



Tufts Independent Data Journal

Some stuff about Tufts you didn't know.

Tufts Trends

The phrase “I just feel like” is used more than 635 times in ‘Tufts Confessions’.

The most reported post-grad job title is “research assistant”. (63*)

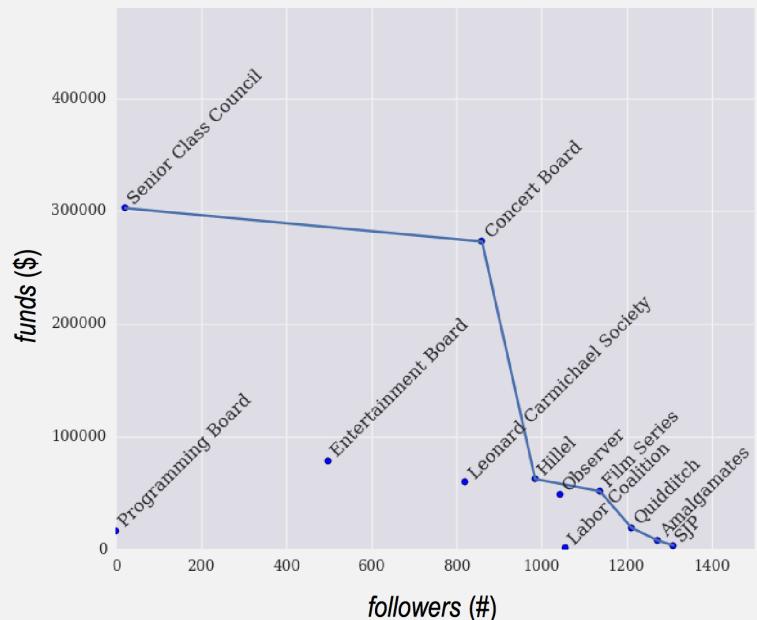
The second most reported post-grad institution is Teach for America . (41*)

Tufts clubs expensed \$159,743 in just food money last year.

Tufts Trends

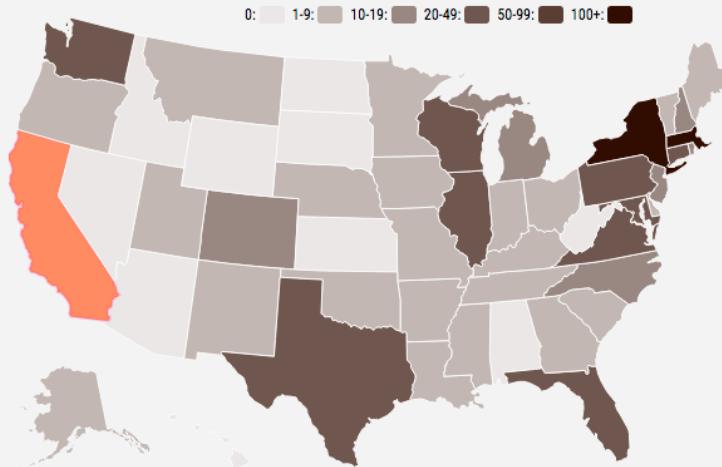
Pareto Frontier of Tufts Clubs

Optimally resourceful university clubs



2013-14 Jumbo Grads in the United States

Domestic first outcomes of Tufts graduates (hover for statistics)



How did we find out the numbers?

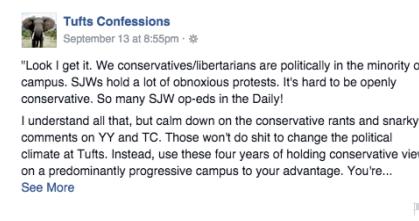
How did we find out the numbers?

Data mining.

Data mining : scrape



International Relations			
Abl Associates	Research Assistant	Cambridge	MA
ACLU	Paralegal	Washington	DC
Acqua	Business Development Representative	Boston	MA
Agahozo Shalom Youth Village	Program Development Intern	New York	NY
AlphaSights	Analyst	New York	NY
American Express	Portfolio Execution Analyst	New York	NY
American Immigration Council	Legal Assistant/Paralegal	Washington	DC
AmeriCorps NCCC - FEMA Corps	Service Member	Vicksburg	MS
Analysis Group	Not specified	Not specified	MA
Analysis Group	Analyst	Not specified	MA
Ann Guyot Emergency Care Consultants	Medical Scribe	Not specified	MN
Asia Initiatives	Research Associate	New York	NY
Atlantic Council	Intern at the Rafik Hariri Center	Washington	DC
Barbara L'Italian for State Senate Campaign	Field Manager	Andover	MA
Best. Corp Hospitality Training Center	NAIP AmeriCorps Service Member	Boston	MA
Bloomberg LP	Financial Product Analyst	New York	NY
BrazilExpt	Brand Ambassador	Sao Paulo	Brazil
Brick Wall Media	Junior Editor	Not specified	NY
Careus Associates	Consultant	Not specified	WA
Cambridge Associates	Consulting Associate	Not specified	MA
Cambridge Mercantile Group	FX Sales Support/Business Development	New York	NY
Capital New York	Business Development Associate	Brooklyn	NY
Catholic Relief Services	Summer Intern	Baltimore	MD
Citigroup	Analyst	Not specified	NY
Clark Lau LLC	Immigration Paralegal	Not specified	MA
Community Chlorinators - The Archimedes Project	Co-Founder	Port-au-Prince	Haiti
Confederation of Indian Industry	Executive	Washington	DC



