Introduction to Writing Analytics: Fundamental Principles and Data Exploration

Jill Burstein

Duolingo

T3 @ LASI21

June 21-25, 2021

duolingo

Tutorial Goals

- Participants walk away with a fundamental understanding of writing analytics
- Participants have sufficient knowledge to develop writing analytics research questions to pursue research projects
- Participants have an interest to continue to develop skills to conduct research

duolingo

Tutorial Outline

- **What is Writing Analytics? (15 minutes)**
  Working Definition & Broad Implications
  Writing Frameworks & Cross-Disciplinary Interaction
  Research Workflow

- **Example Writing Analytics Research Projects (45 minutes)**
  Writing Mentor -- Google Docs Add-on for writing support
  IES WAVES -- US Department of Education, IES-funded research project

- **Data Exploration Activity: What can we explore/learn, and how can writing analytics contribute to education technology innovation? (20 minutes)**
  Review Writing Mentor Process and Product feature data
  Review `IES-WAVES` Data: Attitudes, Genre & Survival
  Example Explorations/Visualizations

duolingo

Schedule

- What is Writing Analytics? (15 minutes)
- Example Writing Analytics Research Projects (40 minutes)
- Break (5 minutes)
- Data Exploration Activity: What can we explore/learn, and how can writing analytics contribute to education technology innovation? (20 minutes)

duolingo

What is Writing Analytics?

Working Definition, Writing Framework/Cross-Disciplinary Interaction, and Data Considerations & Research Workflow

duolingo

Working Definition & Broad Implications

Definition: *Writing analytics* focuses on the collection, measurement and analysis of student writing to investigate patterns in *writing process* (e.g., timestamp) and *product* (e.g., essays) data to provide insights about students' writing achievement at a specific point in time, or over a period of time.

Broad Implications:

- *Yield insights* that may support students, and inform improvements in educational settings for writing, such as writing instruction curriculum, automated personalized instruction.
- Insights may inform workplace success factors outside of educational contexts.

Writing Framework & Cross-Disciplinary Interaction

duolingo

Socio-cognitive framework

Explore relationships between process and product feature data (drawn from student writing samples) and **socio-cognitive writing achievement framework components**.

- *Writing domain knowledge* (e.g., vocabulary choice)
- *General skills* (e.g., critical thinking, reading skill)
- Content knowledge
- *Intrapersonal* factors (e.g., motivation)
- *Interpersonal* factors (e.g., collaboration)
- Neurological factors (e.g., task attendance, vision)

Cross-Disciplinary Interactions

- Writing studies
- Disciplinary constructs (e.g., STEM content knowledge)
- Learning analytics
- Educational measurement
- Computer & Data science
- Natural language processing (NLP)
- Automated writing evaluation (AWE)

Research Workflow

Research Questions, Data Considerations, & Research Methods

duolingo

Motivation: What kinds research questions are we trying to answer?

- Is there a relationship between writing domain knowledge and student interest?
- Are there relationships between writing features and student outcomes, such as course grades?
- Are students using different kinds of language when they write in different academic genres (such as, persuasive vs informational writing)?

duolingo

Data Collection Resources: Where will I get my data?

- **Institutional Digital Resources**
  - Learning Management Systems (e.g.,
  - Canvas)
- **Commercial Writing-Support Applications**
  - Grammarly
  - Turnitin
  - Writing Mentor
- **Publicly-Available Data**
  - IES Writing Achievement Data (ETS Github)
  - Michigan Corpora of Upper Level Student Papers (MICUSP)
- **Student Institutional Data**
  - Student Background & Demographics
  - Enrollment, Grades, Test Scores

Data Considerations

- **Data Privacy**
  - Personally-identifiable information (or, PII)
  - Subject consent for data use
  - Data security
- **File Format**
  - Plain text
  - Text encodings (e.g., UTF-8)
- **Data Format**
  - .csv (e.g., comma-separated, structured data)
  - JSON
  - XML
- **Data Preparation**
  - "As is" (e.g., .csv cells)
  - Chunking/tokenizing: partitioning data into meaningful units (e.g., words, sentences) based on task (e.g., discourse units)
- **Data Feature Generation**
  - *Automated writing evaluation components
- **Data Storage**
  - Local machine (e.g., laptop)
  - Cloud storage (e.g., for massive data sets)

duolingo

Research Methods

- **Study Purpose**
  - *Descriptive*: describe writing characteristics
  - *Predictive*: predict outcomes
  - *Prescriptive*: inform educational processes or systems (e.g., writing instruction curriculum)
- **Method Types**
  - *Exploratory data mining*, e.g., event log data from writing apps
  - *Case studies*, e.g., data from small pilot studies
  - *Experimental methods*, such as quasi-experimental or randomized-control studies

Data Types

Writing Process & Product Data

duolingo

Data Types: Writing Process and Product Data

Writing Process Data

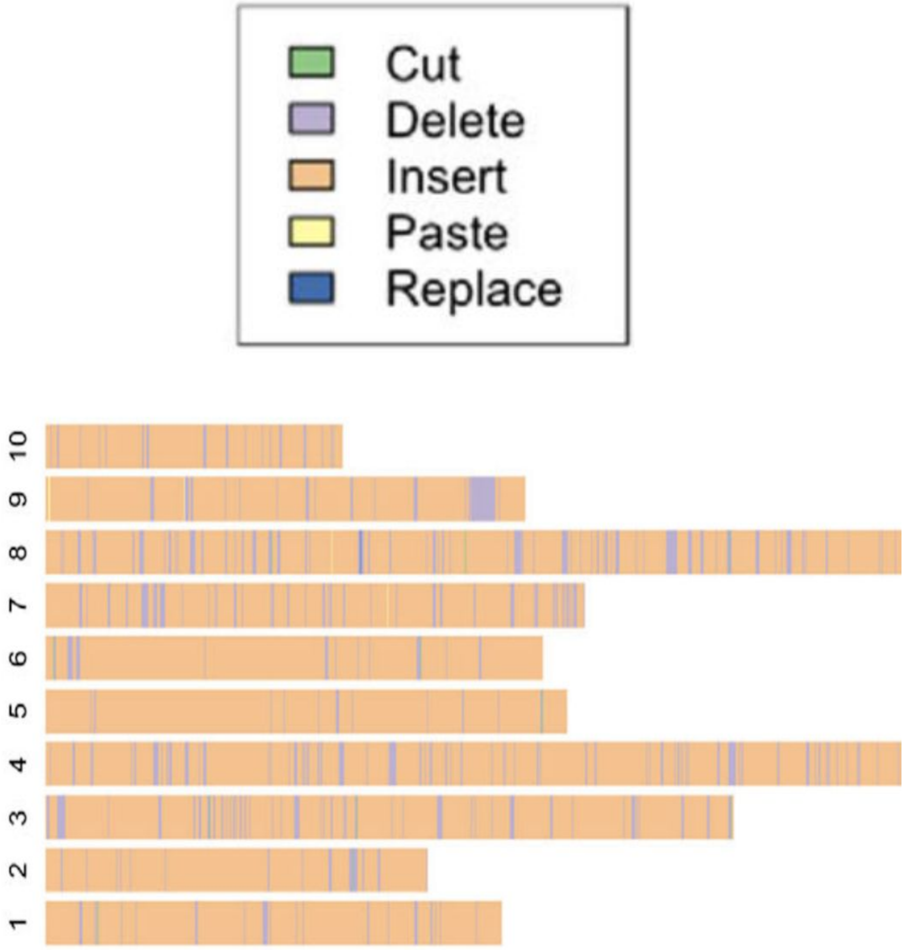
- Time-stamp data
- App navigation (e.g., feature use)
- Keystroke log data

