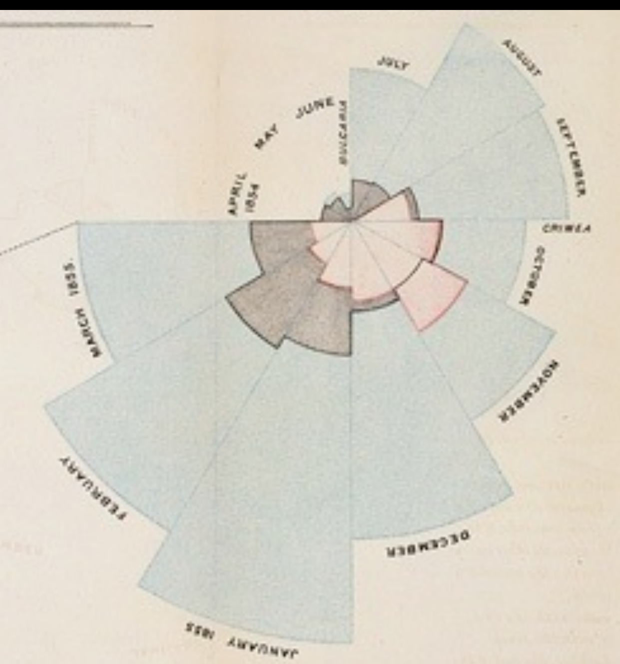
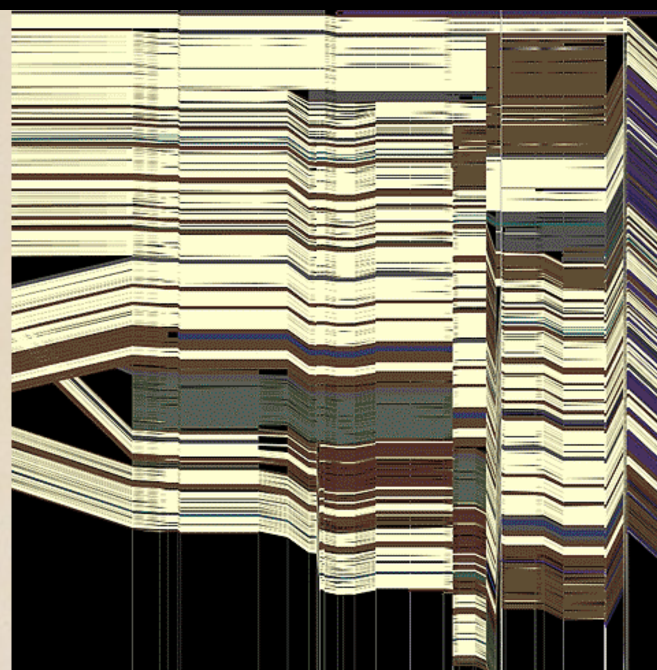
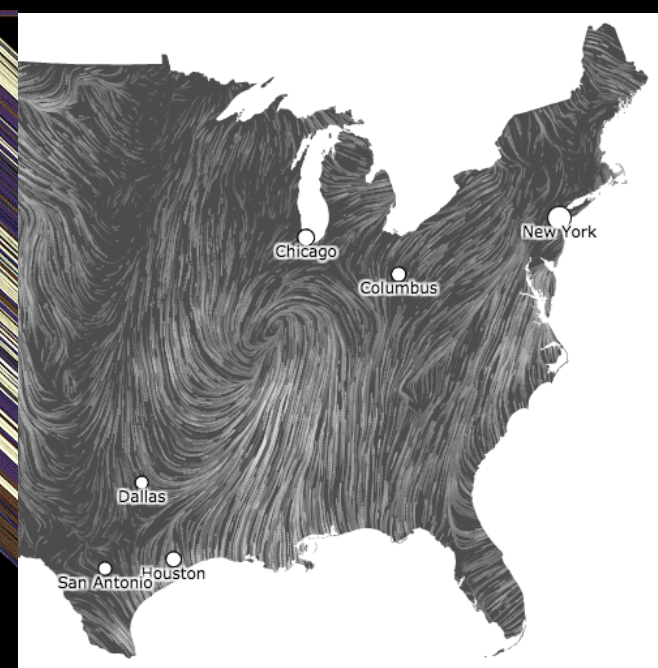
CSE 412 - Intro to Data Visualization

Multidimensional Data



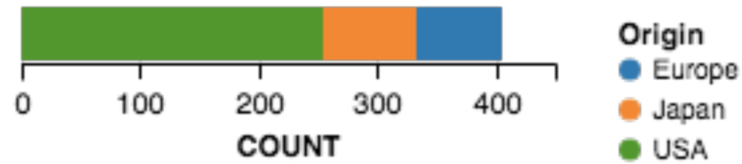




Jane Hoffswell University of Washington

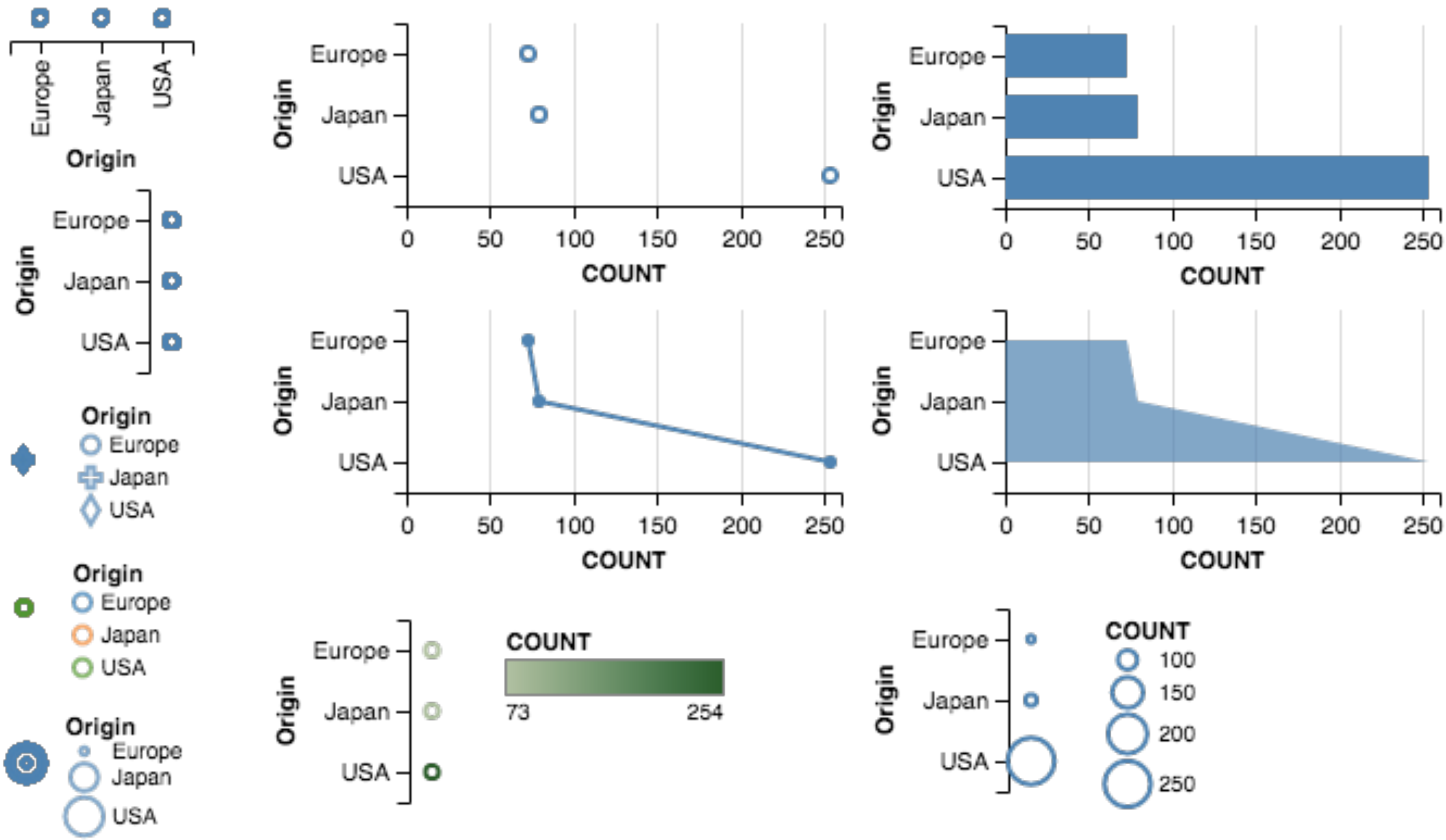
A Design Space of Visual Encodings

1D: Nominal

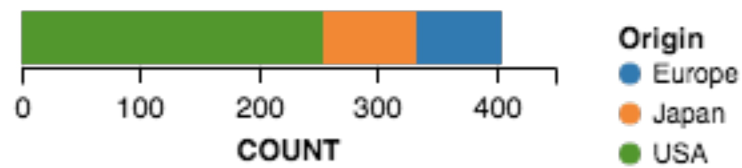


Raw

Aggregate (Count)

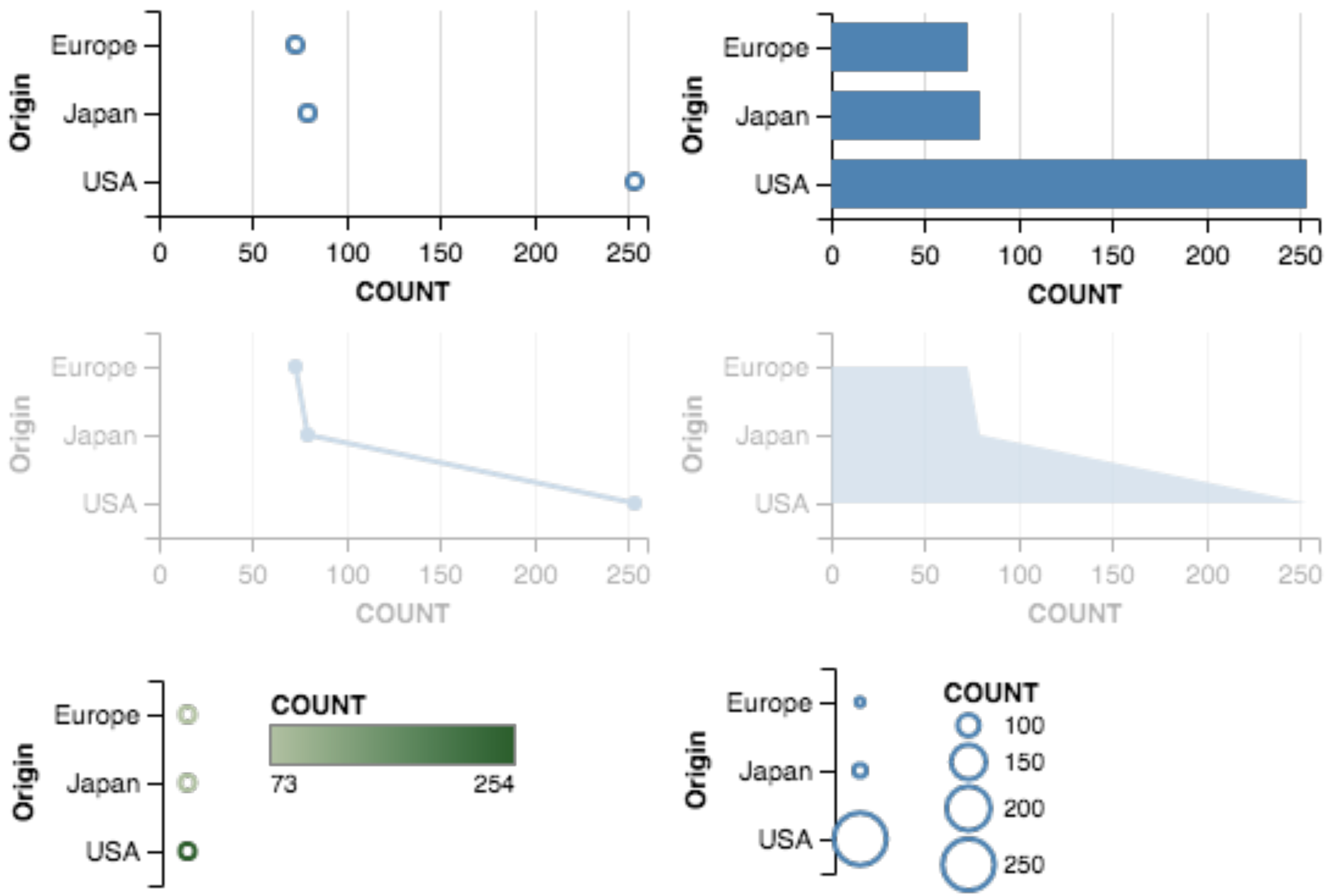


1D (N): Expressive?



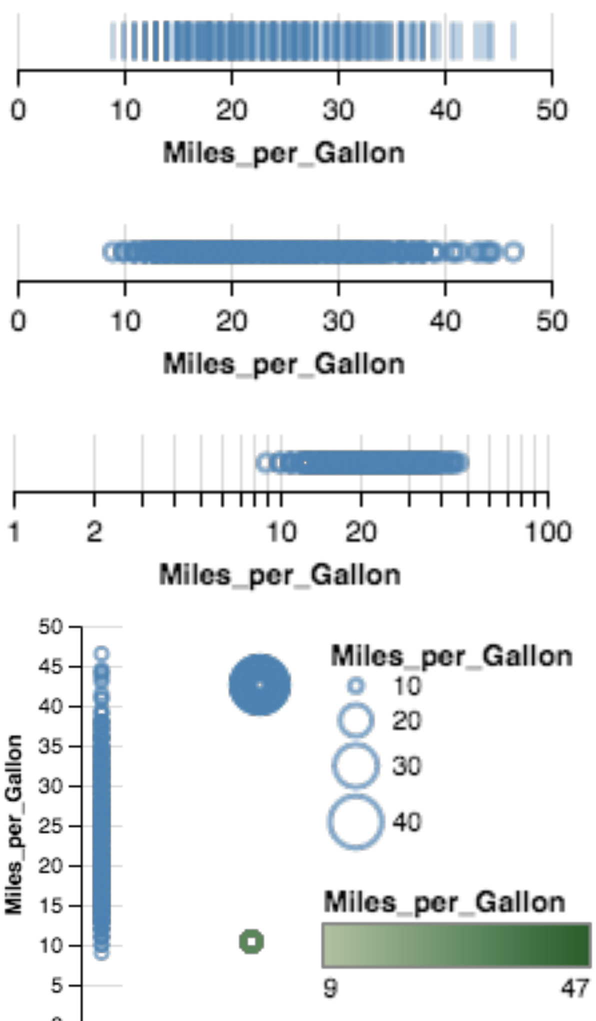
Raw

Aggregate (Count)

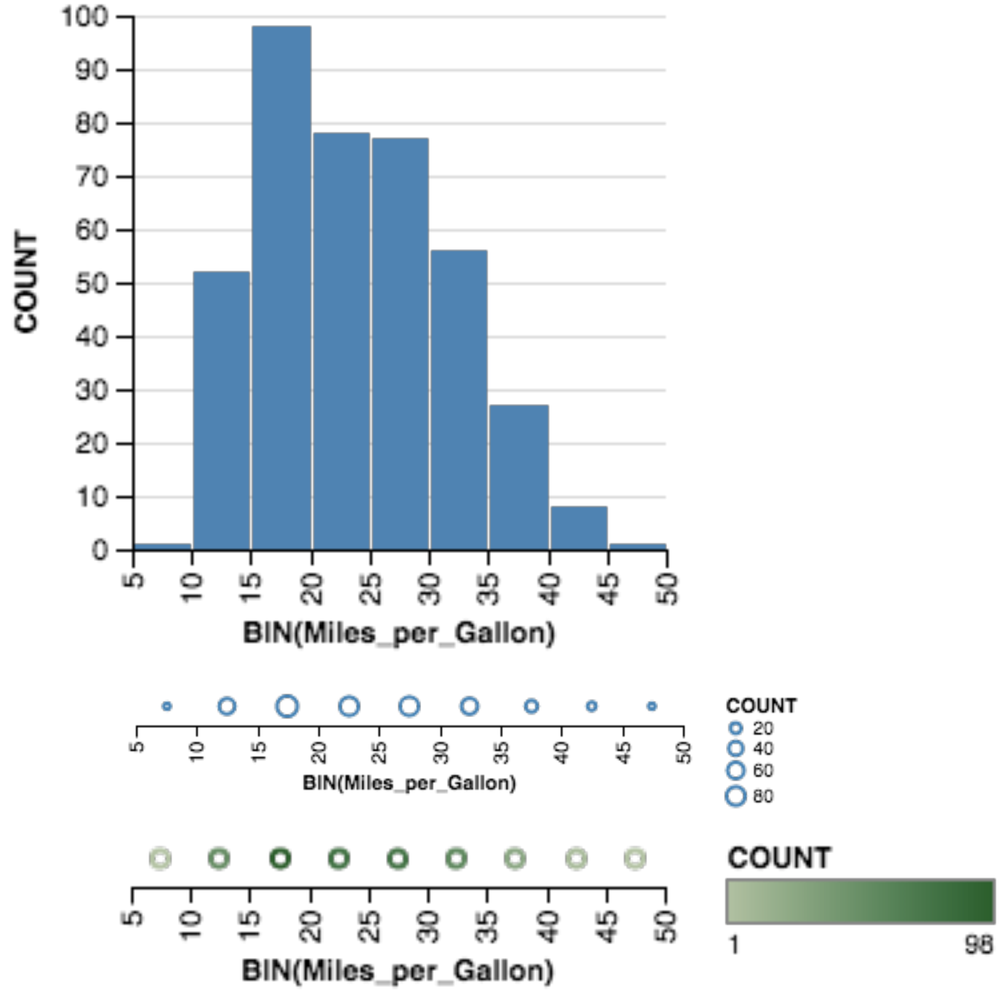


1D: Quantitative

Raw

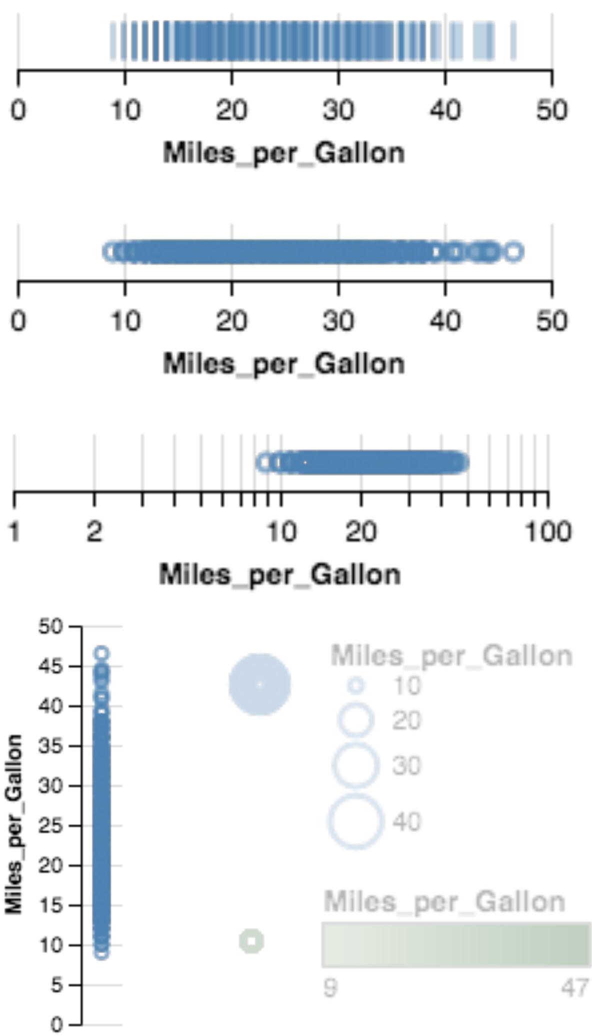


Aggregate (Count)

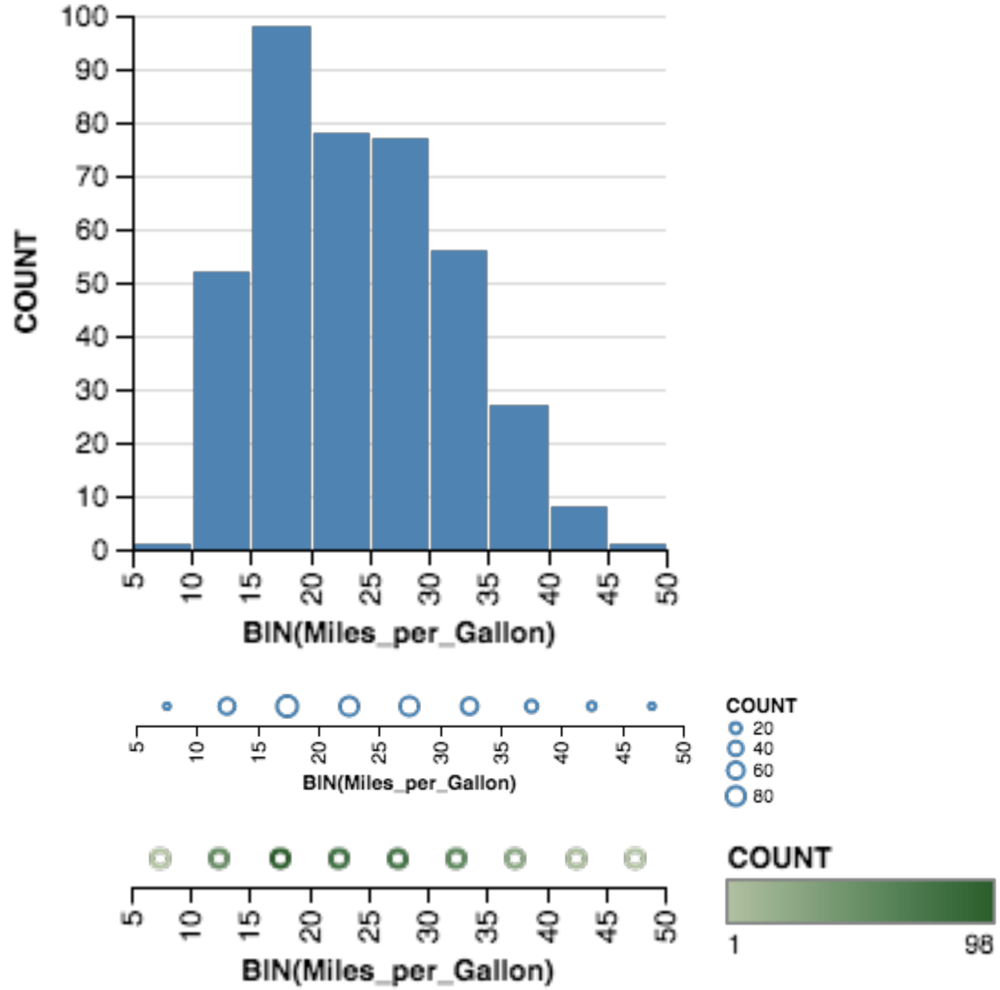


1D: Quantitative - Expressive?

