LESSON 12

Data representation and Visualization for Machine Learning

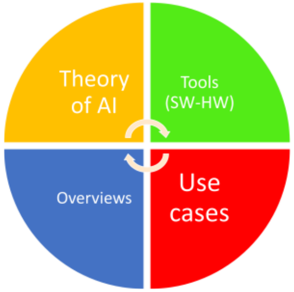




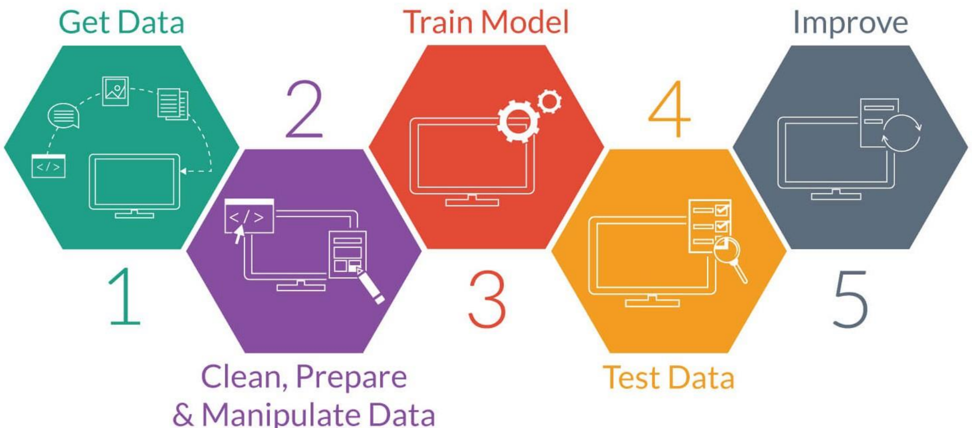


THEORY

Data Visualization for ML



Step 2 of the ML workflow: Substeps: Data visualization



2a) Data Munging/Wrangling

2b) Exploratory Data Analysis

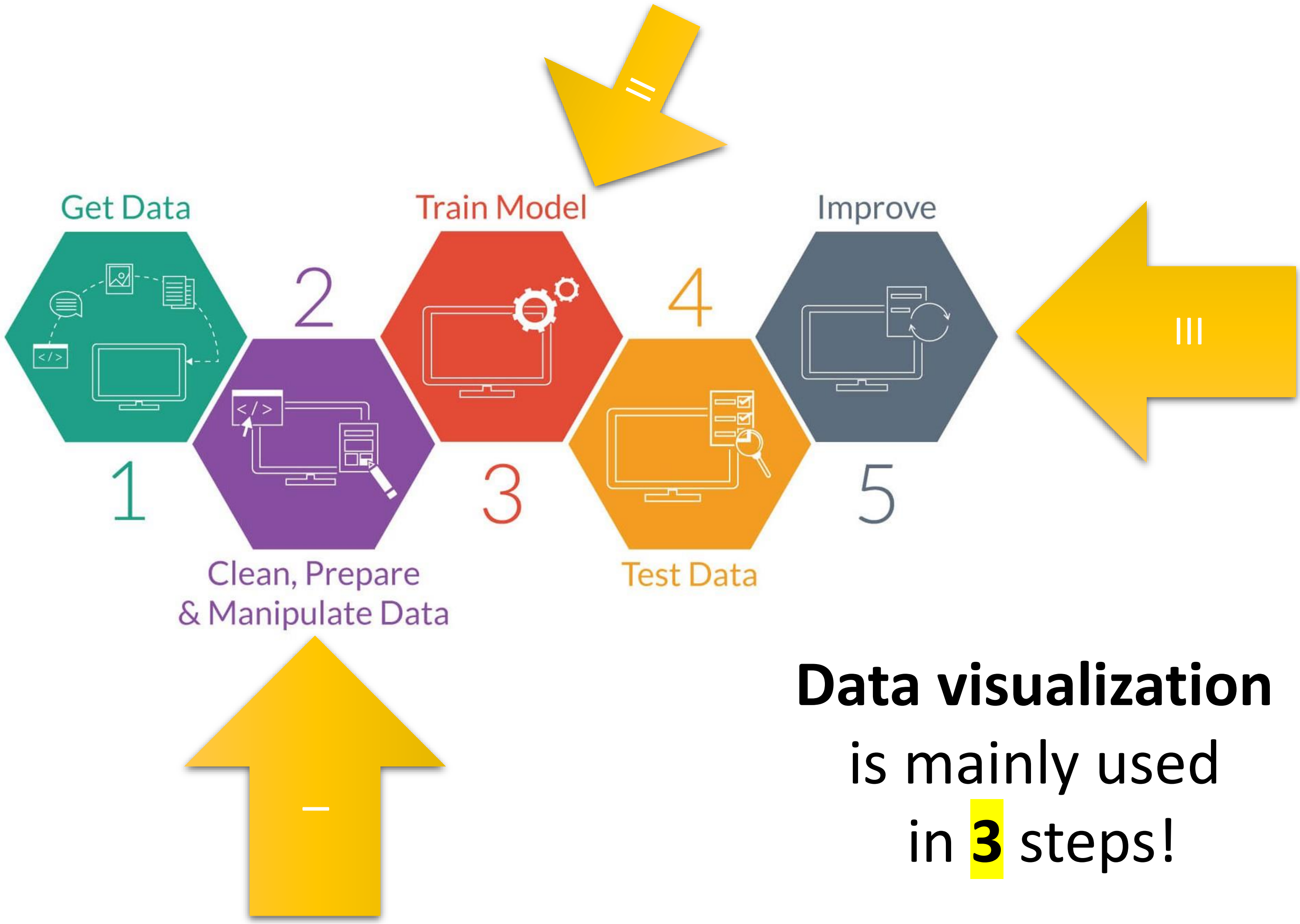
...

Data visualization

....

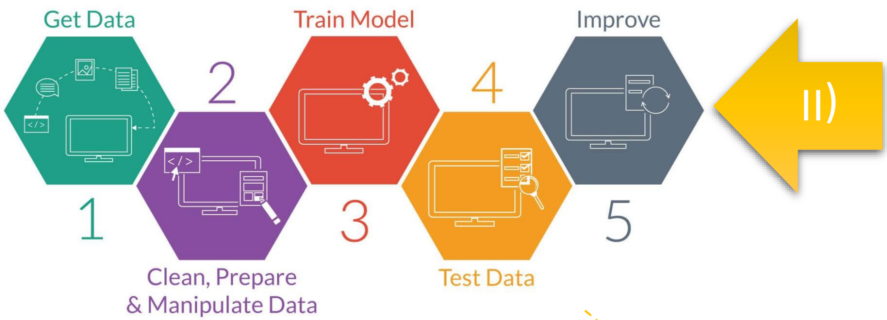
2c) Feature engineering

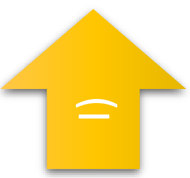
Step 2-5 of the ML workflow

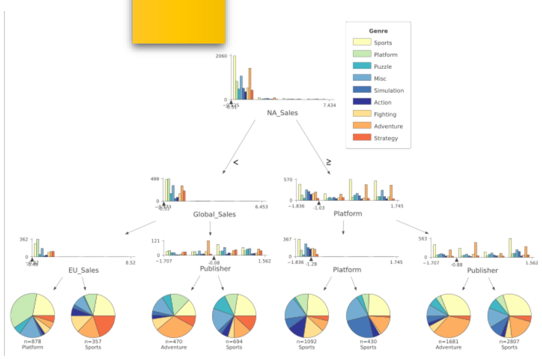


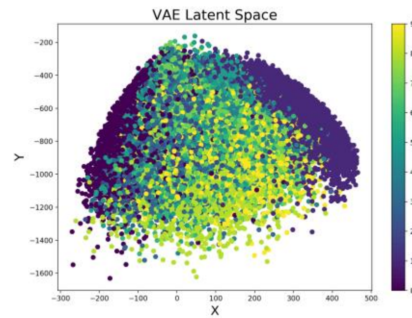
Data visualization
is mainly used
in **3** steps!

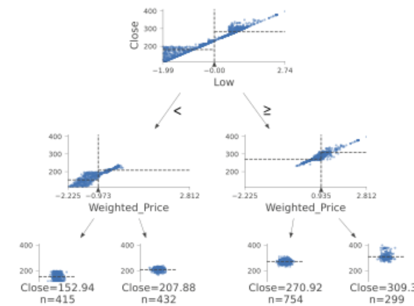
Step 2-5 of the ML workflow

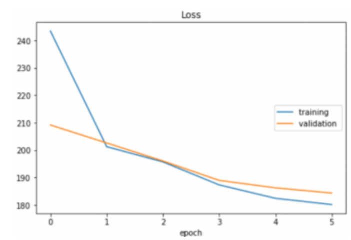


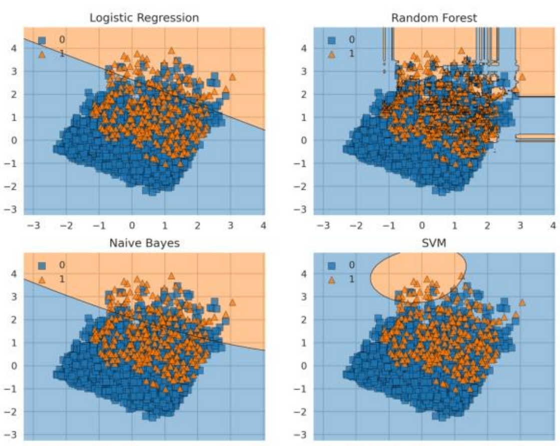


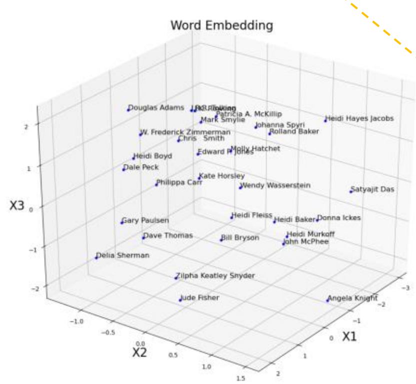








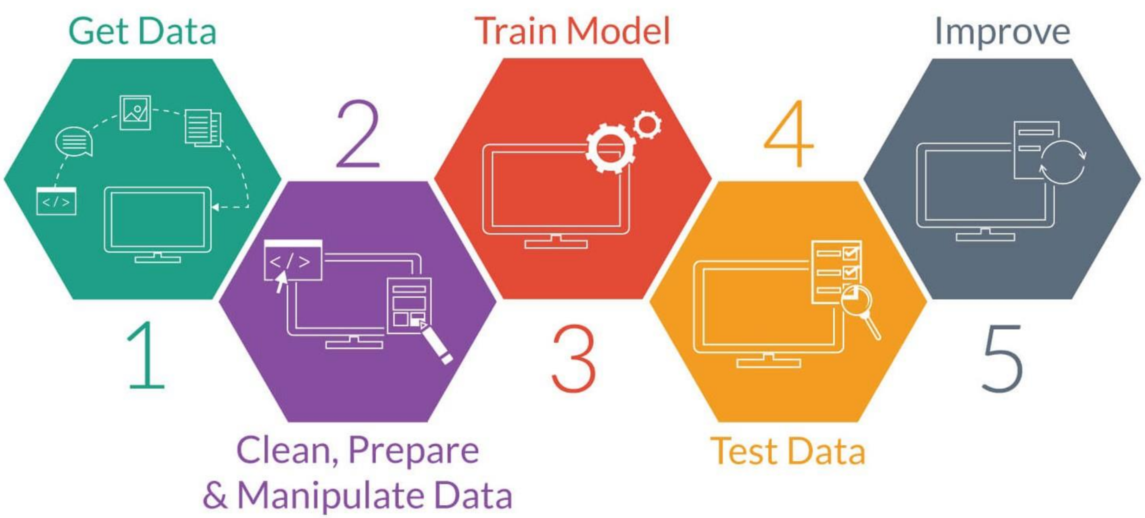


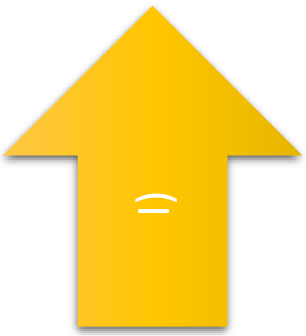


Visualization uses in ML

- Training Data
- Model Performance
- Interpretability + model inspection
- High-dimensional data
- Education and communication

Step 2 of the ML workflow





Y = FUNC(X)
Exploratory Data Analysis / Feature engineering

- Optimizing one single feature at time!

SINGLE FEATURE

Sample, sample, ……

X Features

5.1								
3.5								
1.4								
0.2								

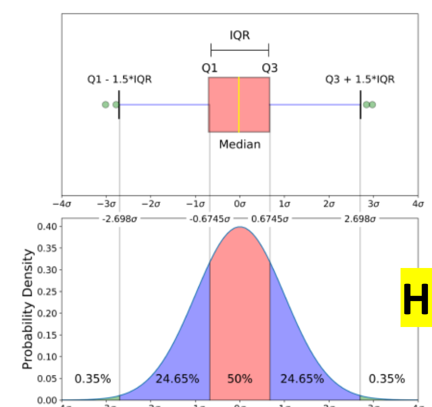
Class ID, or value to be learnt

Y

1	**0**	**3**	**1**					

Descriptive statistics

```
data = [1 2 3 4 50];

        % The arithmetic mean of the data:
  mean(data) → → ans: 12

        %The median of the data:
  median(data) → → ans: 3

        % The standard deviation:
std(data) → → ans: 21.27

        % The smallest value in the data:
  min(data) → → ans: 1

        % The largest value in the data:
  max(data ) → → ans: 50
```



Boxplots

Outliers/Errors

Histograms

$$x' = (x - x_{min})/(x_{max} - x_{min})$$

Simple data processing and normalization

Y = FUNC(X)
Exploratory Data Analysis / Feature engineering

- Exploring or Optimizing all features at the same time!

ALL FEATURES

- X

Sample, sample, ...

Features

5.1								
3.5								
1.4								
0.2								

How can you understand what is happening here in multiple dimensions?

- Y

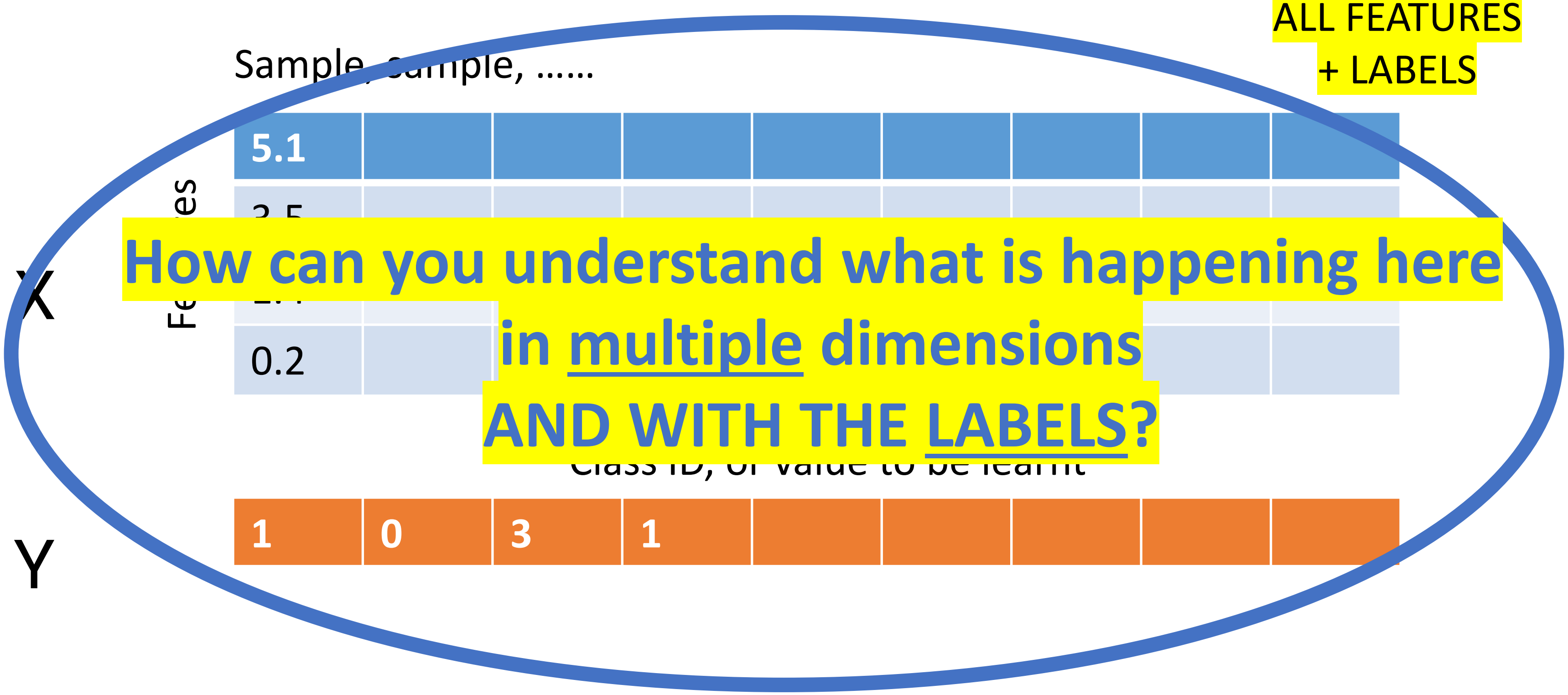
Class ID, or value to be learnt

1	**0**	**3**	**1**					

Y = FUNC(X)
Exploratory Data Analysis / Feature engineering

- Exploring or Optimizing all features + Labels at the same time!

ALL FEATURES
+ LABELS



- X
- Y

Example: Visualizing CIFAR-10

CIFAR-10 and CIFAR-100 datasets

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.

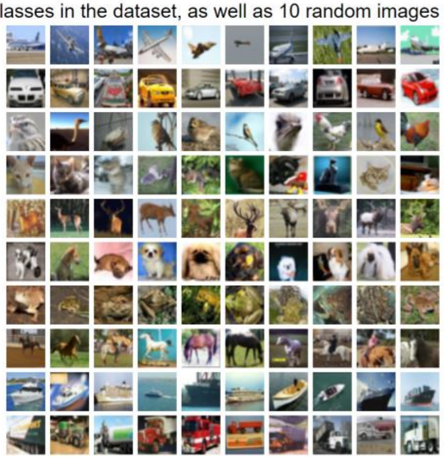
The CIFAR-10 dataset

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:

airplane
automobile
bird
cat
deer
dog
frog
horse
ship
truck



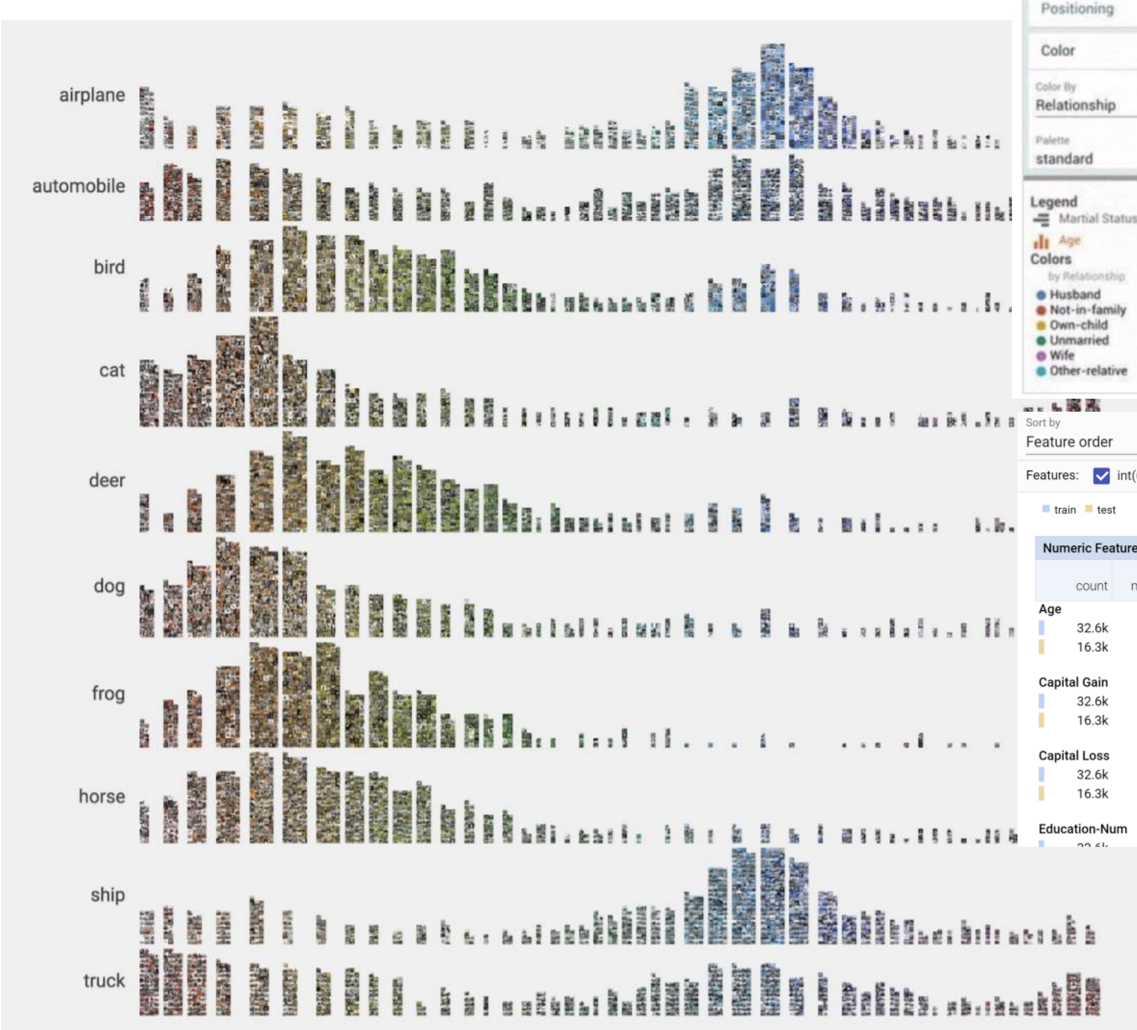
The classes are completely mutually exclusive. There is no overlap between automobiles and trucks. "Automobile" includes sedans, SUVs, things of that sort. "Truck" includes only big trucks. Neither includes pickup trucks.

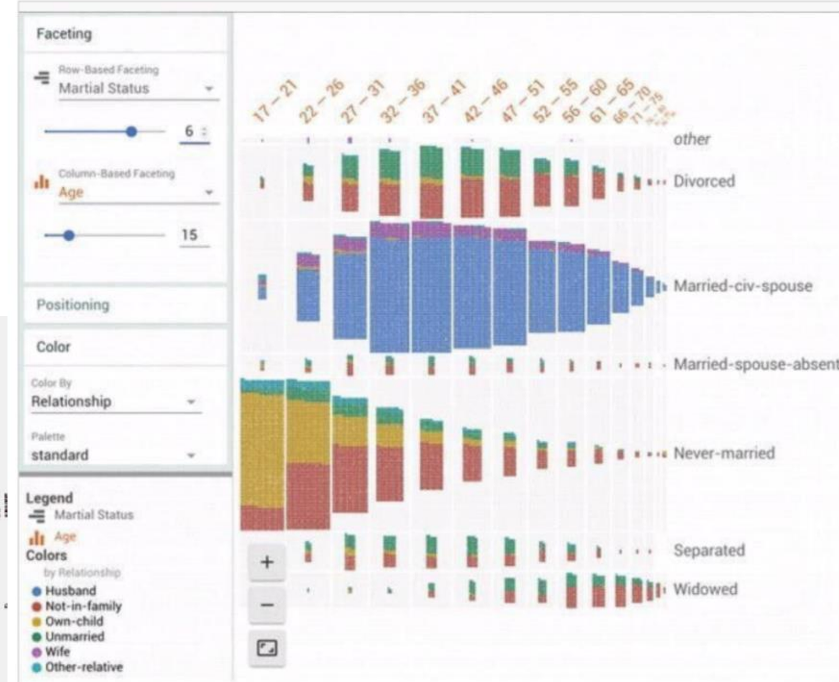
Download

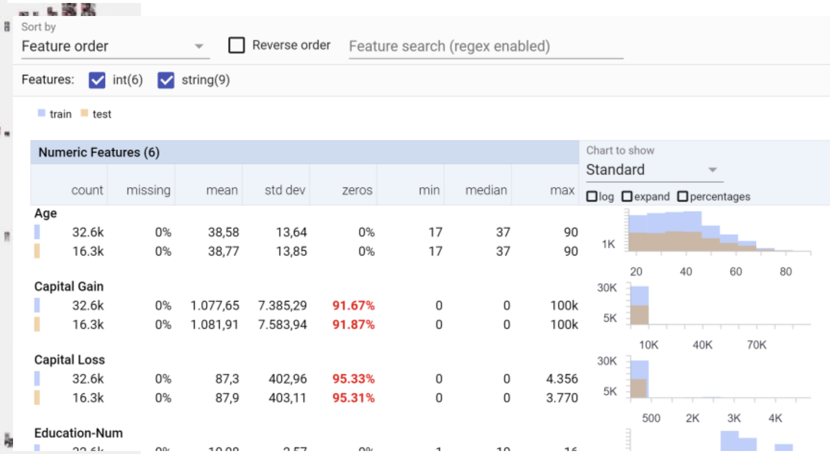
If you're going to use this dataset, please cite the tech report at the bottom of this page.

