Visualizing the Space of Logos of the Fortune 500

Knut M. Synstad
@knutsynstad

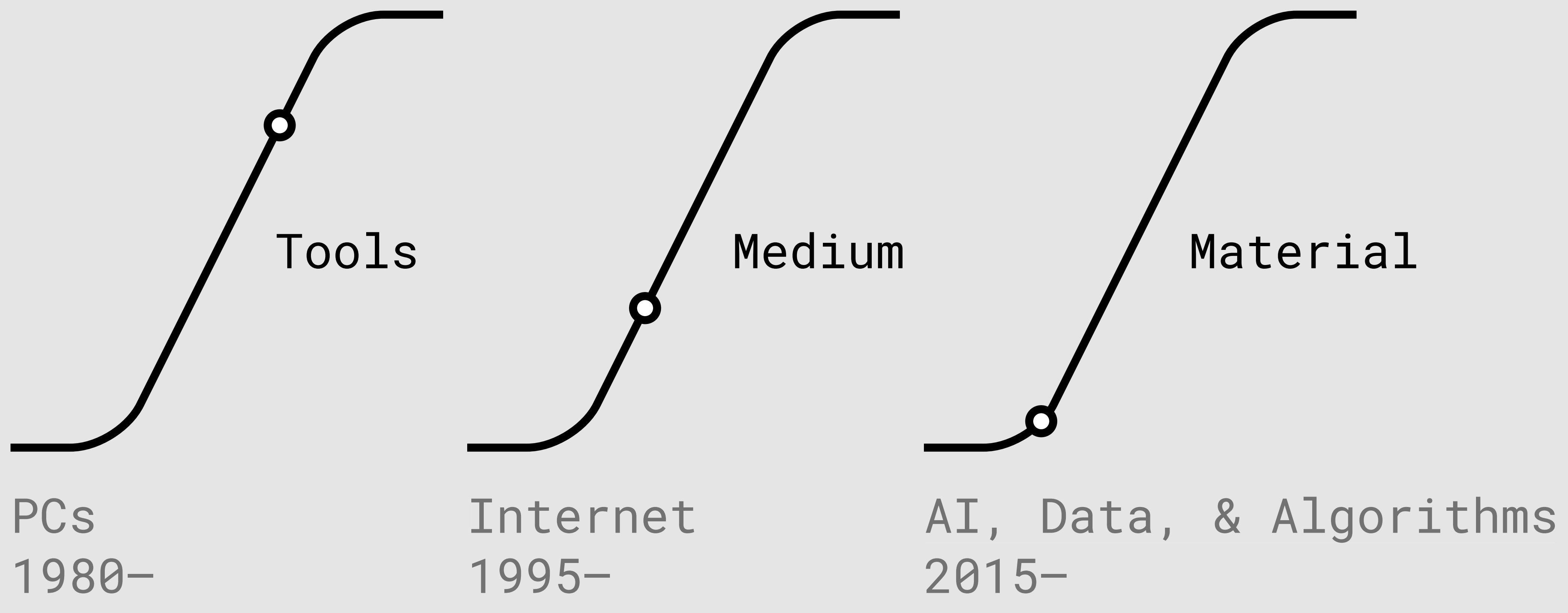
Hello, I’m Knut.
A designer that codes.

Work in systems/UX/UI.
Exploring AI/ML.

Formerly electronics in
Norwegian Navy.

Computation and AI is of increasing importance for design.

As a:
- **Tool**, augmenting the design process.
- **Medium**, for sharing information.
- **Material**, to be shaped into products.



After
Hugh Dubberly

There are many exciting examples …

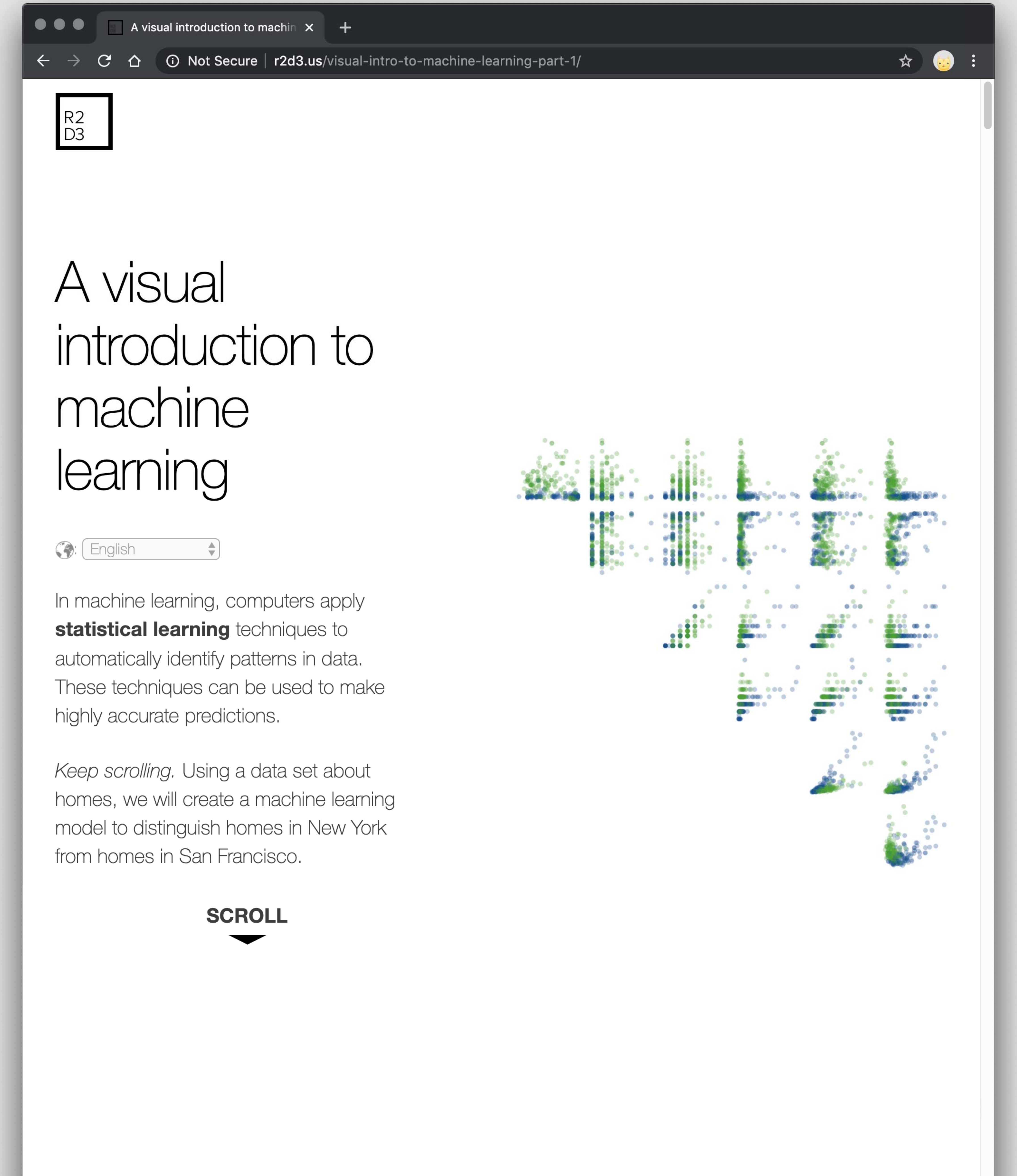
A visual introduction to machine learning
By Stephanie Yee and Tony Chu

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

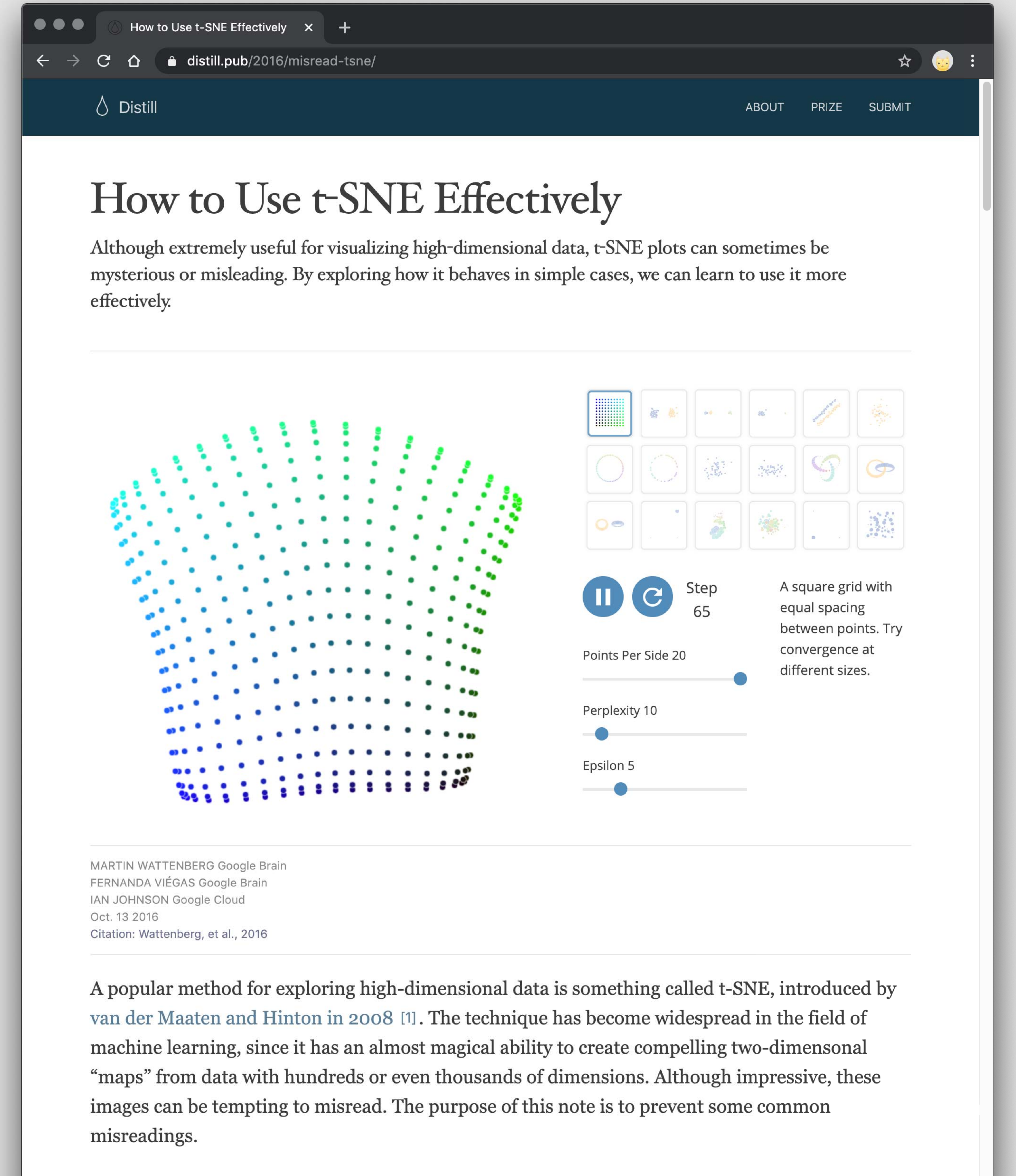
By Martin Wattenberg, Fernanda Viégas, and Ian Johnson

