Toolkits for Visualization and UIs in Data Science

Guest Lecture in Advanced UI Software

Dominik Moritz, April 8, 2020

Hi!
my name is

Dominik Moritz
@domoritz

I build visualization tools

Finished my PhD from UW in 2019

Currently at working on Human Centered ML and Visualization

Starting at CMU HCII in the Fall

dig.cmu.edu

After today, I want you to be able to:

Articulate the difference between chart typologies and component architectures.

Decompose a graphic into the building blocks of visualization.

Know why most visualization toolkits are declarative.

Explain tradeoff between expressiveness and ease-of-use.

Know the history of visualization tools.

Understand the **basics** of D3.

Give examples of manifestations of design principles in Vega-Lite.

Material for this lecture is a combination of:

- Jeffrey Heer’s Tools lecture in CSE 512
- Vega-Lite talk at OpenVis Conf
- My job talk
- Some new slides

Toolkits for Visualization and UIs in Data Science

Guest Lecture in Advanced UI Software

Dominik Moritz, April 8, 2020

Hal Varian, Google’s Chief Economist
The McKinsey Quarterly, Jan 2009

The ability to take data—to be able to **understand** it, to **process** it, to **extract value** from it, to **visualize** it, to **communicate** it—that’s going to be a hugely important skill in the next decades, …
because now we really do have **essentially free and ubiquitous** data. So the complimentary scarce factor is the ability to understand that data and extract value from it.

Four major influences act on data analysis today:

Originally 1965

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

Data Analysis & Statistics. Turkey and Wilk. *1965.*

Effective Data Visualization. Heer. *2015.*

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

Data Analysis & Statistics. Turkey and Wilk. *1965.*

Effective Data Visualization. Heer. *2015.*

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.

Data Analysis & Statistics. Turkey and Wilk. *1965.*

Effective Data Visualization. Heer. *2015.*

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**.

Data Analysis & Statistics. Turkey and Wilk. *1965.*

Effective Data Visualization. Heer. *2015.*

How do people create visualizations?

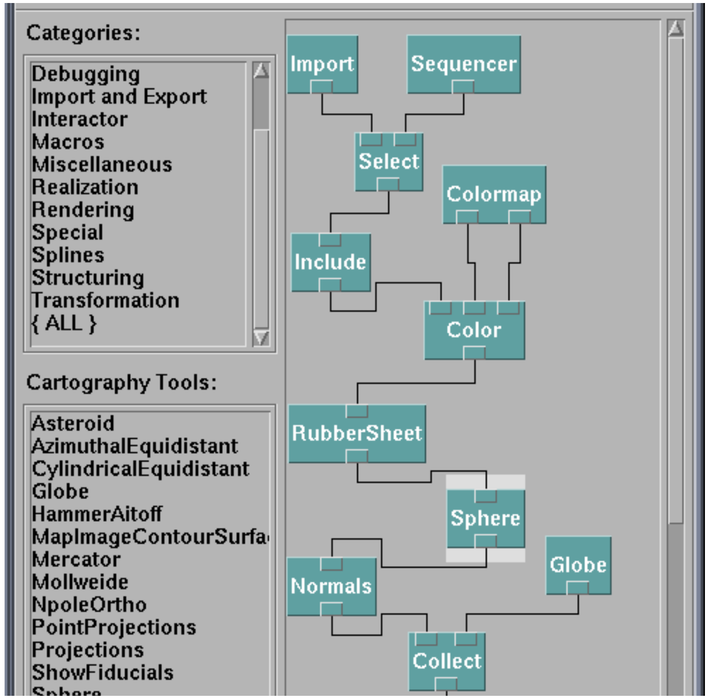


Chart Typology

Pick from a stock of templates

Easy-to-use but limited expressiveness

Prohibits novel designs, new data types

Component Architecture

Permits more combinatorial possibilities

Novel views require new operators, which requires software engineering

Graphics APIs

Processing, OpenGL, Java2D

Drawing Visualizations with Imperative Programs

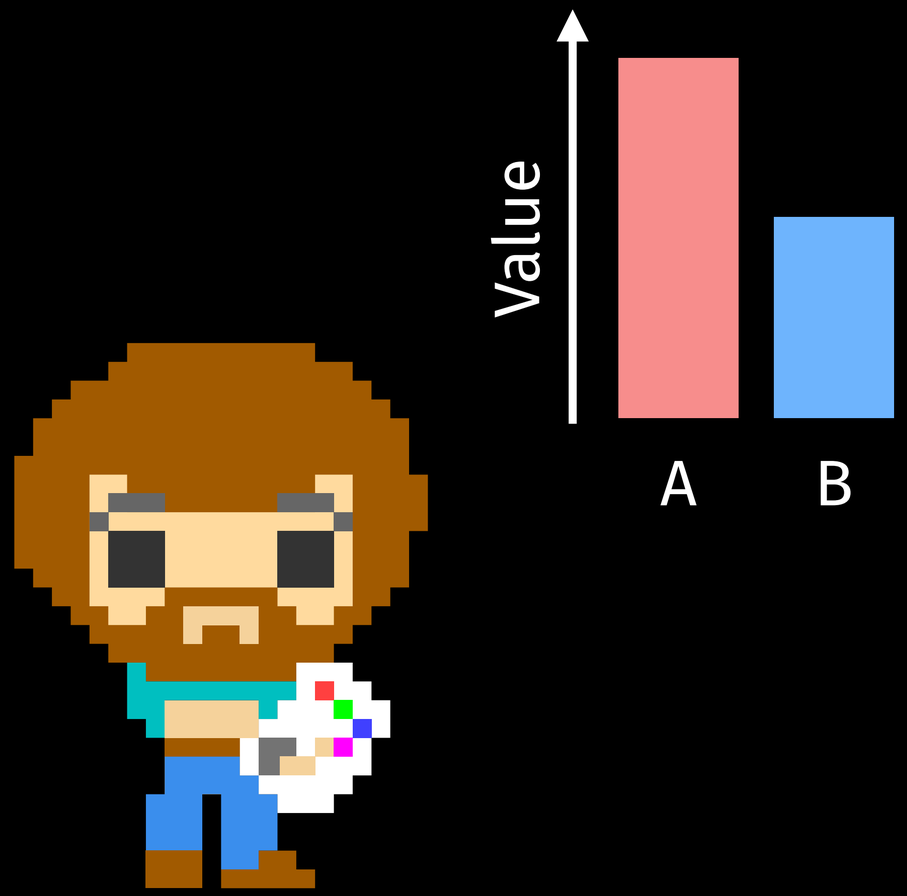
Program by giving explicit steps.

e.g.

"Put a red bar here and a blue bar there."

"Draw a line and some text."

Specification and execution are intertwined.



"You have unlimited power on this canvas. You can literally move mountains." — Bob Ross

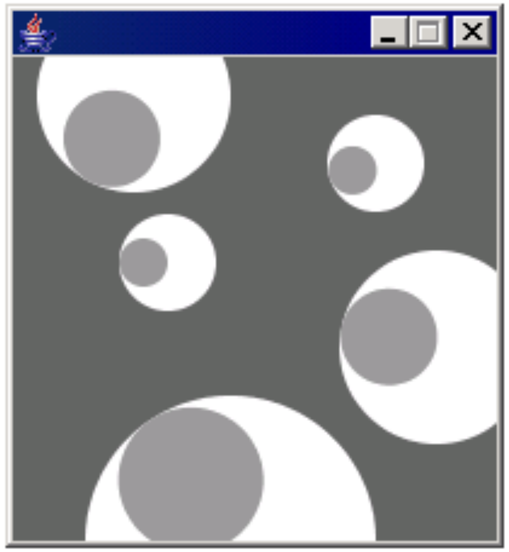
Processing - 0123 Beta

File Edit Sketch Tools Help

Run

sketch_070126a §

```
    ey = y;
    size = s;
  }

  void update(int mx, int my) {
    angle = atan2(my-ey, mx-ex);
  }

  void display() {
    pushMatrix();
    translate(ex, ey);
    fill(255);
    ellipse(0, 0, size, size);
    rotate(angle);
    fill(153);
    ellipse(size/4, 0, size/2, size/2);
    popMatrix();
  }
}
```



59

http://processing.org

US Air Traffic, Aaron Koblin

Graphics APIs

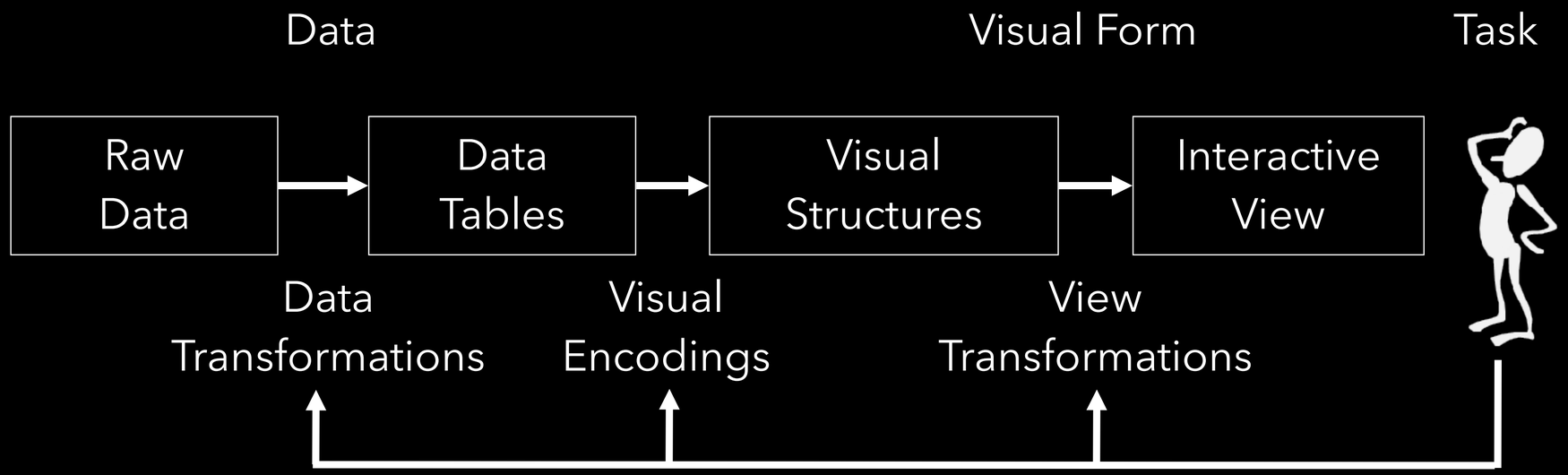
Processing, OpenGL, Java2D

Component Architectures

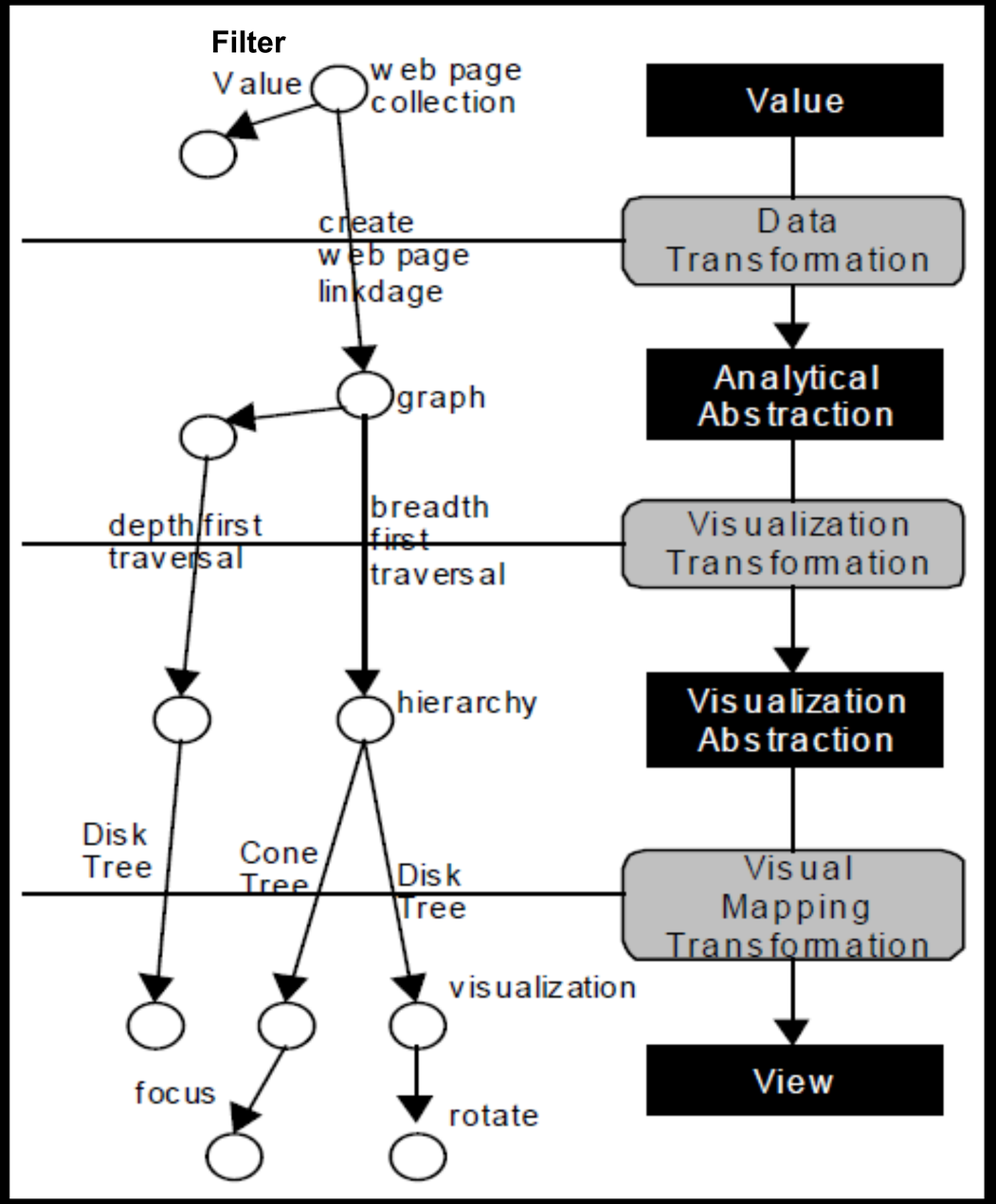
Prefuse, Flare, Improvise, VTK

Graphics APIs

Processing, OpenGL, Java2D


Data
Visual Form
Task
Raw Data
Data Tables
Visual Structures
Interactive View
Data Transformations
Visual Encodings
View Transformations

Data State Model
[Chi 98]



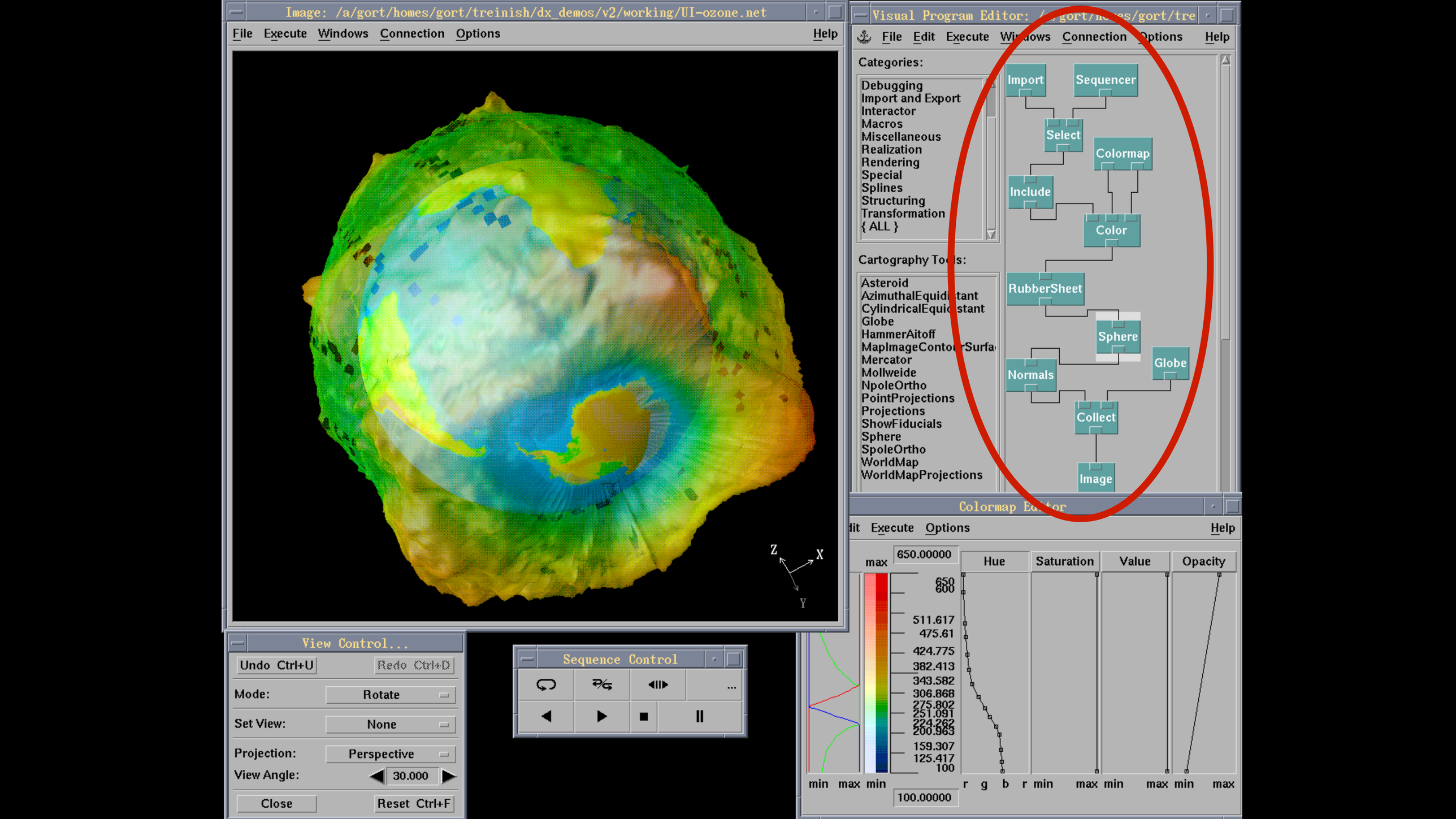
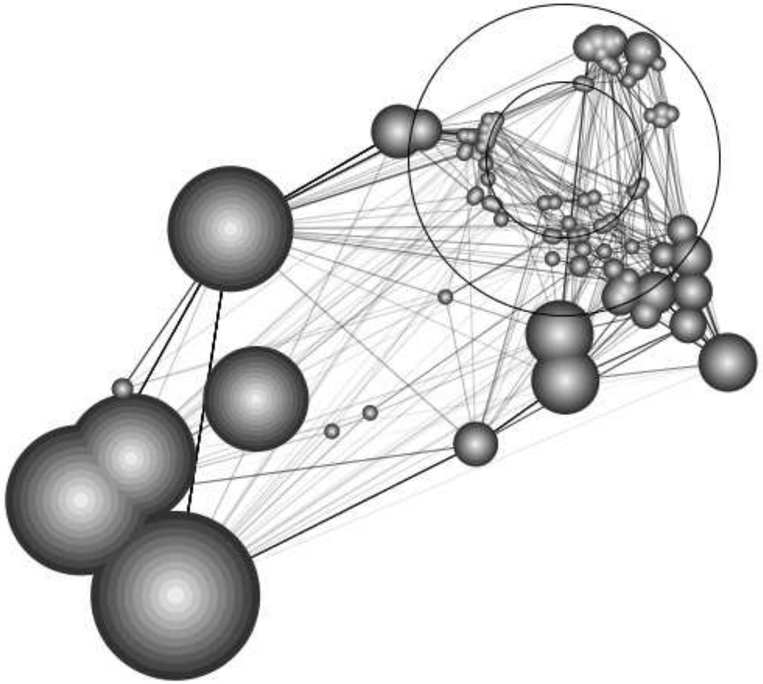


Image: /a/gort/homes/gort/treinish/dx_demos/v2/working/UI-ozone.net
File Execute Windows Connection Options Help
Z
X
Y
Visual Program Editor:
File Edit Execute Windows Connection Options Help
Categories:
Debugging
Import and Export
Interactor
Macros
Miscellaneous
Realization
Rendering
Special
Splines
Structuring
Transformation
{ ALL }
Cartography Tools:
Asteroid
AzimuthalEquidistant
CylindricalEquidistant
Globe
HammerAitoff
MapImageContourSurface
Mercator
Mollweide
NpoleOrtho
PointProjections
Projections
ShowFiducials
Sphere
SpoleOrtho
WorldMap
WorldMapProjections
Import
Sequencer
Select
Colormap
Include
Color
RubberSheet
Sphere
Globe
Normals
Collect
Image
Colormap Editor
Edit Execute Options Help
max
650.00000
Hue
Saturation
Value
Opacity
650
600
511.617
475.61
424.775
382.413
343.582
306.868
275.802
251.091
224.262
200.963
159.307
125.417
100
min max min
r g b r min max min max min max
100.00000
View Control...
Undo Ctrl+U
Redo Ctrl+D
Mode:
Rotate
Set View:
None
Projection:
Perspective
View Angle:
30.000
Close
Reset Ctrl+F
Sequence Control

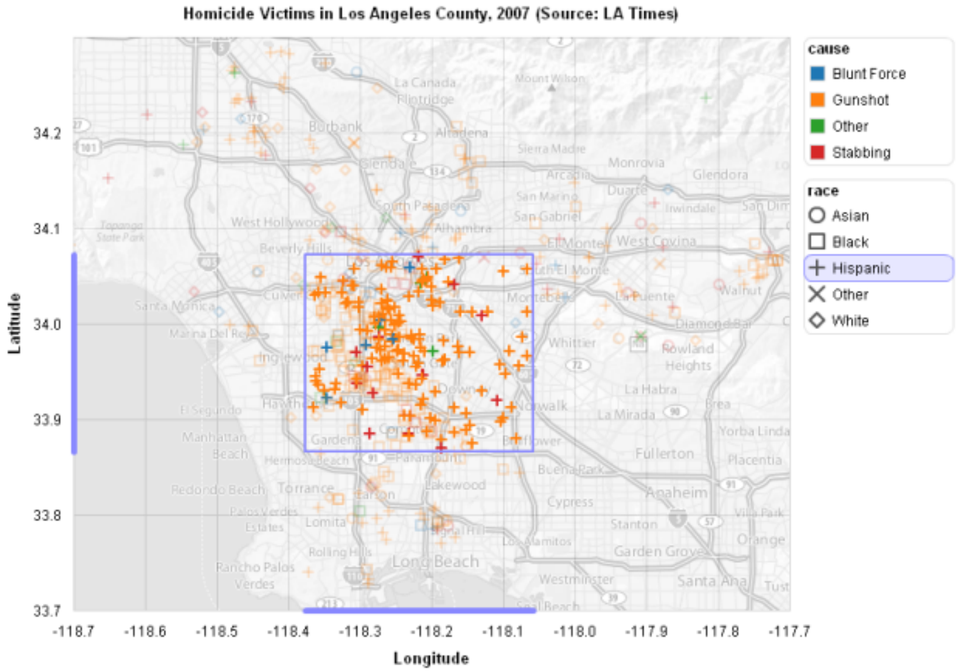
Prefuse & Flare

Operator-based toolkits for visualization design

Vis = (Input Data -> Visual Objects) + Operators



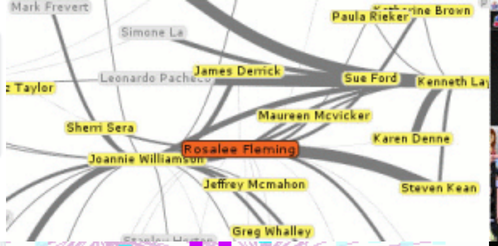
Prefuse (http://prefuse.org)

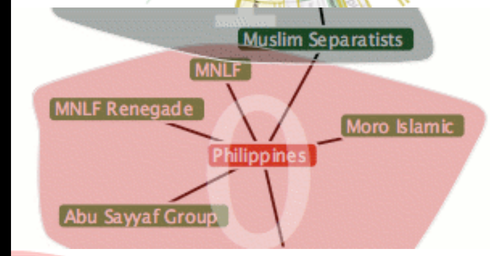


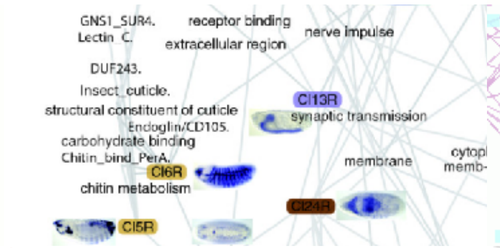
Flare (http://flare.prefuse.org)

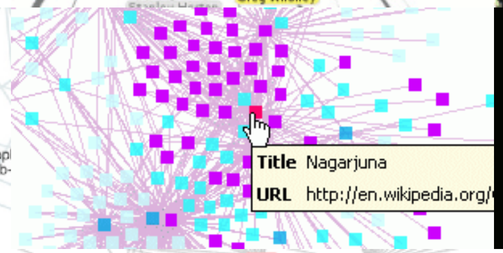

concept
idea
law
plan
kind
principle
model
plan of action
play
system
scheme

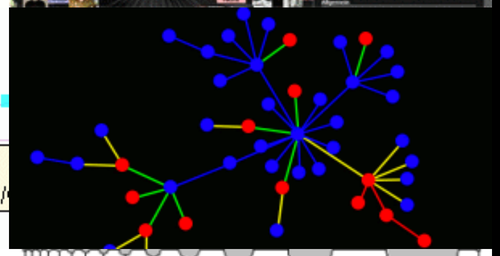

in the
at a conference
national park
in
banff

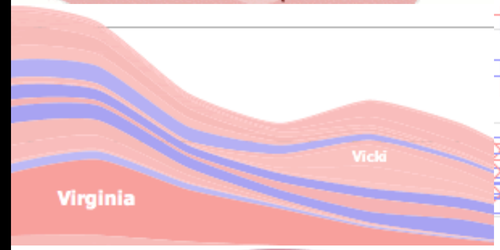

Mark Frevert
Paula Rieker
Simone La
James Derrick
Sue Ford
Leonardo Pacheco
Maureen Mcvicker
Sherri Sera
Karen Denne
Rosalee Fleming
Joannie Williamson
Jeffrey Mcmahon
Steven Kean
Greg Whalley

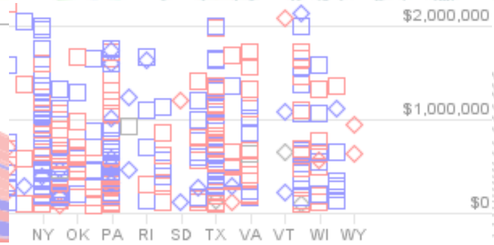

Muslim Separatists
MNLF
MNLF Renegade
Moro Islamic
Philippines
Abu Sayyaf Group

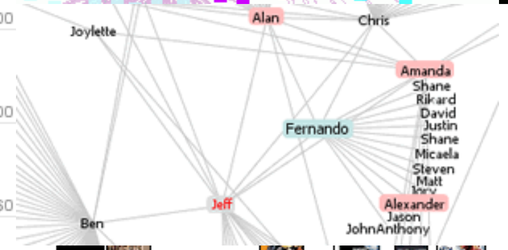

GNS1_SUR4.
Lectin_C.
receptor binding
extracellular region
nerve impulse
DUF243.
Insect_cuticle.
structural constituent of cuticle
Endoglin/CD105.
carbohydrate binding
Chitin_bind_PerA.
chitin metabolism
membrane
Cl13R
synaptic transmission
Cl6R
Cl24H
Cl5R

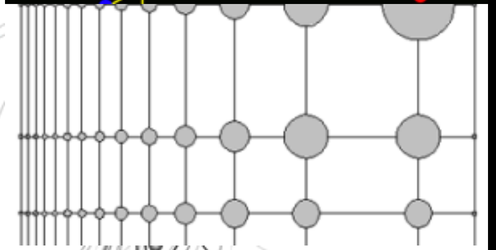

Title Nagarjuna
URL http://en.wikipedia.org/



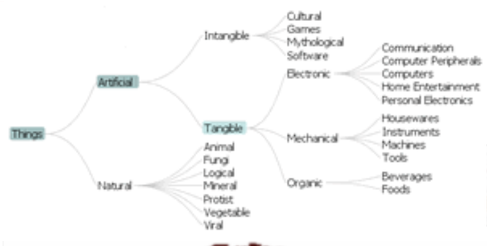

Virginia
Vicki


$2,000,000
$1,000,000
$0
NY OK PA RI SD TX VA VT WI WY

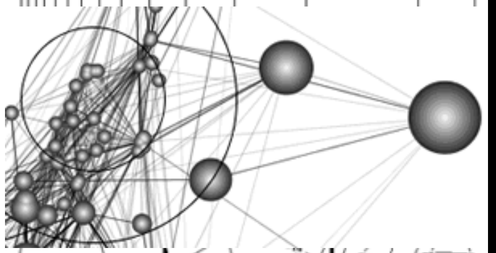

Joylette
Alan
Chris
Amanda
Shane
Rikard
David
Justin
Shane
Micaela
Steven
Matt
Fernando
Jeff
Ben
Alexander
Jason
JohnAnthony

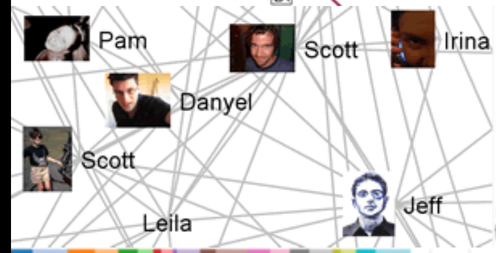



WA
MT
ND
MN
OR
ID
SD
WI
MI
WY
NE
IA
IL
IN
OH
NV
UT
CO
KS
MO
KY
VA
TN
NC
CA
OK
AR
NM
SC
AZ
MS
AL
GA
TX
LA
PA
NJ
MD
NY
VT
NH
ME
RI

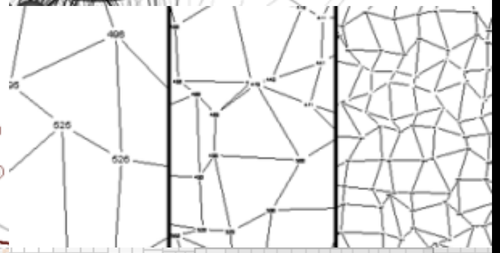

Things
Artificial
Natural
Intangible
Tangible
Electronic
Mechanical
Organic
Cultural
Games
Mythological
Software
Communication
Computer Peripherals
Computers
Home Entertainment
Personal Electronics
Housewares
Instruments
Machines
Tools
Animal
Fungi
Logical
Mineral
Protist
Vegetable
Viral
Beverages
Foods

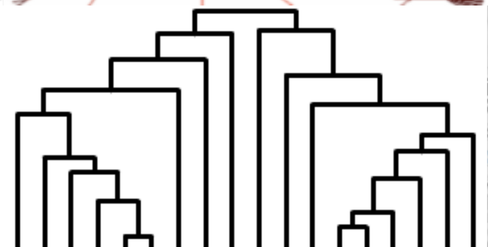

CHASING THE DIME
MICHAEL CONNELLY

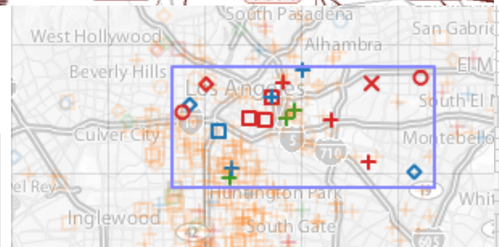



Pam
Scott
Irina
Danyel
Scott
Leila
Jeff


Ideology
Science
Humanities
Technology
Recurring
Knowledge
Generic
Civilizations
Historical
The Millenium
Events
People
Organizations
Categories
Natural
Things
Constellations






West Hollywood
South Pasadena
Alhambra
Beverly Hills
Culver City
Inglewood
South Gate


https://gunsmokeandknitting.files.wordpress.com/2014/12/boston-tea-party-16.jpg

Component Architectures

Prefuse, Flare, Improvise, VTK

Graphics APIs

Processing, OpenGL, Java2D

Chart Typologies

Excel, Many Eyes, Google Charts

Component Architectures

Prefuse, Flare, Improvise, VTK

Graphics APIs

Processing, OpenGL, Java2D

Chart Typologies

Data Sets : State Quick Facts

Uploaded By: zinggoat
Data Source: US Census Bureau
Description:
Tags: people census

Created at: Friday May 18, 3:08 PM

view as text | edit data set

	People QuickFacts	Population 2005 estimate	Population percent change April 1 2000 to July 1 2005	Population 2000	Population percent change 1990 to 2000	Persons under 5 years old percent 2004	Persons under 18 years old percent 2004	Persons 65 years old and over percent 2004
1	Alabama	4557808	0.03	4447100	0.1	0.07	0.24	0.13
2	Alaska	663661	0.06	626932	0.14	0.08	0.29	0.06
3	Arizona	5939292	0.16	5130632	0.4	0.08	0.27	0.13
4	Arkansas	2779154	0.04	2673400	0.14	0.07	0.25	0.14
5	California	36132147	0.07	33871648	0.14	0.07	0.27	0.11
6	Colorado	4665177	0.08	4301261	0.31	0.07	0.26	0.1
7	Connecticut	3510297	0.03	3405565	0.04	0.06	0.24	0.14
8	Delaware	843524	0.08	783600	0.18	0.07	0.23	0.13
9	Florida	17789864	0.11	15982378	0.24	0.06	0.23	0.17
10	Georgia	9072576	0.11	8186453	0.26	0.08	0.26	0.1
11	Hawaii	1275194	0.05	1211537	0.09	0.07	0.24	0.14
12	Idaho	1429096	0.1	1293953	0.29	0.07	0.27	0.11
13	Illinois	12763371	0.03	12419293	0.09	0.07	0.26	0.12

Choosing a visualization type for State Quick Facts

Analyze a text

Tag Cloud

How are you using your words? This enhanced tag cloud will show you the words popularity in the given set of text.