Writing Product Data

- Writing assessment responses
- Coursework assignments (K-12 through postsecondary)
- Writing app submissions

duolingo

Keystroke log analysis records



(Zhu et al, 2019)

Journal Article Critique
Professor Abby Stern
Psych 101

Enhancing Interest and Performance With a Utility Value Intervention

A 2012 study conducted [12]by Hulleman, Godes, Hendricks, and Harackiewicz successfully proves how a utility value intervention can enhance interest and performance in many types of college students. A utility value, in this research, is about how how a task is “useful or relevant for other tasks or aspects of an individual’s life” (2). This research was conducted using two separate studies: studies 1 and 2 conducted in a laboratory and a classroom, respectively.

In study 1, 170 students from the University of Wisconsin, Madison (mostly over 90% Caucasian, but about equal in gender) were given simple math equations (two-digit multiplication) to solve using a new method they had just been taught by an audio demonstration given by the experimenters. The researchers’ hypothesis was that “participants in the relevance condition would be more interested in the math activity at the end of the session than those in the control condition. [They] also expected these effects would be moderated by participants’ performance expectation, such that students with low performance expectation would benefit more from the intervention than those with high performance expectations.” also, “utility value perceptions might also be associated with performance” (4).

The method was such that after “a measure of initial interest in math” (4) the participants were guided through the new method of completing the mathematical tasks as well as how the study would be conducted. After three minutes of practicing the mathematical technique on a problem

Writing Product Data

Automated Writing Evaluation (AWE)

duolingo

What is AWE Technology?

- AWE
  - **Natural language processing for automated detection of linguistic features in text**
    - -rule-based linguistic rules
    - -machine learning methods
    - Feedback generation for assessment
      - Scores
    - Feedback generation for instruction
      - Highlighting language features
    - Writing Analytics

duolingo

Automated Writing Evaluation Feature Types

- **Grammar, Word Usage & Mechanics Errors**
  - Subject-Verb Agreement: *the motel are* …
  - Confused Word: *affect vs. effect*
  - Spelling errors
- **Discourse Structure**
  - Thesis statements, main points, supporting details, conclusion statements
- **Discourse coherence**
  - Measures for flow of ideas
  - Transition terms usage
- **Sentence types**
  - Variety of sentence structures
- **Vocabulary usage**
  - Vocabulary sophistication
  - Metaphor vs literal vocabulary usage
  - Personal reflection vocabulary

duolingo

Example Writing Analytics Research Projects

Writing Mentor Google Docs Add-on for writing support (Zhu et al, 2021)

Writing Achievement Research (McCaffrey, Burstein et al, 2021)

Writing Mentor®: User Behavior Modeling

Writing Mentor video

duolingo

Research Questions

Can we identify user-in-the-wild types from WM patterns of use from a “proxy” middle/high school (MHS) data?

What relationships exist between WM MHS user types and changes in writing construct-relevant measures as captured by automated writing evaluation (AWE) features?

How does WM MHS user type performance differ with regard to writing quality (i.e., human rater scores)?

Are there user-in-the-wild type differences related to writing revision in WM?

duolingo

Writing data extraction & selection criteria

Middle/High School (MHS) Data Set (N=1,857)

- WM Extended Writing mode submissions (~2.5 years)
- Submission header required: *Mr*, *Mrs*, *Ms*, or *Miss*, (suggesting *teacher name*)

Selection Criterion for User Behavior Modeling (N=861)

- Evidence of at least one WM function in event log
- Document had at least two submissions
- Submission >= 10 words

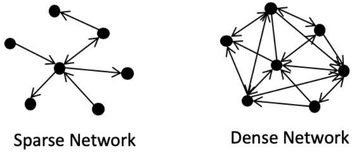
duolingo

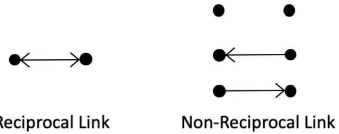
Modeling Features & Analysis Measures

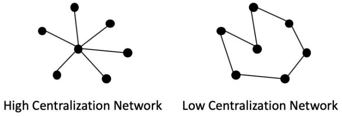
Modeling: WM Event Log Features & Social Network Analysis Modeling Features
Analysis: Automated Writing Evaluation Measures, Human Essay Scores, & Revision Measures

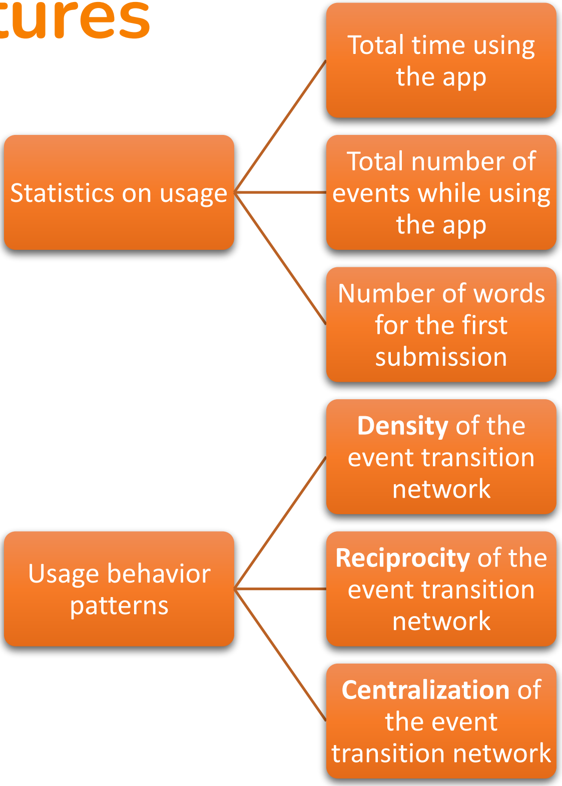
duolingo

User behavior modeling: Process & *social network analysis* features









Human Rater Essay Score Features

Data: 670 first & last "pairs" of MHS essay submissions

Human Raters: 4 raters assigned 3 trait scores to each essay

Rubric: Purpose (0-4), Development (0-4) & Conventions scores (0-3)

