EFFECTIVE
Data Visualization

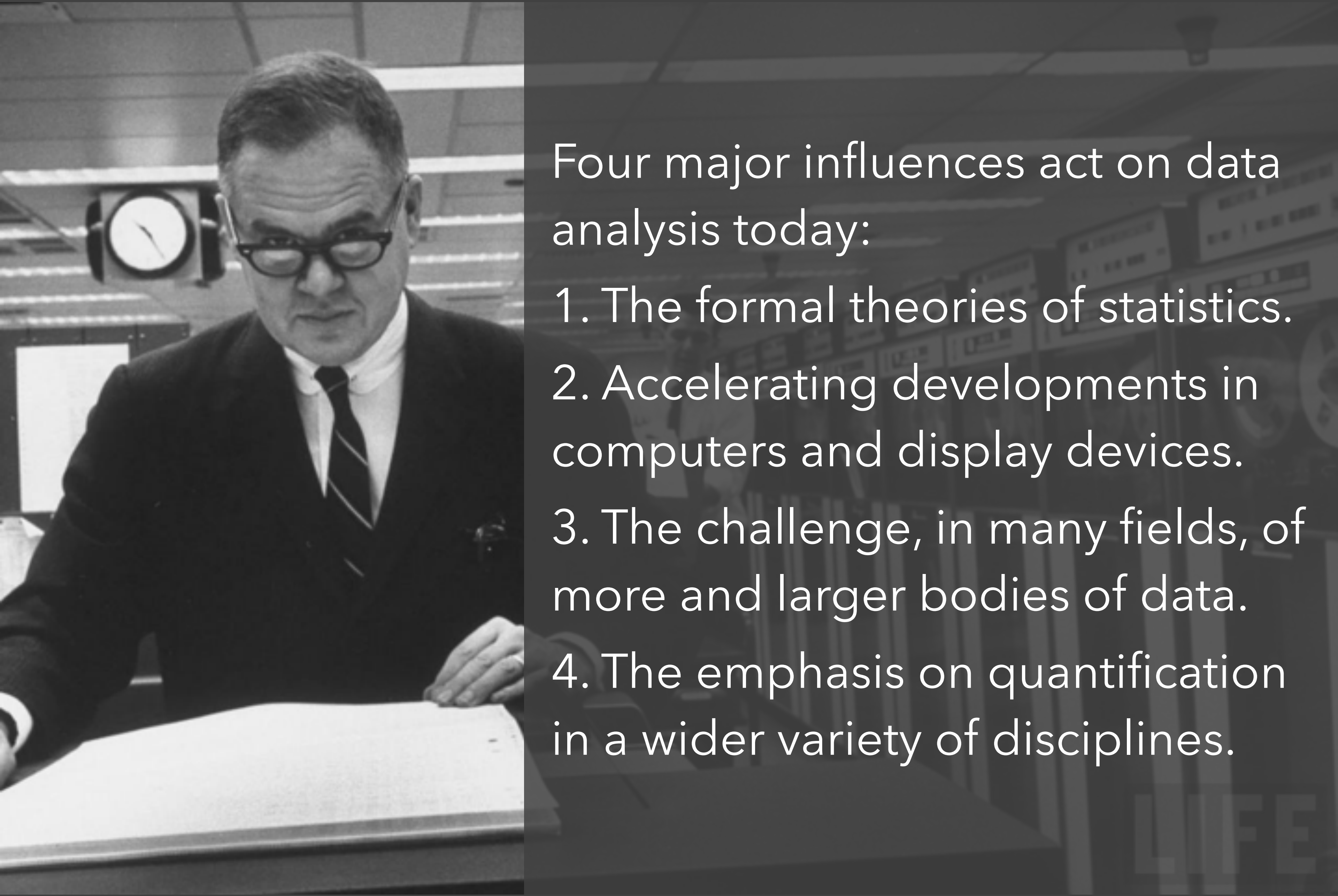
Jeffrey Heer @jeffrey_heer

U. Washington / Trifacta Inc.

Fundamentals

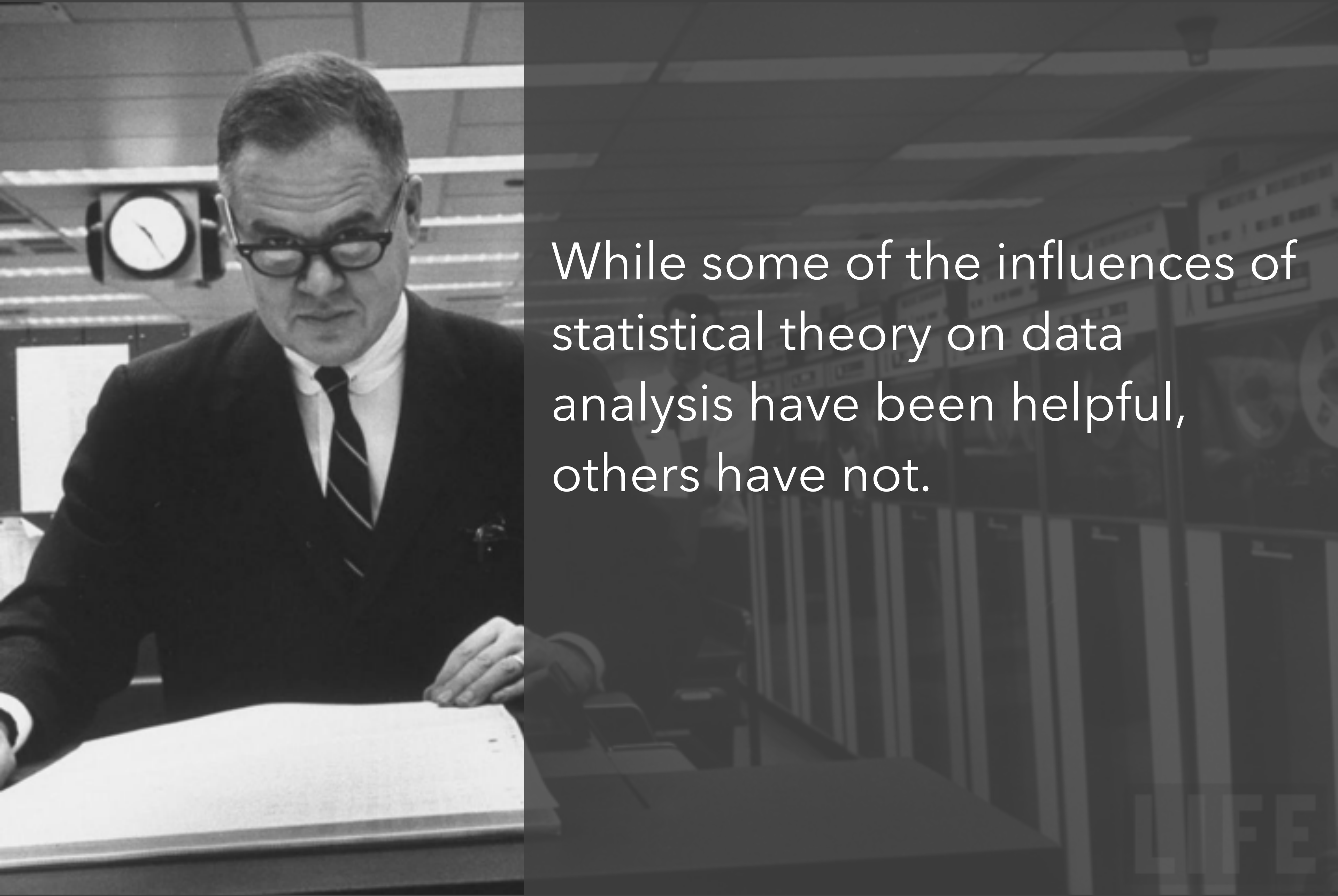
The Value of Visualization

Data Analysis & Statistics, Tukey & Wilk 1965



Four major influences act on data analysis today:

1. The formal theories of statistics.
2. Accelerating developments in computers and display devices.
3. The challenge, in many fields, of more and larger bodies of data.
4. The emphasis on quantification in a wider variety of disciplines.



While some of the influences of statistical theory on data analysis have been helpful, others have not.

Exposure, the effective laying open of the data to display the unanticipated, is to us a major portion of data analysis. Formal statistics has given almost no guidance to exposure; indeed, it is not clear how the **informality** and **flexibility** appropriate to the **exploratory character of exposure** can be fitted into any of the structures of formal statistics so far proposed.

Nothing - not the careful logic of mathematics, not statistical models and theories, not the awesome arithmetic power of modern computers - nothing can substitute here for the **flexibility of the informed human mind**.

Accordingly, both approaches and techniques need to be structured so as to **facilitate human involvement and intervention**.

Set A		Set B		Set C		Set D	
X	Y	X	Y	X	Y	X	Y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.11	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Summary Statistics

$u_X = 9.0$ $\sigma_X = 3.317$

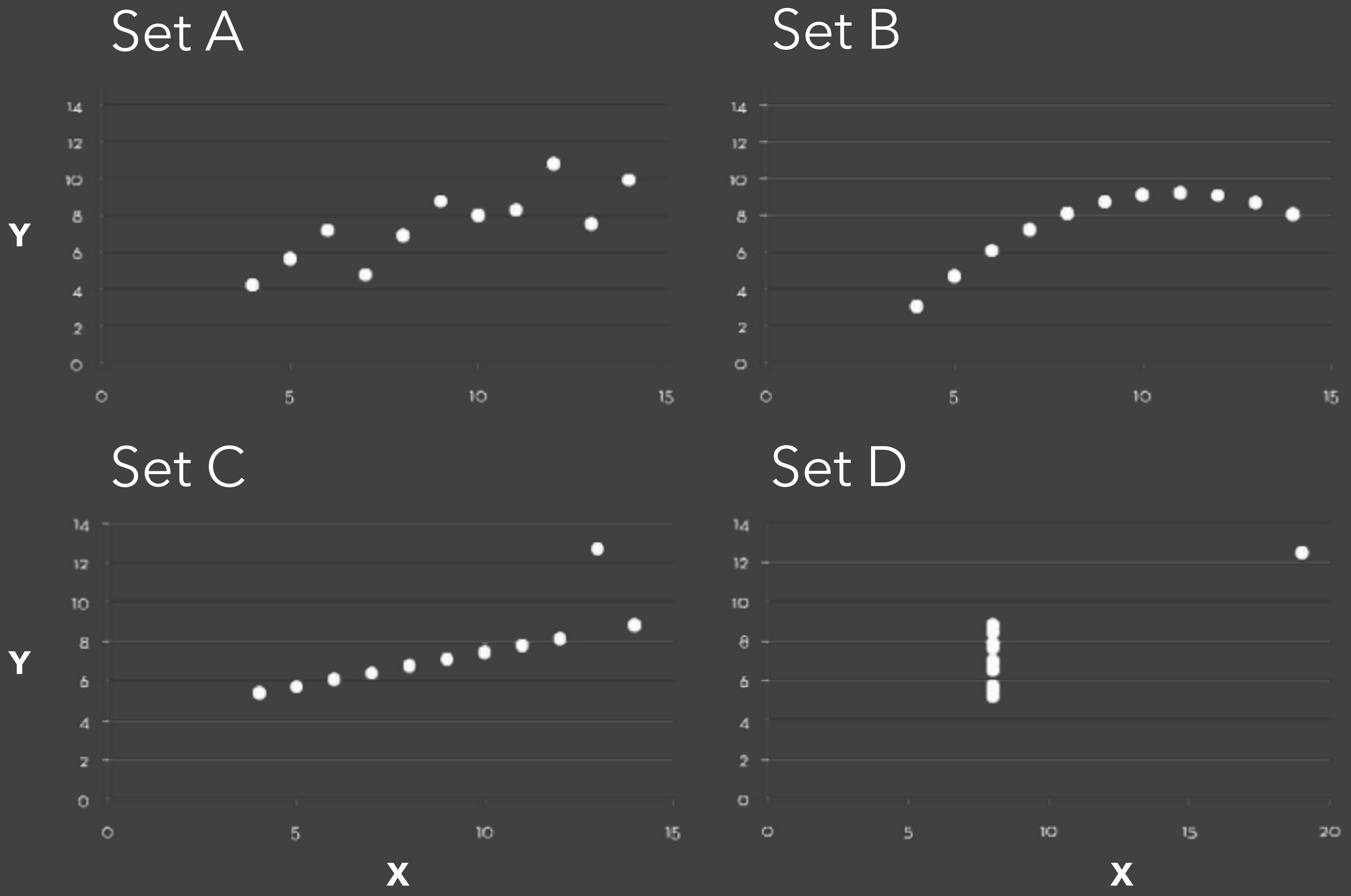
$u_Y = 7.5$ $\sigma_Y = 2.03$

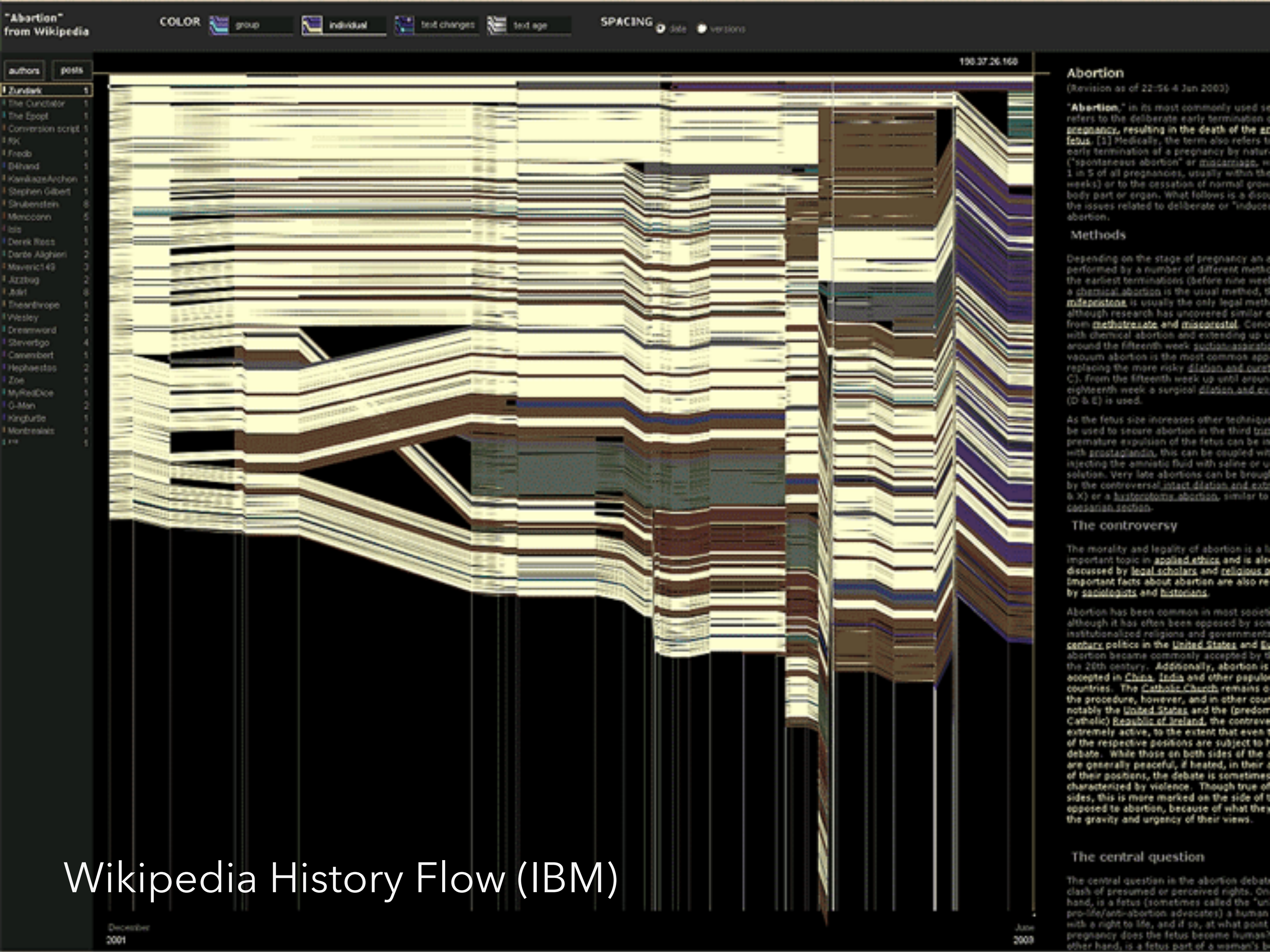
Linear Regression

$Y^2 = 3 + 0.5 X$

$R^2 = 0.67$

[Anscombe 1973]


Set A
Set B
Set C
Set D
Y
Y
X
X
0
2
4
6
8
10
12
14
0
5
10
15
20



Zundark	1
The Cunctator	1
The Epopt	1
Conversion script	1
RK	1
Fredb	1
D4hand	1
KamikazeArchon	1
Stephen Gilbert	1
Strubenstein	8
Mkmcconn	5
Isis	1
Derek Ross	1
Dante Alighieri	2
Maveric149	3
Jizzbug	2
Jtdirl	8
Theanthrope	1
Wesley	2
Dreamword	1
Stevertigo	4
Camembert	1
Hephaestos	2
Zoe	1
MyRedDice	1
O-Man	2
Kingturtle	1
Montrealais	1
[illegible]	1

Abortion

(Revision as of 22:56 4 Jun 2003)

"**Abortion**," in its most commonly used se
refers to the deliberate early termination o
pregnancy, **resulting in the death of the** ej
fetus. [1] Medically, the term also refers t
early termination of a pregnancy by natur
("spontaneous abortion" or miscarriage, w
1 in 5 of all pregnancies, usually within the
weeks) or to the cessation of normal grow
body part or organ. What follows is a disc
the issues related to deliberate or "induce
abortion.

Methods

Depending on the stage of pregnancy an a
performed by a number of different metho
the earliest terminations (before nine wee
a chemical abortion is the usual method, t
mifepristone is usually the only legal meth
although research has uncovered similar e
from methotrexate and misoprostol. Conc
with chemical abortion and extending up u
around the fifteenth week suction-aspirati
vacuum abortion is the most common app
replacing the more risky dilation and curet
C). From the fifteenth week up until aroun
eighteenth week a surgical dilation and ev
(D & E) is used.

As the fetus size increases other techniqu
be used to secure abortion in the third tri
premature expulsion of the fetus can be in
with prostaglandin, this can be coupled wi
injecting the amniotic fluid with saline or u
solution. Very late abortions can be broug
by the controversial intact dilation and ext
& X) or a hysterotomy abortion, similar to
caesarian section.

The controversy

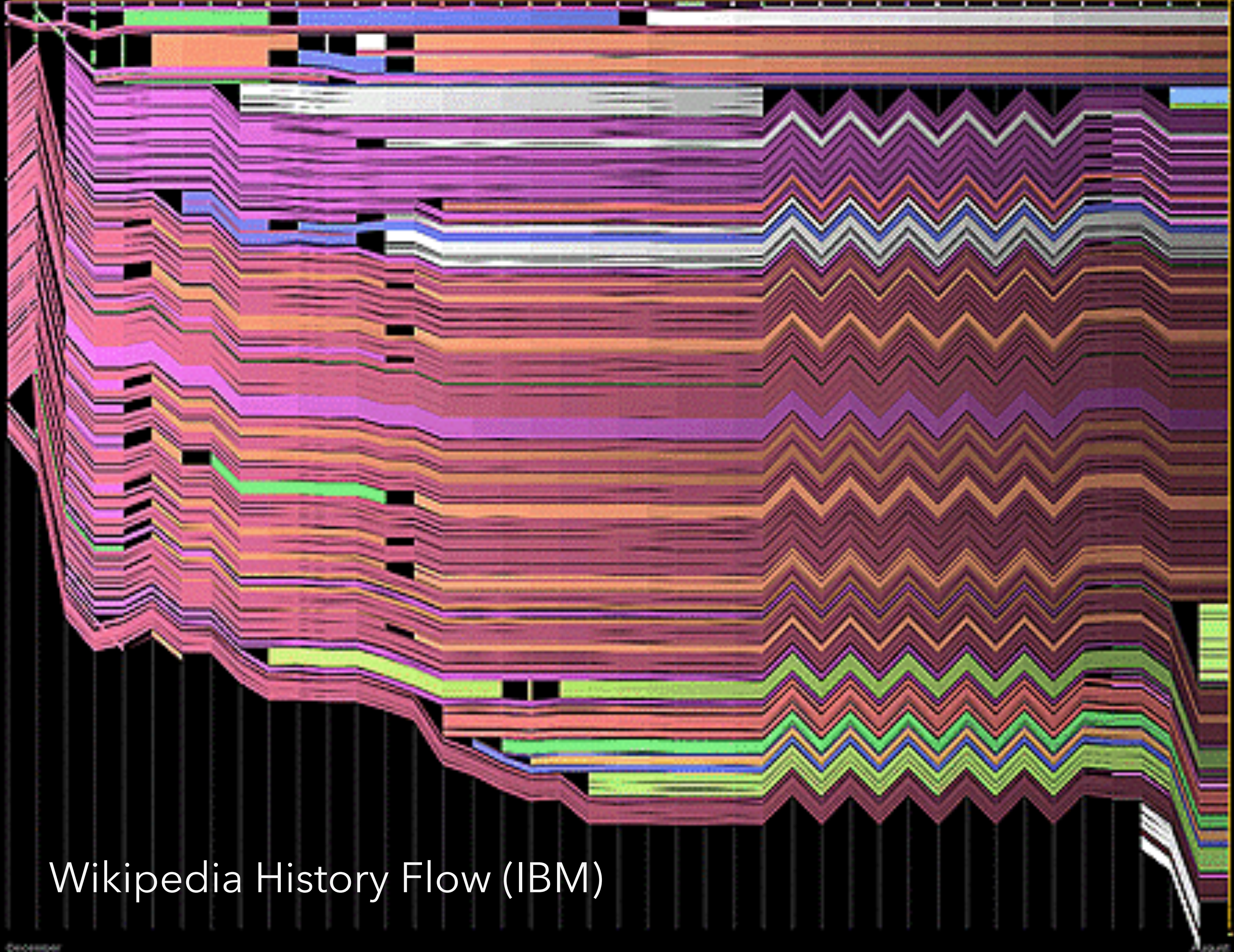
The morality and legality of abortion is a l
important topic in applied ethics and is als
discussed by legal scholars **and** religious p
Important facts about abortion are also re
by sociologists **and** historians.

Abortion has been common in most societ
although it has often been opposed by so
institutionalized religions and government
century **politics in the** United States **and** Eu
abortion became commonly accepted by t
the 20th century. **Additionally, abortion is**
accepted in China, India **and other popula**
countries. The Catholic Church **remains o**
the procedure, however, and in other cou
notably the United States **and the (predom**
Catholic) Republic of Ireland, **the controve**
extremely active, to the extent that even t
of the respective positions are subject to h
debate. While those on both sides of the d
are generally peaceful, if heated, in their a
of their positions, the debate is sometimes
characterized by violence. Though true of
sides, this is more marked on the side of t
opposed to abortion, because of what the
the gravity and urgency of their views.

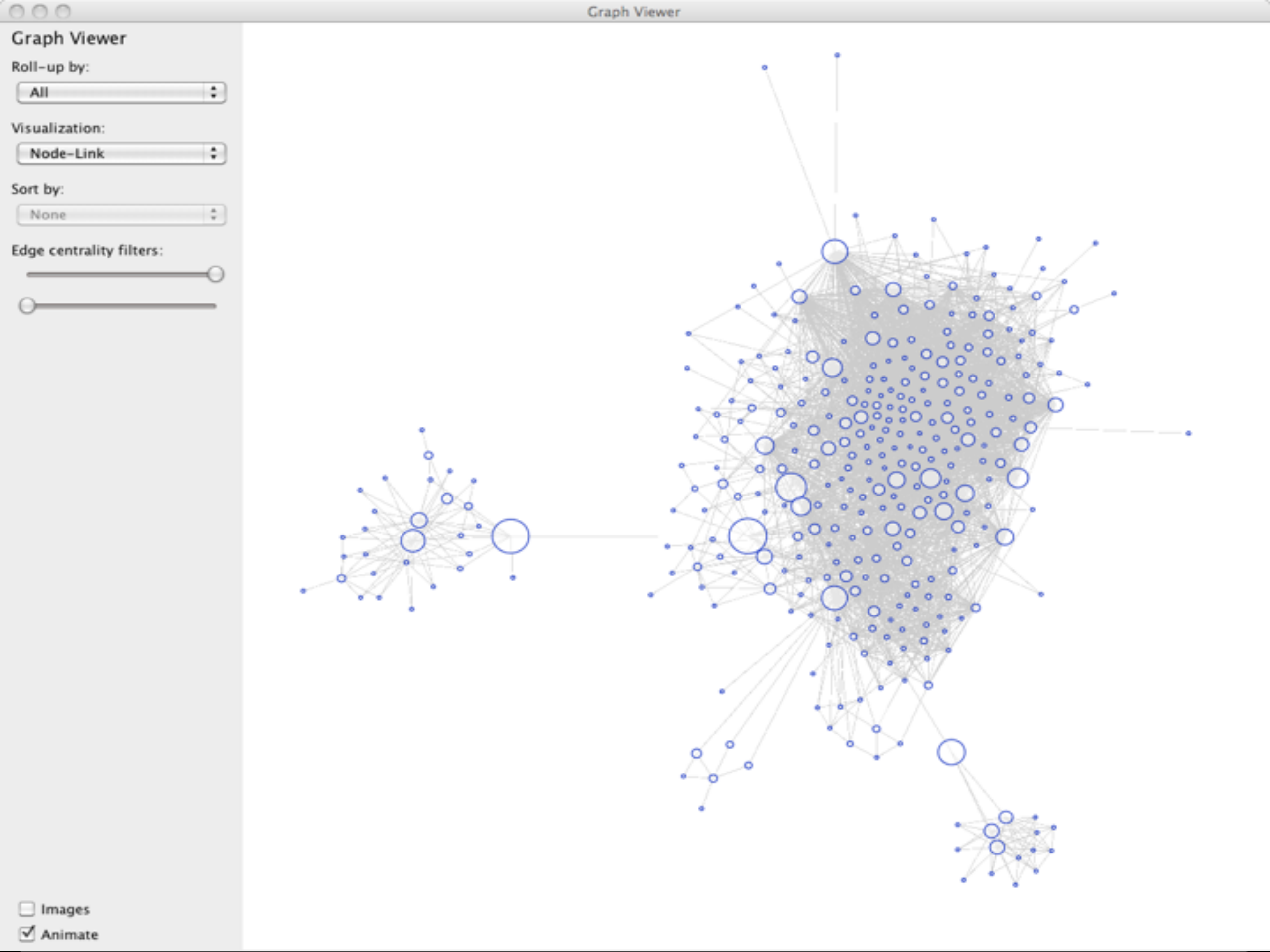
The central question

The central question in the abortion debat
clash of presumed or perceived rights. On
hand, is a fetus (sometimes called the "un
pro-life/anti-abortion advocates) a human
with a right to life, and if so, at what point
pregnancy does the fetus become human?
other hand, is a fetus part of a woman's b

Wikipedia History Flow (IBM)



Wikipedia History Flow (IBM)



Graph Viewer

Roll-up by: All

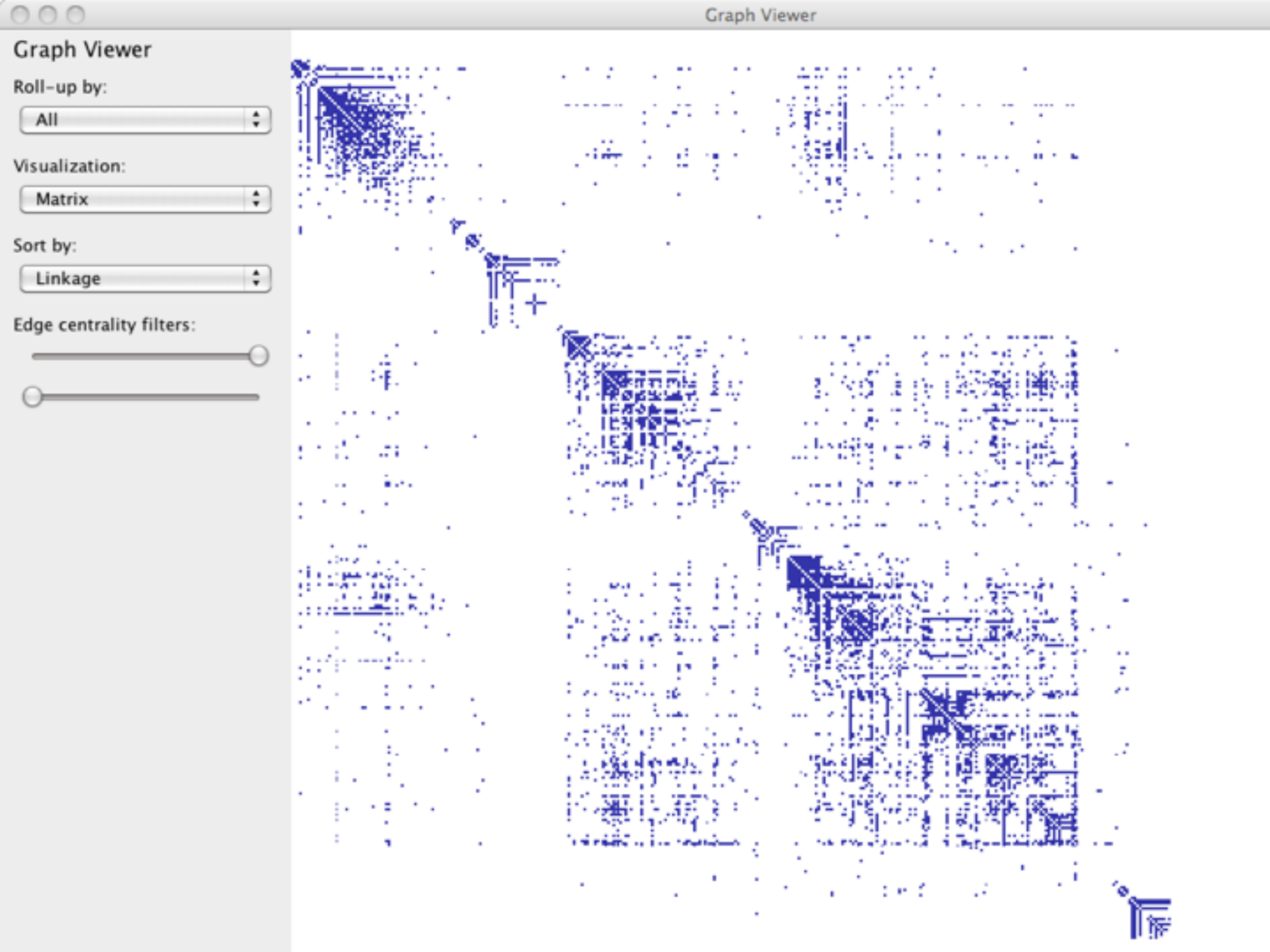
Visualization: Node-Link

Sort by: None

Edge centrality filters:

□ Images

☑ Animate



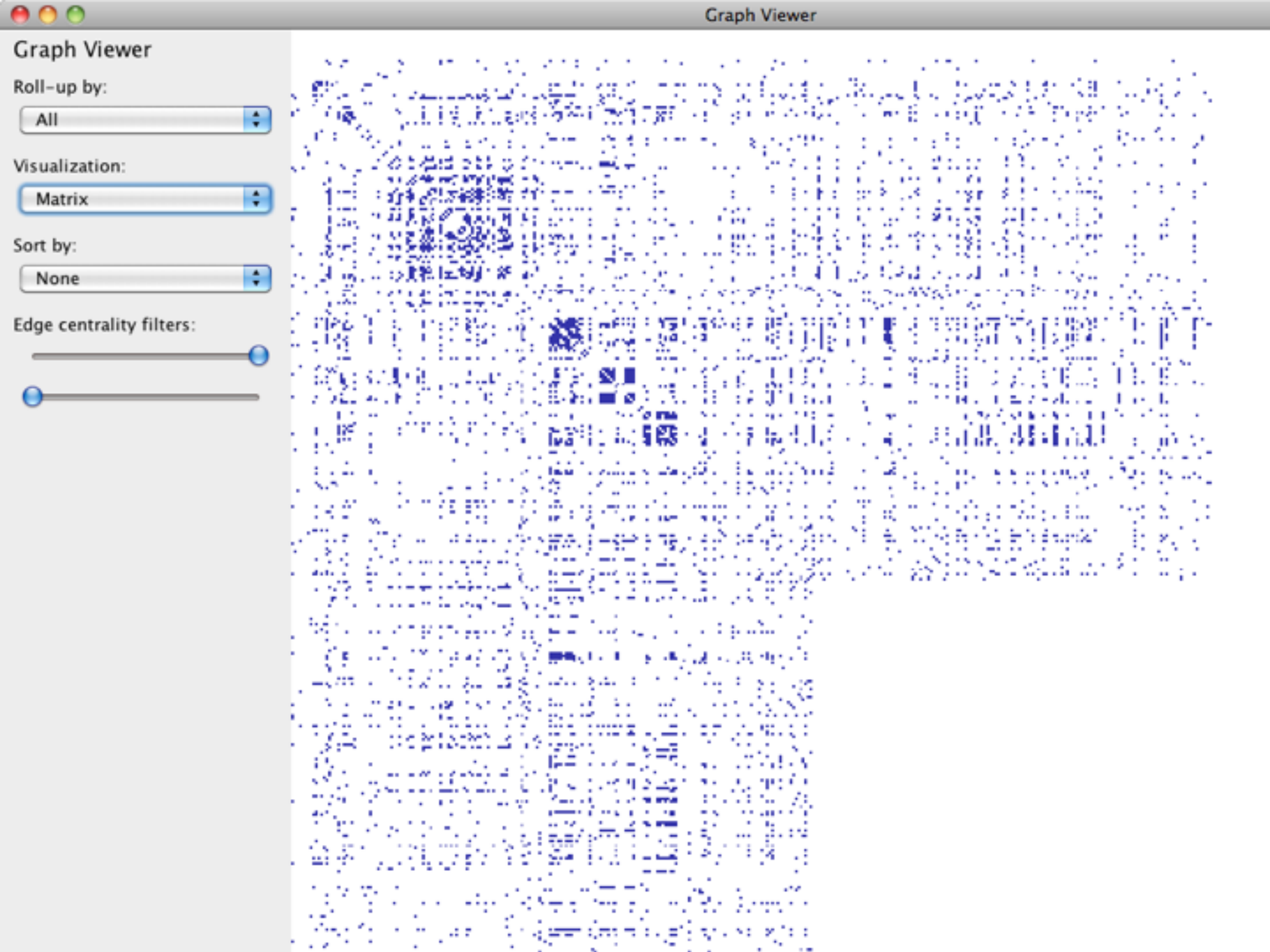
Graph Viewer

Roll-up by: All

Visualization: Matrix

Sort by: Linkage

Edge centrality filters:



Graph Viewer

Graph Viewer

Roll-up by: All

Visualization: Matrix

Sort by: None

Edge centrality filters:

What is Visualization?

“Transformation of the symbolic into the geometric”
[McCormick et al. 1987]

“… finding the artificial memory that best supports our natural means of perception.” [Bertin 1967]

“The use of computer-generated, interactive, visual representations of data to amplify cognition.”
[Card, Mackinlay, & Shneiderman 1999]

Why Create Visualizations?

Why Create Visualizations?

Answer questions (or discover them)

Make decisions

See data in context

Expand memory

Support graphical calculation

Find patterns

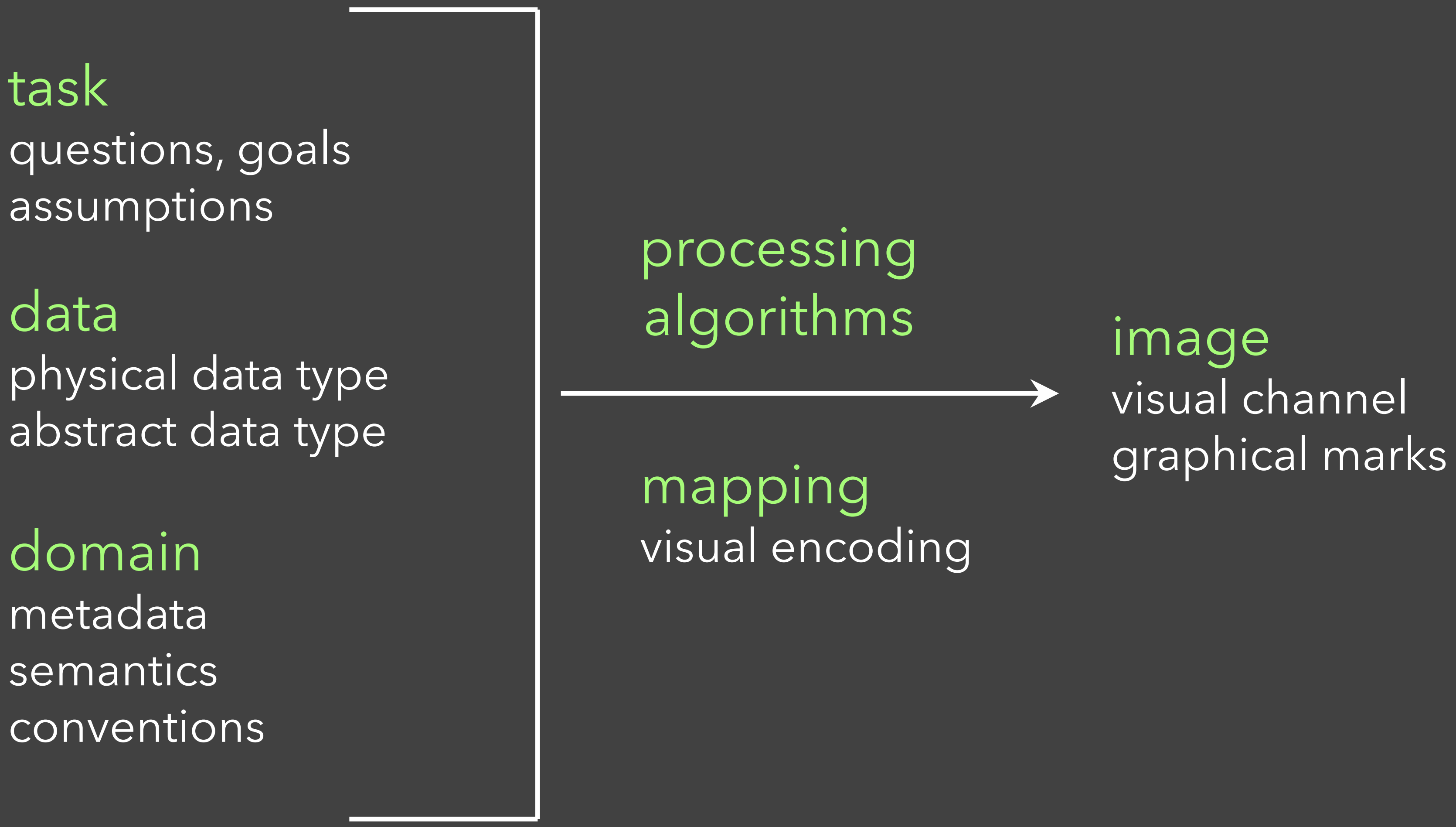
Present argument or tell a story

Inspire

The Value of Visualization

Data & Image Models

Visual Encoding



task

questions, goals
assumptions

data

physical data type
abstract data type

domain

metadata
semantics
conventions

processing
algorithms

mapping

visual encoding

image

visual channel
graphical marks

Nominal, Ordinal and Quantitative

Nominal, Ordinal and Quantitative

N - Nominal (labels or categories)

- Fruits: apples, oranges, …

Nominal, Ordinal and Quantitative

N - Nominal (labels or categories)

- Fruits: apples, oranges, …

O - Ordered

- Quality of meat: Grade A, AA, AAA

Nominal, Ordinal and Quantitative

N - Nominal (labels or categories)

- Fruits: apples, oranges, …

O - Ordered

- Quality of meat: Grade A, AA, AAA

Q - Interval (location of zero arbitrary)

- Dates: Jan, 19, 2006; Location: (LAT 33.98, LONG -118.45)
- Only differences (i.e. intervals) may be compared

Nominal, Ordinal and Quantitative

N - Nominal (labels or categories)

- Fruits: apples, oranges, …

O - Ordered

- Quality of meat: Grade A, AA, AAA

Q - Interval (location of zero arbitrary)

- Dates: Jan, 19, 2006; Location: (LAT 33.98, LONG -118.45)
- Only differences (i.e. intervals) may be compared

Q - Ratio (zero fixed)

- Physical measurement: Length, Mass, Temp, …
- Counts and amounts

Nominal, Ordinal and Quantitative

N - Nominal (labels or categories)

- Operations: =, ≠

O - Ordered

- Operations: =, ≠, <, >

Q - Interval (location of zero arbitrary)

- Operations: =, ≠, <, >, -
- Can measure distances or spans

Q - Ratio (zero fixed)

- Operations: =, ≠, <, >, -, %
- Can measure ratios or proportions

Visual Language is a Sign System

Jacques Bertin

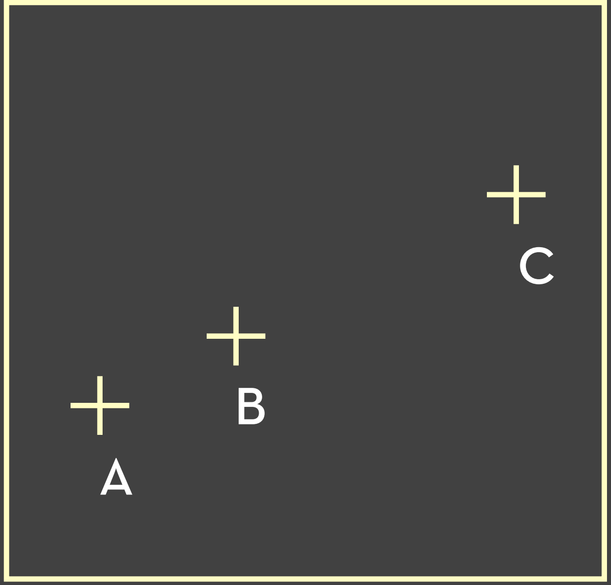
Images perceived as a set of signs

Sender encodes information in signs

Receiver decodes information from signs

Sémiologie Graphique, 1967

Bertin's Semiology of Graphics

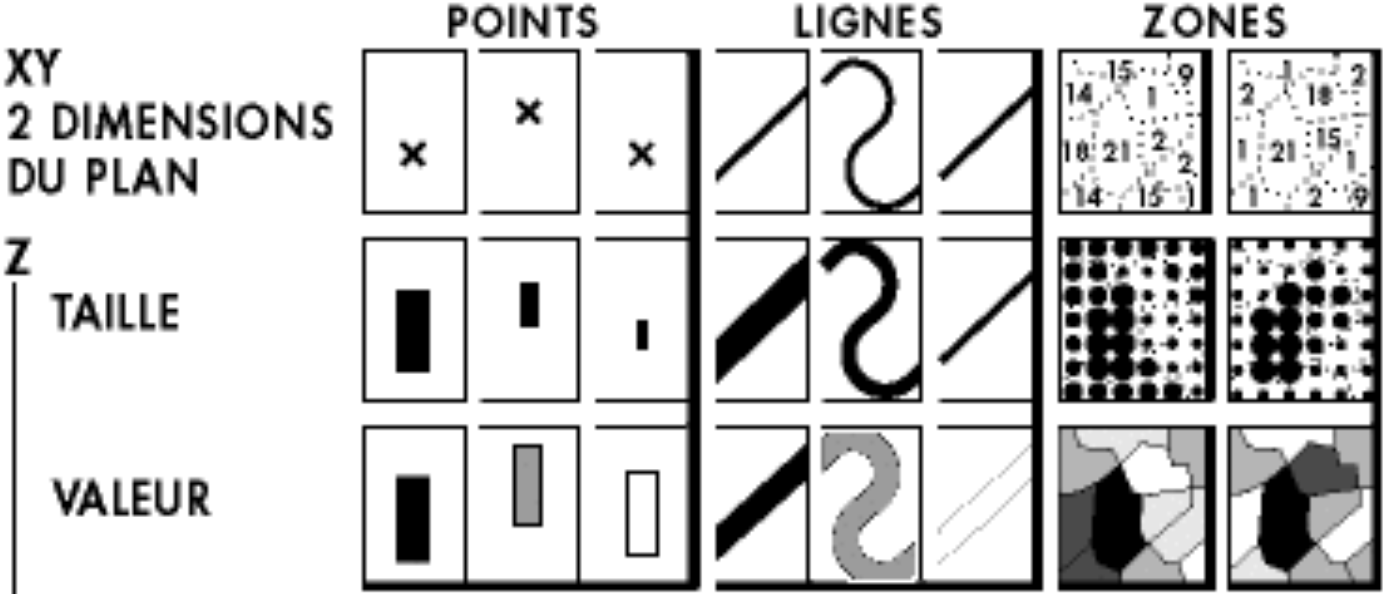


1. A, B, C are distinguishable
2. B is between A and C.
3. BC is twice as long as AB.

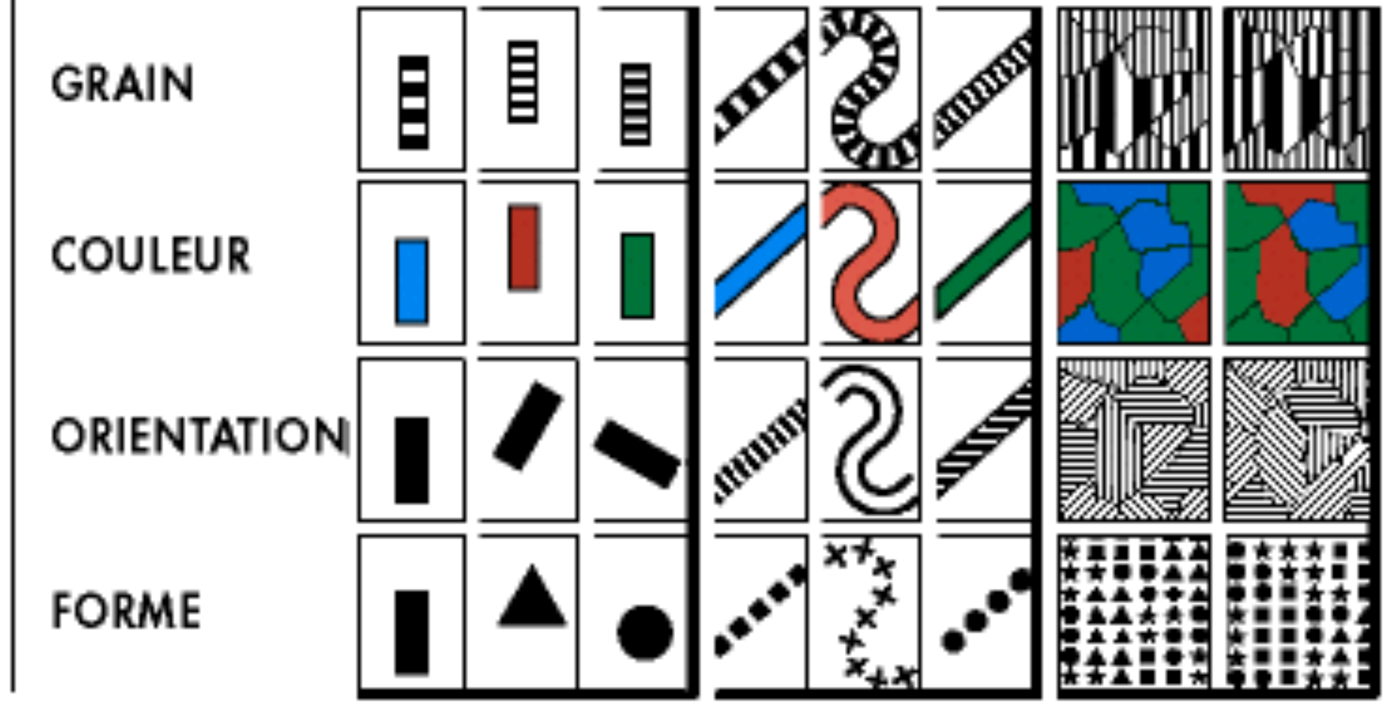
∴ Encode quantitative variables

"Resemblance, order and proportion are the three signfields in graphics." - Bertin

LES VARIABLES DE L'IMAGE

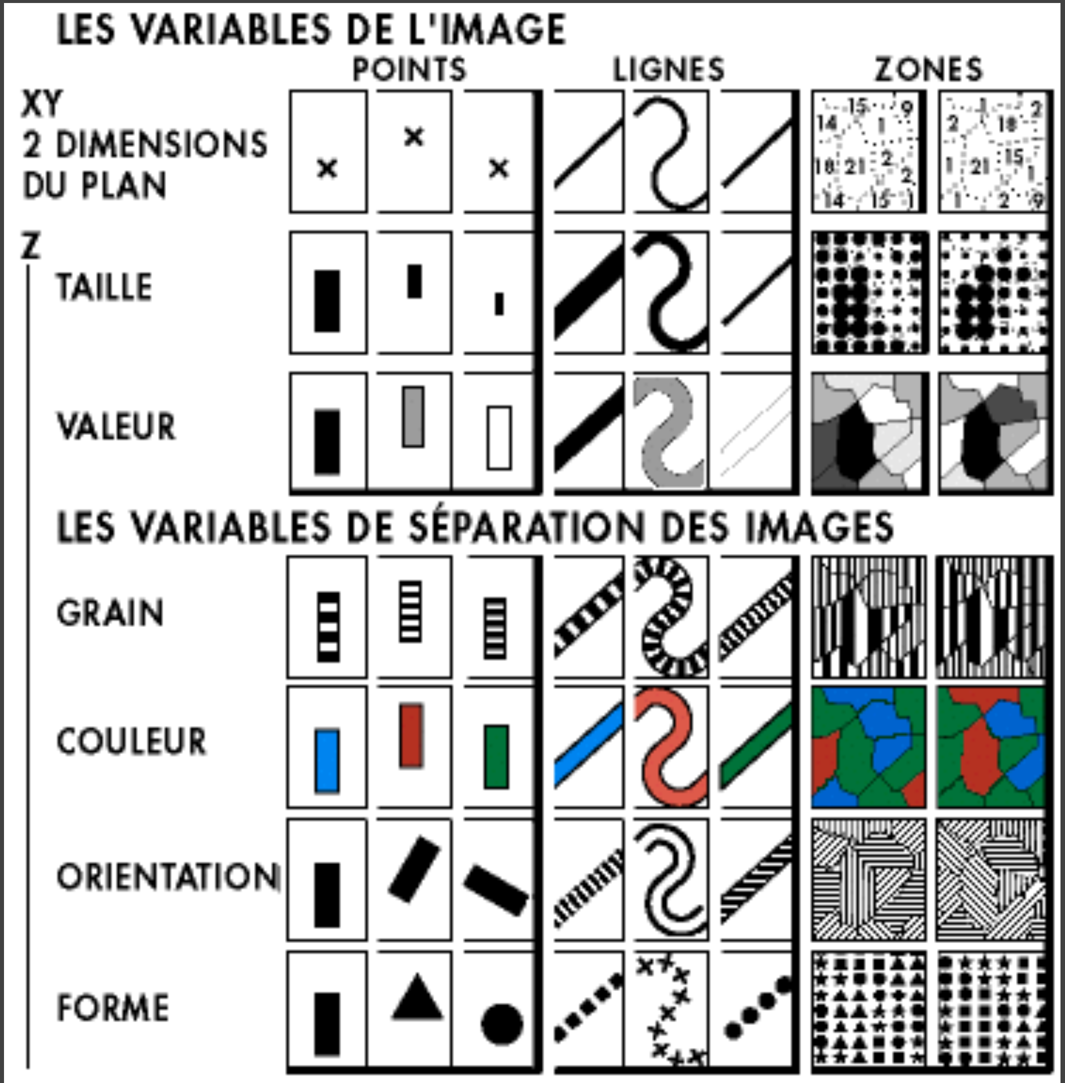


LES VARIABLES DE SÉPARATION DES IMAGES



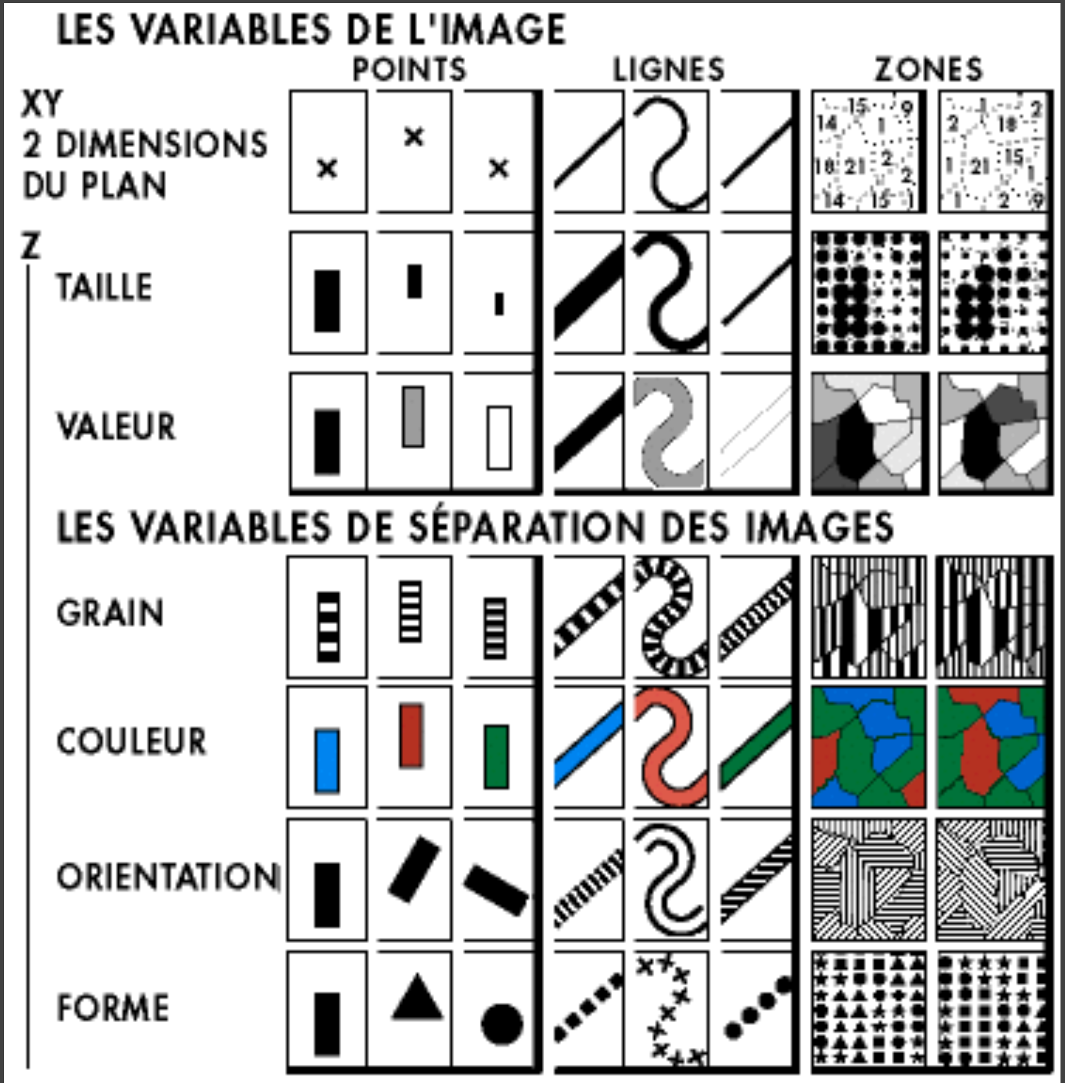
Visual Encoding Variables

Position (x 2)
Size
Value
Texture
Color
Orientation
Shape



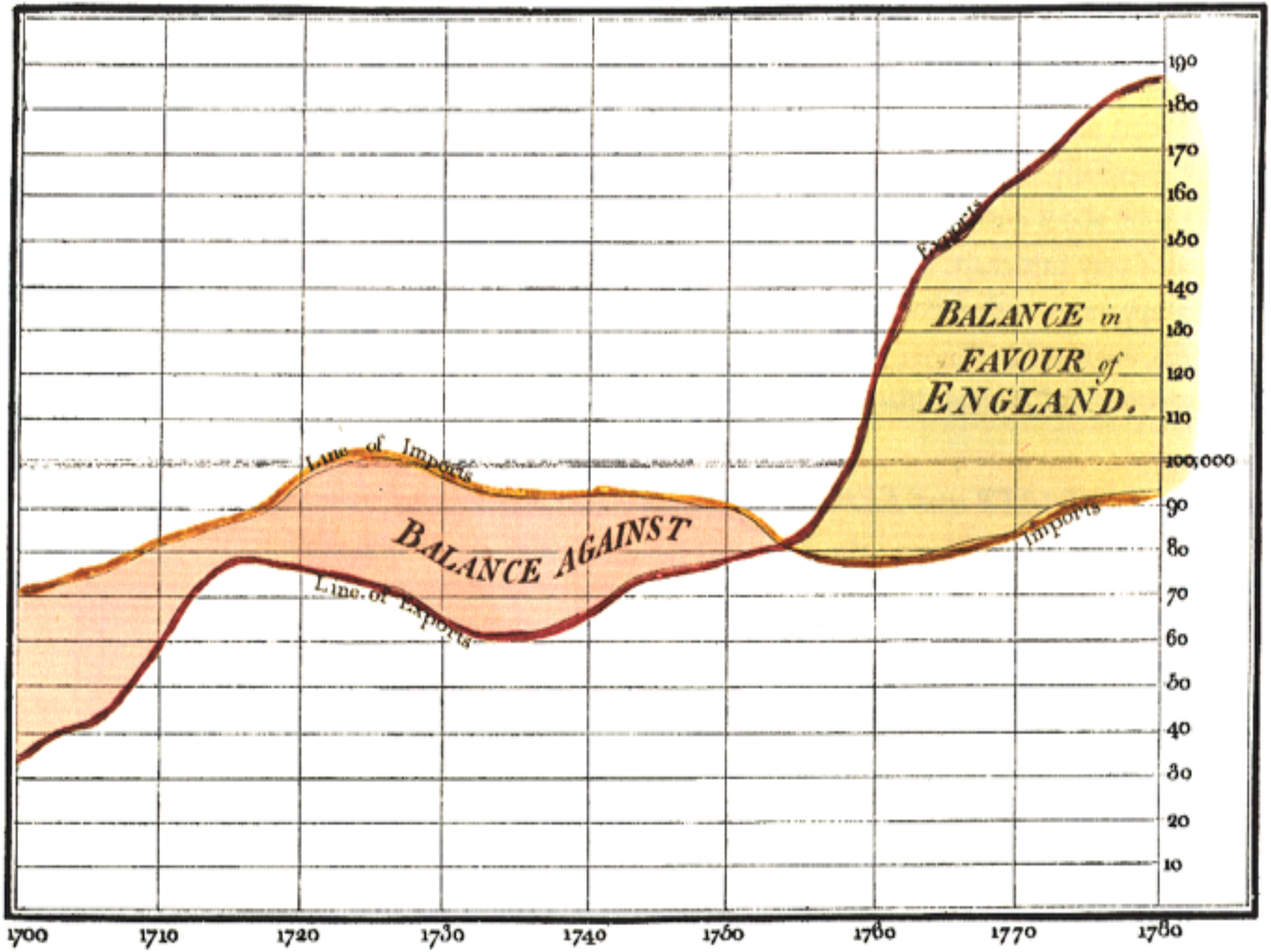
Visual Encoding Variables

Position
Length
Area
Volume
Value
Texture
Color
Orientation
Shape
Transparency
Blur / Focus …

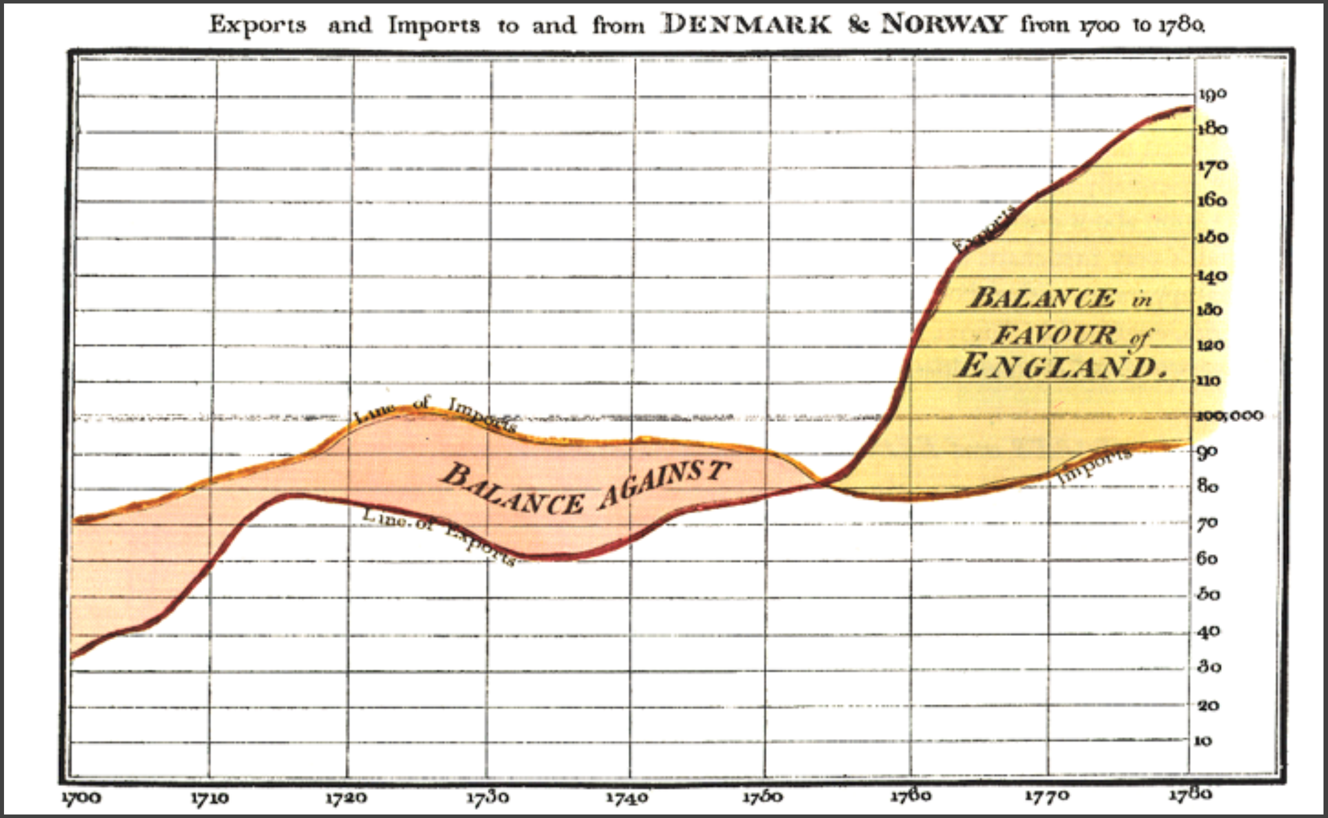


Deconstructions

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



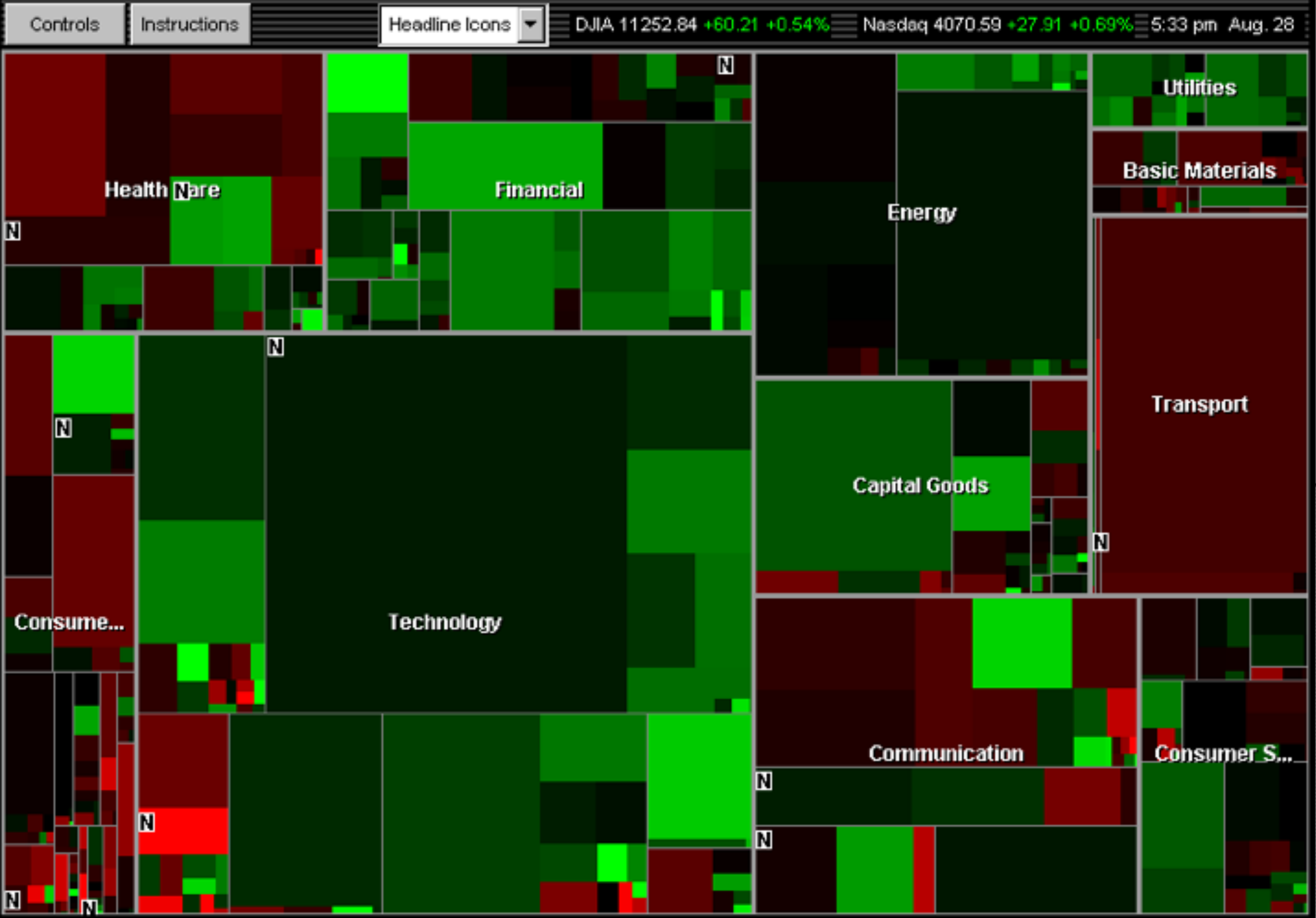
William Playfair, 1786



X-axis: year (Q)

