Universität Bremen

ifib

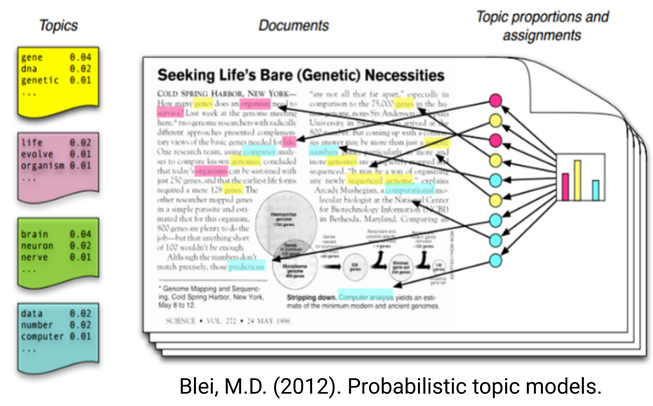
Visualization Needs in Computational Social Sciences

Hendrik Heuer, Anna Polizzotto, Franziska Marx, Andreas Breiter

Motivation & Research Question

- Machine learning techniques are increasingly applied by social scientists
- understanding visualization needs is important for computational social science, which Lazer et al. characterized *"as a field that leverages the capacity to collect and analyze data at scale to examine patterns of individual and group behaviors"*
- Research Question: **What contextual visualization needs do these researchers have when confronted with such complex systems?**

Machine Learning & Topic Modelling



Blei, M.D. (2012). Probabilistic topic models.

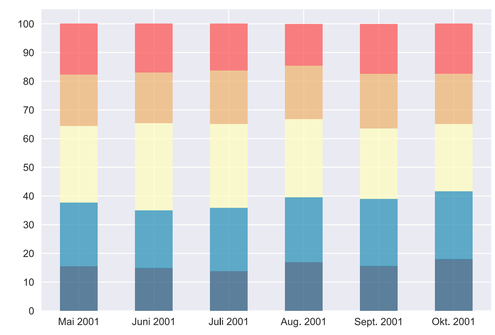
Topics in **956,000 articles** published in the **New York Times** computed using **Latent Dirichlet Allocation (LDA)**

The New York Times

Method

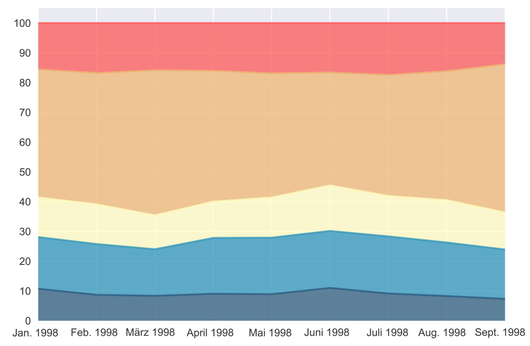
- within-subjects design, where 7 social science and humanities students used visualizations of how the topics in a newspaper changed over time
- Participants performed three tasks:
  - **Reading Values:**
    "Identify the exact proportion of topic X in month M."
  - **Comparing Proportions:**
    "How much is the proportion of topic X in month M in relation to the proportion of topic Y in month N?"
  - **Recognising Patterns:**
    "Name all patterns that you can recognize for all topics.'"
- interviews & qualitative coding with inductive category formulation following Mayring

Visualizations



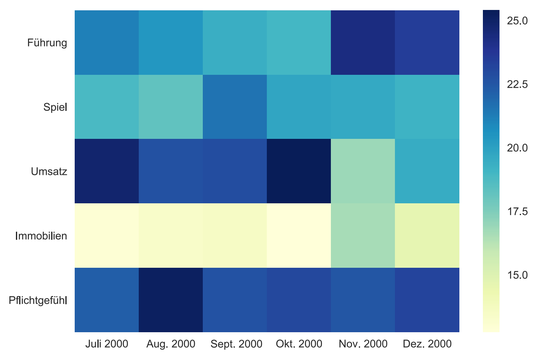
Stacked bar chart

- good readability of exact values
- required subtraction of lower unit values
- not always helpful when comparing individual topics



Stacked area chart

- subtraction of lower unit values necessary
- individual values cannot be precisely identified
- known chart type
- uncertainty about the interpretation strategy



Heat map

- no subtraction of underlying unit values necessary
- intuitive interpretation of the hue
- poor readability of accurate values
- visually appealing through structure
- confusion when comparing topics
- visualization that required the least effort

Visualization Needs

- confirmed universal needs like readability, accuracy, and precision
- found a need to quickly establish an overview and a need for intuitively interpretable visualisations
  - heat maps had a big advantage here since color appeared to be intuitively interpretable by most users
- right visualization technique is highly dependent on the data, the task, and the participant
  - stacked area charts were perceived as difficult for the tasks
- we suggest giving explicit instructions and making interpretation strategies explicit