What Shakespeare Taught Us About Visualization and Data Science

Michael Gleicher

Department of Computer Sciences

University of Wisconsin Madison


Calgary
Ontario
Québec
Vancouver
Winnipeg
Seattle
Washington
Montana
North Dakota
Minnesota
Québec
New Brunswick
Portland
Minneapolis
South Dakota
Ottawa
Montréal
Maine
Oregon
Idaho
Wyoming
Michigan
Toronto
Milwaukee
Detroit
New York
Vermon
Salt Lake City
Iowa
Nebraska
Omaha
Chicago
Cleveland
Pennsylvania
New Ham
Massachuse
Denver
Kansas City
Illinois
Indiana
Ohio
Pittsburgh
New York
Nevada
Utah
Colorado
Sacramento
St Louis
Cincinnati
West Virginia
Rhode Island
San Francisco
California
Kansas
Wichita
Missouri
Connecticut
Kentucky
Virginia
Las Vegas
Tulsa
New Jersey
Albuquerque
Oklahoma
Tennessee
North Carolina
Delaware
Los Angeles
Arizona
Arkansas
Atlanta
Maryland
New Mexico
Mississippi
South Carolina
Phoenix
Dallas
District of Columbia
San Diego
El Paso
Alabama
Tucson
Texas
Georgia
Austin
Louisiana
Jacksonville
Houston
San Antonio
Chihuahua
Florida
Monterrey
Miami
Gulf of Mexico
México
La Habana

Acknowledgements

Visualizing English Print Project

Co-Pis

Michael Witmore (Folger Library)

Jonathan Hope (Strathclyde University)

Robin Valenza (UW English)

VEP Students and Staff

Eric Alexander, Deidre Stuffer, Erin Winter, Hao Fu, Joe Kohlmann, …

Jonathan's students (Heather, Beth, …)

Robin's students (Mattie, Cathy, …)

Other Vis Collaborators

My Students (past and present)

Danielle Szafir, Michael Correll, Alper Sarikaya, Kendall Park, …

Perception and HCI Collaborators

Steve Franconeri, Bilge Mutlu, Chuck Hansen, and their students

Other Domain Collaborators

George Phillips, Dave O'Connor, Colin Dewey, …

Lesson 1

It’s a team effort

Lesson 1b

I’m not the Shakespeare Expert

So… why am I here?

`

Michael Gleicher

Visual Computing Group

Human **Graphics** Interaction

authoring pictures, videos, animations

Human **Robot** Interaction

robots!

Human **Data** Interaction

visualization, visual analytics, interactive learning

~~Data Visualization~~

~~Visual Analytics~~

Human Data Interaction

Lesson 2

Visualization is not (just) **about making pictures**

Data Science is not (just) **about doing statistics**

They are (also) **about Human-Data Interaction**

What Shakespeare Taught Us About Visualization and Data Science

Michael Gleicher

Department of Computer Sciences

University of Wisconsin Madison

Caveats

This is not a talk about Shakespeare

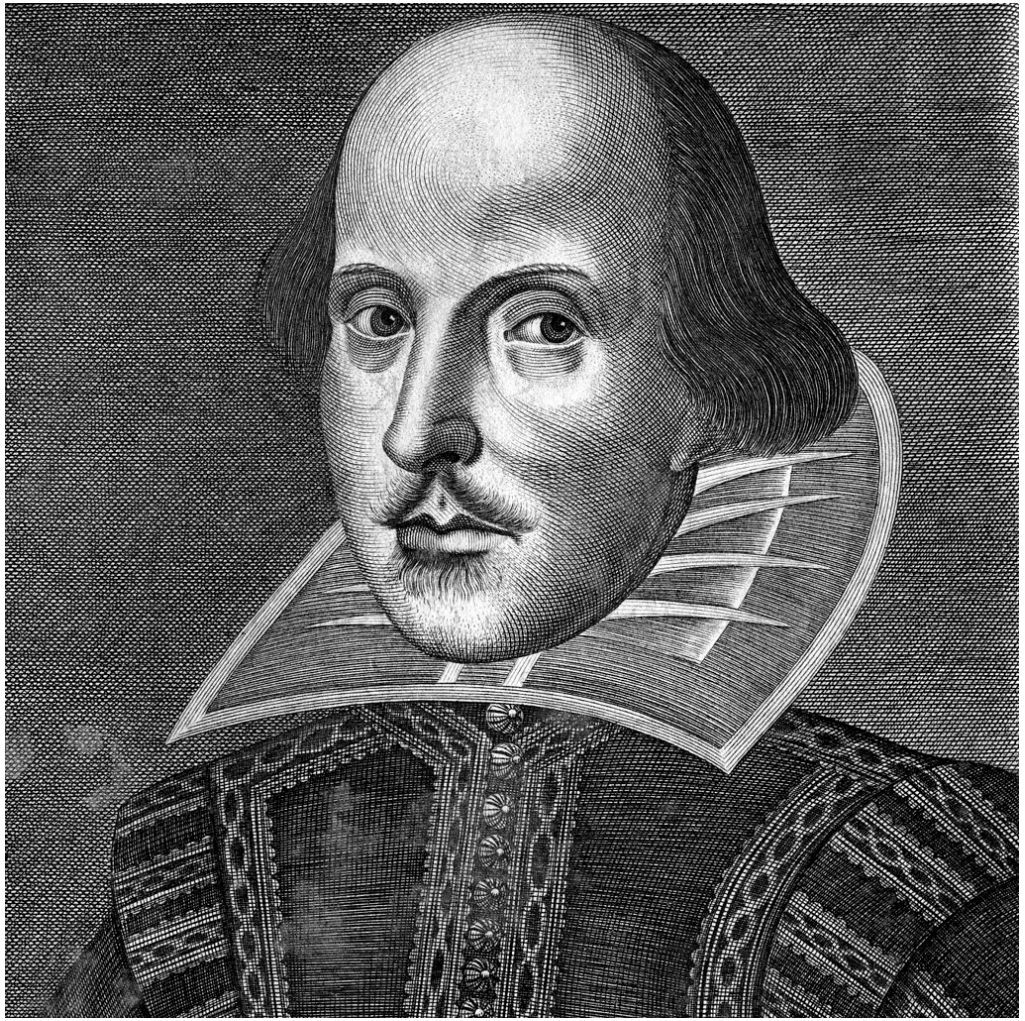
But he makes for a catchy title

It is a talk about a collaboration with **literary scholars**

With the goal of getting beyond Shakespeare

(Early Modern Period 1470-1700)

(Or is that 1470-1660? Or 1800?)



https://commons.wikimedia.org/wiki/File:Shakespeare_engraving_1787cropped.jpg

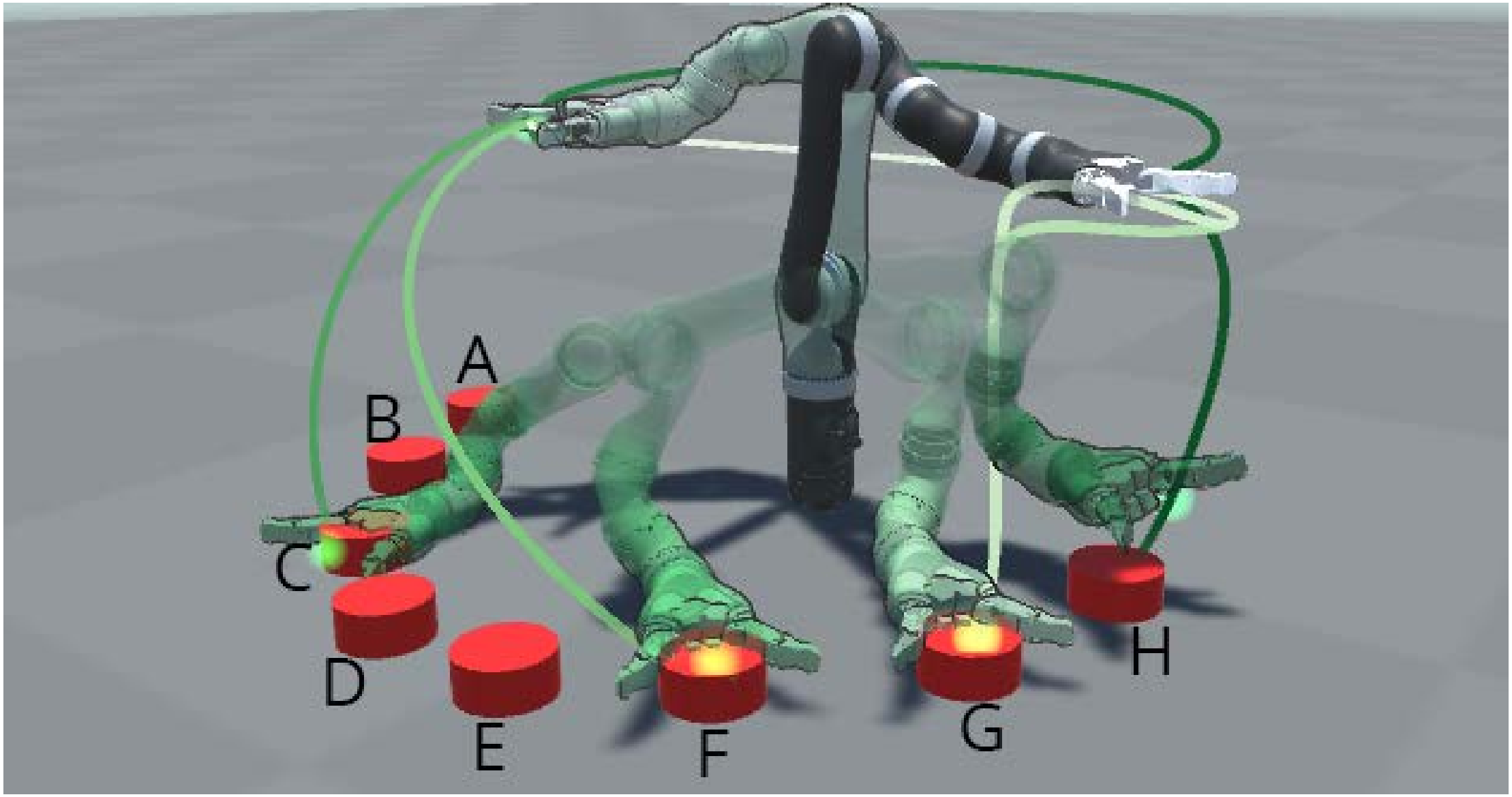
This is not a talk about Digital Humanities

This is about a collaboration with literature scholars.

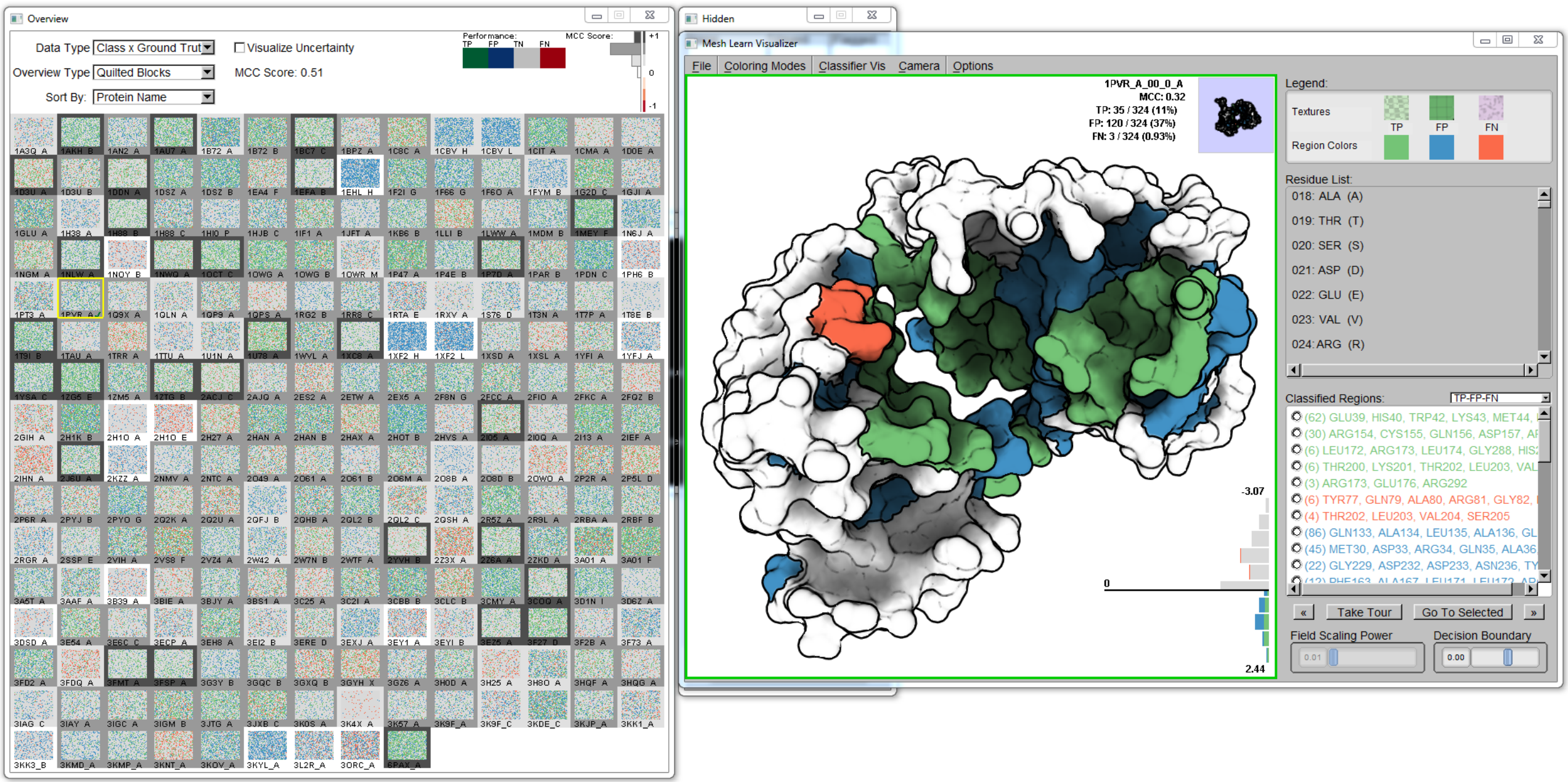
What can the rest of us learn?

Why experience with Literary Scholarship?

Why not experience with Biochemistry, Virology, Robotics, … ?



Rakita, Mutlu and Gleicher. Motion Synopsis for Robot Arm Trajectories,IEEE ROMAN 2016.



Sarikaya, Albers, Mitchel and Gleicher. *Visualizing Validation of Protein Surface Classifiers.* Computer Graphics Forum. Proceedings EuroVis 2014

Lesson 3

It’s nice to have an application that people are interested in

Why experience with Literary Scholarship?

Some lessons we could have gotten elsewhere

but this project made them more clear

Some lessons came from unique aspects of the project

but I think they are more general

Some lessons come from the unique collaboration

Humanist thinking in data analysis!*

* I dislike the term “Humanist” because of what it implies about the rest of us. But it is how “they” self-identify.

Mike's Mantra

Until we take the time to learn about how the other side thinks, we can't really work **together**.

Once we learn how each other thinks, our ways of thinking can infuse each other's.

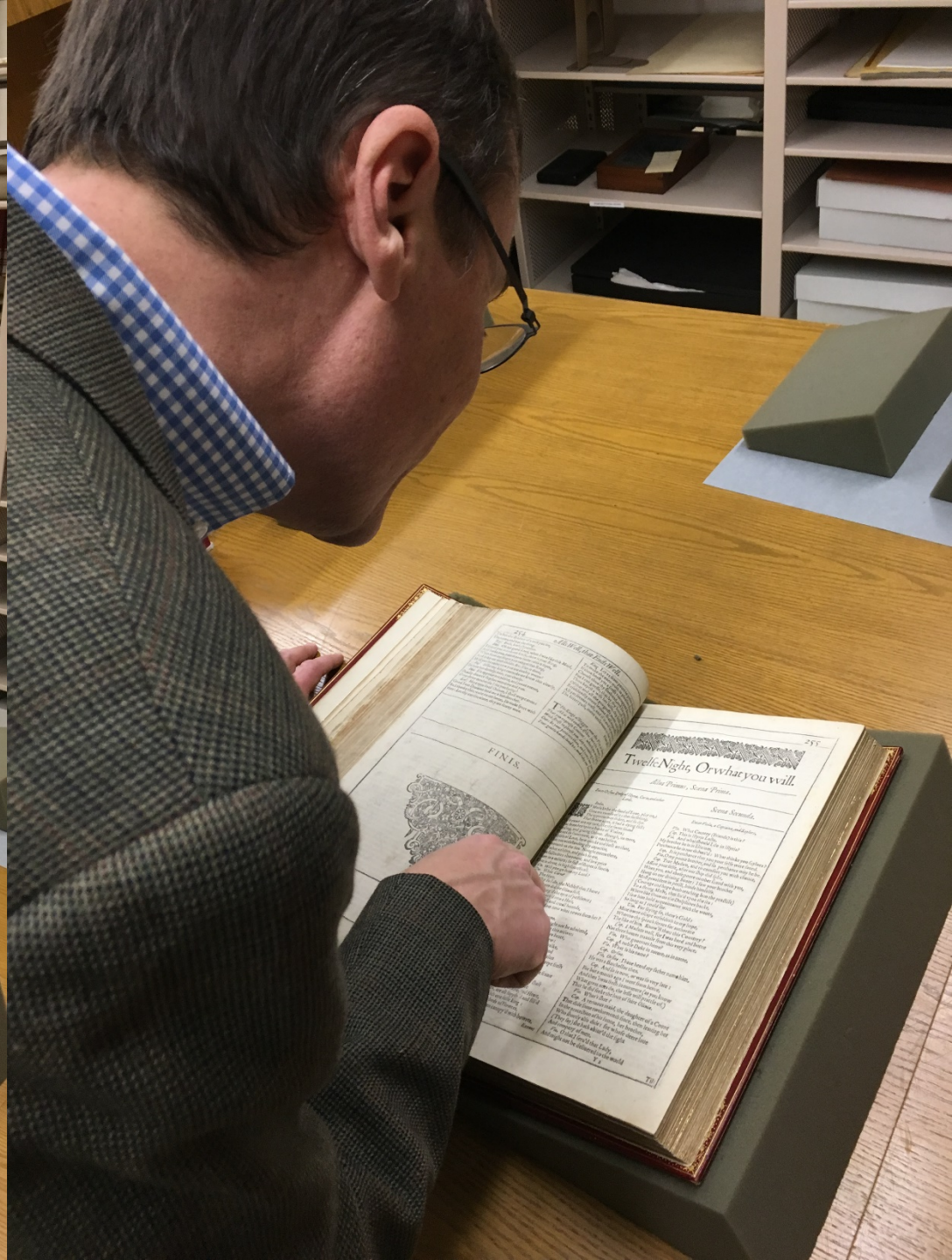
This is not just building tools for our friends.

It's a **lot** more fun and interesting

Working with Literary Scholars

Folger Shakespeare Library


To the Reader.
SHAKESPEARES


FINIS.
Twelfe Night, Or what you will.

THE TRAGEDIE OF MACBETH.

Actus Primus. Scœna Prima.

Thunder and Lightning. Enter three Witches.

1. When shall we three meet againe?
In Thunder, Lightning, or in Raine?
2. When the Hurley-burley's done,
When the Battaile's lost, and wonne.
3. That will be ere the set of Sunne.
1. Where the place?
2. Vpon the Heath.
3. There to meet with *Macbeth*.
1. I come, *Gray-Malkin*.
All. Padock calls anon: faire is foule, and foule is faire,
Houer through the fogge and filthie ayre. *Exeunt.*

Alarum within. Enter King Malcome, Donal-
baine, Lenox, with attendants, meeting

King. O valiant Cousin, worthy Gentleman.
Cap. As whence the Sunne 'gins his reflection,
Shipwracking Stormes, and direfull Thunders:
So from that Spring, whence comfort seem'd to come,
Discomfort swells: Marke King of Scotland, marke,
No sooner Iustice had, with Valour arm'd,
Compell'd these skipping Kernes to trust their heeles,
But the Norweyan Lord, surueying vantage,
With furbusht Armes, and new supplyes of men,
Began a fresh assault.
King. Dismay'd not this our Captaines, *Macbeth* and *Banquoh*?
Cap. Yes, as Sparrowes, Eagles;
Or the Hare, the Lyon:
If I say sooth, I must report they were
As Cannons ouer-charg'd with double Cracks,
So they doubly redoubled stroakes vpon the Foe:
Except they meant to bathe in reeking Wounds,
Or memorize another *Golgotha*,
I cannot tell: but I am faint,
My Gashes cry for helpe.
King. So well thy words become thee, as thy wounds,

When shall we three meet againe?
In Thunder, Lightning, or in Raine?

Some Lessons from “Humanist” Thinking

They got along fine before “data-centric” thinking

Importance of **exemplars** and **outliers**

Importance of going back to the **source** (specific passages)

Value of **multiple** points of view

Contextualized arguments with lots of **background**

Editing and **curation** are scholarly activities

Visualizing English Print

c 1470-~~1800~~ (1700)

What if you had access to all surviving books?

Visualizing English print (VEP)

Large collections of texts becoming available to scholars

How to enable (traditional) literary scholars to use it?

Problem Factory:

What questions should/could one ask?

How do you ask those questions?

What tools to provide to help?

What can we learn about Visualization / Data Science?

