Workshop #3: **Data Visualization.**

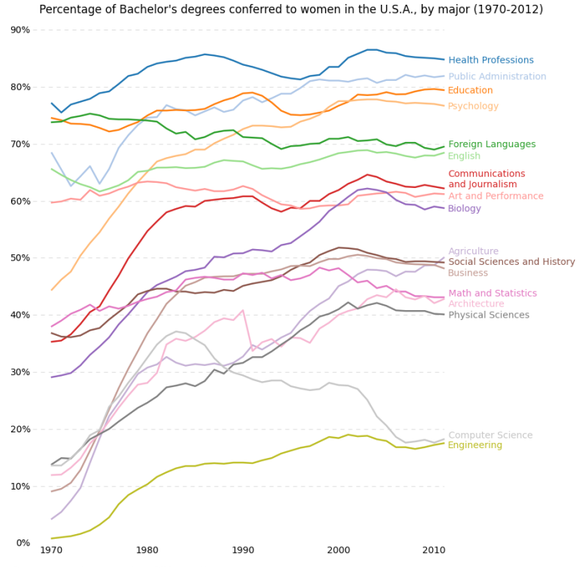
Sam Way
Postdoctoral Researcher
University of Colorado Boulder
[primary author 2018; now at Spotify in NYC]

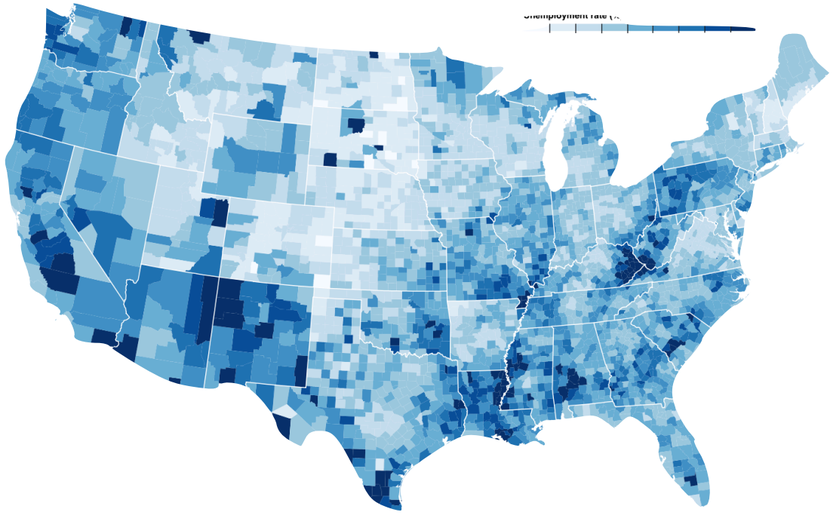
Dan Larremore
Assistant Professor, BioFrontiers & CS
University of Colorado Boulder
[updates and redelivery; 2019]

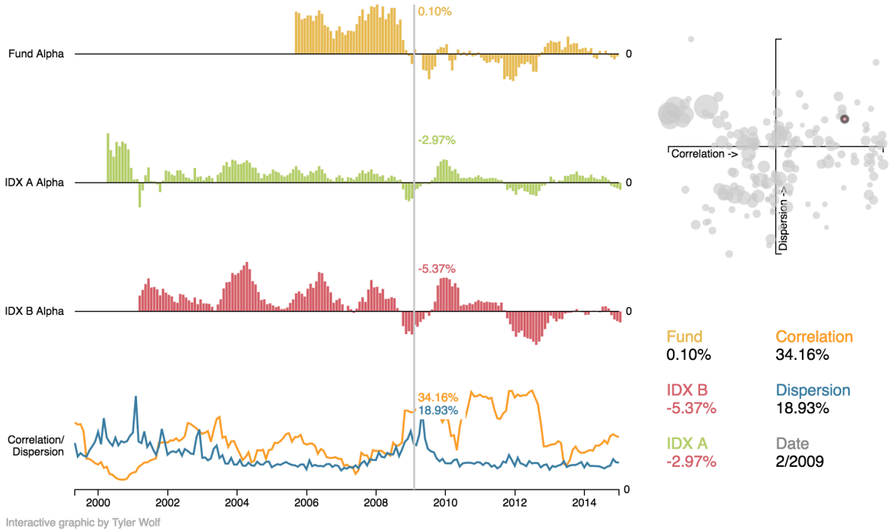
Disclaimer:

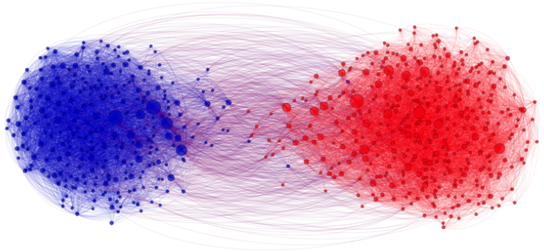
“These are, like, my opinions, man”

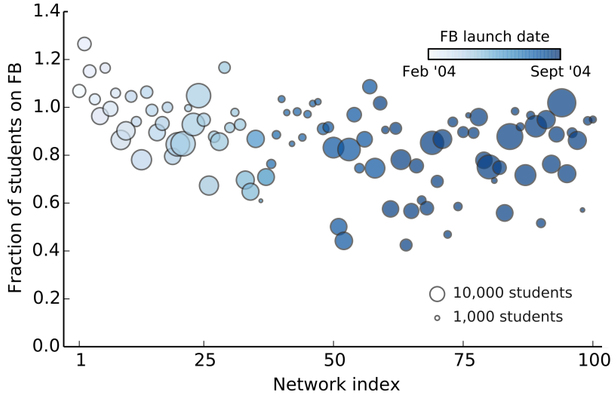
Consider the following as suggestions, not rules.
Developing your own visual style is encouraged.

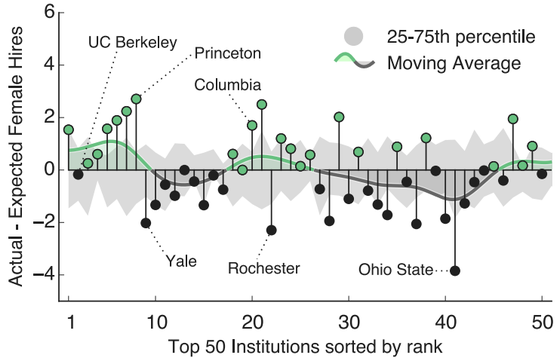

Percentage of Bachelor's degrees conferred to women in the U.S.A., by major (1970-2012)
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1970
1980
1990
2000
2010
Health Professions
Public Administration
Education
Psychology
Foreign Languages
English
Communications and Journalism
Art and Performance
Biology
Agriculture
Social Sciences and History
Business
Math and Statistics
Architecture
Physical Sciences
Computer Science
Engineering
Data source: nces.ed.gov/programs/digest/2013menu_tables.asp
Author: Randy Olson (randalolson.com / @randal_olson)
Note: Some majors are missing because the historical data is not available for them

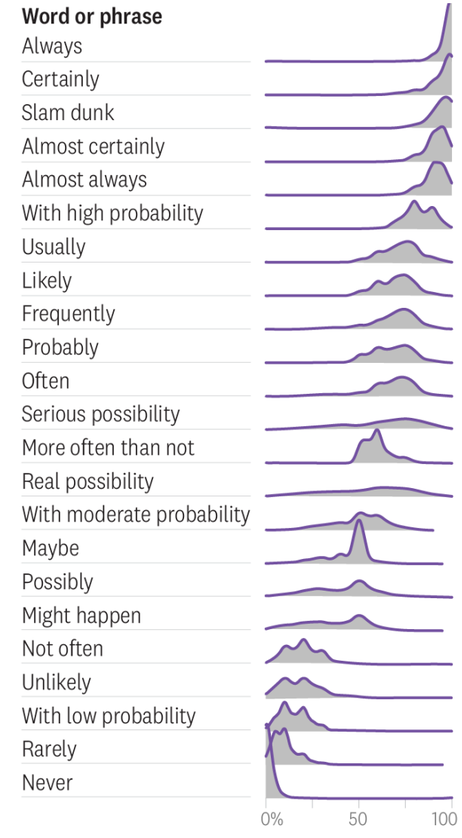


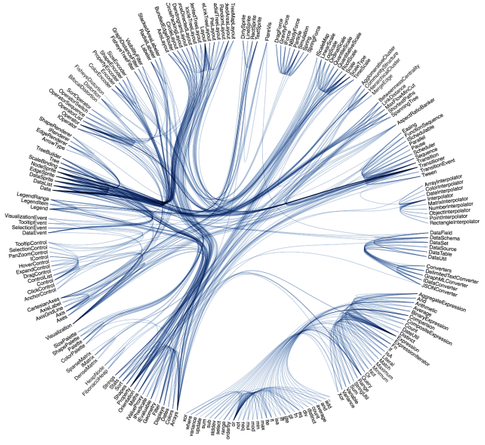

Fund Alpha
0.10%
0
IDX A Alpha
-2.97%
0
IDX B Alpha
-5.37%
0
Correlation/ Dispersion
34.16%
18.93%
0
2000
2002
2004
2006
2008
2010
2012
2014
Interactive graphic by Tyler Wolf
Correlation ->
Dispersion ->
Fund
0.10%
Correlation
34.16%
IDX B
-5.37%
Dispersion
18.93%
IDX A
-2.97%
Date
2/2009

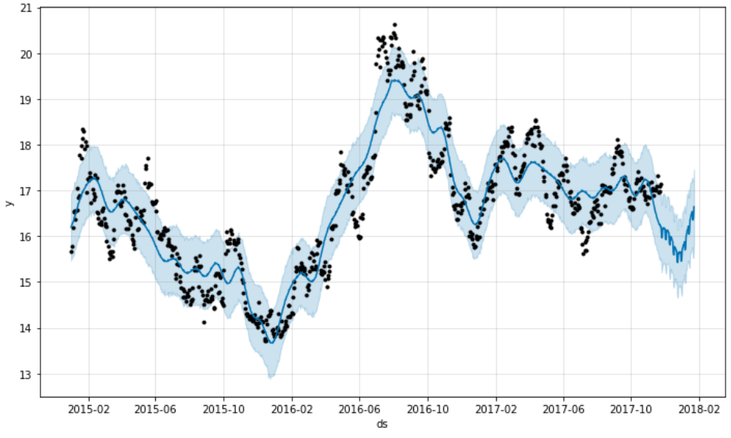


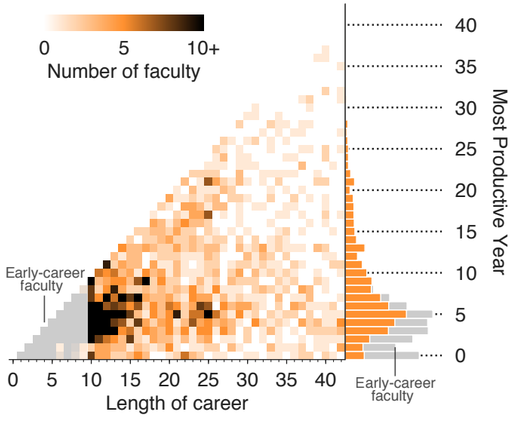

FB launch date
Feb '04
Sept '04
Fraction of students on FB
Network index
10,000 students
1,000 students


25-75th percentile
Moving Average
UC Berkeley
Princeton
Columbia
Yale
Rochester
Ohio State
Actual - Expected Female Hires
Top 50 Institutions sorted by rank


Word or phrase
Always
Certainly
Slam dunk
Almost certainly
Almost always
With high probability
Usually
Likely
Frequently
Probably
Often
Serious possibility
More often than not
Real possibility
With moderate probability
Maybe
Possibly
Might happen
Not often
Unlikely
With low probability
Rarely
Never
0%
50
100




y
ds


Number of faculty
Early-career faculty
Length of career
Most Productive Year
Early-career faculty

Tips for creating visualizations.

During each stage of creation:

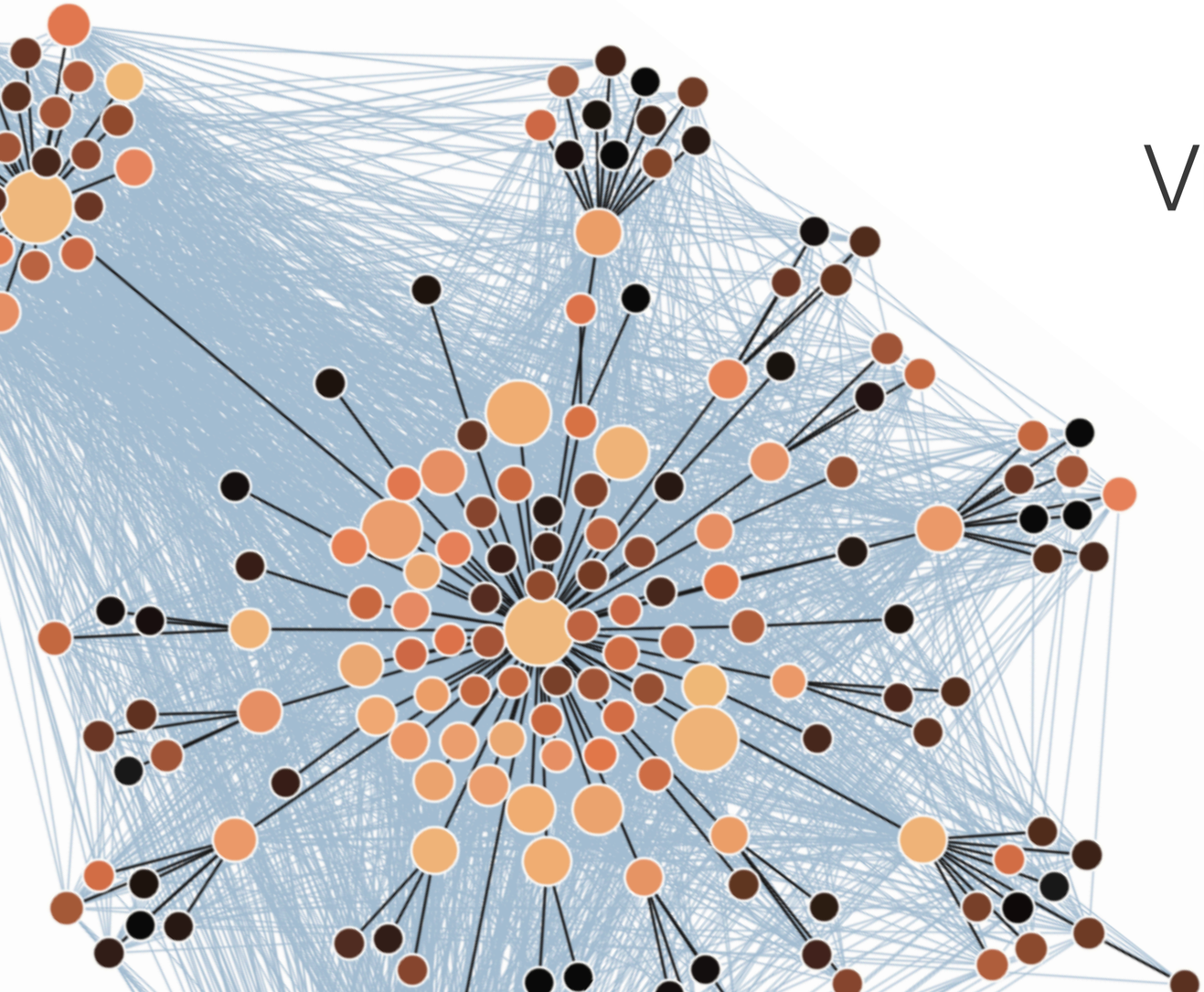
0. Determine your goals
1. Select appropriate type of visualization
2. Build prototypes
3. Gather and address feedback

Tips for creating visualizations.

During each stage of creation:

0. Determine your goals
1. Select appropriate type of visualization
2. Build prototypes
3. Gather and address feedback

First, why make visualizations?



Visualizations can…

clarify your story,

emphasize a message,

build trust with audience,

inform their decisions.

Note: bad visualizations can do the opposite of all these things!

Set goals for your visualization.

Visualizations can serve several purposes.
Think about which goals you have.

Telling people what to see
- **clarify** your story,
- **emphasize** a message,

Allowing people to explore
- **build trust** with audience,
- **inform** their decisions.

Possible differentiation between information visualization and “infographics” (see: https://www.denizcemonduygu.com/2015/06/redefining-infographics/).

Set goals for your visualization.

Visualizations can serve several purposes.
Think about which goals you have.

Bottom line: Make sure your visualization has a purpose.

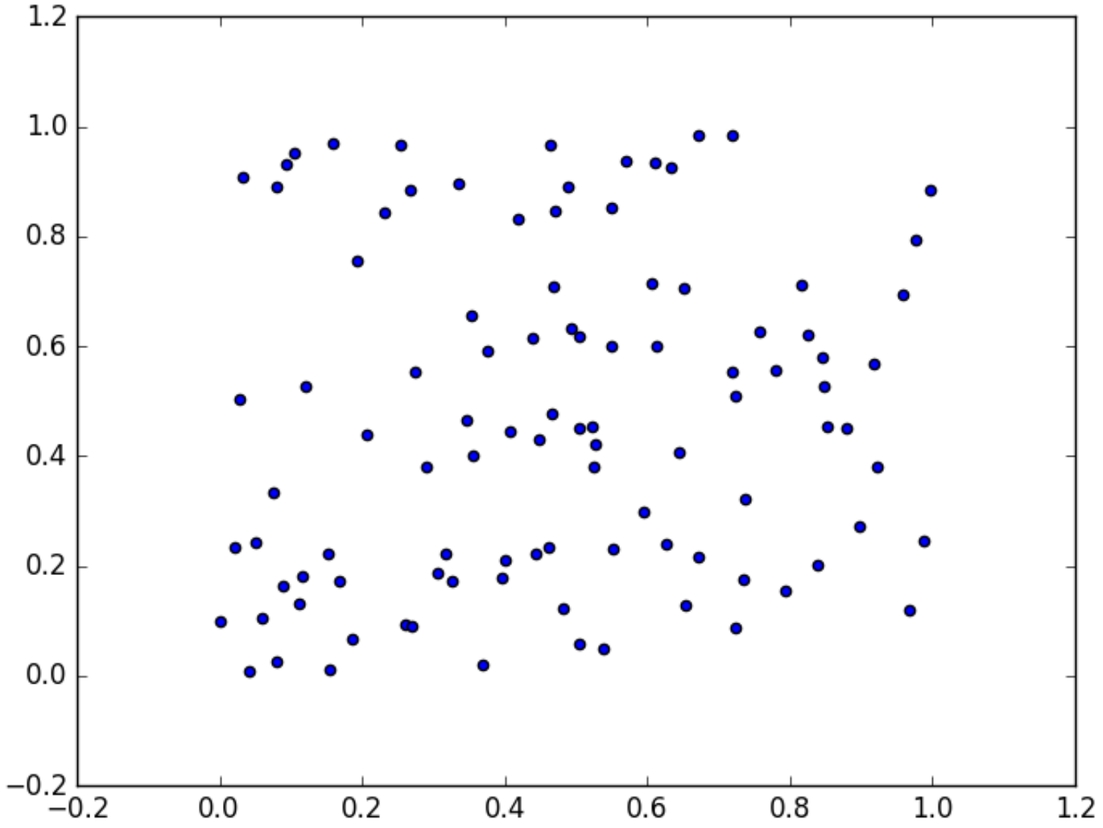
Telling people what to see { clarify your story, emphasize a message,

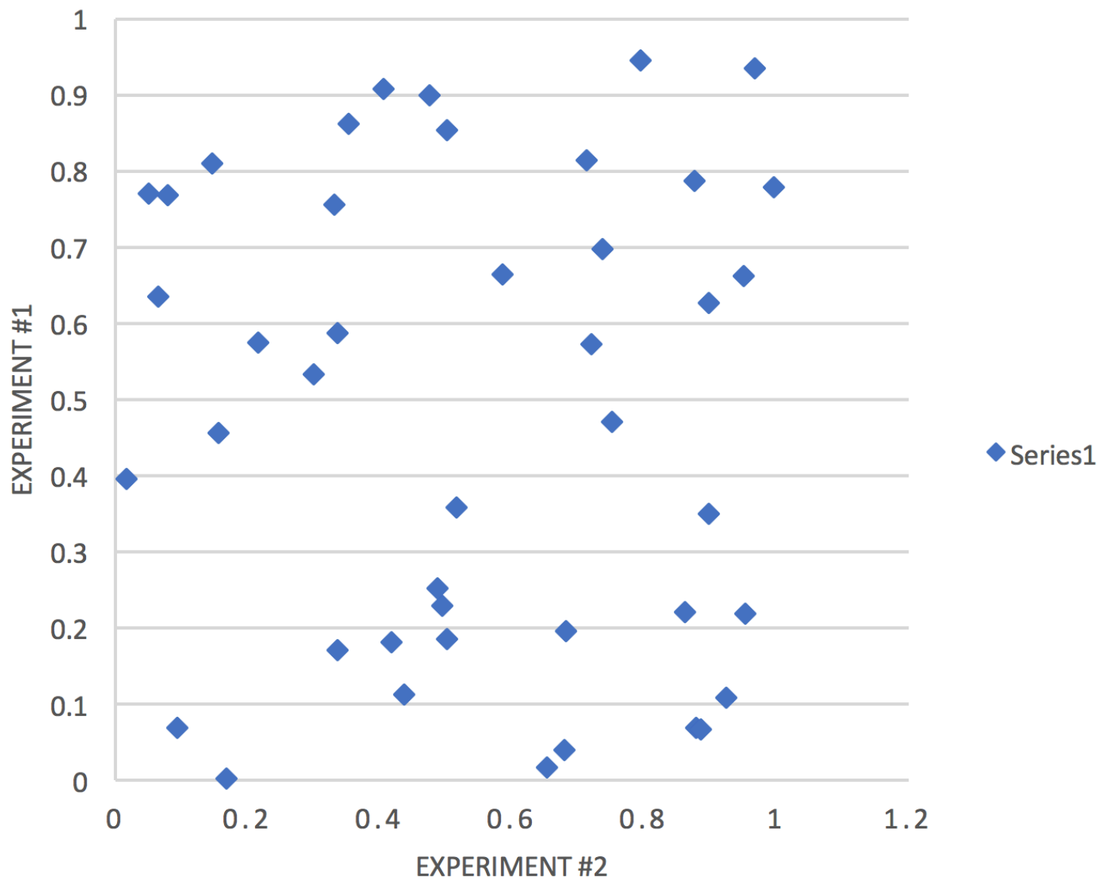
Allowing people to explore { **build trust** with audience, **inform** their decisions.

