Information Visualization Advanced Topics

Tamara Munzner
Department of Computer Science
University of British Columbia

Week 14 sync class, 4 Dec 2025

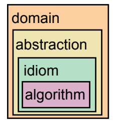
https://www.cs.ubc.ca/~tmm/courses/547-25/

Today: Week 14

- UBC InfoVis group research 1: problem-driven
- break
- UBC InfoVis group research 2: Vis & ML intersections
- next steps

Applying visualization to real-world problems

- today: research highlights from my own group
- nested model: going down, starting from top
  - map from domain to abstraction level
    - crucial & difficult, iterative process
  - select appropriate idioms
    - or create new ones if necessary
- many case studies
  - different domains
  - different methods

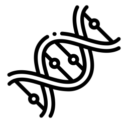


Four case studies of problem-driven work

- in-car networks

- facilities management

- biology (x2)



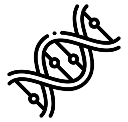
Car by AIGA from the Noun Project Business by Colourcreatype from the Noun Project Biology by tezar tantular from the Noun Project

Four case studies of problem-driven work

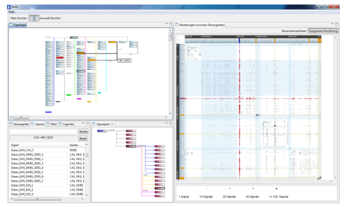
- in-car networks

- facilities management

- biology



Car by AIGA from the Noun Project Business by Colourcreatype from the Noun Project Biology by tezar tantular from the Noun Project



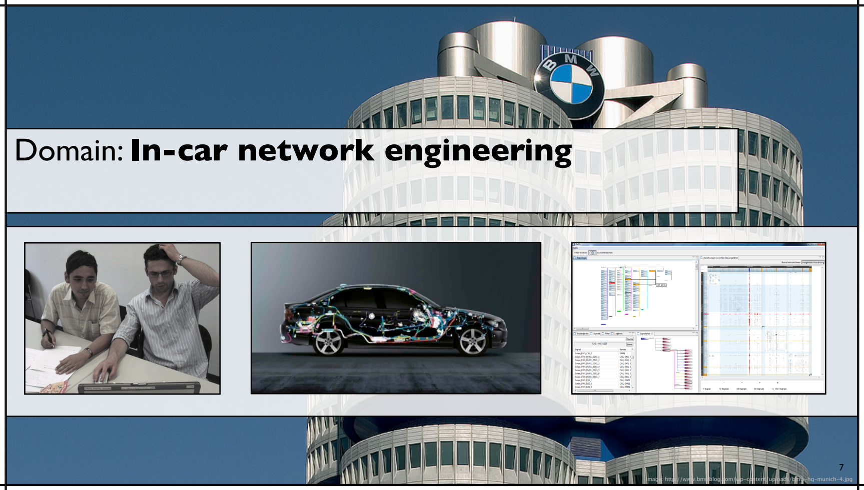
RelEx

Visualization for Actively Changing Overlay Network Specifications

joint work with:
Michael Sedlmair, Annika Frank, Andreas Butz

http://www.cs.ubc.ca/labs/imager/tr/2012/relex/

RelEx: Visualization for Actively Changing Overlay Network Specifications.
Sedlmair, Frank, Butz, Munzner. IEEE TVCG 18(12): 2729-2738, 2012 (Proc. InfoVis 2012).

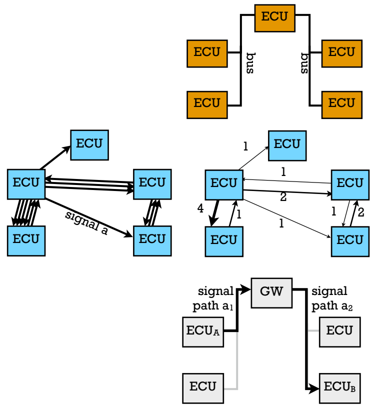


Abstractions

DATA
In-car Electronics

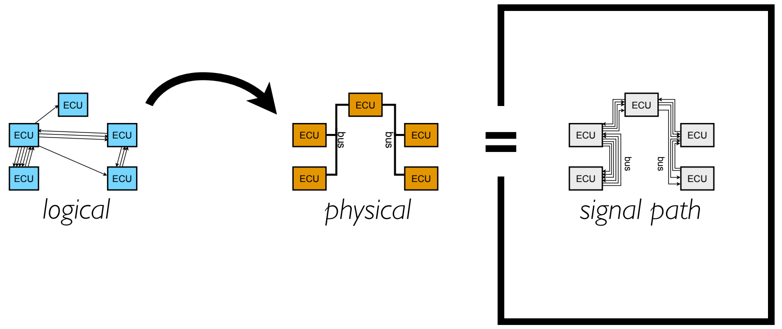
Data Abstraction: 3 Networks

- **physical** network
  - 100 nodes: *Electronic Control Units*
  - 10-15 hyperedges: *bus systems*
  - hardware engineers
- **logical** network
  - same nodes
  - 10,000 multigraph edges: *signals*
  - 1,000 weighted edges: *signal counts*
  - software engineers
- **overlay** network
  - maps logical onto physical
  - 30,000 edges: *signal paths*
  - target engineers