Why? (what is in it for literature)

Consider larger collections of books

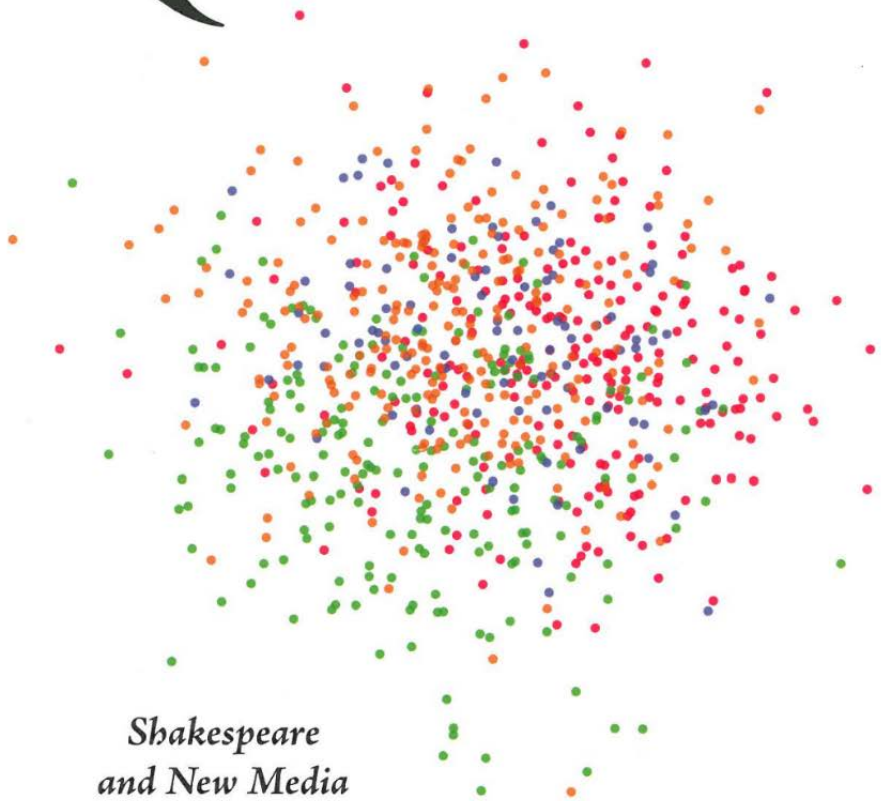
beyond the “canon”

See language develop (independent of content)

See developments over decades

See patterns too subtle for people to pick up

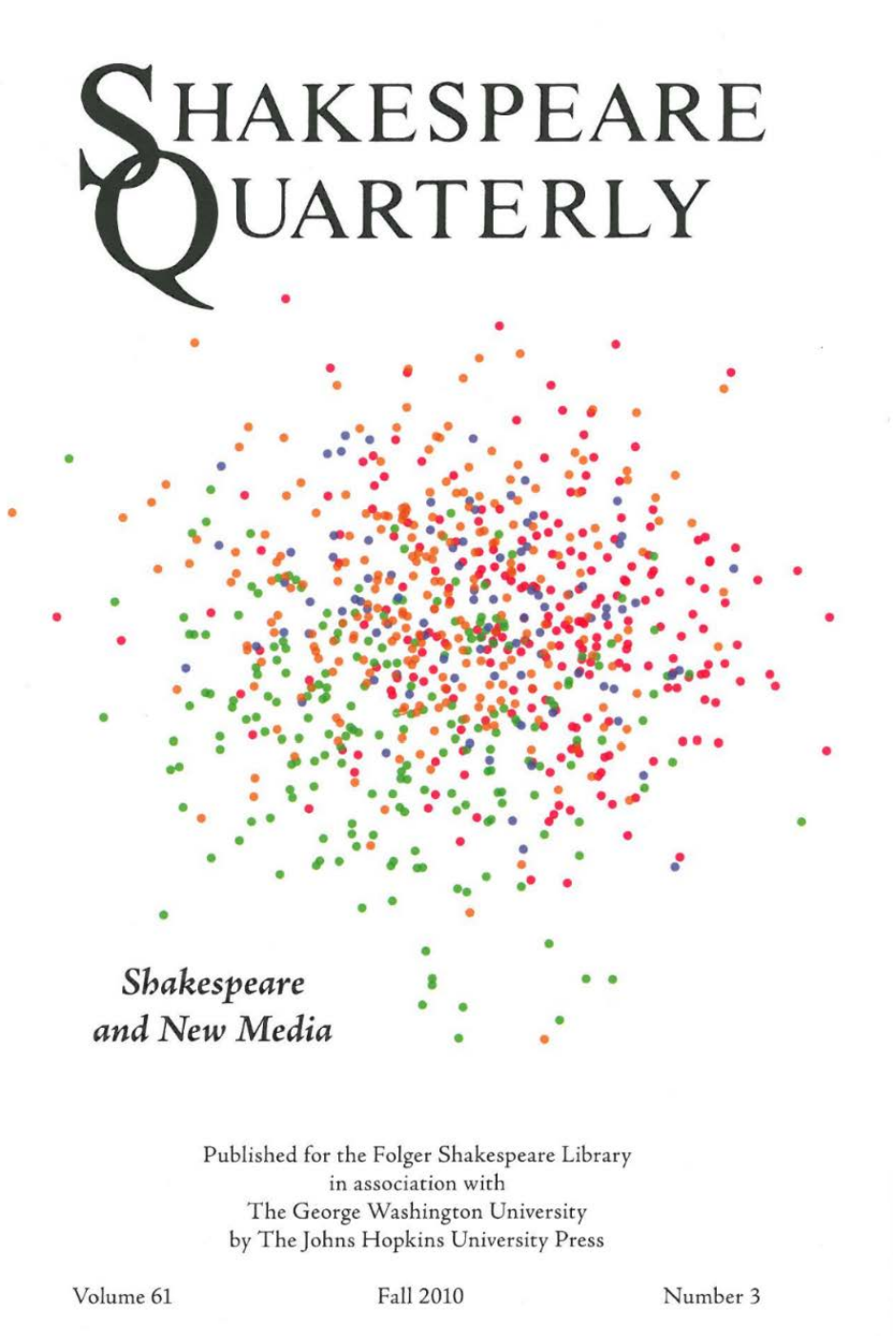
SHAKESPEARE QUARTERLY



Shakespeare and New Media

Published for the Folger Shakespeare Library
in association with
The George Washington University
by The Johns Hopkins University Press

Volume 61 Fall 2010 Number 3



One journal cover image leads to (at least) three challenges

The axes are meaningless!

Explainers – crafted projections
VAST 2013

Can people interpret this?

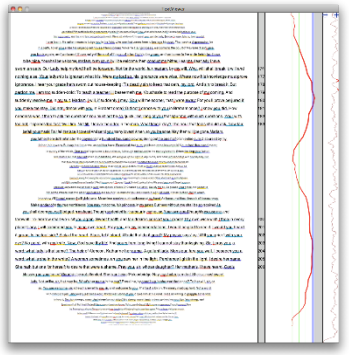
Perception of average value in scatterplots
InfoVis 2013

The scatterplot has too many points!

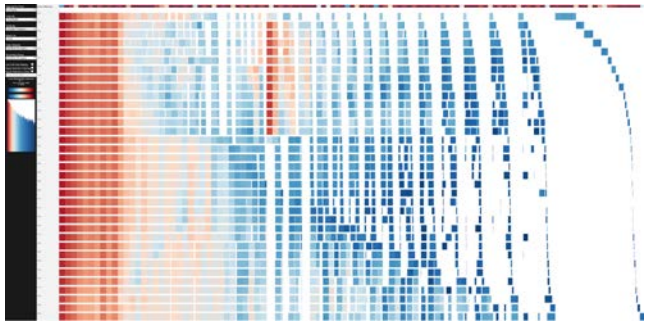
Splatterplots – scalable scatterplots
TVCG 2013

And these lead to many more...

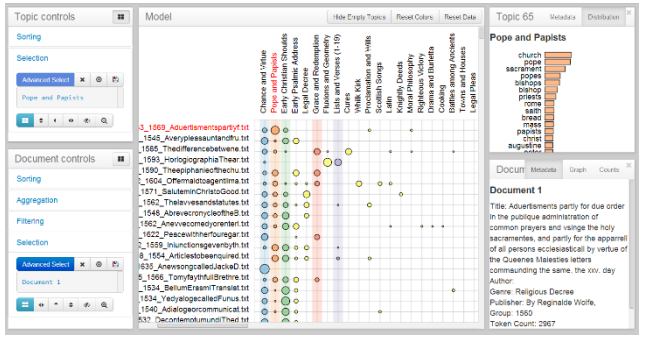
Vis research success stories



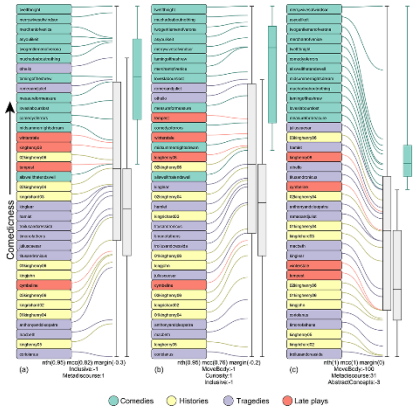
Text Viewer
Correll et al, EuroVis 2011



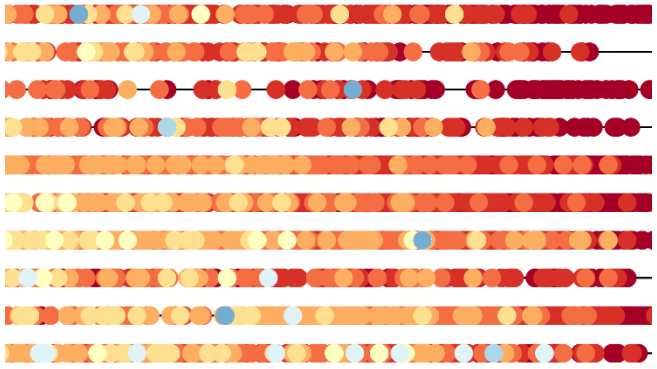
TextDNA
Szafir et al, EuroVis 2016



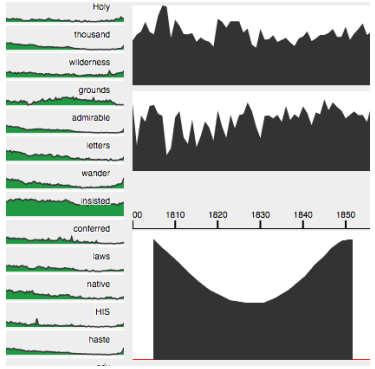
Serendip
Alexander et al, VAST 2014



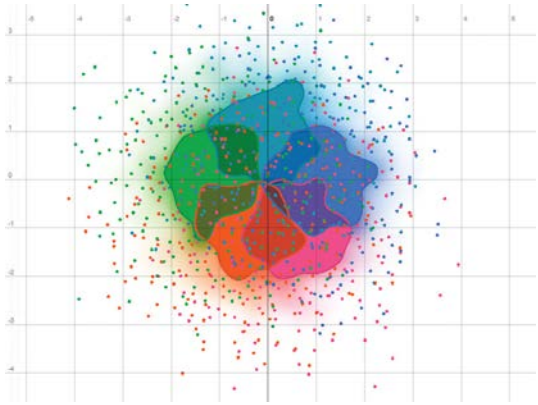
Explainers
Gleicher, VAST 2013



Topic Model Comparison
Alexander et al, VAST 2015



Sketch-based search
Correll et al, VAST 2016

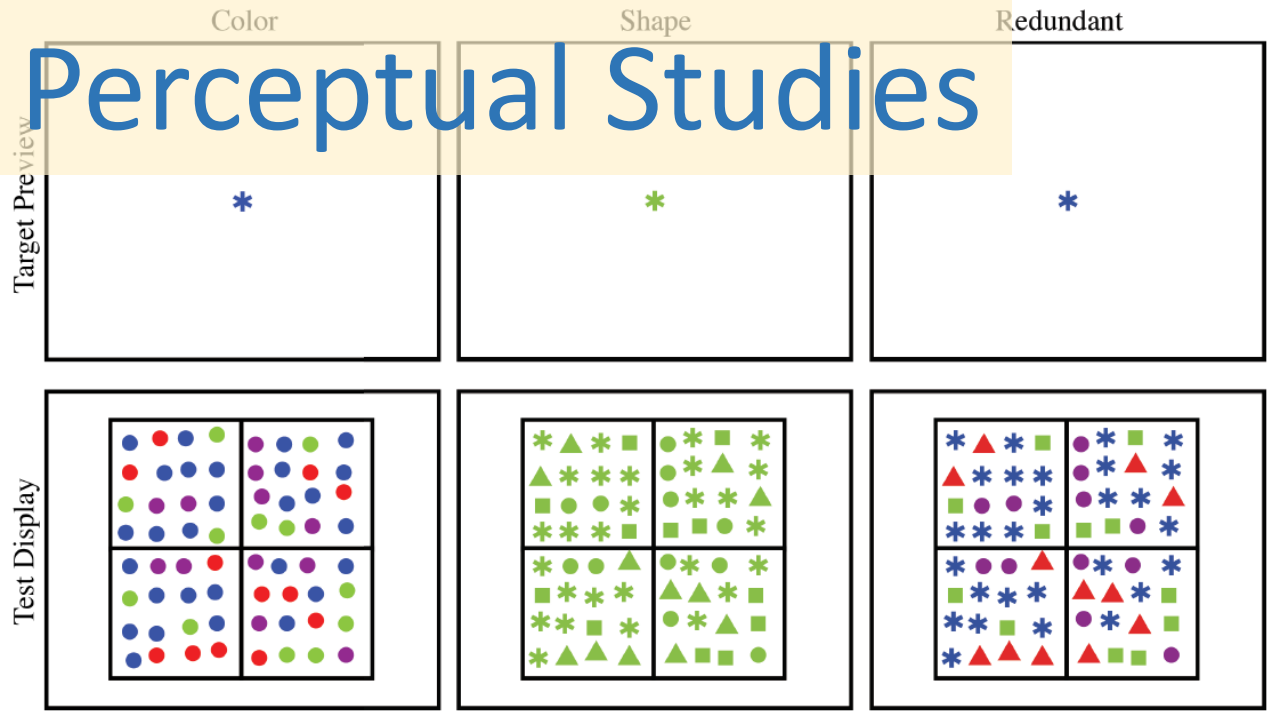


Splatterplots
Mayorga & Gleicher 2013
Sarikaya & Gleicher 2015

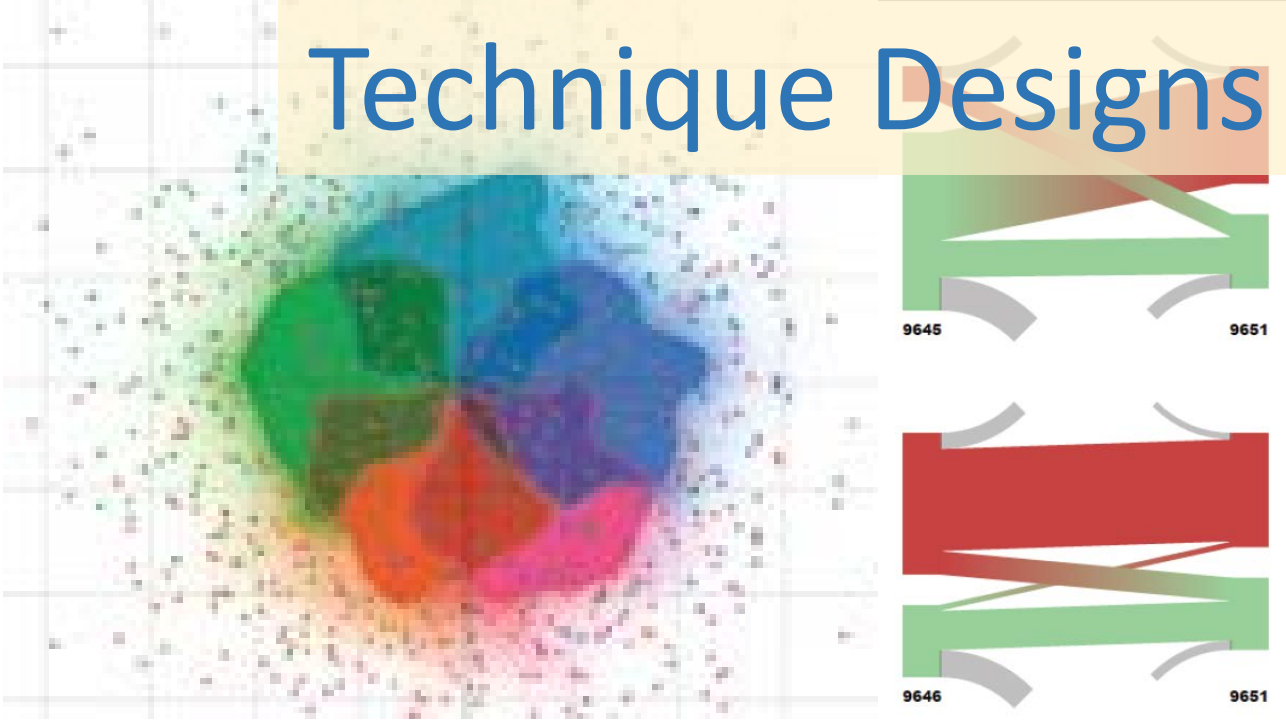
And others…

What is visualization reseach?

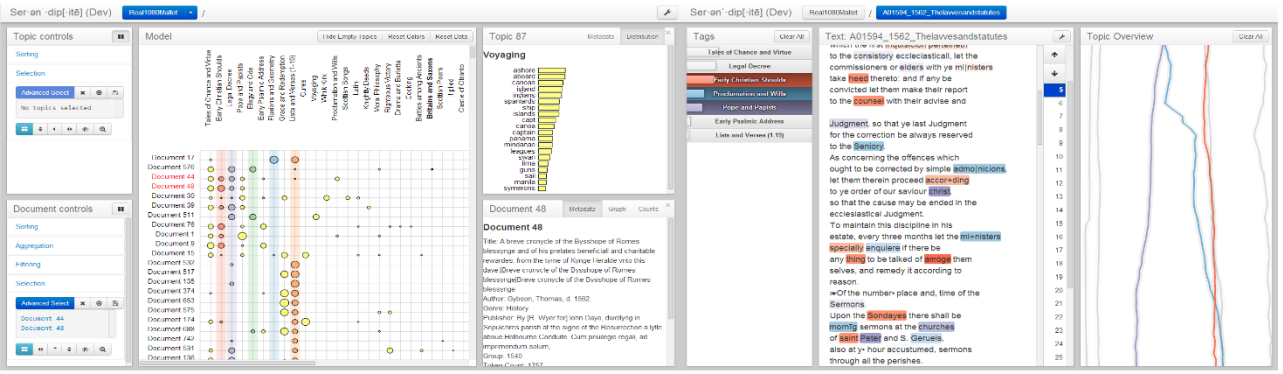
Perceptual Studies

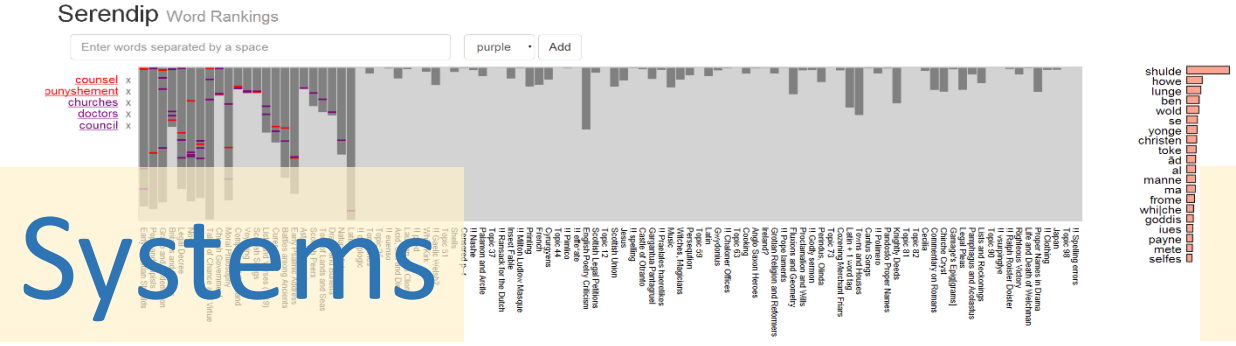


Technique Designs

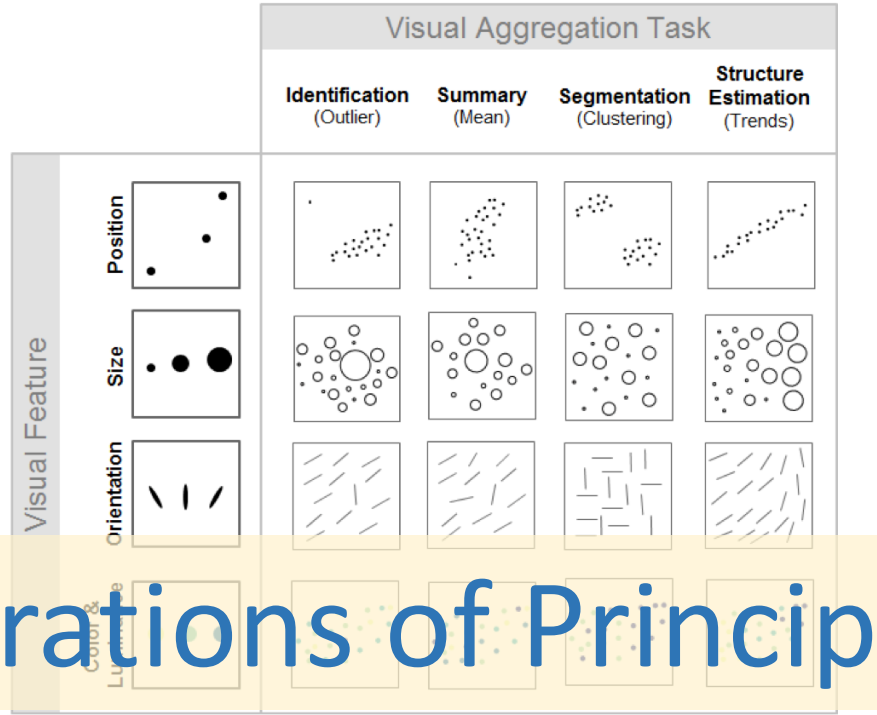


Systems

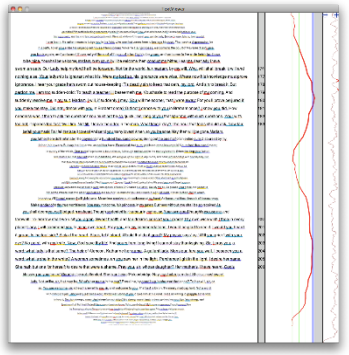




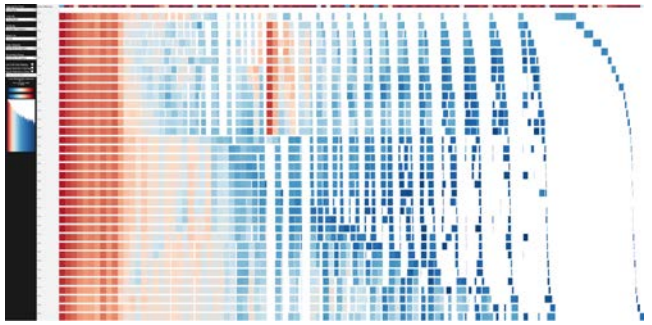
Explorations of Principles



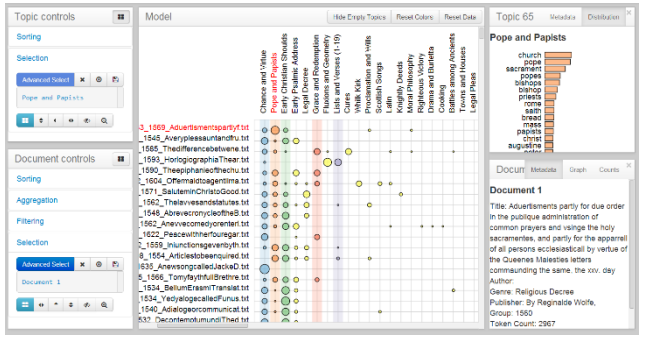
VEP Vis success stories (some of each kind)



