

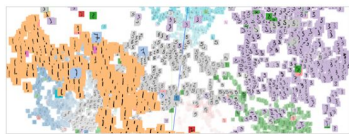
# Visualization for Machine Learning



Fernanda Viégas @viegasf  
Martin Wattenberg @wattenberg  
Google Brain

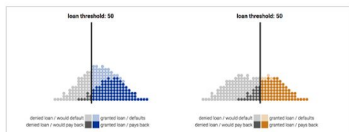
### Embedding Projector

an [open source](#), visualization tool for high-dimensional data



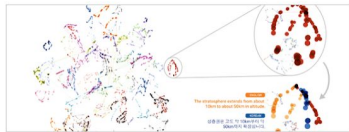
### Fairness in ML

Try different tradeoffs yourself to understand issues around fairness and machine learning.



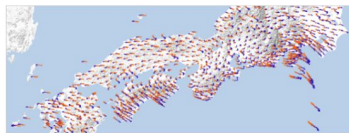
### Machine Translation

Visualizing hints that a translation network learns an "interlingua", or universal language.



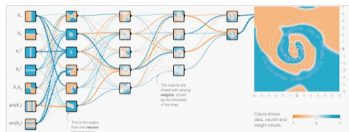
### Geodetic Velocities Visualization

an [open source](#) visualization of earthquake cycle physics



### TensorFlow Playground

an [open source](#), transparent neural net you can play with in your browser



### Unfiltered News

see news coverage around the world and spot underreported stories (a collaboration with [figaw](#))



### TensorFlow Graph Visualizer

an [open source](#), high-level view of TensorFlow computation graphs



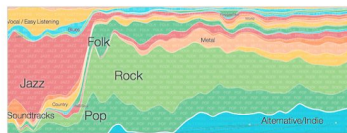
### Periodic Table

a twist on the classic visualization of the atomic elements



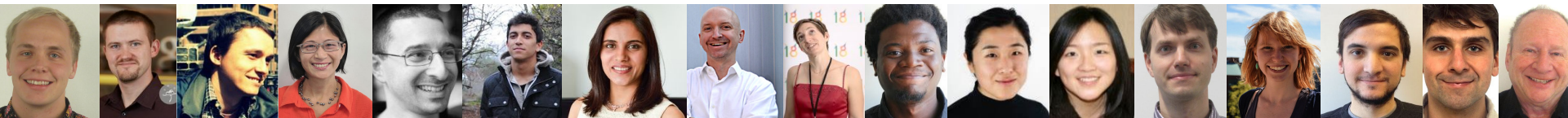
### Music Timeline

see how different musical genres became popular over time, and discover artists in each genre **UPDATED WEEKLY**



### Digital Attack Map

see live data on denial-of-service attacks across the world, and observe historical patterns **UPDATED DAILY**

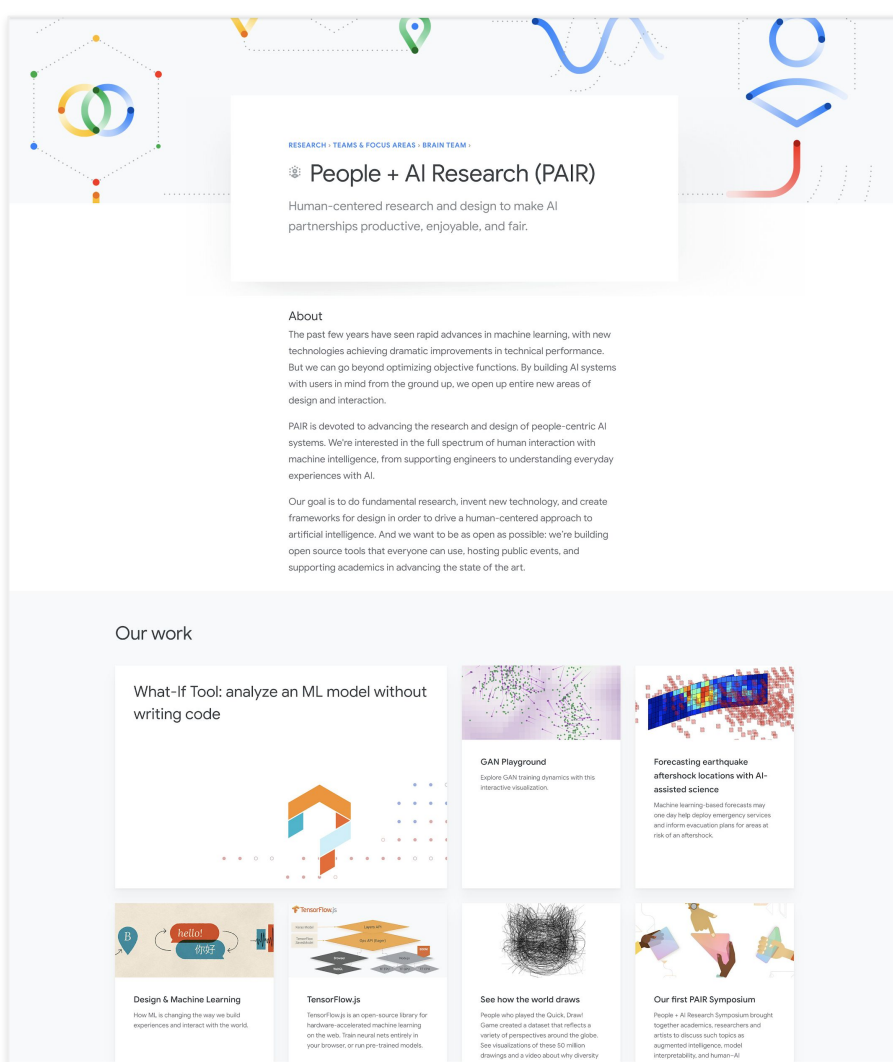


# PAIR

## People + AI Research

Bringing Design Thinking and HCI  
to Machine Learning

[google.ai/pair](https://google.ai/pair)



**RESEARCH • TEAMS & FOCUS AREAS • BRAIN TEAM •**

### People + AI Research (PAIR)

Human-centered research and design to make AI partnerships productive, enjoyable, and fair.

#### About


The past few years have seen rapid advances in machine learning, with new technologies achieving dramatic improvements in technical performance. But we can go beyond optimizing objective functions. By building AI systems with users in mind from the ground up, we open up entire new areas of design and interaction.

PAIR is devoted to advancing the research and design of people-centric AI systems. We're interested in the full spectrum of human interaction with machine intelligence, from supporting engineers to understanding everyday experiences with AI.

Our goal is to do fundamental research, invent new technology, and create frameworks for design in order to drive a human-centered approach to artificial intelligence. And we want to be as open as possible: we're building open source tools that everyone can use, hosting public events, and supporting academics in advancing the state of the art.


#### Our work

##### What-If Tool: analyze an ML model without writing code



##### GAN Playground

Explore GAN training dynamics with this interactive visualization.




##### Forecasting earthquake aftershock locations with AI-assisted science

Machine learning-based forecasts may one day help deploy emergency services and inform evacuation plans for areas at risk of an aftershock.



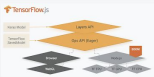
##### Design & Machine Learning

How ML is changing the way we build experiences and interact with the world.




##### TensorFlow.js

TensorFlow.js is an open-source library for hardware-accelerated machine learning on the web. Train neural nets entirely in your browser, or run pre-trained models.




##### See how the world draws

People who played the Quick, Draw! Game created a dataset that reflects a variety of perspectives around the globe. See visualizations of three 10 million drawings and a video about why diversity



##### Our first PAIR Symposium

People + AI Research Symposium brought together academics, researchers and artists to discuss such topics as augmented intelligence, model interpretability, and human-AI



# Today's Agenda

## What is data visualization?

How does it work? What are some best practices?

## How has visualization been applied to ML?

Overview of the landscape

Special case: high-dimensional data

# Goals

## Understand state of the art

Known best practices in visualization

Broad survey of existing applications to ML

## Apply visualizations in your own situation

References to tools and libraries

References to literature

# What is data visualization?

Transform data into visual encodings

## What is it good for?

Data exploration

Scientific insight

Communication

Education

## How to ensure it works well?

Engage the visual system in smart ways

Take advantage of pre-attentive processing

# What is data visualization?

Transform data into visual marks

## What is it good for?

Data exploration

Scientific insight

Communication

Education

## How is it different from statistics?

**Vis:** no specific question necessary

**Classic Stats:** you investigate a specific question\*

**Vis & Stats:** wonderful, complementary partners

## How to ensure it works well?

Engage the visual system in smart ways

Take advantage of pre-attentive processing

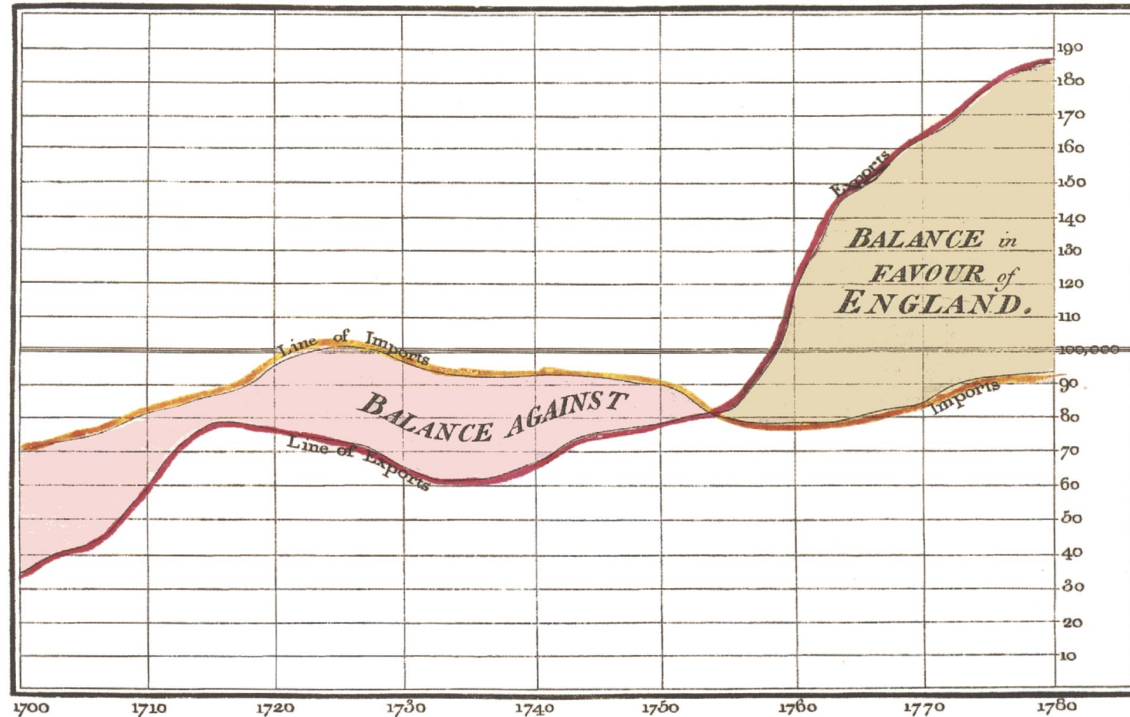
\*OK, maybe not in EDA, but visualization is the key technique there anyway!

Predates computers...



# William Playfair (1786)

Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.



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

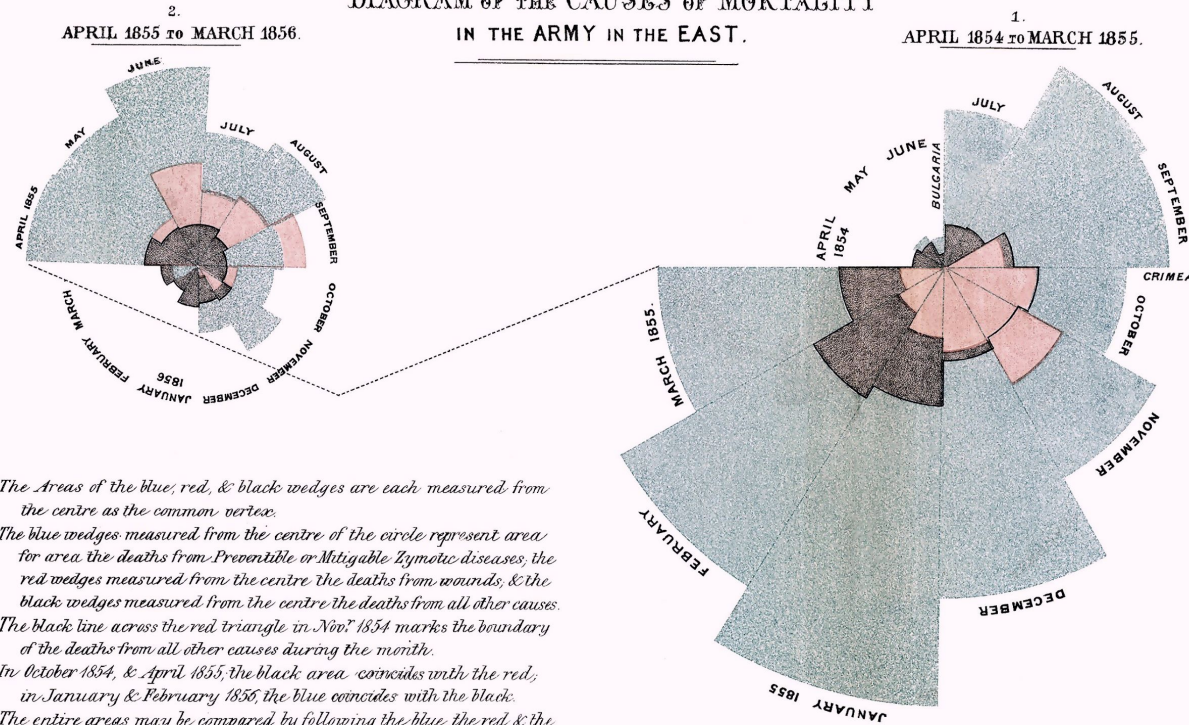
Aside from revolutionizing graphics, Playfair was an economist, engineer, and even a secret agent.

*The Bottom line is divided into Years, the Right hand line into £10,000 each.*  
Published as the Act directs, 1<sup>st</sup> May 1786, by W<sup>m</sup> Playfair. *Neale sculpit 352, Strand, London.*

(Image: Wikipedia)

# Florence Nightingale (1858)

DIAGRAM OF THE CAUSES OF MORTALITY  
IN THE ARMY IN THE EAST.



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.  
The blue wedges measured from the centre of the circle represent area for area the deaths from Preventable or Mitigable Zymotic diseases; the red wedges measured from the centre the deaths from wounds; & the black wedges measured from the centre the deaths from all other causes.  
The black line across the red triangle in Nov? 1854 marks the boundary of the deaths from all other causes during the month.  
In October 1854, & April 1855, the black area coincides with the red; in January & February 1856, the blue coincides with the black.  
The entire areas may be compared by following the blue, the red & the black lines enclosing them.

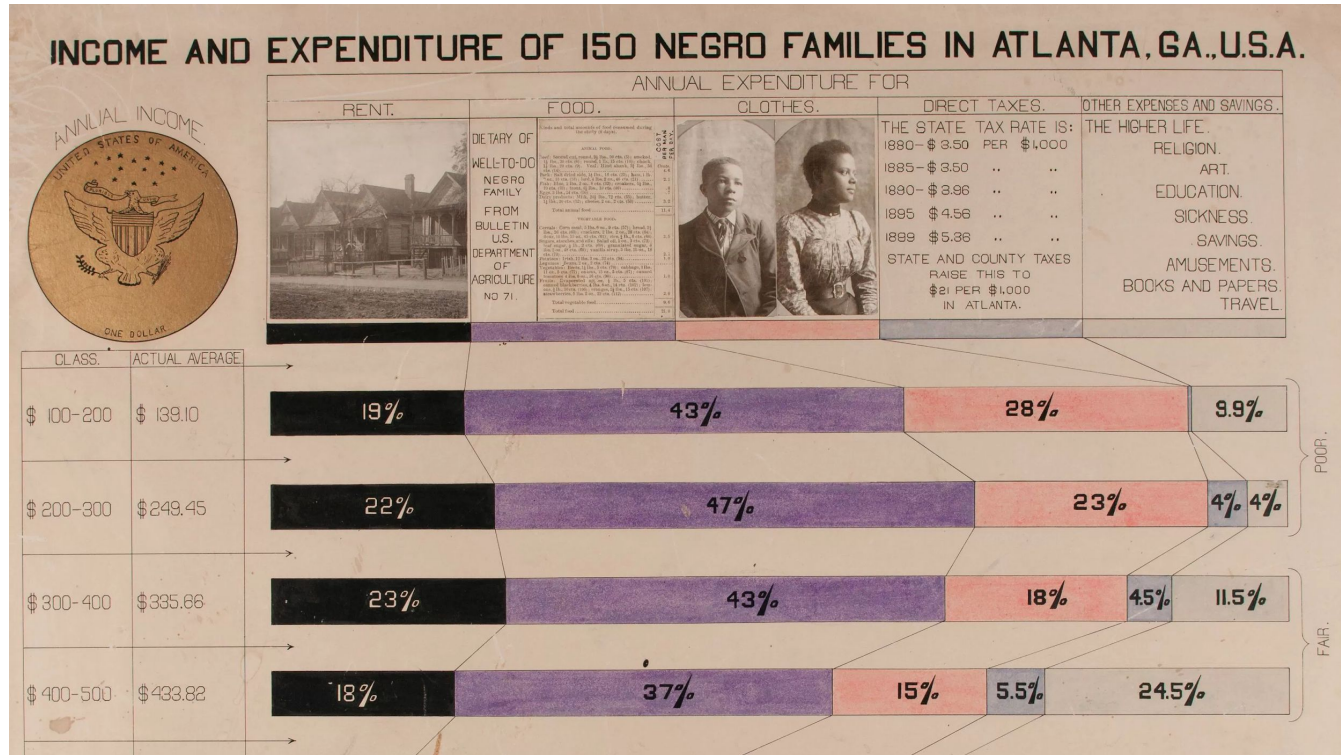
These charts led to the adoption of better hygiene / sanitary practices in military medicine, saving millions of lives.

Arguably the most effective visualization ever!

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

(Image: Wikipedia)

# W. E. B. Du Bois (1900)



For 1900 World's Fair, a compendium of visualizations. Many new chart types!

Excellent example of visualization aimed at political change.

# What do these have in common?

Using special properties of the visual system to help us think.

# What do these have in common?

Using special properties of the visual system to help us think.

Our visual system is like a GPU

- Incredibly good at a few special tasks
- With work, can be repurposed for more general situations

# What do these have in common?

Using special properties of the visual system to help us think.

Our visual system is like a GPU

- Incredibly good at a few special tasks
- With work, can be repurposed for more general situations

**All visualizations are made from a series of compromises.**

How do visualizations work?

# How do visualizations work?

Find visual encodings that

- Guide viewer's attention
- Communicate data to the viewer
- Let viewer calculate with data

On computer

- Interactive exploration



# How do visualizations work?

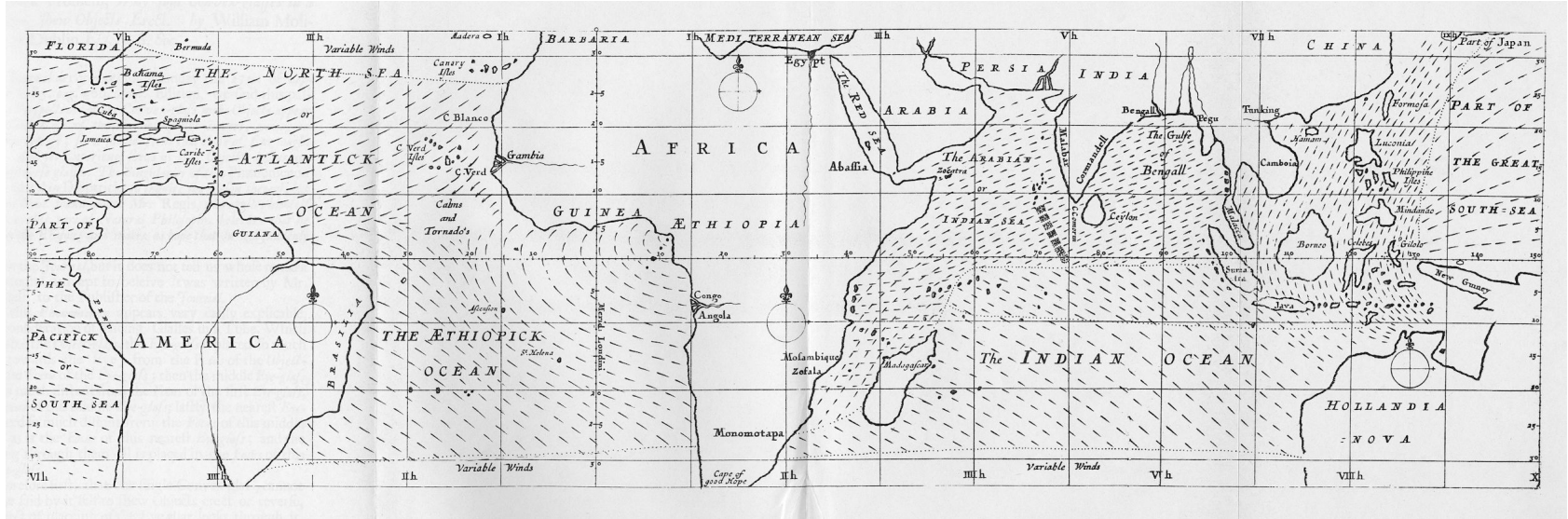
Find visual encodings that

- Guide viewer's attention
- Communicate data to the viewer
- Let viewer calculate with data

On computer

- Interactive exploration

# Encodings: some examples



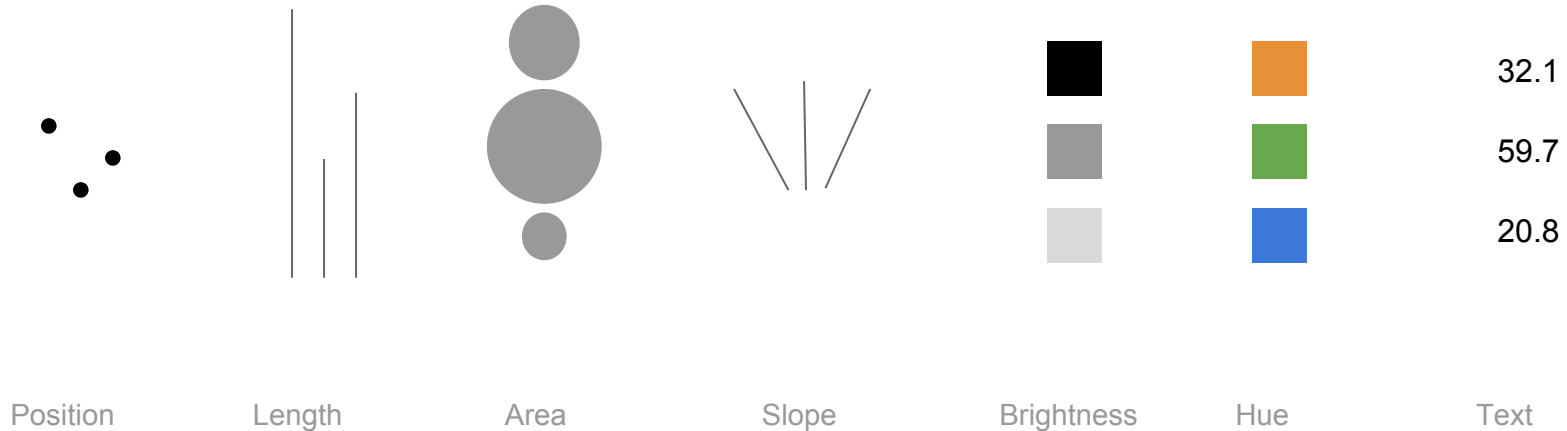
Edmund Halley, 1686

[Comparison A \(2012\)](#): US Wind Map

[Comparison B \(2013\)](#): Earth.nullschool

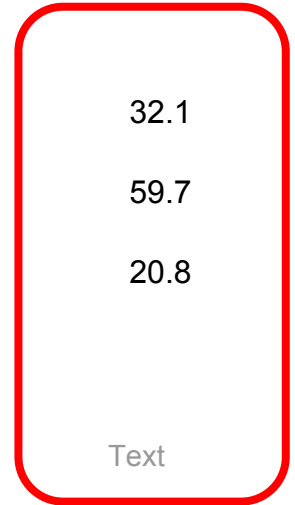
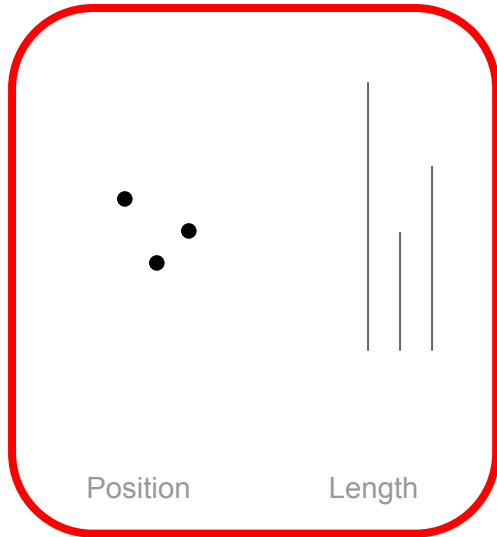
# Encodings: some theory

From perceptual psychology:  
different encodings have different properties.



