VO Informationsvisualisierung (186.141)

Visualization Design & Evaluation

Manuela Waldner

Institute of Visual Computing & Human-Centered Technology, TU Wien, Austria

Information Visualization

"The use of computer-supported, interactive, visual representations of abstract data to amplify cognition"

[Card et al., Readings in Information Visualization: Using Vision to Think, 1999]

[http://d3js.org/]

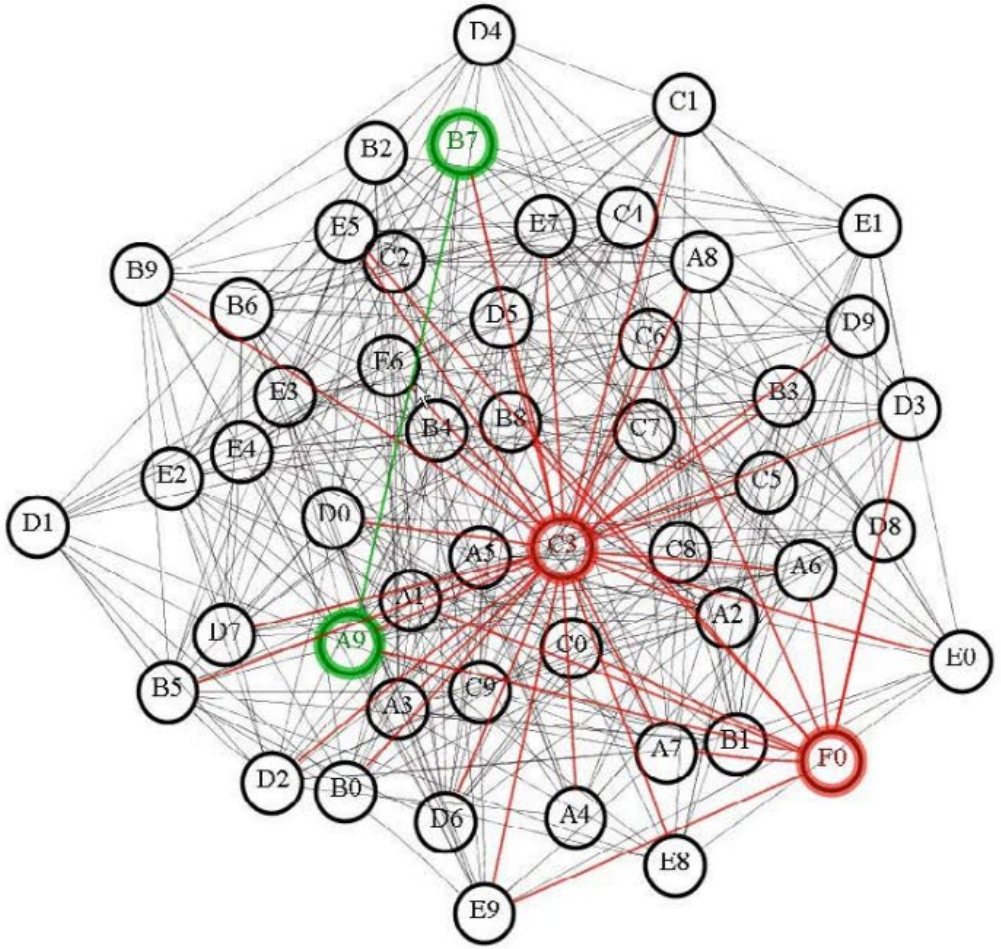
Example: Enterprise Data

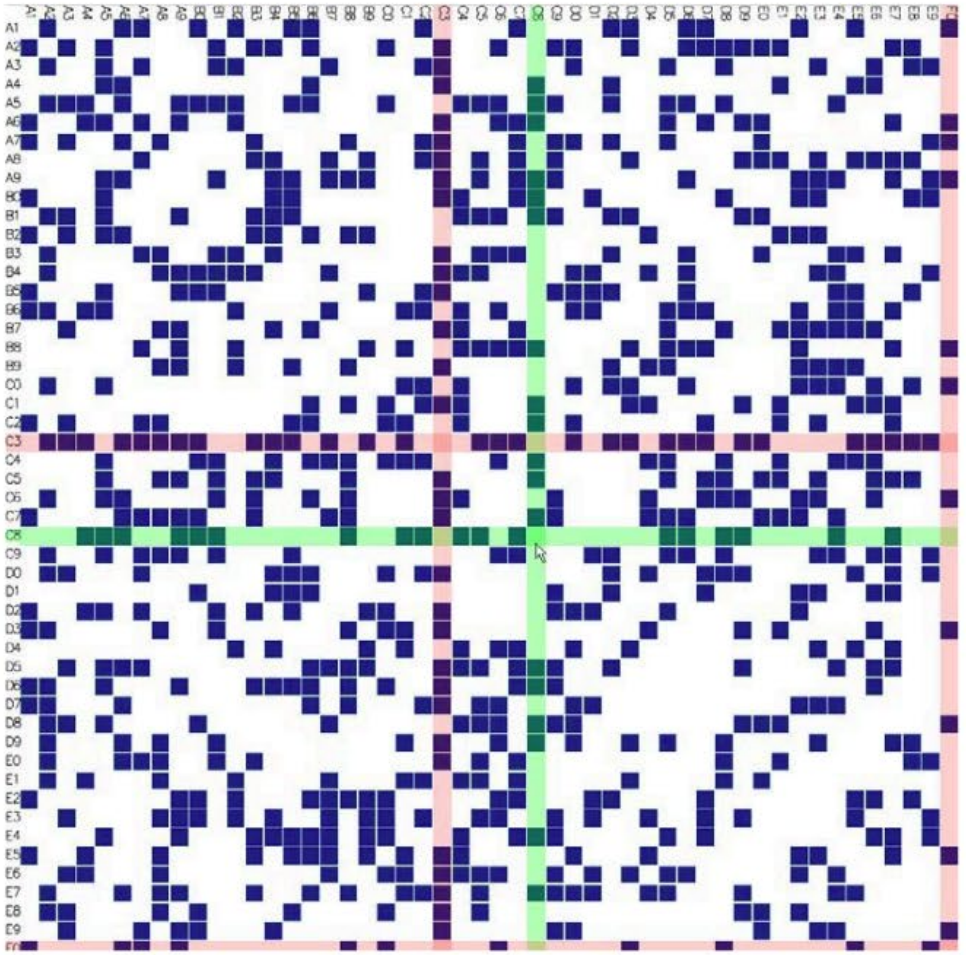
- Many organisations gather large and complex data sets
  - Modeling costumer engagement
  - Improve production
  - Inform sales and business decisions
  - ...
- **Which data analysis and visualization tools could help enterprise analysts?**

[Kandel et al., Enterprise Data Analysis and Visuaization: An Interview Study, TVCG 2012]

Example: Graphs

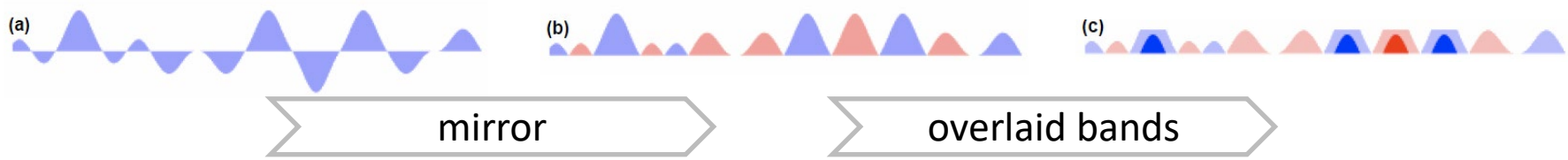
- **Which graph representation is more readable?**





[Ghoniem et al., InfoVis 2004]

Example: Horizon Graphs



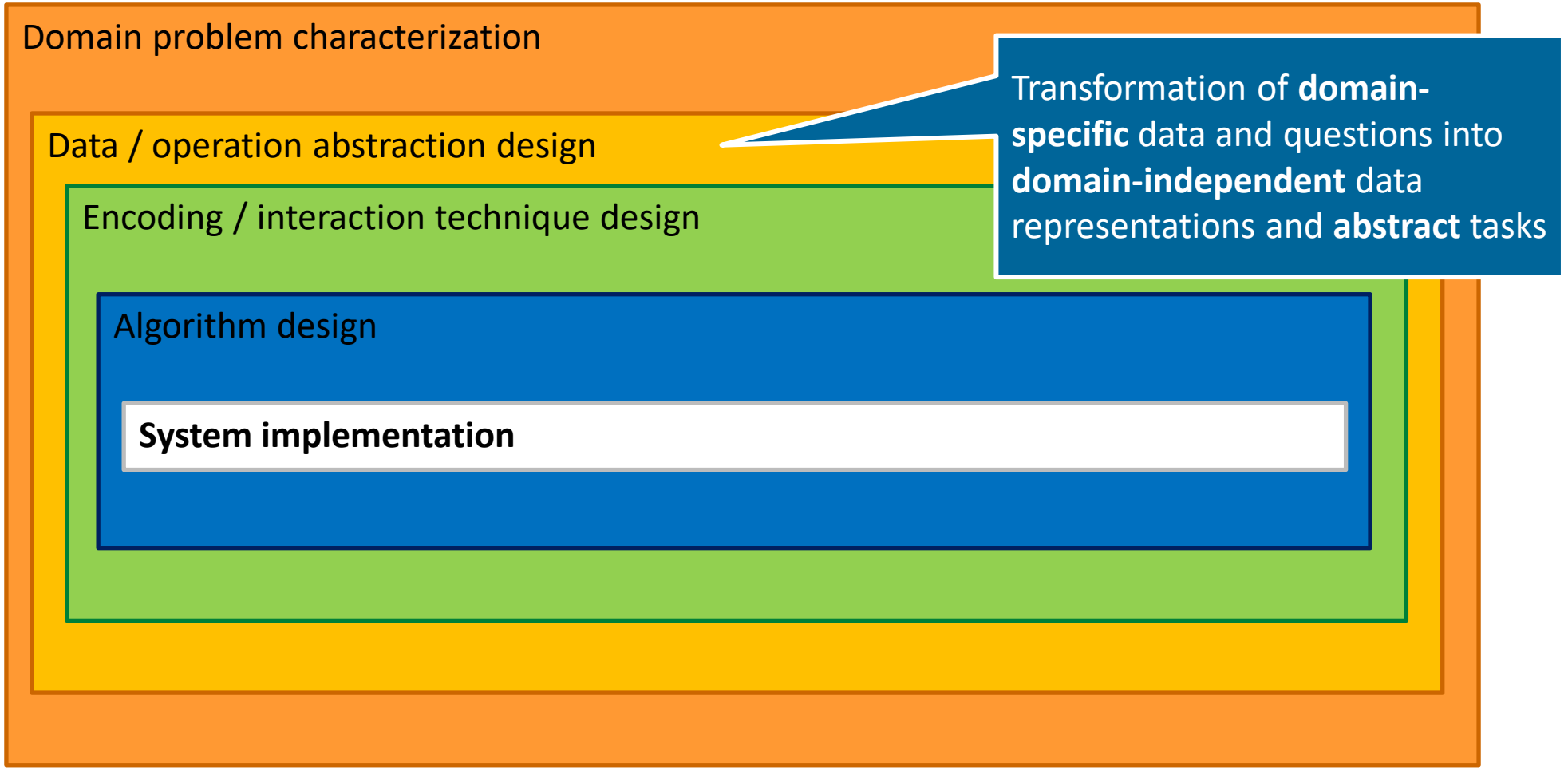
- Space-efficient time-series visualization technique
- Goal: increase data density for time series charts
- **How does this compression affect the interpretation of the time-series data?**

[Heer et al., Sizing the Horizon, CHI 2009]

Example: Thesis

- You: *“I want to visualize this really large and complex data. And here is the super fancy visualization I came up with”*
- Supervisor: *“How will you validate your proposed solution?”*
- You:

Nested Model for Visualization Design



[Munzner, A Nested Model for Visualization Design and Validation, InfoVis 2009]

Example: Media Transparency Database Austria

Public database of advertisement expenses of public authorities in Austria

Meldungen gemäß MedKF-TG für das 3. Quartal 2018

Friends	10.489,50
frisch gekocht	19.278,00
Gourmetro	5.500,00
infoscreen	9.475,00
Kleine Zeitung	65.963,38
Kurier	15.400,00
Meiningers Weinwelt	5.104,44
Meiningers Weinwirtschaft	5.107,44
OÖ Nachrichten	7.274,70
ORF 2	23.250,00
Österreichische BauernZeitung - Gesamtausgabe	9.950,00
schaumagazin	5.500,00
VINARIA	9.000,00
Wein+Markt	6.213,50
Weinseller Journal	5.500,00
weinwelt.at - Magazin	20.000,00
www.martinahohenlohe.com	6.000,00
Bekanntgabe §4 (Förderungen)	
Leermeldung	
Österreich Werbung	
Bekanntgabe §2 (Werbeaufträge und Medienkooperationen)	
www.austrian.com	17.575,00
Bekanntgabe §4 (Förderungen)	
Leermeldung	
Österreichische Agentur für Gesundheit und Ernährungssicherheit GmbH	
Bekanntgabe §2 (Werbeaufträge und Medienkooperationen)	
Leermeldung	
Bekanntgabe §4 (Förderungen)	
Leermeldung	
Österreichische Akademie der Wissenschaften	
Bekanntgabe §2 (Werbeaufträge und Medienkooperationen)	
Der Standard	9.000,00
Die Presse	7.041,71
Bekanntgabe §4 (Förderungen)	
Leermeldung	
Österreichische Akademie für ärztliche und pflegerische Begutachtung (ÖBAK)	
Bekanntgabe §2 (Werbeaufträge und Medienkooperationen)	
Leermeldung	
Bekanntgabe §4 (Förderungen)	
Leermeldung	
Österreichische Apothekerkammer	
Bekanntgabe §2 (Werbeaufträge und Medienkooperationen)	
Die Apotheke	7.800,00
Gesund + Leben in Niederösterreich	8.000,00
Kleine Zeitung	9.323,00
Kronen Zeitung	9.190,61
Kurier	16.086,60
ORF 2	68.893,48
ORF Radio Niederösterreich	6.075,00
ORF Radio Oberösterreich	5.075,40

What: Attribute and Dataset Types

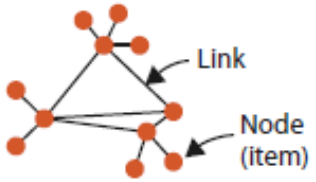
Attribute Types

quantitative

categorical

Dataset Types

➔ Networks



Meldungen gemäß MedKF-TG für das 3. Quartal 2018

Friends	10.489,50
frisch gekocht	19.278,00
Gourmetro	5.500,00
infoscreen	9.475,00
Kleine Zeitung	65.963,38
Kurier	15.400,00
Meiningers Weinwelt	5.104,44
Meiningers Weinwirtschaft	5.107,44
OÖ Nachrichten	7.274,70
ORF 2	23.250,00
Österreichische BauernZeitung - Gesamtausgabe	9.950,00
schaumagazin	5.500,00
VINARIA	9.000,00
Wein+Markt	6.213,50
Weinseller Journal	5.500,00
weinwelt.at - Magazin	20.000,00
www.martinahohenlohe.com	6.000,00
Bekanntgabe §4 (Förderungen)	
Leermeldung	
Österreich Werbung	
	17.575,00
	9.000,00
	7.041,71
Österreichische Akademie für ärztliche und pflegerische Begutachtung (ÖBAK)	
Bekanntgabe §2 (Werbeaufträge und Medienkooperationen)	
Leermeldung	
Bekanntgabe §4 (Förderungen)	
Leermeldung	
Österreichische Apothekerkammer	
Bekanntgabe §2 (Werbeaufträge und Medienkooperationen)	
Die Apotheke	7.800,00
Gesund + Leben in Niederösterreich	8.000,00
Kleine Zeitung	9.323,00
Kronen Zeitung	9.190,61
Kurier	16.086,60
ORF 2	68.893,48
ORF Radio Niederösterreich	6.075,00
ORF Radio Oberösterreich	5.075,40

Amount of money spent on advertisement by a public authority for a specific media

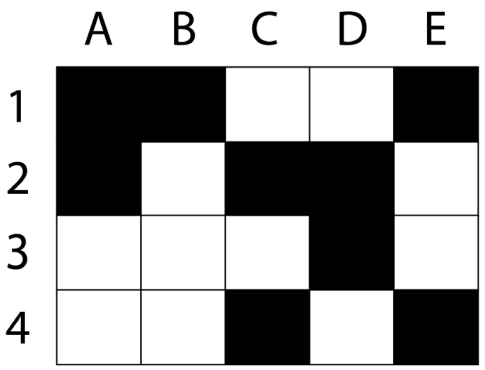
Bipartite graph: a graph where the vertices can be divided into two disjoint sets such that all edges connect a vertex in one set to a vertex in another set

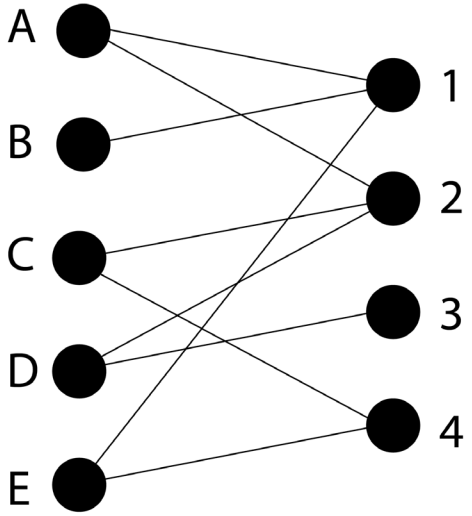
Category 1: public authorities

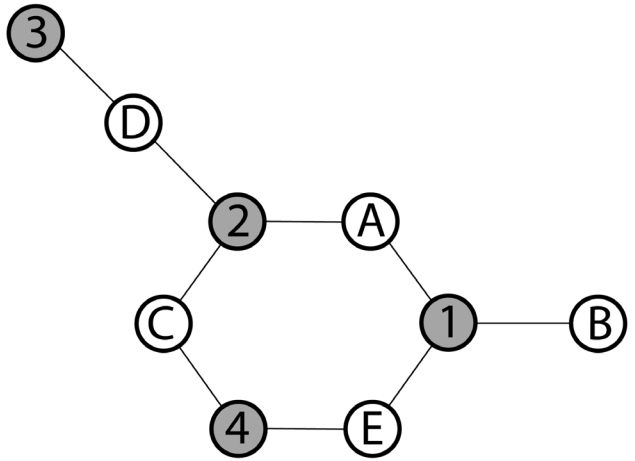
Category 2: media

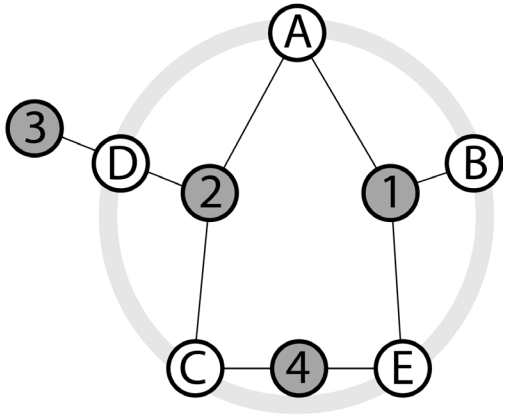
How: Bipartite Graphs

- Known space of visual encodings for bipartite graph:





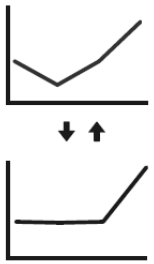




[Waldner et al., 2020]

Why: Low-Level Actions

➔ Identify ➔ Compare ➔ Summarize



Which public authorities spend most money for advertisement?

Which public authorities tend to advertise in similar types of media?

Any obvious clusters or patterns?

MEDIENTRANSPARENZ

Öffentliche Stellen warben im zw um rund 50 Millionen Euro

Ministerien gaben mehr aus als im Vorjahr, der Großteil geht an den Boulevard

Philip Pramer 13. September 2019, 15:30 5 Postings

Ministerien, Länder, Gemeinden und andere öffentliche Stellen und Betriebe haben im zweiten Quartal 2019 49,3 Millionen Euro für Werbung ausgegeben. Das zeigen die Medientransparenzdaten, die am Freitag von der Medienbehörde veröffentlicht wurden. Die Ausgaben waren zwar höher als im Vergleichsquartal des Vorjahres (46,6 Millionen), aber unter dem Höchststand im zweiten Quartal 2013. Damals ließ sich die öffentliche Hand Inserate 58,9 Millionen Euro kosten.

Der österreichische Staat gibt vergleichsweise viel für Werbur

Foto: Standard/Philip Pramer

Medientransparenz-Meldungen

Gesamte Werbeausgaben öffentlicher und staatsnaher Stellen nach Quartal der Meldung. Angaben in Euro.

60.000.000
50.000.000
Werbeausgaben

sen 29.000 Euro, *Alles roger?* konnte sich über
Straches Sportministerium freuen.

Welche Medien am meisten von Regierungswerbung profitieren

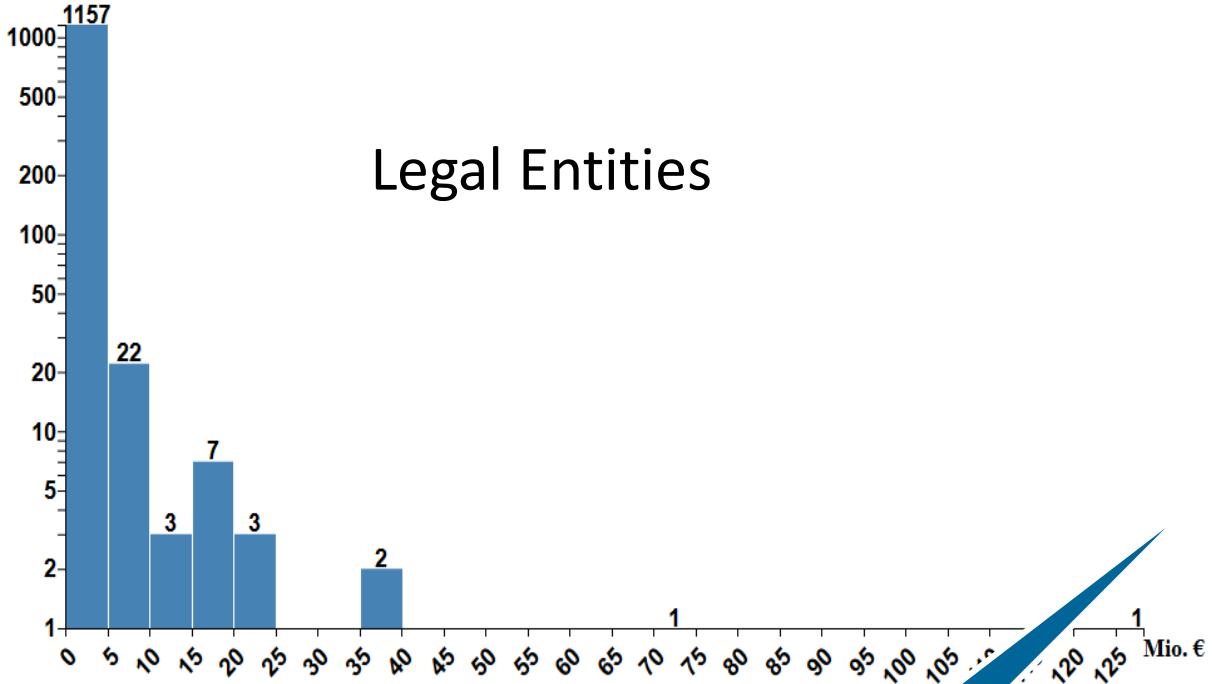
Inserate von Ministerien | von Ländern | sonstige staatsnahe Stellen/Unternehmen

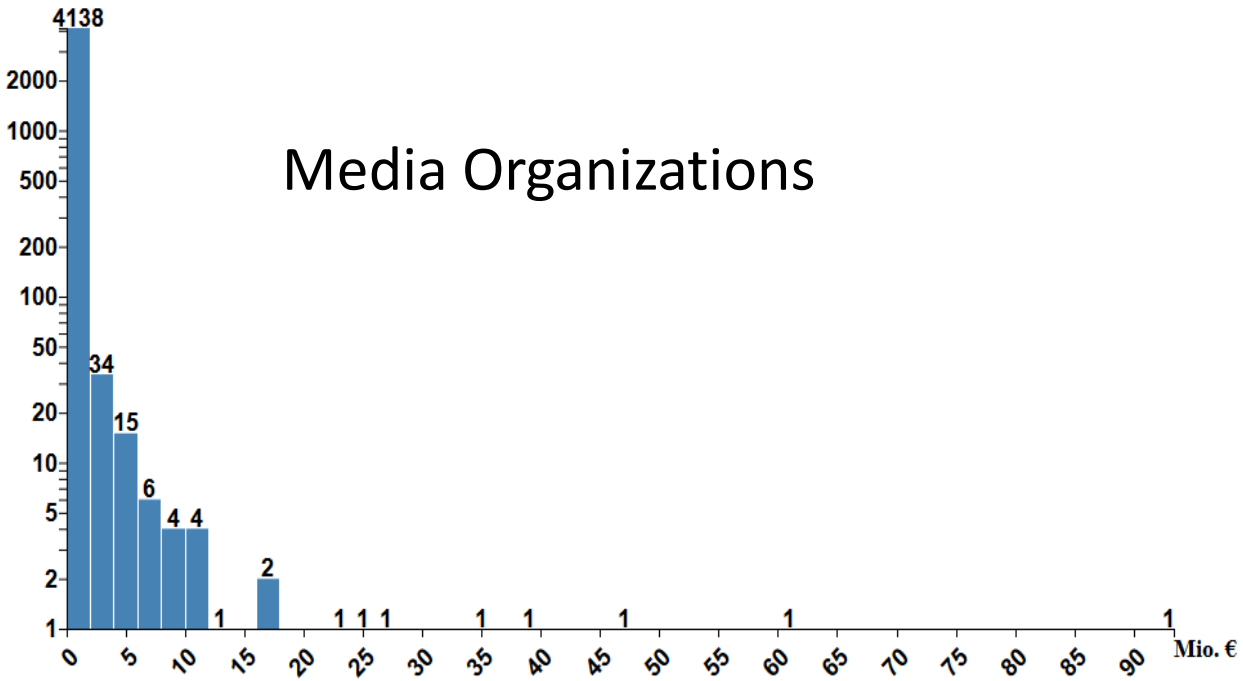
ORF			17.875.615,89
Kronen Zeitung	4.646.732,94	3.354.044	10.793.609,33
Heute	3.499.927,13		6.452.829,58
Österreich	3.638.391,51		5.562.326,87
Kurier			4.935.787,93
Google			6.027.231,41
Kleine Zeitung			4.325.799,67
DER STANDARD			4.065.850,13
RMA			3.675.763,11
Die Presse			