Group Name	Budget
African Student Organization	Download Budget
Asian American Alliance	Download Budget
Association of Latin American Students	Download Budget
Black Student Union	Download Budget
Caribbean Student Organization	Download Budget
Chinese Students Association	Download Budget
German Club	Download Budget
Hong Kong Students Association	Download Budget
International Club	Download Budget
Italian Club	Download Budget
Japanese Culture Club	Download Budget
Korean Student Association	Download Budget

Data mining : munge

International Relations

Abt Associates	Research Assistant	Cambridge	MA
ACLU	Paralegal	Washington	DC
Acqua	Business Development Representative	Boston	MA
Agabuzz Shalom Youth Village	Program Development Intern	New York	NY
AlphaSights	Analyst	New York	NY
American Express	Portfolio Execution Analyst	New York	NY
American Immigration Council	Legal Assistant/Paralegal	Washington	DC
AmeriCorps NCCC - FEMA Corps	Service Member	Vicksburg	MS
Analysis Group	Not specified	Not specified	MA
Analysis Group	Analyst	Not specified	MA
Ann Guyott' Emergency Care Consultants	Medical Scribe	Not specified	MA
Aaa Initiatives	Research Associate	New York	NY
Atlantic Council	Intern at the Rafik Hariri Center	Washington	DC
Barbara L. Italien for State Senate Campaign	Field Manager	Andover	MA
Best. Corp Hospitality Training Center	NAPT AmeriCorps Service Member	Boston	MA
Bloomberg LP	Financial Product Analyst	New York	NY
BrazilExpo	Brand Ambassador	Sao Paulo	Brazil
Brick Media	Junior Editor	Not specified	NY
Careers Associates	Consultant	Not specified	WA
Cambridge Associates	Consulting Associate	Not specified	MA
Cambridge Mercantile Group	FX Sales Support/Business Development	New York	NY
Capital New York	Business Development Associate	Brooklyn	NY
Catholic Relief Services	Summer Intern	Baltimore	MD
Citigroup	Analyst	Not specified	NY
Clark Lau LLC	Immigration Paralegal	Not specified	MA
Community Chlorinators - The Archimedes Project	Co-Founder	Port-au-Prince	Haiti
Confederation of Indian Industry	Executive	Washington	DC



time / geographic data



Tufts Confessions

September 13 at 7:47pm · *

"To the anti-feminists, you people need to chill. I'm guessing most of you are guys. Don't you realize that there is nothing that has done so much to liberate men from traditional gender roles as Feminism? Thanks to feminism's doctrine that women are equal to men in every way, you no longer need to pay for dates, support your fiancee/girlfriend, pay for her rent/car/groceries, etc. etc.

Think about it- when you ask a girl to "go dutch" aka split the bill for a dinner or som... See More

Tufts Confessions

September 13 at 8:55pm · *

"Look I get it. We conservatives/libertarians are politically in the minority on campus. SJWs hold a lot of obnoxious protests. It's hard to be openly conservative. So many SJW op-eds in the Daily!

I understand all that, but calm down on the conservative rants and snarky comments on YY and TC. Those won't do shit to change the political climate at Tufts. Instead, use these four years of holding conservative views on a predominantly progressive campus to your advantage. You're... See More



textual data



Group Name

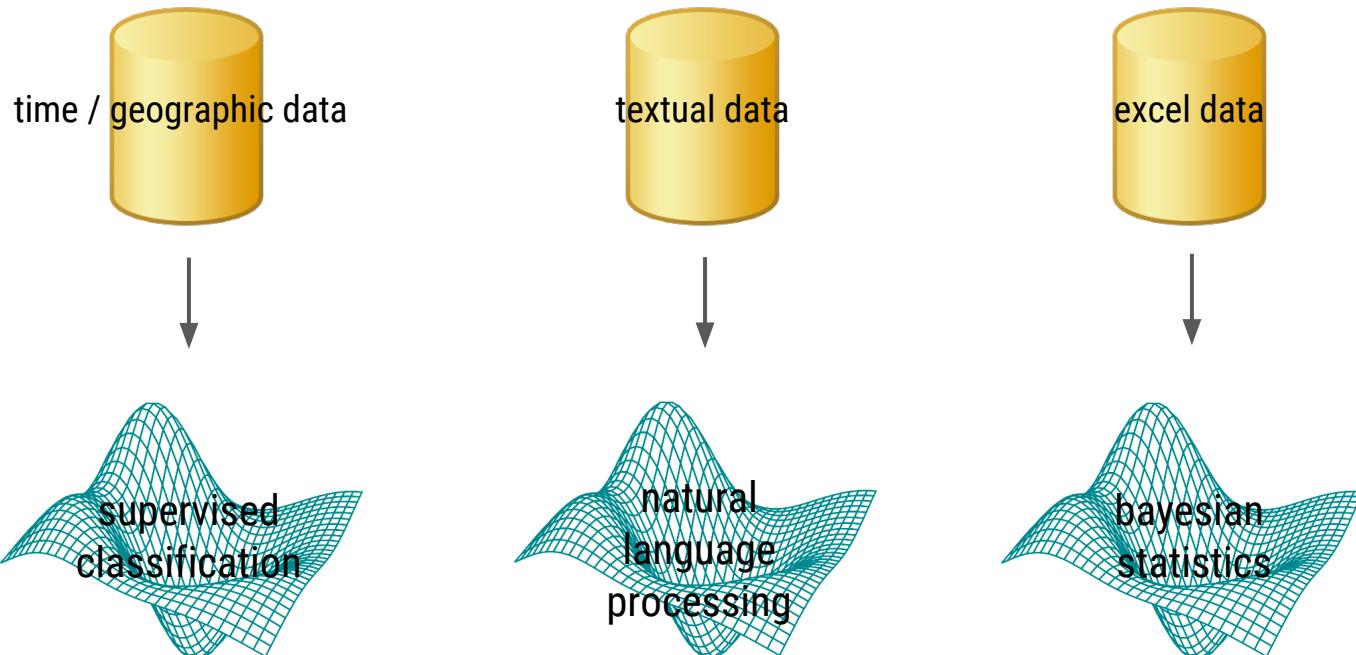
Group Name	Budget
African Student Organization	Download Budget
Asian American Alliance	Download Budget
Association of Latin American Students	Download Budget
Black Student Union	Download Budget
Caribbean Student Organization	Download Budget
Chinese Students Association	Download Budget
German Club	Download Budget
Hong Kong Students Association	Download Budget
International Club	Download Budget
Italian Club	Download Budget
Japanese Culture Club	Download Budget
Korean Student Association	Download Budget



