CENTER FOR TEXTUAL STUDIES AND
DIGITAL HUMANITIES

DIGH 400 - Introduction to Digital Humanities Research

Fall Semester 2014

Week 12

Today's Class

- XSLT Exercise Week 11
- TEI Exercise Week 9
- Arnolfini Portrait
- www.digital-humanities.com
- SciFi Authors - Jules Verne work
- Stylistic Analysis
- GIS in DH

XSLT Exercise - Demos

- Quick show and tell...

TEI Exercise - Sample Solutions

- Sample Solutions

Any questions?

Arnolfini Portrait

- Sample Image
- Working with image and transcription
- TEI examples
  - Facsimile | Facsimile examples
    - Surface | Surface examples
      - Zone | Zone examples

Any questions?

Digital Humanities site

- Digital Humanities site updates?

Verne Digital Corpus

- Many updates
  - Verne Digital Corpus Reference Sheets
  - The Jules Verne Encyclopedia
    - Summary of early US editions
    - Pre-1910 GB editions
    - 19th century French editions
    - Title Cross Reference (Original and variant titles…)
    - GB and US editions
    - Miscellaneous and Unidentified works
    - Overview of Adaptations

Stylistic Analysis

Overview

- consideration of patterns in style
- influence of style on readers' perceptions
- disciplinary concerns of literary and linguistic interpretation
- patterns using computational stylistics

Stylistic Analysis

Analysing Shakespeare (Hugh Craig, 'A Companion to Digital Humanities')

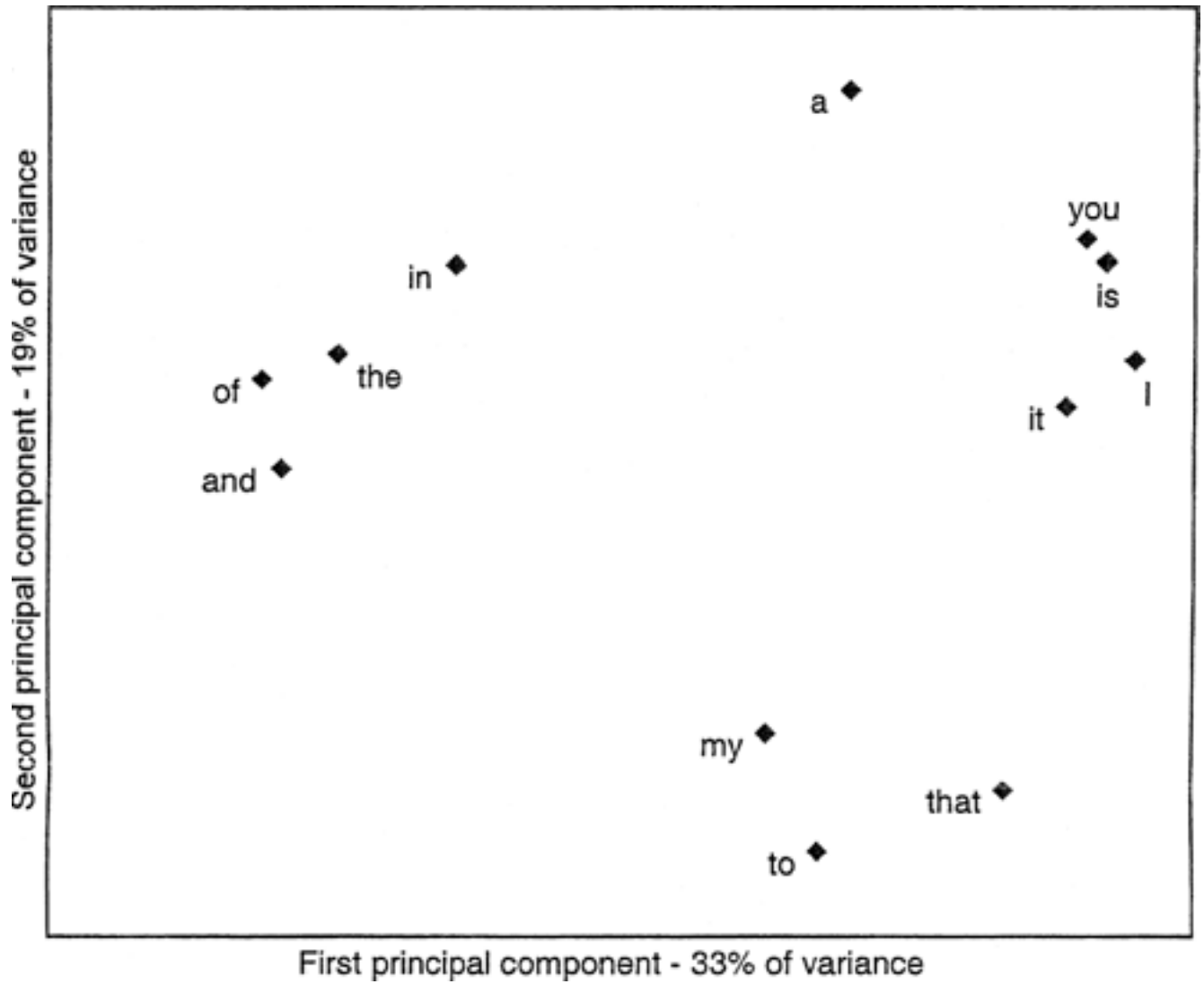
- analysis of 25 of 38 plays in standard Complete Shakespeare
- conduct analysis of the 12 most commonly used words
- Principal Components Analysis (PCA)
- PCA simplifies data by finding new variables that represent most relationships
- new composite variables still represent the variation in a data set
- PCA vectors are an extension of this principle

