


Data Storytelling Studio
analyzing data

CMS.631/831
Rahul Bhargava



Agenda

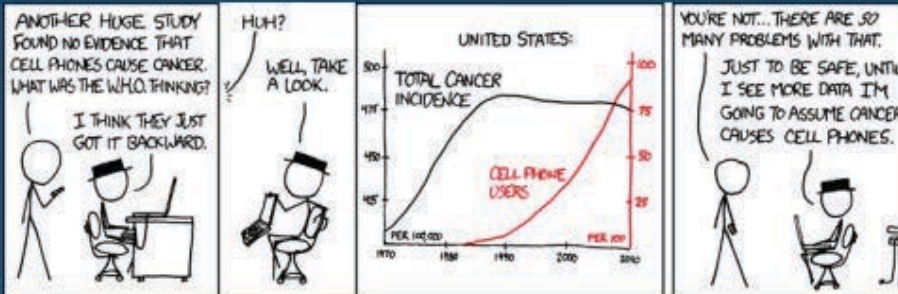
- [10] basic stats
- [10] grad reading presentation
- [20] a taxonomy of story types
- [20] WordCounter storyfinding activity
- [15] tool demonstrations
- [5] MAGIC criteria
- [?] Tableau Playtime?
- [5] homework prep

Basic stats

- mean
- median
- min
- max
- quartiles
- standard deviation
- normalization

Don't be scared by these terms. Most of statistics involves fancy ways of counting, and most folks are good at counting!

Correlation



<https://xkcd.com/925/>

Courtesy of xkcd.com.
License CC BY-NC 2.5

interpretation is *REALLY* hard - let's try it out on <http://bit.ly/538science>

Correlation - how connected two variables are - doesn't imply that one causes the other. Use the linked-to 538 piece to have students in pairs play around with how easy it is to find correlations, or not find them, but picking and choosing your data inputs.

Advanced Topics

(student presentation)

Julia Koschinsky. 2015. Data Science for Good: What Problems Fit? (2015).

Michael Brennan. Can computers be racist? Big data, inequality, and discrimination. Ford Foundation.

Anand Rajaraman and Jeffrey David Ullman. 2011. Data Mining. In Mining of Massive Datasets. New York, N.Y. ; Cambridge: Cambridge University Press.

Have a students present a summary of the assigned readings.

Make sure they cover:

- What algorithmic techniques should we be aware of when analyzing large sets of data? and their limitations?

a taxonomy of types of stories

Lets think about the types of stories you can tell with data.

Story Archetypes

- Changes over Time
- Interesting Factoids
- Surprising Connections
- Personal Experiences
- Revealing Comparisons

This list is useful scaffolding for thinking about the kinds of stories you can find to tell. It isn't a complete list by any means, but is useful to think through as you're digging into data. I'll give a concrete example of each.

