

# Patterns and Outliers Across Cather's Correspondence

---

Gabi Kirilloff (Washington University in St. Louis), Matthew Lavin (Denison University), and Sean McCullough (TCU)

# Agenda

- Computational analysis
- Cather's letters as data
- Our project
  - Connections
  - Time
  - Geography

# Sandbox

[https://gkirilloff.github.io.](https://gkirilloff.github.io)

# Computational Analysis

## Humans reading

The buzz in the street was like the humming of flies.

Photographers stood massed behind barriers patrolled by police, their long-snouted cameras poised, their breath rising like steam. Snow fell steadily on to hats and shoulders; gloved fingers wiped lenses clear.

## Machines reading

The buzz in the street was like the humming of flies.

Photographers stood massed behind barriers patrolled by police, their long-snouted cameras poised, their breath rising like steam. Snow fell steadily on to hats and shoulders; gloved fingers wiped lenses clear.

# Cather's Letters as Data

# Cather's Letters as Data

- 610,000 words written to over 500 unique recipients
- Over 8,900 references to people (2,400 unique)
- Over 4,000 references to places (600 unique)
- Over 3,500 references to works (1,200 unique)

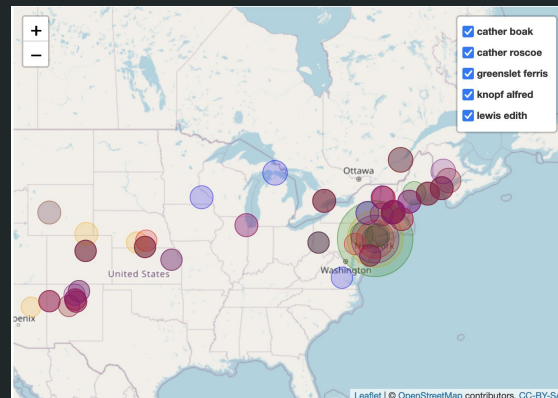
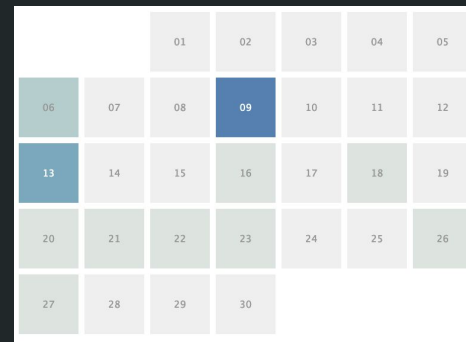
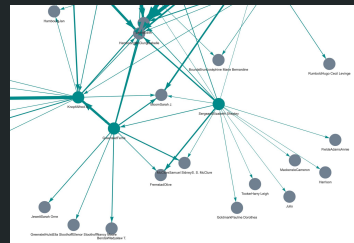
```
<opener>
  <dateline>
    <placeName target="geo:0210">Taos, N.M.</placeName> <lb/>
    <date when="1916-07-08">July 8</date>
  </dateline>
  <salute>My Dear <persName target="psn:0200">Boy</persName>:</salute>
</opener>

<p>I got your letter just as <persName target="psn:0560">Miss Lewis</persName>
and I were leaving <placeName target="geo:0062">Denver</placeName> for
<placeName target="geo:0210">Taos</placeName>. I would have preferred
<placeName target="geo:0106">Lander</placeName>, but Edith wanted very
much to return here, where we had a delightful week last summer. Probably we
```



