



Exploring Aesthetic Communities with Text Mining and Data Visualization:

The Program Era Project and the Iowa Writers' Workshop

Nicholas M Kelly

Department of Rhetoric

The University of Iowa

@newreadia, nicholas-kelly@uiowa.edu

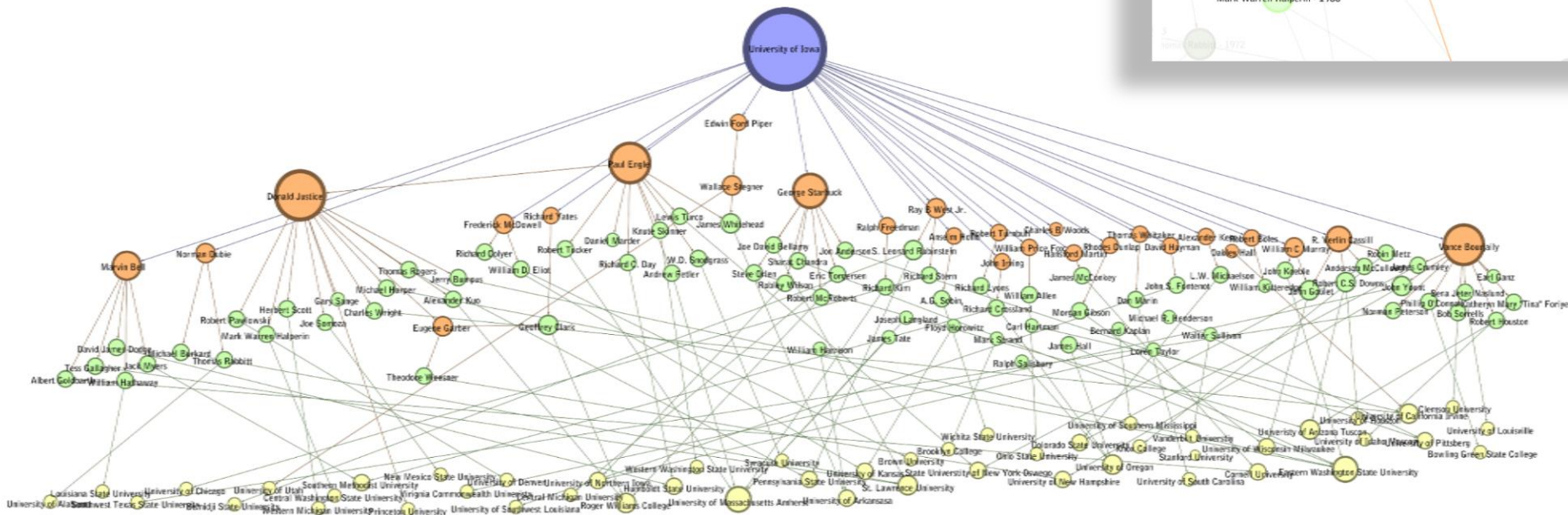
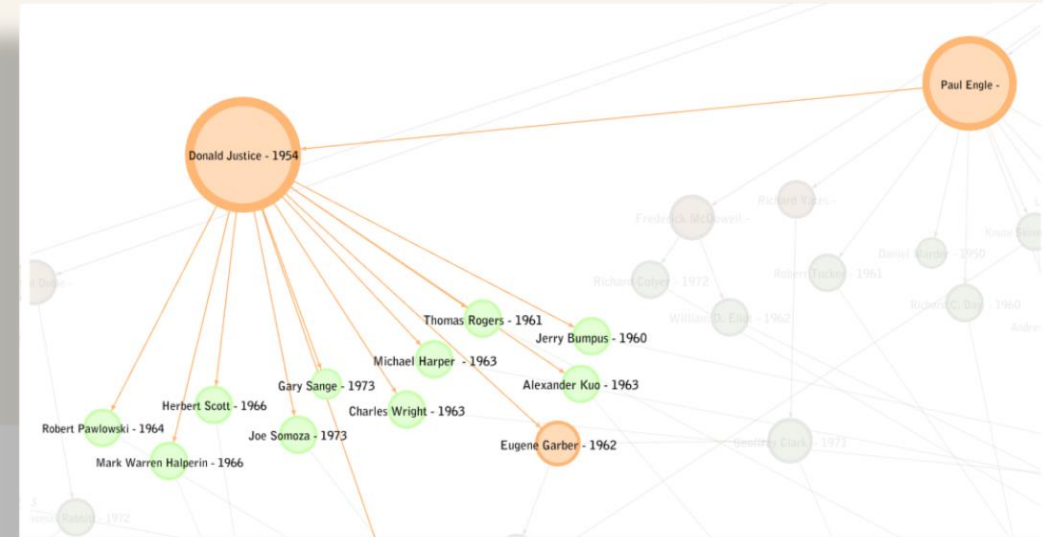
OUR LITERARY HISTORY