Data to Think With

Address	Suffix	Street	Side	Site	X-Coord	Y-Coord	Species	DBH	Condition	Longitude	Latitude	Has Colon?	Base Species	Species Icon	Status Icon
6		ABERDEEN R	Front	1	760669.32	2968011.8	linden: littleleaf		11 Fair	-71.11315	42.391726	TRUE	linden	small_purple	2
9		ABERDEEN R	Front	1	760653.99	2968065.03	linden: littleleaf		12 Fair	-71.113205	42.391872	TRUE	linden	small_purple	2
27		ABERDEEN R	Side	1	760597.78	2968111.87	linden: littleleaf		26 Fair	-71.113413	42.392001	TRUE	linden	small_purple	2
30		ABERDEEN R	Front	1	760576.06	2968176.72	linden: littleleaf		25 Fair	-71.113492	42.392185	TRUE	linden	small_purple	2
31		ABERDEEN R	Front	1	760615.66	2968188.52	linden: littleleaf		13 Good	-71.113345	42.392211	TRUE	linden	small_purple	3
7		ADAMS ST	Front	1	763960.62	2968655.88	linden: littleleaf		32 Fair	-71.100955	42.393479	TRUE	linden	small_purple	2
13		ADAMS ST	Front	1	763972.63	2968693.82	linden: American		15 Poor	-71.10091	42.393556	TRUE	linden	small_purple	1
15		ADAMS ST	Front	1	763997.46	2968747.58	maple: Norway		8 Fair	-71.100817	42.393703	TRUE	maple	small_red	2
16		ADAMS ST	Front	1	763979.04	2968772.4	mountainash: European		7 Poor	-71.100885	42.393771	TRUE	mountainash	meade_grey	1
17 X		ADAMS ST	Front	1	764024.69	2968802.04	maple: Norway		20 Poor	-71.100715	42.393852	TRUE	maple	small_red	1
17 X		ADAMS ST	Front	2	764047.91	2968856.5	maple: silver		21 Poor	-71.100628	42.394003	TRUE	maple	small_red	1
27		ADAMS ST	Front	1	764067.94	2968893.34	maple: Norway		4 Poor	-71.100553	42.394102	TRUE	maple	small_red	1
30		ADAMS ST	Front	1	764068.54	2968952.22	ash: green		9 Fair	-71.100546	42.394291	TRUE	ash	small_green	2
35		ADAMS ST	Front	1	764103.98	2968954.62	serviceberry: downy		3 Poor	-71.100419	42.394297	TRUE	serviceberry	meade_grey	1
39		ADAMS ST	Front	1	764136.01	2969027.89	serviceberry: downy		5 Fair	-71.100299	42.39447	TRUE	serviceberry	meade_grey	2
40		ADAMS ST	Front	1	764107.18	2969036.7	oak: pin		27 Fair	-71.100406	42.394495	TRUE	oak	meade_grey	2
40		ADAMS ST	Front	2	764086.36	2968991.85	ash: green		7 Fair	-71.100483	42.394372	TRUE	ash	small_green	2
43		ADAMS ST	Front	1	764157.64	2969073.54	serviceberry: downy		5 Fair	-71.100218	42.394595	TRUE	serviceberry	meade_grey	2
46		ADAMS ST	Front	1	764156.04	2969136.01	ash: green		8 Fair	-71.100221	42.394766	TRUE	ash	small_green	2
52		ADAMS ST	Front	1	764173.66	2969168.09	oak: pin		19 Fair	-71.100157	42.394854	TRUE	oak	meade_grey	2
53		ADAMS ST	Front	1	764204.09	2969169.65	serviceberry: downy		7 Fair	-71.100044	42.394858	TRUE	serviceberry	meade_grey	2
54		ADAMS ST	Front	1	764212.1	2969246.54	ash: green		9 Fair	-71.100014	42.395069	TRUE	ash	small_green	2
59		ADAMS ST	Front	1	764249.74	2969264.96	pear: callery		11 Fair	-71.099874	42.395115	TRUE	pear	small_yellow	2
60		ADAMS ST	Front	1	764240.93	2969305	oak: pin		22 Fair	-71.099906	42.395229	TRUE	oak	meade_grey	2
66		ADAMS ST	Front	1	764263.36	2969348.25	oak: pin		21 Fair	-71.099822	42.395347	TRUE	oak	meade_grey	2
73		ADAMS ST	Front	1	764309.01	2969381.89	oak: pin		19 Fair	-71.099652	42.395439	TRUE	oak	meade_grey	2
73		ADAMS ST	Side	1	764375.92	2969406.41	sweetgum: American		4 Fair	-71.099404	42.395511	TRUE	sweetgum	meade_grey	2
4		ADRIAN ST	Front	1	764750.63	2962854.33	pear: callery		13 Fair	-71.098131	42.377522	TRUE	pear	small_yellow	2
5		ADRIAN ST	Front	1	764748.59	2962886.98	maple: Norway		13 Fair	-71.098138	42.377611	TRUE	maple	small_red	2
16 X		ADRIAN ST	Front	1	764664.93	2962802.3	pear: callery		13 Fair	-71.098449	42.37738	TRUE	pear	small_yellow	2
16 X		ADRIAN ST	Front	2	764706.25	2962826.79	pear: callery		12 Fair	-71.098296	42.377447	TRUE	pear	small_yellow	2
18		ADRIAN ST	Front	1	764617.69	2962776.79	pear: callery		15 Fair	-71.098625	42.377311	TRUE	pear	small_yellow	2
21		ADRIAN ST	Front	1	764571.07	2962795.67	planetree: London		1 Good	-71.098797	42.377363	TRUE	planetree	meade_whit	3

In 2009 the city of Somerville, MA, USA did an audit of all of its trees. This data includes the exact location, species, health condition and more about each of the ~50,000 trees in the city. We'll use that data to think about the types of stories you can tell.

