Telling an impactful story through Data Visualization

WORKSHOP OVERVIEW

- How it works
- Exploring your data
- Understanding data types.
- Visualization choices and considerations
- Types of scales
- MVD (Minimum Viable Dataset)
- Hands on session

HANDS-ON AGENDA

- Data exploration/information extraction
- Creating a narrative – you're not *always* the expert
- Geolocating information (long, lat)
- Quantifying non numeric values
  - General visualization libraries
- Combining visualization (Or when not to)
- Visual Critique

Hands on Repo: https://github.com/couellette/workshop

https://github.com/couellette/workshop

TECHNOLOGY REQUIREMENTS

Excel or Libre Office or Google Sheet

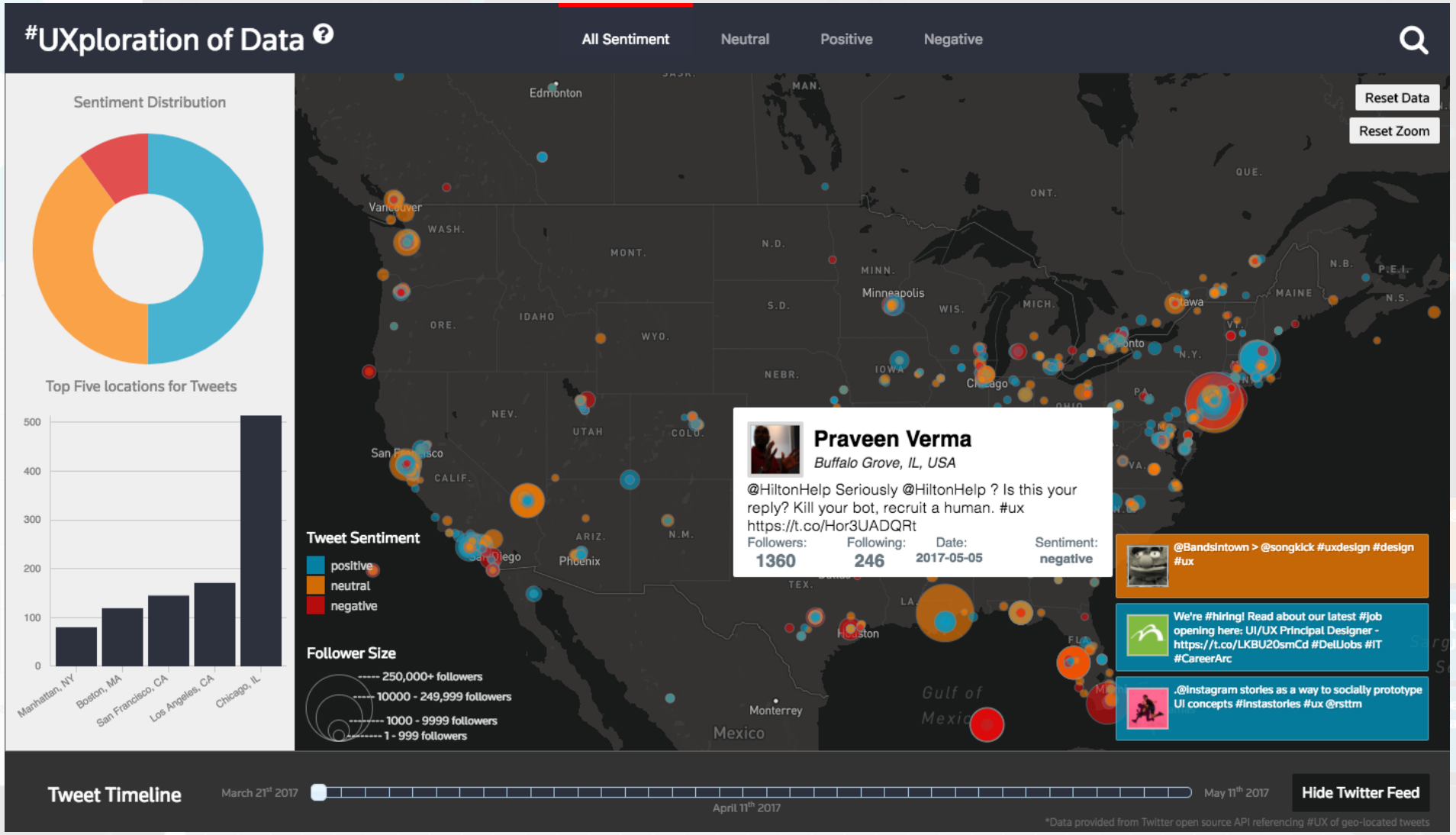
Sublime Text or Atom

Or use *jsfiddle

END GAME

What your end-user sees

END GAME – CLEAN VISUALIZATIONS



END GAME – THE MANICURED DATA SET

- Clean
- Structured
- Organized
- Low margin of error
- Audience specific

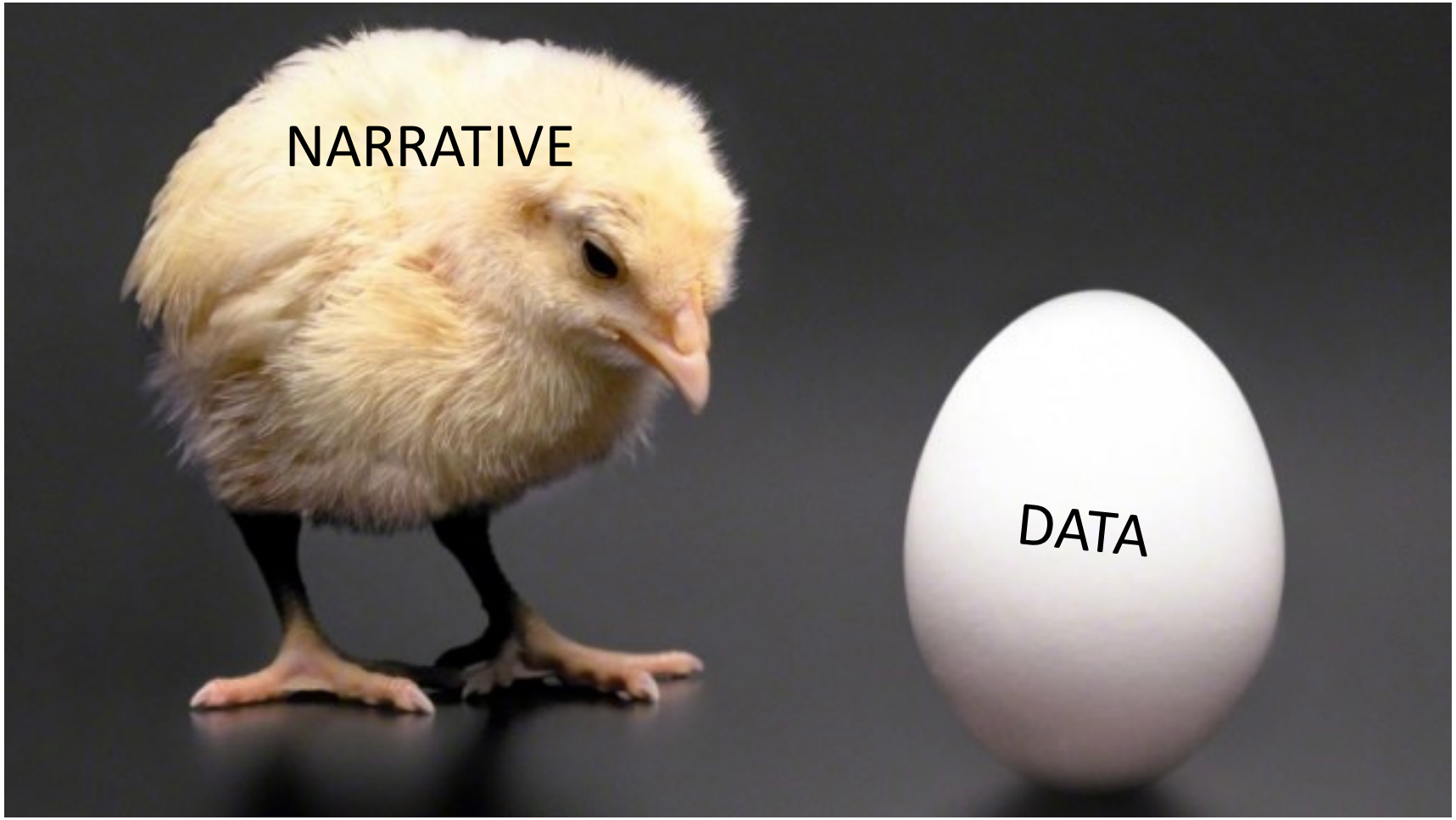
A1 | day

	A	C	D	E	F	G	H	I	J	K
1	day	lng	lat	place_country	place_full_name	place_place_	user_favourites_count	user_name	user_descrip	user_locatio
2	Tue Mar 21	-86.785881	34.578426	United States	Huntsville, AL	city	3114	saketvora	"Hardware (	Santa Cruz, (
3	Tue Mar 21	-122.288439	37.533497	United States	Foster City, CA	city	0	Join Philips	"Live Job Fe	Across North
4	Tue Mar 21	-87.940033	41.644102	United States	Chicago, IL	city	10	CSGI Careers	"CSG Interna	Englewood,
5	Tue Mar 21	-75.586247	39.108566	United States	Dover, DE	city	821	Nneka..._ëÑ	"Yesterday i	DSU19
6	Tue Mar 21	-122.514926	37.708075	United States	San Francisco, CA	city	614	Matthew Ta	"Digital Arch	iPhone: 40.7
7	Tue Mar 21	-93.207783	44.890752	United States	St Paul, MN	city	9895	amandaram	"UX designe	minnesota
8	Tue Mar 21	-122.514926	37.708075	United States	San Francisco, CA	city	2226	Colin Johnst	"Graphic De	San Francisc
9	Tue Mar 21	-84.576827	33.647503	United States	Atlanta, GA	city	0	Atlanta Web	"Follow this	Atlanta, GA
10	Tue Mar 21	-87.940033	41.644102	United States	Chicago, IL	city	10	CSGI Careers	"CSG Interna	Englewood,
11	Tue Mar 21	-76.353876	44.961937	Canada	Ottawa, Ontario	city	2041	Marc Gagnb	"Tech evang	Ottawa
12	Tue Mar 21	-76.762559	39.942947	United States	York, PA	city	73	george corzi	"UX Design -	local
13	Tue Mar 21	-79.639319	43.403221	Canada	Toronto, Ontario	city	122	stickler-ci	"Automatica	Toronto, On