Der Standard

What: Special Challenges

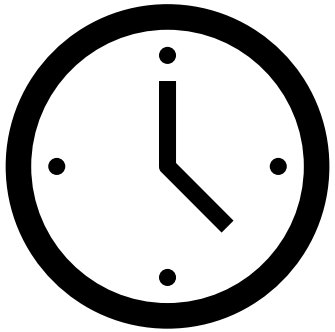
- Challenge 1: thousands of legal entities and media organizations





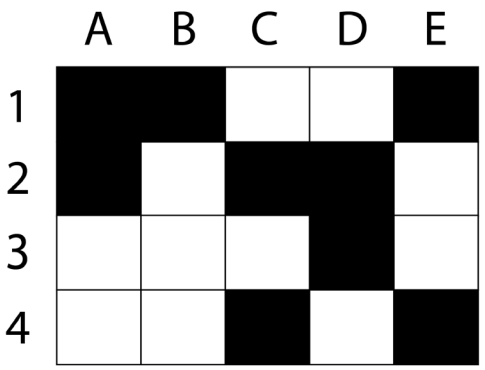
What: Special Challenges

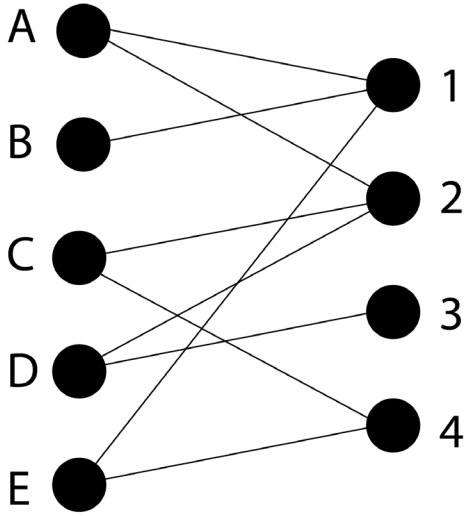
- Challenge 2: time-oriented

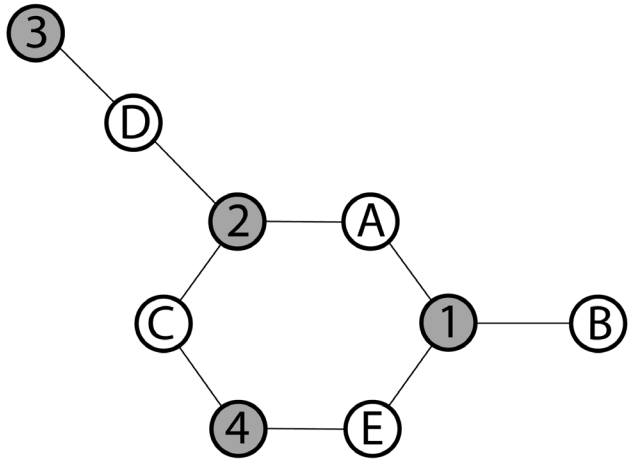


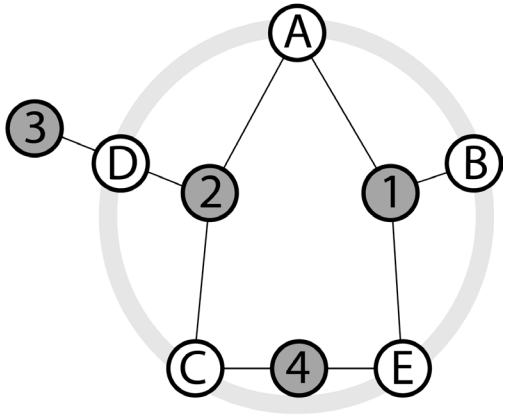
How: Bipartite Graphs

- Thousands of nodes + time-dependency?



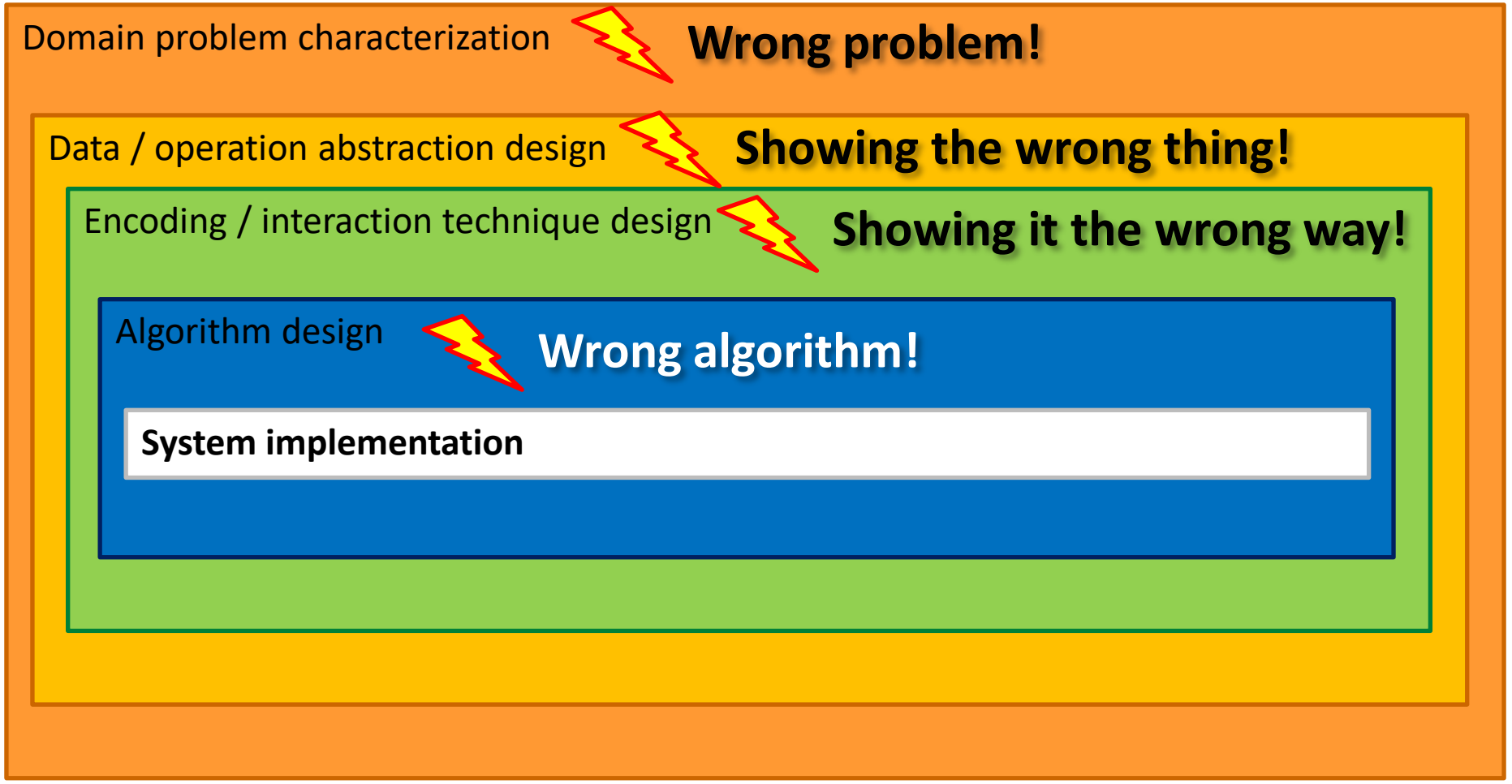






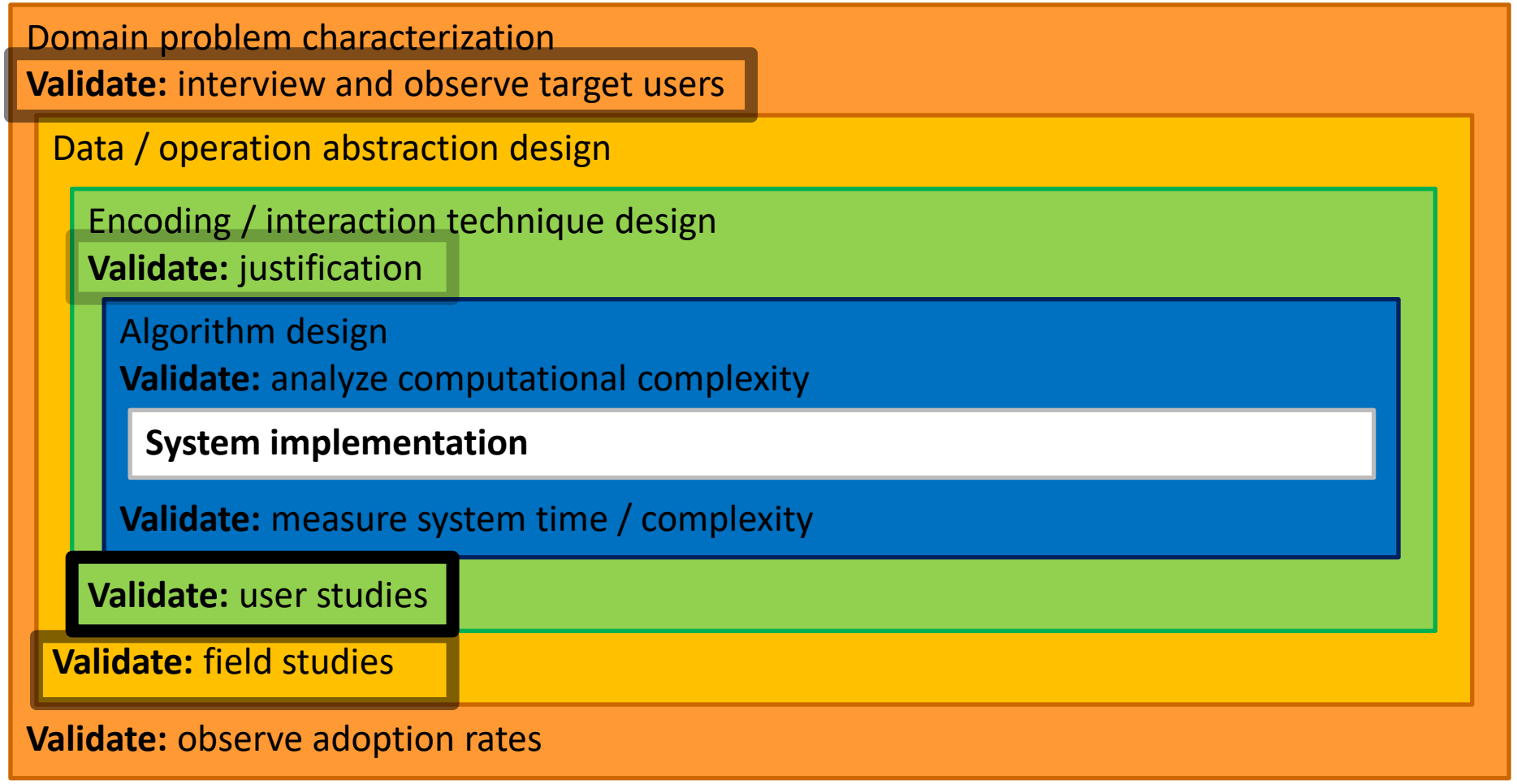
[Waldner et al., 2020]

Nested Model for Visualization Design



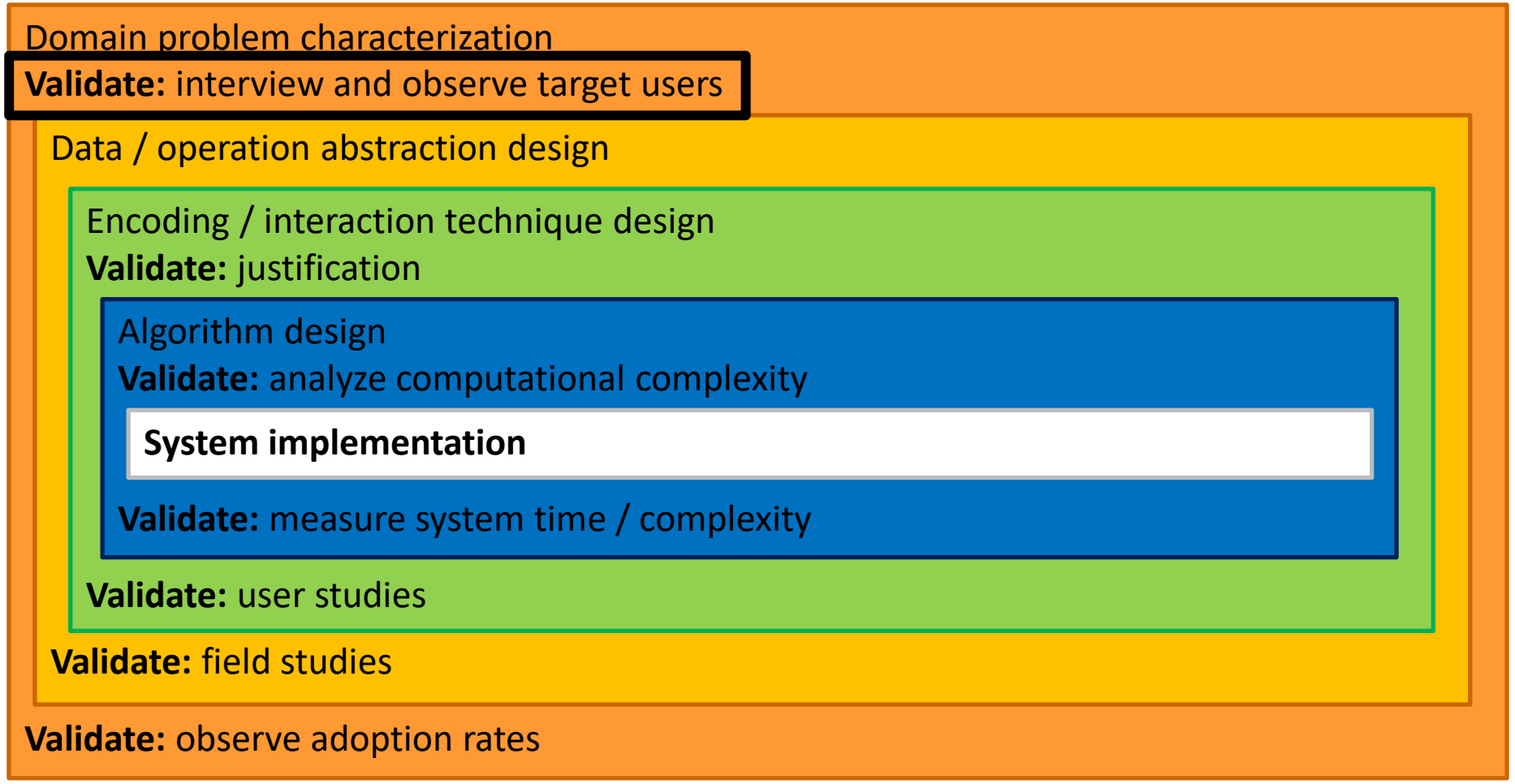
[Munzner, A Nested Model for Visualization Design and Validation, InfoVis 2009]

Nested Model for Visualization Design



[Munzner, A Nested Model for Visualization Design and Validation, InfoVis 2009]

Nested Model for Visualization Design



[Munzner, A Nested Model for Visualization Design and Validation, InfoVis 2009]

Method: Interviews

- Semi-structured interview
  - Some key questions → general area to be explored
  - Freedom to diverge
- Recorded and transcribed
- Coding to analyze data

Method: Coding

- Iterative characterization of qualitative data (interview transcript, video, field notes, …)
- Open coding:
  - Labeling of each observed phenomenon = concept
  - Grouping of concepts into categories
- Quality assurance:
  - Development of a code book
  - Multiple independent coders
    - Inter-rater reliability (degree of agreement between coders)
    - Iteratively trying to reach a certain reliability

Method: Code Book

Feature	Description	Example tweet
Sentiment		
Positive	The tweet contains supportive messages about the HPV vaccine and encourages its uptake	1. Not only does the HPV vaccine protect against human papillomavirus, but it also reduces the risk of cancers 2. #HPV vaccine can be #cancer prevention! Parents, #vaccinate your children at ages 11-12
Negative	The tweet contains disparaging messages about the HPV vaccine or discourages its uptake	1. Healthy 12-year-old girl dies shortly after receiving HPV vaccine 2. RT[a] @CBCHealth: The Gardasil Girls: How Toronto Star story on young women hurt public trust in vaccine http://t.co/...
Neutral	The tweet's text holds no subjective opinions about the vaccine—purely facts repeated from sources	1. State officials unveil campaign for HPV vaccination http://t.co/0l2sAWGXYs 2. RT @DrJenGunter: About 10% boys have received 3 doses HPV vax
No mention	The tweet does not mention the HPV vaccine	1. RT @Forbes: HPV is truly indiscriminate 2. RT @CDCSTD: #Women: get screened & talk w/ your friends about the link between #HPV & cervical #cancer
Side effects	The tweet refers to side effects caused by the HPV vaccine or effects that may be unknown to the user	1. Healthy 12-year-old girl dies shortly after receiving HPV vaccine 2. RT @ksbrowneyedgirl: It can happen to your child...to your family...#OneLess #Gardasil #CDCwhistleblower #vaccine...
Prevention/protection	The tweet refers to the extent to which the HPV vaccine will protect the user	1. Single HPV jab could prevent 70% of cervical cancers (http://t.co/Hg0KSlIk2A) 2. A new HPV vaccine prevents nine strains of the virus http://t.co/ZFGvVqlq0U

[Massey et al., Applying Multiple Data Collection Tools to Quantify Human Papillomavirus Vaccine Communication on Twitter, 2016]

Example: Enterprise Data Analysis

- Semi-structured interviews
  - 35 analysts
  - 25 organizations
  - 15 organization sectors like healthcare, finance...
  - 45 minutes – 2 hours per interview
  - At workplace or via phone / Skype
- Coding of interview data
  - Dimensions: practices, tools, challenges, organizational issues

"For **17 analysts**, finding relevant data distributed across multiple databases, database tables and/or files was very time consuming"

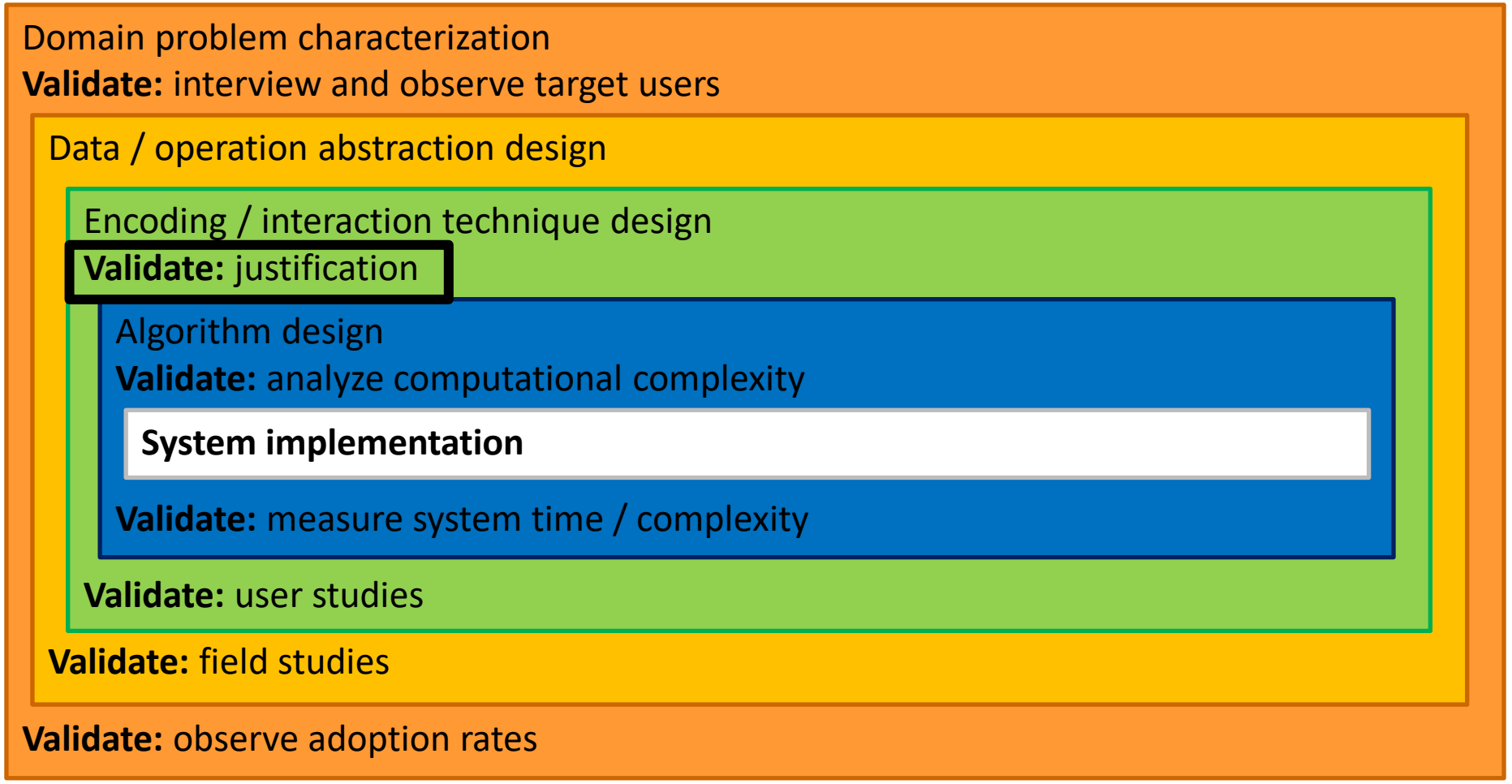
[Kandel et al., Enterprise Data Analysis and Visuaization: An Interview Study, TVCG 2012]

Example: Enterprise Data Analysis

- Findings
  - Identified high-level tasks: discovery, wrangling, profiling, modeling, reporting
    - Data discovery and wrangling most tedious and time-consuming aspects
    - Data discovery and wrangling underserved by existing visualization and analysis tools
  - Identified problems:
    - Integrating data from distributed data sources
    - Scalability issues for large data

[Kandel et al., Enterprise Data Analysis and Visuaization: An Interview Study, TVCG 2012]

Nested Model for Visualization Design



[Munzner, A Nested Model for Visualization Design and Validation, InfoVis 2009]

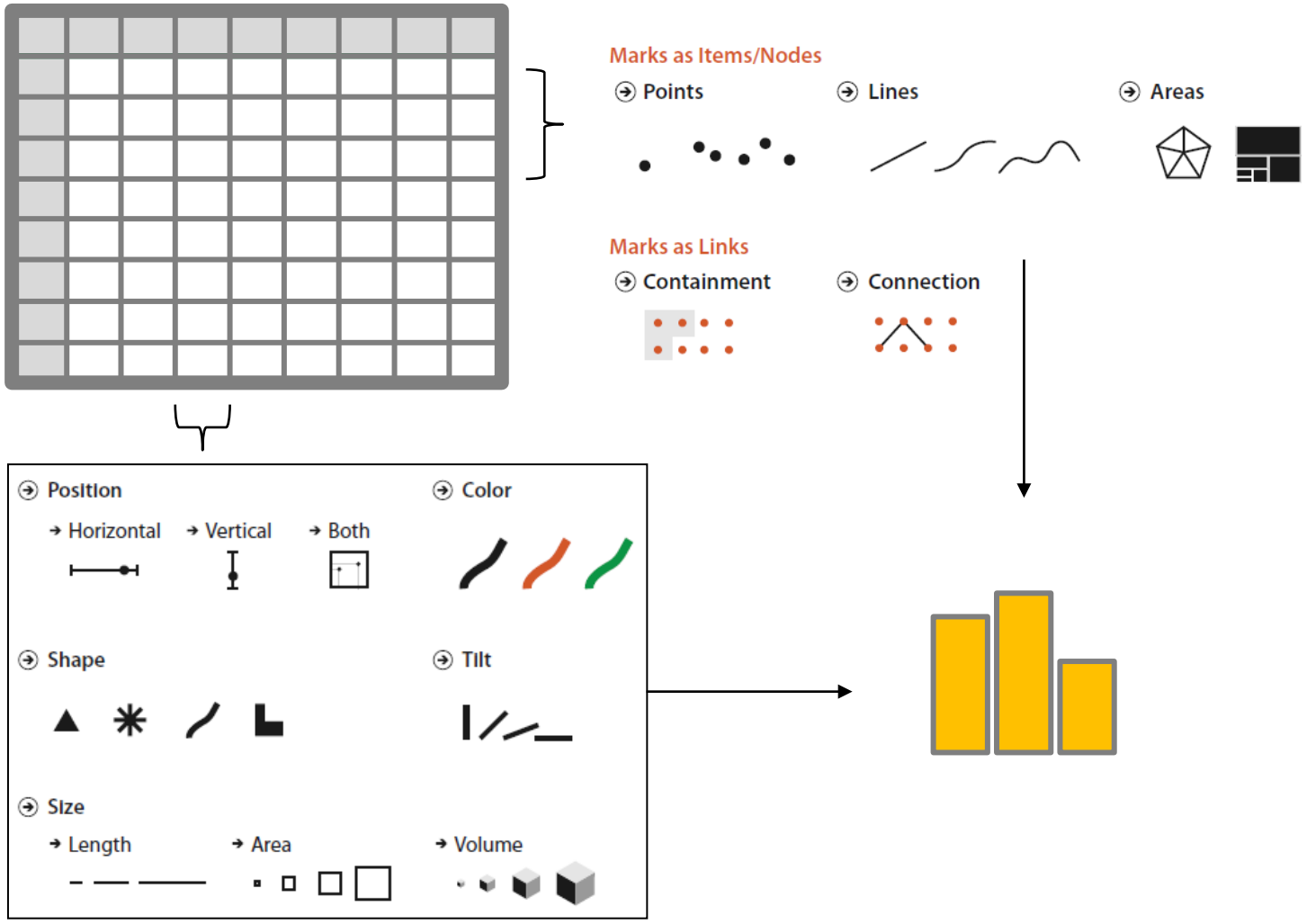
Visualization Design

"Our definition of **design** is the creative process of searching through a vast space of possibilities to select one of many possible good choices from the backdrop of the far larger set of bad choices.