“We need to start documenting this phenomenon, moving out from the illustrious cases of the Iowa Writers’ Workshop and Stanford University to grasp the reality of an enterprise that now numbers some 350 institutional participants and continues to grow. This enterprise is our literary history.”

- Mark McGurl, *The Program Era*

THE PROGRAM ERA PROJECT

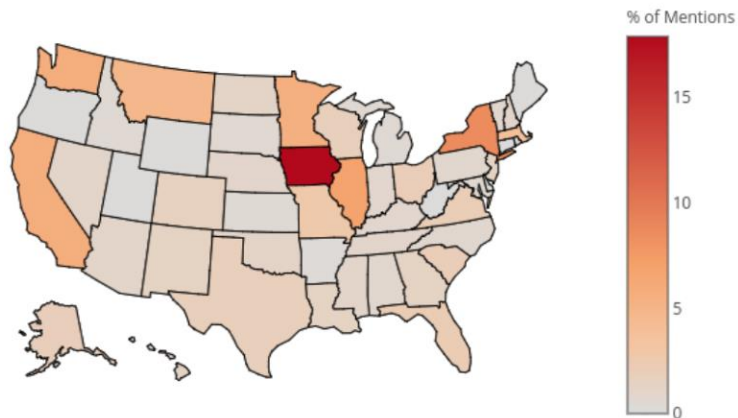
- Develop techniques to study Program Era
- Create public database on Iowa Writers' Workshop
 - Writers (Professional/Biographical info)
 - Workshop writing text metrics
 - Formal features, sentiment, location references, topic modeling...



PEP AND AESTHETIC COMMUNITIES

- PEP, Text Mining, Data Visualization
- “Squatter Regionalism”

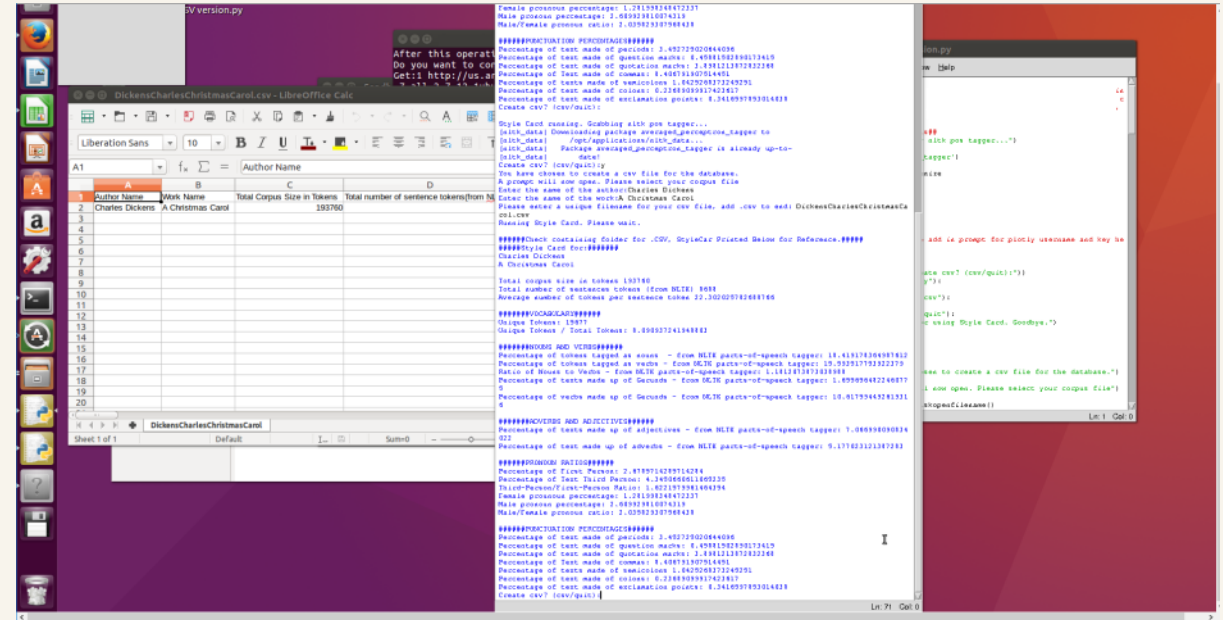
W.P. Kinsella, % of Total Mentions by State