Facets:

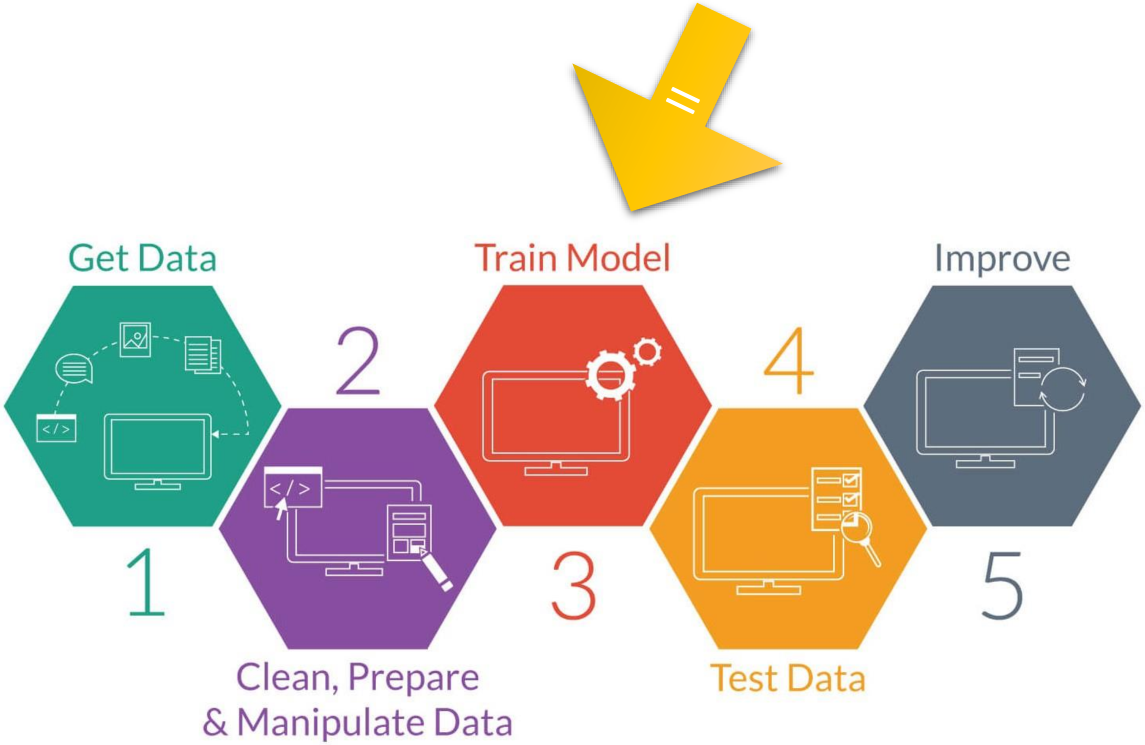
Open Source Visualization Tool for Machine Learning Training Data





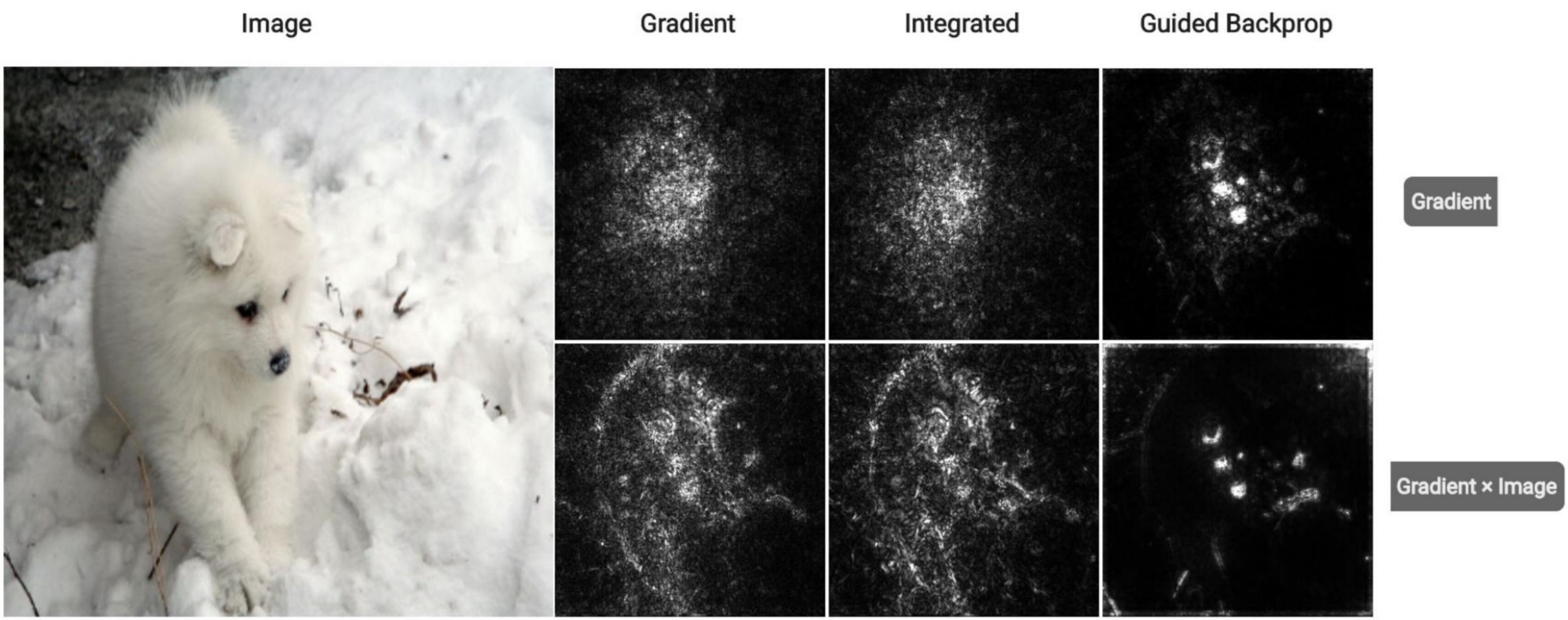


Step 3-4 of the ML workflow



Saliency maps: example

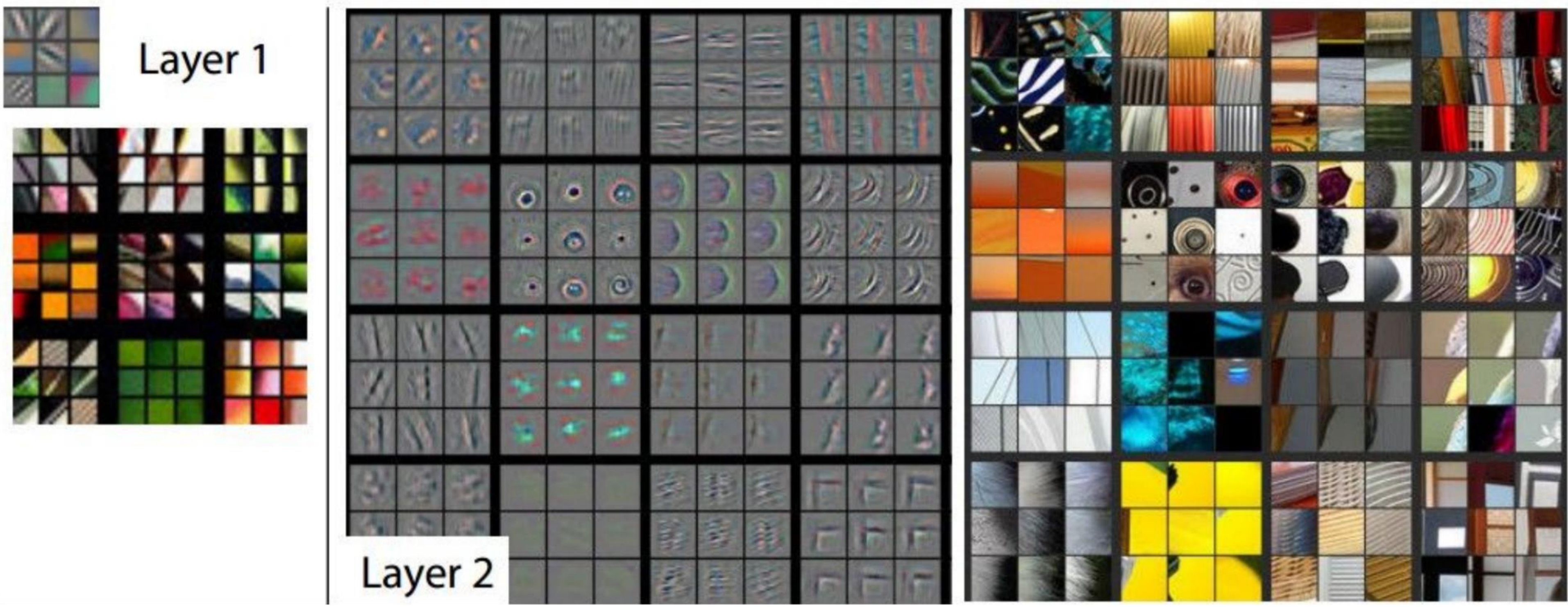
(in the next lessons)….Take the gradient of a class prediction neuron with respect to the input pixels. SmoothGrad technique often significantly denoises the sensitivity mask.
This technique adds pixel-wise Gaussian noise to many copies of the image, and simply averages the resulting gradients.



Determine which features of the input (pixels, for images) were important for the prediction

https://pair-code.github.io/saliency/

Visualizing arbitrary neurons in CNNs



Gray: kernels
Colorful squares: maximal examples from an image data set

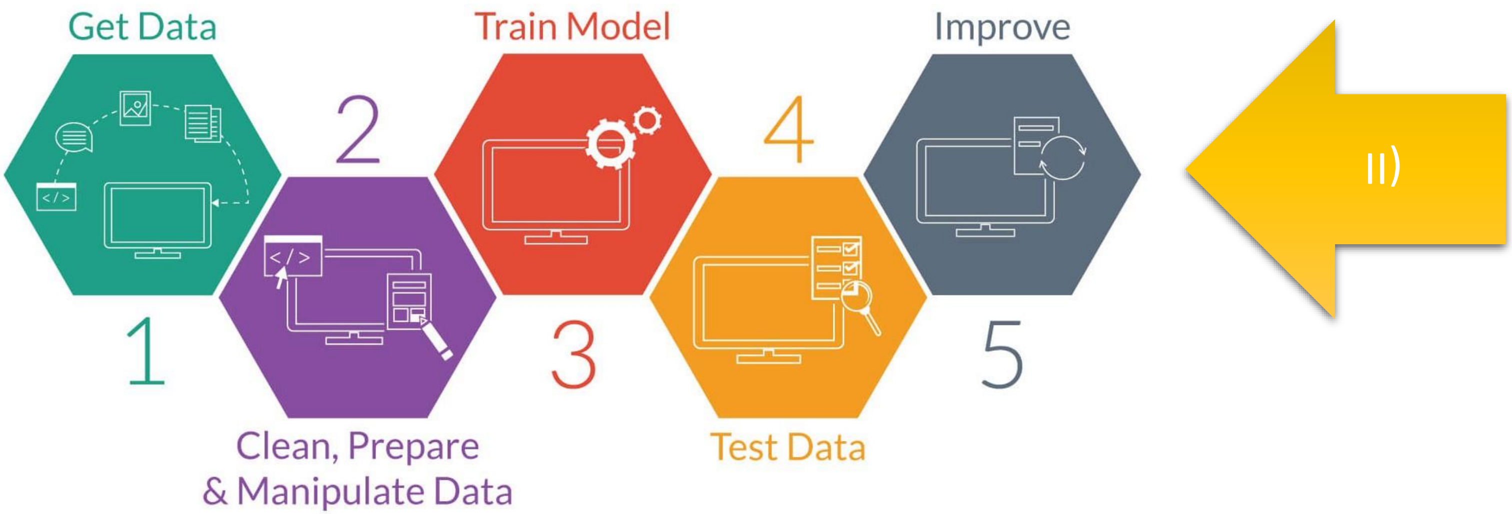
Built your ML model visualization…

Maybe a little too dense?

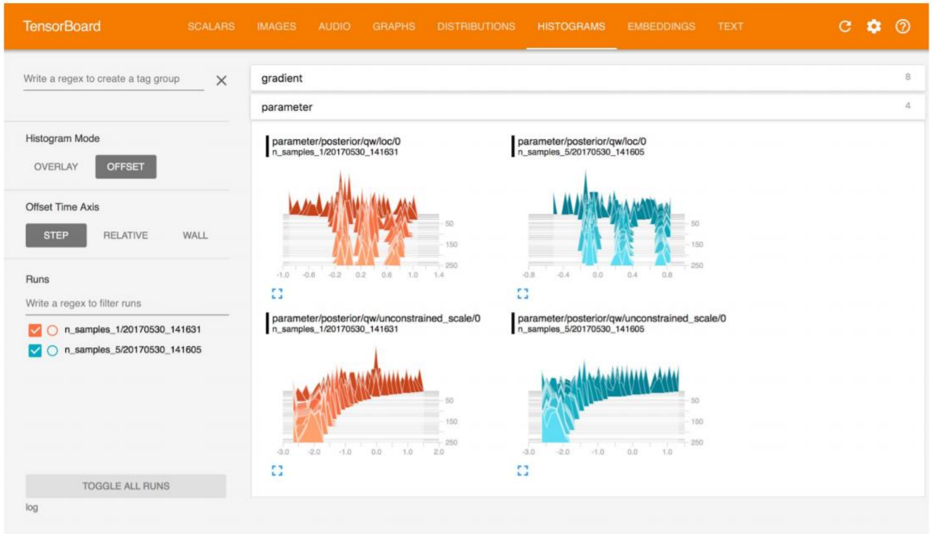
DQNViz: A Visual Analytics Approach to Understand Deep Q-Networks

(DQN) is one type of deep reinforcement learning model

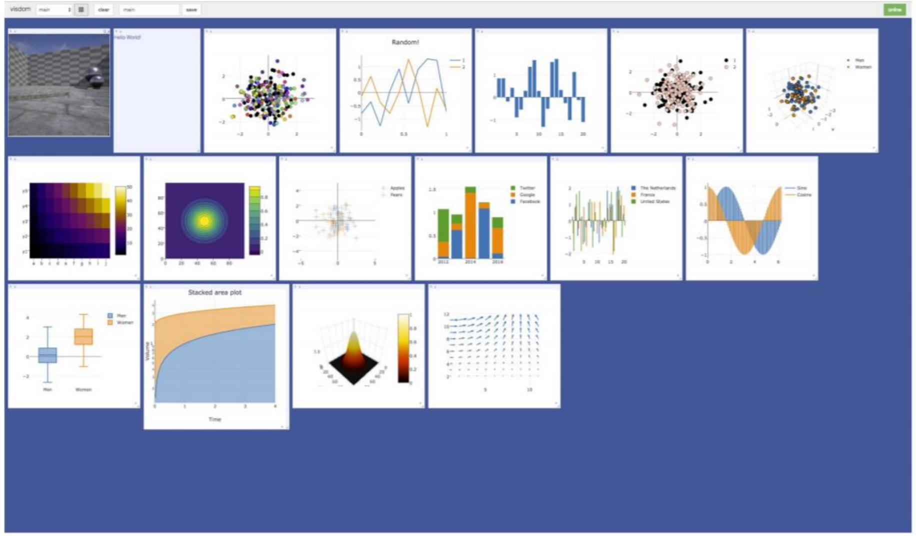
Step 5 of the ML workflow



Monitoring dashboards
Standard visualization tools



TensorBoard



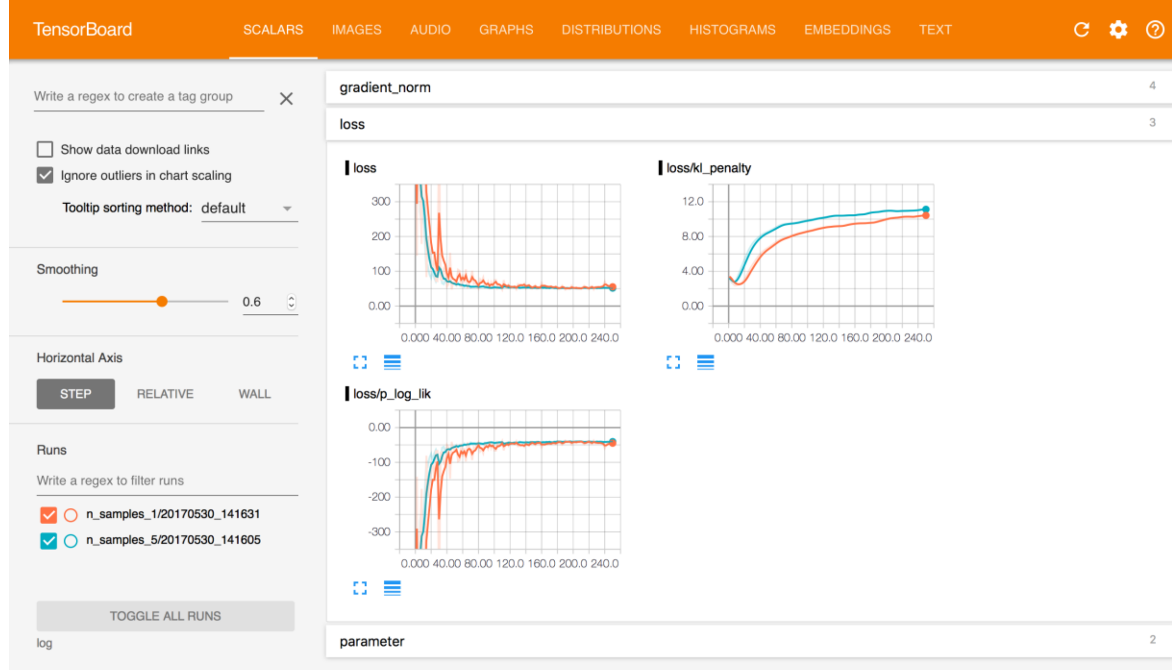
Visdom

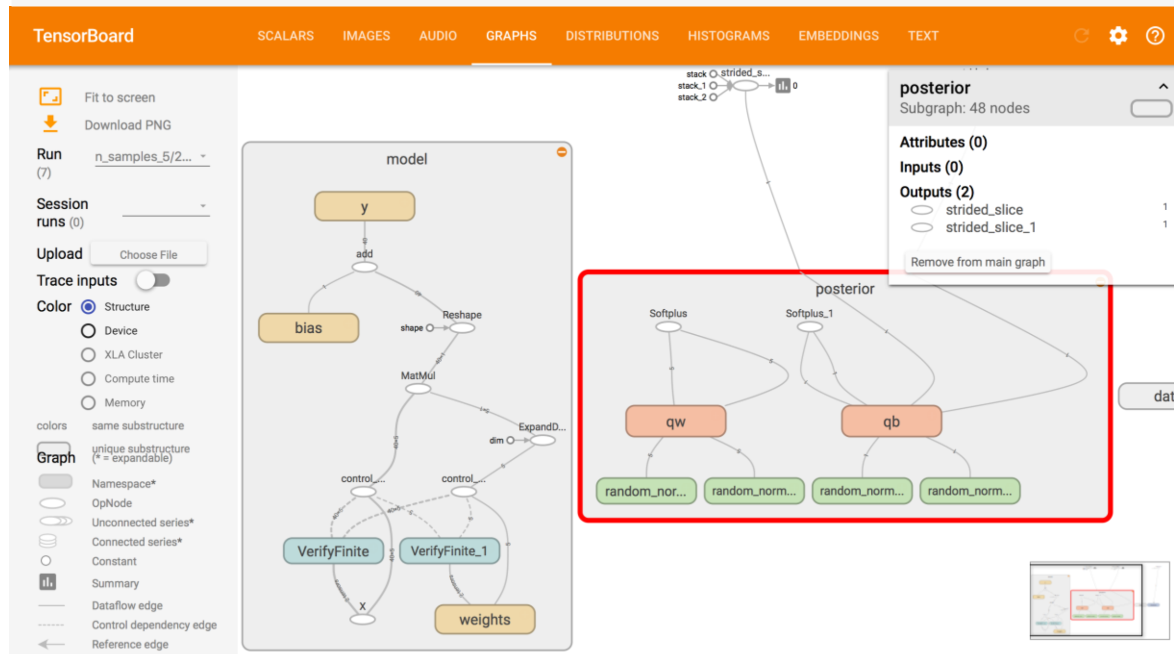
Monitoring dashboards
Standard visualization tools

TensorBoard (Google)

TensorBoard provides a suite of visualization tools to make it easier to understand, debug, and optimize your models

You can use it "to visualize your TensorFlow graph, plot quantitative metrics about the execution of your graph, and show additional data like images that pass through it"





Monitoring dashboards Standard visualization tools

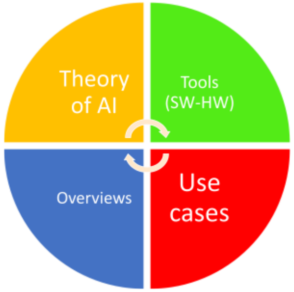
Visdom
(Meta AI)

Visdom is a visualization tool that generates rich visualizations of live data to help researchers and developers stay on top of their scientific experiments that are run on remote servers.
Visualizations in Visdom can be viewed in browsers and easily shared with others.

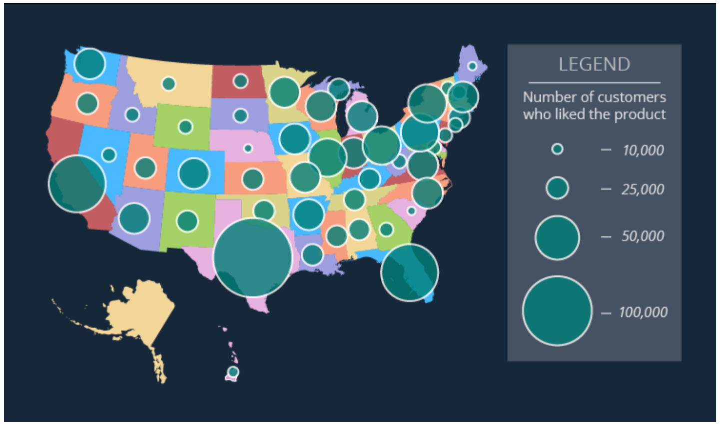



THEORY
Data visualization

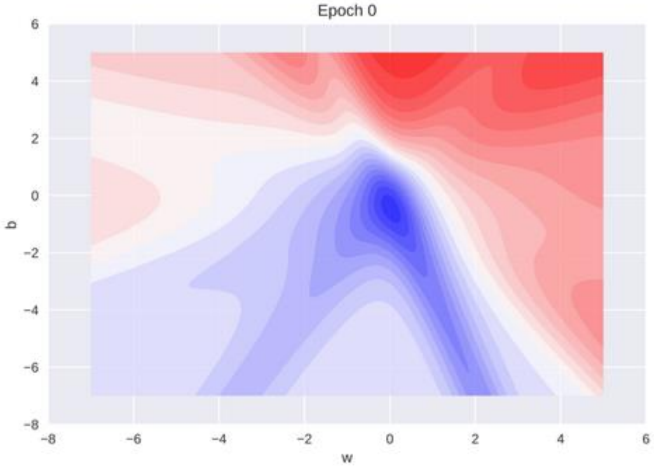
General concepts



Introduction

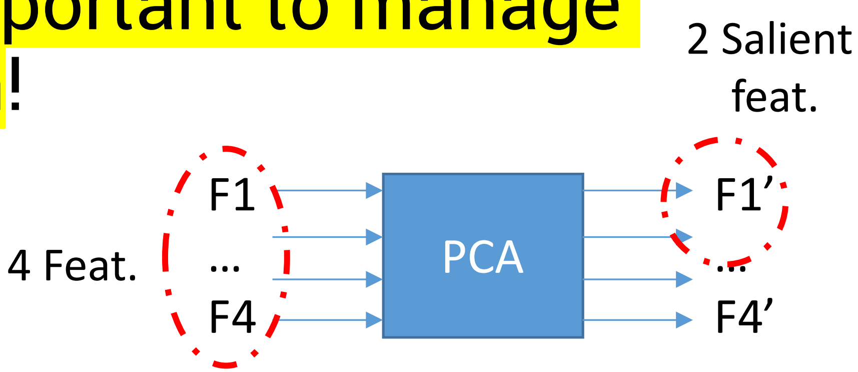


- What is Data Visualization?
- How does Data Visualization Work?
    - Image Theory
    - Definition of and image representation
    - Data Visualization and its use today
- What are the benefits of Data Visualization?
- Examples of Data Visualization



Why (almost) 2D and 3D plots?

