




Data Storytelling Studio

analyzing data

CMS.631/831
Rahul Bhargava



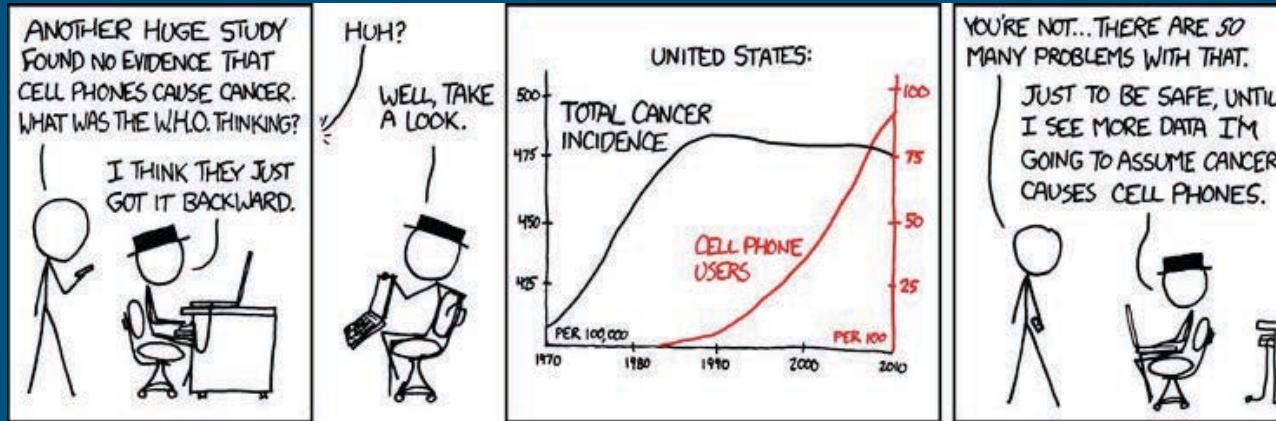
Agenda

- [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

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>

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.

a taxonomy of types of stories

Story Archetypes

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

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	2968178.72	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	2968665.88	linden: littleleaf	32	Fair	-71.100955	42.393479	TRUE	linden	small_purple	2
13		ADAMS ST	Front		1	763972.63	2968693.92	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: Europea	7	Poor	-71.100885	42.393771	TRUE	mountainash	measle_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	42.394001	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	764069.54	2968962.22	ash: green	9	Fair	-71.100546	42.394291	TRUE	ash	small_greener	2
35		ADAMS ST	Front		1	764103.98	2968964.62	serviceberry: downy	3	Poor	-71.100419	42.394297	TRUE	serviceberry	measle_grey	1
39		ADAMS ST	Front		1	764136.01	2969027.89	serviceberry: downy	5	Fair	-71.100299	42.39447	TRUE	serviceberry	measle_grey	2
40		ADAMS ST	Front		1	764107.18	2969036.7	oak: pin	27	Fair	-71.100406	42.394495	TRUE	oak	measle_grey	2
40		ADAMS ST	Front		2	764086.36	2968991.85	ash: green	7	Fair	-71.100483	42.394372	TRUE	ash	small_greener	2
43		ADAMS ST	Front		1	764157.64	2969073.54	serviceberry: downy	5	Fair	-71.100218	42.394595	TRUE	serviceberry	measle_grey	2
46		ADAMS ST	Front		1	764156.04	2969136.01	ash: green	8	Fair	-71.100223	42.394766	TRUE	ash	small_greener	2
52		ADAMS ST	Front		1	764173.66	2969168.05	oak: pin	19	Fair	-71.100157	42.394854	TRUE	oak	measle_grey	2
53		ADAMS ST	Front		1	764204.09	2969169.65	serviceberry: downy	7	Fair	-71.100044	42.394858	TRUE	serviceberry	measle_grey	2
54		ADAMS ST	Front		1	764212.1	2969246.54	ash: green	9	Fair	-71.100014	42.395069	TRUE	ash	small_greener	2
59		ADAMS ST	Front		1	764249.74	2969264.96	pear: callery	11	Fair	-71.099874	42.395119	TRUE	pear	small_yellow	2
60		ADAMS ST	Front		1	764240.93	2969305	oak: pin	22	Fair	-71.099906	42.395229	TRUE	oak	measle_grey	2
66		ADAMS ST	Front		1	764263.36	2969348.25	oak: pin	21	Fair	-71.099822	42.395347	TRUE	oak	measle_grey	2
73		ADAMS ST	Front		1	764309.01	2969381.89	oak: pin	19	Fair	-71.099652	42.395439	TRUE	oak	measle_grey	2
73		ADAMS ST	Side		1	764375.92	2969408.41	sweetgum: American	4	Fair	-71.099404	42.395511	TRUE	sweetgum	measle_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	764617.49	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	measle_whit	3