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**

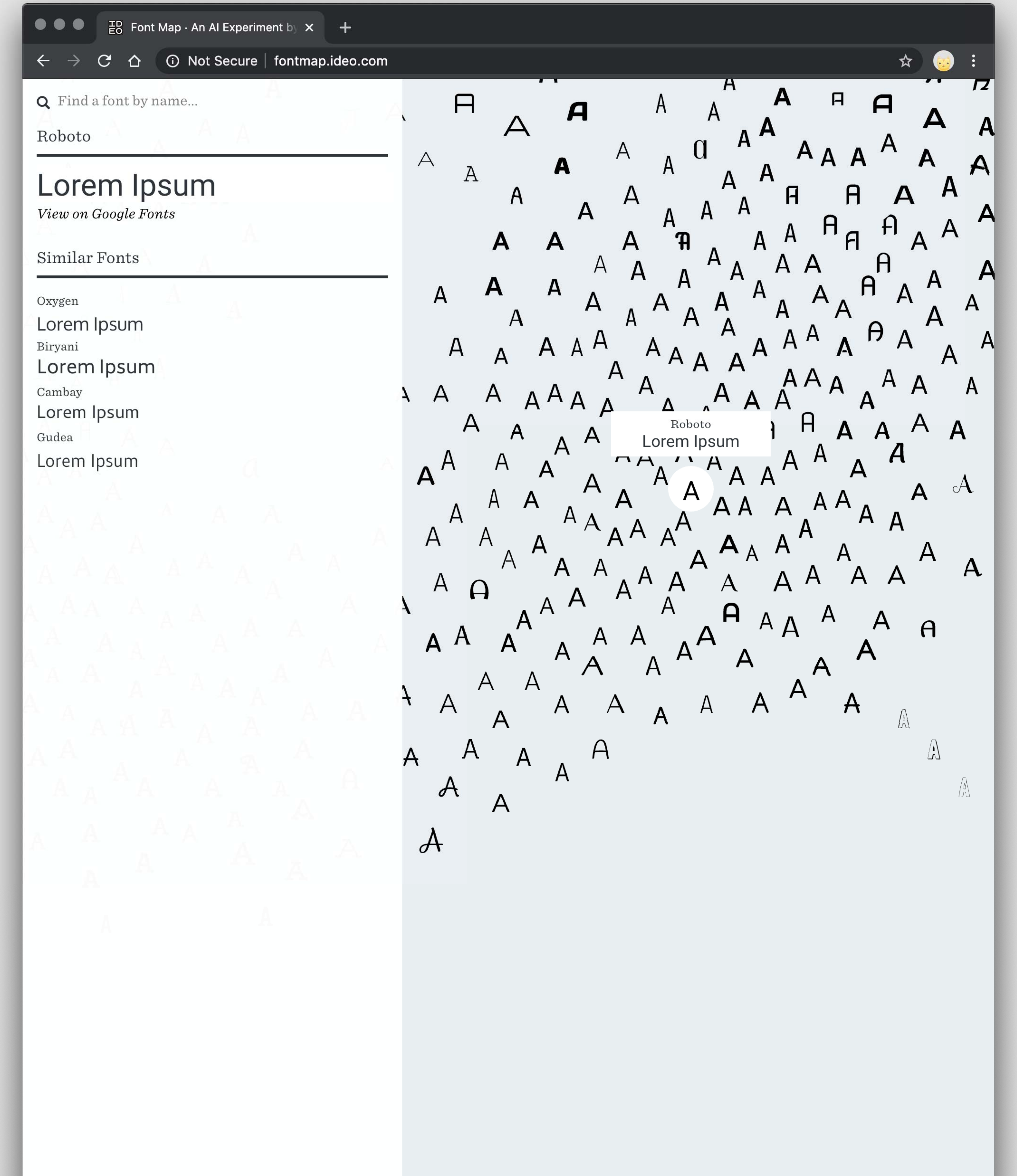
Year
2016

Address
distill.pub/2016/misread-tsne/



Font Map

By Kevin Ho (IDEO)



Year
2017

Address
fontmap.ideo.com/

It made me wonder...

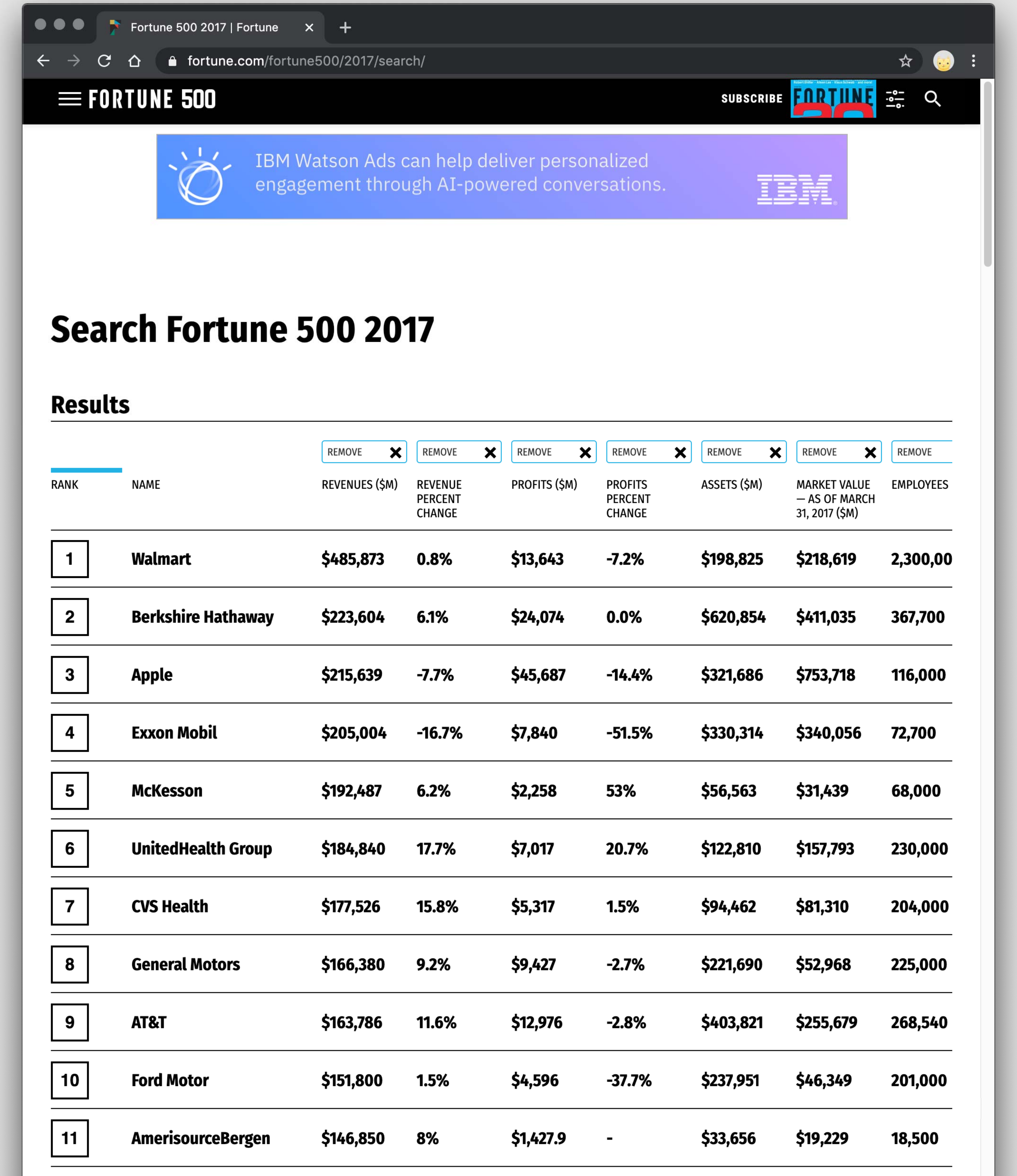
What would it take to organize the logos of the Fortune 500, by visual similarity?

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.



