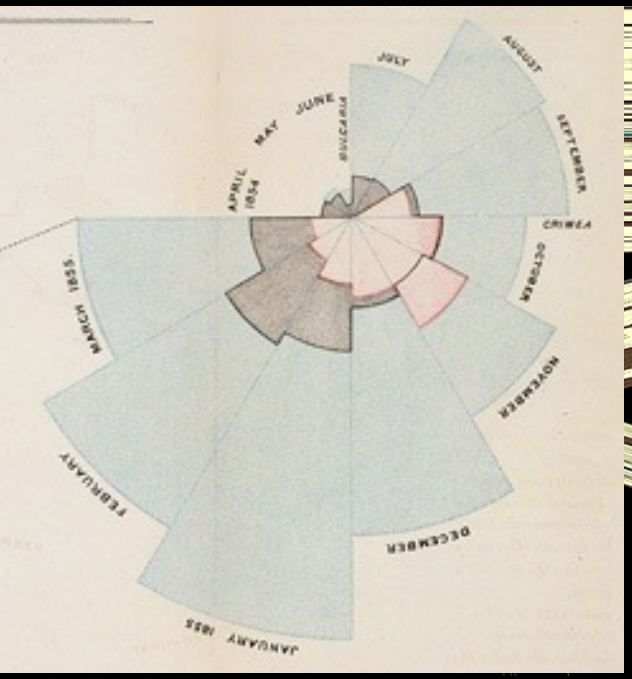
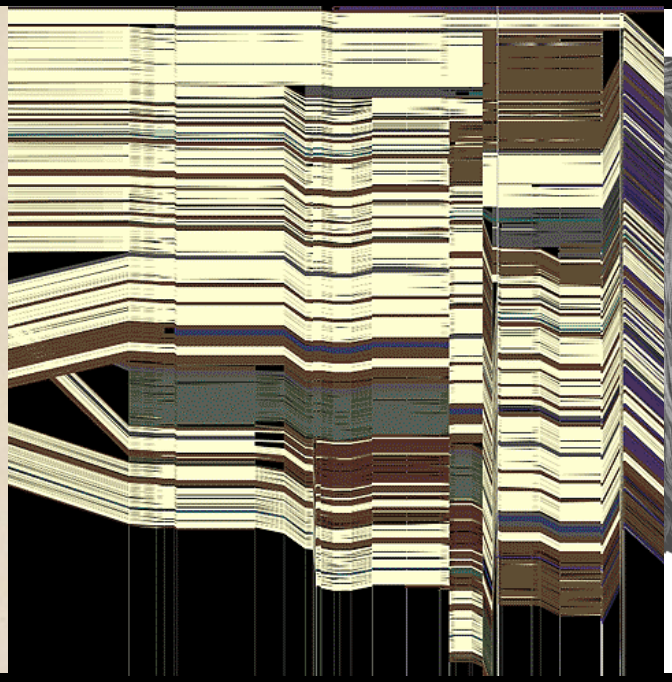
CSE 412 - Intro to Data Visualization

Multidimensional Data





Jane Hoffswell University of Washington

Guest Lecture: Narrative Visualization

This Fri Apr. 16 - Guest: Matt Conlen (UW, NYT)

https://mathisonian.com/

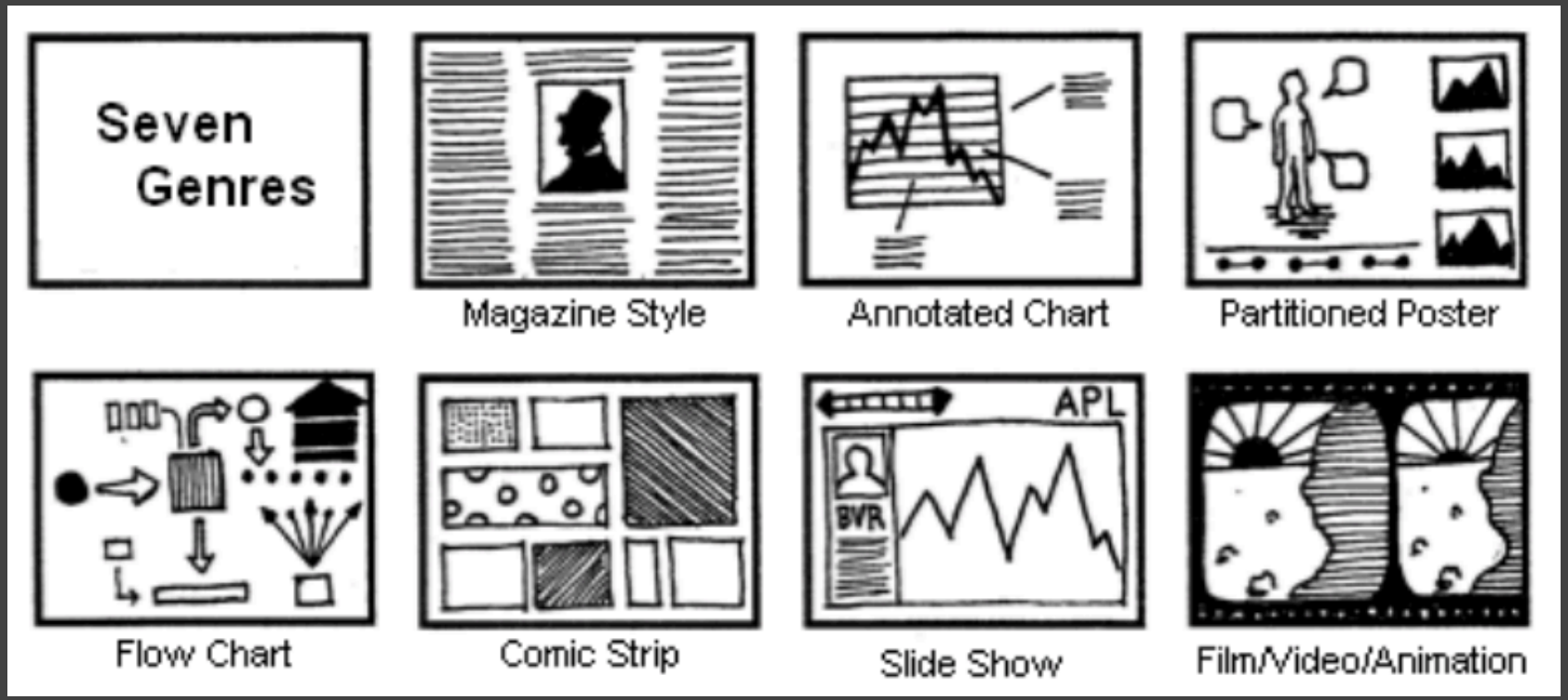
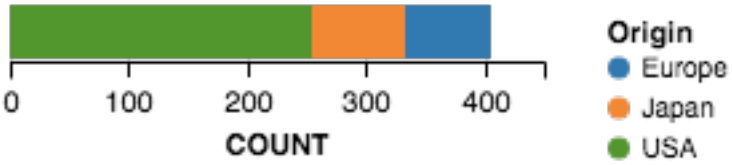


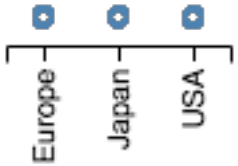
Image: "Narrative Visualization: Telling Stories with Data." Segel & Heer. InfoVis 2010

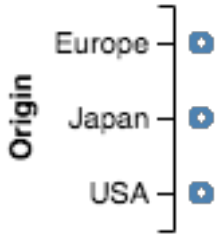
A Design Space of Visual Encodings

1D: Nominal

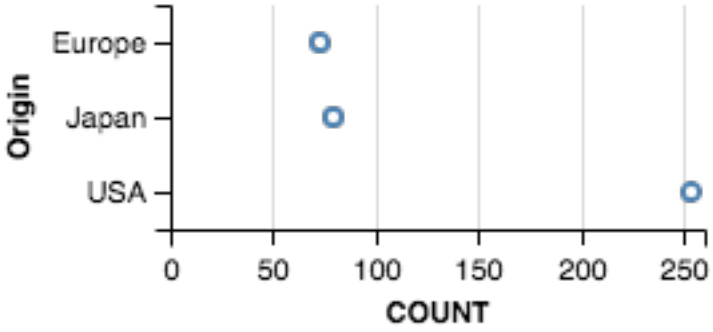


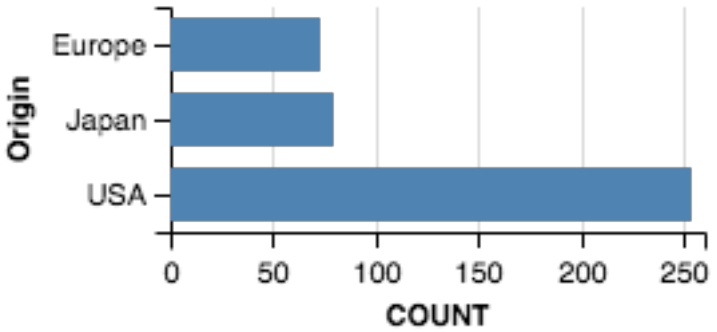
Raw

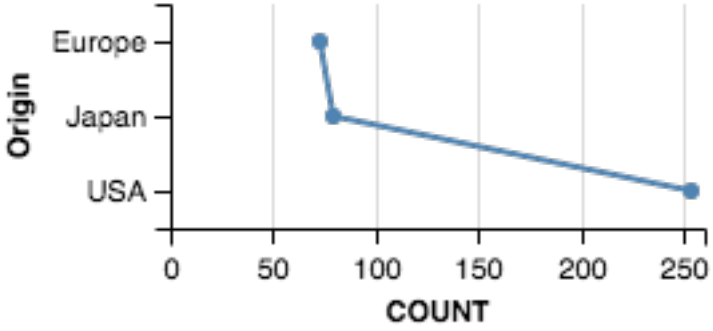


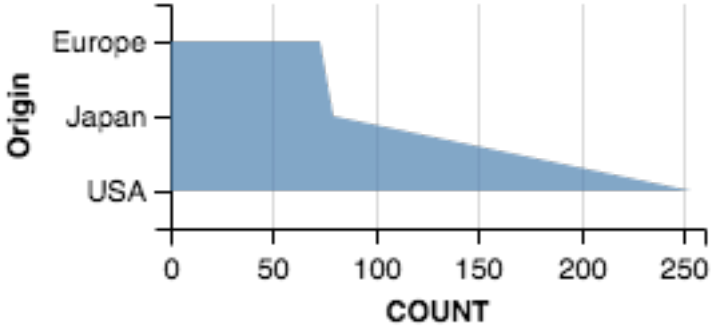


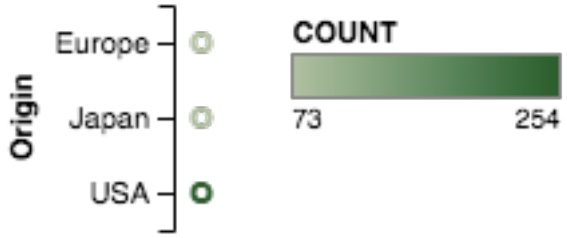
Aggregate (Count)

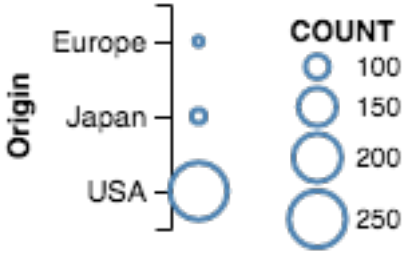








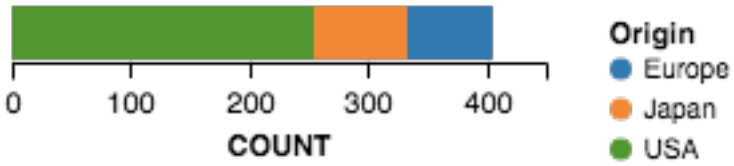


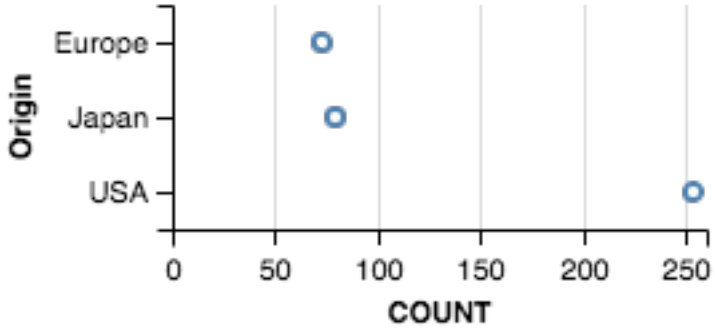


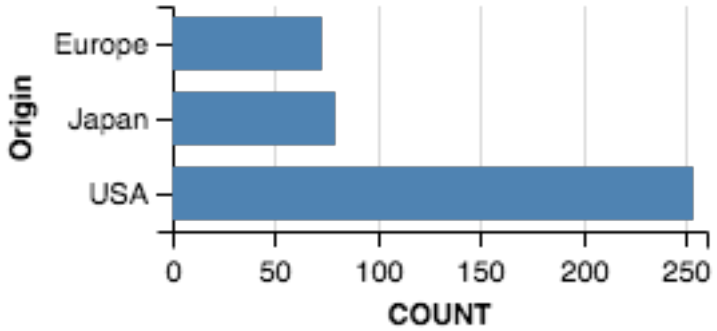
1D (N): Expressive?

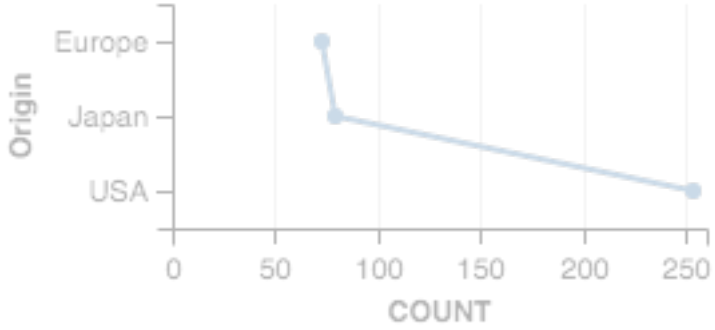
Raw

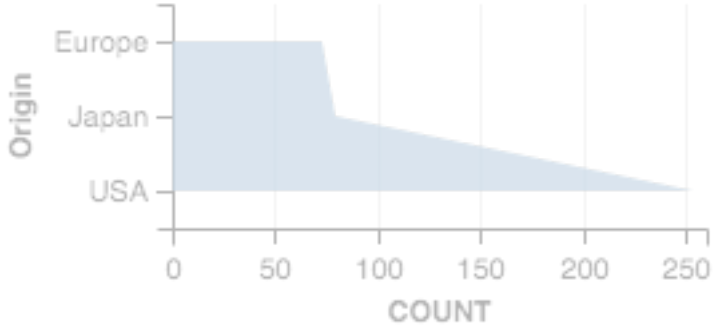
Aggregate (Count)

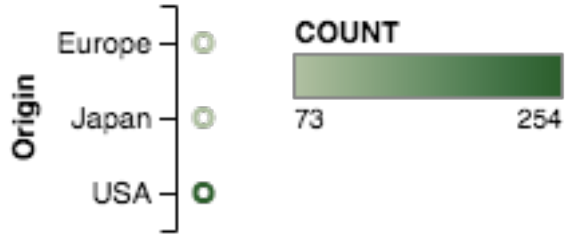


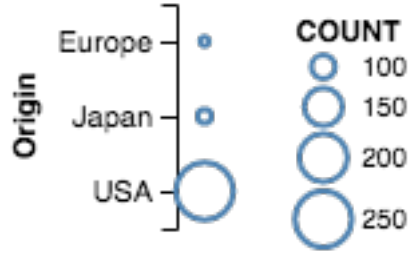






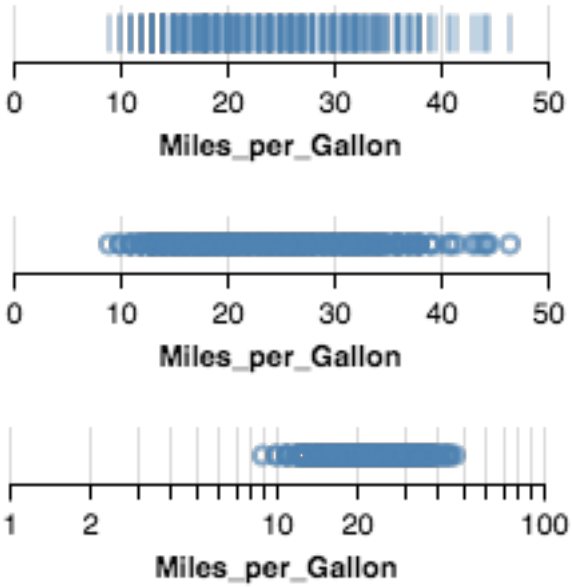


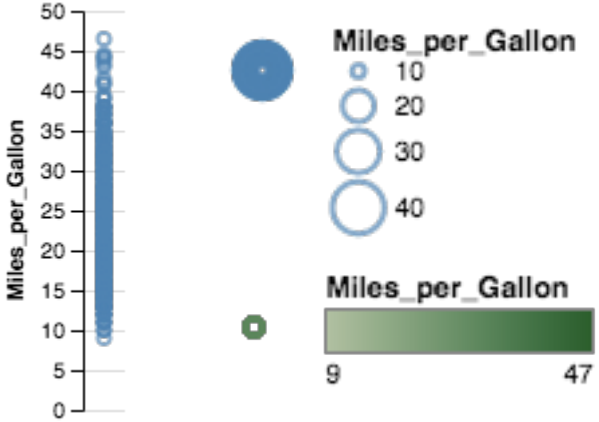




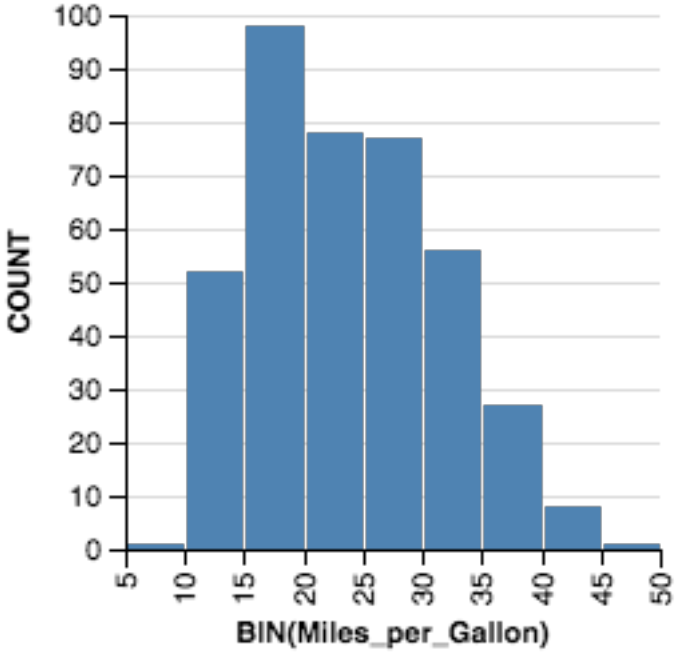
1D: Quantitative

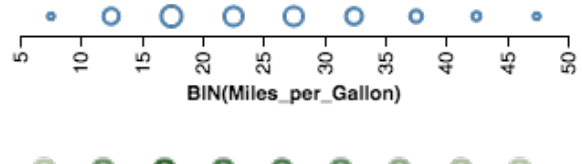
Raw

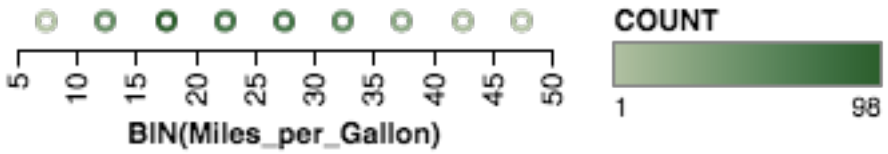




Aggregate (Count)

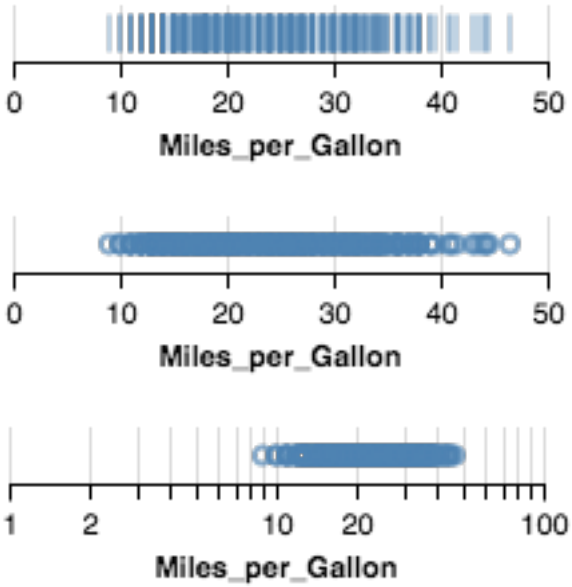


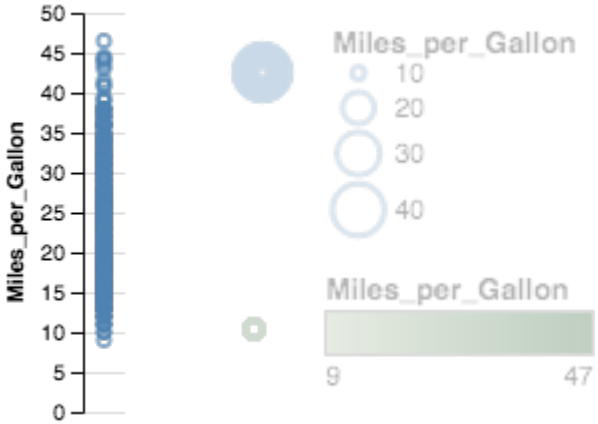




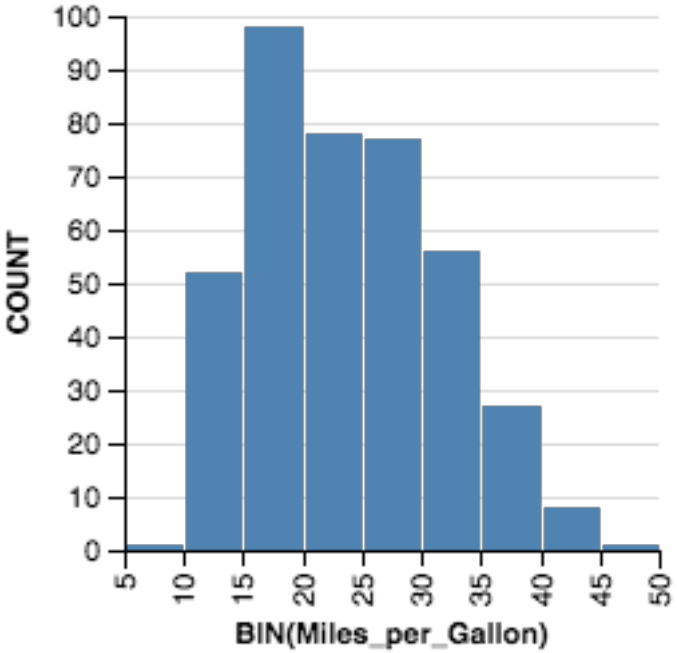
1D: Quantitative - Expressive?