Possible differentiation between information visualization and "infographics" (see: https://www.denizcemonduygu.com/2015/06/redefining-infographics/).

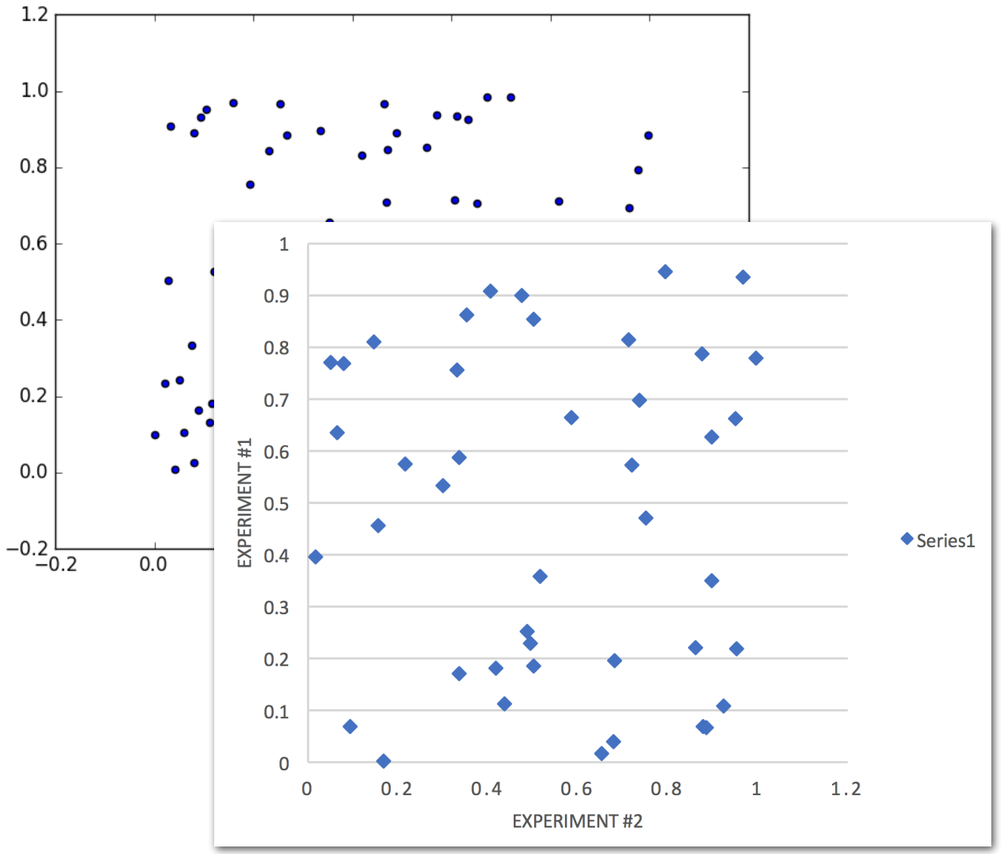
Be aware of unintentional messages.

What can you tell me about the person who made these figures?





Be aware of unintentional messages.



Defaults aren't necessarily bad.

But, they might suggest you put little effort into your visualizations.

In computer science, Excel figures might raise concerns about you.*

Put effort into your visualizations.

* Great science can be and is done in Excel, and the value of science should be independent of the person or tool that created it. That said, be mindful of the biases that may exist in your fields.

Tips for creating visualizations.

During each stage of creation:

0. Determine your goals
1. Select appropriate type of visualization
2. Build prototypes
3. Gather and address feedback

Make use of expected visuals.

People have intuition for standard viz types. Start there.

“The variables are related”

“The distribution is skewed”

“The groups differ in Y”

“Affects the Southwest”

Think about how tweaking even a very standard plot might make it more useful (e.g. https://www.nature.com/articles/d41586-018-06912-1).

Make use of expected visuals.

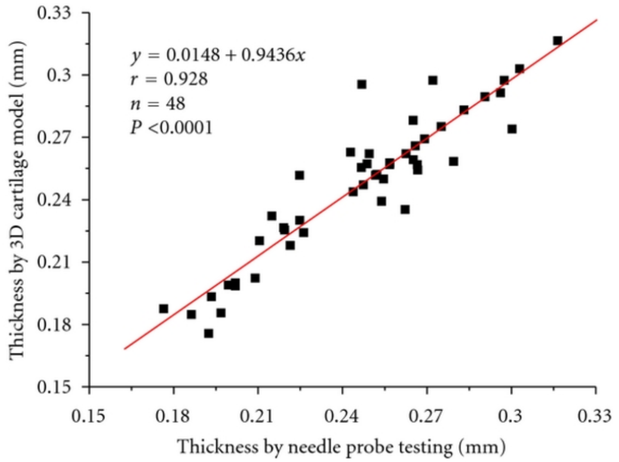
People have intuition for standard viz types. Start there.

"The variables are related"

"The distribution is skewed"

"The groups differ in Y"

"Affects the Southwest"

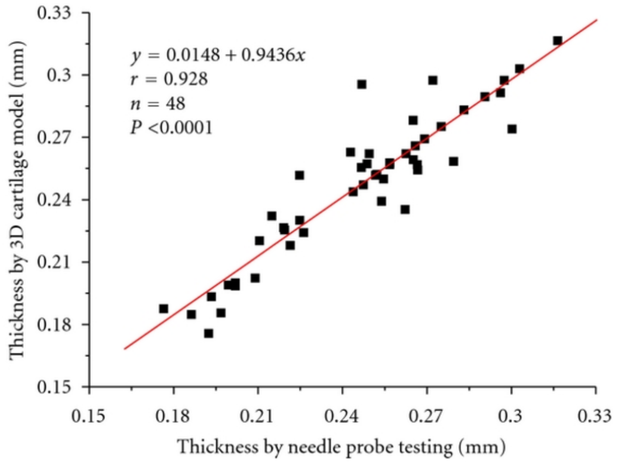


Think about how tweaking even a very standard plot might make it more useful (e.g. https://www.nature.com/articles/d41586-018-06912-1).

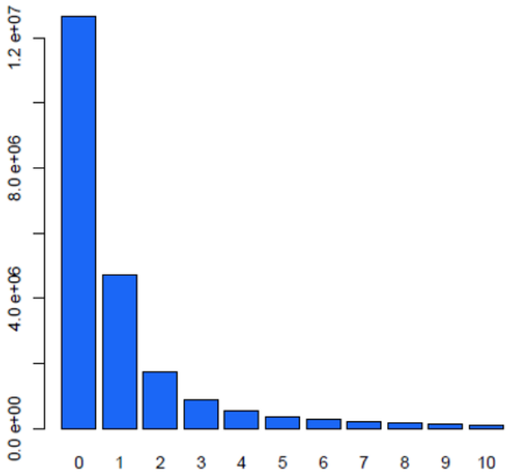
Make use of expected visuals.

People have intuition for standard viz types. Start there.

"The variables are related"



"The distribution is skewed"



"The groups differ in Y"

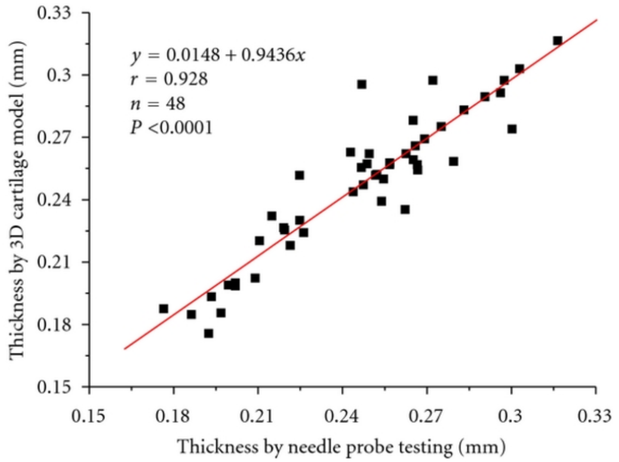
"Affects the Southwest"

Think about how tweaking even a very standard plot might make it more useful (e.g. https://www.nature.com/articles/d41586-018-06912-1).

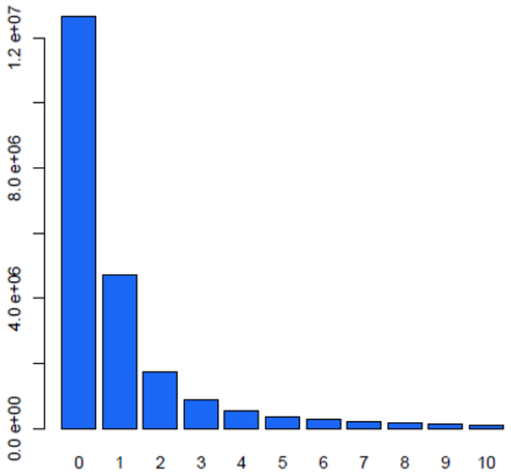
Make use of expected visuals.

People have intuition for standard viz types. Start there.

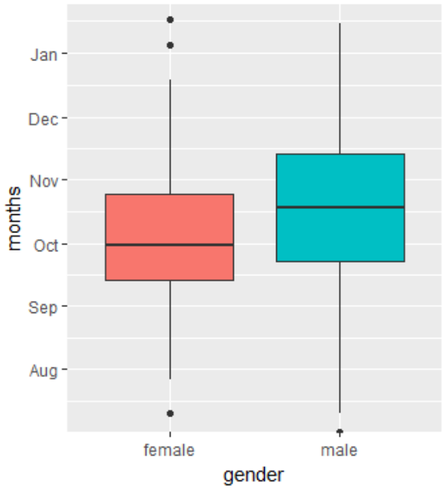
"The variables are related"



"The distribution is skewed"



"The groups differ in Y"



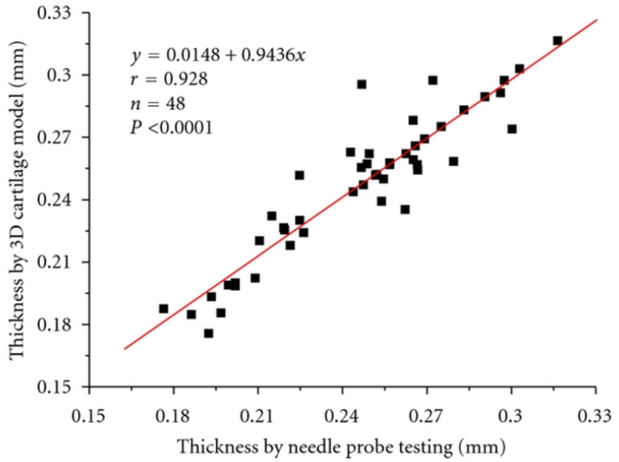
"Affects the Southwest"

Think about how tweaking even a very standard plot might make it more useful (e.g. https://www.nature.com/articles/d41586-018-06912-1).

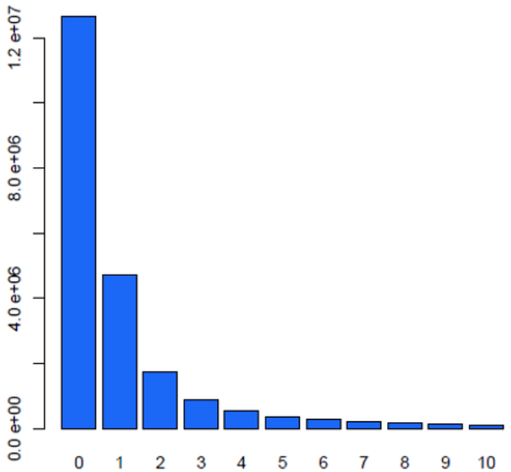
Make use of expected visuals.

People have intuition for standard viz types. Start there.

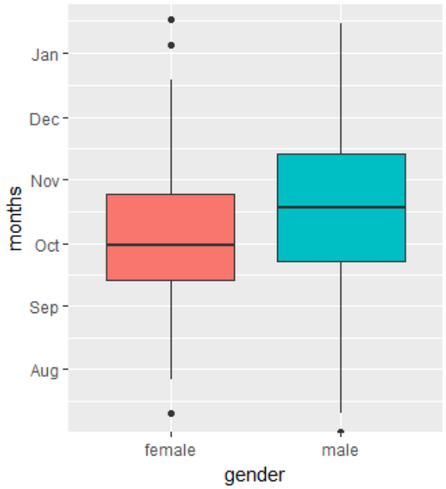
"The variables are related"



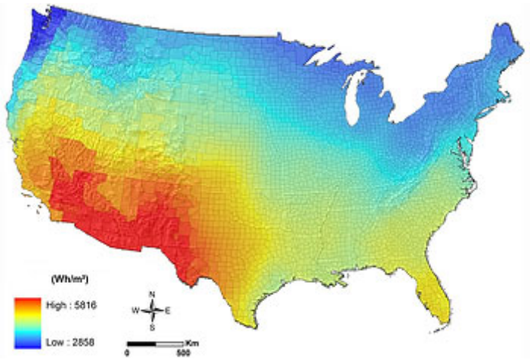
"The distribution is skewed"



"The groups differ in Y"



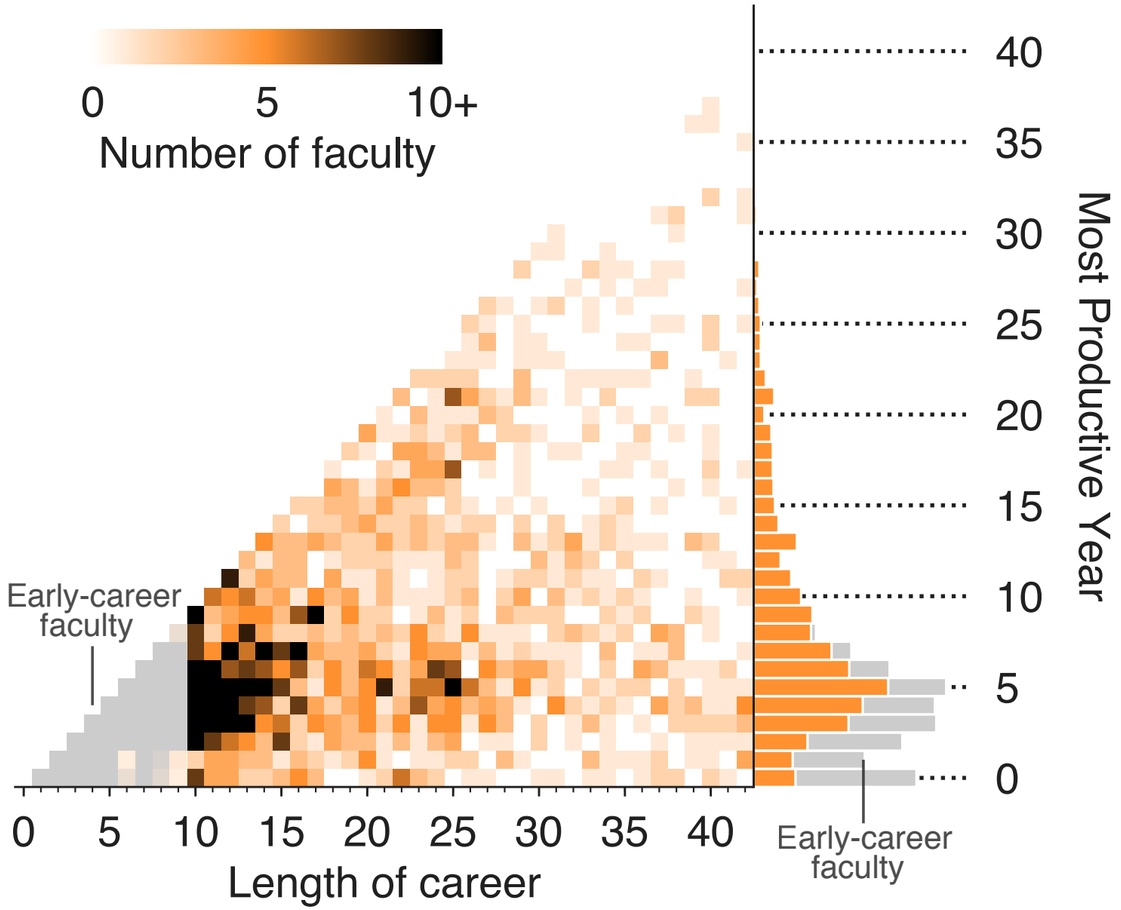
"Affects the Southwest"



Think about how tweaking even a very standard plot might make it more useful (e.g. https://www.nature.com/articles/d41586-018-06912-1).

Combine multiple types.

Example: when are faculty most productive?

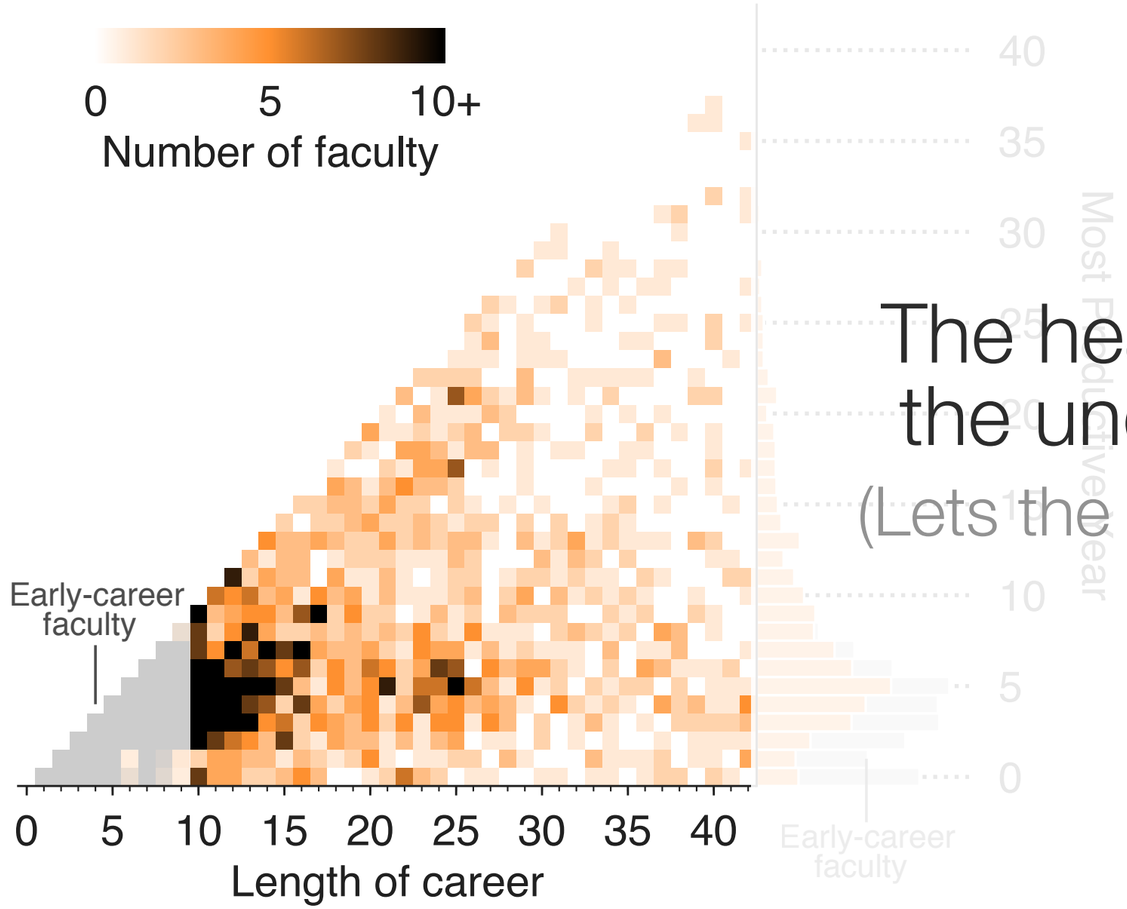


Combine multiple types.

Example: when are faculty most productive?

Combine multiple types.

Example: when are faculty most productive?



The heatmap reveals the underlying data.

(Lets the audience explore)

Tips for creating visualizations.

During each stage of creation:

0. Determine your goals
1. Select appropriate type of visualization
2. Build prototypes
3. Gather and address feedback

Prototype your visualization.

