cogs 105 this week

The FOUR V's of Big Data

BIGDATA

today: LIWC

Talking
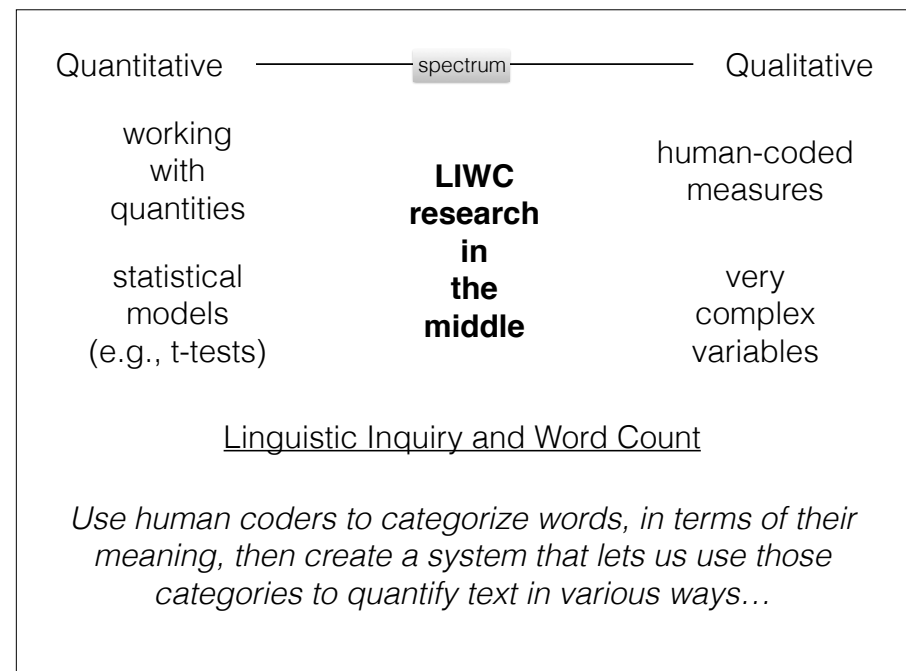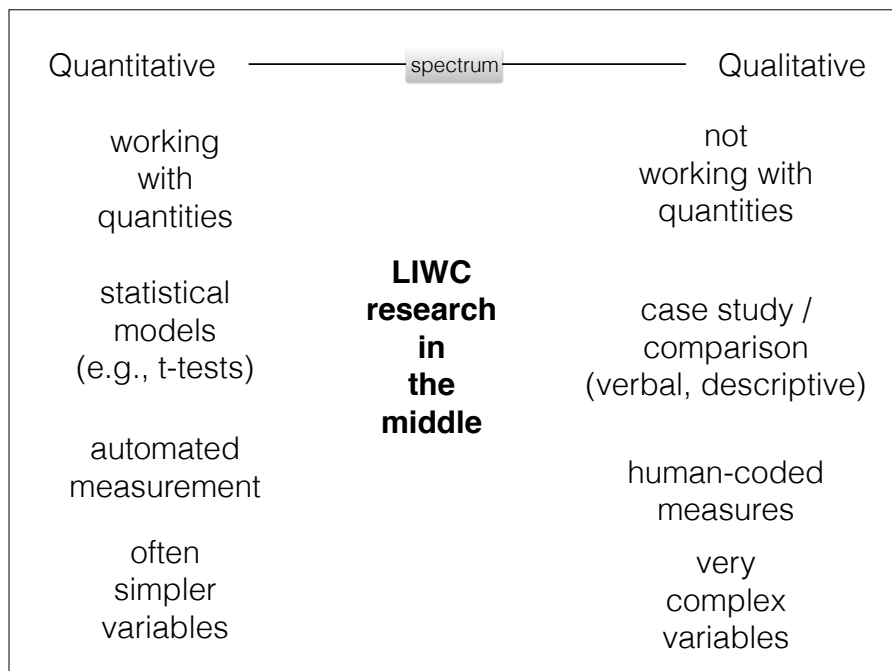
semantics
discourse

body
voice

Today: LIWC

- Last time: LSA, a model of meaning

  - Purely **quantitative**

- Linguistic Inquiry and Word Count (LIWC)

  - A measure of meaning that **mixes qualitative and quantitative**

- Motivation: You are "linguistically leaky" — subtle patterns of word usage may reveal your intentions, emotions, desires, etc.