Author Name	Work Name	IA % of Total Mentions
Cassill, R. V.	<i>The happy marriage, and other stories</i>	51.35135135
McCracken, Elizabeth	<i>Niagara Falls all over again</i>	48.55305466
Irving, John	<i>The water-method man</i>	46.678526
Irving, John	<i>3 by Irving</i>	40.94853409
Cassill, R. V.	<i>Three stories</i>	40
Stuckey-French, Elizabeth	<i>The first paper girl in Red Oak, Iowa ; and other stories</i>	38.11881188
Shields, David	<i>Heroes</i>	33.83806071
Smiley, Jane	<i>At paradise gate</i>	31.94444444
Kinsella, W. P.	<i>Shoeless Joe</i>	31.58414213
Tiempo, Edilberto K.	<i>Finalities : a novelette and five short stories</i>	28.88888889
Irving, John	<i>The Hotel New Hampshire</i>	28.37671402
Grumbach, Doris	<i>The spoil of the flowers.</i>	28.12656956
Yates, Richard	<i>Young hearts crying</i>	28.10110326
Smiley, Jane	<i>A thousand acres</i>	27.7027027
Wittlinger, Ellen	<i>Lombardo's law</i>	25.66666667
Stefaniak, Mary Helen	<i>Self storage and other stories</i>	24.3902439
Schramm, Wilbur Lang	<i>Windwagon Smith and other yarns.</i>	22.64150943
Godwin, Gail	<i>Evenings at five</i>	22.22222222
Irving, John	<i>The 158-pound marriage</i>	20.125
Cheever, John	<i>Oh, what a paradise it seems</i>	20
Halde man, Joe W.	<i>Camouflage</i>	19.62616822
Cassill, R. V.	<i>Collected stories</i>	19.3043674
Louie, David Wong	<i>The barbarians are coming : a novel</i>	18.57923497
Grumbach, Doris	<i>The magician's girl</i>	17.6744186
Kinsella, W. P.	<i>The further adventures of Slugger McBatt : baseball stories</i>	16.93693694

COLLECTING DATA

- Count place names
 - State names (except Washington)
 - State capitols
 - Top 50 U.S. Cities
 - Top four cities
 - Nicknames (Windy City, Big Apple)
- Exceptions
 - Common first names (Norman, Charlotte)
 - Geographic features (Mesa, Butte)
 - Extremely common names (Springfield, Columbus)
 - Geographic features (Mississippi River)
- Assign counts to each state



- Above: HTRC/PEP Data Capsule

DISAMBIGUATION

- Exclude false positives, Collect context
- Weigh context, add ambiguous mentions
 - Total unambiguous references, assign %
 - Distribute ambiguous counts based on proportion
- Aim: Transparency,
 - Comprehend, verify