Advice from Prof. Danielle Szafir:

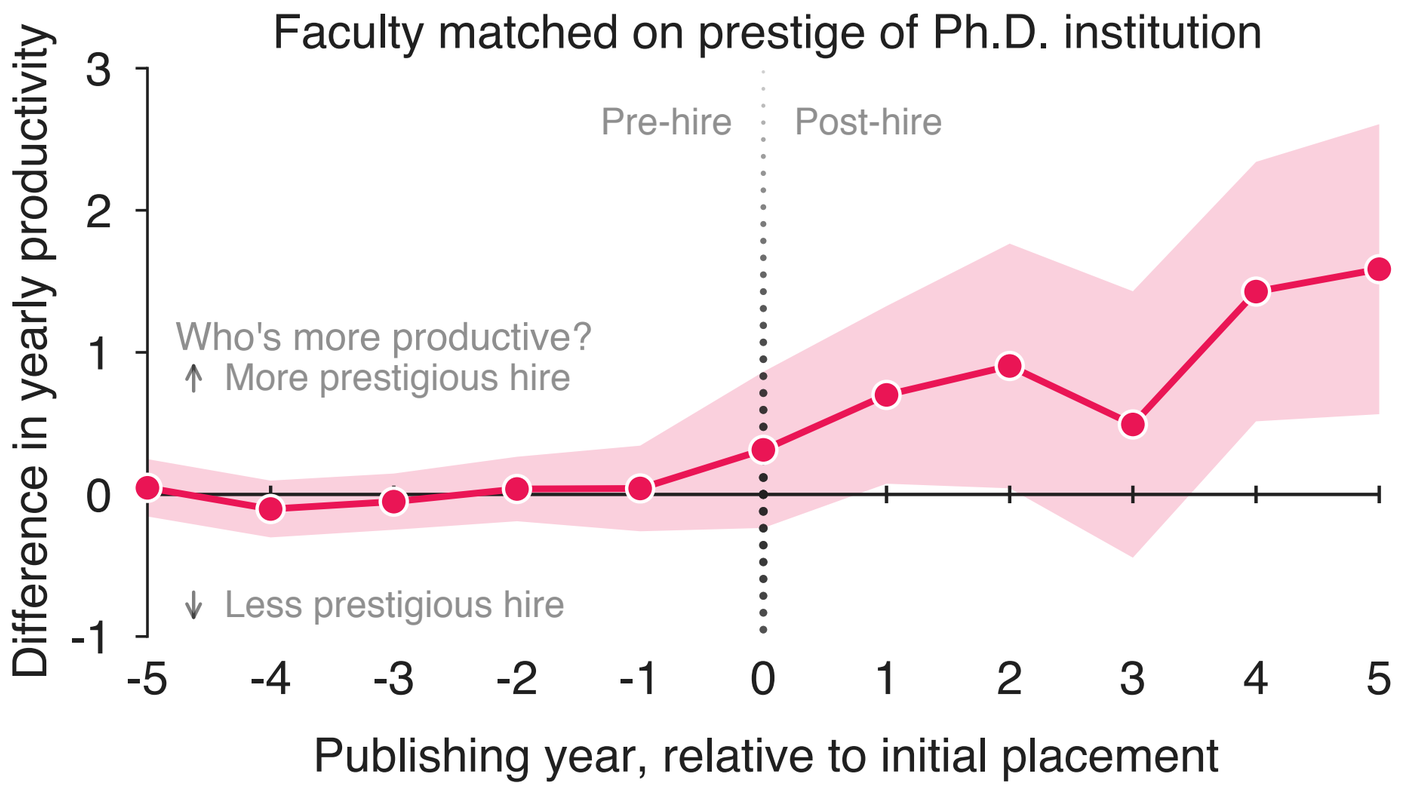
Before you code up anything, get out your colored pencils, and *draw*.

Come up with a handful of concepts and reflect on what you like about each one of them.

Ok, so maybe drawing isn't your thing. Point is: get rough ideas down on paper, matplotlib, whatever. Iterate quickly before polishing, adding interactivity, etc.

Imagine visualizing a sentence.

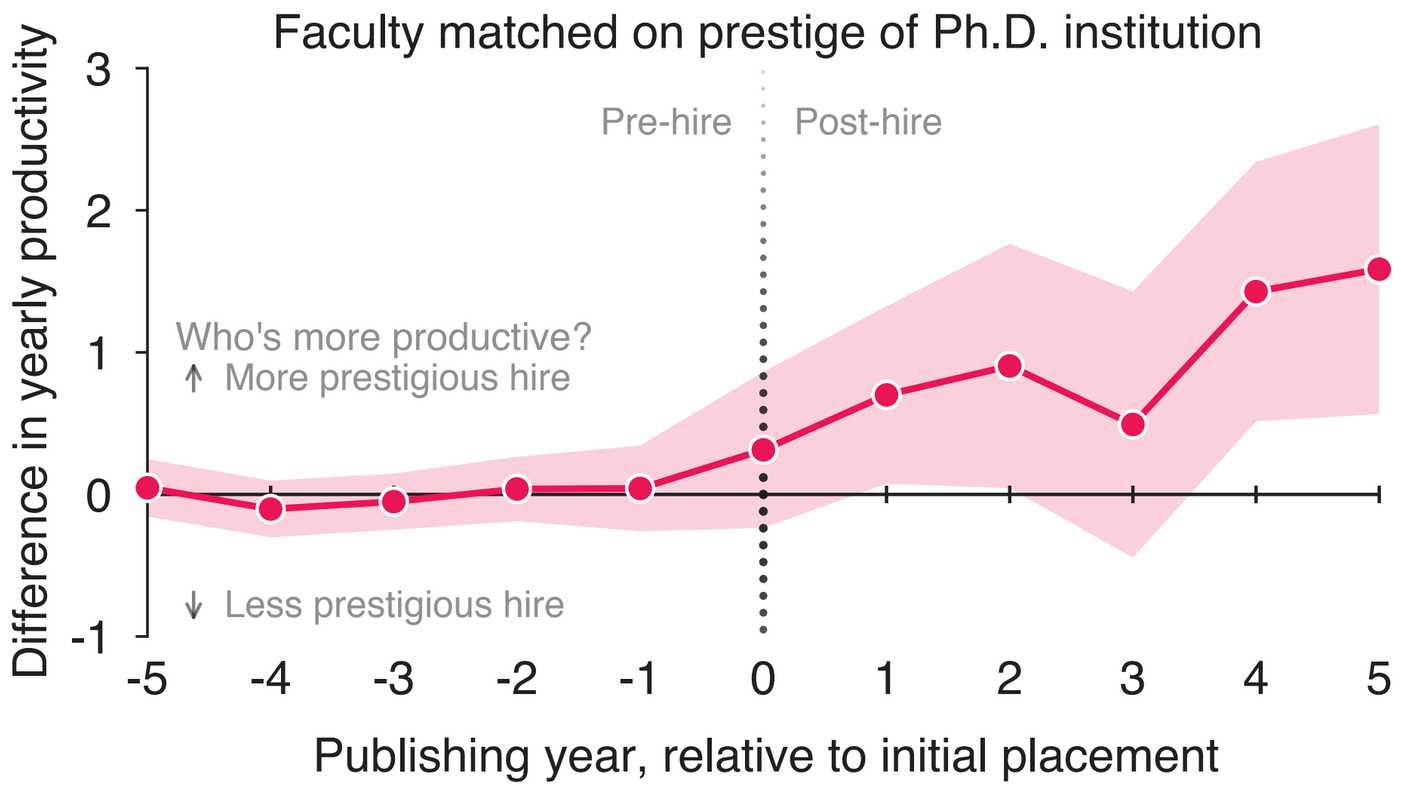
Focus on clearly communicating a single idea.



The sentence here: if you take two researchers with similar training and give one a more prestigious job than the other, that person tends to become more productive.

Imagine visualizing a sentence.

Focus on clearly communicating a single idea.

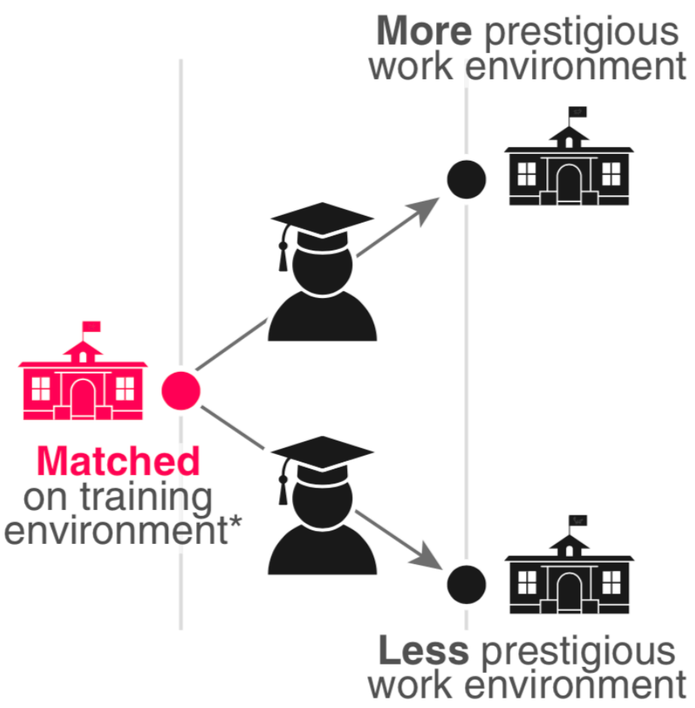


The sentence was not clearly enough communicated. So we iterated…

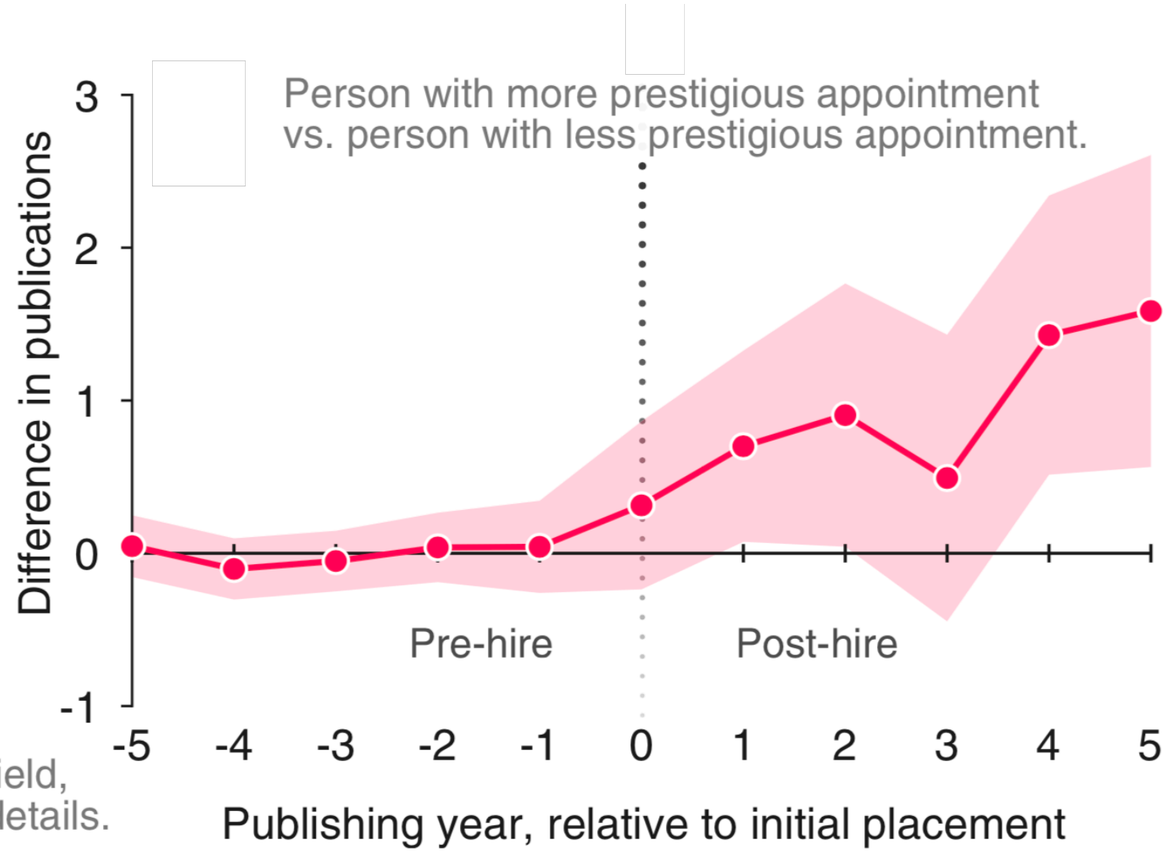
The sentence here: if you take two researchers with similar training and give one a more prestigious job than the other, that person tends to become more productive.

Imagine visualizing a sentence.

Focus on clearly communicating a single idea.



*: Faculty also matched on gender, subfield, and other features. See main text for full details.

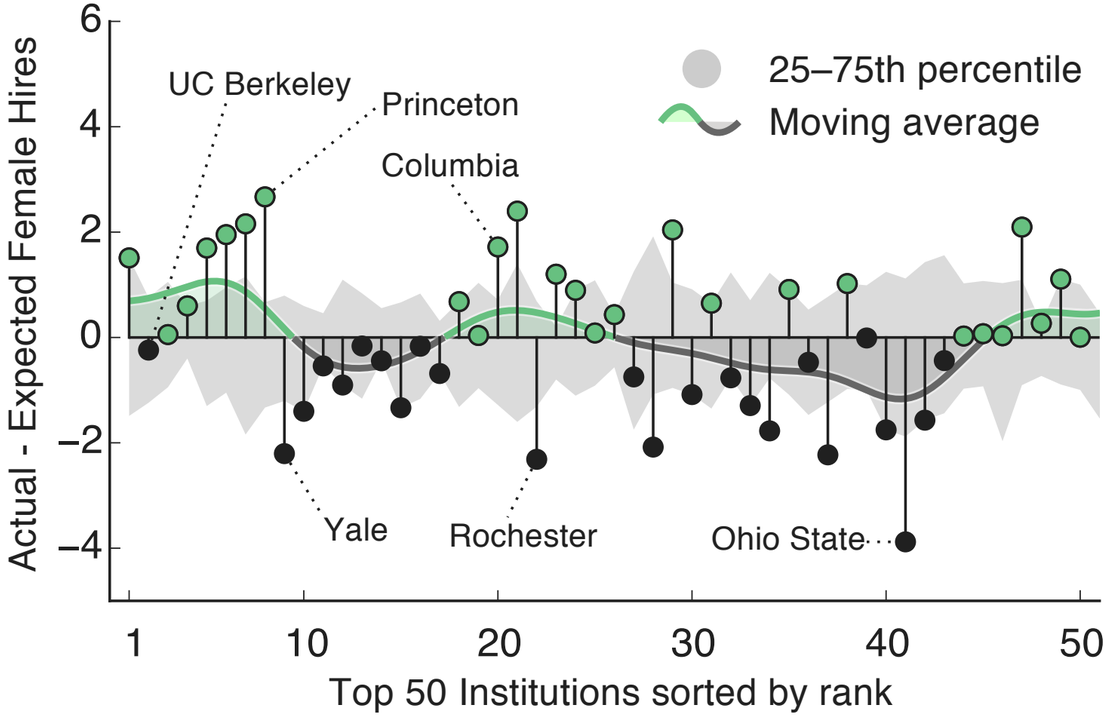


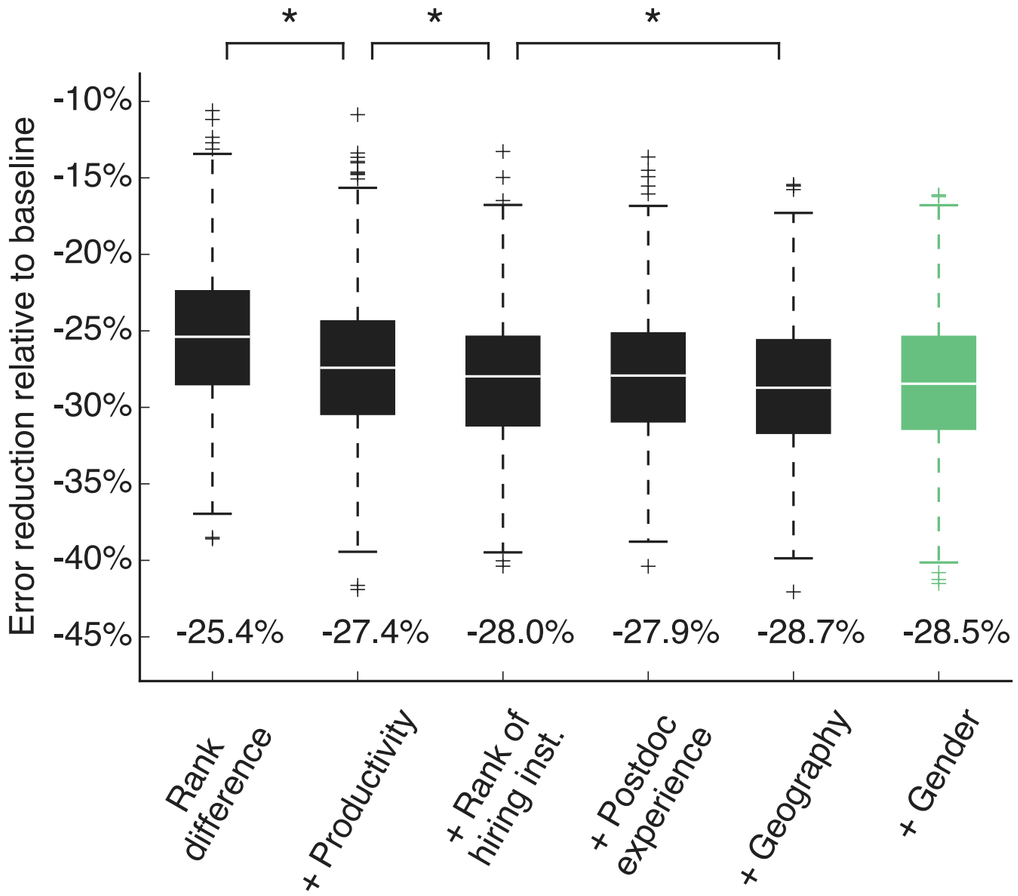
The sentence here: if you take two researchers with similar training and give one a more prestigious job than the other, that person tends to become more productive.

Way, Morgan, Larremore, Clauset. PNAS, 2019. https://www.pnas.org/content/116/22/10729

Use color as an accent.

Almost all of our figures use grayscale + one accent color



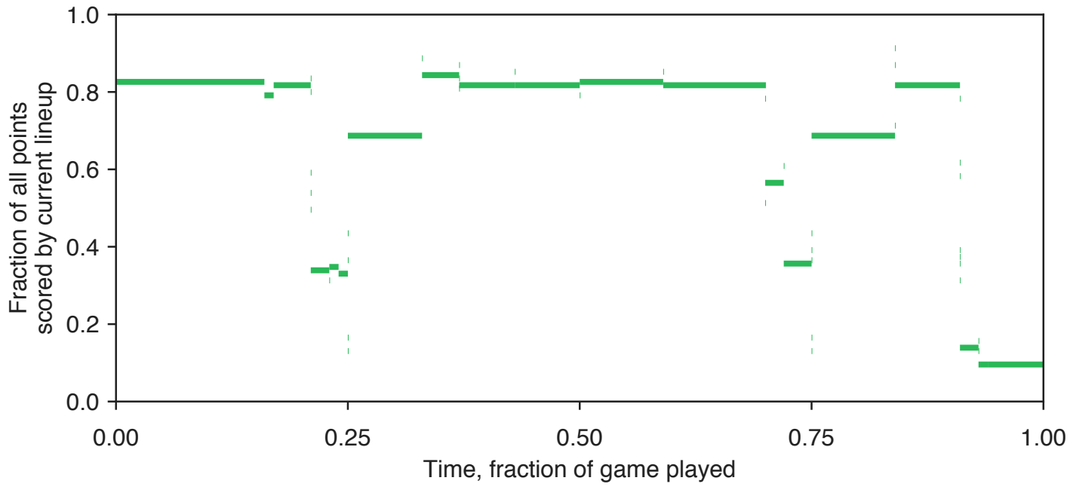


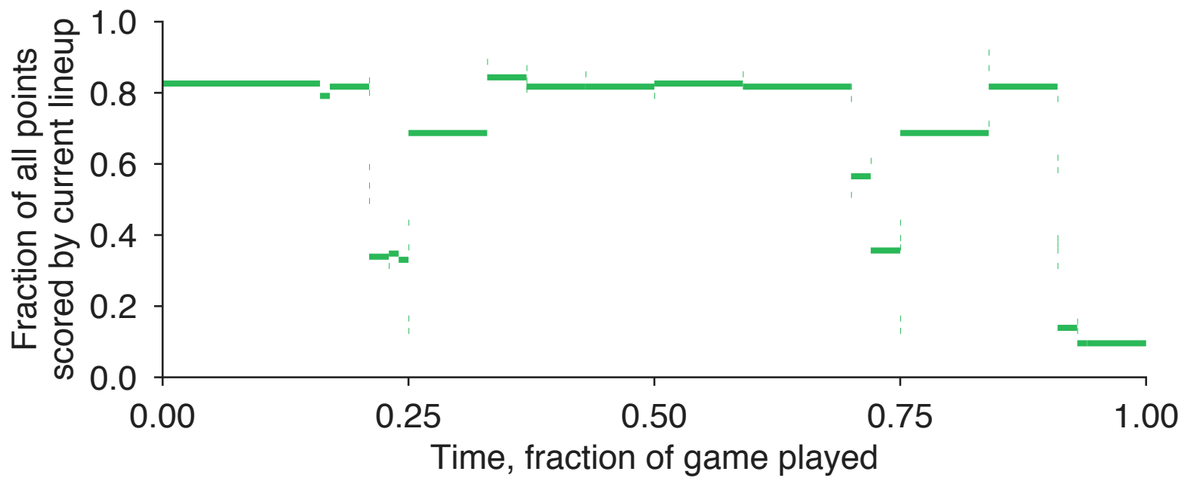
[1] Way et al.. *Gender, Productivity, and Prestige in Computer Science Faculty Hiring Networks*. (2016)

"Usefully restricting": this gives you maybe three grays + a color to play with, which forces you to think about whether you should simplify your visualization.

