Marks Revisited: Beyond Bertin

Tamara Munzner
Department of Computer Science
University of British Columbia

MIT HCI Seminar
13 Feb 2024

http://www.cs.ubc.ca/~tmm/talks.html#mit24

@tamara@vis.social

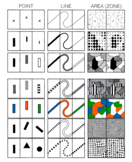
@tamaramunzner

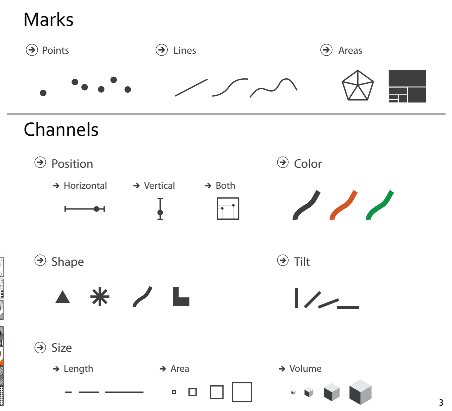
Co-conspirators

- Richard Brath, Uncharted Software
- UBC infovis course TAs
  - Mara Solen, Francis Nguyen
  - Steve Kasica, Ryan Smith

Marks and channels: Foundational model

- decompose visual encoding into marks & channels
  - marks
    - geometric primitives
    - represent items
  - channels
    - control appearance of marks
    - representing attributes
- widely used
  - Bertin 1967
    - Semiology of Graphics



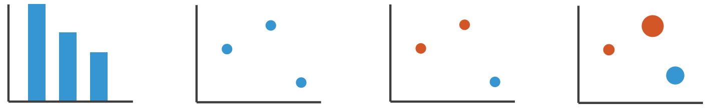


Talk outline

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

Visual encoding

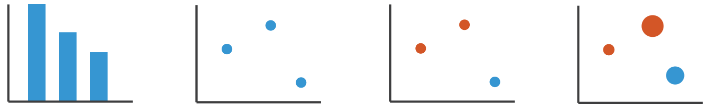
- analyze idiom structure as combination of marks and channels



Visual encoding

- analyze idiom structure as combination of marks and channels

idiom: bar chart



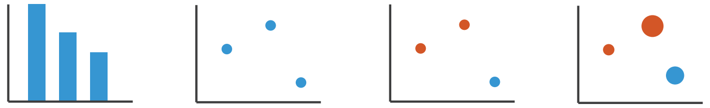
1 channel:
vertical position

mark: line

Visual encoding

- analyze idiom structure as combination of marks and channels

idiom: bar chart | idiom: scatterplot



1 channel:
vertical position

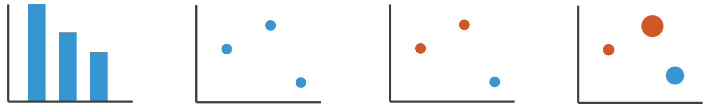
2 channels:
vertical position
horizontal position

mark: line | mark: point

Visual encoding

- analyze idiom structure as combination of marks and channels

idiom: bar chart | idiom: scatterplot



1 channel:
vertical position

2 channels:
vertical position
horizontal position

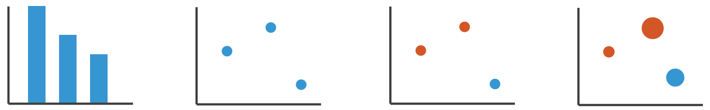
3 channels:
vertical position
horizontal position
color hue

mark: line | mark: point | mark: point

Visual encoding

- analyze idiom structure as combination of marks and channels

idiom: bar chart | idiom: scatterplot



1 channel:
vertical position

2 channels:
vertical position
horizontal position

3 channels:
vertical position
horizontal position
color hue

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

mark: line | mark: point | mark: point | mark: point

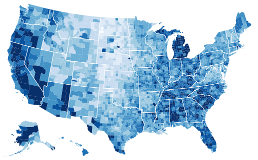
Visual encoding: Spatial data

- marks for given spatial data (boundaries)

idiom: choropleth map

channels:
position
color (saturation)

