Harnessing the power of data visualization: from insight and storytelling to AI explanation and data art

Fernanda Viégas

Gordon McKay Professor of Computer Science, Harvard University

Sally Starling Seaver Professor, Harvard Radcliffe Institute

Principal Scientist, Google

A little about me

From Brazil

Not the usual career path…

Tried several majors, wasn't happy

Scholarship to the US ⟶ "undecided" major

A little about me

From Brazil

Not the usual career path…

Tried several majors, wasn't happy

Scholarship to the US ⟶ "undecided" major

University of Kansas

Graphic design

Art history

MIT Media Lab

Data visualization = graphic design + computation

2003: started collaborating with Martin Wattenberg

Web Seer

will Brazil

- will brazil **win the world cup**
- will brazil **rest players**
- will brazil **win**
- will brazil **be in fifa 23**
- will brazil **play argentina**
- will brazil **become a superpower**
- will brazil **lift covid restrictions**
- will brazil**ian jasmine survive winter**
- will brazil **ever be a developed country**
- will brazil**ian blowout straighten hair**

Google Search I'm Feeling Lucky

Report inappropriate predictions

Web seer

http://hint.fm/seer

Visualization…

Not just for numbers
Not just for individuals
Not just for experts

But first...

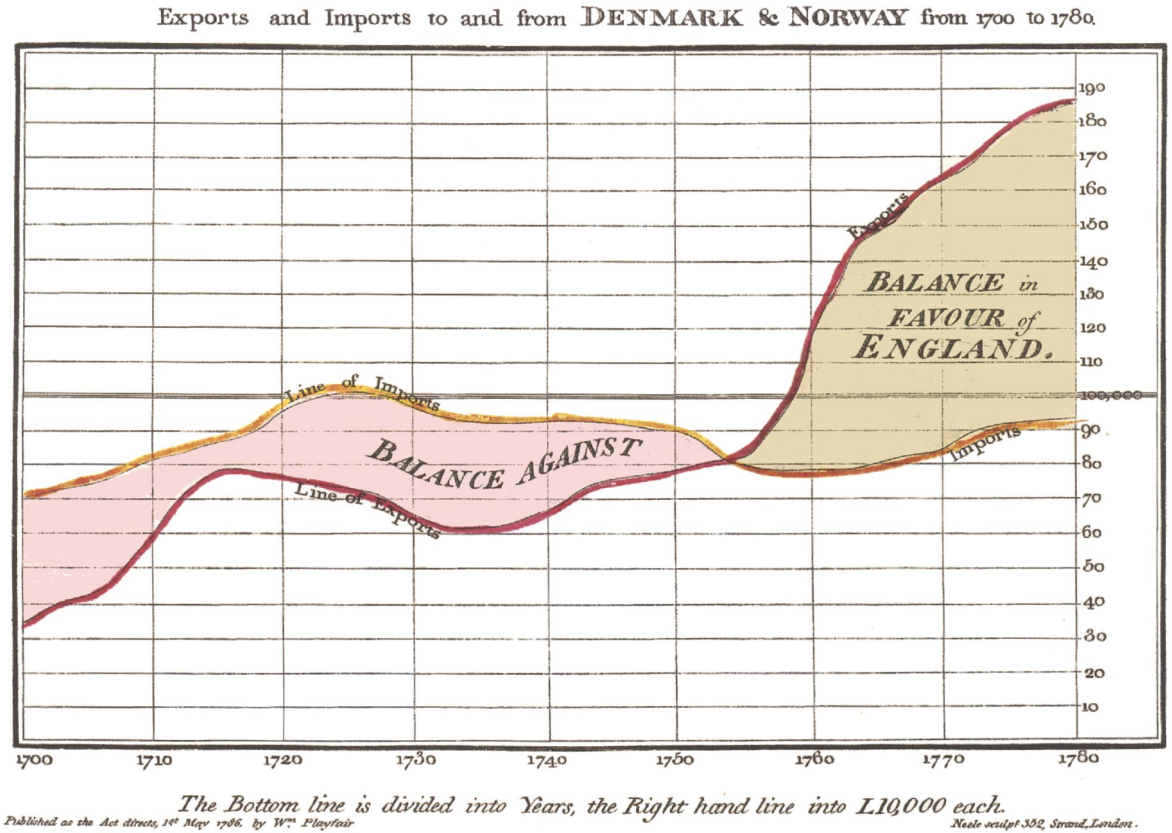
We've been doing this for a while

Ancient visualizations

Babylonian map of the Mesopotamian world
700-500 BCE, probably from southern Iraq

Ptolemy's world map, 2nd century
First use of longitudinal and latitudinal lines on a map!

William Playfair (1786)

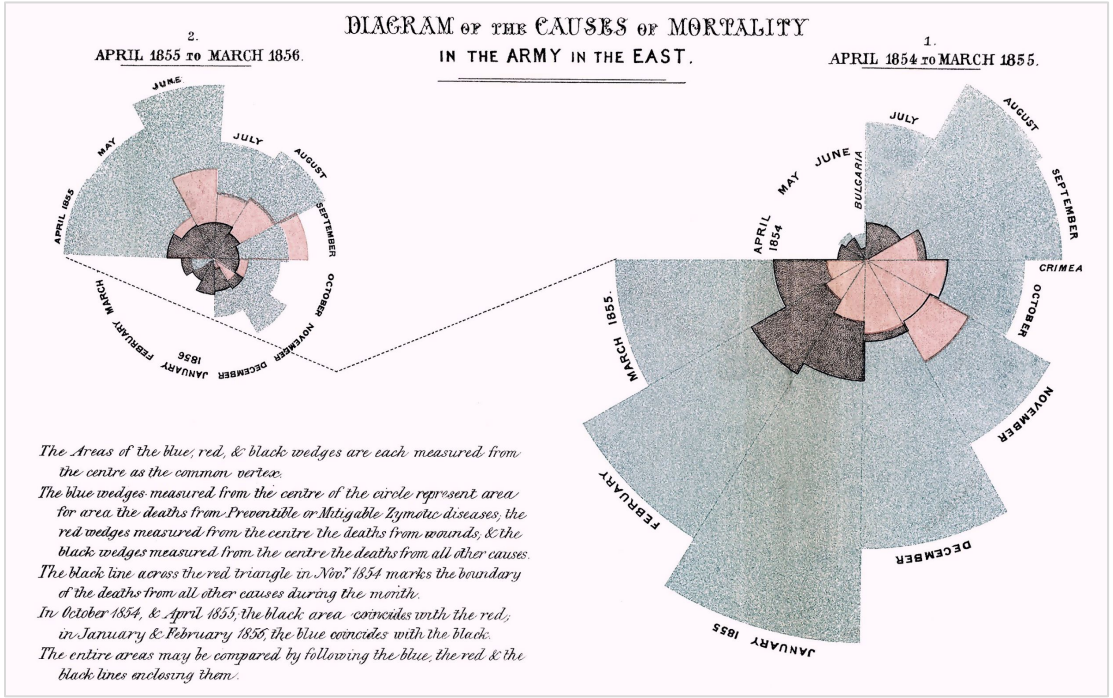


Scottish engineer and political economist

Line, bar, pie charts were all invented by the same person!

Aside from revolutionizing graphics, Playfair was an economist, engineer, and even a secret agent.

Florence Nightingale (1858)



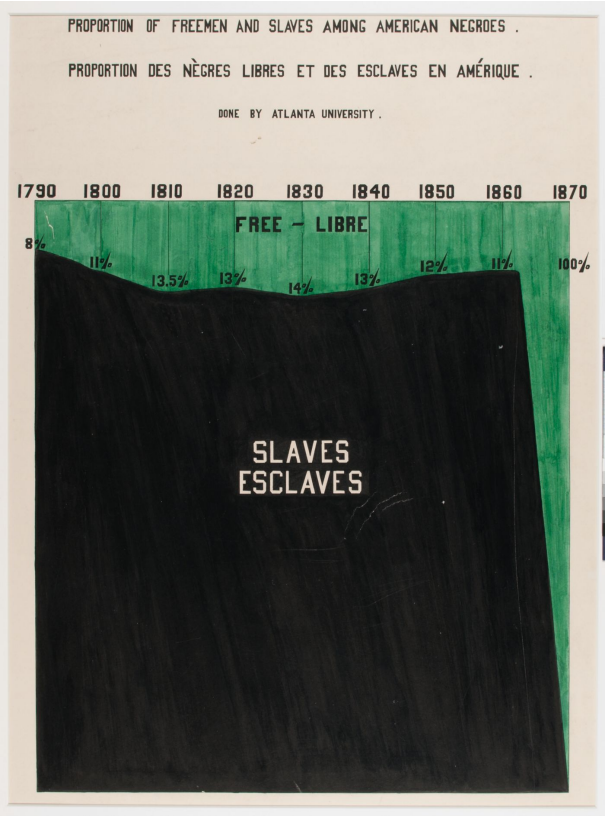
British nurse and statistician

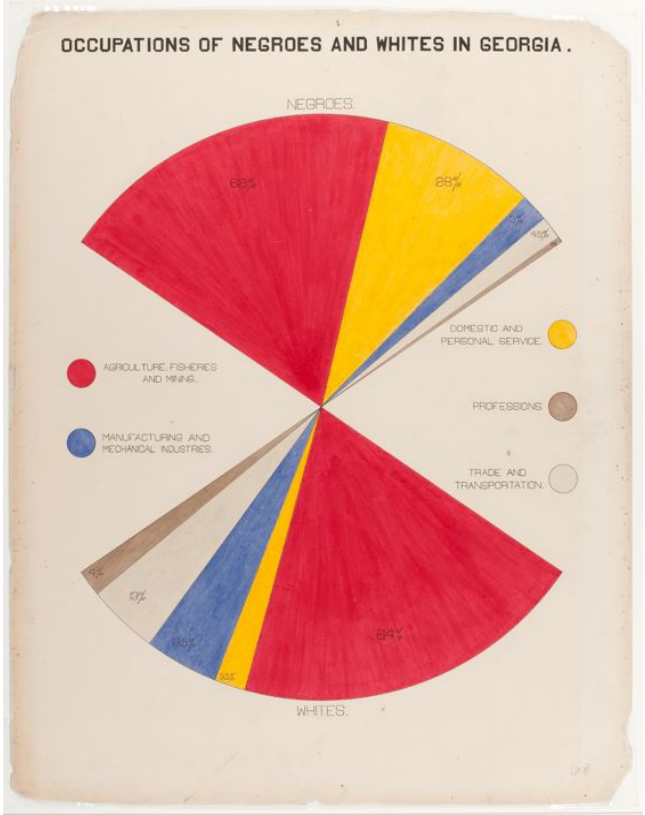
Arguably the most effective visualization ever!

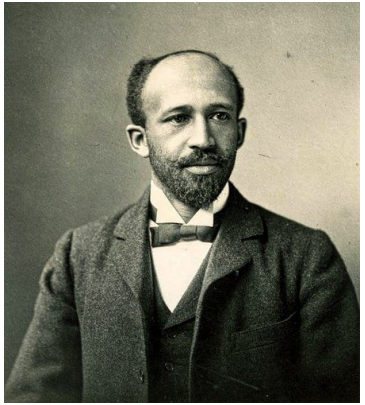
These charts convinced the British Parliament to enact sanitation reforms in hospitals, saving millions of lives.

This particular visualization technique would be frowned on today.
Lesson: technique is less important than having the right data and right message.

W. E. B. Du Bois (1900)







American sociologist,
historian, civil rights activist

Paris World Fair of 1900
"The American Negro" exhibit

A team of Black sociologists creates about 60 visualizations that explain US institutional racism to the world

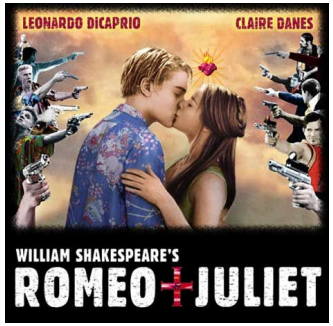
Not just for numbers

Word Tree

Paper: The Word Tree: an Interactive Visual Concordance, FV+MW, IEEE InfoVis.


LEONARDO DICAPRIO
CLAIRE DANES
WILLIAM SHAKESPEARE'S
ROMEO + JULIET

+ Search term: “if love”

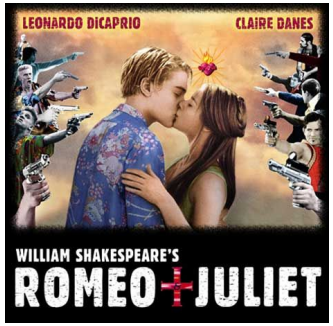


+ Search term: “if love”

if love be rough with you , be rough with love .

if love be blind , love cannot hit the mark .

if love be blind , it best agrees with night .

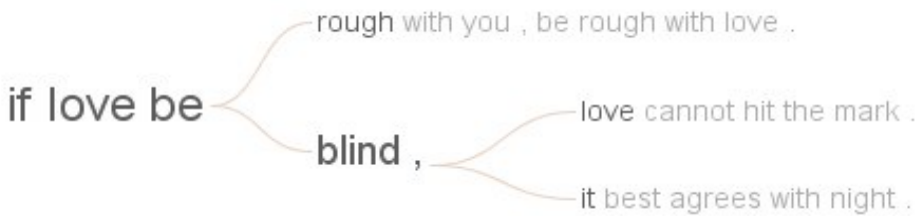


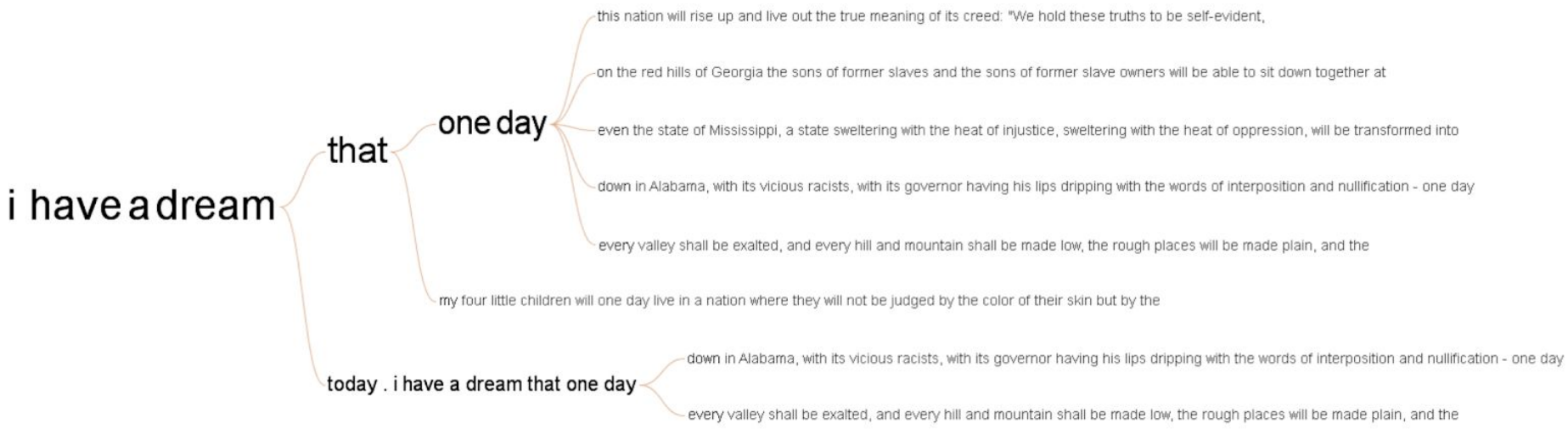
+ Search term: “if love”

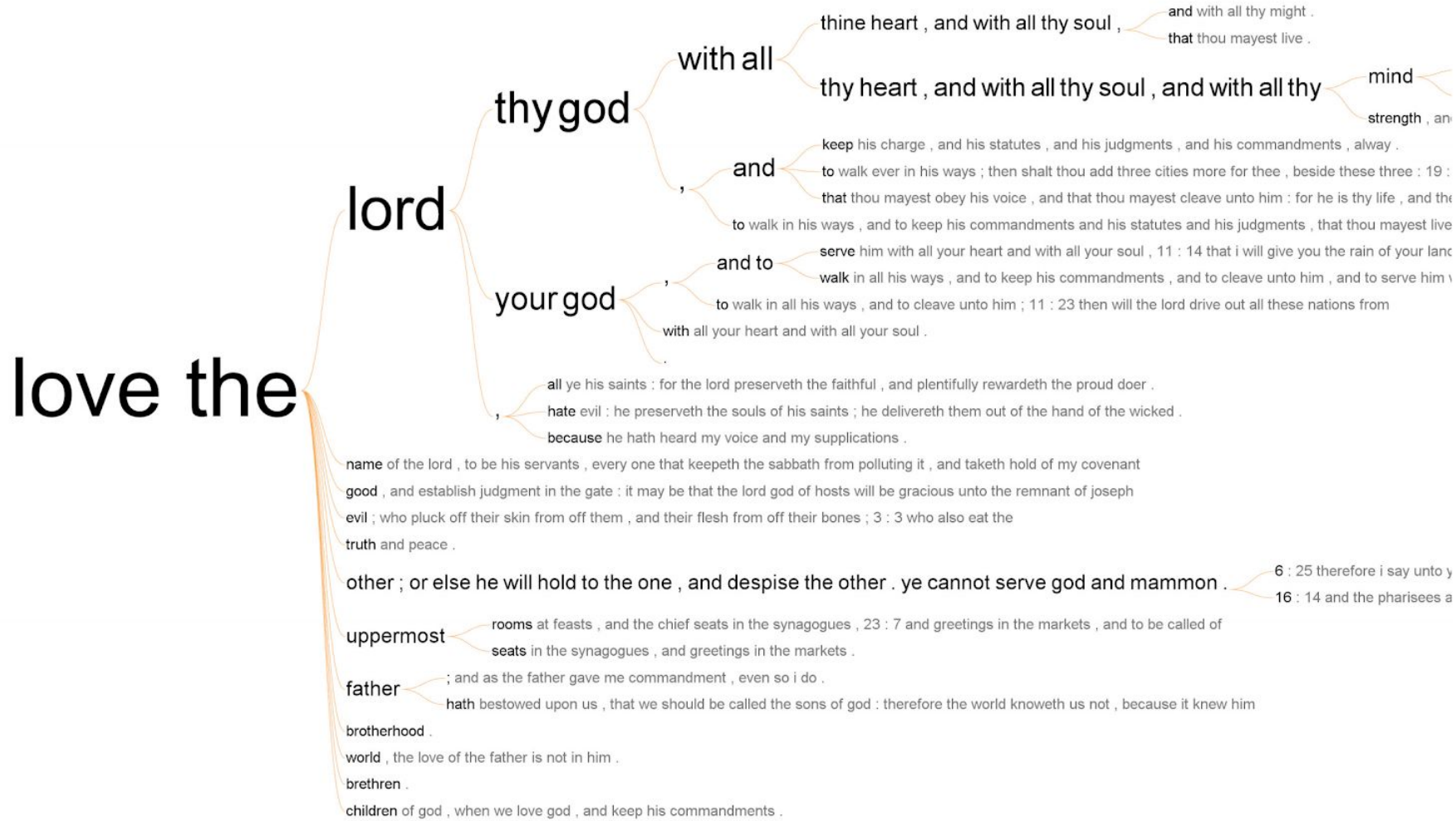
if love be rough with you , be rough with love .

if love be blind , love cannot hit the mark .

if love be blind , it best agrees with night .




i have a dream
that
one day
this nation will rise up and live out the true meaning of its creed: "We hold these truths to be self-evident,
on the red hills of Georgia the sons of former slaves and the sons of former slave owners will be able to sit down together at
even the state of Mississippi, a state sweltering with the heat of injustice, sweltering with the heat of oppression, will be transformed into
down in Alabama, with its vicious racists, with its governor having his lips dripping with the words of interposition and nullification - one day
every valley shall be exalted, and every hill and mountain shall be made low, the rough places will be made plain, and the
my four little children will one day live in a nation where they will not be judged by the color of their skin but by the
today . i have a dream that one day
down in Alabama, with its vicious racists, with its governor having his lips dripping with the words of interposition and nullification - one day
every valley shall be exalted, and every hill and mountain shall be made low, the rough places will be made plain, and the