Stylistic Analysis

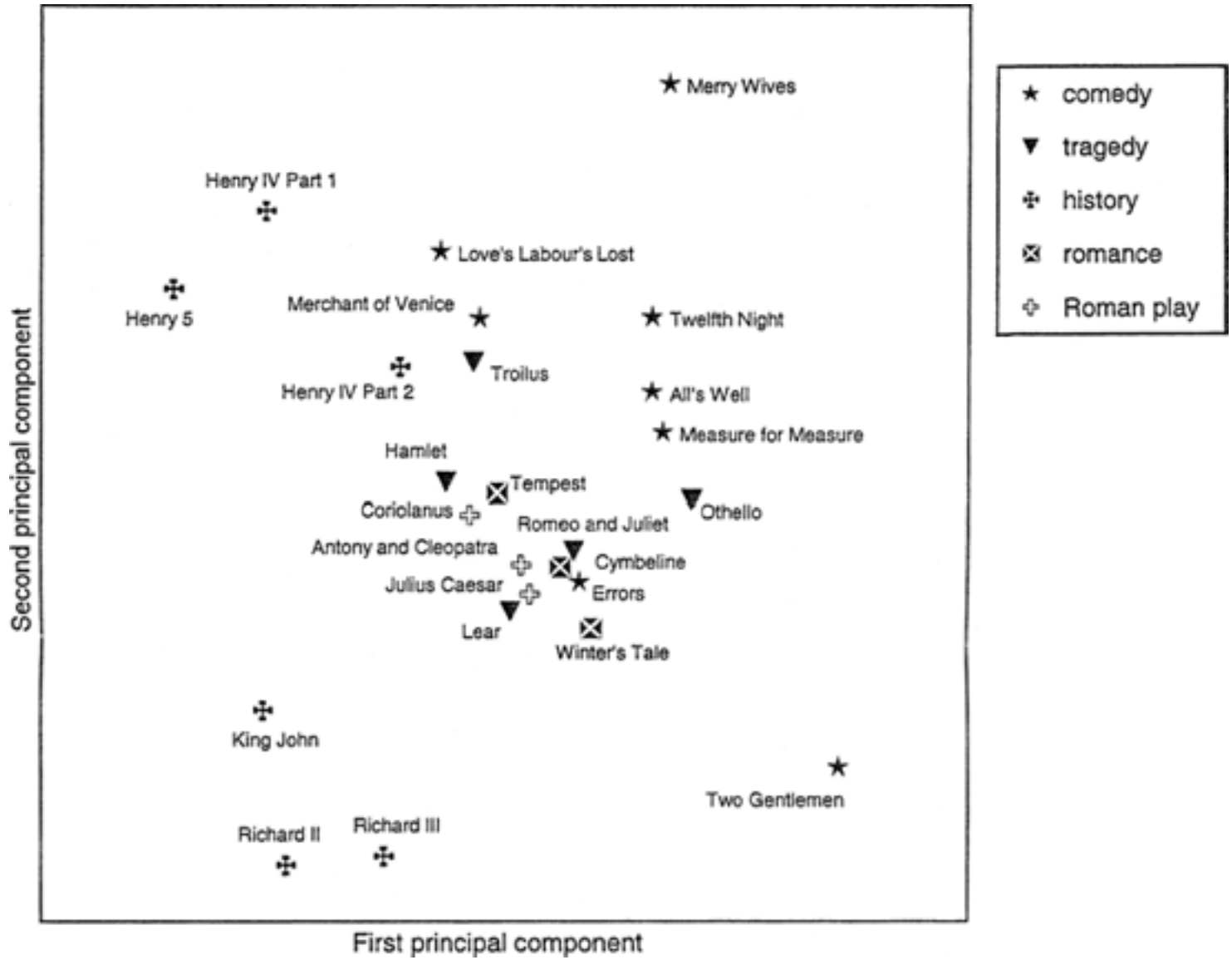
Analysing Shakespeare (Hugh Craig, 'A Companion to Digital Humanities')

