

Marks Revisited: Beyond Bertin

Tamara Munzner

Department of Computer Science
University of British Columbia

*Northeastern Vancouver
12 Jul 2024*

<http://www.cs.ubc.ca/~tmm/talks.html#northeastern24>



 [@tamara@vis.social](https://medium.com/@tamara@vis.social)

 [@tamaramunzner](https://twitter.com/tamaramunzner)

Co-conspirators

- Richard Brath
 - Uncharted Software
- Mara Solen, Francis Nguyen, Ryan Smith
 - UBC CS infovis course TAs
- also useful discussions with
 - Enrico Bertini, Hanspeter Pfister, Arvind Satyanarayan, Maureen Stone, Martin Wattenberg

Marks and channels: Foundational model

- decompose visual encoding into marks & channels

– marks

- geometric primitives
- represent data items

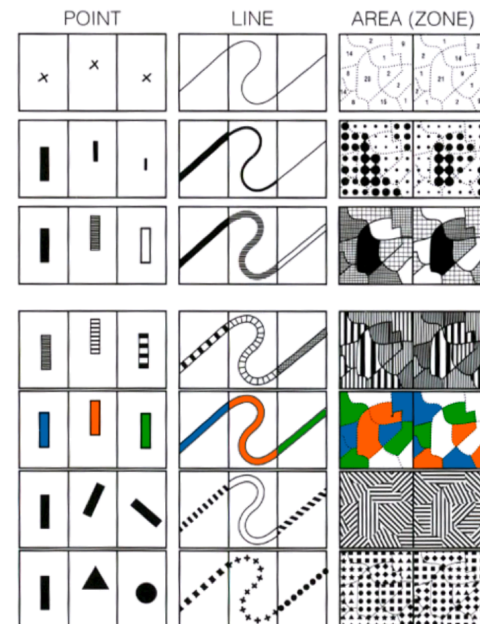
– channels

- control appearance of marks
- representing data attributes

- widely used

– Bertin 1967

- Semiology of Graphics



Marks

→ Points



→ Lines



→ Areas



Channels

→ Position

→ Horizontal



→ Vertical



→ Both



→ Color



→ Shape



→ Tilt



→ Size

→ Length



→ Area

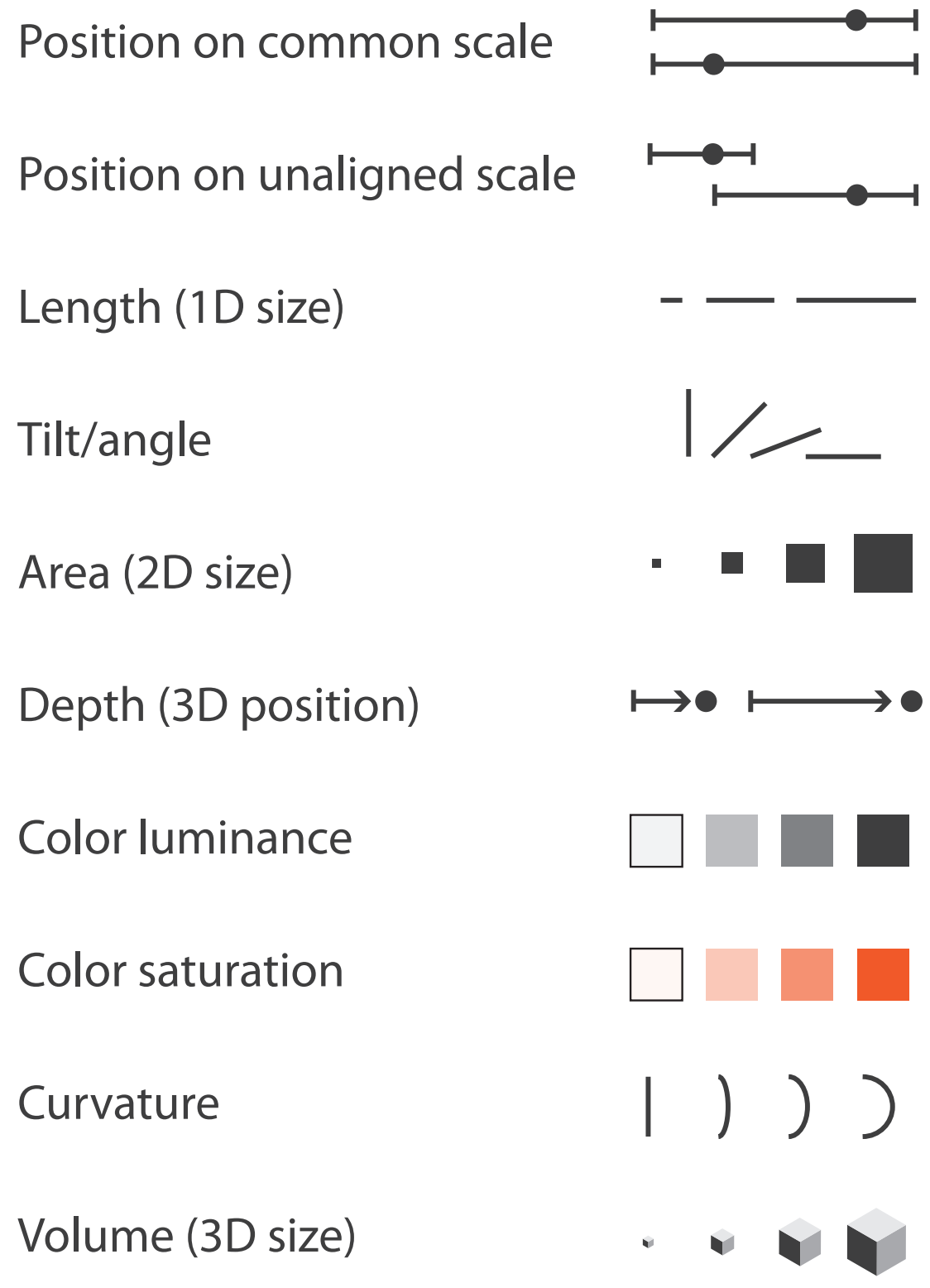


→ Volume



Channels: Rankings

➔ Magnitude Channels: Ordered Attributes



➔ Identity Channels: Categorical Attributes



Best
Effectiveness
Least

- **expressiveness**
 - match channel and data characteristics
- **effectiveness**
 - channels differ in accuracy of perception
 - two-value ratio judgements, Cleveland & McGill 1987

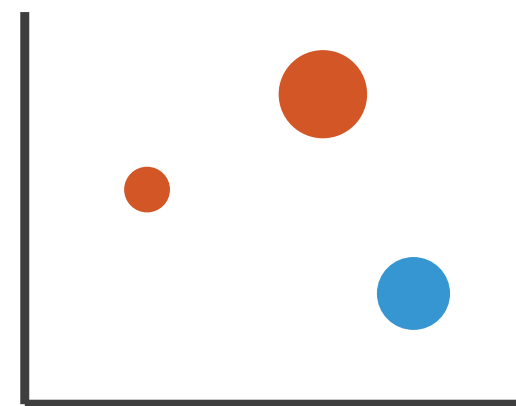
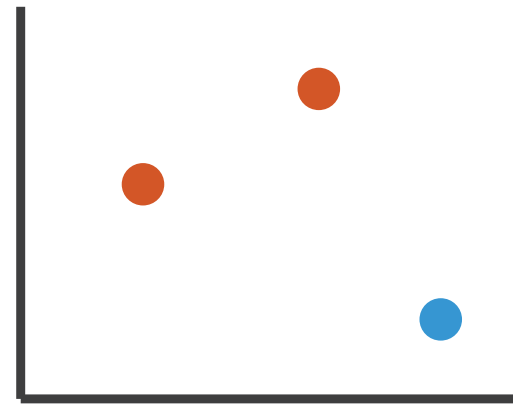
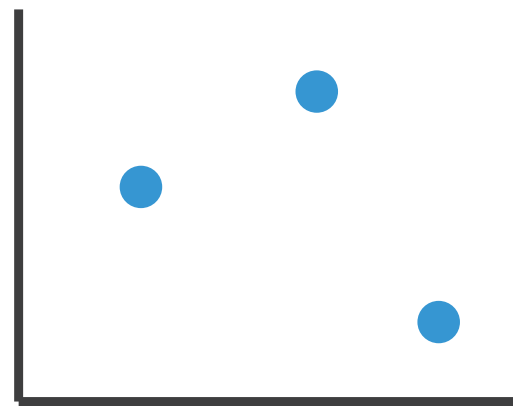
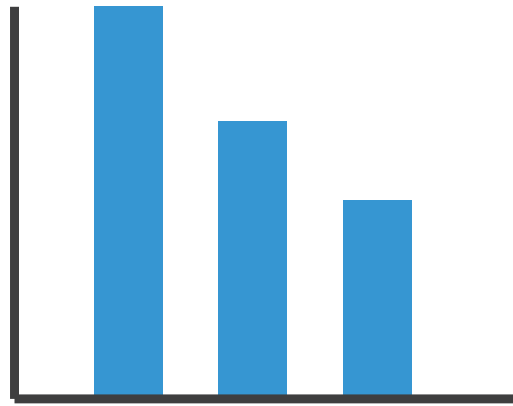
Talk outline

- explain current marks & channels model
- walk through many questions that arise when teaching it
- present preliminary ideas towards an alternative model

Current Marks & Channels Model

Visual encoding model

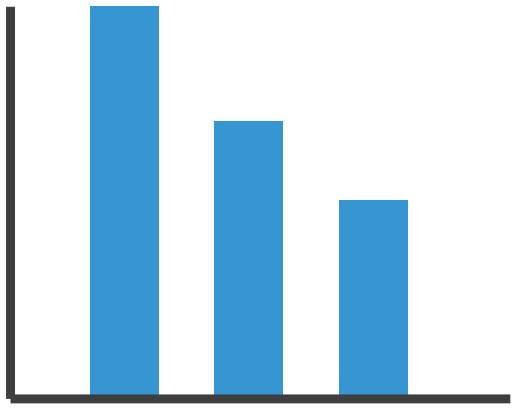
- analyze idiom structure as combination of marks and channels



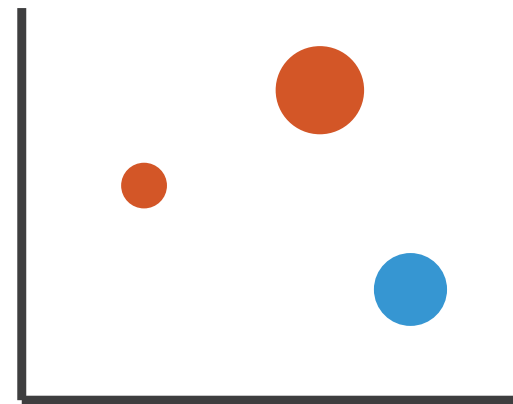
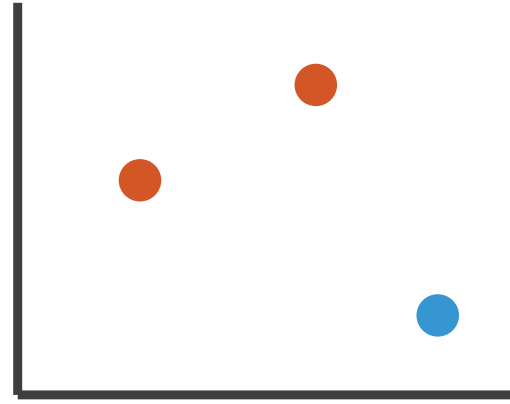
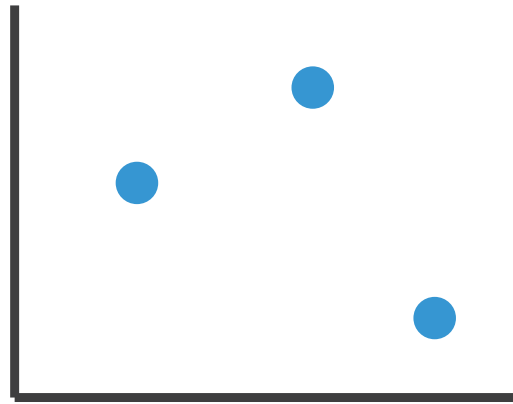
Visual encoding model

- analyze idiom structure as combination of marks and channels

idiom: bar chart



1 channel:
vertical position

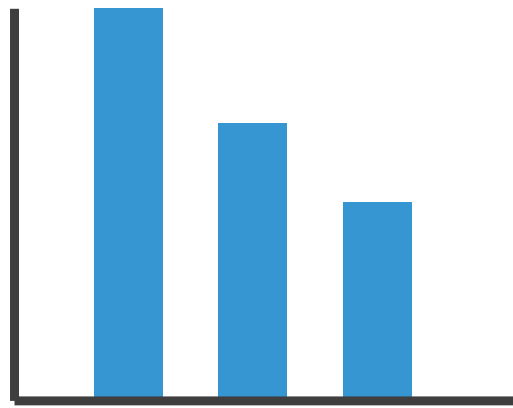


mark: line

Visual encoding model

- analyze idiom structure as combination of marks and channels

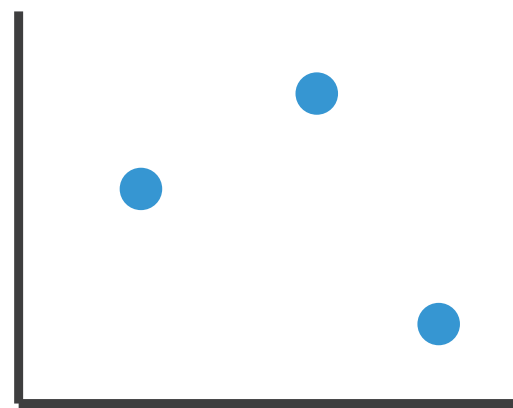
idiom: bar chart



1 channel:
vertical position

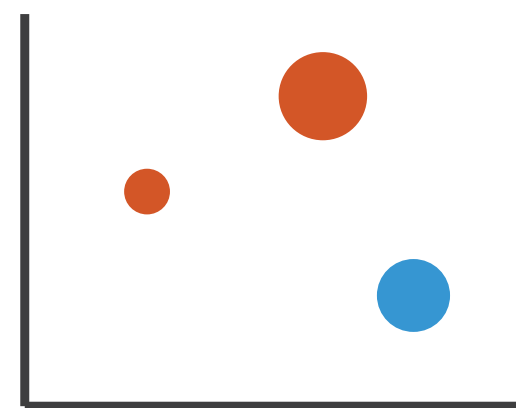
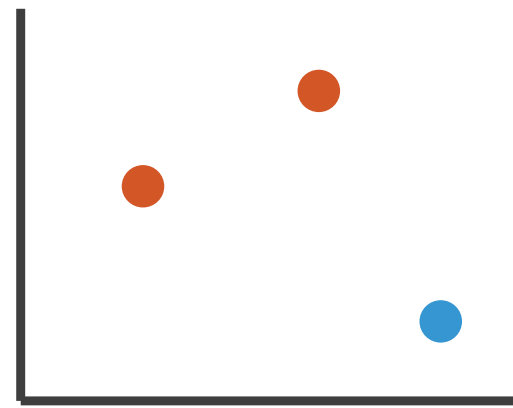
mark: line

idiom: scatterplot



2 channels:
vertical position
horizontal position

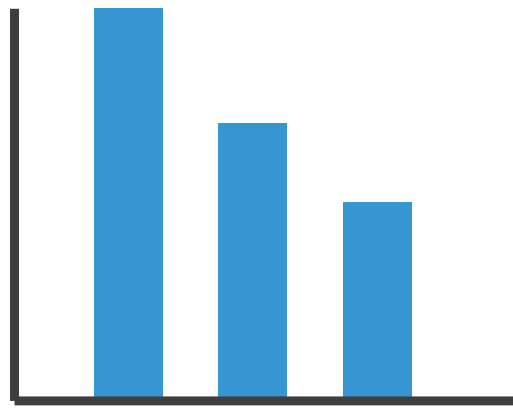
mark: point



Visual encoding model

- analyze idiom structure as combination of marks and channels

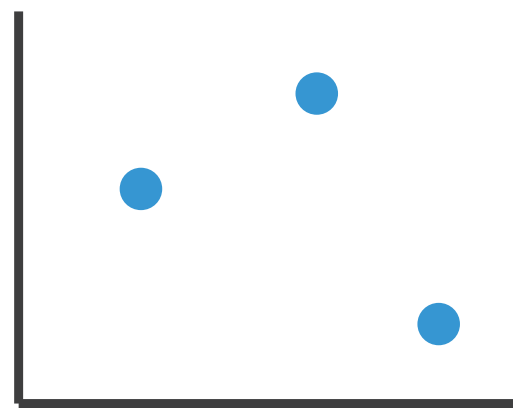
idiom: bar chart



1 channel:
vertical position

mark: line

idiom: scatterplot



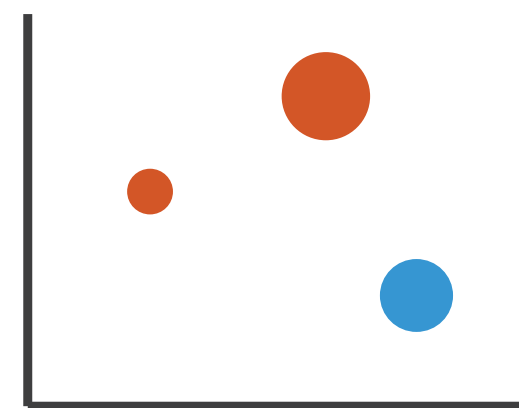
2 channels:
vertical position
horizontal position

mark: point



3 channels:
vertical position
horizontal position
color hue

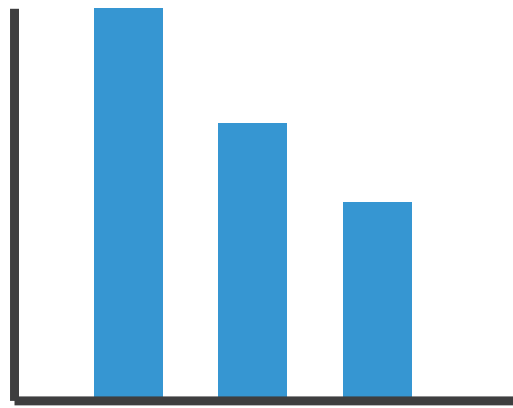
mark: point



Visual encoding model

- analyze idiom structure as combination of marks and channels

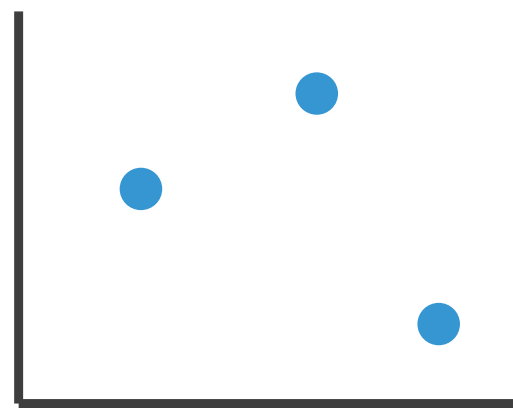
idiom: bar chart



1 channel:
vertical position

mark: line

idiom: scatterplot



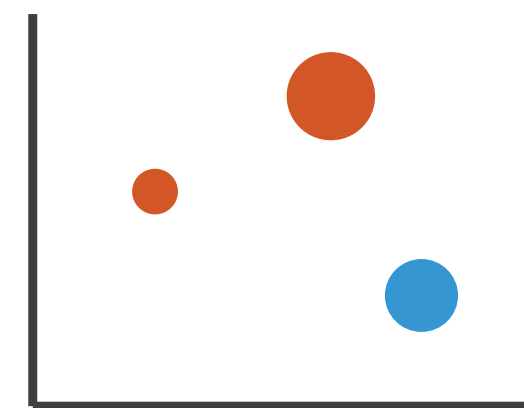
2 channels:
vertical position
horizontal position

mark: point



3 channels:
vertical position
horizontal position
color hue

mark: point



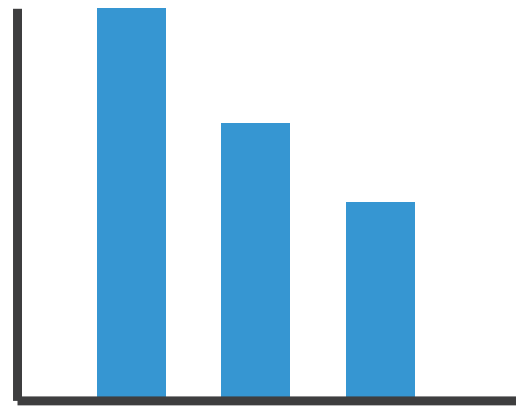
4 channels:
vertical position
horizontal position
color hue
size (area)

mark: point

Visual encoding model: Tabular data

- marks for items of tabular data

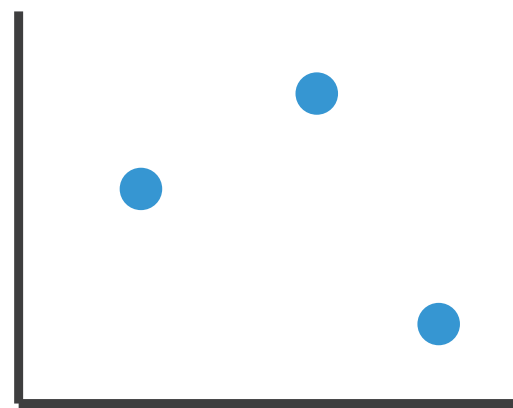
idiom: bar chart



1 channel:
vertical position

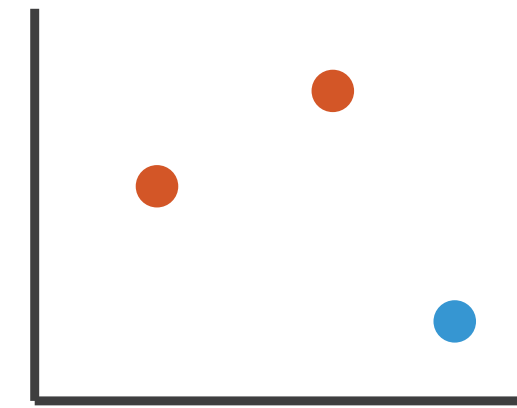
mark: line

idiom: scatterplot



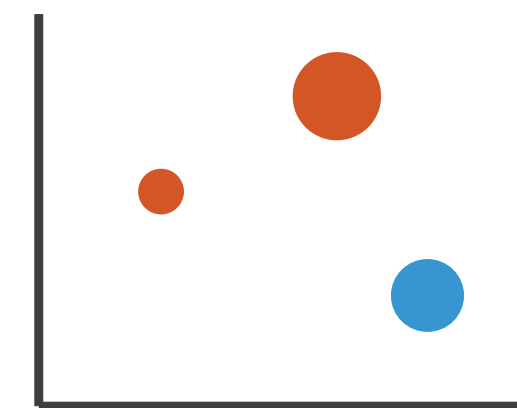
2 channels:
vertical position
horizontal position

mark: point



3 channels:
vertical position
horizontal position
color hue

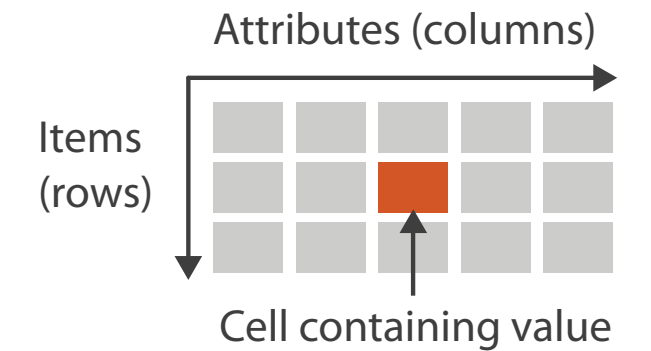
mark: point



4 channels:
vertical position
horizontal position
color hue
size (area)

mark: point

→ Tables



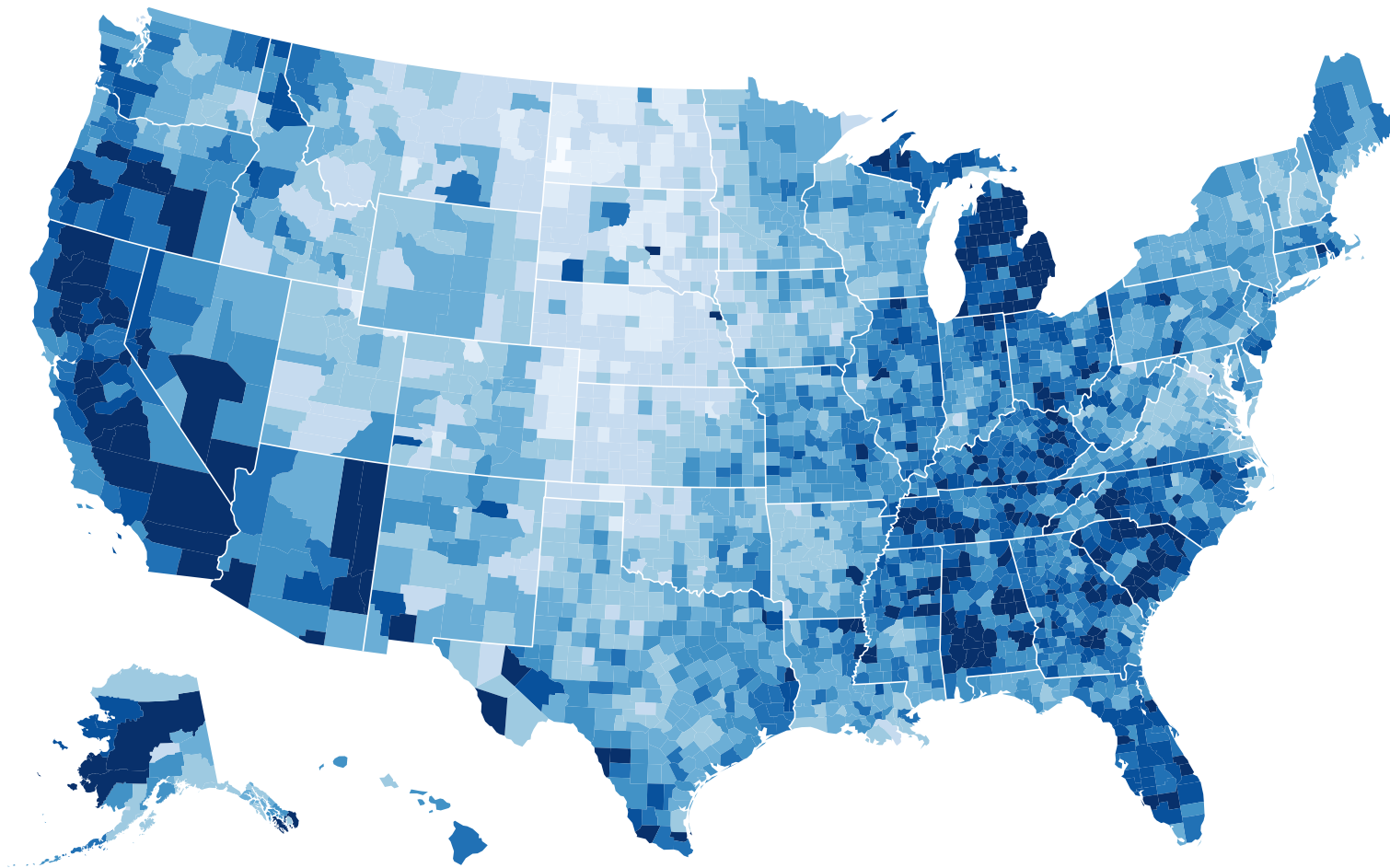
Visual encoding model: Spatial data

- marks for items of spatial data

idiom: choropleth map

channels:
position
color (saturation)

mark: area

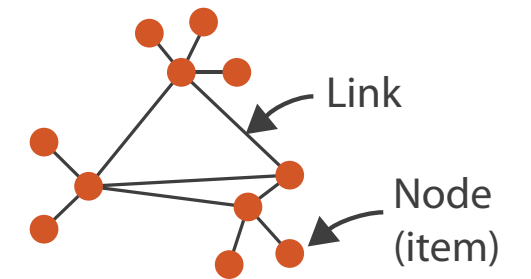


<http://blocks.org/mbostock/4060606>

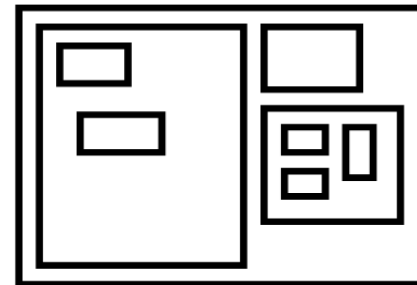
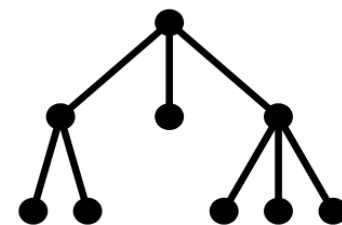
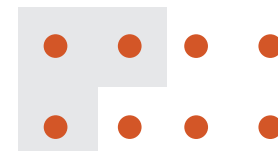
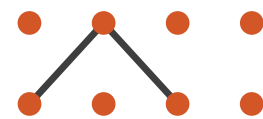
Visual encoding model: Network data