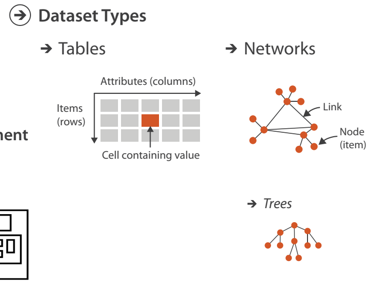
mark: area

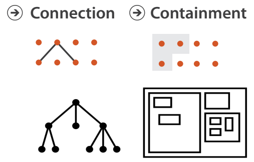


http://bl.ocks.org/mbostock/4060606

Visual encoding: Network data

- marks for items vs marks for links

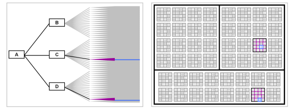




idiom: node-link diagram

channel:
position

position



Node–Link Diagram **Treemap**

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

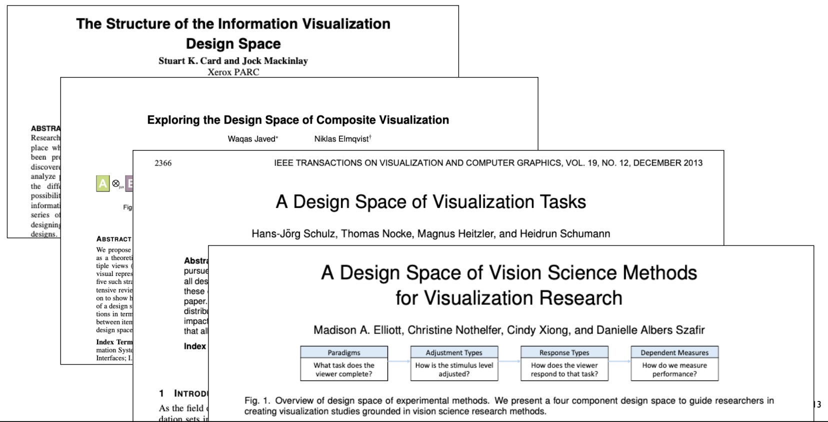
marks:
area for items/nodes,
containment area for parent-child links

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

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...
  - 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
      - *[Ralph. Toward Methodological Guidelines for Process Theories & Taxonomies in Software Engineering. IEEE TSE 2020]*

Design spaces in visualization: continuing theme

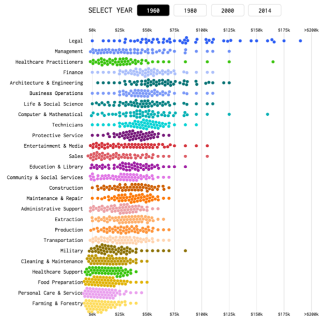


Teaching design space: analyze visual encoding & map to data

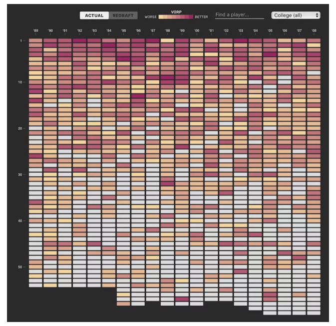
- analyze existing encoding with marks & channels
  - Visual channels used?
    - Channel X encodes attribute Y
    - Channel X encodes attribute Y
  - Marks used?
    - Mark of type X encodes item Y
    - Mark of type X encodes item Y