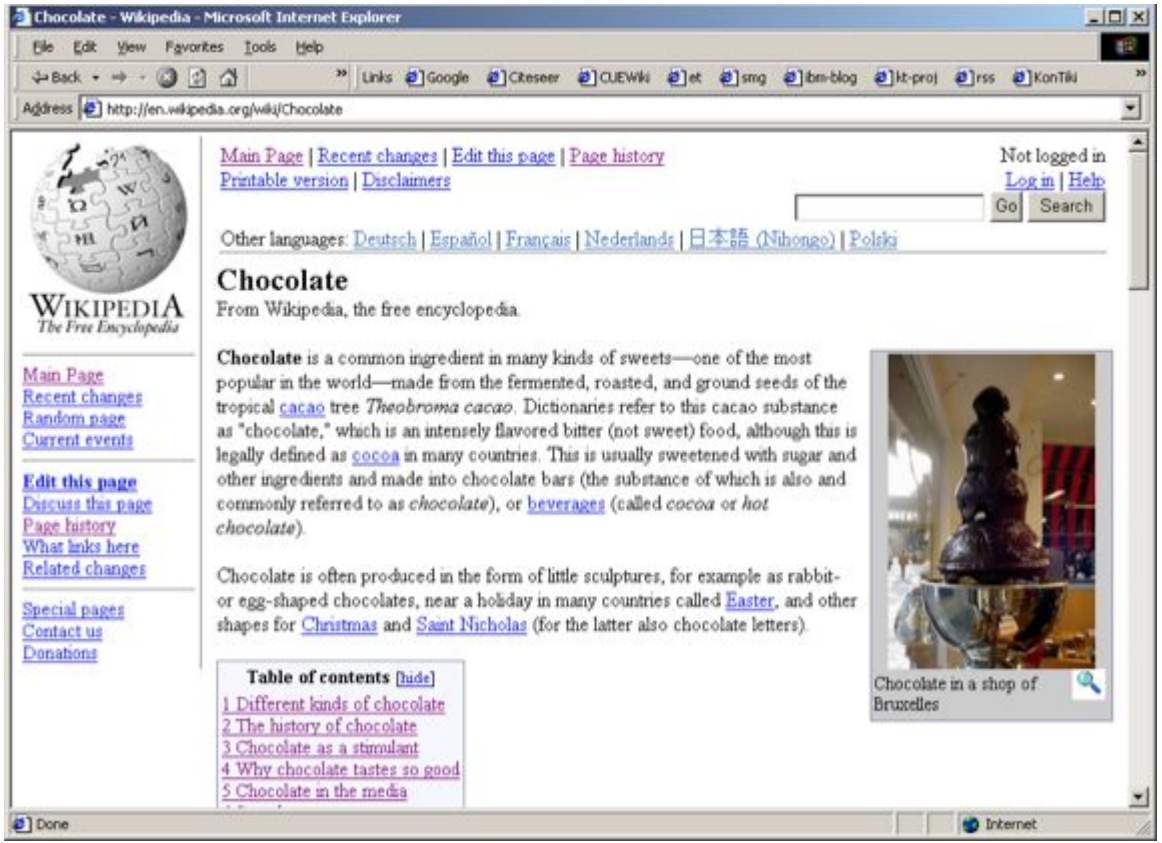

love the
lord
thy god
with all
thine heart , and with all thy soul ,
and with all thy might .
that thou mayest live .
thy heart , and with all thy soul , and with all thy
mind
strength , an
,
and
keep his charge , and his statutes , and his judgments , and his commandments , alway .
to walk ever in his ways ; then shalt thou add three cities more for thee , beside these three : 19 :
that thou mayest obey his voice , and that thou mayest cleave unto him : for he is thy life , and th
to walk in his ways , and to keep his commandments and his statutes and his judgments , that thou mayest live
your god
,
and to
serve him with all your heart and with all your soul , 11 : 14 that i will give you the rain of your lan
walk in all his ways , and to keep his commandments , and to cleave unto him , and to serve him
to walk in all his ways , and to cleave unto him ; 11 : 23 then will the lord drive out all these nations from
with all your heart and with all your soul .
.
,
all ye his saints : for the lord preserveth the faithful , and plentifully rewardeth the proud doer .
hate evil : he preserveth the souls of his saints ; he delivereth them out of the hand of the wicked .
because he hath heard my voice and my supplications .
name of the lord , to be his servants , every one that keepeth the sabbath from polluting it , and taketh hold of my covenant
good , and establish judgment in the gate : it may be that the lord god of hosts will be gracious unto the remnant of joseph
evil ; who pluck off their skin from off them , and their flesh from off their bones ; 3 : 3 who also eat the
truth and peace .
other ; or else he will hold to the one , and despise the other . ye cannot serve god and mammon .
6 : 25 therefore i say unto y
16 : 14 and the pharisees a
uppermost
rooms at feasts , and the chief seats in the synagogues , 23 : 7 and greetings in the markets , and to be called of
seats in the synagogues , and greetings in the markets .
father
; and as the father gave me commandment , even so i do .
hath bestowed upon us , that we should be called the sons of god : therefore the world knoweth us not , because it knew him
brotherhood .
world , the love of the father is not in him .
brethren .
children of god , when we love god , and keep his commandments .

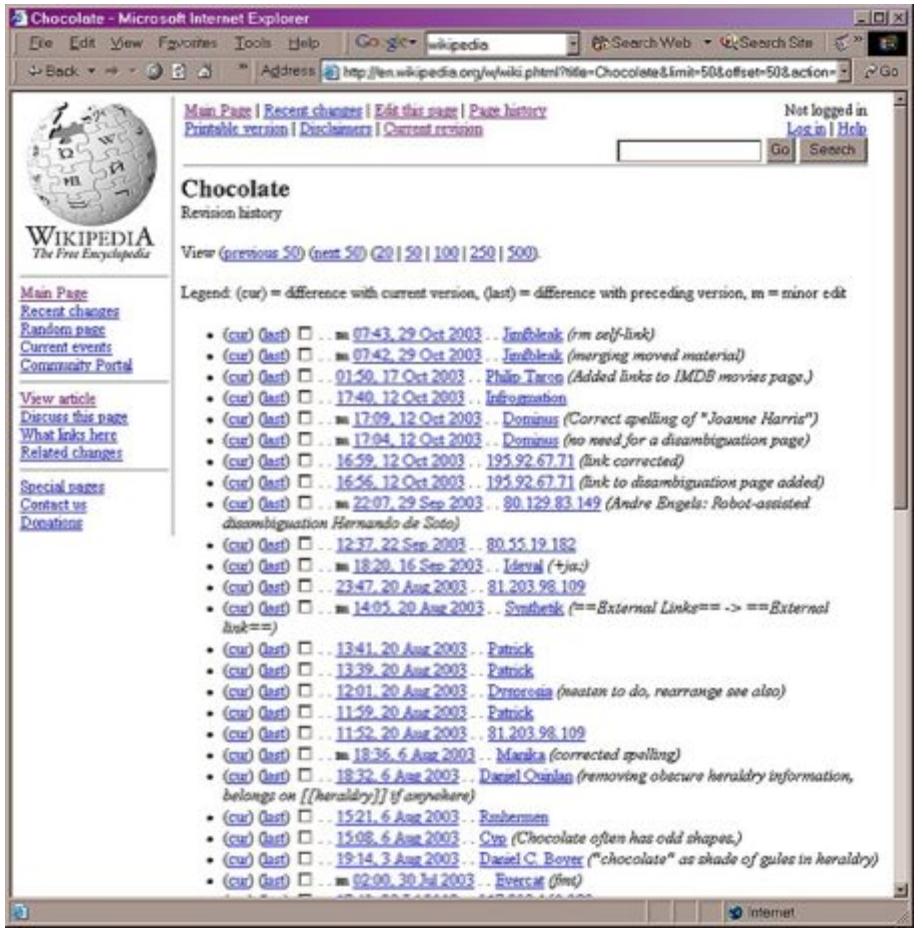
Word tree demo

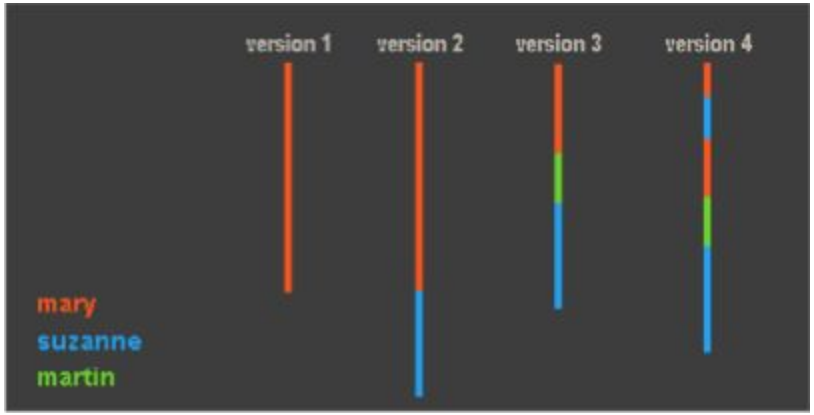
Wikipedia in 2003

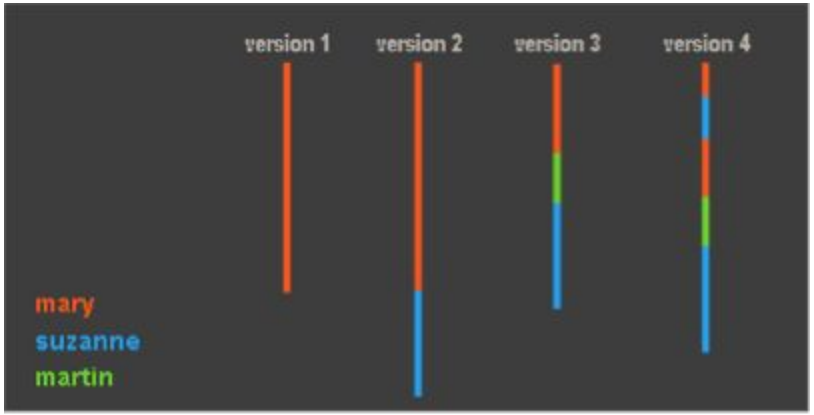
back then...

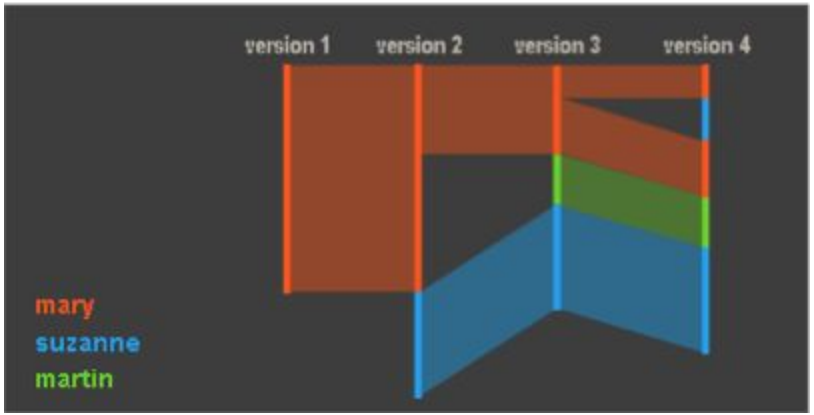
Wikipedia article

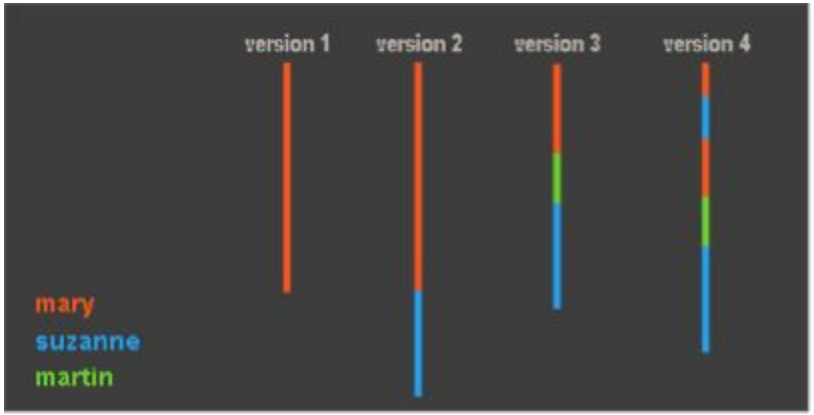


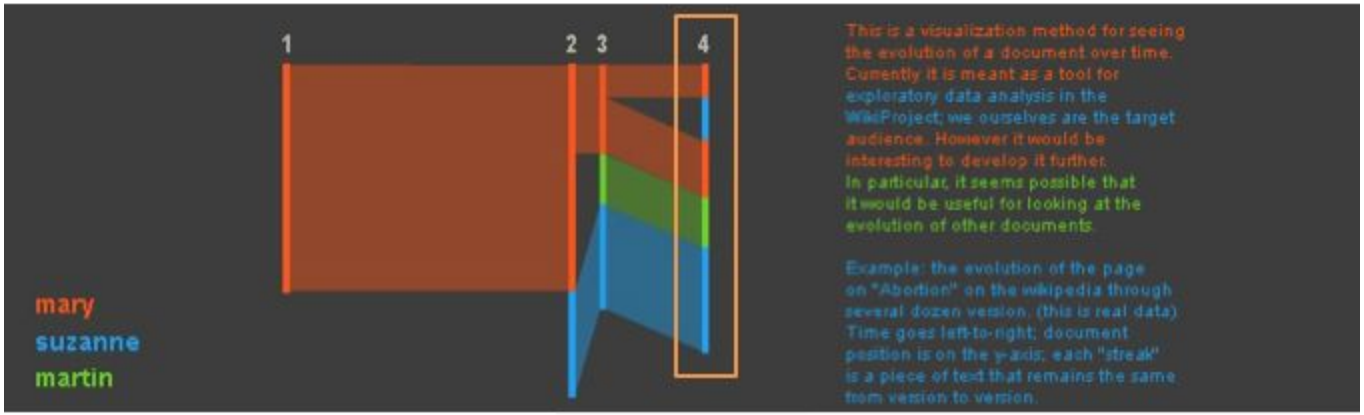

Chocolate - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Address
WIKIPEDIA
The Free Encyclopedia
Main Page
Recent changes
Random page
Current events
Community Portal
View article
Discuss this page
What links here
Related changes
Special pages
Contact us
Donations
Main Page | Recent changes | Edit this page | Page history
Printable version | Disclaimers | Current revision
Not logged in
Log in | Help
Go
Search
Chocolate
Revision history
View (previous 50) (next 50) (20 | 50 | 100 | 250 | 500).
Legend: (cur) = difference with current version, (last) = difference with preceding version, m = minor edit
(cur) (last) m . . 07:43, 29 Oct 2003 . . Jmfbleak (rm self-link)
(cur) (last) m . . 07:42, 29 Oct 2003 . . Jmfbleak (merging moved material)
(cur) (last) . . 01:50, 17 Oct 2003 . . Philip Taron (Added links to IMDB movies page.)
(cur) (last) . . 17:40, 12 Oct 2003 . . Infrogmation
(cur) (last) m . . 17:09, 12 Oct 2003 . . Dominus (Correct spelling of "Joanne Harris")
(cur) (last) m . . 17:04, 12 Oct 2003 . . Dominus (no need for a disambiguation page)
(cur) (last) . . 16:59, 12 Oct 2003 . . 195.92.67.71 (link corrected)
(cur) (last) . . 16:56, 12 Oct 2003 . . 195.92.67.71 (link to disambiguation page added)
(cur) (last) m . . 22:07, 29 Sep 2003 . . 80.129.83.149 (Andre Engels: Robot-assisted disambiguation Hernando de Soto)
(cur) (last) . . 12:37, 22 Sep 2003 . . 80.55.19.182
(cur) (last) m . . 18:20, 16 Sep 2003 . . Ideval (+ja;)
(cur) (last) . . 23:47, 20 Aug 2003 . . 81.203.98.109
(cur) (last) m . . 14:05, 20 Aug 2003 . . Synthetik (==External Links== -> ==External link==)
(cur) (last) . . 13:41, 20 Aug 2003 . . Patrick
(cur) (last) . . 13:39, 20 Aug 2003 . . Patrick
(cur) (last) . . 12:01, 20 Aug 2003 . . Dysprosia (reason to do, rearrange see also)
(cur) (last) . . 11:59, 20 Aug 2003 . . Patrick
(cur) (last) . . 11:52, 20 Aug 2003 . . 81.203.98.109
(cur) (last) m . . 18:36, 6 Aug 2003 . . Maraka (corrected spelling)
(cur) (last) . . 18:32, 6 Aug 2003 . . Daniel Quinlan (removing obscure heraldry information, belongs on [[heraldry]] if anywhere)
(cur) (last) . . 15:21, 6 Aug 2003 . . Rmhermen
(cur) (last) . . 15:08, 6 Aug 2003 . . Cyp (Chocolate often has odd shapes.)
(cur) (last) . . 19:14, 3 Aug 2003 . . Daniel C. Boyer ("chocolate" as shade of gules in heraldry)
(cur) (last) m . . 02:00, 30 Jul 2003 . . Evercat (fmt)
Internet


version 1
version 2
version 3
version 4
mary
suzanne
martin


version 1
version 2
version 3
version 4
mary
suzanne
martin

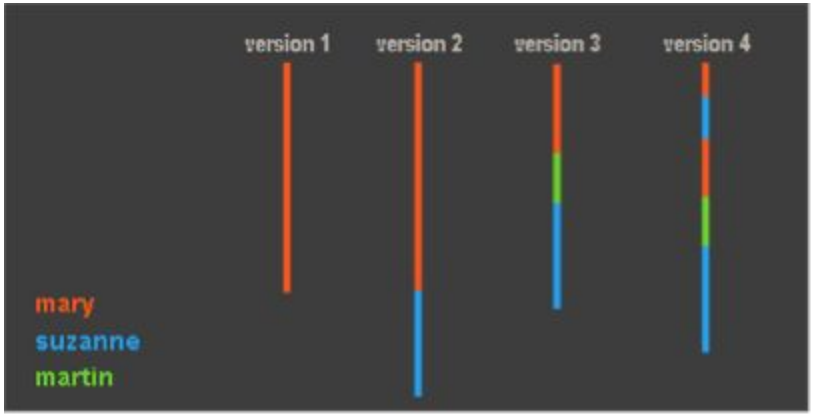

version 1
version 2
version 3
version 4
mary
suzanne
martin



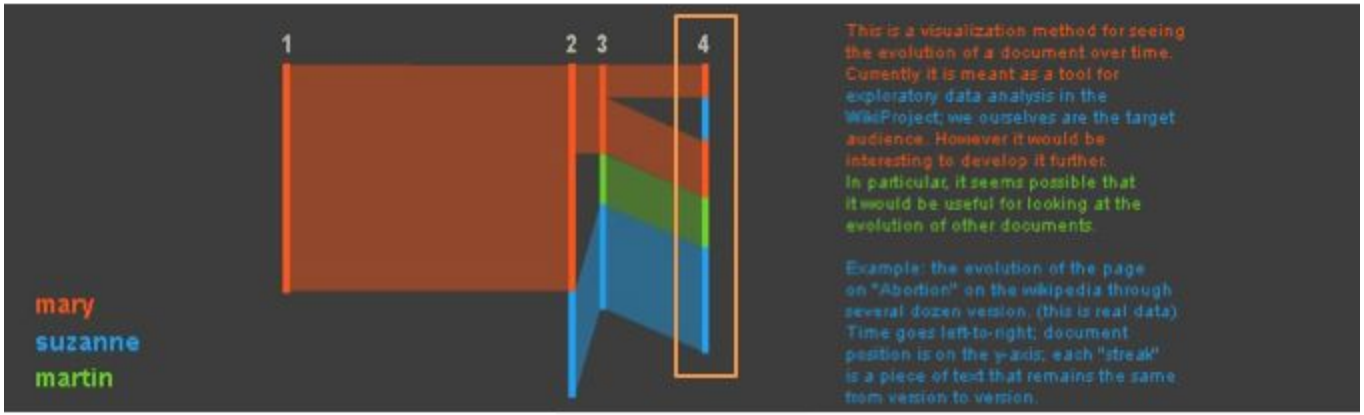


This is a visualization method for seeing the evolution of a document over time. Currently it is meant as a tool for exploratory data analysis in the WikiProject; we ourselves are the target audience. However it would be interesting to develop it further. In particular, it seems possible that it would be useful for looking at the evolution of other documents.

