Visualization and Rhetoric

Key Concerns for Utilizing Big Data in Humanities Research

A Case Study of Vaccination Discourses: 1918-1919
Principal Investigators

- Tom Ewing (History/VT)
- Bernice Hausman (English/VT)
- Bruce Pencek (University Libraries/VT)
- Naren Ramakrishnan (Computer Science/VT)
- Gunther Eysenbach (Centre for Global eHealth Innovation/University of Toronto)

Graduate Research Assistants

- Samah Gad (Computer Science/VT)
- Kathleen Kerr (English/VT)
- Michelle Seref (English/VT)
- Laura West (History/VT)
Methods

• Historical Newspapers
  – *Chronicling America* Database
  – *Peel’s Prairie Provinces* Database

• Computational Analysis
  – Topic modeling and segmentation
  – Tone analysis
  – Reticulation analysis

• Interpretive Analytics (Close/Manual Analysis)
Four Project Case Studies

• Weekly Newspapers
  – 24 papers
  – 1,000+ pages

• Daily Newspapers
  – 16 papers
  – 21,000 pages

• Public Health Officials
  – Royal S. Copeland

• Vaccination-Visualization
  – 90 Papers
  – 1918-1919

Morning Oregonian, November 18, 1918, p. 1
Vaccination-Visualization Study

• Original Research Questions
  • Newspaper reporting on vaccination?
  • Shifts in vaccination discourses?
  • Epidemic impact on vaccination discourses?

• Revised Research Questions
  • How visualizations persuade?
  • Rhetorical work of visualization conventions?
  • Visualization’s impact on analysis?

• Methods
  • No cholera/no black leg
  • No unwanted terms
A Word About Rhetoric

The original Greek, the word Rhetoric comes from the verb *eirô*, "to say."

But rhetoric much more than just talk.

It's a way of thinking about how we talk...

That takes into consideration the entire situation in which that talk takes place.

Rhetoric is about the spaces in which we write, converse, debate, and share ideas.
Rationalist vs. Social Constructionist View

• **Rationalist**
  – Goal to represent data credibly
  – Ideal achievable with right format

• **Social Constructionist**
  – “Correct” format negotiated
  – Needs of both producer and user important
  – Design conventions often naturalized
Visualization Choice #1: Tag Clouds
Analysis = Narrative

- Bounded
- Discrete
- Stories
Visualization Choice #2: ThemeDelta
Analysis = Temporal

- Term Patterns over Time
- Entire Period of Analysis
- Trends in Terms/Clusters
### Table 6: Word Frequency Lists: 1/10/1918 – 3/7/1918 Segment

<table>
<thead>
<tr>
<th>Topic: 1</th>
<th>Topic: 2</th>
<th>Topic: 3</th>
<th>Topic: 4</th>
<th>Topic: 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>call</td>
<td>school</td>
<td>spent</td>
<td>street</td>
<td>typhoid</td>
</tr>
<tr>
<td>german</td>
<td>county</td>
<td>day</td>
<td>camp</td>
<td>vaccination</td>
</tr>
<tr>
<td>club</td>
<td>vaccinate</td>
<td>week</td>
<td>war</td>
<td>smallpox</td>
</tr>
<tr>
<td>inoculate</td>
<td>board</td>
<td>home</td>
<td>time</td>
<td>physician</td>
</tr>
<tr>
<td>cent</td>
<td>day</td>
<td>miss</td>
<td>day</td>
<td>disease</td>
</tr>
<tr>
<td>hay</td>
<td>red</td>
<td>youngstown</td>
<td>committee</td>
<td>vaccine</td>
</tr>
<tr>
<td>jesu</td>
<td>health</td>
<td>visit</td>
<td>arm</td>
<td>vaccinate</td>
</tr>
<tr>
<td>free</td>
<td>color</td>
<td>john</td>
<td>ohio</td>
<td>health</td>
</tr>
<tr>
<td>people</td>
<td>cent</td>
<td>night</td>
<td>special</td>
<td>fever</td>
</tr>
<tr>
<td>inoculation</td>
<td>week</td>
<td>daughter</td>
<td>week</td>
<td>house</td>
</tr>
<tr>
<td>government</td>
<td>smallpox</td>
<td>family</td>
<td>lie</td>
<td>result</td>
</tr>
<tr>
<td>kidney</td>
<td>price</td>
<td>son</td>
<td>son</td>
<td>danger</td>
</tr>
<tr>
<td>propaganda</td>
<td>city</td>
<td>church</td>
<td>school</td>
<td>city</td>
</tr>
<tr>
<td>poison</td>
<td>children</td>
<td>entertain</td>
<td>doctor</td>
<td>ease</td>
</tr>
<tr>
<td>life</td>
<td>bank</td>
<td>entertain</td>
<td>little</td>
<td>army</td>
</tr>
<tr>
<td>country</td>
<td>jasper</td>
<td>entertain</td>
<td>vaccinate</td>
<td>american</td>
</tr>
<tr>
<td>world</td>
<td>lost</td>
<td>entertain</td>
<td>company</td>
<td>dis</td>
</tr>
<tr>
<td>john</td>
<td>public</td>
<td>entertain</td>
<td>town</td>
<td>medical</td>
</tr>
<tr>
<td>house</td>
<td>ship</td>
<td>entertain</td>
<td>home</td>
<td>tell</td>
</tr>
<tr>
<td>record</td>
<td>entertain</td>
<td>entertain</td>
<td>receive</td>
<td>bad</td>
</tr>
</tbody>
</table>
Analysis = Hierarchical

- Term’s Importance in Segment
- Comparison within Segment
- Word Count
Visualization Choices Are Never Neutral

1. They require tacit knowledge.
2. Practices are frequently normalized.
3. Underlying assumptions often un questioned.
4. Different visualizations suggest different interpretations.