Teaching: Bertini in-class exercises, catalyst for questions

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



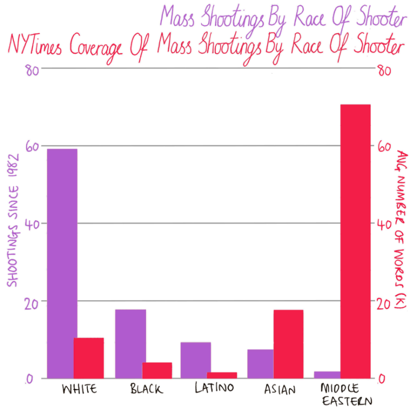
https://flowingdata.com/2016/06/28/distributions-of-annual-income/



https://pudding.cool/2017/03/redraft/

Quiz: Name marks/channels

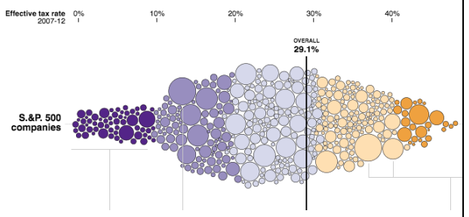
- Shooting Media Coverage
- marks
  - A: points
  - B: lines
  - C: areas
- channels
  - A: position
  - B: color
  - C: motion
  - D: area
  - E: angle



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: motion
  - D: area
  - E: angle

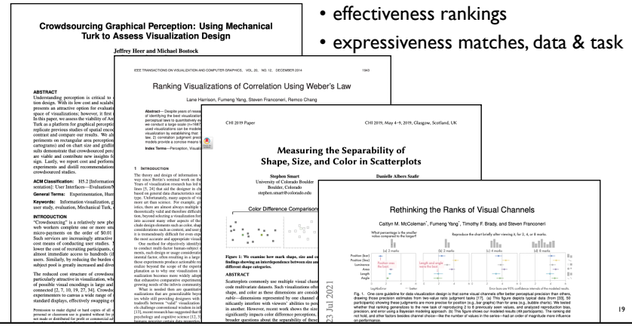


https://archive.nytimes.com/www.nytimes.com/interactive/2013/05/25/sunday-review/corporate-taxes.html

Many, many questions

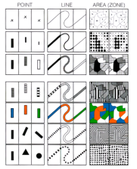
- so what?
  - evidence that this design space could be improved!

Channels: Model evolves, heavily studied

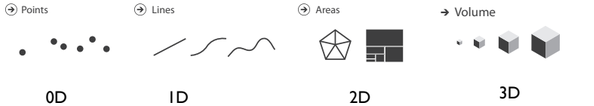


- effectiveness rankings
- expressiveness matches, data & task

Marks: Model stays static

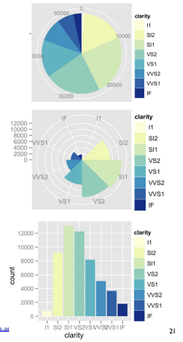


- model inherited from Bertin (Semiology of Graphics, 1967)
  - never questioned
- geometric motivation
  - geometric primitives have dimensions
  - how could we argue with math?!



What do marks do?

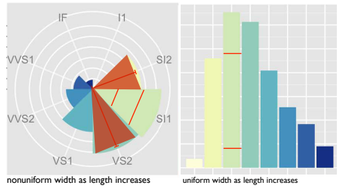
- idiom: pie chart
  - **area** marks with angle channel: **2D area varies**
    - separated & ordered radially, uniform length
  - accuracy: area less accurate than rectilinear aligned line length
  - task: part-to-whole judgements
- idiom: coxcomb chart
  - marks with length channel: **1D length varies**
    - separated & ordered radially, uniform width
  - direct analog to radial bar charts
  - what's the mark type?
    - line, because it's length coded?
    - area, because area varies too?



[A layered grammar of graphics. Wickham, Journ. Computational and Graphical Statistics 19.1 (2010): 3-28

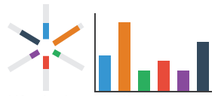
Coxcomb channel analysis

- encode: 1D 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
- mark type: encode or decode?
  - infer designer intent
  - predict viewer response: if channels differ, which "wins"?



nonuniform width as length increases

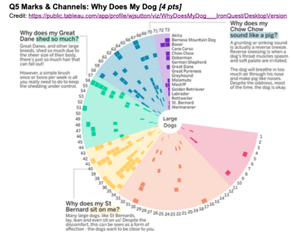
uniform width as length increases



radial & rectilinear bars: uniform width as length increases

Channels used: what does it mean?

