Illinois Institute of Technology, Institute of Design (IIT-ID)
Latham Lecture | Chicago | April 8, 2020

AI + Computing for Design: Data Visualization of Fortune 500 Logos

Knut M. Synstad
@knutsynstad

Hello, I’m Knut Synstad.
I design, and I code.

I design products, services, and systems
at Dubberly Design Office.

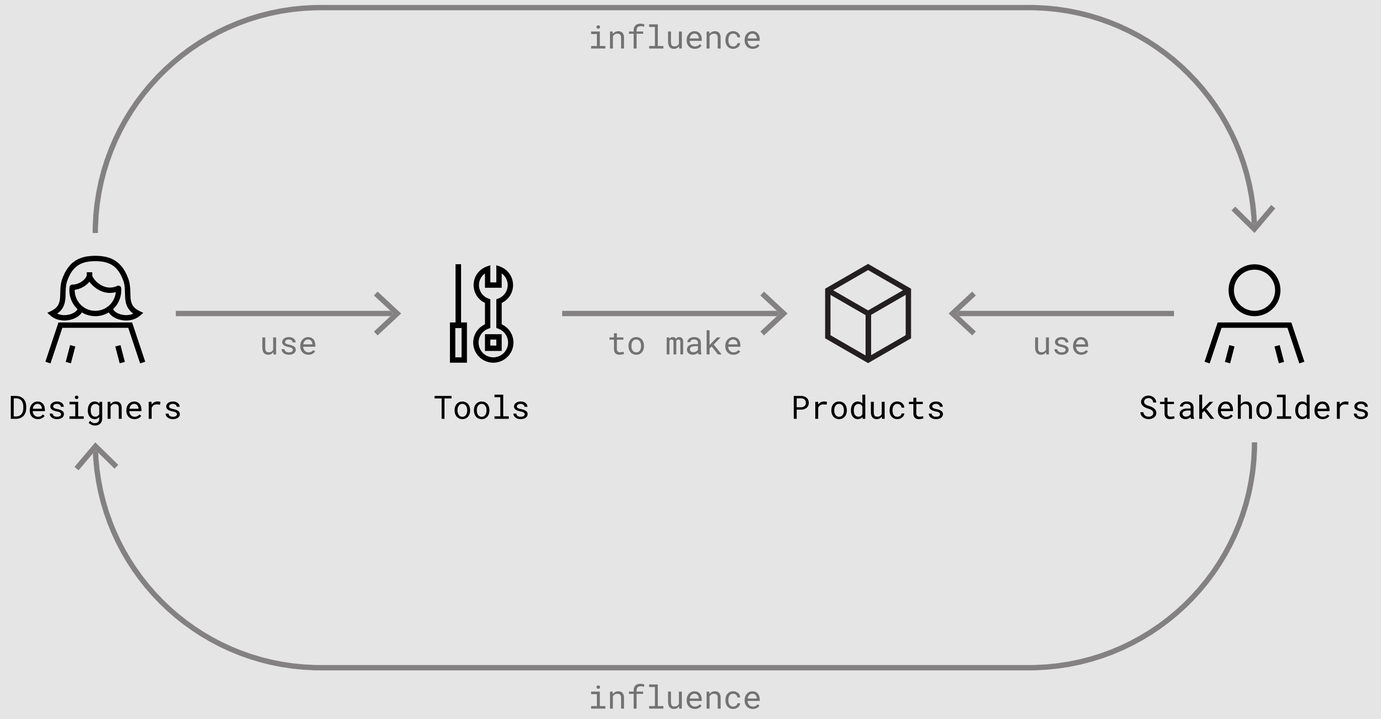
Currently exploring applications of
AI, ML, and computation in design.

Former electronics officer in
the Royal Norwegian Navy.

And former Graphic Designer.

Design practice is continually evolving.

- What we design
- How we design
- Who designs
- Why we design
- Where we design



After
Hugh Dubberly

For more on the evolution of design, see
http://presentations.dubberly.com/IDEEC_SJSU.pdf

In the last 50 years, design's biggest change has been computing.

Computing as Tool:
Augmenting the design process.

1987

1990

1999

(PageMager 1985)

2010

2016

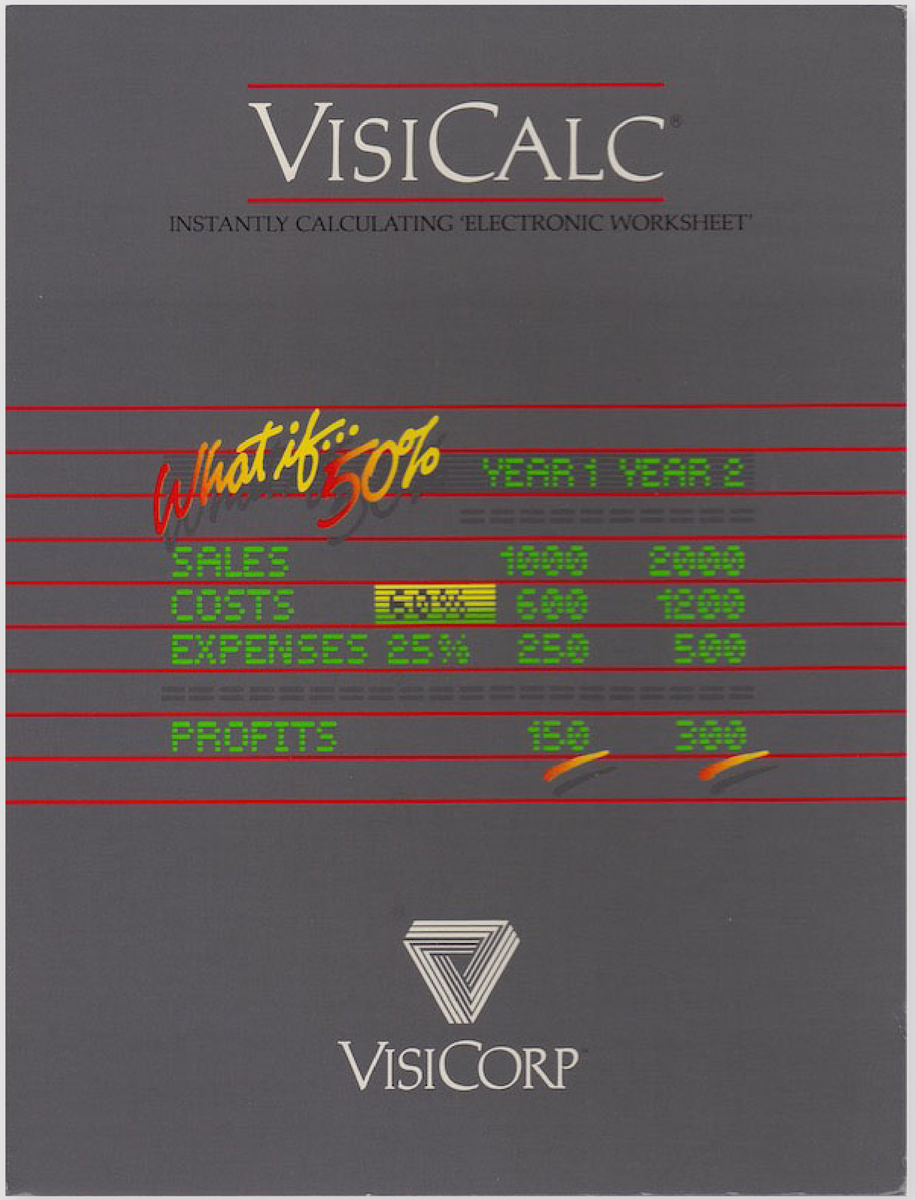
Computing as Medium:
For sharing information.

Computing as Material:
To be shaped into products.

“VisiCalc represented a new idea of a way to use a computer and a new way of thinking about the world.

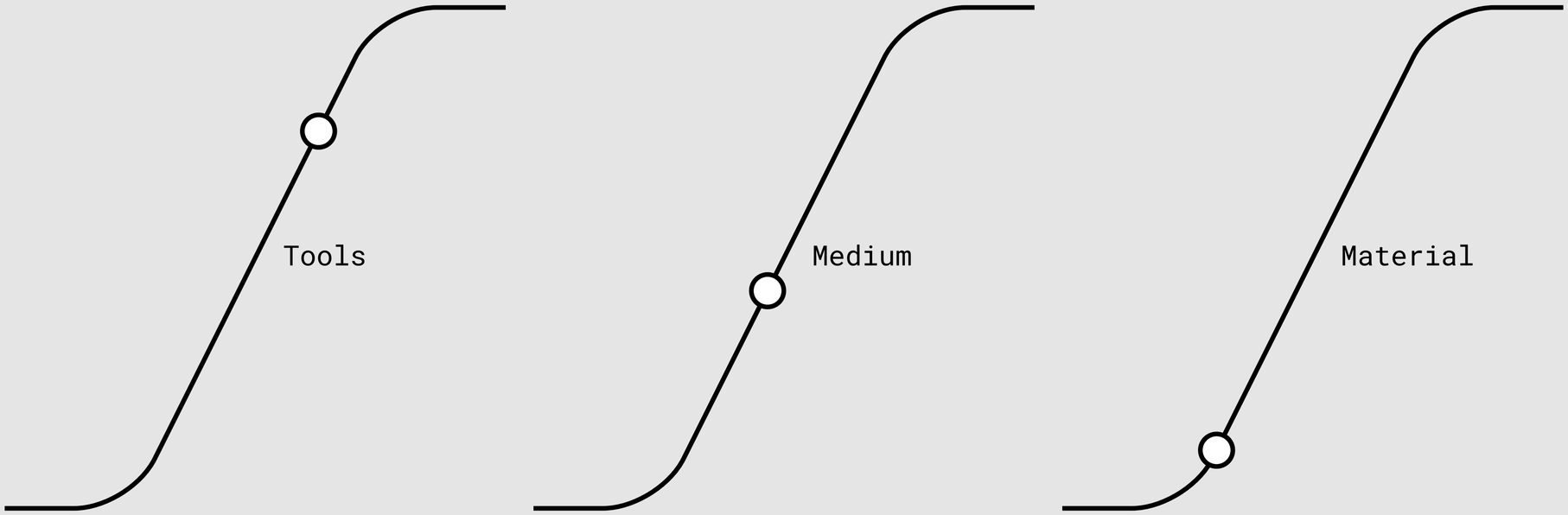
Where conventional programming was thought of as a sequence of steps, this new thing was no longer sequential in effect: When you made a change in one place, all other things changed instantly and automatically.”

—**Ted Nelson**



Initial release
1979 (41 years ago)

Each of these “digital transformations” is at a different stage in the development process.



PCs
1980–

Mobile Devices

Internet
1995–

Internet of Things

AI, Data, & Algorithms
2015–

Cloud Computing

How are designers using computing (or computation) as a medium today?

