Why is validation difficult?
• different ways to get it wrong at each level
• different ways to get it wrong at each level

Analysis: Analyze, Query
• analyze
  • consume
  • discover vs present
    • do not explore vs explain
  • enjoy
  • annotate, record, derive
• query
  • how much data matters?
  • one, some, all
  • independent choices
  • analyze, query, search

Derive: Crucial Design Choice
• don’t just draw what you give!
  • decide what the right thing to show is
  • create it with a series of transformations from the original dataset
draw three
• one of the four major strategies for handling complexity

How to encode: Arrange space, map channels

• marks
  • aspect ratio
  • shape
  • color
  • size
  • orientation
  • trajectory
  • temporal
  • motion
  • visual encoding
  • positional appearance
  • color
  • size
  • animation

Definitions: Marks and channels

Dataset Types
• Table
• Networks

Types: Datasets and data
• Spatial
• List

Analysis example: Derive one attribute
• Strahler number (resilient metric for trees/networks)
  • derived quantitative attribute
  • draw top 5K of 500K for good data
• derive visualization techniques based on the derived attributes

Why analyze?
• imposes a structure on huge design space
  • scaffold to help you think
  • critical for new form

Derive: Crucial Design Choice
• don’t just draw what you’re given!
  • decide what the right thing to show is
  • create it with a series of transformations from the original dataset
draw three
• one of the four major strategies for handling complexity

How to encode: Arrange space, map channels

• marks
  • aspect ratio
  • shape
  • color
  • size
  • orientation
  • trajectory
  • temporal
  • motion
  • visual encoding
  • positional appearance
  • color
  • size
  • animation

Definitions: Marks and channels

Dataset Types
• Table
• Networks

Types: Datasets and data
• Spatial
• List

Analysis example: Derive one attribute
• Strahler number (resilient metric for trees/networks)
  • derived quantitative attribute
  • draw top 5K of 500K for good data
• derive visualization techniques based on the derived attributes

Why analyze?
• imposes a structure on huge design space
  • scaffold to help you think
  • critical for new form

Derive: Crucial Design Choice
• don’t just draw what you’re given!
  • decide what the right thing to show is
  • create it with a series of transformations from the original dataset
draw three
• one of the four major strategies for handling complexity

How to encode: Arrange space, map channels

• marks
  • aspect ratio
  • shape
  • color
  • size
  • orientation
  • trajectory
  • temporal
  • motion
  • visual encoding
  • positional appearance
  • color
  • size
  • animation

Definitions: Marks and channels

Dataset Types
• Table
• Networks

Types: Datasets and data
• Spatial
• List

Analysis example: Derive one attribute
• Strahler number (resilient metric for trees/networks)
  • derived quantitative attribute
  • draw top 5K of 500K for good data
• derive visualization techniques based on the derived attributes

Why analyze?
• imposes a structure on huge design space
  • scaffold to help you think
  • critical for new form

Derive: Crucial Design Choice
• don’t just draw what you’re given!
  • decide what the right thing to show is
  • create it with a series of transformations from the original dataset
draw three
• one of the four major strategies for handling complexity

How to encode: Arrange space, map channels

• marks
  • aspect ratio
  • shape
  • color
  • size
  • orientation
  • trajectory
  • temporal
  • motion
  • visual encoding
  • positional appearance
  • color
  • size
  • animation

Definitions: Marks and channels

Dataset Types
• Table
• Networks

Types: Datasets and data
• Spatial
• List

Analysis example: Derive one attribute
• Strahler number (resilient metric for trees/networks)
  • derived quantitative attribute
  • draw top 5K of 500K for good data
• derive visualization techniques based on the derived attributes

Why analyze?
• imposes a structure on huge design space
  • scaffold to help you think
  • critical for new form

Derive: Crucial Design Choice
• don’t just draw what you’re given!
  • decide what the right thing to show is
  • create it with a series of transformations from the original dataset
draw three
• one of the four major strategies for handling complexity