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



https://public.tableau.com/app/profile/wjnutton/viz/WhyDoesMyDog___IronQuest/DesktopVersion

Channel availability

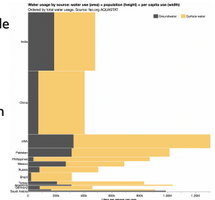
- channels as constraints
  - when does using one channel constrain another channel?
- Channel Availability Model
  - Encoded: which channels directly used to encode attributes?
    - clear meaning
    - multiple channels can be directly used for redundant encoding
  - Unavailable: which channels unavailable / precluded / taken?
    - general dependencies between channels
    - specifics of idiom/algorithm design
  - Free: which channels free to encode another attribute?
    - without changing usability of existing encoding

General dependencies: Position

- need fine-grained ability to specify for adequate descriptive power
  - rectilinear (horizontal and/or vertical)
    - high precision
      - depth (3D position): very low precision
  - radial (angular position and/or radial distance)
    - lower precision
- 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

General dependencies: Size

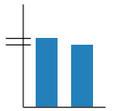
- 1D (length) << 2D (area) << 3D (volume)
- dependencies for unavailability?
  - larger dimension subsumes smaller ones
    - encode with area channel means length channel unavailable
    - volume means area & length unavailable
  - but not vice versa: can augment from length to area
    - add second attribute for 1D size coding in other direction



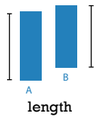
https://www.barchigon.com/blog/2019/12/18/bar-mekiko-chart-study/

General dependencies: Position vs length

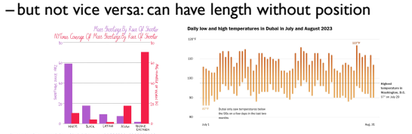
- alignment
  - position (horizontal and/or vertical) is usually shortcut for "aligned position", highest precision channel of all
    - reference frame of explicit axis
    - implicit boundaries of view / window / region
- general dependencies: position (1D) vs length (1D size)?
  - for line marks, position encoded implies length encoded
  - but not vice versa: can have length without position



position along aligned scale



length

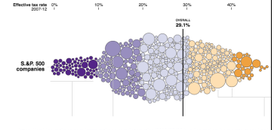


https://twitter.com/MonaChalabi/status/1158779046893879067?s=20

https://www.washingtonpost.com/world/2023/08/10/dubai-heat-anything-else/

Channel availability analysis: Circle packings

- customized circle packing
  - Encoded channels
    - horizontal position: encodes tax rate
    - color: rate, redundant with horizontal position
    - size: market cap
  - Unavailable channels
    - vertical position: used by algorithm to avoid overlap
    - radial positions (angular, distance): precluded by horizontal position channel use
  - Free channels
    - motion
  - unclear
    - shape? orientation?



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

Channel availability analysis: Circle packings

- customized circle packings are special case
  - beeswarm plots

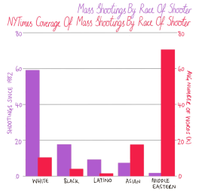
https://flowingdata.com/2014/06/18/distributions-of-annual-income/

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

- general circle packing
  - algorithmic constraint: no overlaps, minimal gaps
    - Unavailable: position used by algorithm

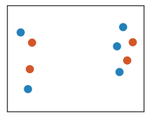
Channel availability analysis: Grouped bar charts

- Encoded
  - vertical position encodes quantitative attributes
    - shootings & coverage counts
  - length (1D size) redundantly encodes same thing
  - color encodes categorical attrib (shooting vs coverage)
  - horizontal position
    - low-level (within group) encodes same as color
    - high-level (across groups) encodes race (shooting & coverage)
- Unavailable
  - any other position channel (radial) precluded
- Free
  - motion, shape, ...