14	Tue Mar 21	-87.940033	41.644102	United States	Chicago, IL	city	10	CSGI Careers	"CSG Interna	Englewood,
15	Tue Mar 21	-118.668404	33.704538	United States	Los Angeles, CA	city	0	LA Marketin	"Follow this	Los Angeles,
16	Tue Mar 21	-115.209254	35.984784	United States	Paradise, NV	city	404753	Evan Kirstel	"#B2B #Solo	Boston, MA
17	Tue Mar 21	-115.209254	35.984784	United States	Paradise, NV	city	192	Dave Schnei	"Chief Mark	Minneapolis
18	Tue Mar 21	-115.209254	35.984784	United States	Paradise, NV	city	192	Dave Schnei	"Chief Mark	Minneapolis
19	Tue Mar 21	-86.348441	39.631677	United States	Indianapolis, IN	city	21485	Cori Faklaris	"Social medi	Indianapolis
20	Tue Mar 21	-87.940033	41.644102	United States	Chicago, IL	city	10	CSGI Careers	"CSG Interna	Englewood,
21	Tue Mar 21	-79.639319	43.403221	Canada	Toronto, Ontario	city	8999	Say Yeah!	"We drive b	Toronto
22	Tue Mar 21	-71.296465	42.564662	United States	Tewksbury, MA	city	338	HawkinsPoir	"Using techr	Chelmsford,
23	Tue Mar 21	-94.733122	38.868002	United States	Kansas City, MO	city	2582	Marjie Gooc	"Digital com	Kansas City
24	Tue Mar 21	-82.620093	27.821353	United States	Tampa, FL	city	272	USF Digital M	"Official Twi	Tampa, FL
25	Tue Mar 21	-82.620093	27.821353	United States	Tampa, FL	city	272	USF Digital M	"Official Twi	Tampa, FL
26	Tue Mar 21	-122.117916	37.356771	United States	Mountain View, CA	city	2543	christina	"Do I contra	palo alto
27	Tue Mar 21	-114.316038	50.84278	Canada	Calgary, Alberta	city	2297	Clarissa Pete	"UX designe	Calgary, Albe
28	Tue Mar 21	-122.514926	37.708075	United States	San Francisco, CA	city	2308	Lane Goldste	"Co-founder	Brooklyn & I
29	Tue Mar 21	-122.514926	37.708075	United States	San Francisco, CA	city	532	Billie Mande	"nerdly spar	Oaktown, CA
30	Tue Mar 21	-87.940033	41.644102	United States	Chicago, IL	city	10	CSGI Careers	"CSG Interna	Englewood,
31	Wed Mar 22	-93.329515	44.889964	United States	Minneapolis, MN	city	239	Ben Wallace	"Marketing S	MN
32	Wed Mar 22	-84.576827	33.647503	United States	Atlanta, GA	city	174	Unbreakable	"Diversity is	Albany, Ga.
33	Wed Mar 22	-71.191421	42.227797	United States	Boston, MA	city	34	Jobs at Cont	None	None
34	Wed Mar 22	-87.940033	41.644102	United States	Chicago, IL	city	10	CSGI Careers	"CSG Interna	Englewood,

HOW IT WORKS

"In Data Visualization it's not always the end deliverable that's most important – *its how you got there*"

HOW IT WORKS – IT ALL STARTS WITH DATA



HOW IT WORKS – IS IT ACCURATE?

