Slide 1

Marks Revisited: Beyond Bertin

Tamara Munzner
Department of Computer Science
University of British Columbia

Northeastern Vancouver
12 Jul 2024

UBC
DESIGNING for PEOPLE
CAIDA
UBC Data Science Institute

@tamara@vis.social
@tamaramunzner

Slide 2 — 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

Slide 3 — 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

Slide 4 — Channels: Rankings

Magnitude Channels: Ordered Attributes
- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)

Identity Channels: Categorical Attributes
- Spatial region
- Color hue
- Motion
- Shape

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

Slide 5 — Talk outline

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

Slide 6

Current Marks & Channels Model
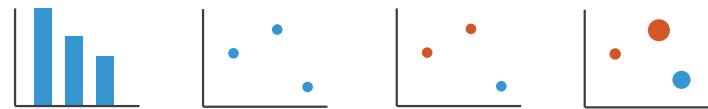
Slide 7 — Visual encoding model

- analyze idiom structure as combination of marks and channels

Slide 8 — Visual encoding model

- analyze idiom structure as combination of marks and channels

idiom: bar chart

1 channel:
vertical position

mark: line

Slide 9 — 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

Slide 10 — 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

Slide 11 — 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

Slide 12 — 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
Attributes (columns)
Items (rows)
Cell containing value

Slide 13 — Visual encoding model: Spatial data

- marks for items of spatial data

idiom: choropleth map

channels:
position
color (saturation)

mark: area

Slide 14 — Visual encoding model: Network data

- marks for items and marks for links

→ Networks
Link
Node (item)

→ Trees

Connection
Containment

idiom: node-link diagram

channel:
position

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

Node–Link Diagram   Treemap

idiom: treemap

channel:
position

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

Slide 15 — 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
      - [Ralph. Toward Methodological Guidelines for Process Theories & Taxonomies in Software Engineering IEEE TSE 2020]

Slide 16 — 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

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

Rethinking book design space: Visualization Analysis & Design 2e

How?

Encode | Manipulate | Facet | Reduce

Encode
- Arrange
  - → Express → Separate
  - → Order → Align
  - → Use
- 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

17

Teaching Challenges

18

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

19

Teaching: Bertini in-class exercises, catalyst for questions

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

20

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

https://twitter.com/MonaChalabi/status/1158779046693679106?s=20
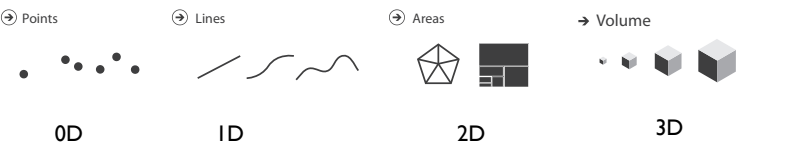
21

Quiz: Name marks/channels

- Tax Rates
- marks
  - A: points
  - B: lines
  - C: areas
- channels
  - A: position
  - B: color
  - C: length
  - D: area
  - E: angle

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

22

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

23

Many, many questions

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

24

Channels: Model evolves, heavily studied

- effectiveness rankings
- expressiveness matches, data & task

25

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

26

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
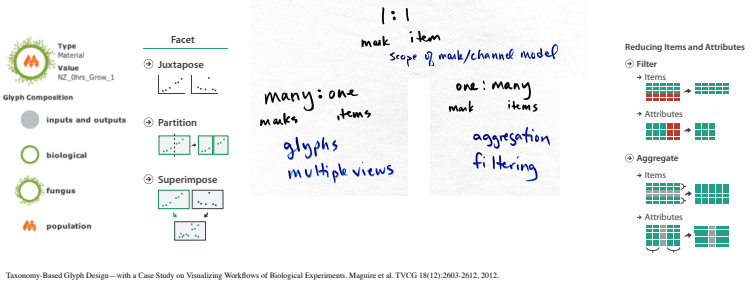
Taxonomy-Based Glyph Design — with a Case Study on Visualizing Workflows of Biological Experiments. Magnire et al. TVCG 18(12):2803-2812, 2012.

27

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!

28

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?