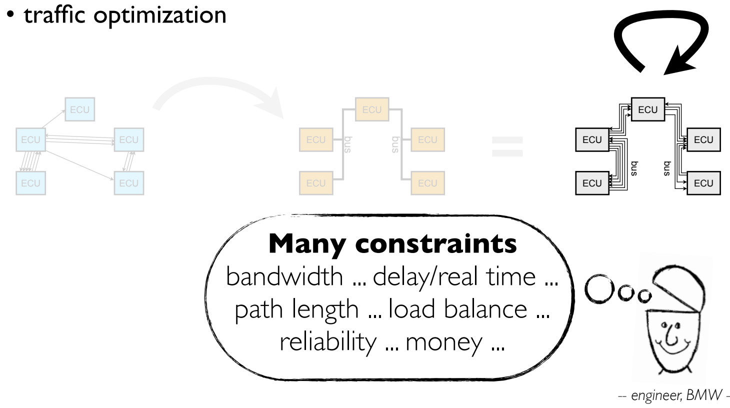
Task Abstraction: Mapping

- specify overlay network that maps logical onto physical



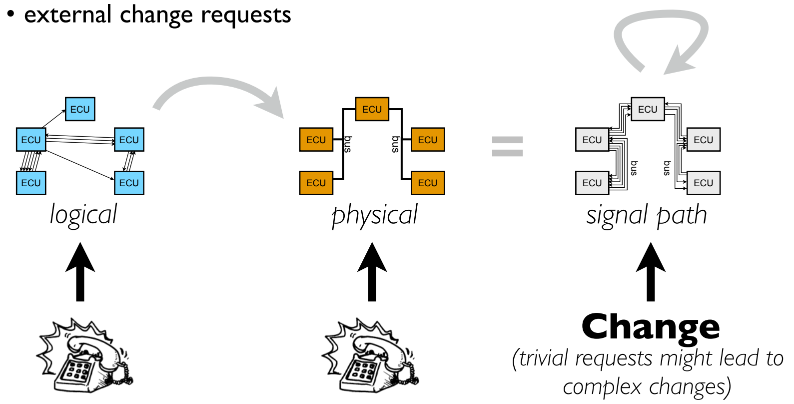
Task Abstraction: Optimizing

- traffic optimization

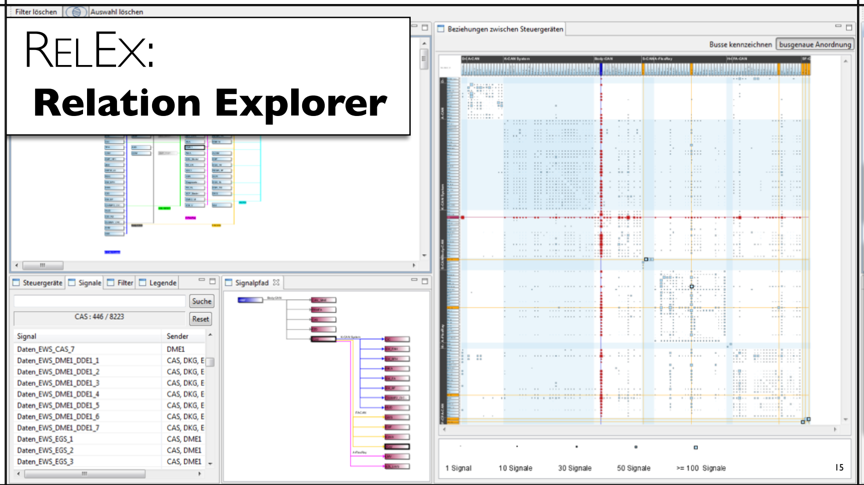


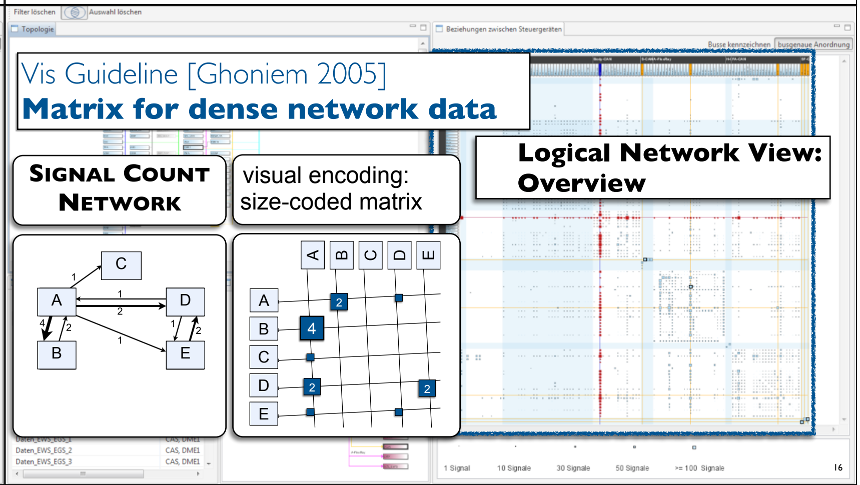
Task Abstraction: Changing

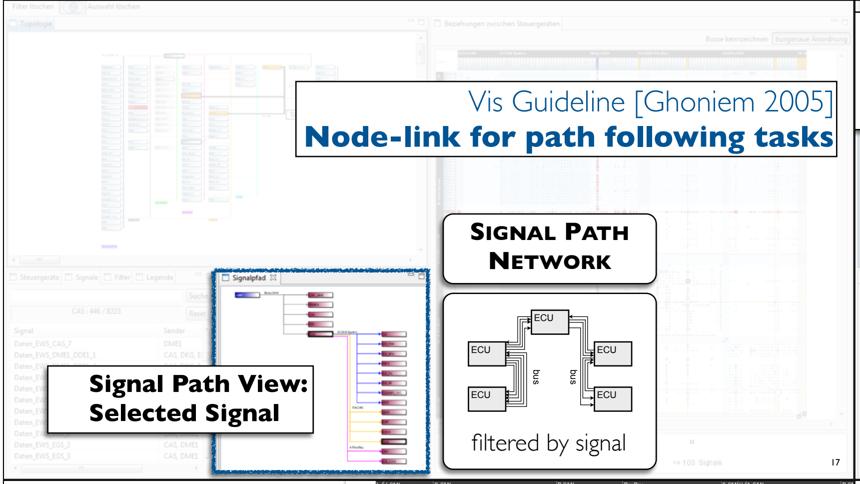
- external change requests



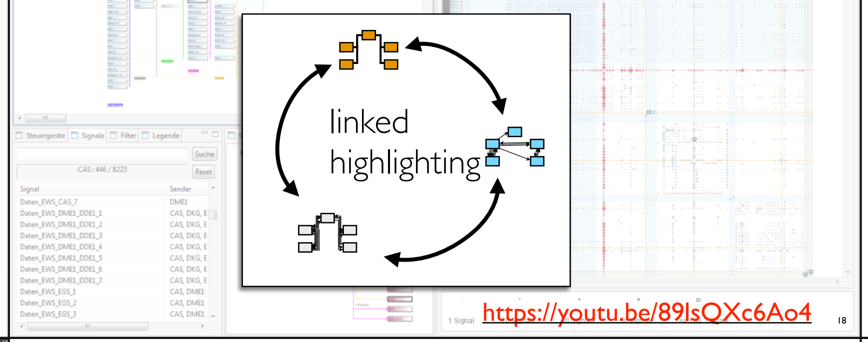
Idioms





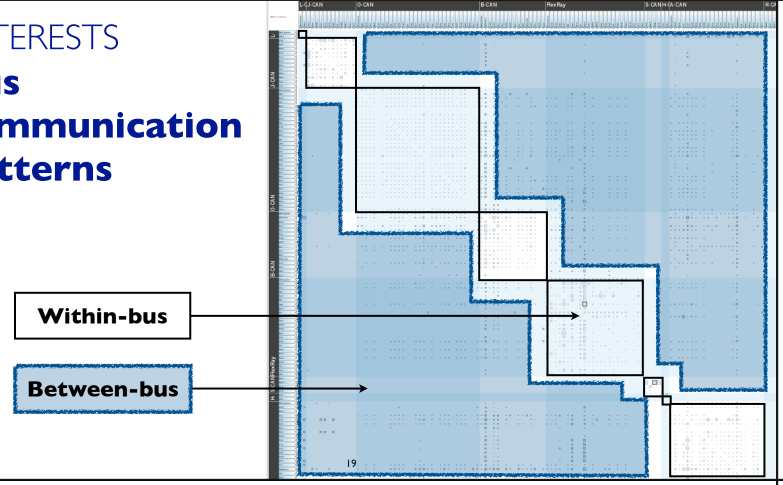


Interaction Idiom: Cross-Network Relations



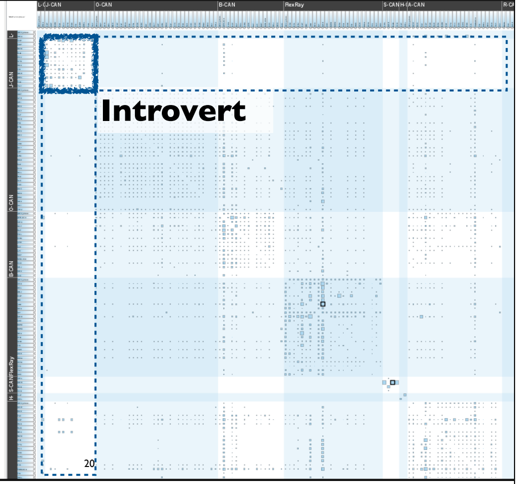
https://youtu.be/89lsQXc6Ao4

Interests
Bus communication patterns



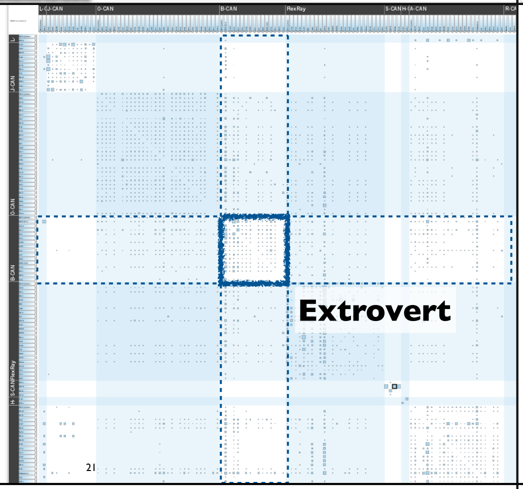
Interests
Bus communication patterns

introvert
vs.
extrovert



Interests
Bus communication patterns

introvert
vs.
extrovert



Abstraction Innovation

Previous Work

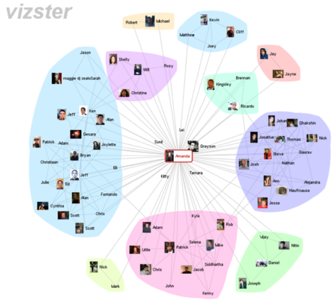
Focus on social network analysis

- radically different task and data abstractions

Task Abstraction

Social Network Analysis Domain

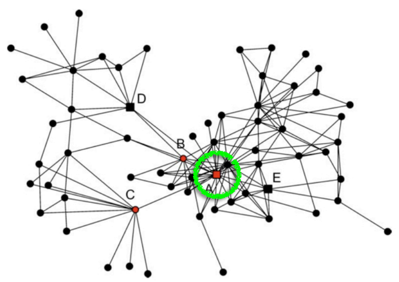
- find clusters



Task Abstraction

Social Network Analysis Domain

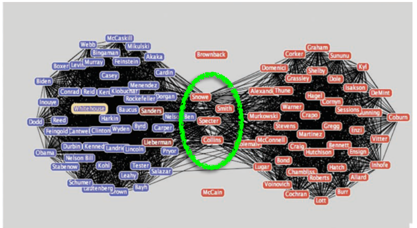
- find clusters
- find high-degree nodes



Task Abstraction

Social Network Analysis Domain

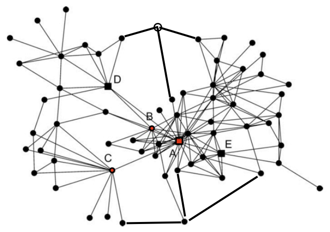
- find clusters
- find high-degree nodes
- find bridge nodes



Task Abstraction

Social Network Analysis Domain

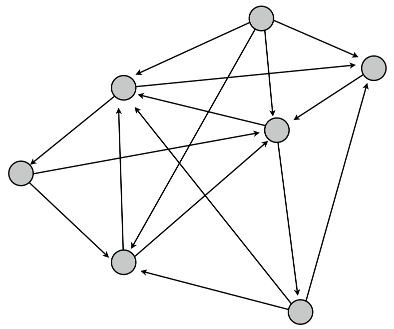
- find clusters
- find high-degree nodes
- find bridge nodes
- understand temporal dynamics
  - passively notice changes



Data Abstraction

Social Network Analysis Domain

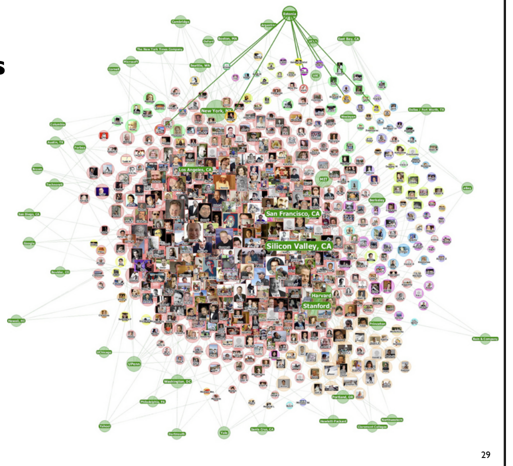
- single graph



Data Abstraction

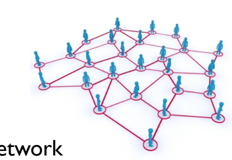
Social Network Analysis

- single graph
- scalability challenge: nodes



Abstraction Differences

Social Network Analysis vs Overlay Network Optimization



- data
  - single network
  - node scalability
    - sparse edges
- task
  - find clusters, high-degree nodes, bridge nodes
  - passive changes 

- data
  - three related networks
    - physical, logical, overlay
  - path scalability
    - dense edges, few nodes
- task
  - traffic optimization
  - active changes 

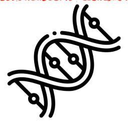
Q&A

Four case studies of problem-driven work

- in-car networks

- facilities management

- biology



Michael Oppermann

Ocupado

Visualizing Location-Based Counts Over Time Across Buildings

http://www.cs.ubc.ca/labs/imager/tr/2020/ocupado/

Ocupado:Visualizing Location-Based Counts Over Time Across Buildings.

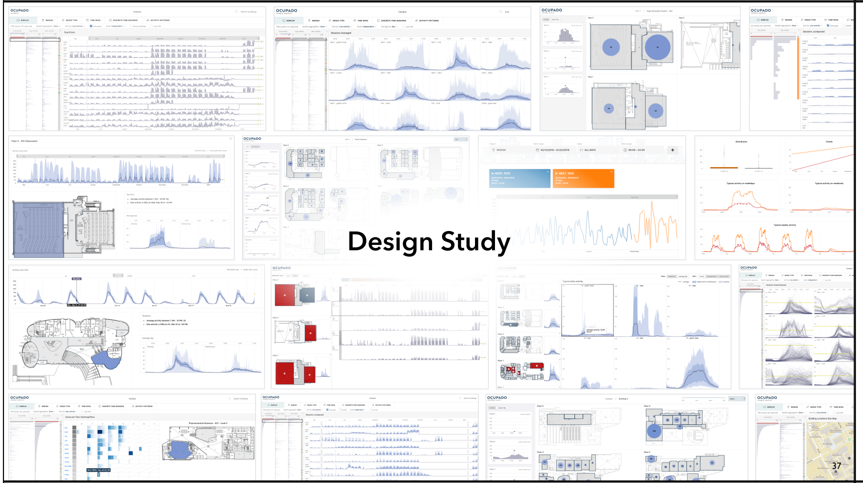
Oppermann and Munzner. Computer Graphics Forum (Proc. EuroVis 2020) 39(3):127-138 2020.

Location-Based Counts

Previous measurement required physical counting or installation of additional hardware.

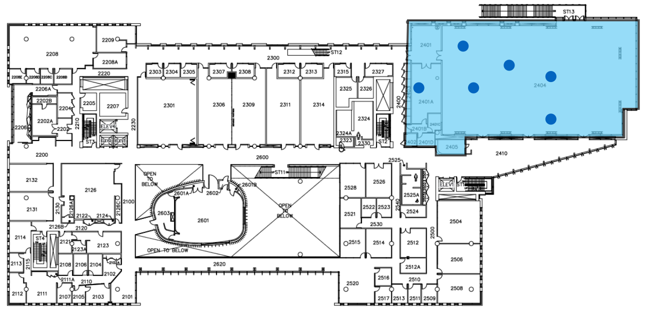
Previous measurement required physical counting or installation of additional hardware.

Previous visualization attempts were limited in space and time.

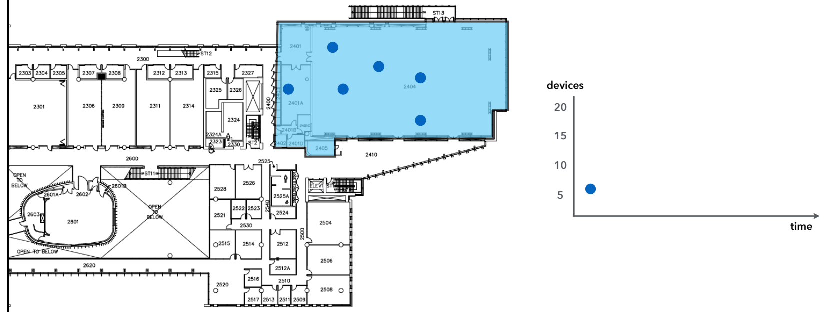


Design Study

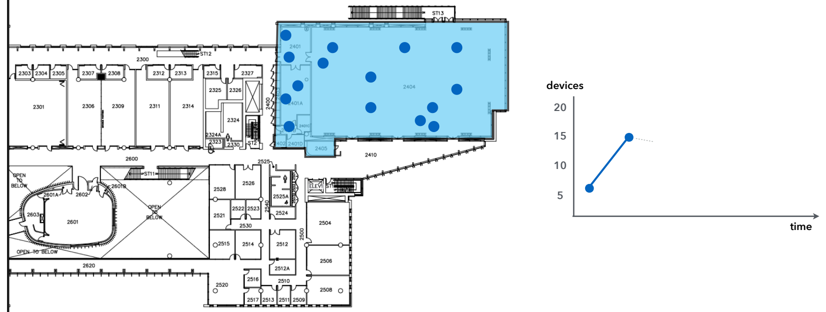
WiFi Connections: Location-Based Counts



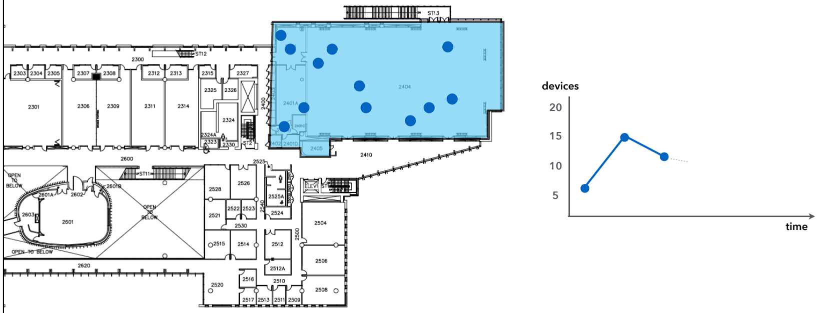
WiFi Connections: Location-Based Counts



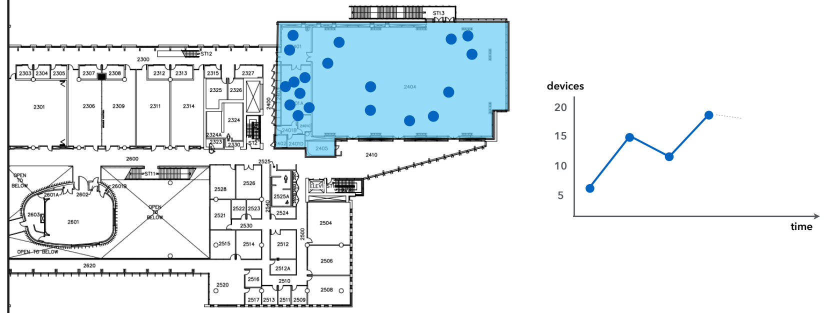
WiFi Connections: Location-Based Counts



WiFi Connections: Location-Based Counts



WiFi Connections: Location-Based Counts



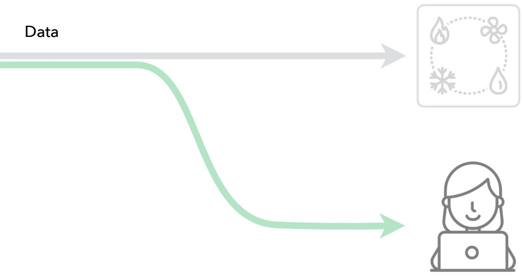
WiFi Connections: Location-Based Counts

WiFi Connections: Location-Based Counts

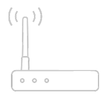
Location-Based Counts

- Regular intervals (e.g., every 5 minutes)
- Spatial hierarchy (Zone → Floor → Building → Campus)
- No trajectories or device identifiers are recorded
- Intrinsic privacy advantages

