Empirical Evaluation of Complex Interactive Visualization Techniques

Alexander Lex
@alexander_lex
http://alexander-lex.net
On sabbatical at datavisyn, in Linz
Data visualization solutions for pharmaceutical industry
20 people and growing!

Marc Streit, CEO
Dominic Giradi, CPO
Alexander Lex
Nils Gehlenborg
Samuel Gratzl
The purpose of computing is insight, not numbers.

visualization

pictures

[Card, Mackinlay, Shneiderman]

[Richard Wesley Hamming]
Banana  M. acuminata
Date  P. dactylifera
Cress  Arabidopsis thaliana
Rice  Oryza sativa
Sorghum  Sorghum bicolor
Brome  Brachypodium distachyon
SO CAN WE DO BETTER?
Universal Set
<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>Cardinality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
</tbody>
</table>
THE BANANA CHART REDESIGNED: UPSET

Largest Intersection Includes All Sets
THE BANANA CHART REDESIGNED: UPSET

Three Leftmost Species Are Most Similar
THE BANANA CHART REDESIGNED: UPSET

Rightmost species is most different
UpSet - Visualizing Intersecting Sets

Choose Dataset: Movies Genres (17 sets, 4 attributes)

Load Data | About UpSet | UpSet for R

http://vcg.github.io/upset/
The canonical way to show set data with > 3 sets
Second-most cited VIS paper of the last decade
> 11 implementations in various languages
See https://upset.app
RESEARCH AREAS
<table>
<thead>
<tr>
<th>TECHNICAL CONTRIBUTIONS</th>
<th>DOMAIN DRIVEN TECHNIQUES</th>
<th>EMPIRICAL &amp; THEORETICAL WORK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Novel Visualization Techniques</td>
<td>Tailored Methods and Systems for High Impact Science Problems</td>
<td>Evaluation Methodology</td>
</tr>
<tr>
<td>Visualization Process Innovations</td>
<td></td>
<td>Design Spaces / Taxonomies</td>
</tr>
<tr>
<td>Data Wrangling Methods</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data Wrangling Methods

Reshaping Networks

TECHNICAL CONTRIBUTIONS

Forest Gump
genre: Drama

Pretty Woman
genre: Comedy

Notting Hill
genre: Comedy

Promote (genre)

Forest Gump
genre: Drama

Pretty Woman
genre: Comedy

Notting Hill
genre: Comedy

Drama

Comedy
**TECHNICAL CONTRIBUTIONS**

Data Wrangling Methods

**DOMAIN DRIVEN TECHNIQUES**

Tailored Methods and Systems for High Impact Science Problems
Tailored Methods and Systems for High Impact Science Problems

Genealogies for Clinical Data Analysis
<table>
<thead>
<tr>
<th>Domain Driven Techniques</th>
<th>Empirical &amp; Theoretical Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tailored Methods and Systems for High Impact Science Problems</td>
<td>Evaluation Methodology</td>
</tr>
<tr>
<td></td>
<td>Design Spaces / Taxonomies</td>
</tr>
</tbody>
</table>
Evaluation
Methodology

Design Spaces / Taxonomies
Empirically Evaluating Complex Interactive Visualization Techniques

Carolina Nobre, Dylan Wootton, Lane Harrison
Perceptual Studies

Static Visualizations

Controlled Interactions

Complex Visualization Techniques

Commonly Evaluated Using Crowdsourcing

Considered not Amenable to Crowdsourced Evaluation
Can we use empirical studies to evaluate complex interactive visualizations?
The State of the Art in Visualizing Multivariate Networks

C. Nobre1, M. Meyer1, M. Streit2, and A. Lex1

1 University of Utah, Salt Lake City, USA
2 Johannes Kepler University Linz, Austria

Abstract
Multivariate networks are made up of nodes and their relationships (links), but also data about those nodes and links as attributes, or several attributes, and may convey data about several attributes of the network. Motivated by this diversity, we introduce a taxonomy of operations and classes and techniques. This taxonomy is a means to improve operations and classes of visualization for multivariate networks. This taxonomy includes an analysis of tasks specific to multivariate networks and give recommendations for which technique to use in which scenario. Finally, we survey application areas and evaluation methodologies.
CONFOUNDERS

HOW CAN WE MAKE SURE THAT WHAT WE TEST IS WHAT WE CARE ABOUT?
Please rate relative to conceivable design alternatives but assuming a node-link diagram as given.