Fortune 500 2017 | Fortune

fortune.com/fortune500/2017/search/

FORTUNE 500 SUBSCRIBE

Search Fortune 500 2017

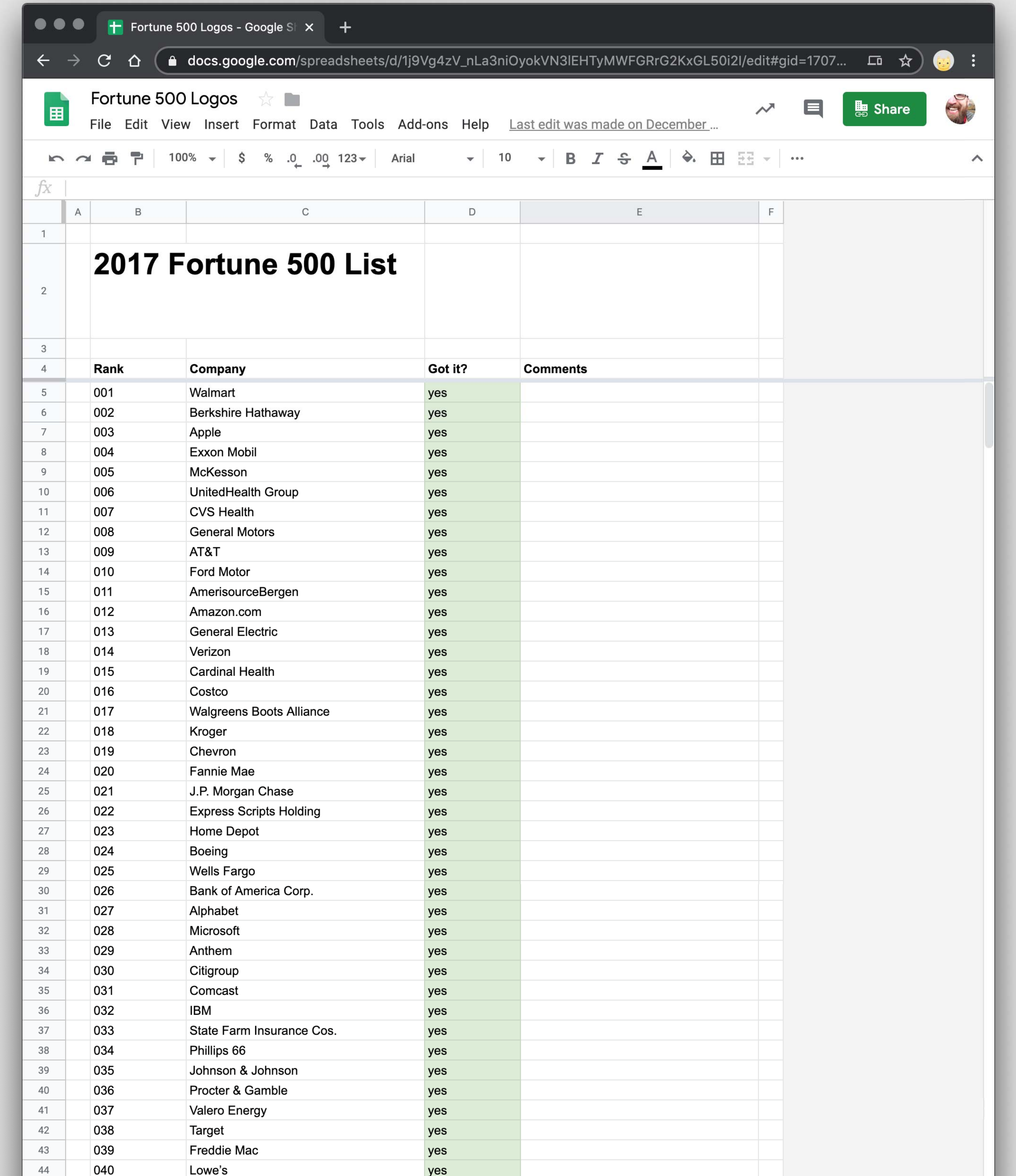
Results

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

Address
fortune.com/fortune500/2017/

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

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

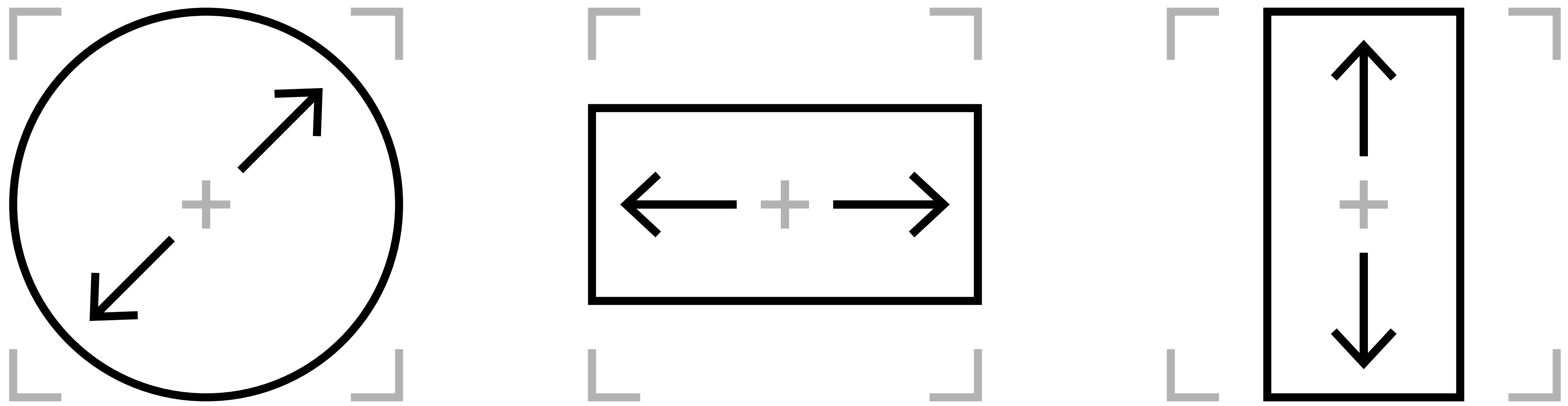


2017 Fortune 500 List

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	

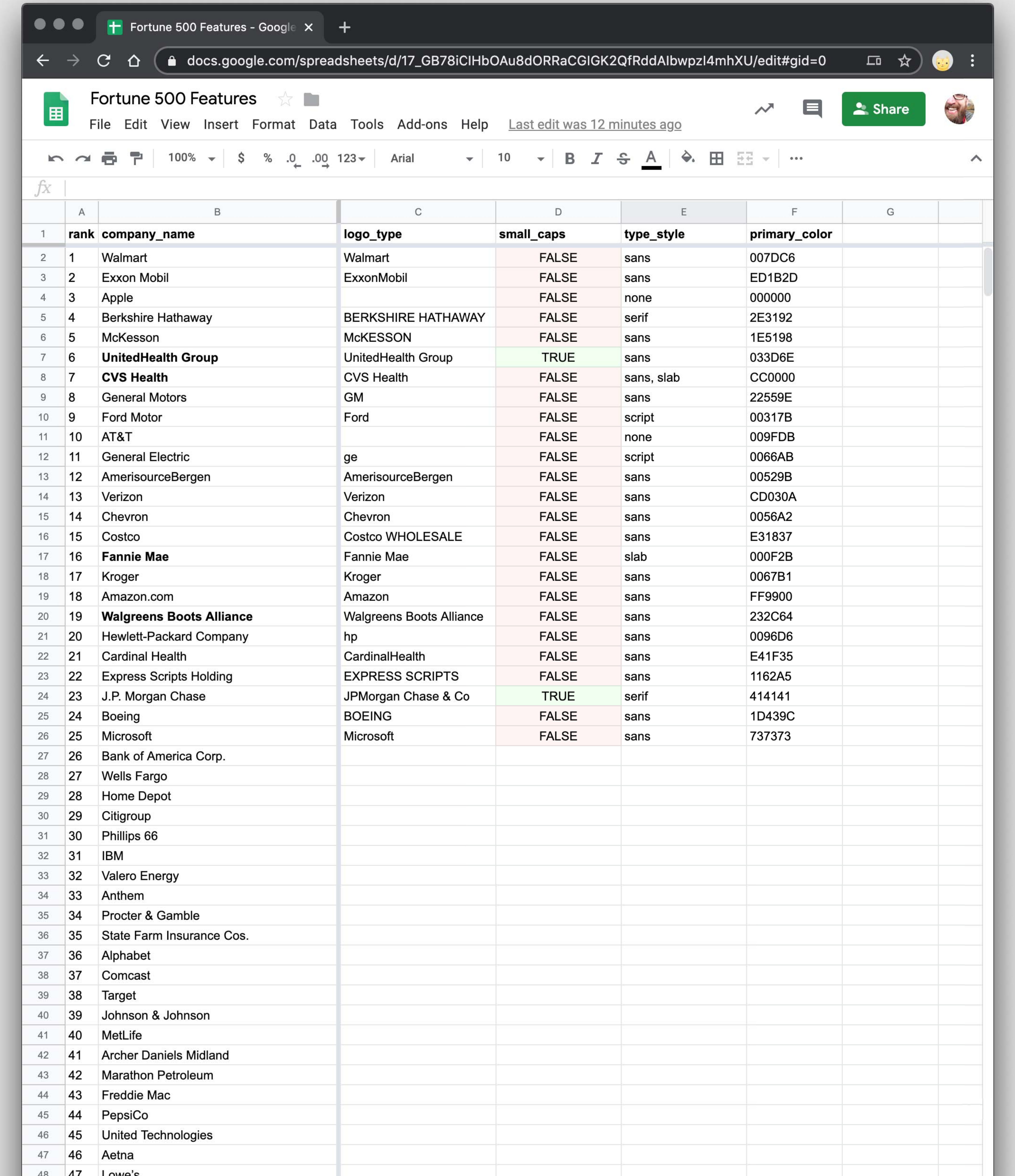
Artwork came from multiple sources, and in multiple file formats, requiring some standardization.

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



2. Feature detection

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



rank	company_name	logo_type	small_caps	type_style	primary_color
1	Walmart	Walmart	FALSE	sans	007DC6
2	Exxon Mobil	ExxonMobil	FALSE	sans	ED1B2D
3	Apple		FALSE	none	000000
4	Berkshire Hathaway	BERKSHIRE HATHAWAY	FALSE	serif	2E3192
5	McKesson	McKESSON	FALSE	sans	1E5198
6	**UnitedHealth Group**	UnitedHealth Group	TRUE	sans	033D6E
7	**CVS Health**	CVS Health	FALSE	sans, slab	CC0000
8	General Motors	GM	FALSE	sans	22559E
9	Ford Motor	Ford	FALSE	script	00317B
10	AT&T		FALSE	none	009FDB
11	General Electric	ge	FALSE	script	0066AB
12	AmerisourceBergen	AmerisourceBergen	FALSE	sans	00529B
13	Verizon	Verizon	FALSE	sans	CD030A
14	Chevron	Chevron	FALSE	sans	0056A2
15	Costco	Costco WHOLESALE	FALSE	sans	E31837
16	**Fannie Mae**	Fannie Mae	FALSE	slab	000F2B
17	Kroger	Kroger	FALSE	sans	0067B1
18	Amazon.com	Amazon	FALSE	sans	FF9900
19	**Walgreens Boots Alliance**	Walgreens Boots Alliance	FALSE	sans	232C64
20	Hewlett-Packard Company	hp	FALSE	sans	0096D6
21	Cardinal Health	CardinalHealth	FALSE	sans	E41F35
22	Express Scripts Holding	EXPRESS SCRIPTS	FALSE	sans	1162A5
23	J.P. Morgan Chase	JPMorgan Chase & Co	TRUE	serif	414141
24	Boeing	BOEING	FALSE	sans	1D439C
25	Microsoft	Microsoft	FALSE	sans	737373
26	Bank of America Corp.				
27	Wells Fargo				
28	Home Depot				
29	Citigroup				
30	Phillips 66				
31	IBM				
32	Valero Energy				
33	Anthem				
34	Procter & Gamble				
35	State Farm Insurance Cos.				
36	Alphabet				
37	Comcast				
38	Target				
39	Johnson & Johnson				
40	MetLife				
41	Archer Daniels Midland				
42	Marathon Petroleum				
43	Freddie Mac				
44	PepsiCo				
45	United Technologies				
46	Aetna				
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.

