Information Visualization
Task Abstraction

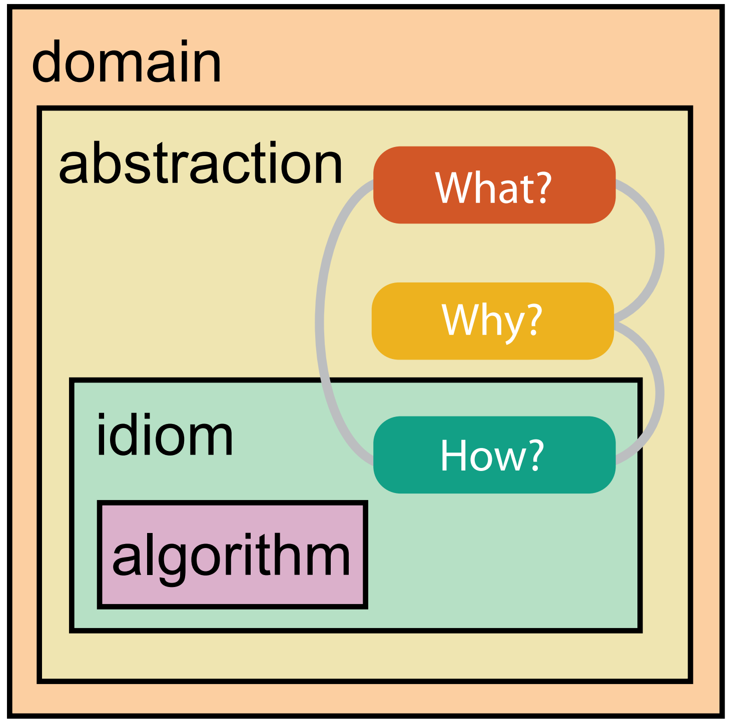
Tamara Munzner
Department of Computer Science
University of British Columbia

Lect 3, 14 Jan 2020

https://www.cs.ubc.ca/~tmm/courses/436V-20

Nested model: Four levels of visualization design

- *domain situation*
  - who are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of visualization
    - **what** is shown? **data** abstraction
    - **why** is the user looking at it? **task** abstraction
      - often must transform data, guided by task
- *idiom*
  - **how** is it shown?
    - **visual encoding** idiom: how to draw
    - **interaction** idiom: how to manipulate
- *algorithm*
  - efficient computation

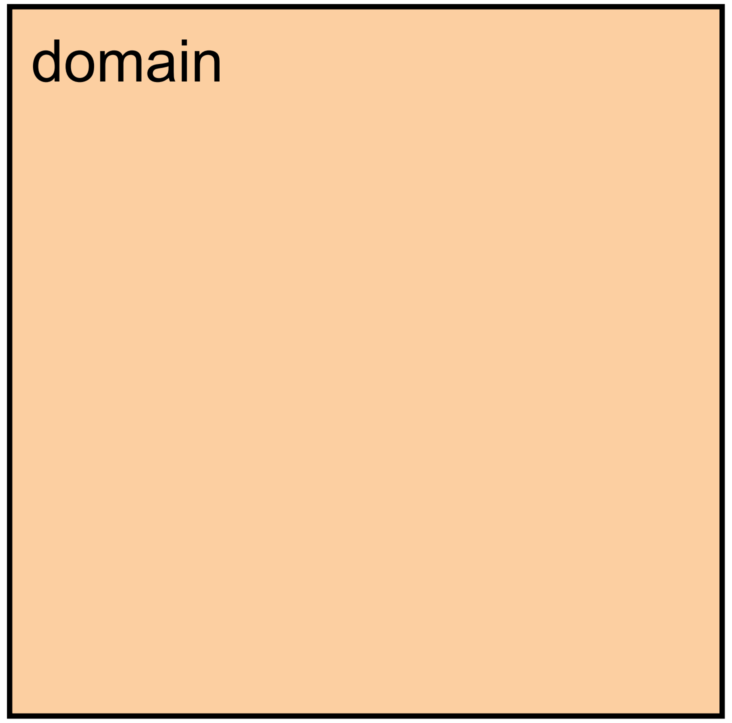


[A Nested Model of Visualization Design and Validation.
Munzner. IEEE TVCG 15(6):921-928, 2009
(Proc. InfoVis 2009).]

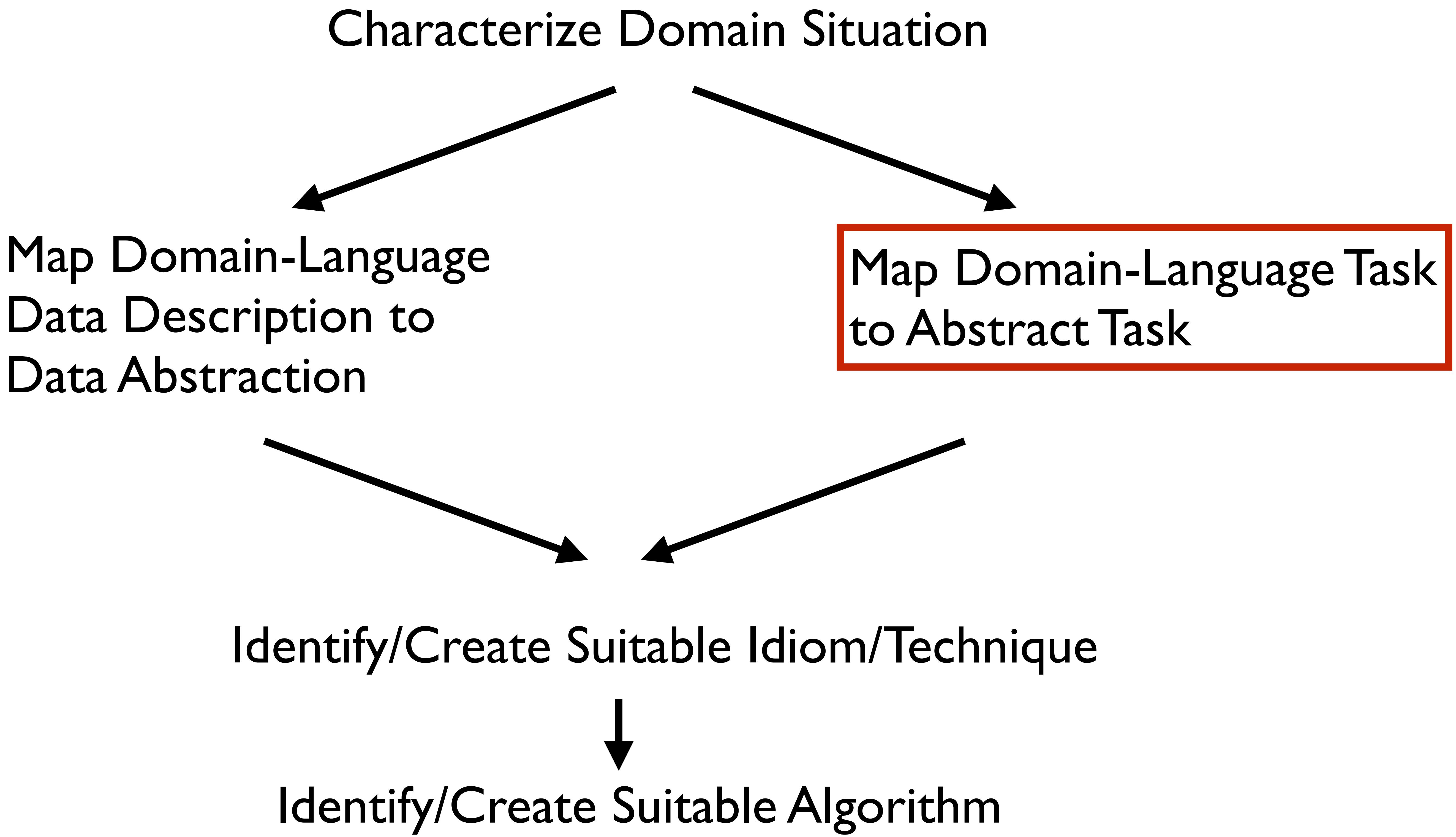
[A Multi-Level Typology of Abstract Visualization Tasks
Brehmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]

Domain characterization

- details of an application domain
- group of users, target domain, their questions, & their data
  - varies wildly by domain
  - must be specific enough to get traction
- domain questions/problems
  - break down into simpler abstract tasks



Design Process

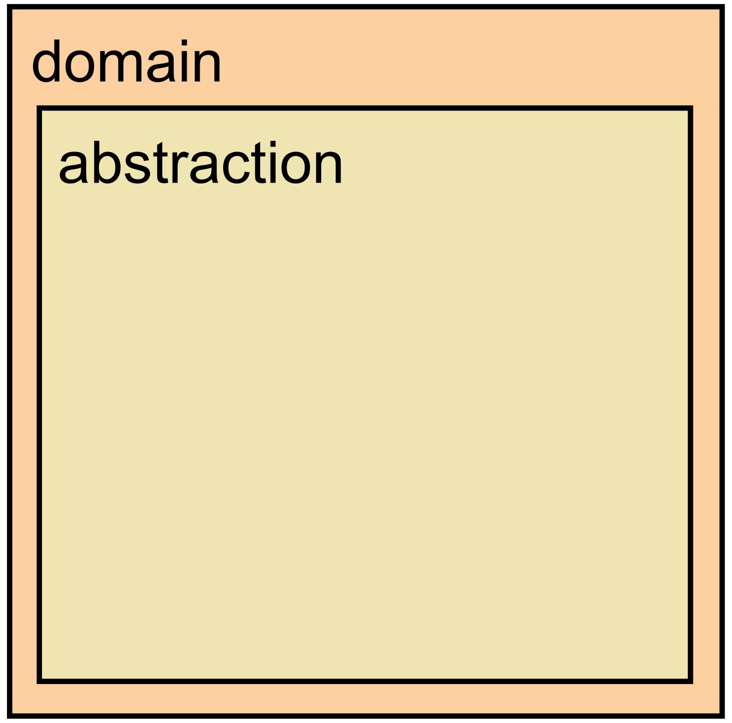


Example: Find good movies

- identify good movies in genres I like
- domain:
  - general population, movie enthusiasts

Abstraction: Data & task

- map *what* and *why* into generalized terms
  - identify tasks that users wish to perform, or already do
  - find data types that will support those tasks
    - possibly transform /derive if need be



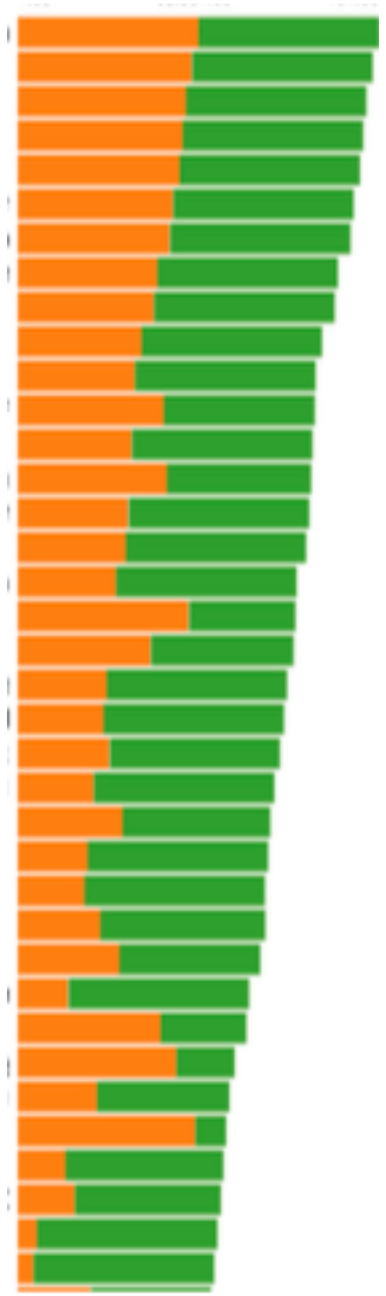
Example: Find good movies

- identify good movies in genres I like
- domain:
  - general population, movie enthusiasts
- task: what is a good movie for me?
  - highly rated by critics?
  - highly rated by audiences?
  - successful at the box office?
  - similar to movies I liked?
  - matches specific genres?
- data: (is it available?)
  - yes! data sources IMDB, Rotten Tomatoes…

Example: Find good movies

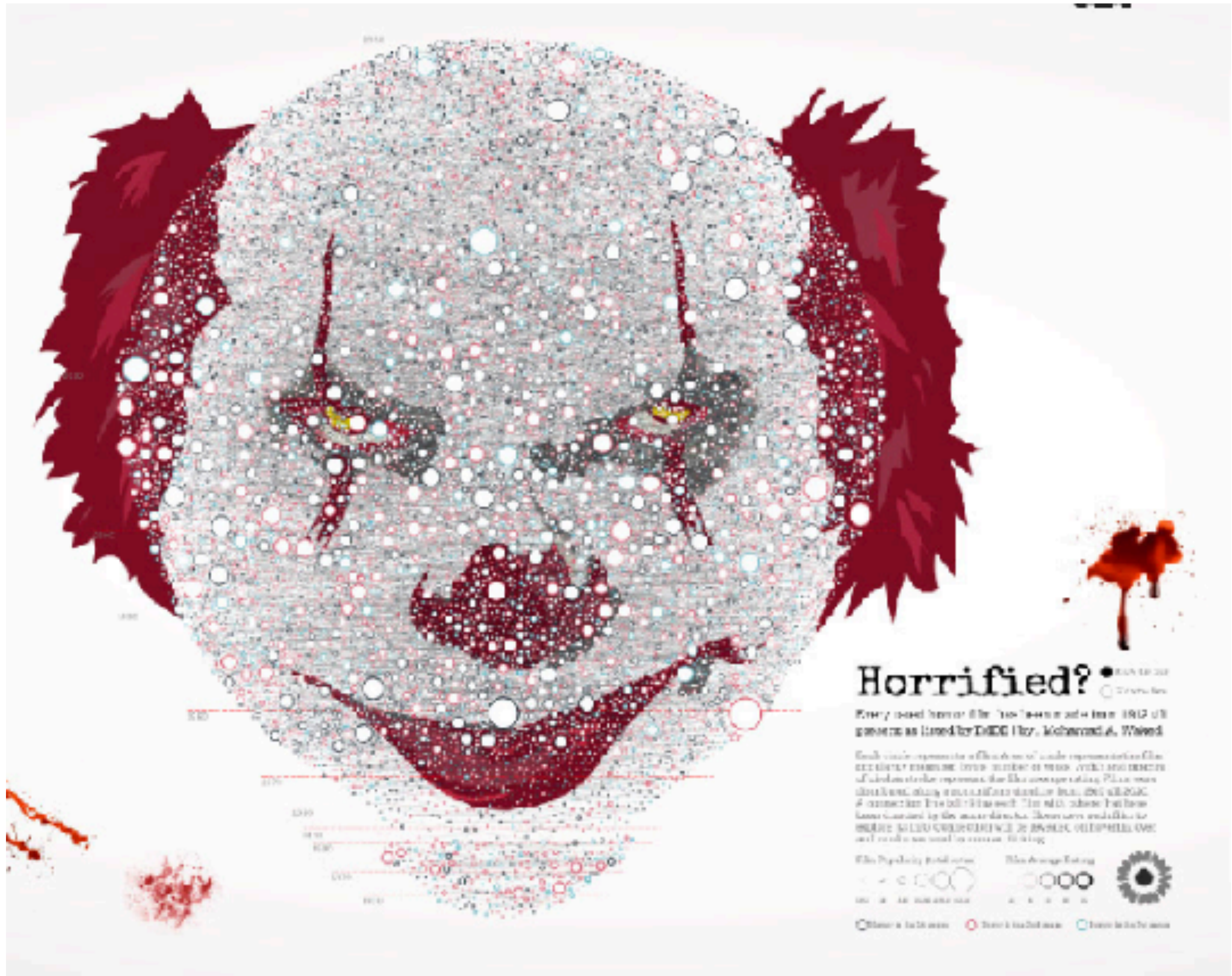
- one possible choice for data and tasks, in domain language
  - data: combine audience ratings and critic ratings
  - task: find high-scoring movies for specific genre
- abstractions?
  - attribute: audience & critic ratings
    - ordinal
      - levels: 3 or 5 or 10...
  - attribute: genre
    - categorical
      - levels: < 20
  - items: movies
    - items: millions
  - task: find high values?