Changes Over Time



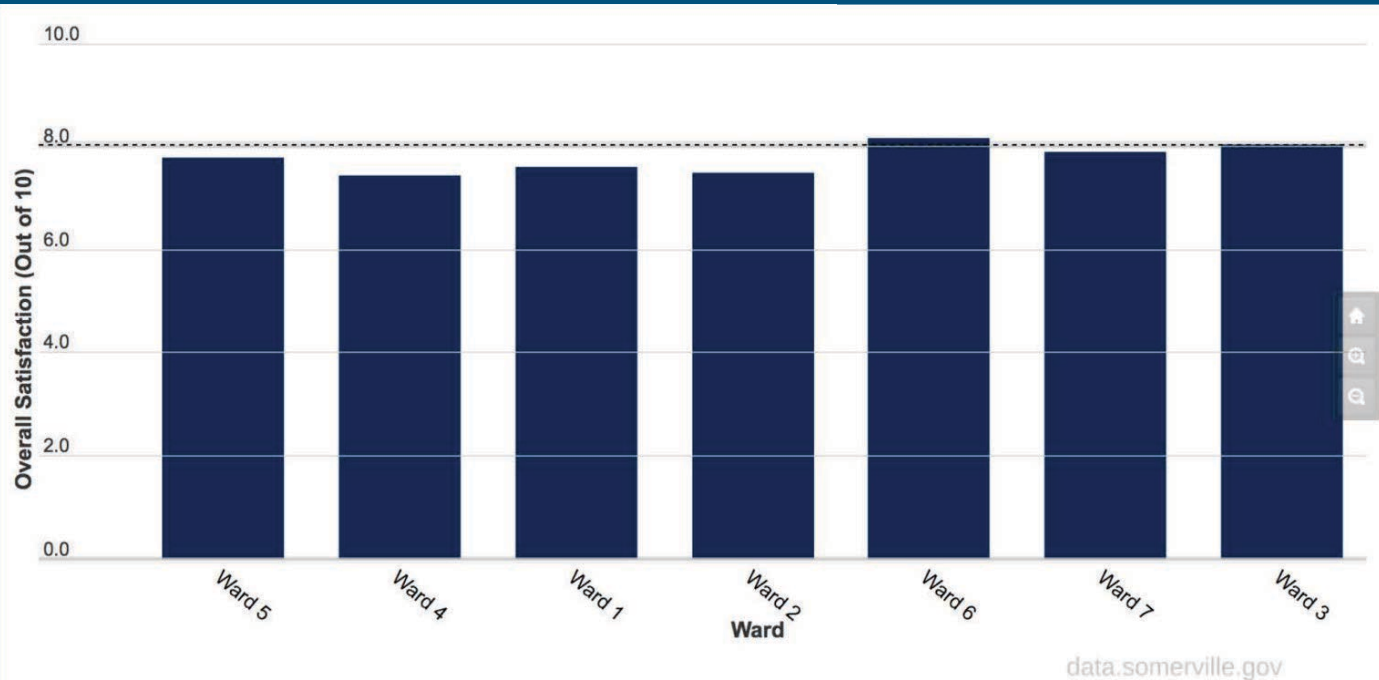
created with Tableau Public

Why are there so many healthy trees on that street?