- marks for items and marks for links

→ Networks



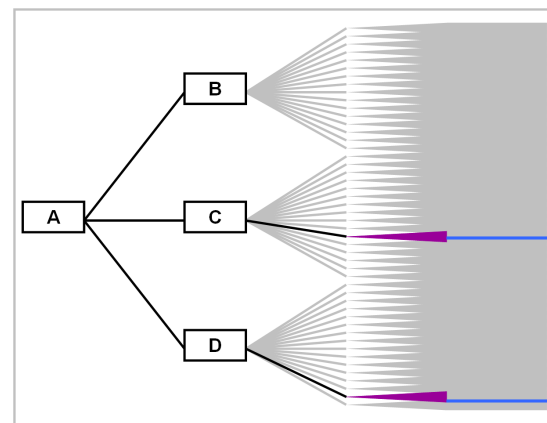
⊕ Connection ⊕ Containment



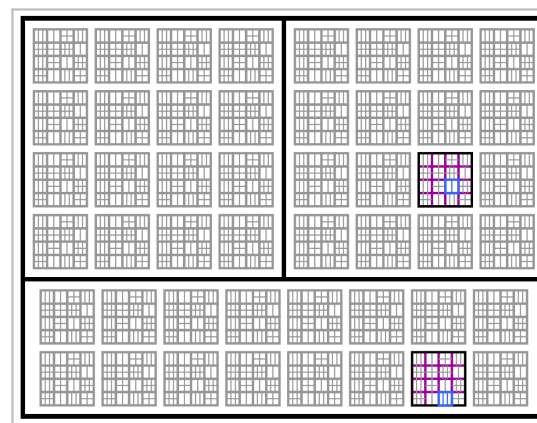
idiom: node-link diagram

channel:
position

marks:
point for items/nodes,
connection line for links

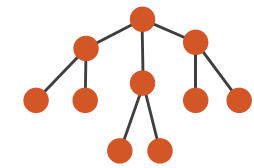


Node-Link Diagram



Treemap

→ Trees



idiom: treemap

channel:
position

marks:
area for items/nodes,
containment area for parent-child links

Why analyze visual encodings?

- marks & channels model is a **design space**
 - **descriptive** power: ability to describe significant range of existing examples
 - **evaluative** power: ability to help assess multiple design alternatives
 - **generative** power: ability to help designers create new designs
 - **criteria:** *Michel Beaudoin-Lafon, Designing Interaction, not Interfaces. AVI 2004.*
- many names: taxonomies, typologies, classifications, frameworks, models, grammars...
 - delineate: axes / dimensions / categories
 - that are cross-cutting / independent / orthogonal
- design spaces help us reason
 - impose systematic & actionable structure on set of possibilities for specific problem
 - to support reasoning about design choices
 - capture the key variables at play
 - increase cognitive efficiency & support inferences by grouping similar instances together to facilitate reasoning about classes

Design spaces in visualization: continuing theme

The Structure of the Information Visualization Design Space

Stuart K. Card and Jock Mackinlay
Xerox PARC

Exploring the Design Space of Composite Visualization

Waqas Javed* Niklas Elmqvist†

2366

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 19, NO. 12, DECEMBER 2013

A Design Space of Visualization Tasks

Hans-Jörg Schulz, Thomas Nocke, Magnus Heitzler, and Heidrun Schumann

A Design Space of Vision Science Methods for Visualization Research

Madison A. Elliott, Christine Nothelfer, Cindy Xiong, and Danielle Albers Szafir



Fig. 1. Overview of design space of experimental methods. We present a four component design space to guide researchers in creating visualization studies grounded in vision science research methods.

ABSTRACT
Research
place wh
been pro
discover
analyze
the diff
possibilit
informati
series of
designing
designs.

A ⊗_{ux} E
Fig

ABSTRACT
We propose
as a theoret
multiple views (v
visual repres
five such stra
tensive review
on to show h
of a design s
tions in term
between item
design space
Index Term
mation Syste
Interfaces; I.

Abstra
pursue
all des
these
paper.
distrib
impact
that all
Index

1 INTRODUCTION

As the field
ation sets in

Rethinking book design space: Visualization Analysis & Design 2e

How?

Encode

➔ Arrange

➔ Express



➔ Order



➔ Use



➔ Separate



➔ Align



➔ Map

from **categorical** and **ordered** attributes

➔ Color

➔ Hue



➔ Saturation



➔ Luminance



➔ Size, Angle, Curvature, ...



➔ Shape



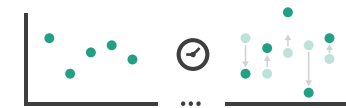
➔ Motion

Direction, Rate, Frequency, ...

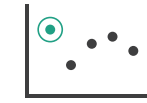


Manipulate

➔ Change



➔ Select



➔ Navigate

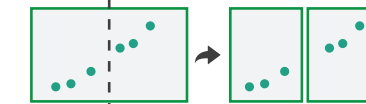


Facet

➔ Juxtapose



➔ Partition



➔ Superimpose



Reduce

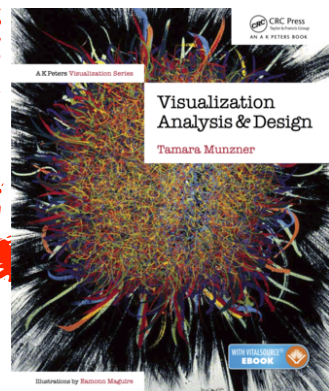
➔ Filter



➔ Aggregate



➔ Embed



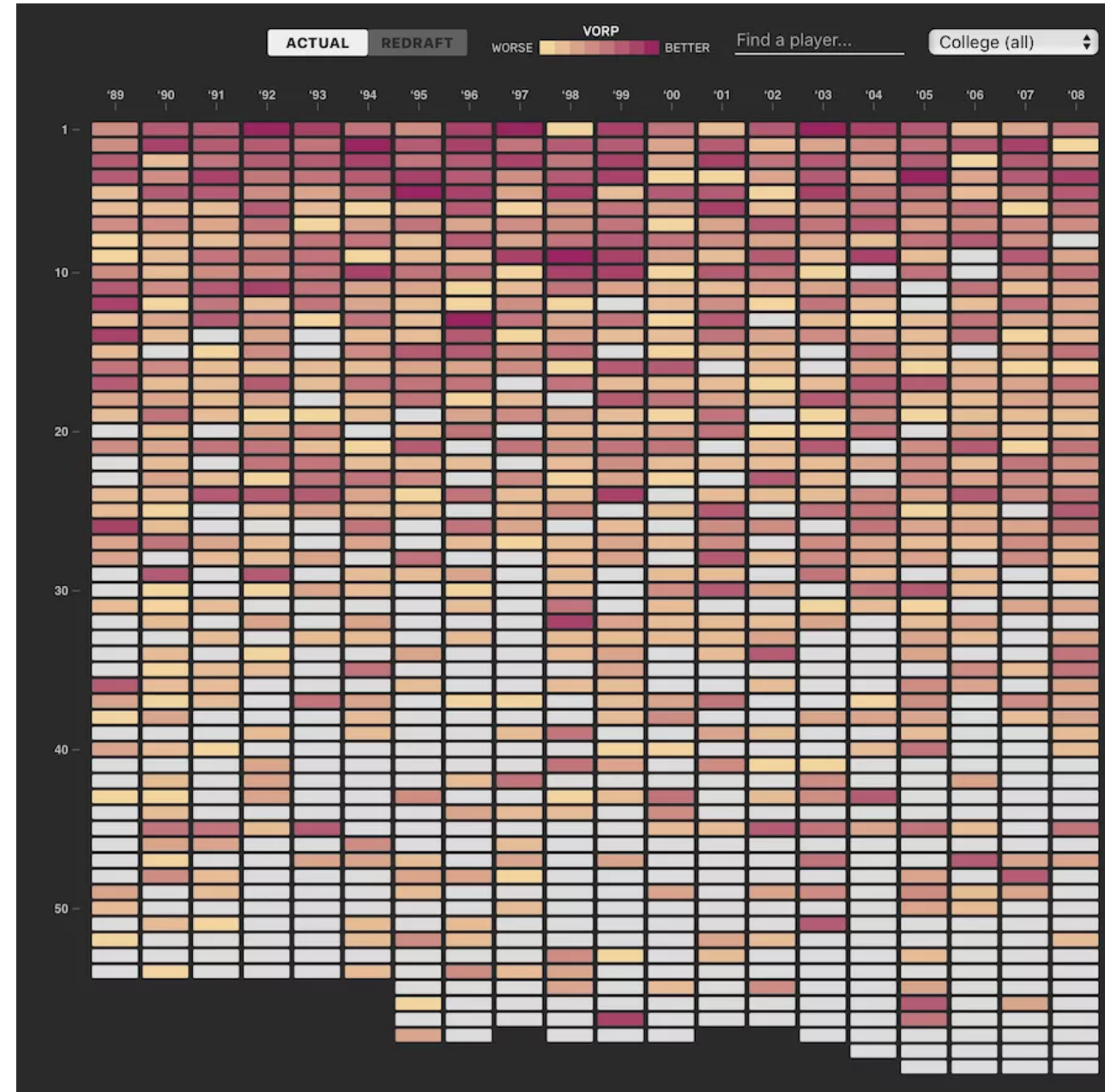
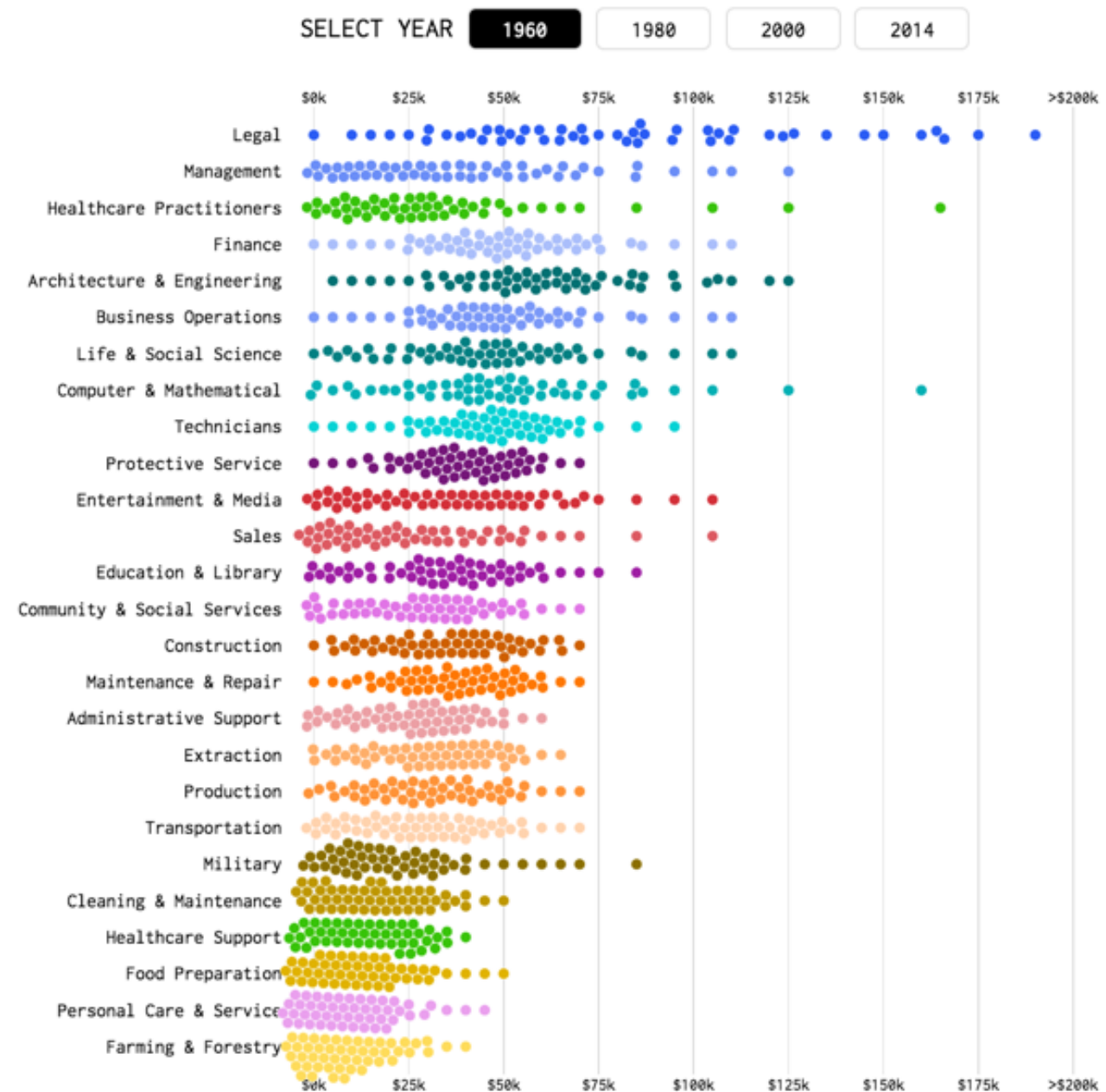
Teaching Challenges

Teaching design space: analyze visual encoding & map to data

- assignment: analyze existing encoding with marks & channels
 - Visual channels used?
 - Channel X encodes attribute Y
 - Channel X encodes attribute Y
 - Marks used?
 - Mark of type X encodes item Y
 - Mark of type X encodes item Y
- rationale
 - reverse-engineering existing designs will help students generate future designs

Teaching: Bertini in-class exercises, catalyst for questions

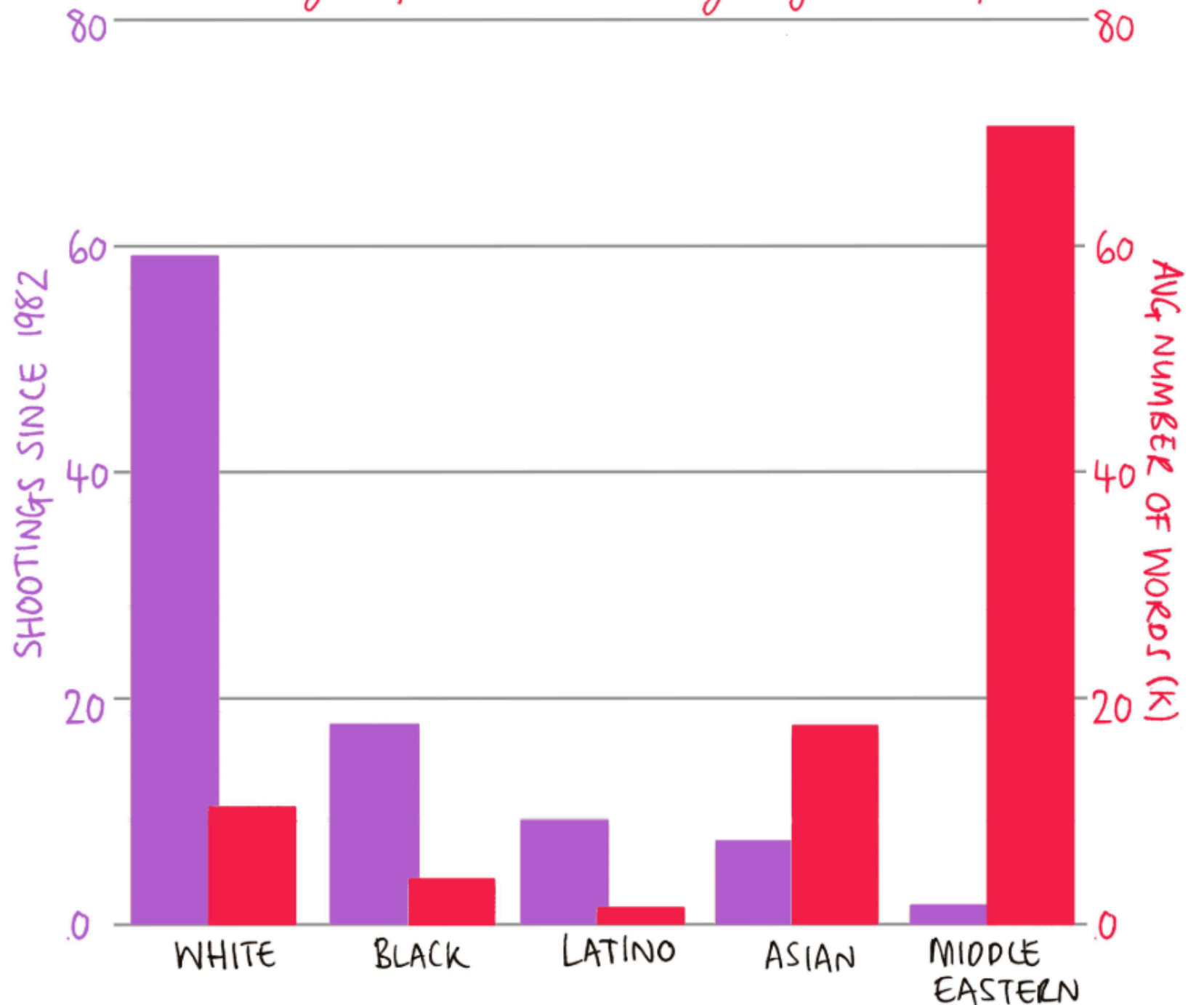
- decoding marks & channels
 - <https://enrico.bertini.io/teaching>



Quiz: Name marks/channels

- Shooting Media Coverage
- marks
 - A: points
 - B: lines
 - C: areas
- channels
 - A: position
 - B: color
 - C: length
 - D: area
 - E: angle

Mass Shootings By Race Of Shooter
NYTimes Coverage Of Mass Shootings By Race Of Shooter



<https://twitter.com/MonaChalabi/status/1158779046693679106?s=20>

Quiz: Name marks/channels

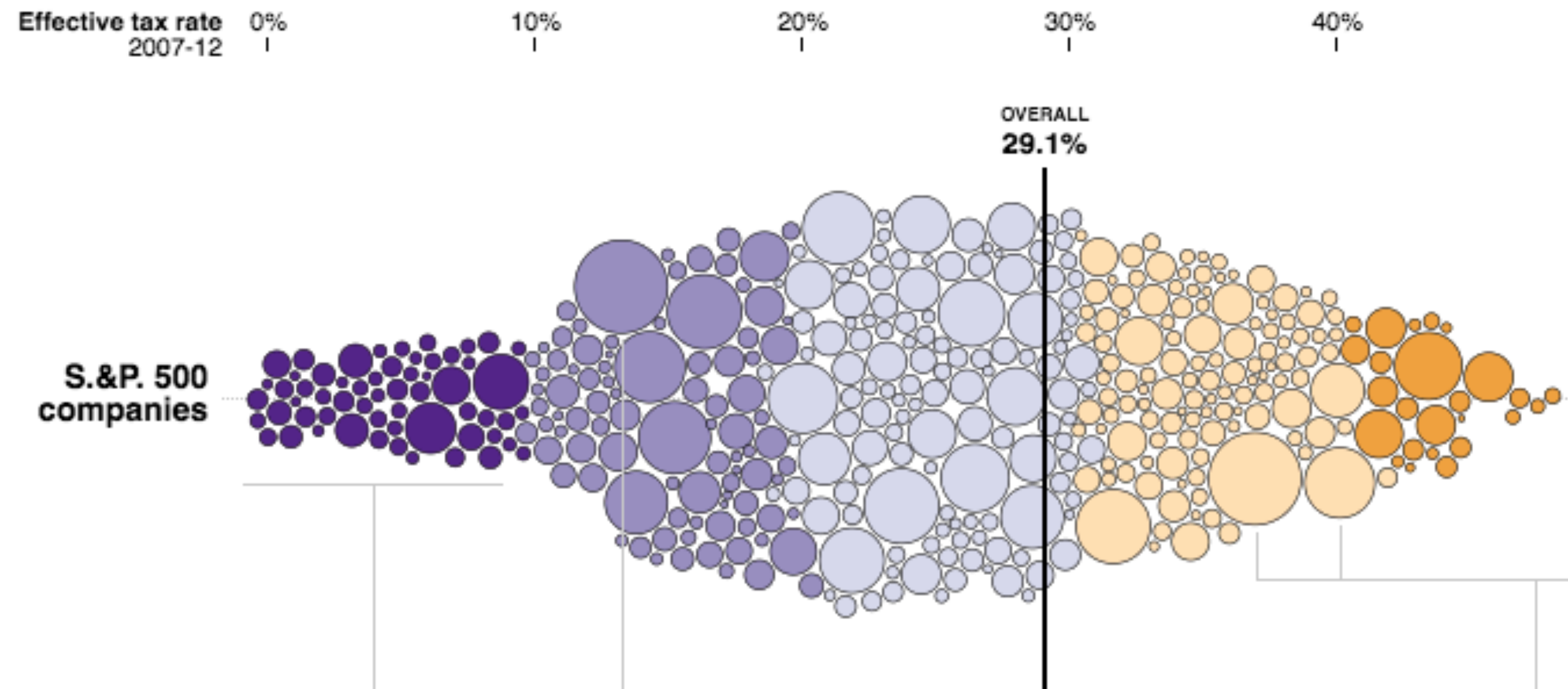
- Tax Rates

- marks

- A: points
- B: lines
- C: areas

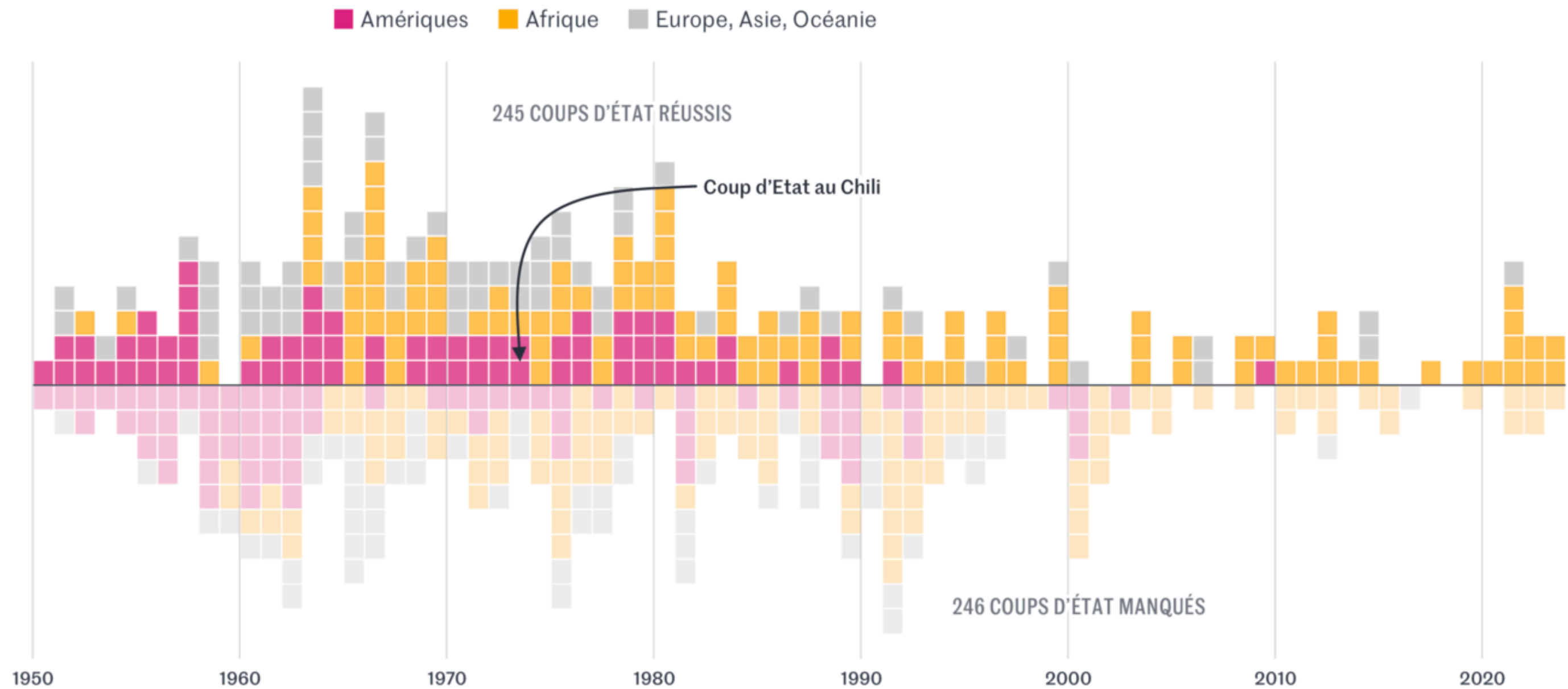
- channels

- A: position
- B: color
- C: length
- D: area
- E: angle



Quiz: Name marks

- points? lines? areas?



https://www.lemonde.fr/les-decodeurs/article/2023/09/11/depuis-1950-pres-de-cinq-cents-coups-d-etat-tentes-ou-reussis-surtout-en-amerique-latine-et-en-afrique_6188906_4355770.html

Many, many questions

- so what?
 - evidence that this design space could be improved!