duolingo

Revision (Editing) Features

Revision Detection

- Identifies editing between pairs of submissions for a single document (e.g., *first-to-last*)
- Detects *Insertion*, *Deletion* & *Substitution* edit types

Revision Types

- Total # of edits (# of *insertions* + # of *deletions* + # of *substitutions*)
- Total # of *edit types*: (a) # of *insertions*, (b) # of *deletions*, & (c) # of *substitutions*
- Normalized *edit types:* (a) # of *insertions*/# of edits, (b) # of *deletions*/# of edits, & (c) # of *substitutions*/# of edits
- # of *edit types* with *one-word revision* (e.g., “teh” -> “the”)
- # of *edit types* with *multi-word revision* (e.g.,“The vase is…”– “A red vase is”)

duolingo

Findings

User Types, Patterns of Use, Writing Change, & Writing Proficiency

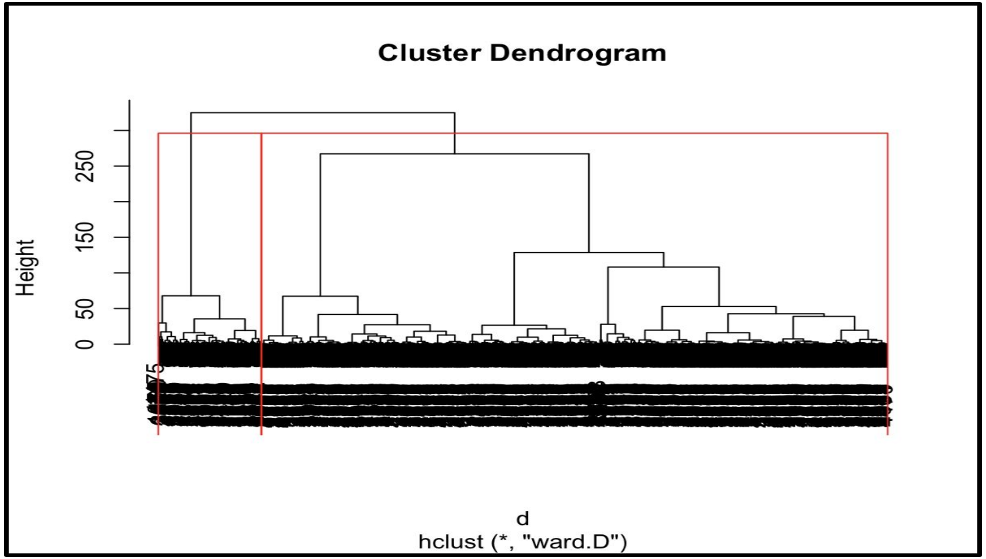
User Types

Can we identify user-in-the-wild types from WM patterns of use from a “proxy” middle/high school (MHS) data?

duolingo

Dendrogram from Hierarchical Clustering Analysis

- Hierarchical Clustering using *Ward's* minimum variance method
- Euclidean distance



CLUSTER	N	TOTAL TIME	EVENT COUNT	WORD COUNT	DENSITY	RECIPROCITY	CENTRALIZATION
Cluster 1 (*Average* User Document)	739	393.53	15.52	615.65	0.01	0.14	0.08
Cluster 2 (*Serious* User Document)	122	1575.66	49.70	553.40	0.02	0.22	0.1
Total/Average	861	561.03	20.37	606.83	0.01	0.15	0.09

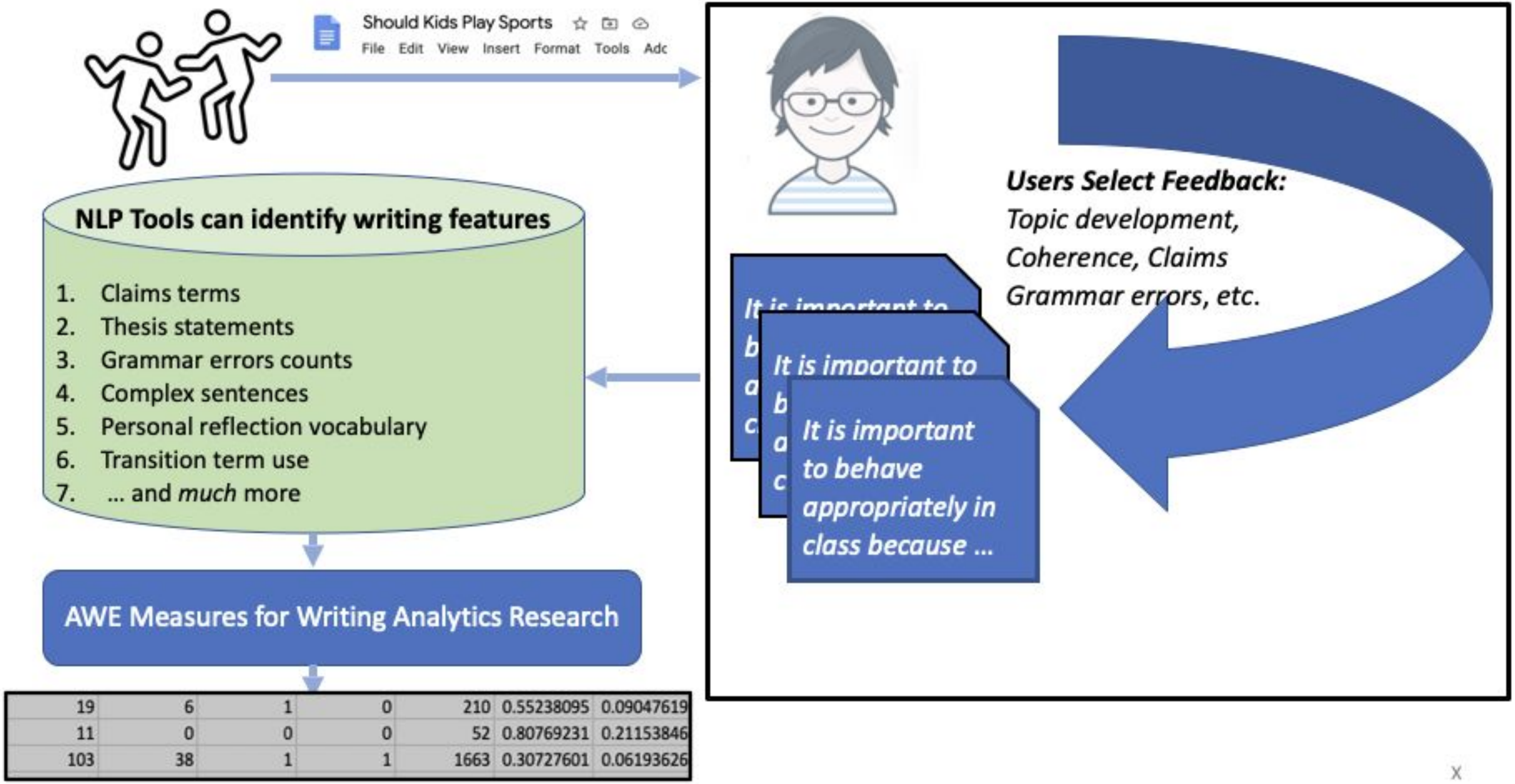
Cluster Comparisons for 6 Features Used in the Cluster Analysis

