate.

The EEA and the European Commission shall assess the manufacturer's corrections and, where appropriate, take them into account in the calculation of the manufacturer's final average CO₂ emissions and specific emission goals. The final data and targets are to be released by 31 October of each year.

The data are required for the Commission to calculate the average specific emissions of CO₂ from new passenger cars and to set the specific emission targets to be met by car manufacturers.

The database contains detailed data without taking into account certain considerations needed for the estimation of defined emission targets, such as phase-in rates, low-emission vehicle credits and E85 fuelled vehicles, and eco-innovation credits. To order to calculate the total annual emissions and the general emission targets, the Commission considers only those vehicles for which the emissions are calculated.

Table 2 EEA CO2 monitoring database [40]

Field name	Field Definition	Data type	Primary key
ID	Identification number	integer	Yes
MS	Member state	varchar(2)	No
Mp	Manufacturer pooling	varchar(50)	No
VFN	Vehicle family identification number	varchar(25)	No
Mh	Manufacturer name EU standard denomination	varchar(50)	No
Man	Manufacturer name OEM declaration	varchar(50)	No
MMS	Manufacturer name MS registry denomination	varchar(125)	No
TAN	Type approval number	varchar(50)	No
T	Type	varchar(25)	No
Va	Variant	varchar(25)	No
Ve	Version	varchar(35)	No
Mk	Make	varchar(25)	No
Cn	Commercial name	varchar(50)	No
Ct	Category of the vehicle type approved	varchar(5)	No
Cr	Category of the vehicle registered	varchar(5)	No
m (kg)	Mass in running order Completed/complete vehicle	integer	No
Mt	WLTP test mass	integer	No
Enedc (g/km)	Specific CO2 Emissions (NEDC)	integer	No
Ewltp (g/km)	Specific CO2 Emissions (WLTP)	integer	No
W (mm)	Wheel Base	integer	No
At1 (mm)	Axle width steering axle	integer	No
At2 (mm)	Axle width other axle	integer	No
Ft	Fuel type	varchar(25)	No
Fm	Fuel mode	varchar(1)	No
ec (cm3)	Engine capacity	integer	No
ep (KW)	Engine power	integer	No
z (Wh/km)	Electric energy consumption	integer	No
IT	Innovative technology or group of innovative technologies	varchar(25)	No

At the same time, with a view to create a story as much complete as possible, data from different open sources such as air pollutants (e.g. NO_x) were searched for and combined. Afterwards, the pre-processing of the data took place where the data was

"cleared" from any wrong / extreme values (outliers). Data was then transformed appropriately in order to easily manage it for further processing in Chart studio, while a statistical processing of the data was carried out, to find some initial interesting information and trends in an initial phase. At this point, the right visual to convey well our messages was selected. This choice was made by comparing different types of graphs. The final story (narration) was created on a dashboard.

4.2 How the passenger car market evolved?

Each year, Member states record all data related to the newly registered passenger cars and submit it to the EEA (European Environmental Agency) which is responsible for collecting the data and for publishing the final EU-28 database. Most of the following graphs are based on this dataset. The data visualization through the following dashboard provides a way to interpret meaningful stories from data, as well as easily recognize trends and patterns.

Few countries are dominating the passenger car market. Germany, United Kingdom, France, Italy and Spain are in descending order, the largest new-car markets in Europe.

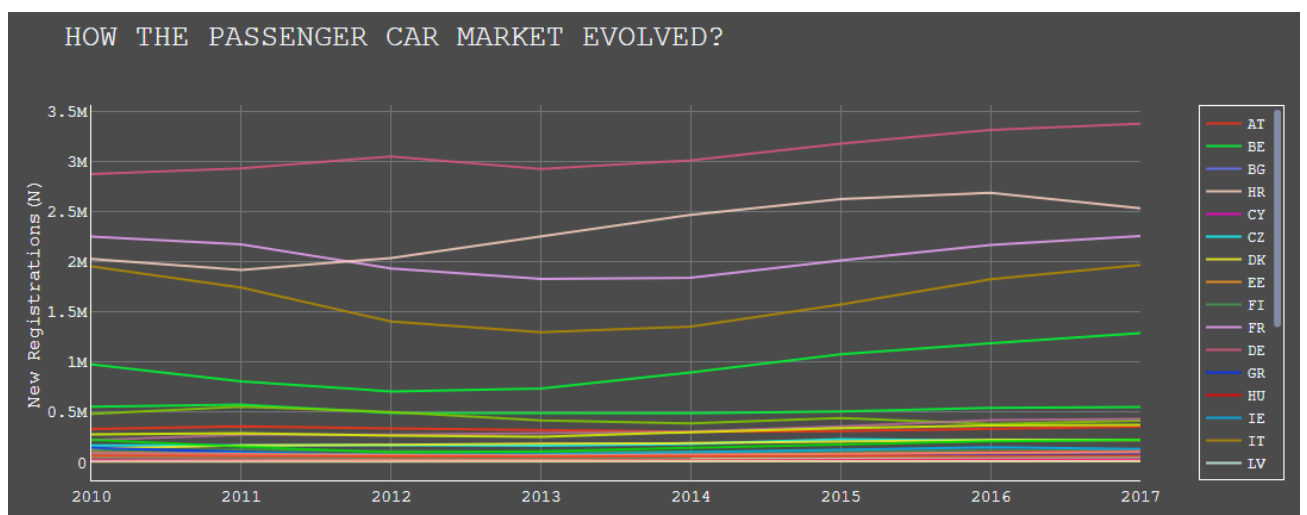


Figure 47 How the passenger car market evolved?

The economic crisis that Europe has suffered in recent years has affected the automotive market. The most affected countries were Greece, Portugal, Cyprus, Italy and

Spain. On the contrary, there were countries where there were increases in sales of new passenger cars such as Hungary, Denmark, Latvia, Lithuania, Poland and Germany.

The comparison of similar population countries such as Greece, Portugal, Sweden and Hungary can reveal a lot of interesting conclusions. It is obvious that the financial situation of the countries also affects the sales of cars.