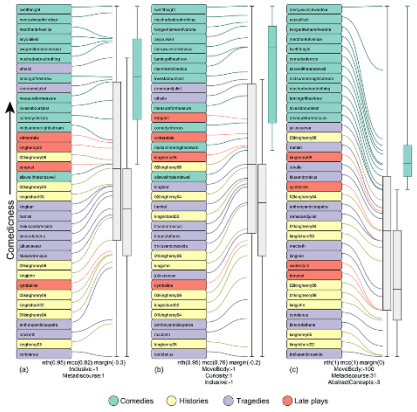
Text Viewer
Correll et al, EuroVis 2011



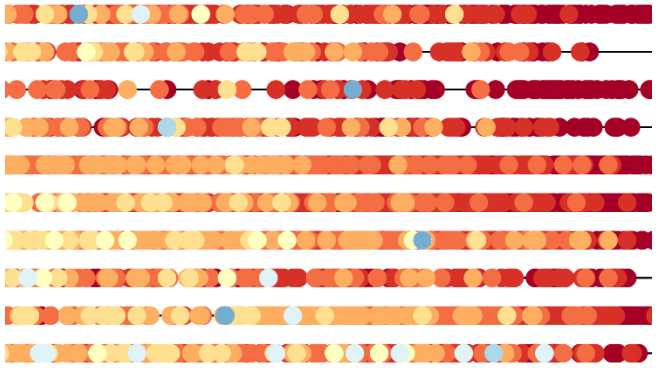
TextDNA
Szafir et al, EuroVis 2016



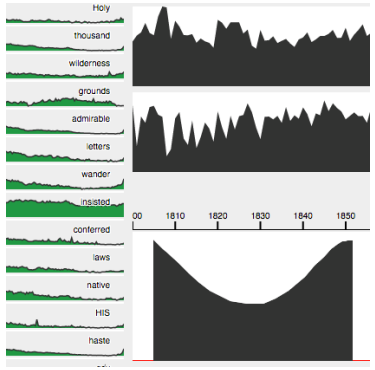
Serendip
Alexander et al, VAST 2014



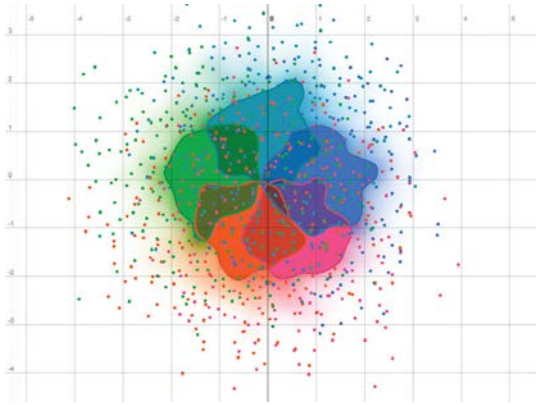
Explainers
Gleicher, VAST 2013



Topic Model Comparison
Alexander et al, VAST 2015



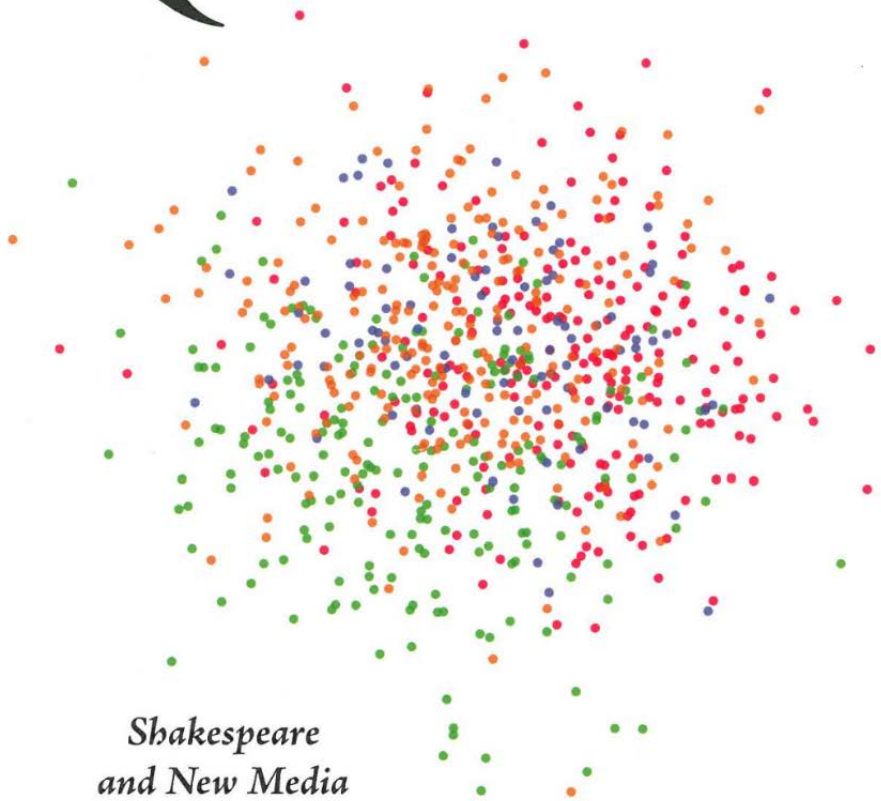
Sketch-based search
Correll et al, VAST 2016



Splatterplots
Mayorga & Gleicher 2013
Sarikaya & Gleicher 2015

And others…

SHAKESPEARE QUARTERLY



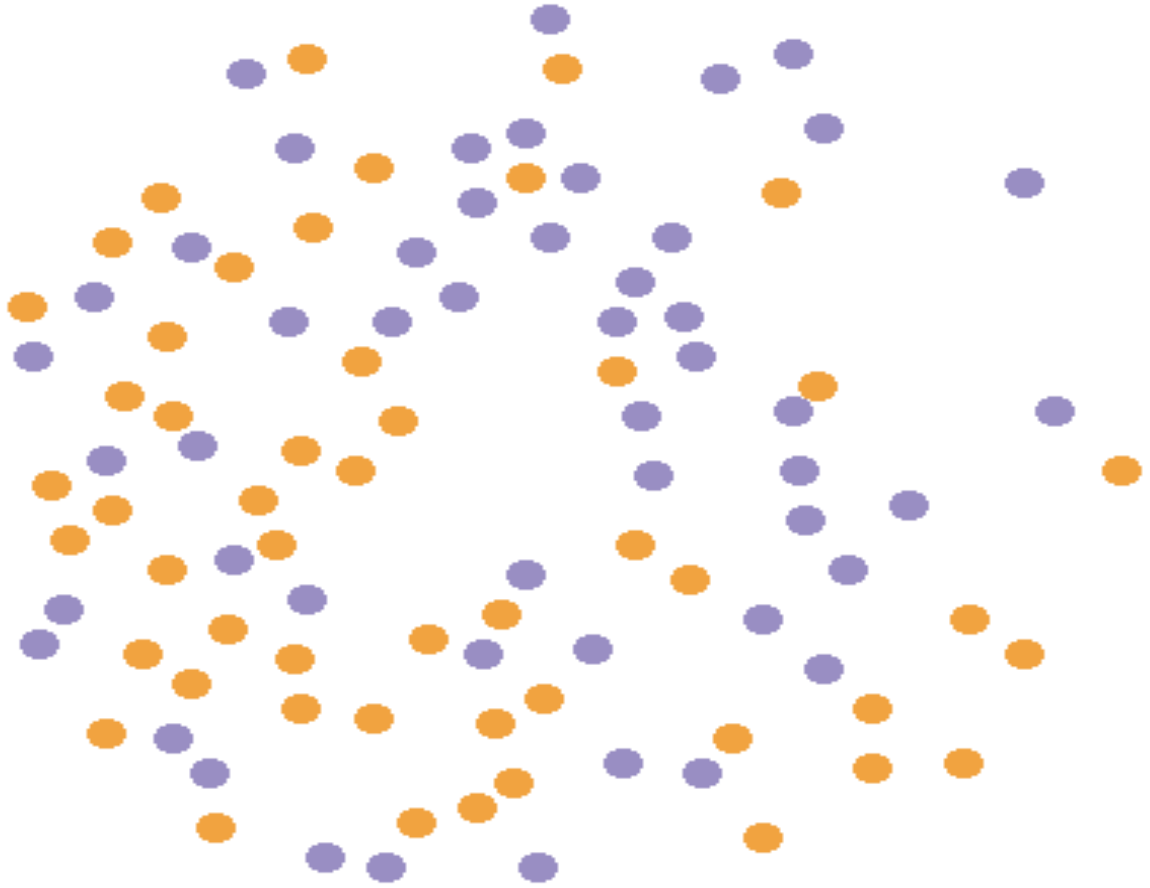
Shakespeare and New Media

Published for the Folger Shakespeare Library
in association with
The George Washington University
by The Johns Hopkins University Press

Volume 61 Fall 2010 Number 3

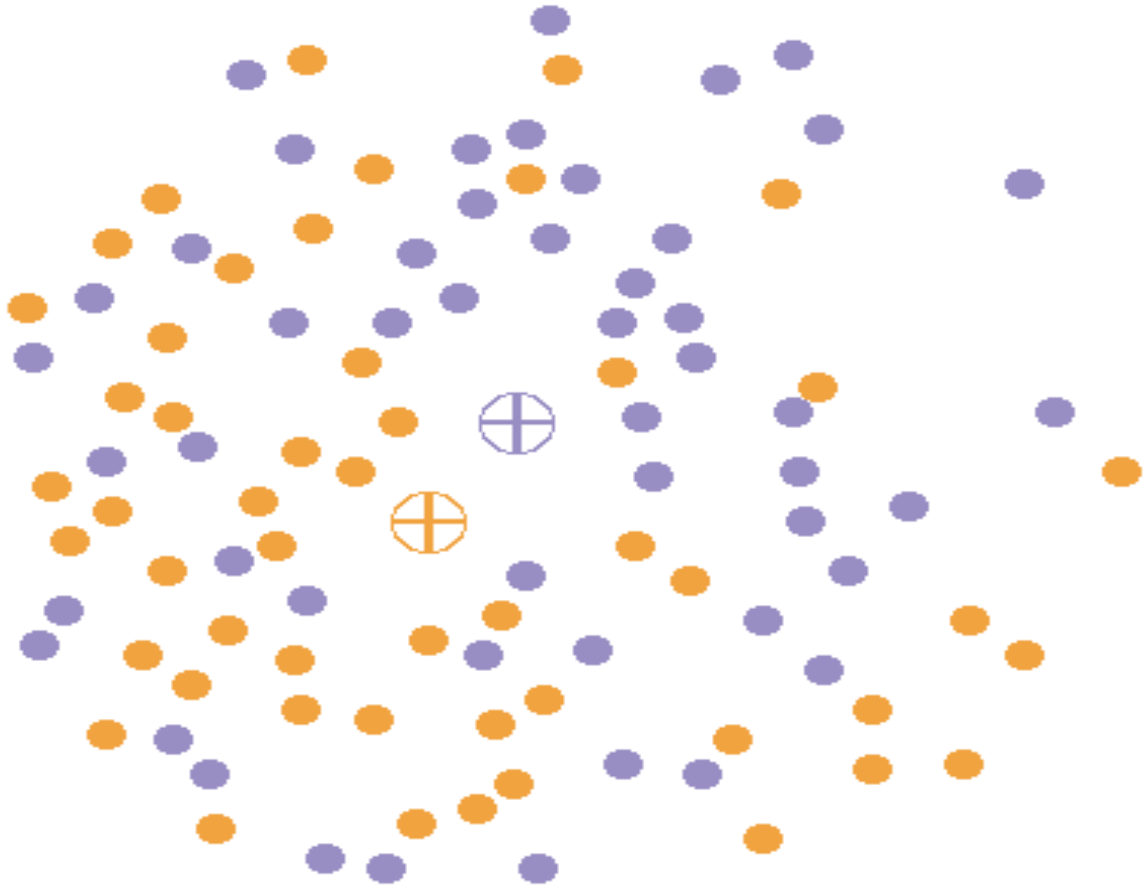
What can you do with too many points?

Which Color Point is Higher on Average?



Gleicher, M., Correll, M., Nothelfer, C. and Franconeri, C. “Perception of Average Value in Multiclass Scatterplots.” InfoVis 2013

How did you do that?



Visual Aggregation

People can extract summary statistics

Which Ones?
Efficiently?
Accurately?
How?

What can we do with it?
Why should we use it?

Visual Aggregation

Empirical Understanding

Averages in Time Series

Correll, et al. CHI 2012

Tagged Text

Correll, et al. CHI 2013

Scatterplot Averages

Gleicher, et al. InfoVis 2013

Other statistics in Time Series

Albers, et al. CHI 2014

Practical Application

Sequence Surveyor (Genetics)

Albers, et al. InfoVis 2011

LayerCake (Virus mutations)

Correll, et al. BioVis 2011

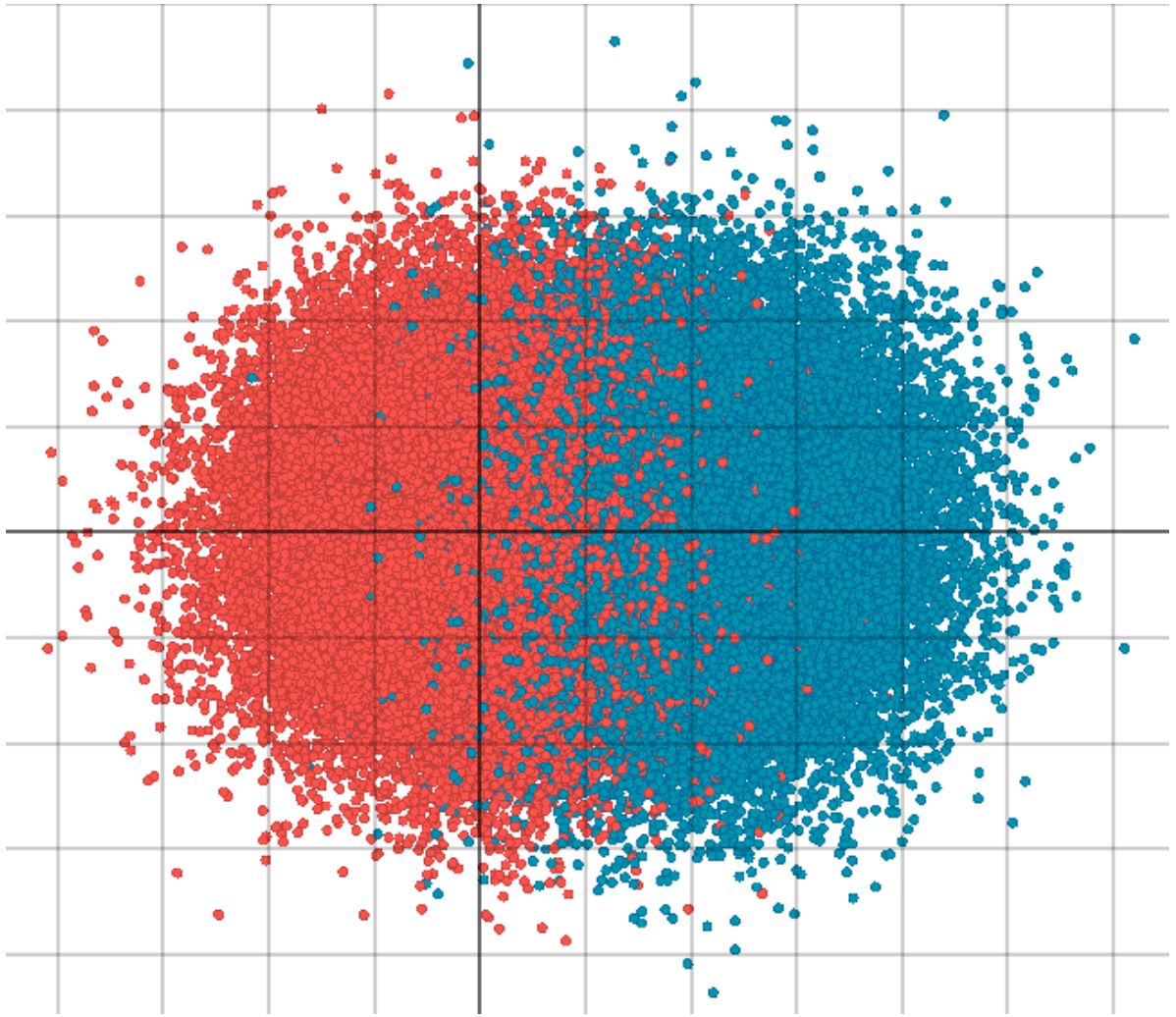
Molecular Surface Experiments

Sarikaya, et al. EuroVis 2014

Decision Making

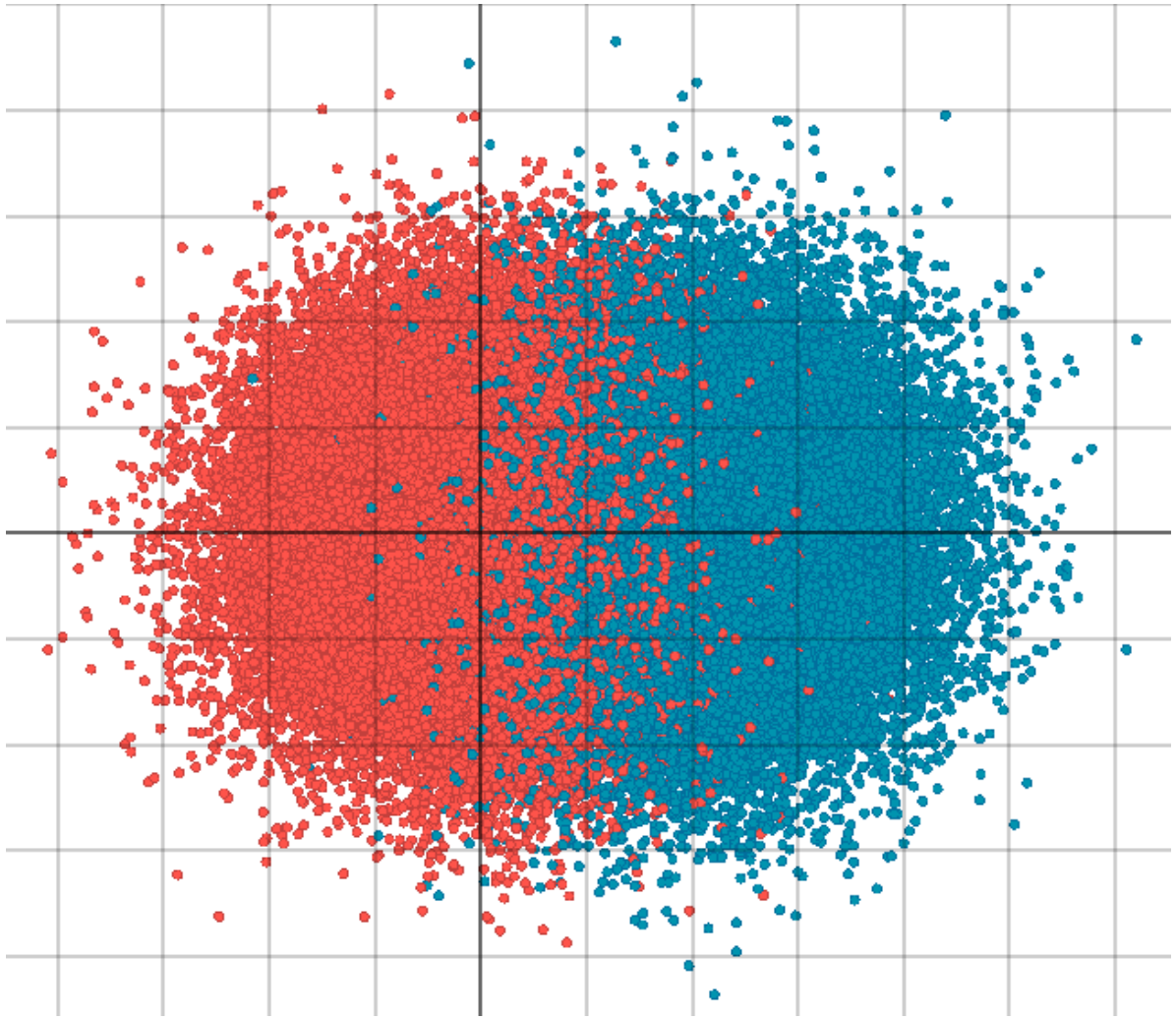
Correll, et al. (submitted)