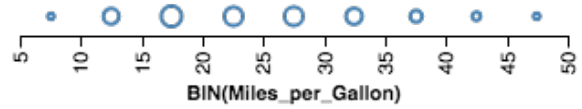
Raw

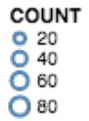




Aggregate (Count)

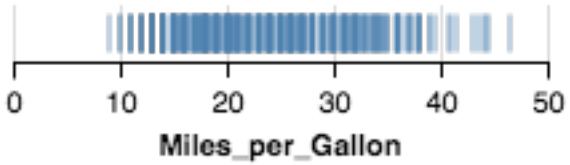


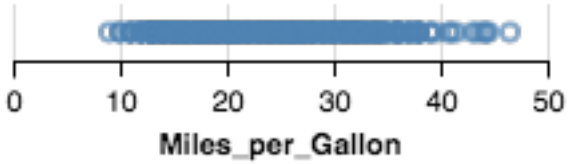


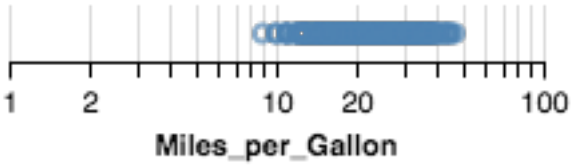


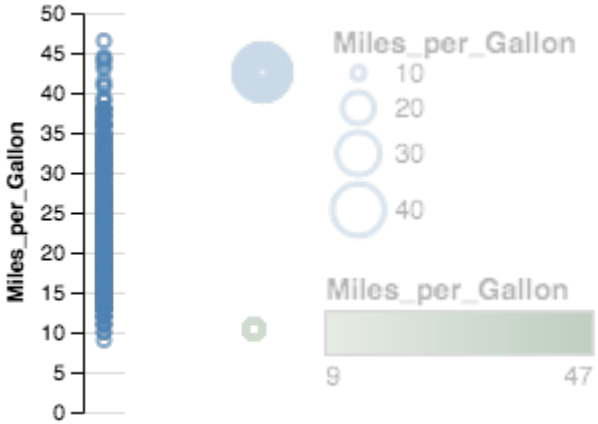
1D: Quantitative - Effective?

Raw

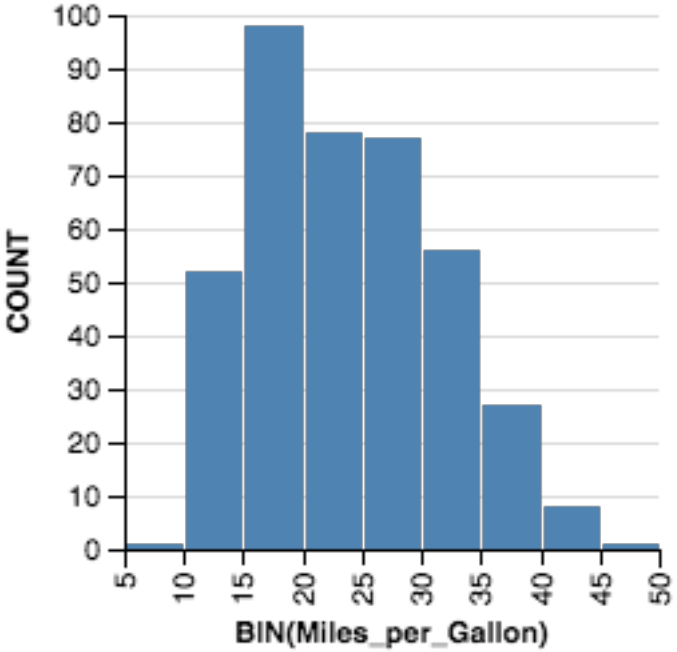


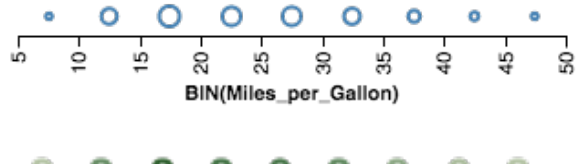


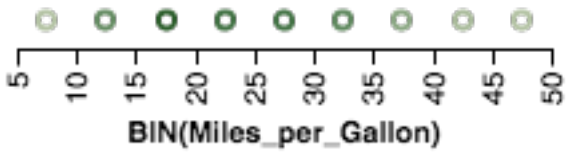




Aggregate (Count)

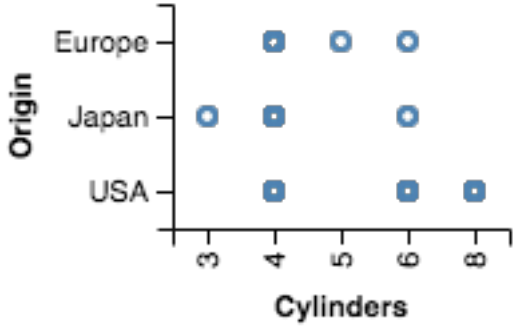


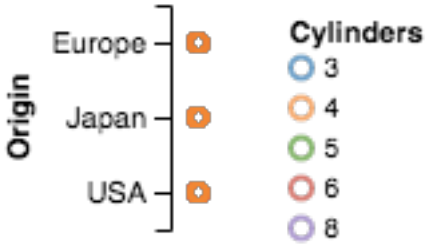


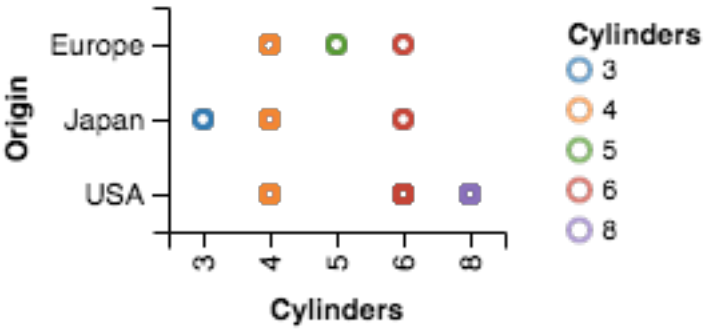


2D: Nominal x Nominal

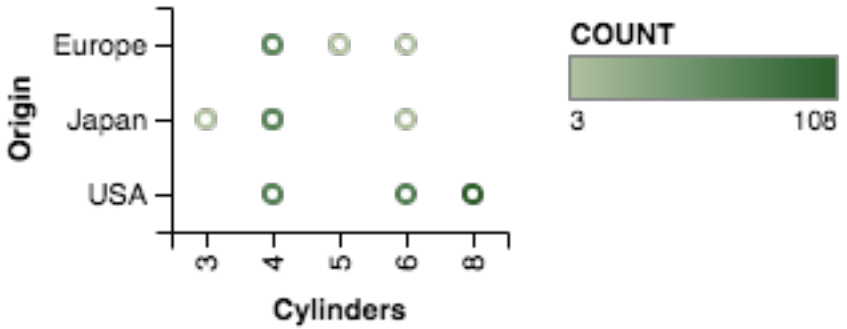
Raw

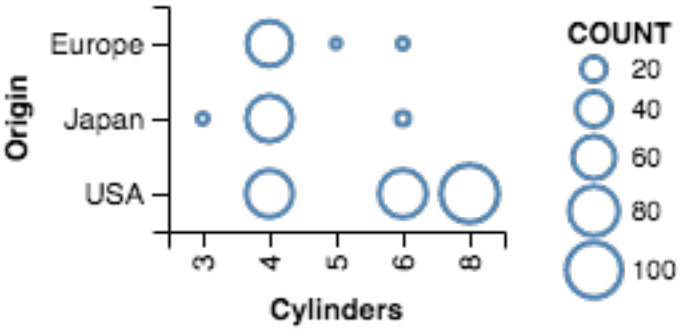


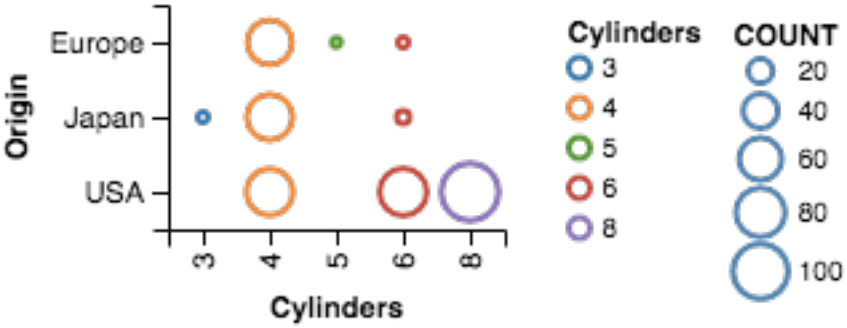




Aggregate (Count)

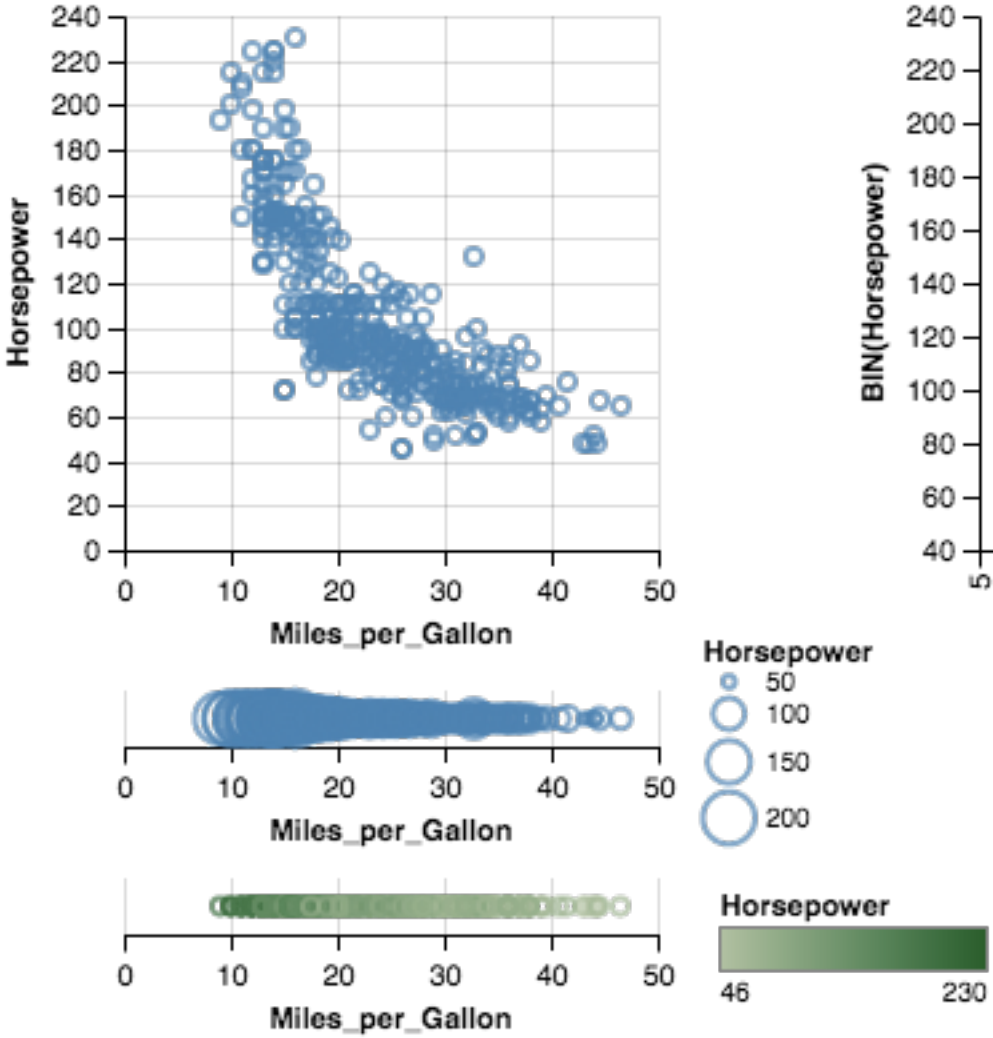




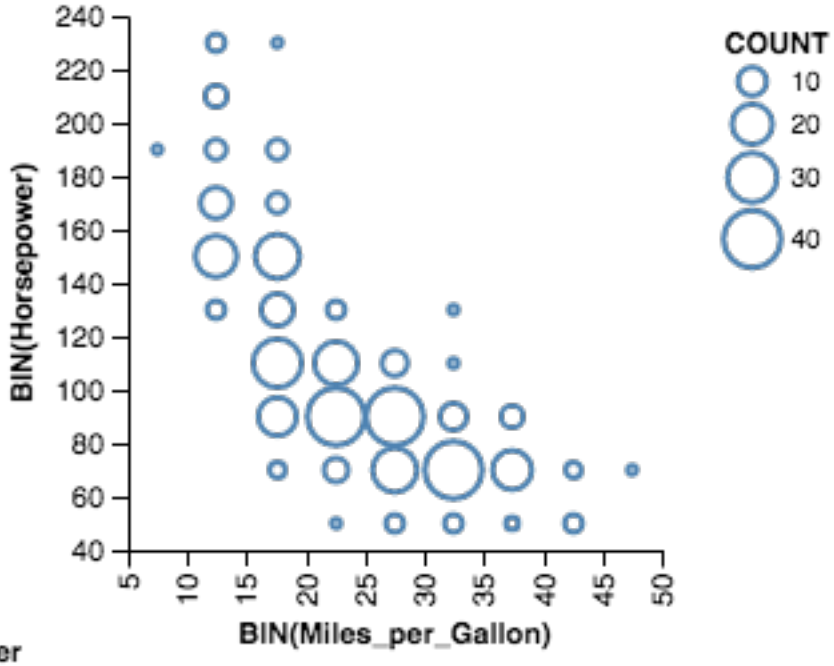


2D: Quantitative x Quantitative

Raw

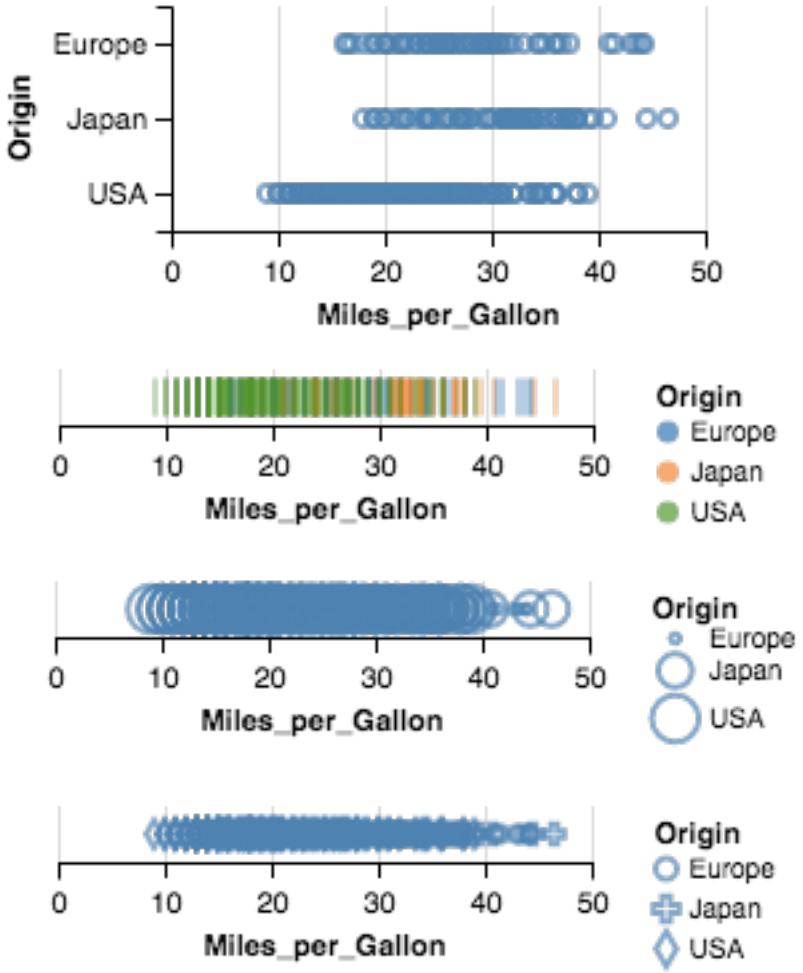


Aggregate (Count)

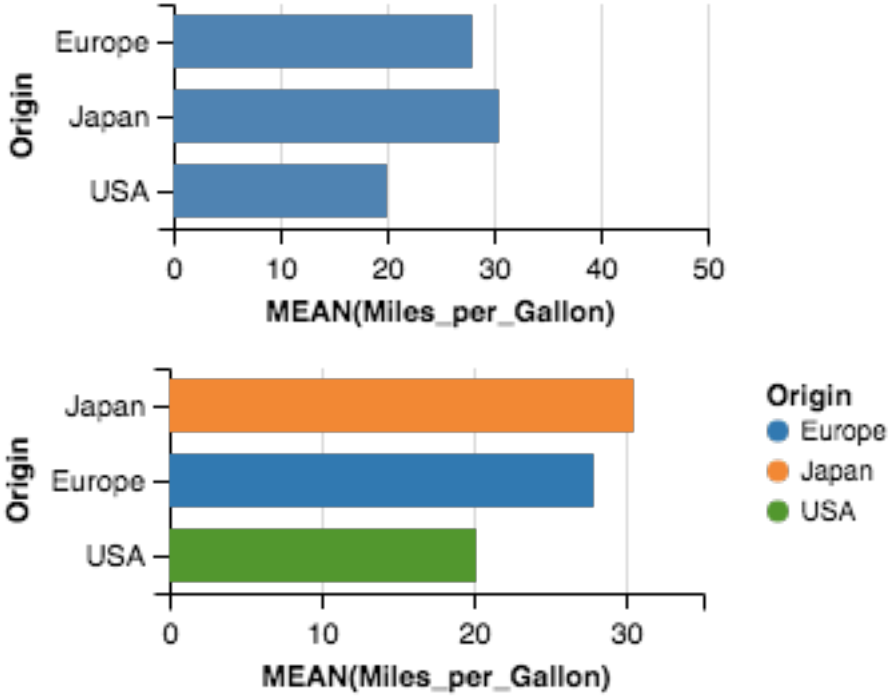


2D: Nominal x Quantitative

Raw



Aggregate (Mean)



3D and Higher

Two variables [x,y]

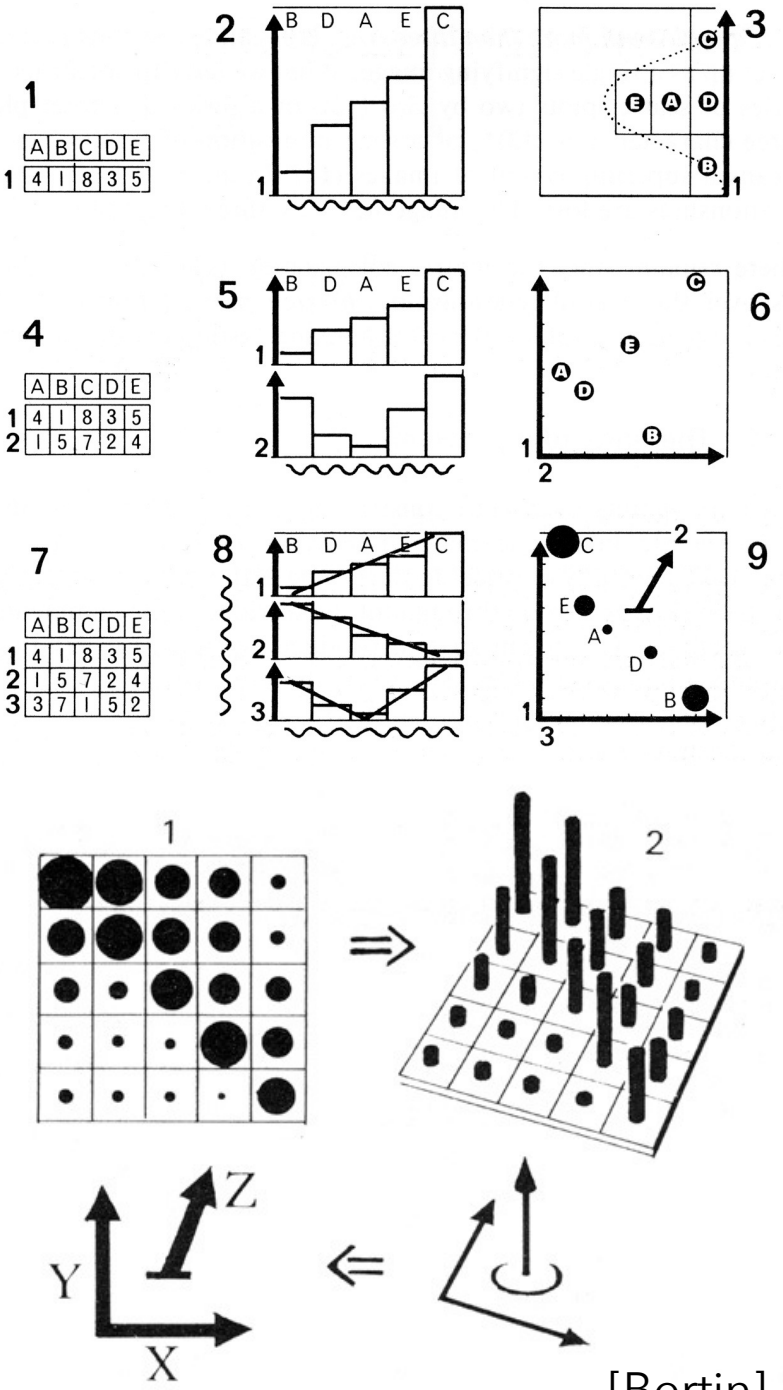
Can map to 2D points.

