# MODULE I

Context of data visualization – Definition, Methodology, Visualization design objectives. Key Factors – Purpose, visualization function and tone, visualization design options – Data representation, Data Presentation, Seven stages of data visualization, widgets, data visualization tools.

# DATA VISUALIZATION –INTRODUCTION

In simple words, data visualization is a graphical representation of any data or information. Visual elements such as charts, graphs, and maps are the few data visualization tools that provide the viewers with an easy and accessible way of understanding the represented information. In this world governed by Big Data, data visualization enables you or decision-makers of any enterprise or industry to look into analytical reports and understand concepts that might otherwise be difficult to grasp.

Why is data visualization important?

By now, you would have understood how data visualization simplifies the way information is presented. However, is that the only power of data visualization? Not really. As the world is changing, the need for information is changing as well. Here are a few benefits of data visualization:

● Easily, graspable information – Data is increasing day-by-day, and it is not wise for anyone to scram through such quantity of data to understand it. Data visualization comes handy then.

● Establish relationships – Charts and graphs do not only show the data but also established co-relations between different data types and information.

● Share – Data visualization is also easy to share with others. You could share any important fact about a market trend using a chart and your team would be more receptive about it.

● Interactive visualization – today, when technological inventions are making waves in every market segment, regardless of big or small, you could also leverage interactive visualization to dig deeper and segment the different portions of charts and graphs to obtain a more detailed analysis of the information being presented.

● Intuitive, personalized, updatable – Data visualization is interactive. You could click on it and get another big picture of a particular information segment. They are also tailored according to the target audience and could be easily updated if the information modifies.

## What are Data Visualization examples?

What better way to understand data visualization, if not with examples? Here are a few for your reference:

● Government Budget – Government budgets are always tough to understand as they number and more numbers. A recent example is a colour-coded treemap that was designed by The White House during Barack Obama's presidency, which visually broke down the US's 2016 the budget for better understanding and put government programs in context.

● World population – How would you present the world population along with their density? Simple, by visual representation. A world map showing the population density is another

data visualization example.
● Profit and loss – Business companies often resort to pie charts or bar graphs showing their annual profit or loss margin.
● Films and dialogues – Out of many characters in the film who will have how many dialogues? Data visualization is the answer here. The makers of popular sitcom 'FRIENDS' used a pie chart during shooting to ensure that every six characters have an equal number of jokes and dialogues.
● Anscombe's quartet – It is one of the most well-known and popular, which has four data sets of identical descriptive statistics, but they appear different when graphed.
All of these four data sets have different distributions and consists of 11 points marked on x and y-axis.

## Categories of Data Visualization ;

Data visualization is very critical to market research where both numerical and categorical data can be visualized that helps in an increase in impacts of insights and also helps in reducing risk of analysis paralysis. So, data visualization is categorized into following categories :

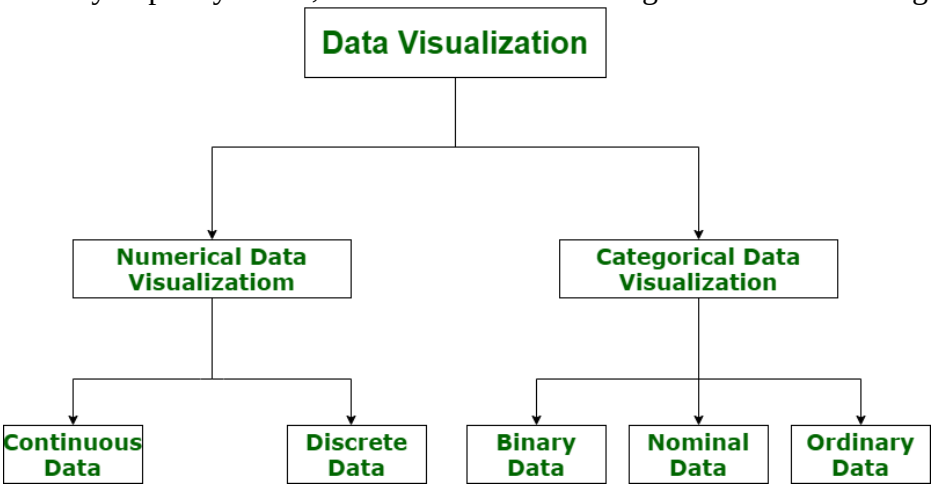



**Figure –** Categories of Data Visualization

1. **Numerical Data :**
   Numerical data is also known as Quantitative data. Numerical data is any data where data generally represents amount such as height, weight, age of a person, etc. Numerical data visualization is easiest way to visualize data. It is generally used for helping others to digest large data sets and raw numbers in a way that makes it easier to interpret into action. Numerical data is categorized into two categories :
   - **Continuous Data –**
     It can be narrowed or categorized (Example: Height measurements).
   - **Discrete Data –**
     This type of data is not "continuous" (Example: Number of cars or children's a household has).

The type of visualization techniques that are used to represent numerical data visualization is Charts and Numerical Values. Examples are Pie Charts, Bar Charts, Averages, Scorecards, etc.

2. **Categorical Data :**

Categorical data is also known as Qualitative data. Categorical data is any data where data generally represents groups. It simply consists of categorical variables that are used to represent characteristics such as a person's ranking, a person's gender, etc. Categorical data visualization is all about depicting key themes, establishing connections, and lending context. Categorical data is classified into three categories :

- **Binary Data –**
  In this, classification is based on positioning (Example: Agrees or Disagrees).
- **Nominal Data –**
  In this, classification is based on attributes (Example: Male or Female).
- **Ordinal Data –**
  In this, classification is based on ordering of information (Example: Timeline or processes).

The type of visualization techniques that are used to represent categorical data is Graphics, Diagrams, and Flowcharts. Examples are Word clouds, Sentiment Mapping, Venn Diagram, etc.

---

# The Context of Data Visualization

The ability to take data—to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it—that's going to be a hugely important skill in the next decades.

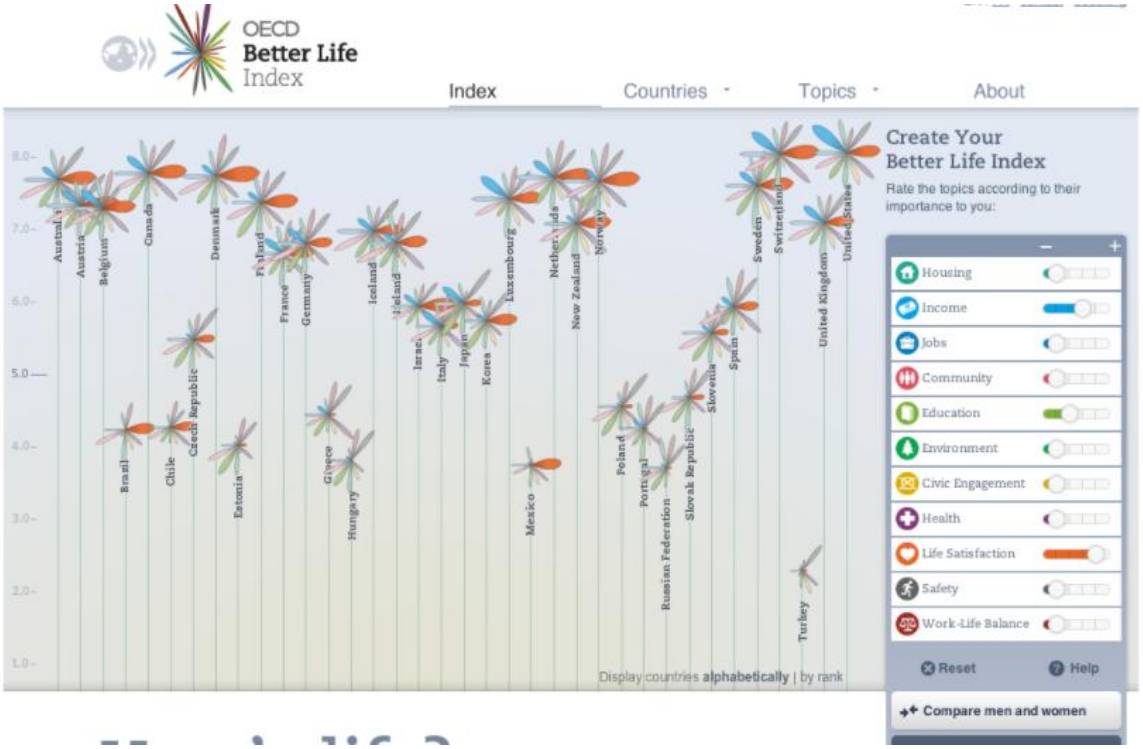
Data visualization is not new; the visual communication of data has been around in various forms for hundreds and arguably thousands of years. Popular methods that still dominate the boardrooms of corporations across the land—the line, bar, and pie charts—originate from the eighteenth century.

Catalyzed by powerful new technological capabilities as well as a cultural shift towards greater transparency and accessibility of data, the field has experienced a rapid growth in enthusiastic participation.

Where once the practice of this discipline would have been the preserve of specialist statisticians, engineers, and academics, the globalized field that exists today is a very active, informed, inclusive, and innovative community of practitioners pushing the craft forward in fascinating directions. The following image shows a screenshot of the OECD 'Better Life Index', comparing

well-being across different countries. This is just one recent example of an extremely successful visual tool emerging from this field.

*Image from "OECD Better Life Index" (http://oecdbetterlifeindex.org), created by Moritz Stefaner*




Data visualization is the multi-talented, boundary-spanning trendy kid that has seen many esteemed people over the past few years, such as Hal Varian, forecasting this as one of the next big things.

Anyone considering data visualization as a passing fad or just another vacuous buzzword is short-sighted; the need to make sense of and communicate data to others will surely only increase in relevance. However, as it evolves from the *next* big thing to the big thing, the field is at an important stage of its diffusion and maturity. Expectancy has been heightened and it does

have a certain amount to prove; something concrete to deliver beyond just experimentation and constant innovation.

## Visualization as a discovery tool

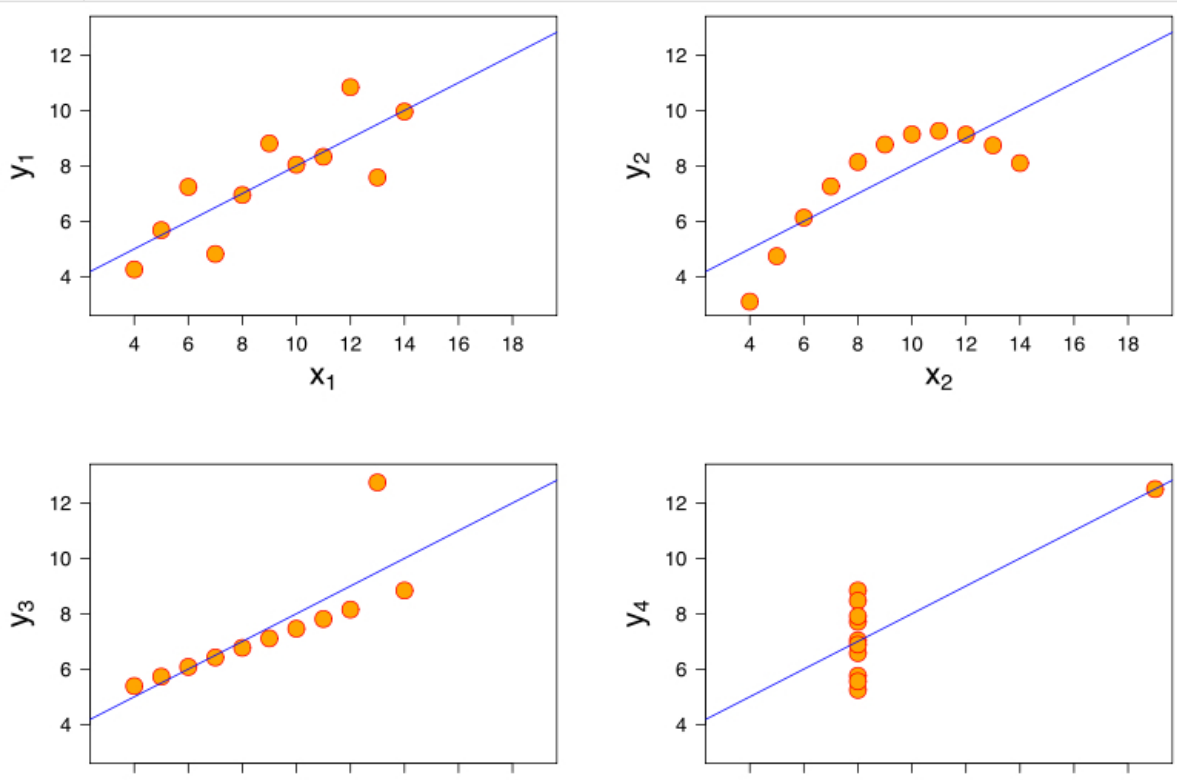
Through visualization, we are seeking to portray data in ways that allow us to see it in a new light, to visually observe patterns, exceptions, and the possible stories that sit behind its raw state. This is about considering visualization as a tool for discovery.

A well known demonstration that supports this notion was developed by noted statistician Francis Anscombe (incidentally, brother-in-law to Tukey) in the 1970s. He compiled an experiment involving four sets of data, each exhibiting almost identical statistical properties including mean, variance, and correlation. This was known as "Anscombe's quartet".

| x1 | y1 | x2 | y2 | x3 | y3 | x4 | y4 |
|---|---|---|---|---|---|---|---|
| 10 | 8.04 | 10 | 9.14 | 10 | 7.46 | 8 | 6.58 |
| 8 | 6.95 | 8 | 8.14 | 8 | 6.77 | 8 | 5.76 |
| 13 | 7.58 | 13 | 8.74 | 13 | 12.74 | 8 | 7.71 |
| 9 | 8.81 | 9 | 8.77 | 9 | 7.11 | 8 | 8.84 |
| 11 | 8.33 | 11 | 9.26 | 11 | 7.81 | 8 | 8.47 |
| 14 | 9.96 | 14 | 8.1 | 14 | 8.84 | 8 | 7.04 |
| 6 | 7.24 | 6 | 6.13 | 6 | 6.08 | 8 | 5.25 |
| 4 | 4.26 | 4 | 3.1 | 4 | 5.39 | 19 | 12.5 |
| 12 | 10.84 | 12 | 9.13 | 12 | 8.15 | 8 | 5.56 |
| 7 | 4.82 | 7 | 7.26 | 7 | 6.42 | 8 | 7.91 |
| 5 | 5.68 | 5 | 4.74 | 5 | 5.73 | 8 | 6.89 |

Ask yourself, what can you *see* in these sets of data? Do any patterns or trends jump out? Perhaps the sequence of eights in the fourth set? Otherwise there's nothing much of interest evident.

So what if we now visualize this data, what can we see then?




Through the previous graphical display, we can immediately see the prominent patterns created by the relationships between the X and Y values across the four sets of data as follows:

- the general tendency about a trend line in **X1**, **Y1**
- the curvature pattern of **X2**, **Y2**
- the strong linear pattern with single outlier in **X3**, **Y3**
- the similarly strong linear pattern with an outlier for **X4**, **Y4**

The intention and value of Anscombe's experiment was to demonstrate the importance of presenting data graphically. Rather than just describing a dataset based on a selection of some of its key statistical properties alone, to make proper sense of data, and avoid forming false conclusions we need to also employ visualization techniques.

It is much easier to discover and confirm the presence (or even absence) of patterns, relationships, and physical characteristics (such as outliers) through a visual display, reinforcing the essence of Tukey's quote about the value of pictures.

## The bedrock of visualization knowledge

Data visualization is not easy. Let's make that clear from the start. It should be genuinely viewed as a craft. It is a unique convergence of many different skills and requires a great deal of practice and experience, which clearly demands time and patience.

This multi-disciplinary recipe unquestionably makes it a challenging subject to master but equally provides an exciting proposition for many. This is evidenced by the field's popular participation, drawing people from many diverse backgrounds.

If we look at this subject convergence at a more summary level, data visualization could be described as an intersection of art and science. This combination of creative and scientific perspectives represents a delicate mixture. Achieving an appropriate balance between these contrasting ingredients is one of the fundamental factors that will determine the success or failure of a designer's work.

The *art* side of the field refers to the scope for unleashing design flair and encouraging innovation, where you strive to design communications that appeal on an aesthetic level and then survive in the mind on an emotional one. Some of the modern-day creative output from across the field is extraordinary and we'll see a few examples of this throughout the chapters ahead.

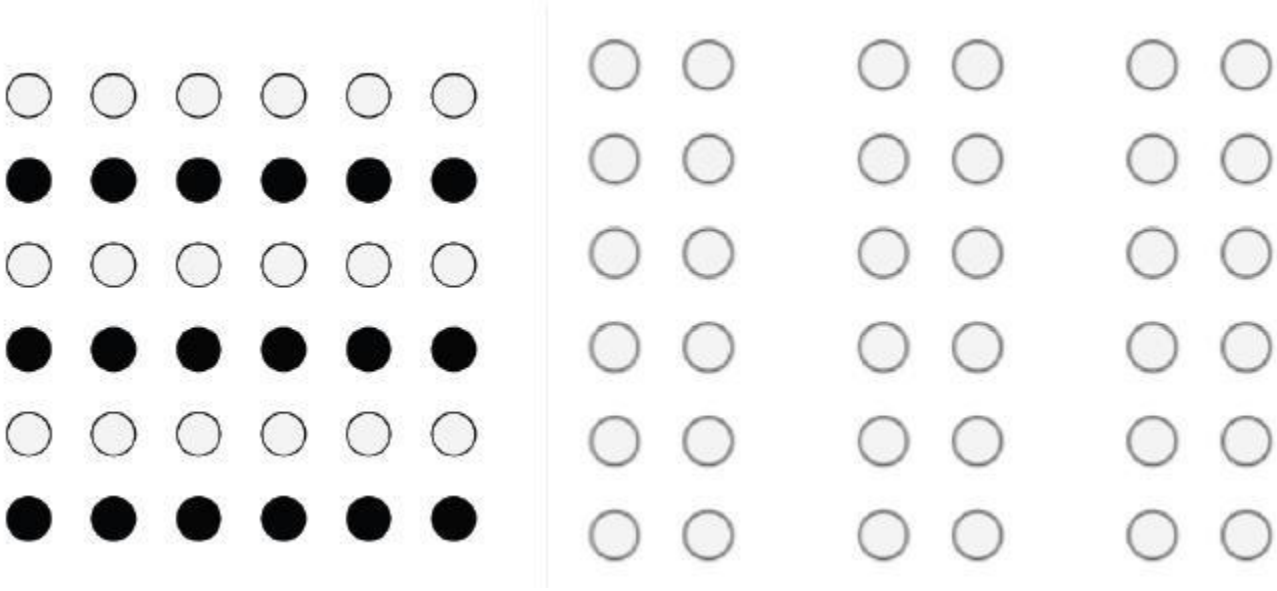
The *science* behind visualization comes in many shapes. I've already mentioned the presence of computer science, mathematics, and statistics, but one of the key foundations of the subject comes through an understanding of cognitive science and in particular the study of visual perception. This concerns how the functions of the eye and the brain work together to process information as visual signals.

One of the other most influential founding studies about visual perception emerged from the Gestalt School of Psychology in the early 1900s, specifically in the shape of the Laws of Perceptual Organization (http://www.interaction-design.org/encyclopedia/data_visualization_for_human_perception.html).

These laws provide an organized understanding about the different ways our eyes and brain inherently and automatically form a global sense of patterns based on the arrangement and physical attributes of individual elements.