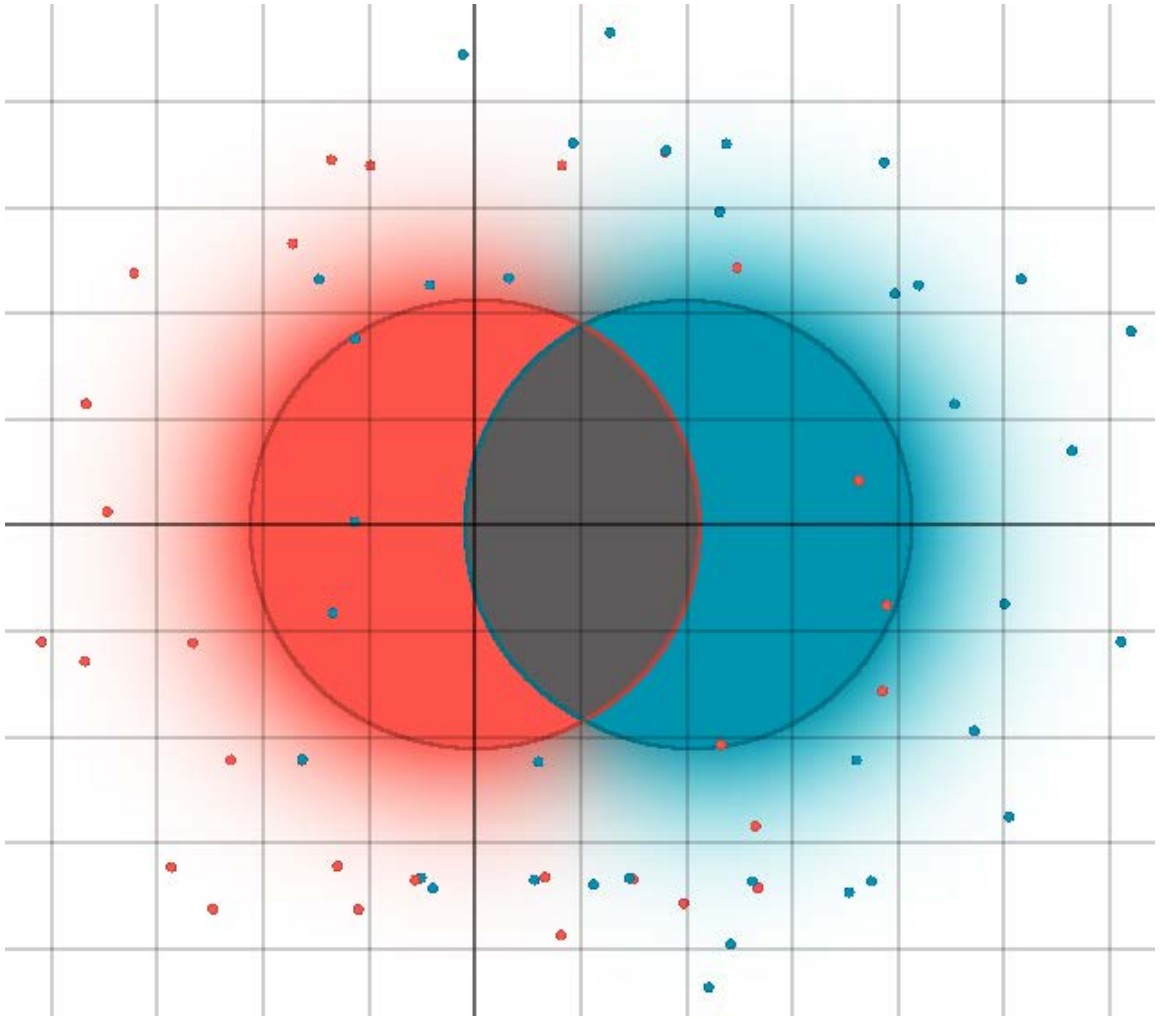
Problem: Scatterplot with way too many points!



Mayorga and Gleicher. Splatterplots : Overcoming Overdraw in Scatter Plots. IEEE Trans Visualization and CG, 19:9, 2013

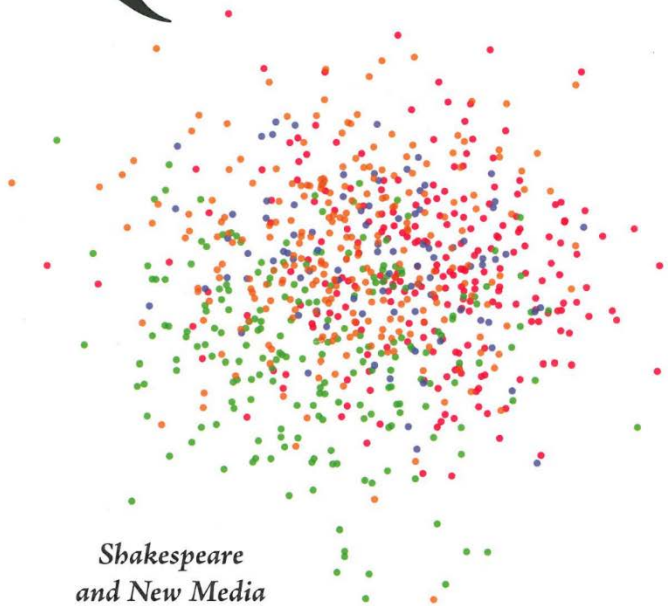
Solution: Splatterplot





Mayorga and Gleicher. Splatterplots : Overcoming Overdraw in Scatter Plots. IEEE Trans Visualization and CG, 19:9, 2013

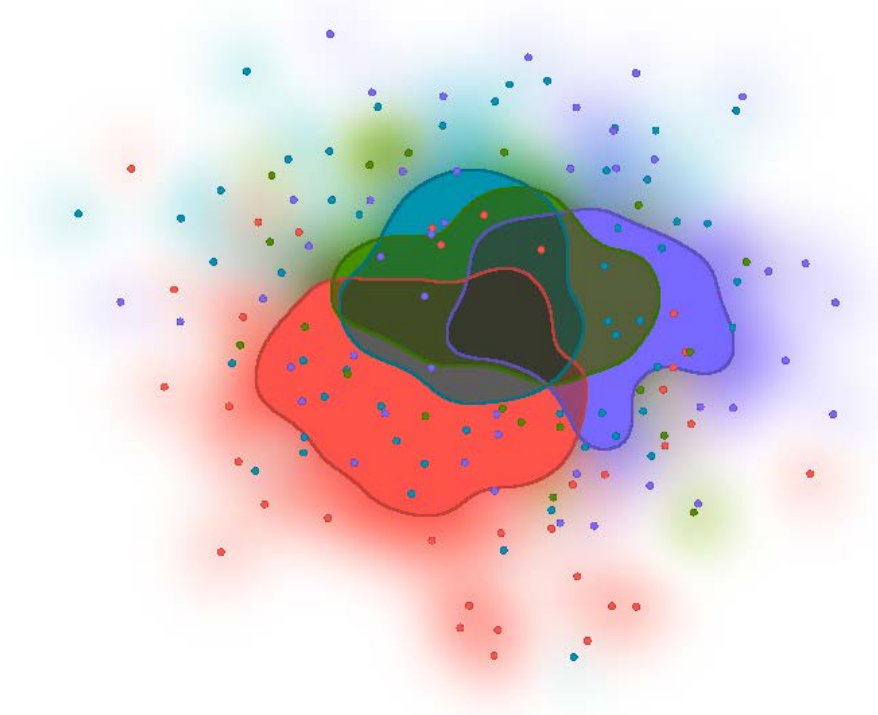
SHAKESPEARE QUARTERLY



Shakespeare and New Media

Published for the Folger Shakespeare Library
in association with
The George Washington University
by The Johns Hopkins University Press

Volume 61 Fall 2010 Number 3

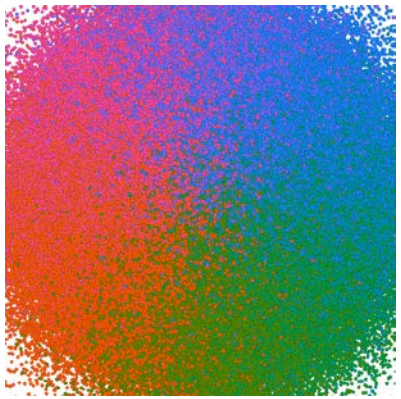


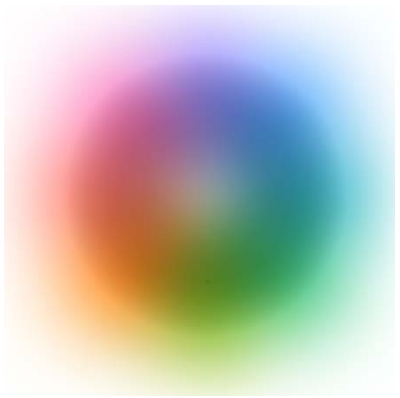
This is pretty!

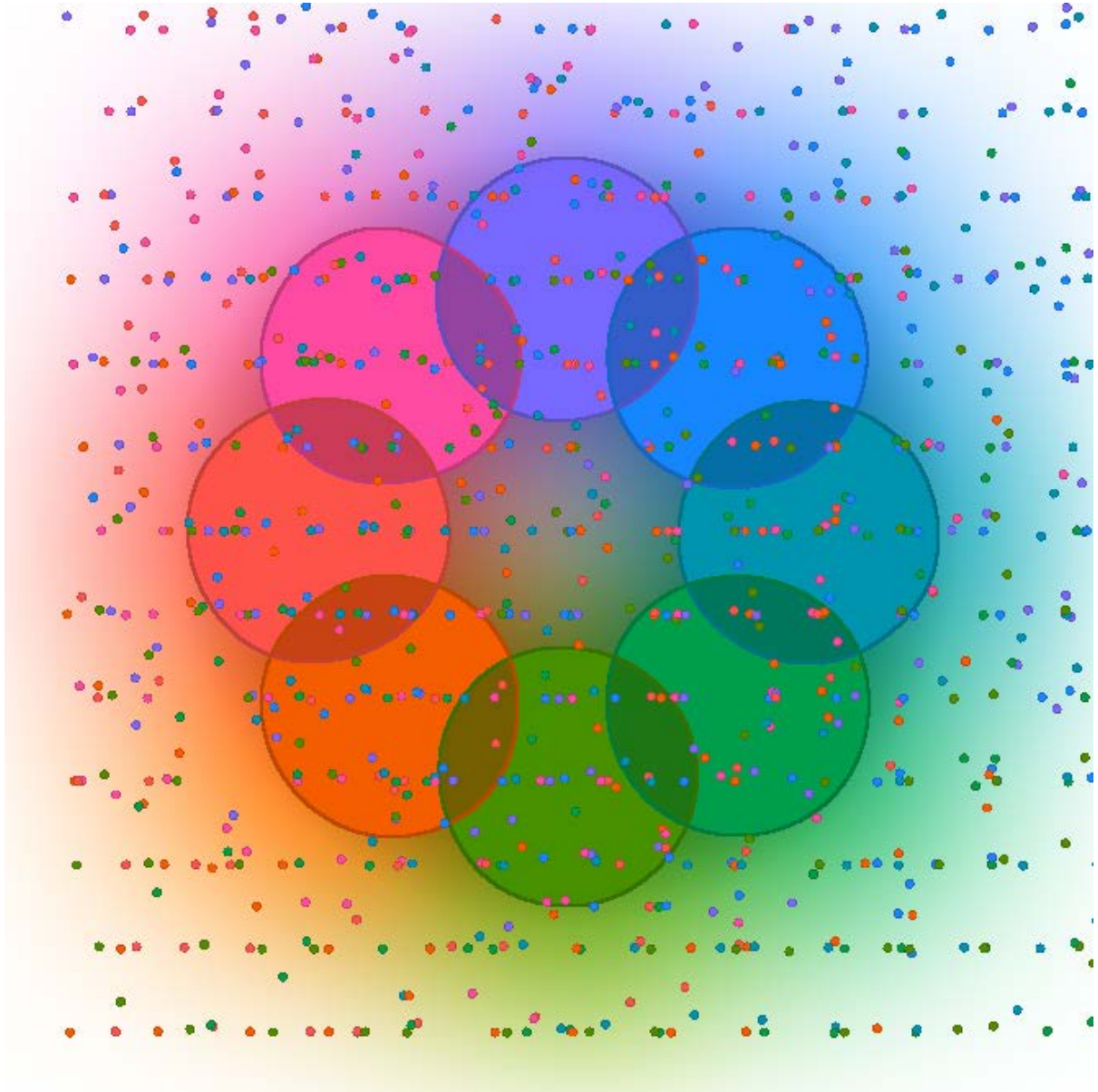
But when should we use it?

Lots of choices

Need actionable advice







Scatterplots are common

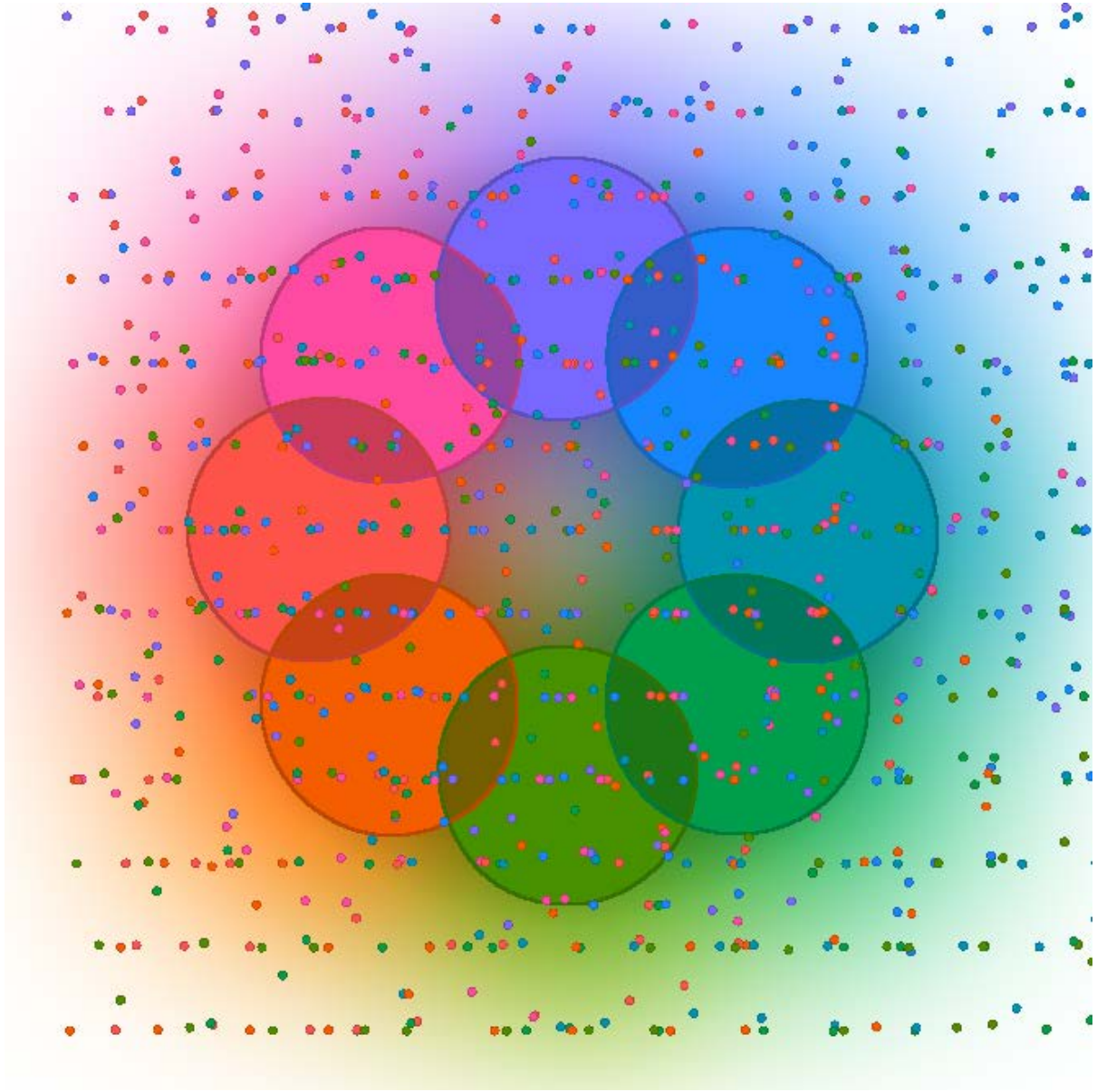
Lots of designs

How do we choose?

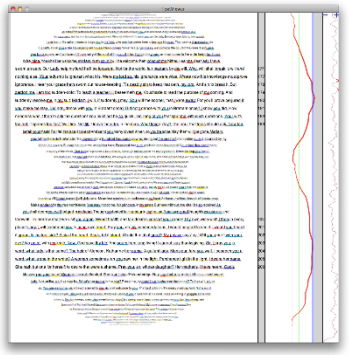
Task oriented analysis

Ask Dr. Scatterplot!

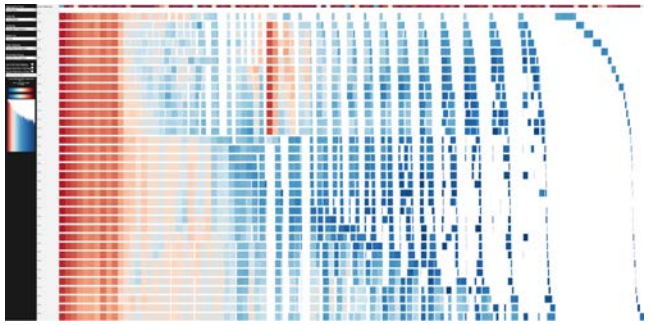
Sarikaya and Gleicher. Scatterplots: Tasks, Data, and Designs. IEEE Trans Vis and CG, 2018.



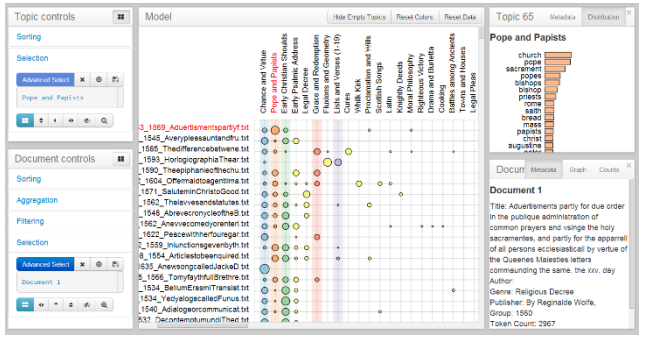
VEP Vis success stories (some of each kind)



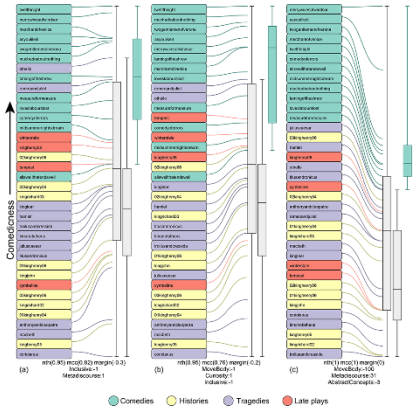
Text Viewer
Correll et al, EuroVis 2011



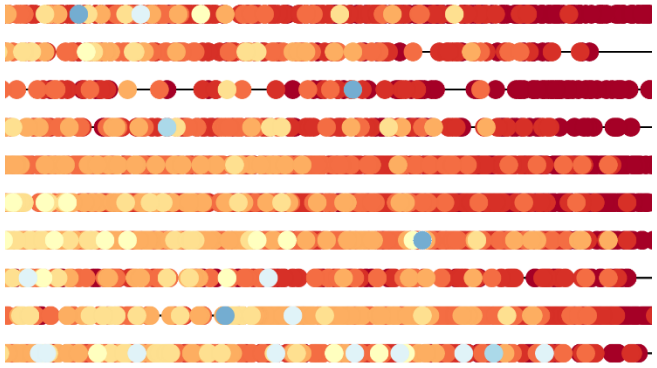
TextDNA
Szafir et al, EuroVis 2016



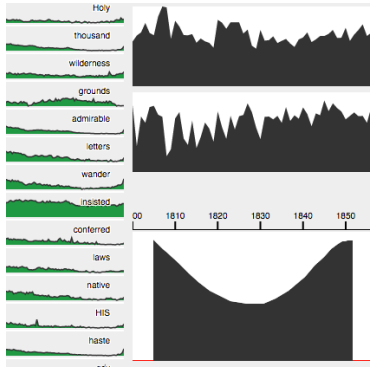
Serendip
Alexander et al, VAST 2014



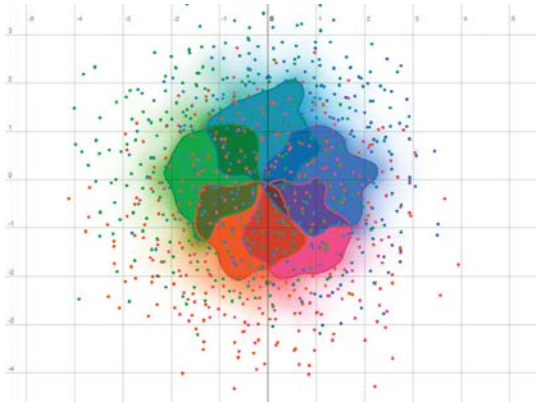
Explainers
Gleicher, VAST 2013



Topic Model Comparison
Alexander et al, VAST 2015



Sketch-based search
Correll et al, VAST 2016



Splatterplots
Mayorga & Gleicher 2013
Sarikaya & Gleicher 2015

And others…

But are there bigger lessons?