Automated HVAC control



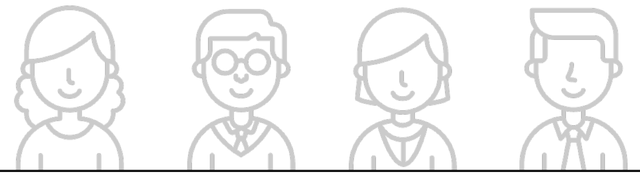
Decision making



WiFi connections as a proxy for occupancy

WiFi connections as a proxy for occupancy

Interviews with potential stakeholders

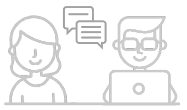


Focus Domains

- Space planning
- Building management
- Custodial services
- Classroom management
- Data quality control

Focus Domains

- Space planning
- Building management
- Custodial services
- Classroom management
- Data quality control



Semi-structured discussions and live demos

Tasks

- ✓ **Confirm** assumptions or previous observations.
  Do students occupy room x in evenings or on weekends?

Tasks

- ✓ **Confirm** assumptions or previous observations.
- ✓ **Monitor** the current/recent utilization rate.
  Which rooms are empty/busy?

Tasks

- ✓ **Confirm** assumptions or previous observations.
- ✓ **Monitor** the current/recent utilization rate.
- ✓ **Communicate** space usage and justify decisions.
  Space usage improved after renovation.

Tasks

- ✓ **Confirm** assumptions or previous observations.
- ✓ **Monitor** the current/recent utilization rate.
- ✓ **Communicate** space usage and justify decisions.
- ✓ **Validate** the data (quality control).
  Check minimum size of a room that can be captured.

Spatial and Temporal Data Granularities

Visualization Prototypes

Sandbox
Data sketches, static data export

Time

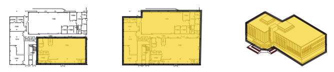
Visualization Prototypes

Sandbox
Data sketches, static data export

- original plan: different interface for each stakeholder
- realization: task & data abstractions match multiple stakeholders
  - if slice by space & time granularity

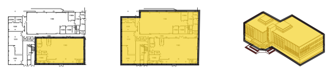
Spatial and Temporal Data Granularities

Regions of interest



Spatial and Temporal Data Granularities

Regions of interest



Periods of interest

Mondays
Weekdays
last 12 hours
Summer term
Fr 8-10am
Weekends

Visualization Prototypes

Sandbox
Data sketches, static data export

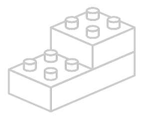
Campus Explorer
Live-data stream, cross-building analysis

Building Recent

Building Long-term

Region Compare

Time



Reusable Visualization Components

Reusable Visualization Components

Layout	Visual Encoding	Facet	Comparisons
	Sparkline	Juxtaposition	Repeating patterns, trends, outliers (contiguous)

Reusable Visualization Components

Layout	Visual Encoding	Facet	Comparisons
	Sparkline	Juxtaposition	Repeating patterns, trends, outliers (contiguous)
	Box-plot-bars	Juxtaposition	Repeating patterns, trends, outliers *(non-contiguous)*

Reusable Visualization Components

Layout	Visual Encoding	Facet	Comparisons
	Sparkline	Juxtaposition	Repeating patterns, trends, outliers (contiguous)
	Box-plot-bars	Juxtaposition	Repeating patterns, trends, outliers *(non-contiguous)*
	Confidence band line chart	Aggregation	Typical utilization profiles

Reusable Visualization Components

Layout	Visual Encoding	Facet	Comparisons
	Sparkline	Juxtaposition	Repeating patterns, trends, outliers (contiguous)
	Box-plot-bars	Juxtaposition	Repeating patterns, trends, outliers *(non-contiguous)*
	Confidence band line chart	Aggregation	Typical utilization profiles
	Superimposed line chart	Superposition	Within-session patterns, outliers

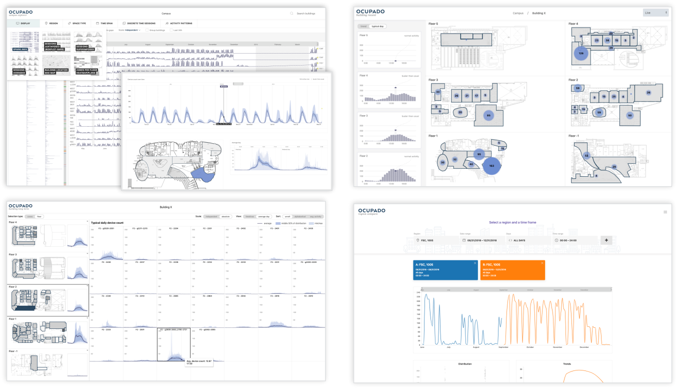
Reusable Visualization Components

	Layout	Visual Encoding	Facet	Comparisons
Temporal		Sparkline	Juxtaposition	Repeating patterns, trends, outliers (contiguous)
		Box-plot-bars	Juxtaposition	Repeating patterns, trends, outliers *(non-contiguous)*
		Confidence band line chart	Aggregation	Typical utilization profiles
		Superimposed line chart	Superposition	Within-session patterns, outliers

Reusable Visualization Components

	Layout	Visual Encoding	Facet	Comparisons
Temporal		Sparkline	Juxtaposition	Repeating patterns, trends, outliers (contiguous)
		Box-plot-bars	Juxtaposition	Repeating patterns, trends, outliers *(non-contiguous)*
		Confidence band line chart	Aggregation	Typical utilization profiles
		Superimposed line chart	Superposition	Within-session patterns, outliers
Spatial		Floor plan with symbols	Superposition	Within local spatial neighborhood
		Spatial heatmap	Containment (nested)	Across distributed regions

Ocupado Interfaces



Ocupado: Visualizing Location-Based Counts Over Time Across Buildings

Michael Oppermann
Tamara Munzner

THE UNIVERSITY OF BRITISH COLUMBIA
DESIGNING for PEOPLE
Project partner: Sensible Building Science

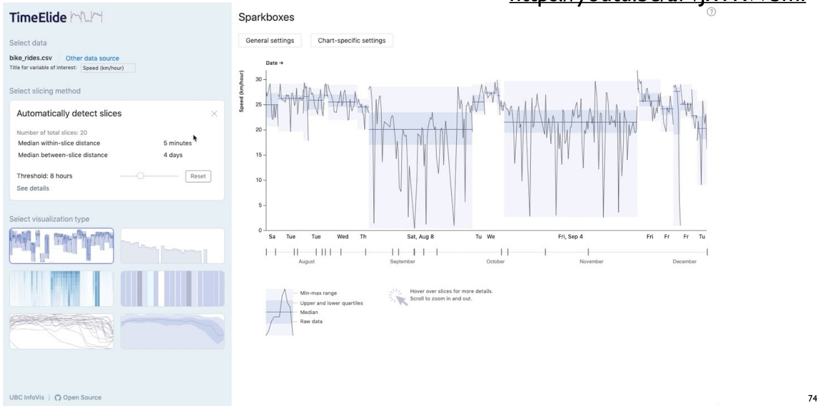
https://youtu.be/KcwjVK8eUdw?t=83

Ocupado Contributions

- Analysis and abstraction of data and tasks for studying space utilization
- Ocupado, a set of visual decision support tools
- Generalizable design choices for visualizing non-trajectory spatiotemporal data relating to large-scale indoor environments

Spinoff toolkit paper: TimeElide

https://youtu.be/a7vJxTRWbml



Q&A

Four case studies of problem-driven work

- in-car networks
- facilities management
- biology

Car by AIGA from the Noun Project Business by Colourcreatype from the Noun Project Biology by tezar tantular from the Noun Project

Zipeng Liu
Shing Hei Zhan

Aggregated Dendrograms

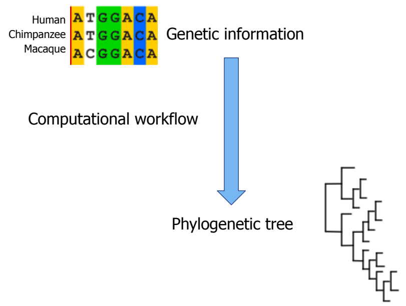
for Visual Comparison Between Many Phylogenetic Trees

http://www.cs.ubc.ca/labs/imager/tr/2019/adview

Aggregated Dendrograms for Visual Comparison Between Many Phylogenetic Trees.
Liu, Zhan, Munzner. IEEE Trans. Visualization and Computer Graphics (TVCG) 26(9):2732-2747, 2019.

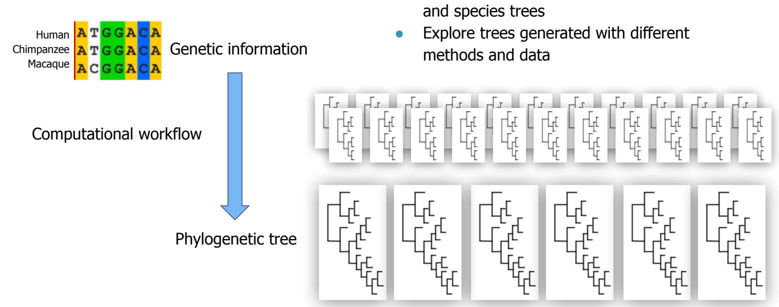
Phylogenetic tree

Evolutionary relationships of organisms

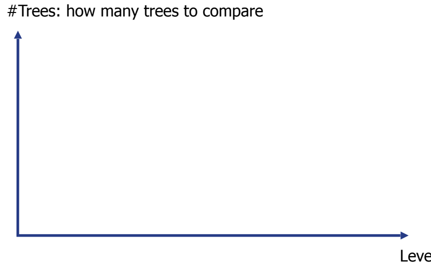


Many phylogenetic trees

- Understand relationships between genes and species trees
- Explore trees generated with different methods and data



Scalability of Existing Tree Comparison Systems

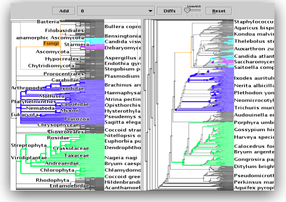


Scalability of Existing Tree Comparison Systems

#Trees: how many trees to compare

Pairs — Few in full

Simplified structure | Full topology | Level of detail (LoD): how much details are visible



TreeJuxtaposer.
Munzner, Guimbretière, Zhang, Zhou.
SIGGRAPH 2003

Scalability of Existing Tree Comparison Systems

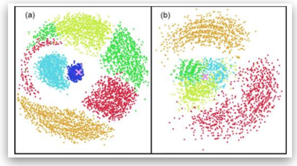
#Trees: how many trees to compare

Thousands / Hundreds / Dozens / Pairs

Many as points

Few in full

Single point | Simplified structure | Full topology | Level of detail (LoD): how much details are visible



Tree space.
Hillis, Health, John.
Systematic Biology 2005.

Scalability of Existing Tree Comparison Systems

#Trees: how many trees to compare

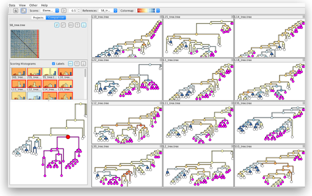
Thousands / Hundreds / Dozens / Pairs

Many as points

Dozens at multi-scale

Few in full

Single point | Simplified structure | Full topology | Level of detail (LoD): how much details are visible



Interactive visual comparison of multiple trees.
Bremm, Landesberger, Heß, Schreck, Weil, Hamacher.
VAST 2011.