Channels: Model evolves, heavily studied

- effectiveness rankings
- expressiveness matches, data & task

Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design

Jeffrey Heer and Michael Bostock

ABSTRACT

Understanding perception is critical to visualization design. With its low cost and scalability, crowdsourcing presents an attractive option for evaluating visualization design space of visualizations; however, it first requires a model of perception. In this paper, we assess the viability of Amazon Mechanical Turk as a platform for graphical perception experiments. We replicate previous studies of spatial encoding (e.g., dot plots, cartograms) and on chart size and gridline density. Our results demonstrate that crowdsourced perception experiments are viable and contribute new insights for visualization design. Lastly, we report cost and performance metrics from our crowdsourced studies.

ACM Classification: H5.2 [Information Systems]: User Interfaces—Evaluation/Assessment

General Terms: Experimentation, Human Factors

Keywords: Information visualization, user study, evaluation, Mechanical Turk, crowdsourcing

INTRODUCTION

“Crowdsourcing” is a relatively new phenomenon where web workers complete one or more small tasks for micro-payments on the order of \$0.01. Such services are increasingly attractive as a low-cost means of conducting user studies. By lowering the cost of recruiting participants, crowdsourcing provides almost immediate access to hundreds of users. Similarly, by reducing the burden on a subject pool is greatly increased and diversified.

The reduced cost structure of crowdsourcing is particularly attractive in visualization, where the space of possible visual encodings is large and complex [2, 7, 10, 19, 27, 34]. Crowdsourcing allows experimenters to canvas a wide range of visualization designs, effectively swapping ex-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that the copy is made for personal or classroom use and is not made or distributed for profit or commercial advertising or promotional purposes, for resale or redistribution. Copyright © 2014 ACM 978-1-4503-2916-6/14/0000-0000\$15.00.

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 20, NO. 12, DECEMBER 2014

1943

Ranking Visualizations of Correlation Using Weber's Law

Lane Harrison, Fumeng Yang, Steven Franconeri, Remco Chang

Abstract—Despite years of research on identifying the best visualization for a given task, we conduct a large scale (n=1687) experiment to assess how well used visualizations can be modeled using Weber's Law. We find that visualization by establishing that: 1) correlation judgment precision follows Weber's Law, 2) correlation judgment precision models provide a concise means to compare visualizations.

Index Terms—Perception, Visualization

1 INTRODUCTION

The theory and design of information visualization have advanced in many ways since Bertin's seminal work on the design of visualization. Years of visualization research has led to a number of design guidelines [5, 24] that aid the designer in choosing a visualization based on general data characteristics such as size, shape, and type. Unfortunately, many aspects of visualization are more art than science. For example, given a set of data, there are almost always multiple theoretically valid and therefore difficult to choose between. Beyond selecting a visualization for a given task, one must also take into account many other aspects of the design, such as color, shape, and size. These considerations such as context, and user experience, make it tremendously difficult for even experts to choose the most accurate and appropriate visualization.

One method for objectively identifying the best visualization is to conduct multi-factor human-subject experiments, each design or usage consideration being a factor, often resulting in a large number of possible visualizations. These experiments produce actionable results that generalize beyond the scope of the experiment. A key challenge in this area is to develop a method for visualization that becomes more widely adopted and that addresses the growing needs of the information visualization community.

What is needed then are quantitative models of visualization that are generalizable beyond a specific task while still providing designers with tradeoffs between “valid” visualization choices. Challenge conventional wisdom in information visualization [13], recent research has suggested that the psychology and cognitive science [12, 3] of human perception of certain data properties

CHI 2019 Paper

CHI 2019, May 4–9, 2019, Glasgow, Scotland, UK

Measuring the Separability of Shape, Size, and Color in Scatterplots

Stephen Smart

University of Colorado Boulder
Boulder, Colorado
stephen.smart@colorado.edu

Danielle Albers Szafir

Color Difference Comparisons

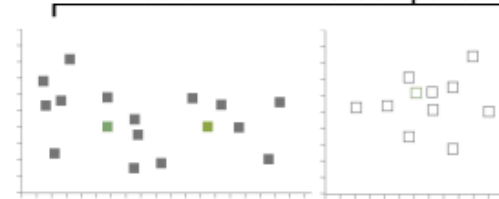


Figure 1: We examine how mark shape, size and color affect color difference judgments, with findings showing an interdependence between size and shape categories.

ABSTRACT

Scatterplots commonly use multiple visual channels to code multivariate datasets. Such visualizations often use shape, size, and color as these dimensions are considered “separable”—dimensions represented by one channel do not significantly interfere with viewers’ abilities to perceive another. However, recent work shows that size significantly impacts color difference perceptions, raising broader questions about the separability of these

23 Jul 2021

Rethinking the Ranks of Visual Channels

Caitlyn M. McColeman*, Fumeng Yang*, Timothy F. Brady, and Steven Franconeri

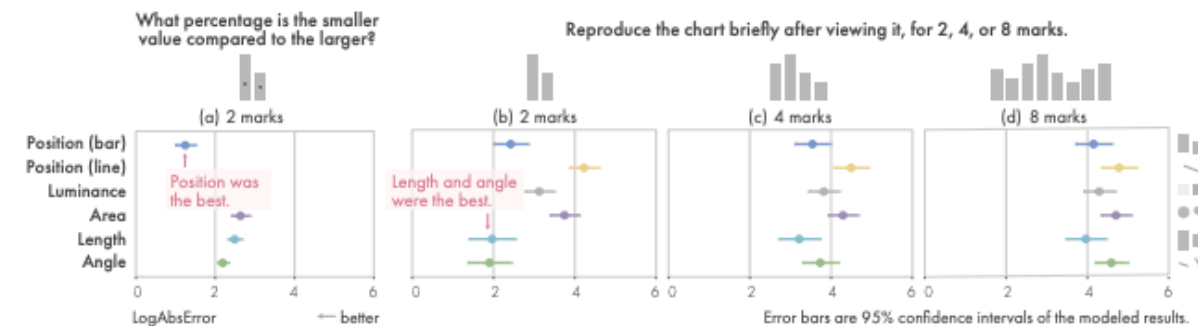
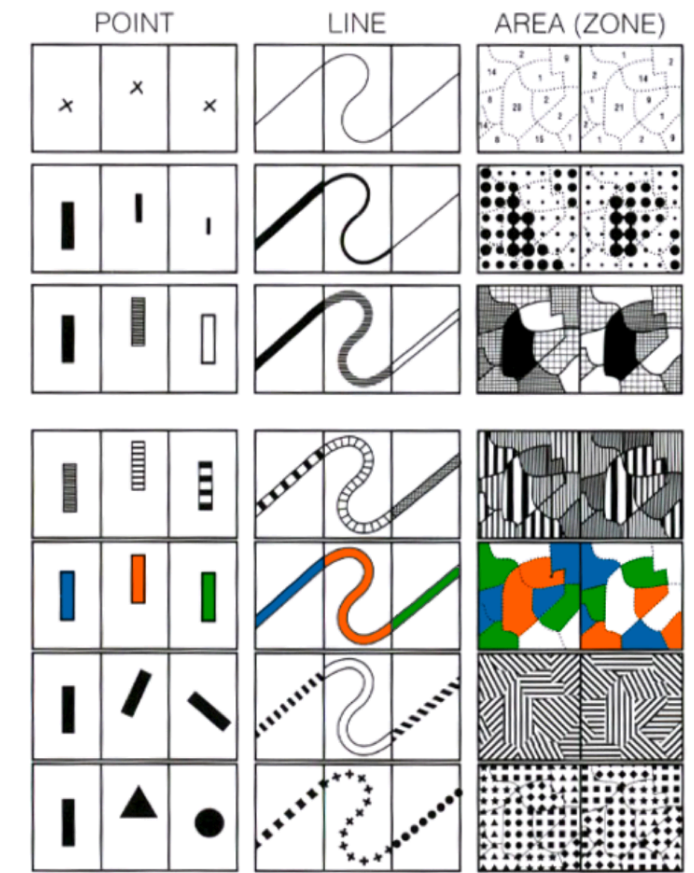


Fig. 1. One core guideline for data visualization design is that some visual channels offer better perceptual precision than others, and drawing those precision estimates from two-value ratio judgment tasks [17]. (a) This figure depicts typical data (from [33], 50 participants) showing these judgments are more precise for position (e.g., bar graphs) than for area (e.g., bubble charts). We tested whether that ranking generalizes to the new task of reproducing 2 to 8 previously seen values, and analyzed reproduction bias, precision, and error using a Bayesian modeling approach. (b) This figure shows our modeled results (49 participants). The ranking did not hold, and other factors besides channel choice—like the number of values in the series—had an order of magnitude more influence on performance.

Marks: Model stays static

- model inherited from Bertin (Semiology of Graphics, 1967)
 - never questioned
- geometric motivation
 - geometric primitives have dimensions
 - how could we argue with math?!



→ Points



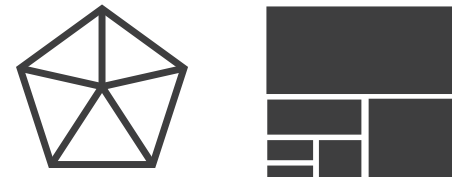
0D

→ Lines



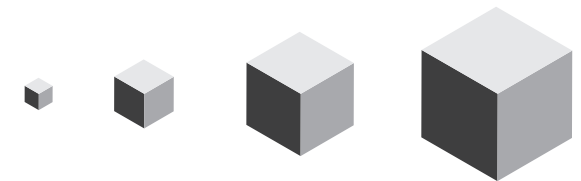
1D

→ Areas



2D

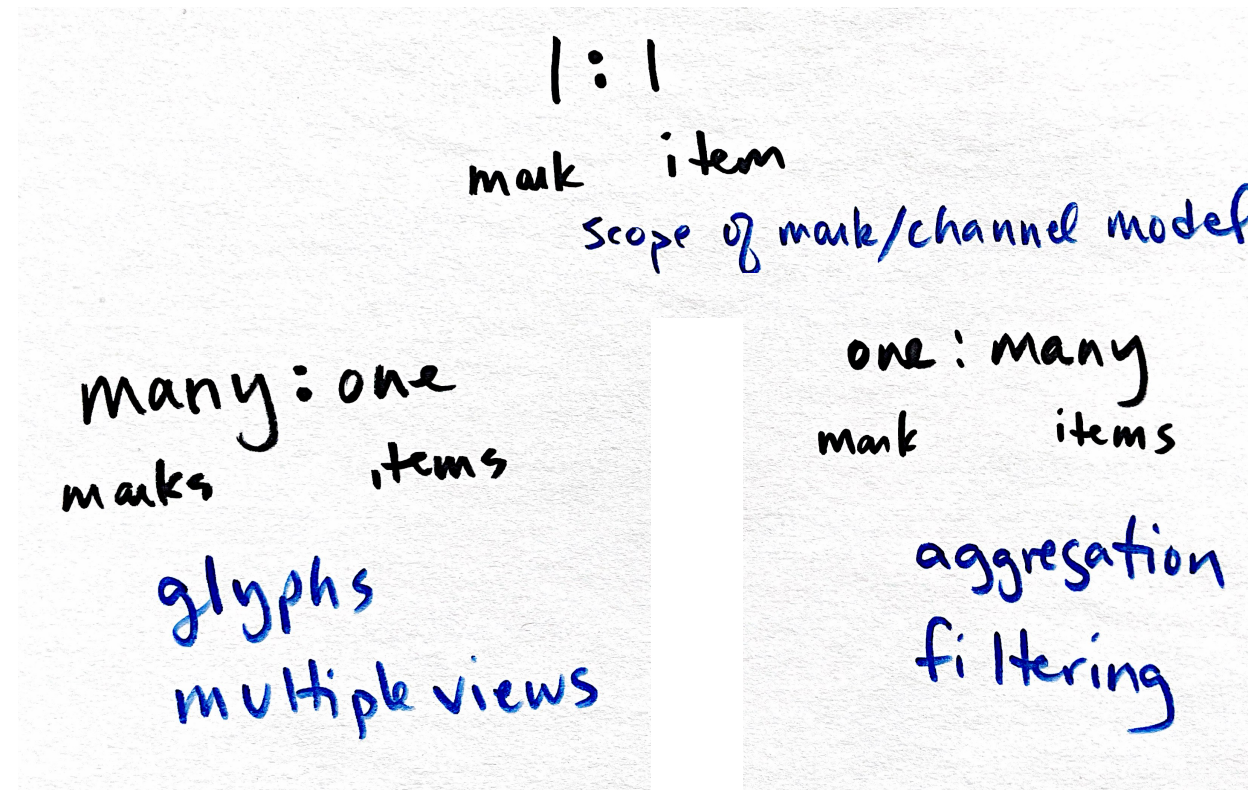
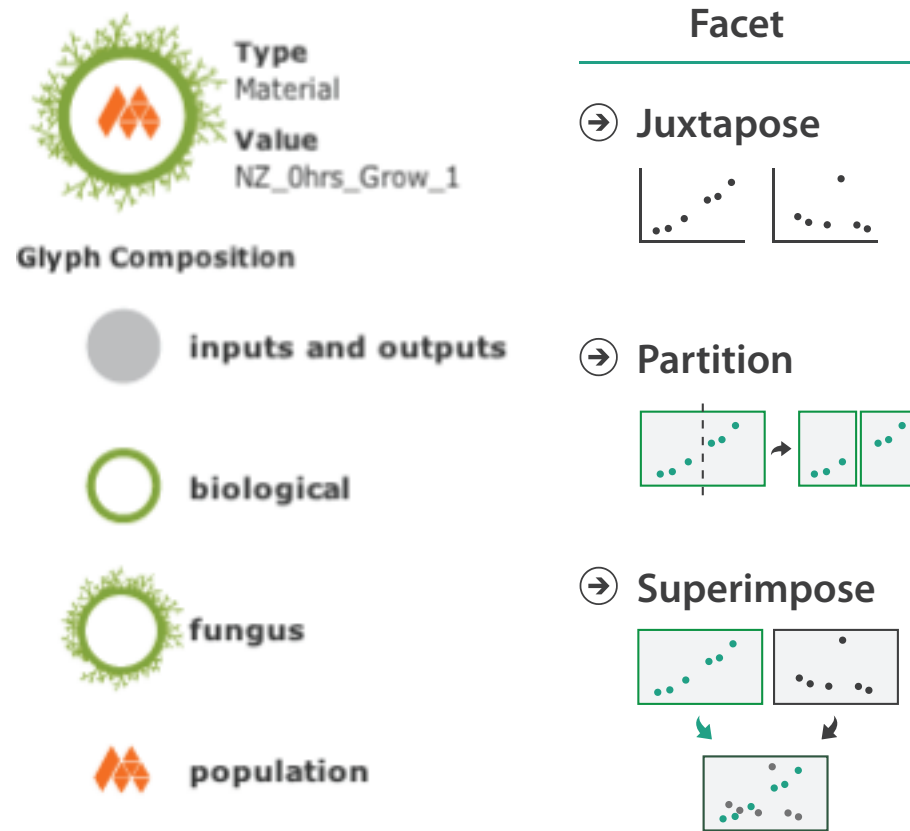
→ Volume



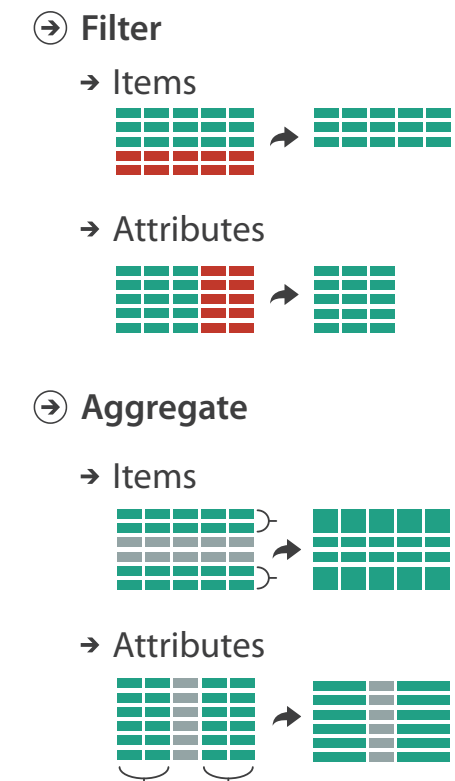
3D

Mark/channel analysis: scope & limits

- model scope: one mark for one data item
 - multiple marks for one item: glyphs, multiple views
 - one mark for many items: aggregation, filtering



Reducing Items and Attributes



Encoding vs decoding models

- **Encoding** model: what should visualization **designer** do?
 - prescriptive model, providing guidance for design
- **Decoding** model: how will visualization **viewer** interpret?
 - predictive model, informed by vision science & perception research
 - predicting viewer response differs from inferring or reverse-engineering designer intent when encoding!

Encode vs decode: Where do models diverge?

- idiom: pie chart

- **encode:** area marks with angle channel: **2D area varies**

- ordered radially, uniform length

- accuracy: area less accurate than rectilinear aligned position/length

- **decode:** not angle! probably arc length, maybe also area

- <https://eagereyes.org/blog/2016/an-illustrated-tour-of-the-pie-chart-study-results>

- idiom: coxcomb chart

- **encode:** marks with length channel: **1D length varies**

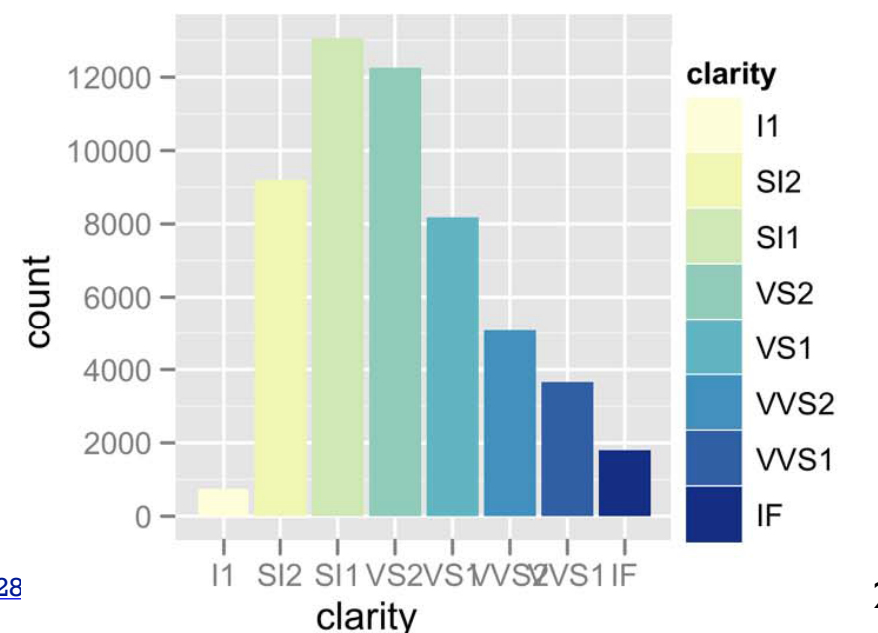
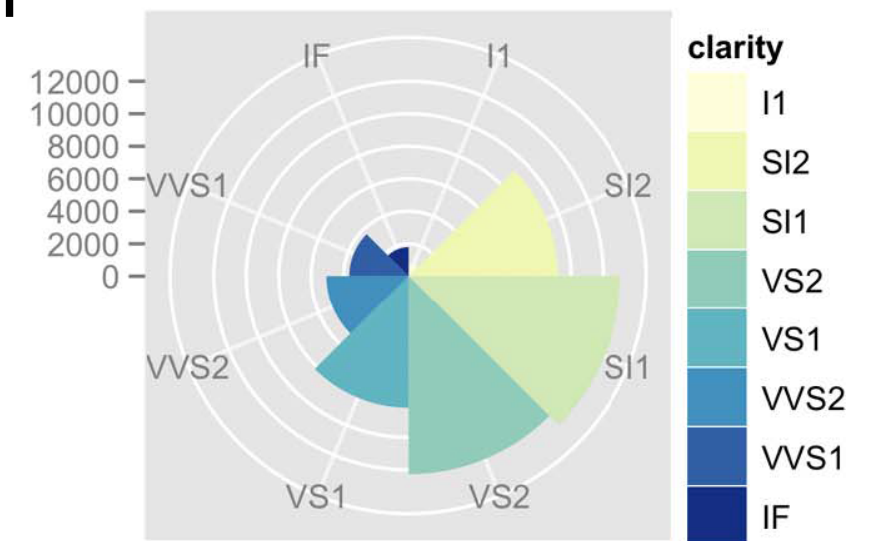
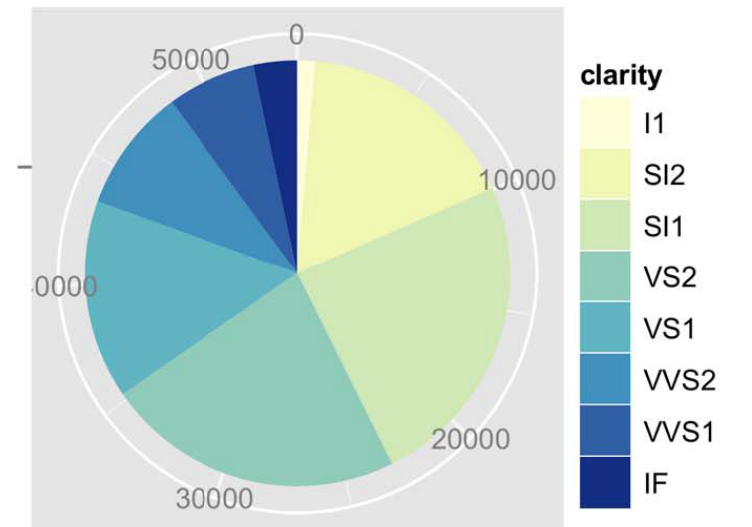
- ordered radially, uniform width

- more direct analog to bar charts, but using radial layout

- what's the mark type?

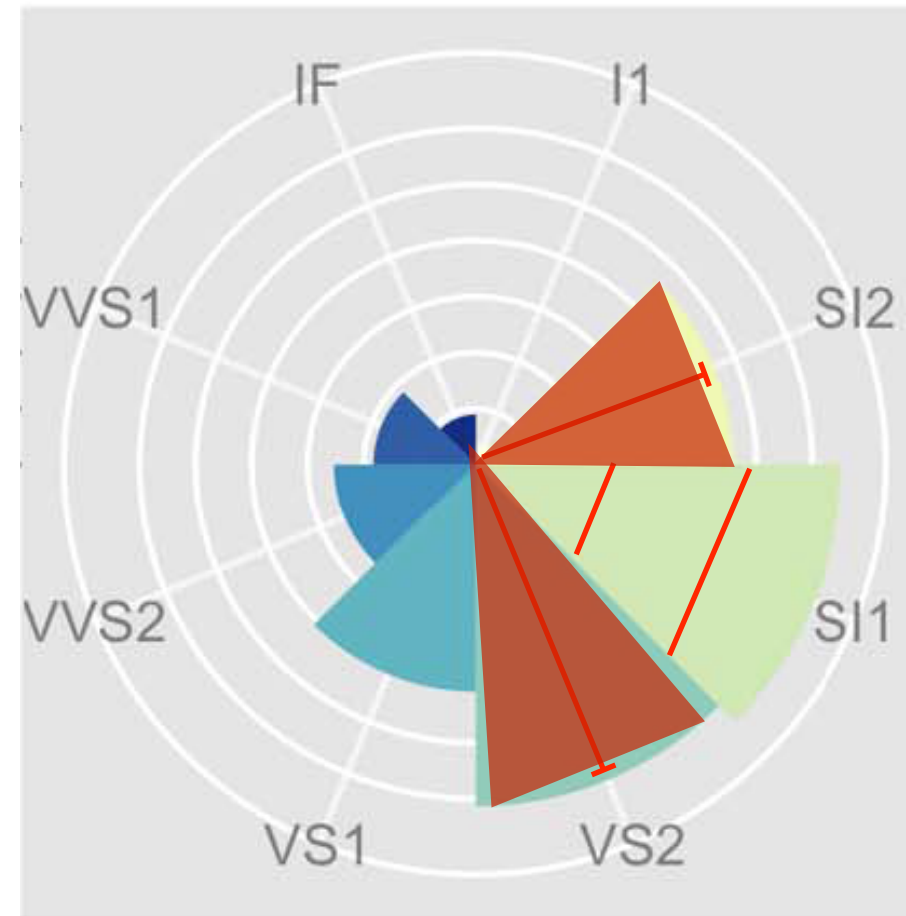
- line, because it's length coded?

- area, because area varies too?

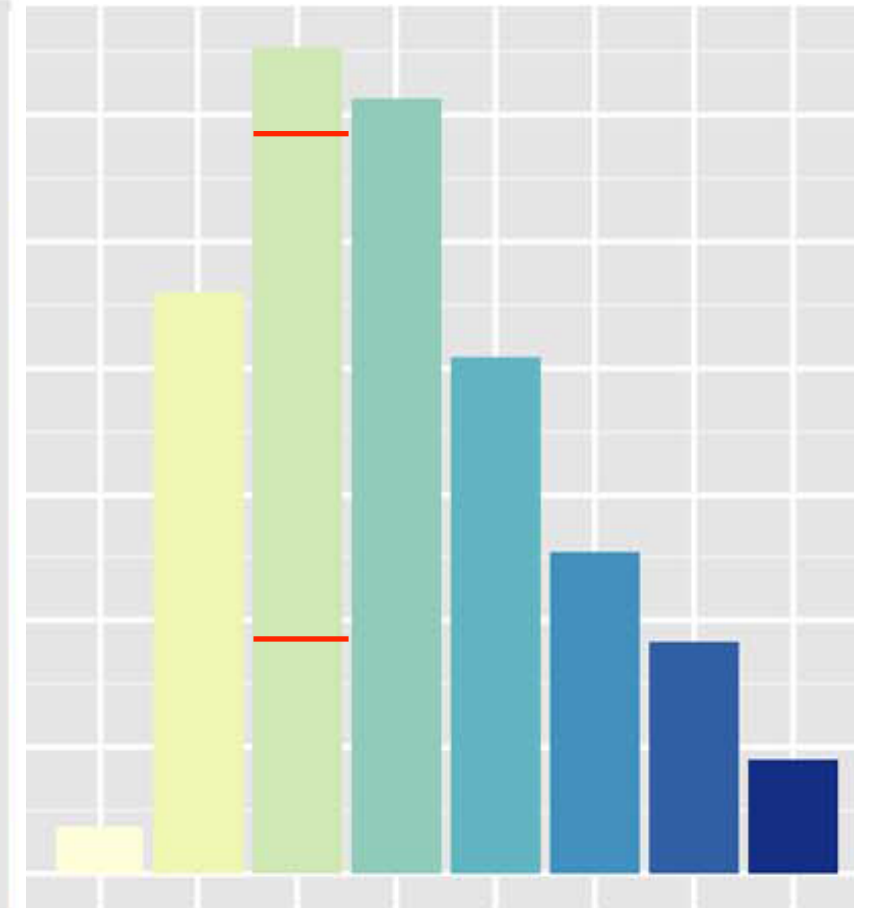


Encode vs decode: Where do models diverge?