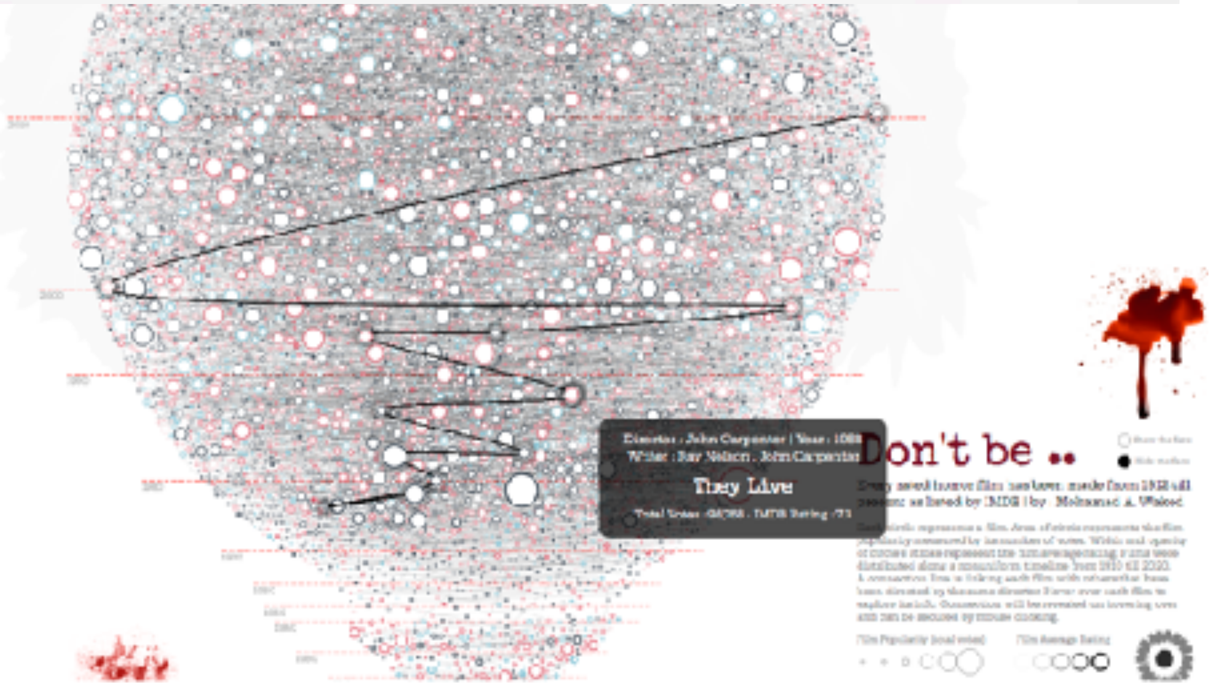
one possible idiom
- stacked bar chart for ratings



Example: Horrified

- same task: high-score movies
- slightly different data
  - 14K rated horror movies from IMDB
- very different visual encoding idiom
  - circle per item (movie)
  - circle area = popularity
  - stroke width/opacity = avg rating
  - year made = vertical position
- interaction idiom
  - lines connect movies w/ same director, on mouseover





http://alhadaqa.com/2019/10/horrified/

Why: Task Abstraction

Task abstraction: Actions and targets

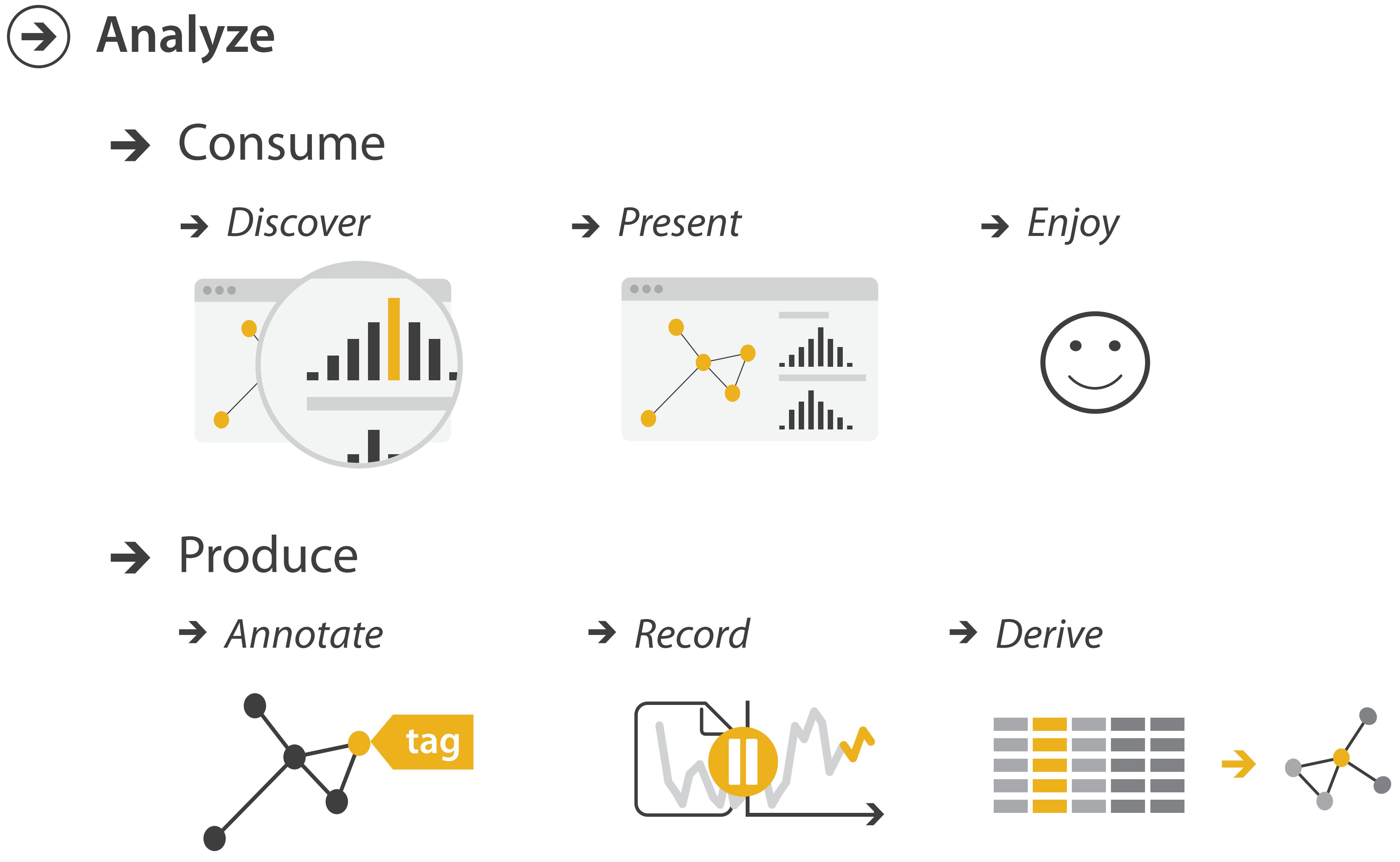
- very high-level pattern
- actions
  - analyze
    - high-level choices
  - search
    - find a known/unknown item
  - query
    - find out about characteristics of item

- {action, target} pairs
  - *discover distribution*
  - *compare trends*
  - *locate outliers*
  - *browse topology*

Actions: Analyze

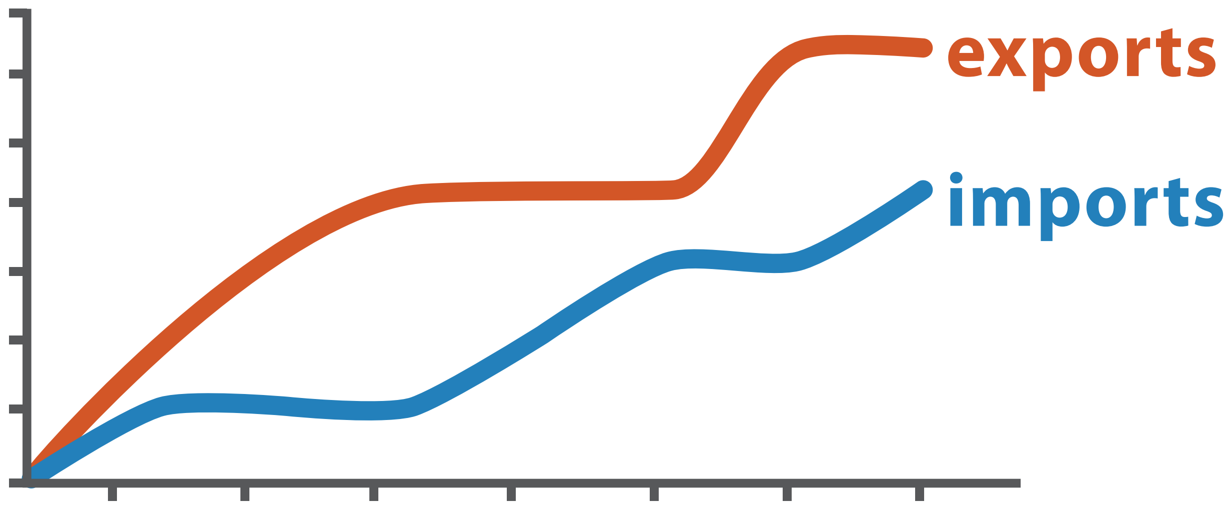
- consume
  - discover vs present
    - classic split
    - aka explore vs explain
  - enjoy
    - newcomer
    - aka casual, social

- produce
  - annotate, record
  - derive
    - crucial design choice

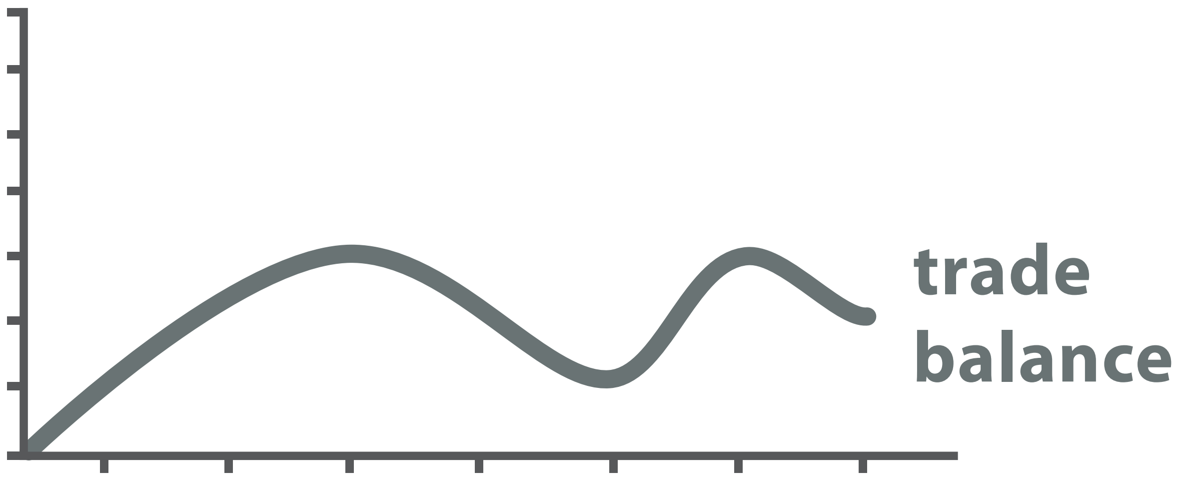


Derive

- 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 that
- *one of the four major strategies for handling complexity*