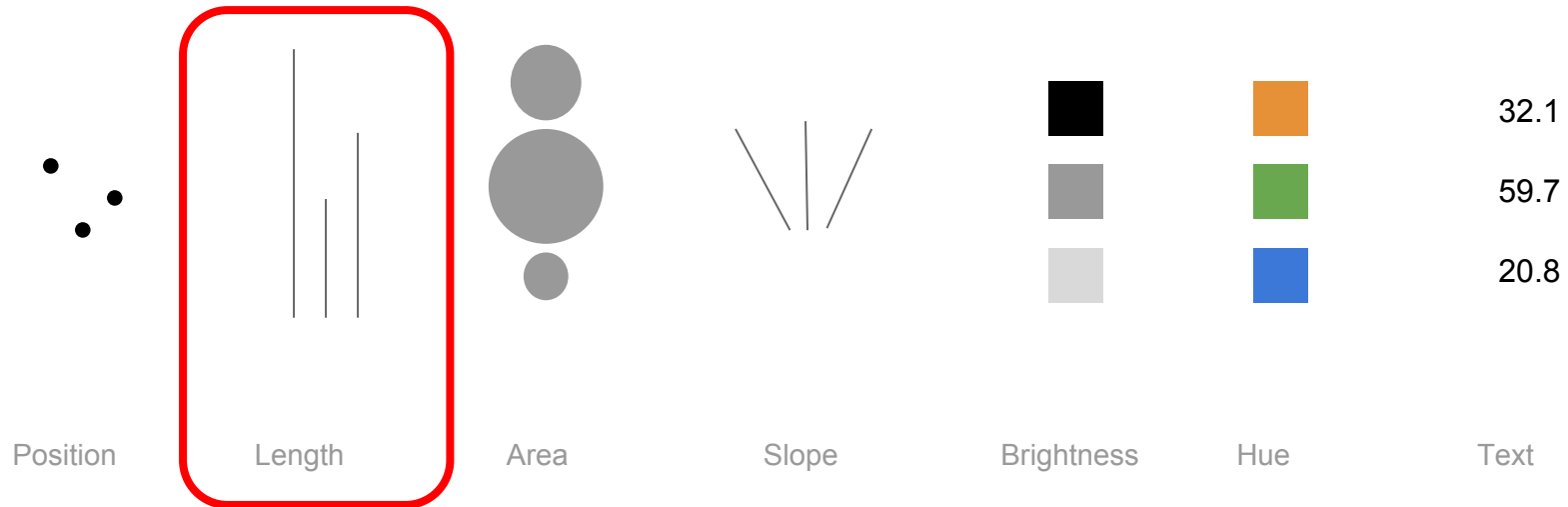
# Encodings: some theory

Good for communicating exact values...



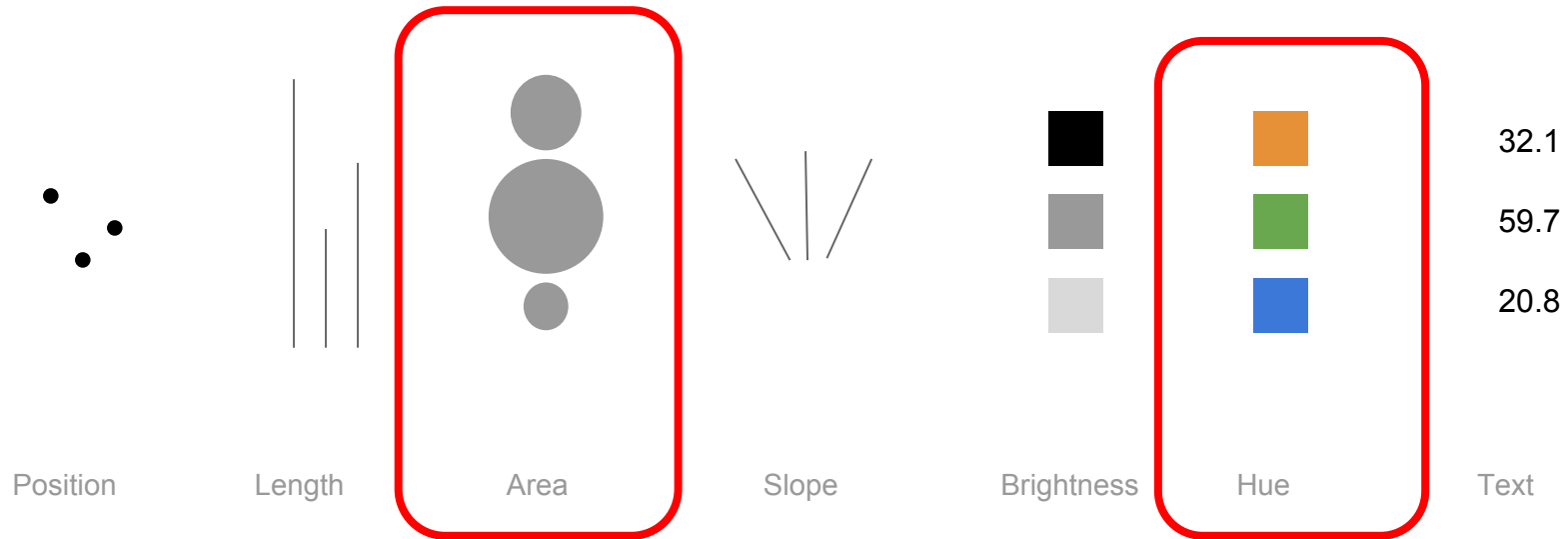
# Encodings: some theory

Good for communicating ratios...



# Encodings: some theory

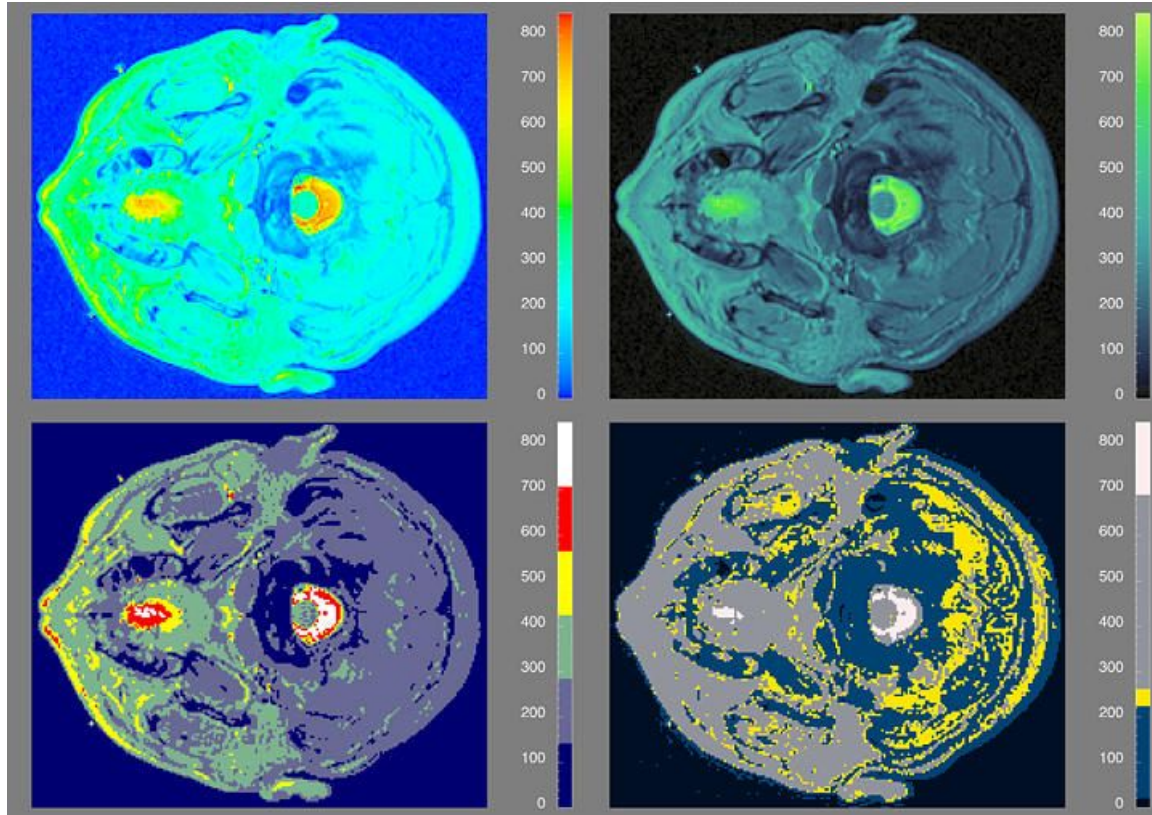
Good for drawing attention...



# Special case: color scales

Intensively studied for decades...

# Rogowitz & Treinish (1996)



Web article:

“Why Should Engineers and Scientists Be Worried About Color?”

Conclusions:

- Rainbow scales: bad
- There is no “best” scale



# Practically speaking...

When in doubt, use the "Color Brewer" site:

<http://colorbrewer2.org>

(Built by Cynthia Brewer, a cartographer)

Number of data classes: 3

how to use | updates | downloads | credits

# COLORBREWER 2.0

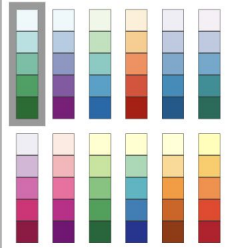
color advice for cartography

## Nature of your data:

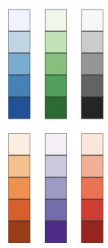
sequential  diverging  qualitative

## Pick a color scheme:

### Multi-hue:



### Single hue:



## Only show:

- colorblind safe
- print friendly
- photocopy safe

## Context:

- roads
- cities
- borders

## Background:

- solid color
- terrain

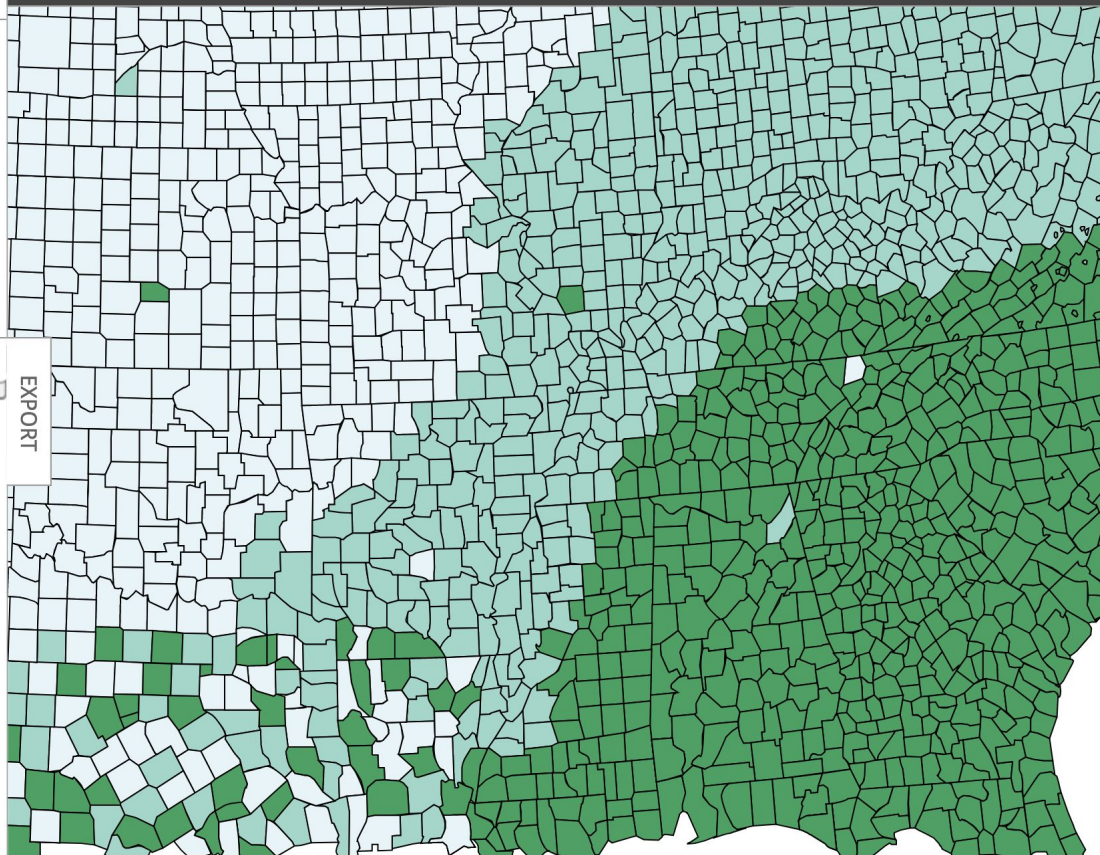
color transparency

## 3-class BuGn



HEX

- #e5f5f9
- #99d8c9
- #2ca25f



And study continues to this day...

A dive into a very recent paper (CHI 2018)

**Somewhere Over the Rainbow: An Empirical Assessment  
of Quantitative Colormaps**

**Yang Liu**  
University of Washington  
Seattle, WA, USA  
yliu0@cs.washington.edu

**Jeffrey Heer**  
University of Washington  
Seattle, WA, USA  
jheer@uw.edu

# Color scales

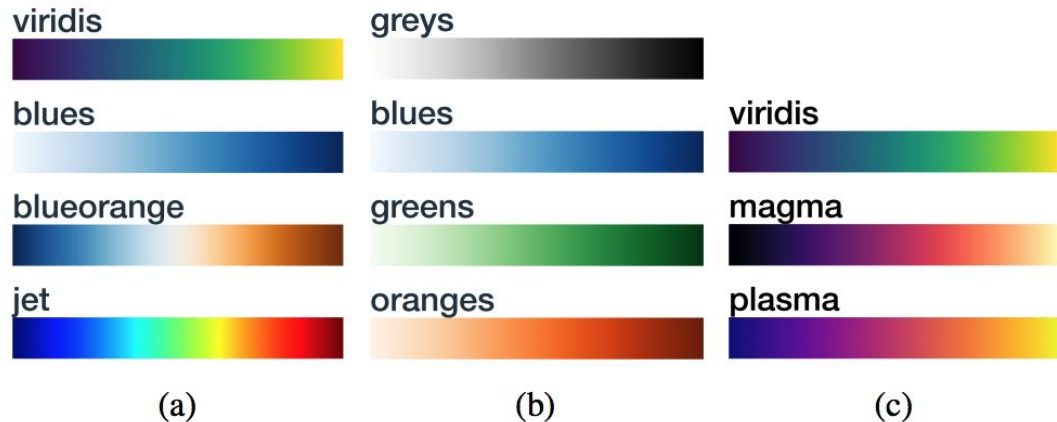


Figure 1: **Colormaps under study.** We evaluate four single-hue, three perceptually-uniform multi-hue, a diverging, and a rainbow colormap(s). We divide them into (a) assorted, (b) single-hue and (c) multi-hue groups, with two colormaps repeated across groups for replication.

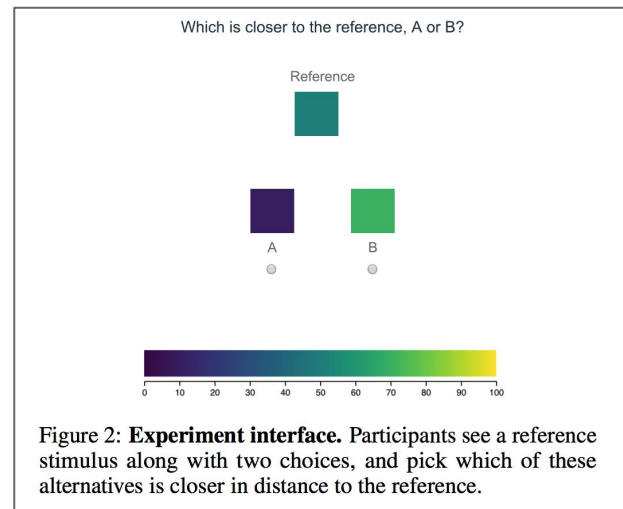
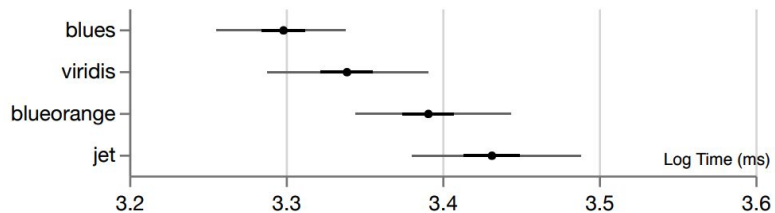
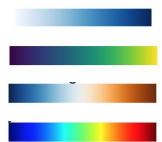
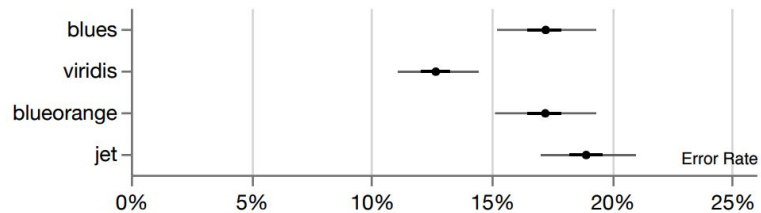


Figure 2: **Experiment interface.** Participants see a reference stimulus along with two choices, and pick which of these alternatives is closer in distance to the reference.

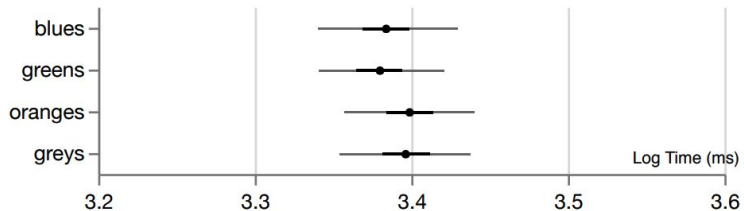
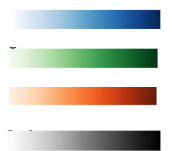
# Color scales



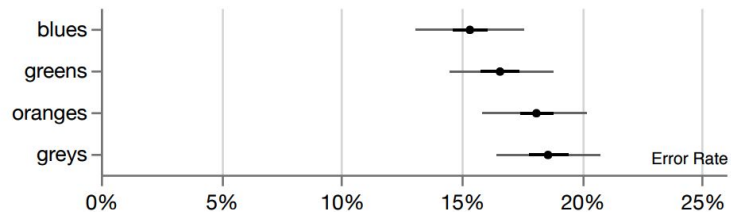
(a) Assorted Colormaps



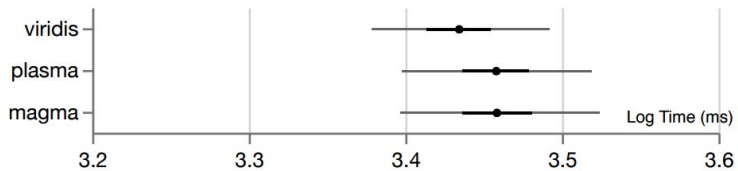
(a) Assorted Colormaps



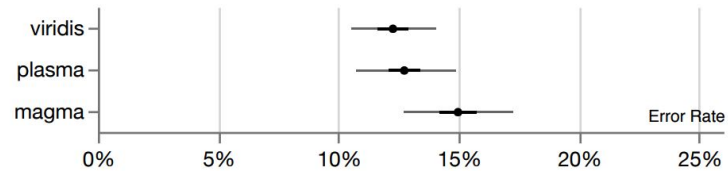
(b) Single-Hue Colormaps



(b) Single-Hue Colormaps



(c) Multi-Hue Colormaps



(c) Multi-Hue Colormaps

# Uh oh, colorblindness... (very common!)

viridis



greys



blues



blues



viridis



blueorange



greens



magma



jet



oranges



plasma



Red-blind protonopia. See <http://www.color-blindness.com/coblis-color-blindness-simulator/>

Guiding attention

Pre-attentive processing

## Count the 5s

987349790275647902894728624092406037070570279072  
803208029007302501270237008374082078720272007083  
247802602703793775709707377970667462097094702780  
927979709723097230979592750927279798734972608027

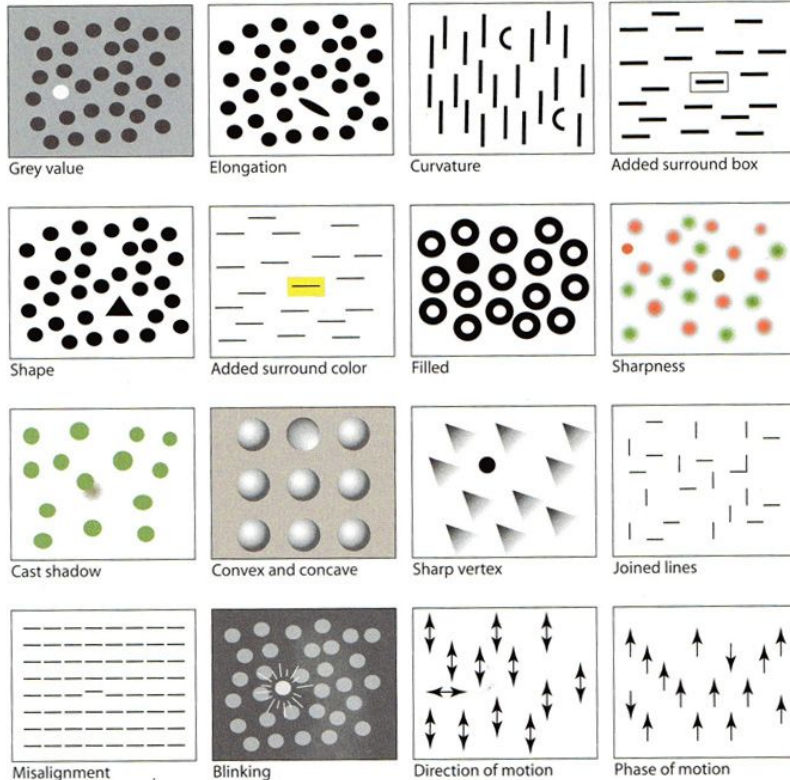


## Count the 5s

987349790275647902894728624092406037070570279072  
803208029007302501270237008374082078720272007083  
247802602703793775709707377970667462097094702780  
927979709723097230979592750927279798734972608027

98734979027**5**647902894728624092406037070**5**70279072  
803208029007302**5**01270237008374082078720272007083  
24780260270379377**5**709707377970667462097094702780  
92797970972309723097**9**5927**5**0927279798734972608027

# Theory: attention



(Colin Ware, Visual Thinking for Design)

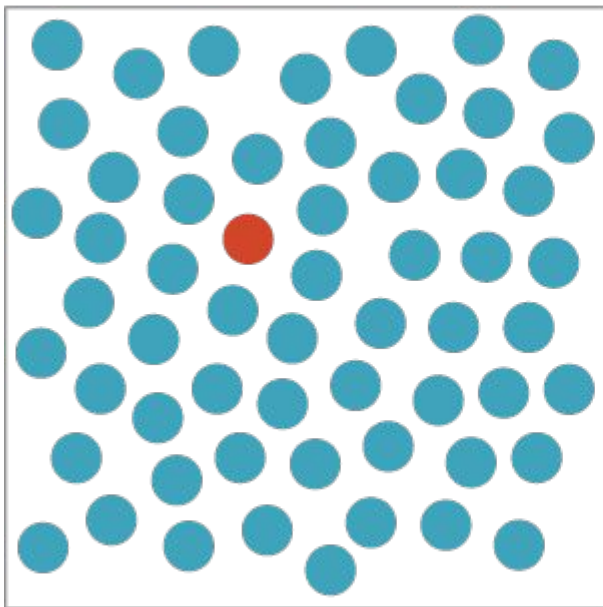
## Pre-attentive processing / "popout"

Under the right circumstances, visual search can be parallel, rather than serial

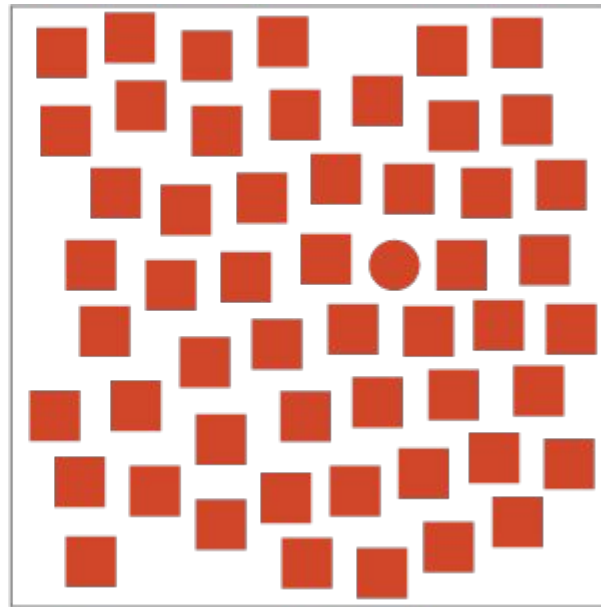
Time to find target does not increase as number of distractors increases

