

Text Mining the Indian Claims Commission:

And other cool ways you can turn digital text into historical insight

Peter Carr Jones

PhD Candidate

Department of History and Art History

George Mason University

Today's Discussion

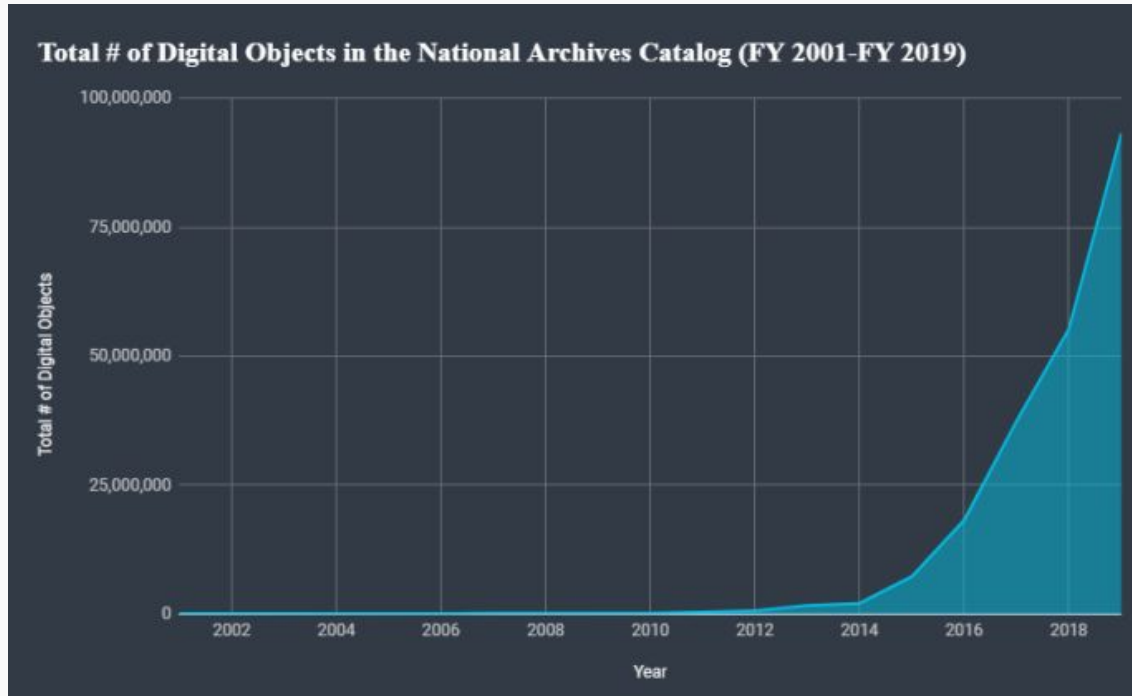
1. Why do we need digital workflows?
2. How I used digital workflows to help answer historical questions?
3. What lessons can you use in your own work?

Physical -> Digital Archives

- Explosion in digital objects

Physical -> Digital Archives

Source:
"Making Access
Happen:
NARA's
Leadership in
the Digital
Decade" [AOTUS
Blog](#) (2019)



Physical -> Digital Archives

- Explosion in digital objects
- NARA FY 2020 Budget:
 - -\$30 million in operations and personnel
 - +\$22 million in digital scanning project

Physical -> Digital Archives

- Reduction in employee hours
 - Description of materials
 - Researcher Facility Hours

Physical -> Digital Archives

- Hathitrust
- Internet Archive
- Digital Public Library of America
- Proprietary Databases

ICC Decisions, 1946-78

1946 - Termination Era

- 2,700+ decisions
- 25,000+ pages in 43 Volumes

1978 - Self-Determination

- 852 claims in 610 dockets.

Digital Analysis

1. **WGET -> downloads decisions (images in PDF form)**
2. **Tesseract -> turns images to text**
3. **RStudio - statistical programming environment ->**
 - a. **ingests text, outputs analysis**

MALLET

1. Text mining program from Andrew McCallum, David Mimno, and others of UMASS <-computer scientists
2. Many types of natural language processing, document classification, topic modeling, etc.
 - a. **Latent Dirichlet Allocation (LDA) <- Topic Modeling**