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]

Designing Charts

Example: Cars

Properties of different models of cars

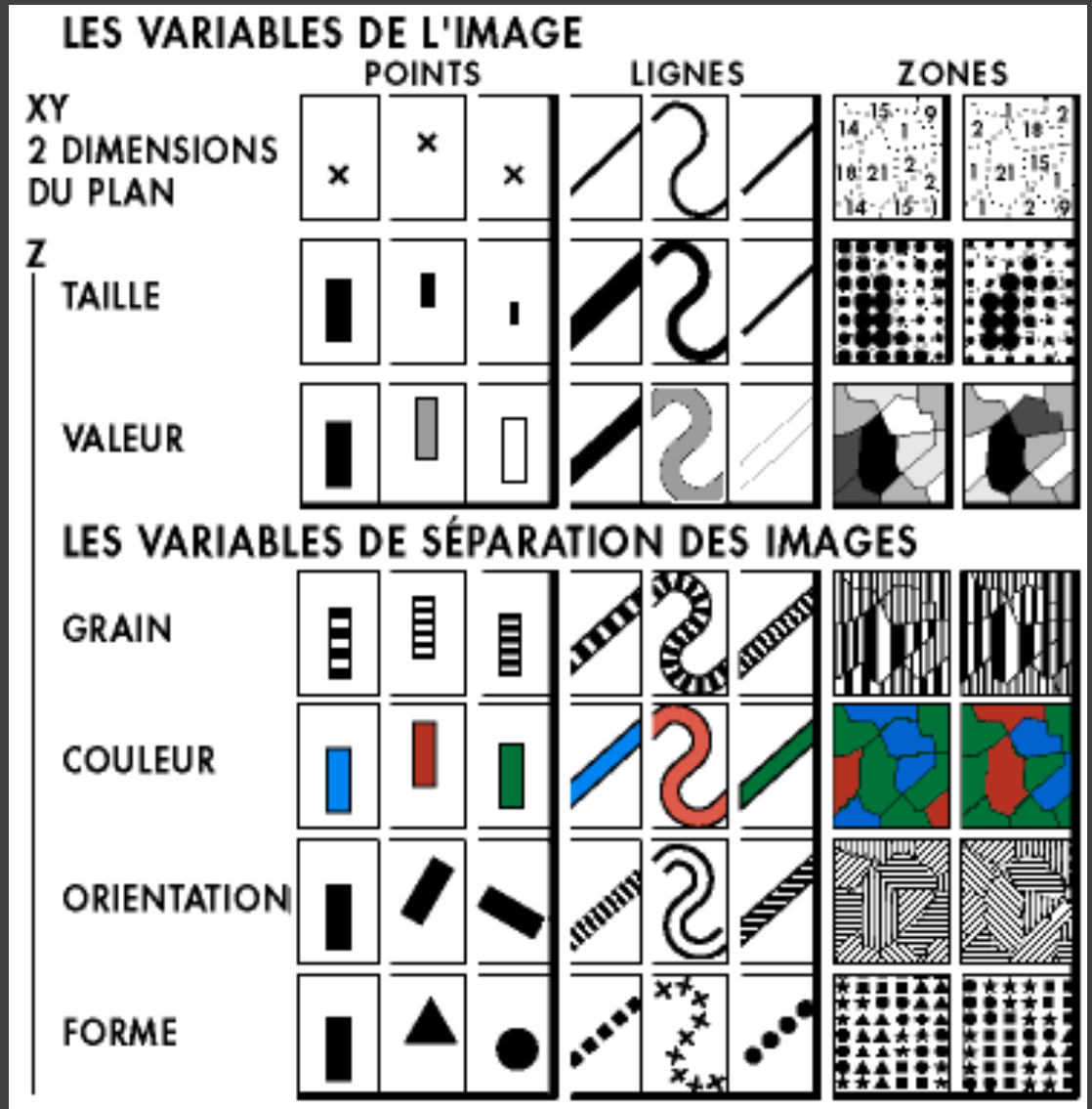
9-Dimensions, 406 rows

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)

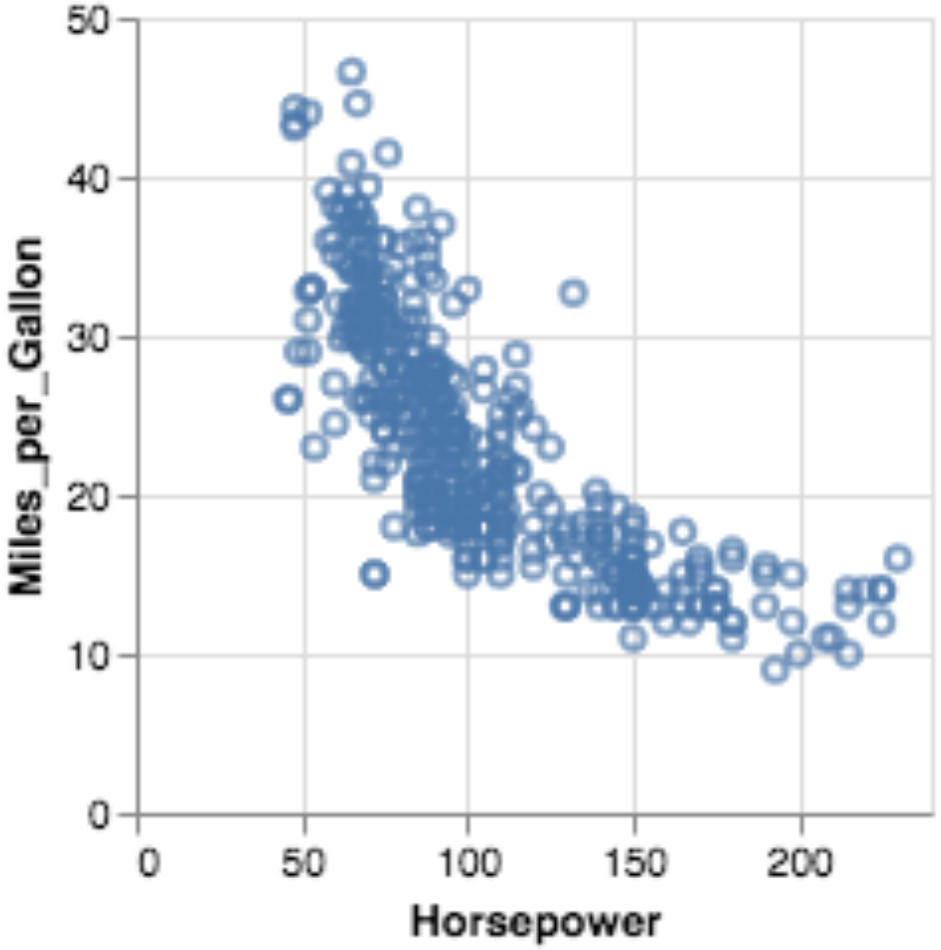
Acceleration

Year

Name

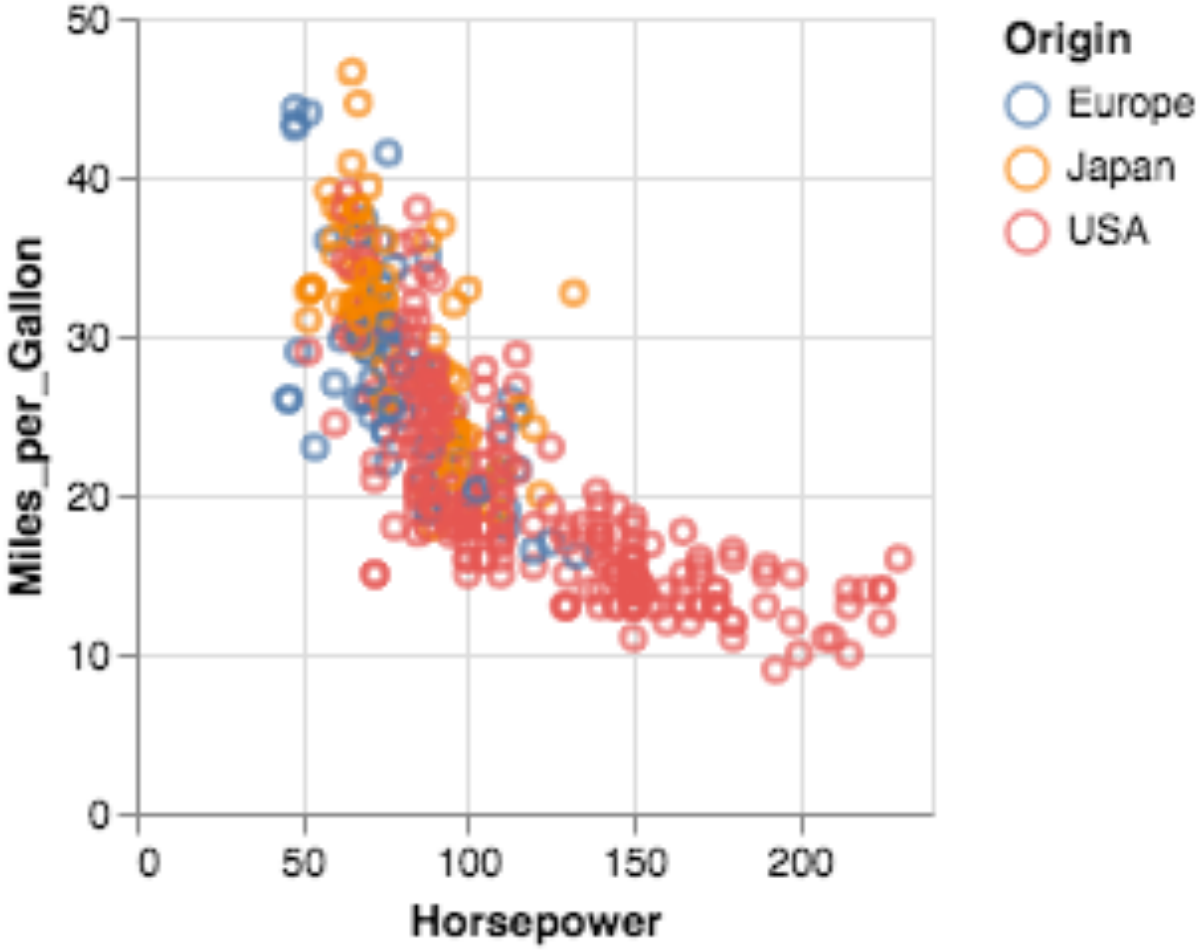
2-Dimensions

Next Up:Origin (N)



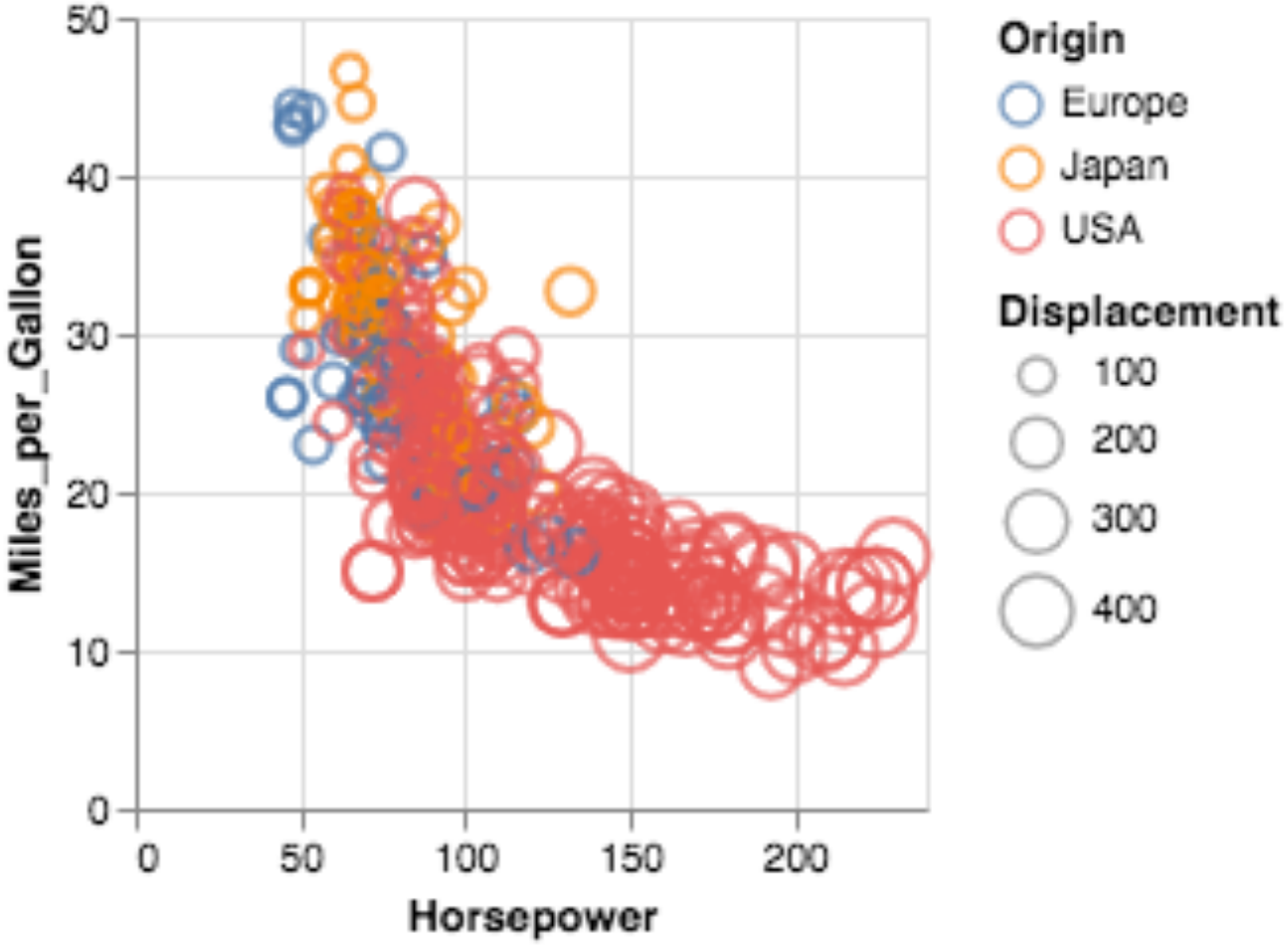
3-Dimensions

Next Up:
Displacement (Q)



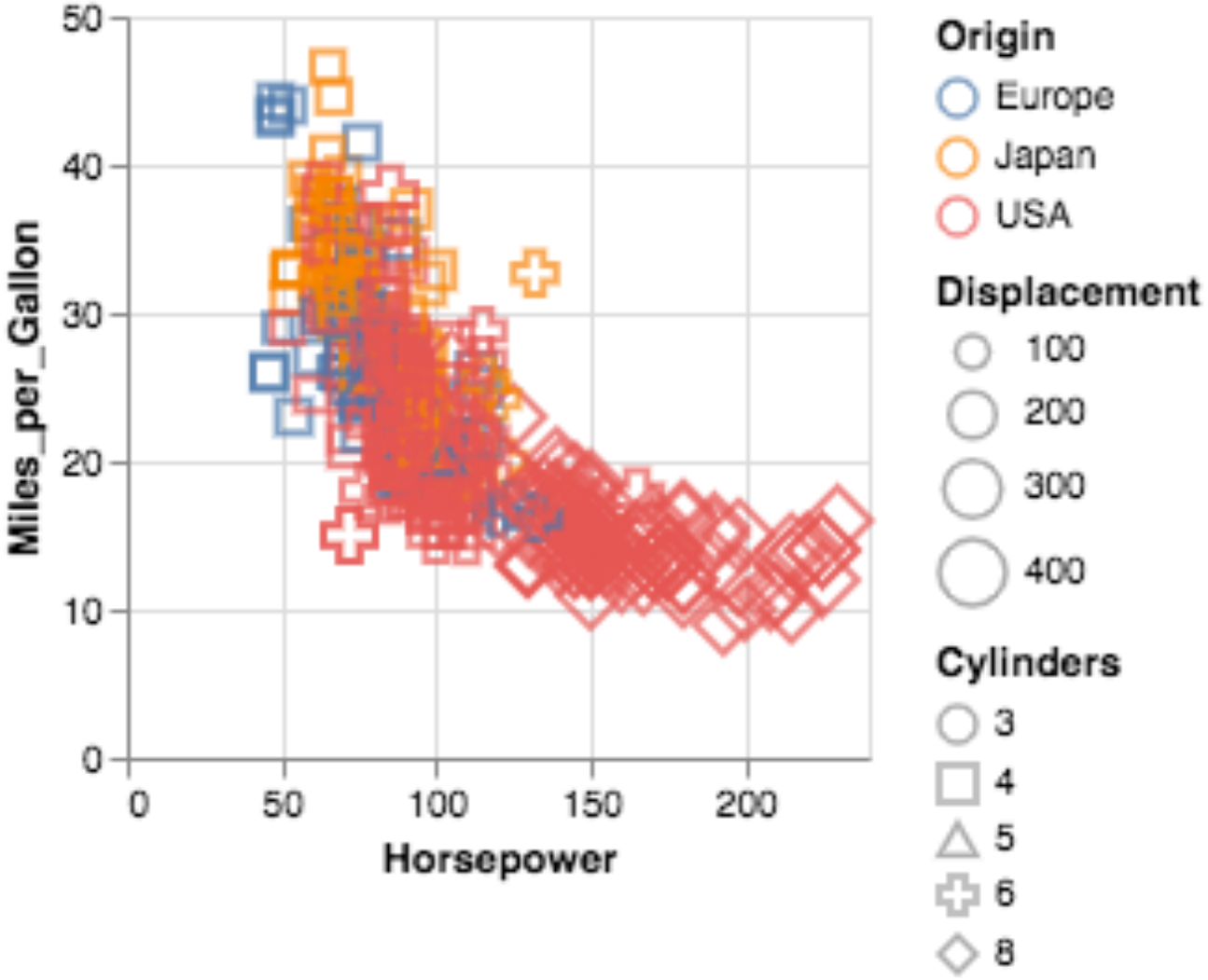
4-Dimensions

Next Up:
Cylinders (O/Q)



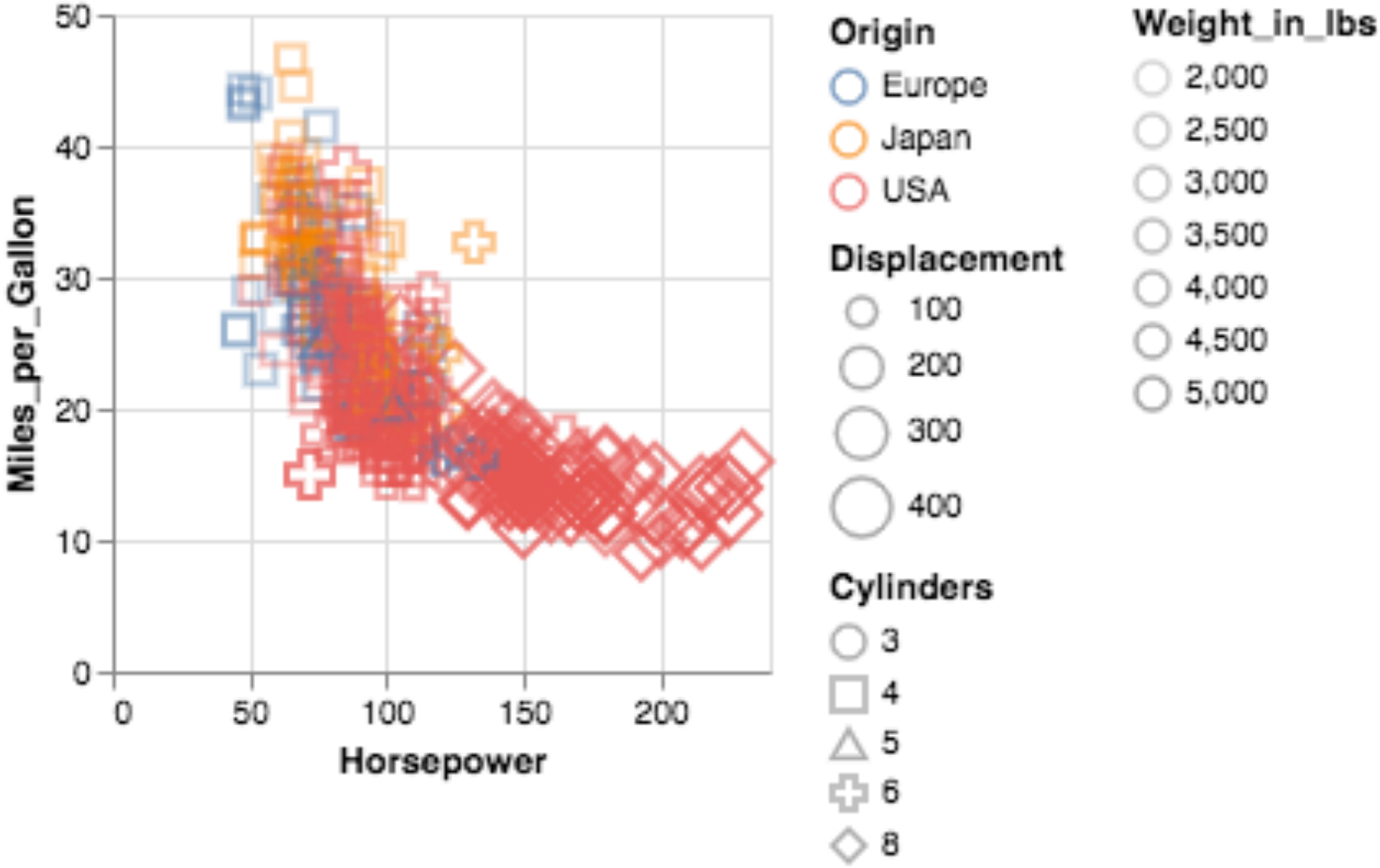
5-Dimensions

Next Up:
Weight (Q)



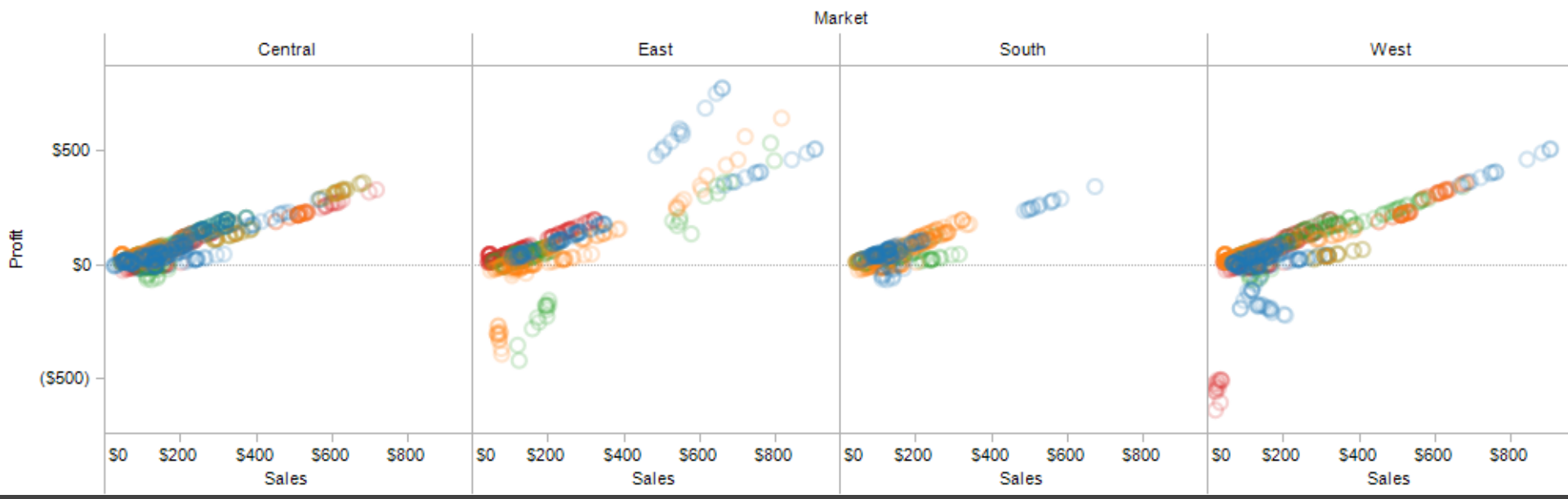
6-Dimensions

Next Up:
Acceleration (Q)



7-Dimensions?