# Pre-Attentive Processing

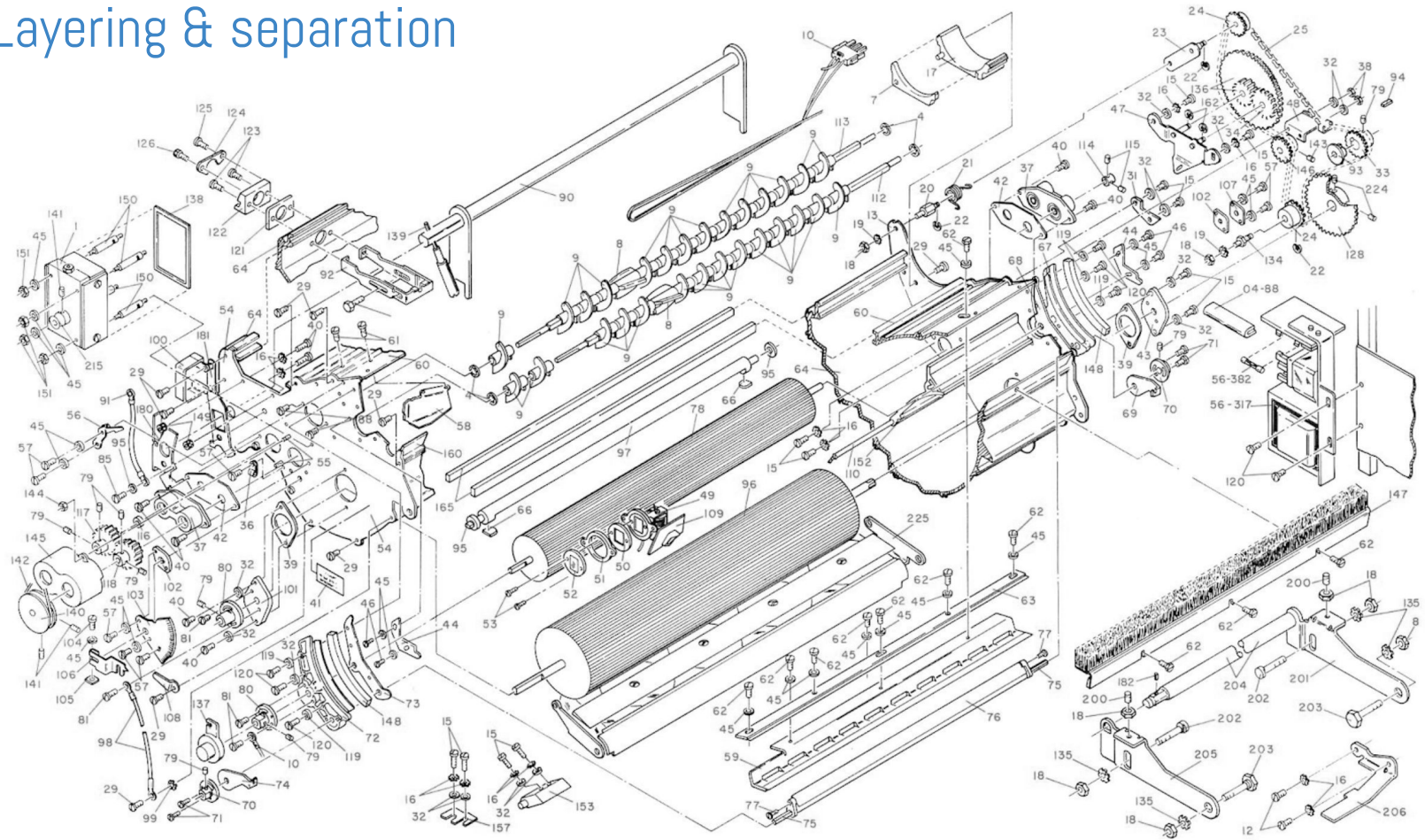
Color



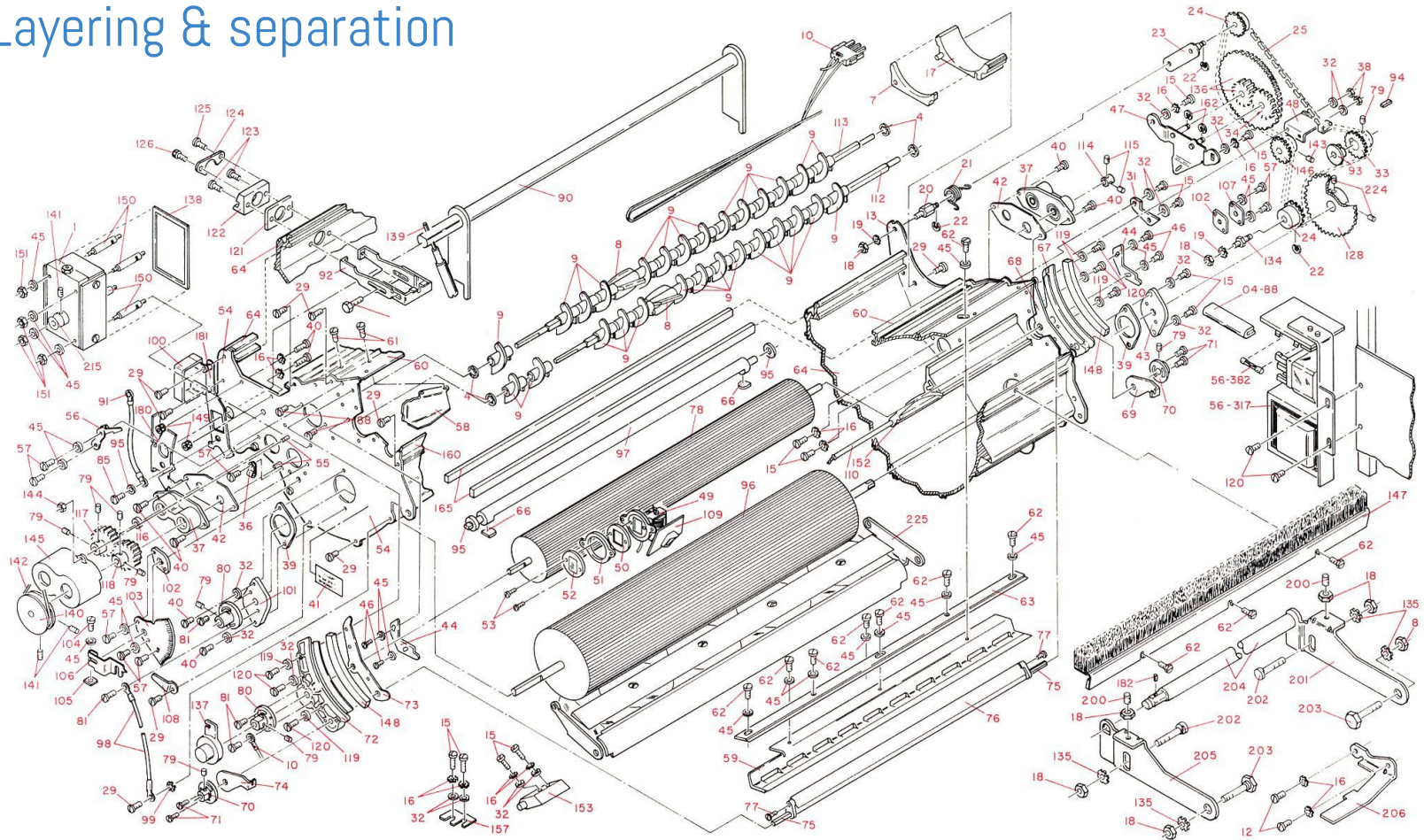
Shape



# Layering & separation



# Layering & separation

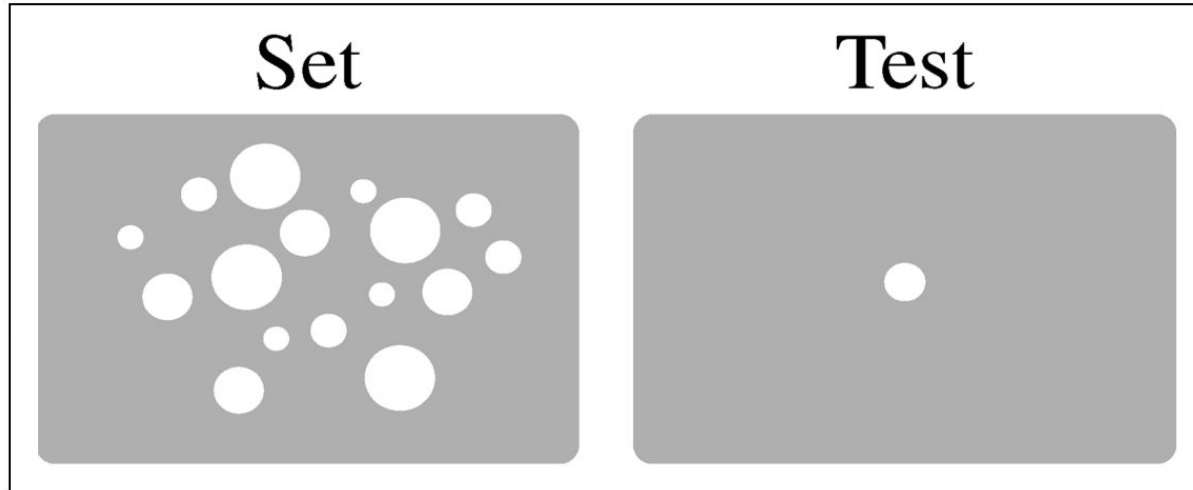


Theory: calculation

# Calculation

Example: we naturally average sizes.

"Seeing Sets: Representation by Statistical Properties." Dan Ariely (2001)



# Calculation

We can do weighted averages, too!

[Example](#)

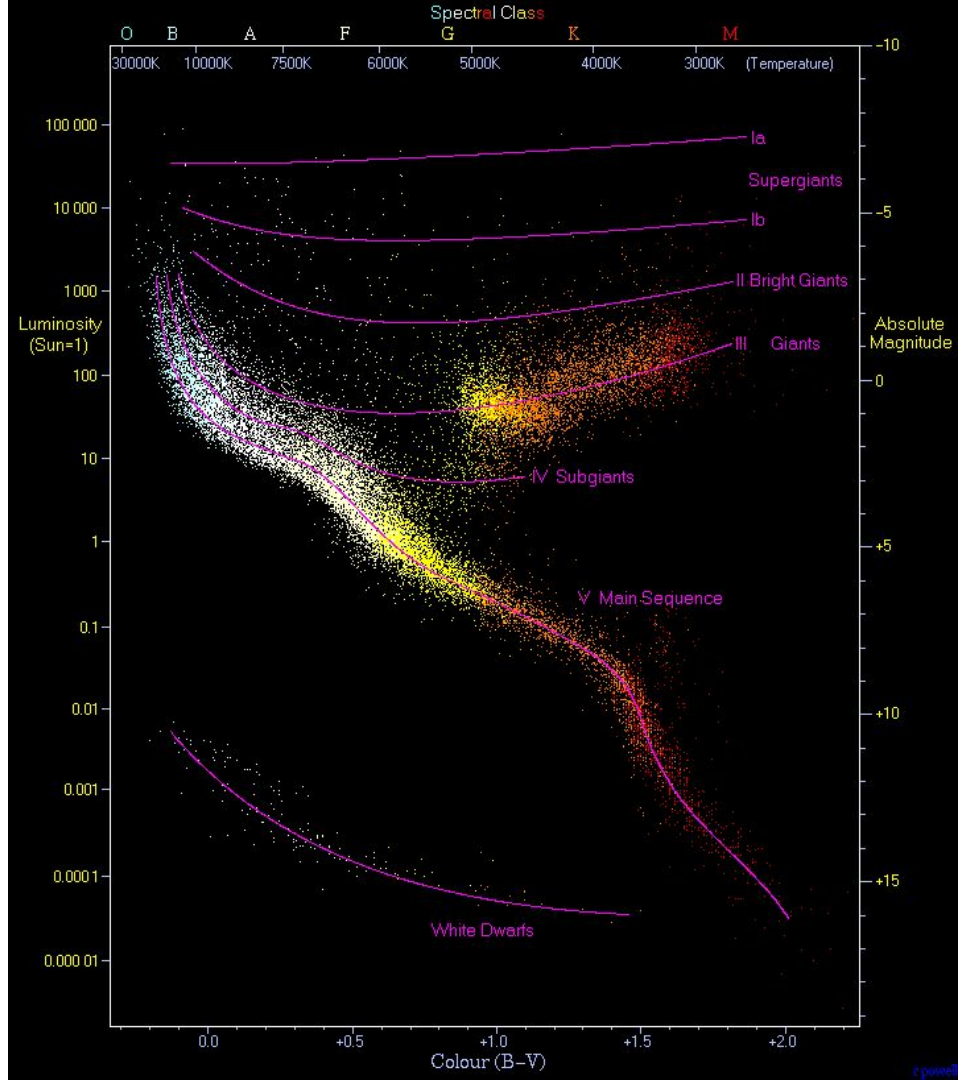


# Calculation

Hertzprung-Russell diagram (via Wikipedia)

Your eye is doing something like kernel density estimation...

Source: Wikipedia



How do visualizations work  
- on computers?

# How do visualizations work - on computers?

Beyond static representations

- Interaction
- Conversation and collaboration

# Theory: interaction

Shneiderman “mantra”:

(1996: “The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations”)

- Overview first
- Zoom and filter
- Details on demand

# Theory: interaction

Shneiderman “mantra”:

(1996: “The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations”)

- Overview first
- Zoom and filter
- Details on demand

Example: [dot maps](#)

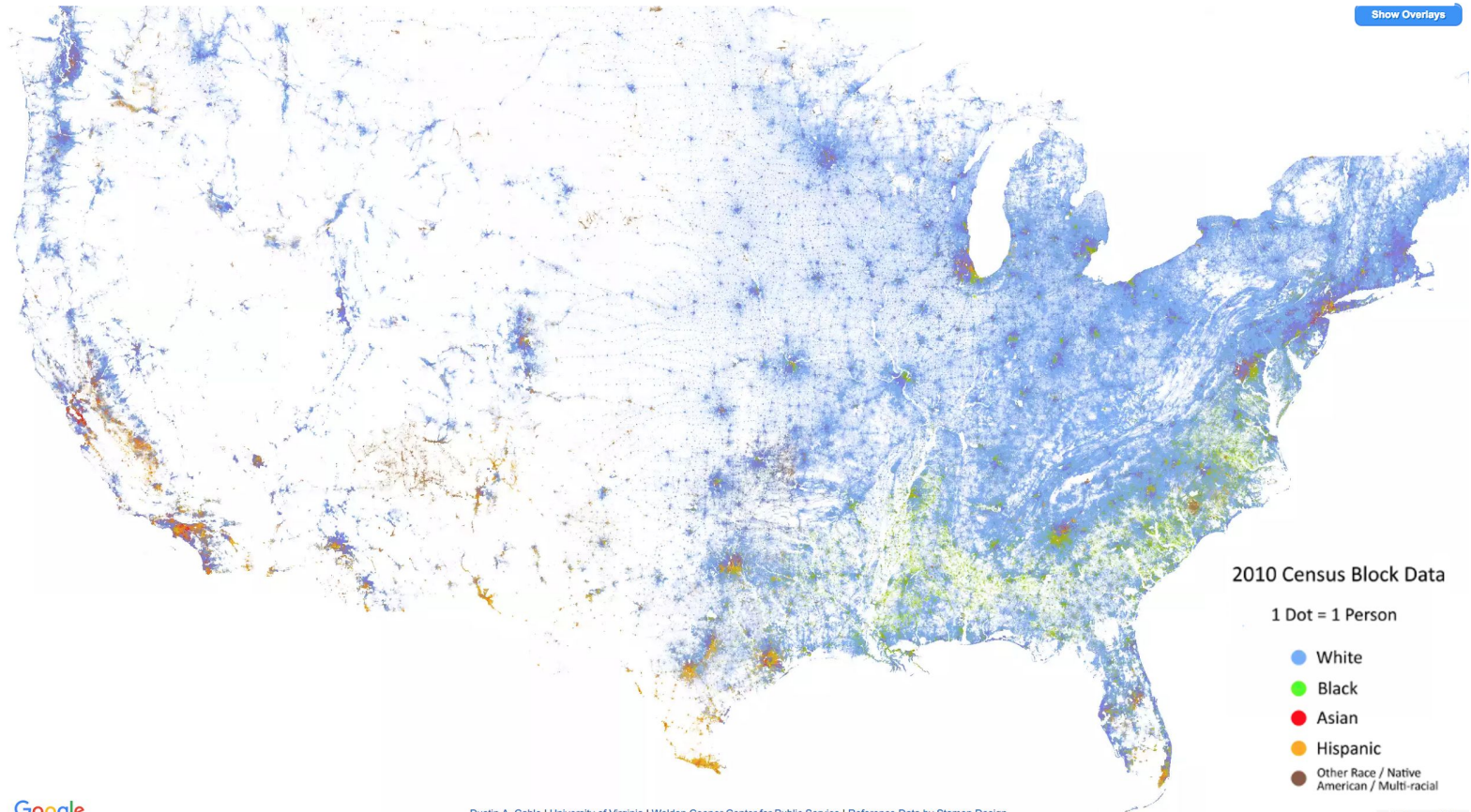
There are many visual design guidelines but the basic principle might be summarized as the Visual Information Seeking Mantra:

Overview first, zoom and filter, then details-on-demand  
Overview first, zoom and filter, then details-on-demand  
Overview first, zoom and filter, then details-on-demand  
Overview first, zoom and filter, then details-on-demand  
Overview first, zoom and filter, then details-on-demand  
Overview first, zoom and filter, then details-on-demand  
Overview first, zoom and filter, then details-on-demand  
Overview first, zoom and filter, then details-on-demand  
Overview first, zoom and filter, then details-on-demand  
Overview first, zoom and filter, then details-on-demand

Each line represents one project in which I found myself rediscovering this principle and therefore wrote it down it as a reminder.

# The Racial Dot Map: One Dot Per Person for the Entire U.S.

[demographics.virginia.edu/DotMap/](http://demographics.virginia.edu/DotMap/)



# Recap: How do visualizations work?

Find visual encodings that

- Guide viewer's attention
- Communicate data to the viewer
- Let viewer calculate with data

On computer

- Interactive exploration

# Some common techniques

That could help in the ML context...

From the simple...



# Case study: the humble table

We've talked to many, many ML teams

Every one of them displayed data in tables

Good design can make a huge difference

# Design thinking in action, a little movie:

## Remove to improve data tables

Joey Cherdarchuk

DarkHorse Analytics

**Remove**  
to improve  
the **data tables** edition

# Key points

- Structure & hierarchy
- Alignment
- Typography
- Color

*These all apply to more complicated visualizations!*

# Some common techniques

That could help in the ML context...

# Data density: small multiples



## Drought's Footprint

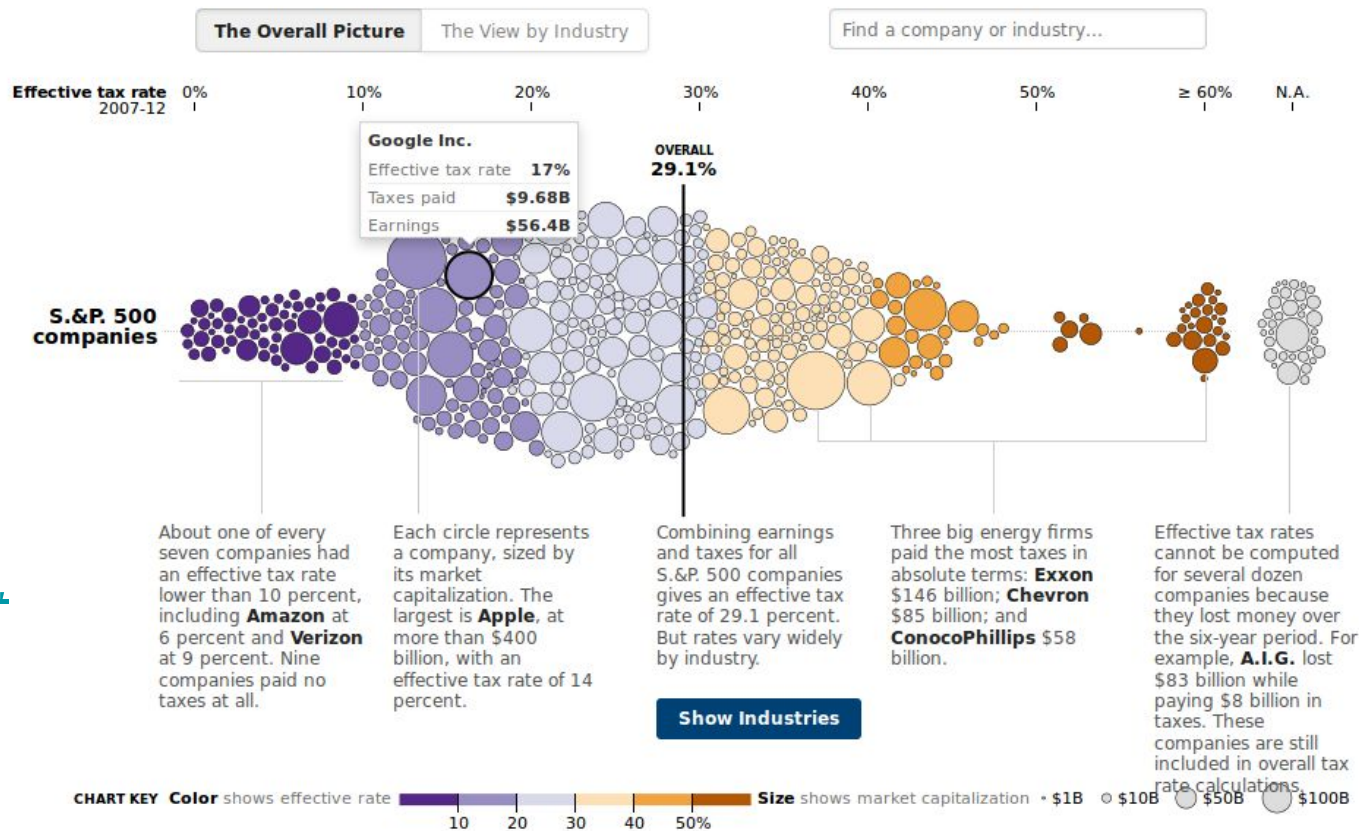
Haeyoun Park, Kevin Quealy

NY Times

# Data faceting

## Across U.S. Companies, Tax Rates Vary Greatly

Last week, in a Congressional hearing, Apple got grilled for its low-tax strategy. But not every business can copy that approach. Here is a look at what S.&P. 500 companies paid in corporate income taxes — federal, state, local and foreign — from 2007 to 2012, according to S&P Capital IQ. [Related Article »](#)

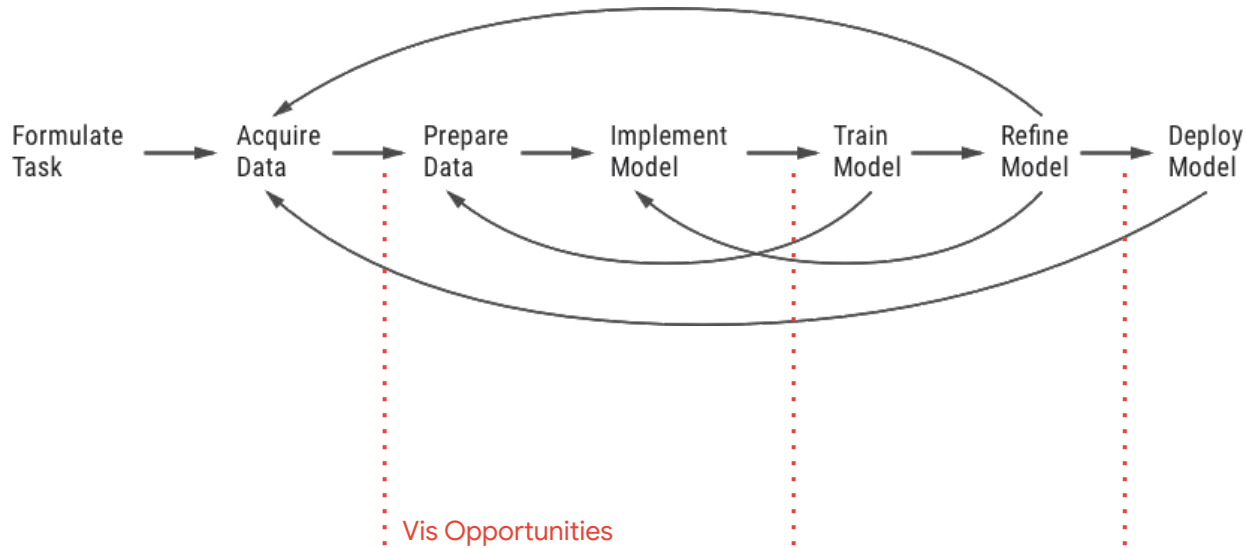


## Across U.S. Companies, Tax Rates Vary Greatly

M. Bostock, M. Ericson, D. Leonhardt, B. Marsh  
NY Times

Back to machine learning!

# Opportunities for Vis





# Framework: visualization uses in ML

1. Training Data
2. Model Performance
3. Interpretability + model inspection
4. High-dimensional data
5. Education and communication

# 1. Visualizing training data

# Visualizing CIFAR-10

CIFAR-10 and CIFAR-100

https://www.cs.toronto.edu/~kriz/cifar.html

< Back to Alex Krizhevsky's home page

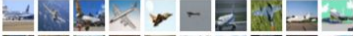









The CIFAR-10 and CIFAR-100 are labeled subsets of the [80 million tiny images](#) dataset. They were collected by Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton.