Here, we can see two visual examples of Gestalt Laws.

On the left-hand side is a demonstration of the "Law of Similarity". This shows a series of rows with differently shaded circles. When we see this our visual processes instantly determine that the similarly shaded circles are related and part of a group that is separate and different to the non-shaded rows. We don't need to think about this and wait to form such a conclusion; it is a preattentive reaction.



*Images republished from the freely licensed media file repository Wikimedia Commons, source: http://en.wikipedia.org/wiki/File:Gestalt_similarity.svg and http://en.wikipedia.org/wiki/File:Gestalt_proximity.svg*

On the right-hand side is a demonstration of the "Law of Proximity". The arrangement of closely packed-together pairs of columns means we assume these to be related and distinct from the other pairings. We don't really view this display as six columns, rather we view them as three clusters or sets.

At the root of visual perception knowledge is the understanding that our visual functions are extremely fast and efficient processes whereas our cognitive processes, the act of thinking, is much slower and less efficient. How we exploit these attributes in visualization has a significant impact on how effectively the design will aid interpretation.

Consider the following examples, both portraying analysis of the placement of penalties taken by soccer players.

When we look at the first image, the clarity of the display allows us to instantly identify the football symbols, their position, and their classifying color. We don't need to think about how to interpret it, we just do. Our thoughts, instead, are focused on the consequence of this information: what do these patterns and insights mean to us? If you're a goalkeeper, you'll be learning that, in general, the penalty taker tends to place their shots to the right of the goal.

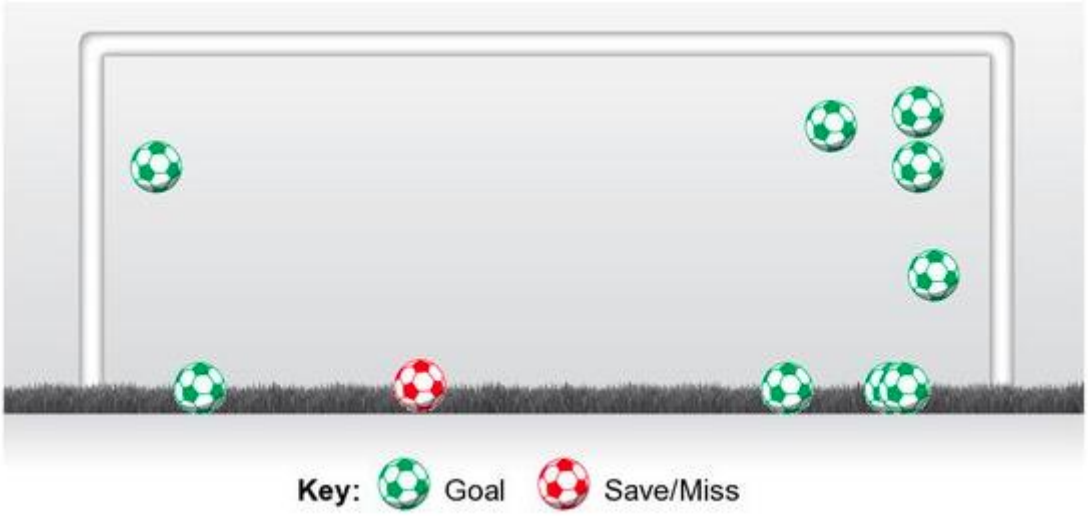



*Image republished under the terms of "fair use", source: http://www.facebook.com/castrolfootball*

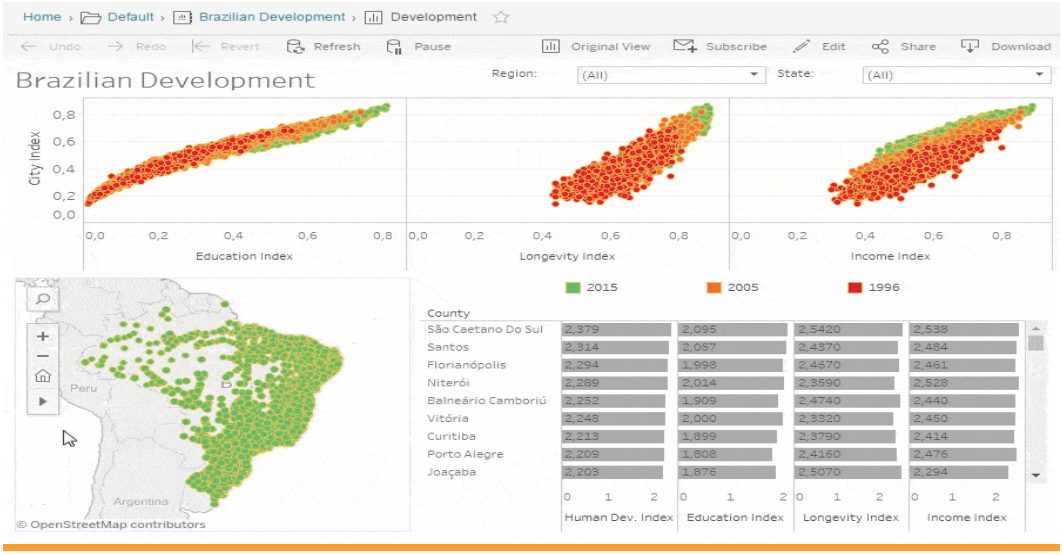
By contrast, this second display's attempt to portray the same type of data presentation causes significant visual clutter and confusion. Rather than using a simple and relatively blank image like the previous one, this display includes strong colors and imagery in the background. The result is that our eyes and brain have to work much harder to spot the footballs and their colors because the data layer has to compete for attention with the background imagery. We are therefore unable to rely on the capabilities of our preattentive visual perception (determined by the Law of Similarity) because we cannot easily perceive the shapes and their attributes representing the data. This delays our interpretative processes considerably and undermines the effectiveness and efficiency of the communication exchange.


---

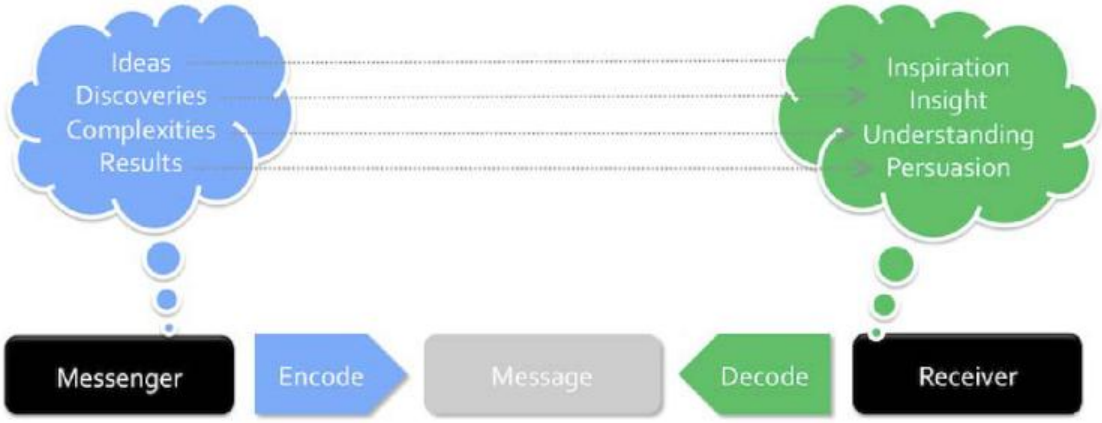
# Defining data visualization

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

In the world of Big Data, data visualization tools and technologies are essential to analyze massive amounts of information and make data-driven decisions.



It is important now to consider a definition of data visualization. To do this, we first need to consider the main agents involved in the exchange of information; namely, the messenger, the receiver, and the message. The relationship between these three is clearly very important, as this illustration explains:




On one side we have a messenger looking to impart results, analysis, and stories. This is the designer. On the other side, you have the receiver of the message. These are the readers or the users of your visualization. The message in the middle is the channel of communication. In our case this is the data visualization; a chart, an online interactive, a touch screen installation, or maybe an infographic in a newspaper. This is the form through which we communicate to the receiver.

The task for you as the designer is to put yourself in the shoes of the reader. Try to imagine, anticipate, and determine what they are going to be seeking from your message. What stories are they seeking? Is it just to learn something new or are they looking for persuasion, something with more emotional impact? This type of appreciation is what fundamentally shapes the best practices in visualization design: considering and respecting the needs of the reader.

The important point is this: to ensure that our message is conveyed in the most effective and efficient form, one that will serve the requirements of the

receiver, we need to make sure we design (or "encode") our message in a way that actively exploits how the receiver will most effectively interpret (or "decode") the message through their visual perception capabilities.

From this illustration we can form the following definition to clarify, at this early stage, what we mean by data visualization:

*The representation and presentation of data that exploits our visual perception abilities in order to amplify cognition.*

Let's take a closer look at the key elements of this definition to clarify its meaning; these are as follows:

- The **representation** of data is the way you decide to depict data through a choice of physical forms. Whether it is via a line, a bar, a circle, or any other visual variable, you are taking data as the raw material and creating a representation to best portray its attributes. We will cover this aspect of design much more in Chapter 4, *Conceiving and Reasoning Visualization Design Options* and Chapter 5, *Taxonomy of Data Visualization Methods*.
- The **presentation** of data goes beyond the representation of data and concerns how you integrate your data representation into the overall communicated work, including the choice of colors, annotations, and interactive features. Similarly, this will be covered in depth in Chapter 4,

- Exploiting our **visual perception abilities** relates to the scientific understanding of how our eyes and brains process information most effectively, as we've just discussed. This is about harnessing our abilities with spatial reasoning, pattern recognition, and big-picture thinking.
- **Amplify cognition** is about maximizing how efficiently and effectively we are able to process the information into thoughts, insights, and knowledge. Ultimately, the objective of data visualization should be to make a reader or users feel like they have become better informed about a subject.

## Visualization skills for the masses

visualization has become both a problem and an opportunity for the masses, which makes the importance and dissemination of effective practice a key imperative.

The quote from Stephen Few will resonate with many of you reading this. If you were to ask yourself "Why do I design visualizations in the way I do?", what would be your answer? Think about any chart or graphic you produce to communicate information to others. How do you design it? What factors do you take into account? Perhaps your response would fall in to one or more of the following:

- You have a certain design style based on personal taste
- You just play around until something emerges that you instinctively like the look of
- You trust software defaults and don't go beyond that in terms of modifying the design
- You have limited software capabilities, so you don't know how to modify a design
- You just do as the boss tells you—"can you do me some fancy charts?"

For many people, the idea of a conscious data visualization design technique isquite new. The absence of any formal coaching, at almost any level of education, in the techniques of visualization means until you become aware of the subject, you have probably never even thought about your visualization design approach.

Before discovering this subject, my own approach to presenting data was certainly not informed by any training or prior knowledge. I'd never even thought about it. Taste and gut-feel were my guiding principles alongside a perceived need to show off technical competencies in tools like Excel. Indeed, I'd like to take this opportunity to apologize for much of my graphical output between 1995 and 2005 where striking gradients and "impressive" 3D were commonplace. The thing is, as I've just said, I didn't realize there was a better way; it simply wasn't on my radar.

In some respects, the reliance on instinct, playing about with solutions that seem to work fine for us, can suffice for most of our needs. However, these days, you often hear the desire being expressed to move beyond devices like the bar chart and find different creative ways to communicate data.

While it is a perfectly understandable desire, just aiming for something different (or even worse, something "cool") is not a good enough motive in itself.

If we want to optimize the way we approach a data visualization design, whether it be a small, simple chart or a complicated interactive graphic, we need to be better equipped with the necessary knowledge and appreciation of the many design and analytical decisions we need to make.

As suggested previously, instinct and taste have got us so far but to move on to a whole new level of effectiveness, we need to understand the key design concepts and learn about the creative process. This is where the importance of a methodology comes in.

---

# The data visualization methodology

The design methodology described in this book is intended to be portable to any visualization challenge. It presents a sequence of important analytical and design tasks and decisions that need to be handled effectively.

As any fellow student of Operational Research (the "Science of Better") will testify, through planning and preparation, and the development and deployment of strategy, complex problems can be overcome with greater efficiency, effectiveness, and elegance. Data visualization is no different.

Adopting this methodology is about recognizing the key stages, considerations, and tactics that will help you navigate smoothly through your visualization project.

# STEP-BY-STEP-DATA VISUALIZATION PROCESS

## COLLECTING DATA

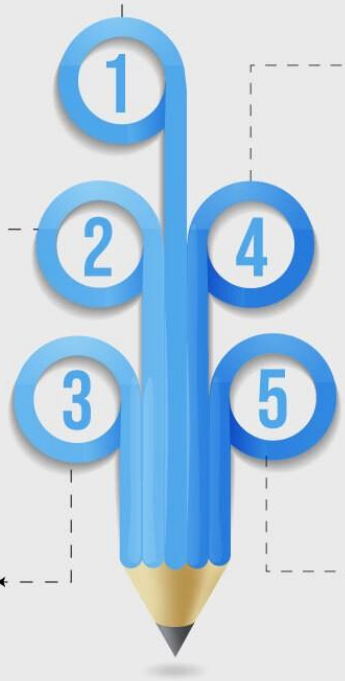
The first and most important step of data visualization is to gather data in large amounts only then we can apply data visualization techniques on the collected data and we will get some helpful insights from it

## CLEAN YOUR DATA

The output received from a data cleaned process is a dataset that is free of errors, missing data, anomalies, etc

## CHOOSE A CHART TYPE

Before choosing a visual chart or graph, it is important to understand your audience or the collected data's domain and then choose a chart or graph accordingly which will best communicate the message

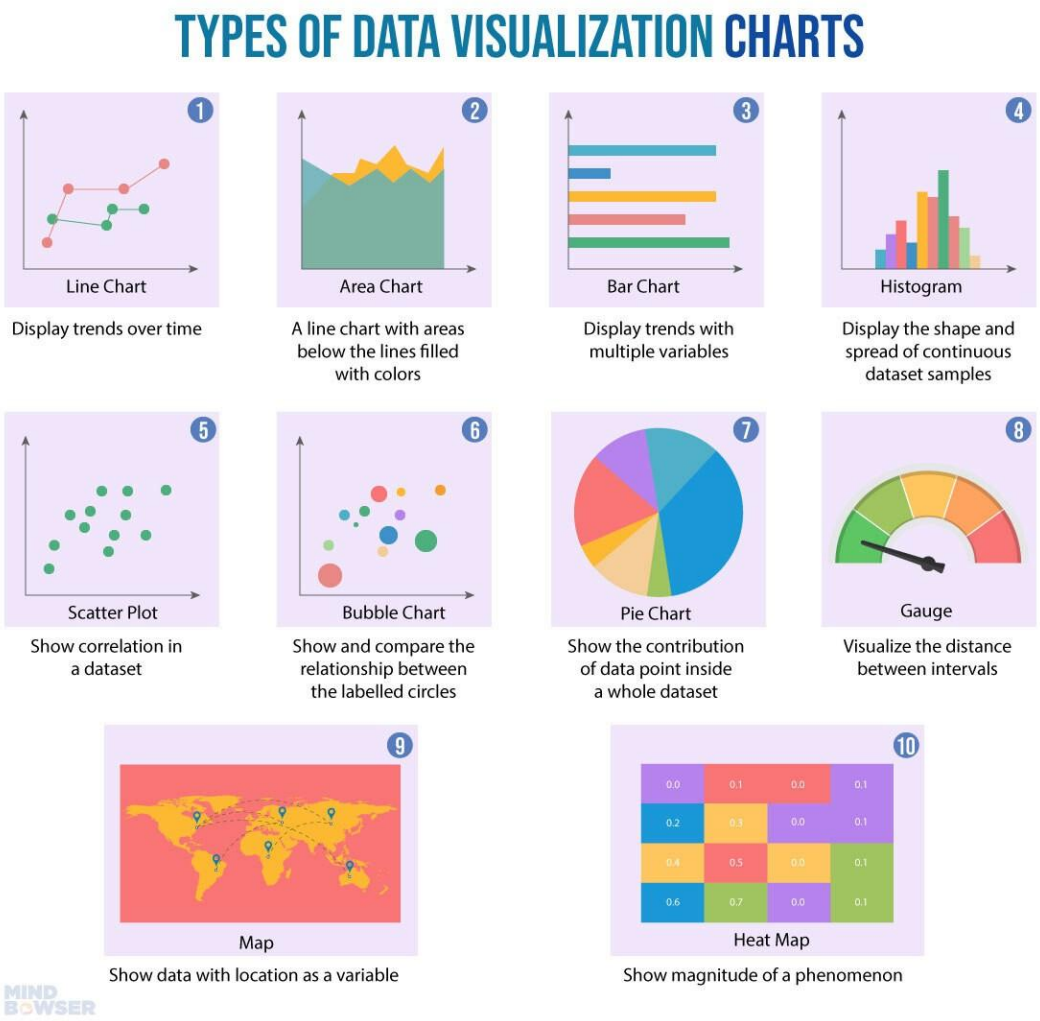



## PREPARE DATA

Data preparation tasks include finding data columns that help to make some decisions out of it and which gives some meaningful insights about data, grouping of data creating aggregate values for groups, combining variables to create new columns etc

## VISUALIZE DATA

In the final step you'll have the required data you need to create visualizations. Now you can apply all your visualizations skills on the prepared data and represent the data in charts or graphs with meaningful insights




Remember, though, design is rarely a neat, linear process and indeed some of the stages may occasionally switch in sequence and require iteration. It is natural that new factors can emerge at any stage and influence alternative solutions, so it is important to be open-minded and flexible. Things might need to be revisited, decisions reversed, and directions changed. What we are trying to do, where possible, is find the best path through the minefield of design choices.

Some may feel uncomfortable at the prospect of following a process to undertake what is fundamentally an iterative, creative design process. But I would argue everyone should find value from working in a more organized and sequenced way especially if it helps to reduce inefficiency and wasted resource.

The design challenges involved in data visualization are predominantly technology related; the creation and execution of a visualization design will typically require the assistance of a variety of applications and programs. However, the focus of this methodology is intended to be technology-neutral, placing an emphasis on the concepting, reasoning, and decision-making.

The variety, evolution, and generally fragmented nature of software in this field (there is no single tool that can do everything) highlights the extra importance of reasoned decision-making, regardless of the richness and power individual solutions can offer.

Another key point to remark on is to emphasize, if it wasn't already clear, that data visualization is not an exact science. There is rarely, if ever, a single right answer or single best solution. It is much more about using heuristic methods to determine the most satisfactory solutions.

On that note, the content of the methodology intentionally avoids any sense of dogmatic instruction, preferring to focus on guidelines over explicit rules; sometimes an ounce of chaos, a certain license to experiment, a leaning on instinct, and a sense of randomness can spark greater creativity and serendipitous discovery.

The methodology is intended to be adopted flexibly, based on your own judgment and discretion, by simply laying out all the important things you need to take into account and proposing some potential solutions for different scenarios.

Finally, as I stressed with my definition of the subject earlier, I'm not suggesting this is a ground-breaking new take on the creative process. It is merely a personal interpretation based on experience and also exposure to the many brilliant people out there who share their own design narratives. It is, though, consistent with how most established observers of the subject would recommend you undertake this task. Moreover, it is an approach that I fundamentally believe works and it has genuinely helped me improve my own work since I've adopted it more deliberately, allowing me to cut through projects with the efficiency and elegance I've always yearned for.

