Two interrelated processes uncover data landscapes:

Data Reconnaissance
- Acquire additional data sources
- Analysis & visualization of available data sources supports acquisition of new data
- Acquire new dataset
- Acquire available, but previously restricted, dataset

Task Wrangling
- New idea: Operational definitions for data reconnaissance and task wrangling

Questions in road trips
- Where are we?
- What’s here?
- Are we there yet? Are we lost?

Domain experts need help uncovering and reasoning about heterogeneous data landscapes

Operational definitions for data reconnaissance and task wrangling

Extended analogy
- Wayfinding through the world with road trips
- Wayfinding through data with visualization

Data Reconnaissance & Task Wrangling

Where are we?
Domain experts need help uncovering and reasoning about heterogeneous data landscapes

What’s here?

Operational definitions for data reconnaissance and task wrangling

New idea:

Data landscape: the very large space of existing heterogeneous and multidimensional datasets that are not yet understood by a specific person

Data Reconnaissance & Task Wrangling

Questions in road trips
- and visualization in data science!
Two interrelated processes uncover data landscapes:

1. Task Wrangling
2. Data Reconnaissance

**Low Task Clarity**
What is this data?

**Evolving Task Clarity**
I think I might want to see geographic patterns.

**Refined Task Clarity**
I want to see the geographic relatedness of connected genomic clusters over time.

Processes influence each other over time. Refined tasks guide the pursuit of data.

A conceptual framework for data reconnaissance and task wrangling:

**New idea:**
A conceptual framework for data reconnaissance and task wrangling.

Existing methods can be slow. Design Study Methodology: Human Centered Design.


Steps in our conceptual framework:

- Acquire
- View
- Assess
- Pursue

From unknown landscape to the final dataset:

Where do we go from here? Building systems suitable for data reconnaissance and task wrangling.

Questions in road trips and visualization in data science:

- where are we?
  -- Uncovering Data Landscapes through Data Reconnaissance & Task Wrangling

- what’s here?
  -- Automatic Encodings through Recommendation

GEViTRec: Data Reconnaissance Through Recommendation Using a Domain-Specific Visualization Prevalence Design Space.

Data Reconnaissance
Data access inspires & refines tasks.
How to connect datasets? Identify shared attributes!

Domain Popularity

Design Space:
Captures full scope of visual encodings used by
defineable set of experts, includes quantitative
estimate for prevalence of each strategy within that
domain
Domain-level answer to question of what’s here?

New Idea: Visually Coherent Chart Combinations Through Gradual Binding

• Automatically coordinating static charts is not trivial
• Cannot change encoding after chart rendered into box of pixels
• Declarative approach of gradual binding
• Initially generate partial specification using template
• Modify specification in discrete stages, to enforce consistency of
channels (color, position) according to desired combination
• Pass final specification to rendering library
• Simply concatenate resulting boxes of pixels to display

Constructing Visually Coherent Chart Combinations Through Gradual Binding

• Automatically coordinating static charts is not trivial
• Cannot change encoding after chart rendered into box of pixels
• Declarative approach of gradual binding
• Initially generate partial specification using template
• Modify specification in discrete stages, to enforce consistency of
channels (color, position) according to desired combination
• Pass final specification to rendering library
• Simply concatenate resulting boxes of pixels to display

GeViTRec algorithm: Overview

Data Type

Data Source

Exploded Fields

Fields

Jaccard Index

Linkage Type

Data Source Graph

Data Ranking

• GEViTRec runs in R Markdown notebooks
• Example: 2013-2016 Ebola outbreak data

Automatically Constructing Visually Coherent Chart Combinations

• GEViTRec runs in R Markdown notebooks
• Example: 2013-2016 Ebola outbreak data

Domain Context: Genomic Epidemiology

A Cross, JI, Giddy, T Maciula:
A systematic method for surveying data visualizations and a
resulting genome-wide data visualization typology (2019)
https://doi.org/10.1093/bioinformatics/bty832

How to show connections for data recon?

Visually Coherent Chart Combinations

that prioritize visual coordination of
shared information between
charts with respect to layout and
consistency among visual
channels (position, color)

How to show connections for data recon?

Visually Coherent Chart Combinations

that prioritize visual coordination of
shared information between
charts with respect to layout and
consistency among visual
channels (position, color)

How to show connections for data recon?

Visually Coherent Chart Combinations

that prioritize visual coordination of
shared information between
charts with respect to layout and
consistency among visual
channels (position, color)
Evaluate GNN: CorGIE idea

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
  - edge prediction
  - link prediction
  - example: product recommendation, protein interactions

Graph neural network (GNN)

- Evaluating GNN quality
  - Two big-picture questions
    - Are we there yet? Should we train / tune more?
    - Are we lost? Does it behave as we expect?

GEViTRec lowers burden to quickly visualize data

- Speeds up the process of data reconnaissance - where are we?
- Automatically shows us what’s here?
  - Identifies connections among datasets
  - Exploits domain prevalence design space
  - Constructs visually coherent chart combinations through gradual binding

VizCommender: Computing Text-Based Similarity in Visualization Repositories for Content-Based Recommendations

Examples of correspondences:

- where are we?
  - Data Reconnaissance & Task Wrangling
  - what’s here?
  - Automatic Encodings through Recommendation
  - To shed light on data landscapes

Questions in road trips - and visualization in data science!

- where are we?
  - Chain of thought: to move forward
  - what’s here?
  - Autoencoding through Recommendation to shed light on data landscapes

• Comparing to existing bespoke tools:
  - Fast: easy to use
  - Unaligned: Have better alignment between chart types
  - Heterogeneity support: handle multiple types of data

Evaluate GNN: CorGIE idea

Graph neural network (GNN)

node embedding

high dim latent space

Examples of correspondences:

- Pick (a cluster)
- Pick (two far-away clusters)
- Pick (a cluster)
- Pick (two nodes sharing many topo neighbors)

Graph neural network (GNN)

node features are aggregated / passed through topological neighborhood

Graph neural network (GNN)

Examples of correspondences:

- Check [similar topology? Similar node features?]
- Check [different topology? Different node features?]
- Pick (a cluster)
- Pick [two nodes sharing many topo neighbors]
Data and tasks
• three data spaces
• tasks
  – specify
  – correspond

CorGIE multi-view interactive interface

CorGIE multi-view interactive interface

CorGIE: Visual Assessment of ML Training Completion & Quality
• Addresses where are we?
  • Visually explore correspondences between input graph and node embedding to show what’s here?
  • Has the GNN training process captured all expected data about k-hop neighborhoods in the input graph, or should we keep going with train/tune?
• Addresses are we there yet?
  • Are the GNN predictions high quality or low quality?
• Addresses are we lost?

Questions in road trips - and visualization in data science!
• one VDS project for each question
  • where are we?
    – Data Reconnaissance & Task Wrangling
  • what’s here?
    – Automatic Encodings through Recommendation
  • are we there yet? are we lost?
    – Visual Assessment of ML Training Completion

More information
• this talk 
  http://www.cs.ubc.ca/~tmm/talks.html#vds23 
• full courses, papers, videos, software, talks 
  http://www.cs.ubc.ca/~tmm/teaching 
  http://www.cs.ubc.ca/~tmm/talks.html 
• book 
  http://www.cs.ubc.ca/~tmm/vadbook 
• VIS23 book table from CRC/Routledge 
  – physical table 

@tamaramunzner
@tamara@vis.social