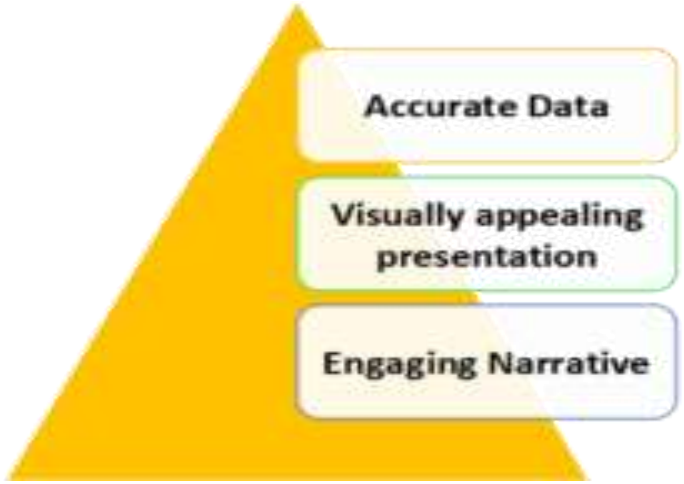
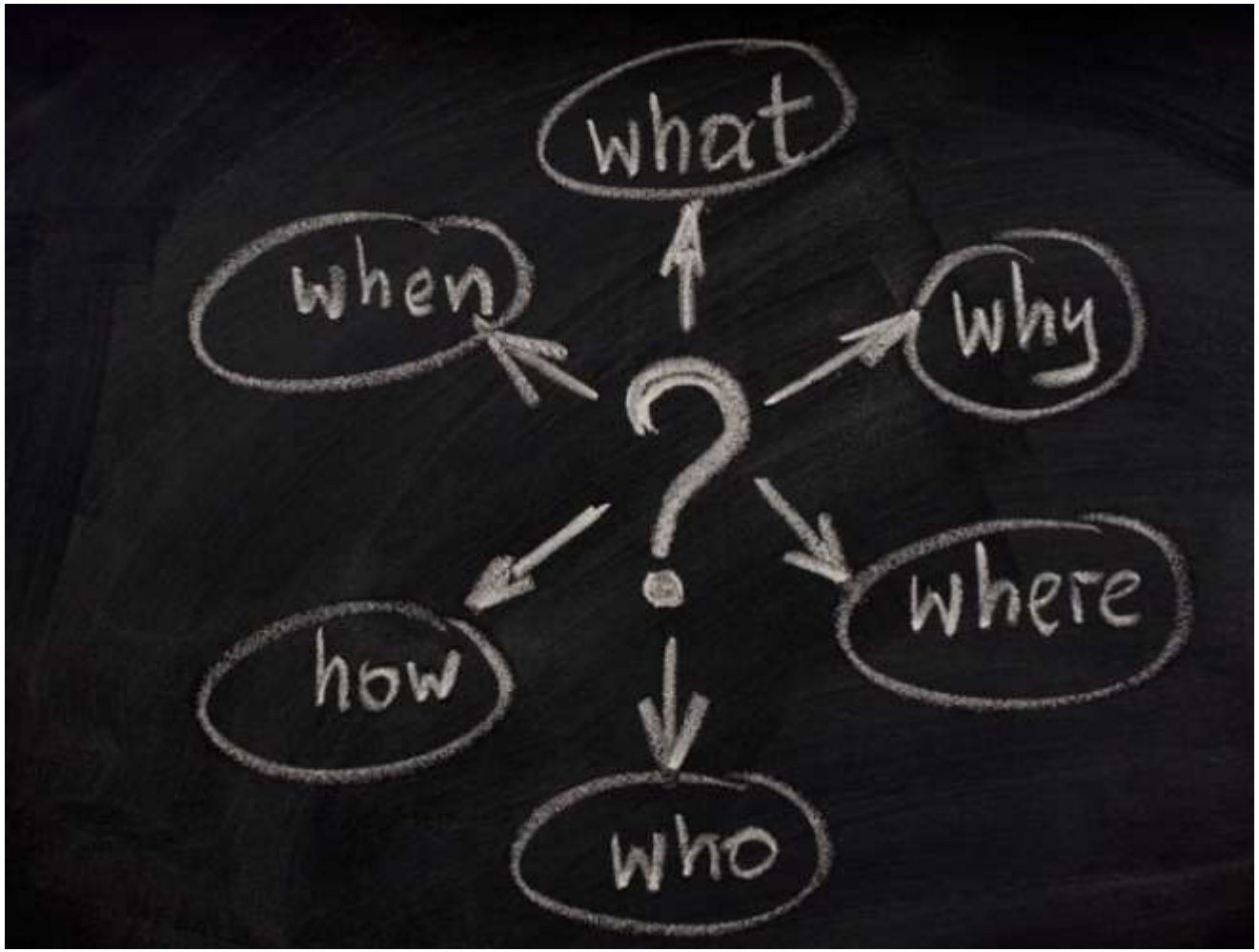
How to create a story?



Hierarchy

Figuring out which one or two or maybe three are the important ones, and eliminating the ones that you don't need, is a great way to get with the hierarchy and the information that you are trying to show to your users.



1-what's wanting? What's missing from the data you have? In the end, you almost always are going to be missing data.

2- what in the world? , a lot of times, you can find other data from other sources.

3-what's wild? , sometimes you want to take a little bit of an out-of-box approach.
Whether it's to capture your user's attention. Or to really tell a new and interesting story. And, it could really mean a couple of different things. One, is it can mean that you're just going to take a very unique visual approach to your data.

Explore data

A predictive data story should connect the dots for the current trend being depicted, find the underlying pattern and reason for the same and predict the future course of events. The story should have a natural flow of events.

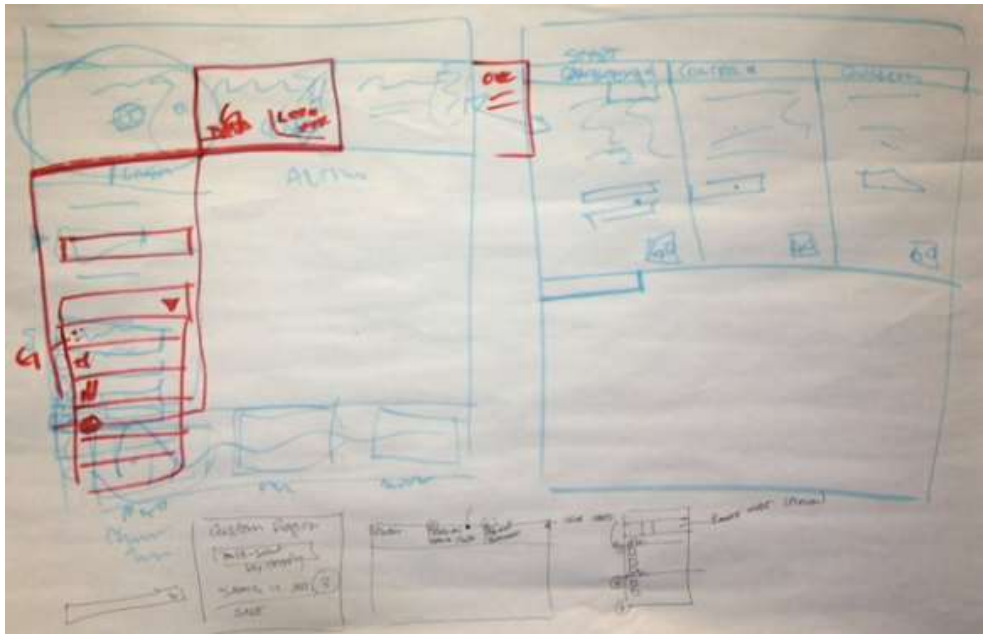
Convert data:common data Adjustment

- calculating indexes and ratios, which is a very core concept to understand(scaling ratios 0..1)
- percentiles (look at data in terms of portion of a whole =(1-rank)/total nb of items)
- aggregating data (aggregate many lines in 1 if possible)
- regrouping data from the way you get it(pivot tables)
- converting from one form to another. For instance, from Excel or comma-delimited files into JSON or XML or MySQL, which are very common forms used for web programming (online converter)

Sketch your idea and wireframe them

Why using data in a very analog way. four reasons

1. Speed,
2. Flexibility,
3. Scale,
4. Getting a good body mind connection



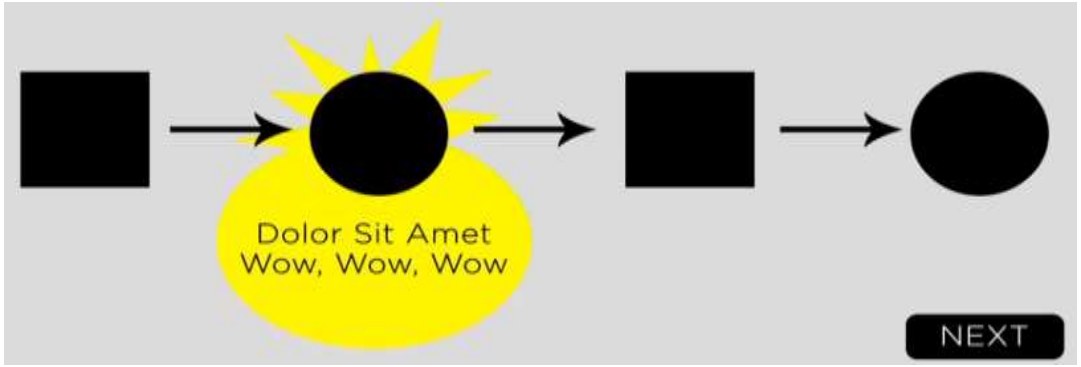
Structure a story in a narrative way

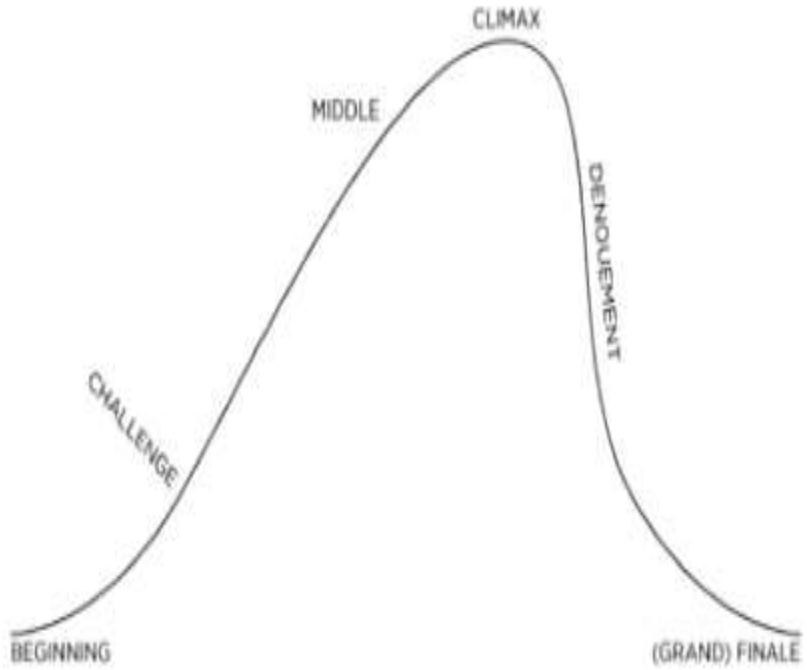
A headline is the most important top line summation of the story you're trying to tell.