Original Data



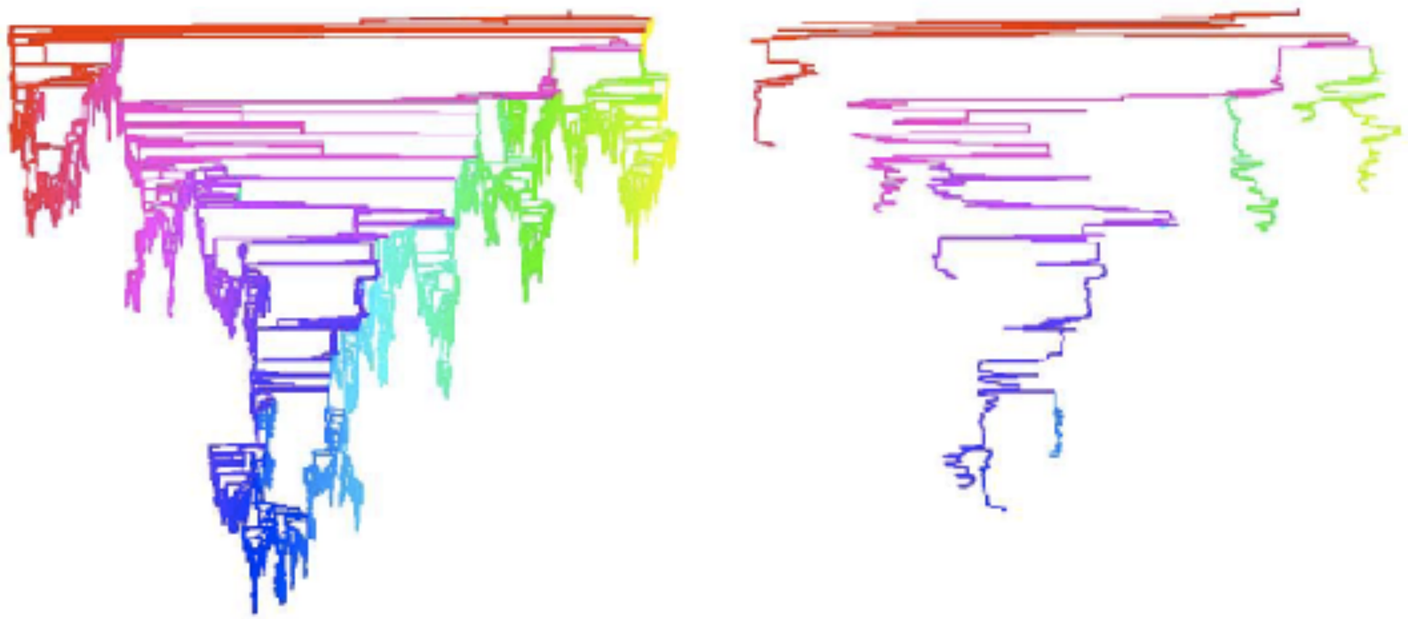
trade balance = exports − imports

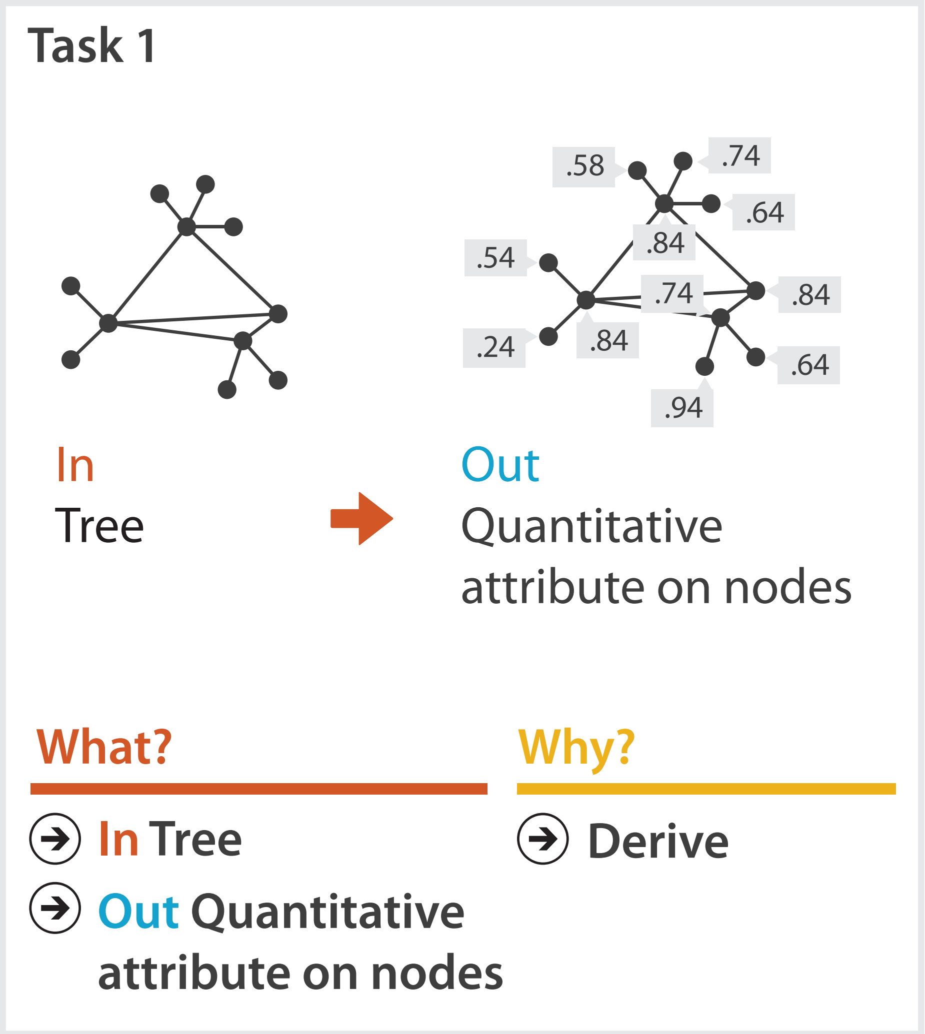
Derived Data

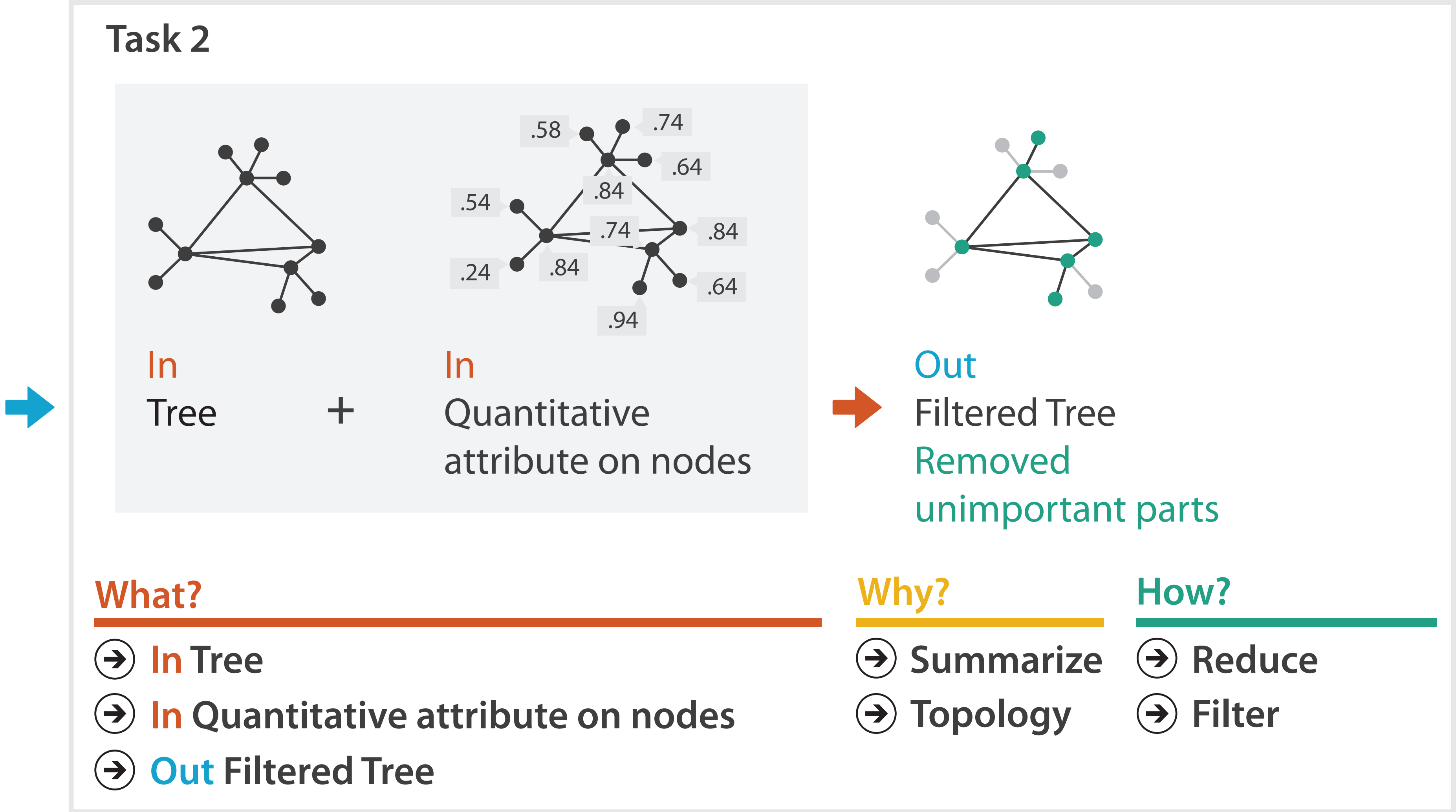
Analysis example: Derive one attribute

- Strahler number
  - centrality metric for trees/networks
  - derived quantitative attribute
  - draw top 5K of 500K for good skeleton

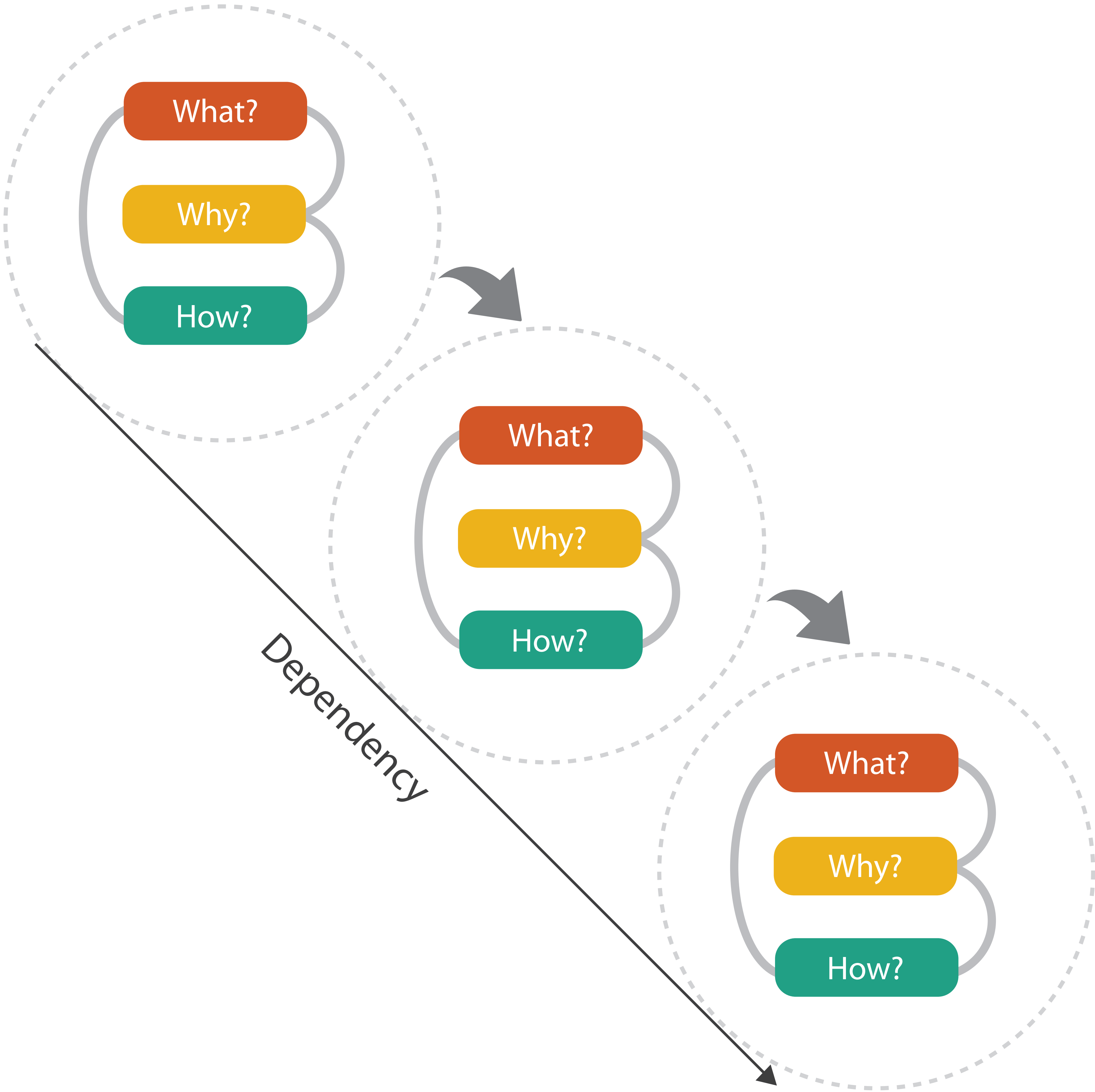
[Using Strahler numbers for real time visual exploration of huge graphs. Auber. Proc. Intl. Conf. Computer Vision and Graphics, pp. 56–69, 2002.]







Means and ends

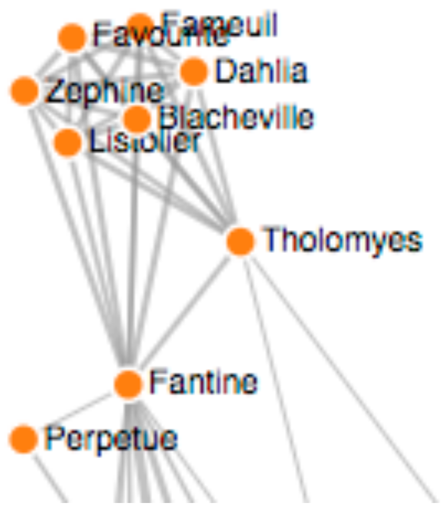


Actions: Search

- what does user know?
  - target, location
- lookup
  - ex: word in dictionary
    - alphabetical order
- locate
  - ex: keys in your house
  - ex: node in network
- browse
  - ex: books in bookstore
- explore
  - ex: cool neighborhood in new city

Search

	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*

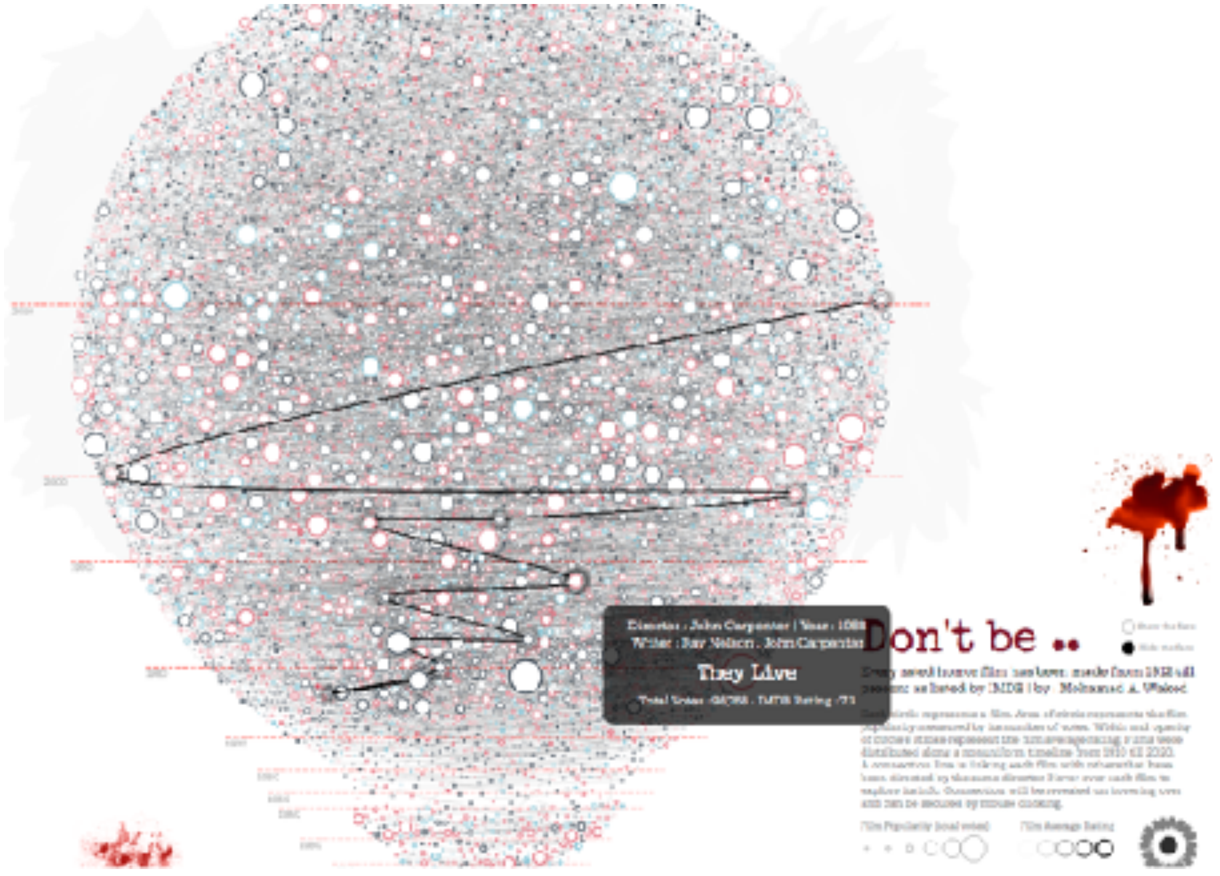


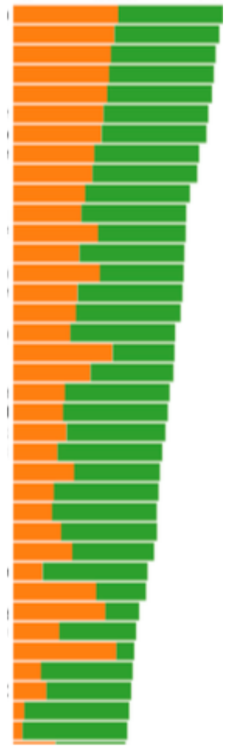
https://bl.ocks.org/heybignick/3faf257bbbbc7743bb72310d03b86ee8

Example: Horrified vs stacked bars

- horrified: browse/explore
- stacked bars: locate/lookup

- which is better?
  - depends on goals / task
    - enjoy, social context, lots of time
    - find 2nd-best rated movie of all time
      - Jeopardy call, < 10 seconds to respond!