I learned a lot from this

Three interesting aspects of VEP

Literary scholarship – and ways of thinking

Literary scholars – with their needs and abilities

300 years of old books – variety and quality

300+ years of books

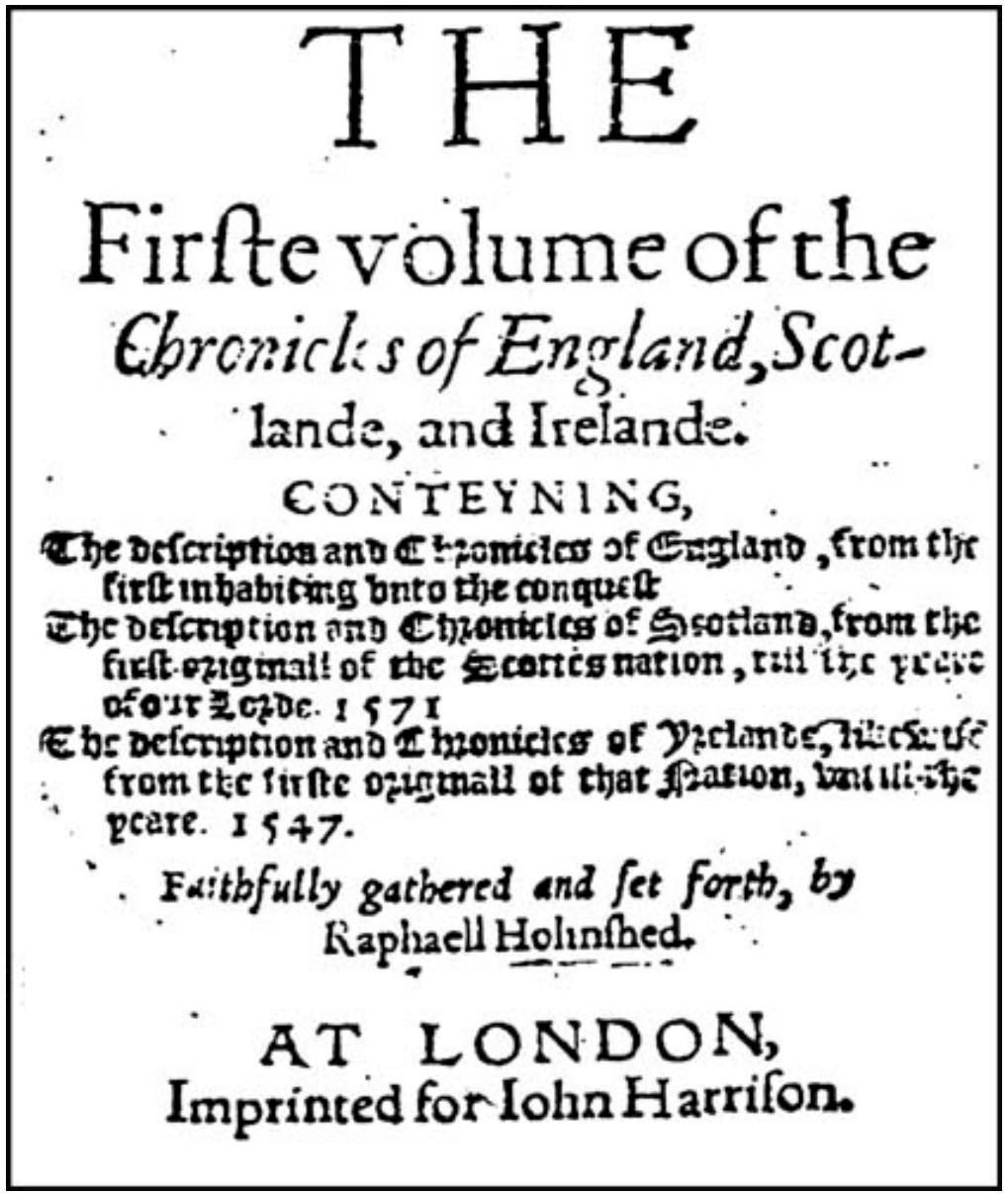
A period of intense changes:
cultural, historical, intellectual, political, …

How does the printed record reflect this?
The language changed a lot
Printing changed a lot
The book business changed a lot
Libraries changed a lot

300+ years of books

Every known surviving book?

Huge historical efforts
- to catalog known books
- to microfilm known books
- to transcribe known books



THE

Firſte volume of the

Chronicles of England, Scot-

lande, and Irelande.

CONTEYNING,

The deſcription and Chronicles of England, from the firſt inhabiting vnto the conqueſt

The deſcription and Chronicles of Scotland, from the firſt originall of the Scottes nation, till the yeare of our Lorde. 1571

The deſcription and Chronicles of Yrelande, likewiſe from the firſte originall of that Nation, vntill the yeare. 1547.

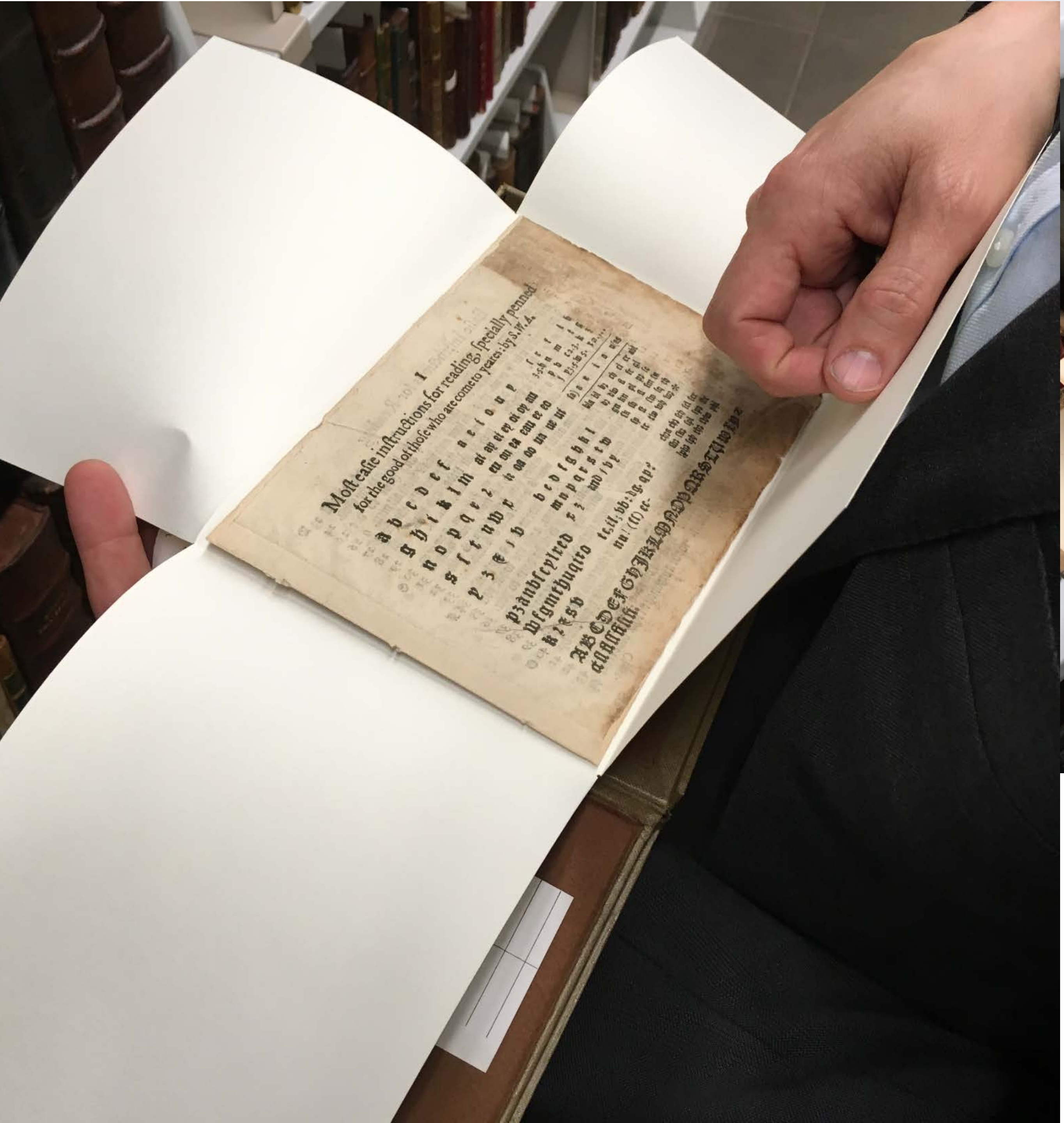
Faithfully gathered and ſet forth, by

Raphaell Holinſhed.

AT LONDON,

Imprinted for Iohn Harriſon.

Text Creation Partnership (**TCP**)
- Hand-keyed about 60,000 books into SGML/XML



1

Most easie instructions for reading, specially penned for the good of those who are come to yeares: by S.W.A.

14 A Discourse concerning

manner, answering, with detestation ...
rour, Will you speake wickedly for God ...
ly for him? will you accept his person ...
tend for God? do you so mocke him? ...
lencie make you affraid? What ...
more emphaticall, or yet more ...
ten, It is impossible for God to lye ...
must we thinke it for God to approue ...
lie, how much lesse wilfull lying? ...
they to forfeit, who stampe a fained ...
were a false face vpon a Miracle ...
nipotencie, which should therefore ...
on it Truth, the first attribute of the God of ...
the perfect Image of the God of ...
that it be a true Miracle?

Iob 13.7.

Heb. 6.18.

Popish Exorcizing. 15

... they) maketh the words, I am no Priest, (otherwise ... in themselues) to be true, as it were one Pro-position. And thus they thinke to auoyd the guilt ... a lie.

This Sophistrie did Master Garnet teach in wri-ting, and practized it egregiously in the Tower: and this Master Parsons did vehemently defend, which other Priests vsed as a Gyges Ring, to become inuisible to men; that no Examiner could espie any of their Priest-hoods.

If they thinke it lawfull thus cunningly to equiuocate, for the preseruation of their owne bodies, and the bodies of others, will they not iudge ... necessary, for the gaining of mens soules, trowe ... ? What credit then shall wee giue vnto such re-ports, which they shall falsly inuent, to draw men vnto their Religion, by saying (as they haue most falsly done) such a Protestant Bishop died in the Ro-mish faith; or such a Boy was exorcized and dispos-sessed of the Deuill, by the helpe of a Romish Priest? Or such a vision of our Lady appeared vnto me, tel-ling me that I cannot take the Oath of Allegeance without mortall sinne? To these men we obiect, Are you so vnconscionable as to lye thus? They readily will answere vs, we lye not, for we did equiuocate. As for example, He died a Catholike, (reseruing) for ... that you shall know: and the Boy was dispossessed ... a Priest, (reseruing) if dispossessed: And the blessed Virgin appeared vnto me, (reseruing) but only in my conceit. This being the professed guile and deceit of these kind of men, we haue iust cause not to be-leeue them.

The

The sixt Aduertisement.

Discouering their doctrine of Mentall Reseruation, which we may iustly suspect in the relating their deceits, whensoeuer they ... their false reports,

THeir Mentall Reseruation is (as they ...) a mixt Proposition, part in speech ... minde: as when a Priest, being in ... be discouered, is examined whether ... mish Priest? who notwithstanding ... I am no Romish Priest, (retaining in his ...) clause) To tell it vnto you; (or) that ... or the like: and the latter clause of the ...

Sources of variation

Variety of topics

Authors' thoughts and ideas

Authors' words

Printers' practicality

Librarians' taking care of books

Transcribers translating books

Modern processing tools handling things correctly

It’s naïve to spell naïve with an i

Did the author use an ï ?

Did the printer have it?

Did the dots get smudged?

Did the microfilming make it look like a dust mark?

Did the transcriber see it?

Did they encode it weirdly that year?

Did our Unicode pipeline mess it up?

Did our spelling standardizer get it wrong?

One form of variation can obscure another

But…

One person's noise is another's data

spelling obscures ideas
or
observe the development of spelling

rotting pages obscure content
or
preserved books tell us what was valued

Data Wrangling

Make data convenient for analysis

Clean away “unwanted” variance

Leave enough signal

Hard choices!

Data Wrangling Lessons

Everyone knows Data Wrangling is a big deal, time consuming, …

Getting good meta-data requires investment

Transparency of data wrangling is valuable

Comprehensibility – the stakeholders need to understand

Curation and editing is part of scholarship.
These are decisions – make them wisely!

A view of Data Wrangling (or cleanup)

It's about variation:

variation is what we're interested in

in order to expose the variation you care about

you need to clean away the variation you don't

Curation and editing are part of scholarship.
These are decisions – make them wisely!

A case study...

Counting “the”

It is interesting at scale

Look for correlations across huge amounts of literature!

It is transparent

we can explain – and check – every step

It’s still hard to get right…

Just because counting is easy – doesn’t mean anything else is

Some literature theory...

Writing about things in the world (Extra-Subjective)

vs.

Writing about abstract things (Intra-Subjective)

How does this play out over 300 years, 60,000 documents, ... ?

How does this play out over plays?

about 1500 documents- that we have lots of meta-data on

Extra-Subjective vs. Intra-Subjective

Is there an easy way to measure it?

Statistical analysis:

The primary variance in the collection is correlated with labels (**science**, **plays**)

The main source of this variance is the most common word

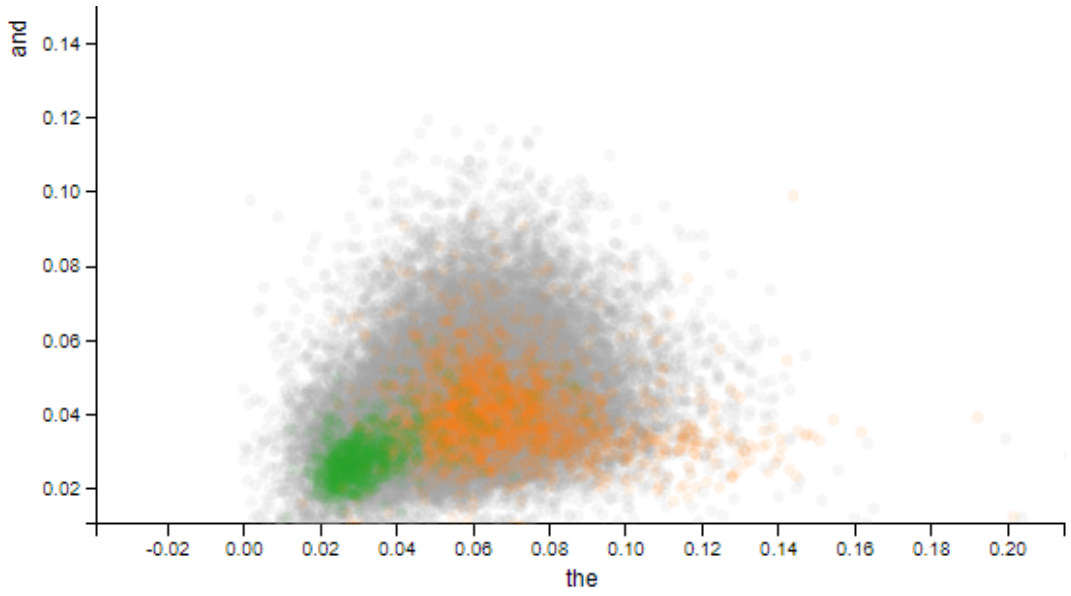
Count “the” ?

Use definite articles to refer to objects in the world

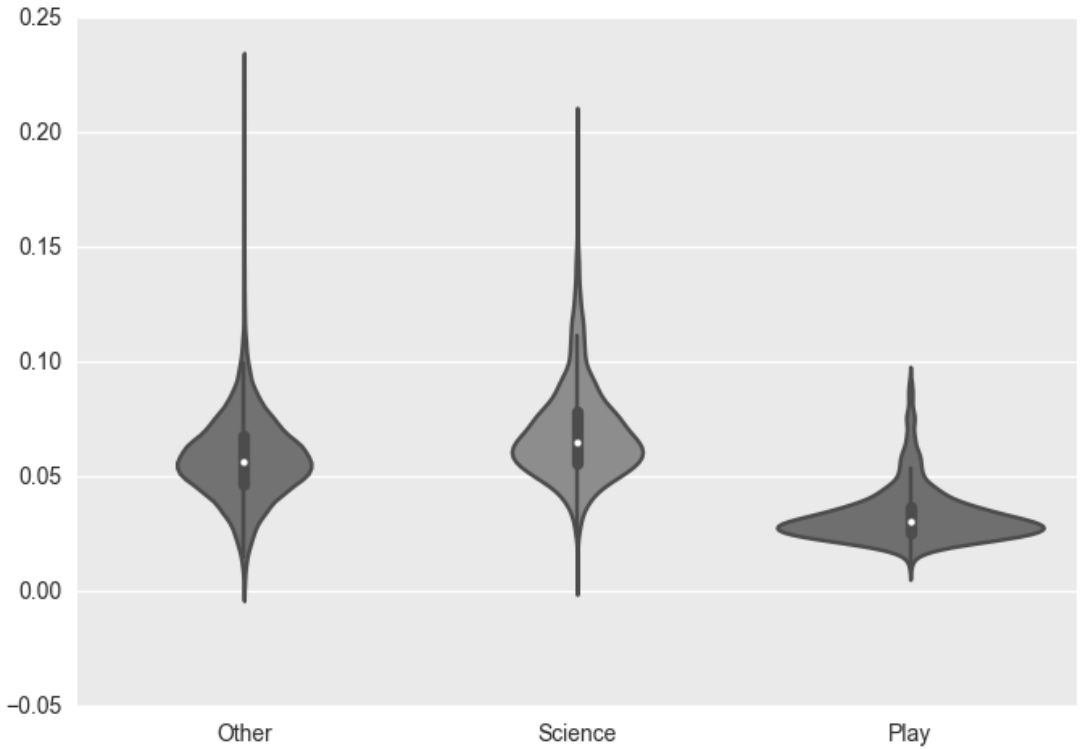
Science uses it more than **Plays**

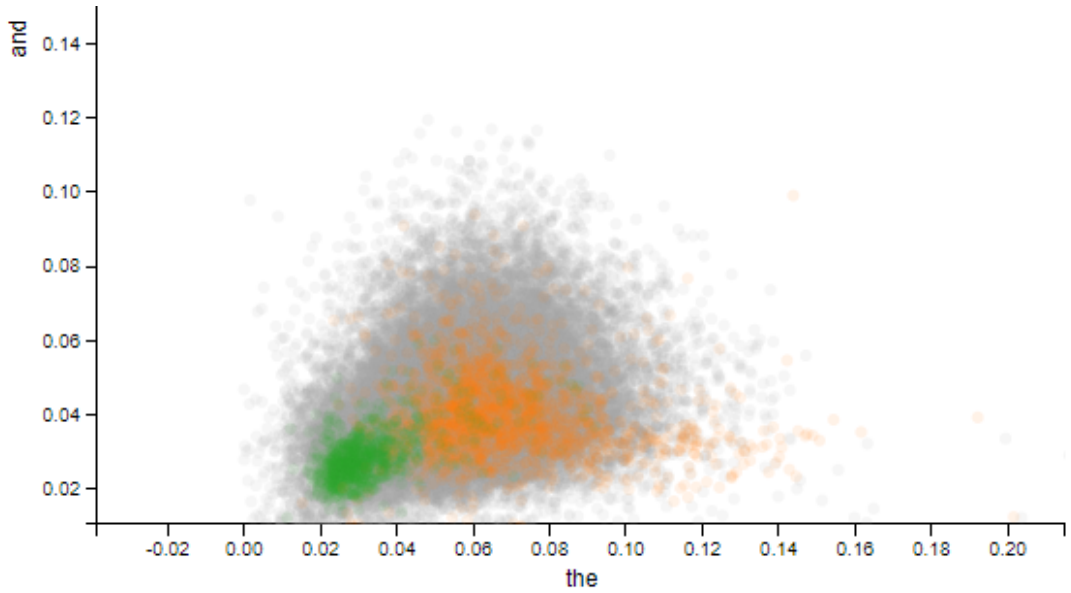
We only have (extensive) labels for **science** and **plays**

Most things are unlabeled



Count “the” ?


0.25
0.20
0.15
0.10
0.05
0.00
−0.05
Other
Science
Play


and
0.14
0.12
0.10
0.08
0.06
0.04
0.02
-0.02
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16
0.18
0.20
the

23% “the” – really?