- Is it statistically accurate?
  - *What is your sample size?*
  - *What is the margin of error? https://www.surveymonkey.com/mp/margin-of-error-calculator/*
- Are you looking at a long enough timeframe?
- When you use a sample to represent an audience, you must make sure that the people in your sample are representative of the audience.
- Always spot check your data when combining multiple datasets for errors

HOW IT WORKS – IS YOUR DATA CLEAN?

- Unstructured text vs. structured text
- File format (csv, tsv, txt, json…)
- Encoding(vs. 😁)

HOW IT WORKS – ARE THERE MISSING PARTS?

- Gaps in time
- Are you exploring all areas of where you can get data?
- Can you aggregate other data sources

EXPLORING YOUR DATA

Unearthing the narrative of your *visualization*

EXPLORING YOUR DATA – WHY IS THERE MISSING DATA?

- Not all datasets are complete
- Some information is confidential
- Consider ways to show missing data

EXPLORING YOUR DATA – VIEWING THE DATA IN DIFFERENT WAYS

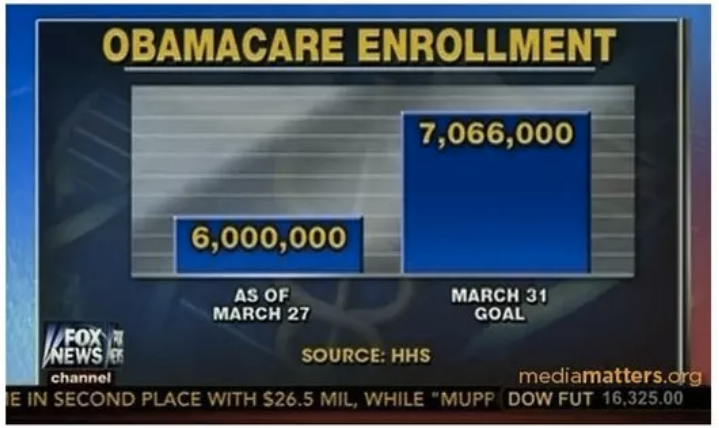
- Review your data – are there trends or patterns?
  - *Time*
  - *Location*
  - *Influence*
  - *Company size*
- Sentiment Analysis – what is the feeling of the data?
- Consider viewing data with different diagrams/visualization

EXPLORING YOUR DATA – ARE YOU TWISTING THE NARRATIVE?

- Hiding the greater picture
  - *Only focusing on a specific time can leave out important comparison information*
  - *Excluding data points to give greater merit to a topic*

- Skewing Visualization in favour of your narrative
  - *Visualizing data can tell false truths when information is not being accurately displayed*

Obamacare Signups, According to Fox News



Source: Media Matters of America

UNDERSTANDING DATA TYPES

NUMERIC VALUES

- Financial
- Census/population
- Aggregated non-numeric values
- Dates/time
- Percentage

When to use:

- Bar graphs
- Scatterplots
- Line graphs
- Tables/data points
- *Pie/Donut

STRING /RESPONSE TEXT

- Open-ended response
- Tweets
- Address information
- Form data

When to use:

- Word cloud
- Text snippet
- Tooltip
- Aggregate information
  - *Sentiment analysis
  - *Natural language processing

SPACIAL / GEOGRAPHIC

- Longitude and Latitude
- IP address
- Line interpolation
- Directions
- Traffic data

When to use:

- GPS
- Points of interest
- Location tracking
- Event detection
- Point to point travel

GROUPED DATA SETS

- API access to consolidated data
  - Twitter
  - RSS
  - News feeds
  - Analytics data
- Survey data
- Databases
- Customer data

When to use:

- Dashboards
- Reports
- Customer analytics

VISUALIZING YOUR DATA

Illustrating Insight

VISUALIZING DATA - CONSIDERATIONS

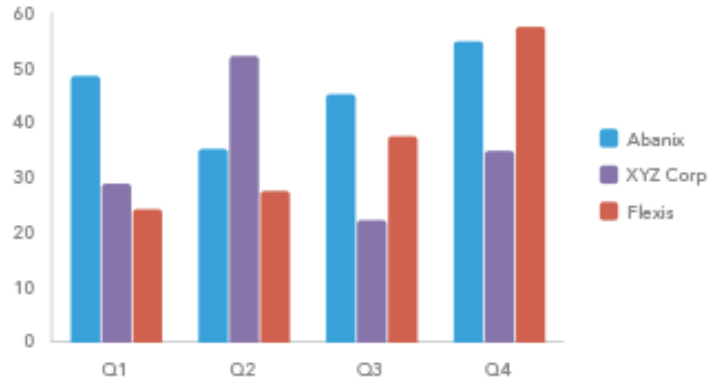
- Who are your audience?
- How much detail do they need?
- What is the margin of error in your data?
- Is the dataset telling the whole story?
- What story does the data tell?
- Why do you need a visualisation?

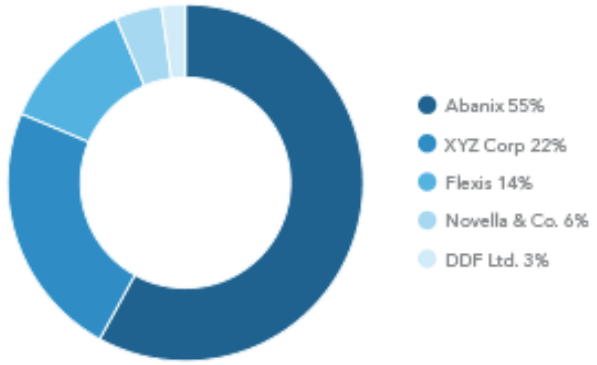
VISUALIZING DATA – HOW TO?