Example: the evolution of the page on "Abortion" on the wikipedia through several dozen version. (this is real data) Time goes left-to-right; document position is on the y-axis; each "streak" is a piece of text that remains the same from version to version.



history flow demo



Machine Learning

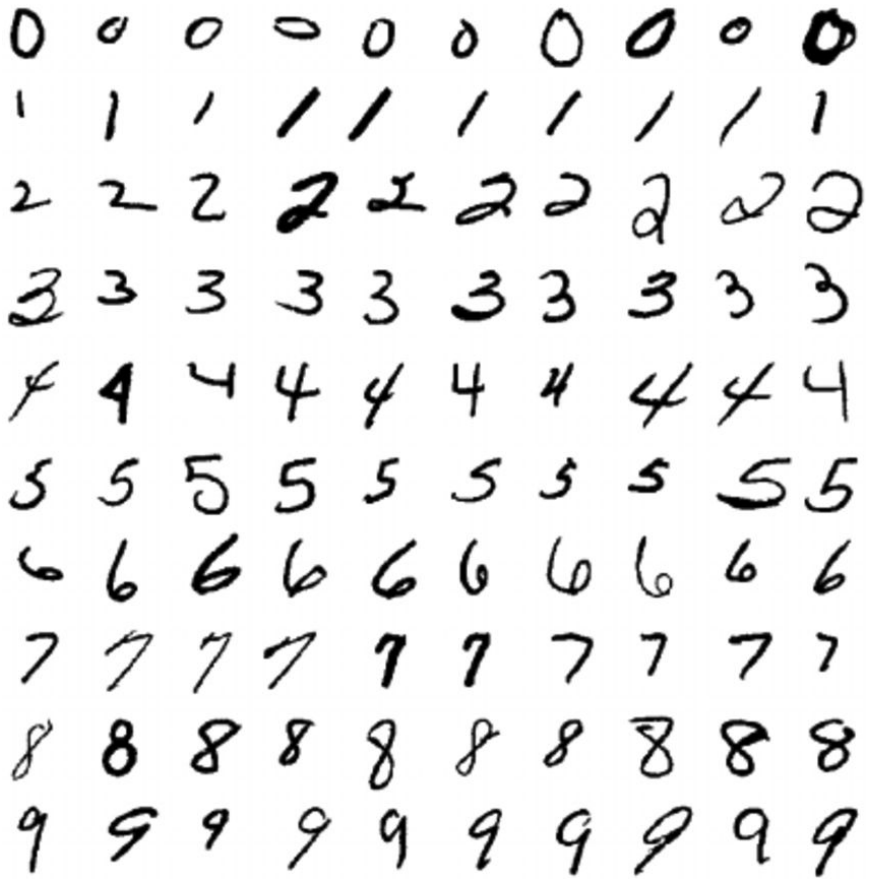
Massively high dimensional spaces

Machine learning & AI: a huge area!

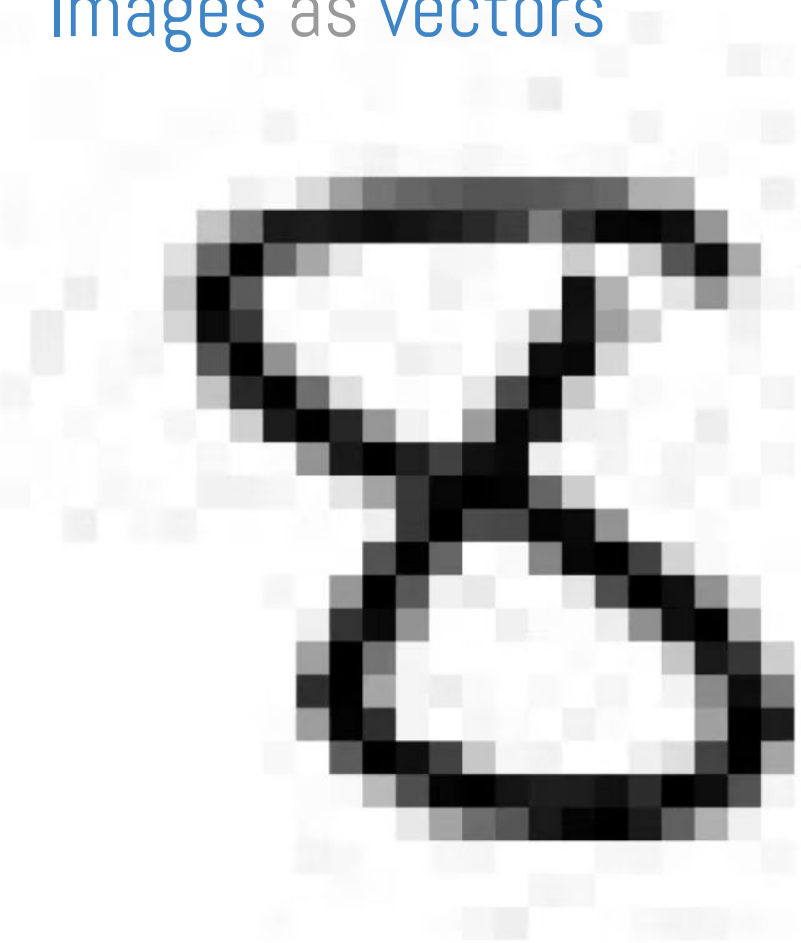
Today we'll focus on just one aspect:
Visualizing high-dimensional data

Today this is an “expert-only” world, but we predict it becomes a standard tool…

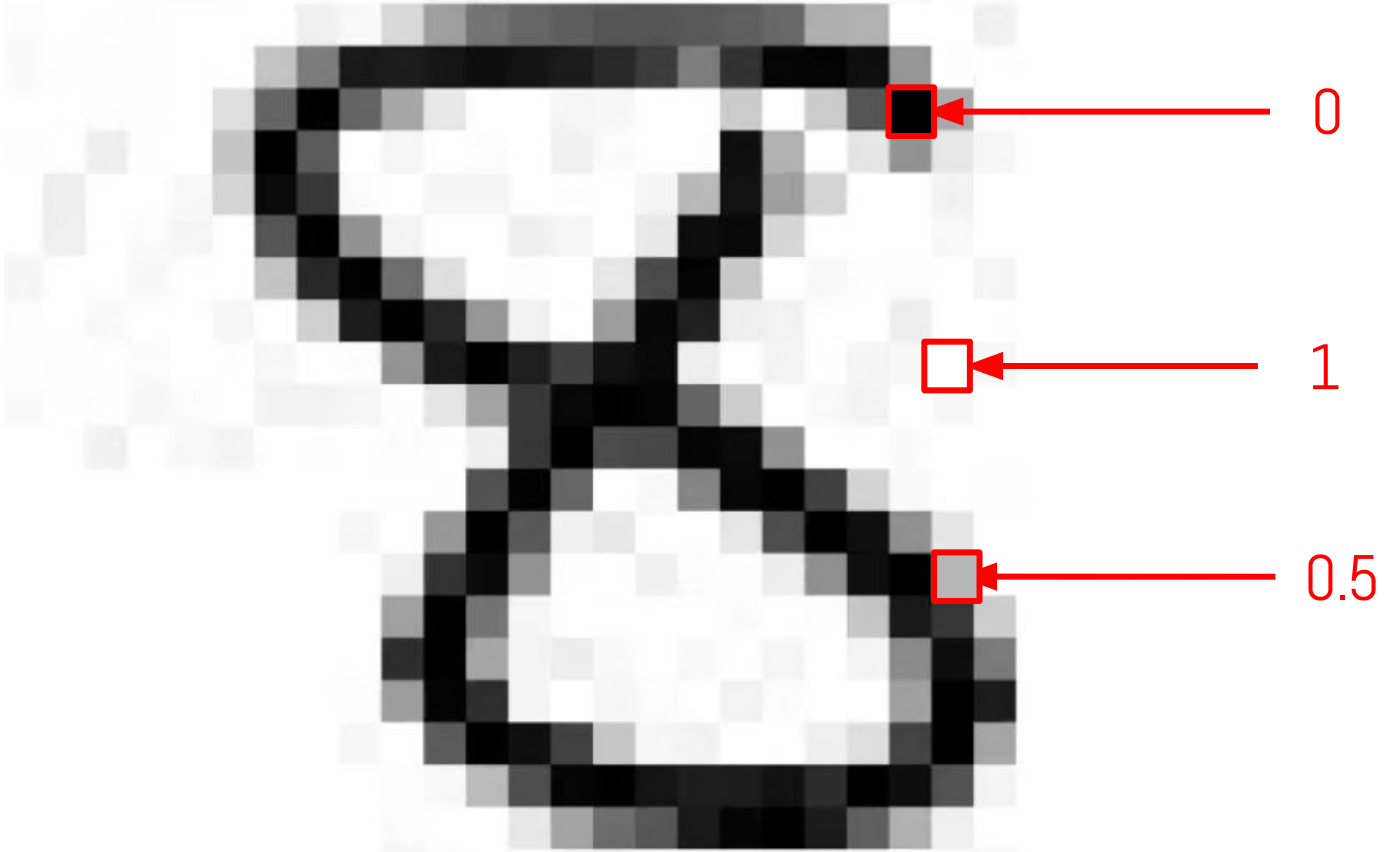
Images are high-dimensional data



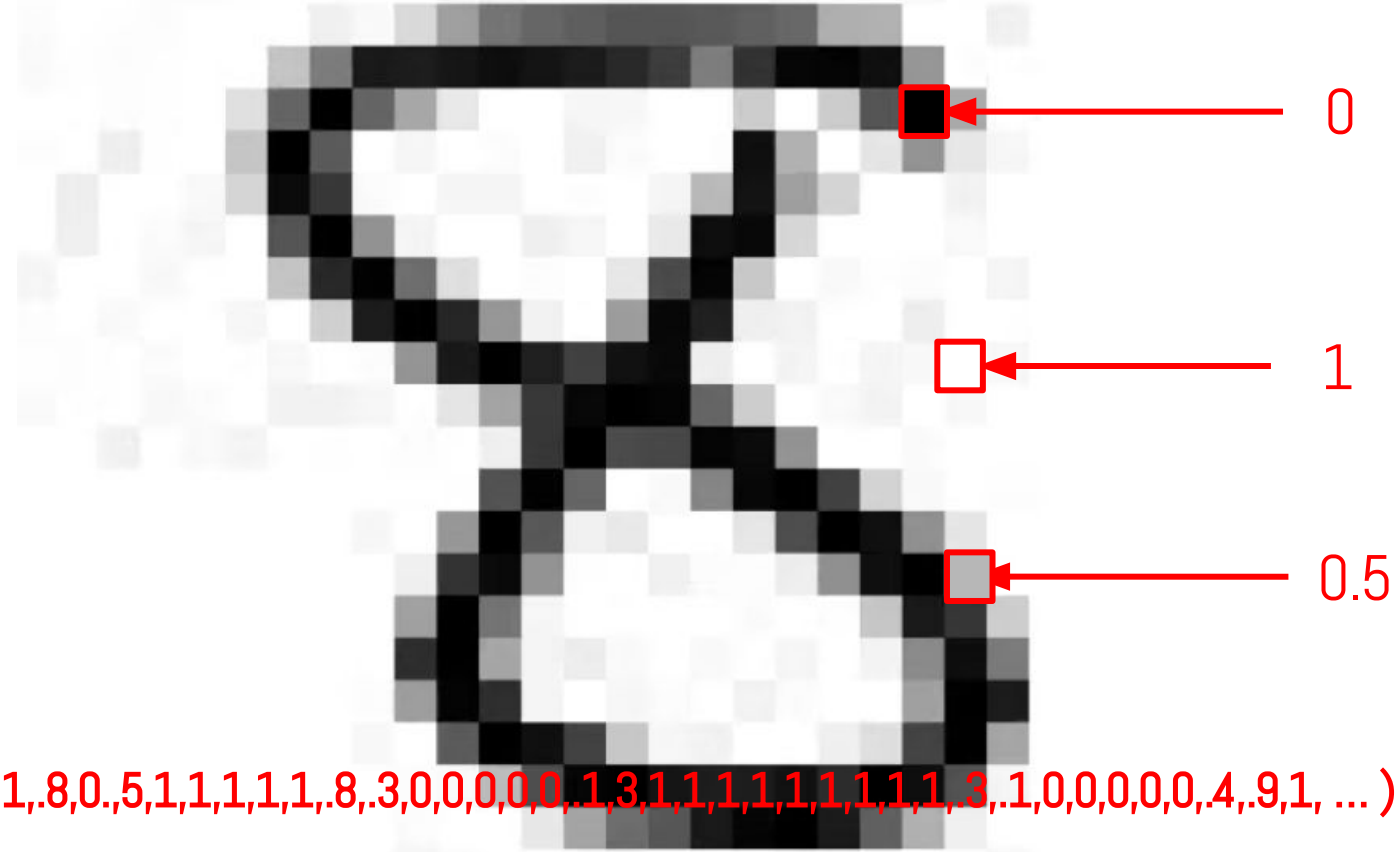
Images as vectors



Images as vectors

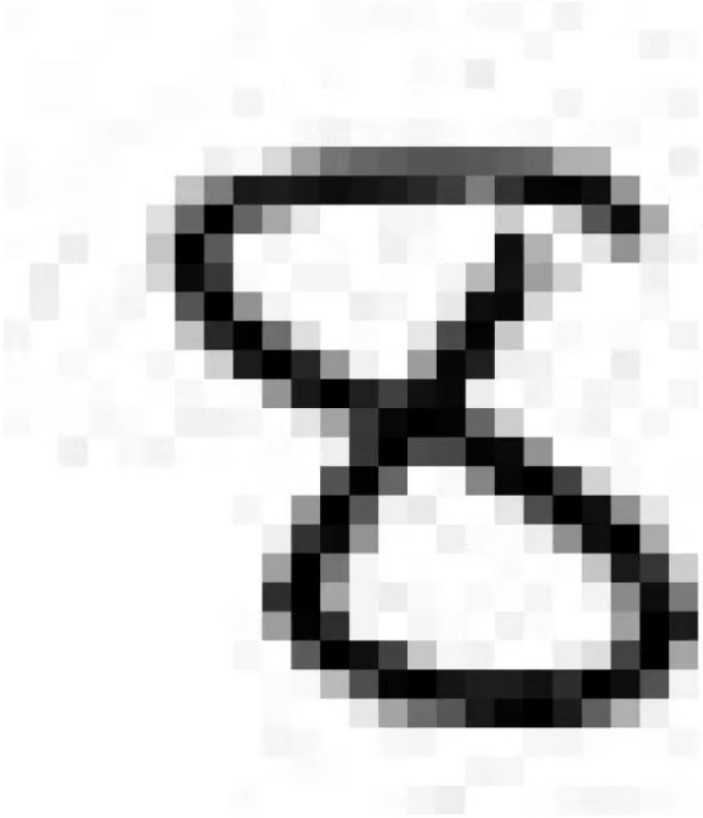


Images as vectors



(1,1,1,1,1,1,1,1,1,.8,0.,5,1,1,1,1,1,.8,.3,0,0,0,0,0,.1,3,1,1,1,1,1,1,1,1,.3,.1,0,0,0,0,0,.4,.9,1, ...)

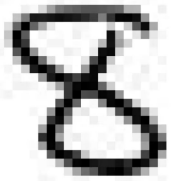
We've turned this image → into a vector



→ (1,1,1,1,1,1,1,1,1,.8,0.,5,1,1,1,...)

784 pixels → 784 dimensions

We've turned this image → into a vector



(1,1,1,1,1,1,1,1,**1,.8,0.,5,1,**1,1,...)

(1,1,1,1,1,1,1,1,**.6,.7,0.,4,1,**1,1,...)

(1,1,1,1,1,1,1,1,**.4,.5,0.,3,.2,**1,1,...)

.
.
.

What does this buy us?

Once we can represent something as a vector, a whole world of powerful math is available.

We can use that math to **map the data!**

Embedding Projector demo

Model interpretability use case

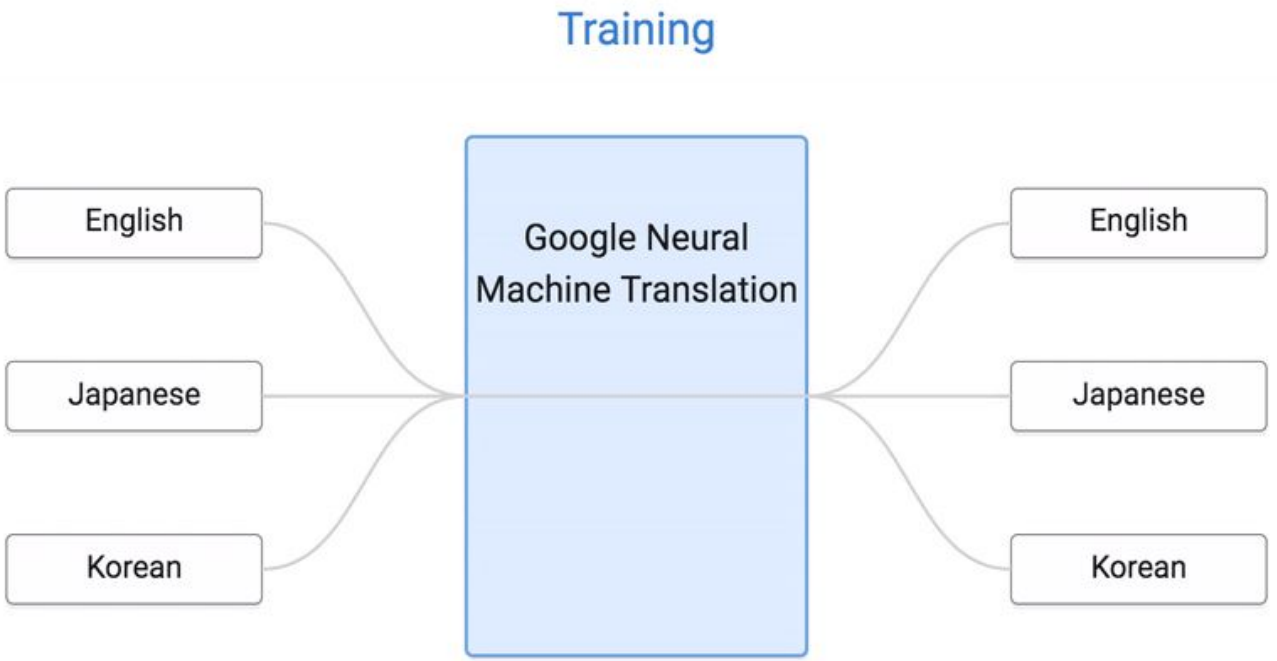
Multi-lingual translation

What does the language embedding space look like?

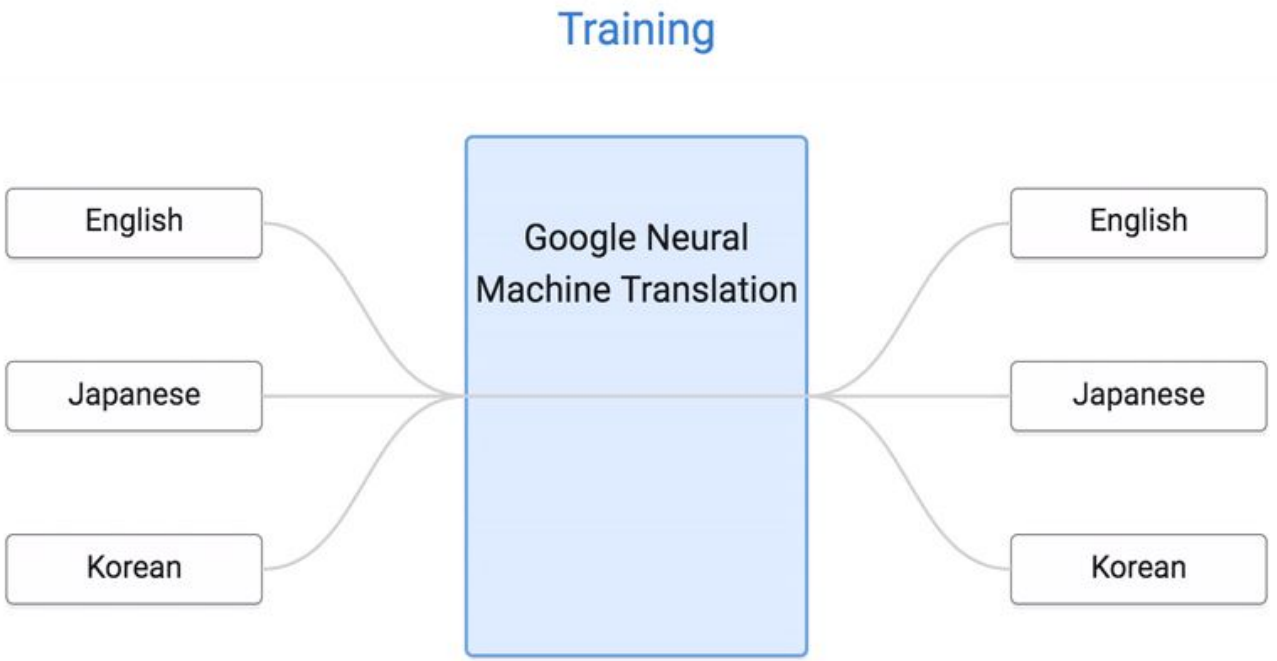
https://arxiv.org/abs/1611.04558
Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation

Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, Jeffrey Dean

Training: English ⟷ Japanese
English ⟷ Korean

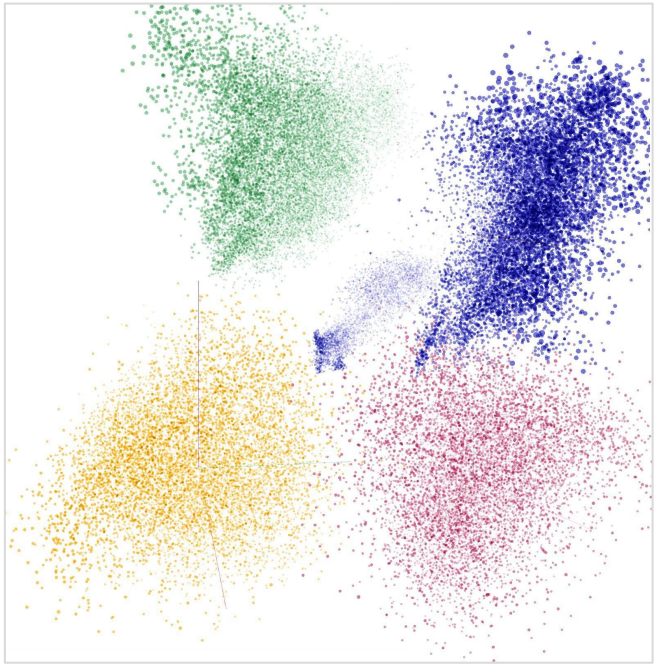


Training: English ⟷ Japanese
English ⟷ Korean
Japanese ⟷ Korean (zero shot)

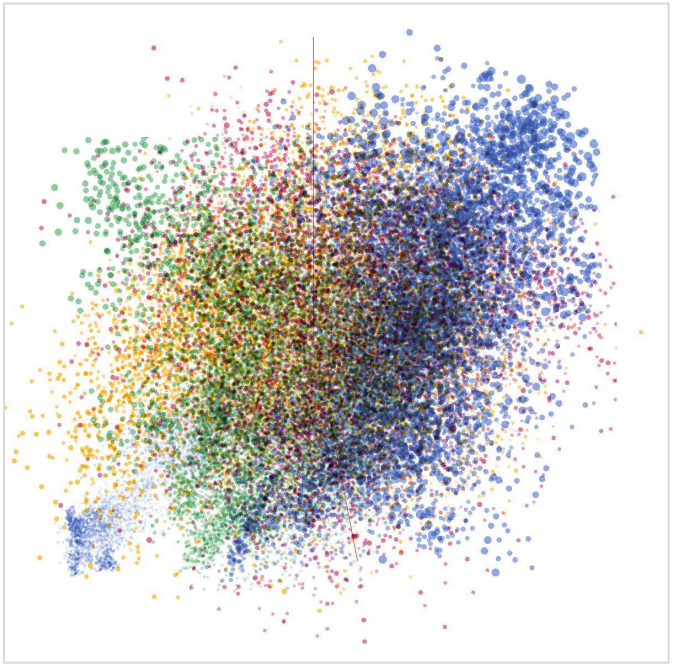


Research question

What does the multi-language embedding space look like?



or



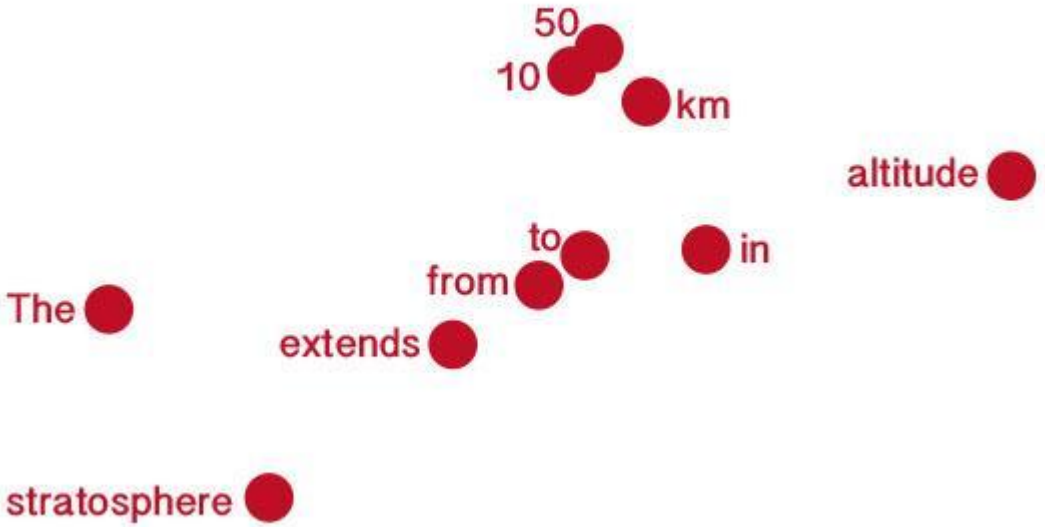
Note: not real data

What does a sentence look like in embedding space?

(points in 1024-dim space: the data that the decoder receives)

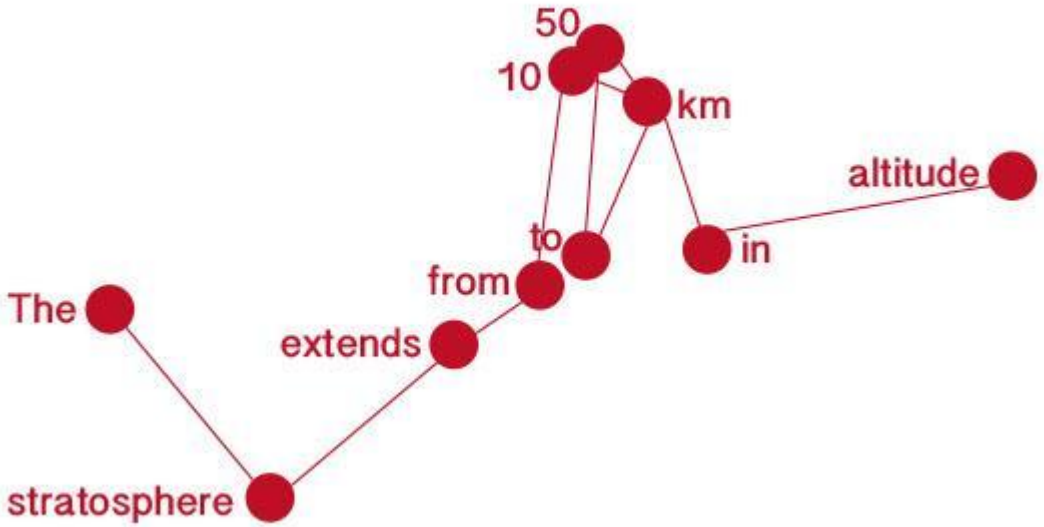
E.g. "*The stratosphere extends from 10km to 50km in altitude*"

What does a sentence look like in embedding space?



Note: simplification of real situation!

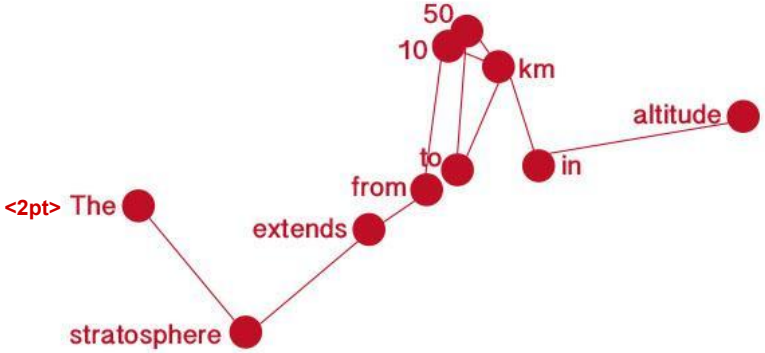
What does a sentence look like in embedding space?

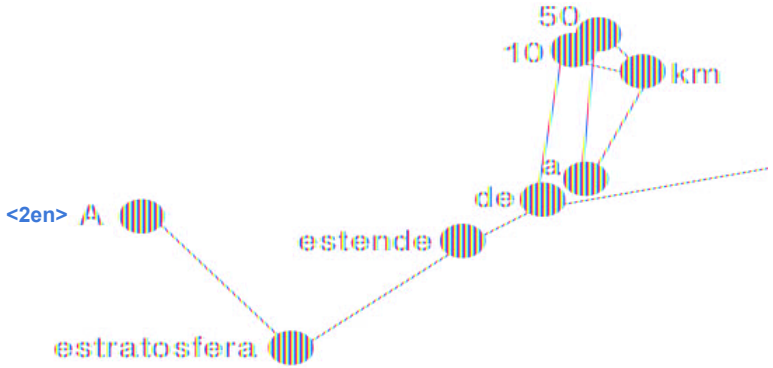


What do parallel sentences look like in embedding space?

(same meaning, different languages)

Like this?





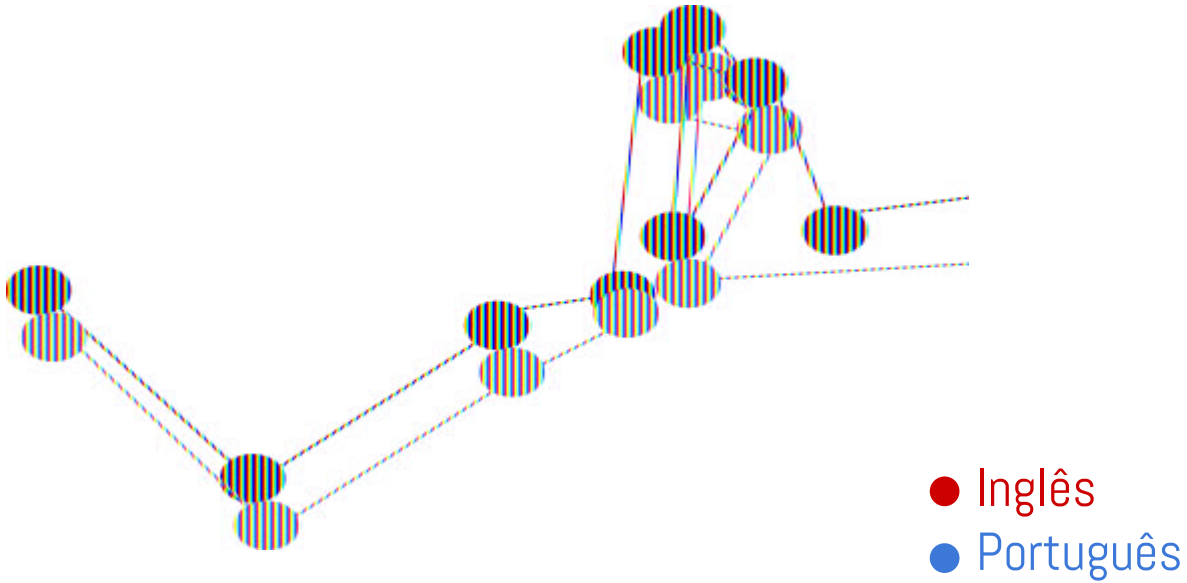
● Inglês

● Português

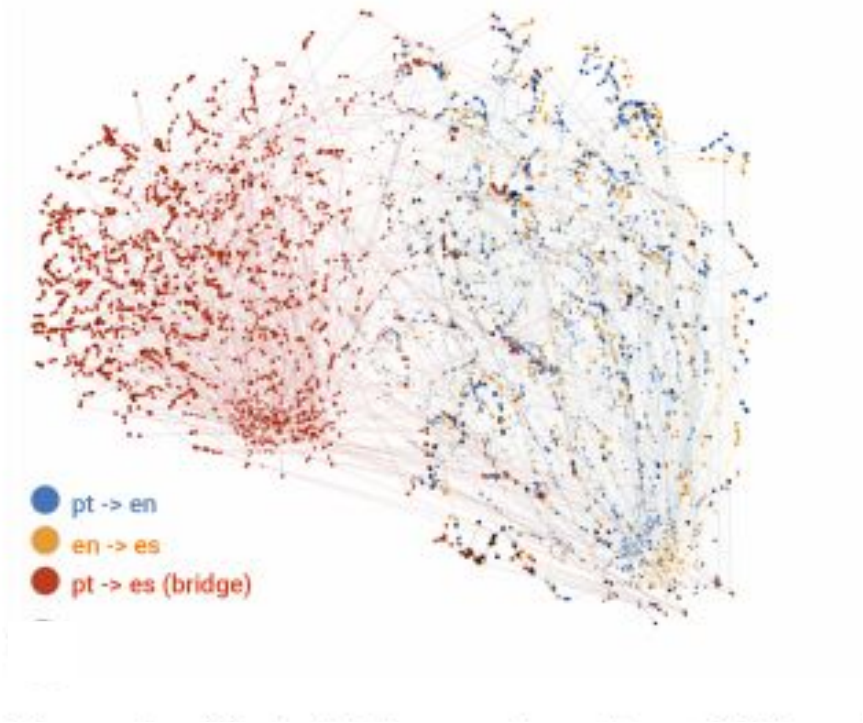
What do parallel sentences look like in embedding space?

(same meaning, different languages)

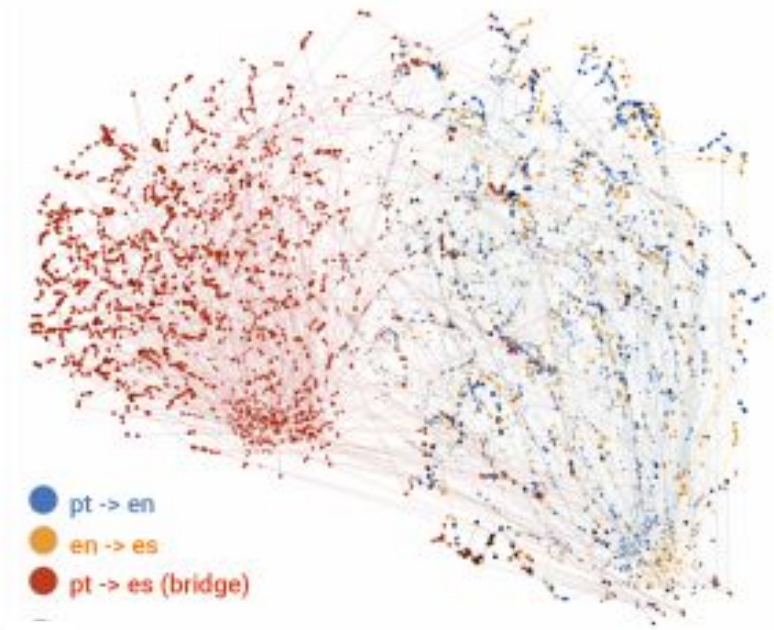
Or like this?

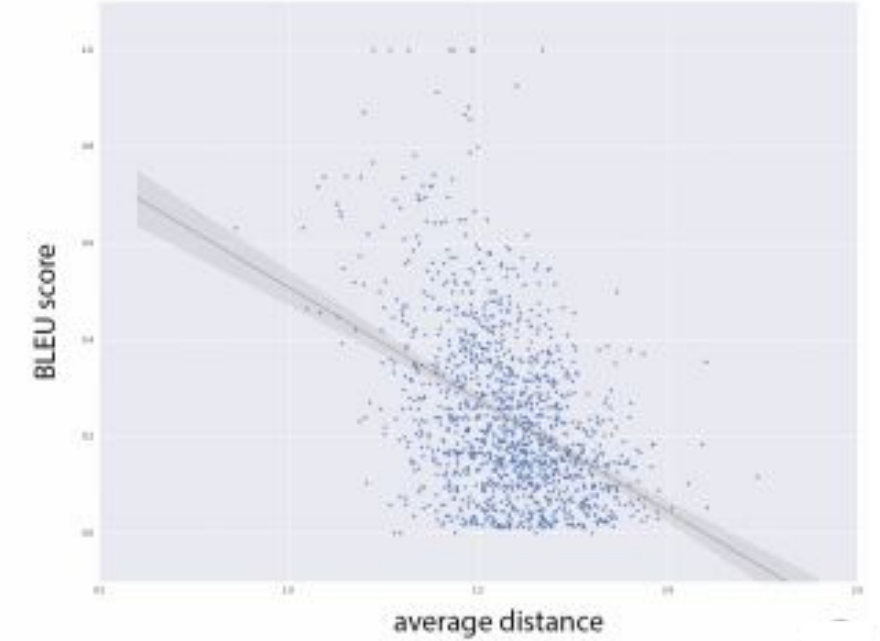


Sister model: Portuguese, Spanish, English

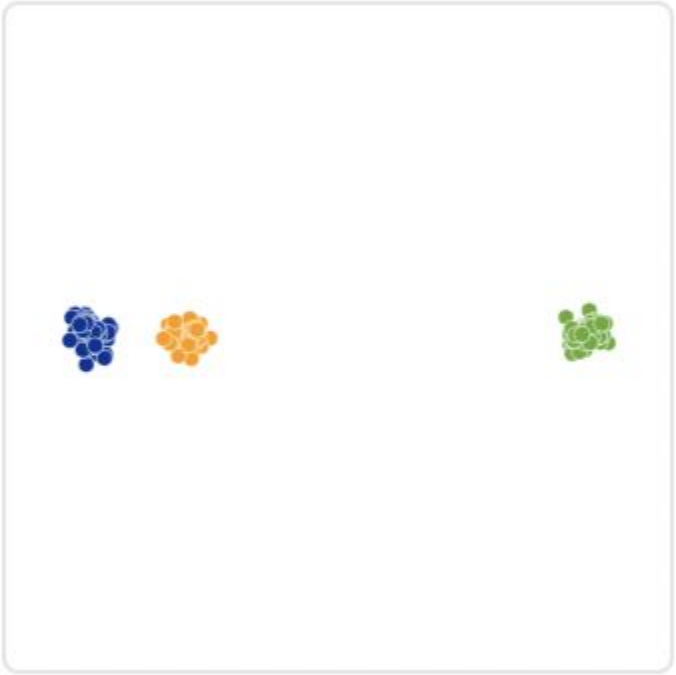


Sister model: Portuguese, Spanish, English



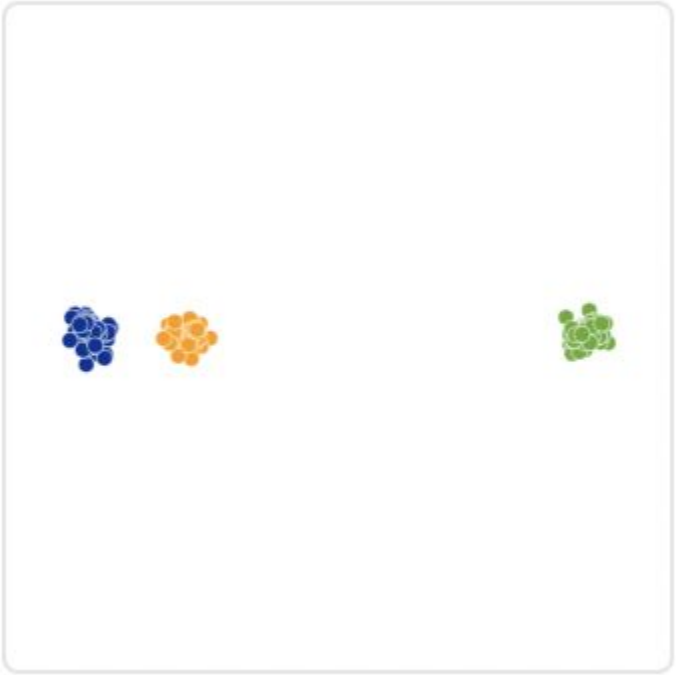


Distances between clusters can be deceptive



Original: synthetic data in 2D

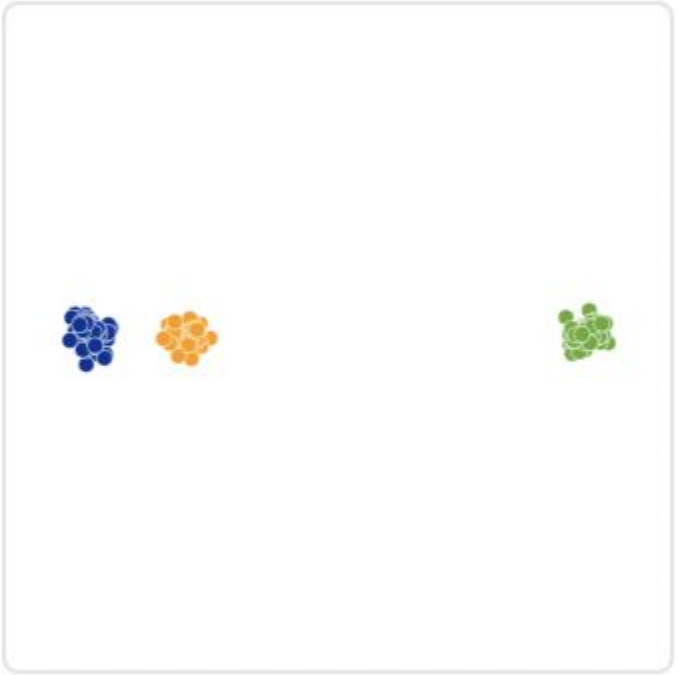
Distances between clusters can be deceptive