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

Language
Python

Model
ResNet50

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

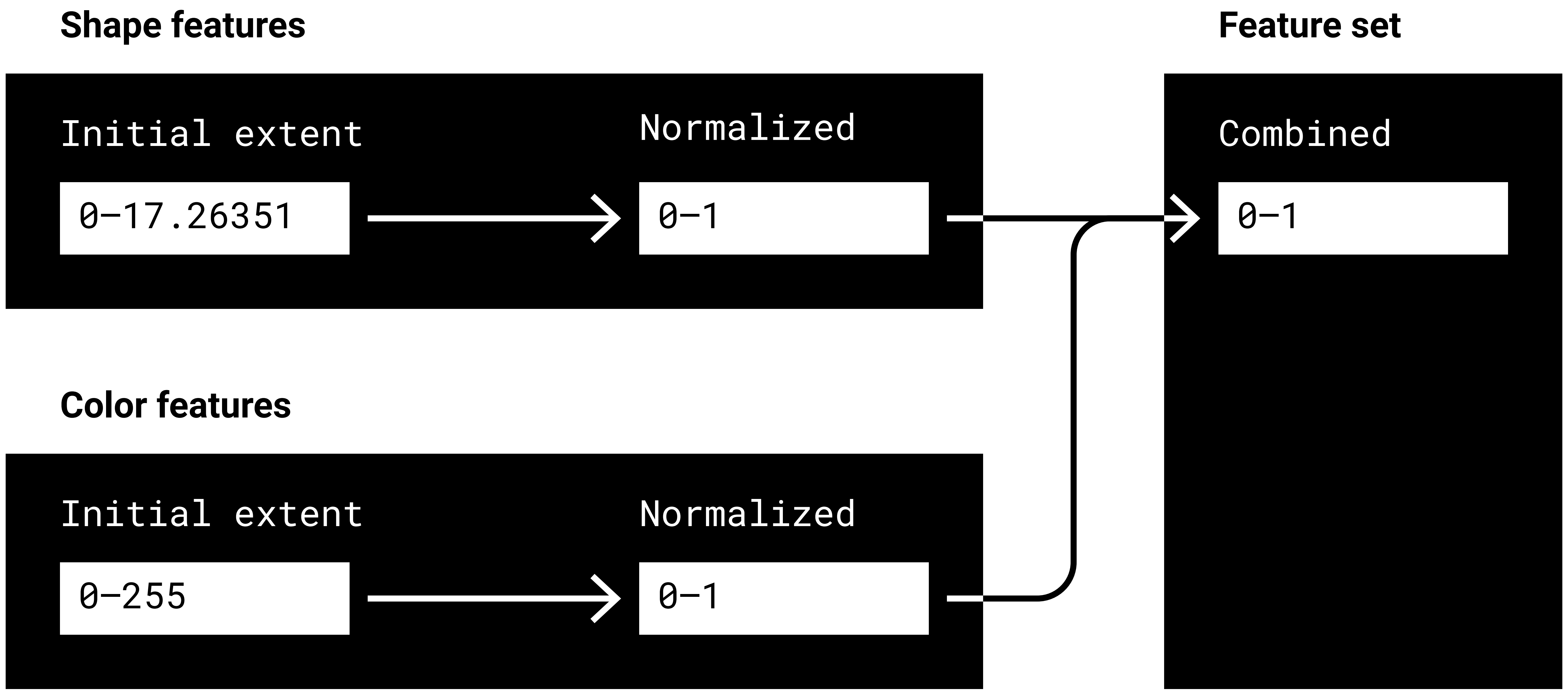
For each logo, thousands of features were generated

```
  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

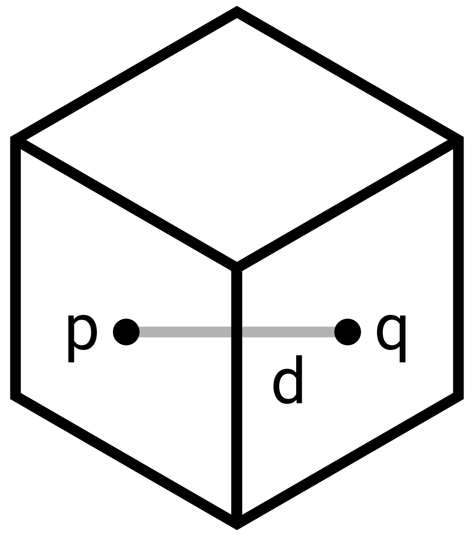
To combine the two feature sets, they must first be normalized independently.



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.

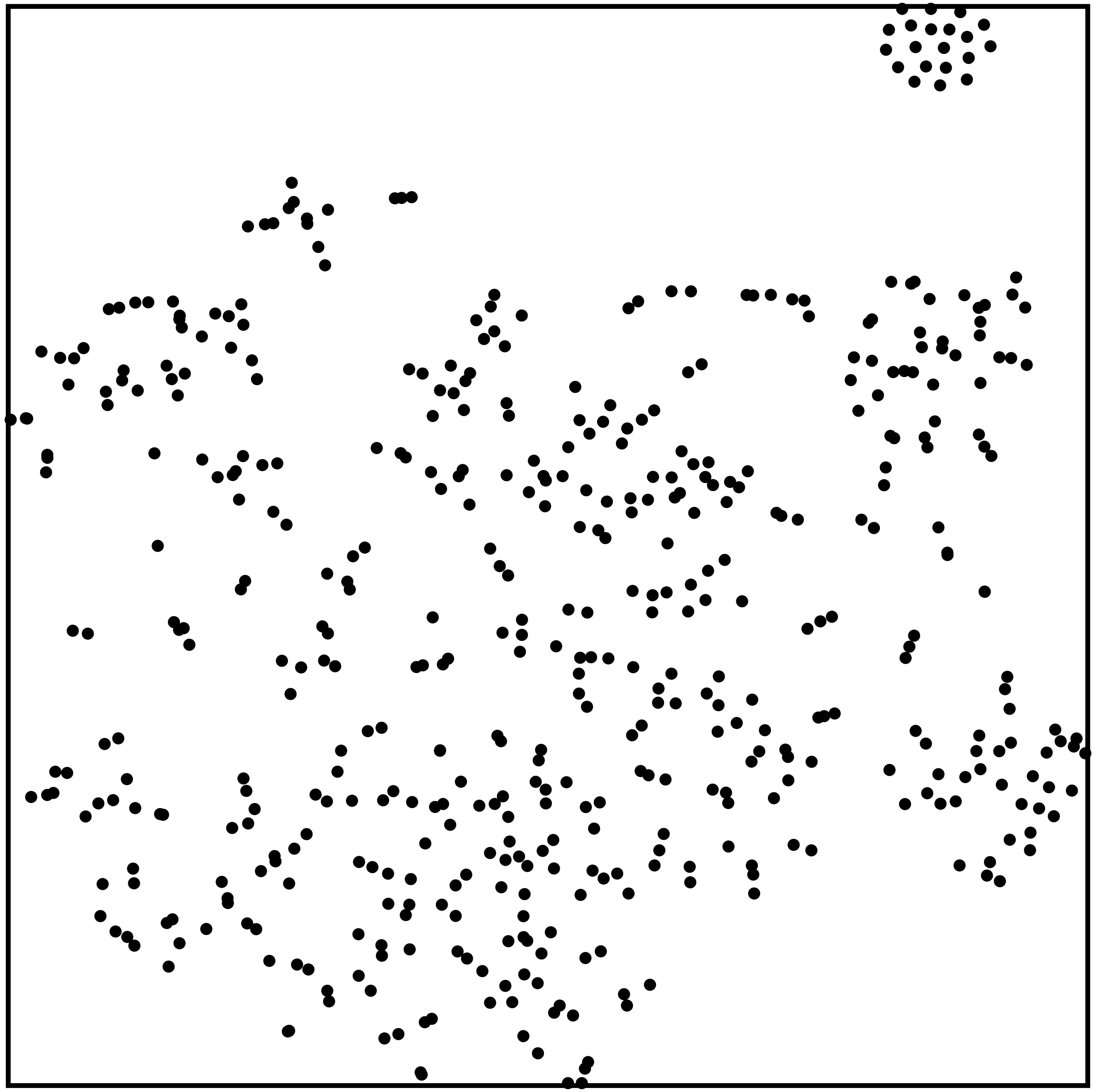
$$d(\mathbf{p},\mathbf{q}) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_i - q_i)^2 + \cdots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}.$$



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
750

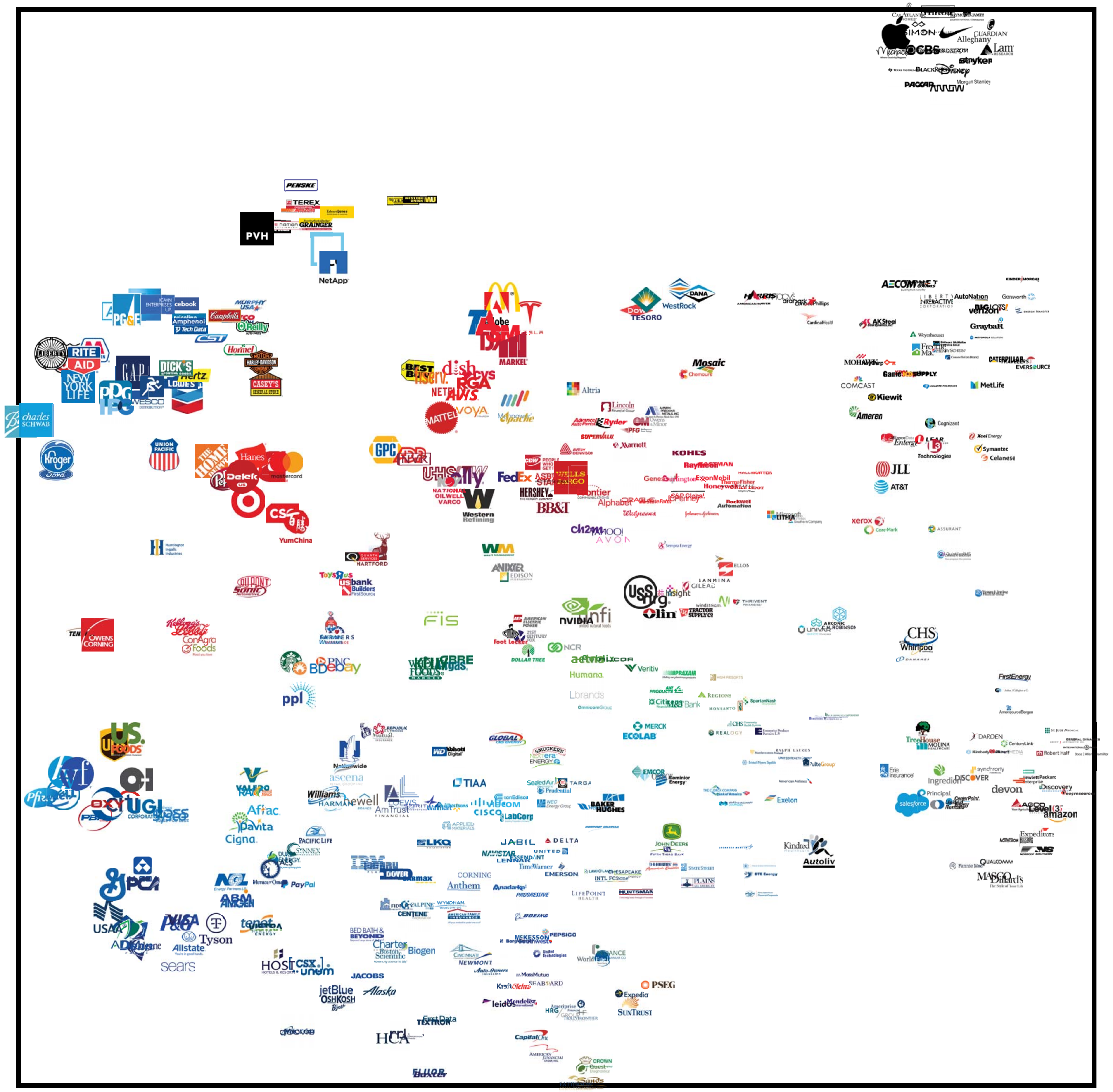
Epsilon
10

Perplexity
20

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

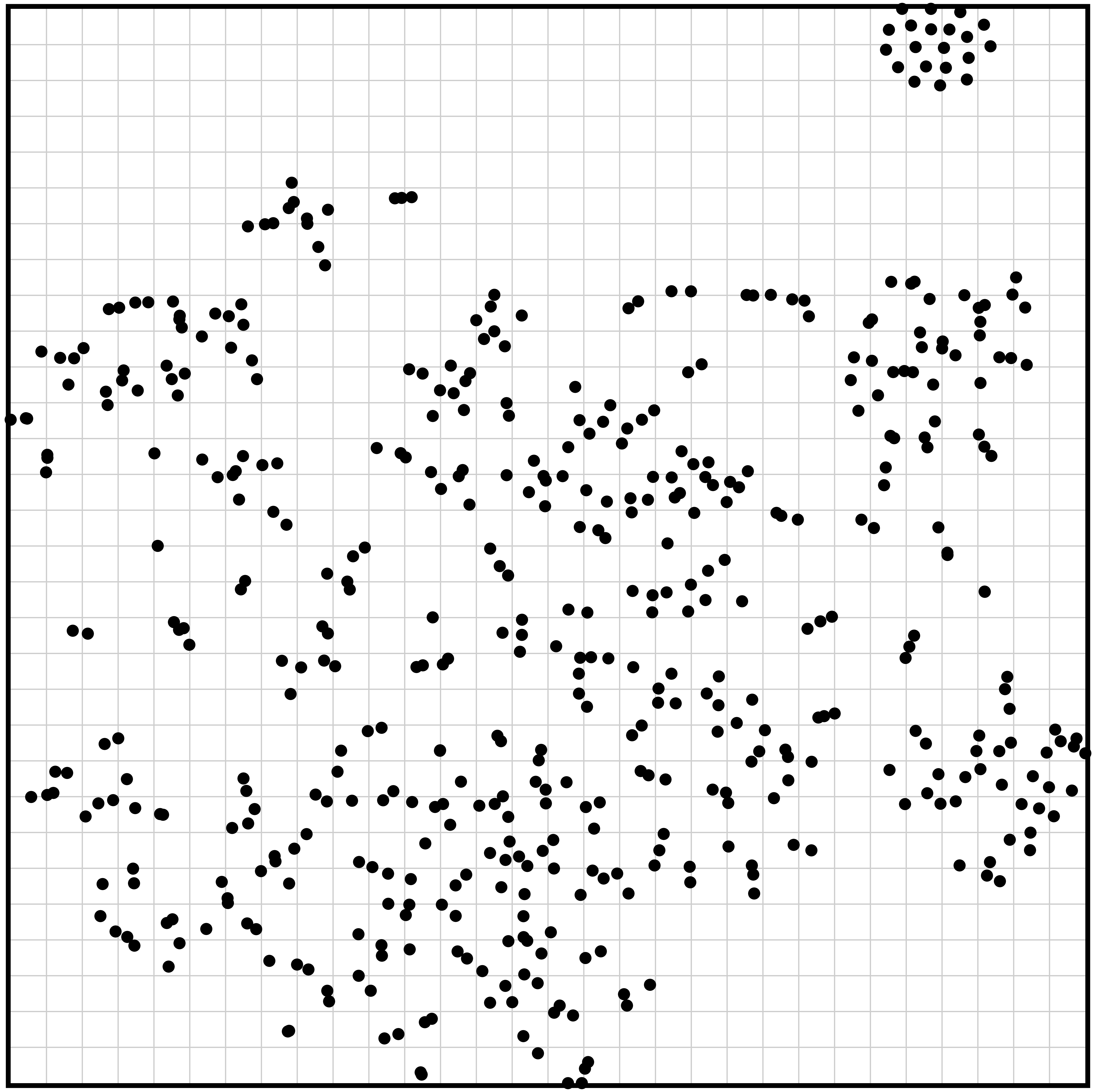
Still need to resolve:

- 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.

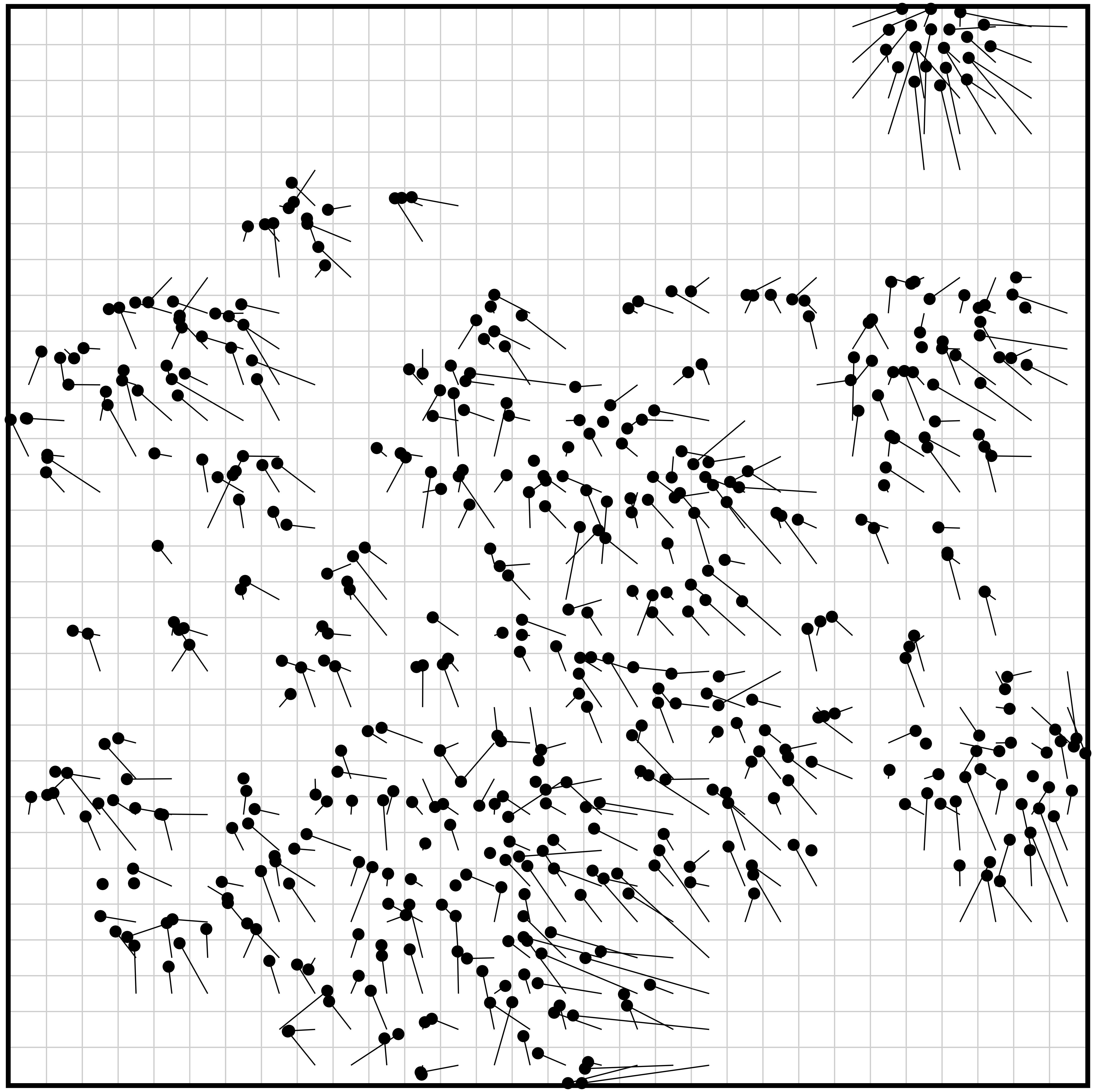


Cells per side
30

Number of cells
900

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

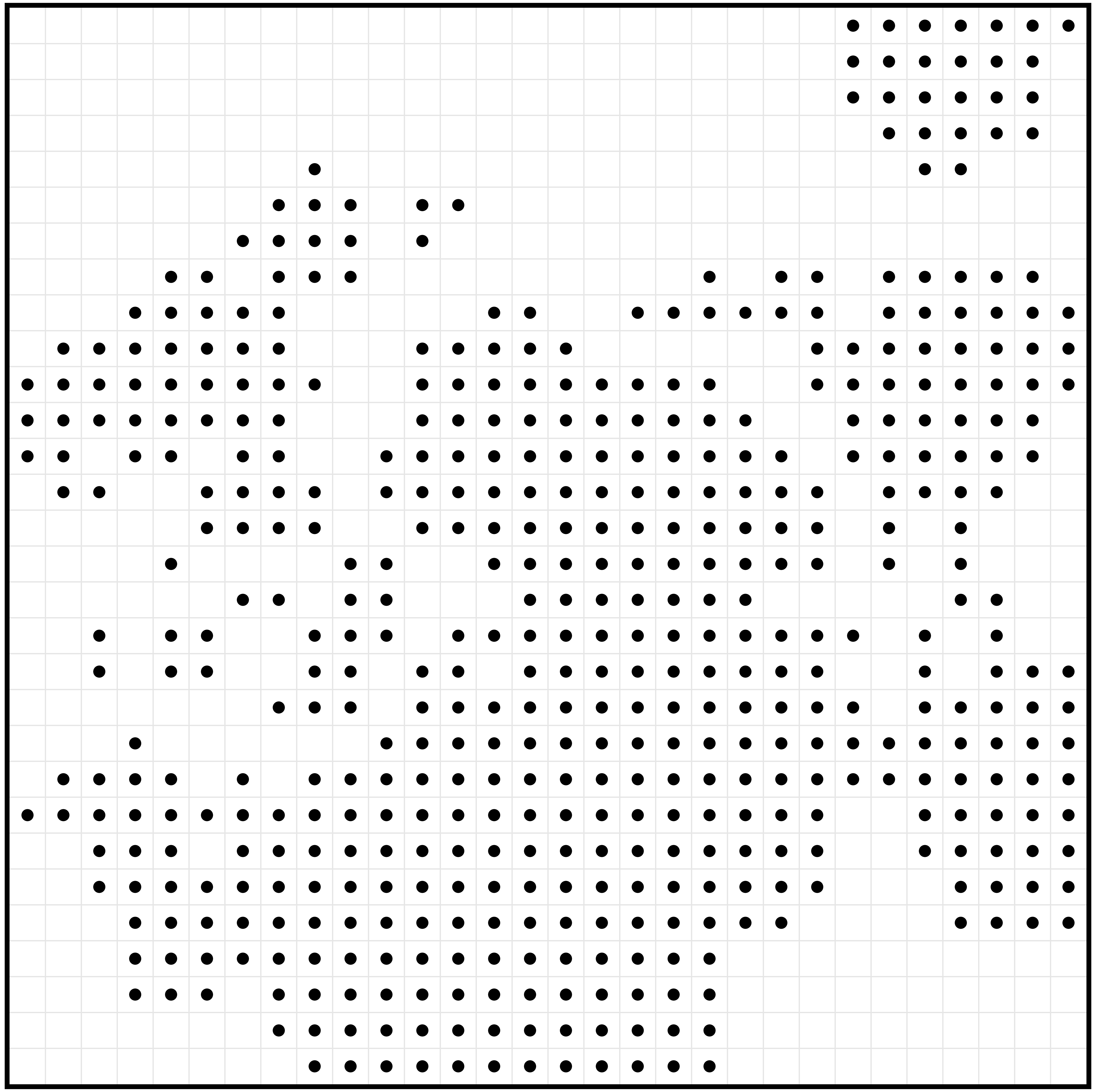
The goal is to achieve
the lowest overall distortion.



Algorithm
Jonker-Volgenant

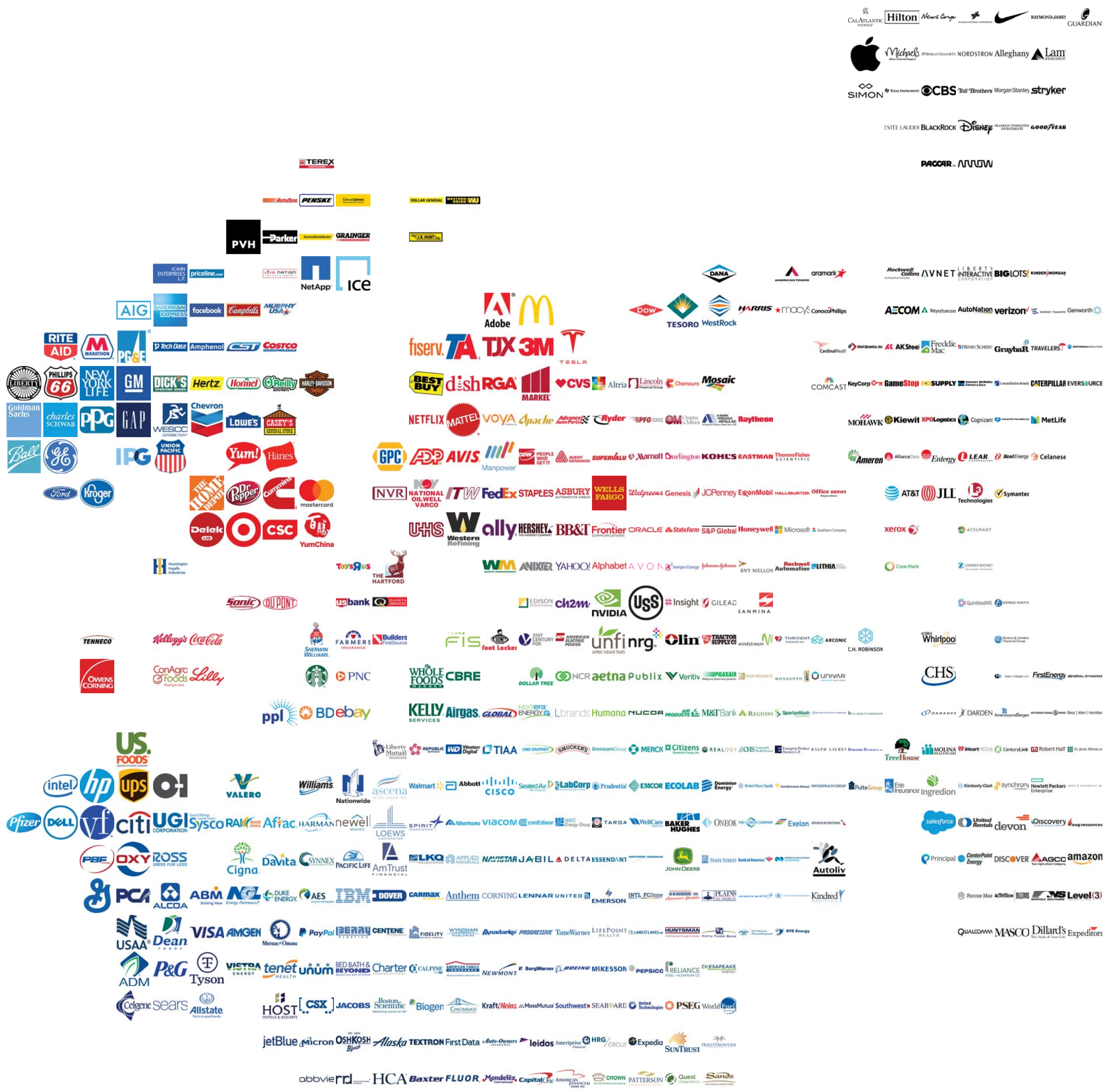