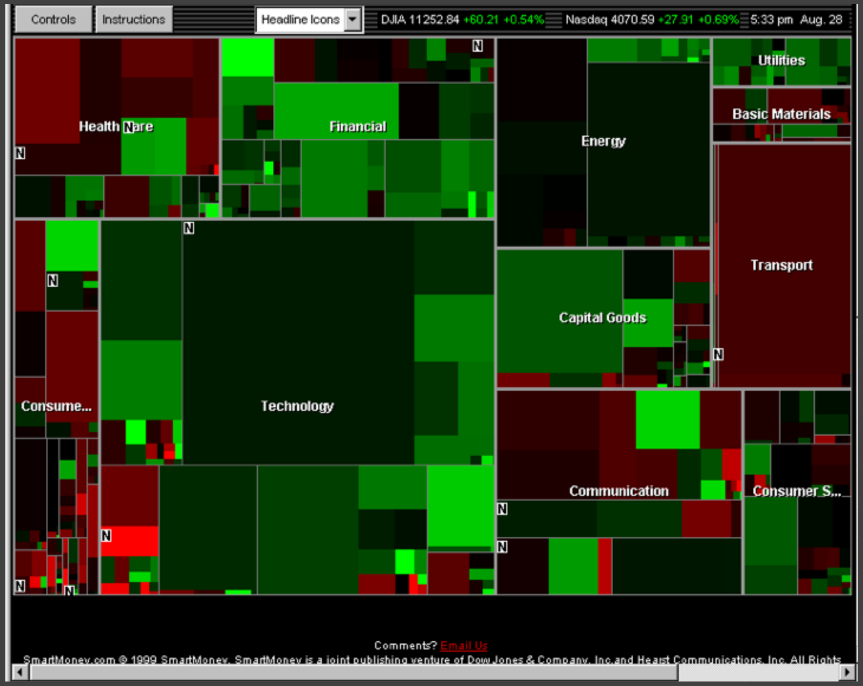
Y-axis: currency (Q)

Color: imports/exports (N, O)



http://www.smartmoney.com/marketmap/

Wattenberg’s Map of the Market



DJIA 11252.84 +60.21 +0.54% Nasdaq 4070.59 +27.91 +0.69% 5:33 pm Aug. 28

Comments? Email Us

SmartMoney.com © 1999 SmartMoney. SmartMoney is a joint publishing venture of Dow Jones & Company, Inc.and Hearst Communications, Inc. All Rights

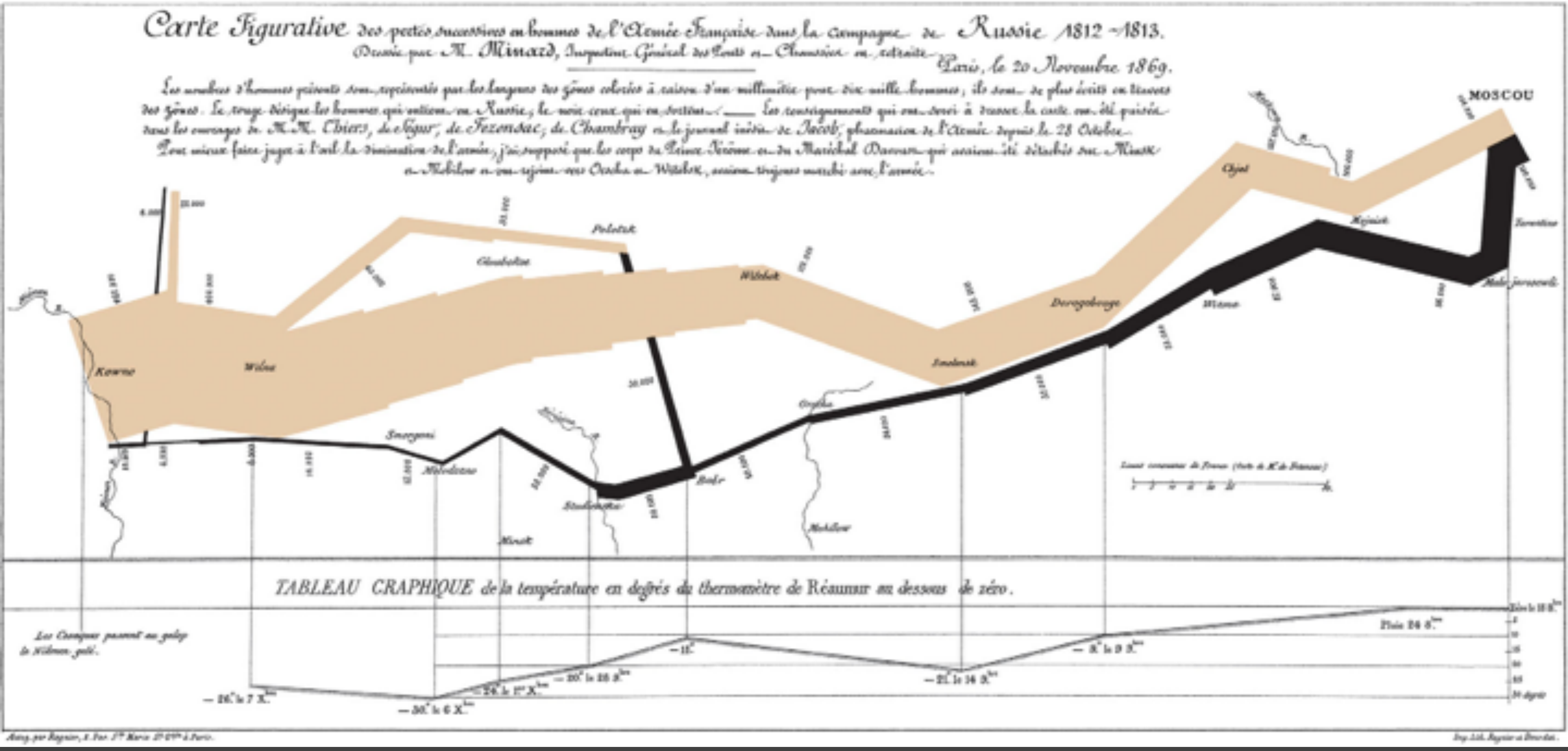
Rectangle Area: market cap (Q)

Rectangle Position: market sector (N), market cap (Q)

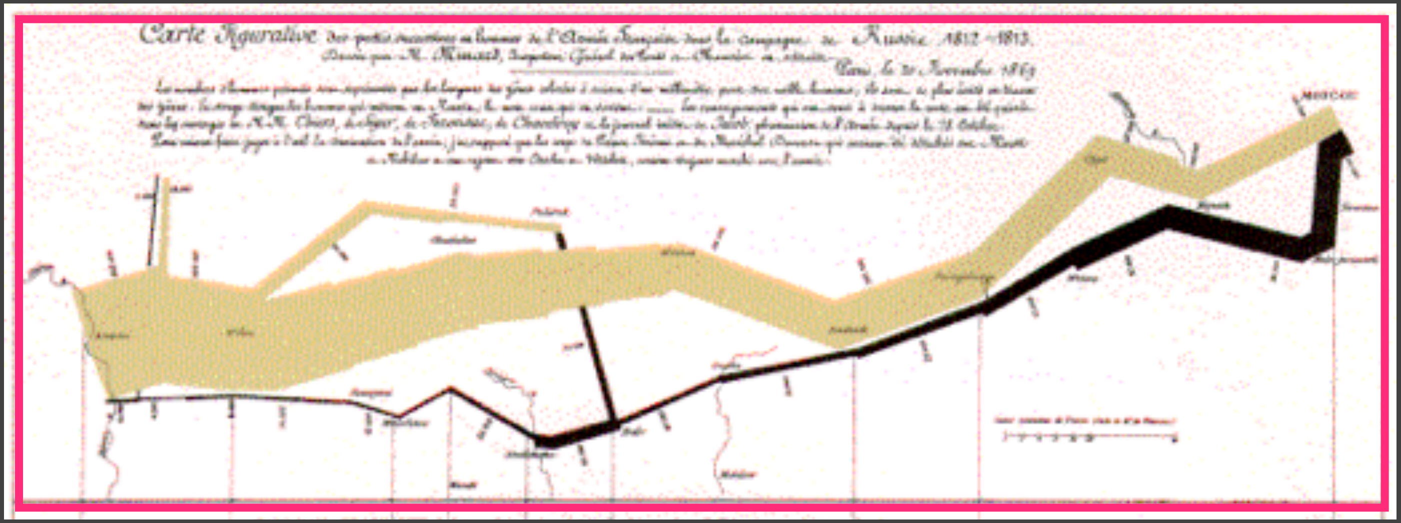
Color Hue: loss vs. gain (N, O)

Color Value: magnitude of loss or gain (Q)

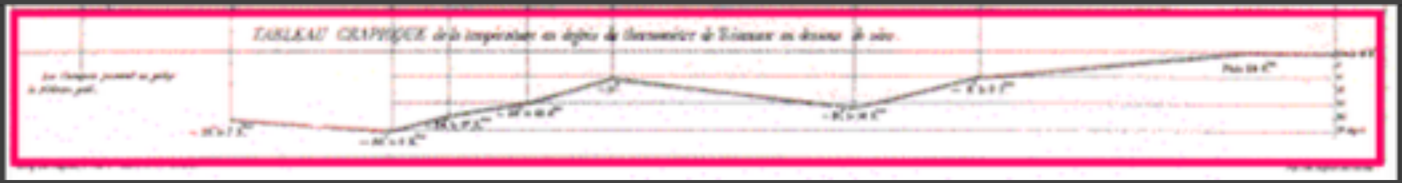
Minard 1869: Napoleon’s March


Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.
Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite.
Paris, le 20 Novembre 1869.
MOSCOU
Kowno
Wilna
Polotzk
Gloubokoe
Witebsk
Smolensk
Dorogobouge
Chjat
Mojaisk
Tarantino
Malo-jarosewli
Wiasma
Orscha
Mohilow
Smorgoni
Molodezno
Studienska
Minsk
Botr
TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.
Les Cosaques passent au galop le Niémen gelé.
Zéro le 18 8.bre
Pluie 24 8.bre
— 9° le 9 9.bre
— 21° le 14 9.bre
— 11°
— 20° le 28 9.bre
— 24° le 1er X.bre
— 30° le 6 X.bre
— 26° le 7 X.bre

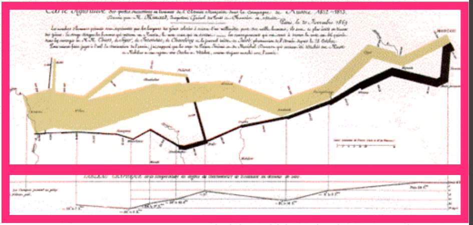
Single-Axis Composition


Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.
MOSCOU

+



=

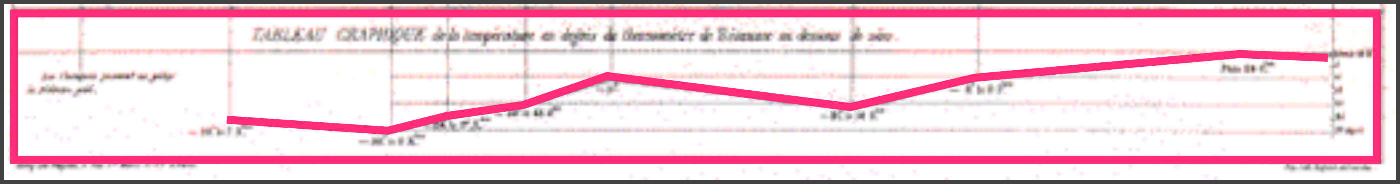


Mark Composition

Y-axis: temperature (Q)

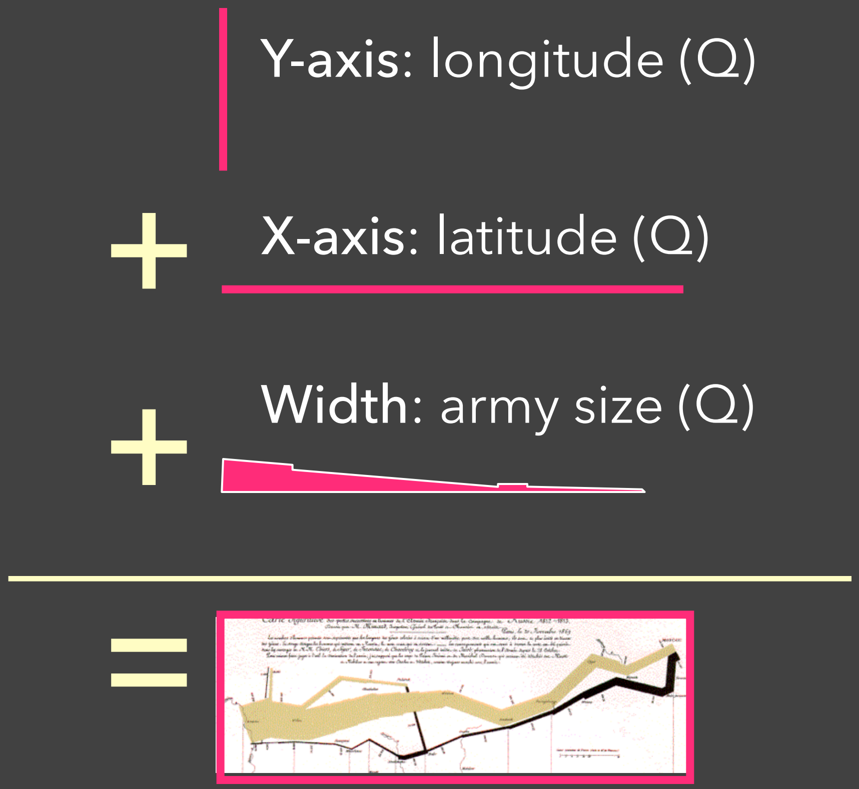
+ X-axis: longitude (Q) / time (O)

=



Temp over space/time (Q x Q)

Mark Composition



Y-axis: longitude (Q)

X-axis: latitude (Q)

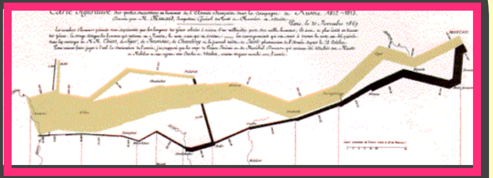
Width: army size (Q)

Army position (Q x Q) and army size (Q)

longitude (Q)

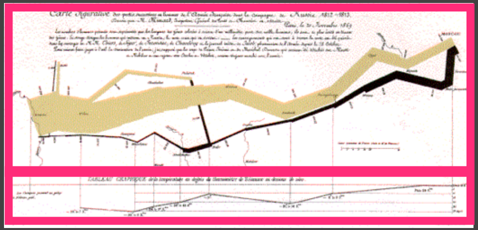
latitude (Q)

army size (Q)

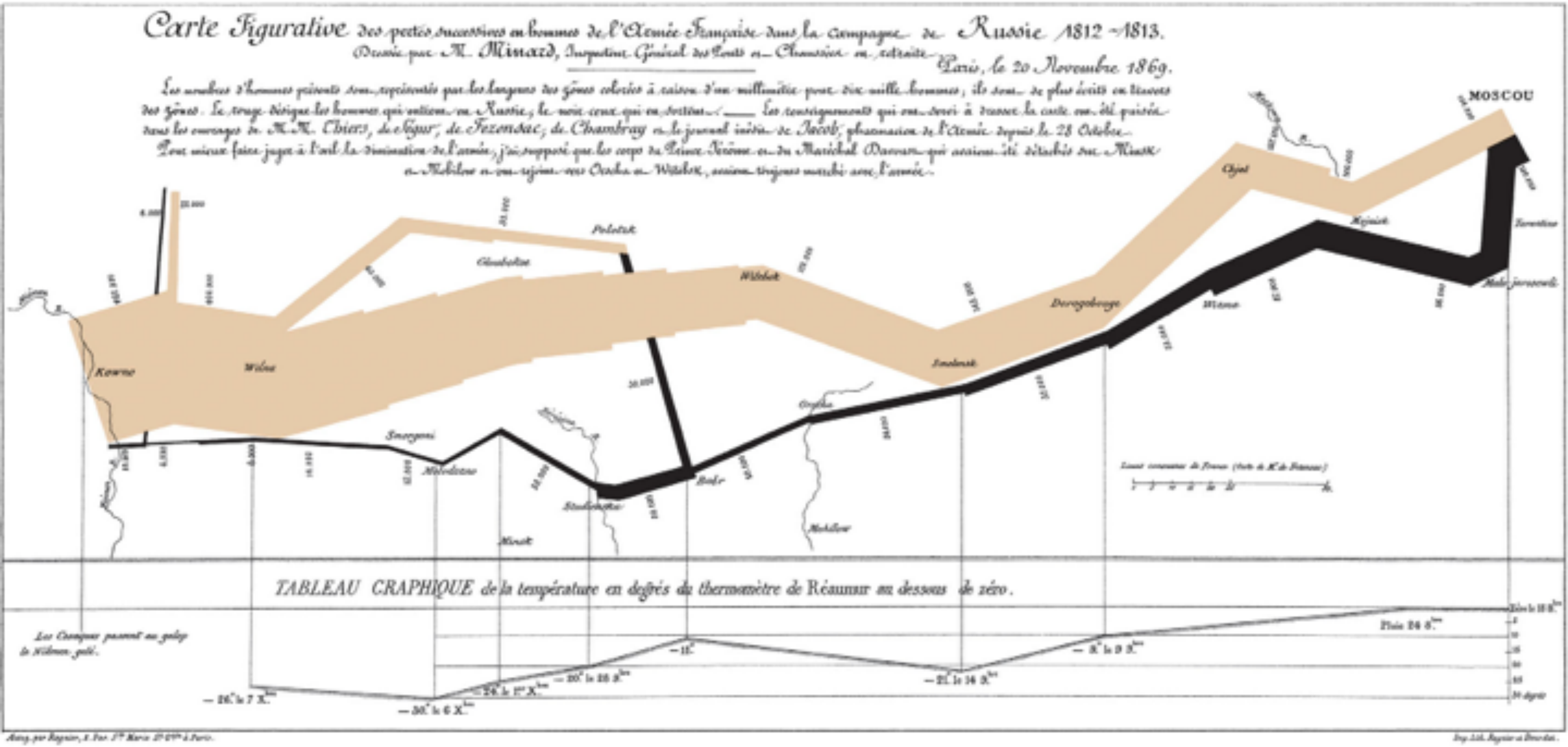


temperature (Q)

latitude (Q) / time (O)



Minard 1869: Napoleon’s March


Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812–1813.
Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite.
Paris, le 20 Novembre 1869.
Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M.M. Thiers, de Ségur, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.
Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davoust qui avaient été détachés sur Minsk et Mobilow et ont rejoint vers Orscha et Witebsk, avaient toujours marché avec l'armée.
MOSCOU
Chjat
Mojaisk
Tarantino
Malo-jarosewli
Wixma
Dorogobouge
Smolensk
Orscha
Mohilow
Witebsk
Polotzk
Gloubokoe
Kowno
Wilna
Smorgoni
Molodezno
Studienska
Bobr
Minsk
TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.
Les Cosaques passent au galop le Niémen gelé.
Zéro le 18 8.bre
Pluie 24 8.bre
— 9° le 9 9.bre
— 11°
— 21° le 14 9.bre
— 20° le 28 9.bre
— 24° le 1er X.bre
— 30° le 6 X.bre
— 26° le 7 X.bre

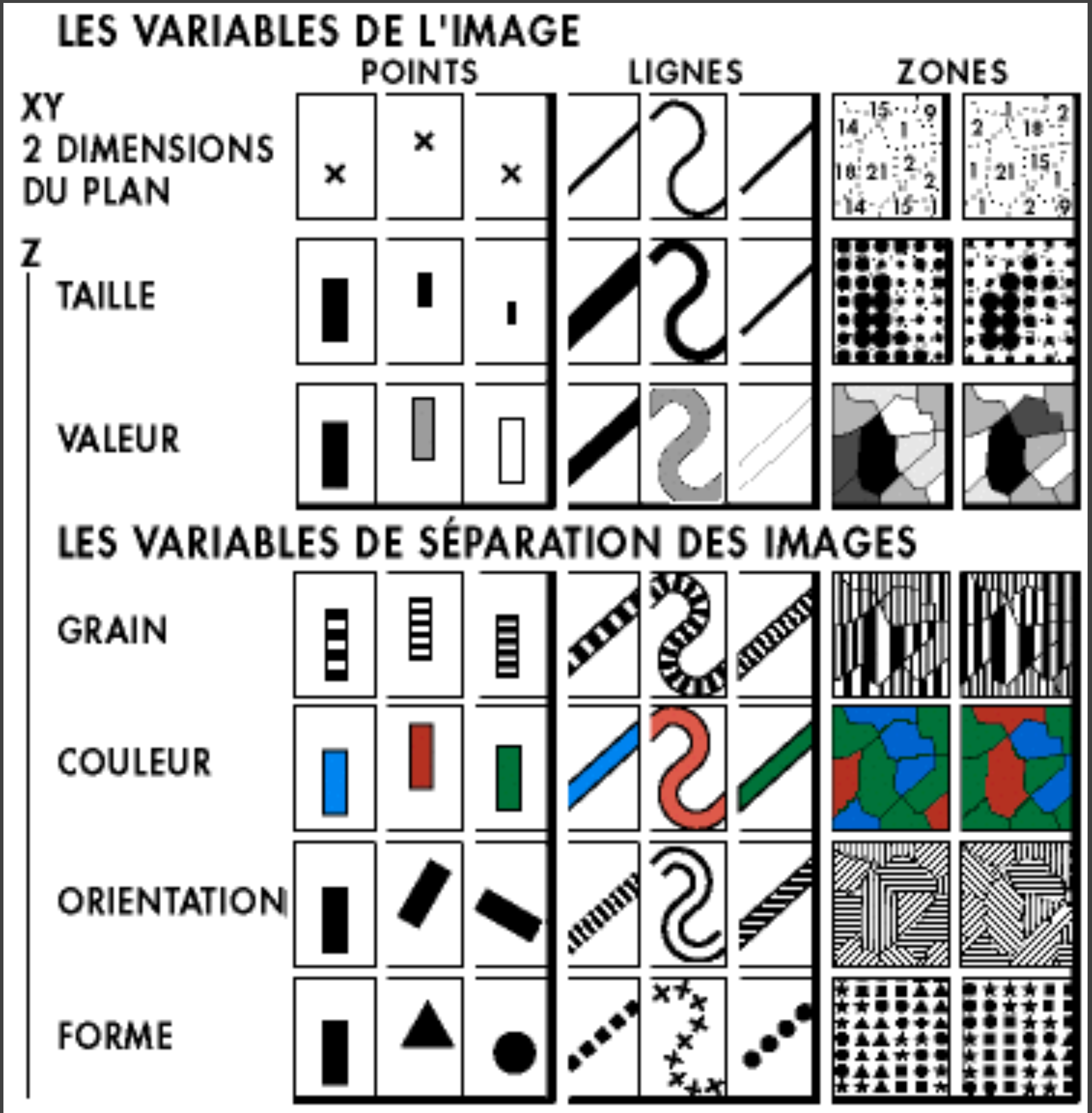
Depicts at least 5 quantitative variables. Any others?

Multidimensional Data

Visual Encoding Variables

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

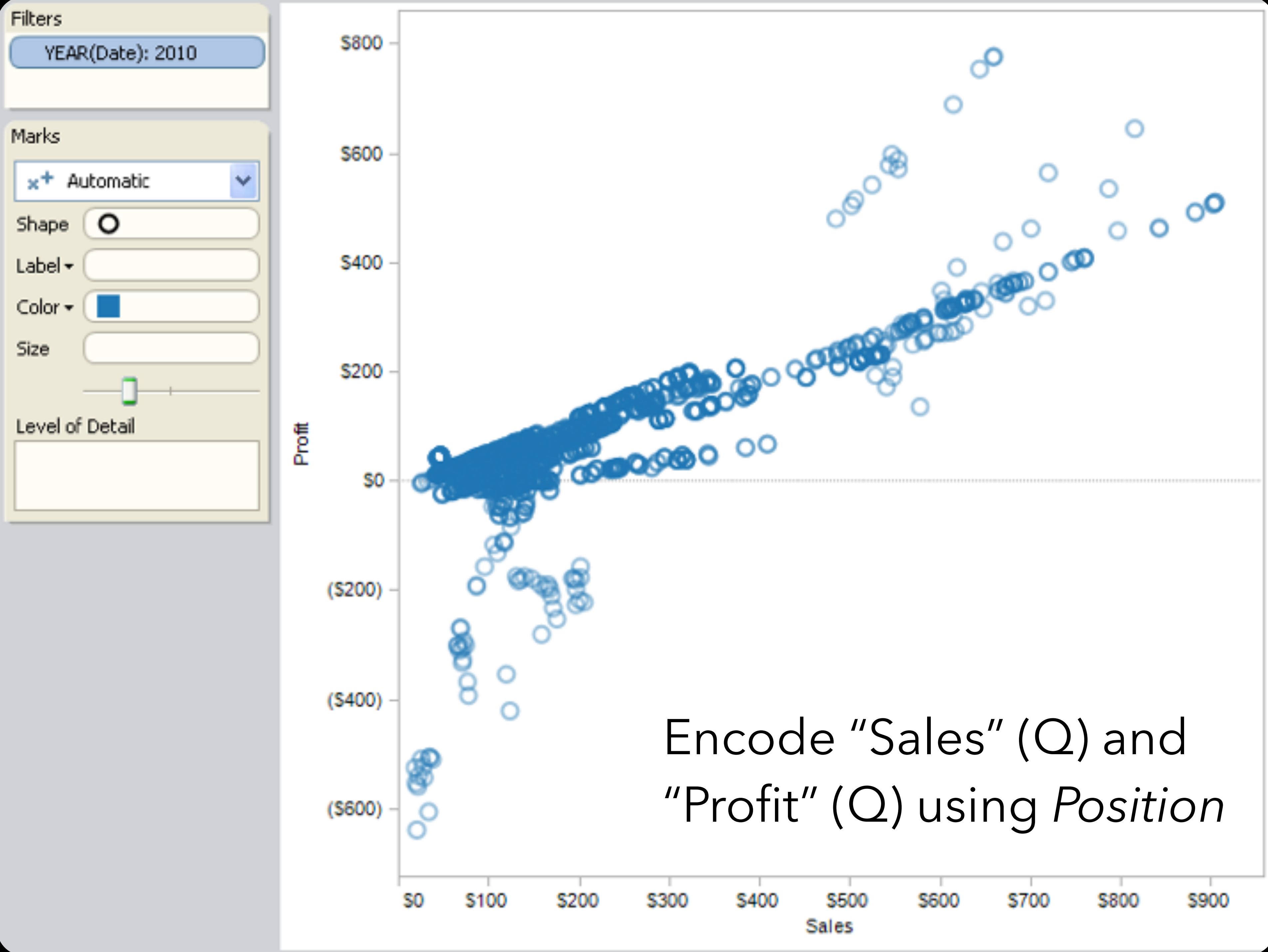
~8 dimensions?



Example: Coffee Sales

Sales figures for a fictional coffee chain:

Sales	Q-Ratio
Profit	Q-Ratio
Marketing	Q-Ratio
Product Type	N {Coffee, Espresso, Herbal Tea, Tea}
Market	N {Central, East, South, West}



Filters

YEAR(Date): 2010

Marks

Automatic

Shape ○

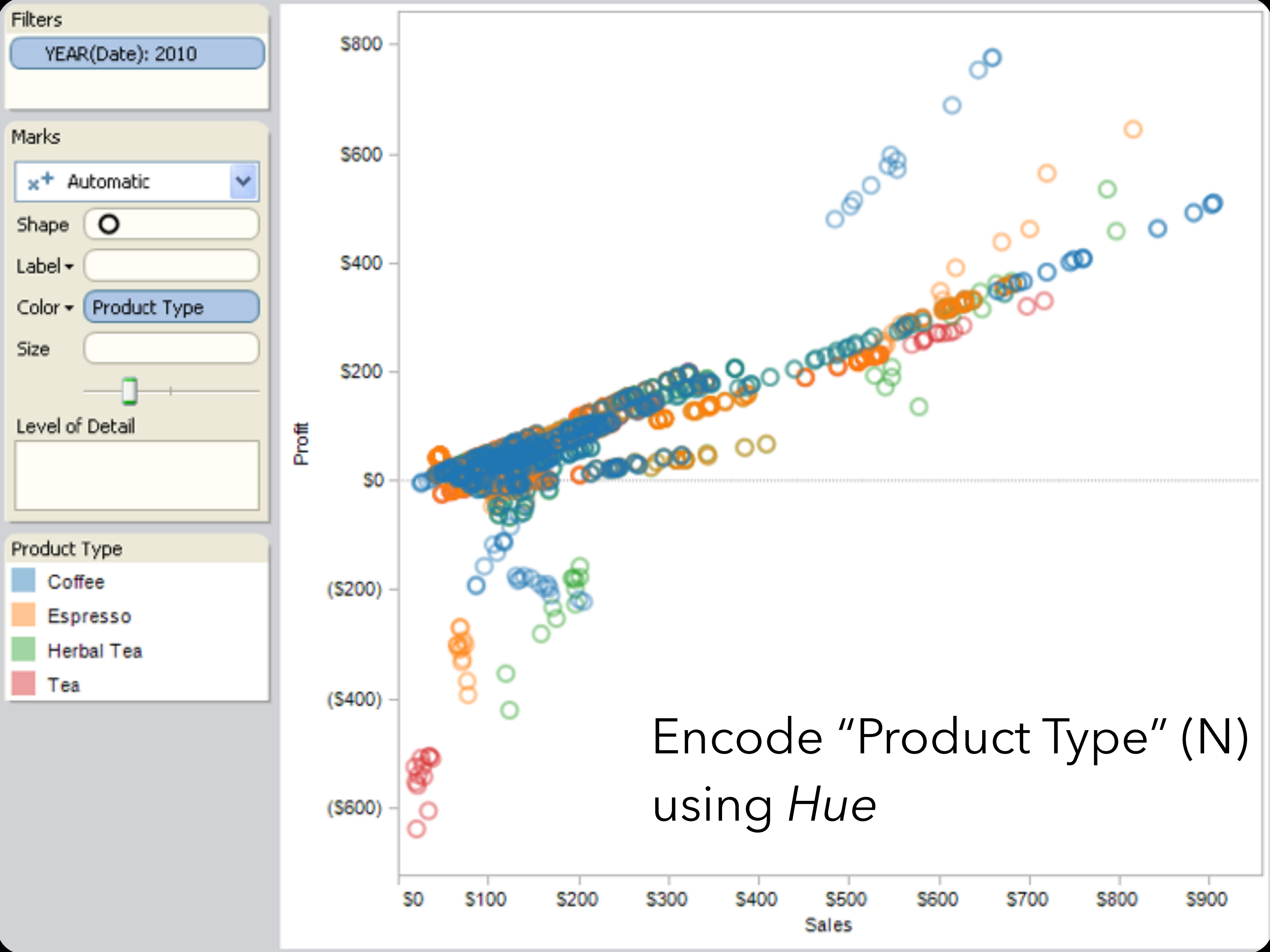
Label

Color 🟦

Size

Level of Detail

Encode “Sales” (Q) and “Profit” (Q) using *Position*



Filters

YEAR(Date): 2010

Marks

Automatic

Shape: ○

Label

Color: Product Type

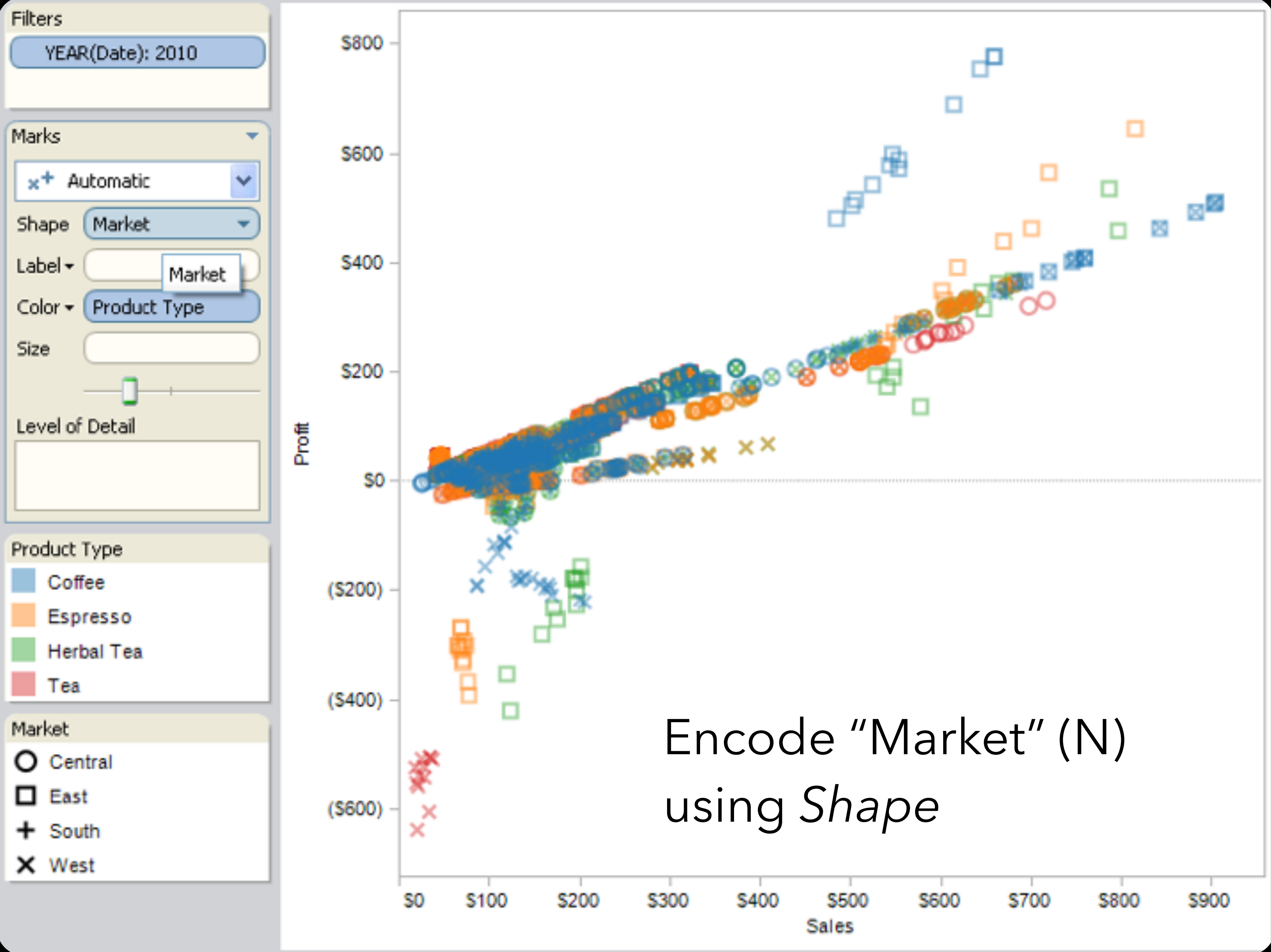
Size

Level of Detail

Product Type

- 🟦 Coffee
- 🟧 Espresso
- 🟩 Herbal Tea
- 🟥 Tea

Encode "Product Type" (N) using *Hue*



Filters

YEAR(Date): 2010

Marks

Automatic

Shape: Market

Label

Market

Color: Product Type

Size

Level of Detail

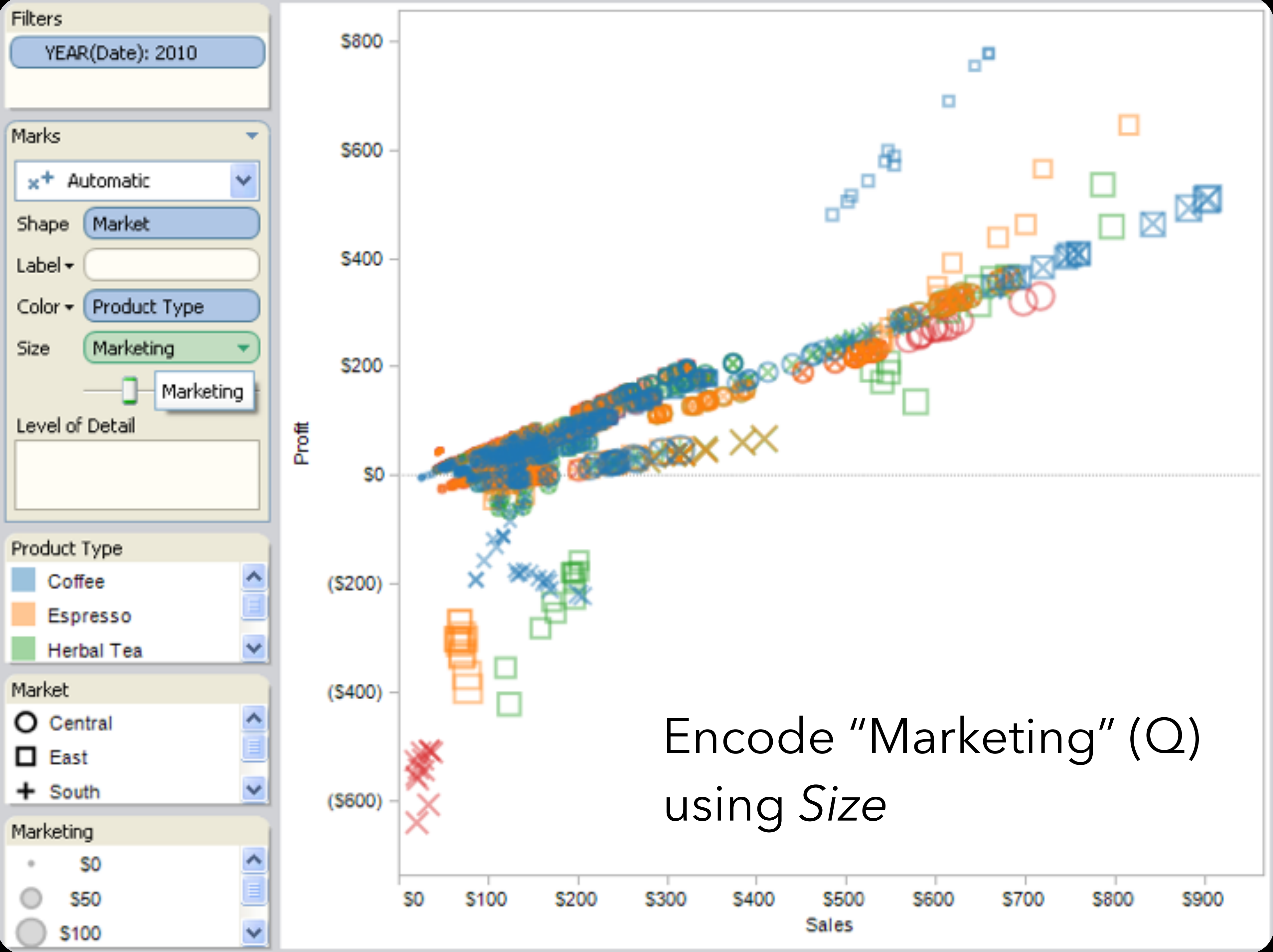
Product Type

🟦	Coffee
🟧	Espresso
🟩	Herbal Tea
🟥	Tea

Market

○	Central
□	East
+	South
×	West

Encode "Market" (N) using *Shape*



Filters

YEAR(Date): 2010

Marks

Automatic

Shape: Market

Label:

Color: Product Type

Size: Marketing

Level of Detail

Product Type

- 🟦 Coffee
- 🟧 Espresso
- 🟩 Herbal Tea

Market

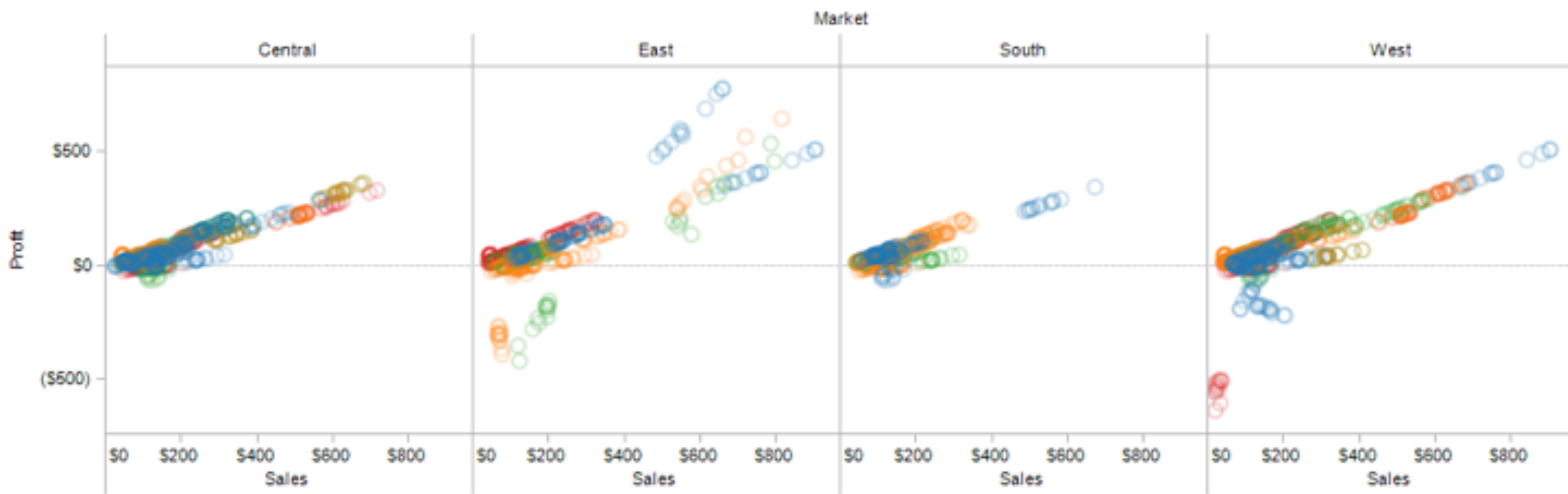
- ○ Central
- □ East
- + South

Marketing

- ● $0
- ● $50
- ● $100

Encode “Marketing” (Q) using *Size*

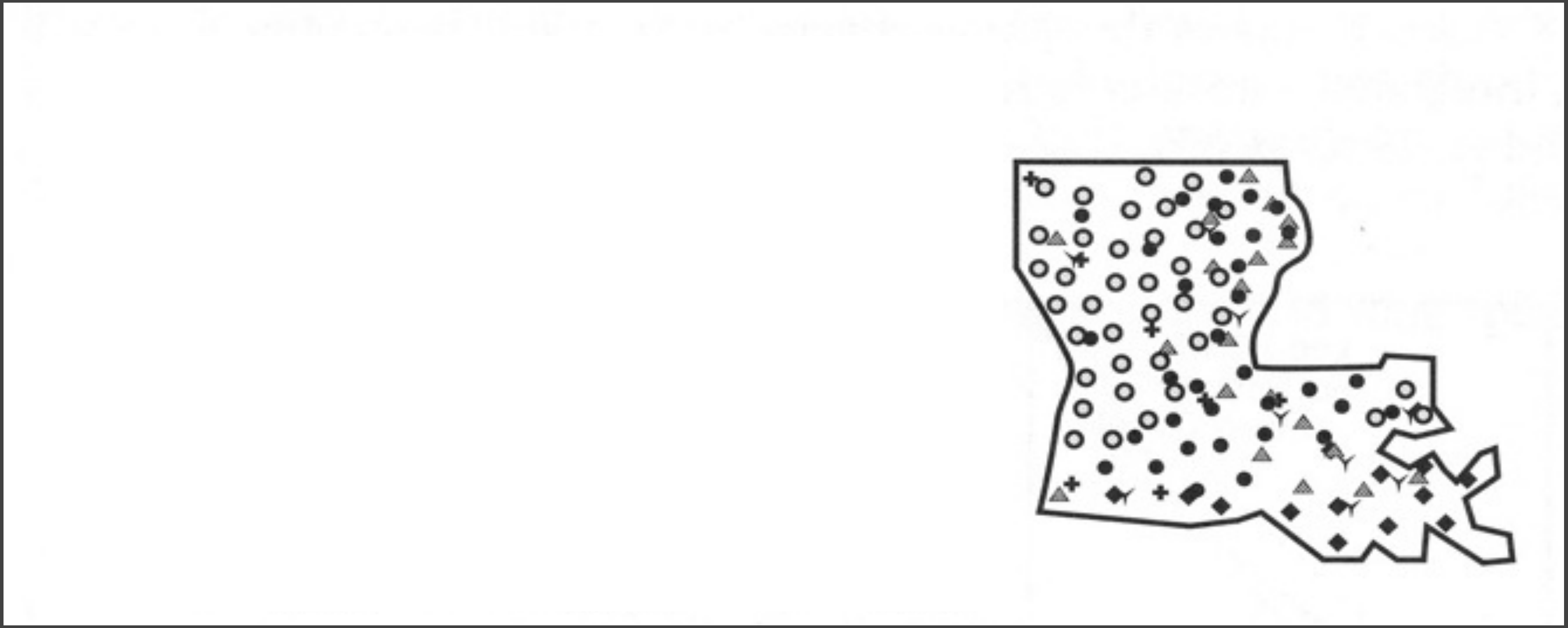
Trellis Plots



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

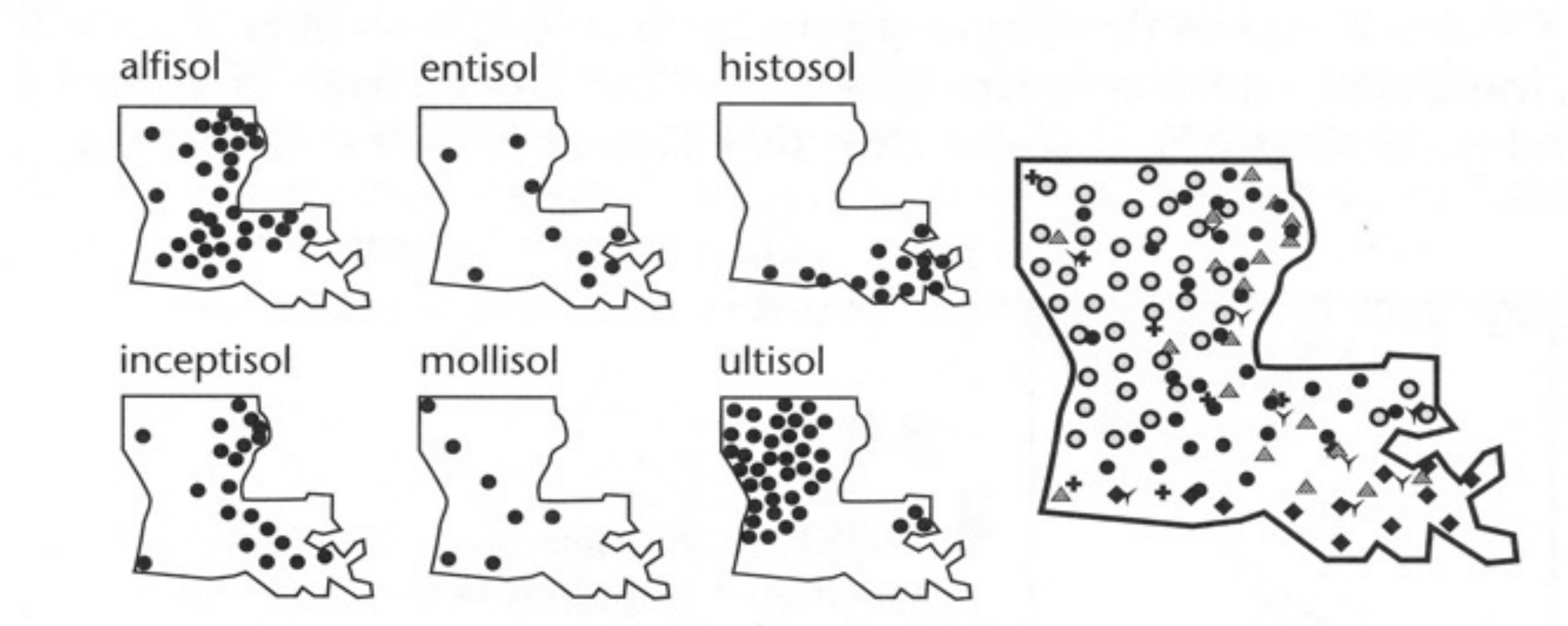
Typically nominal or ordinal variables are used as dimensions for subdivision.

Small Multiples



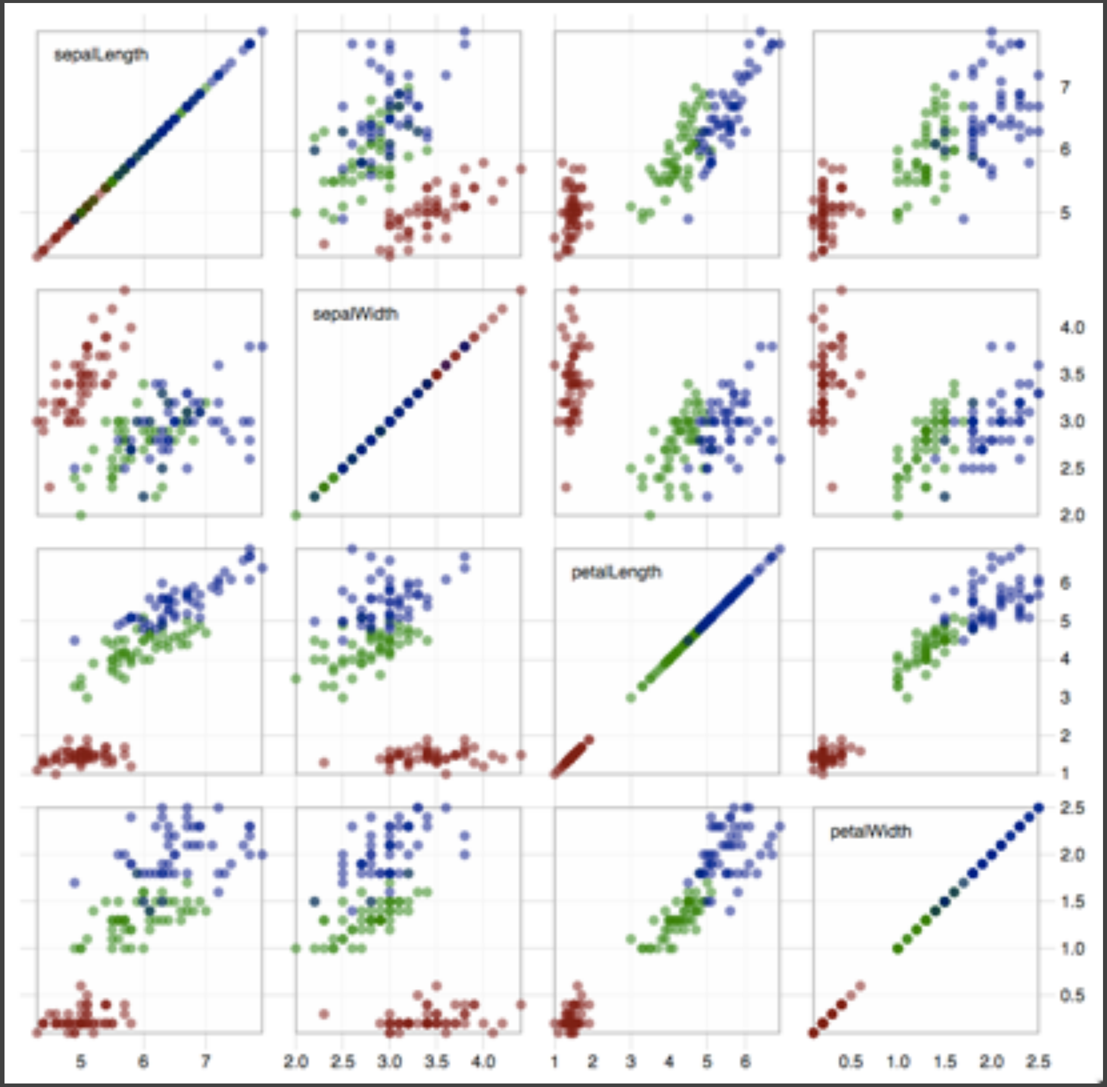
[MacEachren 95, Figure 2.11, p. 38]

Small Multiples

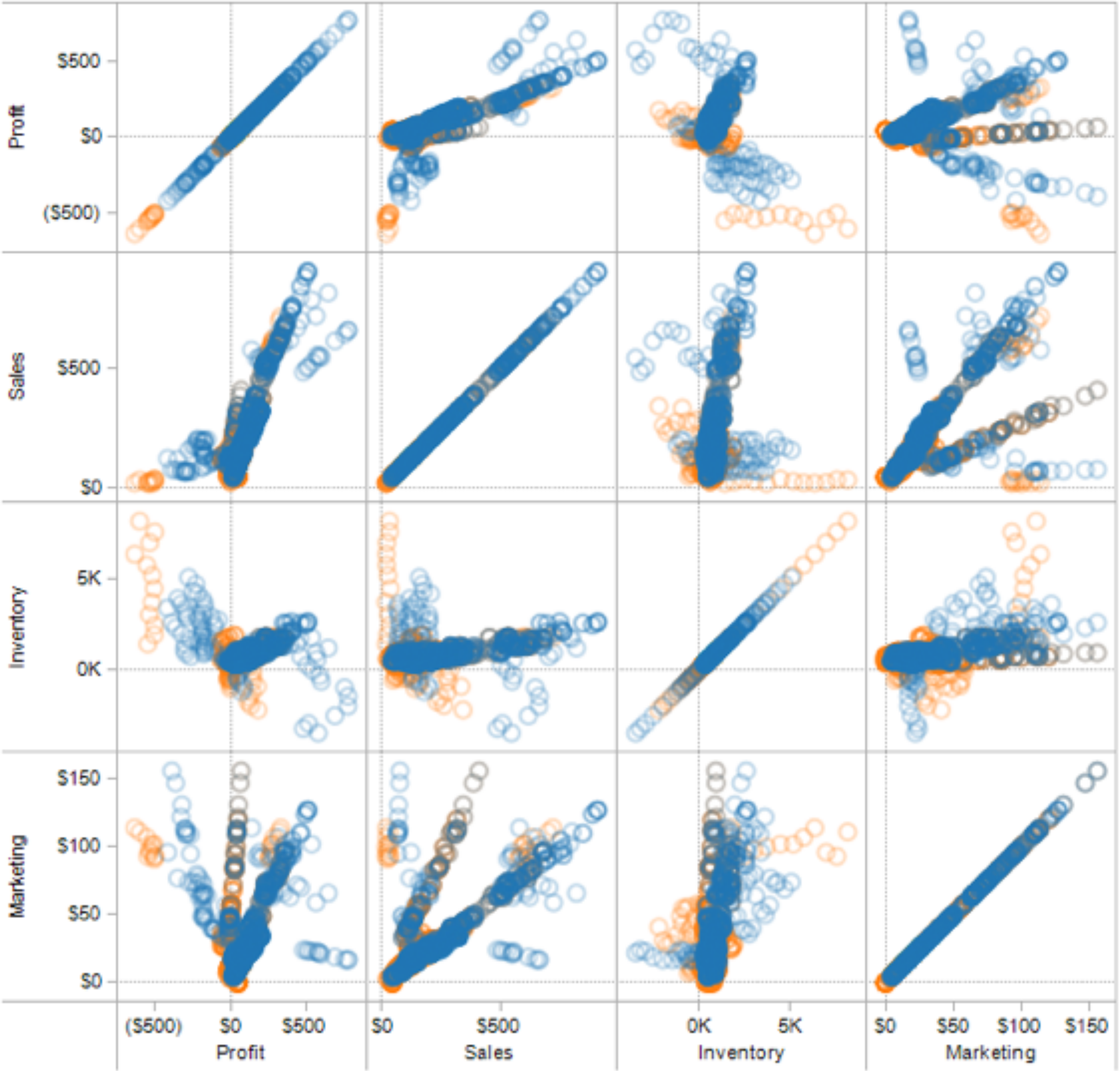


[MacEachren 95, Figure 2.11, p. 38]

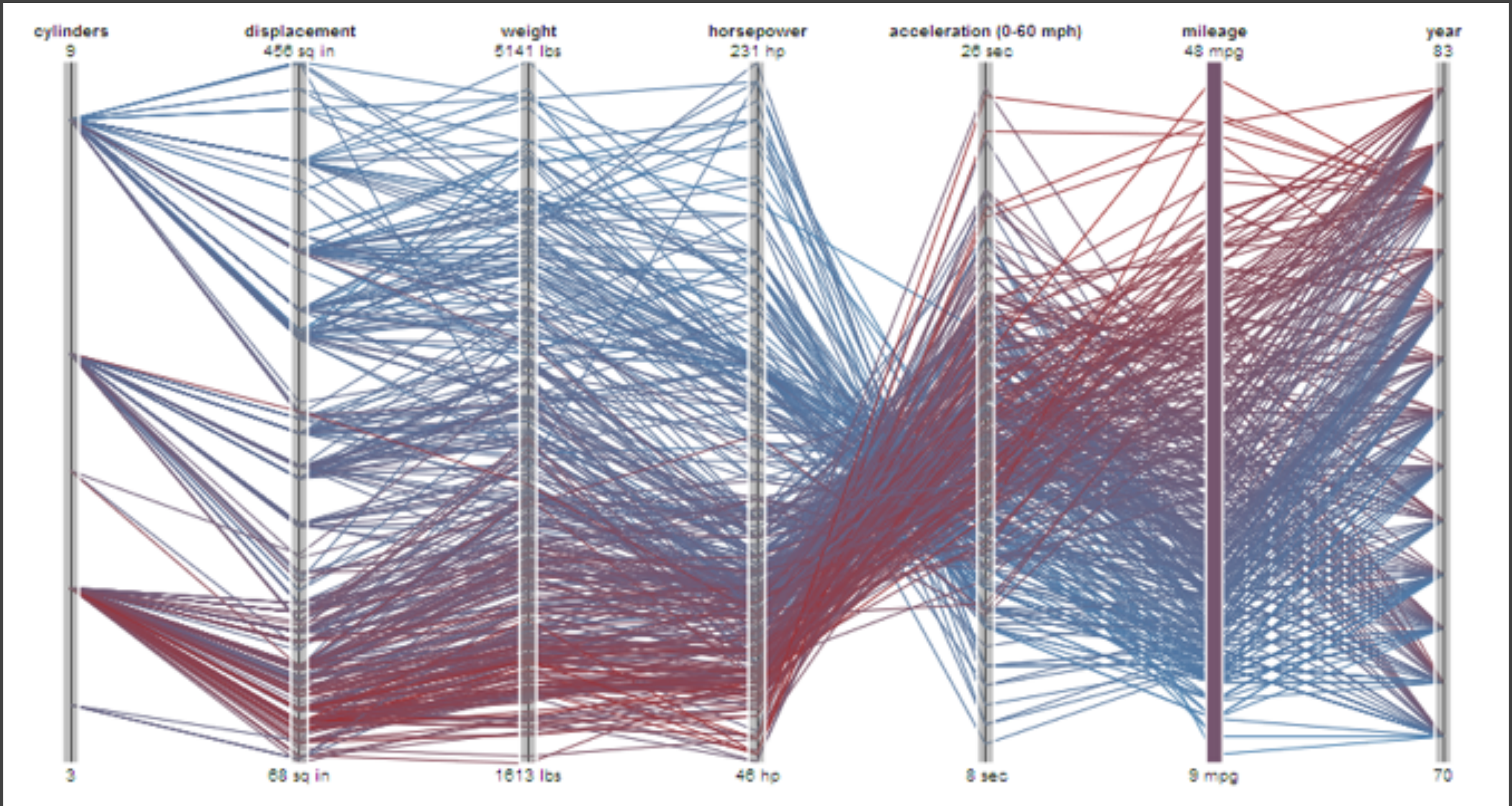
Scatterplot Matrix (SPLOM)



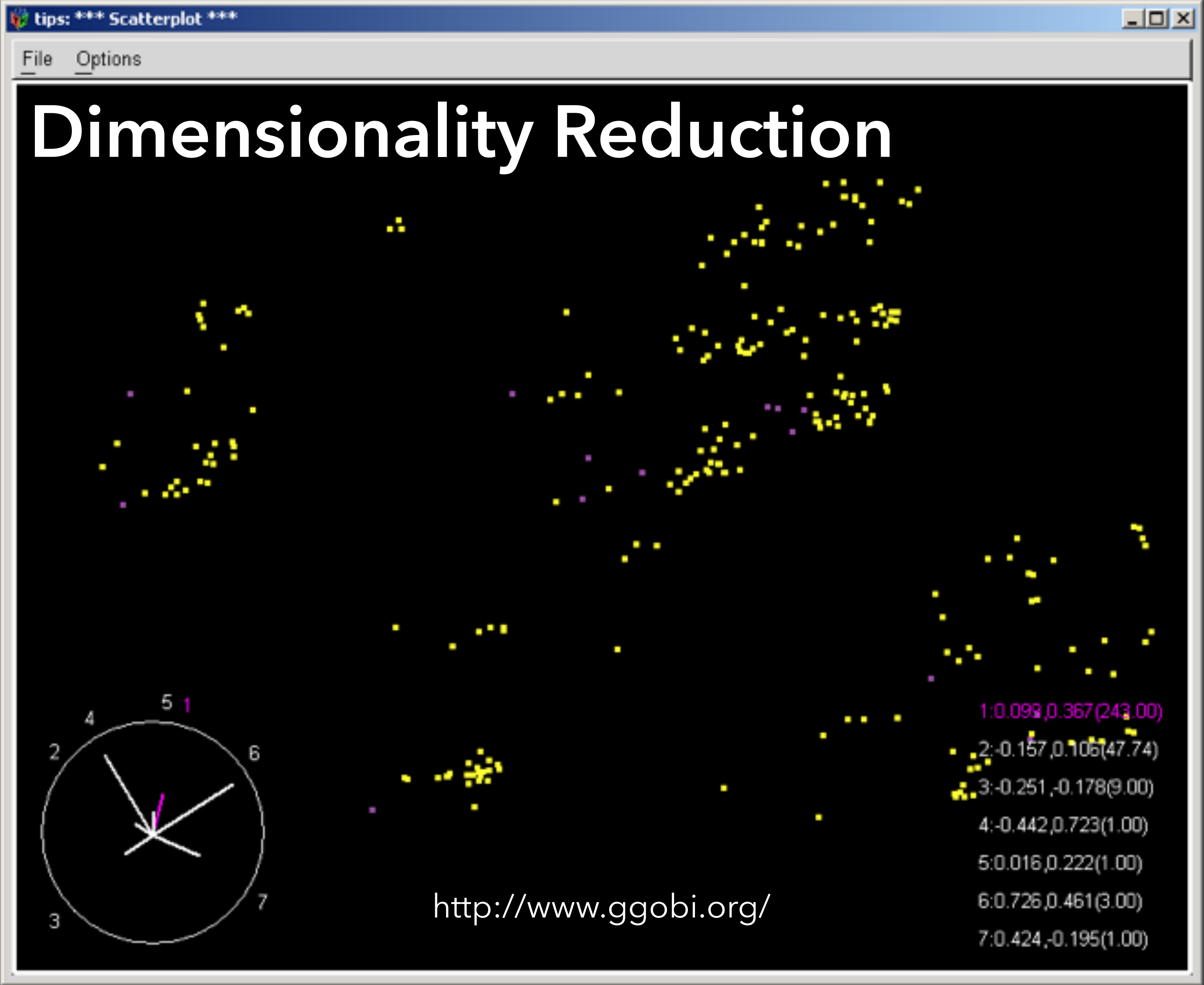
Scatter plots for pairwise comparison of each data dimension.