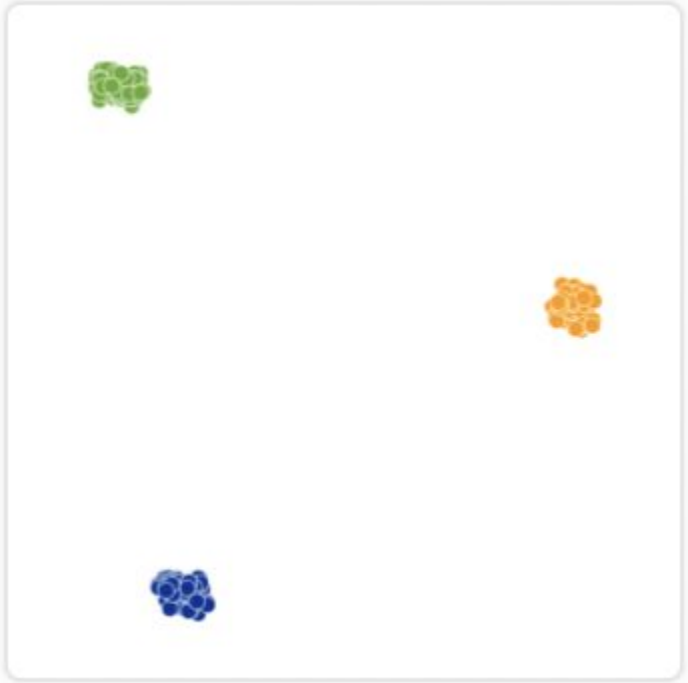
Original: synthetic data in 2D

t-SNE view of the same data

Distances between clusters can be deceptive

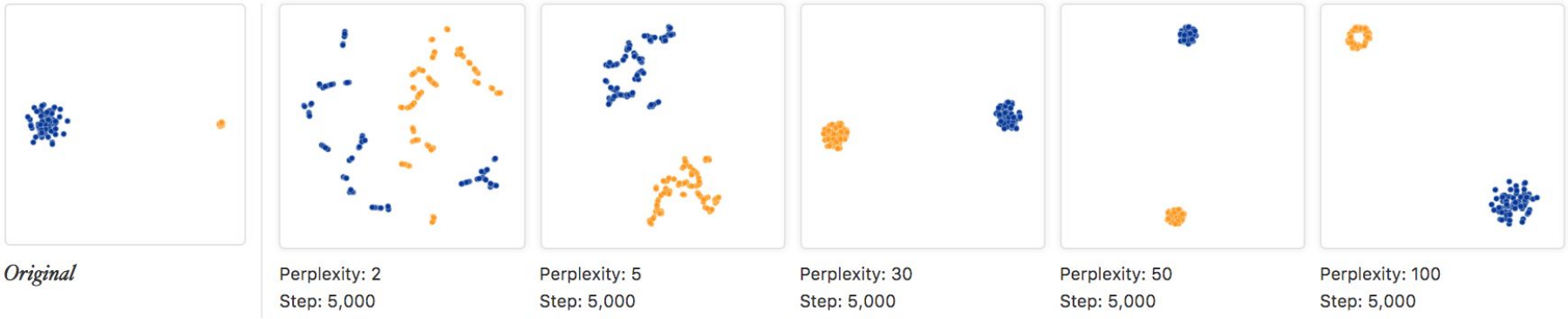


Original: synthetic data in 2D



t-SNE view of the same data

Cluster sizes in a t-SNE plot mean nothing

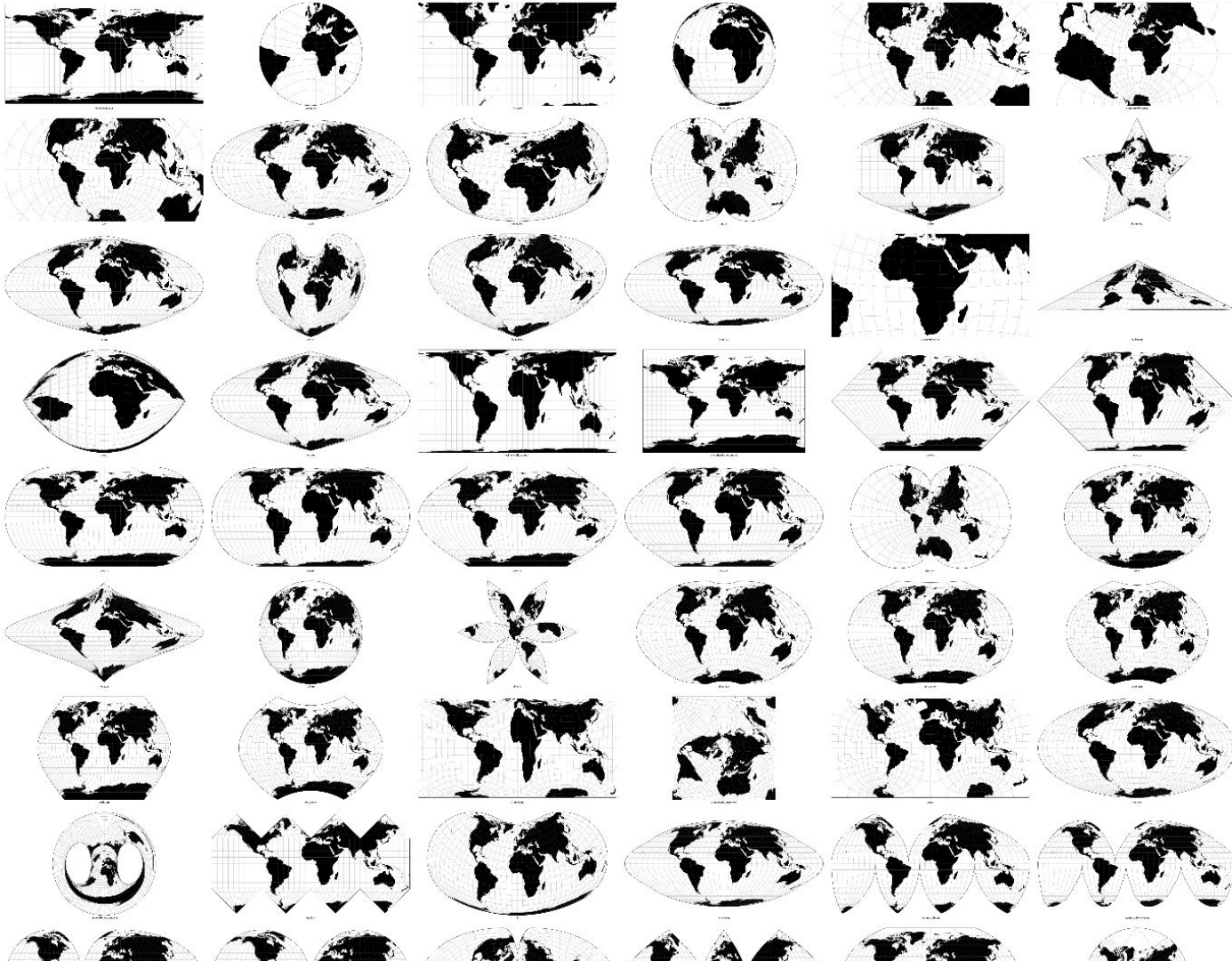


How to Use t-SNE Effectively, Wattenberg, Viégas, Johnson

Projections can fool us

The same data can look many different ways

Not a new problem

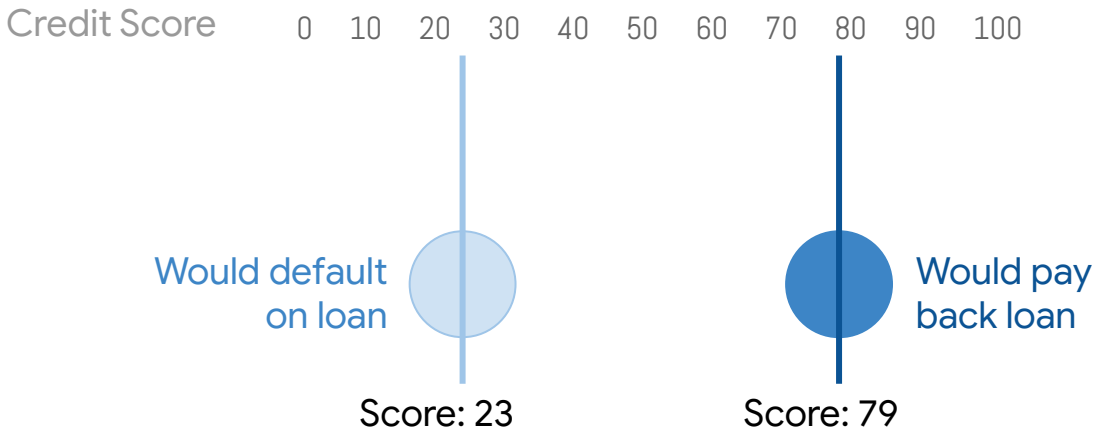


Every Map Projection
Mike Bostock

Visualization to probe machine learning

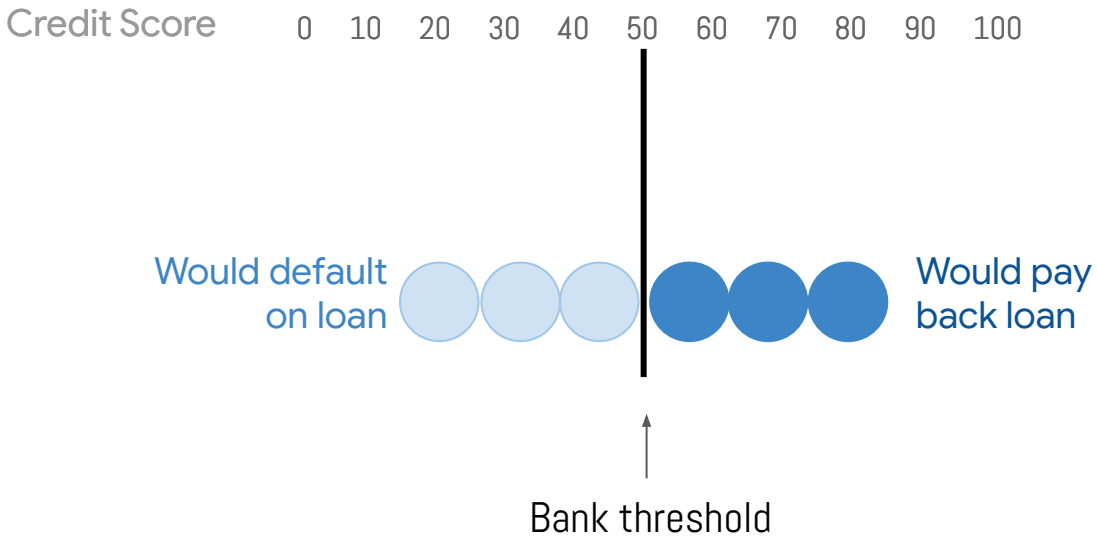
Bank loan scenario: Fairness in machine learning

A hypothetical, idealized example to understand tradeoffs



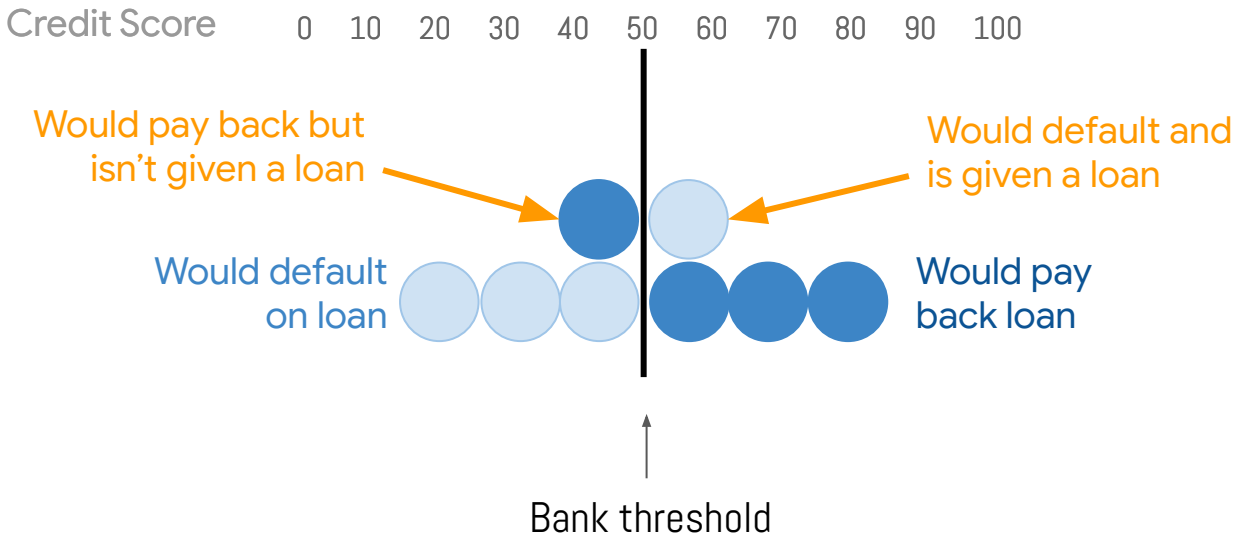
URL: research.google.com/bigpicture/attacking-discrimination-in-ml/

Bank loan scenario



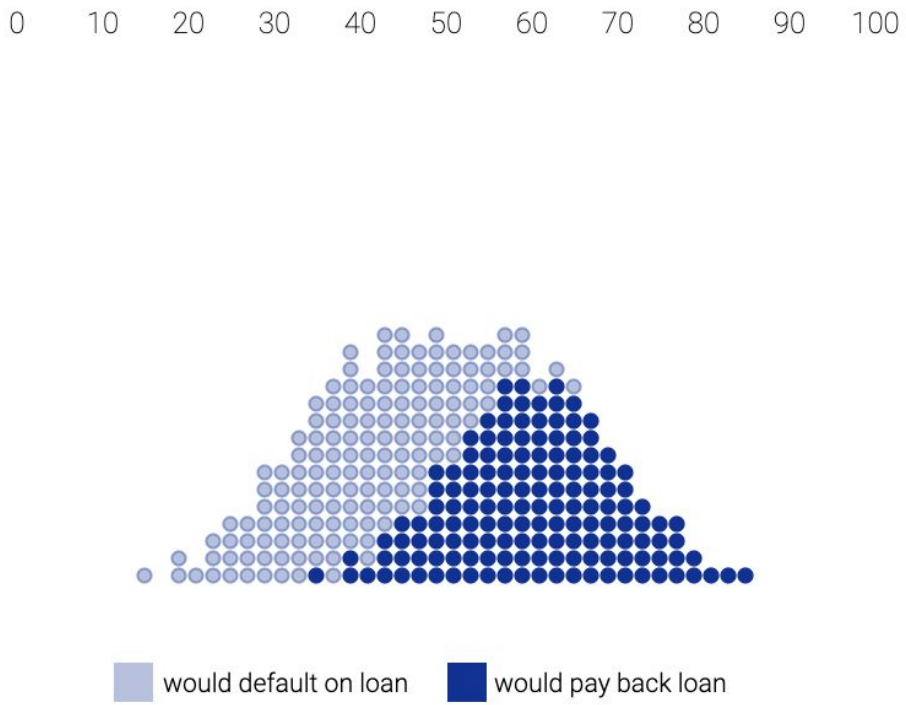
URL: research.google.com/bigpicture/attacking-discrimination-in-ml/