Raw

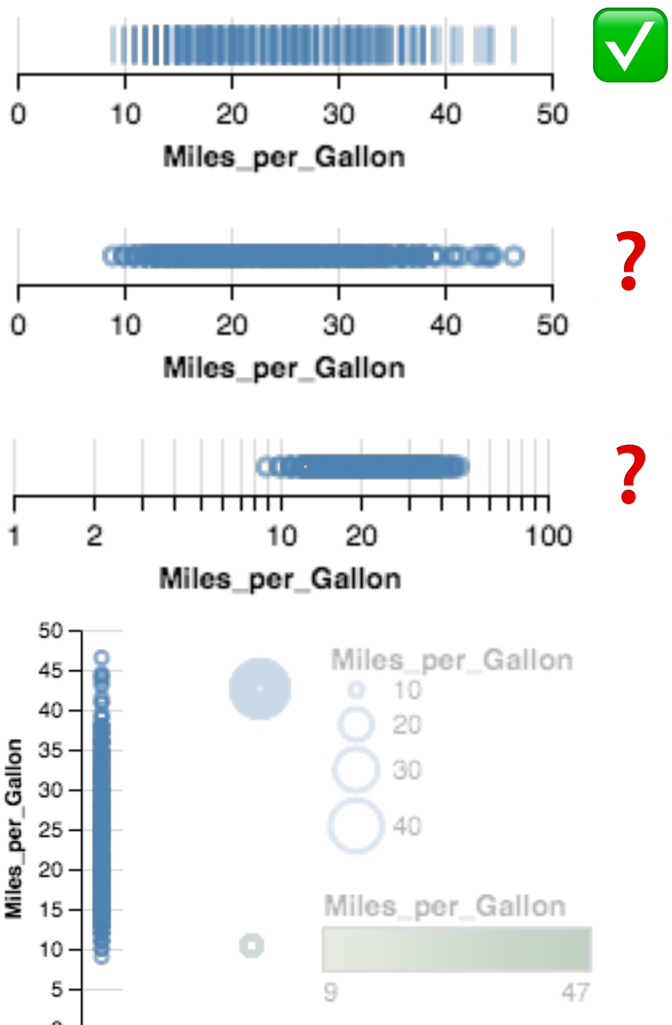


Aggregate (Count)

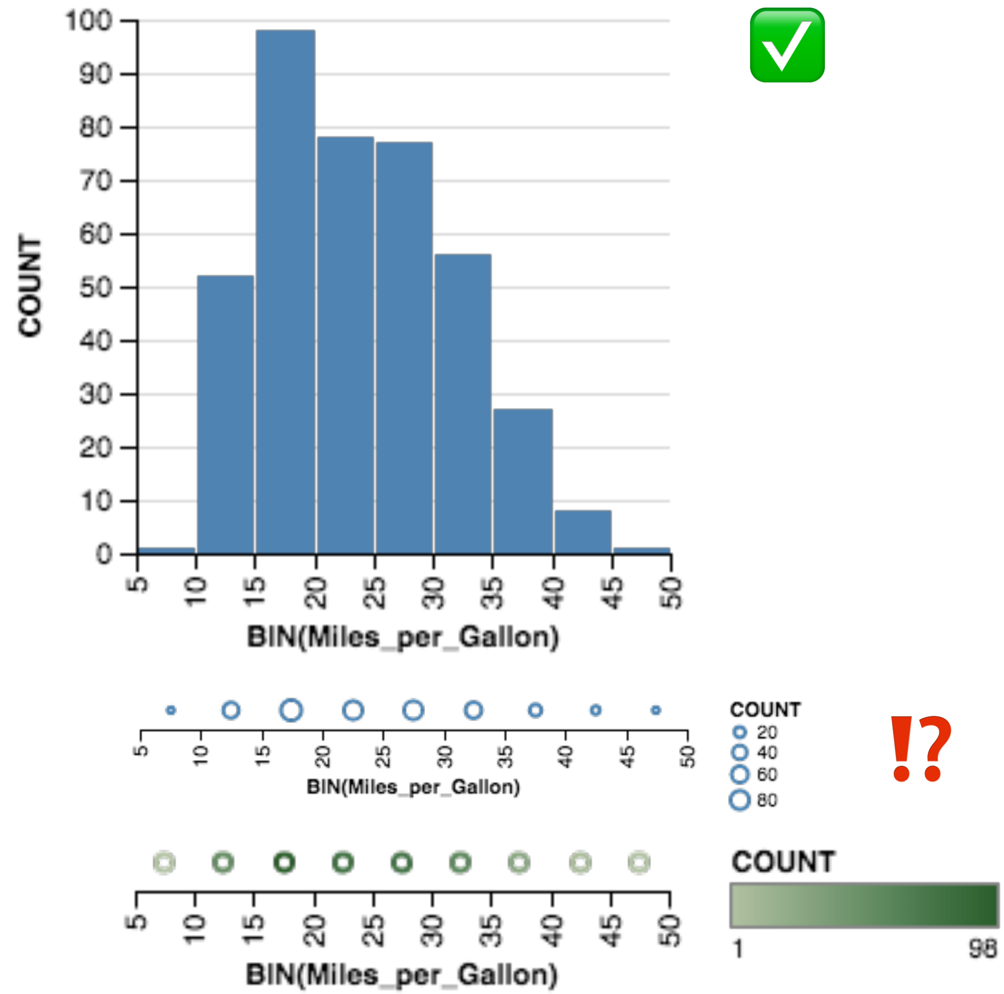


1D: Quantitative - Effective?

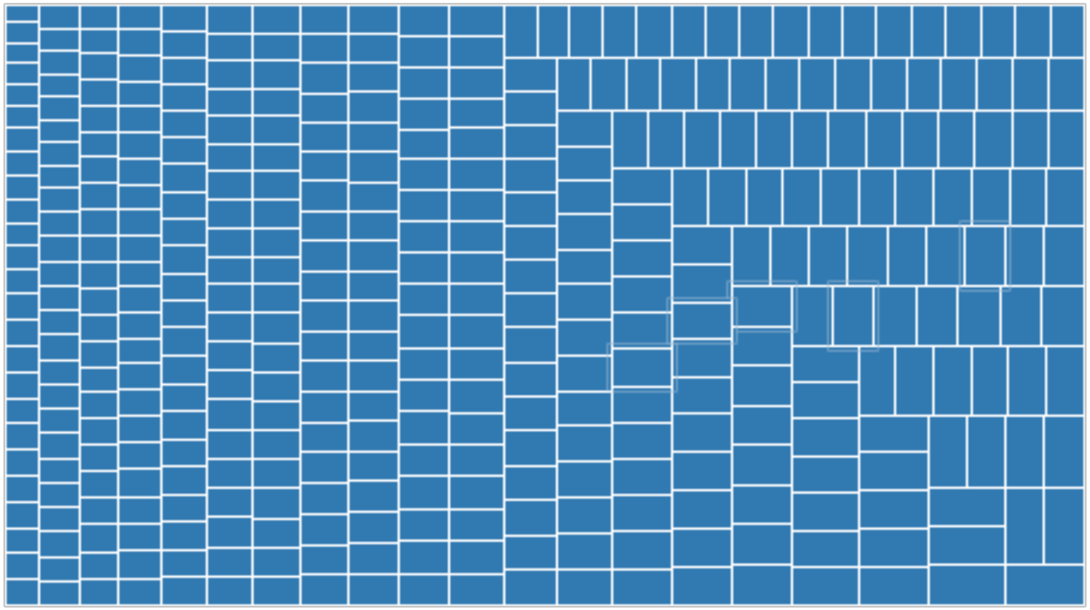
Raw



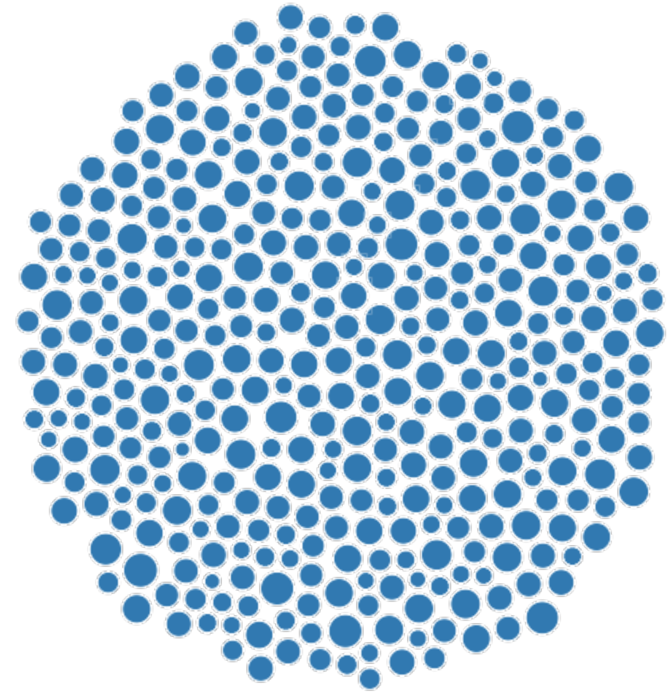
Aggregate (Count)



Raw (with Layout Algorithm)

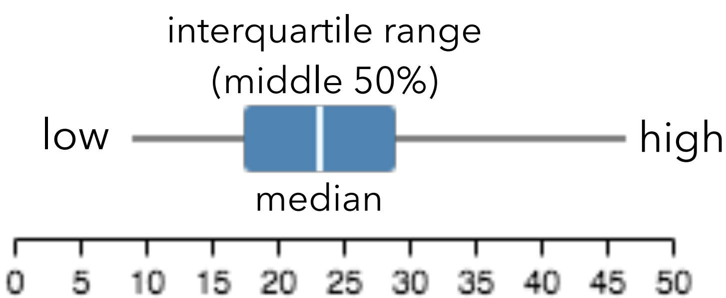


Treemap

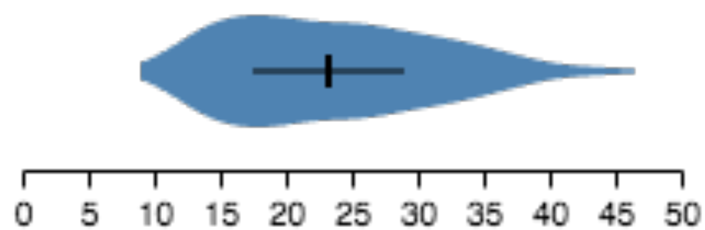


Bubble Chart

Aggregate (Distributions)



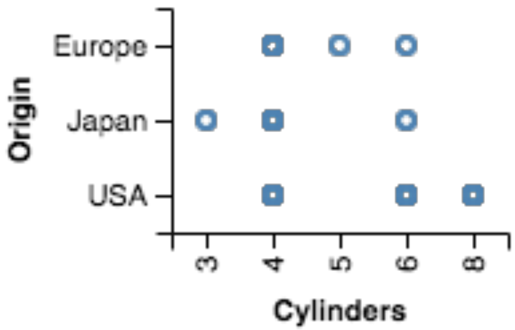
Box Plot

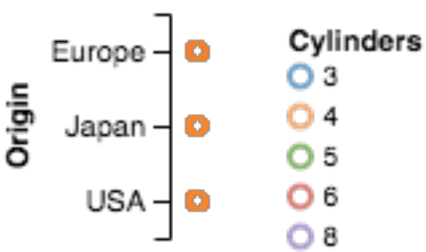


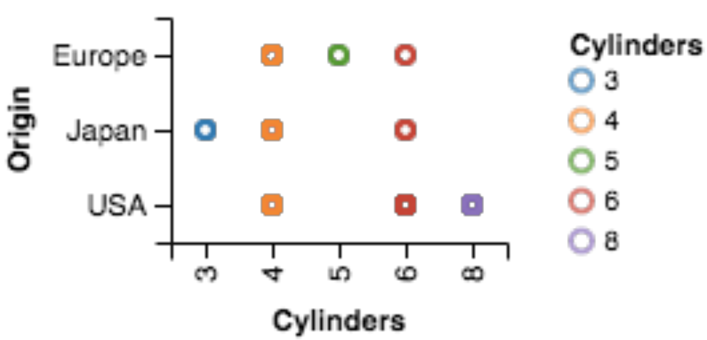
Violin Plot

2D: Nominal x Nominal

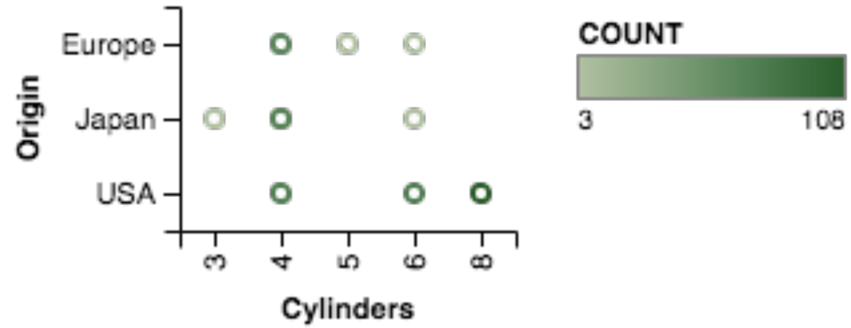
Raw

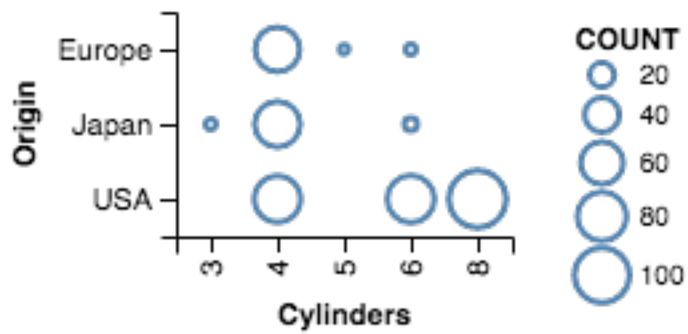


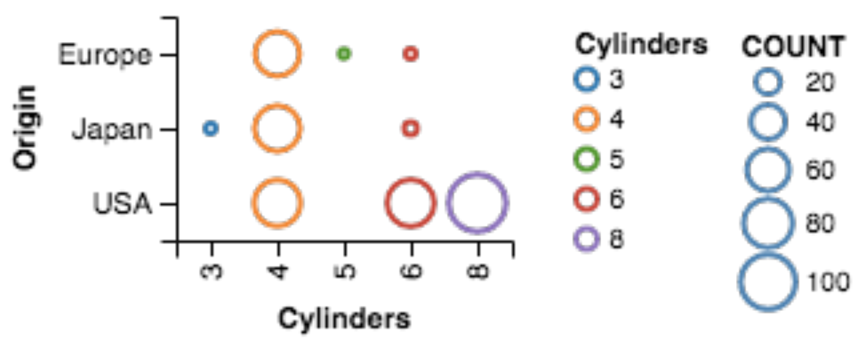




Aggregate (Count)