Trellis Plots



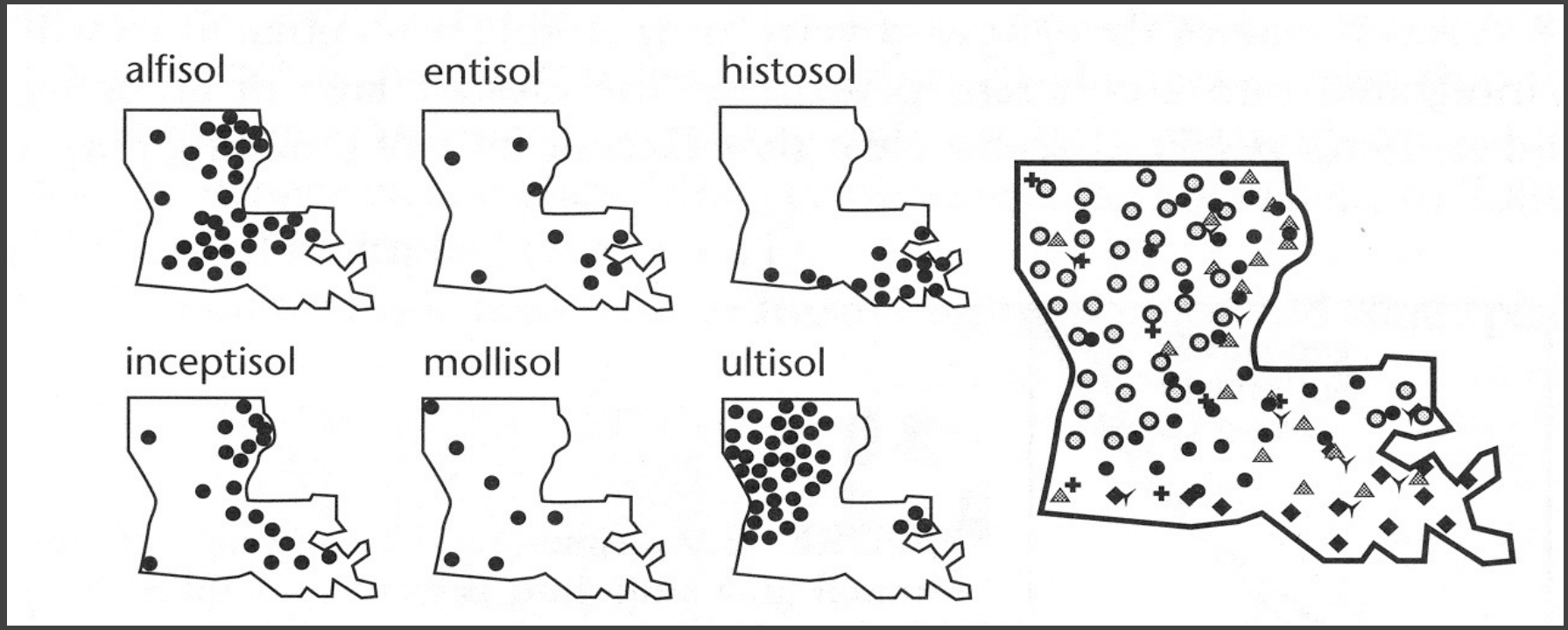
A *trellis plot* subdivides space to enable comparison across multiple plots.

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

Small Multiples

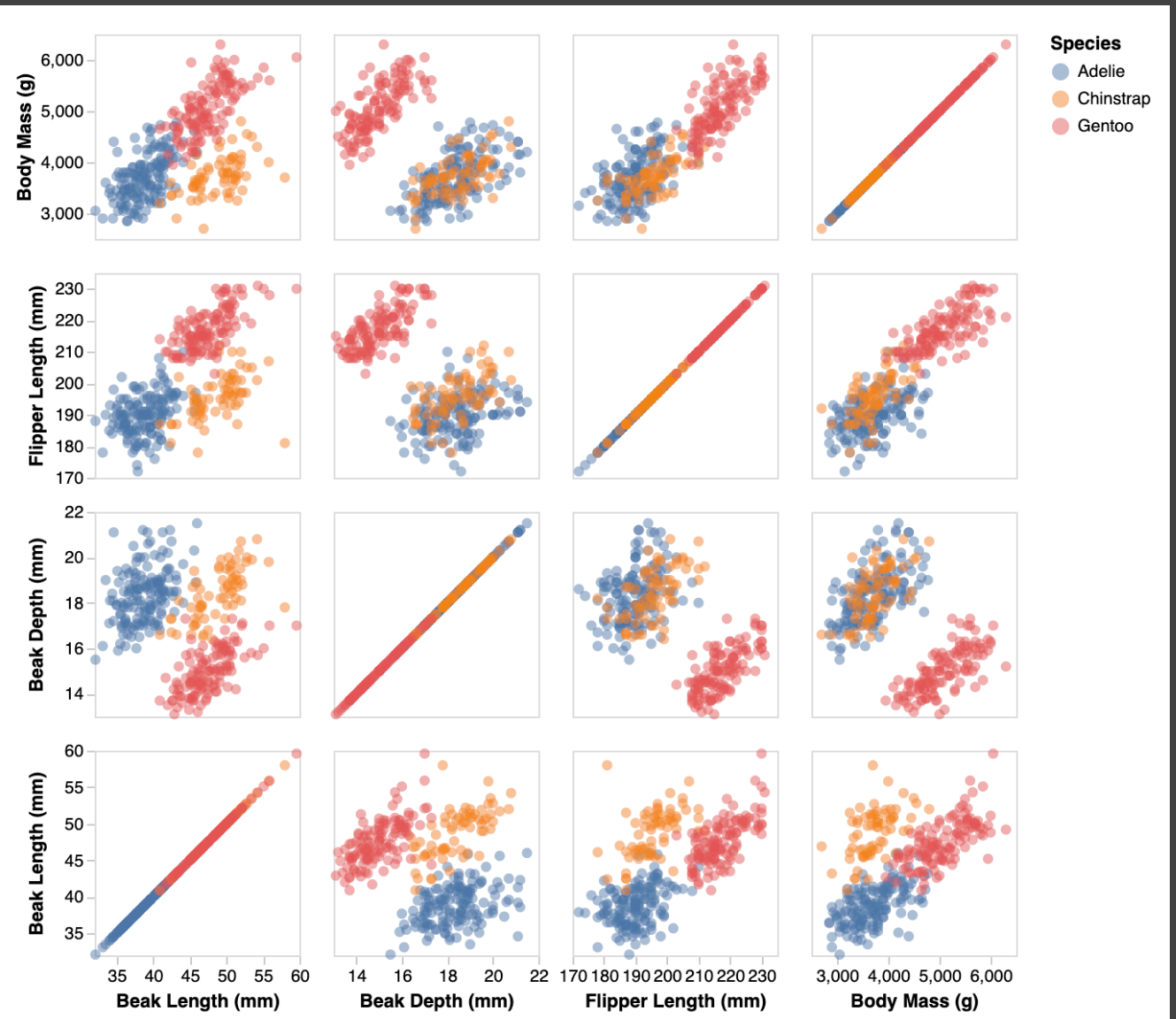
[MacEachren ’95, Figure 2.11, p. 38]

Small Multiples



[MacEachren ’95, Figure 2.11, p. 38]

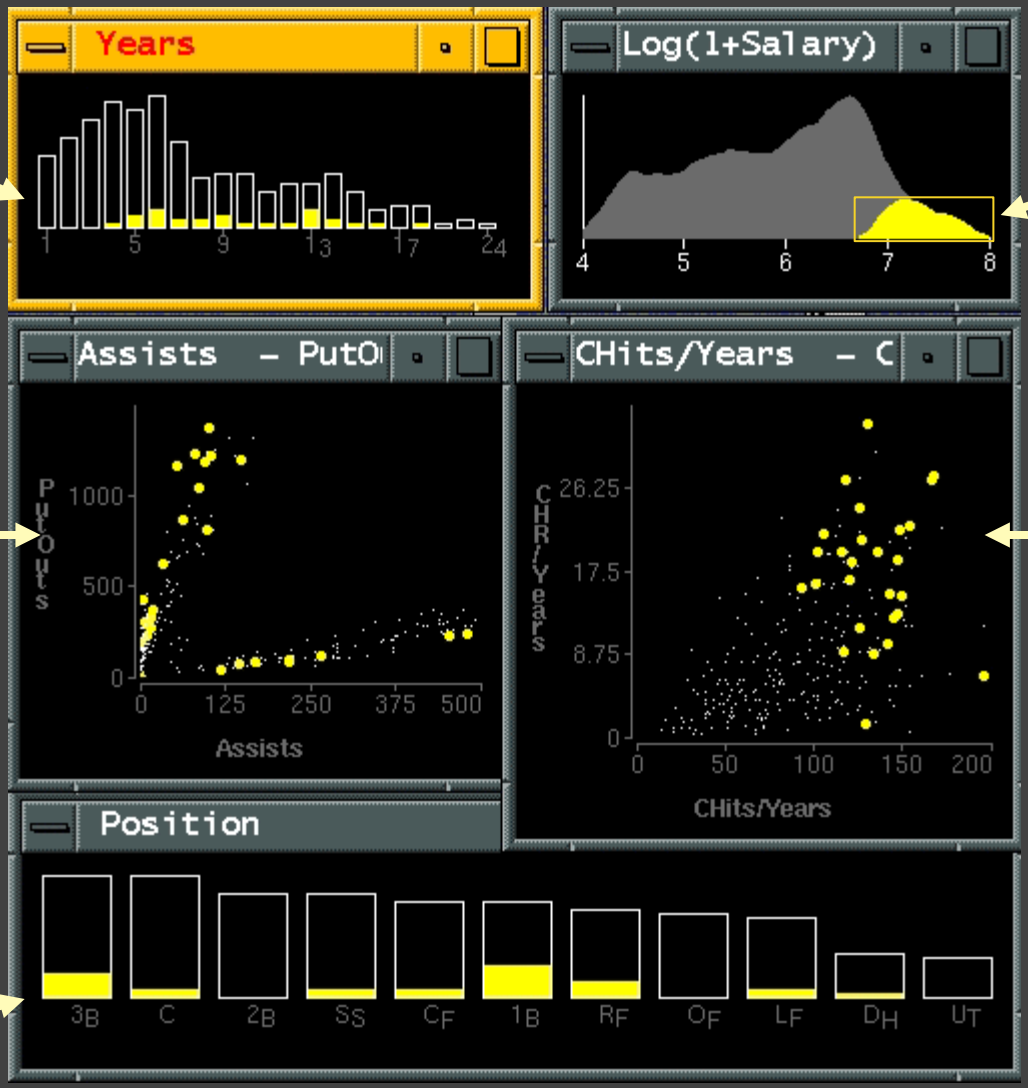
Scatterplot Matrix (SPLOM)



Scatter plots for pairwise comparison of each data dimension.

Multiple Coordinated Views

how long in majors



select high salaries

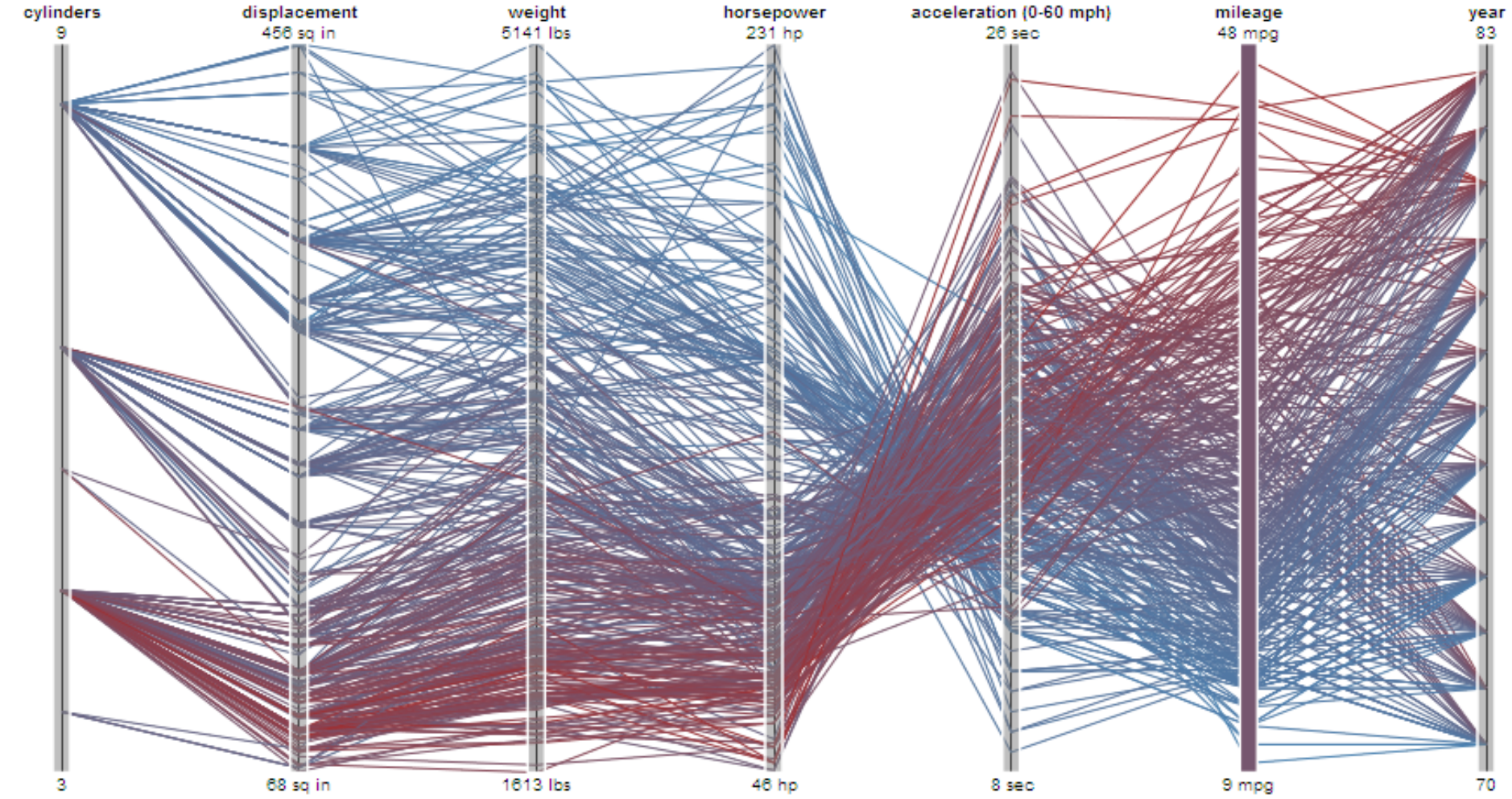
avg assists vs avg putouts (fielding ability)

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

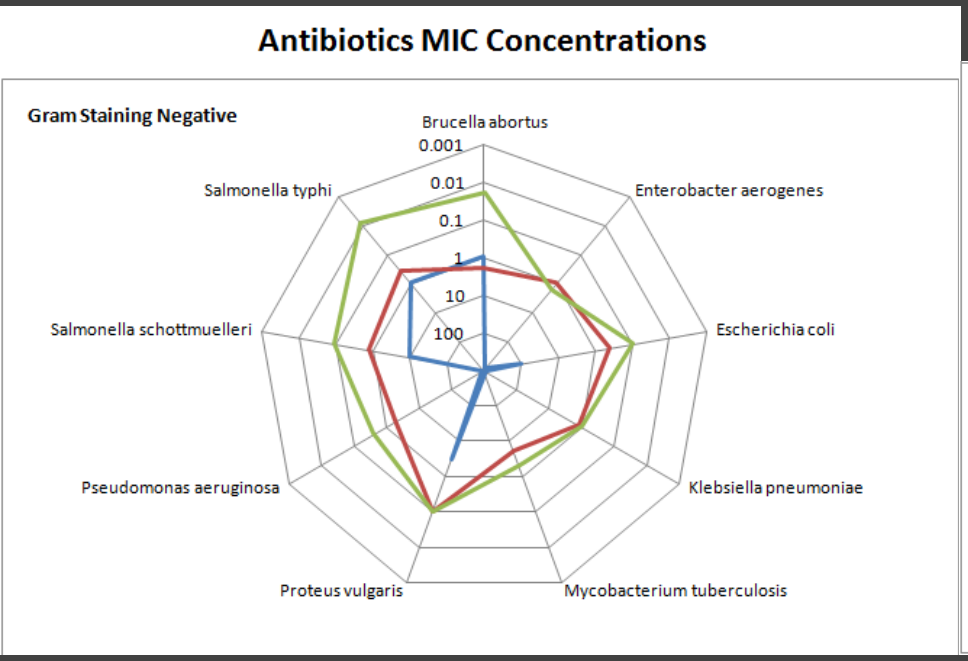
Between adjacent axes: line crossings imply neg. correlation, shared slopes imply pos. correlation.

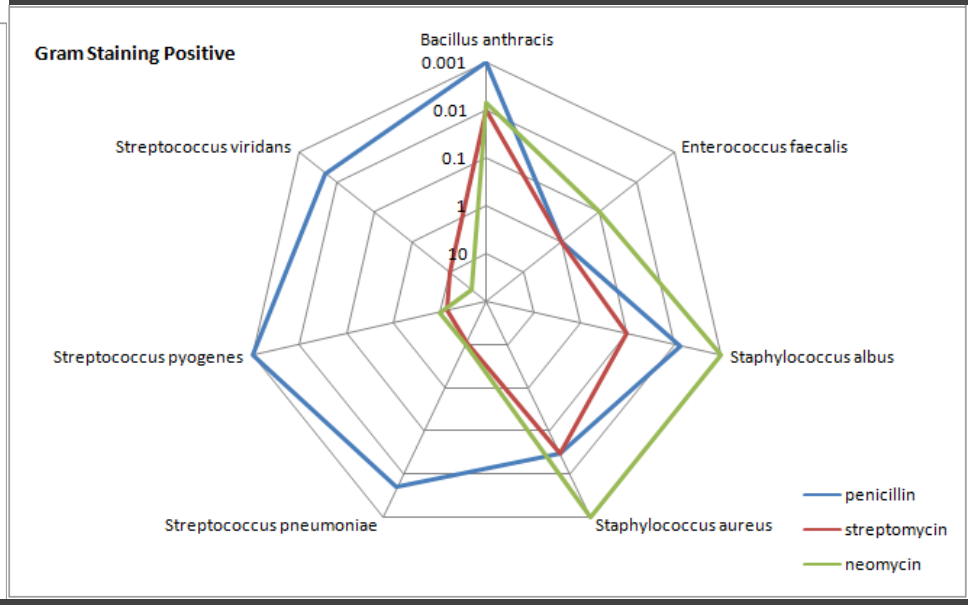
Full plot can be cluttered. **Interactive selection** can be used to assess multivariate relationships.

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).