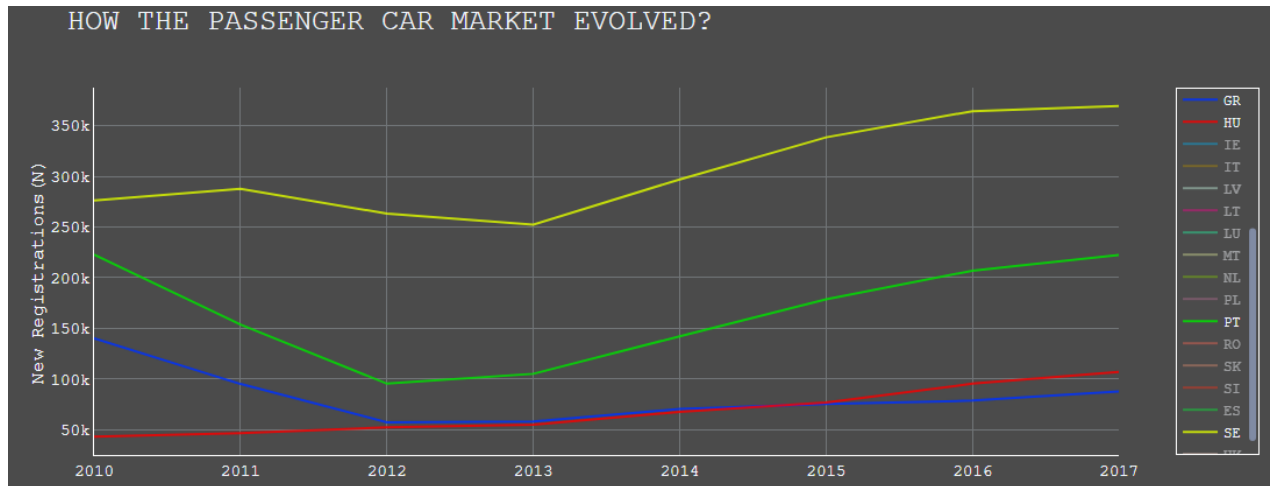


Figure 48 How the passenger car market evolved? The cases of Greece, Portugal, Hungary and Sweden

The Greek passenger car market has been suffered by the financial crisis as sales in 2012 have fallen by almost three times compared to 2010. The crisis has affected Portugal in a similar way, though in absolute terms car sales are well above Greece. Hungary did not appear to be affected by the economic crisis, with upward trend in car sales. Finally, it seems that Sweden has been slightly affected by the crisis and has multiple passenger car sales than all other countries.

4.3 Average CO2 emissions in 2017

The highest emitting cars were sold in Estonia. This high emission rate is probably due to the Estonians' love for large cars [41].

Norway is the frontrunner among European countries as it has the lowest average CO2 emissions from new passenger cars. This is due to the explosion of electric vehicles' sales. Financial and other incentives provided by the Norwegian government have led to

this performance as the cost of purchasing electric vehicle models is similar to conventional ones.

Greece holds the 5th best performance (108.8 g / km). The lifting of the diesel vehicles ban in Athens and Thessaloniki (2012) and the poor economic situation of the country are the main reasons for this good CO₂ performance. Low-priced and therefore small models were preferred by consumers, which generally lead to lower emissions.

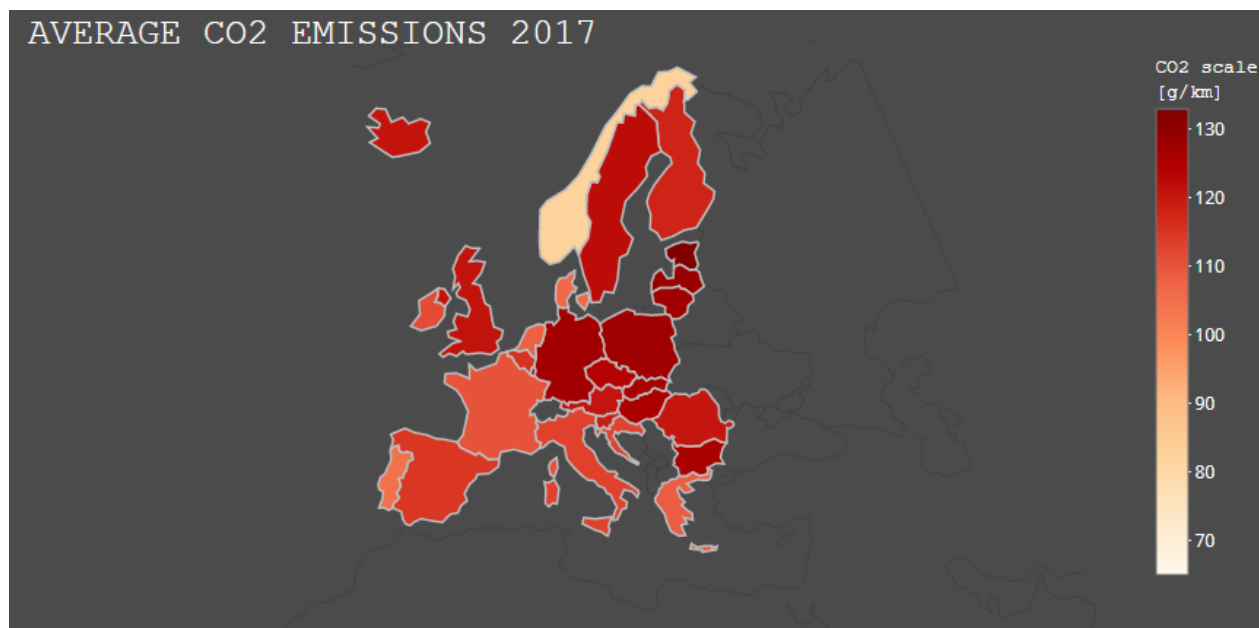


Figure 49 Average CO₂ emissions in 2017 from passenger cars in Europe

4.4 Historical CO₂ emissions and adopted targets

Average official CO₂ emissions (NEDC driving cycle) showed a significant drop of approximately 22 g/km in the period 2010-2017 (average annual reduction 3.1 g/km). Indeed, the target set for 2015 (130 g / km) was achieved two years earlier than it was expected. For the 2021 target (95 g / km), the rate of emission reduction should be increased to around 5.9 g / km per year.

However, the official test procedure that was used for monitoring and reporting of CO₂ emissions of light duty vehicles in Europe (NEDC) has been recognised non-representative of real-world vehicle usage, fuel consumption and CO₂ emissions leading to a gap between officially reported emissions and the ones experienced during real world operation. To address this gap, a new test protocol was designed (WLTP: Worldwide

Harmonized Light duty vehicle Test) that is expected to provide more realistic fuel consumption and emissions results.

Real-world CO2 emissions are significantly higher and dropped just 4.5% over the period 2010-2017. Indeed, the gap between real-world emissions and official emissions is rising over the years. This is attributable to the progressively greater exploiting of technical tolerances and loopholes in the official testing procedure, by manufacturers.

From 2021, the EU fleet-wide average emission target for new cars will be 95 g CO2/km. In addition, new EU fleet-wide CO2 emission targets are set for the years 2025 and 2030, for newly registered passenger cars. These targets are defined as a percentage reduction from the 2021 starting points: -15% reduction from 2025 on and 37.5% reduction from 2030.