duolingo

User Types & Writing Change

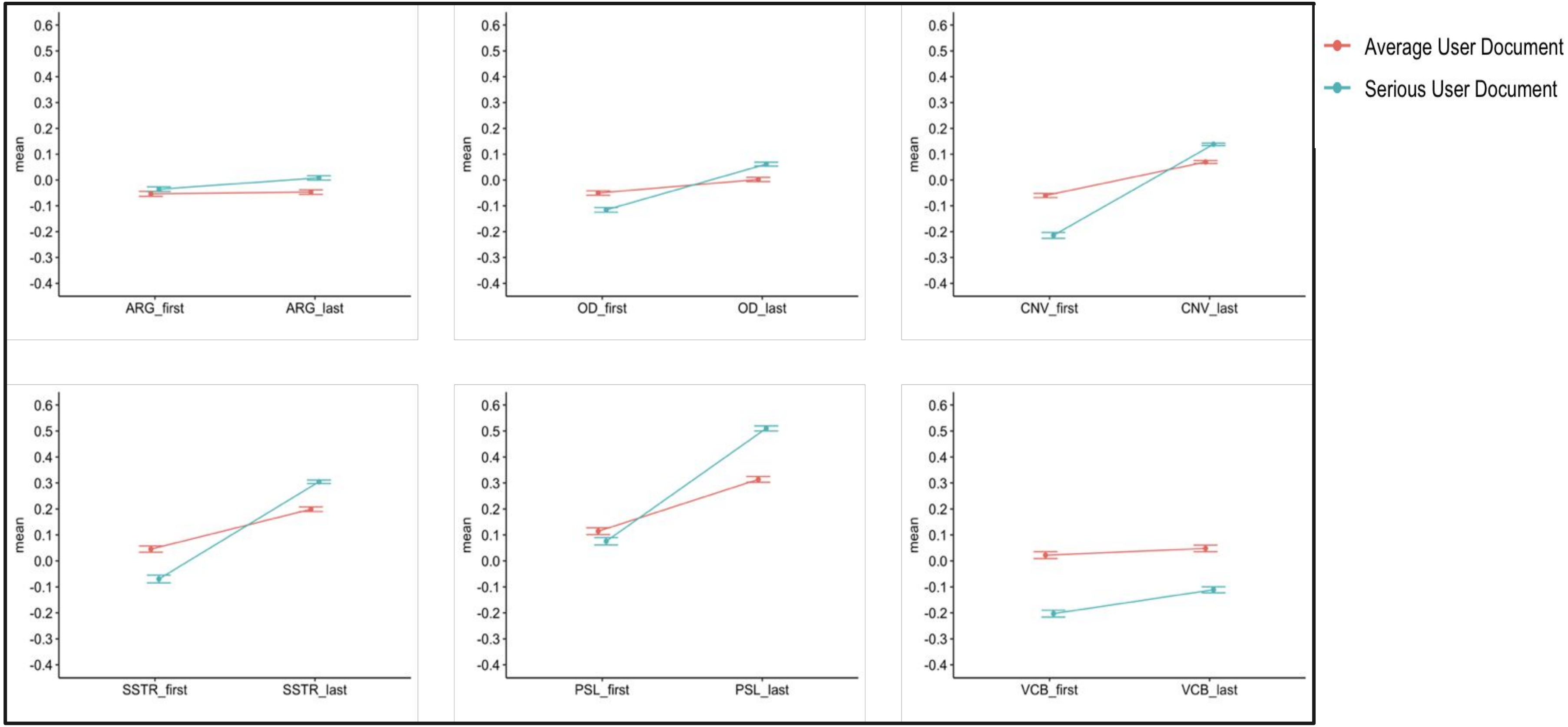
1. What relationships exist between WM MHS user types and changes in writing construct-relevant measures as captured by AWE features?
2. Are there user-in-the-wild type differences related to writing revision in WM?

duolingo


Should Kids Play Sports
File Edit View Insert Format Tools Adc
NLP Tools can identify writing features
1. Claims terms
2. Thesis statements
3. Grammar errors counts
4. Complex sentences
5. Personal reflection vocabulary
6. Transition term use
7. … and much more
AWE Measures for Writing Analytics Research
19 6 1 0 210 0.55238095 0.09047619
11 0 0 0 52 0.80769231 0.21153846
103 38 1 1 1663 0.30727601 0.06193626
Users Select Feedback:
Topic development,
Coherence, Claims
Grammar errors, etc.
It is important to behave appropriately in class because …