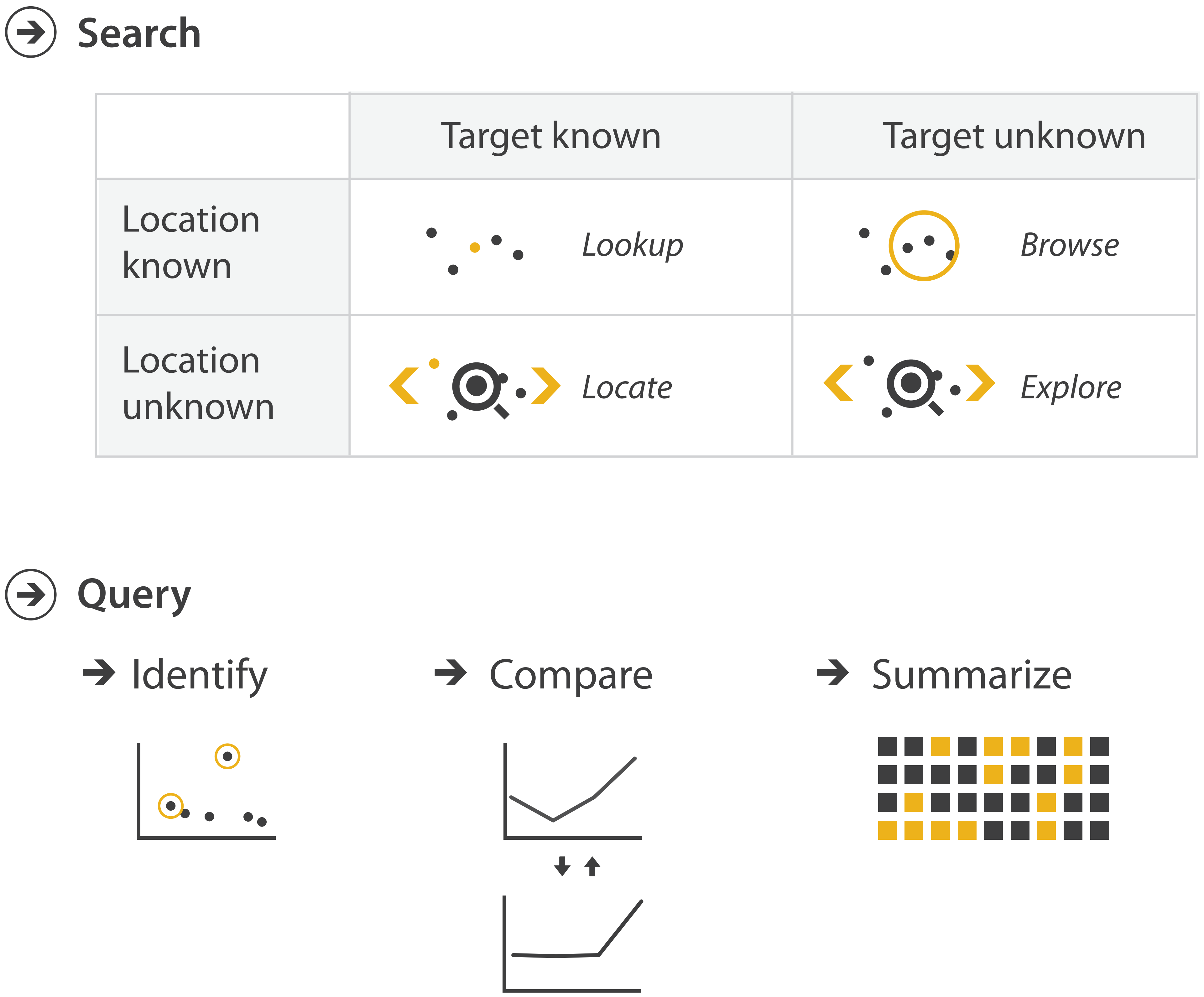




http://alhadaqa.com/2019/10/horrified/

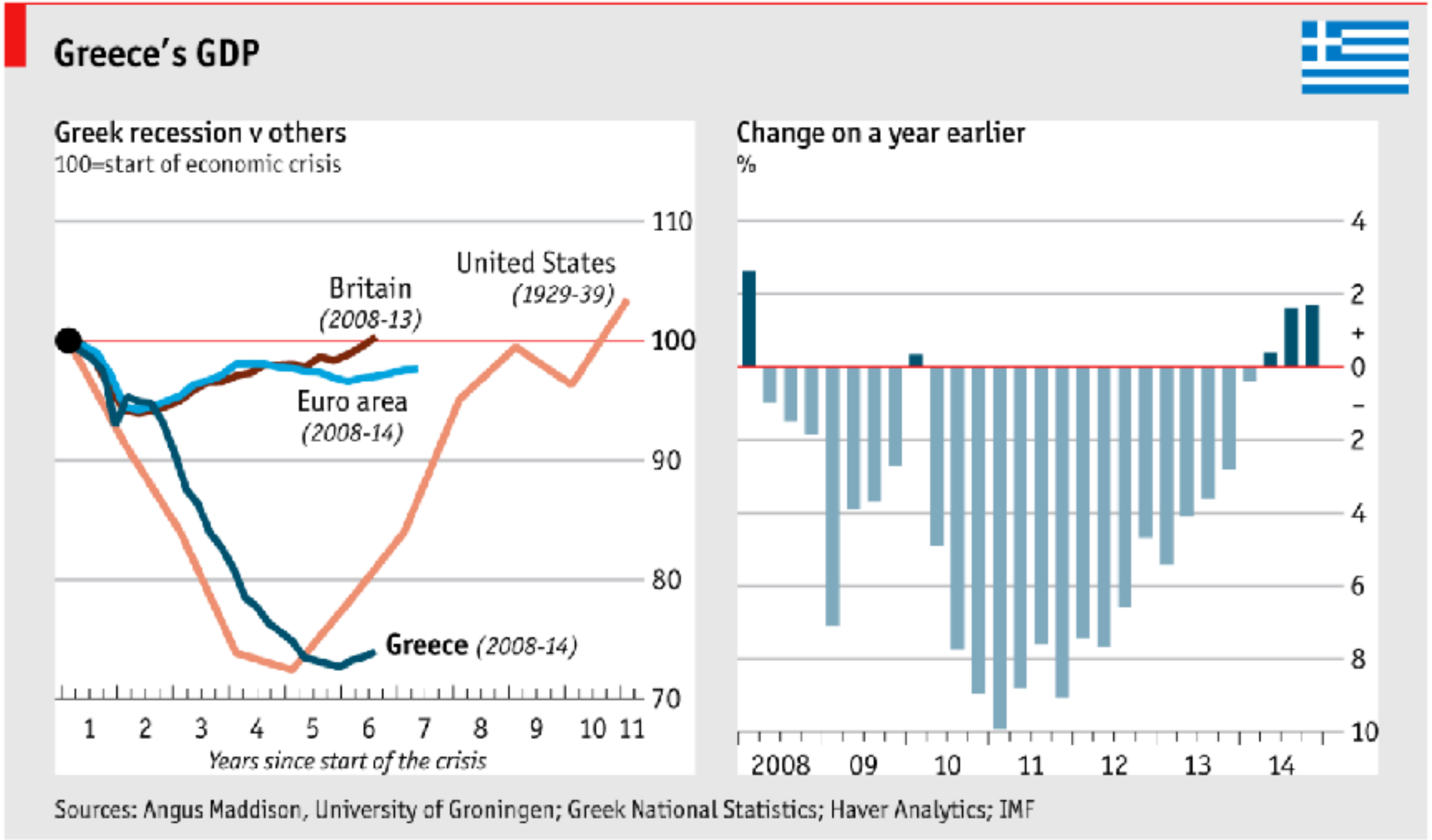
Actions: Search, query

- what does user know?
  - target, location
- how much of the data matters?
  - one, some, all
- independent choices for each of these three levels
  - analyze, search, query
  - mix and match



Example: Economics

- task: compare and derive
- data: derive change

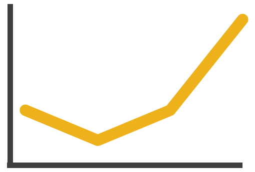


The Economist

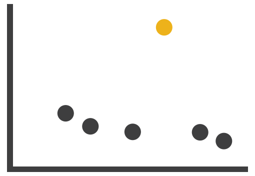
Task abstraction: Targets

➔ All Data

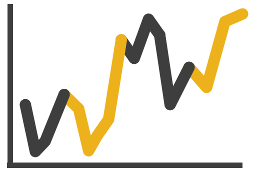
- ➔ Trends



- ➔ Outliers

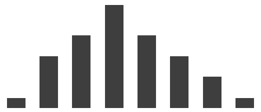


- ➔ Features

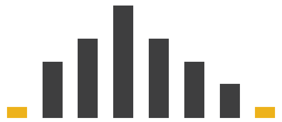


➔ Attributes

- ➔ One
  - ➔ *Distribution*

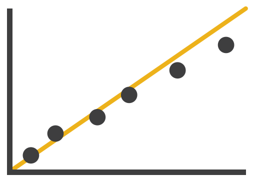


  - ➔ *Extremes*

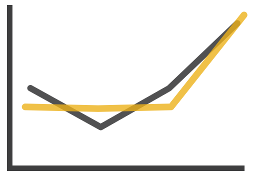


- ➔ Many
  - ➔ *Dependency*

  - ➔ *Correlation*

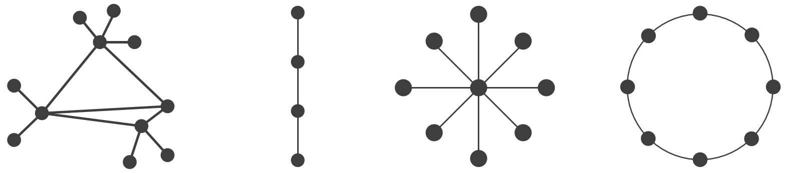


  - ➔ *Similarity*

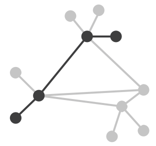


➔ Network Data

- ➔ Topology

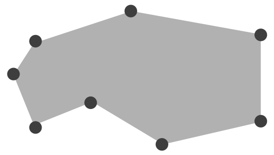


  - ➔ *Paths*



➔ Spatial Data

- ➔ Shape

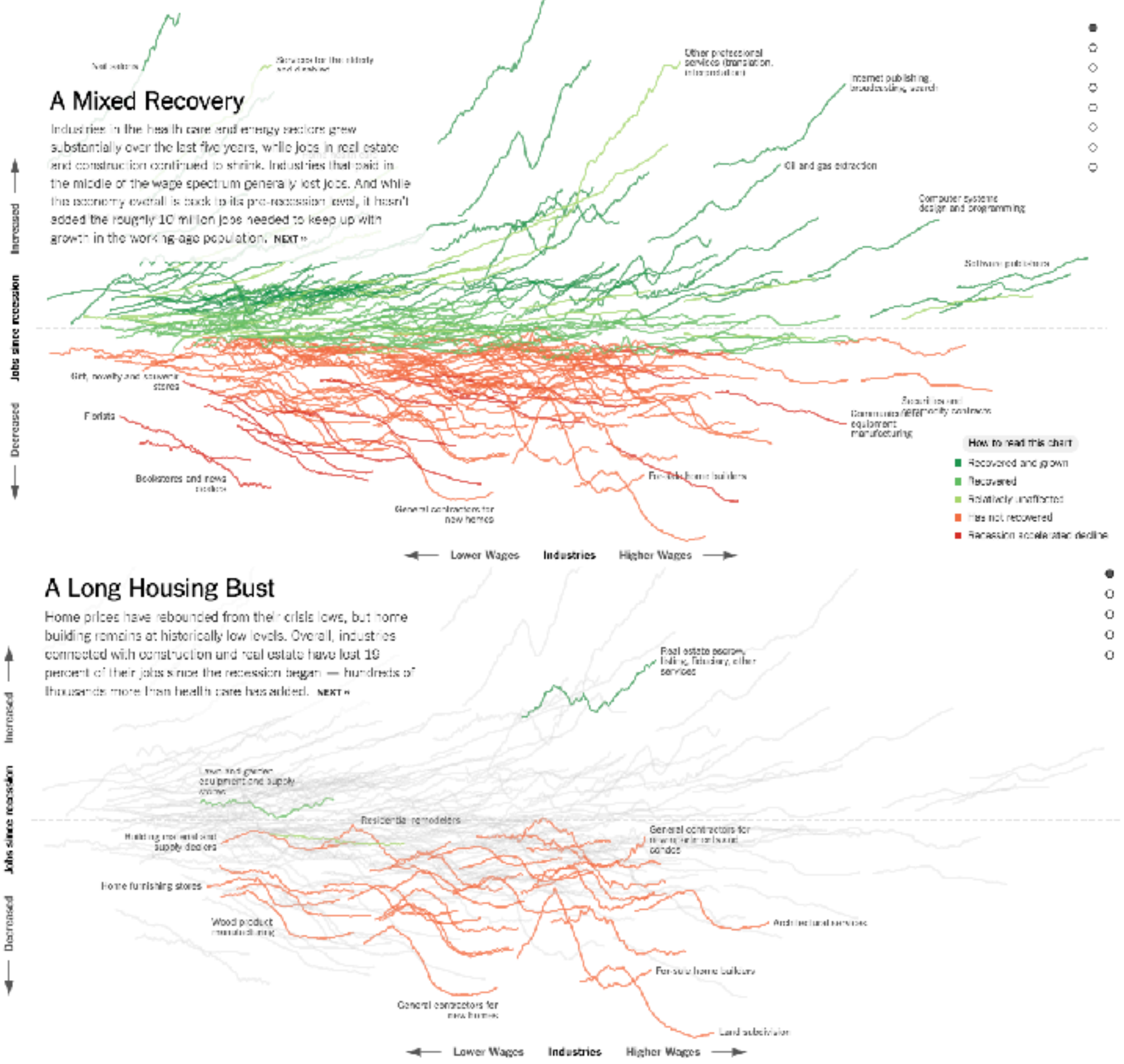


Abstraction

- these {action, target} pairs are good starting point for vocabulary
  - but sometimes you'll need more precision!
- rule of thumb
  - systematically remove all domain jargon
- interplay: task and data abstraction
  - need to use data abstraction within task abstraction
    - to specify your targets!
    - but task abstraction can lead you to transform the data
  - iterate back and forth
    - first pass data, first pass task, second pass data, ...