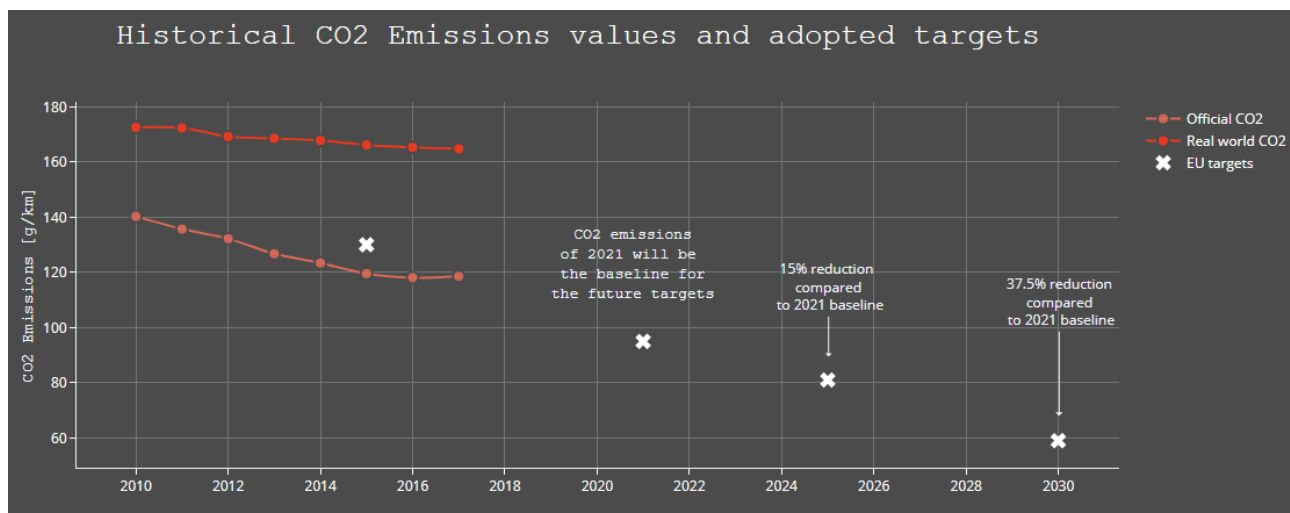


Figure 50 Historical CO2 emissions and adopted targets

4.5 Average parameters 2017

There is a correlation of engine power, mass, engine capacity and CO2 emissions. Countries where small vehicles are sold (in terms of mass, power and capacity) have generally lower CO2 emissions (coloured with pink), while those in which powerful cars are sold (e.g. Luxembourg), these vehicles are heavier, with larger capacity and they consume much more energy.

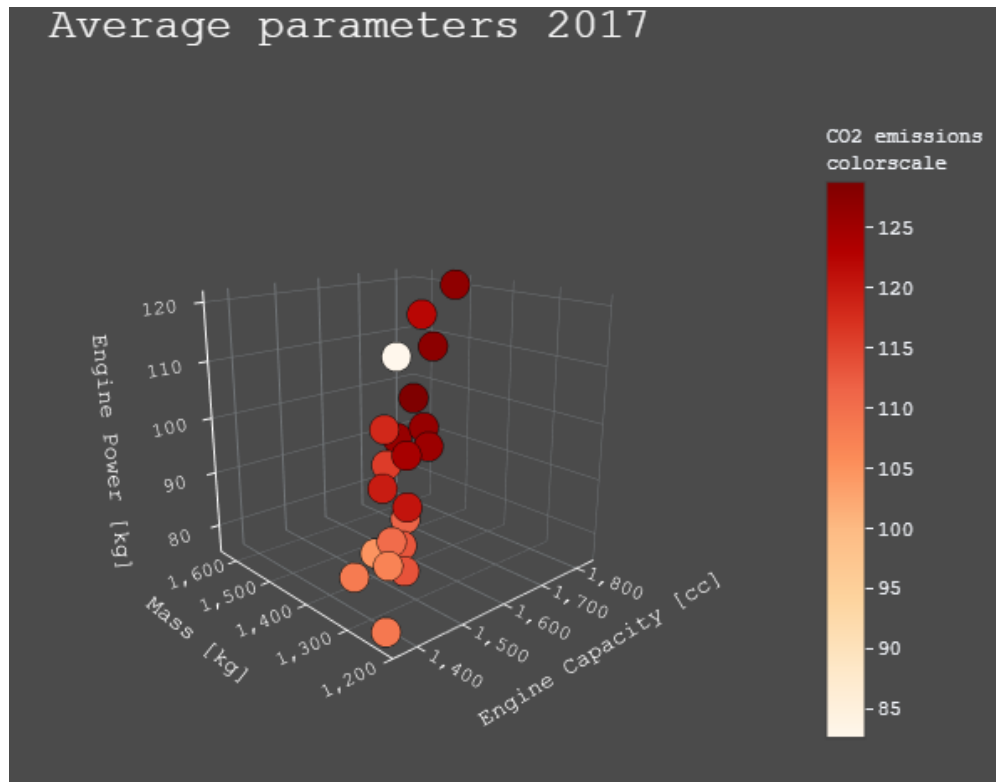


Figure 51 Average mass, engine power and capacity in 2017

The exception is the case of Norway as it has the highest average weight but it is coloured white (i.e. it has the smallest emissions). In general, electric vehicles are heavier than their conventional competitors, mainly because of the batteries weight. Electric cars also require a charger, inverter, and motor which all add to the weight of an electric car.

4.6 Heatmaps – Average parameters

The average **mass** values of the new vehicles followed a slight increasing trend. Vehicle mass is a critical parameter in vehicle design as it is linked closely to the fuel consumption. Increasing the mass of the vehicle leads to a rise in the energy needed to accelerate the vehicle and an increase in the rolling resistance resulting in increased consumption (and thus CO2 emissions). In order to examine the quantitative effect of weight, several studies have been carried out which demonstrate that for additional mass of 50 – 200 kg over various cycles and operating conditions, the increase of fuel consumption ranges from 5 to 9% [Fuel consumption and CO2 emissions from passenger cars in Europe – Laboratory versus real-world emissions]

Engine displacement has generally declined over the years. Turbocharging and advances in engine technology denote that some of today's smaller engines are able to generate more power than certain larger, more old-fashioned engines. Manufacturers can therefore reduce the number of cylinders of an engine and generally reduce engine capacity and that happened in the previous years.

High competition among manufacturers has led to a constant increase in **engine power** of passenger cars over the years.