Remove unnecessary design elements.

If it doesn’t have a purpose, take it out.

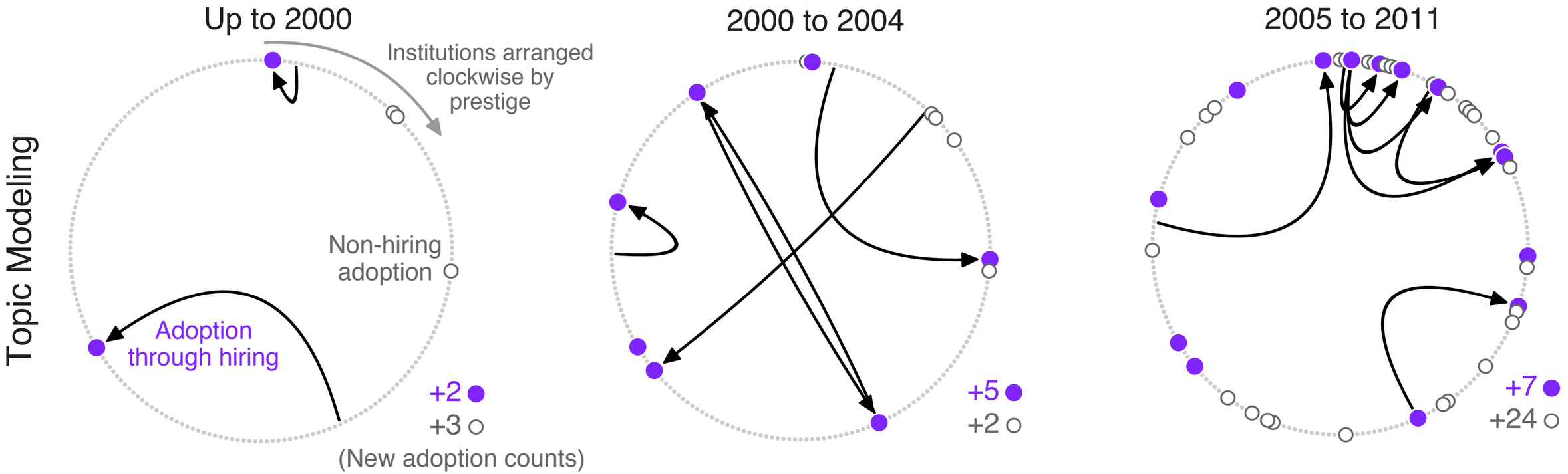




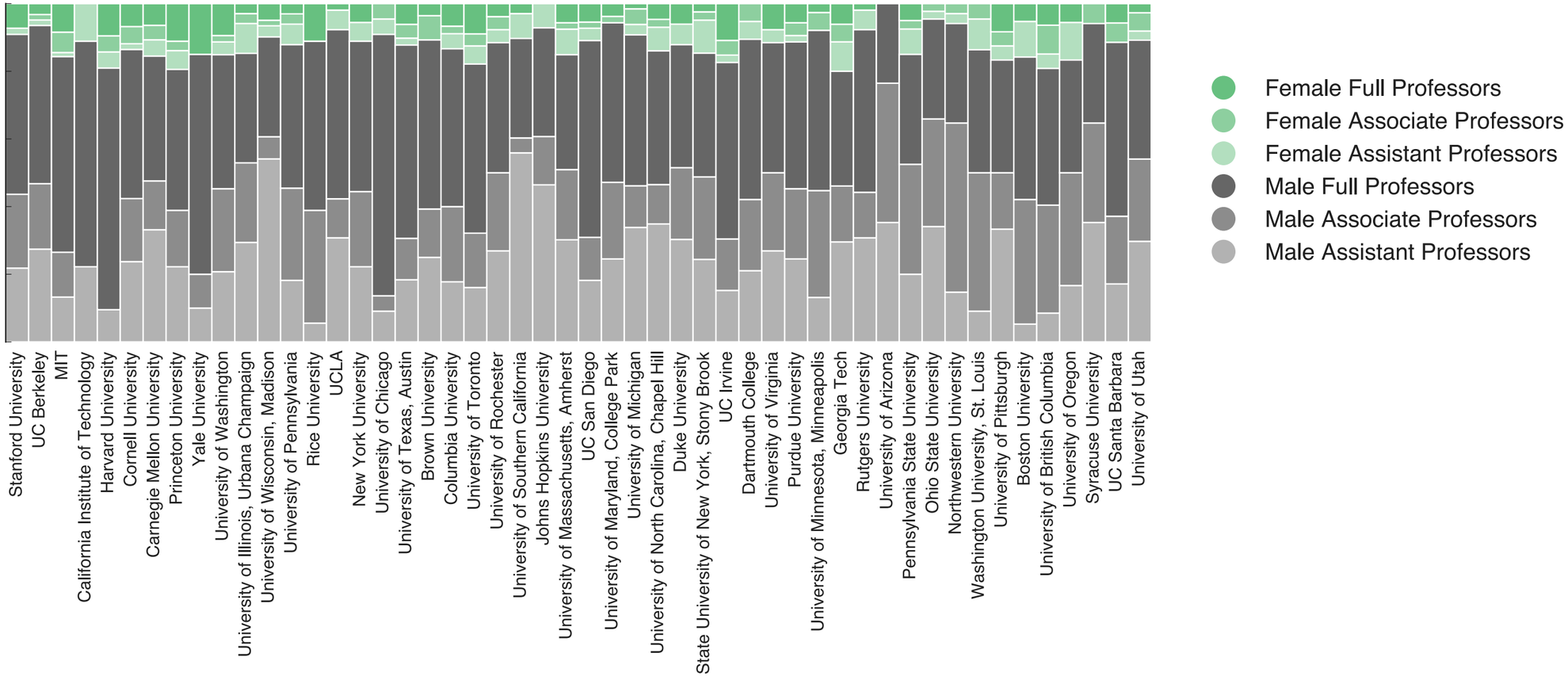
This goes for the design and the data itself: what are you trying to say? Is X required or helpful for saying it?

Give the audience instructions.

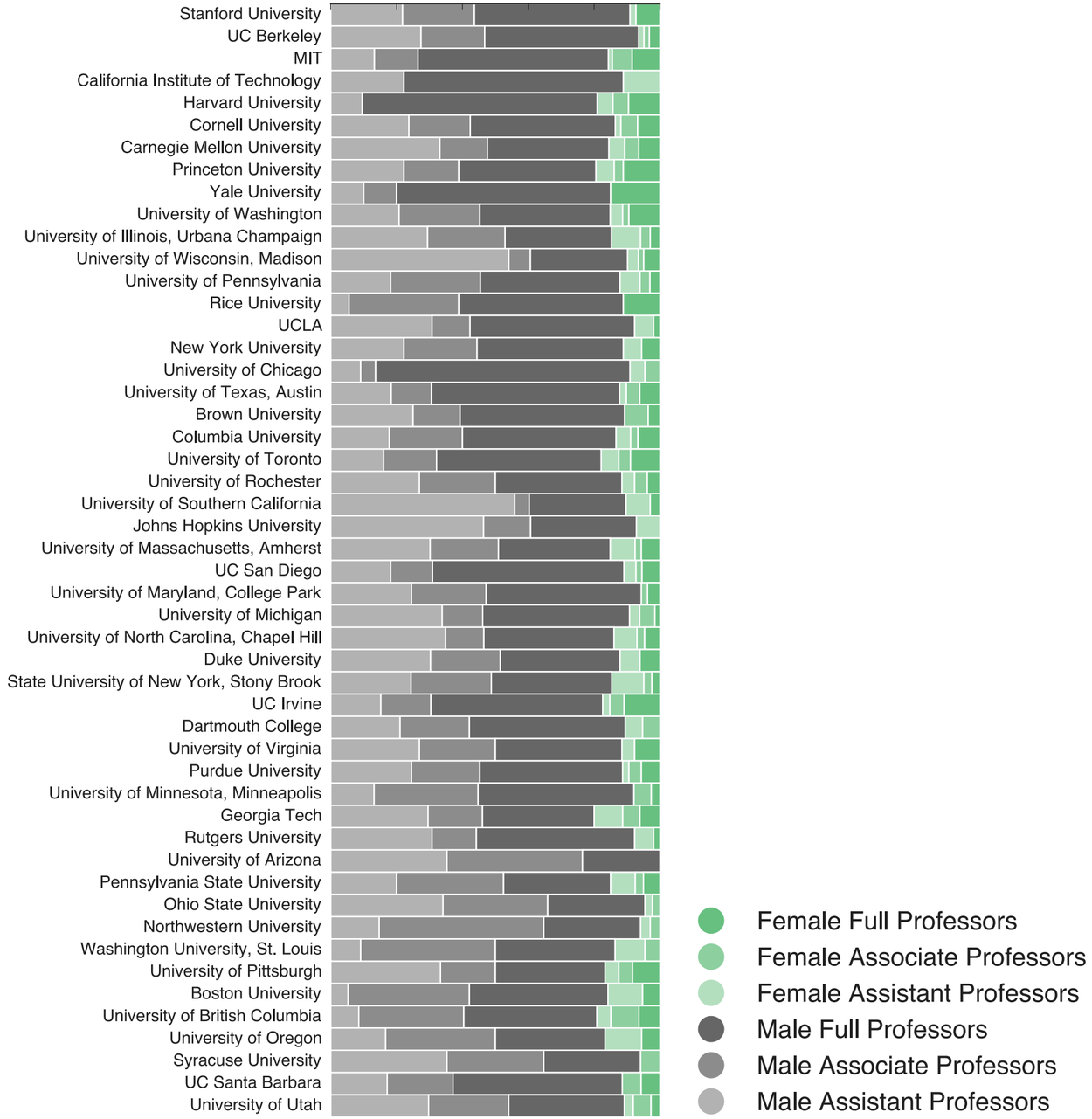
If possible, save them from having to consult the main text or caption.



Simulate your audience.



What's the first thing that we do when we see a list of Universities?


Stanford University
UC Berkeley
MIT
California Institute of Technology
Harvard University
Cornell University
Carnegie Mellon University
Princeton University
Yale University
University of Washington
University of Illinois, Urbana Champaign
University of Wisconsin, Madison
University of Pennsylvania
Rice University
UCLA
New York University
University of Chicago
University of Texas, Austin
Brown University
Columbia University
University of Toronto
University of Rochester
University of Southern California
Johns Hopkins University
University of Massachusetts, Amherst
UC San Diego
University of Maryland, College Park
University of Michigan
University of North Carolina, Chapel Hill
Duke University
State University of New York, Stony Brook
UC Irvine
Dartmouth College
University of Virginia
Purdue University
University of Minnesota, Minneapolis
Georgia Tech
Rutgers University
University of Arizona
Pennsylvania State University
Ohio State University
Northwestern University
Washington University, St. Louis
University of Pittsburgh
Boston University
University of British Columbia
University of Oregon
Syracuse University
UC Santa Barbara
University of Utah
Female Full Professors
Female Associate Professors
Female Assistant Professors
Male Full Professors
Male Associate Professors
Male Assistant Professors

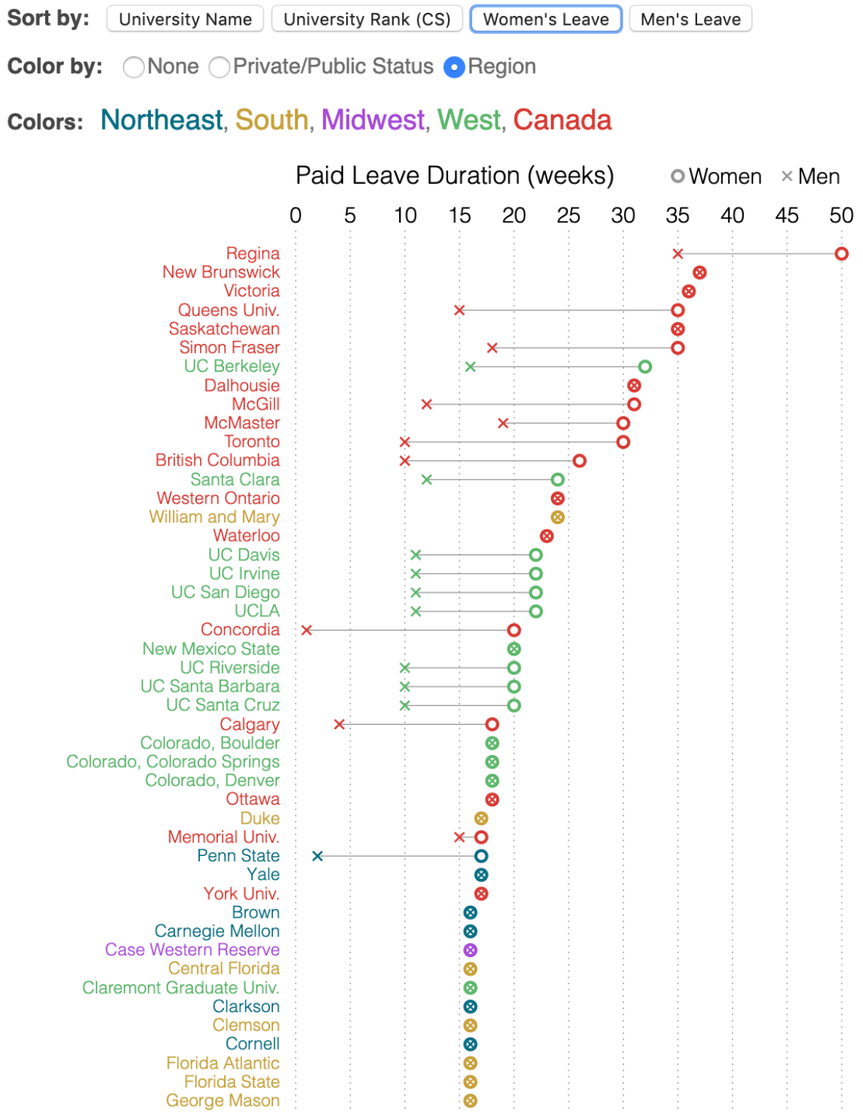
Simulate your audience. (part 2)

Paid Parental Leave at US and Canadian Universities

A dataset of parental leave policies. Equal contributions by Allison C. Morgan, Samuel F. Way, Mirta Galesic, Daniel B. Larremore, and Aaron Clauset.

View the Project on GitHub
aaronclauset/parental-leave

Download ZIP File	Download TAR Ball	View On GitHub



When given the ability to sort a list, what is the first thing people do?

aaronclauset.github.io/parental-leave/

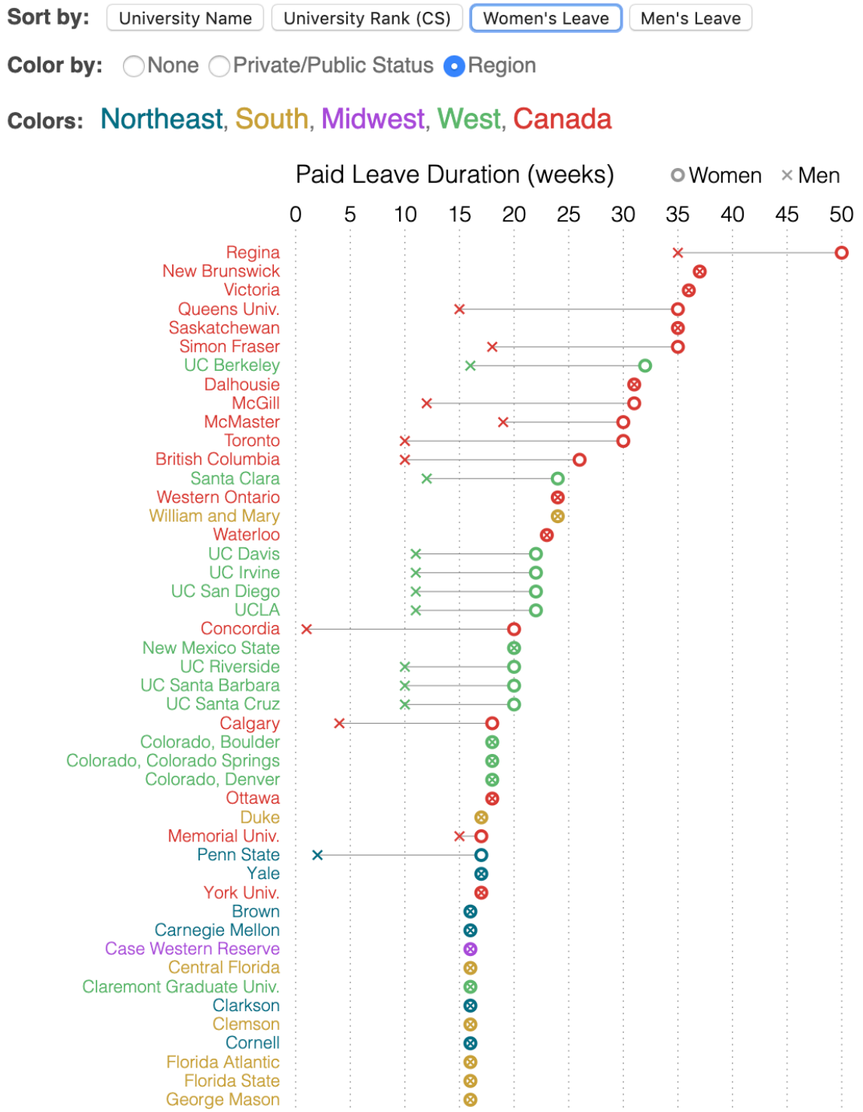
Simulate your audience. (part 2)

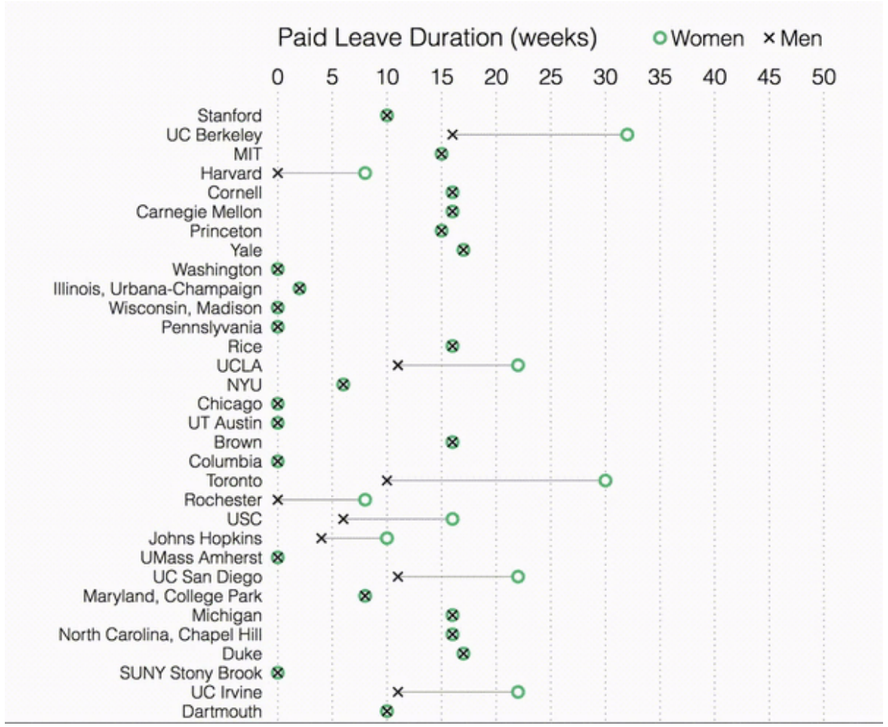
Paid Parental Leave at US and Canadian Universities

A dataset of parental leave policies. Equal contributions by Allison C. Morgan, Samuel F. Way, Mirta Galesic, Daniel B. Larremore, and Aaron Clauset.