## The CIFAR-10 dataset

The CIFAR-10 dataset consists of 60000 32x32 colour images in 10 classes, with 6000 images per class. There are 50000 training images and 10000 test images.

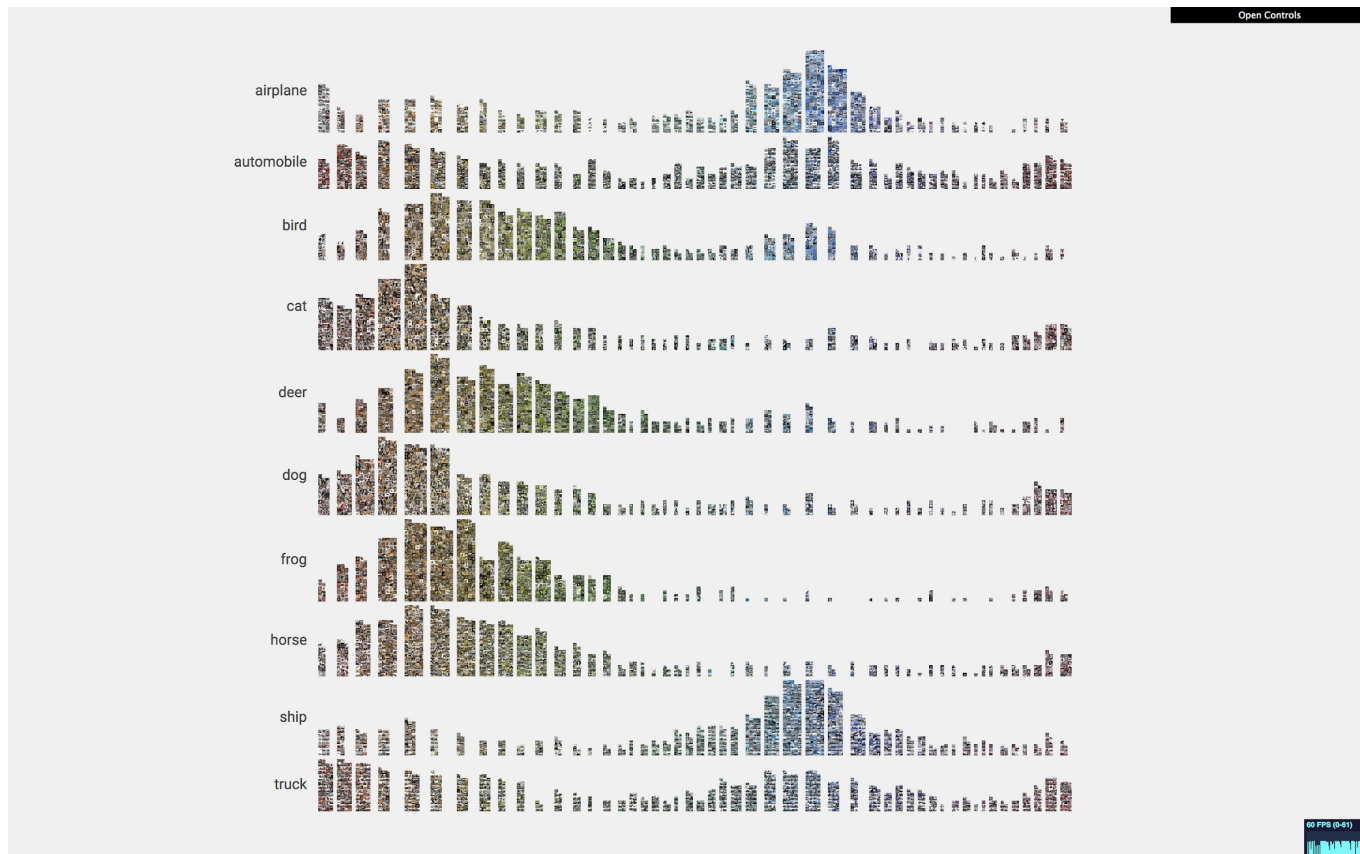
The dataset is divided into five training batches and one test batch, each with 10000 images. The test batch contains exactly 1000 randomly-selected images from each class. The training batches contain the remaining images in random order, but some training batches may contain more images from one class than another. Between them, the training batches contain exactly 5000 images from each class.

Here are the classes in the dataset, as well as 10 random images from each:

<b>airplane</b>	
<b>automobile</b>	
<b>bird</b>	
<b>cat</b>	
<b>deer</b>	
<b>dog</b>	
<b>frog</b>	
<b>horse</b>	
<b>ship</b>	
<b>truck</b>	

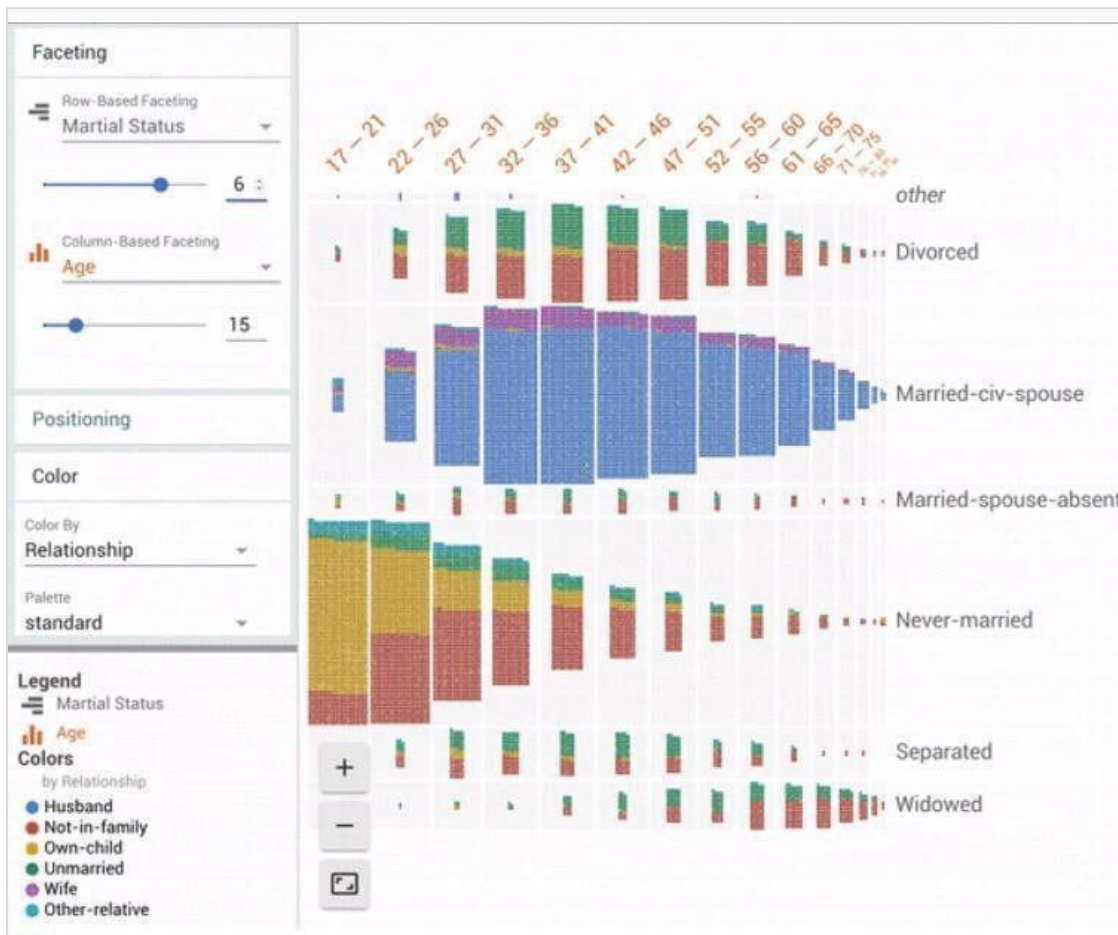
The classes are completely mutually exclusive. There is no overlap between automobiles and trucks. "Automobile" includes sedans, SUVs, things of that sort. "Truck" includes only big trucks. Neither includes pickup trucks.

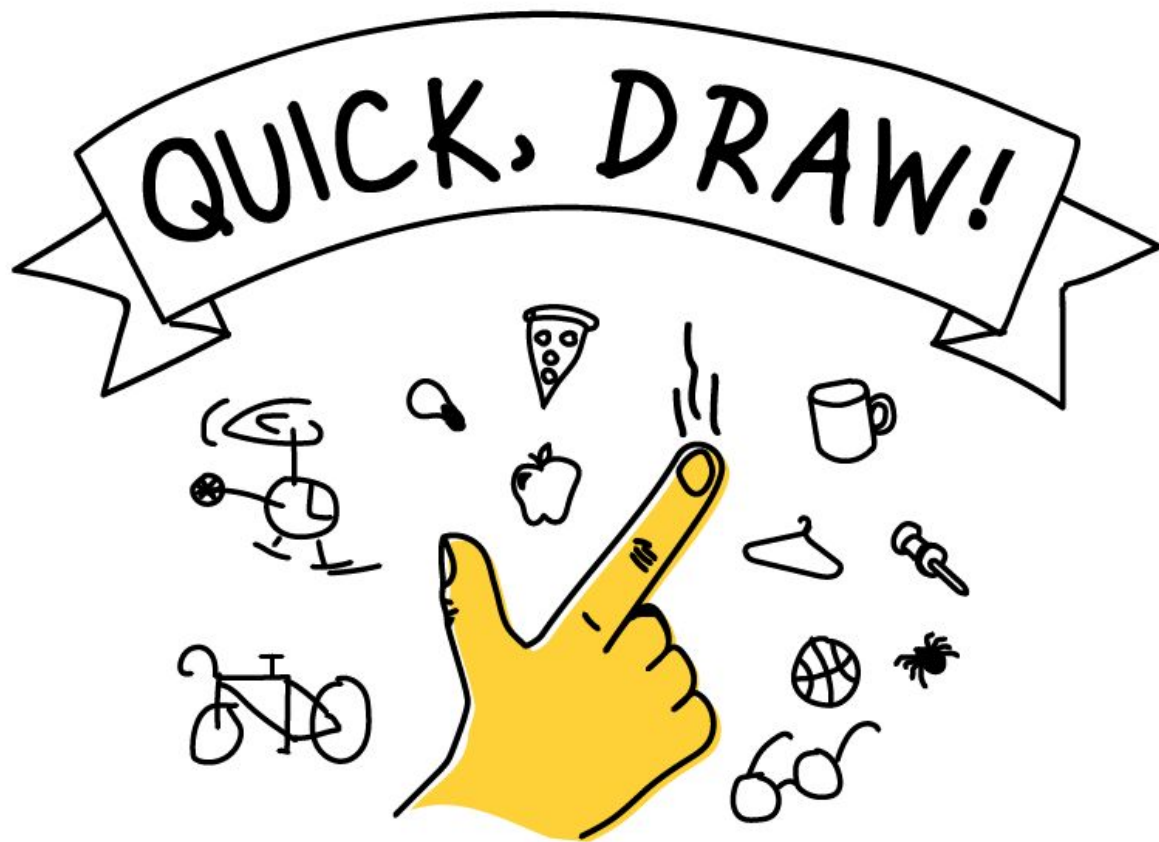
# CIFAR-10 Facets Demo



# Facets

Open-source visualization  
[pair-code.github.io/facets](https://pair-code.github.io/facets)





Google Creative Lab

<https://quickdraw.withgoogle.com/>

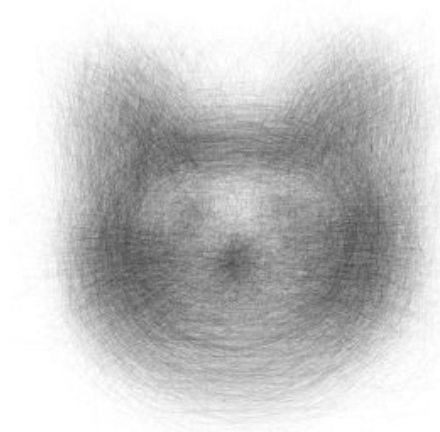


**Quick Draw, the data. .**

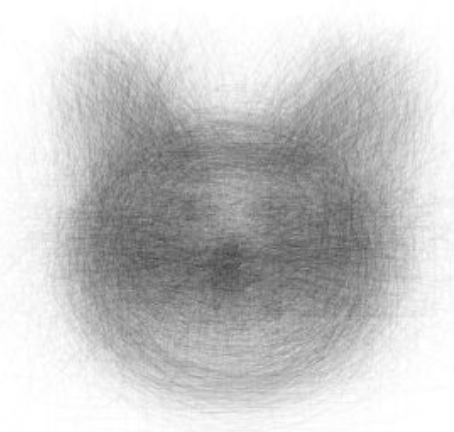
**[https://quickdraw.withgoogle.com/data. .](https://quickdraw.withgoogle.com/data)**



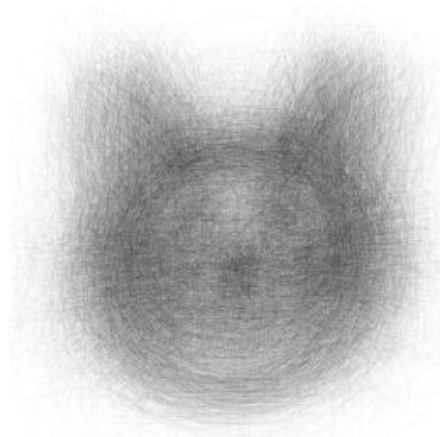
When things look alike  
across cultures



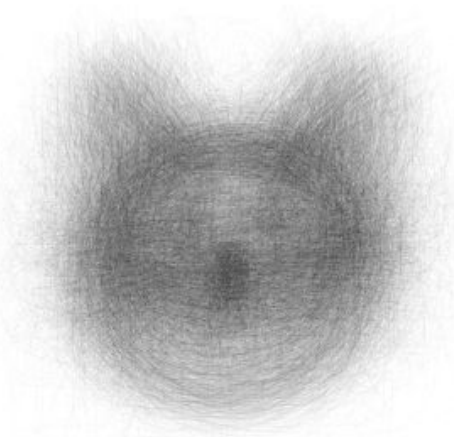
Korea



Germany



South Africa



United States

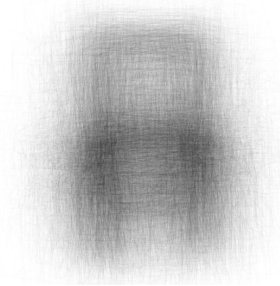
[Machine Learning for Visualization](#)

Let's Explore the Cutest Big Dataset

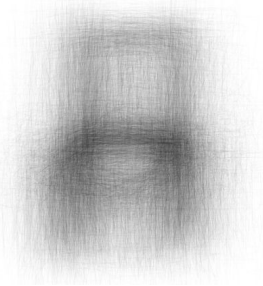
Ian Johnson



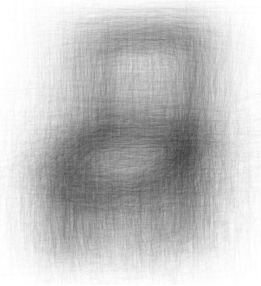
## And when they don't



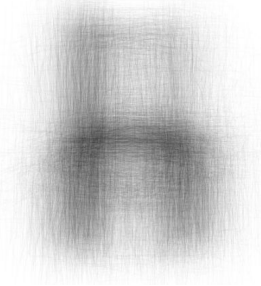
South Africa



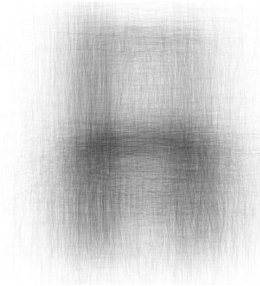
Russia



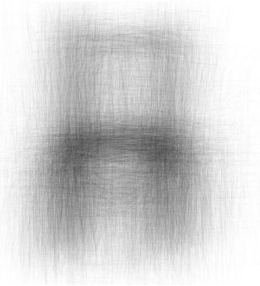
Korea



Brazil



United States

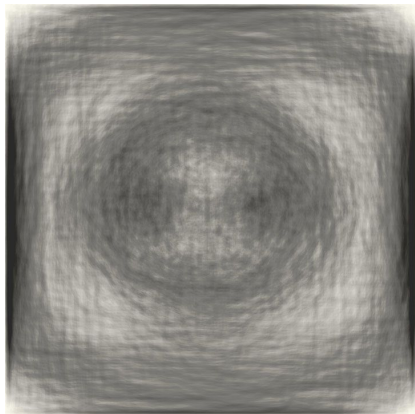


Germany

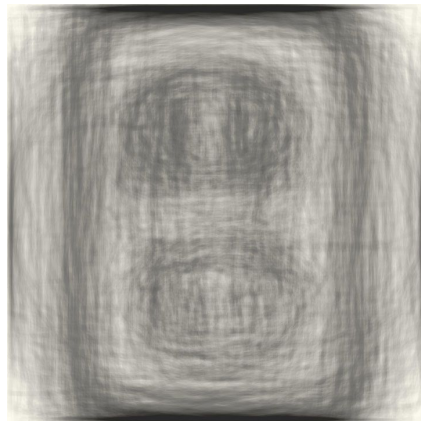
[Visual Averages by Country](#)

Kyle McDonald

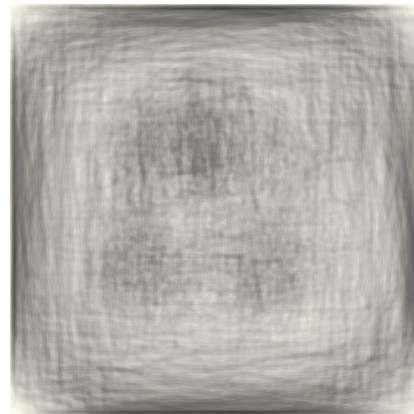
# Outlets



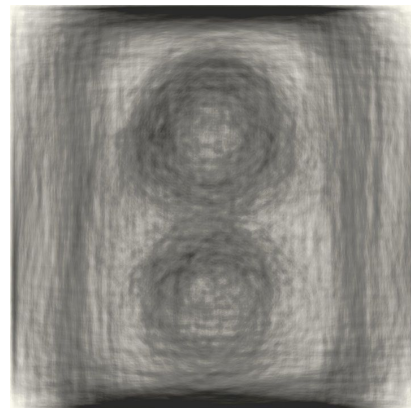
Germany



Japan



Malaysia

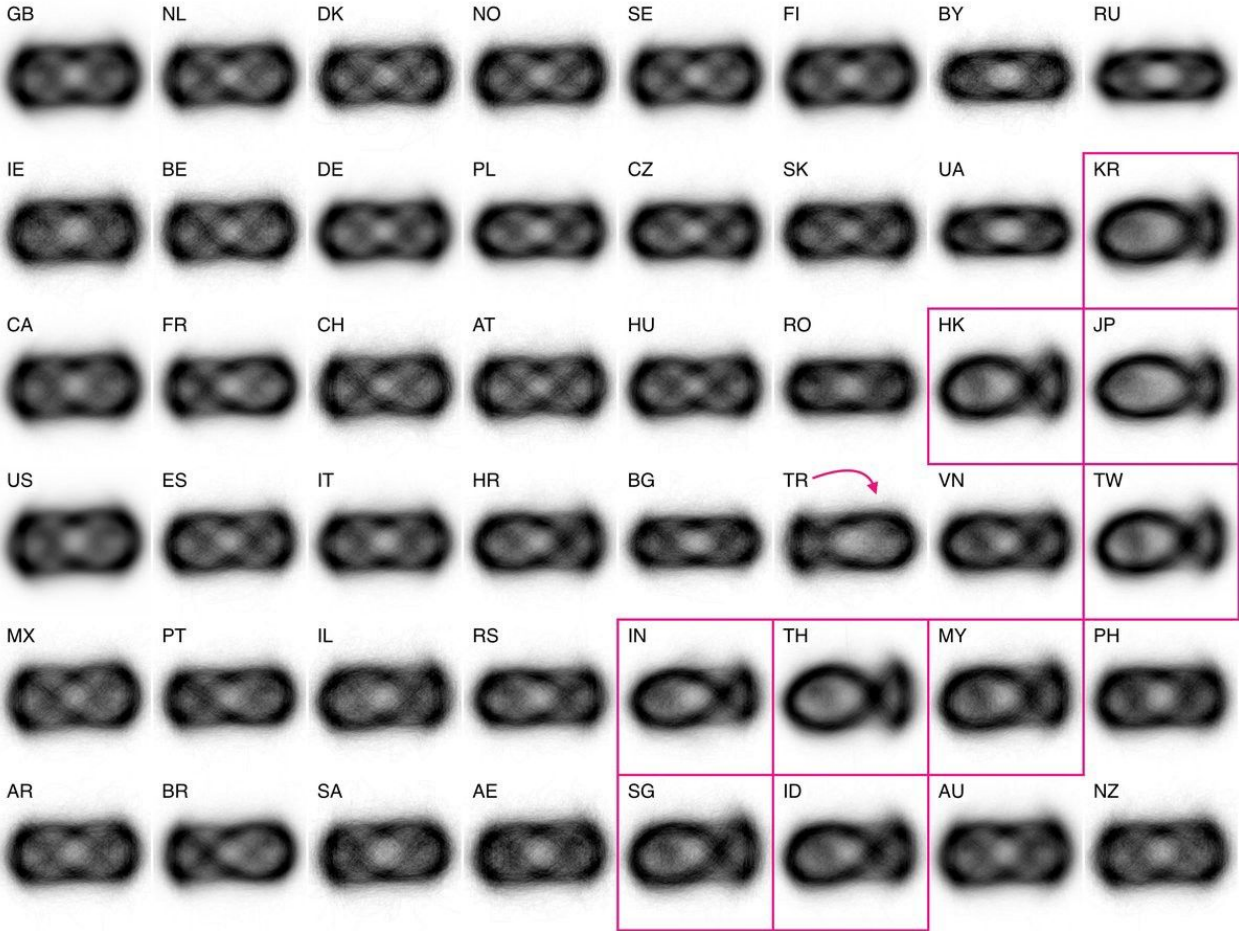


Sweden

[Visual Averages by Country](#)

Kyle McDonald

# Finding nemo: small multiples

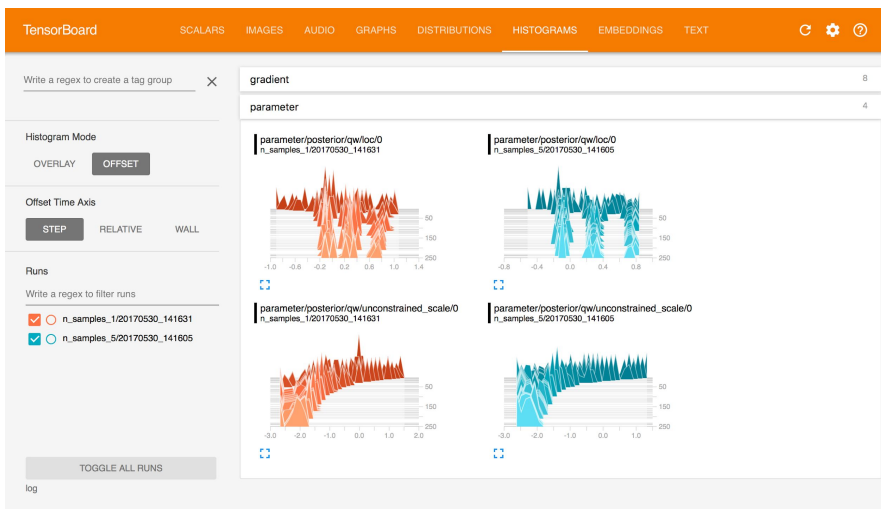


[Visual Averages by Country](#)

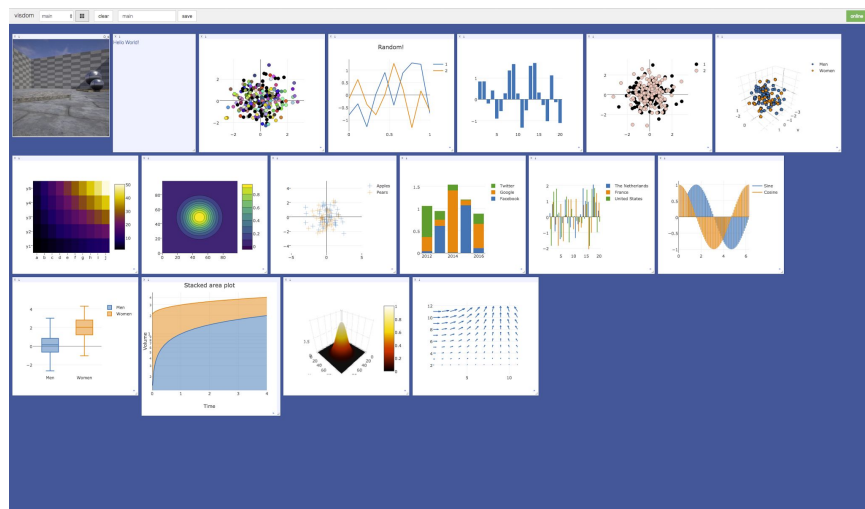
Kyle McDonald

## 2. Performance monitoring (very briefly!)

# Monitoring dashboards - apply standard visualization tools!



TensorBoard



Visdom

Two examples among many...