Comparing many phylogenetic trees

#Trees: how many trees to compare

Thousands / Hundreds / Dozens / Pairs

Many as points

Hundreds / thousands at multi-scale?

Dozens at multi-scale

Few in full

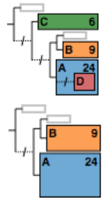
Single point | Simplified structure | Full topology | Level of detail (LoD): how much details are visible

Contributions include idiom & algorithm levels

- Data and task abstractions for comparison of phylogenetic trees

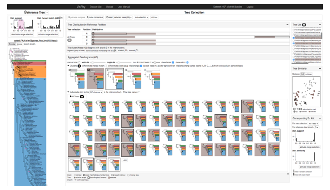
Contributions include idiom & algorithm levels

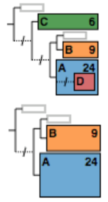
- Data and task abstractions for comparison of phylogenetic trees
- A new visual encoding: **Aggregated Dendrogram**
  - Compact tree representation that focuses on selected subtrees
  - Adapts to available screen space



Contributions include idiom & algorithm levels

- Data and task abstractions for comparison of phylogenetic trees
- A new visual encoding: **Aggregated Dendrogram**
  - Compact tree representation that focuses on selected subtrees
  - Adapts to available screen space
- A multi-view interactive tool: **ADView**
  - Covers multiple levels of details for tree comparison



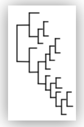


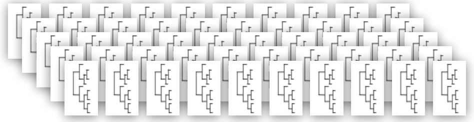
Data & Tasks

- Tree data
- Two crucial tasks

Tree data

Reference tree vs. Tree collection





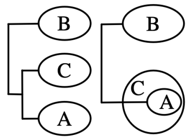
Two crucial tasks

Topological relationships between subtrees / leaf nodes

Two crucial tasks

Topological relationships between subtrees / leaf nodes

Separated Nested



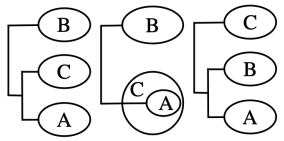
Two crucial tasks

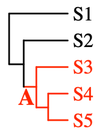
Topological relationships between subtrees / leaf nodes

- Topological distance

Leaf node memberships compared to reference tree

Separated Nested





Reference

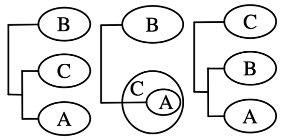
Two crucial tasks

Topological relationships between subtrees / leaf nodes

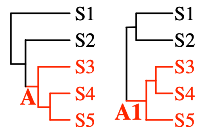
- Topological distance

Leaf node memberships compared to reference tree

Separated Nested



Exact match



Reference Tree1

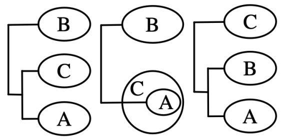
Two crucial tasks

Topological relationships between subtrees / leaf nodes

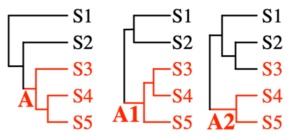
- Topological distance

Leaf node memberships compared to reference tree

Separated Nested



Exact match Partial match



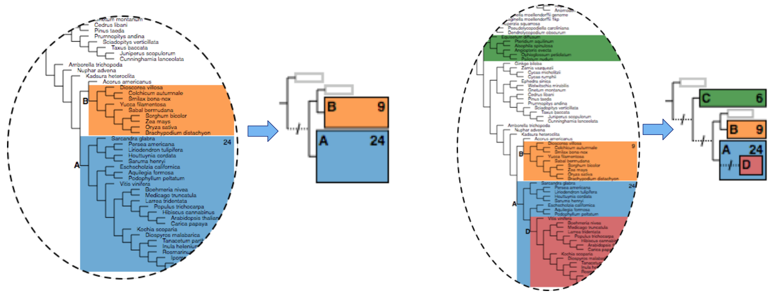
Reference Tree1 Tree2

Aggregated Dendrogram (AD)

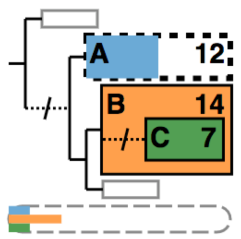
- Intuition
- Visual design

Intuition

Use glyphs to compress a tree according to user selections

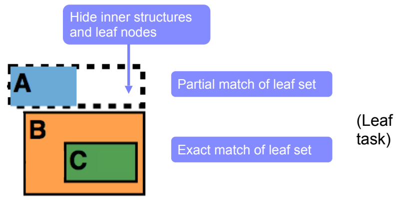


Visual design: focus + context



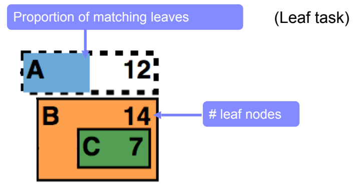
Visual design: focus + context

- Focus
  - Selected subtrees



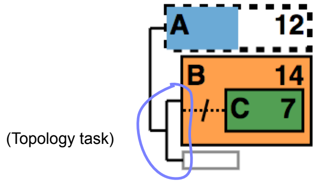
Visual design: focus + context

- Focus
  - Selected subtrees



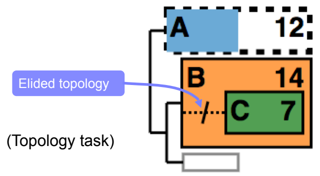
Visual design: focus + context

- Focus
  - Selected subtrees
  - Topological relationships between them



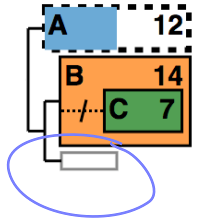
Visual design: focus + context

- Focus
  - Selected subtrees
  - Topological relationships between them



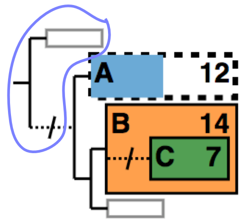
Visual design: focus + context

- Focus
  - Selected subtrees
  - Topological relationships between them
- Context
  - Neighboring subtrees



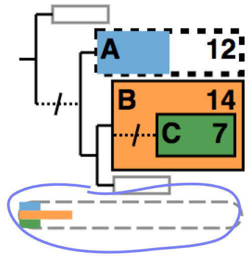
Visual design: focus + context

- Focus
  - Selected subtrees
  - Topological relationships between them
- Context
  - Neighboring subtrees
  - Upstream topology and root



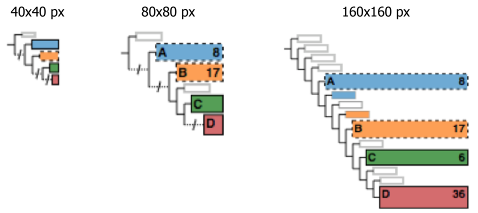
Visual design: focus + context

- Focus
  - Selected subtrees
  - Topological relationships between them
- Context
  - Neighboring subtrees
  - Upstream topology and root
  - Missing leaf nodes



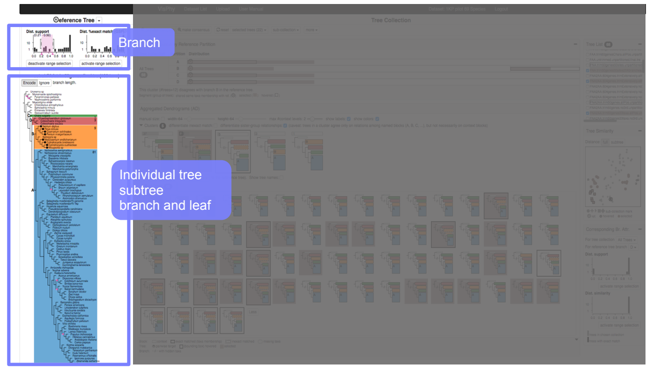
Visual design: algorithm adapts to space

- Show more info when space permitted
  - Labels
  - #leaf nodes
  - Neighboring blocks



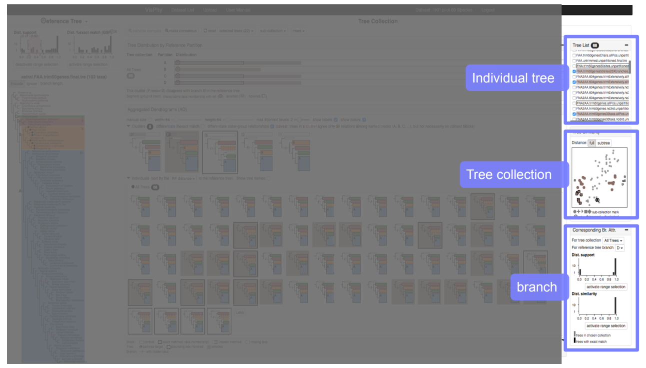
ADView Interface: Multi-level structure across views

Multi-level structure across views



Interface walkthrough: tree collection main views

Interface walkthrough: tree collection aux. views

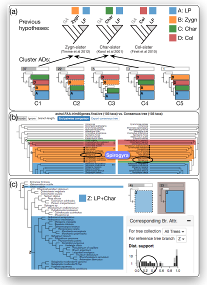


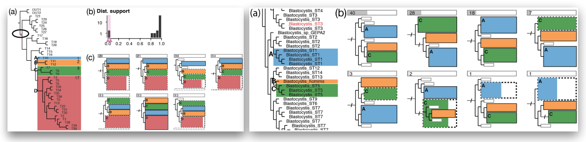
Validation with many biologists

- Work closely with a biology PhD student (second author)
- Demos, interviews and discussions
  - 10 biologists at different times throughout project

Validation with many biologists

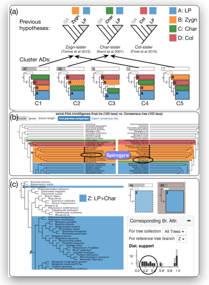
- Work closely with a biology PhD student (second author)
- Demos, interviews and discussions
  - 10 biologists at different times throughout project
- User study sessions
  - 5 biologists
  - Using their own datasets

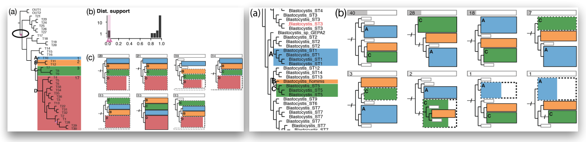




Validation with many biologists

- Work closely with a biology PhD student (second author)
- Demos, interviews and discussions
  - 10 biologists at different times throughout project
- User study sessions
  - 5 biologists
  - Using their own datasets
- Biologists confirmed
  - Validity of data and task abstractions
  - Utility of ADView





Q&A

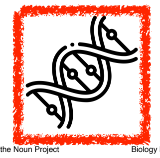
Break: 3:40-3:50

Four case studies of problem-driven work

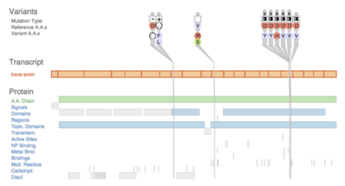
- in-car networks

- facilities management

- biology



Car by AIGA from the Noun Project Business by Colourcreatype from the Noun Project Biology by tezar tantular from the Noun Project



Variant View

Visualizing Sequence Variants in their Gene Context

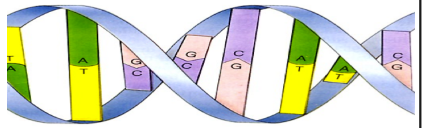
joint work with:
Joel Ferstay, Cydney Nielsen

http://www.cs.ubc.ca/labs/imager/tr/2012/VariantView/

Variant View: Visualizing Sequence Variants in their Gene Context.
Ferstay, Nielsen, Munzner. IEEE TVCG 19(12): 2546-2555, 2013 (Proc. InfoVis 2013).

Sequence Variant Definition

- Sequence variants
  - Difference between reference and given genome



Reference Genome DNA: ATA TGA TCA ACA CTT

Sample 1 Genome DNA: ATA TGG TCA ATA CTT Harmful?