View the Project on GitHub
aaronclauset/parental-leave

Download **ZIP File**	Download **TAR Ball**	View On **GitHub**





Show users they can sort.

aaronclauset.github.io/parental-leave/

Link to Movie [Twitter @samfway]

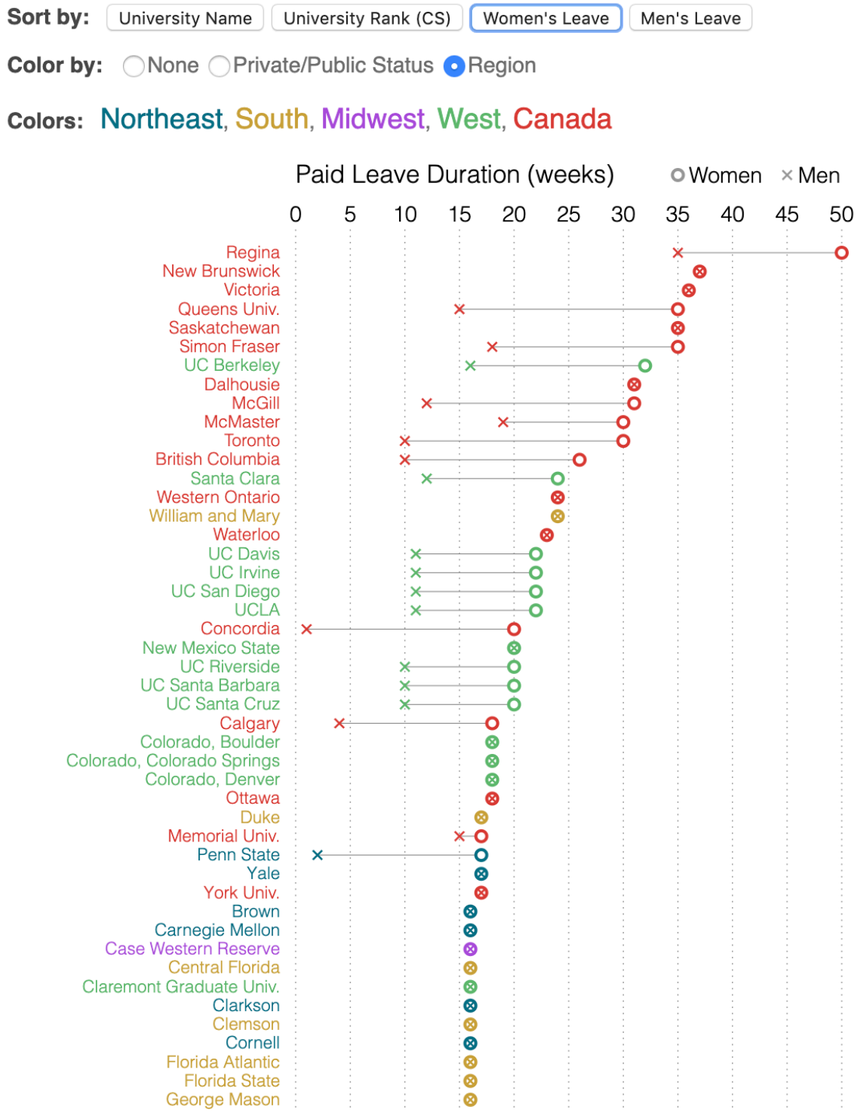
Simulate your audience. (part 2)

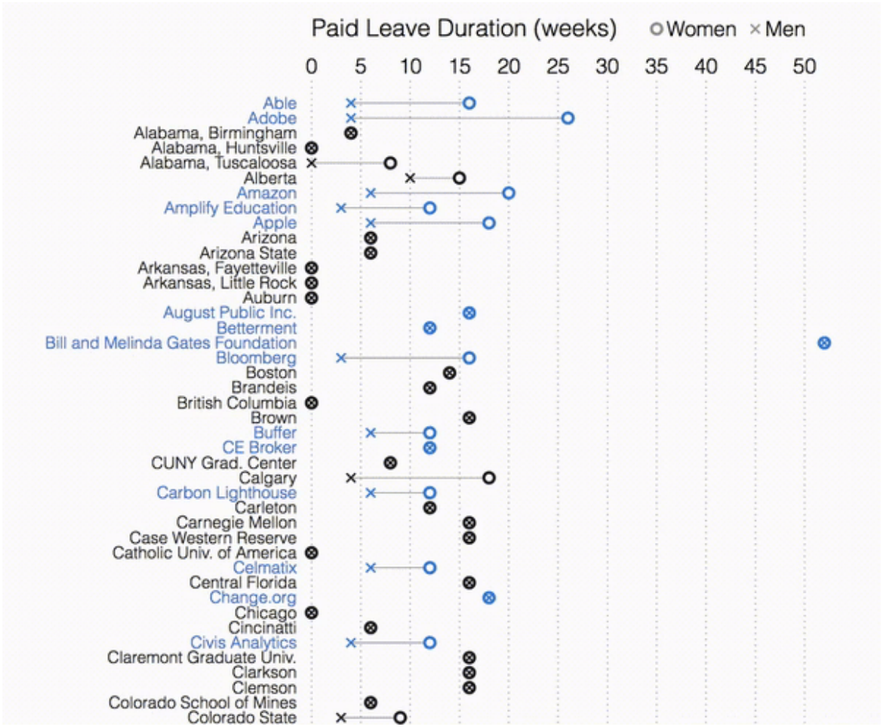
Paid Parental Leave at US and Canadian Universities

A dataset of parental leave policies. Equal contributions by Allison C. Morgan, Samuel F. Way, Mirta Galesic, Daniel B. Larremore, and Aaron Clauset.

View the Project on GitHub
aaronclauset/parental-leave

Download **ZIP File**	Download **TAR Ball**	View On **GitHub**





Add the data to compare to Tech.

aaronclauset.github.io/parental-leave/

aaronclauset.github.io/parental-leave/industry_academia.html

Link to Movie [Twitter @alliecmorgan]

Simulate your audience. (part 3)

	coef	se	zvalue	p
(Intercept)	-4.320906e-17	0.05499178	-7.857367e-16	1.000000e+00
scale(prestige)	2.625988e-01	0.08632739	-3.041894e+00	2.350947e-03
scale(private)	1.387510e-01	0.07156172	1.938900e+00	5.251353e-02
scale(dept.size)	1.983602e-01	0.07589478	2.613622e+00	8.958812e-03
scale(jr.sr.ratio)	-9.558408e-03	0.06365449	-1.501608e-01	8.806378e-01
scale(gender.ratio)	-7.084867e-02	0.06756081	-1.048665e+00	2.943323e-01
scale(phd.gender.ratio)	9.798437e-03	0.06294453	1.556678e-01	8.762949e-01
scale(phd.per.fac)	3.848203e-01	0.07000914	5.496715e+00	3.869303e-08
scale(bs.per.fac)	-5.444218e-02	0.06352855	-8.569719e-01	3.914604e-01
scale(support.per.fac)	9.543509e-02	0.06547985	1.457473e+00	1.449859e-01
scale(exfund.per.fac)	6.856958e-02	0.07071276	9.696918e-01	3.322002e-01
scale(deptsup.per.fac)	-1.674358e-01	0.08082410	-2.071608e+00	3.830202e-02
scale(teaching.load)	-7.490268e-02	0.07069245	-1.059557e+00	2.893462e-01
scale(avg.asst.sal)	4.094305e-02	0.07299193	5.609258e-01	5.748481e-01
scale(avg.asst.sal.rel)	-1.789111e-02	0.06982525	-2.562270e-01	7.977756e-01
scale(space.per.fac)	1.098649e-01	0.07586357	1.448190e+00	1.475638e-01
scale(grad.sup)	-6.055612e-04	0.06713530	-9.020012e-03	9.928032e-01
scale(local.pop)	6.320005e-02	0.06512214	9.704849e-01	3.318048e-01
scale(parent.sup)	1.580210e-02	0.09118763	1.732921e-01	8.624218e-01
scale(parent.sup.plus)	-3.885033e-02	0.09084825	-4.276398e-01	6.689134e-01

	coef	se	zvalue	p
(Intercept)	-1.015627e-16	0.05746459	-1.767396e-15	1.000000e+00
scale(dept.size)	2.772717e-01	0.07150747	3.877520e+00	1.055266e-04
scale(jr.sr.ratio)	-6.417544e-03	0.06646028	-9.656210e-02	9.230742e-01
scale(gender.ratio)	-8.493457e-02	0.06781088	-1.252521e+00	2.103800e-01
scale(phd.gender.ratio)	-4.790964e-03	0.06482548	-7.390558e-02	9.410855e-01
scale(phd.per.fac)	3.924791e-01	0.07038592	5.576103e+00	2.459666e-08
scale(bs.per.fac)	-1.196690e-01	0.06366251	-1.879740e+00	6.014349e-02
scale(support.per.fac)	1.757839e-01	0.06512311	2.699255e+00	6.949487e-03
scale(exfund.per.fac)	1.186341e-01	0.06999055	1.695001e+00	9.007517e-02
scale(deptsup.per.fac)	-1.856224e-01	0.08165860	-2.273152e+00	2.301704e-02
scale(teaching.load)	-1.284644e-01	0.07203570	-1.783344e+00	7.453031e-02
scale(avg.asst.sal)	8.159199e-02	0.07472268	1.091931e+00	2.748636e-01
scale(avg.asst.sal.rel)	-1.627540e-02	0.07217724	-2.254922e-01	8.215964e-01
scale(space.per.fac)	1.329687e-01	0.07762836	1.712888e+00	8.673307e-02
scale(grad.sup)	3.940688e-02	0.06831598	5.768325e-01	5.640526e-01
scale(local.pop)	8.609404e-02	0.06596881	1.305072e+00	1.918684e-01
scale(parent.sup)	6.581023e-02	0.09416931	6.988501e-01	4.846457e-01
scale(parent.sup.plus)	-3.837969e-02	0.09509670	-4.035859e-01	6.865172e-01

Here are two of 16 regression tables from https://www.pnas.org/content/116/22/10729. How do scientists make sense of these tables? How do we do some legwork for them?

Simulate your audience. (part 3)

	coef	se	zvalue	p
(Intercept)	-4.320906e-17	0.05499178	-7.857367e-16	1.000000e+00
scale(prestige)	2.625988e-01	0.08632739	-3.041894e+00	2.350947e-03
scale(private)	1.387510e-01	0.07156172	1.938900e+00	5.251353e-02
scale(dept.size)	1.983602e-01	0.07589478	2.613622e+00	8.958812e-03
scale(jr.sr.ratio)	-9.558408e-03	0.06365449	-1.501608e-01	8.806378e-01
scale(gender.ratio)	-7.084867e-02	0.06756081	-1.048665e+00	2.943323e-01
scale(phd.gender.ratio)	9.798437e-03	0.06294453	1.556678e-01	8.762949e-01
scale(phd.per.fac)	3.848203e-01	0.07000914	5.496715e+00	3.869303e-08
scale(bs.per.fac)	-5.444218e-02	0.06352855	-8.569719e-01	3.914604e-01
scale(support.per.fac)	9.543509e-02	0.06547985	1.457473e+00	1.449859e-01
scale(exfund.per.fac)	6.856958e-02	0.07071276	9.696918e-01	3.322002e-01
scale(deptsup.per.fac)	-1.674358e-01	0.08082410	-2.071608e+00	3.830202e-02
scale(teaching.load)	-7.490268e-02	0.07069245	-1.059557e+00	2.893462e-01
scale(avg.asst.sal)	4.094305e-02	0.07299193	5.609258e-01	5.748481e-01
scale(avg.asst.sal.rel)	-1.789111e-02	0.06982525	-2.562270e-01	7.977756e-01
scale(space.per.fac)	1.098649e-01	0.07586357	1.448190e+00	1.475638e-01
scale(grad.sup)	-6.055612e-04	0.06713530	-9.020012e-03	9.928032e-01
scale(local.pop)	6.320005e-02	0.06512214	9.704849e-01	3.318048e-01
scale(parent.sup)	1.580210e-02	0.09118763	1.732921e-01	8.624218e-01
scale(parent.sup.plus)	-3.885033e-02	0.09084825	-4.276398e-01	6.689134e-01

	coef	se	zvalue	p
(Intercept)	-1.015627e-16	0.05746459	-1.767396e-15	1.000000e+00
scale(dept.size)	2.772717e-01	0.07150747	3.877520e+00	1.055266e-04
scale(jr.sr.ratio)	-6.417544e-03	0.06646028	-9.656210e-02	9.230742e-01
scale(gender.ratio)	-8.493457e-02	0.06781088	-1.252521e+00	2.103800e-01
scale(phd.gender.ratio)	-4.790964e-03	0.06482548	-7.390558e-02	9.410855e-01
scale(phd.per.fac)	3.924791e-01	0.07038592	5.576103e+00	2.459666e-08
scale(bs.per.fac)	-1.196690e-01	0.06366251	-1.879740e+00	6.014349e-02
scale(support.per.fac)	1.757839e-01	0.06512311	2.699255e+00	6.949487e-03
scale(exfund.per.fac)	1.186341e-01	0.06999055	1.695001e+00	9.007517e-02
scale(deptsup.per.fac)	-1.856224e-01	0.08165860	-2.273152e+00	2.301704e-02
scale(teaching.load)	-1.284644e-01	0.07203570	-1.783344e+00	7.453031e-02
scale(avg.asst.sal)	8.159199e-02	0.07472268	1.091931e+00	2.748636e-01
scale(avg.asst.sal.rel)	-1.627540e-02	0.07217724	-2.254922e-01	8.215964e-01
scale(space.per.fac)	1.329687e-01	0.07762836	1.712888e+00	8.673307e-02
scale(grad.sup)	3.940688e-02	0.06831598	5.768325e-01	5.640526e-01
scale(local.pop)	8.609404e-02	0.06596881	1.305072e+00	1.918684e-01
scale(parent.sup)	6.581023e-02	0.09416931	6.988501e-01	4.846457e-01
scale(parent.sup.plus)	-3.837969e-02	0.09509670	-4.035859e-01	6.865172e-01

Which relationships are significant?
What's the strength of these relationships?

Here are two of 16 regression tables from https://www.pnas.org/content/116/22/10729. How do scientists make sense of these tables? How do we do some legwork for them?

Simulate your audience. (part 3)

	coef	se	zvalue	p
(Intercept)	-4.320906e-17	0.05499178	-7.857367e-16	1.000000e+00
scale(prestige)	2.625988e-01	0.08632739	-3.041894e+00	2.350947e-03
scale(private)	1.387510e-01	0.07156172	1.938900e+00	5.251353e-02
scale(dept.size)	1.983602e-01	0.07589478	2.613622e+00	8.958812e-03
scale(jr.sr.ratio)	-9.558408e-03	0.06365449	-1.501608e-01	8.806378e-01
scale(gender.ratio)	-7.084867e-02	0.06756081	-1.048665e+00	2.943323e-01
scale(phd.gender.ratio)	9.798437e-03	0.06294453	1.556678e-01	8.762949e-01
scale(phd.per.fac)	3.848203e-01	0.07000914	5.496715e+00	3.869303e-08
scale(bs.per.fac)	-5.444218e-02	0.06352855	-8.569719e-01	3.914604e-01
scale(support.per.fac)	9.543509e-02	0.06547985	1.457473e+00	1.449859e-01
scale(exfund.per.fac)	6.856958e-02	0.07071276	9.696918e-01	3.322002e-01
scale(deptsup.per.fac)	-1.674358e-01	0.08082410	-2.071608e+00	3.830202e-02
scale(teaching.load)	-7.490268e-02	0.07069245	-1.059557e+00	2.893462e-01
scale(avg.asst.sal)	4.094305e-02	0.07299193	5.609258e-01	5.748481e-01
scale(avg.asst.sal.rel)	-1.789111e-02	0.06982525	-2.562270e-01	7.977756e-01
scale(space.per.fac)	1.098649e-01	0.07586357	1.448190e+00	1.475638e-01
scale(grad.sup)	-6.055612e-04	0.06713530	-9.020012e-03	9.928032e-01
scale(local.pop)	6.320005e-02	0.06512214	9.704849e-01	3.318048e-01
scale(parent.sup)	1.580210e-02	0.09118763	1.732921e-01	8.624218e-01
scale(parent.sup.plus)	-3.885033e-02	0.09084825	-4.276398e-01	6.689134e-01

	coef	se	zvalue	p
(Intercept)	-1.015627e-16	0.05746459	-1.767396e-15	1.000000e+00
scale(dept.size)	2.772717e-01	0.07150747	3.877520e+00	1.055266e-04
scale(jr.sr.ratio)	-6.417544e-03	0.06646028	-9.656210e-02	9.230742e-01
scale(gender.ratio)	-8.493457e-02	0.06781088	-1.252521e+00	2.103800e-01
scale(phd.gender.ratio)	-4.790964e-03	0.06482548	-7.390558e-02	9.410855e-01
scale(phd.per.fac)	3.924791e-01	0.07038592	5.576103e+00	2.459666e-08
scale(bs.per.fac)	-1.196690e-01	0.06366251	-1.879740e+00	6.014349e-02
scale(support.per.fac)	1.757839e-01	0.06512311	2.699255e+00	6.949487e-03
scale(exfund.per.fac)	1.186341e-01	0.06999055	1.695001e+00	9.007517e-02
scale(deptsup.per.fac)	-1.856224e-01	0.08165860	-2.273152e+00	2.301704e-02
scale(teaching.load)	-1.284644e-01	0.07203570	-1.783344e+00	7.453031e-02
scale(avg.asst.sal)	8.159199e-02	0.07472268	1.091931e+00	2.748636e-01
scale(avg.asst.sal.rel)	-1.627540e-02	0.07217724	-2.254922e-01	8.215964e-01
scale(space.per.fac)	1.329687e-01	0.07762836	1.712888e+00	8.673307e-02
scale(grad.sup)	3.940688e-02	0.06831598	5.768325e-01	5.640526e-01
scale(local.pop)	8.609404e-02	0.06596881	1.305072e+00	1.918684e-01
scale(parent.sup)	6.581023e-02	0.09416931	6.988501e-01	4.846457e-01
scale(parent.sup.plus)	-3.837969e-02	0.09509670	-4.035859e-01	6.865172e-01

Which relationships are significant?
What's the strength of these relationships?

Here are two of 16 regression tables from https://www.pnas.org/content/116/22/10729. How do scientists make sense of these tables? How do we do some legwork for them?

Simulate your audience. (part 3)

	coef	se	zvalue	p
(Intercept)	-4.320906e-17	0.05499178	-7.857367e-16	1.000000e+00
scale(prestige)	2.625988e-01	0.08632739	-3.041894e+00	2.350947e-03
scale(private)	1.387510e-01	0.07156172	1.938900e+00	5.251353e-02
scale(dept.size)	1.983602e-01	0.07589478	2.613622e+00	8.958812e-03
scale(jr.sr.ratio)	-9.558408e-03	0.06365449	-1.501608e-01	8.806378e-01
scale(gender.ratio)	-7.084867e-02	0.06756081	-1.048665e+00	2.943323e-01
scale(phd.gender.ratio)	9.798437e-03	0.06294453	1.556678e-01	8.762949e-01
scale(phd.per.fac)	3.848203e-01	0.07000914	5.496715e+00	3.869303e-08
scale(bs.per.fac)	-5.444218e-02	0.06352855	-8.569719e-01	3.914604e-01
scale(support.per.fac)	9.543509e-02	0.06547985	1.457473e+00	1.449859e-01
scale(exfund.per.fac)	6.856958e-02	0.07071276	9.696918e-01	3.322002e-01
scale(deptsup.per.fac)	-1.674358e-01	0.08082410	-2.071608e+00	3.830202e-02
scale(teaching.load)	-7.490268e-02	0.07069245	-1.059557e+00	2.893462e-01
scale(avg.asst.sal)	4.094305e-02	0.07299193	5.609258e-01	5.748481e-01
scale(avg.asst.sal.rel)	-1.789111e-02	0.06982525	-2.562270e-01	7.977756e-01
scale(space.per.fac)	1.098649e-01	0.07586357	1.448190e+00	1.475638e-01
scale(grad.sup)	-6.055612e-04	0.06713530	-9.020012e-03	9.928032e-01
scale(local.pop)	6.320005e-02	0.06512214	9.704849e-01	3.318048e-01
scale(parent.sup)	1.580210e-02	0.09118763	1.732921e-01	8.624218e-01
scale(parent.sup.plus)	-3.885033e-02	0.09084825	-4.276398e-01	6.689134e-01

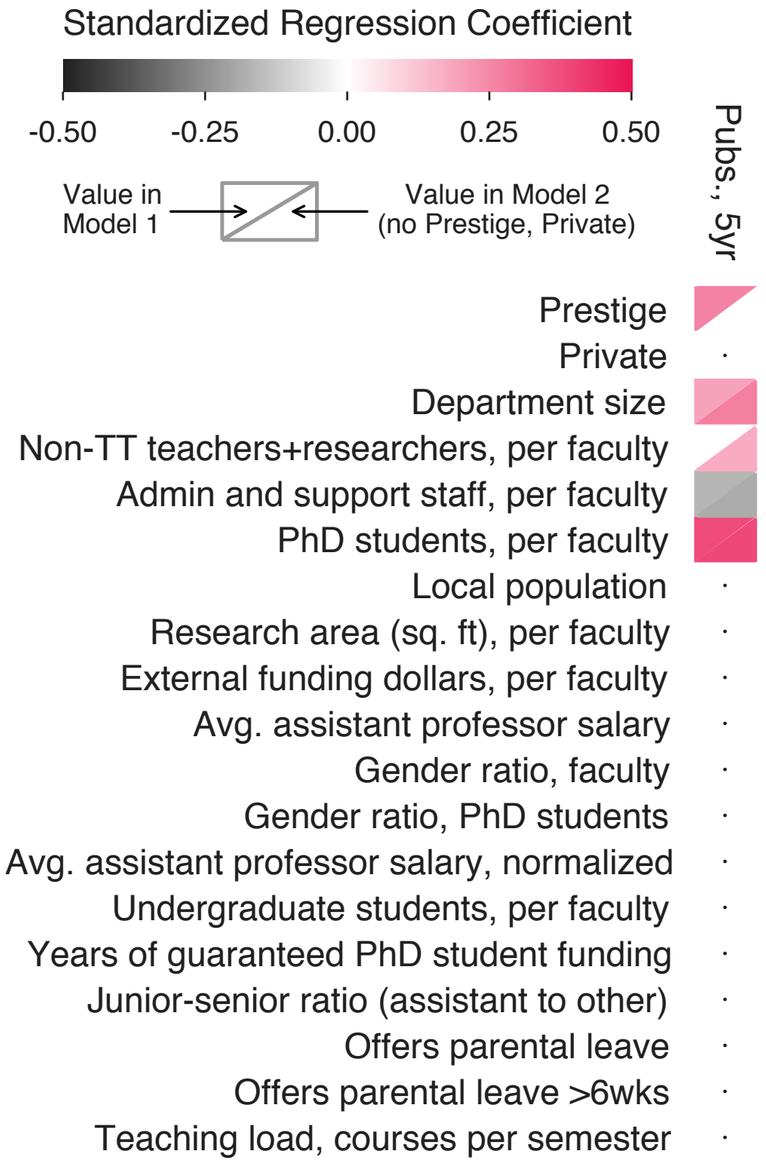
	coef	se	zvalue	p
(Intercept)	-1.015627e-16	0.05746459	-1.767396e-15	1.000000e+00
scale(dept.size)	2.772717e-01	0.07150747	3.877520e+00	1.055266e-04
scale(jr.sr.ratio)	-6.417544e-03	0.06646028	-9.656210e-02	9.230742e-01
scale(gender.ratio)	-8.493457e-02	0.06781088	-1.252521e+00	2.103800e-01
scale(phd.gender.ratio)	-4.790964e-03	0.06482548	-7.390558e-02	9.410855e-01
scale(phd.per.fac)	3.924791e-01	0.07038592	5.576103e+00	2.459666e-08
scale(bs.per.fac)	-1.196690e-01	0.06366251	-1.879740e+00	6.014349e-02
scale(support.per.fac)	1.757839e-01	0.06512311	2.699255e+00	6.949487e-03
scale(exfund.per.fac)	1.186341e-01	0.06999055	1.695001e+00	9.007517e-02
scale(deptsup.per.fac)	-1.856224e-01	0.08165860	-2.273152e+00	2.301704e-02
scale(teaching.load)	-1.284644e-01	0.07203570	-1.783344e+00	7.453031e-02
scale(avg.asst.sal)	8.159199e-02	0.07472268	1.091931e+00	2.748636e-01
scale(avg.asst.sal.rel)	-1.627540e-02	0.07217724	-2.254922e-01	8.215964e-01
scale(space.per.fac)	1.329687e-01	0.07762836	1.712888e+00	8.673307e-02
scale(grad.sup)	3.940688e-02	0.06831598	5.768325e-01	5.640526e-01
scale(local.pop)	8.609404e-02	0.06596881	1.305072e+00	1.918684e-01
scale(parent.sup)	6.581023e-02	0.09416931	6.988501e-01	4.846457e-01
scale(parent.sup.plus)	-3.837969e-02	0.09509670	-4.035859e-01	6.865172e-01

How do they compare between
Model 1 (left) and Model 2 (right)?