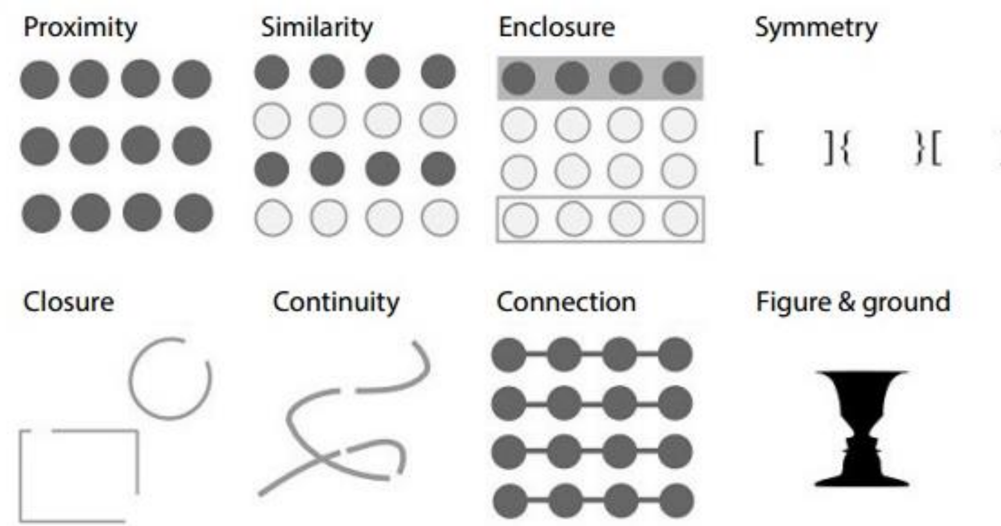
- We live in the *Space-Time*
  - a mathematical model fusing the three dimensions of space and the one dimension of time into a single four-dimensional continuum….
- Ok, but our perception is limited…
  - Only 2D and (rarely) 3D plots can be easely evaluated and understood
- That is also why is so important to manage dimensionality reduction!



What is Data Visualization?

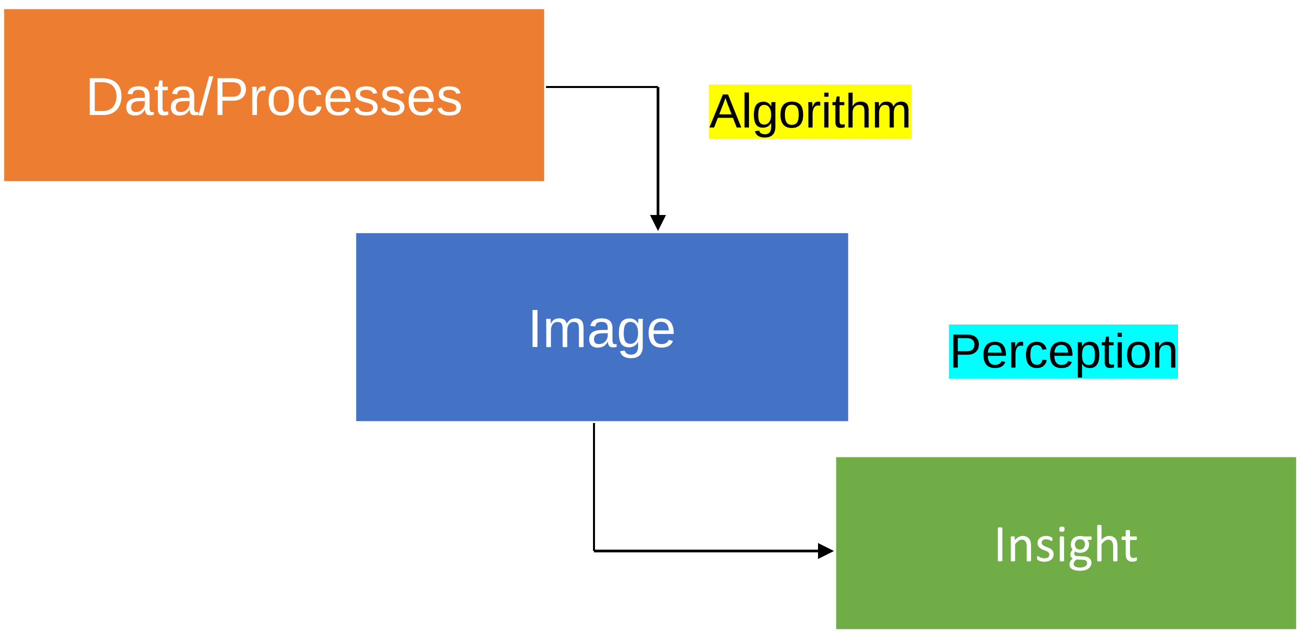
- Data visualization is the process of converting raw data into easily understood pictures of information that enable fast and effective decisions.

- Early in the 20th-century, gestalt psychologists observed that when elements were gathered into a figure, the figure took on a perceptual salience that exceeded the sum of its parts;



Data -> Easily Understood Pictures

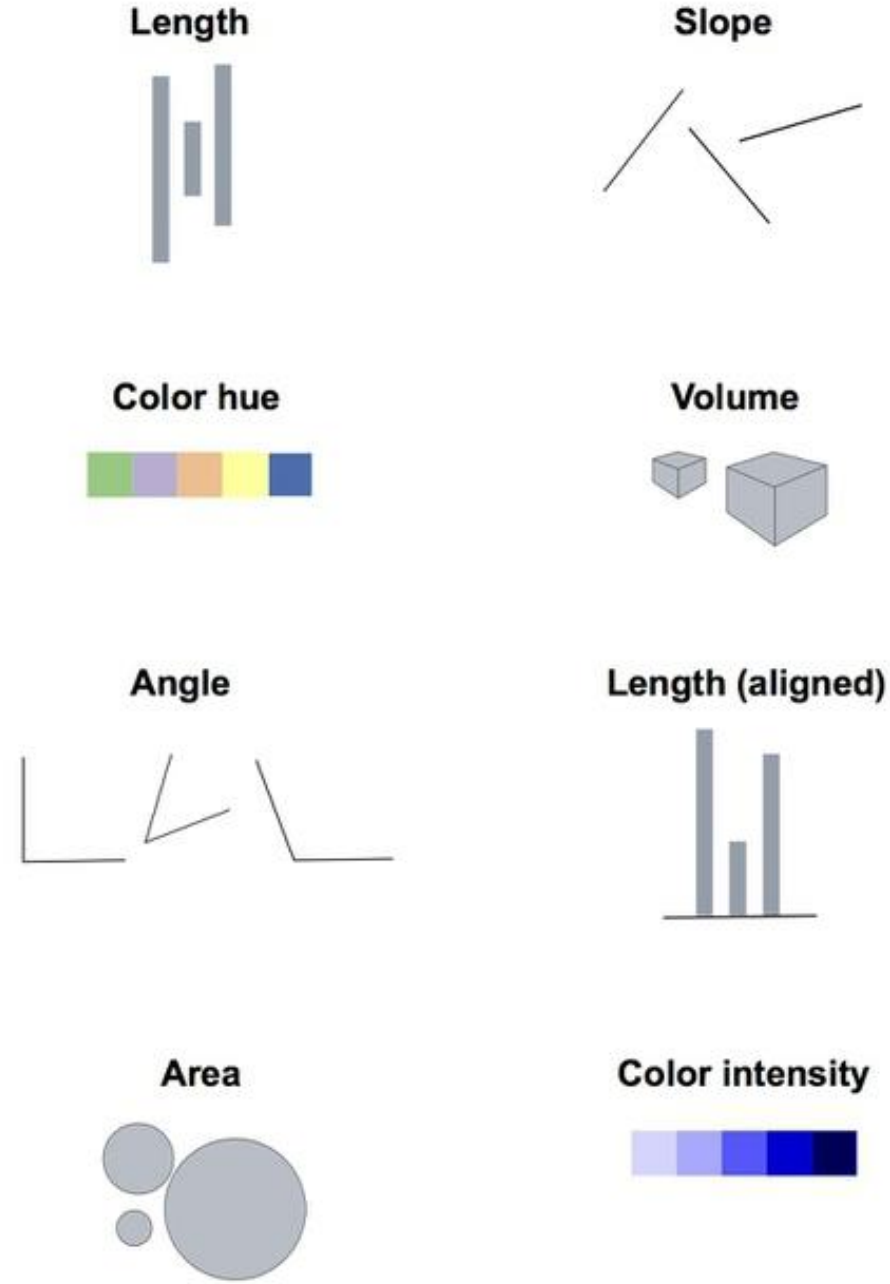
- Jacques Bertin who wrote the classic works of graphical visualization “Semiology of Graphics” states that the “transformation from numbers to insight requires two stages.”



Bertin’s 7 Visual Variables

- 7 Visual Variables
  - position
  - form
  - orientation
  - color
  - texture
  - value
  - size
- combined with a visual semantics for linking data attributes to visual elements

Visualization: encoding data using visual cues



Not all the visualization features are of the same kind…

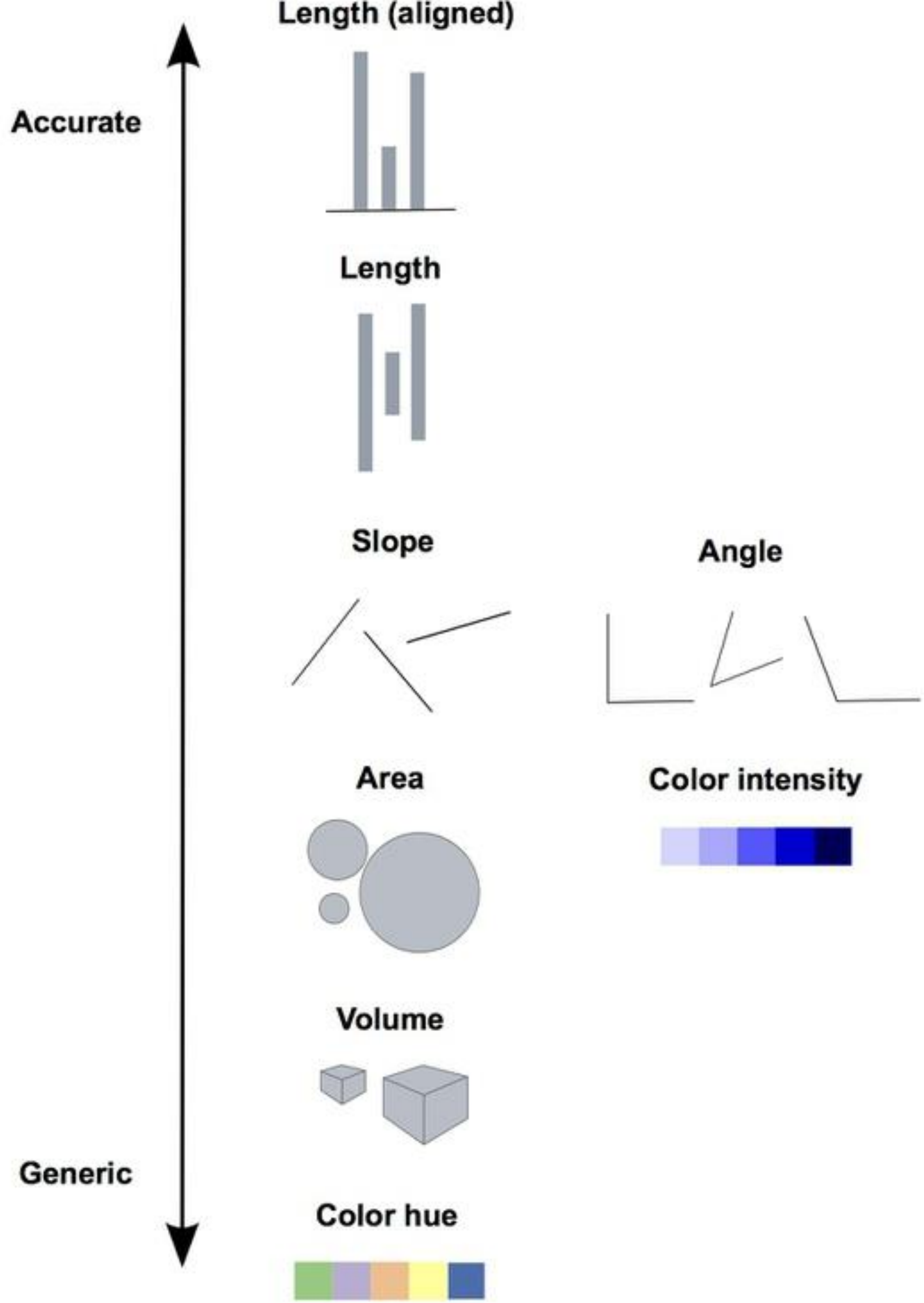
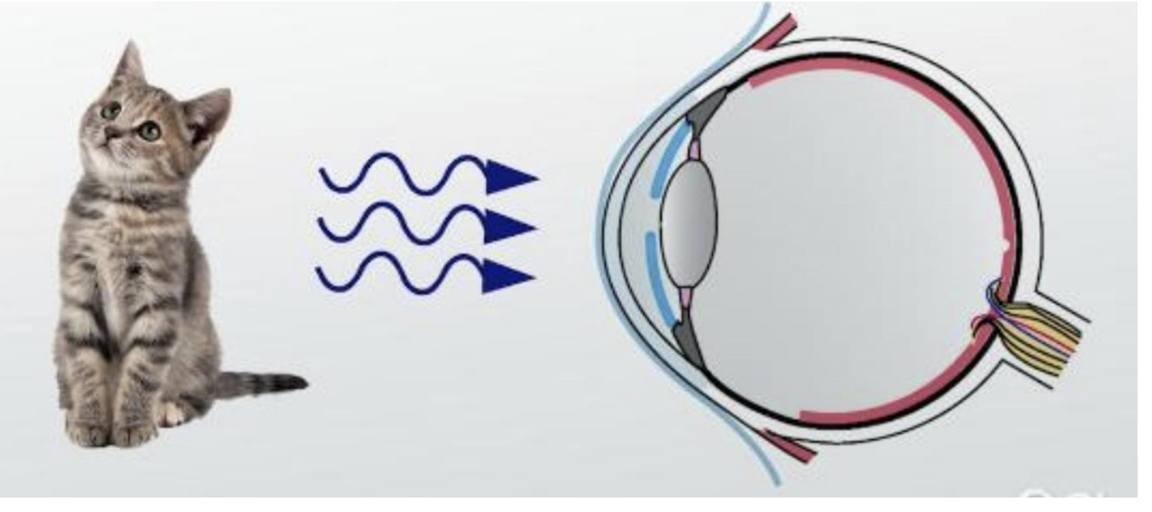
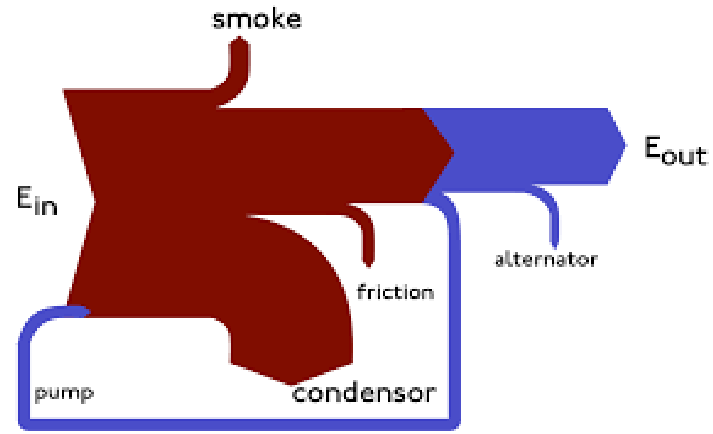


Image Theory

- Visual Processing occurs in 3 steps.
  1. Formation of the retinal image
  2. Decomposition of the retinal image information into an array of specialized representations
  3. Reassembly of the information into object perception.



Image

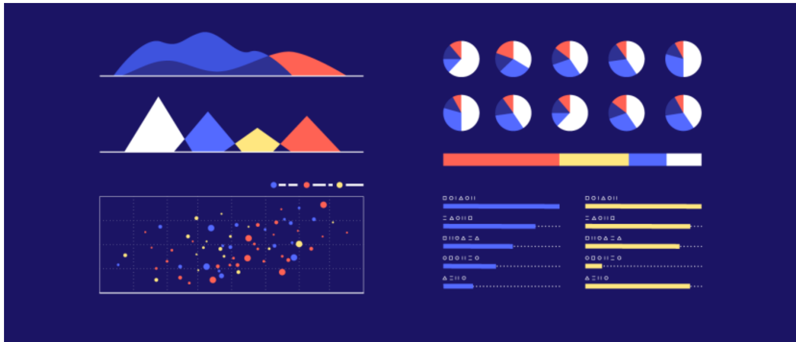


Where are you loosing energy?

- Bertin's key concept is the *image*, from which the theory derives its name.
- Simplifying…. an *image* is the fundamental perceptual unit of a visualization.
  - An ideal visualizations will contain only a single *image* in order to optimize "efficiency" the speed with which observer can extract the information

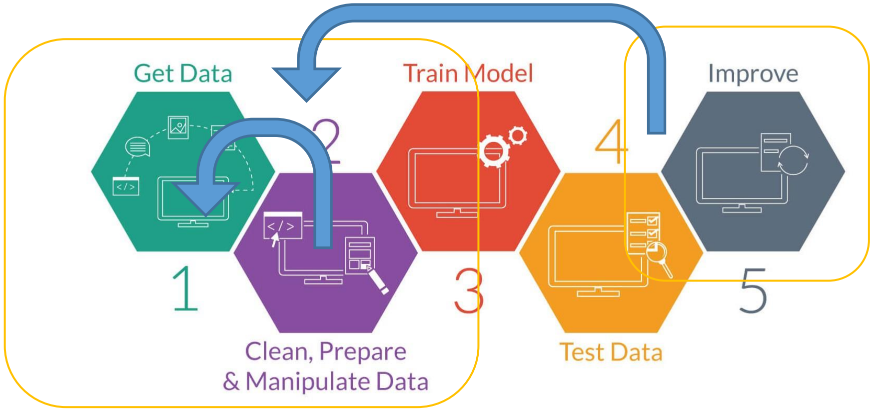
A perfect fusion of methods for your applications

- Information visualization is emerging as an important fusion of
  - graphics,
  - scientific visualization,
  - database management,
  - human-computer interaction.
- In your applications use Data Visualization to convey complex results as understandable images, improving your HMI



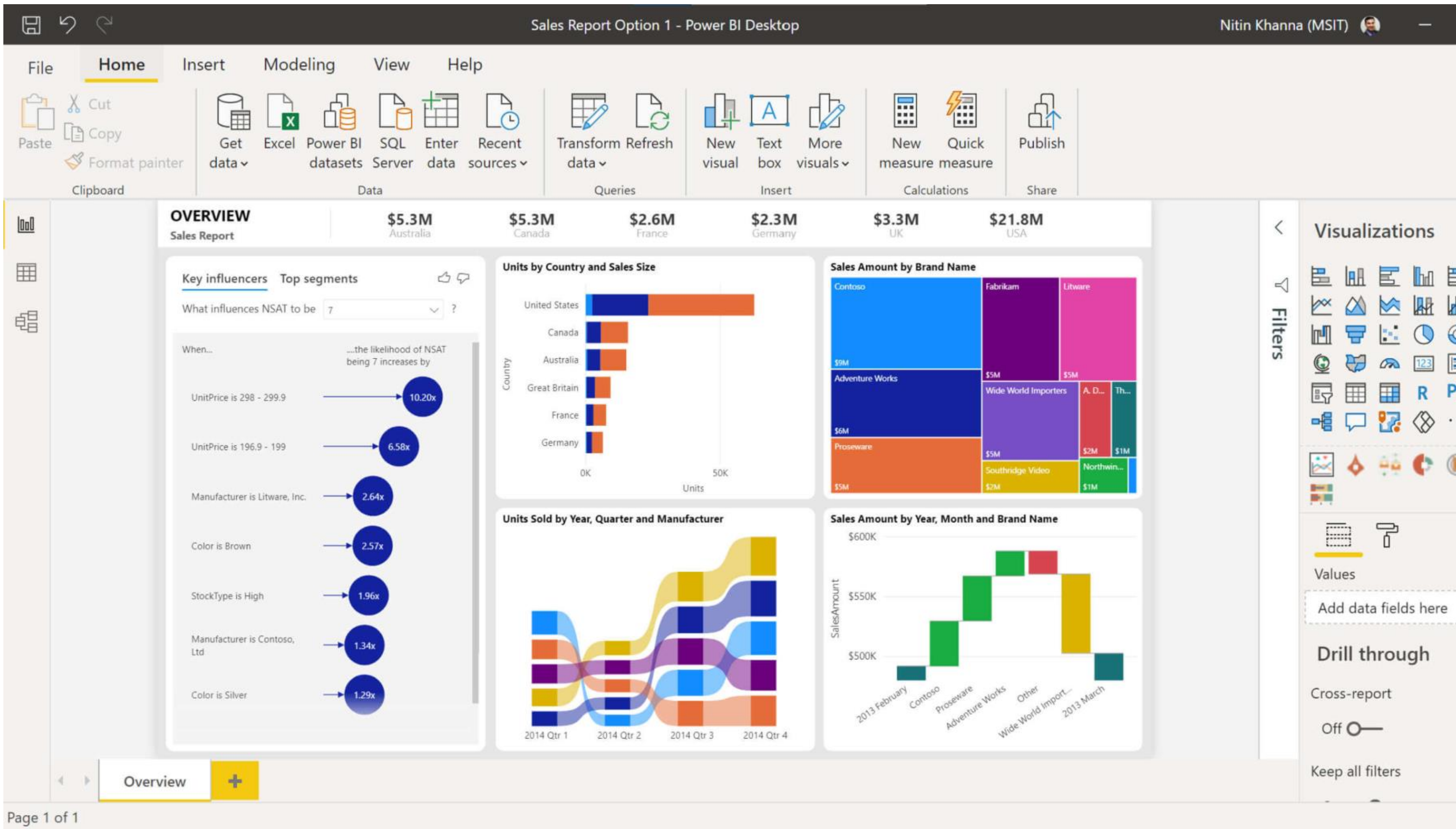
Important for users and the designers of intelligent systems

- Data visualization is used in software applications to provide an intuitive graphical interface.
- It is applied to many areas to enable users to glean useful information from their data for faster, more informed decision making.
- Users but also you, the designer!

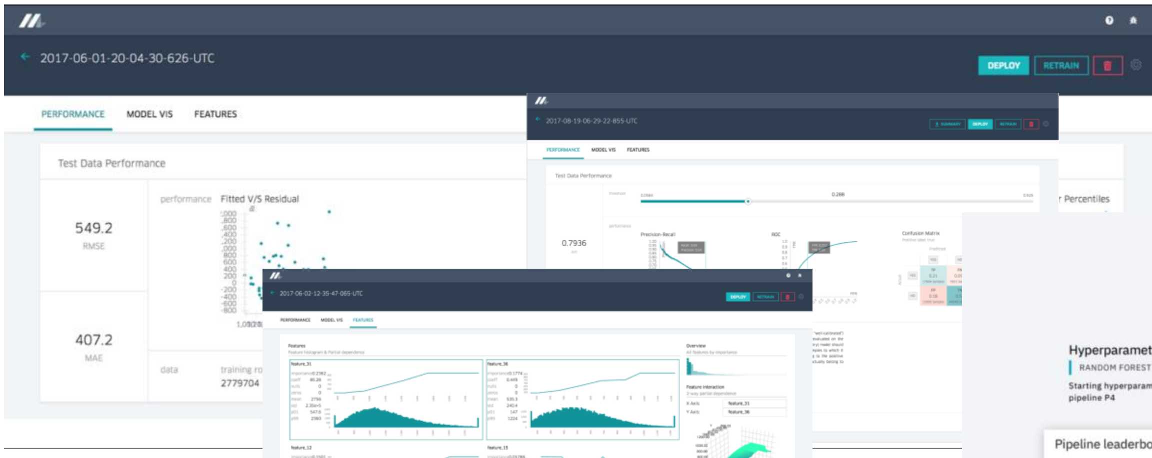


Example: MS Power Bi

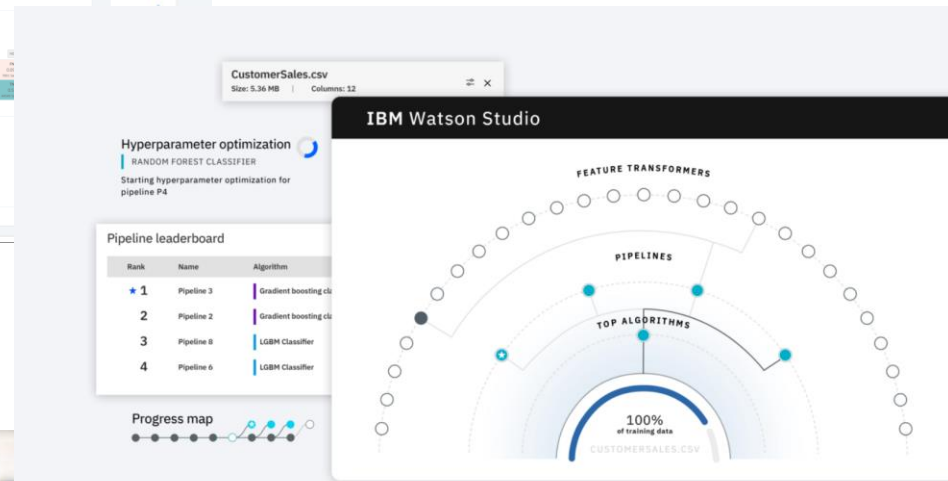
"Connect to and visualize any data using the unified, scalable platform for self-service and enterprise business intelligence (BI) that's easy to use and helps you gain deeper data insight."