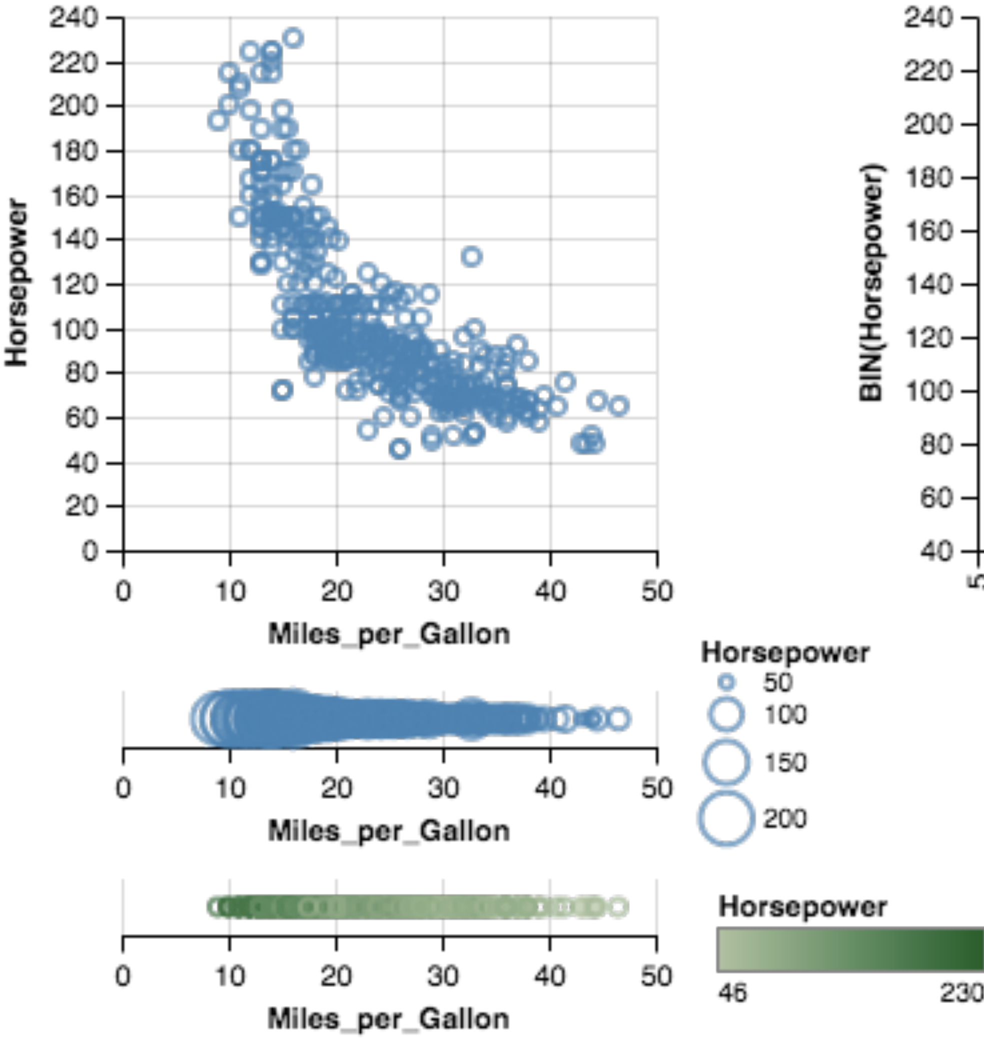




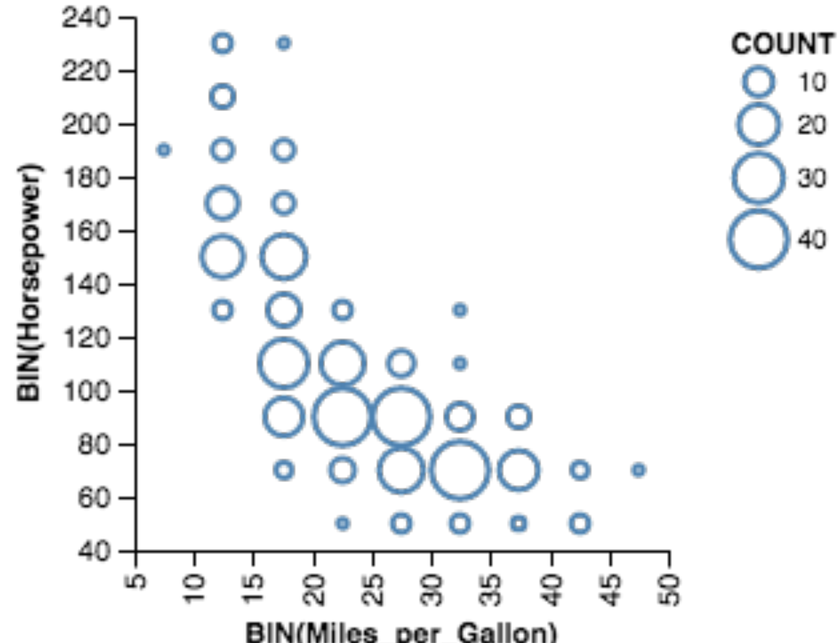


2D: Quantitative x Quantitative

Raw

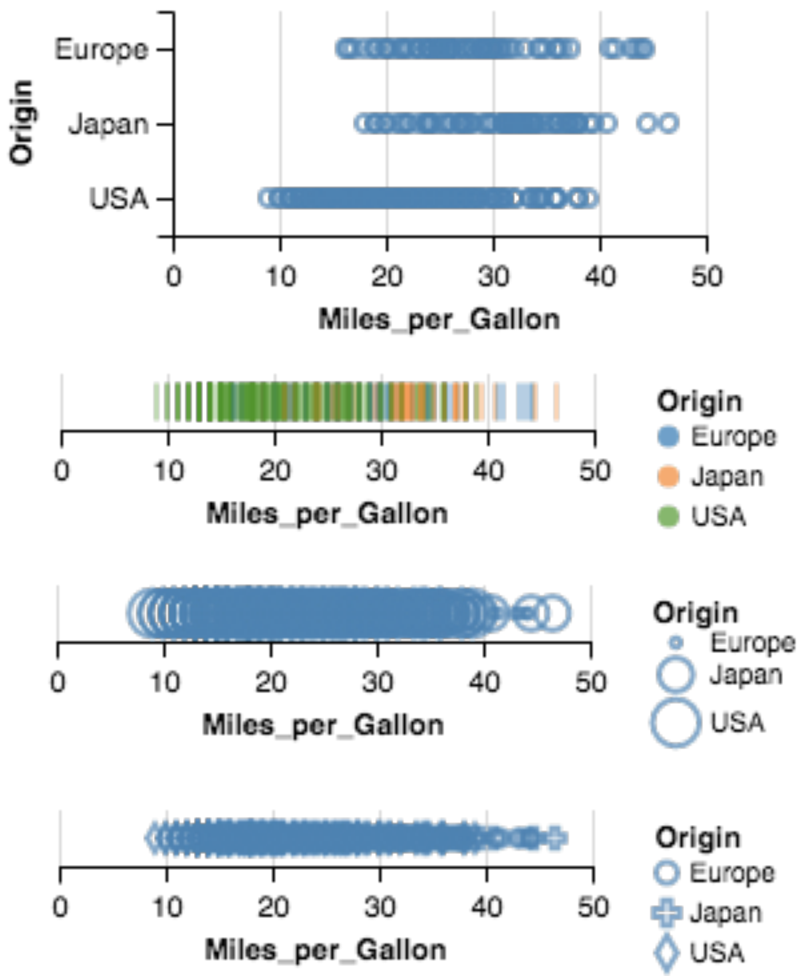


Aggregate (Count)

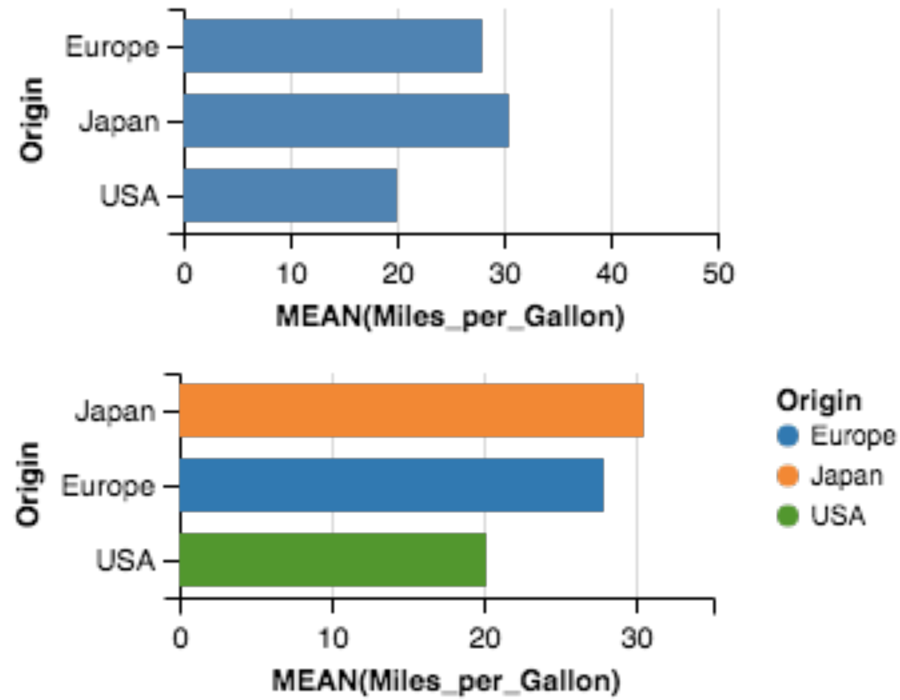


2D: Nominal x Quantitative

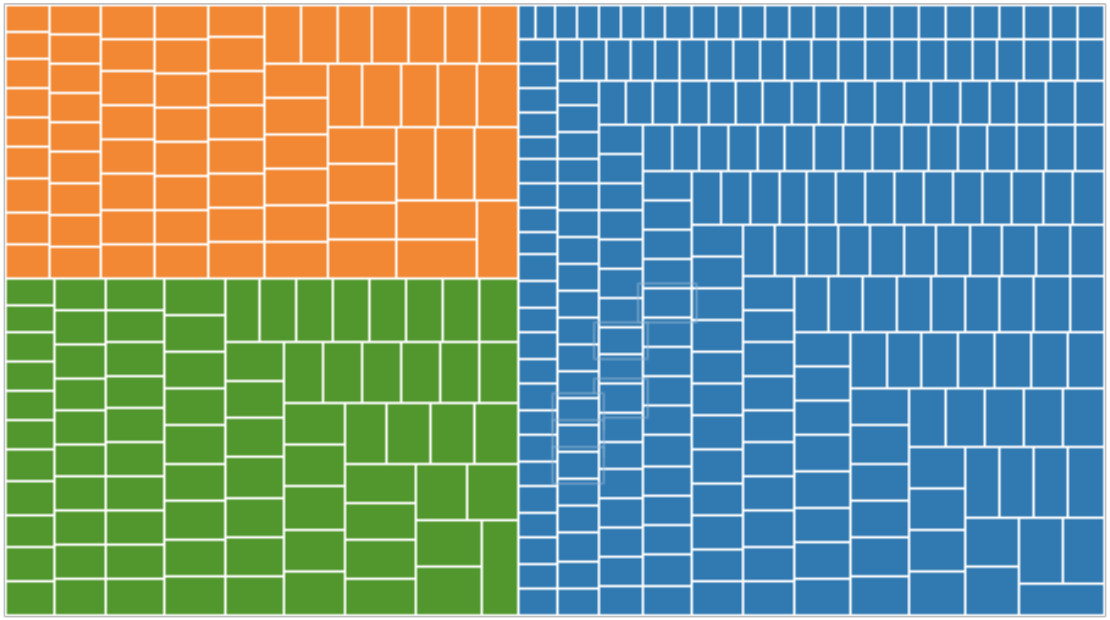
Raw



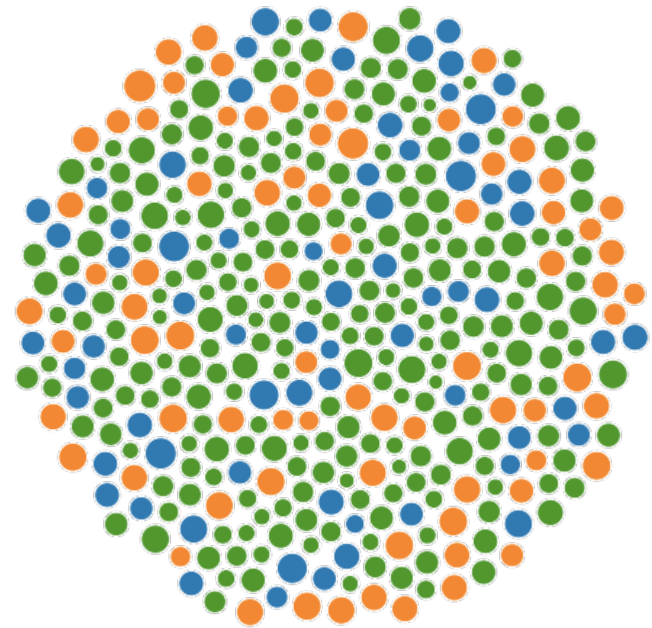
Aggregate (Mean)



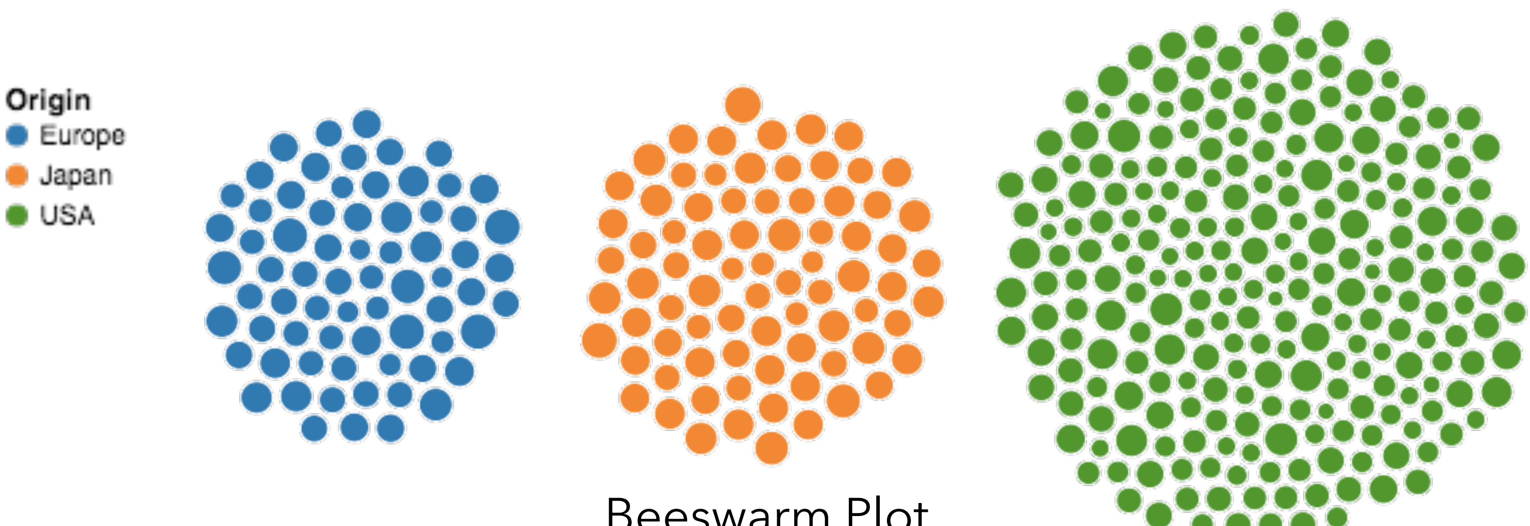
Raw (with Layout Algorithm)



Treemap



Bubble Chart



Beeswarm Plot

3D and Higher

Two variables [x,y]

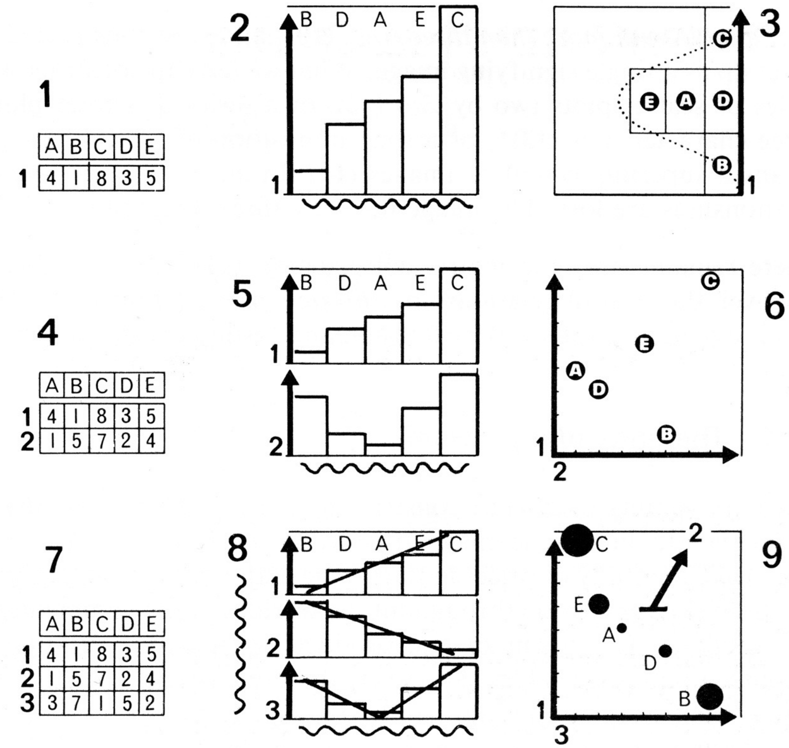
Can map to 2D points.

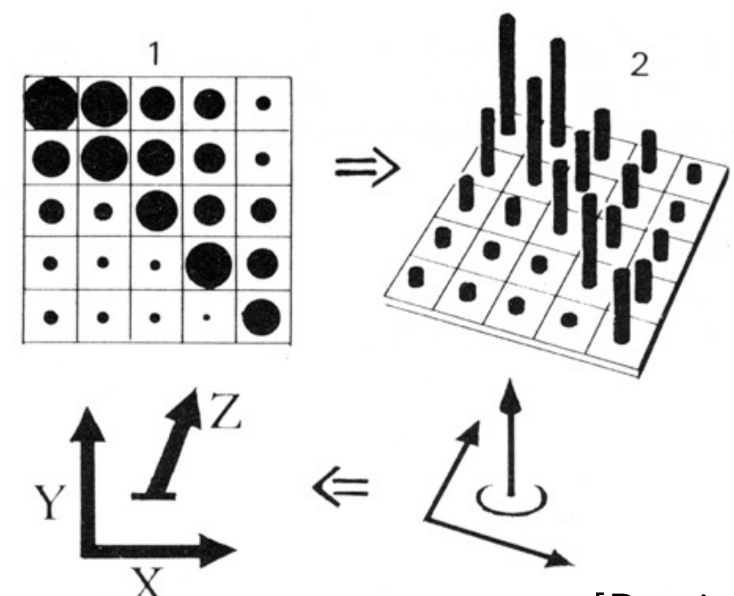
Scatterplots, maps, …

Third variable [z]

Often use one of size, color, opacity, shape, *etc*. Or, one can further partition space.

What about 3D rendering?





[Bertin]

Encoding Effectiveness

Effectiveness Rankings [Mackinlay 86]

QUANTITATIVE	ORDINAL	NOMINAL
Position	Position	Position
Length	Density (Value)	Color Hue
Angle	Color Sat	Texture
Slope	Color Hue	Connection
Area (Size)	Texture	Containment
Volume	Connection	Density (Value)
Density (Value)	Containment	Color Sat
Color Sat	Length	Shape
Color Hue	Angle	Length
Texture	Slope	Angle
Connection	Area (Size)	Slope
Containment	Volume	Area
Shape	Shape	Volume

Effectiveness Rankings [Mackinlay 86]

QUANTITATIVE	ORDINAL	NOMINAL
Position	**Position**	**Position**
Length	Density (Value)	Color Hue
Angle	Color Sat	Texture
Slope	Color Hue	Connection
Area (Size)	Texture	Containment
Volume	Connection	Density (Value)
Density (Value)	Containment	Color Sat
Color Sat	Length	Shape
Color Hue	Angle	Length
Texture	Slope	Angle
Connection	Area (Size)	Slope
Containment	Volume	Area
Shape	Shape	Volume

Effectiveness Rankings [Mackinlay 86]

QUANTITATIVE	ORDINAL	NOMINAL
Position	Position	Position
Length	Density (Value)	**Color Hue**
Angle	Color Sat	Texture
Slope	**Color Hue**	Connection
Area (Size)	Texture	Containment
Volume	Connection	Density (Value)
Density (Value)	Containment	Color Sat
Color Sat	Length	Shape
Color Hue	Angle	Length
Texture	Slope	Angle
Connection	Area (Size)	Slope
Containment	Volume	Area
Shape	Shape	Volume

Mapping Data to Visual Variables

Assign **data fields** (e.g., with *N, O, Q* types) to **visual channels** (*x, y, color, shape, size, …*) for a chosen **graphical mark** type (*point, bar, line, …*).

Additional concerns include choosing appropriate **encoding parameters** (*log scale, sorting, …*) and **data transformations** (*bin*, *group, aggregate, …*).

These options define a large combinatorial space, containing both useful and questionable charts!

Designing Charts

Example: Cars

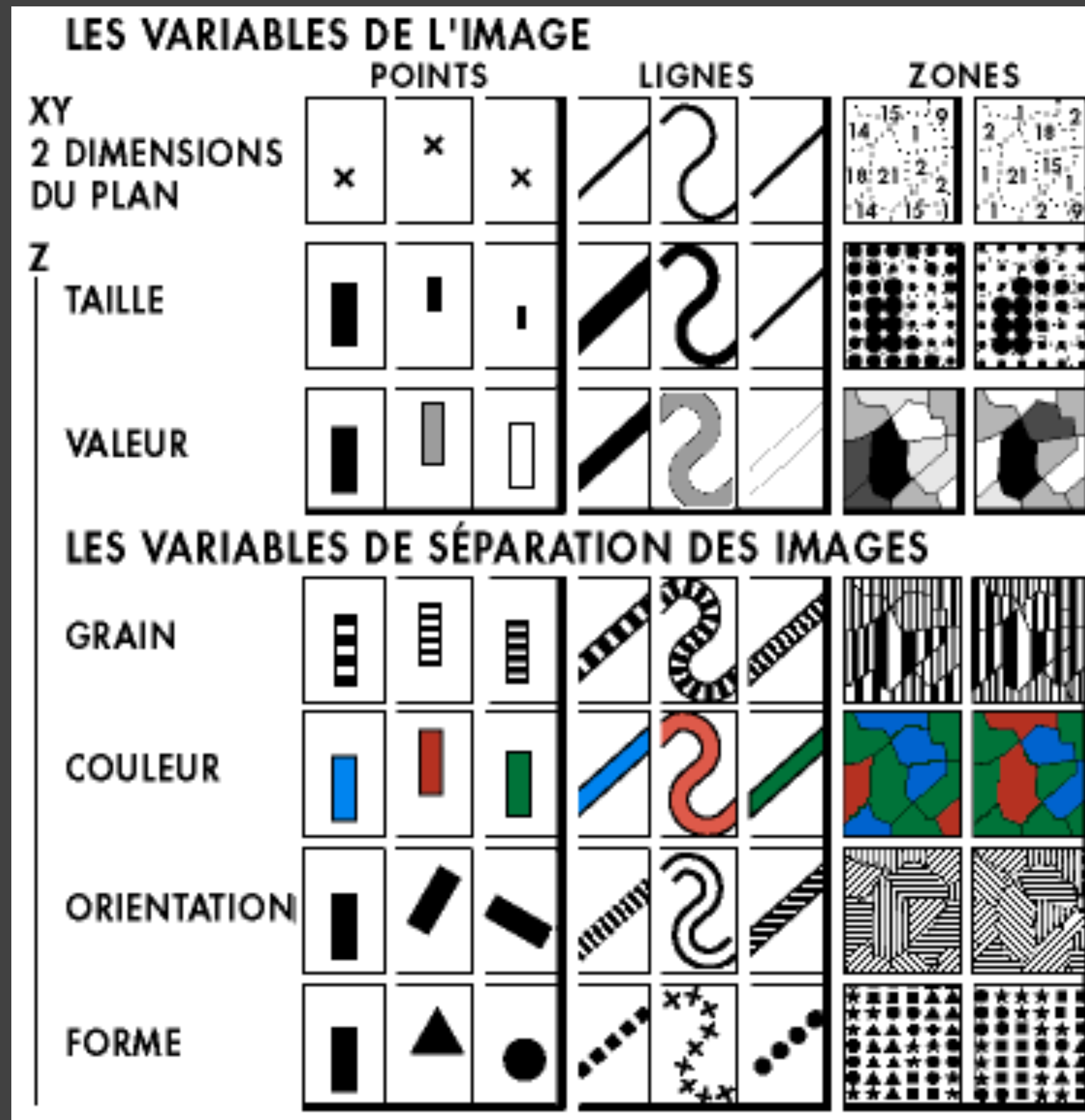
Properties of different models of cars

9-Dimensions, 406 rows

Name	N	Horsepower	Q
MPG	Q	Weight (lbs)	Q
Cylinders	Q	Acceleration	Q
Displacement	Q	Year	T
Origin	N {USA, Europe, Japan}		

Visual Encoding Variables

Position (X)
Position (Y)
Area
Value
Texture
Color
Orientation
Shape



~9 dimensions?

