Lineage: Visualizing Multivariate Clinical Data in Genealogy Graphs
We develop data visualization solutions for applications in pharmaceutical and biomedical R&D.

PRODUCTS

TARGET DISCOVERY PLATFORM

Our Target Discovery Platform is a web-based visual data analysis solution designed to score, rank, filter and visualize datasets that provides all the data and visualizations needed to identify analysis targets.
Research Areas
Large, Multivariate (Biological) Networks
Multidimensional Data

Multivariate Rankings – Lineup

Set Visualization – UpSet
Genomic Data
Cancer Subtypes / Omics
Clustering and Stratification

Alternative Splicing / mRNA-seq

Cancer Subtypes / Omics
Clustering and Stratification
Reproducibility, Storytelling, Annotation, and Integration in Computational Workflows
Genealogies & Clinical Data
The purpose of computing is insight, not numbers.

- Richard Wesley Hamming

visualization

Card, Mackinlay, Shneiderman

- Richard Wesley Hamming
Banana  M. acuminata
Date    P. dactylifera
Cress   Arabidopsis thaliana
Rice    Oryza sativa
Sorghum Sorghum bicolor
Brome   Brachypodium distachyon
Good Data Visualization

... makes data **accessible**
... combines strengths of **humans and computers**
... enables **insight**
... **communicates**
Can We Trust Statistics?

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>8.04</td>
<td>10</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>6.95</td>
<td>8</td>
<td>8</td>
<td>6.77</td>
</tr>
<tr>
<td>13</td>
<td>7.58</td>
<td>13</td>
<td>13</td>
<td>8.14</td>
</tr>
<tr>
<td>9</td>
<td>8.81</td>
<td>9</td>
<td>9</td>
<td>8.77</td>
</tr>
<tr>
<td>11</td>
<td>8.33</td>
<td>11</td>
<td>11</td>
<td>9.26</td>
</tr>
<tr>
<td>14</td>
<td>9.96</td>
<td>14</td>
<td>14</td>
<td>9.13</td>
</tr>
<tr>
<td>6</td>
<td>7.24</td>
<td>6</td>
<td>6</td>
<td>6.13</td>
</tr>
<tr>
<td>4</td>
<td>4.26</td>
<td>4</td>
<td>4</td>
<td>3.1</td>
</tr>
<tr>
<td>12</td>
<td>10.84</td>
<td>12</td>
<td>12</td>
<td>8.15</td>
</tr>
<tr>
<td>7</td>
<td>4.82</td>
<td>7</td>
<td>7</td>
<td>7.26</td>
</tr>
<tr>
<td>5</td>
<td>5.45</td>
<td>5</td>
<td>5</td>
<td>5.68</td>
</tr>
</tbody>
</table>

Mean x: 9  y: 7.50  
Variance x: 11 y: 4.122  
Correlation x – y: 0.816  
Linear regression: y = 3.00 + 0.500x
Anscombe’s Quartet

Mean x: 9 y: 7.50
Variance x: 11 y: 4.122
Correlation x – y: 0.816
Linear regression: y = 3.00 + 0.500x
Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing, CHI 2017, Justin Matejka, George Fitzmaurice
Visualization in the Data Science Process

[Diagram showing the data science process with stages like Raw Data is Collected, Data is Processed, Clean Data, Exploratory Data Analysis, Machine Learning Algorithms, Statistical Models, Make Decisions, Communicate Visualizations, Report Findings, Build Data Product, and Real World.]
Visualization = Human Data Interaction
Lineage
Visualizing Clinical Data in Genealogy Graphs

Carolina Nobre
Nils Gehlenborg
Hilary Coon
Alexander Lex
Motivation

Understand Complex Psychiatric Conditions
Discover Genetic Risk Factors
Dataset: 118k people, 19k suicide cases, 550 families based on Utah Population Database
Specific Goals

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

Proofreading the Data!
Multivariate Attributes and Graphs

How can we deal with graphs that contain rich attribute data?

[McDonnel2009]  [Gehlenborg2010]
Genealogy with ~400 members rendered with Progeny
1. De-cycle and linearize graph

2. Plot attributes in table
De-Cycling
De-Cycling
Linearization

Diagram showing the process of linearization, with nodes and arcs representing the structure before and after linearization.
Linearization
Can't show many people

Lots of missing data
Aggregation

People of Interest
Aggregation

One row for every person of interest

Others have to share a row
### Aggregated Rows

| KindredID | RelativeID | sex | deceased | suicide | Bipolar | MaxBMI | Depression | AgeID_Depression | cause_death |
|-----------|------------|-----|----------|---------|---------|--------|-----------|------------------|-------------|-------------|
| 1         | 903988     | 59% | 45%      | 9%      | 47%     | 50%    | Total:59   |                  |             |             |
| 3         | 903988     |     |          |         |         |        | Total:32   |                  |             |             |
| 1         | 903988     | #60996 |          |         |         |        |           |                  |             |             |
| 9         | 903988     | ... |          |         |         |        |           |                  | gunshot woun... |             |
| 1         | 903988     | #13789 |          |         |         |        |           |                  | asphyxia     |             |
| 7         | 903988     | ... |          |         |         |        |           |                  | carbon monox... |             |
| 1         | 903988     | #26695 |          |         |         |        |           |                  |             |             |
| 1         | 903988     | #40329 |          |         |         |        |           |                  |             |             |
More Aggressive: Hiding
Implicit Encoding of Family

No Aggregation

Aggregation

Parents

Children
Search and Filter Strategies

Find families enriched for a trait

Scan relevant cases for relationships
Live Demo
Next Steps

Show variants and other genetic data

Figure mode

Extend to other datasets

UPDB Users: Cancer, other psychiatric, cardiovascular, etc.

Genealogical datasets becoming more common

Smaller pedigrees, like trios?

Phylogenies, ...
Could we use something like Lineage for general Multivariate Networks?
Juniper: A Tree+Table Approach to Multivariate Graph Visualization

Carolina Nobre  Marc Streit

Alexander Lex
We’re hiring PostDocs and accept PhD Students!

Miriah Meyer  Alexander Lex
Thanks to: Carolina Nobre, Hilary Coon, Marc Streit, Nils Gehlenborg

Funded by the Utah Genome Project, NIH, NSF and DoD