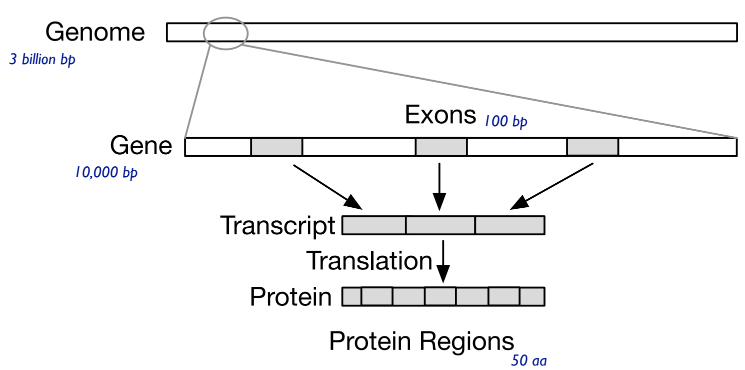
Sample 2 Genome DNA: ATA TGA TGA ACA CCT Harmless?

Cancer Research

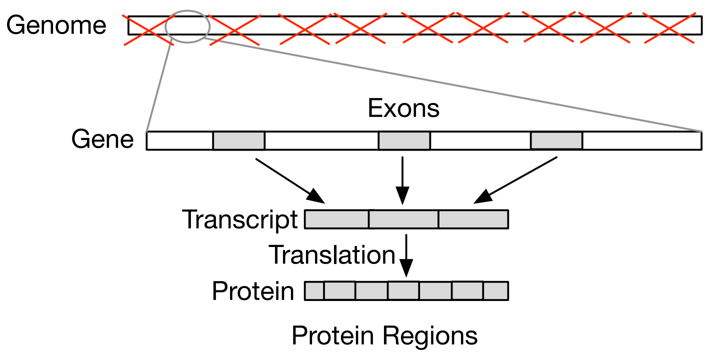
- collaboration with analysts at BC Genome Sciences Center
  - studying genetic basis of leukemia
- driving task
  - discover new candidate genes with harmful variants
- two big questions
  - what to show
    - data abstraction
    - challenge: enormous range of scales in the data
  - how to show it
    - visual encoding idiom

Abstractions

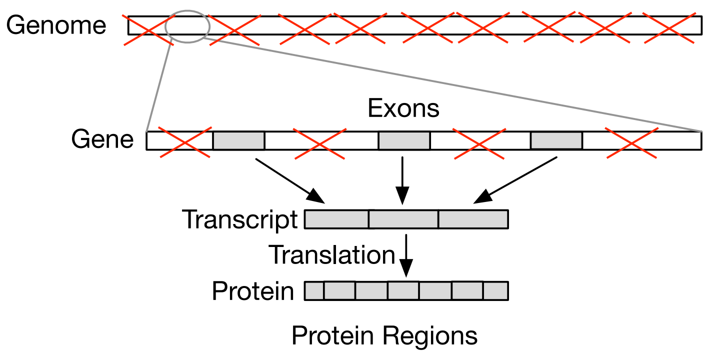
Data: Filtering to relevant biological levels and scales



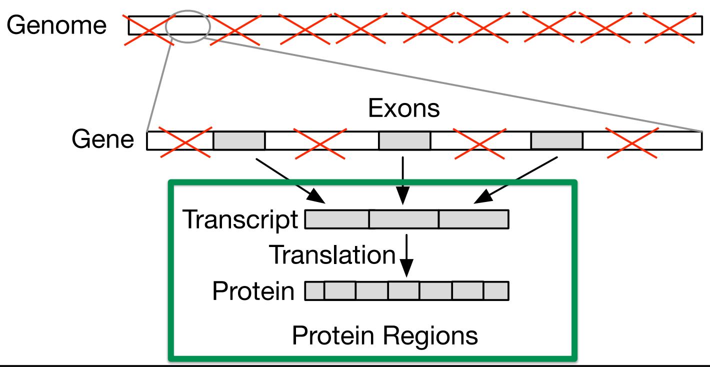
Filter out whole genome; keep genes



Filter out non-exon regions

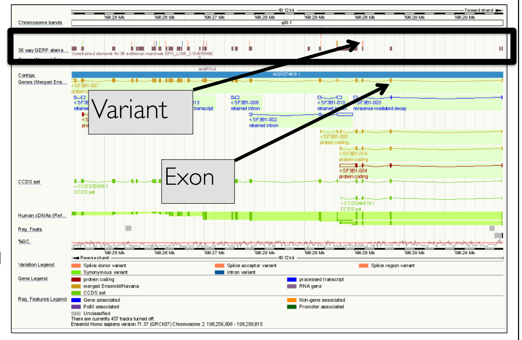


Data abstraction: highly filtered scope



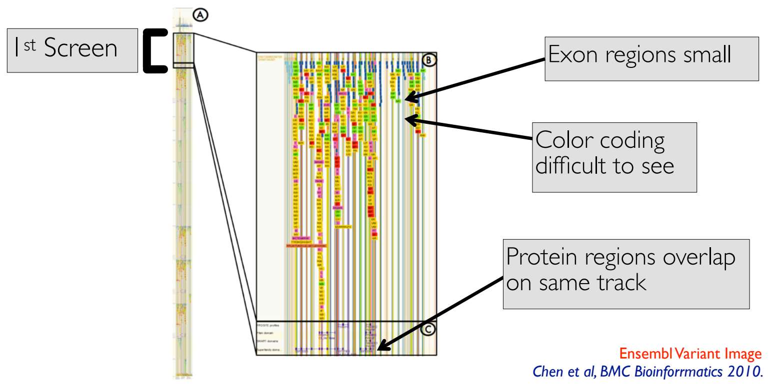
Dominant paradigm: genome browsers

- strengths: flexible and powerful
  - horizontal tracks: user data
  - shared coordinate system: genome coordinates (bp)
- problems
  - tiny features of interest spread out across large extent
    - must zoom far in to inspect known feature, then zoom out and pan to locate next
    - high cognitive load for interaction
    - must already know where to look



representative example: Ensembl
Chen et al, BMC Bioinformatics 2010.

Features of interest small even in variant-specific view

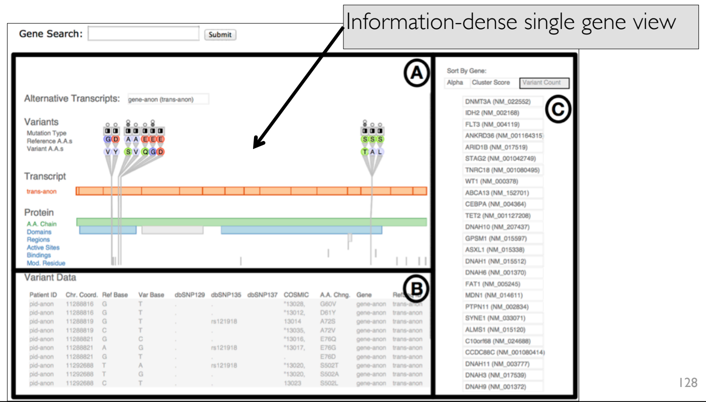


Ensembl Variant Image
Chen et al, BMC Bioinformatics 2010.

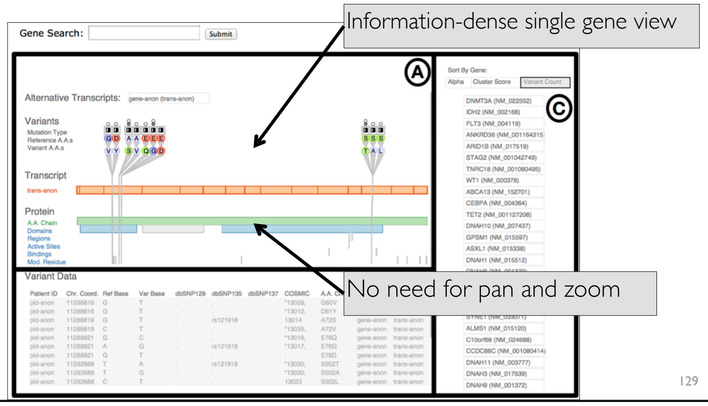
Idioms

Variant View

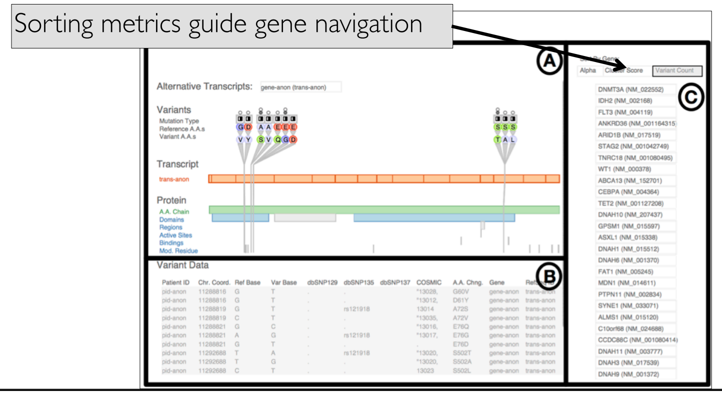
Variant View



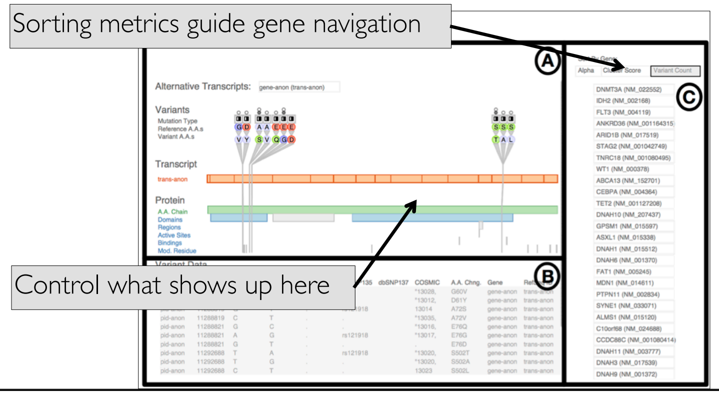
Variant View



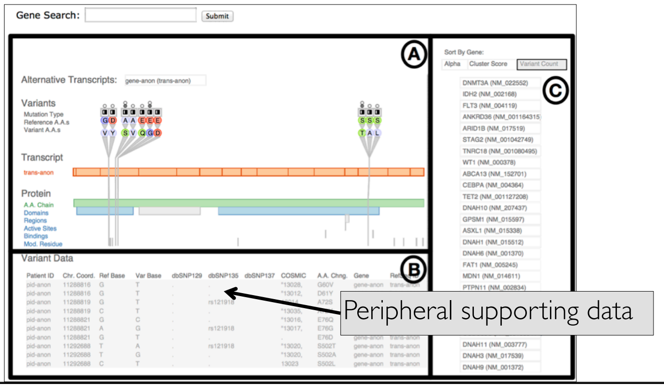
Variant View



Variant View

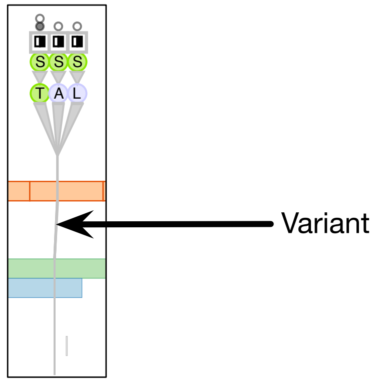


Variant View

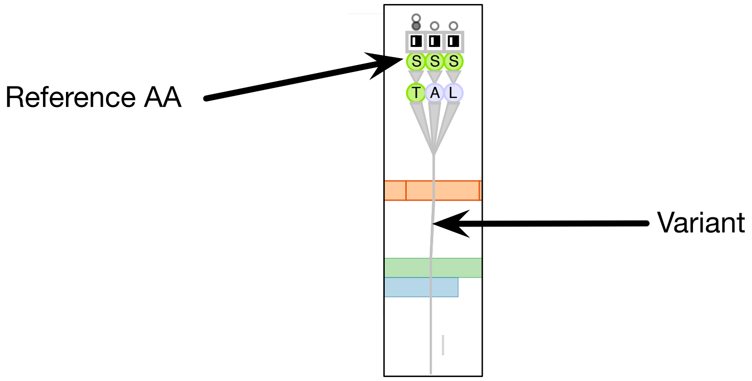


Design information-dense visual encoding

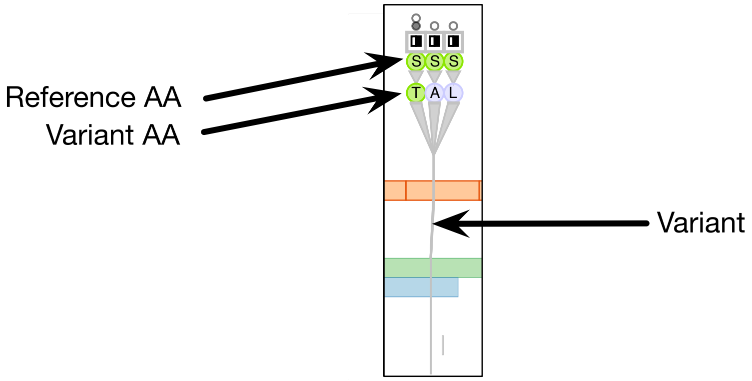
- show all attributes necessary for variant analysis
  - match salience with importance for analysis task
- variant not just a thin line!
- emphasize with high salience
  - collocated variants fan out at top
  - grey variant vertical stroke intersects horizontal colored protein regions