Example: Cars

Properties of different models of cars

9-Dimensions, 406 rows

Name	N	Horsepower	Q
MPG	Q	Weight (lbs)	Q
Cylinders	Q	Acceleration	Q
Displacement	Q	Year	T
Origin	N {USA, Europe, Japan}		

Example: Cars

Horsepower

MPG

Origin

Displacement

Cylinders

Weight (lbs)

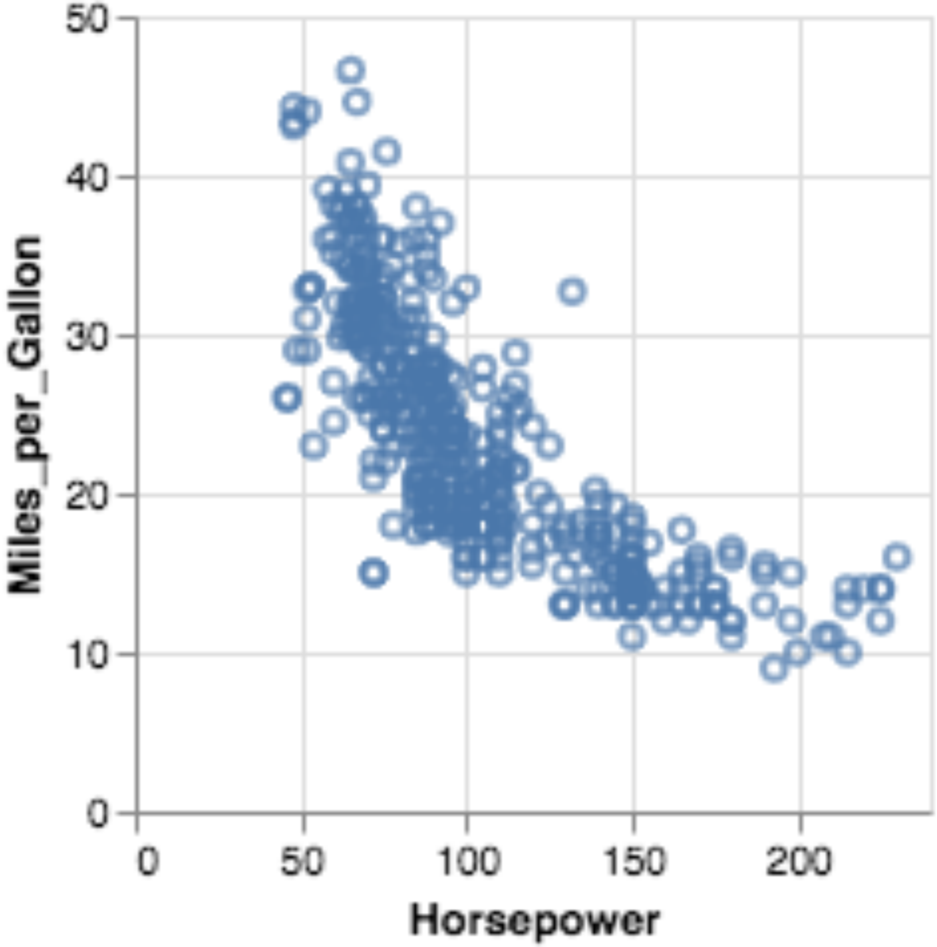
Acceleration

Year

Name

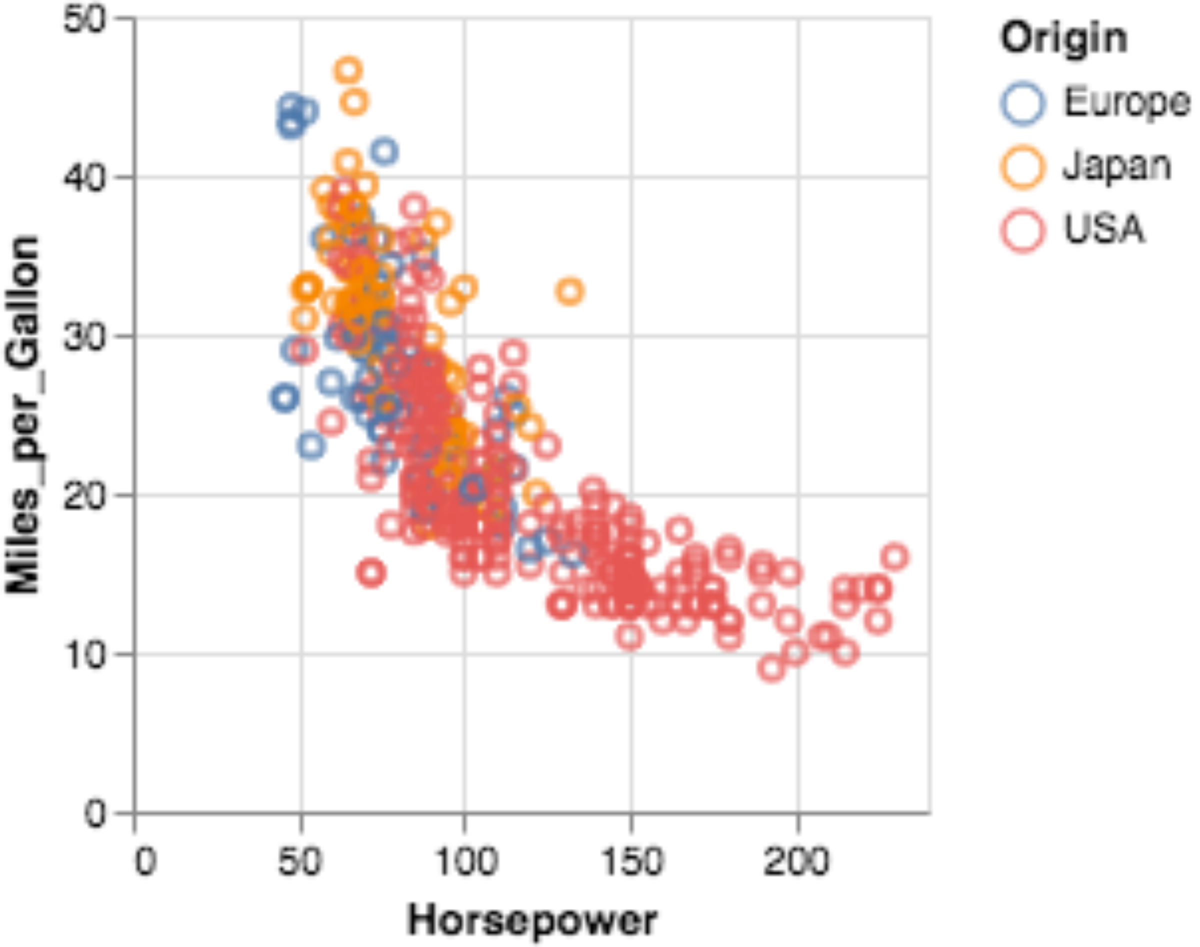
2-Dimensions

Next Up:Origin (N)



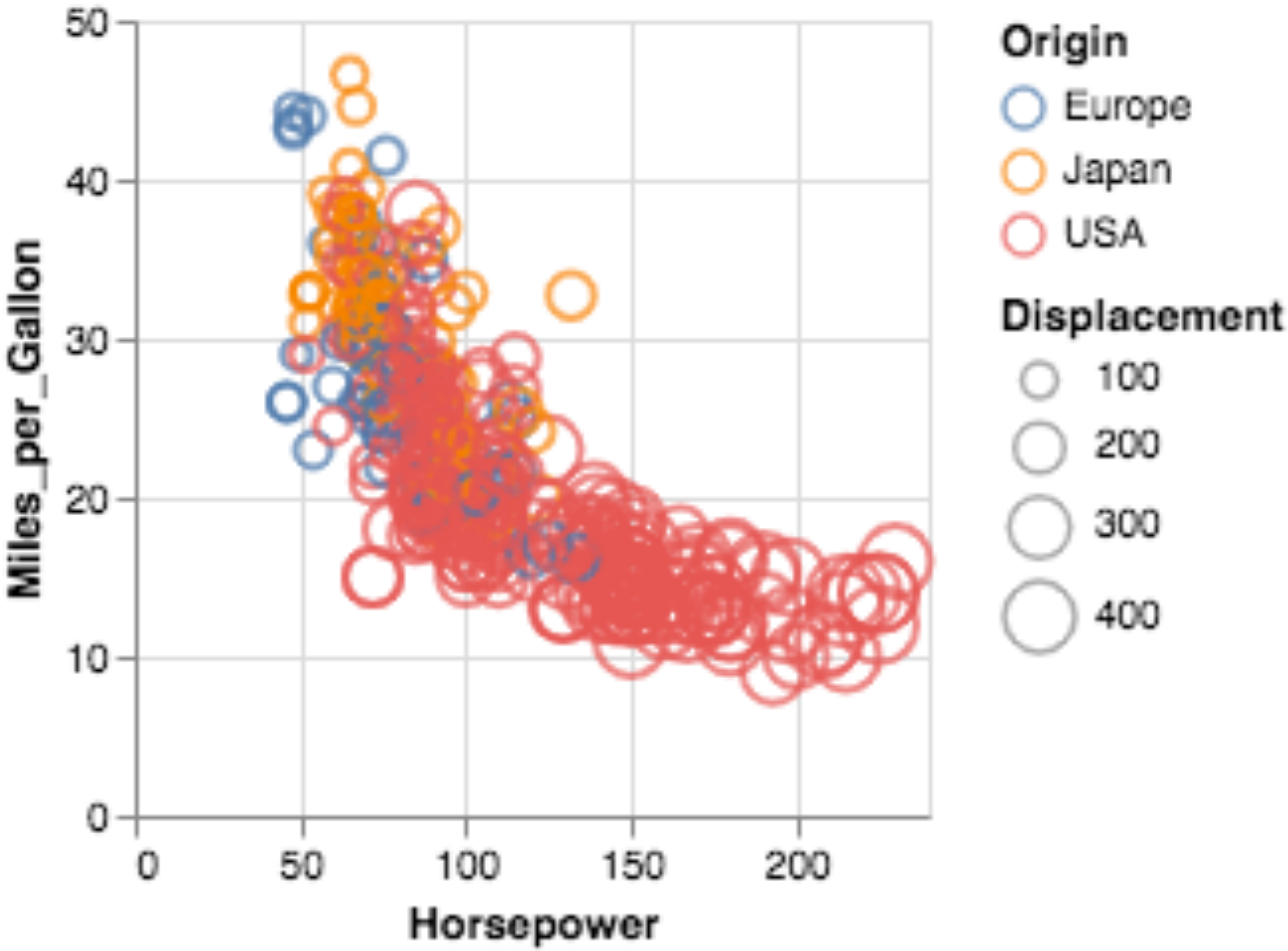
3-Dimensions

Next Up: Displacement (Q)



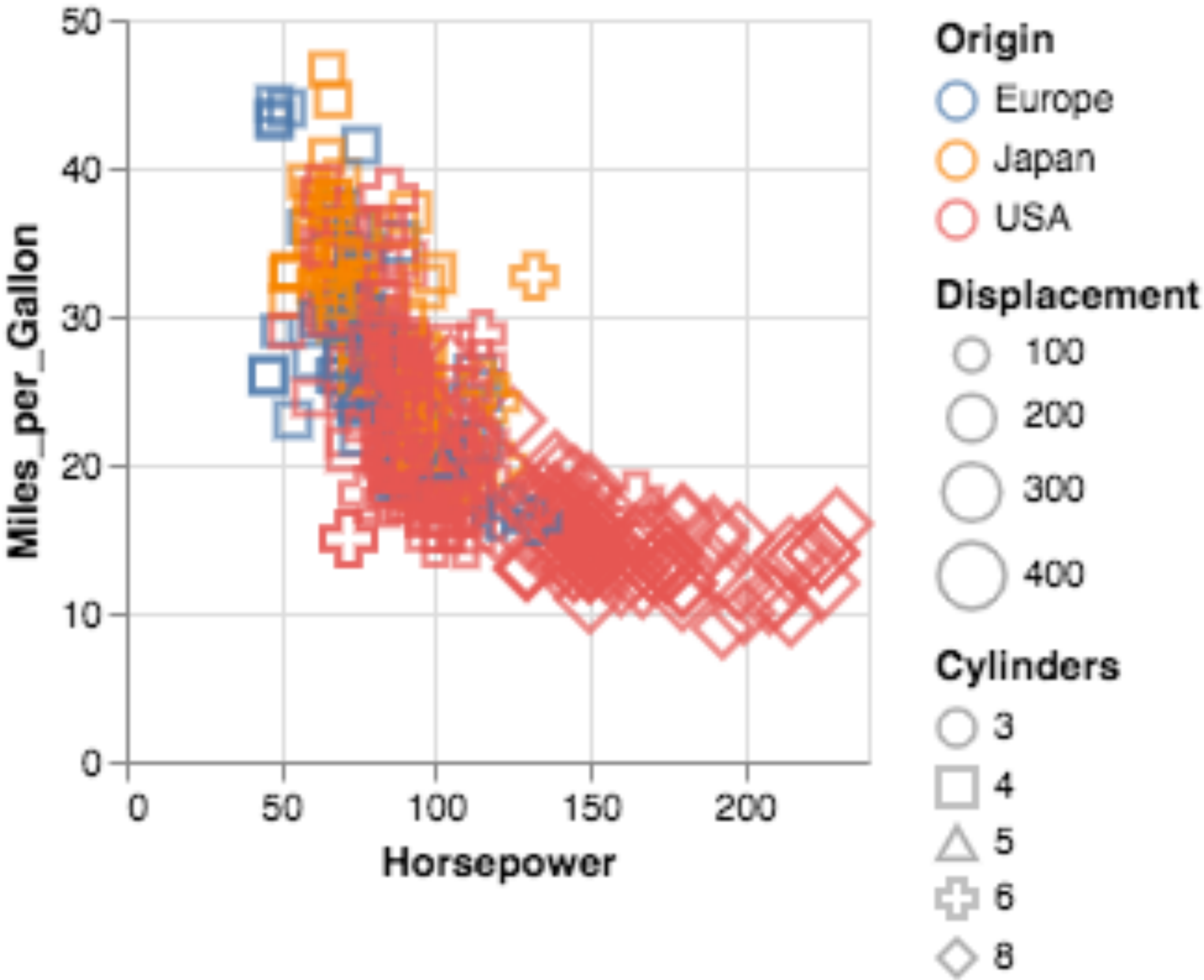
4-Dimensions

Next Up:
Cylinders (O/Q)



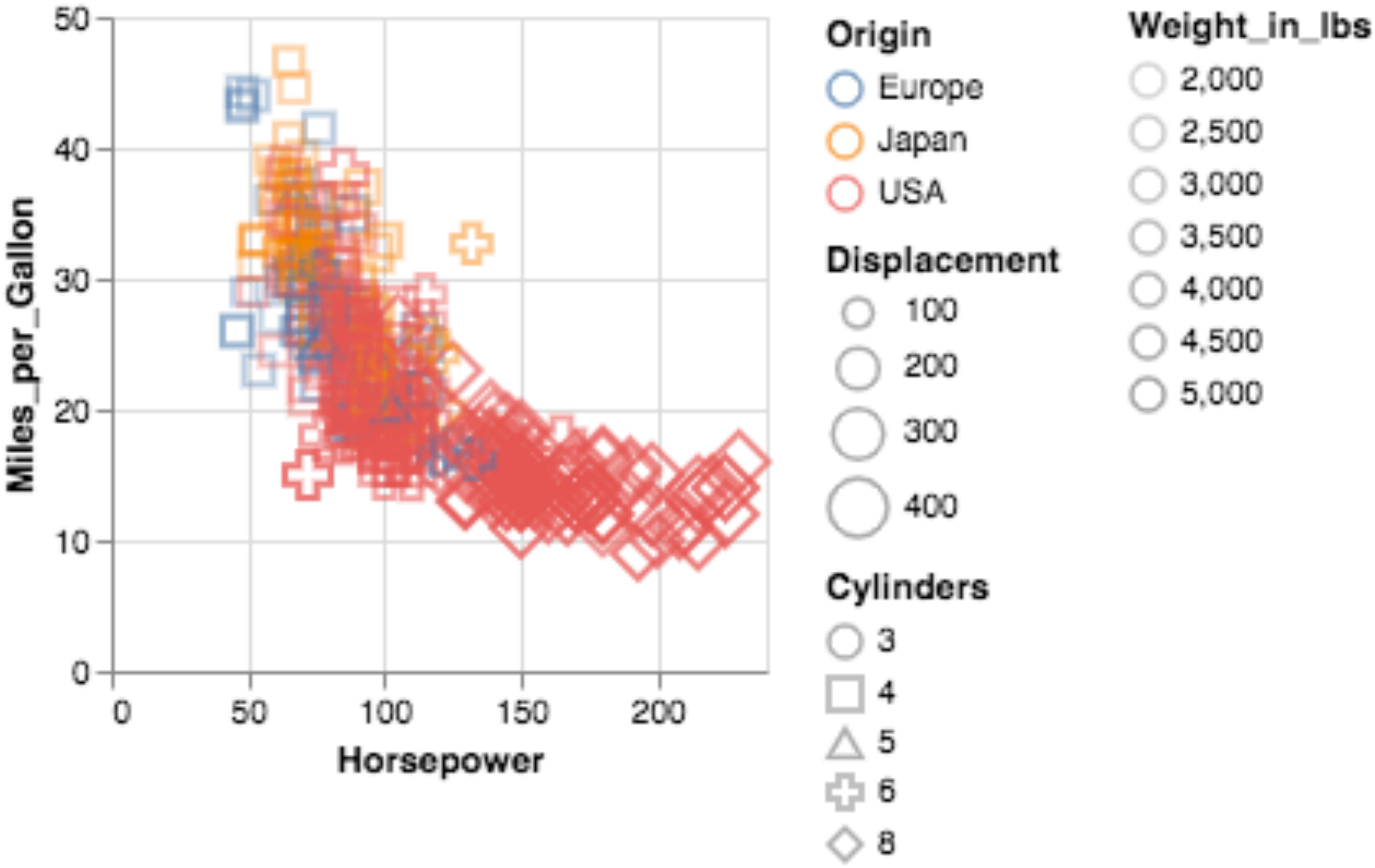
5-Dimensions

Next Up:
Weight (Q)



6-Dimensions

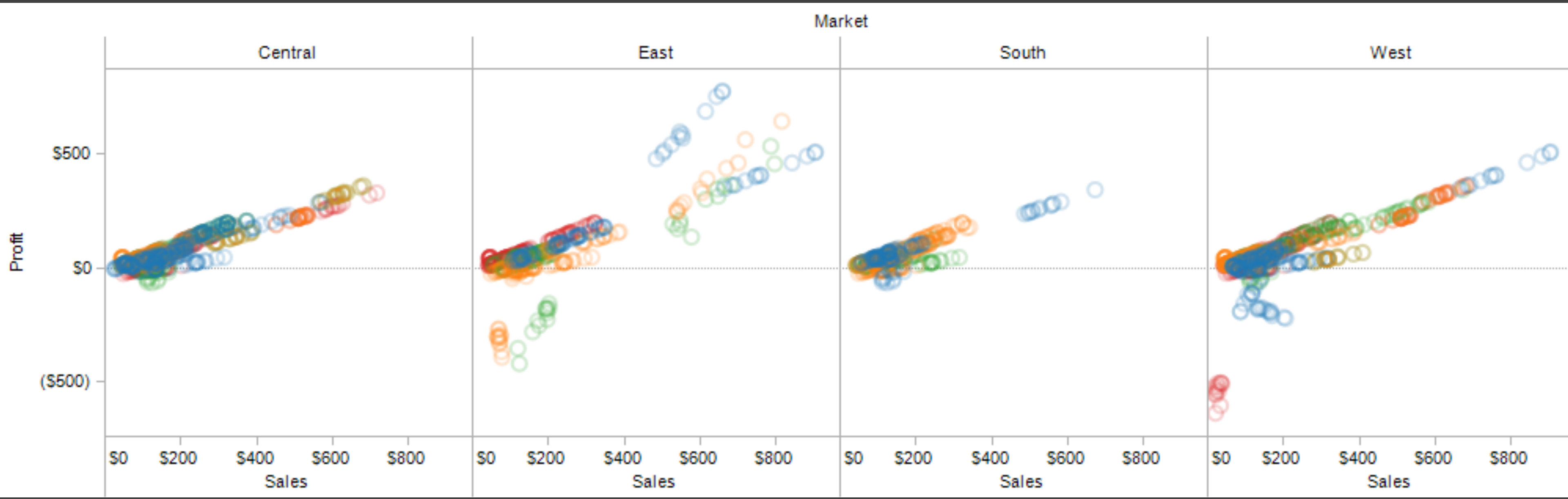
Next Up: Acceleration (Q)



7-Dimensions?