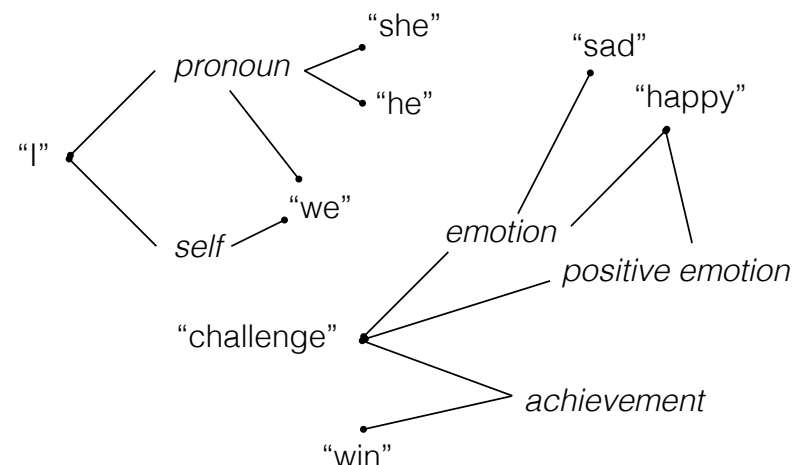# Types of Research

- Qualitative vs. … quantitative

observe, annotate

Slide 1

Quantitative ———— spectrum ———— Qualitative

working
with
quantities

not
working with
quantities

**LIWC
research
in
the
middle**

statistical
models
(e.g., t-tests)

case study /
comparison
(verbal, descriptive)

automated
measurement

human-coded
measures

often
simpler
variables

very
complex
variables

Slide 2

Quantitative ———— spectrum ———— Qualitative

working
with
quantities

human-coded
measures

**LIWC
research
in
the
middle**

statistical
models
(e.g., t-tests)

very
complex
variables

Linguistic Inquiry and Word Count

*Use human coders to categorize words, in terms of their
meaning, then create a system that lets us use those
categories to quantify text in various ways…*

Slide 3

Examples

- What the word "I" means can be described in terms of the categories in which this word can be placed. E.g.: *pronoun*, *self*

- How about the word "challenge": *emotion*, *positive emotion*, *achievement*

- Each word can be in many categories; each category has many words.

- LIWC now has almost 5,000 words!

Slide 4

Examples

"I"

pronoun — "she"
pronoun — "he"

self — "we"

"challenge"

emotion
positive emotion
achievement

"sad"
"happy"

"win"

LIWC Is, in Part, a Dictionary

Word	Category
I	pronoun, self
we	pronoun, self
she	pronoun, self, female
challenge	emotion, positive emotion, achievement
win	achievement, etc.
happy	emotion, positive emotion
see	senses, vision, etc.

Qualitative Methods

- LIWC used qualitative coding by human participants in order to get these dictionaries…

- Human judges (about 5) worked together to help (i) pick the words, (ii) pick the categories, and (ii) **categorize the words**.

- Judges achieved inter-rater reliability:

  - The process of comparing judge categorizations to determine how closely judges agreed. Where they do not have a majority vote, they either removed a word or negotiated to come to agreement. Ultimately, LIWC's scores are based on over 90% agreement by judges.
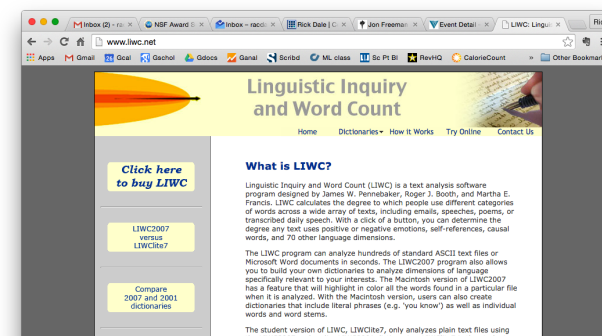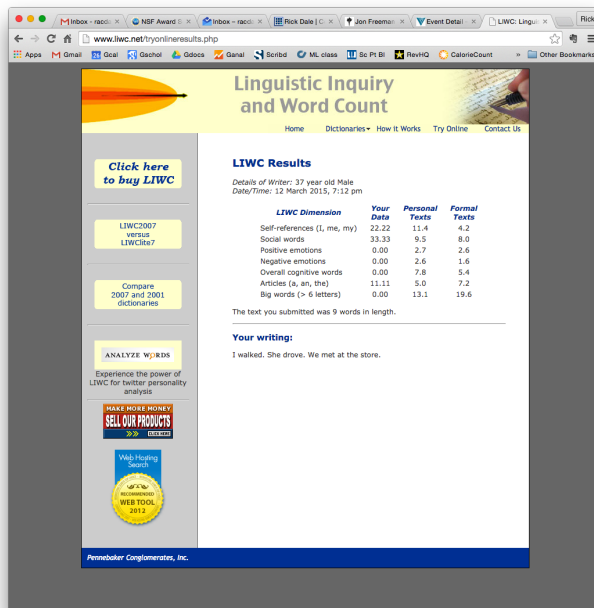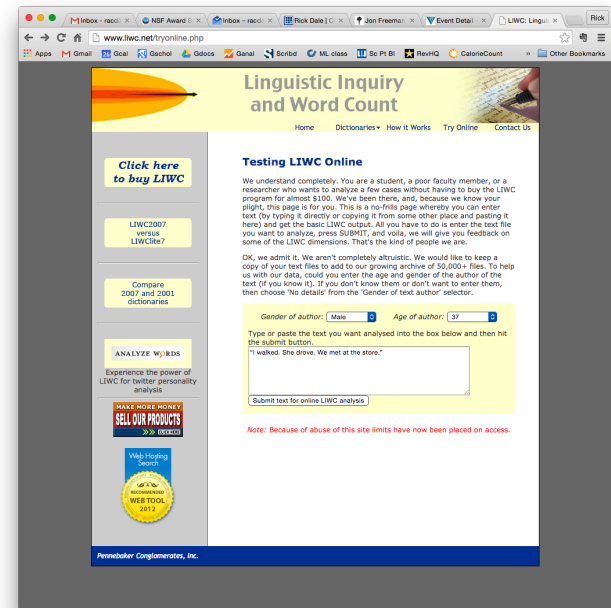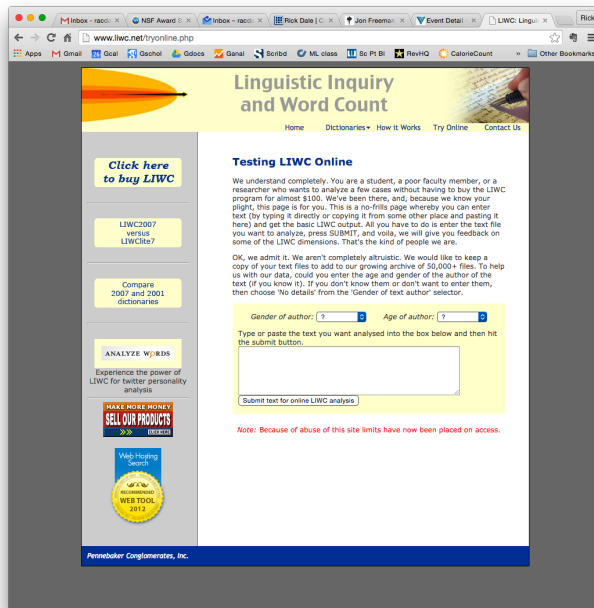
Cognitive Questions…

- After LIWC has **hand-coded** these categories, we can now use them to **quantify** text.

- Questions we might ask:

  - How much do people refer to themselves?

  - How sad are people? How negative?

  - How much do they talk about their job?

  - How often do they make social references?

LIWC Is Quite Transparent

"I walked. She drove. We met at the store."

9 words.
3 pronouns (33%)
2 self words (22%)

Linguistic Inquiry and Word Count

Home   Dictionaries▾   How It Works   Try Online   Contact Us

LIWC Results

Details of Writer: 37 year old Male
Date/Time: 12 March 2015, 7:12 pm

LIWC Dimension	Your Data	Personal Texts	Formal Texts
Self-references (I, me, my)	22.22	11.4	4.2
Social words	33.33	9.5	8.0
Positive emotions	0.00	2.7	2.6
Negative emotions	0.00	2.6	1.6
Overall cognitive words	0.00	7.8	5.4
Articles (a, an, the)	11.11	5.0	7.2
Big words (> 6 letters)	0.00	13.1	19.6

The text you submitted was 9 words in length.

Your writing:

I walked. She drove. We met at the store.

Commencement 2015

UCM Provost EVC <provostevc@ucmerced.edu>   Mar 3 (9 days ago)
to

Dear members of the UC Merced Faculty:

As you know, Commencement is our most visible public recognition of the accomplishments of our students.   As an educator and mentor at UC Merced, you have played a vital role in the success of our students. Your personal participation in the Commencement ceremony is very important to your students.