Examples: Job market

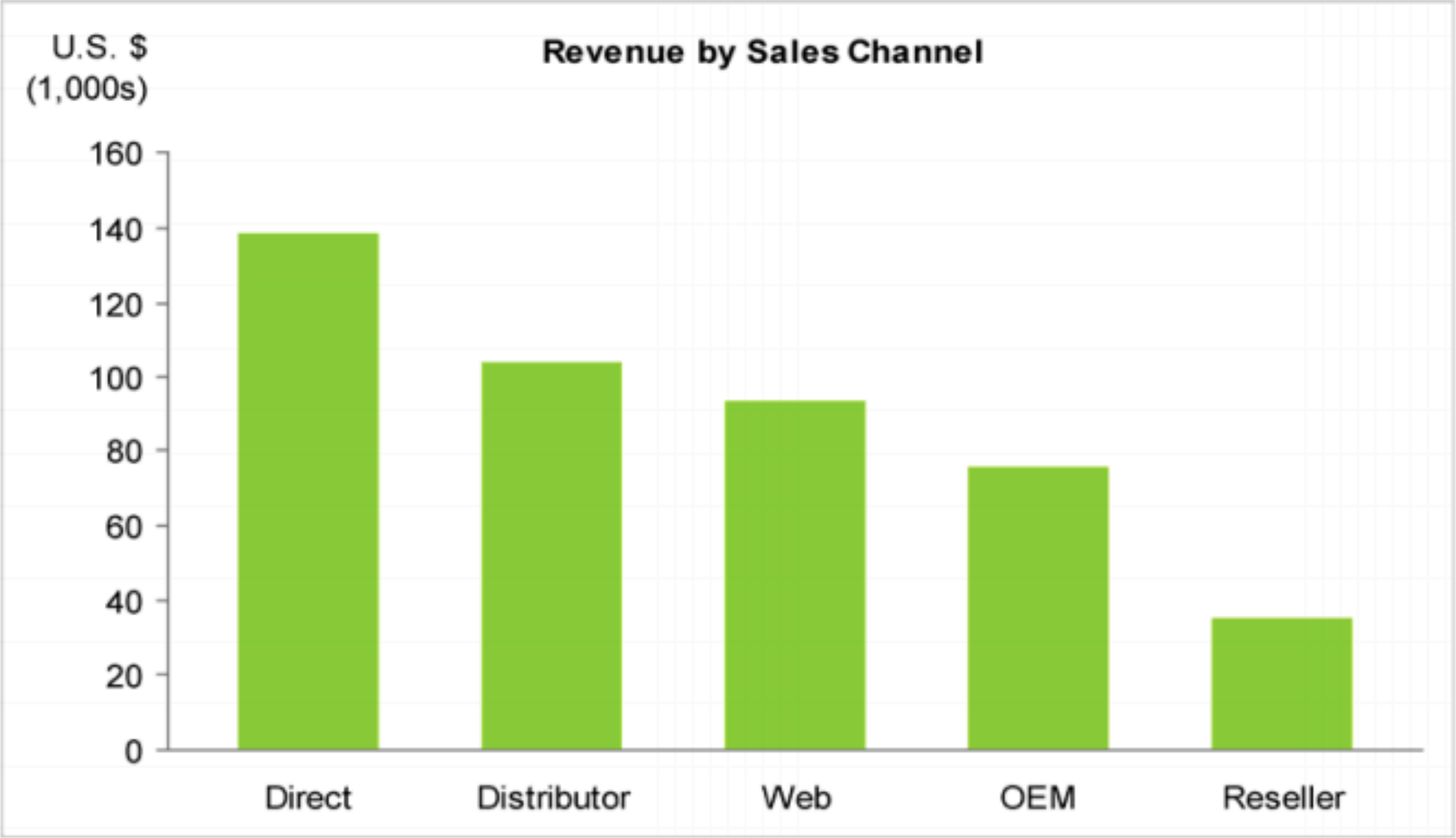
- trends
  - how did job market develop since recession overall?
- outliers
  - real estate related jobs

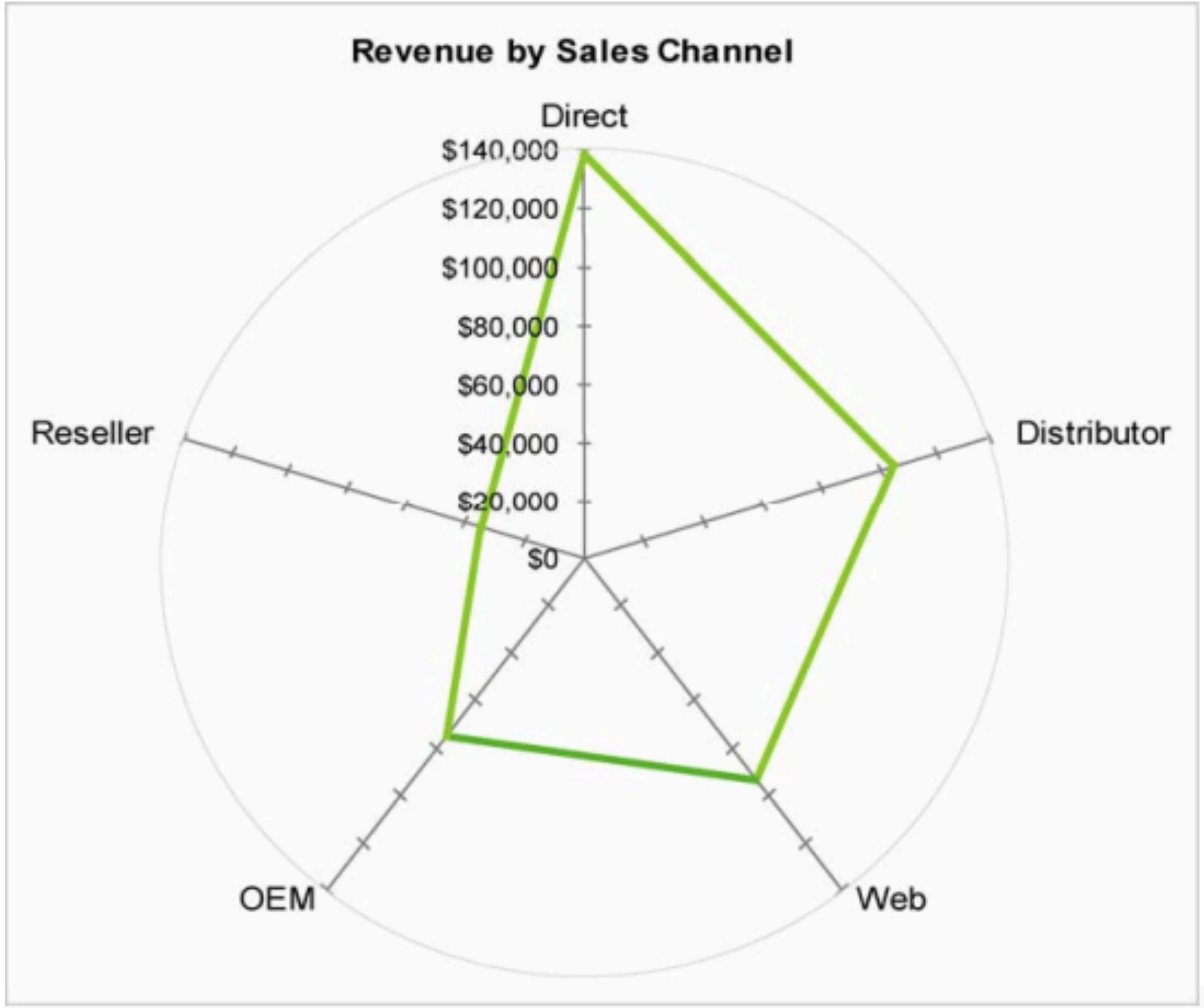


https://www.nytimes.com/interactive/2014/06/05/upshot/how-the-recession-reshaped-the-economy-in-255-charts.html

Exercise: Rating Charts for Tasks

- task A: sort attributes

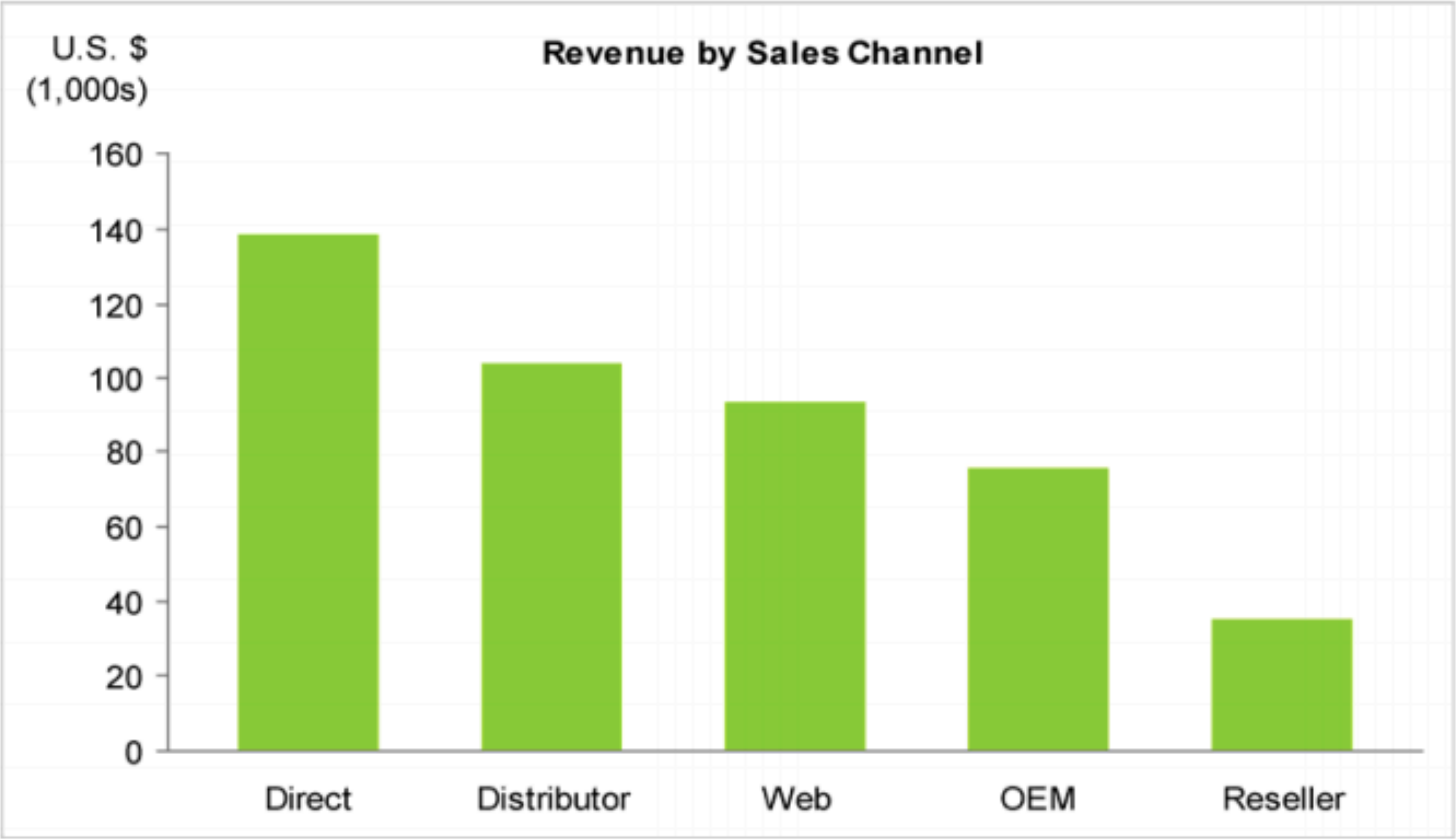


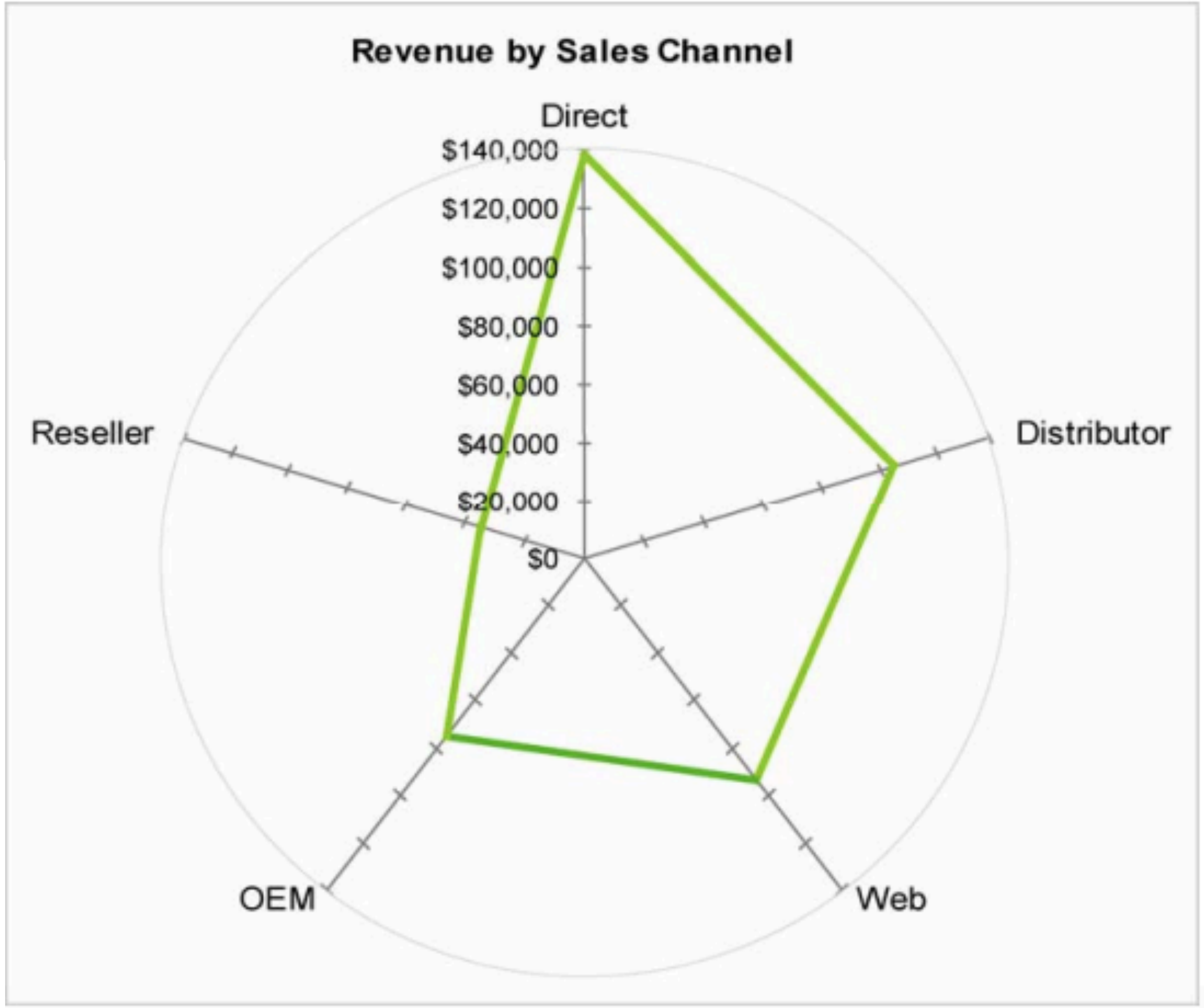


https://www.perceptualedge.com/articles/dmreview/radar_graphs.pdf

Exercise: Rating Charts for Tasks

- task B: compare pair of attributes (Direct vs Distributor)