There are many exciting examples …

A visual introduction to machine learning

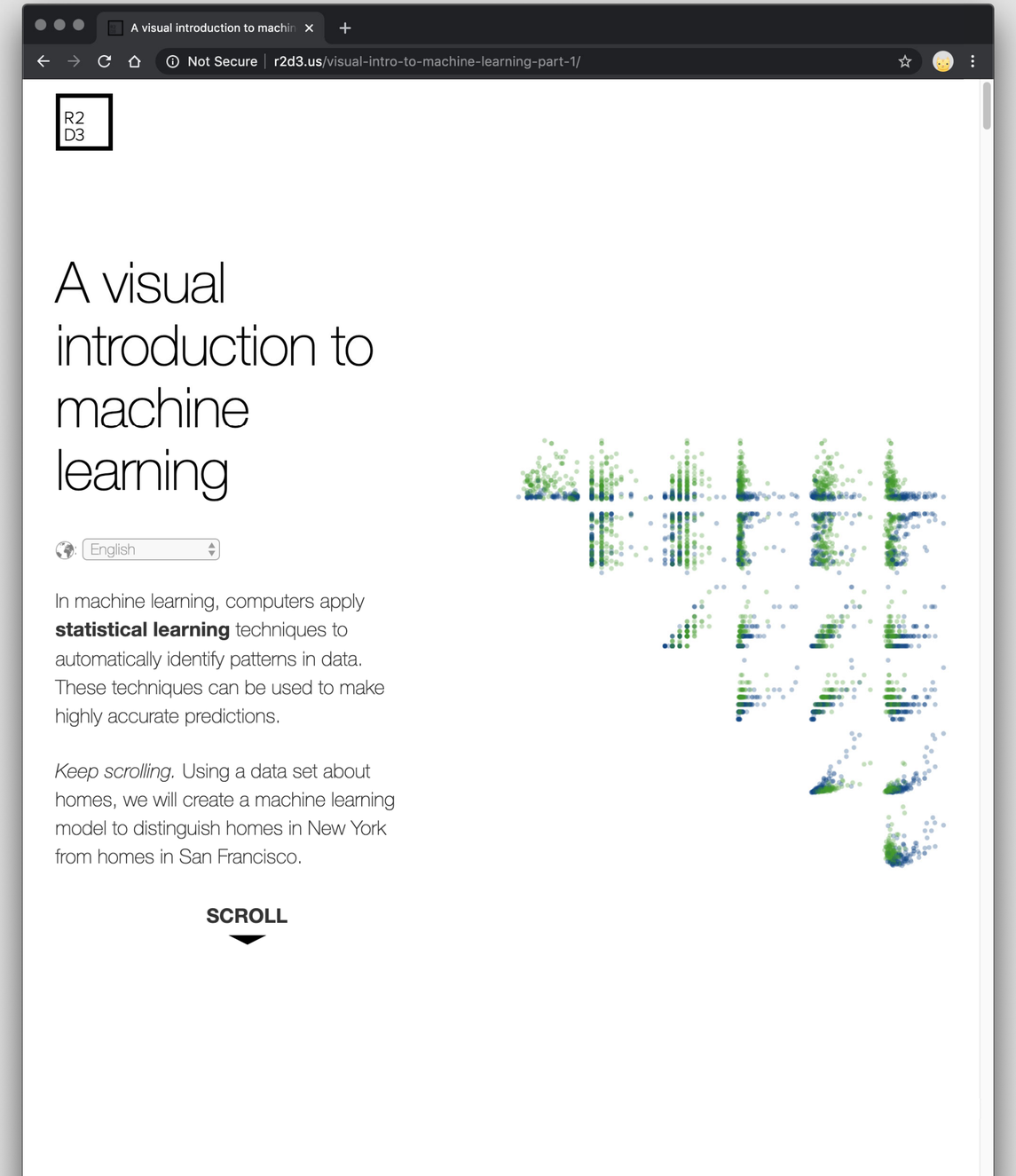
By Stephanie Yee and Tony Chu

Machine learning is …

“the science of getting computers to act without being explicitly programmed.”
—**Stanford University**

Year
2015

Address
r2d3.us/visual-intro-to-machine-learning-part-1/



How to Use t-SNE Effectively

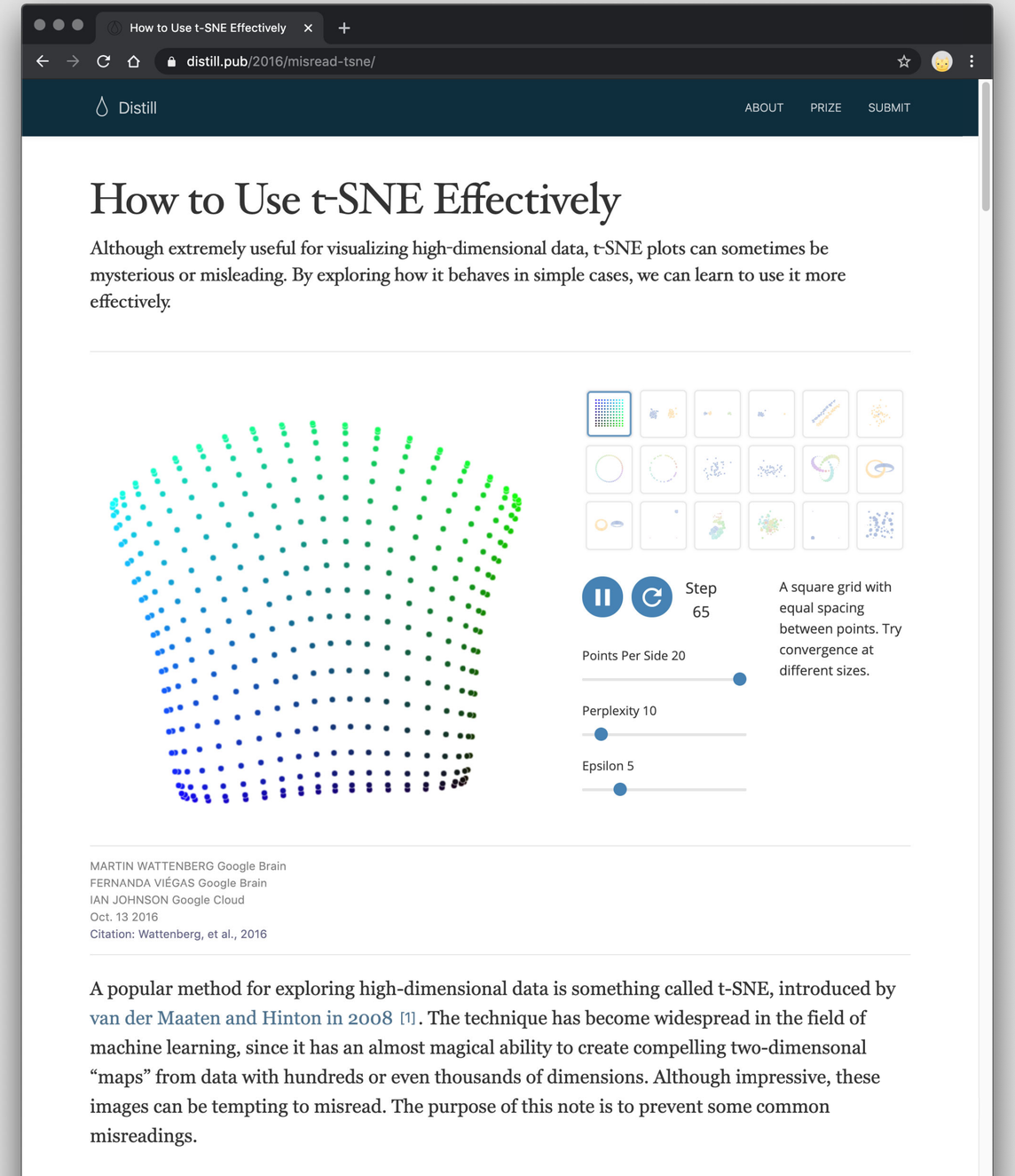
By Martin Wattenberg, Fernanda Viégas, and Ian Johnson

t-Distributed Stochastic Neighbor Embedding (t-SNE) is …

“a technique for dimensionality reduction that is particularly well suited for the visualization of high-dimensional datasets.”
—**Laurens van der Maaten**

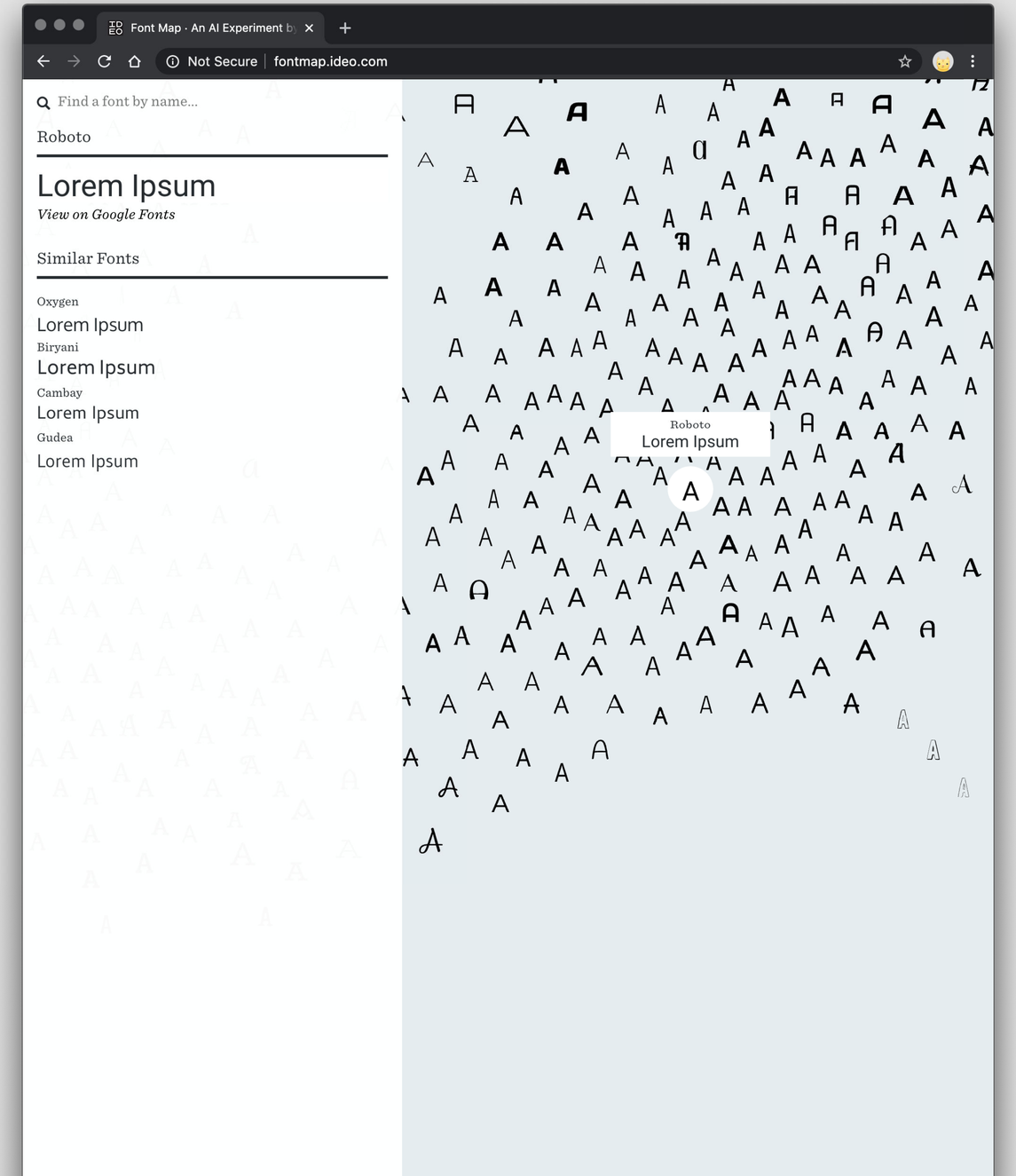
Year
2016

Address
distill.pub/2016/misread-tsne/



Font Map

By Kevin Ho (IDEO)



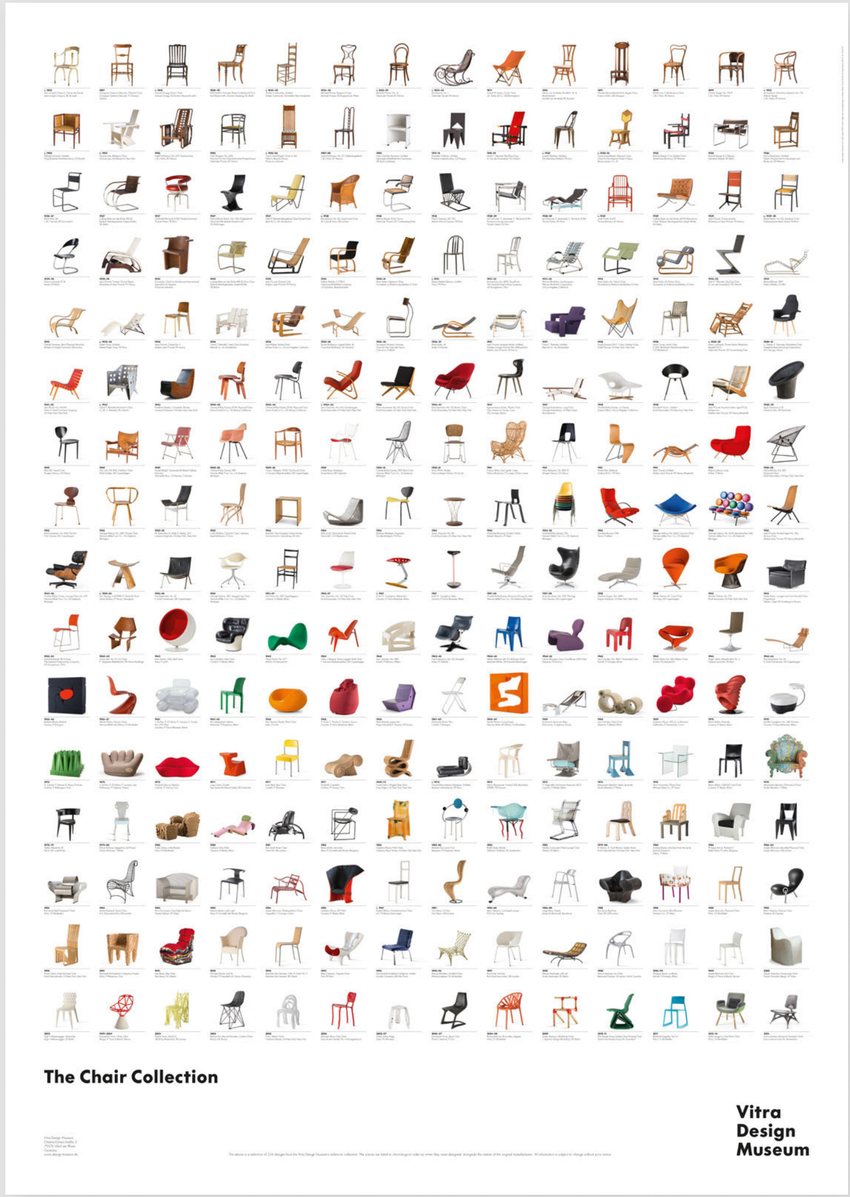
Year
2017

Address
fontmap.ideo.com/

The idea of a solution space goes back at least to Herbert Simon.