# An Opportunity

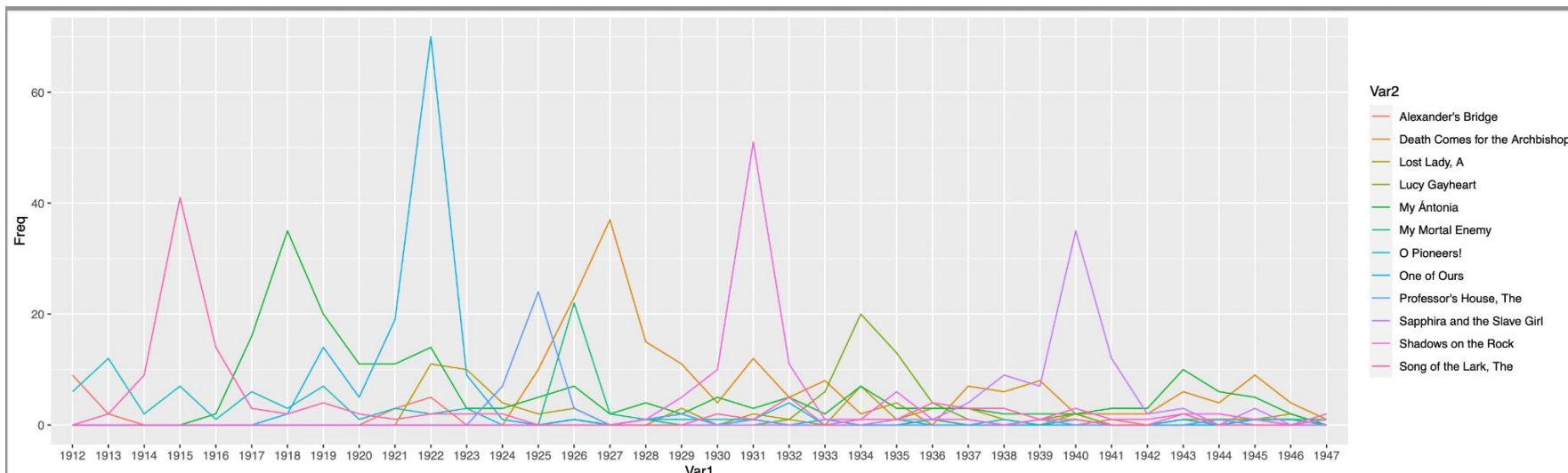
1. Visualize connections and gaps in the data.
2. Use data to provoke new interpretations/insights.