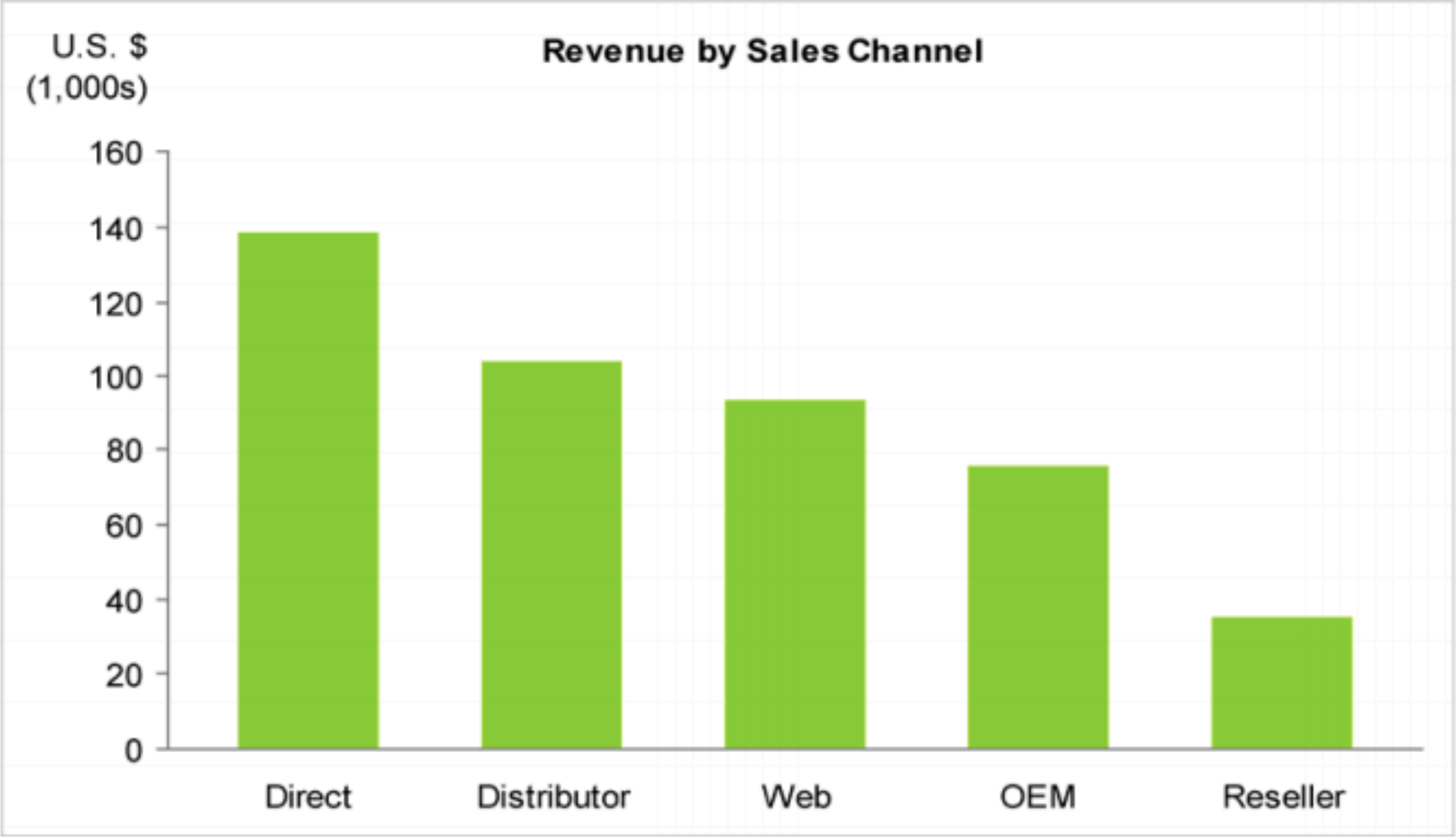


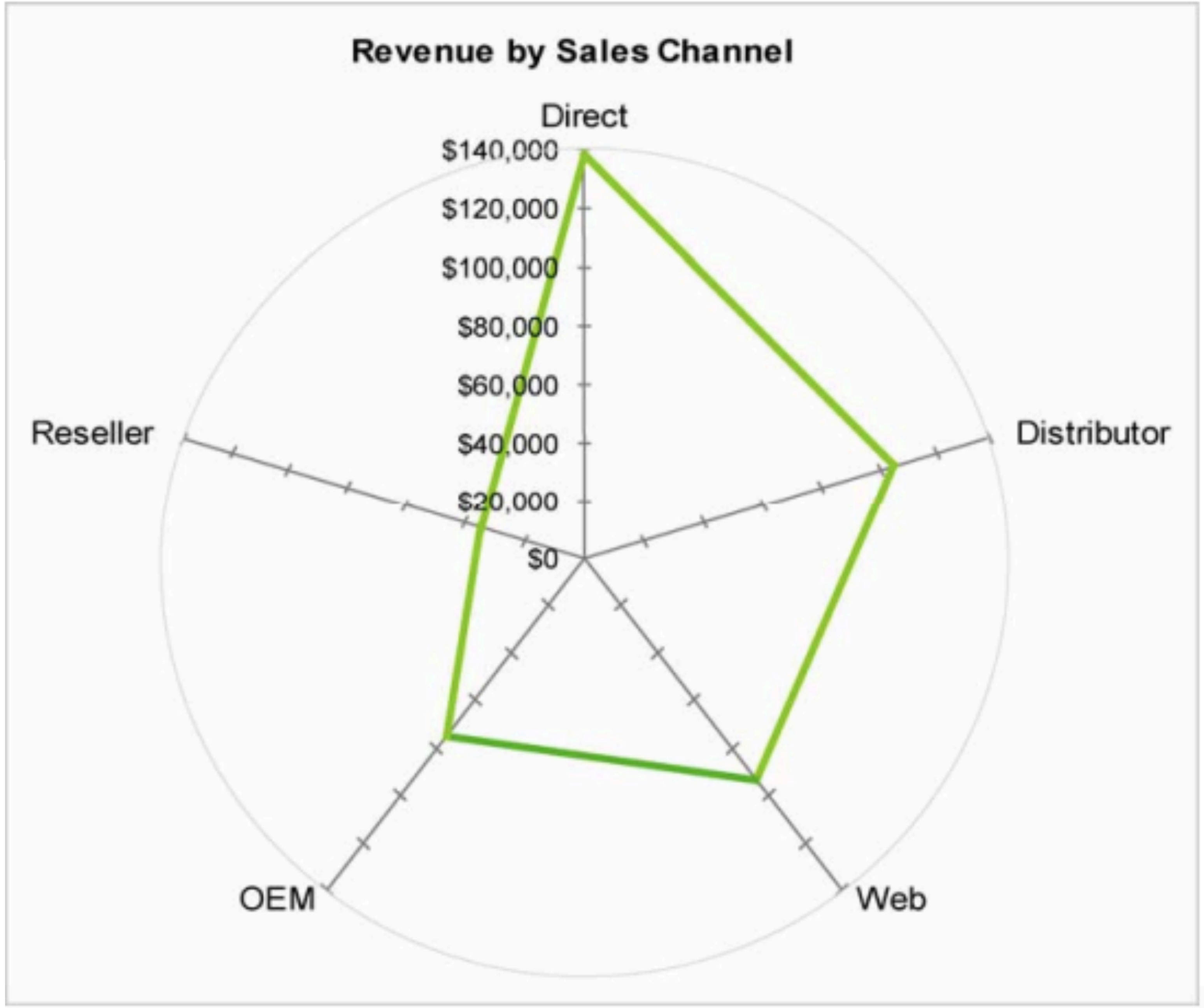


https://www.perceptualedge.com/articles/dmreview/radar_graphs.pdf

Exercise: Rating Charts for Tasks

- task C: compare pair of attributes (Distributor vs OEM)