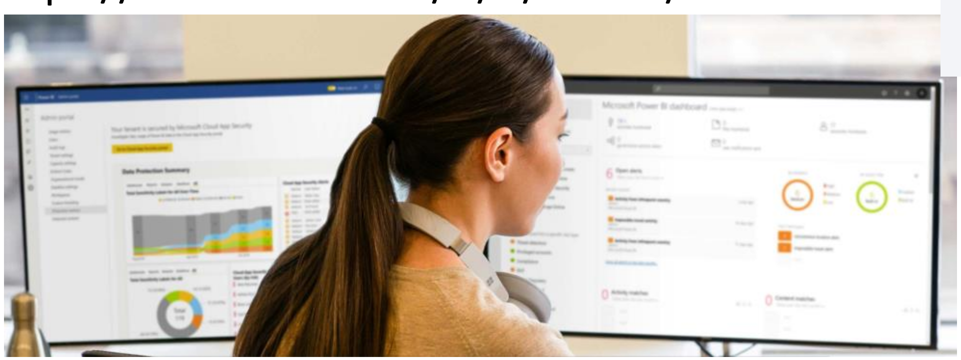
Example: training of complex models in cluster



Uber Michelangelo
ML-as-a-service platform
https://www.uber.com/it/it/uberai/

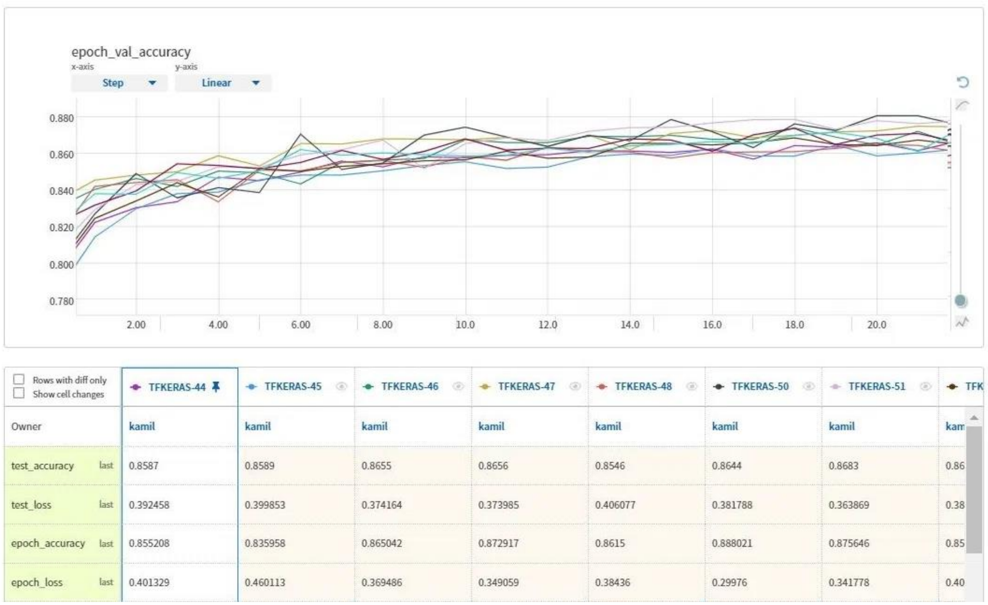


IBM Watson Studio
https://www.ibm.com/cloud/watson-studio



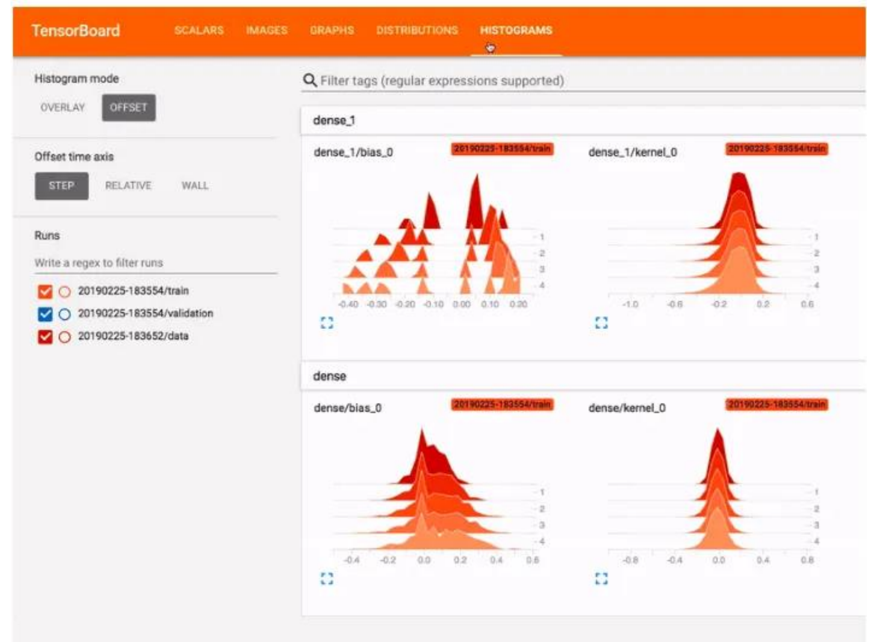
Example of tools to monitor ML experiments

Neptune



Experiment tracking and model registry
Fast UI with to organize runs in groups, save custom dashboard views and share, works with scripts (Python, R, other), notebooks (local, Google Colab, AWS SageMaker) on different infrastructure (cloud, laptop, cluster)

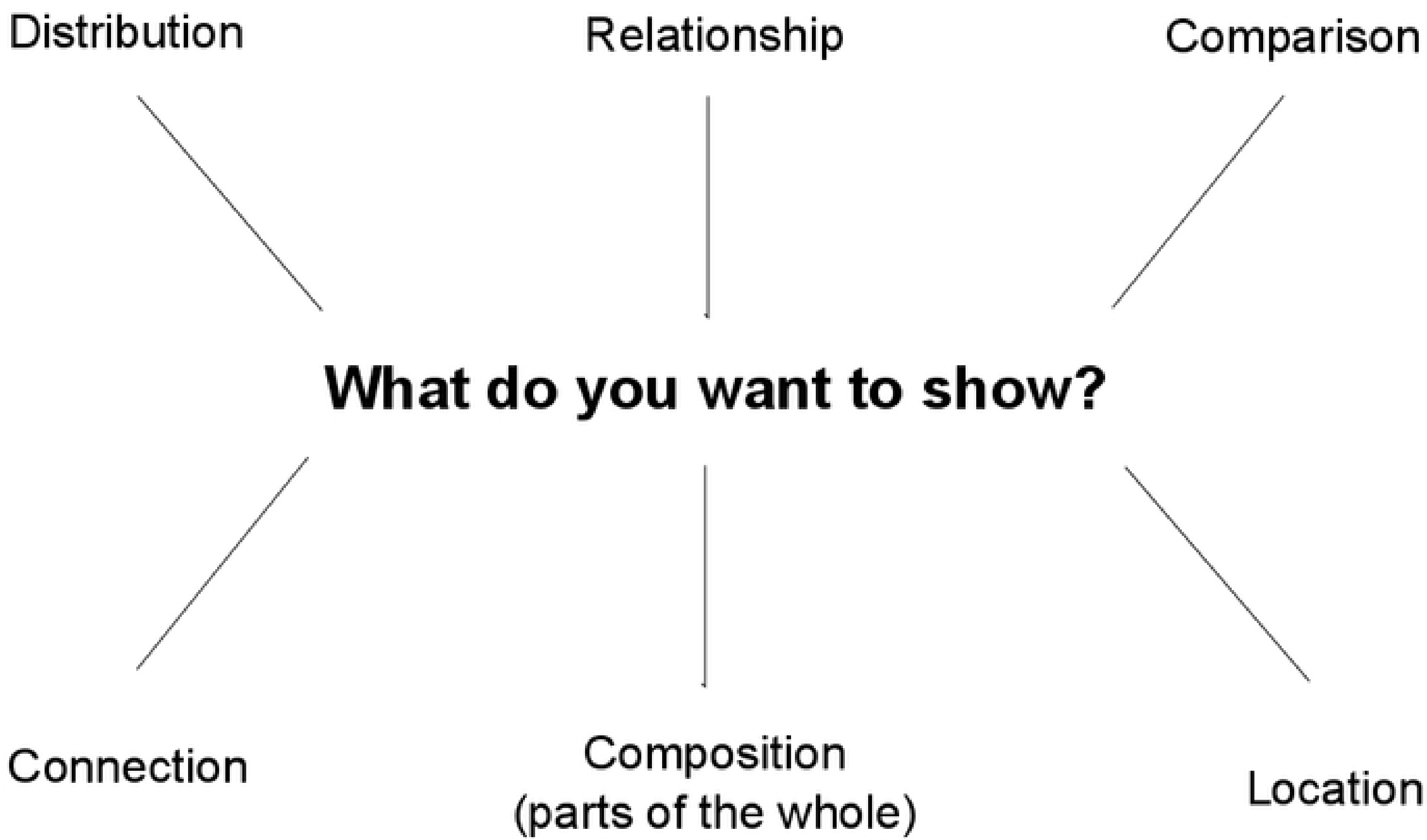
TensorBoard



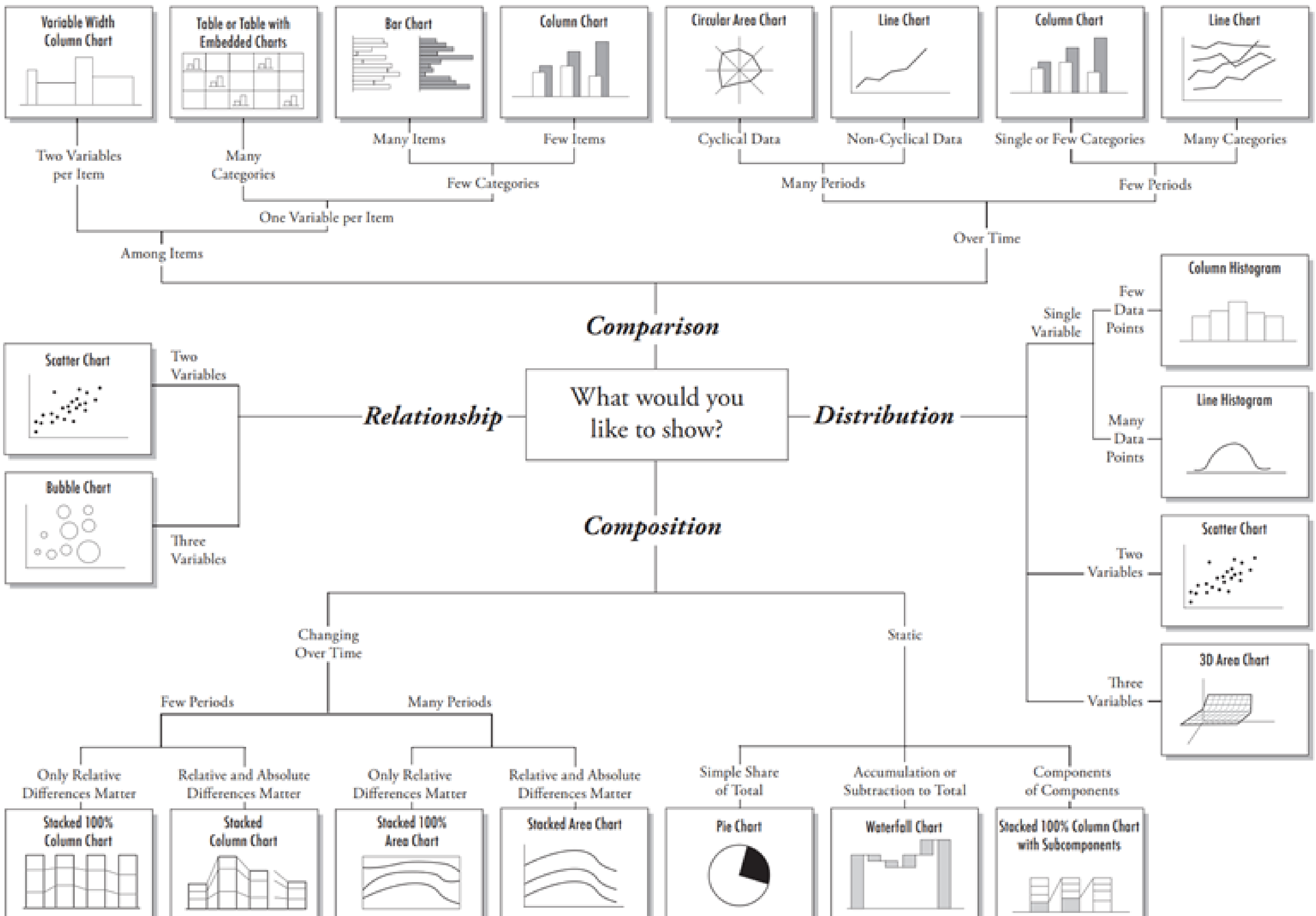
TensorBoard is a visualization toolkit for TensorFlow that lets you analyze model training runs

https://neptune.ai/

What chart?


Distribution
Relationship
Comparison
What do you want to show?
Connection
Composition
(parts of the whole)
Location

What chart? (cont..)

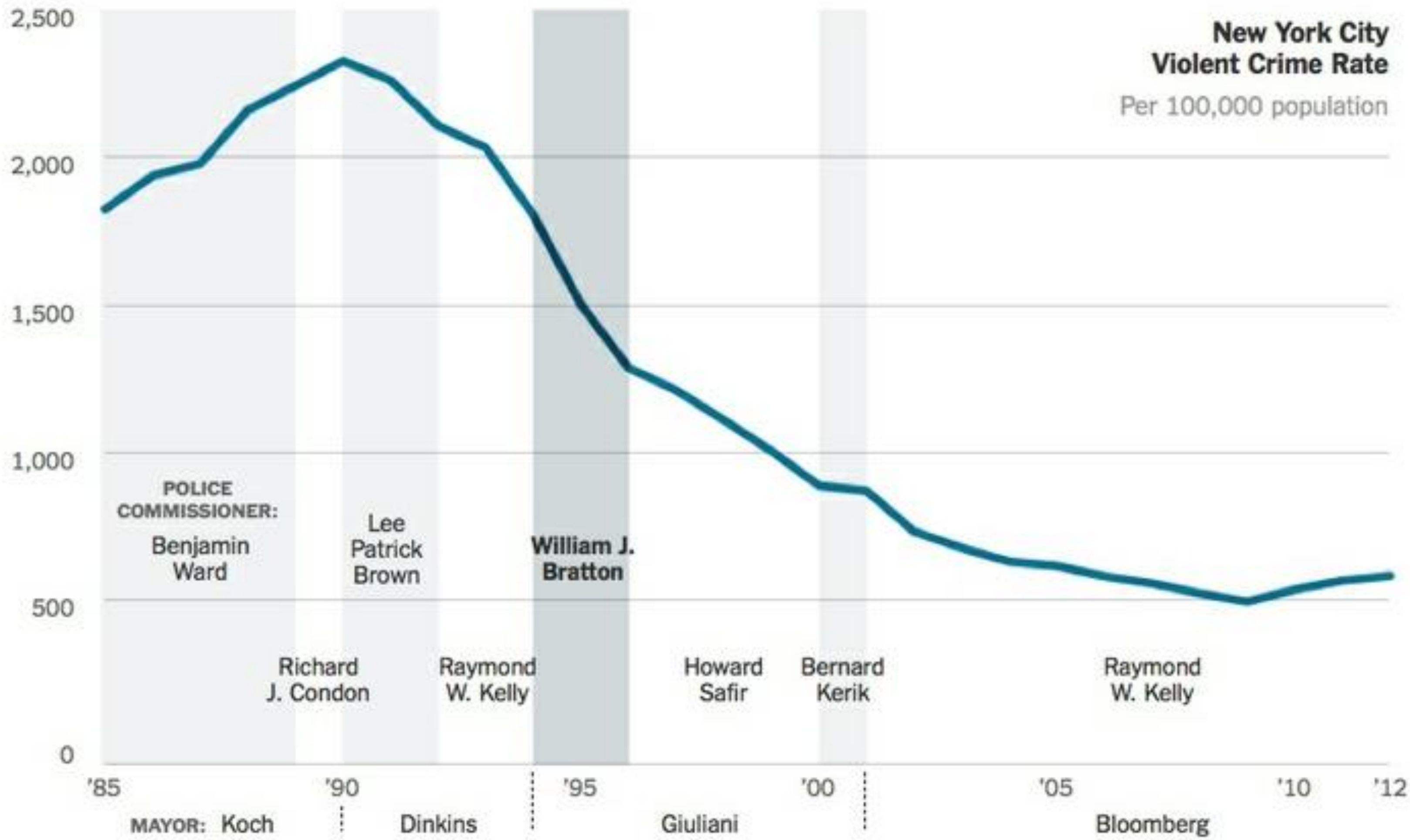


What are the benefits of Data Visualization?

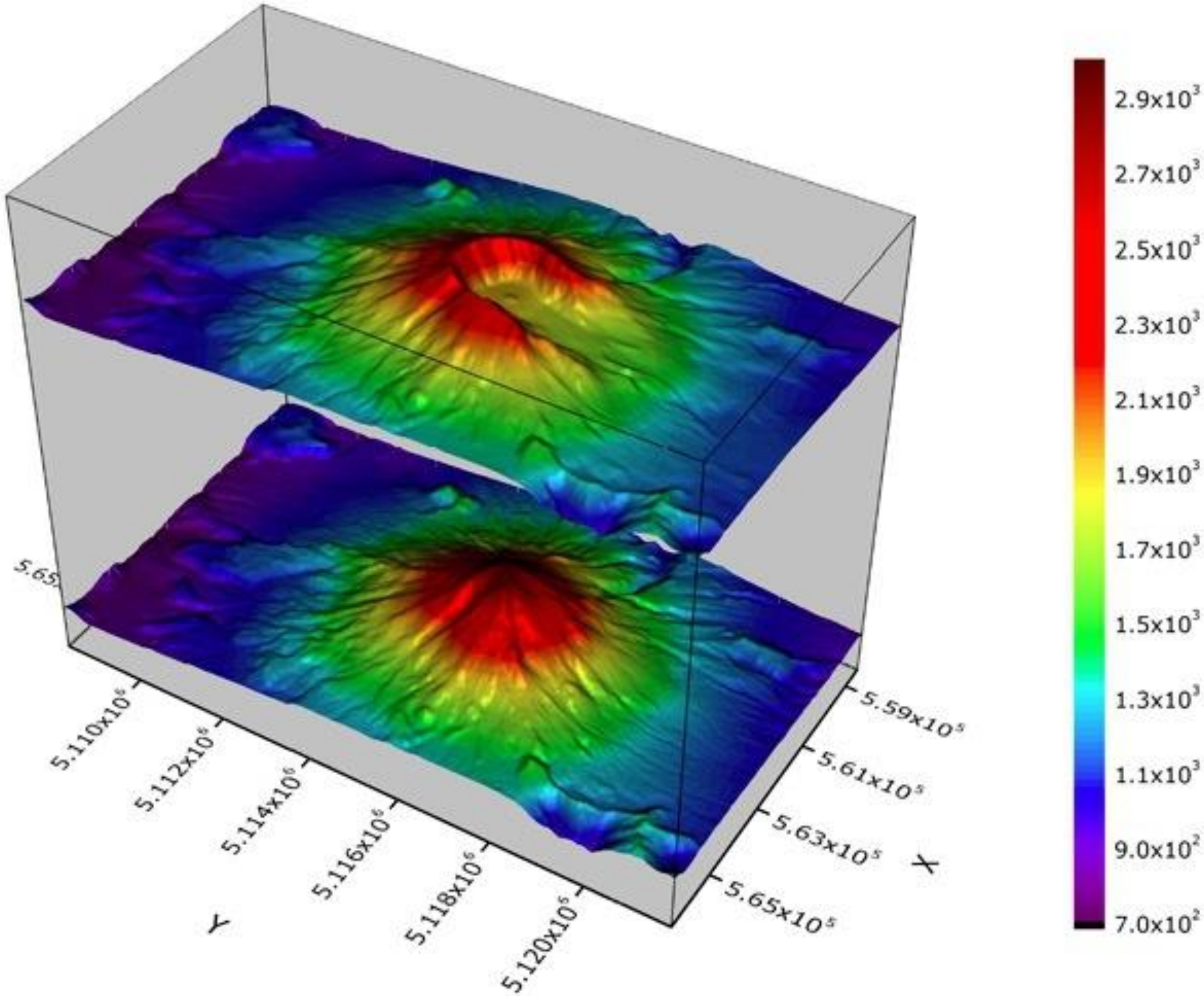
- Data visualization allows users see several different perspectives of the data.
- Data visualization makes it possible to interpret vast amounts of data
- Data visualization offers the ability to note exceptions in the data.
- Data visualization allows the user to analyze visual patterns in the data.
- Exploring trends within a database through visualization by letting analysts navigate through data and visually orient themselves to the patterns in the data.