the beginning of our story is the headline, and the introduction (establishing the premise and the context for the visualization, before diving into the data. And it belongs below the headline at the top of the page to draw the eye there first.)

the challenge to visualization. the story you're trying to tell, or the problem that's being addressed, or the question that your visualization is meant to answer.

introduce the climax? The, the sort of pivotal moment in the experience? And we're going to look at two different scenarios.
One is a static infographic and one is an interactive chart.

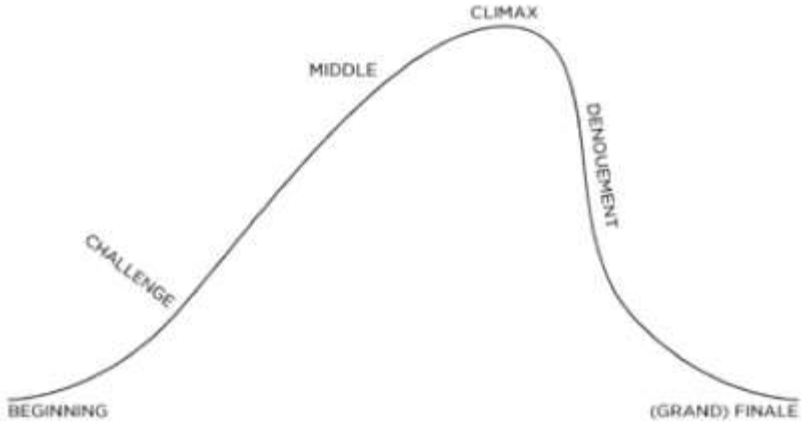




Structure a story in a narrative way ..continue

Go through deep exploration process. Where did this data come from? Why should I believe you? Why is it valid? That's sort of part of the conclusion in a story structure

Introduce conclusion (short video may be)

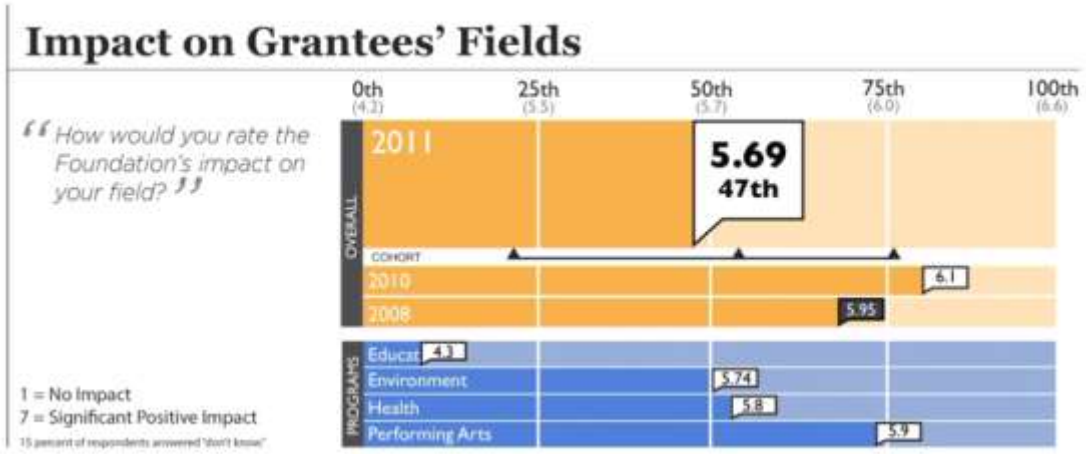


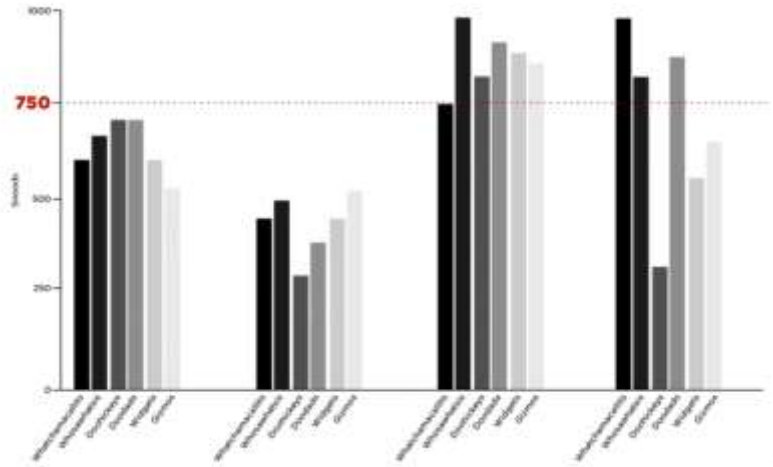
Connect with your audience

Make it relatable and visual as well as on topic

Communicate visually but do not over do it

1- use illustration and iconography
2- typography in visualization (emphasis things with high accuracy)
(So there are really for basic types of type in data visualization.
So there's axes and legends, and labels and callouts and
infographics that are sort of their own category
the balance between accuracy and readability and aesthetics !)

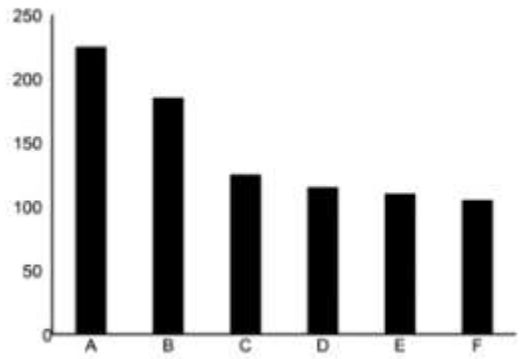


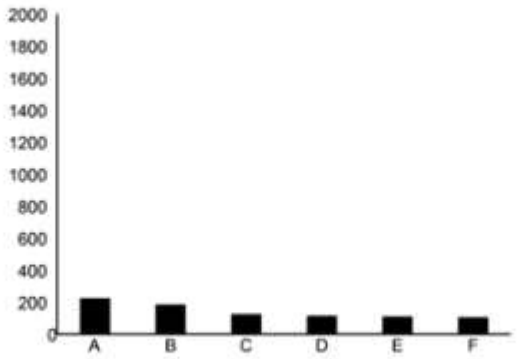


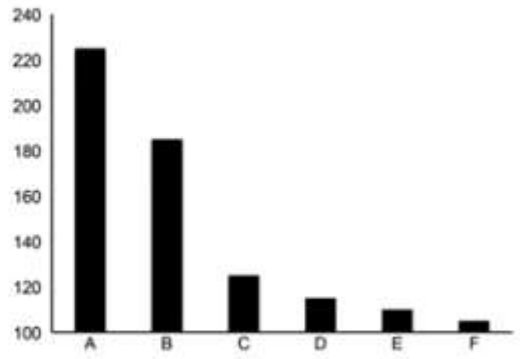
“The new graphic designer no longer creates visualizations by choosing a rigid collection of shapes, positions and color, but rather by choosing the rules needed for data to breathe form. Into geometric abstractions”

Five ways to show differentiation.
Position, size, color, contrast and shape
(draw the eye of the audience)

Importance of changing scale







Type chart (bar , line…
Internal reference
External reference
Fair and unbiased

Ask yourself are the skeptics going to believe this? Are they going to buy it, are they going to think that this is an honest representation of the data? And then the same thing on the believer's side.

Clear legends and sources

If you can, get to a point where you have, great user empathy for people who know nothing about what you're showing them, then that will help you understand, what's hard for them to understand which will help you figure out what you should include in the legend.

Include how to read the data

Add resources

Right chart ? figuring out which charts to use in which situation

Basic charts

Bar charts, boring and old but widely used, very effective an emphasis on a comparison within groups,

if you want to think of the elements of data as a part of a whole, like a category, then a stacked bar

If you want to emphasize not the total value for each, but the relative value, so how much each category influences the total value, then use a stacked percentage bar.

Line charts, to show the trends over continuous time

Area charts, hard to find the relative value; stacked percentage area chart. Which again is a really effective way of showing the relative strength of categories across time or whatever the x axis represents but as a portion of a whole.