Changes Over Time



Humans experience the world through time, so before/after stories are naturally relatable.

This dot-map shows one area of town. Each dot is one tree, colored based on its health (red=poor, yellow=ok, green=good). Like most dot maps this suffers from the occlusion problem (where some dots sit on top of others, blocking them from view). However, we can see one street that stands out as having lots of healthy trees and not many unhealthy ones. Why?

With my knowledge of the neighborhood, I know that this is Somerville Ave., which just before this data collection effort was repaved and redone to beautify it. This story demonstrates a change-over-time narrative. In this case it could be used to discuss the role of city infrastructure planning in creating a healthier treescape.

Interesting Factoids

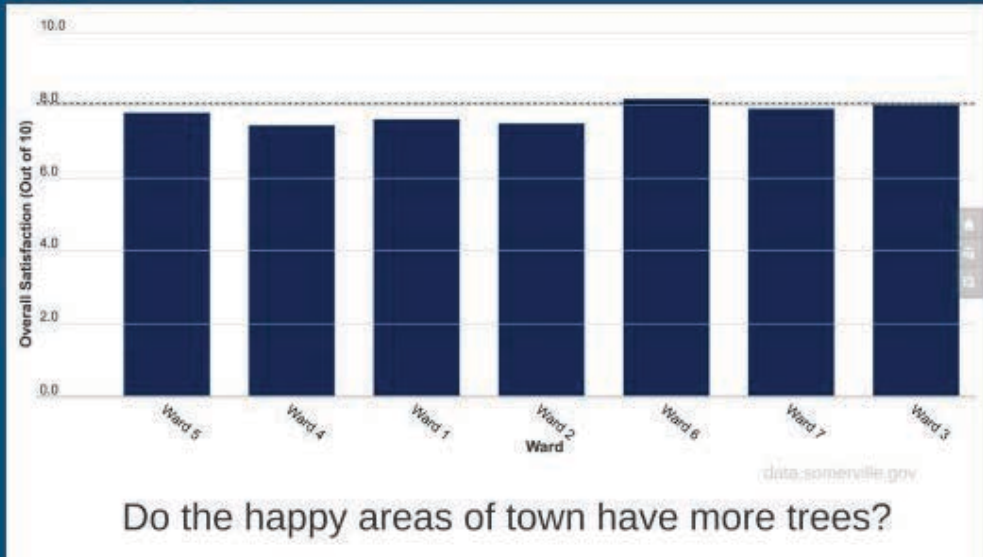


Sometimes the odd (aka outlier) datapoint is the the one to dig into for a story.

We often try to start examining data by looking at it in aggregate. Don't forget that sometimes the solitary story of one data point can be the most compelling. This is a picture of a Hazelnut tree. There is only one in the whole city of Somerville. How did it get there?

This kind of story is an interesting "Factoid" - a single data point that can be the spark of a longer story.

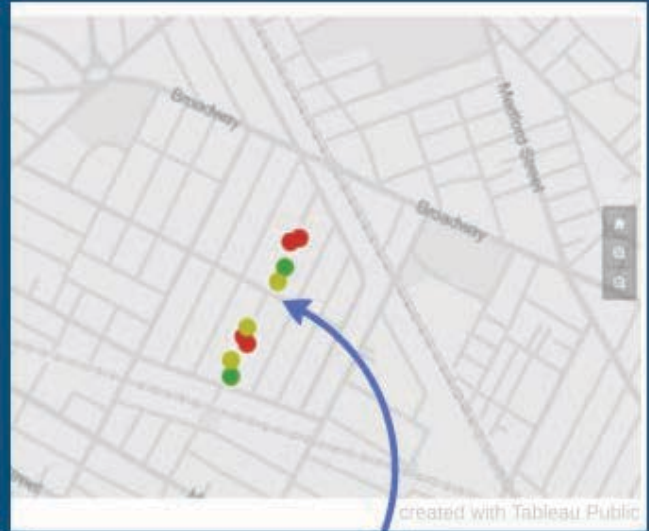
Surprising Connections



People often look for connections between different parts of the data - in statistics there are called correlations. We talked about these before, where I cautioned that a correlation doesn't imply that one thing causes another.