David Mimno's Website

<https://mimno.infosci.cornell.edu/jsLDA/jslda.html>

[16] government
american states
america national
such work nation
peace war

government federal war



government war economic



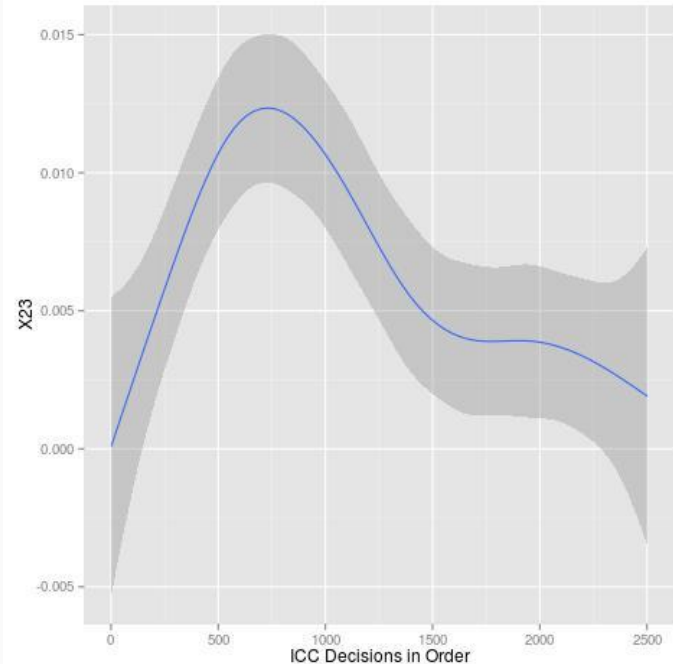
[17] government
federal america
peace country united
war economic
american any

government country national



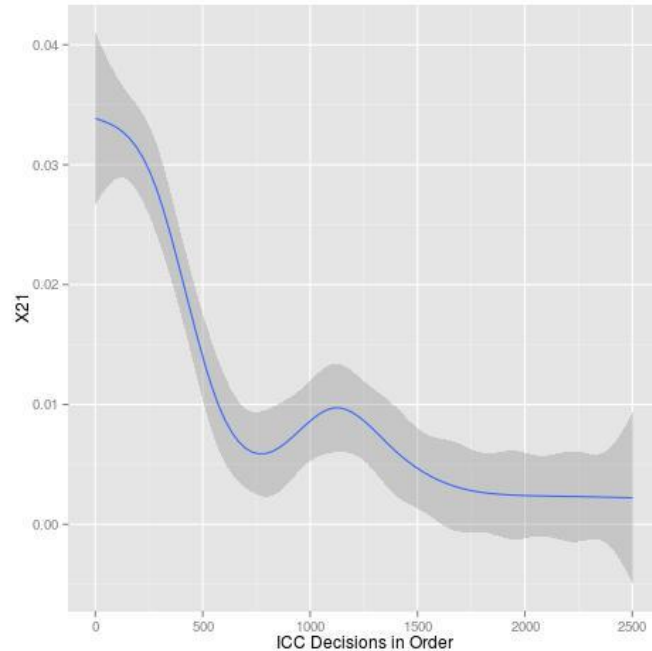
Topic - Expert Witnesses

compensation
witnesses
testified work employed
days
hearing expert paid
witness employment
testimony counsel
hearings agreement



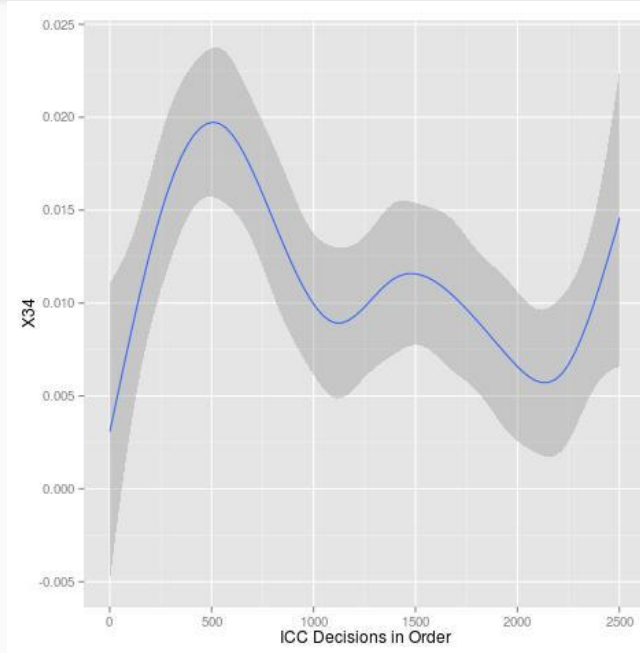
Typical Case -Phase 1

recognition
exclusive occupied
decided claimant
area recognized
territory title original
rights
occupancy
held boundaries
claimed