Automated writing evaluation features

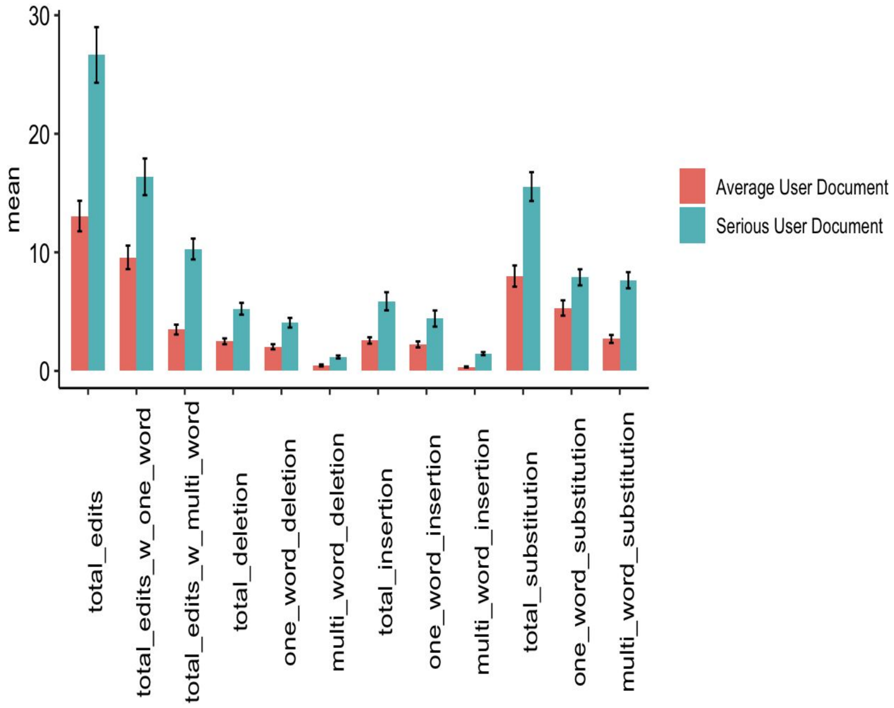
ARGUMENTATION	• Features related to argumentation, e.g., presence of claims and sources
ORGANIZATION & DEVELOPMENT	• Features related to organization & development, e.g., presence of discourse units (such as thesis), discourse coherence quality measure
ENGLISH CONVENTIONS	• Features related to knowledge of conventions, e.g., grammar errors, use of contractions
UTILITY VALUE	• Features related to personal reflection, e.g., use of 1st person pronouns
SENTENCE STRUCTURE	• Features related to phrasal and sentence structure, e.g., sentence variety, complex noun phrases
VOCABULARY	• Features related to vocabulary usage, e.g., metaphor use, rare words



Significant differences for *Serious Users* on the AWE features (at .05 level) were found for OD, CNV, SSTR, and PSL, but not for ARG and VCB.

duolingo

Revision & User Type



- **Measured by differences between *first* and *last* submissions**
- ***Serious* Users tend to make more changes, overall and across edit types**
- **Significant differences on the *edit type* features between two clusters for**
  - ***Total_edits* (*t*(50) = 2.29, *p* = .03)**
  - ***Total_deletion* (*t*(49) = 2.18, *p* = .03)**
  - ***Total_substitution* (*t*(55) = 2.40, *p* = .02)**

duolingo

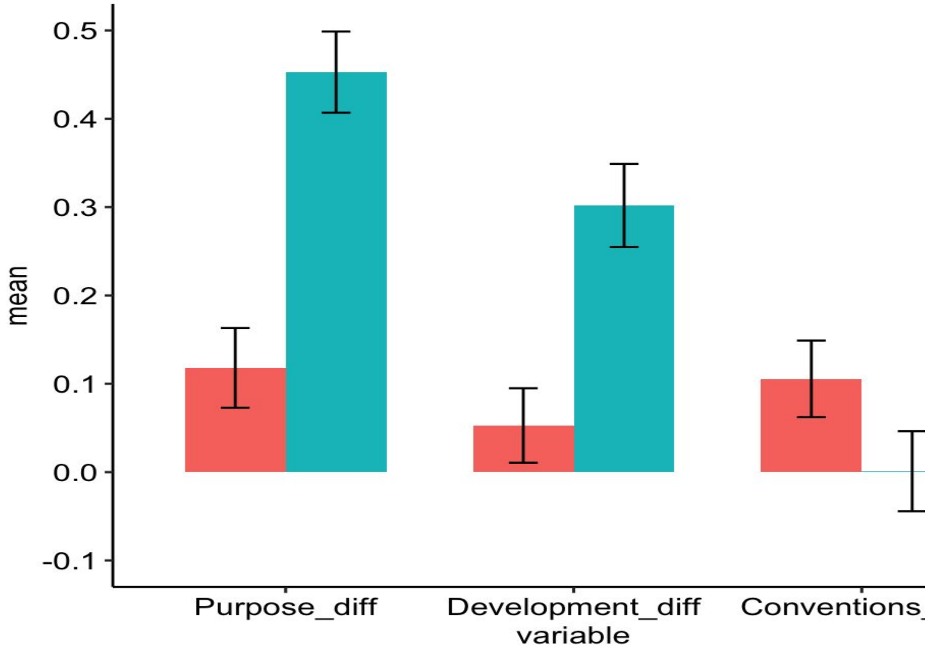
User Type & Writing Quality

How does WM MHS user type performance differ with regard to writing quality (i.e., human rater scores)?

duolingo

Trait Score Mean Differences Comparisons

- There were significant differences on the score differences between the two clusters for *Purpose* ($t(70) = 2.55$, $p = .01$), and marginally significant for *Development* ($t(66) = 1.87$, $p = .06$); but not for *Conventions* ($t(69) = 0.82$, $p = .41$).
- # of *Average* User Document = 322 (red)
- # of *Serious* User Document = 53 (blue)
- Total N users= 375



duolingo

What did we learn about middle/high school Writing Mentor users from the writing analytics study?

- **Two user types: *Serious* & *Average***
- **User Types & Writing Quality**
  - User types associated with differential gains in writing quality – i.e., *Serious* Users associated with greater gains
- **User Types & Writing Revision**
  - Differences in how *Serious* Users navigate and revise in WM as compared to *Average* Users.
- **Implications for system development & personalized learning for writing**

duolingo

IES-funded Exploratory Writing Achievement & College Retention

Student Writing, Writing Attitudes & Broader Success Outcomes

Motivation

- Low U.S. 4-year college retention is a national concern
- Across US institutions, ~65% of first-time students in 2012 graduated in 6 years
- Writing skills are critical to college success
- Nearly 75% of grade 8 and 12 students are below *Proficient* on NAEP writing
- No literature explicitly examines relationship between writing domain skills and university retention

Goals

- Explore the relationship between writing skills and retention in university using AWE to assess writing
- Find writing skills that may be for analytics to identify students at risk for dropout
- Generate hypotheses for future studies about essential writing skills for postsecondary success

duolingo

Research Questions

What are the relationships between writing attitudes and AWE features, & college success factors?