excel data



Data mining : model



Data mining : visualize

supervised classification

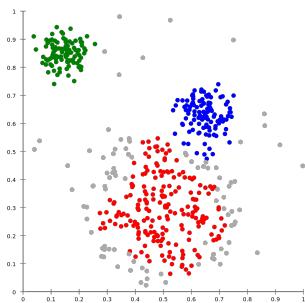
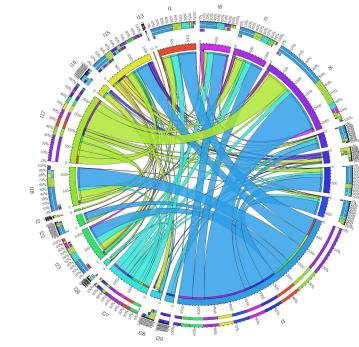


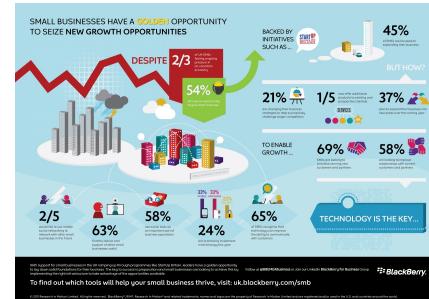
chart / graph

natural language processing



interactive viz

bayesian statistics

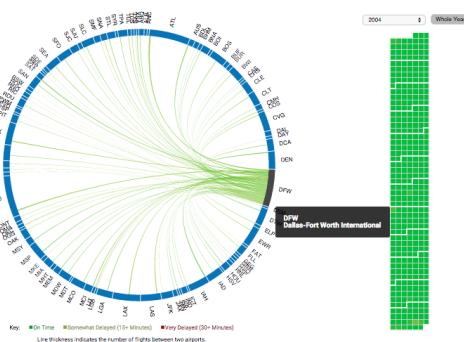
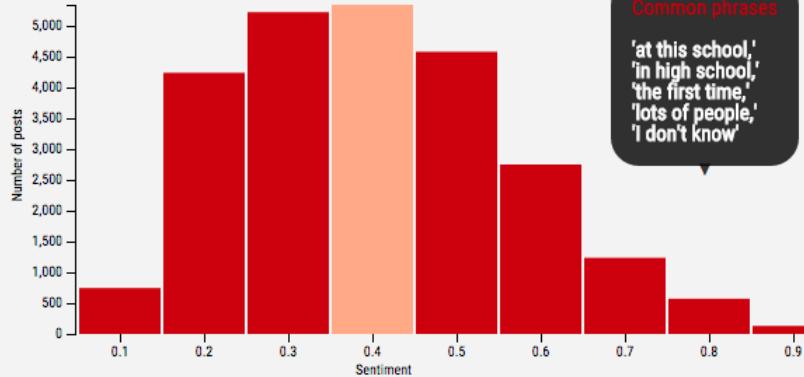


infographic

Data mining : examples

The Sentiment Spectrum

Confessions classified by sentiment [negative to positive]