Successful design typically requires the explicit consideration of multiple alternatives and a thorough knowledge of the space of possibilities."

[Sedlmair et al., Design Study Methodology, TVCG 2012]

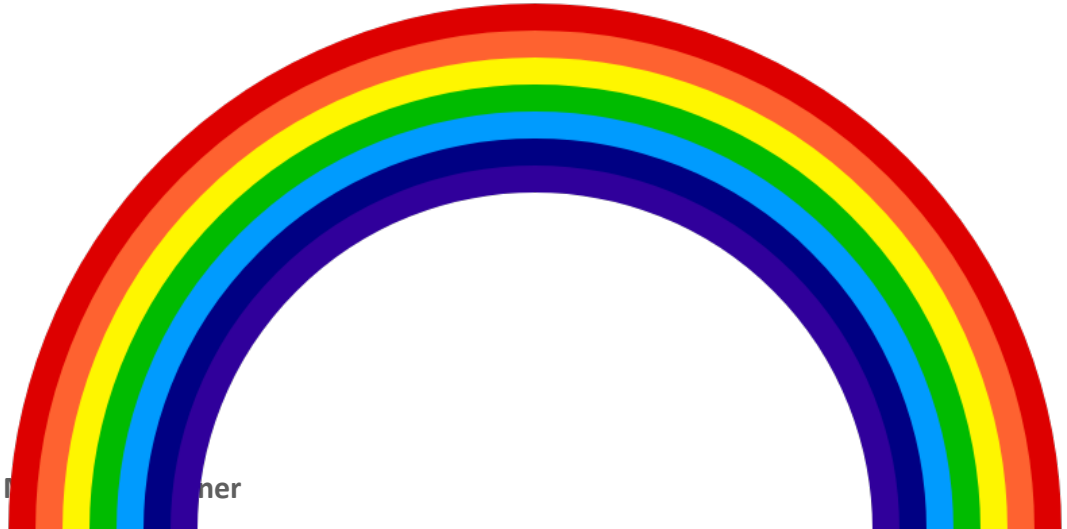
How: Marks and Channels

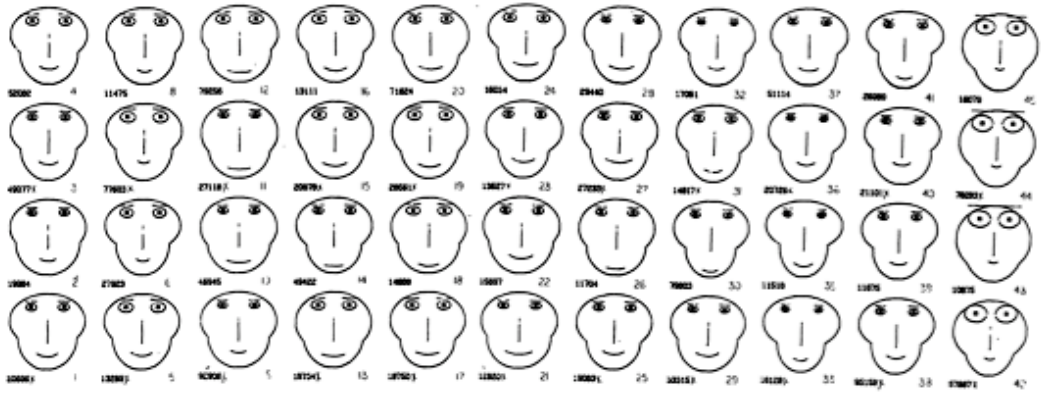


Big design space!

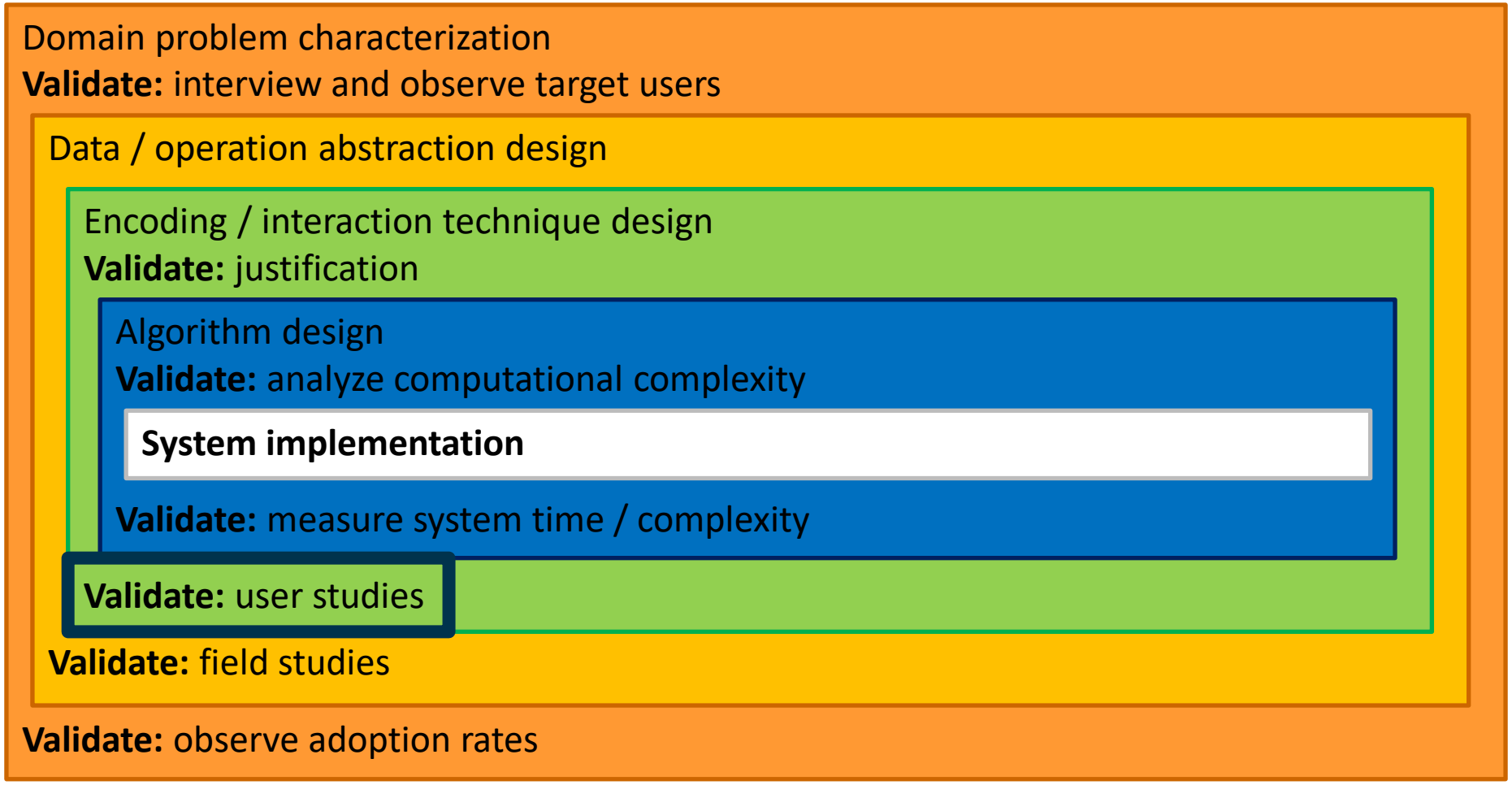
Expressiveness & Effectiveness

- Expressiveness
  - Ordered attributes
    → magnitude channels
  - Categorical attributes
    → identity channels
- Effectiveness
  - Faster to interpret
  - More distinctions
  - Fewer errors





Nested Model for Visualization Design



[Munzner, A Nested Model for Visualization Design and Validation, InfoVis 2009]

Methods: Validating Encoding / Interaction Design

- Controlled user study
  - Comparative lab / crowdsourcing study
  - Eye tracking study
  - Insight-based evaluation
- Inspections performed by experts
  - Heuristic evaluation (informal, holistic)
  - Cognitive walk-throughs (specific tasks)
- Qualitative results inspection

Method: Controlled User Studies

- **Hypothesis**: logical, precise, testable, justifyable

- **Independent variables** / factors
  - Examples: visualization method, task, data size, population...
  - Levels: number of variations for each factor = test conditions
- **Dependent variables**: measured user performance
  - Examples: task completion time, accuracy, number of interaction steps, subjective ratings...

Method: Controlled User Studies

- **Control variables**: held constant during the experiment
  - Examples: display size, data size, data characteristics
- **Random variables**: randomly varied across all conditions
  - Examples: data size, data variations

- **Confounding variables**: unmeasured variable that unintentionally correlates with an independent variable
  - Examples: label readability, visual appeal, interaction design, …
  - Need to be avoided to ensure **internal validity** of the experiment

Method: Controlled User Studies

- Experimental Design:
  - **Within-subjects design**: every condition is performed by every participant (= repeated measures)
    - Be aware of learning effect!
  - **Between-subjects design**: users are divided into groups, performing different variations of one independent variable
  - **Mixed design**: users are divided in groups performing different variations of one independent variable and all variations of others

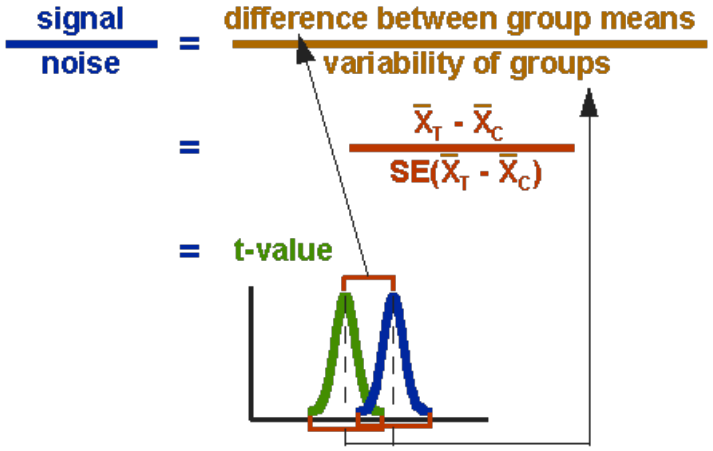
Example: Graph Readability

- **Hypothesis**: “we expect the number of nodes and the link density to influence greatly the readability of this representation”
- **Independent variables**:
  - Graph representation (node-link diagram, adjacency matrix)
  - Size / number of nodes (20, 50, 100)
  - Link density = edges / nodes² (0.2, 0.4, 0.6)
- **Dependent variables**:
  - Accuracy
  - Completion times
- **Design**: within-subjects

[Ghoniem et al., InfoVis 2004]

Method: Controlled User Studies

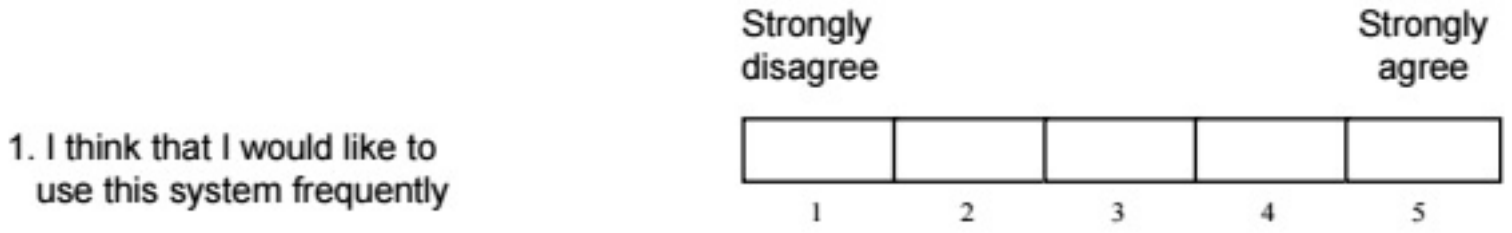
- **Analysis**
  - How large is the difference?
  - Is the difference statistically significant?
    - t-tests
    - **An**alysis **o**f **Va**riance (= ANOVA)
  - Is the difference of practical significance?



http://www.socialresearchmethods.net/kb/stat_t.php

Method: Controlled User Studies

- Nested qualitative inquiry: put quantitative data into social context
  - **Experimenter observations** → detect unexpected events, explanations for outliers
  - **„think-aloud“**
  - **User opinions**
    - Semi-structured interviews
    - Questionnaires (quantification through Likert scale)



http://usabilitygeek.com/how-to-use-the-system-usability-scale-sus-to-evaluate-the-usability-of-your-website/

Method: Controlled User Studies

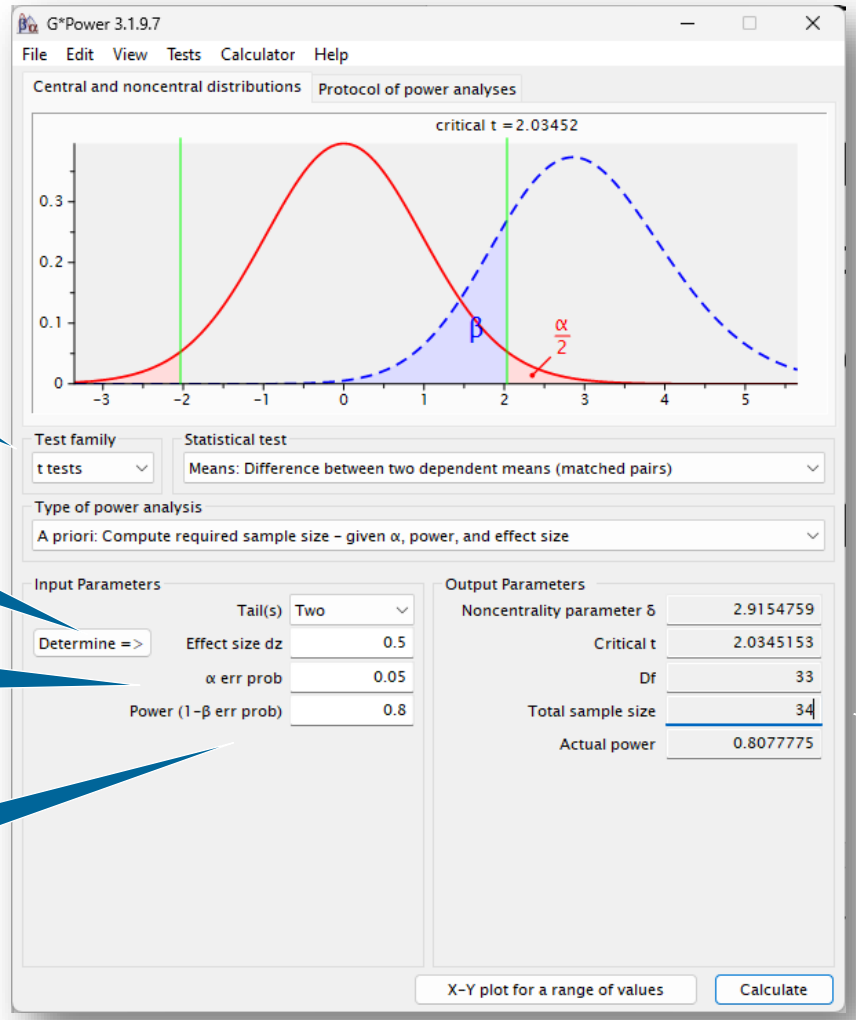
- How many users do I need?
- **Power analysis:**

Within-subjects t-test

Medium **effect size**

α: probability of Type I error

β: probability of Type II error



Number of required participants for desired power

Method: Controlled User Studies

- How to increase power:
  - Increase sample size
  - Keep significance level high / number of conditions low
  - Increase effect size
    - Increase mean difference between measurements of treatments
    - Decrease variance

      Homogeneous population

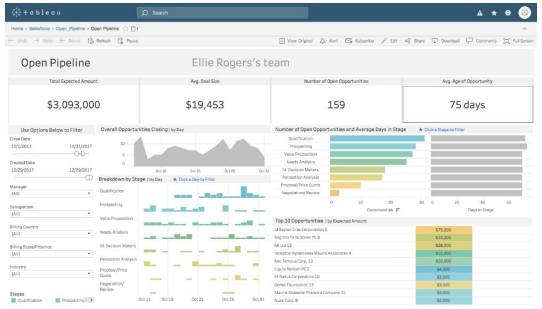
      Removing outliers

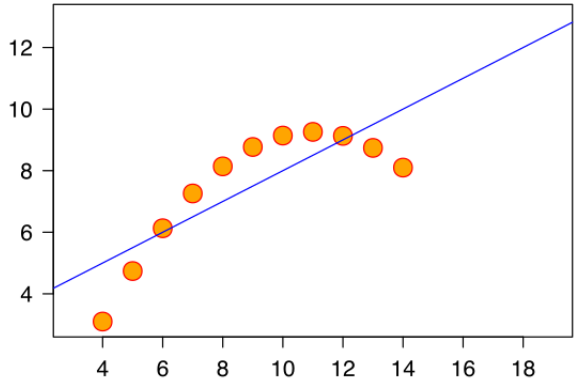
      Control variables instead of random variables

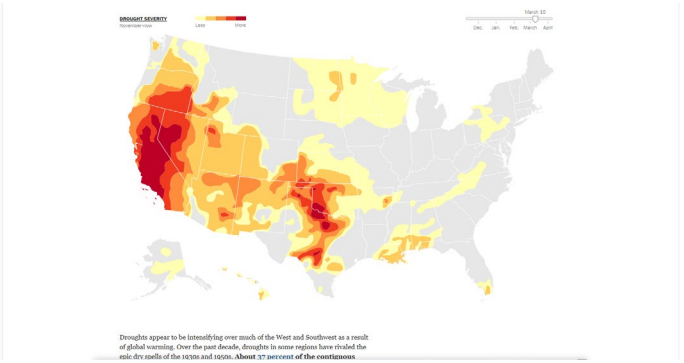
      Controlled conditions

More power = more effort / costs and / or loss of generalizability

Recall Why: High-Level Goals







interactivity

Exploration

Searching and analyzing data without prior hypothesis to find potentially useful information

Confirmation

Goal-oriented examination of hypotheses

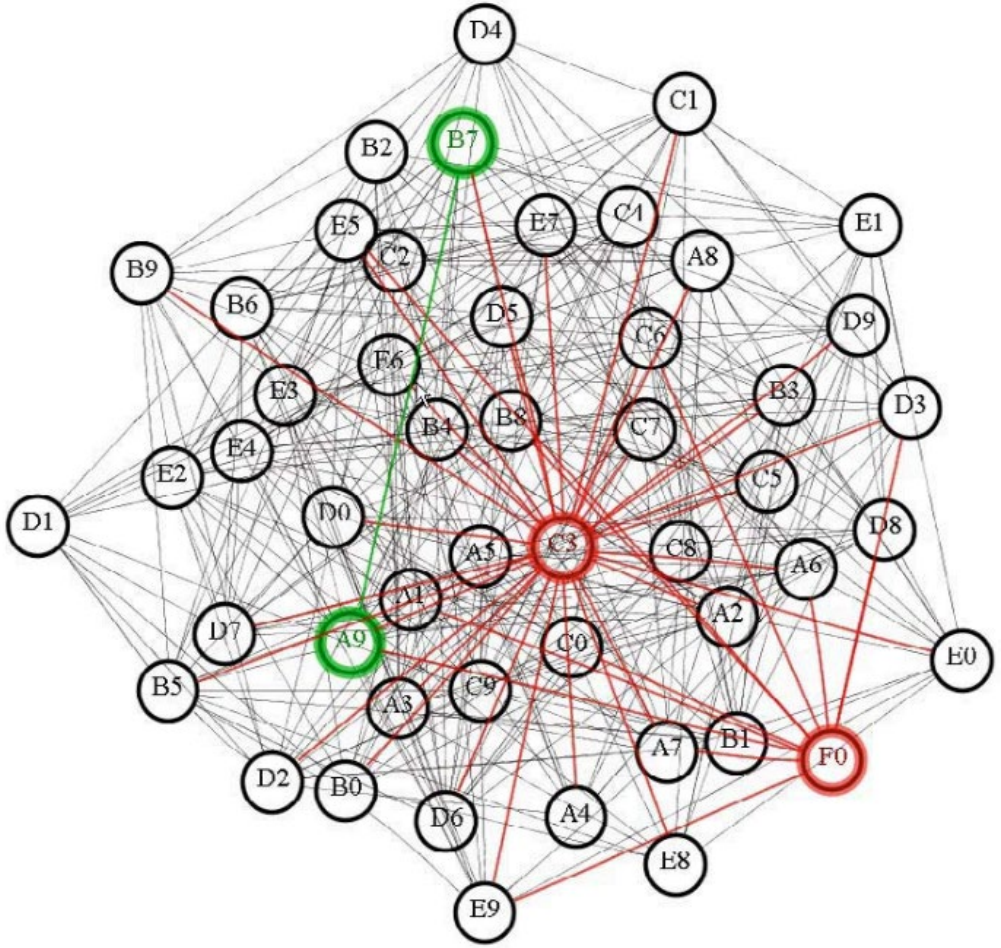
How to measure in a controlled study?

Presentation

Efficiently and effectively communicate facts

Example: Graphs

- How **readable** is this graph?

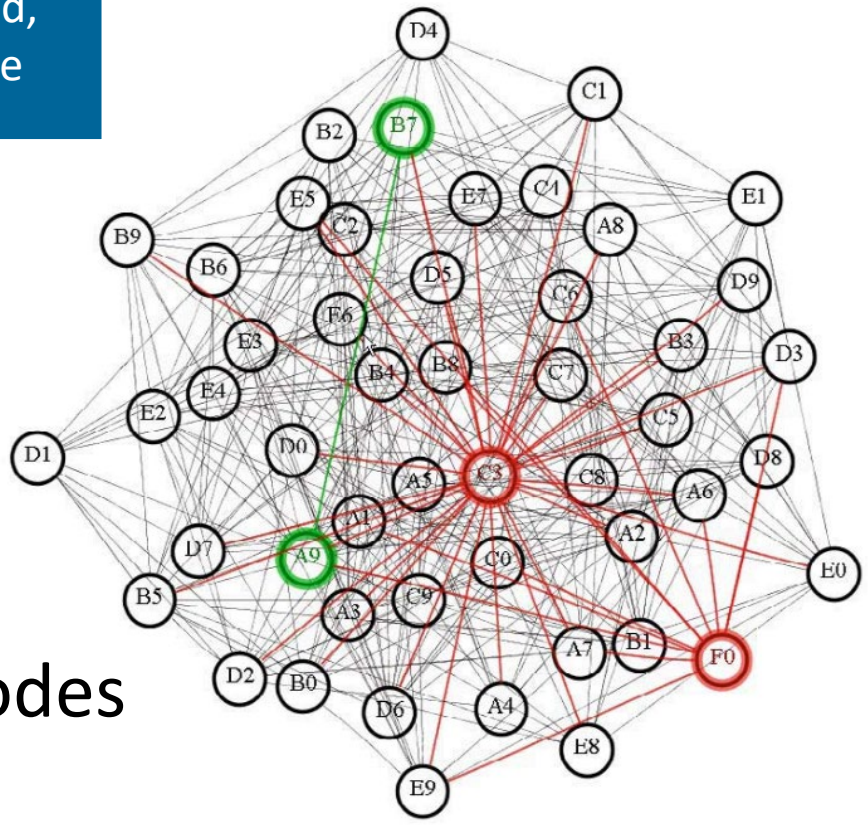


[Ghoniem et al., InfoVis 2004]

Method: Controlled User Studies

Easy to understand, short, measurable

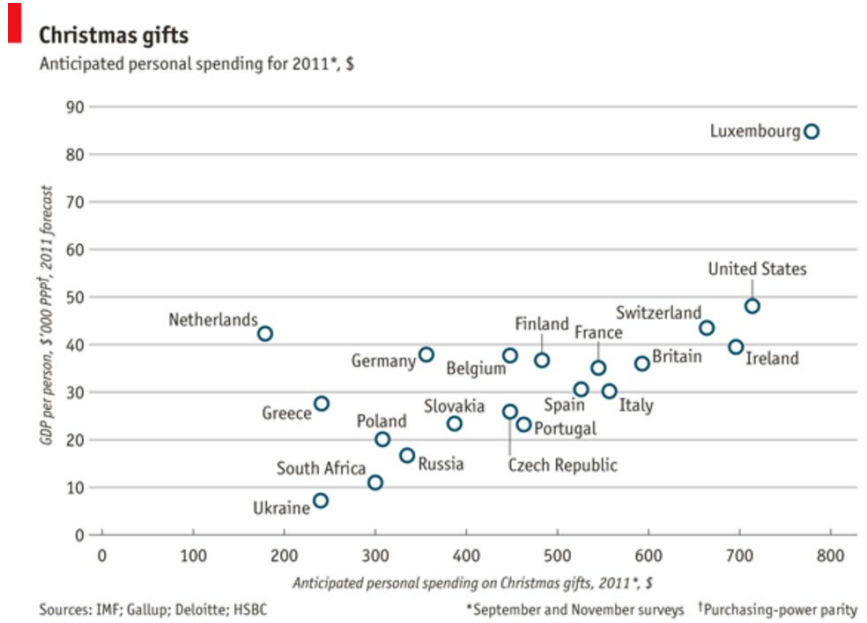
- **Low-level tasks**:
  - Estimate number of nodes / links
  - Finding most connected node
  - Finding node with a given label
  - Finding link between two given nodes
  - Finding common neighbor between two nodes
  - Finding a path between two nodes



[Ghoniem et al., InfoVis 2004]

Method: Controlled User Studies

- **Task taxonomy** of low-level analytical tasks:
  - Retrieve a value
  - Filter
  - Compute derived value (e.g., average, median, count)
  - Find extremum
  - Sort
  - Determine range
  - Characterize distribution
  - Find anomalies
  - Cluster
  - Correlate