?

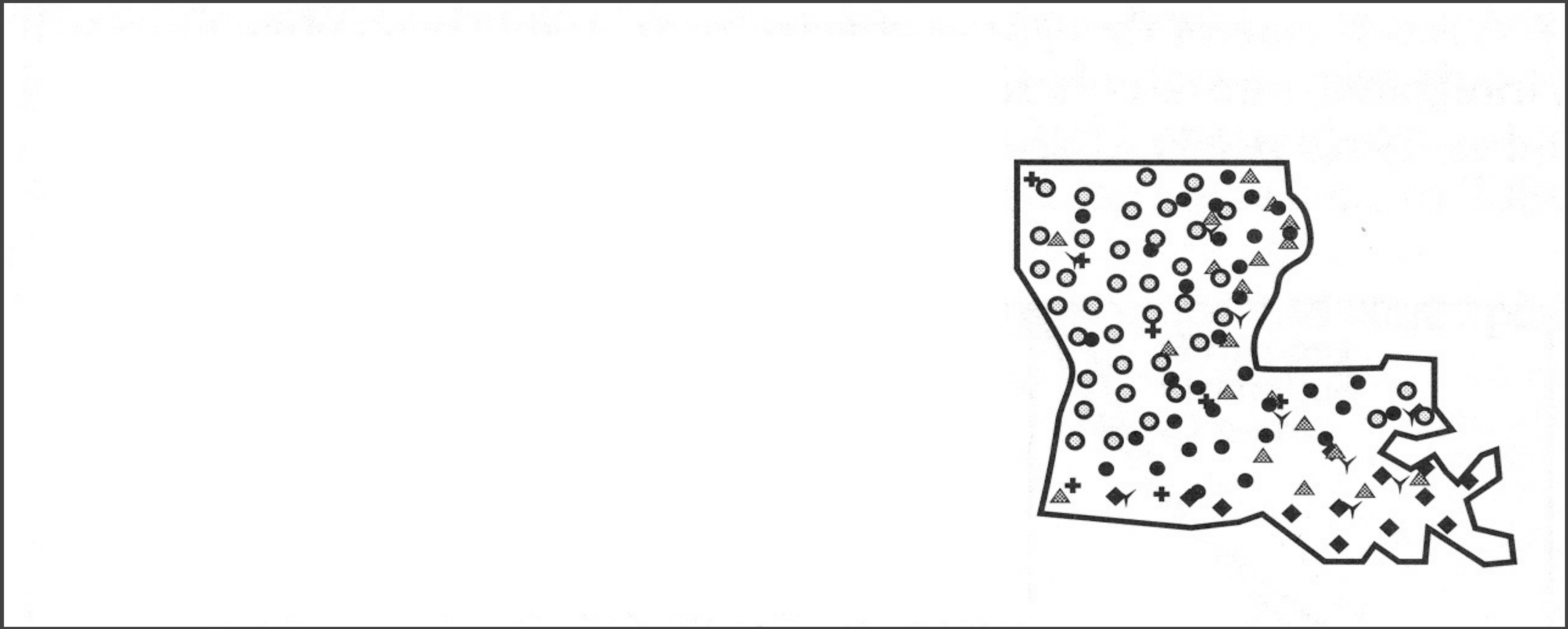
Trellis Plots



A *trellis plot* subdivides space to enable comparison across multiple plots.

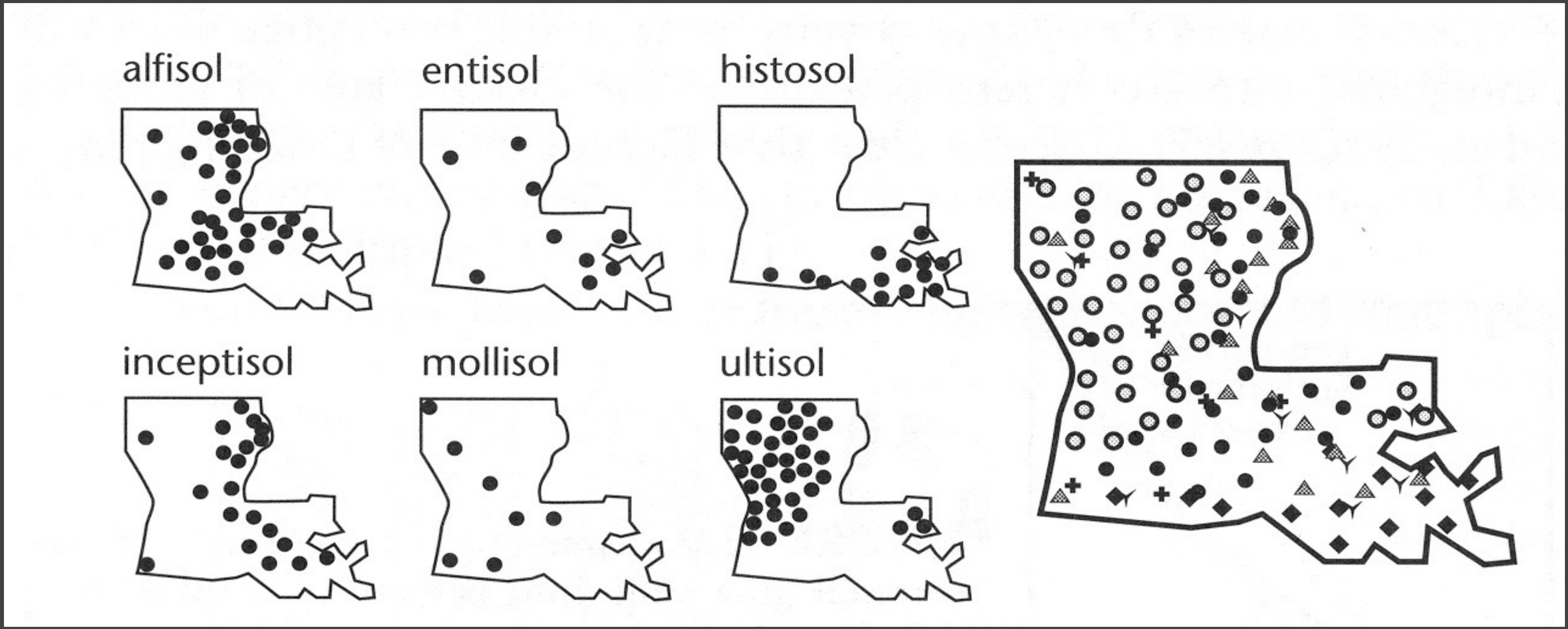
Typically nominal or ordinal variables are used as dimensions for subdivision.

Small Multiples



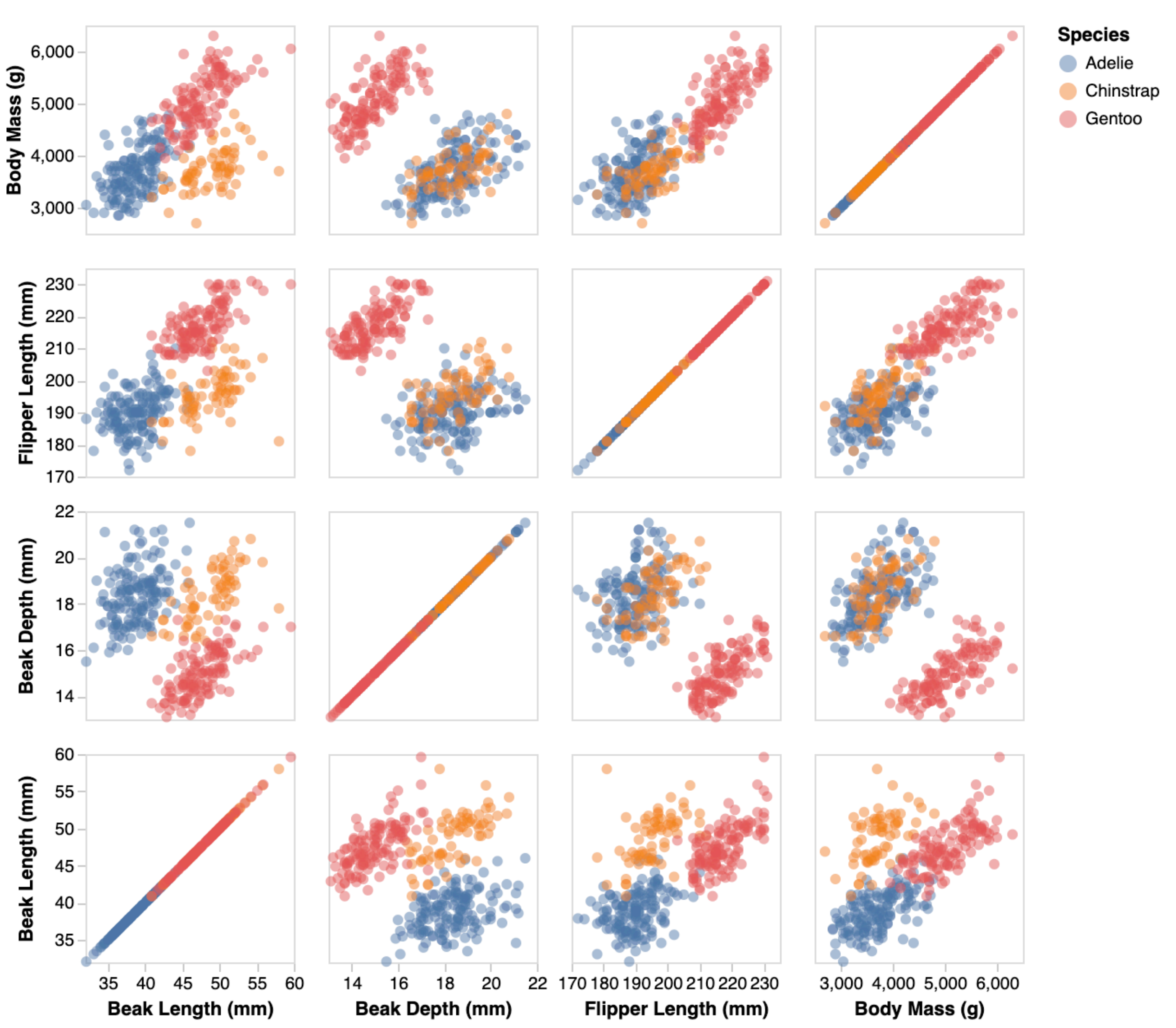
[MacEachren ’95, Figure 2.11, p. 38]

Small Multiples



[MacEachren '95, Figure 2.11, p. 38]

Scatterplot Matrix (SPLOM)



Scatter plots for pairwise comparison of each data dimension.

Multiple Coordinated Views

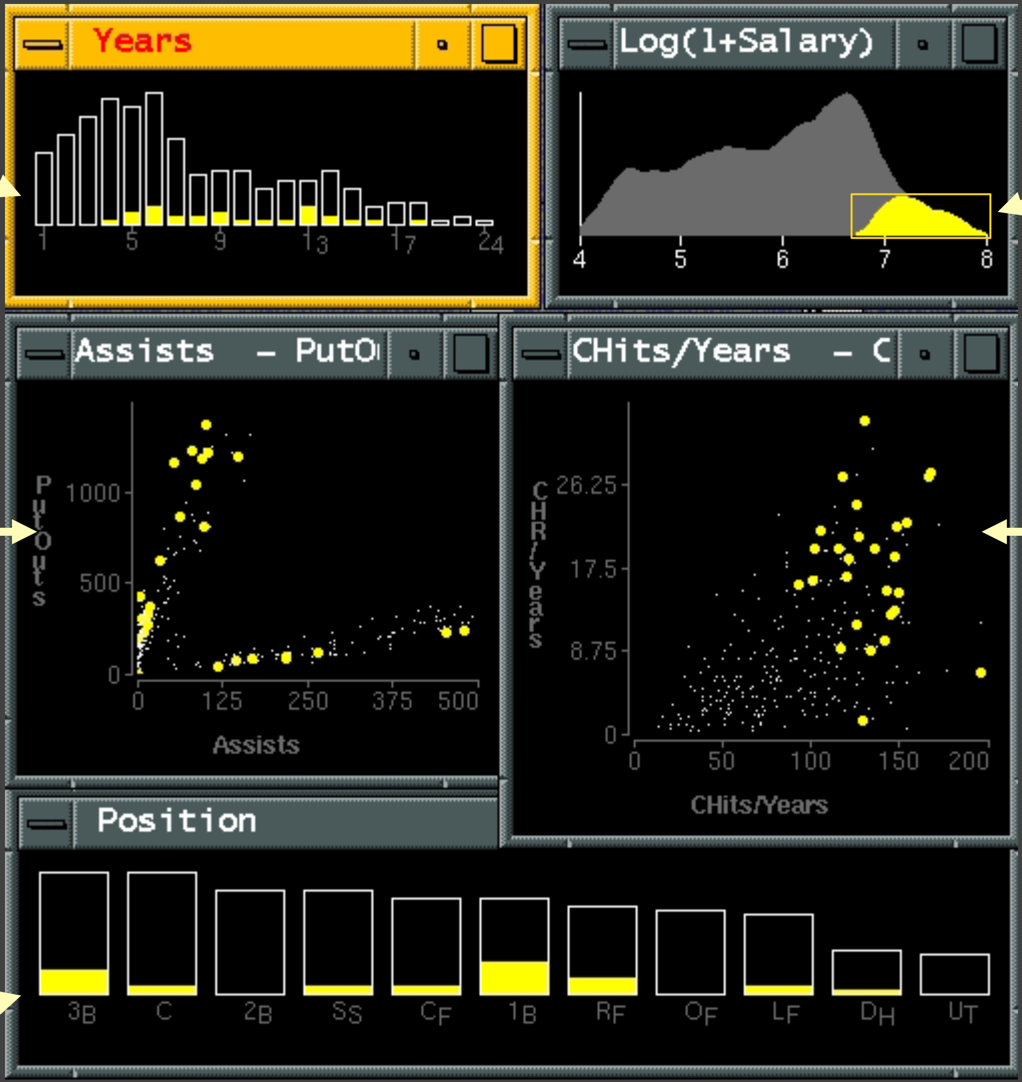
how long in majors

select high salaries

avg assists vs avg putouts (fielding ability)

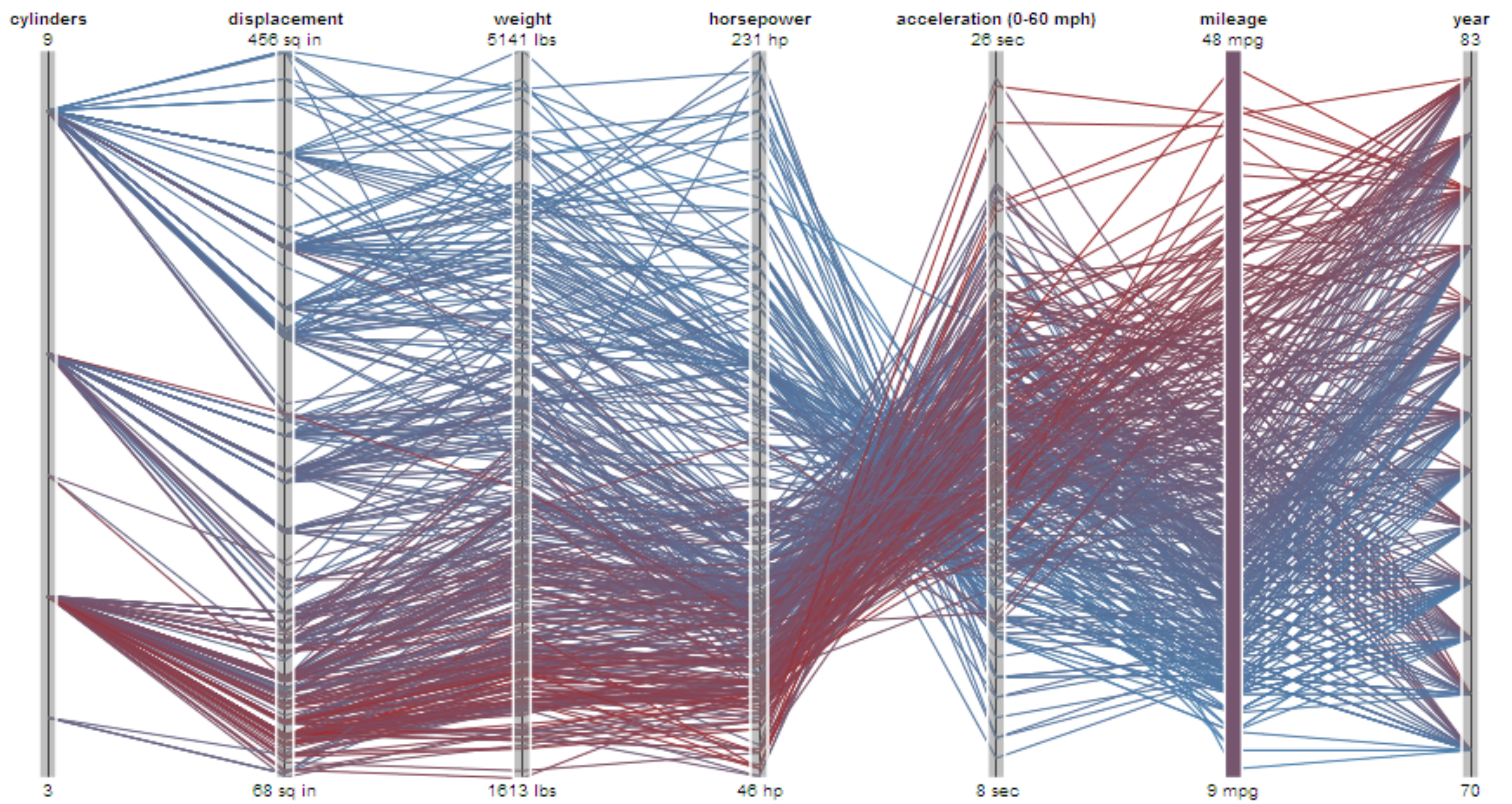
avg career HRs vs avg career hits (batting ability)

distribution of positions played



Parallel Coordinates

Parallel Coordinates [Inselberg]



Parallel Coordinates [Inselberg]

Visualize up to ~two dozen dimensions at once

1. Draw parallel axes for each variable
2. For each tuple, connect points on each axis

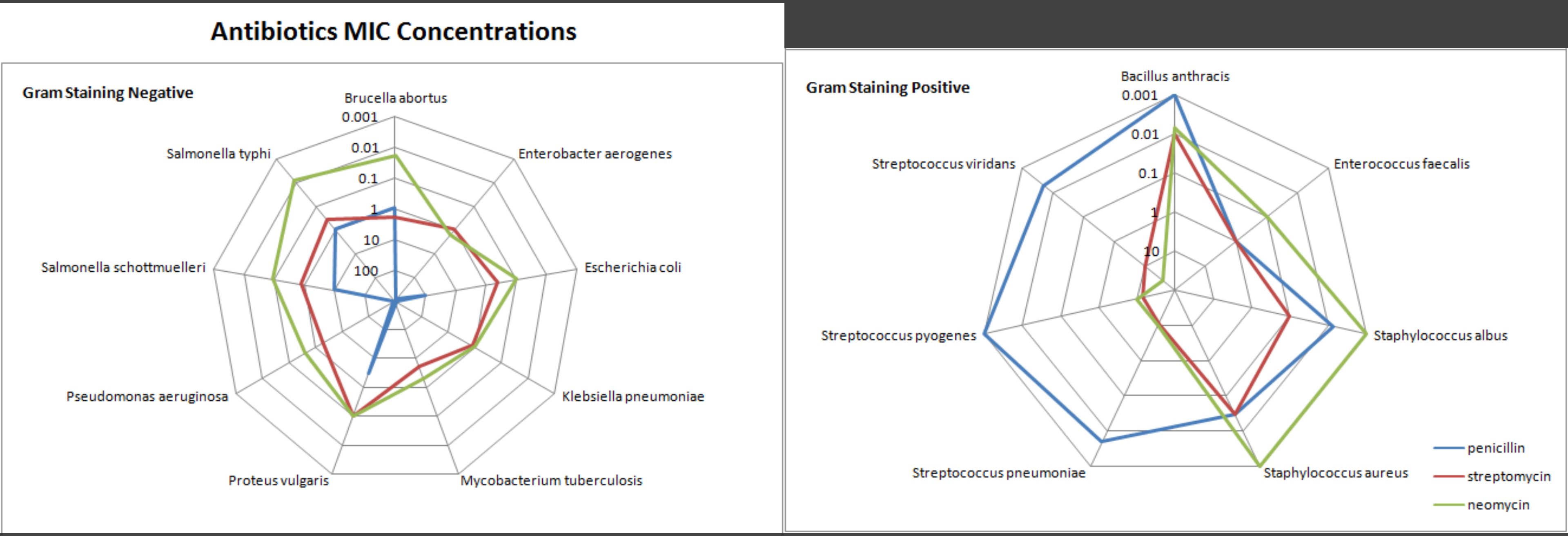
Between adjacent axes: line crossings imply neg. correlation, shared slopes imply pos. correlation.