We’ve all seen examples;
yet representing a solution space by hand is a lot of work.

How might computing help?



Year
2018

Design
Schärer de Carli and BOROS

It made me wonder...

What would it take to organize the logos of the Fortune 500, by visual similarity?

Walmart
Berkshire Hathaway
ExxonMobil
McKesson
UnitedHealth Group
CVS
GM
AT&T
Ford
AmerisourceBergen
Amazon
GE
Verizon
CardinalHealth
Costco
Walgreens
Kroger
Chevron
Fannie Mae
JPMorgan Chase & Co
Express Scripts
The Home Depot
Boeing
Wells Fargo
Bank of America
Alphabet
Microsoft
Anthem
Citi
Comcast
IBM
State Farm
Phillips 66
Johnson & Johnson
P&G
Valero
Target
Freddie Mac
Lowe's
Dell
MetLife
Aetna
PepsiCo
ADM
UPS
Intel
Prudential
Albertsons
United Technologies
Marathon
Disney
Humana
Pfizer
AIG
Sysco
FedEx
Hewlett Packard Enterprise
Cisco
HP
Dow
HCA
Coca-Cola
New York Life
Centene
American Airlines
Nationwide
Merck
Cigna
Delta
Best Buy
Honeywell
Caterpillar
Liberty Mutual Insurance
Morgan Stanley
MassMutual
Goldman Sachs
Energy Transfer
TIAA
Oracle
Tyson
United
Allstate
Publix
American Express
TJX
Nike
Exelon
General Dynamics
Rite Aid
Gilead
CHS
3M
Time Warner
Charter
Northwestern Mutual
Facebook
Travelers
Capital One
21st Century Fox
USAA
World Fuel Services
John Deere
Kraft Heinz
Tech Data
Avnet
Mondelēz International
Macy's
AbbVie
McDonald's
DuPont
Northrop Grumman
ConocoPhillips
Raytheon
Tesoro
Arrow
Qualcomm
Progressive
Duke Energy
Enterprise Products Partners L.P.
Amgen
US Foods
US Bank
Aflac
Sears
Dollar General
AutoNation
CHS Community Health Systems
Starbucks
Lilly
International Paper
Tenet Healthcare
Abbott
Dollar Tree
Whirlpool
Southwest
Emerson
Staples
Plains All American
Penske
Union Pacific
Danaher
Southern Company
Manpower
Bristol-Myers Squibb
Altria
Fluor
Kohl's
Lear
Jabil
The Hartford
Thermo Fisher Scientific
Kimberly-Clark
Molina Healthcare
PG&E
Supervalu
Cummins
CenturyLink
AECOM
Xerox
Marriott
Paccar
Publix
PNC
American Electric Power
ICAHN Enterprises L.P.
Nucor
NextEra Energy
PFG
PBF Energy
Halliburton
CarMax
Freeport-McMoRan
Whole Foods Market
BNY Mellon
Gap
OmnicomGroup
GPC
DaVita
Colgate-Palmolive
PPG
Goodyear
Synchrony Financial
Dish
Visa
Nordstrom
INTL FCStone
WestRock
XPO Logistics
Aramark
CBS
AES
WellCare
FirstEnergy
ConAgra Foods
Synnex
Pep Boys
Textron
Waste Management
ITW
Office Depot
Monsanto
Cognizant
Texas Instruments
Lincoln Financial Group
Newell Brands
Land O'Lakes
Marsh & McLennan Companies
Ecolab
C.H. Robinson
Loews Corporation
CBRE
Kinder Morgan
Kellogg's
Western Digital
Guardian
Ross Dress for Less
L Brands
JCPenney
Farmers Insurance
RAI
Viacom
BD
Micron
Principal
Arconic
NRG
Vf
Devon
Dr Pepper
Bed Bath & Beyond
ConEdison
Edison International
Sherwin-Williams
NGL Energy Partners LP
Dominion Energy
Ameriprise Financial
ADP
Hilton
First Data
Henry Schein
Toys R Us
BB&T
RGA
Core-Mark
Biogen
Sands
Stanley Black & Decker
Parker
Stryker
Estée Lauder
Celgene
BlackRock
Xcel Energy
CSX
Unum
Jacobs
Lennar
Entergy
PayPal
Applied Materials
Voya
Mastercard
Priceline.com
Liberty Interactive Corporation
AutoZone
State Street
DTE Energy
Technologies
HollyFrontier
Praxair
UHS
Discover
Occidental Petroleum
US Steel
Sempra Energy
Baxter
Grainger
Autoliv
Norfolk Southern
Baker Hughes
Ally
Sonic Automotive
Owens & Minor
Huntsman
LabCorp
Murphy USA
Advance Auto Parts
Fidelity National Information Services
Air Products
Hormel
Hertz
MGM Resorts
Corning
Republic Services
Alcoa
FIS
Pacific Life
SunTrust
LKQ
BorgWarner
Ball
CST Brands
PSEG
Eastman
eBay
Mohawk
ONEOK
Frontier Communications
Netflix
American Family Insurance
Thrivent Financial
Expedia
Lithia Motors
Avis
Reliance Steel & Aluminum Co.
GameStop
Tenneco
O'Reilly Auto Parts
Kiewit
unfi
Salesforce
Boston Scientific
Newmont
Genworth
Live Nation Entertainment
Veritiv
News Corp
Crown Holdings
Global
PVH
Level 3
Navistar
Univar
Campbell's
Dick's Sporting Goods
Weyerhaeuser
Mutual of Omaha
Chesapeake Energy
Anadarko
IPG
Smucker's
Rivet Dynamics, Inc.
Foot Locker
Western Refining
SpartanNash
Dean Foods
Zimmer Biomet
PulteGroup
Quanta Services
Reg Resources
Charles Schwab
Eversource
Anixter
EMCOR
Assurant
CenterPoint Energy
Harris
HD Supply
ppl
Quest Diagnostics
Williams
WEC Energy Group
Hershey
AGCO
Ralph Lauren
Masco
Wesco Distribution
LifePoint Health
National Oilwell Varco
Kindred Healthcare
Mosaic
Alliance Data
CSC
Huntington Ingalls Industries
leidos
Erie Insurance
Tesla
ascena
Darden
Harman
NVIDIA
RRD
Fifth Third Bank
Quintiles IMS
JLL
Dover
Spirit AeroSystems
Ryder
AAR
Tractor Supply Co.
Sealed Air
Auto-Owners Insurance
YumChina
Calpine
Owens-Illinois
Targa
jetBlue
Edward Jones
Franklin Templeton Investments
Activision Blizzard
J.B. Hunt
Constellation Brands
NCR
Asbury Automotive Group
American Financial Group
Discovery Communications
Berry Plastics
Sanmina
CalAtlantic Homes
Dr Pepper
Dillard's
HRG Group
CMS Energy
Graybar
Builders FirstSource
Yum!
Casey's General Stores
Amphenol
Oshkosh
iHeartMedia
TreeHouse
Alleghany
Expeditors
Avery Dennison
Ameren
Hanes
Motorola Solutions
St. Jude Medical
Harley-Davidson
Regions
ICE
Alaska
Lam Research
AK Steel
Rockwell Automation
Adobe
Avon
Terex
NVR
Dana
Realogy
American Tower
PCA
Citizens Financial Group, Inc.
United Rentals
The Clorox Company
Genesis
M&T Bank
Ingredion
UGI Corporation
Owens Corning
S&P Global
Markel
Wyndham Worldwide
Arthur J. Gallagher & Co.
Burlington
First American Financial Corporation
Symantec
Patterson Companies
Olin
NetApp
Raymond James
TA
fiserv.
Host Hotels & Resorts
Insight
Mattel
AmTrust Financial
Cincinnati Financial Corporation
Simon
Western Union
KeyCorp
Delek US
Booz | Allen | Hamilton
Chemours
Western & Southern Financial Group
Celanese
Windstream
Seaboard
Essendant
Apache
Airgas
Kelly Services
Liberty
Rockwell Collins
Robert Half
ch2m
Big Lots
Michaels
Toll Brothers
Yahoo!
Vistra Energy
ABM

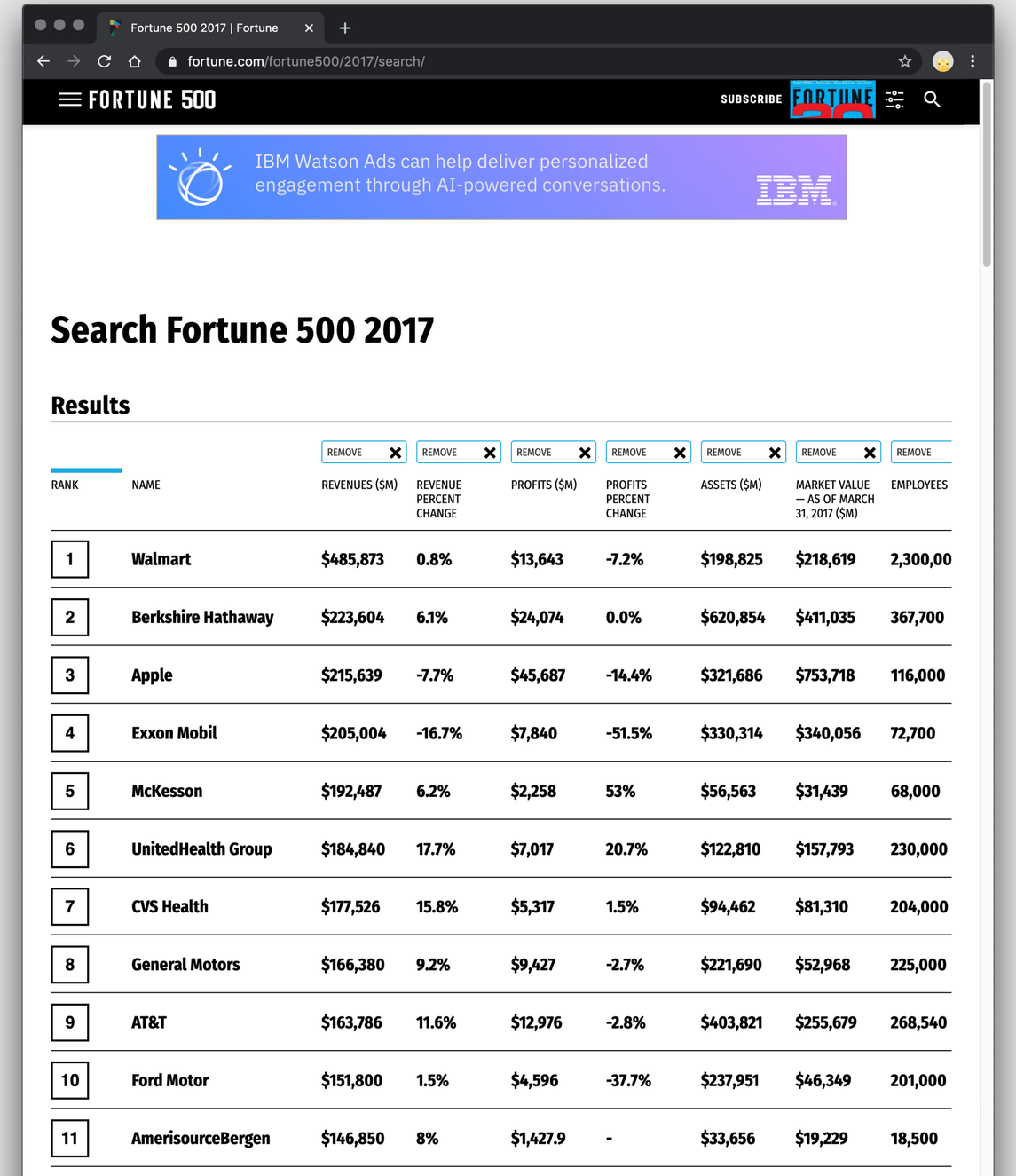
Turns out, a lot.
A sequence of data transformations are needed.