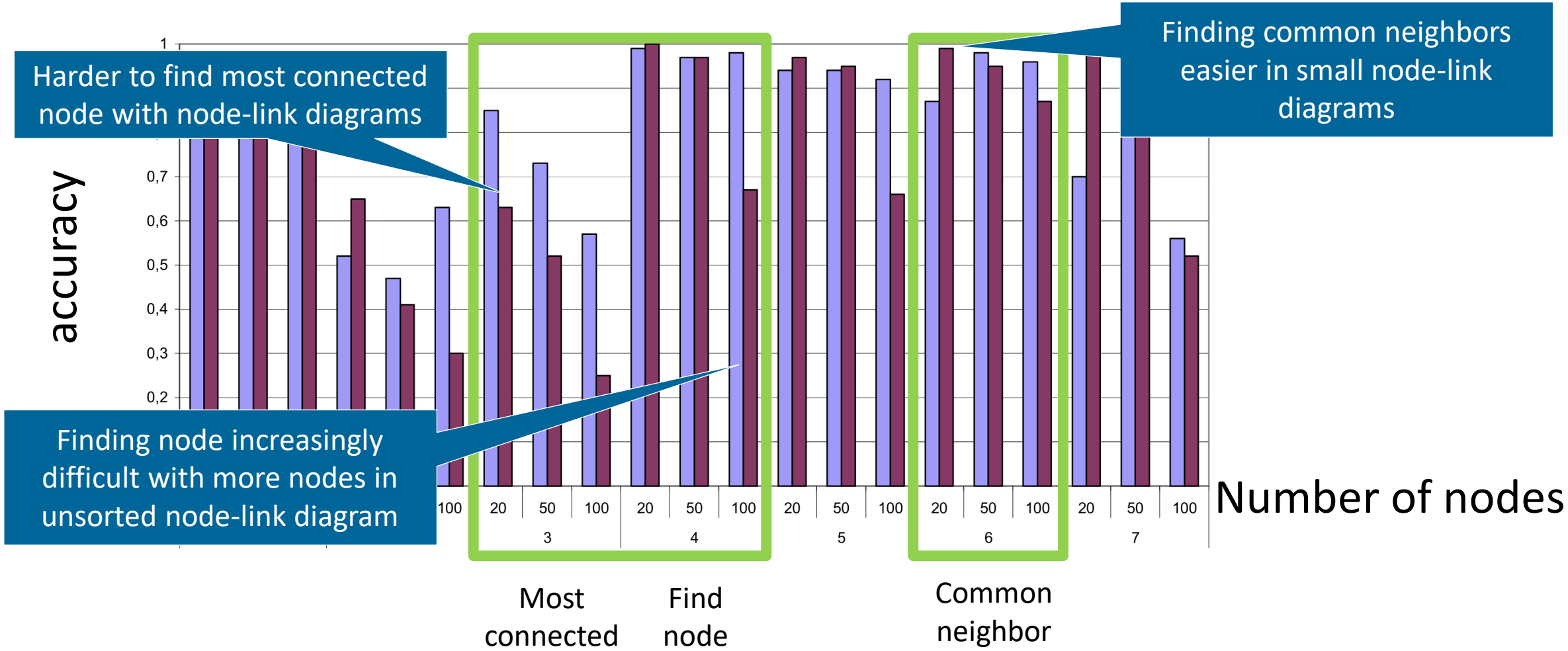


http://massvis.mit.edu/

[Amar et al., Low-Level Components of Analytic Activity in Information Visualization, InfoVis 2005]

Example: Graphs

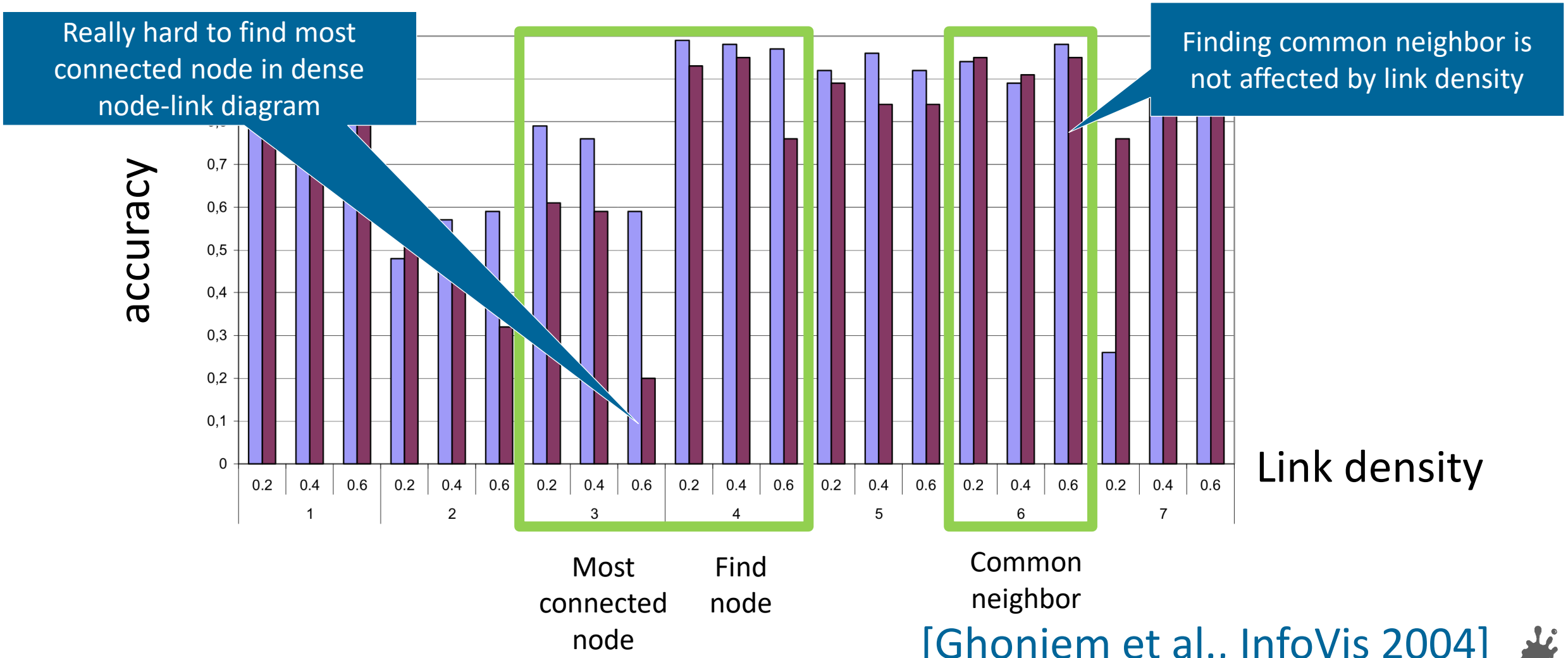
- Results for adjacency matrix and node-link diagrams:



[Ghoniem et al., InfoVis 2004]

Example: Graphs

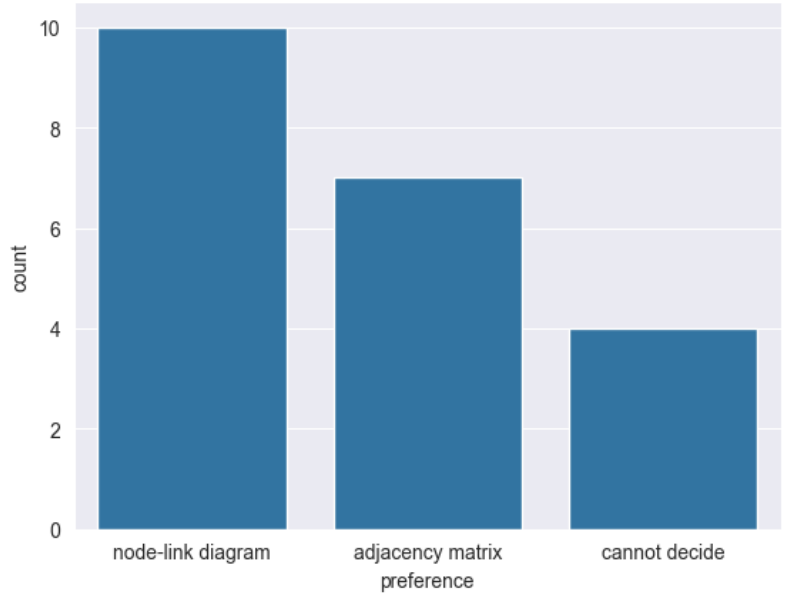
- Results for adjacency matrix and node-link diagrams:

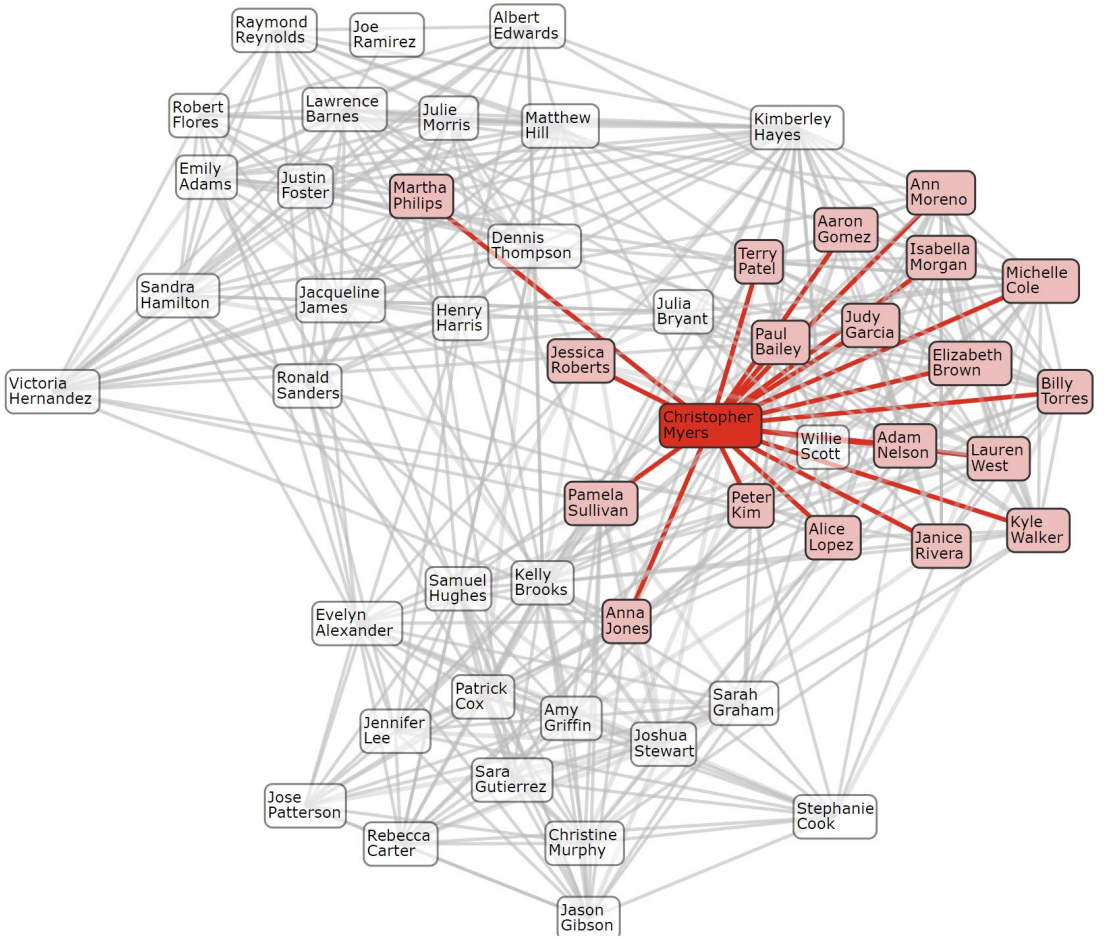


[Ghoniem et al., InfoVis 2004]

Example: Graphs

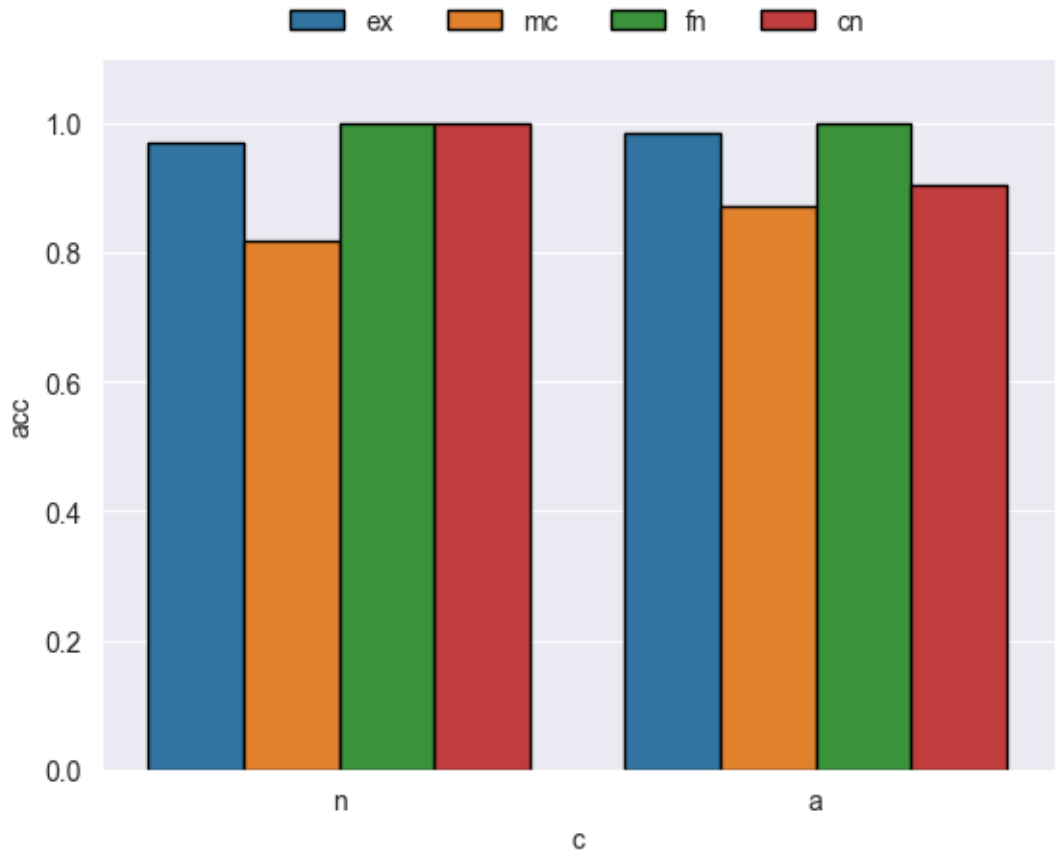
- Results:
  - 22 participants, 21 finished
  - 48% preferred node-link diagrams
  - 33% preferred adjacency matrices

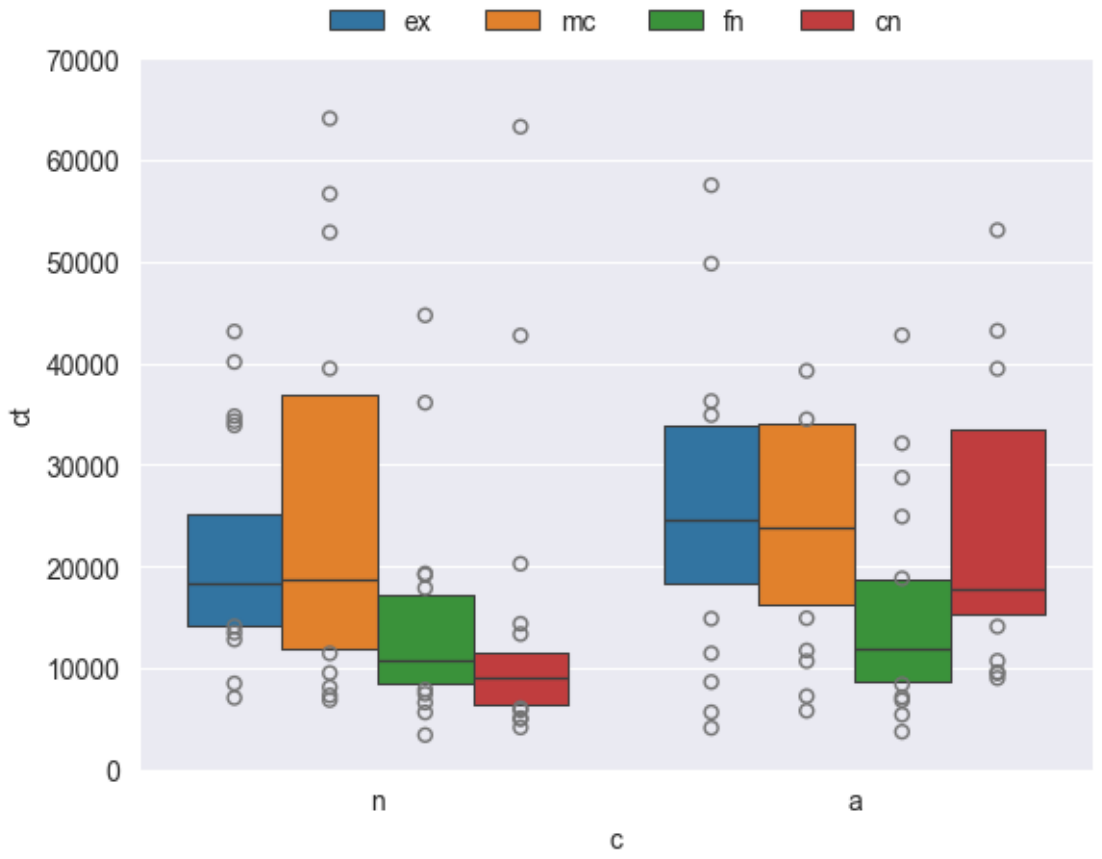




Example: Graphs

- Results:
  - Accuracy and completion times:



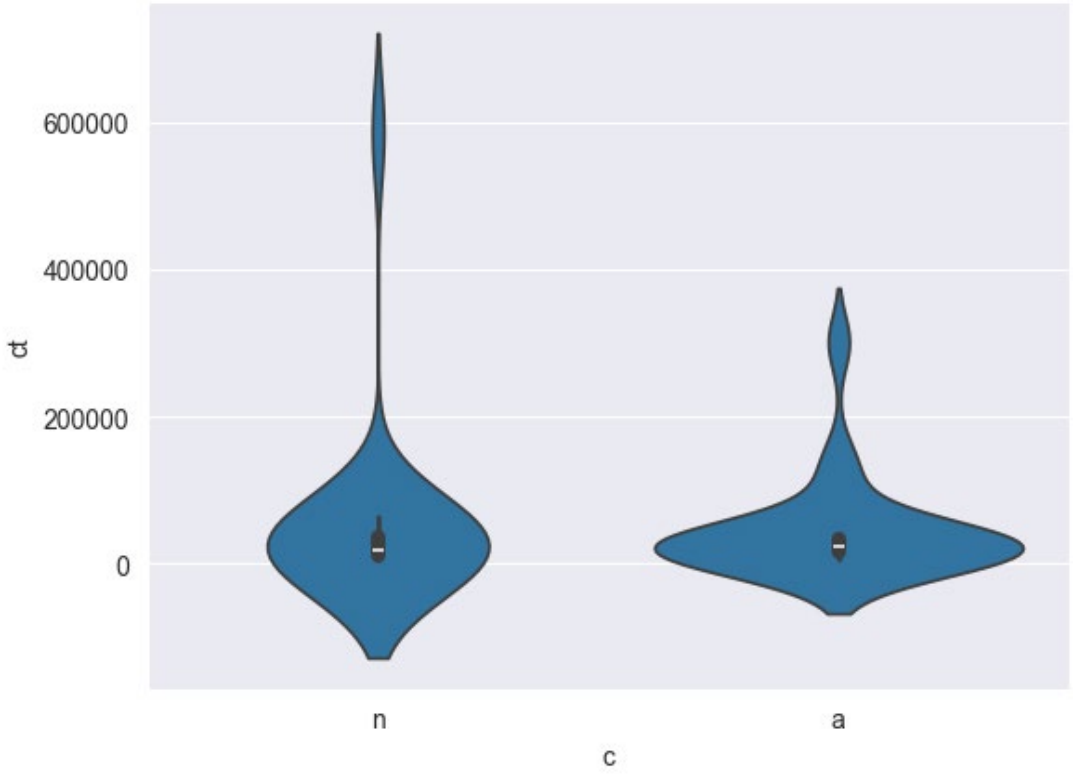


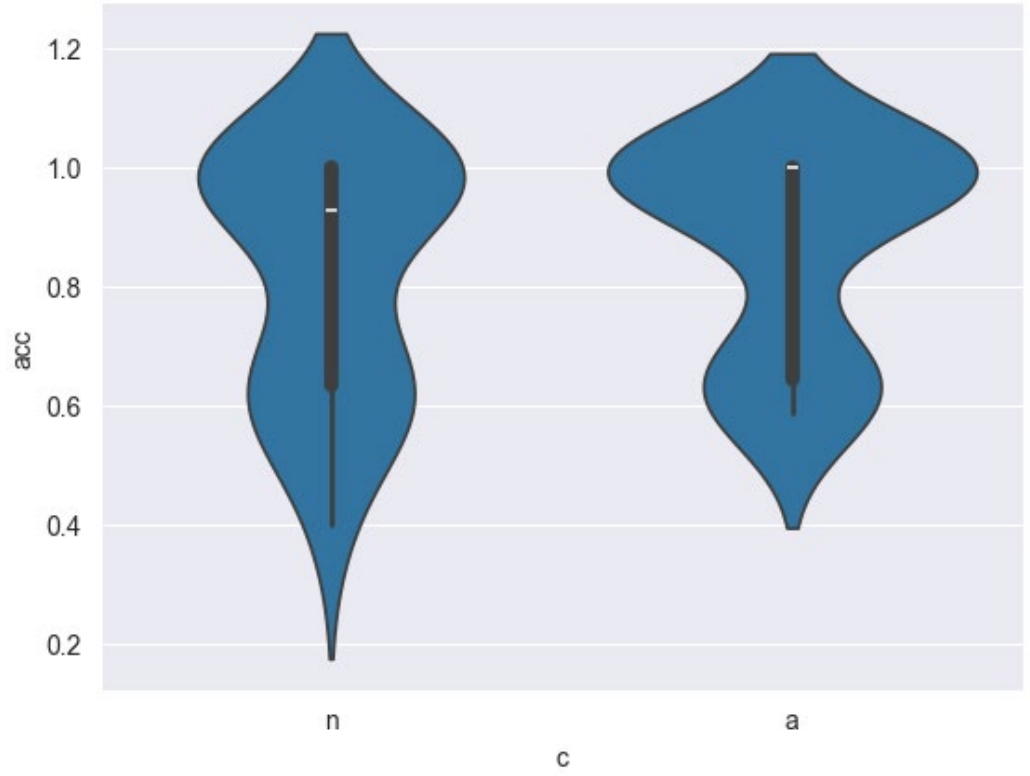
Example: Graphs

- Statistical evaluation:
  - Shapiro test for normal distribution
    - ~~Passed: **matched t-test**~~
    - Not passed: **Wilcoxon signed-rank test**

Example: Graphs

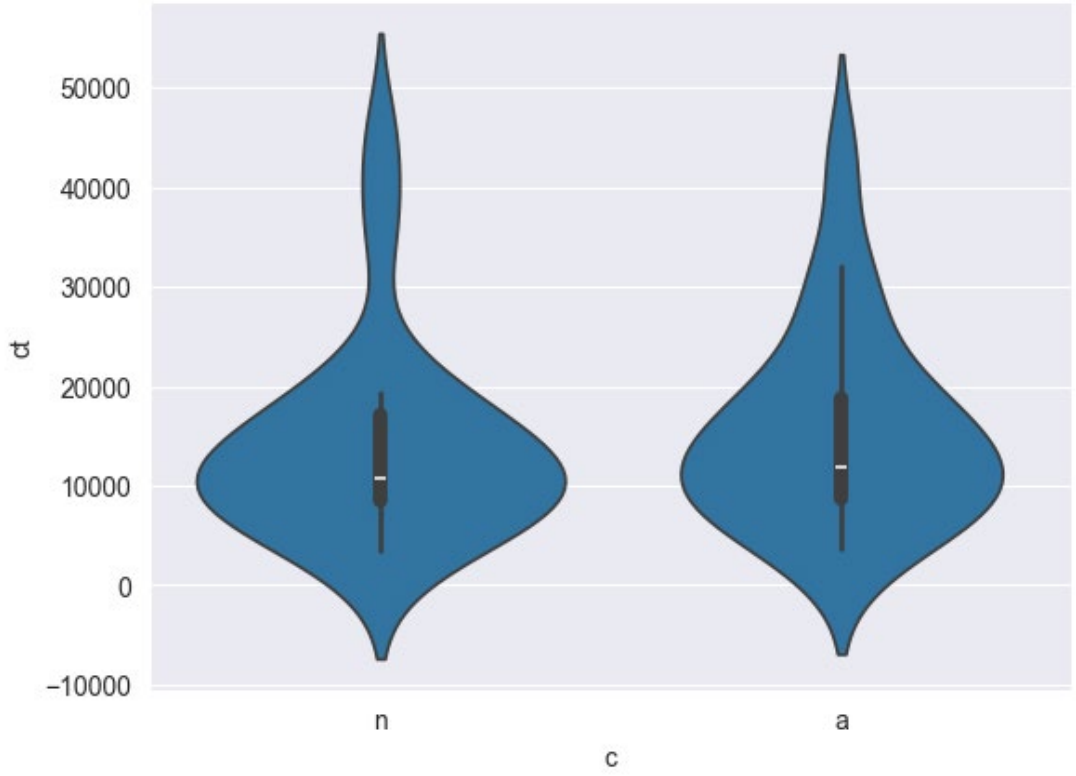
- Task: finding most connected node – no significant differences

