



THE UNIVERSITY OF CHICAGO

**DATA SCIENCE
INSTITUTE**

Data Science Clinic

with **University of Northern Iowa**

Wina Aaron, Sana Fessuh, Hanyun Liu, Abigail Starr



Highlighted sections represent the primary coding contributions I focused on within this collaborative project.



Taraneh Matloob, PhD

Associate Professor of Children's Literature
at University of Northern Iowa

RESEARCH QUESTIONS

- Can we find meaningful changes and patterns in word usage within Young Adult (YA) African American Novels?
- Do **power** and **danger** occur differently in YA African American novels than in YA novels by White authors?

WE CARE BECAUSE...

- ...children's' literature acts as a window and mirror
 - **mirror** — reflects reader's experiences
 - **window** — provides insights into others' experiences
- ...most literature studies are **non-computational**
 - data science tools —> new ways to analyze and interpret text

RELEVANT BACKGROUND

“Power - Danger Framework” by Dodds, et al. (2021)

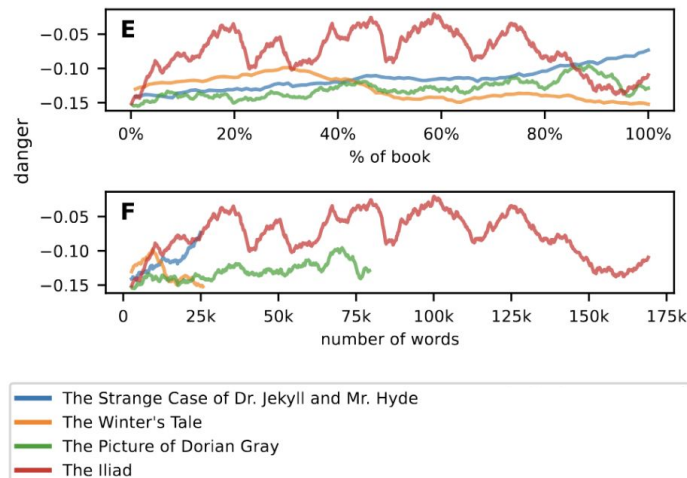
CLAIMS

- Meaning can be broken down into dimensions
 - previously, Valence, Arousal, & Dominance (VAD)
- VAD axes are **non-orthogonal**
 - Singular Value Decomposition (SVD) creates new, orthogonal axes
 - “Power” replaces dominance
 - “Danger” replaces arousal & valence
- There is a bias towards **weak-safe language**

Applications & Implications of Power-Danger Framework by Fudolig, et al. (2022)

CLAIMS

- Fluctuations in word usage across narrative time are not due to random noise
 - instead — result of **sub-arcs** within longer narratives
- Detailed methodology for creating power-danger time series



[example of time series from Fudolig, 2022]

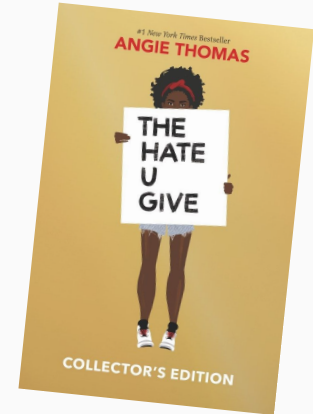
- **Two-dimensional** histograms that show distributions of meaning
 - think of a heat-map!

OUR DATA

- Textual data — Newbery Award Winners
 - 29 African American Young Adult Novels & Poetry
 - 27 White Young Adult Novels
- Table to map words to Power-Danger scores, provided by Fudolig

Word	Power Score	Danger Score
<i>attack</i>	0.230	0.471
<i>sleep</i>	-0.204	-0.448

Range = (-0.5, 0.5)



deliverables:

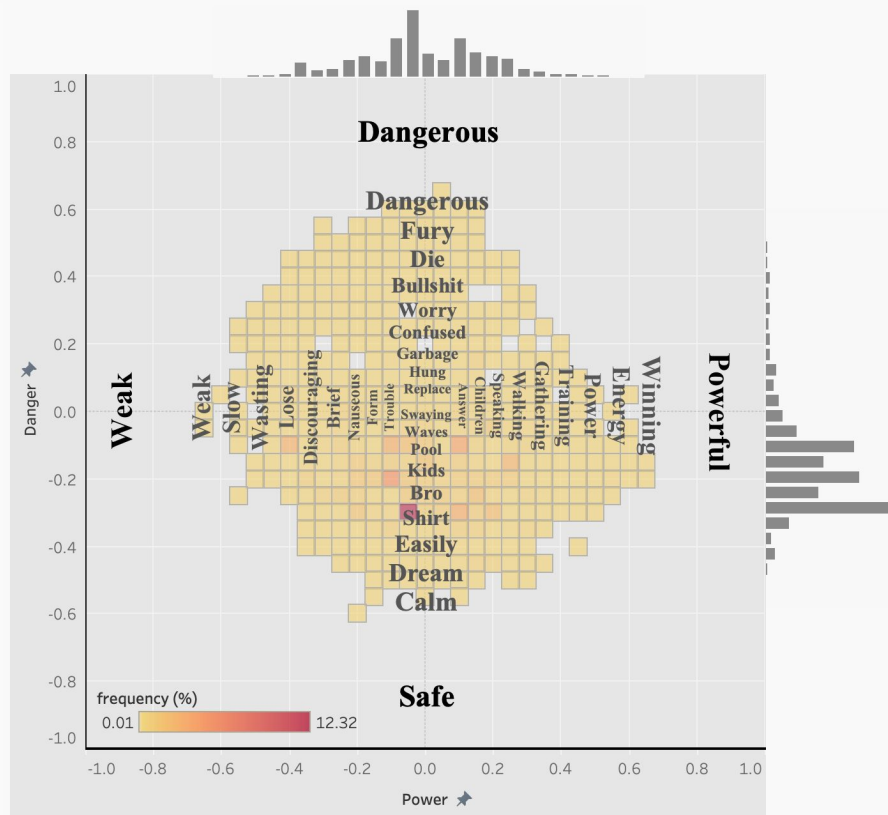
END-TO-END PIPELINE

- **Input:** raw book files
- **Processing:**
 - data cleaning and formatting
 - text segmentation
 - power-danger score application
- **Output:** visualizations & summary statistics



OUSIOGRAMS

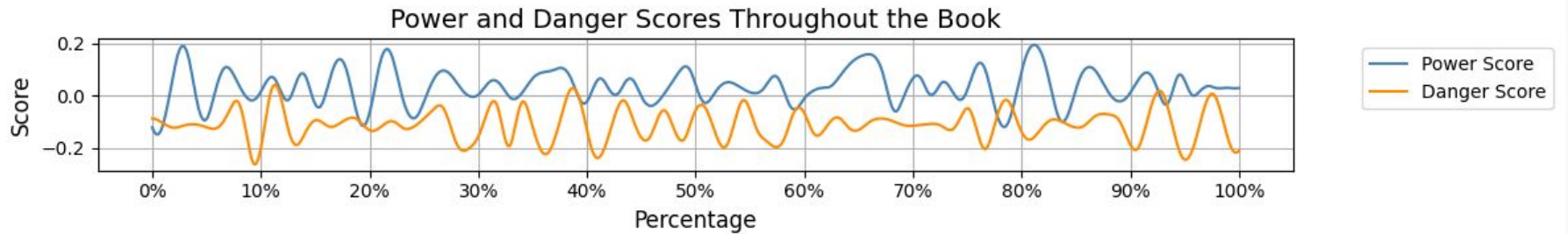
- Clustered around low-mid combinations of PD scores
- Histograms align with axes and emphasize concentration around “**neutral**” scores
- Only evidence for weak-safe bias