https://twitter.com/MonaChalabi/status/1158779046893879067?s=20

Separability vs integrality: Existing framework

Position + Hue (Color)	Size + Hue (Color)	Width + Height	Red + Green
			
Fully separable	Some interference	Some/significant interference	Major interference
2 groups each	2 groups each	3 groups total: integral area	4 groups total: integral hue

Channel availability model augments, not replaces

Channel availability model

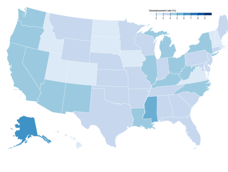
- Channel availability
  - Encoded
    - clear meaning
  - Unavailable: which channels unavailable / taken?
    - general dependencies
    - idiom/algorithm design
  - Free: which channels free to encode another attribute?
    - without changing usability of existing encoding
- how does this idea relate to mark types?

Area marks: Rethinking

- area marks is a terrible name
  - other marks all have graphical area too
    - allowing us to encode with color
  - there's also an "area" channel, which is confusingly different
  - area is not the only channel in play with these marks!

Area marks

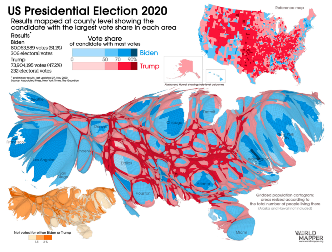
- obvious example: choropleth maps
- what can we do to California? could we encode additional data?
  - cannot shrink/grow (size channel)
  - cannot translate (position channel)
  - cannot rotate (orientation channel)
  - cannot reshape (shape channel)
  - why not?
    - would lose meaning of that mark: 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

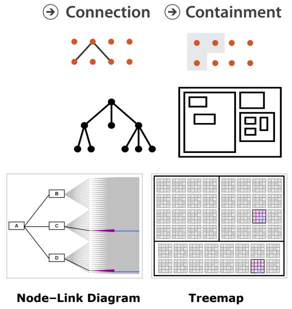
- many channels locked down with interlocking marks
  - boundary encodes meaning
  - cannot change size, shape, position, orientation
  - mark type as a constraint
- but... what about cartograms?
  - cannot change just one mark (California)
  - but could change them all!
- interlocking marks as global 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

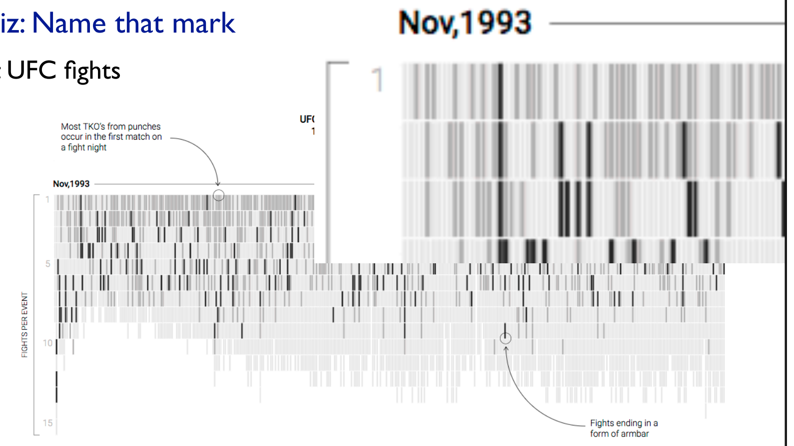
- example with non-spatial data?
- treemaps
  - show hierarchy with containment, not connection
  - encode additional attribute with area/size
- again, cannot change just one mark alone
  - but could recompute layout to change all at once
- combined layout of all marks together carries meaning
  - unlike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