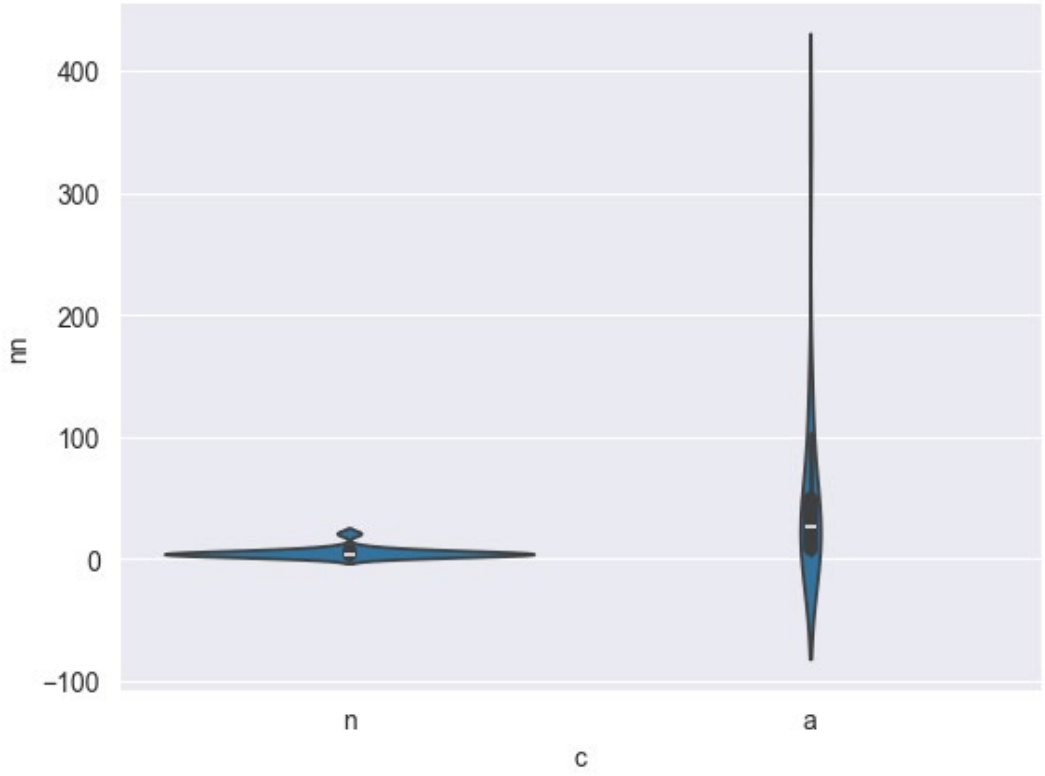




Example: Graphs

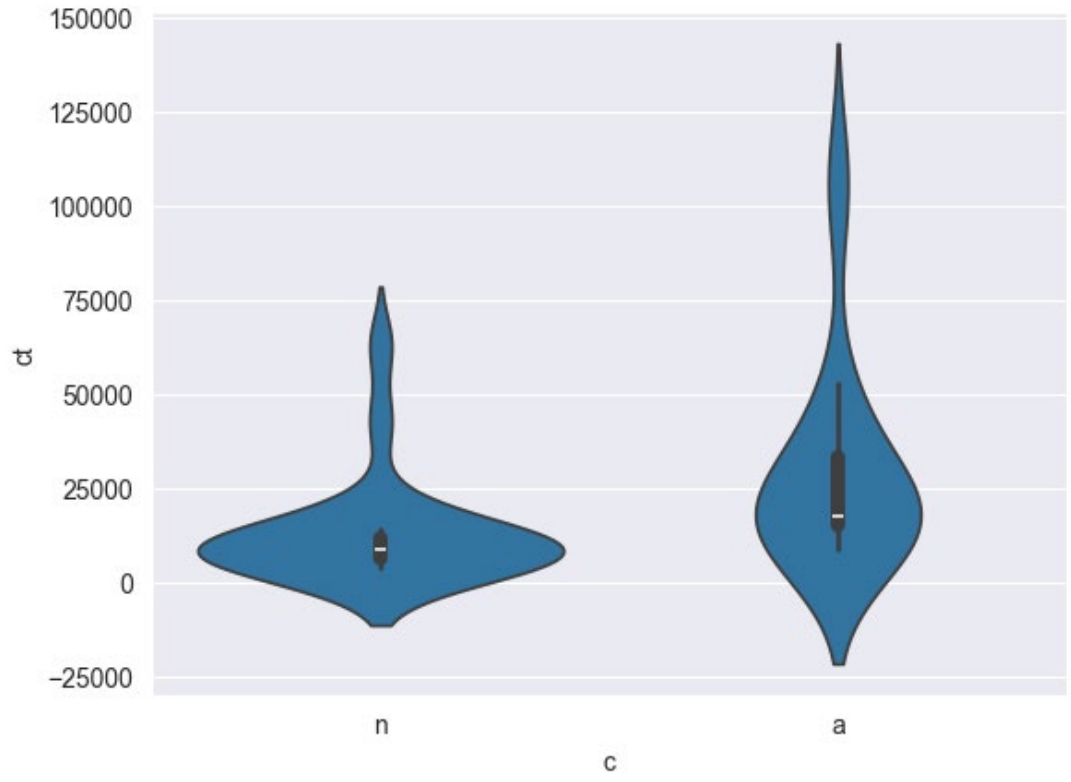
- Task: finding node by name – no significant time differences

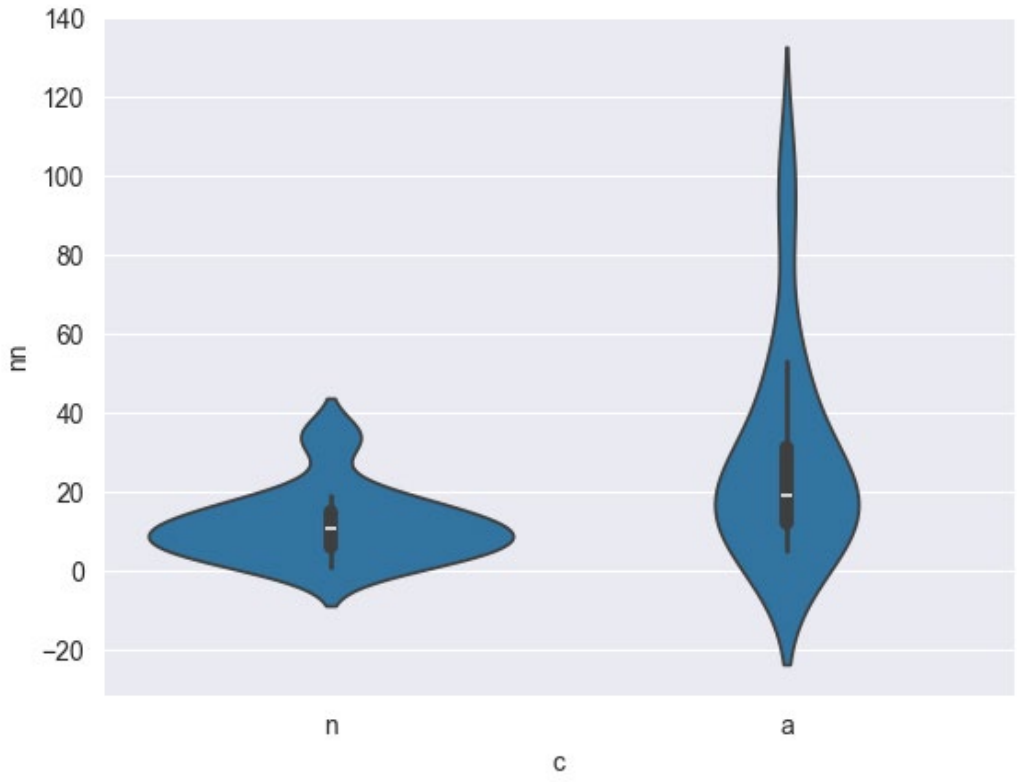




Example: Graphs

- Task: find common neighbor – significantly faster using node-link diagram





Example: Graphs

Using the node-link diagram, it is somewhat tough to identify common connections and their corresponding neighbours. But with the adjacency matrix, it is relatively easier to find out connections as they are laid out in a grid.

That depends on the task. In the adjacency matrix, the names could perhaps have been sorted alphabetically.

it was confusing for me with both visualizations, that the highlighting was gone when hovering

An auto scroll to the top for every question would have been nice. Also the copied task description for the switch between adjacency matrix and node-link diagram was a little confusing. A text like "We have the same situation as before but now it is visualised differently" would have been great.

Making the highlight toggle-able would have helped. Also, pre-clustering in the adjacency matrix helped immensely. Random ordering would result in having to count all connections by hand.

To the question of "which one did you prefer?" I must reply with: to do what exactly?

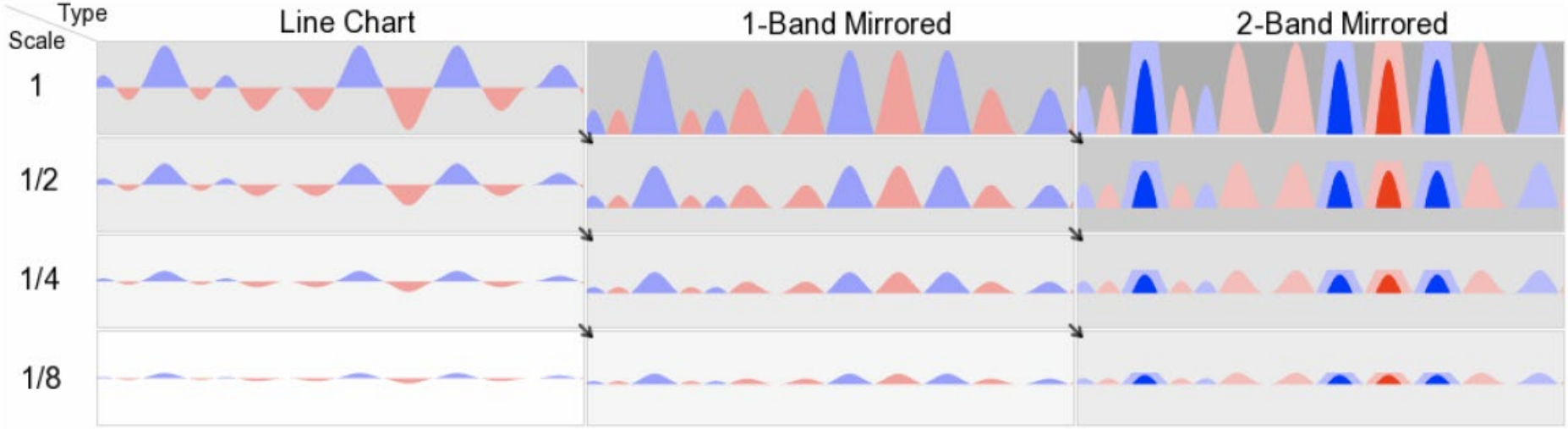
Example: Graphs

- Node-link diagram: hard to see common connections and neighbors
- Adjacency matrix:
  - Grid layout helps to see neighbors
  - Sorting:
    - According to connectivity helpful
    - Not according to names difficult
- Subjective difficulty depends on the task (2x)
- Asking for more / improved interactivity (2x)

Example: Graphs

- Replication study:
  - Harder to find most connected node in node-link diagram ✗
  - Harder to find node with given name in node-link diagram ✗
  - Finding common neighbors easier in node-link diagram ✓

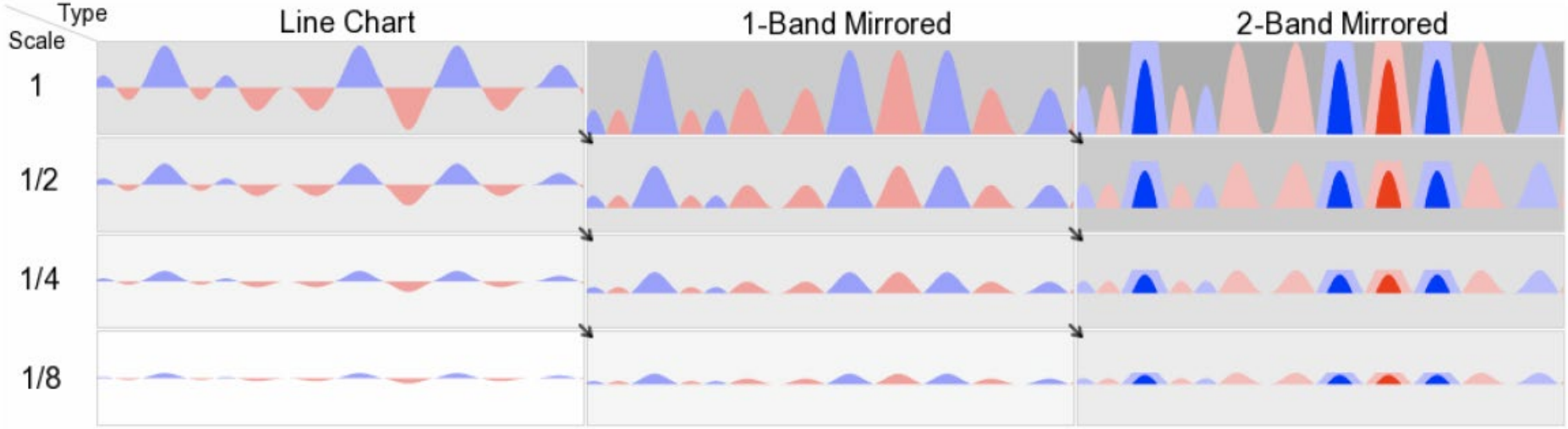
Example: Horizon Graphs



- **Research questions:**
  - How do mirroring and layering affect estimation time and accuracy compared to line charts?
  - How does chart size affect estimation time and accuracy?

[Heer et al., Sizing the Horizon, CHI 2009]

Example: Horizon Graphs

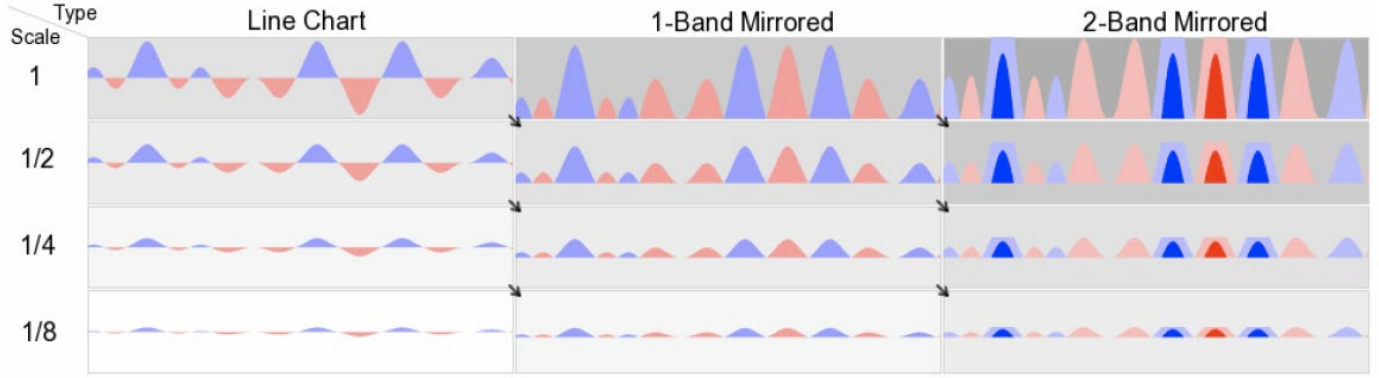


- **Hypotheses:**
  - Large: line charts faster & more accurate than mirror charts
  - Large: 1-band mirror faster & more accurate than 2-band mirror
  - Smaller: 2-band mirror faster & more accurate than others

[Heer et al., Sizing the Horizon, CHI 2009]

Example: Horizon Graphs

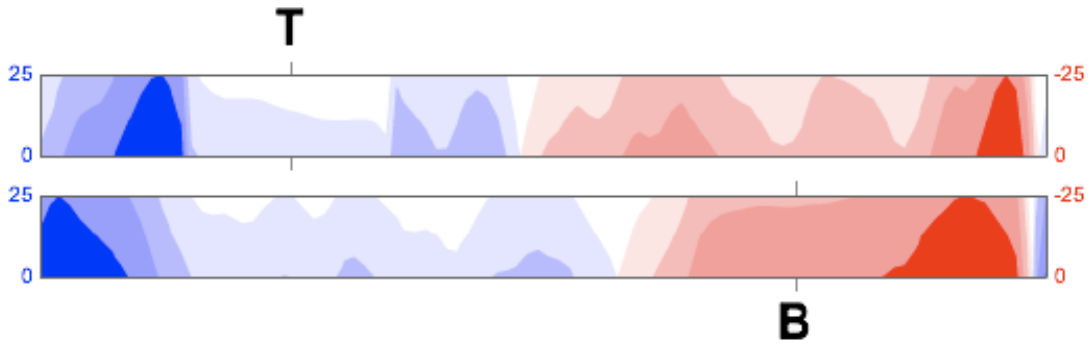
- Method:
  - **Independent variables:** chart type, chart height
  - **Dependent variables:** error, time
  - 30 users (undergraduate students)
  - 3 (chart) x 4 (size) **within-subjects design**
  - 10 repetitions → 120 trials per participants



[Heer et al., Sizing the Horizon, CHI 2009]

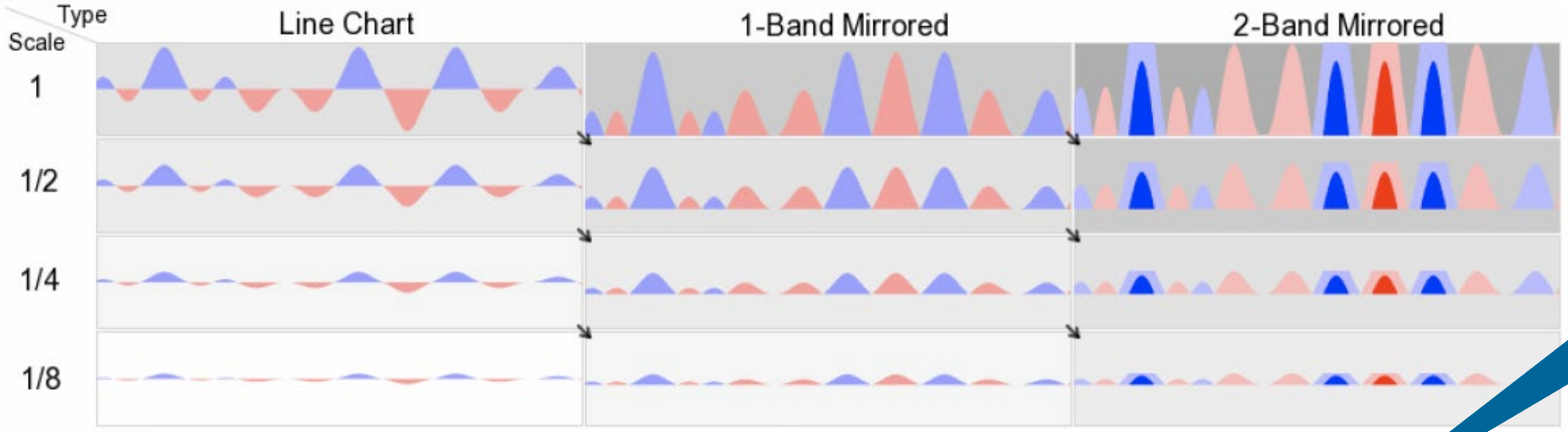
Example: Horizon Graphs

- Tasks:
  - Report whether T or B is higher
  - Estimate absolute difference between T and B
- Analysis:
  - Repeated measures multivariate analysis of variance (RM-MANOVA)



[Heer et al., Sizing the Horizon, CHI 2009]

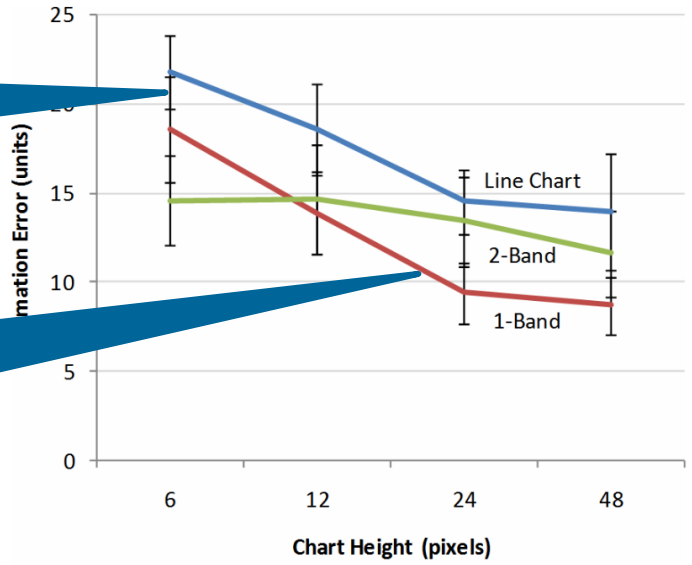
Example: Horizon Graphs

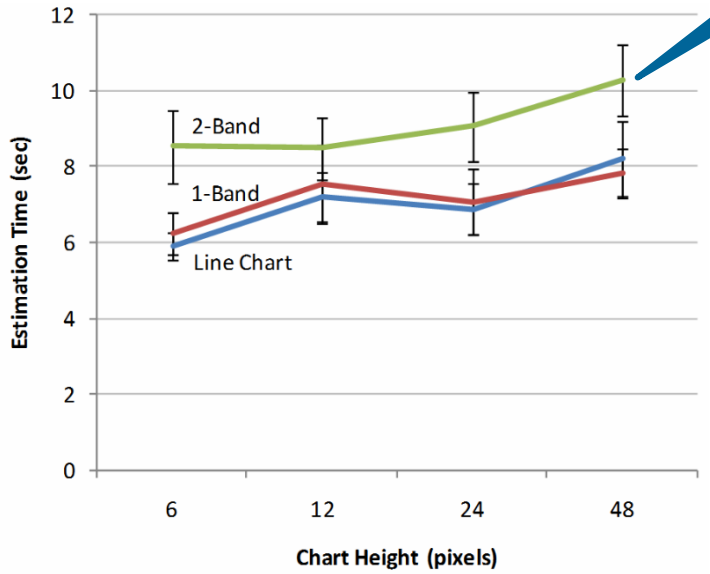


2-band: slower than 1-band and line chart

Line charts: higher error than 1-band

2-band: higher error than 1-band for larger charts





[Heer et al., Sizing the Horizon, CHI 2009]

Method: Crowdsourcing Visualization Evaluation

- Web workers perform small tasks for micro-payments
  - Amazon Mechanical Turk
  - Prolific
- Advantages:
  - More diverse population → ecological validity
  - Large population → large design space evaluation
- Disadvantages:
  - Lack of control over experimental conditions
  - Subject motivation?

[Heer and Bostock, Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design, CHI 2010]

Method: Eye Tracking

- Recording of spatio-temporal eye movement data
- Raw gaze data → filters → metrics for statistical analysis:
  - Fixations:
    - Example: number of fixations in an **area of interest** (**AOI**)
  - Saccades:
    - Example: saccade amplitude
  - Scanpath (series of fixations and saccades):
    - Example: transition matrix between pairs of AOIs

[Kurzhals et al., Evaluating Visual Analytics with Eye Tracking, BELIV 2014]

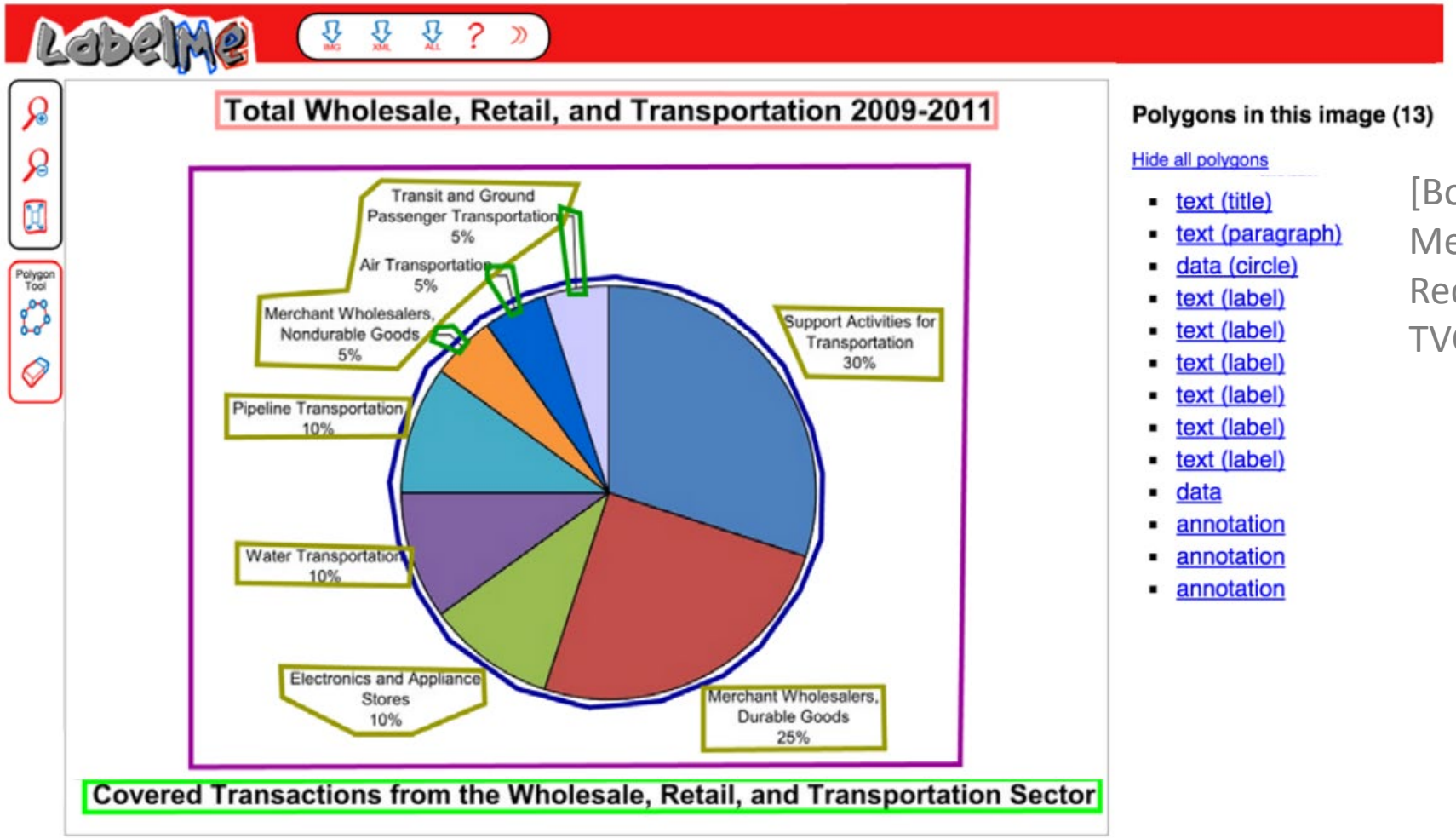
Example: Task-Dependent Areas of Interest

- Retrieve value task: „What is the attendance of Universal Studios Hollywood?“

[Polatsek et al., Exploring visual attention and saliency modeling for task-based visual analysis, C&G 2018]