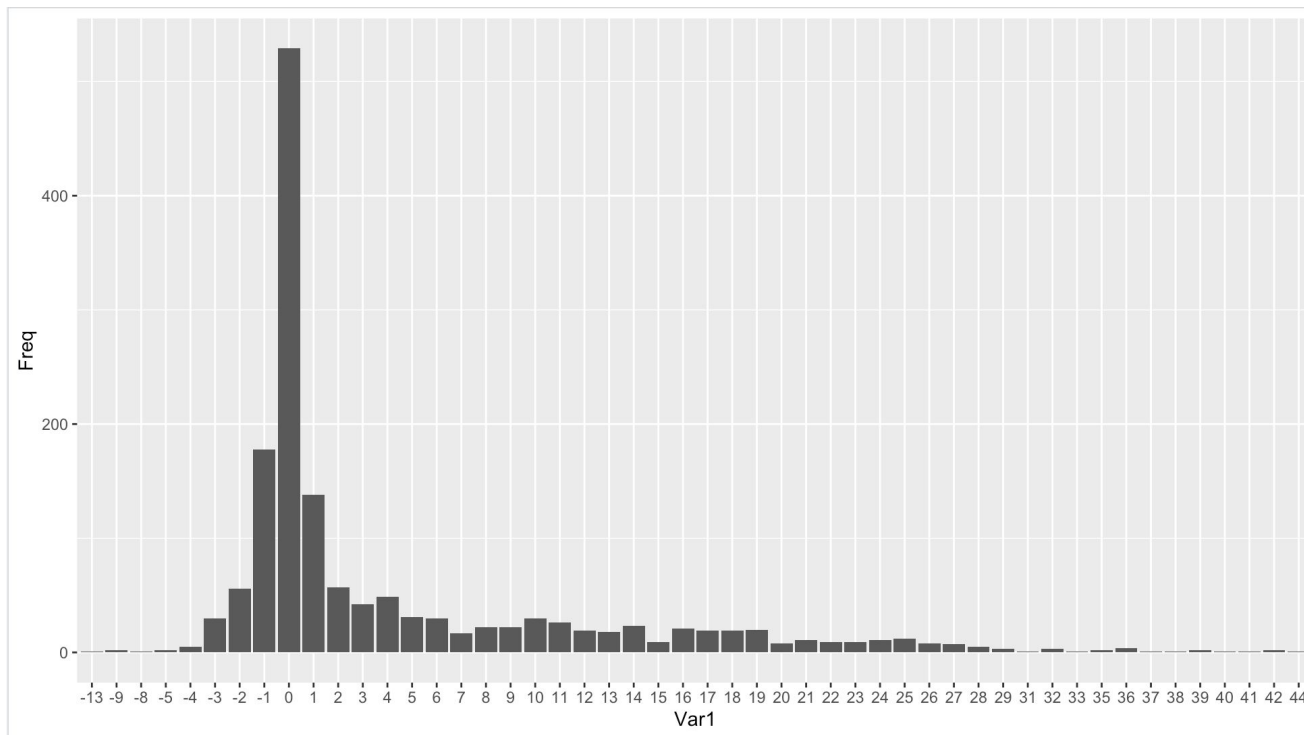


# Connections

# Big Picture

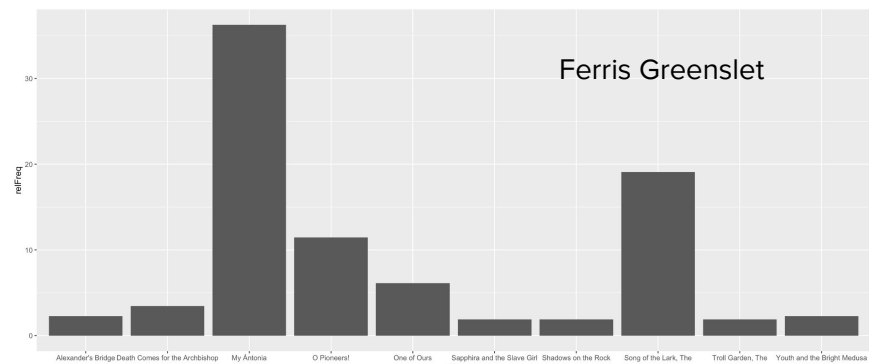
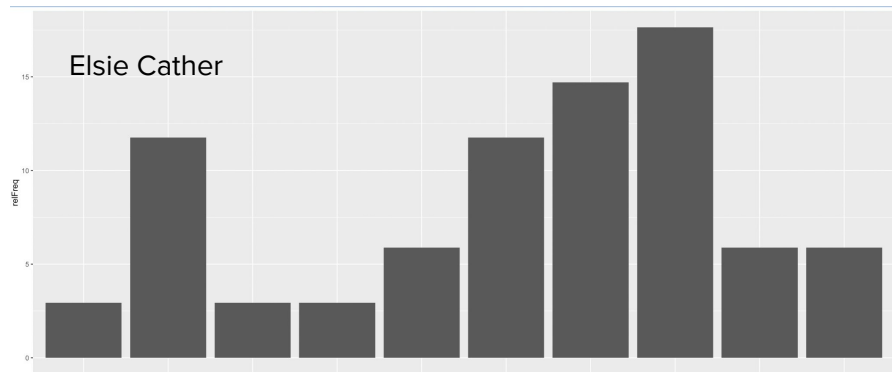


# Big Picture



# Big Picture

	psn	number_of_letters	name	frequency	relFreq
6	psn:0401	262	GreensletFerris	262	100.00000
7	psn:0518	170	KnopfAlfred A.	158	92.94118
5	psn:0311	96	FisherCanfieldDorothy	79	82.29167
9	psn:0842	78	SergeantElizabeth Shepley	52	66.66667
3	psn:0198	41	CatherBoakMary Virginia Jennie	27	65.85366
8	psn:0520	91	KnopfWolfBlanche	57	62.63736
1	psn:0016	101	AkinsZoë	50	49.50495
2	psn:0187	86	CatherElsie	34	39.53488
4	psn:0200	182	CatherRoscoe	62	34.06593
10	psn:0849	106	SherwoodMinerCarrie	35	33.01887