[A layered grammar of graphics. Wickham. Journal of Computational and Graphical Statistics 19:1 (2010), 3-28]
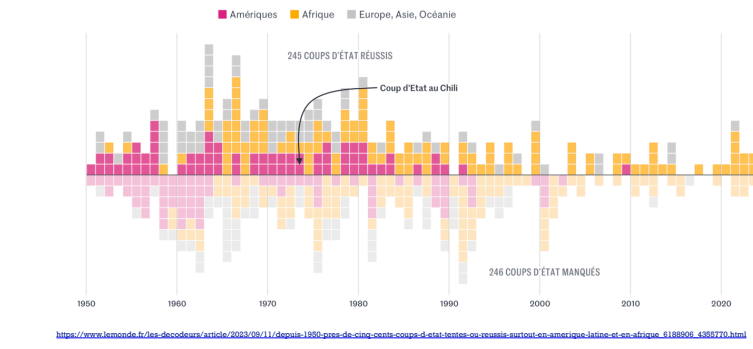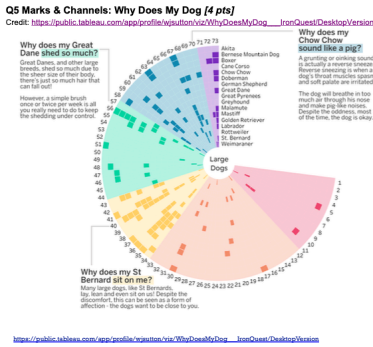
29

Encode vs decode: Where do models diverge?

- 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
- 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

30

Constraints

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

Alternative Ideas

31

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/xjsutton/viz/WhyDoesMyDog____IronQuest/DesktopVersion

https://public.tableau.com/app/profile/xjsutton/viz/WhyDoesMyDog____IronQuest/DesktopVersion

33

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?

34

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!

35

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://clear.robledo.com/@43ue-state-choropleth
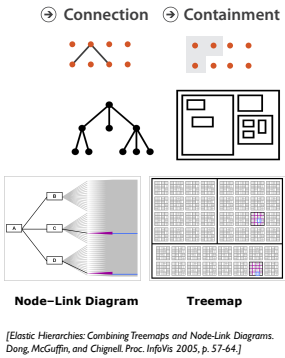
36
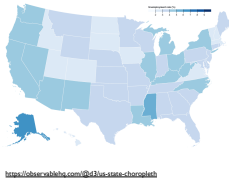
Interlocking (area) marks

- 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

US Presidential Election 2020

https://worldmapper.org/us-presidential-election-2020/
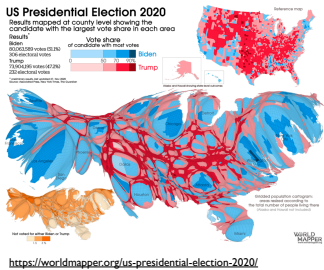
37

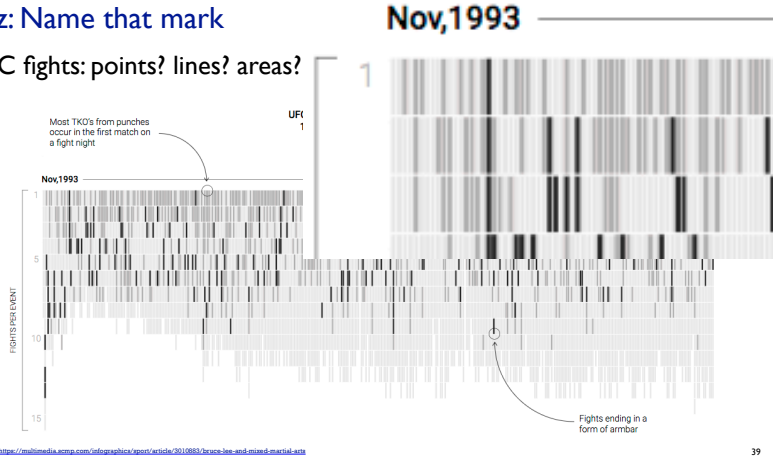## Interlocking marks: Non-spatial

- 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
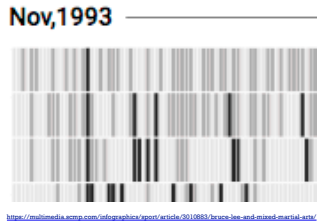
Node–Link Diagram   Treemap

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

38

Quiz: Name that mark

- UFC fights: points? lines? areas?