TESTS

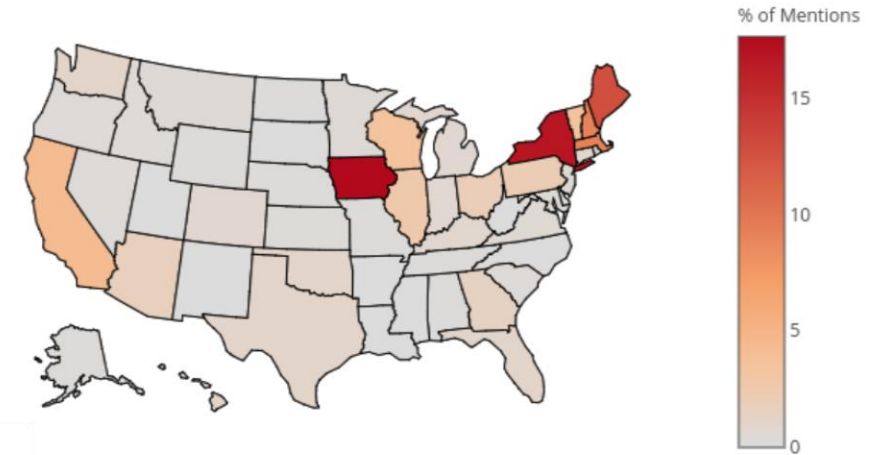
- **Classic Regionalists:**

- Hamlin Garland
- Sarah Orne Jewett
- Willa Cather

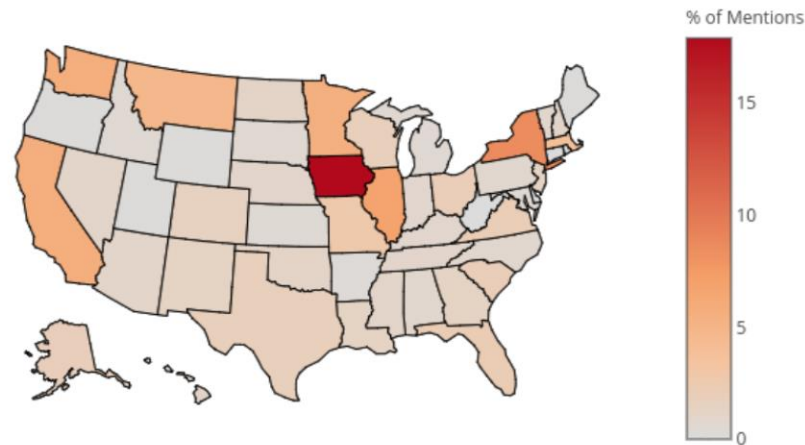
- **Iowa Regionalism:**

- John Irving
- W.P. Kinsella

John Irving, % of Total Mentions by State



W.P. Kinsella, % of Total Mentions by State



- Above: John Irving (14 volumes), born and raised New Hampshire, IWW MFA (1967).
- Left: W.P. Kinsella (9 volumes), born and raised Alberta, Canada, IWW MFA (1978).

CORPUS

- Criteria:
 - Fiction only
 - > 10 Location references
- Trimming:
 - Cut first, last five pages
 - Limit publication, bio material
- Final size:
 - ~ 150 Authors
 - ~ 400 Volumes

FINDINGS - RANKINGS

- New York, California
 - NY: 18% total, 20.7% average
 - CA: 9% total, 10% average
- Iowa 5th, 6th in average/total
 - Higher than .97%, 27th (pop.)

State	% of Total Mentions
NY	18.16061595
CA	9.010267993
IL	5.510935414
MA	5.377919894
IA	4.217383434
VA	4.04132286
GA	3.065163101
TX	3.011978612
FL	2.747485075
MO	2.677888202
PA	2.552503584
TN	2.356026864
KY	1.91900747
ME	1.915344864
WA	1.865112638

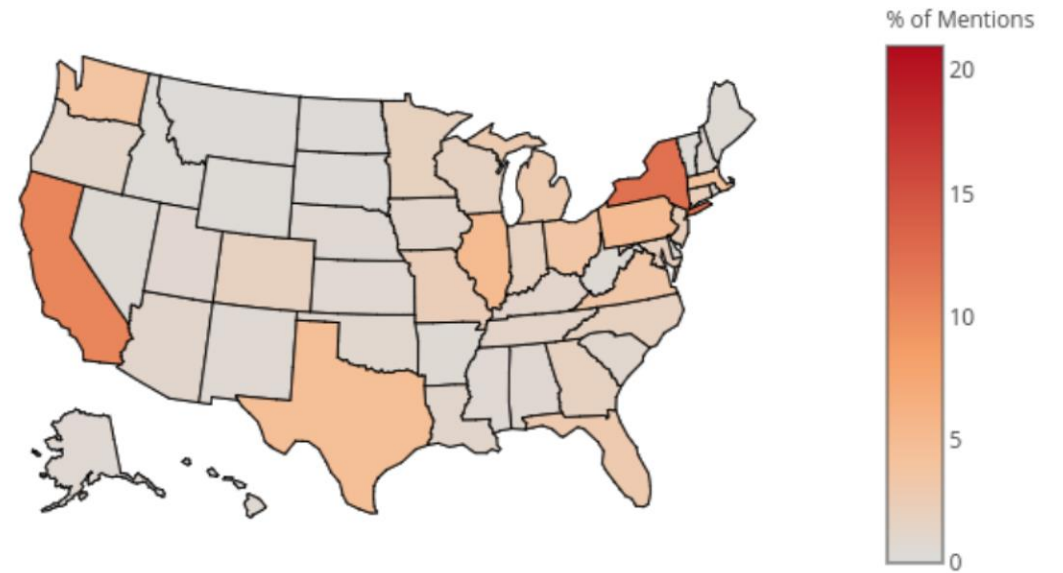
State Name	Avg. % of State Mentions
NY	20.78969801
CA	10.08578956
IL	5.625532571
MA	5.615473166
FL	3.336102142
IA	3.223680295
PA	2.979747509
TX	2.804666588
VA	2.698287641
MO	2.061339615
GA	1.998031815
WA	1.90351601
TN	1.821698227
OH	1.780736881
MI	1.721523164