# Big Picture

recipient	name	letters	dif_people	relFreq
psn:0369	GereMariel	34	135	3.9705882
psn:0198	CatherBoakMary Virginia Jennie	41	126	3.0731707
psn:0187	CatherElsie	86	189	2.1976744
psn:0311	FisherCanfieldDorothy	96	185	1.9270833
psn:0849	SherwoodMinerCarrie	106	197	1.8584906
psn:0842	SergeantElizabeth Shepley	78	135	1.7307692
psn:0200	CatherRoscoe	182	272	1.4945055
psn:0016	AkinsZoë	101	114	1.1287129
psn:0401	GreensletFerris	262	259	0.9885496
psn:0518	KnopfAlfred A.	170	141	0.8294118

recipient	name	letters	Edith_mentions	relFreq
psn:0585	LuhanDodgeGansonMabel	20	18	90.000000
psn:0650	MenuhinYaltah	16	8	50.000000
psn:0198	CatherBoakMary Virginia Jennie	41	20	48.780488
psn:0044	MellenAuldMary Virginia	9	4	44.444444
psn:0197	ShannonCatherMargaret	32	14	43.750000
psn:0182	CatherCharles DouglasDouglass	7	3	42.857143
psn:0187	CatherElsie	86	35	40.697674
psn:0052	AustermannGeorge	10	4	40.000000
psn:0872	SouthwickCatherHelen Louise	13	5	38.461538
psn:0199	CatherSchaperMeta	39	13	33.333333

# Networks

[https://gkirilloff.github.io/catherworks\\_to\\_otherworks.html](https://gkirilloff.github.io/catherworks_to_otherworks.html)

[https://gkirilloff.github.io/recipients\\_to\\_people.html](https://gkirilloff.github.io/recipients_to_people.html)



HOW GOOD some of those old traveling companies were, and how honestly they did their work and tried to put on a creditable performance. There was the Andrews Opera company, for example; they usually had a good voice or two among them, a small orchestra and a painstaking **conductor, who was also the pianist**. What good luck for a country child to hear those tuneful old operas . . .

(Willa Cather to Harvey E. Newbranch, October 27, 1929)

Mr. Gayheart set off through the snow flurry, a daughter on either arm. He liked to reach the Opera House early and watch the people come in. (The theatre in every little Western town was then called an opera house.) . . . When the **conductor, who was also the pianist**, appeared, Mr. Gayheart settled back with satisfaction, and the curtain rose on the hunting scene.

(Lucy Gayheart, 1934)

---

[W]hy do you so scorn Loti? If anyone told me they saw a trace of Loti in anything I had done, I'd swoon with joy. And here you are indignant at such a suggestion. I don't know the new one you read but "Le Pecheur d'Island," "Le Desert," "Le Roman d'unspahi" and a dozen more are surely among **the most beautiful things in the world**. They stir one up like music and steal your senses away in the same fashion.

(Willa Cather to Harrison G. Dwight, January 12, [1907])

They're going to let me sing Sieglinde . . . A thousand things may happen between this and Friday to put me out. Something will happen. If that part were sung well, as well as it ought to be, it would be one of **the most beautiful things in the world**. That's why it never is sung right, and never will be." She clenched her hands and opened them despairingly, looking out of the open window. "It's inaccessibly beautiful!" she brought out sharply.