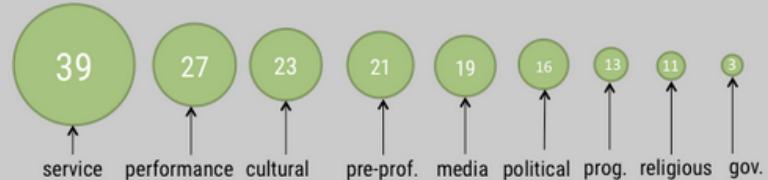


Tufts University Clubs By the Numbers

Data from officially recognized 2015 Tufts student organizations

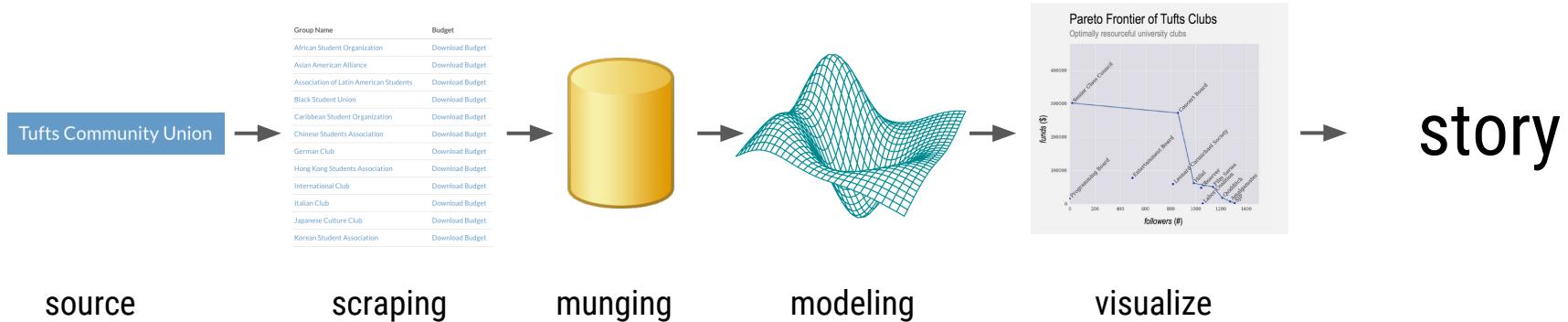


Councils Ranked by Number of Member Clubs



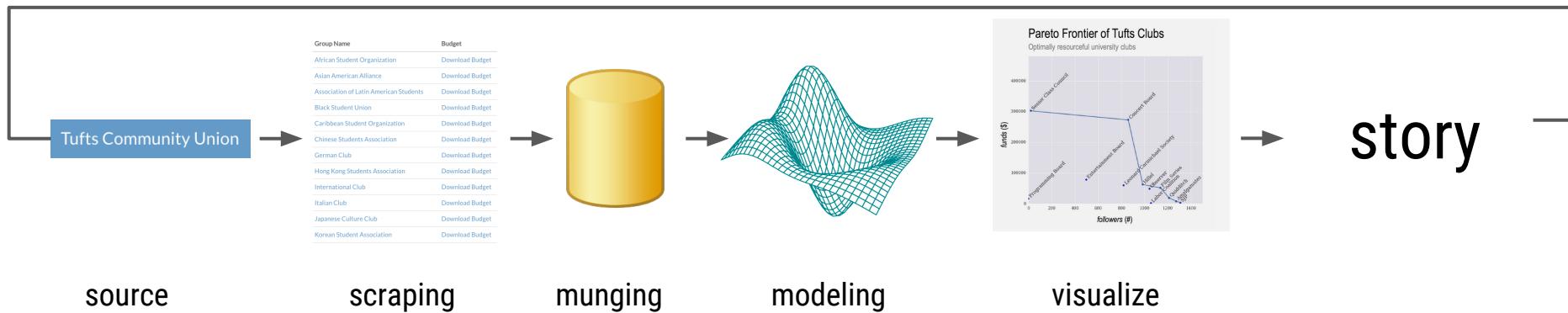
Data mining → journalism

“increased role that numerical data is used in the production and distribution of information in the digital era.”



Data mining → journalism

“increased role that numerical data is used in the production and distribution of information in the digital era.”

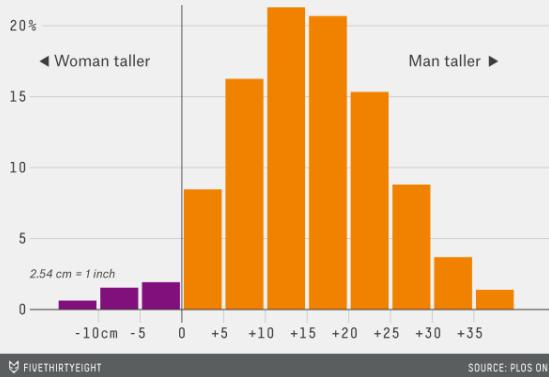


Who uses data journalism?

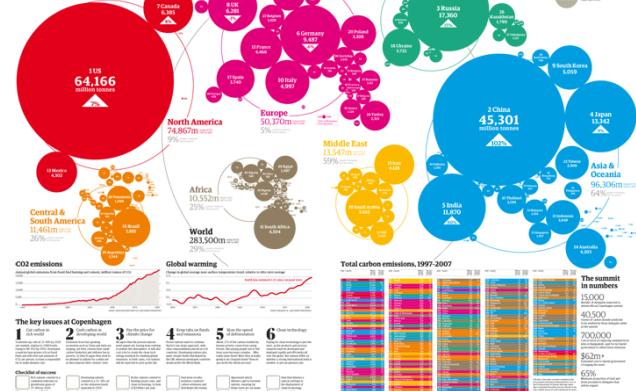
 FiveThirtyEight

Height Of Romance

Gap (centimeters) between the man and woman in a relationship; distribution among 12,502 U.K. couples with babies born in 2000



Global emissions since Kyoto



the guardian

The New York Times

Four Ways to Slice Obama's 2013 Budget Proposal

Explore every nook and cranny of President Obama's federal budget proposal.

All Spending Types of Spending Changes Department Totals

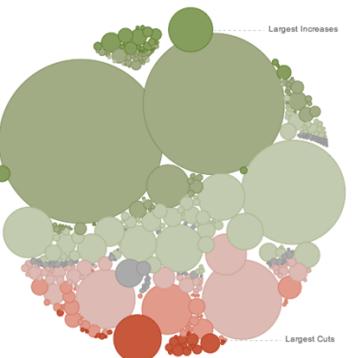
How \$3.7 Trillion Is Spent

Mr. Obama's budget proposal includes \$3.7 trillion in spending in 2013, and forecasts a \$901 billion deficit.