- relative weightings used to create vector which accounts for greatest proportion of variance
- then the second vector, and so on...
- strong associations between variables will lead to the first few composite variables accounting for most of the results

DIGH 400 - Introduction to Digital Humanities Research



DIGH 400 - Introduction to Digital Humanities Research



Stylistic Analysis

Considerations

- consider how constant such patterns are when we remove or add a play
- effect of modifying variables on the results
- why did we choose the given parameters for this particular study?

Stylistic Analysis

Computational Stylistics

- extensive and perhaps best suited to large scale comparisons
- changes in language over time, eg: a writer's career, historical period...
- provision of a class of evidence not otherwise accessible
- not a solution by itself, requires knowledge of humanities and statistical techniques

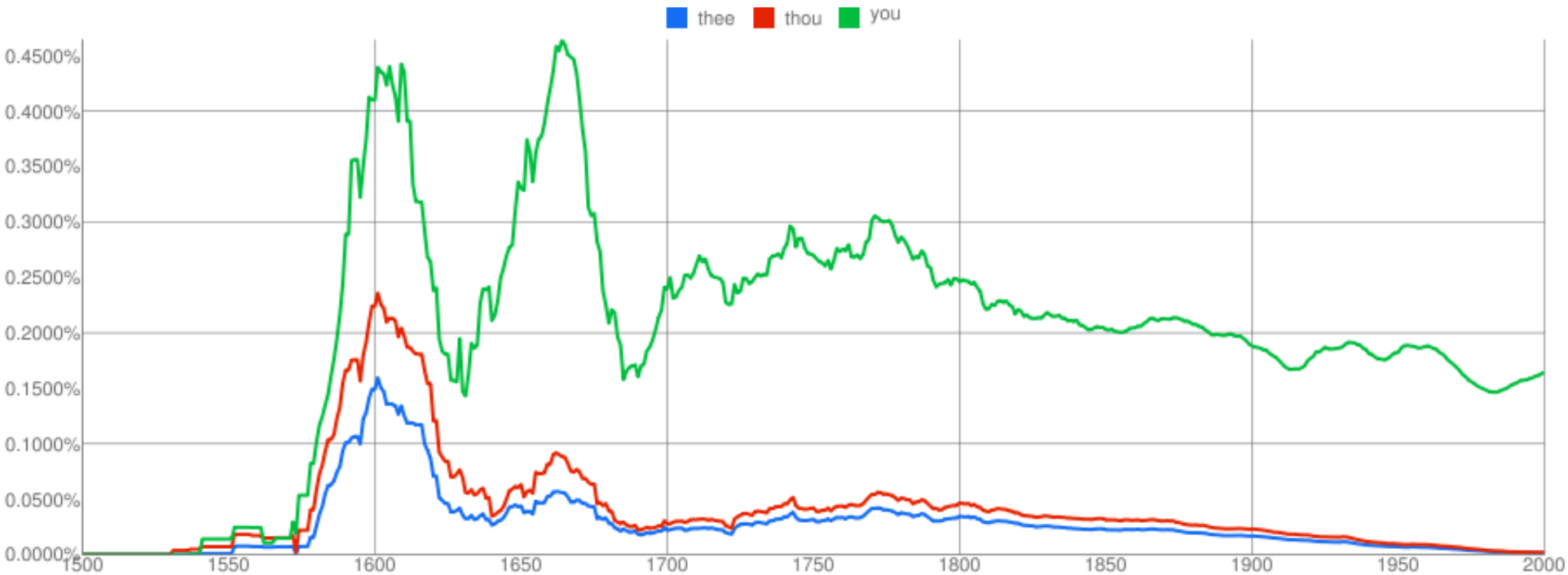
Google Books Ngram Viewer

What does it actually do?