Cells per side
30

Result of the linear assignment.

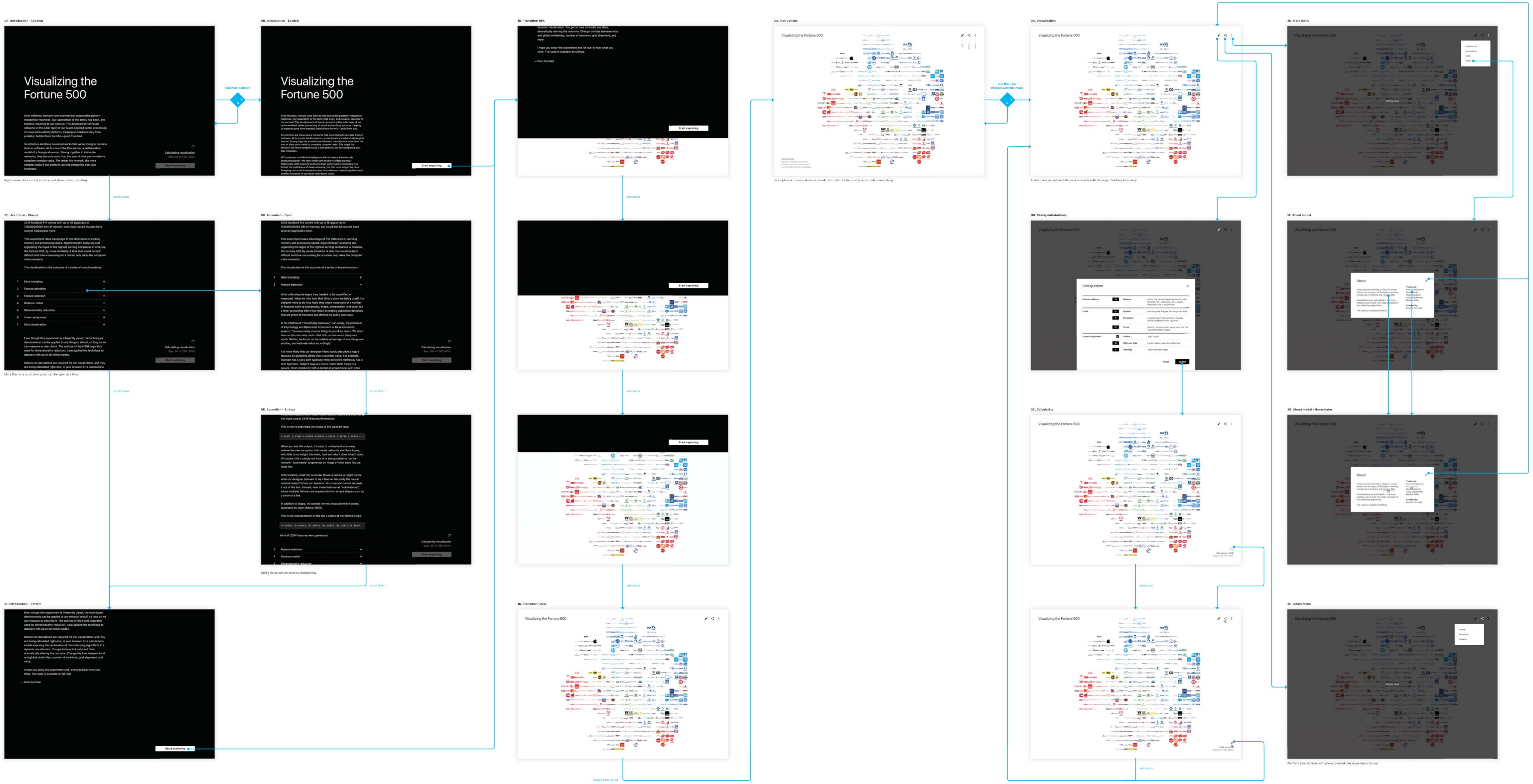


This output is one of many in a “space of solutions.”

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

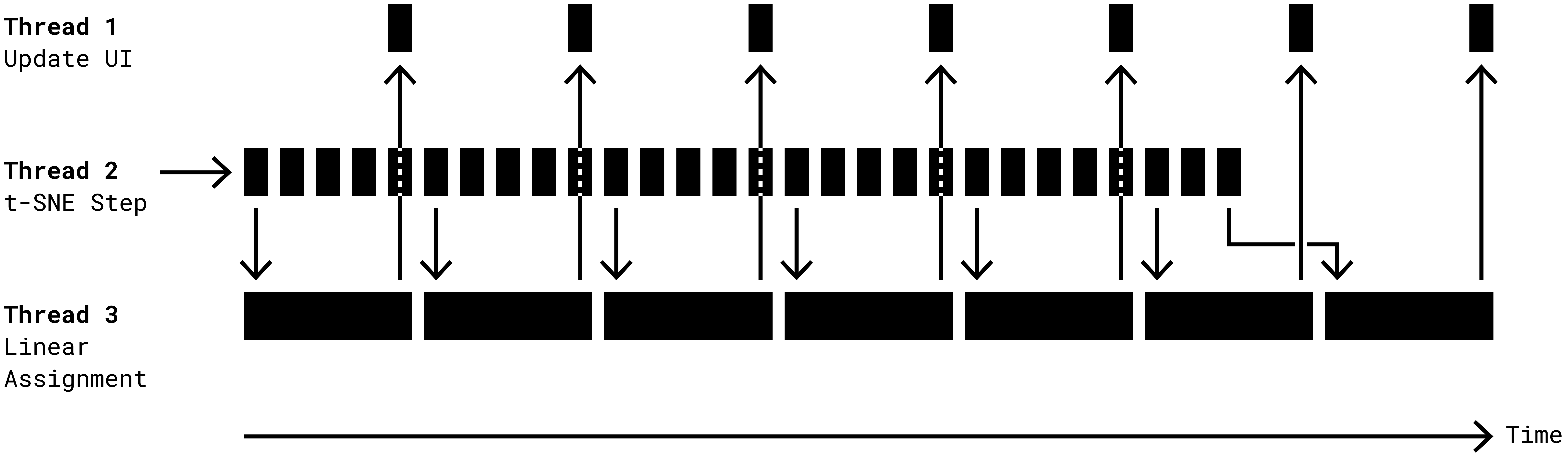
Pre-calculated

---

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

Live

**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

Special thanks to
Hugh Dubberly
Ryan Keisler
Cody Wackerman

Knut Synstad
@knutsynstad

Presentation posted at
knutsynstad.com/fortune500.pdf

Appendix

Visualizing the Space of Logos of the Fortune 500

Abstract
This interactive visualization uses a neural network to analyze a set of logos (in this case the Fortune 500), find similarities and differences, and display the results. Users can adjust the parameters of the display format. The related paper describes how the code works (the analysis and display process). Code is available, enabling others to replicate the process, apply it to their own data sets (other sets of symbols), or to extend it to other uses.

Introduction
Much of what designers are called upon to do involves looking at how artifacts or features "fit-in" and "stand-out". That is, designers are often concerned with similarity and difference—seeing them, understanding them, and manipulating them. For example, a new icon must fit into a system and also be distinct from it.