Circles are sized according to the proposed spending.

\$100 billion
\$10 billion
\$1 billion

Largest Increases
Largest Cuts

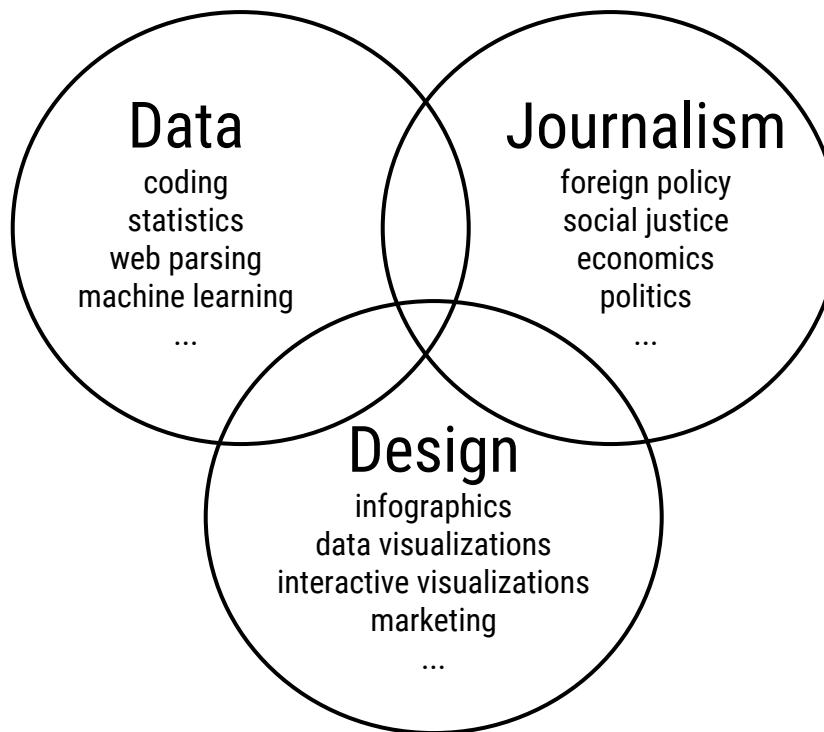


Who are we looking for?

Who are we looking for?

Everyone.

Who are we looking for?



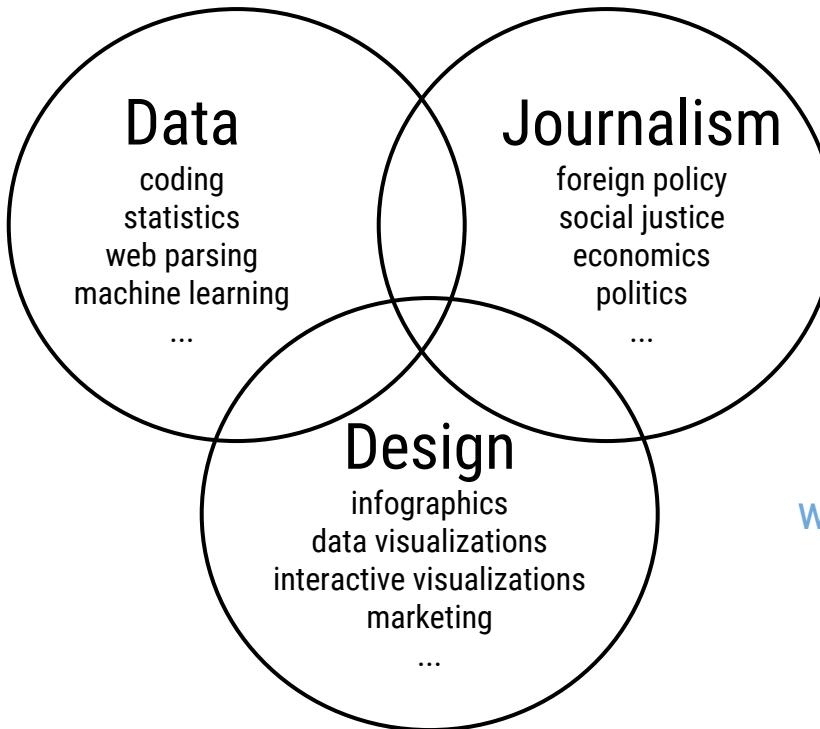
Who are we looking for?

passionate about an issue or domain.

work well in a team.

interdisciplinary + eagerness to learn.

want to use data to make a difference.



Enigma : Tufts Independent Data Journal

1. choose a topic / project.
2. form a team / go solo.
3. learn from our weekly workshops.
4. work on a long-form data journalism article.
5. get published.

