Developing library-based computational research skill instruction programs

William Okech, PhD

05/15/2023

Overall Goal

To outline a methodology for developing a **computational research skill instruction program** centered in **libraries** that will **support learners with diverse disciplinary and educational backgrounds**.

Outline

1. Importance of computational research skills

2. Effective teaching of computational research skills

3. Sample lesson and problem

4. Challenges associated with teaching computational research skills

5. Conclusions

Importance of Computational Research Skills

Importance of Computational Research Skills

Why are computational research skills important for everyone?

- “Computational thinking is a fundamental skill for everyone, not just computer scientists.” (Wing, 2006)
- Major components of computational thinking include:
  - **Decomposition** – break down a complex problem to multiple simpler problems.
  - **Pattern recognition** – identify patterns in the data or information.
  - **Abstraction** – identify and use the relevant aspects of the problem.
  - **Algorithmic thinking** – design algorithms to solve problems.

1. Wing, J (2006) Comm. ACM, 49 (3):33-35
2. Grover, S and Pea, R (2012) Educational Researcher, 42 (1):38-43

Importance of Computational Research Skills

Why are computational research skills important for everyone?

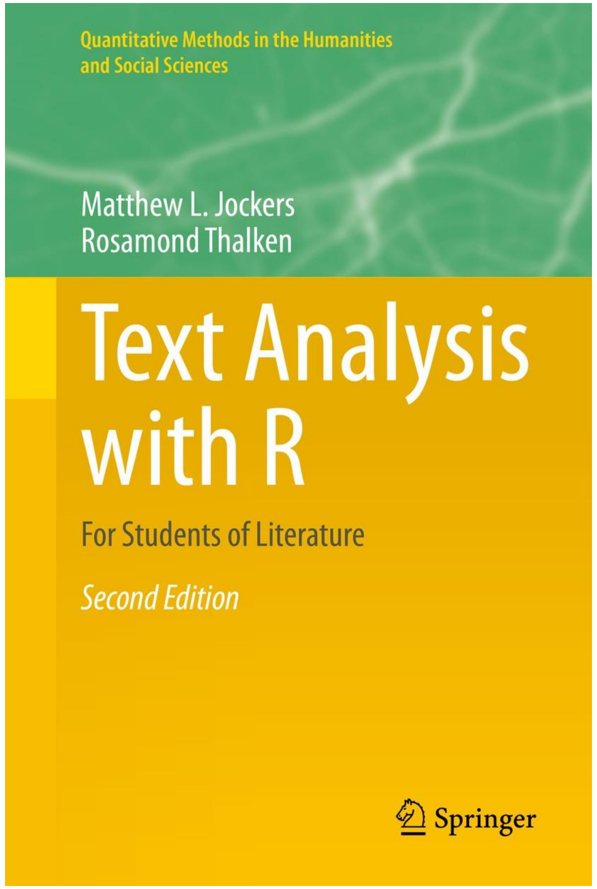
- New career opportunities and increased competitiveness in the job market.
    - Employment of software developers is projected to grow by 25% (~400k jobs) between 2021 and 2031 (Bureau of Labor Statistics).
- Subject-matter experts with computational research skills are required to build new tools/technologies. Examples include:
    - Computational linguists – develop new language-learning apps.
    - Data journalists – develop fact-checking and misinformation detectors.

https://www.bls.gov/ooh/computer-and-information-technology/software-developers.htm#tab-6

Importance of Computational Research Skills

Where can computational research be applied in the humanities?

- Text analysis for Literature Students



Quantitative Methods in the Humanities and Social Sciences

Matthew L. Jockers
Rosamond Thalken

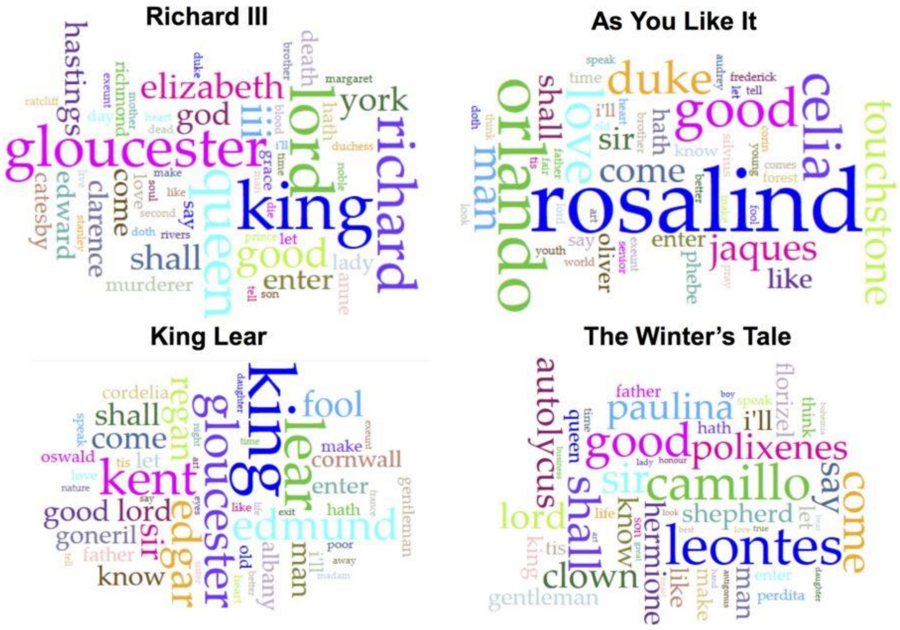
Text Analysis with R

For Students of Literature

Second Edition

Springer

Or you can compare several plays and look for patterns or differences.