Nov,1993

Most TKO's from punches occur in the first match on a fight night

Fights ending in a form of armbar

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

39

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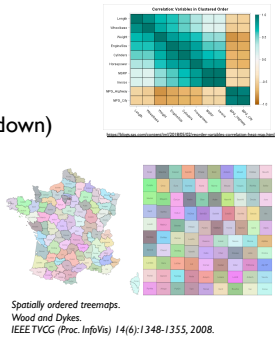
- what type of mark?
  - line?
    - no, not length coded
  - point mark with rectangular shape?
    - 2020: yes!
    - 2023: no!
      - cannot change position / size / orientation
  - area?
    - 2020: no, area/shape does not convey meaning
    - 2023: yes!
      - fully interlocking
      - position,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

Nov,1993

https://multimedia.scmp.com/infographics/sport/article/3010883/bruce-lee-and-mixed-martial-arts
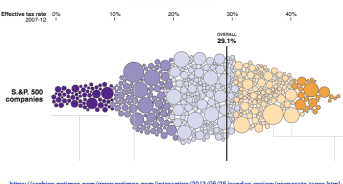
40

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

41

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

42

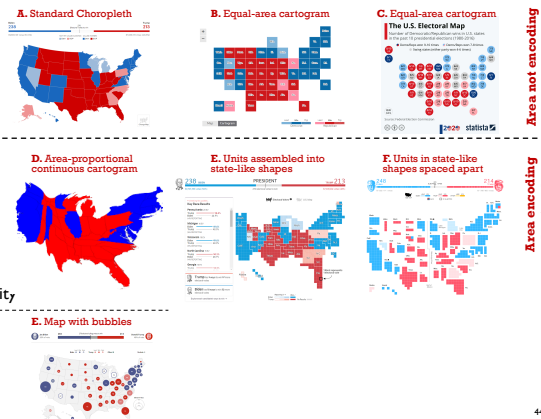## Interlocking marks: Circle packings

- 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/distribution-of-annual-income/
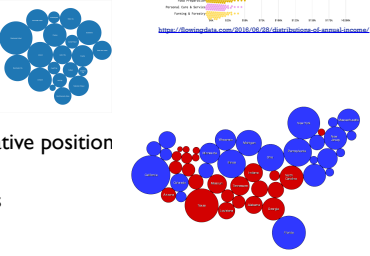
43

## 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

A. Standard Choropleth   B. Equal-area cartogram   C. Equal-area cartogram The U.S. Electoral Map

D. Area-proportional continuous cartogram   E. Units assembled into state-like shapes   F. Units in state-like shapes spaced apart

Area not encoding election data

Area encoding election data

E. Map with bubbles

https://www.washingtonpost.com/graphics/politics/2016/dynamic-maps-us-time-zones/

44

## 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?
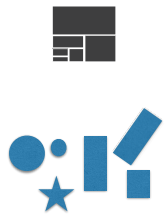
45

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
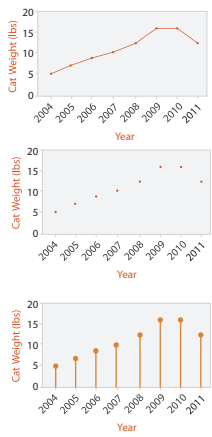
46

## 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

Lines

segments   paths
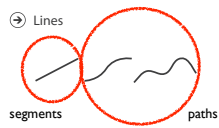
47

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

Points   Lines
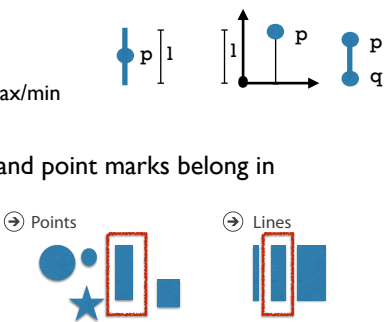
Line marks vs area marks

- 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.

49
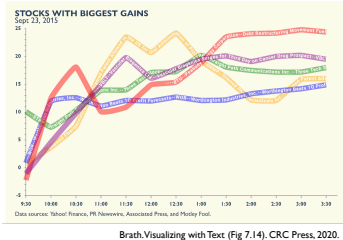
Line charts vs filled area charts

- 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).]

50

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

51

Alternative mark types model

52

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?**

53

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

54

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

- size is local, in contrast to shared position
  - can be independent across marks

https://www.fastlogo.com/blog/2013/12/18/bar-mobile-chart-details/ 55

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/1188779046883679510879820
https://www.washingtonpost.com/world/2023/08/10/dubai-heat-staying-cool/ 56

Channel Dependencies/Hierarchy proto-model musings

57

Shared vs local example

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

position along aligned scale

length

58

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/2015/06/26/distributions-of-annual-incomes/ 59

Multi-level analysis needed: 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 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/1188779046883679510879820 60

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

https://blocks.roadtolarissa.com/mbostock/4679202 61

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/2015/06/26/distributions-of-annual-incomes/ 62

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)

63

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?

64

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**

🐦 **@tamaramunzner**