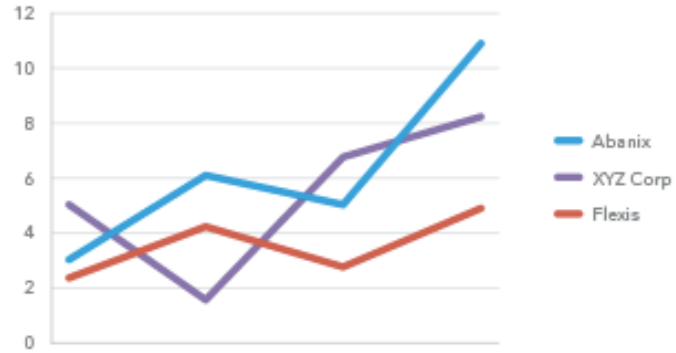
- Test different ways of visualizing your data
  - *Is this for analysis or for story telling?*
  - *Multiple ways of seeing the same information can help reinforce*
- Consider the scales and dimensions on what your visualizing
- Leverage the use of *colours

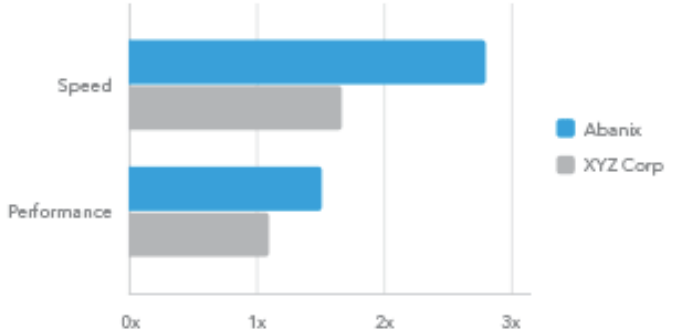
VISUALIZING DATA – CONSIDERATIONS

- Clear
- Specific
- Keep it simple
- To the point
- Inline with audience



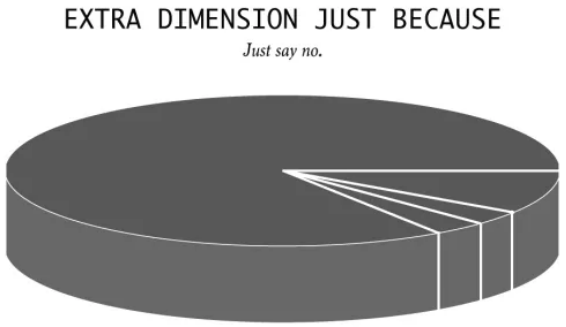


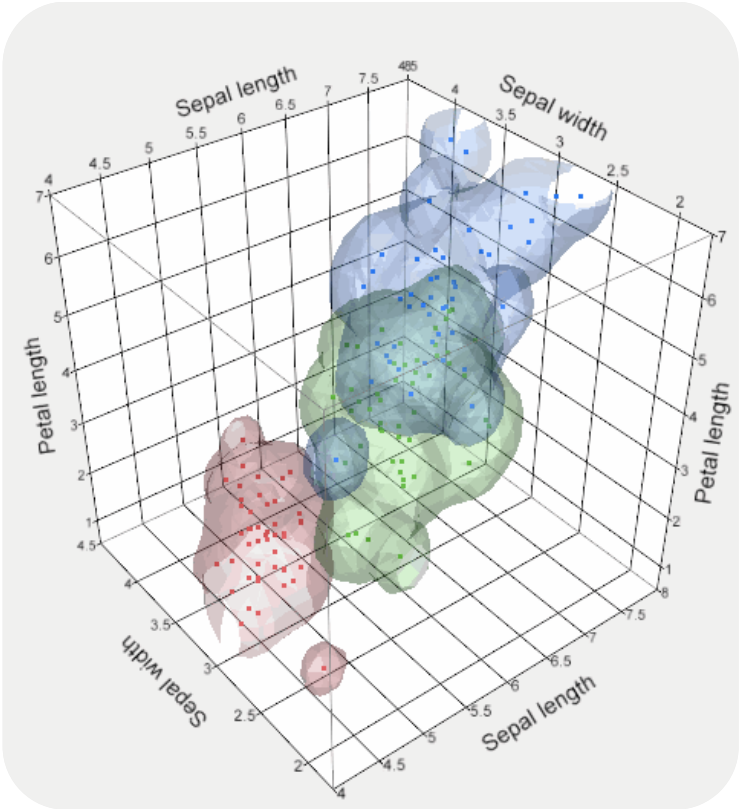




VISUALIZING DATA – THINGS TO AVOID

- 3D – novelty vs. insight
- Visualizing data for the sake of visualizing
- Over complicating the information
- Avoid graphical distortion – pick the right scale
- Too many colours (no more than 6)
- Reduce the need for math