Profit
$500
$0
($500)
Sales
$500
$0
Inventory
5K
0K
Marketing
$150
$100
$50
$0
($500)
$0
$500
Profit
$0
$500
Sales
0K
5K
Inventory
$0
$50
$100
$150
Marketing

Parallel Coordinates [Inselberg]



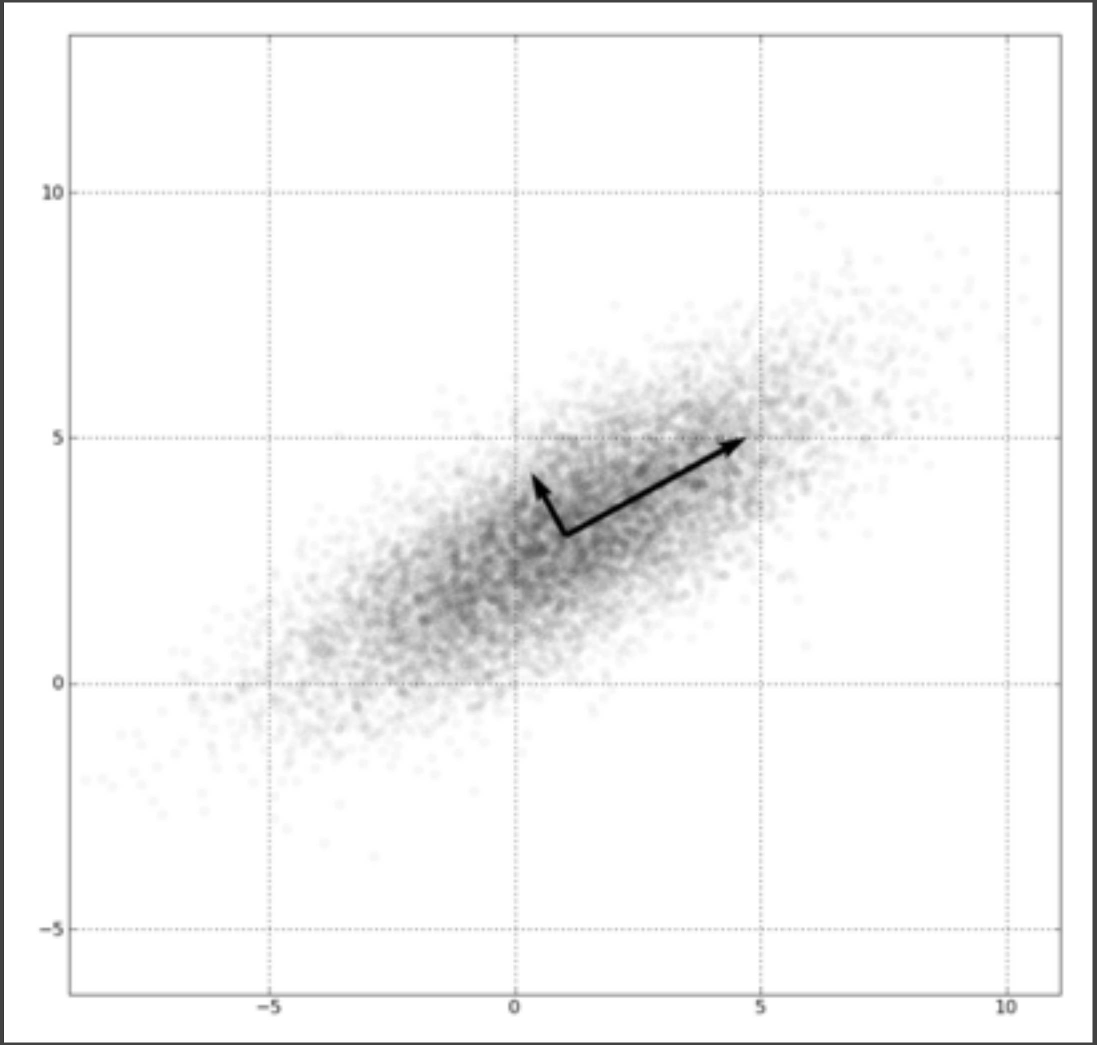
Dimensionality Reduction



```
1:0.099,0.367(243.00)
2:-0.157,0.106(47.74)
3:-0.251,-0.178(9.00)
4:-0.442,0.723(1.00)
5:0.016,0.222(1.00)
6:0.726,0.461(3.00)
7:0.424,-0.195(1.00)
```

http://www.ggobi.org/

Principal Components Analysis

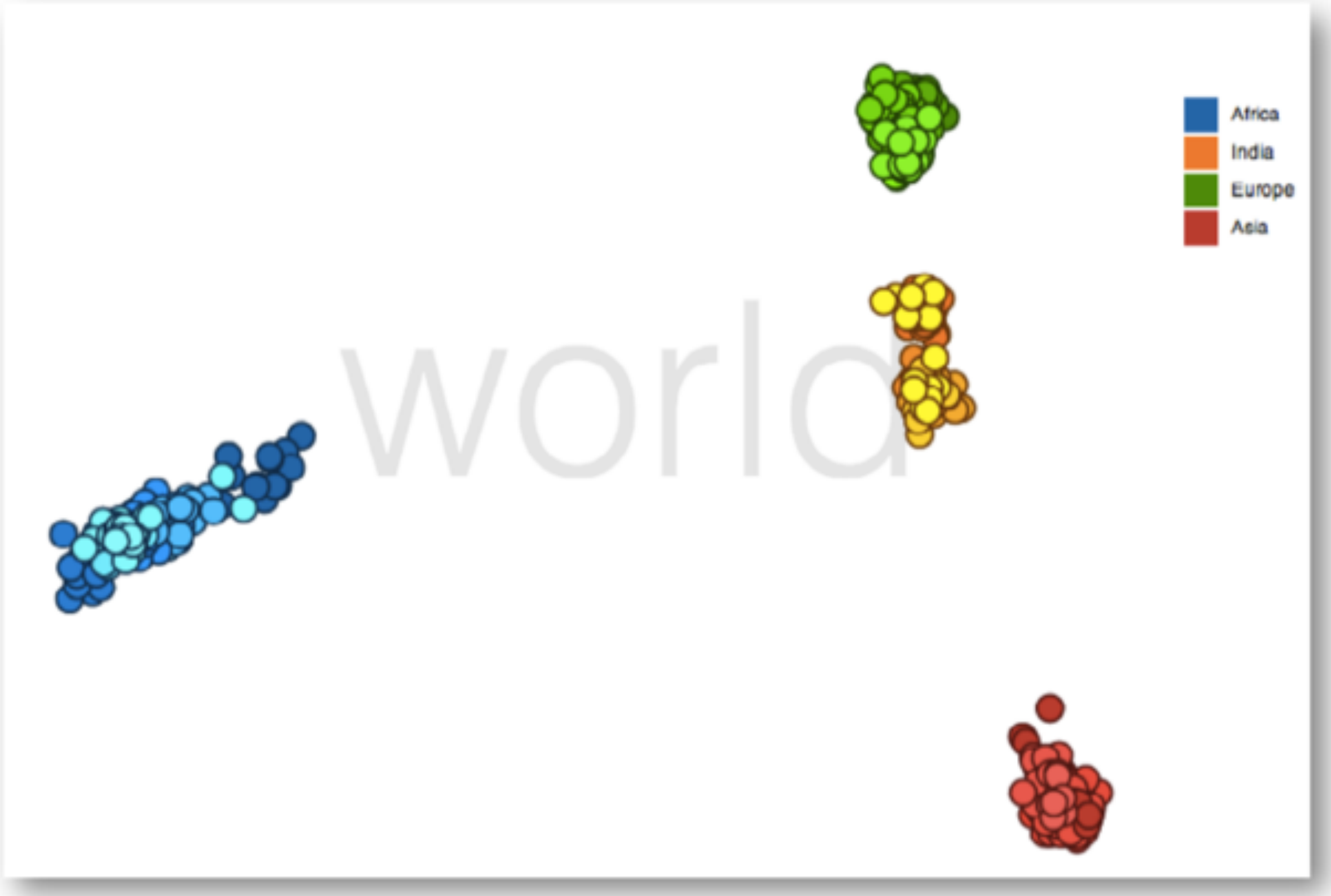


1. Mean-center the data.

2. Find ⊥ basis vectors that maximize the data variance.

3. Plot the data using the top vectors.

PCA on Genetic Sequences



Visualizing Multiple Dimensions

Strategies:
Avoid “over-encoding”
Use space and small multiples intelligently
Reduce the problem space
Use interaction to generate *relevant* views

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

Perception

Design Principles

What makes a visualization **“good”**?

Design Principles [Mackinlay 86]

Expressiveness

A set of facts is *expressible* in a visual language if the sentences (i.e. the visualizations) in the language express all the facts in the set of data, and only the facts in the data.

Effectiveness

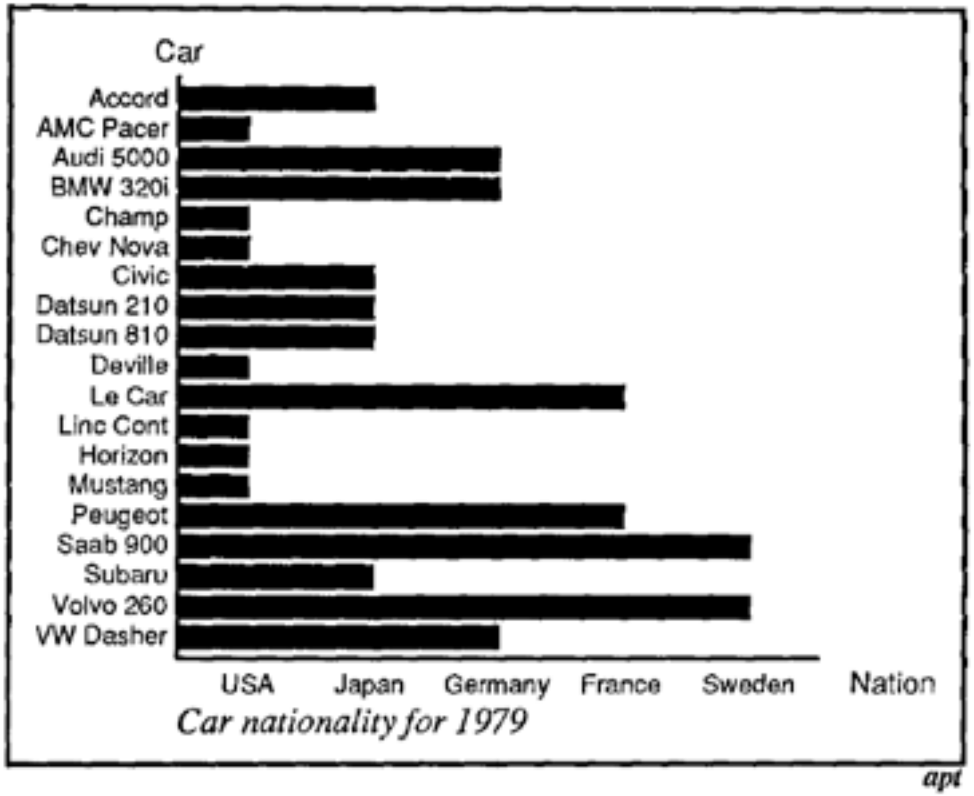
A visualization is more *effective* than another visualization if the information conveyed by one visualization is more readily perceived than the information in the other visualization.

Design Principles [Mackinlay 86]

Expressiveness
A set of facts is *expressible* in a visual language if the sentences (i.e. the visualizations) in the language express all the facts in the set of data, and only the facts in the data.

Effectiveness
A visualization is more *effective* than another visualization if the information conveyed by one visualization is more readily perceived than the information in the other visualization.

Expresses facts not in the data



apt

Fig. 11. Incorrect use of a bar chart for the *Nation* relation. The lengths of the bars suggest an ordering on the vertical axis, as if the USA cars were longer or better than the other cars, which is not true for the *Nation* relation.

A length is interpreted as a quantitative value.

Design Principles [Mackinlay 86]

Expressiveness

A set of facts is *expressible* in a visual language if the sentences (i.e. the visualizations) in the language express all the facts in the set of data, and only the facts in the data.

Effectiveness

A visualization is more *effective* than another visualization if the information conveyed by one visualization is more readily perceived than the information in the other visualization.

Design Principles [Mackinlay 86]

Expressiveness
A set of facts is *expressible* in a visual language if the sentences (i.e. the visualizations) in the language express all the facts in the set of data, and only the facts in the data.

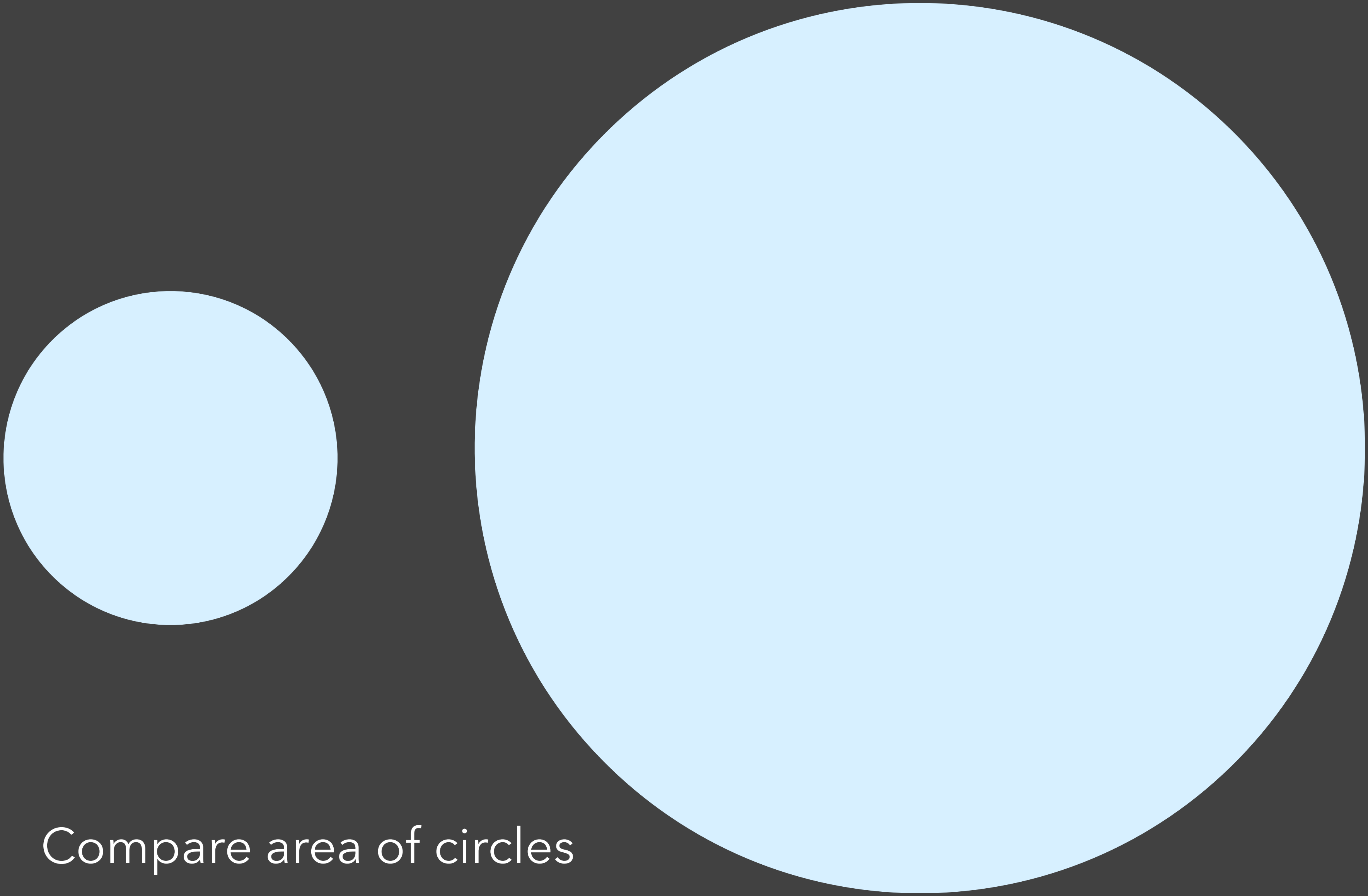
Effectiveness
A visualization is more *effective* than another visualization if the information conveyed by one visualization is more readily perceived than the information in the other visualization.

Design Principles *Translated*

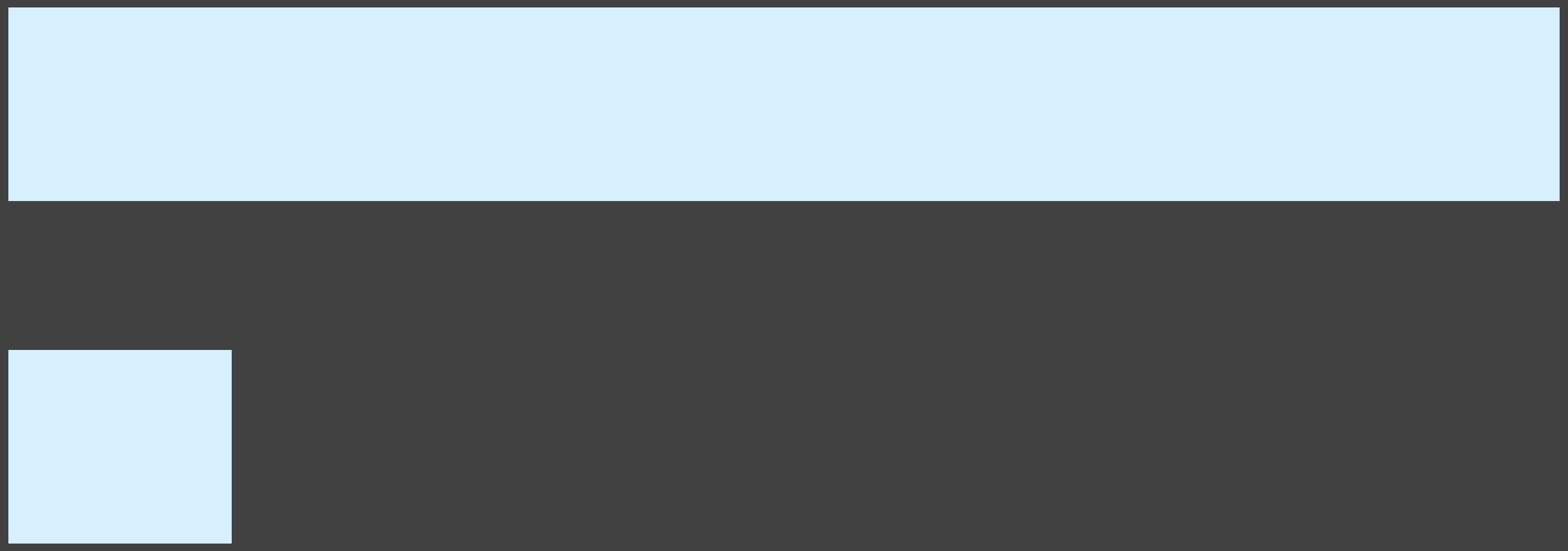
Tell the truth and nothing but the truth
(don’t lie, and don’t lie by omission)

Use encodings that people decode better
(where better = faster and/or more accurate)

Graphical Perception



Compare area of circles



Compare length of bars

Steven’s Power Law

Exponent

(Empirically Determined)

↓

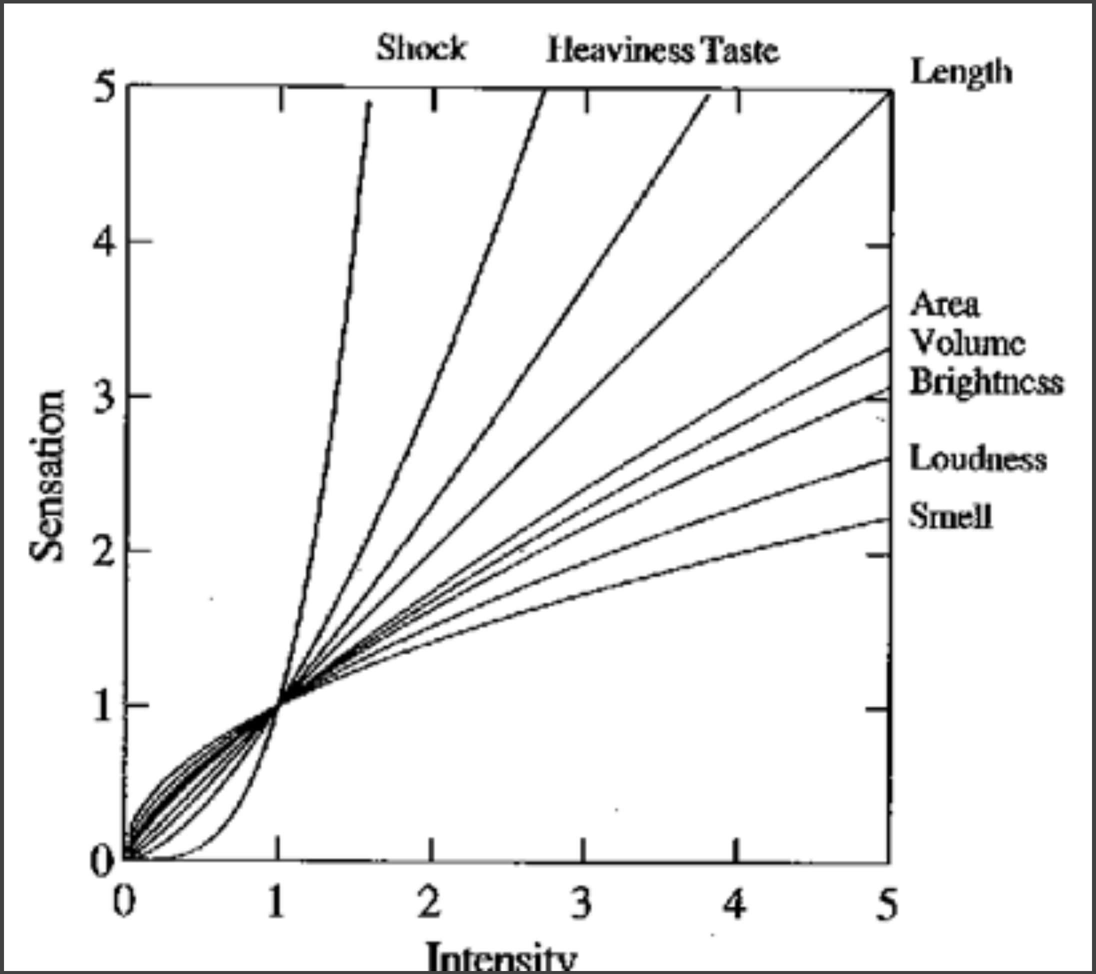
$$S = I^p$$

↑ ↑

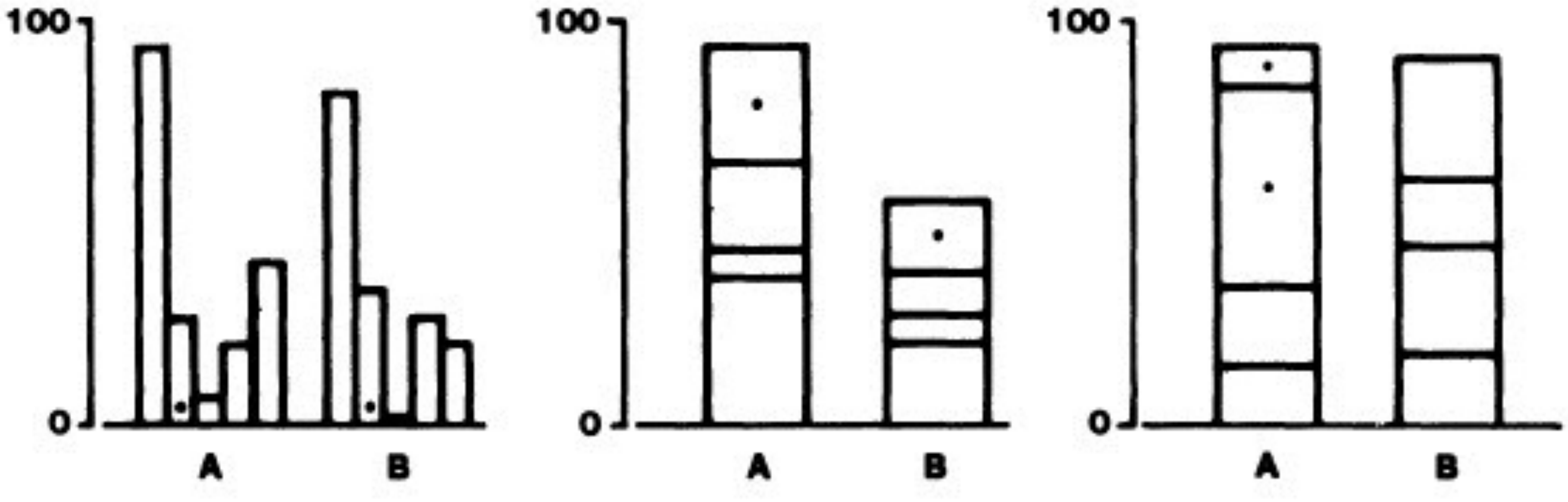
Perceived Sensation

Physical Intensity

Predicts bias, not necessarily accuracy!



[Graph from Wilkinson 99, based on Stevens 61]



Graphical Perception [Cleveland & McGill 84]

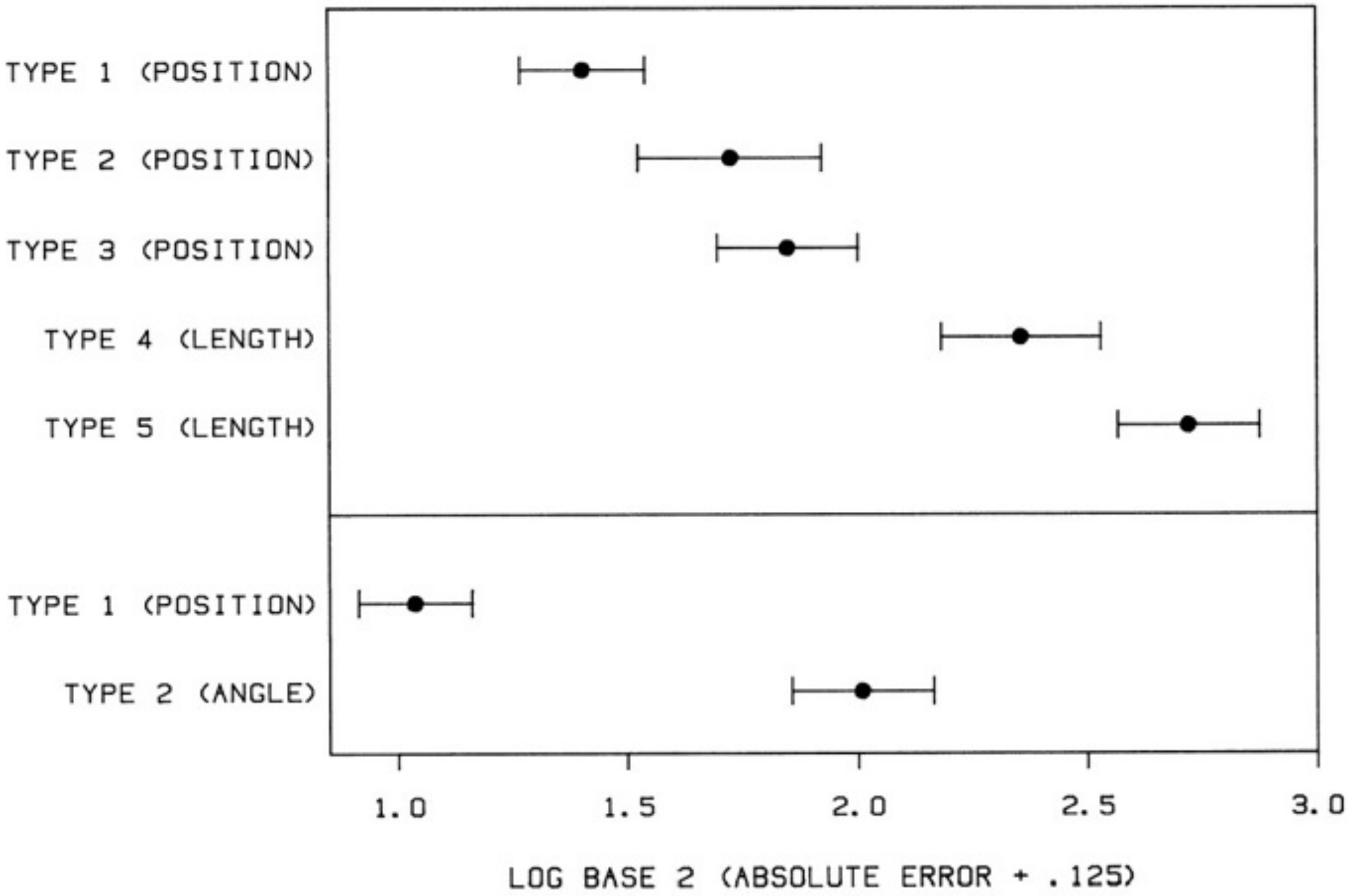
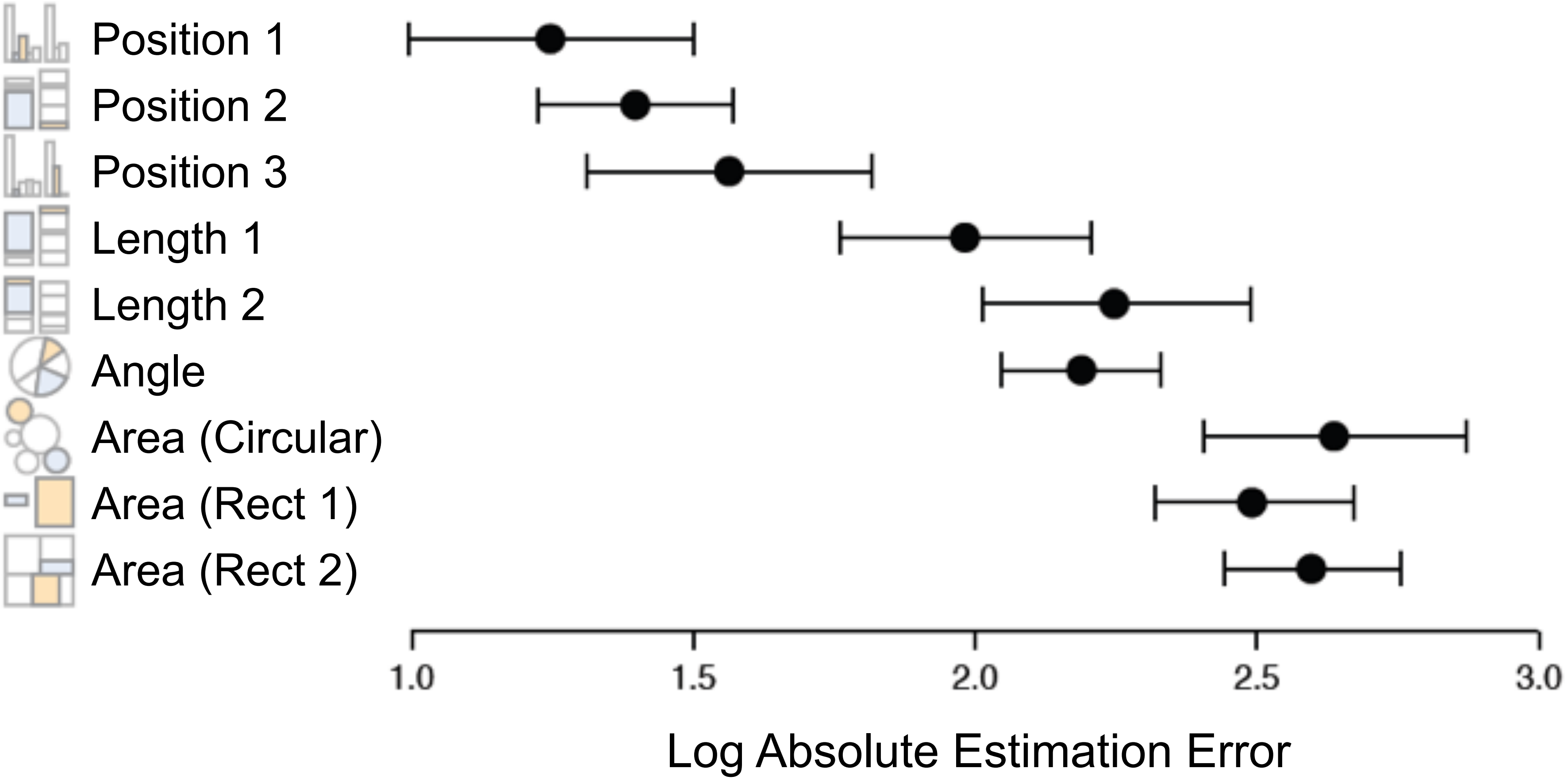


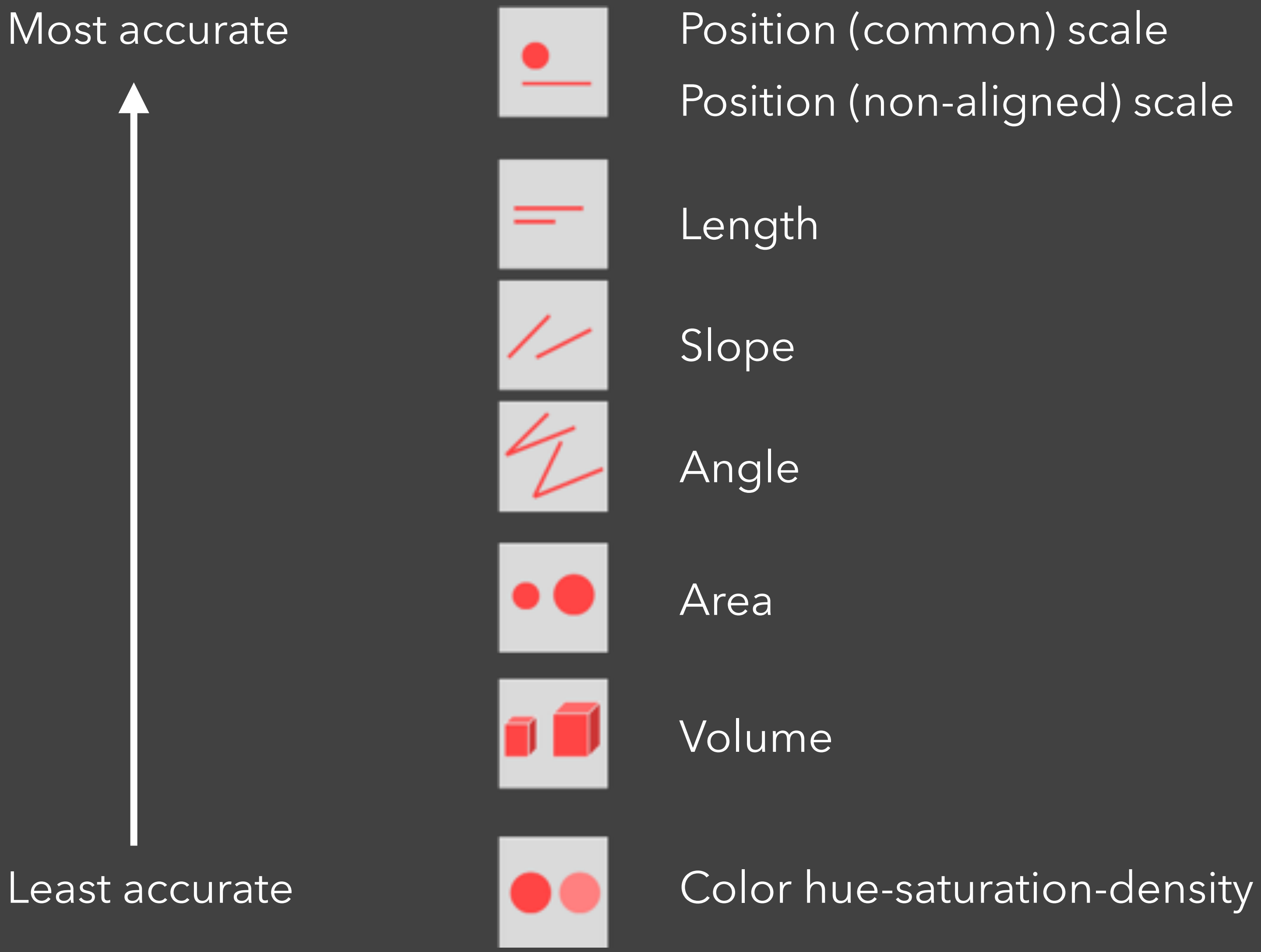
Figure 16. Log absolute error means and 95% confidence intervals for judgment types in position–length experiment (top) and position–angle experiment (bottom).



Graphical Perception Experiments

Empirical estimates of encoding effectiveness

Relative Magnitude Estimation



Most accurate

Position (common) scale

Position (non-aligned) scale

Length

Slope

Angle

Area

Volume

Color hue-saturation-density

Least accurate

Effectiveness Rankings [Mackinlay 86]

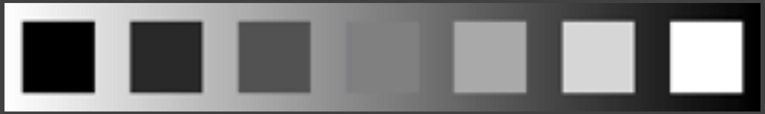
QUANTITATIVE	ORDINAL	NOMINAL
Position	Position	Position
Length	Density (Value)	Color Hue
Angle	Color Sat	Texture
Slope	Color Hue	Connection
Area (Size)	Texture	Containment
Volume	Connection	Density (Value)
Density (Value)	Containment	Color Sat
Color Sat	Length	Shape
Color Hue	Angle	Length
Texture	Slope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

Color

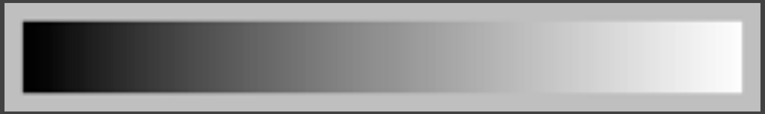
Encoding Data with Color

Value is perceived as ordered

∴ Encode ordinal variables (O)

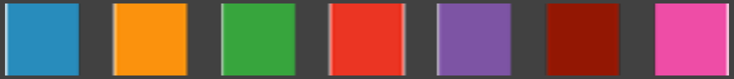


∴ Encode continuous variables (Q) [not as well]



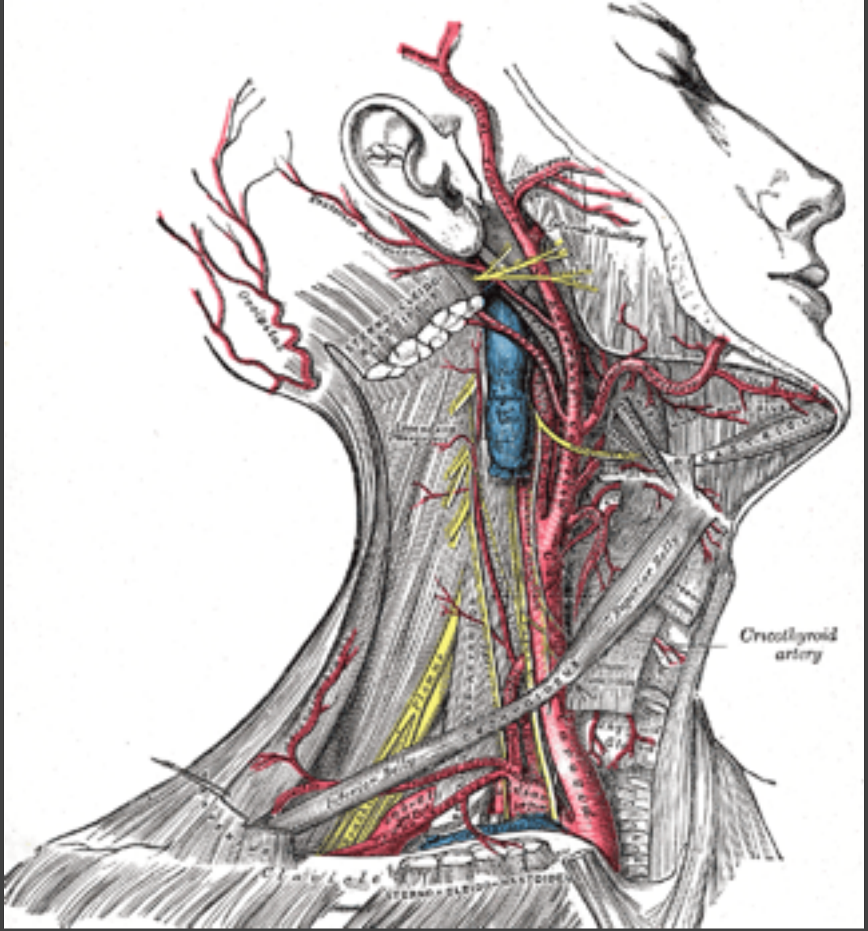
Hue is normally perceived as unordered

∴ Encode nominal variables (N) using color



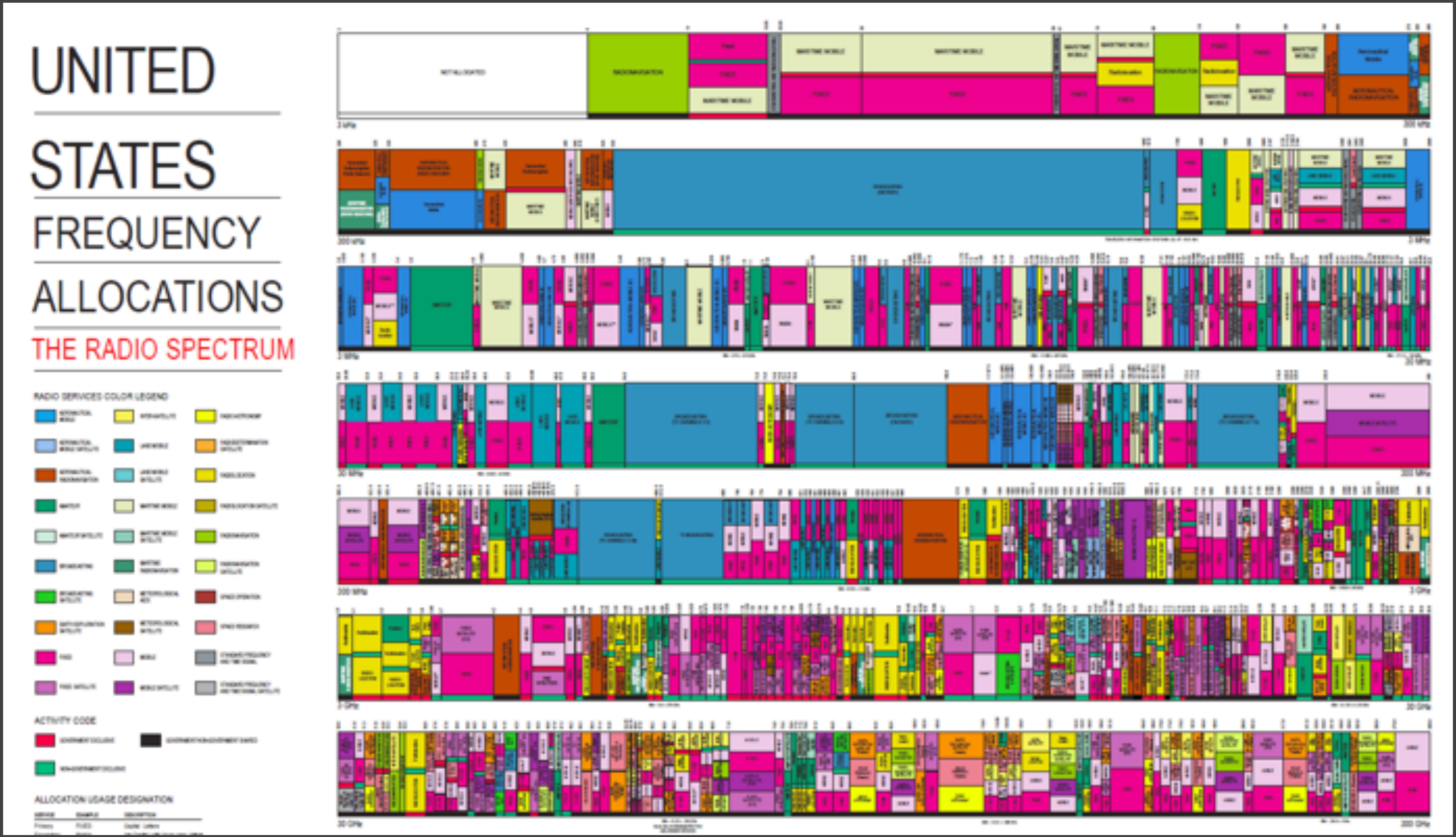
Categorical Color

Gray’s Anatomy



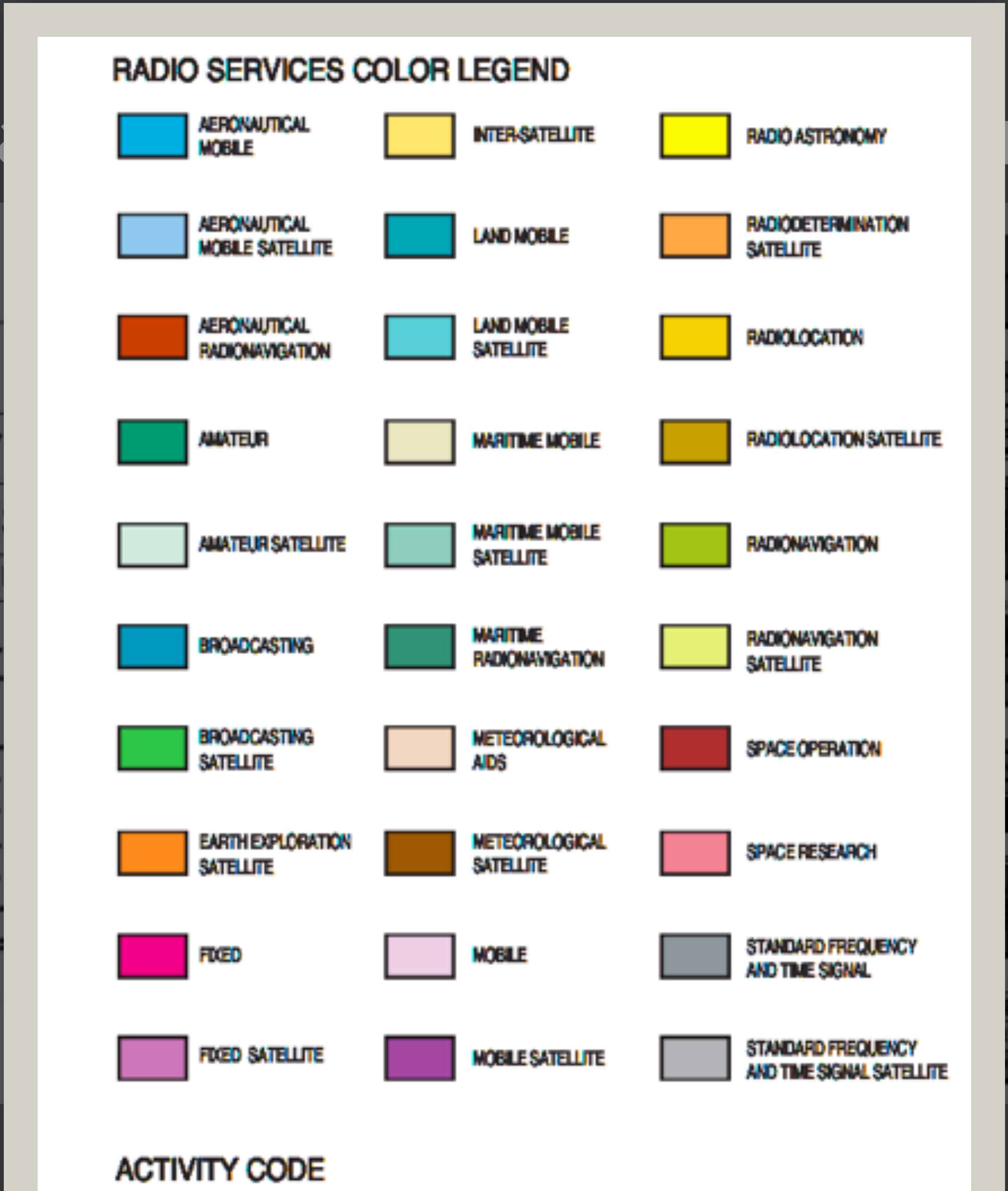
Superficial dissection of the right side of the neck, showing the carotid and subclavian arteries. (http://www.bartleby.com/107/illus520.html)

Allocation of the Radio Spectrum



http://www.ntia.doc.gov/osmhome/allochrt.html

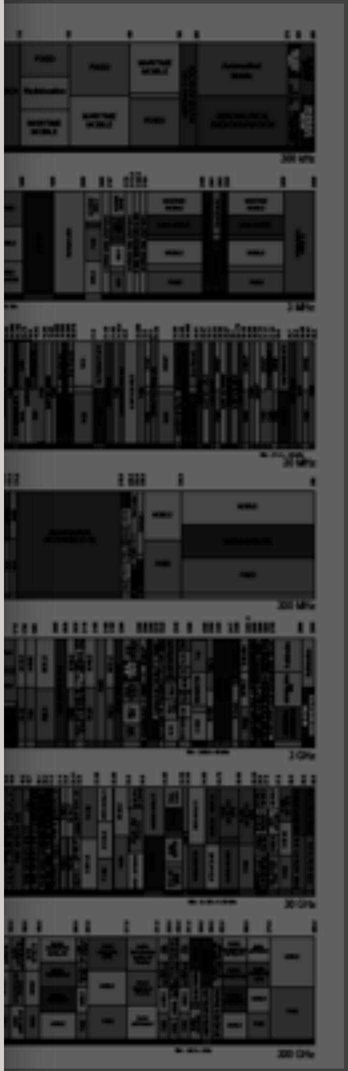
Alloca rum



RADIO SERVICES COLOR LEGEND

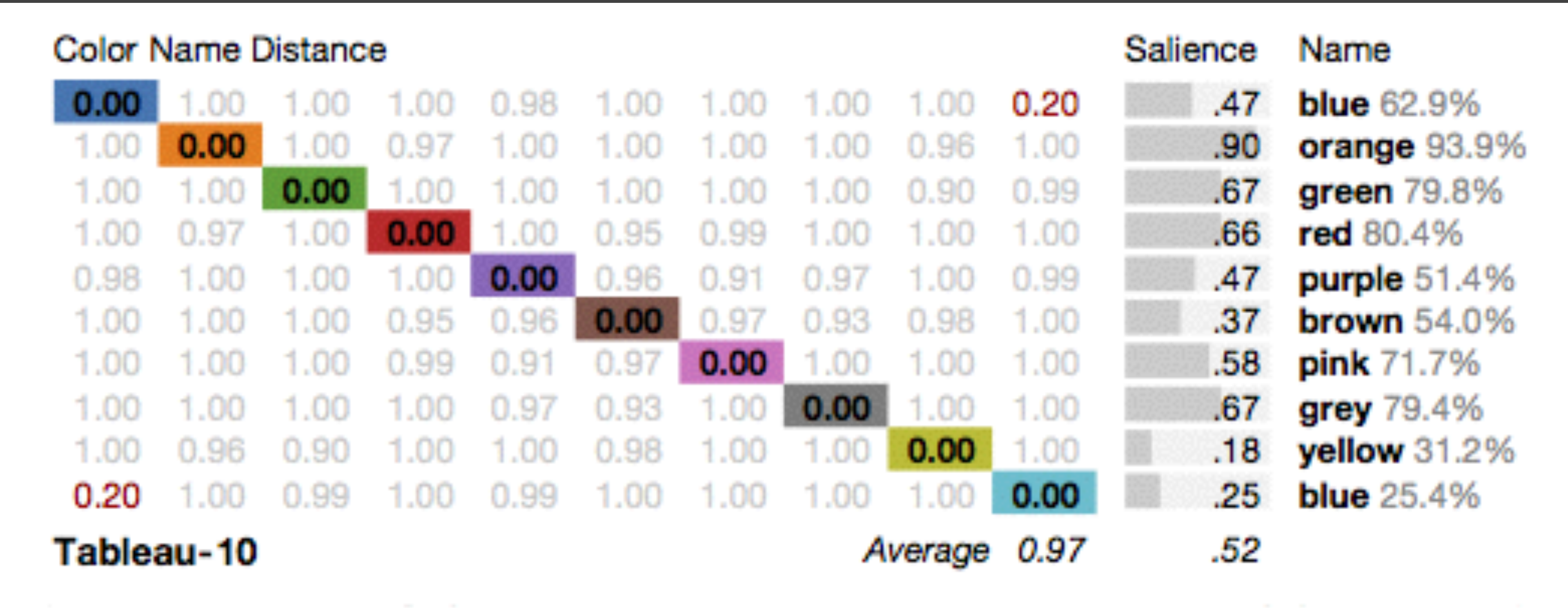
- 🟦 AERONAUTICAL MOBILE
- 🟦 AERONAUTICAL MOBILE SATELLITE
- 🟥 AERONAUTICAL RADIONAVIGATION
- 🟩 AMATEUR
- ⬜ AMATEUR SATELLITE
- 🟦 BROADCASTING
- 🟩 BROADCASTING SATELLITE
- 🟧 EARTH EXPLORATION SATELLITE
- 🟪 FIXED
- 🟪 FIXED SATELLITE
- 🟨 INTER-SATELLITE
- 🟦 LAND MOBILE
- 🟦 LAND MOBILE SATELLITE
- ⬜ MARITIME MOBILE
- 🟩 MARITIME MOBILE SATELLITE
- 🟩 MARITIME RADIONAVIGATION
- ⬜ METEOROLOGICAL AIDS
- 🟫 METEOROLOGICAL SATELLITE
- ⬜ MOBILE
- 🟪 MOBILE SATELLITE
- 🟨 RADIO ASTRONOMY
- 🟧 RADIODETERMINATION SATELLITE
- 🟨 RADIOLOCATION
- 🟫 RADIOLOCATION SATELLITE
- 🟩 RADIONAVIGATION
- 🟨 RADIONAVIGATION SATELLITE
- 🟥 SPACE OPERATION
- 🟥 SPACE RESEARCH
- ⬜ STANDARD FREQUENCY AND TIME SIGNAL
- ⬜ STANDARD FREQUENCY AND TIME SIGNAL SATELLITE

ACTIVITY CODE



Palette Design & Color Names

Minimize overlap and ambiguity of colors.

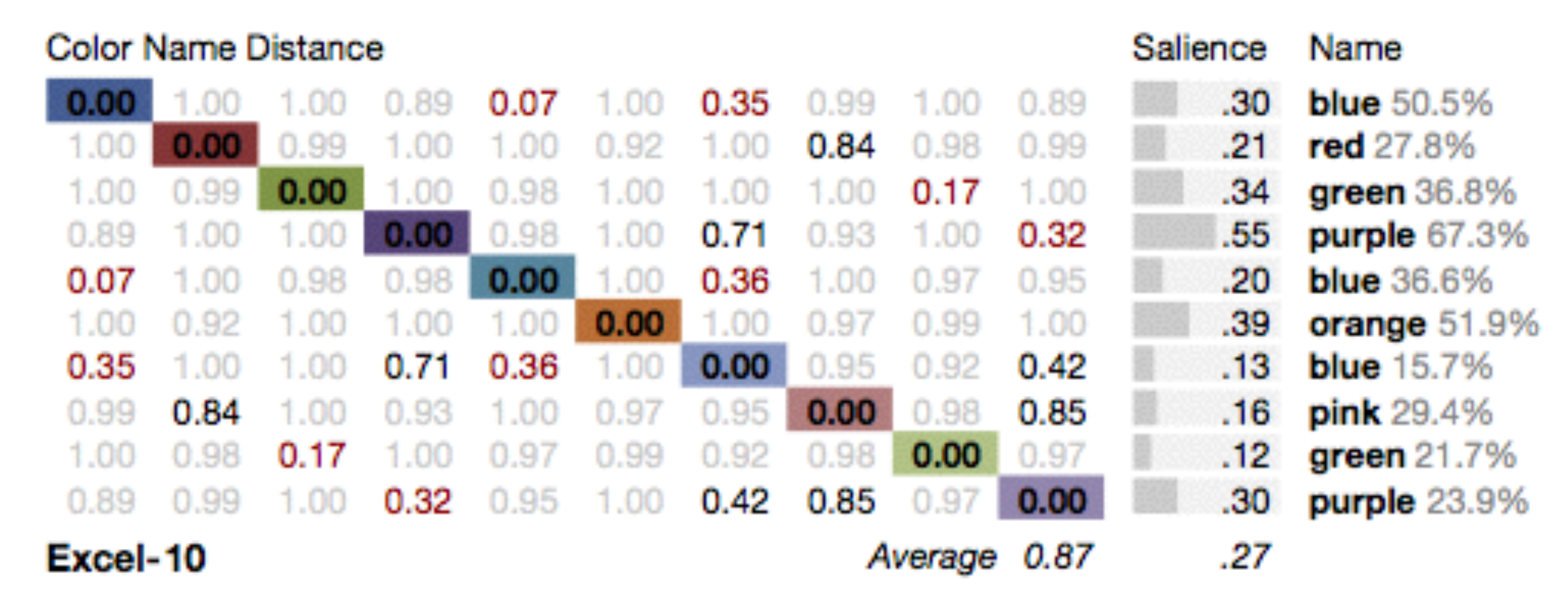


Color Name Distance										Salience	Name
0.00	1.00	1.00	1.00	0.98	1.00	1.00	1.00	1.00	0.20	.47	**blue** 62.9%
1.00	0.00	1.00	0.97	1.00	1.00	1.00	1.00	0.96	1.00	.90	**orange** 93.9%
1.00	1.00	0.00	1.00	1.00	1.00	1.00	1.00	0.90	0.99	.67	**green** 79.8%
1.00	0.97	1.00	0.00	1.00	0.95	0.99	1.00	1.00	1.00	.66	**red** 80.4%
0.98	1.00	1.00	1.00	0.00	0.96	0.91	0.97	1.00	0.99	.47	**purple** 51.4%
1.00	1.00	1.00	0.95	0.96	0.00	0.97	0.93	0.98	1.00	.37	**brown** 54.0%
1.00	1.00	1.00	0.99	0.91	0.97	0.00	1.00	1.00	1.00	.58	**pink** 71.7%
1.00	1.00	1.00	1.00	0.97	0.93	1.00	0.00	1.00	1.00	.67	**grey** 79.4%
1.00	0.96	0.90	1.00	1.00	0.98	1.00	1.00	0.00	1.00	.18	**yellow** 31.2%
0.20	1.00	0.99	1.00	0.99	1.00	1.00	1.00	1.00	0.00	.25	**blue** 25.4%
Tableau-10								*Average*	*0.97*	*.52*	

http://vis.stanford.edu/color-names

Palette Design & Color Names

Minimize overlap and ambiguity of colors.

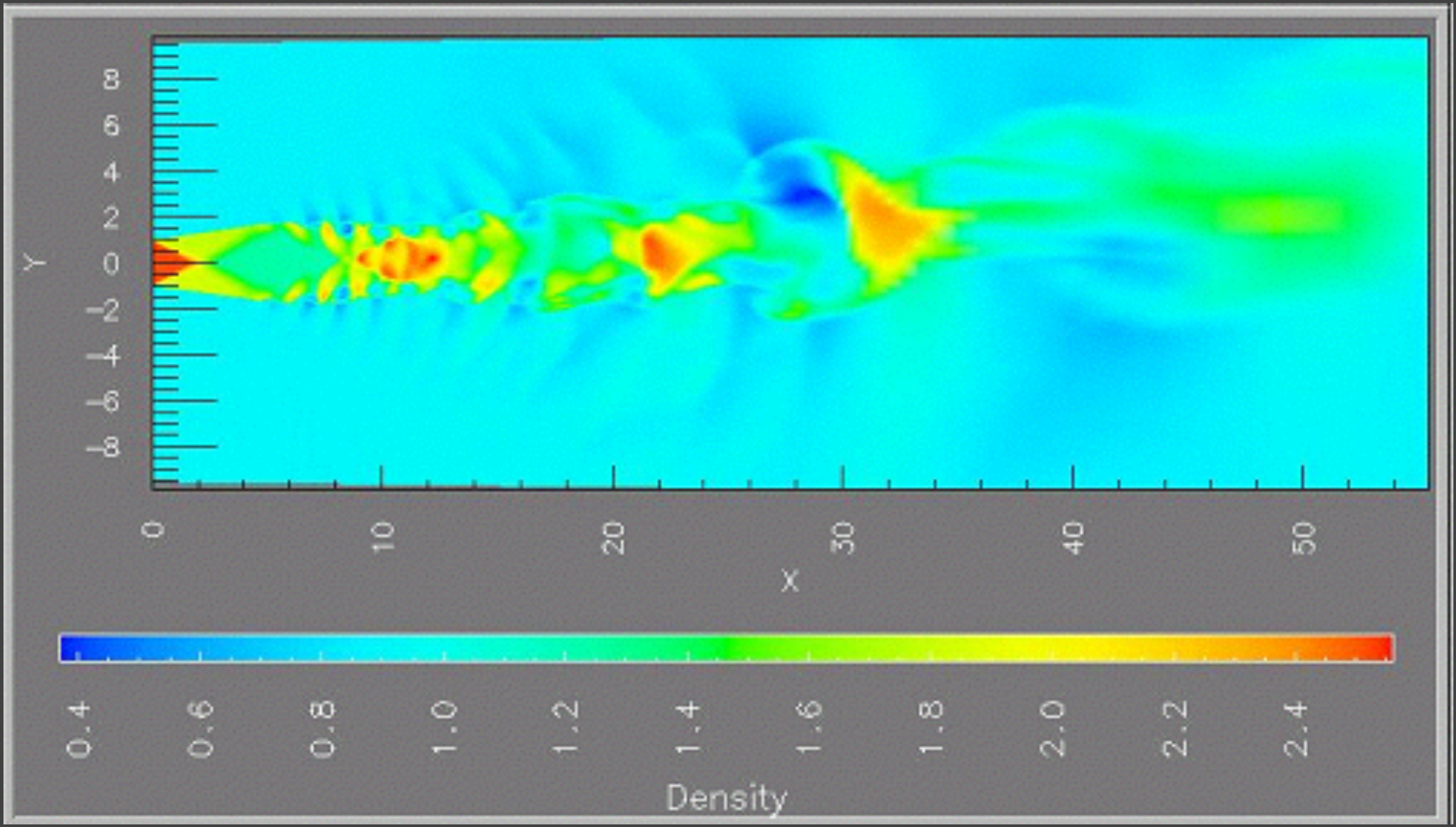


Color Name Distance										Salience	Name
0.00	1.00	1.00	0.89	0.07	1.00	0.35	0.99	1.00	0.89	.30	**blue** 50.5%
1.00	0.00	0.99	1.00	1.00	0.92	1.00	0.84	0.98	0.99	.21	**red** 27.8%
1.00	0.99	0.00	1.00	0.98	1.00	1.00	1.00	0.17	1.00	.34	**green** 36.8%
0.89	1.00	1.00	0.00	0.98	1.00	0.71	0.93	1.00	0.32	.55	**purple** 67.3%
0.07	1.00	0.98	0.98	0.00	1.00	0.36	1.00	0.97	0.95	.20	**blue** 36.6%
1.00	0.92	1.00	1.00	1.00	0.00	1.00	0.97	0.99	1.00	.39	**orange** 51.9%
0.35	1.00	1.00	0.71	0.36	1.00	0.00	0.95	0.92	0.42	.13	**blue** 15.7%
0.99	0.84	1.00	0.93	1.00	0.97	0.95	0.00	0.98	0.85	.16	**pink** 29.4%
1.00	0.98	0.17	1.00	0.97	0.99	0.92	0.98	0.00	0.97	.12	**green** 21.7%
0.89	0.99	1.00	0.32	0.95	1.00	0.42	0.85	0.97	0.00	.30	**purple** 23.9%
Excel-10								*Average*	*0.87*	*.27*	

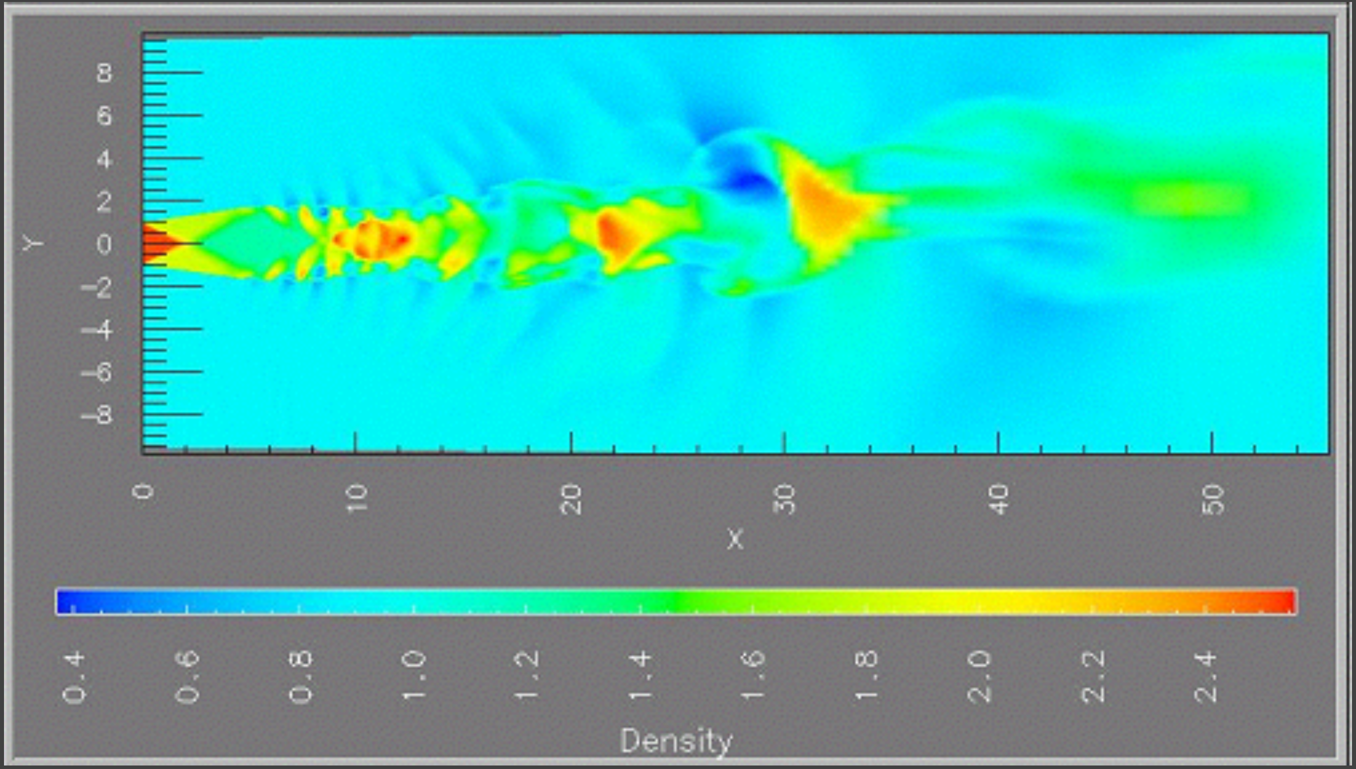
http://vis.stanford.edu/color-names

Quantitative Color

Rainbow Color Maps



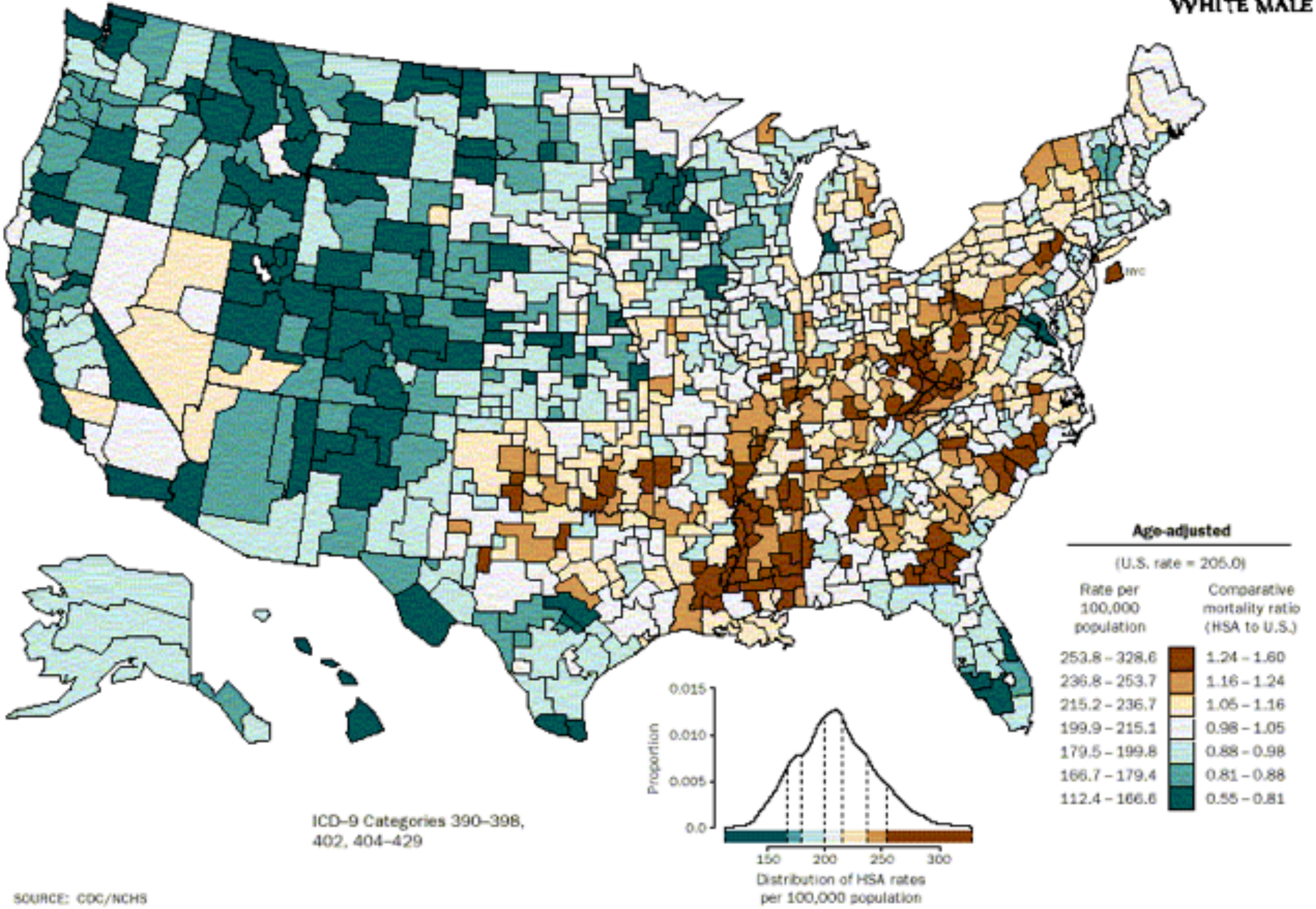
Be Wary of Rainbows!