Learn more

Wordle

Wordle is a toy for generating "word clouds" from text that you provide. The clouds give greater prominence to words that appear more frequently in the source text.

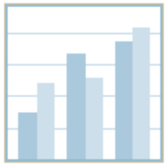
Learn more

Word Tree

See a branching view of how a word or phrase is used in a text. Navigate the text by zooming and clicking.

Learn more

Compare a set of values



Bar Chart

How do the items in your data set stack up? A bar chart is a simple and recognizable way to compare values. You can display several sets of bars for multivariate comparisons.

Learn more

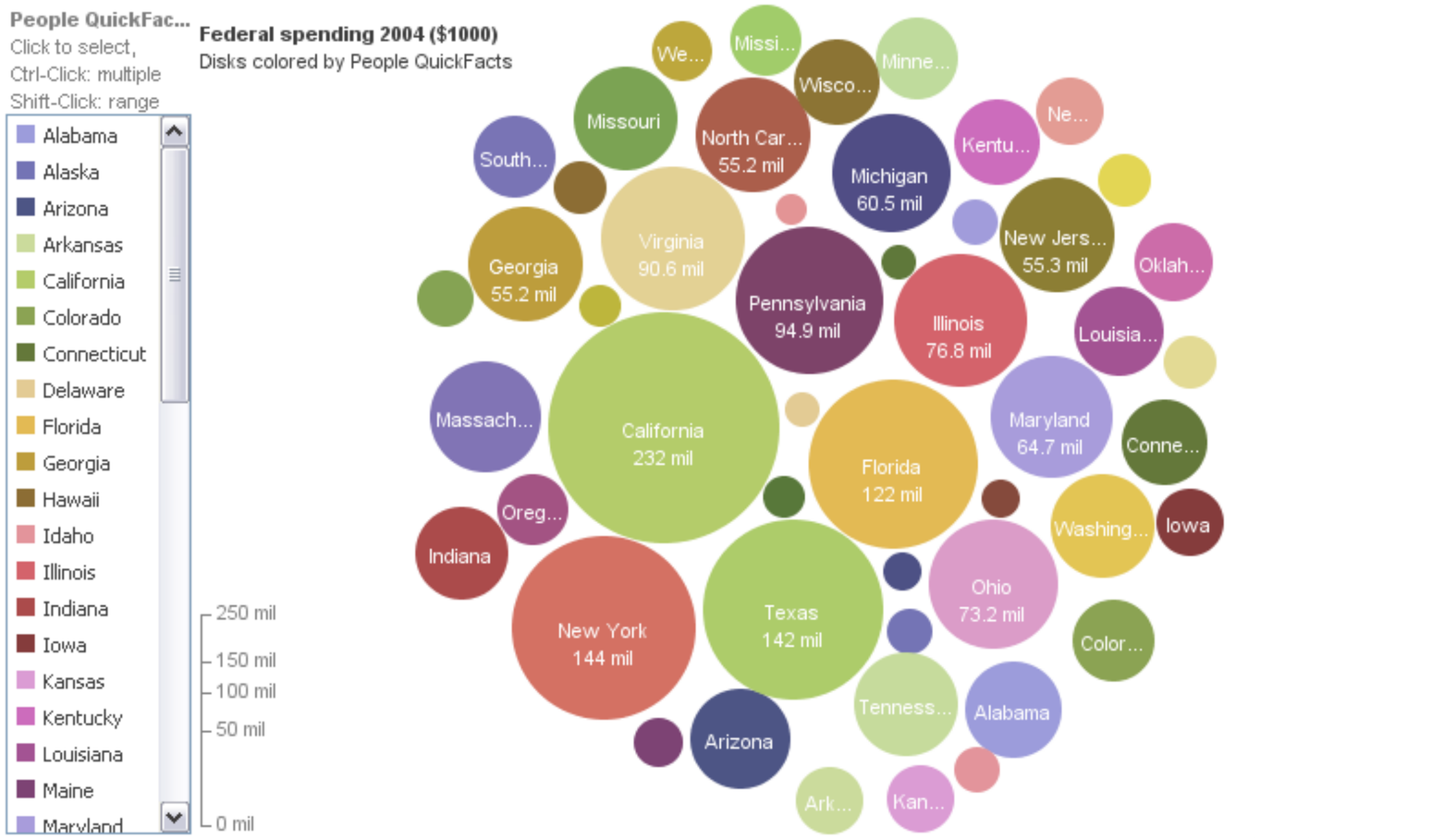
Block Histogram

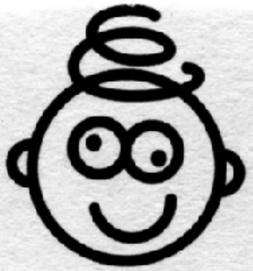
This versatile chart lets you get a quick sense of how a single set of data is distributed. Each item in the data is an individually identifiable block.

Learn more

Visualizations : Federal Spending by State, 2004

Creator: Anonymous
Tags: census people





MAD LIBS®

MY MUSIC LESSON

Every Wednesday, when I get home from school, I have a piano lesson. My teacher is a very strict house (NOUN). Her name is Hillary Clinton (CELEBRITY (FEMALE)). Our piano is a Steinway Concert tree (NOUN) and it has 88 cups (PLURAL NOUN). It also has a soft pedal and a/an Smily (ADJECTIVE) pedal. When I have a lesson, I sit down on the piano AIBERTO (NOUN) and play for 16 minutes (PERIOD OF TIME). I do scales to exercise my cats (PLURAL NOUN), and then I usually play a minuet by Johann Sebastian Washington (CELEBRITY (LAST NAME)). Teacher says I am a natural Haunted House (NOUN) and have a good musical leg (PART OF THE BODY). Perhaps when I get better I will become a concert vet (PROFESSION) and give a recital at Carnegie hospital (TYPE OF BUILDING).

https://en.wikipedia.org/wiki/DeLorean_time_machine#/media/File:TeamTimeCar.com-BTTF_DeLorean_Time_Machine-OtoGodfrey.com-JMortonPhoto.com-07.jpg

Grammars

- *Context-free grammars* good for describing textual dialogs which have a formal syntax.
  - "BNF" for languages (Backus–Naur Form)
  - alphabet of symbols
  - sentences - legal sequences of symbols
  - language - set of all legal sentences
  - grammar - set of terminals, set of non-terminals, set of productions, start symbol
  - context-free-grammar: all productions of the form <a> ::= B where <a> is a single nontermimal, and B is a non-empty string of terminals and/or non-terminals

Slide from 2/17/ 2020

The Grammar of Graphics. Leland Wilkinson. 1999

https://www.youtube.com/watch?v=9eQl6dmugqY

[M]ost charting packages channel user requests into a **rigid array of chart types**. To atone for this lack of flexibility, they offer a kit of post-creation editing tools to return the image to what the user originally envisioned. **They give the user an impression of having explored data rather than the experience**.

Chart Typologies

Excel, Many Eyes, Google Charts

Component Architectures

Prefuse, Flare, Improvise, VTK

Graphics APIs

Processing, OpenGL, Java2D

Chart Typologies

Excel, Many Eyes, Google Charts

Visual Analysis Grammars

VizQL, ggplot2

Component Architectures

Prefuse, Flare, Improvise, VTK

Graphics APIs

Processing, OpenGL, Java2D

Tableau - Book2

File Edit View Format Data Analysis Table Bookmark Window Help

Show Me!

Schema

congress.csv Connection

Find:

Dimensions

- Abc Candidate
- Abc Candidate ID
- Abc General Elec Status
- Abc Incumbent/Challenger/Open-Seal
- # Party
- Abc Party Desig
- Abc Primary Elec Status
- Abc Runoff Elec Status
- Abc Spec Elec Status
- Abc State Code
- # Year
- Abc *Measure Names*

Measures

- # District
- # General Elec Pct
- # Total Receipts
- # *Measure Values*

Groups

Filters:

Level of Detail:

Mark: Automatic

Text:

Color: Party

Size:

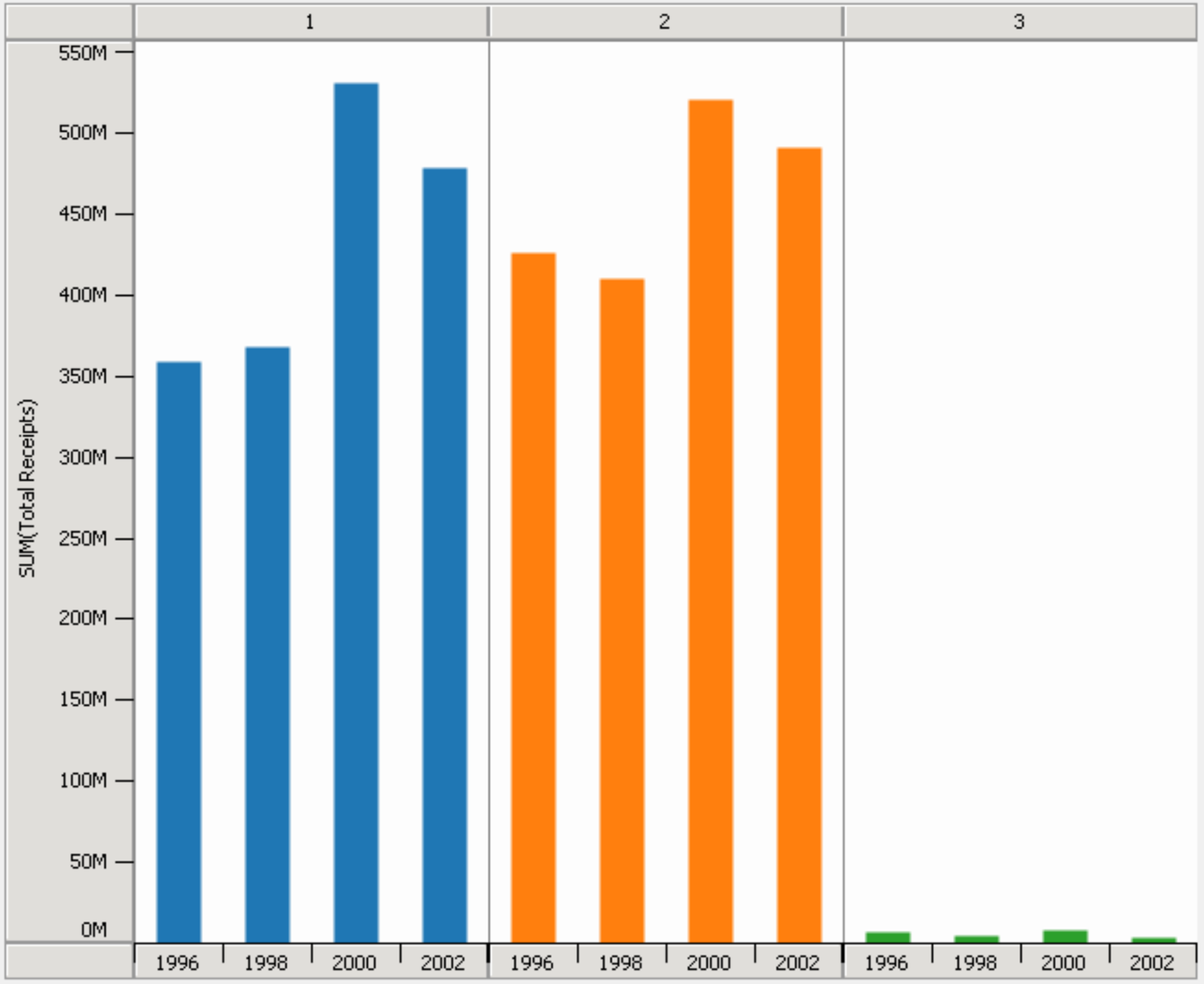
Legend:

1
2
3

Size:

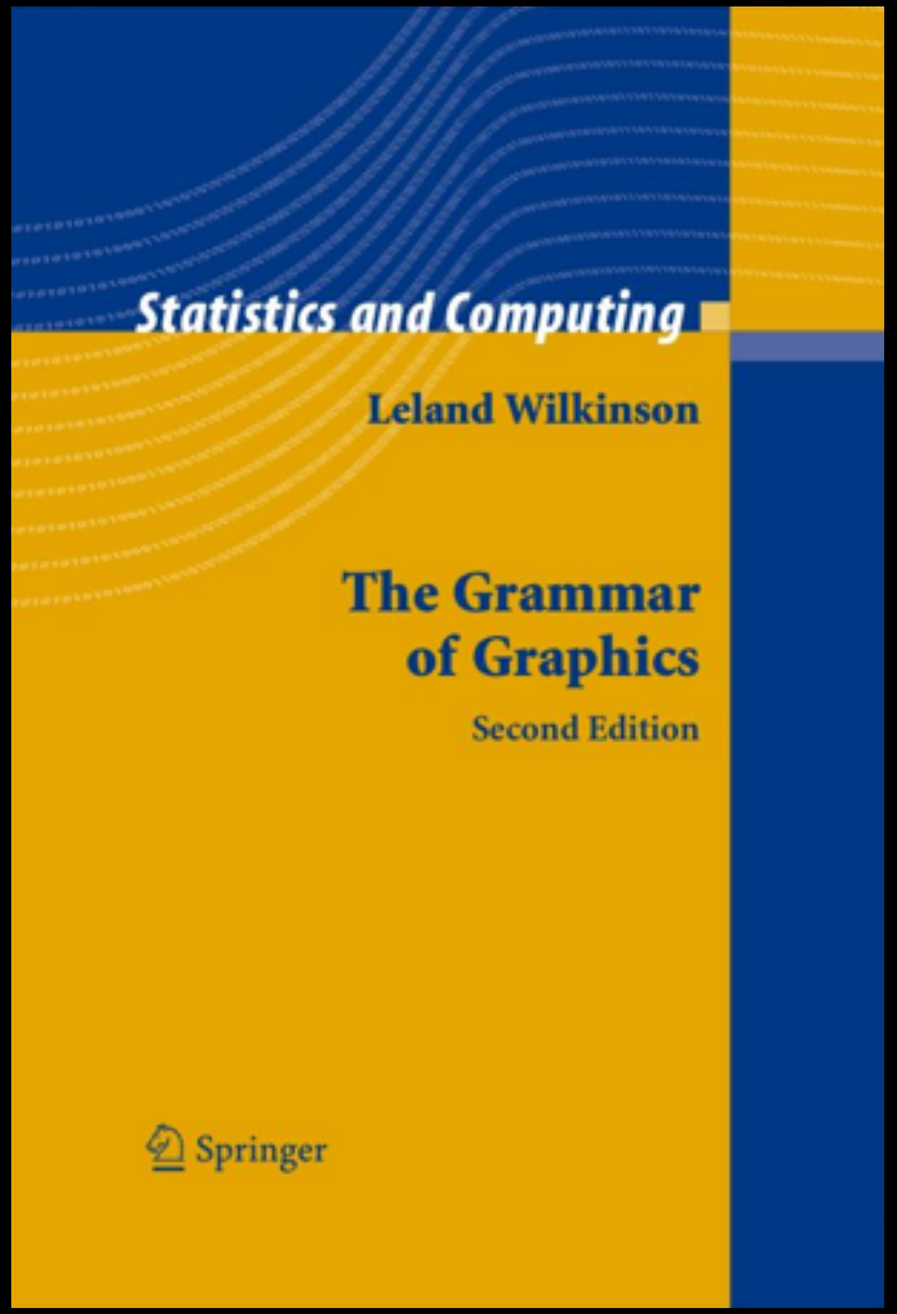
Columns: Party | Year

Rows: SUM(Total..

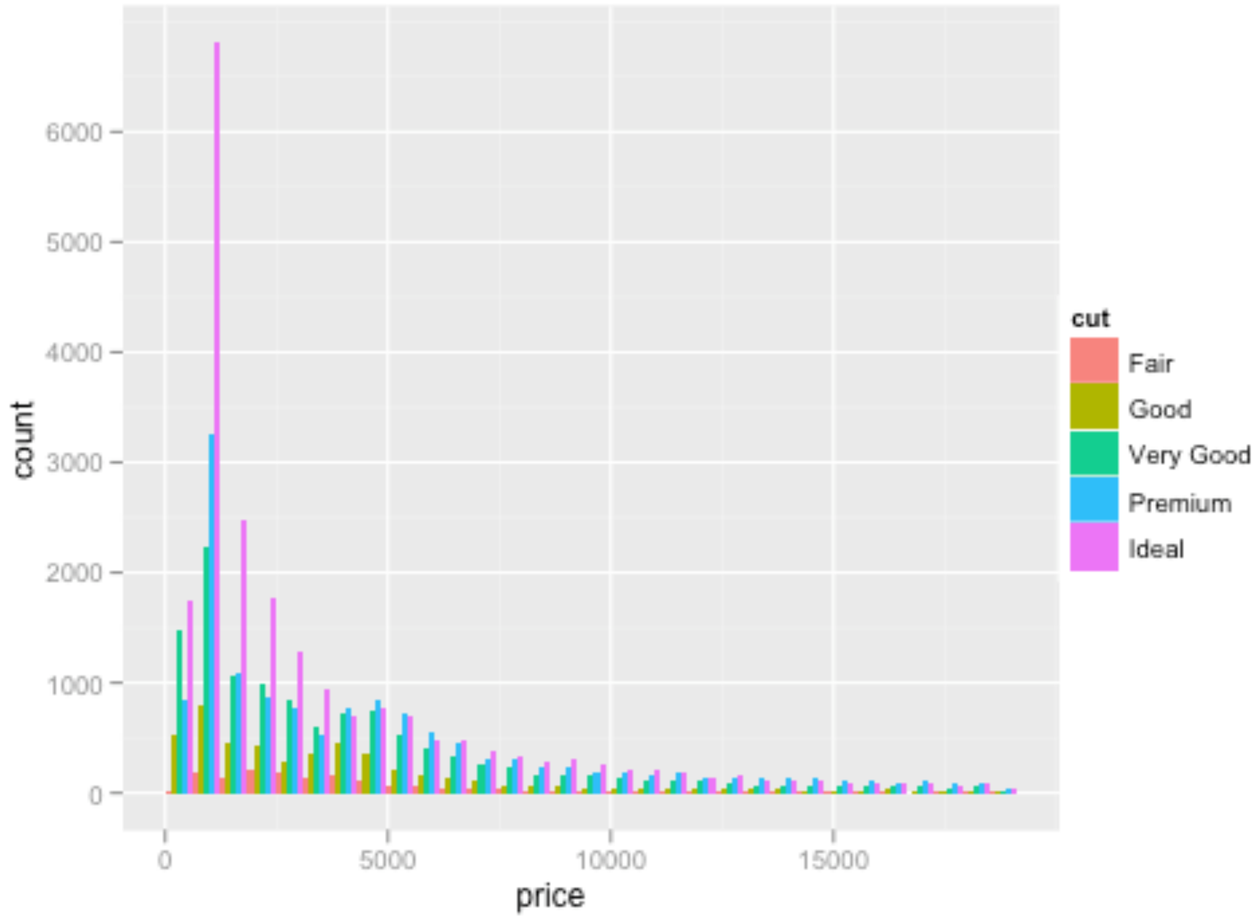


Sheet 1

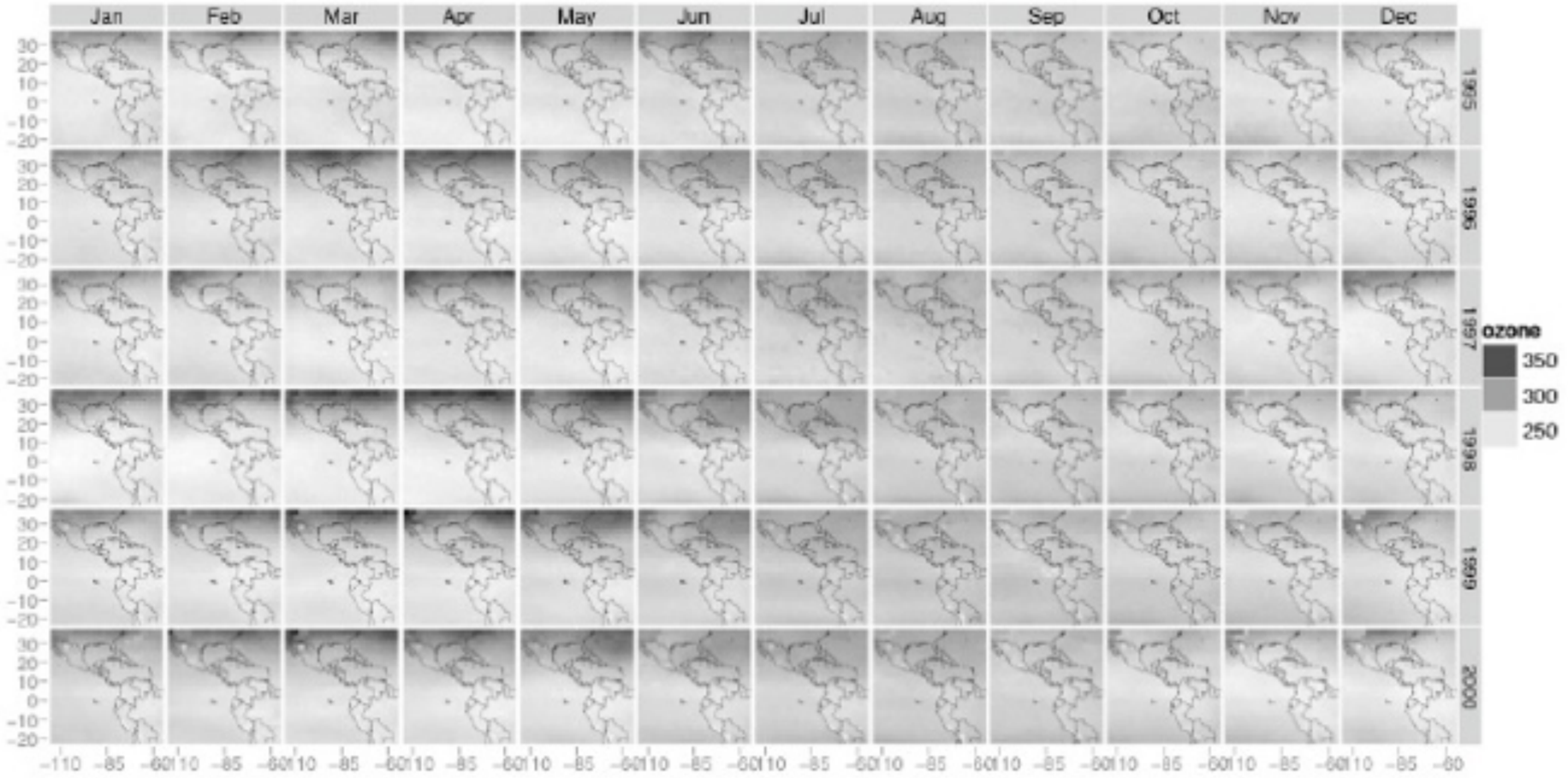
Ready


Statistics and Computing
Leland Wilkinson
The Grammar of Graphics
Second Edition
Springer

```
ggplot(diamonds, aes(x=price, fill=cut))
+ geom_bar(position="dodge")
```




```
ggplot(diamonds, aes(x=price, fill=cut))
+ geom_bar(position="dodge")
```




```
qplot(long, lat, data = expo, geom = "tile", fill = ozone,
  facets = year ~ month) +
scale_fill_gradient(low = "white", high = "black") + map
```

Chart Typologies

Excel, Many Eyes, Google Charts

Visual Analysis Grammars

VizQL, ggplot2

Component Architectures

Prefuse, Flare, Improvise, VTK

Graphics APIs

Processing, OpenGL, Java2D

Chart Typologies

Excel, Many Eyes, Google Charts

Visual Analysis Grammars

VizQL, ggplot2

Component Architectures

Prefuse, Flare, Improvise, VTK

Graphics APIs

Processing, OpenGL, Java2D

Chart Typologies

Excel, Many Eyes, Google Charts

Visual Analysis Grammars

VizQL, ggplot2

Component Architectures

Prefuse, Flare, Improvise, VTK

Graphics APIs

Processing, OpenGL, Java2D

Chart Typologies

Excel, Many Eyes, Google Charts

Visual Analysis Grammars

VizQL, ggplot2

?

Component Architectures

Prefuse, Flare, Improvise, VTK

Graphics APIs

Processing, OpenGL, Java2D

Chart Typologies

Excel, Many Eyes, Google Charts

Visual Analysis Grammars

VizQL, ggplot2

Visualization Grammars

Protovis, D3.js

Component Architectures

Prefuse, Flare, Improvise, VTK

Graphics APIs

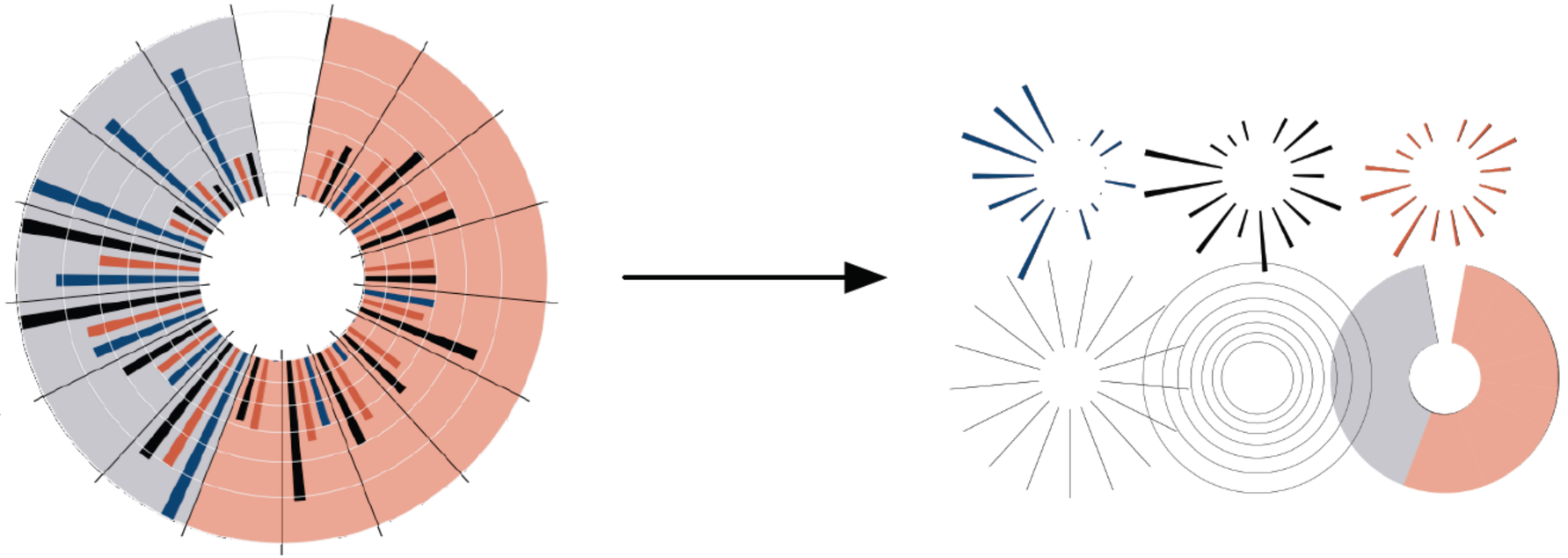
Processing, OpenGL, Java2D

Protovis & D3

Today's first task is not to invent wholly new [graphical] techniques, though these are needed. Rather we need most vitally to recognize and reorganize the **essential of old techniques, to make easy their assembly in new ways, and to modify their external appearances to fit the new opportunities**.

Data Analysis & Statistics. Turkey and Wilk. *1965.*

Protovis: A Grammar for Visualization



A graphic is a composition of data-representative marks.