Here are two of 16 regression tables from https://www.pnas.org/content/116/22/10729. How do scientists make sense of these tables? How do we do some legwork for them?

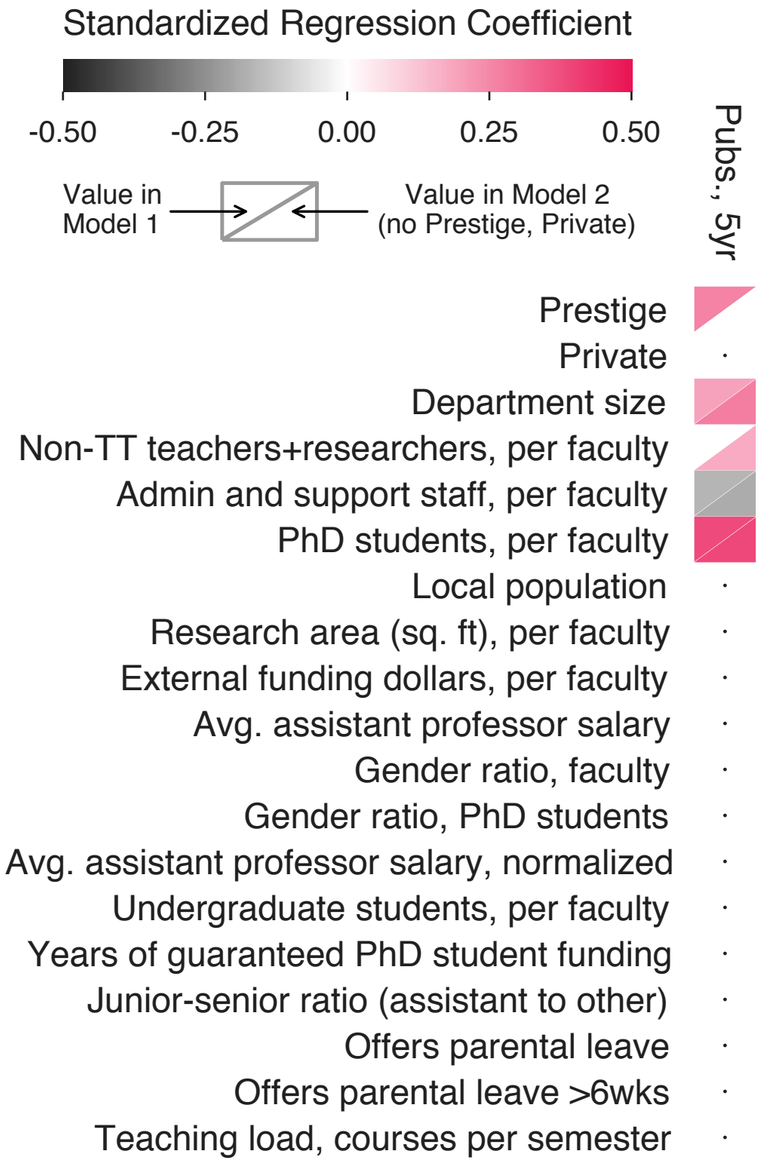
	coef	se	zvalue	p
(Intercept)	-1.015627e-16	0.05746459	-1.767396e-15	1.000000e+00

	coef	se	zvalue	p
(Intercept)	-4.320906e-17	0.05499178	-7.857367e-16	1.000000e+00
scale(prestige)	-2.625988e-01	0.08632739	-3.041894e+00	2.350947e-03
scale(private)	1.387510e-01	0.07156172	1.938900e+00	5.251353e-02
scale(dept.size)	1.983602e-01	0.07589478	2.613622e+00	8.958812e-03
scale(jr.sr.ratio)	-9.558408e-03	0.06365449	-1.501608e-01	8.806378e-01
scale(gender.ratio)	-7.084867e-02	0.06756081	-1.048665e+00	2.943323e-01
scale(phd.gender.ratio)	9.798437e-03	0.06294453	1.556678e-01	8.762949e-01
scale(phd.per.fac)	3.848203e-01	0.07000914	5.496715e+00	3.869303e-08
scale(bs.per.fac)	-5.444218e-02	0.06352855	-8.569719e-01	3.914604e-01
scale(support.per.fac)	9.543509e-02	0.06547985	1.457473e+00	1.449859e-01
scale(exfund.per.fac)	6.856958e-02	0.07071276	9.696918e-01	3.322002e-01
scale(deptsup.per.fac)	-1.674358e-01	0.08082410	-2.071608e+00	3.830202e-02
scale(teaching.load)	-7.490268e-02	0.07069245	-1.059557e+00	2.893462e-01
scale(avg.asst.sal)	4.094305e-02	0.07299193	5.609258e-01	5.748481e-01
scale(avg.asst.sal.rel)	-1.789111e-02	0.06982525	-2.562270e-01	7.977756e-01
scale(space.per.fac)	1.098649e-01	0.07586357	1.448190e+00	1.475638e-01
scale(grad.sup)	-6.055612e-04	0.06713530	-9.020012e-03	9.928032e-01
scale(local.pop)	6.320005e-02	0.06512214	9.704849e-01	3.318048e-01
scale(parent.sup)	1.580210e-02	0.09118763	1.732921e-01	8.624218e-01
scale(parent.sup.plus)	-3.885033e-02	0.09084825	-4.276398e-01	6.689134e-01



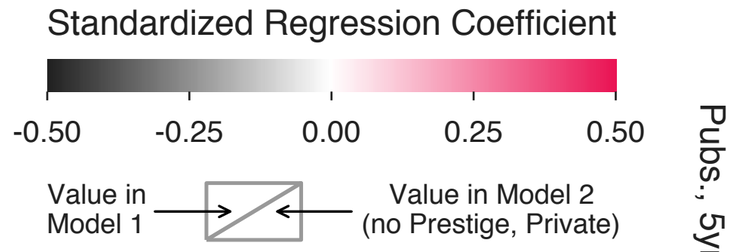
	coef	se	zvalue	p
(Intercept)	-1.015627e-16	0.05746459	-1.767396e-15	1.000000e+00

	coef	se	zvalue	p
(Intercept)	-4.320906e-17	0.05499178	-7.857367e-16	1.000000e+00
scale(prestige)	-2.625988e-01	0.08632739	-3.041894e+00	2.350947e-03
scale(private)	1.387510e-01	0.07156172	1.938900e+00	5.251353e-02
scale(dept.size)	1.983602e-01	0.07589478	2.613622e+00	8.958812e-03
scale(jr.sr.ratio)	-9.558408e-03	0.06365449	-1.501608e-01	8.806378e-01
scale(gender.ratio)	-7.084867e-02	0.06756081	-1.048665e+00	2.943323e-01
scale(phd.gender.ratio)	9.798437e-03	0.06294453	1.556678e-01	8.762949e-01
scale(phd.per.fac)	3.848203e-01	0.07000914	5.496715e+00	3.869303e-08
scale(bs.per.fac)	-5.444218e-02	0.06352855	-8.569719e-01	3.914604e-01
scale(support.per.fac)	9.543509e-02	0.06547985	1.457473e+00	1.449859e-01
scale(exfund.per.fac)	6.856958e-02	0.07071276	9.696918e-01	3.322002e-01
scale(deptsup.per.fac)	-1.674358e-01	0.08082410	-2.071608e+00	3.830202e-02
scale(teaching.load)	-7.490268e-02	0.07069245	-1.059557e+00	2.893462e-01
scale(avg.asst.sal)	4.094305e-02	0.07299193	5.609258e-01	5.748481e-01
scale(avg.asst.sal.rel)	-1.789111e-02	0.06982525	-2.562270e-01	7.977756e-01
scale(space.per.fac)	1.098649e-01	0.07586357	1.448190e+00	1.475638e-01
scale(grad.sup)	-6.055612e-04	0.06713530	-9.020012e-03	9.928032e-01
scale(local.pop)	6.320005e-02	0.06512214	9.704849e-01	3.318048e-01
scale(parent.sup)	1.580210e-02	0.09118763	1.732921e-01	8.624218e-01
scale(parent.sup.plus)	-3.885033e-02	0.09084825	-4.276398e-01	6.689134e-01

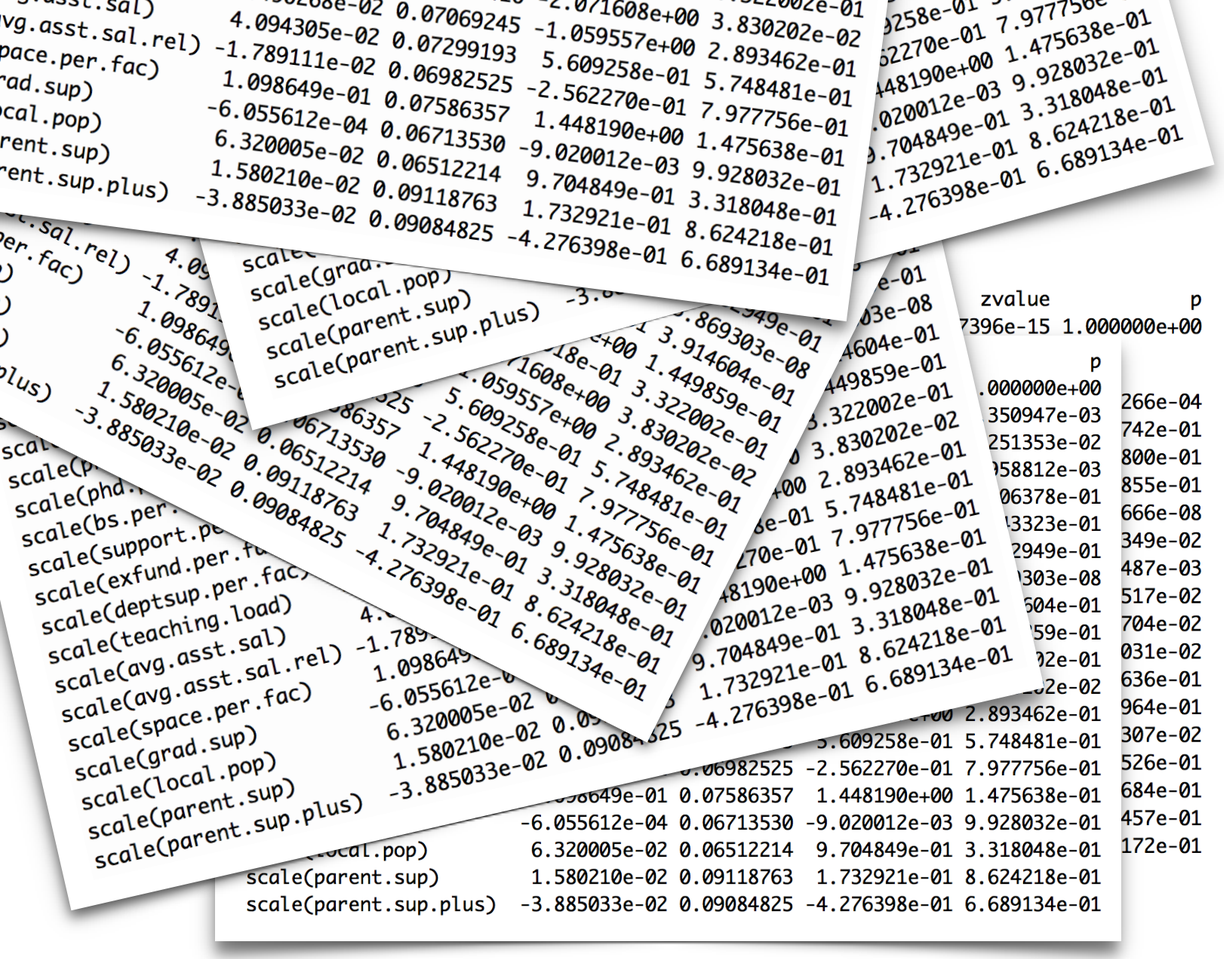


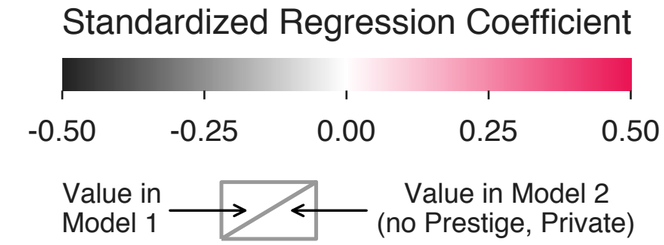
	coef	se	zvalue	p
(Intercept)	-1.015627e-16	0.05746459	-1.767396e-15	1.000000e+00

	coef	se	zvalue	p
(Intercept)	-4.320906e-17	0.05499178	-7.857367e-16	1.000000e+00
scale(prestige)	-2.625988e-01	0.08632739	-3.041894e+00	2.350947e-03
scale(private)	1.387510e-01	0.07156172	1.938900e+00	5.251353e-02
scale(dept.size)	1.983602e-01	0.07589478	2.613622e+00	8.958812e-03
scale(jr.sr.ratio)	-9.558408e-03	0.06365449	-1.501608e-01	8.806378e-01
scale(gender.ratio)	-7.084867e-02	0.06756081	-1.048665e+00	2.943323e-01
scale(phd.gender.ratio)	9.798437e-03	0.06294453	1.556678e-01	8.762949e-01
scale(phd.per.fac)	3.848203e-01	0.07000914	5.496715e+00	3.869303e-08
scale(bs.per.fac)	-5.444218e-02	0.06352855	-8.569719e-01	3.914604e-01
scale(support.per.fac)	9.543509e-02	0.06547985	1.457473e+00	1.449859e-01
scale(exfund.per.fac)	6.856958e-02	0.07071276	9.696918e-01	3.322002e-01
scale(deptsup.per.fac)	-1.674358e-01	0.08082410	-2.071608e+00	3.830202e-02
scale(teaching.load)	-7.490268e-02	0.07069245	-1.059557e+00	2.893462e-01
scale(avg.asst.sal)	4.094305e-02	0.07299193	5.609258e-01	5.748481e-01
scale(avg.asst.sal.rel)	-1.789111e-02	0.06982525	-2.562270e-01	7.977756e-01
scale(space.per.fac)	1.098649e-01	0.07586357	1.448190e+00	1.475638e-01
scale(grad.sup)	-6.055612e-04	0.06713530	-9.020012e-03	9.928032e-01
scale(local.pop)	6.320005e-02	0.06512214	9.704849e-01	3.318048e-01
scale(parent.sup)	1.580210e-02	0.09118763	1.732921e-01	8.624218e-01
scale(parent.sup.plus)	-3.885033e-02	0.09084825	-4.276398e-01	6.689134e-01



Way, Morgan, Larremore, Clauset. PNAS, 2019. https://www.pnas.org/content/116/22/10729





Way, Morgan, Larremore, Clauset. PNAS, 2019. https://www.pnas.org/content/116/22/10729

Tips for creating visualizations.

During each stage of creation:

0. Determine your goals
1. Select appropriate type of visualization
2. Build prototypes
3. Gather and address feedback

Test-drive your visualizations.

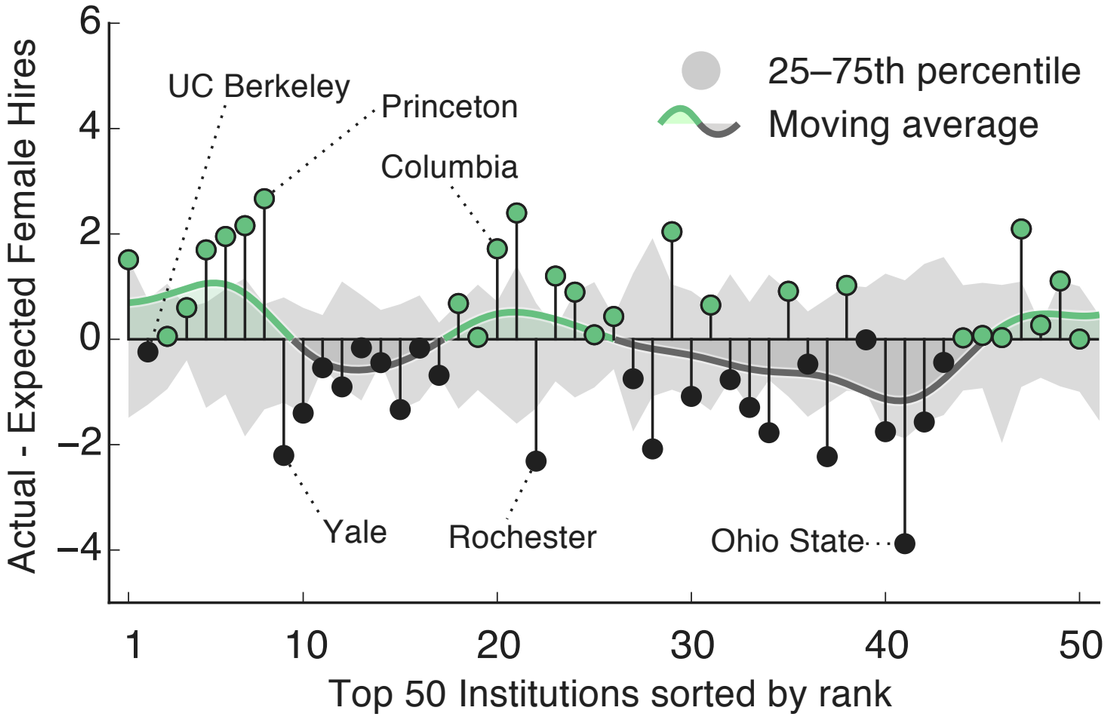
Ask for feedback on drafts.

Did your message come across?

Were they confused by anything?

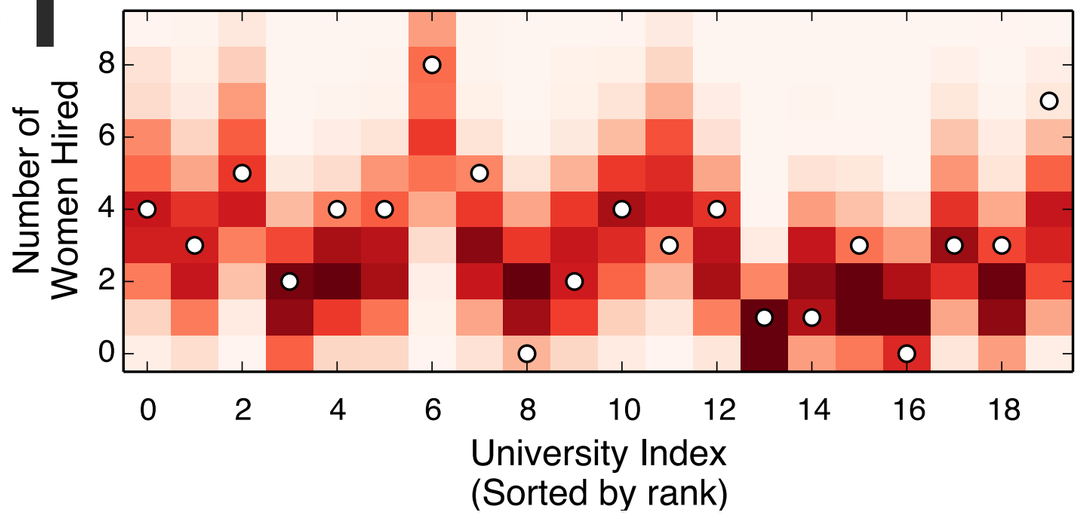
Iterate on the design.