1. People segment colors into classes
2. Hues are not naturally ordered
3. Different lightness emphasizes certain scalar values
4. Low luminance colors (blue) hide high frequencies

Age-adjusted death rates by HSA, 1988-92

Heart disease
White male



Age-adjusted

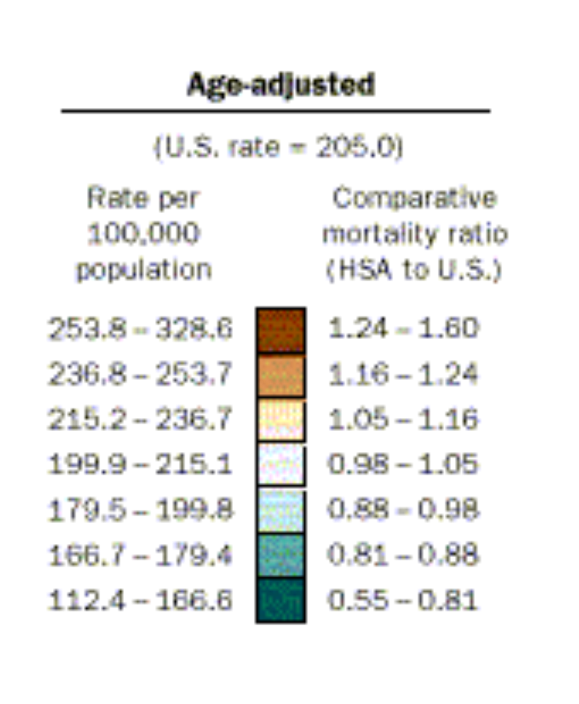
(U.S. rate = 205.0)

Rate per 100,000 population		Comparative mortality ratio (HSA to U.S.)
253.8 – 328.6	🟫	1.24 – 1.60
236.8 – 253.7	🟫	1.16 – 1.24
215.2 – 236.7	🟨	1.05 – 1.16
199.9 – 215.1	⬜	0.98 – 1.05
179.5 – 199.8	🟦	0.88 – 0.98
166.7 – 179.4	🟩	0.81 – 0.88
112.4 – 166.6	🟩	0.55 – 0.81

ICD–9 Categories 390–398,
402, 404–429

SOURCE: CDC/NCHS

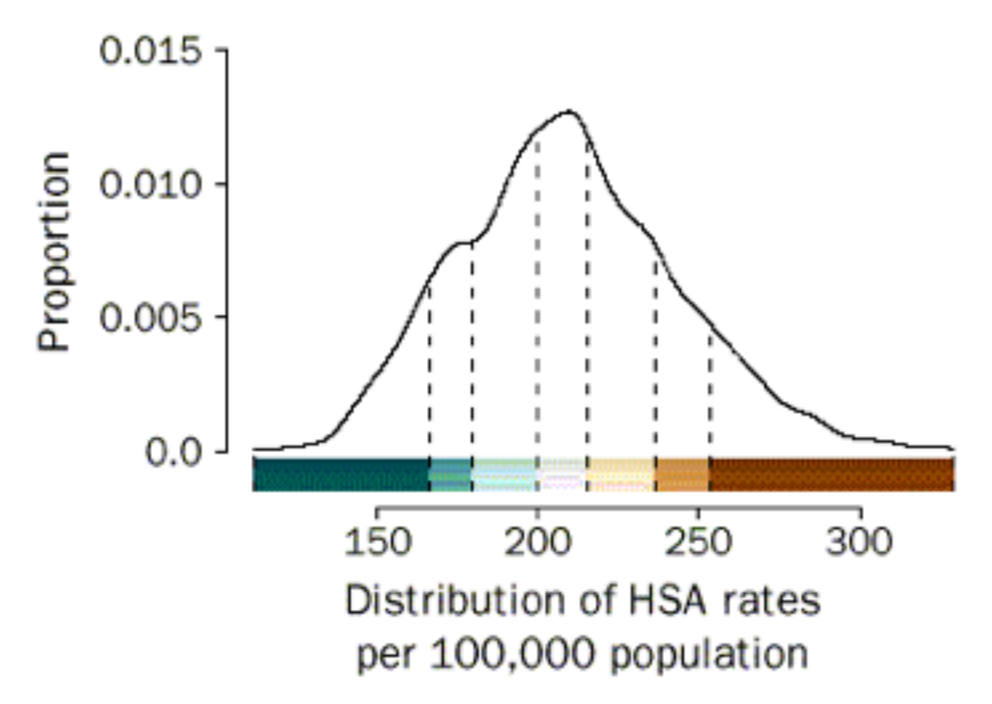
Classing Quantitative Data



Age-adjusted

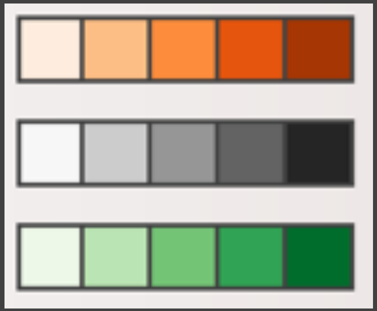
(U.S. rate = 205.0)

Rate per 100,000 population		Comparative mortality ratio (HSA to U.S.)
253.8 – 328.6	🟫	1.24 – 1.60
236.8 – 253.7	🟫	1.16 – 1.24
215.2 – 236.7	🟨	1.05 – 1.16
199.9 – 215.1	⬜	0.98 – 1.05
179.5 – 199.8	🟦	0.88 – 0.98
166.7 – 179.4	🟦	0.81 – 0.88
112.4 – 166.6	🟩	0.55 – 0.81



Age-adjusted mortality rates for the United States.
Common option: break into 5 or 7 quantiles.

Quantitative Color Encoding



Sequential color scale

Constrain hue, vary luminance/saturation

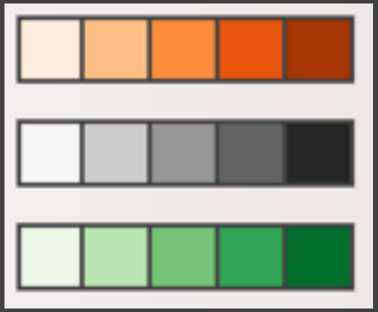
Map higher values to darker colors

Quantitative Color Encoding

Sequential color scale

Constrain hue, vary luminance/saturation

Map higher values to darker colors

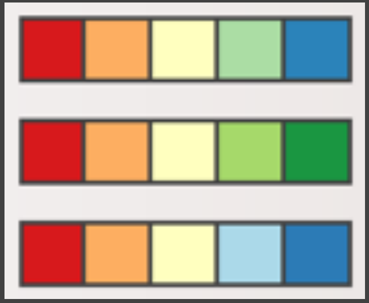


Diverging color scale

Useful when data has meaningful "midpoint"

Use neutral color (e.g., grey) for midpoint

Use saturated colors for endpoints

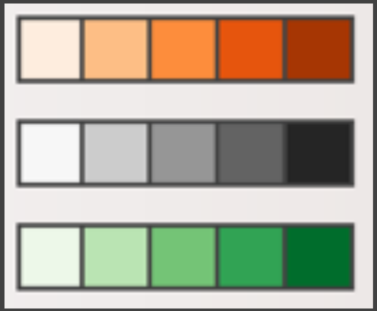


Quantitative Color Encoding

Sequential color scale

Constrain hue, vary luminance/saturation

Map higher values to darker colors

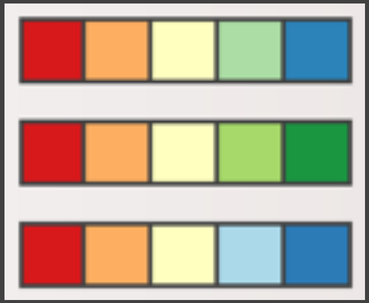


Diverging color scale

Useful when data has meaningful “midpoint”

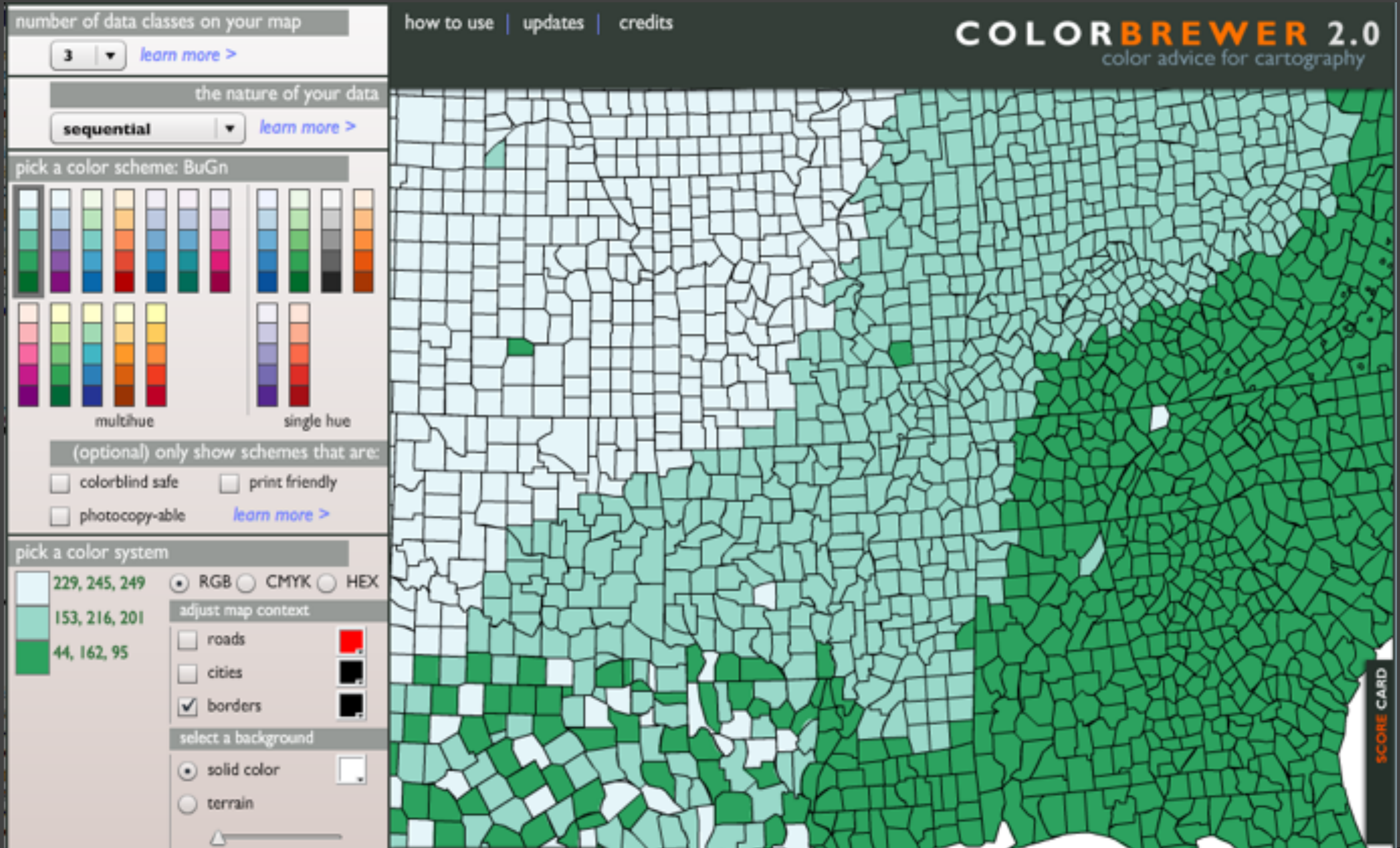
Use neutral color (e.g., grey) for midpoint

Use saturated colors for endpoints



Limit number of steps in color to 3-9

Color Brewer: Palettes for Maps



number of data classes on your map

3 ▾ learn more >

the nature of your data

sequential ▾ learn more >

pick a color scheme: BuGn

multihue

single hue

(optional) only show schemes that are:

□ colorblind safe □ print friendly

□ photocopy-able learn more >

pick a color system

⬜	229, 245, 249
🟩	153, 216, 201
🟩	44, 162, 95

● RGB ○ CMYK ○ HEX

adjust map context

□ roads

□ cities

☑ borders

select a background

● solid color

○ terrain

Hints for the Colorist

Hints for the Colorist

Use **only a few** colors (~6 ideal)

Hints for the Colorist

Use **only a few** colors (~6 ideal)

Colors should be **distinctive** and **named**

Hints for the Colorist

Use **only a few** colors (~6 ideal)

Colors should be **distinctive** and **named**

Strive for color **harmony** (natural colors?)

Hints for the Colorist

Use **only a few** colors (~6 ideal)

Colors should be **distinctive** and **named**

Strive for color **harmony** (natural colors?)

Use **cultural conventions**; appreciate symbolism

Hints for the Colorist

Use **only a few** colors (~6 ideal)

Colors should be **distinctive** and **named**

Strive for color **harmony** (natural colors?)

Use **cultural conventions**; appreciate symbolism

Get it right in **black and white**

Hints for the Colorist

Use **only a few** colors (~6 ideal)

Colors should be **distinctive** and **named**

Strive for color **harmony** (natural colors?)

Use **cultural conventions**; appreciate symbolism

Get it right in **black and white**

Respect the **color blind**

Hints for the Colorist

Use **only a few** colors (~6 ideal)

Colors should be **distinctive** and **named**

Strive for color **harmony** (natural colors?)

Use **cultural conventions**; appreciate symbolism

Get it right in **black and white**

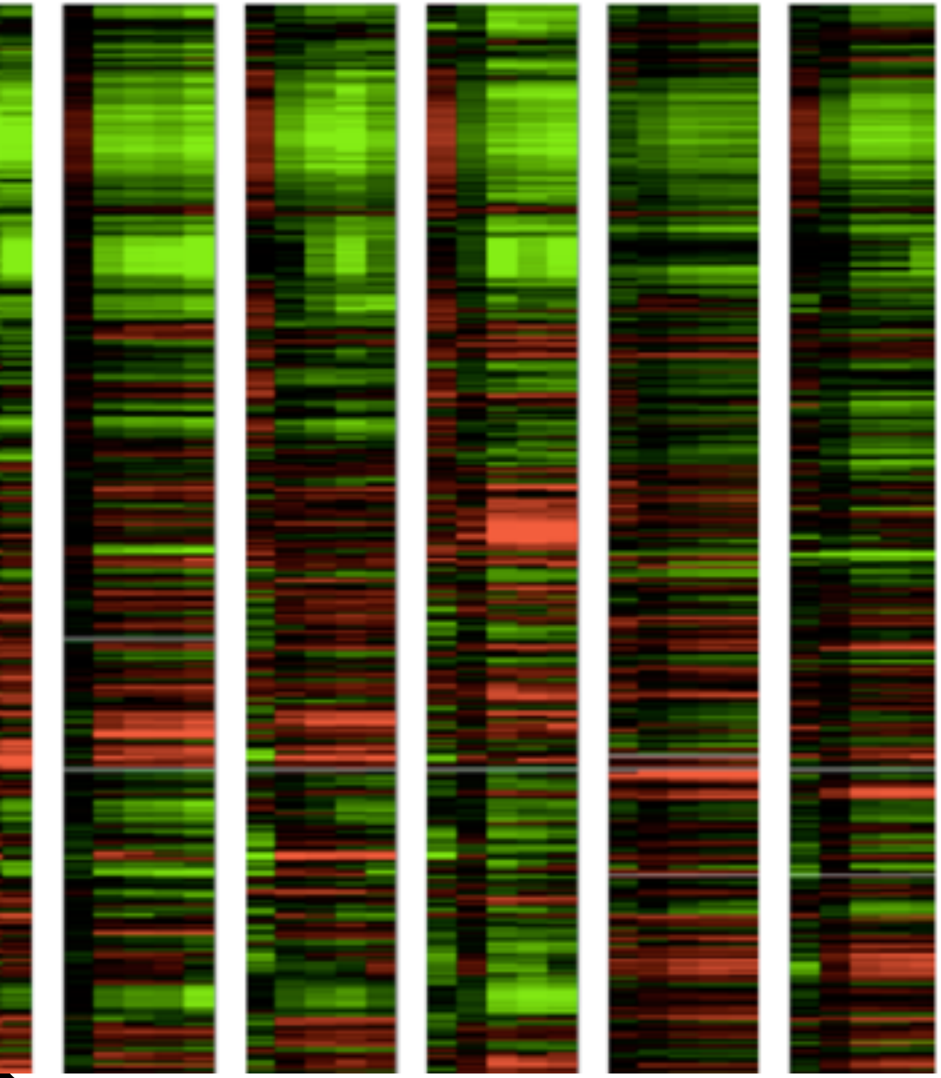
Respect the **color blind**

Take advantage of **perceptual color spaces**

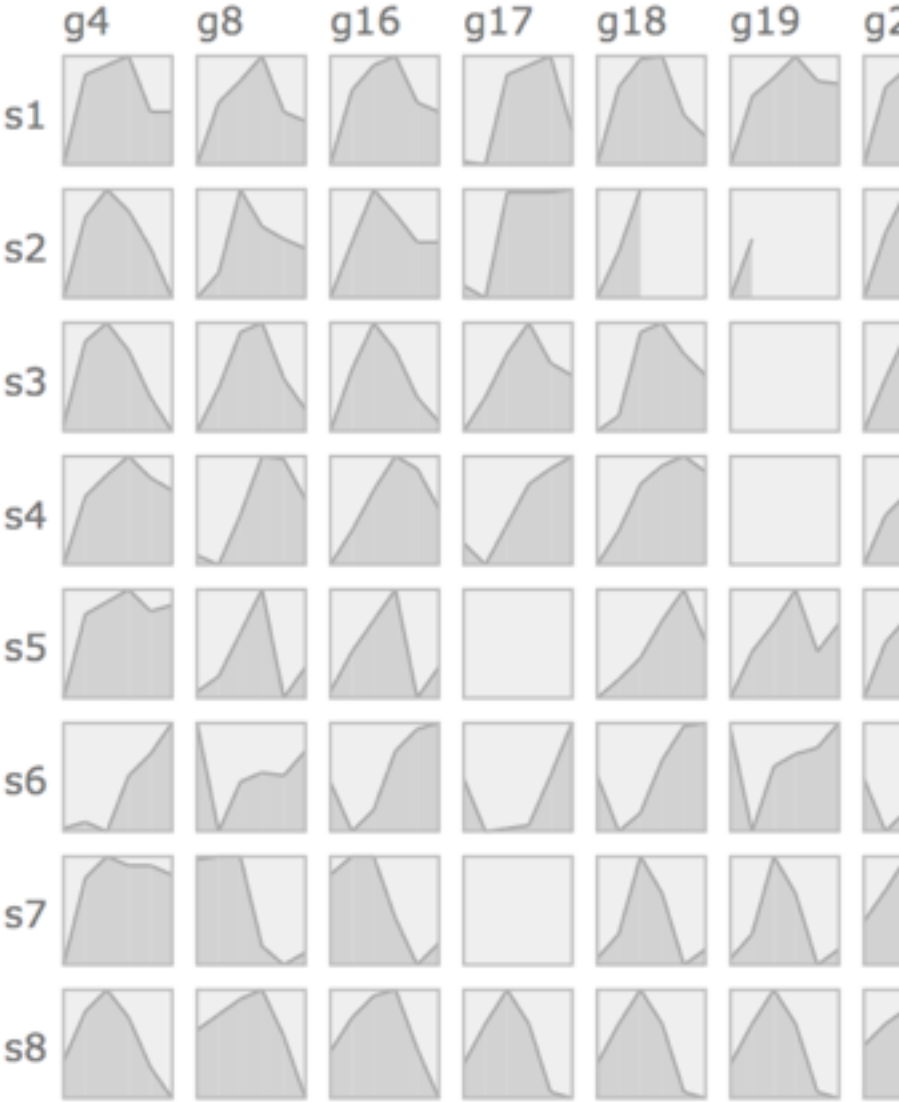
Perceptual Re-designs

Gene Expression Time-Series [Meyer et al 11]

Color Encoding



Position Encoding



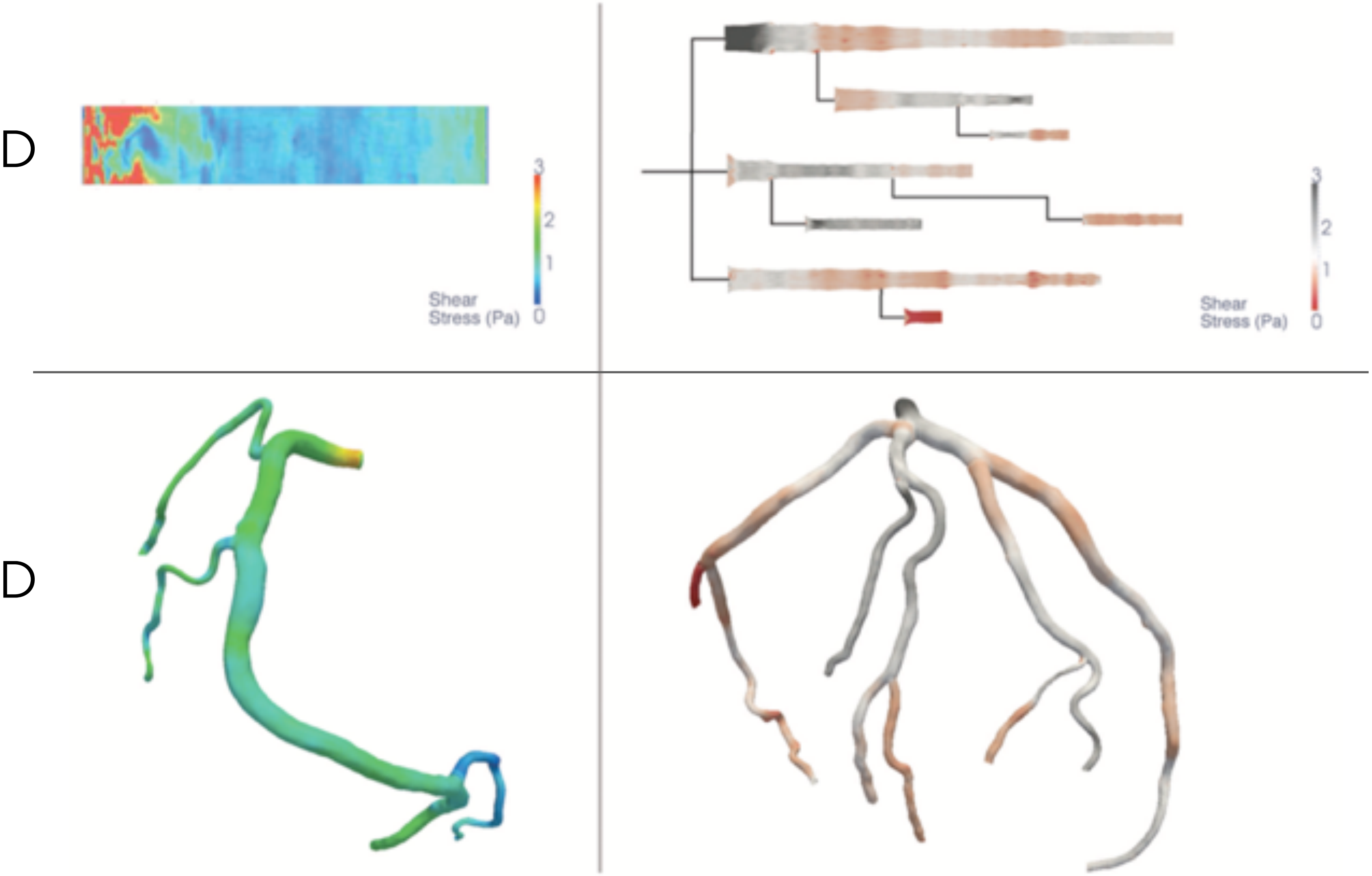
Artery Visualization [Borkin et al 11]

Rainbow Palette

Diverging Palette

2D

3D



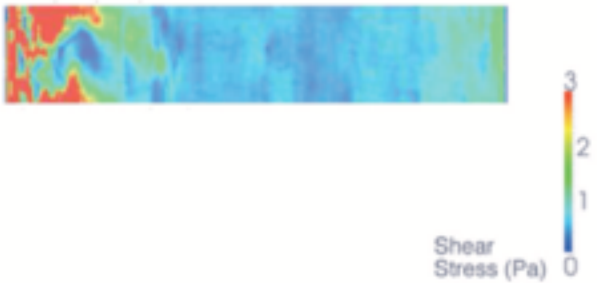
Artery Visualization [Borkin et al 11]

Rainbow Palette

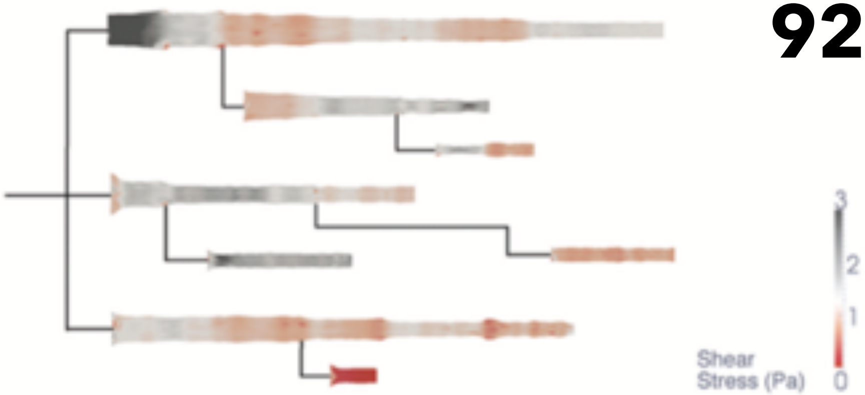
Diverging Palette

Accuracy: **62%**

2D

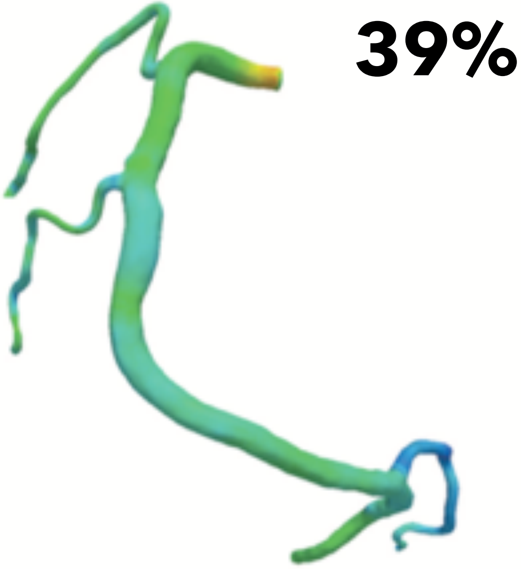


92%

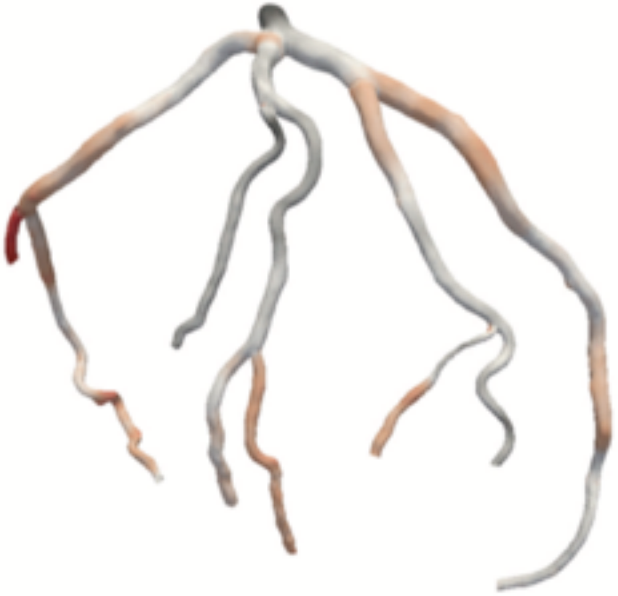


3D

39%



71%



Interaction

Taxonomy of Interactions

Taxonomy of Interactions

Data and View Specification

Visualize, Filter, Sort, Derive

Taxonomy of Interactions

Data and View Specification

Visualize, Filter, Sort, Derive

View Manipulation

Select, Navigate, Coordinate, Organize

Taxonomy of Interactions

Data and View Specification

Visualize, Filter, Sort, Derive

View Manipulation

Select, Navigate, Coordinate, Organize

Process and Provenance

Record, Annotate, Share, Guide

Selection

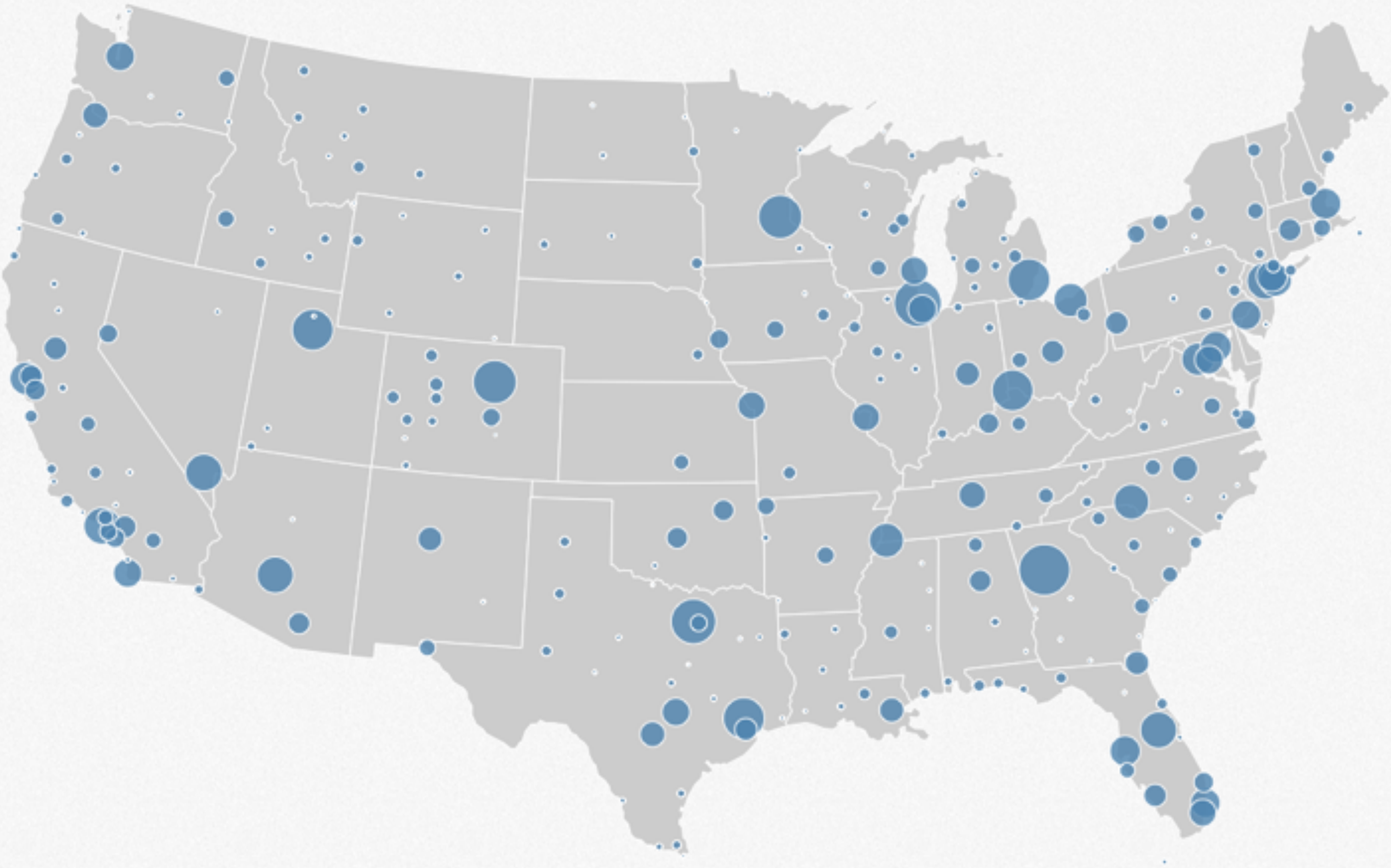
Basic Selection Methods

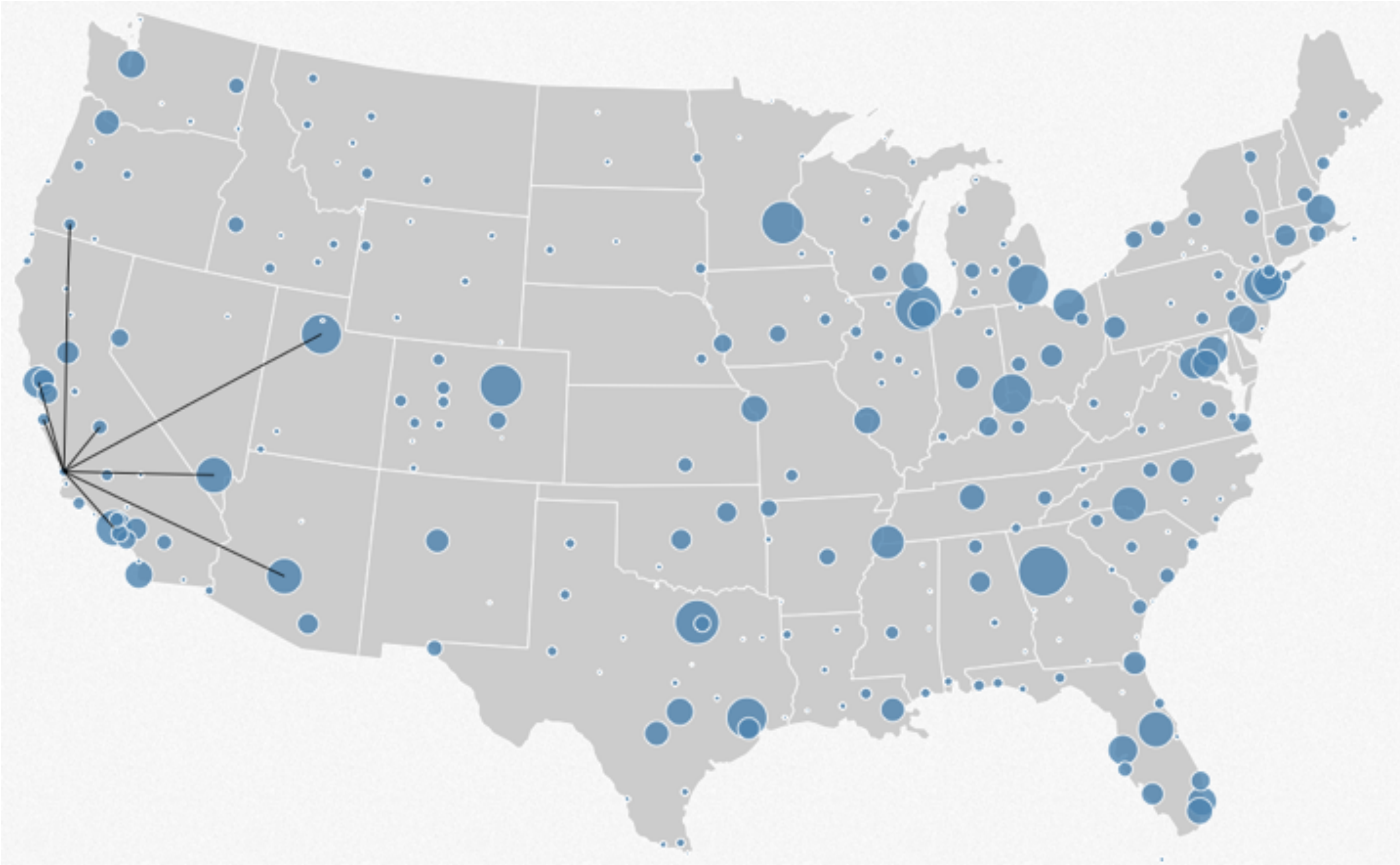
Point Selection

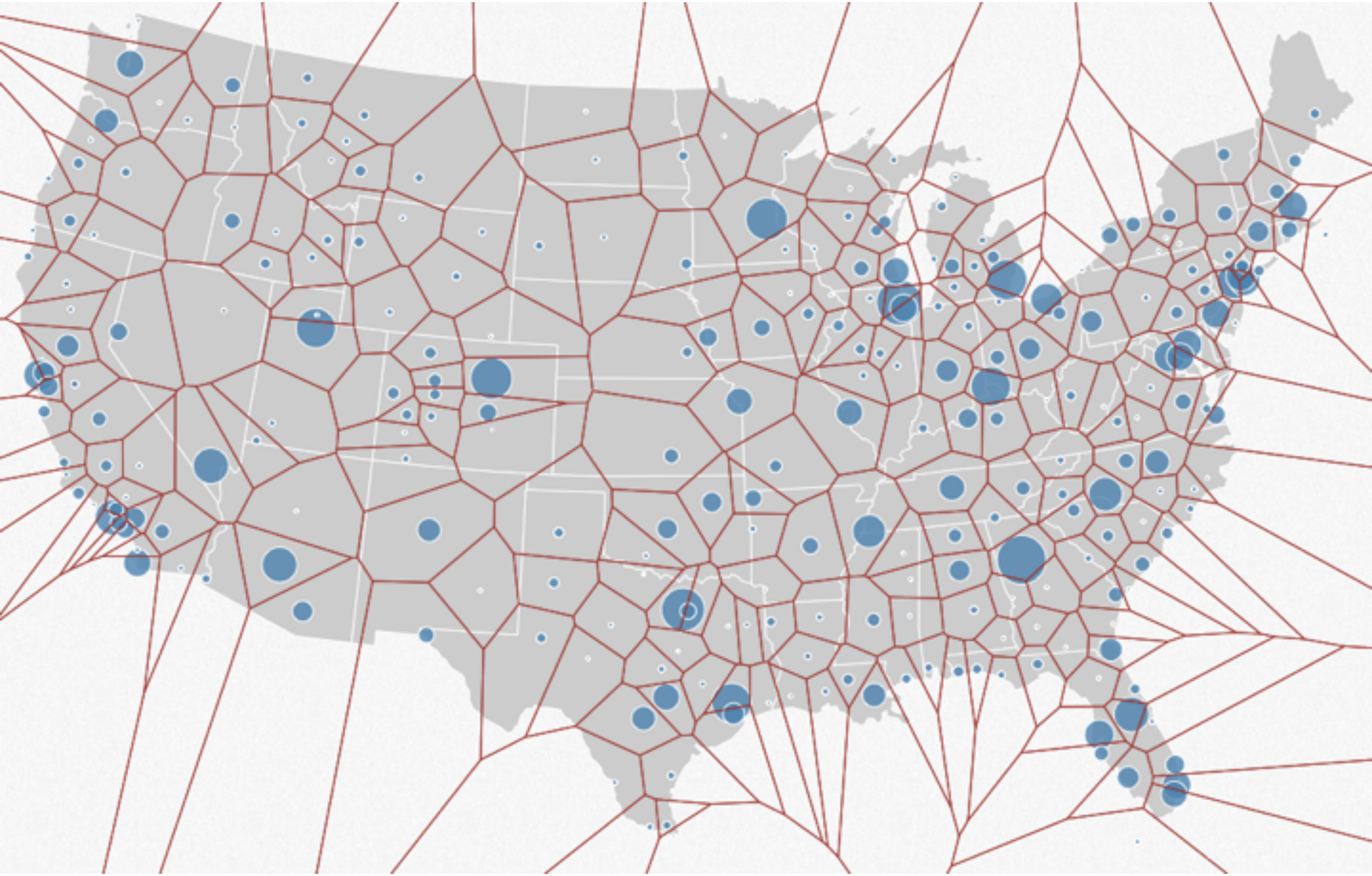
Mouse Hover / Click

Touch / Tap

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







Basic Selection Methods

Point Selection

Mouse Hover / Click

Touch / Tap

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

Region Selection

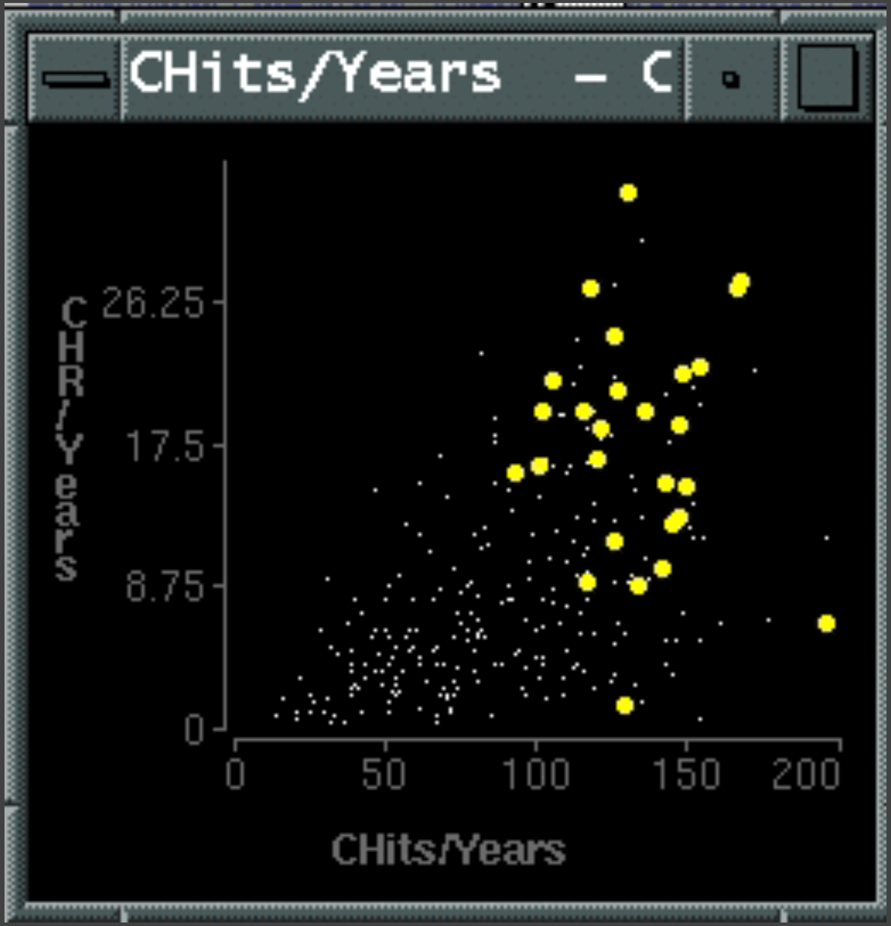
Rubber-band (rectangular) or Lasso (freehand)

Area cursors (“brushes”)

Brushing & Linking

Brushing

Direct attention to a subset of data [Wills 95]

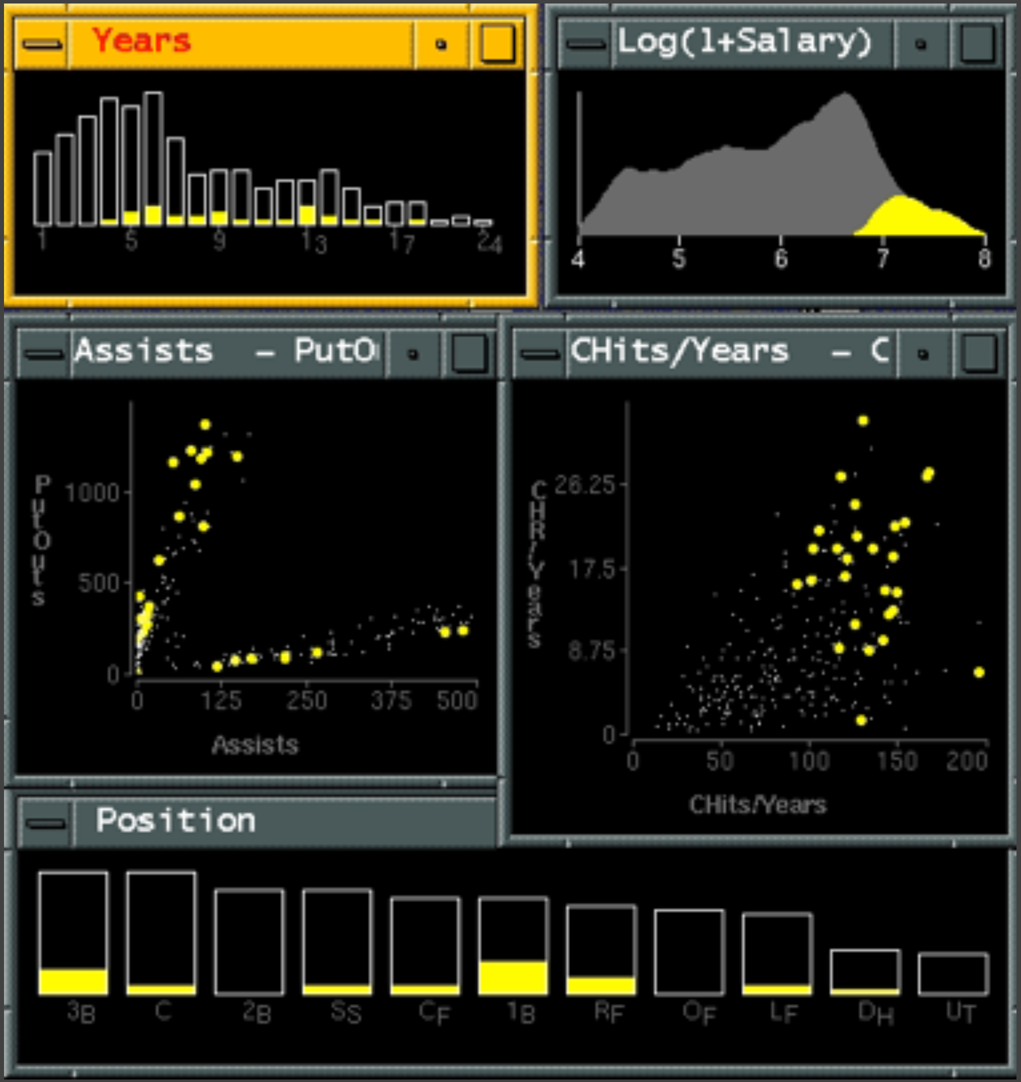


Brushing & Linking

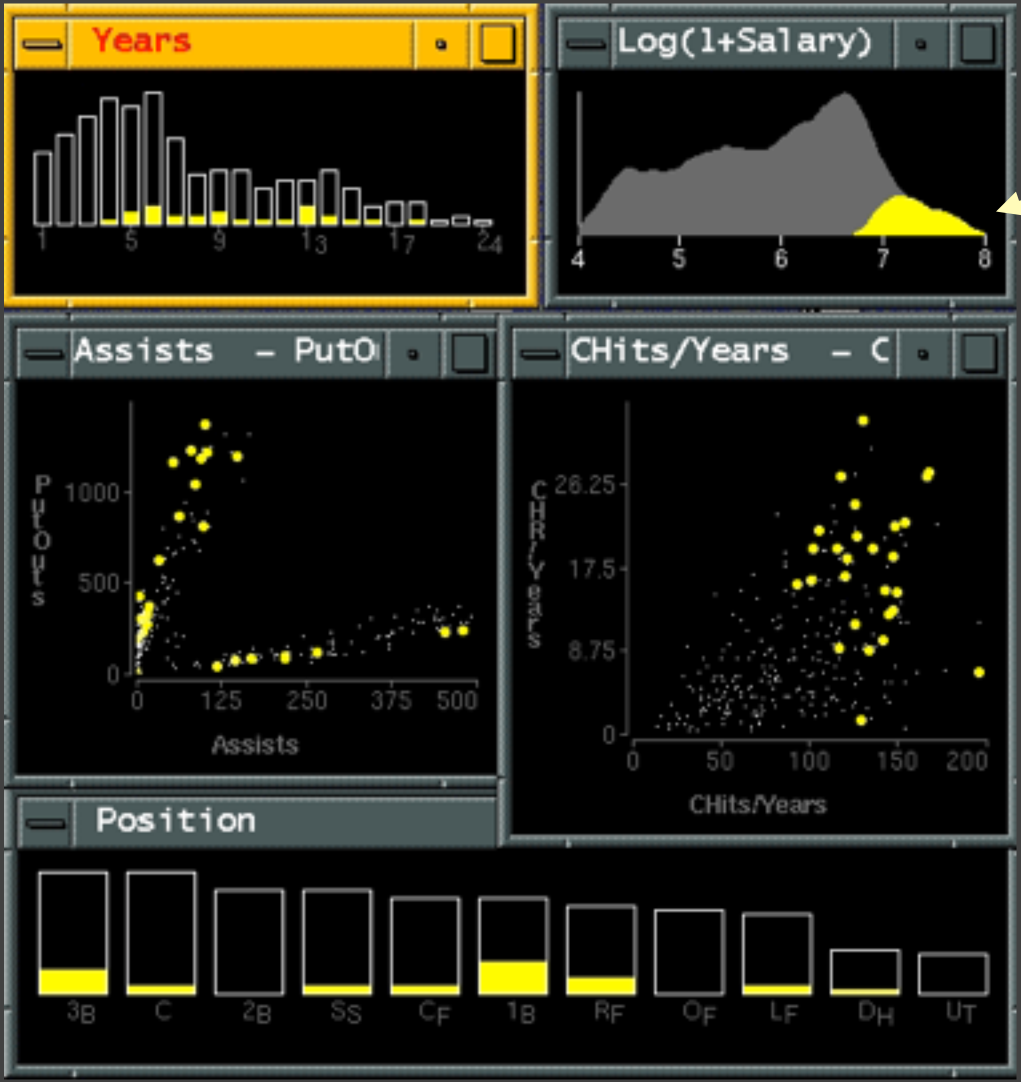
Select (“***brush***”) a subset of data
See selected data in other views

The components must be ***linked***
by *tuple* (matching data points), or
by *query* (matching range or values)

Baseball Statistics [Wills 95]

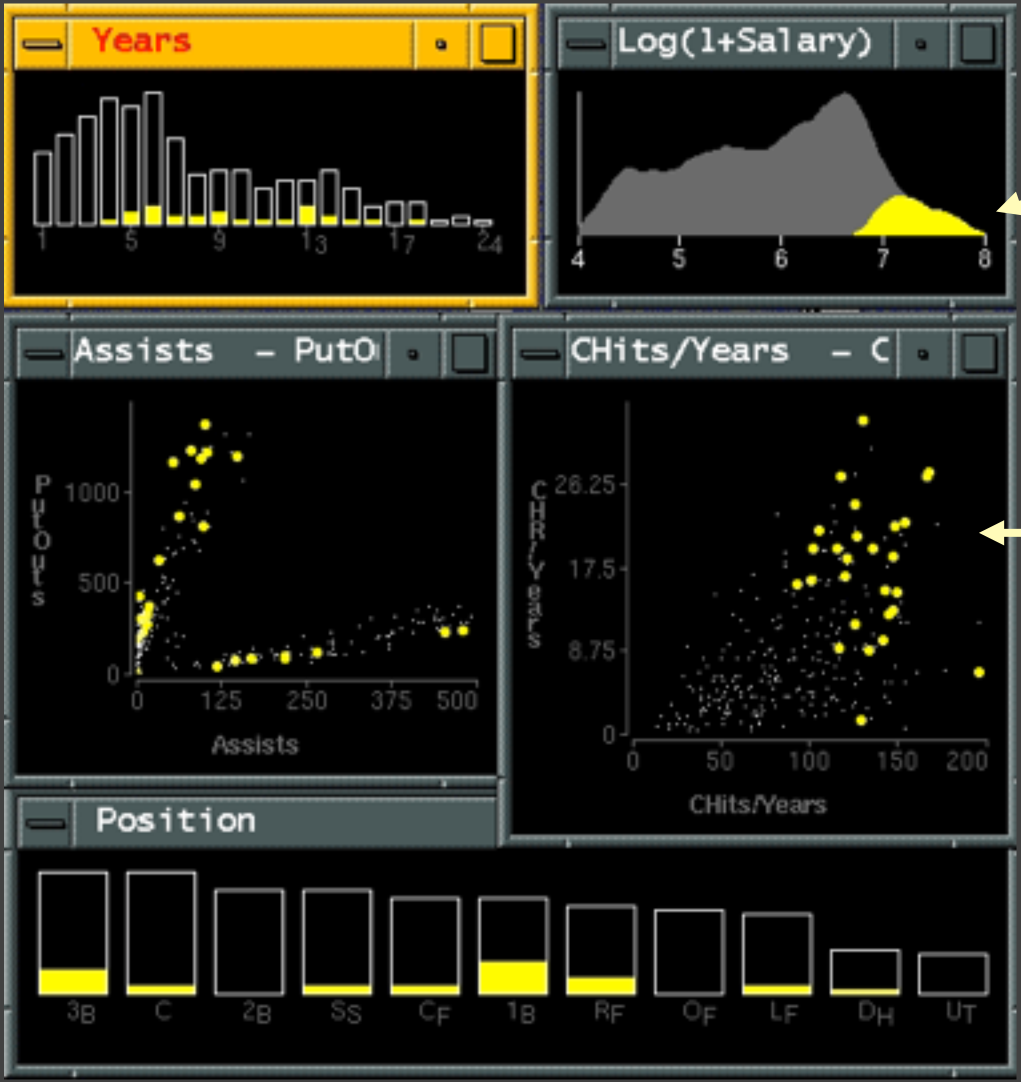


Baseball Statistics [Wills 95]



select high salaries

Baseball Statistics [Wills 95]

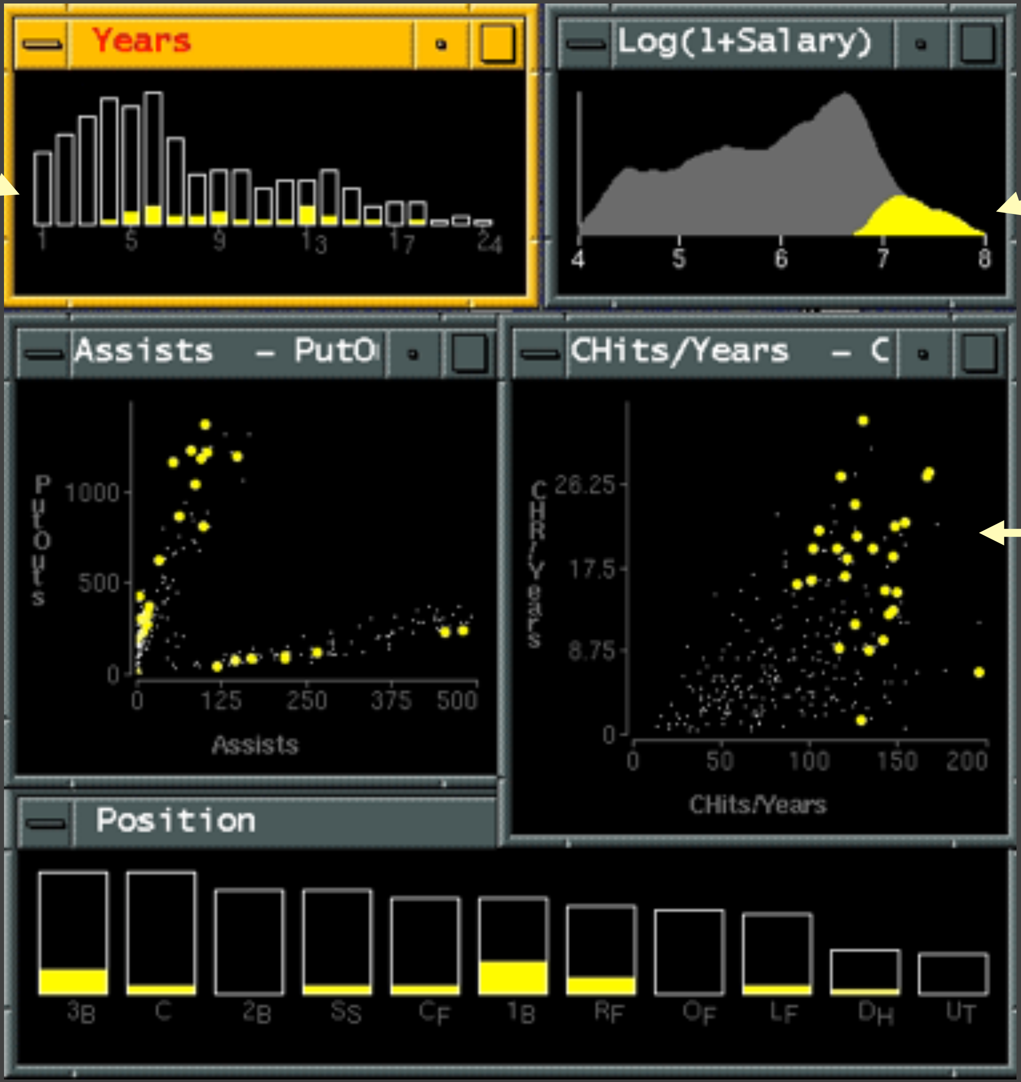


select high salaries

avg career HRs vs avg career hits (batting ability)

Baseball Statistics [Wills 95]

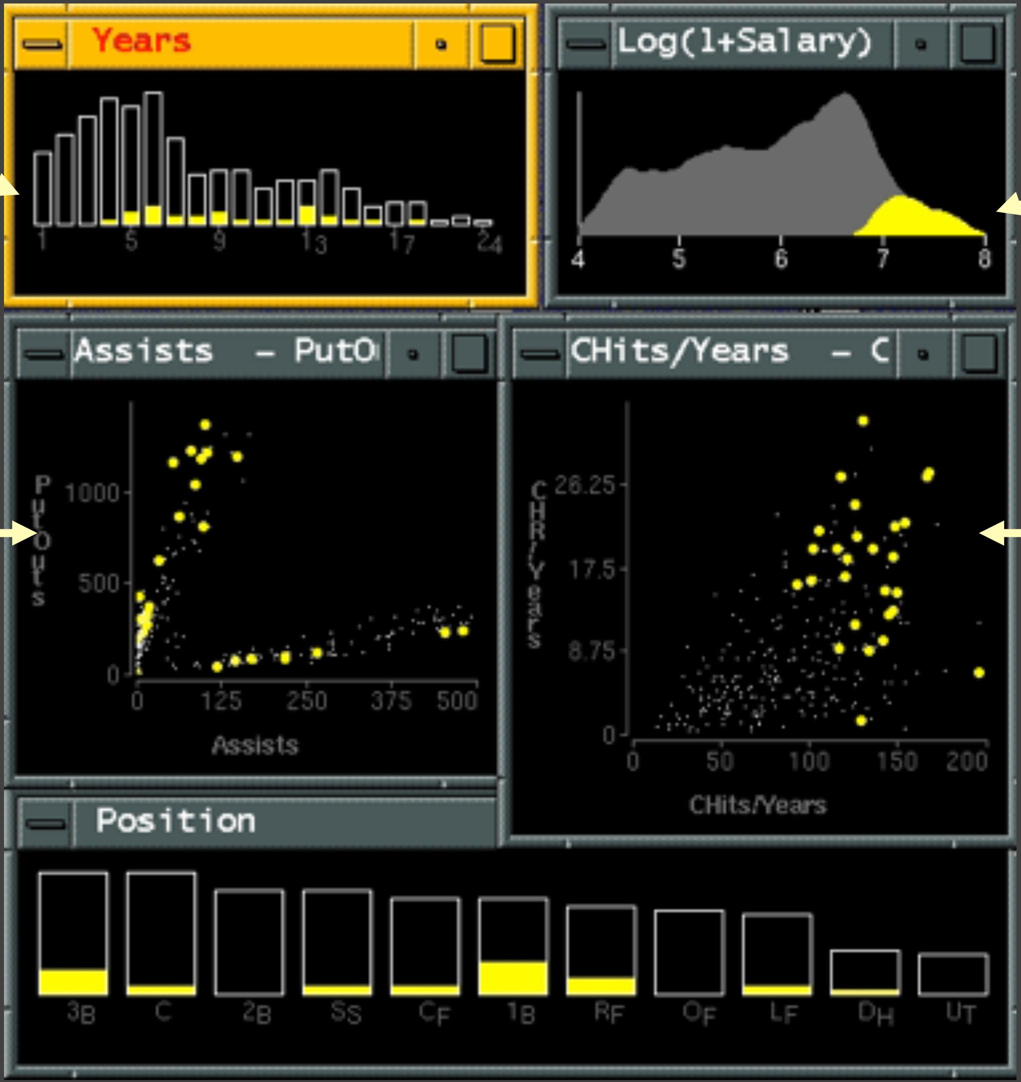
how long in majors



select high salaries

avg career HRs vs avg career hits (batting ability)

Baseball Statistics [Wills 95]