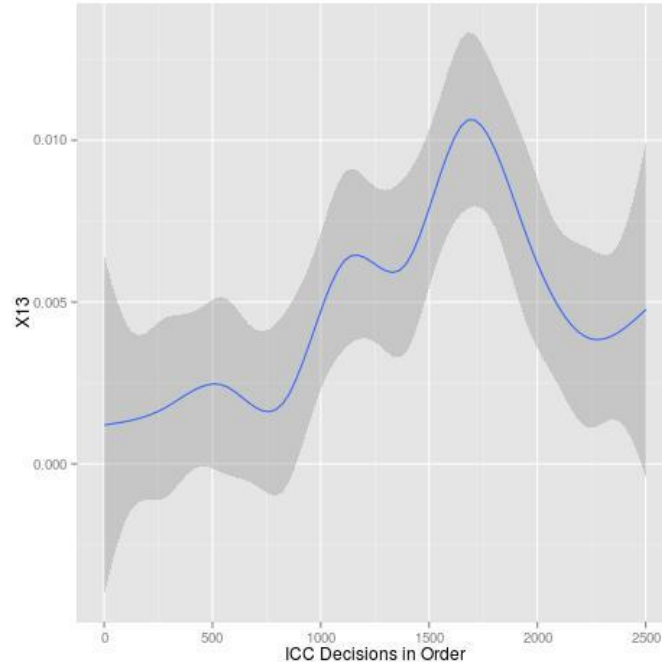
Typical Case -Phase 2

price public
average acre sale
settlers soil period
market land year
acres years
sales sold