A **picture** is worth a thousand words …

English language idiom



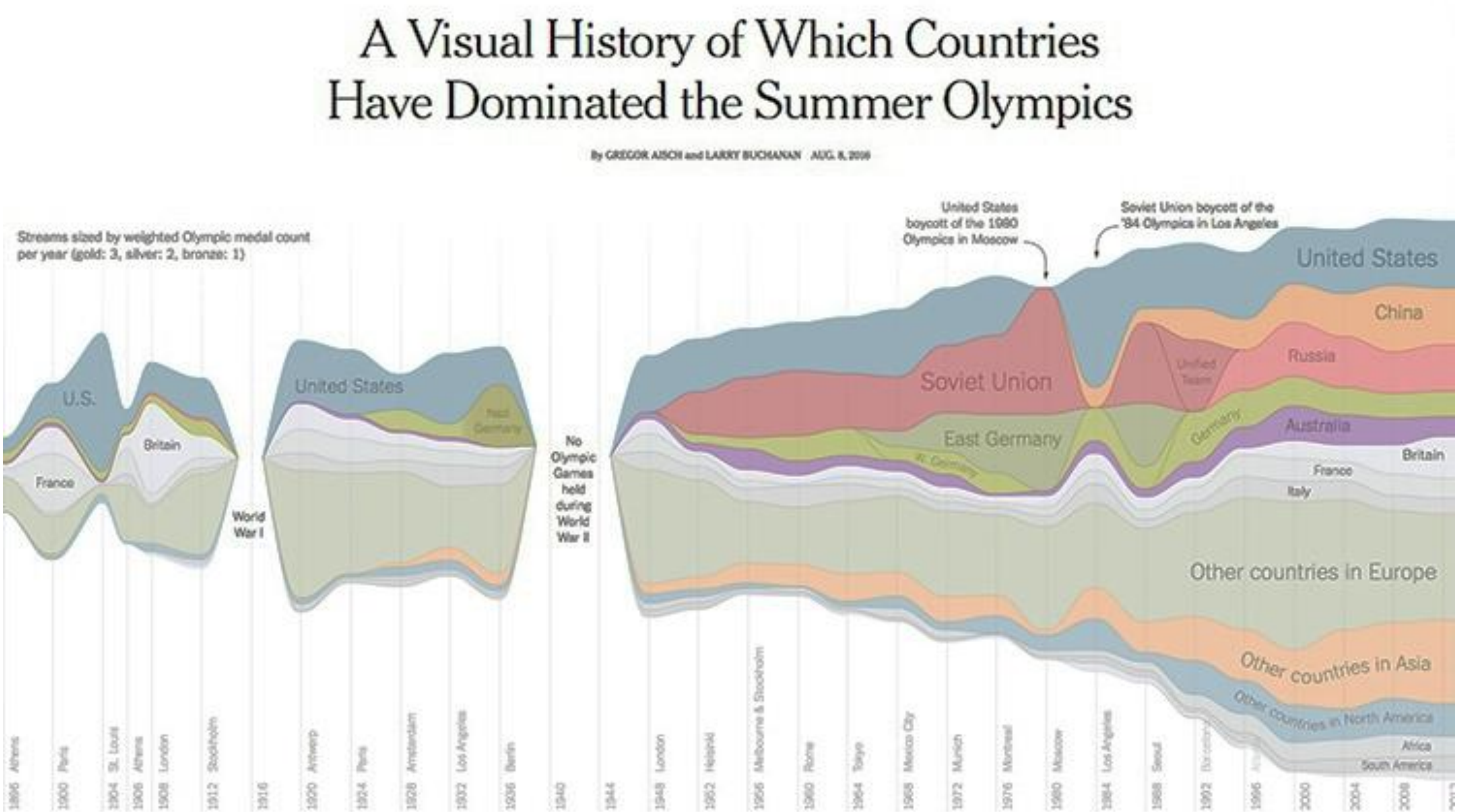
3D data in time



Representations of
Mount St. Helens
in Washington State,
before and after its 1980 eruption

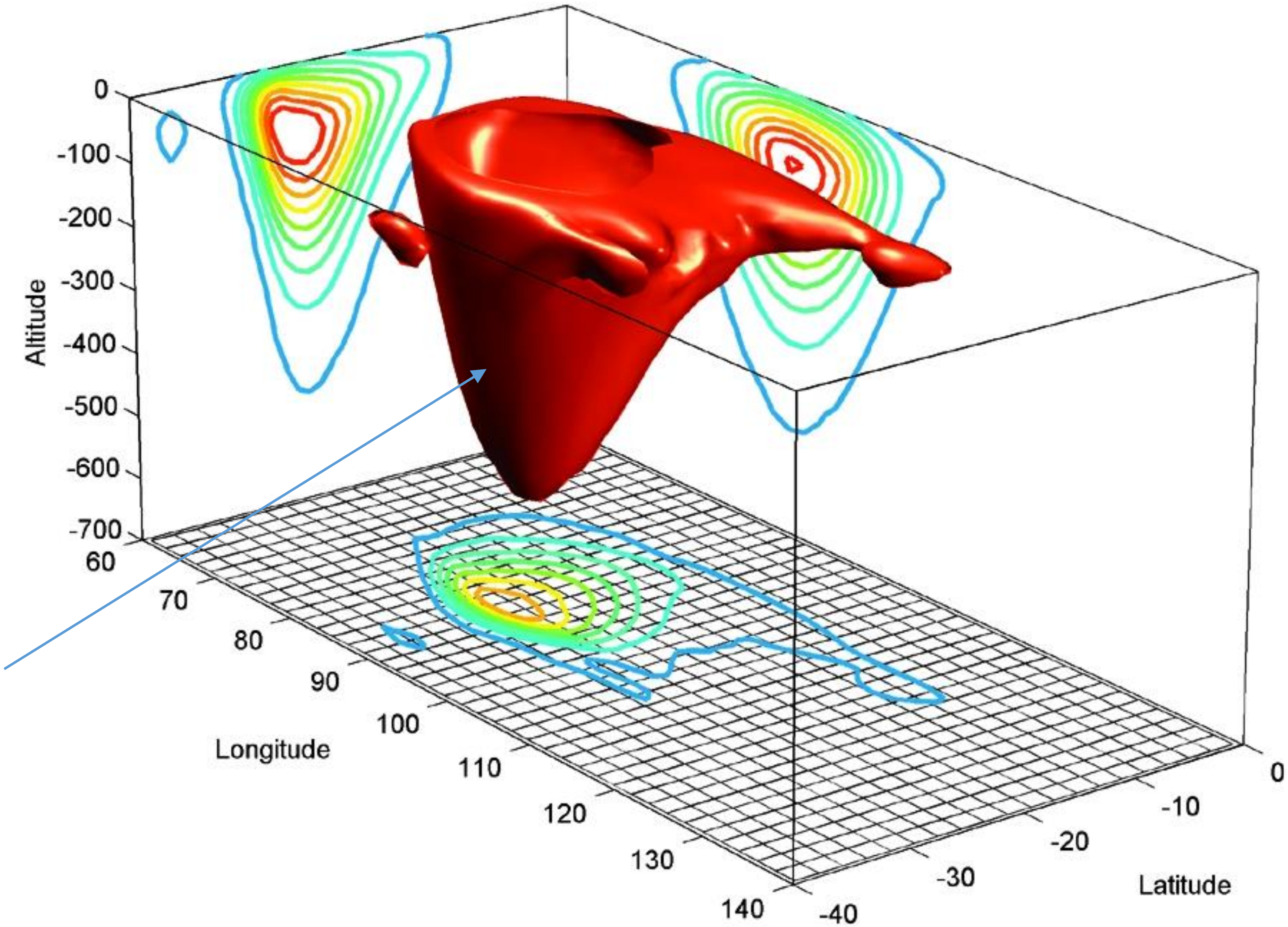
Distributions in time

No data, winners, events…



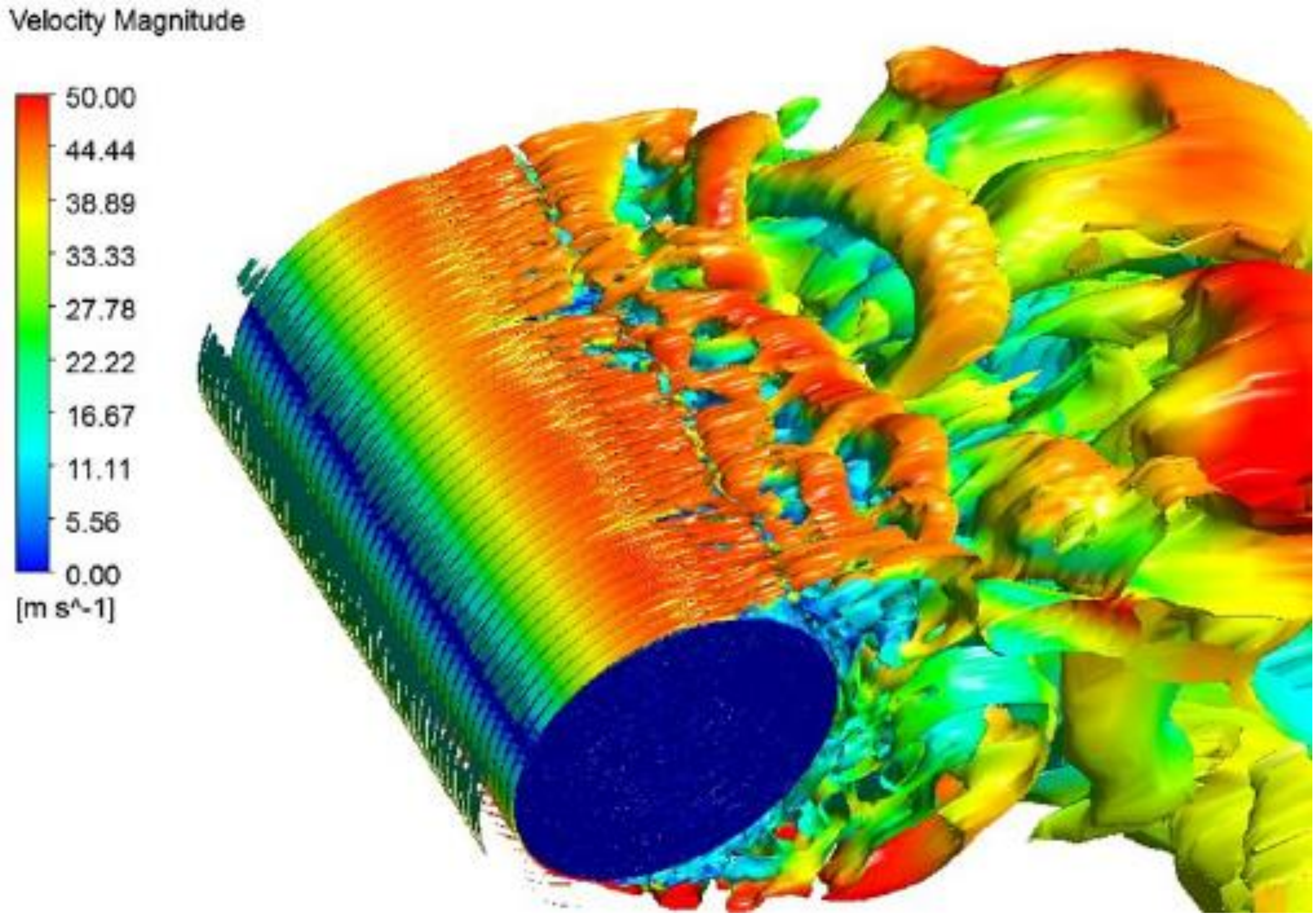
3D Surfaces

Intelligent systems decisions
can be represented
as surfaces in internal input spaces



isosurface

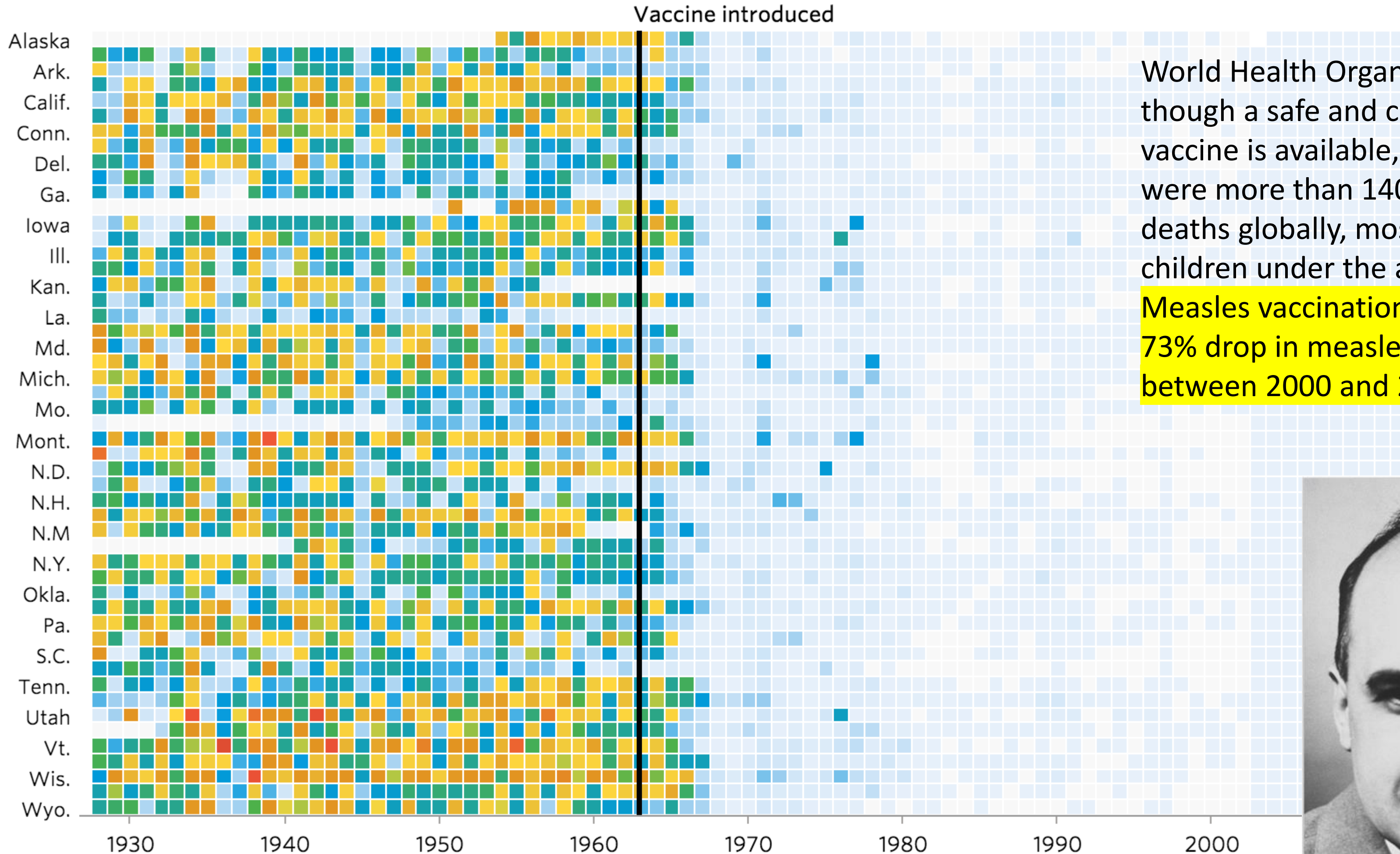
Complex phenomena…

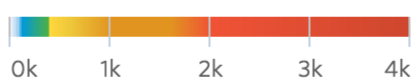


The heat maps below show number of cases per 100,000 people.

Measles (Morbilli)

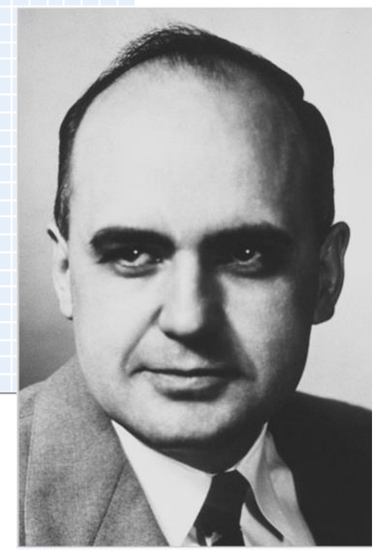
No vax?!





A percentage will die

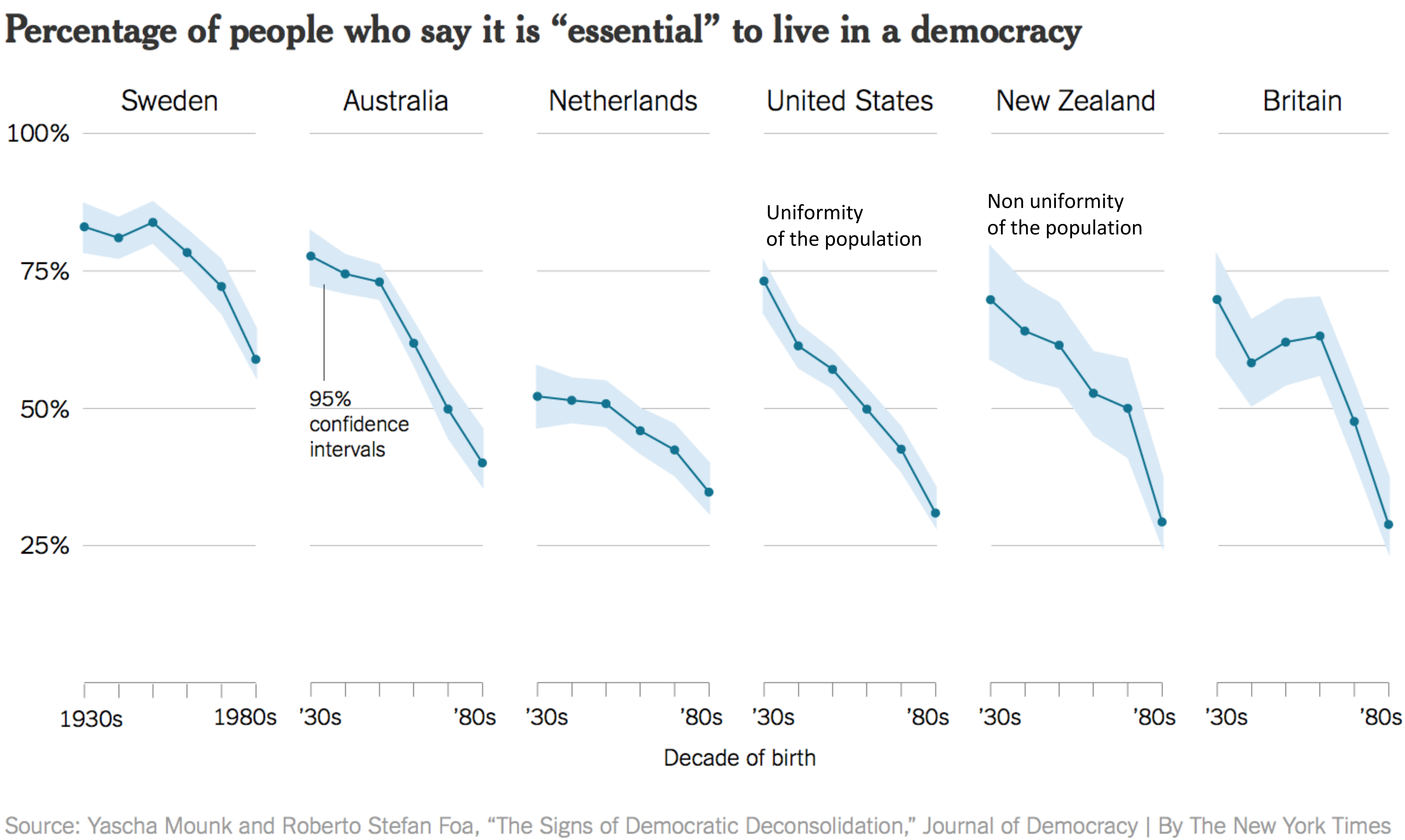
World Health Organization: Even though a safe and cost-effective vaccine is available, in 2018, there were more than 140 000 measles deaths globally, mostly among children under the age of five. Measles vaccination resulted in a 73% drop in measles deaths between 2000 and 2018 worldwide



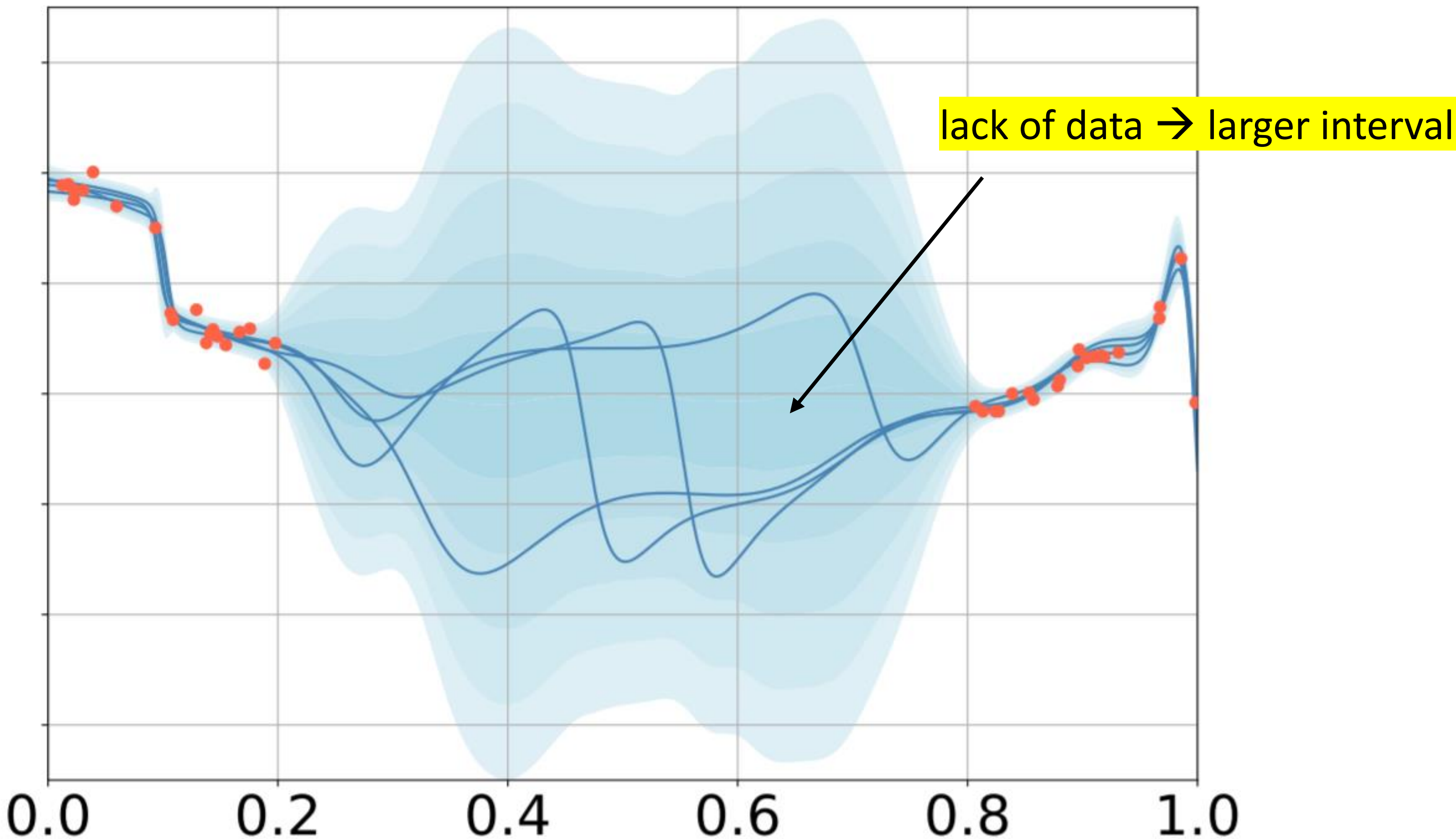
Maurice Hilleman's measles vaccine is estimated to prevent 1 million deaths per year.[117]

Note: CDC data from 2003-2012 comes from its Summary of Notifiable Diseases, which publishes yearly rather than weekly and counts confirmed cases as opposed to provisional ones.

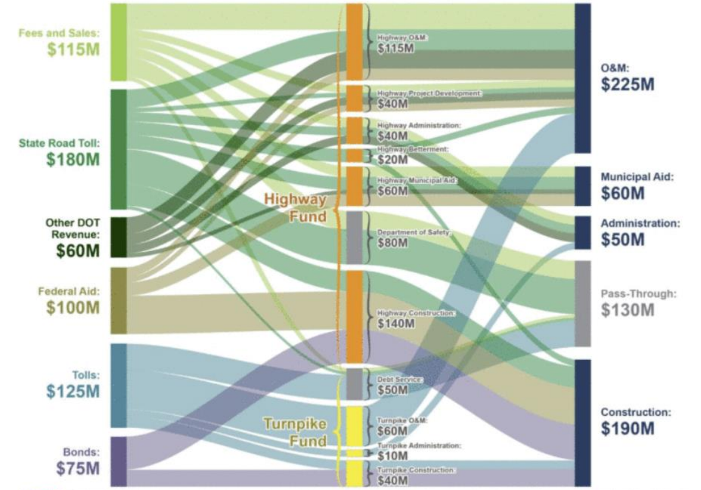
Confidence…


Percentage of people who say it is "essential" to live in a democracy
Sweden
Australia
Netherlands
United States
New Zealand
Britain
100%
75%
50%
25%
95% confidence intervals
Uniformity of the population
Non uniformity of the population
1930s
1980s
'30s
'80s
Decade of birth
Source: Yascha Mounk and Roberto Stefan Foa, "The Signs of Democratic Deconsolidation," Journal of Democracy | By The New York Times

Showing the confidence intervals of trained NN



Benefits (2)



- Data visualization can help translate data patterns into insights, making it a highly effective decision-making tool.
  - reduce the time and difficulty it takes to move from data to decision.

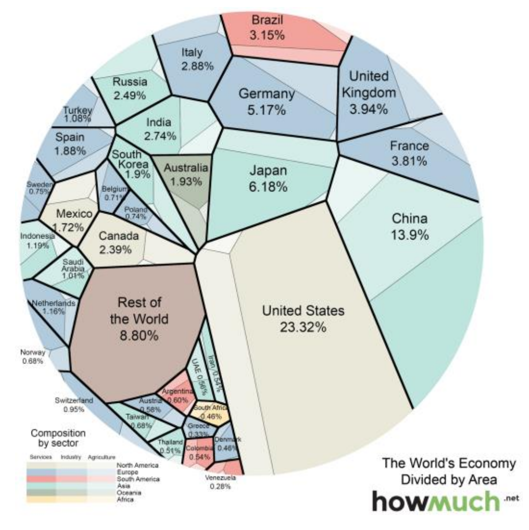
- Data visualization equips users with the ability to see influences (outputs VS inputs) that would otherwise be difficult to find.

- With all the data available, it is difficult to find the nuances that can make a difference.

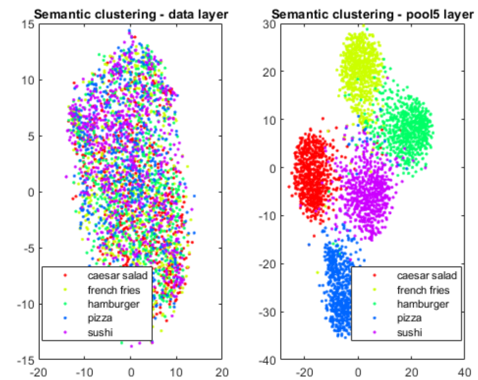
Conclusions: Graphical displays should…

Too much graphic design…

- Show the data
- Induce the viewer to think about the substance rather than about methodology, graphic design, the technology of graphic production, or something else
- Avoid distorting what the data have to say
- Present many numbers in a small space
- Make large data sets coherent
    - Homogenization/Data harmonization.