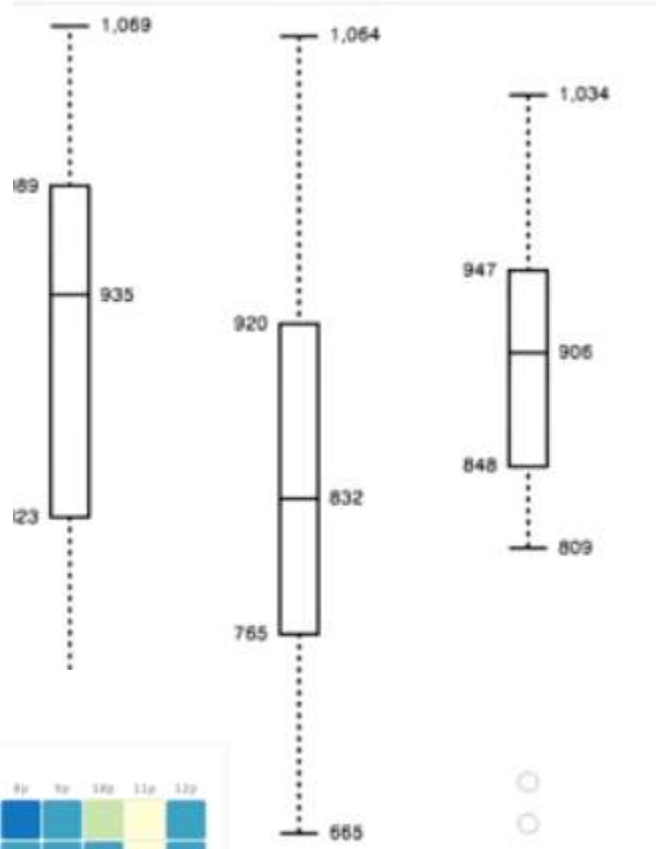
Scatter plots, Scatter plots are great at showing correlation. E.g as x increases, as things get further to the right, y also increases.

Bubble charts, Bubble charts are great at showing three variables, it's really just a scatter plot, but now we have a third variable (size of the dot)

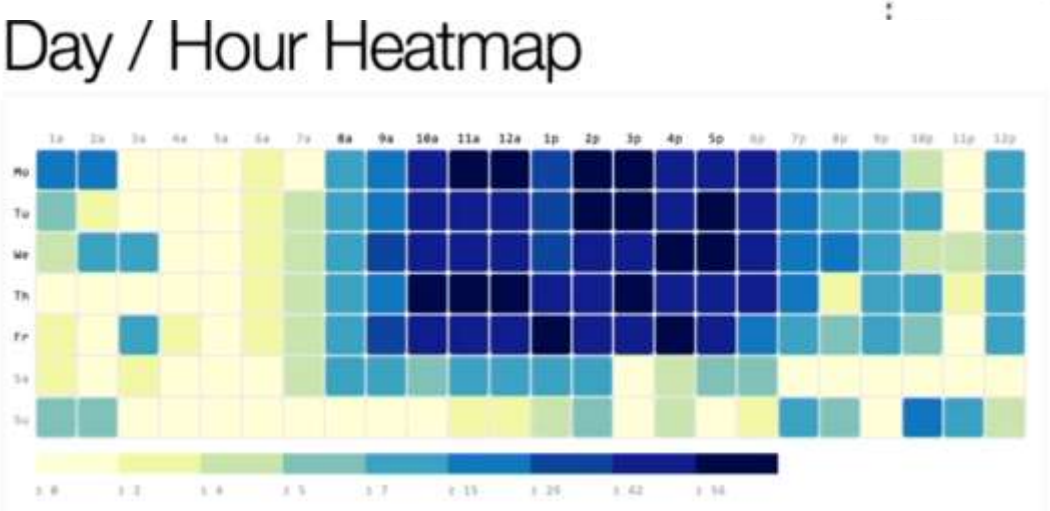
Pie charts ,hard to parse, bad at showing slight variance between variable but effective in comparing 2 variables

Alternative chart ?

Box plots with whiskers



Heatmaps are a great way at looking at, really a limited list amount of data, across a number of variables. And to see how one compares to the other, or how groups compare to other groups

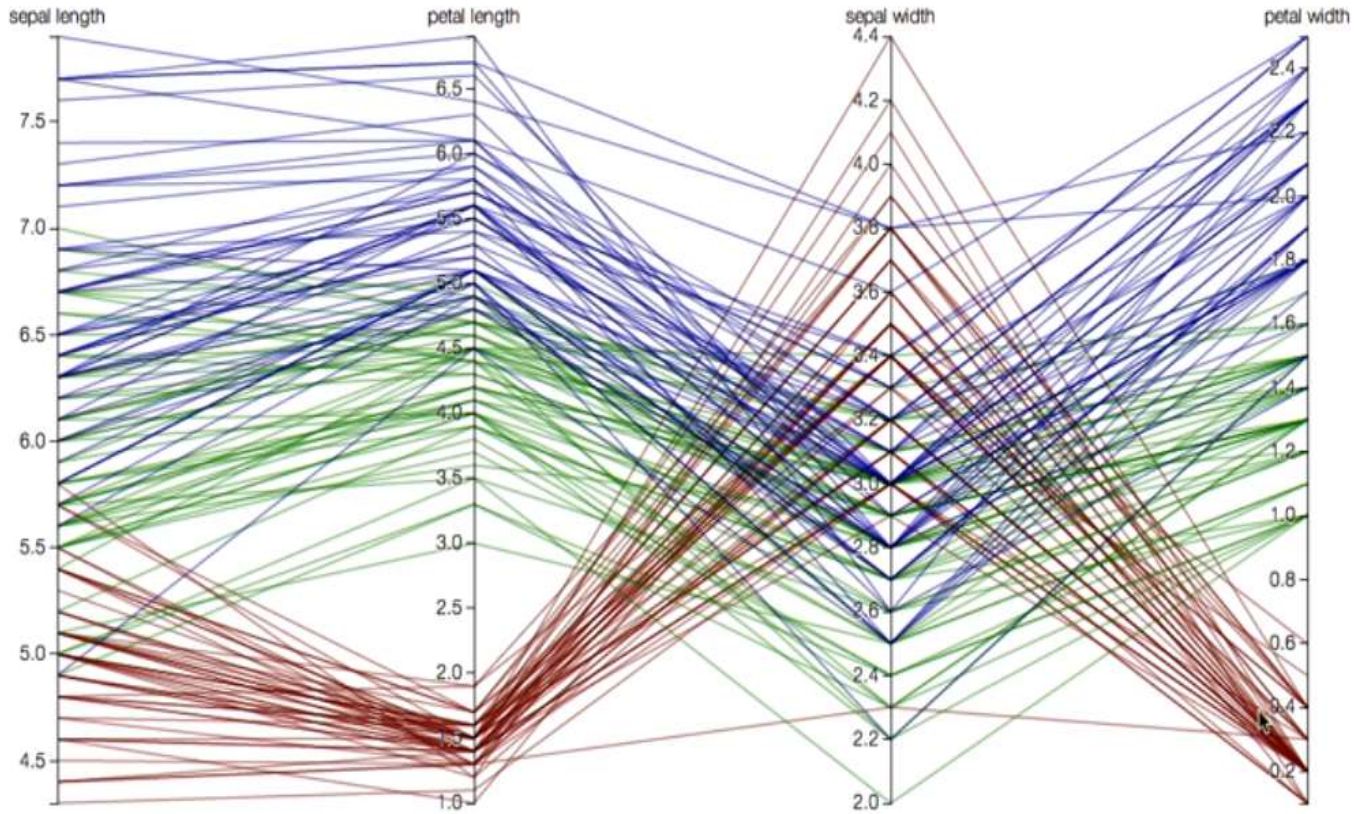


Radar or spider charts , hard to understand when there is a lot of data

Parallel coordinates Chart

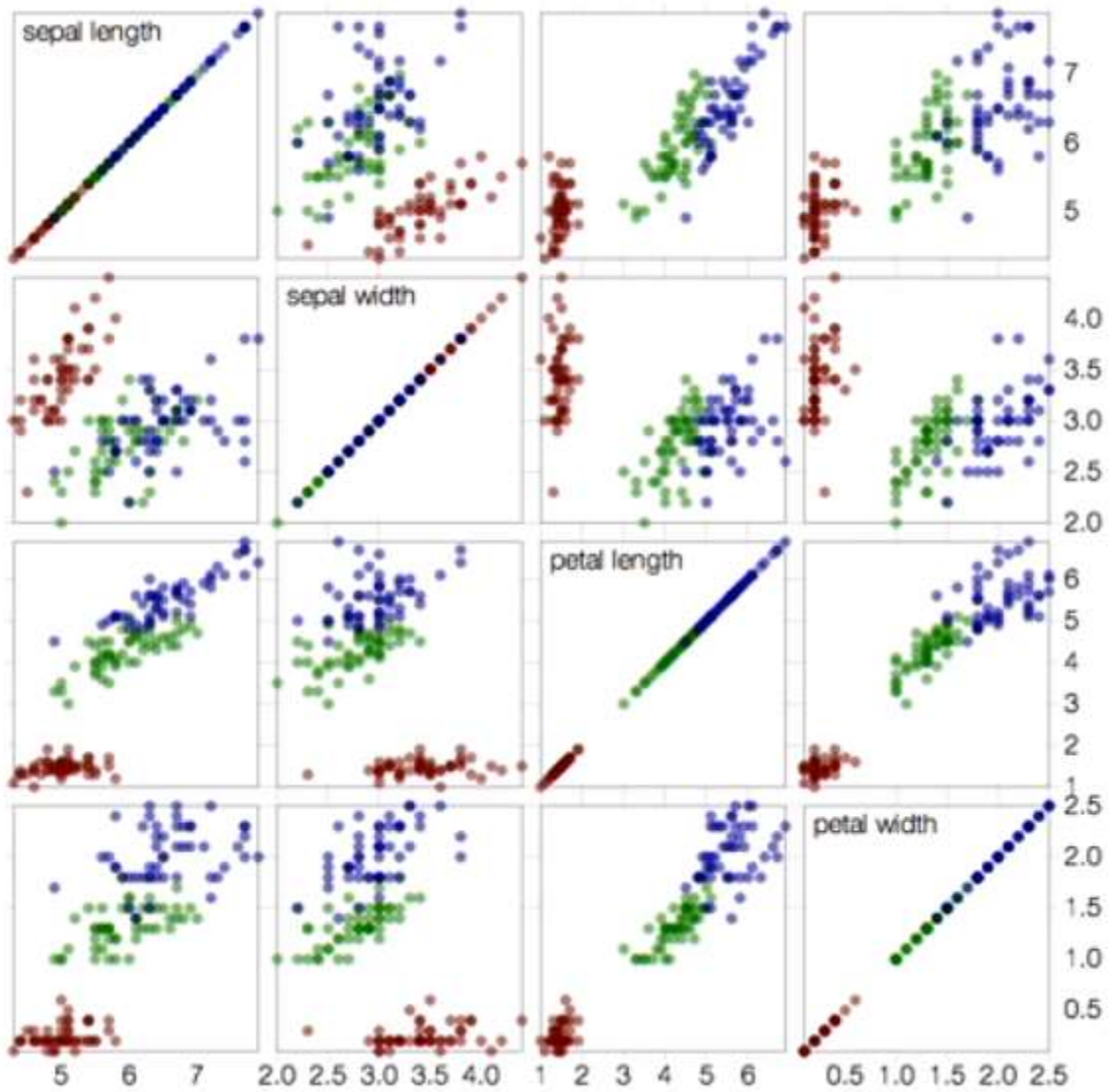
Good at showing relationships between variables. So for instance, that there is some sort of correlation or relationship between the second variable, and the fourth variable.
Petal length and width amongst this red category, but less correlation when looking at sepal length versus sepal width for **that category**