# 3. Interpretability + model inspection

# Convolutional NNs

# Image classification: interpretability petri dish

Image classifiers are effective in practice

Exactly what they're doing is somewhat mysterious

- And their failures (e.g. adversarial examples) add to mystery

**But:** Way easier to inspect what's going on in artificial classifiers than in human classifiers ;-)

Since these are visual systems, it's natural to use visualization to inspect them

- What features are these networks really using?
- Do individual units have meaning?
- What roles are played by different layers?
- How are high-level concepts built from low-level ones?



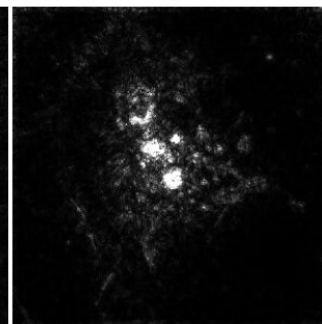
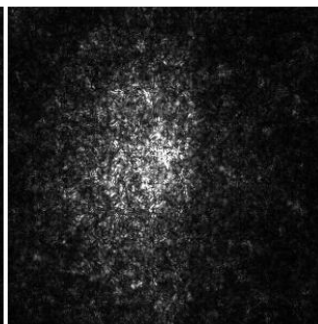
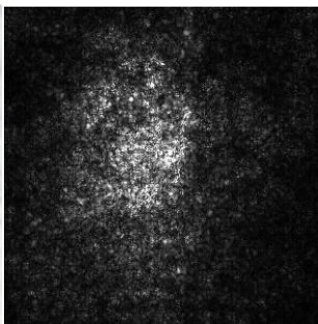
# Saliency maps - examples

Image

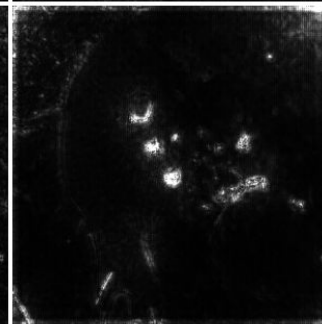
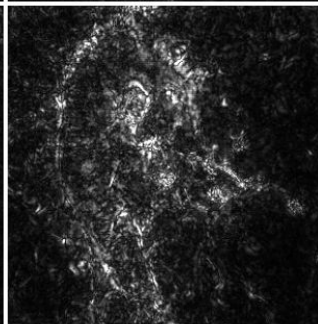
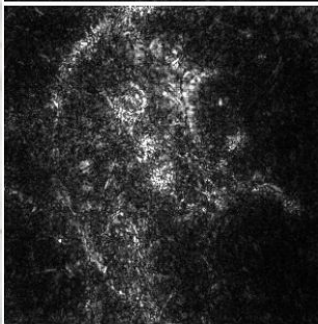
Gradient

Integrated

Guided Backprop



Gradient



Gradient  $\times$  Image

# Saliency maps

(a.k.a. "Sensitivity maps")

Idea: consider sensitivity of class to each pixel

i.e.  $\text{grad}(f)$ , where  $f$  is function from pixels to class score.

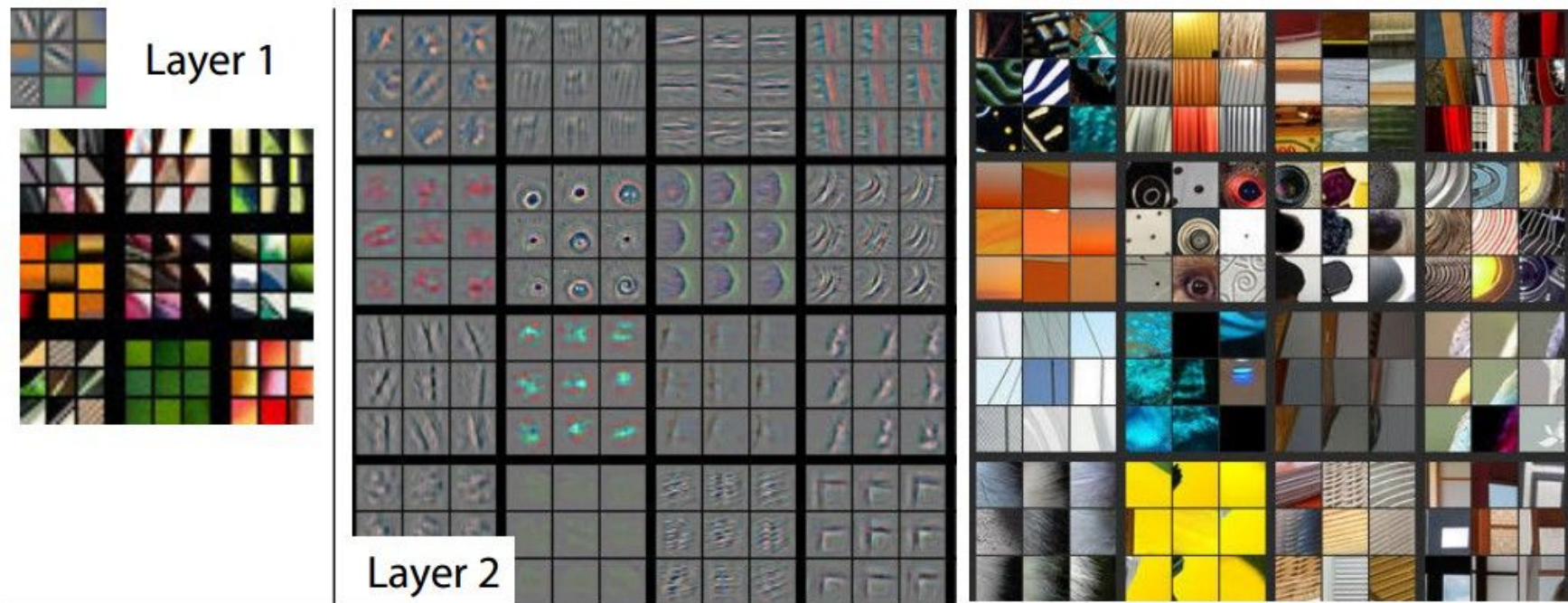
**Many** ways to extend basic idea!

- Layer-wise relevance propagation (Binder et al.)
- Integrated gradients (Sundararajan et al.)
- Guided backprop (Springenberg et al.)
- etc.

Yet interpretation is slippery (Adebayo et al., Kindermans et al.)

- Tend to be visually noisy. Are these sometimes Rorschach tests?
- Are some of these methods essentially edge detectors?

# Visualizing arbitrary neurons along the way to the top...



Gray: trying to maximize neural response. Colorful squares: maximal examples from an image data set

Visualizing and Understanding Convolutional Networks

Zeiler & Fergus, 2013

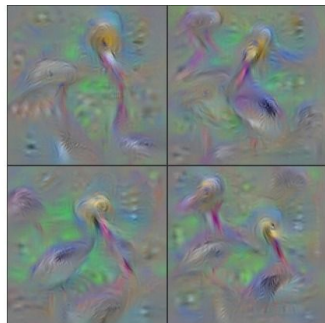
# Understanding Neural Networks Through Deep Visualization

Yosinski et al. , 2015

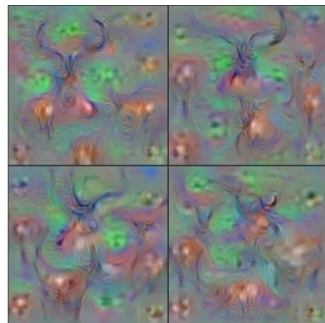
<http://yosinski.com/deepvis>



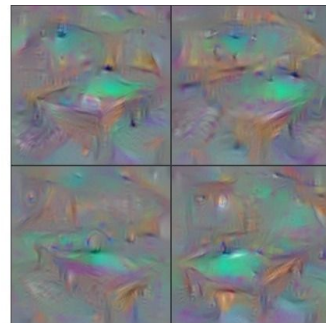
Flamingo



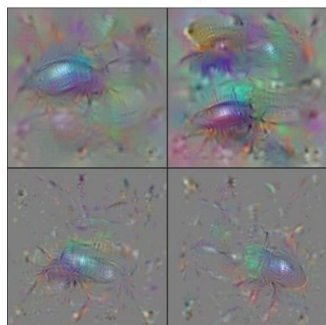
Pelican



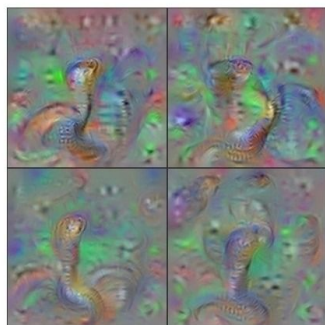
Hartebeest



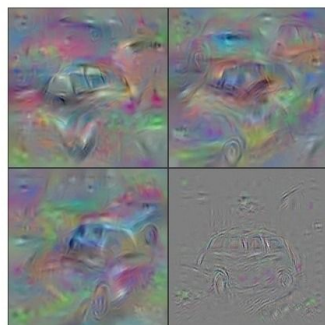
Billiard Table



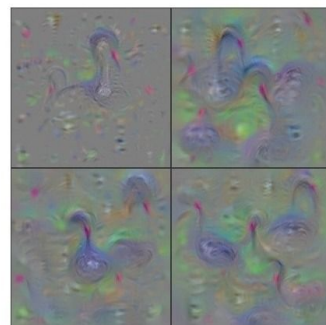
Ground Beetle



Indian Cobra



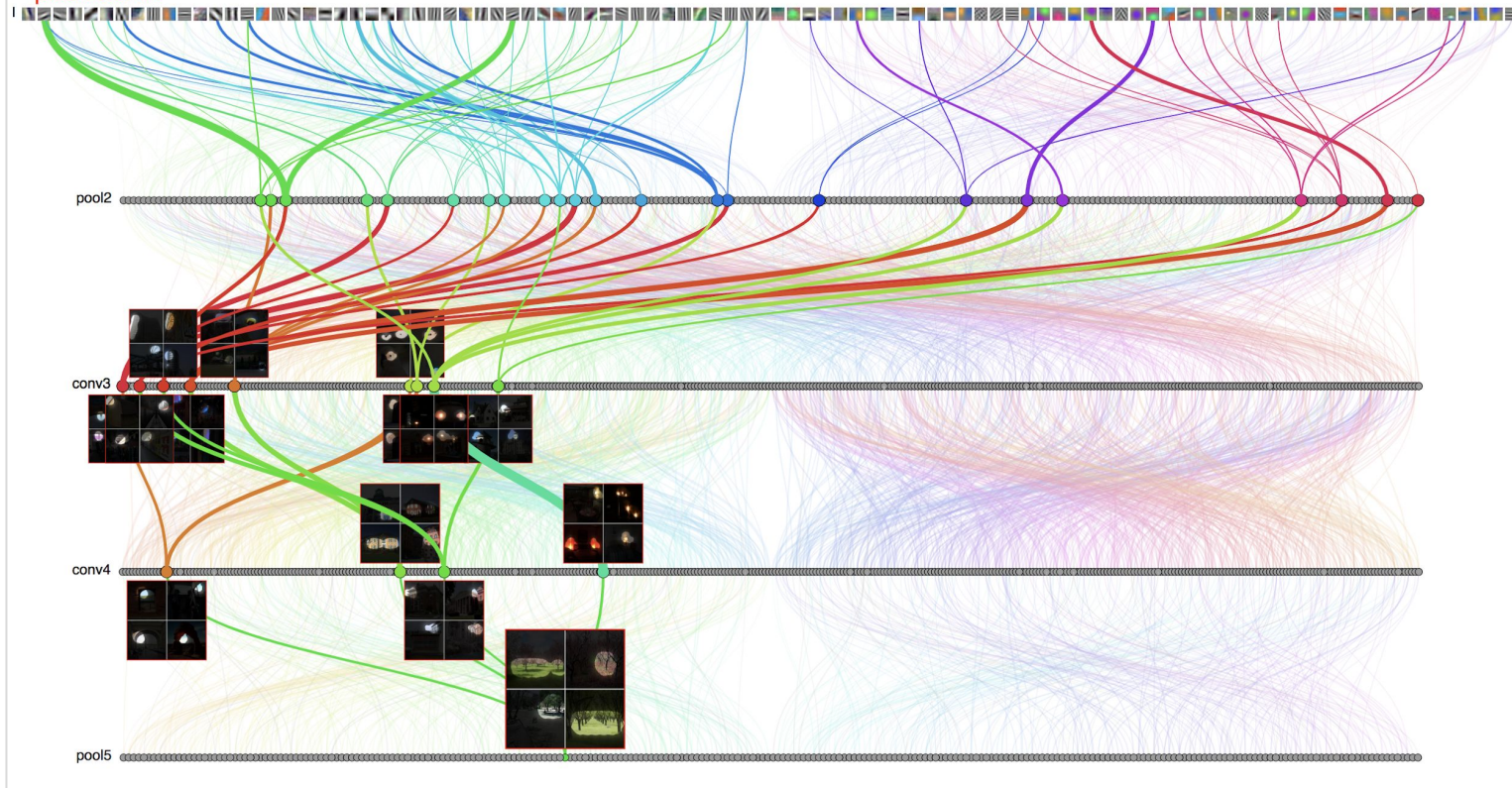
Station Wagon



Black Swan

# drawNet

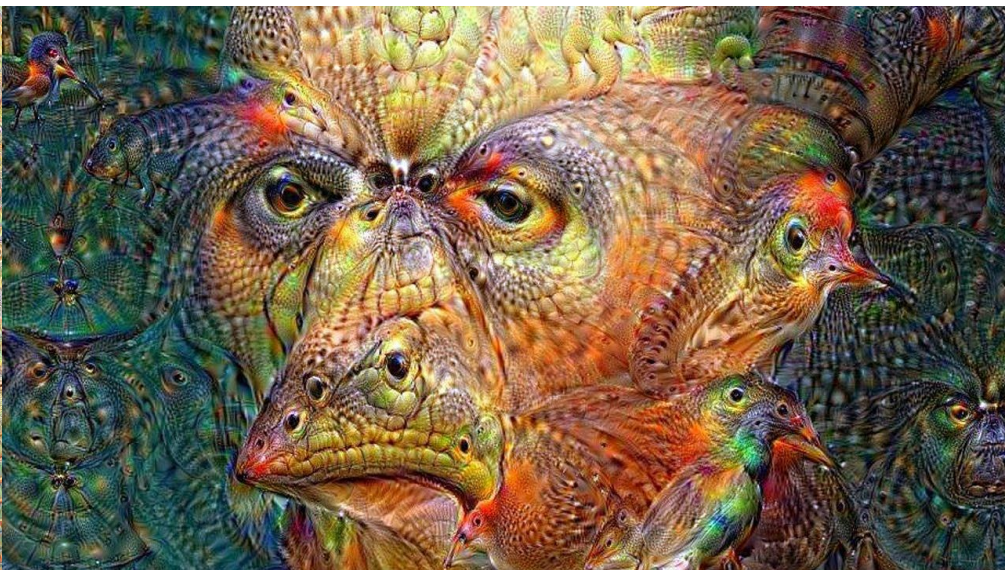
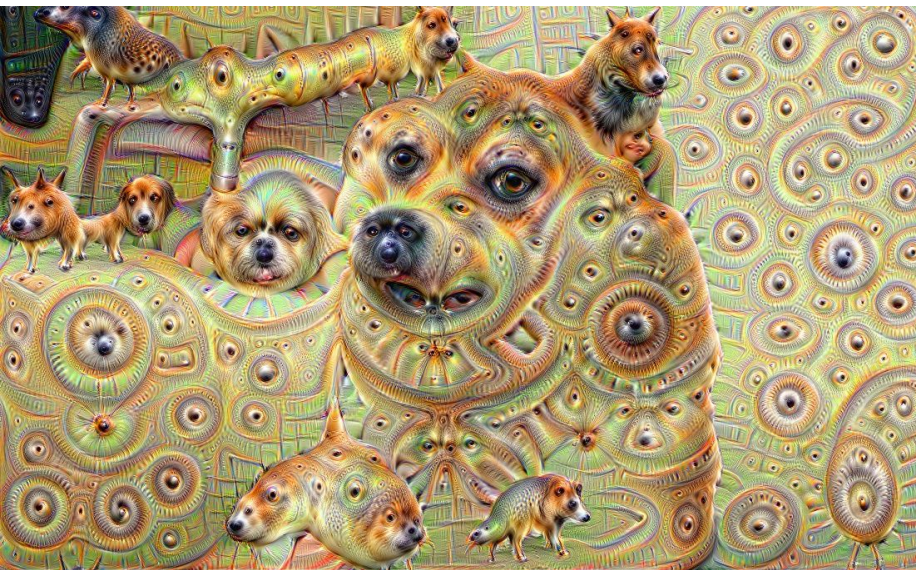
placesCNN



drawNet

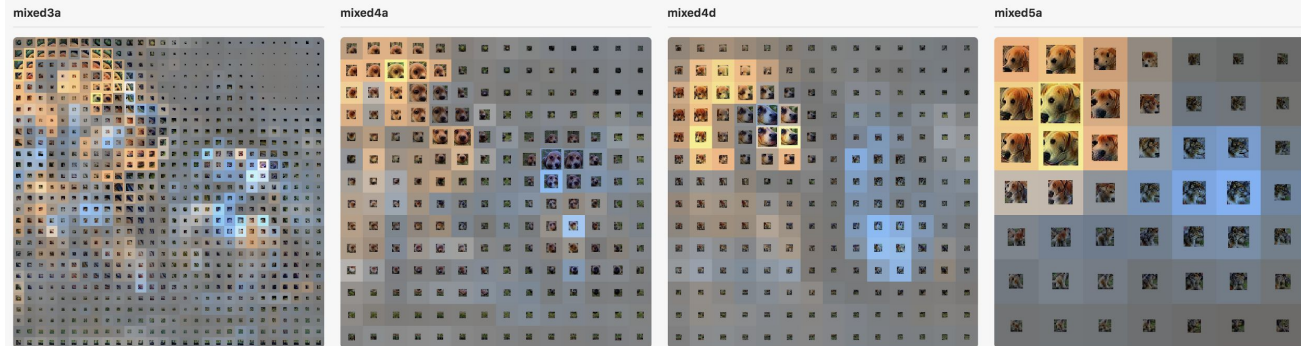
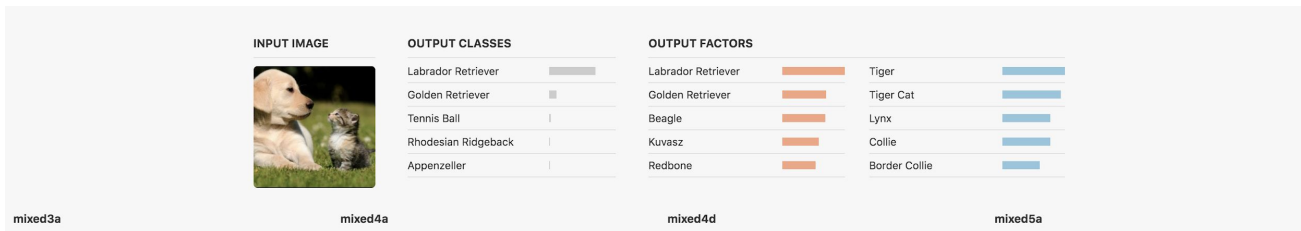
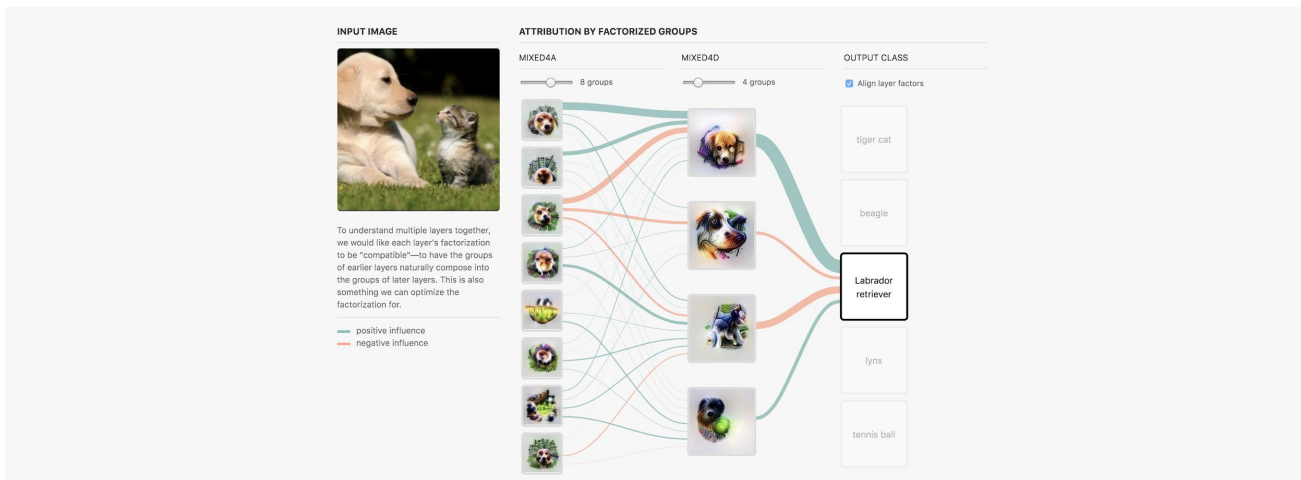
Torralba

# Deep Dream



deepdream  
Mordvintsev, Tyka, Olah

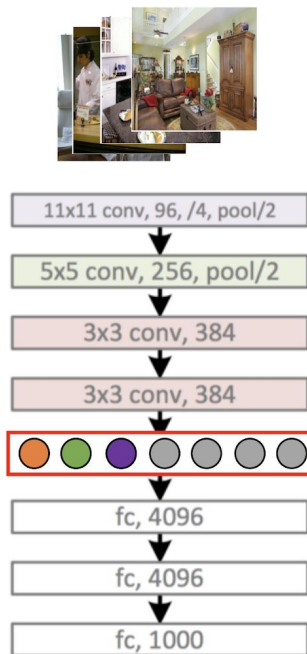
# Combining these interpretability ideas to create new visualizations



## The Building Blocks of Interpretability

Olah, Satyanarayan, Johnson, Carter, Schubert, Ye, Mordvintsev

# Going From Visualization to Interpretation



Unit 1



Top Activated Images

Interpretation: head

Score: 0.23



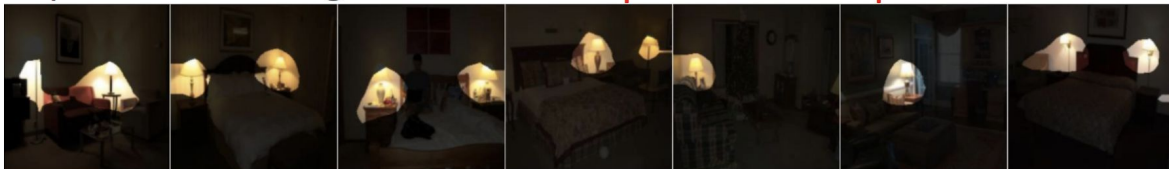
Unit 2



Top Activated Images

Interpretation: lamp

Score: 0.15



Unit 3



Top Activated Images

Interpretation: car

Score: 0.02





RNNs

# Visualizing text sequences, colored by activations of a cell

## Cell sensitive to position in line:

The sole importance of the crossing of the Berezina lies in the fact that it plainly and indubitably proved the fallacy of all the plans for cutting off the enemy's retreat and the soundness of the only possible line of action--the one Kutuzov and the general mass of the army demanded--namely, simply to follow the enemy up. The French crowd fled at a continually increasing speed and all its energy was directed to reaching its goal. It fled like a wounded animal and it was impossible to block its path. This was shown not so much by the arrangements it made for crossing as by what took place at the bridges. When the bridges broke down, unarmed soldiers, people from Moscow and women with children who were with the French transport, all--carried on by vis inertiae--pressed forward into boats and into the ice-covered water and did not, surrender.

## Cell that turns on inside quotes:

"You mean to imply that I have nothing to eat out of.... On the contrary, I can supply you with everything even if you want to give dinner parties," warmly replied Chichagov, who tried by every word he spoke to prove his own rectitude and therefore imagined Kutuzov to be animated by the same desire.

Kutuzov, shrugging his shoulders, replied with his subtle penetrating smile: "I meant merely to say what I said."

Cell that robustly activates inside if statements:

```
static int __dequeue_signal(struct sigpending *pending, sigset_t *mask,
                           siginfo_t *info)
{
    int sig = next_signal(pending, mask);
    if (sig) {
        if (current->notifier) {
            if (sigismember(current->notifier_mask, sig)) {
                if (!(current->notifier)(current->notifier_data)) {
                    clear_thread_flag(TIF_SIGPENDING);
                    return 0;
                }
            }
        }
        collect_signal(sig, pending, info);
    }
    return sig;
}
```

# Seq2Seq Vis

die längsten reisen fangen an , wenn es auf den straßen dunkel wird .

Enc words: die längsten reisen fangen an , wenn es auf den straßen dunkel wird .