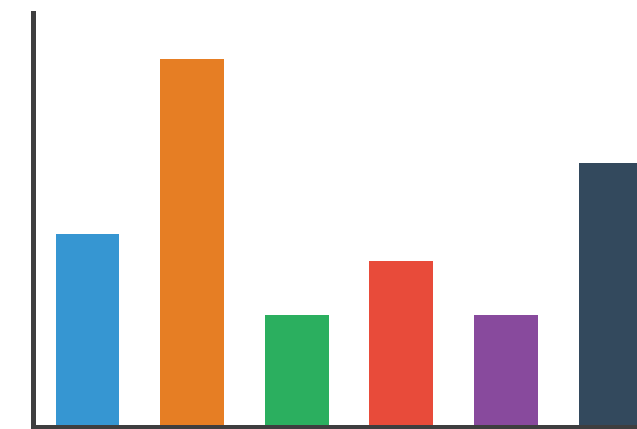
- encode: **ID size (length)**
- decode/perceive: **2D area**
- nonuniform line/sector width as length increases
 - so area variation is nonlinear wrt mark length!
- bar chart safer
 - uniform width, so area is linear with mark length
 - both radial & rectilinear cases
- **encode vs decode divergence**
 - if channels differ, which "wins"?



nonuniform width as length increases



uniform width as length increases



radial & rectilinear bars: uniform width as length increases

Alternative Ideas

Constraints

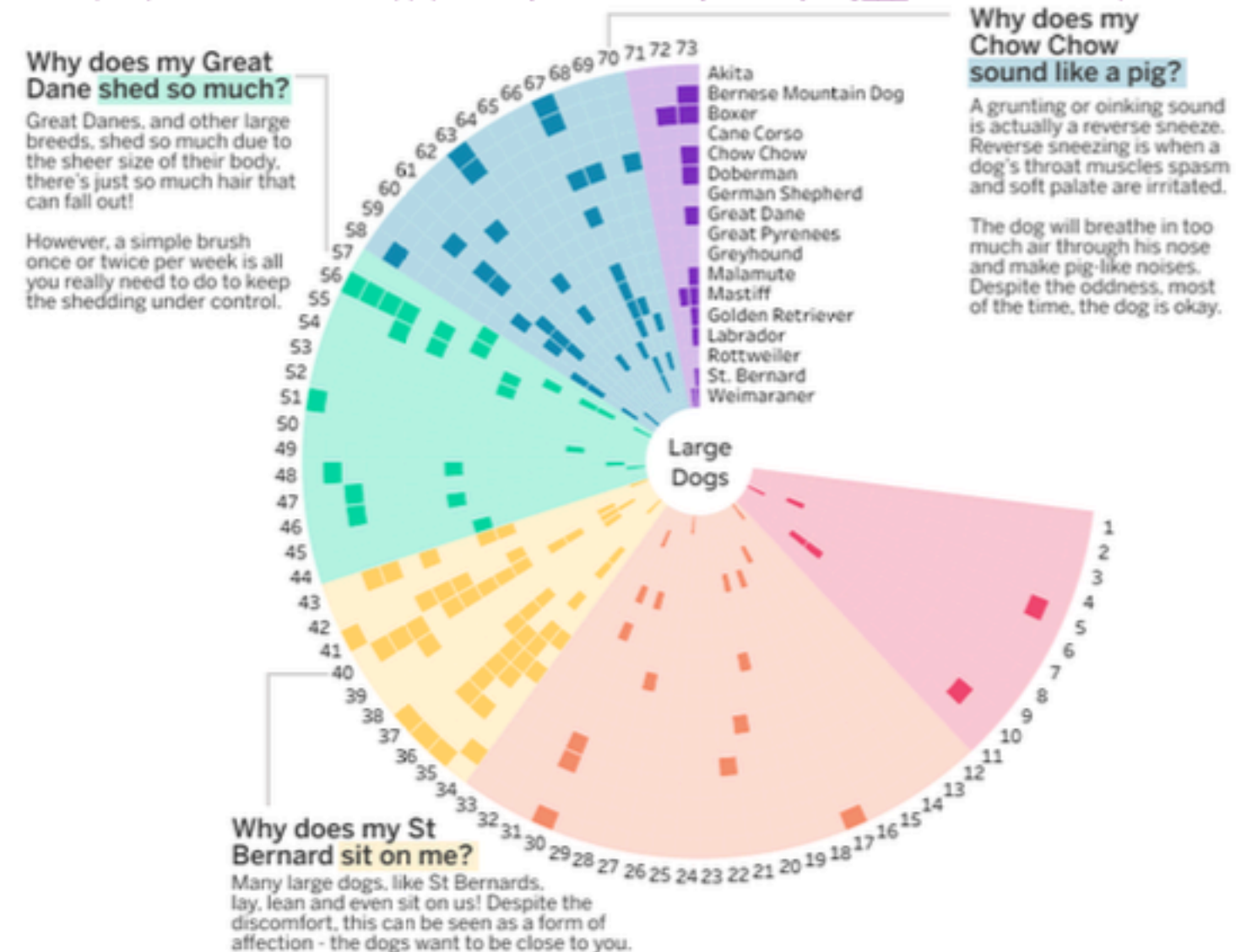
- consider marks and channels as imposing constraints
 - when does mark type constrain channel use?
 - when does using one channel constrain another channel?

Channel use: what does it mean?

- Does channel size encode attribute?
 - yes? sizes differ
 - according to dog name in alphabetical order
 - no! size differences not meaningful
 - just emerges from choice of layout, radial vs rectilinear
 - not a "real" attribute encoding
- Can we use size channel to encode another attribute?
 - no! not free
 - it's "taken" already, would change meaning
- Size channel is Unavailable

Q5 Marks & Channels: Why Does My Dog [4 pts]

Credit: https://public.tableau.com/app/profile/wjsutton/viz/WhyDoesMyDog_IronQuest/DesktopVersion



https://public.tableau.com/app/profile/wjsutton/viz/WhyDoesMyDog_IronQuest/DesktopVersion

Channel Availability Model

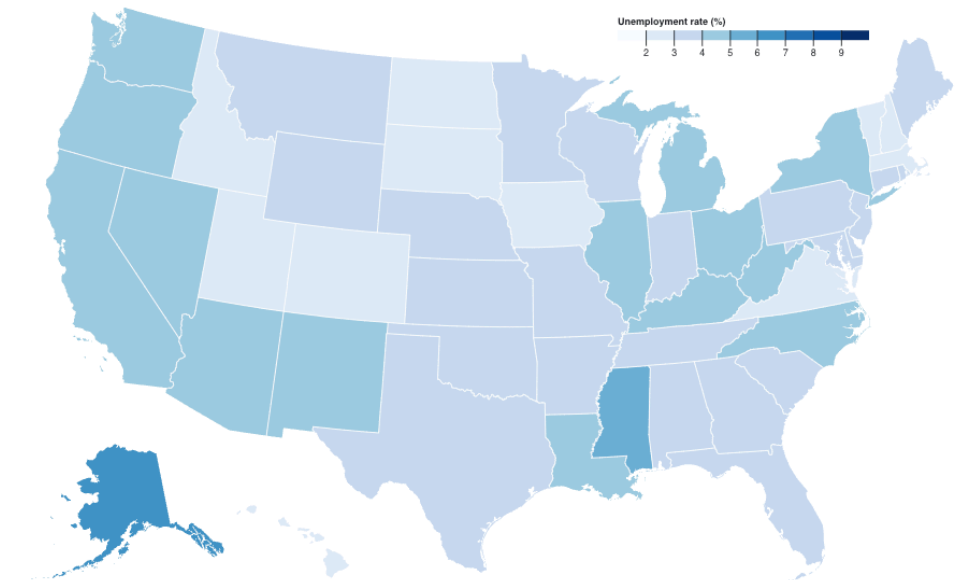
- **Encoded:** which channels directly used to encode attributes?
 - clear meaning
 - multiple channels can be directly used for redundant encoding
- **Free:** which channels free to encode another attribute?
 - without changing usability of existing encoding
- **Unavailable:** which channels unavailable / precluded / taken?
 - because of mark type?
 - because of idiom/algorithm design specifics?
 - because other channels used?

Area marks: Rethinking

- area marks is a terrible name
 - other marks all have graphical area too
 - allowing us to encode with color
 - computer graphics point of view: they're all just polygons
 - there's also an "area" channel, which is confusingly different
 - area is not the only channel in play with these marks!

Area marks

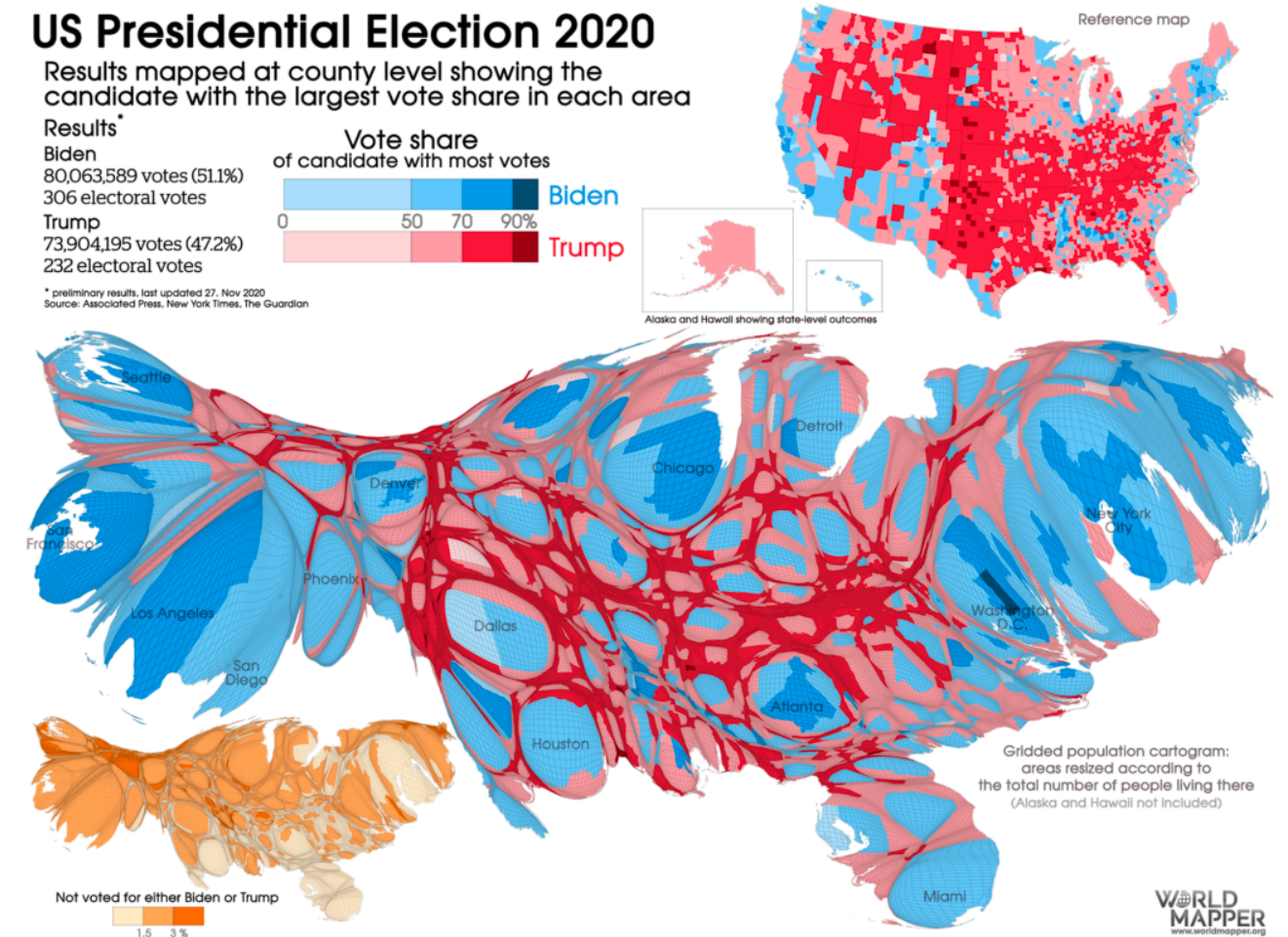
- obvious example: choropleth maps
- what can we do to California? could we encode additional data?
 - cannot shrink/grow (size channel)
 - cannot translate (position channel)
 - cannot rotate (orientation channel)
 - cannot reshape (shape channel)
 - why not?
 - would lose meaning of that mark: boundary is the data
 - also lose meaning for other occluded marks
- "area" mark is not specific enough
 - AreaPositionOrientationShape mark??? nah...
 - idea: **interlocking**



<https://observablehq.com/@d3/us-state-choropleth>

Interlocking (area) marks

- many channels locked down with interlocking marks
 - boundary encodes meaning
 - cannot change size, shape, position, orientation
 - **mark type imposes constraints**
- but...
 - what about cartograms?
 - cannot change just one mark (California)
 - but could change them all!
- interlocking marks as global constraint:
 - cannot change just one independently
 - but can change all simultaneously!
 - typically with algorithm

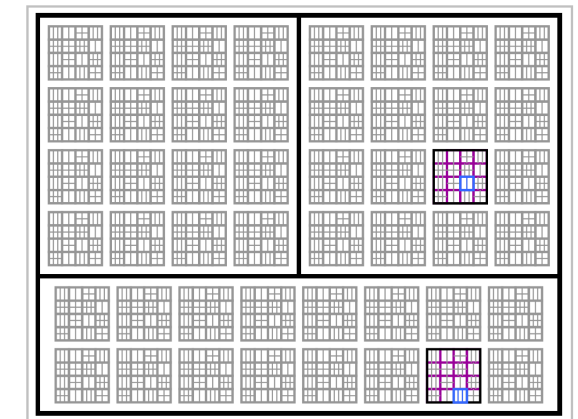
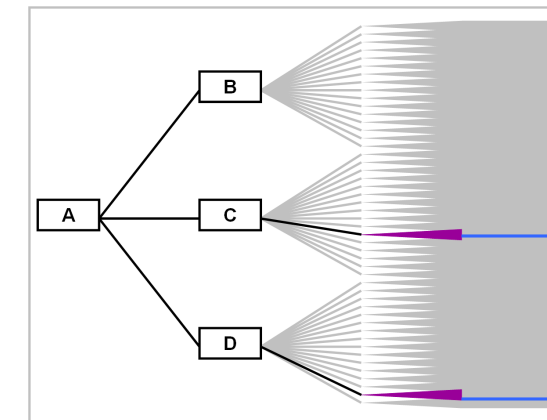
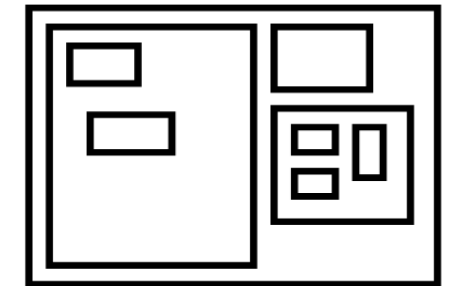
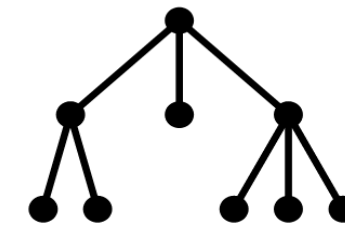
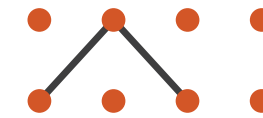


<https://worldmapper.org/us-presidential-election-2020/>

Interlocking marks: Non-spatial

- example with non-spatial data?
- treemaps
 - show hierarchy with containment, not connection
 - encode additional attribute with area/size
- again, cannot change just one mark alone
 - but could recompute layout to change all at once
- combined layout of all marks together carries meaning
 - unlike spatial data mark boundaries
 - individual mark boundaries have no intrinsic meaning

→ Connection → Containment



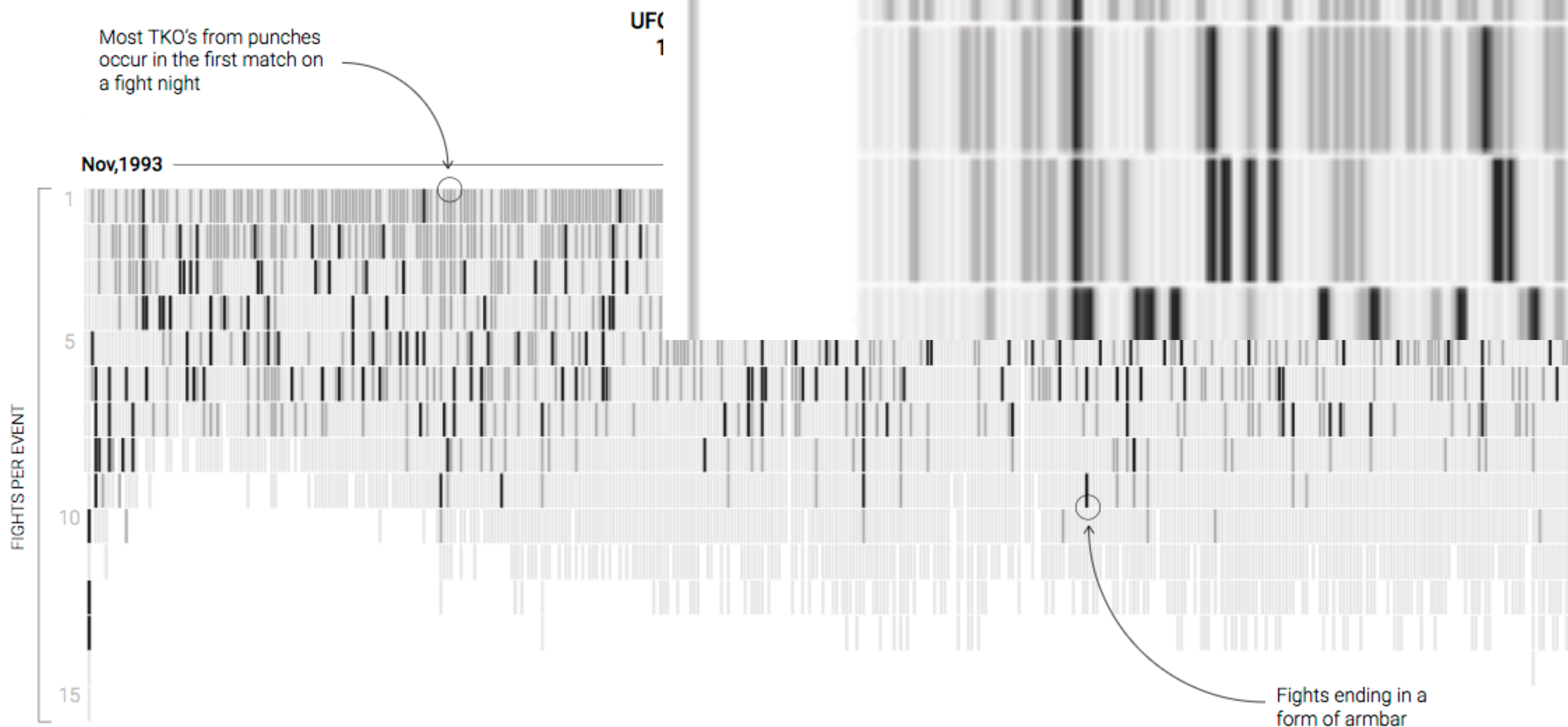
Node-Link Diagram

Treemap

[Elastic Hierarchies: Combining Treemaps and Node-Link Diagrams. Dong, McGuffin, and Chignell. Proc. InfoVis 2005, p. 57-64.]

Quiz: Name that mark

- UFC fights: points? lines? areas?

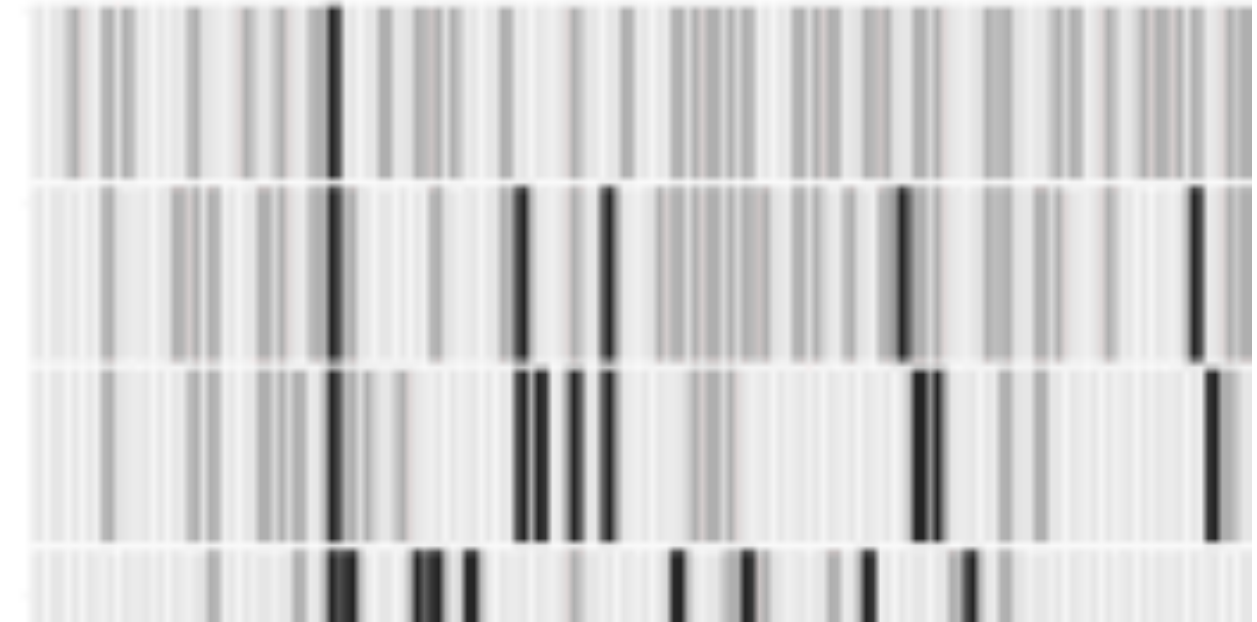


Nov, 1993

Analyzing marks

- what type of mark?
 - line?
 - no, not length coded
 - point mark with rectangular shape?
 - 2020: yes!
 - 2023: no!
 - cannot change position / size / orientation
 - area?
 - 2020: no, area/shape does not convey meaning
 - 2023: yes!
 - fully interlocking
 - position, size, shape, orientation all locked

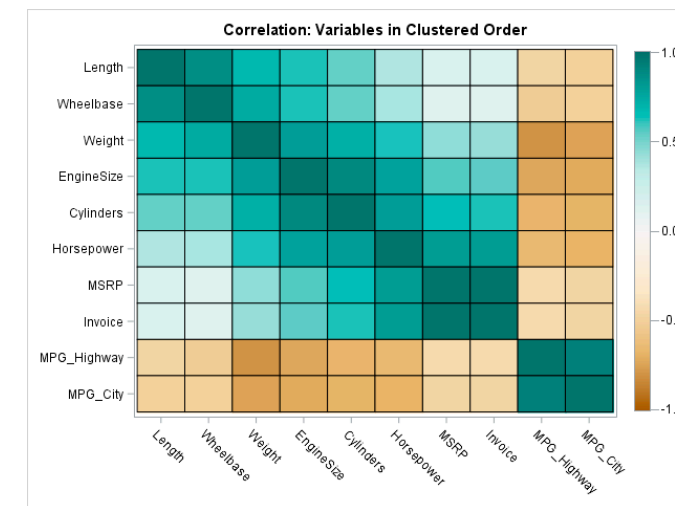
Nov, 1993



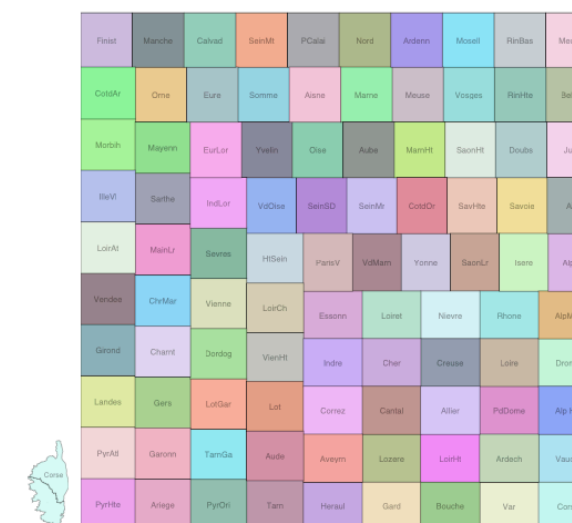
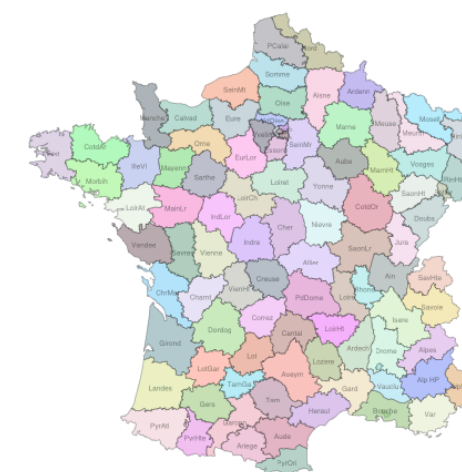
<https://multimedia.scmp.com/infographics/sport/article/3010883/bruce-lee-and-mixed-martial-arts/>

Interlocking marks: Tile heatmaps

- 2D matrix/grid as index
 - position in use as index
 - size/area & shape & orientation all equal (& locked down)
- simplest possible case of interlocking marks?
 - more regular than choropleths or treemaps
 - but underlying similarities
- full extent of cell used for color coding
 - different from using a point mark within the cell



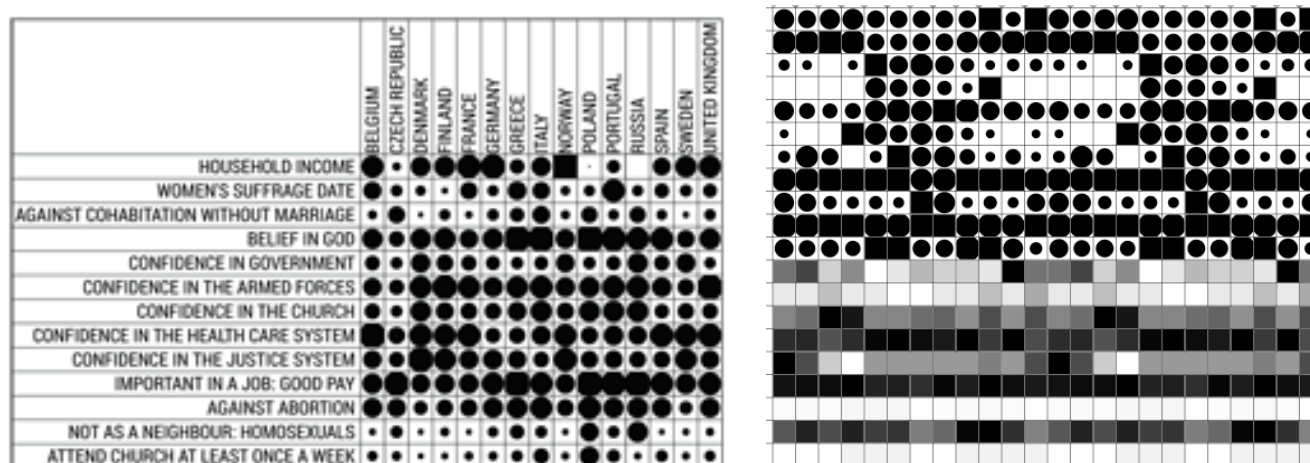
<https://blogs.sas.com/content/iml/2018/05/02/reorder-variables-correlation-heat-map.html>



Spatially ordered treemaps.