A common step in developing products and identity systems is to collect examples from competitors and discuss them with clients. Now, software can

Calculating visualization
Step 186 of 250 (74%)

Start exploring

Visualizing the Space of Logos of the Fortune 500

Abstract
This interactive visualization uses a neural network to analyze a set of logos (in this case the Fortune 500), find similarities and differences, and display the results. Users can adjust the parameters of the display format. The related paper describes how the code works (the analysis and display process). Code is available, enabling others to replicate the process, apply it to their own data sets (other sets of symbols), or to extend it to other uses.

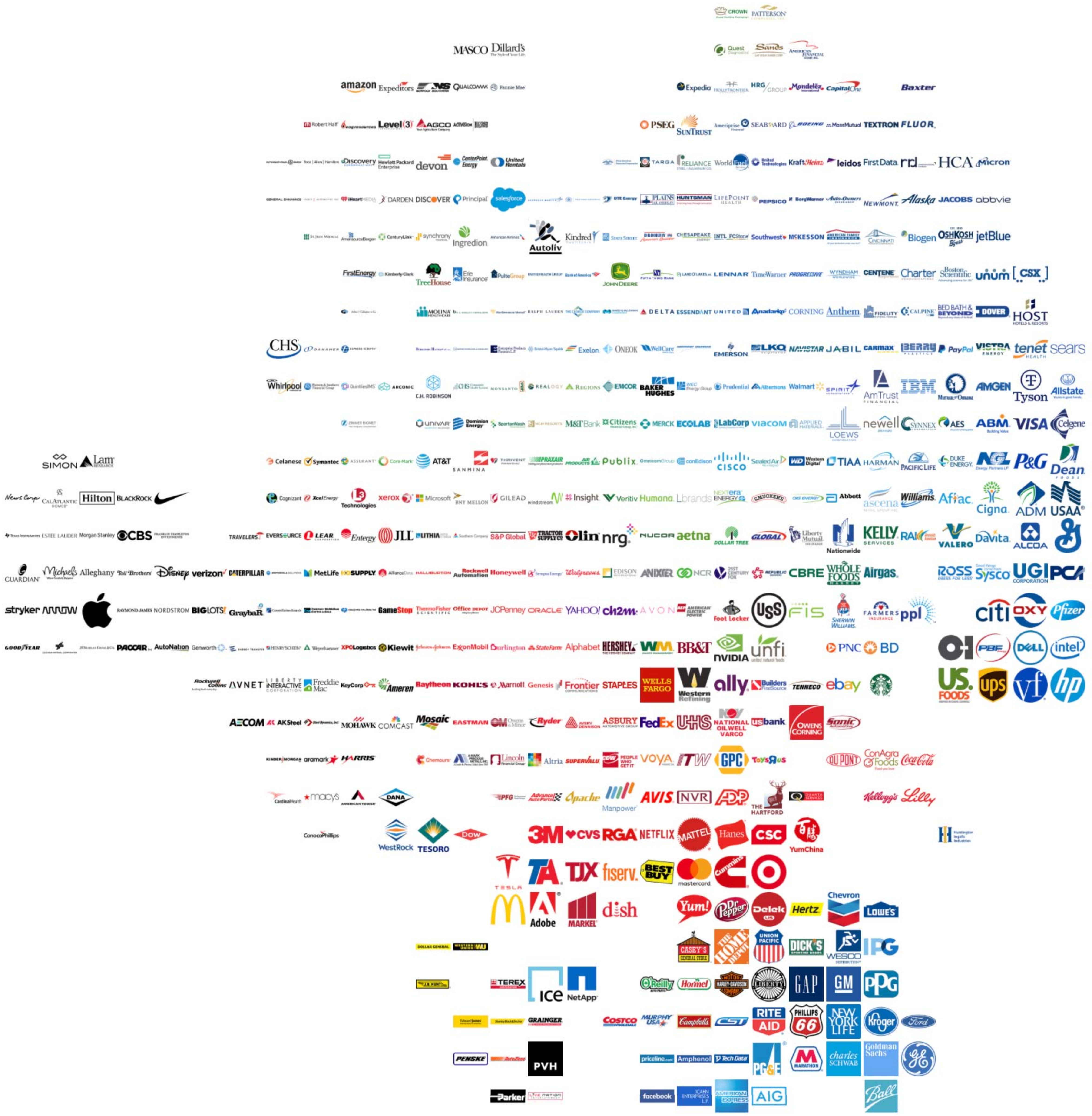
Introduction
Much of what designers are called upon to do involves looking at how artifacts or features "fit-in" and "stand-out". That is, designers are often concerned with similarity and difference—seeing them, understanding them, and manipulating them. For example, a new icon must fit into a system and also be distinct from it.