Full plot can be cluttered. **Interactive selection** can be used to assess multivariate relationships.

Highly sensitive to axis **scale** and **ordering**.

Expertise required to use effectively!

Radar Plot / Star Graph



"Parallel" dimensions in polar coordinate space

Best if same units apply to each axis

[Loren Yu, CS448B 2009]

Dimensionality Reduction

Dimensionality Reduction (DR)

Project nD data to 2D or 3D for viewing. Often used to interpret and sanity check high-dimensional representations fit by machine learning methods.

Different DR methods make different trade-offs: for example to **preserve global structure** (e.g., PCA) or **emphasize local structure** (e.g., nearest-neighbor approaches, including t-SNE and UMAP).

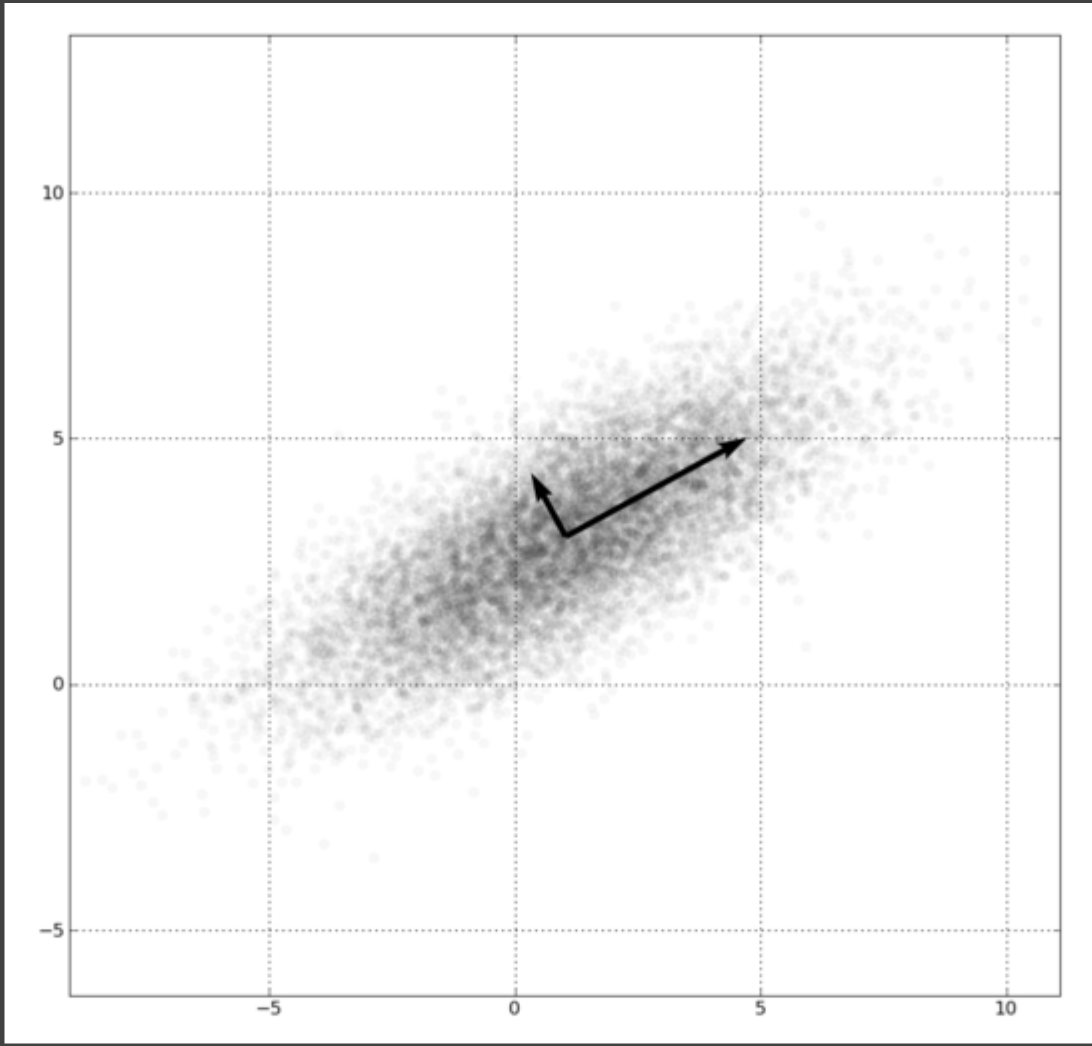
Reduction Techniques

Principal Components Analysis (PCA)
Linear transformation of basis vectors, ordered by amount of data variance they explain.

t-Dist. Stochastic Neighbor Embedding (t-SNE)
Probabilistically model distance, optimize positions.

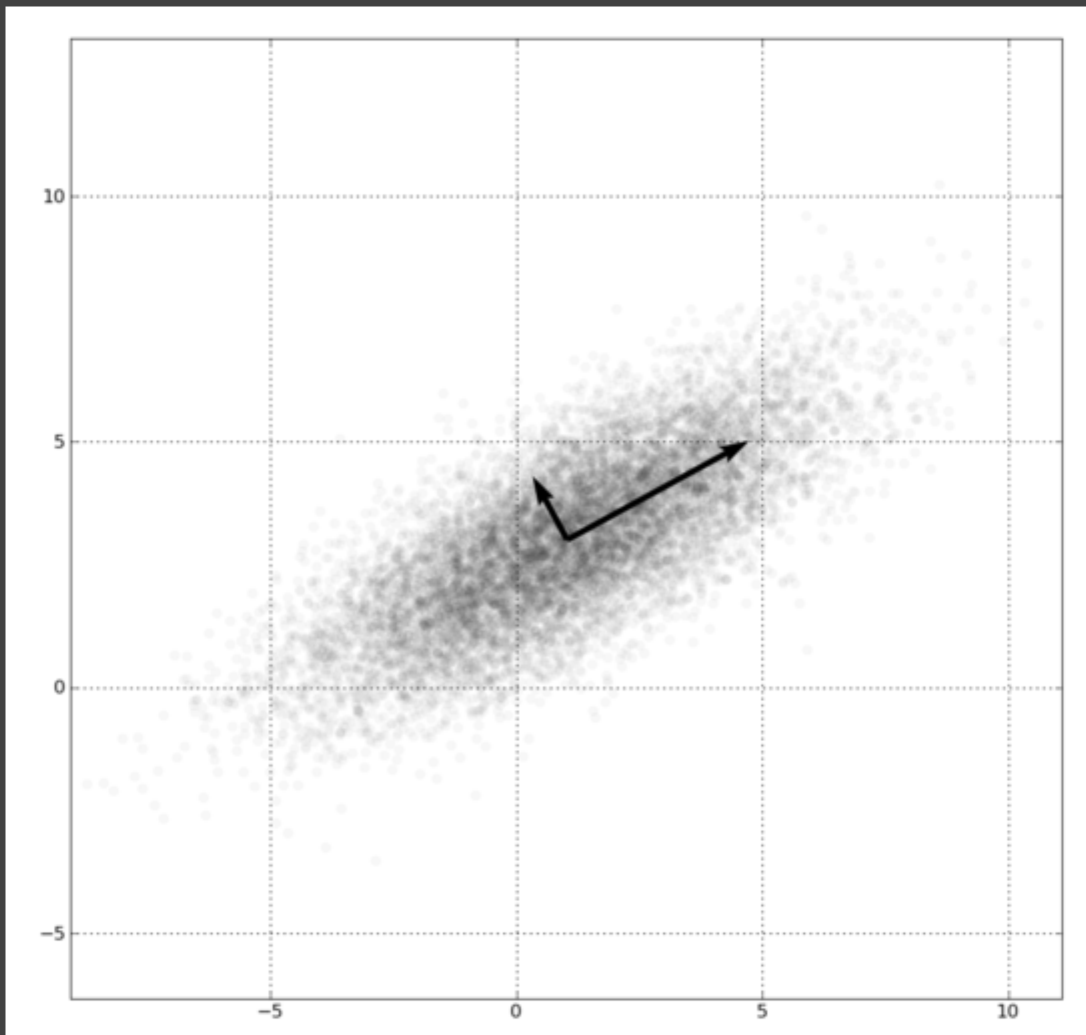
Uniform Manifold Approx. & Projection (UMAP)
Identify local manifolds, then stitch them together.

Principal Components Analysis



1. Mean-center the data.
2. Find ⊥ basis vectors that maximize the data variance.
3. Plot the data using the top vectors.

Principal Components Analysis

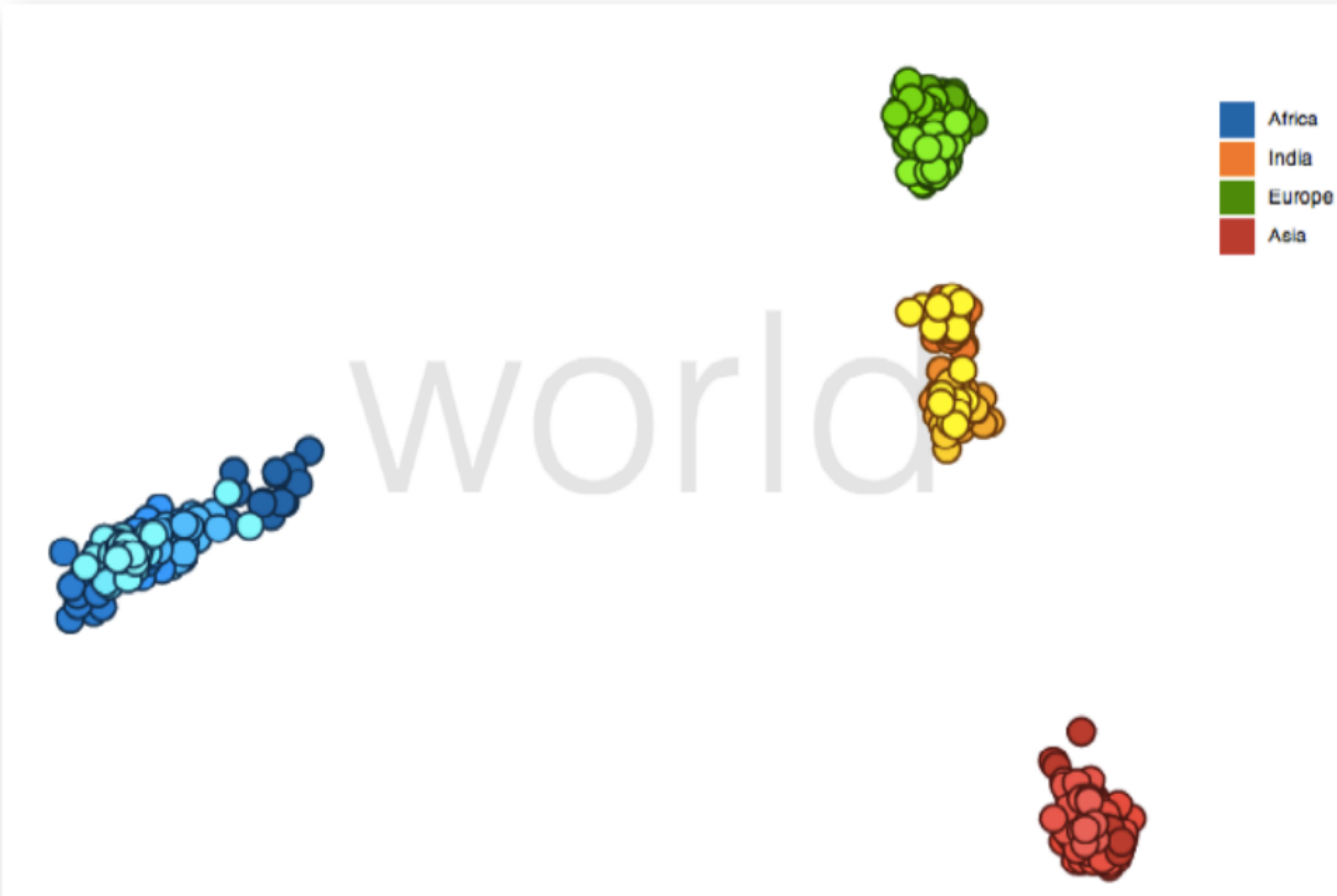


Linear transform: scale and rotate original space.

Lines (vectors) project to lines.

Preserves global distances.

PCA of Genomes [Demiralp et al. '13]



Non-Linear Techniques

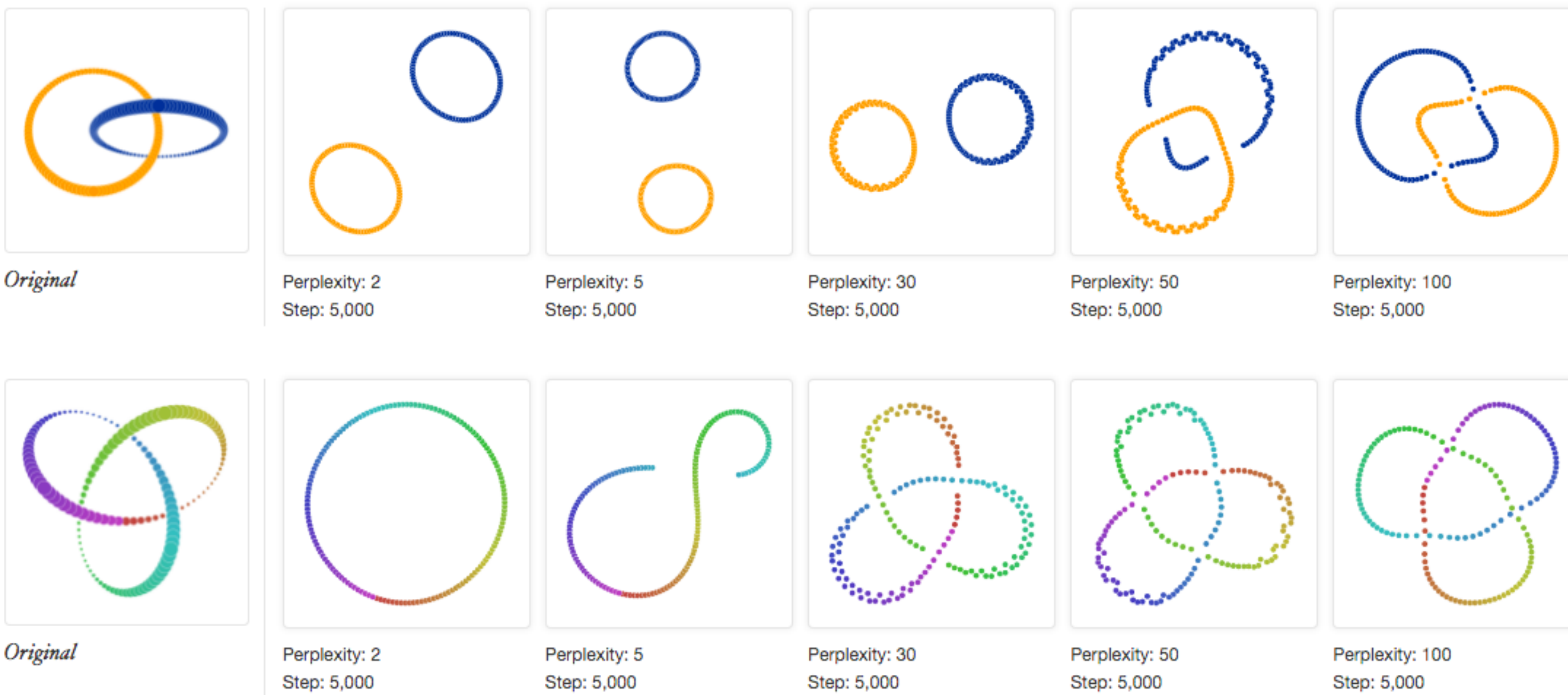
Distort the space, trade-off preservation of global structure to emphasize local neighborhoods. Use topological (nearest neighbor) analysis.

Two popular contemporary methods:

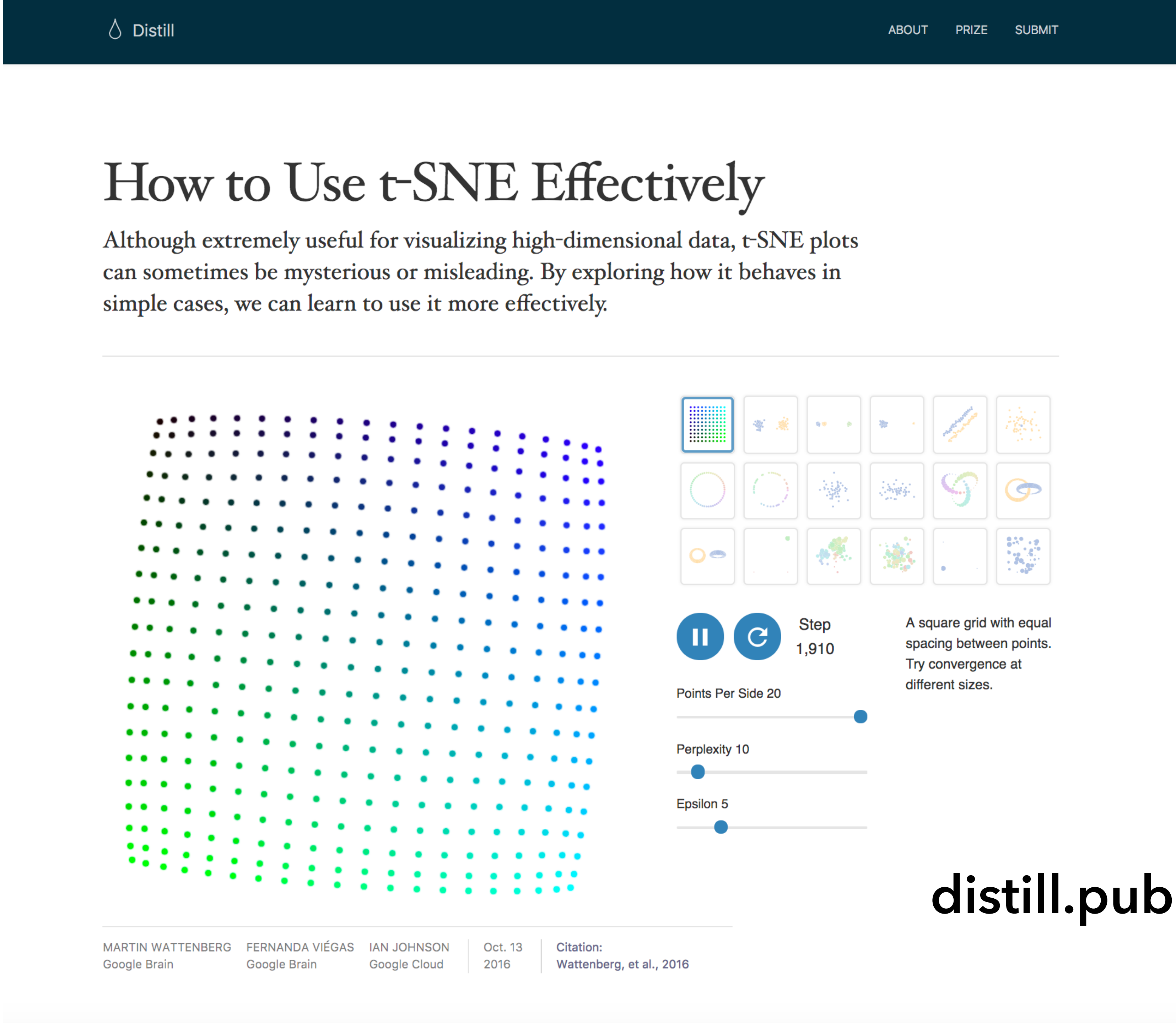
t-SNE - probabilistic interpretation of distance