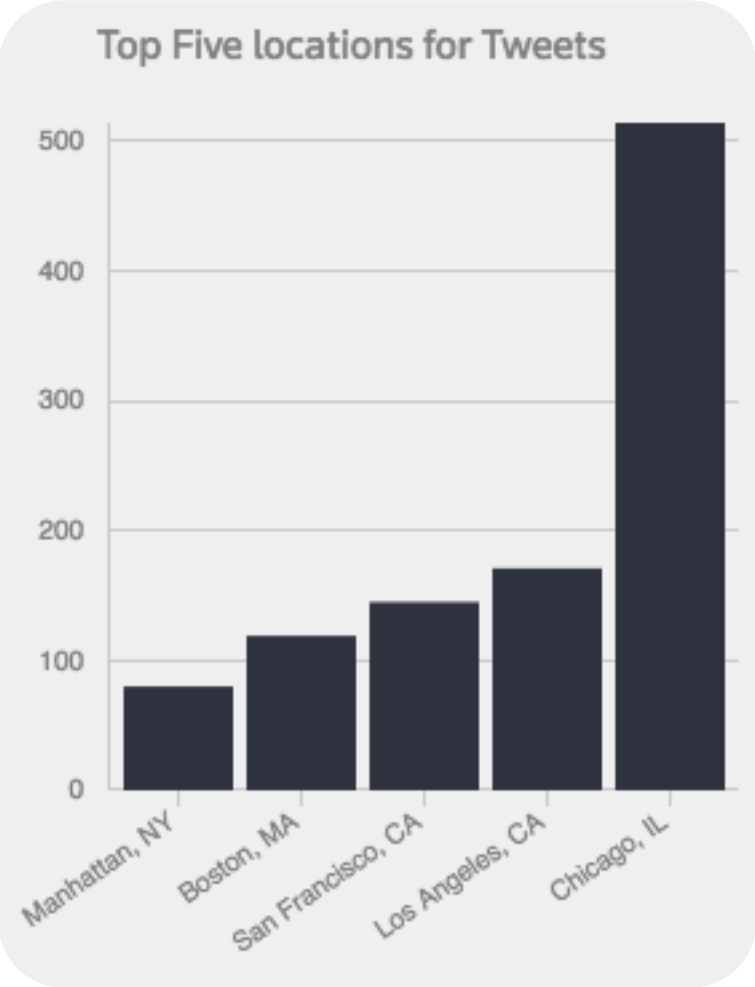




TYPES OF SCALES

Don’t open the opportunity for *misinterpretation*

TYPES OF SCALES – LINEAR SCALE



- Equally distributed
- Based on addition

✓Best used with less complicated data

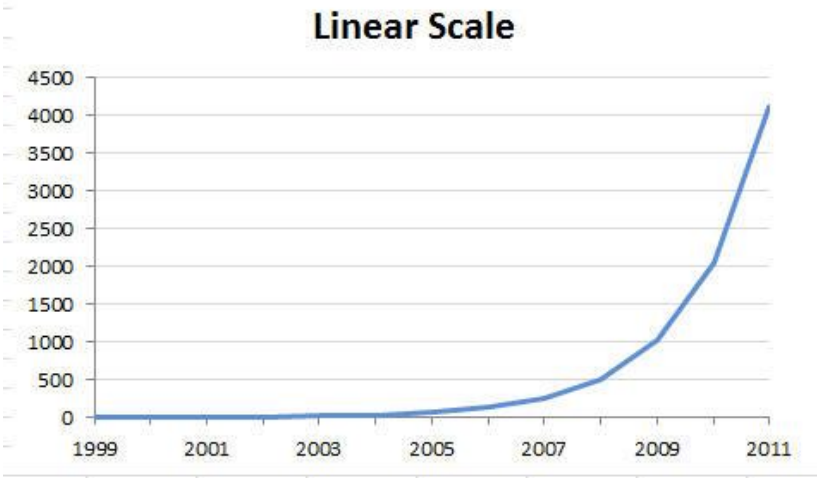
✓Simpler graph visualization

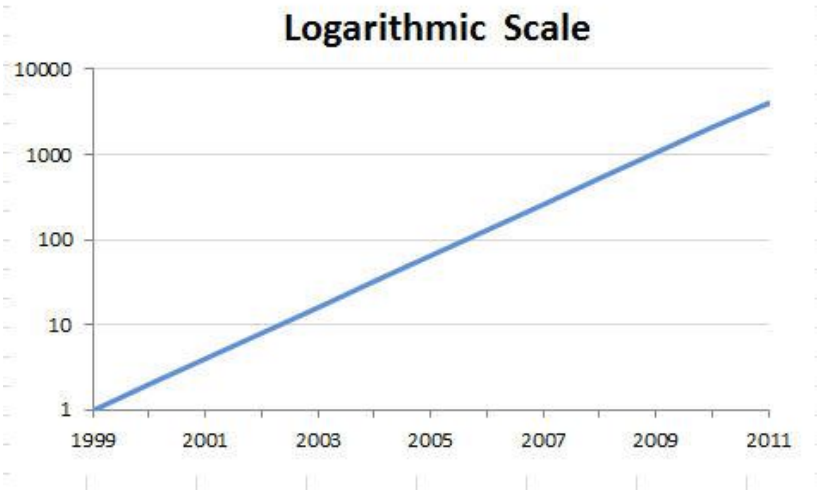
TYPES OF SCALES – LOGARITHMIC SCALE

- Next point is multiplied by point previous
- Based on multiplication

✓Best used when there is a large skew in value with the scale

✓Larger sets of data

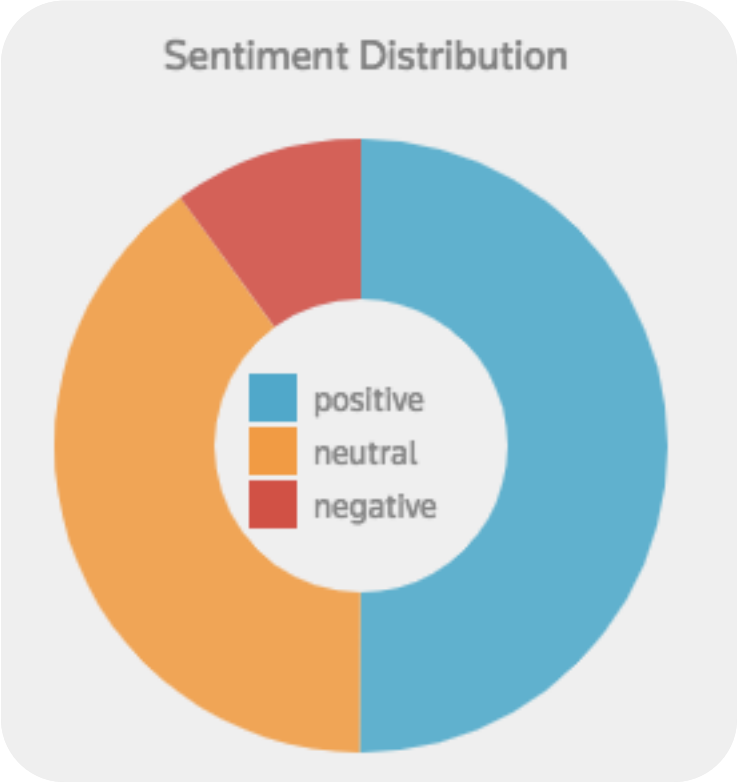




TYPES OF SCALES – CATEGORICAL SCALE

- Text/Word based measurement
  - *Ex. Satisfied vs Unsatisfied*
  - *Gender*
  - *Location*