What are the relationships between student writing data and indicators of college retention (survival)?

Intrapersonal factors & success

What are the relationships between writing attitudes and AWE features, & college success factors?

duolingo

Writing Attitudes & College Success

Data

- **Response data to a 50-item writing attitudes survey**
  - *Writing Goals (Mastery, Performance, Avoidance)*
  - *Writing Confidence*
  - *Writing Beliefs (Importance of Content & Conventions)*
  - *Writing Affect (Feelings)*
- **Institutional outcomes data: Cumulative GPA**

Participants

566 students from six 4-year institutions

duolingo

Writing Confidence & Success Indicators

Higher Cumulative GPAs are significantly correlated with writing confidence

- *p < 0.05
- **p < 0.01

Academic Performance Indicators	n	Confidence
Academic Measures		
Study Semester Cumulative GPA	563	.043
Study Semester Plus One Semester Cumulative GPA	526	.069
Study Semester Plus Two Semesters Cumulative GPA	**456**	**.138****[**]**
Study Semester Plus Three Semesters Cumulative GPA	**364**	**.117***[*]**
Study Semester Plus Four Semesters Cumulative GPA	215	.101

duolingo

AWE Features & Attitudes

Data

- 997 **coursework** writing assignments
- 366 students completed **standardized writing assessment**
- AWE Feature Composite Scores

Participants

418 students from six 4-year institutions

duolingo

Intercorrelations Between Writing Motivation Subconstructs & AWE Components

AWE Features	Mastery	Avoidance	Confidence	Content (Belief)	Conventions (Belief)
Argumentation	–.00	–.05	.03	.00	–.03
Organization and Development	–.02	–.07	.00	.00	–.07
Personal Reflection	.012	–.06	–.01	**.12***	–.023
Sentence Structure	–.04	.03	–.05	–.02	.07
Vocabulary	.00	**–.12***	**.15****	–.04	**–.12***
Convention	**.15****	**–.14***	**.18****	**.15****	**–.12***

*** $p < .05$.**

**** $p < .01$.**

duolingo

AWE & College Retention

What are the relationships between student writing data and indicators of college retention (survival)?

duolingo

AWE Features and College Retention

Data

- 997 coursework writing assignments
- 366 students completed standardized argumentative writing assessment
- AWE Feature Composite Scores
- Institutional data: Enrollment status between 3 - 5 semesters after study semester

Participants

418 students from six 4-year institutions

Survival Analysis

Analysis

Random effects Cox proportional hazards regression used to model dropout as a function of the AWE feature composite score

Results

Suggests *vocabulary usage* might identify students *at risk* of dropping out

- ***Utility-Value Language* feature** predicted increased dropout for coursework ($p < .05$) and assessment ($p < .10$) writing
- ***Vocabulary* feature** predicted decreased dropout ($p < .10$) in the standardized assessment

duolingo

What did we learn about relationships between writing achievement & college retention?

- **New Insights**
  - Student writing attitudes and writing characteristics
  - Student retention and writing characteristics
- **Using Insights**
  - *Writing Attitudes & Success*
    - Stakeholder understanding of student writing attitudes could inform individualized instruction and support for students early on
  - *AWE & Retention*
    - AWE integration into a learning management system could provide *not only* personalized learning feedback for writing, but also retention analytics for stakeholders

duolingo

Data Exploration Activities

What tools are available?

duolingo

Which tools do I need?

1) What are my **writing analytics research questions**?
2) What **kinds of data** do I have?
    a) Continuous numerical (number values, e.g., GPA)
    b) Categorical (names or classifications, e.g., undergraduate vs. rising_senior, novice vs. expert)
3) What **kinds of analyses** do I need to conduct?
4) How can I create **visualizations** to illustrate my findings?

duolingo

Example Data Exploration & Visualization Tools

No programming

- Microsoft Excel or Google Sheets -- Tools for numerical and categorical data analysis & visualizations
- TAACO - Cohesion Indices (Crossley et al, 2016)
- TAALES - Lexical Sophistication (Kyle & Crossley, 2015)

Programming

- Python libraries
  - Pandas
    - Open source data analysis and manipulation tool, built on top of the Python programming language.
  - Visualization libraries
    - Matplotlib, Seaborn, Plotly
- R language
- Jupyter Notebook (Python & R)
  - Interactive data science environment
  - Easy access through Anaconda

duolingo

IES writing achievement study data

https://github.com/EducationalTestingService/ies-writing-achievement-study-data

duolingo

Data Exploration Activity

signs	Inform/Explore	43
try	Inform/Explore	111
call	Inform/Explore	61
national	Inform/Explore	101
hotline	Inform/Explore	18
sloppy	Persuade	190
people	Persuade	2017
what	Persuade	925
is	Persuade	4949
your	Persuade	388

Research Questions:
What can we learn about vocabulary usage by writing genre? And, by subject area?

Data

1. Coursework writing assignment plain text documents
2. Human-assigned genre labels: a) inform/explore, b) persuade, c) reflect