Attention:



← change:

word attn

→ compare:

sentence

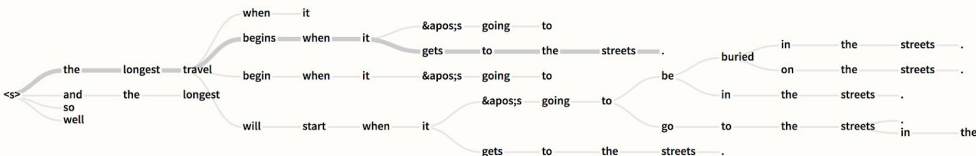
swap:

↔

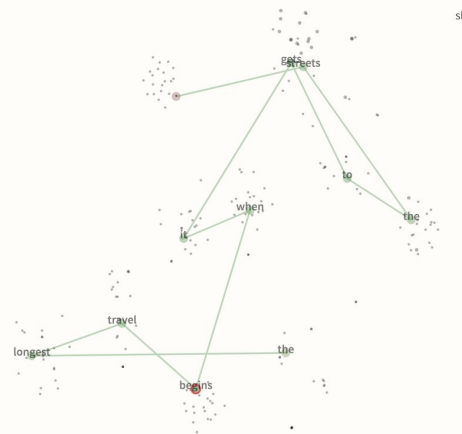
topK:

the	longest	travel	when	when	it	&apos;	to	the	streets	.
and	oldest	trips	will	if	they	gets	dark	a	roads	in
so	tallest	journeys	begins	,	the	becomes	buried	shore	road	of
well	russians	travels	begin	as	there	grows	into	heaven	street	,
you	icons	journey	start	in	this	comes	in	its	city	to

pivot



decoder ▼



show: src tgt highlight: -1 0 +1

- das sicherheitstheater wird aufgedeckt , wenn es offensichtlich ist , dass es nicht richtig funktioniert .
- <S> security theater &apos;s exposed when it &apos;s obvious that it &apos;s not working properly . </S>
- &quot; was würde passieren , wenn ich für jedes <unk> paar dieser schuhe , genau das gleiche paar jemandem gebe , der gar keine schuhe besitzt ? &quot;
- <S> &quot; what would happen if every time someone bought a pair of these shoes i gave exactly the same pair to someone who doesn't even own a pair of shoes ? &quot;
- letzten endes kann das nur erreicht werden , wenn unsere internationalen institutionen gut genug arbeiten und das schaffen können .
- <S> now , in the end , that can only be achieved if your international institutions work well enough to be able to do so . </S>
- &quot; verwenden tiere ihre schwänze , wenn sie mauern <unk> ? &quot;
- <S> &quot; do animals use their tails when they climb up walls ? &quot; </S>
- genial ist , was passiert , wenn ich diesen klavierschaltkreis verschiedenen leuten gebe .
- <S> what &apos;s awesome is what happens when you give the piano circuit to people . </S>
- das passiert wenn jesse daran denkt seine hand zu <unk> und zu schließen , oder ellbogen beugen oder strecken .
- <S> this is what happens when jesse thinks open and close his hand , or bend or <unk> your elbow . </S>
- was würde passieren , wenn sie an die unterseite des blattes kletterten und es käme eine <unk> oder wir würden das blatt schütteln ? &quot;
- <S> and what would happen if they climbed on the underside of that leaf , and there was some wind , or we shook it ?
- das untere bild zeigt , was passiert , wenn der körper sich nicht mehr bewegt und das tier beginnt , eine <unk> gerät oder einen <unk> zu kontrollieren .

## Seq2Seq-Vis: Visual Debugging Tool for Sequence- to-Sequence Models

Strobel et al., 2018

Examine model  
decisions

Connect decisions to  
previous examples

Test alternative  
decisions



Linking multiple views...

DQNViz: A Visual Analytics Approach to Understand Deep Q-Networks

Wang et al  
VAST 2018.

Fig. 1. DQNViz: (a) the *Statistics* view presents the overall training statistics with line charts and stacked area charts; (b) the *Epoch* view shows epoch-level statistics with pie charts and stacked bar charts; (c) the *Trajectory* view reveals the movement and reward patterns of the DQN agent in different episodes; (d) the *Segment* view reveals what the agent really sees from a selected segment.

## 4. High-dimensional data

# Why high-dimensional data?

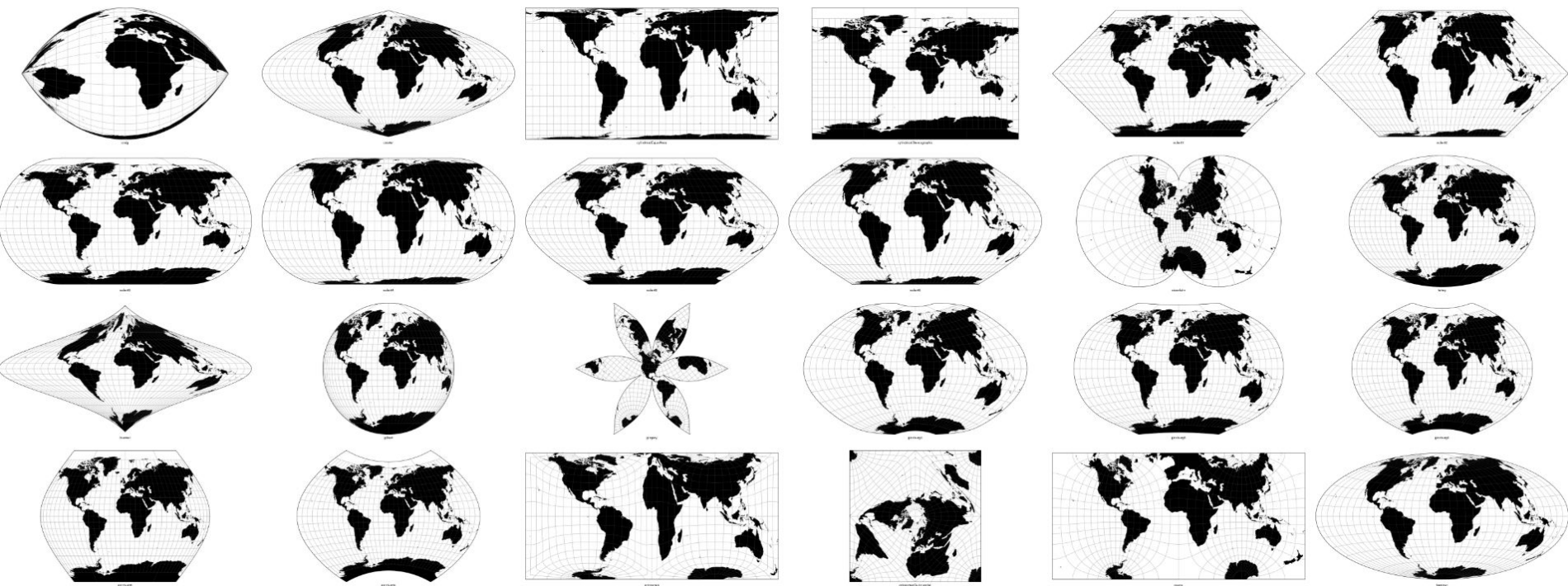
Vectors spaces are the lingua franca of much of ML these days

- Data such as images, audio, video is naturally high-dimensional
- Dense representations of discrete data (e.g. word embeddings) have had major successes

Why is it hard? Because it's impossible



# Why is it hard? Because it's impossible



See [Every Map Projection](#), Bostock.

# Main approaches

## Linear

- Principal Component Analysis
- [Visualization of Labeled Data Using Linear Transformations](#) (Koren & Carmel)

## Non-linear (just a few of many)

- Multidimensional scaling
- Sammon mapping
- Isomap
- **t-SNE**
- **UMAP**

# Main approaches

## Linear

- Principal Component Analysis (show as much variation in data as possible)
- Visualization of Labeled Data Using Linear Transformations (clusters match labels)

## Non-linear (just a few of many)

- Multidimensional scaling
- Sammon mapping
- Isomap
- **t-SNE**
- **UMAP**



Minimize distortion, according to some metric

# t-SNE

## Visualizing Data using t-SNE

**Laurens van der Maaten**

*TiCC*

*Tilburg University*

*P.O. Box 90153, 5000 LE Tilburg, The Netherlands*

LVDMAATEN@GMAIL.COM

**Geoffrey Hinton**

*Department of Computer Science*

*University of Toronto*

*6 King's College Road, M5S 3G4 Toronto, ON, Canada*

HINTON@CS.TORONTO.EDU

# t-SNE

Fairly complex non-linear technique

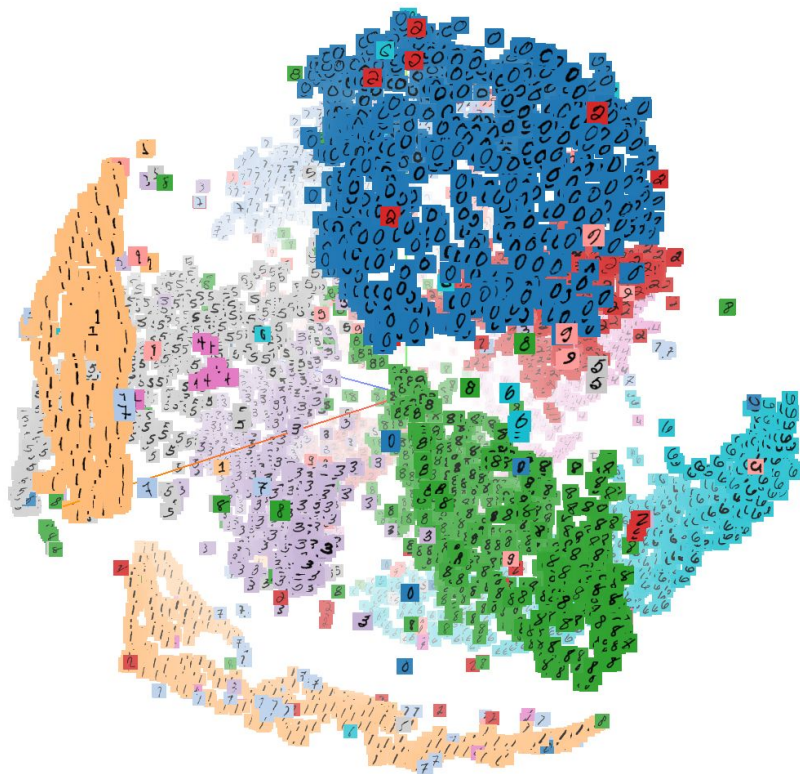
Uses an adaptive sense of "distance." Translates well between geometry of high- and low-dimensional space

Has become a standard tool, so we'll spend some time discussing how to read it.

# Demo: MNIST visualization

## Embedding Projector

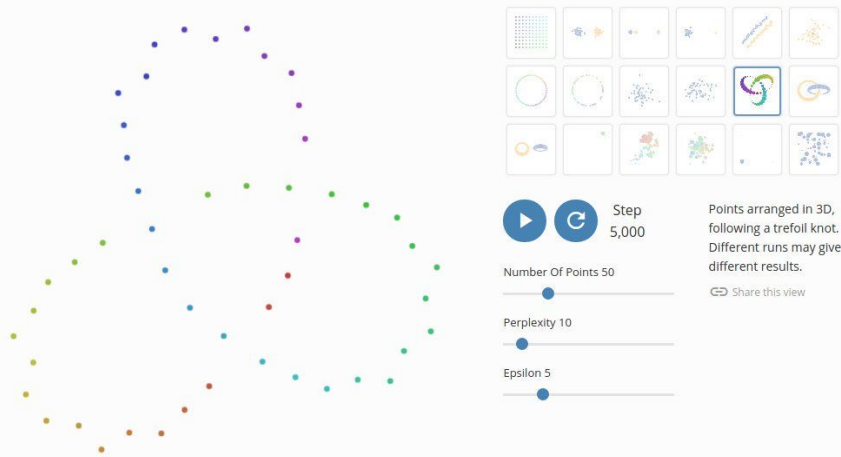
Open Source visualization tool  
Also available on Tensorboard  
[projector.tensorflow.org/](https://projector.tensorflow.org/)



# "Close reading" a visualization technique

## How to Use t-SNE Effectively

Although extremely useful for visualizing high-dimensional data, t-SNE plots can sometimes be mysterious or misleading. By exploring how it behaves in simple cases, we can learn to use it more effectively.



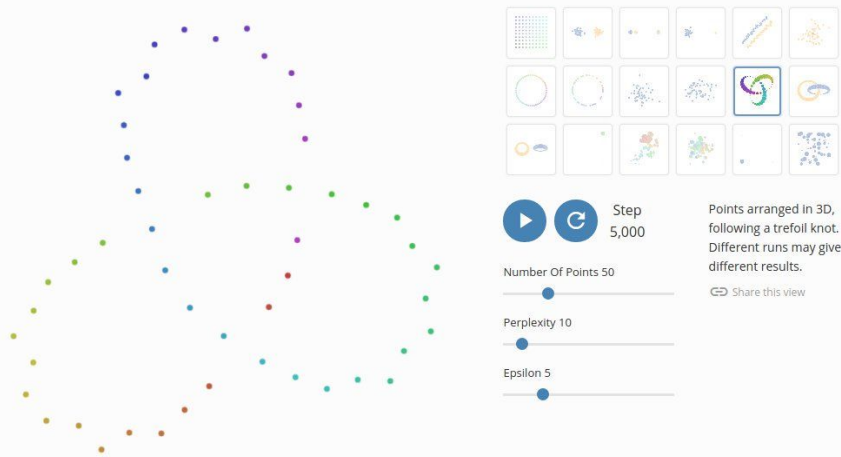
What's the right way to understand a "magic" visualization technique?

See [Distill article](#)

# "Close reading" a visualization technique

## How to Use t-SNE Effectively

Although extremely useful for visualizing high-dimensional data, t-SNE plots can sometimes be mysterious or misleading. By exploring how it behaves in simple cases, we can learn to use it more effectively.

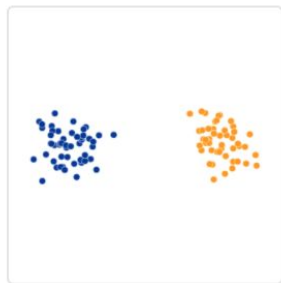


What's the right way to understand a "magic" visualization technique?

**More visualization, of course!**

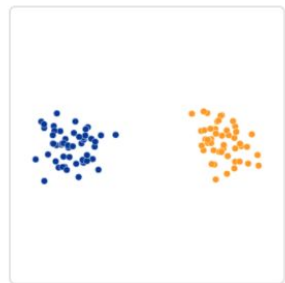


# Those hyperparameters really matter

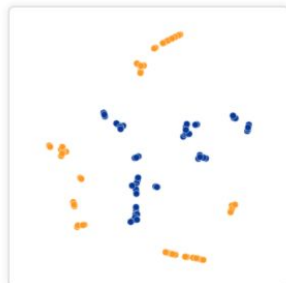


*Original*

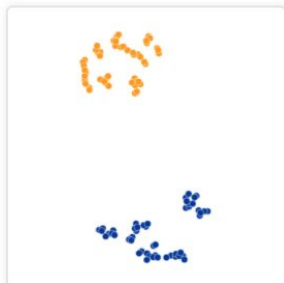
# Those hyperparameters really matter



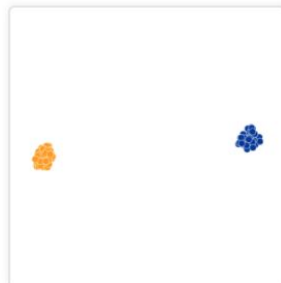
*Original*



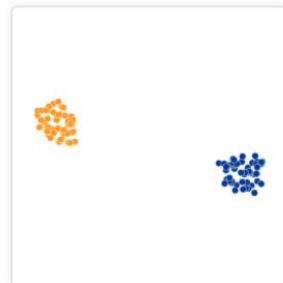
Perplexity: 2  
Step: 5,000



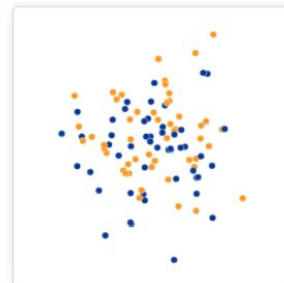
Perplexity: 5  
Step: 5,000



Perplexity: 30  
Step: 5,000



Perplexity: 50  
Step: 5,000



Perplexity: 100  
Step: 5,000

# Cluster sizes in a t-SNE plot mean nothing

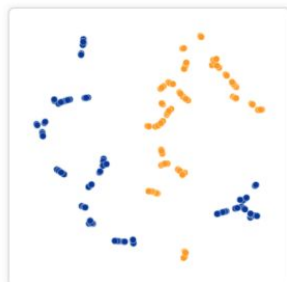


*Original*

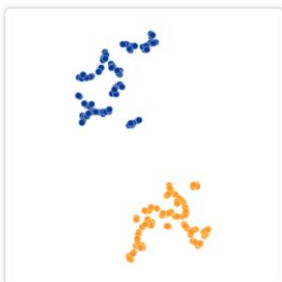
# Cluster sizes in a t-SNE plot mean nothing



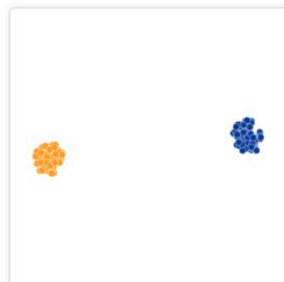
*Original*



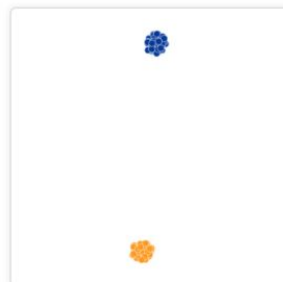
Perplexity: 2  
Step: 5,000



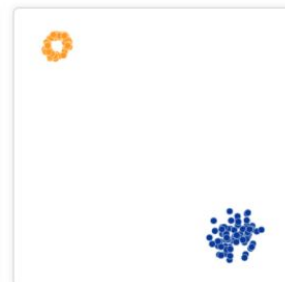
Perplexity: 5  
Step: 5,000



Perplexity: 30  
Step: 5,000

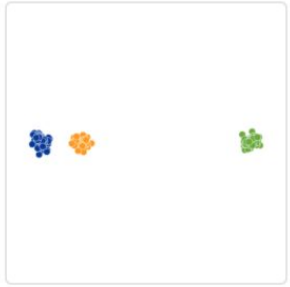


Perplexity: 50  
Step: 5,000



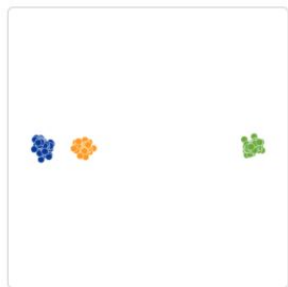
Perplexity: 100  
Step: 5,000

# Distances between clusters may not mean much

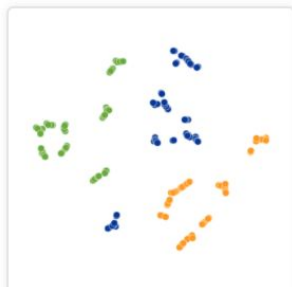


*Original*

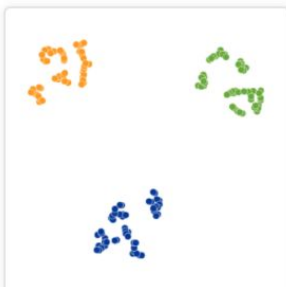
# Distances between clusters may not mean much



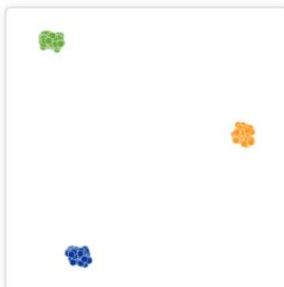
*Original*



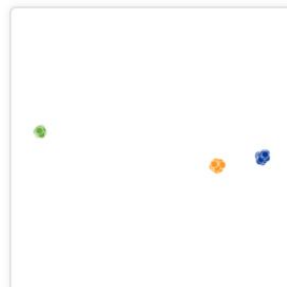
Perplexity: 2  
Step: 5,000



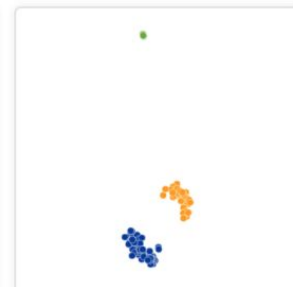
Perplexity: 5  
Step: 5,000



Perplexity: 30  
Step: 5,000

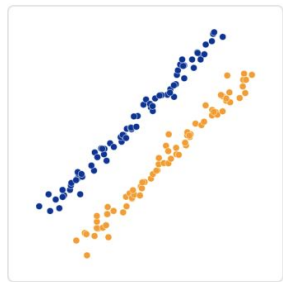
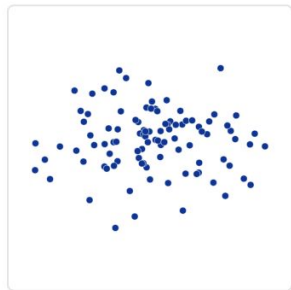


Perplexity: 50  
Step: 5,000



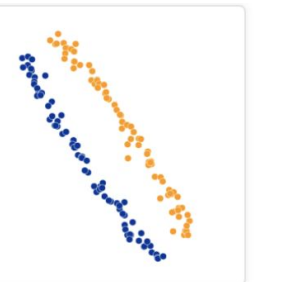
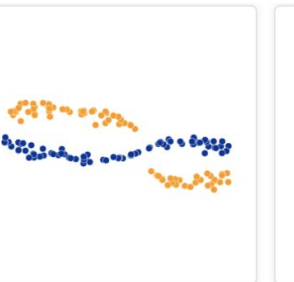
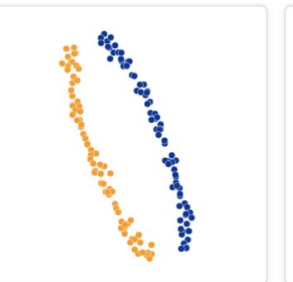
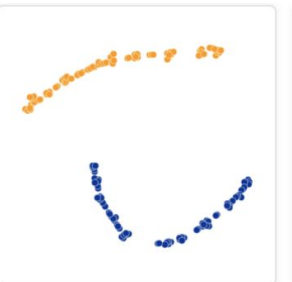
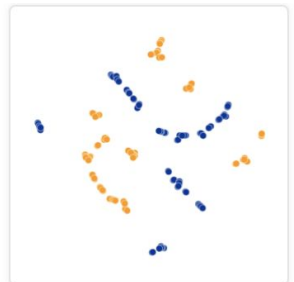
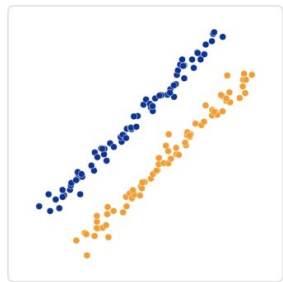
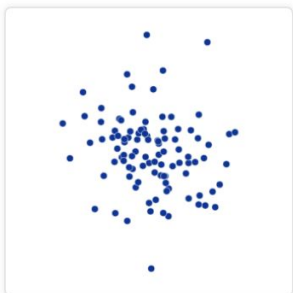
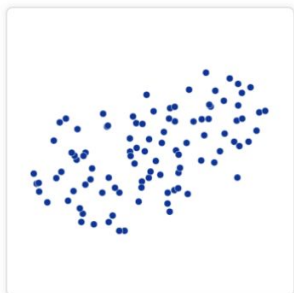
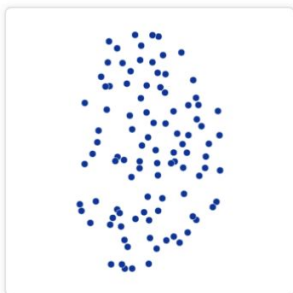
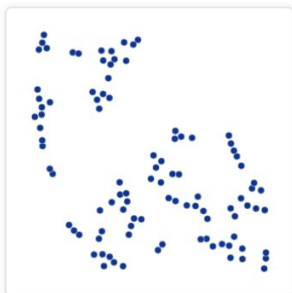
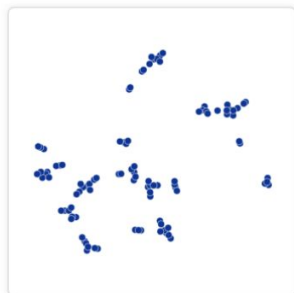
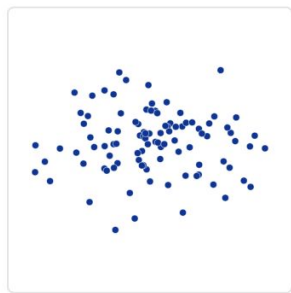
Perplexity: 100  
Step: 5,000