This year we will be hosting two school-based ceremonies, with the School of Engineering and the School of Natural Sciences ceremony on Saturday, May 16, and the School of Social Sciences, Arts and Humanities on Sunday, May 17.
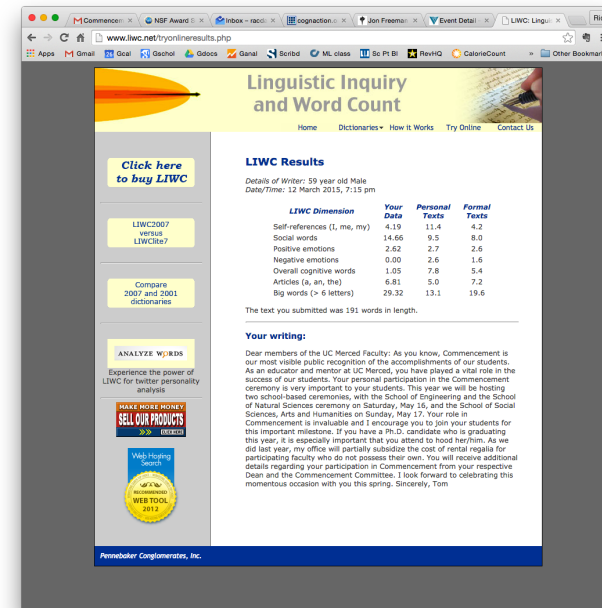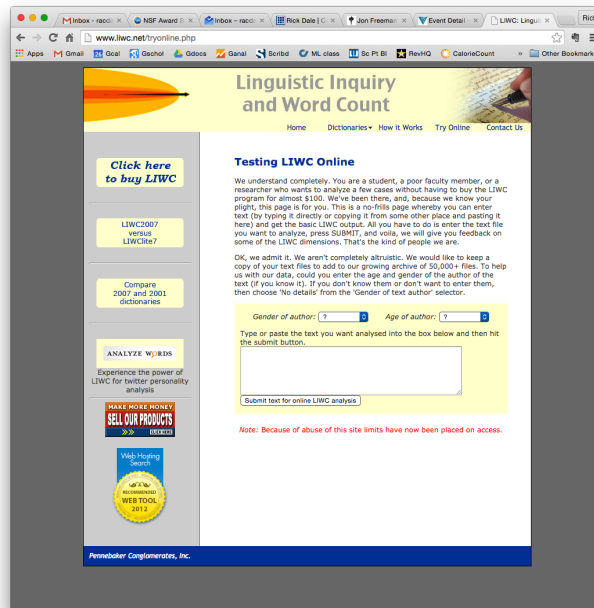
Your role in Commencement is invaluable and I encourage you to join your students for this important milestone. If you have a Ph.D. candidate who is graduating this year, it is especially important that you attend to hood her/him. As we did last year, my office will partially subsidize the cost of rental regalia for participating faculty who do not possess their own.

You will receive additional details regarding your participation in Commencement from your respective Dean and the Commencement Committee. I look forward to celebrating this momentous occasion with you this spring.

Sincerely,

Tom

Thomas W. Peterson
Provost and Executive Vice Chancellor
UNIVERSITY OF CALIFORNIA, MERCED
5200 North Lake Road, Merced, CA 95343
209-228-4439 | twpeterson@ucmerced.edu

LIWC as "Forensic Linguistics"

- LIWC has been applied all over the place. The reading provides some very nice details…

- The idea is that subtle linguistic usage, especially such things as stylistics (pronouns, prepositions, etc.) might serve as a kind of "linguistic forensic" device, that could detect some psychological states.

Cognitive Question

- How much do people refer to themselves?

  - Turns out, liars tend to use language in a way to minimize first-person singular.

    - LIWC has been used to quantify text that contains lies. Fewer "I"s!

  - Also find that depressives use more first-person singular.

Cognitive Questions

- How sad are people? How negative?