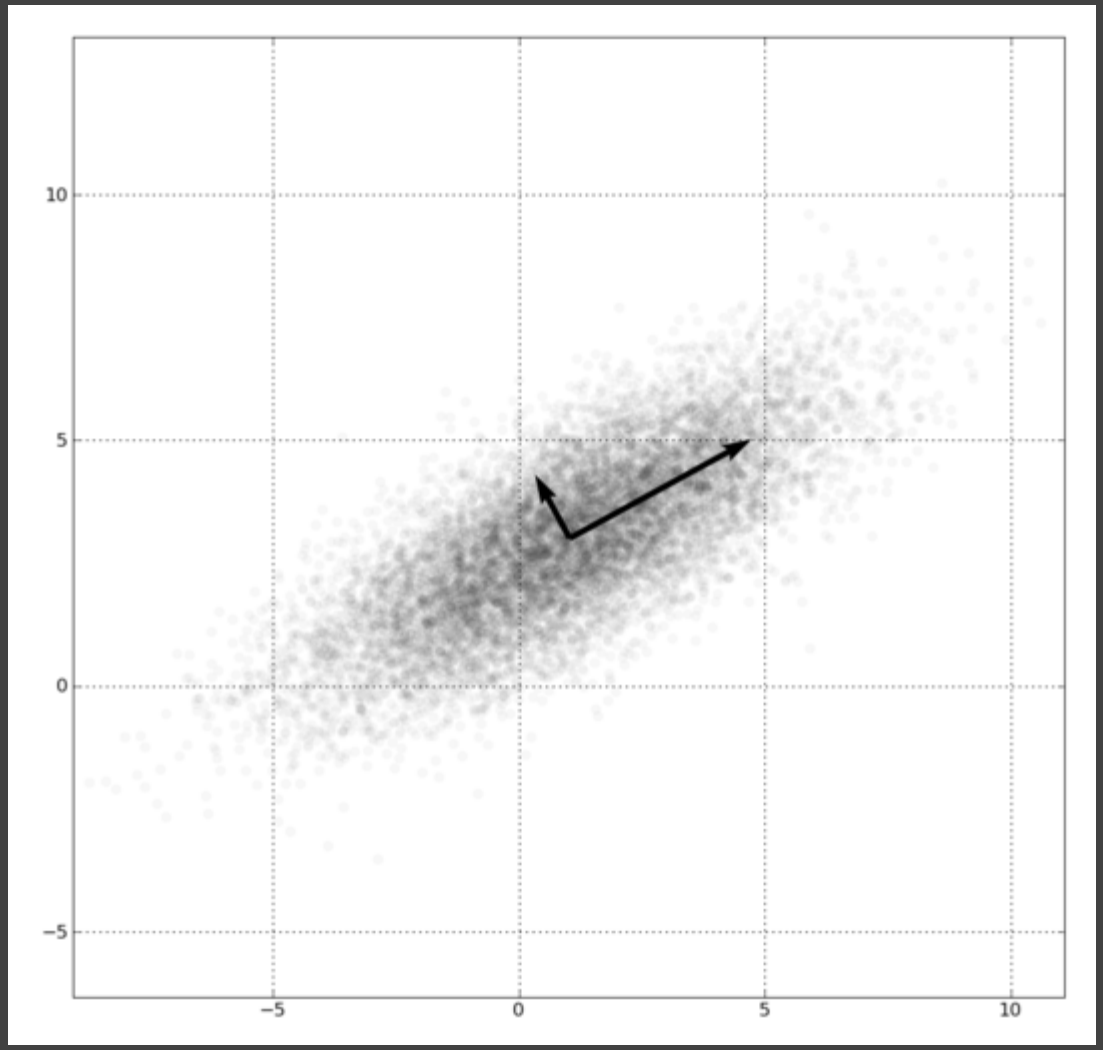
Reduction Techniques

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.

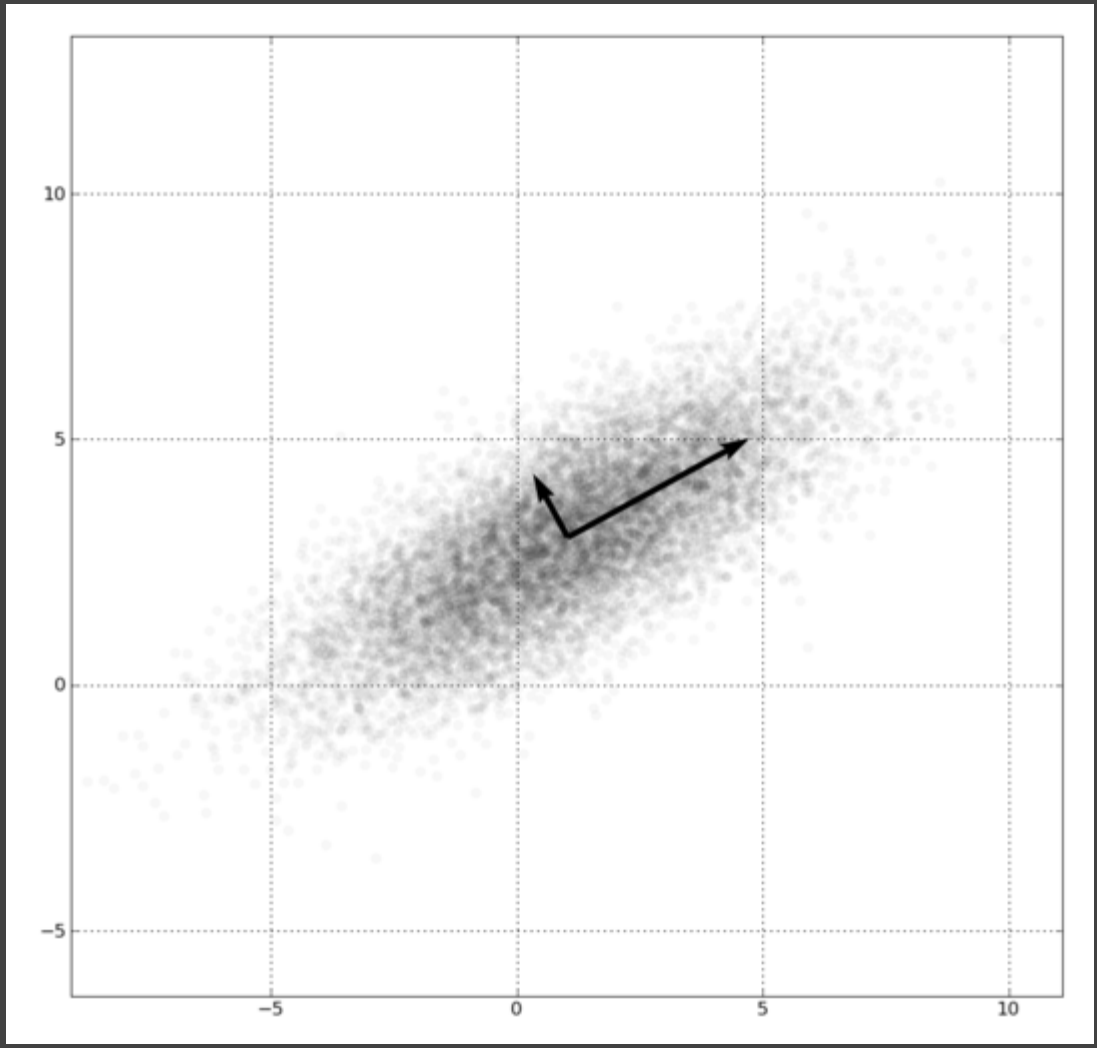
Uniform Manifold Approx. & Projection (UMAP)
Identify local manifolds, then stitch them together.

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

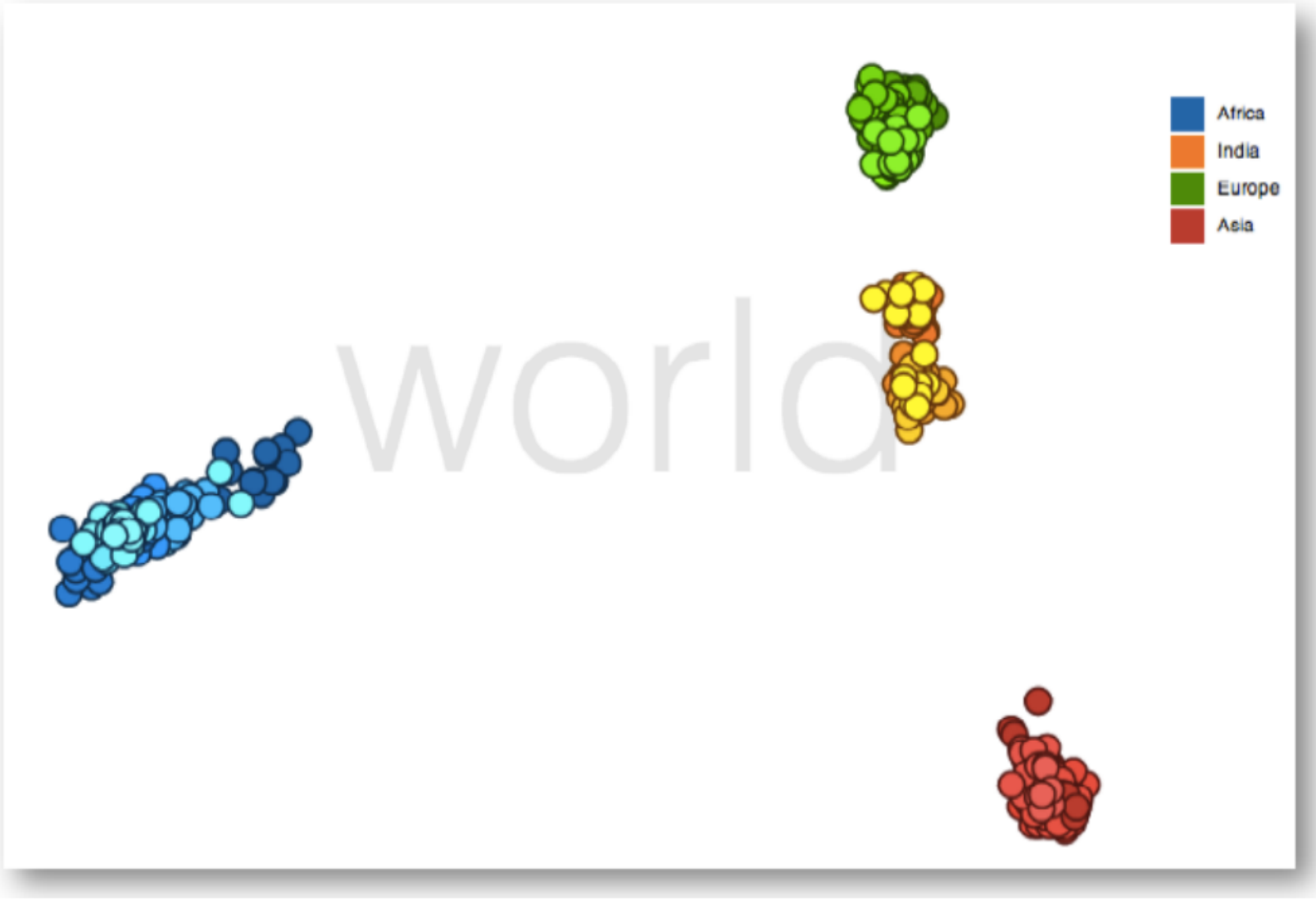


Linear transform: scale and rotate original space.

Lines (vectors) project to lines.

Preserves global distances.

PCA of Genomes [Demiralp et al. '13]

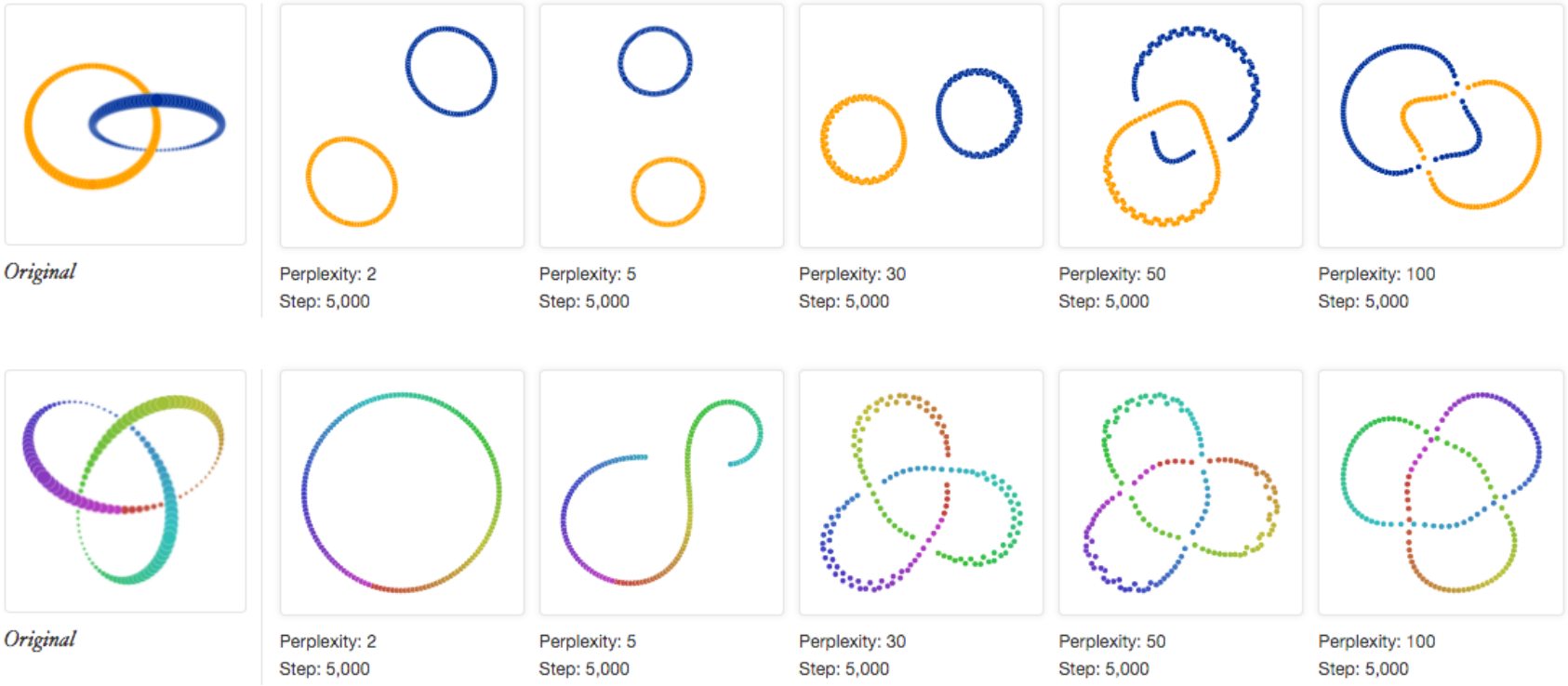


Non-Linear Techniques

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

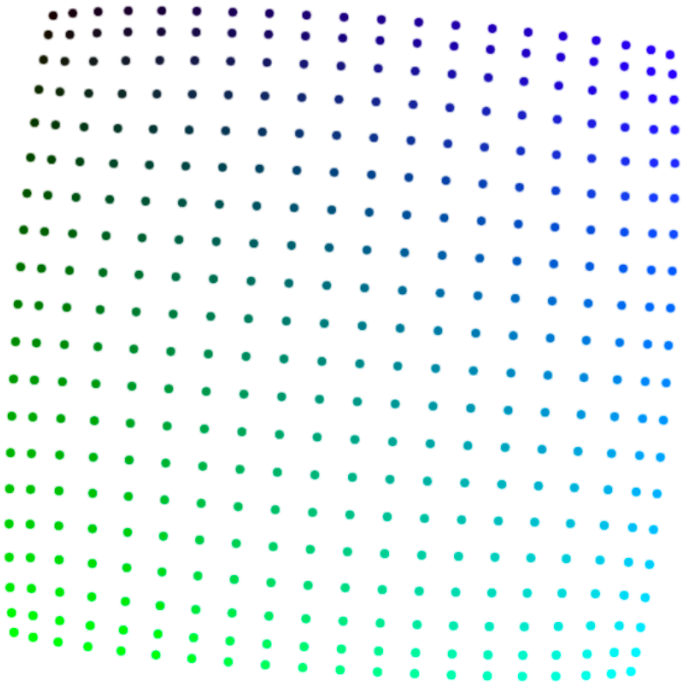
Visualizing t-SNE [Wattenberg et al. '16]



Results can be highly sensitive to the algorithm parameters!

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

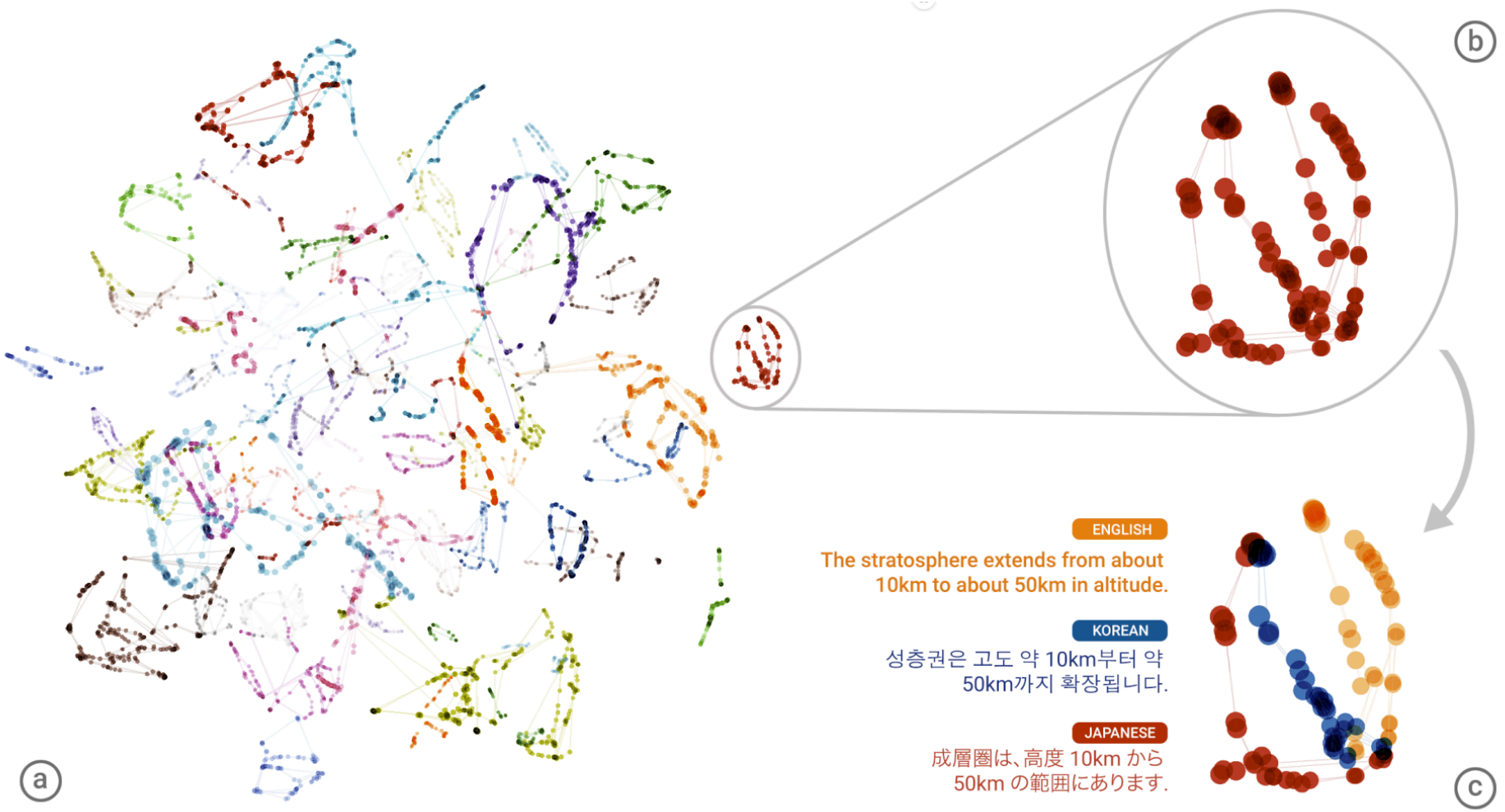
IAN JOHNSON
Google Cloud

Oct. 13
2016

Citation:
Wattenberg, et al., 2016

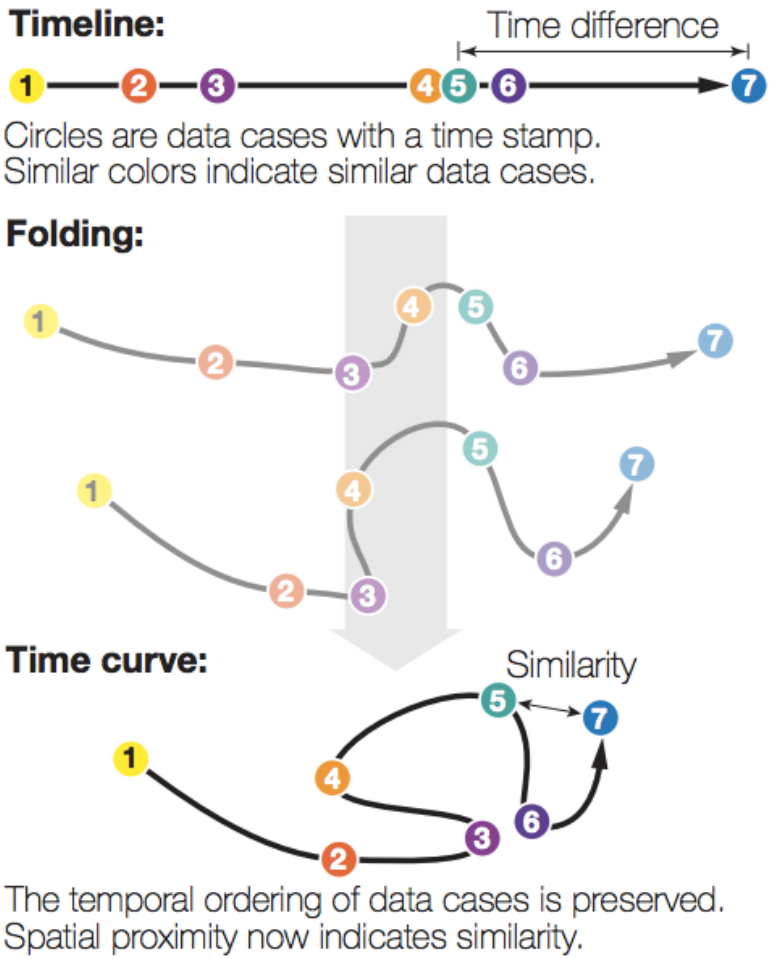
distill.pub

MT Embedding [Johnson et al. 2018]

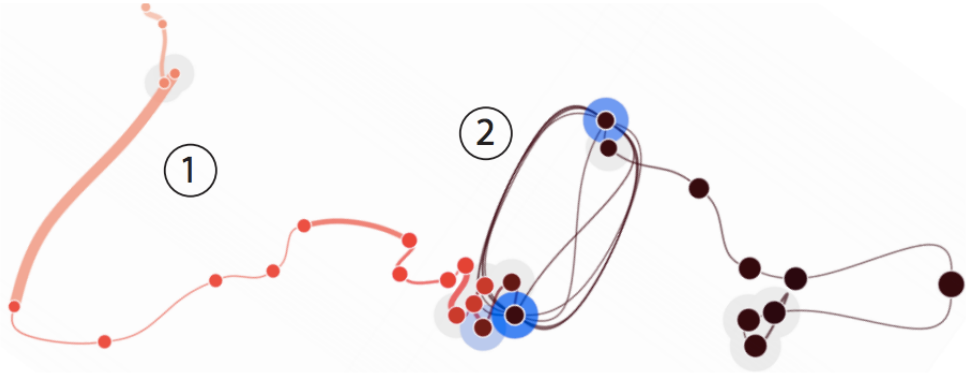


t-SNE projection of latent space of language translation model.