# Visualization design objectives

Before we launch in to the first stages of the methodology in Chapter 2, *Setting the Purpose and Identifying Key Factors*, it is important to acknowledge a handful of key, overriding design objectives that should provide you with a framework to test your progress and the suitability of your design decisions.

Whereas the methodology will introduce a number of key thoughts and decisions at each stage of the process, these objectives transcend any individual step and highlight the intricate issues you have to handle throughout your process.

The key objectives are as follows:

1. **Strive for form and function**
2. **Creating accessibility through intuitive design**
3. **Never deceive the receiver**

## Strive for form and function

The following is a quote from Frank Lloyd Wright:

*"Form follows function—that has been misunderstood. Form and function should be one, joined in a spiritual union."*

The first objective brings us immediately face-to-face with the age-old debate of form versus function or style over substance. As Frank Lloyd Wright proposed, all the way back in 1908, these are aspects of design that should be combined and brought together in harmony, not at the sacrifice of one or the other. There's room and a need for both.

It is a very difficult balancing act to achieve, as I've already alluded to in the discussion about art and science, but our aim should be to hit that sweet-spot where something is aesthetically inviting and functionally effective.

The designer and author Don Norman (http://www.jnd.org/dn.mss/emotion_design.html) talks about how we're more tolerant about things that are attractive and more likely to want them to perform well. Indeed, there is a school of thought that suggests how we think cannot be separated from how we feel.

Norman goes on to describe how well-executed aesthetics can naturally create favorable emotional and mental responses, but emotional affection can also come from the experience of good usability and the accomplishment of insight. Fundamentally, attractive form enhances function and the function portrays beauty through its effect.

Throughout this book, we will see examples of designs that have succeeded in creating elegance in form and in function. The following image is taken from an animated wind map developed by Fernanda Viégas and Martin Wattenberg. It is a beautiful piece of work, exceptionally well designed and executed but it also serves its purpose as a way of informing users about the wind patterns, strength, and directions occurring across the United States. This is form and function in spiritual union:

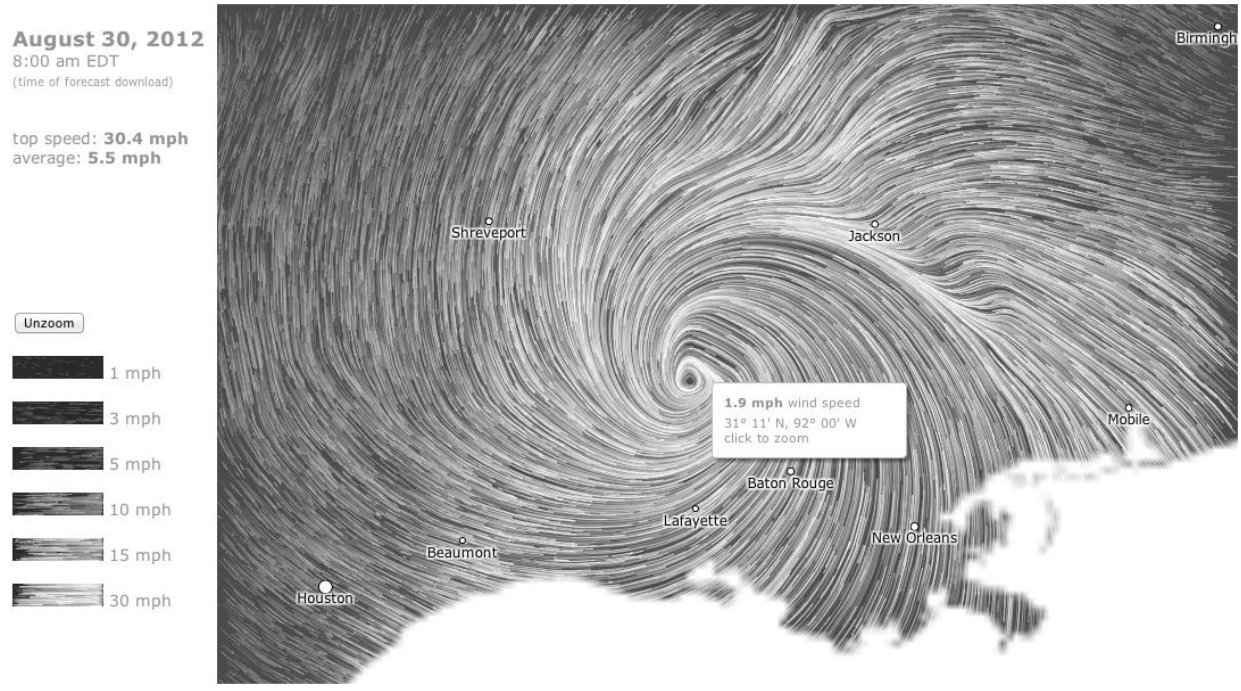



*Image from "Wind Map" (http://hint.fm/wind/) created by Fernanda Viégas and Martin Wattenberg*

The general advice, especially for beginners, is to initially focus on securing the functional aspects of your visualization. First, try to achieve the foundation of something that informs—that functions—before exploring the ways of enhancing its

form. The simplest analogy would be build the house before decorating it, but I wouldn't want to create too much separation between the two as they are often intrinsically linked. Over time, you will be much more confident and capable of synthesizing the two demands in harmony. We shall discuss this in more depth in Chapter 4, *Conceiving and Reasoning Visualization Design Options*.

In many ways, the central idea behind the methodology is encouraging you to determine that everything you do is thoroughly planned, understood, and reasoned.

This particular objective is about recognizing and responding to the scoping information that you will gather at the start of the methodology, to ensure that everything undertaken thereafter serves the purpose of our work and the needs of the audience.

Here, we should consider the idea of deliberate design, which means that the inclusion, exclusion, and execution of every single mark, characteristic, and design feature is done for a reason.

When we reach the stage of designing, concepting, and construction, you should be prepared to challenge everything; the use of a shape, the selection of a color pallet, the position of a label, or the use of an interaction.

In this next example, when displaying a section of a tree-hierarchy work by data illustrator, Stefanie Posavec, every visible property presented is used to communicate data, whether it be the use of color, the arc lengths of the petals, the position and sequence of stems; nothing is redundant and everything is deliberate.


Image from "Literary Organism" (`http://itsbeenreal.co.uk/index.php?/wwwords/literary-organism/`), created by Stefanie Posavec

It is also important to make sure that any visual property that is included, but does not represent data, such as shading, labels, colors, and axes among other properties, should only be included to aid the process of visual perception, not hinder it.

Furthermore, for interactive and animated visualizations, remember Amanda Cox's quote—"just because you can, doesn't mean you should." Don't succumb to the belief (like I did for many years) of thinking a visualization is a platform solely to showcase your technical competence.

Cluttering visualizations with fancy interactive features is a trap that is easy to fall into and leads to projects that look nice or are impressive technically but fail to serve their intended purpose. Instead, they interfere with the efficiency and effectiveness of the information exchange thus demonstrating a failure to synthesize form and function.

## Creating accessibility through intuitive design

When you next happen to be in a town or city center, take a look around you and observe how often people are confused by and struggle with the basic operation of correctly opening and entering doors into a store. Notice how the accessibility and function of a door—the simple act of opening and walking through it—is often impaired through a lack of intuitive design.

The method of opening a door should be straightforward, but often the aesthetics of features such as stylish door handles means we pull when we should push and we push when we should pull. This is a flaw in the intuitiveness and logic of the design, a failure in perceived affordance—it doesn't do what it looks like it should do.

This idea is an important concept to translate into visualization. As we have already outlined, we are trying to exploit the inherent spatial reasoning and pattern recognition functions of visual perception. We don't want people to have to spend unnecessary time thinking about how to use or how to read and interpret something.

When you are creating a visualization, you are integrating visual design with a subject matter's data. The former is the window into the latter, and it is the design and execution of this window that creates the accessibility.

But it is important to create a distinction between accessibility and immediacy. The speed with which you are able to read or interpret a visualization should be determined by the complexity of the subject and the purpose of the project, not by the ineffectiveness of design.

---

Sometimes subjects are fundamentally simple and the portrayal of the data is straightforward and intuitive. This in turn means the reader's task of interpreting the data should be relatively easy.

On other occasions, a data framework might be more complex. Your challenge will be to respect the complexity and avoid simplifying, diluting, or reducing the essence of this subject. This might mean something is not immediately easy to interpret. Some visualizations will require effort to be put in, forcing the reader to undertake a certain amount of experiential practice in order for the eye and mind to essentially become trained in reading the display.

Think of it being like muscle memory, but for the eye and the brain. We are so used to reading bar charts and line charts that they have become entrenched and programmed into our interpretative toolkit. But when we are faced with something new, something different or seemingly complex, its not always immediately clear how we are supposed to handle it.

In the following example, we see a demonstration of what is quite a complex data framework. This is an image of a legend that was used to explain how to read an innovative visualization to portray three separate indicators of a movie's success. On the left-hand side of the image is the aggregate reviews (the higher the value, the better) and on the right-hand side of the image are both the budget and gross takings (the bigger the gap, the better):

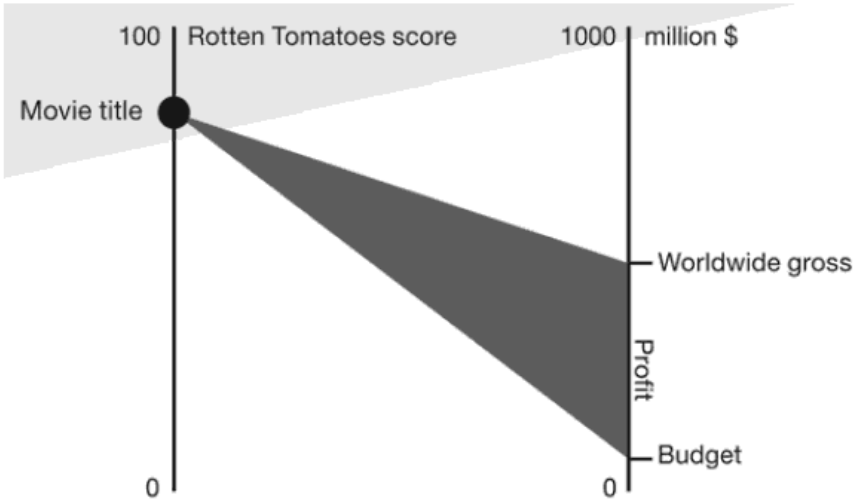



Image from "Spotlight on Profitability" (`http://www.szucskrisztina.hu`),

It is an unusual representation of data, not something as preprogrammed as the bar or line chart, and so it takes a short while to learn how to read and interpret the resulting shapes formed by the movie data shown across piece. This is absolutely legitimate as an effective approach to visualizing this data so long as the efforts that go into learning how to read it eventually leads the user to understand it.

Take another example, which portrays the key events in a couple of soccer matches showing completed passes (green lines), shots (blue triangles), and goals (red dots) as shown in the following image:

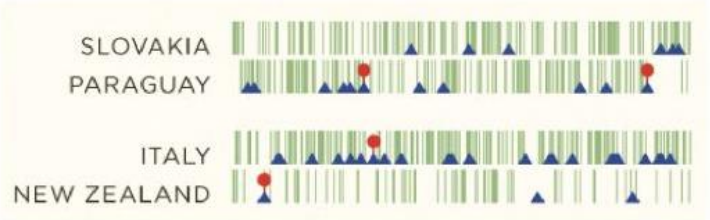



Image from "Umbro World Cup Poster" (`http://www.mikemake.com/Umbro-s-World-Cup-Poster`), created by Michael Deal

Once the reader has mastered the understanding of what each shape and its position means, these displays provide a powerful and rewarding insight in to the key incidents and the general ebb and flow of each game.

In simple terms, so long as you can avoid all the negative characteristics that Edward Tufte mentions at the top of this section, you should succeed in giving people an accessible route in to the data. Make sure that the efforts needed from the reader or user to understand how to use and interpret a visualization are ultimately rewarded

# Never deceive the receiver

Visualization ethics relates to the potential deception that can be created, intentionally or otherwise, from an ineffective and inappropriate representation of data. Sometimes it can be through a simple lack of understanding of visual perception.

In the following diagram, we see a 2D pie chart and a 3D version. When the eye interprets a graphic like this, what it is actually doing is perceiving the proportion of visible pixels:

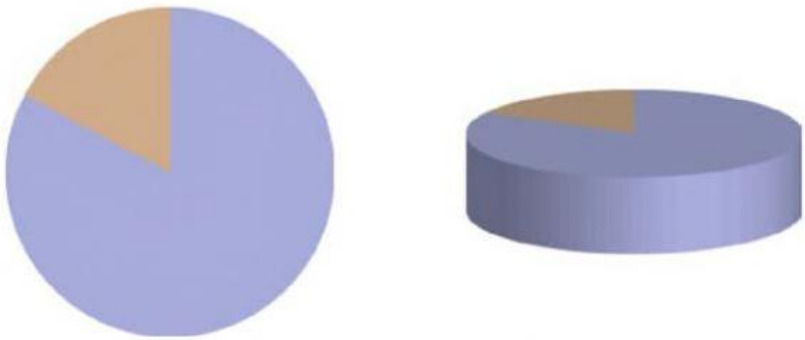


Image from "The Curious Incident of Kevins in Zurich...and other stories" (`http://www.researchobservatories.org.uk/EasysiteWeb/getresource.axd?AssetID=38334`) by Alan Smith.

On the left-hand side of the diagram, we see a blue segment representing 82 percent and an orange segment representing 18 percent. These are the actual values. However, when we introduce a third dimension on the right—incidentally, a dimension which is purely decorative and has no relationship with data values—our eyes are deceived because we are not capable of easily adjusting our interpretation of the values across this isometric projection. With the introduction of the extra dimension and the visible height of the pie itself, we now perceive 91 percent of the visible area as blue and only 9 percent orange. This is clearly a hugely distorted reading of the values.

Another similar example comes from a Wikipedia fundraising campaign from a few years ago and a progress bar depicting the status of their efforts; as shown in the following screenshot:


Image published under the terms of "Creative Commons Attribution-Share Alike", source: `https://donate.wikimedia.org/`

As with the pie chart, for a bar chart we perceive the visible pixels as being representative of the values. The label indicates a total of **$0.8M USD** had been raised (10.7 percent towards target) but if you calculate the actual length of the bar displayed, this occupies 24.6 percent of the overall bar length. Once again, a significant distortion of the truth.

This next example is a demonstration of where aesthetics and style completely hijack a visualization. Here, we have a still showing a 3D bar chart that swooshes impressively onto the screens of those watching soccer on TV in the UK:


But what have we here? There is a yellow **Drawn** bar representing the value **1** and this appears to be more than half the length of a red **Aston Villa** bar representing **4**. How can that be?

The designers of this visual have chosen to include the category labels within the bar's length, thus completely distorting the values being represented. Now, this is possibly one of the least interesting statistics you'll come across, and I'm assured the world will not stop turning as a result of this graphical misdemeanor, but it should demonstrate the pitfalls of decoration and overly stylized design.

Obeying visualization ethics is clearly an objective for any project, but really it is just about basic, good practice, respect for your readers, and attention to detail.

---

# Purpose of DV PROJECT :

## Clarifying the purpose of your project

We start at the very beginning. Why are we doing this project? What is its purpose?

These might seem like blunt questions to ask but it is so important to establish this type of clarity before we go much further. You might think: "please can we just get on with it?" and you're probably itching to get on to a computer and start playing with some data, but these initial few stages of this methodology are very deliberately designed to get you in to the habit of this type of reflective or analytical assessment.

If you look at a dictionary definition for "purpose", it will usually say something similar to: reason for existing, intended effect. These two dimensions neatly capture the focus of our definition work at this point in the journey.

## The reason for existing

Let's start with the reason for existing. This is about recognizing the trigger behind the project or the origin from where it emerged. This gives us an idea of the scope and context of what we are about to undertake, how much creative control we might have, whether we've been encouraged to follow a particular creative direction and what ideas have already been formed.

A project will typically form in one of following two ways: you've either been asked to do it or you've decided to do something yourself. You might think that's obvious, but these are very different scenarios for working creatively.

If it's the former, a project that's essentially been commissioned, you will have most likely been given the task by a colleague or a manager or from a client you are working with/for. A further source may be something like an invited assignment such as tendering for some work or even entering a design contest.

You will have received or read a brief and possibly had some initial discussions that provided you with an outline of the requirements. You might have some instructions and a general idea of what they are seeking.

From gathering this contextual information, you should have a reasonably clear idea about the background to the project, what you're being asked to do, why you're doing it and who you're doing it for. It may be quite loose and open-ended, in which case you've got a greater range of possibilities, but it's more likely to be quite defined and specific.

You've probably also experienced the pain of hearing some of the initial ideas flying round, as your creative soul dies a slow death in response to requests such as "cool charts" and "make it fancy" or "I want an Edward Tufte style piece".

# The intended effect

Whatever the motivation and background for doing the project, you will inevitably start to form a vision in your mind of what you might be about to create, how it might look, and what it might do. This is a natural instinct as you embark on a creative process.

This vision might leap into your mind the minute you start to think around the task, regardless of its origin. You might recall certain influential or inspirational works that you've seen in the past or remember concepts you previously developed that went no further.

It's important to capture these thoughts if they do form. Make sure you keep notes, in your sketchbook, on your tablet, or on a cigarette packet—it doesn't matter where, just do it before you forget. While we don't want to be closed off and commit ourselves to the pursuit of the first thing we think of, these instinctive thoughts could prove valuable later on.


If we look closely at the verbs and the overall language being used, we can start to recognize quite a range of differing effects that might be sought.

For example, a visualization to assist with the monitoring of signals or facilitating a visual lookup of data will be very different from a design that is intended to grab attention or change behavior. Similarly, presenting arguments and telling a story is a very different setting to conducting analysis or 'playing' with data.

What we have here is evidence of different dimensions of intent. Identifying your intended effect means deciding what you're aiming to achieve and how you're going to achieve it.

---

# VISUALIZATION FUNCTION AND TONE:

:

## Establishing intent – the visualization's function

The intended function of a data visualization concerns the functional experience you create between your design, the data, and the reader/user.