A plain and easy rule to rigge any ship by the length of his masts, and yards, without any further trouble

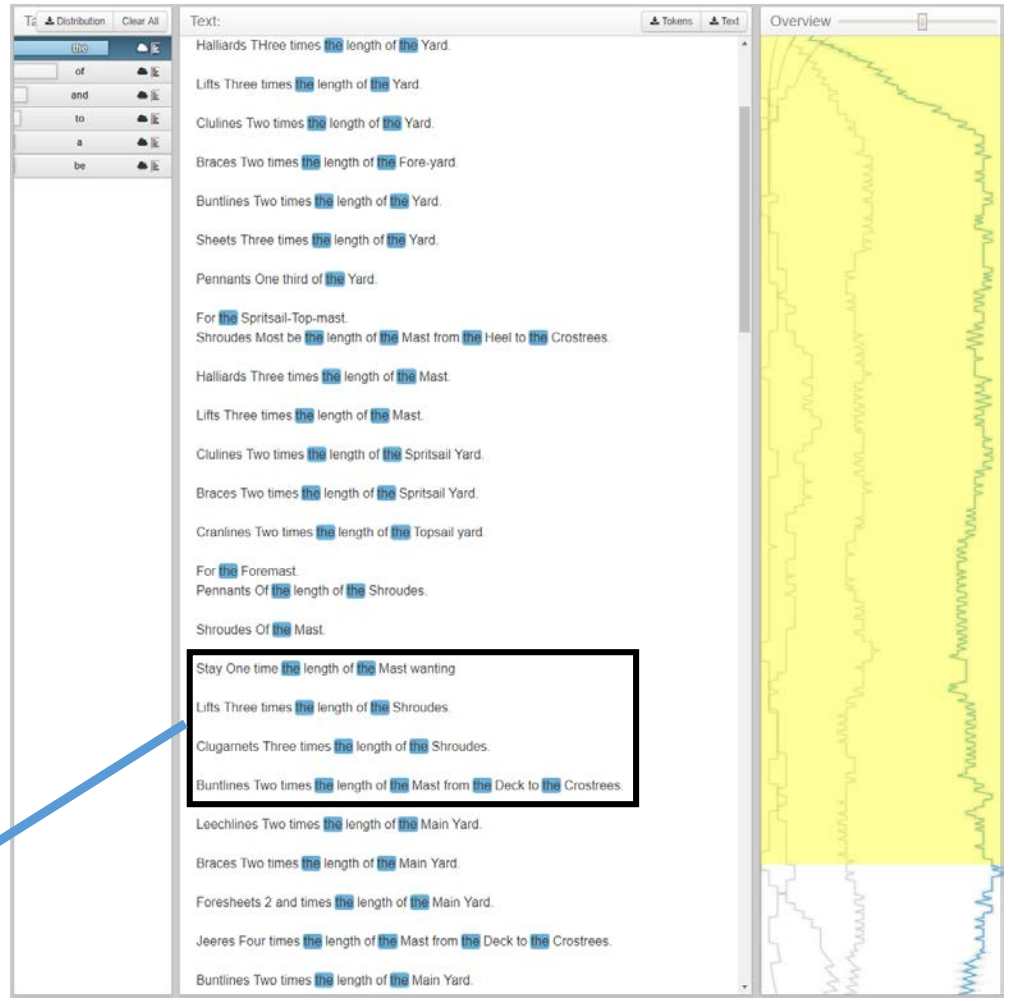
23% “the” ?

Stay One time the length of the Mast wanting

Lifts Three times the length of the Shroudes.

Clugarnets Three times the length of the Shroudes.

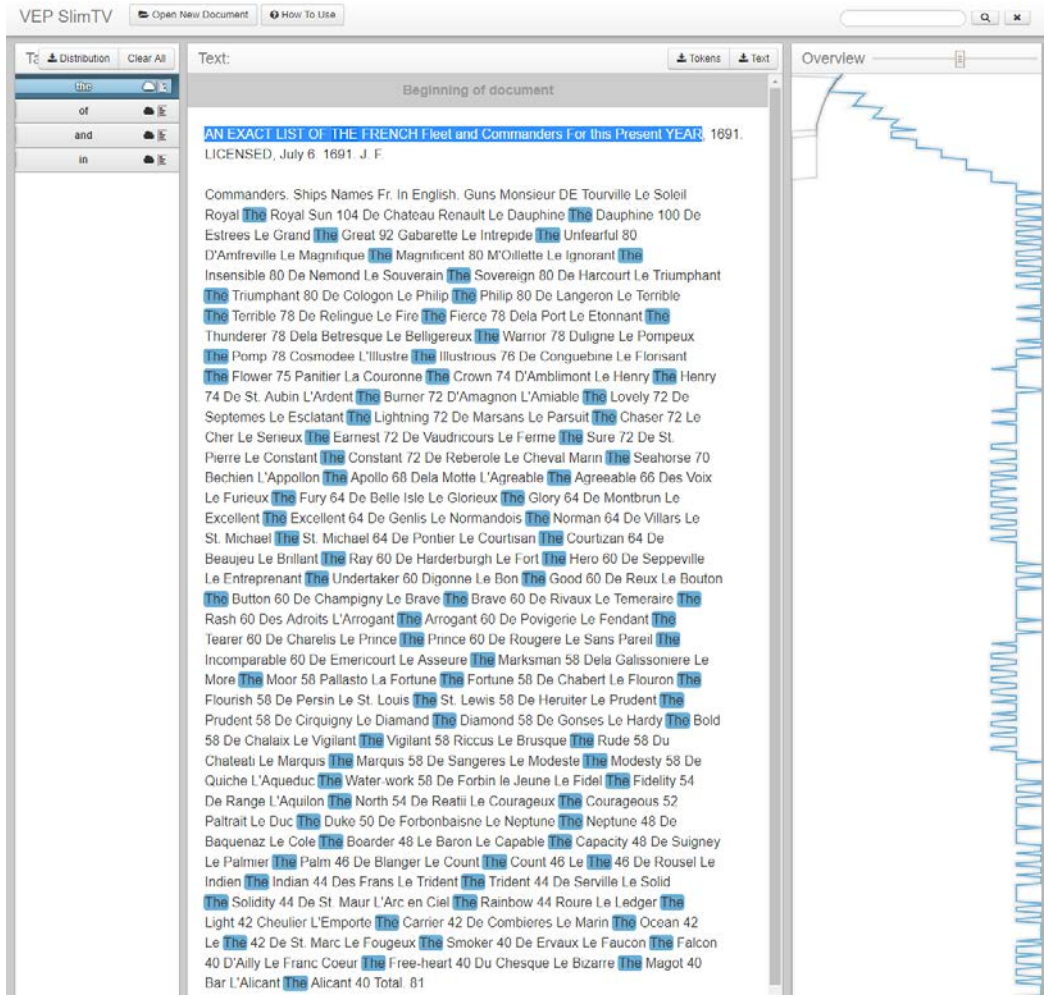
Buntlines Two times the length of the Mast from the Deck to the Crostrees.



51% “the” if you count differently

AN EXACT LIST OF THE FRENCH Fleet and Commanders For this Present YEAR

51% of “words” – only counting things we recognize as words



Why is this hard?

We need to carefully treat each book in a uniform way

Which words do we extract off the pages?

We need to carefully treat each word in a uniform way

How to deal with bad characters and spelling variation?

We need to analyze in a uniform and transparent way

We need good meta-data

The ***scholars*** (and their audience) must understand each step!

We need to get **back to the sources**!

Big lesson:

Data Science is a process
Challenges are everywhere!

Big lesson:

We must help all users understand all phases of the process

Big lesson:

We must help all users understand all phases of the process

Towards Comprehensibility in Modeling

The other title for this talk

Data Science is a Process

There are lots of steps

Many provide challenges

Many provide opportunities

Identify Problem

Gather Data

Clean Data

Abstract

Design

Build

Evaluate/Validate

Interpret

Act

Disseminate

An Aside...

Can we automate this process?
Make it possible for subject experts?

Data Driven Discovery of Models

Smart systems to do steps

Smart user experience to guide users

Some steps seem really hard

Identify Problem
Gather Data
Clean Data
Abstract
Design
Build
Evaluate/Validate
Interpret
Act
Disseminate

d3mapp

https://d3m-app.firebaseapp.com

Apps | sitekey | clientId US | Boards and Beyond... | clientId Updated | INJCT | ANN | Rails | React | Writing | Design | Business | Recipes | Adventure | Go | Other Bookmarks

- Approve Problems
- Wait for Results
- Select Metric
  Selected: specificity
- Select Model
  Default: Model 1
  - Examine model results in detail
  - Compare models

Model 1

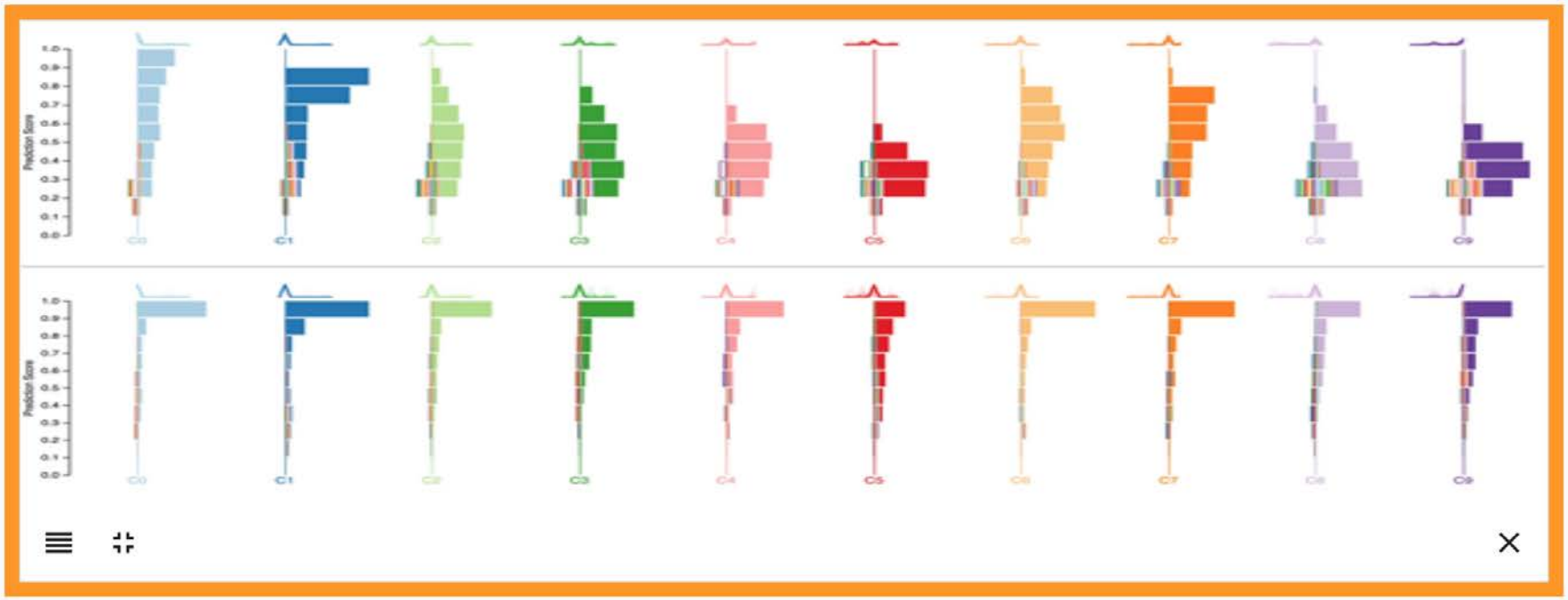
SELECT MODEL

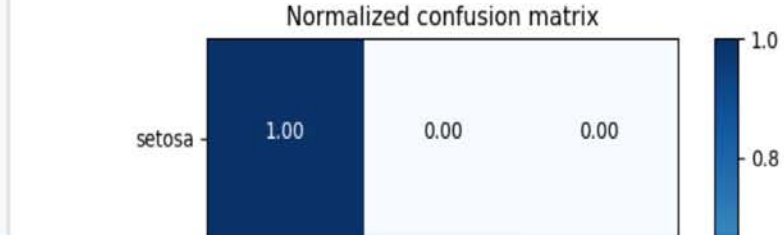
5 Create Executable

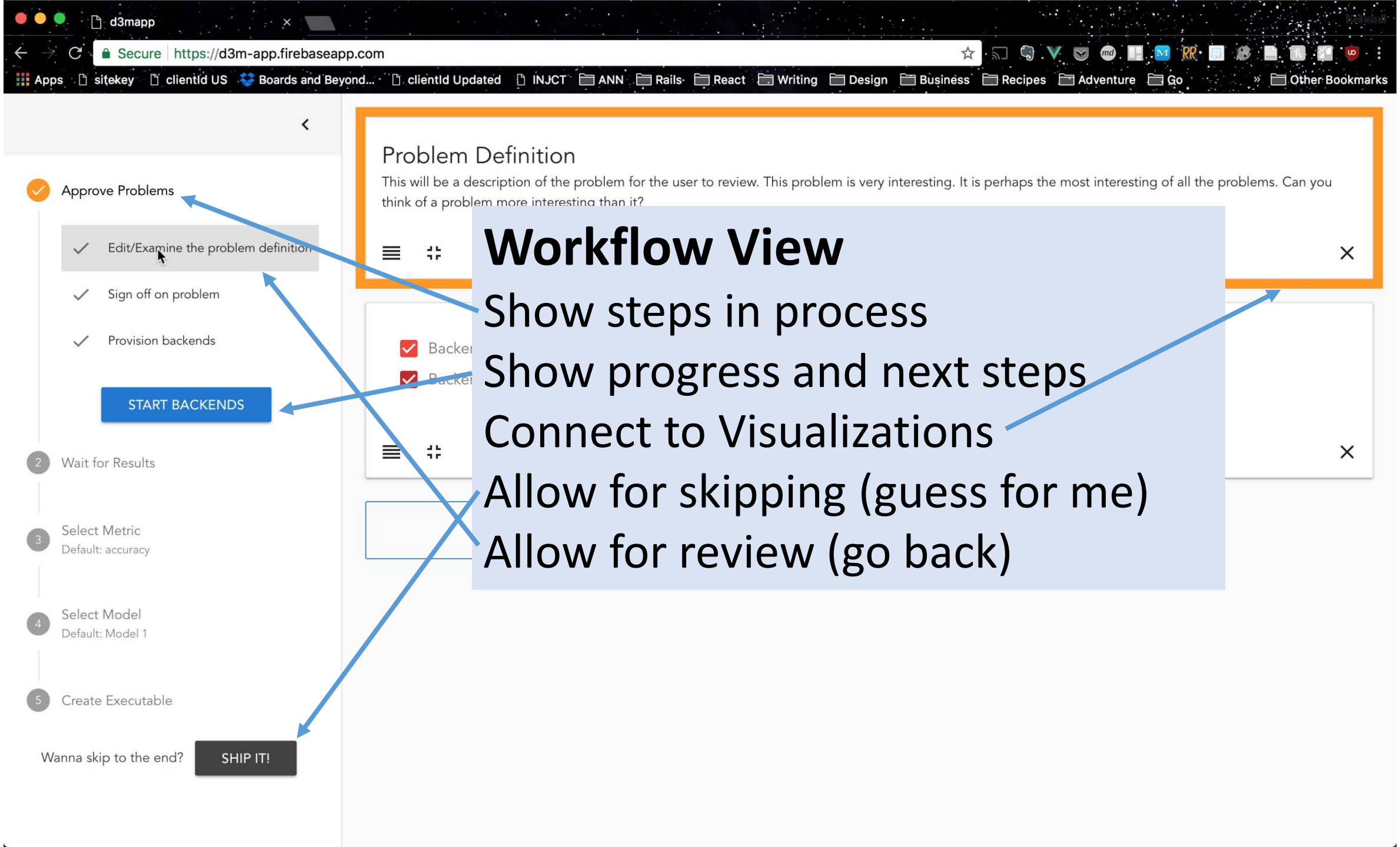
Wanna skip to the end? SHIP IT!

Problem Definition

This will be a description of the problem for the user to review. This problem is very interesting. It is perhaps the most interesting of all the problems. Can you think of a problem more interesting than it?






d3mapp
Secure https://d3m-app.firebaseapp.com
Apps sitekey clientId US Boards and Beyond... clientId Updated INJCT ANN Rails React Writing Design Business Recipes Adventure Go Other Bookmarks
Approve Problems
Edit/Examine the problem definition
Sign off on problem
Provision backends
START BACKENDS
2 Wait for Results
3 Select Metric
Default: accuracy
4 Select Model
Default: Model 1
5 Create Executable
Wanna skip to the end?
SHIP IT!
Problem Definition
This will be a description of the problem for the user to review. This problem is very interesting. It is perhaps the most interesting of all the problems. Can you think of a problem more interesting than it?
Workflow View
Show steps in process
Show progress and next steps
Connect to Visualizations
Allow for skipping (guess for me)
Allow for review (go back)

Comprehensibility across the Process

Who?

Stakeholders

Developers

Data Scientists

Domain Experts

Audience

Subjects

Why?

Reason for Wanting

Improve Performance

Build Theory

Extend/Characterize

Build Trust

Actionability

Convince

Where?

Phase of the **Process**

Identify Problem

Gather Data

Clean Data

Abstract

Design

Build

Evaluate

Interpret

Disseminate

Gleicher, M. A Framework for Comprehensibility in Modeling.
Big Data, January 2016.

Who? Why? Where?
(for How? and What? See the paper)

Who?

Stakeholders

Developers

Data Scientists

Domain Experts

Audience

Subjects

Why?

Reason for Wanting

Improve Performance

Build Theory

Extend/Characterize

Build Trust

Actionability

Convince

Where?

Phase of the **Process**

Identify Problem

Gather Data

Clean Data

Abstract

Design

Build

Evaluate

Interpret

Disseminate

Gleicher, M. A Framework for Comprehensibility in Modeling.
Big Data, January 2016.

Some phases for comprehensibility

Data Wrangling

understand what data the data is and what it can do

Model Building

understand classifiers to build theory / identify items

Validation Experiments

understand experiments to identify items / build trust

Towards Comprehensibility in Modeling

This is not just comprehensible models

genre

Categorization given by Shakespeare's contemporaries

Comedy **Tragedy** **History**

Category for plays written after that

Late Plays

data

Count words

Count kinds of words – using dictionaries

- Docuscope (Igarashi and Kaufer)
- Lists of phrases for each “Language Action Type”
- Simple and Transparent dictionary matching
  - (once the dictionaries are made)

Fancier methods (Topic Models, Machine Learning, …)

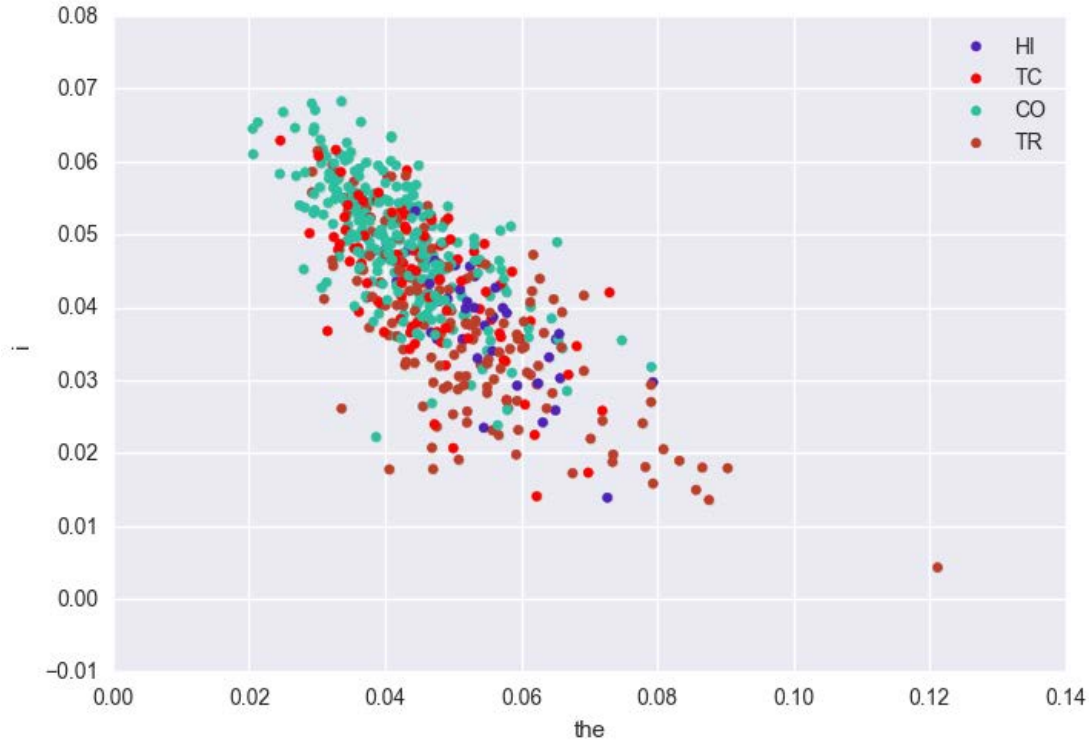
Is “the” good enough?