UMAP - tries to balance local/global trade-off

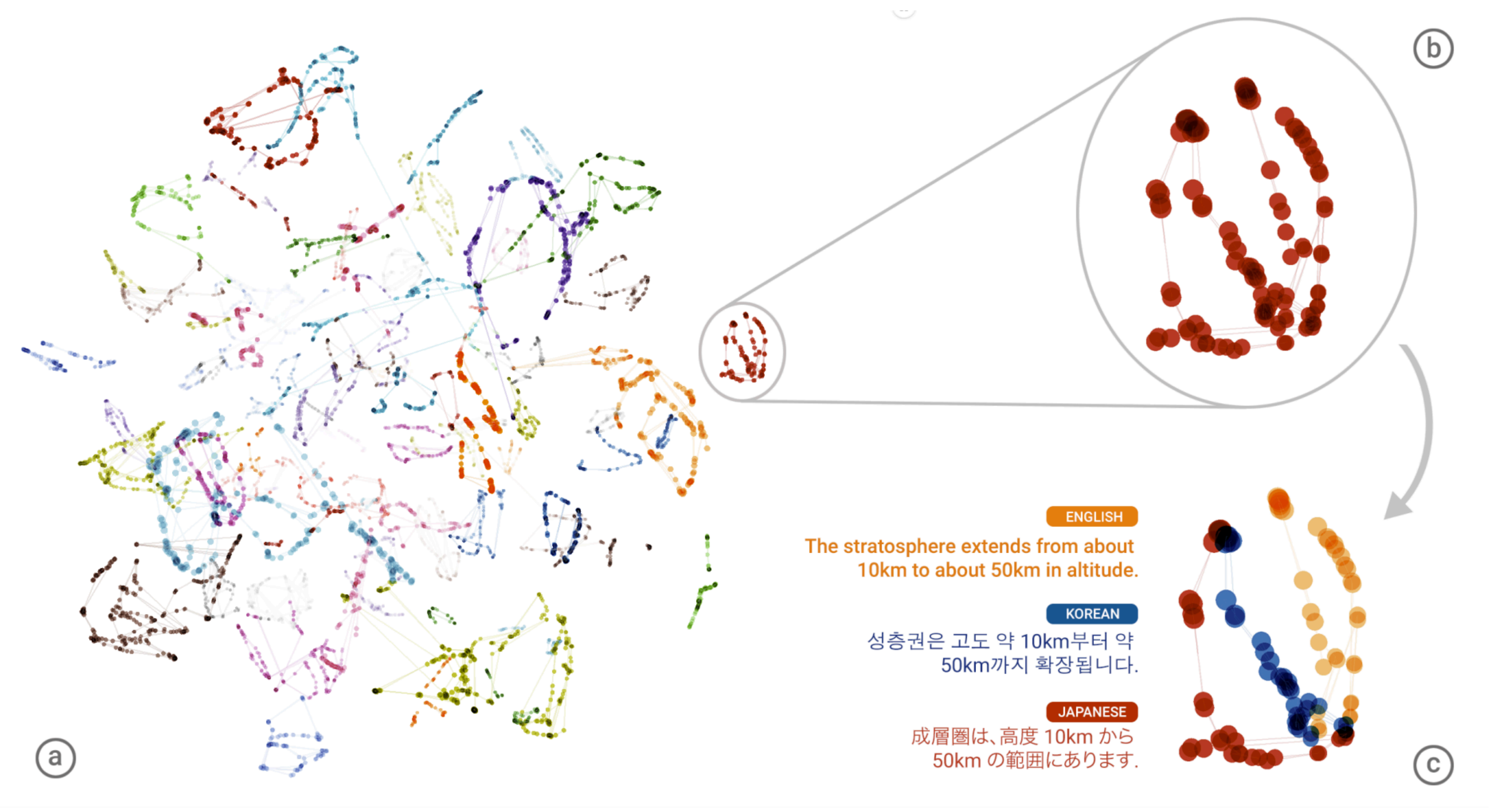
Visualizing t-SNE [Wattenberg et al. '16]



Results can be highly sensitive to the algorithm parameters!

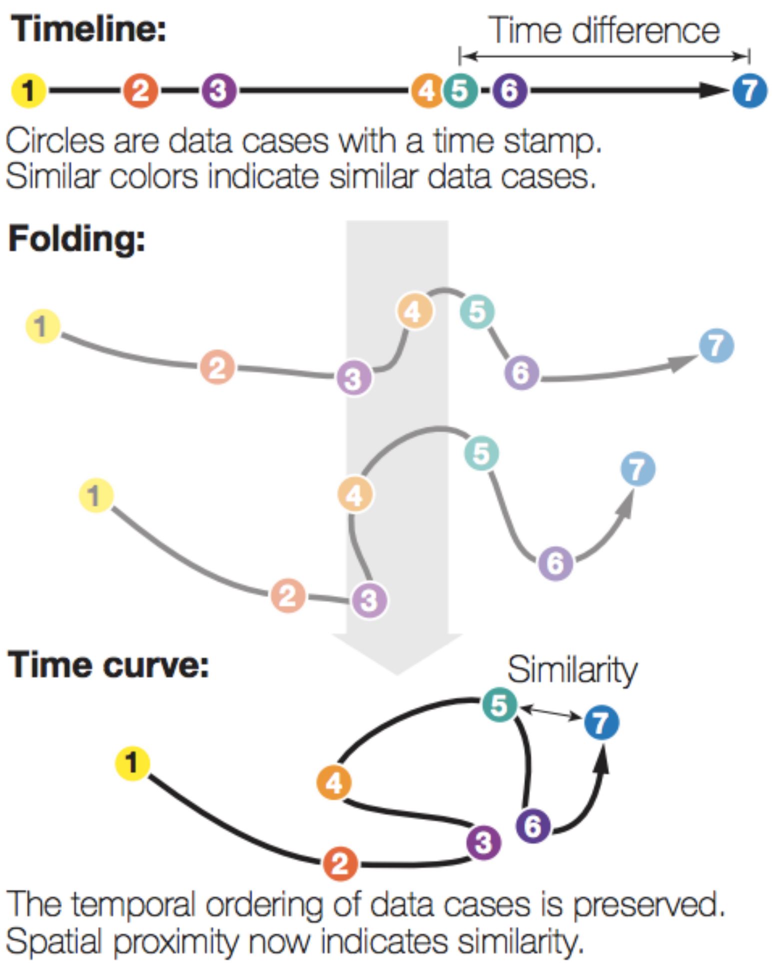

Distill
ABOUT
PRIZE
SUBMIT
How to Use t-SNE Effectively
Although extremely useful for visualizing high-dimensional data, t-SNE plots can sometimes be mysterious or misleading. By exploring how it behaves in simple cases, we can learn to use it more effectively.
Step
1,910
A square grid with equal spacing between points. Try convergence at different sizes.
Points Per Side 20
Perplexity 10
Epsilon 5
MARTIN WATTENBERG
Google Brain
FERNANDA VIÉGAS
Google Brain
IAN JOHNSON
Google Cloud
Oct. 13
2016
Citation:
Wattenberg, et al., 2016
distill.pub

MT Embedding [Johnson et al. 2018]

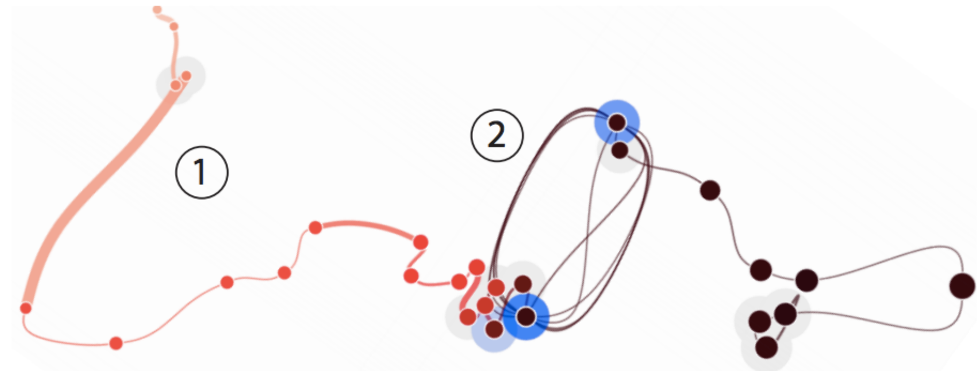


t-SNE projection of latent space of language translation model.

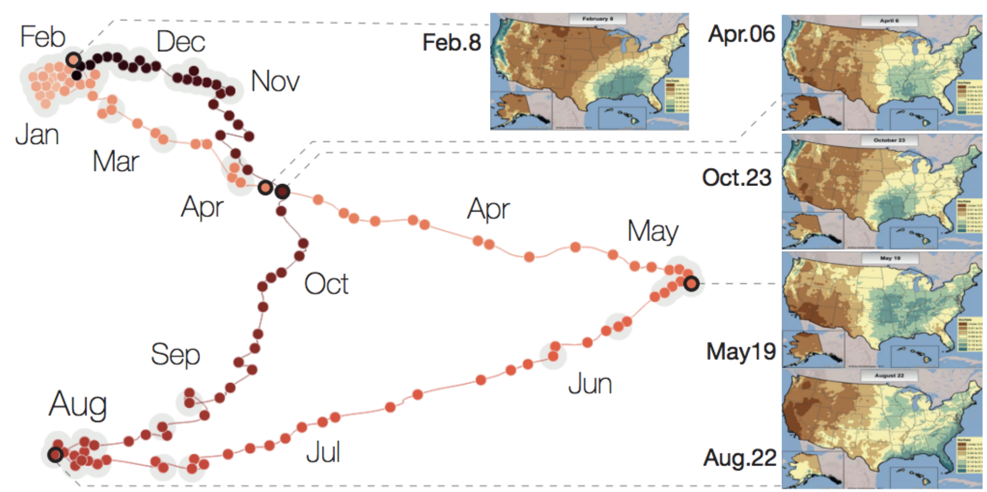
Time Curves [Bach et al. '16]



(a) Folding time



Wikipedia “Chocolate” Article



U.S. Precipitation over 1 Year

Administrivia

Final Project

Initial Project Prototype due **tonight Feb. 26th**
Prototype Deliverables: must **submit link on Canvas**

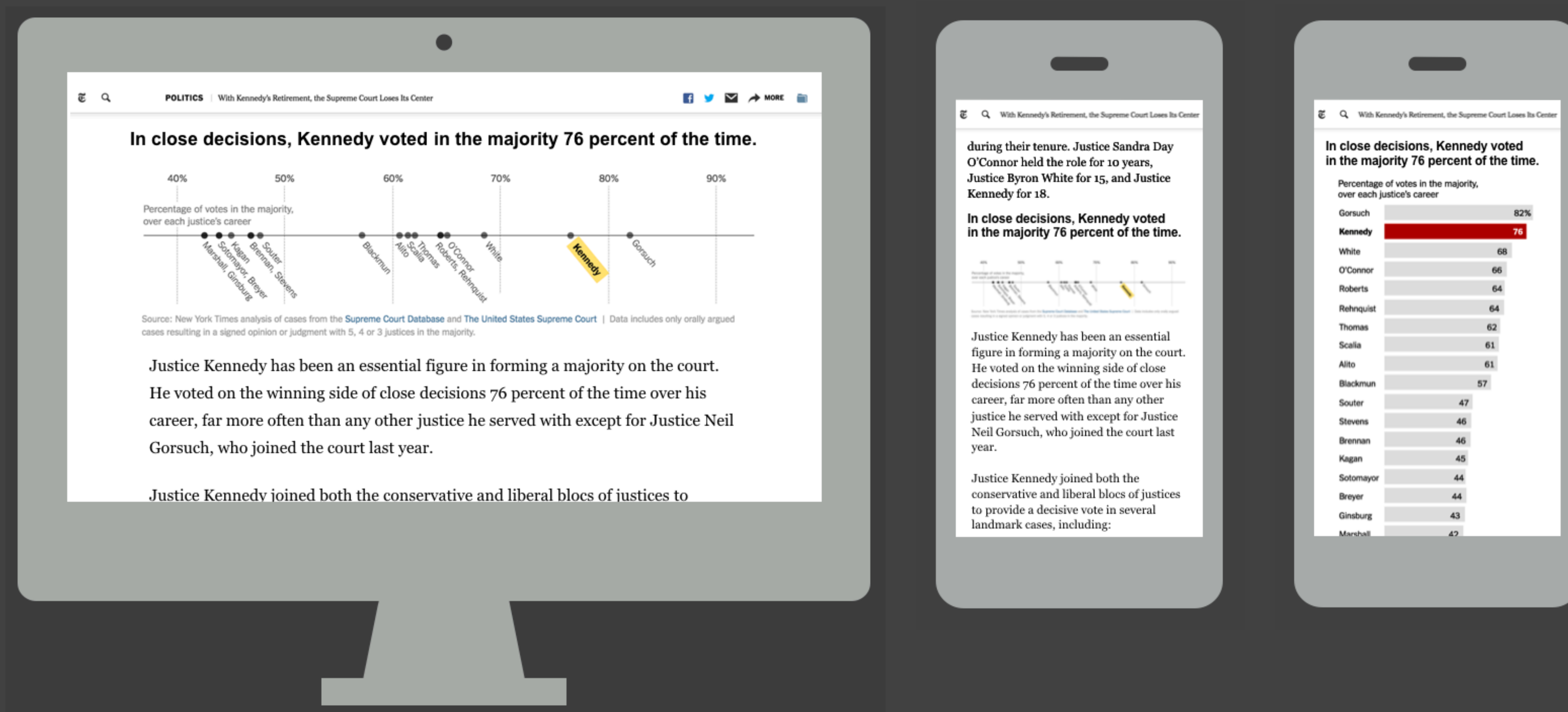
Prototype Expectations:
Outline of the overall project structure
Rough prototypes of visualizations and interactions
Basic descriptive (narrative) text
Discussion of any concerns or plans for next steps

The more content you have on your page, the more specific feedback we can give to refine your project.

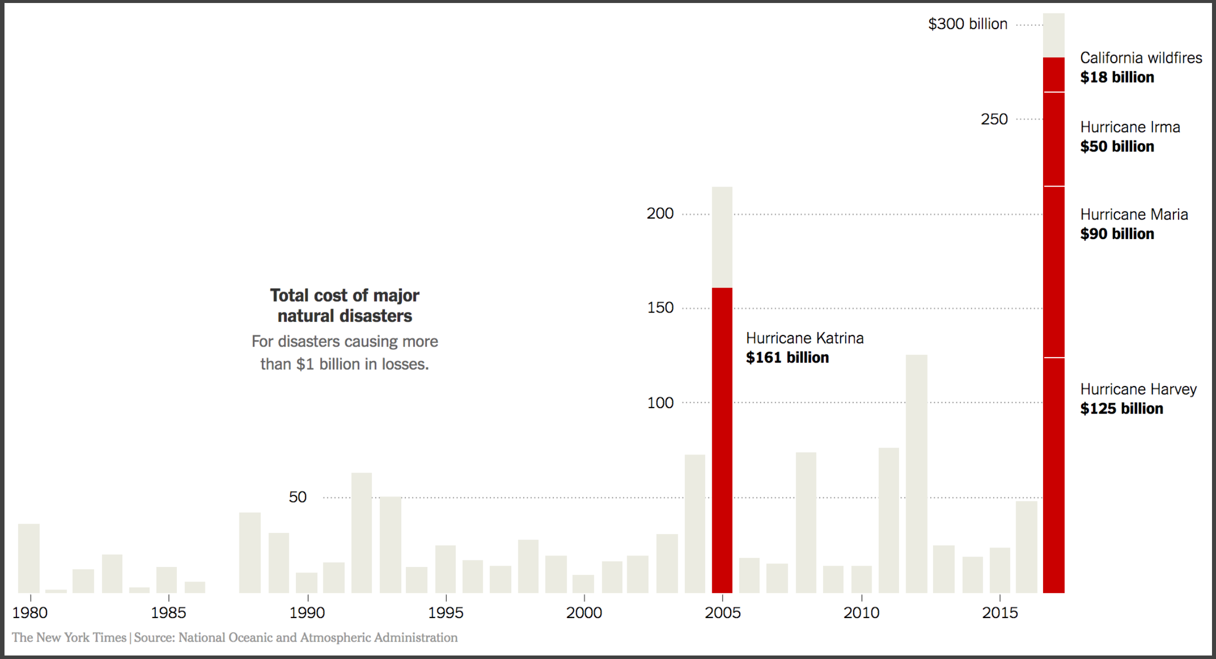
BONUS TOPIC

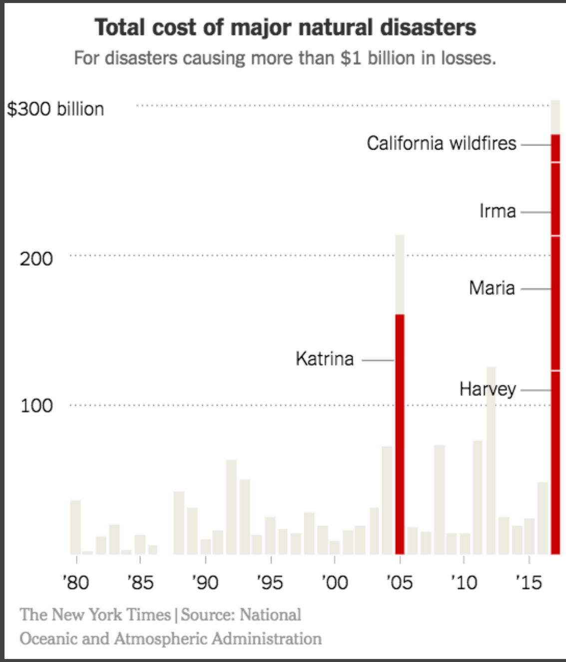
Responsive Visualization

Responsive Visualization

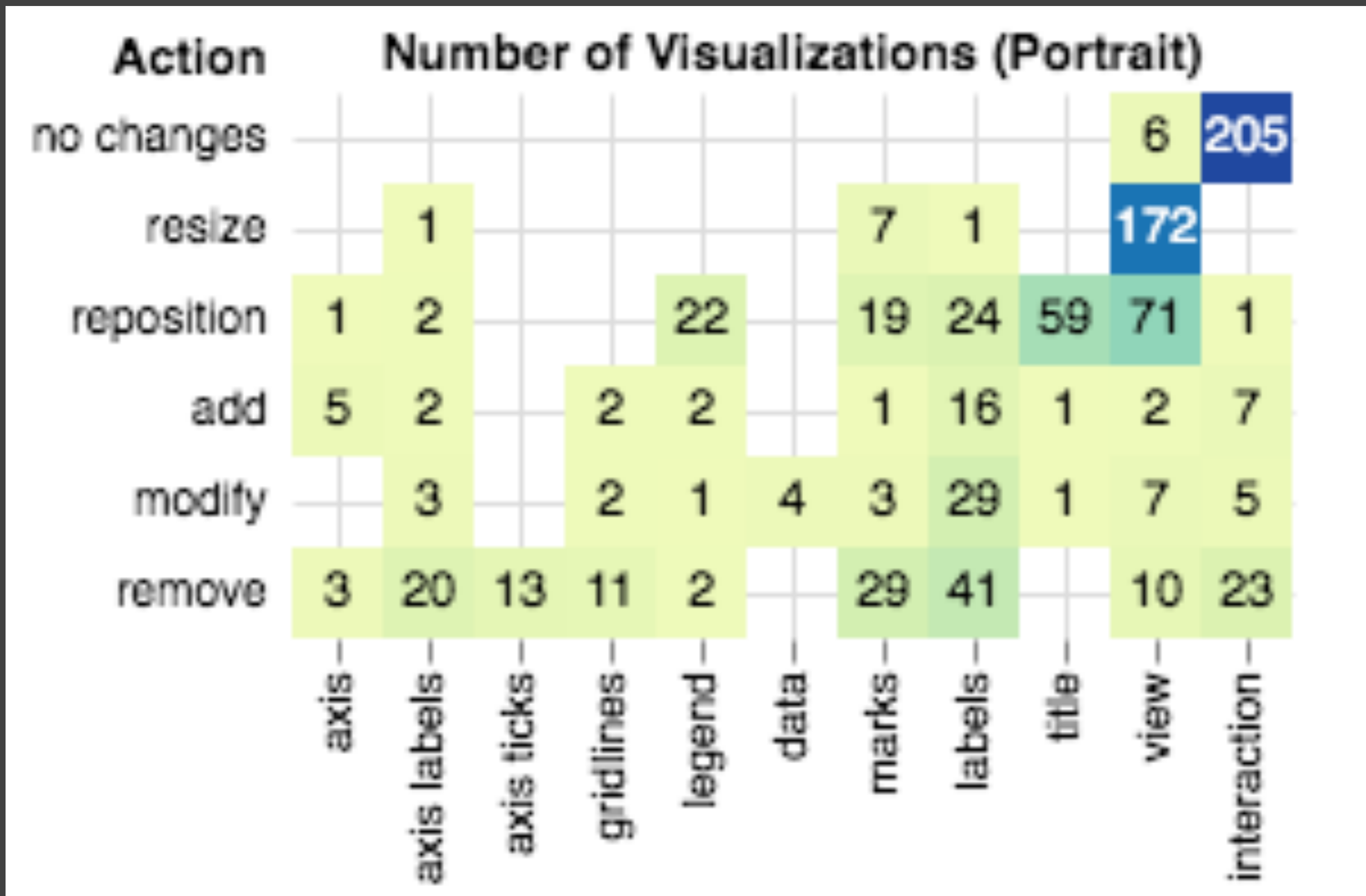


Responsive Visualization [Hoffswell et al. 20]