Scrubbing :cut part of chart



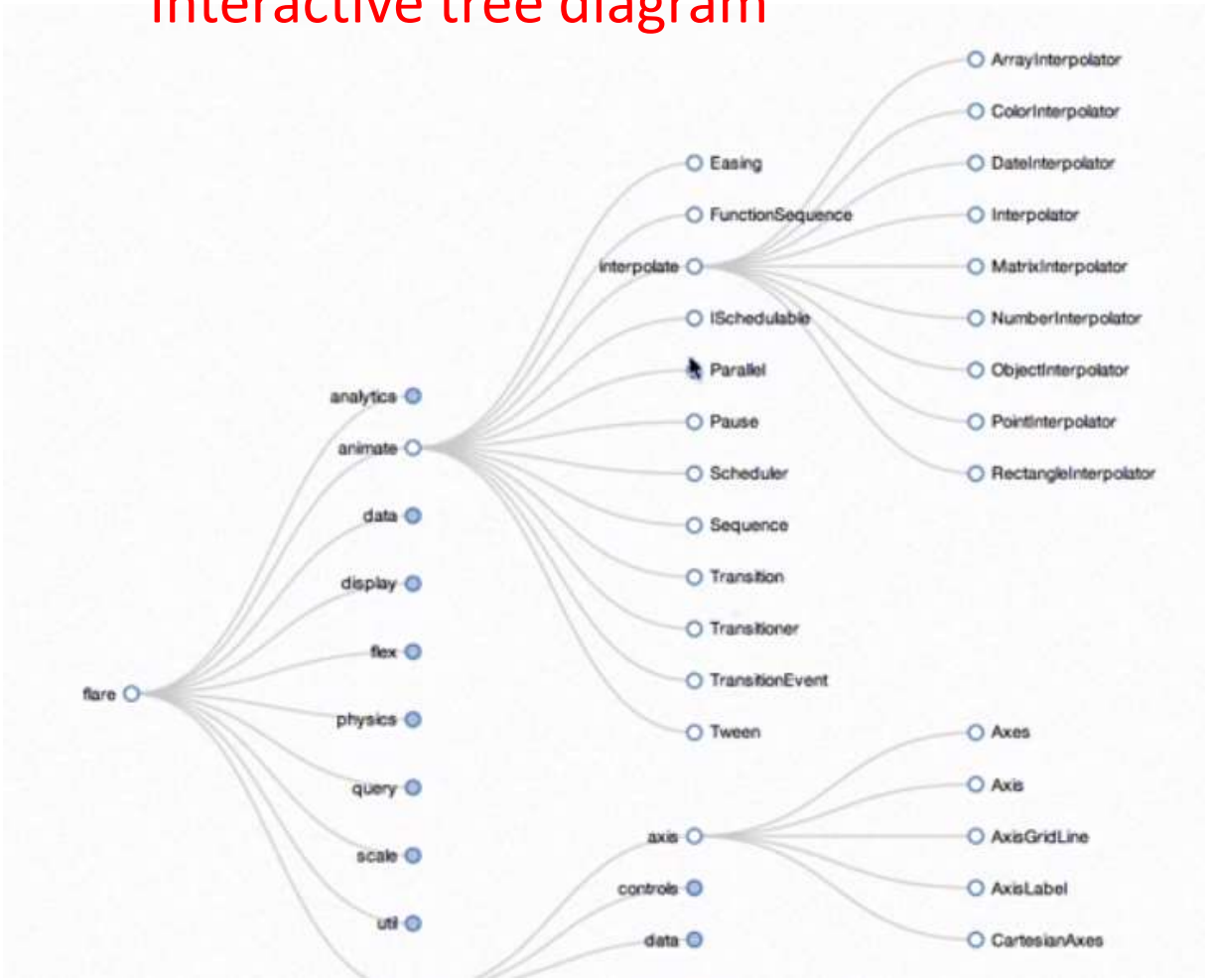
Scatter Matrix Chart

Scatter plot of each of these four related just to each other. So for instance, in this corner, I have sepal length, and sepal length. I have a row of sepal length and a column of sepal length.

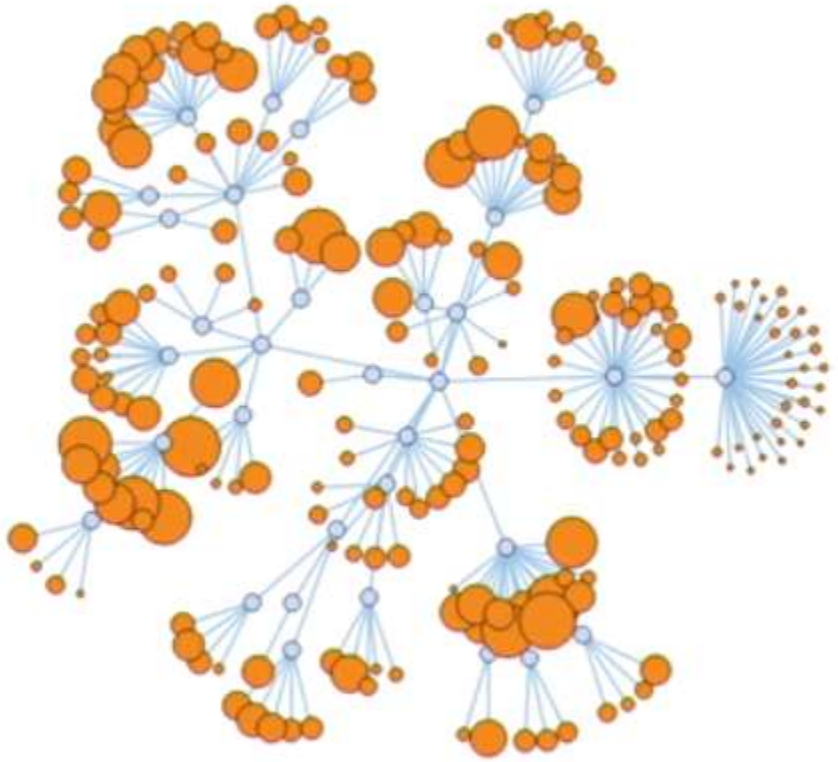


Data visualization is often about focusing on showing the connections between, and the hierarchy of objects. D3 is Data-Driven Documents. It's a JavaScript library, for doing data visualization (interactivity)

Interactive tree diagram



Node link diagram



Maps ……Google maps

So as a developer,
1- you have access to:

Satellite view
Street view

2- Ability to show places and do custom markers on those different places.

3- use the routing information, how do I get from point A to point B and Google maps will figure that out and how to show it.

4- Access to data visualization tool sets via the JavaScript API.

5-Ability to put objects on Google Maps in layers and show them interacting with real data and real time.

6- Ability to represent data as dots on the screen. And the size and the colors all represents something such as population density around the world.

7- Ability to do other interesting things with Google maps. Such a puzzle where you have different shaped objects on the map, and you have to click and drag them and put them into place. And if you get it right, it snaps into place, and changes color.

Five standard ways to show data on maps

1-Markers are for pointing points out of interest on a map.

iTouchMap.com and type in either the address or just a more general location like this, and it'll actually show you the lat, long, that coordinates for that particular region

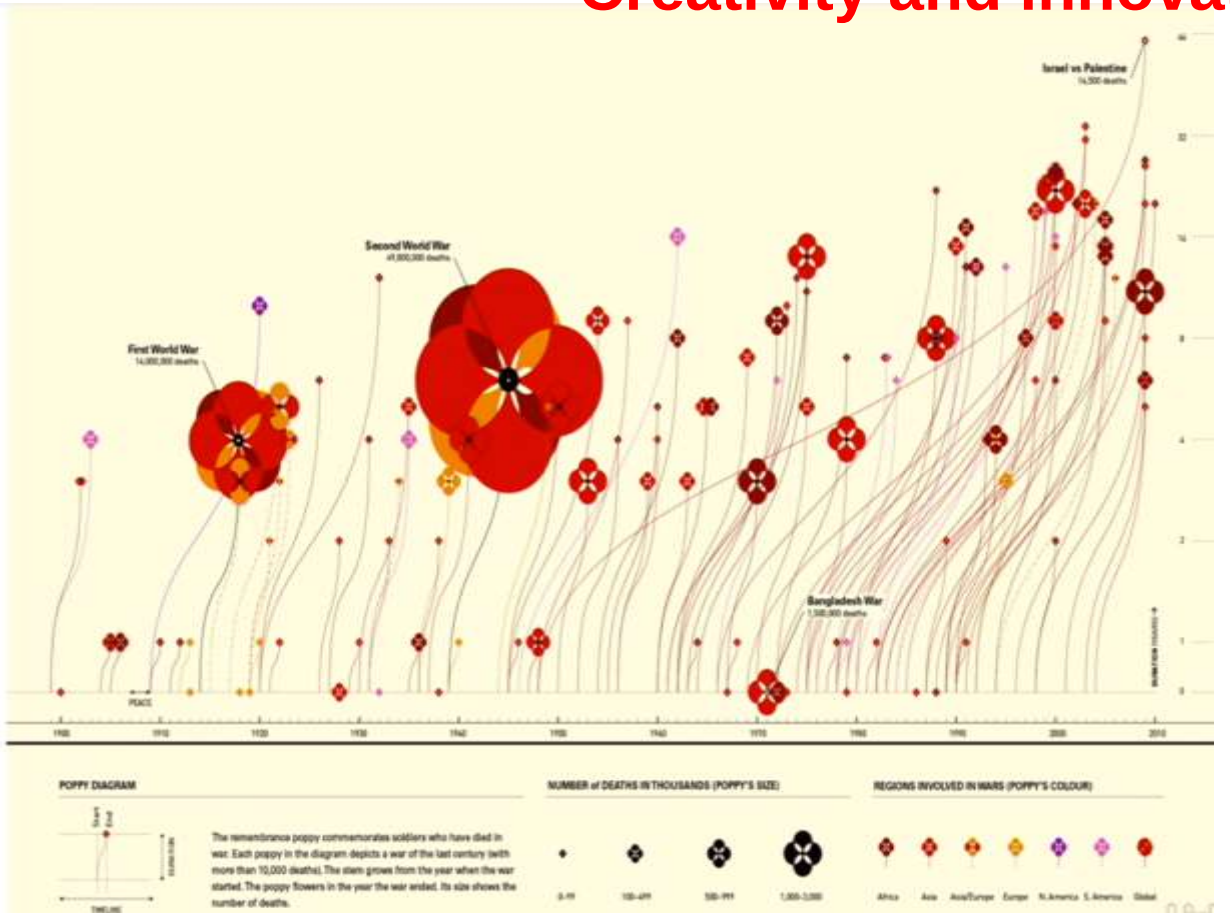
2-Another basic approach to showing data is layers on top of a map to indicate the data associated with the region. It's great for showing the data itself