If we revisit the range of phrases presented earlier, it is possible to form three separate clusters or categories of function. While there is always a chance of slight overlap, there will be a significant difference in your design choices depending on whether the function of your visualization is to:

- Convey an **explanatory** portrayal of data to a reader
- Provide an interface to data in order to facilitate visual **exploration**
- Use data as an **exhibition** of self-expression

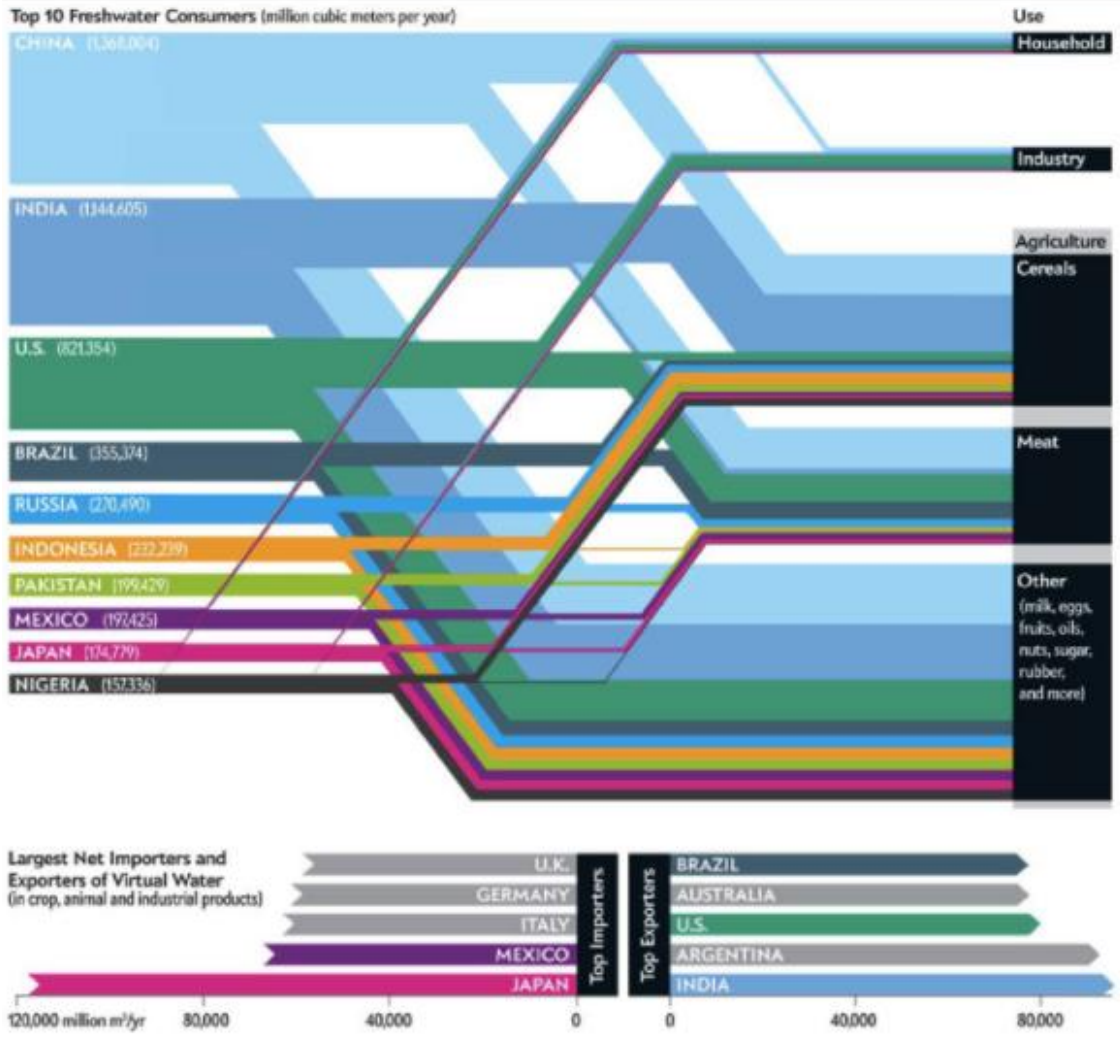
## When the function is to explain

Explanatory data visualization is about conveying information to a reader in a way that is based around a specific and focused narrative. It requires a designer-driven, editorial approach to synthesize the requirements of your target audience with the key insights and most important analytical dimensions you are wishing to convey.

There are many ways in which you can "explain" data. It could be through an information dashboard in a corporate setting, where you are conveying the latest performance figures and highlighting the key issues requiring attention. It might be a graphic in a newspaper, explaining the complexity and severity of the problems around the economic crisis. It could be an animated design to display patterns of population migration over time. It could also be a physical or ambient visualization designed to draw attention to the sugar content of certain drinks.

The end result is typically a visual experience built around a carefully constructed narrative. Your objective as the designer is to create a graphical display, made accessible through intuitive, visual design that clearly portrays the narrative you are seeking to impart.

Here is an example of an explanatory visualization, based on a chart type called a **Sankey** diagram, which portrays analysis of the top ten freshwater-consuming countries and the breakdown of its usage:




e from "Top 10 Freshwater Consumers" (`http://www.scientificamerican.com/article.`

# When the function is to explore

Exploratory data visualization design is a slightly different matter compared to creating an explanatory piece. Here, we are seeking to facilitate the familiarization and reasoning of data through a range of user-driven experiences. In contrast to explanatory-based functions, exploratory data visualizations lack a specific, single narrative. They are more about visual analysis than just the visual presentation of data.

Exploratory solutions aim to create a tool, providing the user with an interface to visually explore the data. Through this they can seek out personal discoveries, patterns, and relationships, thereby triggering and iterating curiosities. It also opens up the possibility for chance or serendipitous findings caused by forming different combinations of variable displays.

Really, the key feature that differentiates an exploratory piece from an explanatory piece is the amount of work you have to do as a reader to discover insights. For explanatory pieces, the designer should do the hard work and create a clear portrayal of the interesting stories and analysis from a dataset. An exploratory piece will be more about the readers doing the analysis themselves, putting the effort in to discover things that strike them as being significant or interesting.

In the following image, we see a **scatterplot matrix** visualization: a method used to reveal correlations across a multivariate dataset, enabling the eye to efficiently scan the entire matrix to quickly identify variable pairings with strong or weak relationships. This is a perfect example of an exploratory visualization design:

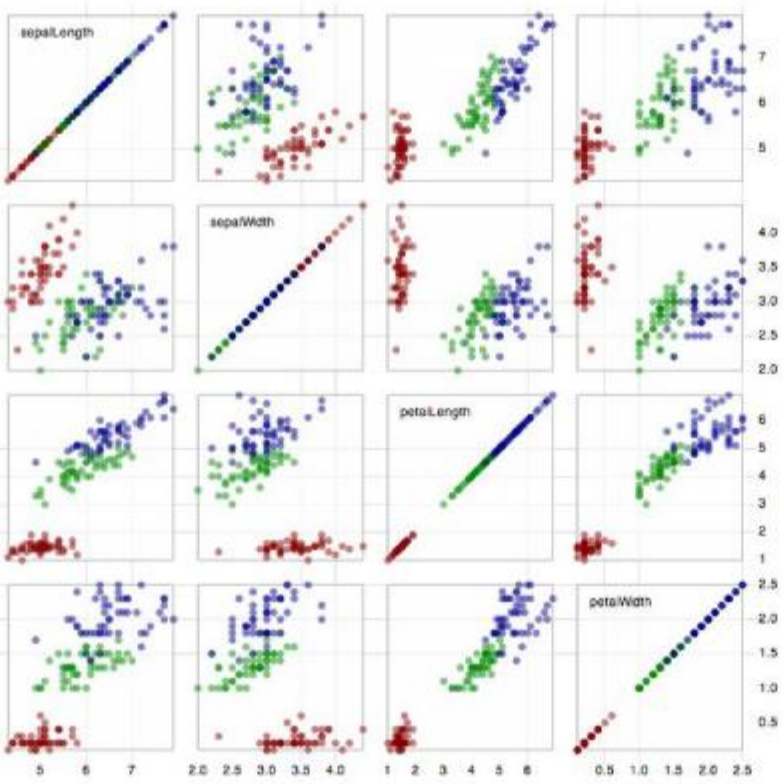



Image from "Scatterplot Matrix" (`http://mbostock.github.com/d3/ex/splom.html`), created by Mike Bostock.

Exploratory visualizations are not limited to being interactive. Visual analysis can be facilitated through static portrayals of data. The previous example is actually interactive but a static version would still offer a discovery of the relationships and patterns of the dataset.

# When the function is to exhibit data

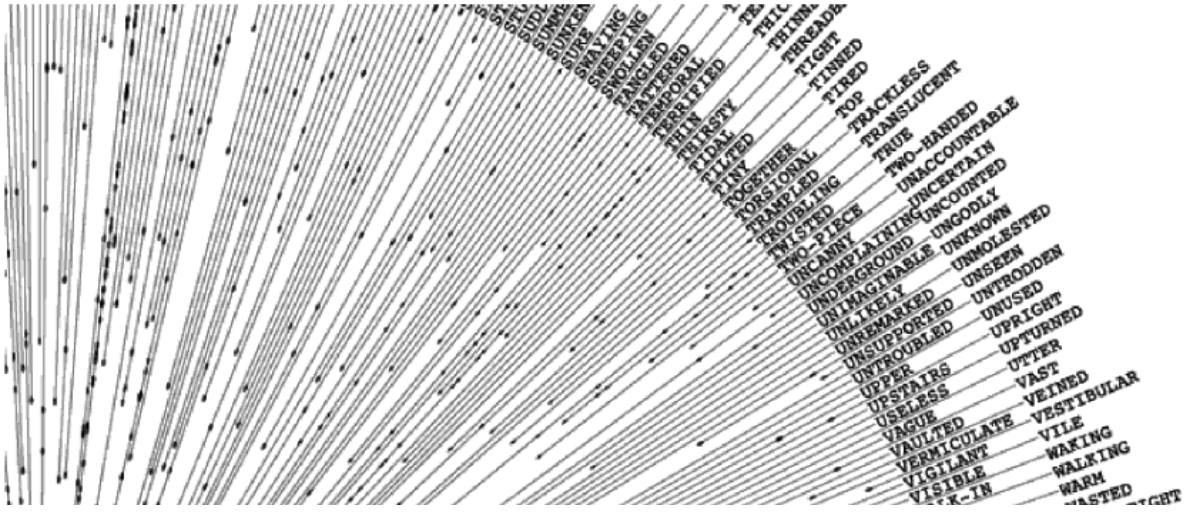
The final classification of intended function is in some respects a controversial one, because including the exhibiting of data as an intended function of visualization will not be consistent with many people's definition of data visualization.

We're not talking controversy on the level of a political scandal. Let's put it into some perspective: within the context of this field, this is a big deal and the cause of so much debate.

As with any attempt at classification, there is a spectrum of variety within and so clear boundaries are difficult to establish and very much open to personal interpretation.

We are talking here about designs that use data as the raw material, but where the intention is perhaps somewhat removed from a pure desire to inform. Rather, the objective is closer to a form of exhibition or self-expression through data representation. This genre of work embodies the term "data art".

In the following example, we see an example of data art (as defined by the creator himself) that visualizes all the adjectives used in Cormac McCarthy's book *The Road*. The adjectives are arranged radially in alphabetical order and each line represents a timeline of the book, beginning at the perimeter:



# Establishing intent – the visualization's tone

Earlier, we looked at a collection of different phrases that articulate realistic intentions for creating a data visualization. We've just proposed three classifications for how you might organize these terms according to their function.

There will sometimes be an overlap between these otherwise distinct categories, but you should be able to determine relatively clearly where your work intends to fit on the scale between explanation, exploration, and exhibition.

Setting the function is just one part of the "intent" equation. The clarity of your potential design pathway will be much more apparent as we now consider the second dimension of intent—tone.

Establishing a suitable tone goes beyond function and more towards the style of the design experience. It concerns the type of stimulus or desired emotional response that you are trying to create. It is therefore important for you, as the designer, to be able to reason what sort of design will achieve that tone.

With this type of judgment to make, you will inevitably find yourself juggling creative and scientific perspectives. This dynamic poses a significant challenge for any data visualization designer to reason and resolve. Indeed, the design objectives outlined at the end of the previous chapter highlight the intricate issues you have to handle.

However, sometimes there has to be mutual recognition that for different scenarios there might be good reason for leaning more towards one direction than the other.

Let's look at the language of two potential motives behind creating a data visualization:

- "We need a chart to help monitor..."
- "We need to present this in a way that persuades people..."

Here, we have two situations both aiming to better inform a reader or a user, but the intended effect or outcome from the experience will be different.

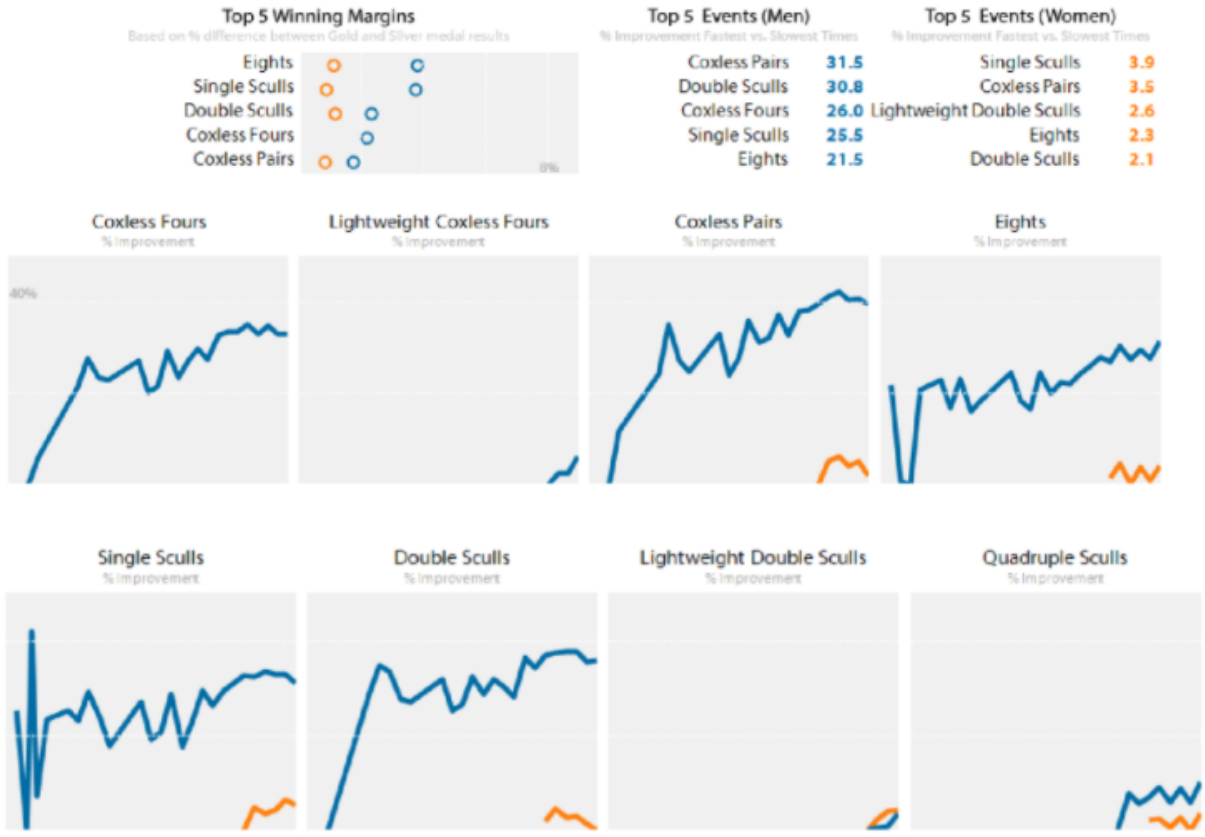
The reaction of a user reading, for example, a dashboard full of bar charts and line charts to help monitor monthly performance will be quite analytical and pragmatic in style. It is unlikely to involve or stir much emotion (unless things are suddenly and unexpectedly plummeting). The style of the visualization design will be consistent with the intended nature of this particular type of engagement, probably quite sober and with an emphasis on the precision of perception.

Compare that with the intended impact of a presentation that depicts how many lives could be saved if a charity was able to achieve a certain level of fundraising. The setting and intent will be more about persuasion making it emotionally charged. It will need to attempt to create an experience that is much more personal and more impactive.

Here, we see potentially two ends of a spectrum for judging the right tone. Yet they fundamentally share the overall motivation of wishing to inform people about a subject through the visual representation of data.

Pragmatic and analytical

Designs that fit this classification will often involve data being represented through the use of bar charts, line charts and dot plots, for example. Stylistically, they will be characterized by a rather clinical look-and-feel that is consistent with the next sample image, taken from a project analyzing Olympic results over the years:




Creating a visualization with a pragmatic tone is about recognizing a need for a design that delivers fast, efficient and precise portrayals of data. Typically, you will have a captive audience, a readership who want to or need to interact and learn from the data. This could be a corporate environment, where people need to simply learn about recent performance of operational activity or undertake visual analysis to discover potentially revealing patterns.

In these cases, there is no value placed on attempting to draw attention to the visualization, or trying to encourage somebody to read or interact with a graphic by employing aesthetic novelty. Furthermore, it's not about trying to inject any emotional or metaphorical connection with the data stories presented.

The purity and impact of its function—the satisfaction that comes from an efficient intake of understanding—fulfils the aesthetics of the charting methods deployed. Therein lies the elegance of pragmatic work.

Sometimes you just want to and need to move beyond bars, straight lines, and right angles and more towards curves, circles, and other bendy things. As we will see in *Chapter 4, Preparing and Familiarizing with Data*, there are consequences to this choice, in respect of the known reduction in the accuracy of value perception this will cause. That is a sacrifice you as the designer need to juggle and justify.

Abstract visualization, in terms of its tone, is more about creating an aesthetic that portrays a general story or sense of pattern. You might not be able to pick out every data point or category, but there is enough visual information to give you a feel for the physicality of the data.

This next image is taken from a project to visualize the global airline transportation network consisting of all commercial flights worldwide. The routes highlighted are those flights in and out of Toronto Pearson airport. The project was designed to assess the threat of infectious diseases.


Image from "Toronto Flight Lines" (`http://www.biodiaspora.com/`) created by Bio.Diaspora 2012

The design does not intend to offer an analytical summary of air travel statistics. Instead it creates a more immersive experience in to the data, offering a visual interface to establish a greater sense of how interconnected the world is through air travel. It causes us to imagine just how easy it could be for diseases to spread across the globe in a short period of time.

For more emotive visualizations, you might be seeking to generate a different type of emotional connection with the design. This connection comes both at the start of the engagement—creating attraction and appeal—as well as after the engagement—the outcome.

In contrast to pragmatic works, as we described in the introduction, here we might be seeking to achieve impact, to emphasize issues, and perhaps to engender shock. We might also seek to generate a certain amount of visual attention through novelty and innovation in a way that more pragmatic approaches would not be able to achieve.

In the following image, we see a section taken from a newspaper infographic that depicted Iraq's bloody toll. While the chart method is nothing more complex than an upside down bar chart, the tone is very impactive and metaphorically emphatic, creating a strong emotional impact to the story it portrays:

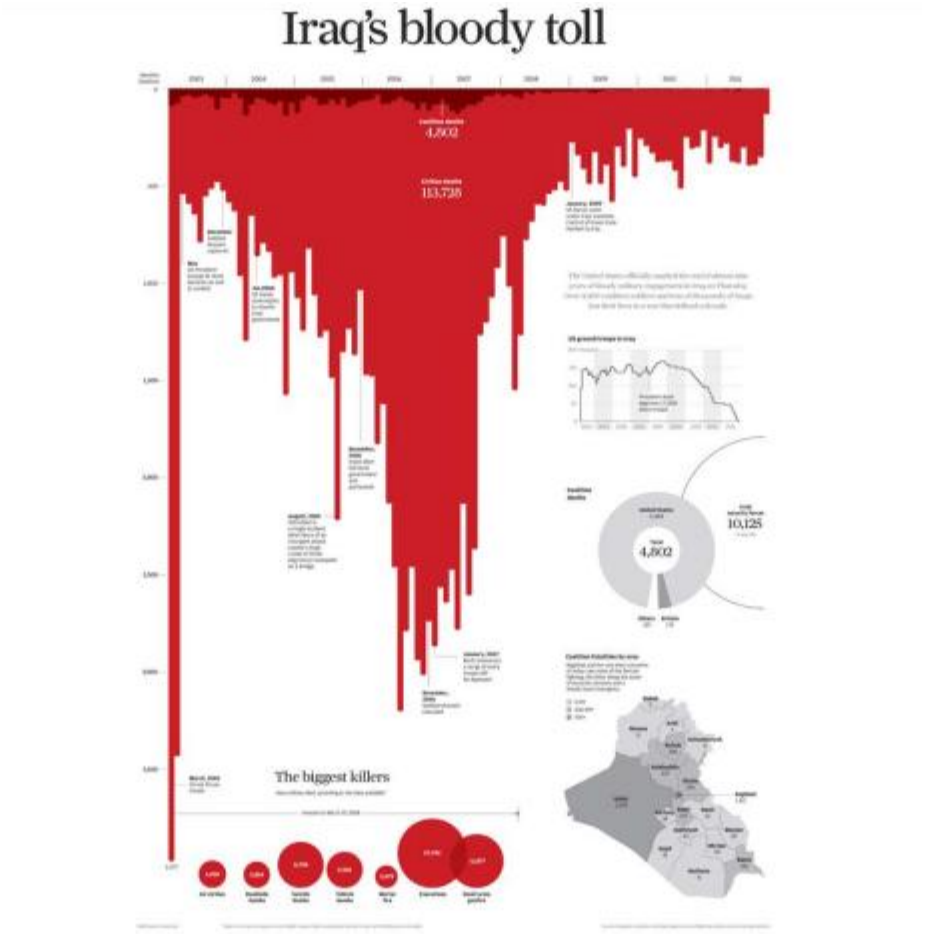



Image from "Iraq's Bloody Toll", published in the South China Morning Post on Saturday, December 13, 2011 (http://graphics-info.blogspot.hk/2012/09/malofiej-20-look-at-our-participation.html), created by Simon Scarr.