- Combined with other types of scale
- Group text into numeric values

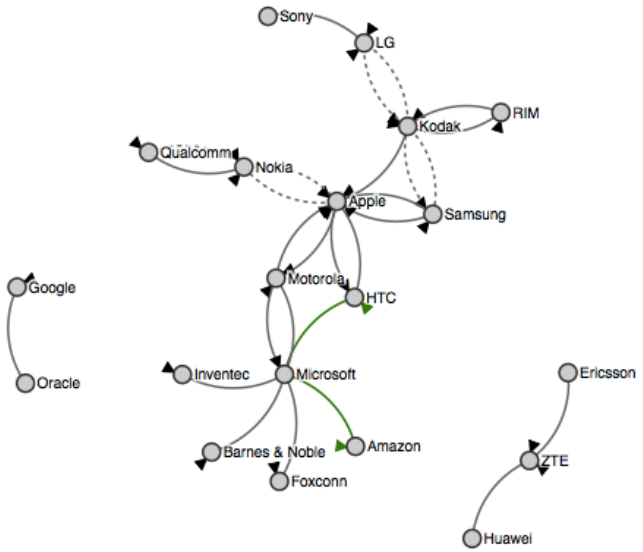


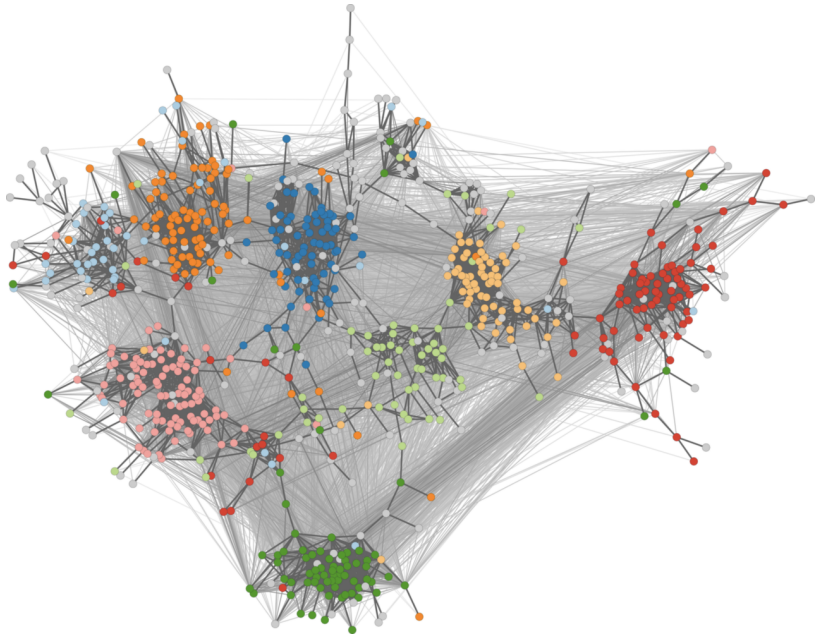
MVD (MINIMUM VIABLE DATASET)

Extraction of information while maintaining
accuracy

MVD - CONSIDERATIONS

- Avoid overloading a visualization
- If it doesn't add value – remove it
- Not specific to narrative/end-user consider removing





HANDS ON

Lets go *explore*

HANDS-ON AGENDA

- Data exploration/information extraction
- Creating a narrative – you're not the expert
- Geolocating information (long, lat)
- Quantifying non numeric values
  - General visualization libraries
- Combining visualization (Or when not to)
- Visual Critique

Hands on Repo: https://github.com/couellette/workshop

TECHNOLOGY REQUIREMENTS

Excel or Libre Office

Sublime Text or Atom

DATA CLEANING AND EXPLORATION

ux-data.csv

DATA CLEANING

- Look for empty fields
- Encoding – are there garbled characters?
- Is all the data relevant?
- Source checking
  - Accuracy of data
  - Quality of data
  - Privacy or confidentiality concerns
- Standardize information
- Structuring/Organize the data

DATA EXPLORATION

- Are there trends in the data?
  - Start initial visualizations in excel/simple graph tools
- Are there outliers?
- Combine datasets
  - Corresponding data that would be of value for comparison
- Deeper Analysis
  - Sentiment analysis
  - NLP

VISUALIZATION LIBRARIES

D3

PROS

- Large community of developers
- Mass customization
- Huge library

CONS

- Relatively large learning curve
- Require knowledge of JavaScript, HTML, and CSS

TABLEAU

PROS

- Lower learning curve compared to other technologies
- Free to try
- Out of the box ready

CONS

- Paid version required for commercial user
  - Requires a server for larger companies
- Not as easy to use outside of CSV and excel

DESIGN SOFTWARE



PROS

- Complete flexibility of abstract design
- No code knowledge required
- Smaller learning curve

CONS

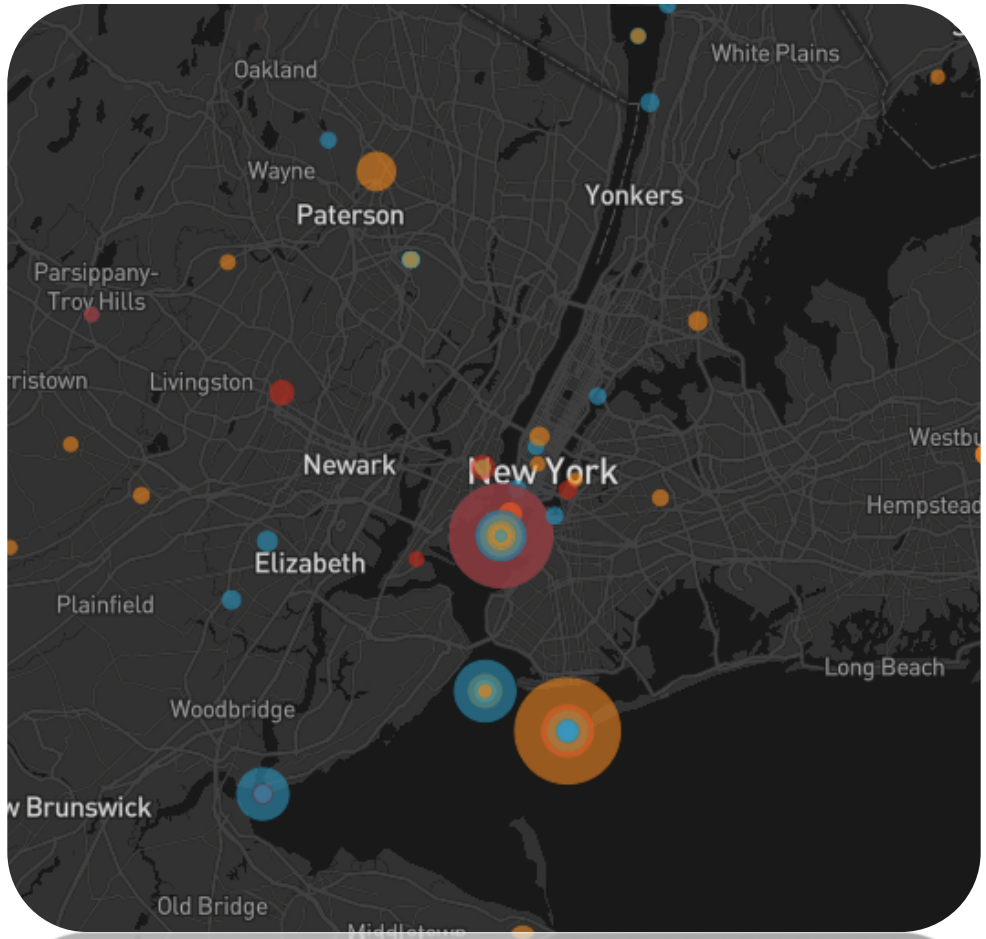
- Limited use beyond demonstrations
- Not interactive
- Needs to be handed of to a developer to build concepts

GEOLOCATING INFORMATION

What do Bigfoot and a map key have in common?