Bank loan scenario

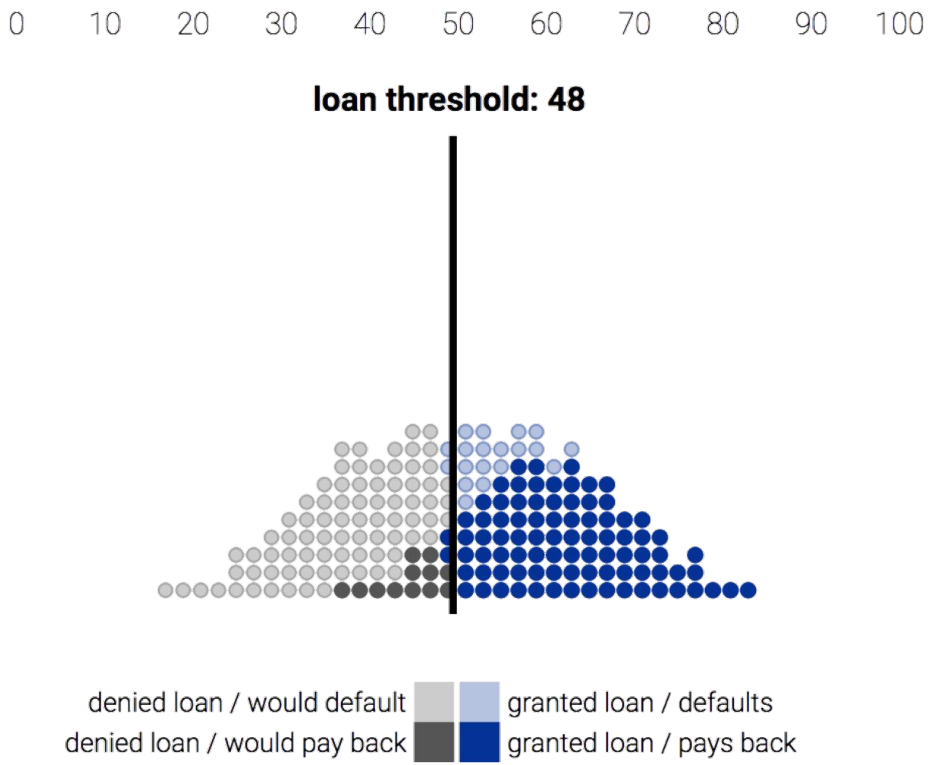


URL: research.google.com/bigpicture/attacking-discrimination-in-ml/

Bank loan scenario

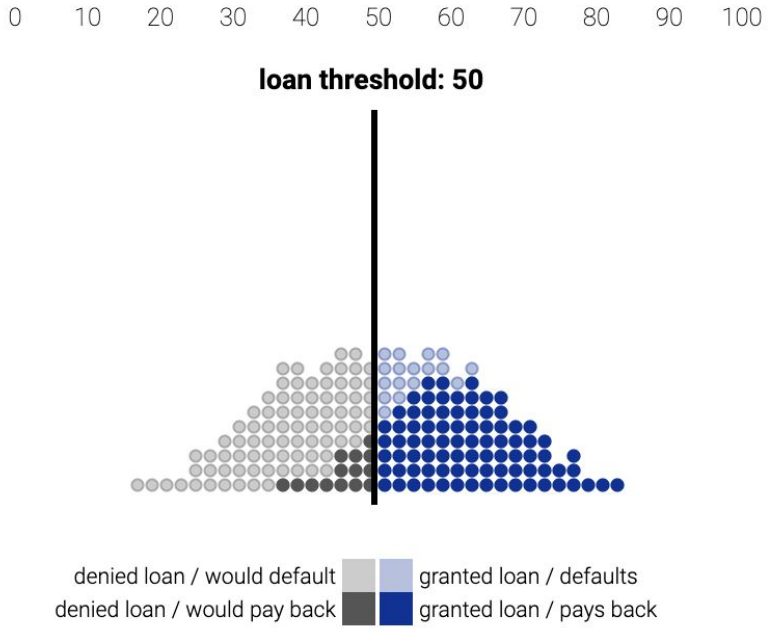


Bank loan scenario



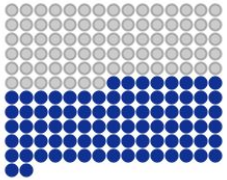
Bank loan scenario

Threshold Decision



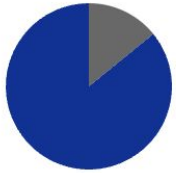
Outcome

Correct 84%
loans granted to paying applicants and denied to defaulters



Incorrect 16%
loans denied to paying applicants and granted to defaulters

True Positive Rate 86%
percentage of paying applications getting loans

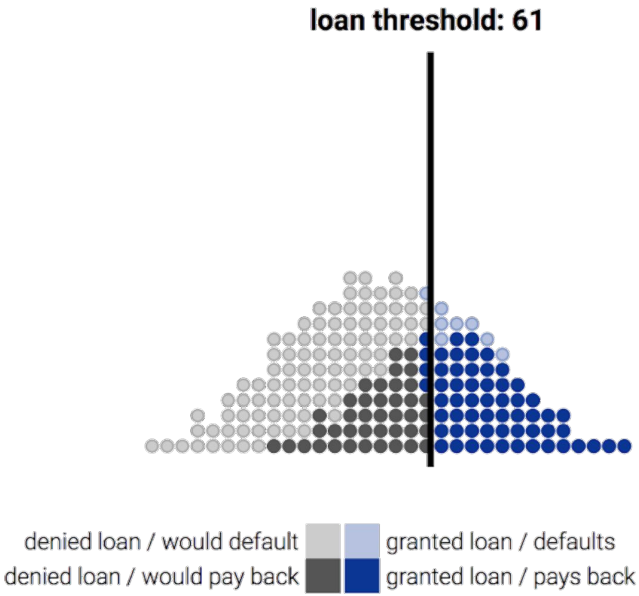


Positive Rate 52%
percentage of all applications getting loans

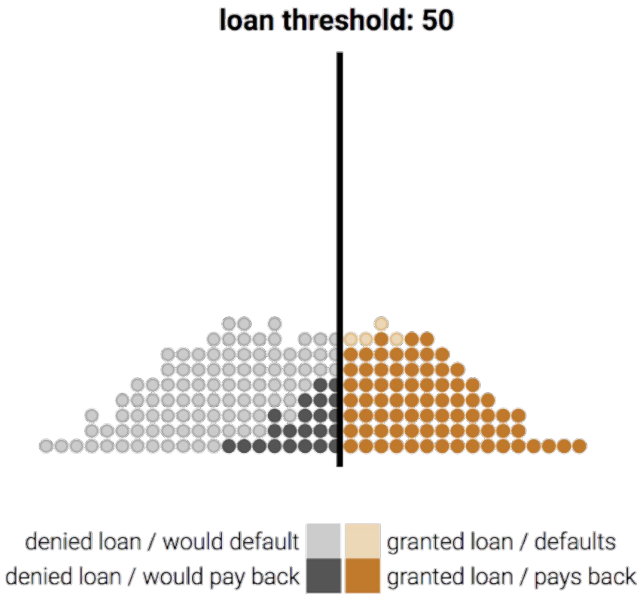
Profit: **13600**

Bank loan scenario

Blue Population

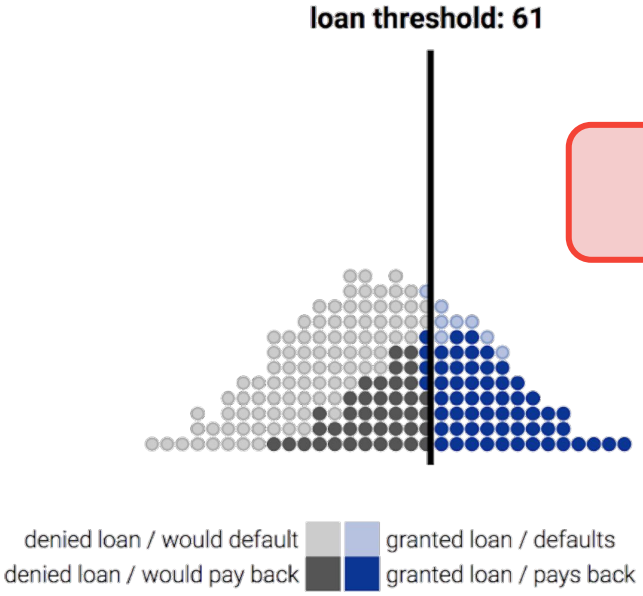


Orange Population



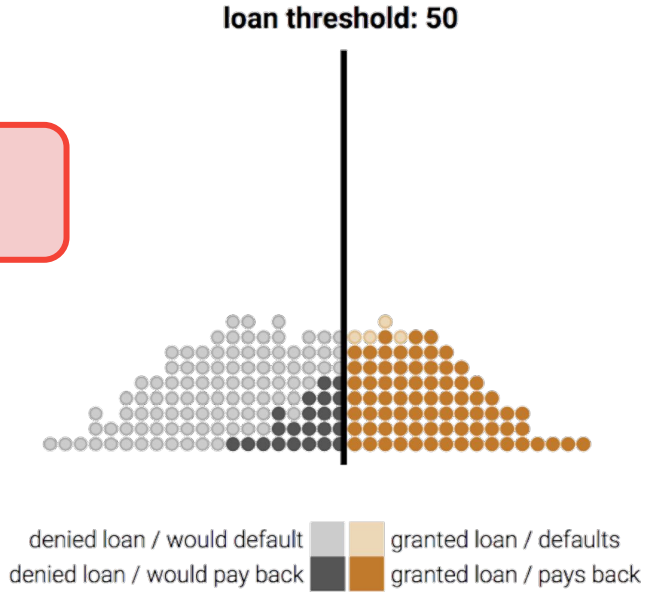
Bank loan scenario

Blue Population



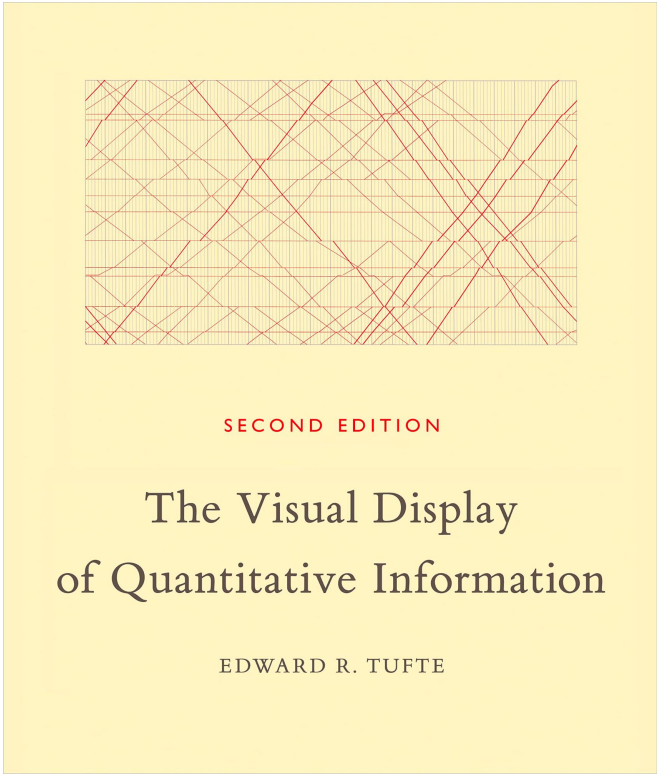
demo

Orange Population

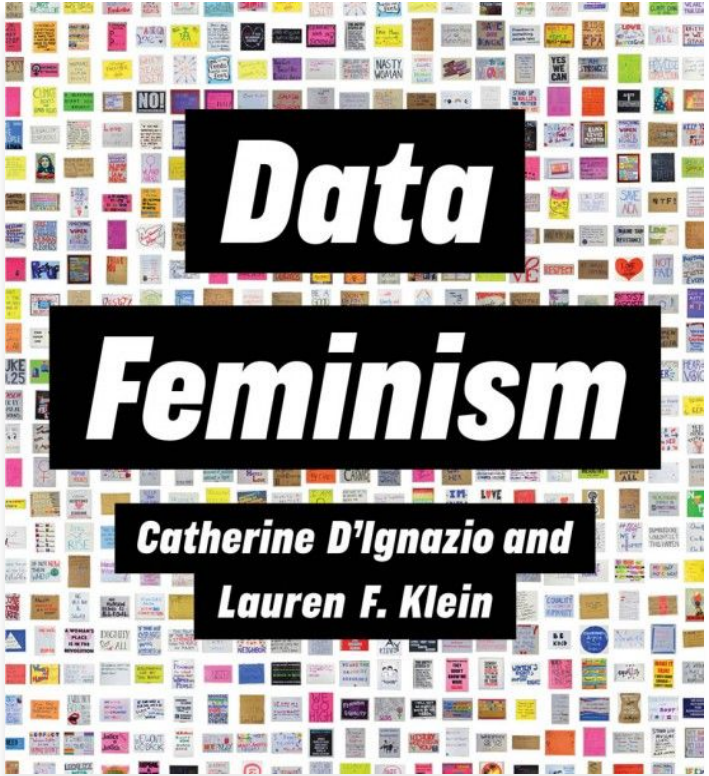


Different perspectives and uses of data vis

Different voices



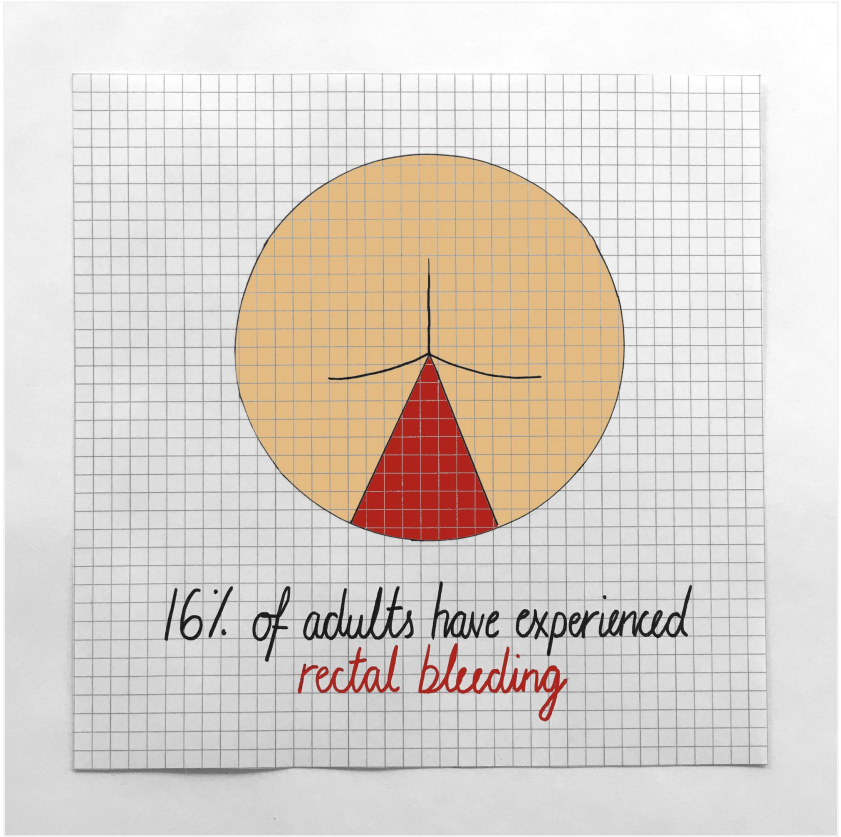
Data-ink ratio, neutral point of view



Elevate emotion in data visualization

“If it’s about farts, draw a butt for god’s sakes."

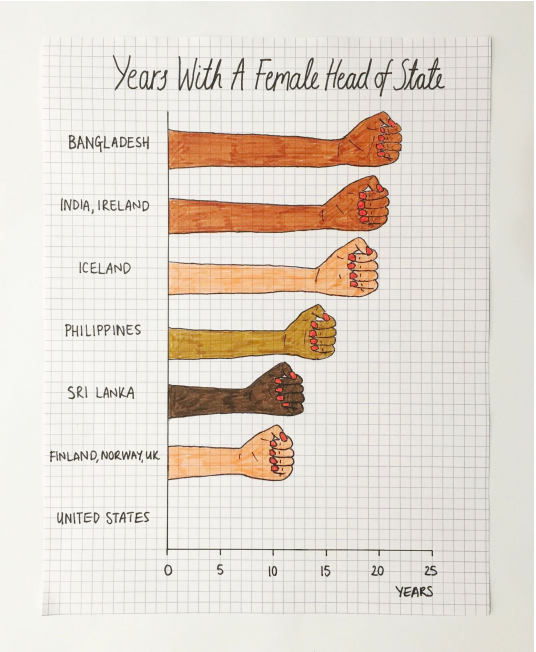
Mona Chalabi, journalist
The Guardian



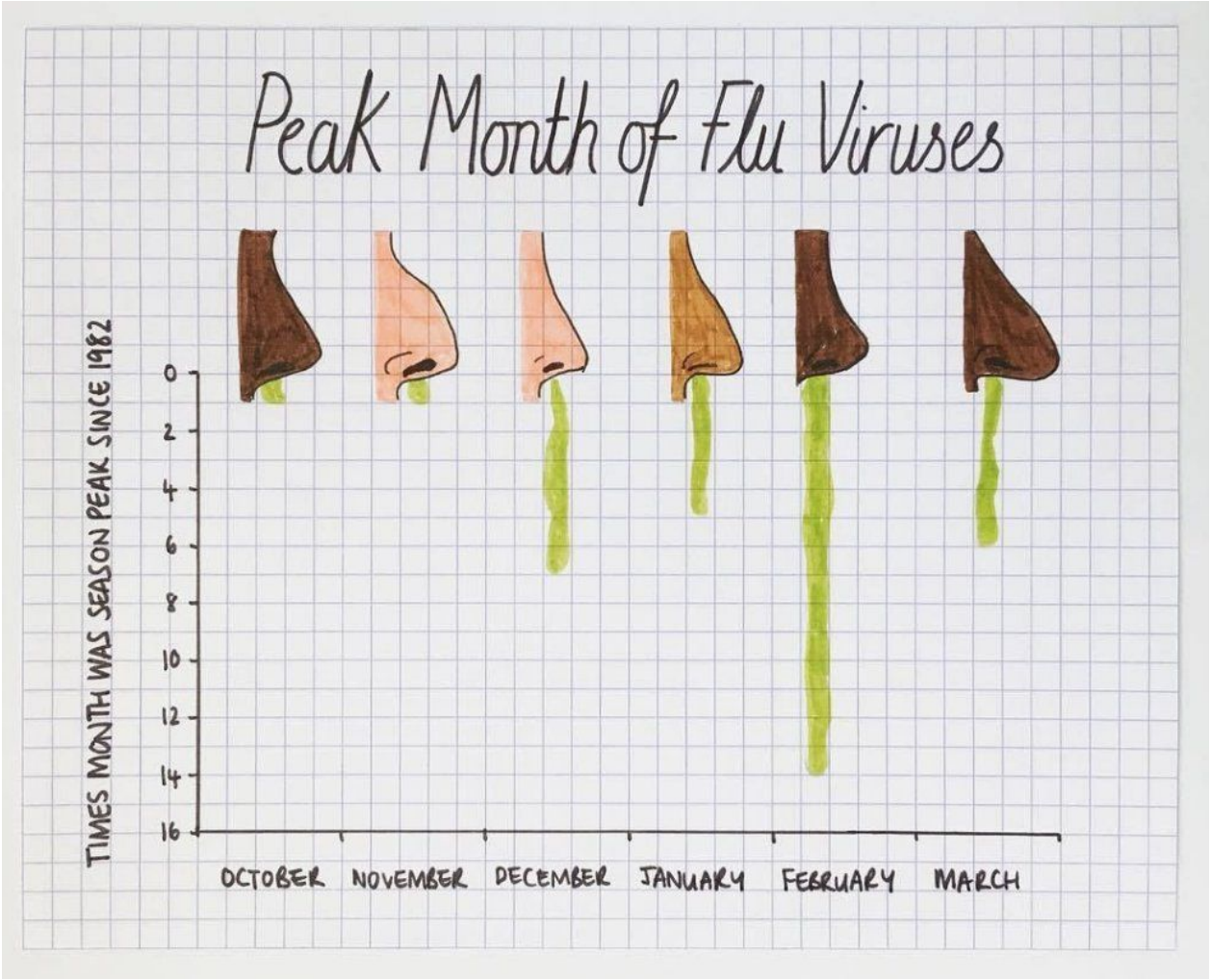
“If it’s about farts, draw a butt for god’s sakes."

Mona Chalabi, journalist
The Guardian

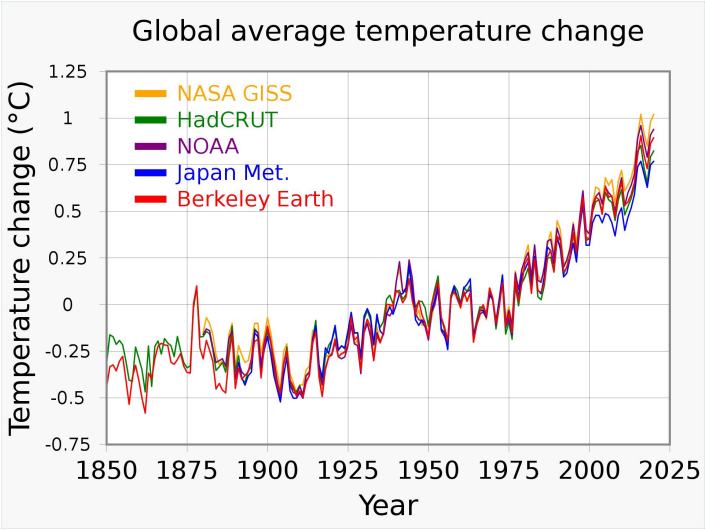
For most charts, the only clues you have about the subject are the labels. But I think that’s wrong. If it’s about sexual harassment, draw some pervs. I think data purists think that charts should be neutral, that they should elicit no emotion but I think that’s a mistaken understanding of objectivity. There is no such thing in the social sciences — if you look at a chart of mine about the rise of the KKK and think it looks creepy, well, good.

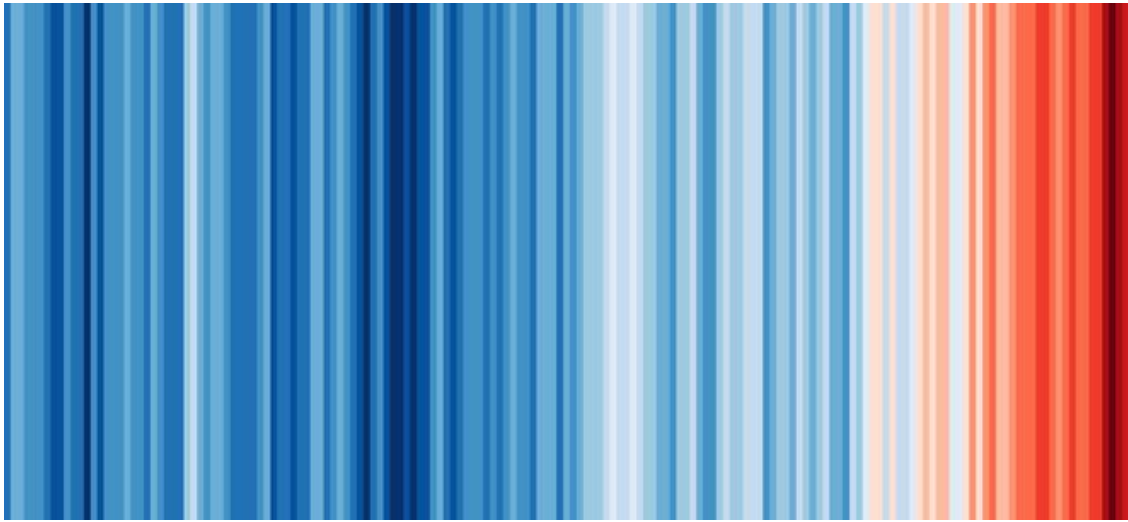


https://www.itsnicethat.com/articles/mona-chalabi-illustration-internationalwomensday-080318


Peak Month of Flu Viruses
TIMES MONTH WAS SEASON PEAK SINCE 1982
0
2
4
6
8
10
12
14
16
OCTOBER
NOVEMBER
DECEMBER
JANUARY
FEBRUARY
MARCH

Exploration vs. communication



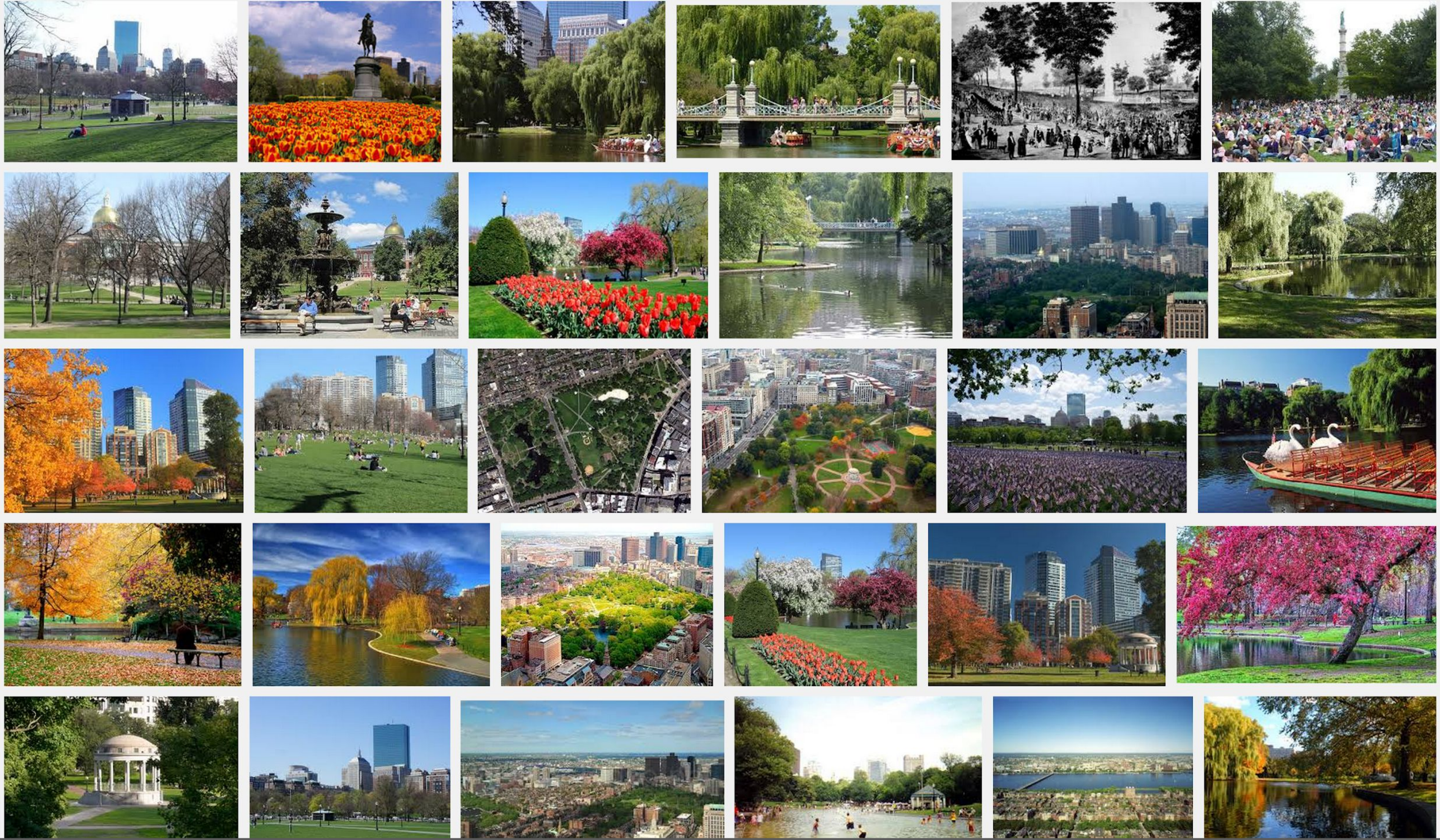


See Wikipedia article on “Warming Stripes”. Original stripe image due to Ed Hawkins.