Some sample themes

Quantifying the Syrian Refugee Crisis

An Analysis of the 2016 Presidential Candidates

Homelessness in the Boston / Somerville Areas

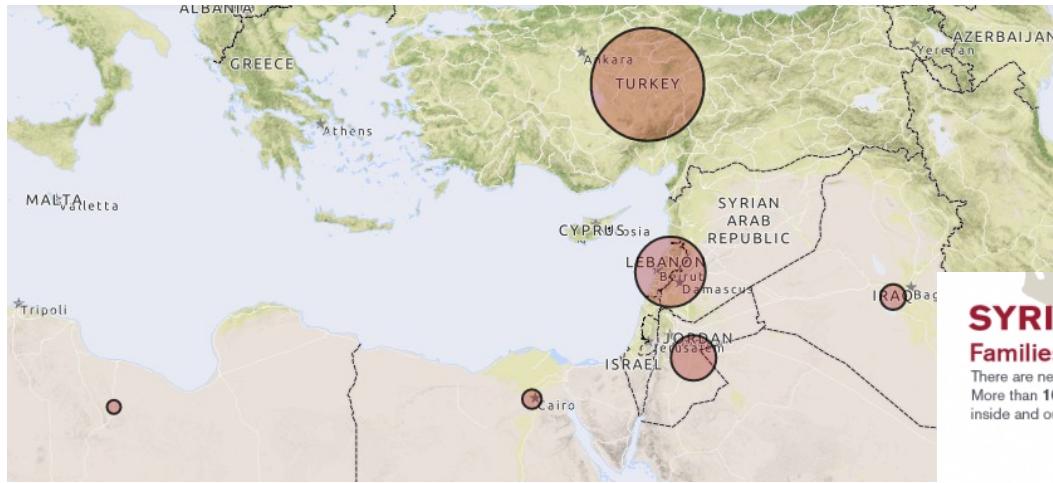
A Demographic Analysis of LGBTQ+ Groups in the U.S.

University Workforce : Statistical Trends on Inequity of University Workers

Using Predictive Analytics for Data-Driven Criminal Justice

The Natural Language of Racial Bias in News Media (e.g. "terrorism")

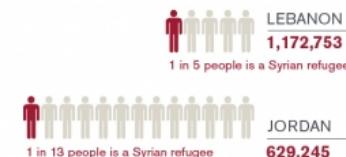
Quantifying the Syrian Refugee Crisis



SYRIAN REFUGEE CRISIS

Families Fleeing Violence

There are nearly 4 million Syrian refugees in five host countries. More than 16 million people are in need of assistance inside and outside Syria.



As of September 4, 2015
Source: UNHCR <http://data.unhcr.org/syrianrefugees/regional.php>; <http://www.unhcr.org/cgi-bin/texis/vtx/search?page=search&docid=550d648a9&query=syrian%20refugees>

Quantifying the Syrian Refugee Crisis

- Human rights/government organizations have an abundance of public data
 - UN High Commission for Refugees (UNHCR)
 - Amnesty International
 - Migration Policy Centre
 - Worldvision
- Many publications and organizations have analyzed and communicated trends in this data but we have an opportunity to delve deeper and use hard data science to find out more
- Our connections with the EECS and other departments could allow collaboration and sharing of resources/skills
 - Dr. Karen Panetta sponsors projects related to this area

An Analysis of the 2016 Presidential Candidates

- 2016 political race is proving to be very controversial; fluctuation in polls
- Donald Trump, Carly Fiorina, Bernie Sanders, Hillary Clinton, Dr. Ben Carson, Jeb Bush, etc.
- But who is promising what and will they actually follow through?
- Let's analyze some of their statements, shall we?
- Here is an example: Jeb Bush proposes to raise the growth rate (of GDP or output of the country) to 4%. Currently, growth sits at 2%. Let's see if what he is proposing is actually possible...

*Familiarity and Favorable Ratings of Potential 2016 Presidential Candidates,
Based on National Adults*

Ranked by % familiar

	% Familiar (have an opinion)	% With favorable opinion	% With unfavorable opinion	Net favorable (pct. pts.)
Hillary Clinton	89	50	39	+11
Joe Biden	78	39	39	0
Jeb Bush	68	35	33	+2
Chris Christie	65	31	34	-3
Mike Huckabee	57	30	27	+3
Rick Perry	57	25	32	-7
Rand Paul	53	28	25	+3
Ted Cruz	50	22	28	-6
Marco Rubio	47	26	21	+5
Rick Santorum	47	20	27	-7
Elizabeth Warren	41	22	19	+3
Scott Walker	38	20	18	+2
Bobby Jindal	35	18	17	+1
Ben Carson	28	20	8	+12
Bernie Sanders	24	12	12	0
Jim Webb	24	11	13	-2