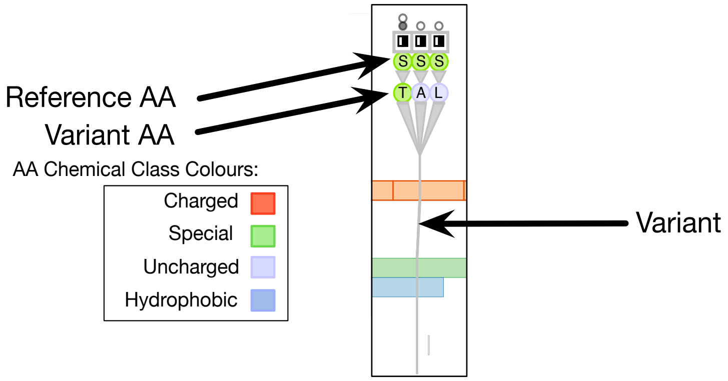
Design information-dense visual encoding



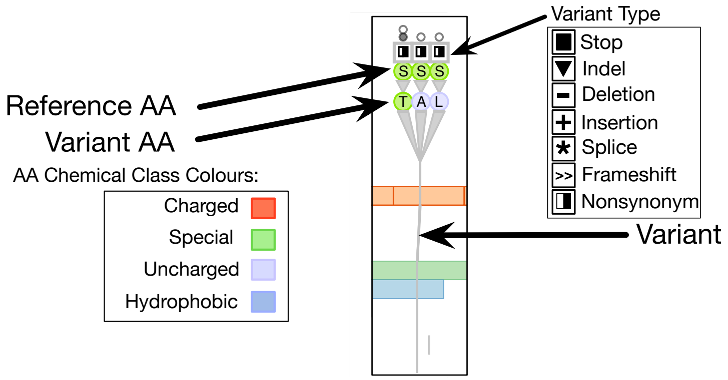
Design information-dense visual encoding



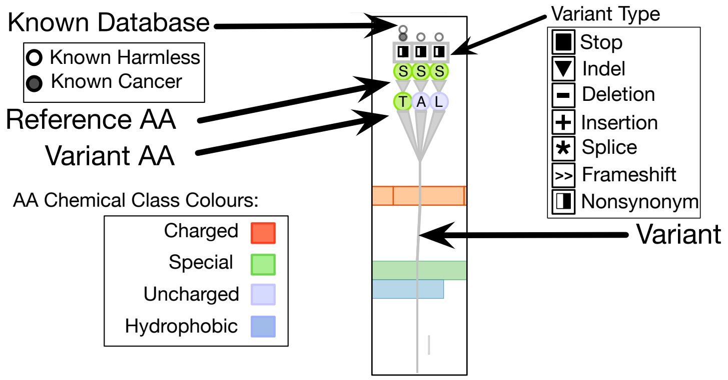
Design information-dense visual encoding



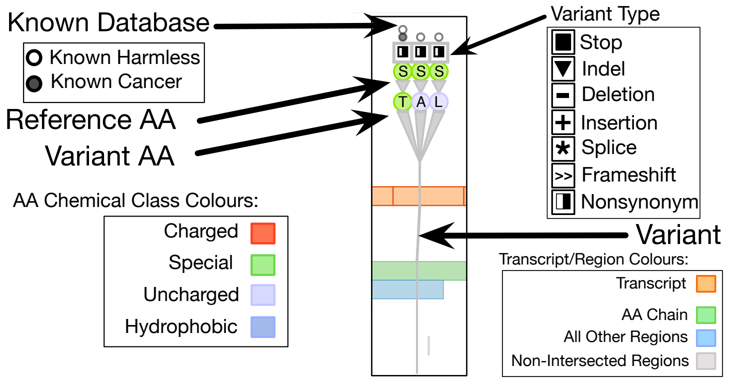
Design information-dense visual encoding



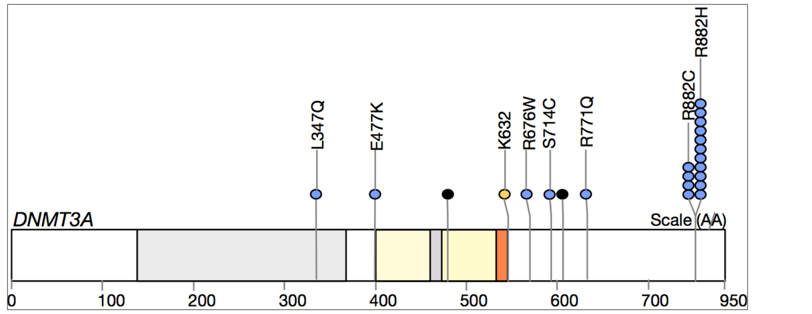
Design information-dense visual encoding



Design information-dense visual encoding

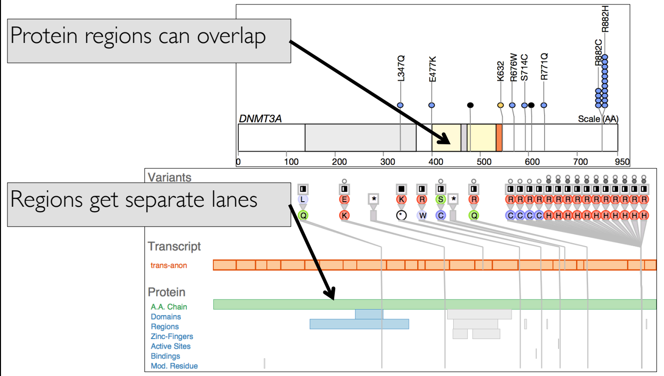


Previous work targeted at variant analysis: MuSiC

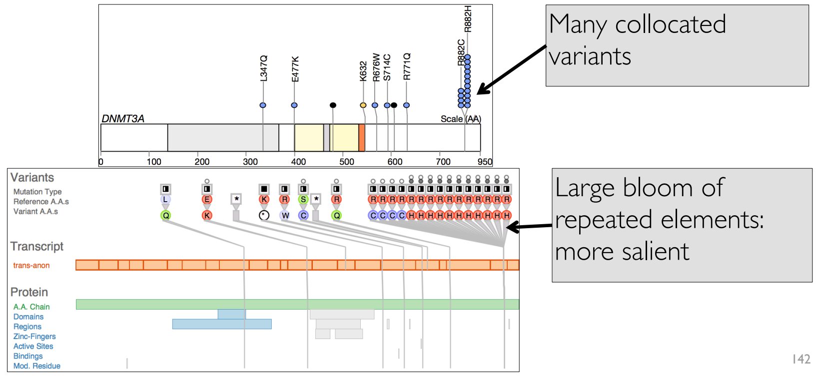


MuSiC
[Dees et al., Genome Research 2012]

Side-by-side comparison: MuSiC vs Variant View

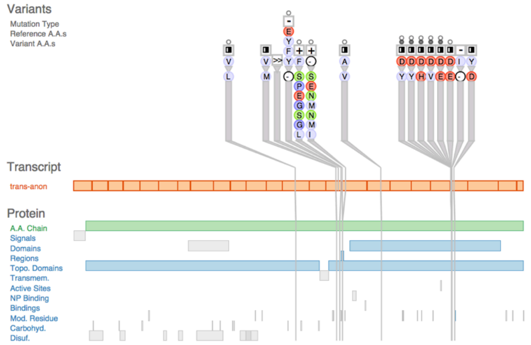


Side-by-side comparison: MuSiC vs Variant View

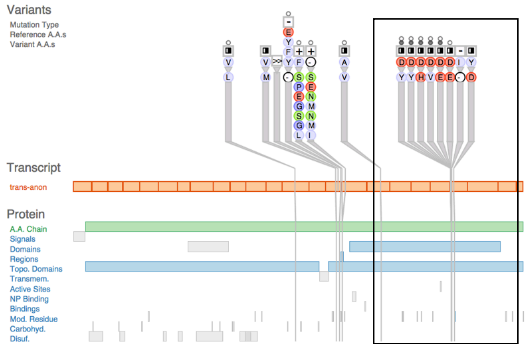


Results

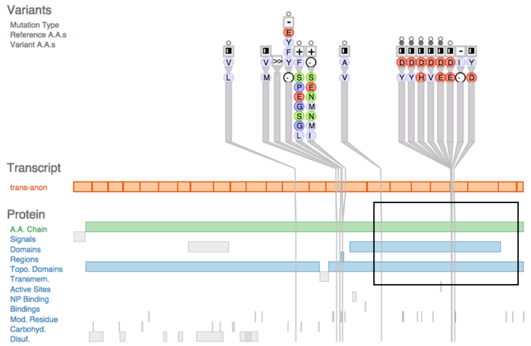
Verify known leukemia gene: Highly scored by sorting metric



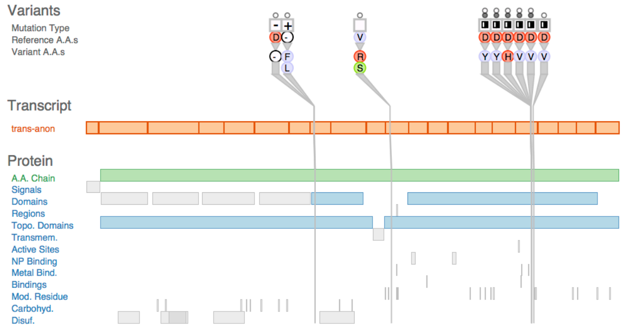
Visual inspection reveals collocation of variants



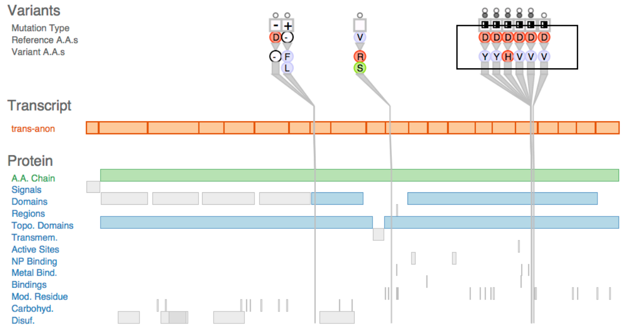
Several functional protein regions affected



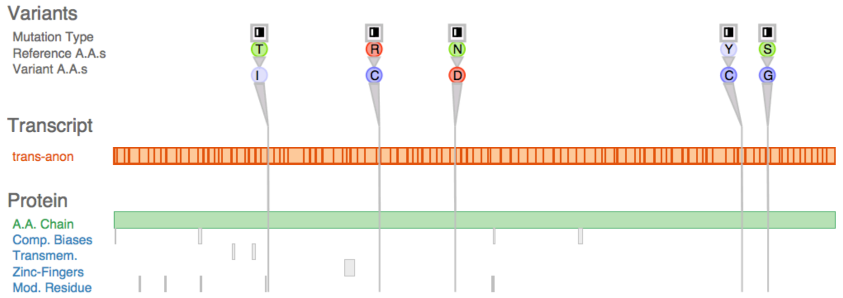
Highly scored by metric: not previously known, good candidate