1. Data wrangling
2. Feature detection
3. Feature selection
4. Distance matrix
5. Dimensionality reduction
6. Linear assignment
7. Data visualization

Let’s go through them step-by-step.

Every summer, Fortune Magazine publishes its Fortune 500 list online.

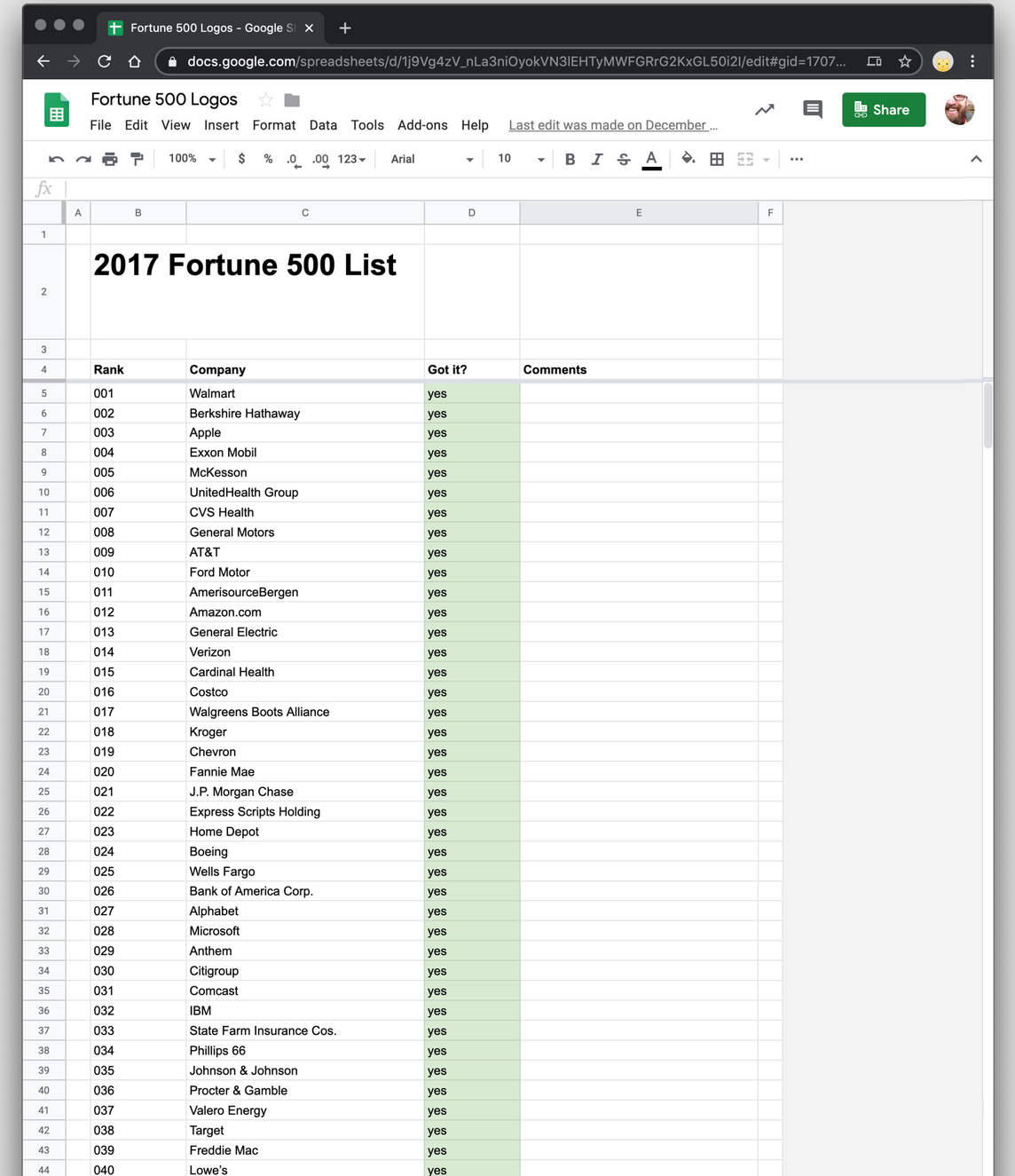
Address
fortune.com/fortune500/2017/



RANK	NAME	REVENUES ($M)	REVENUE PERCENT CHANGE	PROFITS ($M)	PROFITS PERCENT CHANGE	ASSETS ($M)	MARKET VALUE — AS OF MARCH 31, 2017 ($M)	EMPLOYEES
1	Walmart	$485,873	0.8%	$13,643	-7.2%	$198,825	$218,619	2,300,00
2	Berkshire Hathaway	$223,604	6.1%	$24,074	0.0%	$620,854	$411,035	367,700
3	Apple	$215,639	-7.7%	$45,687	-14.4%	$321,686	$753,718	116,000
4	Exxon Mobil	$205,004	-16.7%	$7,840	-51.5%	$330,314	$340,056	72,700
5	McKesson	$192,487	6.2%	$2,258	53%	$56,563	$31,439	68,000
6	UnitedHealth Group	$184,840	17.7%	$7,017	20.7%	$122,810	$157,793	230,000
7	CVS Health	$177,526	15.8%	$5,317	1.5%	$94,462	$81,310	204,000
8	General Motors	$166,380	9.2%	$9,427	-2.7%	$221,690	$52,968	225,000
9	AT&T	$163,786	11.6%	$12,976	-2.8%	$403,821	$255,679	268,540
10	Ford Motor	$151,800	1.5%	$4,596	-37.7%	$237,951	$46,349	201,000
11	AmerisourceBergen	$146,850	8%	$1,427.9	-	$33,656	$19,229	18,500

With a spreadsheet as a reference, vector logos were collected.

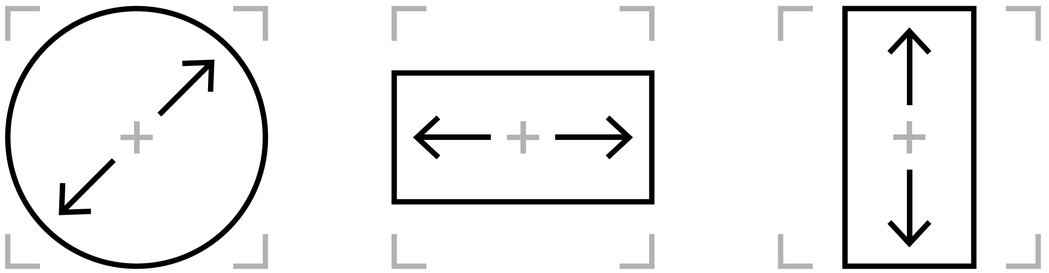
Many came form WikiMedia, some from corporate websites, and even annual reports.



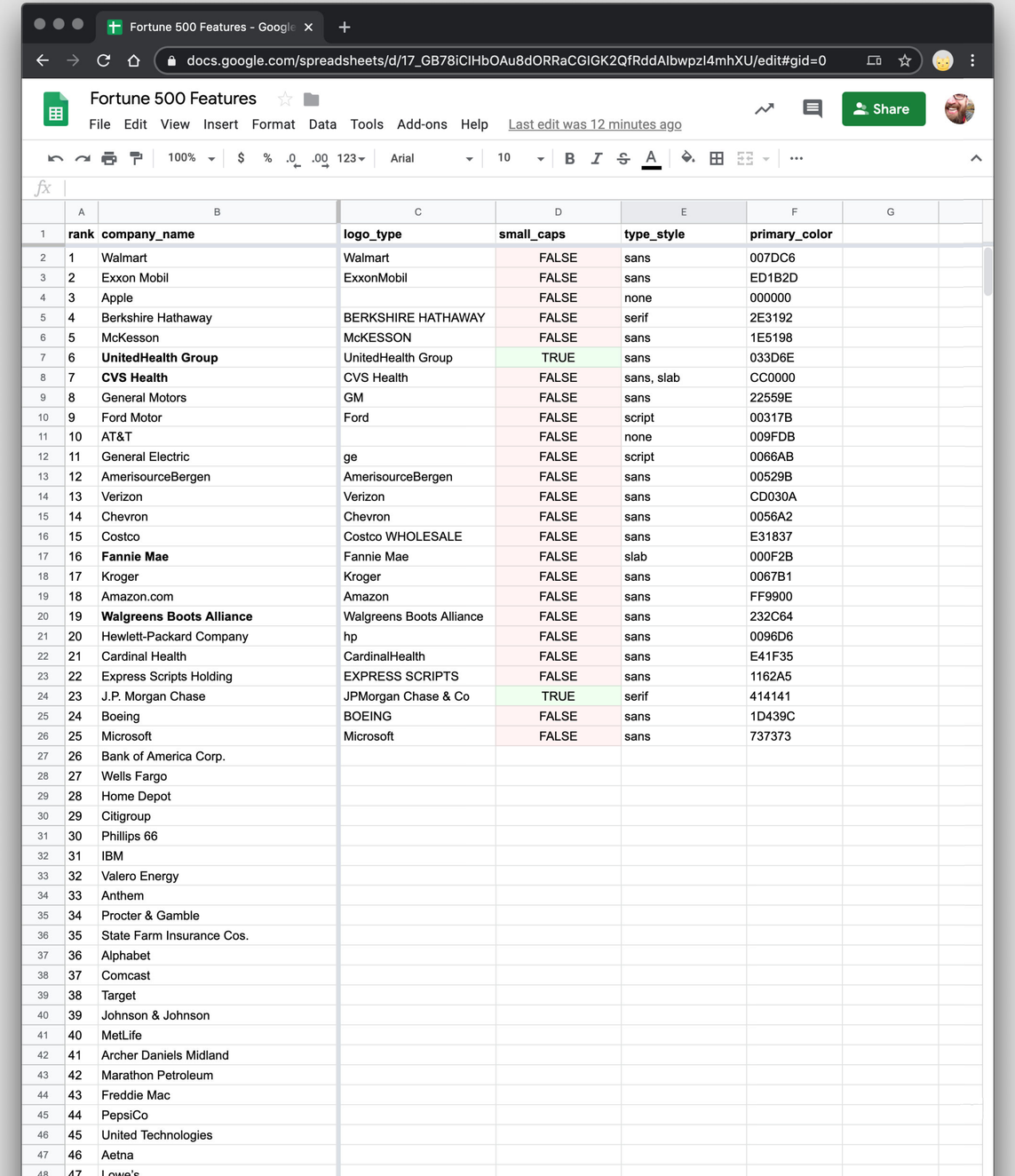
Rank	Company	Got it?	Comments
001	Walmart	yes	
002	Berkshire Hathaway	yes	
003	Apple	yes	
004	Exxon Mobil	yes	
005	McKesson	yes	
006	UnitedHealth Group	yes	
007	CVS Health	yes	
008	General Motors	yes	
009	AT&T	yes	
010	Ford Motor	yes	
011	AmerisourceBergen	yes	
012	Amazon.com	yes	
013	General Electric	yes	
014	Verizon	yes	
015	Cardinal Health	yes	
016	Costco	yes	
017	Walgreens Boots Alliance	yes	
018	Kroger	yes	
019	Chevron	yes	
020	Fannie Mae	yes	
021	J.P. Morgan Chase	yes	
022	Express Scripts Holding	yes	
023	Home Depot	yes	
024	Boeing	yes	
025	Wells Fargo	yes	
026	Bank of America Corp.	yes	
027	Alphabet	yes	
028	Microsoft	yes	
029	Anthem	yes	
030	Citigroup	yes	
031	Comcast	yes	
032	IBM	yes	
033	State Farm Insurance Cos.	yes	
034	Phillips 66	yes	
035	Johnson & Johnson	yes	
036	Procter & Gamble	yes	
037	Valero Energy	yes	
038	Target	yes	
039	Freddie Mac	yes	
040	Lowe's	yes	

Artwork came from multiple sources, and in multiple file formats, requiring some standardization.

