



“TO CODE, OR NOT TO CODE”

**USING DATA SCIENCE TOOLS TO ENHANCE
YOUR LITERARY ANALYSIS**



WHAT ARE WE DOING TODAY?

- 01 Throwing a Fish at a Painting
- 02 What distant reading can and cannot do
- 03 Trying out Sentiment Analysis
- 04 A bit of Topic modeling

Text Analysis: “Like Throwing a Fish at a Painting”



TEXT ANALYSIS: WHAT IS IT GOOD FOR?



Handling Large Corpora

Analyze vast collections of texts (such as novels, plays, speeches) quickly and efficiently



Support Traditional Analysis:

It can provide quantitative data that can supplement close reading and help validate hypotheses



Uncover Patterns

Identify recurring words, themes, or sentiment shifts that may not be immediately visible, offering fresh insights and avenues for interpretation



TEXT ANALYSIS: SOME KEY PRINCIPLES



SOCIAL SCIENCE THEORIES AND SUBSTANTIVE KNOWLEDGE ARE ESSENTIAL FOR RESEARCH DESIGN



TEXT ANALYSIS DOES NOT REPLACE HUMANS—IT AUGMENTS THEM



THE BEST METHOD DEPENDS ON THE TASK & VALIDATIONS ARE ESSENTIAL





A TYPICAL DATA ANALYSIS WORKFLOW

IDENTIFY
CORPUS

WRANGLE,
CLEAN,
TRANSFORM

EXPLORE

ANALYZE,
CREATE MODELS

PRESENT
RESULTS





Sentiment Analysis:

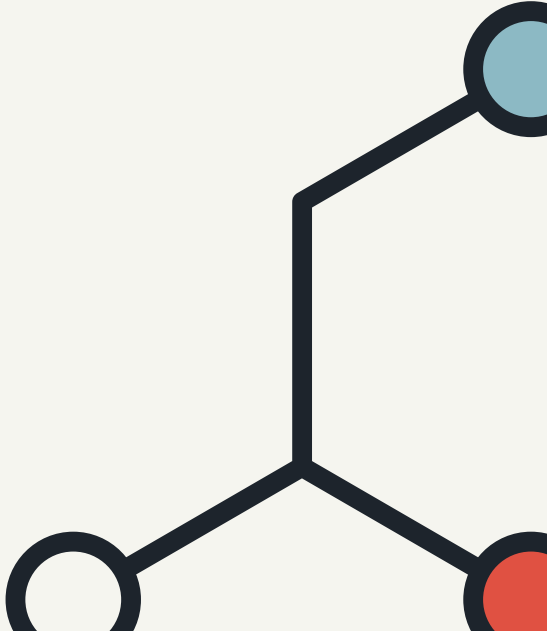
“Don’t be afraid to catch feels”

- **Sentiment analysis** assesses the emotional tone (positive, neutral, negative) expressed in texts.
- Can be useful when exploring character relationships, narrative tone shifts, and authorial perspective in literary texts.
- It is also used to analyze feedback loops and monitor the effect of news on public opinion, among others.




Sentiment Analysis:

“Don’t be afraid to catch feels”

- **Sentiment analysis** is rule-based; we use algorithms to compare each word in the text to a sentiment dictionary
- 




Topic Modeling:

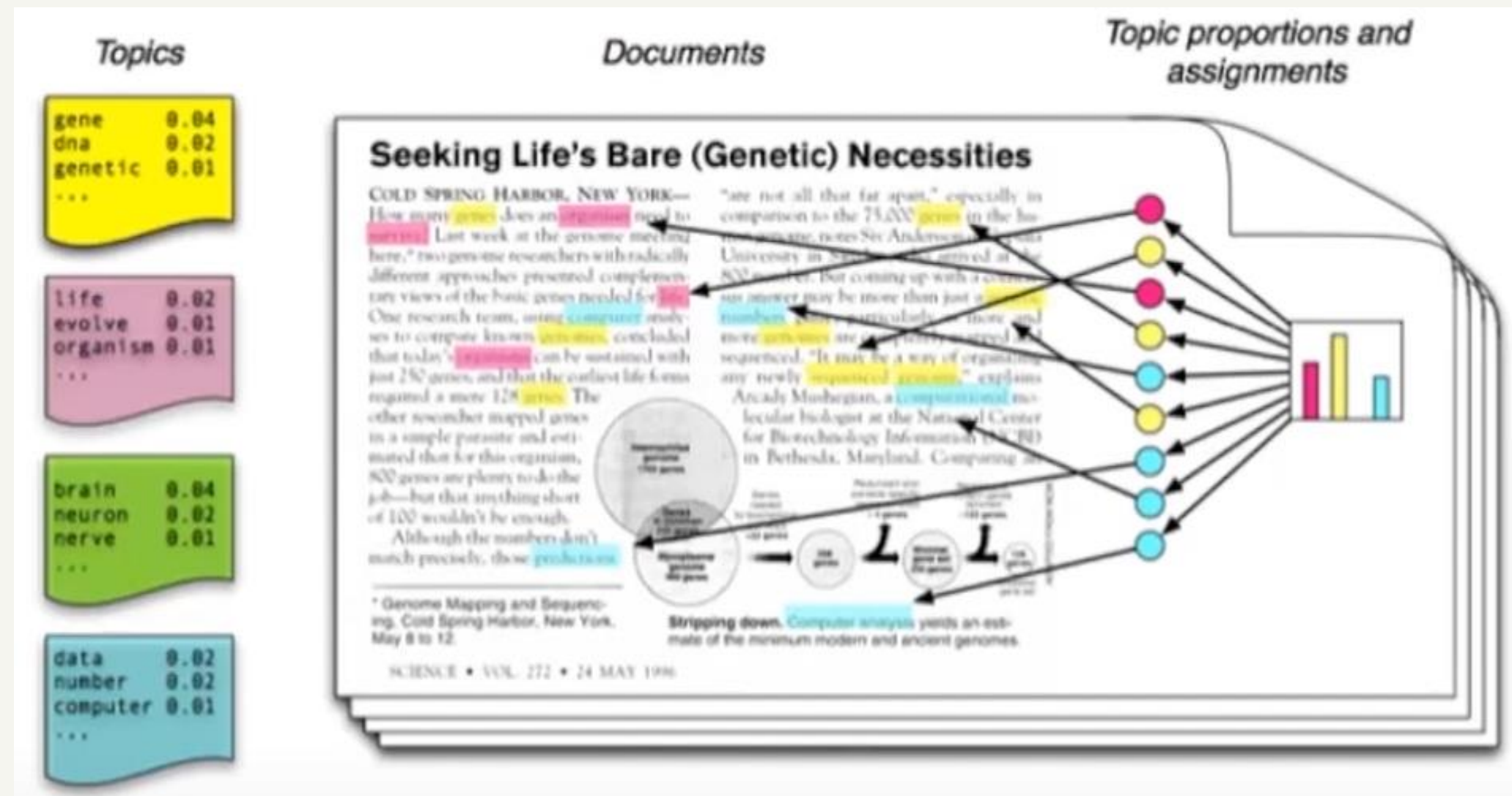
- Topic modeling (LDA) identifies clusters of related words that represent underlying themes or topics in a text.
 - LDA can reveal latent themes across a body of work, track thematic changes over time, or even compare thematic focus between different authors or genres.
- 



Topic Modeling:

- **Thematic Tracking:** Identify how themes like emerge and evolve across texts.
 - **Analyze a larger corpus for overarching themes** that might not be visible when reading individually.
- 

Topic Modeling:





Topic Modeling:

How does Topic modeling with LDA work?

- **Input**: LDA starts with a corpus and a predefined number of topics you want to find.
- Assigning Topics: LDA assigns a topic based on probabilities for each word; it looks at the words around it and determines which topic it likely belongs to. The process is iterative, adjusting the topic assignments until stable groupings of words are found.
- **Output**: A list of topics, each characterized by the most frequent words associated with it, allowing you to understand the main themes in your documents.

USING R FOR TEXT ANALYSIS



**Long-standing and
Reproducibility**

Researchers can create fully reproducible documents ,
code, analyses, and narrative



**Comprehensive Text Analysis
Packages**

Rich ecosystem of packages specifically designed for text
mining and natural language processing.



Data Visualization

Visuals like topic distribution graphs and sentiment
timelines enhance understanding



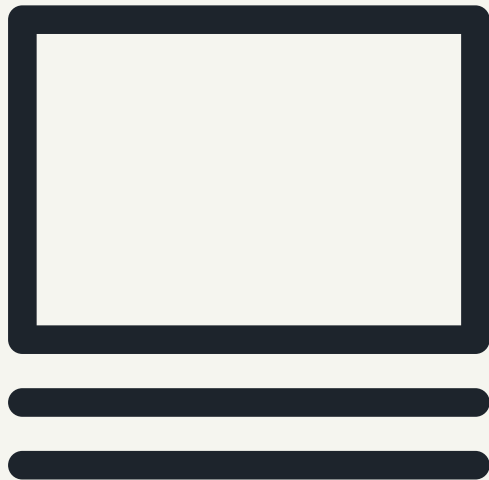
COMMUNITY & SUPPORT

Open-source codes, free online resources, forums, books,
lots of Youtube videos





PACKAGES



WHAT'S A PACKAGE?