# You can see some shapes, sometimes



*Original*

# You can see some shapes, sometimes



*Original*

Perplexity: 2  
Step: 5,000

Perplexity: 5  
Step: 5,000

Perplexity: 30  
Step: 5,000

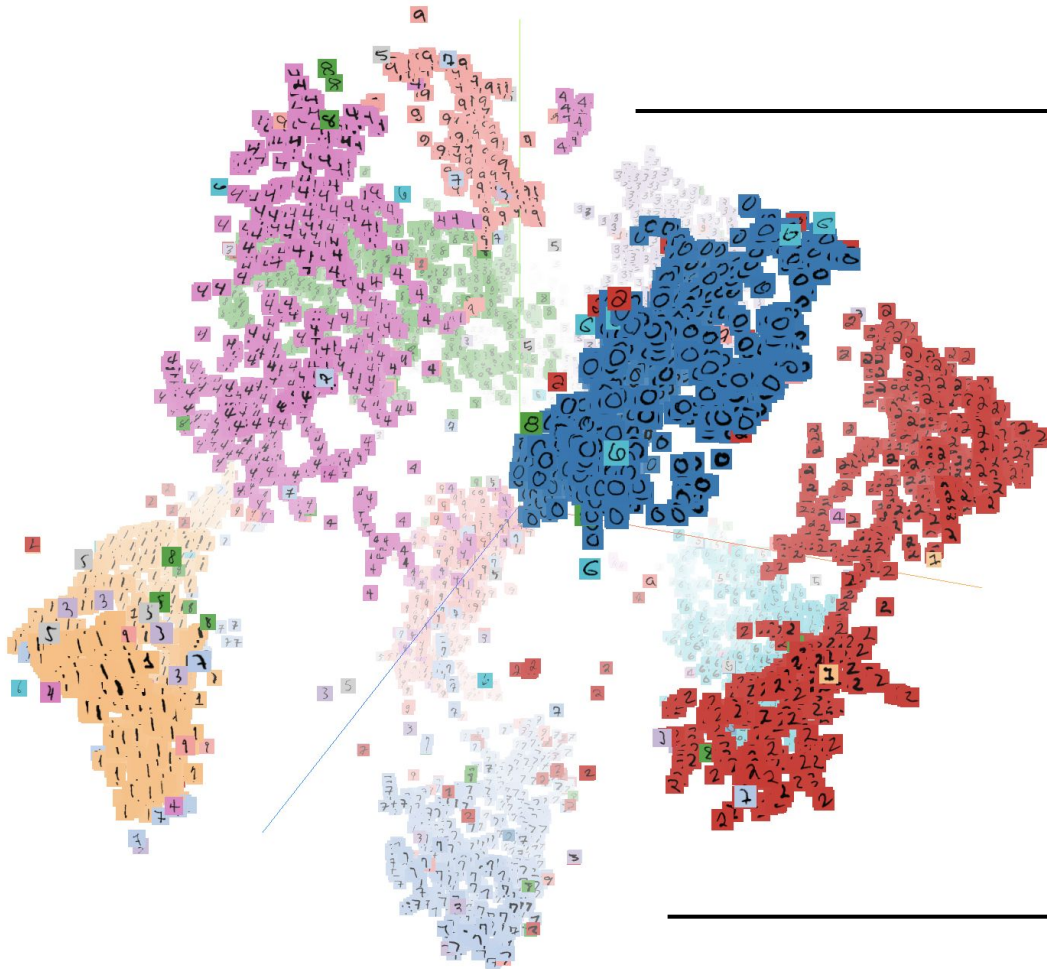
Perplexity: 50  
Step: 5,000

Perplexity: 100  
Step: 5,000



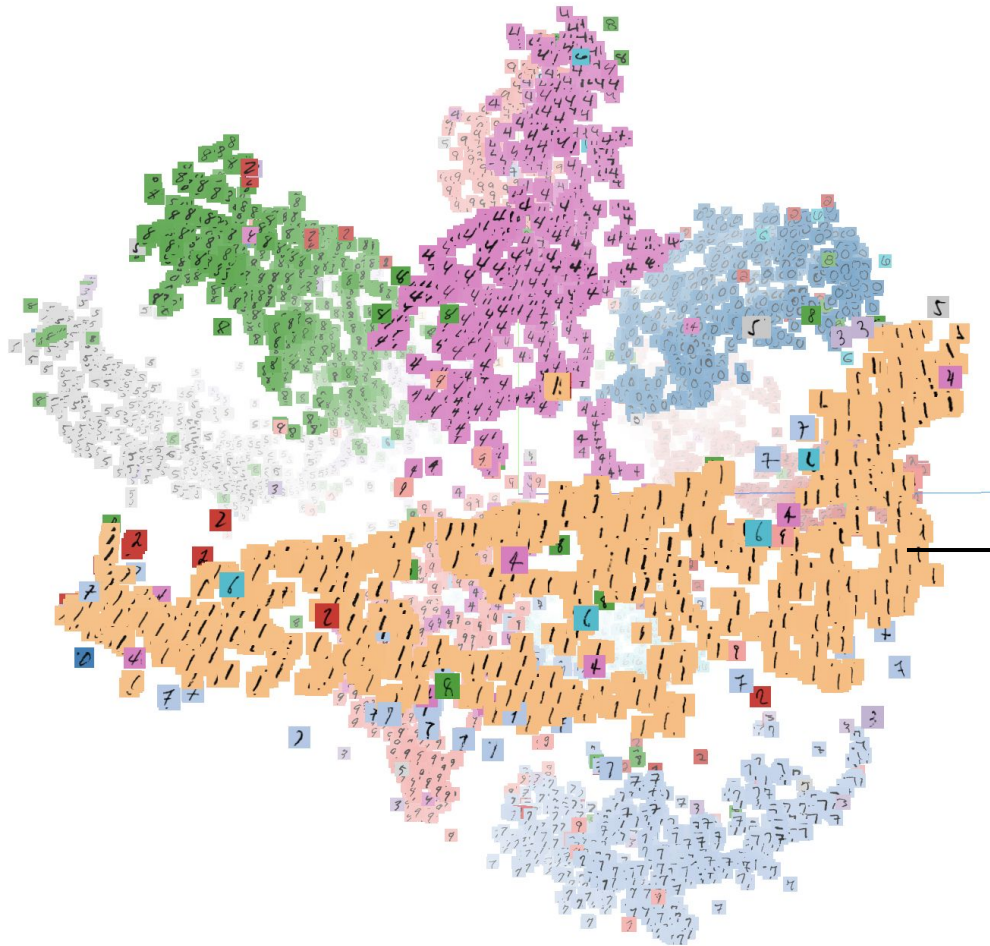
Let's try this out with MNIST





The 4's may not be separated into two clusters.

Clusters seem about equally far apart in 3D; may not actually be.



The clusters of 1's probably is long and thin.

# UMAP: New kid on the block

## **UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction**

Leland McInnes and John Healy

Tutte Institute for Mathematics and Computing

leland.mcinnes@gmail.com

jchealy@gmail.com

February 13, 2018

# UMAP: New kid on the block

## Practical value

- Faster than t-SNE
- Can efficiently embed into high dimensions (i.e. useful not just for visualization)
- Often seems to capture global structure better

# UMAP: New kid on the block

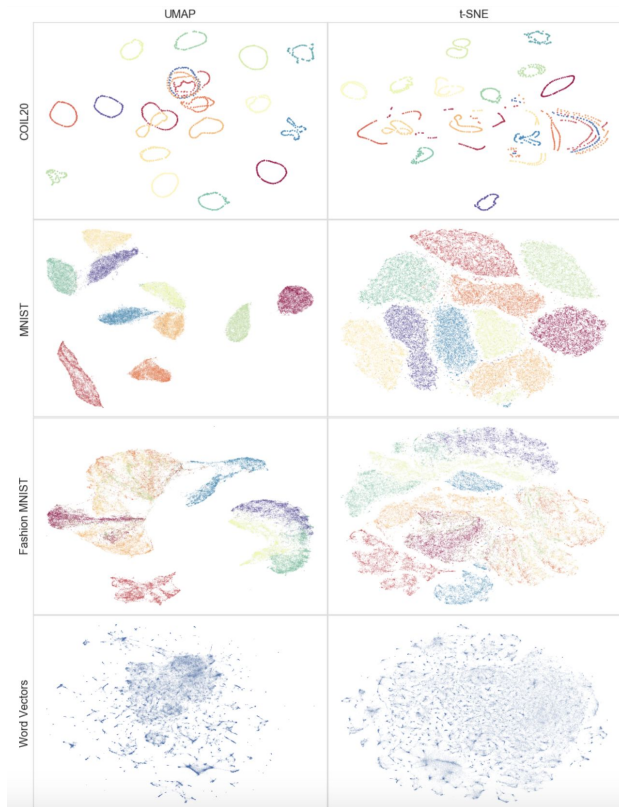
## Practical value

- Faster than t-SNE
- Can efficiently embed into high dimensions (i.e. useful not just for visualization)
- Often seems to capture global structure better

## Theory

- Roughly: manifold learning combined with explicit topology
- In detail: I don't completely understand the theory!
  - This note does an amazing job of extracting key bits of UMAP paper:  
<https://www.math.upenn.edu/~jhansen/2018/05/04/UMAP/>

# UMAP: New kid on the block



Comparison of UMAP (left) and t-SNE (right) from McInnes & Healy.

Global structure does seem to emerge more in UMAP.

## For more

Let's compare in real-time on an audio data set!

[Comparative Audio Analysis With Wavenet, MFCCs, UMAP, t-SNE and PCA](#)

(Leon Fedden)



# Putting this together

The Beginner's Guide to Dimensionality Reduction  
Matthew Conlen and Fred Hohman

<https://idyll.pub/post/dimensionality-reduction-293e465c2a3443e8941b016d/>

(just Google "Beginner's Guide to Dimensionality Reduction")

# Pitfalls of high-dimensional space

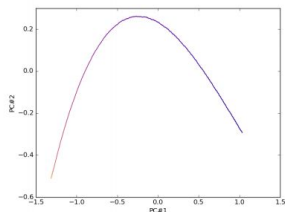
Geometry of high-dimensional space holds many surprises...

Be careful about interpreting visualizations!

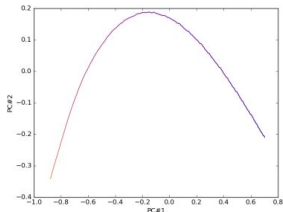
Adding "**usually**," "**most**," and "**approximately**" where appropriate:

- Two random vectors are perpendicular
- A standard Gaussian distribution is just a uniform distribution on a sphere
- A random matrix is a scalar multiple of an orthogonal matrix
- Random walks all have the same shape

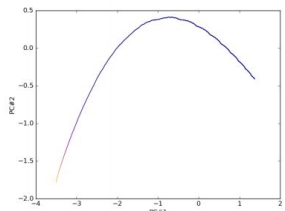
# Example: PCA of gradient descent trajectories



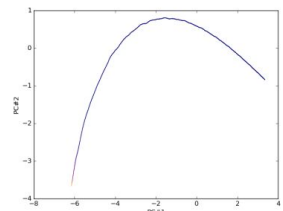
(a)  $\mu = 0.5$ ; PCs 1, 2



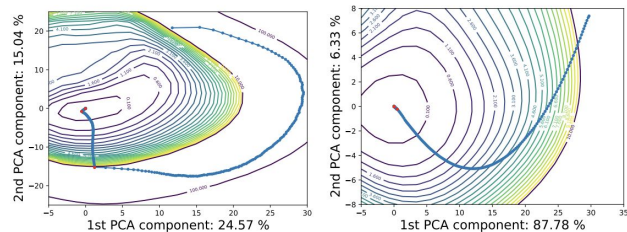
(b)  $\alpha = 0.0001$ ; PCs 1, 2



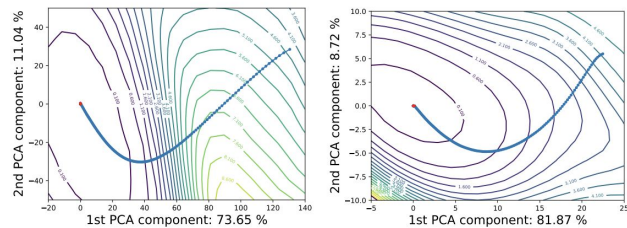
(d) MNIST  $\mu = 0.5$ ; PCs 1, 2



(e) MNIST  $\mu = 0.9$ ; PCs 1, 2



(a) SGD,  $WD=5e-4$

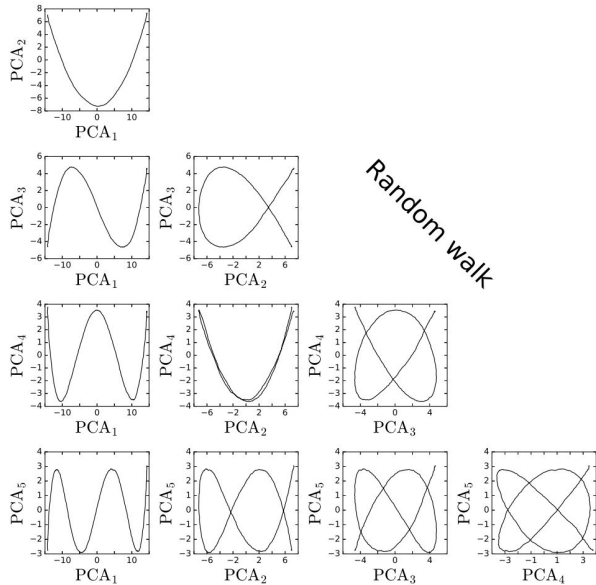


(c) SGD,  $WD=0$

Lorch, Visualizing Deep Network Training Trajectories, 2017

Li et al, Visualizing the Loss Landscape of Neural Nets, 2018

# How to interpret? Compare random walks



It turns out that principal components of a random walk in a high-dimensional space are (probably, approximately) cosines of various frequencies! (Antognini, Sohl-Dickstein)

Can also see this via Karhunen-Loeve theorem for Brownian motion.

**Important:** This doesn't invalidate work that uses PCA to look at SGD trajectories. But it changes how we read the visualizations: the interesting parts are differences from Lissajous patterns, not similarities.

# Lesson

If you see something interesting in high-dimensional space...

**compare to a random baseline!**

# Model interpretability example

Multi-lingual translation

What does the language embedding space look like?

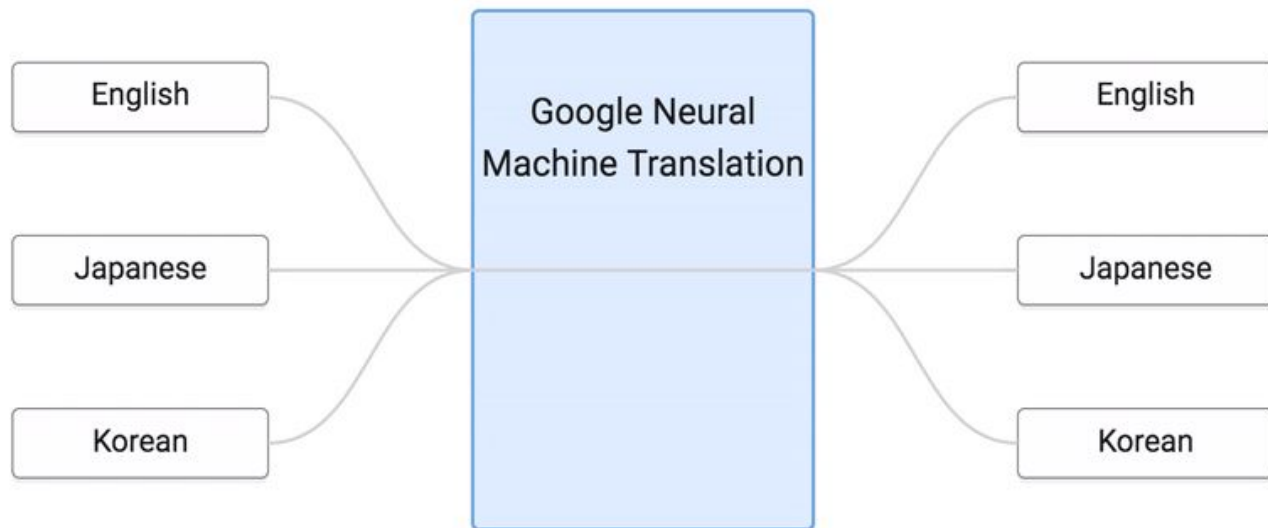
<https://arxiv.org/abs/1611.04558>

Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation

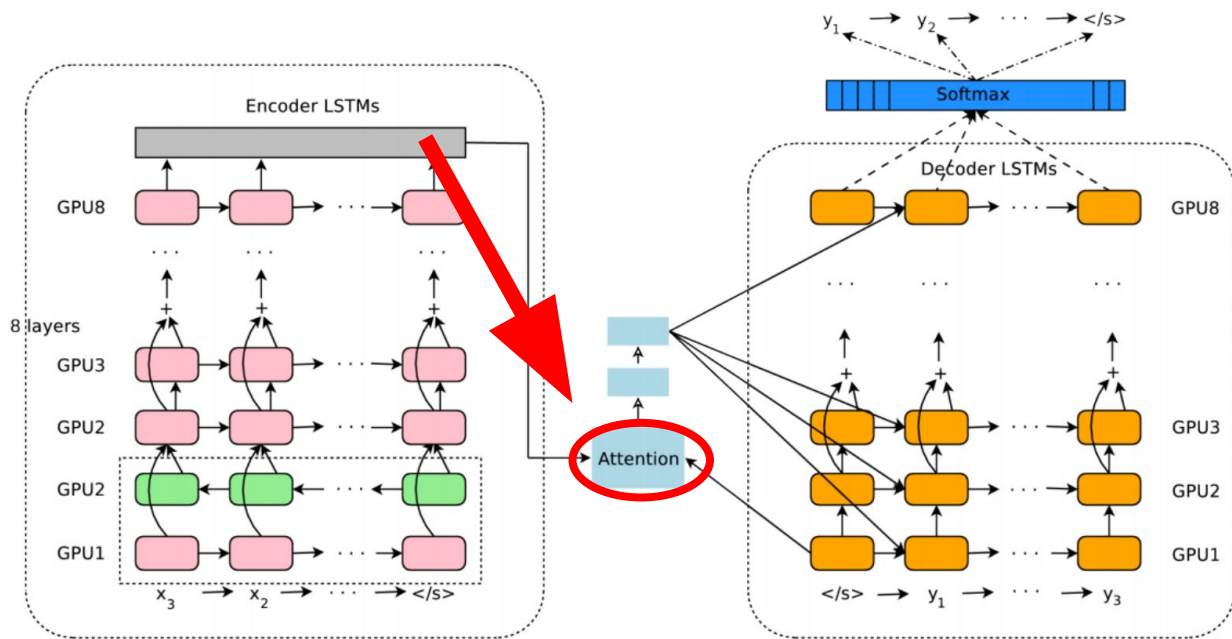
Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, Jeffrey Dean

**Training:** English  $\leftarrow \rightarrow$  Japanese  
English  $\leftarrow \rightarrow$  Korean  
Japanese  $\leftarrow \rightarrow$  Korean (zero shot)

Training



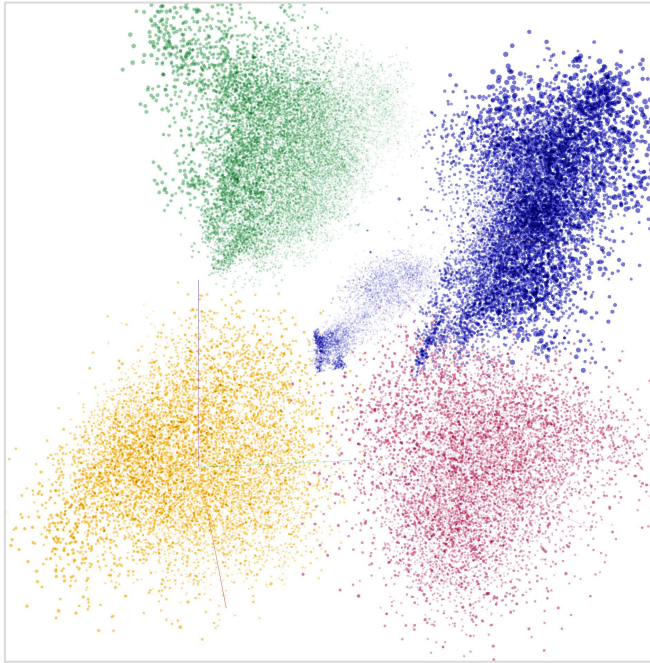
# Visualize internal representation ("embedding space")





# Research question

What does the multi language embedding space look like?



or

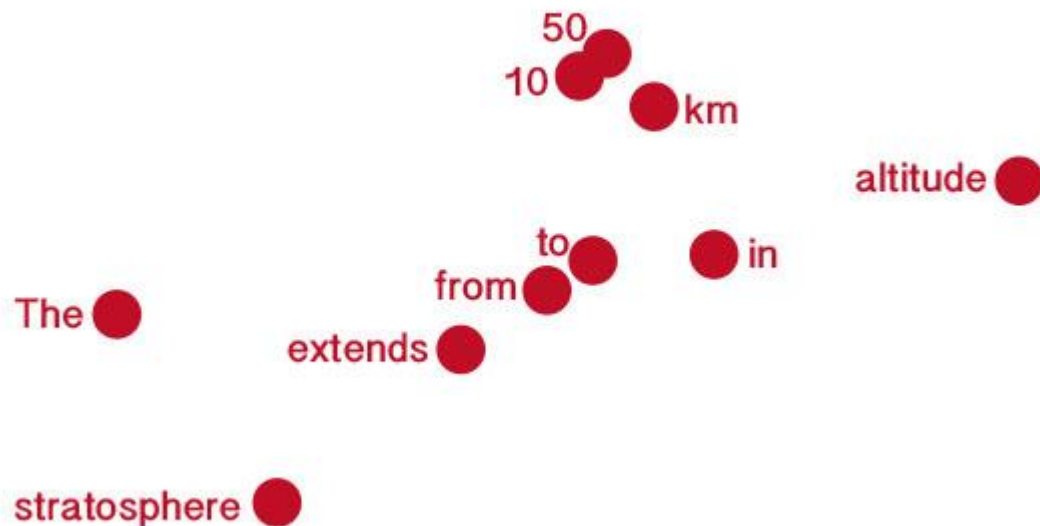


Note: not real data

What does a sentence look like in embedding space?  
(points in 1024-dim space: the data that the decoder receives)

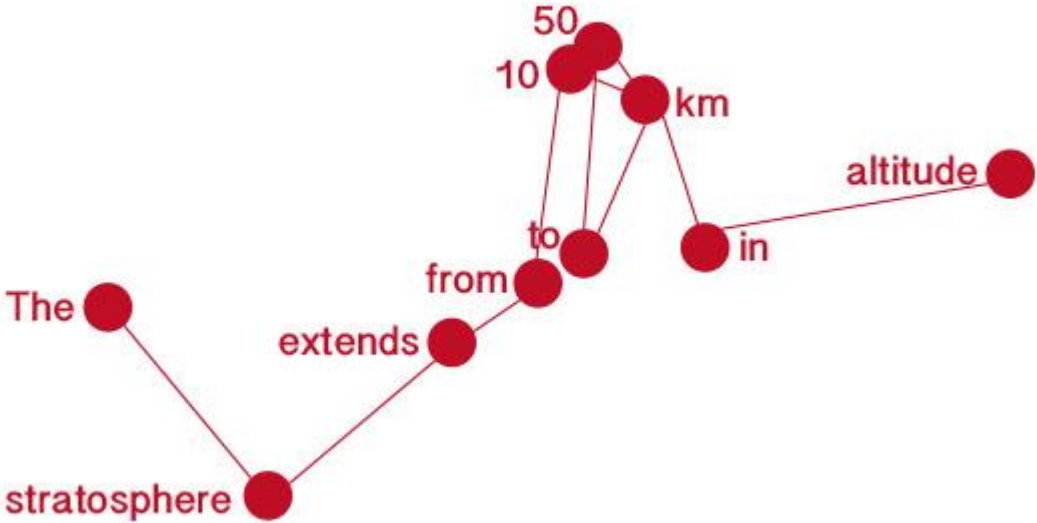
E.g. “*The stratosphere extends from 10km to 50km in altitude*”

# What does a sentence look like in embedding space?



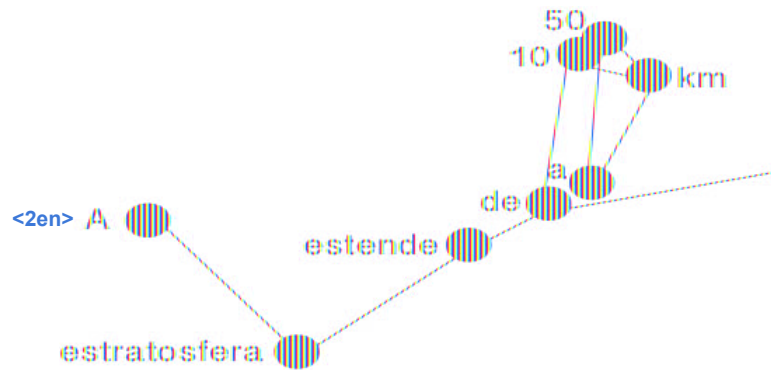
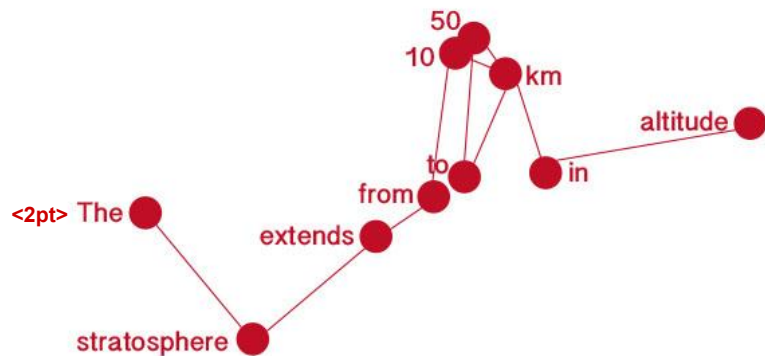
Note: simplification of real situation!