Jeffrey Heer, **Mike Bostock** & **Vadim Ogievetsky**

Visualization Grammar

Visualization Grammar

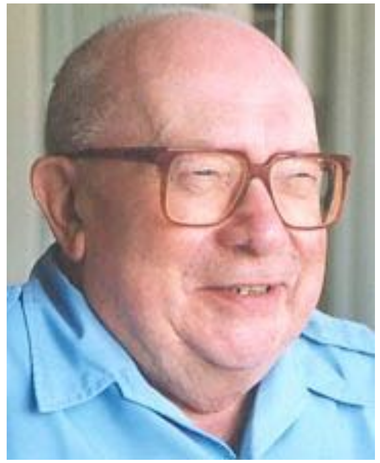
Data	Input data to visualize

Visualization Grammar

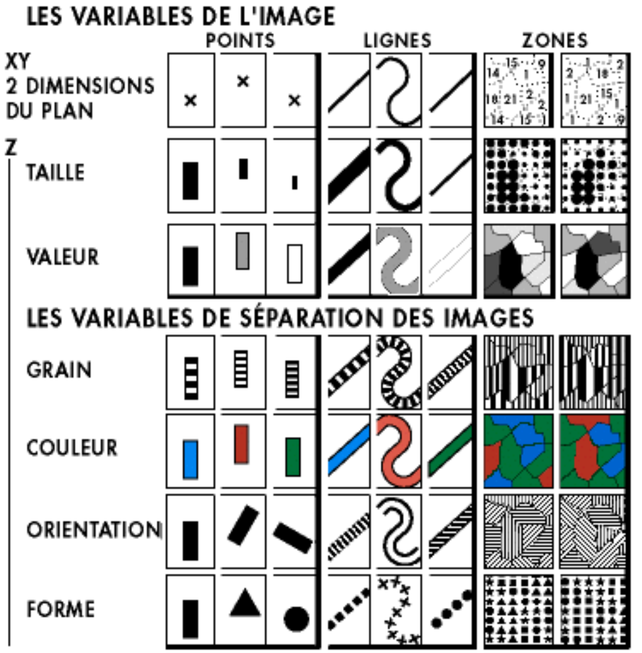
Data	Input data to visualize
Transforms	Grouping, stats, projection, layout

Visualization Grammar

Data	Input data to visualize
Transforms	Grouping, stats, projection, layout
Scales	Map data values to visual values



Jacques Bertin
Sémiologie Graphique, 1967

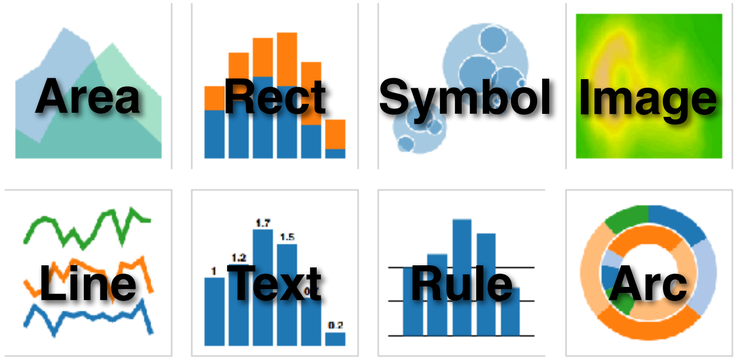


Visualization Grammar

Data	Input data to visualize
Transforms	Grouping, stats, projection, layout
Scales	Map data values to visual values
Guides	Axes & legends visualize scales

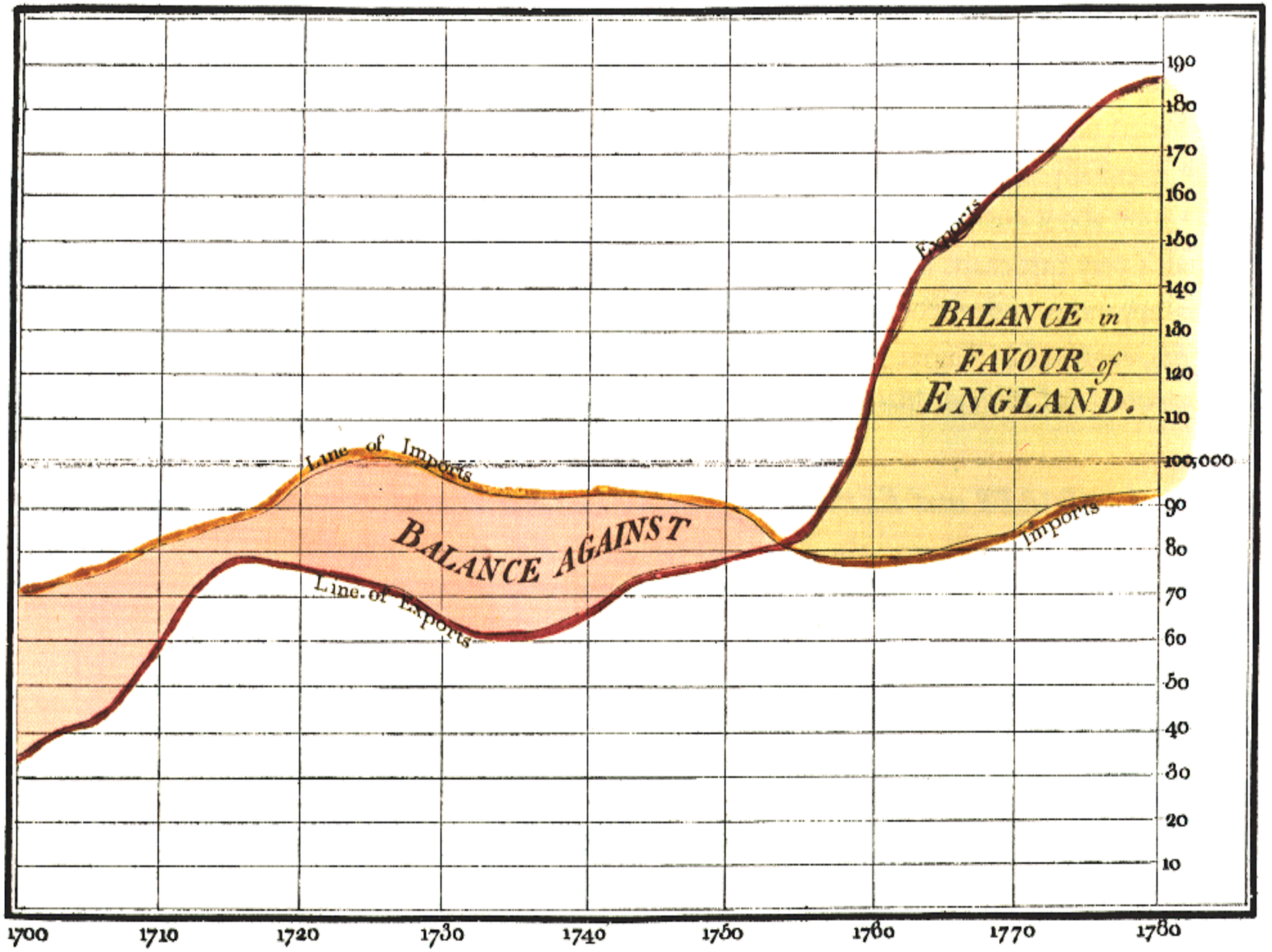
Visualization Grammar

Data	Input data to visualize
Transforms	Grouping, stats, projection, layout
Scales	Map data values to visual values
Guides	Axes & legends visualize scales
Marks	Data-representative graphics

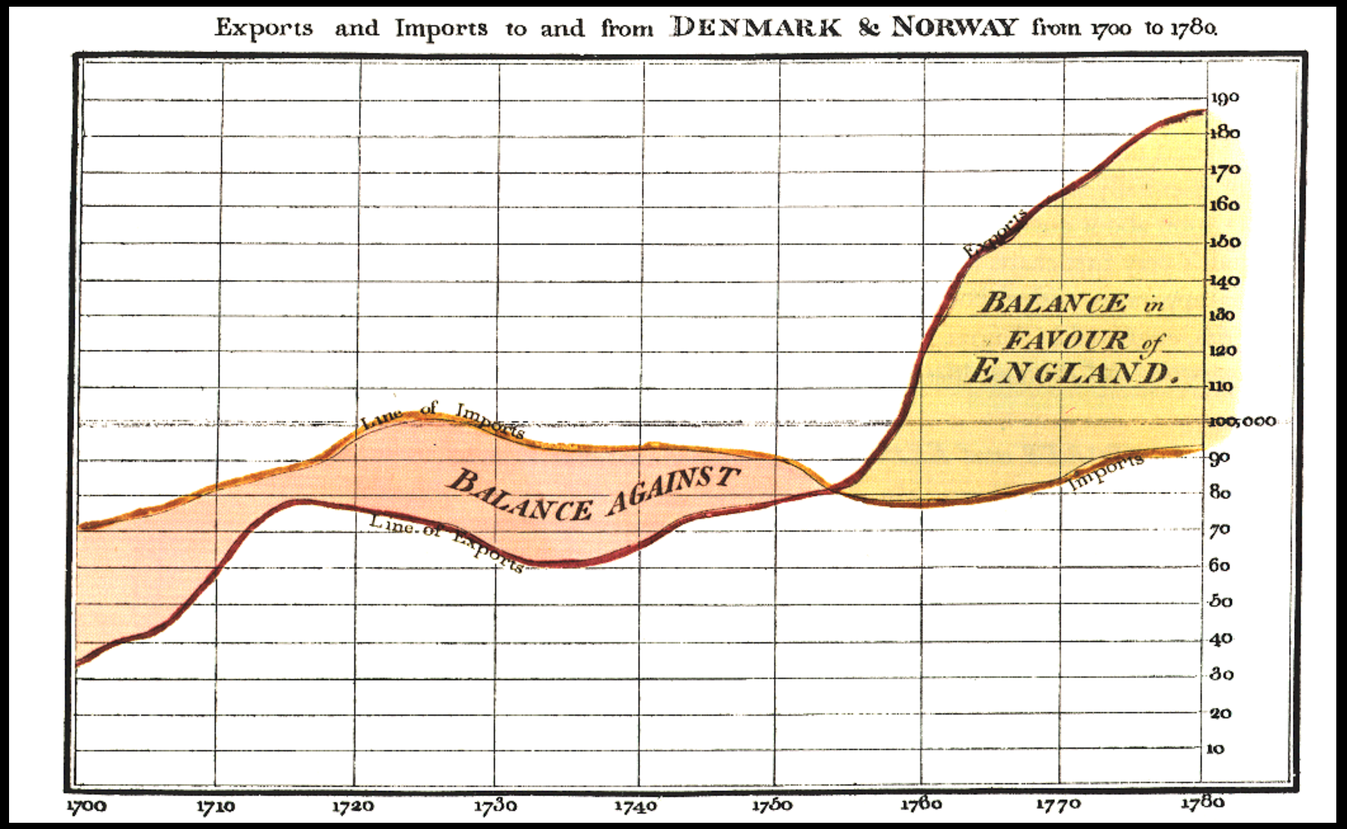


Deconstructions

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



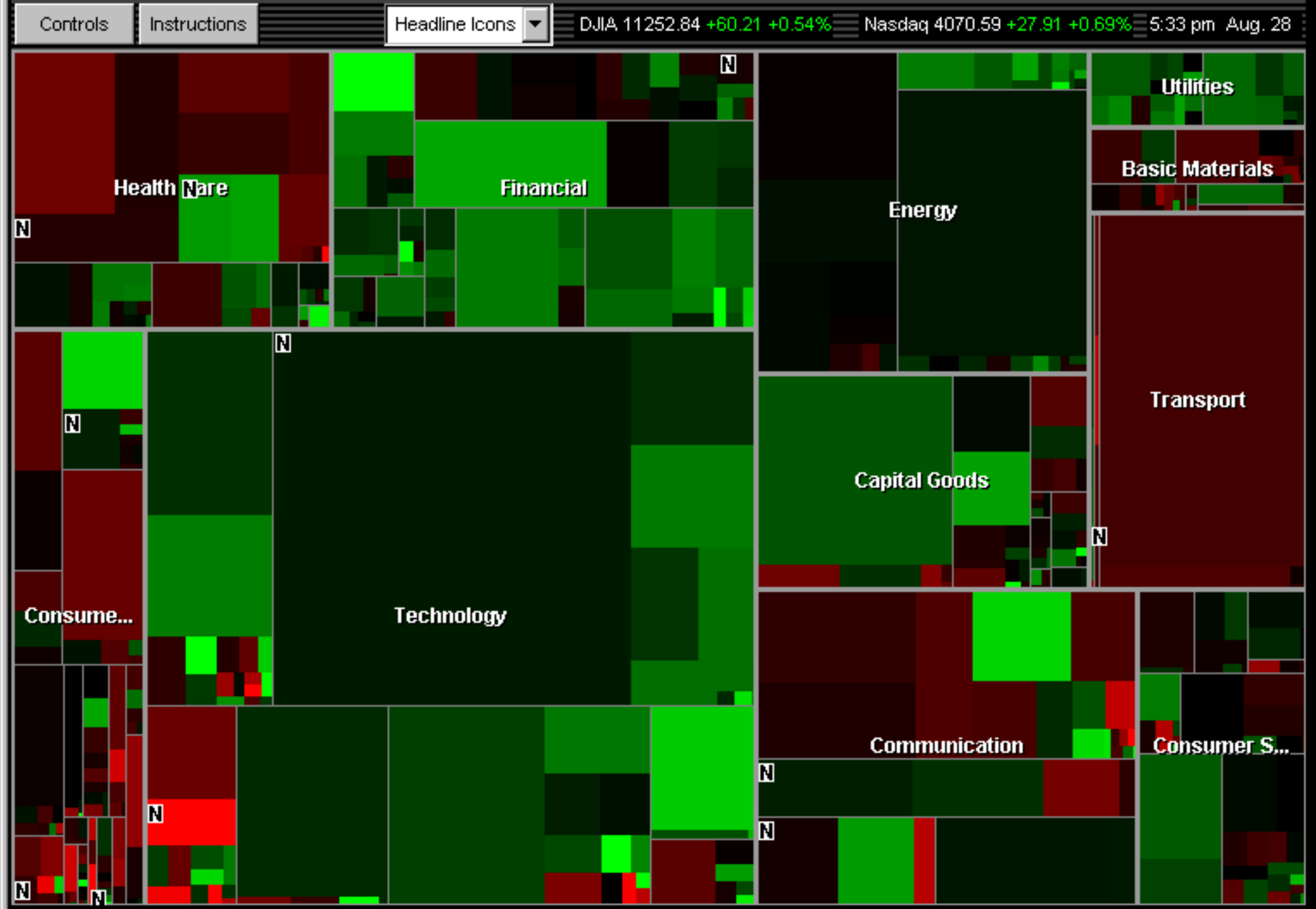
William Playfair, 1786



X-axis: year (Q)

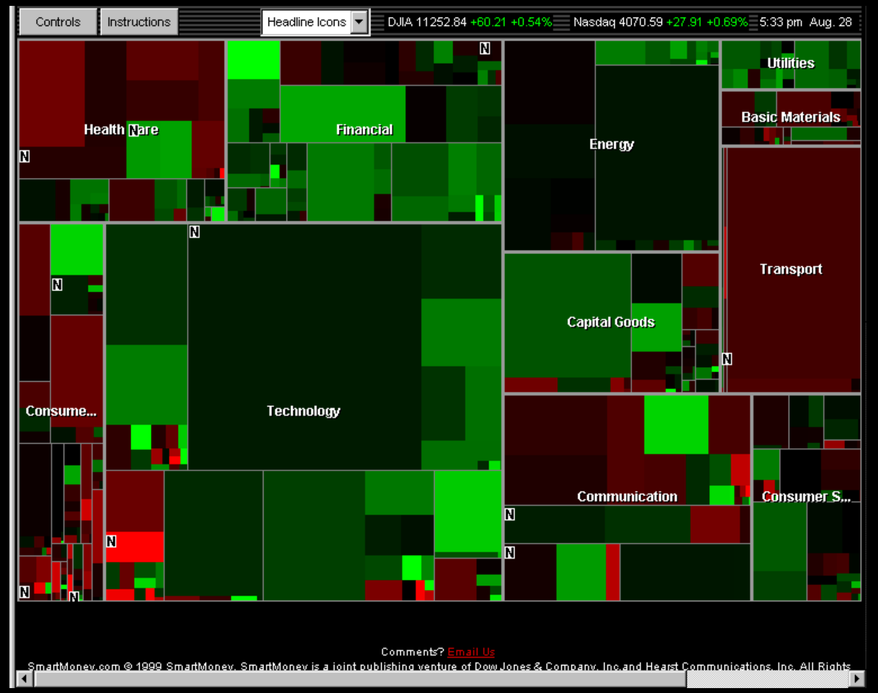
Y-axis: currency (Q)

Color: imports/exports (N, O)



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

Wattenberg's Map of the Market



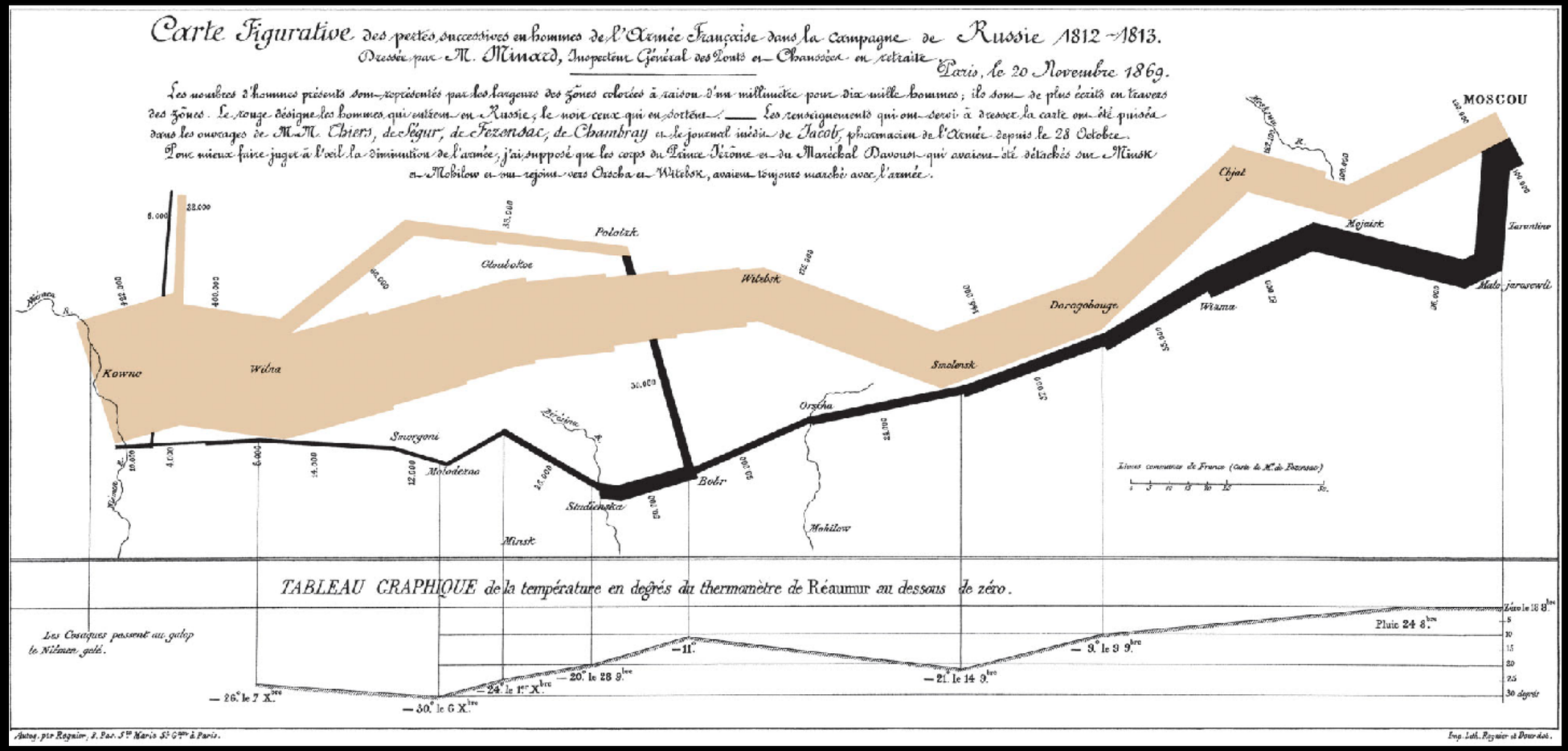
Rectangle Area: market cap (Q)

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

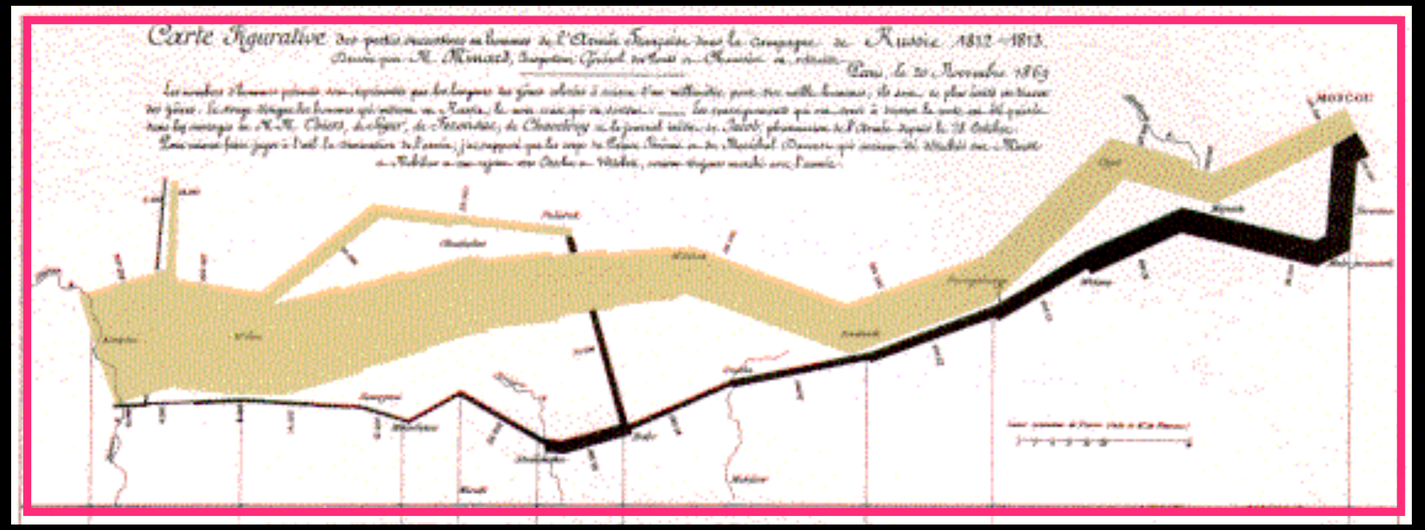
Color Hue: loss vs. gain (N, O)

Color Value: magnitude of loss or gain (Q)

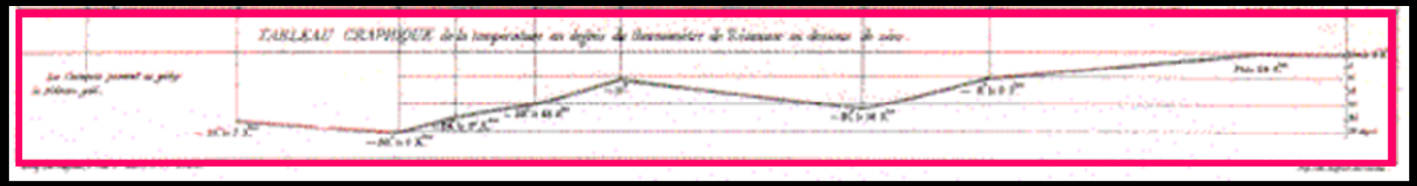
Minard 1869: Napoleon's March



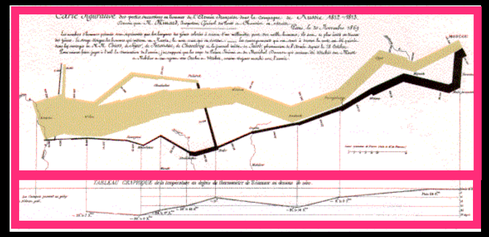
Single-Axis Composition



+



=

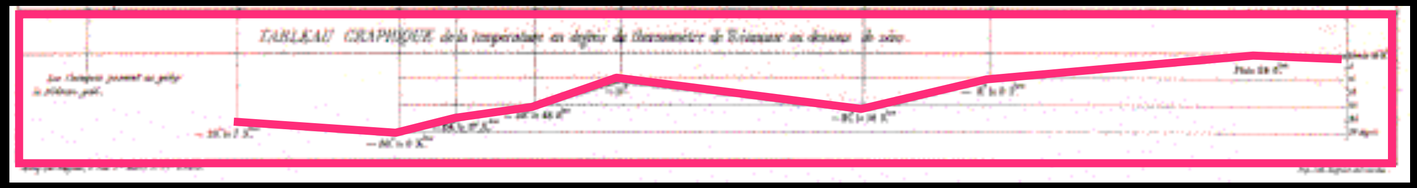


Mark Composition

Y-axis: temperature (Q)

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

\=



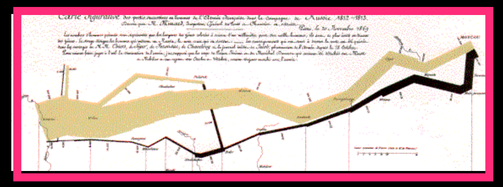
Temp over space/time (Q x Q)

Mark Composition

Y-axis: longitude (Q)

\+ **X-axis**: latitude (Q)

\+ **Width**: army size (Q)

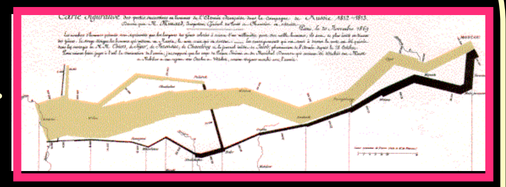


Army position (Q x Q) and army size (Q)

longitude (Q)

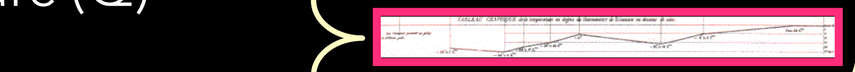
latitude (Q)

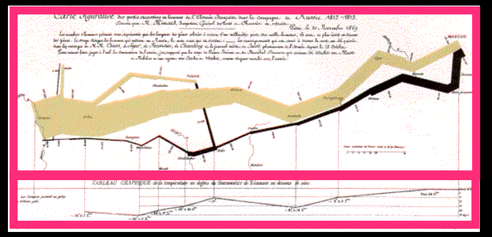
army size (Q)



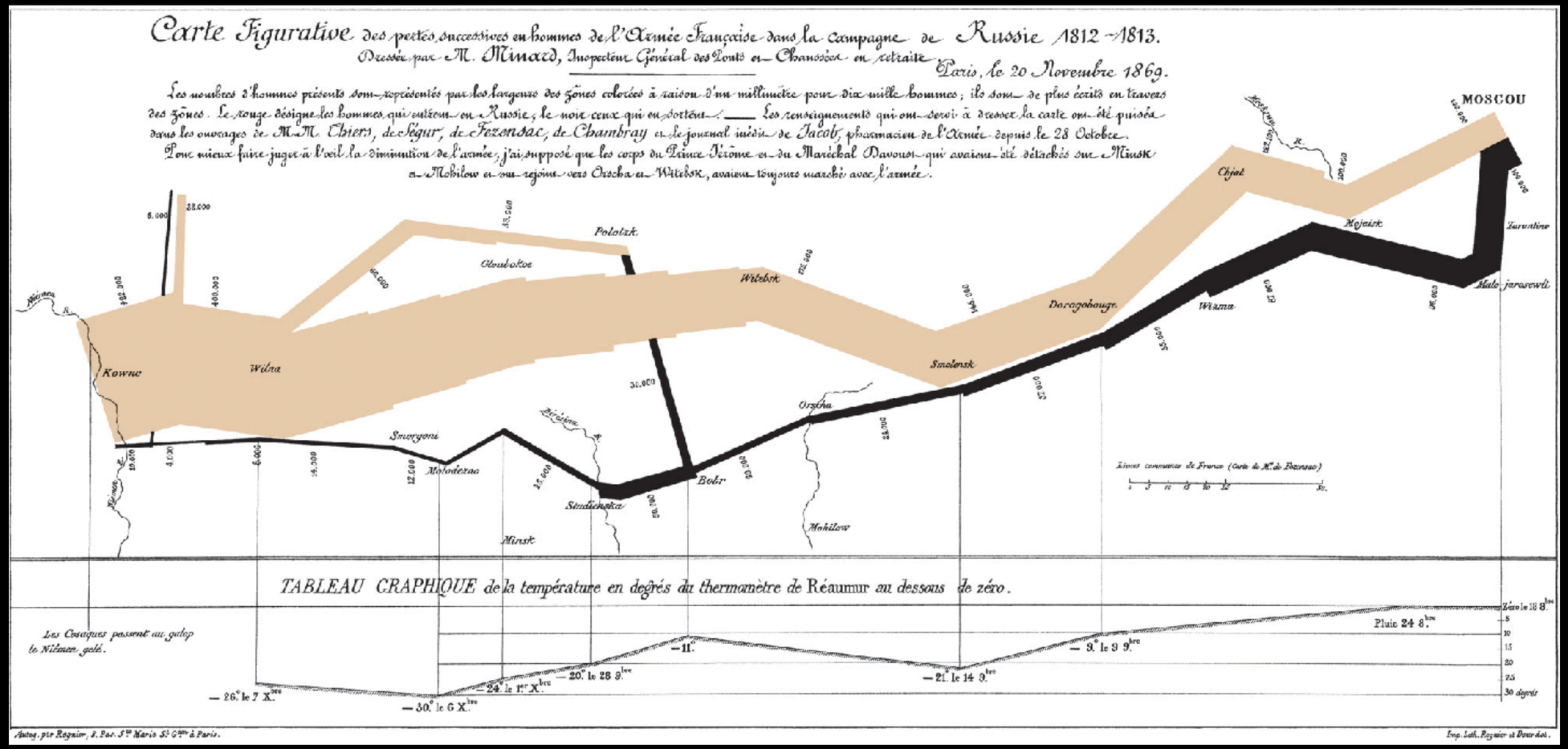
temperature (Q)

latitude (Q) / time (O)

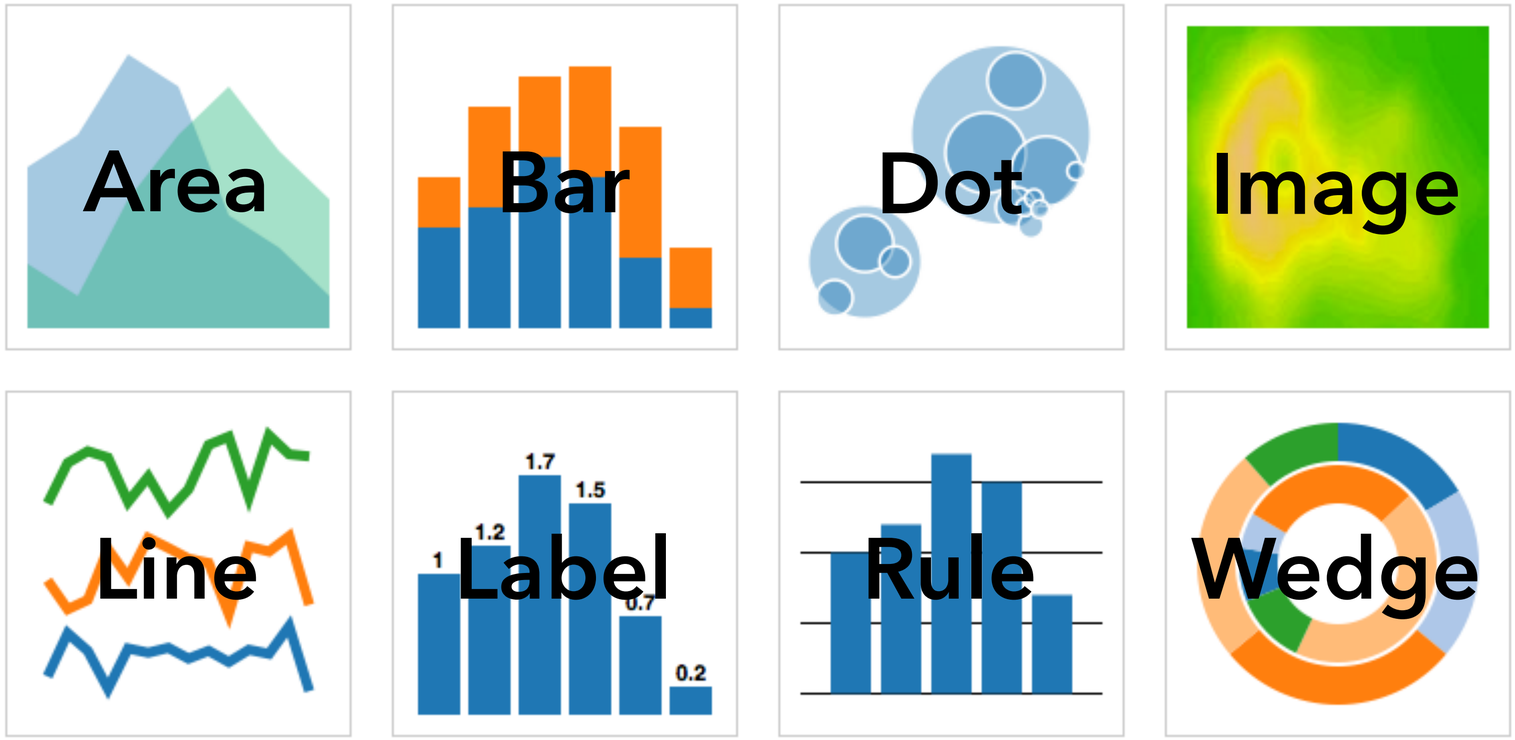




Minard 1869: Napoleon's March

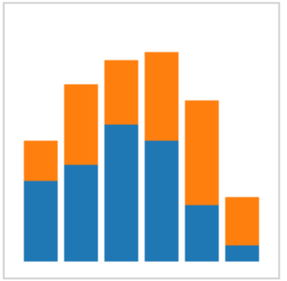


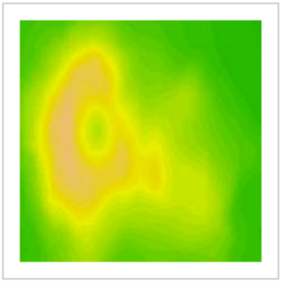
Depicts at least 5 quantitative variables. Any others?