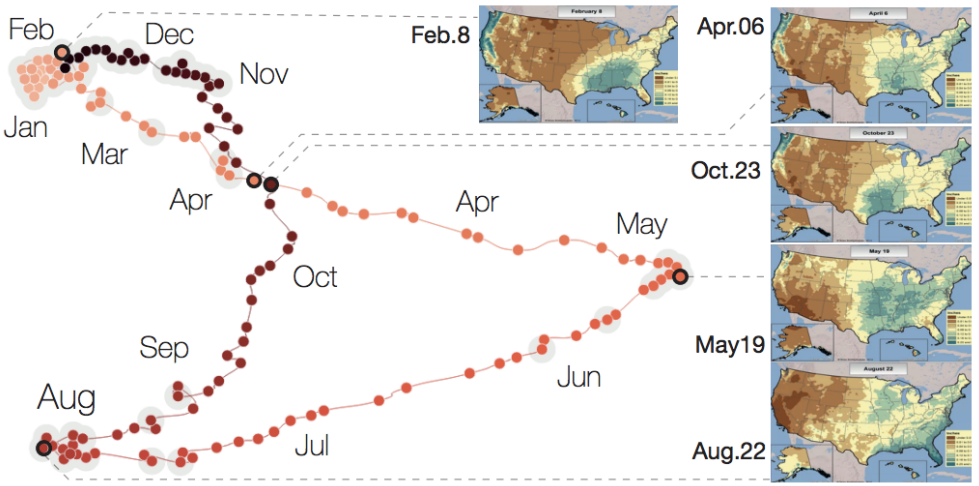
Time Curves [Bach et al. '16]



(a) Folding time



Wikipedia "Chocolate" Article



U.S. Precipitation over 1 Year

Visual Encoding Design

Use **expressive** and **effective** encodings
Reduce the problem space
Avoid **over-encoding**
Use **space** and **small multiples** intelligently
Use **interaction** to generate *relevant* views

Rarely does a single visualization answer all questions. Instead, the ability to generate appropriate visualizations quickly is critical!

About the design process…

Visualization draws upon both science and art!

Principles like expressiveness & effectiveness are not hard-and-fast rules, but can assist us to guide the process and articulate alternatives.

They can lead us to think more deeply about our design rationale and prompt us to reflect.

It helps to know "the rules" in order to wisely bend (*or break*) them at the right times!

Administrivia

A2: Exploratory Data Analysis

Use visualization software to form & answer questions

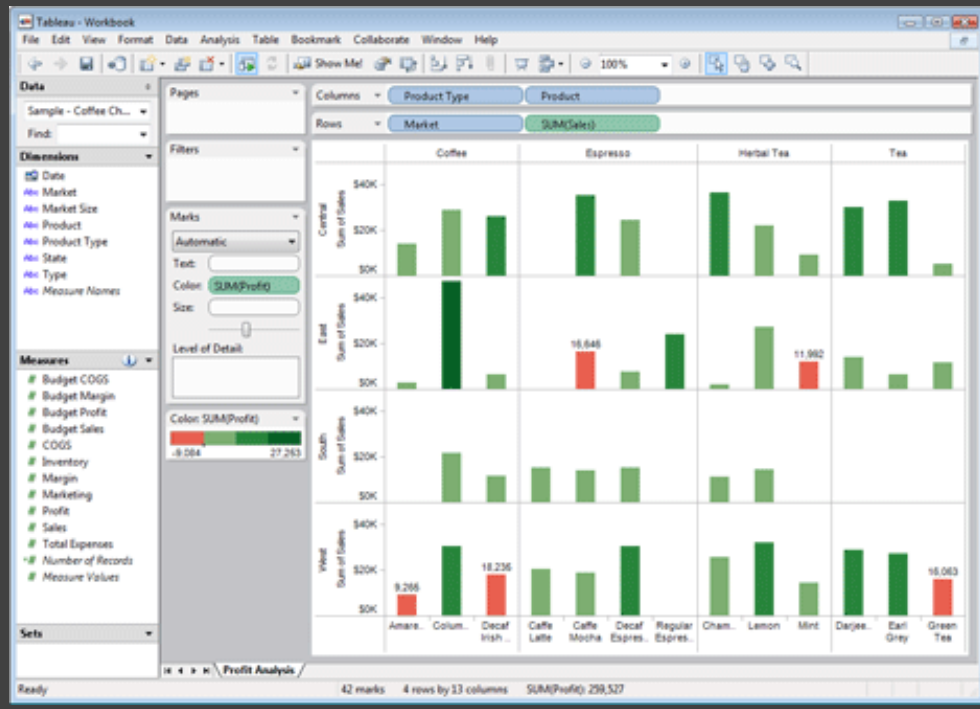
First steps:

Step 1: Pick domain & data

Step 2: Pose questions

Step 3: Profile the data

Iterate as needed



Create visualizations

Interact with data

Refine your questions

Author a report

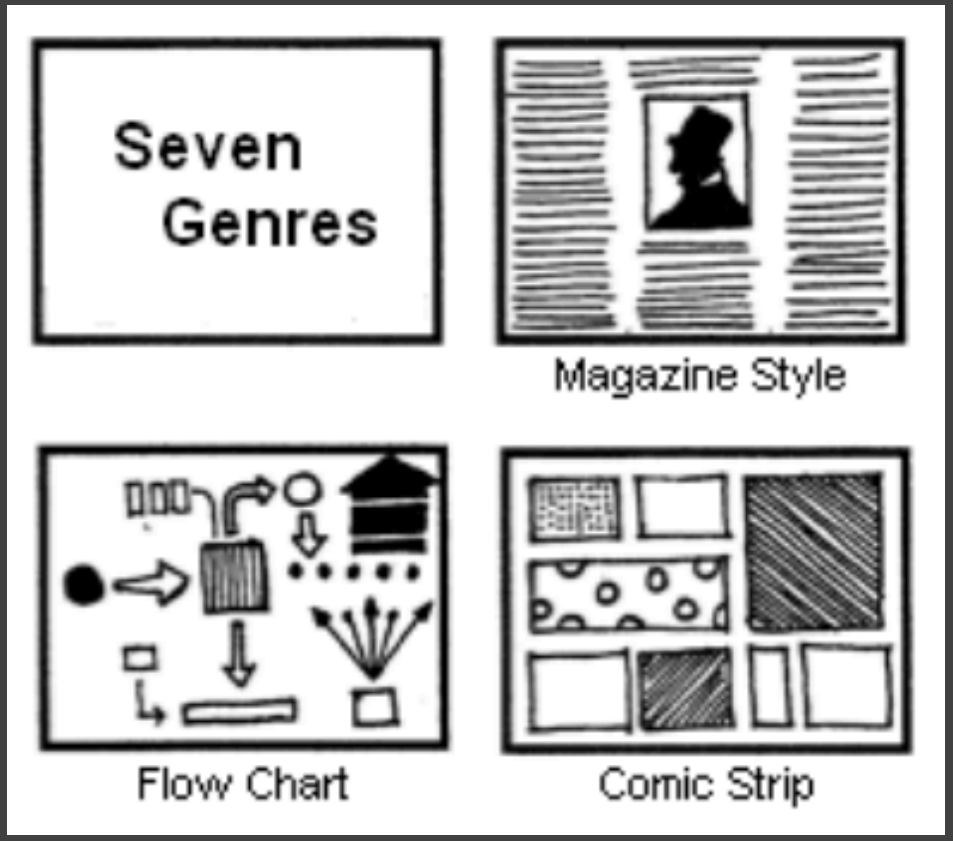
Screenshots of most insightful views *(8+)*

Include titles and captions for each view

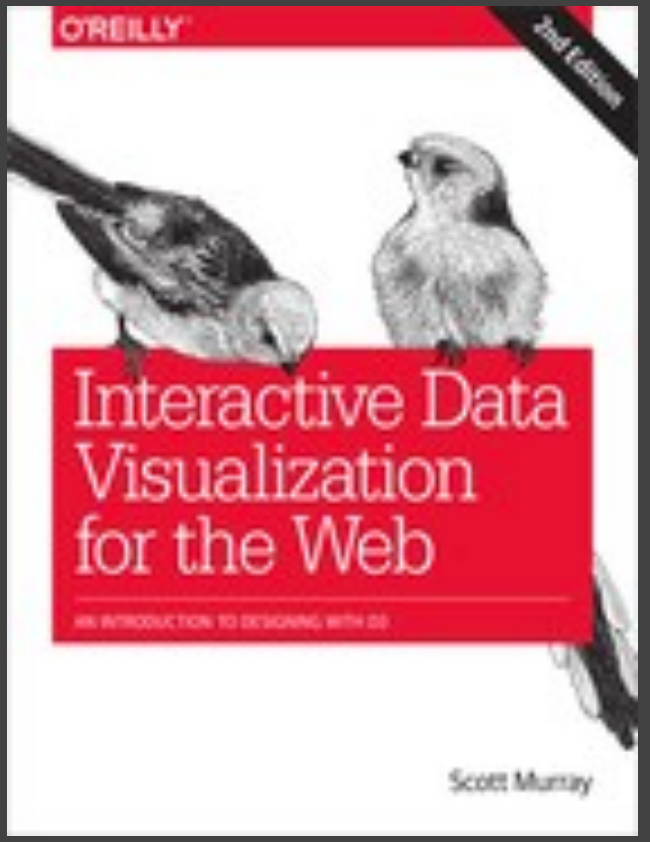
Due by 11:59pm

Monday, Apr 19

Required Readings for Fri 4/16

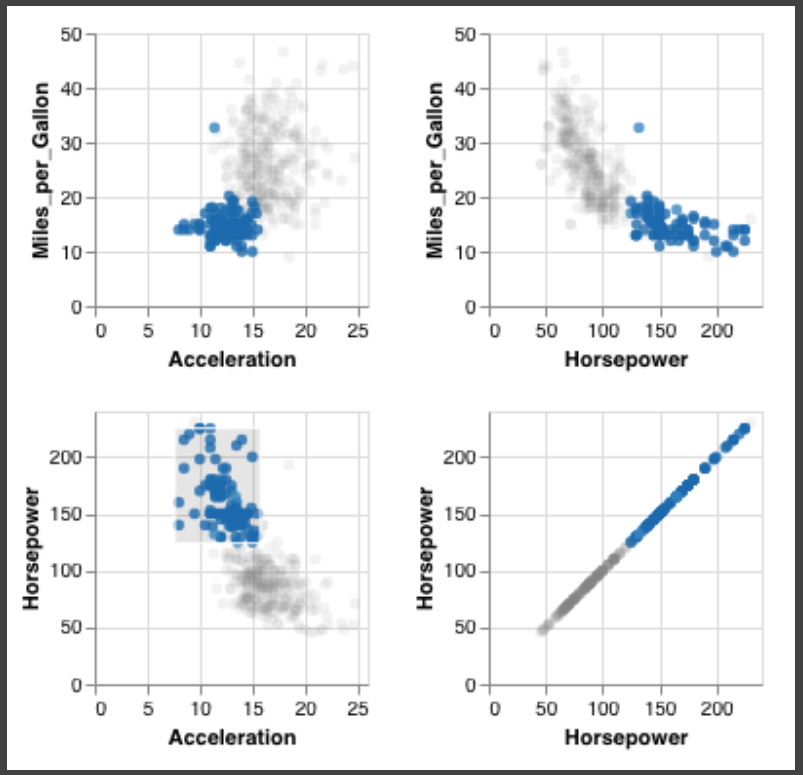


Narrative Visualization: Telling Stories with Data. Edward Segel and Jeffrey Heer. IEEE InfoVis. 2010.

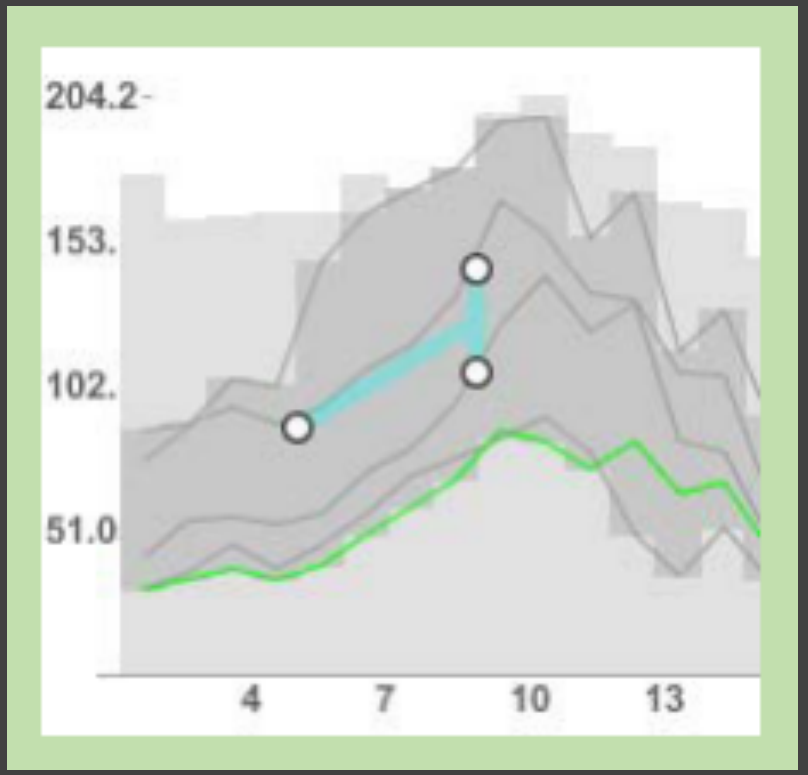


Chapters 2, 4, 5 in Interactive Data Visualization for the Web. Scott Murray.

Required Readings for Mon 4/19

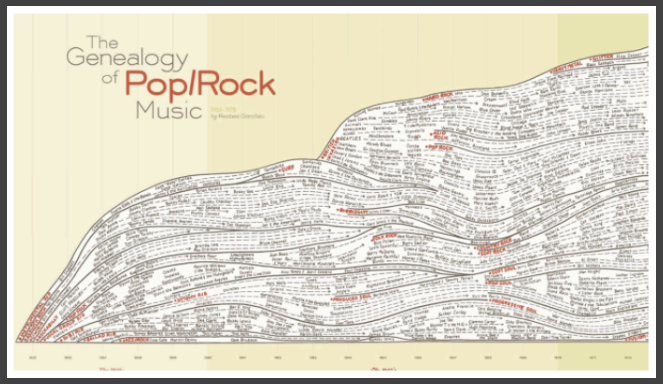


Notebook: Interaction.



Interactive Dynamics for Visual Analysis. Jeffrey Heer and Ben Shneiderman. 2012.

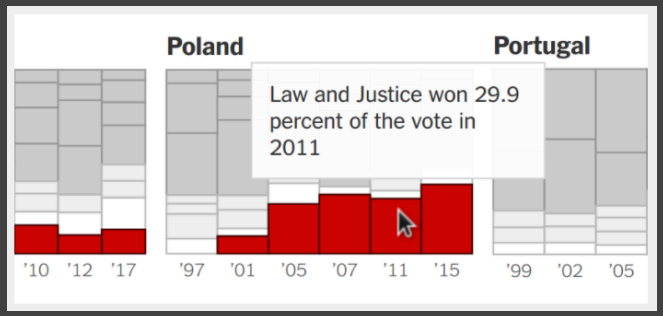
Optional Readings for Week 4



MON The Death of Interactive Infographics?

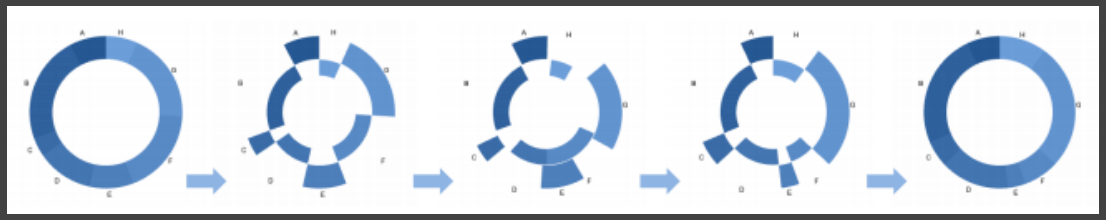
FRI Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts.

FRI The Work that Visualisation Conventions Do.

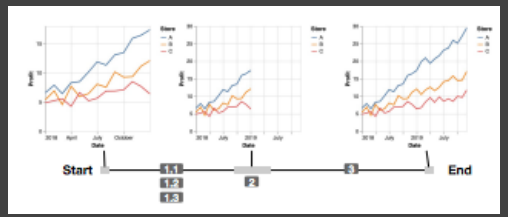


MON In Defense of Interactive Graphics.

WED Effectiveness of Animation in Trend Visualization.



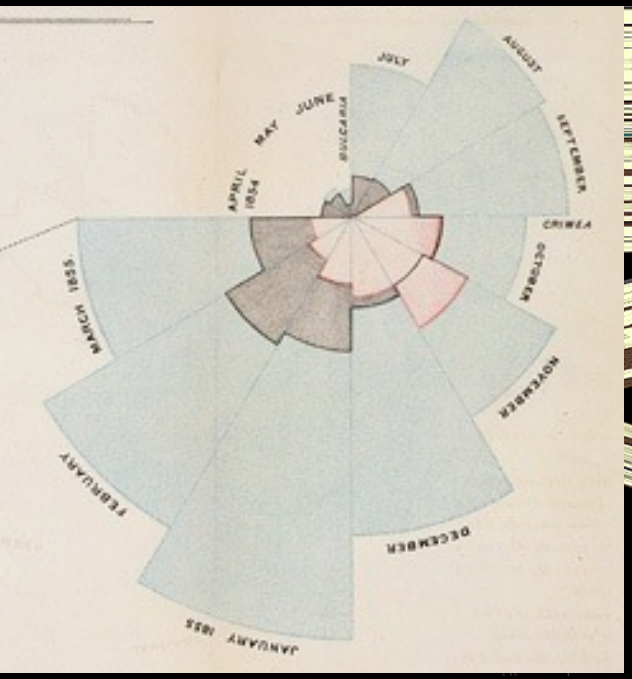
WED Animated Transitions in Statistical Data Graphics.

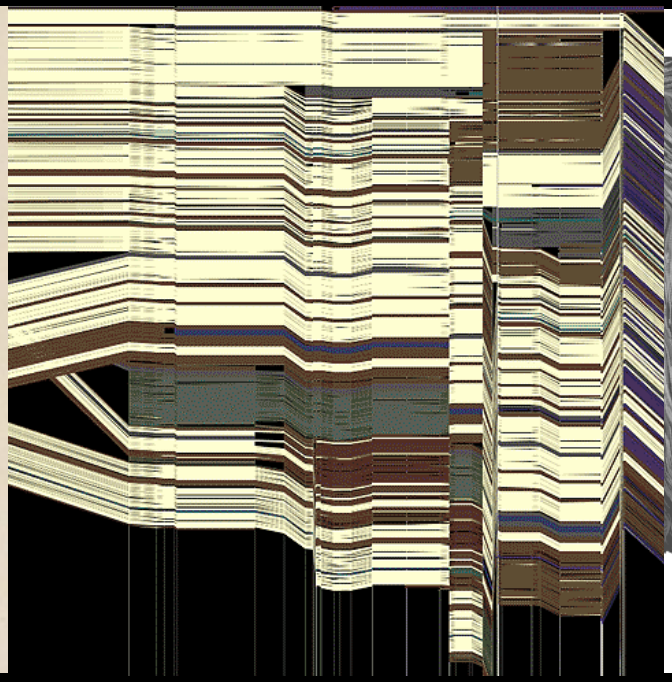


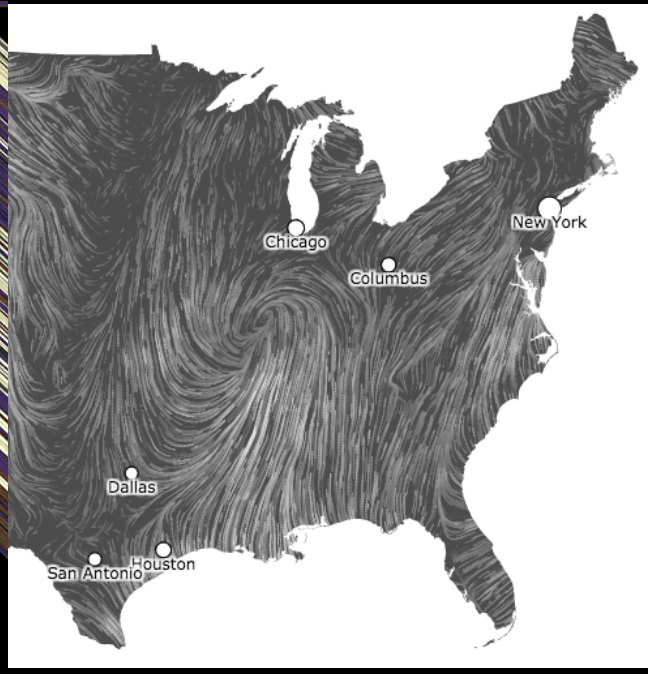
WED Gemini: A Grammar and Recommender System for Animated Transitions in Statistical Graphics.