A common step in developing products and identity systems is to collect examples from competitors and discuss them with clients. Now, software can speed up the process, while also enabling inclusion of much larger data sets

Start exploring

Visualizing the Space of Logos of the Fortune 500

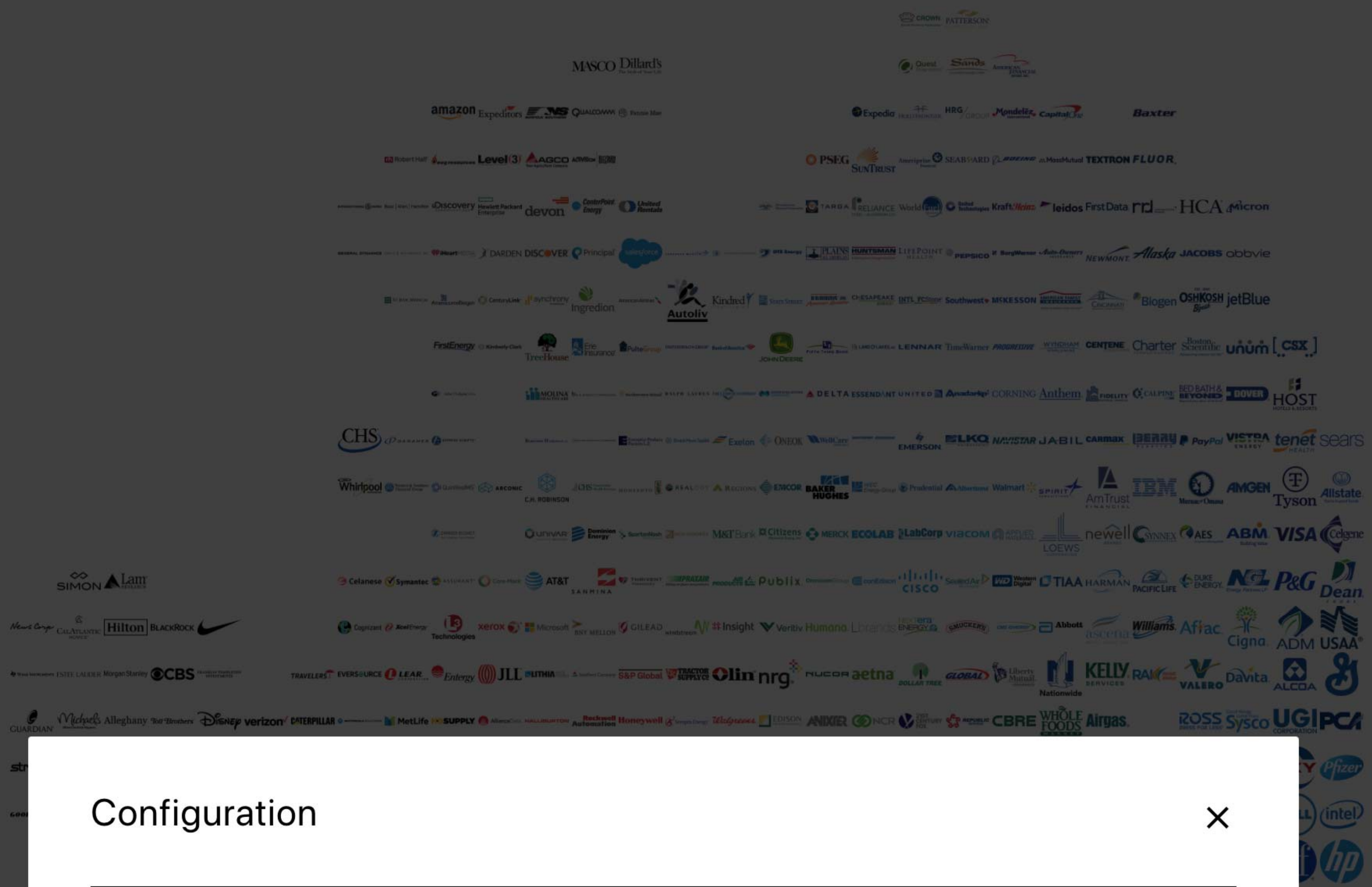
Edit
Share
More



Instructions
- Scroll to change zoom level
- Click-and-drag to move (pan)
- Click the edit icon to manipulate

Visualizing the Space of Logos of the Fortune 500

Visualizing the Space of Logos of the Fortune 500

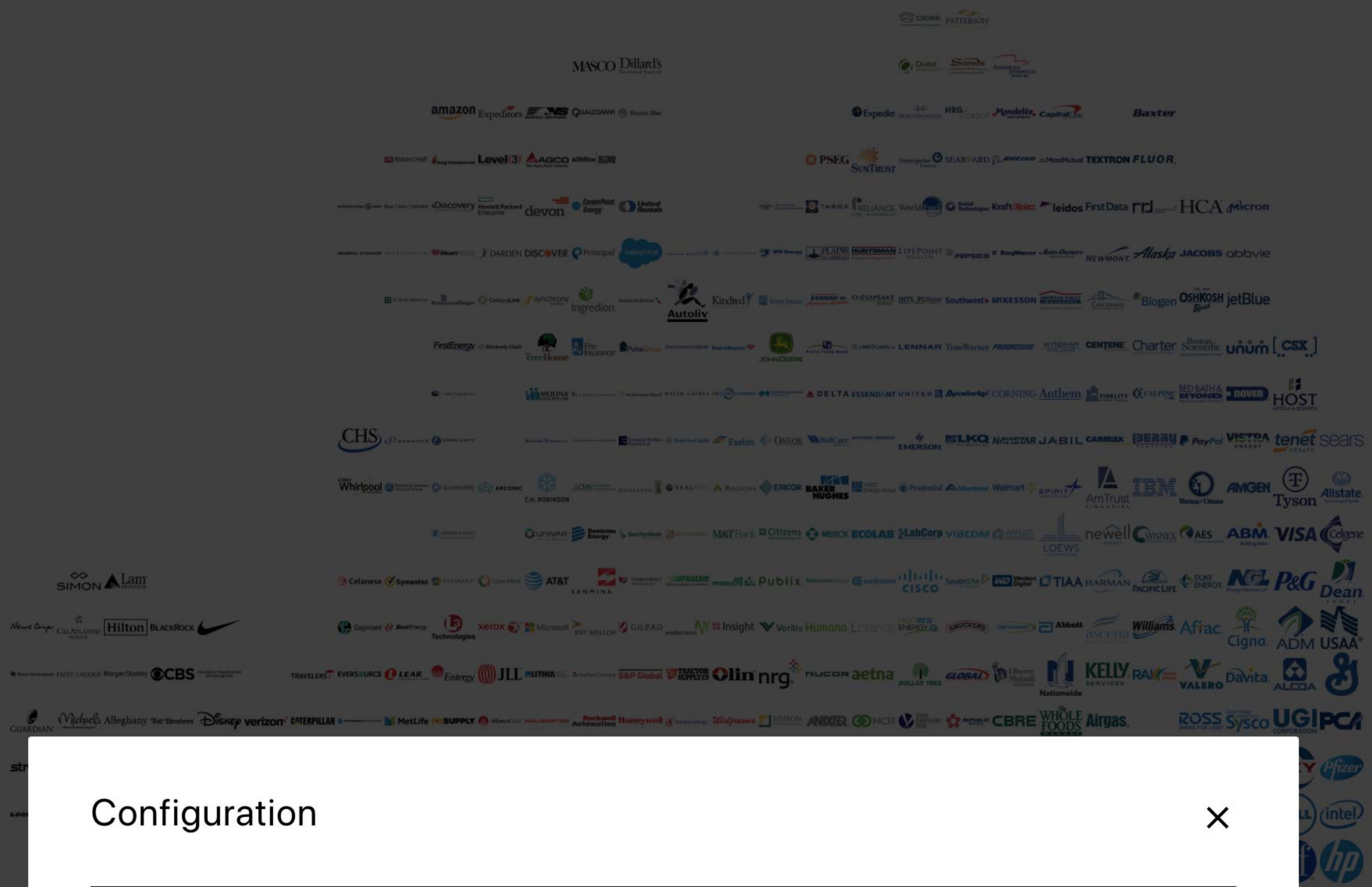


Configuration

t-SNE	10	**Epsilon**	Learning rate.
	15	**Perplexity**	A guess about the number of visually similar neighbors each logo has.
	250	**Steps**	Solution improves with every step, but the calculation takes longer.
Linear assignment	☑	**Enable**	Align to grid.
	30	**Cells per side**	Larger values mean less distortion
	2	**Padding**	Space between logos

Reset **Update**

Visualizing the Space of Logos of the Fortune 500



Configuration

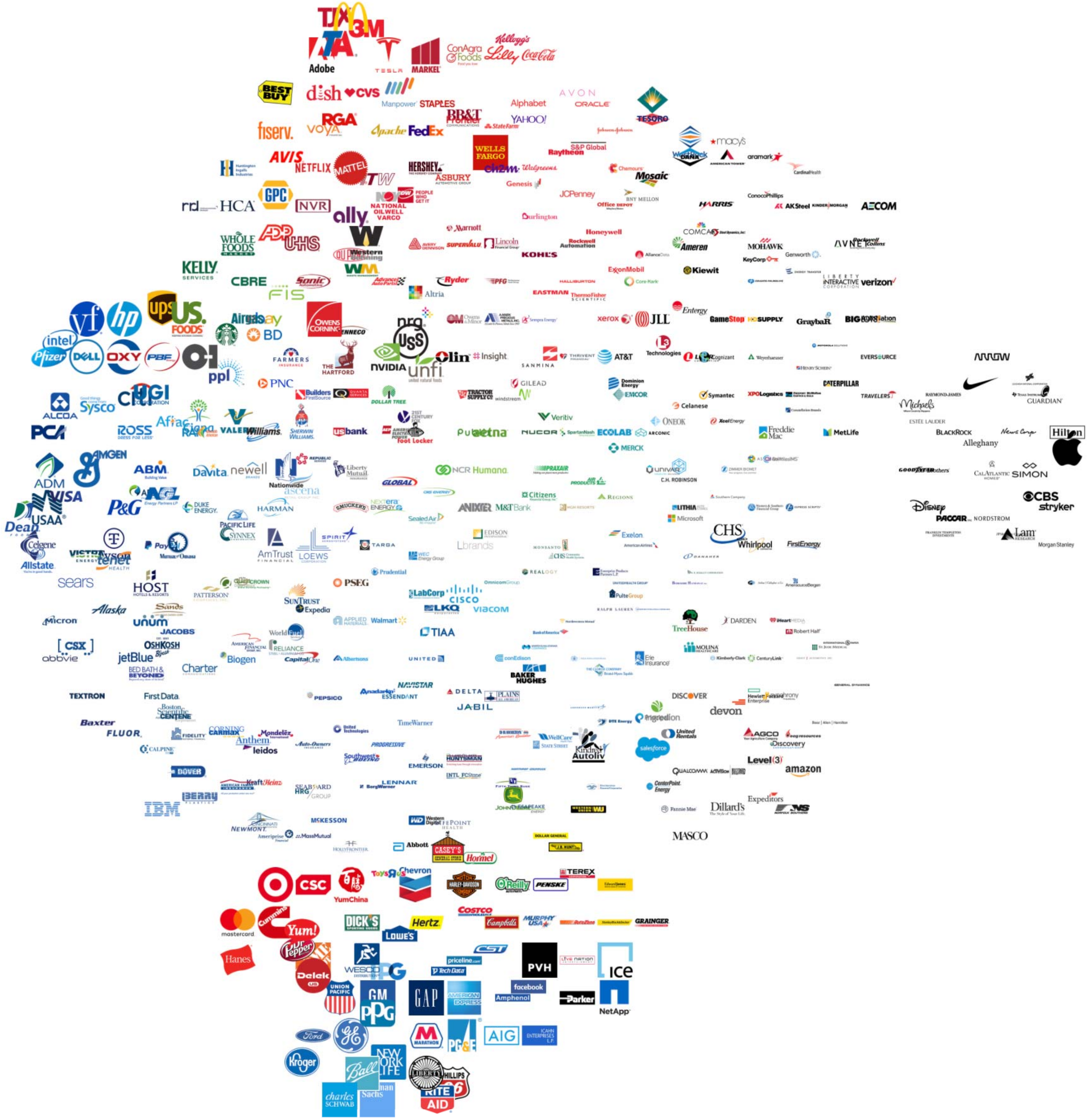
t-SNE	10	**Epsilon**	Learning rate.
	10	**Perplexity**	A guess about the number of visually similar neighbors each logo has.
	750	**Steps**	Solution improves with every step, but the calculation takes longer.