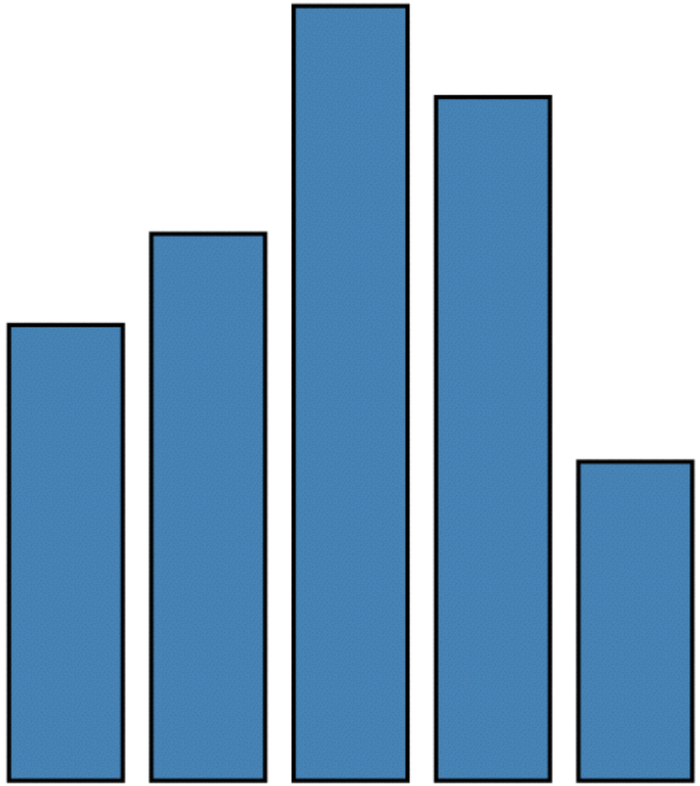
MARKS: Protovis graphical primitives

MARK	$\lambda : D \rightarrow R$
data	λ
visible	λ
left	λ
bottom	λ
width	λ
height	λ
fillStyle	λ
strokeStyle	λ
lineWidth	λ
...	λ

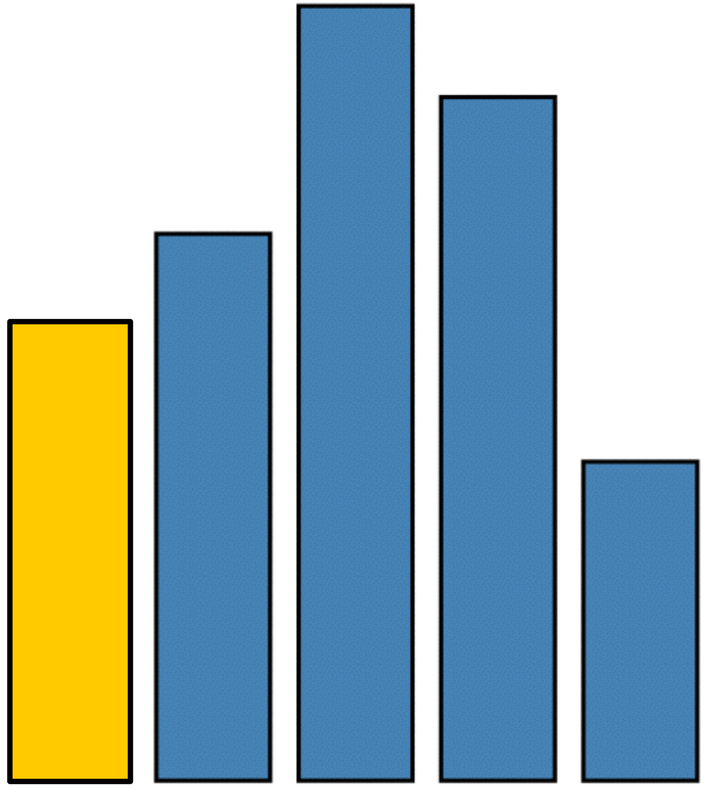


RECT	$\lambda : D \rightarrow R$				
data	1	1.2	1.7	1.5	0.7
visible	true				
left	λ: index * 25				
bottom	0				
width	20				
height	λ: datum * 80				
fillStyle	blue				
strokeStyle	black				
lineWidth	1.5				
...	...				



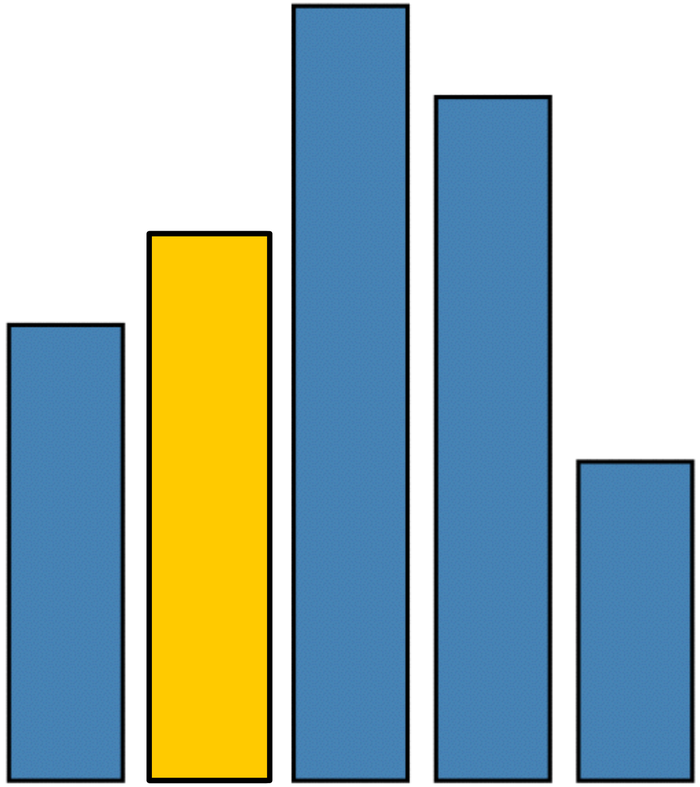
RECT $\lambda : D \rightarrow R$

data	1	1.2	1.7	1.5	0.7
visible	true				
left	0 * 25				
bottom	0				
width	20				
height	1 * 80				
fillStyle	blue				
strokeStyle	black				
lineWidth	1.5				
...	...				



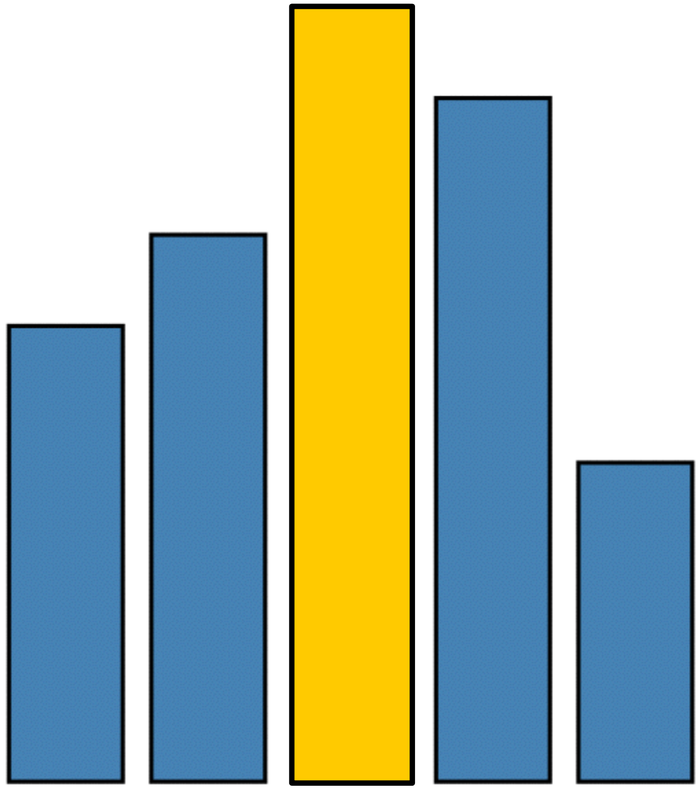
RECT $\lambda : D \rightarrow R$

data	1	1.2	1.7	1.5	0.7
visible	true				
left	1 * 25				
bottom	0				
width	20				
height	1.2 * 80				
fillStyle	blue				
strokeStyle	black				
lineWidth	1.5				
...	...				



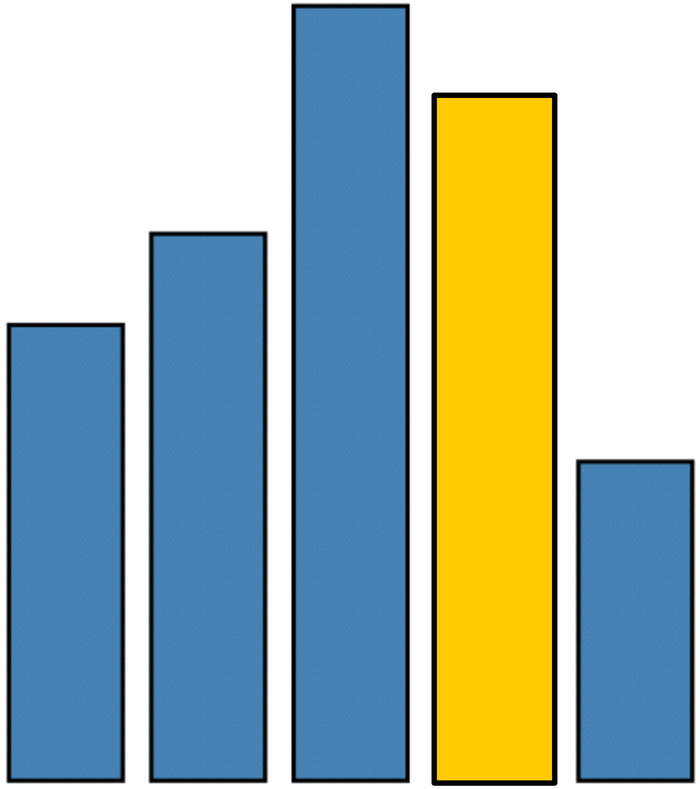
RECT $\lambda : D \rightarrow R$

data	1	1.2	1.7	1.5	0.7
visible	true				
left	2 * 25				
bottom	0				
width	20				
height	1.7 * 80				
fillStyle	blue				
strokeStyle	black				
lineWidth	1.5				
...	...				



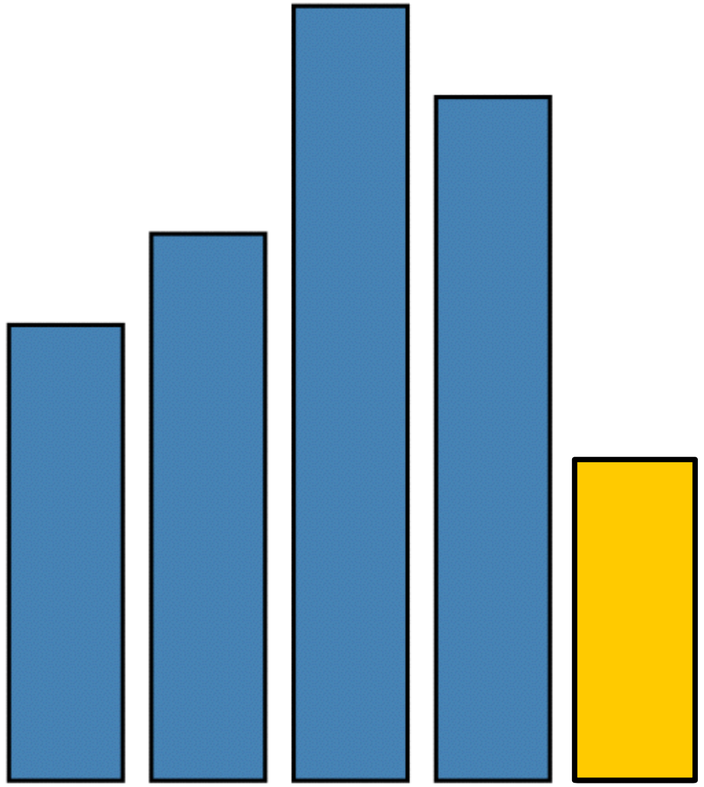
RECT $\lambda : D \rightarrow R$

data	1	1.2	1.7	1.5	0.7
visible	true				
left	3 * 25				
bottom	0				
width	20				
height	1.5 * 80				
fillStyle	blue				
strokeStyle	black				
lineWidth	1.5				
...	...				



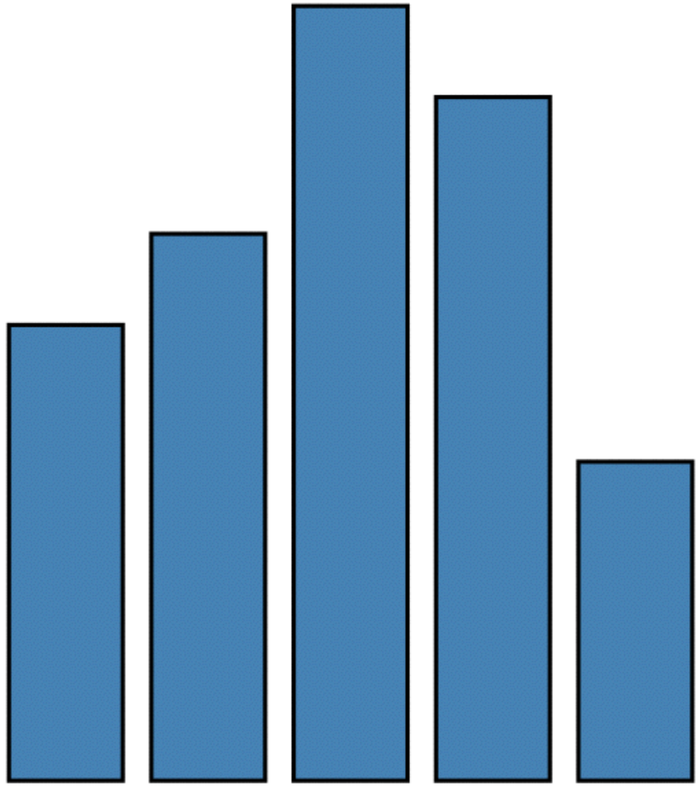
RECT $\lambda : D \rightarrow R$

data	1	1.2	1.7	1.5	0.7
visible	true				
left	4 * 25				
bottom	0				
width	20				
height	0.7 * 80				
fillStyle	blue				
strokeStyle	black				
lineWidth	1.5				
...	...				

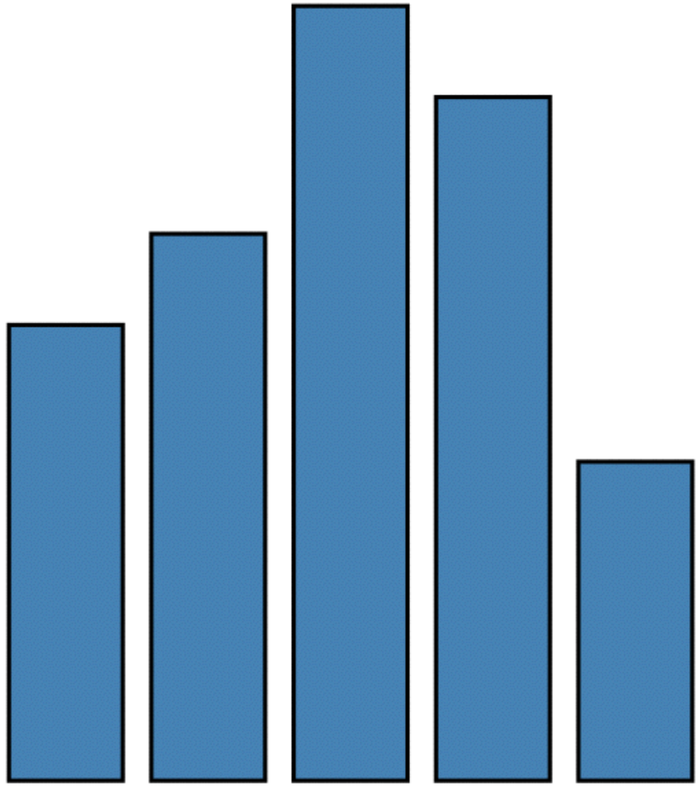


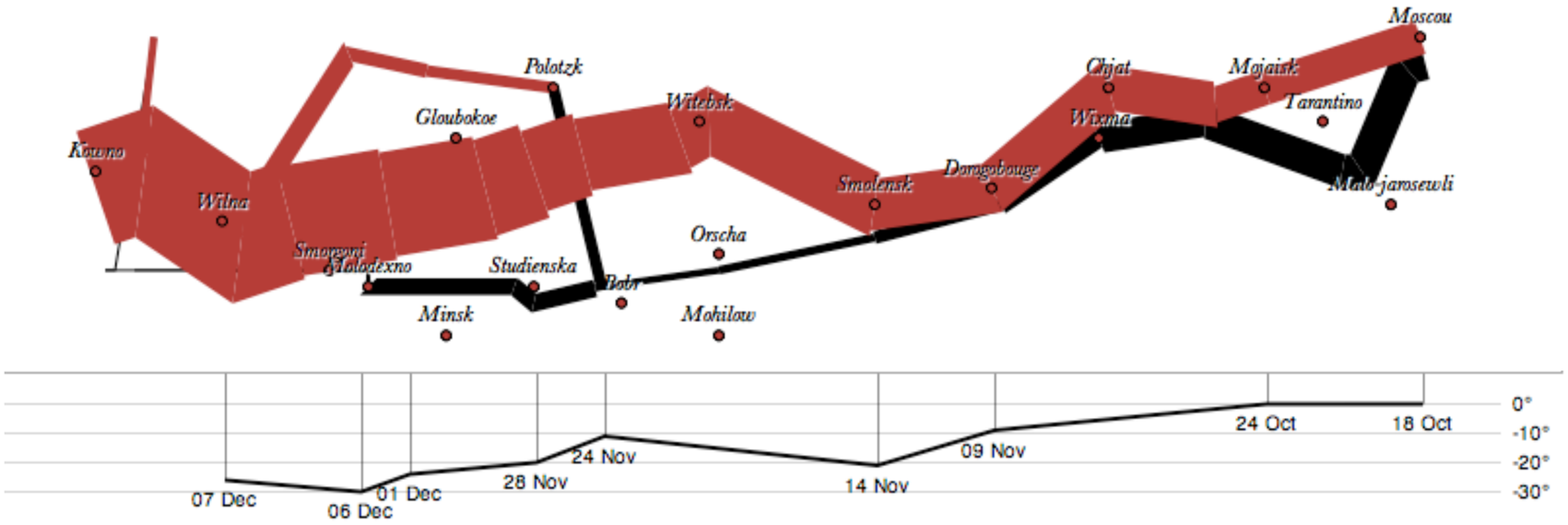
RECT $\lambda : D \rightarrow R$

data	1	1.2	1.7	1.5	0.7
visible	true				
left	λ: index * 25				
bottom	0				
width	20				
height	λ: datum * 80				
fillStyle	blue				
strokeStyle	black				
lineWidth	1.5				
...	...				



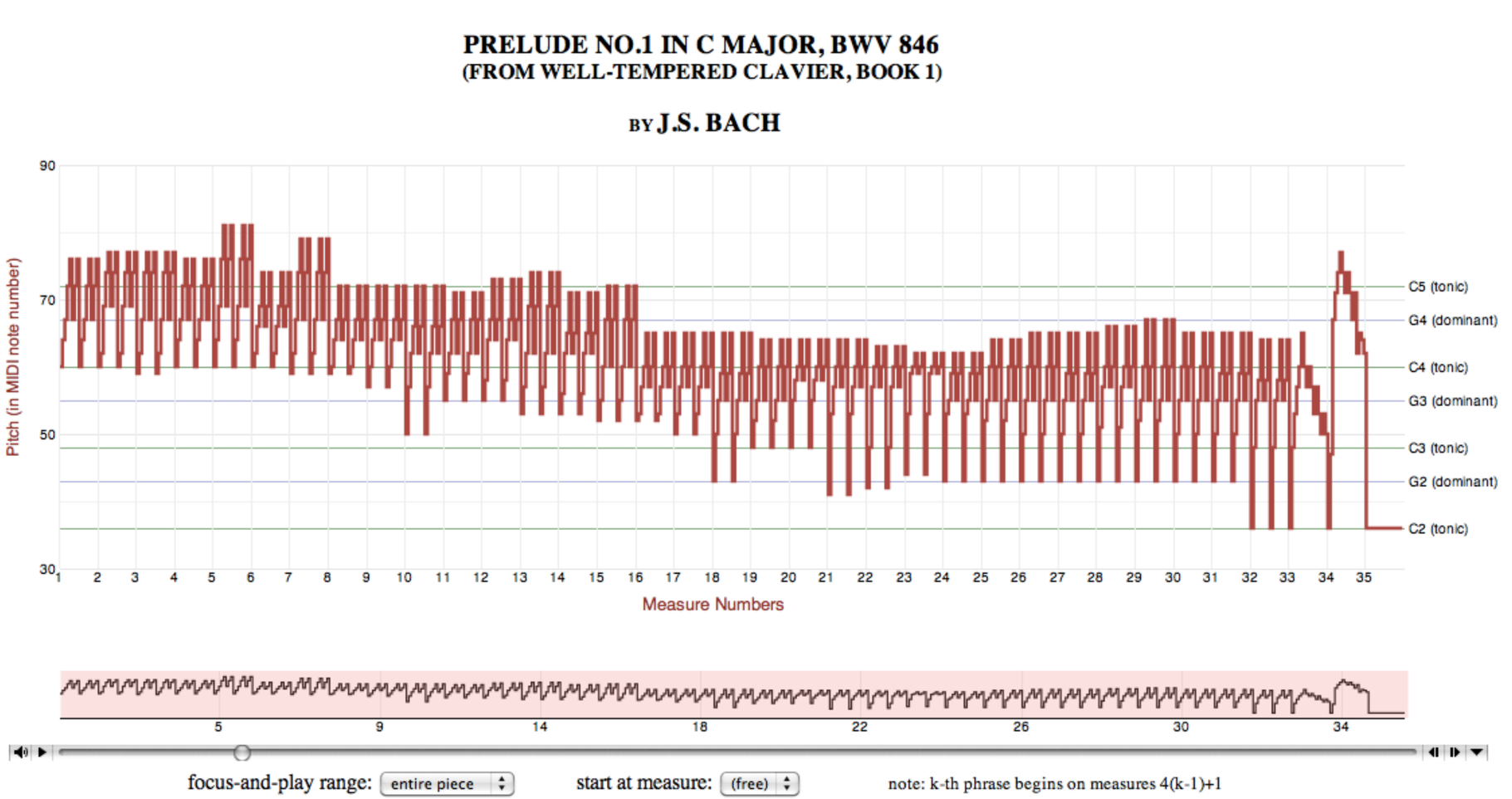
```
var vis = new pv.Panel();
vis.add(pv.Bar)
  .data([1, 1.2, 1.7, 1.5, 0.7])
  .visible(true)
  .left((d) => this.index * 25);
  .bottom(0)
  .width(20)
  .height((d) => d * 80)
  .fillStyle(“blue”)
  .strokeStyle(“black”)
  .lineWidth(1.5);
vis.render();
```



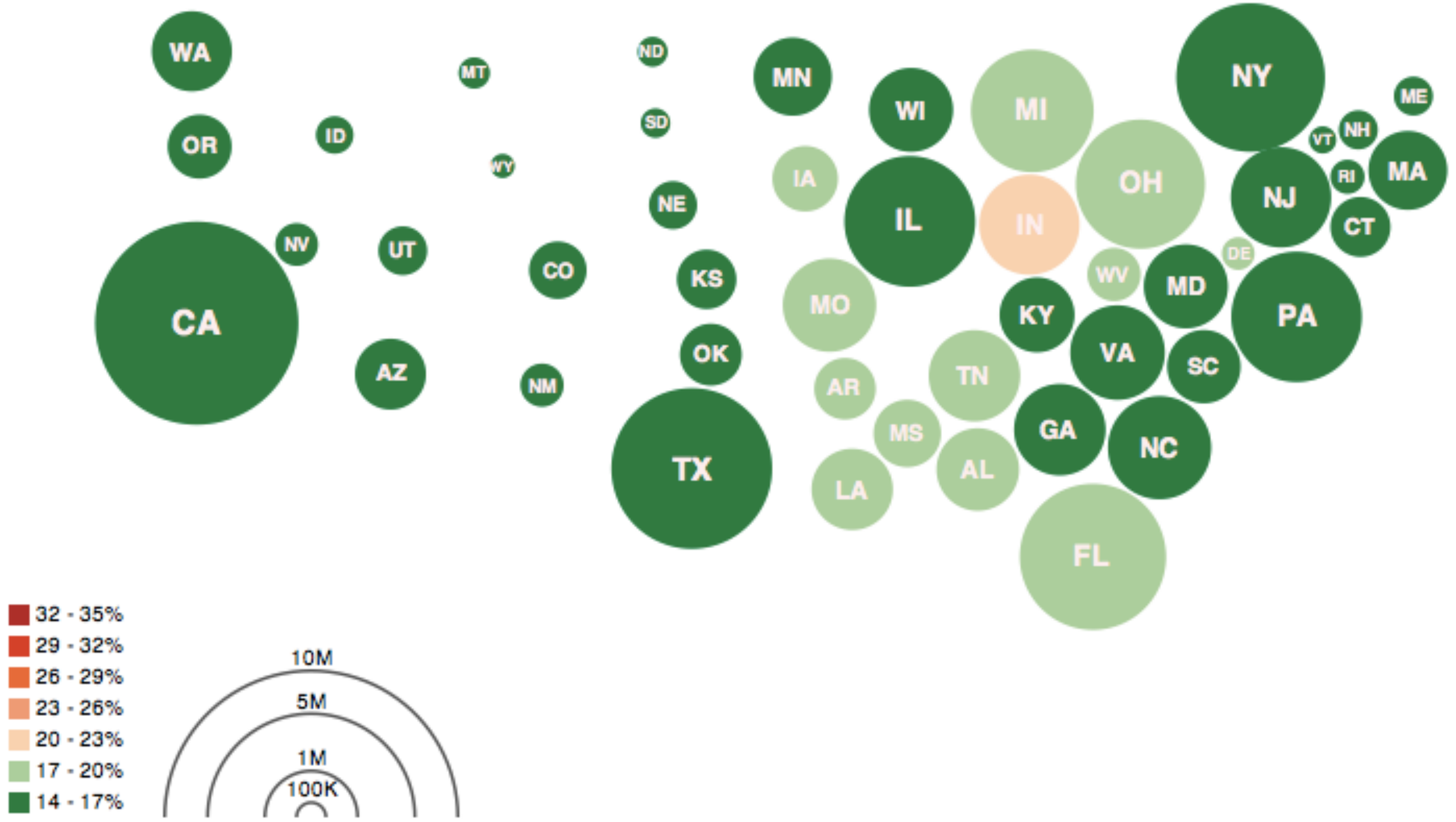



```
var army = pv.nest(napoleon.army, "dir", "group“);
var vis = new pv.Panel();

var lines = vis.add(pv.Panel).data(army);
lines.add(pv.Line)
  .data(() => army[this.idx])
  .left(lon).top(lat).size((d) => d.size/8000)
  .strokeStyle(() => color[army[paneIndex][0].dir]);

vis.add(pv.Label).data(napoleon.cities)
  .left(lon).top(lat)
  .text((d) => d.city).font("italic 10px Georgia")
  .textAlign("center").textBaseline("middle");
```

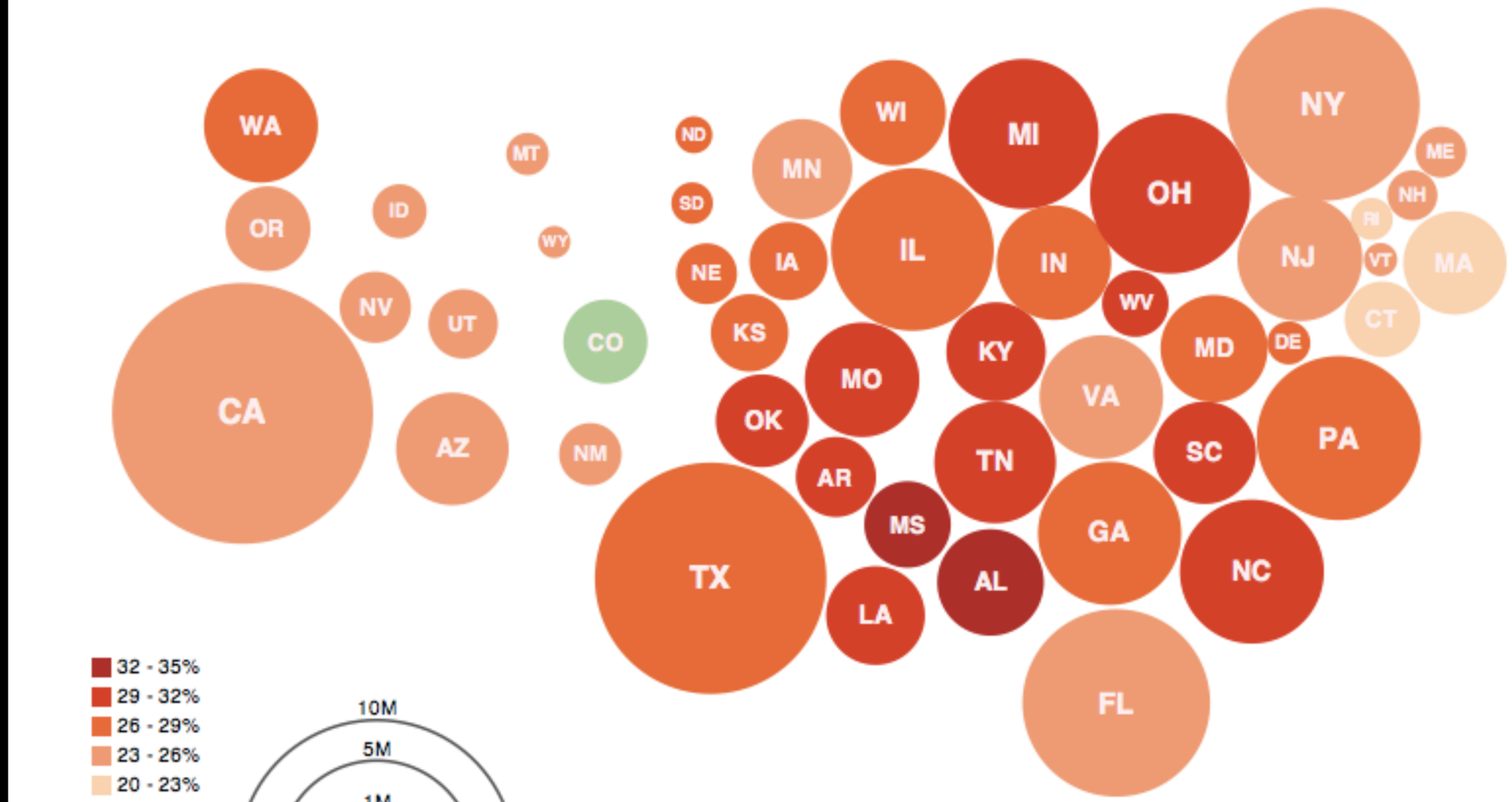
```
vis.add(pv.Rule).data([0,-10,-20,-30])
   .top((d) => 300 - 2*d - 0.5).left(200).right(150)
   .lineWidth(1).strokeStyle("#ccc")
   .anchor("right").add(pv.Label)
     .font("italic 10px Georgia")
     .text((d) => d+"°").textBaseline("center");

vis.add(pv.Line).data(napoleon.temp)
   .left(lon).top(tmp) .strokeStyle("#0")
  .add(pv.Label)
    .top((d) => 5 + tmp(d))
    .text((d) => d.temp+"° "+d.date.substr(0,6))
```