26. Embedded bar charts are well suited to encode multiple numerical attributes.
   Mark only one oval.
   
   |   |   |   |   |   |   |   |
   | Strongly Disagree |   |   |   |   |   | Strongly Agree |

27. Embedded colored glyphs are well suited to encode multiple categorical attributes.
   Mark only one oval.
   
   |   |   |   |   |   |   |   |
   | Strongly Disagree |   |   |   |   |   | Strongly Agree |
Quickly find research participants you can trust.

Launch your study to tens of thousands of trusted participants in minutes. Recruit niche or representative samples on-demand. Prolific builds the most powerful and flexible tools for online research. Sign up for free.

Research
Collect high quality responses from people around the world within minutes. Learn more

Participate
Take part in engaging research, earn cash, and help improve human knowledge. Learn more

SIGN UP TO RESEARCH
SIGN UP TO PARTICIPATE

Find any research participant, anywhere in the world

Our participant pool is profiled, high quality and fast. The average study is completed in under 2 hours. Filter participants by demographics, interests, or role.
NOVICE USERS

NOVICE USERS DON’T KNOW ABOUT ADVANCED VISUALIZATIONS
NOVICE USERS
TRAINING CAN GIVE USERS THE EXPERTISE NECESSARY TO COMPLETE THE TASKS.
NOVICE USERS TRAINING CAN GIVE USERS THE EXPERTISE NECESSARY TO COMPLETE THE TASKS.
INCENTIVES
HOW CAN WE GET USERS TO TRY HARD AND TO PARTICIPATE IN AN EXPERIMENT THAT TAKES ~1H
AN INTERESTING PROBLEM

MONEY

~ $4,500 in 2h
VALIDATION

HOW CAN WE MAKE SURE THIS ALL WORKS?

DETAILED PROVENANCE TRACKING

MULTIPLE PILOTS
A library for reproducible tracking

Tracks “differential state”

Tracks 2 levels:
  Application State
  Study Metadata (responses, etc.)

https://vd़l.sci.utah.edu/trrack/
Allows a full re-hydrate of every analysis session

Great for debugging pilots - what went wrong for this person?

Great for detailed analysis

Provenance data vis to spot problems e.g. with tasks
PATHS

Is NL or AM better for Path Tasks?
NL outperforms AM for path based tasks
What types of insight do NL and AM representations support?
“INSTITUTIONS HAVE MUCH FEWER TWEETS IN GENERAL THAN PERSON ACCOUNTS”
"IT DOES SEEM A BIT ODD THAT JEFFREY ALEX AND ROB HAVE SUCH LARGE NETWORKS WITH THEIR LOWER THEN AVERAGE TWEETING."
Yes We Can!

- Picking the right techniques
- Evidence-based design
- Design validation
- Careful training
- Good compensation
- Interesting Tasks

**Pushing the boundary of what can be evaluated using crowdsourcing**
Should this be the new gold standard to evaluate systems?  
**NO!**

Needs established techniques  
Needs specifically designed and instrumented systems  

Our instrumentation can be used broadly
ADVANCED ANALYSIS OF STUDY DATA
What else can we do with this rich data?