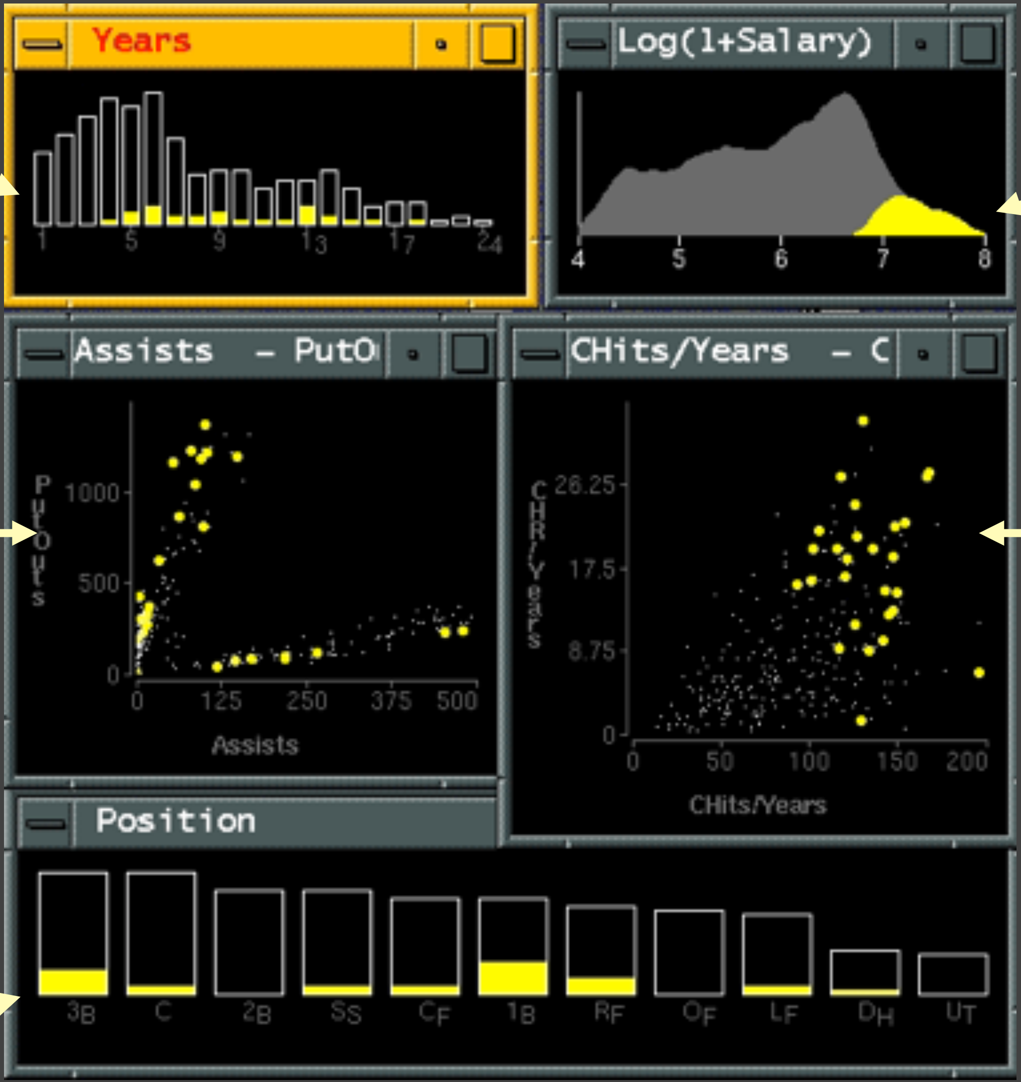
how long in majors

select high salaries

avg assists vs avg putouts (fielding ability)

avg career HRs vs avg career hits (batting ability)

Baseball Statistics [Wills 95]



how long in majors

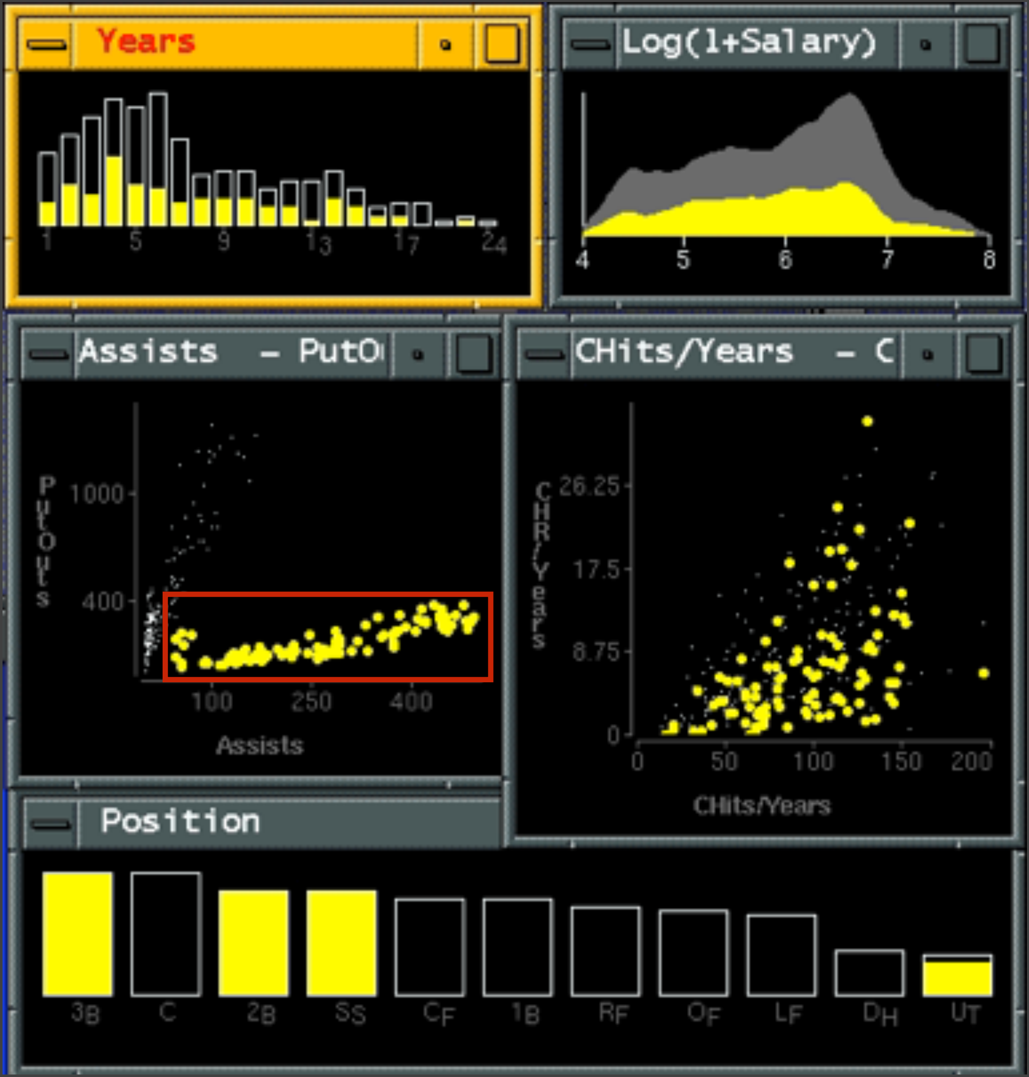
select high salaries

avg assists vs avg putouts (fielding ability)

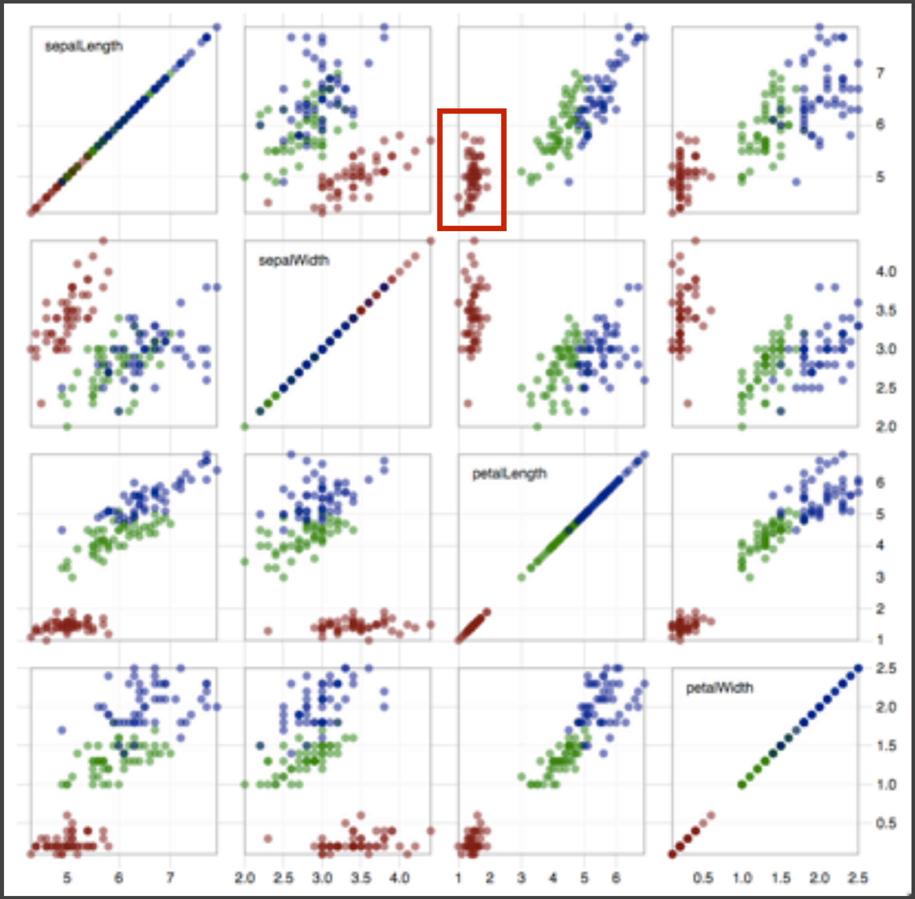
avg career HRs vs avg career hits (batting ability)

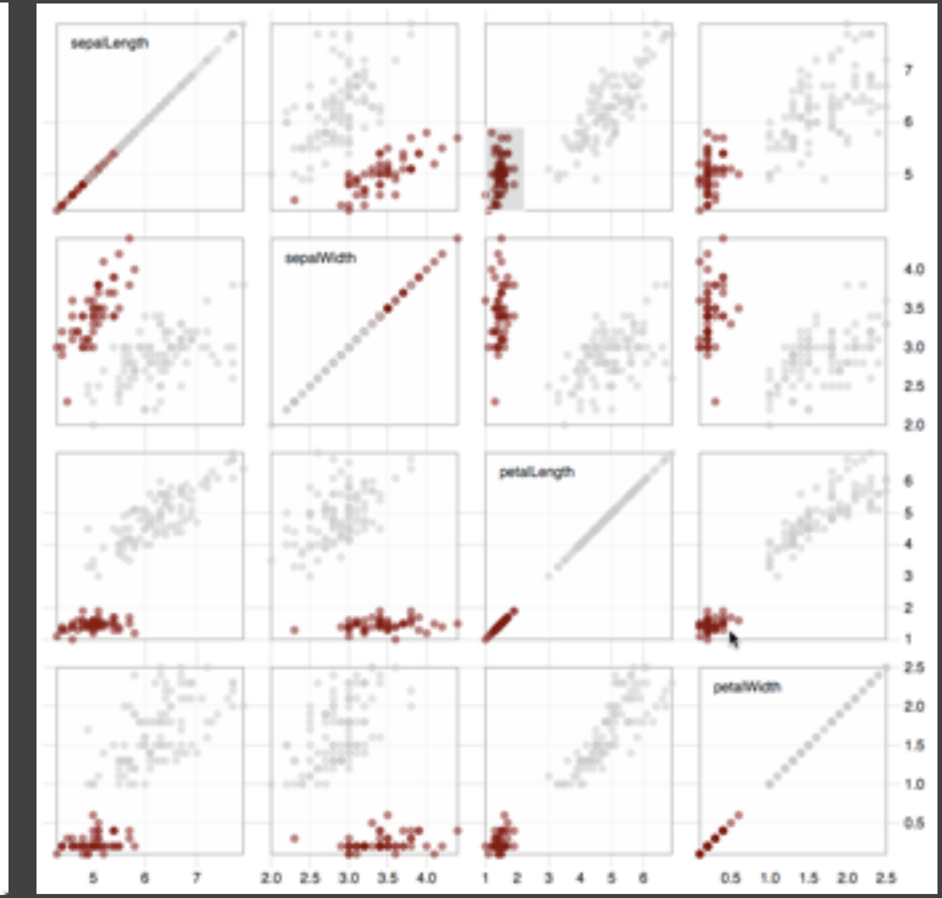
distribution of positions played

Linking Assists to Positions



Brushing Scatterplots





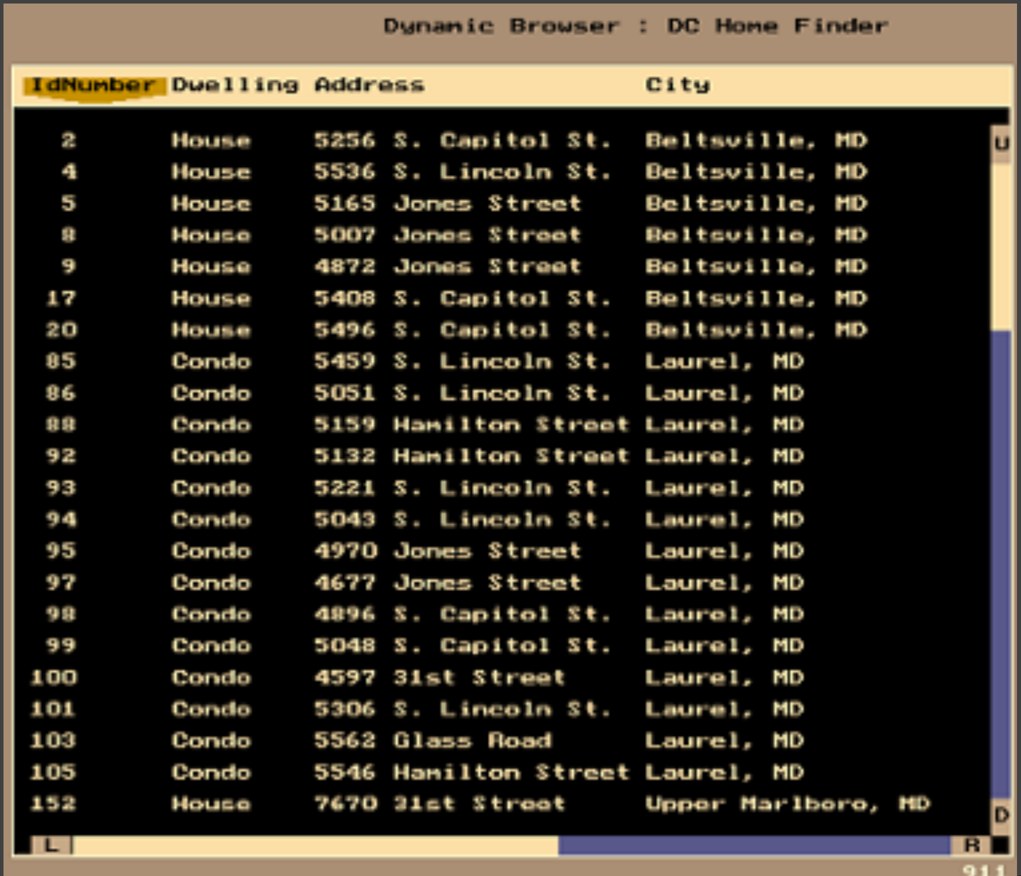
Dynamic Queries

Query & Results

SELECT house FROM seattle_homes

WHERE price < 1,000,000 AND bedrooms > 2

ORDER BY price



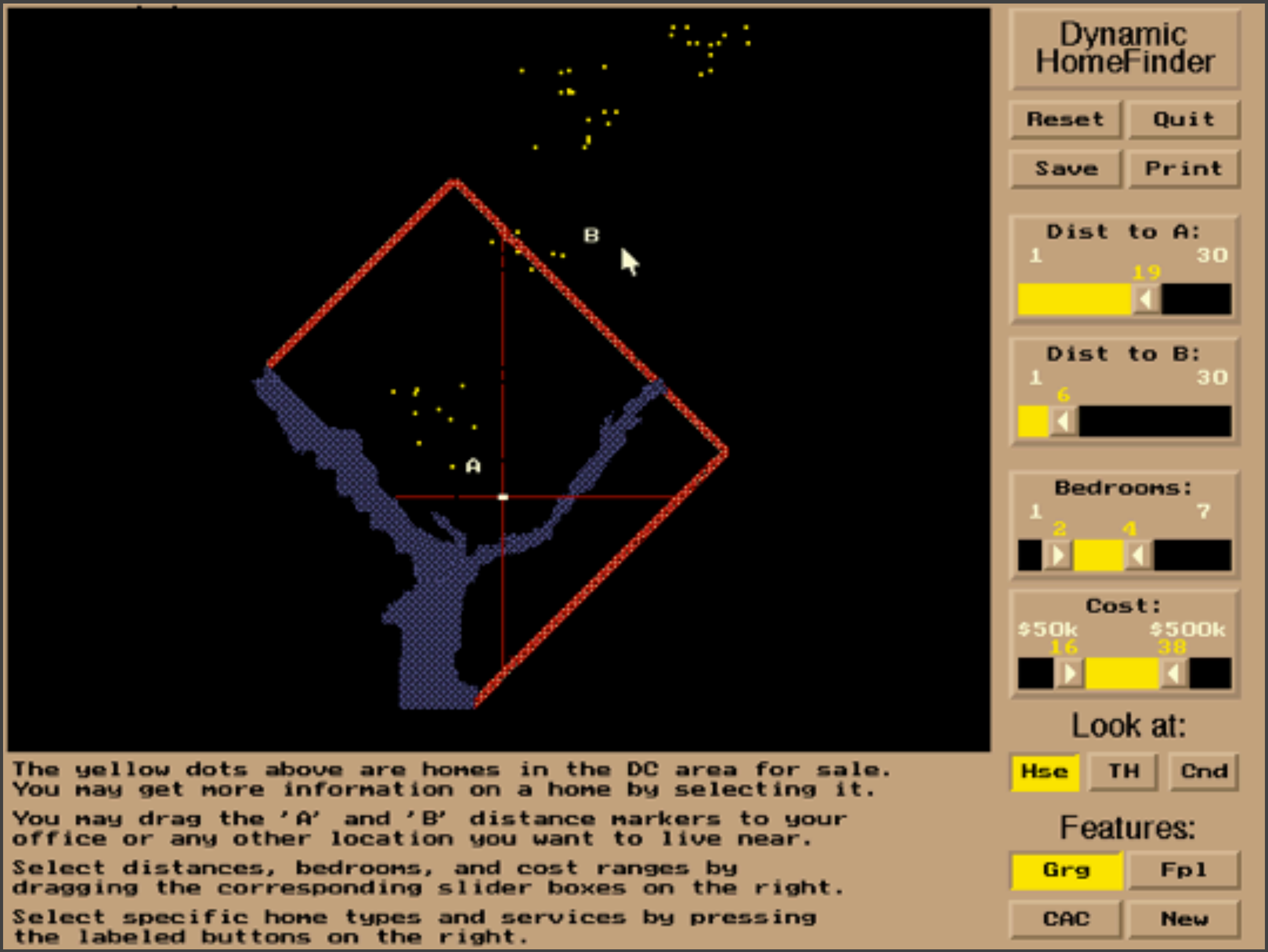
Dynamic Browser : DC Home Finder

IdNumber	Dwelling	Address	City
2	House	5256 S. Capitol St.	Beltsville, MD
4	House	5536 S. Lincoln St.	Beltsville, MD
5	House	5165 Jones Street	Beltsville, MD
8	House	5007 Jones Street	Beltsville, MD
9	House	4872 Jones Street	Beltsville, MD
17	House	5408 S. Capitol St.	Beltsville, MD
20	House	5496 S. Capitol St.	Beltsville, MD
85	Condo	5459 S. Lincoln St.	Laurel, MD
86	Condo	5051 S. Lincoln St.	Laurel, MD
88	Condo	5159 Hamilton Street	Laurel, MD
92	Condo	5132 Hamilton Street	Laurel, MD
93	Condo	5221 S. Lincoln St.	Laurel, MD
94	Condo	5043 S. Lincoln St.	Laurel, MD
95	Condo	4970 Jones Street	Laurel, MD
97	Condo	4677 Jones Street	Laurel, MD
98	Condo	4896 S. Capitol St.	Laurel, MD
99	Condo	5048 S. Capitol St.	Laurel, MD
100	Condo	4597 31st Street	Laurel, MD
101	Condo	5306 S. Lincoln St.	Laurel, MD
103	Condo	5562 Glass Road	Laurel, MD
105	Condo	5546 Hamilton Street	Laurel, MD
152	House	7670 31st Street	Upper Marlboro, MD

Issues with Textual Queries

1. For programmers
2. Rigid syntax
3. Only shows exact matches
4. Too few or too many hits
5. No hint on how to reformulate the query
6. Slow question-answer loop
7. Results returned as table

HomeFinder



Dist to A: 1 30 19

Dist to B: 1 30 6

Bedrooms: 1 7 2 4

Cost: $50k $500k 16 38

The yellow dots above are homes in the DC area for sale. You may get more information on a home by selecting it.

You may drag the 'A' and 'B' distance markers to your office or any other location you want to live near.

Select distances, bedrooms, and cost ranges by dragging the corresponding slider boxes on the right.

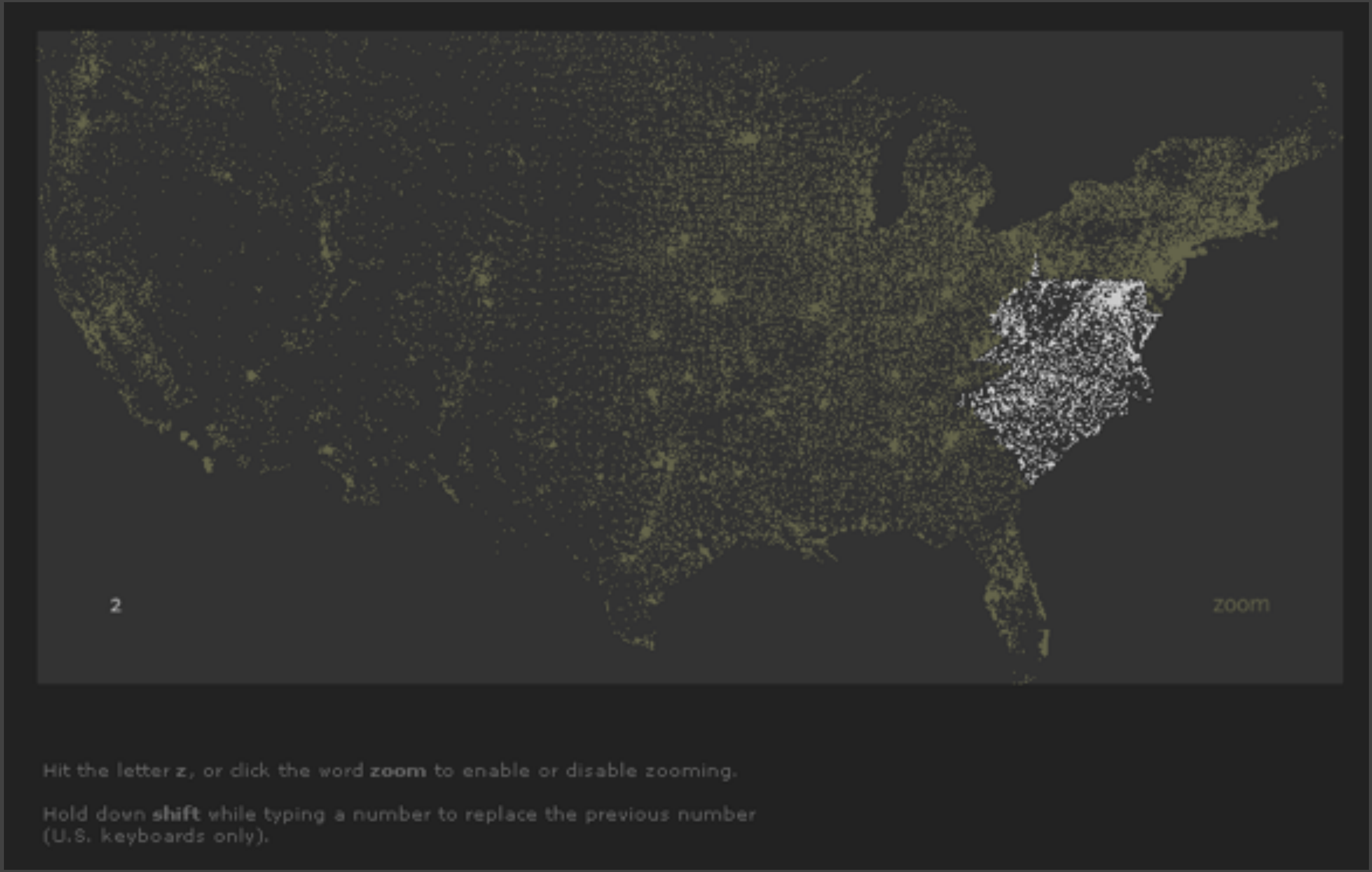
Select specific home types and services by pressing the labeled buttons on the right.

[Williamson and Shneiderman 92]

Direct Manipulation

1. Visual representation of objects and actions
2. Rapid, incremental and reversible actions
3. Selection by pointing (not typing)
4. Immediate and continuous display of results

Zipdecode [Fry 04]

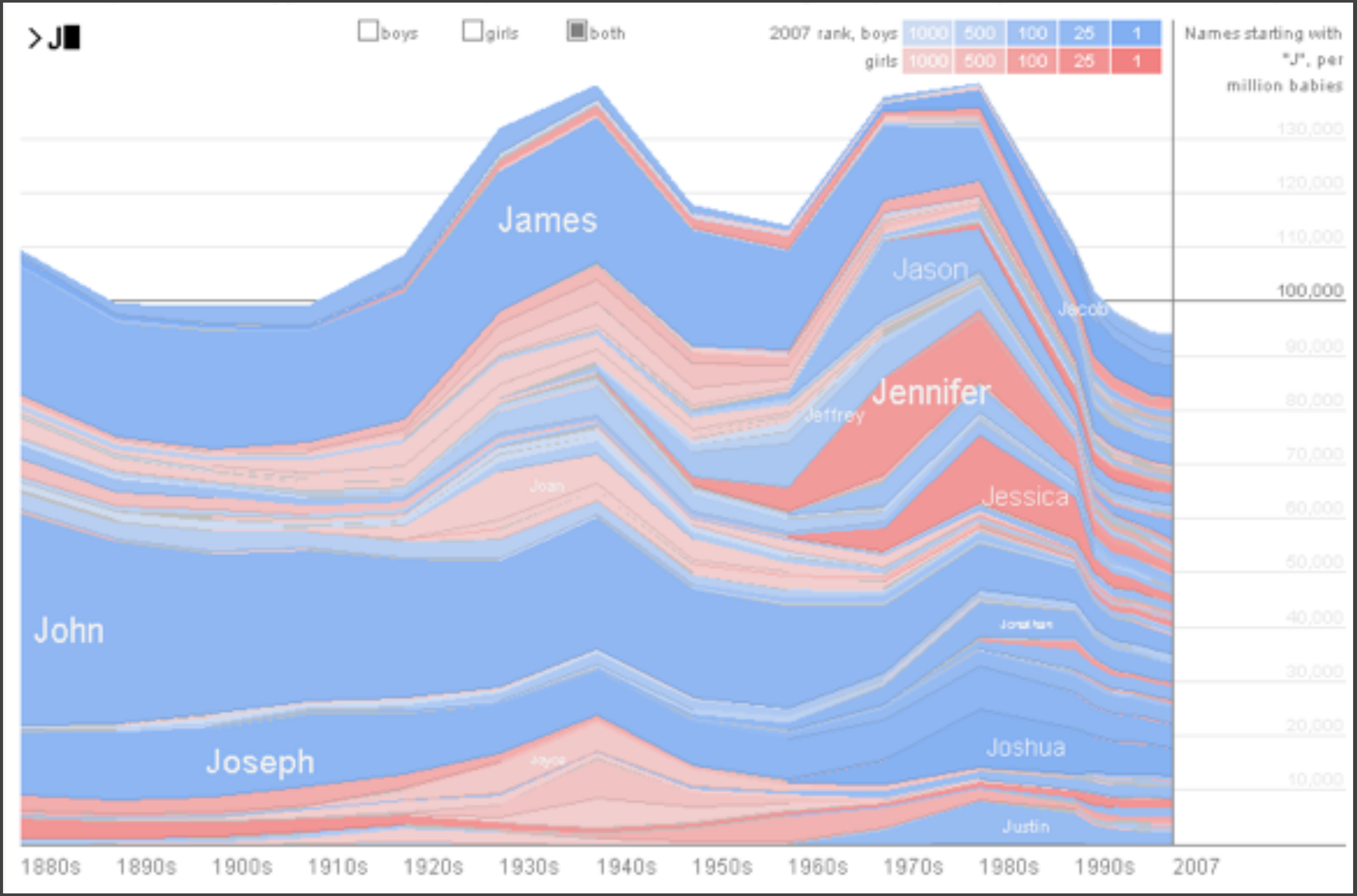


Hit the letter **z**, or click the word **zoom** to enable or disable zooming.

Hold down **shift** while typing a number to replace the previous number (U.S. keyboards only).

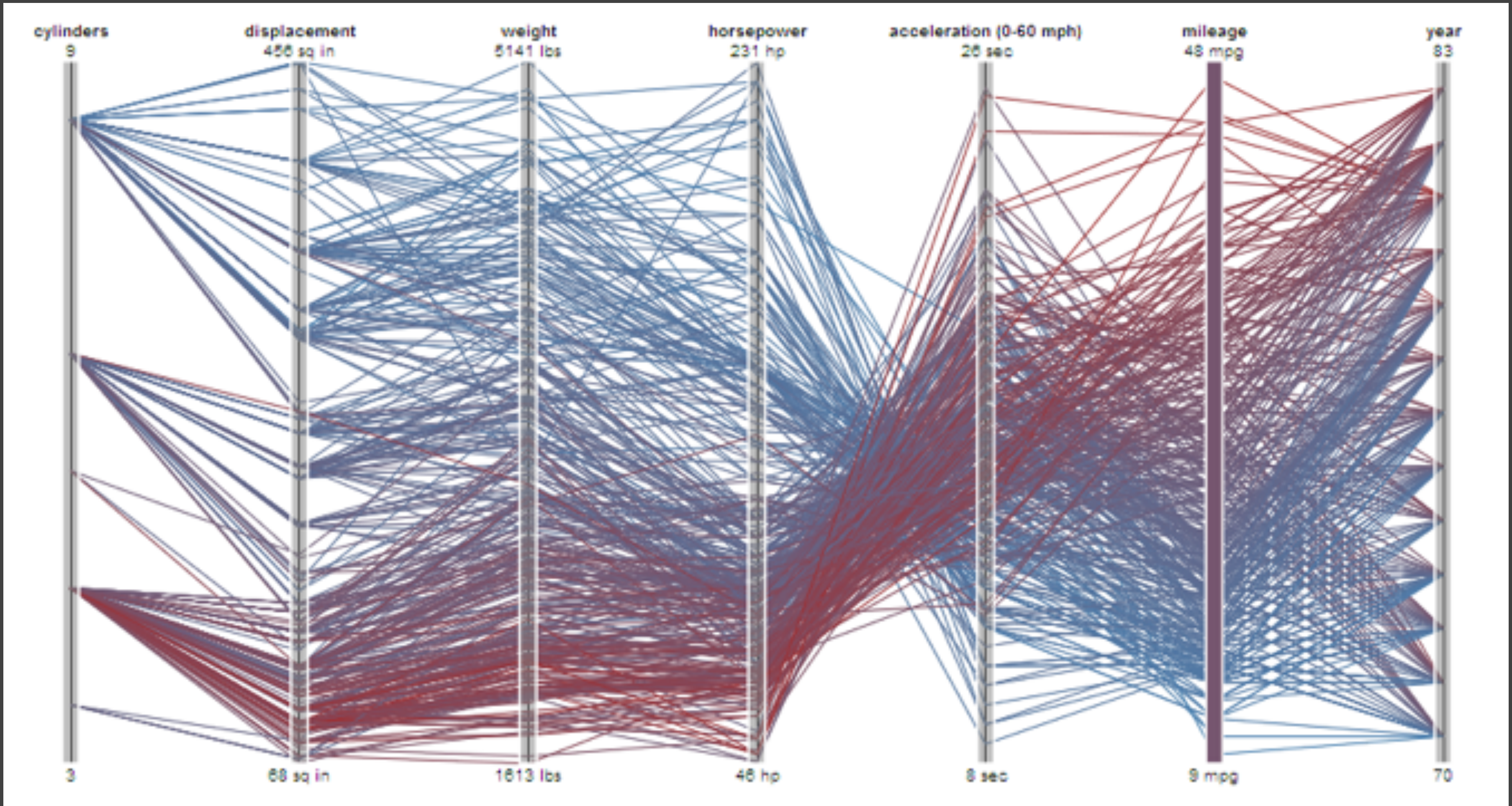
http://benfry.com/zipdecode/

NameVoyager [Wattenberg 06]

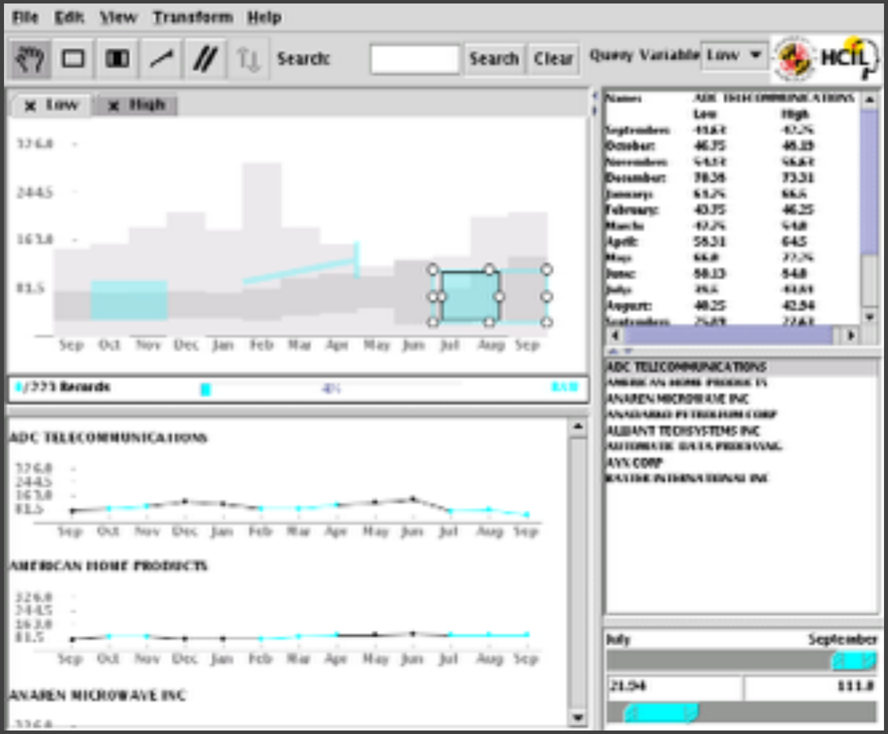


http://www.babynamewizard.com/voyager

Parallel Coordinates [Inselberg]



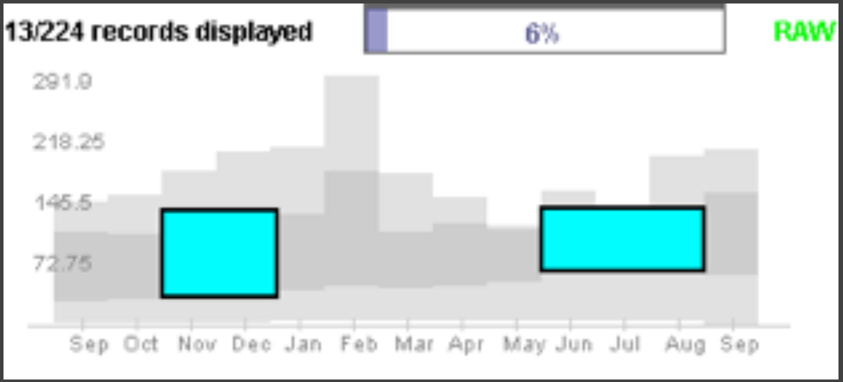
TimeSearcher [Hocheiser 02]

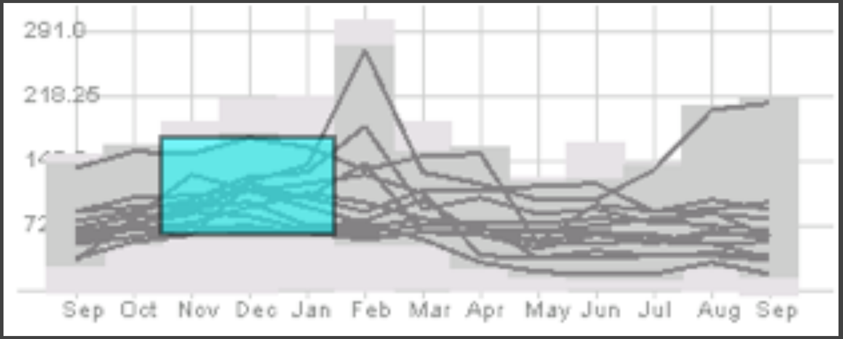


Names:	ADC TELECOMMUNICATIONS	
	Low	High
September:	41.63	47.25
October:	46.75	48.19
November:	54.13	56.63
December:	70.38	73.31
January:	61.25	66.6
February:	43.75	46.25
March:	47.25	54.8
April:	58.31	64.5
May:	66.8	77.25
June:	58.13	54.8
July:	39.6	44.44
August:	40.25	42.94
September:	25.69	27.63

ADC TELECOMMUNICATIONS
AMERICAN HOME PRODUCTS
ANAREN MICROWAVE INC
ANADARKO PETROLEUM CORP
ALLIANT TECHSYSTEMS INC
AUTOMATIC DATA PROCESSING
AVX CORP
BAXTER INTERNATIONAL INC

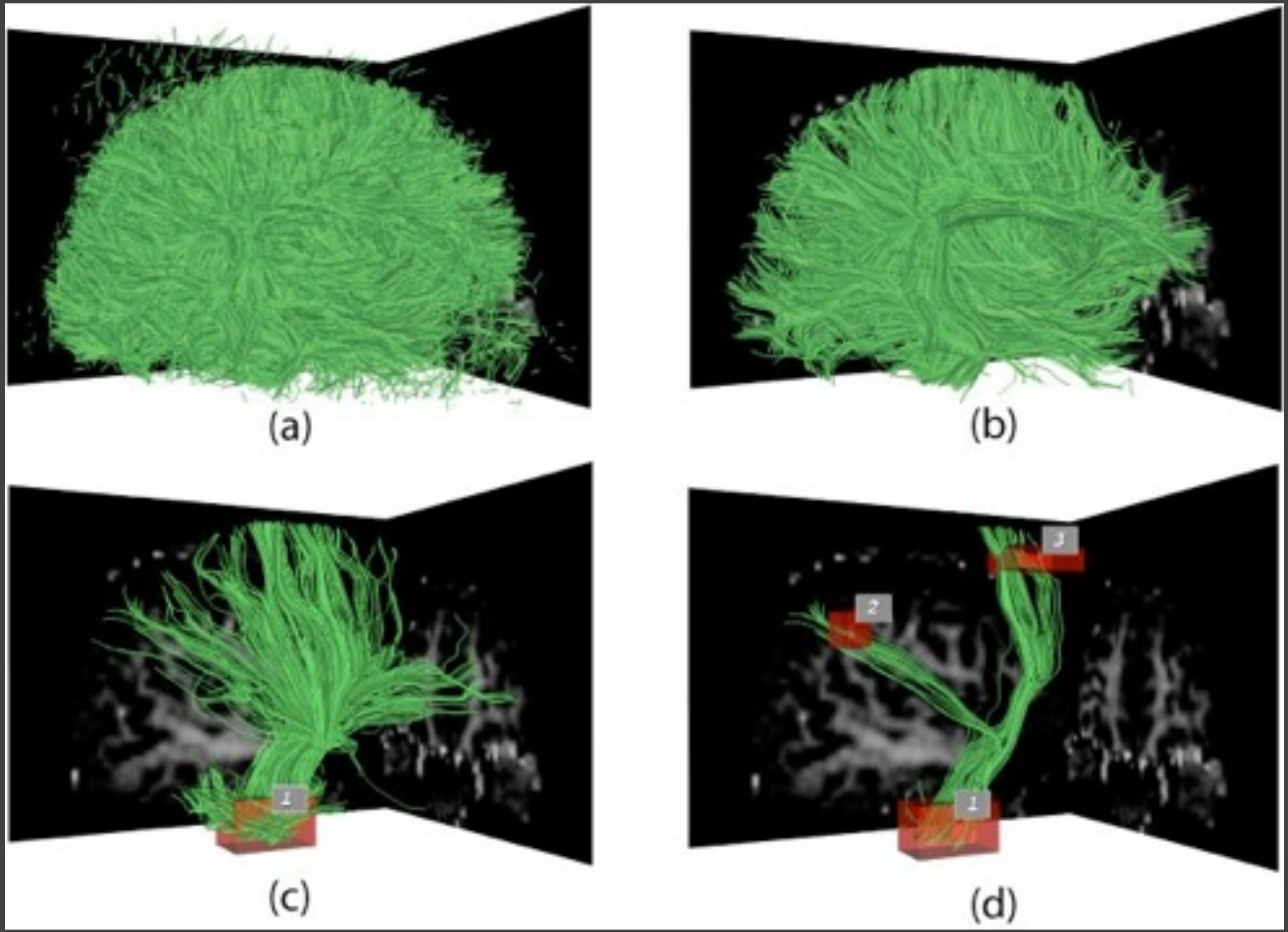
21.94 | 111.8



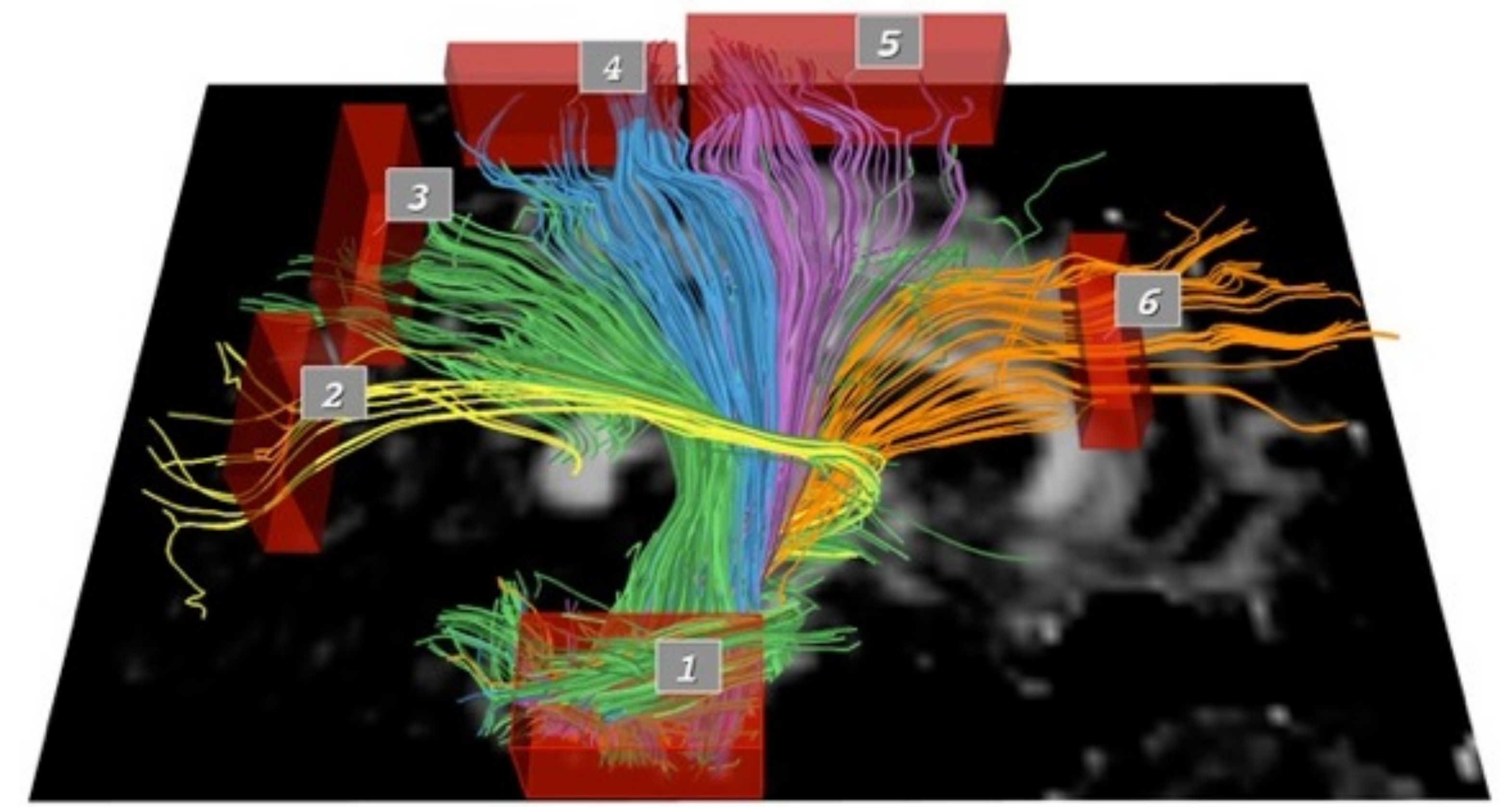


Builds on Wattenberg's [2001] idea for sketch-based queries of time-series data.

3D Dynamic Queries [Akers 04]



3D Dynamic Queries [Akers 04]



Pros & Cons

Pros

Controls useful for both novices and experts

Quick way to explore data

Pros & Cons

Pros

Controls useful for both novices and experts

Quick way to explore data

Cons

Simple queries

Lots of controls

Amount of data shown limited by screen space

Who would use these kinds of tools?

Examples

Analysis Example: MTurk Participation

Data Set: Turker Participation

Turker ID	String (N)
Avg. Completion Rate	Number [0,1] (Q)

Collected in 2009 by Heer & Bostock.

What questions might we ask of the data?
What charts might provide insight?

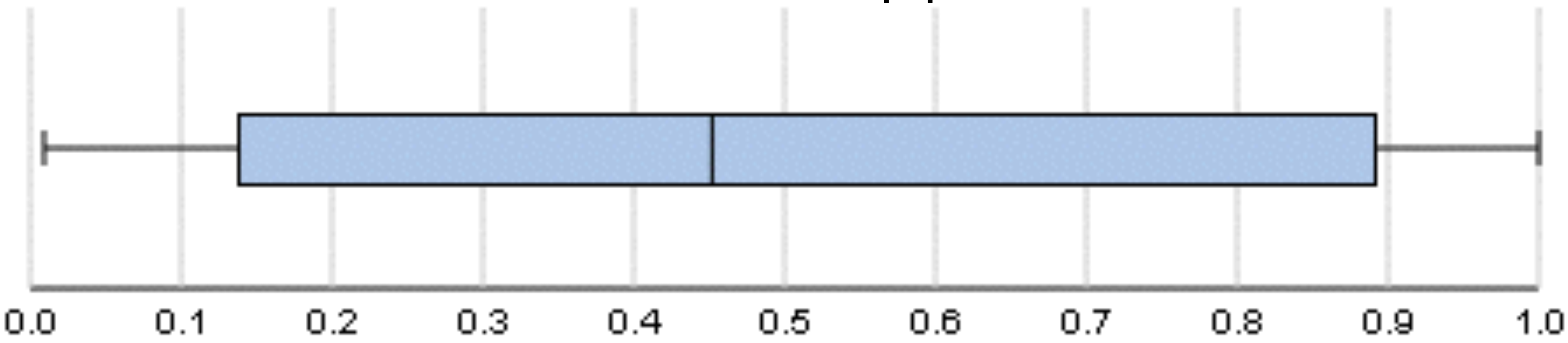
Min

Median

Max

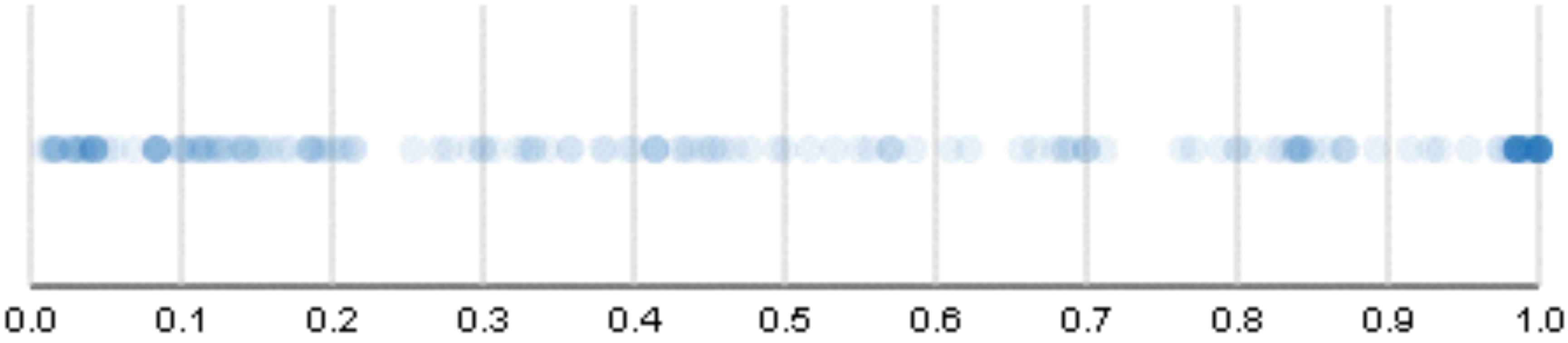
Lower Quartile

Upper Quartile



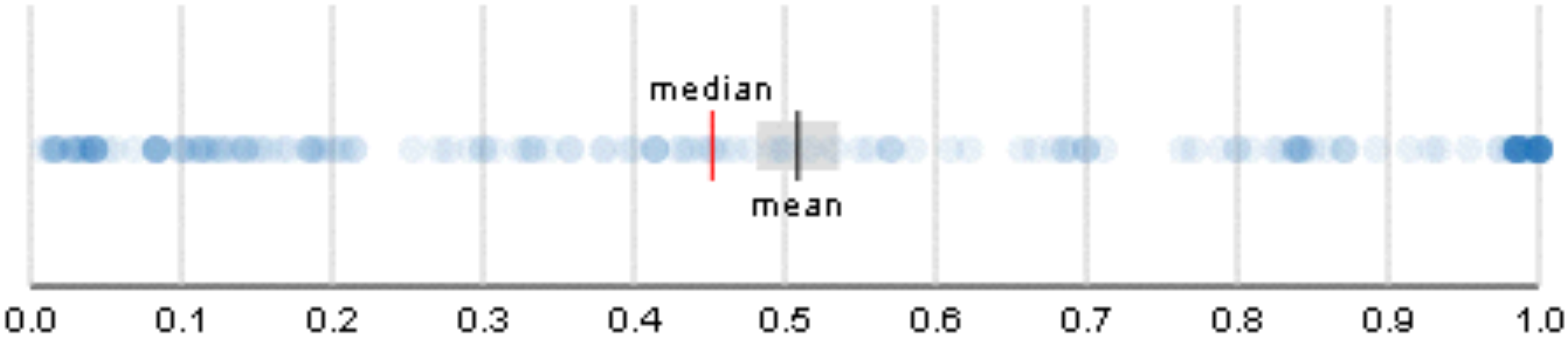
Turker Completion Percentage

Box Plot



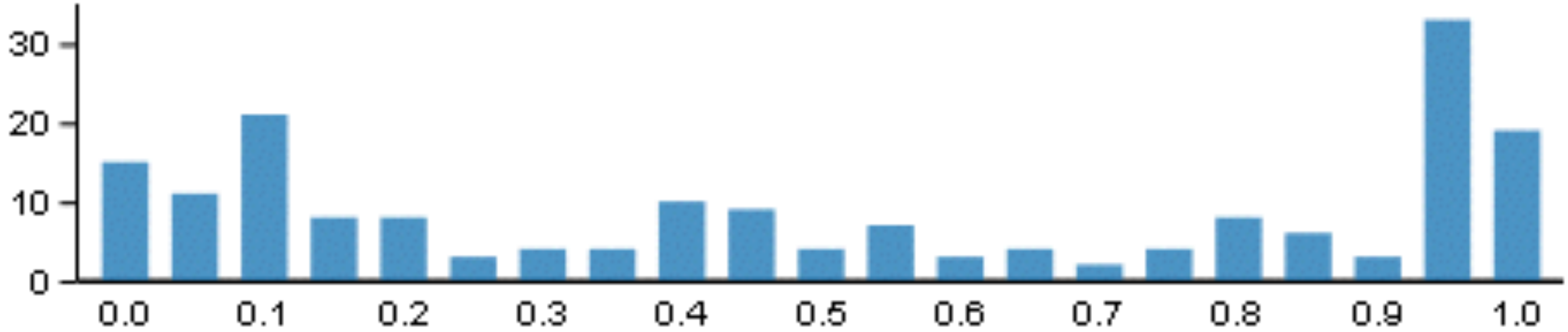
Turker Completion Percentage

Dot Plot (with transparency for overlap)



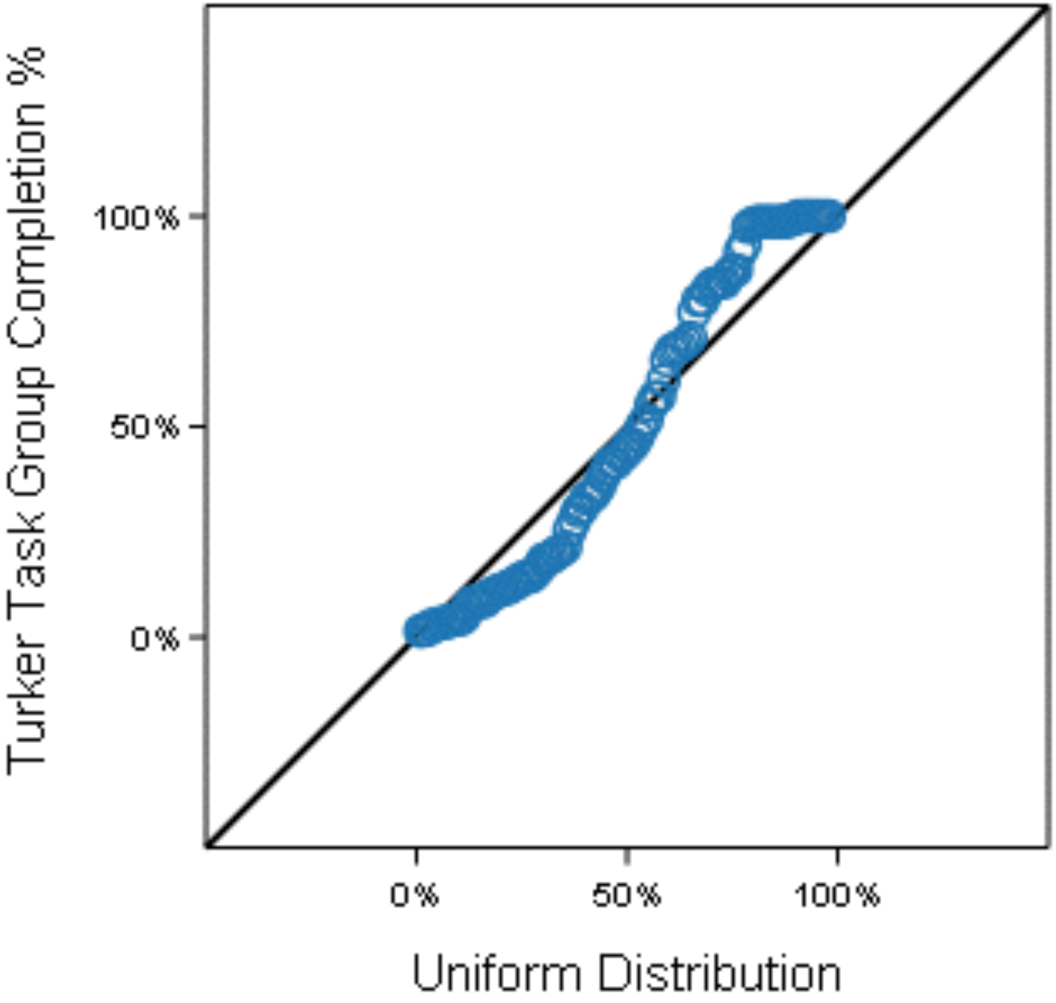
Turker Completion Percentage

Dot Plot (with Reference Lines)



Turker Completion Percentage

Histogram (binned counts)

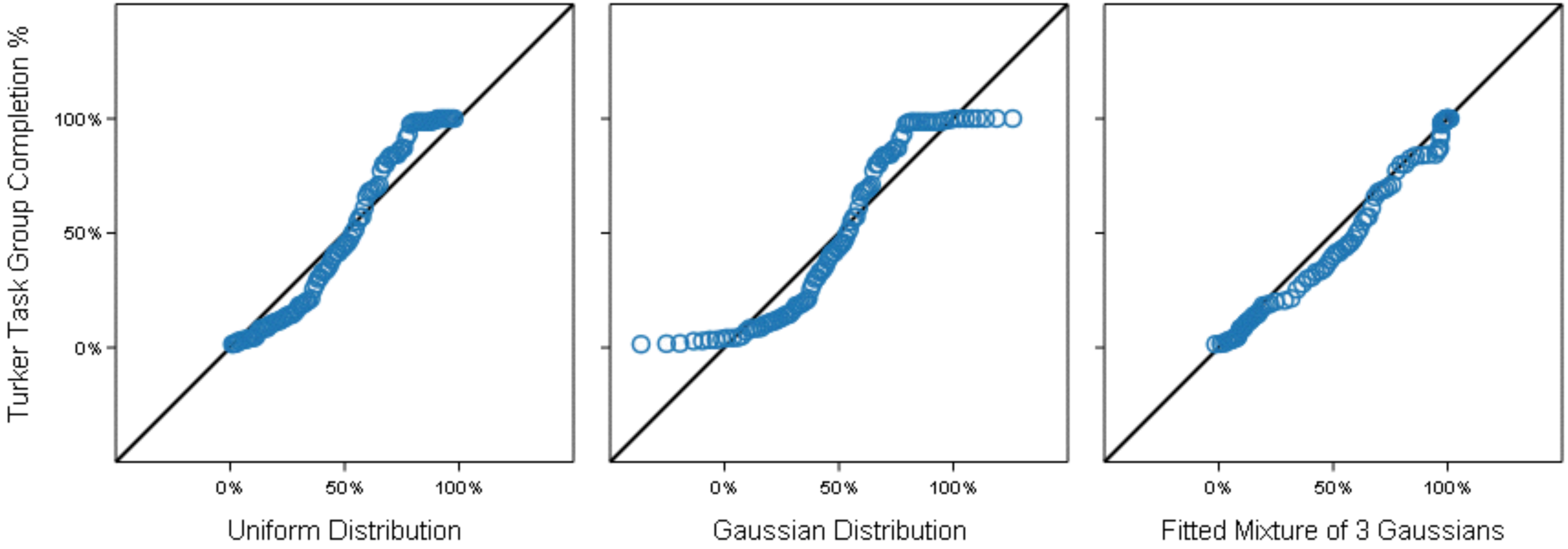


Quantile-Quantile Plot

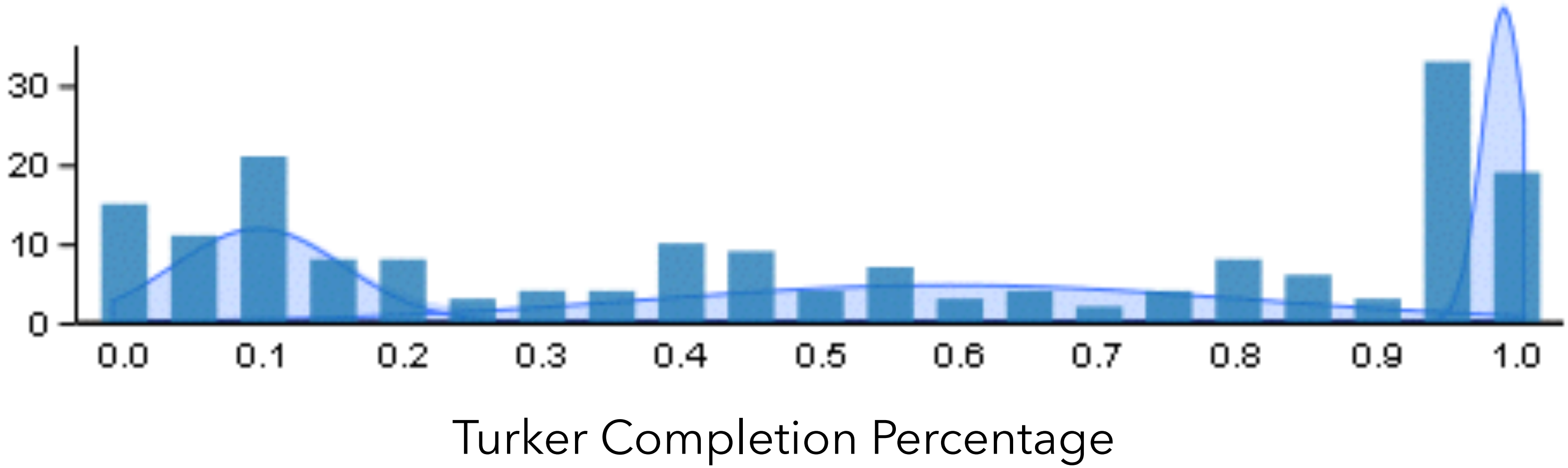
Used to compare two distributions; in this case, one actual and one theoretical.

Plots the quantiles (here, the percentile values) against each other.

Similar distributions lie along the diagonal. If linearly related, values will lie along a line, but with potentially varying slope and intercept.



Quantile-Quantile Plots



Histogram (+ Fitted Mixture of 3 Gaussians)

Data Set: Turker Participation

Even for “simple” data, a variety of graphics might provide insight. Tailor the choice of graphic to the questions being asked, but be open to surprises.

Graphics can be used to understand and help assess the quality of statistical models.

Premature commitment to a model and lack of verification can lead an analysis astray.

Analysis Example: Antibiotic Effectiveness

Data Set: Antibiotic Effectiveness

Genus of Bacteria	String (N)
Species of Bacteria	String (N)
Antibiotic Applied	String (N)
Gram-Staining?	Pos / Neg (N)
Min. Inhibitory Concent. (g)	Number (Q)

Collected prior to 1951.

What questions might we ask?