# What does a sentence look like in embedding space?



What do **parallel** sentences look like in embedding space?  
(same meaning, different language)

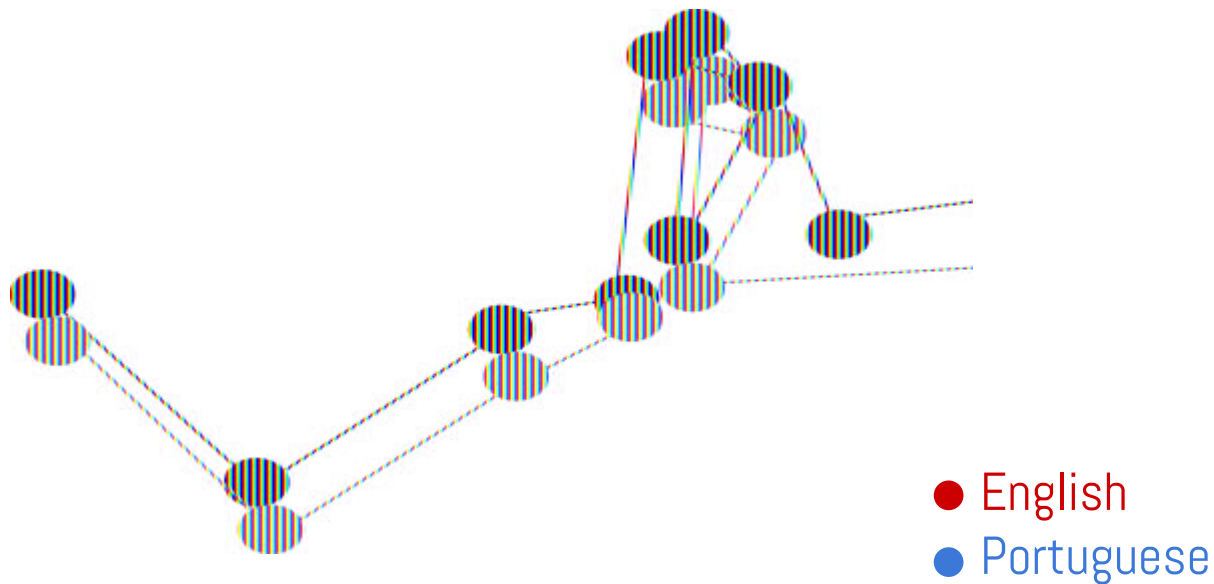
like this?



- English
- Portuguese

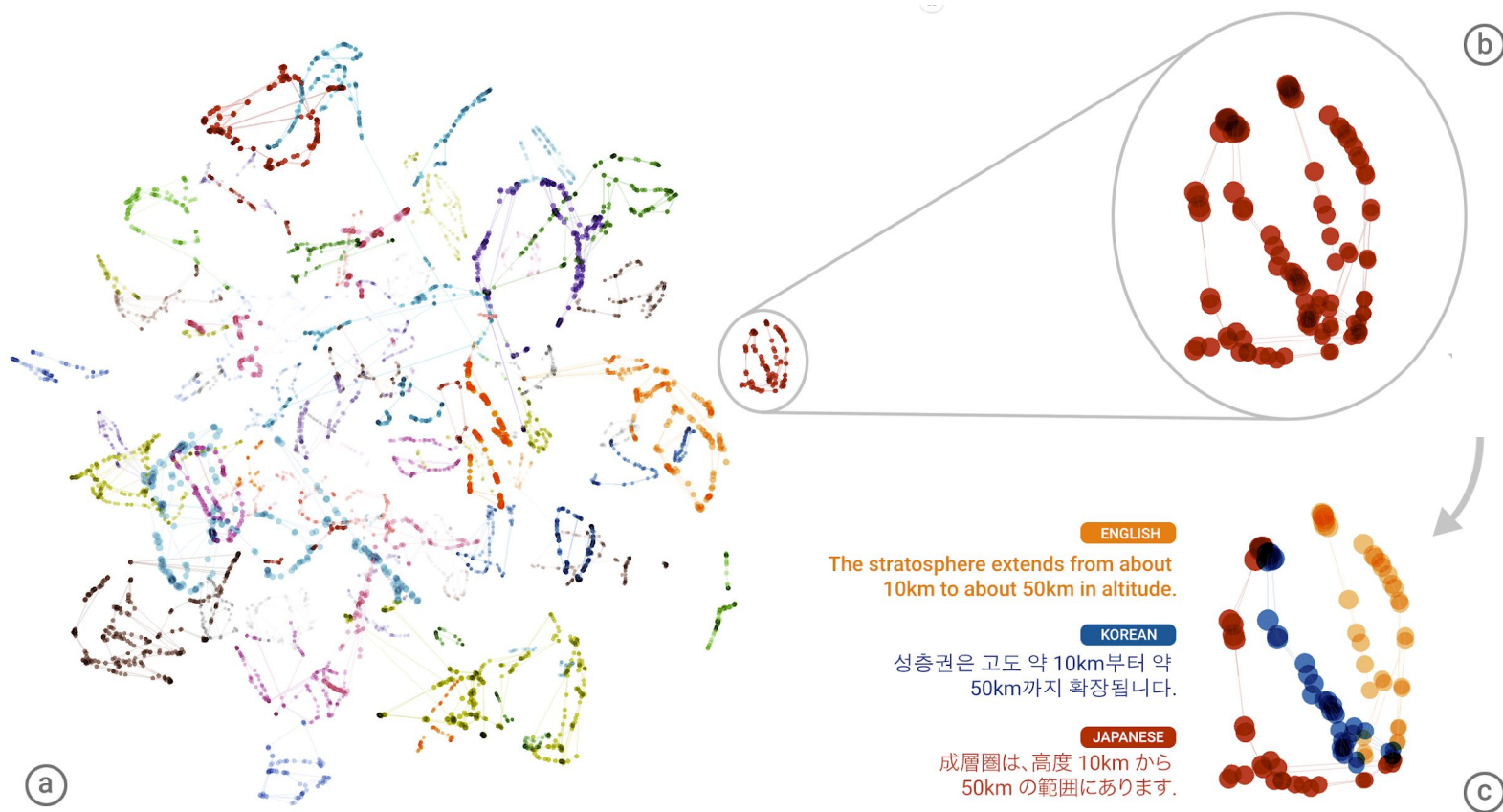
What do **parallel** sentences look like in embedding space?  
(same meaning, different language)

or like this?



# Interlingua?

Sentences with the same meaning mapped to similar regions regardless of language!



Distance between bridge / non-bridge sentences is inversely related to translation quality

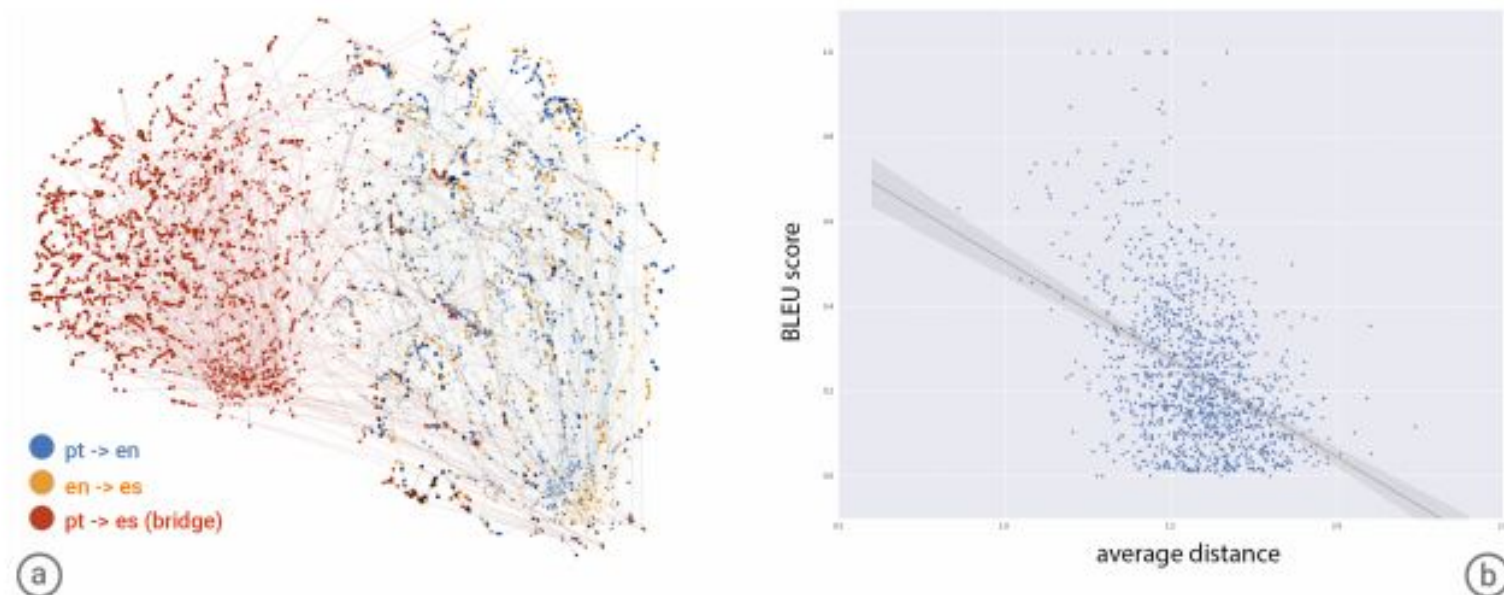


Figure 3: (a) A bird's-eye view of a t-SNE projection of an embedding of the model trained on Portuguese→English (blue) and English→Spanish (yellow) examples with a Portuguese→Spanish zero-shot bridge (red). The large red region on the left primarily contains the zero-shot Portuguese→Spanish translations. (b) A scatter plot of BLEU scores of zero-shot translations versus the average point-wise distance between the zero-shot translation and a non-bridged translation. The Pearson correlation coefficient is  $-0.42$ .



## 5. Education and communication

Education & communication  
for technical audiences

# TensorFlow Playground

playground.tensorflow.org

## DATA

Which dataset do you want to use?



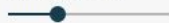
Ratio of training to test data: 50%



Noise: 0



Batch size: 10



REGENERATE

## INPUT

Which properties do you want to feed in?

$X_1$   
 $X_2$   
 $X_1^2$   
 $X_2^2$   
 $X_1 X_2$   
 $\sin(X_1)$   
 $\sin(X_2)$



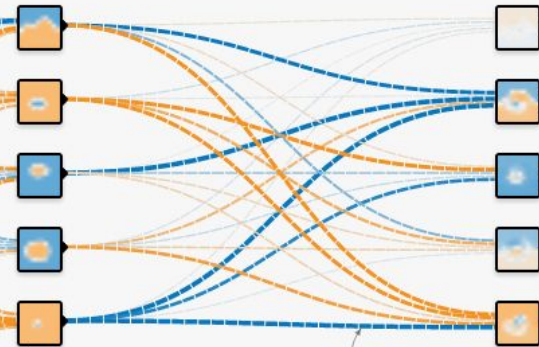
+ - 2 HIDDEN LAYERS

+ -

5 neurons

+ -

5 neurons

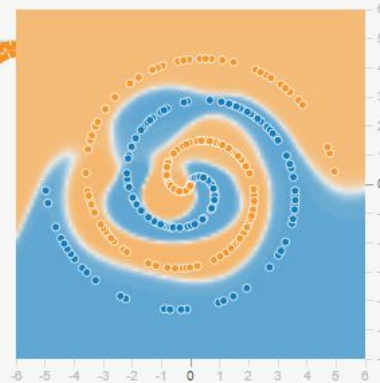


This is the output from one neuron. Hover to see it larger.

The outputs are mixed with varying weights, shown by the thickness of the lines.

## OUTPUT

Test loss 0.013  
Training loss 0.013



Colors shows data, neuron and weight values.



Show test data

Discretize output

# GAN Lab

<https://poloclub.github.io/ganlab/>



Aug. 14, 2018

## Distill Update 2018

EDITORIAL

An Update from the Editorial Team

July 25, 2018

## Differentiable Image Parameterizations

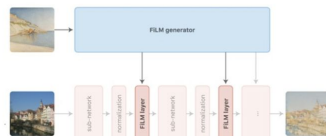
A powerful, under-explored tool for neural network visualizations and art.



July 9, 2018

## Feature-wise transformations

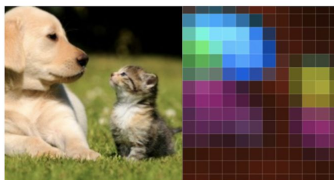
A simple and surprisingly effective family of conditioning mechanisms.



March 6, 2018

## The Building Blocks of Interpretability

Interpretability techniques are normally studied in isolation. We explore the powerful interfaces that arise when you combine them—and the rich structure of this combinatorial space.



Dec. 4, 2017

## Using Artificial Intelligence to Augment Human Intelligence

COMMENTARY

By creating user interfaces which let us work with the



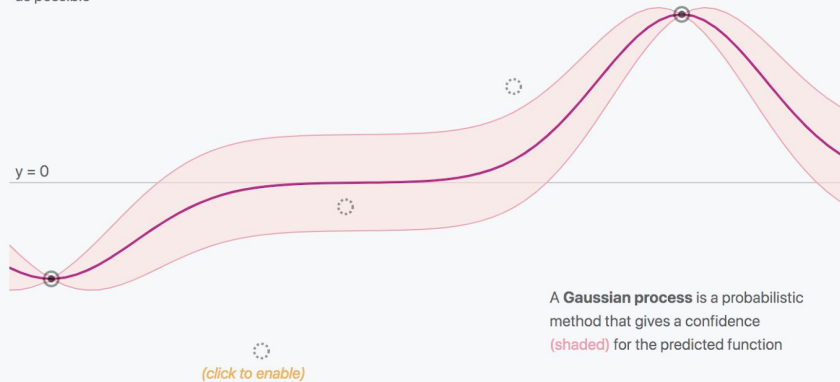
# Distill.pub

Editors: Carter, Olah, Satyanarayan

# A Visual Exploration of Gaussian Processes

How to turn a collection of small building blocks into a versatile tool for solving regression problems.

**Regression** is used to find a function (line) that represents a set of data points as closely as possible



Education & communication  
for non-technical audiences

# Attacking discrimination with smarter machine learning

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

## Simulating loan decisions for different groups

Drag the black threshold bars left or right to change the cut-offs for loans.  
Click on different preset loan strategies.

### Loan Strategy

Maximize profit with:

**MAX PROFIT**

No constraints

**GROUP UNAWARE**

Blue and orange thresholds are the same

**DEMOGRAPHIC PARITY**

Same fractions blue / orange loans

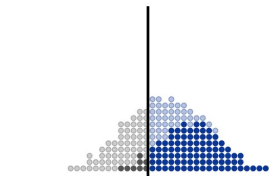
**EQUAL OPPORTUNITY**

Same fractions blue / orange loans to people who can pay them off

### Blue Population

0 10 20 30 40 50 60 70 80 90 100

loan threshold: 50

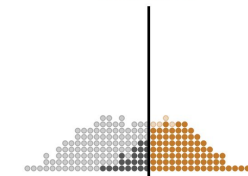


denied loan / would default    granted loan / defaults  
denied loan / would pay back    granted loan / pays back

### Orange Population

0 10 20 30 40 50 60 70 80 90 100

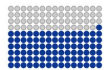
loan threshold: 50



denied loan / would default    granted loan / defaults  
denied loan / would pay back    granted loan / pays back

Total profit = 19600

**Correct 76%**  
loans granted to paying applicants and denied to defaulters



**Incorrect 24%**  
loans denied to paying applicants and granted to defaulters



**True Positive Rate 92%**  
percentage of paying applications getting loans

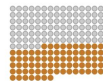


Profit: -700

**Positive Rate 66%**  
percentage of all applications getting loans



**Correct 87%**  
loans granted to paying applicants and denied to defaulters



**Incorrect 13%**  
loans denied to paying applicants and granted to defaulters



**True Positive Rate 78%**  
percentage of paying applications getting loans



Profit: 20300

**Positive Rate 41%**  
percentage of all applications getting loans



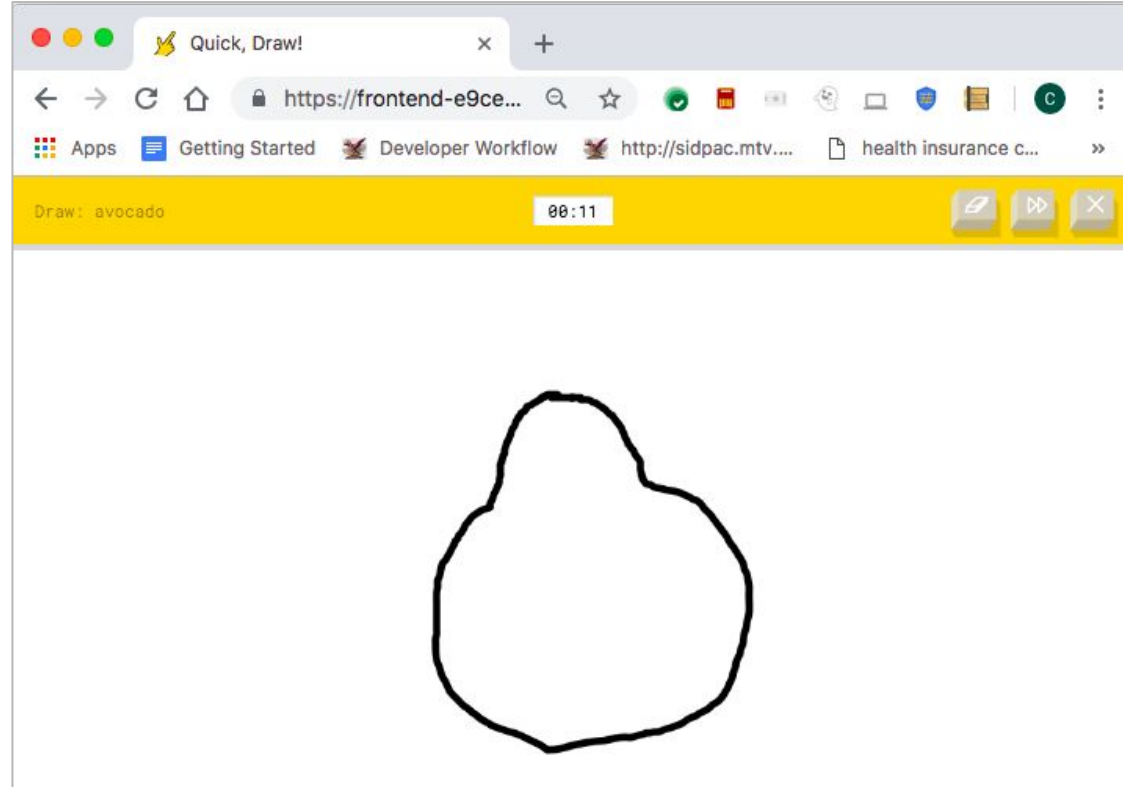
Transform math into a visual, interactive simulation that can be used by a broader set of stakeholders such as policymakers and regulators.

Wattenberg, Viégas, Hardt. 2016





On Quickdraw, users draw common objects (e.g. avocado), then see if the algorithm has correctly recognized the object.



You were asked to draw avocado, and the neural net did not recognize it.


After users see the recognition result, Quickdraw shows **visual examples** to help users understand the algorithm's reasoning.

For example, it shows examples of what typical avocados look like.











You were asked to draw avocado

You drew this, and the neural net didn't recognize it.



---


What does it think avocado looks like?  
It learned by looking at these examples drawn by other people.

It also shows a **visual diff** between the user's drawing and the most-similar drawings from alternative classes.




You were asked to draw avocado

You drew this, and the neural net didn't recognize it.



---

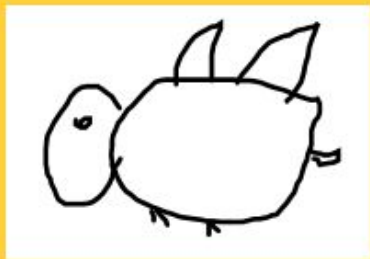
It thought your drawing looked more like these:

<p>Closest match <i>pear</i></p> 	<p>2<sup>nd</sup> closest match <i>onion</i></p> 	<p>3<sup>rd</sup> closest match <i>potato</i></p> 
---	--	---



# You were asked to draw bee

You drew this, and the neural net didn't recognize it.



---

It thought your drawing looked more like these:

Closest match  
sea turtle



2<sup>nd</sup> closest match  
mouse



3<sup>rd</sup> closest match  
shark



Compare user input to classes system thought were closest

---

What does it think bee looks like?

It learned by looking at these examples drawn by other people.



Show examples of what the system expected for the class in question

Illustrate latent space to users

# Visual Analytics in Deep Learning: An Interrogative Survey for the Next Frontiers

Hohman, Kahng, Pienta, Chau

Work	WHY				WHO			WHAT					HOW					WHEN		WHERE	
	4.1	4.2	4.3	4.4	5.1	5.2	5.3	6.1	6.2	6.3	6.4	6.5	7.1	7.2	7.3	7.4	7.5	7.6	8.1	8.2	9.2
Abadi, et al., 2016 [27]	■	■	■		■	■					■			■					■	■	arXiv
Bau, et al., 2017 [28]	■		■		■														■	■	CVPR
Bilal, et al., 2017 [29]	■	■			■					■									■	■	TVCG
Bojarski, et al., 2016 [30]	■	■			■				■										■	■	arXiv
Bruckner, 2014 [31]	■	■			■			■	■				■						■		MS Thesis
Carter, et al., 2016 [32]	■			■	■	■	■			■	■	■							■	■	Distill
Cashman, et al., 2017 [33]	■	■			■	■			■	■	■								■	■	VADL
Chae, et al., 2017 [34]	■	■			■					■				■					■		VADL
Chung, et al., 2016 [35]	■	■			■			■	■	■	■		■	■	■	■			■		FILM
Goyal, et al., 2016 [36]	■						■		■										■	■	arXiv
Harley, 2015 [37]	■			■			■	■	■				■						■	■	ISVC
Hohman, et al., 2017 [38]	■		■	■			■						■						■	■	CHI
Kahng, et al., 2018 [39]	■	■			■	■		■					■	■					■	■	TVCG
Karpathy, et al., 2015 [40]	■				■					■	■	■		■					■	■	arXiv
Li, et al., 2015 [41]	■				■	■				■	■	■		■					■	■	arXiv
Liu, et al., 2017 [14]	■	■			■			■	■	■	■	■	■						■		TVCG
Liu, et al., 2018 [42]	■	■			■			■	■	■	■	■	■						■		TVCG
Ming, et al., 2017 [43]	■		■		■					■									■	■	VAST
Norton & Qi, 2017 [44]	■	■		■	■	■	■												■	■	VizSec
Olah, 2014 [45]	■			■	■		■				■		■						■	■	Web
Olah, et al., 2018 [46]	■			■	■	■	■	■		■	■	■		■					■	■	Distill
Pezzotti, et al., 2017 [47]	■	■			■					■	■	■	■	■	■	■			■		TVCG

# Resources

ML-specific	General visualization & design	Implementation
<p data-bbox="104 369 409 412"><a href="#">Stanford CS 231</a></p> <p data-bbox="104 430 309 473">Sequences</p> <ul data-bbox="142 484 382 565" style="list-style-type: none"><li>- <a href="#">Seq2Seq-vis</a></li><li>- <a href="#">LSTMvis</a></li></ul> <p data-bbox="104 635 486 679"><a href="#">Embedding Projector</a></p> <p data-bbox="104 696 235 740"><a href="#">Facets</a></p> <p data-bbox="104 757 235 801"><a href="#">Lobe.ai</a></p> <p data-bbox="104 871 614 976"><a href="#">A Survey: Visual Analytics in Deep Learning</a> (Hohman et al)</p>	<p data-bbox="699 369 1116 412"><a href="#">Tableau</a> (desktop app)</p> <ul data-bbox="738 430 1070 554" style="list-style-type: none"><li>- Commercial</li><li>- State of the art</li><li>- Industrial-strength</li></ul> <p data-bbox="699 565 1031 609"><a href="#">RawGraphs</a> (web)</p> <p data-bbox="699 620 1097 663"><a href="#">Flourish.studio</a> (web)</p> <p data-bbox="699 740 942 784"><a href="#">Color Brewer</a></p> <p data-bbox="699 801 819 845"><a href="#">Coblis</a></p> <ul data-bbox="738 856 1155 888" style="list-style-type: none"><li>- Colorblindness simulator</li></ul>	<p data-bbox="1298 369 1356 412"><a href="#">D3</a></p> <ul data-bbox="1336 430 1676 463" style="list-style-type: none"><li>- See also <a href="#">blocks.org</a></li></ul> <p data-bbox="1298 478 1503 521">Notebooks</p> <ul data-bbox="1336 532 1561 613" style="list-style-type: none"><li>- <a href="#">Observable</a></li><li>- <a href="#">Jupyter'</a></li></ul> <p data-bbox="1298 624 1491 668"><a href="#">Matplotlib</a></p> <p data-bbox="1298 685 1452 729"><a href="#">Three.js</a></p> <p data-bbox="1298 746 1464 790"><a href="#">Kepler.gl</a></p> <p data-bbox="1298 801 1414 845"><a href="#">Plotly</a></p>



# Visualization for Machine Learning



Fernanda Viégas @viegasf  
Martin Wattenberg @wattenberg  
Google Brain