- searches a selected corpus of books for a user selected set of phrases
- select years for search
- apply smoothing to specify moving average of results returned
- search Google Books from returned set of results
- use the returned raw data to create your own visualisations, tests...

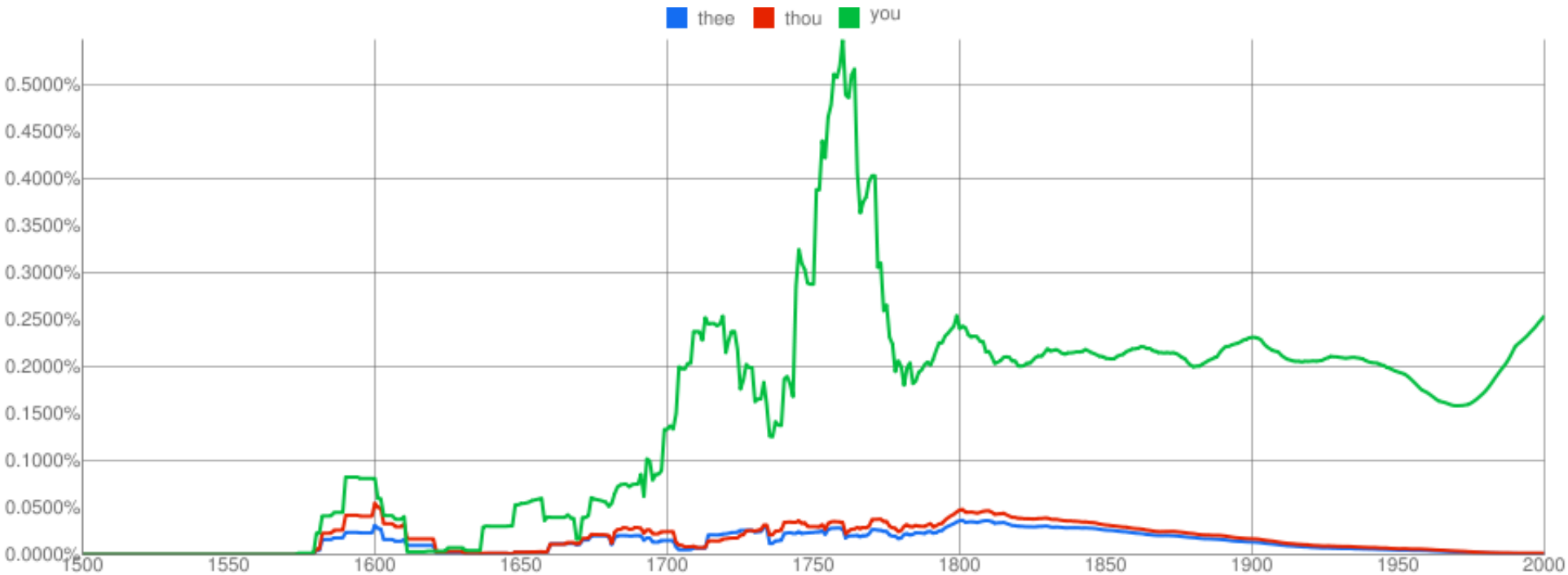
Try a few tests

DIGH 400 - Introduction to Digital Humanities Research



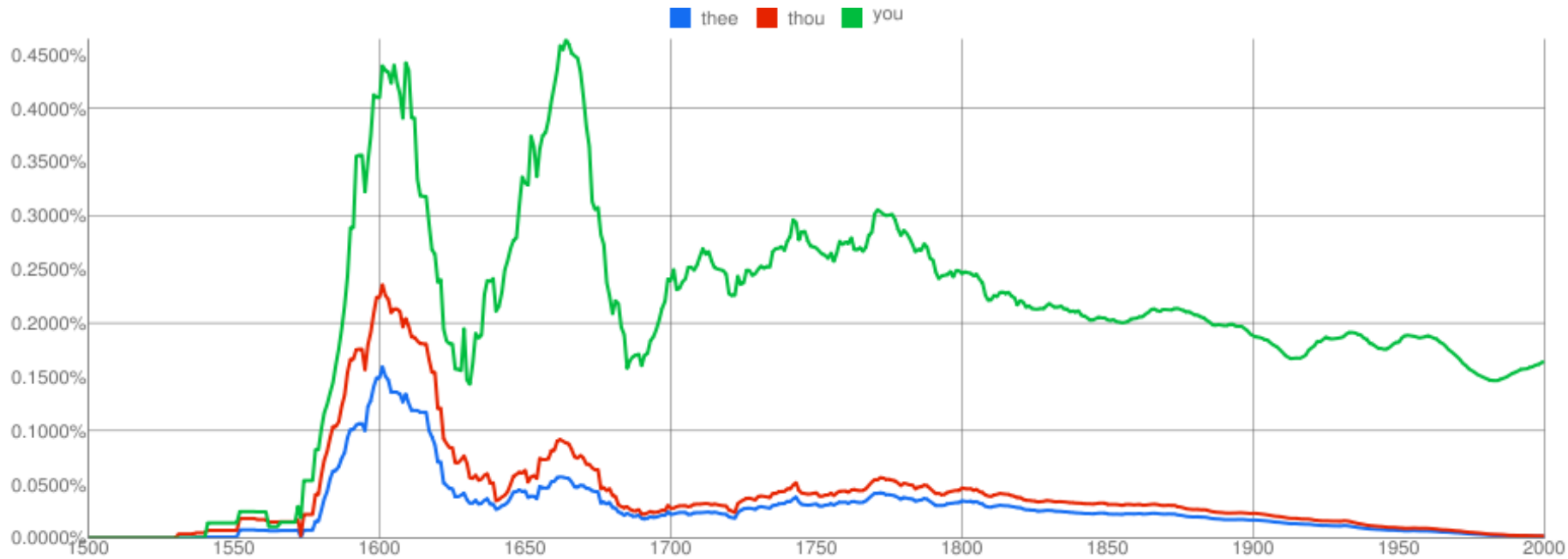
British English Corpus results for 1500 - 2000