Histories and **Tragedies** are more external

Comedies are more internal

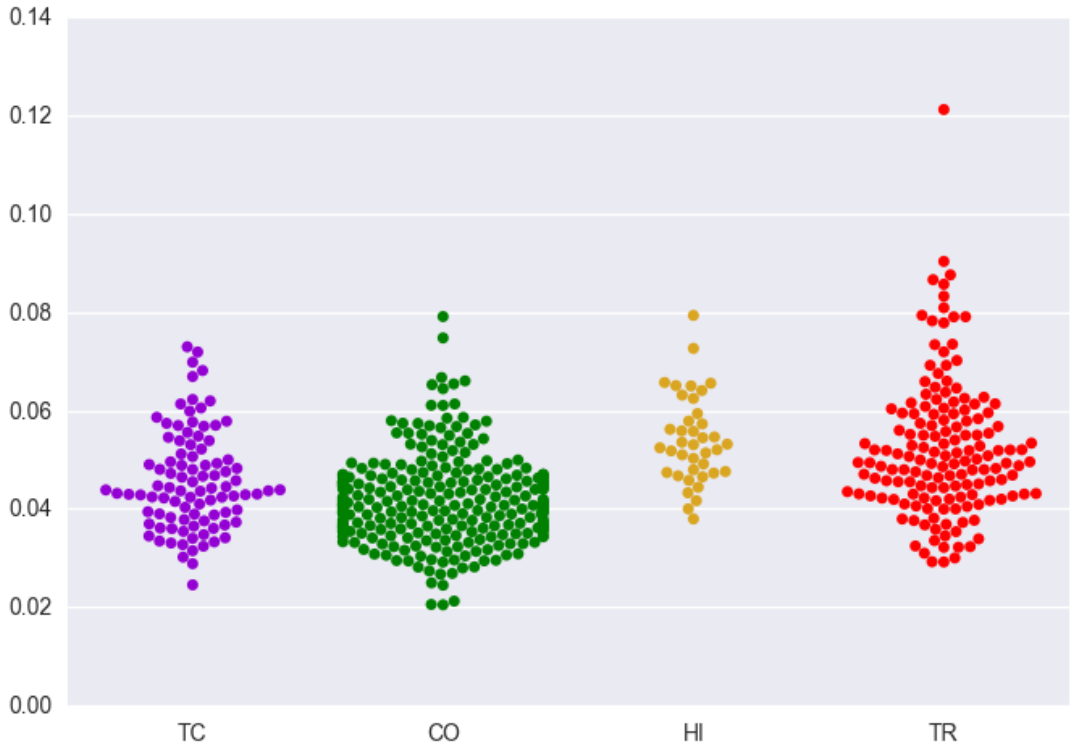
The second most common word, “I”, might be better (it is for internal)

600 plays of Shakespeare’s era

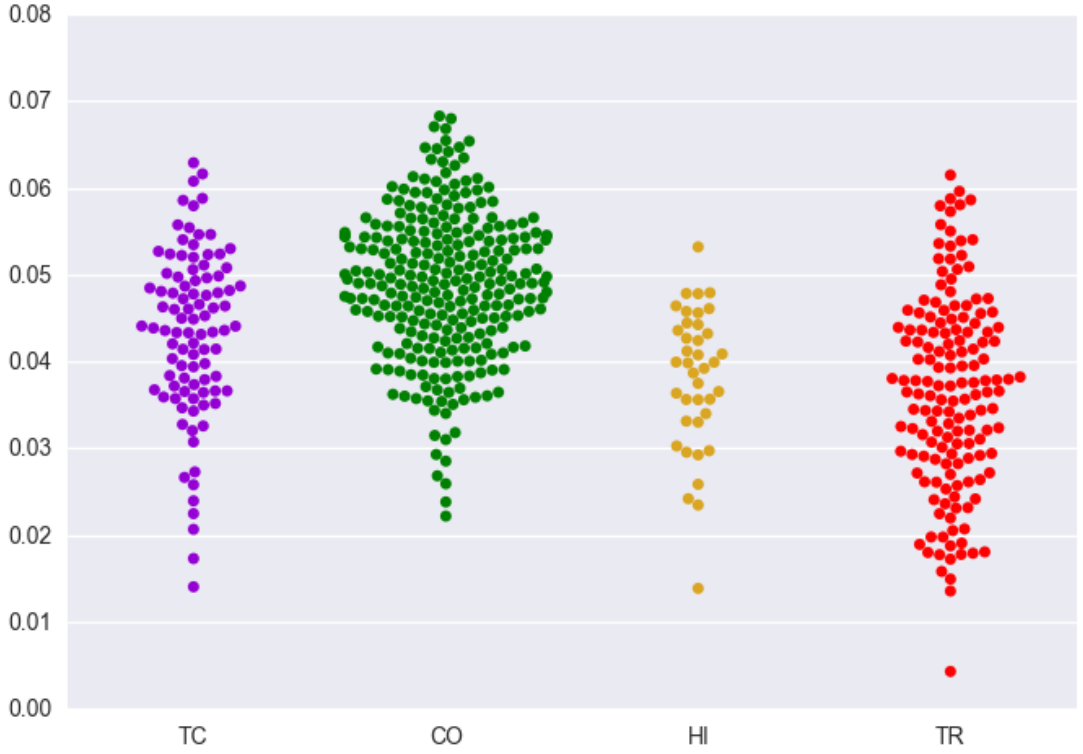


Neither is great, maybe try combining them?

The



I



Mathematical Models

“the” – “I” --- pretty good
(“the” + “to”) – (“I” + “a”) --- very good

How do we come up with such equations?
What do we do with them?

Machine Learning Classification:

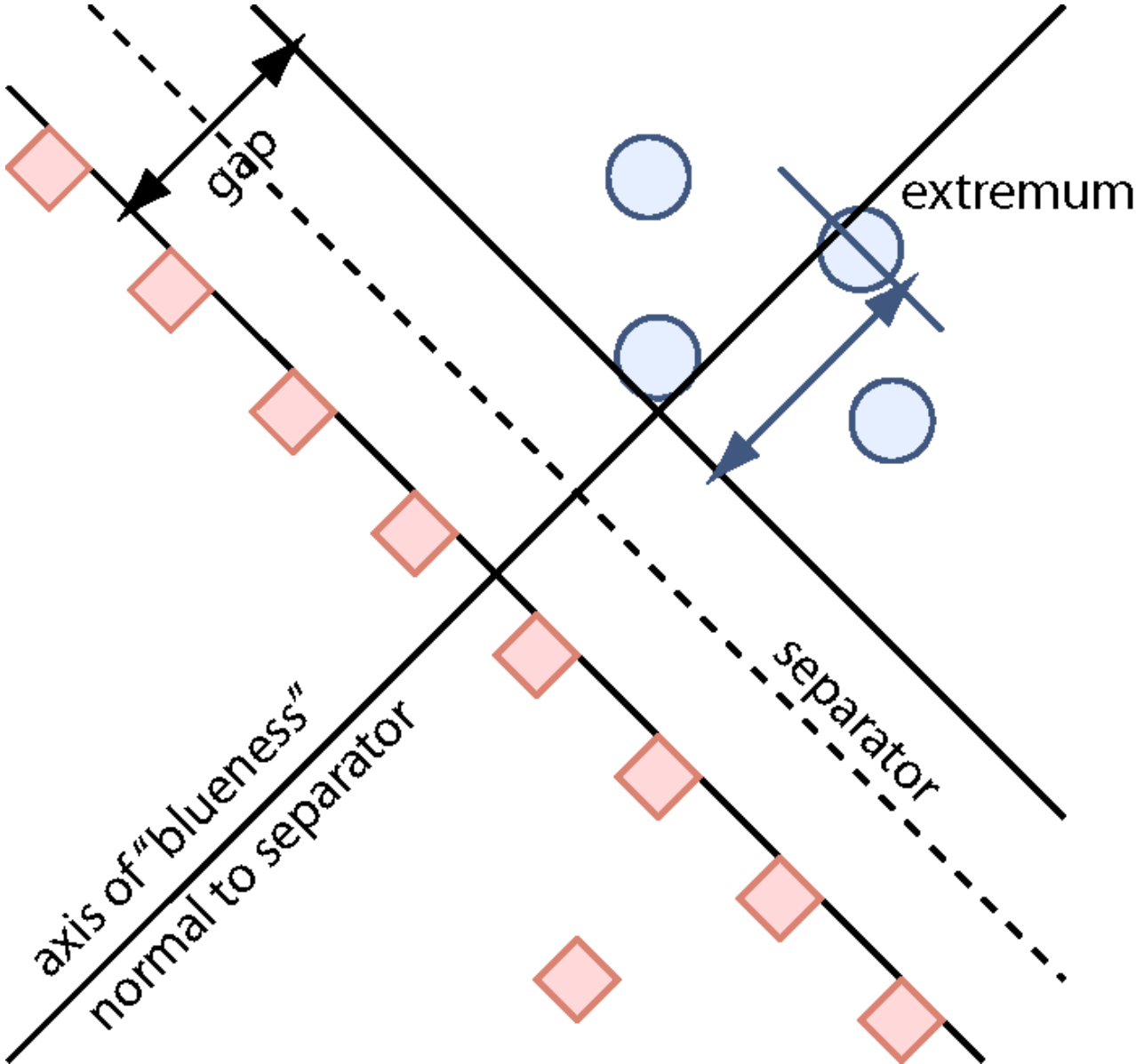
Make up these function so that one “class” is high
The other classes are low

Comprehensibility in modeling?

They may never understand
Support Vector Machines

(or choose your favorite buzzword)

But they might understand
what it does…



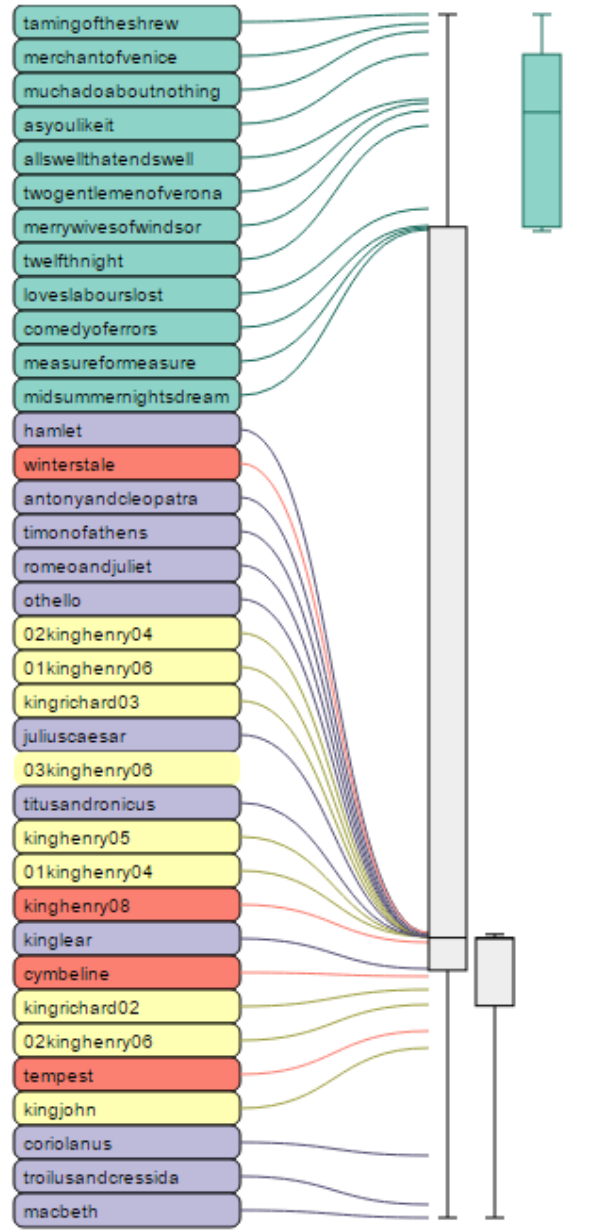
Use Models for Insight

We know the answers!
Teaching the machine can help us:

Can you pick apart known groups?

Does the data capture the concepts?
Language features vs. genre

Can we quantify and organize? (comedic-ness)

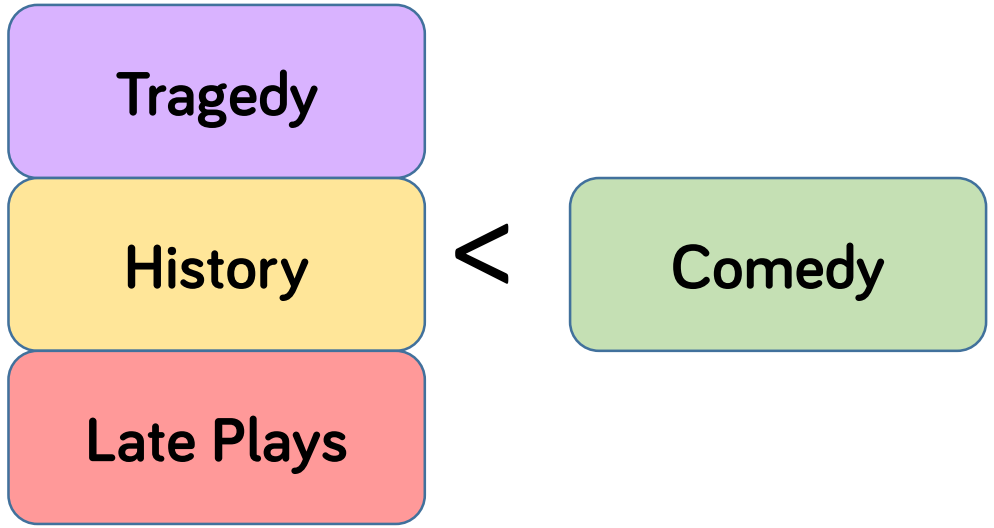


nth(1) mcc(1) margin(0.58)
Docuscope.Precedent_Defending:-125.36
Docuscope.Apology:-45.61
Docuscope.Procedures:25.52
Docuscope.OpenQuery:10.66
and 19 others...

Comedicness

A measure of how much of a comedy something is

It's the "stuff" comedies have more of
Where "stuff" has to be in the data



Organization:

What is most/least comedic?

Explanation:

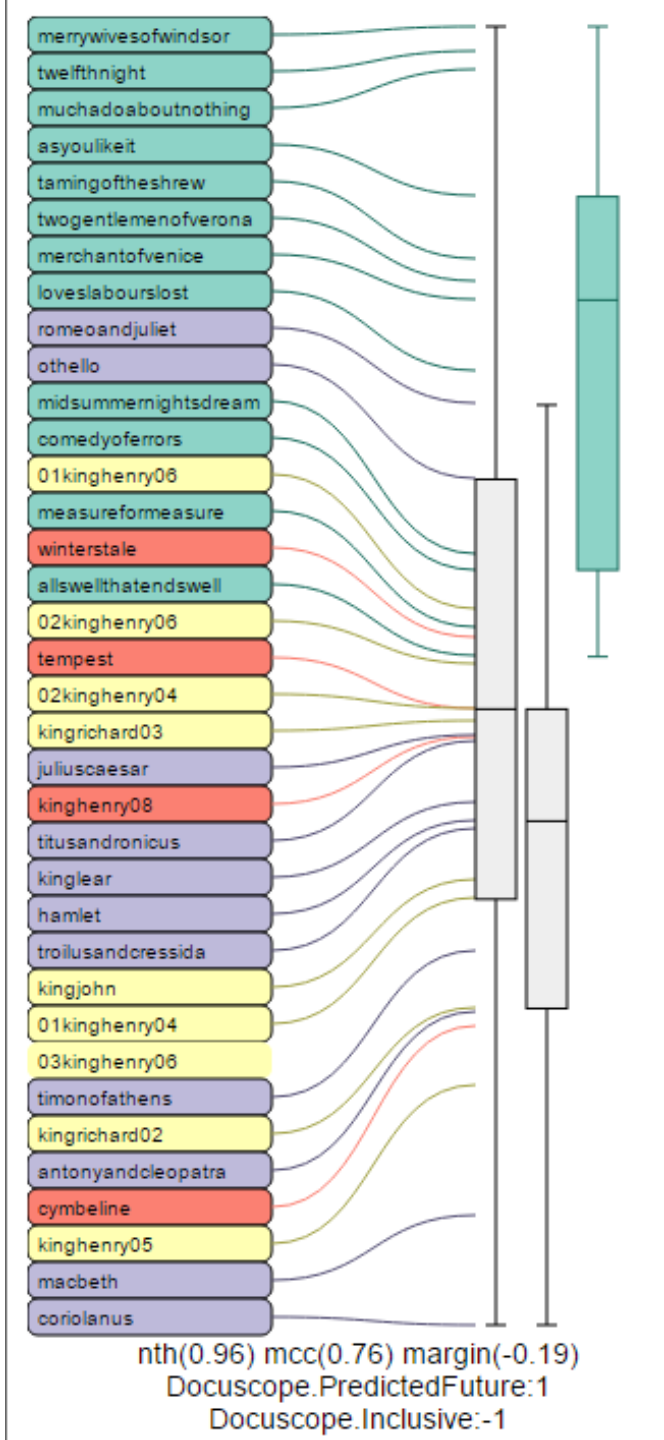
How is the word usage (measured stuff) different in comedies?

Use Simple Models for Insight

Create models for comprehensibility

See what you get wrong

Look “inside” models to see how they work



Gleicher. *Explainers: Expert Explorations with Crafted Projections.* IEEE TVCG 19(12), Dec 2013. Proceedings VAST 2013.

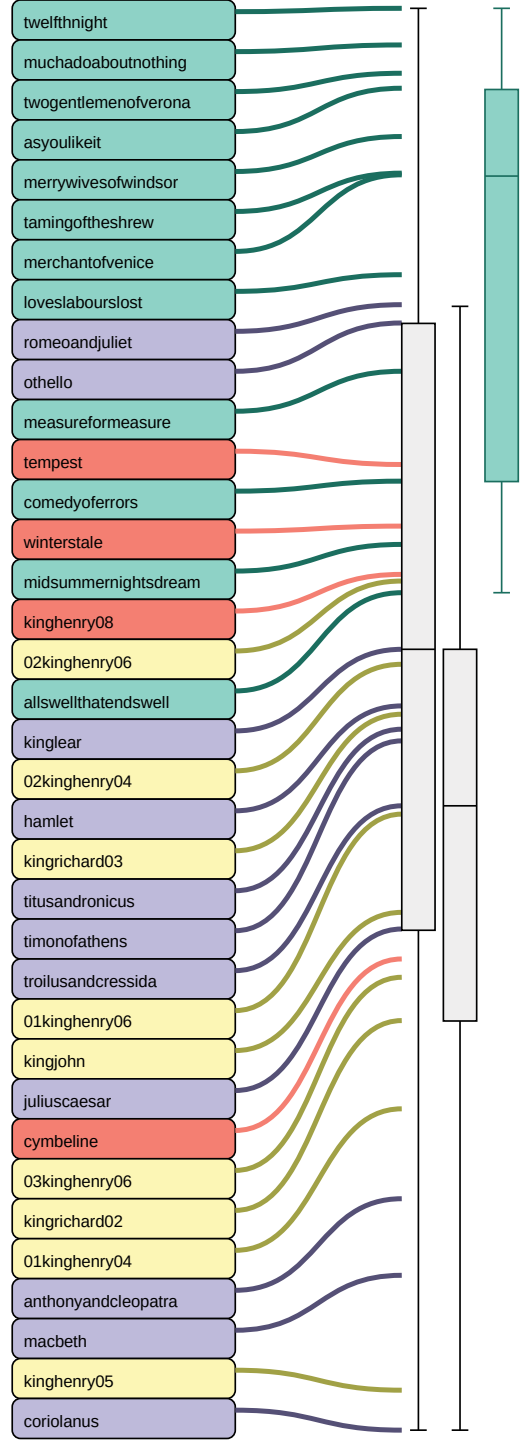
comedicness = M − I

f(V) = V[39] − V[42]

M = Predicted Future

I = Inclusiveness

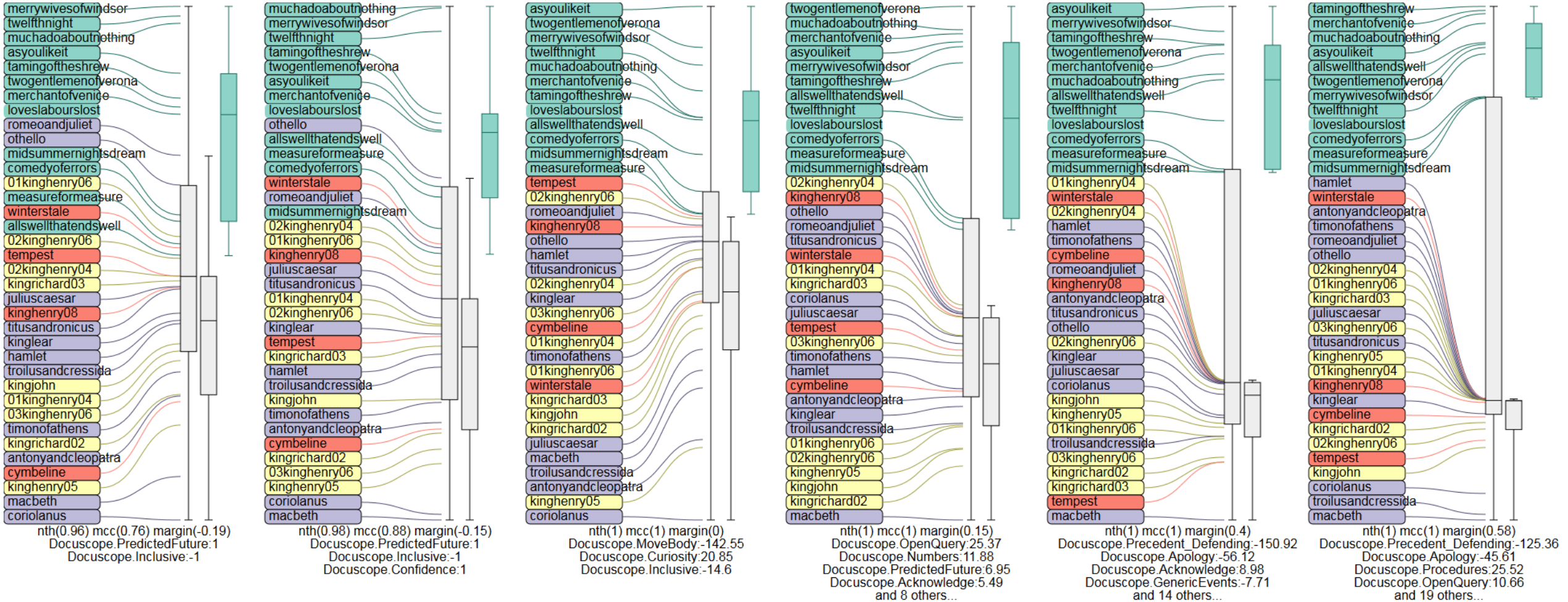
36 plays of Shakespeare, Docuscope Features