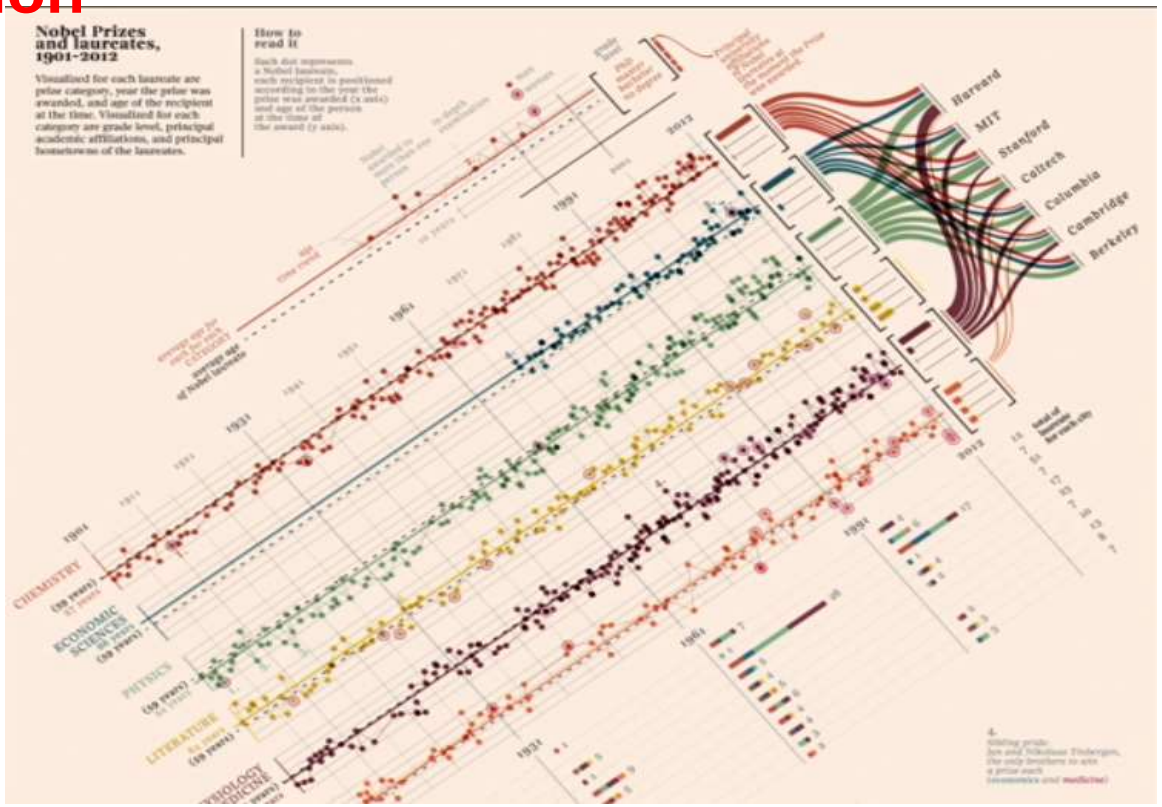
3-Another very common approach in mapping data is called a choropleth (voting red/blue)

4- Heat map. This is similar to a choropleth in that you're showing data associated with regions. But this is really great to show the concentration of data on fairly micro-level

5- Show flows between regions by putting arrows on things, (exports or imports of some goods between various countries)

Creativity and innovation





Typical data story Interactive chart

http://drones.pitchinteractive.com/

http://listen.hatnote.com/

https://demographics.virginia.edu/DotMap/index.html

categories

dotted line to indicate average age winner for that category.

If the dot has a little circle around it, that means it's a women, the other ones are men.

Summary: Challenge Objectives

1. Understand and reorganize your data.
2. Think about a visual approach.
3. Sketch in a form you could present.
4. Defensibility of ideas.

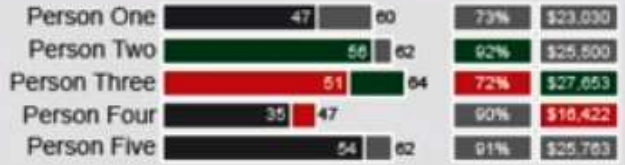
Acme Widgets

Sales Performance Ranking Report: May 2014

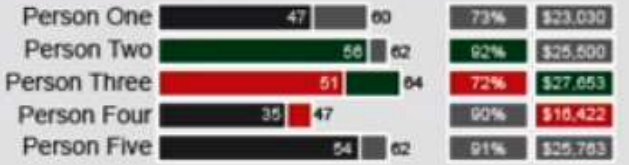
	Person One	Person Two	Person Three	Person Four	Person Five
Product Index	1.53	1.73	1.55	1.26	1.32
Combined Index	2.67	2.44	1.93	2.00	2.62
PRU	$1,118	$1,686	$1,087	$1,322	$1,447
Overall Rank	2	1	4	5	3

Instead of Excel sheet

Whatsits

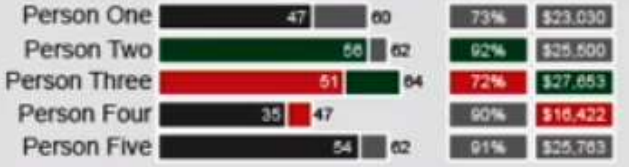


Doohickeys

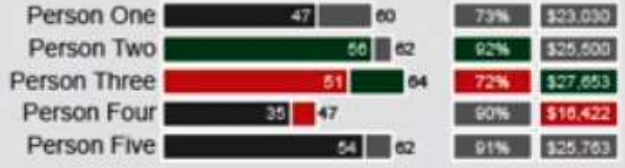


Gadgets

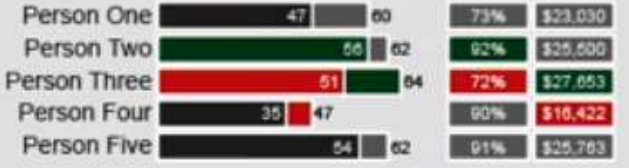
Bahbooms



Heehaws



Doodads



Adding motion with interactivity

http://msbarry.github.io/raceviz/2013-ashland-half-marathon/

Why?

Interactive feedback
Draw the eye
Show change overtime

When?

Always

Interactivity

1. Makes the result tangible and more memorable
2. Makes your data story into their data story
3. Increases Credibility
4. Get past editorial limitations

Why ?

Need of simplicity (good storytelling becomes a very crucial component of business intelligence (BI) tools)

narrative (Data storytelling goes beyond Data Visualization. A story with a real-life example or personal experience will help the end-users to find a context to the numbers and lend more meaning to the analysis)

Need to understand the logic behind it.

With the use of attractive visuals, right format and a strong narrative, decision makers will gain required clarity of the issue and understand the most appropriate action to be taken.

What are the skills required to be great Data storytellers?

1. Knowing the audience and adapting the story to their needs

One of the biggest mistakes in Data storytelling is making an assumption of one size fits all. The story has to be tailored to the sensibilities of the audience and their level of understanding. It is still acceptable if the story doesn't engage audience, but if they are unable to understand it, then there is a problem with storytelling. So knowing the target audience is of utmost importance.

2. Understanding the business problem

Before weaving a story to address a business problem, it is important for a Data Scientist to understand the problem very well. Without proper understanding of the issue, the scientist will not be able to present the story in the correct context. Storytelling is done with the sole purpose of expediting the decision making process, so it imperative that the business problem is broken down to the most granular level and to facilitate making a clear and concise story.

3. Identifying the probable questions and preparing answers

If the story presented is interesting enough, it is sure to attract loads of questions from the audience as well. As a good storyteller you need to be always prepared with the answers to the possible questions. A good story spurs inquisitiveness and the Data Scientist should have answers backed by relevant data to drive decision making.

4. Getting the right data at hand

You need to always ensure that you have the right data to tell your story. While a good and compelling story may be great for persuasion and striking an emotional chord, incorrect data will spoil its credibility. As a good storyteller, it is of utmost importance for you to balance beauty of a storyline alongside factual accuracy.

5. Good presentation skills

A good storyteller knows how to present the story with appropriate visualisations such as charts, graphs, maps, infographics etc. This makes it easier for the audience to spot trends, patterns and the get underlying message. Along with enhancing the visual appeal of the story, it makes persuasion easy.

The Final Word

Storytelling is not as easy as it appears. It takes a great deal of visualization, creativity and good presentation skills to ace the art of Data Storytelling. Most Data Scientists are trained to just stick to numbers and analytical skills and are not quite aware of the requirement to weave a good narrative around their analysis. However, it does pay to develop this critical skill because it adds a great deal of value to your analysis. Good Storytelling only helps decision makers to make sense of your analysis faster, be fully convinced with your findings and forecasts, and take the right course of action at the right time.