Wood and Dykes.

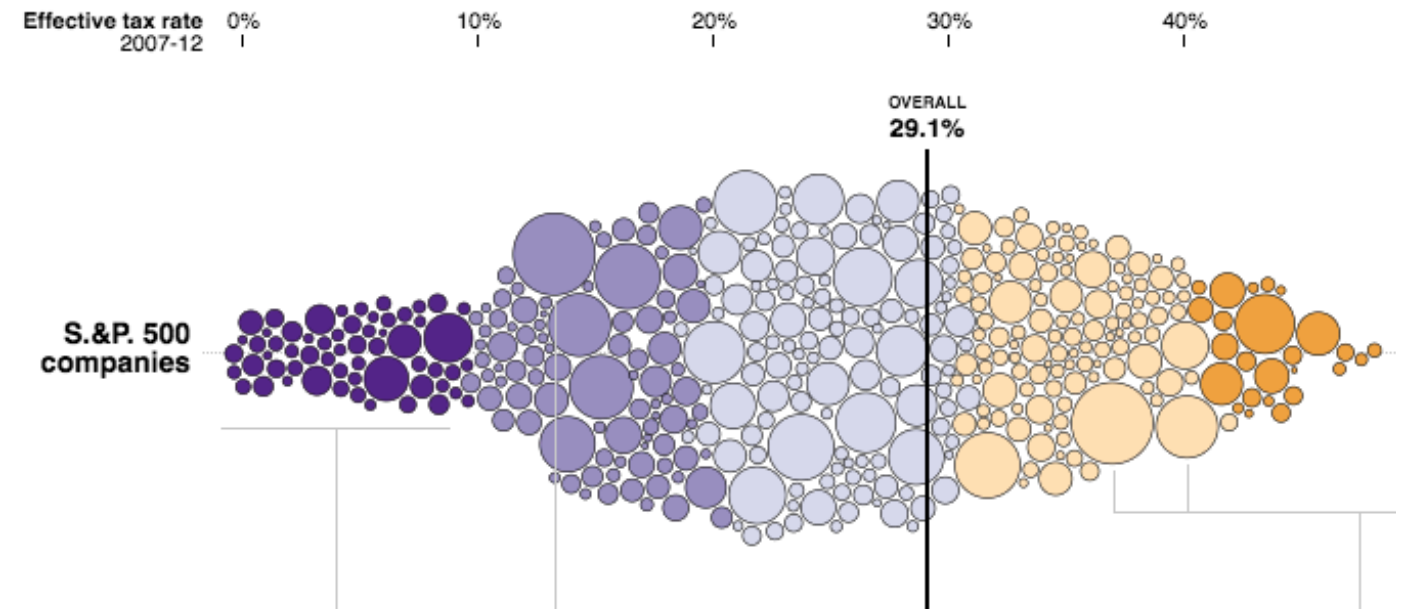
IEEE TVCG (Proc. InfoVis) 14(6):1348-1355, 2008.



Bertier, Charles Perin, Pierre Dragicevic, Jean-Daniel Fekete (2014). [Revisiting Bertin's Matrices: New Interactions for Crafting Tabular Visualizations](#). TVCG, VIS' 2014.

Interlocking marks: Circle packings

- also are interlocking marks, **not** size-coded point marks
 - more like treemap than scatterplot!
- channel availability analysis: customized circle packing
 - Encoded channels
 - horizontal position: encodes tax rate
 - color: rate, redundant with horizontal position
 - size (2D area): market cap
 - Free channels
 - motion
 - Unavailable channels
 - vertical position: used by algorithm to avoid overlap & minimize gaps
 - shape & orientation equal and unavailable: can't just change, would need to redo layout



<https://archive.nytimes.com/www.nytimes.com/interactive/2013/05/25/sunday-review/corporate-taxes.html>

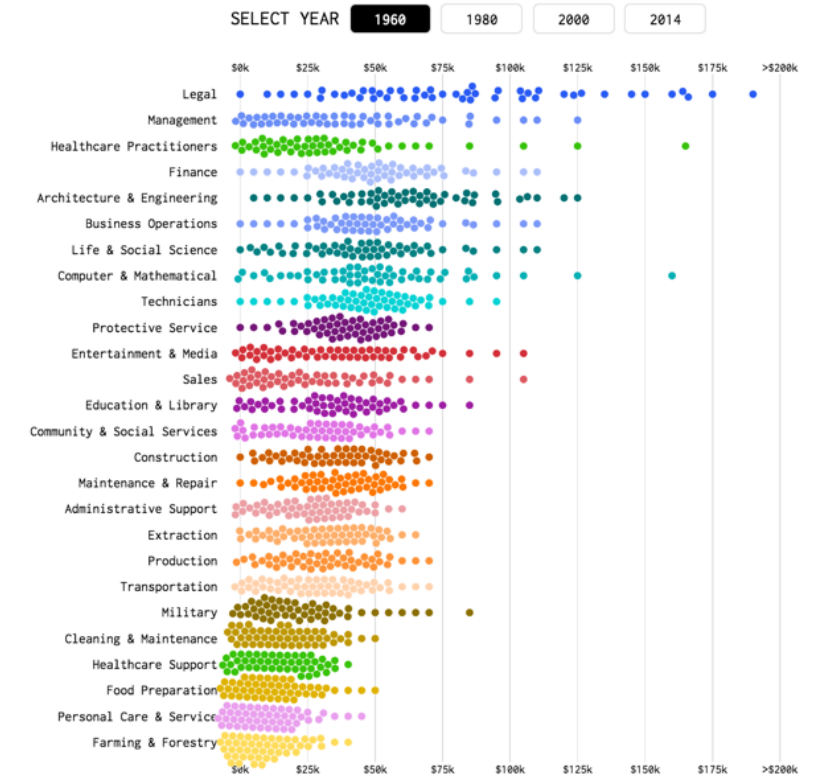
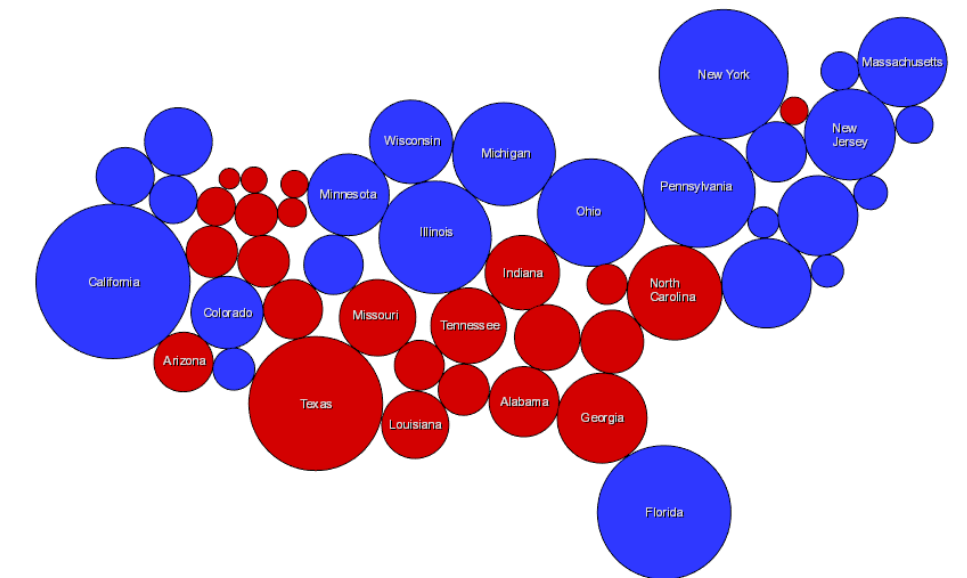
Interlocking marks: Circle packings

- customized circle packings are special case
 - including beeswarm plots
- general circle packing
 - algorithmic constraint: no overlaps, minimal gaps
 - position unavailable since used by algorithm



- Dorling cartogram

- can treat as special case of circle packing, with additional constraints to maintain relative position from geographic location
- throw away shape by regularizing to circles
- add size coding

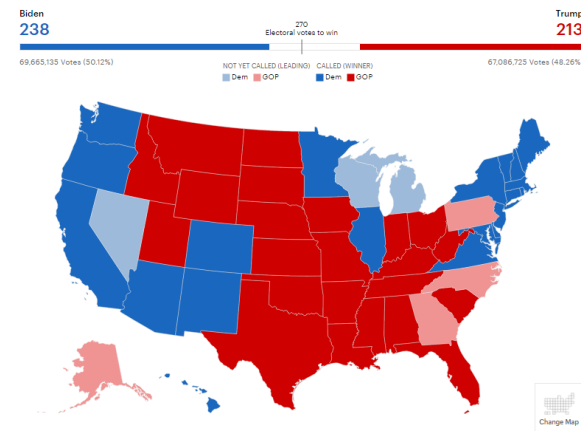


<https://flowingdata.com/2016/06/28/distributions-of-annual-income/>

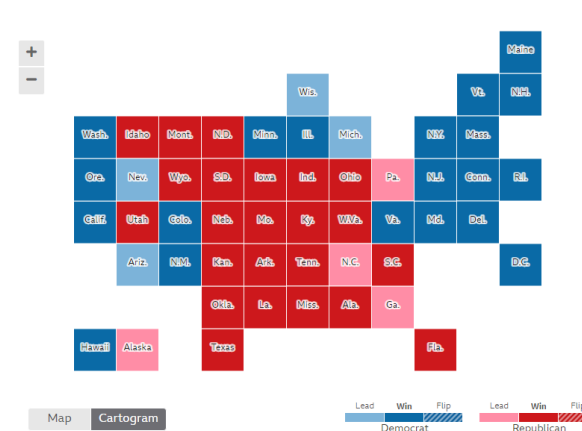
Interlocking? Election maps roundup

- yes interlocking
 - A: already covered
 - B/C: equal-area alg algorithm simplifies shape

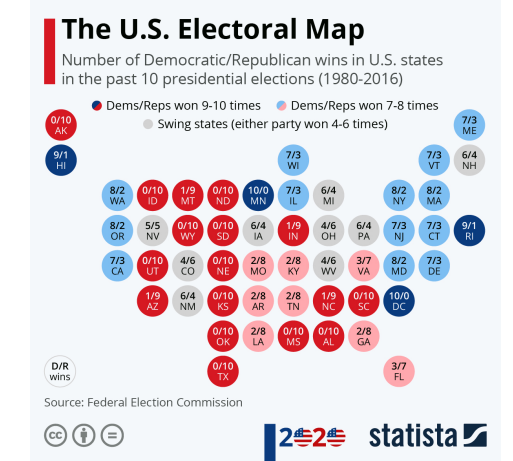
A. Standard Choropleth



B. Equal-area cartogram



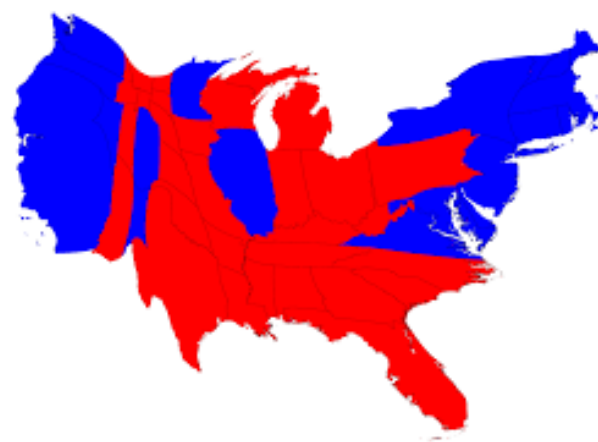
C. Equal-area cartogram



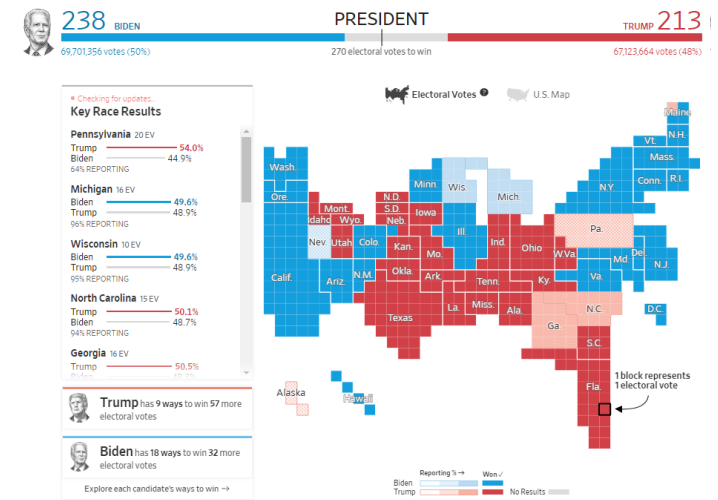
Area not encoding election data

- yes interlocking
 - E/F multi-level
 - top level: interlocking marks
 - bottom level: square units
 - E/F: countability for votes
 - F whitespace: population density

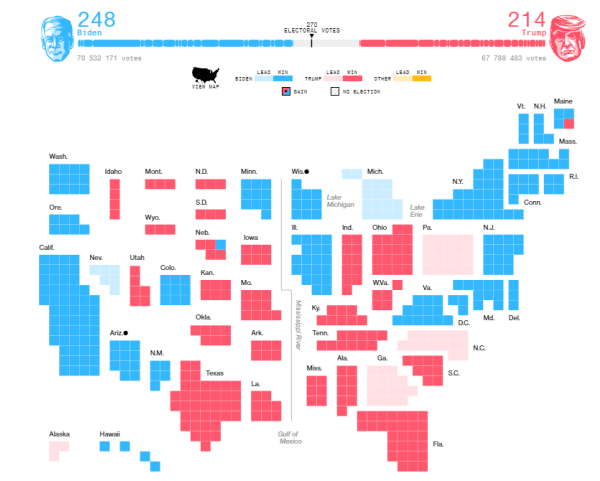
D. Area-proportional continuous cartogram



E. Units assembled into state-like shapes



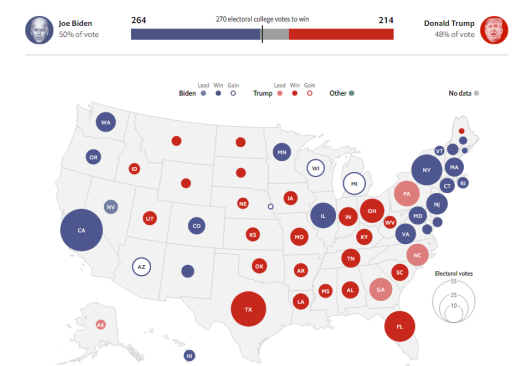
F. Units in state-like shapes spaced apart



Area encoding election data

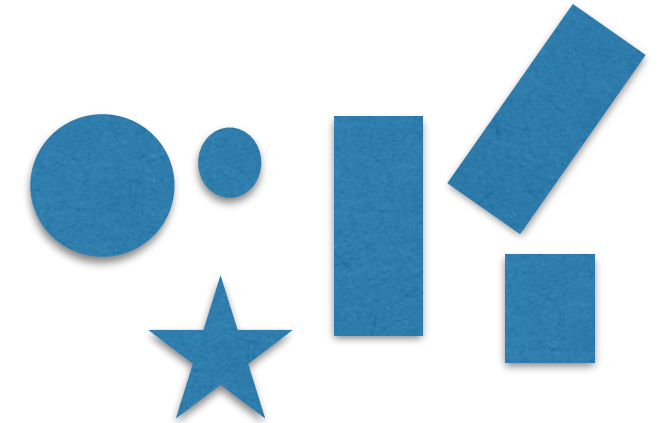
- no, point marks
 - size coded by area

E. Map with bubbles



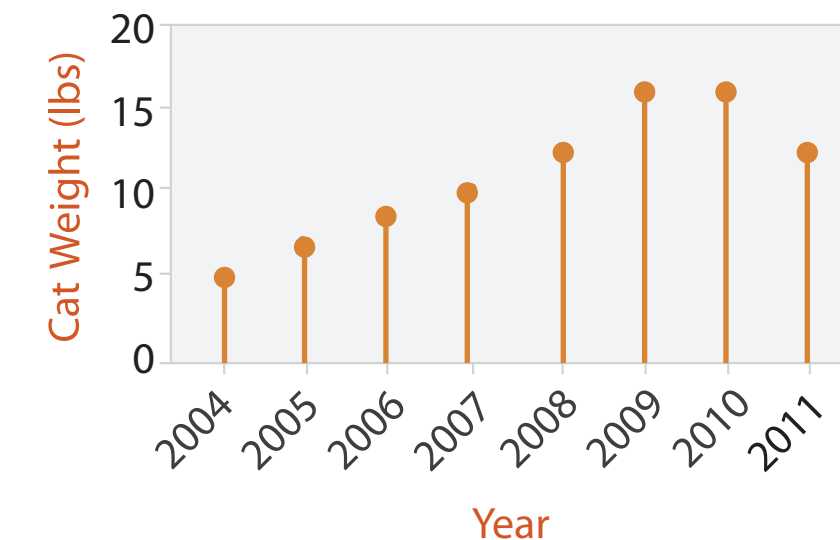
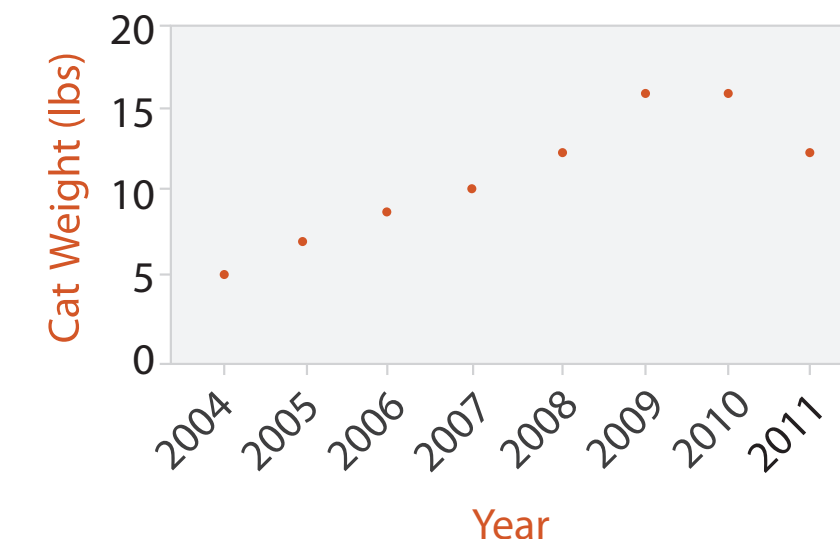
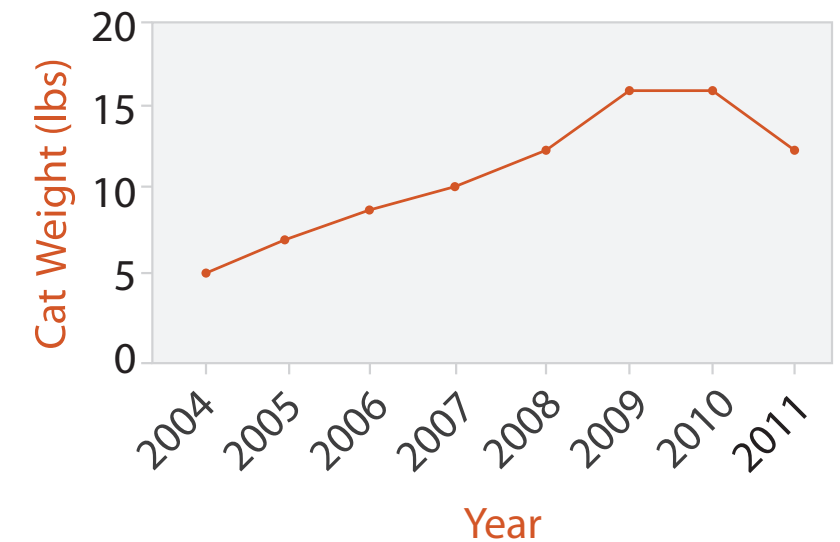
Distinguishing marks through constraints

- highly constrained: interlocking marks
 - many channels unavailable: size, position, shape, orientation
 - proposal: rename from "area" to "interlocking"
- unconstrained: point marks
 - can encode more info with any channel at all!
 - size, position, shape, orientation
 - color, motion, ...
 - does "point" imply circular shape?
 - proposal: is "unconstrained" a better / more evocative name?
- so... what about line marks?



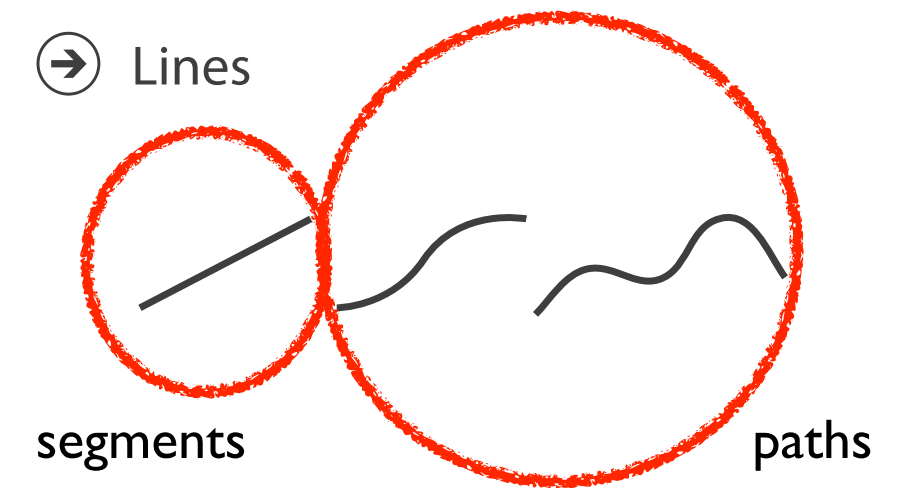
Line marks: Rethinking

- do line charts use line marks?
 - construct connections between points
 - trend task: emphasize relationships between items
 - may or may not show points explicitly
 - no! not like bar charts or lollipop charts do...
 - **line chart encodes many items, not just one**
 - with many piecewise-linear segments or smooth curve



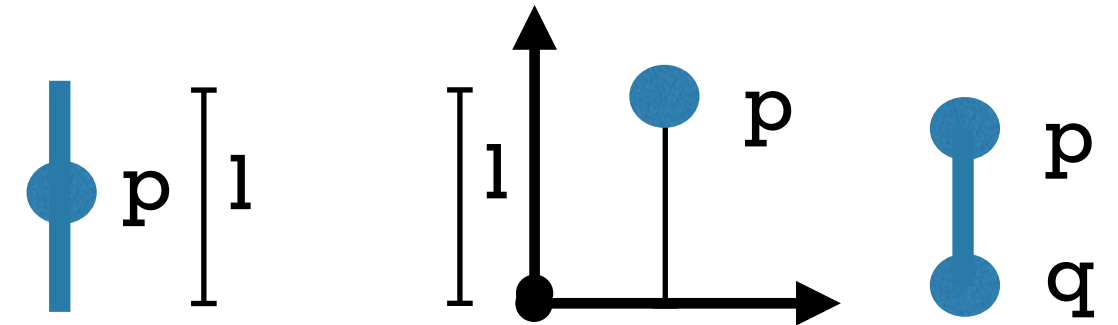
Line marks: Naming two cases separately?