# Key factors surrounding a visualization project

The following is a quotation from Edward Tufte's book, *The Visual Display of Quantitative Information*:

> "*Most principles of design should be greeted with some skepticism... we may come to see only through the lenses of word authority rather than with our own eyes.*"

While establishing the purpose of the visualization project sets the desired tone of the design and its function, there are inevitably many other factors that will have a significant influence on the shape and direction of our visualization design.

It is especially important to identify and recognize the impact of the contextual conditions, within and around your project that will affect what you can and can't achieve and how you might achieve it.

Here are some of the most important factors to consider and to evaluate their potential impact:

- **The aim**: As we have seen already, there are different origins and triggers for a project. We mentioned the self-initiated ones as being almost free of external constraint and essentially framed by our own capabilities and intentions. The important thing worth reinforcing here is the need to take responsibility when a project involves a brief, commissioned by a client or a colleague. You must demonstrate excellent communication skills to ensure you seek and gather as much of an understanding as possible of what it is they are aiming to achieve. Sometimes, you might be provided with a very open brief because a client may not even know what it is they are seeking. In these situations, your responsibility needs to extend to assist them in the scoping and requirements of the work. On other occasions you will be asked to create something that goes against your general practice (for example, the subject matter or requested style) or it might even be simply impossible to deliver (perhaps due to the desired design or available resources). Here again your communication skills are going to be required to manage the expectations. It is easy to be shy and delay asking vital questions but this will only cause you pain later.

**Time pressures**: Common to just about every commissioned design project will be the pressure of time and deadlines. Most projects have clear timescales, from in-day quick turnaround pieces to longer-term grand projects. The challenge of maintaining objective creativity in the face of diminishing time is something that will severely test designers of all experiences. Whatever your situation, you have to use your time effectively and that's where value will come from following the tactics in this methodology. Plan your work and create a balanced layout of the things you need to accomplish, so that you avoid disproportionately spending time on tasks that are less important than others. Often you will find yourself undertaking a visualization project in parallel with many other commitments. Not only will your capacity be limited, the momentum and duration of your focus will be impacted. This is where project management skills come to the fore as well as a realistic appreciation of what you should and shouldn't commit to undertaking. It also highlights the importance of keeping notes so that you can move seamlessly between projects and not lose track of your thoughts, ideas, or progress.

**Costs**: The issue of financial resource will unquestionably emerge, especially for large-scale projects. Costs will significantly influence the time you are able to commit to a project, the scope for bringing in additional collaborators, and the range of tools or technical resources you might be able to utilize. Once again, the planning and preparation stages will be invaluable to surface all potential issues around financial matters.

- **Client pressures**: Aside from time pressures, you need to anticipate and reduce the impact of potential unexpected pressures and interruptions coming from your client or colleagues. This might be changes in requirements, new demands, interference in the design solution, and generally annoying things that get in the way of your progress. A further manifestation of the pressure that can come from clients is the insistence on observing organizational visual or brand identities, layout rules, editorial guidelines, and technical frameworks. All of these will shape the scope of your design choices. You have to be prepared for and capable of managing this relationship, and the mutual expectations, effectively so always be open with your client, keep them regularly updated with progress and, where applicable, involve them in the key decision moments throughout the process.
- **Format**: From a design perspective this will be a significant influencing aspect. Are you creating a static or an interactive design? Maybe it's a multifaceted project and you are looking to create both. If it is an interactive design, what platform do you need to achieve compatibility with? Will it be for the Web, a tablet, and/or smartphones? If it is a static design, will it be a small graphic in a publication, a full-page spread, or a large poster display? Maybe it will be a video animation or an ambient display out in the wild, or a large touch screen installation in a museum. This is a vital consideration that needs to be cleared up at the earliest possible stage. Another factor to take into account will be the likely frequency of the project—is it a one-off piece or will it be something that needs to be replicable and/or scalable? That could hugely affect what you can or can't deliver.

- **Technical capabilities**: Aside from your own technical capabilities, what are the technical resources to which you have access? For example, are you limited to free tools or can you access more premium software? Do you have the most appropriate technical infrastructure, such as server speed and capacity if it is an online project? Depending on your format choices, what frameworks are you going to deploy, what browsers do you need to have it working on, what backend database technologies are you going to require? This is a wide-ranging and very technical set of decisions that will likely require a specialist technician to determine.

# Conceiving and Reasoning Visualization Design Options

So far the focus of our attention has been on the uncelebrated but vital preparatory and scoping stages of the data visualization methodology. We have established the purpose of our design and the key factors surrounding the project have been identified and weighed-up. We have also acquired and prepared our data and begun exploring it to identify the key data stories and analytical slices around which we may form our editorial focus.

These contextualizing activities are often neglected because they are understandably deemed not as fun as the design stage. Yet, they will save you time and pain, helping your work to proceed more efficiently by avoiding blind alleys and creative misjudgments.

In this chapter, we will be taking a forensic look at the many design choices involved in the process of establishing an effective visualization solution. We will tackle these choices by working through the anatomy of a visualization design, separating our thoughts into the complementary dimensions of the representation and presentation of data.

For rookie and experienced designers alike, the framework of design considerations outlined here should help you successfully navigate through the creative challenges and opportunities you are facing.

## Data visualization design is all about choices

To frame this discussion, do a quick image search in Google for the term "data visualization" and scroll through the first few screens. You will see just a snapshot of evidence of the innumerable variety of ways in which you can represent data. Some

are good, some are bad. Some are really bad. Some shouldn't even be connected with the term data visualization.

Additionally, have a glance through the collection of submissions uploaded on to sites that run some of the main data visualization design contests (for example, www.visualizing.org, www.infobeautyawards.com). Choose a particular contest and explore the spectrum of proposed solutions, all typically emerging from the same dataset and responding to the same brief.

What can we learn from this? What does this evidence of the variety of ways in which people interpret visualization design challenges reveal to us?

The first thing to say is that there is never a single path towards a "best" solution. The inherent creativity and individualism of design work ensures that. An idealistic desire for a single and simple set of rules to achieve a guaranteed effective solution is simply unreasonable due to the many different factors that will shape the scope and intention of any given project.

There is, however, an established body of theoretical and practical evidence that guides us to understand which techniques work better for certain situations and less well for others. Importantly, these guides transcend instinct or personal taste and help us frame many of our design options, influencing the choices we make.

Beyond that it is more about managing trade-offs, about trusting your judgment to make sense of the problem context in which you are working, the requirements you are responding to, and keeping in mind the overall objectives of visualization design, as outlined in Chapter 1, *Context of Data Visualization*.

The second key observation is to remark that the very moment we take on a visualization challenge, and start our journey towards a design solution, we are commencing a unique creative route formed by numerous permutations of choices. Nobody else will go through the same experience nor arrive at exactly the same solution.

You won't always get there easily. That's important to recognize too. As Jer Thorp expressed in his quote, even the best make mistakes and end up wasting time following ideas that lead nowhere and having to change course halfway through. However, by following the approach we have outlined in this book, and specifically the framework of considerations for this chapter, we hope to reduce the waste and eliminate inefficiency. This allows us to fail faster and recover more quickly.

A useful way to look at a data visualization challenge is to recognize that we are actually seeking to reduce choices. This is achieved through recognizing influential factors, by considering the desired function and tone of our work, familiarizing with our data and identifying stories. We are building clarity through selection and rejection. We are reducing the problem by enhancing our clarity.

The reasoning involved in eliminating options is just as important a skill as determining those we shall pursue. This lets us control our work, it helps us plan better, and prepare for the creative avenues down which we may proceed.

In many ways you could equate this design process with the responsibilities of being a film director, managing the dramatic, artistic, and technical aspects of a film. A director has to create the film's vision, direct the cast, manage the crew, oversee the script, coordinate the choice of locations, the music, and the post-production effects. All these different perspectives require separate attention and unique treatments until they are brought together into a cohesive single product: a movie. We're trying a similar approach with our visualization design.

As we now move into the creative stage it is helpful to follow a framework that will help you to understand the many different design options about which you will have to make a decision.

An effective way to think about this is to consider the visualization "anatomy". By definition, anatomy refers to "the bodily structure of an organism", so we are appropriating the term to apply it to the structural layers of a data visualization design.

In the first chapter, you will recall the proposed definition of data visualization and how it separated the idea of representation and presentation of data. We see these as separate dimensions of our design task:

- **Data representation**: This is the foremost layer, how we give form to our data through the use of "visual variables" to construct chart or graph types.
- **Data presentation**: This is the delivery format, appearance, and synthesis of the entire design. It concerns the layers of color use, interactivity, annotation, and the arrangement of all elements

# The visualization anatomy – data representation

The process of identifying the most effective and appropriate solution for representing our data is unquestionably the most important feature of our visualization design. Working on this layer involves making decisions that cut across the artistic and scientific foundations of the field.

Here we find ourselves face-to-face with the demands of achieving that ideal harmony of form and function that was outlined in the objectives section of Chapter 1, *Context of Data Visualization*. We need to achieve the elegance of a design that aesthetically suits our intent and the functional behavior required to fulfill the effective imparting of information.

What we're doing here is determining how we are going to show what it is we want to say. It is a difficult skill to master—something of a dark art—particularly given the set of factors we need to consider and the trade-offs we might need to make. DATA representation task involves following:

- Choosing the correct visualization "method" for the stories we're telling
- Accommodating the physical properties of your data
- Facilitating the desired degree of precision
- Creating an appropriate metaphor to depict our subject stylistically

# The visualization anatomy – data presentation

The presentation of data involves thinking about pretty much every other design feature that might be included in our visualization. Here, we are determining the following:

- The use of color
- The potential of interactive features
- The explanatory annotation
- The architecture and arrangement

The decisions we make about these layers should be focused on delivering extra meaning, intuitiveness, and depth of insight to our readers or users.

One of the key concepts throughout our judgment of presentation-related design options is to seek to make the visible, invisible. In contrast to data representation, where our objective is to make the invisible stories and insights, visible, data presentation features should almost feel invisible so that the portrayal of the data maintains visual dominance. Therefore, try to bear the following two things in mind:

- **Visual inference means data inference**: If it looks like data, it should be data. If it isn't data then you've incorrectly conveyed a sense of representation where there isn't any and design refinement is required. An example might be the use of a color to represent a certain sentiment. If that color is used on a bar chart or is picked for the background of a label or call-out, but it is no longer connected to the representation of any sentiment meaning, this may trick the reader who has programmed their visual sense to spot this inference.
- **Facilitating the resemblance of data**: Let the data breathe. We talked about this in the discussion about Jacques Bertin's interpretative acts, but the presentation layers of your visualization will have a great impact on this. Ensuring a reader can discriminate between data categories and values is usually influenced by the background artifacts and surrounding apparatus. Throughout your design, make sure your data stands out clearly as the principle visual component.

# DATA VISUALIZATION TOOLS

### 1. Tableau

Tableau is a data visualization tool that can be used to create interactive graphs, charts, and maps. It allows you to connect to different data sources and **create visualizations in minutes.**

Tableau Desktop is the original product. It's made for creating static visualizations that can be published on one or more web pages, but it doesn't create interactive maps.

Tableau Public is the free distribution of the Desktop product with some limitations.

Learning Tableau takes time and practice, but there are plenty of resources out there to help you learn how to use it. As a data scientist, Tableau has to be the number one tool for you to learn and use in your everyday tasks.

### 2. QlikView

QlikView is not just another data visualization tool, It is a **data discovery platform** that empowers the users to make faster, more informed decisions by accelerating analytics, revealing new business insights, and increasing the accuracy of results.

It has been an intuitive software development kit that has been used in organizations around the world for many years. It can combine various kinds of data sources with visualizations in color-coded tables, bar charts, line graphs, pie charts, and sliders.

It has been developed on a **"drag and drops" visualization interface,** allowing users to easily add data from many different sources, such as databases or spreadsheets, without having to write any code. These characteristics also make it a relatively easier tool to learn and grasp.

### 3. Microsoft Power BI

The Microsoft Power BI is the data visualization tool that is used for business intelligence type of data. It is and can be used for reporting, self-service analytics, and predictive analytics.

> According to Microsoft, there are more than 200K+ organizations that are using this tool for their Business Intelligence needs.

Furthermore, it provides an end-user platform to create reports and share insights with others in their organization. It **acts as a centralized repository** for all your business data which can be accessed by all your business users.

*On top of all this, Power BI also provides for integration with other SaaS products like Google Analytics, MailChimp, Office 365, etc.*

Through such integrations, the reports created can be shared within the organization, and thus making it a very important tool for organizations looking for a centralized data reporting system.

### 4. Datawrapper

Datawrapper is an online data visualization tool that can be used in various contexts. It is very easy to use, and it has a **clean and intuitive user interface**.

Datawrapper allows users to create charts and maps directly in the browser by uploading their data files. The charts and maps created in Datawrapper are responsive and designed for all kinds of devices, so readers will be able to view them on any device that they are using.

Datawrapper is free for everyone; however, there are certain limitations in the free version.

**For example**, it only lets you upload 500 rows of data and one sheet (or 5MB) of data at a time. The available chart types include Line, Bar, Area, Column, Pie, and Scatter. The data files that can be uploaded are .csv,.tsv or .txt files.

### 5. Plotly

Plotly is a data visualization tool that is used to create interactive graphs, charts, and maps. You can also use Plotly to create a visualization of a dataset, then share the link of that visualization with your readers on social media or on your blog.

Graphs made on Plotly are interactive and have a unique URL, so they're easy for you to share. Readers can explore how you created them by hovering over data points and viewing information about them.

Readers can also explore all the data interactively instead of trying to decipher your code, which makes it perfect for sharing both interactive plots and datasets with your audience.

Plotly's interface is easy to use, so you can create beautiful graphs in less time than ever before. Also, Plotly features a large library of open-source visualization types, allowing you to choose from a variety of plots and maps.

### 6. Sisense

Sisense is a data visualization tool that allows you to easily create interactive visualizations from your data. With Sisense, you can quickly and easily create extensive, informative dashboards that will help you understand your data better.

It has a very powerful yet simple and intuitive interface that allows you to drag-and-drop your data onto the canvas and create visualizations with a few clicks of a mouse.

*It is also fully integrated with several BI tools such as Microsoft Excel, BIRT, Pentaho, Qlikview & Tableau.*

Sisense utilizes multi-dimensional in-memory technology that is designed for Big Data. It also has an embedded artificial intelligence engine with predictive analytics, allowing you to easily visualize data trends and discover hidden patterns in your data.

### 7. Excel

Microsoft Excel is a data visualization tool that has an easy interface, so it doesn't have to be difficult to work with.

There are many different ways of visualizing data in Excel. One of them is by **using scatter plots,** Scatter plots display the relationship between two datasets that you would want to compare. You can also see how different variables are related to one another in order to determine if they're connected or not.

Many data analysts use scatter plots to analyze statistical, scientific, medical, and economic data for purposes such as market research or financial planning.

### 8. Zoho analytics

Zoho Analytics is a data visualization and reporting tool that can help you to easily create custom reports and dashboards. With Zoho Analytics, you can:

- quickly create custom reports and dashboards with drag-and-drop ease.
- Get insights into your data with interactive charts and graphs.
- Share reports and dashboards with colleagues or customers in just a few clicks.

Apart from this, it is part of the Zoho Office Suite, which also includes Zoho Writer, Zoho Sheet, and Zoho Show. You can use Zoho Analytics to report on data from any of these applications, as well as from external sources such as MySQL, SQL Server, Oracle, and Google Sheets.

# Data Visualization Tools

## 1. Tableau

- Tableau is a data visualization tool. You can create graphs, charts, maps, and many other graphics.

## 2. infogram

An infogram is a representation of information in a graphic format designed to make the data easily understandable in a view

## 3. Chartblocks

Chartblocks is an easy way to use online tool which required no coding and builds visualization from databases, spreadsheets, and live feeds.

## 4. Datawrapper

Datawrapper is easy visualization tool, and it requires zero codings. You can upload your data and easily create and publish a map or a chart

## 5. Plotly

Plotly is very user-friendly visualization tool which is quickly started within a few minutes.The guys use Plotly at Google and also by the US Air Force, Goji and The New York University.

## 6. RAW

RAW creates the missing link between spreadsheets and vector graphics on its home page.

## 7. Visual.ly

- Visual.ly is a visual content service. It has a dedicated data visualization service and their impressive portfolio that includes work for **Nike, VISA, Twitter, Ford, The Huffington post**, and the national geographic.

## 8. D3.js

- D3.js is a best data visualization library for manipulating documents. D3.js runs on JavaScript, and it uses **CSS**, **html**, and **SVG**. D3.js is an open-source and applies a data-driven transformation to a webpage. It's only applied when data is in **JSON** and **XML** file.