Dataset example :stock price from the NASDAQ 100 tech companies

The columns that we selected for our analysis from each year were:

1. **Headlines.Securities.Symbol :** We used this column to filter out and group our data according to the rows we needed for each specific company code.
2. **Headlines.Securities.CategoryorIndustry:** To understand which news articles were relevant according to their industry. For example, if a market sentiment was only for a specific set of companies or for all in general.
3. **Headlines.Title:** One of the important factors required for understanding the state of the market on that date.
4. **Headlines.Date:** To clearly segregate our news based on a particular month or date to understand what was the outlying sentiment in the timeframe.
5. **Headlines.Source:** To see which journal or news source was influential and prompt in reporting news about the company or stock.
6. **Headlines.Url:** To scrape news articles from the respective websites, because today's news titles tend to incompletely or very vaguely convey the complete sentiment of the content it encompasses.

1-Scripting down your ideas and flow before you start structuring your story

Aristotle’s classic five point plan that helps deliver strong impacts is:

1.Deliver a story or statement that arouses the audience’s interest.

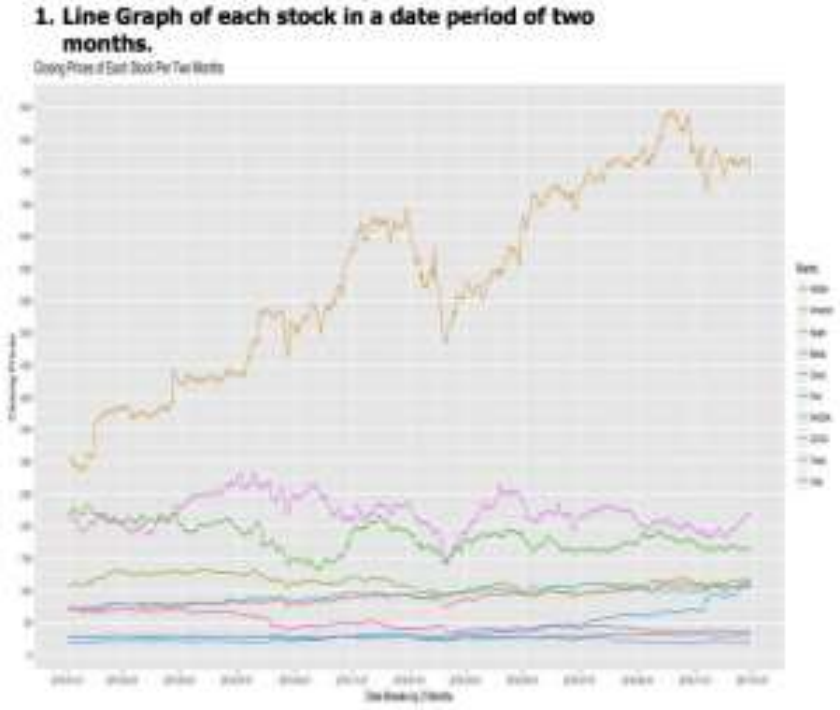
2.Pose a problem or question that has to be solved or answered.

3.Offer a solution to the problem you raised.

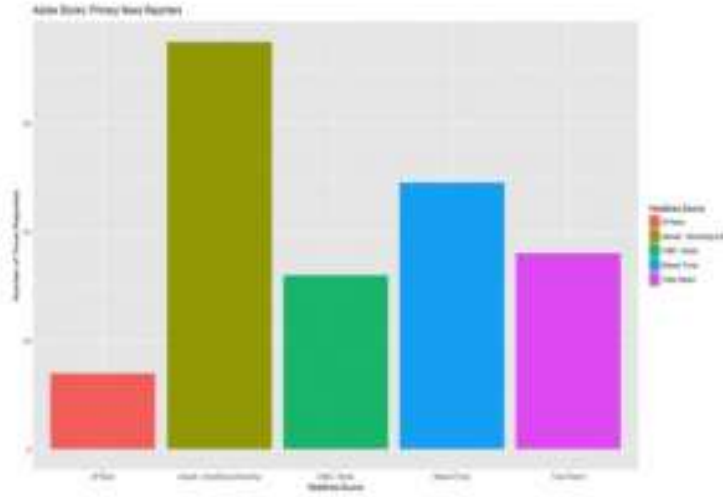
4.Describe specific benefits for adopting the course of action set forth in your solution.

5.State a call to action.

Line Graph To Find Out The Trend Lines Of Specific Stock Prices



Identifying which news source reported most about a particular stock, ➔is a good source for the specific stock.



2- How the story can help in making decision ?

Develop a personal “passion statement.”

Tell your prospects and why you are genuinely excited about working with them. Your passion statement will be remembered long.

Amazon is a good stock ➔analyze many companies

Lower priced stock vs higher value

Market analysis by dates

line plot analysis by date

Bar Plot use the right news sources to scrape from

3. Use powerful headings

Create your heading, a one-sentence statement for your story, visual, or analysis. The most effective headlines are concise, specific, and offer a personal benefit.

Remember, your heading is a statement that offers your audience a vision of a better understanding. It's not about you. It's about them.

4. Design a Road-Map

Create a list of all the key points you want your audience to know about your story, visual, or analysis

Categorize the list until you are left with only three major message points. This group of three will provide the verbal road map for your story.

Under each of your three key messages, add supporting evidences to enhance the narrative. These could include some or all of the following: personal stories, facts, examples, analogies etc.

Stock, volume, trend, evaluation

5. Conclude with brevity

Conclusion should be short and powerful

Hot picks for January –Mars 2017

Amazon

Adobe

Nvidia

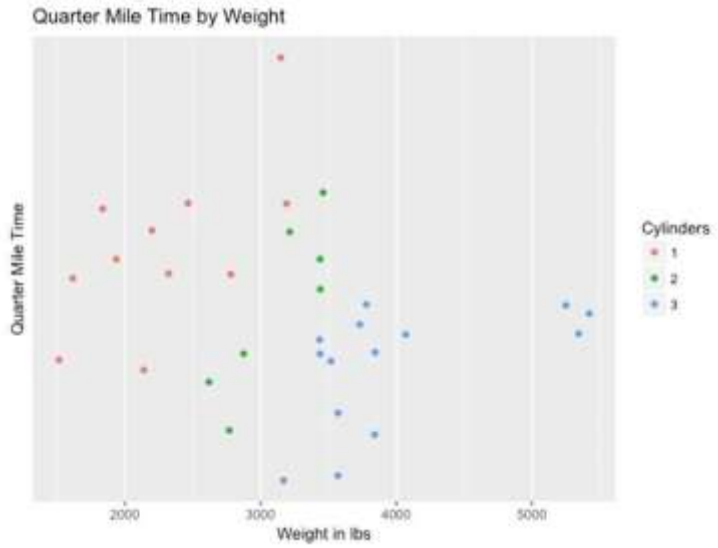
Type of Data vs chart type

1-textual data

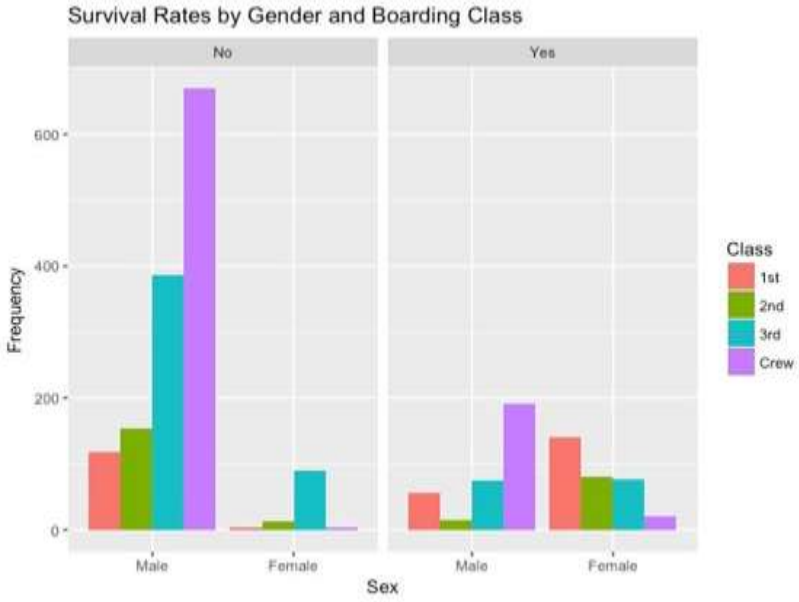
usually good to be finding how ofte has been used or what the sentime text is. Stories can be told best usir form of data

2. Mixed Data

Numeric or any other variety of formats, we need to know which ones are important and give us better insights from our dataset.
multivariate plot. The dataset in use for this plot is the Car Performance and Specifications dataset