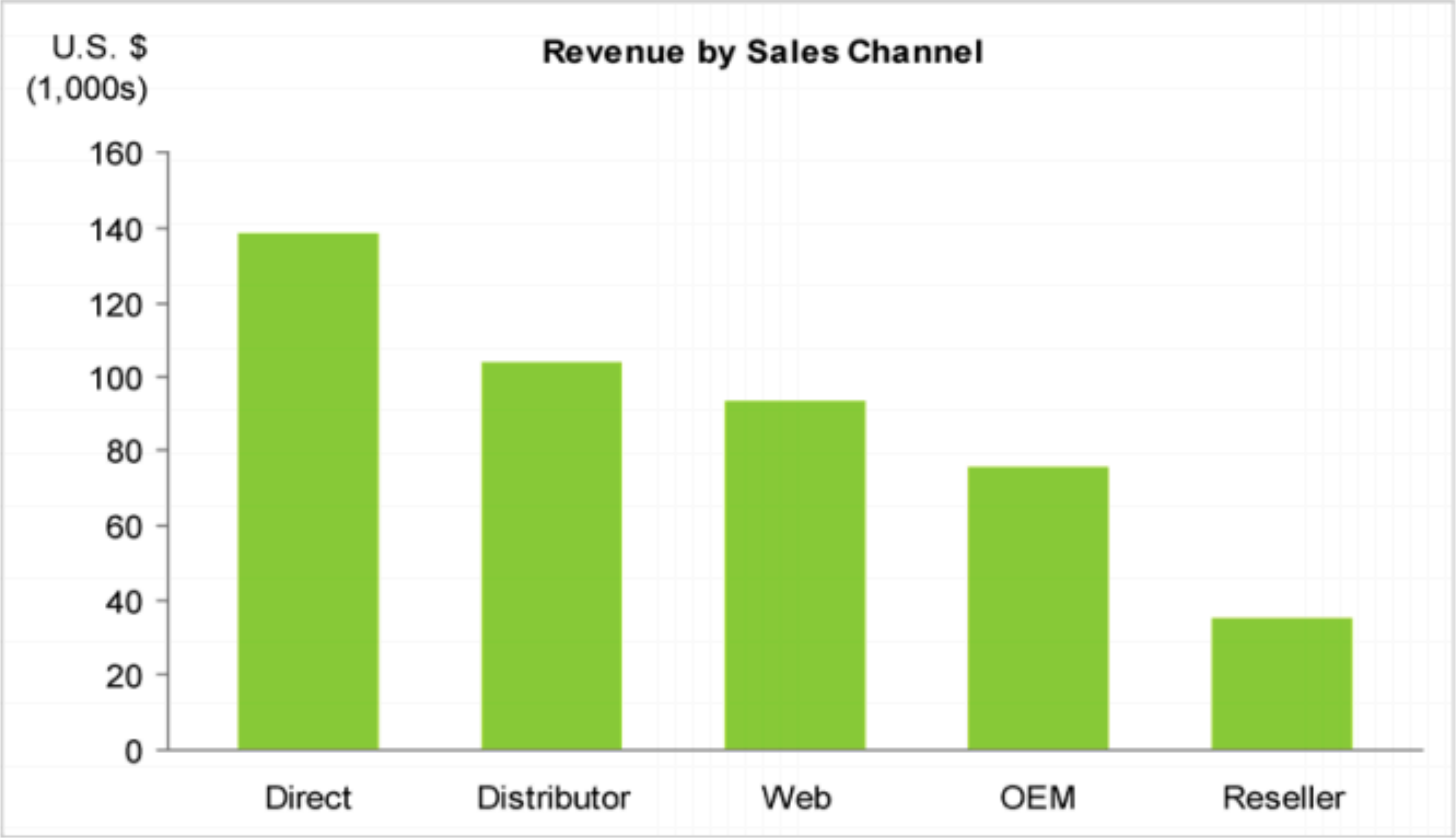


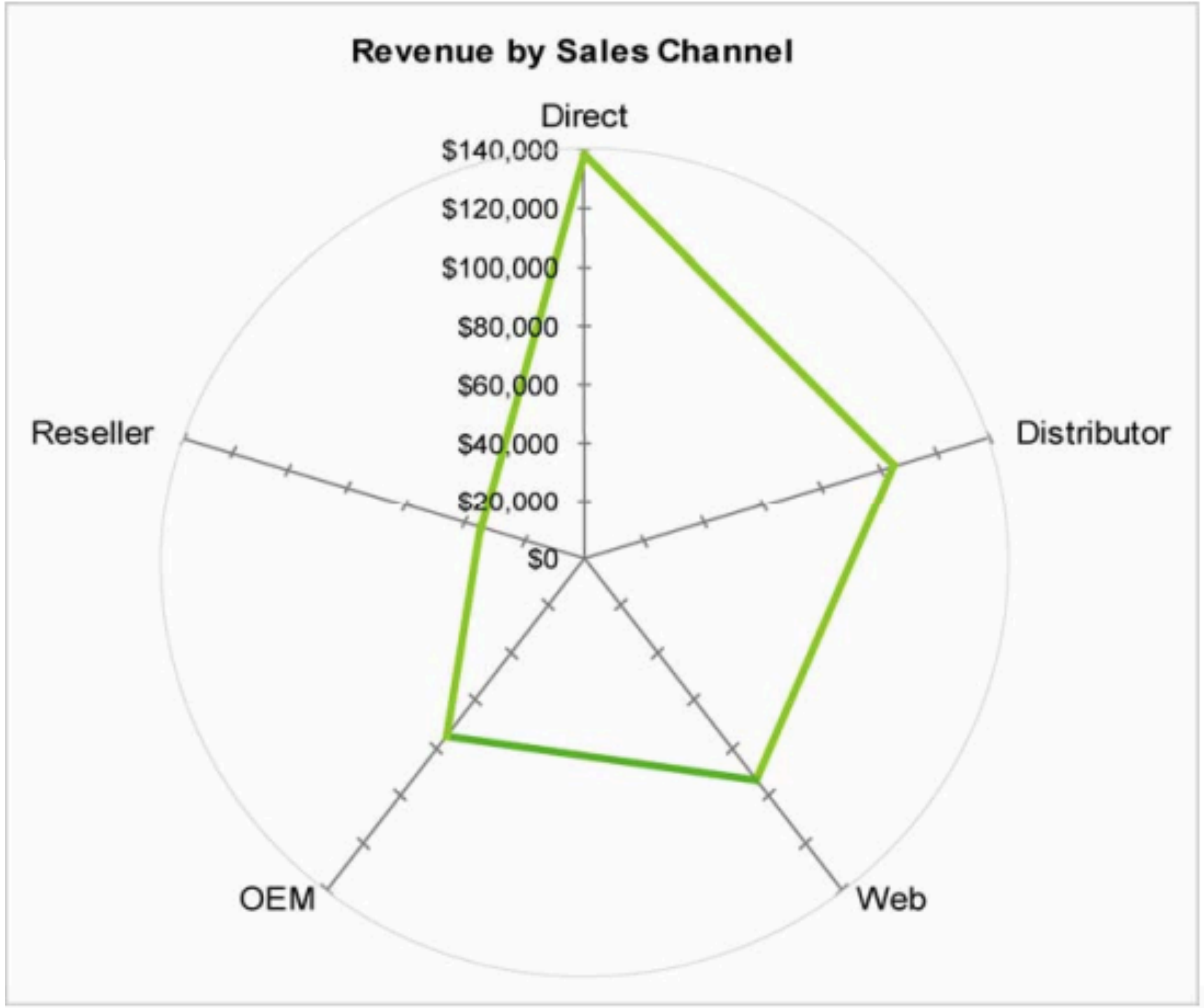


https://www.perceptualedge.com/articles/dmreview/radar_graphs.pdf

Exercise: Rating Charts for Tasks

- task D: present trends across all attributes

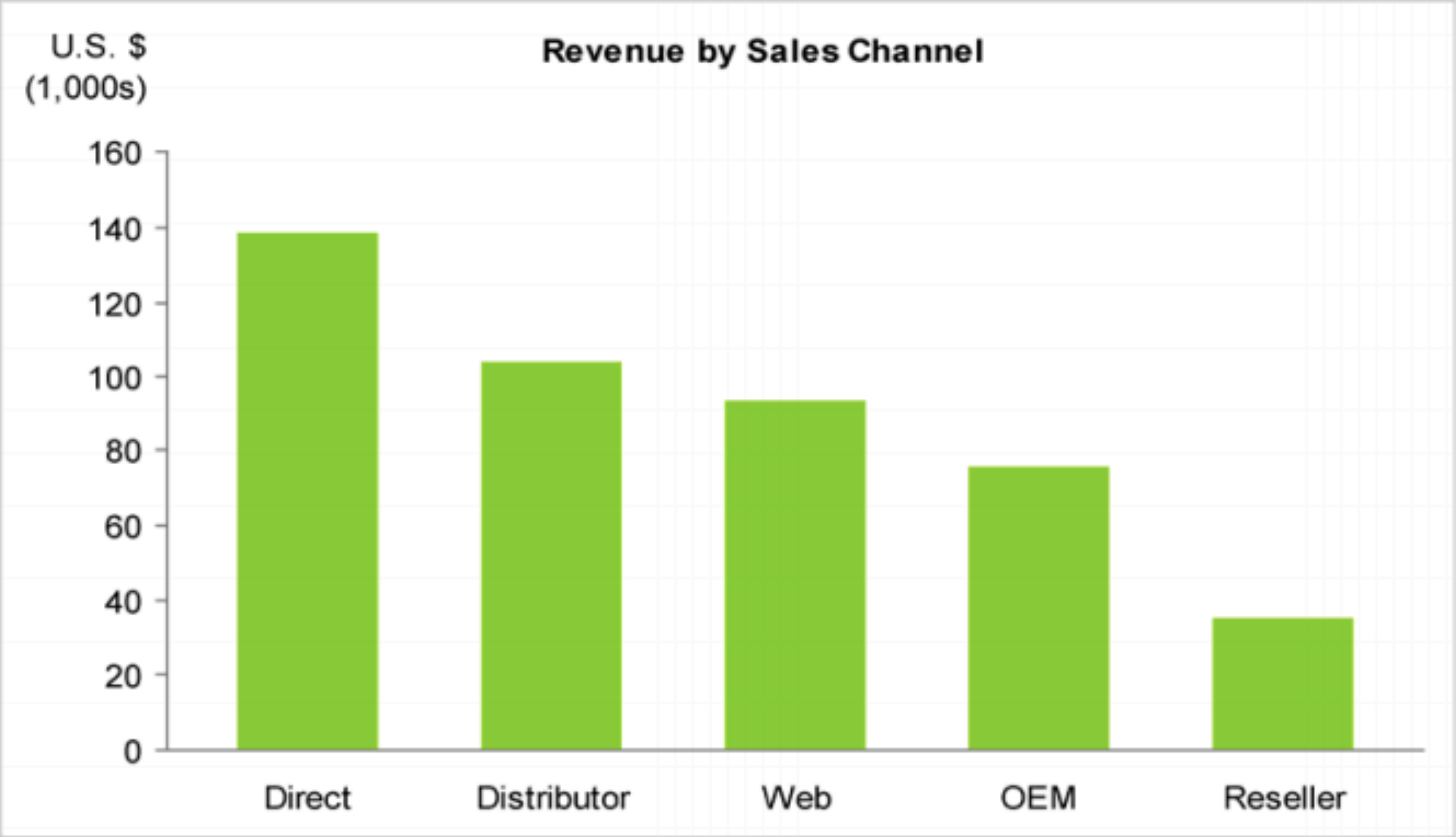


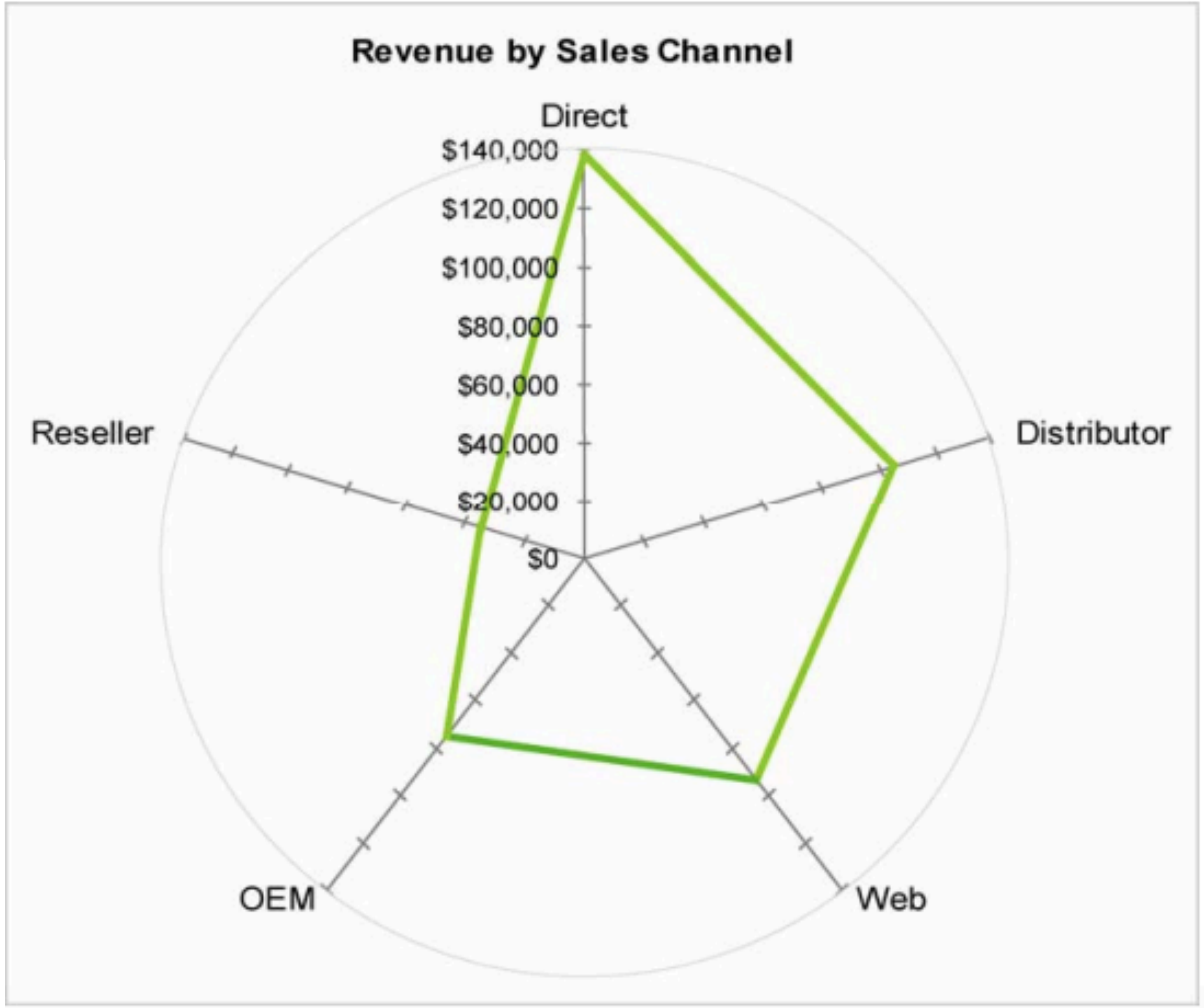


https://www.perceptualedge.com/articles/dmreview/radar_graphs.pdf

Exercise: Rating Charts for Tasks

- task E: spot outlier attributes

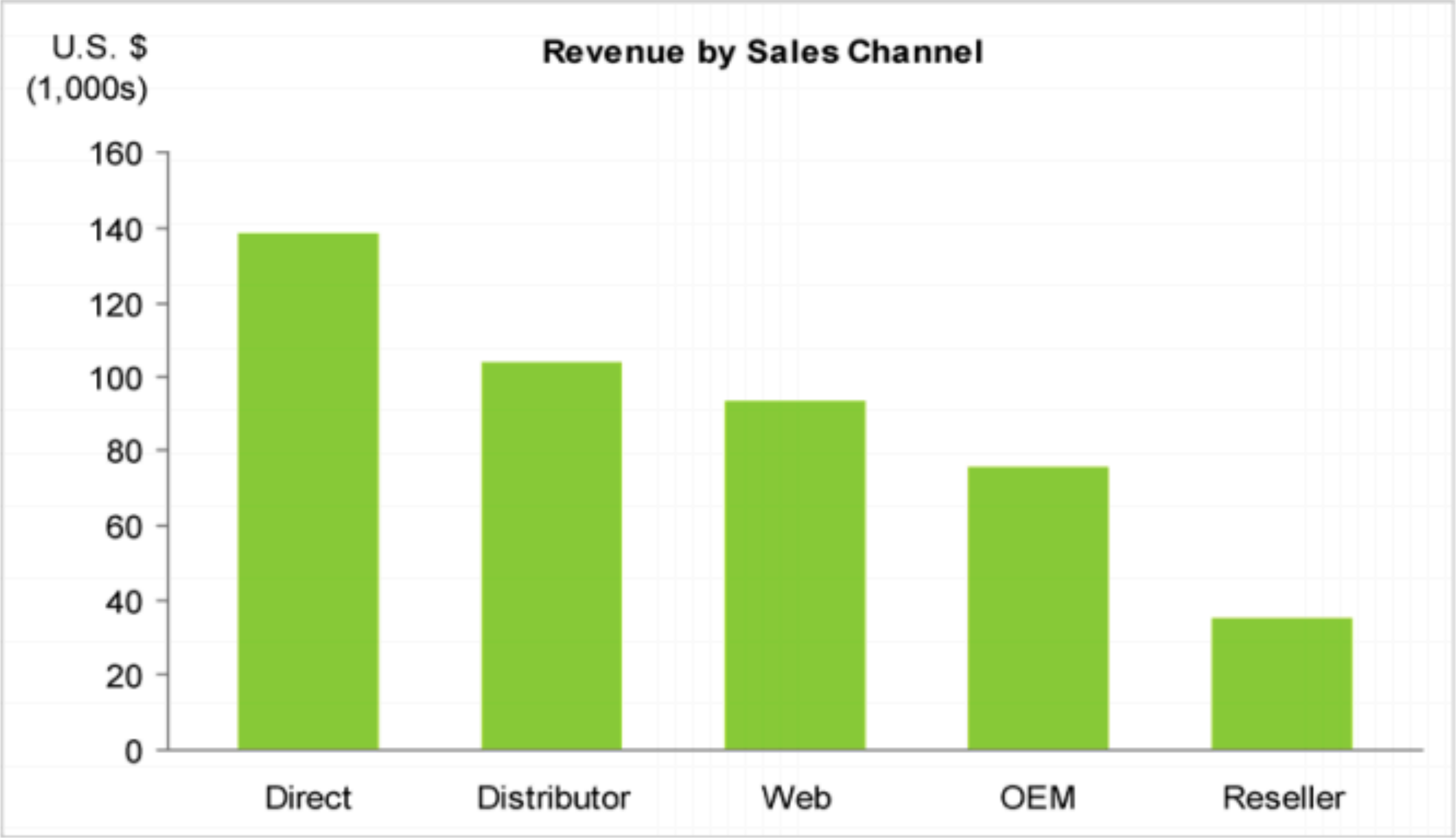


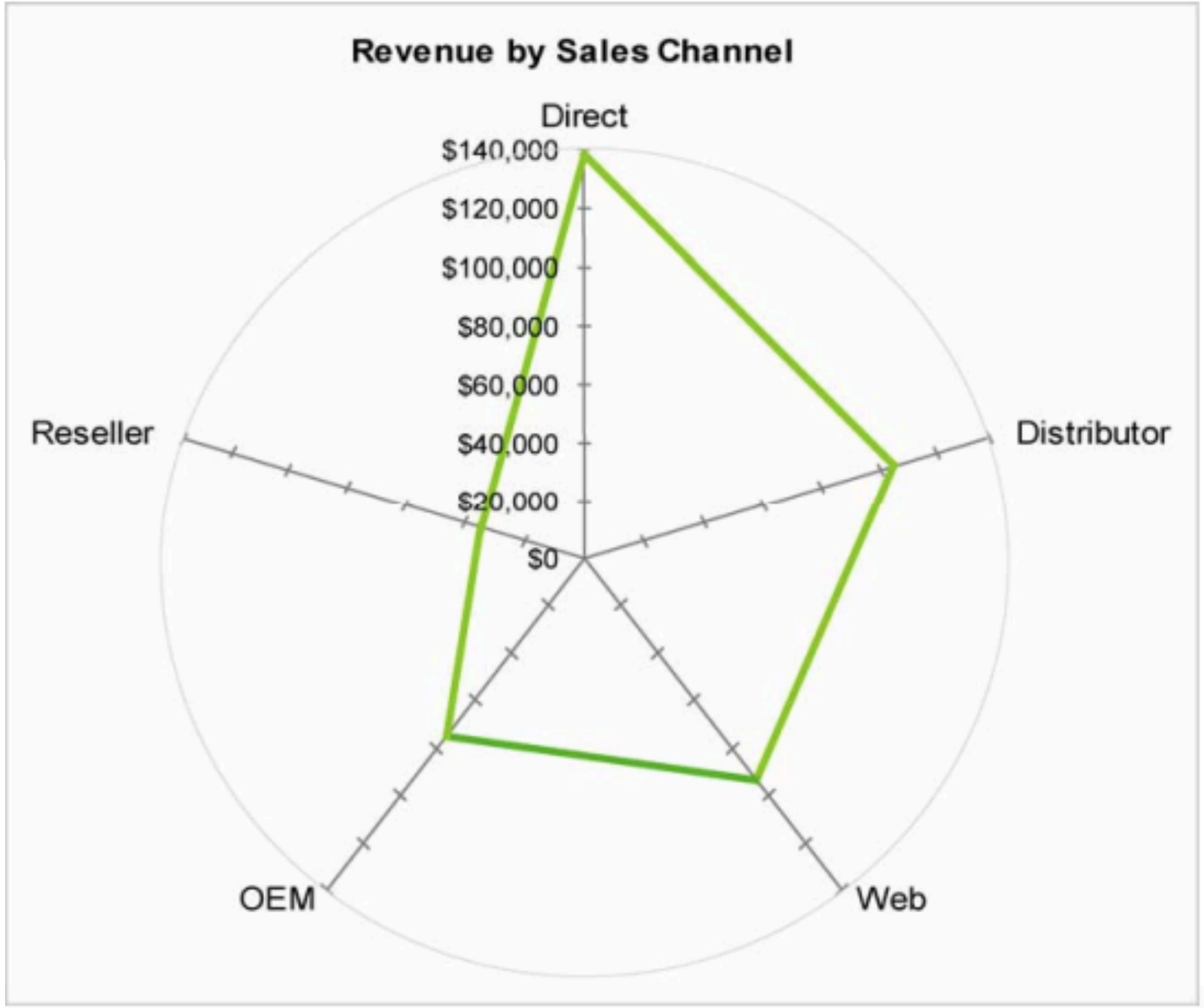


https://www.perceptualedge.com/articles/dmreview/radar_graphs.pdf

Exercise: Rating Charts for Tasks

- task F: enjoy / engage





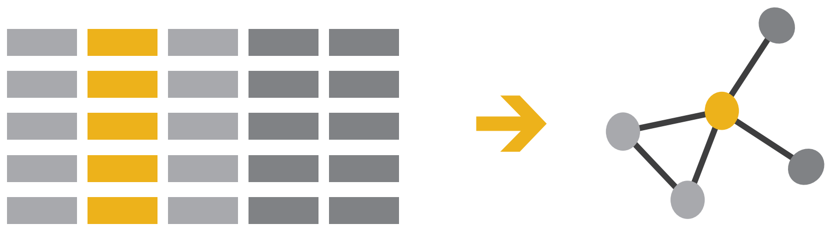
https://www.perceptualedge.com/articles/dmreview/radar_graphs.pdf