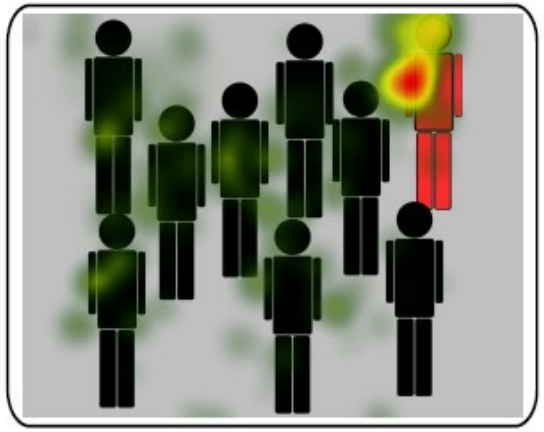
Example: Task-Independent Areas of Interest

- Data and different text categories



[Borkin et al., Beyond Memorability: Visualization Recognition and Recall, TVCG 2016]

Visual Analysis of Eye Tracking Data…



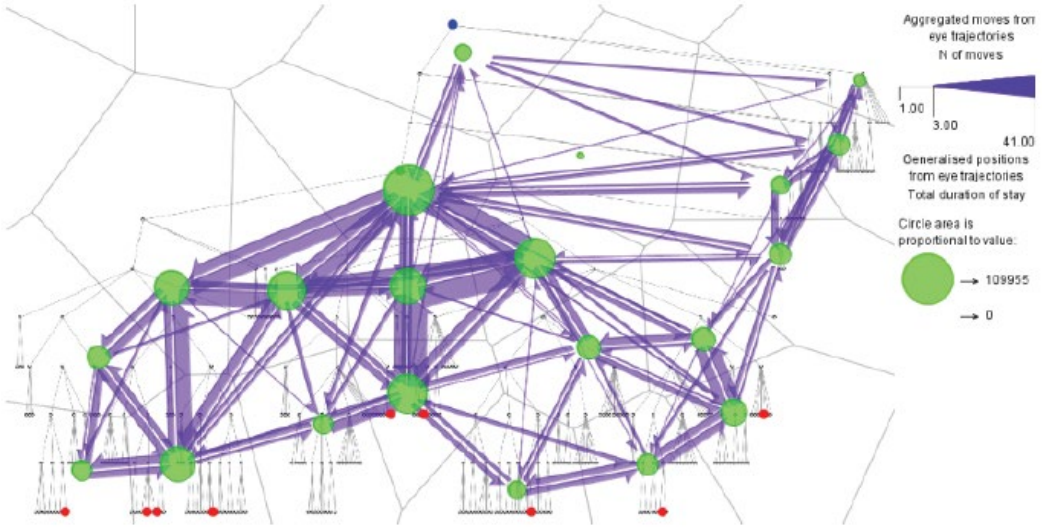
Attention map

Gaze plot

[Kurzhals et al., Evaluating Visual Analytics with Eye Tracking, BELIV 2014]

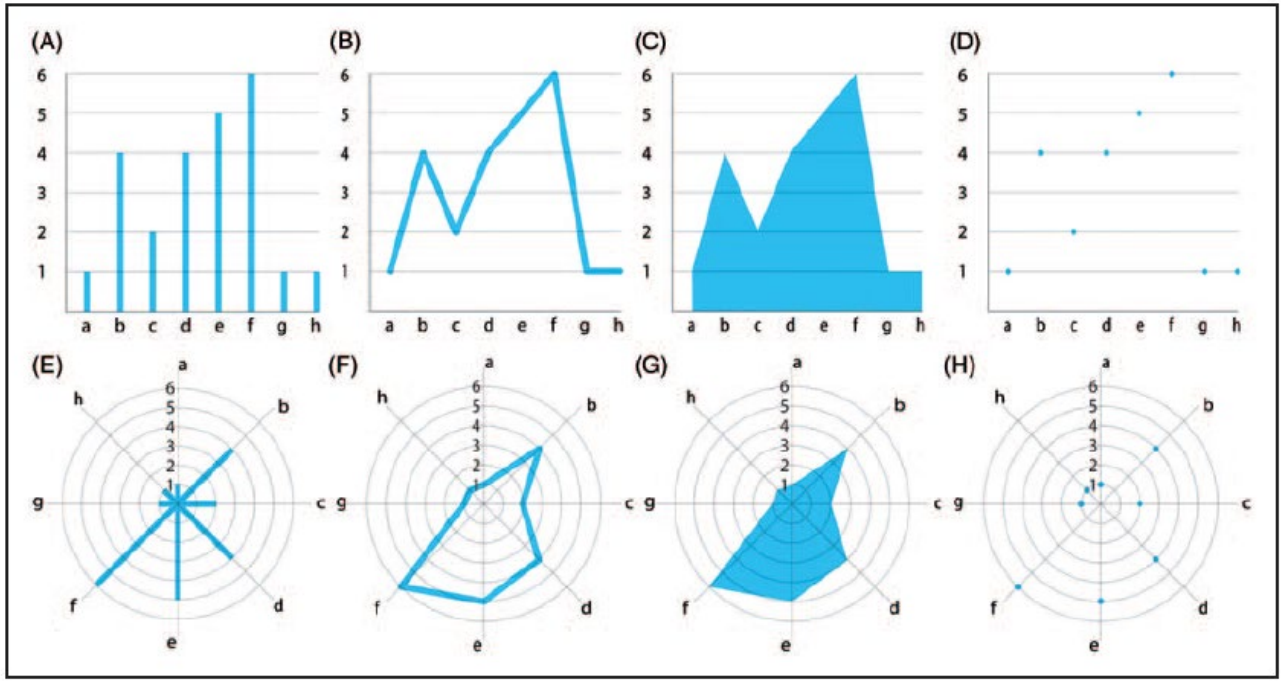
Visual analytics methods from geographic movement data analysis

[Andrienko et al., Visual Analytics Methodology for Eye Movement Studies, TVCG 2012]



Example: Linear vs. Radial Graphs

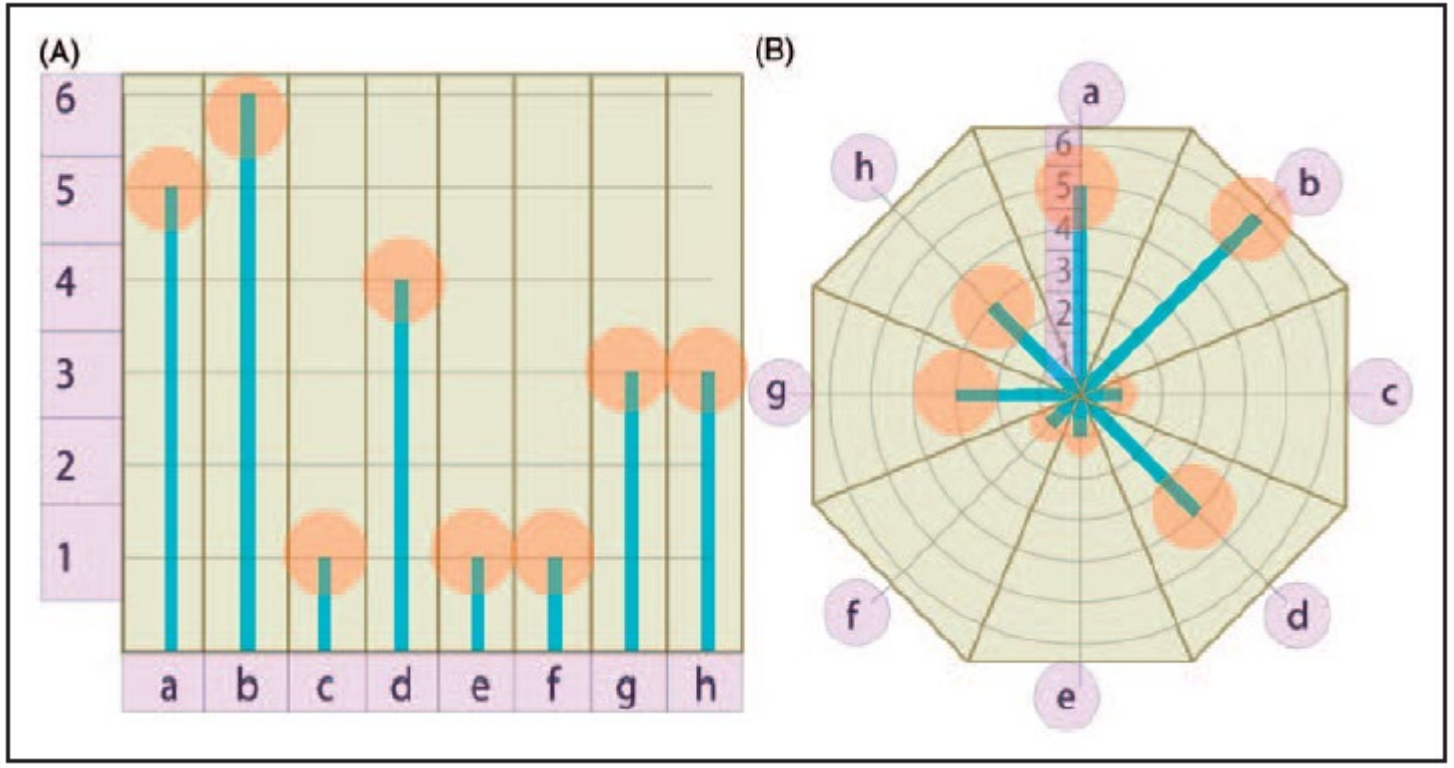
- Comparison of „retrieve value“ task with four graph types (bar, line, area, scatter) and two graph styles (linear, radial)



[Goldberg and Helfman, Eye tracking for visualization evaluation: Reading values on linear versus radial graphs, Information Visualization 2011]

Example: Linear vs. Radial Graphs

- Areas of Interest (AOIs)



[Goldberg and Helfman, Eye tracking for visualization evaluation: Reading values on linear versus radial graphs, Information Visualization 2011]

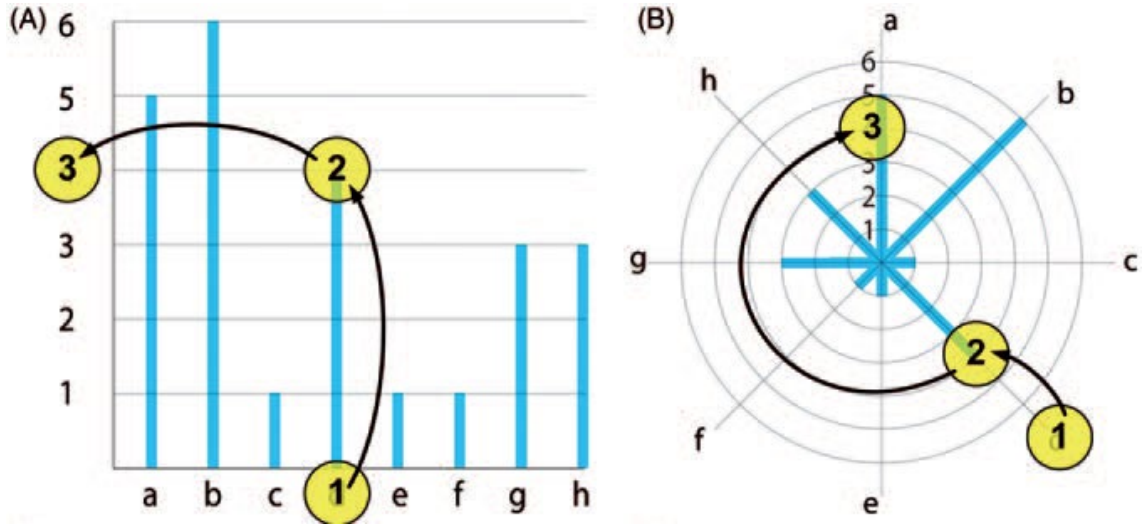
Example: Linear vs. Radial Graphs

- 32 experienced users
- **Dependent variables:**
  - Task completion time
  - First fixation time: initial time a fixation was made within an AOI
  - Minimum time: instant a participant had completed initial fixations within all required AOIs
  - ....
- **Analysis:** ANOVA with Tukey's pairwise comparisons

[Goldberg and Helfman, Eye tracking for visualization evaluation: Reading values on linear versus radial graphs, Information Visualization 2011]

Example: Linear vs. Radial Graphs

- **Results:**
  - Radial graph response time slower
  - Mapping datapoint to value (step 3) especially slow for radial graphs



[Goldberg and Helfman, Eye tracking for visualization evaluation: Reading values on linear versus radial graphs, Information Visualization 2011]

Method: Heuristic Evaluation

- „discount evaluation method“:
  - 5 evaluators find > 75% of problems [Nielson]
  - Detect positive and negative aspects of a visualization based on set of heuristics
- Information visualization heuristics
  - Shneiderman‘s information seeking mantra („overview first, zoom and filter...“)
  - Task-based
  - Perception-based
    (e.g., „consider people with color blindness“)

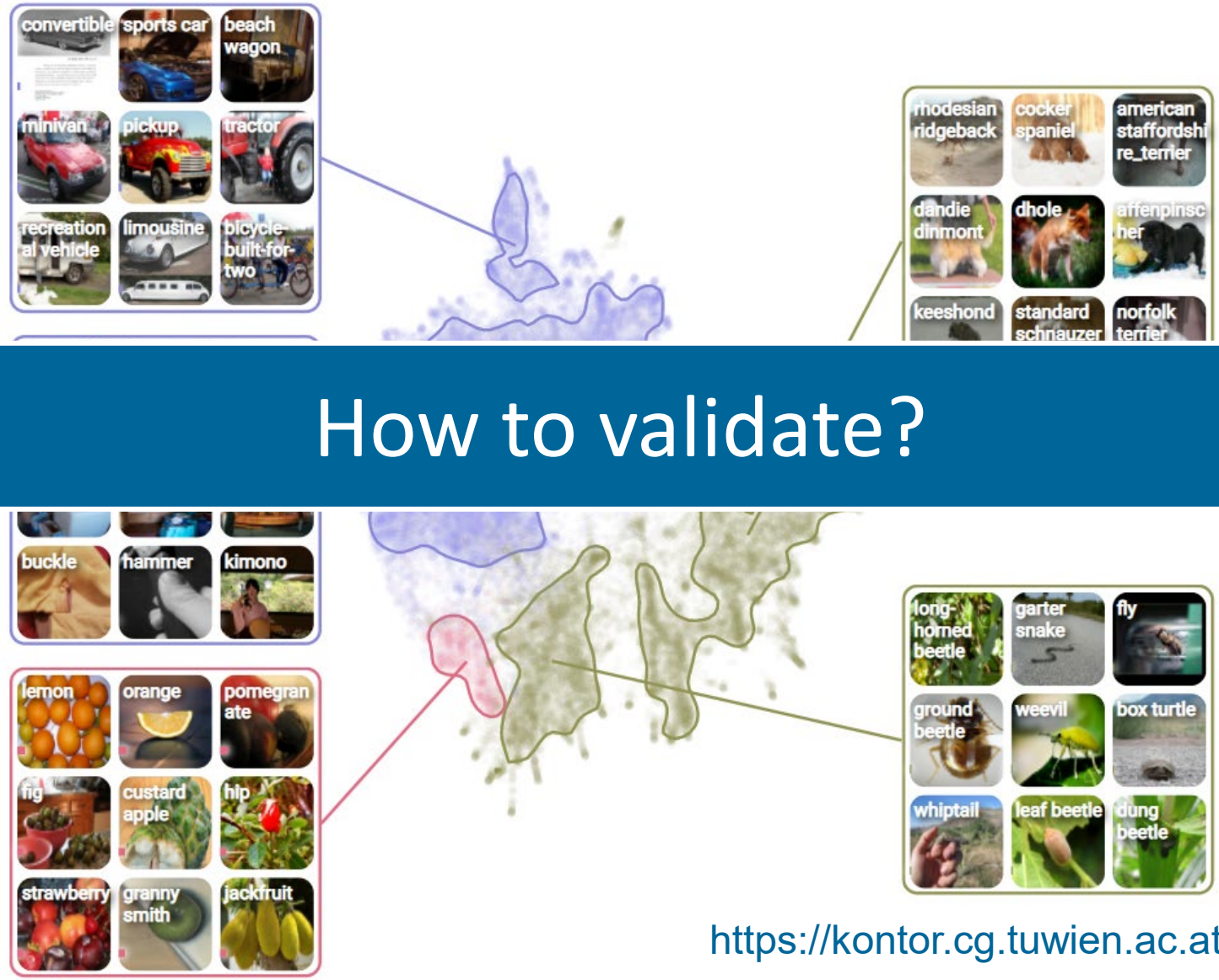
[Zuk et al., **Heuristics for Information Visualization Evaluation**, BELIV 2006]

Method: Qualitative Result Inspection / Case Study

- No experiment
- Reader inspects result image or walkthrough
  - Comparative
  - Isolated

[Isenberg et al., A Systematic Review on the Practice of Evaluating Visualization, TVCG 2013]

Example: ConceptSplatters



How to validate?

https://kontor.cg.tuwien.ac.at/ConceptSplatters/

[Grossmann et al., Computers & Graphics 2022]

Example: ConceptSplatters

Do not **cherry-pick**: replicate „non-visualized“ cases from the past

“The tench class, a large fish species, is often identified by fingers on front of a greenish background. Closer inspection revealed that tench images typically feature the fish hold up like a trophy, thus making the hand and fingers holding it a very predictive image feature.”

Published as a conference paper at ICLR 2019

APPROXIMATING CNNS WITH BAG-OF-LOCAL-FEATURES MODELS WORKS SURPRISINGLY WELL ON IMAGENET

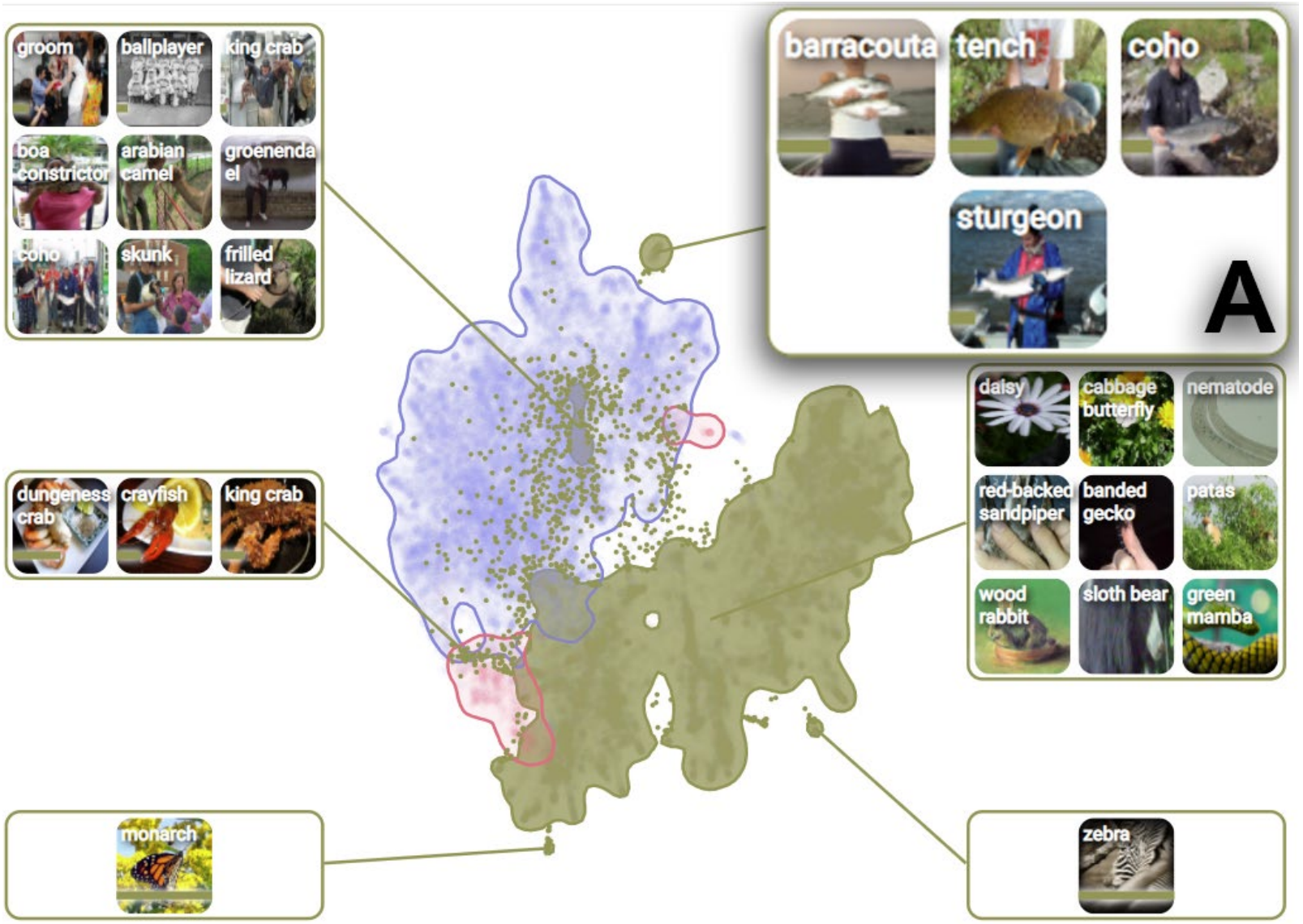
Wieland Brendel and Matthias Bethge

Eberhard Karls University of Tübingen, Germany
Werner Reichardt Centre for Integrative Neuroscience, Tübingen, Germany
Bernstein Center for Computational Neuroscience, Tübingen, Germany
{wieland.brendel, matthias.bethge}@bethgelab.org

ABSTRACT

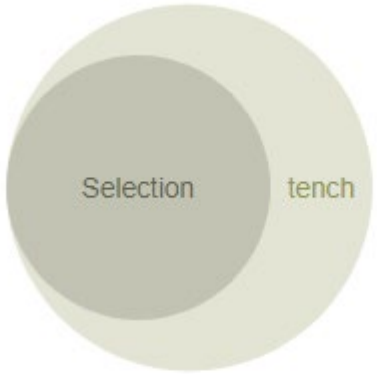
Deep Neural Networks (DNNs) excel on many complex perceptual tasks but it has proven notoriously difficult to understand how they reach their decisions. We here introduce a high-performance DNN architecture on ImageNet whose decisions are considerably easier to explain. Our model, a simple variant of the ResNet-50 architecture called BagNet, classifies an image based on the occurrences of small local image features without taking into account their spatial ordering. This strategy is closely related to the bag-of-feature (BoF) models popular before the onset of deep learning and reaches a surprisingly high accuracy on ImageNet (87.6% top-5 for 33 × 33 px features and Alexnet performance for 17 × 17 px features). The constraint on local features makes it straight-forward to analyse how exactly each part of the image influences the classification. Furthermore, the BagNets behave similar to state-of-the art deep neural networks such as VGG-16, ResNet-152 or DenseNet-169 in terms of feature sensitivity, error distribution and interactions between image parts. This suggests that the improvements of DNNs over previous bag-of-feature classifiers in the last few years is mostly achieved by better fine-tuning rather than by qualitatively different decision strategies.

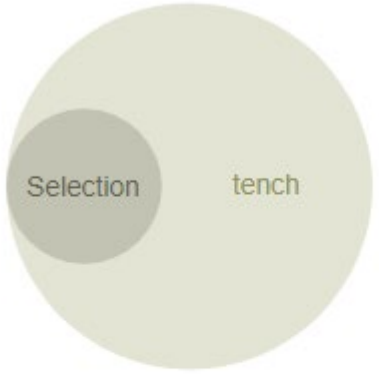
Example: ConceptSplatters



[Grossmann et al., Computers & Graphics 2022]

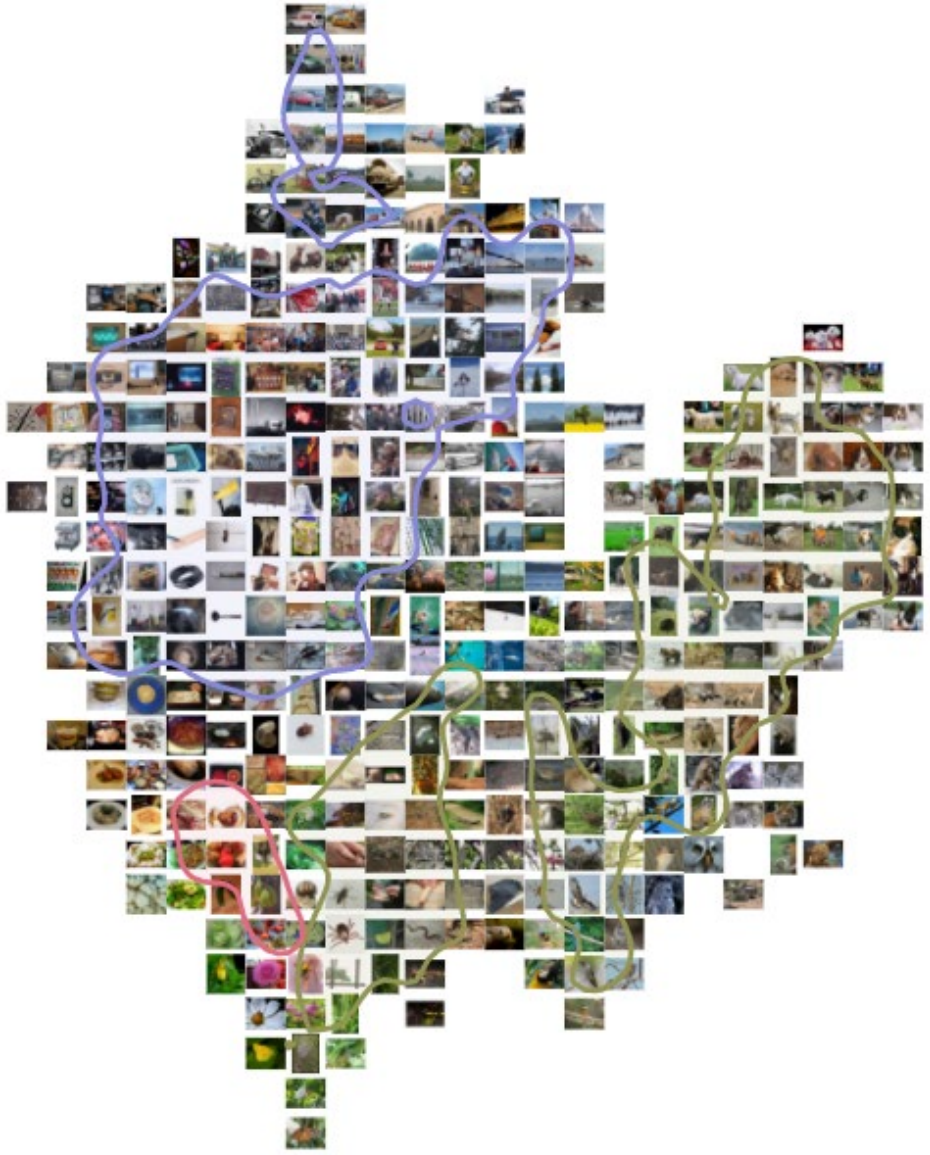
Example: ConceptSplatters





Show insights beyond what is already known / reported.