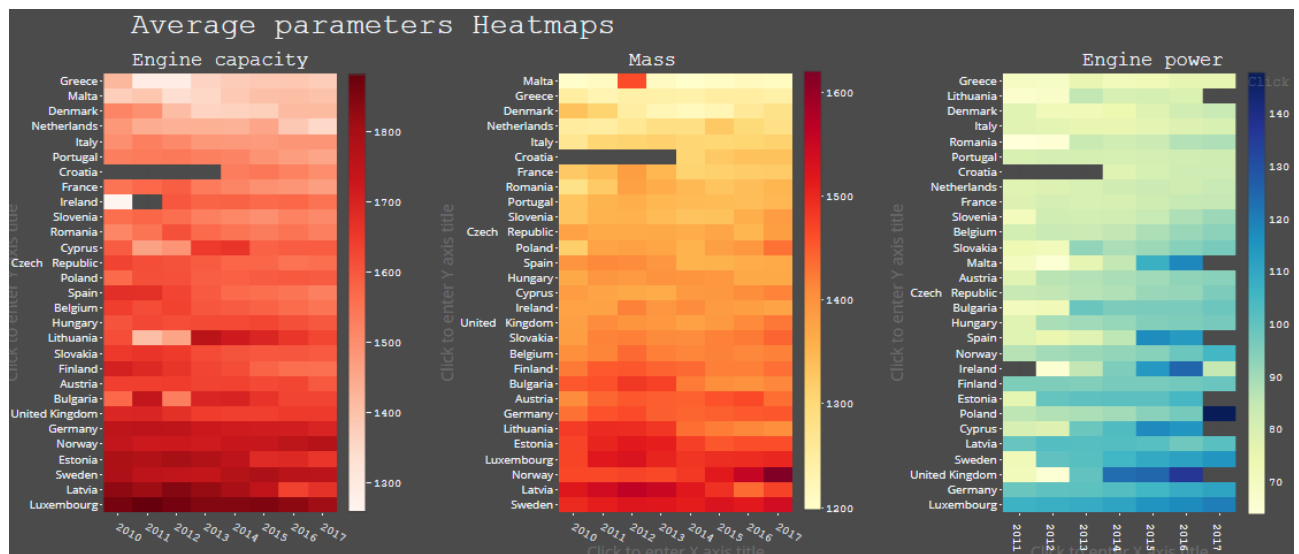


Figure 52 Average parameters through a heatmap

4.7 Market share of new car sold in the EU per fuel

Diesel vehicles were the predominant choice of consumers for the time period 2010 - 2016. In September 2015, the Volkswagen emissions scandal or "Dieselgate" breaks out, making consumers more cautious towards diesel. The United States Environmental Protection Agency (EPA) found that Volkswagen had intentionally programmed turbocharged direct injection (TDI) diesel engines to activate their emissions controls only during laboratory emissions testing which caused the vehicles' NOx output to meet US standards during regulatory testing, but emit up to 40 times more NOx in real-world driving. This software has been applied in about eleven million cars worldwide, from 2009 through 2015.

Following the Dieselgate scandal, some European countries (UK, France, Ireland, Spain) proposed that the sale of diesel cars and petrol cars will be banned in the near future (2030-2040) [42]–[44]

Dieselgate speeded the drop in the share sharply, reaching about 44.5% in 2017. In 2017 the dominant choice of consumers was petrol vehicles.

Dieselgate affected market shares:

Diesel vehicles decreased by more than 7 percentage points and Petrol increased of about 6 percentage points since 2015

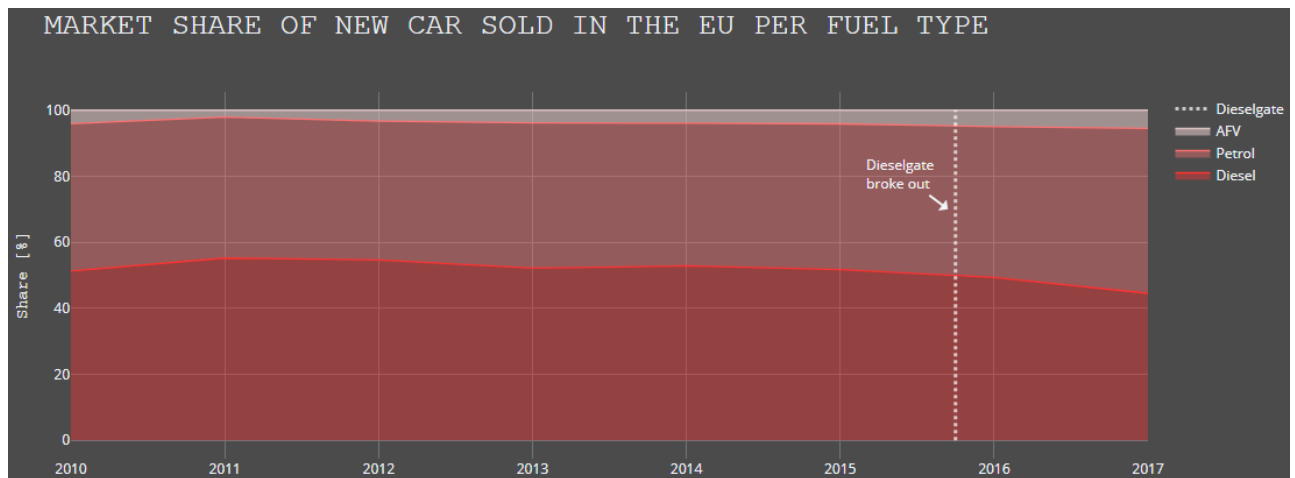


Figure 53 Market share of new car sold in the EU per fuel

4.8 Market share evolution for top manufacturers

How badly has the VW diesel scandal hurt VW?

The Diesel Gate scandal had a clear negative impact on Volkswagen's market share. Manufacturers who seem to be favoured from the scandal were Daimler and Renault.