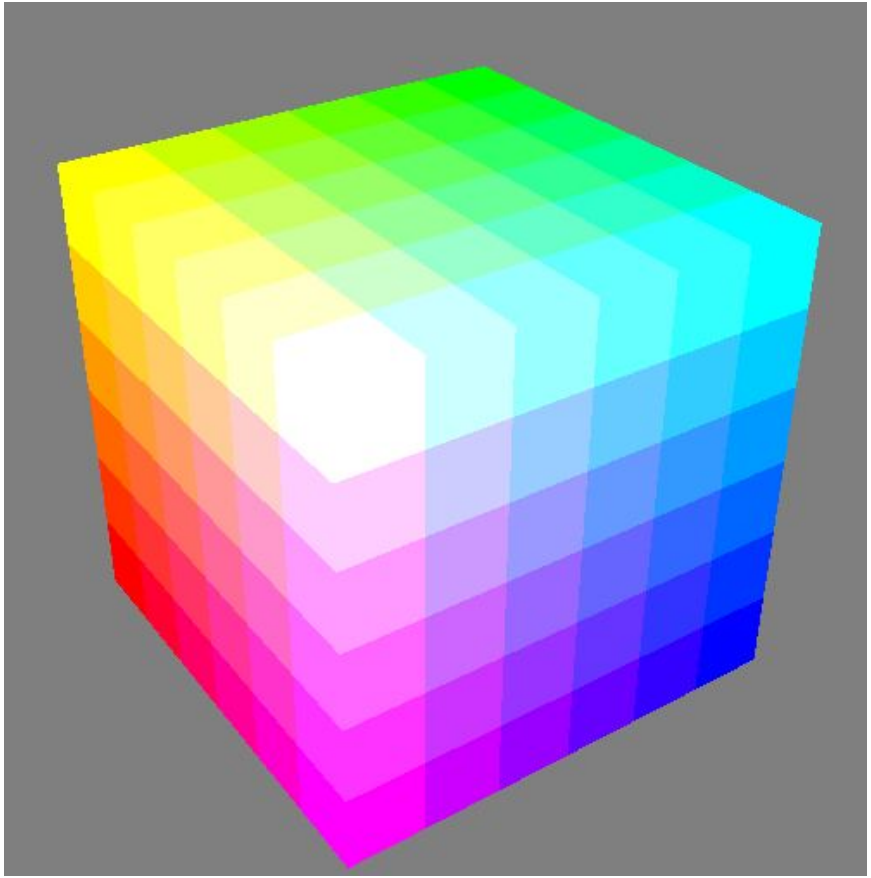
Data visualization and art

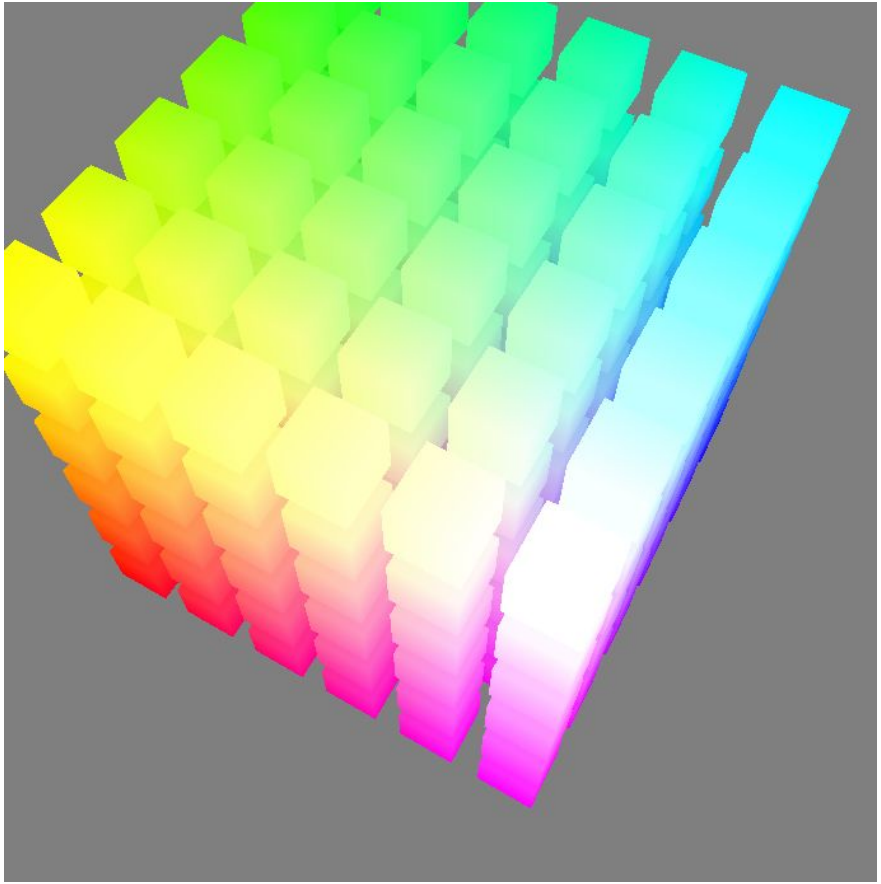
Flickr Flow

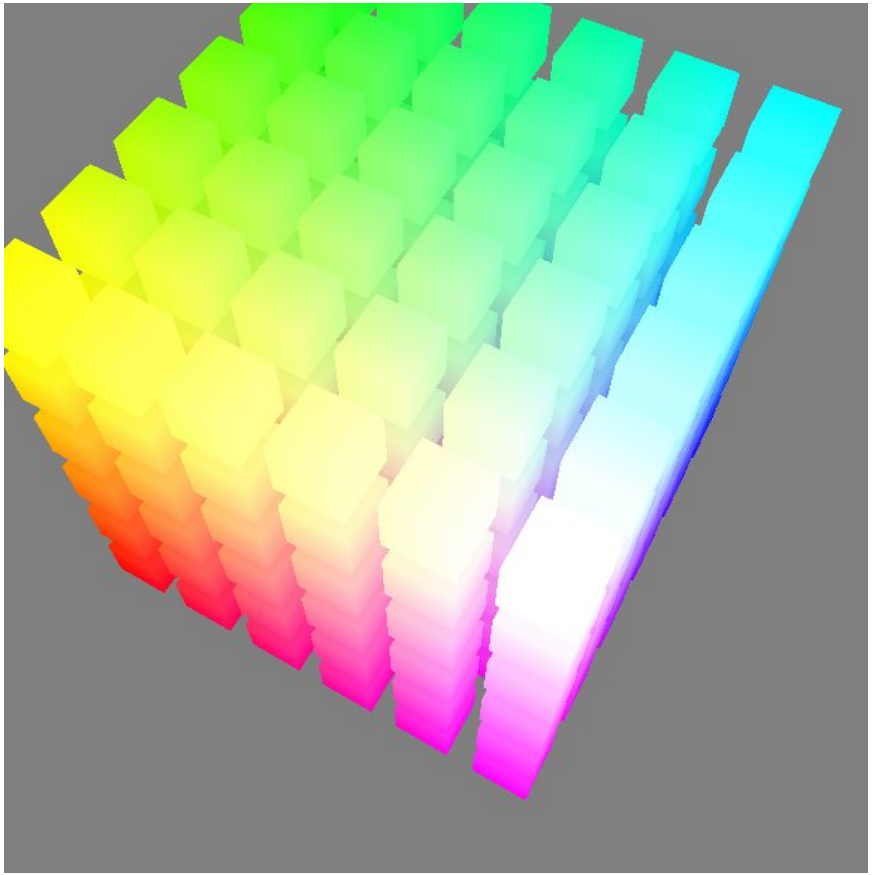
How would you visualize Boston?

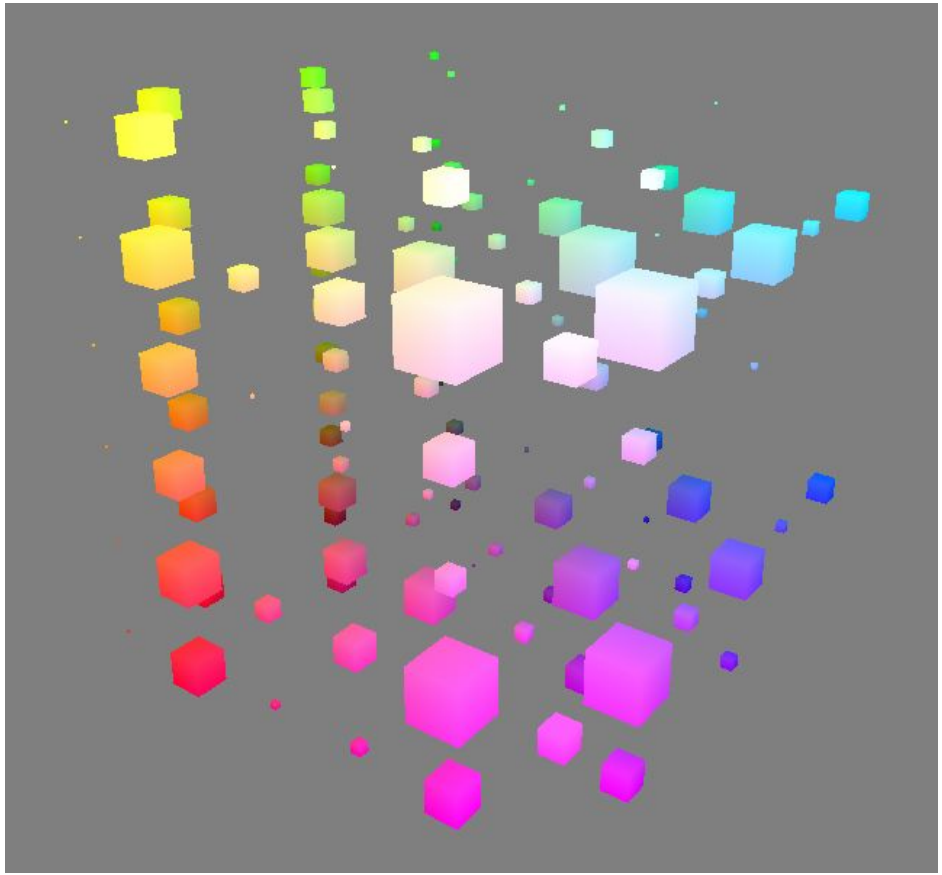


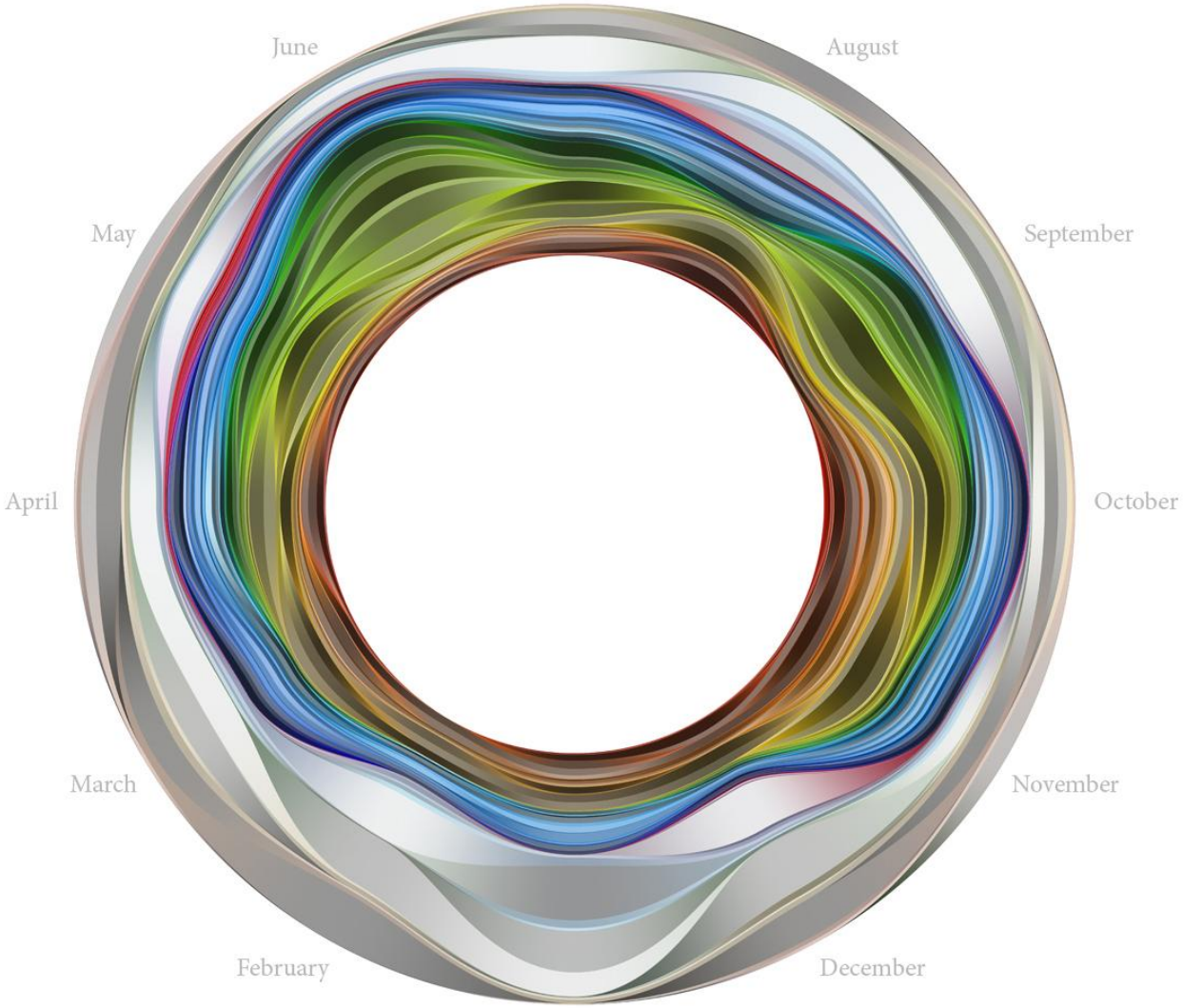
Boston Common public photos under CC license



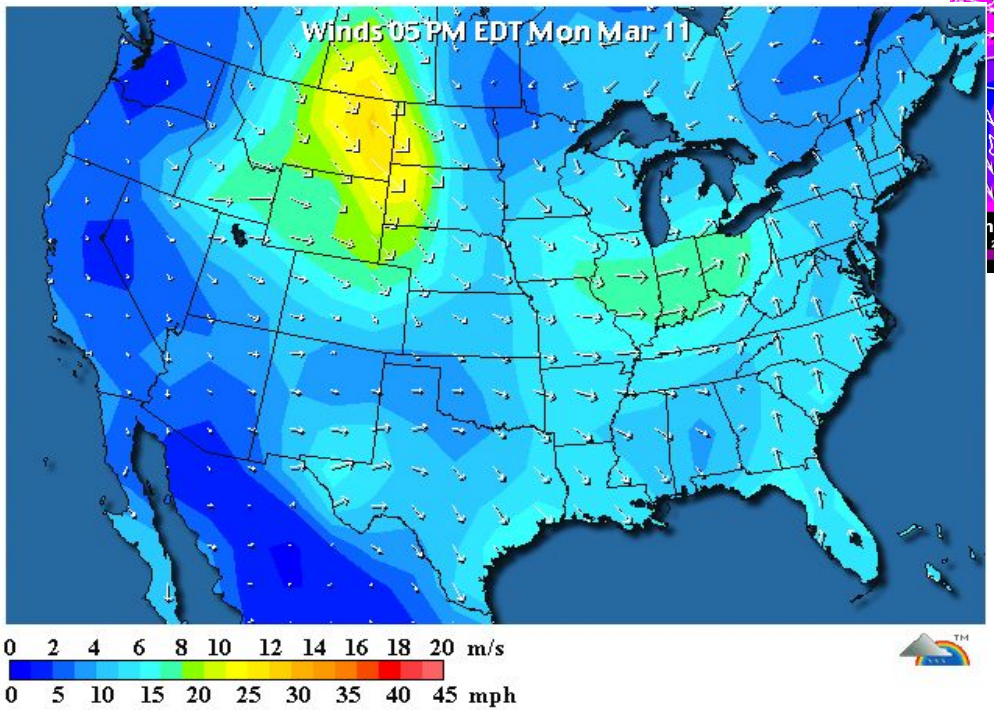


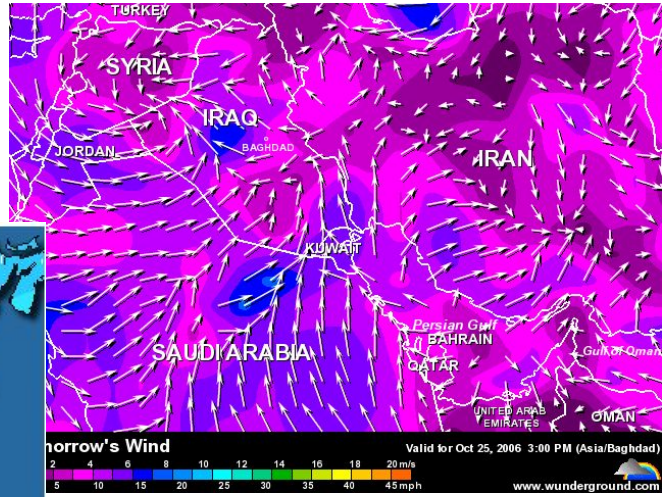


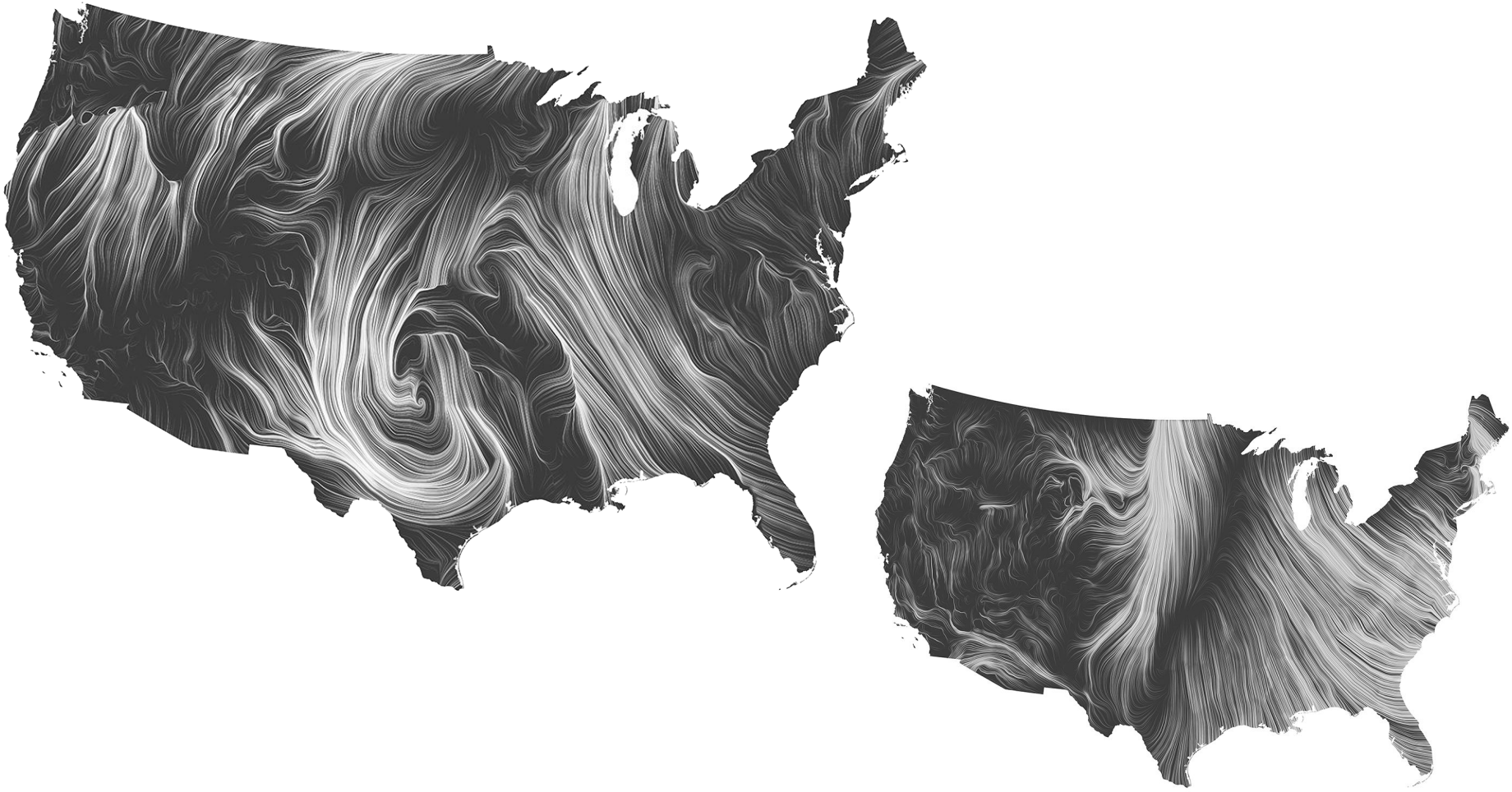



June
August
May
September
April
October
March
November
February
December

Seeing the wind


Winds 05 PM EDT Mon Mar 11
0 2 4 6 8 10 12 14 16 18 20 m/s
0 5 10 15 20 25 30 35 40 45 mph


TURKEY
SYRIA
IRAQ
BAGHDAD
JORDAN
IRAN
KUWAIT
Persian Gulf
BAHRAIN
SAUDI ARABIA
QATAR
UNITED ARAB EMIRATES
OMAN
Gulf of Oman
Valid for Oct 25, 2006 3:00 PM (Asia/Baghdad)
2 4 6 8 10 12 14 16 18 20 m/s
5 10 15 20 25 30 35 40 45 mph
www.wunderground.com



I am a farmer in Oklahoma and just found this map a few days ago. I find it extremely useful for planning when I spray and do other field work like making hay.