Responsive Visualization [Hoffswell et al. 20]



Responsive Visualization [Hoffswell et al. 20]

Basic Selection Methods

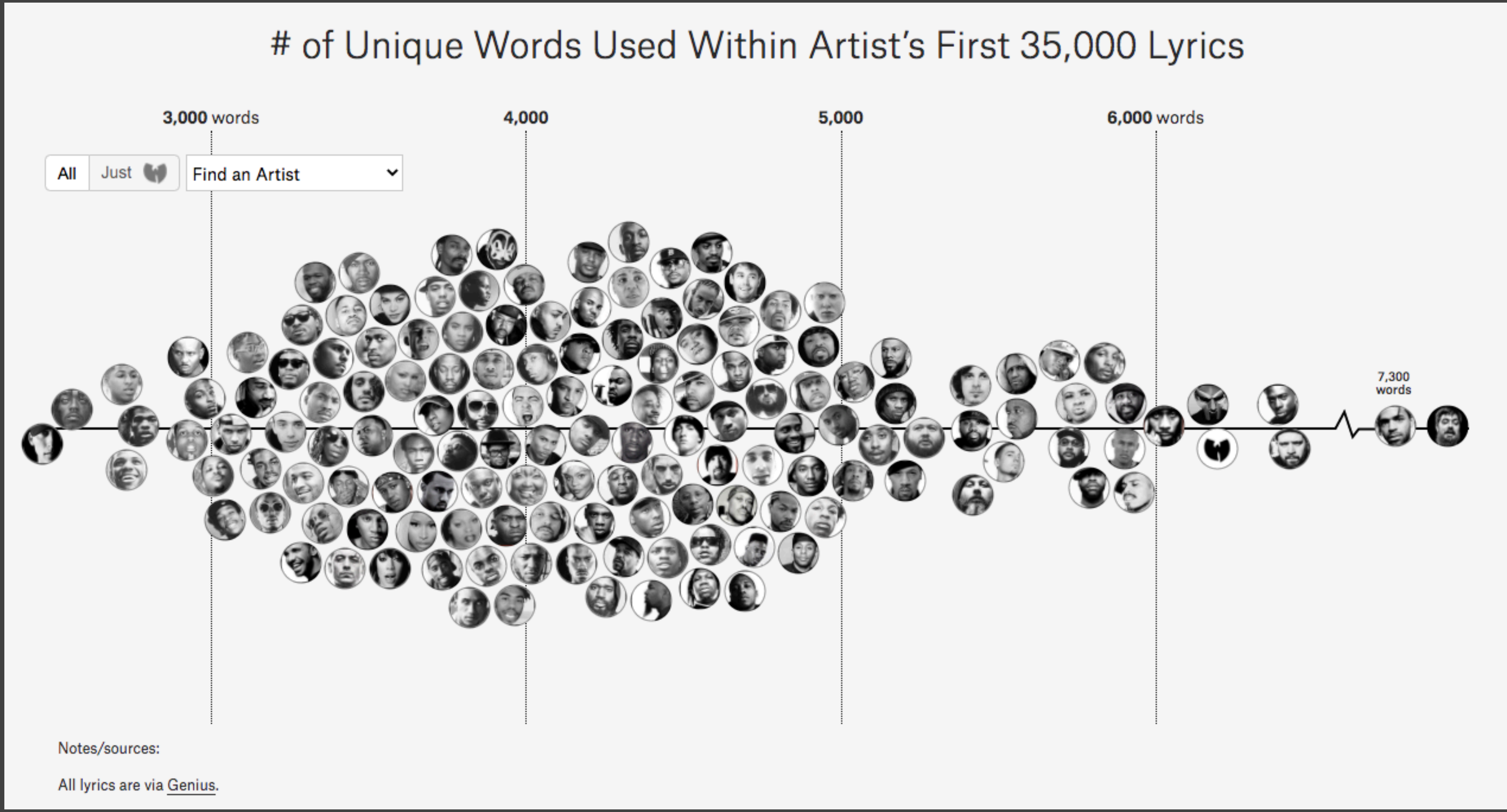
Point Selection

Mouse Hover / Click

Touch / Tap

Select Nearby Element (e.g., Bubble Cursor)

Desktop vs. Mobile Tooltips



Right now we have at least 50%, sometimes 60% or 70%… of our readers that come through mobile phones to our site.

Gregor Aisch, *Information+ Conference 2016*

Right now we have at least 50%, sometimes 60% or 70%... of our readers that come through mobile phones to our site… Nobody is interacting with news graphics… it’s like 10% of all users click that button.

Gregor Aisch, *Information+ Conference 2016*

82% of mobile readers advanced through at least some of the content, even though they needed to dismiss a warning about download size; however, only 34% attempted to tune the guitar and just 6% tuned all six strings.

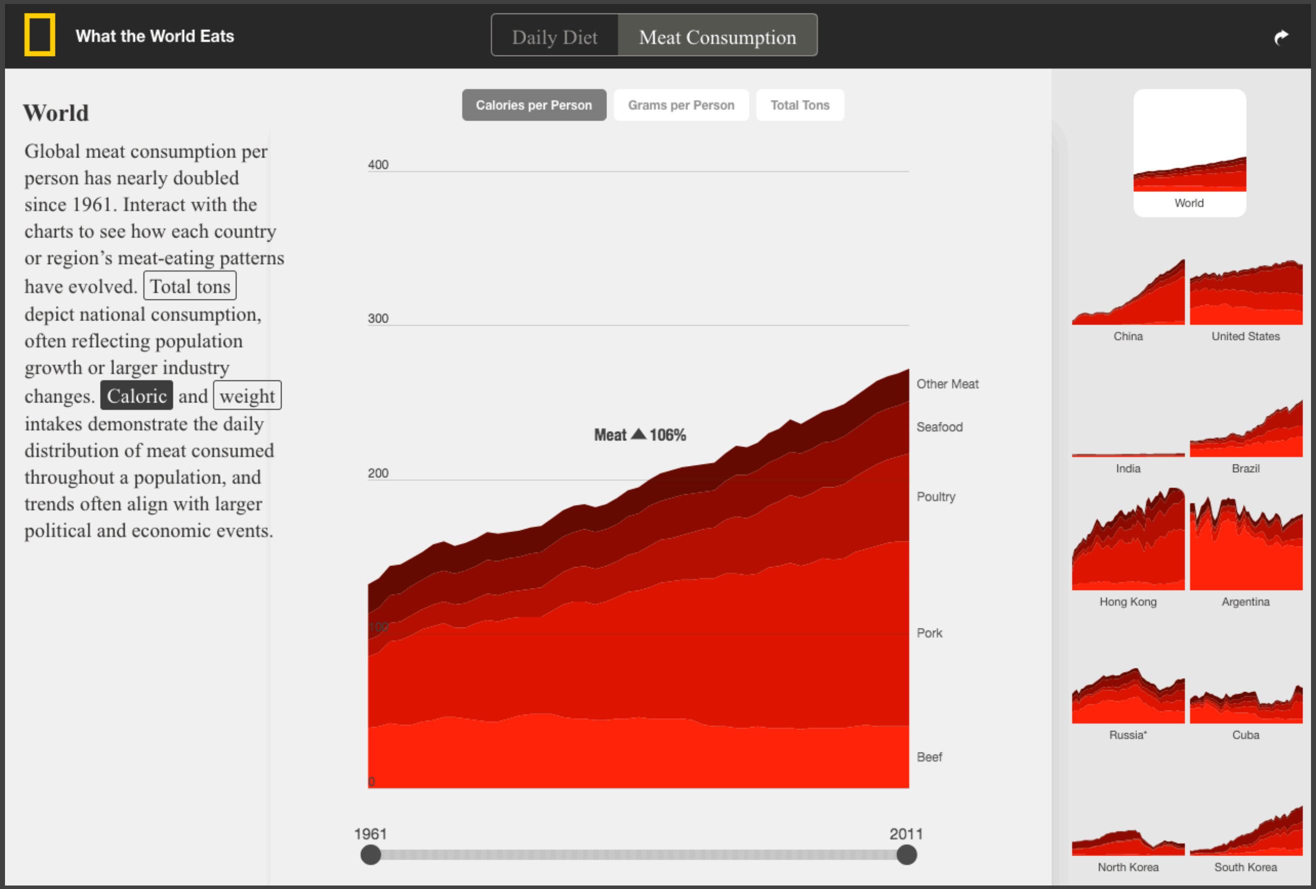
Conlen et al., *EuroVis 2019*

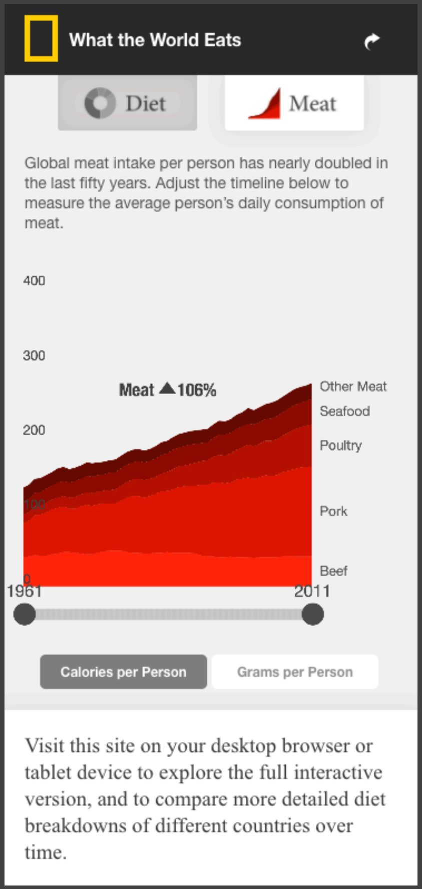
82% of mobile readers advanced through at least some of the content, even though they needed to dismiss a warning about download size; however, only 34% attempted to tune the guitar and just 6% tuned all six strings.

These observations suggest that mobile users are willing to engage with interactive content, and that the specific interactions should have been refined to better accommodate them.

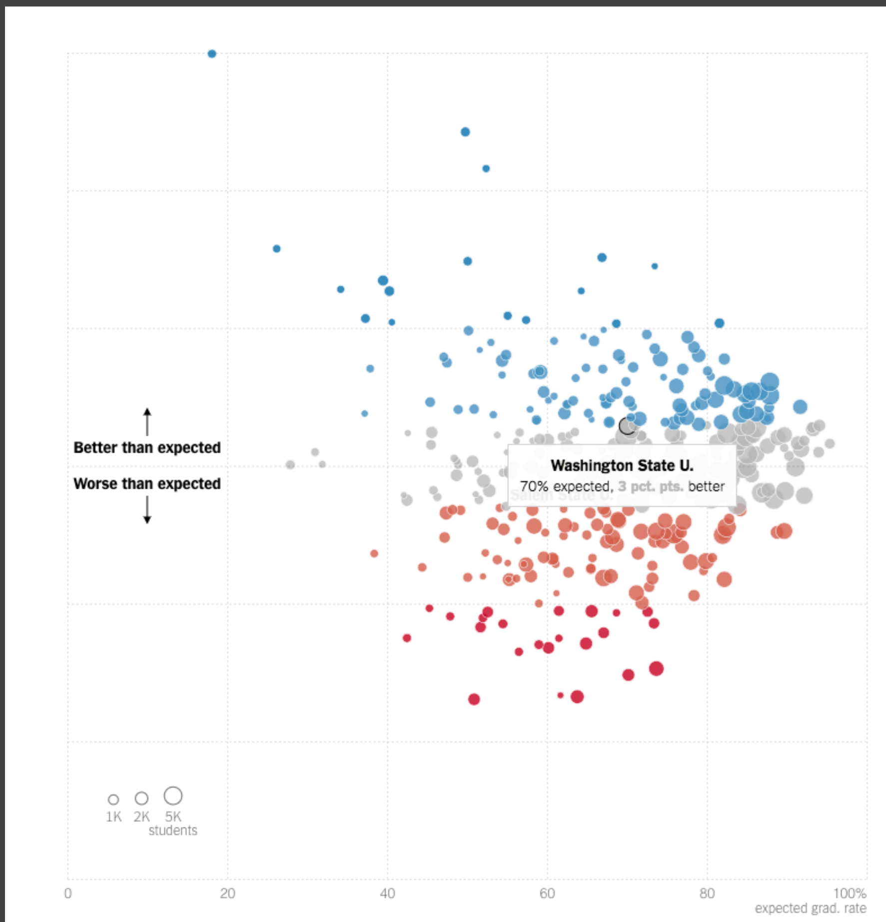
Conlen et al., *EuroVis 2019*

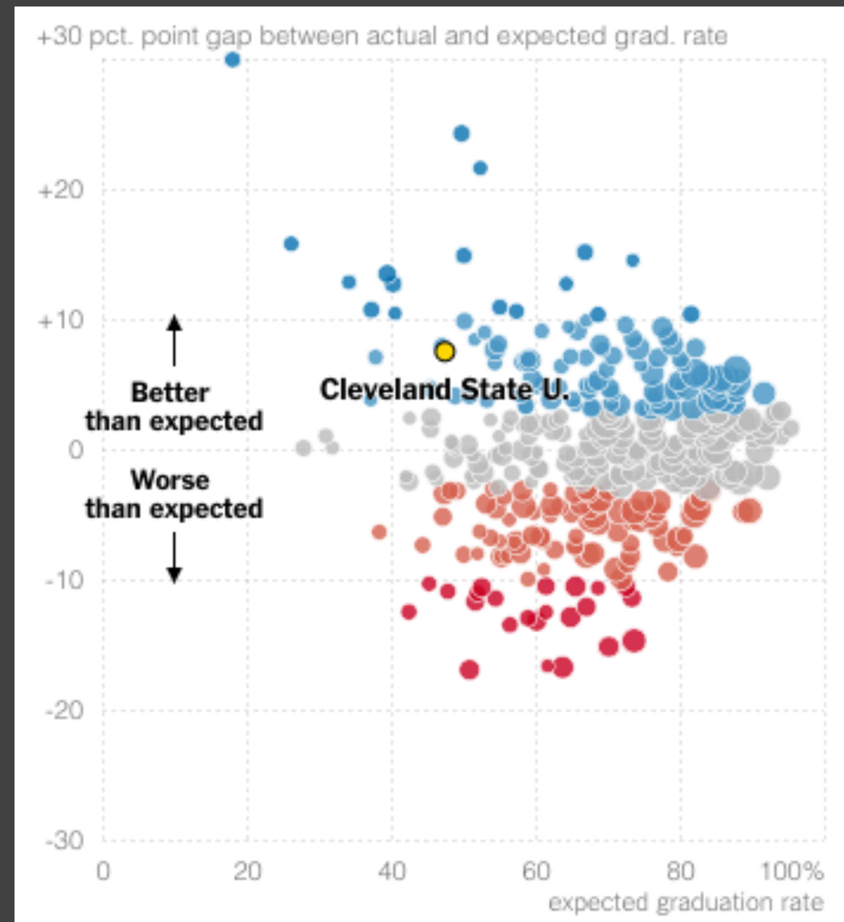
Interactions Disabled



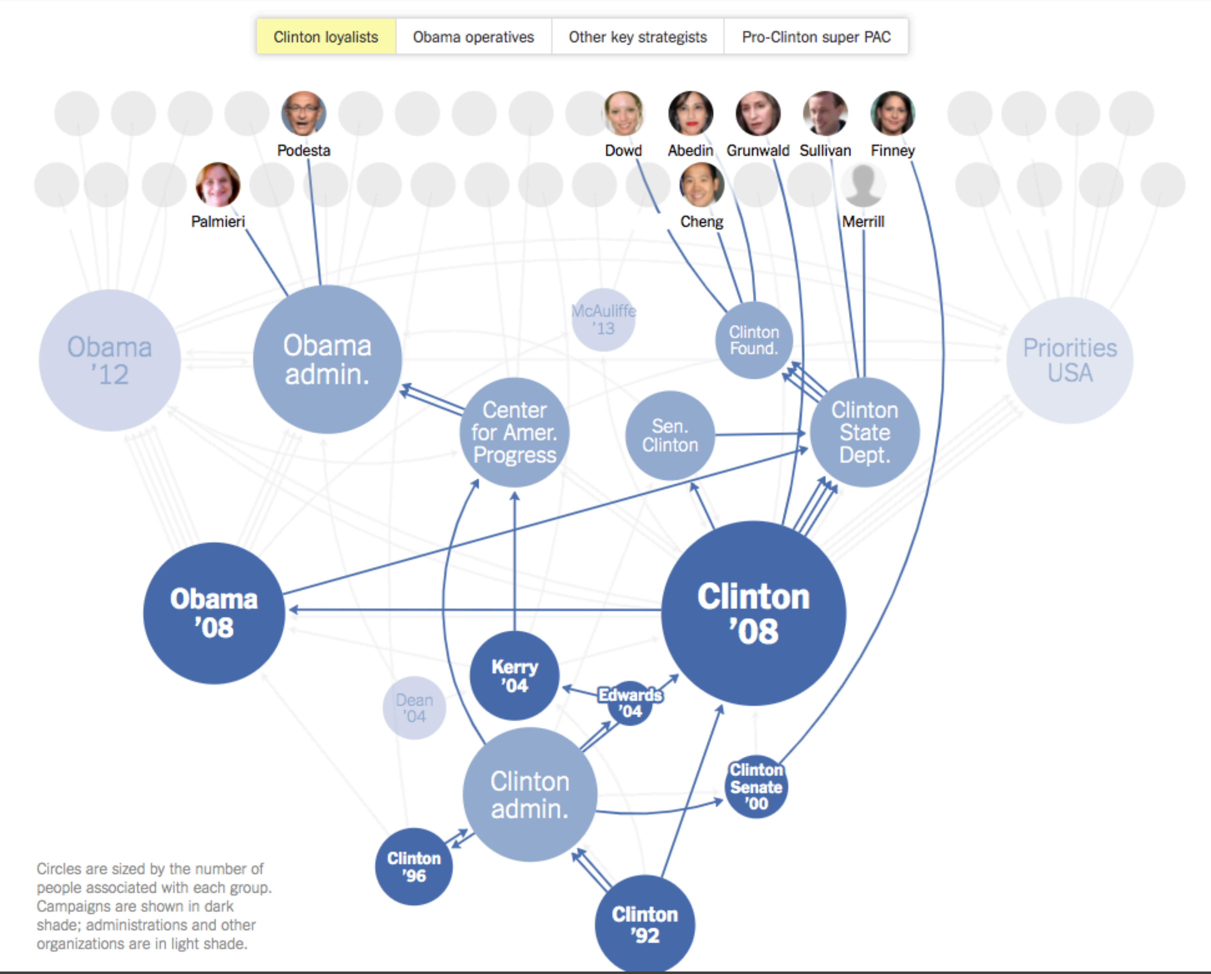


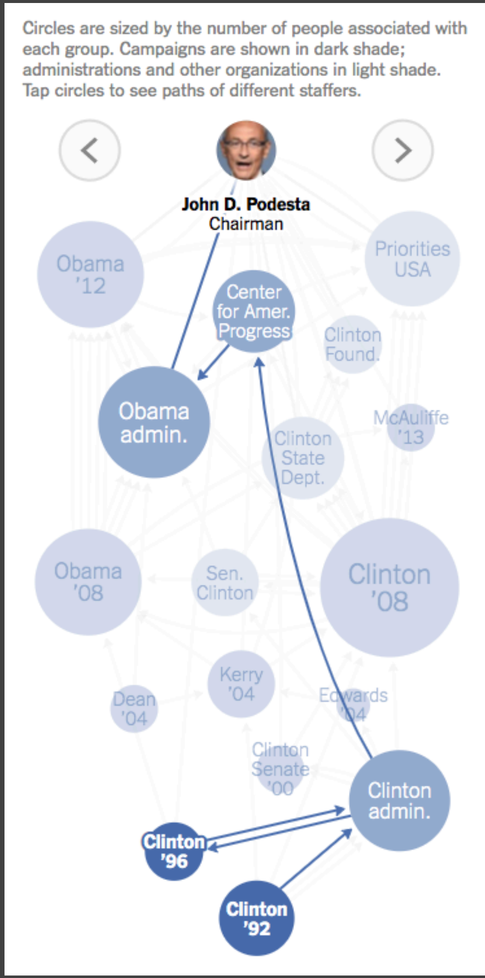
Interactions Previewed





Interactions Simplified





Responsive Visualization Summary

Good visualizations are task dependent

Who is the audience and what is the task?

Pick the right interaction technique

Visualizations are not one size fits all

Context might change user goals