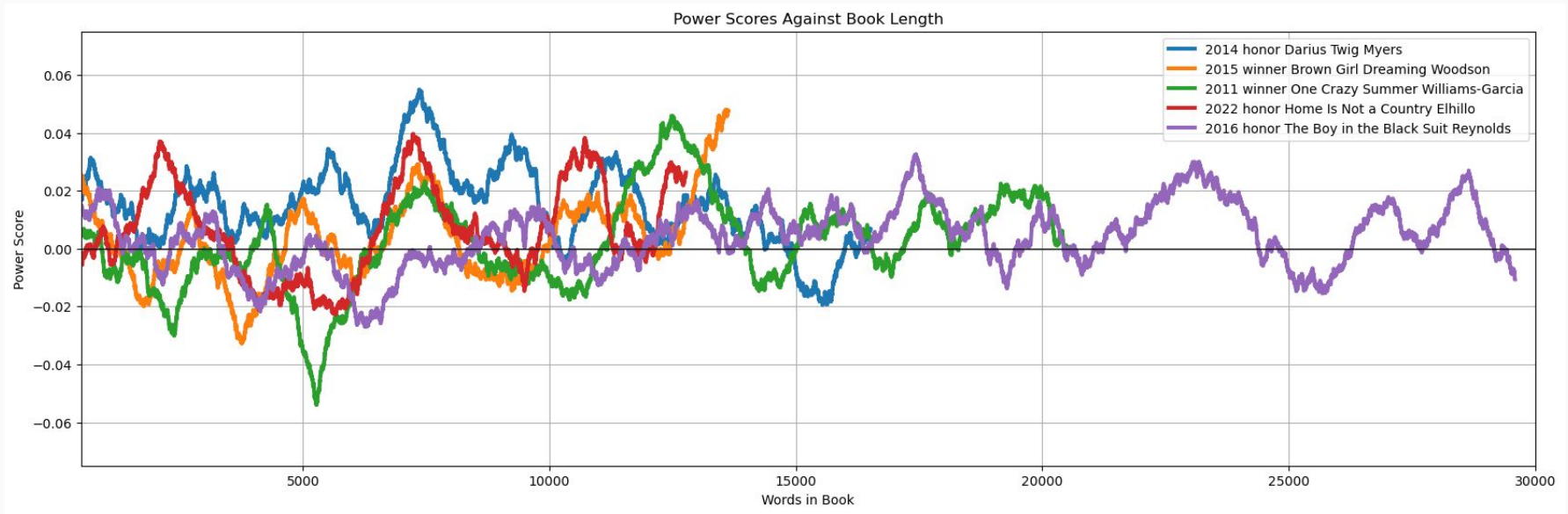
[Ousiogram for Darius & Twig By Walter Dean Myers, 2014]

TIME SERIES

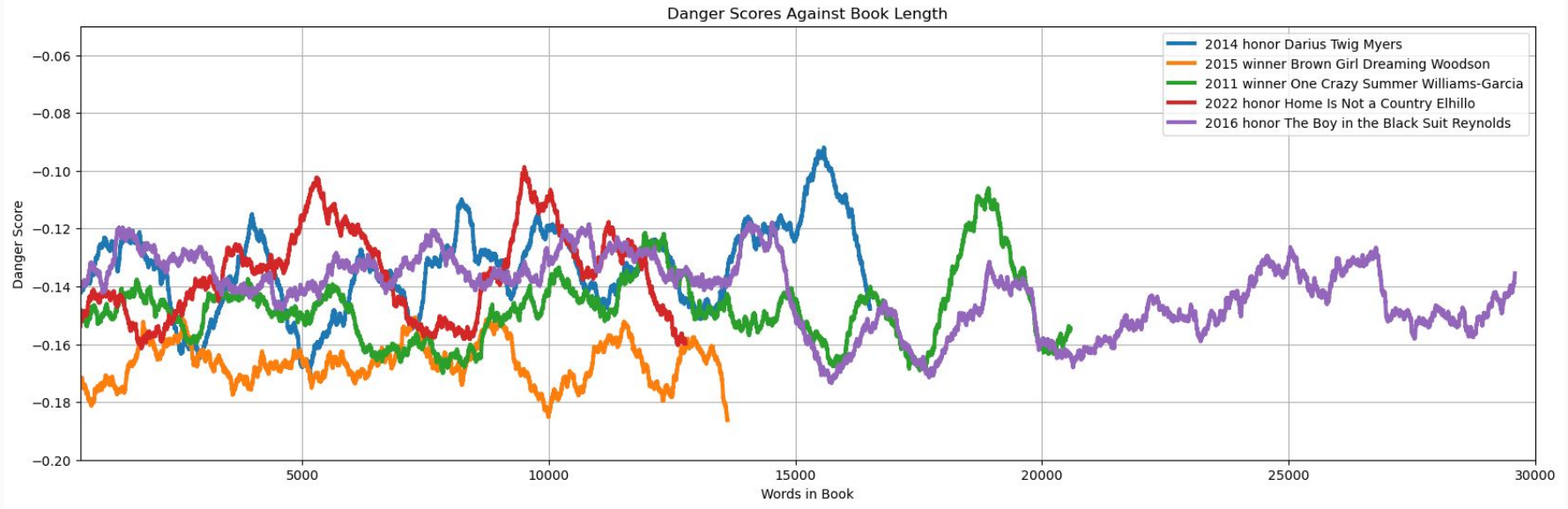
- Power scores are **mostly positive**
- Danger scores are **consistently negative**
- Fluctuate but maintain respective positions above and below zero