- line segments showing single item, vs curved lines showing multiple items
 - should we reason about them separately instead of analyzing them together?
 - line segment: express single quantitative attribute for one item with length
 - single mark represents single item of data
 - proposal: call these "segments"
 - curved / complex lines
 - proposal: call these "paths"
 - single mark represents many items of data

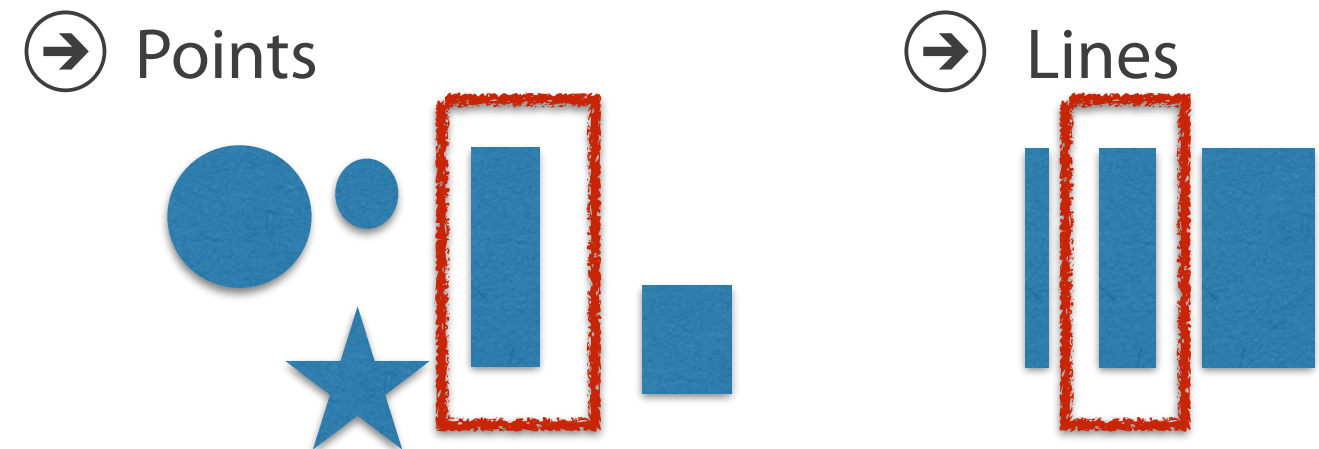


Line marks vs point marks

- what's relationship between length channel and length of line?
 - exactly the same? confusingly different?
- how does line segment differ from "length-coded point mark"?
 - two numbers, either centroid/length or max/min

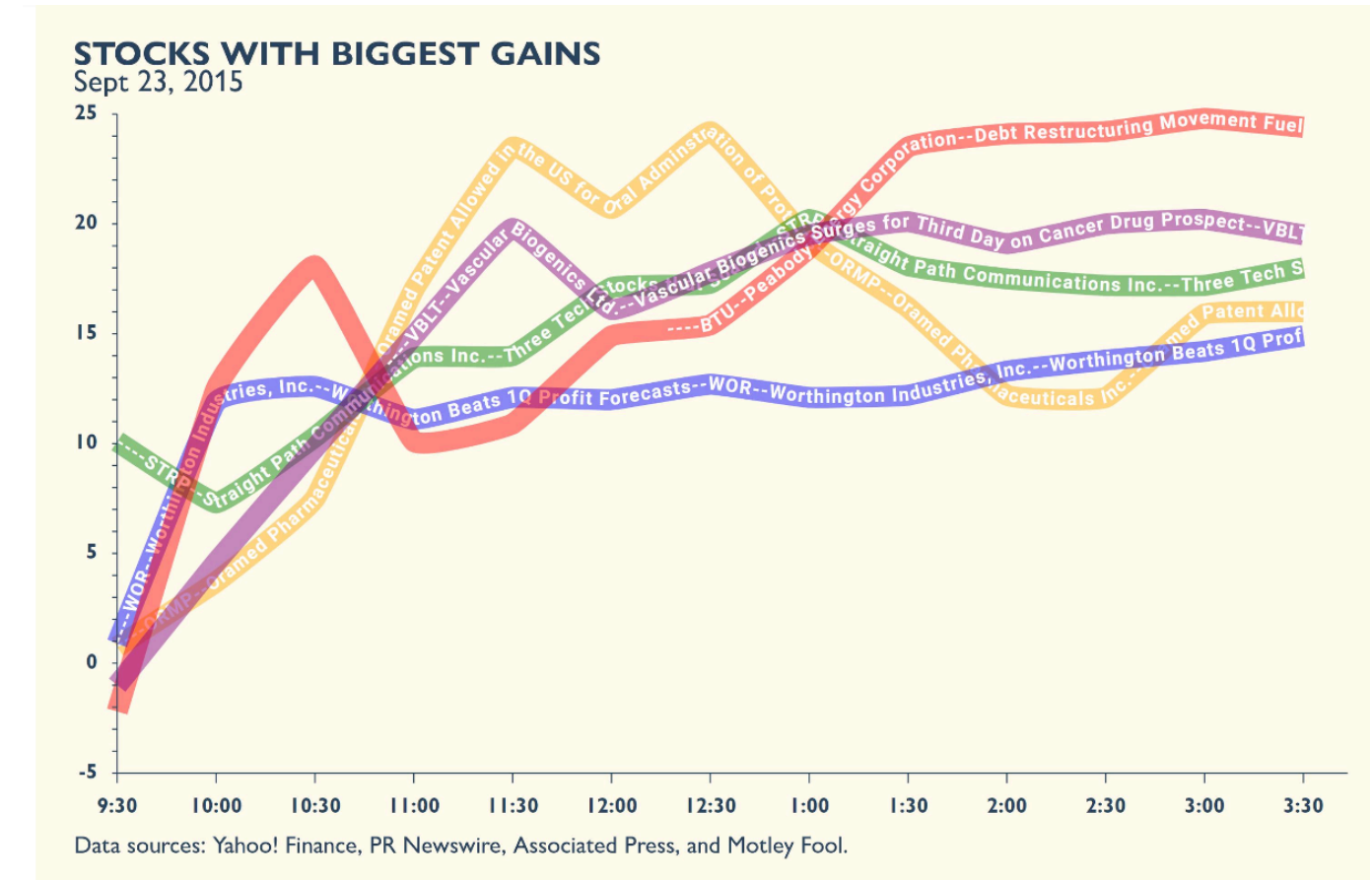


- proposal: what if line segment marks and point marks belong in same "singleton" category?
 - to distinguish from multi-item marks



Line marks vs area marks

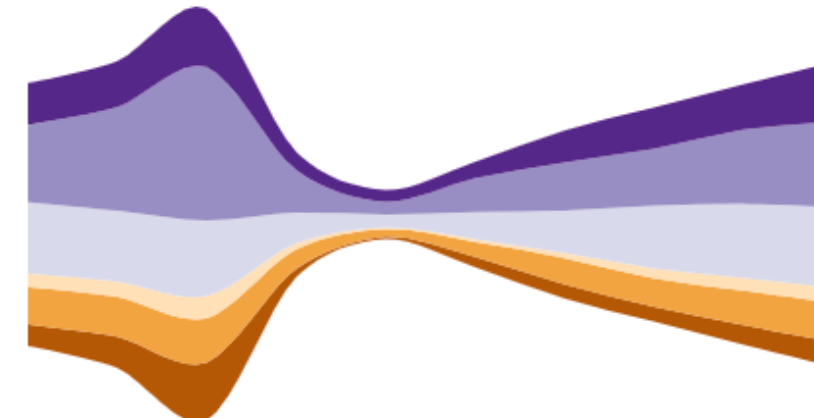
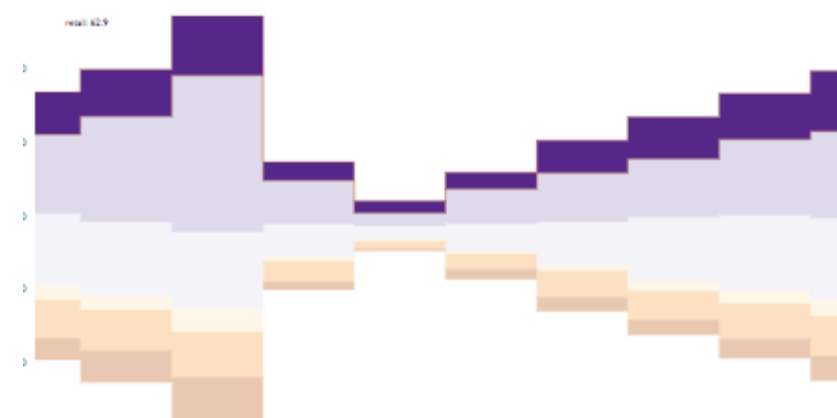
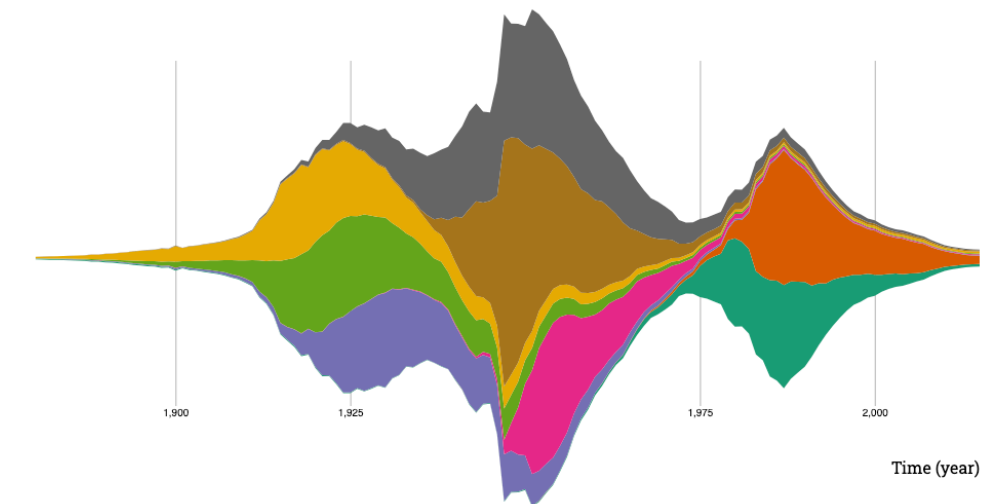
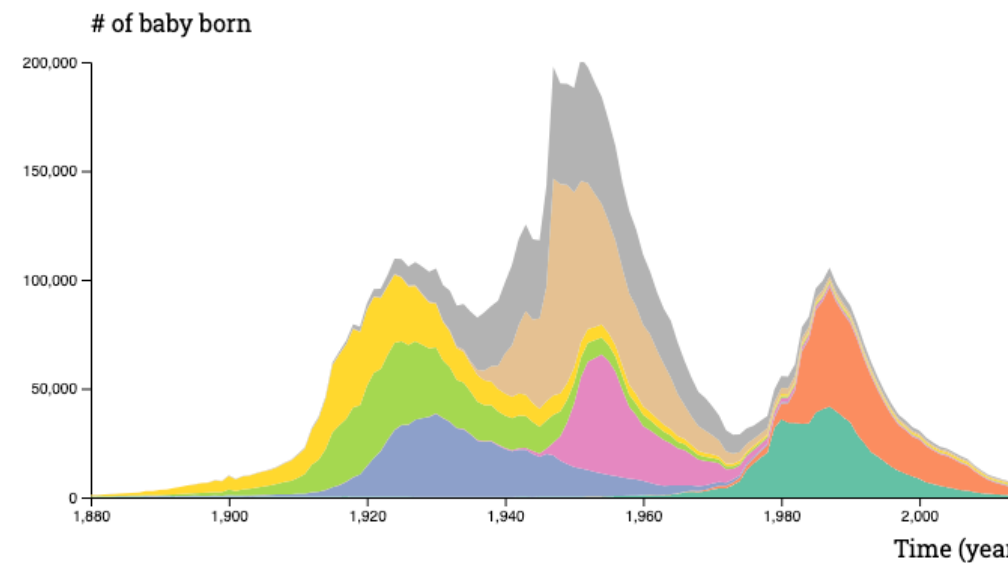
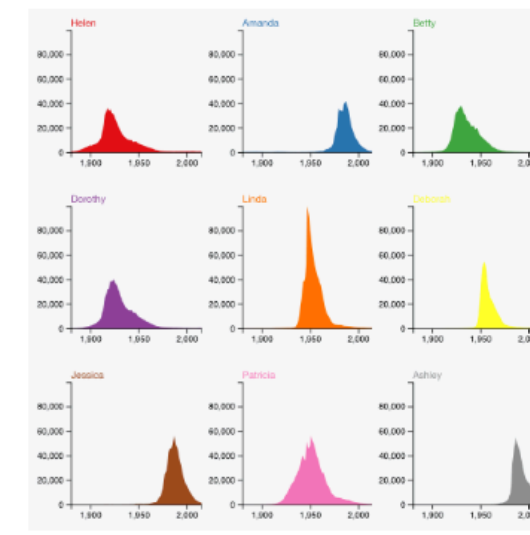
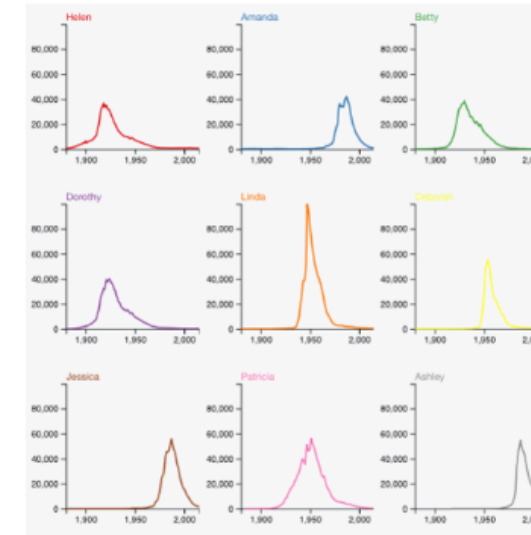
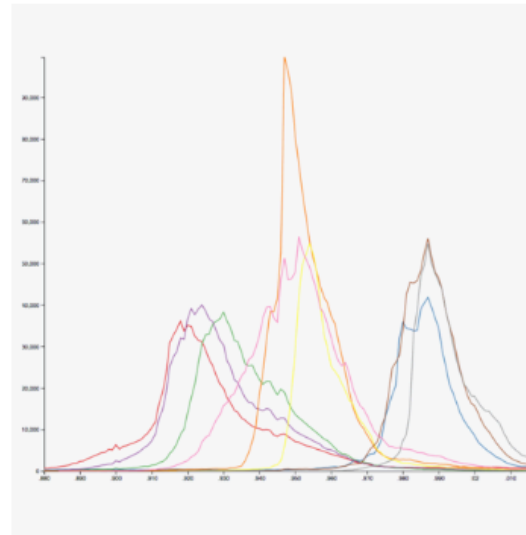
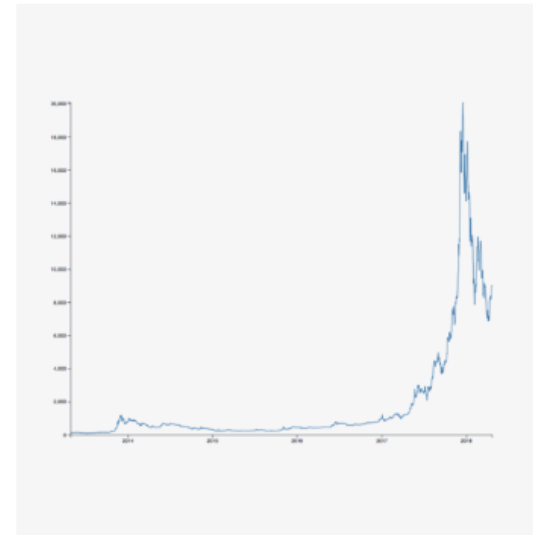
- what's the border between line path marks and area marks?
 - if path is wide enough, is it an area?
 - what if there's information shown in region inside path, within its boundary?
 - different color for inside vs outside, or even text?
- what about the region below path?...



Brath. Visualizing with Text (Fig 7.14). CRC Press, 2020.

Line charts vs filled area charts

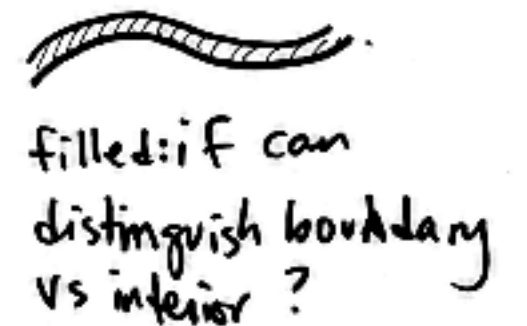
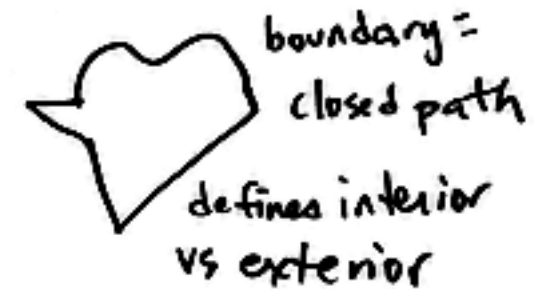
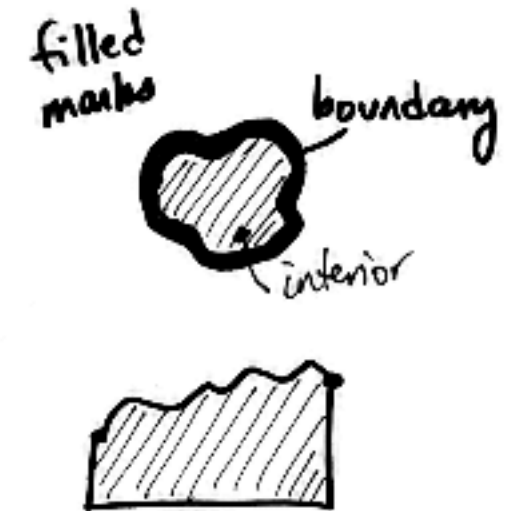
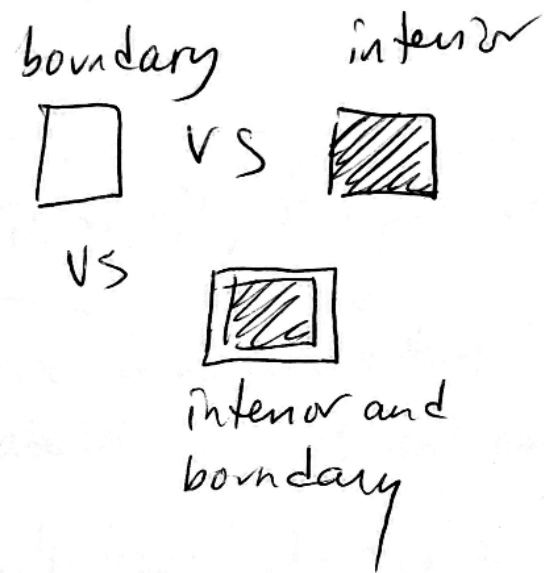
- should we reason differently about
 - line chart boundary vs filled area chart interior?
 - stacked area charts vs streamgraphs?
 - discrete stacked bar charts vs continuous streamgraphs?
- what matters?
 - boundary vs interior?
 - discrete vs continuous
 - occlusion?



<https://d3-graph-gallery.com/line.html>
<https://d3-graph-gallery.com/area.html>
<https://d3-graph-gallery.com/stackedarea.html>
<https://d3-graph-gallery.com/streamgraph.html>

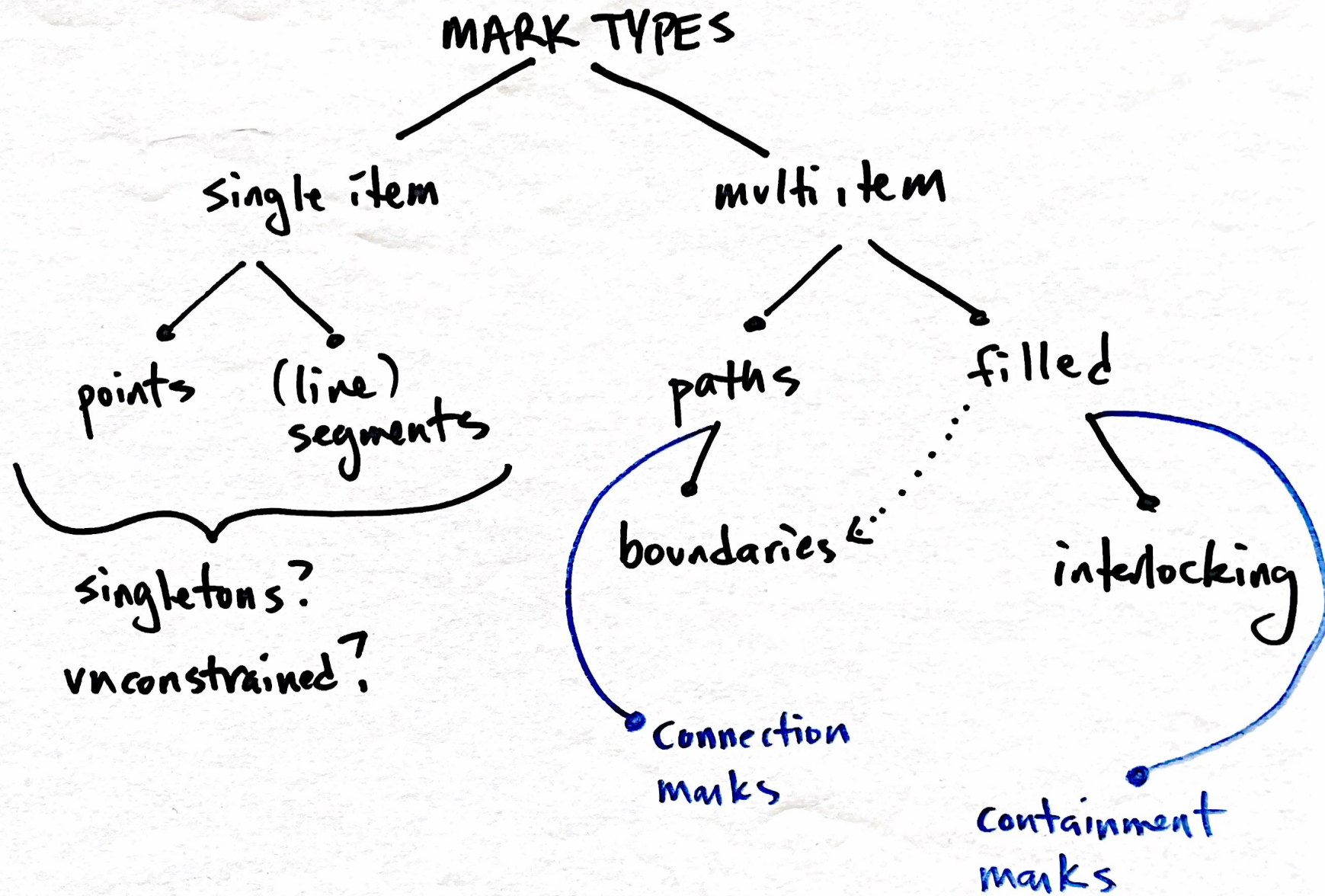
Boundaries vs interiors

- proposal: if path is closed, call it a "boundary"
 - (maybe also if path is infinite)
 - all boundaries also define "interior" region
 - distinction may or may not be visually highlighted



- proposal: use name "filled" mark instead of "area"
 - then interlocking marks are a special case of filled marks

Alternative mark types model



1 mark:
per item
single



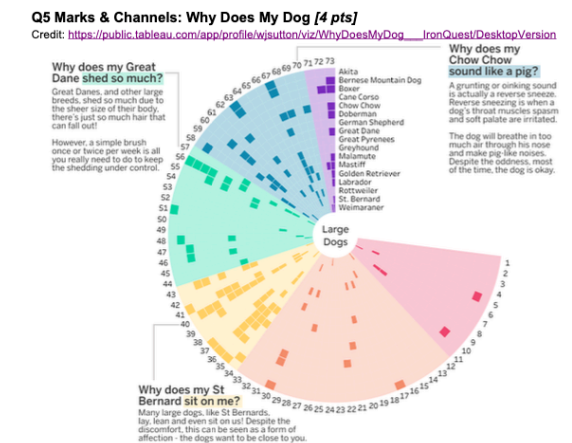
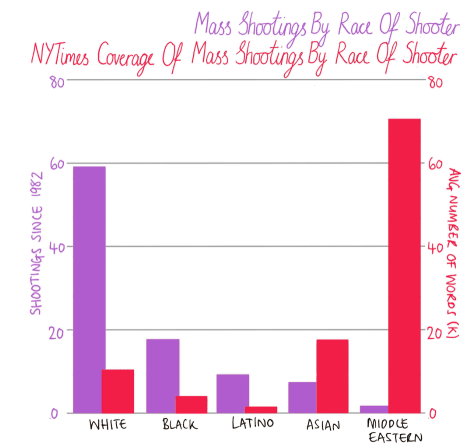
1 mark:
multiple
items

Constraints & Channel Availability

- consider marks and channels as imposing constraints
 - when does mark type constrain channel use?
 - **when does using one channel constrain another channel?**
- Channel Availability Model
 - Encoded: which channels directly used to encode attributes?
 - clear meaning
 - multiple channels can be directly used for redundant encoding
 - Free: which channels free to encode another attribute?
 - without changing usability of existing encoding
 - Unavailable: which channels unavailable / precluded / taken?
 - because of mark type?
 - because of idiom/algorithm design specifics?
 - **because other channels used?**

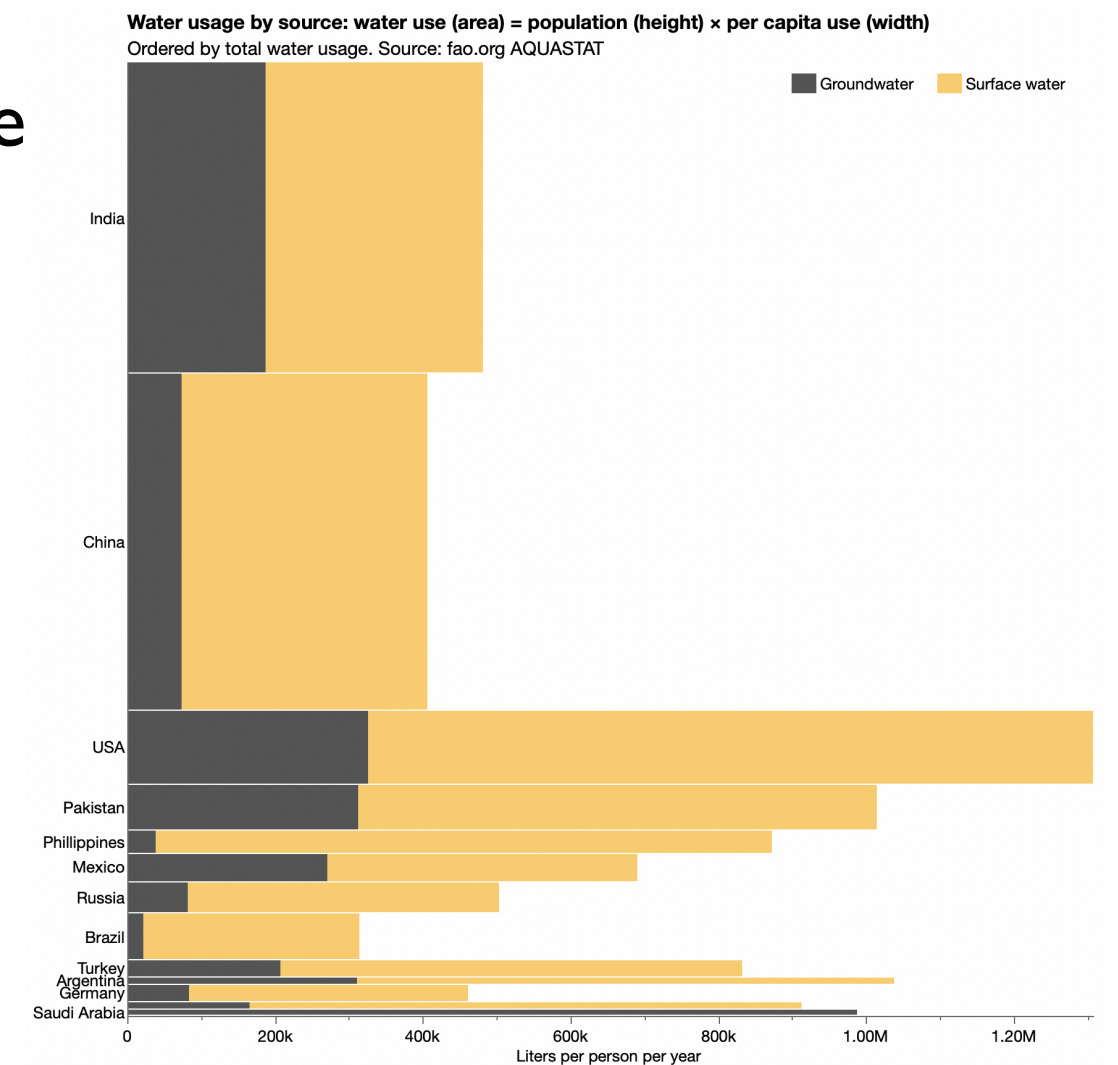
General dependencies: Position

- need fine-grained ability to specify for adequate descriptive power
 - rectilinear (horizontal and/or vertical)
 - high precision because perceptually aligned
 - depth (3D position): very low precision
 - radial (angular position and/or radial distance)
 - lower precision, no perceptual alignment
- general dependencies for unavailability?
 - cannot use both rectilinear and radial simultaneously
 - in same layer, using one type precludes other
 - but horizontal doesn't preclude vertical & vice versa
- position is shared / global
 - with respect to specific coordinate frame shared across many marks