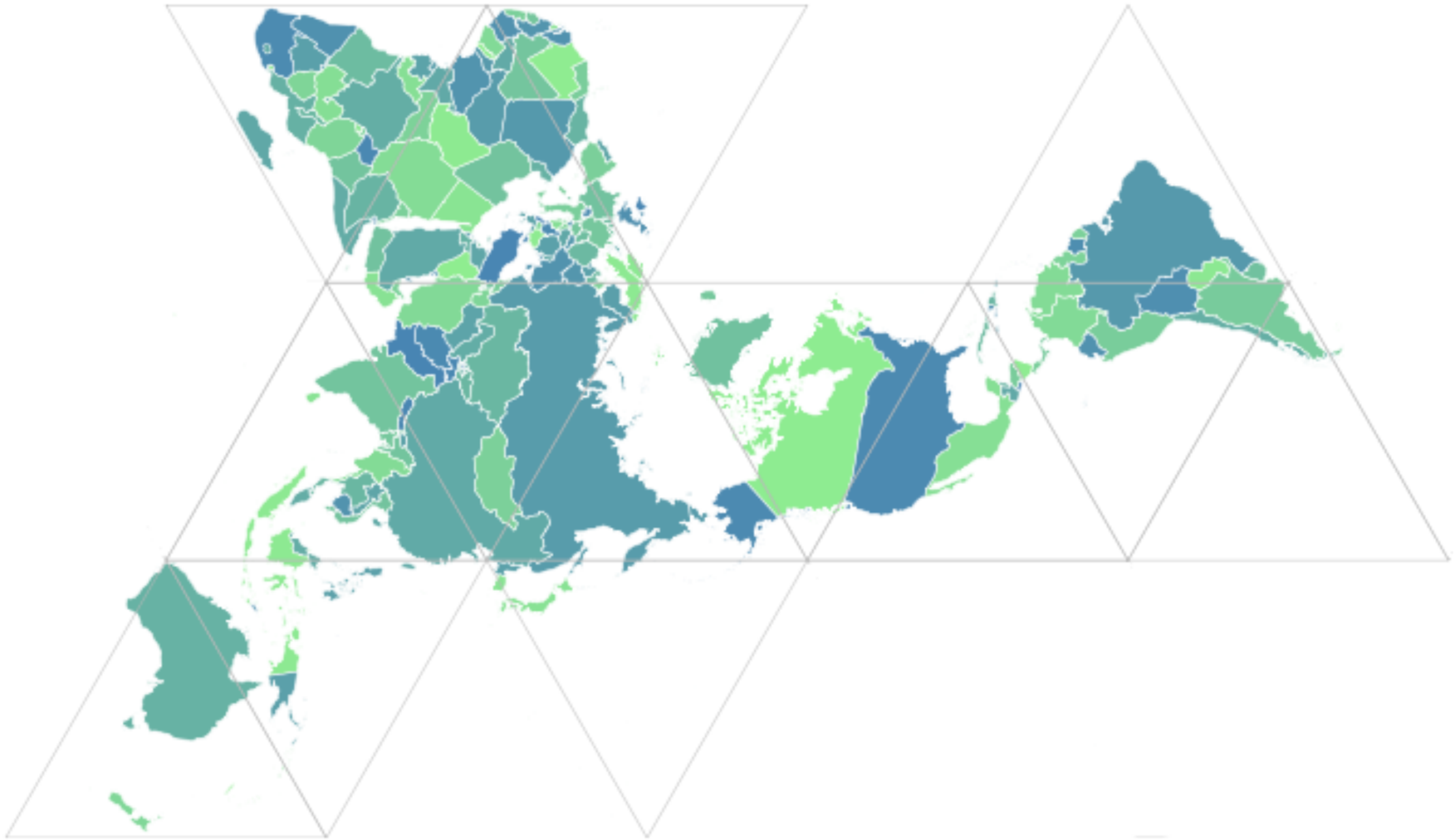
Bach's Prelude #1 in C Major | Jieun Oh



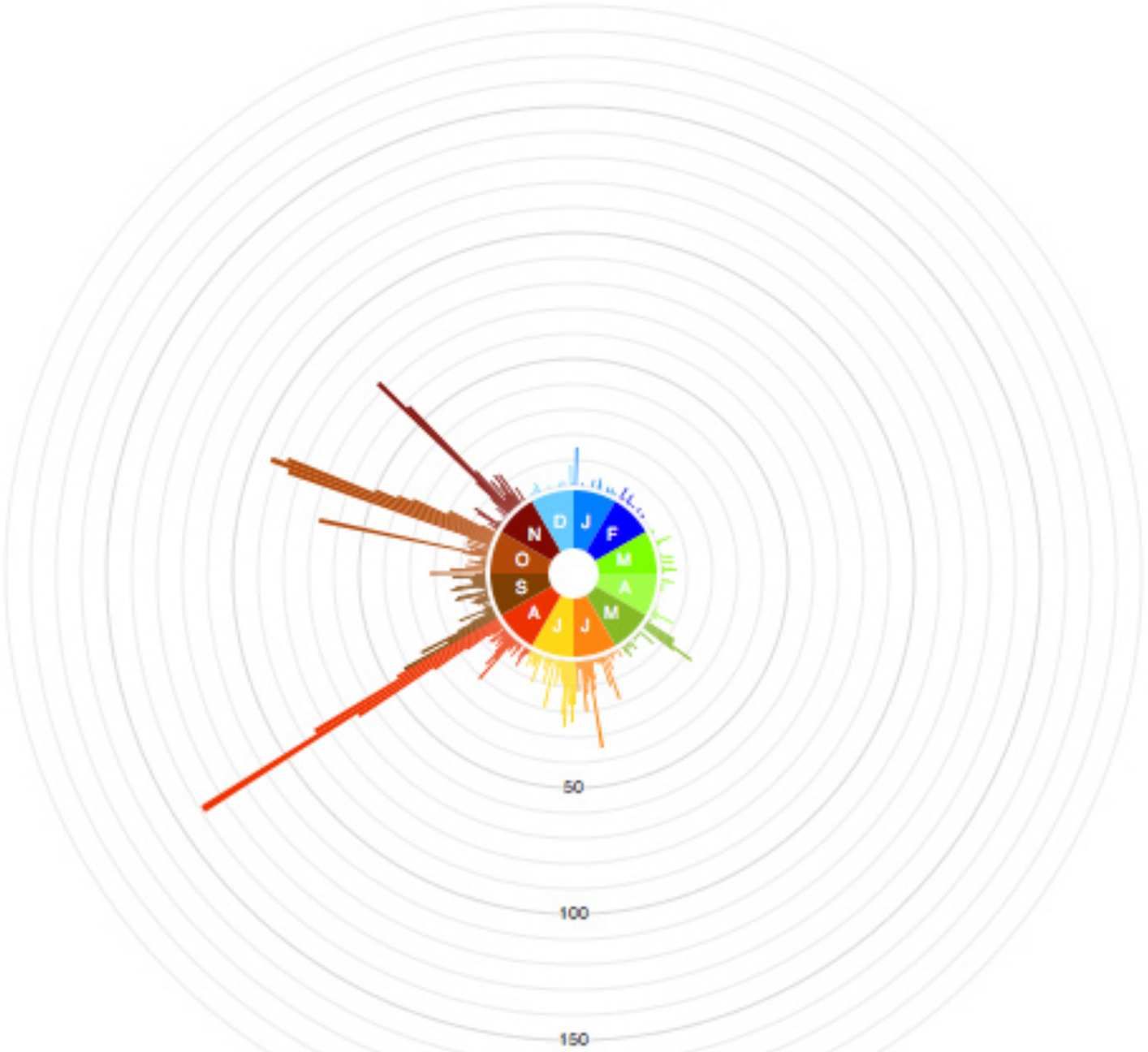
Obesity Map | Vadim Ogievetsky



Obesity Map | Vadim Ogievetsky

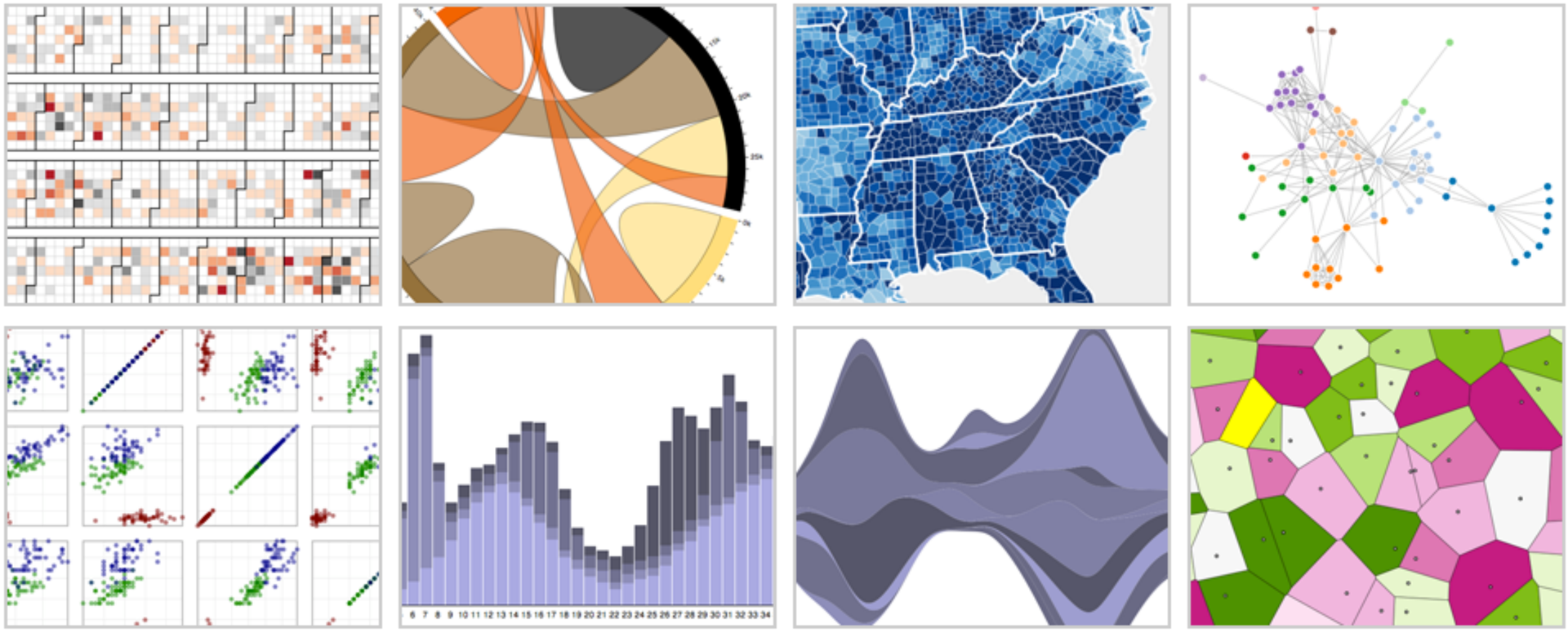


Dymaxion Maps | Vadim Ogievetsky



FlickrSeason | Ken-Ichi Ueda

d3.js Data-Driven Documents



with **Mike Bostock** & Vadim Ogievetsky

Protovis

Specialized mark types

+ Streamlined design
- Limits expressiveness
- More overhead (slower)
- Harder to debug
- Self-contained model

Specify a scene (nouns)

+ Quick for static vis
- Delayed evaluation
- Animation, interaction are more cumbersome

Protovis

Specialized mark types

+ Streamlined design
- Limits expressiveness
- More overhead (slower)
- Harder to debug
- Self-contained model

Specify a scene (nouns)

+ Quick for static vis
- Delayed evaluation
- Animation, interaction are more cumbersome

D3

Bind data to DOM

- Exposes SVG/CSS/…
+ Exposes SVG/CSS/…
+ Less overhead (faster)
+ Debug in browser
+ Use with other tools

Transform a scene (verbs)

- More complex model
+ Immediate evaluation
+ Dynamic data, anim, and interaction natural

D3 Selections

The core abstraction in D3 is a ***selection***.

D3 Selections

The core abstraction in D3 is a ***selection***.

```
// Add and configure an SVG element
var svg = d3.append(“svg”)      // add new SVG to page body
    .attr(“width”, 500)         // set SVG width to 500px
    .attr(“height”, 300);       // set SVG height to 300px
```

D3 Selections

The core abstraction in D3 is a ***selection***.

```
// Add and configure an SVG element
var svg = d3.append(“svg”)      // add new SVG to page body
    .attr(“width”, 500)         // set SVG width to 500px
    .attr(“height”, 300);       // set SVG height to 300px

// Select & update existing rectangles contained in the SVG element
svg.selectAll(“rect”)           // select all SVG rectangles
    .attr(“width”, 100)         // set rect widths to 100px
    .style(“fill”, “steelblue”); // set rect fill colors
```

Data Binding

Selections can ***bind*** **data and DOM elements**.

```
var values = [ {…}, {…}, {…}, … ]; // input data as JS objects
```

Data Binding

Selections can ***bind*** **data and DOM elements**.

```
var values = [ {…}, {…}, {…}, … ]; // input data as JS objects

// Select SVG rectangles and bind them to data values.
var bars = svg.selectAll(“rect.bars”).data(values);
```

Data Binding

Selections can ***bind*** **data and DOM elements**.

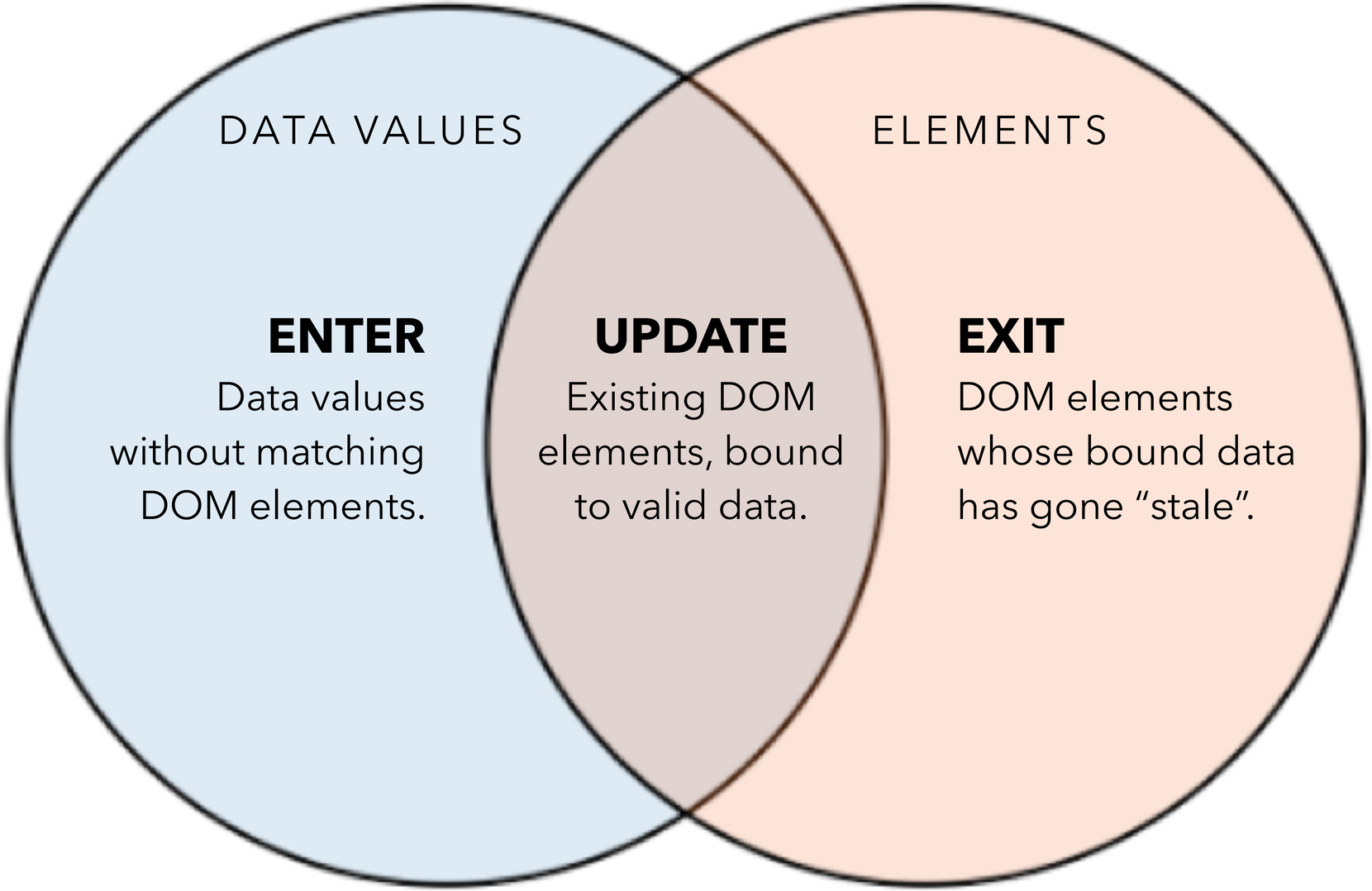
```
var values = [ {…}, {…}, {…}, … ]; // input data as JS objects

// Select SVG rectangles and bind them to data values.
var bars = svg.selectAll(“rect.bars”).data(values);

// What if the DOM elements don’t exist yet? The enter set represents data
// values that do not yet have matching DOM elements.
bars.enter().append(“rect”).attr(“class”, “bars”);
```

Data Binding

Selections can ***bind*** **data and DOM elements**.

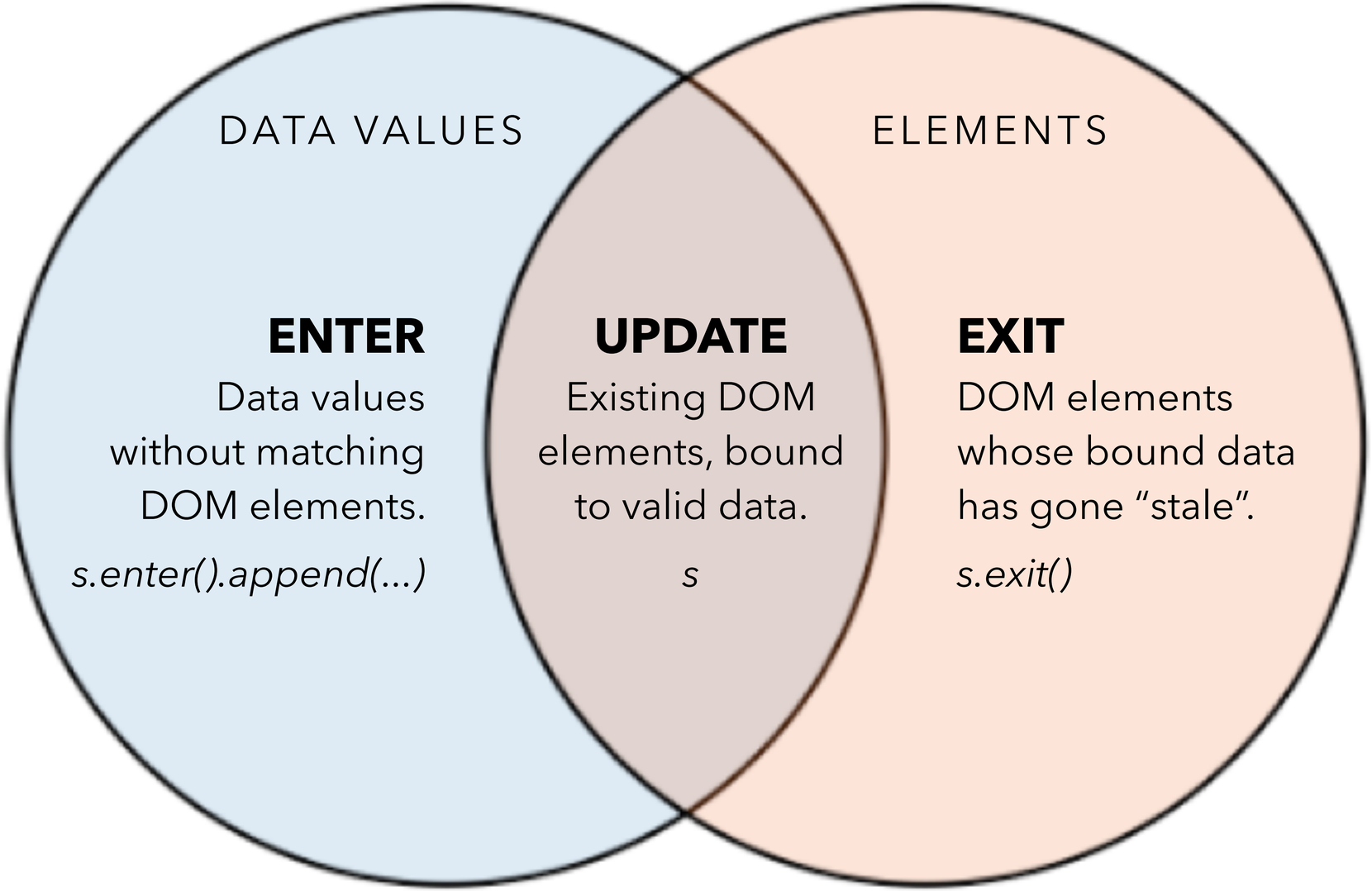
```
var values = [ {…}, {…}, {…}, … ]; // input data as JS objects

// Select SVG rectangles and bind them to data values.
var bars = svg.selectAll(“rect.bars”).data(values);

// What if the DOM elements don’t exist yet? The enter set represents data
// values that do not yet have matching DOM elements.
bars.enter().append(“rect”).attr(“class”, “bars”);

// What if data values are removed? The exit set is a selection of existing
// DOM elements who no longer have matching data values.
bars.exit().remove();
```

The Data Join



The Data Join

var s = d3.selectAll(...).data(...)



D3 Modules

Data Parsing / Formatting (JSON, CSV, …)
Shape Helpers (arcs, curves, areas, symbols, …)
Scale Transforms (linear, log, ordinal, …)
Color Spaces (RGB, HSL, LAB, …)
Animated Transitions (tweening, easing, …)
Geographic Mapping (projections, clipping, …)
Layout Algorithms (stack, pie, force, trees, …)
Interactive Behaviors (brush, zoom, drag, …)

Chart Typologies

Excel, Many Eyes, Google Charts

Visual Analysis Grammars

VizQL, ggplot2

Visualization Grammars

Protovis, D3.js

Component Architectures

Prefuse, Flare, Improvise, VTK

Graphics APIs

Processing, OpenGL, Java2D

A Visualization Tool Stack

Chart Typologies

Excel, Many Eyes, Google Charts

Visual Analysis Grammars

VizQL, ggplot2

Visualization Grammars

Protovis, D3.js

Component Architectures

Prefuse, Flare, Improvise, VTK

Graphics APIs

Processing, OpenGL, Java2D

Chart Typologies

Excel, Many Eyes, Google Charts

Charting Tools

Visual Analysis Grammars

VizQL, ggplot2

Visualization Grammars

Protovis, D3.js

Declarative Languages

Component Architectures

Prefuse, Flare, Improvise, VTK

Graphics APIs

Processing, OpenGL, Java2D

Programming Toolkits

Chart Typologies

Excel, Many Eyes, Google Charts

Charting Tools

Visual Analysis Grammars

VizQL, ggplot2

Visualization Grammars

Protovis, D3.js

Declarative Languages

Component Architectures

Prefuse, Flare, Improvise, VTK

Graphics APIs

Processing, OpenGL, Java2D

Programming Toolkits

https://en.wikipedia.org/wiki/DeLorean_time_machine#/media/File:TeamTimeCar.com-BTTF_DeLorean_Time_Machine-OtoGodfrey.com-JMortonPhoto.com-07.jpg

Declarative Languages

- For dialog boxes, forms and menus, just list the contents
- But need to also list *a lot* of properties for each field
- Used by
  - CMU's COUSIN system (~1985), and
  - Apollo's Domain/Dialog (later HP/Apollo's Open Dialog) ~1985-87.
  - Html forms (to some extent)
- Goals:
  - less overall effort
  - Better separation UI and application
  - Multiple interfaces for same application
  - Consistency since UI part uses same package

Slides from 2/17/ 2020

What is a Declarative Language?

What is a Declarative Language?

Programming by describing *what*, not *how*

What is a Declarative Language?

Programming by describing *what*, not *how*

Separate **specification** (*what you want*) from **execution** (*how it should be computed*)

What is a Declarative Language?

Programming by describing *what*, not *how*

Separate **specification** (*what you want*) from **execution** (*how it should be computed*)

In contrast to **imperative programming**, where you must give explicit steps.

What is a Declarative Language?

Programming by describing *what*, not *how*

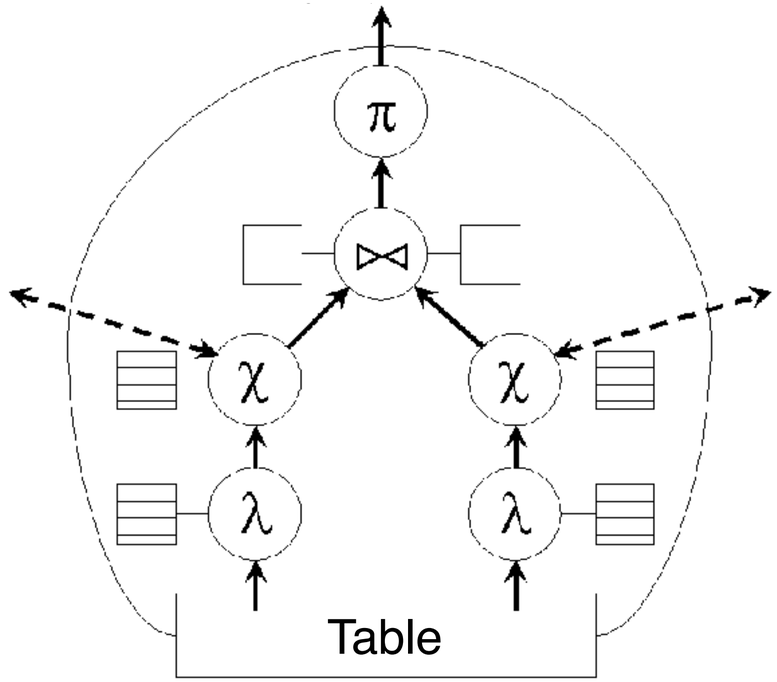
Separate **specification** (*what you want*) from **execution** (*how it should be computed*)

In contrast to **imperative programming**, where you must give explicit steps.

```
d3.selectAll("rect")
  .data(my_data)
 .enter().append("rect")
  .attr("x", (d) => xscale(d.foo))
  .attr("y", (d) => yscale(d.bar))
```


```
<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
"http://www.w3.org/TR/html4/loose.dtd">
<!--[if IE]><![endif]-->
▼<html>
  ▶<head>…</head>
  ▼<body id="home" style="visibility: visible; ">
    <script src="http://connect.facebook.net/en_US/all.js"></scrip
    <div id="fb-root"></div>
    <a name="top"></a>
    ▼<div id="shell">
      ▶<ul id="memberTools">…</ul>
        <!-- ADXINFO classification="text_ad" campaign="nyt2010-circ-
      ▶<div class="tabsContainer">…</div>
        <!-- close .tabsContainer -->
      ▶<div id="page" class="tabContent active">…</div>
        <!--close page -->
    </div>
    <!--close shell -->
  ▶<script type="text/javascript" language="JavaScript">…</script
    <img src="/adx/bin/clientside/
    1b1a1755Q2F,.tWdC,KiQ2BCQ5C6iSM.KCYQ2FMadYYQ24xN!Q2B!xQ24WN4aN
    <script src="http://graphics8.nytimes.com/js/test/yto/wt_capi.
    charset="utf-8"></script>
    <span id="to_script_wt">an>
  ▶<script type="text/javascript">…</script
    <img src="http://graphics8.nytimes.com/ads/blank.gif">
    <script type="text/javascript" src="http://graphics8.nytimes.c
```

HTML / CSS




```
SELECT customer_id, customer_name,
    COUNT(order_id) as total
FROM customers
INNER JOIN orders ON
    customers.customer_id
    = orders.customer_id
GROUP BY customer_id, customer_name
HAVING COUNT(order_id) > 5
ORDER BY COUNT(order_id) DESC
```

SQL

Why Declarative Languages?

Better visualization? *Smart defaults.*

Reuse. *Write-once, then re-apply.*

Performance. *Optimization, scalability.*

Portability. *Multiple devices, renderers, inputs.*

Programmatic generation.
Write programs which output visualizations.
Automated search & recommendation.

Chart Typologies

Excel, Many Eyes, Google Charts

Charting Tools

Visual Analysis Grammars

VizQL, ggplot2

Visualization Grammars

Protovis, D3.js

Declarative Languages

Component Architectures

Prefuse, Flare, Improvise, VTK

Graphics APIs

Processing, OpenGL, Java2D

Programming Toolkits

Chart Typologies

Excel, Many Eyes, Google Charts

Charting Tools

Visual Analysis Grammars

VizQL, ggplot2, ***Vega-Lite***

Visualization Grammars

Protovis, D3.js, ***Vega***

Declarative Languages

Component Architectures

Prefuse, Flare, Improvise, VTK

Graphics APIs

Processing, OpenGL, Java2D

Programming Toolkits

Chart Typologies

Excel, Many Eyes, Google Charts

?!

Charting Tools

Visual Analysis Grammars

VizQL, ggplot2, ***Vega-Lite***

Visualization Grammars

Protovis, D3.js, ***Vega***

Declarative Languages

Component Architectures

Prefuse, Flare, Improvise, VTK

Graphics APIs

Processing, OpenGL, Java2D

Programming Toolkits

Visual Analysis Grammars

VizQL, ggplot2, ***Vega-Lite***

Visualization Grammars

Protovis, D3.js, ***Vega***

Declarative Languages

Component Architectures

Prefuse, Flare, Improvise, VTK

Graphics APIs

Processing, OpenGL, Java2D

Programming Toolkits

Interactive Data Exploration

Tableau, ***Lyra, Voyager***

Graphical Interfaces

Visual Analysis Grammars

VizQL, ggplot2, ***Vega-Lite***

Visualization Grammars

Protovis, D3.js, ***Vega***

Declarative Languages

Component Architectures

Prefuse, Flare, Improvise, VTK

Graphics APIs

Processing, OpenGL, Java2D

Programming Toolkits

Vega-Lite

vega.github.io/vega-lite

Kanit "Ham" Wongsuphasawat

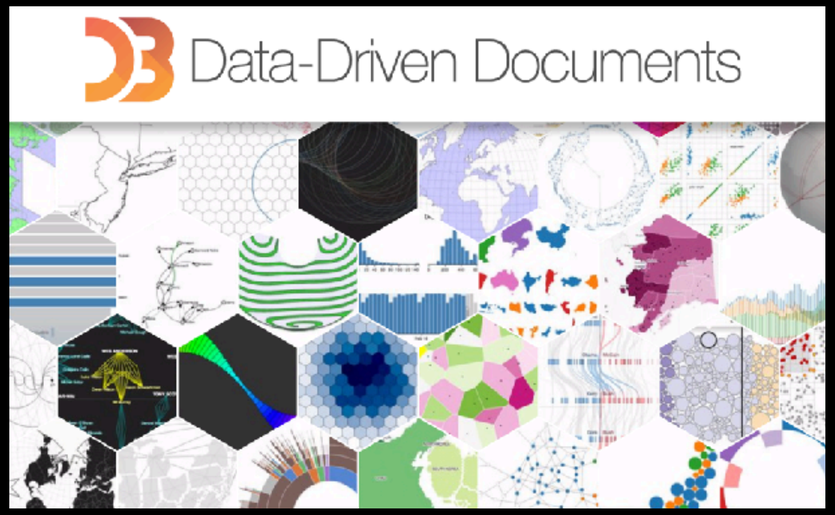
Dominik Moritz

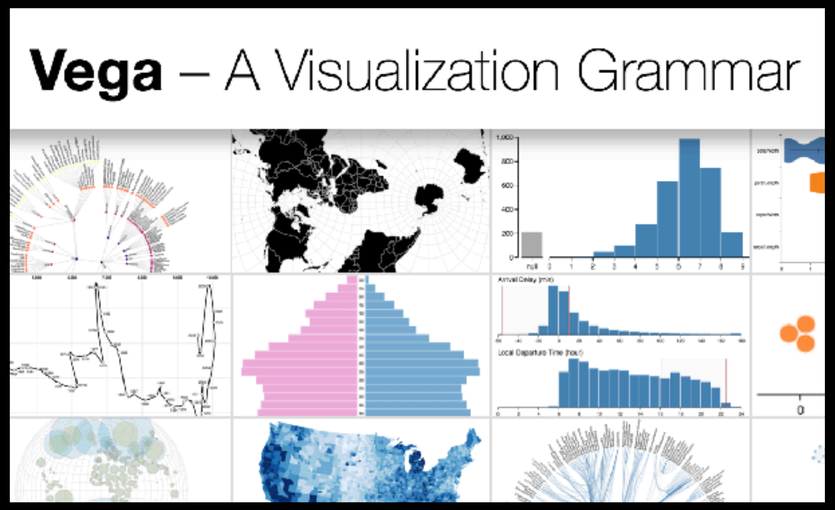
Jeffrey Heer

Arvind Satyanarayan

And many others...