DIGH 400 - Introduction to Digital Humanities Research

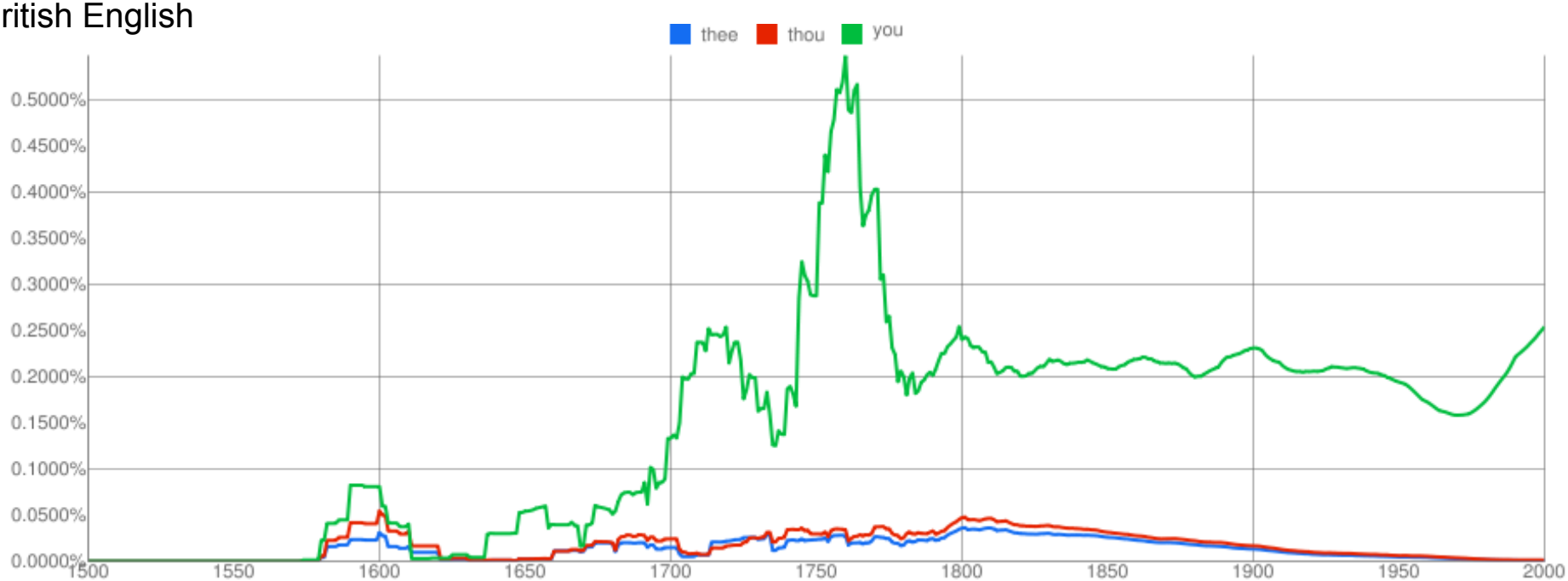


American English Corpus results for 1500 - 2000

DIGH 400 - Introduction to Digital Humanities Research



British English



American English

Google+ Ripples

- allows users to quickly and easily visualise post sharing