Conclusions: Graphical displays should (2)…



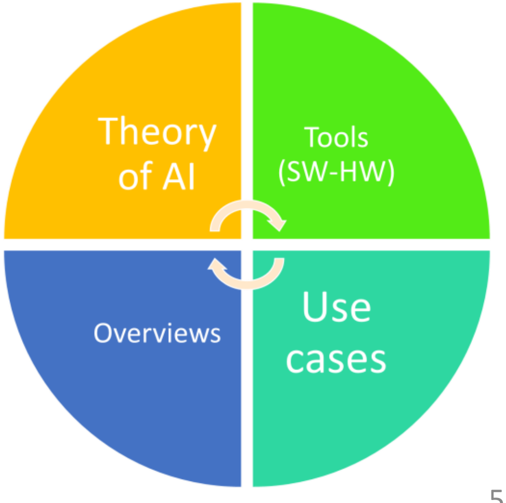
- Encourage the eye to compare different pieces of data
- Reveal the data at several levels of detail, from a broad overview to the fine structure
- Serve a reasonably clear purpose:
  - description,
  - exploration,
  - tabulation,
  - or decoration
- Be closely integrated with the statistical and verbal descriptions of a data set




Toolboxes

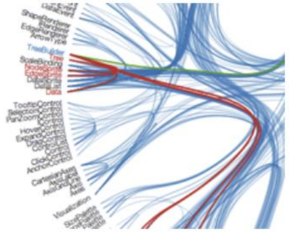
Data visualization

How to check/plot/ look at the data

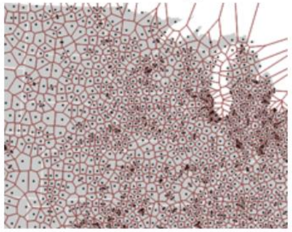


On-line tools
Example: observablehq.com

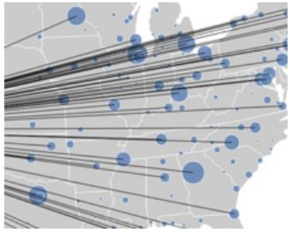
Hierarchical Edge Bundling



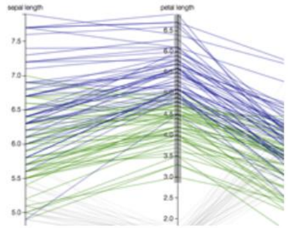
Voronoi Diagram



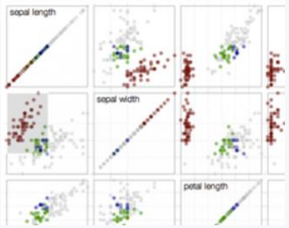
Bubble Map



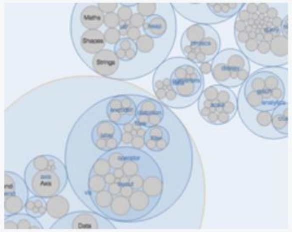
Parallel Coordinates



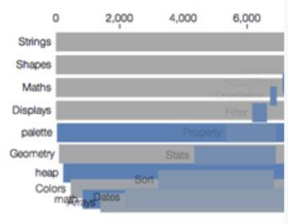
Scatterplot Matrix



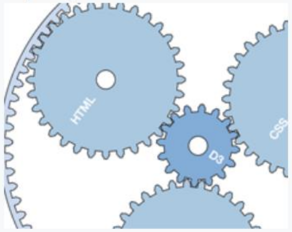
Zoomable Pack Layout



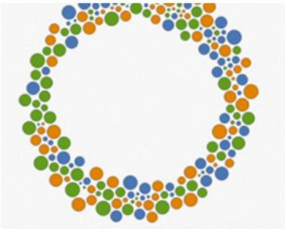
Hierarchical Bars



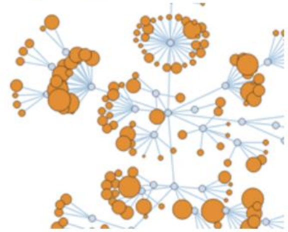
Epicyclical Gears



Collision Detection



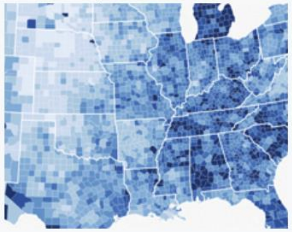
Collapsible Force Layout



Force-Directed States

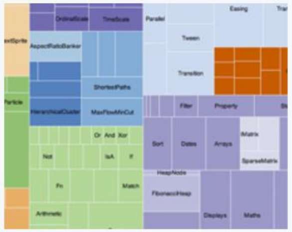
Versor Dragging

Choropleth



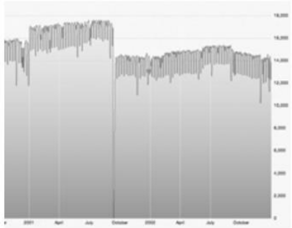
Collapsible Tree Layout

Zoomable Treemap

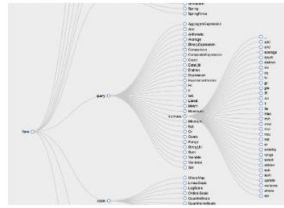


Zoomable Icicle

Zoomable Area Chart

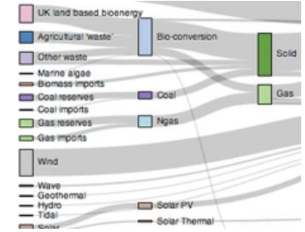


Drag and Drop Collapsible Tree Layout

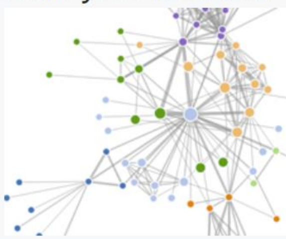


Radial Cluster Layout

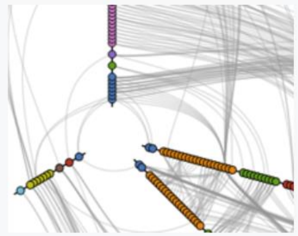
Sankey Diagram



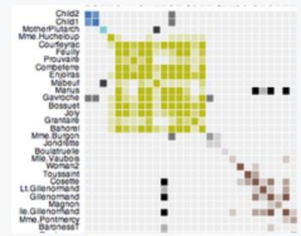
Fisheye Distortion



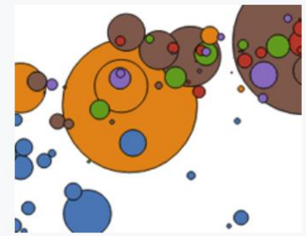
Hive Plot



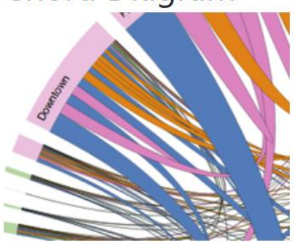
Co-occurrence Matrix



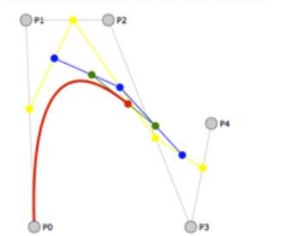
Motion Chart



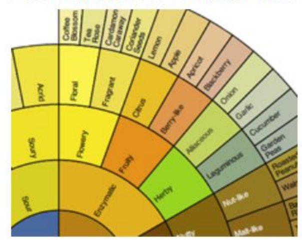
Chord Diagram



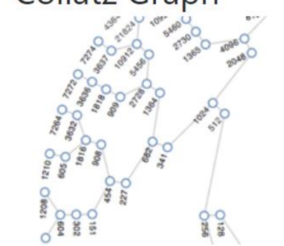
Animated Béziers



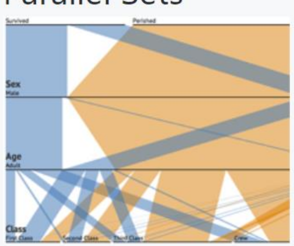
Zoomable Sunburst



Collatz Graph

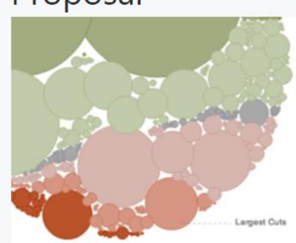


Parallel Sets

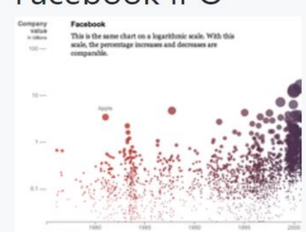


Word Cloud

Obama's Budget Proposal



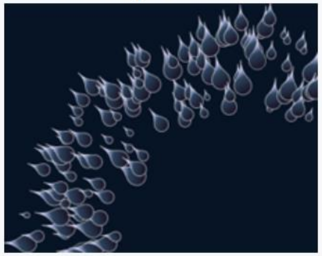
Facebook IPO



Voronoi Picking

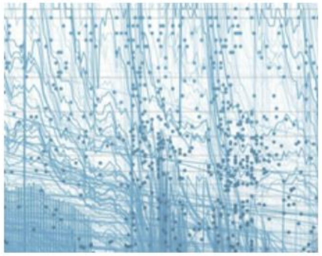
Zoomable Map

Raindrops

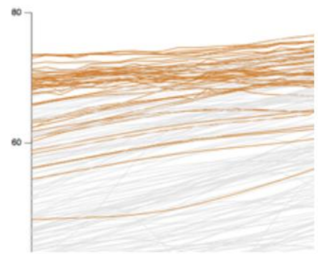


Color

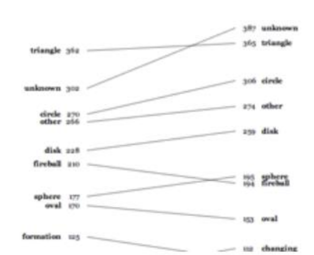
Hacker News Popularity



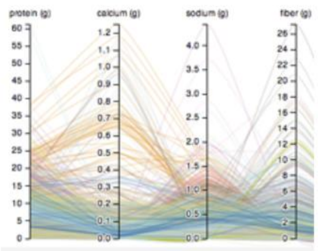
Life Expectancy



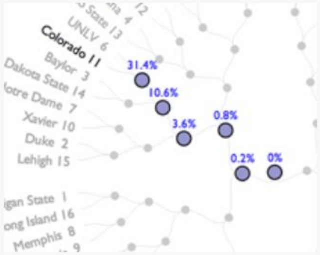
Slopegraphs



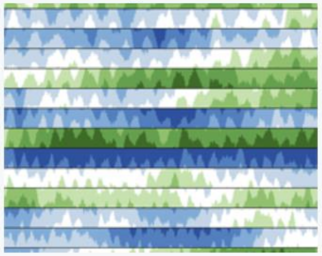
Parallel Coordinates



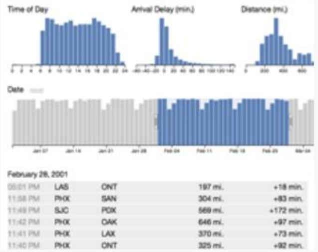
NCAA Predictions



Cubism.js

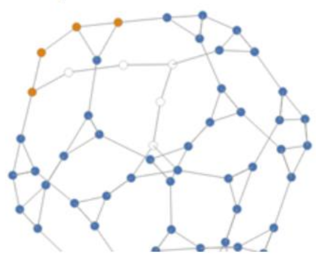


Crossfilter.js

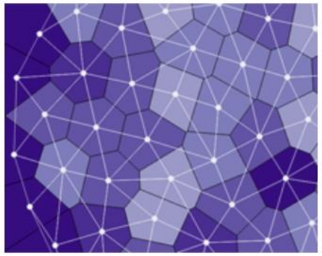


Wind History

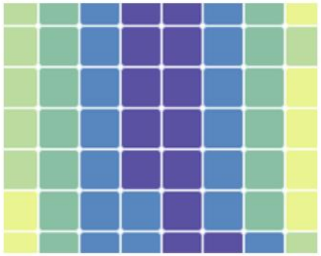
Cubic Hamiltonian Graphs



Force-Directed Voronoi



Trulia Trends



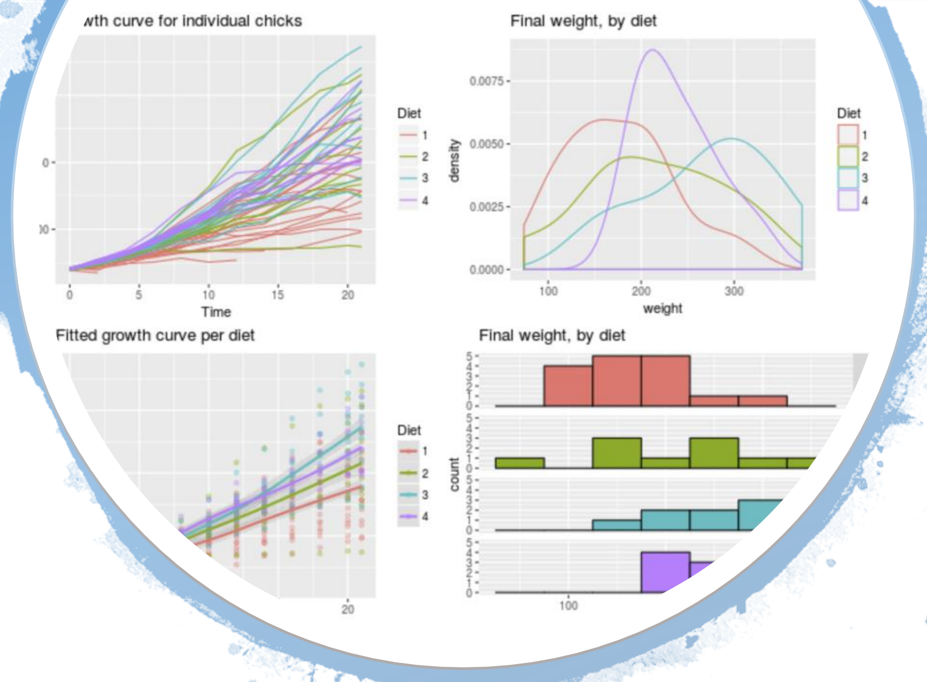
Trulia Trends

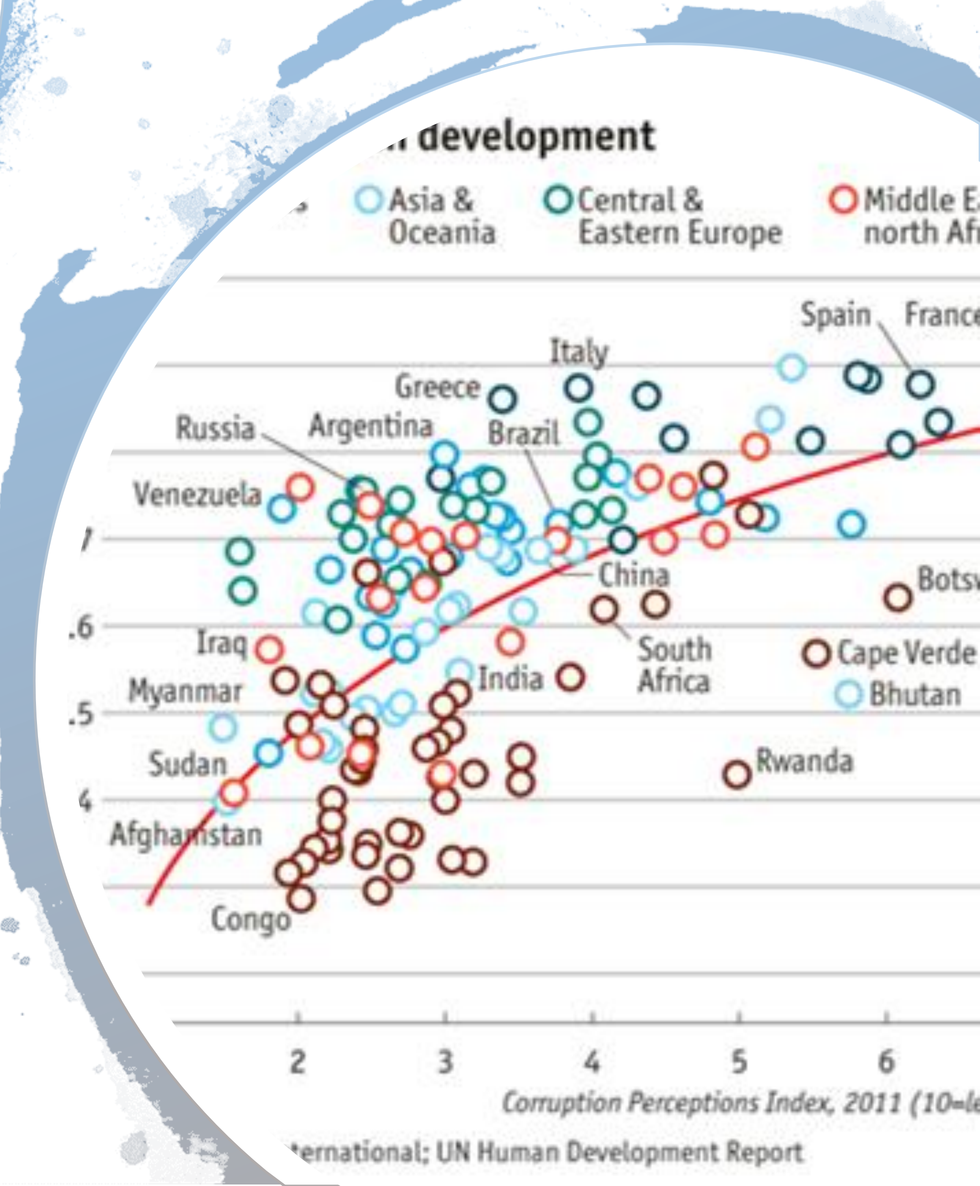
Fancy plots…

- The graphic power of the plots and charts is useles if….
  - data are not representative
  - data are corrupted
  - data are not properly preprocessed
  - data are FAKE!!
    (e.g., too much or bad augmentation)

R's ggplot2 library

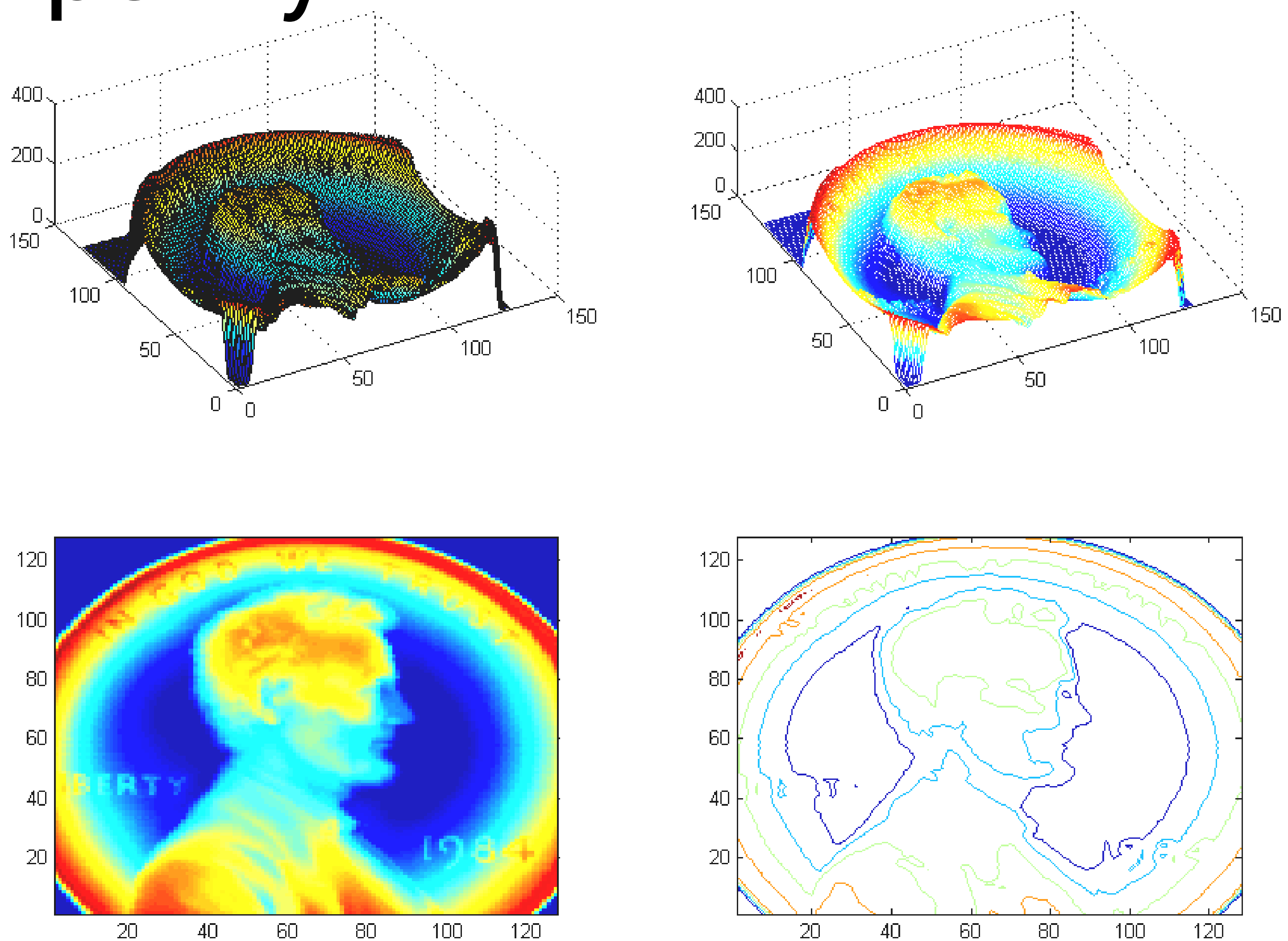
- R is a programming language and free software environment for statistical computing and graphics supported by the R Foundation for Statistical Computing.