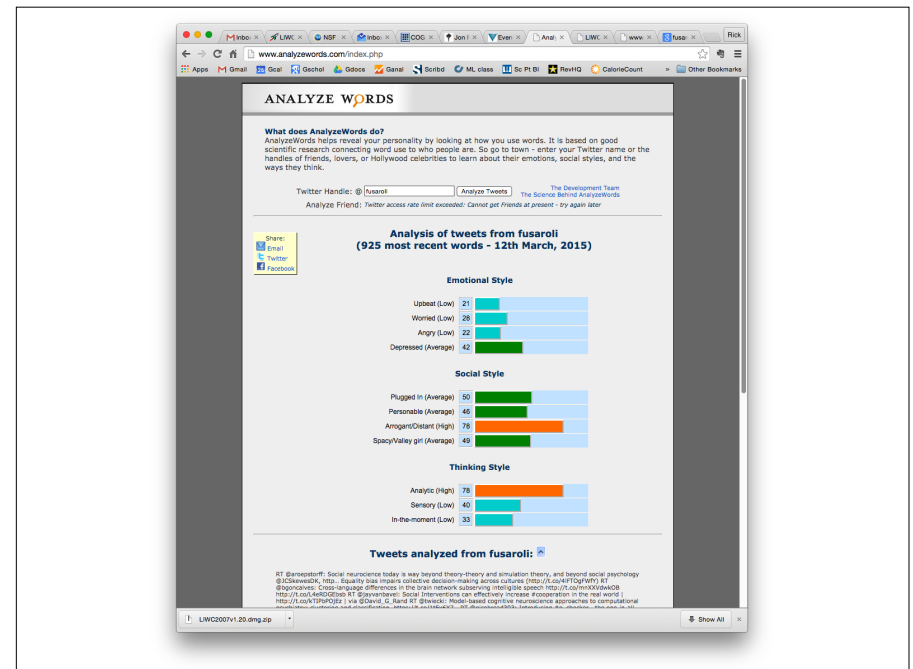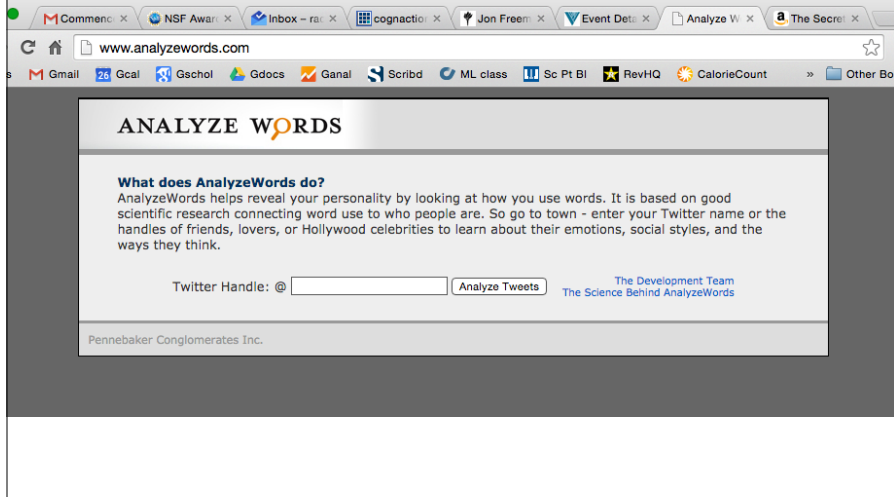 - Turns out, participants who are more **neurotic** tend to have correlated **negative word** usage.

ranging between 0.10 and 0.16. Overall, neuroticism was positively correlated with use of negative emotion words and negatively with positive emotion words; extraversion correlated positively with positive emotion words and words indicative of social processes; agreeableness was positively related to positive emotion and negatively to negative emotion words. In addition, neuroticism was characterized by a more frequent use of first person singular, a finding that is consistent with the idea that excessive use of first person pronouns reflects a high degree of self-involvement (e.g., Davis & Brock 1975, Ickes et al. 1986, Scherwitz & Canick

---

Cognitive Questions

consistent changes in their linguistic styles. With increasing age, individuals used more positive emotion words, fewer negative emotion words, fewer first person singular self-references, more future tense, and fewer past tense verbs. Age was also positively correlated with an increase in cognitive complexity (e.g., causation words, insight words, long words). In addition to challenging some of the cultural stereotypes on aging, these results suggest that language use can serve as a subtle linguistic age marker.

LIWC could serve as a kind of linguistic diagnostic tool for different stages in life…

---

http://www.analyzewords.com/

Grades

- Final paper
  - Our coverage of diverse material is partly designed to help expose you to many techniques.
  - In the final paper, you will choose one method, and expand a lab into a paper project.
  - Choose something that is related to **career interests**.
    - E.g., philosophy grad school - build a thought experiment?
    - E.g., forensic psychology - LIWC?
    - E.g., CogSci grad school - RT experiment?
    - E.g., cog neuro interests - Neurosynth exploration?

Sections

- Going to lab ensures you understand the task — you can even check to make sure it is correct before you submit it to your TA.

  - You can ask *any questions* about the task, and make sure you understand the procedures.

  - Ask your TA about careers in cognitive sc

- Don't forget extra SONA credit.

Exam 2 / Next Week

- Date for next exam: **March 31st**, after spring break.

  - Review guide posted well before spring break.

- One more section before exam two: **neural network models**

  - From computer analysis in big data to computational modeling using neural network