Interesting Factoids

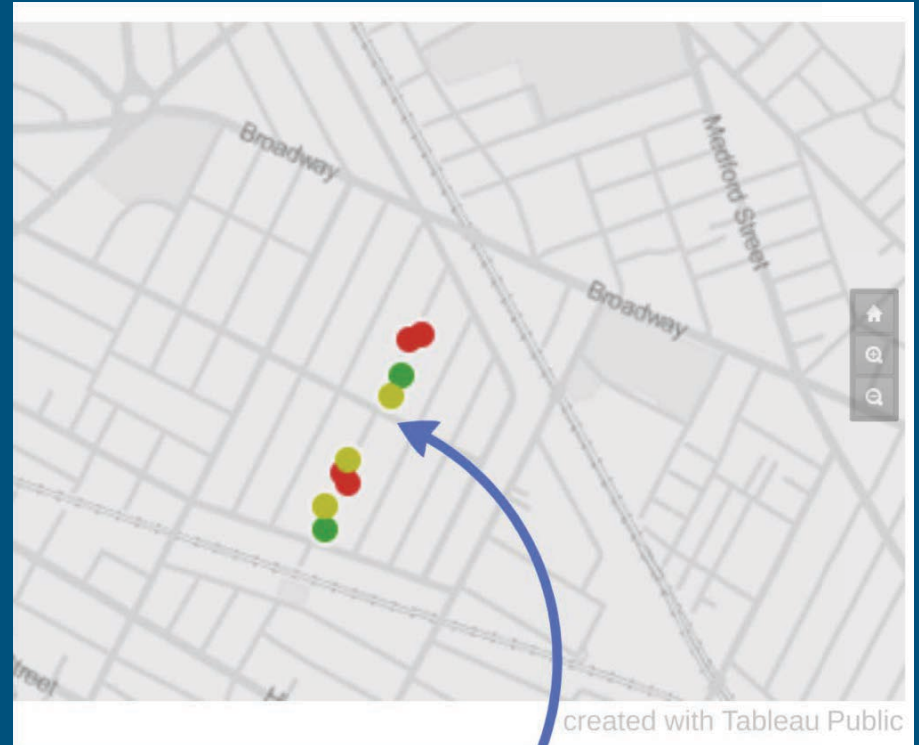


Surprising Connections

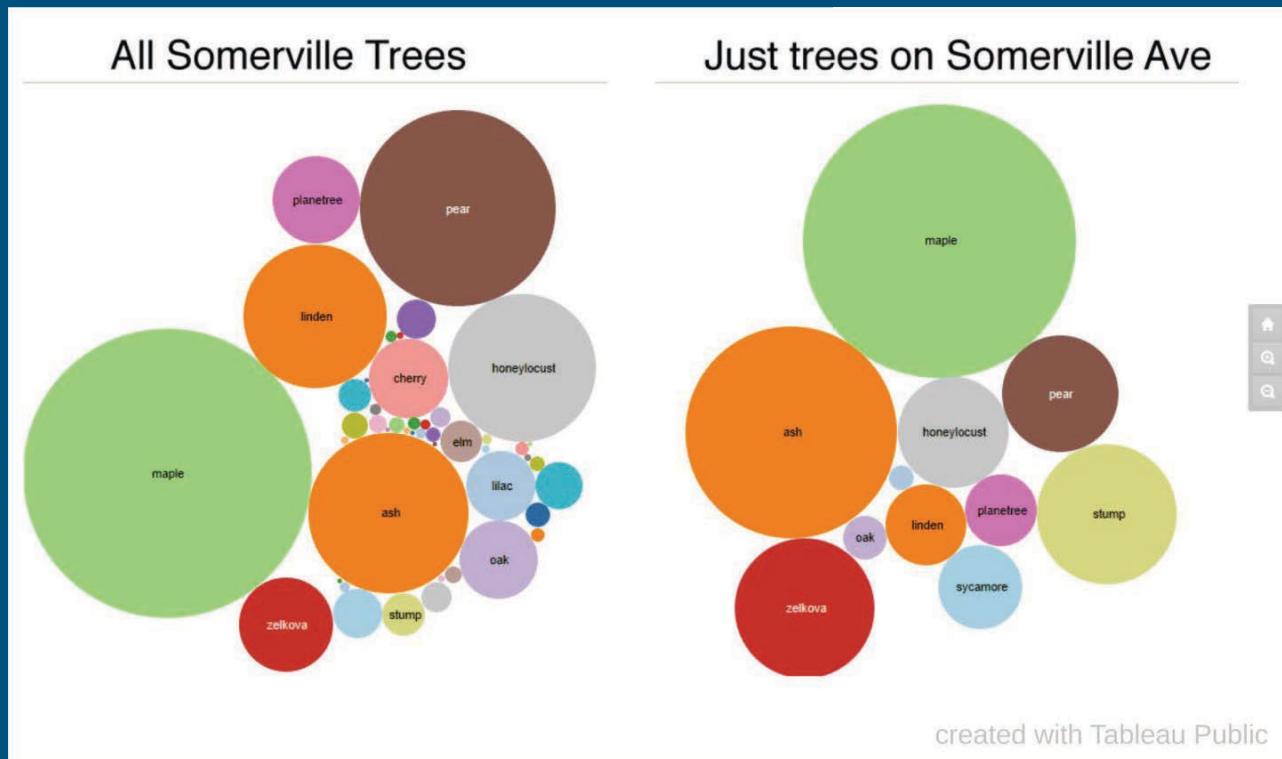


Do the happy areas of town have more trees?

Personal Experiences



Revealing Comparisons



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
-

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)

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?

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>.