(*Song of the Lark* 1915)

---

Time

# Goals

- To capture an important aspect of how Cather scholars read her letters (by anchoring to the letter's known date and making connections to events that are described)
- To visualize how many days, weeks, and months are mentioned in a letter from a different day
- To visualize how many absences there are among these days, weeks, and months
- To bring readers closer to the inference, conjecture, and guess work in biography (that is often described by scholars as fact)

# Process So Far: Initial Round of Qualitative Analysis

- What kinds of words and phrases related to the past and future does Cather tend to use?

Reference	Count
day	1145
yesterday	251
the day	157
every day	141
a day	140
some day	78
all day	72
telegrams and day letters	47

Reference	Count
week	787
a week	290
next week	137
last week	119
for a week	88
this week	61

Reference	Count
morning	443
this morning	134
the morning	88
tomorrow	207
evening	176
in the morning	59

# Process So Far: Qualitative Analysis Continued

- To what degree are these phrases subject to interpretation?
- In how many references can we identify the following:
  - A unit of time (day, week, month)
  - A numerical quantity (one, more than one, a specific number)
  - A temporal direction (past or future)
- See [https://www.novetta.com/2020/04/dates\\_with\\_nlp/](https://www.novetta.com/2020/04/dates_with_nlp/) for a discussion of the ideas behind this approach

## Some Examples

- **Clear Referent:** “A week ago today I and my family lost the kindest of friends” (Cather to Mary Miner Creighton, [August 14, 1934])
- **More Abstract:** “I some how need your counsel and encouragement more than ever I needed it before. But since I must fight it out alone a few weeks longer, don’t *please* think unkindly of me.” (Cather to Dorothy Canfield, [March 28, 1903])
- **Letterhead:** “The filing time as shown in the date line on full rate telegrams and day letters, and the time of receipt at destination as shown on all messages, is STANDARD TIME.” (Cather to Alfred A. Knopf, June 3, 1926)

**Note:** We can parse the XML to exclude letterhead text, but it’s included here to show the need for care and attention

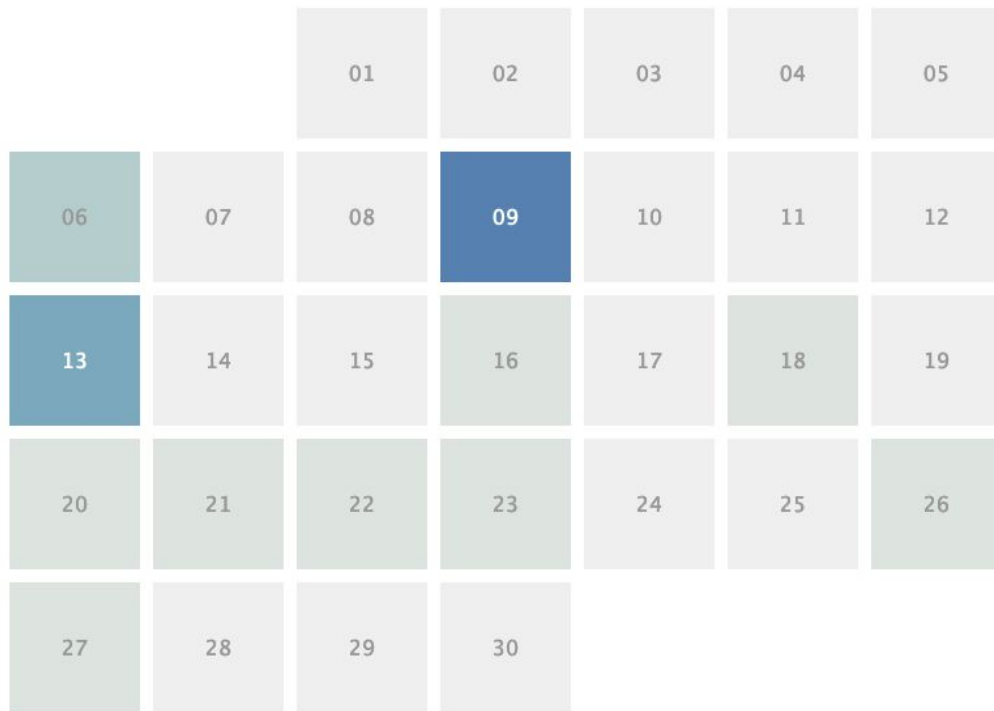
# Process So Far: Natural Language Processing