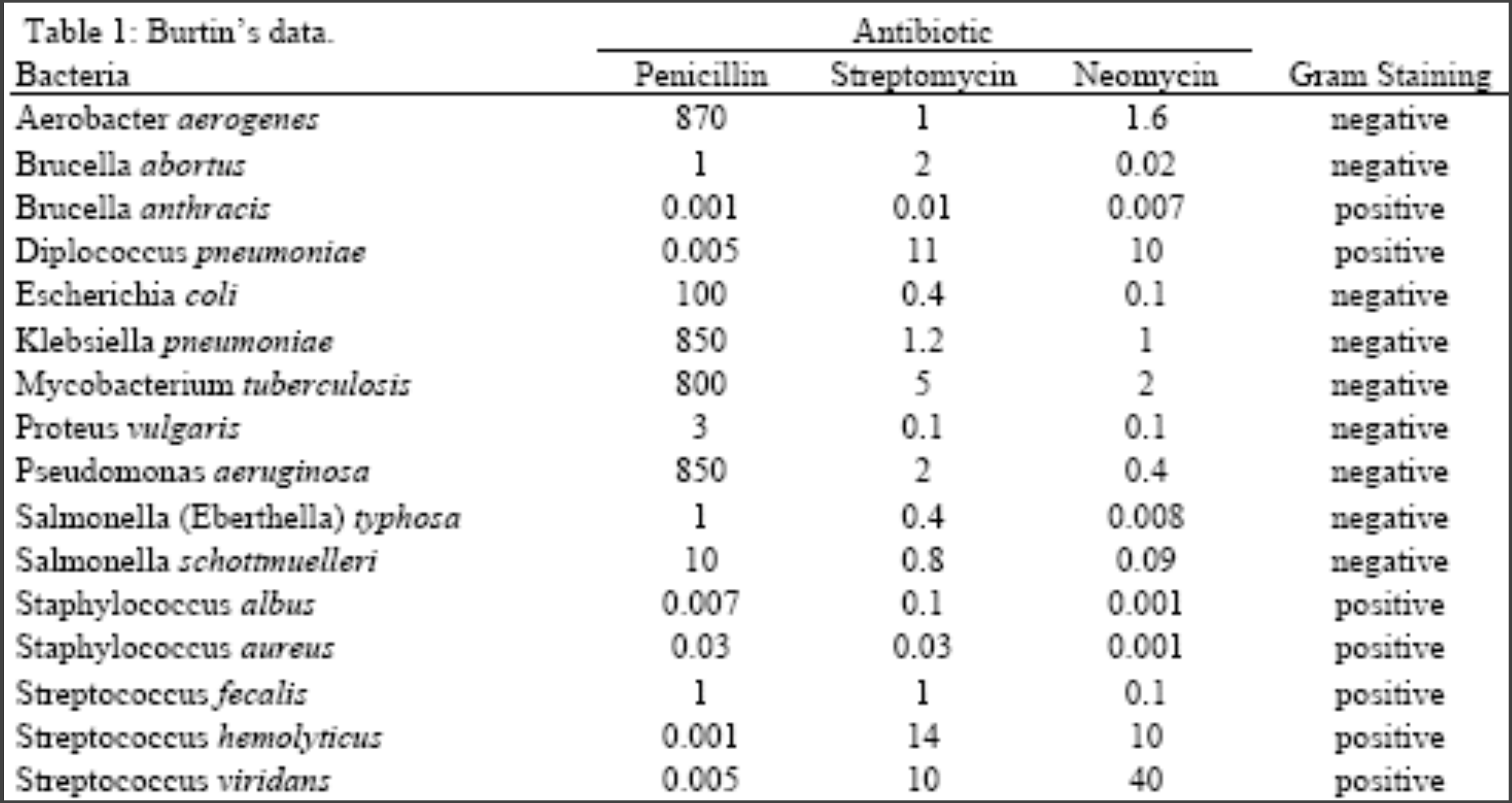
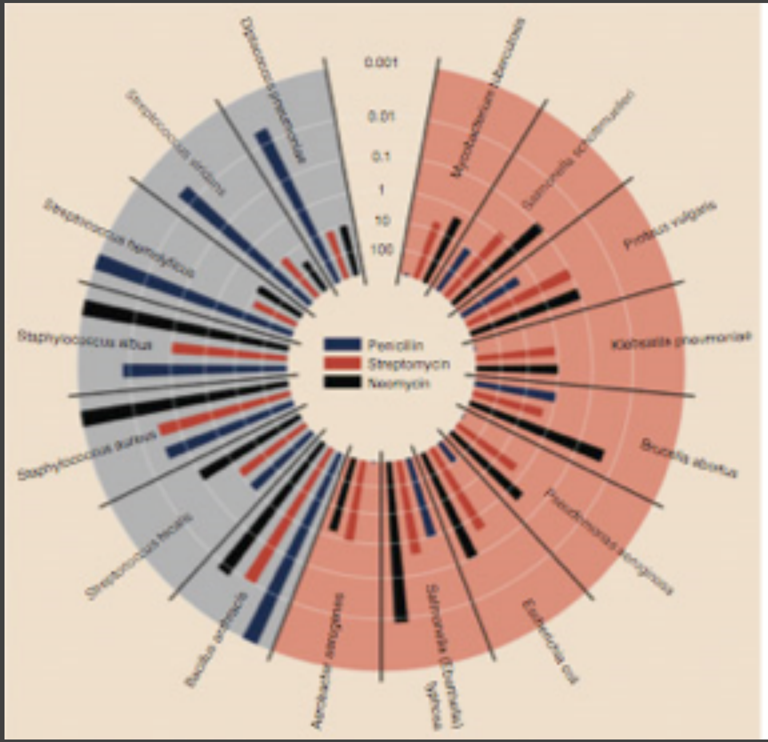
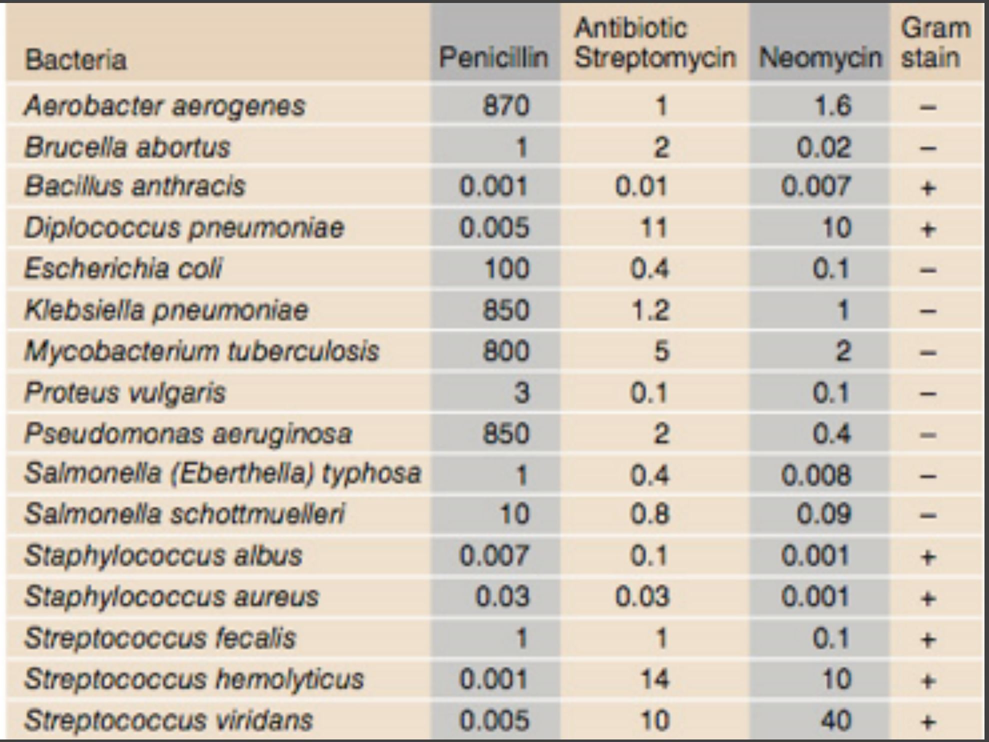


Table 1: Burtin's data.

Bacteria	Antibiotic			Gram Staining
	Penicillin	Streptomycin	Neomycin	
Aerobacter *aerogenes*	870	1	1.6	negative
Brucella *abortus*	1	2	0.02	negative
Brucella *anthracis*	0.001	0.01	0.007	positive
Diplococcus *pneumoniae*	0.005	11	10	positive
Escherichia *coli*	100	0.4	0.1	negative
Klebsiella *pneumoniae*	850	1.2	1	negative
Mycobacterium *tuberculosis*	800	5	2	negative
Proteus *vulgaris*	3	0.1	0.1	negative
Pseudomonas *aeruginosa*	850	2	0.4	negative
Salmonella (Eberthella) *typhosa*	1	0.4	0.008	negative
Salmonella *schottmuelleri*	10	0.8	0.09	negative
Staphylococcus *albus*	0.007	0.1	0.001	positive
Staphylococcus *aureus*	0.03	0.03	0.001	positive
Streptococcus *fecalis*	1	1	0.1	positive
Streptococcus *hemolyticus*	0.001	14	10	positive
Streptococcus *viridans*	0.005	10	40	positive

How do the drugs compare?

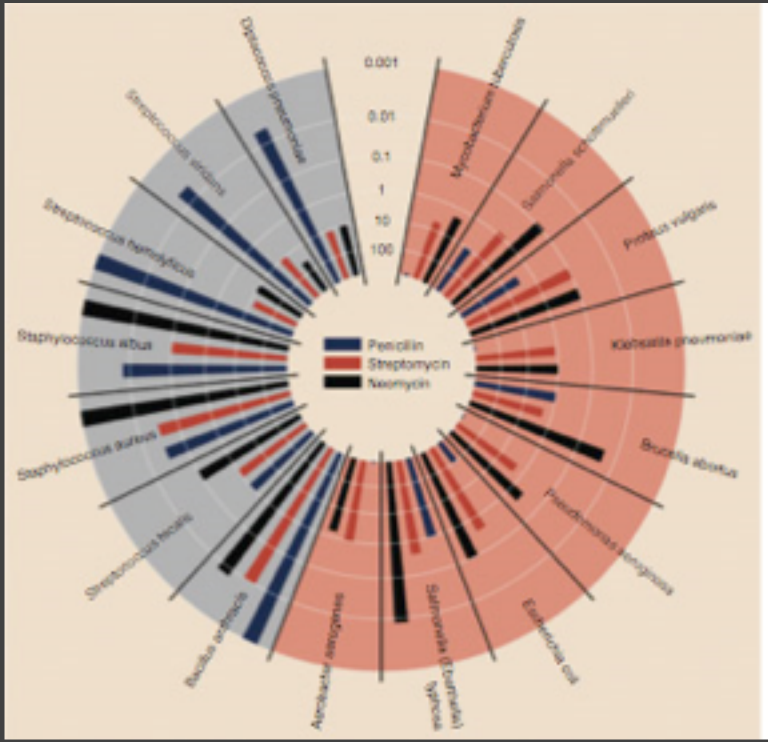


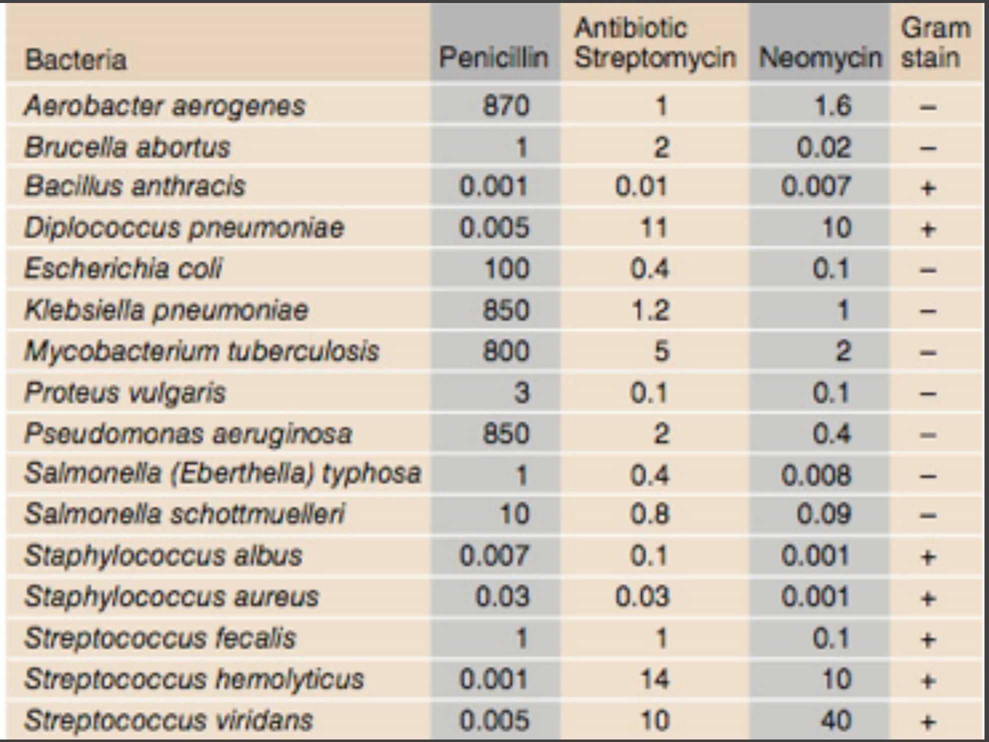


		Antibiotic		Gram
Bacteria	Penicillin	Streptomycin	Neomycin	stain
Aerobacter aerogenes	870	1	1.6	–
Brucella abortus	1	2	0.02	–
Bacillus anthracis	0.001	0.01	0.007	+
Diplococcus pneumoniae	0.005	11	10	+
Escherichia coli	100	0.4	0.1	–
Klebsiella pneumoniae	850	1.2	1	–
Mycobacterium tuberculosis	800	5	2	–
Proteus vulgaris	3	0.1	0.1	–
Pseudomonas aeruginosa	850	2	0.4	–
Salmonella (Eberthella) typhosa	1	0.4	0.008	–
Salmonella schottmuelleri	10	0.8	0.09	–
Staphylococcus albus	0.007	0.1	0.001	+
Staphylococcus aureus	0.03	0.03	0.001	+
Streptococcus fecalis	1	1	0.1	+
Streptococcus hemolyticus	0.001	14	10	+
Streptococcus viridans	0.005	10	40	+

Original graphic by Will Burtin, 1951

How do the drugs compare?





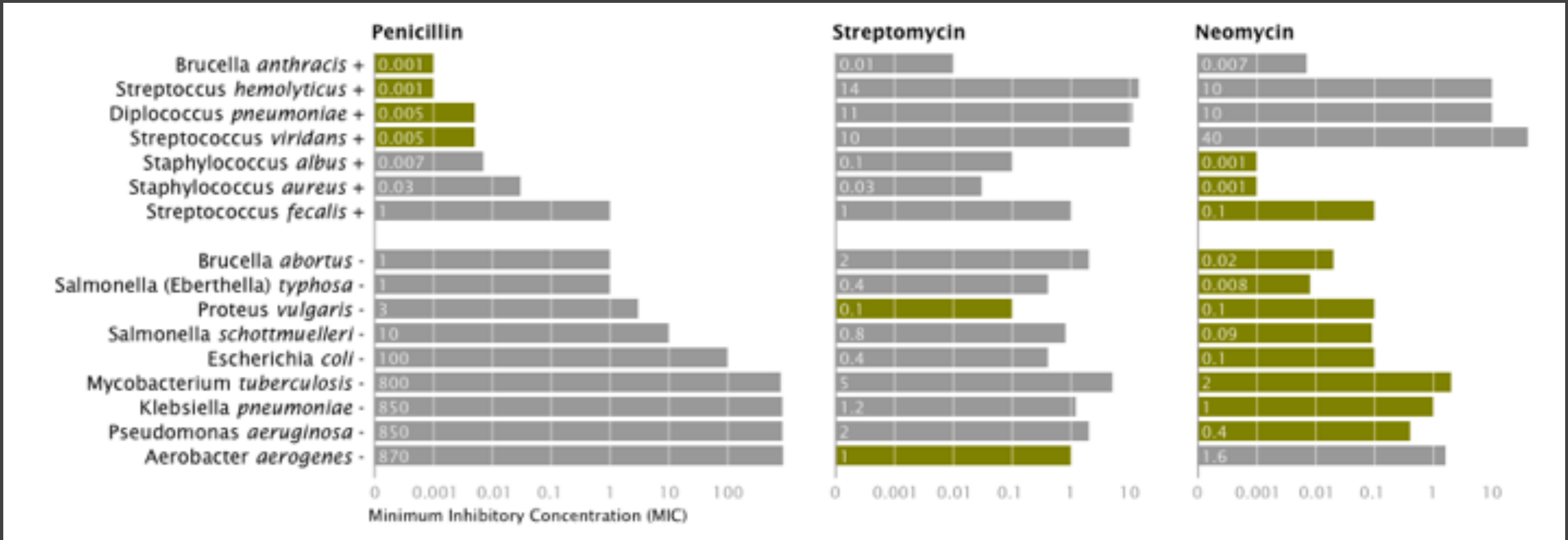
Bacteria	Penicillin	Antibiotic Streptomycin	Neomycin	Gram stain
Aerobacter aerogenes	870	1	1.6	–
Brucella abortus	1	2	0.02	–
Bacillus anthracis	0.001	0.01	0.007	+
Diplococcus pneumoniae	0.005	11	10	+
Escherichia coli	100	0.4	0.1	–
Klebsiella pneumoniae	850	1.2	1	–
Mycobacterium tuberculosis	800	5	2	–
Proteus vulgaris	3	0.1	0.1	–
Pseudomonas aeruginosa	850	2	0.4	–
Salmonella (Eberthella) typhosa	1	0.4	0.008	–
Salmonella schottmuelleri	10	0.8	0.09	–
Staphylococcus albus	0.007	0.1	0.001	+
Staphylococcus aureus	0.03	0.03	0.001	+
Streptococcus fecalis	1	1	0.1	+
Streptococcus hemolyticus	0.001	14	10	+
Streptococcus viridans	0.005	10	40	+

Radius: 1 / log(MIC)

Bar Color: Antibiotic

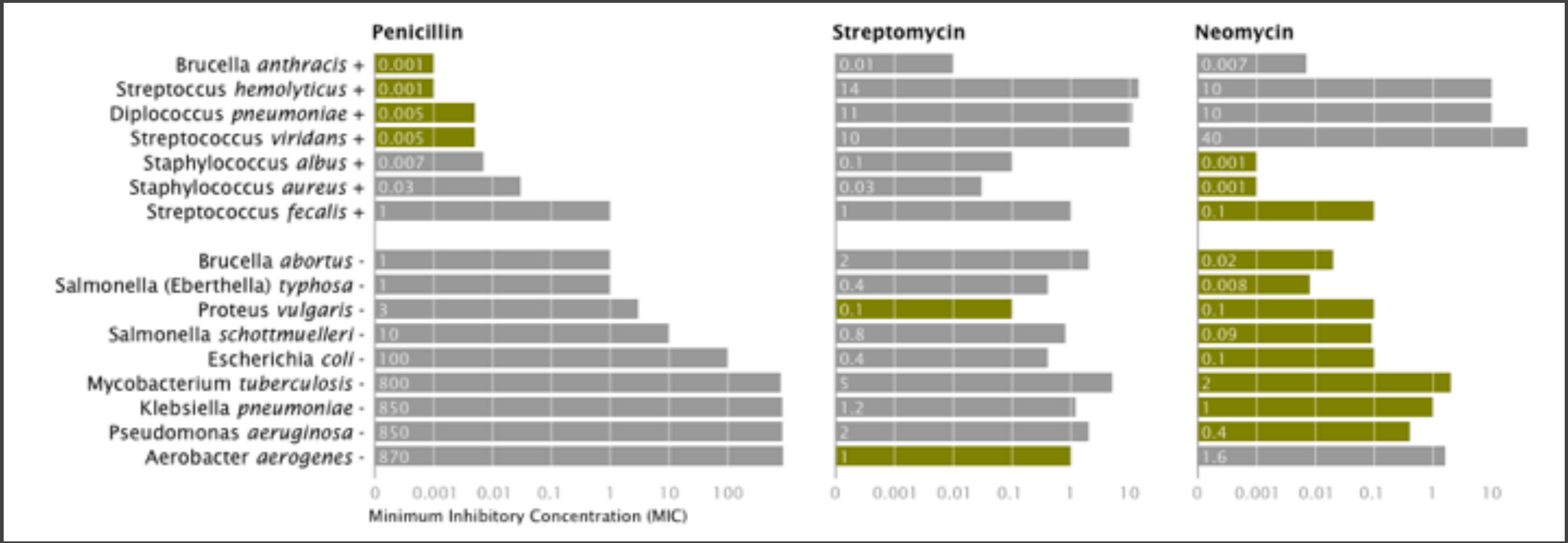
Background Color: Gram Staining

How do the drugs compare?

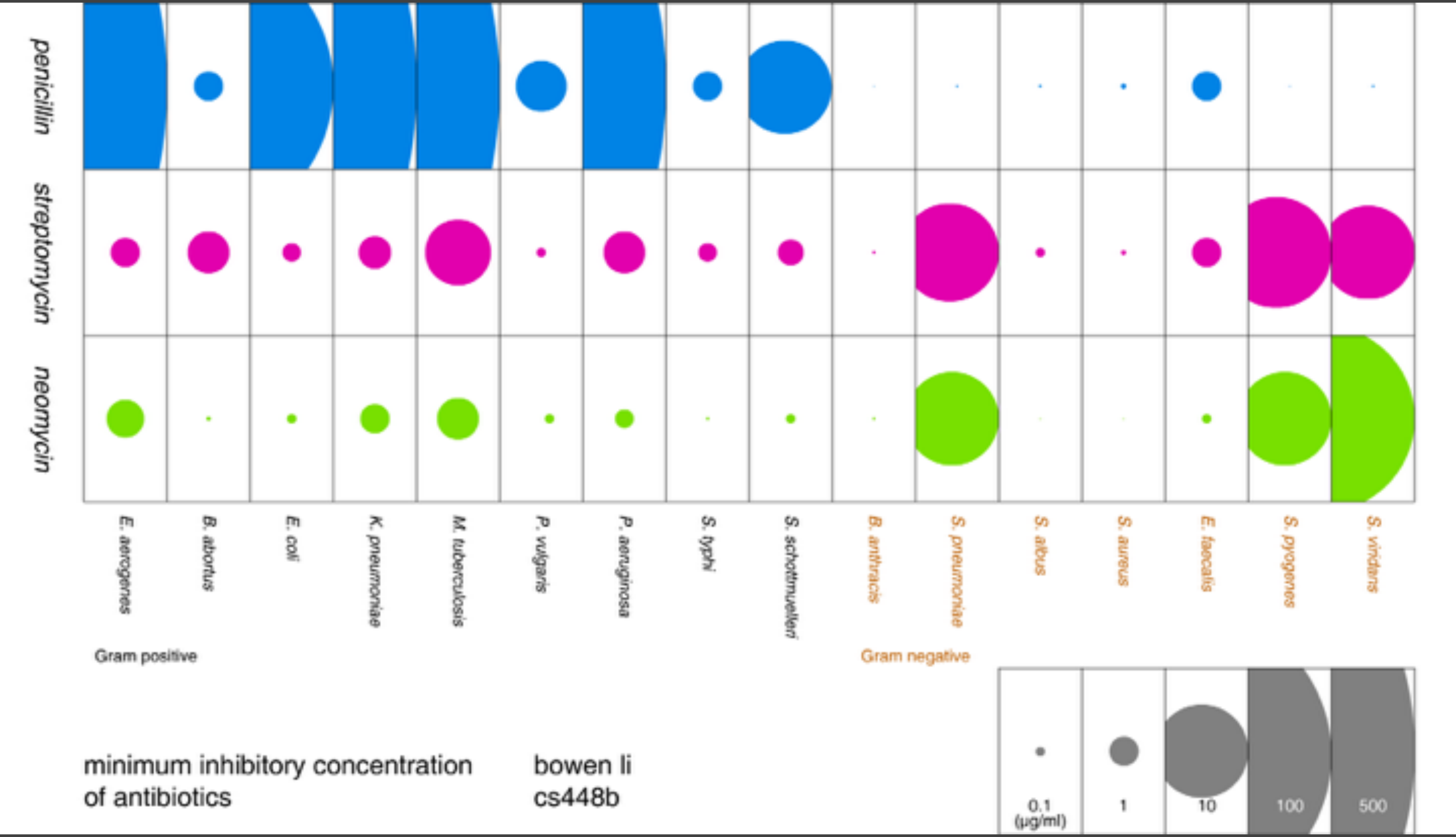


Mike Bostock
Stanford CS448B, Winter 2009

How do the drugs compare?

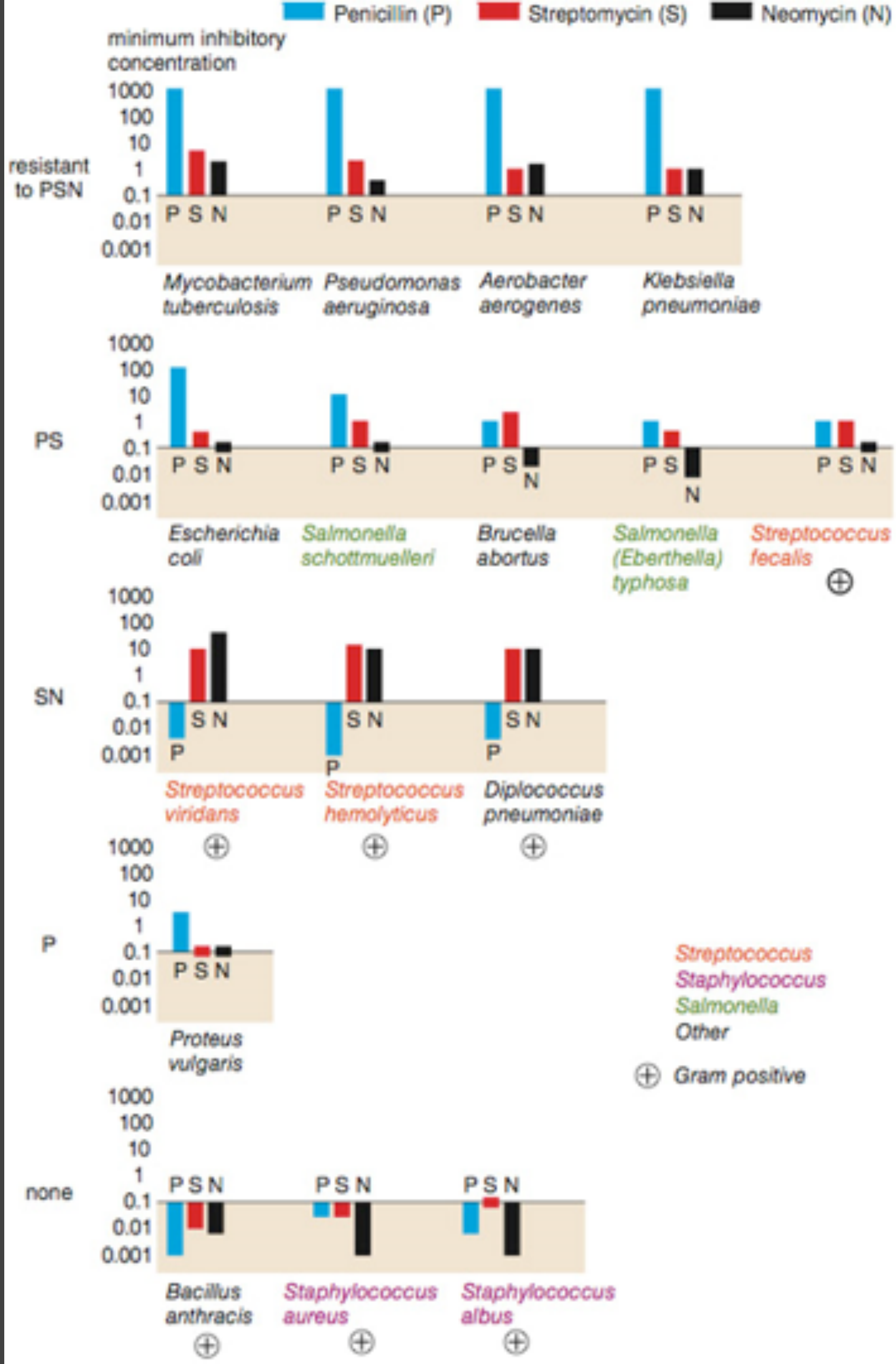


X-axis: Antibiotic | log(MIC)
Y-axis: Gram-Staining | Species
Color: Most-Effective?



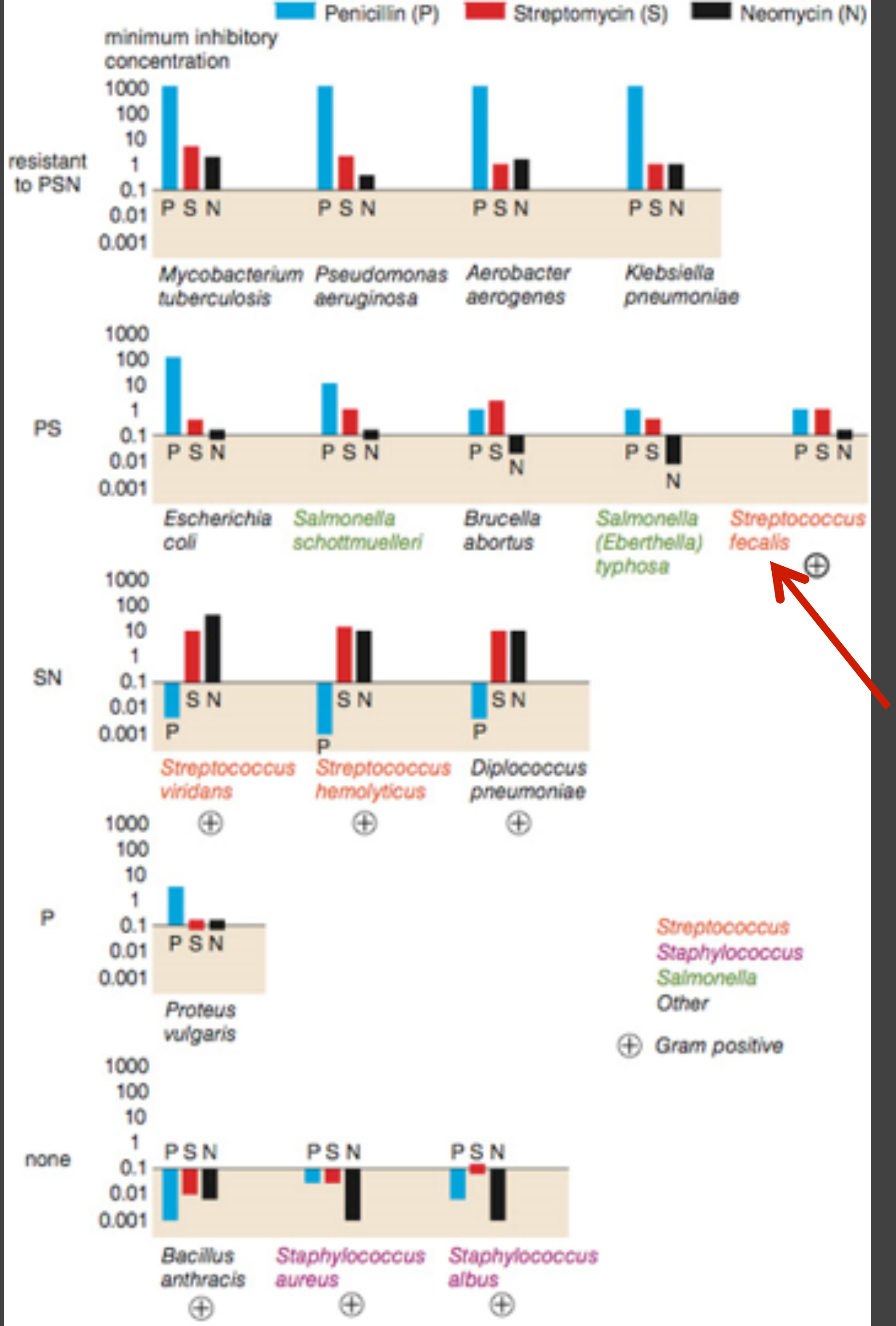
Bowen Li
Stanford CS448B, Fall 2009

Do the bacteria group by antibiotic resistance?



Do the bacteria group by antibiotic resistance?

Wainer & Lysen
American Scientist, 2009

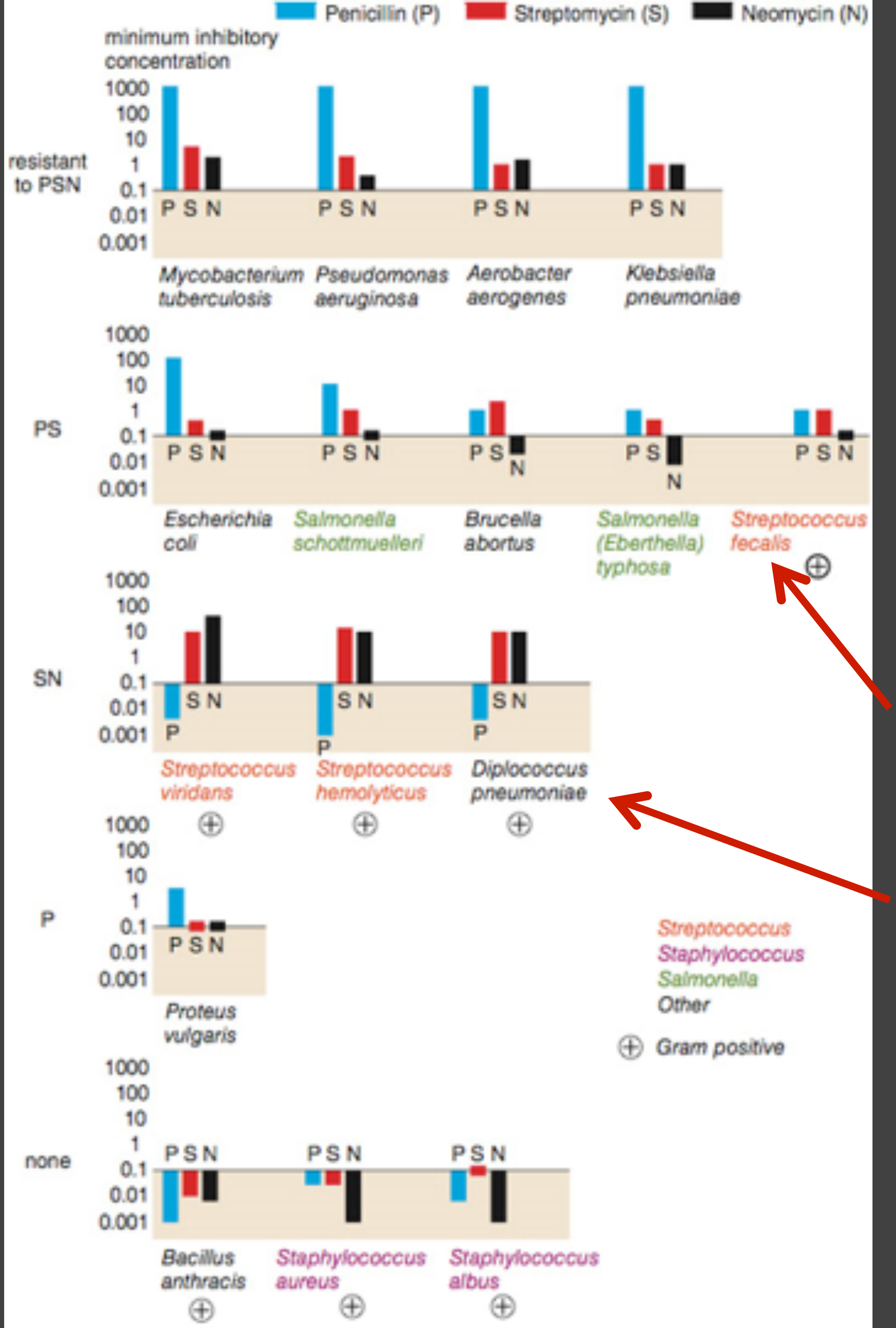


Do the bacteria group by antibiotic resistance?

Not a streptococcus! (realized ~30 yrs later)

Wainer & Lysen
American Scientist, 2009

Do the bacteria group by antibiotic resistance?

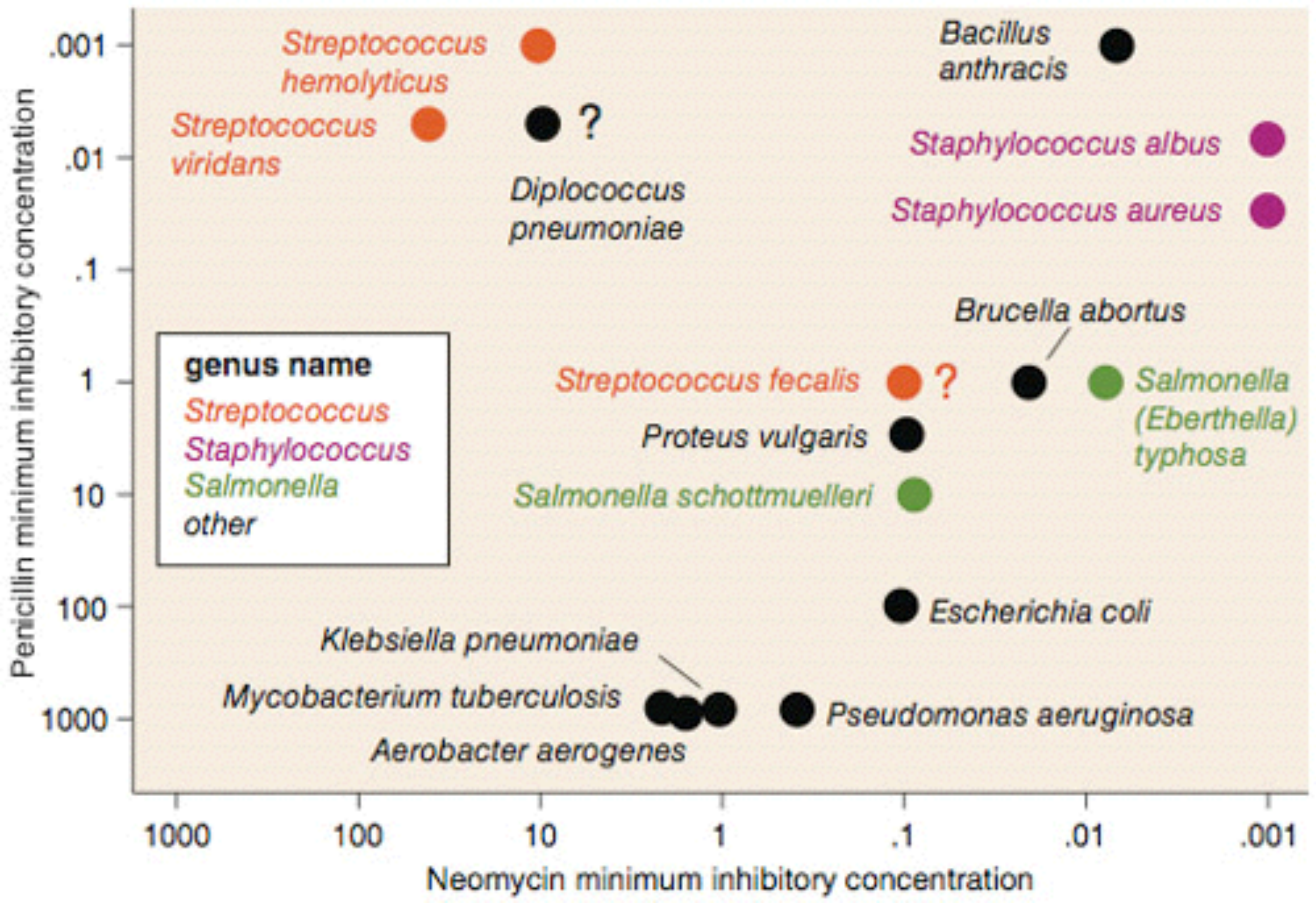


Not a streptococcus! (realized ~30 yrs later)

Really a streptococcus! (realized ~20 yrs later)

Wainer & Lysen
American Scientist, 2009

Do the bacteria group by resistance?
Do different drugs correlate?



Do the bacteria group by resistance?

Do different drugs correlate?

Wainer & Lysen
American Scientist, 2009

Lesson: Iterative Exploration

Exploratory Process

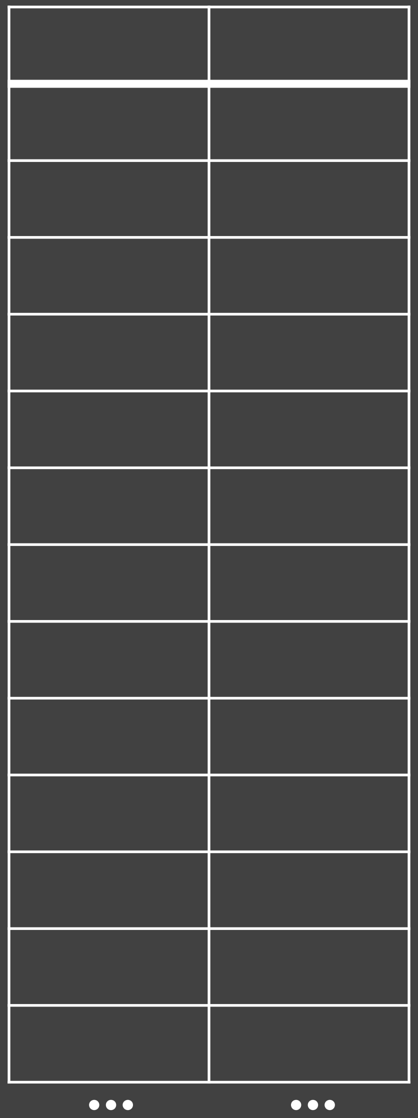
1 Construct graphics to address questions
2 Inspect “answer” and assess new questions
3 Repeat…

Transform data appropriately (e.g., invert, log)

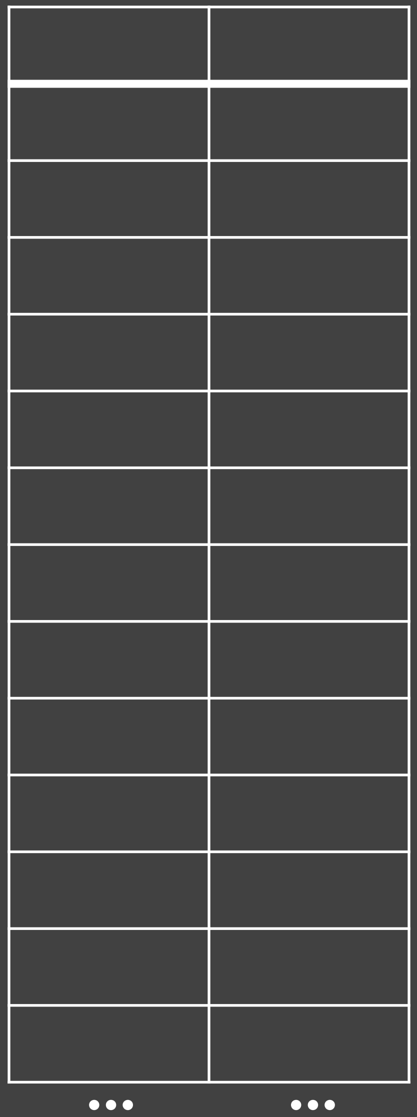
“Show data variation, not design variation” -Tufte

Visualizing Big Data

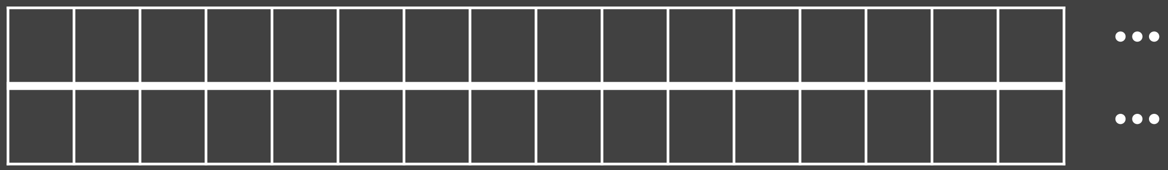
Tall data



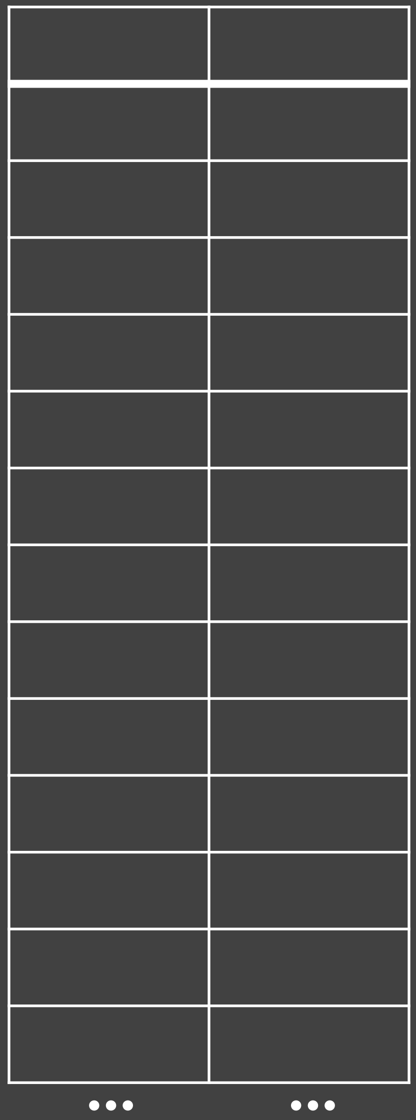
Tall data



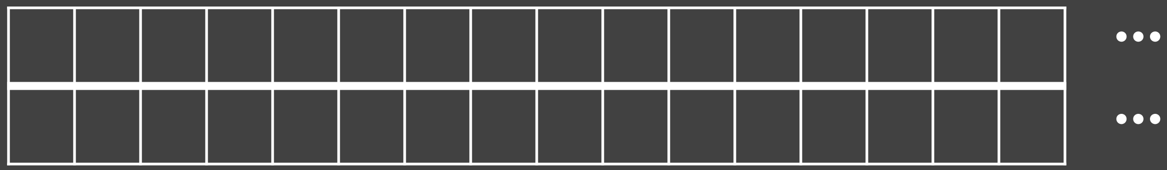
Wide data



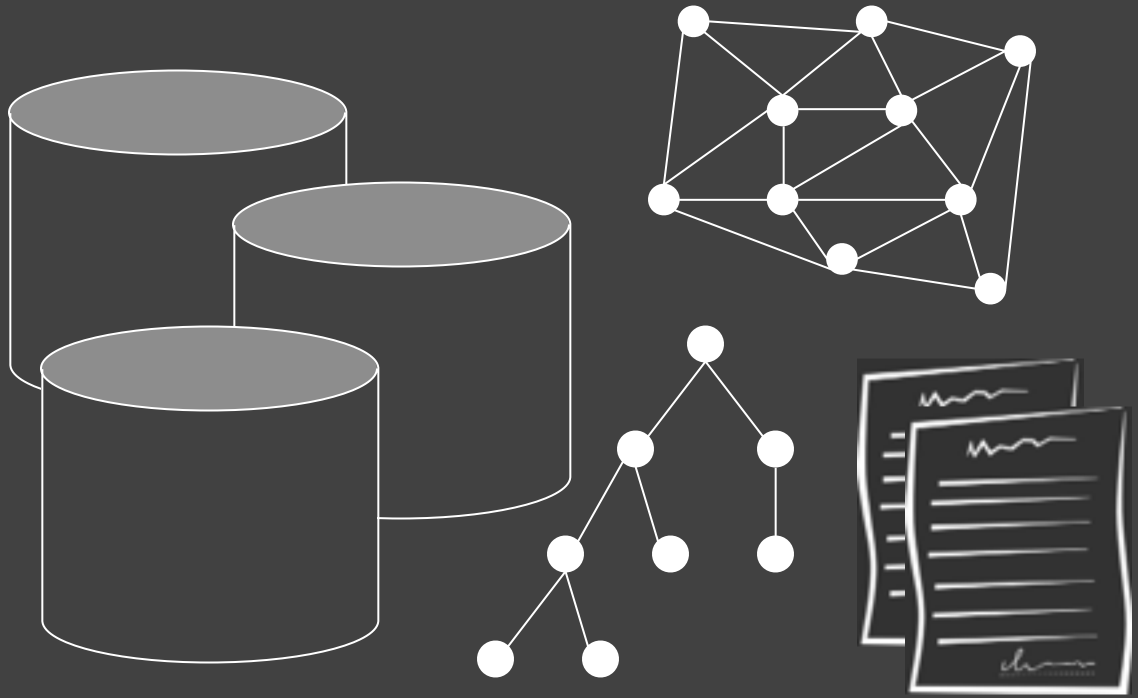
Tall data



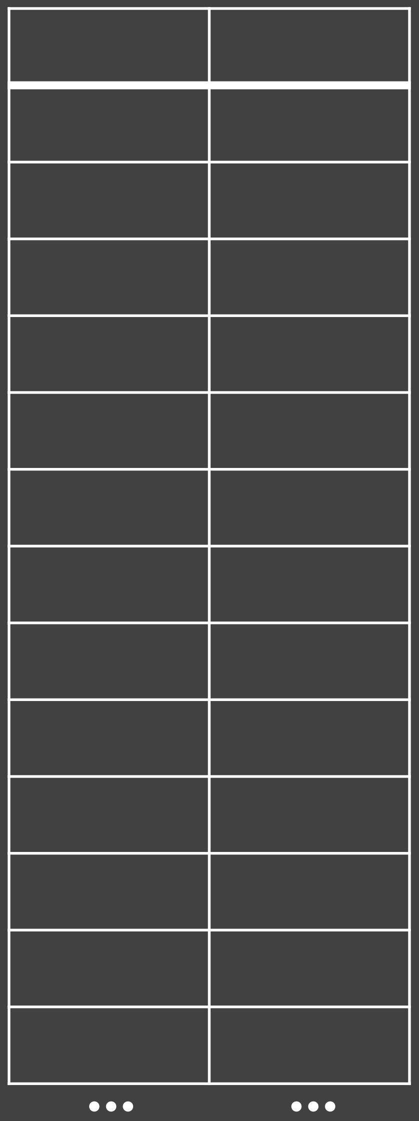
Wide data



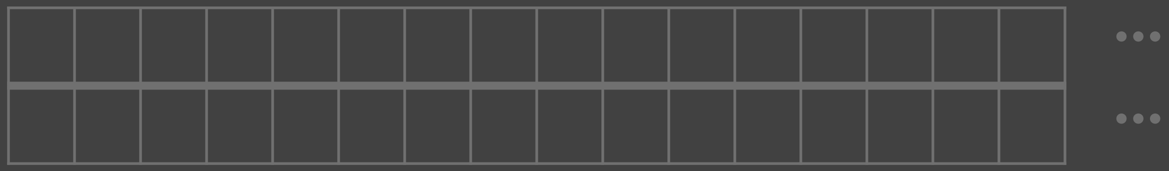
Diverse data



Tall data



Wide data



Diverse data

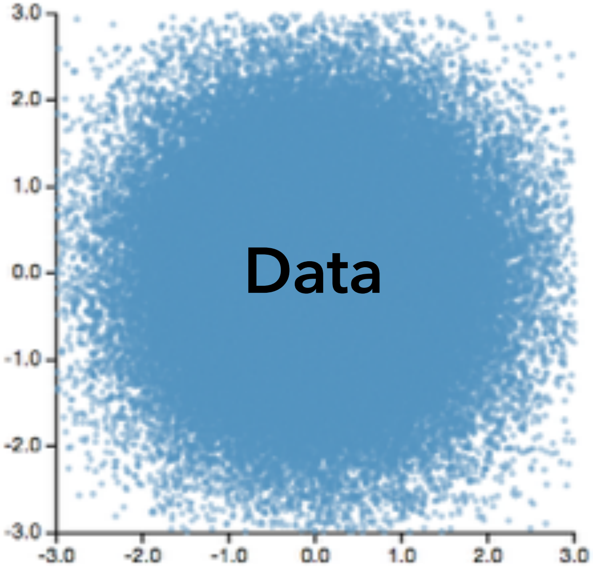
How can we visualize and interact with **billion+ record** databases in real-time?

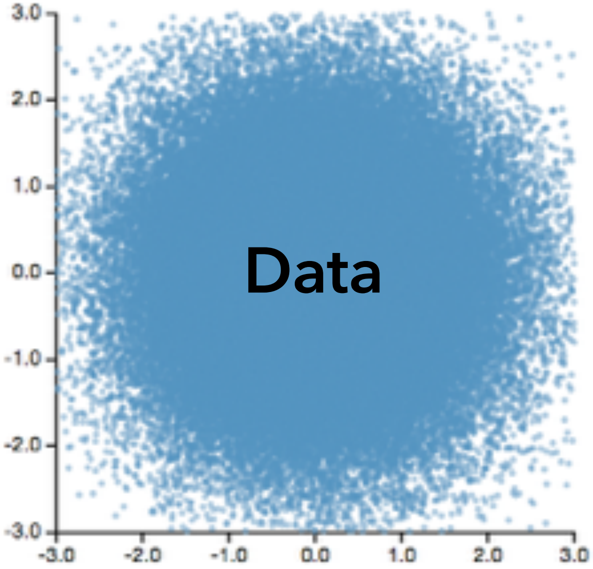
Two Challenges:

1. Effective **visual encoding**
2. Real-time **interaction**

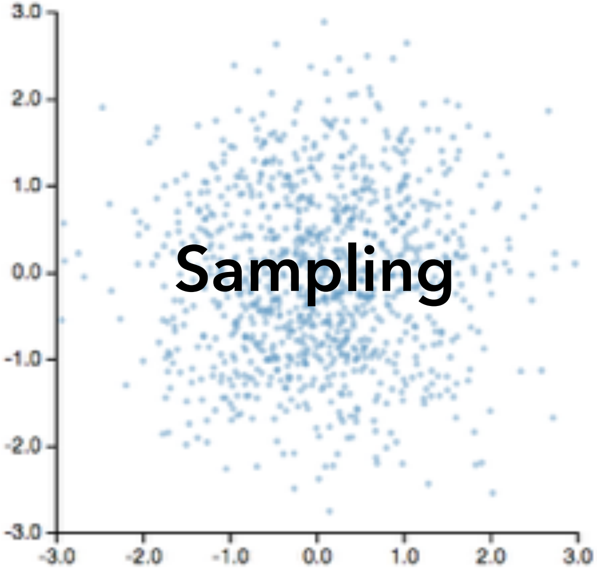
Perceptual and interactive scalability should be limited by the **chosen resolution** of the visualized data, not the number of records.

Perception

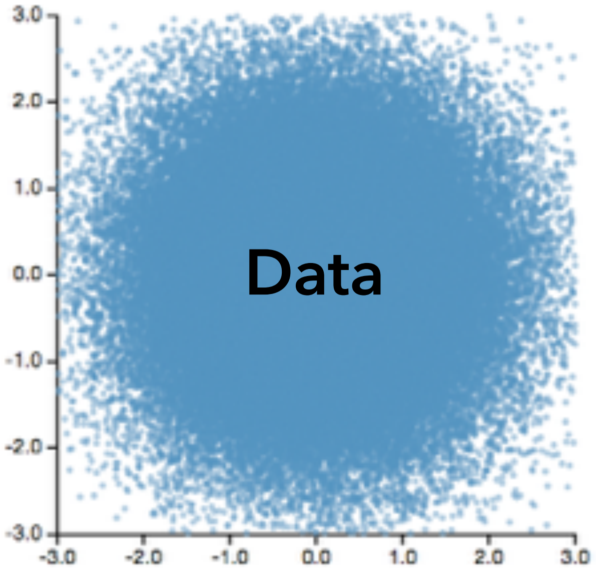

Data
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
-3.0
-2.0
-1.0
0.0
1.0
2.0
3.0



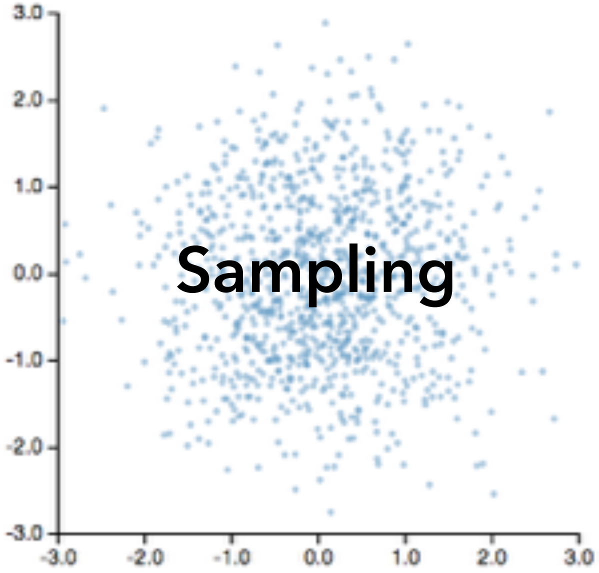
Data



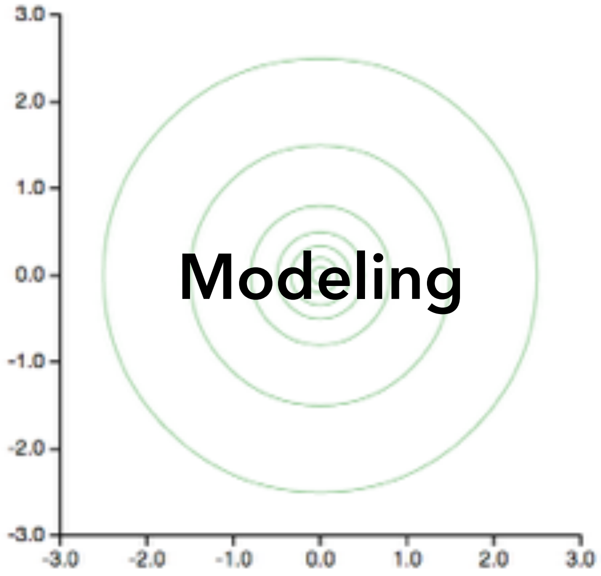
Sampling



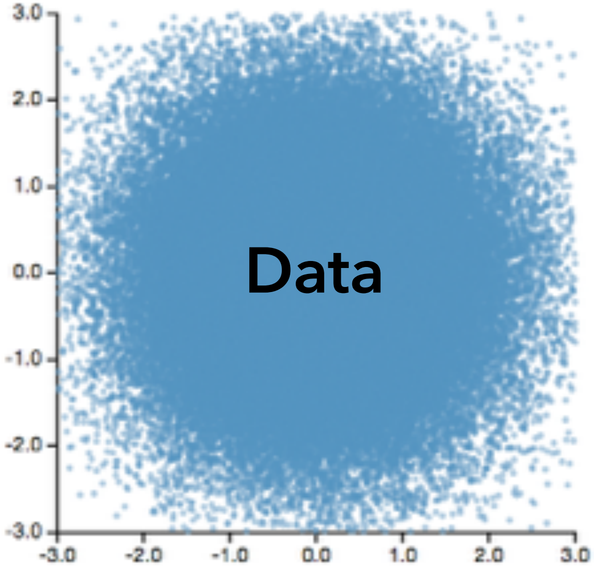
Data



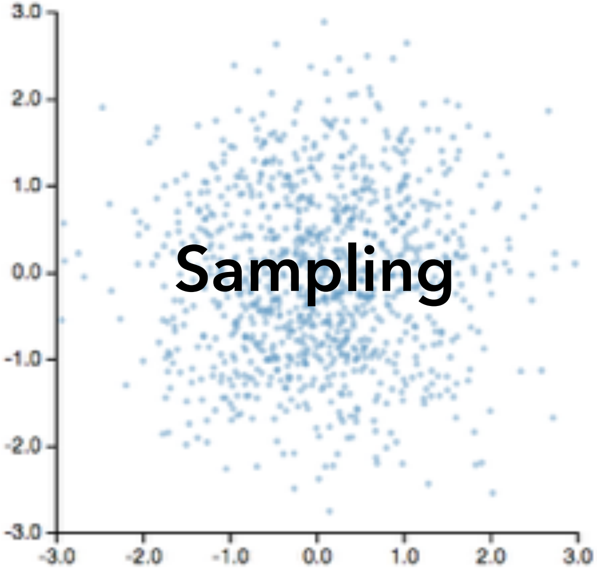
Sampling



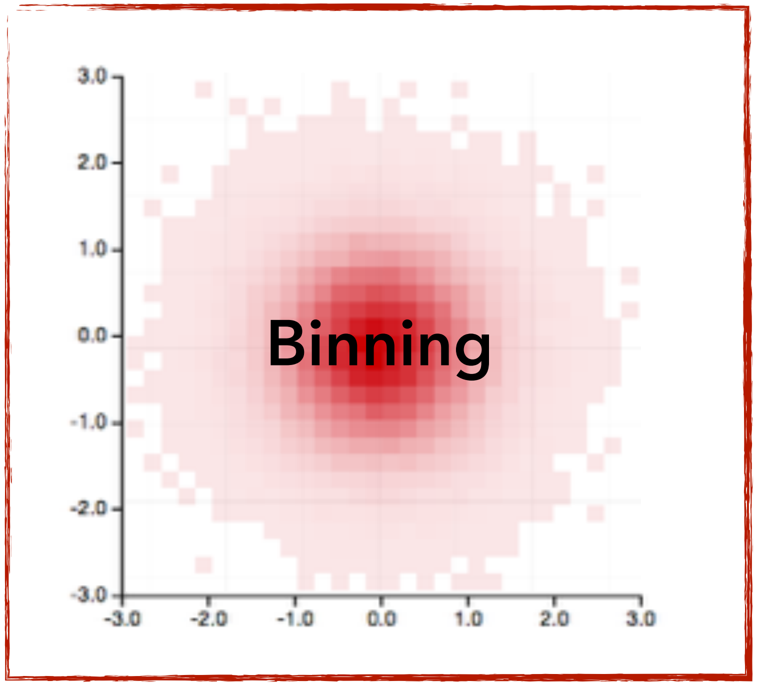
Modeling



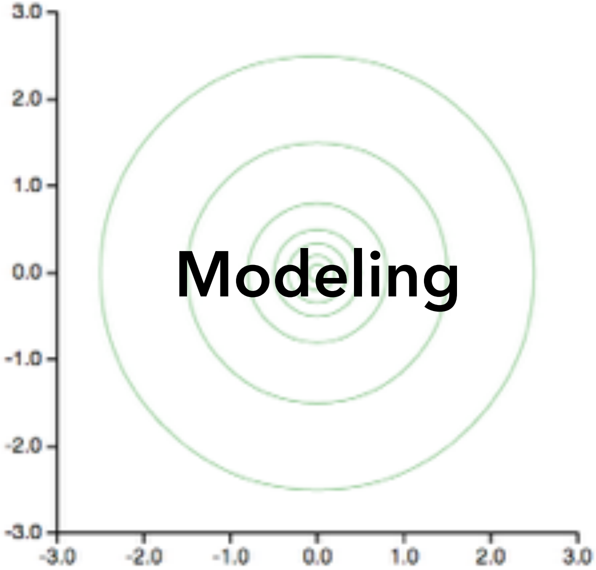
Data



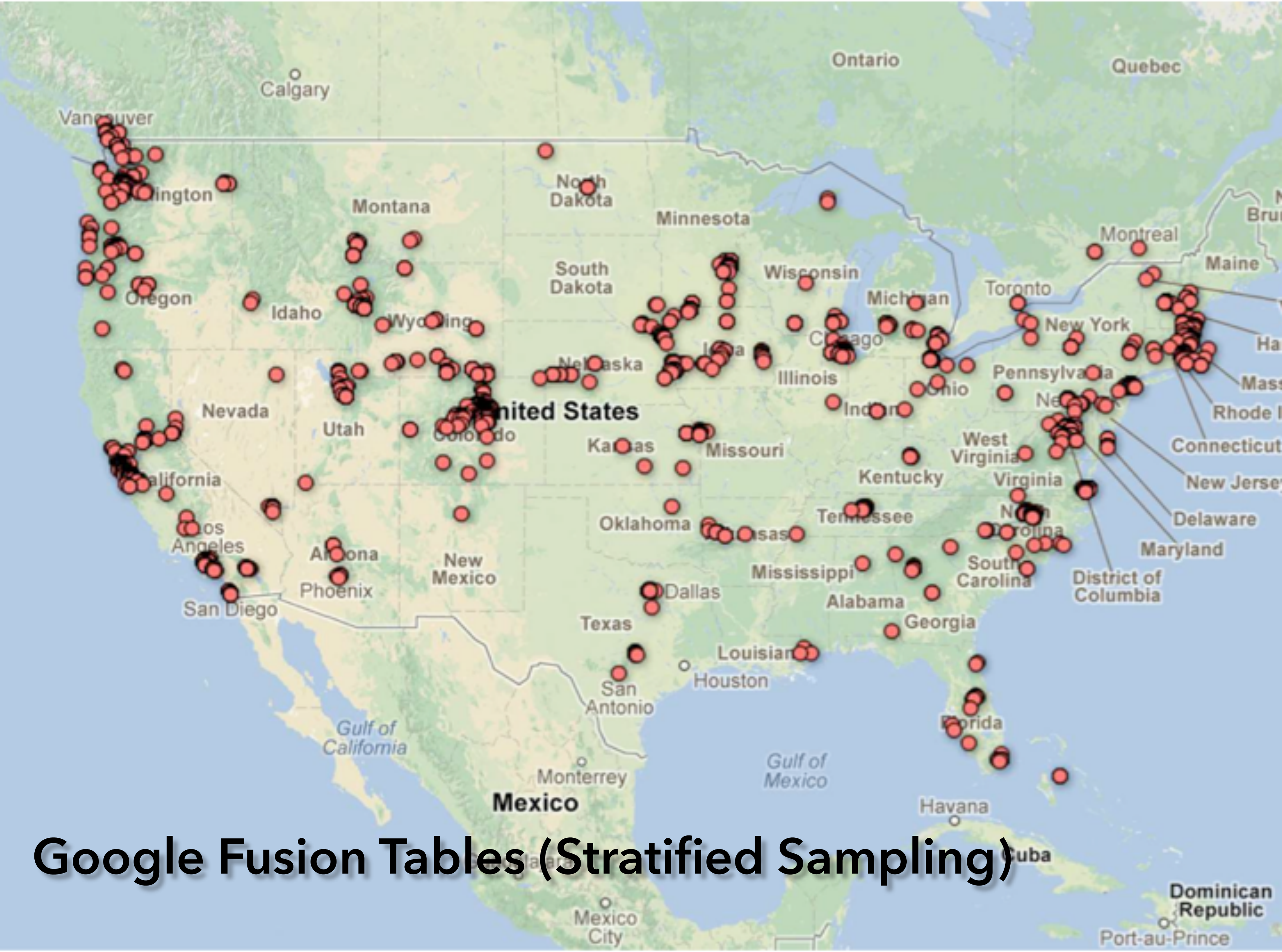
Sampling



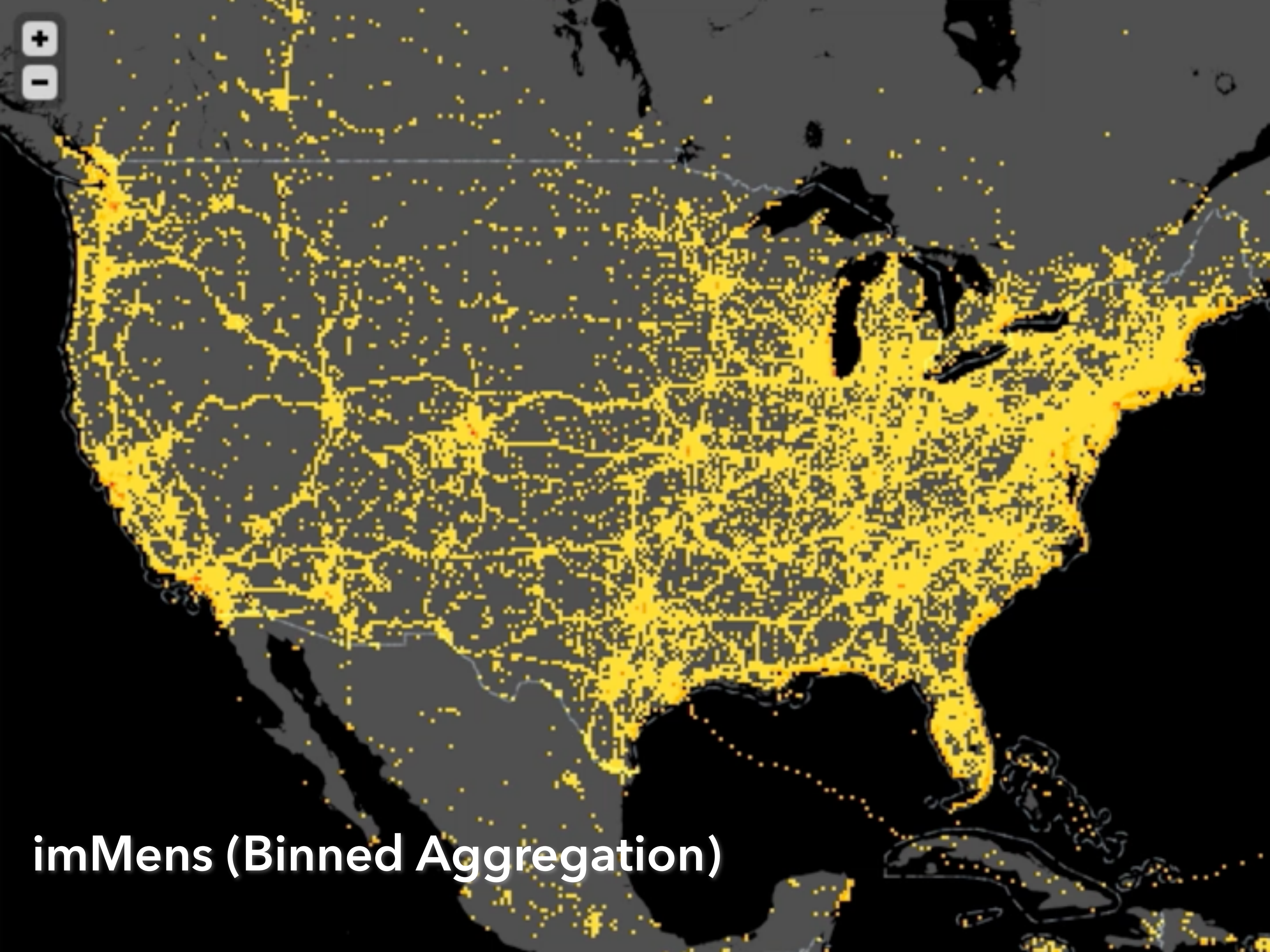
Binning



Modeling



Google Fusion Tables (Stratified Sampling)



imMens (Binned Aggregation)

Bin > Aggregate (> Smooth) > Plot

1. Bin Divide data domain into discrete “buckets”

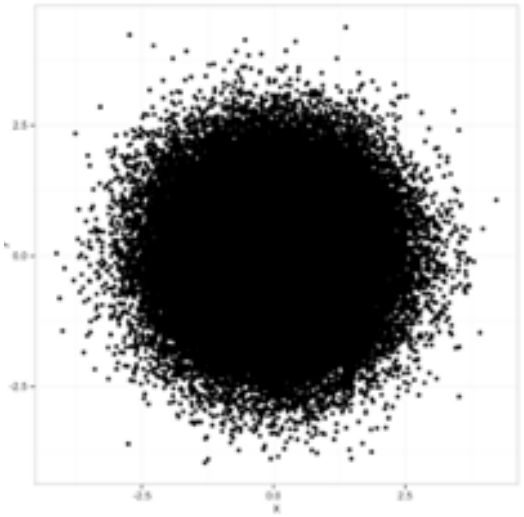
Categories: Already discrete (but check cardinality)

Numbers: Choose bin intervals (uniform, quantile, ...)