CSE 412 - Intro to Data Visualization

Responsive Visualization

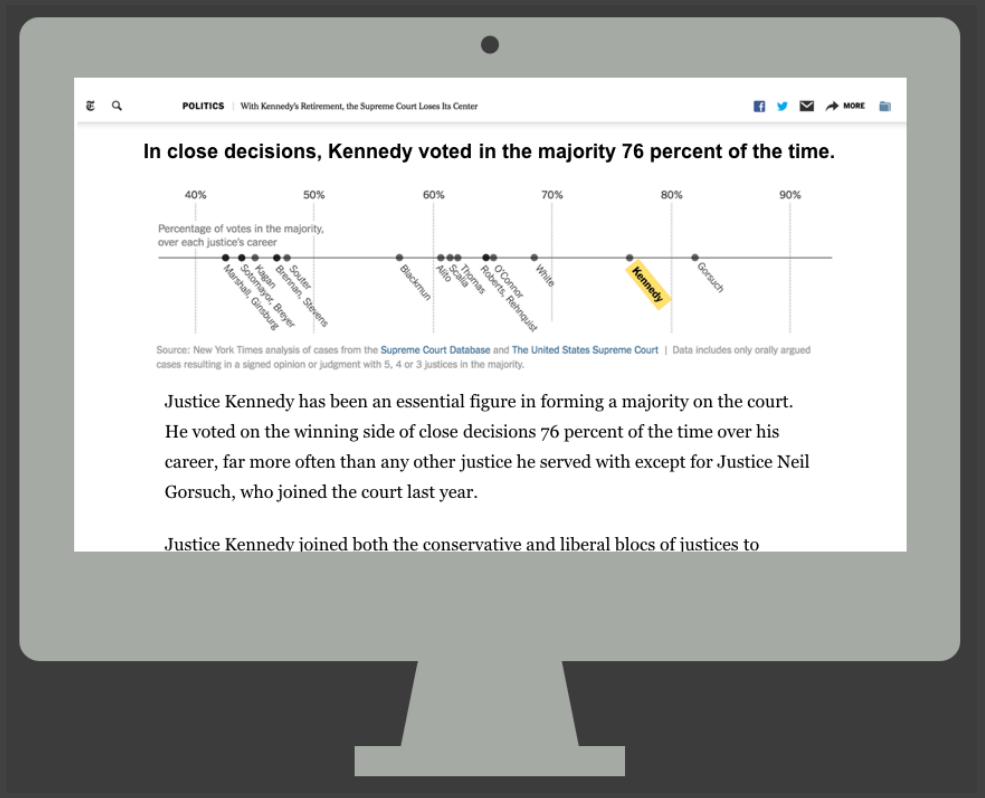


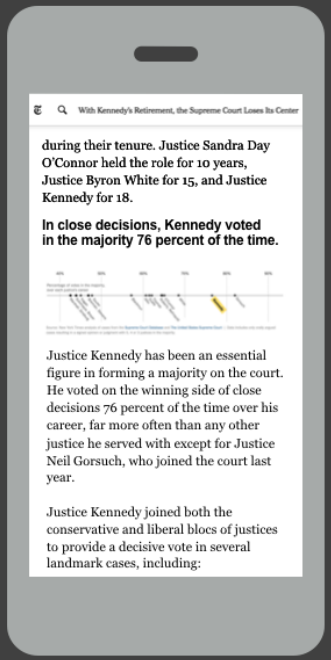


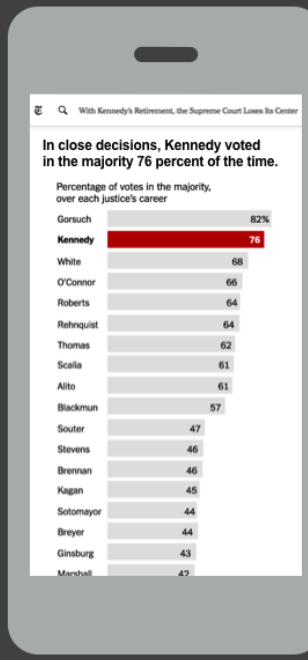


Jane Hoffswell University of Washington

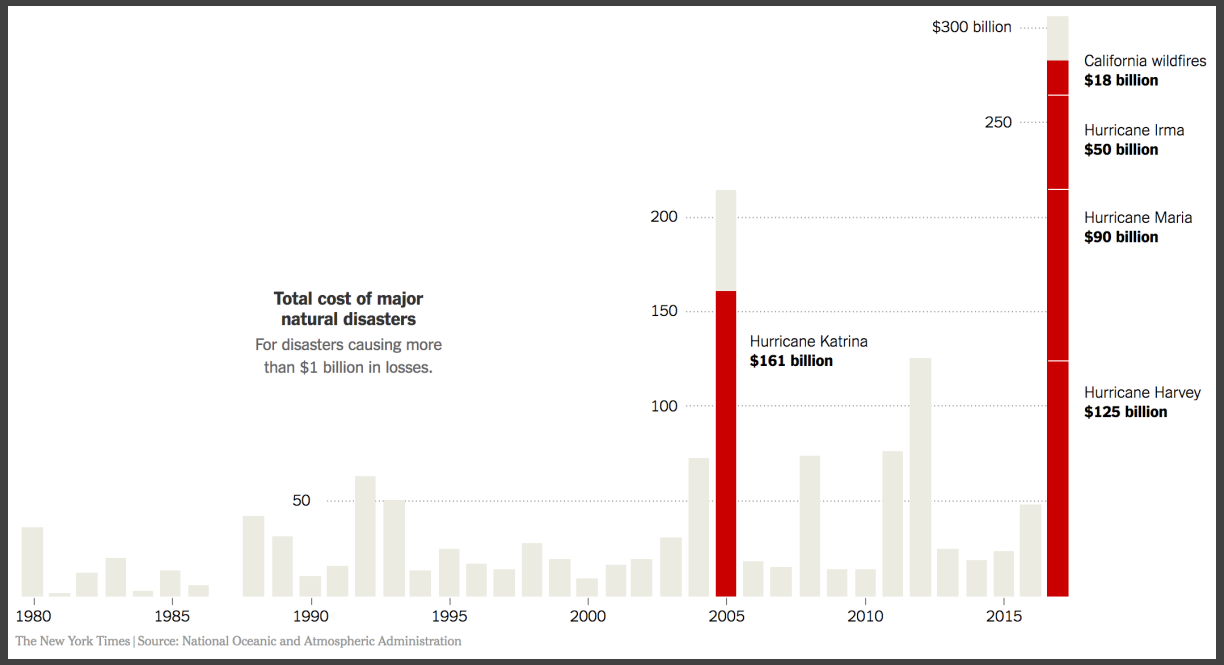
Responsive Visualization

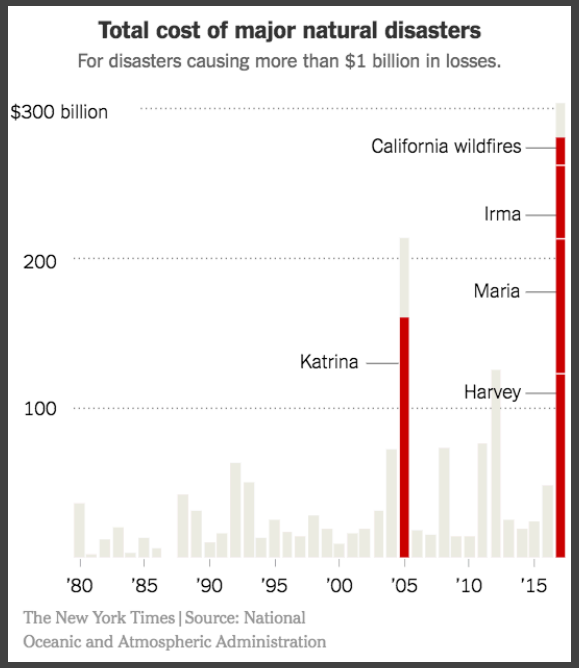




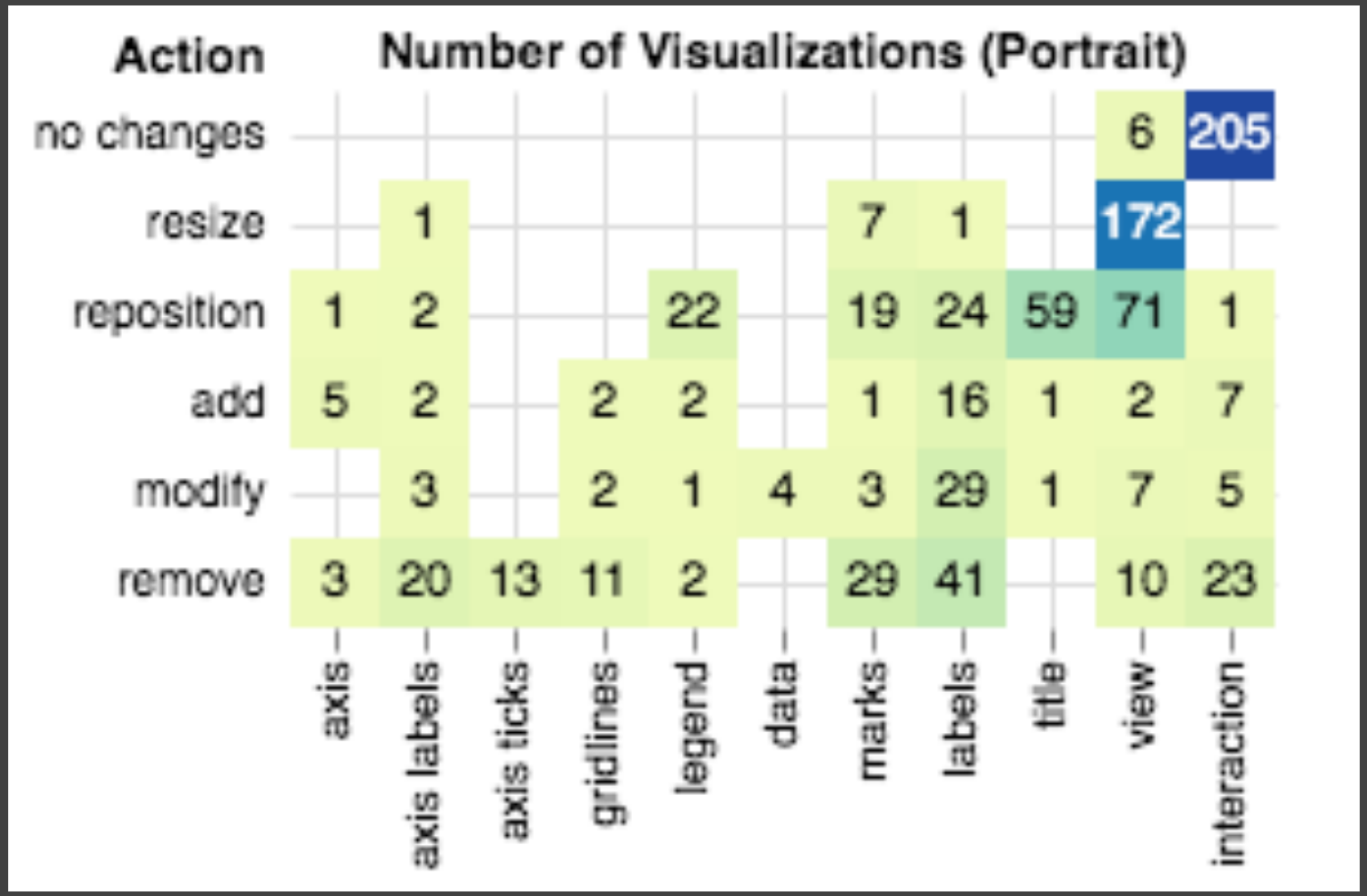


Responsive Visualization [Hoffswell et al. 20]





Responsive Visualization [Hoffswell et al. 20]



Responsive Visualization [Hoffswell et al. 20]

Basic Selection Methods

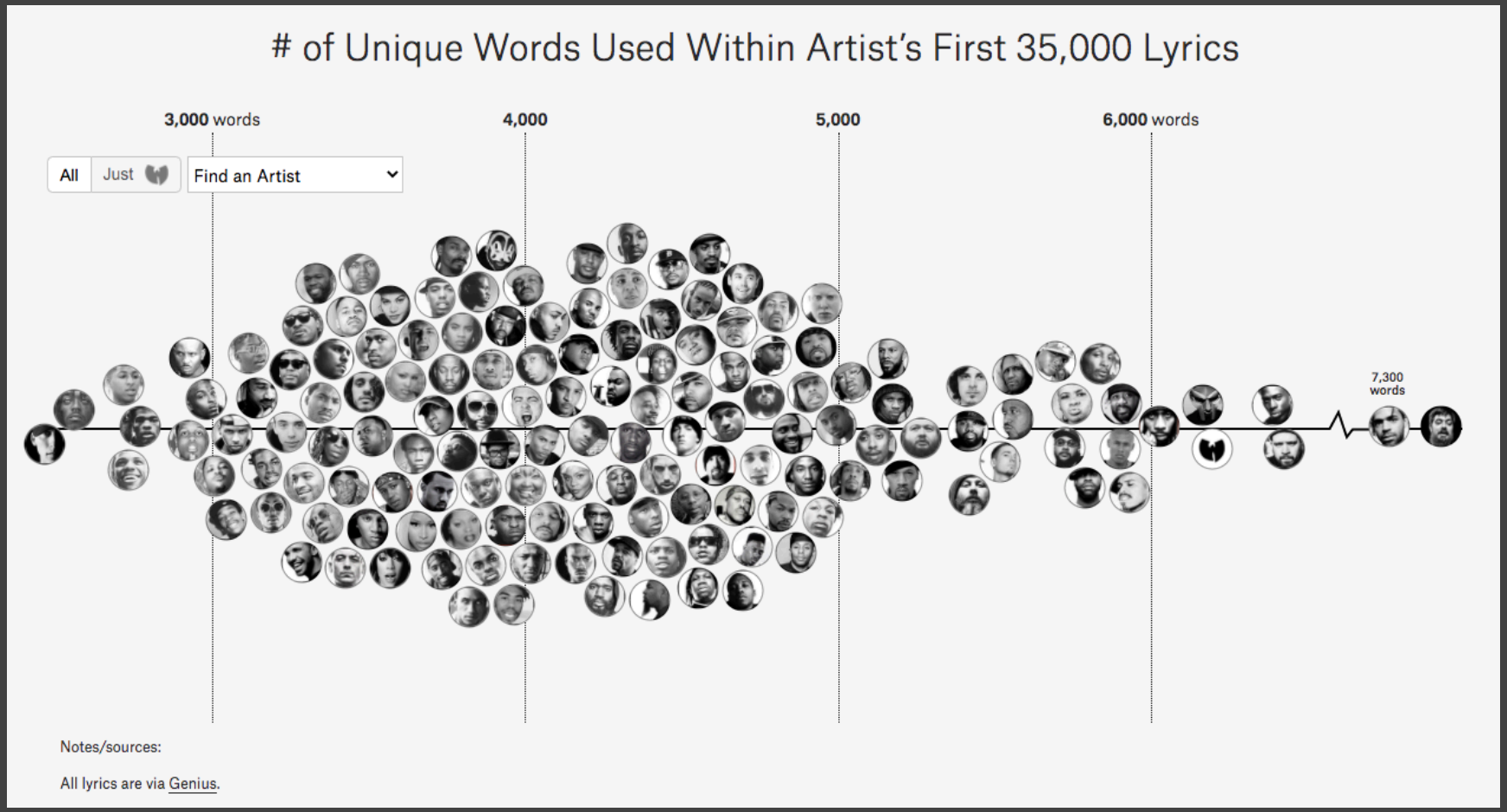
Point Selection

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*

Eurographics Conference on Visualization (EuroVis) 2019
M. Gleicher, H. Leitte, and I. Viola
(Guest Editors)

Volume 38 (2019), Number 3

Capture & Analysis of Active Reading Behaviors for Interactive Articles on the Web

Matthew Conlen[1], Alex Kale[1], and Jeffrey Heer[1]

[1] Paul G. Allen School of Computer Science & Engineering, University of Washington

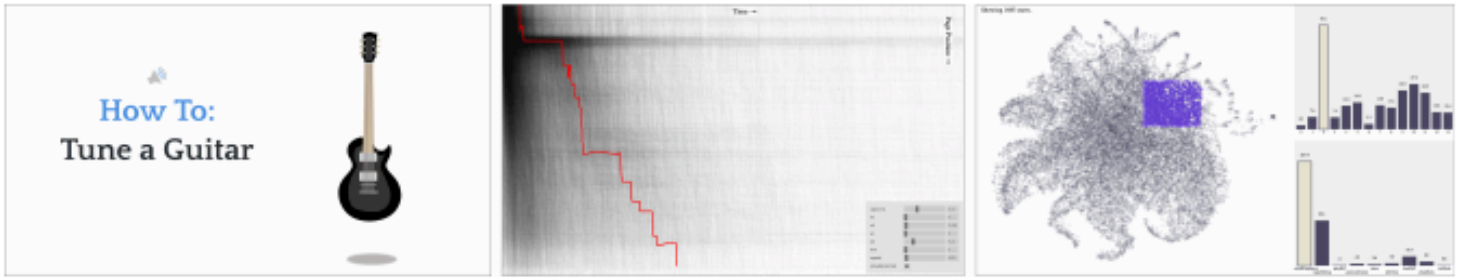


Figure 1: *An online interactive article* How To: Tune A Guitar *(left) and visualizations of collected reader activity data.* HopScroll *(center) visualizes reader progress over time, revealing reading patterns and fixation points.* Readuction *(right) uses dimensionality reduction of reader feature vectors to enable nuanced segment analysis; linked views show timing and event information for selected points. Along with these tools we present Idyll language extensions for automating the collection of detailed log data, and discuss reading patterns discovered.*

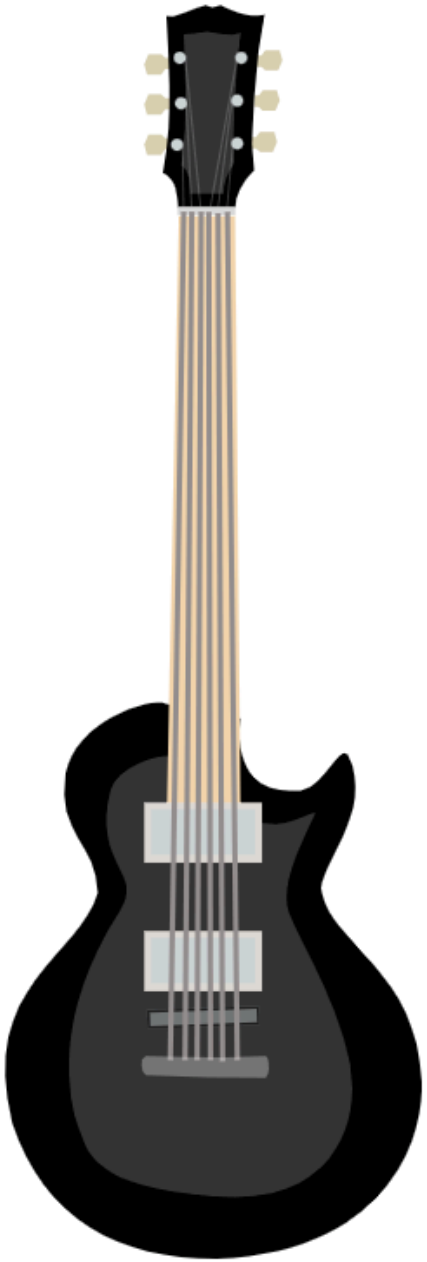
Abstract

Journalists, educators, and technical writers are increasingly publishing interactive content on the web. However, popular analytics tools provide only coarse information about how readers interact with individual pages, and laboratory studies often fail to capture the variability of a real-world audience. We contribute extensions to the Idyll markup language to automate the detailed instrumentation of interactive articles and corresponding visual analysis tools for inspecting reader behavior at both micro- and macro-levels. We present three case studies of interactive articles that were instrumented, posted online, and promoted via social media to reach broad audiences, and share data from over 50,000 reader sessions. We demonstrate the use of our tools to characterize article-specific interaction patterns, compare behavior across desktop and mobile devices, and reveal reading patterns common across articles. Our contributed findings, tools, and corpus of behavioral data can help advance and inform more comprehensive studies of narrative visualization.