Grammar of Graphics *for Customized Designs*

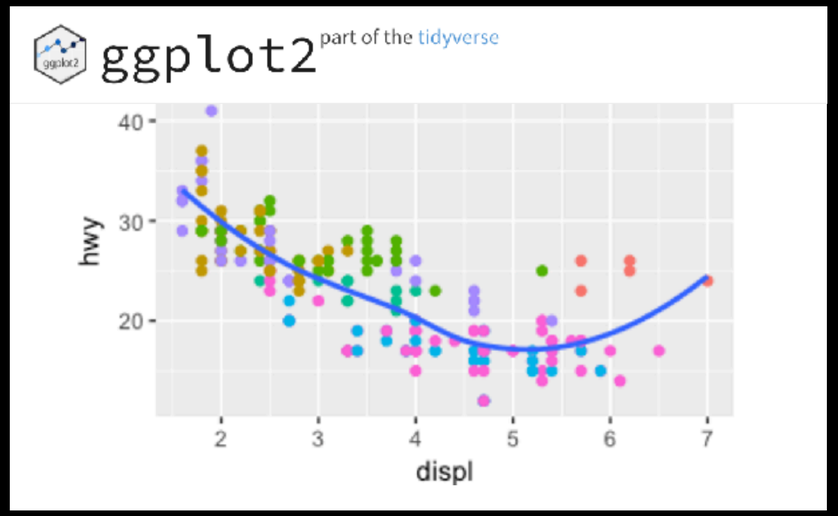


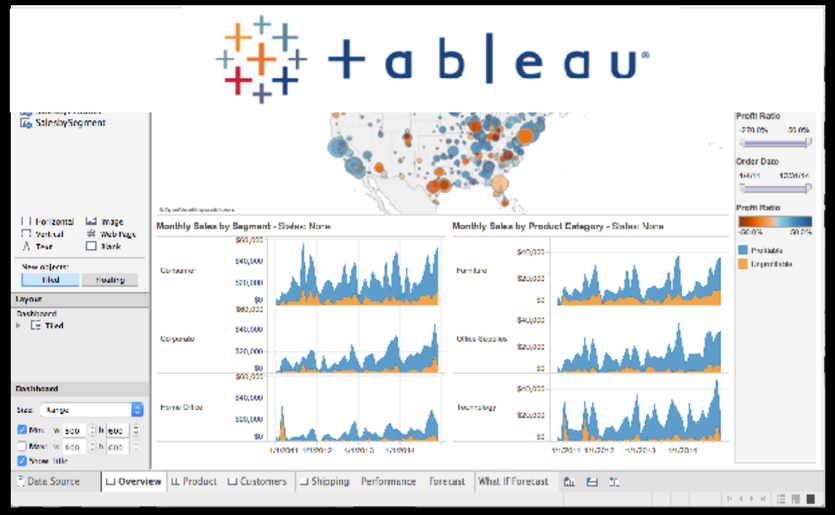


Offer **fine-grained control** for composing interactive graphics.

But require technical expertise and **verbose** specifications
(e.g., 100 LOCs for a bar chart)

Grammar of Graphics *for Exploration*





Facilitate **rapid exploration**
with **concise** specifications
by omitting low-level details.

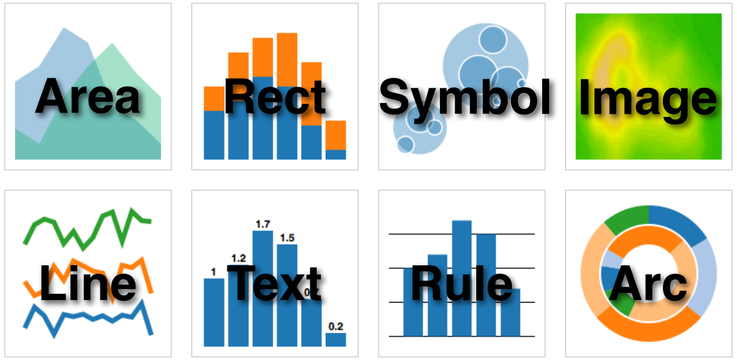
Infer **sensible defaults** and
allow customization
by overriding defaults.

Vega-Lite's Mission

Facilitate exploratory data analysis
with an ***expressive*** yet ***concise*** language
to specify ***interactive multi-view graphics***

Visualization Grammar

Data	Input data to visualize
Transforms	Grouping, stats, projection, layout
Scales	Map data values to visual values
Guides	Axes & legends visualize scales
Marks	Data-representative graphics



Specifying Visualizations in Vega-Lite

Abstract Data  Visual Representation

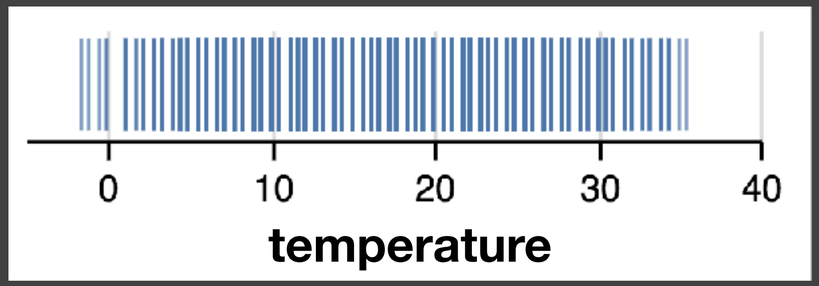
Specifying Visualizations in Vega-Lite

Abstract Data → Visual Representation

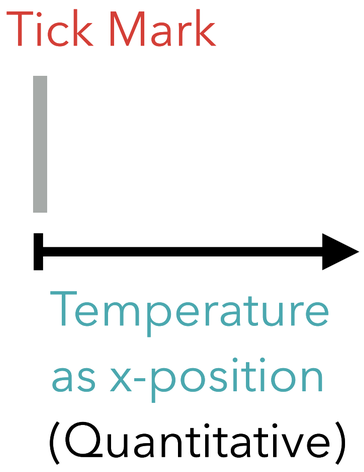
Weather Data for Seattle

date	temperature	precipitation	weather
1/1	10.6	10.9	"rain"
1/2	11.7	0.8	"drizzle"
1/3	12.2	10.2	"rain"
…	…	…	…

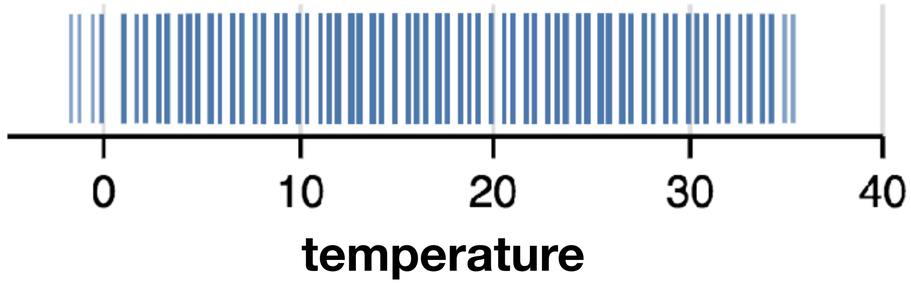
Strip Plot of Temperature



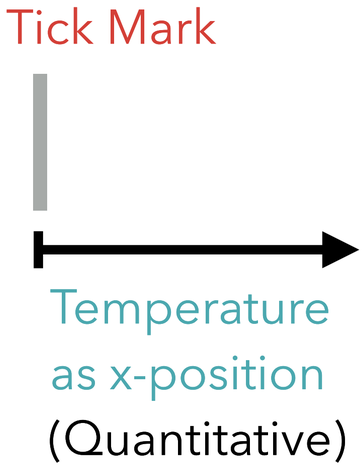
Strip Plot = (Tick with x=field)



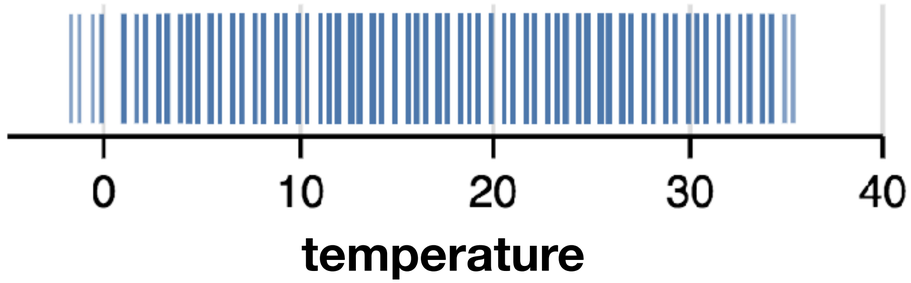

```
{
  data: {url: "weather-seattle.json"},
  mark: "tick",
  encoding: {
    x: {
      field: "temperature",
      type: "quantitative"
    }
  }
}
```



Strip Plot = (Tick with x=field)



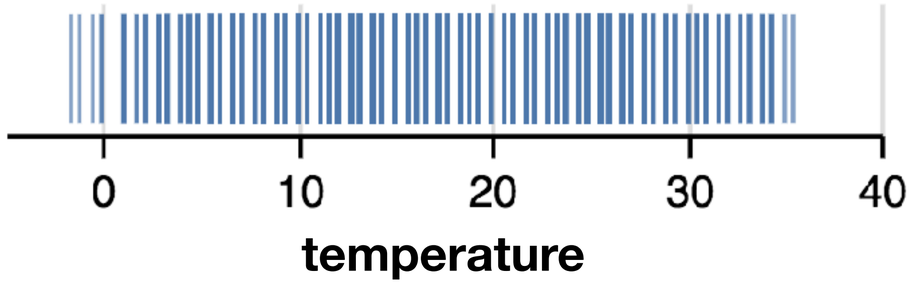

```
{
  data: {url: "weather-seattle.json"},
  mark: "tick",
  encoding: {
    x: {
      field: "temperature",
      type: "quantitative"
    }
  }
}
```



Vega-Lite is portable JSON specification

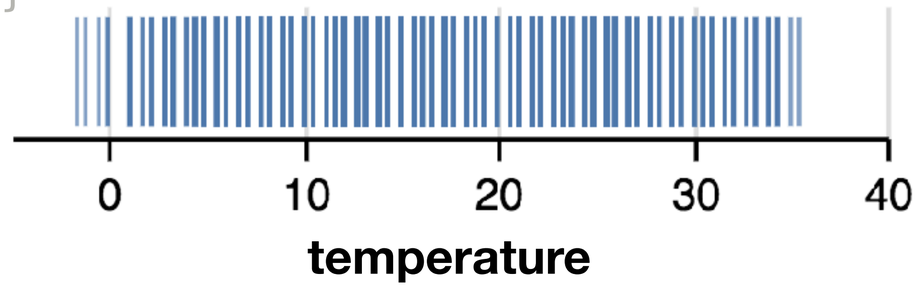
Strip Plot = (Tick with x=field)

```
{
  data: {url: "weather-seattle.json"},
  mark: "tick",
  encoding: {
    x: {
      field: "temperature",
      type: "quantitative"
    }
  }
}
```



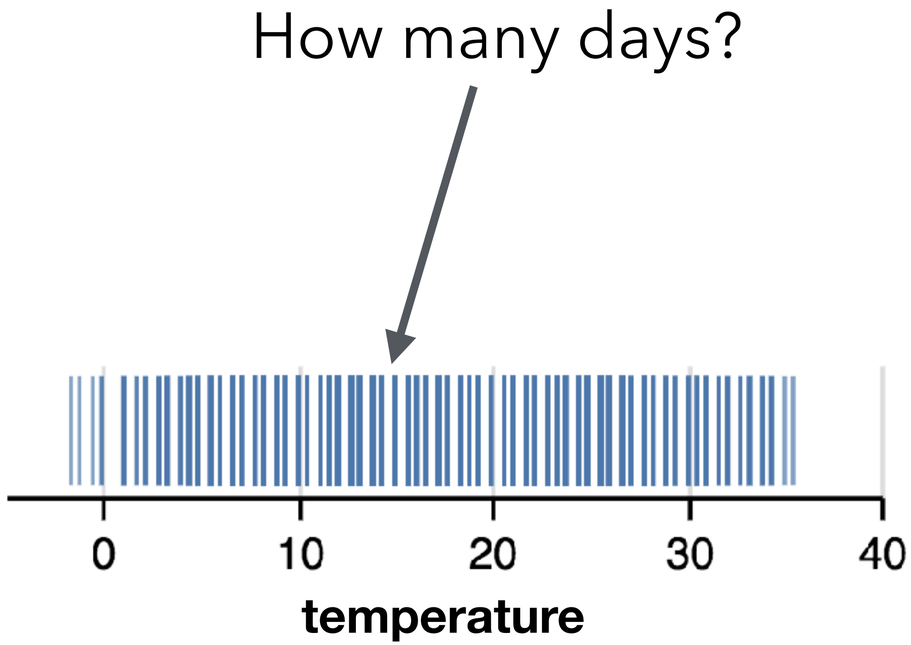
Strip Plot: Default Scales and Axes

```
{
  data: {url: "weather-seattle.json"},
  mark: "tick",
  encoding: {
    x: {
      field: "temperature",
      type: "quantitative",
      scale: {type: "linear", domain: [-10, 40], ...}
      axis: {title: "temperature", grid: true, ...}
    }
  }
}
```



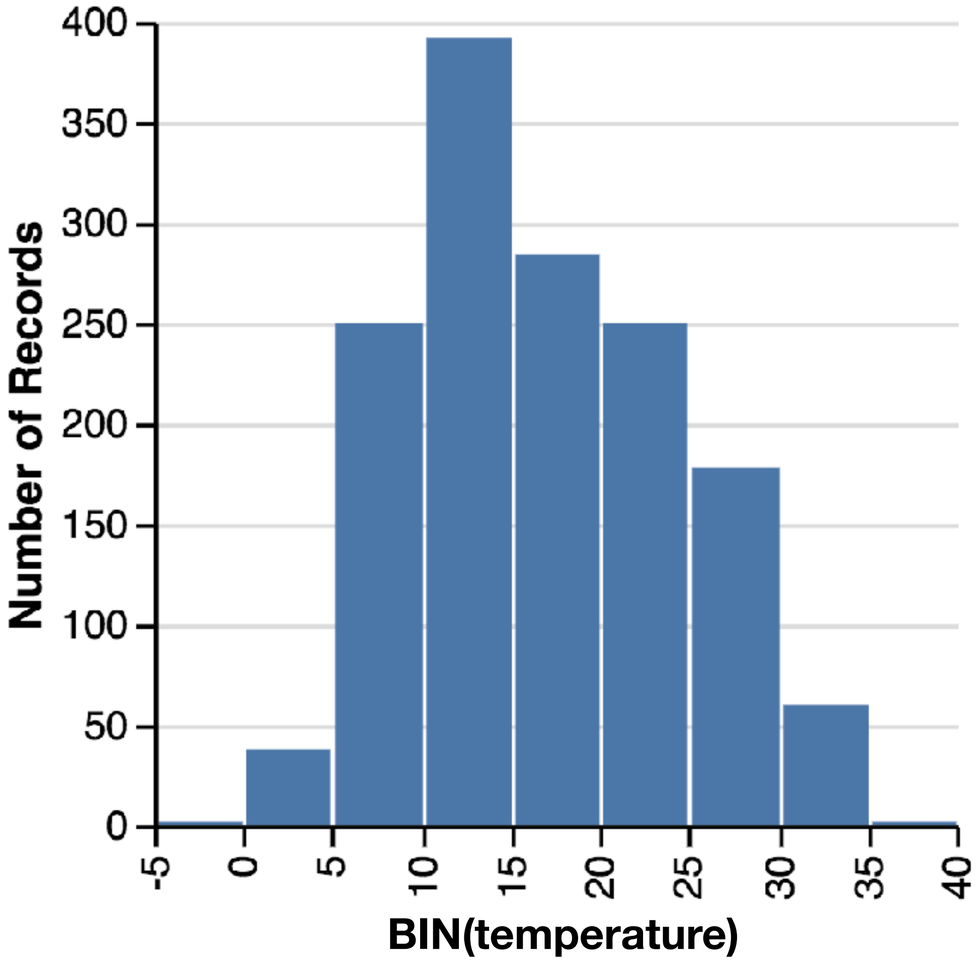
Strip Plot

```
{
  data: {url: "weather-seattle.json"},
  mark: "tick",
  encoding: {
    x: {
      field: "temperature",
      type: "quantitative"
    }
  }
}
```



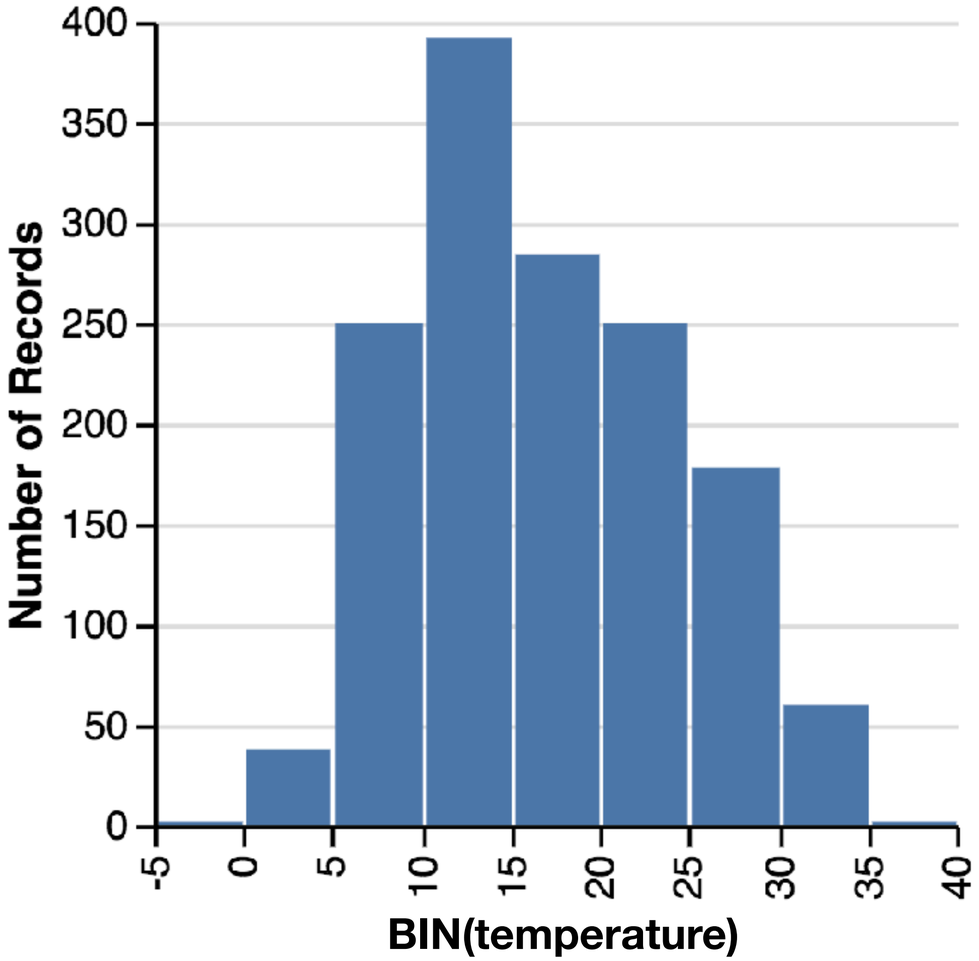
Histogram

Goal



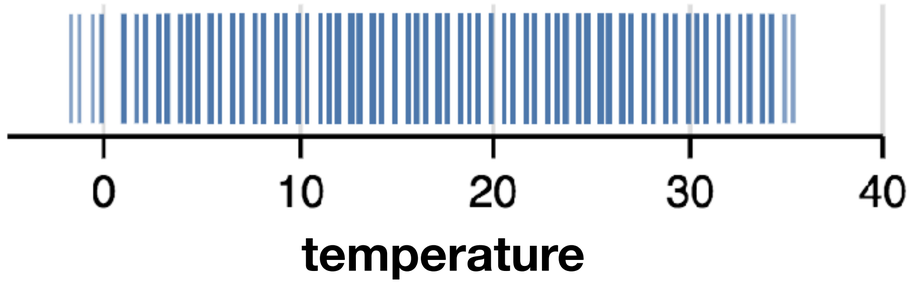
Histogram = (Bar with x=binned field, y=count)

Goal



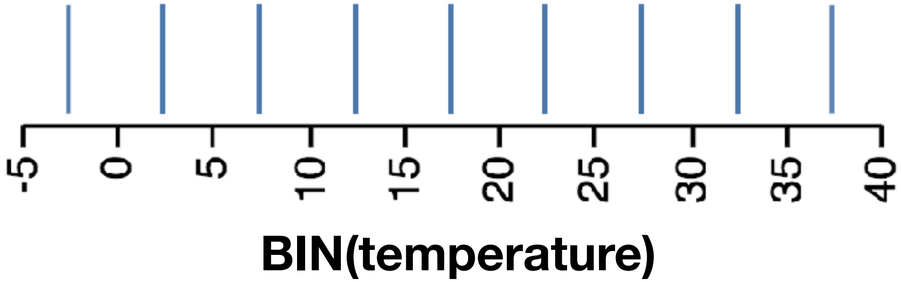
Histogram = (Bar with x=binned field, y=count)

```
{
  data: {url: "weather-seattle.json"},
  mark: "tick",
  encoding: {
    x: {
      field: "temperature",
      type: "quantitative"
    }
  }
}
```



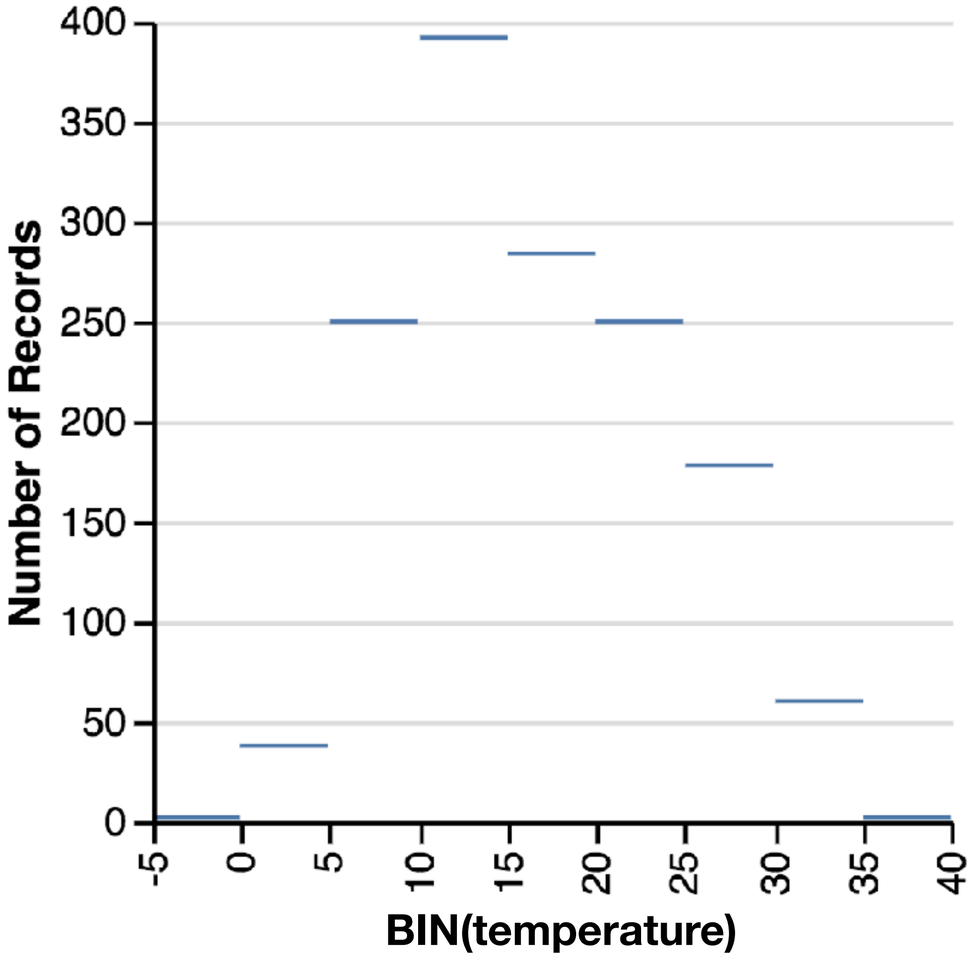
Histogram = (Bar with x=binned field, y=count)

```
{
  data: {url: "weather-seattle.json"},
  mark: "tick",
  encoding: {
    x: {
      bin: true,
      field: "temperature",
      type: "quantitative"
    }
  }
}
```



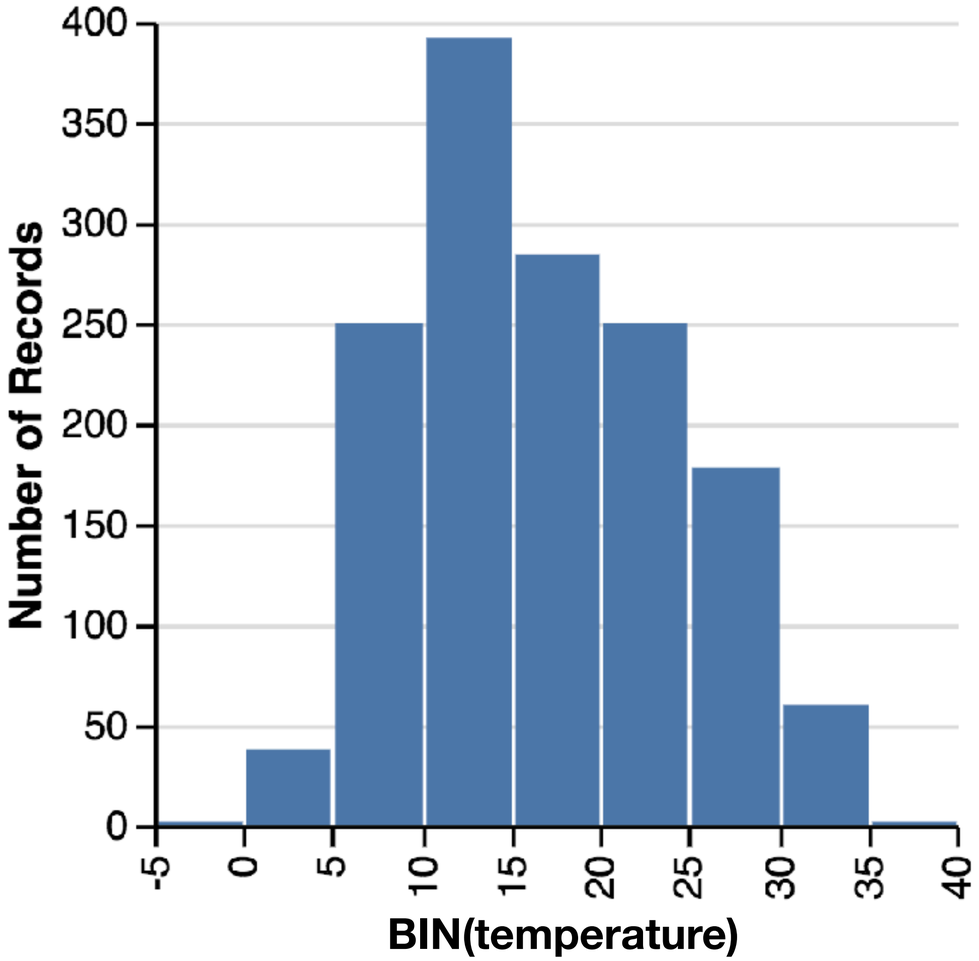
Histogram = (Bar with x=binned field, y=count)

```
{
  data: {url: "weather-seattle.json"},
  mark: "tick",
  encoding: {
    x: {
      bin: true,
      field: "temperature",
      type: "quantitative"
    },
    y: {
      aggregate: "count",
      type: "quantitative"
    }
  }
}
```



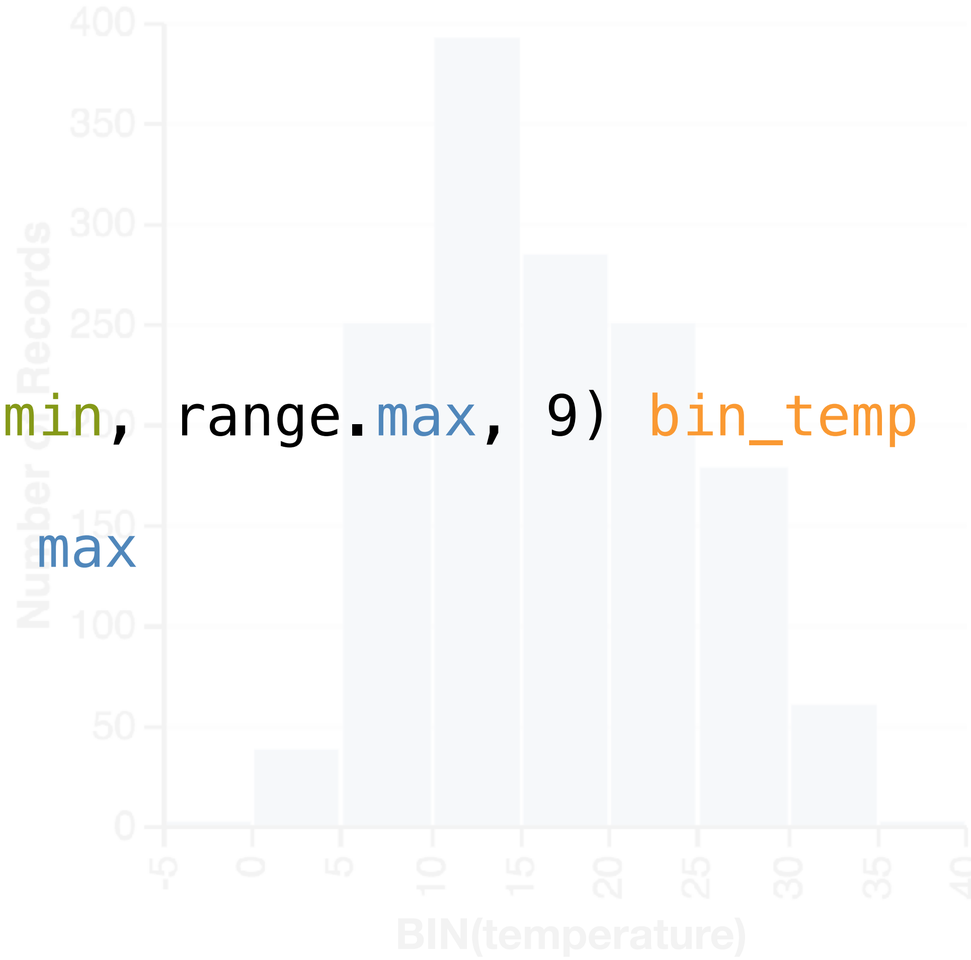
Histogram = (Bar with x=binned field, y=count)

```
{
  data: {url: "weather-seattle.json"},
  mark: "bar",
  encoding: {
    x: {
      bin: true,
      field: "temperature",
      type: "quantitative"
    },
   y: {
      aggregate: "count",
      type: "quantitative"
    }
  }
}
```



Histogram = (Bar with x=binned field, y=count)

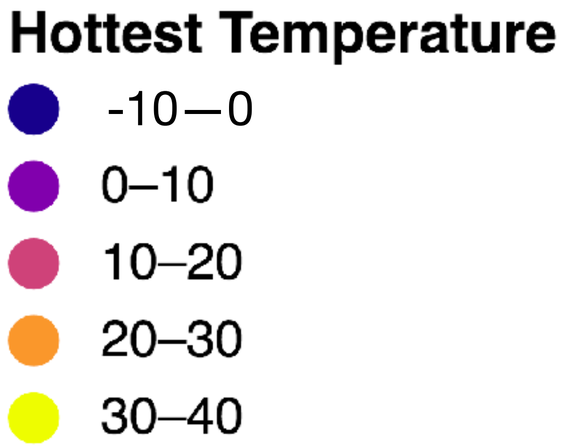
```
SELECT bin_temp, count(*)
FROM (
  SELECT floor(weather.temp, range.min, range.max, 9) bin_temp
  FROM weather, (
    SELECT min(temp) min, max(temp) max
  ) range
)
```