[Time Series for Darius & Twig By Walter Dean Myers, 2014]

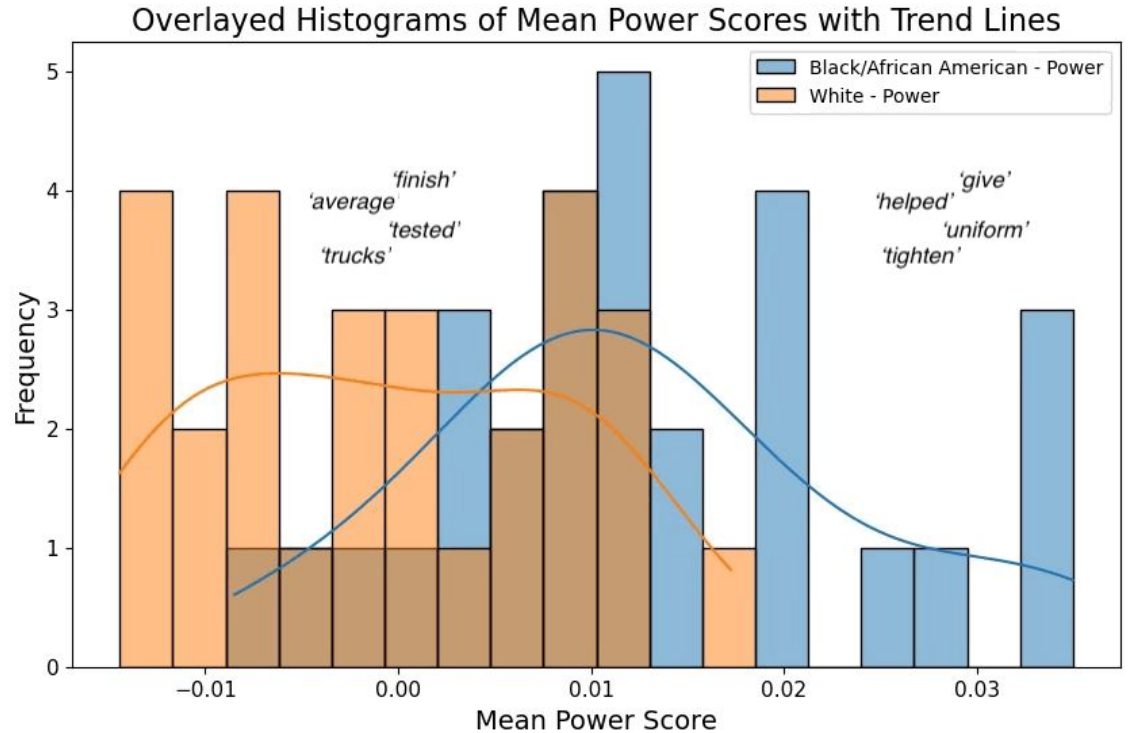


- Degree of variation does not differ between shorter vs. longer books
 - (e.g., *Home Is Not a Country* vs. *The Boy in the Black Suit*).`