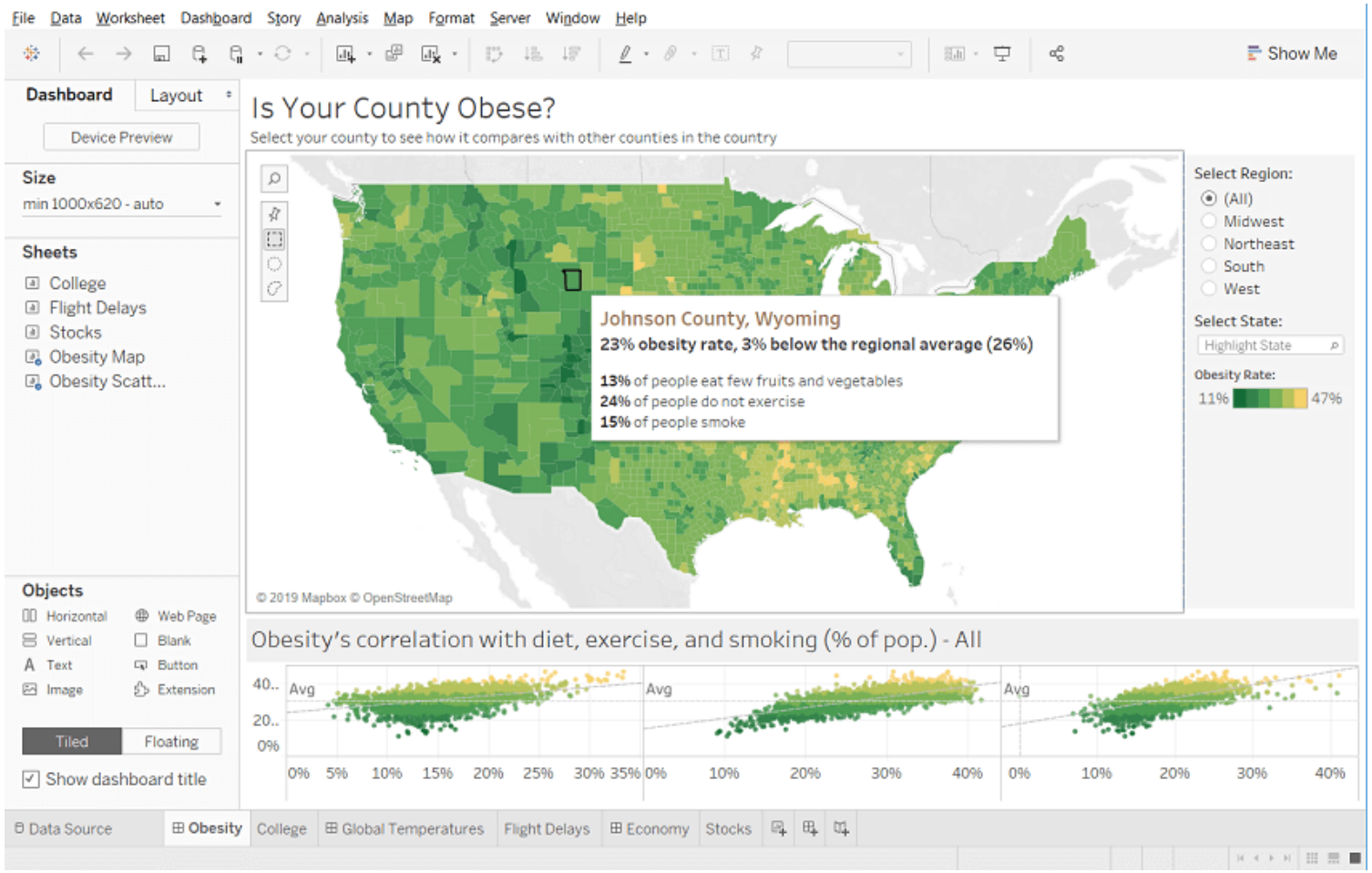
## 9. Ember Charts

- Ember charts are based on the ember.js and D3.js framework, and it uses the D3.js under the hood. It also applied when the data is in **JSON** and **XML** file.

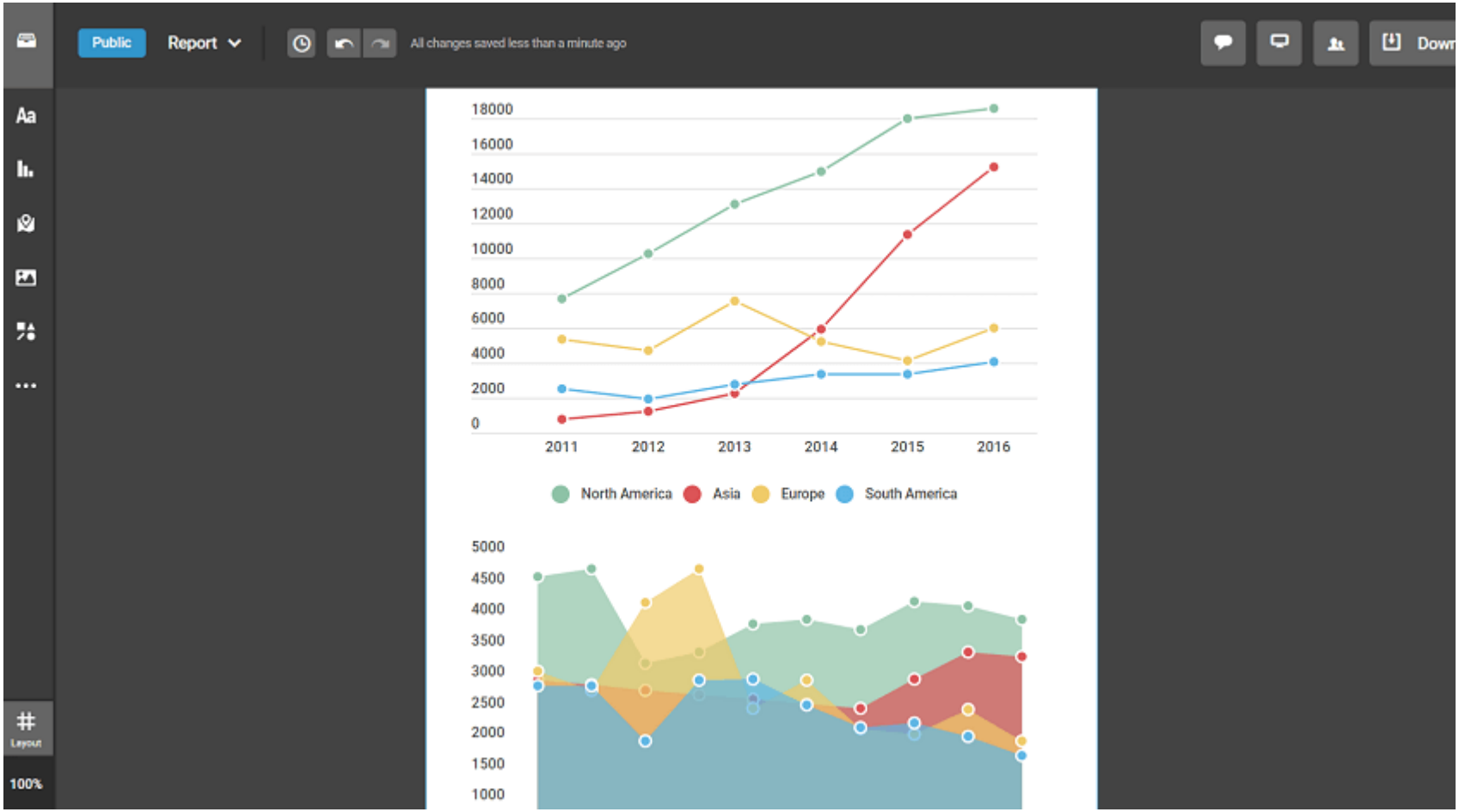
## 10. NVD3

- NVD3 is a project that attempts to build reusable charts and components. This project is to keeps all your charts neat and customizable.

# 1.Tableau

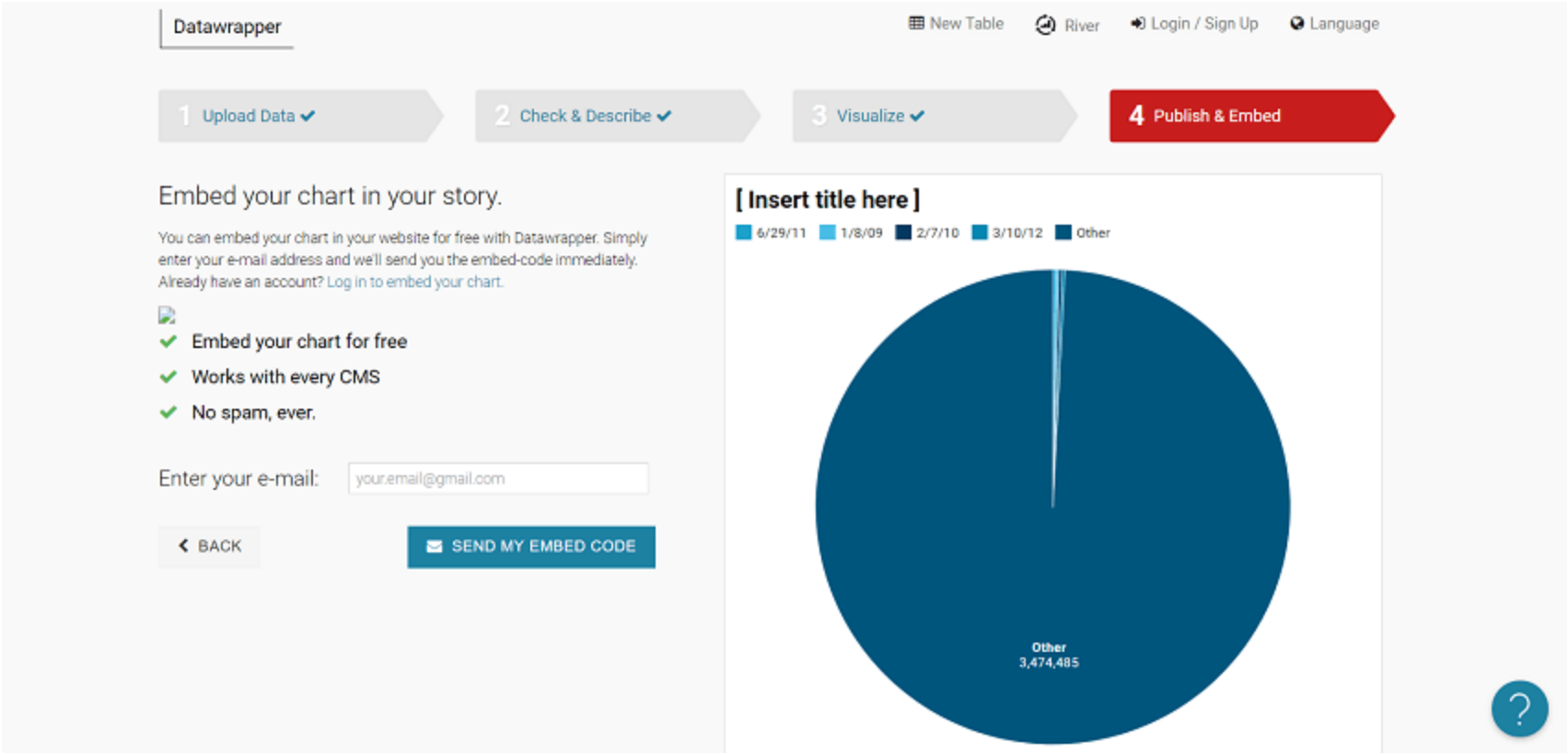

File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Dashboard
Layout
Device Preview
Size
min 1000x620 - auto
Sheets
College
Flight Delays
Stocks
Obesity Map
Obesity Scatt...
Objects
Horizontal
Vertical
Text
Image
Web Page
Blank
Button
Extension
Tiled
Floating
Show dashboard title
Is Your County Obese?
Select your county to see how it compares with other counties in the country
Johnson County, Wyoming
23% obesity rate, 3% below the regional average (26%)
13% of people eat few fruits and vegetables
24% of people do not exercise
15% of people smoke
© 2019 Mapbox © OpenStreetMap
Select Region:
(All)
Midwest
Northeast
South
West
Select State:
Highlight State
Obesity Rate:
11%
47%
Obesity's correlation with diet, exercise, and smoking (% of pop.) - All
40..
20..
0%
Avg
0% 5% 10% 15% 20% 25% 30% 35%
0% 10% 20% 30% 40%
0% 10% 20% 30% 40%
Data Source
Obesity
College
Global Temperatures
Flight Delays
Economy
Stocks

# 2. infogram

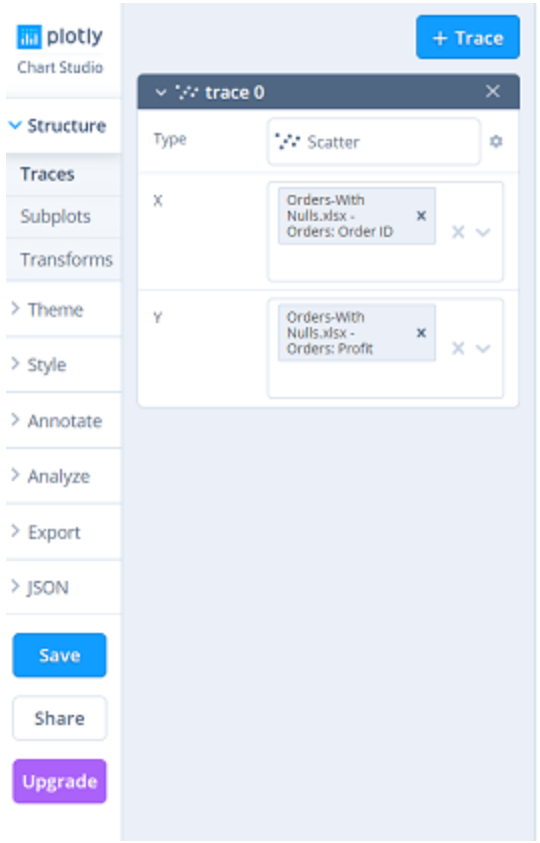

Public
Report
All changes saved less than a minute ago
Down
Aa
Layout
100%
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
2011
2012
2013
2014
2015
2016
North America
Asia
Europe
South America
5000
4500
4000
3500
3000
2500
2000
1500
1000

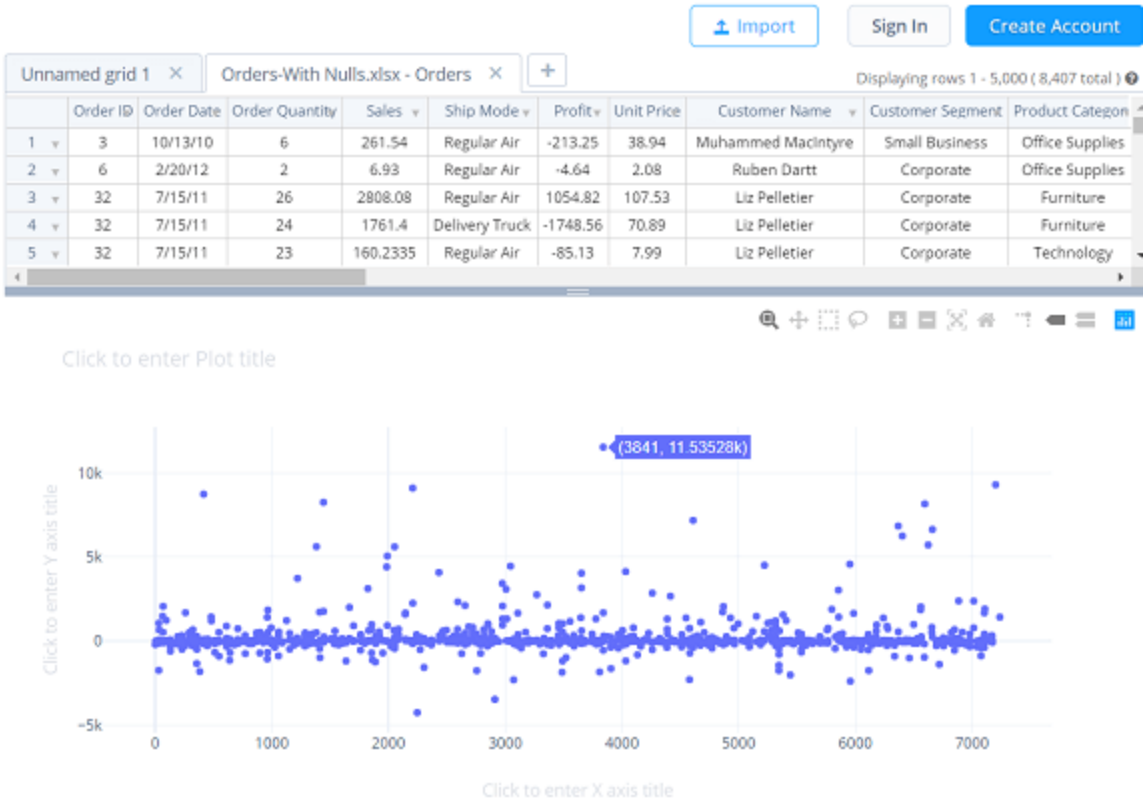
# 3. Chartblocks

# 4. Datawrapper

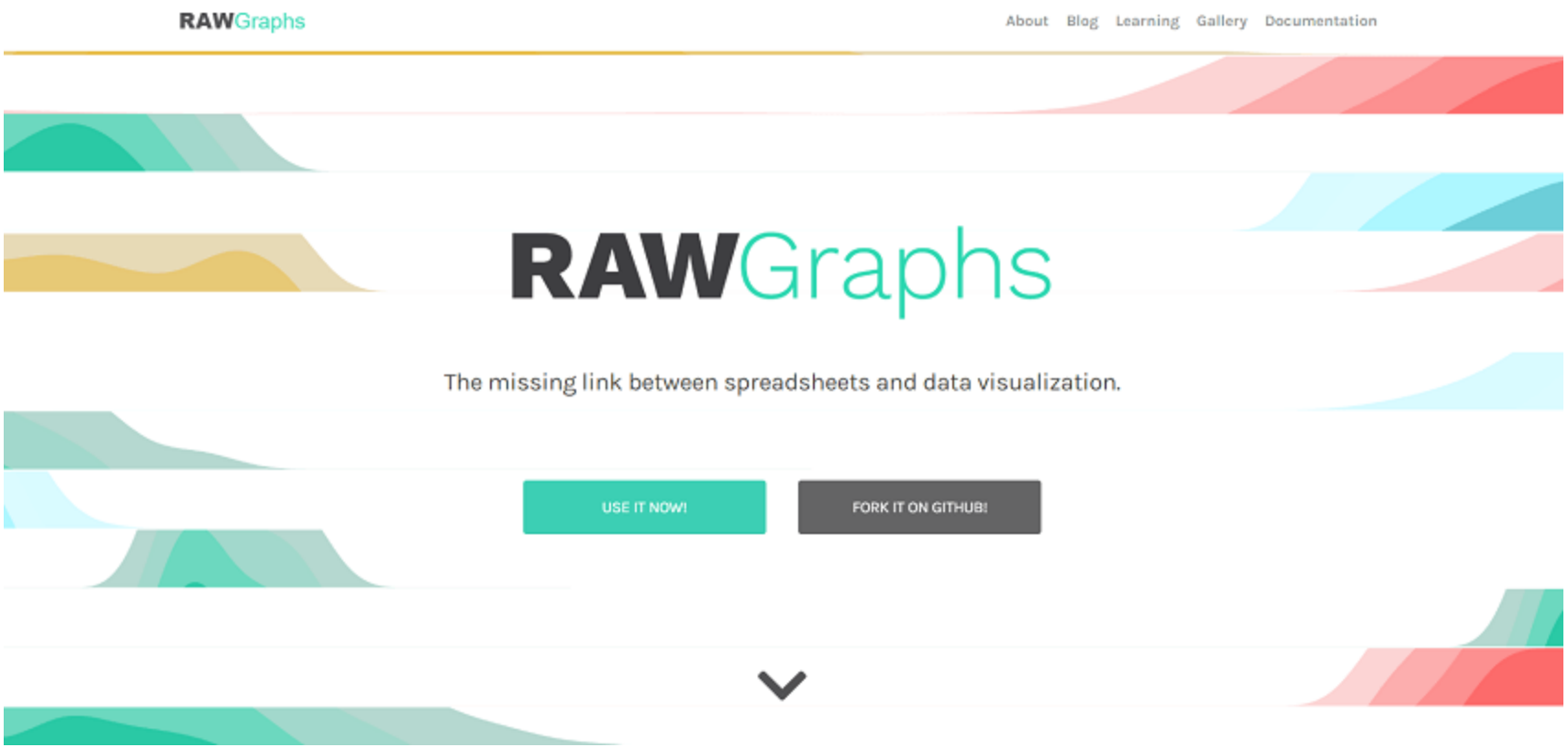


# 5. Plotly


plotly
Chart Studio
+ Trace
trace 0
Type
Scatter
X
Orders-With Nulls.xlsx - Orders: Order ID
Y
Orders-With Nulls.xlsx - Orders: Profit
Structure
Traces
Subplots
Transforms
Theme
Style
Annotate
Analyze
Export
JSON
Save
Share
Upgrade