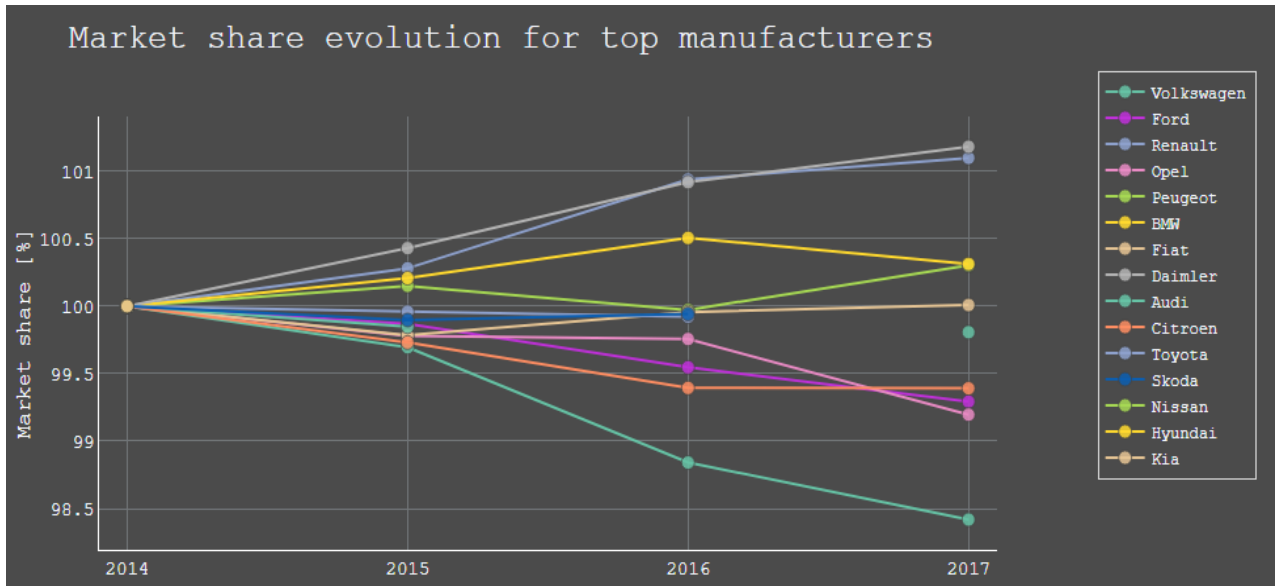


Figure 54 Market share evolution for top manufacturers

4.9 CO2 emissions per powertrain

Are diesel cars really more polluting than petrol cars?

As part of the EU's reaction to the 1997 Kyoto Protocol to decrease greenhouse gas pollution, Diesel was introduced as a more eco-friendly fuel. A diesel car's overall CO2 emissions tend to be smaller.

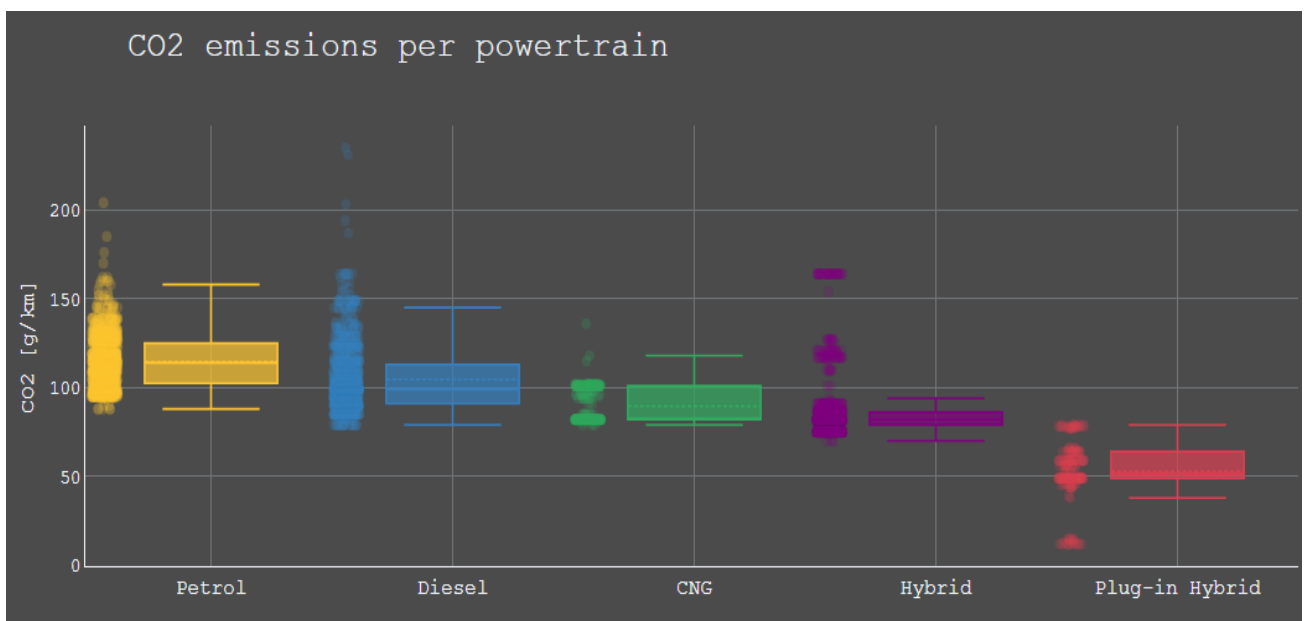


Figure 55 CO2 emissions per powertrain

However, other toxic emissions, particularly nitrogen oxides (NOx) and particulates (PM), are emitted in considerably higher levels by diesel vehicles. In built-up urban areas, these emissions mean that diesel engines are usually the major source of roadside air pollution.

Long-term exposure to NOx can considerably boost the risk of harming respiratory system and particulate matter (PM) have also been accused of causing cancer.

4.10 Real world vs Euro standards limits NOx emissions per fuel type

Diesel vehicles have generally higher NOx emissions limits per Euro standard technology compared to Petrol vehicles. However, real-world NOx emissions are 7.5 times more than the legal limit. Petrol emissions, on the other hand, are within the limits of Euro 5 and Euro 6 vehicles.

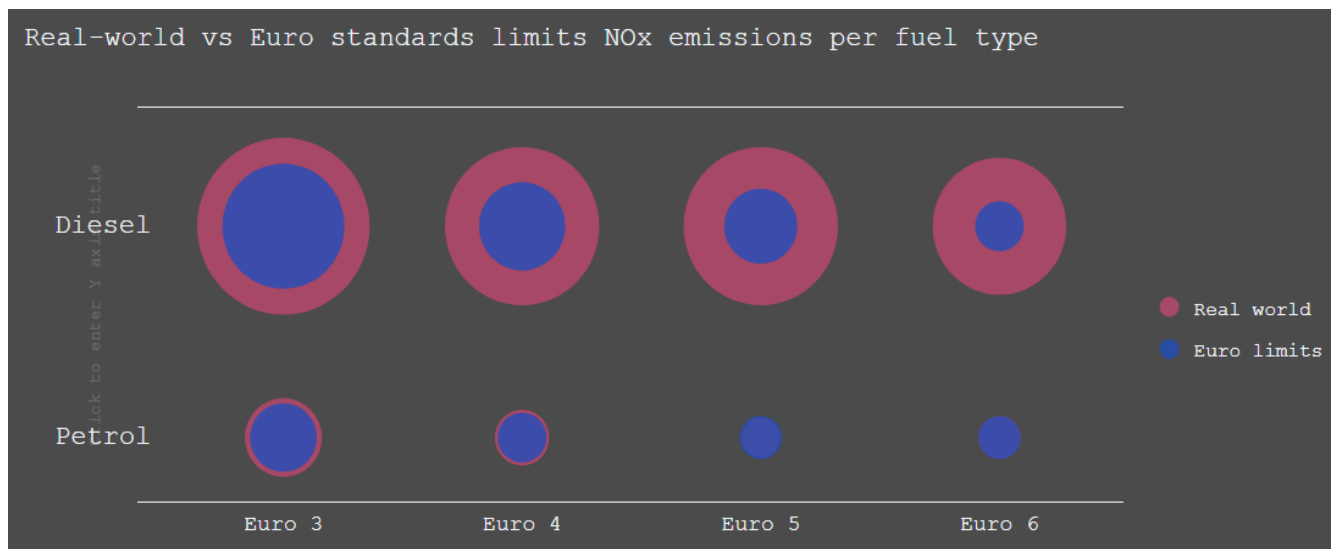


Figure 56 Real world vs Euro standards limits NOx emissions per fuel type

In order to accomplish the goals of emission reduction, new powertrain technologies like BEV and PHEV need to be deployed progressively.