- Above: Top 15 states by avg. % of State Mentions, HTRC Workshop Corpus
- Left: Top 15 states by % of total mentions, HTRC Workshop corpus

TXTGEO COMPARISON

- TxtGeo
 - Location references in 10 million volumes
 - Comparison corpus
 - 1930 – 2015
 - LOC “Language and Literature”
- TxtGeo Corpus:
 - Iowa = 1.058% of total
 - 1.1x Iowa pop.
- Workshop Corpus:
 - Average: 3.2%
 - 3x TxtGeo
 - Total: 4.2% of
 - 4x TxtGeo

TxtGeo Corpus, % of Total Mentions by State



IOWA VOLUMES

- Top 25
 - Iowa as % of total state mentions.
- Repeat authors
 - John Irving
 - W.P. Kinsella
- New Names
 - Jane Smiley
 - Doris Grumbach
- One Iowan
 - R.V. Cassill

Author Name	Work Name	IA % of Total Mentions
Cassill, R. V.	<i>The happy marriage, and other stories</i>	51.35135135
McCracken, Elizabeth	<i>Niagara Falls all over again</i>	48.55305466
Irving, John	<i>The water-method man</i>	46.678526
Irving, John	<i>3 by Irving</i>	40.94853409
Cassill, R. V.	<i>Three stories</i>	40
Stuckey-French, Elizabeth	<i>The first paper girl in Red Oak, Iowa ; and other stories</i>	38.11881188
Shields, David	<i>Heroes</i>	33.83806071
Smiley, Jane	<i>At paradise gate</i>	31.94444444
Kinsella, W. P.	<i>Shoeless Joe</i>	31.58414213
Tiempo, Edilberto K.	<i>Finalities : a novelette and five short stories</i>	28.88888889
Irving, John	<i>The Hotel New Hampshire</i>	28.37671402
Grumbach, Doris	<i>The spoil of the flowers.</i>	28.12656956
Yates, Richard	<i>Young hearts crying</i>	28.10110326
Smiley, Jane	<i>A thousand acres</i>	27.7027027
Wittlinger, Ellen	<i>Lombardo's law</i>	25.66666667
Stefaniak, Mary Helen	<i>Self storage and other stories</i>	24.3902439
Schramm, Wilbur Lang	<i>Windwagon Smith and other yarns.</i>	22.64150943
Godwin, Gail	<i>Evenings at five</i>	22.22222222
Irving, John	<i>The 158-pound marriage</i>	20.125
Cheever, John	<i>Oh, what a paradise it seems</i>	20
Halde man, Joe W.	<i>Camouflage</i>	19.62616822
Cassill, R. V.	<i>Collected stories</i>	19.3043674
Louie, David Wong	<i>The barbarians are coming : a novel</i>	18.57923497
Grumbach, Doris	<i>The magician's girl</i>	17.6744186
Kinsella, W. P.	<i>The further adventures of Slugger McBatt : baseball stories</i>	16.93693694

SIGNIFICANCE

- IOWA part of larger trend?
 - Implications for post-war writing
- Starting point for future scholarship
 - Suggestions, comments?
- Expand, diversify toolkit
 - Examine Program Era at scale