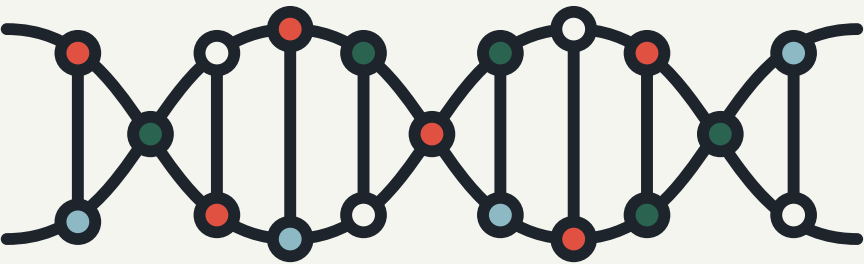
A collection of pre-written codes that you can use to add extra functionality to your programs without writing everything from scratch.

Think of packages like toolkits—they provide tools (code) for specific tasks, such as working with data, creating graphs, or making web requests.

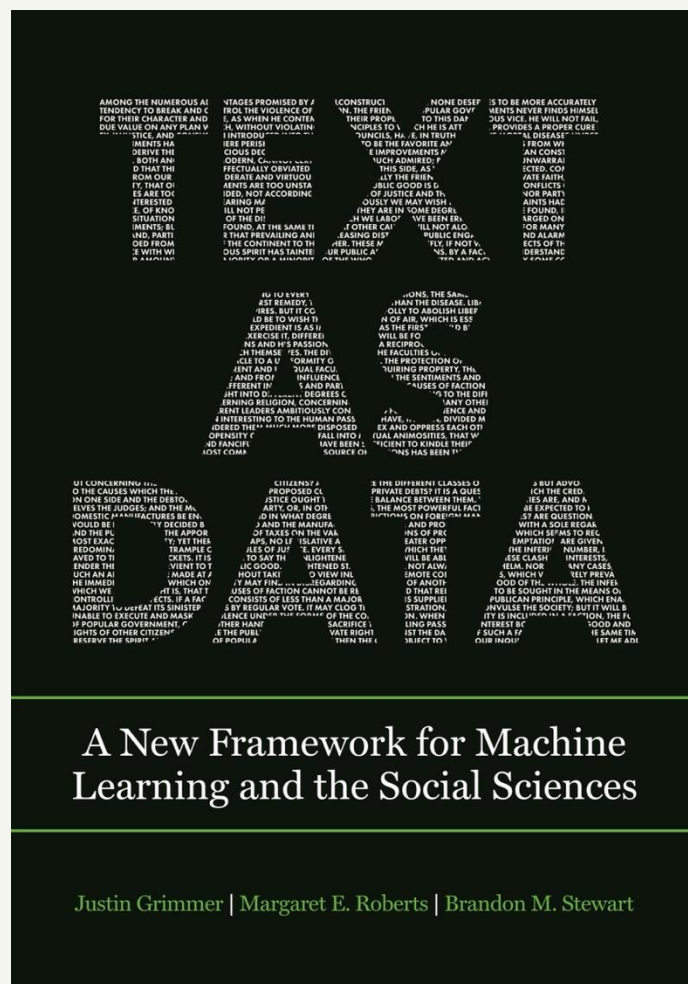
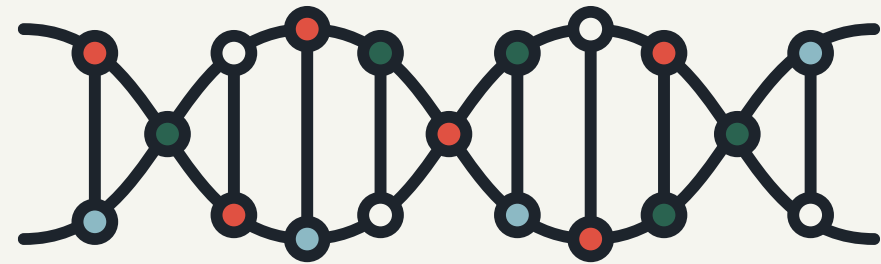
USING R FOR TEXT ANALYSIS