- describes itself as

"a way to visualize the impact of any public post."

- nested circles show generations of the shared post
- use the circles to view how the post was shared over time

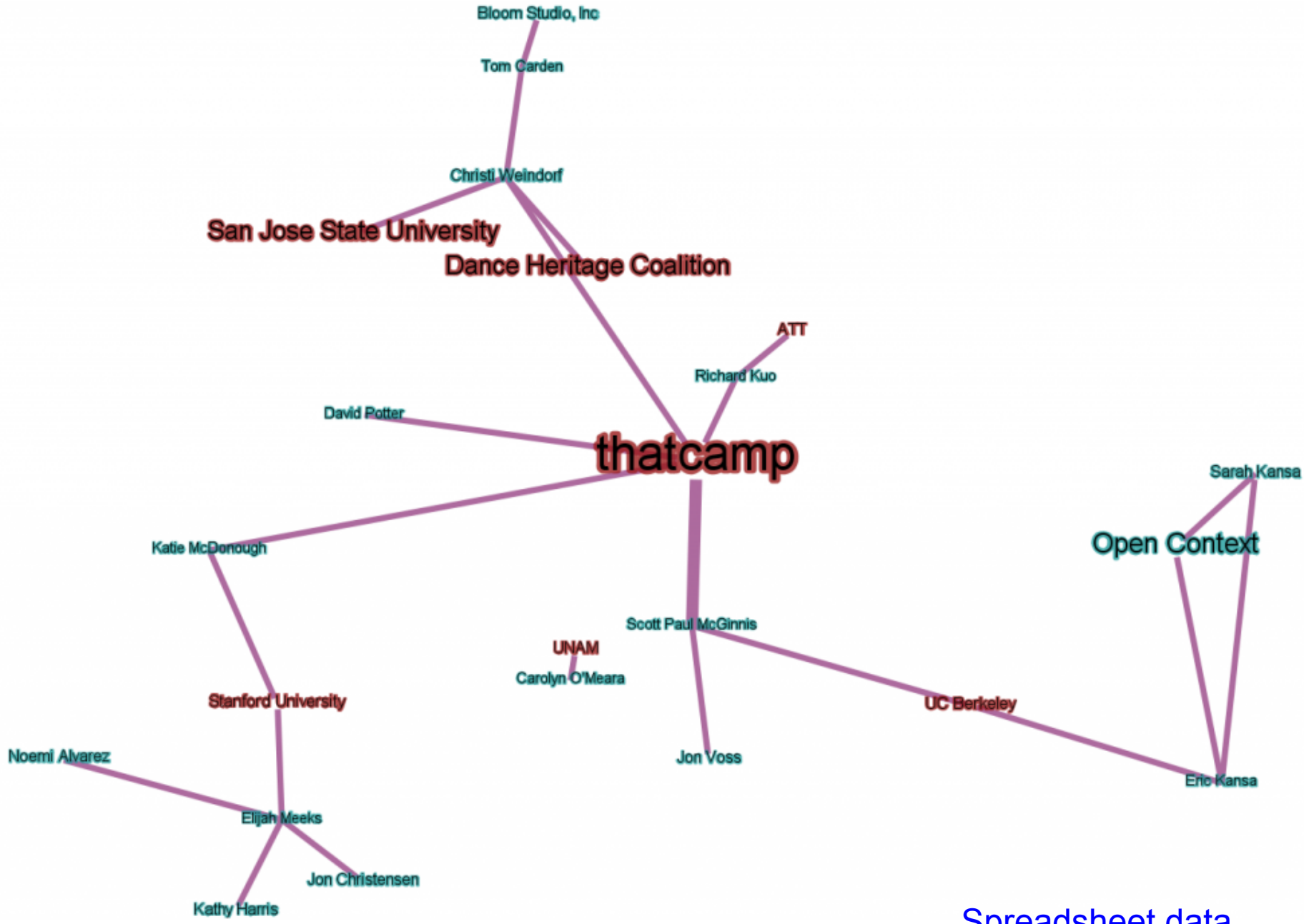
Further Info

Gephi

What does it actually do? (see https://gephi.org/about/)