- Crop to fit.
- Convert to a uniform file format (SVG)
- Remove non-essential tags (e.g., vendor tags)
- Generate uniformly sized bitmap images (PNG)



Intimidated by neural networks, I tried, and immediately failed, to describe the logos by hand …



	A	B	C	D	E	F	G
1	rank	company_name	logo_type	small_caps	type_style	primary_color	
2	1	Walmart	Walmart	FALSE	sans	007DC6	
3	2	Exxon Mobil	ExxonMobil	FALSE	sans	ED1B2D	
4	3	Apple		FALSE	none	000000	
5	4	Berkshire Hathaway	BERKSHIRE HATHAWAY	FALSE	serif	2E3192	
6	5	McKesson	McKESSON	FALSE	sans	1E5198	
7	6	**UnitedHealth Group**	UnitedHealth Group	TRUE	sans	033D6E	
8	7	**CVS Health**	CVS Health	FALSE	sans, slab	CC0000	
9	8	General Motors	GM	FALSE	sans	22559E	
10	9	Ford Motor	Ford	FALSE	script	00317B	
11	10	AT&T		FALSE	none	009FDB	
12	11	General Electric	ge	FALSE	script	0066AB	
13	12	AmerisourceBergen	AmerisourceBergen	FALSE	sans	00529B	
14	13	Verizon	Verizon	FALSE	sans	CD030A	
15	14	Chevron	Chevron	FALSE	sans	0056A2	
16	15	Costco	Costco WHOLESALE	FALSE	sans	E31837	
17	16	**Fannie Mae**	Fannie Mae	FALSE	slab	000F2B	
18	17	Kroger	Kroger	FALSE	sans	0067B1	
19	18	Amazon.com	Amazon	FALSE	sans	FF9900	
20	19	**Walgreens Boots Alliance**	Walgreens Boots Alliance	FALSE	sans	232C64	
21	20	Hewlett-Packard Company	hp	FALSE	sans	0096D6	
22	21	Cardinal Health	CardinalHealth	FALSE	sans	E41F35	
23	22	Express Scripts Holding	EXPRESS SCRIPTS	FALSE	sans	1162A5	
24	23	J.P. Morgan Chase	JPMorgan Chase & Co	TRUE	serif	414141	
25	24	Boeing	BOEING	FALSE	sans	1D439C	
26	25	Microsoft	Microsoft	FALSE	sans	737373	
27	26	Bank of America Corp.					
28	27	Wells Fargo					
29	28	Home Depot					
30	29	Citigroup					
31	30	Phillips 66					
32	31	IBM					
33	32	Valero Energy					
34	33	Anthem					
35	34	Procter & Gamble					
36	35	State Farm Insurance Cos.					
37	36	Alphabet					
38	37	Comcast					
39	38	Target					
40	39	Johnson & Johnson					
41	40	MetLife					
42	41	Archer Daniels Midland					
43	42	Marathon Petroleum					
44	43	Freddie Mac					
45	44	PepsiCo					
46	45	United Technologies					
47	46	Aetna					
48	47	Lowe's					

I was bound to fail.

“Humans rarely choose things in absolute terms.
We don’t have an internal value meter that
tells us how much things are worth.

Rather, we focus on the relative advantage
of one thing over another,
and estimate value accordingly.”

—**Dan Ariely**
Professor of Psychology and Behavioral Economics at Duke University
“Predictably Irrational” 2018

I used Keras, Tensorflow, and a pre-trained general-purpose image classification model, to describe each logo's shape.

Language
Python

Model
ResNet50

Input: 1024×1024 px png

Output: Quantified presence of 2048 features

```
0.02772 0.00000 0.00000 0.00000 1.09299 0.00000
0.01938 0.55155 0.00009 0.00014 0.00000 0.00000
0.33905 0.10659 1.46157 0.08142 0.00411 0.00000
0.00000 0.36360 0.00000 0.03051 1.01659 0.05828
0.10766 0.00000 0.68964 0.00040 0.00000 1.51671
0.01628 0.00000 0.01213 0.00000 0.00000 0.00000
0.06907 0.00000 0.00892 0.00000 0.06411 0.08882
0.00000 0.00000 0.00000 0.02236 0.00000 0.05602
0.06424 0.00000 1.22857 1.15797 0.20735 0.49470
0.09517 0.11458 0.00000 0.00105 0.09719 1.76852
0.05036 0.00000 0.00000 0.00509 0.00000 0.00000
0.00000 0.14901 0.00000 0.00000 0.07446 0.25466
0.40862 0.00000 0.00000 0.01263 0.00183 0.12038
0.00457 0.00000 0.06357 0.37006 0.34136 0.02781
0.05190 0.05535 0.00000 0.00242 0.00000 0.00000
0.00065 0.00000 0.00000 0.01920 0.00000 0.00000
0.23270 0.00000 0.10789 0.00000 0.00000 4.78532
```

The top two colors were found using k-means clustering.

K-means clustering attempts to partition all colored pixels into k clusters of “similar” appearance.

Language
Python

Input: 1024×1024 px png

Output: Two most prominent colors as RGB values

0 124.86442 197.96974
#007DC6

255 194.16913 31.88876
#FFC220

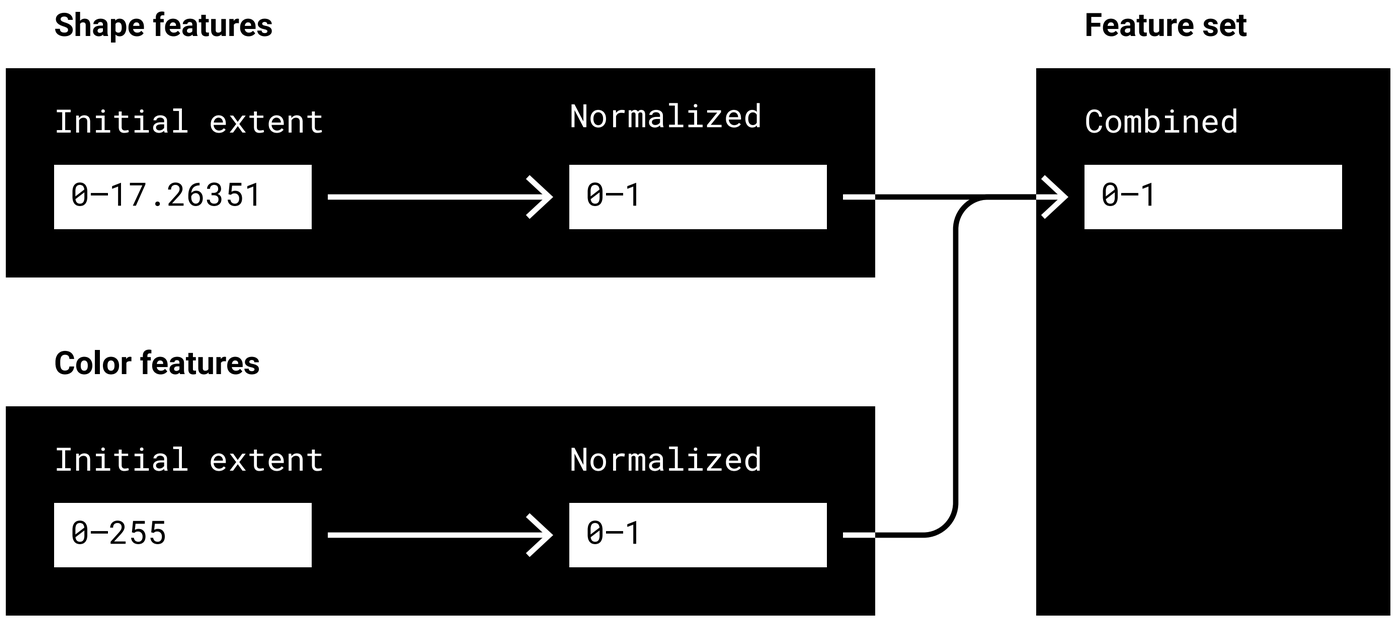
For each logo, thousands of features were generated

```
   2048  from neural network
+     6  from k-means clustering
= 2054  features
```

The neural network also produced some “noise.”
This was removed through principal component analysis (PCA).

```
  2048  from neural network
- 1998  PCA
=   50  features describing shape
+    6  features describing color
=   56  features
```

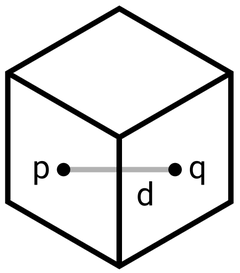
To combine the two feature sets, they must first be normalized independently.



Each feature is a dimension for describing a logo.
Together, they form an n-dimensional space—the “space of logos.”

This enables calculating the Euclidean distance
between two points (logos) within the space—
a measurement of their similarity.

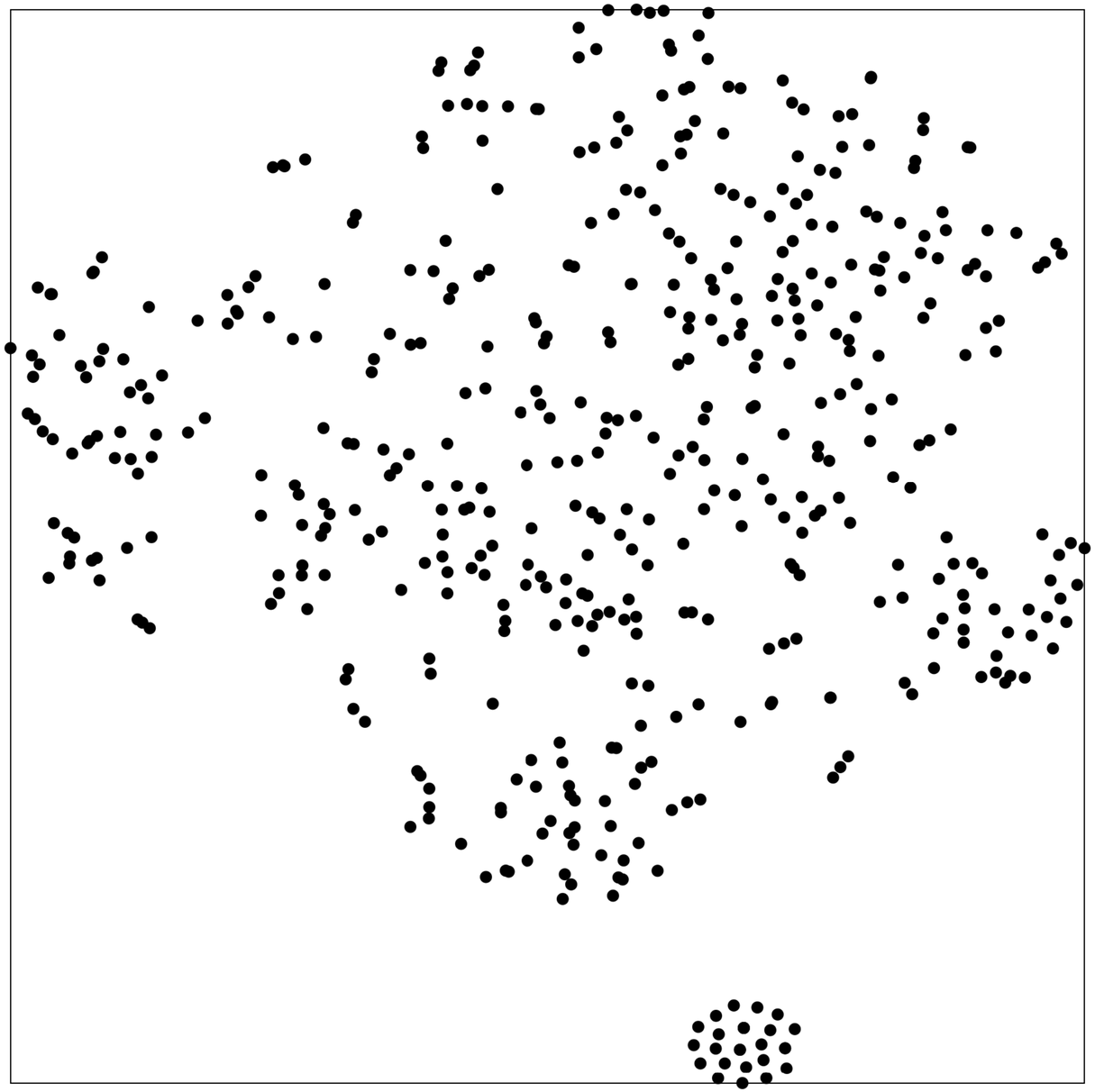
$$d(\mathbf{p}, \mathbf{q}) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_i - q_i)^2 + \cdots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}.$$



Calculating all pairwise distances yields a distance matrix.