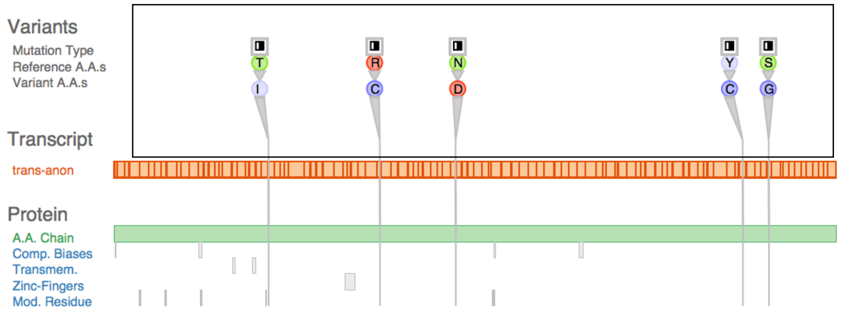
Protein chemical class change evident



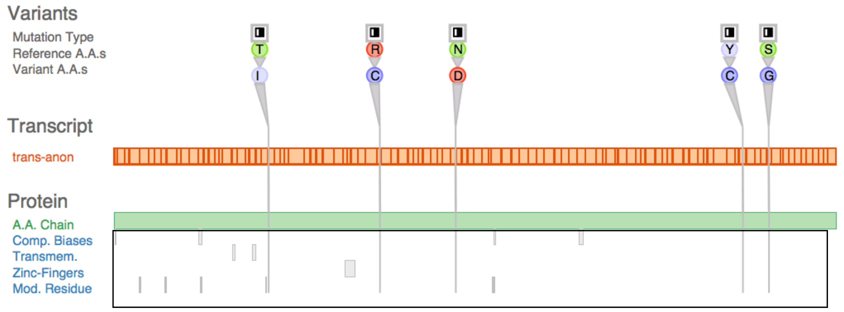
In contrast, low scoring gene



No collocation of variants



Mostly unaffected protein regions

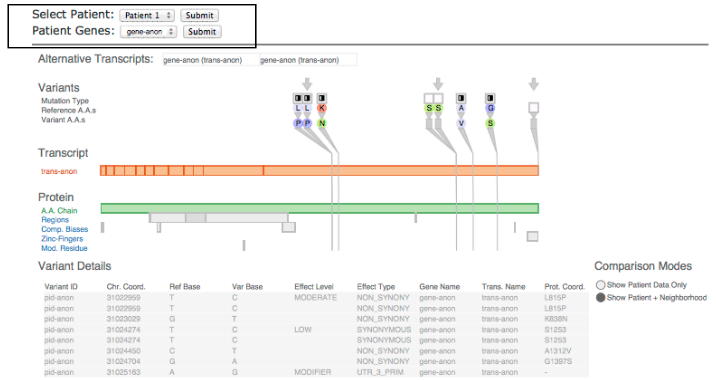


Additional tasks

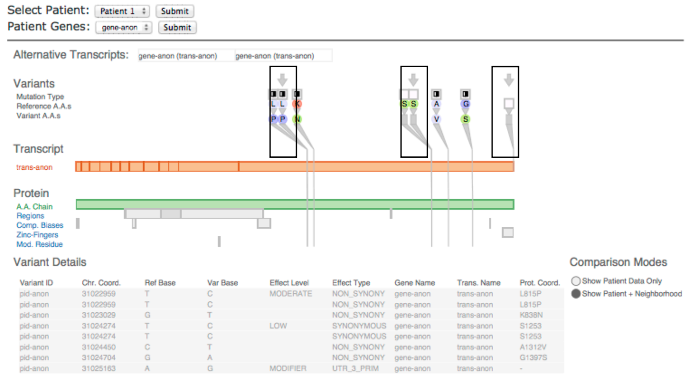
- task 2: compare patients
  - clinical setting application
  - compare patient data to known harmful variants
  - challenge
    - similarity is loosely understood rather than fully characterized
    - visual inspection for what constitutes a match

Adapted Variant View with minimal changes

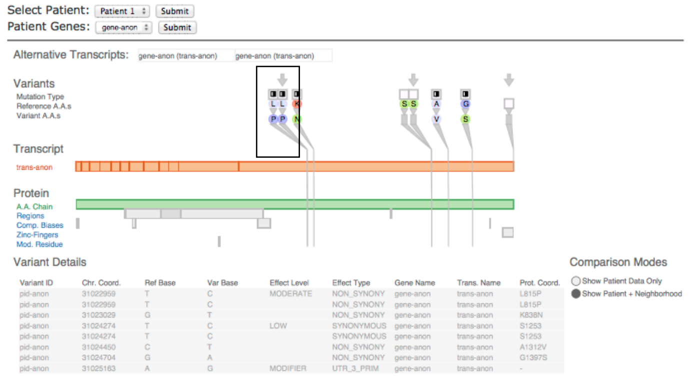
Navigate through patient data with list



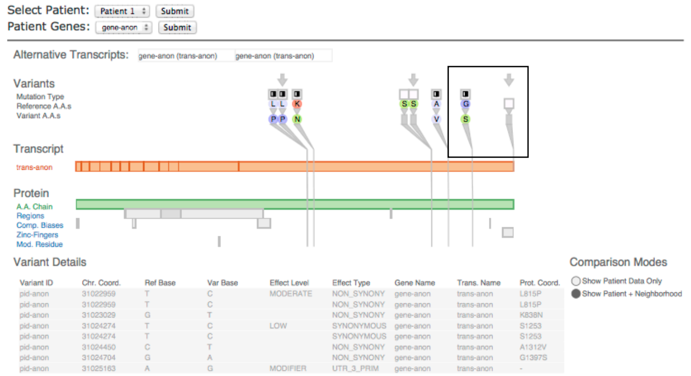
Patient data emphasized with arrows



Patient has same harmful L to P mutation



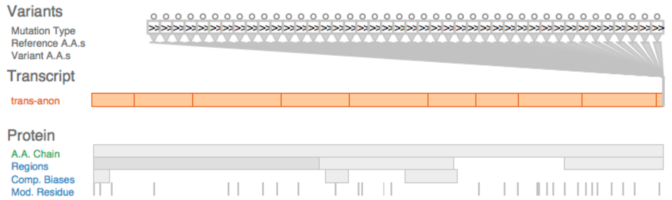
Nonmatching variants



Additional tasks

- task 3: debug pipeline
  - data cleansing before analysis
  - analysts originally thought pipeline fully debugged
    - no perceived need for vis support

Tool revealed errors in the data



The tool exposed artifacts in the data that slid past at least two rounds of quality metric filtering … this type of problem would not have been caught by our previous, automated methods.

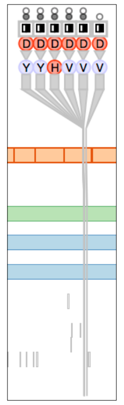
- Analyst 3

Reflections: vis design guidelines

- transferrable to other domains
  - specialize first, generalize later
    - good for domains where with complex, multi-scale data
    - difficult to judge a priori which design elements will generalize
  - high-level considerations
    - identifying scales of interest
    - what to visually encode directly vs what to support through interaction
    - when (and how) to eliminate navigation

Conclusions

- visual variant impact assessment
  - designed, implemented, and deployed tool for
- originally designed for Discover Genes task
  - adapted to two others with minimal changes
- features
  - navigation-free main overview at gene level
  - reveal genes of interest through sorting by new derived metrics
- major considerations
  - what to show
    - filtering data scope
  - how to show it
    - carefully selected visual encodings



Q&A

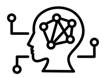
Mixed-initiative: Human in the loop with ML

- visualizing imperfect models

Two case studies of visualizing imperfect models

- NLP for temporal data

- ML for graph data

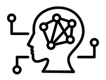


time by Wayne Middleton from the Noun Project machine learning by Eli Magaziner from the Noun Project

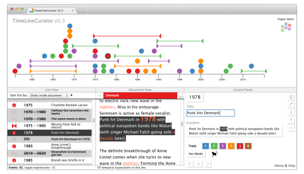
Two case studies of visualizing imperfect models

- NLP for temporal data

- ML for graph data



time by Wayne Middleton from the Noun Project machine learning by Eli Magaziner from the Noun Project



TimeLineCurator

Interactive Authoring of Visual Timelines from Unstructured Text

http://about.timelinecurator.org
http://timelinecurator.org

Johanna Fulda @jofu_

Matthew Brehmer @mattbrehmer

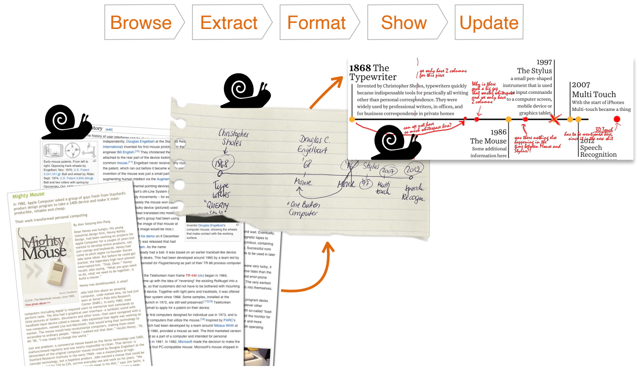
Tamara Munzner @tamaramunzner

TimeLineCurator: Interactive Authoring of Visual Timelines from Unstructured Text.
Fulda, Brehmer, Munzner. IEEE Trans. Visualization and Computer Graphics (Proc IEEE VAST 2015) 22(1):300-309, 2015.

TimeLineCurator
visual & browser-based

https://vimeo.com/jofu/tlc

Manual creation process



Structured creation process

TimelineJS
timeline.knightlab.com/

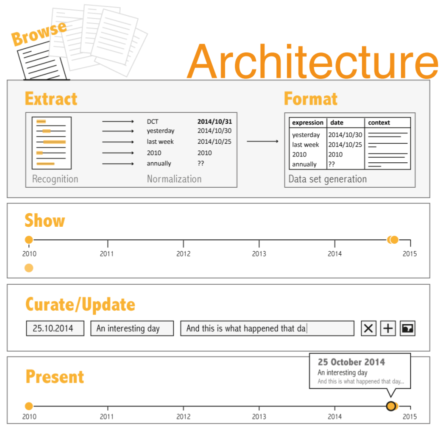
Timeline authoring model

- time required for each task

	Browse	Extract	Format	Show	Update
Manual Drawing	slow	slow		slow	slow
Structured Creation	slow	slow	slow	automated	fast
TimeLine Curator	fast	automated	automated	fast	fast

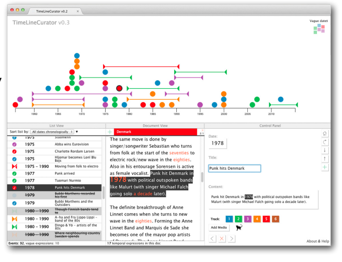
The general case for curation

- build for human in the loop as continuing need
  - automatic processing to accelerate not replace
  - **assume automatic computational results good but not perfect**
    - for the indefinite future!
  - visual feedback to accelerate



The importance of being brisk

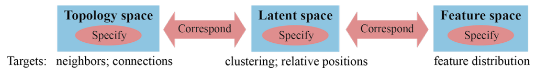
- cool use case: eureka moment
  - success: enable what was impossible before
  - vis tools for new insights & discoveries
- workhorse use case: workflow speedup
  - success: vis tools accelerate your prior workflow
    - sometimes enables the previously infeasible
- TLC use cases
  - started with speedup use case, for presentation
    - make this doc into a timeline now!
  - two other use cases nudge towards exploration
    - comparison between multiple timelines
    - speculative browsing



TimeLineCurator: Speculative Browsing

speculative browsing

https://vimeo.com/jofu/tlc



Visualizing Graph Neural Networks with CorGIE:

Corresponding a Graph to Its Embedding

http://www.cs.ubc.ca/group/infovis/pubs/2021/corgie/

Zipeng Liu
UBC/Beihang

Yang Wang
Uber/Facebook

Jürgen Bernard
UBC/Zurich

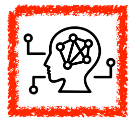
Tamara Munzner
UBC

Visualizing Graph Neural Networks with CorGIE: Corresponding a Graph to Its Embedding.
Liu, Wang, Bernard, Munzner. IEEE TVCG 28(6):2500-2516 2002.