Import
Sign In
Create Account
Unnamed grid 1
Orders-With Nulls.xlsx - Orders
Displaying rows 1 - 5,000 ( 8,407 total )
Order ID | Order Date | Order Quantity | Sales | Ship Mode | Profit | Unit Price | Customer Name | Customer Segment | Product Categor
1 | 3 | 10/13/10 | 6 | 261.54 | Regular Air | -213.25 | 38.94 | Muhammed Macintyre | Small Business | Office Supplies
2 | 6 | 2/20/12 | 2 | 6.93 | Regular Air | -4.64 | 2.08 | Ruben Dartt | Corporate | Office Supplies
3 | 32 | 7/15/11 | 26 | 2808.08 | Regular Air | 1054.82 | 107.53 | Liz Pelletier | Corporate | Furniture
4 | 32 | 7/15/11 | 24 | 1761.4 | Delivery Truck | -1748.56 | 70.89 | Liz Pelletier | Corporate | Furniture
5 | 32 | 7/15/11 | 23 | 160.2335 | Regular Air | -85.13 | 7.99 | Liz Pelletier | Corporate | Technology
Click to enter Plot title
(3841, 11.53528k)
10k
5k
0
-5k
0
1000
2000
3000
4000
5000
6000
7000
Click to enter Y axis title
Click to enter X axis title

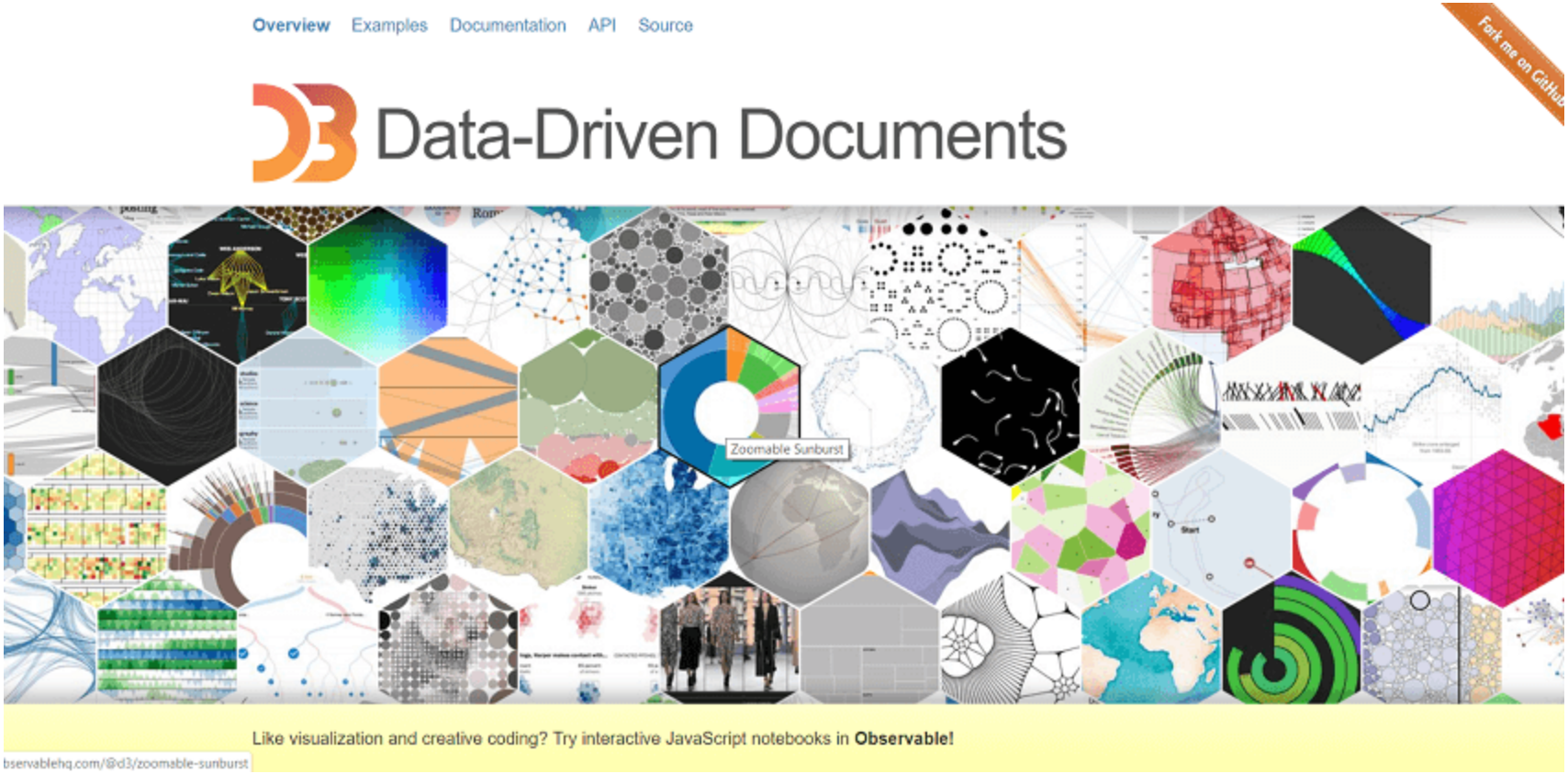
# 6. RAW


RAWGraphs
About
Blog
Learning
Gallery
Documentation
RAWGraphs
The missing link between spreadsheets and data visualization.
USE IT NOW!
FORK IT ON GITHUB!

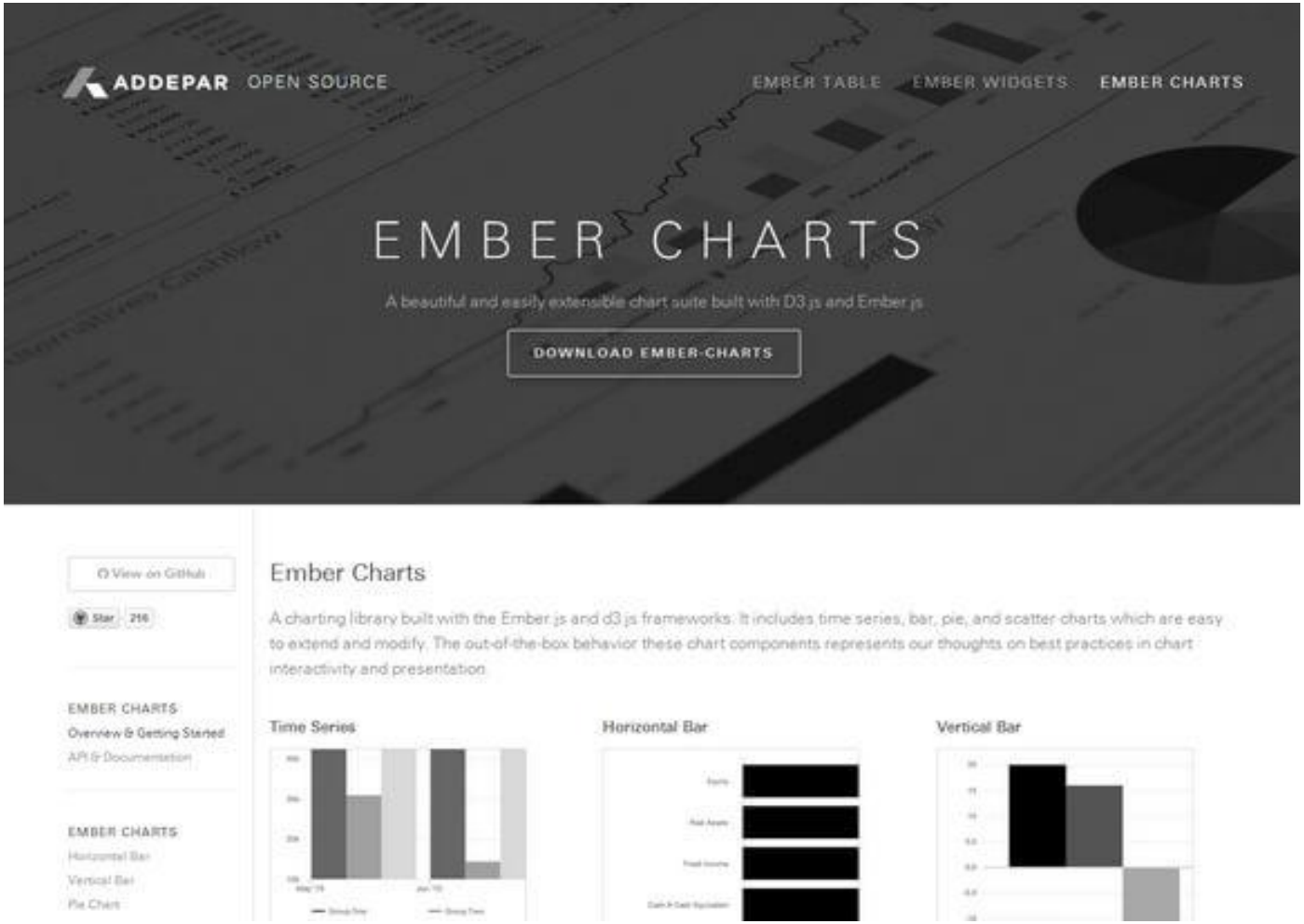
# 7. Visual.ly


visually
PRODUCTS
SOLUTIONS
HOW IT WORKS
PORTFOLIO
LOG IN
GET A QUOTE
Visual Content for Modern Marketers
A new way to think about content creation
Specialized creative talent, online and on-demand
GET A QUOTE
LEARN MORE

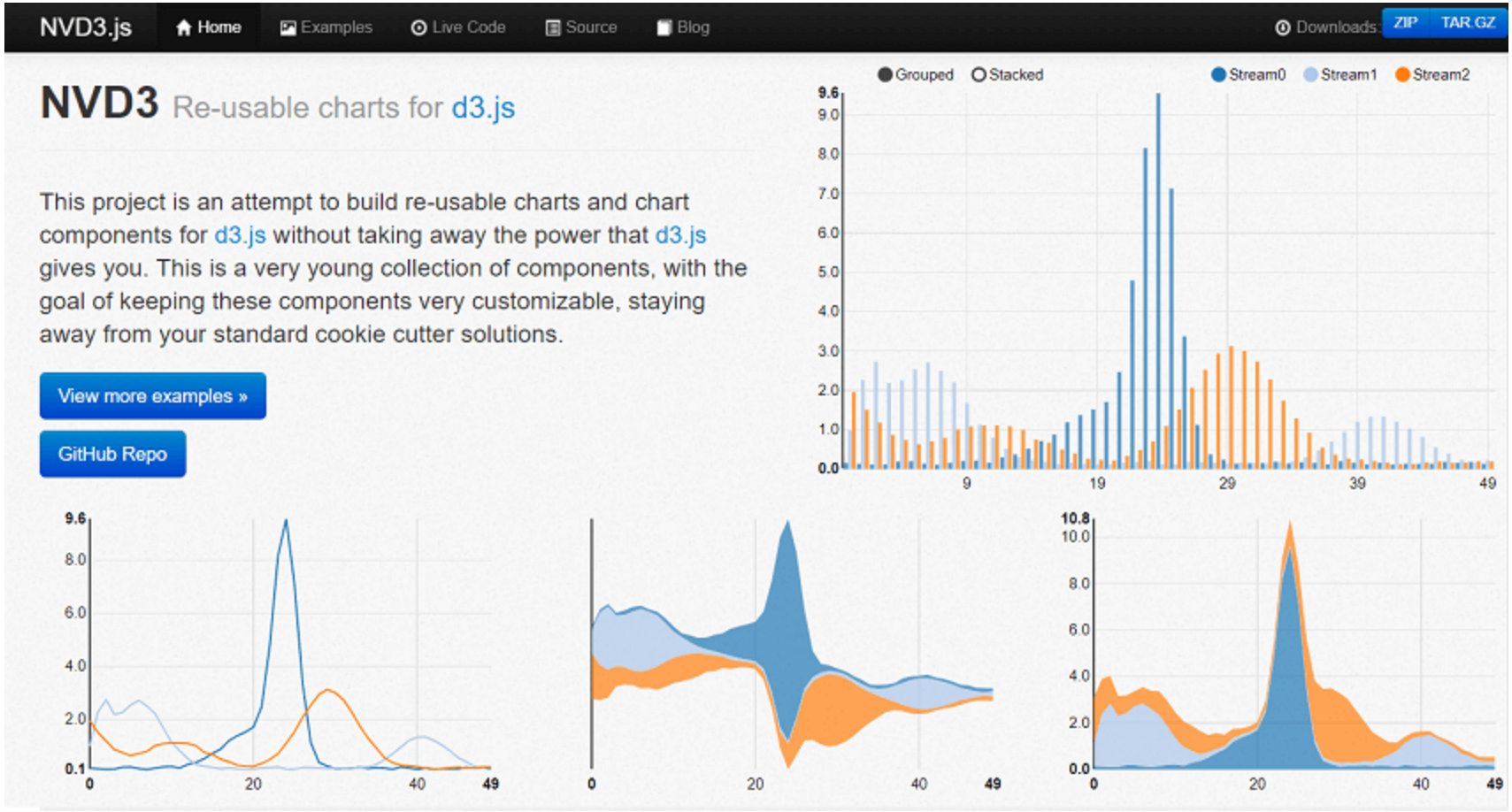
# 8. D3.js


Overview Examples Documentation API Source
Fork me on GitHub
Data-Driven Documents
Zoomable Sunburst
Like visualization and creative coding? Try interactive JavaScript notebooks in Observable!
bservablehq.com/@d3/zoomable-sunburst

# 9. Ember Charts



# 10. NVD3


NVD3.js
Home
Examples
Live Code
Source
Blog
Downloads:
ZIP
TAR.GZ
NVD3 Re-usable charts for d3.js
This project is an attempt to build re-usable charts and chart components for d3.js without taking away the power that d3.js gives you. This is a very young collection of components, with the goal of keeping these components very customizable, staying away from your standard cookie cutter solutions.
View more examples »
GitHub Repo
Grouped
Stacked
Stream0
Stream1
Stream2

# SEVEN STAGES OF data visualization

# SEVEN STAGES OF data visualization

*Acquire*

Obtain the data, whether from a file on a disk or a source over a network.

*Parse*

Provide some structure for the data's meaning, and order it into categories.

*Filter*

Remove all but the data of interest.

*Mine*

Apply methods from statistics or data mining as a way to discern patterns or place the data in mathematical context.

*Represent*

Choose a basic visual model, such as a bar graph, list, or tree.

*Refine*

Improve the basic representation to make it clearer and more visually engaging.

*Interact*

Add methods for manipulating the data or controlling what features are visible.

# Acquire

- The acquisition step involves obtaining the data. Like many of the other steps, this can be either extremely complicated (i.e., trying to glean useful data from a large system)
- or very simple (reading a readily available text file).

# Example

- A copy of the zip code listing can be found on the U.S. Census Bureau web site
- frequently used for geographic coding of statistical data.
- The listing is a freely available file with approximately 42,000 lines, one for each of the codes, a tiny portion of which is shown :

# *Zip codes in the format provided by the U.S. Census Bureau*

| | | | | | | |
|---|---|---|---|---|---|---|
| 00210 | +43.005895 | -071.013202 | U | PORTSMOUTH | 33 | 015 |
| 00211 | +43.005895 | -071.013202 | U | PORTSMOUTH | 33 | 015 |
| 00212 | +43.005895 | -071.013202 | U | PORTSMOUTH | 33 | 015 |
| 00213 | +43.005895 | -071.013202 | U | PORTSMOUTH | 33 | 015 |
| 00214 | +43.005895 | -071.013202 | U | PORTSMOUTH | 33 | 015 |
| 00215 | +43.005895 | -071.013202 | U | PORTSMOUTH | 33 | 015 |
| 00501 | +40.922326 | -072.637078 | U | HOLTSVILLE | 36 | 103 |
| 00544 | +40.922326 | -072.637078 | U | HOLTSVILLE | 36 | 103 |
| 00601 | +18.165273 | -066.722583 | | ADJUNTAS | 72 | 001 |
| 00602 | +18.393103 | -067.180953 | | AGUADA | 72 | 003 |
| 00603 | +18.455913 | -067.145780 | | AGUADILLA | 72 | 005 |
| 00604 | +18.493520 | -067.135883 | | AGUADILLA | 72 | 005 |
| 00605 | +18.465162 | -067.141486 | P | AGUADILLA | 72 | 005 |
| 00606 | +18.172947 | -066.944111 | | MARICAO | 72 | 093 |
| 00610 | +18.288685 | -067.139696 | | ANASCO | 72 | 011 |
| 00611 | +18.279531 | -066.802170 | P | ANGELES | 72 | 141 |
| 00612 | +18.450674 | -066.698262 | | ARECIBO | 72 | 013 |
| 00613 | +18.458093 | -066.732732 | P | ARECIBO | 72 | 013 |
| 00614 | +18.429675 | -066.674506 | P | ARECIBO | 72 | 013 |
| 00616 | +18.444792 | -066.640678 | | BAJADERO | 72 | 013 |

# Parse

- After you acquire the data, it needs to be parsed—changed into a format that tags each part of the data with its intended use.
- Each line of the file must be broken along its individual parts; in this case, it must be delimited at each tab character.
- Then,each piece of data needs to be converted to a useful format.


*Figure 1-2. Structure of acquired data*

Each field is formatted as a data type that we'll handle in a conversion program:

*String*

A set of characters that forms a word or a sentence. Here, the city or town name is designated as a string.

*Float*

A number with decimal points (used for the latitudes and longitudes of each location). The name is short for *floating point, from programming nomenclature* that describes how the numbers are stored in the computer's memory.

*Character*

A single letter or other symbol. In this data set, a character sometimes designates special post offices.

*Integer*

A number without a fractional portion, and hence no decimal points (e.g., –14,0, or 237).

*Index*

Data (commonly an integer or string) that maps to a location in another table of data. In this case, the index maps numbered codes to the names and two-digit abbreviations of states.

- With the completion of this step, the data is successfully tagged and consequently
- more useful to a program that will manipulate or represent it in some way.

# Filter

- The next step involves filtering the data to remove portions not relevant to our use.
- In this example, for the sake of keeping it simple, we'll be focusing on the contiguous
- 48 states, so the records for cities and towns that are not part of those states—
- Alaska, Hawaii, and territories such as Puerto Rico—are removed.

# Mine

This step involves math, statistics, and data mining. The data in this case receives only a simple treatment: the program must figure out the minimum and maximum values for latitude and longitude by running through the data (as shown in Figure )

so that it can be presented on a screen at a proper scale. Most of the time,this step will be far more complicated than a pair of simple math operations.