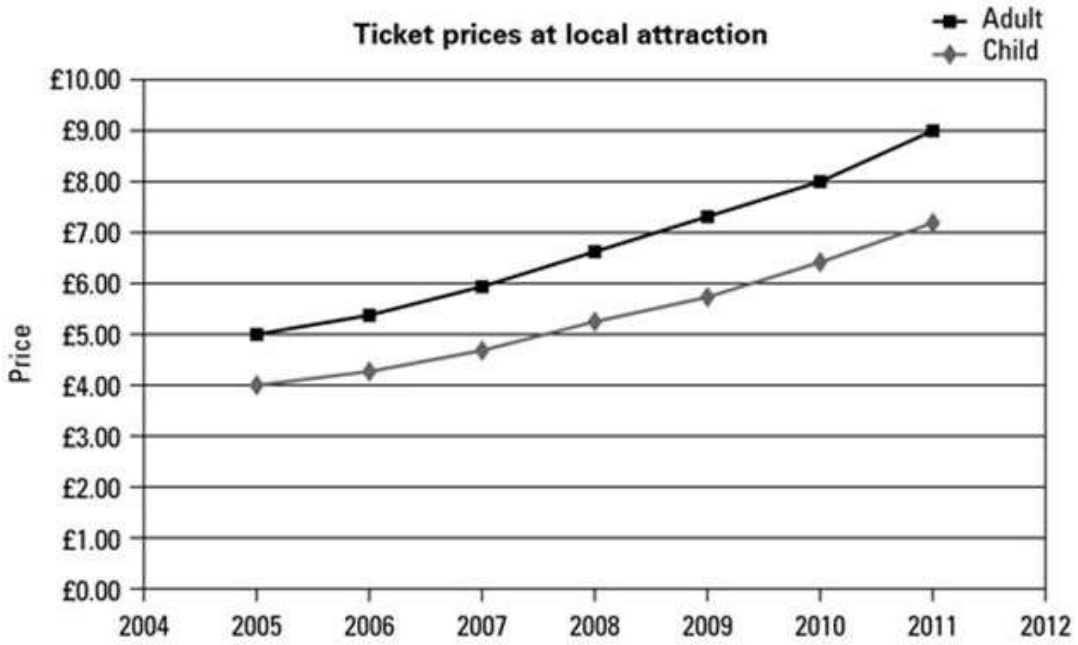


facet grids for the data



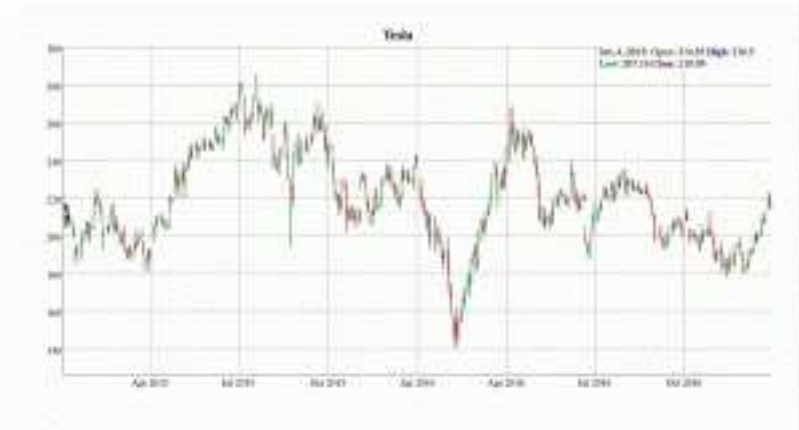
3. Numeric Data

usually looking for trends or lines that depict numbers. The visual that would suit numeric data best would be a line or a step graph.



4. Stocks

candlestick charts can be used to maneuver across each date



5. Geographic Data

data pertaining to specific locations and areas

Summary

During all stages of predictive modeling, storytelling could be a vital addition to any analysis.

1. Data Exploration

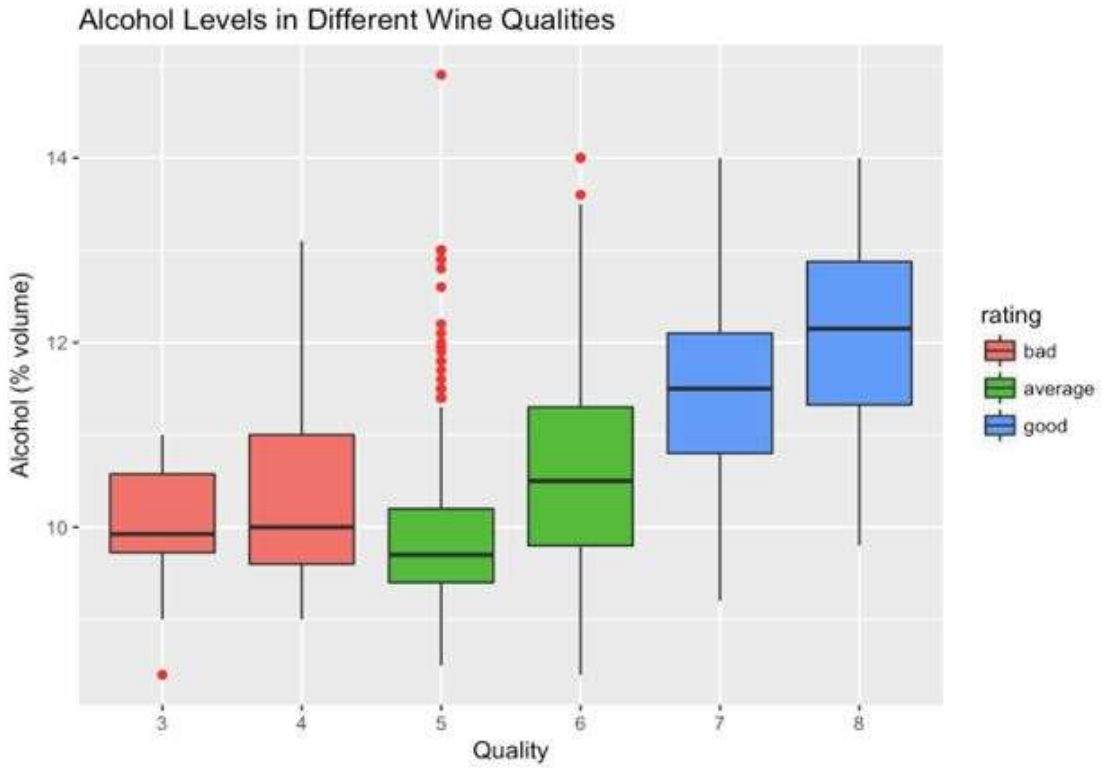
The first step of model building is understanding your data

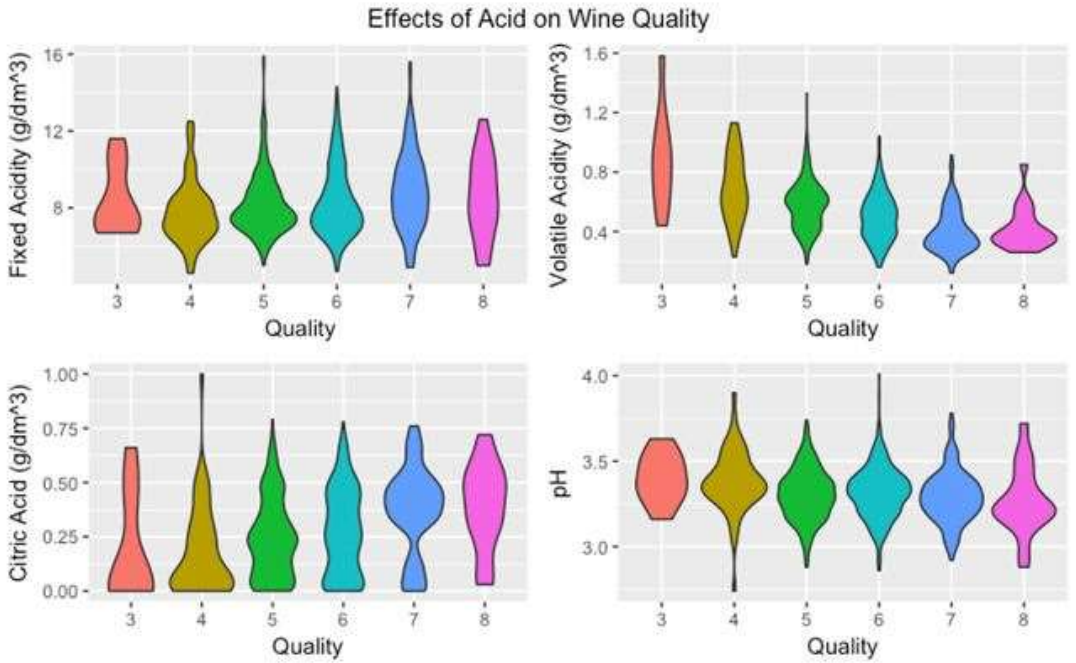
associated summary statistics (info(), describe()

For example Finding out how Wine Quality relates to Alcohol content.
there is any correlation between alcohol volumes and wine qualities, how do we do it?
We could either compute Pearson's 'r'. It would help us in building a model, but would not help us in analyzing much.

This shows a very strong correlation between Alcohol content and wine quality. But does it tell you anything else?
Ideally, it doesn't. So, what does?

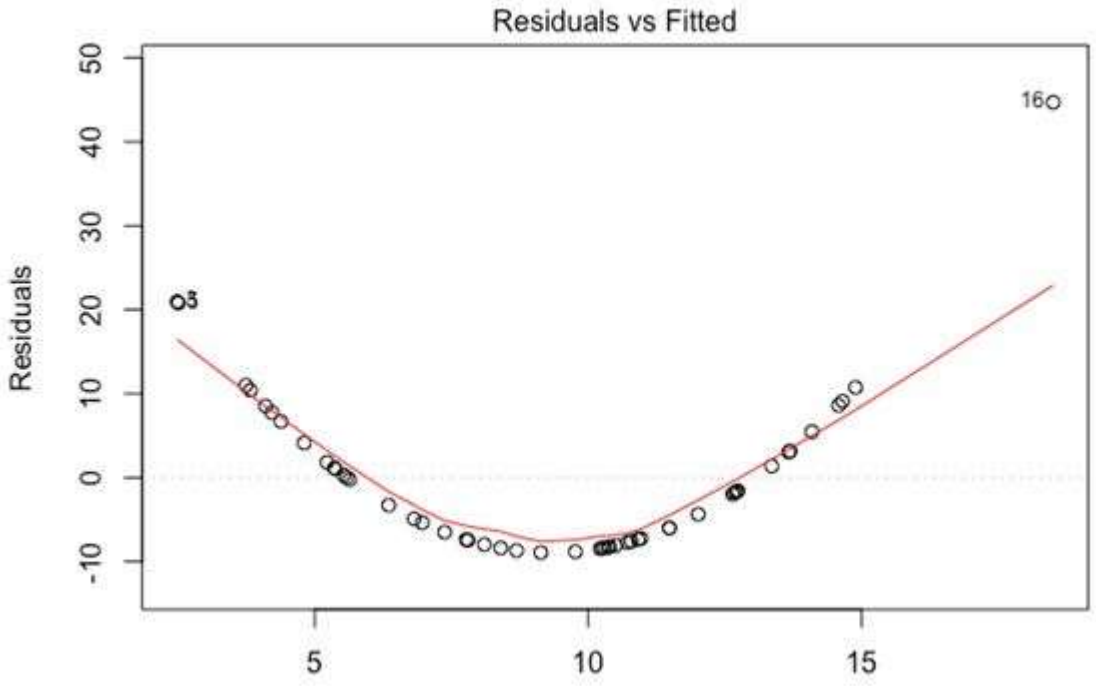
Use graphs


Alcohol Levels in Different Wine Qualities
Alcohol (% volume)
14
12
10
3
4
5
6
7
8
Quality
rating
bad
average
good