Data Preparation

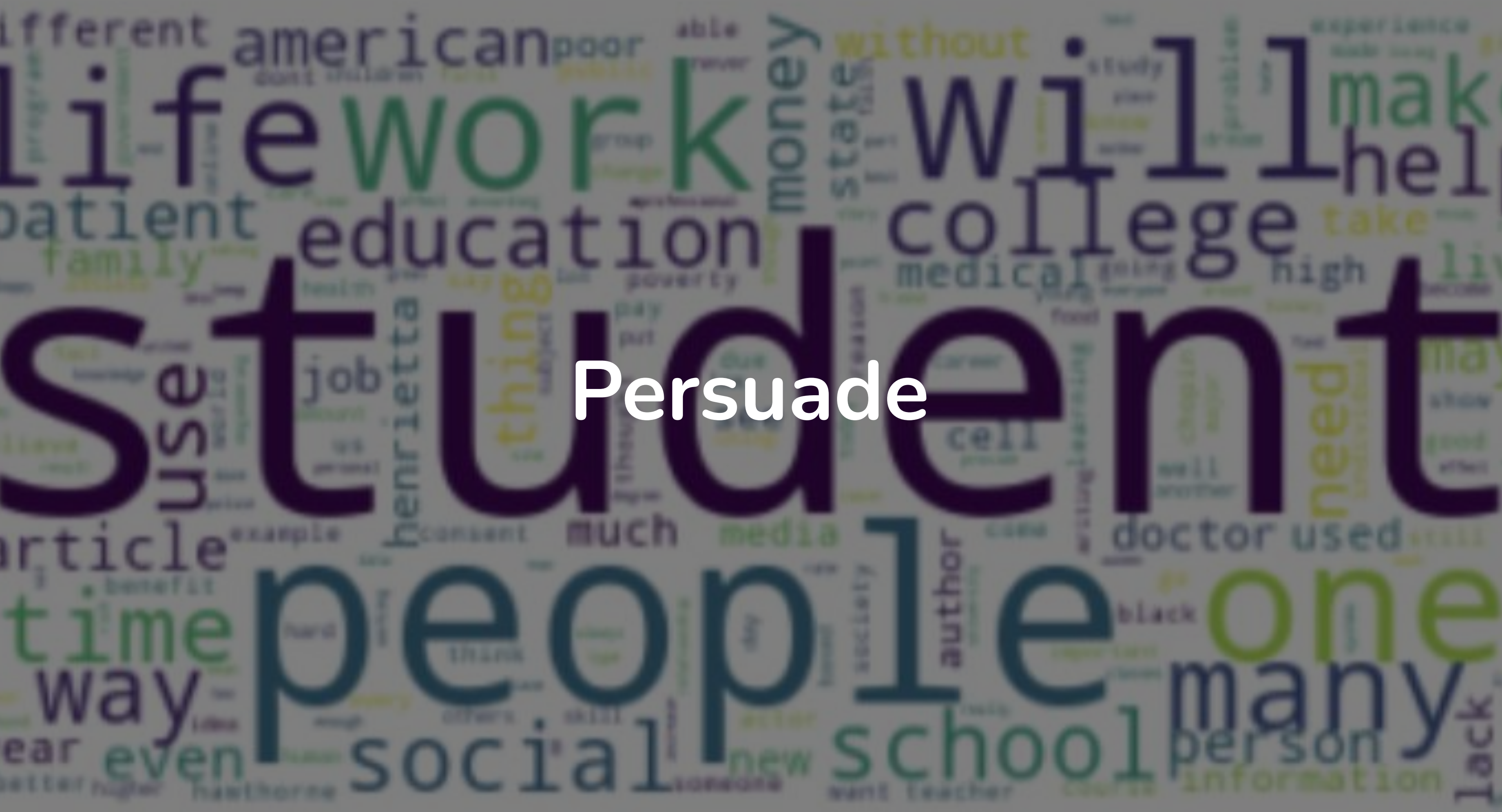
1. Word-tokenize text (identify all words)
2. For all documents in a genre, provide word counts in.csv for analysis
3. Create visualizations to examine word use by genre.

Example Jupyter Notebook for Analysis

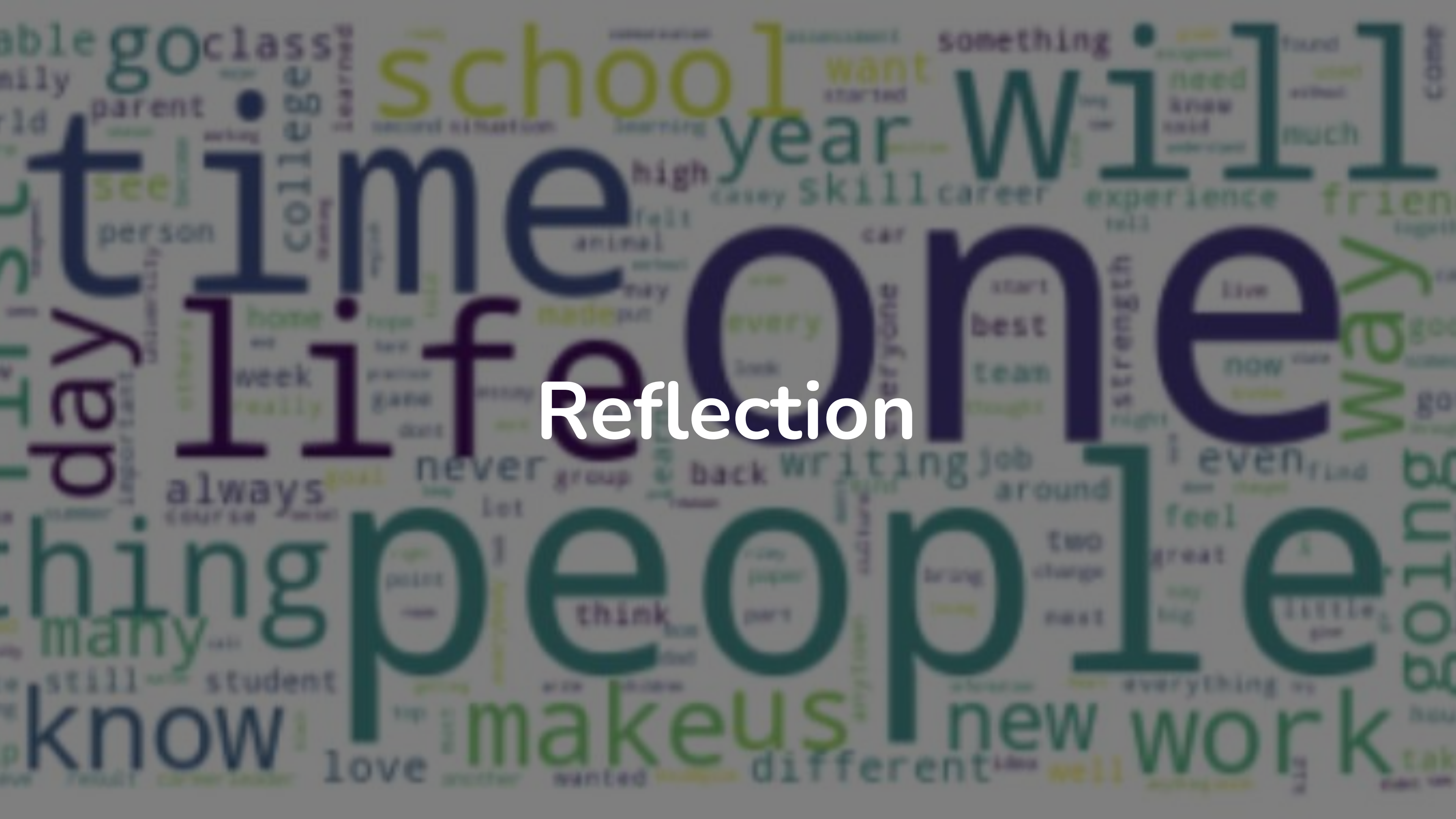
1. Open Anaconda to access Jupyter Notebook
2. Create New or Open Existing Notebooks

duolingo

Inform/Explore


student
people
will
work
life
college
education
american
patient
money
state
medical
henrietta
job
use
thing
need
much
doctor
used
media
article
time
way
one
many
school
social
person
author
information
even
lack

Persuade



Reflection

Inform/Explore: Biology

Inform/Explore: Business

Rough Insights

Writing in different genres might offer students opportunities to....