- Rule-based NLP
  - For example, look for all occurrences of “tomorrow”
  - Code as one day from letter date
  - If preceded by the words “day after” code as two days from letter date
- Named Entity Recognition (using a pre-trained machine learning classifier)
  - Uses training data to learn patterns
  - Predicts words and phrases categorized as TIME, DATE, PERSON, ORG, MONEY, EVENT, etc.



# Time

- Visualize letter dates and date mentions as a heatmap calendar
- Click on a date to view a list of letters and mentions
- Zoom controls
- Filter by date type, letter text, recipient, letters with “not before” and “not after” dates?



# Geography

# Maps

<https://gkirilloff.github.io/catherDataMappingResults.html>

# Questions

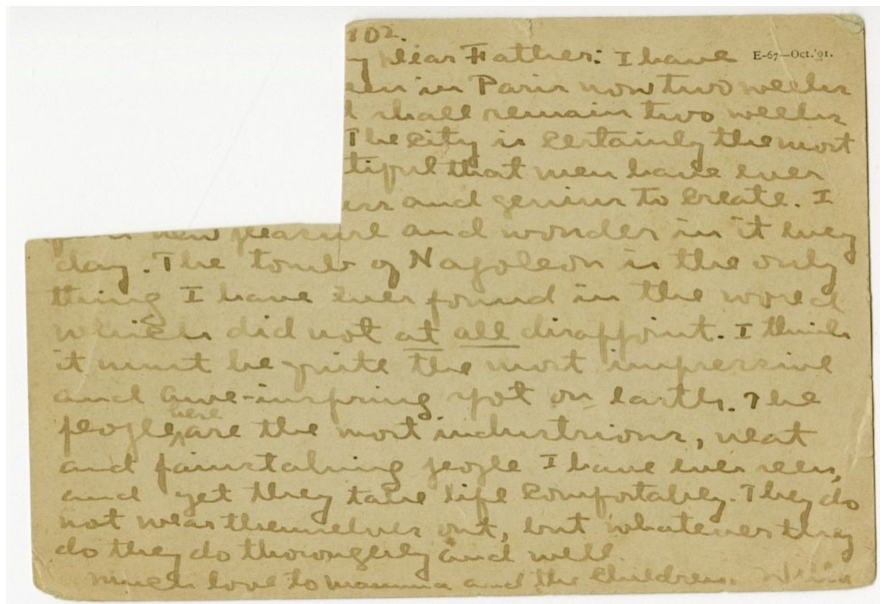
## Questions for the Audience

- Most of what we accomplished seems to be coming up with different ways for users to view the letters. As researchers, what would you like to see or explore in relation to the topics we have covered?
- Given that some of this work is incorporating sentiment analysis, what sentiments do you typically associate with Cather's letters and writing?

# Networks

I always do the marketing and all the housekeeping when I am here. I like it, but it does take time, and now I've got behind with my book [My Antonia] and am getting nervous. I'm going to cut out parties for the first two months of the New year.  
(Cather to Mary Virginia Boak Cather, December 30, 1916)

# Small Moments



people_name	recipient	year
mama	psn:0897	1889
mamma	psn:0369	1893
mamma	psn:0369	1893
mamma	psn:0367 psn:0368	1895
mamma	psn:0369	1896
mother	psn:0369	1896
mother	psn:0840 psn:0841	1897
mother	psn:0369	1897
my mother	psn:0355	1898
mother	psn:0355	1898
mother	psn:0369	1899
mother	psn:0493	1900
mother	psn:0370	1901
mother	psn:0840	1901
mamma	psn:0180	1902
mother	psn:0311	1903
mother	psn:0180	1906
mother	psn:0849	1906
mother	psn:0191	1910