How To:

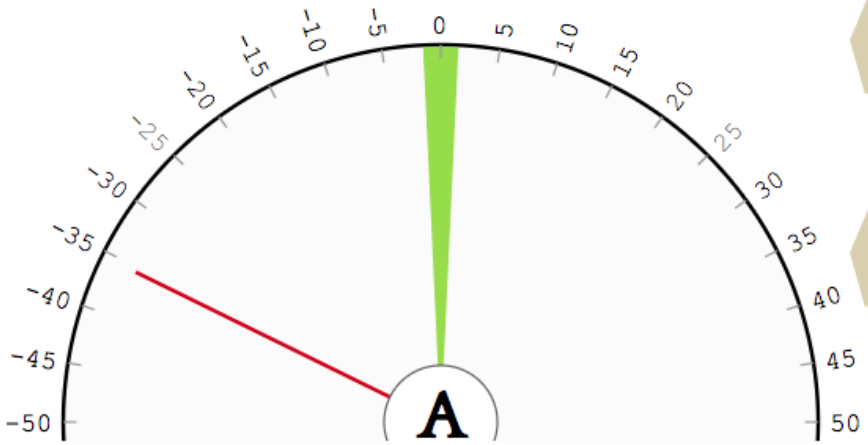
Tune a Guitar

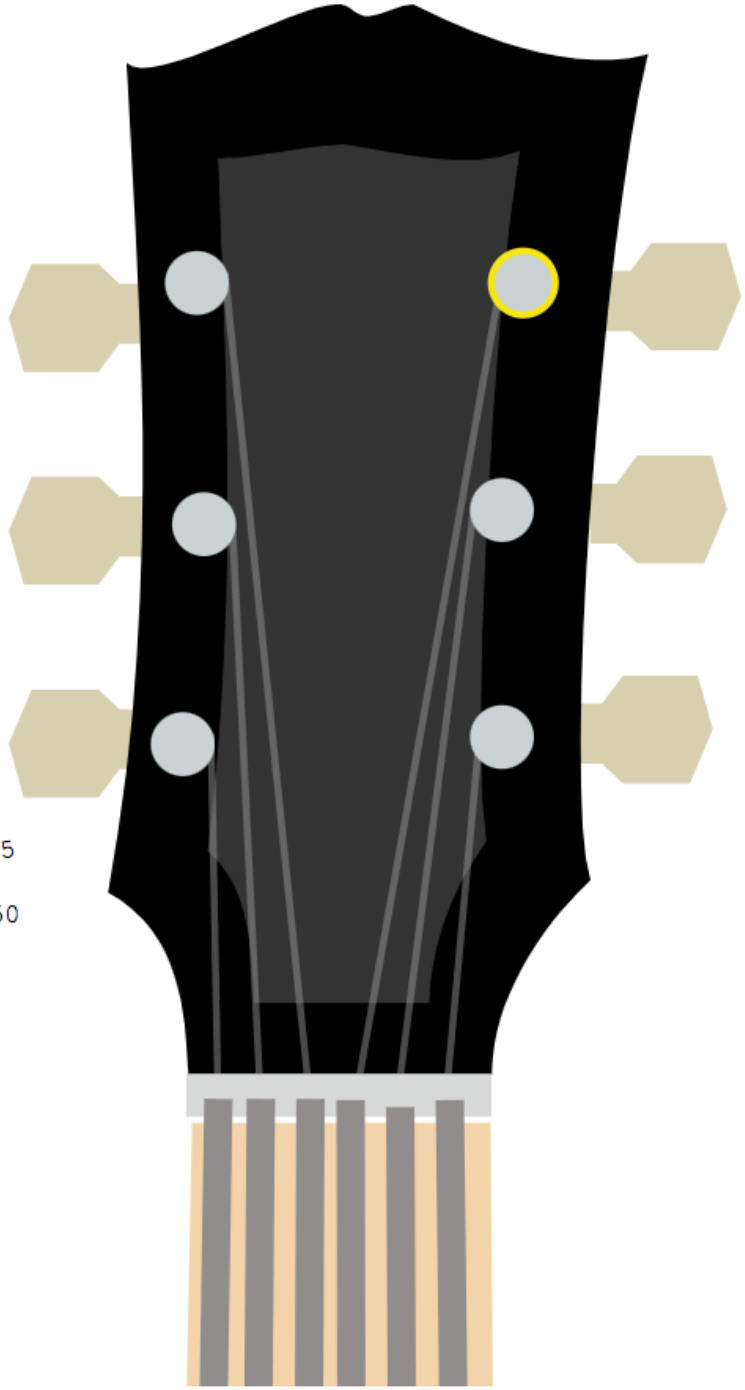
By Matthew Conlen and Alex Kale



Electric Tuner to the Rescue.

Tune the guitar using the tuner. Click and drag the tuning knobs on the right to tighten and loosen the strings.





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.

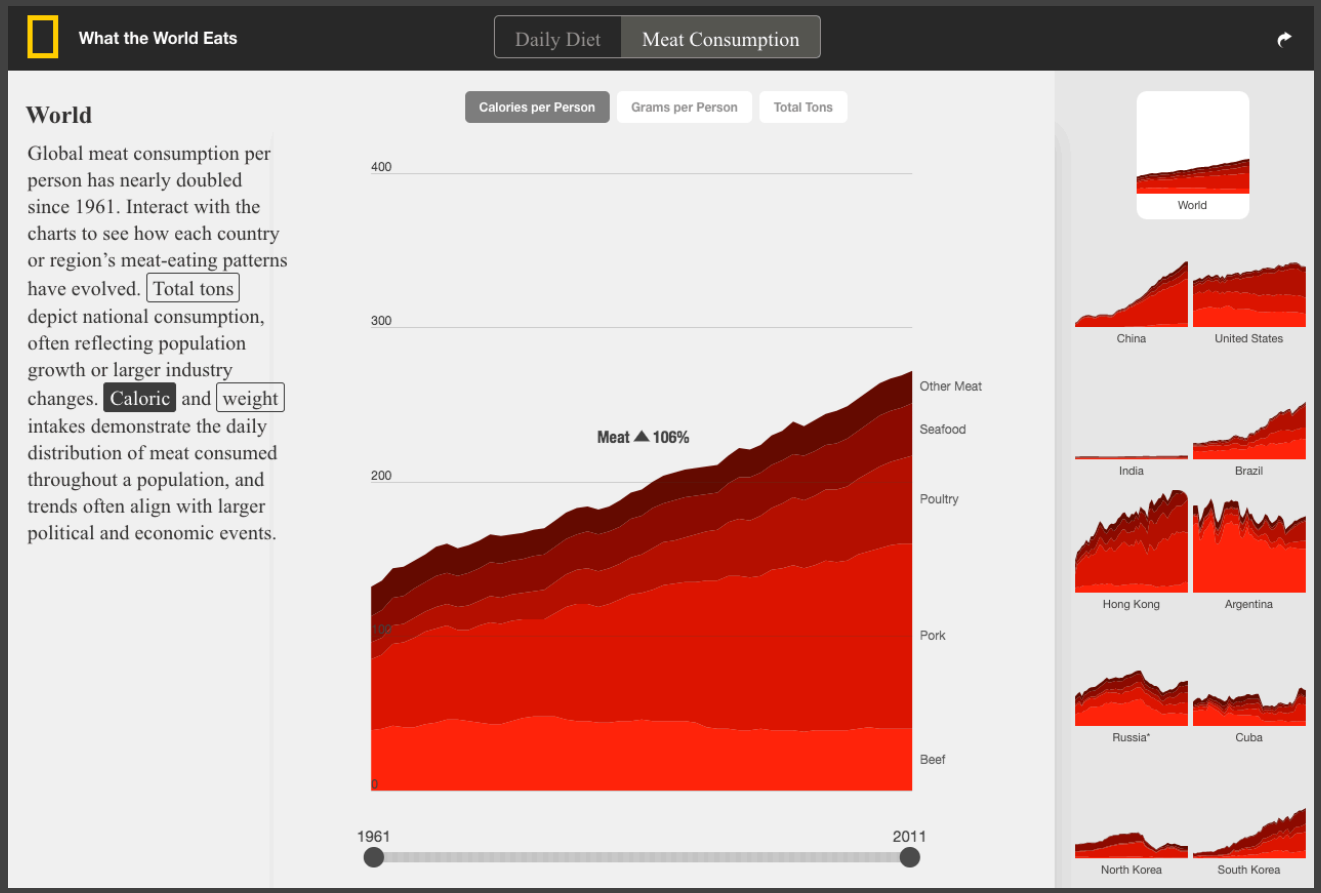
Conlen et al., *EuroVis 2019*

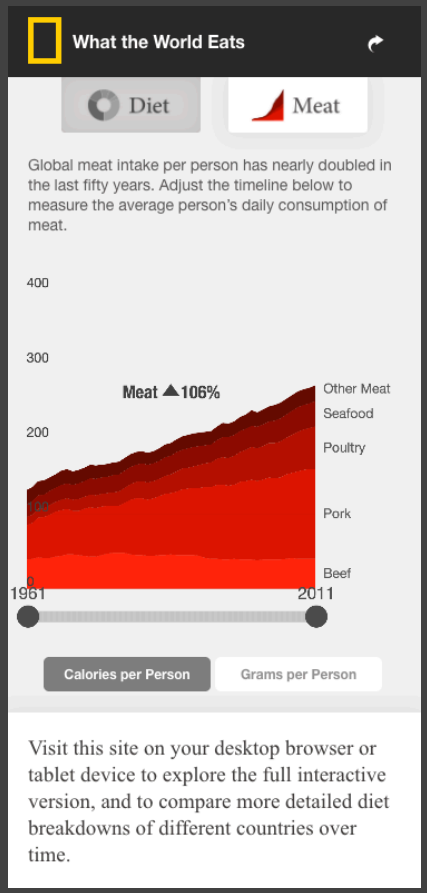
82% of mobile readers advanced through at least some of the content, even though they needed to dismiss a warning about download size; however, only 34% attempted to tune the guitar and just 6% tuned all six strings.

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.

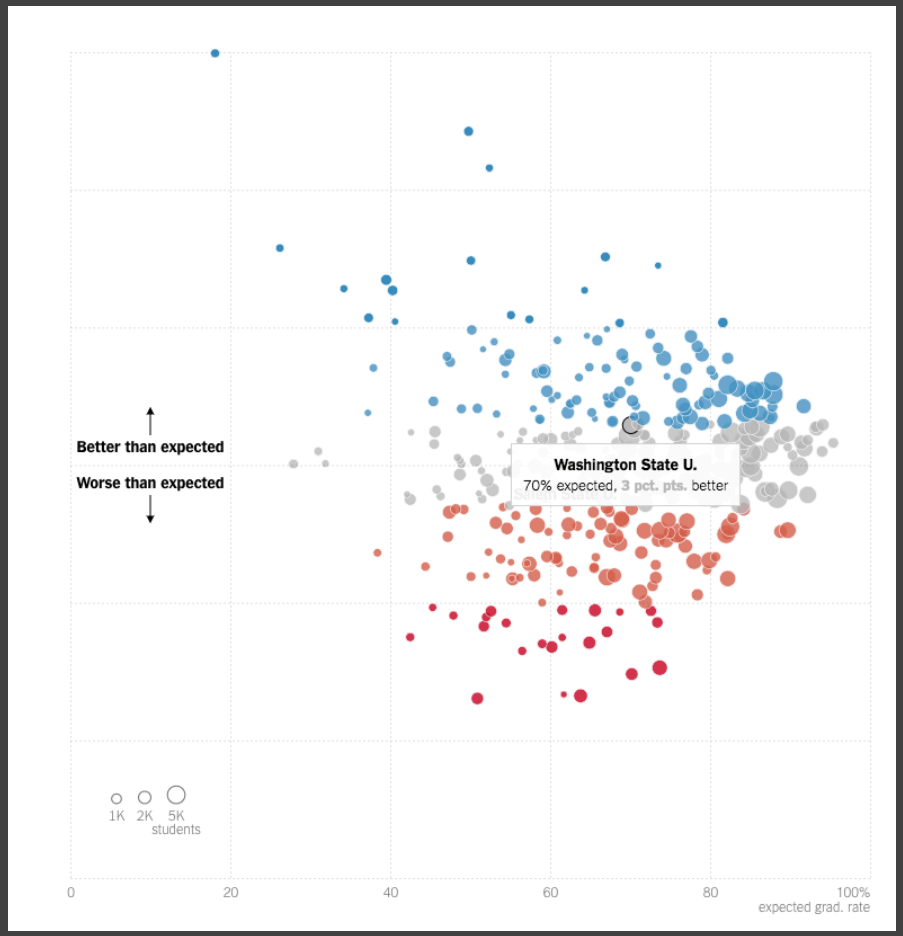
Conlen et al., *EuroVis 2019*

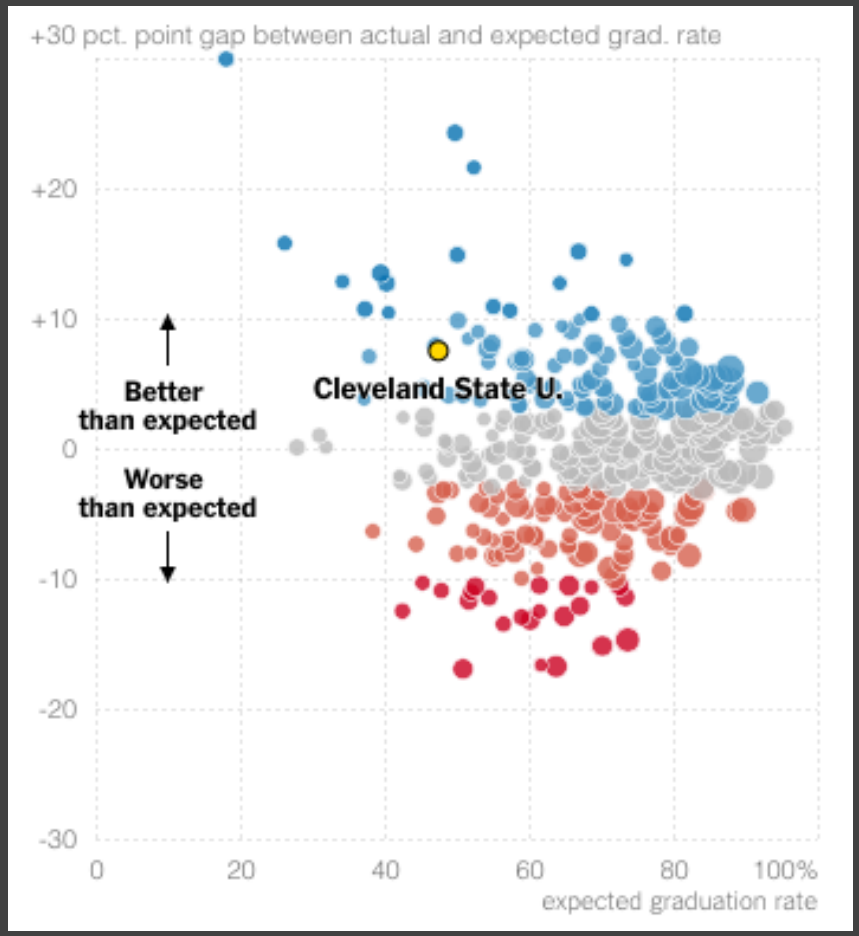
Interactions Disabled



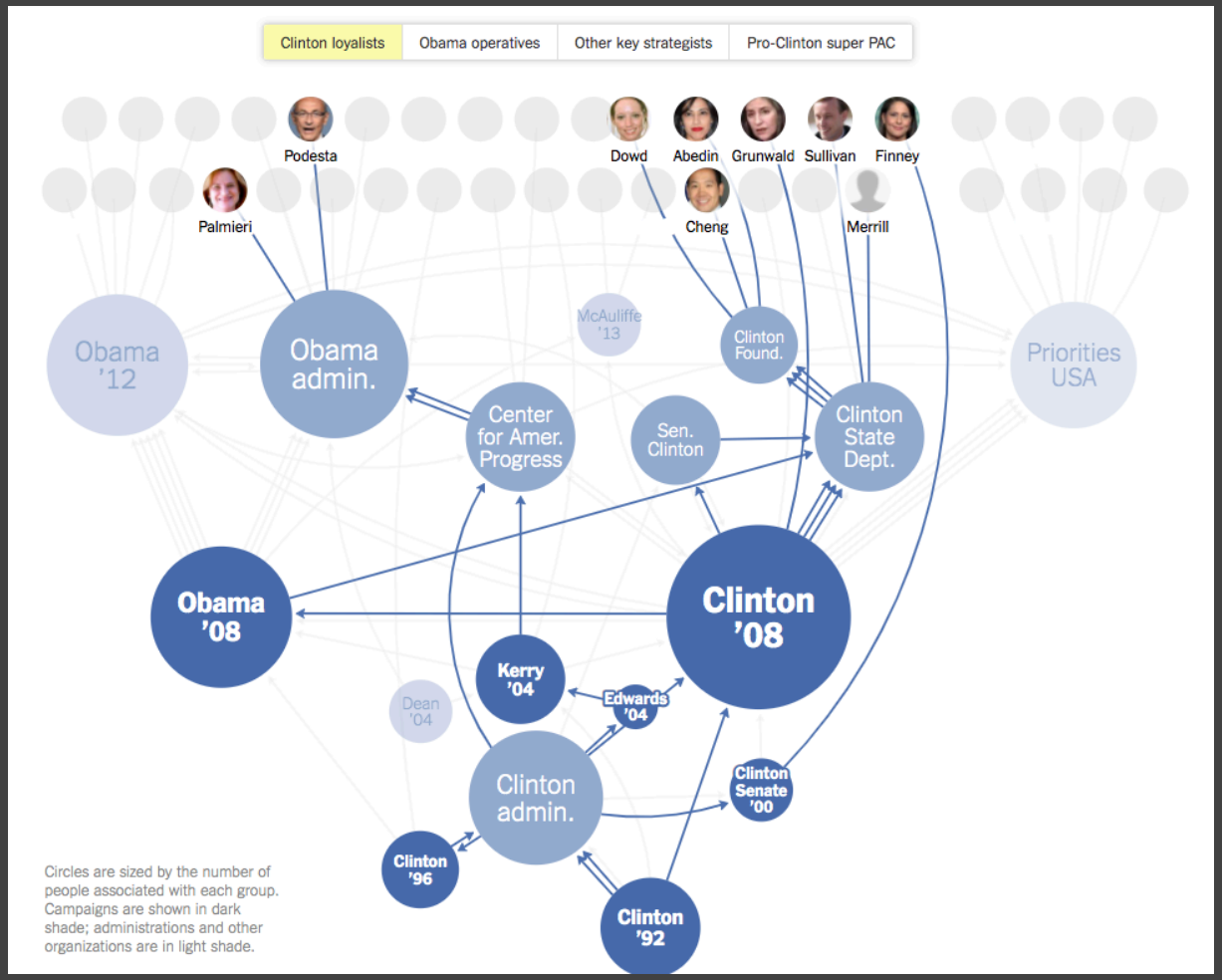


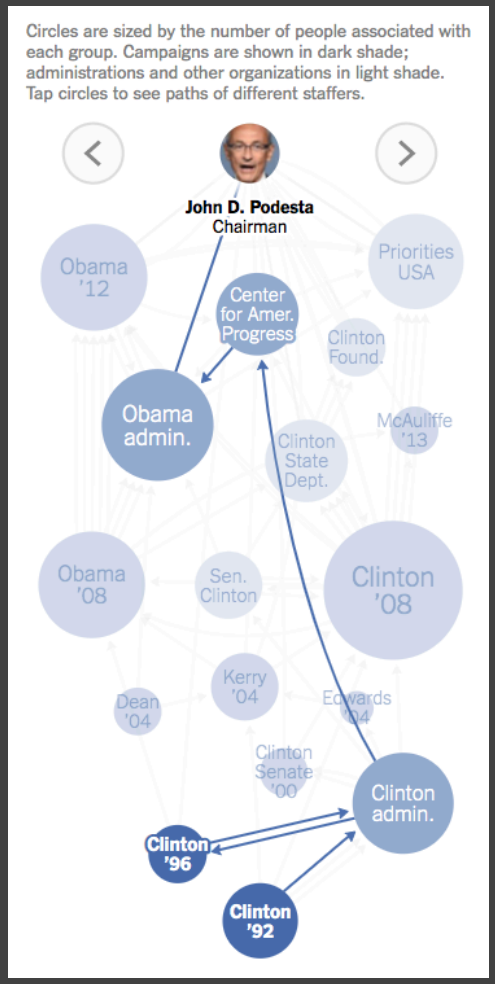
Interactions Previewed





Interactions Simplified





Responsive Visualization Summary

Good visualizations are task dependent
Who is the audience and what is the task?
Pick the right interaction technique

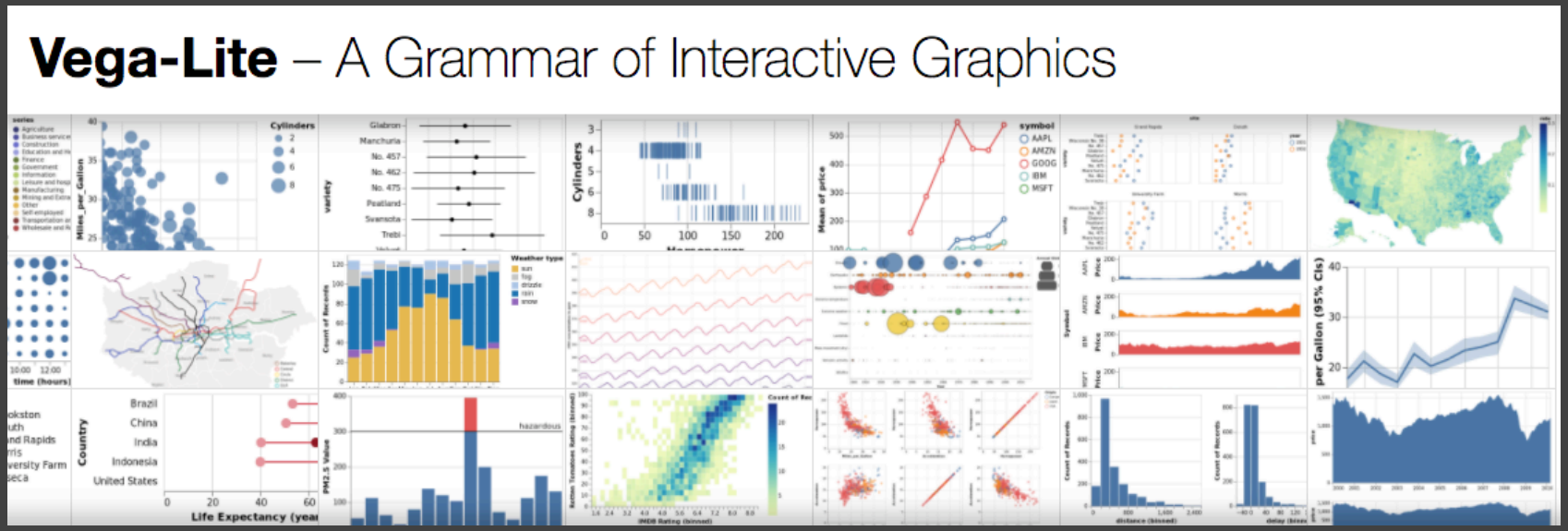
Visualizations are not one size fits all
Context might change user goals

Quiz Section: Vega-Lite

Tomorrow, Thursday April 15th

Hands-on experience with Vega-Lite

Come prepared with questions!



Up Next: Jane's Office Hour (link on Canvas)