Be willing to change it*.



* Ask for feedback from someone you trust will be honest and critical. Don't wait until you're already married to an idea/design.

#1



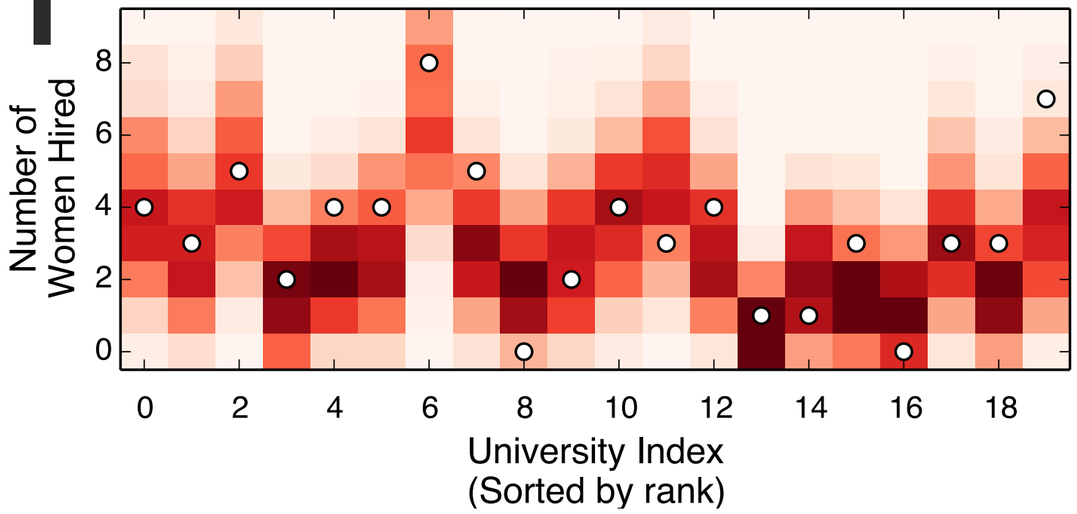
#2

#3

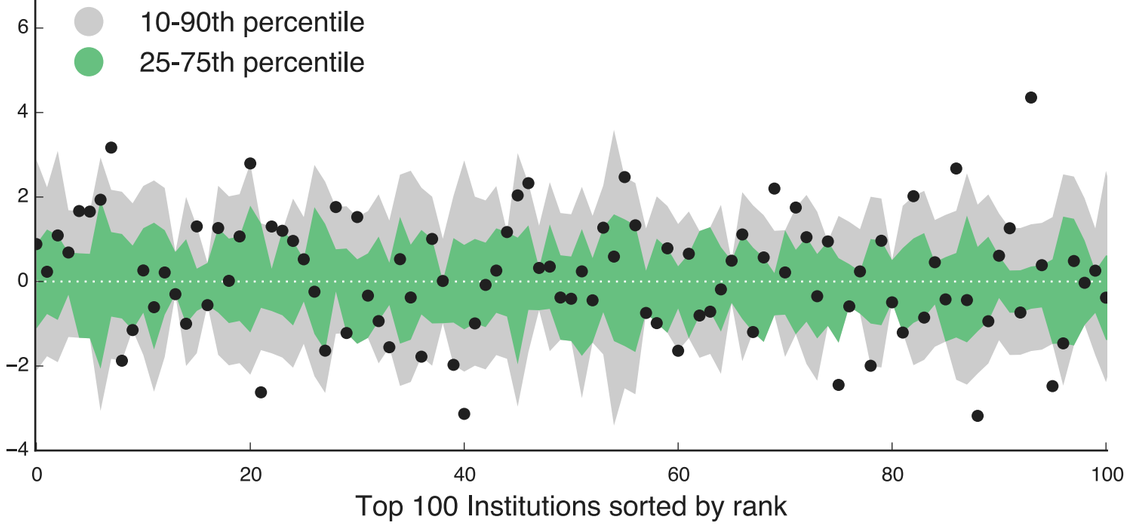
#4

How do actual hires compare to expectations under our model? Which schools are above/below expectations? Are these differences reasonable?

#1



#2

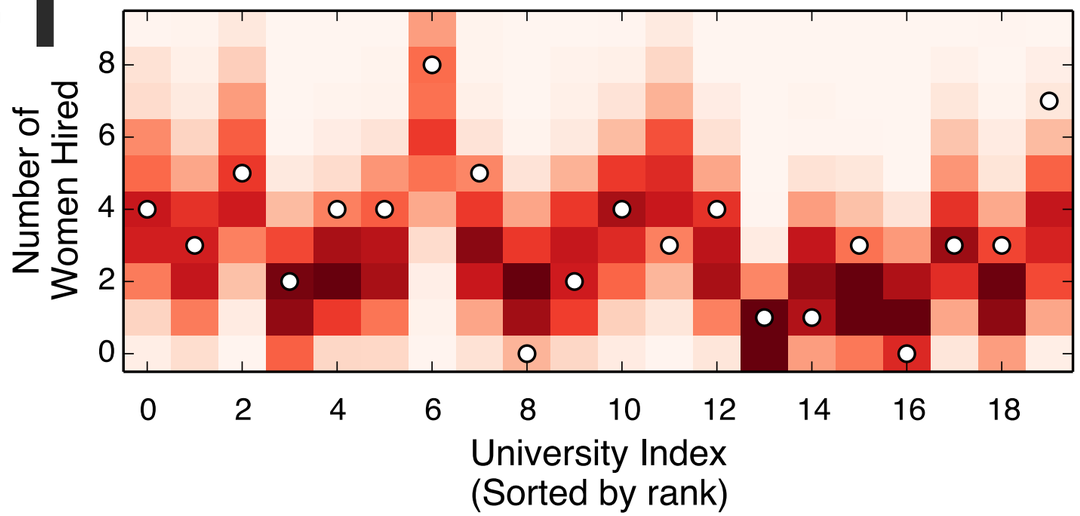


#3

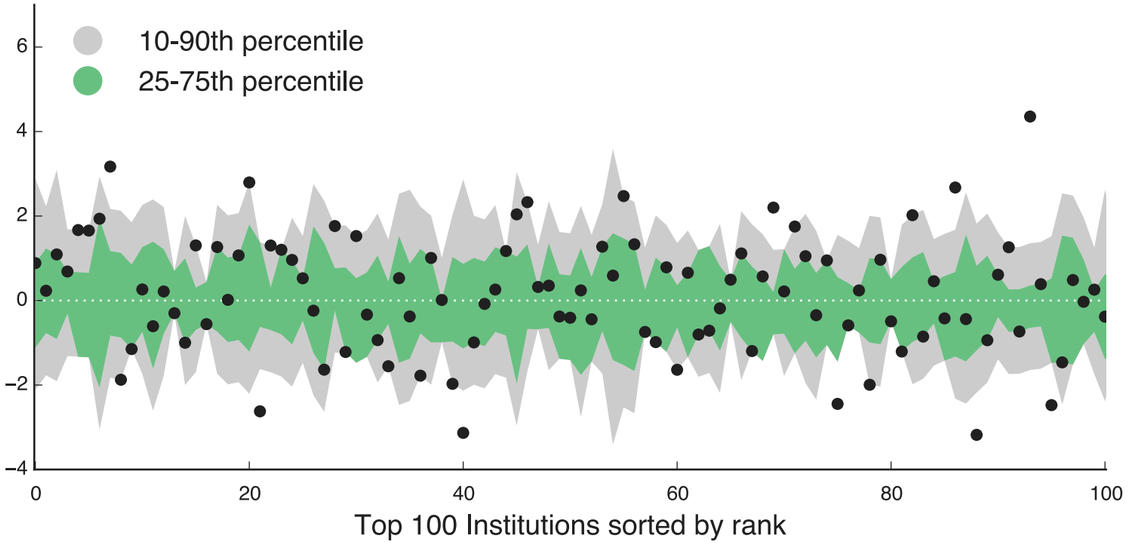
#4

How do actual hires compare to expectations under our model? Which schools are above/below expectations? Are these differences reasonable?

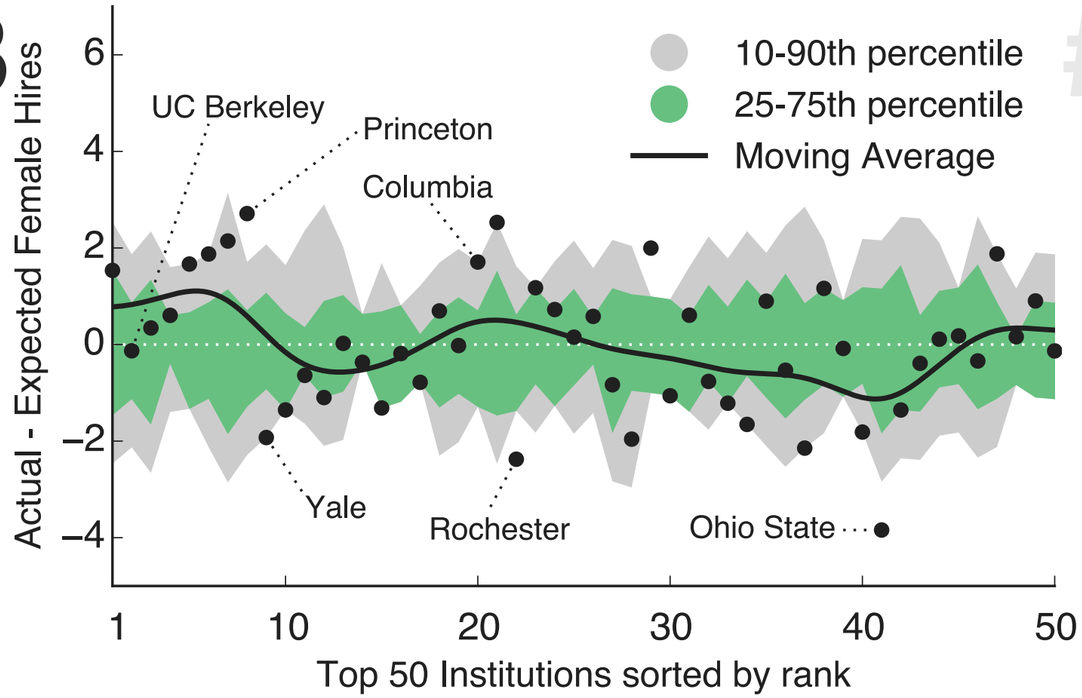
#1



#2



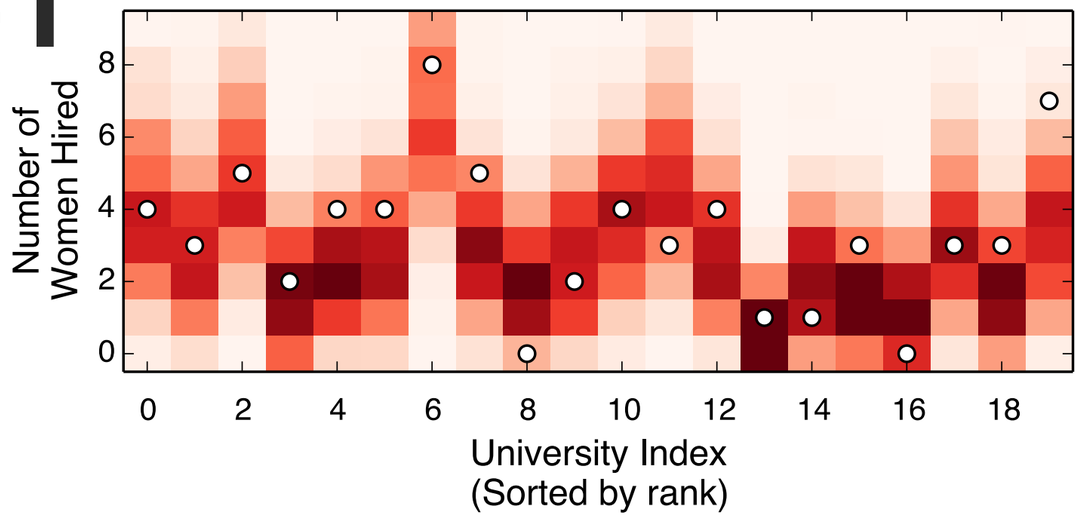
#3



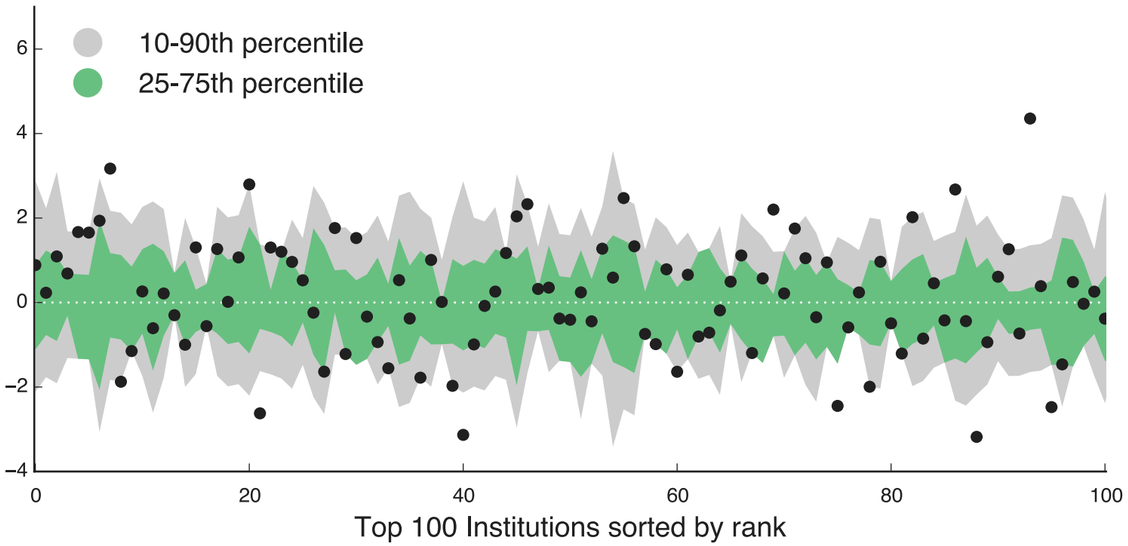
#4

How do actual hires compare to expectations under our model? Which schools are above/below expectations? Are these differences reasonable?

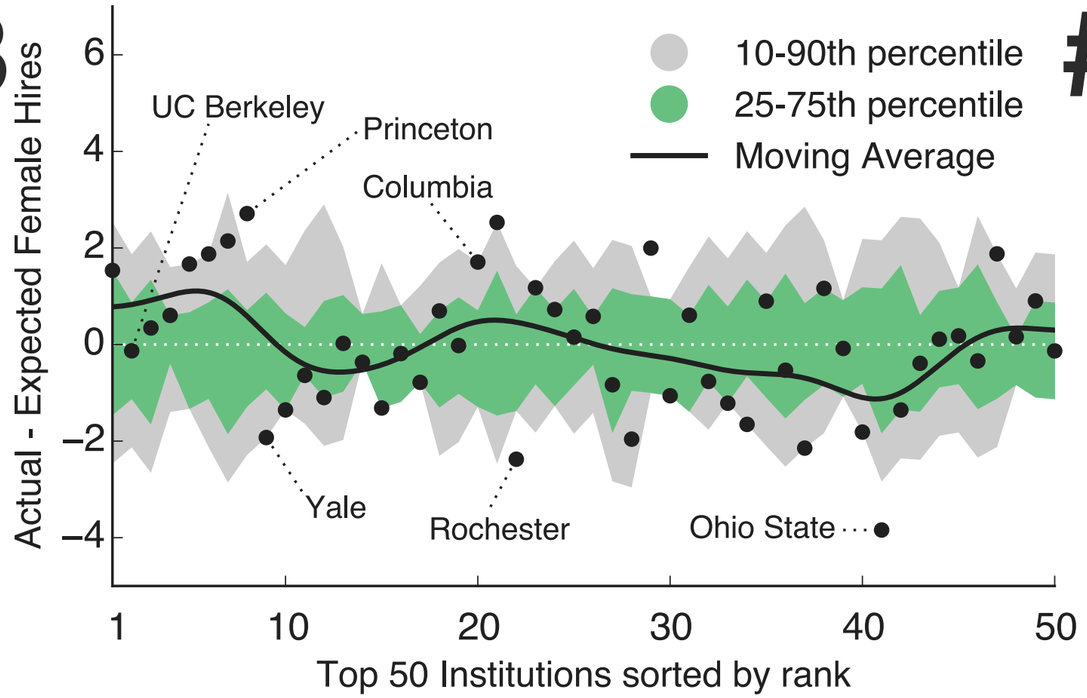
#1



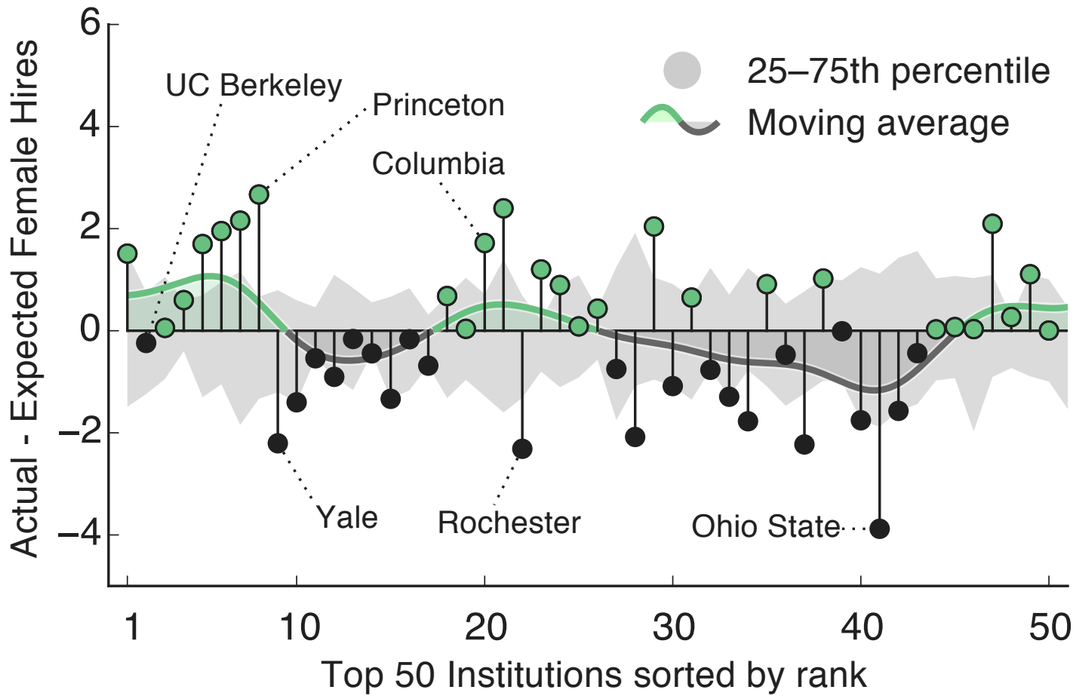
#2



#3



#4



How do actual hires compare to expectations under our model? Which schools are above/below expectations? Are these differences reasonable?

Other rules of thumb*

Don’t use hard-to-read fonts.

https://pubs.acs.org/doi/pdfplus/10.1021/acs.chemmater.6b00306

Try Helvetica instead.

https://olgabotvinnik.com/blog/how-to-set-helvetica-as-the-default-sans-serif-font-in/
Make text large enough to be read…

* There are *absolutely* situations where violating these rules makes sense. Use your discretion.

Other rules of thumb*

Don’t use hard-to-read fonts.

https://pubs.acs.org/doi/pdfplus/10.1021/acs.chemmater.6b00306

Try Helvetica instead.

https://olgabotvinnik.com/blog/how-to-set-helvetica-as-the-default-sans-serif-font-in/
Make text large enough to be read…

Don’t use true black (#000000).

https://ianstormtaylor.com/design-tip-never-use-black/

Try something less harsh (#222222).

http://uxmovement.com/content/why-you-should-never-use-pure-black-for-text-or-backgrounds/

* There are *absolutely* situations where violating these rules makes sense. Use your discretion.

Other rules of thumb*

Don’t use hard-to-read fonts.

https://pubs.acs.org/doi/pdfplus/10.1021/acs.chemmater.6b00306

Try Helvetica instead.

https://olgabotvinnik.com/blog/how-to-set-helvetica-as-the-default-sans-serif-font-in/
Make text large enough to be read…

Don’t use true black (#000000).

https://ianstormtaylor.com/design-tip-never-use-black/

Try something less harsh (#222222).

http://uxmovement.com/content/why-you-should-never-use-pure-black-for-text-or-backgrounds/

Don’t use rainbow colormaps.

http://people.cs.vt.edu/~npolys/IT/FDI/spring2011/color_07.pdf

Try something more functional.

http://people.cs.vt.edu/~npolys/IT/FDI/spring2011/color_07.pdf
https://jiffyclub.github.io/palettable/

* There are *absolutely* situations where violating these rules makes sense. Use your discretion.