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

Quiz: Name that mark

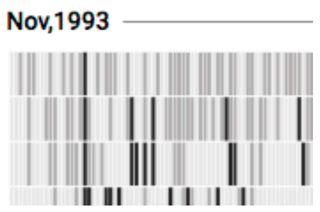
- C: UFC fights



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

Analyzing marks

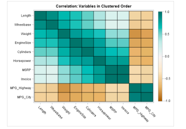
- what type of mark?
  - line?
    - no, not length coded
  - point mark with rectangular shape?
    - 2020: yes!
    - 2023: no!
      - cannot change position / size / orientation
  - area?
    - 2020: no, area/shape does not convey meaning
    - 2023: yes!
      - fully interlocking
      - position,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

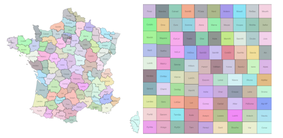


https://multimedia.scmp.com/infographics/sport/article/3010883/bruce-lee-and-mixed-martial-arts/

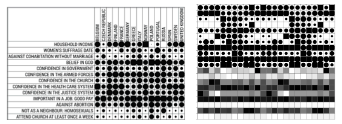
Interlocking marks: Tile heatmaps

- 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





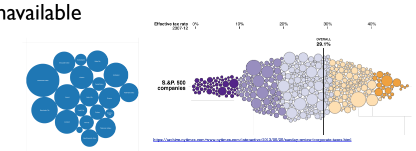
Spatially ordered treemaps. Wood and Dykes. IEEE TVCG (Proc. InfoVis) 14(6):1348-1355, 2008.



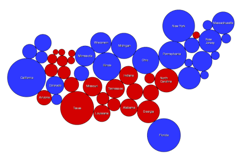
Bertifier Charles Perin, Pierre Dragicevic, Jean-Daniel Fekete (2014). Revisiting Bertin's Matrices: New Interactions for Crafting Tabular Visualizations. TVCG, VIS' 2014.

Interlocking marks: Circle packings

- interlocking marks
  - not size-coded point marks (with circle shape)
  - shape / position / orientation: equal & unavailable
  - more like treemap than scatterplot

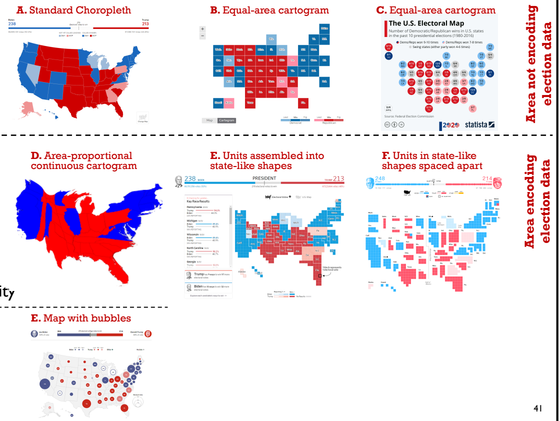


- Dorling cartogram
  - can treat as special case of circle packing, with additional constraints based
  - maintain relative positioning from geographic location
  - throw away shape by regularizing to circles
  - add size coding



Interlocking? Election maps roundup

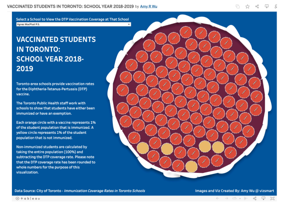
- yes interlocking
  - A: already covered
  - B/C: equal-area alg algorithm simplifies shape
- yes interlocking
  - E/F multi-level
  - top level: interlocking marks
  - bottom level: square units
    - E/F: countability for votes
    - F whitespace: population density
- no, point marks
  - size coded by area