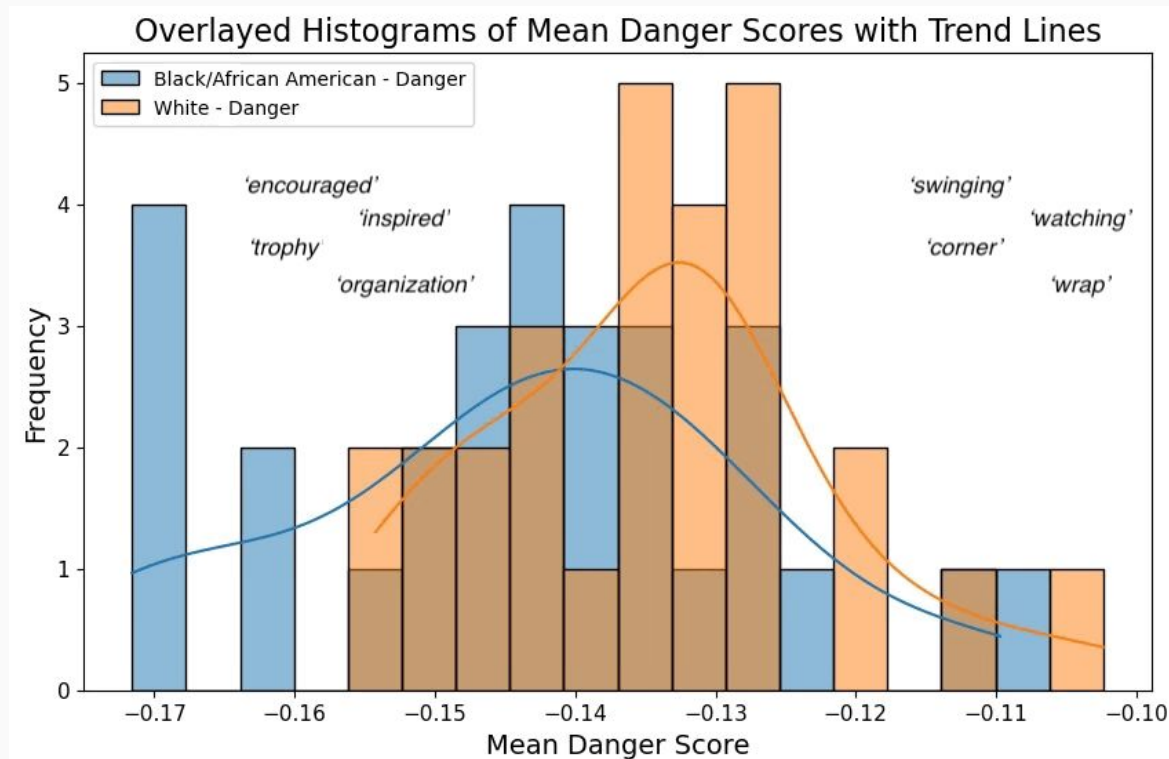


- Danger scores pattern similar to power
 - i.e. size & frequency of fluctuations do not change with length

- All novels centered around 0
- Books by African American authors
—> trend toward **power**
- Books by White authors
—> trend towards **weak**



- All novels cluster within narrow range centered around -0.14
- African American authors —> trend towards **safety**
- White authors —> trend towards **danger**



CONCLUSIONS

- Weak-safety bias exists in all data
- Shorter books showed similar evidence of sub-arcs
- Novels by African American authors trend towards **safe-powerful** words
 - may emphasize resilience, empowerment, community
- Novels by White authors trend towards **dangerous-weak** words
 - may emphasize external threats, isolation, individual vulnerability