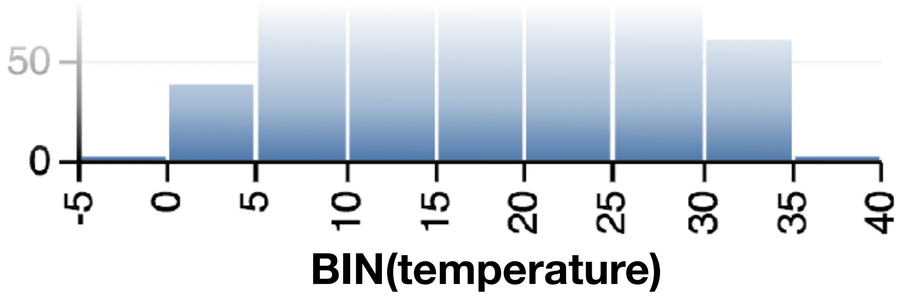


Sensible Defaults for Binning

Channel determines guide and bin parameters

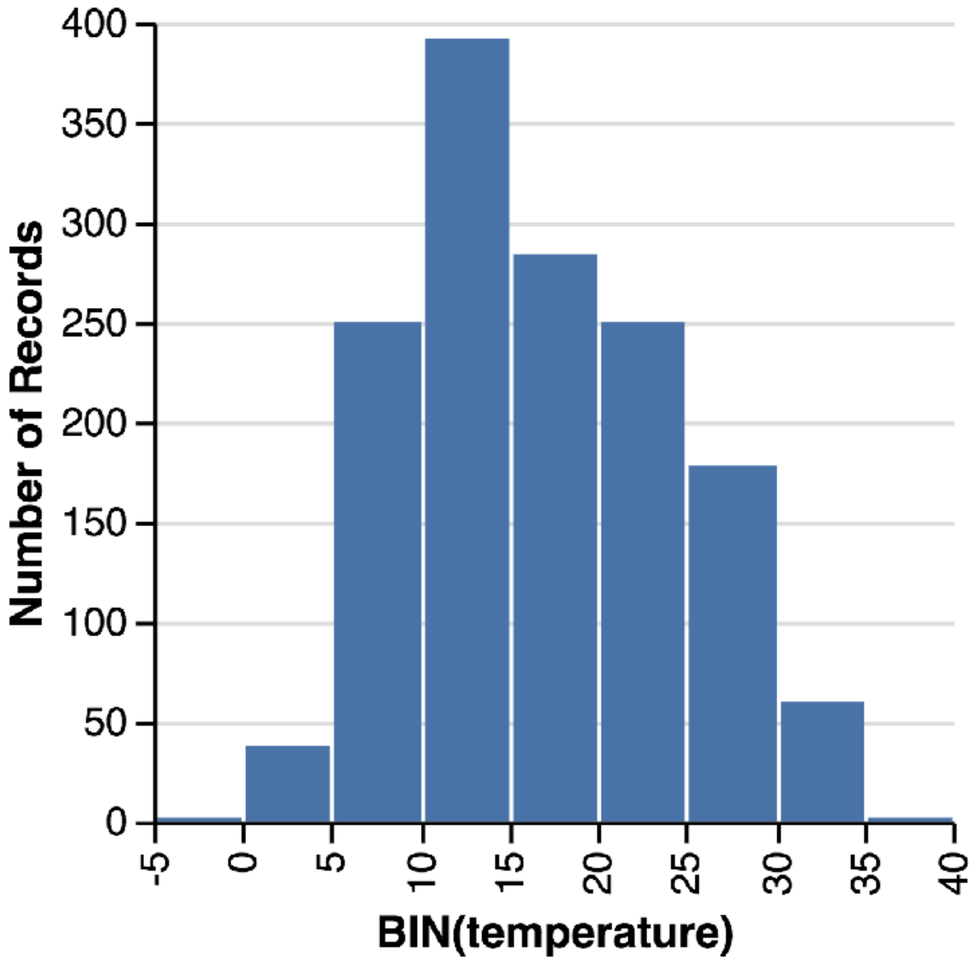
	Color/Opacity/Shape	Position
Guide	Legend with range labels	Quantitative axis
# of Bins	Fewer bins	Many bins





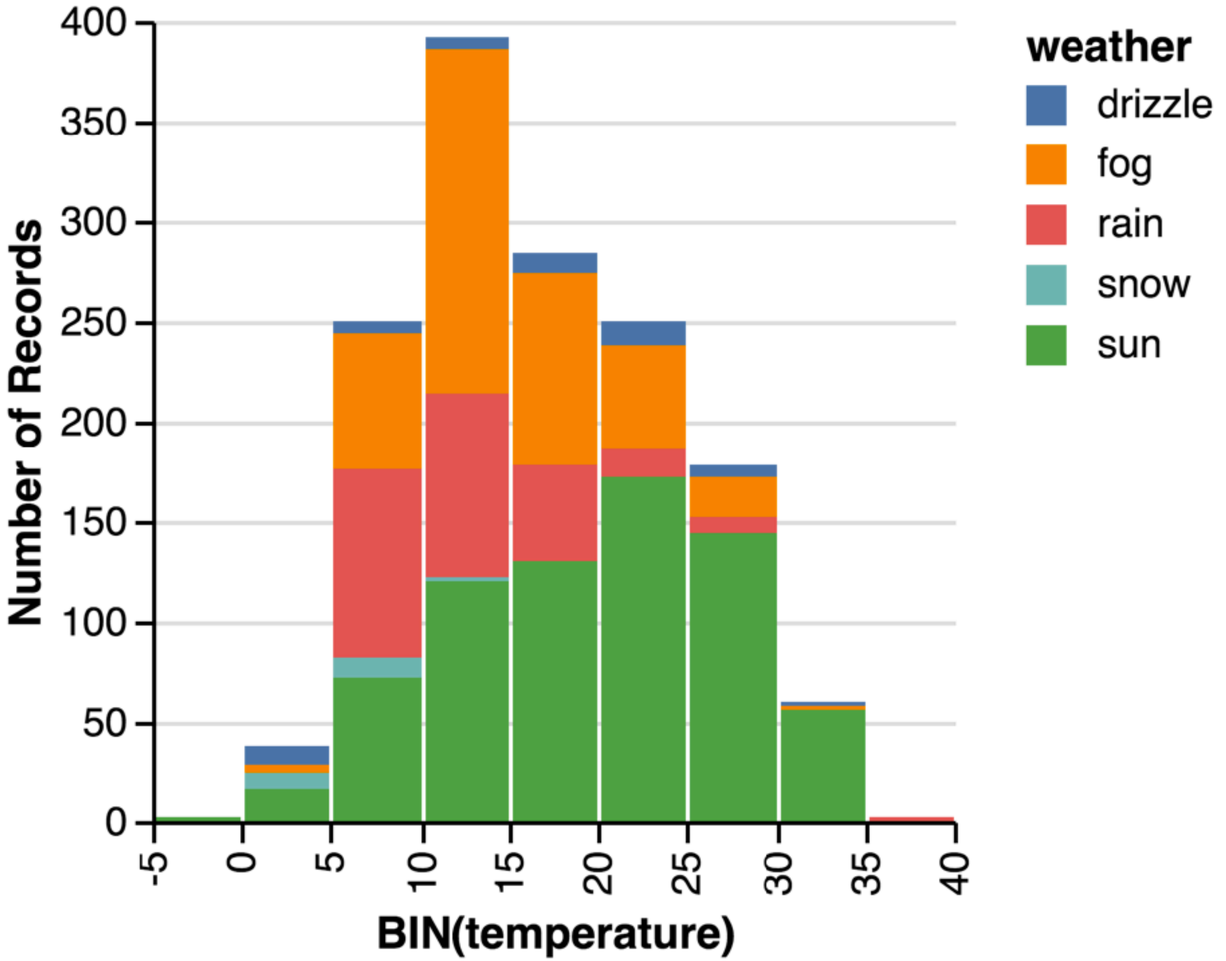
Histogram

```
{
  data: {url: "weather-seattle.json"},
  mark: "bar",
  encoding: {
    x: {
      bin: true,
      field: "temperature",
      type: "quantitative"
    },
   y: {
      aggregate: "count",
      type: "quantitative"
    }
  }
}
```



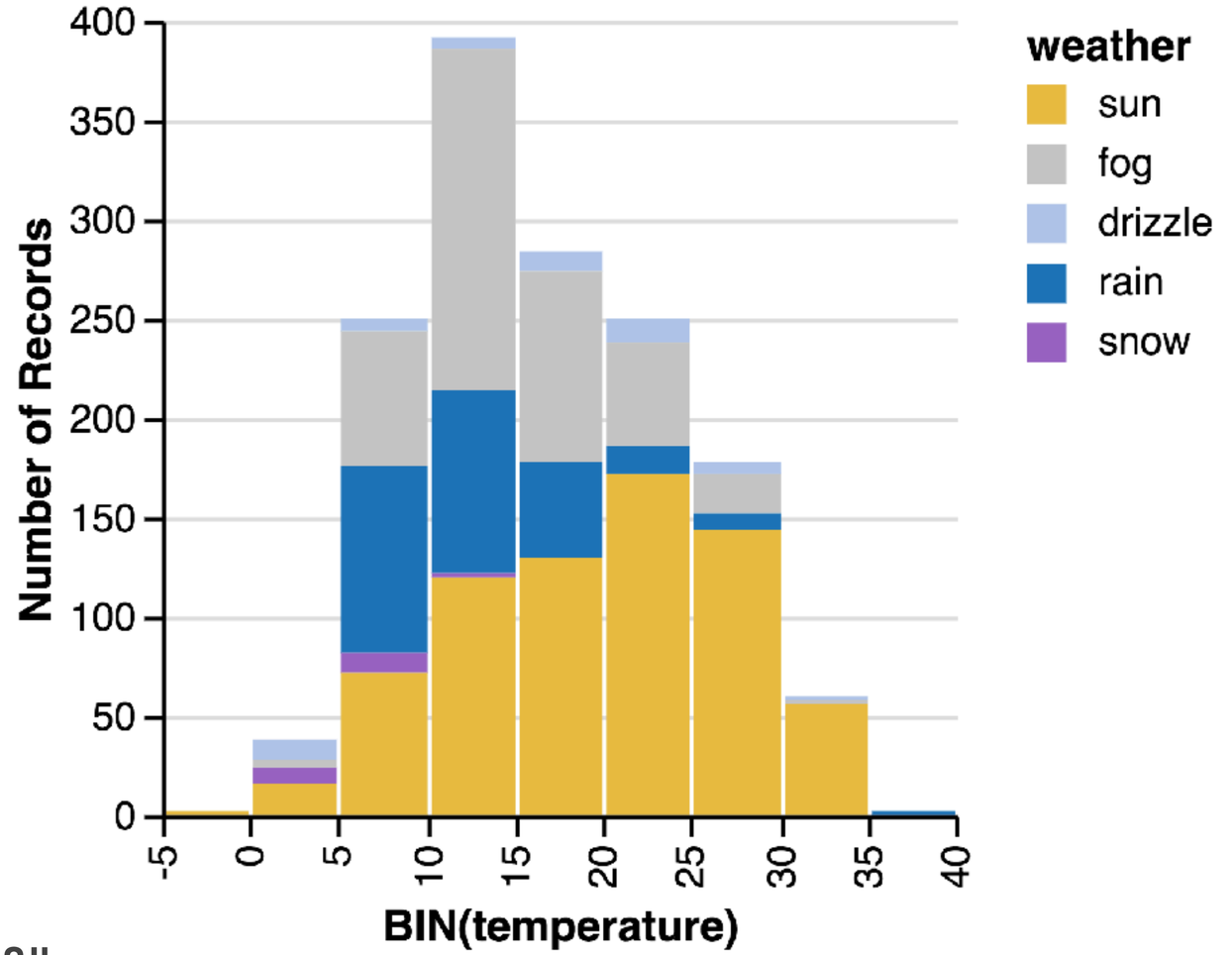
Histogram + Color

```
{
  data: {url: "weather-seattle.json"},
  mark: "bar",
  encoding: {
    x: {
      bin: true,
      field: "temperature",
      type: "quantitative"
    },
    y: {
      aggregate: "count",
      type: "quantitative"
    },
    color: {
      field: "weather",
      type: "nominal"
    }
  }
}
```



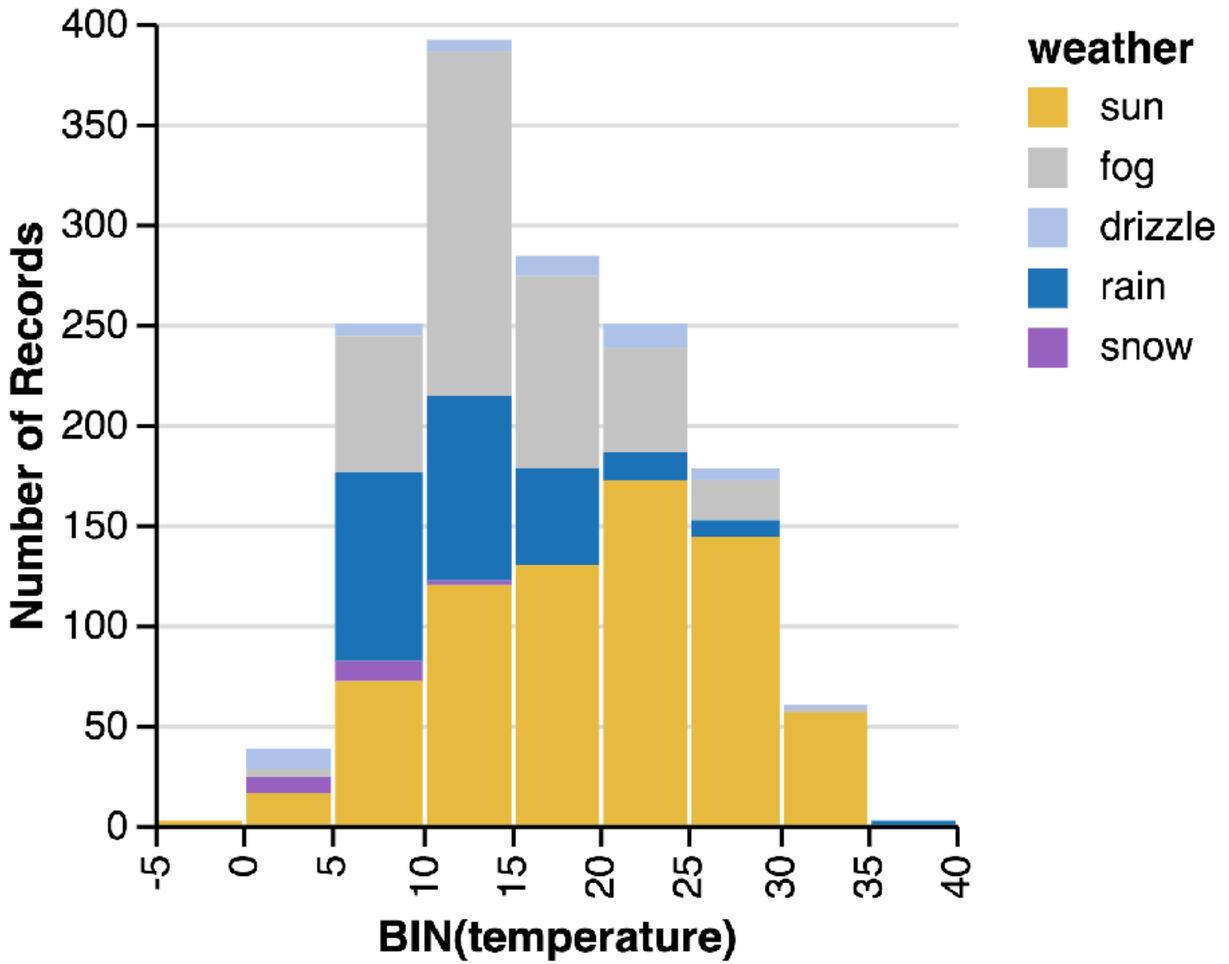
Histogram + Color

```
{
  data: {url: "weather-seattle.json"},
  mark: "bar",
  encoding: {
    x: {
      bin: true,
      field: "temperature",
      type: "quantitative"
    },
    y: {
      aggregate: "count",
      type: "quantitative"
    },
    color: {
      field: "weather",
      type: "nominal",
      "scale": {
        "domain": ["sun","fog","drizzle",
                   "rain","snow"],
        "range": ["#e7ba52","#c7c7c7","#aec7e8",
                  "#1f77b4","#9467bd"]
      }
    }
```



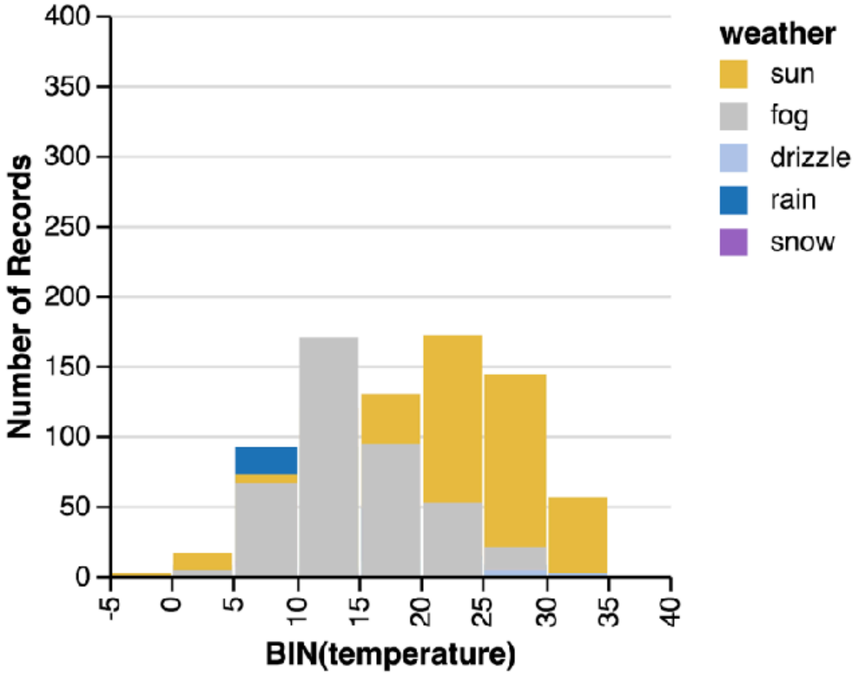
Histogram + Color = Stacked Histogram

```
{
  data: {url: "weather-seattle.json"},
  mark: "bar",
  encoding: {
    x: {
      bin: true,
      field: "temperature",
      type: "quantitative"
    },
    y: {
      aggregate: "count",
      type: "quantitative"
    },
    color: {
      field: "weather",
      type: "nominal",
      ...
      }
    }
  }
}
```

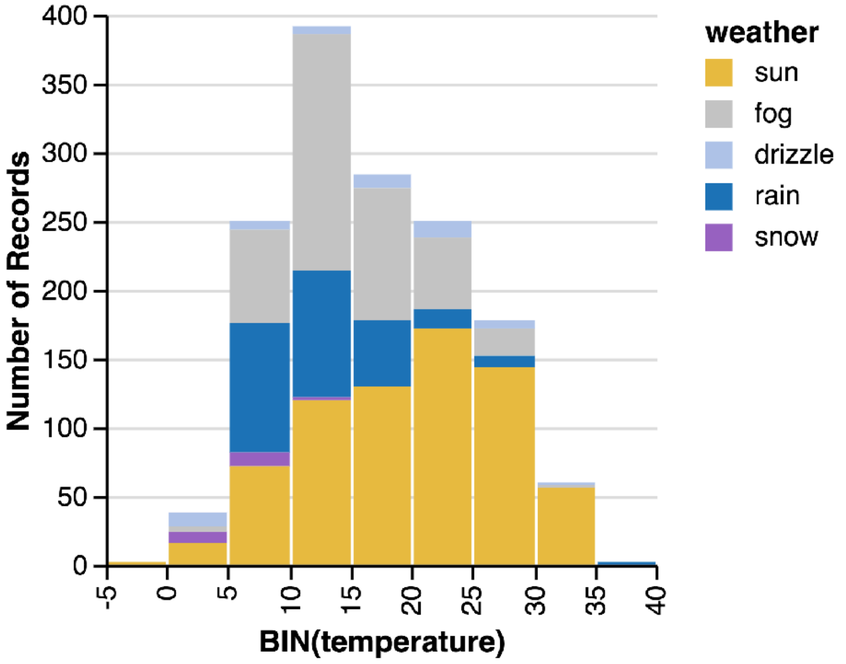


Stacked Histogram: Sensible Defaults

no stack



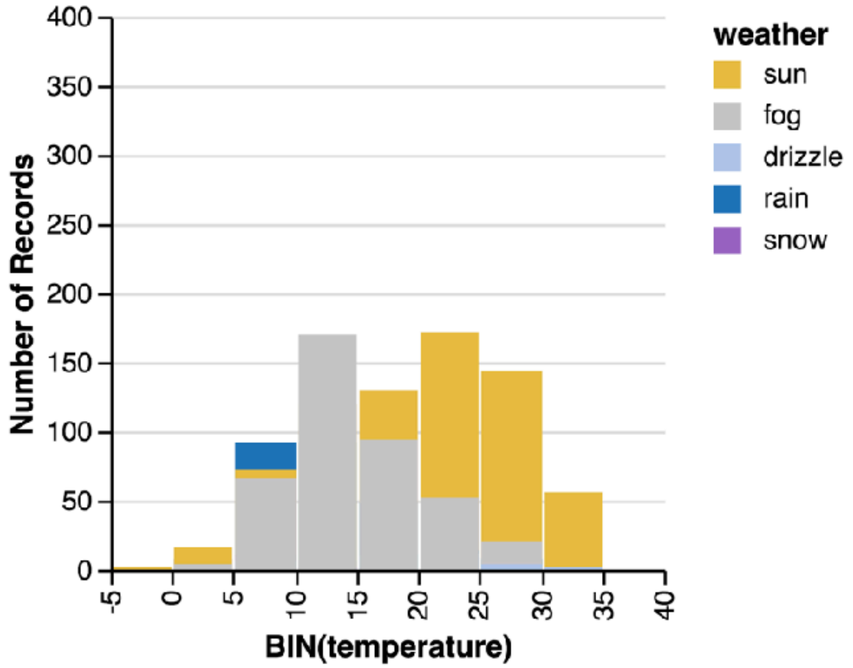
stack (default)



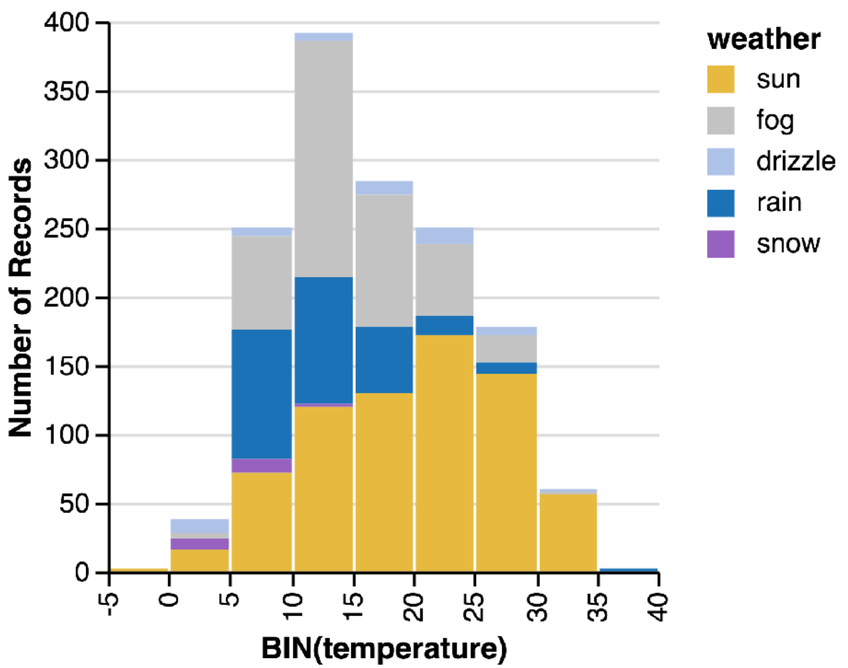
Stacked Histogram: Sensible Defaults

Channel (color) + Mark (bar) automatically enables stacking: a layout transform.

no stack



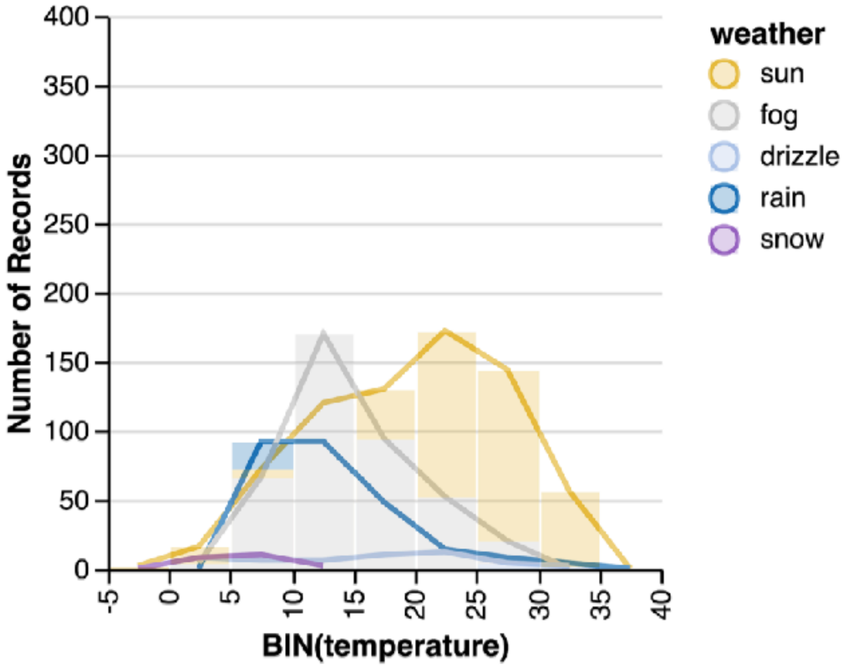
stack (default)



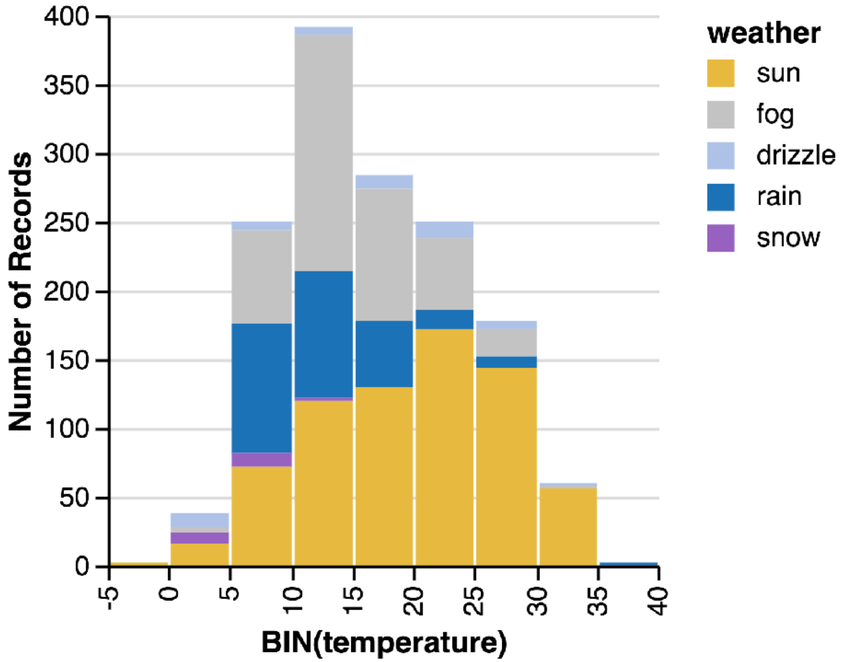
Stacked Histogram: Sensible Defaults

Channel (color) + Mark (bar) automatically enables stacking: a layout transform.

no stack → overlap

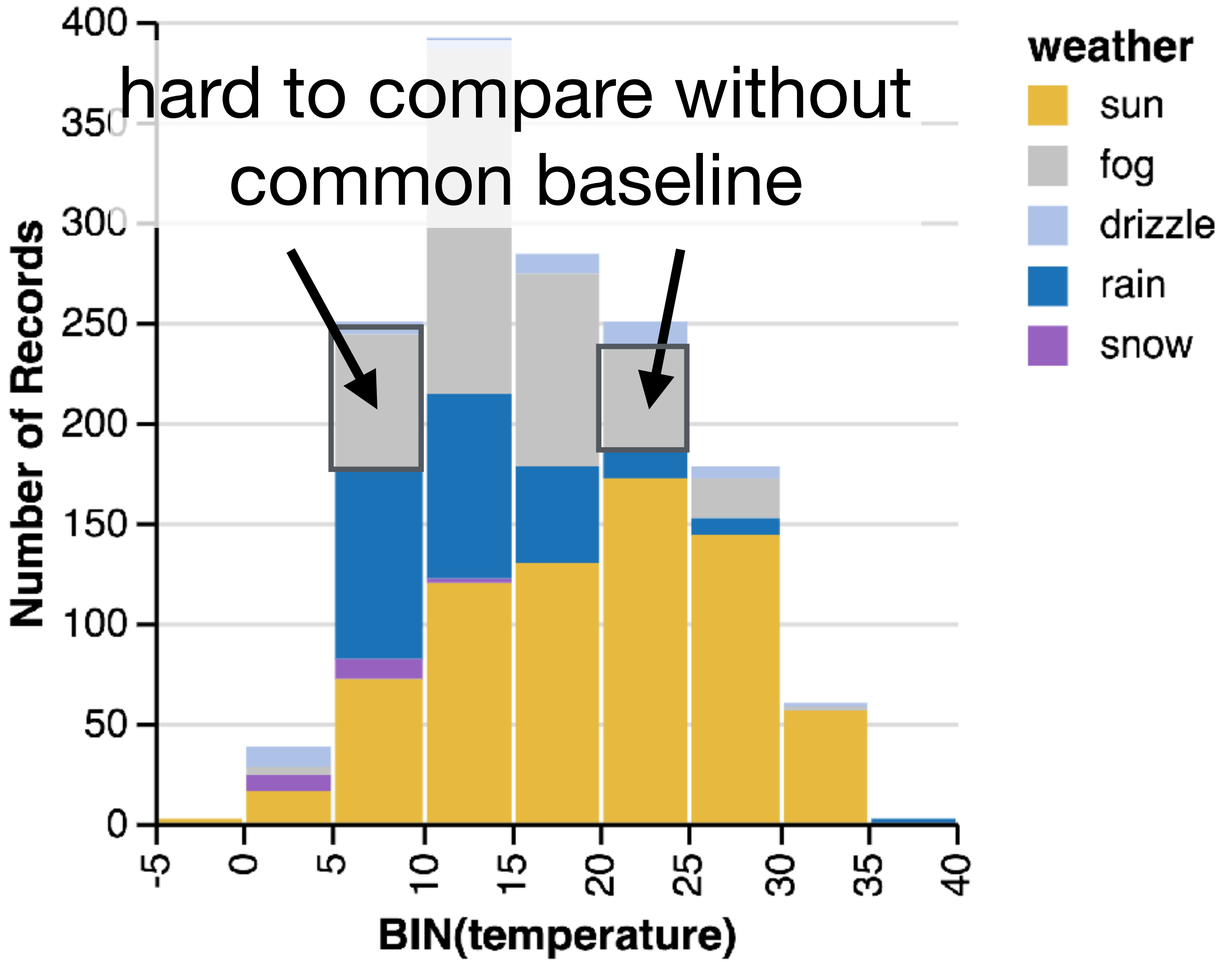


stack (default)



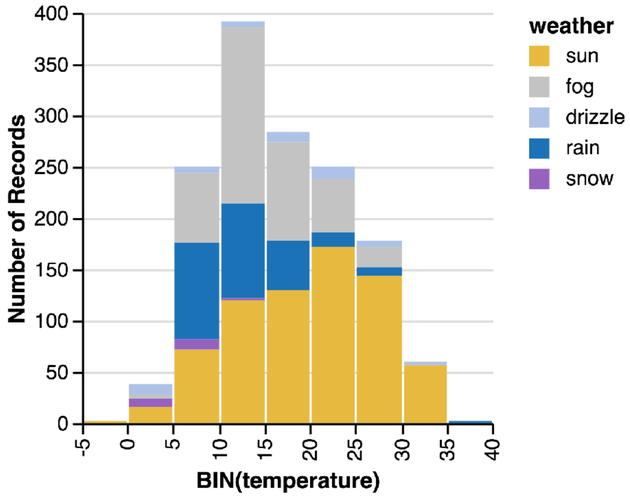
Histogram + Color = Stacked Histogram

```
{
  data: {url: "weather-seattle.json"},
  mark: "bar",
  encoding: {
    x: {
      bin: true,
      field: "temperature",
      type: "quantitative"
    },
    y: {
      aggregate: "count",
      type: "quantitative"
    },
    color: {
      field: "weather",
      type: "nominal"
    }
  }
}
```