Walmart	Berkshire Hathaway	Apple	Exxon Mobil	McKesson	UnitedHealth Group	CVS Health	
0	22.65585	30.05652	20.74215	14.98730	19.48658	26.64625	**Walmart**
22.65585t	0	20.30818	15.83581	17.60751	11.07304	32.29748	**Berkshire Hathaway**
30.05652	20.30818	0	25.22732	28.32987	20.34397	37.80716	**Apple**
20.74215	15.83581	25.22732	0	17.84799	16.52740	26.67482	**Exxon Mobil**
14.98730	17.60751	28.32987	17.84799	0	14.83402	30.27049	**McKesson**
19.48658	11.07304	20.34397	16.52740	14.83402	0	32.93077	**UnitedHealth Group**
26.64625	32.29748	37.80716	26.67482	30.27049	32.93077	0	**CVS Health**
27.11042	31.06151	37.25250	29.44294	28.88743	31.78418	31.17721	**General Motors**
49.49596	53.56491	54.49421	53.04318	51.71548	52.81218	51.24819	**AT&T**

Reducing the dimensionality of the distance matrix results in a point cloud that can be plotted.



Algorithm
t-SNE

Steps
600

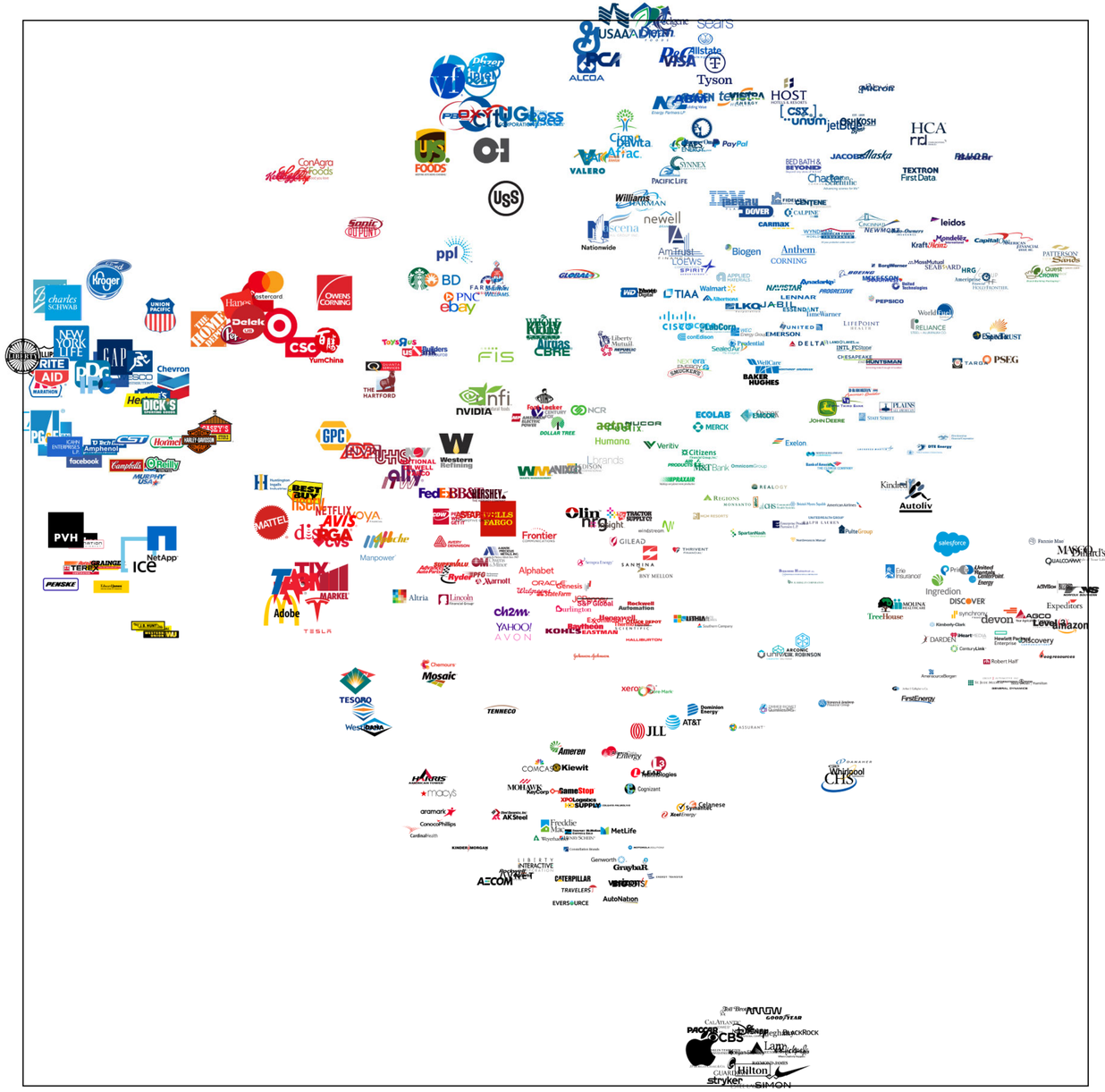
Epsilon
10

Perplexity
15

Each point corresponds with a Fortune 500 company logo and can be replaced by their logo.

Still need to resolve:

- Boundary violations
- Overlapping logos

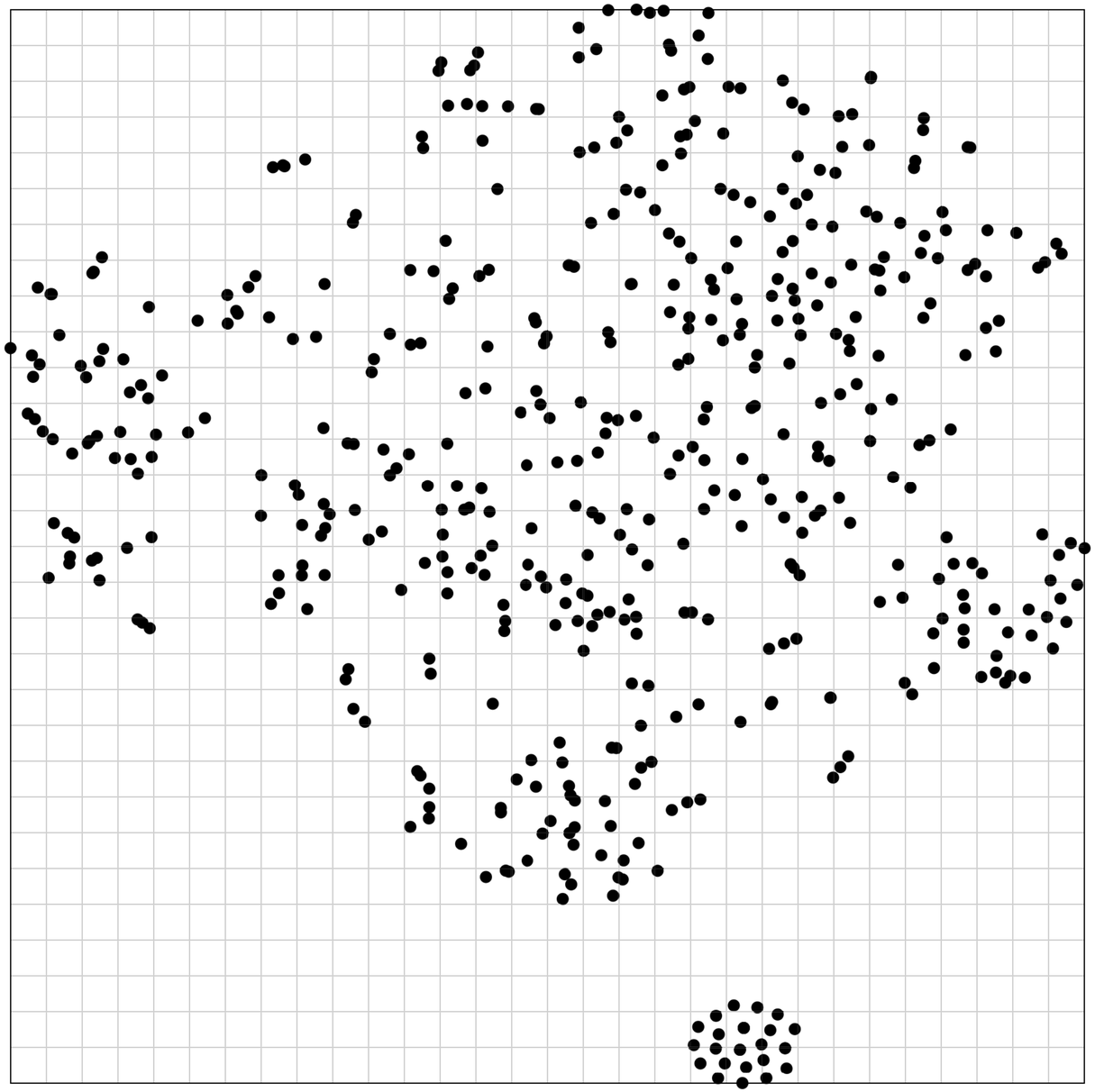


Introducing a grid to align points solves both problems.

Pairing cells and points
is an assignment problem—
a form of optimization.

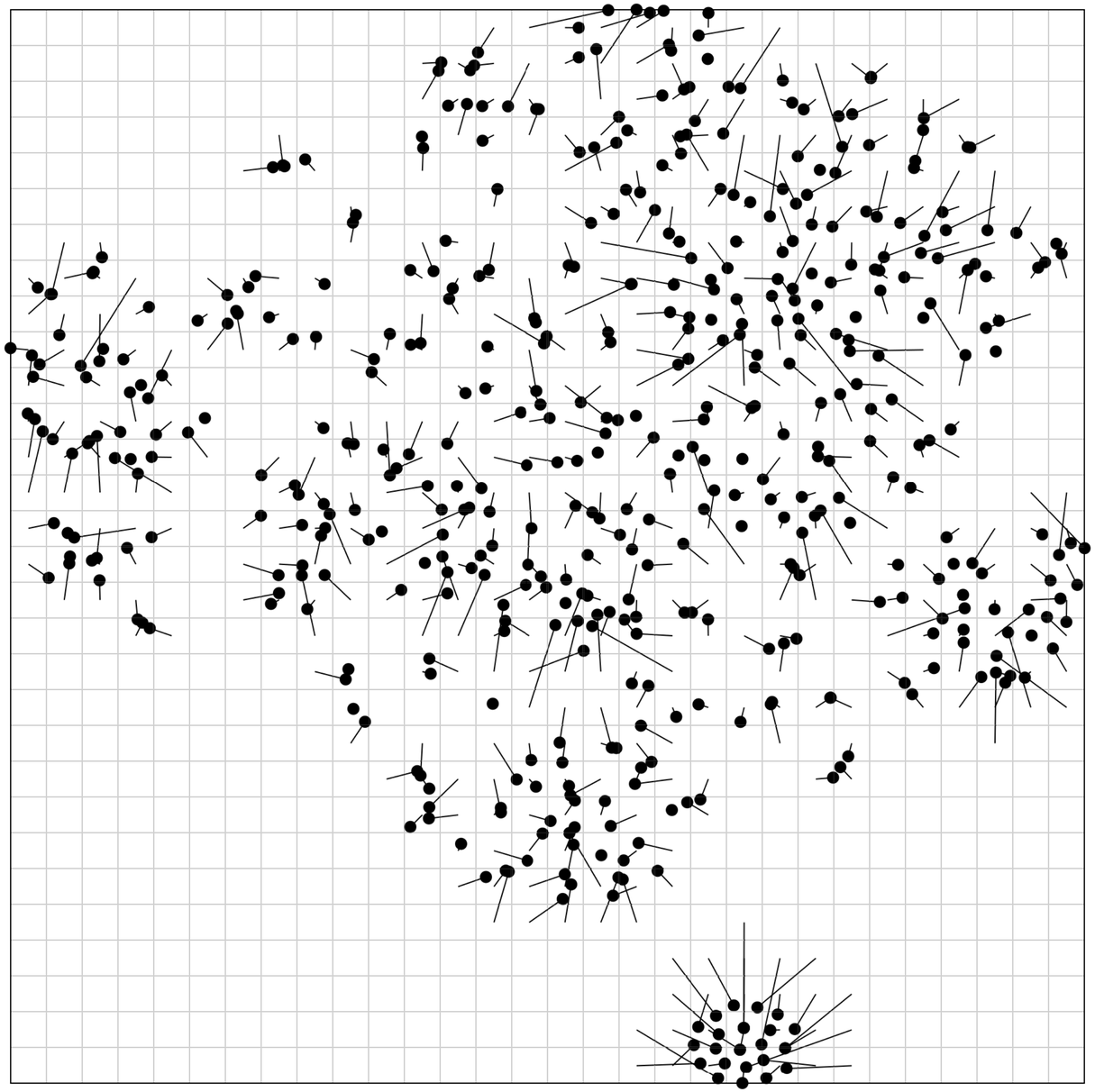
Cells per side
30

Number of cells
900



Each point is assigned to a cell.
The further a point moves,
the more it distorts the data.