Other rules of thumb*

Don’t use hard-to-read fonts.

https://pubs.acs.org/doi/pdfplus/10.1021/acs.chemmater.6b00306

Try Helvetica instead.

https://olgabotvinnik.com/blog/how-to-set-helvetica-as-the-default-sans-serif-font-in/
Make text large enough to be read…

Don’t use true black (#000000).

https://ianstormtaylor.com/design-tip-never-use-black/

Try something less harsh (#222222).

http://uxmovement.com/content/why-you-should-never-use-pure-black-for-text-or-backgrounds/

Don’t use rainbow colormaps.

http://people.cs.vt.edu/~npolys/IT/FDI/spring2011/color_07.pdf

Try something more functional.

http://people.cs.vt.edu/~npolys/IT/FDI/spring2011/color_07.pdf
https://jiffyclub.github.io/palettable/

Don’t make pie charts.

https://www.perceptualedge.com/articles/visual_business_intelligence/save_the_pies_for_dessert.pdf

Try… literally anything else.

https://github.com/d3/d3/wiki/Gallery

* There are *absolutely* situations where violating these rules makes sense. Use your discretion.

General resources

visualisingdata.com/resources/
D3 (js), matplotlib (python), seaborn (python), ggplot (R, python)
Storytelling with data:
https://www.amazon.com/Storytelling-Data-Visualization-Business-Professionals/dp/1119002257/ref=nodl_
Caveats to data visualization:
https://www.data-to-viz.com/caveats.html
Randal Olson's matplotlib tips:
http://www.randalolson.com/2014/06/28/how-to-make-beautiful-data-visualizations-in-python-with-matplotlib/

Colors

coolors.co, palettable.io (custom color palettes)
jiffyclub.github.io/palettable (colors in Python)
colororacle.org (color blind test app)
ianstormtaylor.com/design-tip-never-use-black

Science as Art

worrydream.com/ScientificCommunicationAsSequentialArt
r2d3.us/visual-intro-to-machine-learning-part-1
r-graph-gallery.com/portfolio/data-art/

Misc

Spotify dot art: tinyurl.com/y9w2p5fr
tableau.com/learn/articles/best-beautiful-data-visualization-examples
webwebpage.github.io (one-click network visualization from Python & MATLAB)

Thank you.

Original Slides: Sam Way

http://samfway.com
samfway@gmail.edu
@samfway

Modifications: Dan Larremore

http://danlarremore.com
daniel.larremore@colorado.edu
@danlarremore

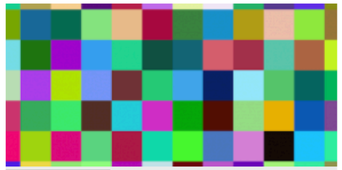
Not Secure | www.visualisingdata.com/resources/

visualising data

HOME BLOG **RESOURCES** TRAINING BOOK ABOUT

18 OCT BEST OF THE VISUALISATION WEB... JULY 2018 >>

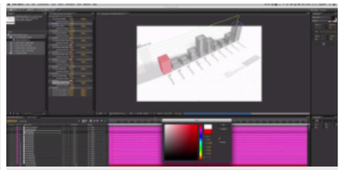
DATA HANDLING | CHARTING | PROGRAMMING | MULTIVARIATE | MAPPING | WEB-BASED | SPECIALIST | COLOUR


0 TO 255


ABBYY

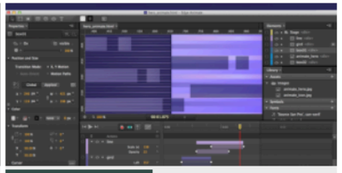
ABLE2EXTRACT


ADIOMA

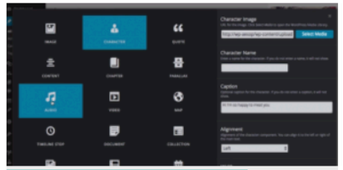

ADOBE AFTER EFFECTS


ADOBE ANIMATE


ADOBE COLOR

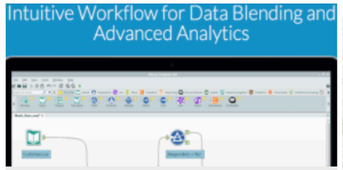

ADOBE EDGE


ADOBE ILLUSTRATOR

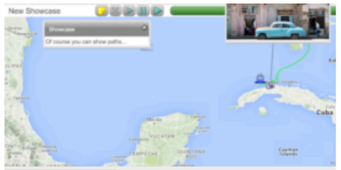

AESOP STORY ENGINE


AFFINITY DESIGNER


AI2HTML



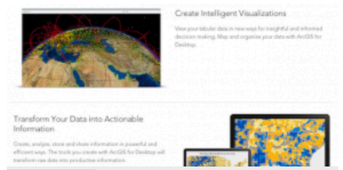
ALTERYX


ANIMAPS

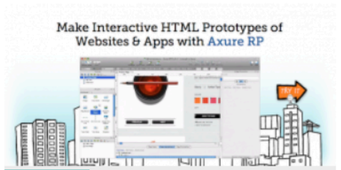

ANYCHART


APPS FOR EXCEL


ARBOR.JS


ARCGIS

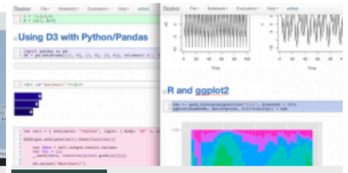

AUTODRAW



AXURE


BALSAMIQ


BATCHGEO



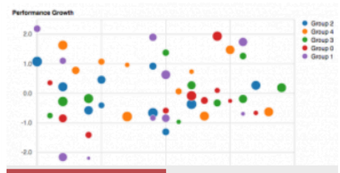
BEAKER


BERTIFIER


BIGML

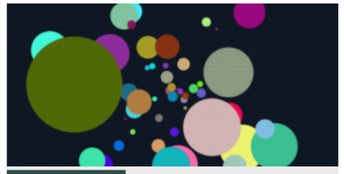

BIGPICTURE


BLENDER



BLOCKSPRING


BLUESHIFT


BONSAI.IS

Bonus: Critique some visualizations.

What do you like, or not like?

Context: Recent publication in EPJ Data Science on predicting AirBnB's penetration from market attributes.

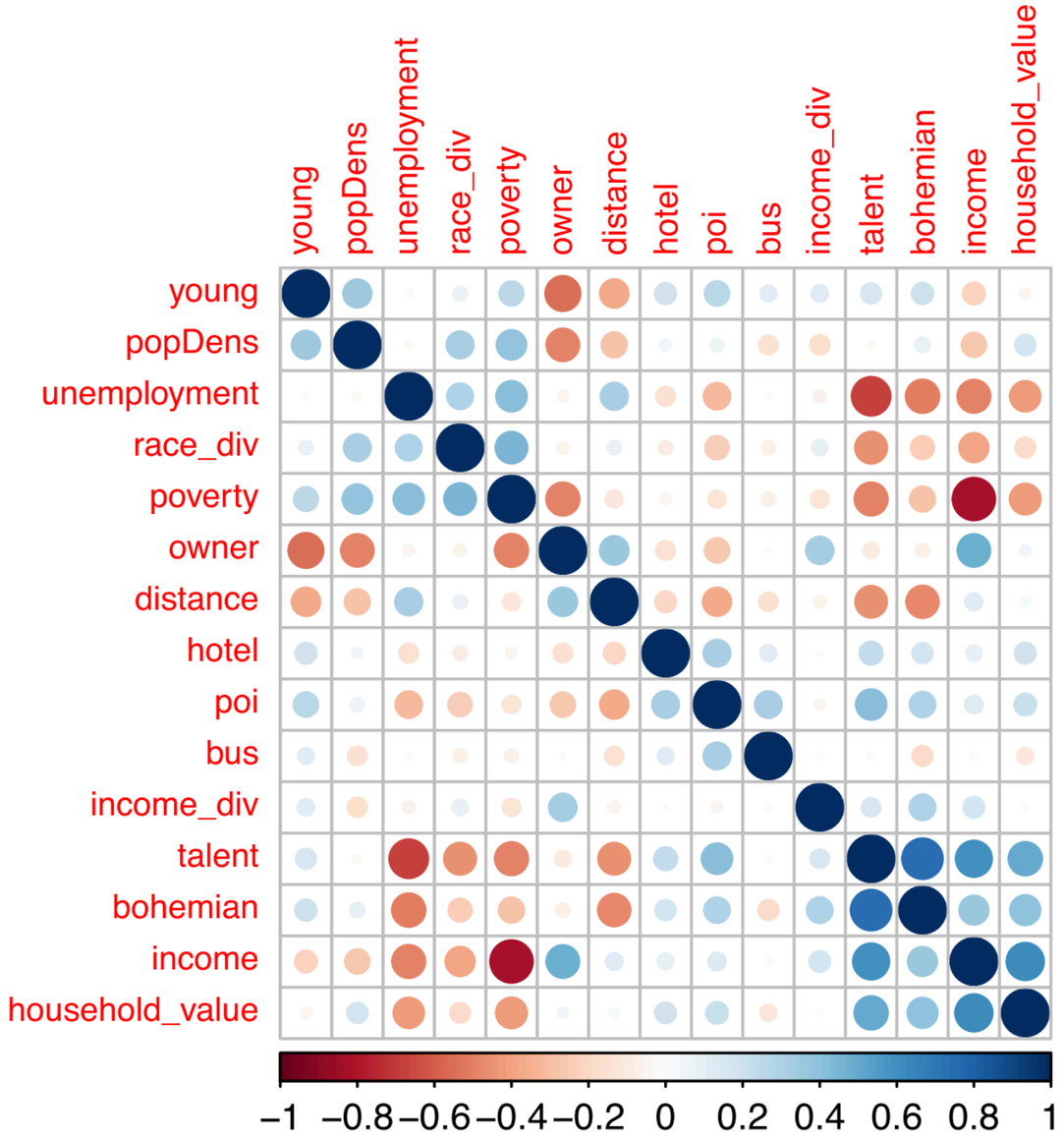


Figure 3 Pairwise Spearman correlation between explanatory variables for the eight considered U.S. cities

https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-018-0156-6

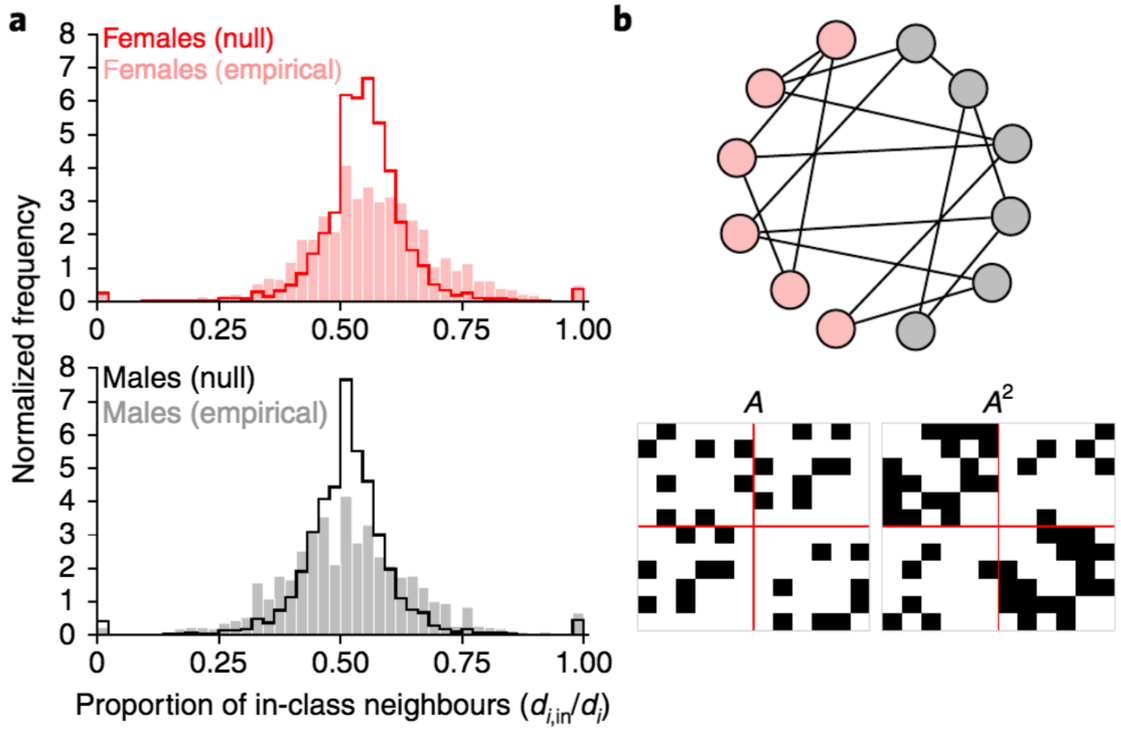


Fig. 1 | Overdispersion in attribute preferences. a, Amherst College Facebook network. Empirical distribution (filled bars) of in-class preferences for females and males compared with a null distribution (solid lines) based on preferences with binomial variation (see Methods). We observe overdispersion for females and males as the observed empirical variance is greater than under the null. **b**, A sample network without homophily or heterophily, but with monophily. We also show the link structure of the adjacency matrix (A) and the two-hop adjacency matrix (A^2). The matrices are grouped by attribute class where the red line separates classes. Monophily results in a block structure in the ties between friends-of-friends, but not between friends.

Context: Recent publication in Nature Human Behavior on monophily in social networks.

https://www.nature.com/articles/s41562-018-0321-8

Context: Recent publication in PNAS using peoples' language and usage statistics on Facebook to predict depression.

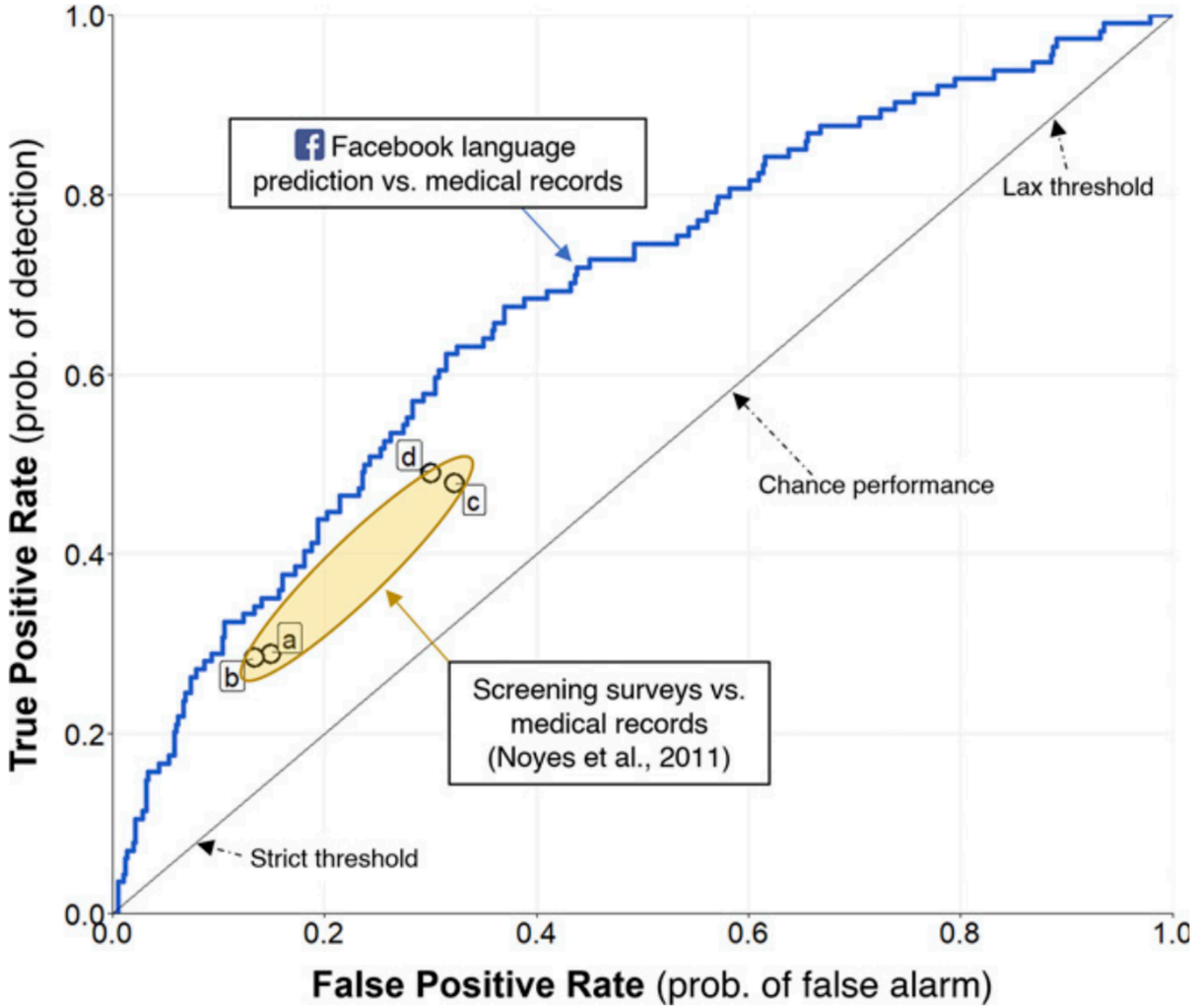
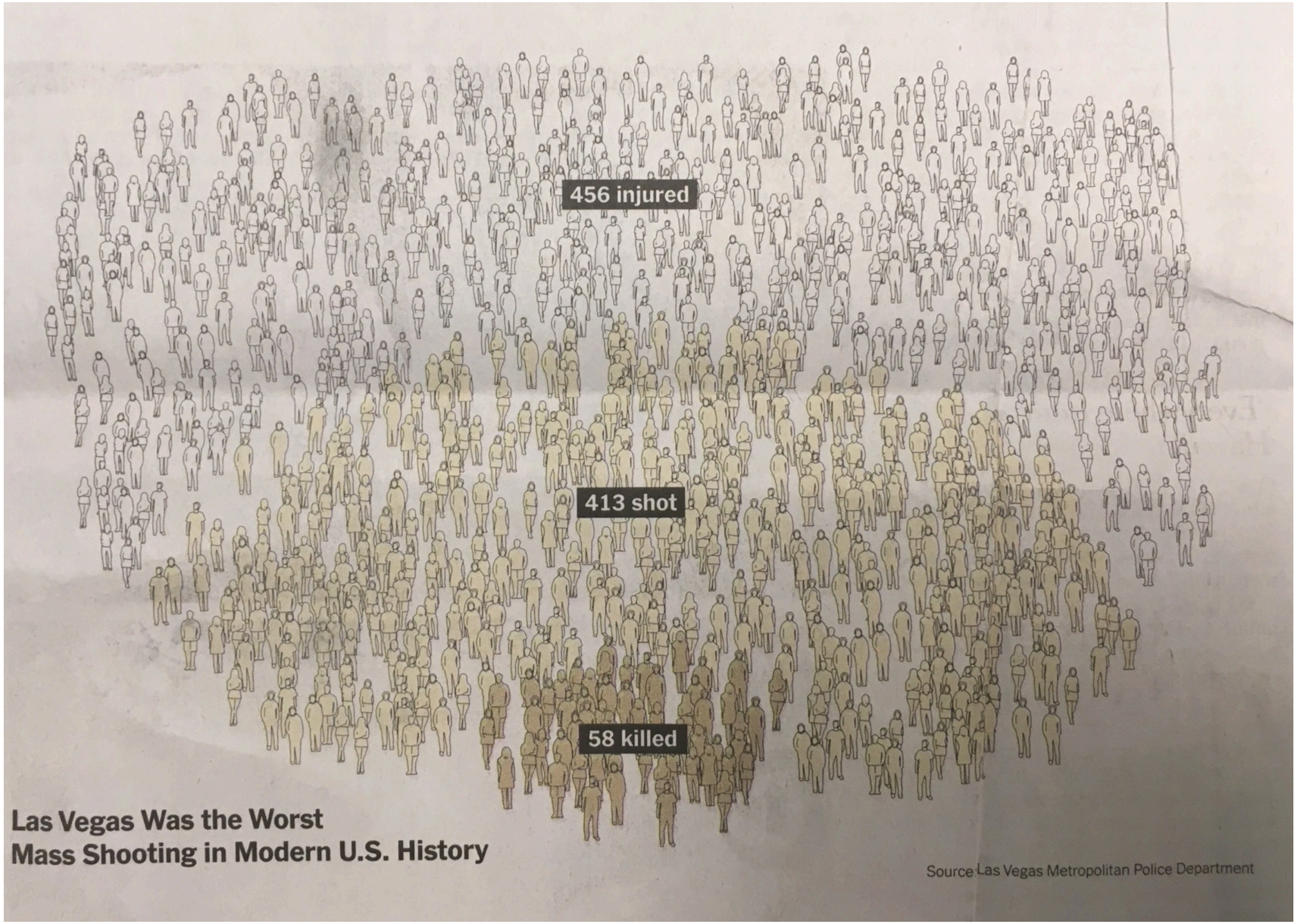


Fig. 2. ROC curve for a Facebook activity-based prediction model (all predictors combined; blue), and points as combinations of true and false positive rates reported by Noyes et al. (17) for different combinations of depression surveys (*a* and *b*, 9-item Mini-International Neuropsychiatric Interview–Major Depressive Episode Module; *c* and *d*, 15-item Geriatric Depression Scale with a cutoff >6) and time windows in Medicare claims data (*a* and *c*, within 6 mo before and after survey; *b* and *d*, within 12 mo).

http://www.pnas.org/content/early/2018/10/09/1802331115



Context: Recent infographic about the Las Vegas shooting.

https://dataanddragons.wordpress.com/2018/10/08/time-and-space/