4.11 New registrations of PHEVs

A plug-in hybrid electric vehicle (PHEV) is a hybrid electric vehicle that combines an internal combustion engine with an electric motor and a large battery that can be recharged by plugging it into an outlet. It has two operational modes 1) all-electric mode in which electric motor and battery provides the energy for the propulsion and 2) hybrid mode in which both electricity and fossil fuel are employed. PHEVs combine the fuel and cost efficiency of hybrid models along with the all electric potentials of BEV or FCEV vehicles. On top of that, PHEVs address one of the big concerns about electric vehicles, which is the limited range. Although plug-in hybrids typically emit less emissions than conventional vehicles, the total amount produced depends partly on how the electricity is produced. For example, nuclear and hydroelectric plants are cleaner than coal-fired power plants.

Many of the well-known car manufacturers already brought PHEV models into the market and the number of the PHEV models is expected to upsurge in the coming years. Available PHEV models have risen to a total of about 30. Those that stand out in sales are Mitsubishi Outlander and Golf / Passat.

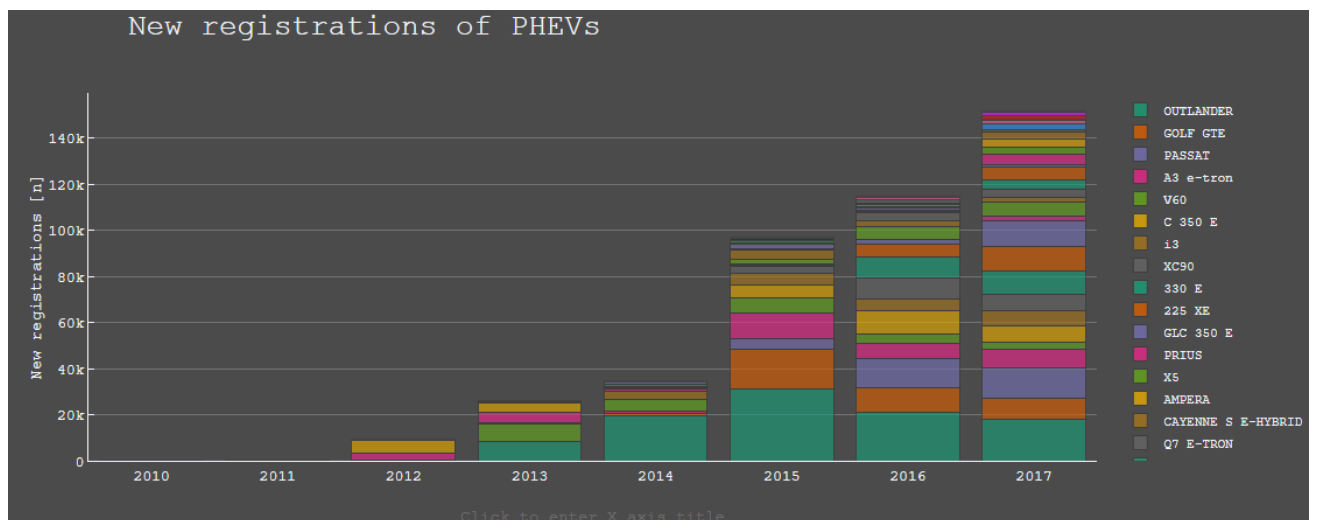


Figure 57 New registrations of PHEVs

4.12 New registrations of BEVs

Battery electric vehicles (BEV) use a battery-powered electric motor drive system to drive the vehicle. BEVs don't produce exhaust emissions at all, contributing in significant decrease in local air pollution, and thus they considered as an environmentally

friendly technology. However, the electricity they use may produce GHG and other pollution at the source of its generation.

Recharging an EV lasts considerably longer than refuelling a conventional vehicle. BEVs have a similar number of sales to PHEVs. The major BEV models in terms of sales are ZOE, LEAF, MODEL-S, i3 and e-GOLF.

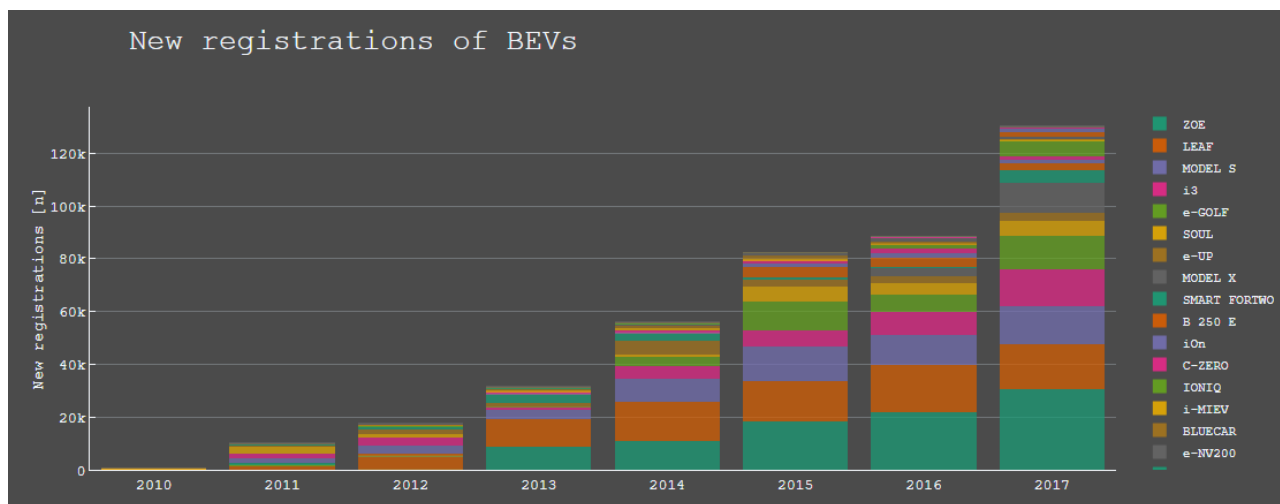


Figure 58 New registrations of BEVs

4.13 Technology shares

Despite the promotion of the electrified vehicles in some Member states the sales figures are disappointing. Aside from a few outliers (e.g. Norway), the electrified vehicles (BEV / PHEV / FCEV) sold in European Union are below 2% of the total sales in 2017. The main reasons for the slow introduction of EVs are high purchase cost, limited driving range and longer charging time compared to conventional cars, along with lack of sufficient infrastructure.