How to encode: Arrange space, map channels

• marks
  • aspect ratio
  • shape
  • color
  • size
  • orientation
  • trajectory
  • temporal
  • motion
  • visual encoding
  • positional appearance
  • color
  • size
  • animation

Definitions: Marks and channels

Dataset Types
• Table
• Networks

Types: Datasets and data
• Spatial
• List

Analysis example: Derive one attribute
• Strahler number (resilient metric for trees/networks)
  • derived quantitative attribute
  • draw top 5K of 500K for good data
• derive visualization techniques based on the derived attributes

Why analyze?
• imposes a structure on huge design space
  • scaffold to help you think
  • critical for new form

Derive: Crucial Design Choice
• don’t just draw what you’re given!
  • decide what the right thing to show is
  • create it with a series of transformations from the original dataset
draw three
• one of the four major strategies for handling complexity

How to encode: Arrange space, map channels

• marks
  • aspect ratio
  • shape
  • color
  • size
  • orientation
  • trajectory
  • temporal
  • motion
  • visual encoding
  • positional appearance
  • color
  • size
  • animation

Definitions: Marks and channels

Dataset Types
• Table
• Networks

Types: Datasets and data
• Spatial
• List

Analysis example: Derive one attribute
• Strahler number (resilient metric for trees/networks)
  • derived quantitative attribute
  • draw top 5K of 500K for good data
• derive visualization techniques based on the derived attributes

Why analyze?
• imposes a structure on huge design space
  • scaffold to help you think
  • critical for new form

Derive: Crucial Design Choice
• don’t just draw what you’re given!
  • decide what the right thing to show is
  • create it with a series of transformations from the original dataset
draw three
• one of the four major strategies for handling complexity

How to encode: Arrange space, map channels

• marks
  • aspect ratio
  • shape
  • color
  • size
  • orientation
  • trajectory
  • temporal
  • motion
  • visual encoding
  • positional appearance
  • color
  • size
  • animation

Definitions: Marks and channels

Dataset Types
• Table
• Networks

Types: Datasets and data
• Spatial
• List

Analysis example: Derive one attribute
• Strahler number (resilient metric for trees/networks)
  • derived quantitative attribute
  • draw top 5K of 500K for good data
• derive visualization techniques based on the derived attributes

Why is validation difficult?
• different ways to get it wrong at each level

Analysis: Analyze, Query
• analyze
  • consume
  • discover vs present
    • do not explore vs explain
  • enjoy
  • annotate, record, derive
• query
  • how much data matters?
  • one, some, all
  • independent choices
  • analyze, query, search
Partitioning: Recursive subdivision

- split by neighborhood
- then by type
- then time
  - years as rows
  - months as columns
- color by price

- neighborhood patterns
  - where it’s expensive
  - where you pay much more for detached type

Reducing items and attributes

- reduce/increase inverses
- filter
  - pro: straightforward and intuitive
  - con: out of sight, out of mind
  - aggregation
  - pro: inform about whole set
  - con: difficult to avoid losing signal
- not mutually exclusive
  - combine filter, aggregate
  - combine reduce, facet, change, derive

Idiom: boxplot

- static item aggregation
- task: find distribution
  - data: table
  - derived data
    - 5 number summary
    - lower and upper quartile boxes
    - lower and upper fences
    - outliers beyond fence cutoffs explicitly shown

More Information

- this talk
  [http://www.cs.ubc.ca/~tmm/talks.html#vad16pacvis](http://www.cs.ubc.ca/~tmm/talks.html#vad16pacvis)
- book page (including tutorial lecture slides)
  [http://www.cs.ubc.ca/group/infovis](http://www.cs.ubc.ca/group/infovis)
  - 20% promo code for book+ebook combo: HIVE20
- illustrations: Eamonn Maguire
- papers, videos, software, talks, full courses
  [http://www.cs.ubc.ca/~tmm](http://www.cs.ubc.ca/~tmm)