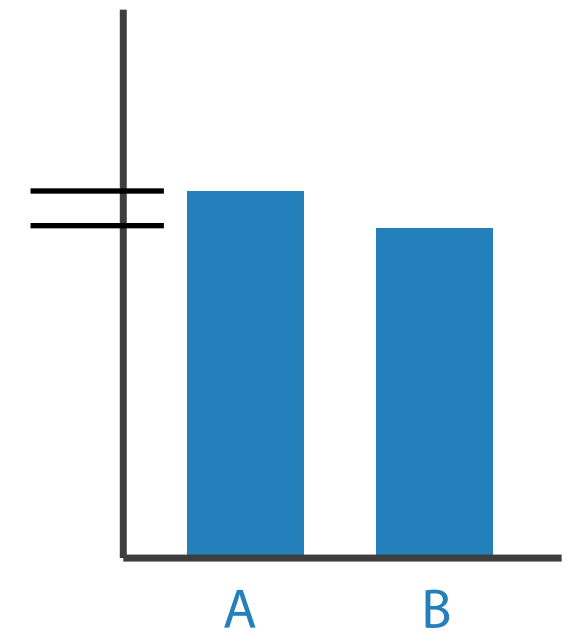
General dependencies: Size

- 1D (length) \ll 2D (area) \ll 3D (volume)
- dependencies for unavailability?
 - larger dimension subsumes smaller ones
 - encode with area channel means length channel unavailable
 - volume means area & length unavailable
 - but not vice versa: can augment from length to area
 - add second attribute for 1D size coding in other direction
- size is local, in contrast to shared position
 - can be independent across marks

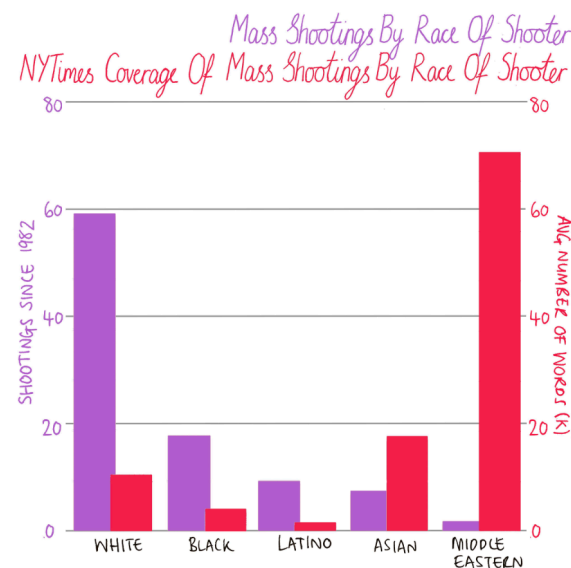


General dependencies: Position vs length

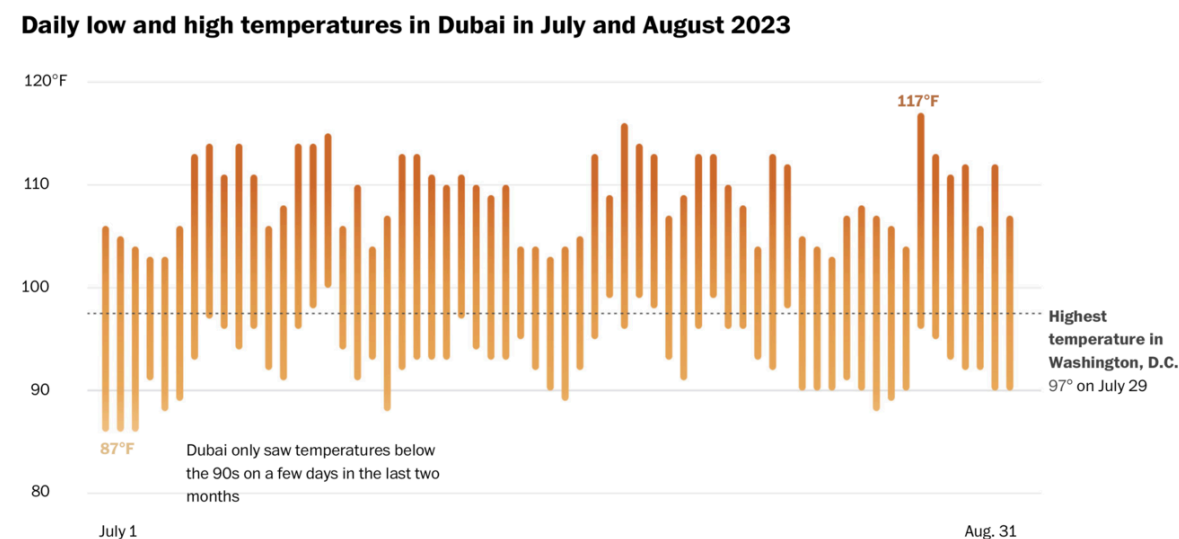
- alignment
 - position (horizontal and/or vertical) is usually shortcut for "aligned position", highest precision channel of all
 - reference frame of explicit axis
 - implicit boundaries of view / window / region
- general dependencies: position (ID) vs length (ID size)?
 - for line marks, position encoded implies length encoded
 - but not vice versa: can have length without position



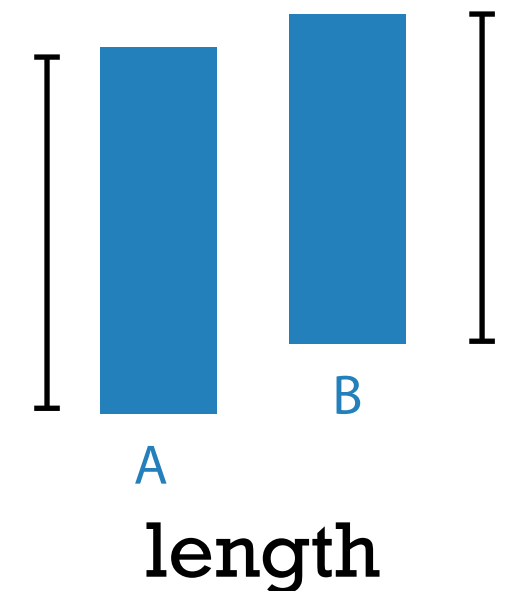
position along aligned scale



<https://twitter.com/MonaChalabi/status/1158779046693679106>

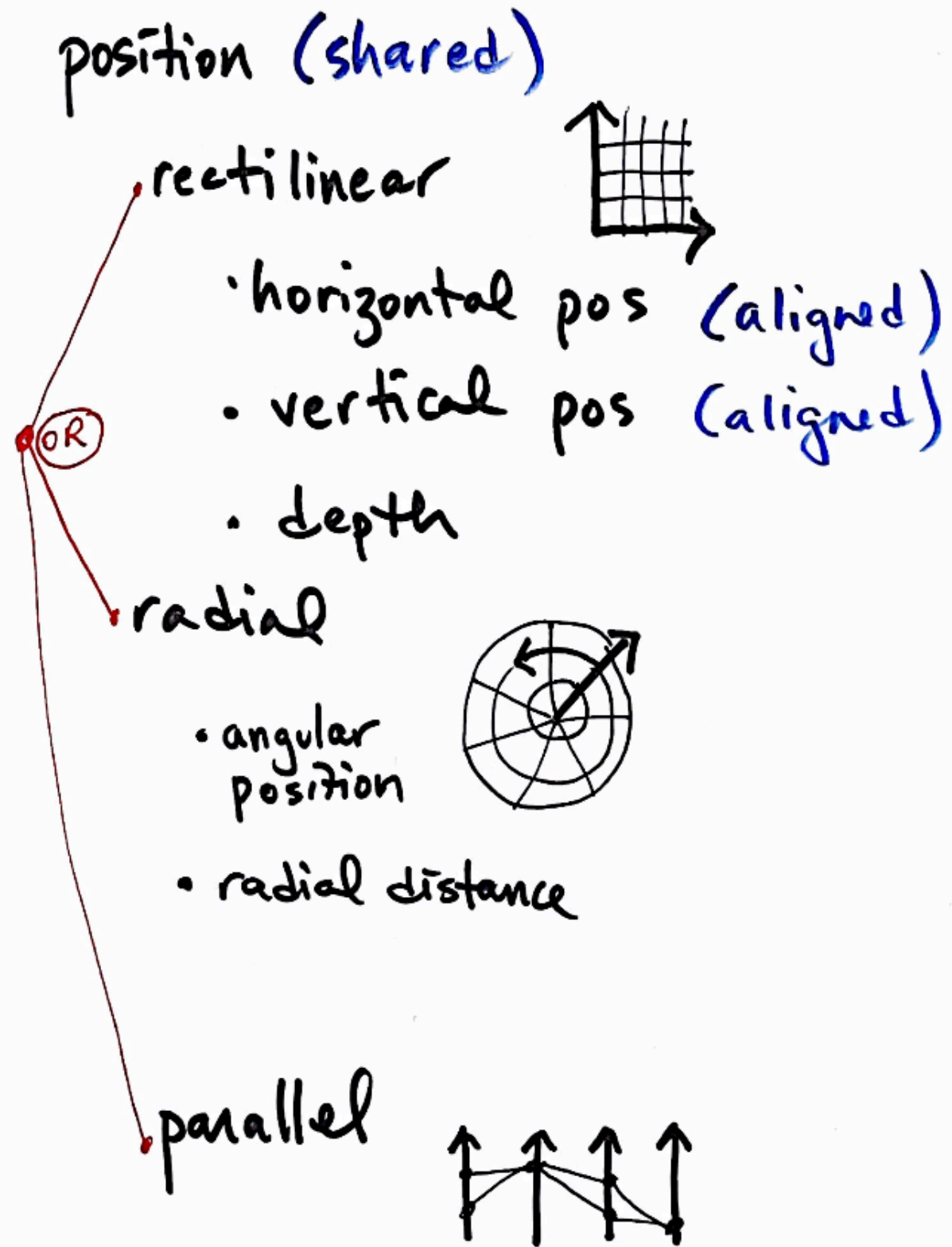


<https://www.washingtonpost.com/world/2023/09/10/dubai-heat-staying-cool>






length

Channel Dependencies/Hierarchy proto-model musings



• size (local)

- 1D length 
 - 2D area 
 - 3D volume 
- uses

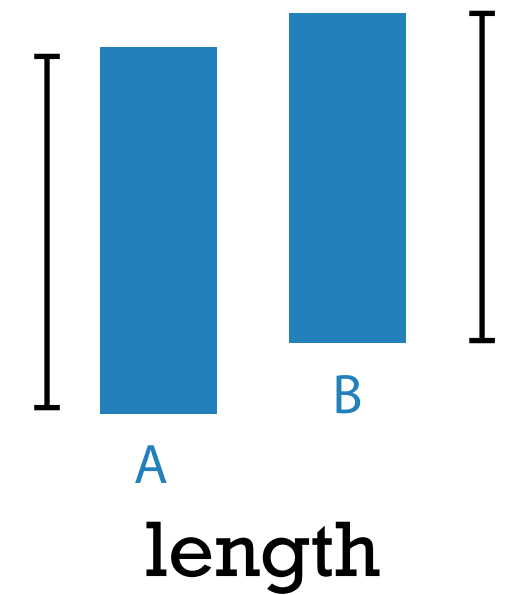
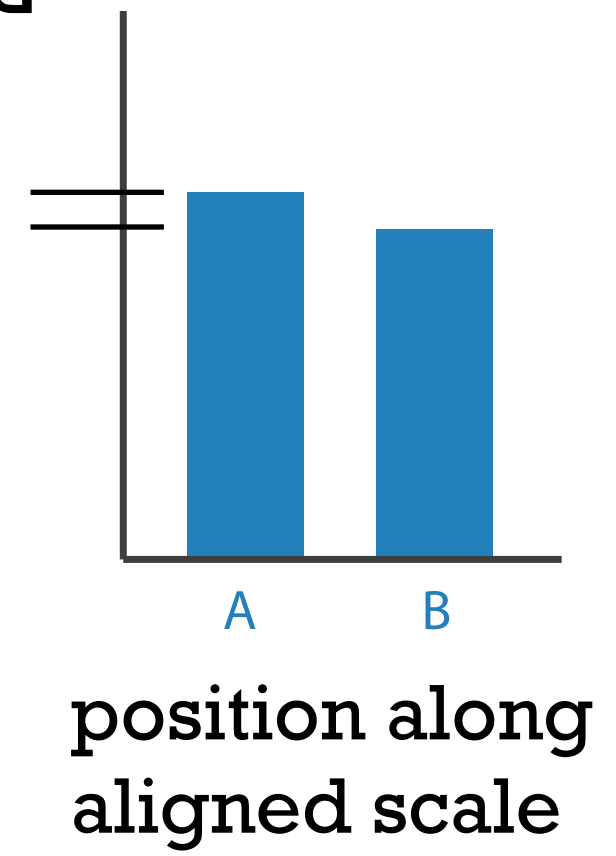
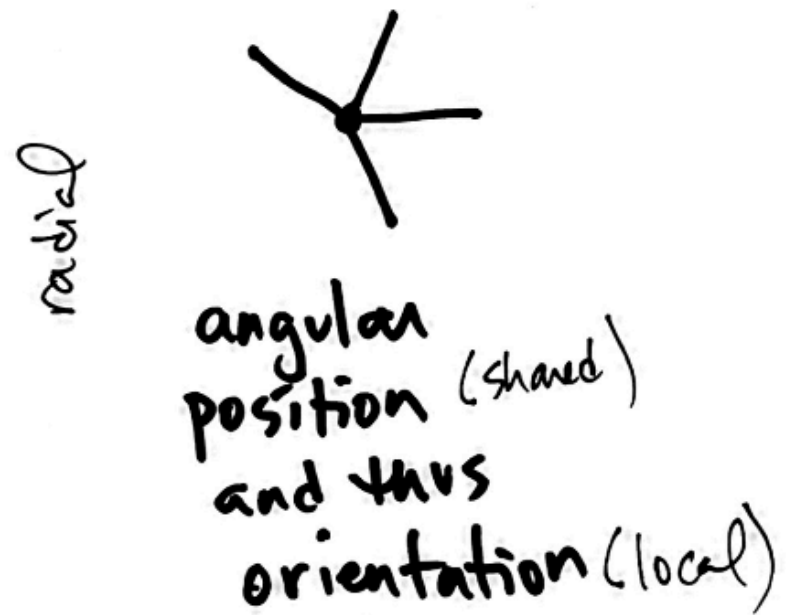
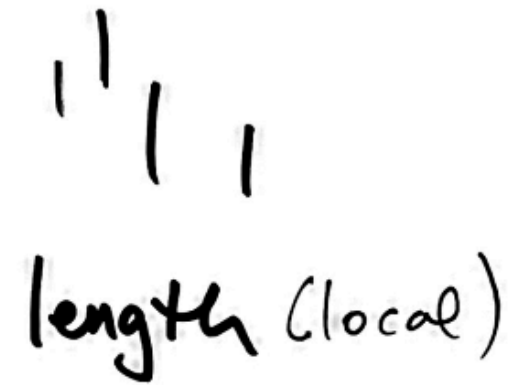
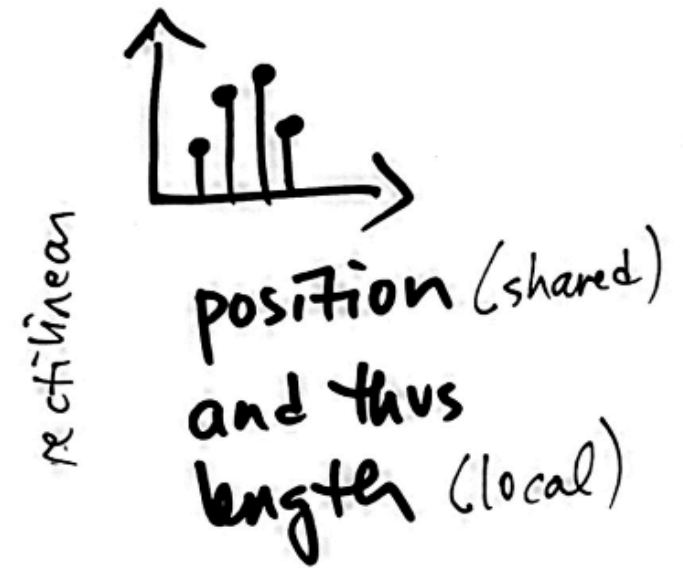
• orientation (local)

- color
- hue
 - saturation
 - luminance

- shape
- motion

Shared vs local example

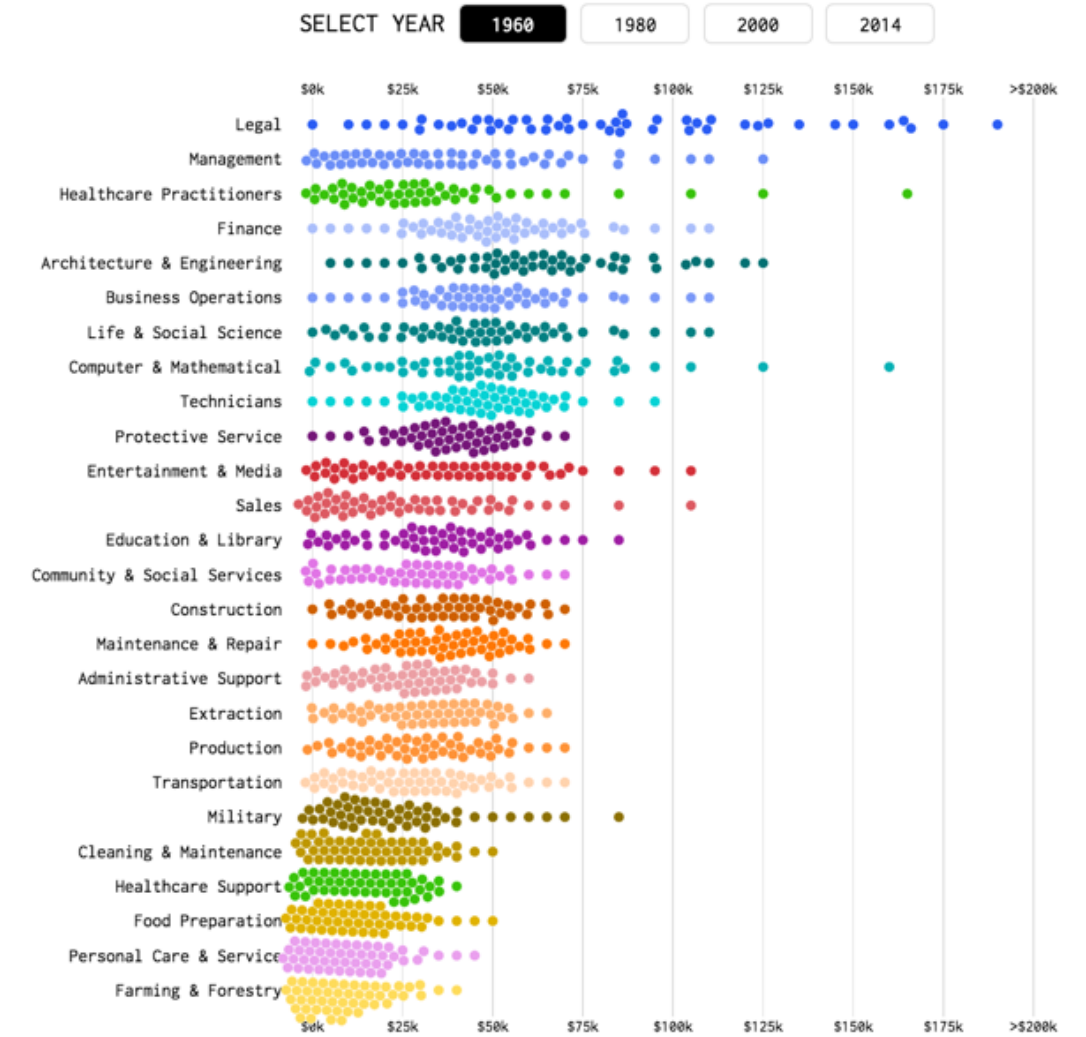
- shared locks down local, but can use local without shared



Multi-level analysis required for many cases

– small multiples: juxtaposed views

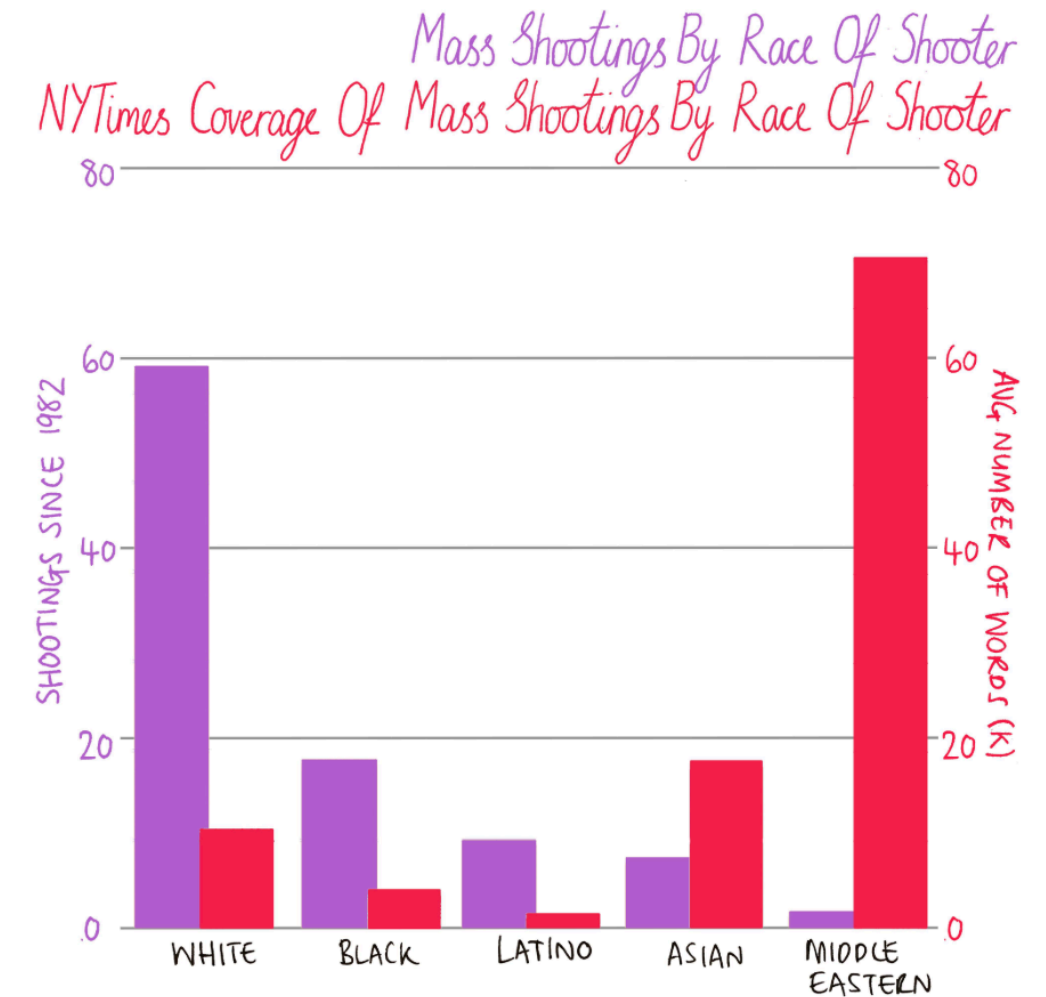
- vertical position within row: algorithmic, avoid occlusion
- vertical position across rows: encodes job type attribute



<https://flowingdata.com/2016/06/28/distributions-of-annual-income/>

Multi-level analysis needed: Grouped bar charts

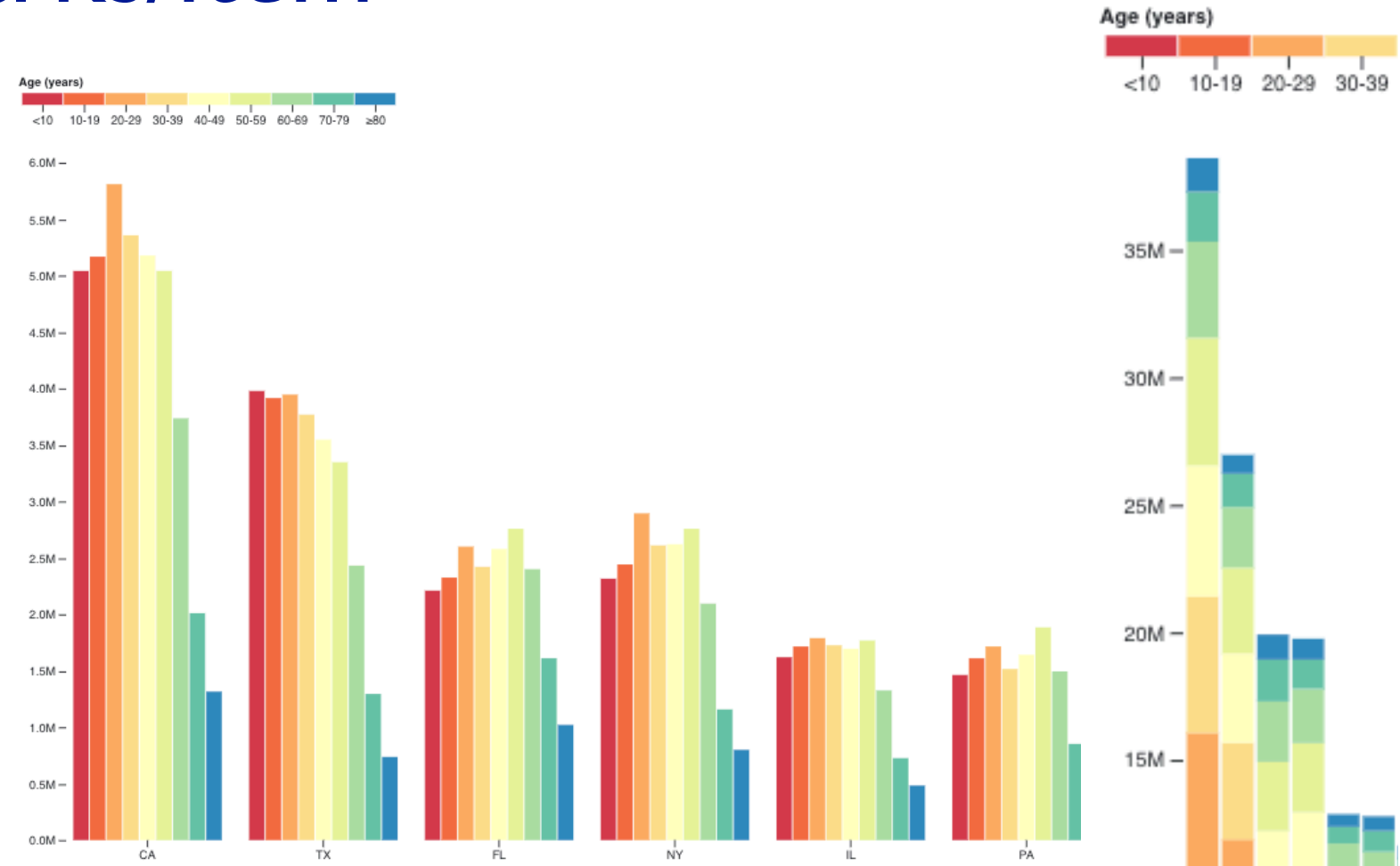
- Encoded
 - vertical position encodes quantitative attributes
 - shootings & coverage counts
 - length (ID size) redundantly encodes same thing
 - color encodes categorical attrib (shooting vs coverage)
 - horizontal position
 - low-level (within group) encodes counts, redundant w/ color
 - high-level (across groups) encodes race (shooting & coverage)
- Unavailable
 - any other position channel (radial) precluded
- Free
 - motion, shape, ...