M – I	(5 wrong)
C – B – I	(4 wrong)
C – I – 10 M	(1 wrong)
31 D – 100 M – 3 A	(none wrong)

“standard” L1 SVM (none wrong, reasonable margin)

25.3698 Q + 11.8823 U + 6.9492 F + 5.4897 A + 4.1489 P - 3.3765 N + 2.6392 D + 2.0172 F - 1.5404 I + 1.1864 R - 0.7958 C + 0.7272 D


merrywivesofwindsor
twelfthnight
muchadoaboutnothing
asyoulikeit
tamingoftheshrew
twogentlemenofverona
merchantofvenice
loveslabourslost
romeoandjuliet
othello
midsummernightsdream
comedyoferrors
01kinghenry06
measureformeasure
winterstale
allswellthatendswell
02kinghenry06
tempest
02kinghenry04
kingrichard03
juliuscaesar
kinghenry08
titusandronicus
kinglear
hamlet
troilusandcressida
kingjohn
01kinghenry04
03kinghenry06
timonofathens
kingrichard02
antonyandcleopatra
cymbeline
kinghenry05
macbeth
coriolanus
nth(0.96) mcc(0.76) margin(-0.19)
Docuscope.PredictedFuture:1
Docuscope.Inclusive:-1
muchadoaboutnothing
merrywivesofwindsor
twelfthnight
tamingoftheshrew
twogentlemenofverona
asyoulikeit
merchantofvenice
loveslabourslost
othello
allswellthatendswell
measureformeasure
comedyoferrors
winterstale
romeoandjuliet
midsummernightsdream
02kinghenry04
01kinghenry06
kinghenry08
juliuscaesar
titusandronicus
01kinghenry04
02kinghenry06
kinglear
tempest
kingrichard03
hamlet
troilusandcressida
kingjohn
timonofathens
antonyandcleopatra
cymbeline
kingrichard02
03kinghenry06
kinghenry05
coriolanus
macbeth
nth(0.98) mcc(0.88) margin(-0.15)
Docuscope.PredictedFuture:1
Docuscope.Inclusive:-1
Docuscope.Confidence:1
asyoulikeit
twogentlemenofverona
merrywivesofwindsor
twelfthnight
muchadoaboutnothing
merchantofvenice
tamingoftheshrew
loveslabourslost
allswellthatendswell
comedyoferrors
midsummernightsdream
measureformeasure
tempest
02kinghenry06
romeoandjuliet
kinghenry08
othello
hamlet
titusandronicus
02kinghenry04
kinglear
03kinghenry06
cymbeline
01kinghenry04
timonofathens
01kinghenry06
winterstale
kingrichard03
kingjohn
kingrichard02
juliuscaesar
macbeth
troilusandcressida
antonyandcleopatra
kinghenry05
coriolanus
nth(1) mcc(1) margin(0)
Docuscope.MoveBody:-142.55
Docuscope.Curiosity:20.85
Docuscope.Inclusive:-14.6
twogentlemenofverona
muchadoaboutnothing
merchantofvenice
asyoulikeit
merrywivesofwindsor
tamingoftheshrew
allswellthatendswell
twelfthnight
loveslabourslost
comedyoferrors
measureformeasure
midsummernightsdream
02kinghenry04
kinghenry08
othello
romeoandjuliet
titusandronicus
winterstale
01kinghenry04
kingrichard03
coriolanus
juliuscaesar
tempest
03kinghenry06
timonofathens
hamlet
cymbeline
antonyandcleopatra
kinglear
troilusandcressida
01kinghenry06
02kinghenry06
kinghenry05
kingjohn
kingrichard02
macbeth
nth(1) mcc(1) margin(0.15)
Docuscope.OpenQuery:25.37
Docuscope.Numbers:11.88
Docuscope.PredictedFuture:6.95
Docuscope.Acknowledge:5.49
and 8 others...
asyoulikeit
merrywivesofwindsor
tamingoftheshrew
twogentlemenofverona
merchantofvenice
muchadoaboutnothing
allswellthatendswell
twelfthnight
loveslabourslost
comedyoferrors
measureformeasure
midsummernightsdream
01kinghenry04
winterstale
02kinghenry04
hamlet
timonofathens
cymbeline
romeoandjuliet
kinghenry08
antonyandcleopatra
titusandronicus
othello
02kinghenry06
kinglear
juliuscaesar
coriolanus
kingjohn
kinghenry05
01kinghenry06
troilusandcressida
03kinghenry06
kingrichard02
kingrichard03
tempest
macbeth
nth(1) mcc(1) margin(0.4)
Docuscope.Precedent_Defending:-150.92
Docuscope.Apology:-56.12
Docuscope.Acknowledge:8.98
Docuscope.GenericEvents:-7.71
and 14 others...
tamingoftheshrew
merchantofvenice
muchadoaboutnothing
asyoulikeit
allswellthatendswell
twogentlemenofverona
merrywivesofwindsor
twelfthnight
loveslabourslost
comedyoferrors
measureformeasure
midsummernightsdream
hamlet
winterstale
antonyandcleopatra
timonofathens
romeoandjuliet
othello
02kinghenry04
01kinghenry06
kingrichard03
juliuscaesar
03kinghenry06
titusandronicus
kinghenry05
01kinghenry04
kinghenry08
kinglear
cymbeline
kingrichard02
02kinghenry06
tempest
kingjohn
coriolanus
troilusandcressida
macbeth
nth(1) mcc(1) margin(0.58)
Docuscope.Precedent_Defending:-125.36
Docuscope.Apology:-45.61
Docuscope.Procedures:25.52
Docuscope.OpenQuery:10.66
and 19 others...

Simpler Functions

Easier to Understand

More likely
to lead to Theory

Less Expressive

Less Likely
to be Accurate
(but more likely to overfit)

Tradeoffs

Give the **user** control over the tradeoffs

But how do we help them make informed choices?

Paths to model usability?

Interpretable models

simplify the models so they can be understood

Examinable models

look inside the models and hope you understand

Instance-based explanations

pick some decisions and try to understand them

Experiment/Outcome Examination

look at the right input/outputs from the black box

Wrong?

Interesting Outliers

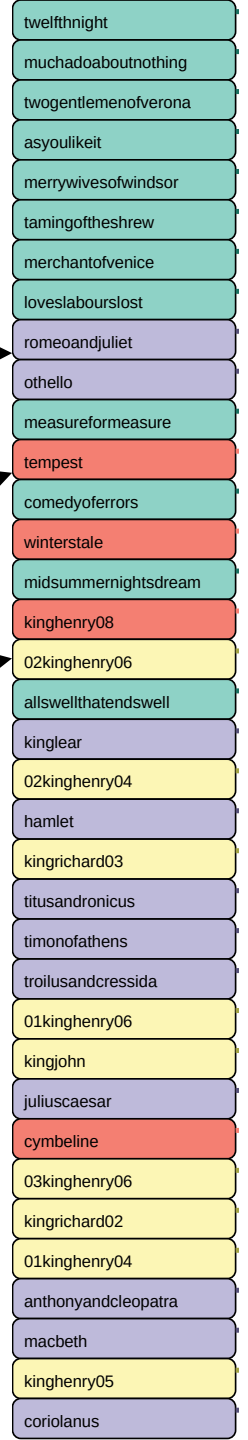
"Romeo and Juliet" is pretty comedic

Ambiguous Classifications

Late Plays are called Tragi-Comedies

Near-Misses

A tiny shift, and this would be different

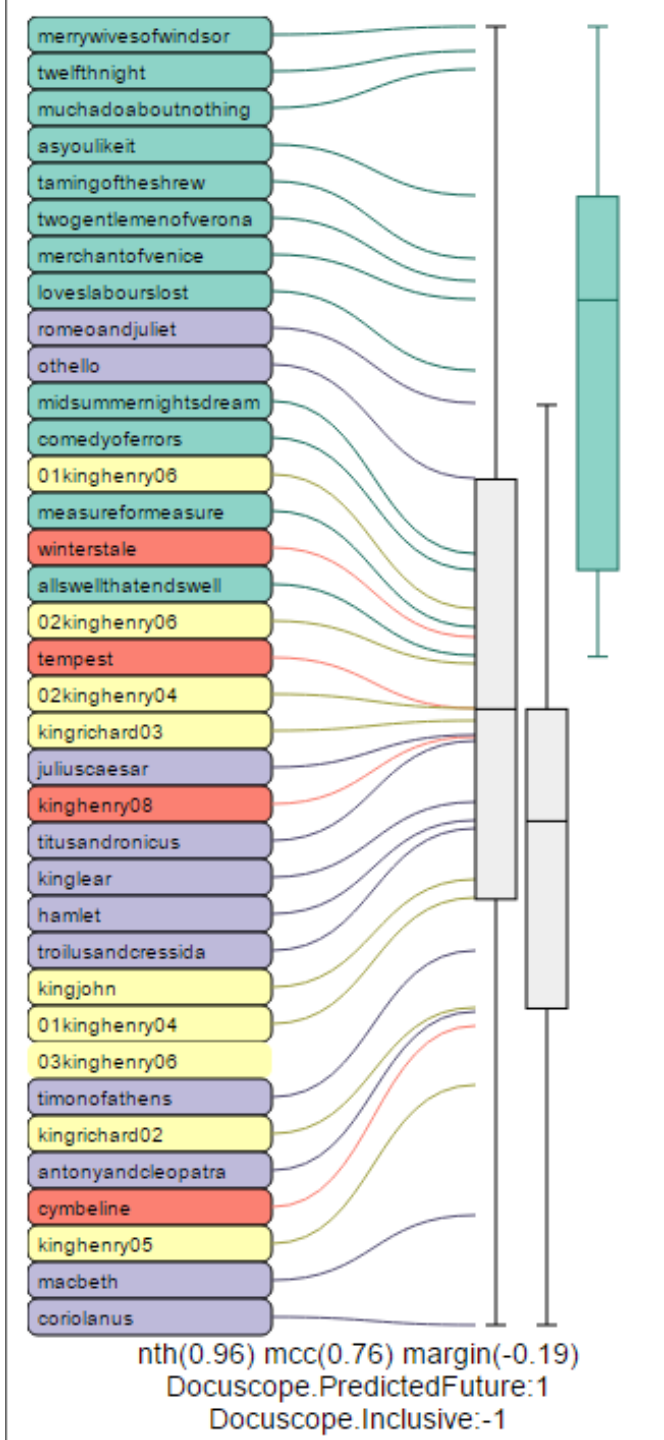


Use Simple Models for Insight

Create models for comprehensibility

See what you get wrong

Look “inside” models to see how they work



Gleicher. *Explainers: Expert Explorations with Crafted Projections.* IEEE TVCG 19(12), Dec 2013. Proceedings VAST 2013.

Lesson

Good tools for exploring your results are useful.

Is literature unique?

Large library has evolved

Understanding different kinds of variation

Bringing in other knowledge

Look at specific examples and outliers

Drill into details

Learn from validation experiments

Get **stakeholders** involved in
all **data science phases**
by helping them **understand**

Some Lessons from “Humanist” Thinking

They got along fine before “data-centric” thinking

Importance of **exemplars** and **outliers**

Importance of going back to the source (specific passages)

Value of multiple points of view

Contextualized arguments with lots of background

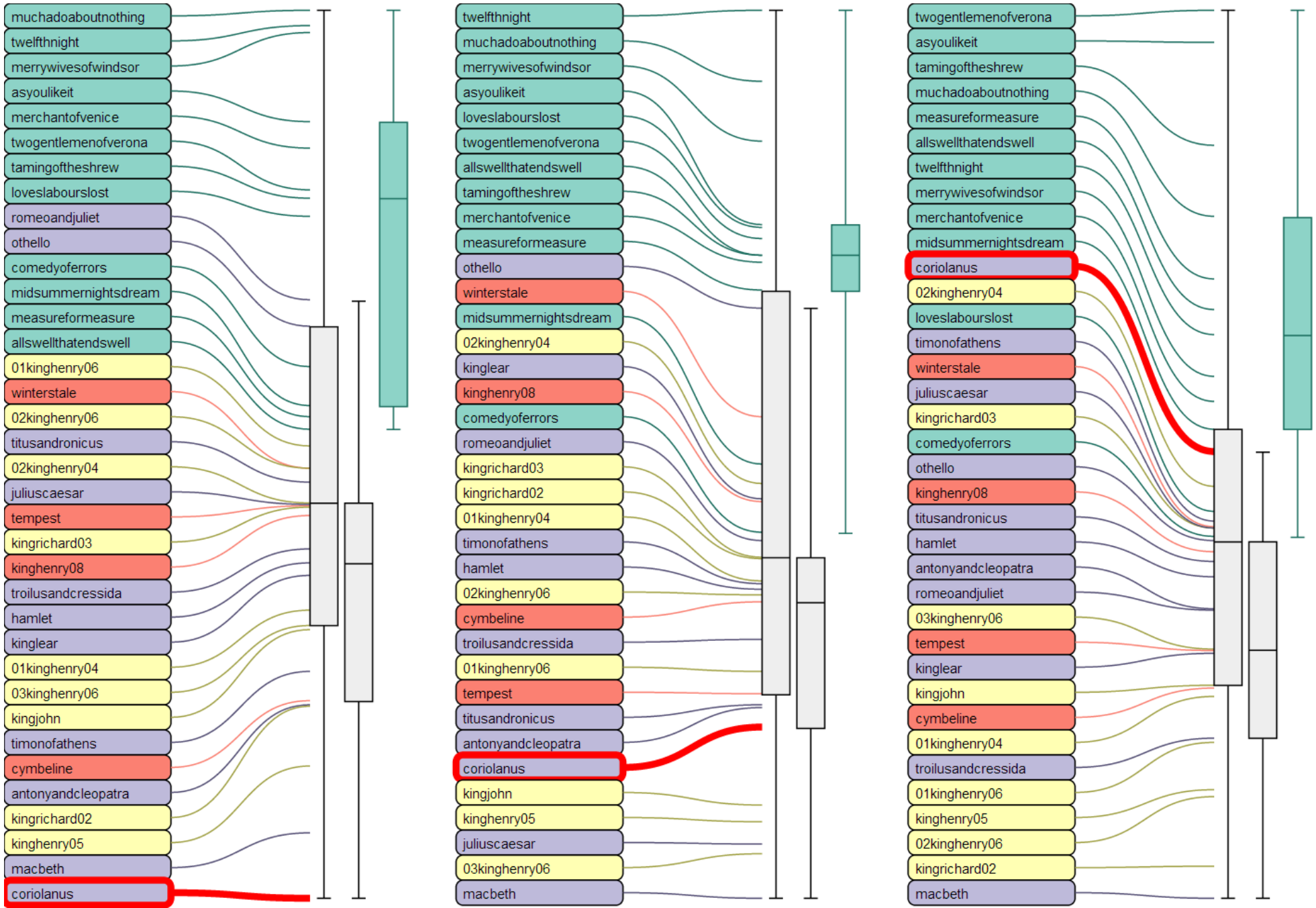
Editing and curation are scholarly activities

Diversity

F + Q − I
C − M − I
P + N + D

Same:
Correctness
Simplicity

Different:
Explanations
Orderings



Maybe Literature Scholars aren't so weird?

They got along fine before "data-centric" thinking

Importance of **exemplars** and **outliers**

Importance of going back to the **source** (specific passages)

Value of **multiple** points of view

Contextualized arguments with lots of **background**

Editing and **curation** are scholarly activities

OK, some things are weird

Queſtion.

Náttoohtemâuwetoowunk.

How prove you that there is a God?

Oohgôdje korámen neh átta Atandouh

Anſwer.

Anaſquetáuweten.

From the univerſal and conſtant agree-

Wutche wéwinnakommuk quah yeiache wérrawân-

ment of all Nations, and perſons

wunk wutche wame arkèes, quah ſkeetambâwg mit-

in the world, who are not void of

tâuhkuk terre, owwànnak matta ſâuwaioóguk wutche

right reaſon and humanity.

ſompâio penauwáuwuk quah renôowunk.

For the things which are grounded

Wutche ai akquíiks chawgwunſh wekakontamoo-

upon particular mens fancies

awk skeje nanſcêawk rénwawk róytammoúnganſh

and opinions are not acknow-

quah wàrramawóytammounganſh matta wérramattau-

ledged of all men, and are

ôomunks wutche wame rènawawk, quah wegonje

often changed but this notion that

àſſowunnamanôoſh: webe (youh côytámmoounk) neh

there is a God is common to all men, nor it it chan-

Mandoo nànnarwee re wâme rènawawk matta áſſowú-

ged by the changes of times;

numôoanas ſpe aſſowunnàmoúnganſh quompaious;

therefore it muſt arise from

règouche youh paughke môuche ſongème wutche

ſome light, which is common to all

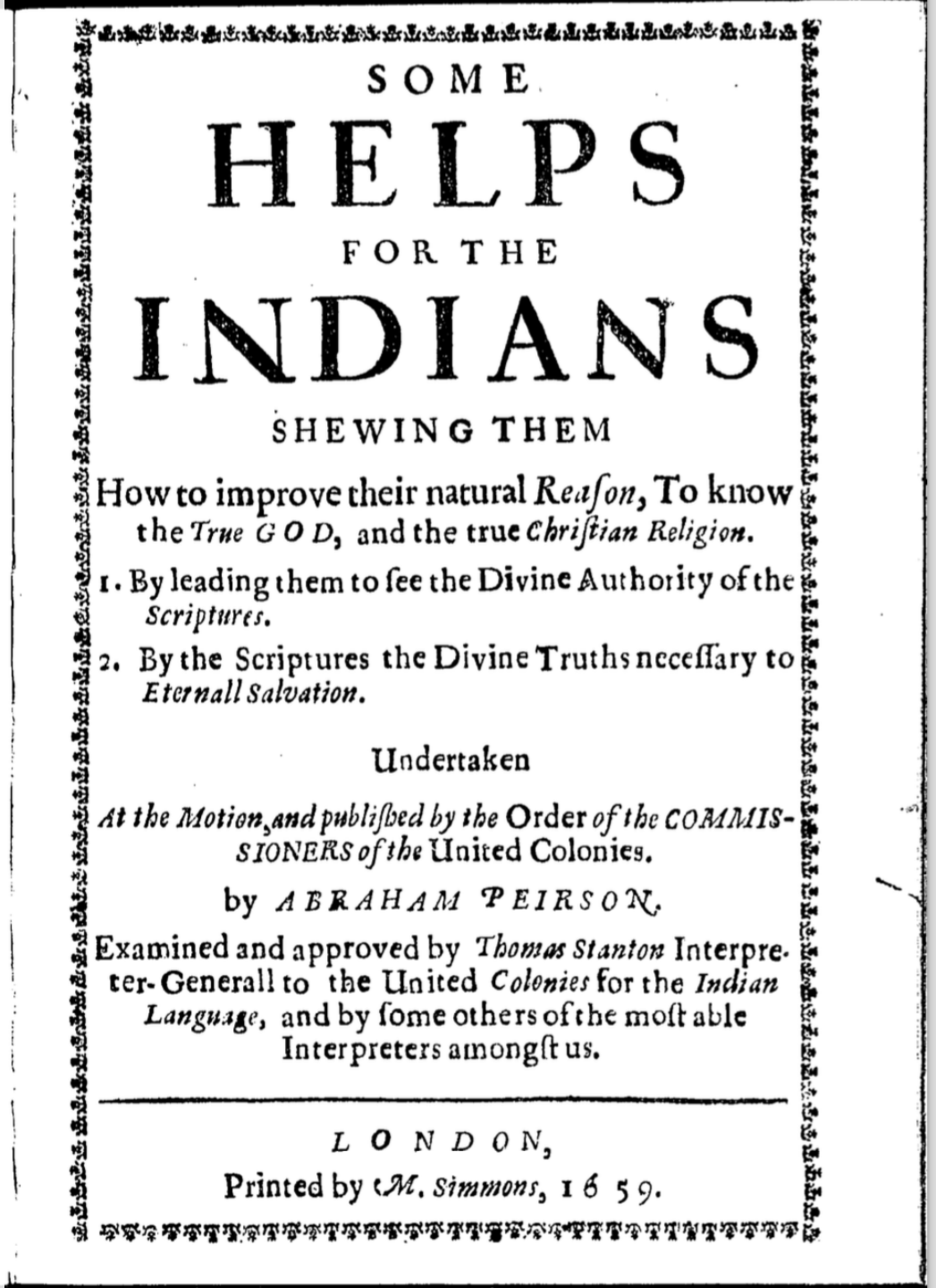
chawgun nowèta wequâ-ai, teou nannarwe re wame

rèn-

But, this might be a lesson …

Don’t just give “them” your methods

If you listen carefully,
they might have things to teach you



SOME

HELPS

FOR THE

INDIANS

SHEWING THEM

How to improve their natural *Reason*, To know the *True GOD*, and the true *Christian Religion*.

1. By leading them to see the Divine Authority of the *Scriptures*.
2. By the Scriptures the Divine Truths necessary to *Eternall Salvation*.

Undertaken

At the Motion, and published by the Order *of the COMMISSIONERS of the* United Colonies.

by *ABRAHAM PEIRSON*.

Examined and approved by *Thomas Stanton* Interpreter-Generall to the United *Colonies* for the *Indian Language*, and by some others of the most able Interpreters amongst us.

LONDON,

Printed by *M. Simmons*, 1659.

Thanks!

To you for listening.
To my students and collaborators.
To the NSF, NIH, DARPA and Mellon Foundation for funding.

What Shakespeare taught us about (Visual) Data Science

Michael Gleicher

University of Wisconsin Madison

`gleicher@cs.wisc.edu`