Example: ConceptSplatters



Pick a useful baseline and show that this would be a user study with a **foregone conclusion.**

Example: ConceptSplatters

"Each dimension of the vectors encodes a different aspect of words. In embedding spaces, semantically close words are likely to cluster together and form semantic cliques."

Do not limit yourself to just one example. Demonstrate the **generalizability** of your visualization!

Neurocomputing
Volume 174, Part B, 22 January 2016, Pages 806-814

Semantic expansion using word embedding clustering and convolutional neural network for improving short text classification

Peng Wang [a], Bo Xu [a], Jiaming Xu [a], Guanhua Tian [a], Cheng-Lin Liu [a b], Hongwei Hao [a]

Show more

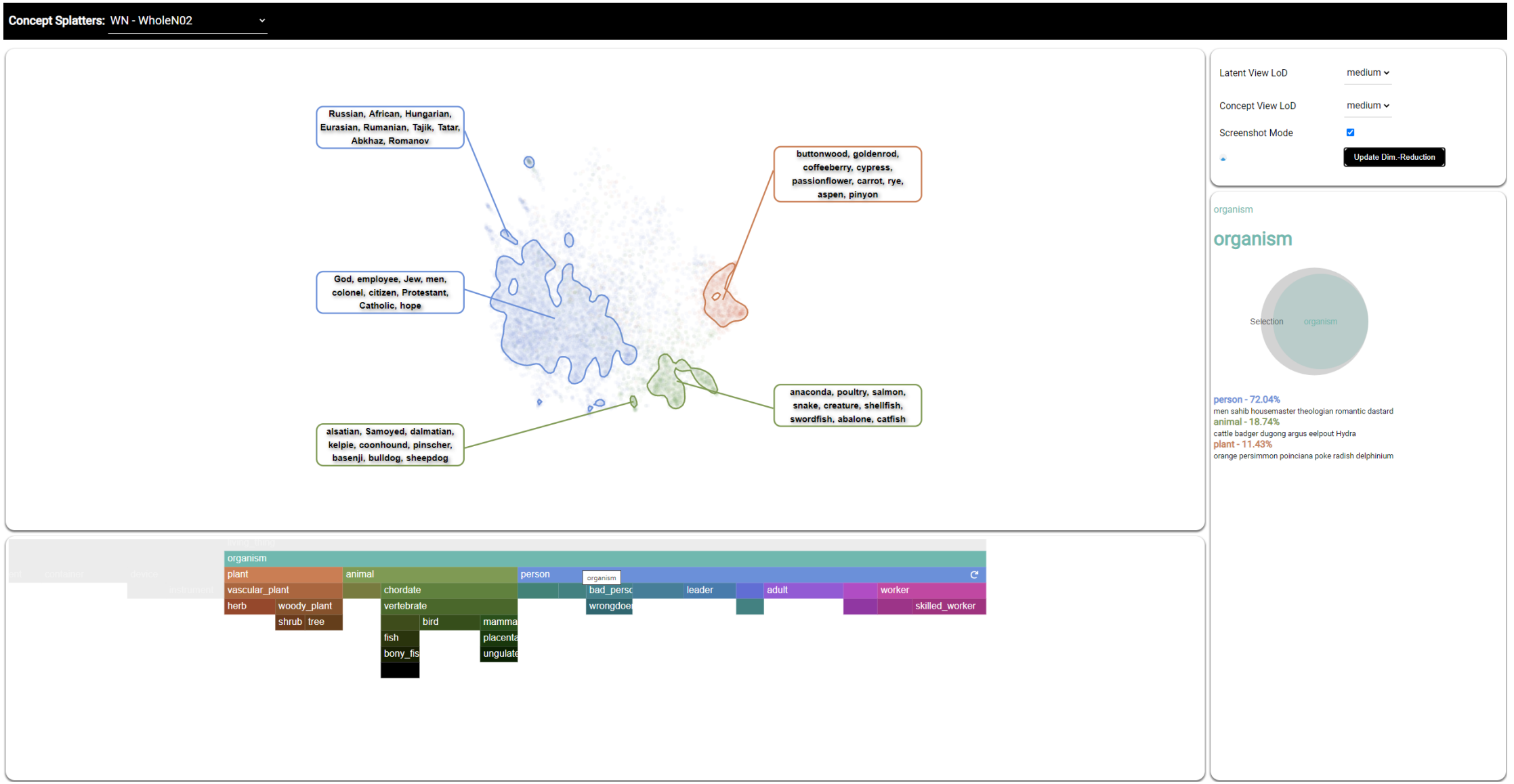
+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.neucom.2015.09.096 Get rights and content

Abstract

Text classification can help users to effectively handle and exploit useful

Example: ConceptSplatters



Which Method?

- Useless user studies:
  - Studies with foregone conclusions
  - Fishing for results
  - Confounding factors
  - Experimenter bias
  - Lack of external validity
  - Lack of practical significance
  - Lack of power

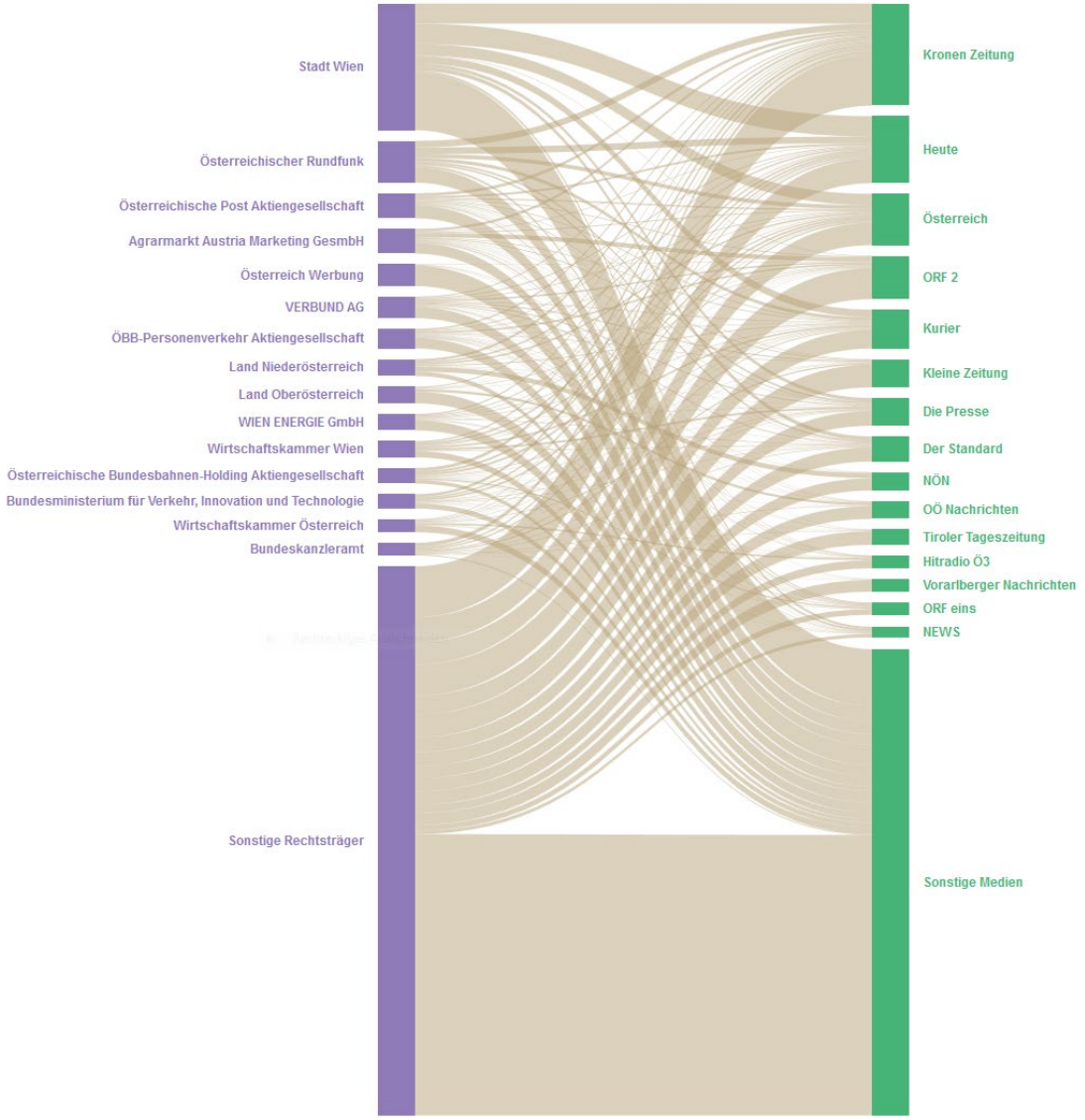
Define your research questions and goals / success criteria cleary **before** your start. Then, picking the appropriate validation method is usually straight-forward!

Method: Insight-Based Study

- Users explore data until they feel they have learned all they can from the data
- No predefined tasks
- **Think-aloud protocol**
- Quantification of insights through **coding**
- Tasks are treated as dependent variables in experiment

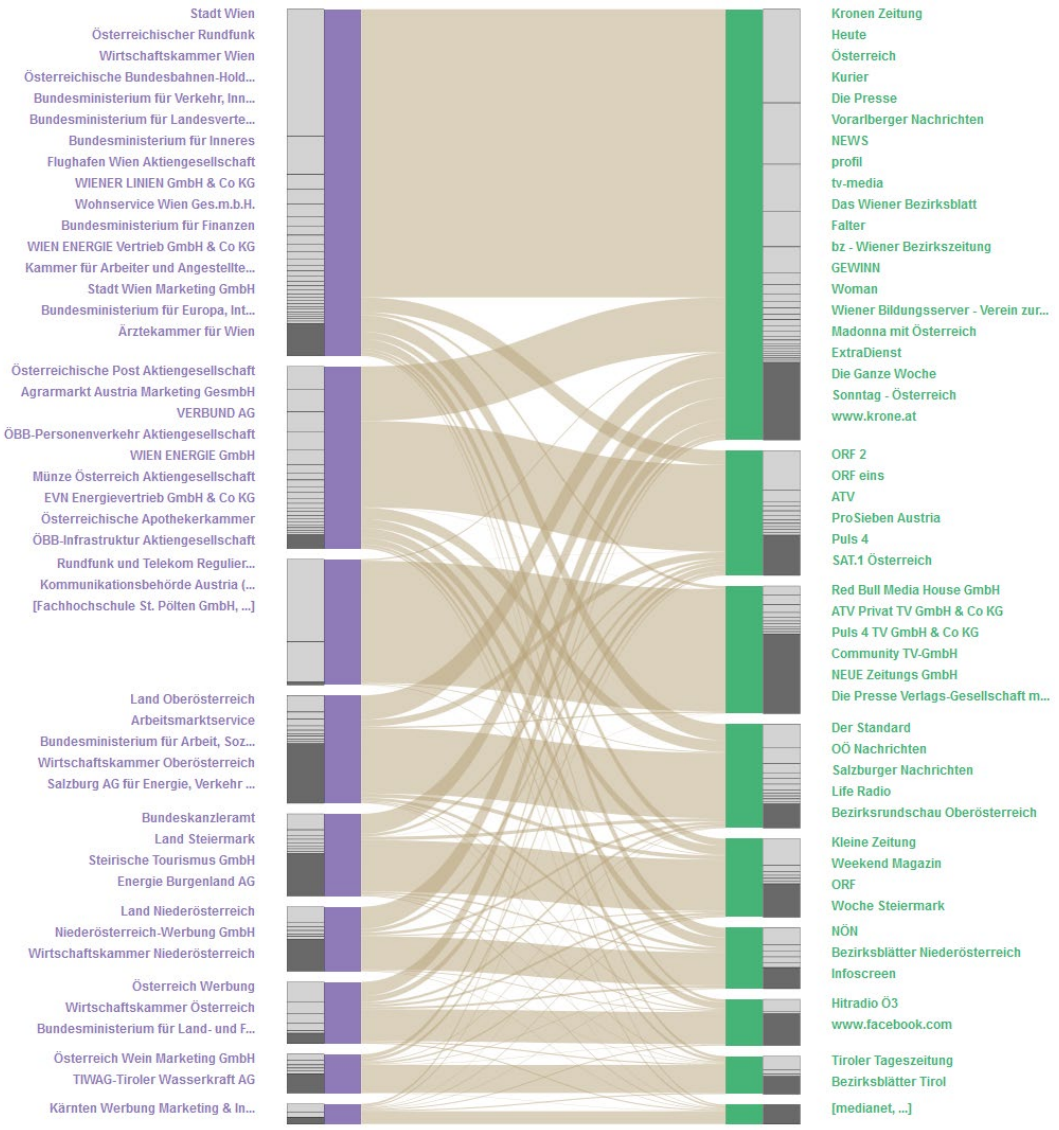
[North, Towards Measuring Visualization Insight, Computer Graphics and Applications 2006]

Example: Bipartite Graphs



[Steinböck et al., 2018]

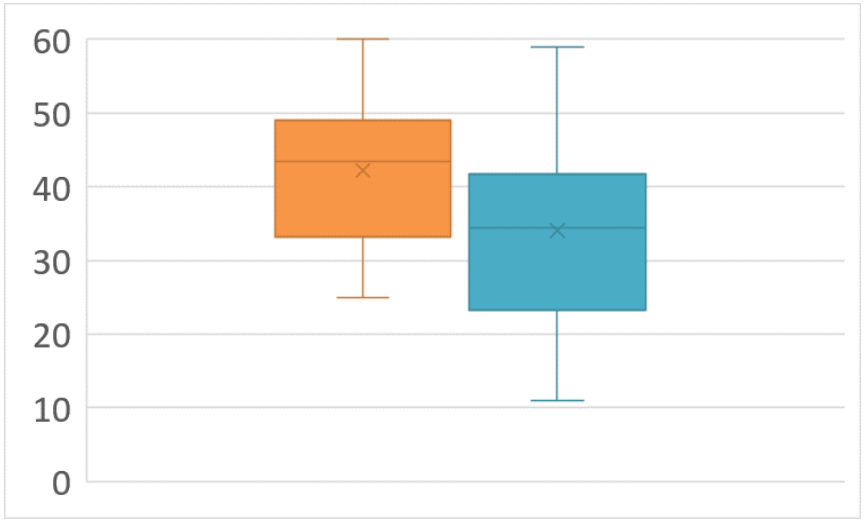
Example: Bipartite Graphs



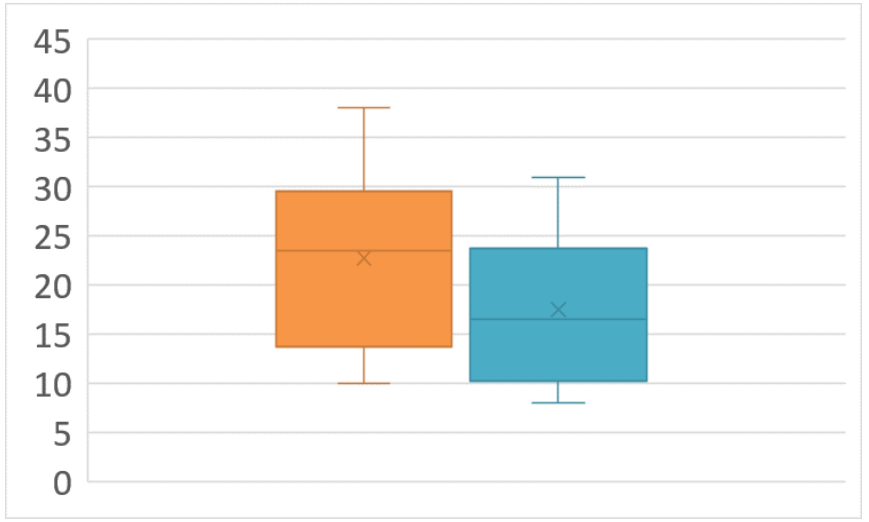
[Steinböck et al., 2018]

Example: Bipartite Graphs

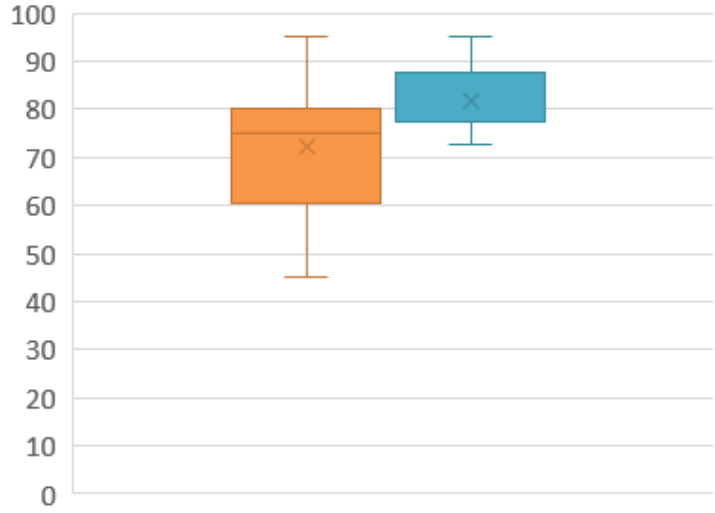
- Method:
  - 12 users from Austria, familiar with political and media landscape
  - Within-subjects (two data subsets, counter-balancing order)
  - Insight-based evaluation + System Usability Score (questionnaire)



Number of unique entities mentioned

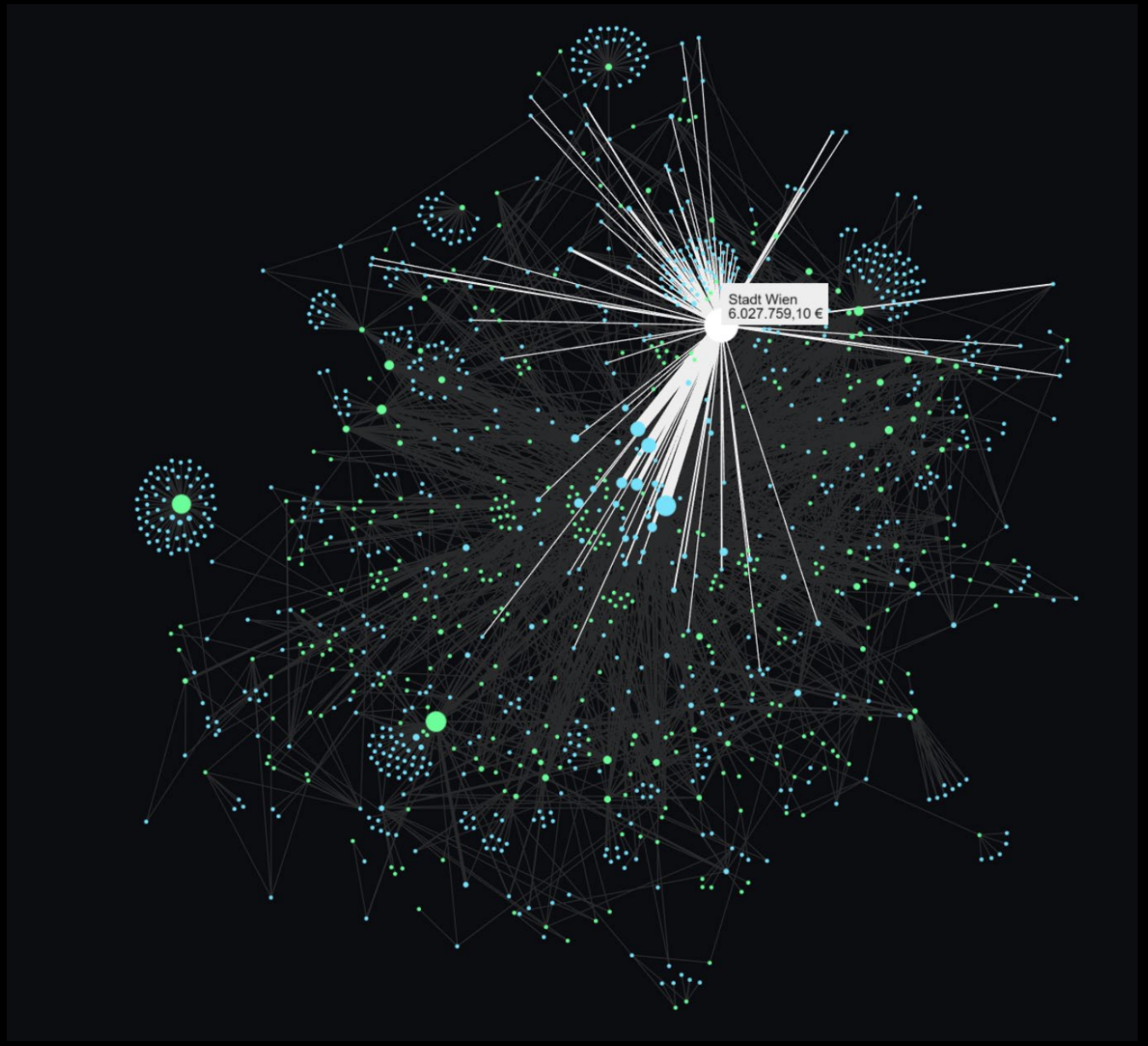


Exploration time



System Usability Score

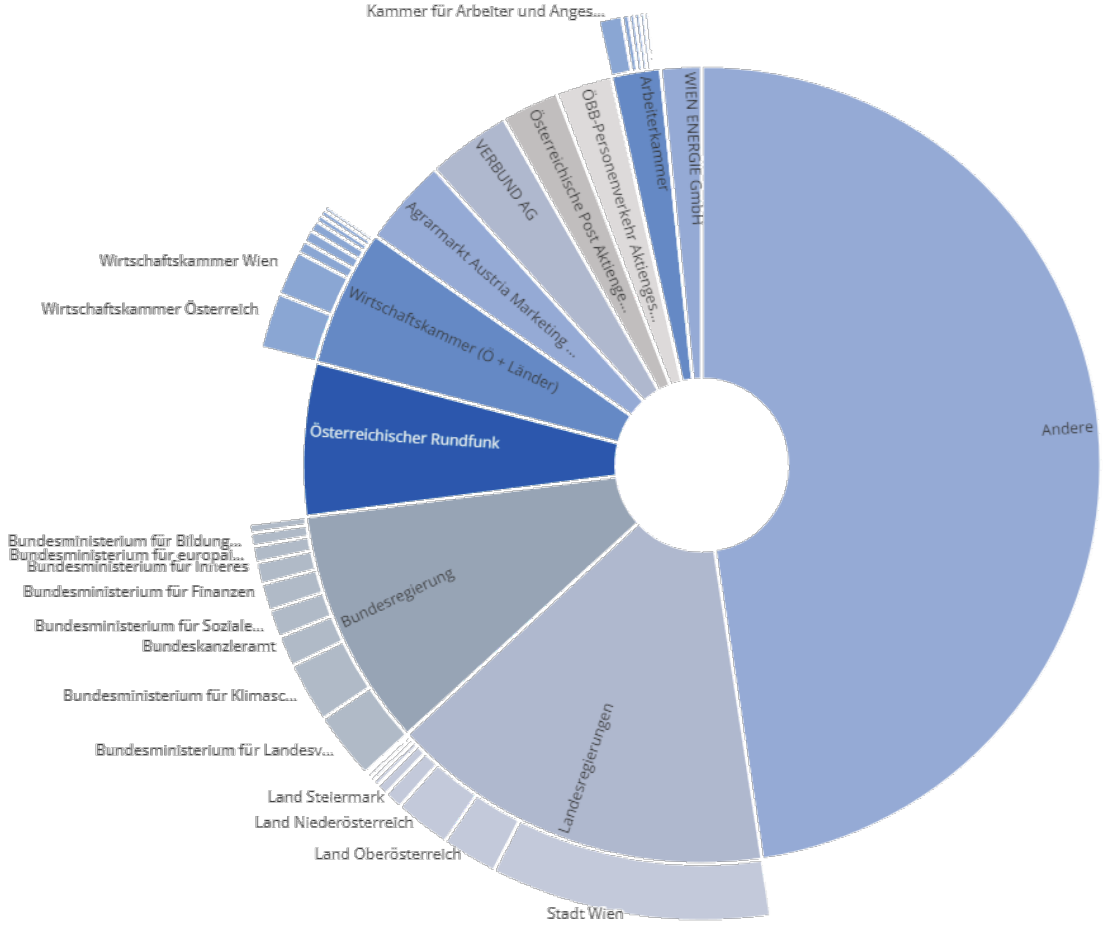
Example: Bipartite Graphs



[Unger 2019]