*Mining the data: just compare values to find the minimum and maximum*

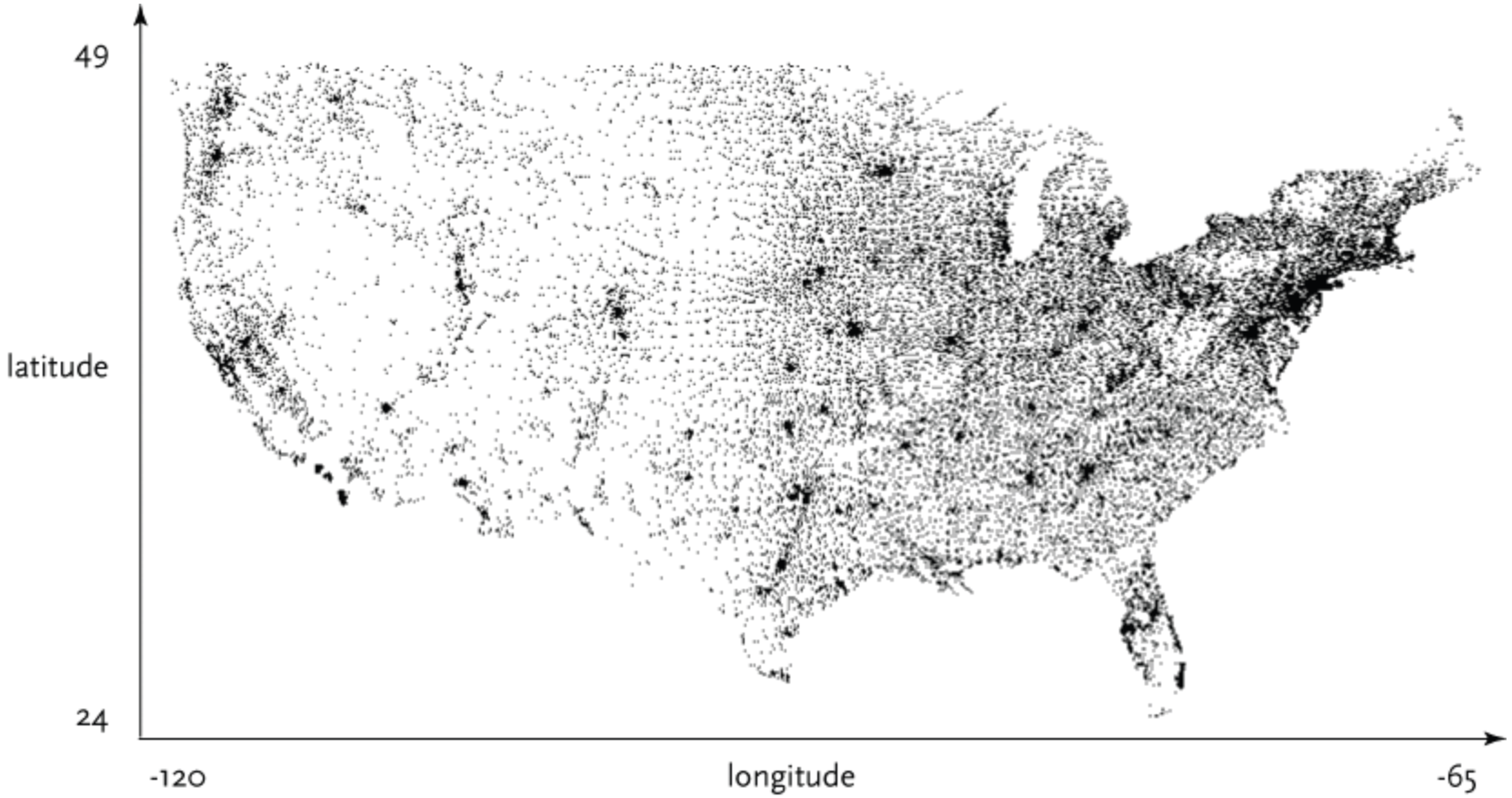
| | | | | |
|---|---|---|---|---|
| 00210 | 43.005895 | -71.013202 | PORTSMOUTH | NH |
| 00211 | 43.005895 | -71.013202 | PORTSMOUTH | NH |
| 00212 | 43.005895 | -71.013202 | PORTSMOUTH | NH |
| 00213 | 43.005895 | -71.013202 | PORTSMOUTH | NH |
| 00214 | 43.005895 | -71.013202 | PORTSMOUTH | NH |
| 00215 | 43.005895 | -71.013202 | PORTSMOUTH | NH |
| 00501 | 40.922326 | -72.637078 | HOLTSVILLE | NY |
| 00544 | 40.922326 | -72.637078 | HOLTSVILLE | NY |
| . | . | . | . | . |
| . | . | . | . | . |
| . | . | . | . | . |

# Represent

This step determines the basic form that a set of data will take. Some data sets are shown as lists, others are structured like trees, and so forth.

In this case, each zip code has a latitude and longitude, so the codes can be mapped as a two-dimensional plot, with the minimum and maximum values for the latitude and longitude used for the start and end of the scale in each dimension.

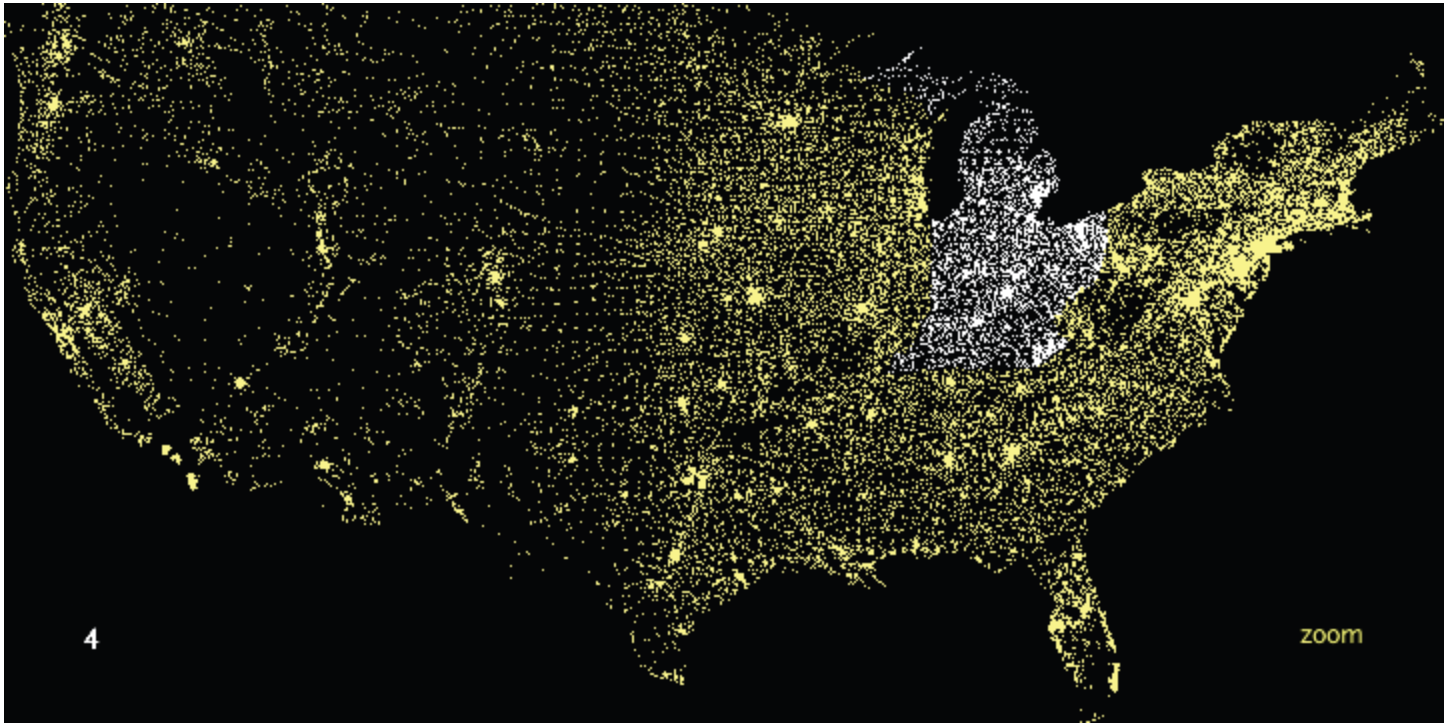
# *Basic visual representation of zip code data*



# Refine

- In this step, graphic design methods are used to further clarify the representation by calling more attention to particular data (establishing hierarchy) or by changing attributes (such as color) that contribute to readability.
- Hierarchy is established in Figure shown, for instance, by coloring the background deep gray and displaying the selected points (all codes beginning with four) in white andthe deselected points in medium yellow.

# *Using color to refine the representation*

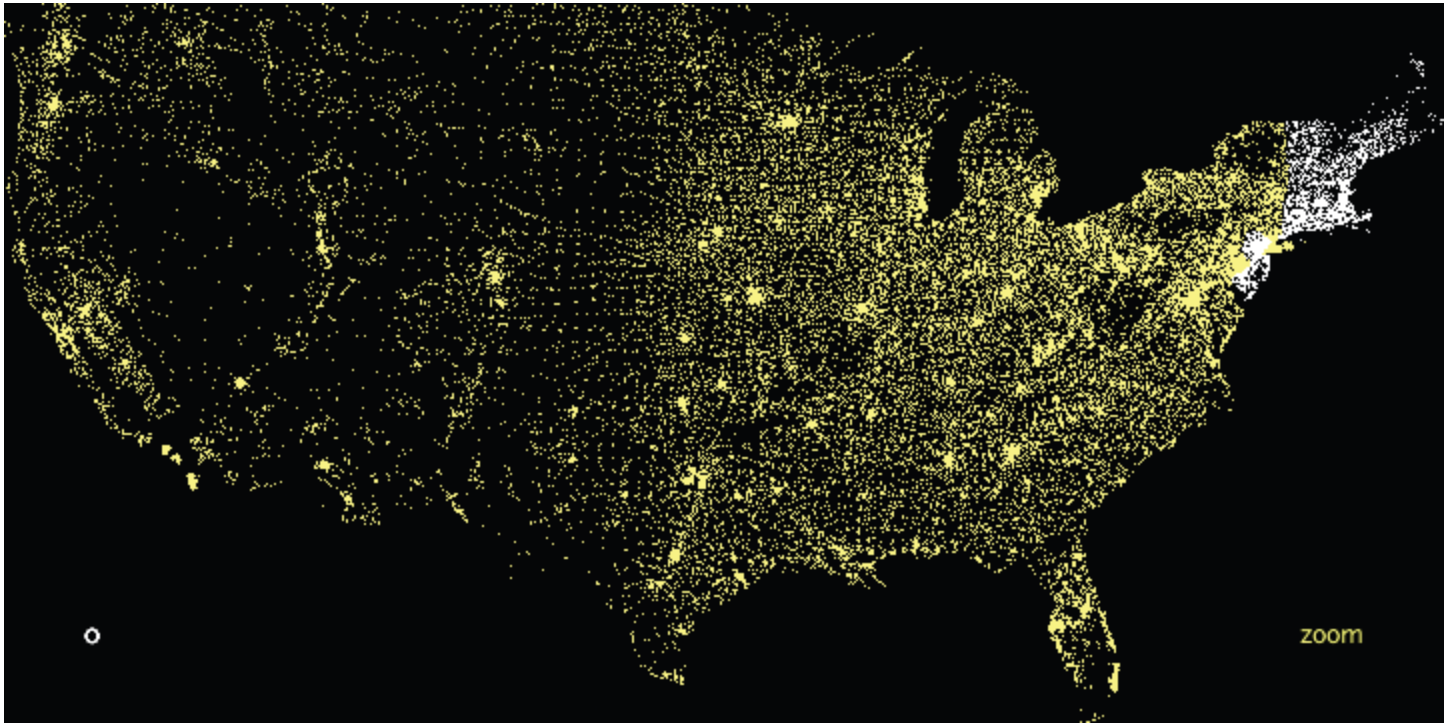


# Interact

- The next stage of the process adds interaction, letting the user control or explore the data. Interaction might cover things like selecting a subset of the data or changing the viewpoint.
- As another example of a stage affecting an earlier part of the process,this stage can also affect the refinement step, as a change in viewpoint might require the data to be designed differently.

- In the Zipdecode project, typing a number selects all zip codes that begin with that number. Figures show all the zip codes beginning with zero and nine,respectively.
- Another enhancement to user interaction (not shown here) enables the users to traverse the display laterally and run through several of the prefixes. After typing part or all of a zip code, holding down the Shift key allows users to replace the last number typed without having to hit the Delete key to back up.

*The user can alter the display through choices (zip codes starting with 0)*

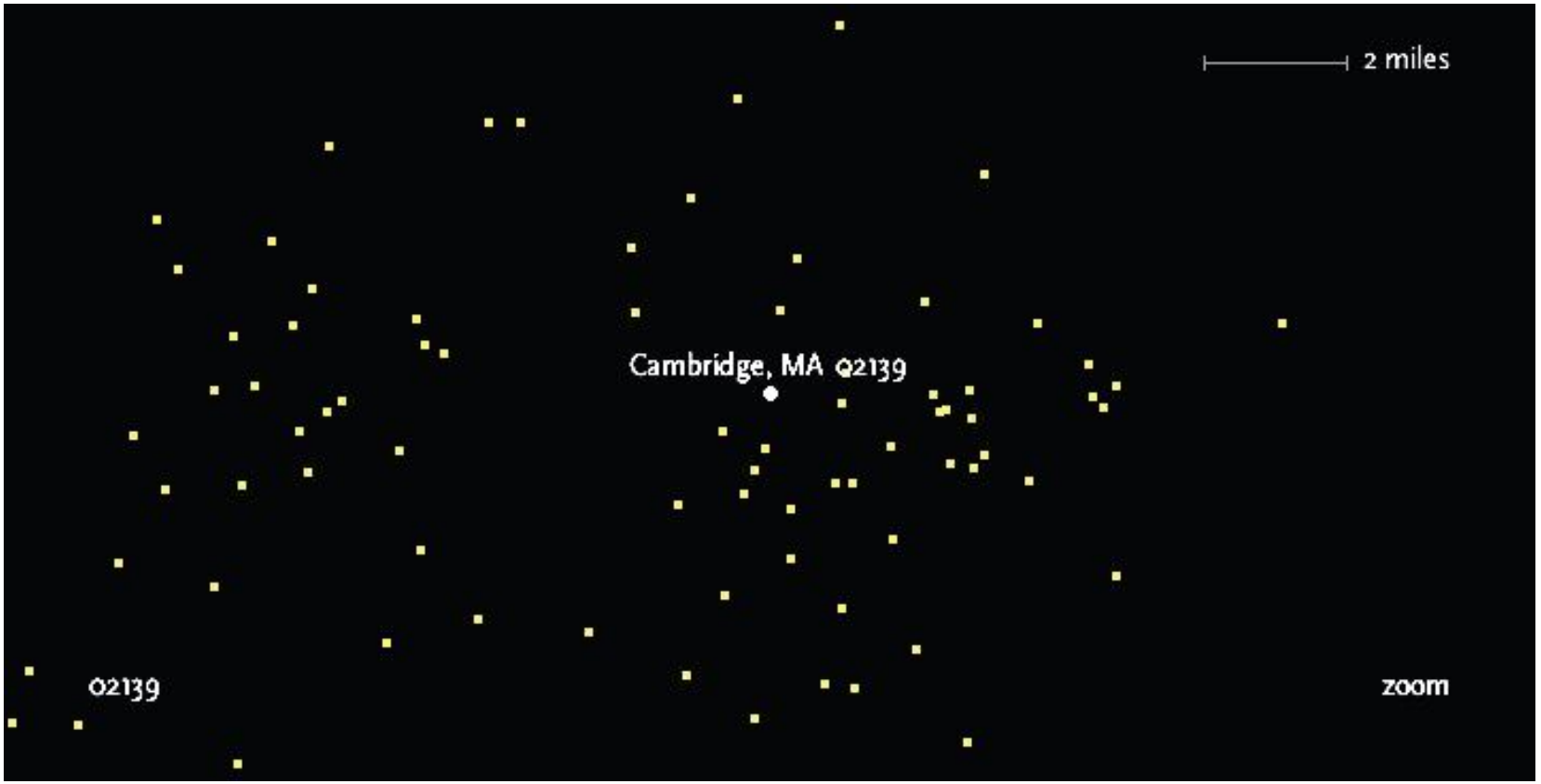


Typing is a very simple form of interaction, but it allows the user to rapidly gain an understanding of the zip code system's layout. Just contrast this sample application

with the difficulty of deducing the same information from a table of zip codes and

city names.The viewer can continue to type digits to see the area covered by each subsequent set of prefixes.

# *Honing in even further with the full zip code (02139)*



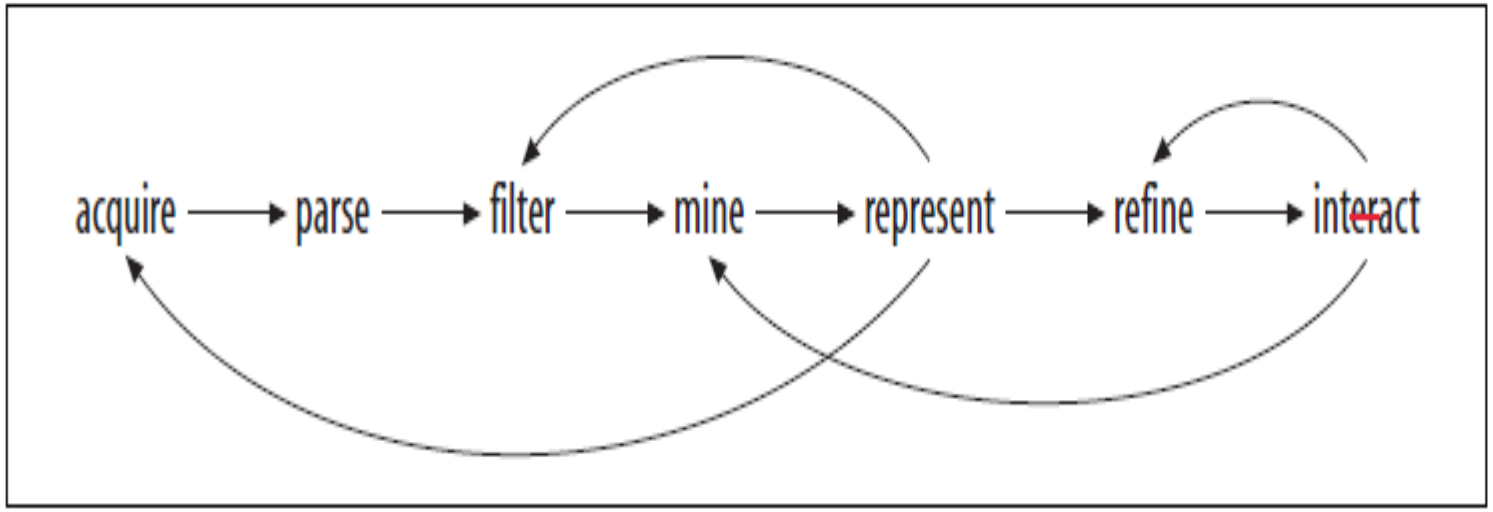



*Figure 1-12. Interactions between the seven stages*

- Acquire
- Parse
- Filter
- Mine
- Represent
- Refine
- Interact

- **Acquire** Obtain the data, whether from a file on a disk or a source over a network.
- **Parse** Provide some structure for the data's meaning, and order it into categories.
- **Filter** Remove all but the data of interest.
- **Mine** Apply methods from statistics or data mining as a way to discern patterns or place the data in mathematical context.

- **Represent** Choose a basic visual model, such as a bar graph, list, or tree.
- **Refine** Improve the basic representation to make it clearer and more visually engaging.
- **Interact** Add methods for manipulating the data or controlling what features are visib