Personally, I love to use it to study wind patterns for nocturnal bird migration. It's so much more readable than standard weather service maps.

Mesmerizing. I used it to teach my kids about wind patterns yesterday and we forecasted tornados in the midwest given the large low pressure area and cold/warm fronts converging there. Today I read that texas indeed had tornados yesterday and the kids are awestruck that they figured that out.

I'm a pilot and you have with one simple map utterly destroyed what the government provides to pilots for pre-flying planning. [...] your Wind Map could be an amazing resource for pilots.

I am a powered paragliding pilot and I am going to start using the map to see where the winds are exactly.

Disclaimer

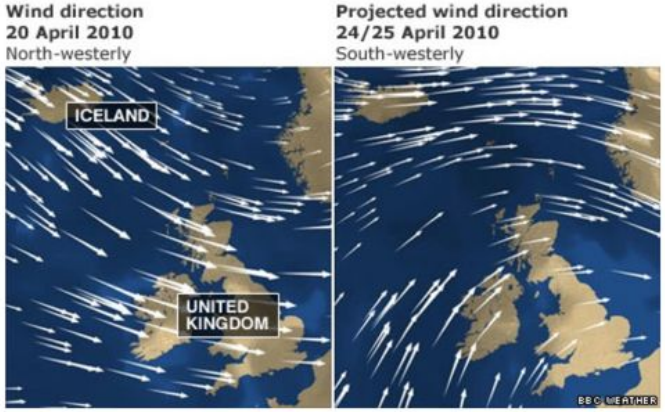
Please do not use the map or its data to fly a plane, sail a boat, or fight wildfires :-)

Disclaimer

Please do not use the map or its data to fly a plane, sail a boat, or fight wildfires :-)

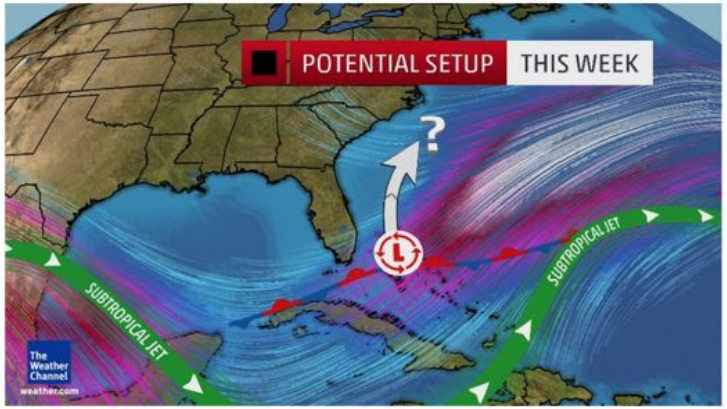
Response

We appreciate your disclaimer that it is inappropriate to use your map to fight wildfires, but respect the power of this visualization in promoting the prevention of wildfires



BBC Weather

CNN Weather



Weather Channel

weather.com

MoMA

Never Alone: Video Games and Other Interactive Design, 2022
Triennale di Milano, 2019
This is for Everyone, 2016
A Collection of Ideas, 2015
Applied Design, 2013

Visualization…

Not just for numbers
Not just for individuals
Not just for experts

A broad and expressive medium that invites a variety of stakeholders into the world of data.

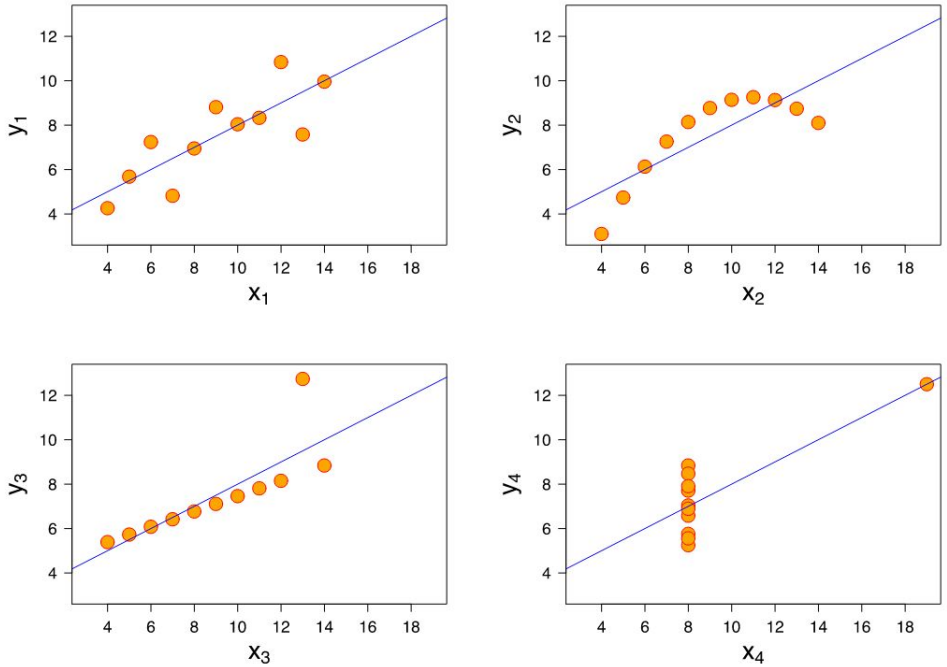
Thank you!

Fernanda Viégas

viegasf@

Anscombe's quartet

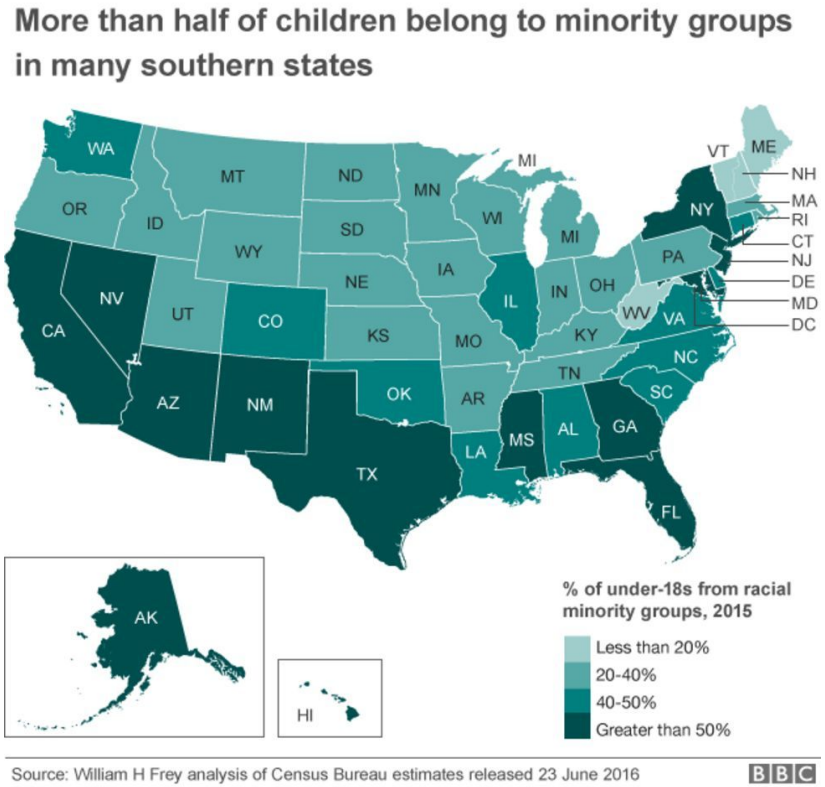
- nearly identical summary statistics
- very different distributions
- appear very different when graphed



Created in 1973 by the statistician Francis Anscombe to demonstrate the importance of graphing data when analyzing it, and the effect of outliers on statistical properties.

Communication vs. exploration

Communication



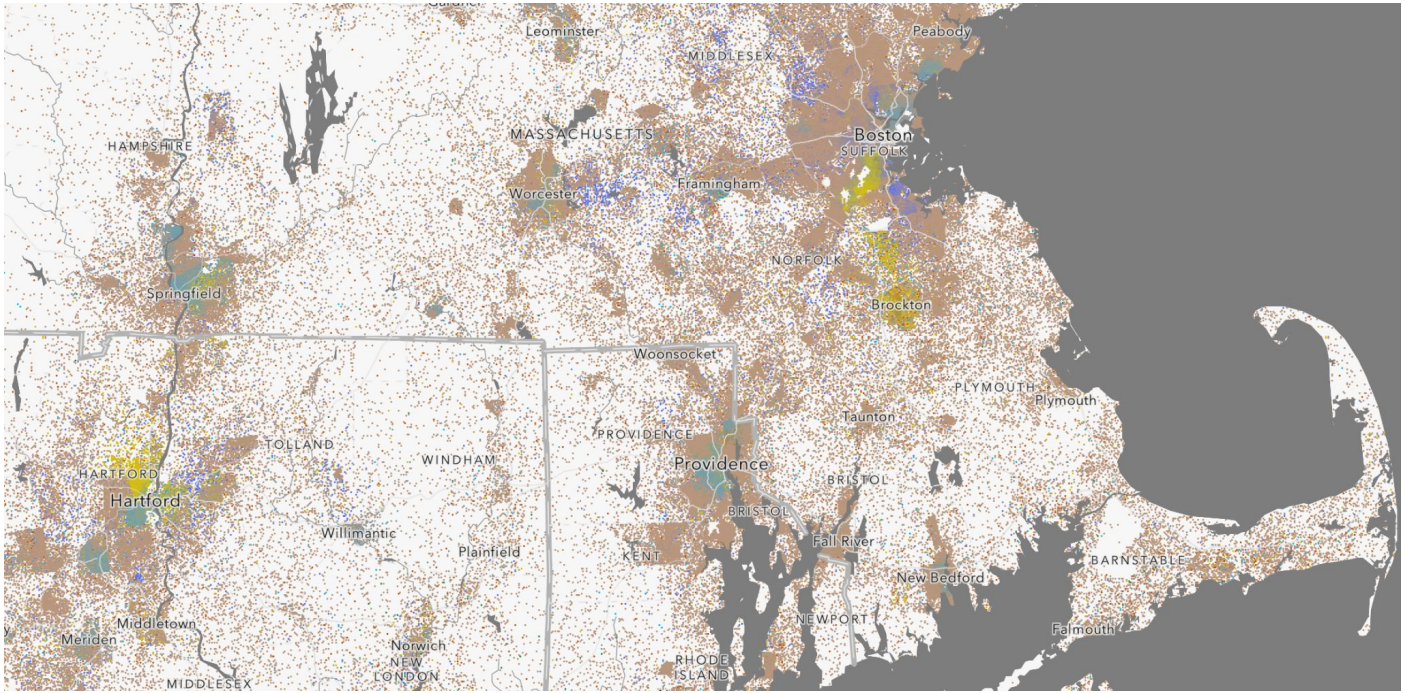
US state-level chart;
BBC News 2017

High-level overview, clear message supporting a news story.

“How young Americans are set to change the US forever”

https://www.bbc.com/news/world-us-canada-40461666

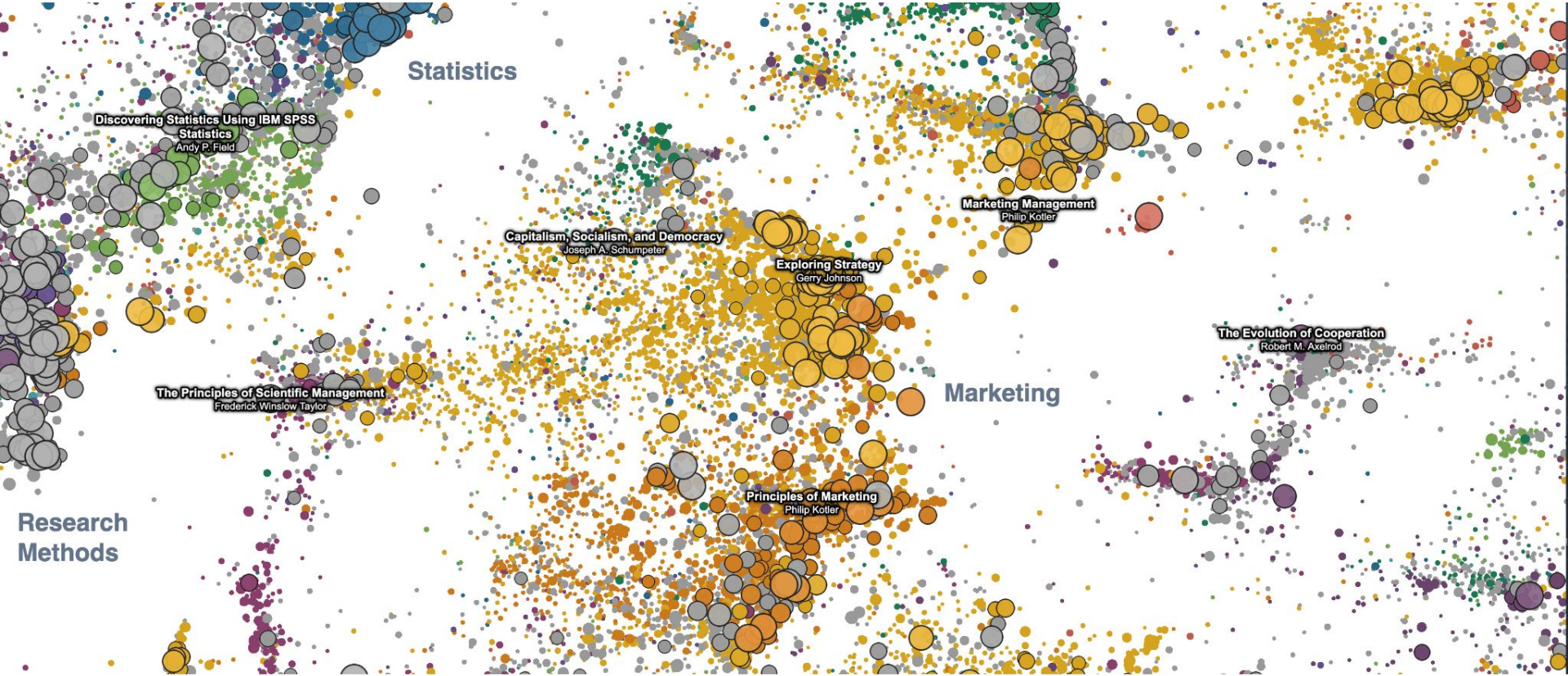
Exploration



https://www.arcgis.com/home/item.html?id=30d2e10d4d694b3eb4dc4d2e58dbb5a5
or Google for [arcgis racial dot map]

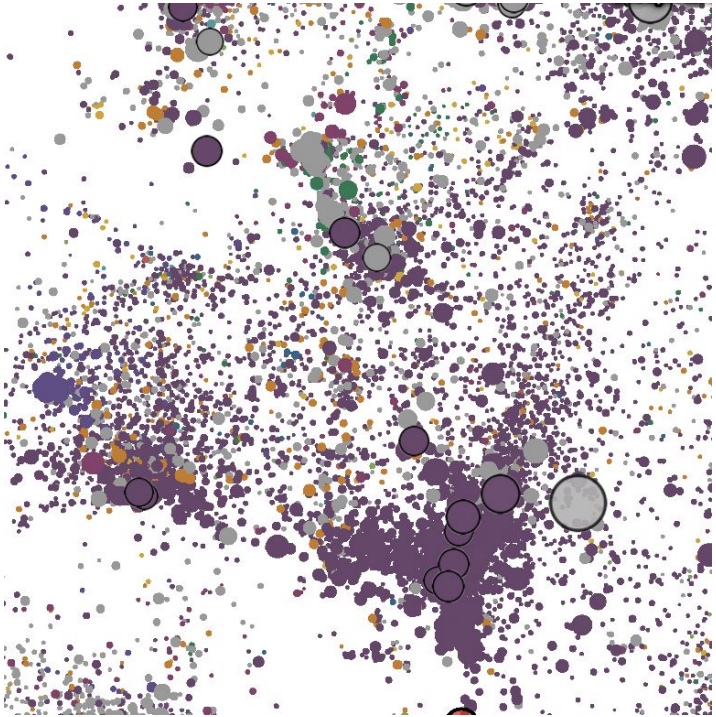
From machine learning to the broader world…
Example: OpenSyllabus

1,138,841 most frequently assigned texts in the Open Syllabus corpus, a database of 7,292,573 college course syllabi. https://galaxy.opensyllabus.org/ (McClure et al.)

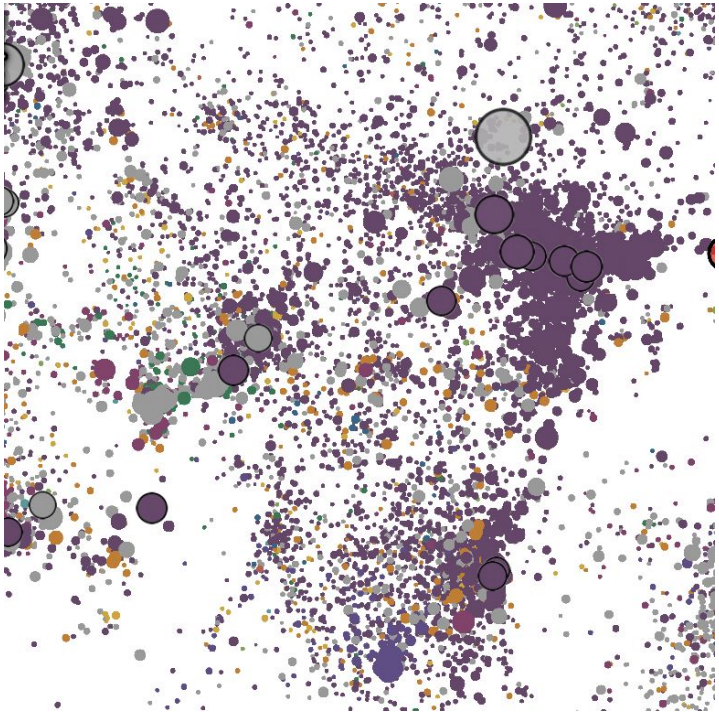


A word of caution:
Be careful in interpreting these plots…

Axes mean nothing! All that matters is relative position



View 1



View 2: same as #1, but rotated 90 degrees