R's ggplot2 library

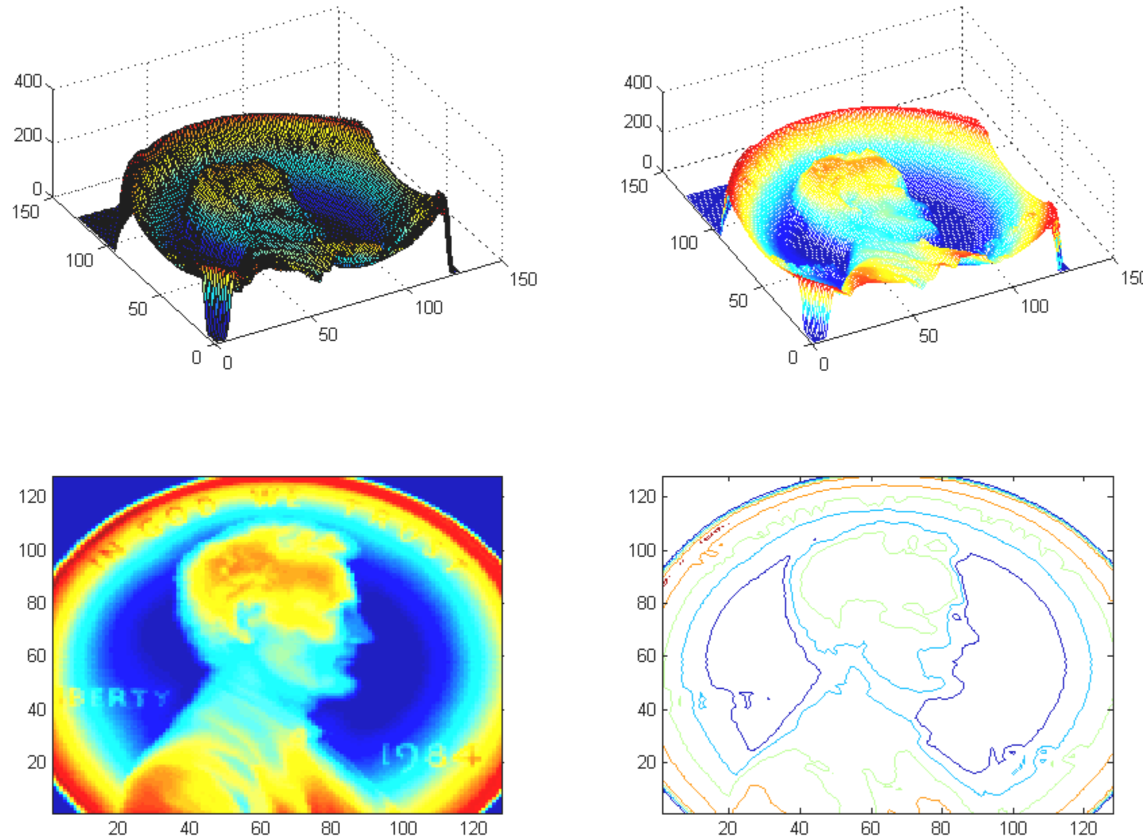
The scanned 3D penny



Some very simple examples with 2D data (matlab)

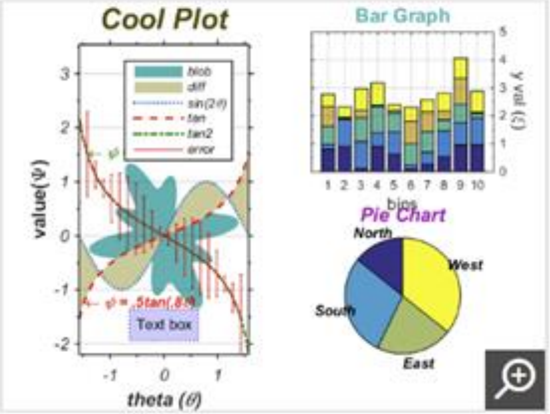
4 different plots of a penny

```
load penny  ;
P = flipud(P);
subplot(2,2,1)
   surf(P);

subplot(2,2,2)
  mesh(P);

subplot(2,2,3)
  pcolor(P);
  shading flat


subplot(2,2,4)
  contour(P);
```



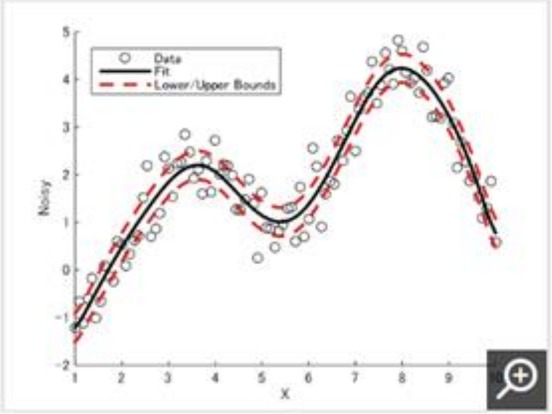
Annotations, colormaps bounds

Area Bar Pie Charts with Annotations



Colormap Chart

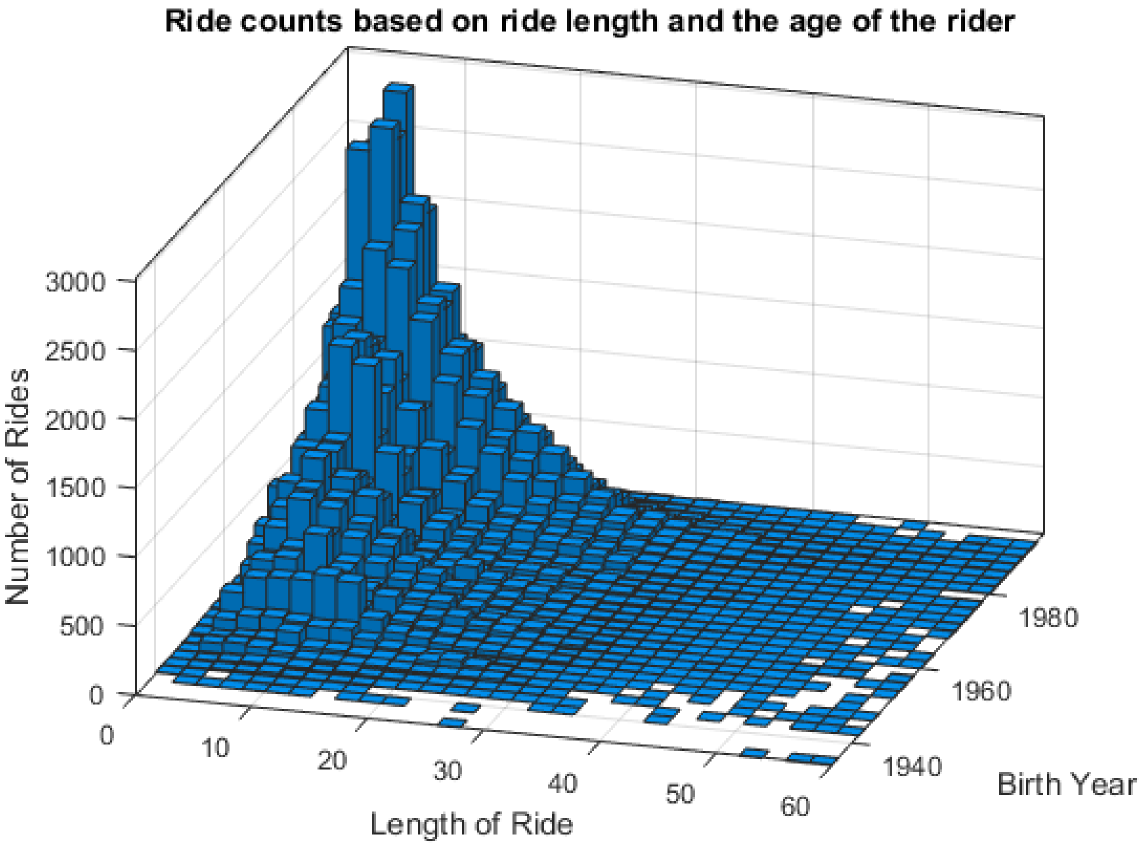
Curve with Lower and Upper Bounds

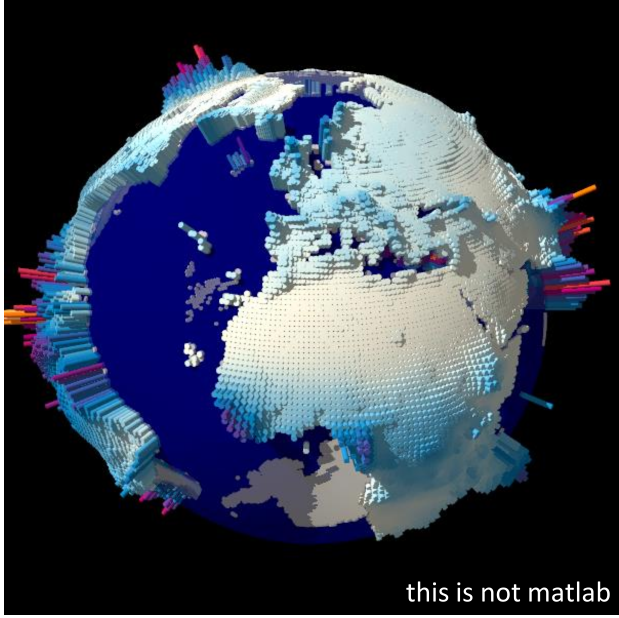


https://it.mathworks.com/products/matlab/plot-gallery.html

3D Barchart

(Data from Boston's bike sharing program)

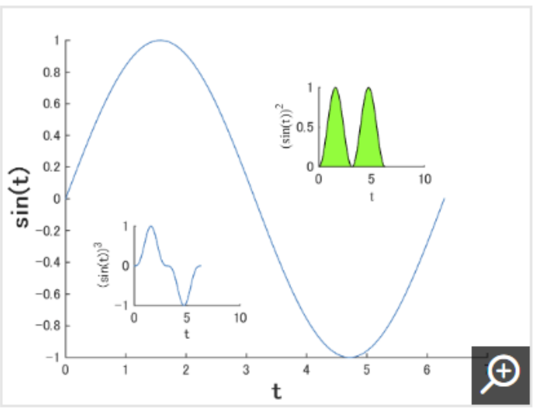





```
% Create bivariate histogram plot using the histogram2 function
histogram2(rideData.Duration, rideData.birth_date, 'BinWidth', [2 2])
xlabel('Length of Ride')
ylabel('Birth Year')
zlabel('Number of Rides')
title('Ride counts based on ride length and the age of the rider')
% Adjust view
view(17,30)
```

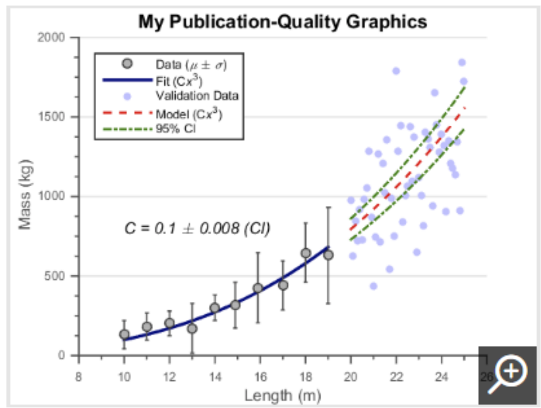
Plot in plot, 3D line vectors

Plot in Plot



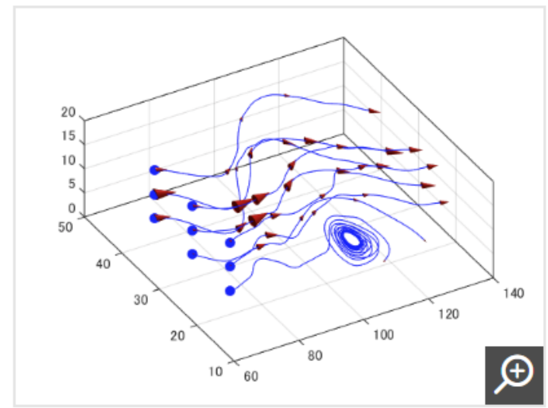
View source

Publication Quality Graphics



View source

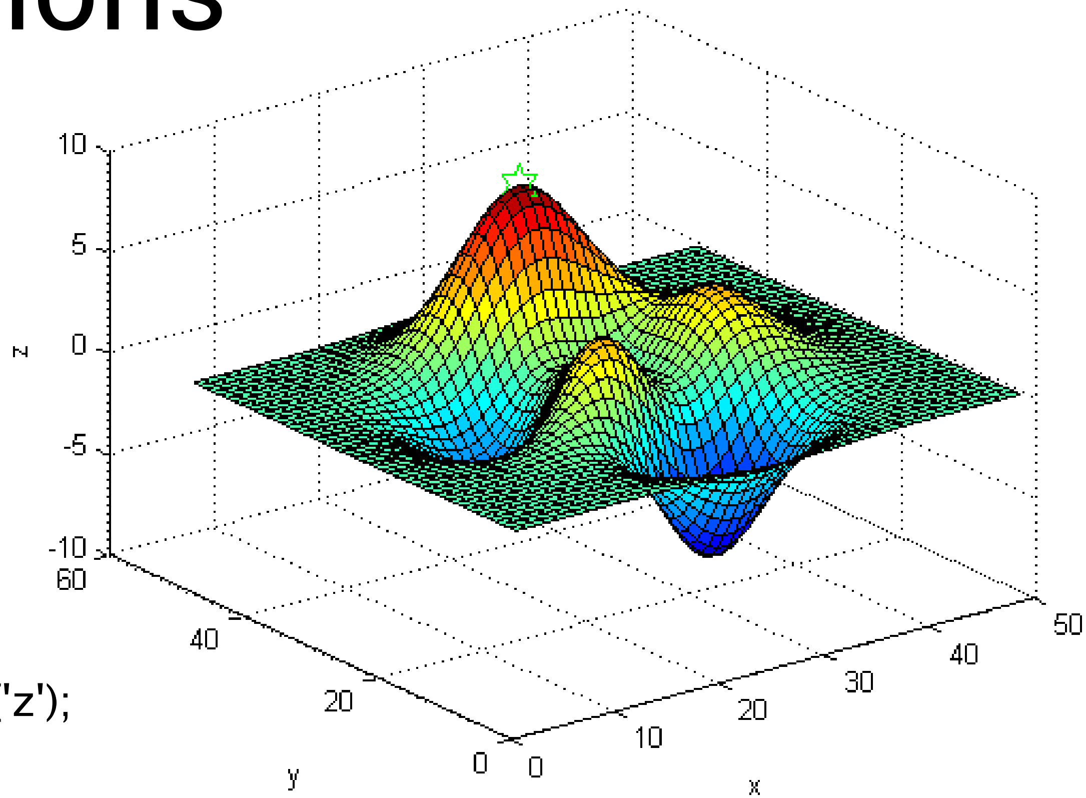
Streamline



View source

Find maximum in a 2D functions

Learning means OPTIMIZATION that means find a minimum or a maximum in a function



Homework:
consider the function «peaks»,
plot and locate the maximum value

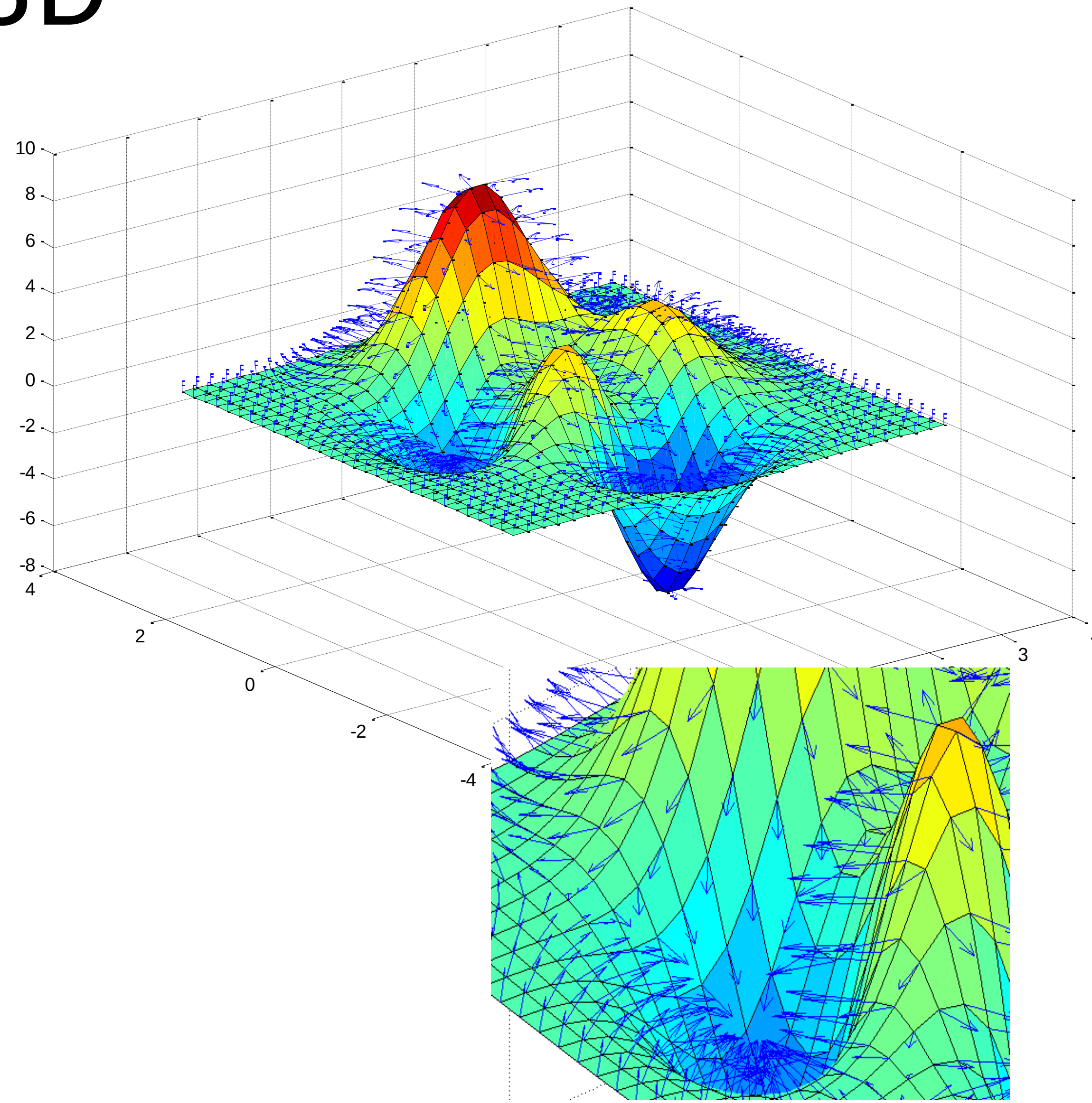
```
[Z] = peaks;
% abbiamo tutti i valori di X, Y, Z gia' precalcolati

surf(Z);    xlabel('x'); ylabel('y'); zlabel('z');

Zmax = max(max(Z))
[ymax, xmax] = find(Z == Zmax)

%
hold on;
plot3( xmax, ymax, Zmax, 'hg', 'Markersize', 15 );
hold off;
```

Vectors in 3D

Showing a complex property of a single point such as vectors

Plot the othogonal vectors of z = x .* exp(-x.^2 - y.^2)

```
[x,y] = meshgrid(-2:.2:2,-1:.15:1);
z = x .* exp(-x.^2 - y.^2);

[u,v,w] = surfnorm(x,y,z);
```




```
quiver3(x,y,z,u,v,w);

hold on,
surf(x,y,z),
hold off
```

Showing a property in 3D

Map with a proper pallette the gradient of z = x .* exp(-x.^2 - y.^2)

[x,y] = meshgrid([-2:.1:2]);

Z = x.*exp(-x.^2-y.^2);

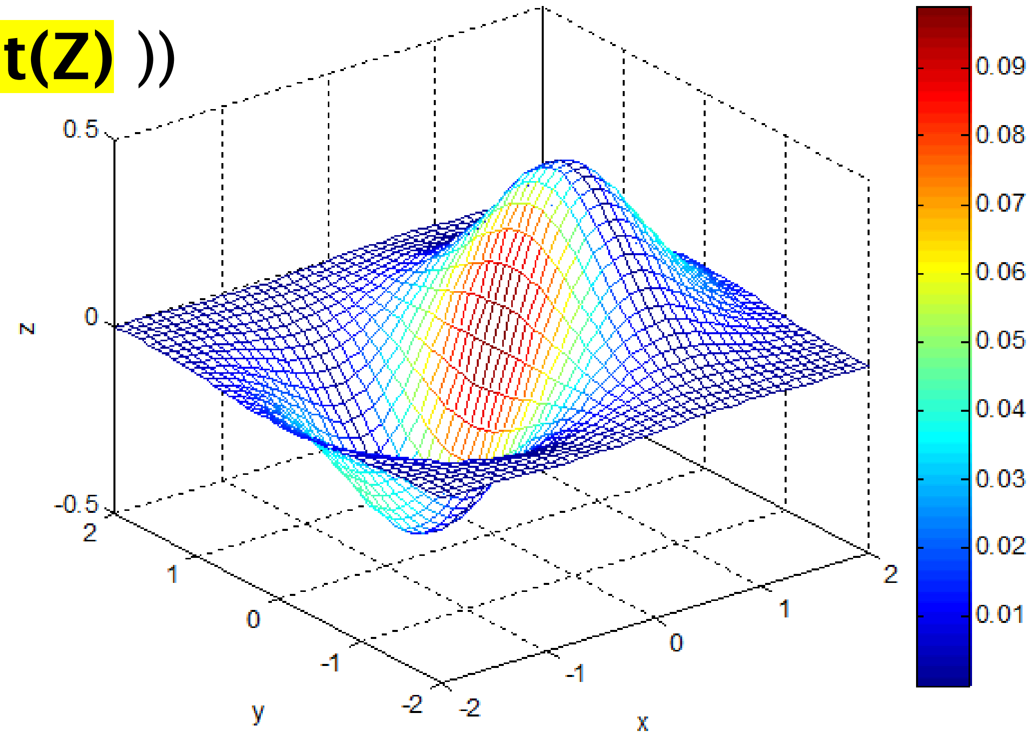
The color is the slope gradient

mesh(x,y,Z,**abs(gradient(Z)**))

xlabel('x'); ylabel('y'); zlabel('z');

title(‘ …. ‘);

colorbar; colormap jet ;

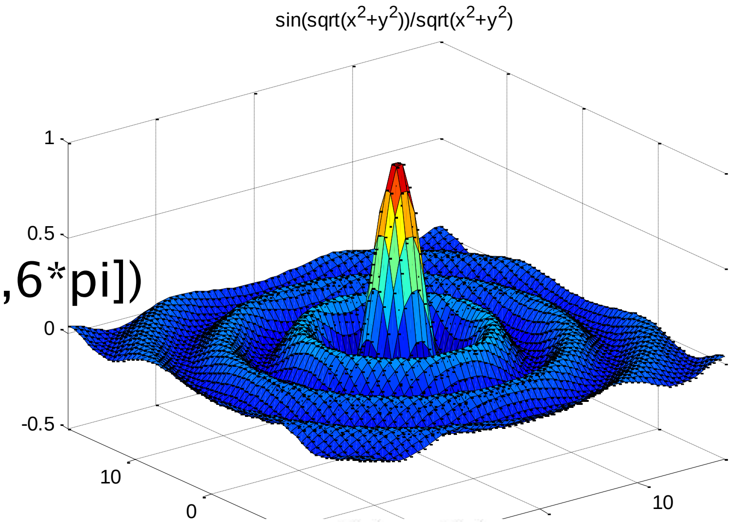


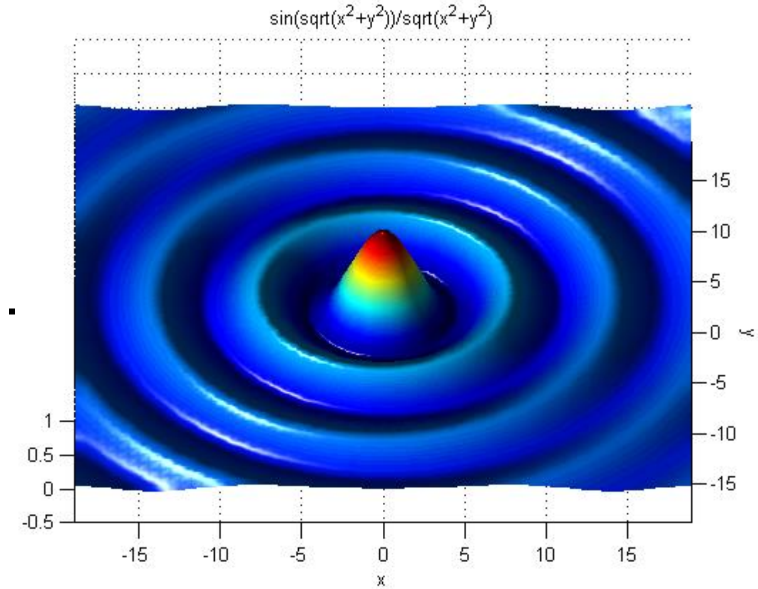
Inspect decision surfaces adding lights

ezsurf('funzione simbolica, dominio)

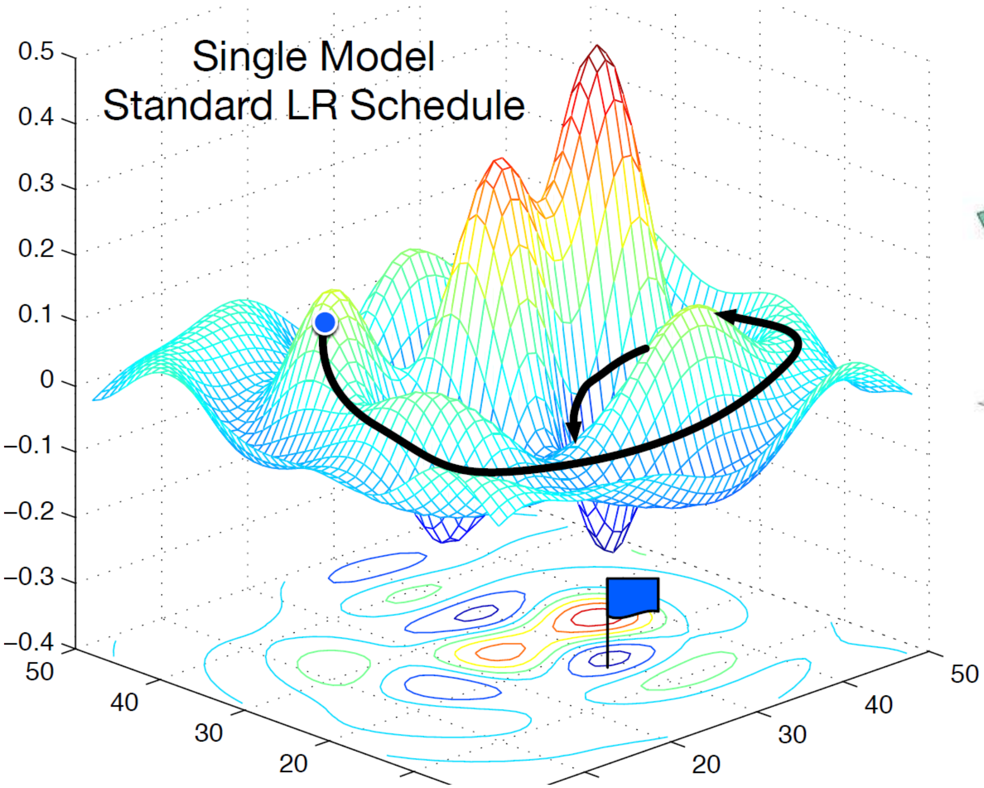
es:

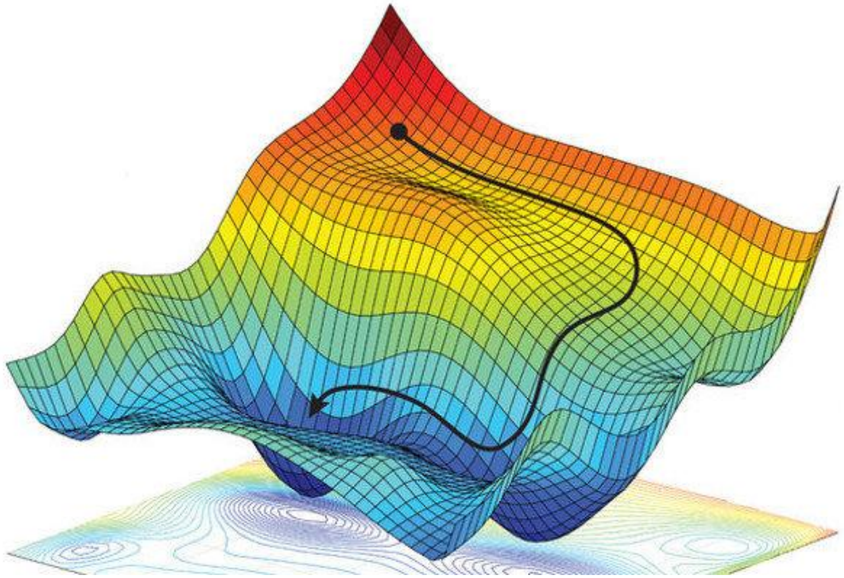
```
ezsurf('sin(sqrt(x^2+y^2))/sqrt(x^2+y^2)',[-6*pi,6*pi])

view(0,75)
shading interp
lightangle(-45,30)
set(gcf,'Renderer','zbuffer')
set(findobj(gca,'type','surface'),...
    'FaceLighting','phong',...
    'AmbientStrength',.3,'DiffuseStrength',.8,...
    'SpecularStrength',.9,'SpecularExponent',25,...
    'BackFaceLighting','unlit')
```





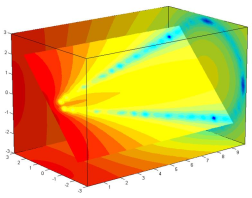
Showing the optimization process





Example of
stochastic gradient descent
in non-convex optimization

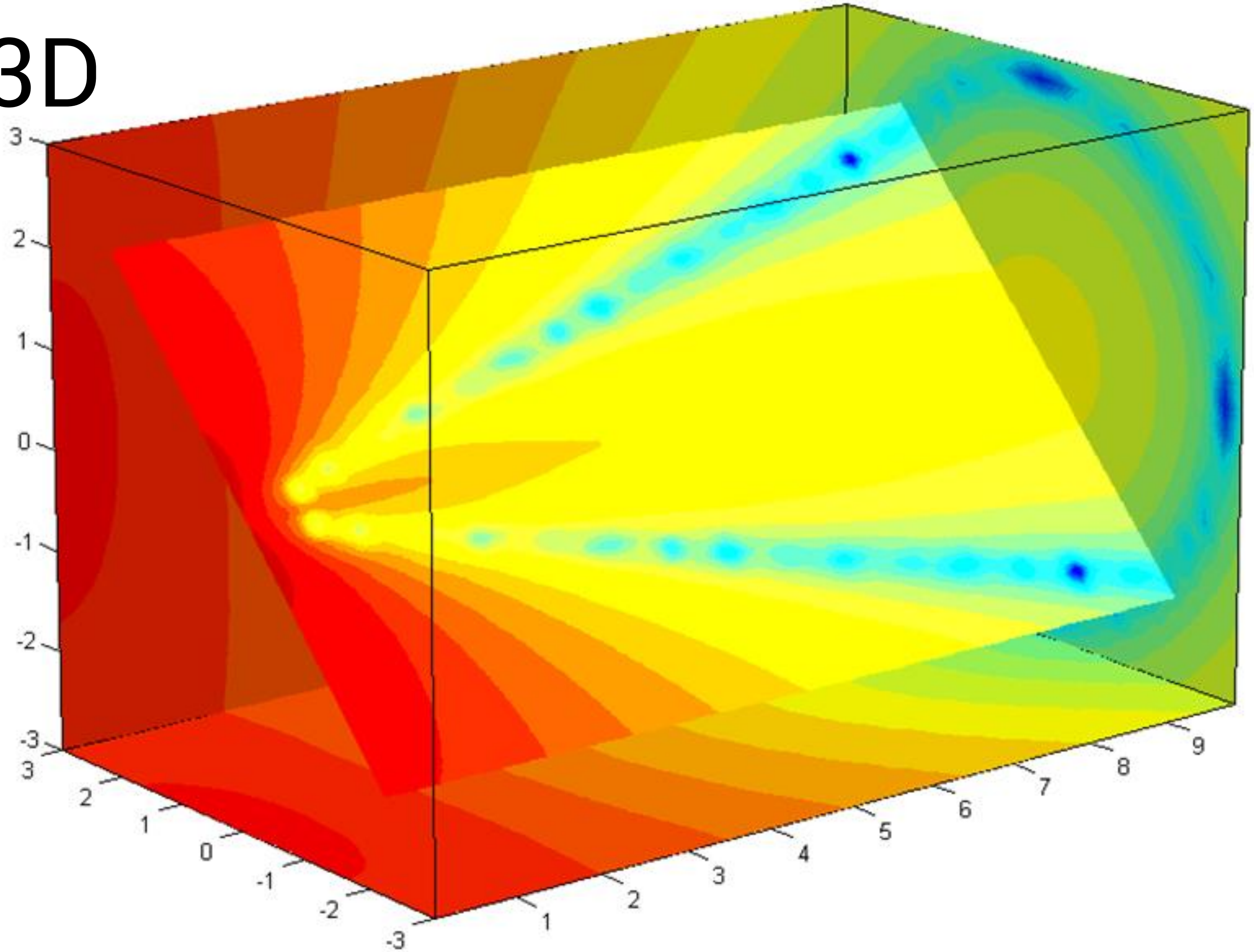
3D data



Speed vectors in a fluid jet in water (with no boundaries)

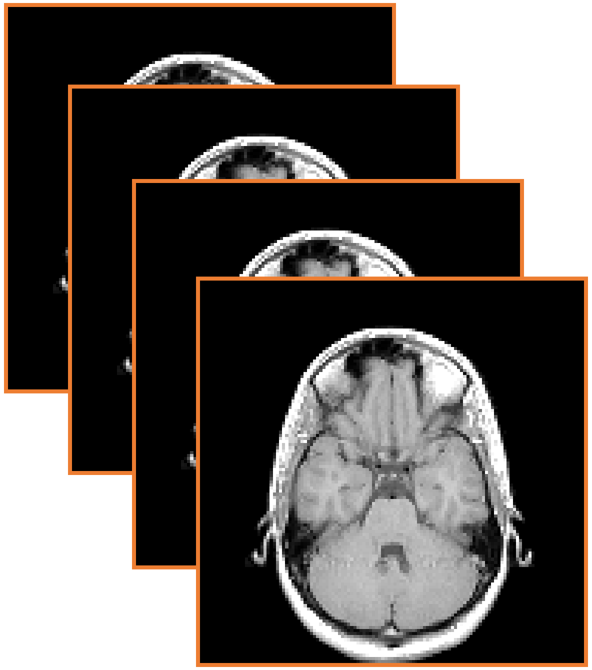
```
% Generate the volume data with the command
[x,y,z,v] = flow;

% Determine the range of the volume by finding the minimum and maximum of the coordinate data.
xmin = min(x(:));  ymin = min(y(:));  zmin = min(z(:));
xmax = max(x(:));  ymax = max(y(:));  zmax = max(z(:));
hslice = surf(linspace(xmin,xmax,100),    linspace(ymin,ymax,100),    zeros(100));
rotate(hslice,[-1,0,0],-45)
xd = get(hslice,'XData'); yd = get(hslice,'YData'); zd = get(hslice,'ZData'); delete(hslice)
h = slice(x,y,z,v,xd,yd,zd); set(h,'FaceColor','interp',    'EdgeColor','none',     'DiffuseStrength',.8)

hold on
hx = slice(x,y,z,v,xmax,[],[]); set(hx,'FaceColor','interp','EdgeColor','none')
hy = slice(x,y,z,v,[],ymax,[]); set(hy,'FaceColor','interp','EdgeColor','none')
hz = slice(x,y,z,v,[],[],zmin); set(hz,'FaceColor','interp','EdgeColor','none')

daspect([1,1,1])
axis tight
box on
view(-38.5,16)
camzoom(1.4)
camproj perspective
lightangle(-45,45)
colormap (jet(24))
set(gcf,'Renderer','zbuffer');
title('Profilo di velocita'' in un getto di fluido in un contenitore infinito');
```

3D



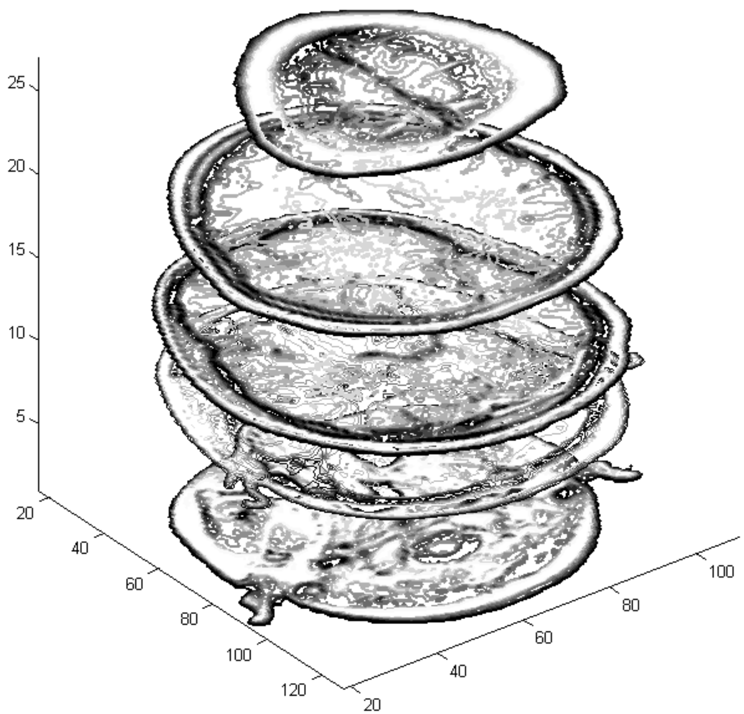
3D and surfaces

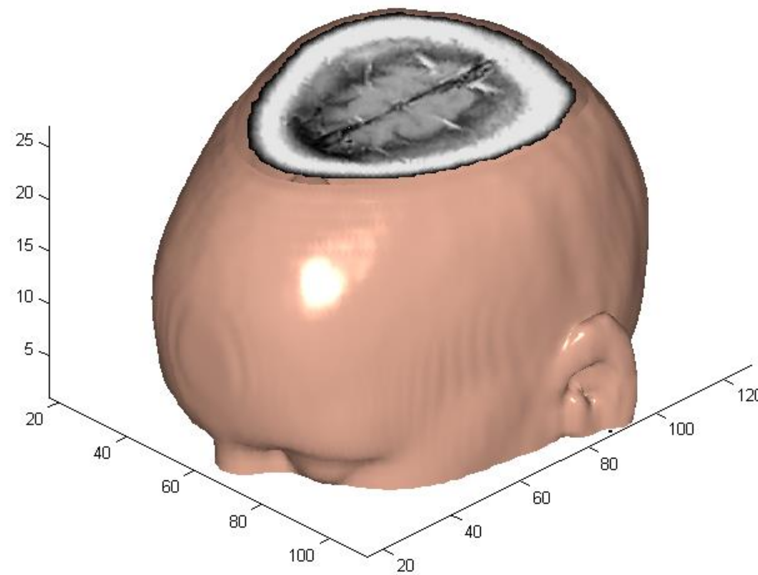
```
clear; close all; load mri; D = squeeze(D); image_num = 8;

subplot(1,2,1);   %%%%%%%%%%%%%%%%%%%%%%%%
phandles = contourslice(D,[],[],[1,12,19,27],8);
view(3); axis tight
set(phandles,'LineWidth',2)
axis square

subplot(1,2,2)   %%%%%%%%%%%%%%%%%%%%%%%%
Ds = smooth3(D);
hiso = patch(isosurface(Ds,5),...
    'FaceColor',[1,.75,.65],...
    'EdgeColor','none');

hcap = patch(isocaps(D,5),...
    'FaceColor','interp',...
    'EdgeColor','none');
colormap(map)

view(45,30)
axis tight
daspect([1,1,.4])

lightangle(45,30);
set(gcf,'Renderer','zbuffer'); lighting phong
isonormals(Ds,hiso)
set(hcap,'AmbientStrength',.6)
set(hiso,'SpecularColorReflectance',0,'SpecularExponent',50)
```



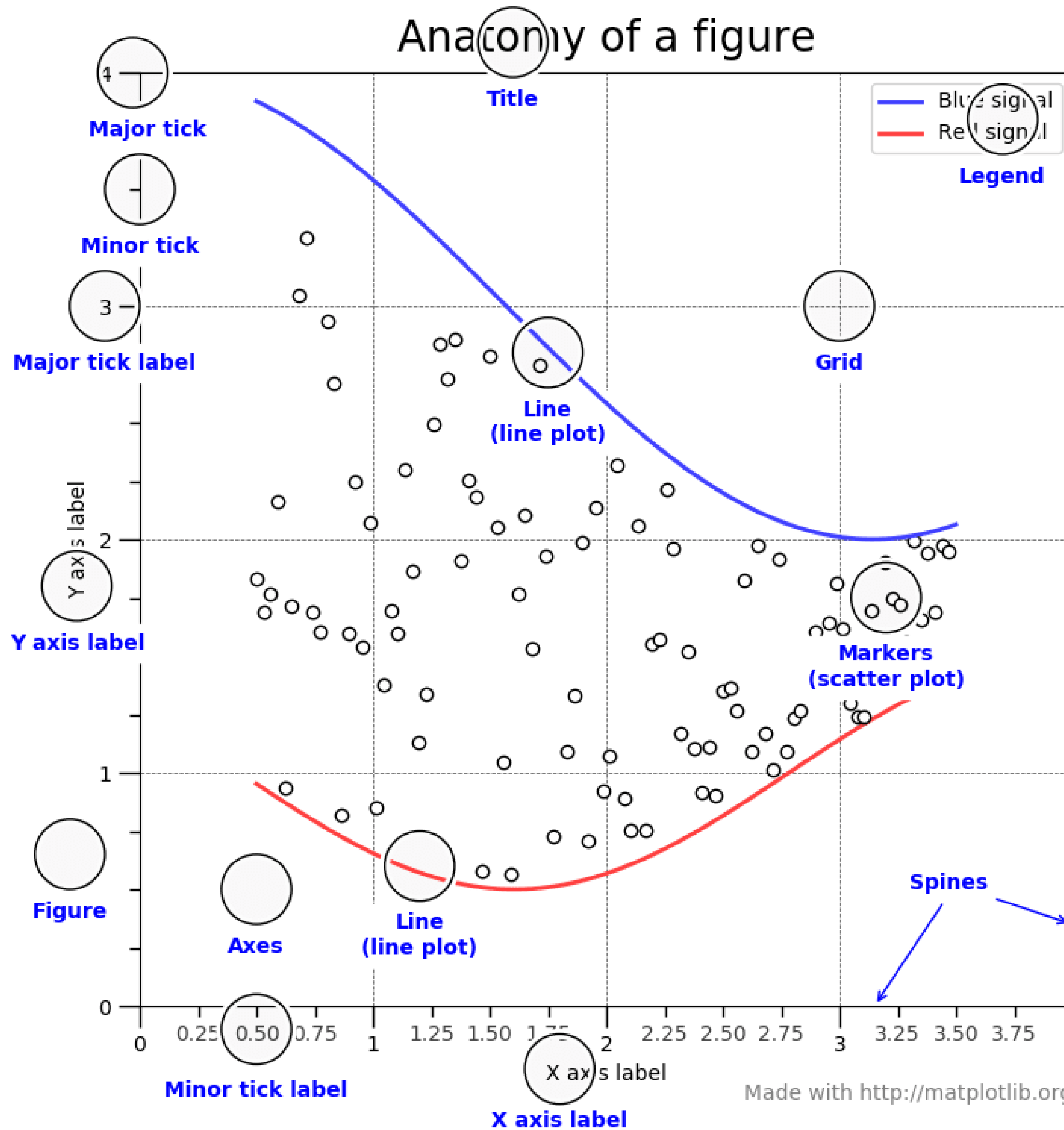
Starting from a 3D matrix ***D*** of MRI [128x128x27]

3D and surfaces





The structure of a very basic (but complete) plot



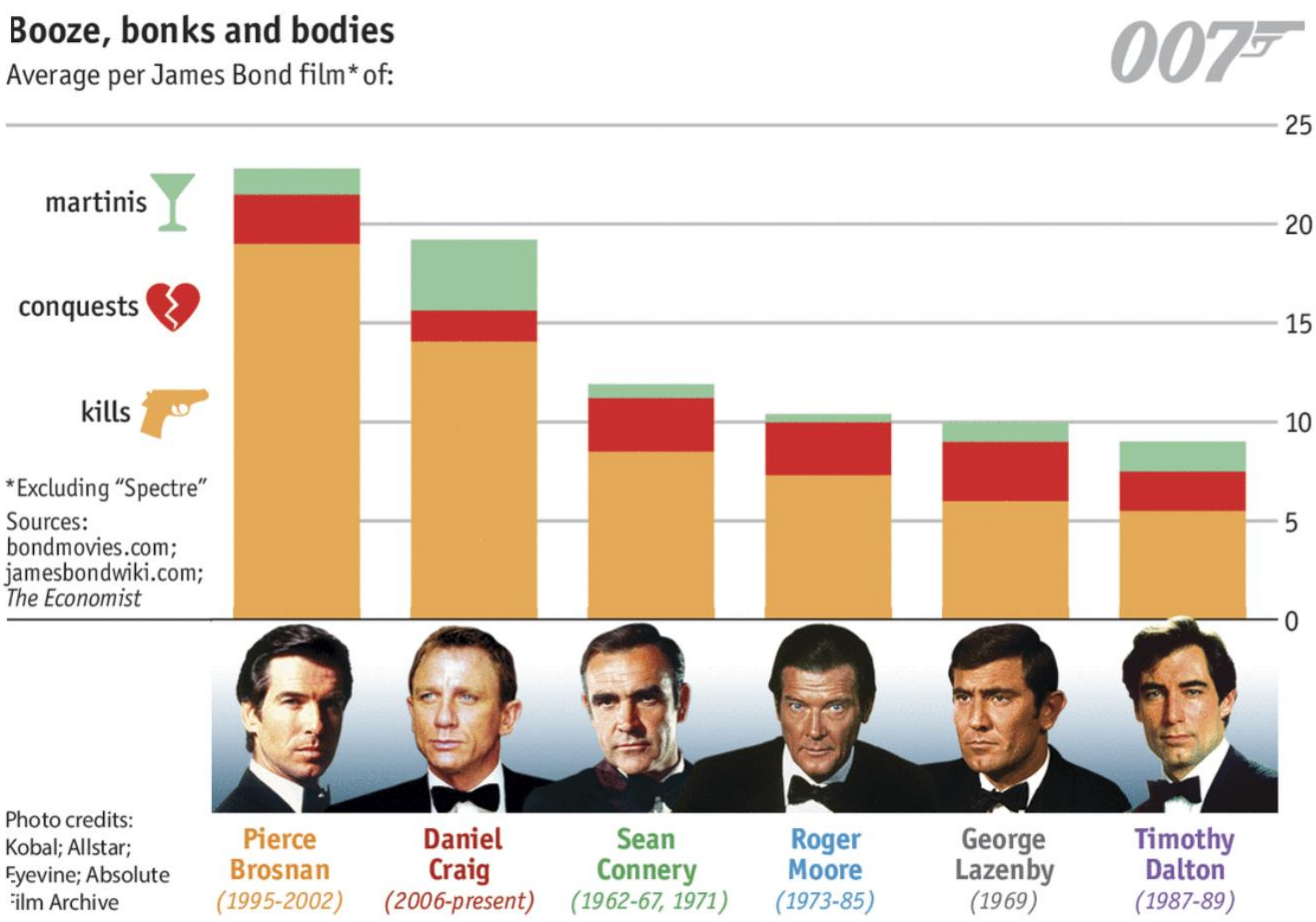
Now, just some funny slides before to close the lesson

- NOT IN THE EXAM!
- Can I compare the different 007s?
- Are you creative?
  Can I compare different lifestyles?

Eccessive use of data visualization ☺

(humor, not in the exam)

Even for strange «applications» the choice of a proper input space permits comparisons and to «think»



Are you creative?

THE DAILY ROUTINES OF

FAMOUS CREATIVE PEOPLE

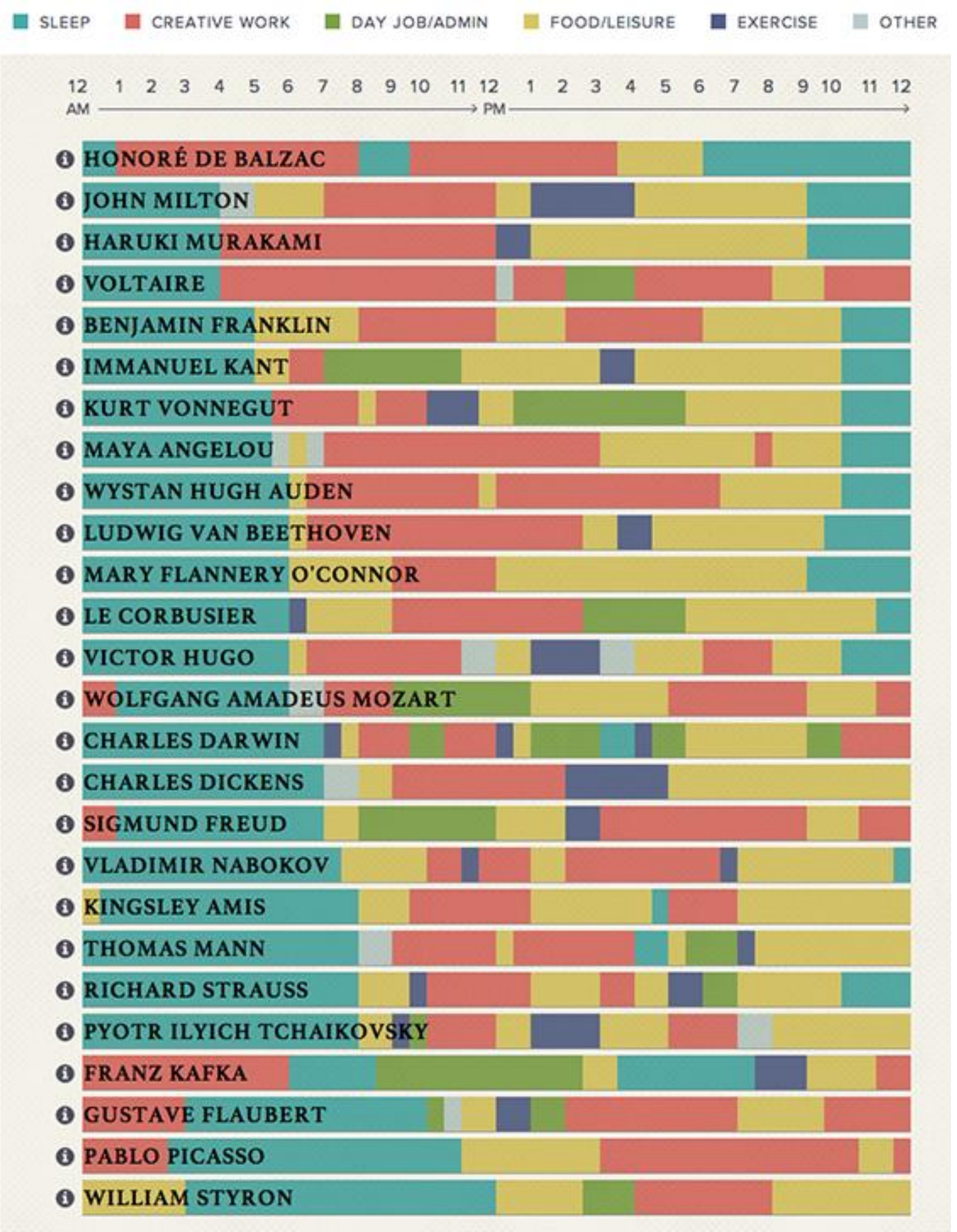
Turns out great minds don't think alike. Discover how some of the world's most original artists, writers and musicians structured their day.

Alignment is critical here,
Try to think this plot without it…

How to find the one
who is the most similar to you?

1) Decide the input space
2) Decide the metrics (distance)
3) Loop to find the closest creative!

This way you are finding the «**First Nearest Neighbor**» which is a typical ML method



Main points

Avoid funny plot and choose wisely the visual features!

One concept at time in 1 powerful image