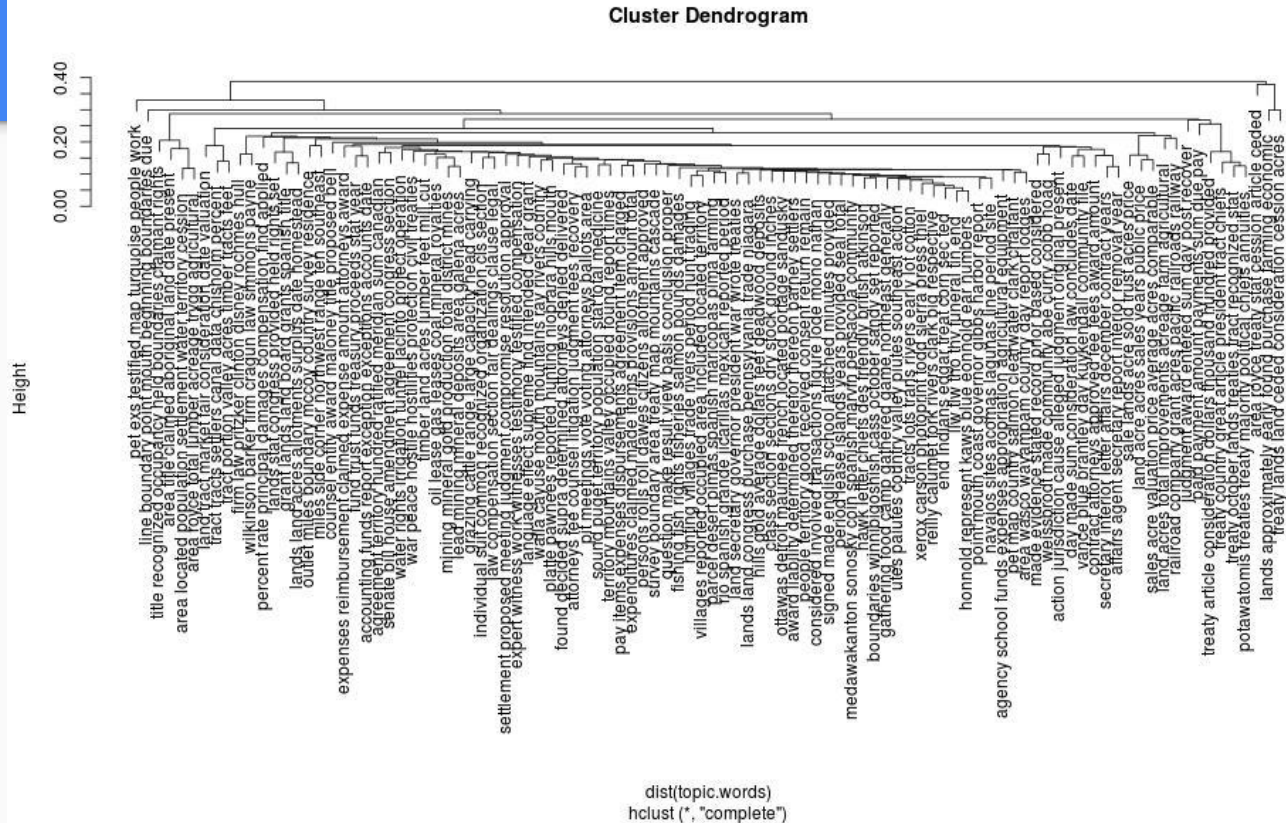


Typical Case -Phase 3

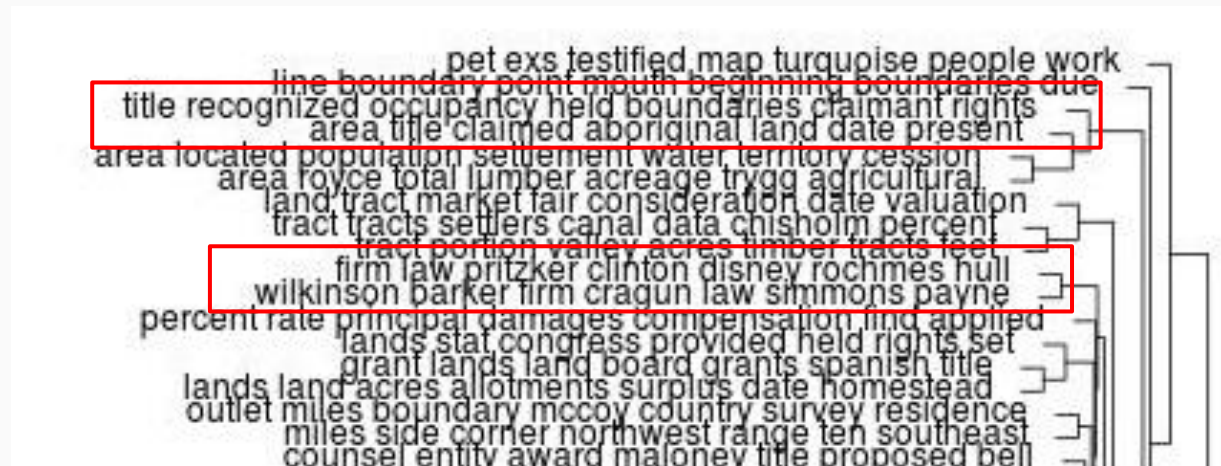
resulting simple
compensation
damages rate formula
percent
principal find
applied net
annum computed
award percentage