March 2-4, 2015

GALLUP®

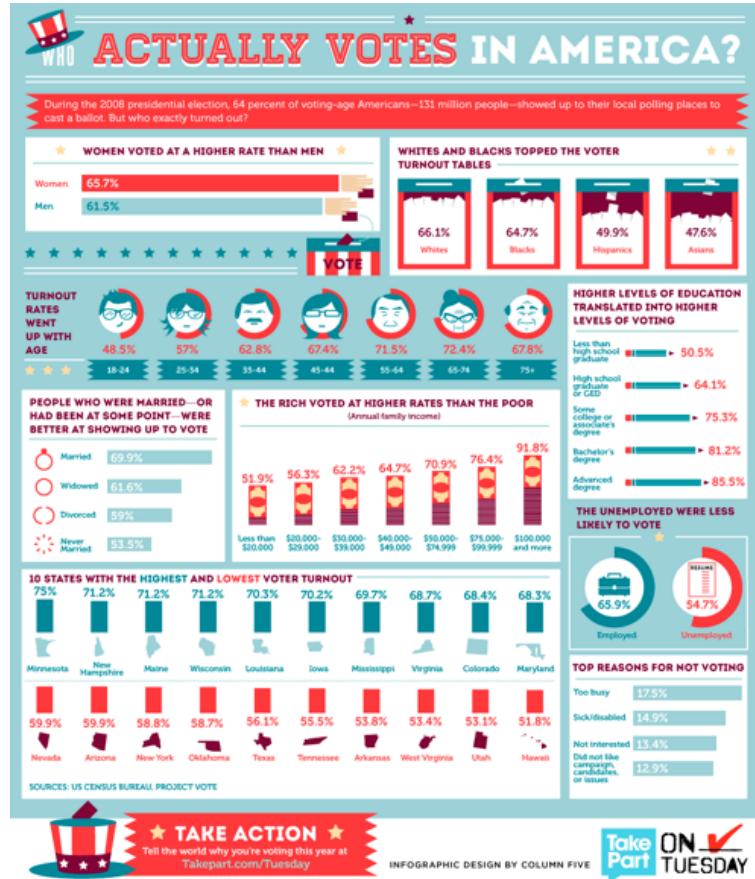
Analyzing Campaign Statements about Reform:

- Historically, this has never been done (and bingo! this is where you could add a **data visualization** to analyze this point further! How many candidates have followed through in the past in regards to this?)
- In Contrast: Obama promised to send people to the moon by 2020 (oh, and also Mars!)
 - It was too expensive, lacked developments in innovation, over the budget and behind schedule (Make an infographic about the history of space innovation and costs)
 - You could make an infographic about past candidates promises' about space exploration over the course of American history → how many have actually achieved this goal?)

Political Race 2016

- Also think about the **stats!** Infographic/Data Visualization Examples:
 - What percentage of women currently say they would vote for Carly Fiorina over Donald Trump?
 - What percentage of candidates are promising education and health reform?
 - What are the most important topics in this race? How is this subdivided? (what percentage of the middle class, for example, believes that tax is the most important issue?)

****And who will be likely to vote in this election?
 (Here is a GREAT example of a data visualization by the Huffington Post) →



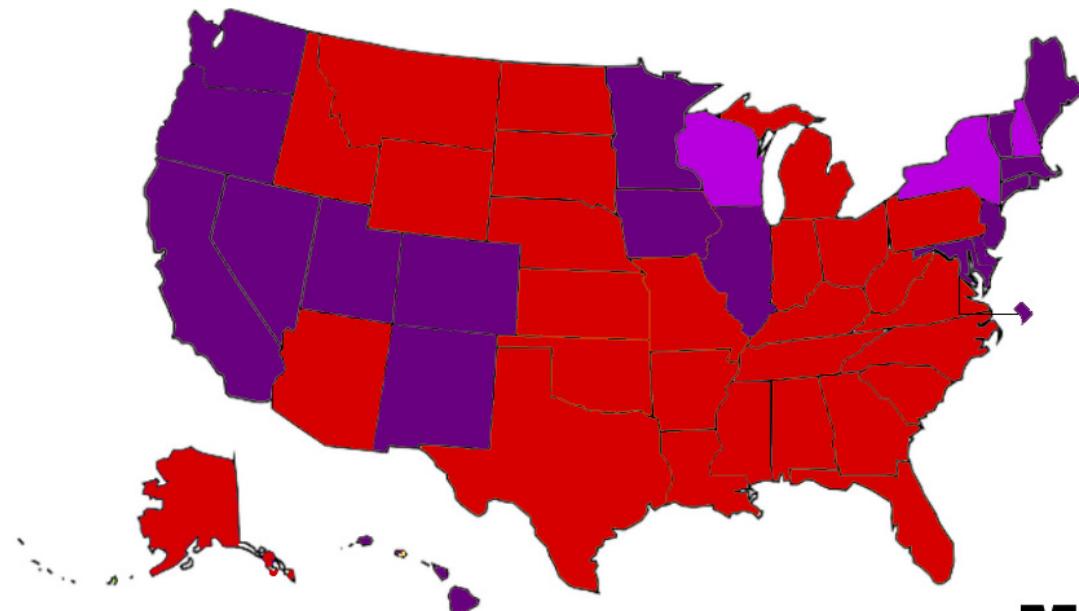
Homeless Youth in the Boston / Somerville Areas

- Boston has the highest homeless population in all of the 25 major cities in the USA. We want to do a qualitative, data-centered approach to humanize these populations; they are people, not facts.
- Interview homeless youth for in-depth stories, shelters for more general information, collect statistics online from academic resources and journal articles.
- Questions to ask youth: How long have you been on the streets? How did you get here? What is your history? Can you map out the places you have travelled over the past couple weeks? What are the things you carry with you? What is your relationship with your family? Who do you trust?
- Questions for shelters: What kinds of visitors do you see? Are they victims of domestic/inter-family violence? Are they people of color? Do they identify as LGBTQ+? Do they struggle with drug addiction?
- Inspired by an interactive history of the Atlantic Slave Trade: http://www.slate.com/articles/life/the_history_of_american_slavery/2015/06/animated_interactive_of_the_history_of_the_atlantic_slave_trade.html
- And a map of Infinite Jest: <http://infiniteatlas.com/location/1440-North-End>
- The goal is to visualize how these stigmas contribute to the structural violence that is poverty & homelessness, while still bringing light to their individual stories and narratives.

A Demographic Analysis of LGBT Groups in the U.S.