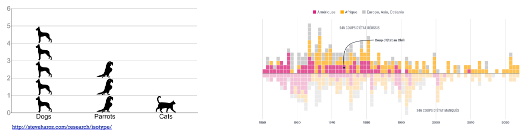


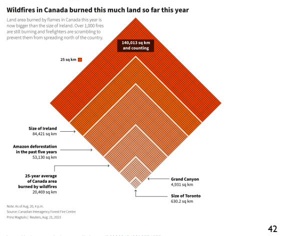
https://www.anychart.com/blog/2020/11/06/election-maps-us-vote-live-results/

Unit encodings

- point marks
  - general case: quantity only!
  - position channel not necessarily in use
- often constrained by idiom
  - 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)



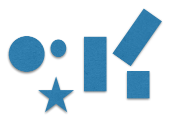




Distinguishing marks through constraints

- highly constrained: interlocking marks
  - many channels unavailable: size, position, shape, orientation
  - proposal: rename from "area" to "interlocking"

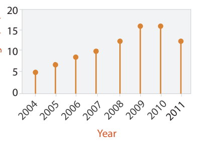
- unconstrained: point marks
  - can encode more info with any channel at all!
    - size, position, shape, orientation
    - color, motion, ...
  - does "point" imply circular shape?
    - proposal: is "unconstrained" a better / more evocative name?



- so... what about line marks?

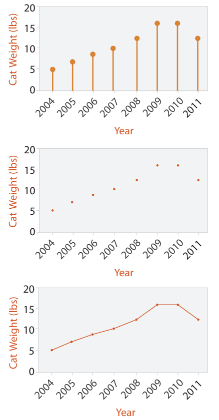
Line marks: Also confusing

- curved lines are not line marks!
  - model: one mark represents one item of data
  - line mark: express single quantitative attribute for one item with length
- also misnamed?
  - proposal: call them "segment" marks instead!



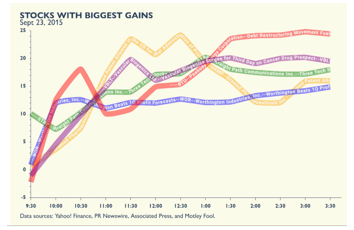
Line marks: Also confusing

- do line charts use line marks?
  - no
    - at least, not exactly like bar charts do
  - connection line segments between points
    - trend task: emphasize relationships between items
  - line chart encodes many items, not just one
    - with many piecewise-linear segments
    - or smoothed into curve



Line marks: Also confusing

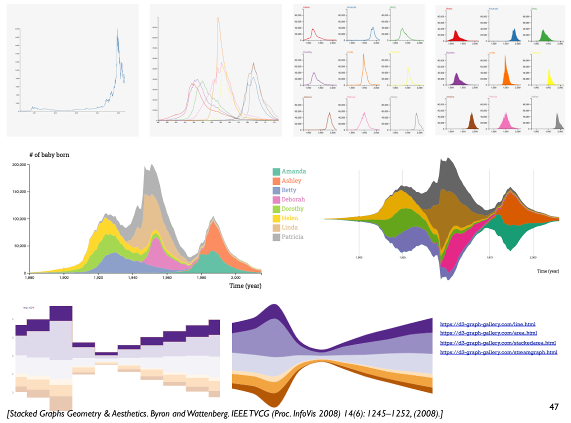
- when does region inside curved "line" boundary act like an area mark?
  - can encode information within its boundary
    - color, even text
  - only if it's wide enough?
  - or always!?
- what about the region below it?...



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

Line charts vs filled area charts

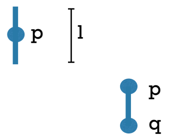
- should we reason differently about
  - line chart boundary vs filled area chart interior?
  - stacked area charts vs streamgraphs?
  - discrete stacked bar charts vs continuous streamgraphs?
- what matters?
  - boundary vs interior?
  - discrete vs continuous
  - occlusion?



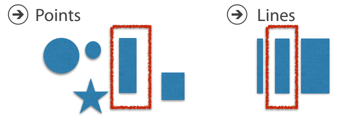
[Stacked Graphs Geometry & Aesthetics. Byron and Wattenberg. IEEE TVCG (Proc. InfoVis 2008) 14(6): 1245–1252, (2008).]