Exercise: Task abstraction in genomics I

You have been approached by a geneticists to help with a visualization problem. She has **gene expression data** (data that measures the activity of the genes) for **30 cancer tissue samples**. She is applying an experimental drug to **see whether the cancer tissue dies** as she hopes, but she finds that **only some samples show the desired effect**. She believes that the difference between the samples is caused by differential expression (**different activity) of genes in a particular pathway**, i.e., an interaction network of genes. She would like to understand **which genes are likely to cause the difference**, and **what role they play in that pathway.**

Exercise: Task abstraction in genomics I

➔ *Derive*

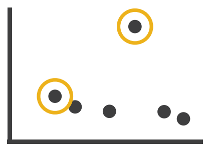


- *... only some samples show the desired effect*
  - **derive two groups of samples**

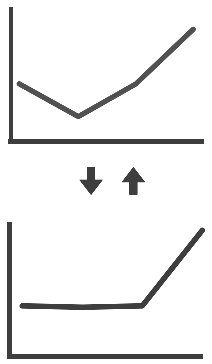
- *... the difference between the samples is caused by differential expression* ***(different activity) of genes*** *in a particular pathway. She would like to understand* ***which genes are likely to cause the difference***
  - **identify those genes**
  - **compare gene expression of pathway genes between two groups**
  - **identify the outliers**

➔ **Query**

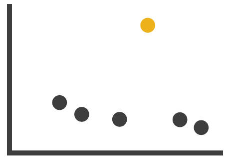
➔ Identify



➔ Compare



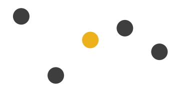
➔ Outliers



Exercise: Task abstraction in genomics, I

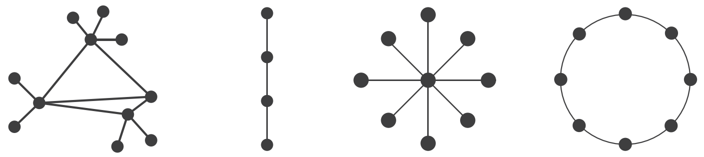
- … *which genes are likely to cause the difference, and* ***what role they play in that pathway***.
  - **locate the outlier in the network**
  - **explore the topology**

Search

	Target known
Location known	 *Lookup*
Location unknown	 *Locate*

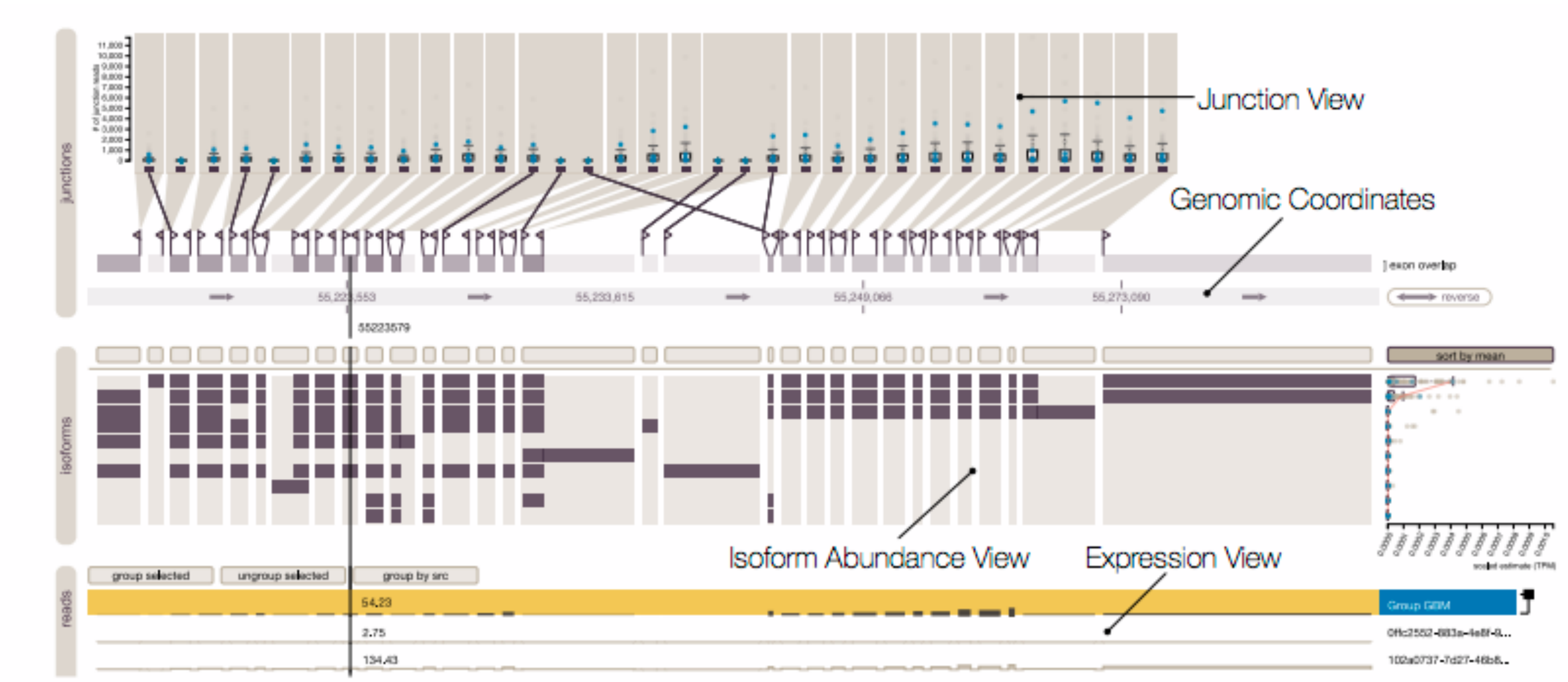
Network Data

→ Topology



Example: Genomics II

- goal: control data quality for gene splicing data
- tasks
  - judge magnitude of sample
  - compare samples, identify within-group variance & outliers
  - compare groups, identify between-group variance



3.1 Tasks

From this set of domain goals we infer two groups of tasks: those that are primarily concerned with the tabular experimental data (expression, junction support, isoform abundance; enumerated with T), and those that are concerned with the composition of isoforms (C). In the following, we describe these tasks and state the related goals.

For each of the three data types isoform abundance, exon expression, and junction support, we identify the same **tasks for the tabular experimental data** (**T**).

T1: Judge the magnitude of a sample or group (e.g., is the isoform highly expressed for a given sample?) [G1, G4]
T2: Compare samples and identify within-group variance and outliers (e.g., is the junction support different between samples?, is the junction support within a group of samples consistent?) [G1, G4]
T3: Compare groups, i.e., identify between-group variance (e.g., is an exon expressed differently between the groups?) [G1, G4]

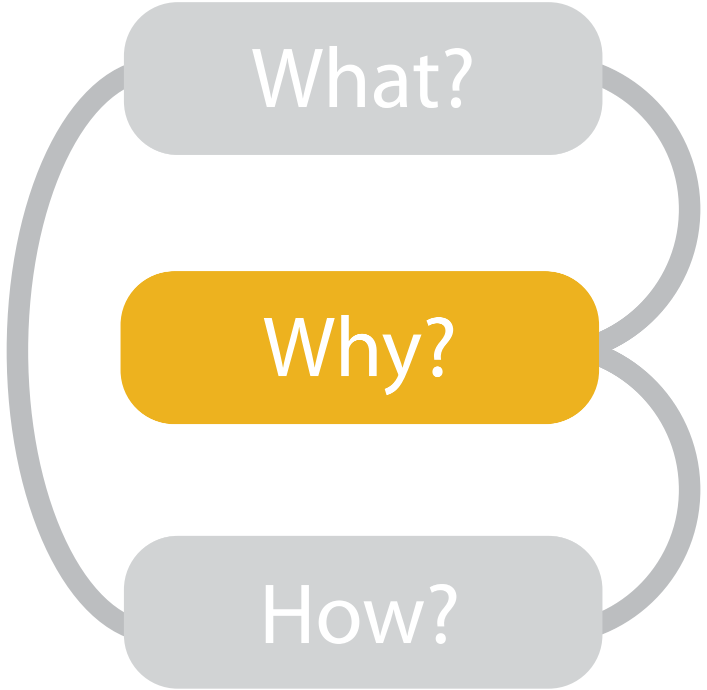
The **tasks related to the composition of isoforms** (**C**) bridge the data types. The composition tasks are:

C1: Identify the exons/junction that are part of an isoform. [G2, G3]
C2: Identify the relationships between isoforms, e.g., find out whether they include the same or similar exons. [G2, G3]
C3: Identify evidence for novel exons or isoforms that are not in the reference data. [G2]

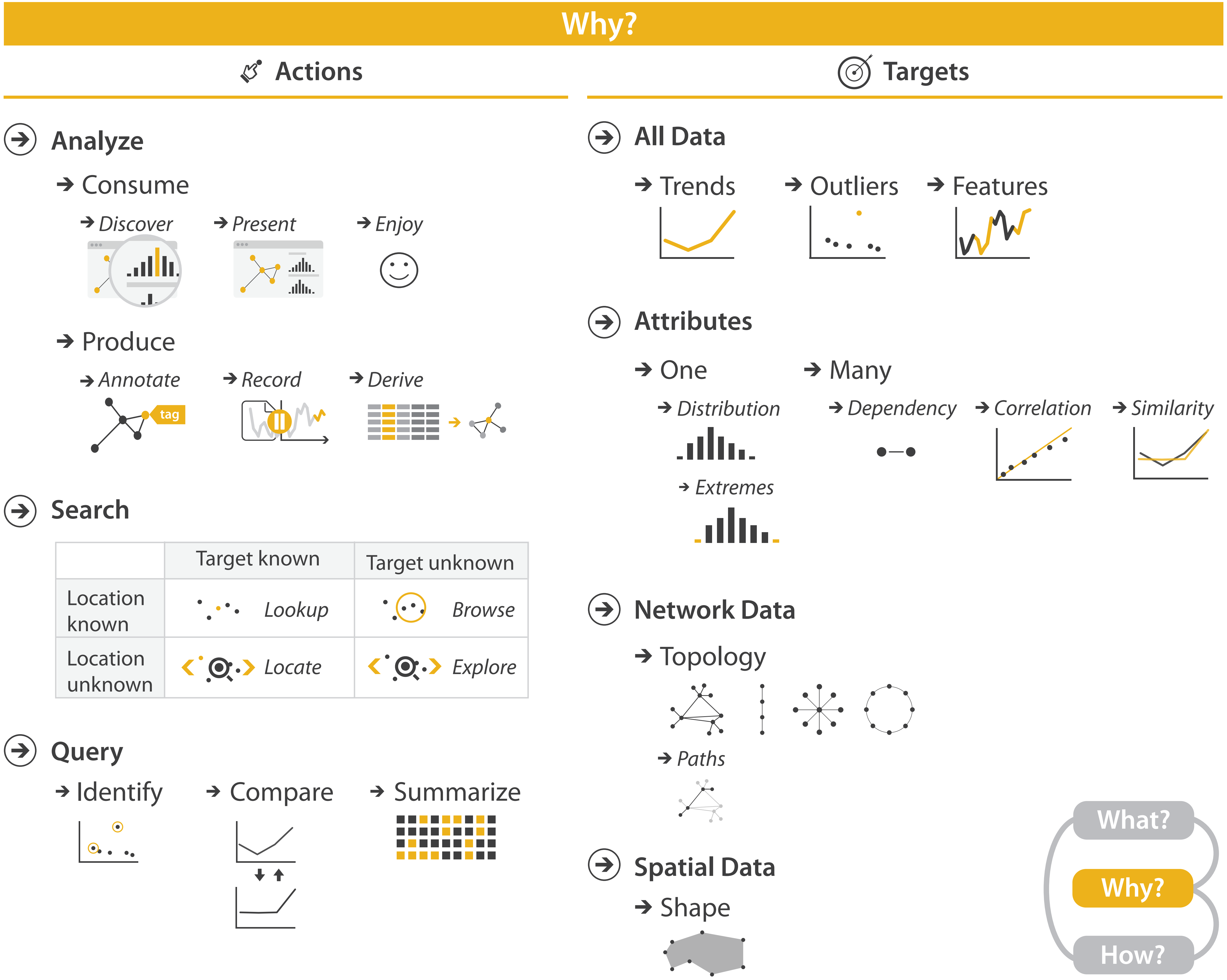
Finally, there is the supporting task of defining sample groupings, either based on user knowledge or through data (**GR**).

As is evident from this list, comparing between groupings and exploring the connections of multiple data types are critical for this type of analysis. We have designed Vials to address these tasks so that our collaborators can answer their higher-level questions.

proposed idiom
- Vials [Strobelt et al 2016]



- {action, target} pairs
  - *discover distribution*
  - *compare trends*
  - *locate outliers*
  - *browse topology*



Todo this week

- D3 videos to watch this week
  - Making a Bar Chart with D3 and SVG [30 min]
- Quiz 2 to do this week, due by Fri Jan 17, 8am
- labs start this week!
  - Fri 9-10, 11-12, 4-5
  - strongly recommended but optional: we do not track attendance
  - TA office hours for individual consultation and help
    - TAs will typically alternate weeks
  - if you can't register, try attending the one you want
    - seats for registered students first, but may be room
- Foundations Exercise 2 out next time (Thu Jan 16)
  - due Wed Jan 22
- Programming Exercise 1 out next time (Thu Jan 16)
  - due Wed Jan 29

Credits

- Visualization Analysis and Design (Ch 3)
- Alex Lex & Miriah Meyer, http://dataviscourse.net/
- Marti Hearst, exercise (tasks & charts)
  - Teaching as Coaching (VIS 2015 panel on Vis, The Next Generation)