When your code finally runs after 257
attempts

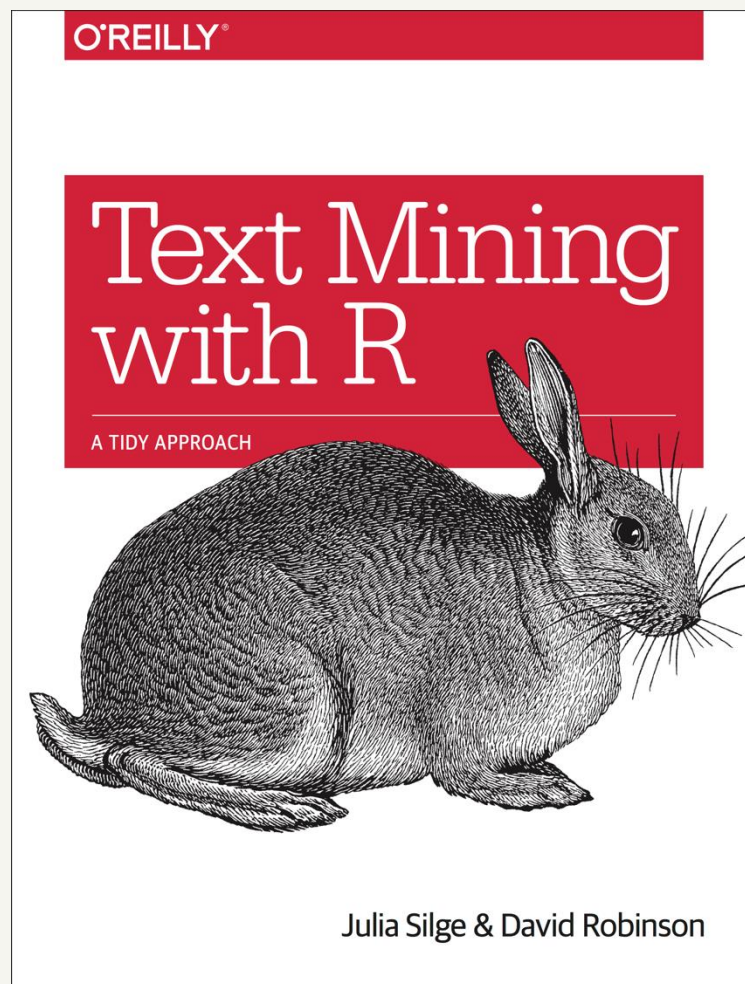




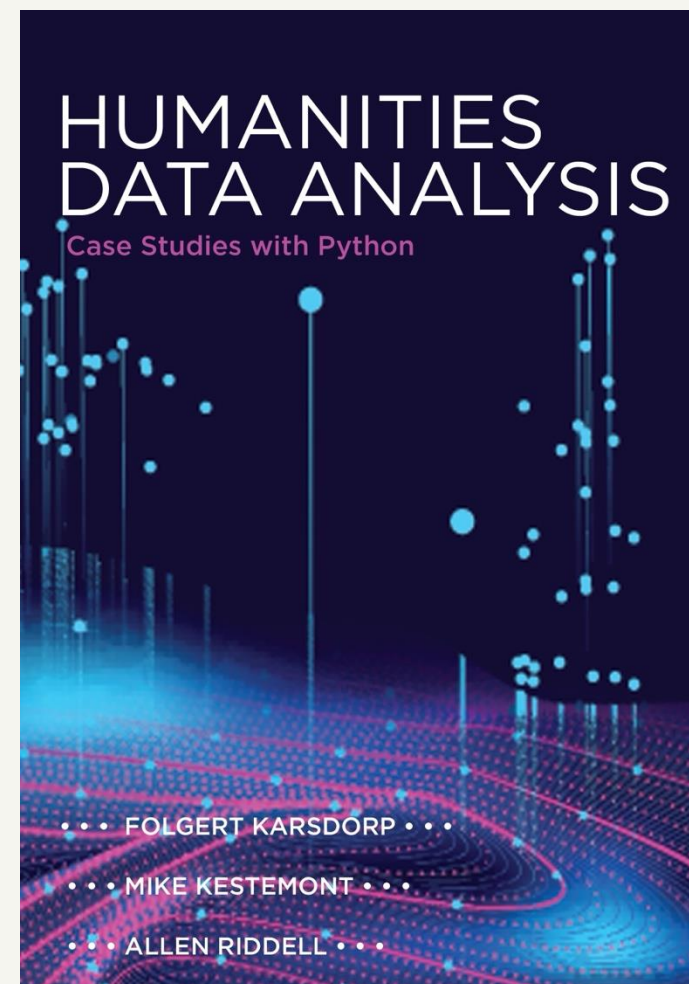
SOME RECOMENDED READS



TEXT AS DATA



TEXT MINING WITH R



HUMANITIES DATA ANALYSIS



ENHORABUENA!

You took your first steps in text analysis!





THANK YOU!