Writing in different subject areas might offer students opportunities to ...

duolingo

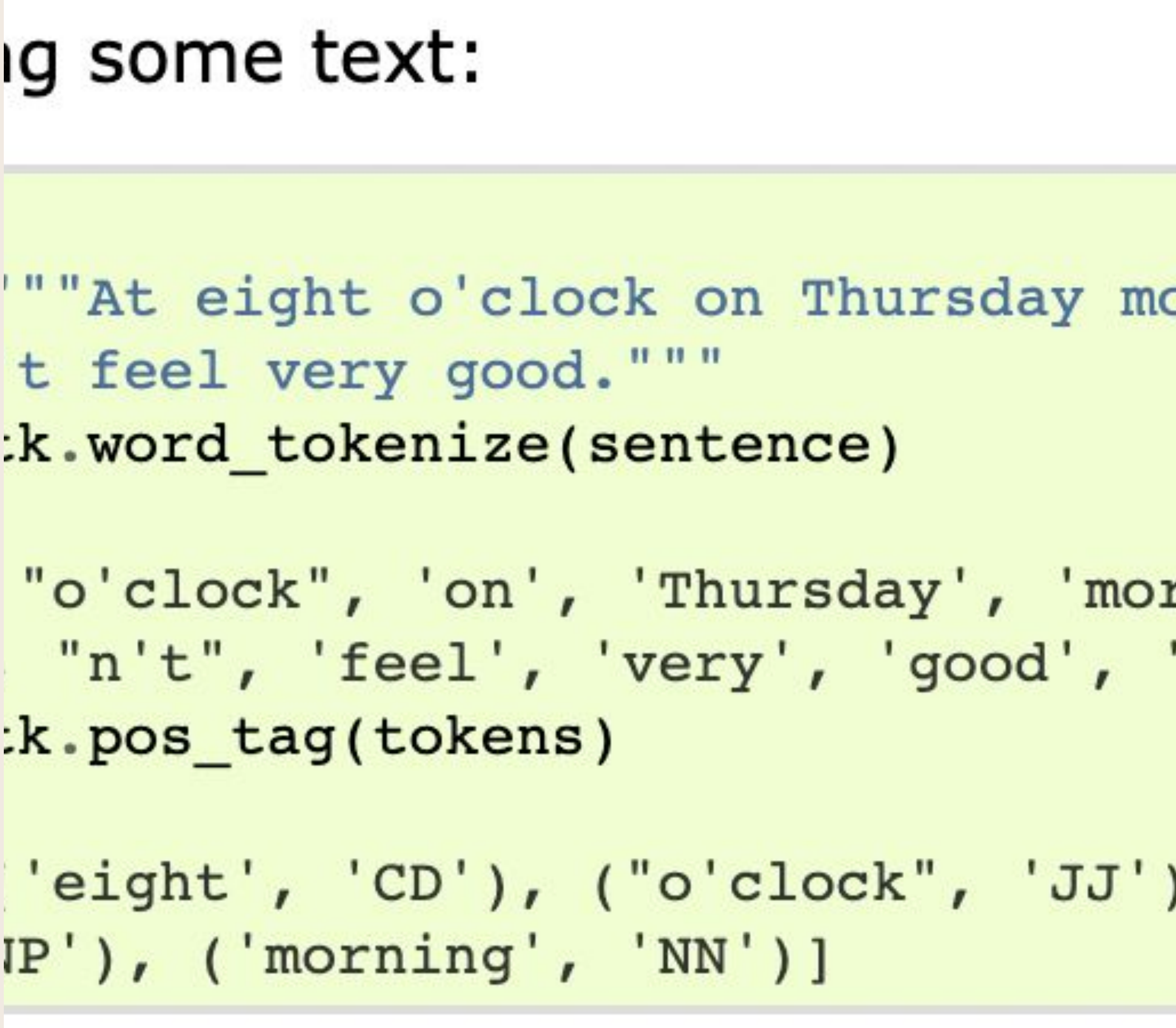
Suggested Resources

[Python for Everybody Specialization - Coursera]

[Data Visualization with Python - Coursera]

[Natural Language Tool Kit (NLTK)]

[spaCy]



References

- Burstein, J. Elliot, N., Beigman Klebanov, B., Madnani, N., Napolitano, D., Schwartz, M., Houghton, P. & Molloy, H. (2018). Writing Mentor: Writing Progress Using Self-Regulated Writing Support. *Journal of Writing Analytics*. Vol. 2: 285-313.
- Ling, Guangming, Elliot, N., Burstein, J. C., McCaffrey, D. F., MacArthur, C. A., Holtzman, S. (2021). Writing motivation: A validation study of self-judgement and performance. *Assessing Writing*, 48.
- McCaffrey, D., Holtzman, S., Burstein, J., and Beigman Klebanov, B. (to appear). What can we learn about college retention from student writing? In *Companion Proceedings in the 11th International Conference on Learning Analytics & Knowledge (LAK21).*
- Zhu, M., Zhang, M., & Deane, P. (2019). Analysis of keystroke sequences in writing logs. *ETS Research Report Series*, *2019*(1), 1-16.
- Zhu, M., Burstein, J., Hao., J., & Livne, O. (2021) Uncovering Patterns of Use in the Writing Mentor® App through Cluster Analysis, Presented at the Advancing Digital Instruction and Assessment with Natural Language Processing & Learning Analytics, Coordinated Symposium to be held virtually at the 2021 Annual Meeting of the National Council on Measurement in Education (NCME), June 2021.

duolingo

Materials from the tutorial can be found here:
https://github.com/jillburstein/LASI21-Intro-to-Writing-Analytics

Thank you!

Questions?

jill@duolingo.com