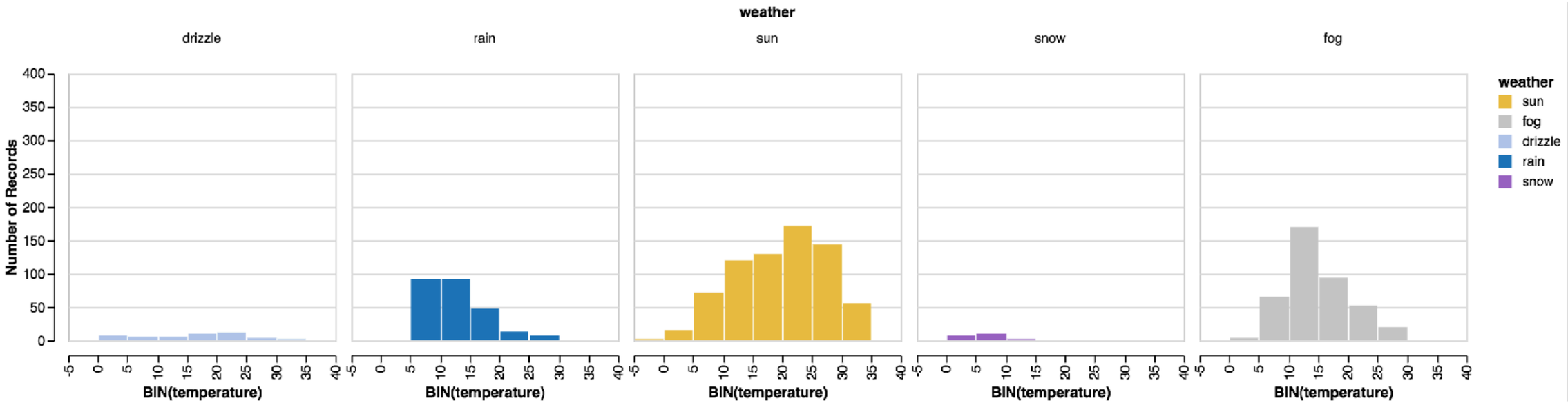
Histogram + Color = Stacked Histogram

```
{
  data: {url: "weather-seattle.json"},
  mark: "bar",
  encoding: {
    x: {bin: true, field: "temperature", type: "quantitative"},
    y: {aggregate: "count",              type: "quantitative"},
    color: {       field: "weather",     type: "nominal"}
  }
}
```

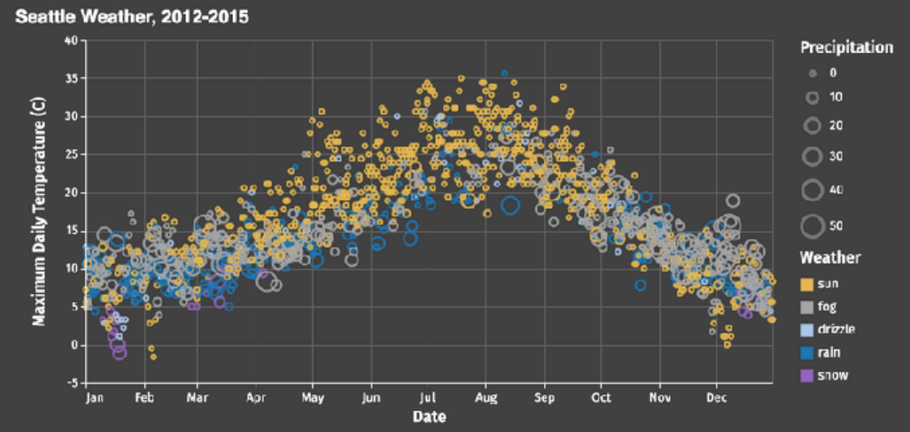


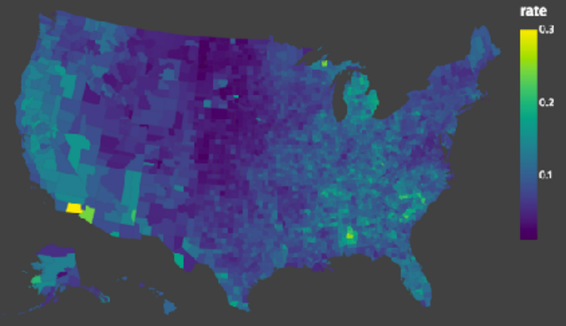
Histogram + Column = Trellis Histogram

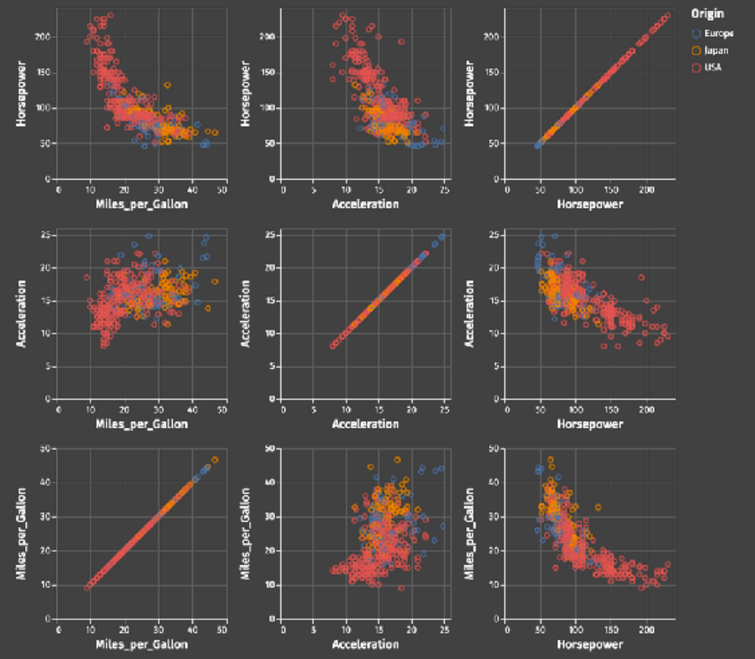
```
{
  data: {url: "weather-seattle.json"},
  mark: "bar",
  encoding: {
    x: {bin: true, field: "temperature", type: "quantitative"},
    y: {aggregate: "count",              type: "quantitative"},
    column: {      field: "weather",     type: "nominal"}
  }
}
```

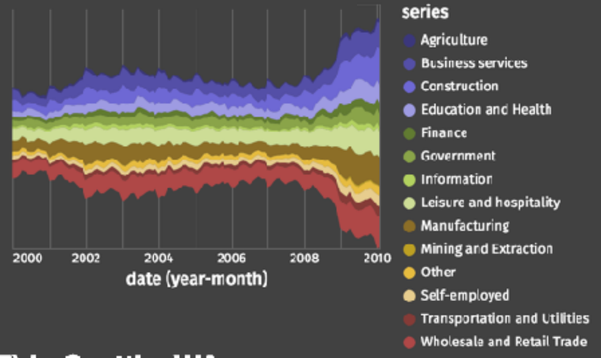


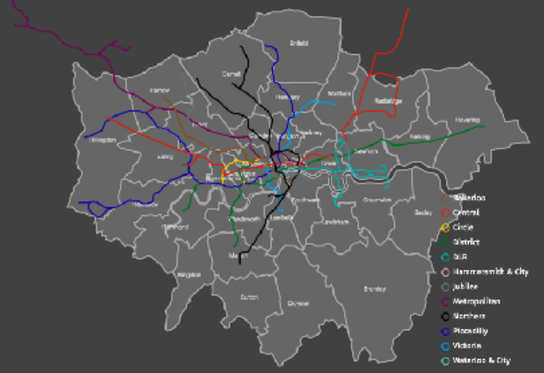
Vega-Lite is an Expressive Language.

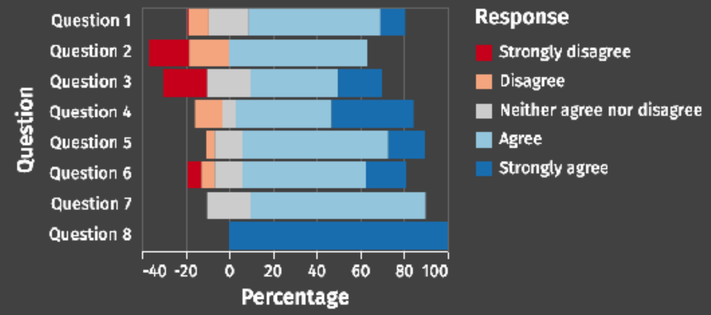


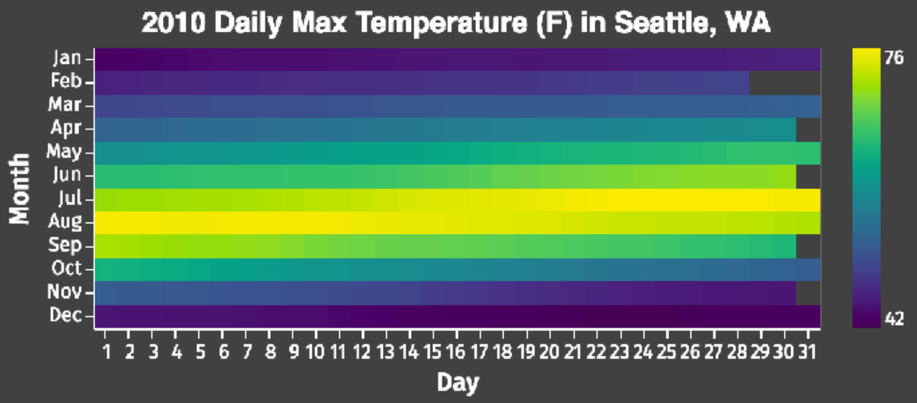


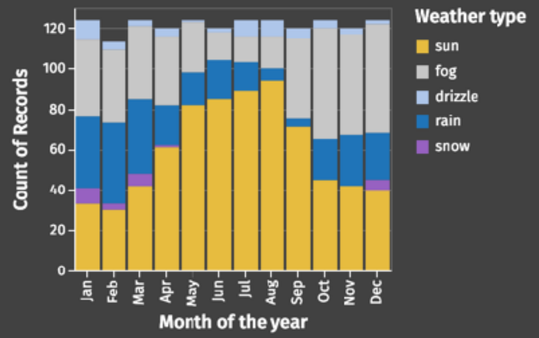


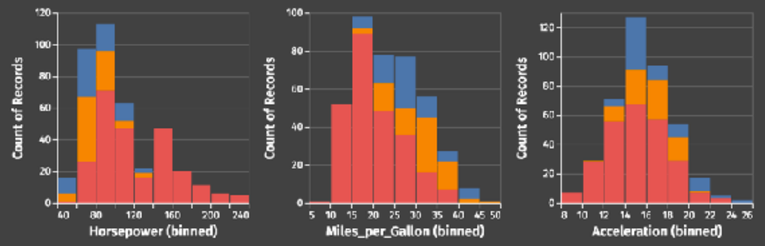




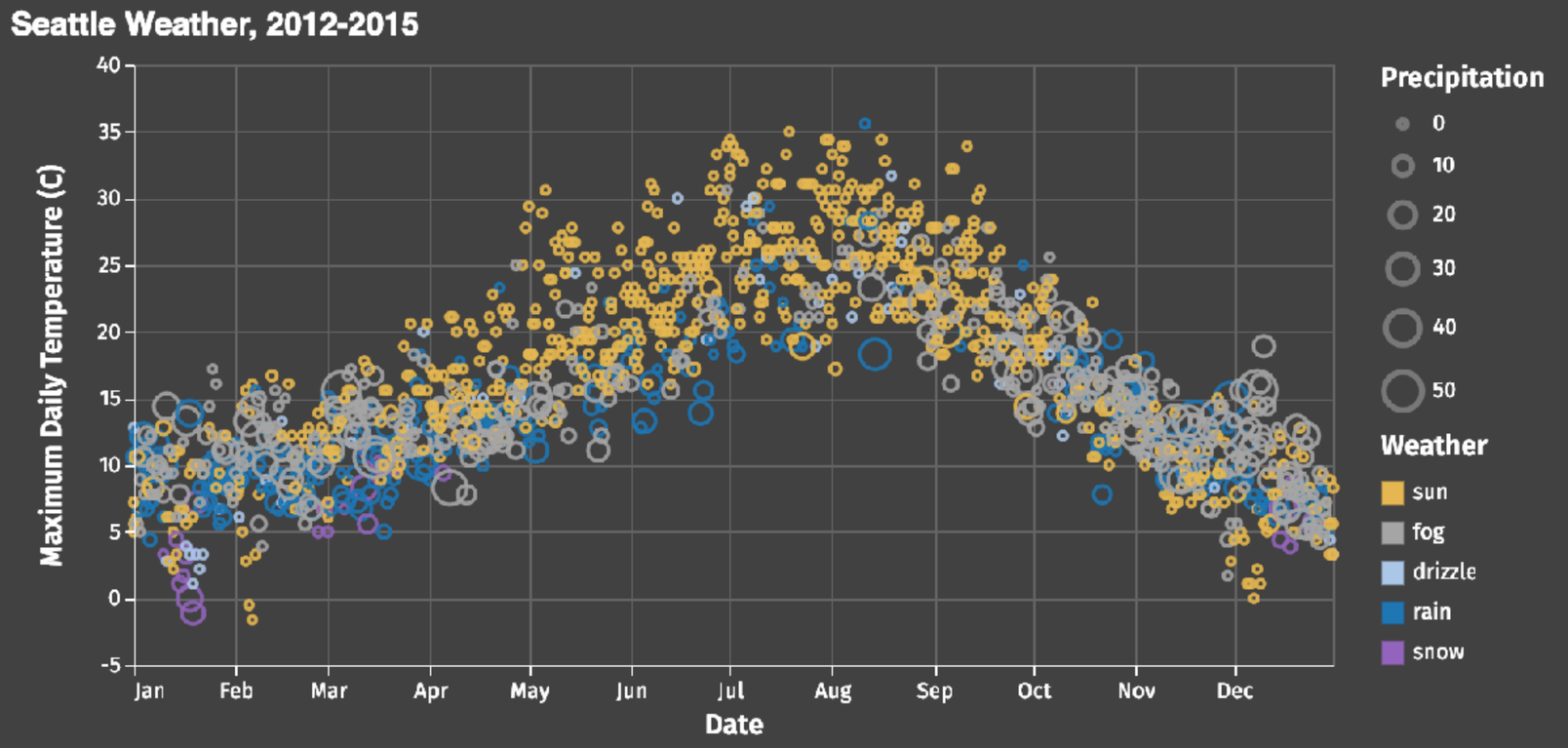






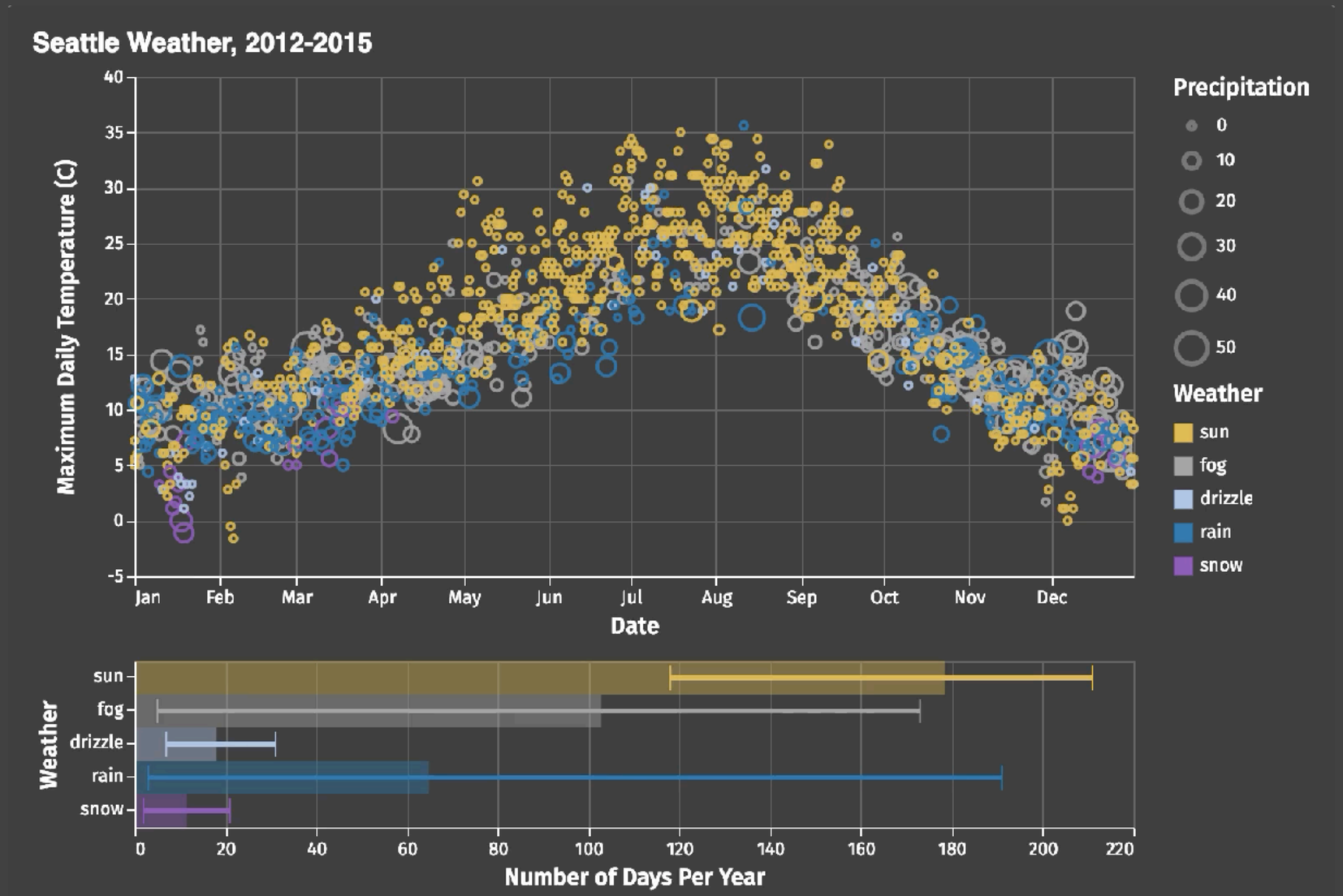


Vega-Lite is an Expressive Language for **Statistical** Graphics.



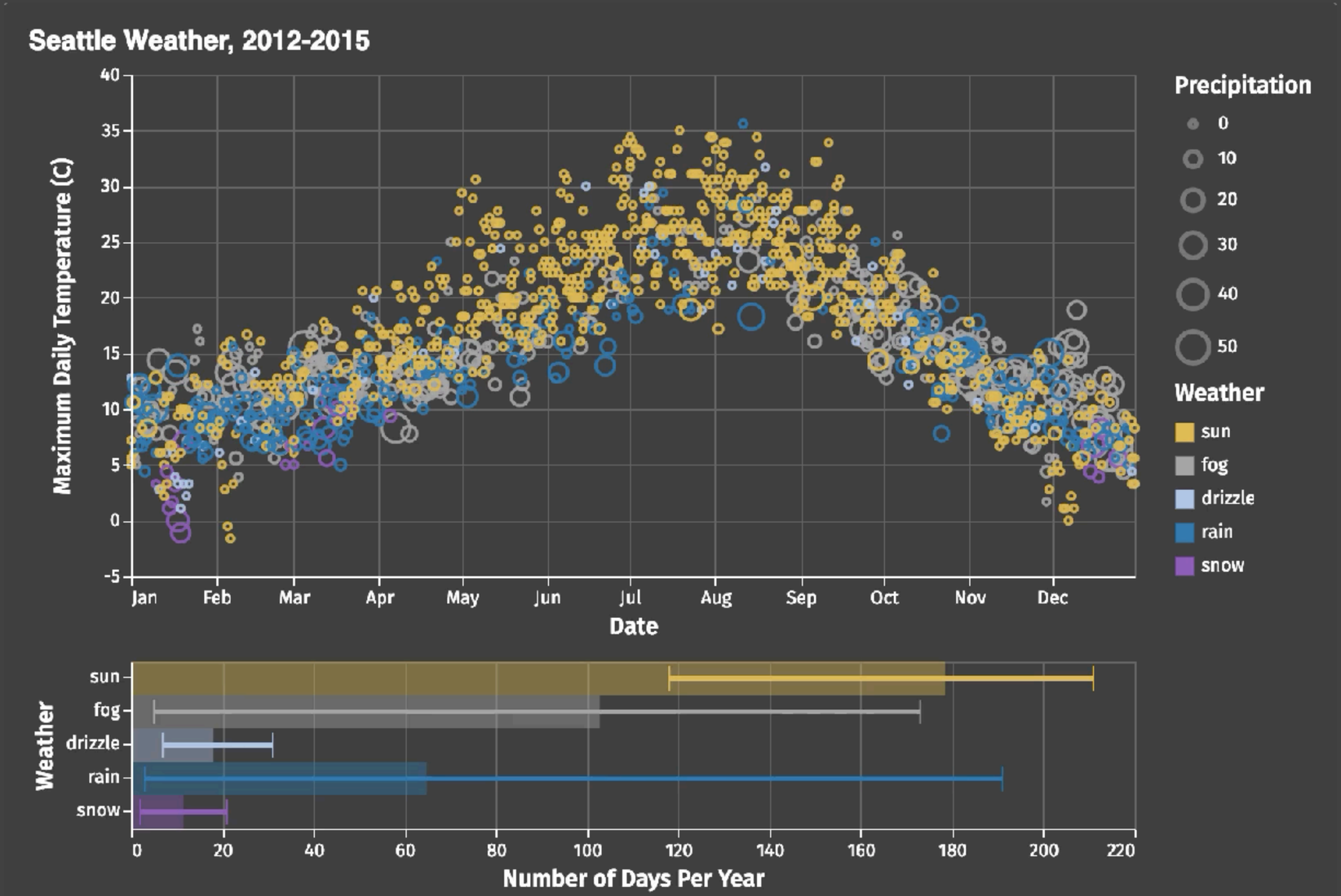
Vega-Lite is an Expressive Language for Statistical **Multi-View** Graphics.

Concatenation

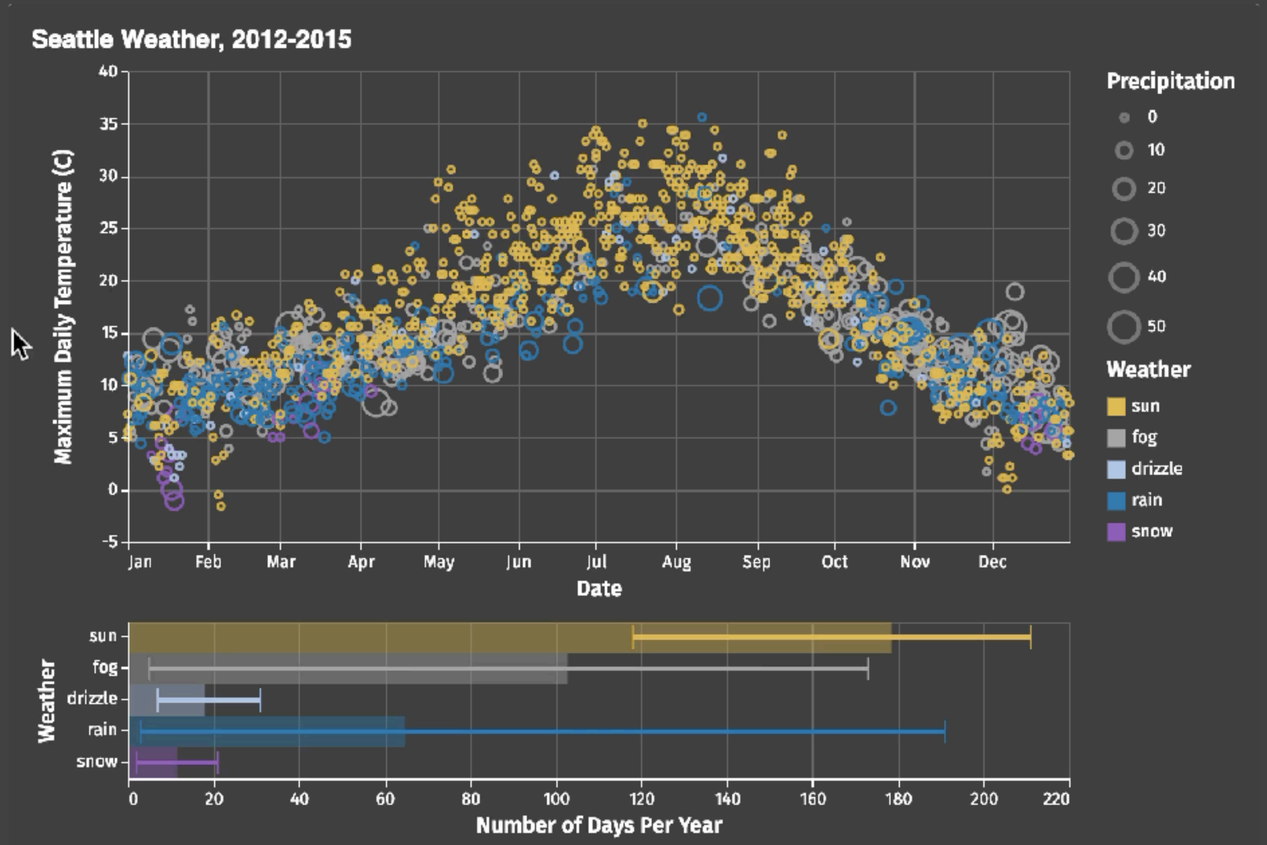


Layering & Error Marks

Vega-Lite is an Expressive Language for Statistical **Interactive** Multi-View Graphics.

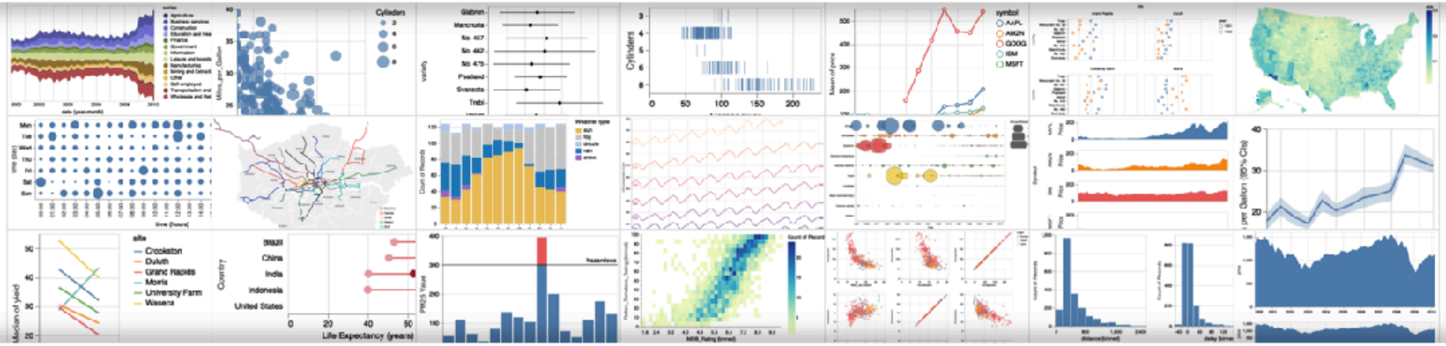


```
{
  "$schema": "https://vega.github.io/schema/vega-lite/v3.json",
  "title": "Seattle Weather, 2012-2015",
  "data": {
    "url": "data/seattle-weather.csv"
  },
  "vconcat": [
    {
      "width": 600,
      "height": 300,
      "mark": "point",
      "selection": {"brush": {"encodings": ["x"], "type": "interval"}},
      "transform": [{"filter": {"selection": "click"}}],
      "encoding": {
        "color": {
          "condition": {
            "field": "weather",
            "selection": "brush",
            "type": "nominal"
          },
          "value": "lightgray"
        },
        "size": {
          "field": "precipitation",
          "scale": {"domain": [-1, 50]},
          "type": "quantitative"
        },
        "x": {
          "axis": {"title": "Date", "format": "%b"},
          "field": "date",
          "timeUnit": "monthdate",
          "type": "temporal"
        },
        "y": {
          "field": "temp_max",
          "scale": {"domain": [-5, 40]},
          "type": "quantitative"
        }
      }
    },
    {
      "width": 600,
      "mark": "bar",
      "selection": {"click": {"encodings": ["color"], "type": "multi"}},
      "transform": [{"filter": {"selection": "brush"}}],
      "encoding": {
        "color": {
          "condition": {
            "field": "weather",
            "selection": "click",
            "type": "nominal"
          },
          "value": "lightgray"
        },
        "x": {"aggregate": "count", "type": "quantitative"},
        "y": {"field": "weather", "type": "nominal"}
      }
    }
  ]
}
```



Vega-Lite Examples Tutorials Documentation Usage Ecosystem GitHub Try Online

Vega-Lite – A Grammar of Interactive Graphics



Vega-Lite is a high-level grammar of interactive graphics. It provides a concise JSON syntax for rapidly generating visualizations to support analysis. Vega-Lite specifications can be compiled to Vega specifications.

Vega-Lite specifications describe visualizations as mappings from data to **properties of graphical marks** (e.g., points or bars). The Vega-Lite compiler **automatically produces visualization components** including axes, legends, and scales. It then determines properties of these components based on a set of **carefully designed rules**. This approach allows specifications to be succinct and expressive, but also provide user control. As Vega-Lite is designed for analysis, it supports **data transformations** such as aggregation, binning, filtering, sorting, and **visual transformations** including stacking and faceting. Moreover, Vega-Lite specifications can be **composed** into layered and multi-view displays, and made **interactive with selections**.

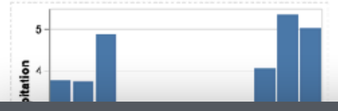
Get started
Latest Version: 4.9.0

Try online

Read our introduction article to Vega-Lite v2 on Medium, watch our OpenVis Conf talk about the new features in Vega-Lite v2, check out the documentation and take a look at our example gallery. Follow us on Twitter at @vega_vis to stay informed about updates.

Example

With Vega-Lite, we can start with a **bar chart of the average monthly precipitation** in Seattle, overlay a rule for the overall yearly average, and have it represent an interactive moving average for a dragged region. Next step



```
{
  "data": {"url": "data/seattle-weather.csv"},
  "mark": "bar",
  "encoding": {
    "x": {
```

Open in Vega Editor

vega.github.io/vega-lite

Recap: Vega-Lite Design Principles
As a Tool

Favor composition over templates

Provide sensible defaults, but allow customization

Automatic optimization of visual presentation and processing

Support programmatic generation, sharing, and reuse

https://medium.com/hci-design-at-uw/introducing-vega-lite-438f9215f09e

https://medium.com/@uwdata/introducing-vega-lite-2-0-de6661c12d58

Recap: Vega-Lite Design Principles

As a Language

Approachable. You don’t need to understand everything before you can understand anything.

Consistent. Building blocks can be reused.

Explains itself. Affordances like signposts.

Teaches. Vega-Lite implements best practices.

For humans. Reduce surprises and anticipate the need for learning and debugging.

https://medium.com/@mbostock/what-makes-software-good-943557f8a488

Is Vega-Lite really easier?

Let’s make a histogram

Vega-Lite

```
{
  data: {url: "weather-seattle.json"},
  mark: "bar",
  encoding: {
    x: {
      bin: true,
      field: "temperature",
      type: "quantitative"
    },
    y: {
      aggregate: "count",
      type: "quantitative"
    }
  }
}
```

Chart Typology (e.g. Seaborn)

```
sns.distplot(x);
```

The code looks simple but brushes complexity under the rug.

It does not teach you that a histogram is a bar chart.

Cannot be adapted to use different aggregation (e.g. sum).

What's next?

Work we didn't have time to talk about today

Protovis, D3, Vega, Vega-Lite, etc. allow us to create a combinatorial number of visualizations. Can tools help us make **good** charts?
→ Draco [Moritz et. al. 2018]

Use in data science environments (e.g. Python) → Altair (altair-viz.github.io)

Make animated visualizations.

Scale visualizations to large data.

Visualization tools (Tableau, Voyager, etc…).

Debugging tools → [Hoffswell et al. EuroVis'16]

What you can do now:

Articulate the difference between chart typologies and component architectures.

Decompose a graphic into the building blocks of visualization.

Know why most visualization toolkits are declarative.

Explain tradeoff between expressiveness and ease-of-use.

Know the history of visualization tools.

Understand the basics of D3.

Give examples of manifestations of design principles in Vega-Lite.