- software for network visualisation and analysis
- helps data analysts to intuitively reveal patterns and trends...
- highlight outliers and tell stories with data
- displays large graphs in real time to speed up exploration
- built-in functionalities and flexible architecture for networks to
    - explore, analyse, spatialise, filter, cluster, manipulate & export

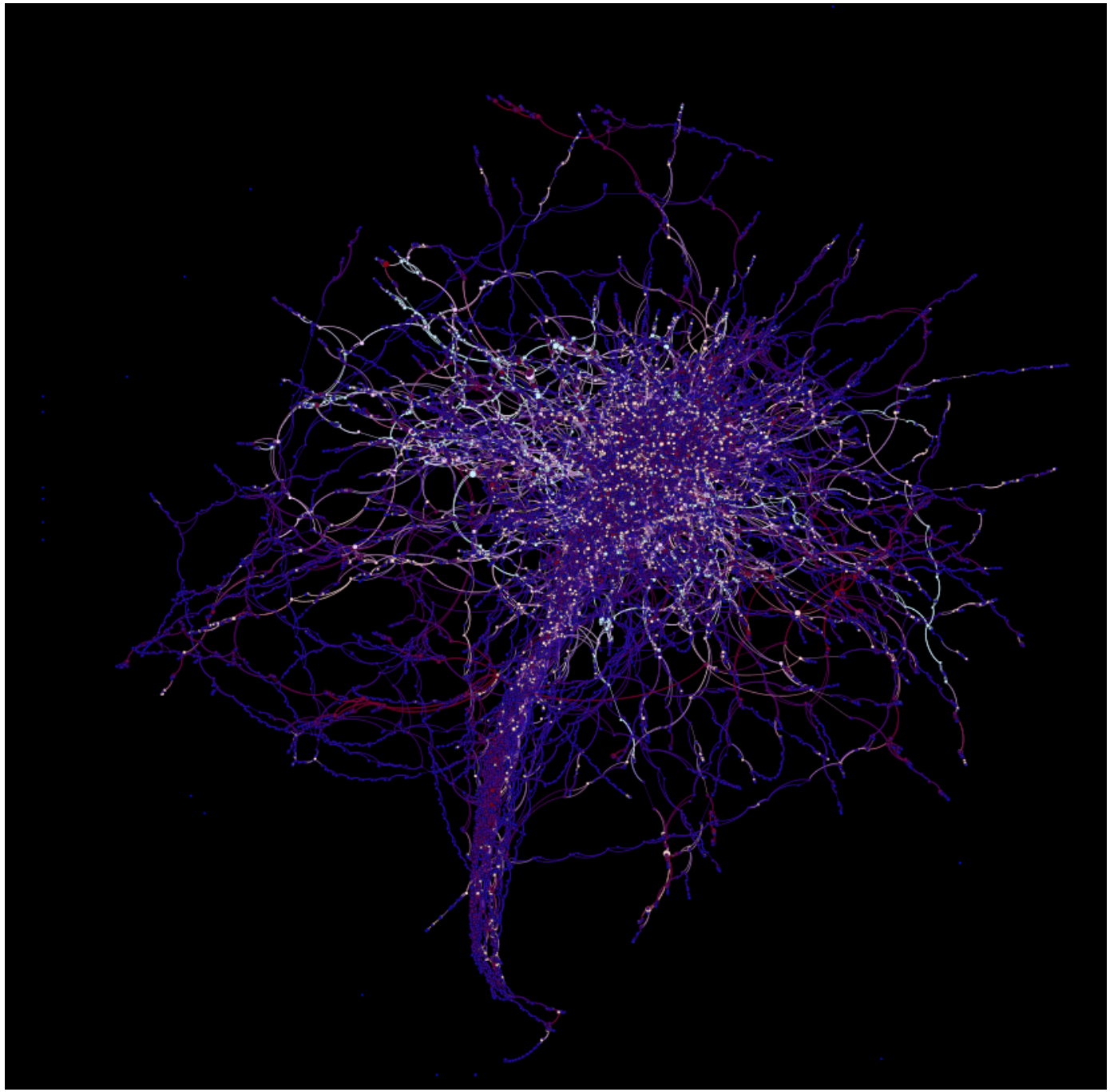
DIGH 400 - Introduction to Digital Humanities Research



[Spreadsheet data]

DIGH 400 - Introduction to Digital Humanities Research

Further Info



And now for something completely different...but somewhat relevant, and a lot of fun

Twitterology (?) - a bit of fun...