Cluster Dendrogram



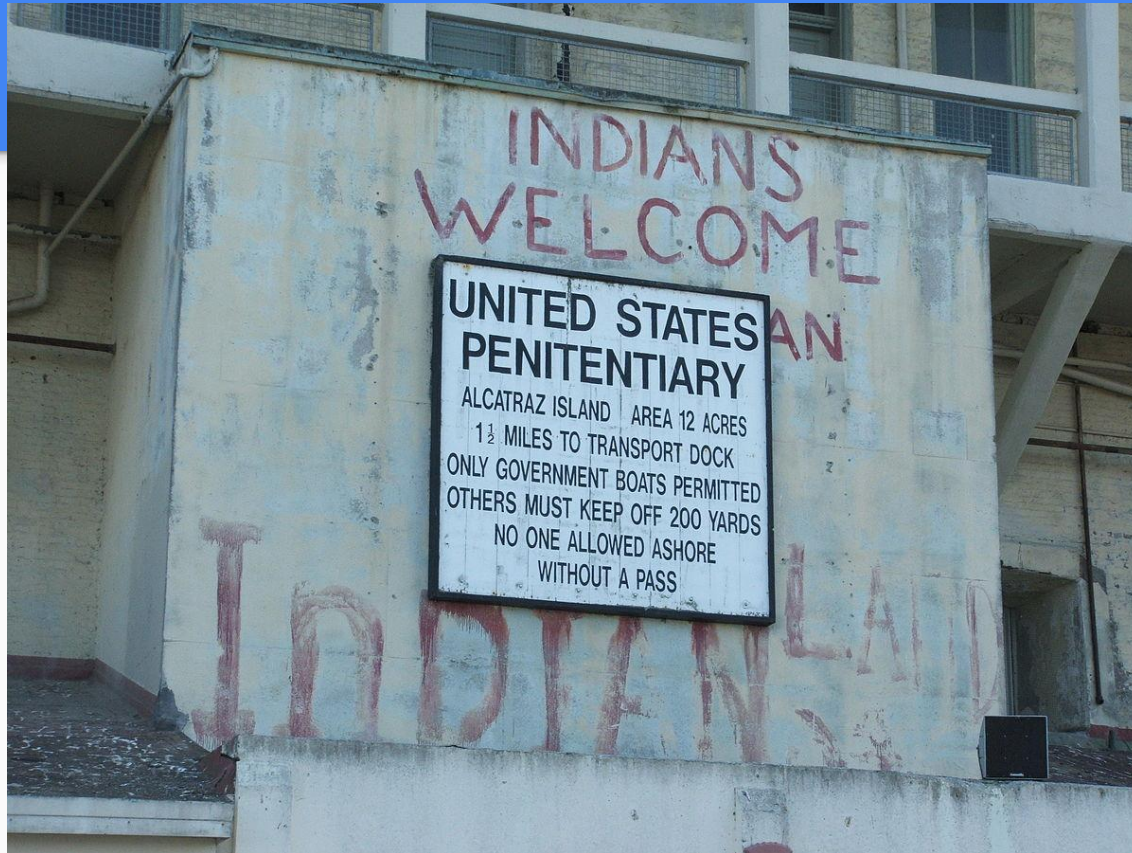
Cluster Dendrogram



Expert Witness - Near Neighbor

grazing cattle range large capacity head carrying
individual suit common recognized organization cis section
law compensation section fair dealings clause legal
settlement proposed meeting judgment compromise resolution approval
expert witness work witnesses testimony testified compensation
language effect supreme find intended clear grant
walla cayuse mouth mountains ray rivers country
. . . platte pawnees reported hunting niopara hills mouth

Occupation of Alcatraz

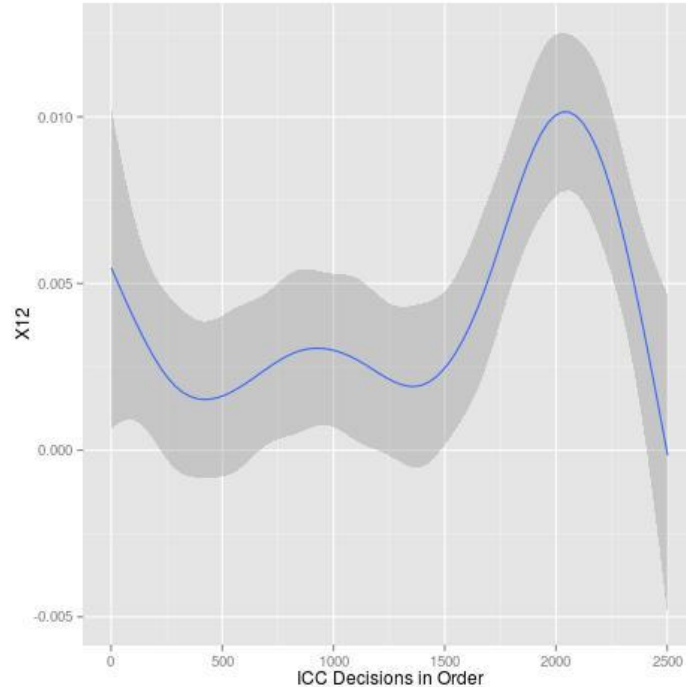


Symbolic Judgment

“I’ll always say it was that Indian Money that freed us from bondage... because so many of those who had been so down on the Indians had to face up to us over that money.” -Creek Nation Member

Topic - Trust Accounting

appropriation
rate treasury
stat trust
year fund
interior interior moneys
funds held congress
proceeds law
secretary



Reconciliation / Transitive Justice

- 1. Critical Reflection**
- 2. Symbolic/Material Recognition**
- 3. Political Participation**

Best Practices

1. Ask Historical Questions First.

Best Practices

1. Ask Historical Questions First.
2. Experiment with available data

Best Practices

1. Ask Historical Questions First.
2. Experiment with available data
3. Reproducible Results