….they're both legends

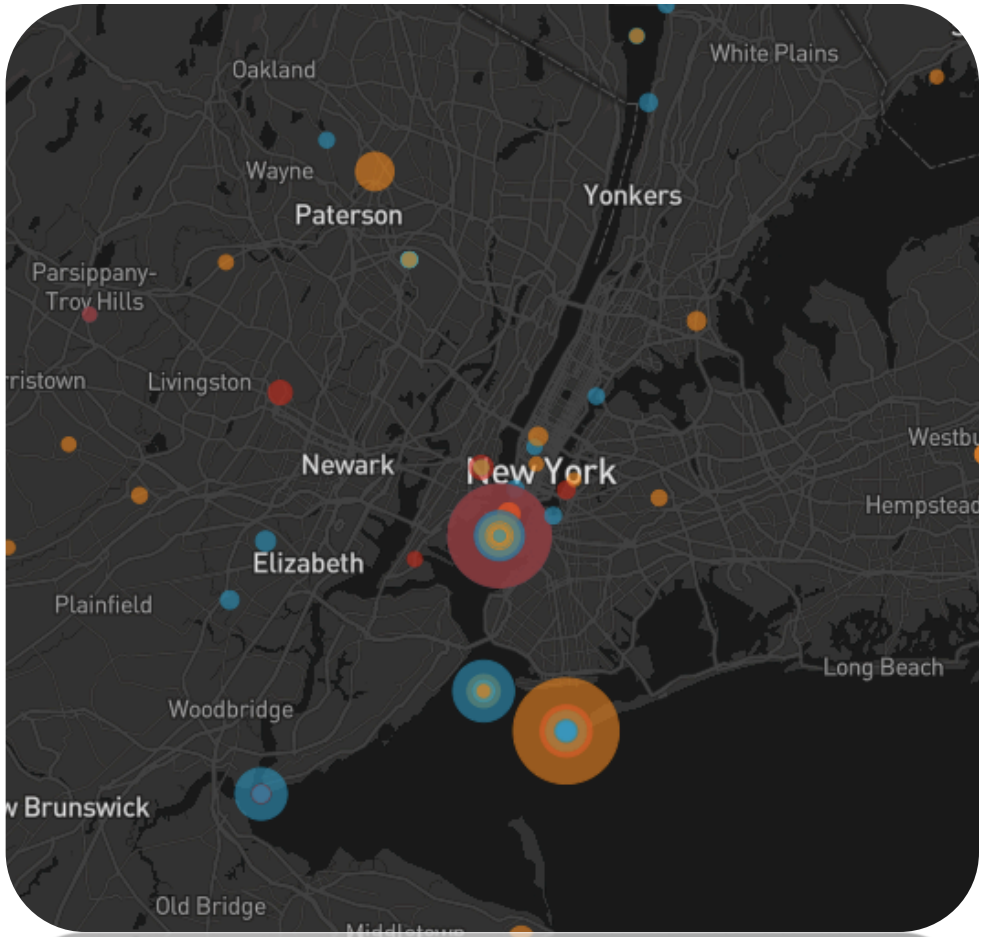
RESOURCES NEEDED



Go to the map folder in your repo:

- index.html
- map.js

Map



- Define graph location [10]
- Define map style [11]
- Define the data location [35]
- Set the dot location [47]
- Define legend location[62]
- Define legend data [84]

MAPBOX STYLES

- **Comic:** mapbox://styles/couellette/cj6qihl4m3jkx2qnsh1fyar3s
- **Basic:** mapbox://styles/couellette/cj1nok0fi00242rnrai70hfxk
- **Dark:** mapbox://styles/mapbox/dark-v9
- **Navigation:** mapbox://styles/mapbox/navigation-guidance-day-v2
- **Satellite:** mapbox://styles/mapbox/satellite-v9
- **Light:** mapbox://styles/mapbox/light-v9

MAPBOX

- Free individual licence
- Very user friendly
- Low initial learning curve

https://www.mapbox.com

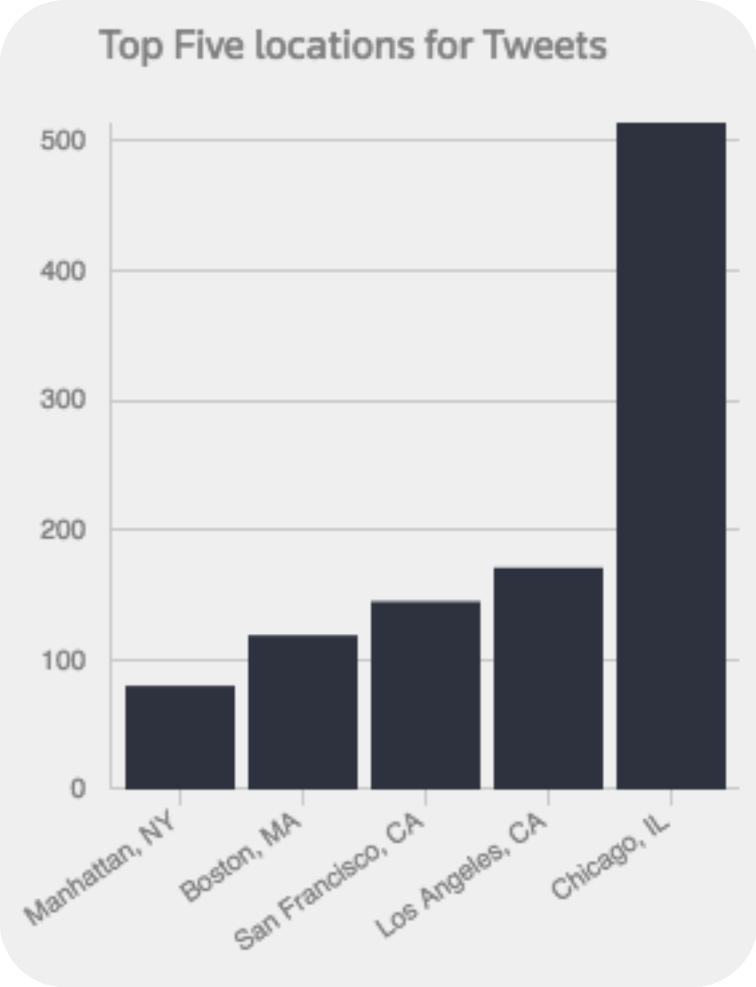
QUANTIFYING NON-NUMERIC VALUES

VISUALIZATION RESOURCES

In the main data folder we will use:

- Bar
  - *index.html and bar.js*
- Pie
  - *index.html and pie.js*
- Line
  - *index.html and line.js*

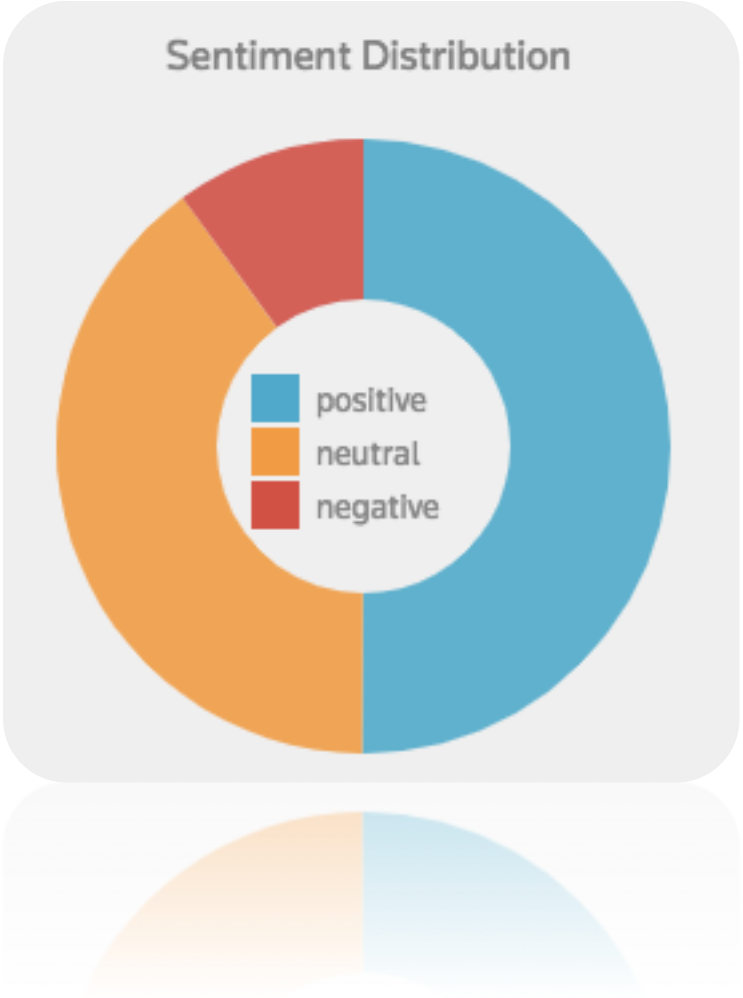
Bar Graph



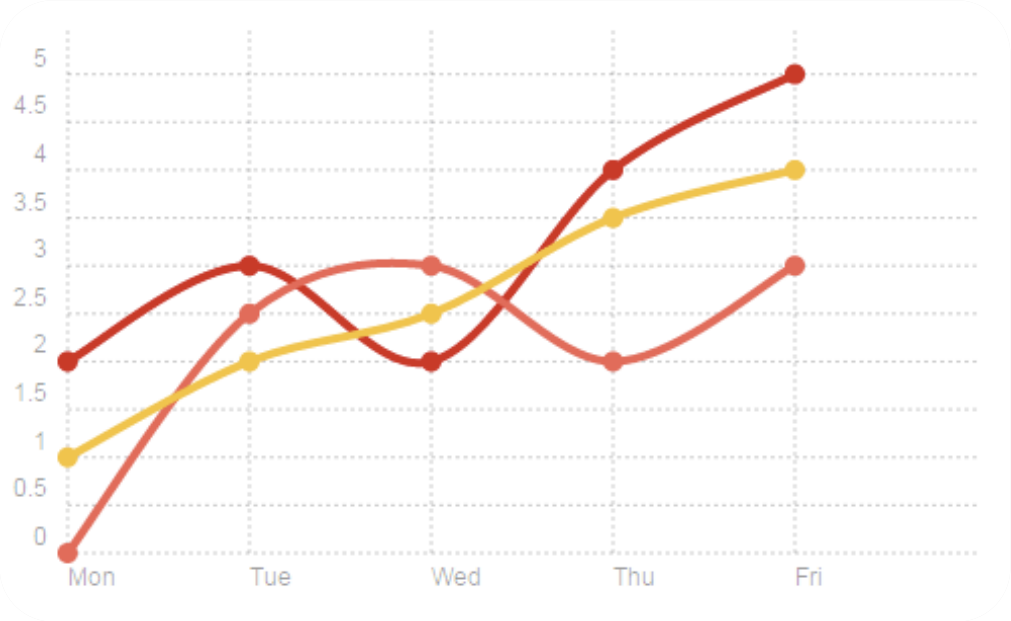
- Define graph location [12]
- Define the data location [43]
- Quantify non-numeric values [56]
- Define x-axis [70]
- Define bar data [98]
- Visualize and review

Pie/Donut Graph

- Define graph location [22]
- Define the data location [28]
- Quantify non-numeric values [31]
- Group values by category [39]
- Set colour by category [50]
- Visualize and review



Line Graph



- Define the graph location [23]
- Define the data location [30]
- Define timeframe [34]
- Quantify/group non-numeric values [38,39]
- Define x-axis [49]
- Visualize and review

BRINGING IT ALL TOGETHER

ALL ON ONE PAGE

In the main data folder we will use:

- Dashboard
  - Open index.html

VISUAL CRITIQUE

Open dialog on *good* and *bad* visualizations

LETS COMPARE

Houston Floods

- Houston's floodwaters are receding, but they remain dangerously high in many areas
- Hell or High Water

Rio Olympics

- The Countries Where Women Won More Medals Than Men in Rio
- Rio 2016 Olympic Medal Count

AVAILABLE TECHNOLOGIES

- Vega D3 - https://vega.github.io/vega/
- D3 Live - http://d3js.live/
- Chart JS - http://www.chartjs.org/samples/latest/
- Tableau - https://public.tableau.com/en-us/s/
- Google Charts - https://google-developers.appspot.com/chart/interactive/docs/gallery

A good read...

Interactive Data Visualizations for the Web – 2nd Edition

http://shop.oreilly.com/product/0636920037316.do

THANK YOU