- used in diverse fields such as linguistics, sociology, and psychology
- analyse and examine language usage, patterns, location specific terms...
- immediacy and immensity
- University of Texas research into tweets and streams emanating from Libya and Egypt
- noticeable increase in usage following certain political events
- language patterns could also be discerned relative to such events

NYTimes Article on "Twitterology: A New Science"

Carnegie Mellon Article

Twitter API - what can we do?

- work with a user's timeline
  - status updates
  - search tweets…
- manage your own account
  - post, delete, maintain tweets
  - check friends, followers, IDs…
- search Twitter
  - search feeds and hashtags
  - search friends, user IDs
  - get tweets by geo location

and lots more…

Twitter Stream | Twitter API

GIS in Digital Humanities - What is GIS?

- Geographic Information System...
- system to obtain, store, manipulate, analyse, display...data
- allows us to view, understand, question, interpret, visualise data
- reveals relationships, patterns, trends...

GIS in Digital Humanities - Using GIS

A few ideas - part 1

- map positions, places, events...
- add contextual value to a project
- locate quantities and usage relative to geographical locations
- map densities and their distribution in geographical contexts
    - uses a uniform areal unit, which is arbitrarily assigned
    - often used to show political affiliation, household numbers...
    - often relative to census tracts, counties...
    - can raise issues of manipulation of data
    - problem known as 'The Modifiable Areal Unit Problem'

GIS in Digital Humanities - Using GIS

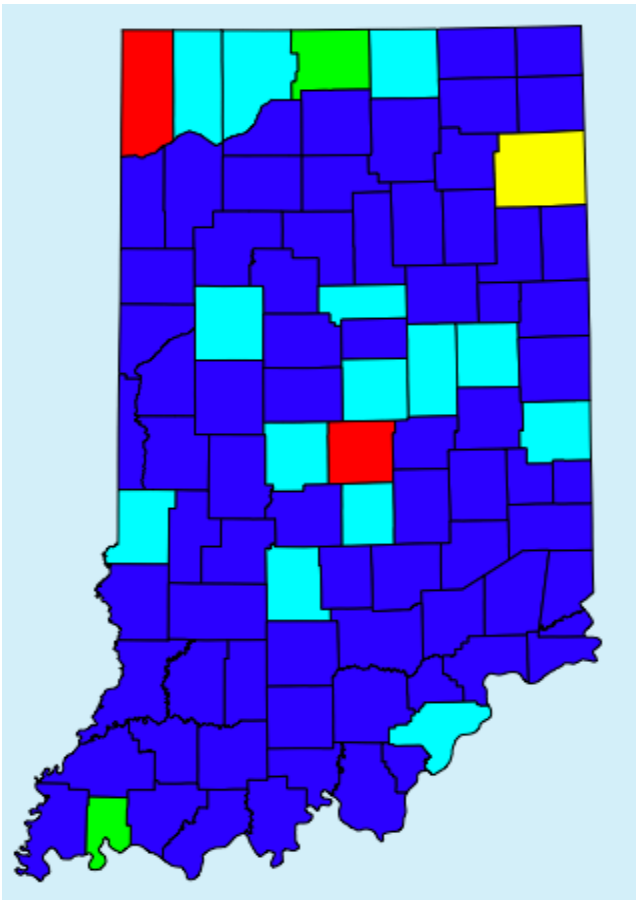
'The Modifiable Areal Unit Problem' - Openshaw, S. 1984

- referenced by Openshaw in 1984
- classic example is 'Gerrymandering'
- units of analysis are not based on geographic principles
    - instead based on political and social biases
- findings can change relative to change in arbitrarily defined units
- two important features: zone and scale
- zonal problem

GIS in Digital Humanities - Using GIS

'The Modifiable Areal Unit Problem' - Openshaw, S. 1984

Counties



Census Tracts