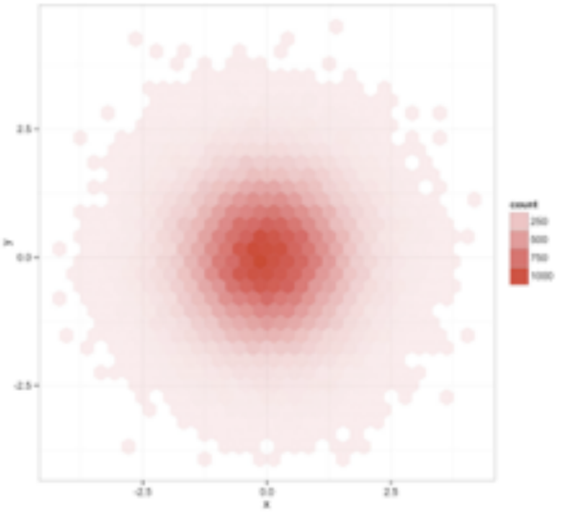
Time: Choose time unit: Hour, Day, Month, etc.

Geo: Bin x, y coordinates *after* cartographic projection

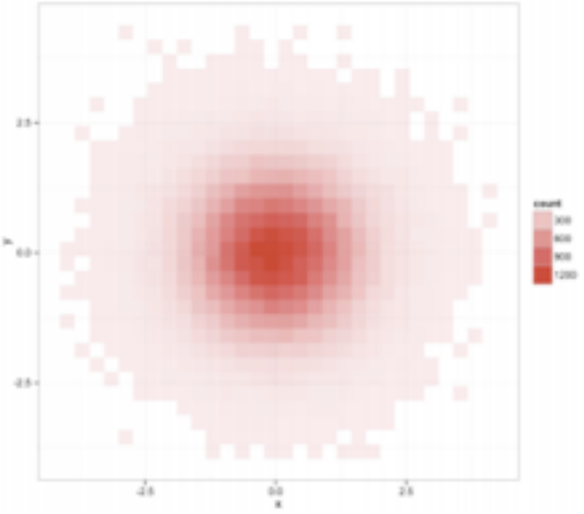
Hexagonal or Rectangular Bins?



100,000 Data Points



Hexagonal Bins



Rectangular Bins

Hex bins better estimate density for 2D plots, but **the improvement is marginal** [Scott 92], while rectangles support **reuse** and **query processing**.

Bin > Aggregate (> Smooth) > Plot

1. Bin Divide data domain into discrete “buckets”

Categories: Already discrete (but check cardinality)

Numbers: Choose bin intervals (uniform, quantile, ...)

Time: Choose time unit: Hour, Day, Month, etc.

Geo: Bin x, y coordinates *after* cartographic projection

Bin > Aggregate (> Smooth) > Plot

1. Bin Divide data domain into discrete "buckets"

Categories: Already discrete (but check cardinality)

Numbers: Choose bin intervals (uniform, quantile, ...)

Time: Choose time unit: Hour, Day, Month, etc.

Geo: Bin x, y coordinates *after* cartographic projection

2. Aggregate Count, Sum, Average, Min, Max, ...

Bin > Aggregate (> Smooth) > Plot

1. Bin Divide data domain into discrete "buckets"

Categories: Already discrete (but check cardinality)

Numbers: Choose bin intervals (uniform, quantile, ...)

Time: Choose time unit: Hour, Day, Month, etc.

Geo: Bin x, y coordinates *after* cartographic projection

2. Aggregate Count, Sum, Average, Min, Max, ...

(**3. Smooth** Optional: smooth aggregates [Wickham '13])

Bin > Aggregate (> Smooth) > Plot

1. Bin Divide data domain into discrete “buckets”

Categories: Already discrete (but check cardinality)

Numbers: Choose bin intervals (uniform, quantile, ...)

Time: Choose time unit: Hour, Day, Month, etc.

Geo: Bin x, y coordinates *after* cartographic projection

2. Aggregate Count, Sum, Average, Min, Max, ...

(**3. Smooth** Optional: smooth aggregates [Wickham ’13])

4. Plot Visualize the aggregate summary values

Plot: Visual Encoding

Use Most Effective Encoding [Cleveland & McGill ’84]

1D Plot -> Position or Length Encoding

- Histograms, line charts, etc.

Plot: Visual Encoding

Use Most Effective Encoding [Cleveland & McGill ’84]

1D Plot -> Position or Length Encoding

Histograms, line charts, etc.

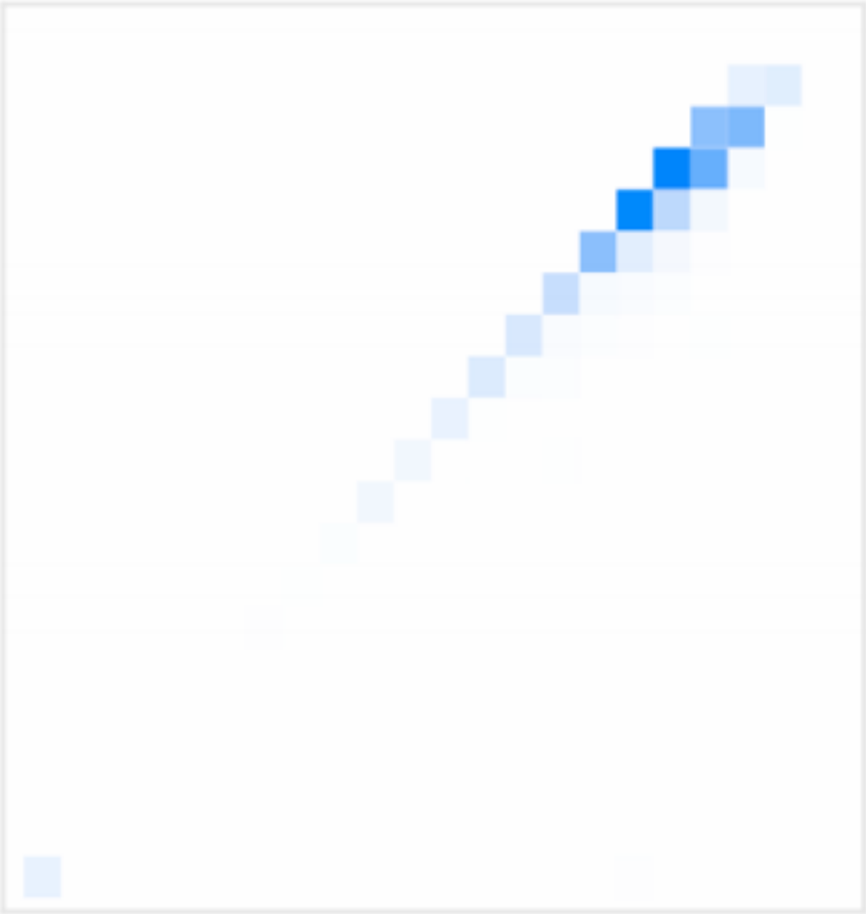
2D Plot -> Area or Color Encoding

Spatial dimensions (x, y) already allocated.

While less effective than **area** for magnitude estimation, **color** can be used at the per-pixel level and provides an overall “gestalt”

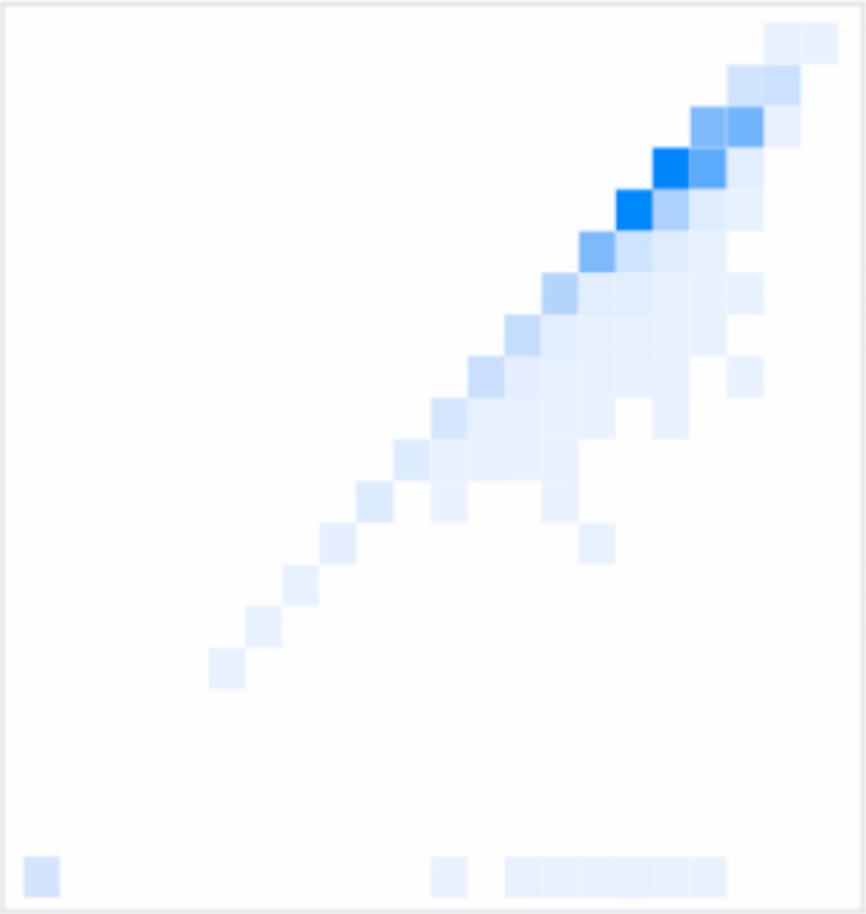
Design Space of Binned Plots

	Numeric	Ordinal	Temporal	Geographic



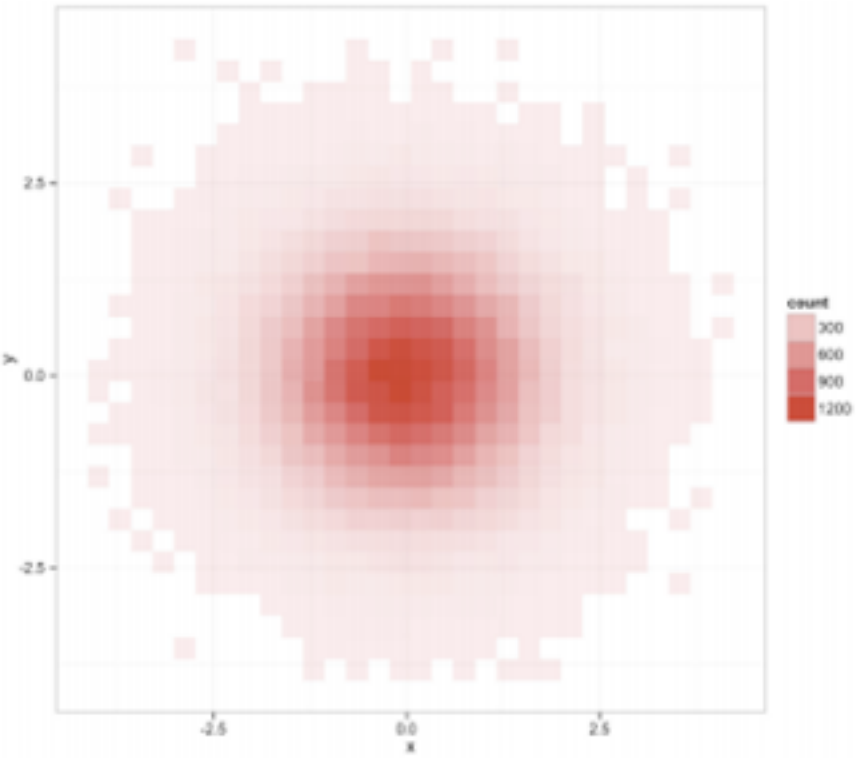
Standard Color Ramp

Counts near zero are white.
-> Outliers are missed

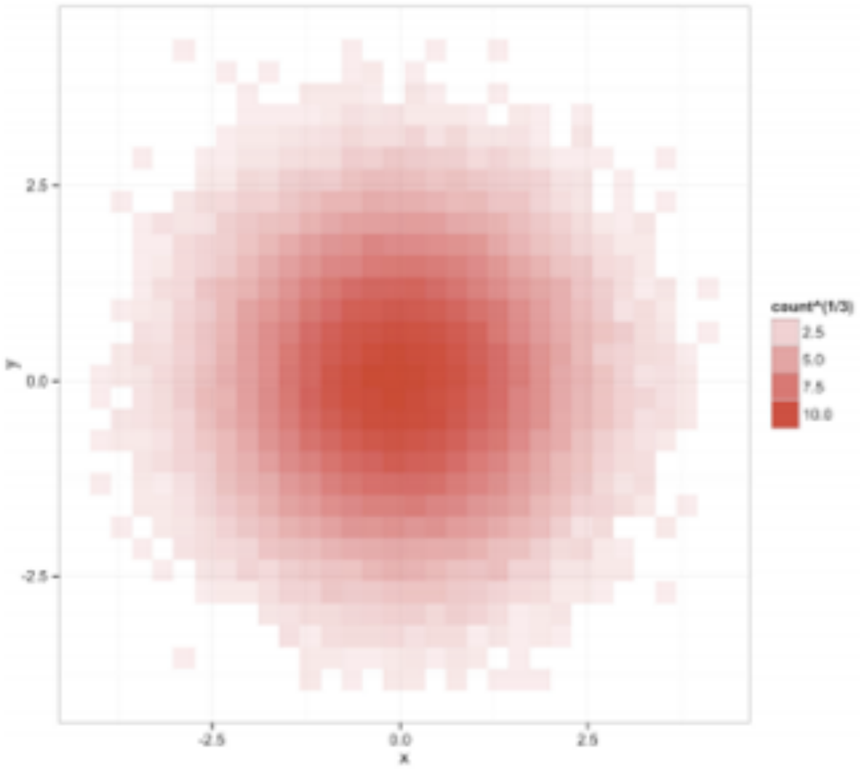


Add Discontinuity after Zero

Counts near zero remain visible.
-> Outliers can be seen



Linear Alpha Interpolation
is not *perceptually* linear.



Cube-Root Alpha Interpolation
approximates perceptual linearity.

Color Encoding

Data Value ($x > 0$, $x >= x_{min}$, $x <= x_{max}$)

$$Y = \alpha + \left(\frac{\hat{x} - x_{min}}{x_{max} - x_{min}} \right)^{\gamma} (1 - \alpha)$$

Luminance (in range 0-1)

Color Encoding

Min. Non-Zero Intensity (α=0.15) [1]

Perceptual Scaling (γ=1/3) [2]

$$Y = \alpha + \left(\frac{\hat{x} - x_{min}}{x_{max} - x_{min}} \right)^{\gamma} (1 - \alpha)$$

User-Adjustable Min/Max Values [3]

[1] Keep small non-zero values visible (outliers!)

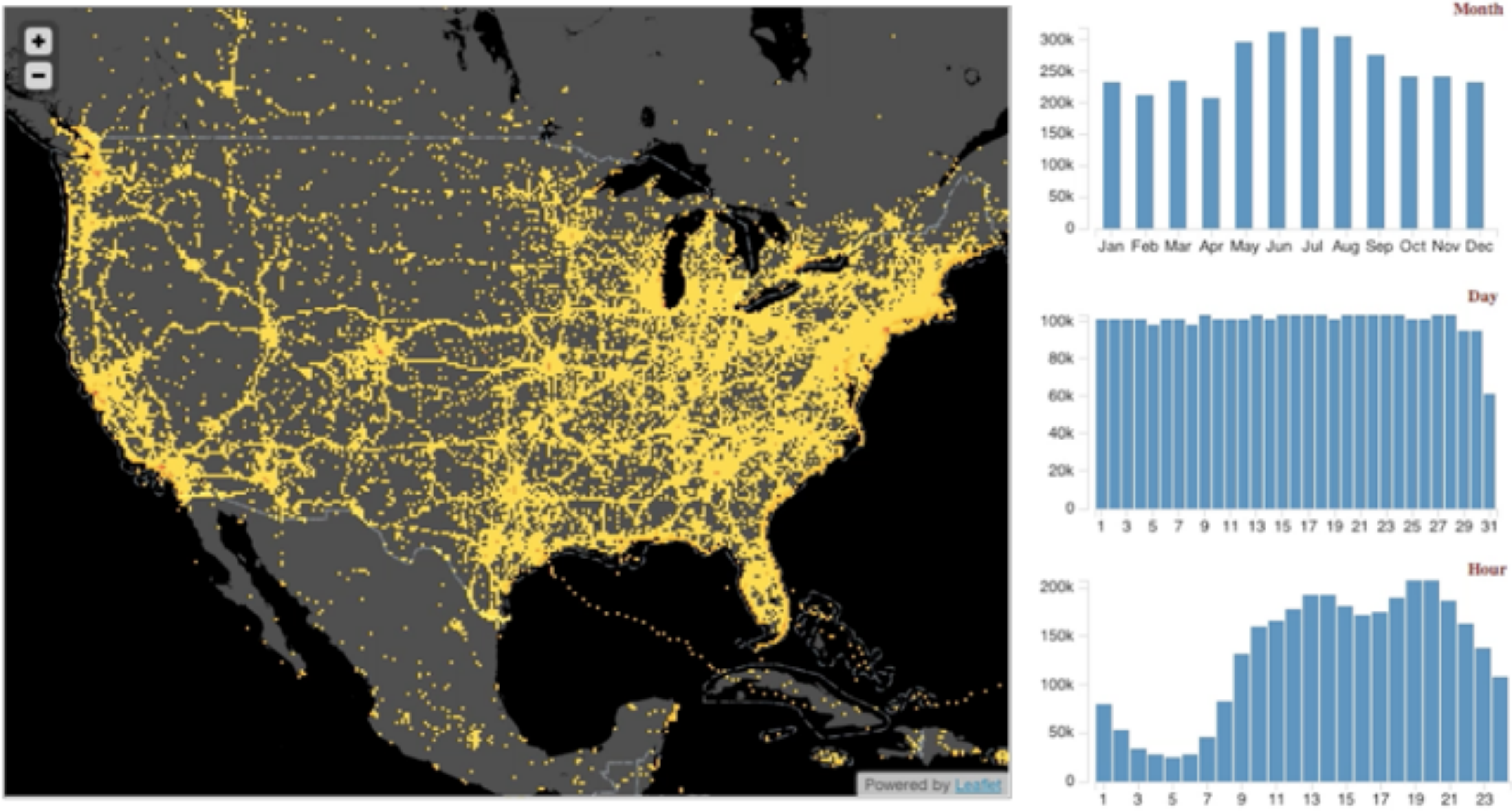
[2] Match color ramp to perceptual distances

[3] Enable exploration across value ranges

Interaction

Interaction Techniques?

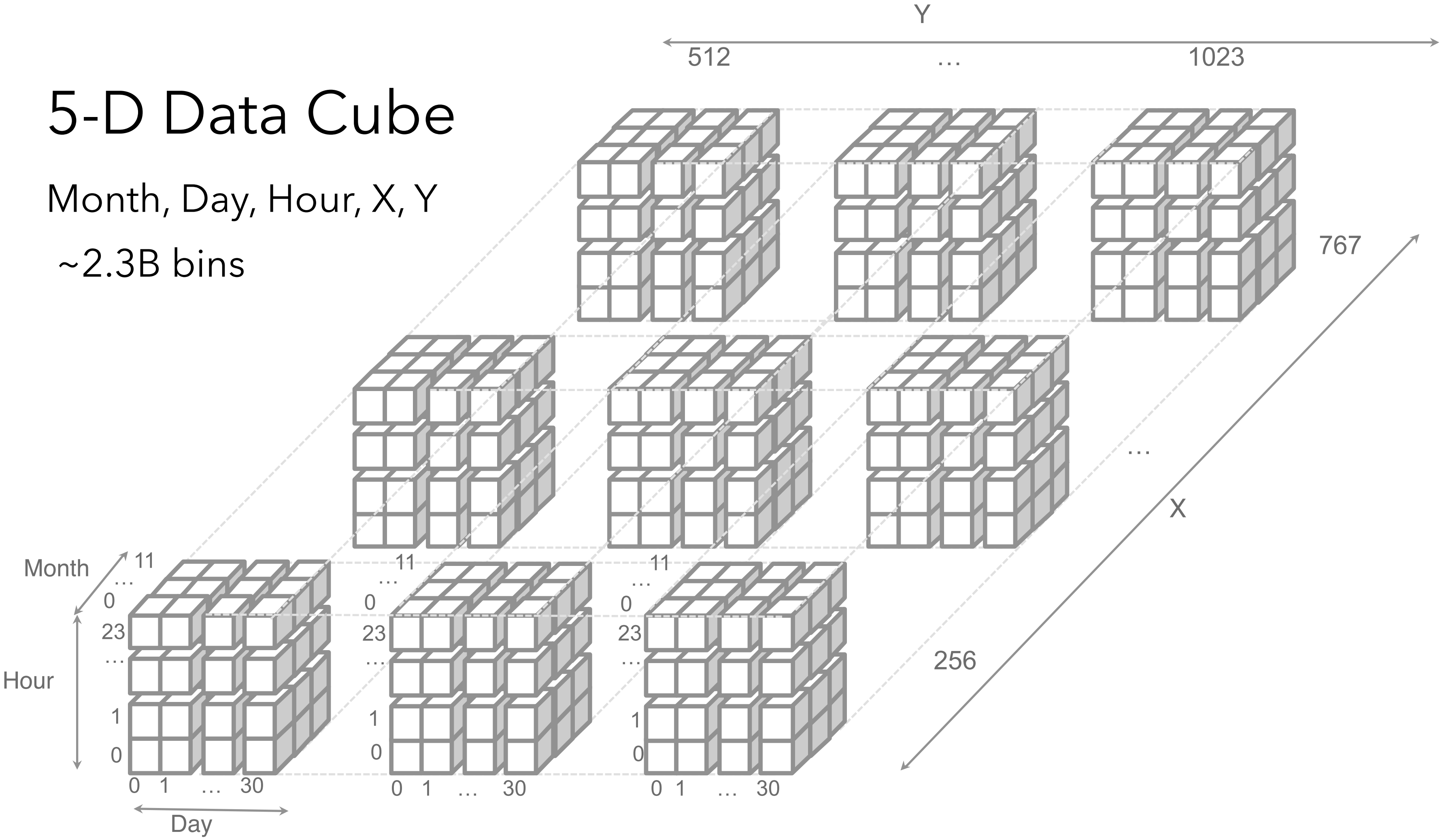
1. **Select**	Detail-on-Demand
2. **Navigate**	Pan & Zoom
3. **Query**	Brush & Link


Powered by Leaflet
Month
300k
250k
200k
150k
100k
50k
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Day
100k
80k
60k
40k
20k
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31
Hour
200k
150k
100k
50k
0
1 3 5 7 9 11 13 15 17 19 21 23

5-D Data Cube

Month, Day, Hour, X, Y

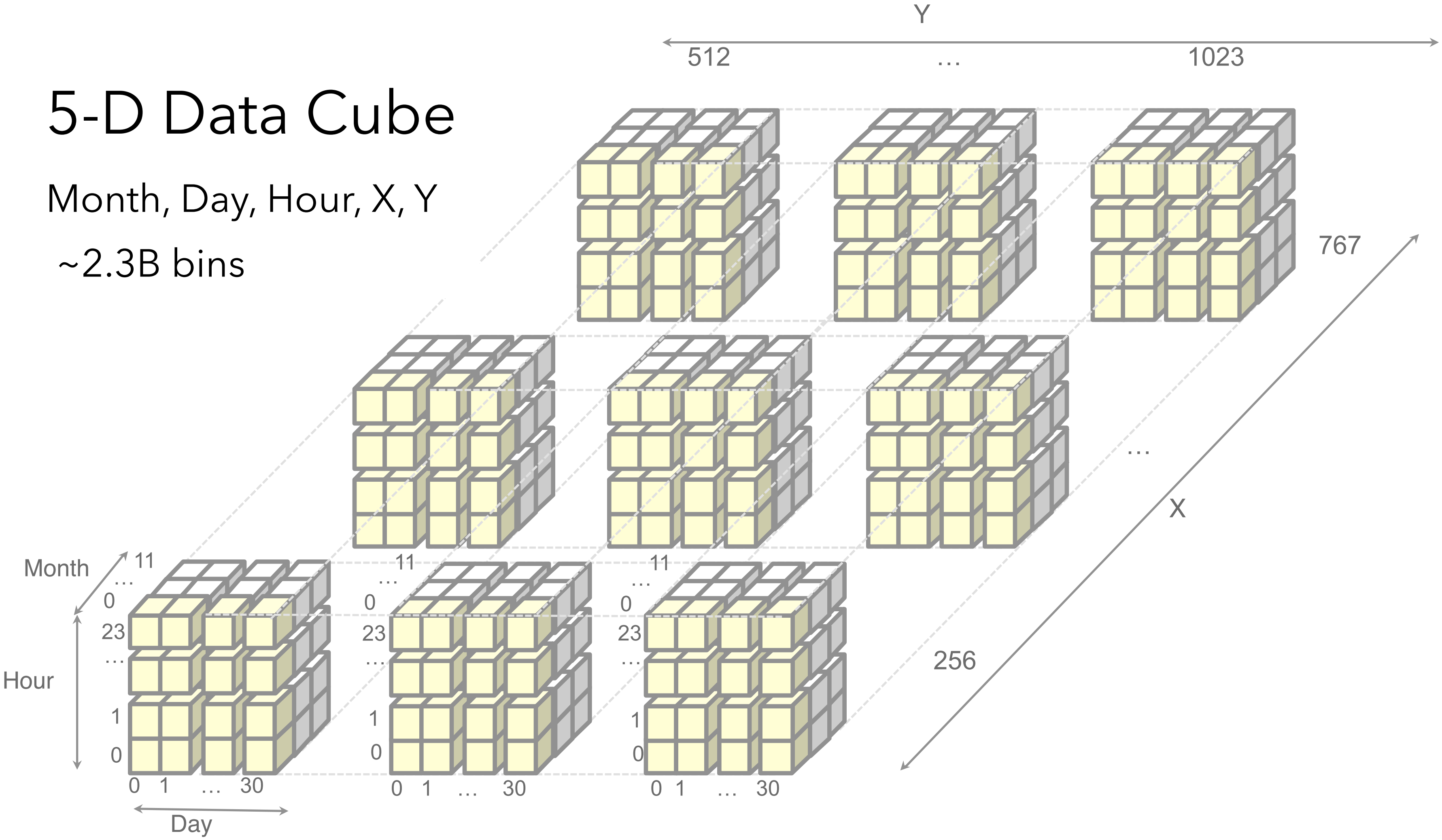
~2.3B bins

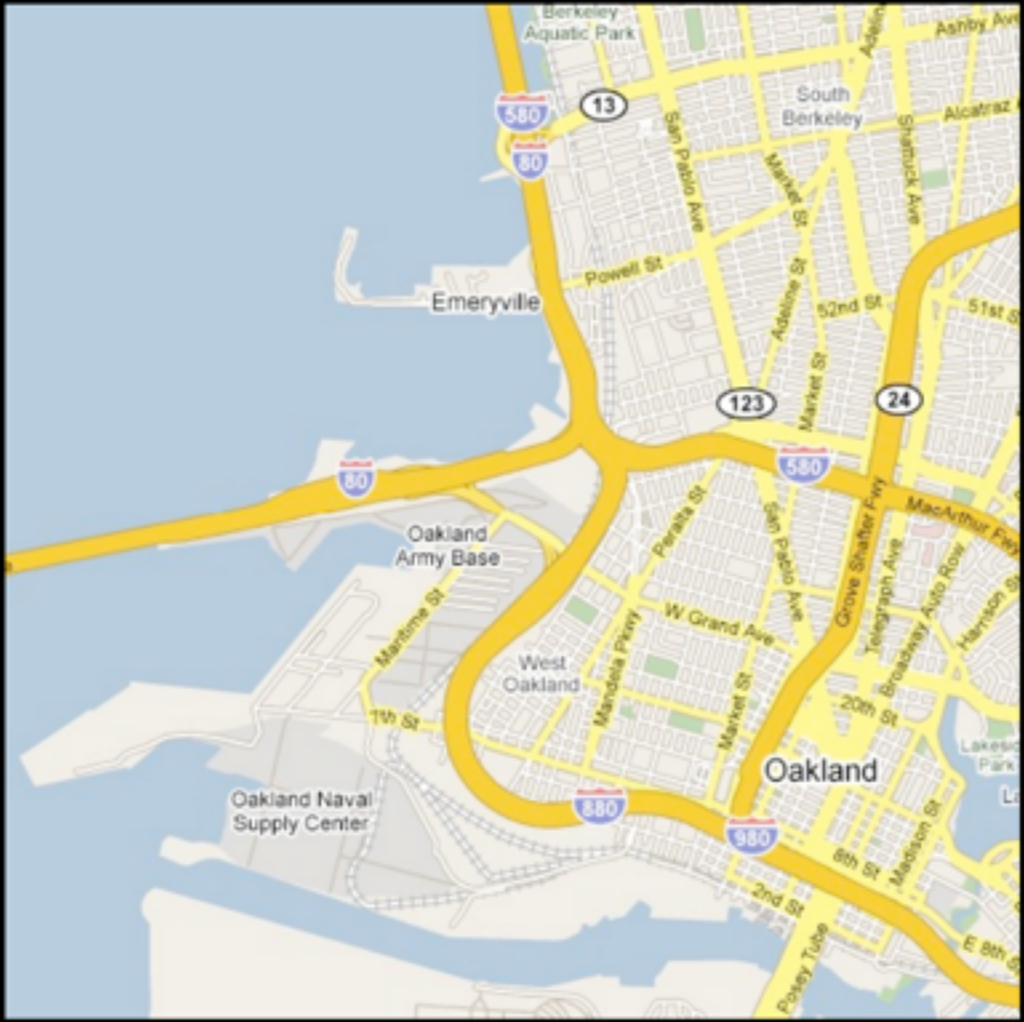


5-D Data Cube

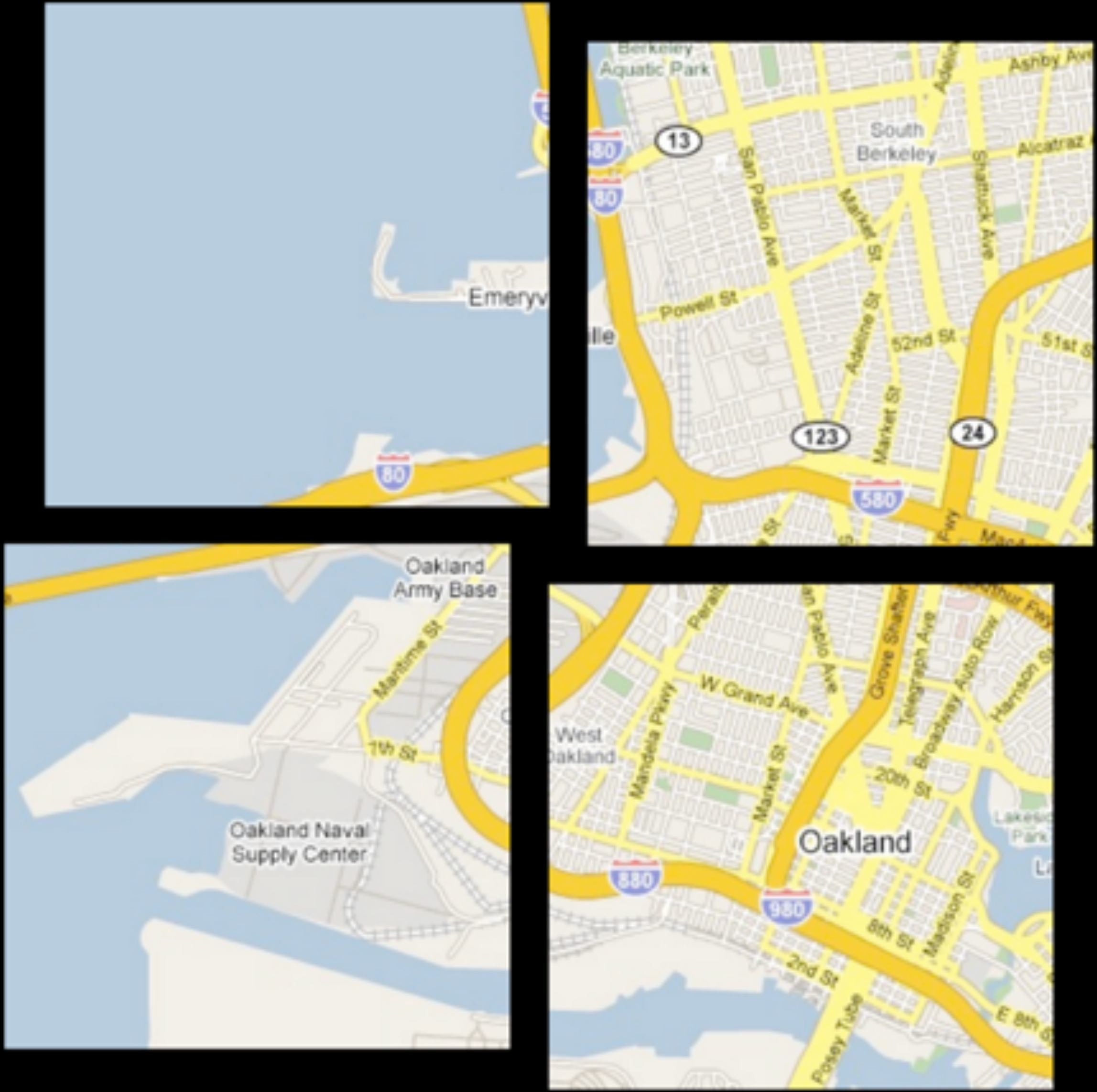
Month, Day, Hour, X, Y

~2.3B bins





Berkeley
Aquatic Park
Ashby Ave
Adeline
13
580
80
South
Berkeley
Alcatraz
San Pablo Ave
Market St
Shattuck Ave
Powell St
Emeryville
Adeline St
52nd St
51st St
Market St
123
24
580
80
MacArthur Fwy
Grove Shafter Fwy
Oakland
Army Base
Peralta St
San Pablo Ave
Telegraph Ave
Broadway Auto Row
Harrison St
W Grand Ave
Maritime St
West
Oakland
Mandela Pkwy
Market St
20th St
7th St
Lakeside
Park
Oakland
Oakland Naval
Supply Center
880
980
8th St
Madison St
2nd St
E 8th St
Posey Tube

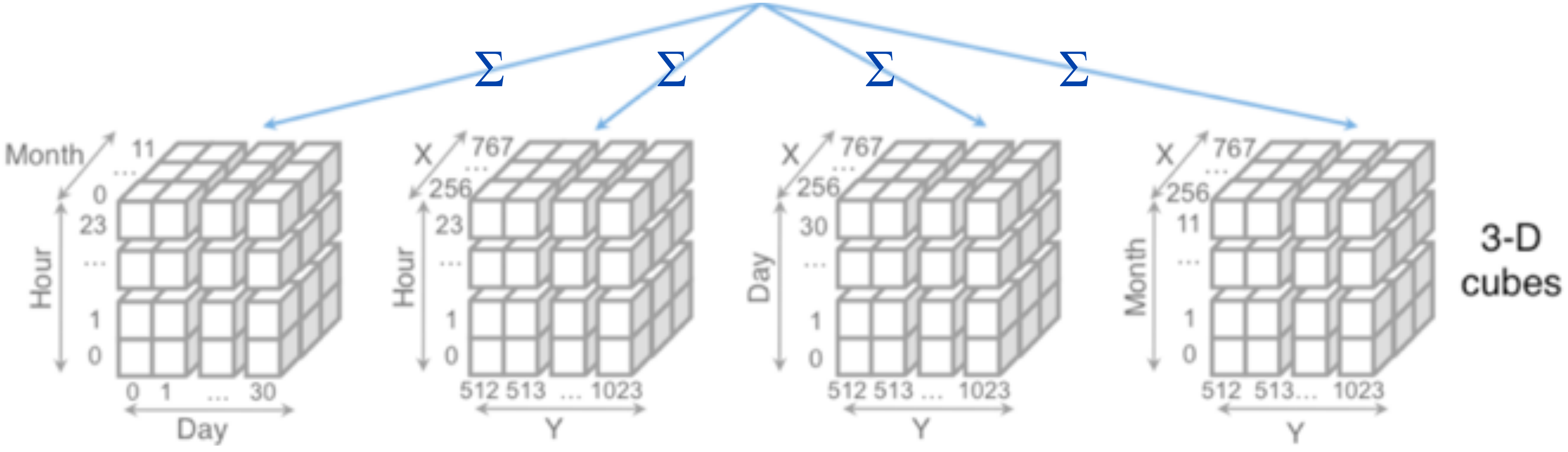

Berkeley
Aquatic Park
Ashby Ave
13
880
80
South
Berkeley
Alcatraz
San Pablo Ave
Market St
Adeline
Shattuck Ave
Emeryv
lle
Powell St
Adeline St
52nd St
51st St
123
24
Market St
580
Oakland
Army Base
Maritime St
7th St
Oakland Naval
Supply Center
80
W Grand Ave
West
akland
Mandela Pkwy
Market St
Grove Shafter
Telegraph Ave
Broadway Auto Row
Harrison St
20th St
Oakland
Lakeside
Park
880
980
8th St
Madison St
2nd St
Posey Tube
E 8th St
MacArthur Fwy

Multivariate Data Tiles

1. Send data, not pixels
2. Embed multi-dim data

Full 5-D Cube

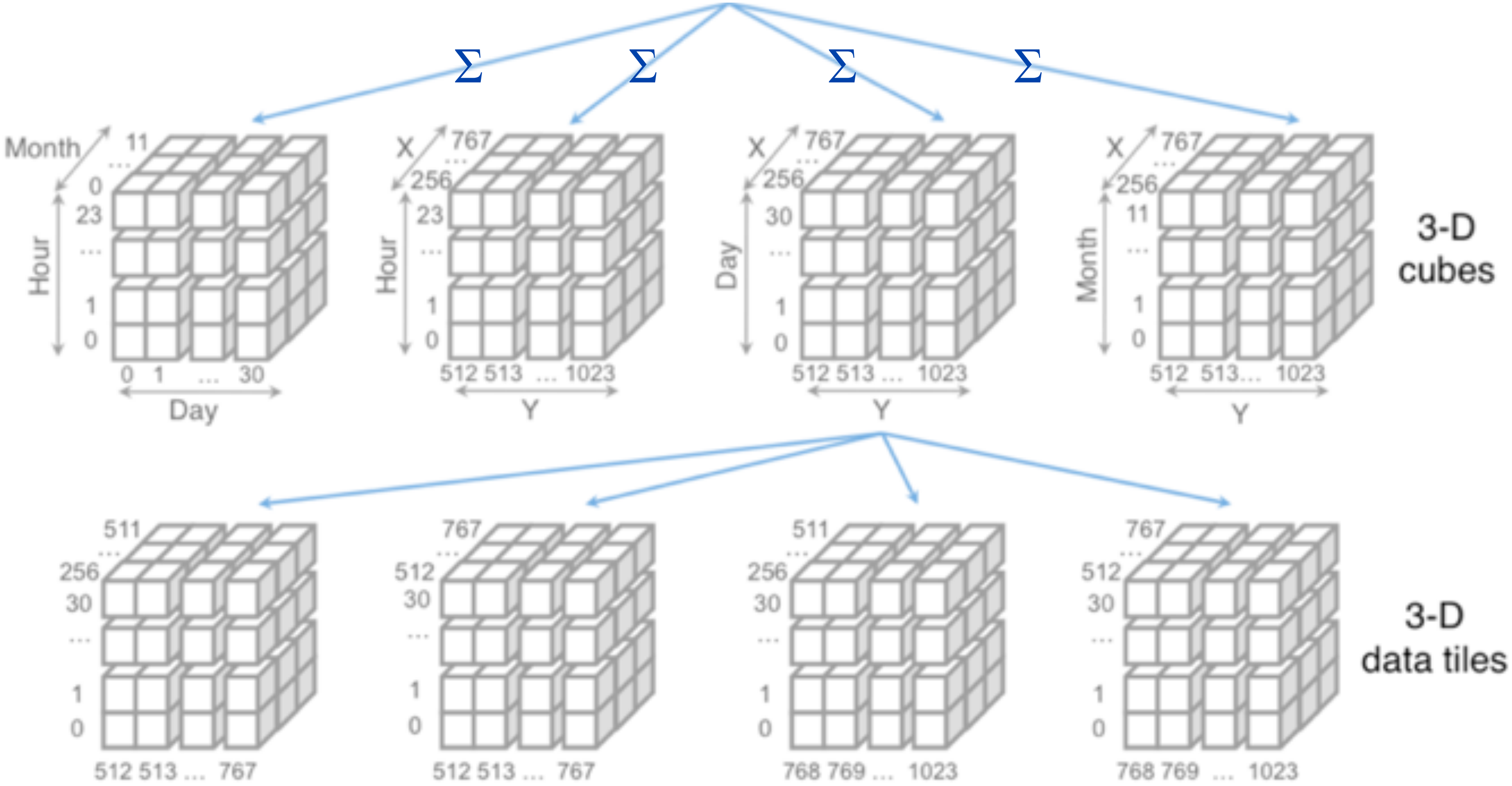
Full 5-D Cube



3-D cubes

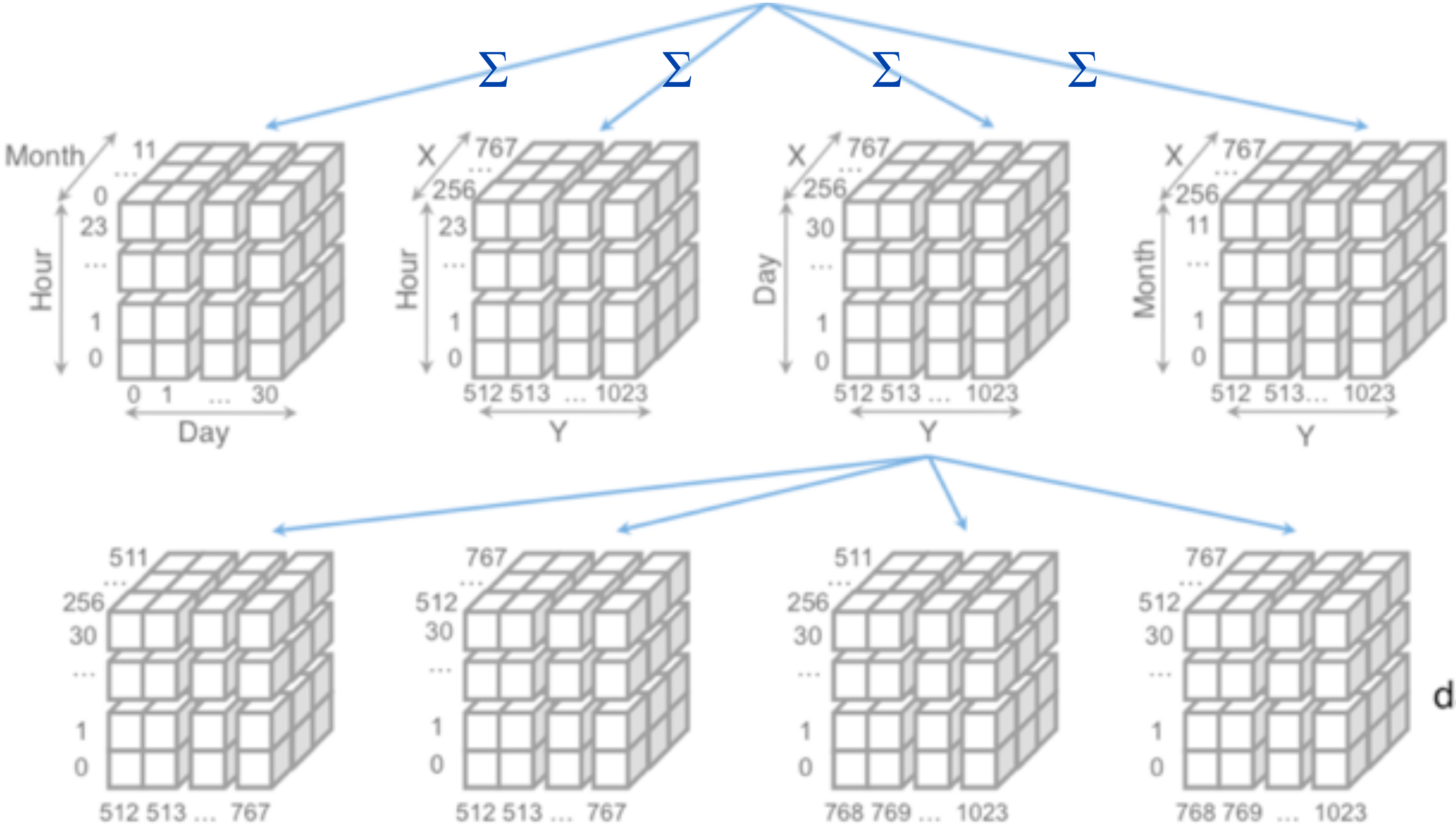
For any pair of 1D or 2D binned plots, the maximum number of dimensions needed to support brushing & linking is **four**.

Full 5-D Cube



13 3-D Data Tiles

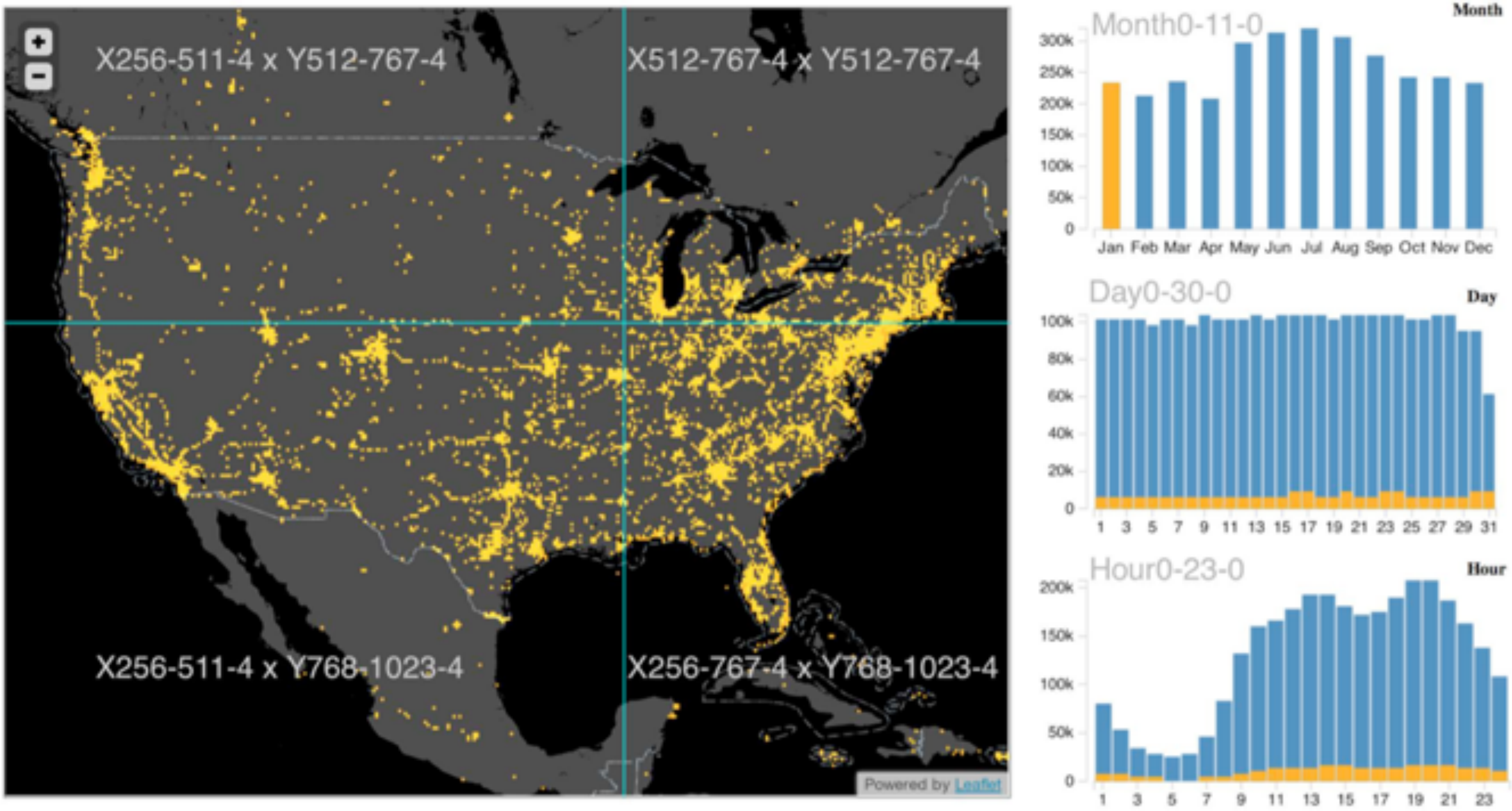
Full 5-D Cube → ~2.3B bins

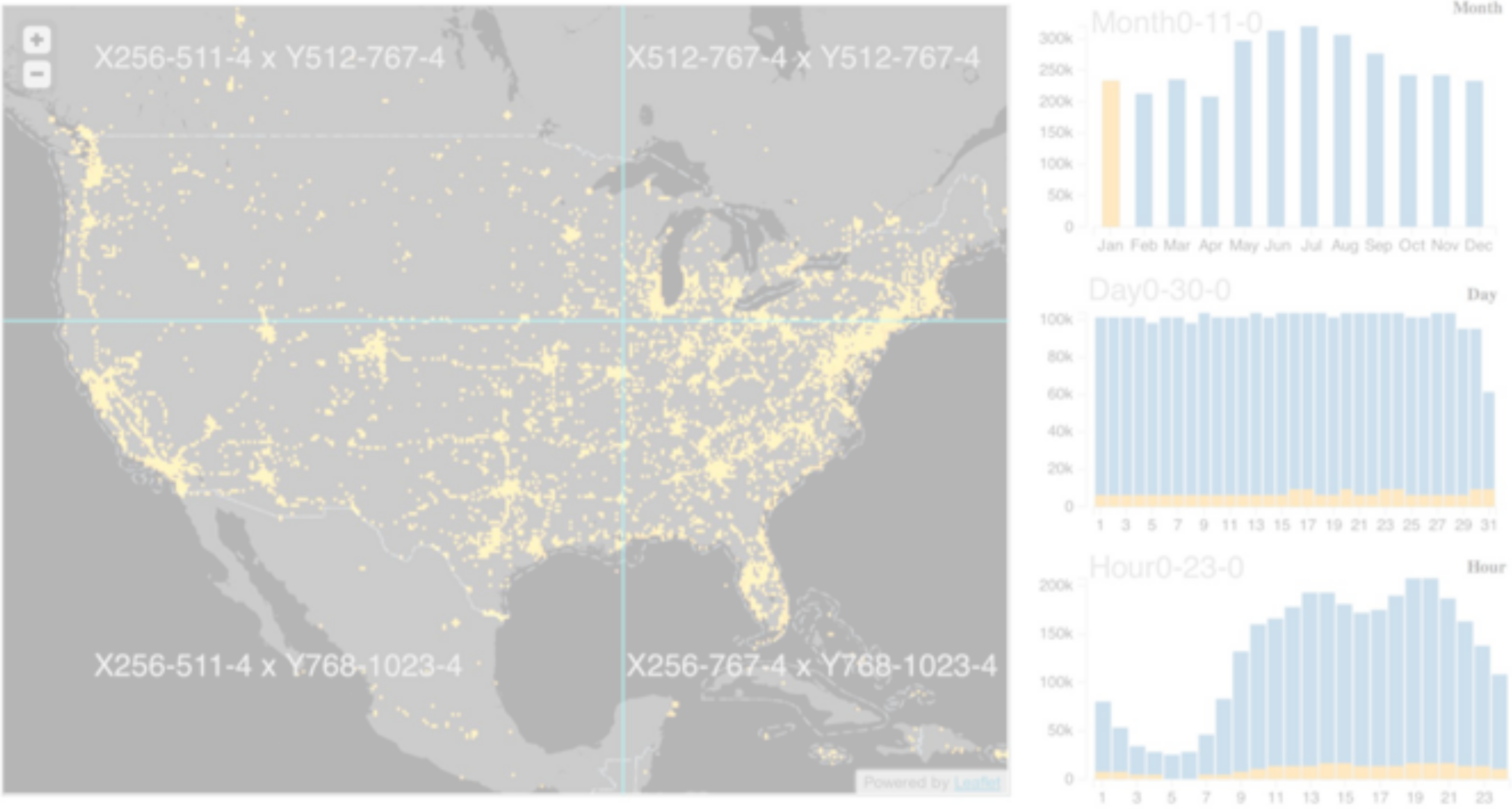


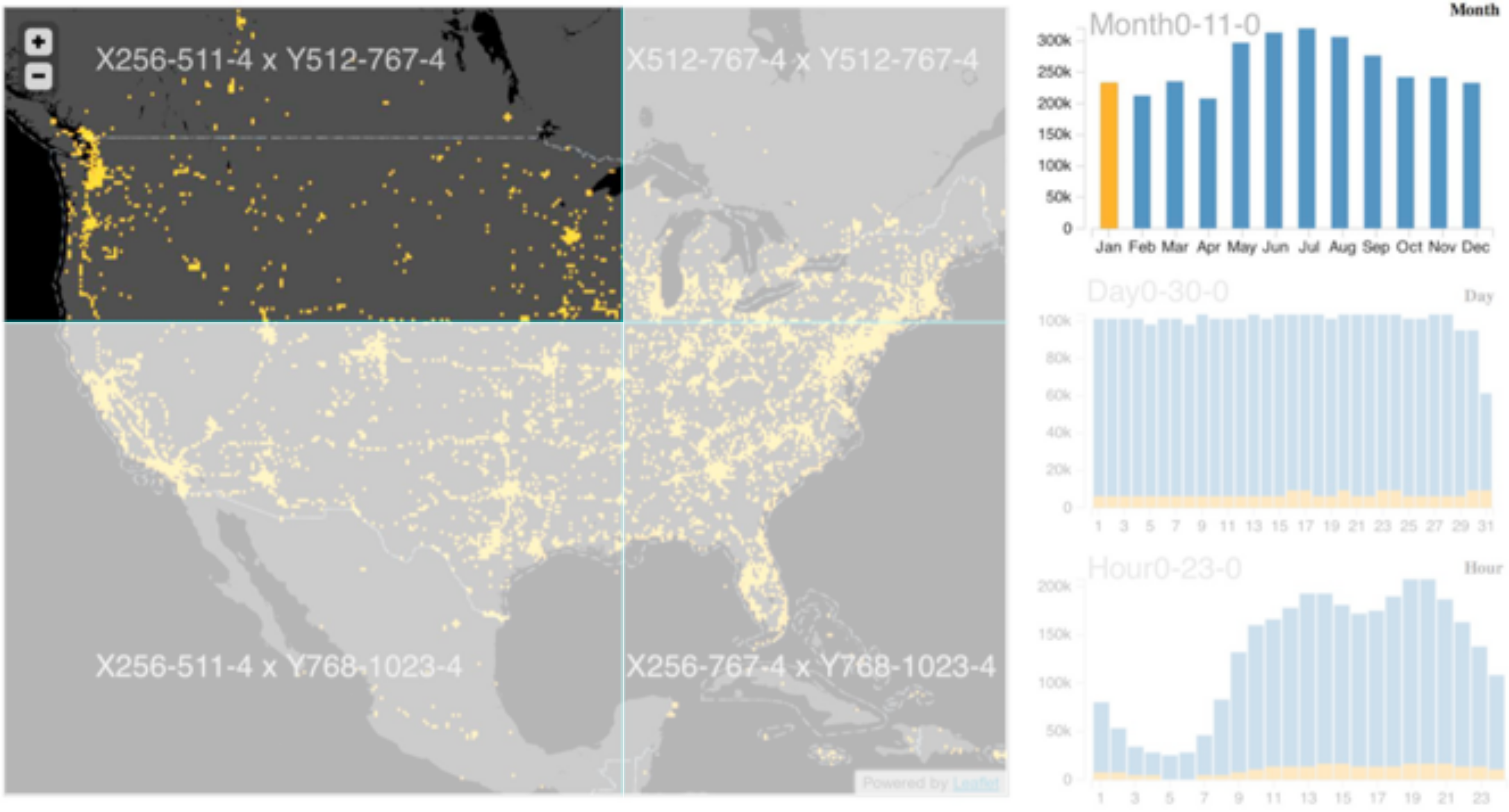
3-D cubes

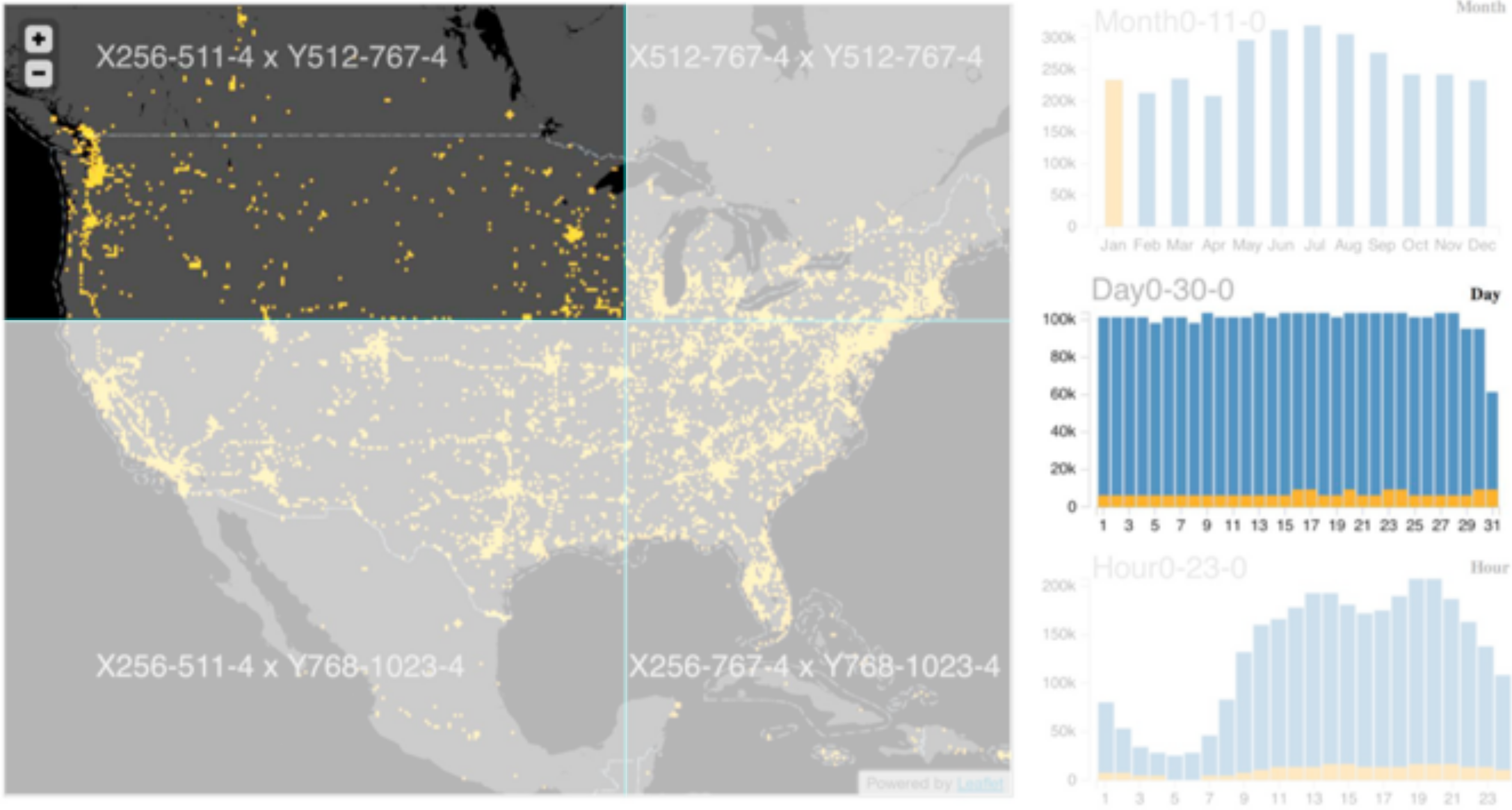
3-D data tiles

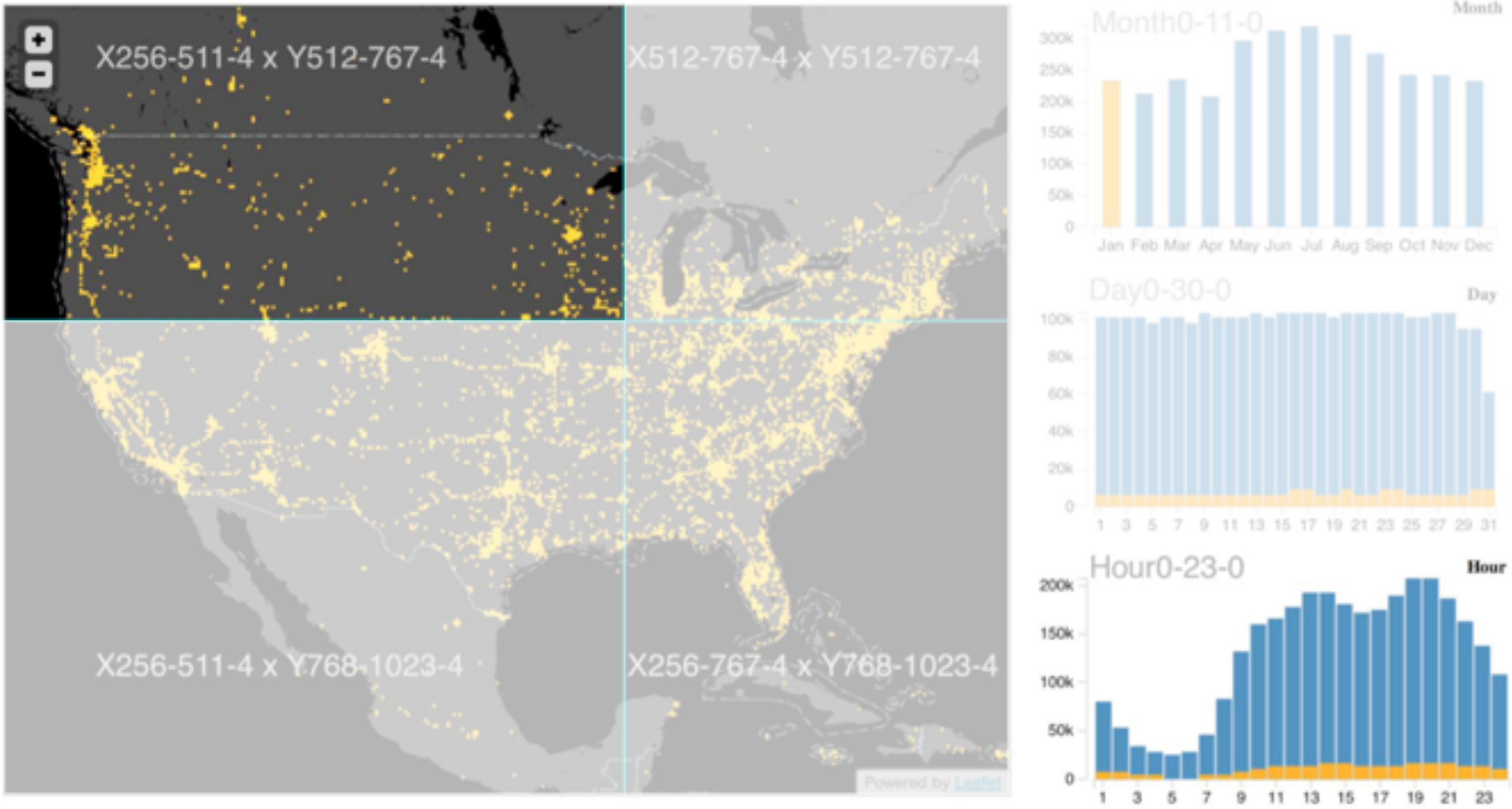
13 3-D Data Tiles → ~17.6M bins (in 352KB!)

