Background

- Humanities and social sciences historically dominated by qualitative methods.
- Researchers in these fields increasingly turning to quantitative methods and computational resources to answer larger questions.
- Still much resistance, often driven by a "technological familiarity gap".

So why expect users to start with this?

Programming is great! But...

- Users often lack the background and have little access to training.
- No institutional rewards for developing technical skills.
- Programming vs GUI: recall vs. recognition.

TAG: The Text Analytics Gateway

Basic Text Analysis Tools Without the Programming

- Written mostly in R.
- Extensive re-use of existing R packages.
- UI served via shiny [1] and shiny server.
- Some custom high-performance C.
- Help files written in markdown, rendered on the fly.

Offer full analysis pipeline, by combining:

- Data retrieval
- Preprocessing
- Analysis
- Reproducibility

into a single, complete workflow.

TAG Features

Detailed summary statistics

Includes LDAvis [2] for powerful topic modeling visualization:

And more...

- **Processing**: Text filtering and transformations.
- **Vis**: Dispersion plots and wordclouds.
- **Analysis**: LDA and n-gram modeling.
- **Reproducibility**: State management, script generation.

Future Work

- Stand up gateway on Comet.
- Incorporate multidocument support.
- Add sentiment analysis feature.
- Develop userbase and community.

References


Acknowledgements

This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number OCI-1053575 and the XSEDE Science Gateways Program.