However, most of the manufacturers have already announced investments in the electrified vehicles in order to achieve their specific emissions targets.

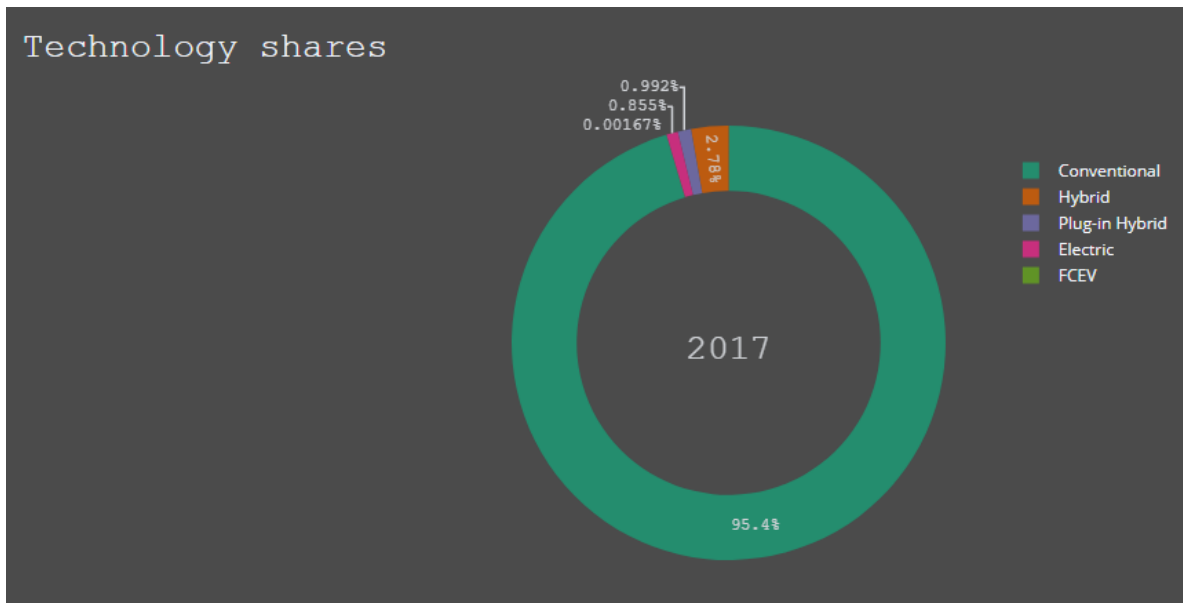


Figure 59 Technology of passenger car shares in 2017

Electric vehicles are identified as the most prominent solution from the car makers as offer a quick solution to two significant problems: reducing GHG from transport, and addressing air pollution in city centres.

5 Conclusions

The present thesis focused on data visualization and data storytelling with modern tools. In the literature part of the thesis, a review was made exploring key aspects of the scientific areas, trends emerging in recent years and interesting examples explaining the best practices to be applied.

Complementary, a case study was developed where open data is exploited, visualized and a story is drawn from which interesting conclusions can be drawn.

5.1 Research overview and findings

It's often said that a picture is worth a thousand words and data visualization works the same way. A visual representation or a chart may contain a lot more information than text for a similar amount of space on a page/slide etc and at the same time to be much more engaging to the audience. In this thesis, an in-depth analysis was made for the data

visualisation. In particular, there are some significant conclusions arising from this analysis:

- Static visualisation technique should be preferred for relatively simple and straightforward stories that intended to present a single point, whereas, interactive visualisations should be chosen for the presentation of the results from large datasets that contain complex relationships.
- Exploratory analysis includes everything that need to be done by the author to find interesting points of data and then should be presented to the audience in a smart way (explanatory analysis).
- The author should be very careful and the display of the data should be done as objectively as possible, to avoid conveying wrong messages to the audience.
- Hearing a story leads to the release of dopamine, which in turn causes euphoria. This is a bio-chemical reason that makes people prefer stories than raw data.
- There different types of story structures, author-driven, viewer-driven and a mix of them called martini glass.

In the case study part, Plotly's Chart studio was used. This tool seems to offer very good visualization capabilities since it covers a very large number of charts while combining many desired interactivity features that help the user to analyze and explore the charts and draw important conclusions. Even, in the case study it became obvious that it is a tool that provides the user with a very user friendly interface. The advantages of Plotly also includes the fact that its basic version is free while in this basic version the user may find almost everything that is necessary. Finally, it is worth noting that there is also the Plotly Community forum which can provide many answers from experts to various issues that may arise during the creation of a chart.

As disadvantages of Plotly chart studio can be stated that the user cannot extract the code behind the chart that he created. Another drawback is that it does not provide the user with the opportunity to create "animation" charts.

5.2 Limitations

As for the case study, it would be nice if we could implement an “animation” diagram that was not possible. Previously this feature was available to users but has now been removed.

5.3 Future research

Further research can be done to evaluate alternative data visualization tools. It would be nice to make a comparison of the most important data visualization tools available so that the users can make the best decision according to their needs.

6 References

- [1] “IBM Software | What is big data.” [Online]. Available: <http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>. [Accessed: 08-Oct-2019].
- [2] BSA, “What ’ s the Big Deal With Data ? Executive Summary,” 2018.
- [3] United Nations, “Making Data Meaningful - data presentation,” vol. 1, 2015.
- [4] S. Murray, “Interactive Data Visualization for the Web: AN INTRODUCTION TO DESIGNING WITH D3,” *O’Reilly Media*. p. 447, 2017.
- [5] C. N. Knaflic, *Storytelling with data - a data vizualization guide for business professionals*. 2015.
- [6] V. Friedman, “Data Visualization and Infographics — Smashing Magazine,” 2008. [Online]. Available: <https://www.smashingmagazine.com/2008/01/monday-inspiration-data-visualization-and-infographics/>.
- [7] “Humans Process Visual Data Better.” [Online]. Available: <http://www.t-sciences.com/news/humans-process-visual-data-better>. [Accessed: 15-Oct-2019].
- [8] D. Hyerle, “Thinking maps: Visual tools for activating habits of mind,” 2000.
- [9] L. Jukes, I. McCain, T. D. E., Crockett, *Understanding the digital generation: Teaching and learning in the new digital landscape*. 2010.
- [10] E. Jensen, “Brain-Compatible Learning,” *Int. Alliance Learn.*, vol. 3, 1996.
- [11] “The Power of Visuals in eLearning Infographic - e-Learning Infographics.” [Online]. Available: <https://elearninginfographics.com/power-visuals-in-elearning-infographic/>. [Accessed: 15-Oct-2019].
- [12] W. Playfair, *The Commercial and Political Atlas: Representing, by Means of Stained Copper-Plate Charts, the Progress of the Commerce, Revenues, Expenditure and Debts of England during the Whole of the Eighteenth Century*. .
- [13] Snow John, “On the Mode of Communication of Cholera,” 1855. [Online]. Available: http://matrix.msu.edu/~johnsnow/images/online_companion/chapter_images/fig12-5.jpg. [Accessed: 15-Oct-2019].
- [14] “Key Figures in the History of Data Visualization - Infogram - Medium.” [Online]. Available: <https://medium.com/@Infogram/key-figures-in-the-history-of-data-visualization-30486681844c>. [Accessed: 15-Oct-2019].