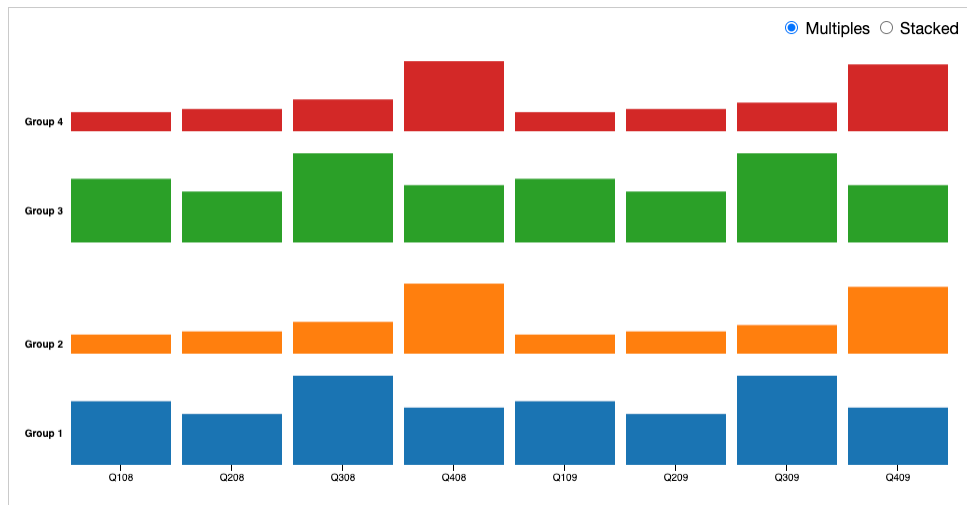
<https://twitter.com/MonaChalabi/status/1158779046693679106?s=20>

From marks to glyphs: multiple marks/item

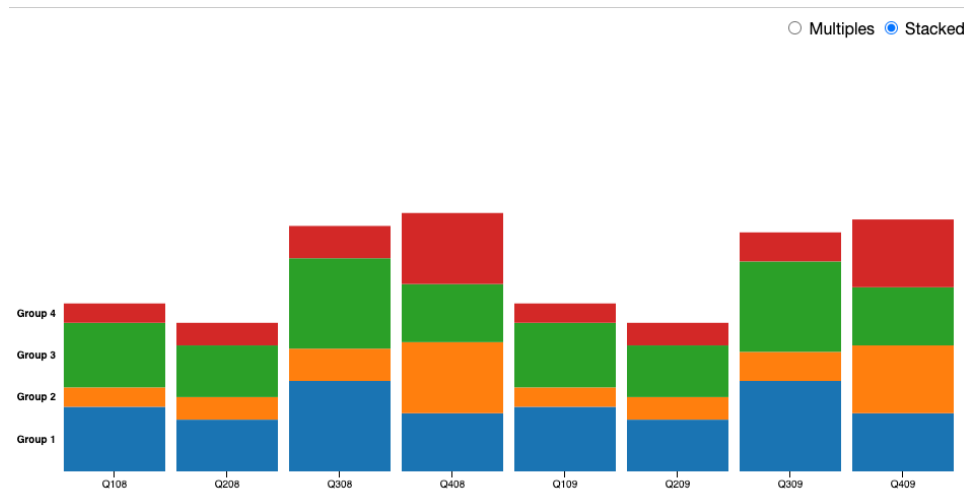
- glyphs: more than one mark per item
 - grouped bars
 - stacked bars
- multiple views
 - bar chart small multiples



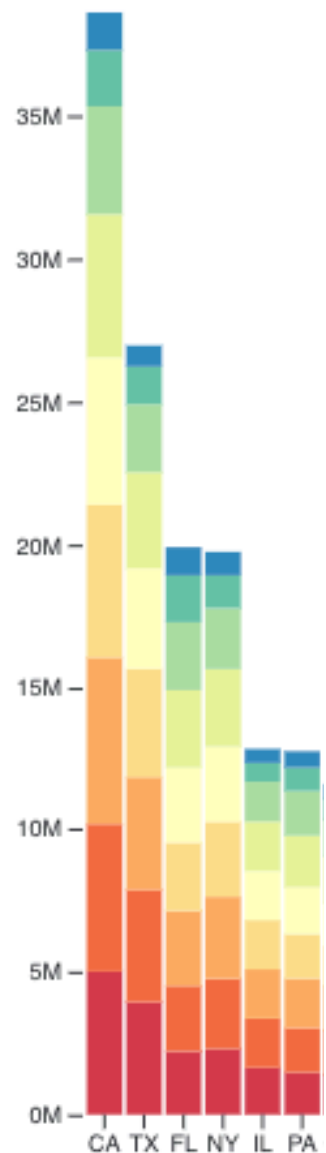
<https://observablehq.com/@d3/grouped-bar-chart/2>



<https://blocks.roadtolarissa.com/mbostock/4679202>



<https://observablehq.com/@d3/stacked-bar-chart/2>

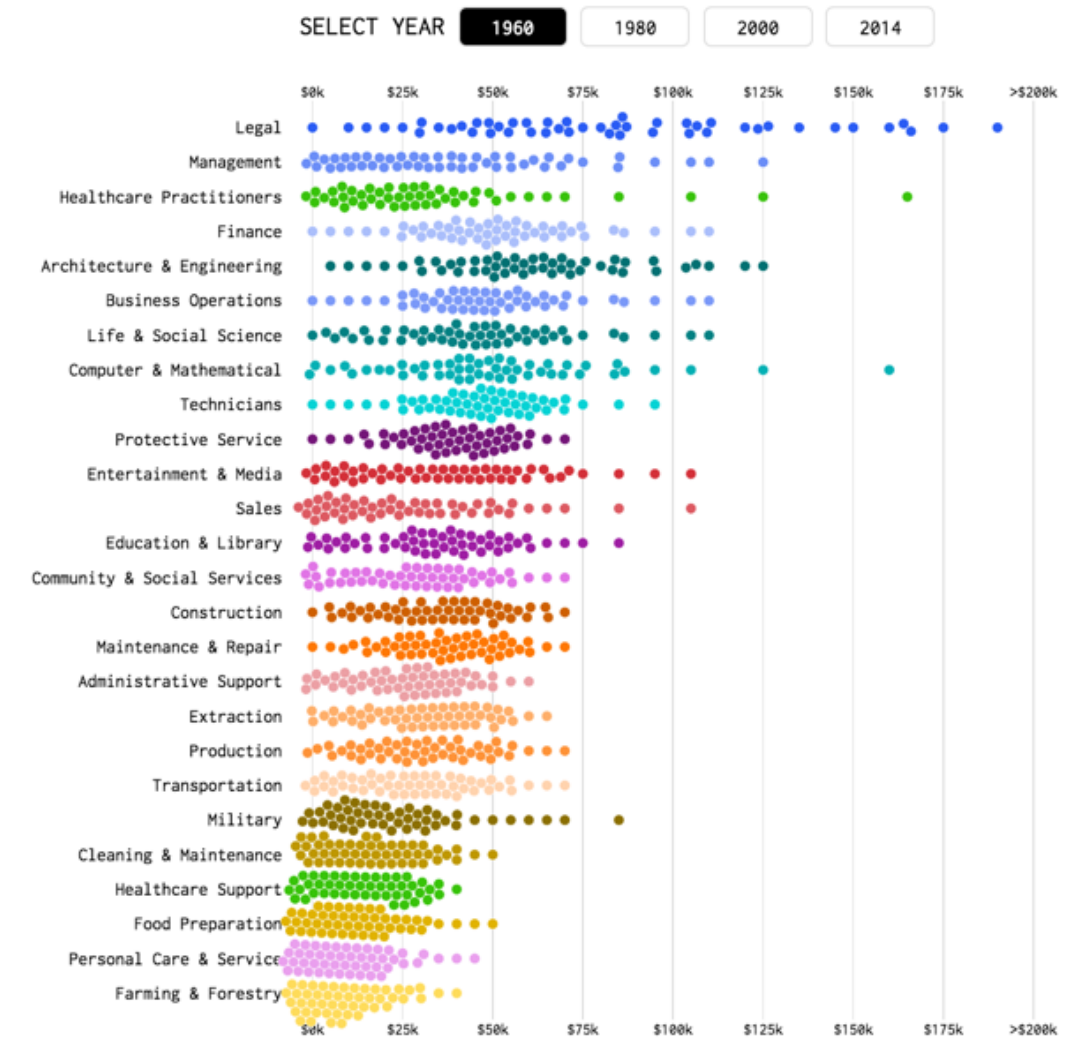


Multi-level analysis required for many cases

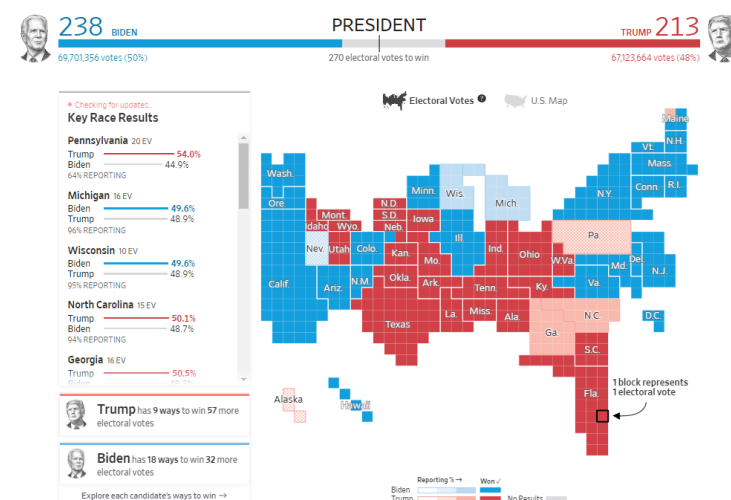
– small multiples: juxtaposed views

- vertical position within row: algorithmic, avoid occlusion
- vertical position across rows: encodes job type attribute

– nesting:
multi-scale
views / glyphs



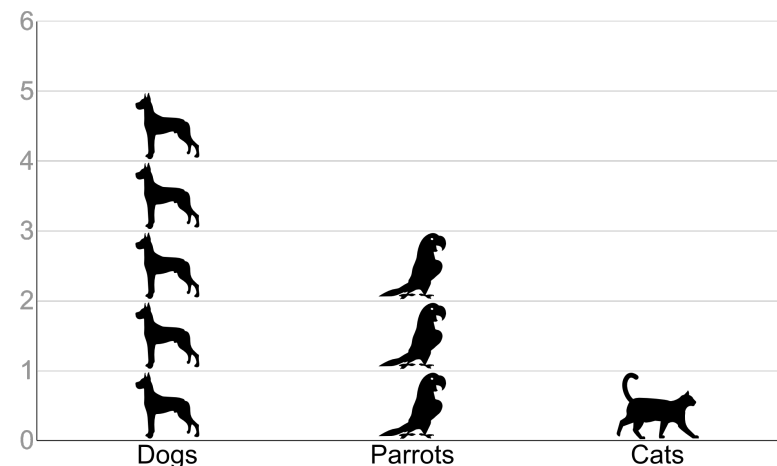
<https://flowingdata.com/2016/06/28/distributions-of-annual-income/>



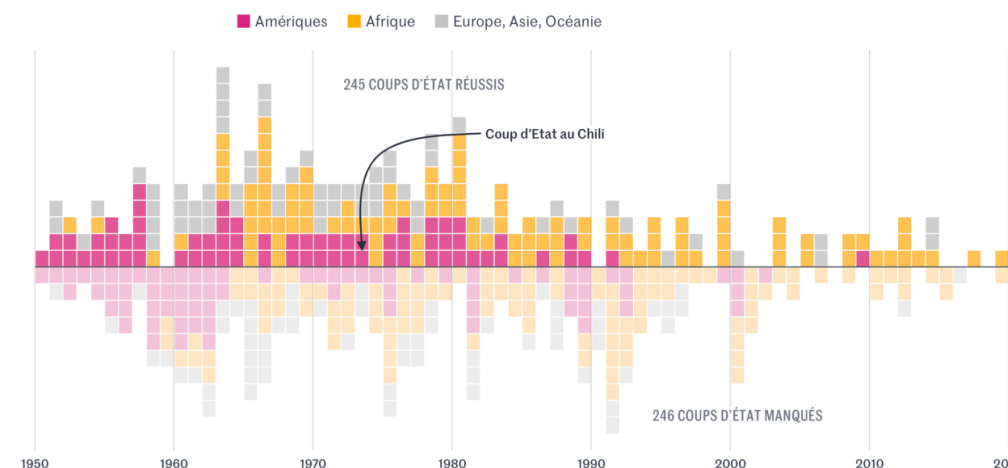
<https://www.anychart.com/blog/2020/11/06/election-maps-us-vote-live-results/>

Unit encodings

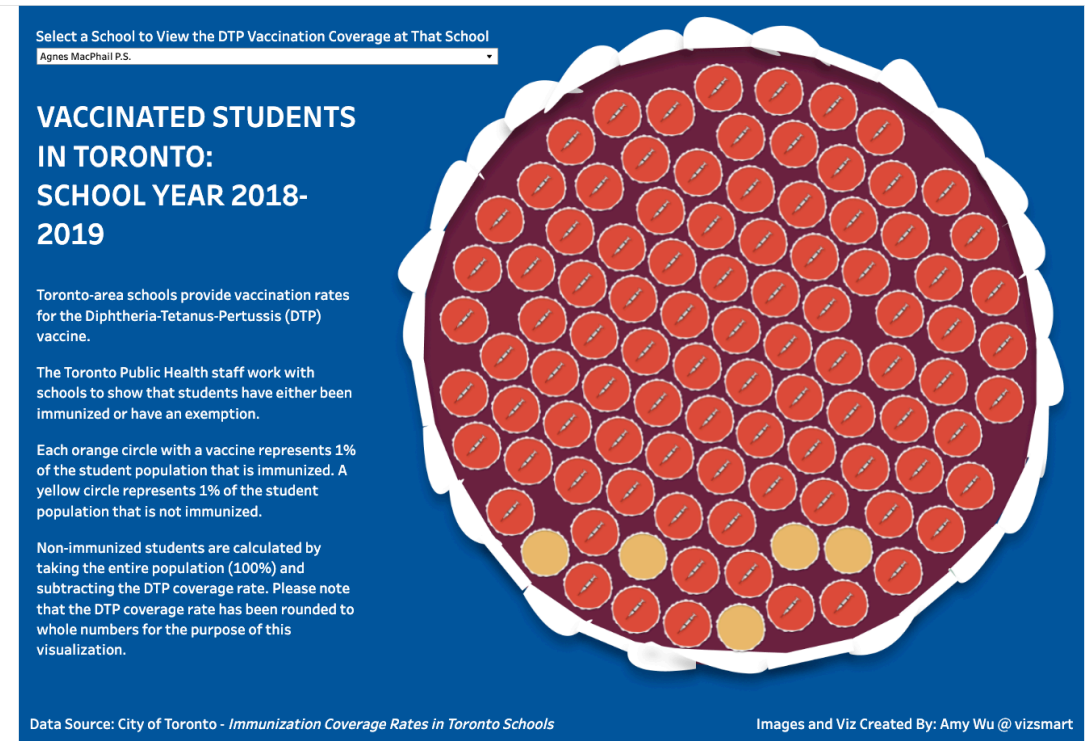
- point marks
 - general case: quantity only!
 - position channel not necessarily in use
- often constrained by idiom
 - then need **multi-level interpretation**
 - top level: interlocking mark
 - rectilinear: support counting width & height separately
 - bottom level: unit point marks
 - can be independently color-coded (or interactively highlighted)



<http://steveharoz.com/research/isotype/>



https://www.lemonde.fr/les-decodeurs/article/2023/09/11/depuis-1950-pres-de-cinq-cents-coups-d-etat-tentes-ou-reussis-surtout-en-amerique-latine-et-en-afrique_6188906_4355770.html



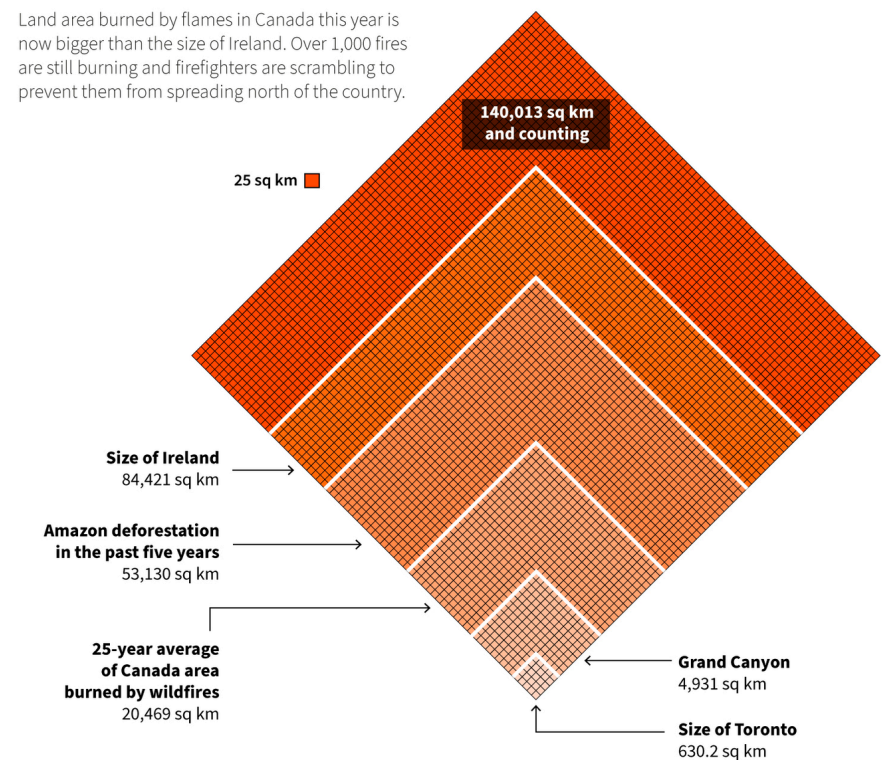
Data Source: City of Toronto - Immunization Coverage Rates in Toronto Schools
Images and Viz Created By: Amy Wu @vizsmart

<https://www.informationisbeautifulawards.com/showcase/6163-vaccinated-students-in-toronto-school-year-2018-2019>

https://public.tableau.com/app/profile/amy.r.wu/viz/DTP_16715915193810/DTPDashboard

Wildfires in Canada burned this much land so far this year

Land area burned by flames in Canada this year is now bigger than the size of Ireland. Over 1,000 fires are still burning and firefighters are scrambling to prevent them from spreading north of the country.



Note: As of Aug. 20, 4 p.m.
Source: Canadian Interagency Forest Fire Centre
Prinz Magtulis | Reuters, Aug. 21, 2023

<https://twitter.com/prinzmagtulis/status/1693684342818574675>

Conclusion: Preliminary steps towards answers?

- Old marks/channels models
 - marks, based on Bertin's geometry: 0D points, 1D lines, 2D areas, 3D volumes
 - channel rankings by accuracy, based on Cleveland & McGill two-value ratio task
- Alternative marks/channels proto-models
 - channel-based constraint analysis, channel availability: Encoded, Free, Unavailable
 - mark types model, mark-based constraints
 - Unconstrained / Singleton, for single items (points & simple line segments)
 - Paths, for multiple items (complex lines)
 - Filled & Interlocking, with boundary & interiors (areas)
 - channel dependency proto-model
 - distinguish coordinate frame positions as shared, vs size & orientation as local
 - multi-level mark type analysis
- do these help think and reason about design space of visual encodings?

More stuff

- this talk

<http://www.cs.ubc.ca/~tmm/talks.html#northeastern24>

–more questions? thoughts on answers??

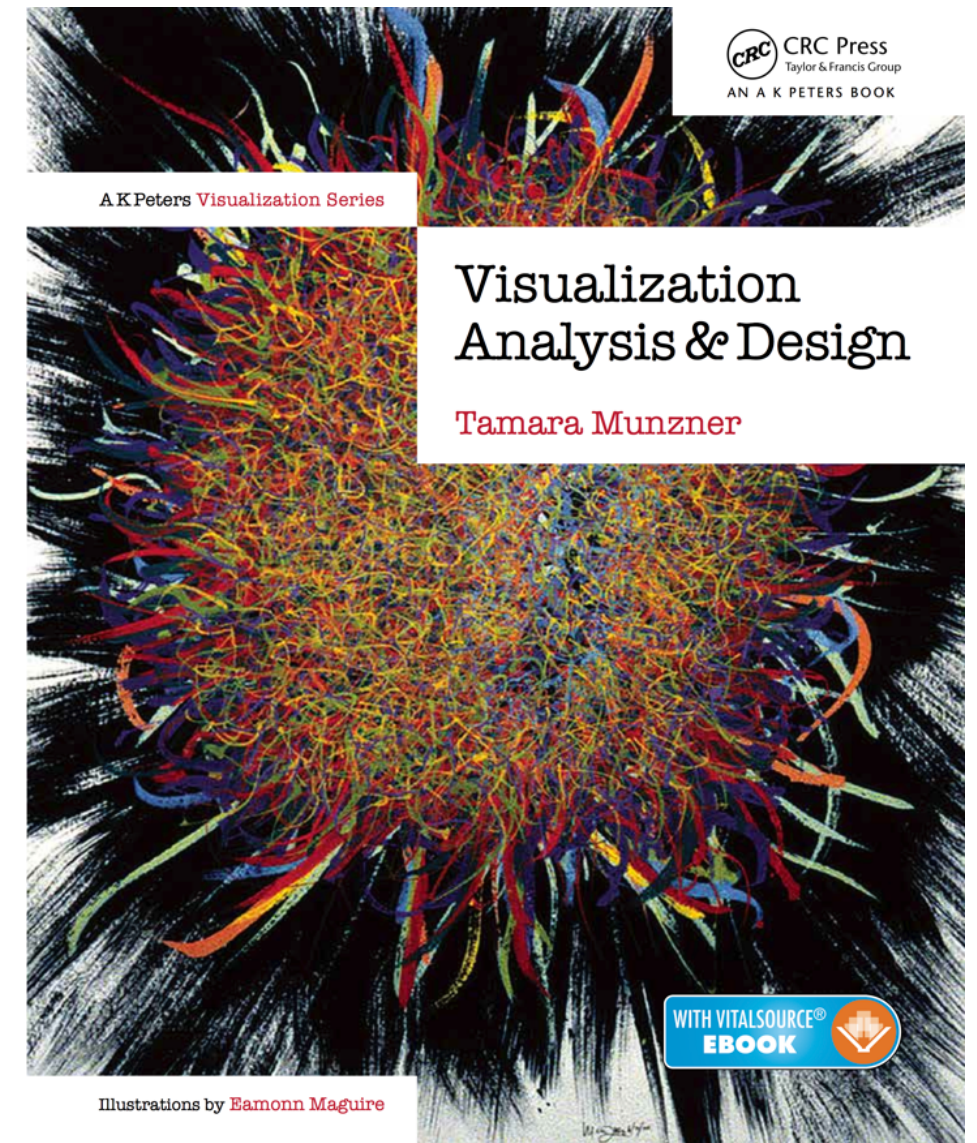
- book

<http://www.cs.ubc.ca/~tmm/vadbook>

- full courses, papers, videos, software, talks

<http://www.cs.ubc.ca/group/infovis>

<http://www.cs.ubc.ca/~tmm>



Visualization Analysis and Design. Munzner.
CRC Press, AK Peters Visualization Series, 2014.

 [@tamara@vis.social](https://medium.com/@tamara@vis.social)

 [@tamaramunzner](https://twitter.com/tamaramunzner)