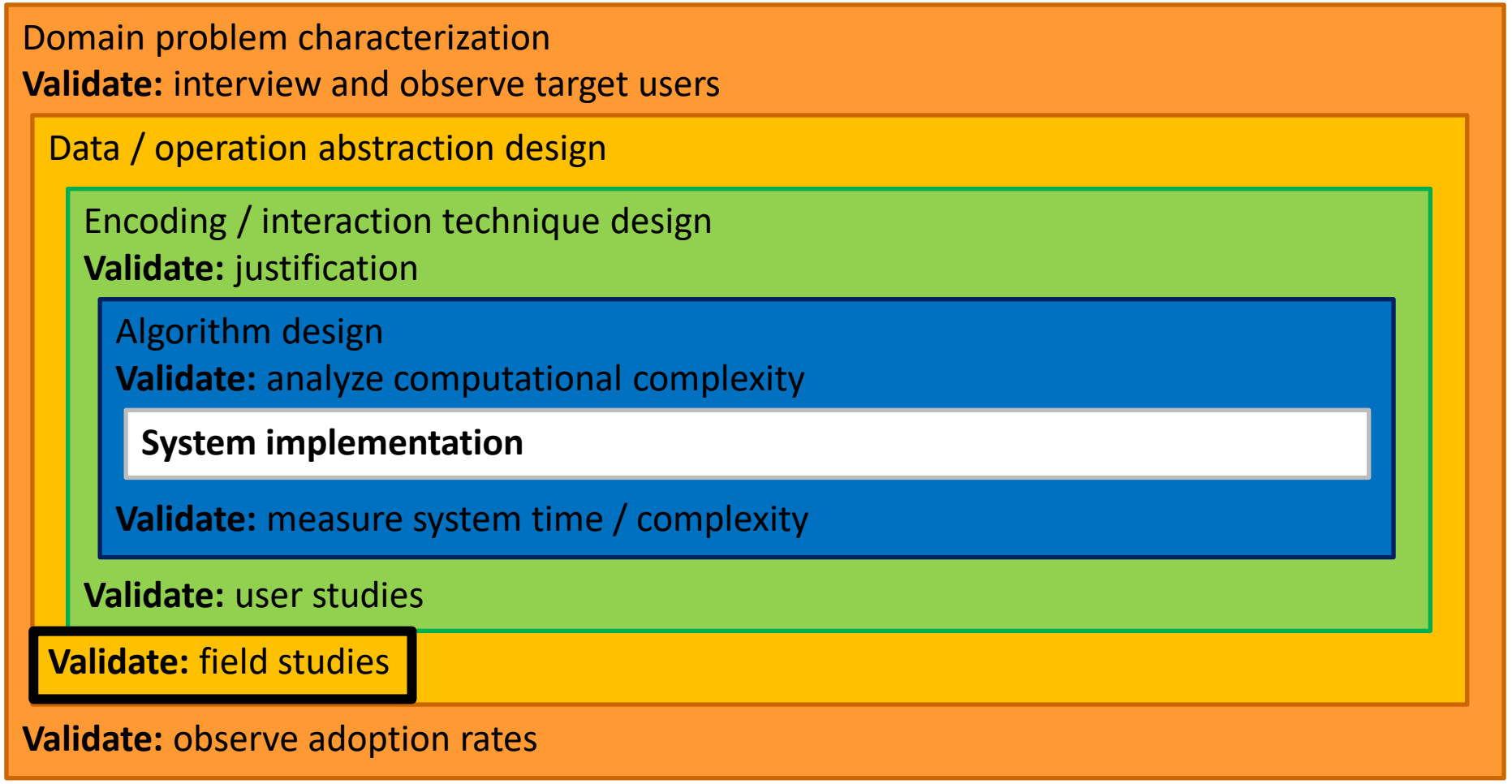
Example: Bipartite Graphs

- The visualization by RTR itself:



https://visualisierung.medientransparenz.rtr.at/top/transfers

Nested Model for Visualization Design



[Munzner, A Nested Model for Visualization Design and Validation, InfoVis 2009]

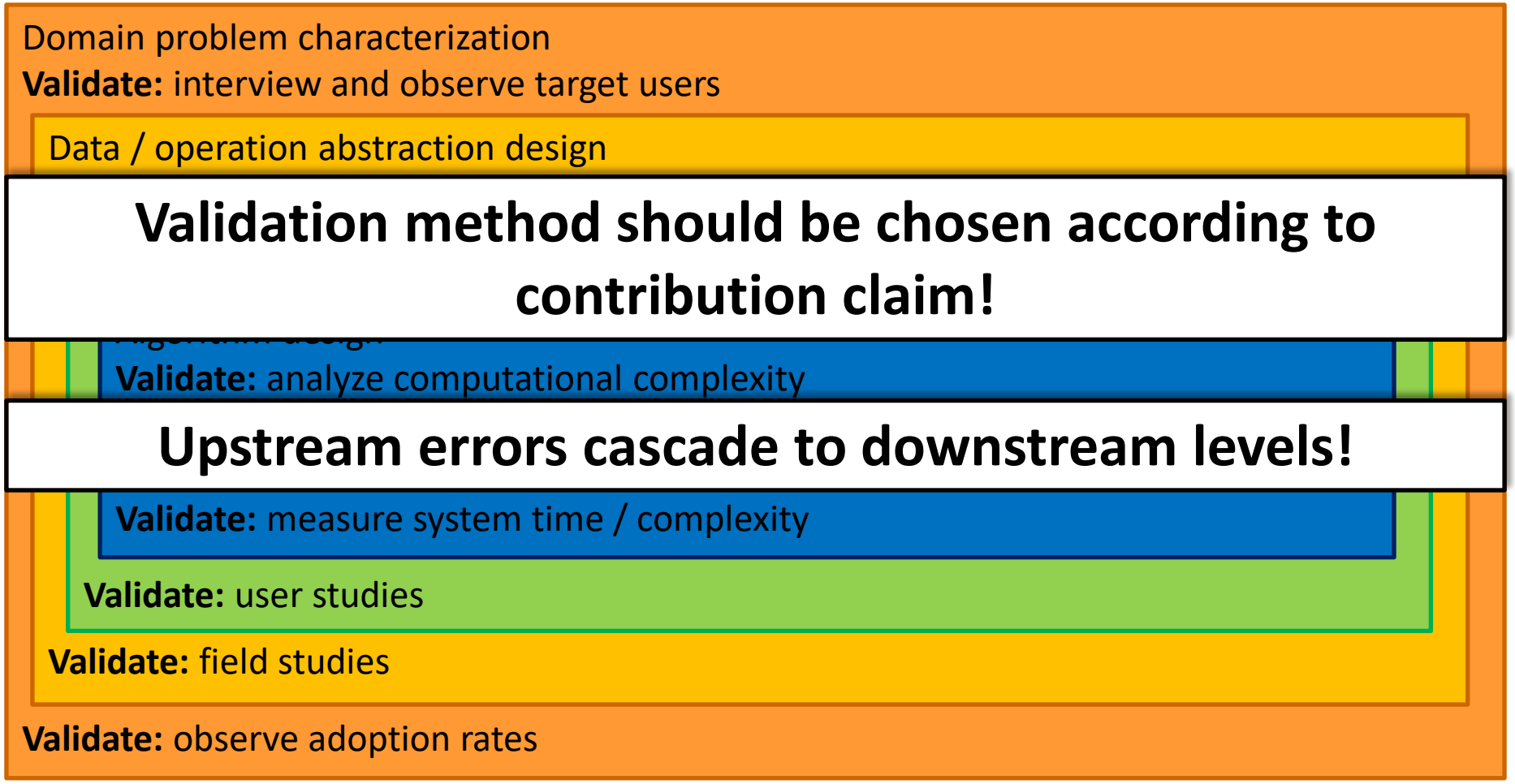
Methods: Expert Validation

- Field studies
  - Qualitative on-site
  - Users work on their own problems in their normal environment
- Usability study
  - Qualitative lab study
  - Pre-defined task
  - Observations & interview
- Field experiment
  - On-site „controlled“ study
- Informal evaluation
  - Demo to domain experts in lab
  - No pre-defined task

Summary

- Designing and validating visualizations
  - Understanding data analysis processes
    - Field and interview studies
  - Design choices & justifications
    - Expressiveness & effectiveness
  - Validation
    - Visual encoding and interaction techniques
      - Controlled user studies (lab, online, eye tracking, insight-based, …)
      - Inspection methods
      - Qualitative result inspections / case studies
    - Abstraction design
      - Field studies
      - Usability studies
      - ...

Conclusion



[Munzner, A Nested Model for Visualization Design and Validation, InfoVis 2009]

Methods Overview

	Sample	Duration	Environment	Data	Task	Realism	Precision
Controlled User Study	Large, usually non-experts	Short	Lab / Online	Quantitative (+ nested qualitative data)	Simple and short	Low	High
Usability Study	Small, domain experts	Short	Lab	Mostly qualitative	Medium complexity	Medium	Medium
Heuristic evaluation	Small, visualization experts	Short	Lab	Quantitative + qualitative	Medium complexity	Low	Medium
Informal evaluation	Small, domain experts	Short	Lab	Qualitative	---	Low	Low
Field experiment	Small, domain experts	Short	Field	Mostly qualitative	Medium complexity	Medium	Medium
Case / field study	Small, domain experts	Long	Field	Mostly qualitative	---	High	Low
Field logs	Large, domain experts	Long	Field	Mostly quantitative	---	High	Medium

General References

- Munzner, **A nested model for visualization design and validation**, TVCG 2009
- Sedlmair et al. **Design study methodology: Reflections from the trenches and the stacks**, TVCG 2012
- Lam et al., **Empirical Studies in Information Visualization: Seven Scenarios**, TVCG 2012
- Isenberg et al., **A Systematic Review on the Practie of Evaluating Visualization**, TVCG 2013
- Carpendale, **Evaluating Information Visualizations**, Information Visualization, 2008
- Ellis and Dix, **An explorative analysis of user evaluation studies in information visualisation**, BELIV 2006
- North, **Towards Measuring Visualization Insight**, Computer Graphics and Applications 2006
- Andrews, **Evaluation comes in many guises**, BELIV 2008
- Ware, **Information Visualization: Perception for Design**, 2013 (3rd edition)
- Ehrenstein and Ehrenstein, **Psychophysical Methods**, Modern Techniques in Neuroscience Research, 1999

waldner@cg.tuwien.ac.at

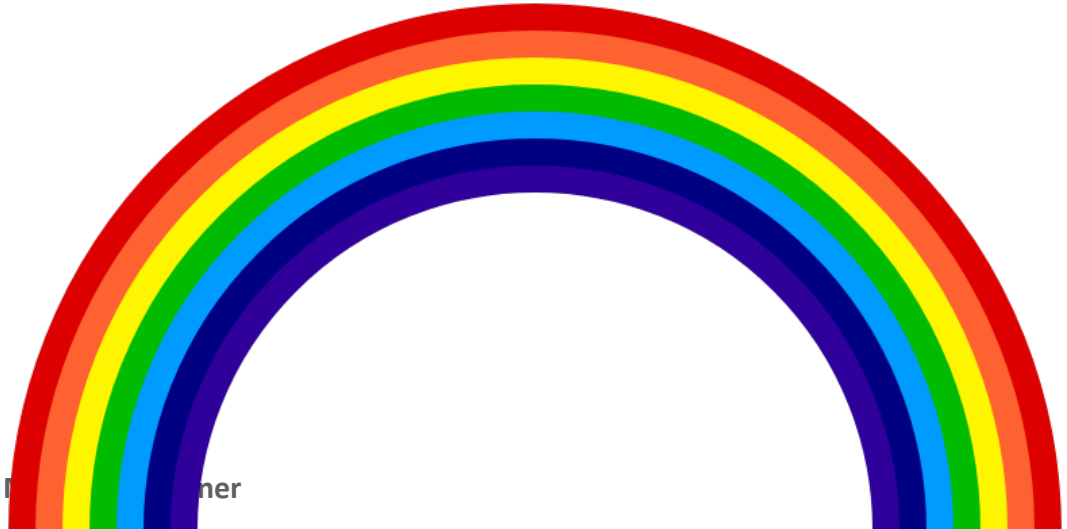
Effectiveness

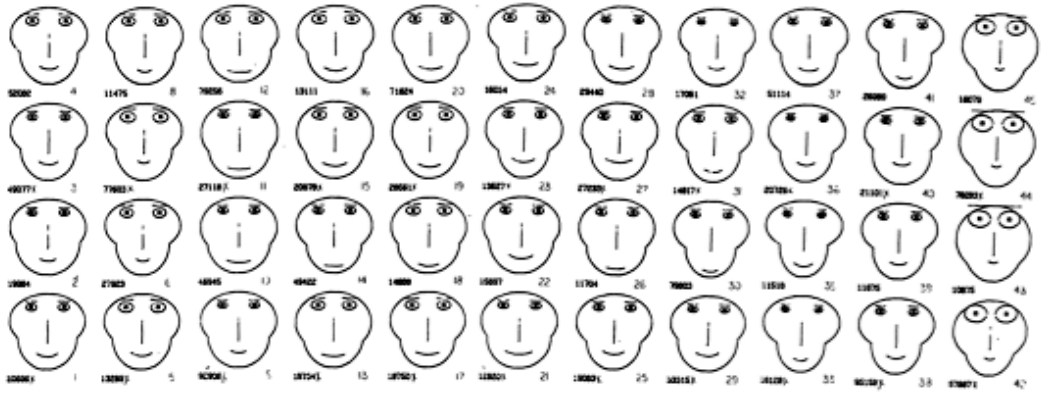
Methods from Psychophysics

Expressiveness & Effectiveness

- Expressiveness
  - Ordered attributes
    → magnitude channels
  - Categorical attributes
    → identity channels

- Effectiveness
  - Faster to interpret
  - More distinctions
  - Fewer errors





Method: Psychophysical Experiments

- Methods to measure human sensation triggered by physical stimuli

- Examples:
  - Finding a flicker frequency so that signal is perceived as steady (**absolute threshold**)
  - Finding a just noticeable difference (JND) between two colors (**difference threshold**)

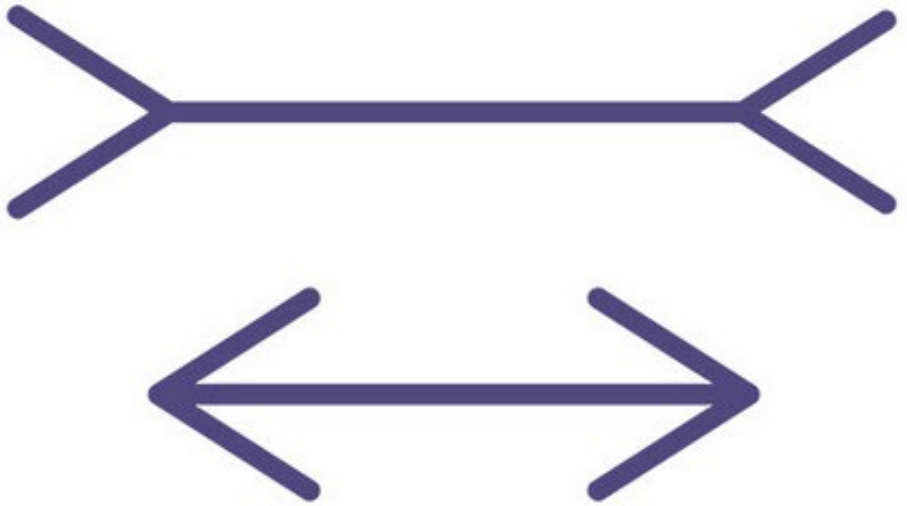
Method: Psychophysical Experiments

- Precise physical definition of stimulus pattern
- Assuming no / low instructional bias
  → often very low number of participants

- Methods for threshold measurements
  - Method of adjustment
  - Staircase procedure
  - …

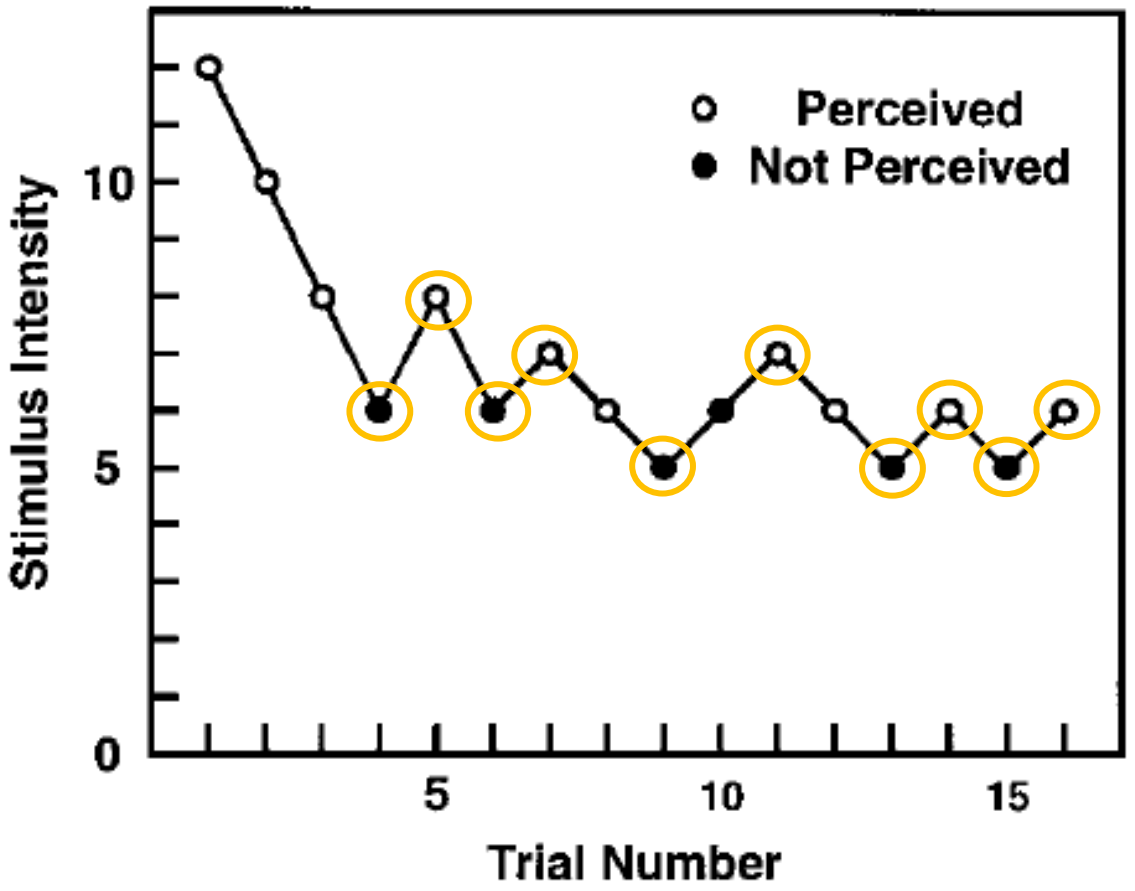
Method of Adjustment

- Observer can adjust stimulus intensity
- Until it becomes just unnoticeable / just noticeably different from a standard stimulus
- Example: user adjusts length of lower bar until it perceives to be equally long

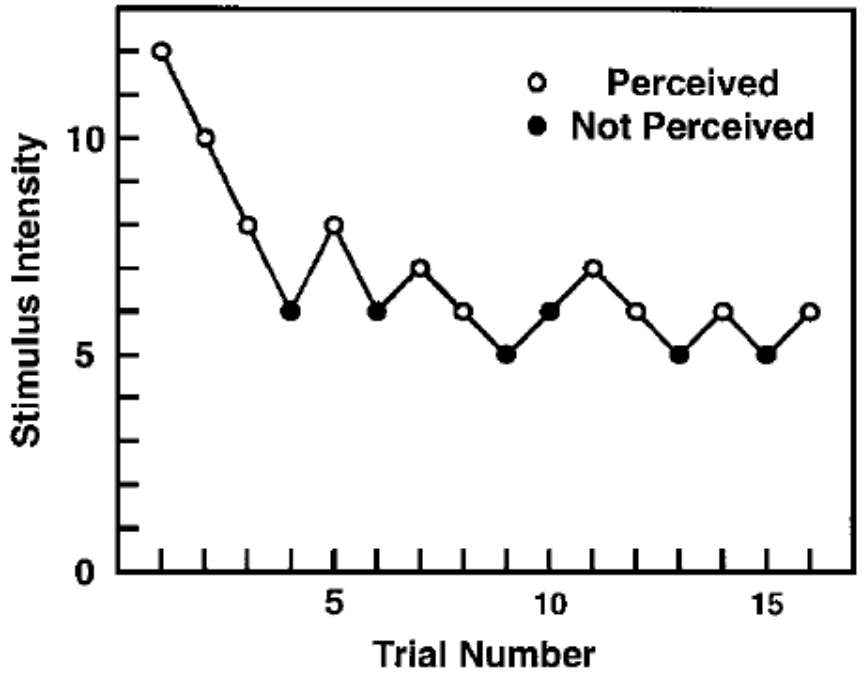


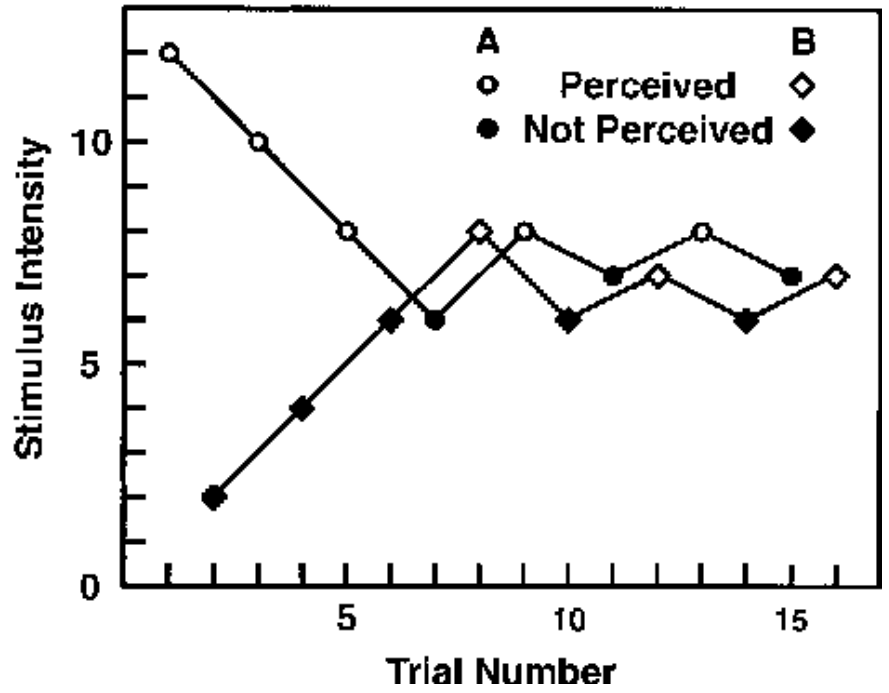
Method: Staircase Procedure

- Dynamic adaptation of stimulus level based on observer‘s response (adaptive method)
- Threshold: average stimulus intensities when observer response changed



Method: Interleaved Staircase

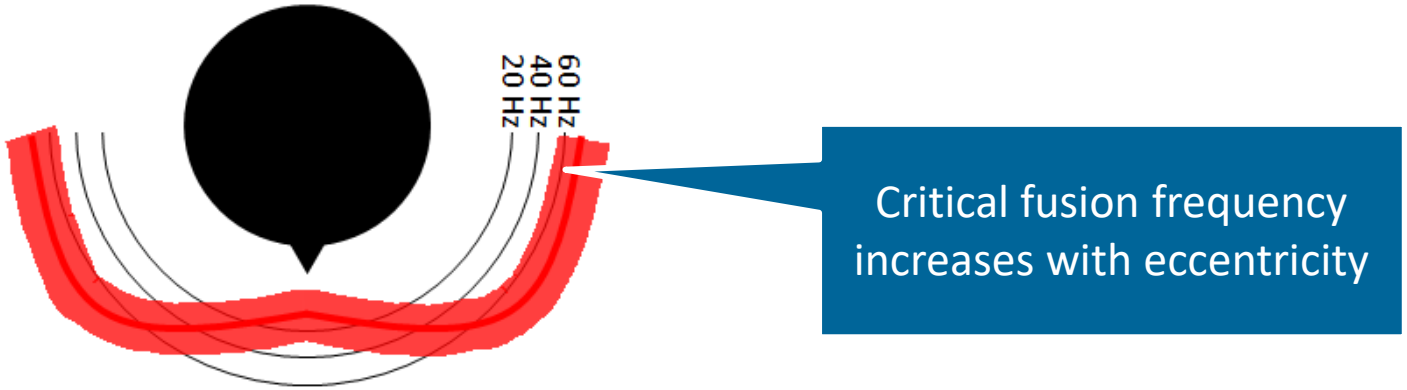




[Ehrenstein & Ehrenstein, Psychophysical Methods]

Example: High-Frequency Flicker for Guidance

- Critical fusion frequency: flicker fuses to continuous signal
- Critical fusion frequency varies across the retina

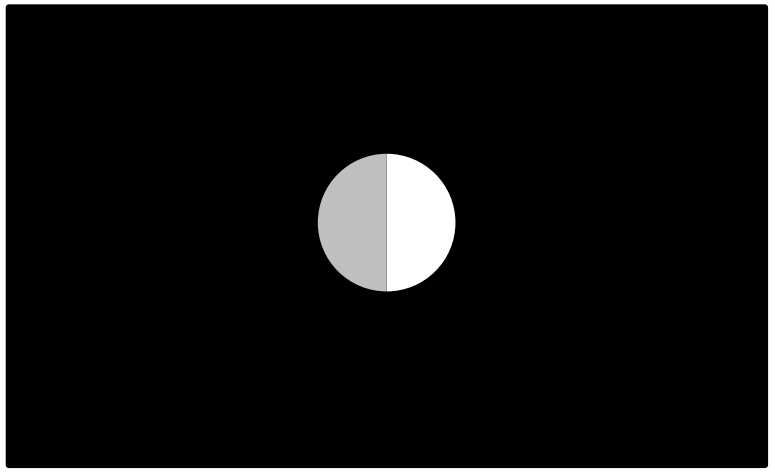


- Research question: can we use this propertly to effectively guide the user‘s attention without them noticing any visual changes to the image?

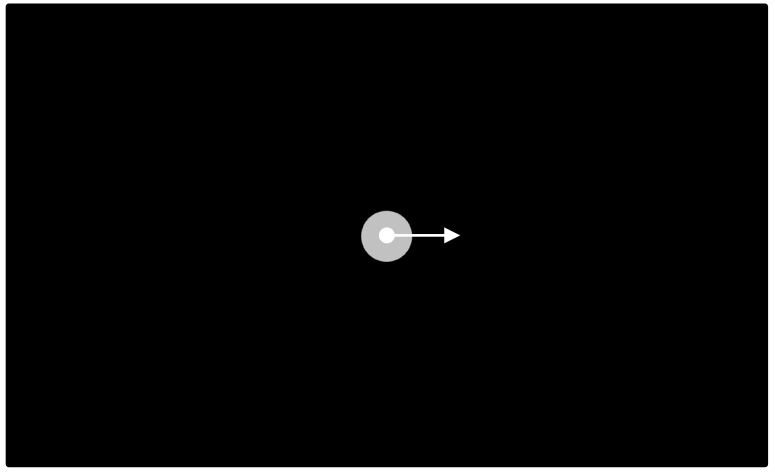
[Waldin et al., CGF 2017]

Example: High-Frequency Flicker for Guidance

- Challenge 1:
  - High-frequency flicker increases perceived brightness
  - Method of adjustment



- Challenge 2:
  - Fovea is very small
  - Method: staircase procedure



[Waldin et al., CGF 2017]