Line marks: what good are they?

- are line/segment marks an unnecessary construct?
  - how does segment differ from "length-coded point mark"?
  - is segment just two point marks in a trenchcoat???
    - two position-coded point marks, connected by segment



- radical proposal: eliminate line marks as a separate mark type!
  - merge lines and points together, into "unconstrained" marks



Many, many questions

- including
  - Size-coded point marks vs area marks?
  - Area marks vs area channel?
  - What kind of marks are in
    - a tile heatmap?
    - a circle packing?
    - a cartogram?
    - a multi-level thingie with units?
  - Line mark: line segment vs curved path?
    - Do line charts use line marks?
  - Line chart boundaries vs filled area charts interiors?
  - Length-coded point marks vs line marks?

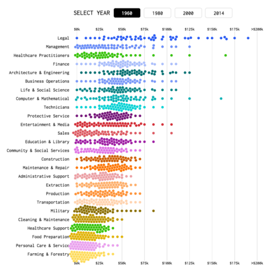
Alternative models: preliminary steps towards answers?

- Old mark/channel model: geometry-based marks
  - 0D points, 1D lines, 2D areas, 3D volumes
- Alternative mark/channel model
  - channel-based analysis: channel availability model
    - Encoded, Unavailable, Free
  - mark-based analysis: mark constraint model
    - Unconstrained (points), Interlocking (areas)

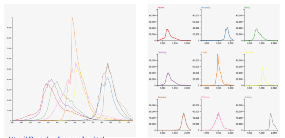
- what best helps us think and reason about design space of visual encoding?
  - combination of both? just one?
  - another alternative?
    - are there other interesting emergent properties arising from bottom-up channel analysis?

Beyond marks and channels

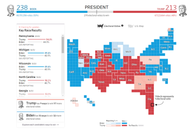
- multi-level analysis requires larger design space
  - small multiples: juxtaposed views
    - vertical position within row: algorithmic, avoid occlusion
    - vertical position across rows: encodes job type attribute
  - superposition: layered views
  - nesting: multi-scale views / glyphs



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

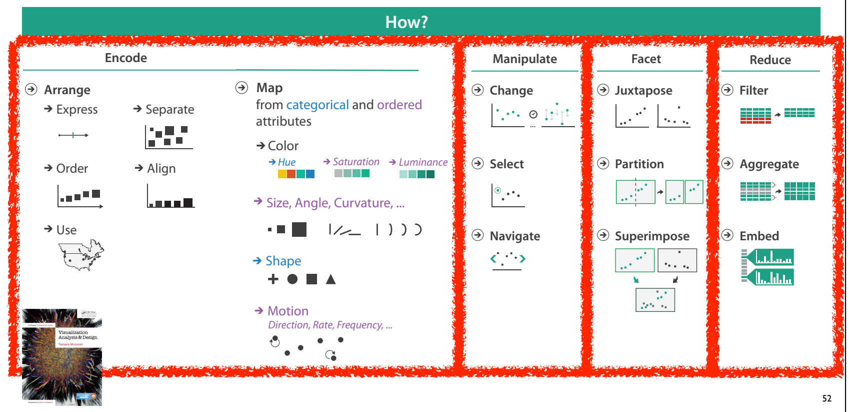


https://d3-graph-gallery.com/line.html



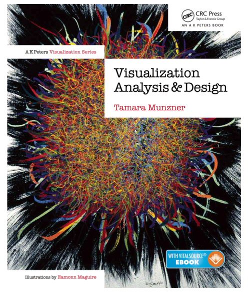
https://www.anychart.com/blog/2020/11/06/election-maps-us-vote-live-results/

Rethinking book design space: Visualization Analysis & Design 2e



More stuff

- this talk
  http://www.cs.ubc.ca/~tmm/talks.html#mit24
  - 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

@tamaramunzner