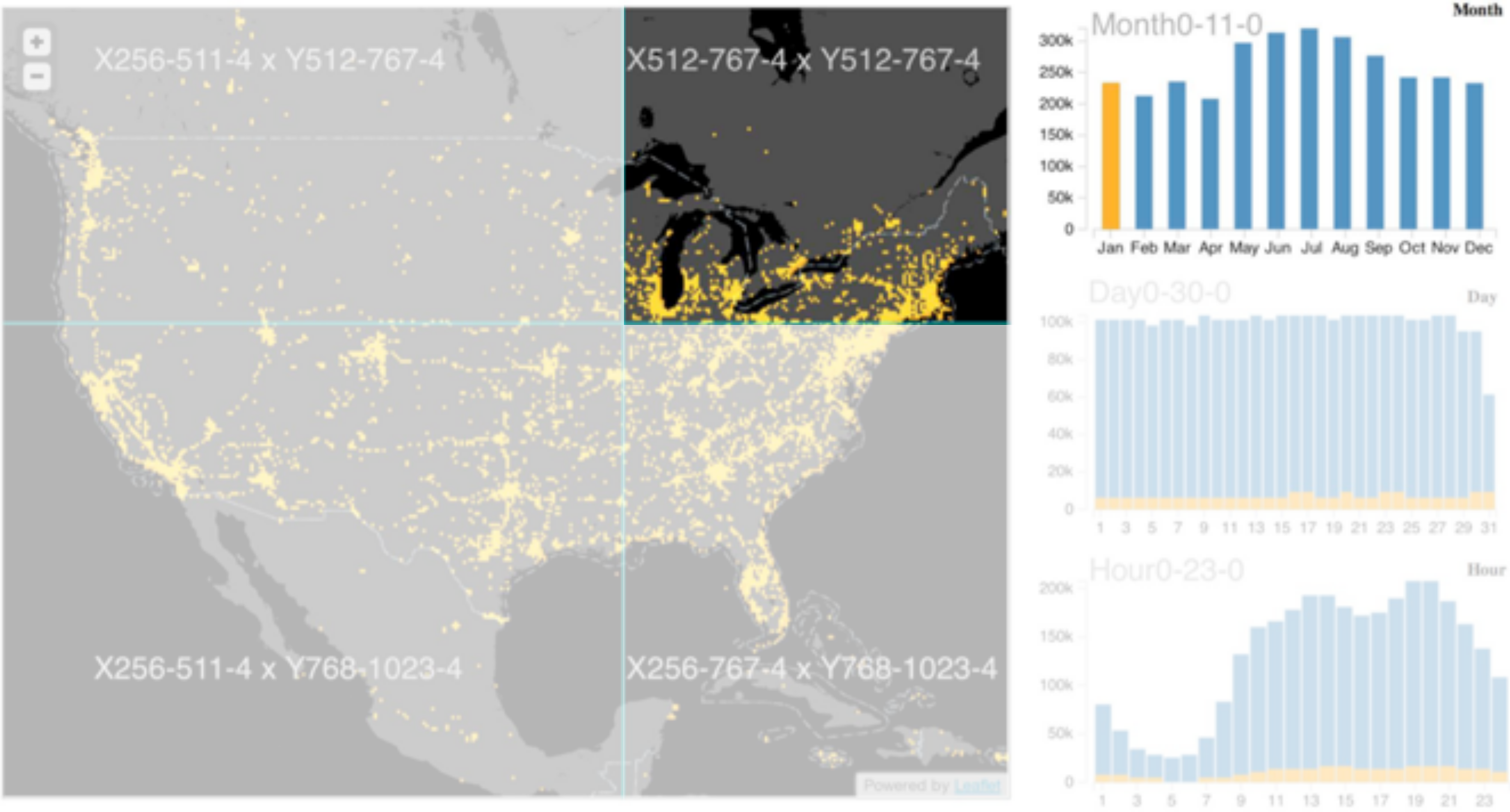

+
−
X256-511-4 x Y512-767-4
X512-767-4 x Y512-767-4
X256-511-4 x Y768-1023-4
X256-767-4 x Y768-1023-4
Powered by Leaflet
Month0-11-0
Month
300k
250k
200k
150k
100k
50k
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Day0-30-0
Day
100k
80k
60k
40k
20k
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31
Hour0-23-0
Hour
200k
150k
100k
50k
0
1 3 5 7 9 11 13 15 17 19 21 23

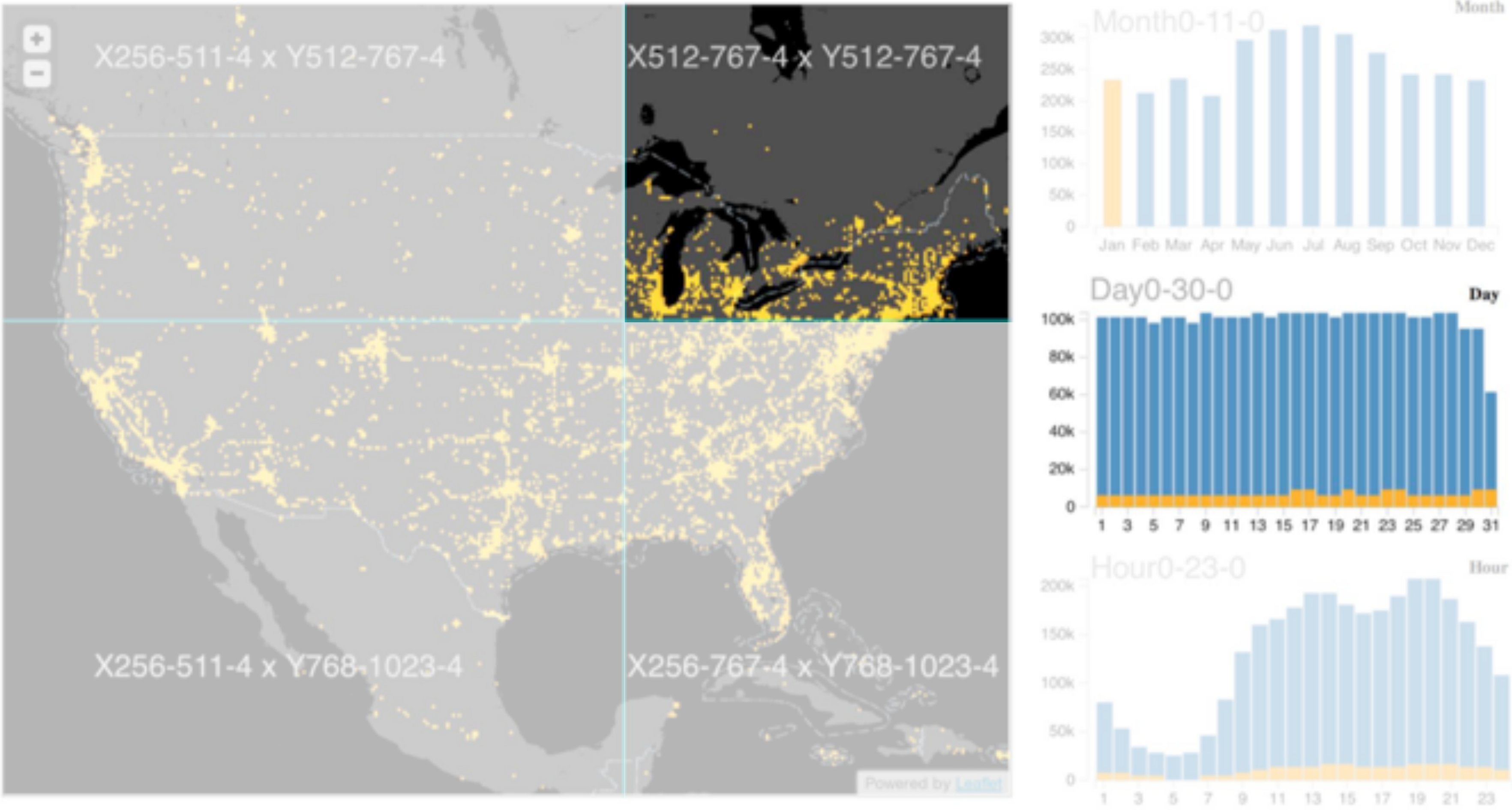

+
−
X256-511-4 x Y512-767-4
X512-767-4 x Y512-767-4
X256-511-4 x Y768-1023-4
X256-767-4 x Y768-1023-4
Powered by Leaflet
Month0-11-0
Month
300k
250k
200k
150k
100k
50k
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Day0-30-0
Day
100k
80k
60k
40k
20k
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31
Hour0-23-0
Hour
200k
150k
100k
50k
0
1 3 5 7 9 11 13 15 17 19 21 23

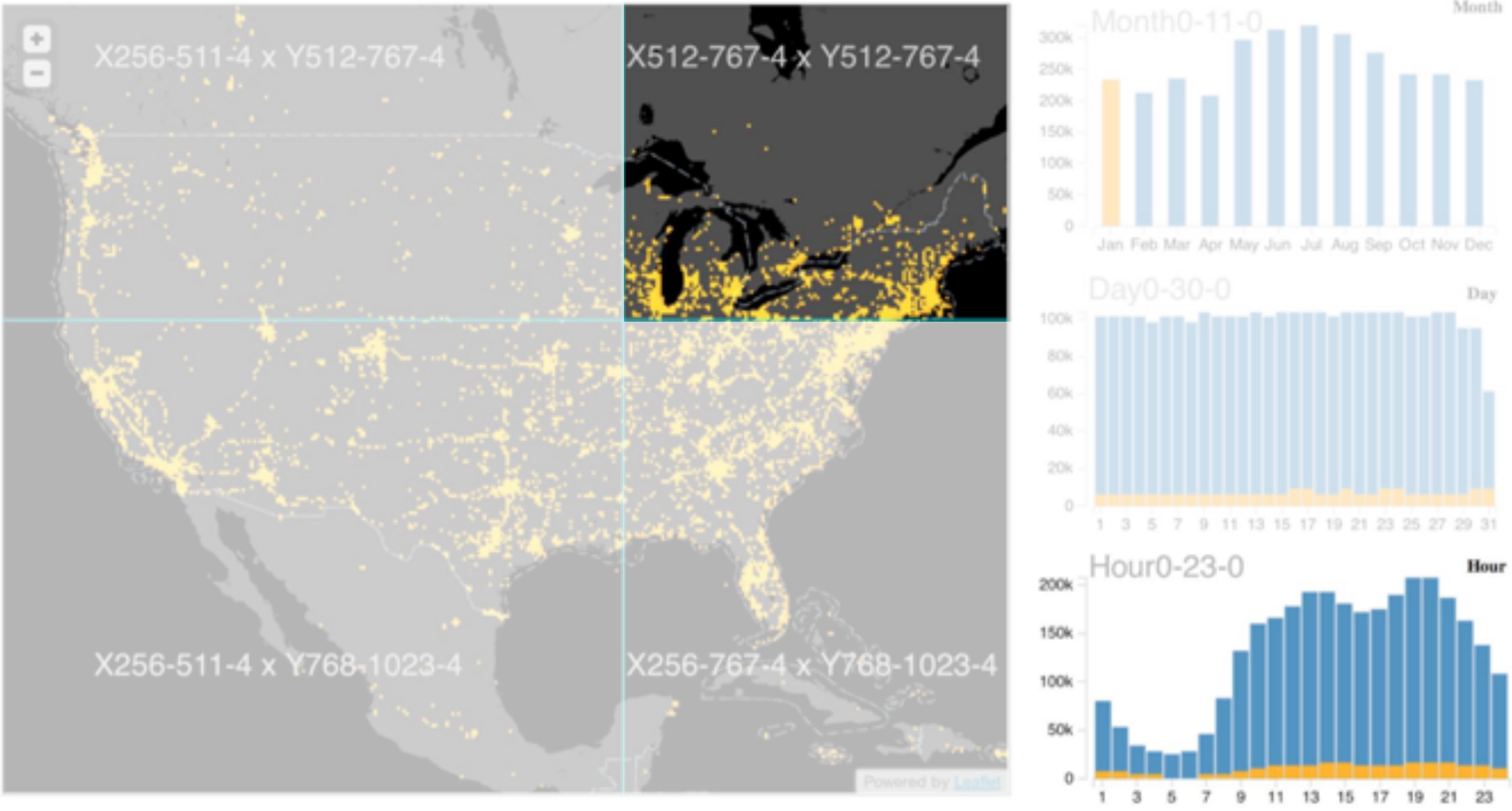

X256-511-4 x Y512-767-4
X512-767-4 x Y512-767-4
X256-511-4 x Y768-1023-4
X256-767-4 x Y768-1023-4
Powered by Leaflet
Month0-11-0
Month
300k
250k
200k
150k
100k
50k
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Day0-30-0
Day
100k
80k
60k
40k
20k
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31
Hour0-23-0
Hour
200k
150k
100k
50k
0
1 3 5 7 9 11 13 15 17 19 21 23

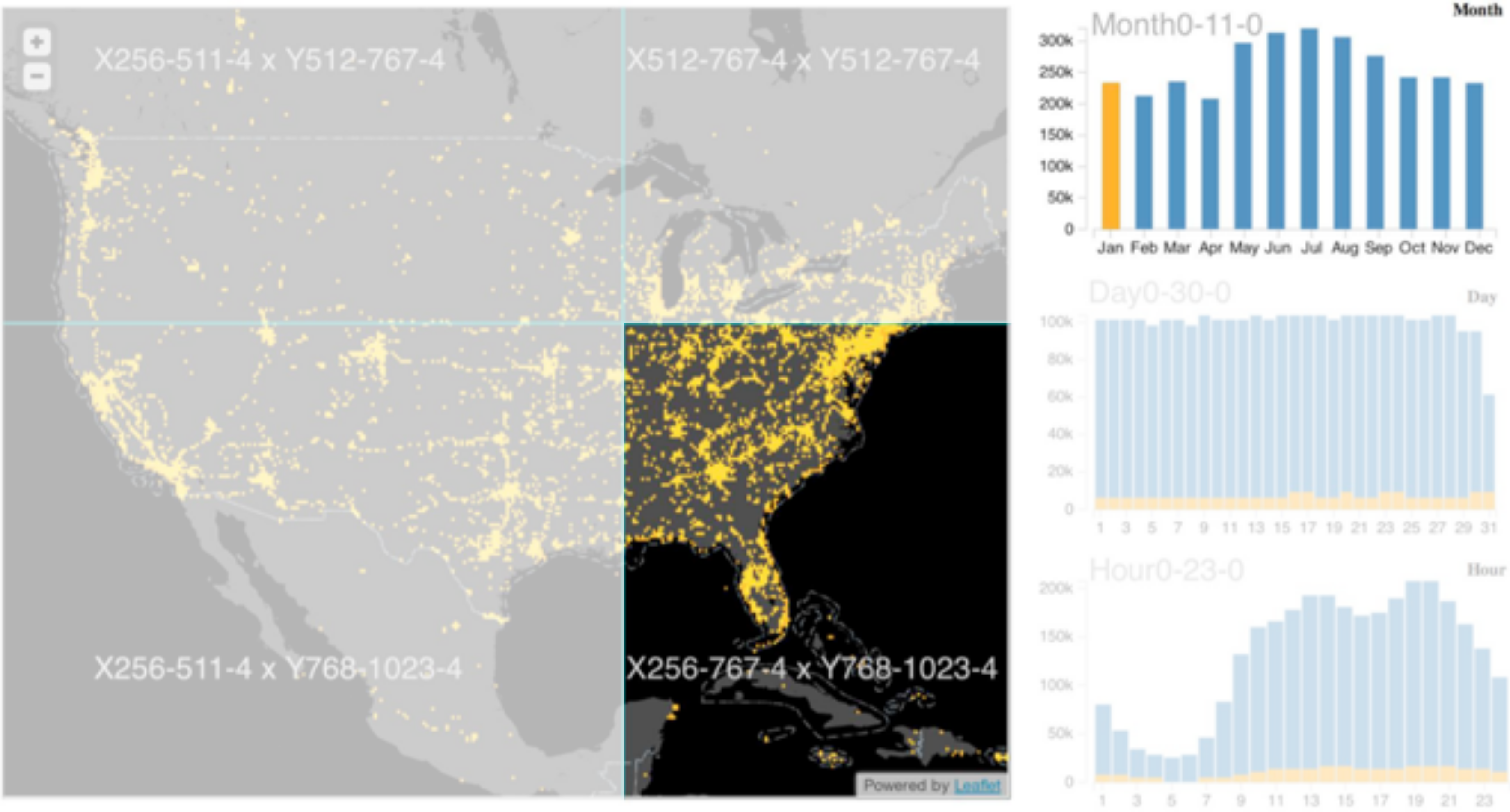

+
−
X256-511-4 x Y512-767-4
X512-767-4 x Y512-767-4
X256-511-4 x Y768-1023-4
X256-767-4 x Y768-1023-4
Powered by Leaflet
Month0-11-0
Month
300k
250k
200k
150k
100k
50k
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Day0-30-0
Day
100k
80k
60k
40k
20k
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31
Hour0-23-0
Hour
200k
150k
100k
50k
0
1 3 5 7 9 11 13 15 17 19 21 23

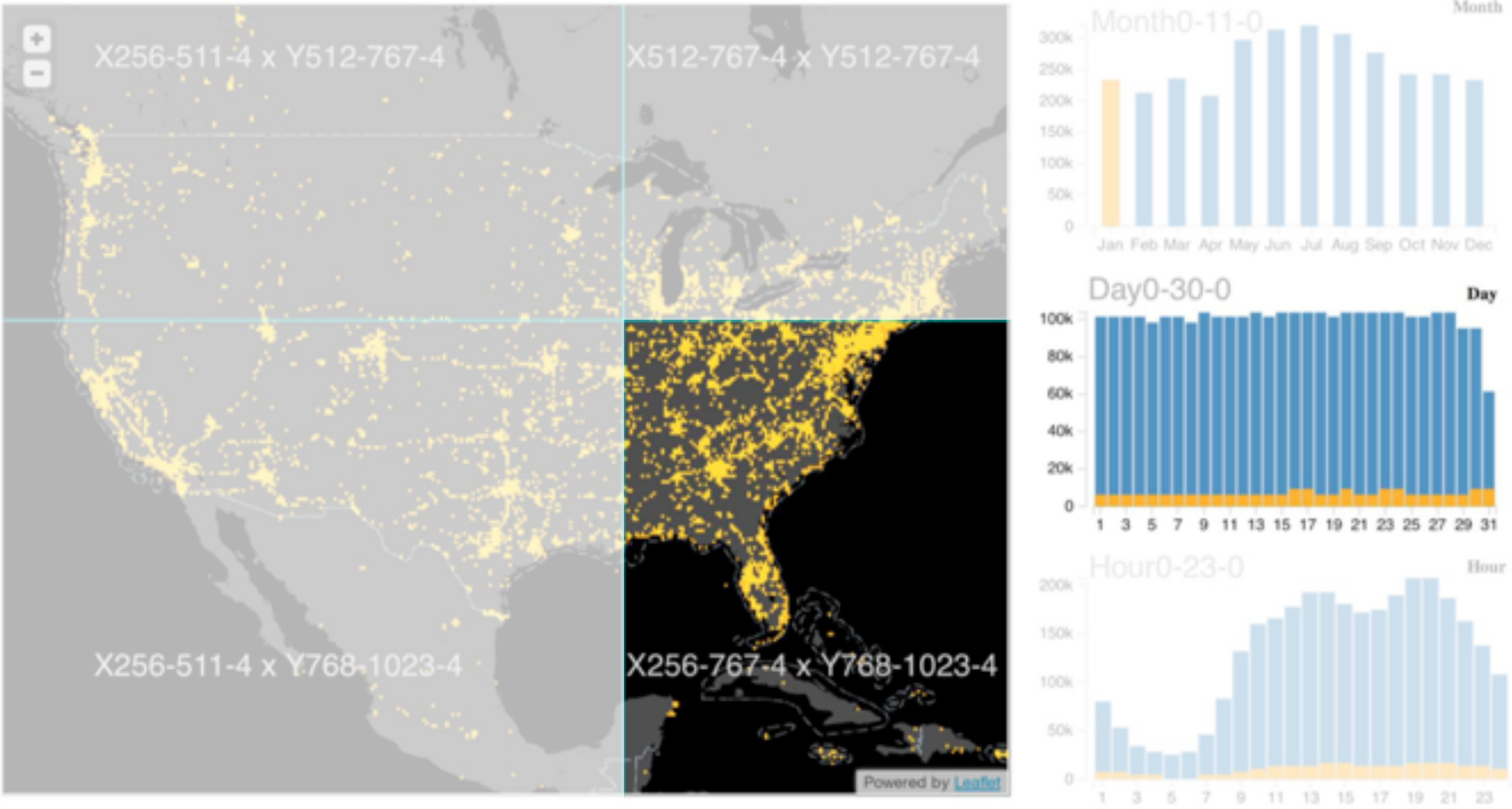

+
−
X256-511-4 x Y512-767-4
X512-767-4 x Y512-767-4
X256-511-4 x Y768-1023-4
X256-767-4 x Y768-1023-4
Powered by Leaflet
Month0-11-0
Month
300k
250k
200k
150k
100k
50k
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Day0-30-0
Day
100k
80k
60k
40k
20k
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31
Hour0-23-0
Hour
200k
150k
100k
50k
0
1 3 5 7 9 11 13 15 17 19 21 23

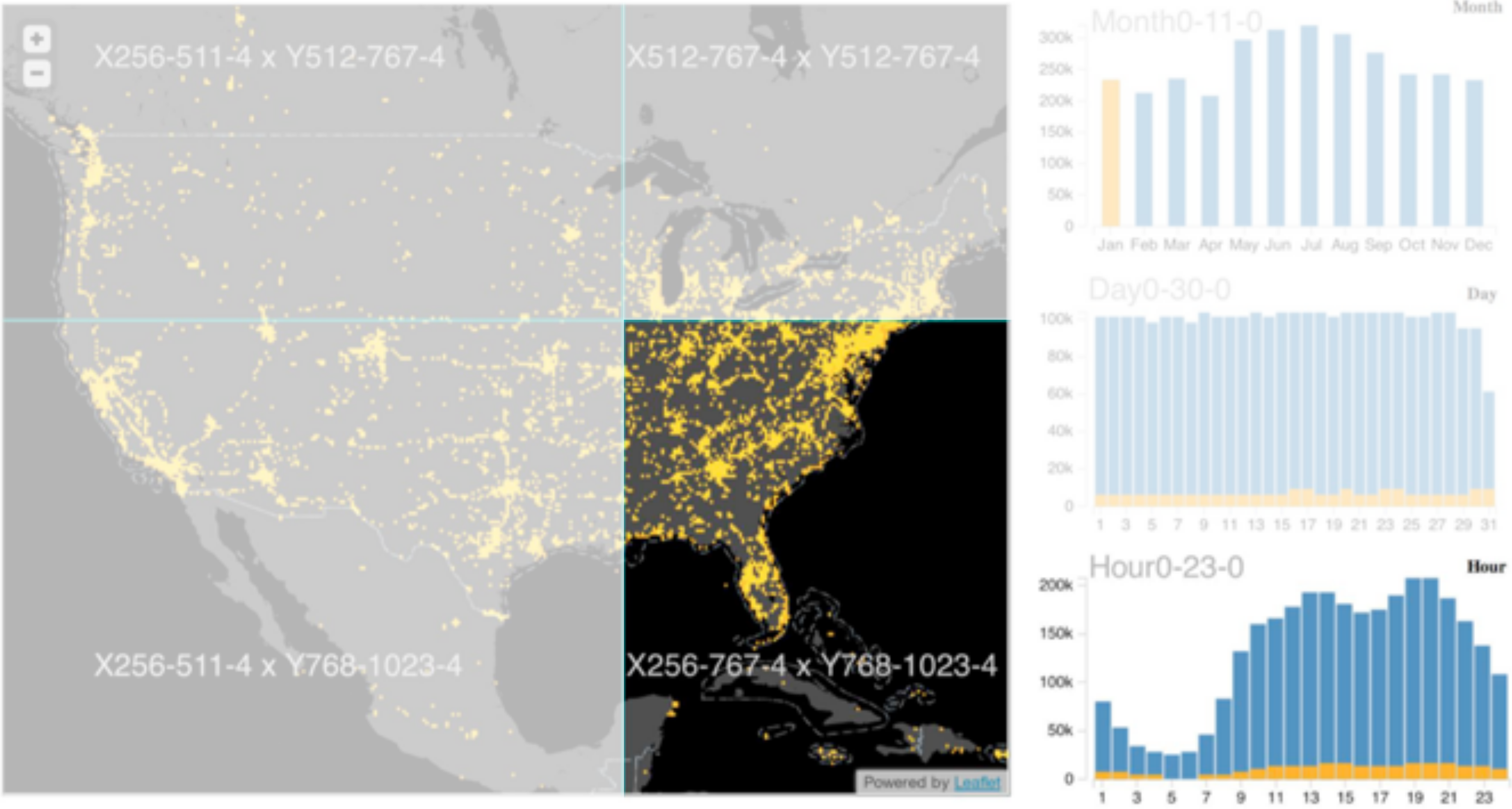

+
−
X256-511-4 x Y512-767-4
X512-767-4 x Y512-767-4
X256-511-4 x Y768-1023-4
X256-767-4 x Y768-1023-4
Powered by Leaflet
Month0-11-0
Month
300k
250k
200k
150k
100k
50k
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Day0-30-0
Day
100k
80k
60k
40k
20k
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31
Hour0-23-0
Hour
200k
150k
100k
50k
0
1 3 5 7 9 11 13 15 17 19 21 23

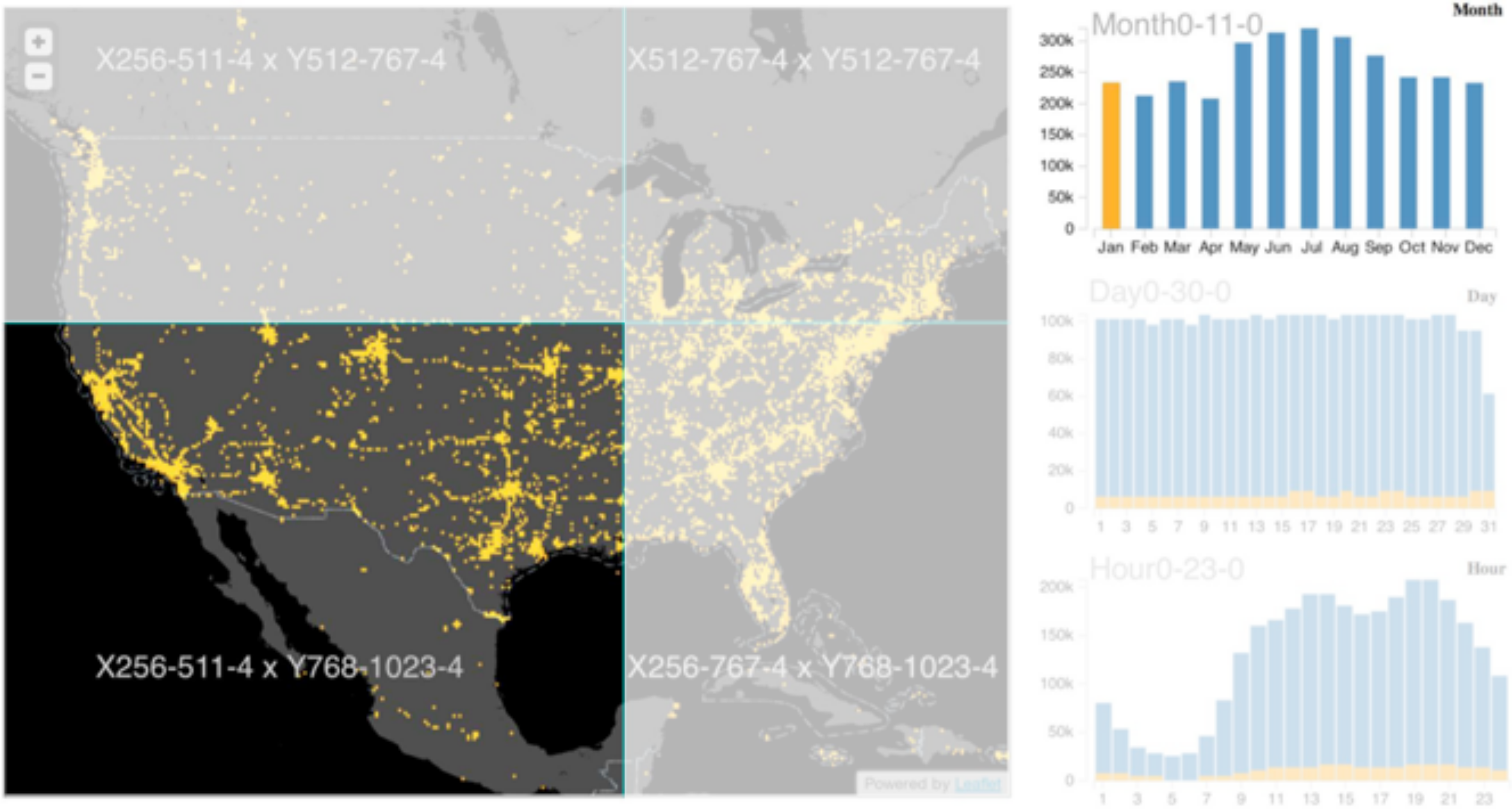

X256-511-4 x Y512-767-4
X512-767-4 x Y512-767-4
X256-511-4 x Y768-1023-4
X256-767-4 x Y768-1023-4
Powered by Leaflet
Month0-11-0
Month
300k
250k
200k
150k
100k
50k
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Day0-30-0
Day
100k
80k
60k
40k
20k
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31
Hour0-23-0
Hour
200k
150k
100k
50k
0
1 3 5 7 9 11 13 15 17 19 21 23

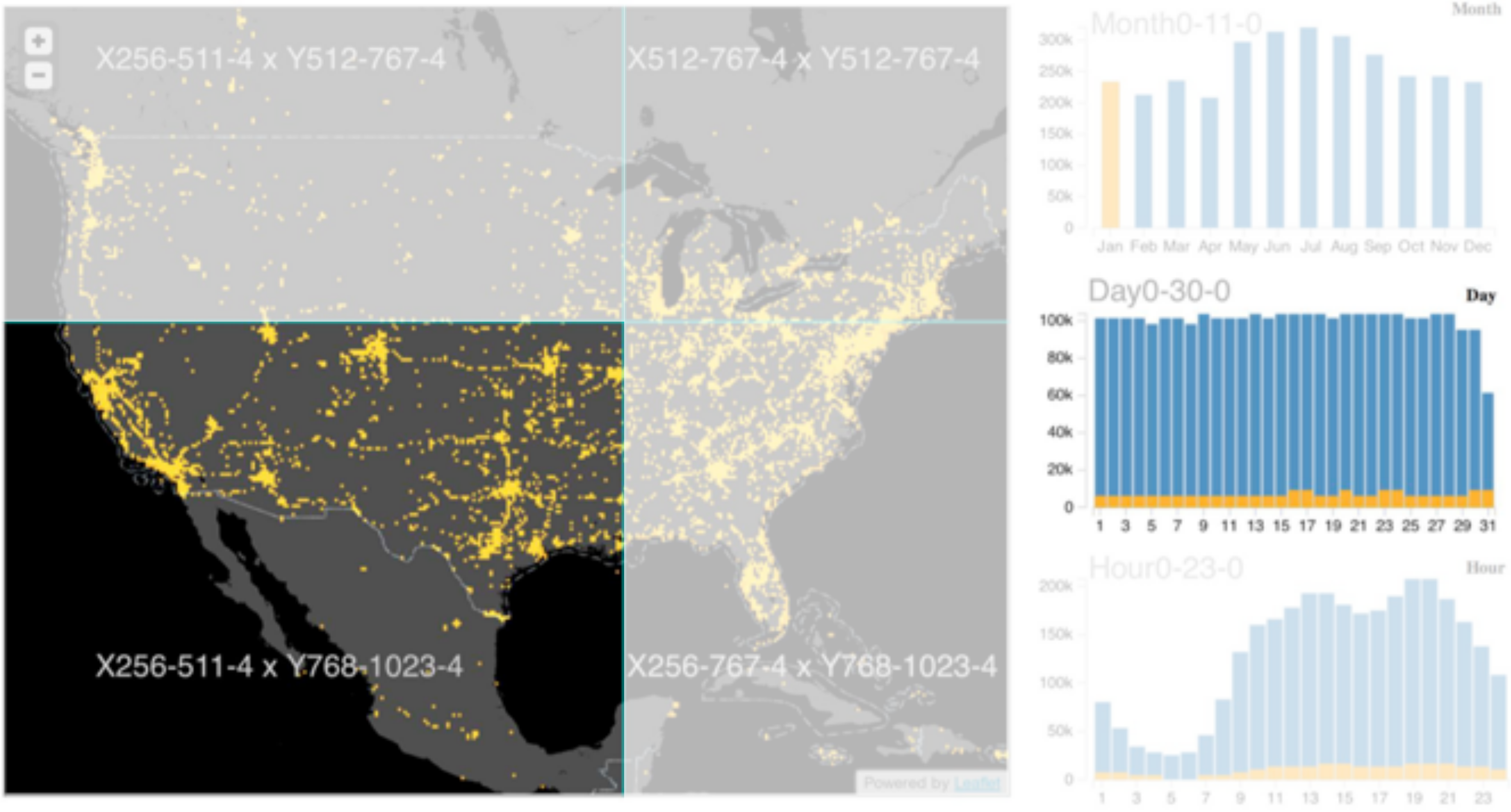

+
−
X256-511-4 x Y512-767-4
X512-767-4 x Y512-767-4
X256-511-4 x Y768-1023-4
X256-767-4 x Y768-1023-4
Powered by Leaflet
Month0-11-0
Month
300k
250k
200k
150k
100k
50k
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Day0-30-0
Day
100k
80k
60k
40k
20k
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31
Hour0-23-0
Hour
200k
150k
100k
50k
0
1 3 5 7 9 11 13 15 17 19 21 23


X256-511-4 x Y512-767-4
X512-767-4 x Y512-767-4
X256-511-4 x Y768-1023-4
X256-767-4 x Y768-1023-4
Powered by Leaflet
Month0-11-0
Month
300k
250k
200k
150k
100k
50k
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Day0-30-0
Day
100k
80k
60k
40k
20k
0
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
31
Hour0-23-0
Hour
200k
150k
100k
50k
0
1
3
5
7
9
11
13
15
17
19
21
23

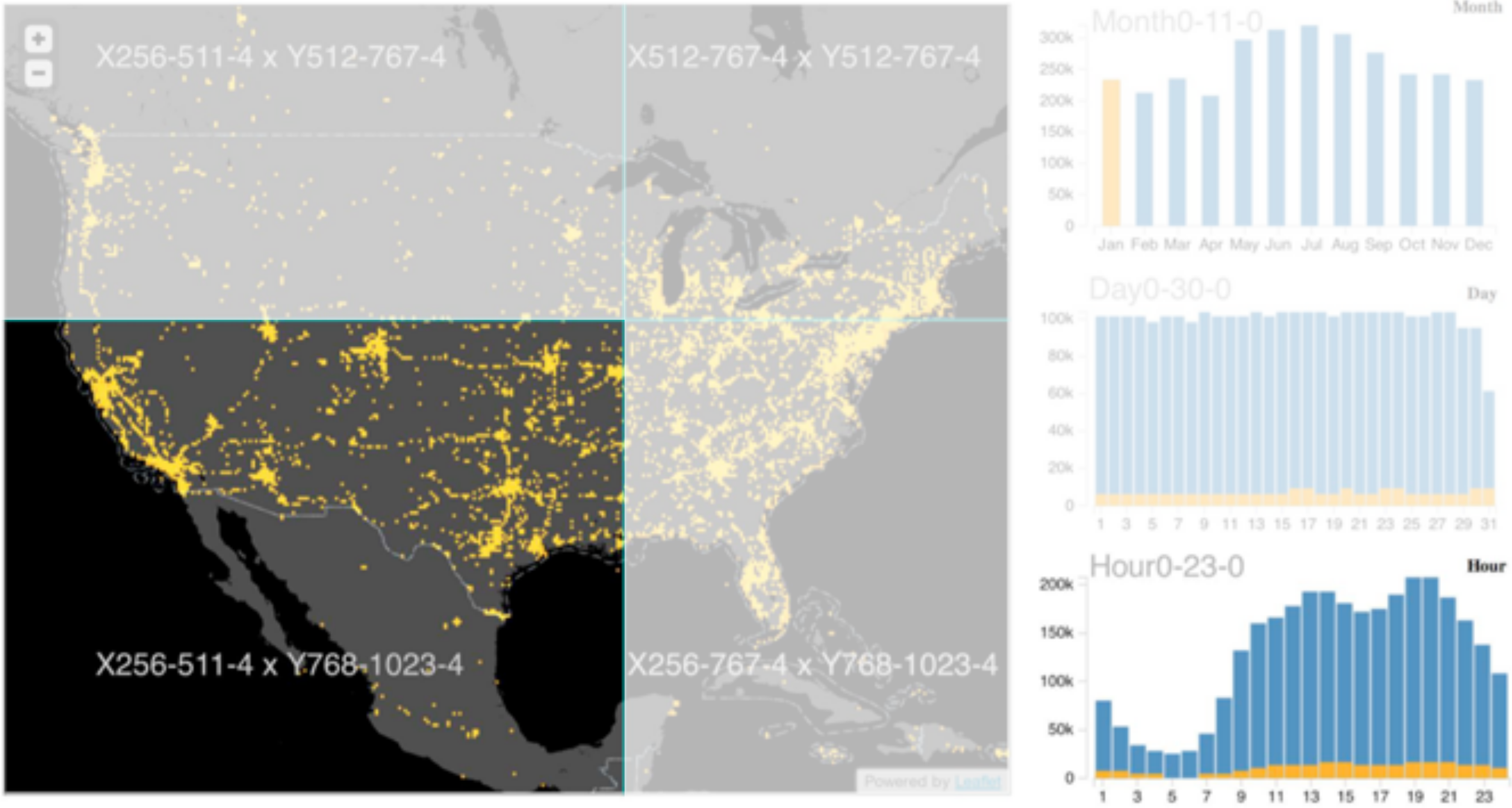

X256-511-4 x Y512-767-4
X512-767-4 x Y512-767-4
X256-511-4 x Y768-1023-4
X256-767-4 x Y768-1023-4
Powered by Leaflet
Month0-11-0
Month
300k
250k
200k
150k
100k
50k
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Day0-30-0
Day
100k
80k
60k
40k
20k
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31
Hour0-23-0
Hour
200k
150k
100k
50k
0
1 3 5 7 9 11 13 15 17 19 21 23

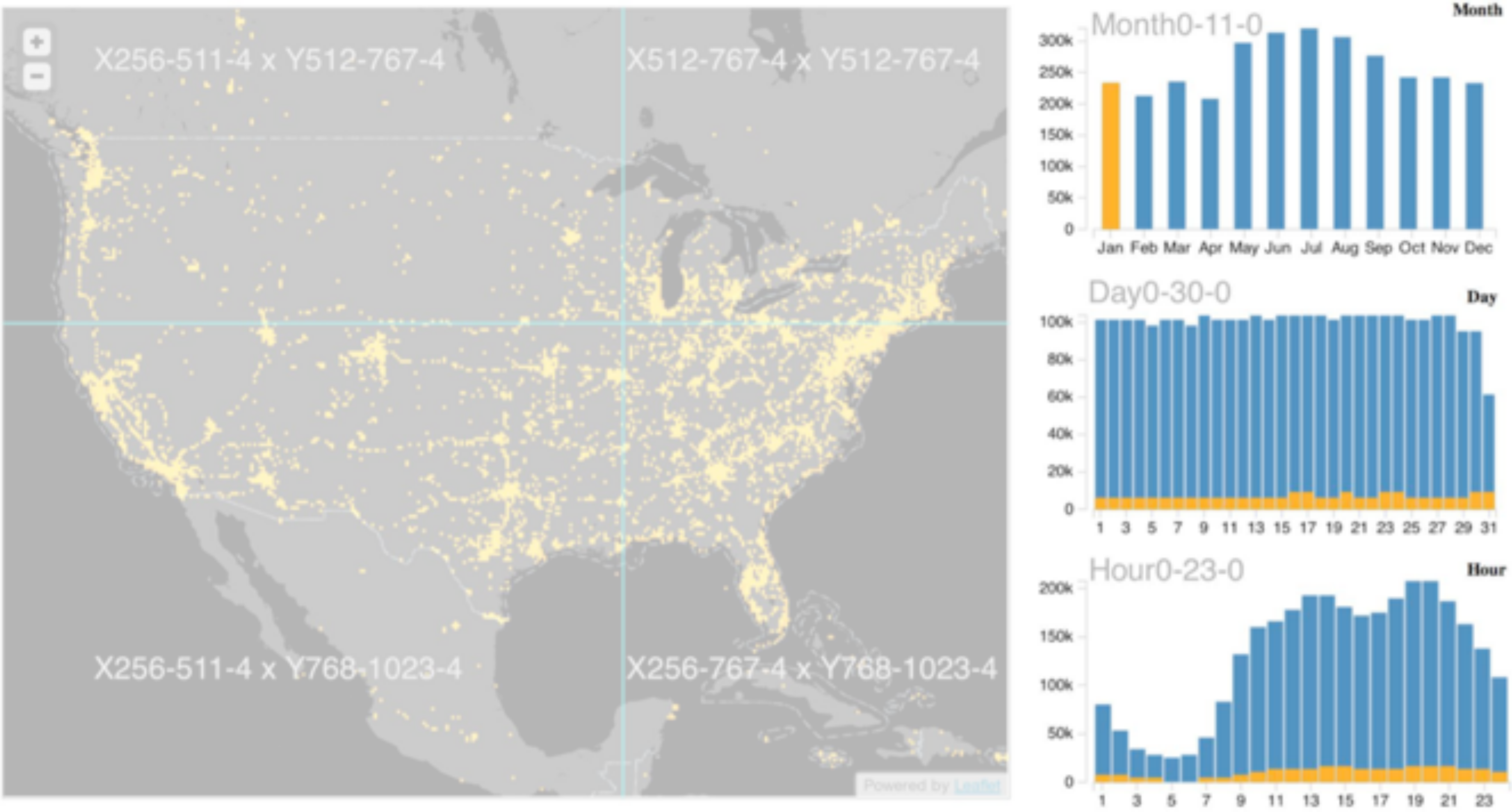

X256-511-4 x Y512-767-4
X512-767-4 x Y512-767-4
X256-511-4 x Y768-1023-4
X256-767-4 x Y768-1023-4
Powered by Leaflet
Month0-11-0
Month
300k
250k
200k
150k
100k
50k
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Day0-30-0
Day
100k
80k
60k
40k
20k
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31
Hour0-23-0
Hour
200k
150k
100k
50k
0
1 3 5 7 9 11 13 15 17 19 21 23

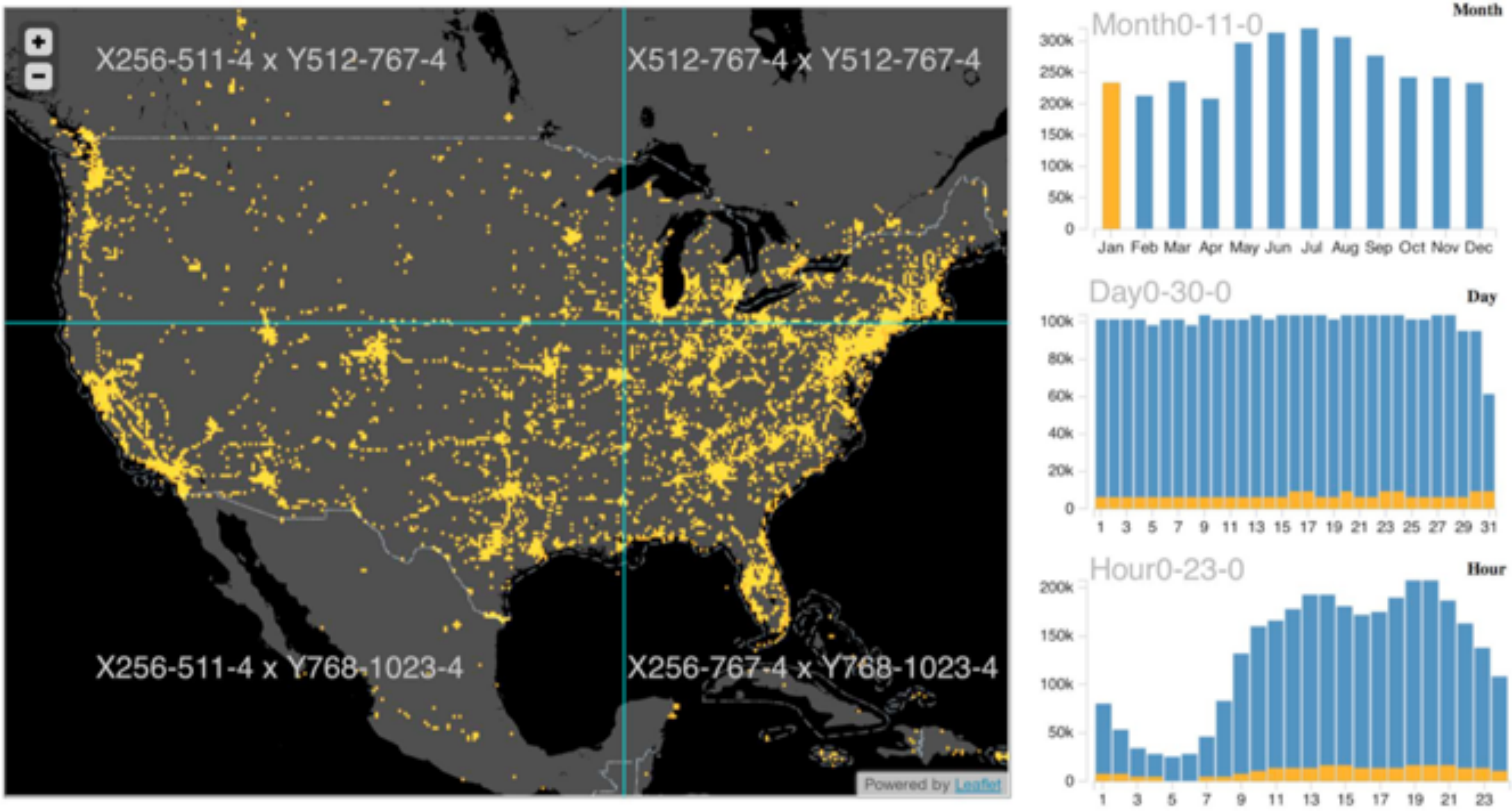

+
−
X256-511-4 x Y512-767-4
X512-767-4 x Y512-767-4
X256-511-4 x Y768-1023-4
X256-767-4 x Y768-1023-4
Powered by Leaflet
Month0-11-0
Month
300k
250k
200k
150k
100k
50k
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Day0-30-0
Day
100k
80k
60k
40k
20k
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31
Hour0-23-0
Hour
200k
150k
100k
50k
0
1 3 5 7 9 11 13 15 17 19 21 23



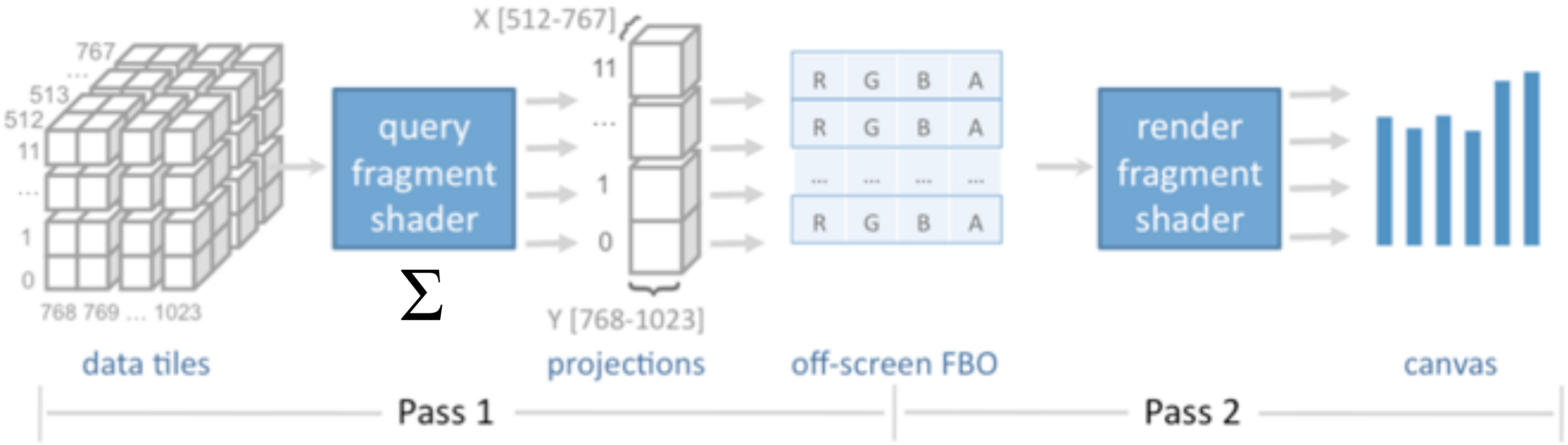
X256-511-4 x Y512-767-4
X512-767-4 x Y512-767-4
X256-511-4 x Y768-1023-4
X256-767-4 x Y768-1023-4
Powered by Leaflet
Month0-11-0
Month
0
50k
100k
150k
200k
250k
300k
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Day0-30-0
Day
0
20k
40k
60k
80k
100k
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
31
Hour0-23-0
Hour
0
50k
100k
150k
200k
1
3
5
7
9
11
13
15
17
19
21
23


X256-511-4 x Y512-767-4
X512-767-4 x Y512-767-4
X256-511-4 x Y768-1023-4
X256-767-4 x Y768-1023-4
Powered by Leaflet
Month0-11-0
Month
300k
250k
200k
150k
100k
50k
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Day0-30-0
Day
100k
80k
60k
40k
20k
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31
Hour0-23-0
Hour
200k
150k
100k
50k
0
1 3 5 7 9 11 13 15 17 19 21 23


X256-511-4 x Y512-767-4
X512-767-4 x Y512-767-4
X256-511-4 x Y768-1023-4
X256-767-4 x Y768-1023-4
Powered by Leaflet
Month0-11-0
Month
300k
250k
200k
150k
100k
50k
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Day0-30-0
Day
100k
80k
60k
40k
20k
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31
Hour0-23-0
Hour
200k
150k
100k
50k
0
1 3 5 7 9 11 13 15 17 19 21 23

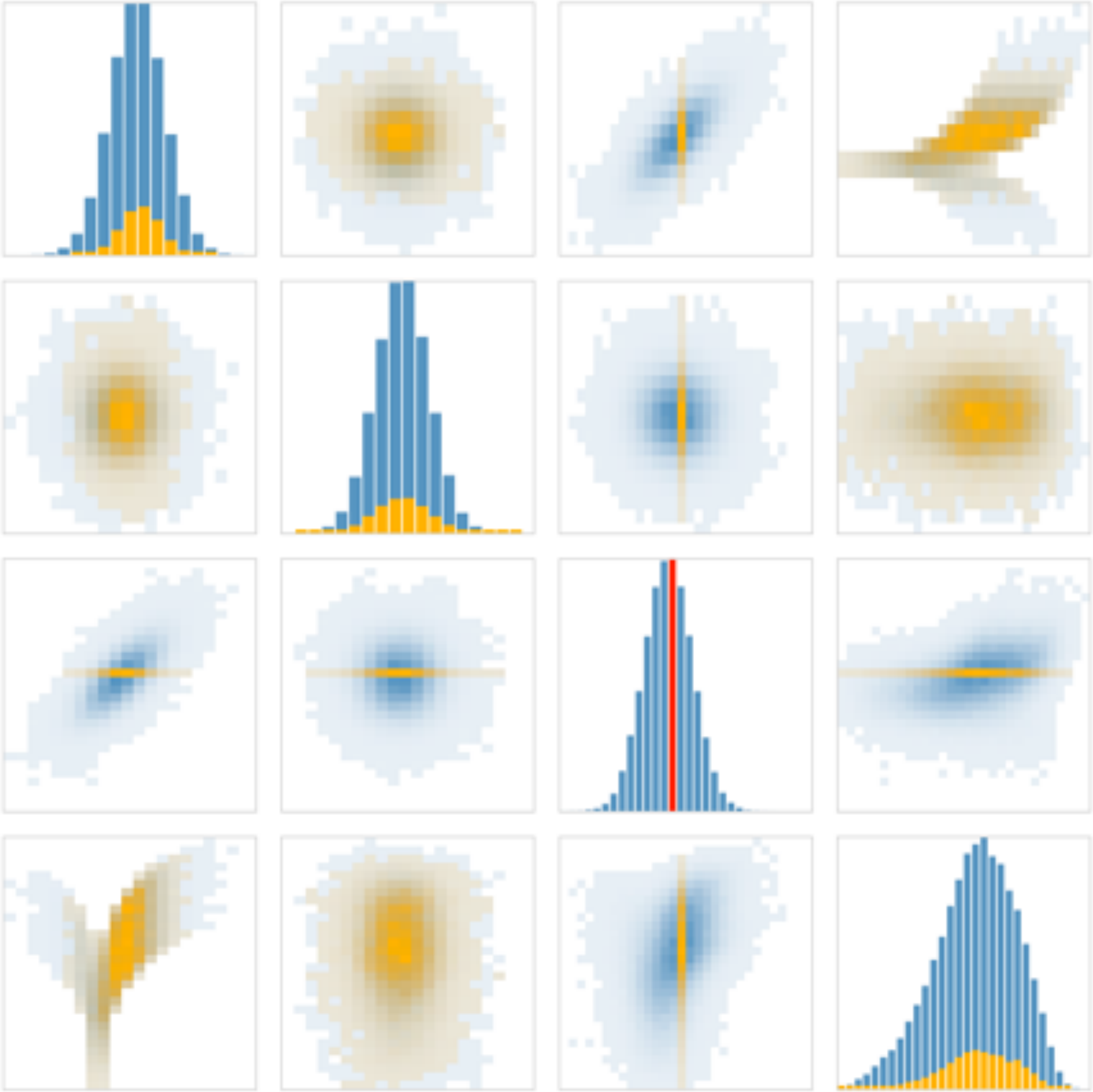

+
−
X256-511-4 x Y512-767-4
X512-767-4 x Y512-767-4
X256-511-4 x Y768-1023-4
X256-767-4 x Y768-1023-4
Powered by Leaflet
Month0-11-0
Month
0
50k
100k
150k
200k
250k
300k
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Day0-30-0
Day
0
20k
40k
60k
80k
100k
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
31
Hour0-23-0
Hour
0
50k
100k
150k
200k
1
3
5
7
9
11
13
15
17
19
21
23

Query & Render on GPU (WebGL)



Pre-compute tiles & send from server.
Bind data tiles as image textures.
Execute queries in parallel on GPU.

Performance Benchmarks



Simulate interaction: brushing & linking across binned plots.

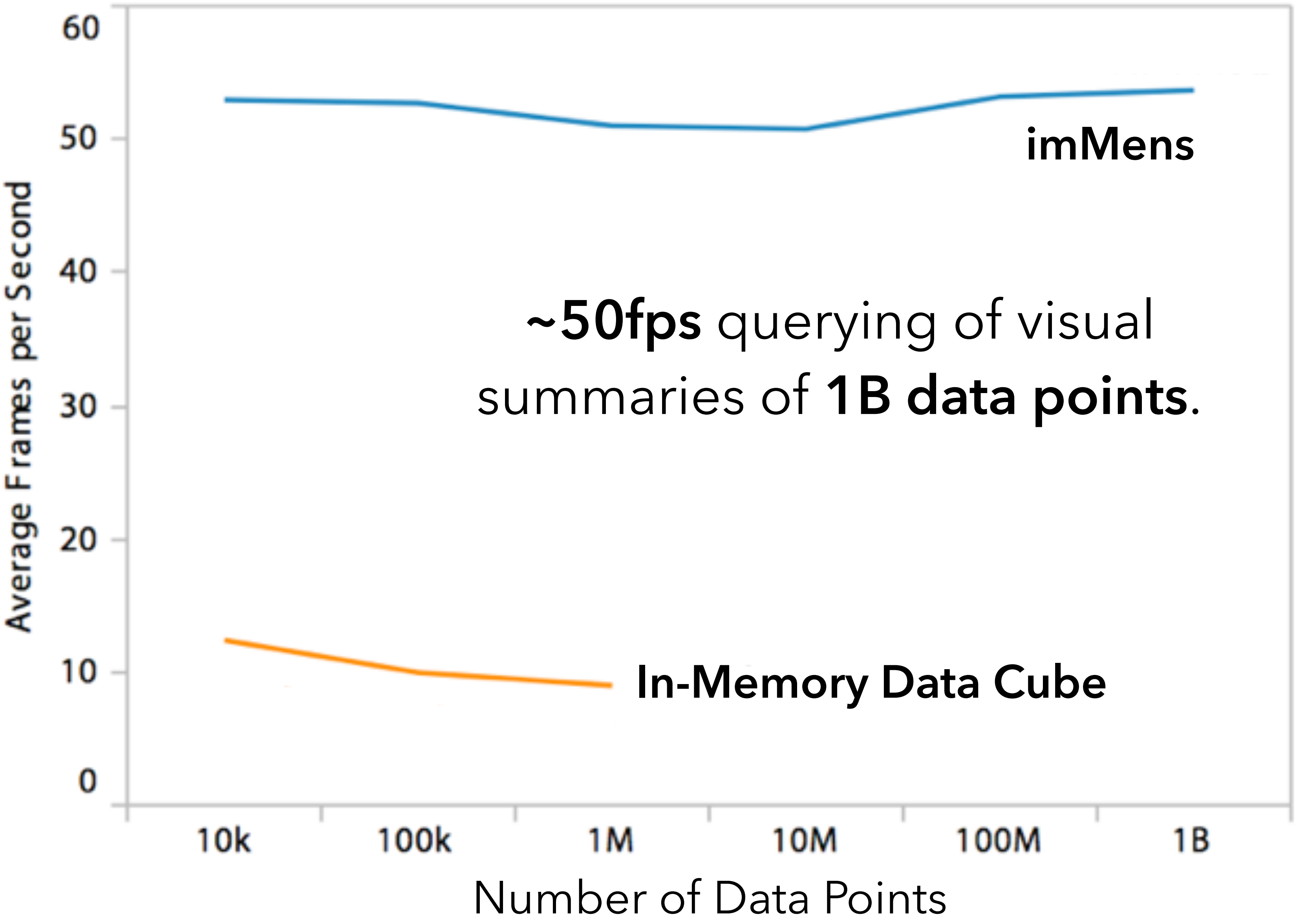
- 4x4 and 5x5 plots
- 10 to 50 bins

Measure time from selection to render.

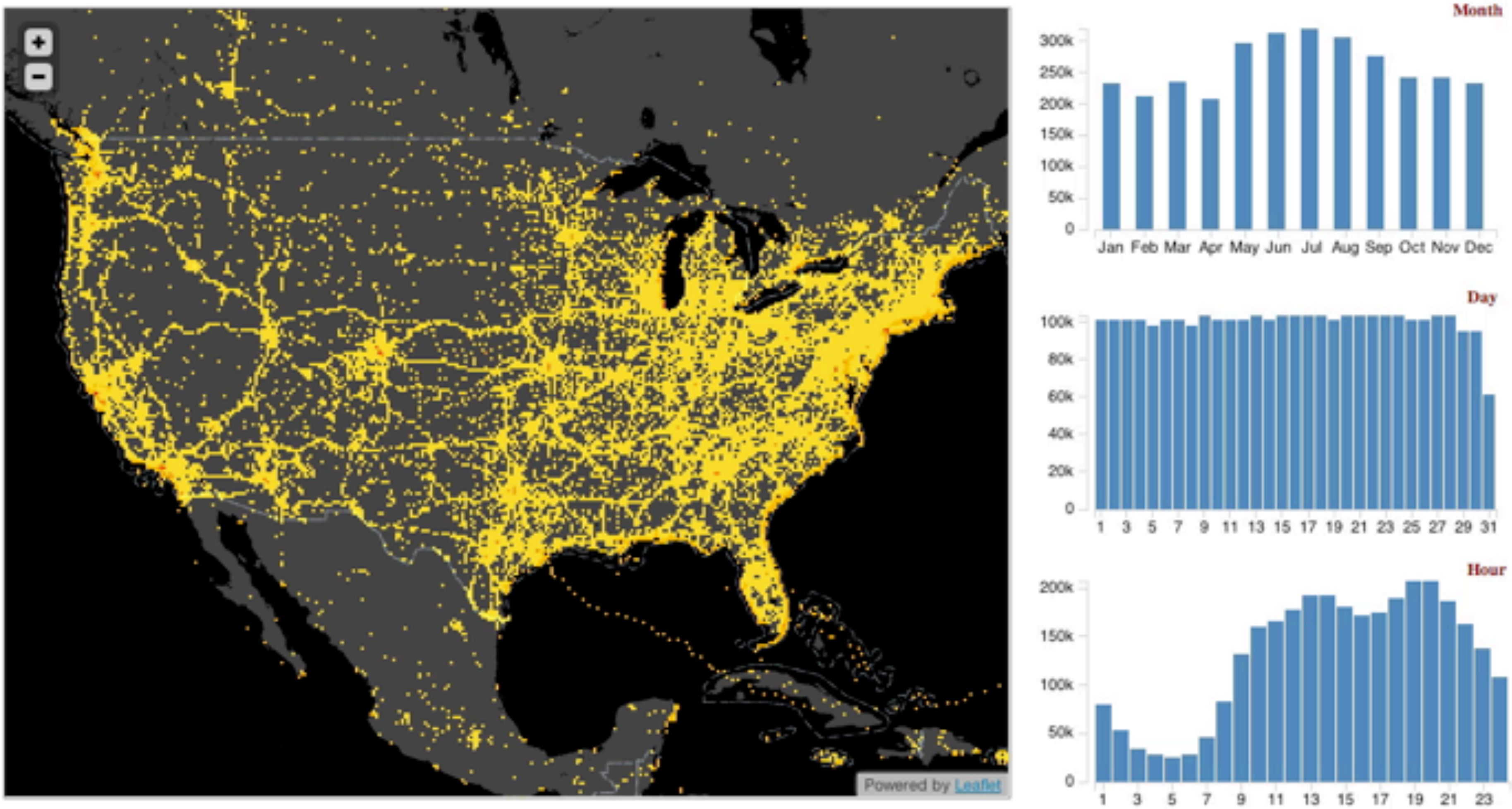
Test setup:

2.3 GHz MacBook Pro
NVIDIA GeForce GT 650M
Google Chrome v.23.0

5 dimensions x 50 bins/dim x 25 plots



~50fps querying of visual summaries of **1B data points**.


+
−
Powered by Leaflet
Month
300k
250k
200k
150k
100k
50k
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Day
100k
80k
60k
40k
20k
0
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
31
Hour
200k
150k
100k
50k
0
1
3
5
7
9
11
13
15
17
19
21
23

Visualizing Big Data

Acknowledgments

Zhicheng “Leo” Liu, Biye Jiang

Sean Kandel, Lars Grammel

Mike Bostock

Maneesh Agrawala, Pat Hanrahan