Two case studies of visualizing imperfect models

- NLP for temporal data

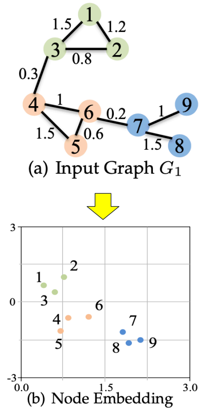
- ML for graph data



time by Wayne Middleton from the Noun Project machine learning by Eli Magaziner from the Noun Project

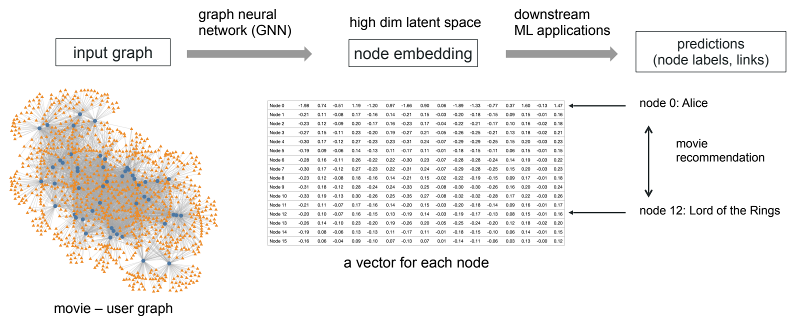
Graph neural network (GNN)

- machine learning (ML) models for graphs
  - like CNN for images
  - like Transformer for text
- many real-world graph-related applications
  - node classification
    - examples: fraud detection, disease classification
  - link prediction
    - examples: product recommendation, protein interactions

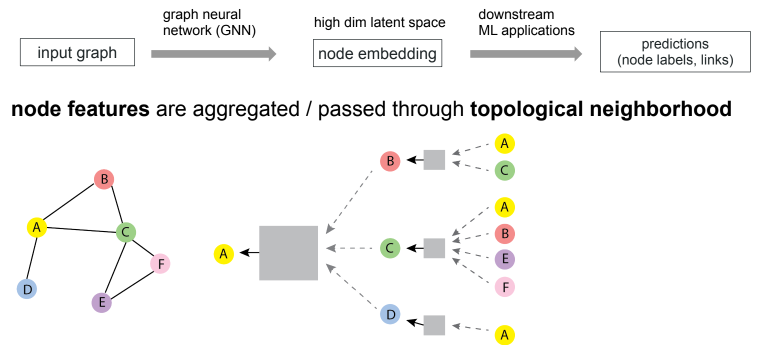


[Cai et al. TKDE'18]

Graph neural network (GNN)



Graph neural network (GNN)



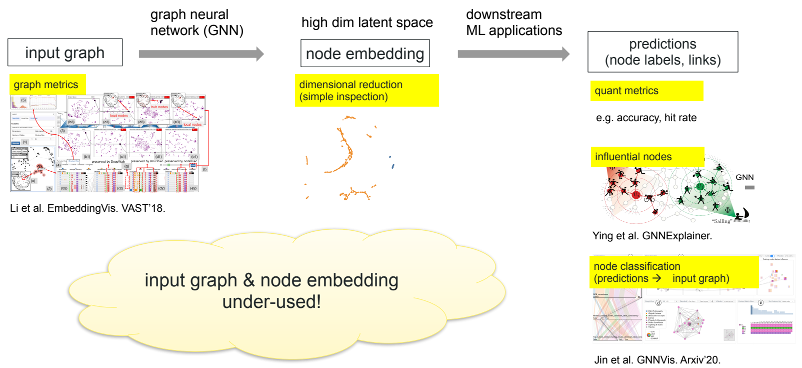
Remake from https://snap-stanford.github.io/cs224w-notes/machine-learning-with-networks/graph-neural-networks

Evaluate GNN

Two big-picture questions

- "Are we there yet?": should we train / tune more?
- "Are we lost?": does it behave as we expect?

Evaluate GNN: Previous approaches

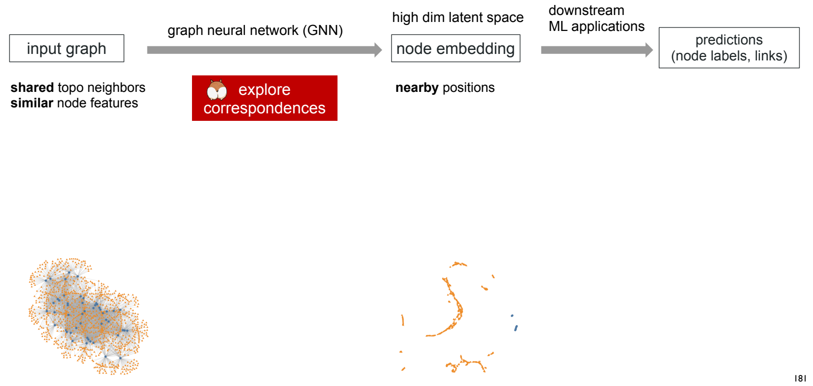


Li et al. EmbeddingVis. VAST'18.

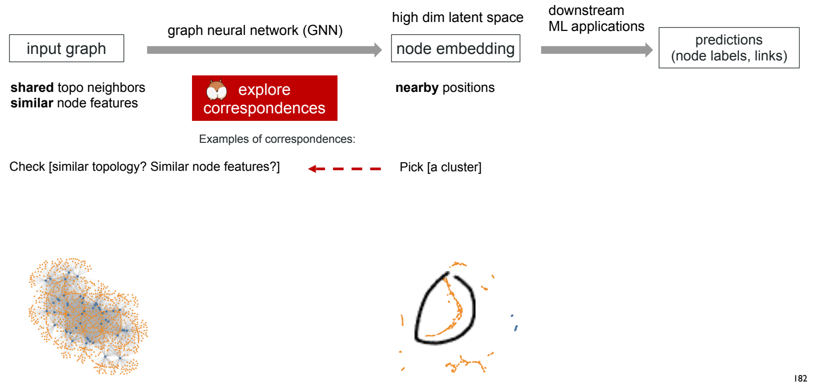
Ying et al. GNNExplainer.

Jin et al. GNNVis. Arxiv'20.

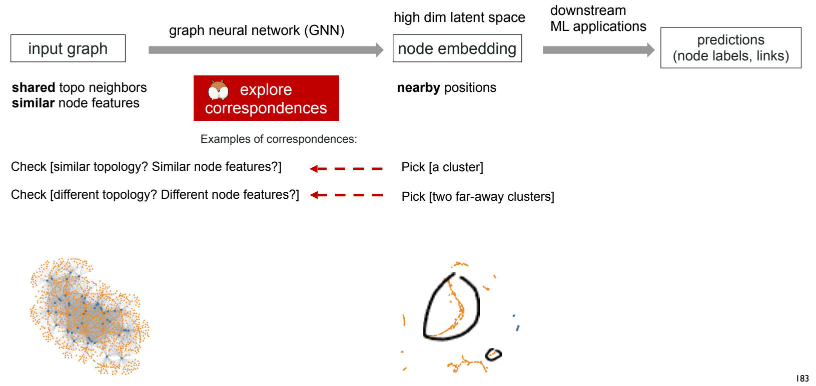
Evaluate GNN: CorGIE idea



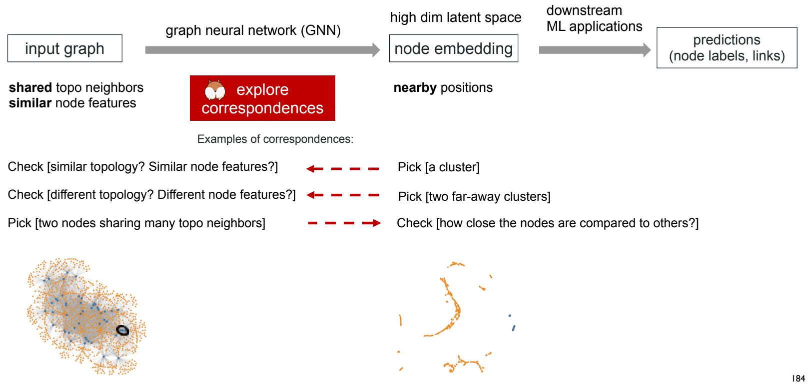
Evaluate GNN: CorGIE idea



Evaluate GNN: CorGIE idea



Evaluate GNN: CorGIE idea



Data and task abstraction

Topology space | **Latent space** | **Feature space**

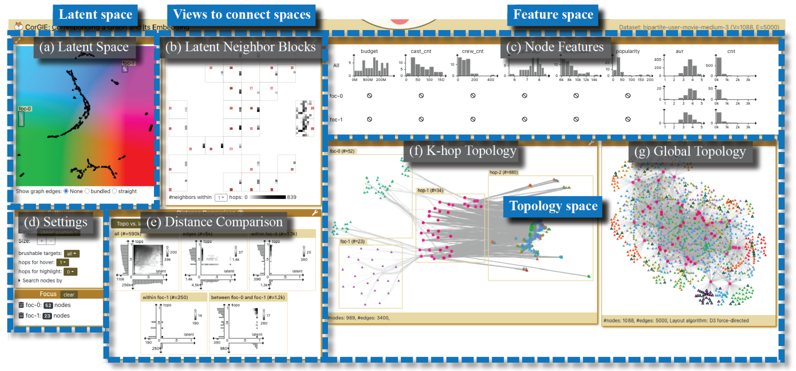
- data spaces

Data and task abstraction

Topology space (Specify) ⟷ Correspond ⟷ **Latent space** (Specify) ⟷ Correspond ⟷ **Feature space** (Specify)

- data spaces
- tasks
  - specify
  - correspond
- task iteration
  - levels in grouping structure of nodes

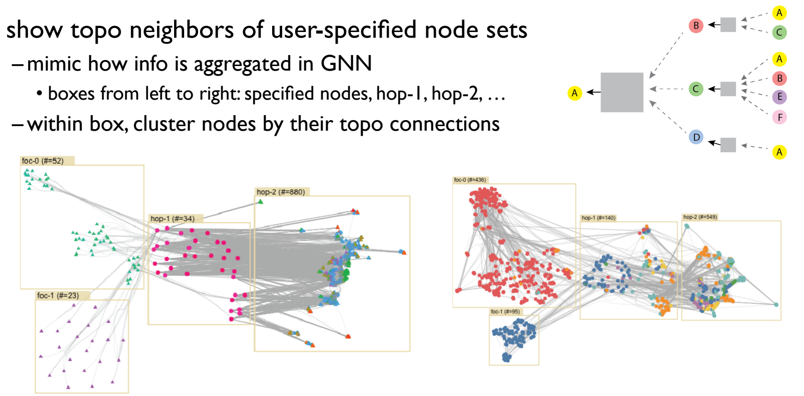
Contribution: Multi-view interactive interface



Video: https://osf.io/j56hu/

Contribution: K-hop layout

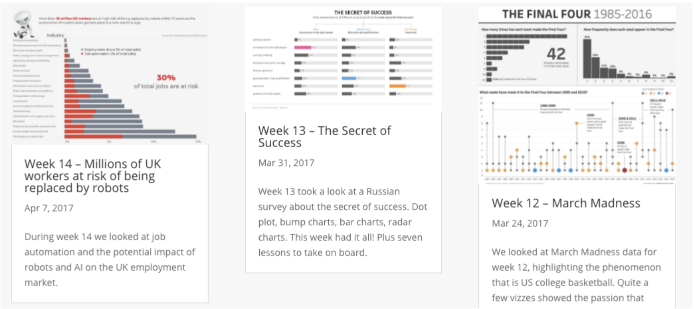
- show topo neighbors of user-specified node sets
  - mimic how info is aggregated in GNN
    - boxes from left to right: specified nodes, hop-1, hop-2, …
  - within box, cluster nodes by their topo connections



Next Steps

Learning through Redesign En Masse: **Makeover Mondays**

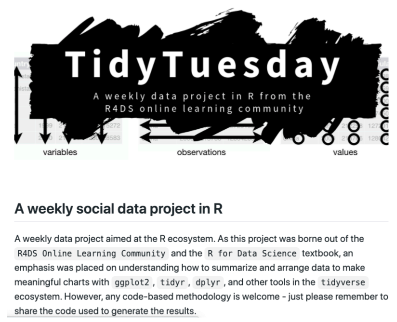
- easy entry point (Tableau focus)



http://www.makeovermonday.co.uk

Learning through Redesign En Masse: **Tidy Tuesdays**

- easy entry point (R focus)



https://github.com/rfordatascience/tidytuesday

Visual Design Process In Depth: **Dear Data**

- inspiring celebration of data humanism

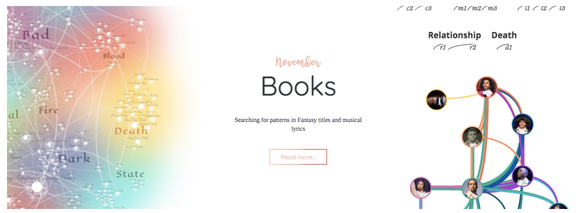
Giorgia Lupi and Stefanie Posavec

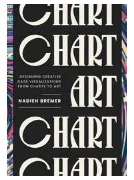
http://www.dear-data.com/by-week/

https://giorgialupi.com/dear-data https://www.stefanieposavec.com/dear-data

Visual Design Process In Depth: **Data Sketches**

- detailed process notes, from sketching through coding Shirley Wu & Nadieh Bremer





new book out from Bremer: CHART
Designing Creative Data Visualizations from Charts to Art

http://www.datasketch.es/

Pathways for more participation: organizations

- join Vancouver Visualization meetup (4K+ members)
  - https://www.meetup.com/Vancouver-Data-Visualization/
- join Data Visualization Society
  - https://www.datavisualizationsociety.org
  - four years old, 23K+ members around the world
  - **jobs board:** full-time, part-time and contract positions worldwide https://jobs.datavisualizationsociety.org/
  - many other resources, super-active Slack incl local groups, challenges, ...
  - articles: Nightingale
  - conferences: Outlier
  - awards: Information Is Beautiful

Visualization jobs

- spectrum
  - visualization as main/core focus
  - visualization as occasional task
  - visualization skills add strength to your portfolio even if no immediate duties
- local companies
  - Tableau Vancouver is largest company focused on visualization
  - many smaller ones have visualization / data science needs

Upcoming

- see you next week for final presentations
- enjoy visualization, for those who keep going down this path!
  - now or later...