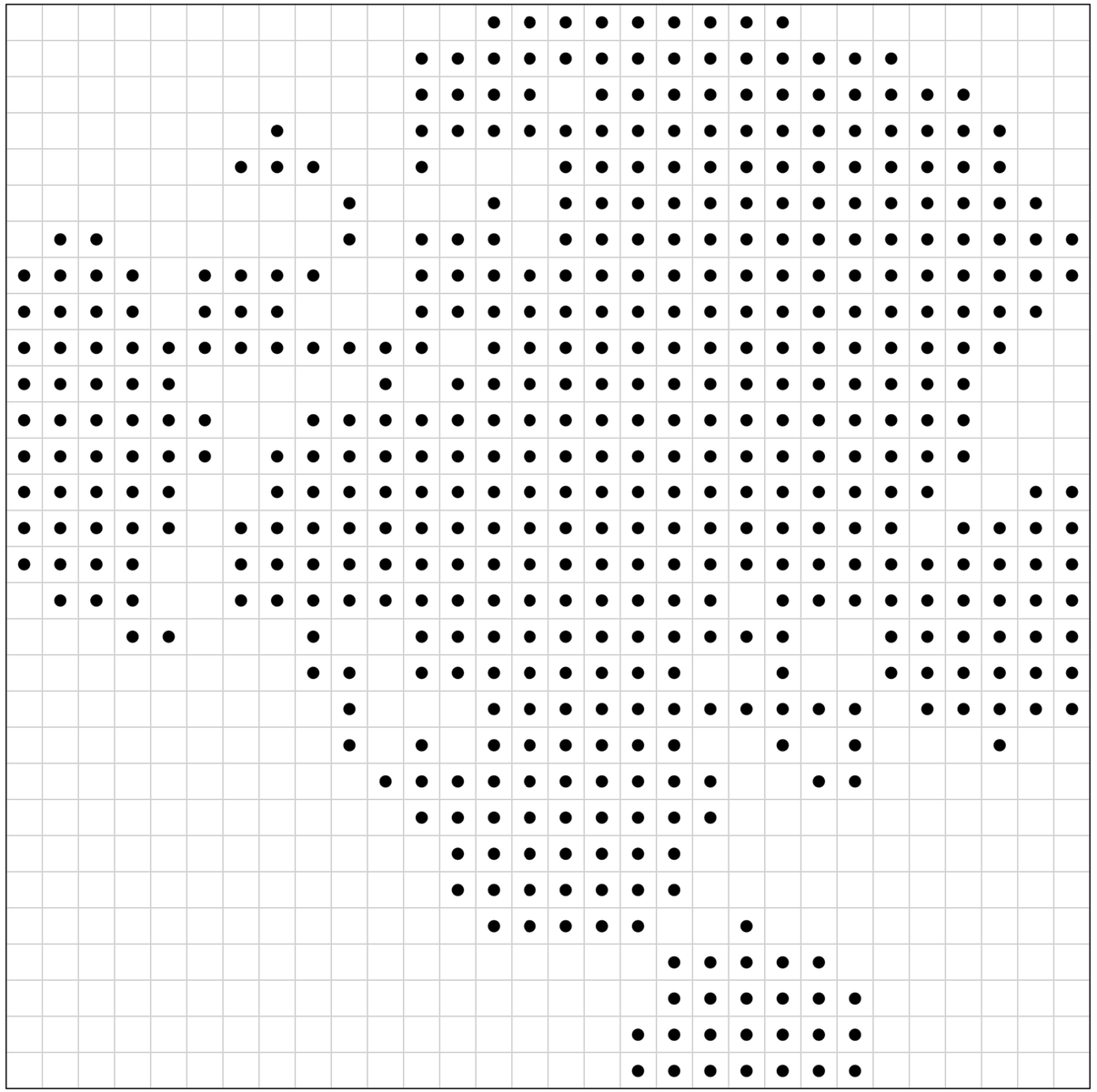
The goal is to achieve
the lowest overall distortion.



Algorithm
Jonker-Volgenant

Cells per side
30

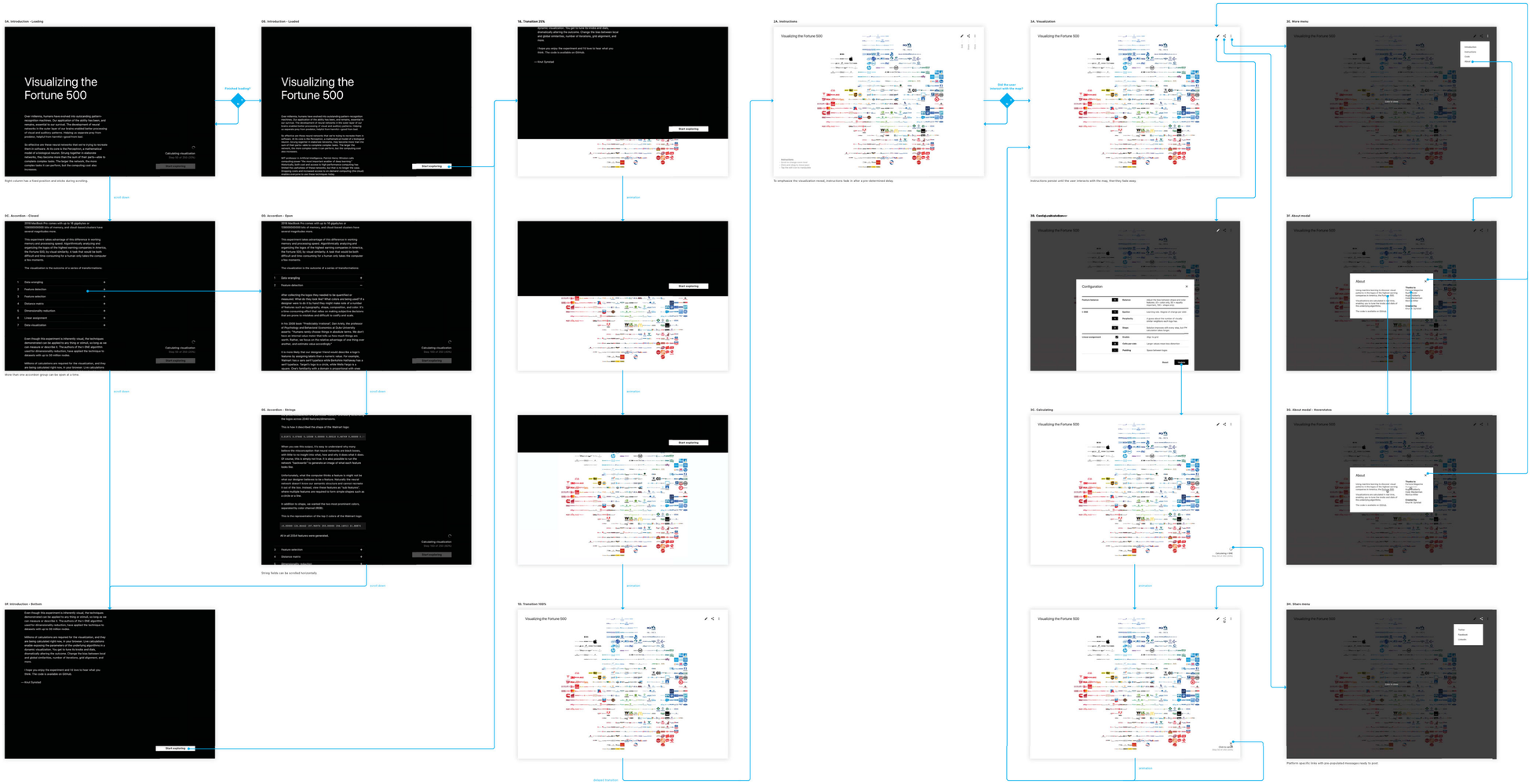
Result of the linear assignment.



This output is one of many in a “space of solutions.”

By surfacing parameters, users can manipulate the outcome.

For manipulation of the solution space to be possible, it must be wrapped in a larger context—a user interface.



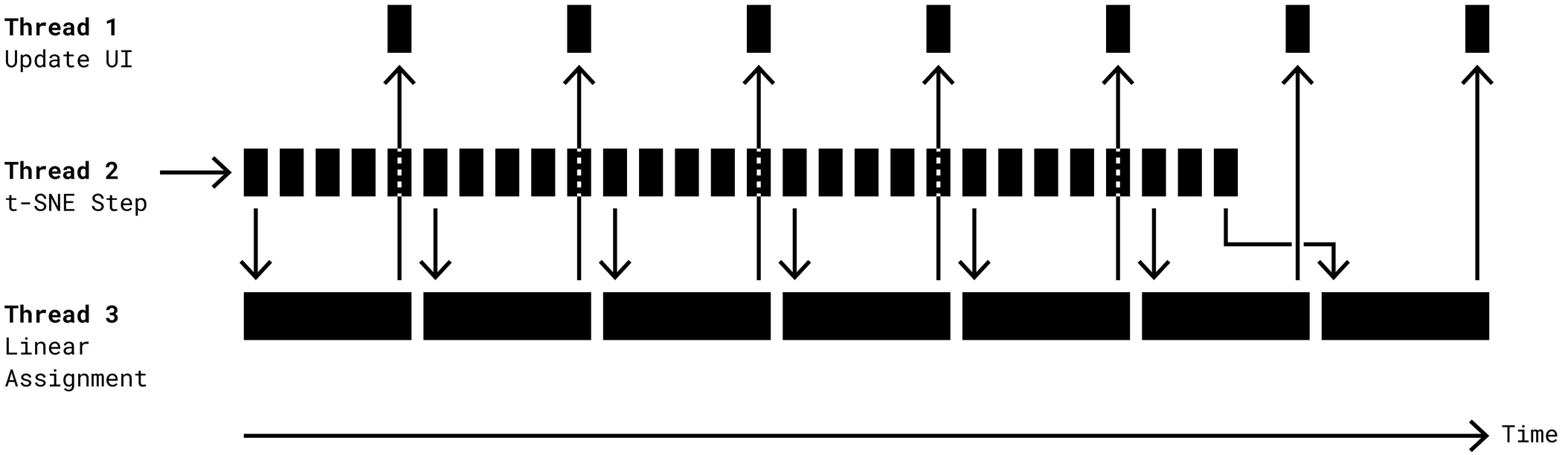
An interactive exhibit requires some transformations to be calculated live—in the browser.

1. Data wrangling
2. Feature detection
3. Feature selection
4. Distance matrix — Pre-calculated

---

5. Dimensionality reduction — Live
6. Linear assignment
7. Data visualization

It simply isn’t feasible to calculate it all in the main UI thread—it would block the UI, or worse, crash the browser.
The heavy lifting must be moved to additional threads.