Linear assignment	☑	**Enable**	Align to grid.
	30	**Cells per side**	Larger values mean less distortion
	2	**Padding**	Space between logos

Reset **Update**

Visualizing the Space of Logos of the Fortune 500

Calculating visualization
Step 278 of 750 (37%)

Visualizing the Space of Logos of the Fortune 500



Calculating visualization
Step 627 of 750 (84%)

Visualizing the Space of Logos of the Fortune 500

Visualizing the Space of Logos of the Fortune 500

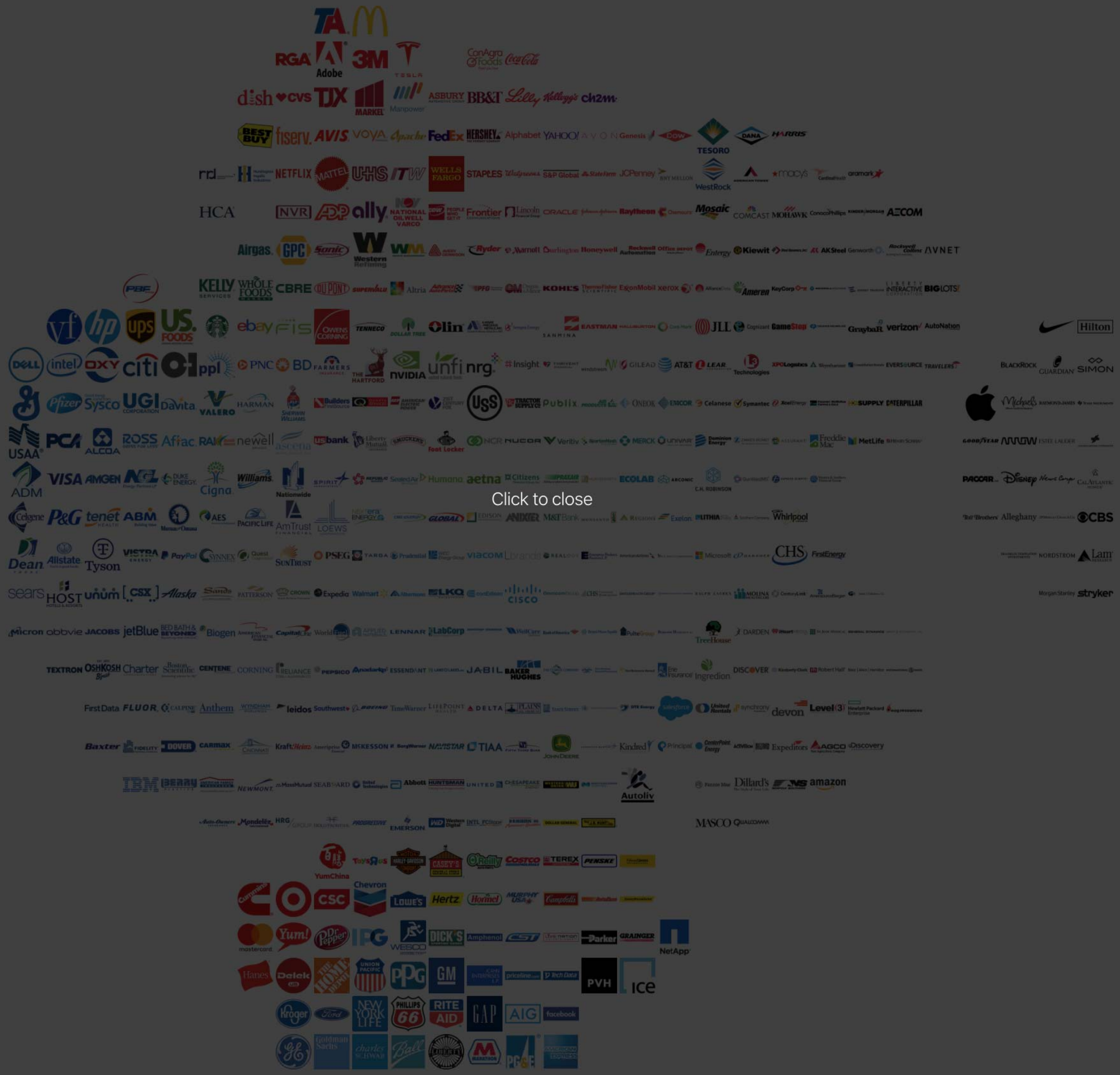
Twitter

Facebook

LinkedIn

Pinterest

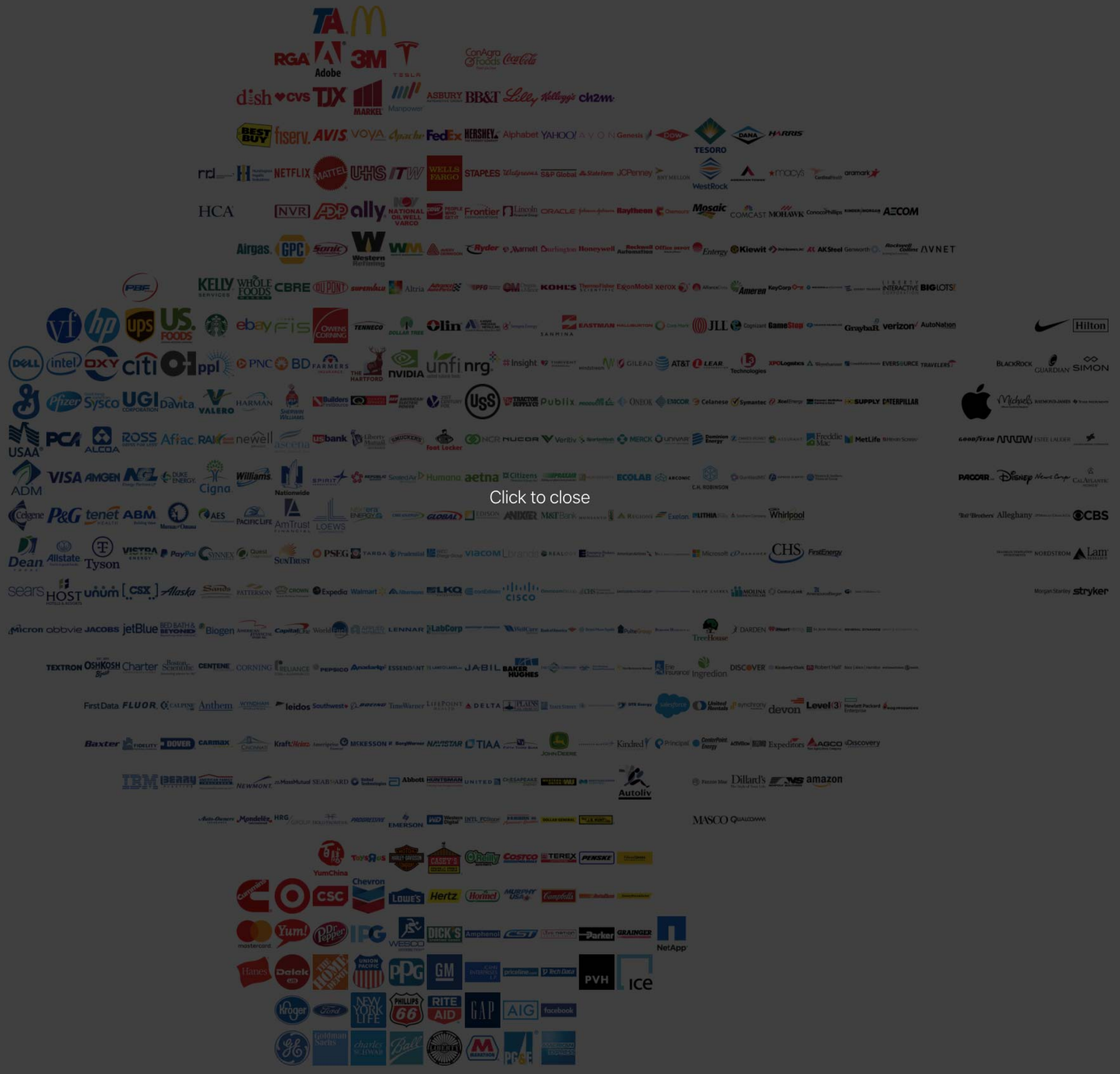
Email



Click to close

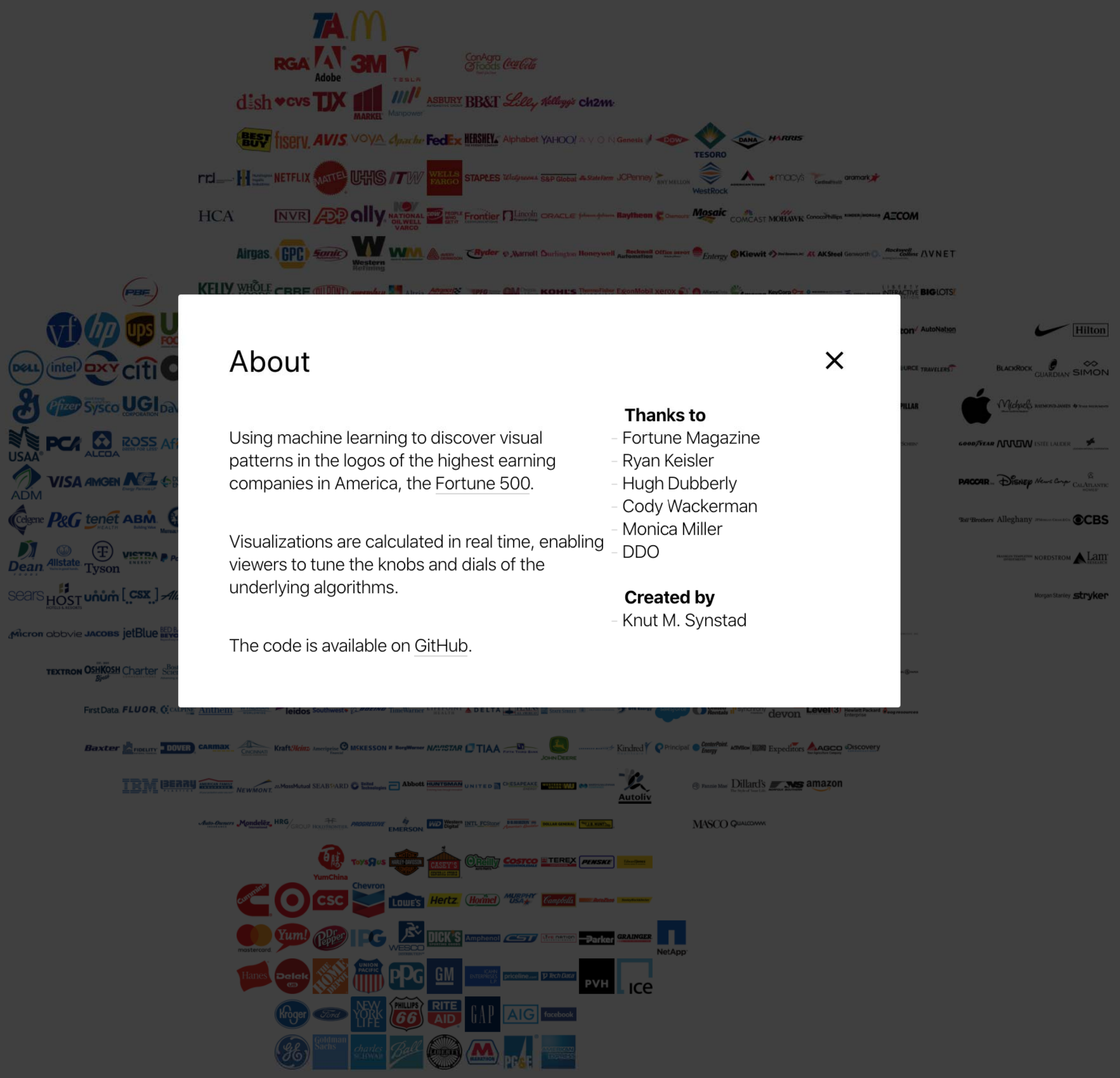
Visualizing the Space of Logos of the Fortune 500

Visualizing the Space of Logos of the Fortune 500



Click to close

Visualizing the Space of Logos of the Fortune 500



About

Using machine learning to discover visual patterns in the logos of the highest earning companies in America, the Fortune 500.

Visualizations are calculated in real time, enabling viewers to tune the knobs and dials of the underlying algorithms.

The code is available on GitHub.

Thanks to

- Fortune Magazine
- Ryan Keisler
- Hugh Dubberly
- Cody Wackerman
- Monica Miller
- DDO

Created by

- Knut M. Synstad