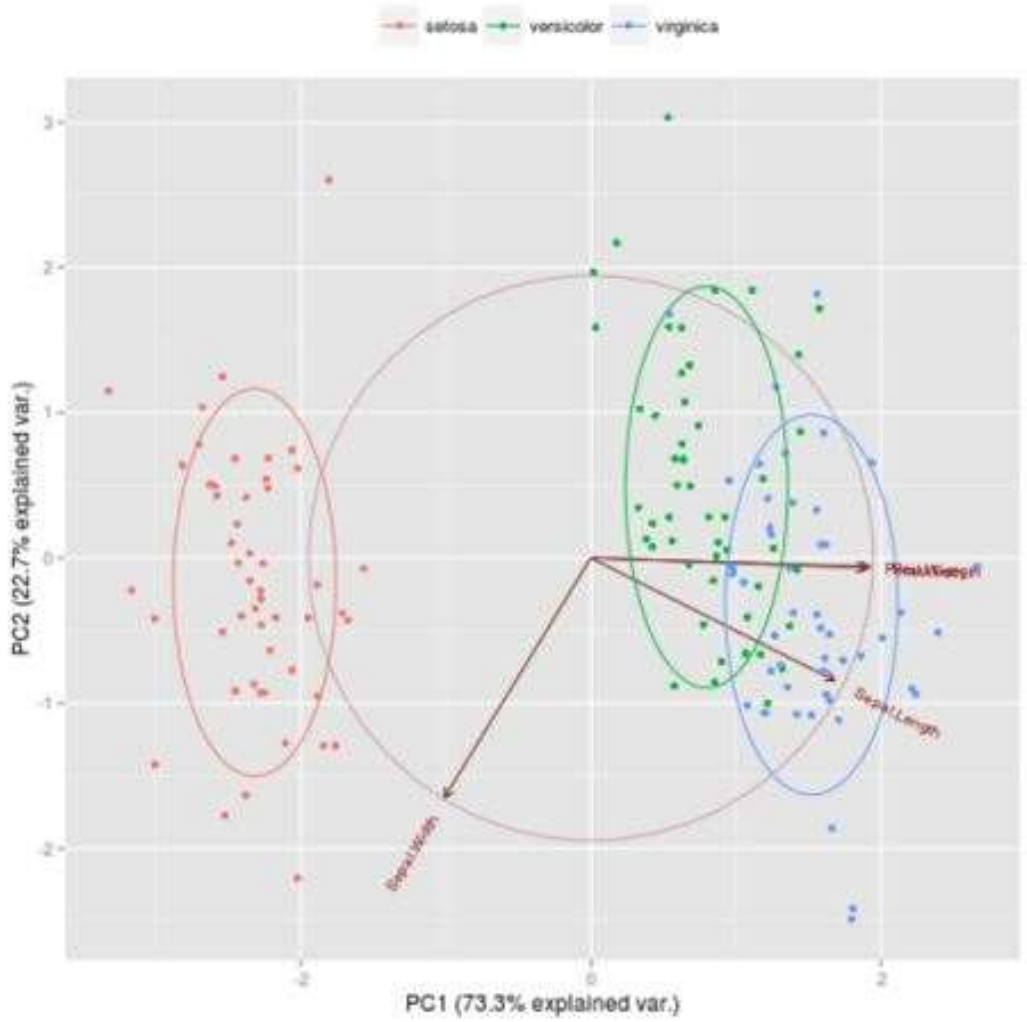

Effects of Acid on Wine Quality
Fixed Acidity (g/dm^3)
16
12
8
Volatile Acidity (g/dm^3)
1.6
1.2
0.8
0.4
Citric Acid (g/dm^3)
1.00
0.75
0.50
0.25
0.00
pH
4.0
3.5
3.0
3
4
5
6
7
8
Quality

2. Feature Visualizing

After you generate features, how do you see how well one is predicting?



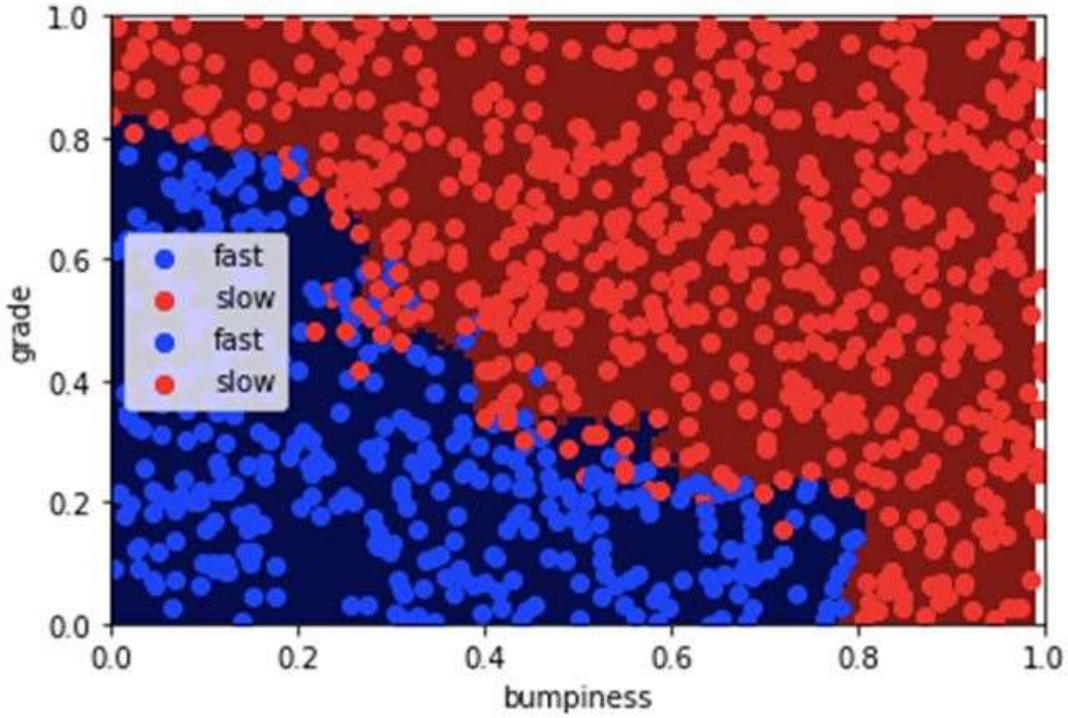
Principal Component Analysis



3. Model Creation and Comparison

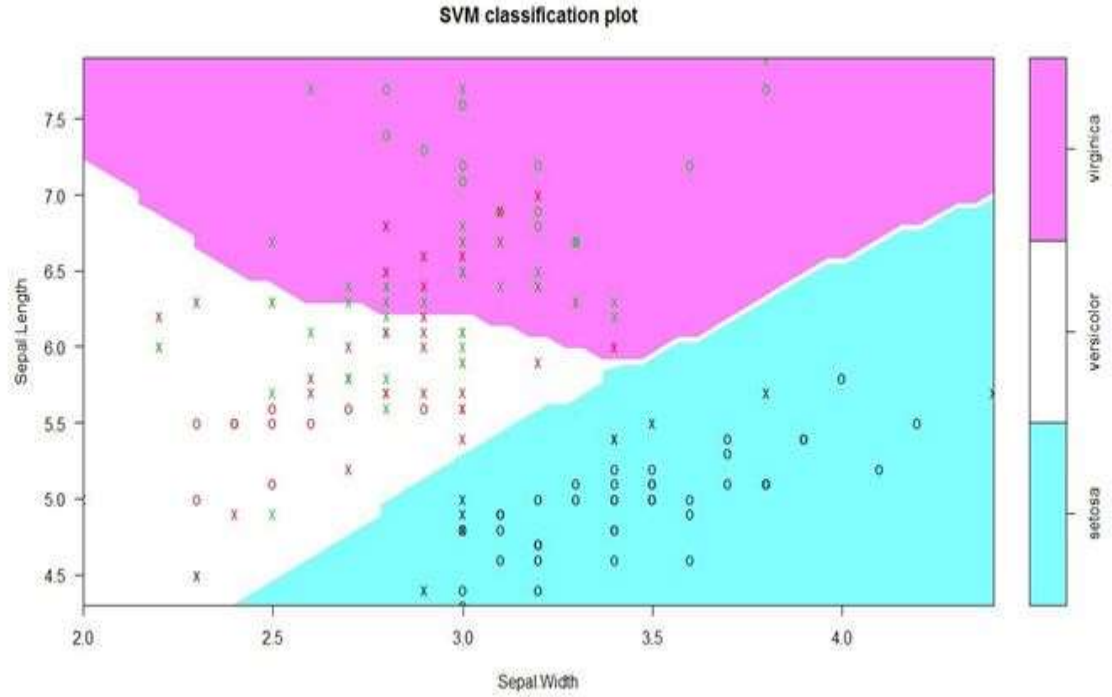
Coming to the model creation phase, we usually find the need to understand how our data is being fitted

Model that predicts whether the car should go fast or slow, based on the grade of the road and bumpiness.



Classifies most of the data but an accuracy of 88.21% doesn't tell much of a story

Plot shows us a clear classification boundary where the Species separate from each other.



Best Practices for Story Telling

•Always **label your axes and give the heading** of your plot.
•Use **legends** where necessary.
•Use **colors that are lighter** on the eye and in proportion.
•**Avoid adding unnecessary detail** to your visualization like backgrounds or themes that don't allow good readability.
•**Only a point can be used** to simultaneously encode two quantitative values based on a horizontal and vertical location.
•**Never use points** for visualization if you are doing time series encoding.

•Understand the importance of context
•Determine the appropriate type of graph
•Recognize and eliminate the clutter
•Direct your audience's attention
•Think like a designer when visualizing data
•Leverage the power of storytelling to help your message resonate with your audience

Use stories and charts to uncover

- Insights from data that you might have missed before.
- Relations between features and data that numbers can never clearly depict