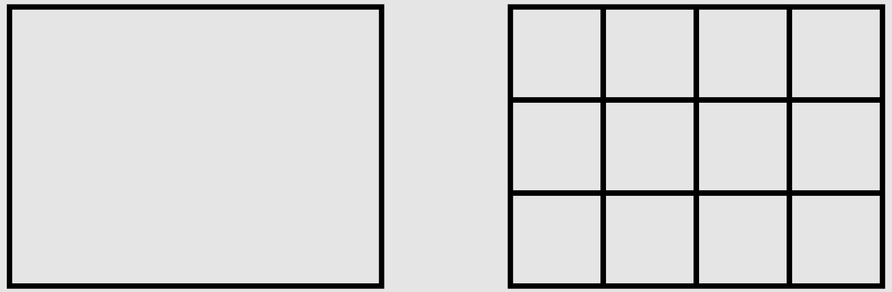
The result is an interactive visualization, where users can adjust the parameters, to find similarities and differences, in the logos of the Fortune 500.

Let’s take a look:
fortune500.knutsynstad.com

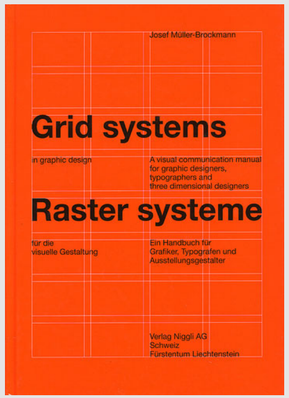
Source code
github.com/knutsynstad/fortune500

If you have a moment,
I’d like to share another example.

In 2006, Bill Drenttel and Jessica Helfand patented the 3 × 4 grid.



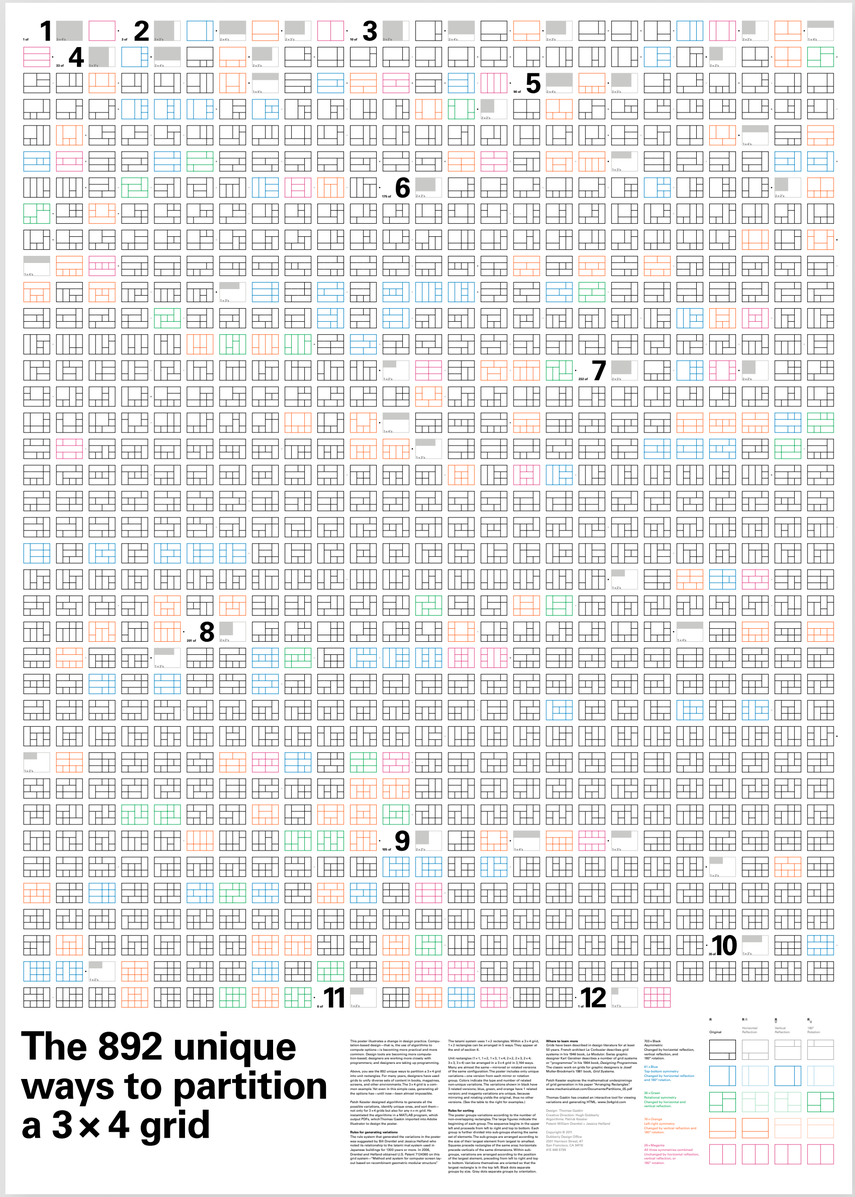
There may have been prior art



The 3 × 4 grid forms a solution space with a finite number of permutations.

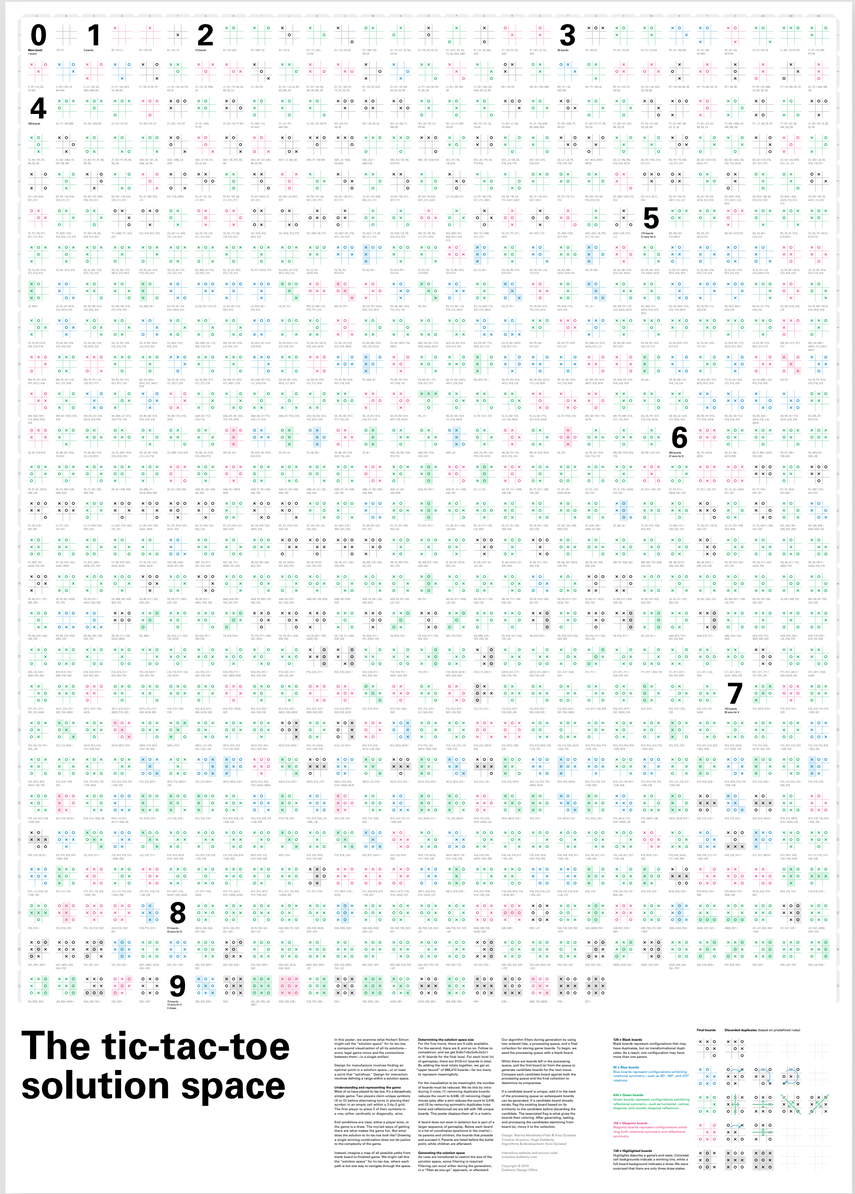
How do you find them all?
It seemed like a simple counting problem.

Turned out not to be.



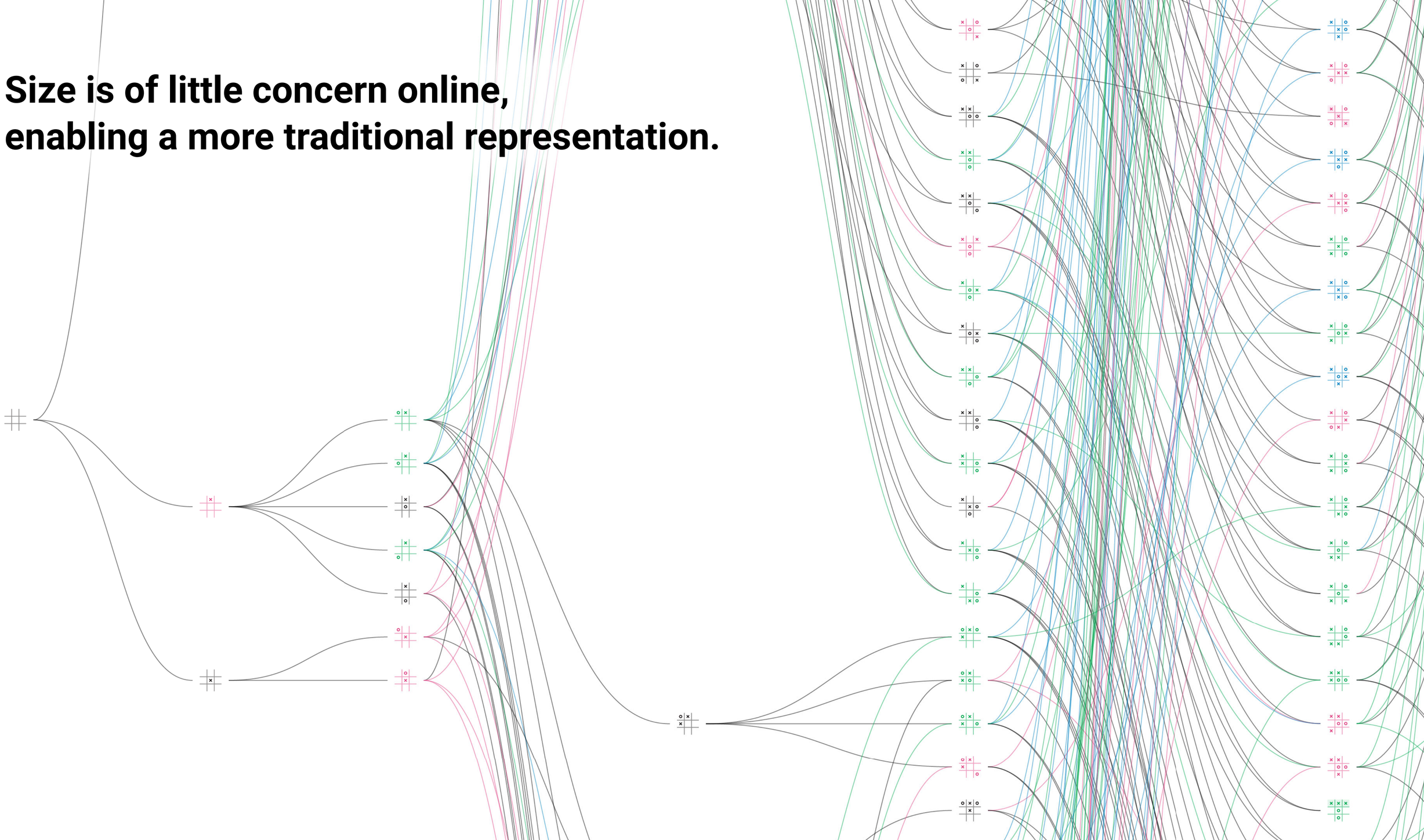
Last year, I explored the Tic-Tac-Toe solution space.

This poster includes all 765 unique game board configurations of the Tic-Tac-Toe solution space and the relationships between them.



0
Move (level)
1 board
1
3 boards
2
12 boards
3
38 boards
4
108 boards
5
174 boards
21 wins for X
6
204 boards
21 wins for O

Size is of little concern online,
enabling a more traditional representation.



I’ll end with this …

Increasingly, products are delivered as services,
and increasingly services require data and computation.

As a result, data and computation are media for designing.

And design and code have begun converging.

Special thanks to
John Cain
Hugh Dubberly
Kristin Gecan
Ryan Keisler
Cody Wackerman

Knut M. Synstad
@knutsynstad

Presentation posted at
knutsynstad.com/latham.pdf