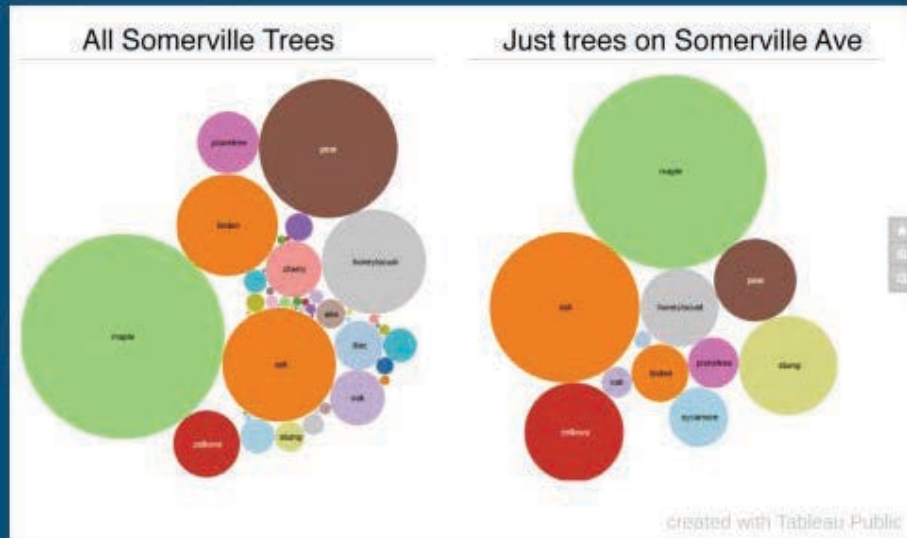
Here is a chart of the overall satisfaction of residents in each part of town. One could wonder if there is a connection between the tree density in one part of town and the overall happiness.

Personal Experiences



Another type of story to tell involves personal experiences. Here I've highlighted one street in town. All the dots shown are spruce trees - this one street has all the spruce trees in town! You could connect to people on that street by talking about how unique their street is, and maybe even spark a campaign to rename their street to "Spruce St"!

Revealing Comparisons



The last type of story in my list is the revealing comparison. The idea here is to compare one slice of the data to another slice.

Here I'm showing two bubble charts. The left shows the population of each species of tree in town. You can see Maple and Pear are the most common. On the right I'm showing the population of trees just along Somerville Ave (one of the main streets). They look similar, and if you do the counting you'll find Somerville Ave to be the most similar to the overall population of trees in town. So if someone wanted to visit and see the tree leaves changing color in fall, you could point them towards Somerville Ave as a reasonable representation of the whole town.

Let's have some fun!

databasic.io/wordcounter

- team up
 - find a story in some lyrics
 - grab some paper and crayons
 - sketch a visual of your story
-

Run the [WordCounter Sketch-a-Story activity \(PDF\)](#).

Analysis

- Exploring quantitative data with [Tableau](#) (demo)
 - More stats-oriented people could try [JMP](#) or [Minitab](#) (but they don't give much for free)
- Exploring large qualitative data with [overview](#)
- [Data Science Toolkit](#) virtual machine for API-level access
- OpenRefine for [finding correlations](#)
- Sentiment analysis [is popular](#), but has limitations
- Qualitative text analysis? ([nineteen](#) is promising)

There are lots of tools to help you here.

Abelson's MAGIC

- **Magnitude** - size of the claim
- **Articulation** - how precise is your claim
- **Generality** - is it valid in multiple contexts
- **Interestingness** - can this change beliefs in a way that matters?
- **Credibility** - do you believe it?

Reivew Abelson's MAGIC criteria from the reading. Remember these are things to keep in mind while making your arguments.

homework

- read stuff
- grad student to present reading on visual narratives

MIT OpenCourseWare
<https://ocw.mit.edu/>

CMS.631 Data Storytelling Studio: Climate Change
Spring 2017

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.