States With LGBT Employment Protection

- Employment discrimination for orientation and gender identity
- Employment discrimination for orientation
- No employment nondiscrimination laws



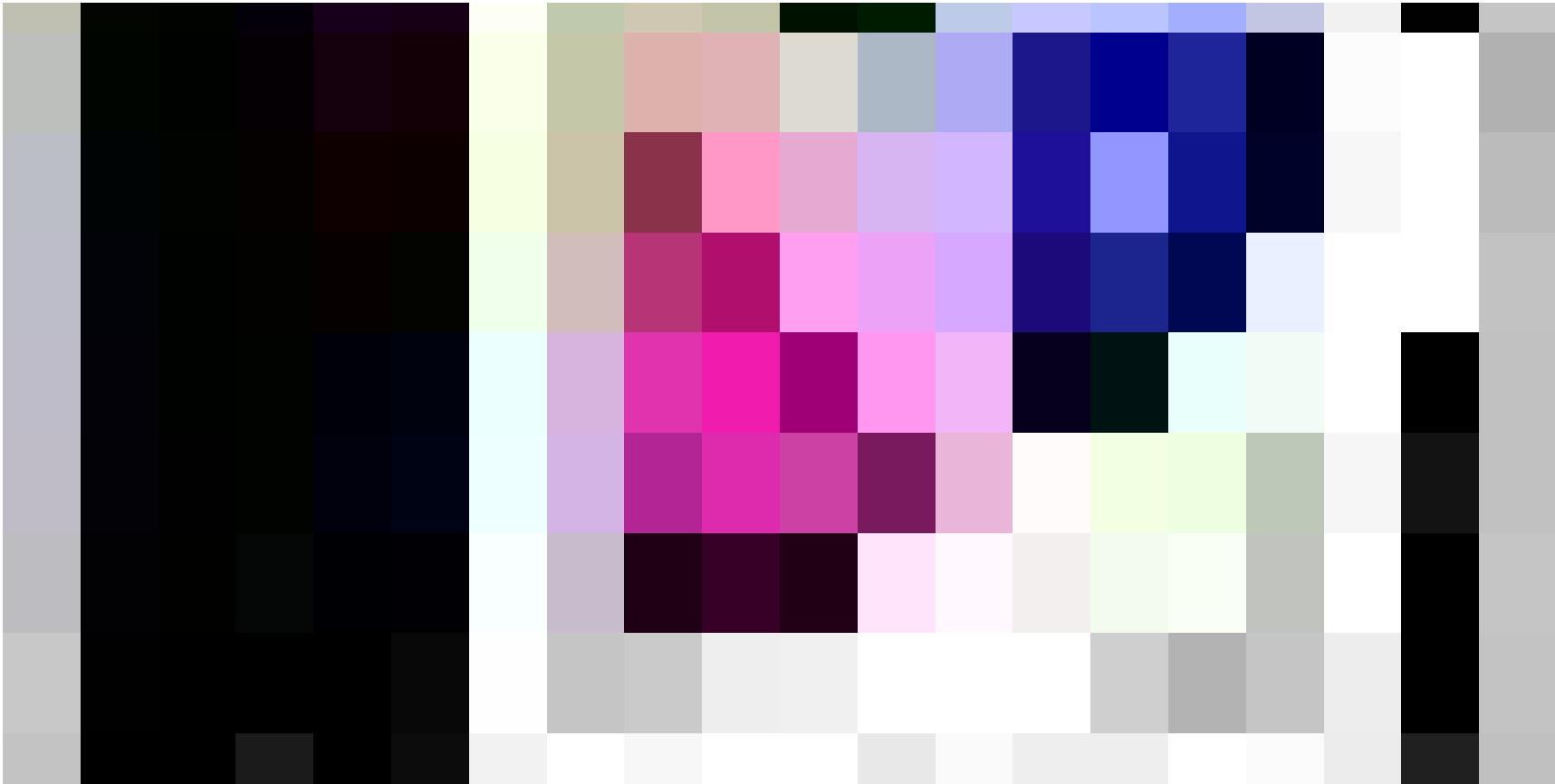
Source: ACLU

.Mic

E



E



Partnership with Tufts Observer



Work on a data journalism piece for the Tufts Observer!

Issue 1

Travel / Tourism

Issue 2

Order / Disorder

Issue 3

Fear / Phobias

Issue 4

Language / Words

Issue 5

Omissions / Lacunae / Gaps

Issue 6

Bodies

Data journalism workshop series



"Learn the skills you need to work on a topic you love."

- | | |
|----------------------------------------------------------------------|-------------------------|
| Getting the Data : Scraping Public Sources in Python | <i>Wed, 9/30 : 8pm</i> |
| Data Science : Answering Journalism Questions with Statistics | <i>Wed, 10/7 : 8pm</i> |
| Natural Language Processing : Trends in Textual Data | <i>Wed, 10/14 : 8pm</i> |
| Data Warehousing : Storing Data for Future Insights | <i>Wed, 10/21 : 8pm</i> |
| Infographics and Design : Visualizing Trends | <i>Wed, 10/28 : 8pm</i> |
| Journalism and Digital Storytelling | <i>Wed, 11/4 : 8pm</i> |
| Symposium | <i>Wed, 12/2 : 8pm</i> |

*Tell us about yourself
and we'll put you in touch with the right people.*

<http://bit.ly/1FswScq>

Thanks!



tuftsenigma@gmail.com