Do different analysis strategies result in different results?
How do analysis strategies impact user performance when using interactive visualizations?
Pilot Iterations

- Study Design
- Instrumentation
- User Study
- Analysis Outcomes

Data Analysis

- Quantitative Analysis
- Qualitative Analysis
- Event Sequence Mining
- Event Playback in Study

reVISIt

Sequence Aggregation
Data Analysis

Quantitative Analysis

Robert together being connected to with Jon Jeffrey Alex and Rob have such large
to most important nodes in the graph. It is not that surprising. -- There
average than the people -- nothing I ca

Qualitative Analysis

Event Sequence Mining

Sequence Aggregation

Event Playback in Study
**FUTURE WORK: BETTER SCAFFOLDING FOR STUDIES**

**reVISit: Scalable Empirical Evaluation of Interactive Visualizations**

<table>
<thead>
<tr>
<th>Community Input</th>
<th>User Services and Resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaborators</td>
<td>Documentation / Examples</td>
</tr>
<tr>
<td>Core Community</td>
<td>Replications</td>
</tr>
<tr>
<td>Community Workshops</td>
<td>Community Engagement</td>
</tr>
<tr>
<td>Broader Community</td>
<td></td>
</tr>
</tbody>
</table>

---

**Core Infrastructure: Software Components**

**Inform Software Development**

**Synergistic Activities**

**Aim 1: Study Infrastructure and Multilevel Instrumentation**

- Study Scaffolding
- Integration with Crowdsourcing Platforms
- Component Registry
- Capturing Insights / Rationale
- Provenance Tracking
- Designing Trainings

**Aim 2: Data Transformation and Visualization Methods**

- Data Transformation
- Event Sequence Visualization
- Trials Visualization
- Study Results Visualization
- Study Replay Interface
- Qualitative Coding Tools
LINEAGE:
VISUALIZING CLINICAL DATA IN GENEALOGY GRAPHS

Carolina Nobre, Nils Gehlenborg, Hilary Coon, Alexander Lex
THE WORLD HEALTH ORGANIZATION ESTIMATES
ONE PERSON DIES OF SUICIDE EVERY 40 SECONDS
SUICIDE IS THE SECOND LEADING CAUSE OF DEATH IN YOUTHS BETWEEN 15 AND 29 YEARS OLD
AVERAGE SUICIDE RATE BY STATE
(National average 13.26 per 100,000)

- **Top 5 states with the highest suicide rates**
- **States with suicide rates above the national average**
- **States with suicide rates below the national average**
Utah has the 5th highest suicide rate in the country, with between 500-600 cases per year.
Acute Air Pollution Exposure and Risk of Suicide Completion

Original Contribution

The Role of Social Isolation in Suicide

Deborah L. Trout M.A.


Cited by: 95

The author wishes to thank Dr. Charles Neuringer for his assistance with the preparation of this manuscript.
Understand Complex Conditions
Discover Genetic Risk Factors

Dataset:
118k people, 19k suicide cases, ~2k with genomic data, 550 families
Based on Utah Population Database
Find familial cases that also have an “interesting” phenotype e.g., predominantly female, associated with rare psychiatric disease, etc.

Prioritize those cases for analysis of shared genomic sequences

Proofread the Data!
Age
Sex
Race
Bipolar
Depression
Asthma
Obesity
Schizophrenia
Cause of Death
Weapon Used
...
GENEALOGY WITH ~400 MEMBERS RENDERED WITH PROGENY
Family Selector
Pedigree Visualization

Family Selector
Family Selector

Pedigree Visualization

Attribute Table
LINEARIZING
Can’t show many people

Lots of missing data
Currently used by team of Psychiatry researchers on a daily basis

Widespread interest from other labs working with UPDB data

Integration of other data types
Geospatial, Environmental, Genomic
Thanks to: Carolina Nobre, Dylan Wootton, Kiran Gadhave, Zach Cutler, Marc Streit, Jochen Görtler, Oliver Deussen, Miriah Meyer, Jeff Phillips, Samuel Gratzl, Holger Stitz, Nils Gehlenborg, Hendrik Strobelt, Romain Vuillemot, Hanspeter Pfister, and many others!