- [15] “exploratory vs explanatory analysis — storytelling with data.” [Online]. Available: <http://www.storytellingwithdata.com/blog/2014/04/exploratory-vs-explanatory-analysis>. [Accessed: 15-Oct-2019].
- [16] “Ranking Popular JavaScript Visualization Packages for Data Science | The Data Incubator.” [Online]. Available: <https://blog.thedataincubator.com/2018/03/ranking-popular-javascript-visualization-packages-for-data-science/>. [Accessed: 15-Oct-2019].
- [17] “D3.js - Data-Driven Documents.” [Online]. Available: <https://d3js.org/>. [Accessed: 15-Oct-2019].
- [18] “Chart.js samples.” [Online]. Available: <https://www.chartjs.org/samples/latest/>. [Accessed: 15-Oct-2019].
- [19] “Interactive JavaScript charts for your webpage | Highcharts.” [Online]. Available: <https://www.highcharts.com/>. [Accessed: 15-Oct-2019].
- [20] “Examples - Plottable.” [Online]. Available: <http://plottablejs.org/examples/>. [Accessed: 15-Oct-2019].
- [21] “plotly.js | JavaScript Graphing Library.” [Online]. Available: <https://plot.ly/javascript/>. [Accessed: 15-Oct-2019].
- [22] “Three Common Ways to Lie with Data Visualization.” [Online]. Available: <https://heap.io/blog/data-stories/how-to-lie-with-data-visualization>. [Accessed: 15-Oct-2019].
- [23] “Lies, damn lies, and Steve Jobs keynotes | Technology | The Guardian.” [Online]. Available: <https://www.theguardian.com/technology/blog/2008/jan/21/liesdamnliesandstevej-obs>. [Accessed: 15-Oct-2019].
- [24] “After Babylon.” [Online]. Available: <http://www.puffpuffproject.com/languages.html>. [Accessed: 15-Oct-2019].
- [25] “What’s Really Warming the World? Climate deniers blame natural factors; NASA data proves otherwise.” [Online]. Available: <https://www.bloomberg.com/graphics/2015-whats-warming-the-world/>. [Accessed: 15-Oct-2019].
- [26] “Bus Bunching Explained Visually.” [Online]. Available: <http://setosa.io/bus/>. [Accessed: 15-Oct-2019].
- [27] “How People Like You Spend Their Time | FlowingData.” [Online]. Available:

- <https://flowingdata.com/2016/12/06/how-people-like-you-spend-their-time/>.
[Accessed: 15-Oct-2019].
- [28] “The Pudding.” [Online]. Available: <https://pudding.cool/>. [Accessed: 15-Oct-2019].
- [29] “Information is Beautiful.” [Online]. Available: <https://informationisbeautiful.net/>. [Accessed: 15-Oct-2019].
- [30] “FlowingData.” [Online]. Available: <https://flowingdata.com/>. [Accessed: 15-Oct-2019].
- [31] “Tableau Blog | Tableau Software.” [Online]. Available: <https://www.tableau.com/about/blog>. [Accessed: 15-Oct-2019].
- [32] “FiveThirtyEight | Nate Silver’s FiveThirtyEight uses statistical analysis — hard numbers — to tell compelling stories about politics, sports, science, economics and culture.” [Online]. Available: <https://fivethirtyeight.com/>. [Accessed: 15-Oct-2019].
- [33] “data storytelling - Google Trends.” [Online]. Available: [https://trends.google.com/trends/explore?q=data storytelling&geo=US](https://trends.google.com/trends/explore?q=data%20storytelling&geo=US). [Accessed: 15-Oct-2019].
- [34] K. Caine, R., Caine, G., McClintic, C., & Klimic, *12 Brain/mind learning principles in action*. 2005.
- [35] K. K. Richter M., “How to increase retention through storytelling - TechRepublic,” 2000. [Online]. Available: <https://www.techrepublic.com/article/how-to-increase-retention-through-storytelling/>. [Accessed: 16-Oct-2019].
- [36] G. Rodriguez, “This Is Your Brain On Storytelling: The Chemistry Of Modern Communication,” 2017. [Online]. Available: <https://www.forbes.com/sites/giovannirodriguez/2017/07/21/this-is-your-brain-on-storytelling-the-chemistry-of-modern-communication/#7c253e9dc865>. [Accessed: 16-Oct-2019].
- [37] C. Abrahamson, *Storytelling as a pedagogical tool in higher education*. 1998.
- [38] H. J. Watson, “Data Visualization, Data Interpreters, and Storytelling,” *Bus. Intell. J.*, vol. 22, no. 1, pp. 5–11, 2015.
- [39] P. Man, “Narrative Structures in Data Visualizations to Improve Storytelling,” *Masters of Media*, no. 9, pp. 1–3, 2011.
- [40] EEA, “Monitoring of CO2 emissions from passenger cars – Regulation (EC) No

443/2009,” 2019. [Online]. Available: <https://www.eea.europa.eu/data-and-maps/data/co2-cars-emission-16>.

- [41] “Estonians love high-emission cars. Or do they? | Economy | ERR.” [Online]. Available: <https://news.err.ee/115659/estonians-love-high-emission-cars-or-do-they>. [Accessed: 18-Oct-2019].
- [42] “Diesel crisis deepens as western Europe sends hundreds of thousands of banned cars to Poland | The Independent.” [Online]. Available: <https://www.independent.co.uk/news/world/politics/diesel-cars-europe-germany-poland-air-pollution-dieselgate-motoring-a8591641.html>. [Accessed: 16-Oct-2019].
- [43] “Bans on petrol, diesel cars to intensify in Europe.” [Online]. Available: <https://www.businesslive.co.za/bd/life/motoring/2018-11-22-bans-on-petrol-diesel-cars-to-intensify-in-europe/>. [Accessed: 16-Oct-2019].
- [44] “Ireland to ban new petrol and diesel vehicles from 2030 - BBC News.” [Online]. Available: <https://www.bbc.com/news/world-europe-48668791>. [Accessed: 16-Oct-2019].