Richard III

As You Like It

King Lear

The Winter's Tale

Word clouds from Voyant-tools.org
CC BY 4.0 Voyant Tools by Stéfan Sinclair & Geoffrey Rockwell

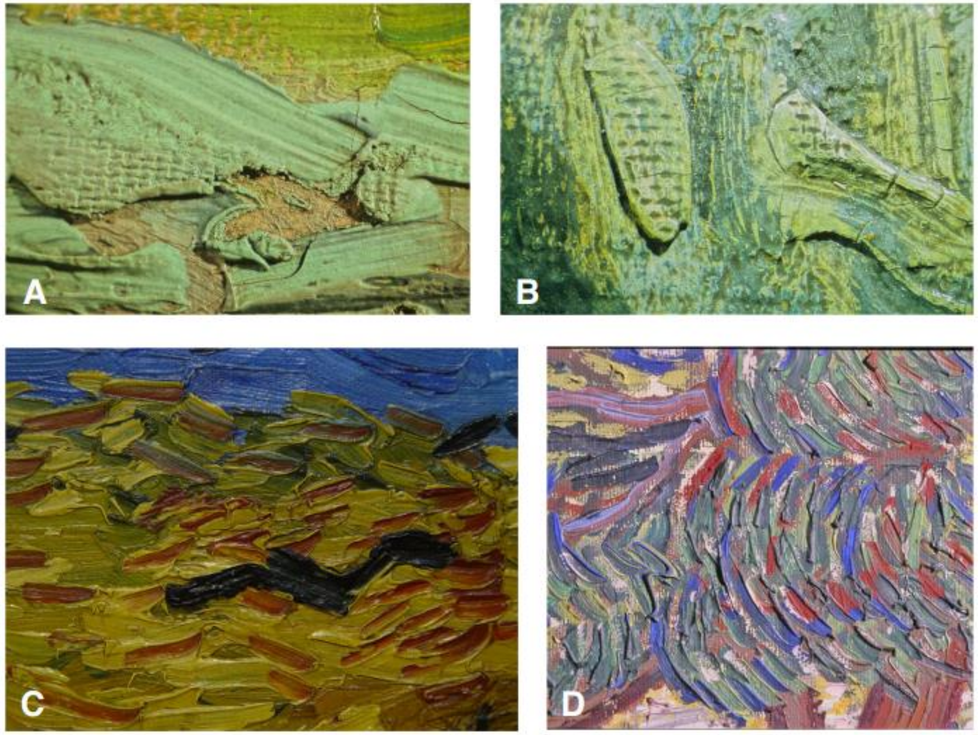
Plays by Shakespeare

1. Jockers and Thalken (2020) Text Analysis with R for Students of Literature
2. https://kennedyhq.com/wp/2016/11/19/shakespeare-through-digital-humanities-textual-analysis/

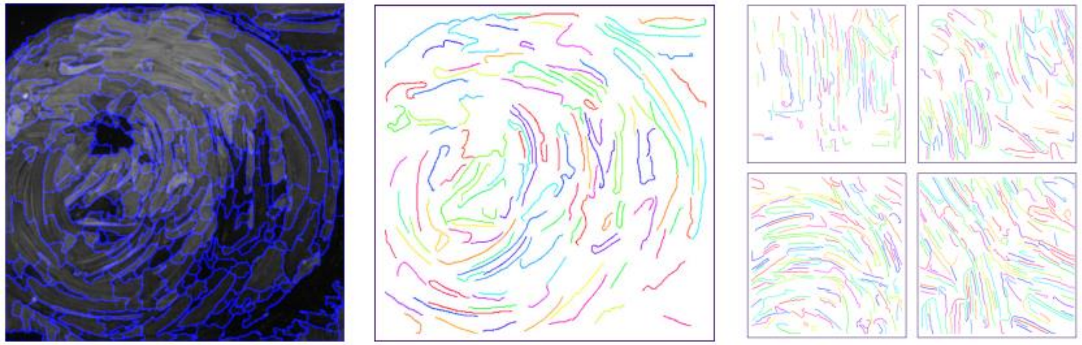
Importance of Computational Research Skills

Where can computational research be applied in the humanities?

- Image Processing for Artist Identification (useful in Fine Arts)



Portions of van Gogh’s paintings



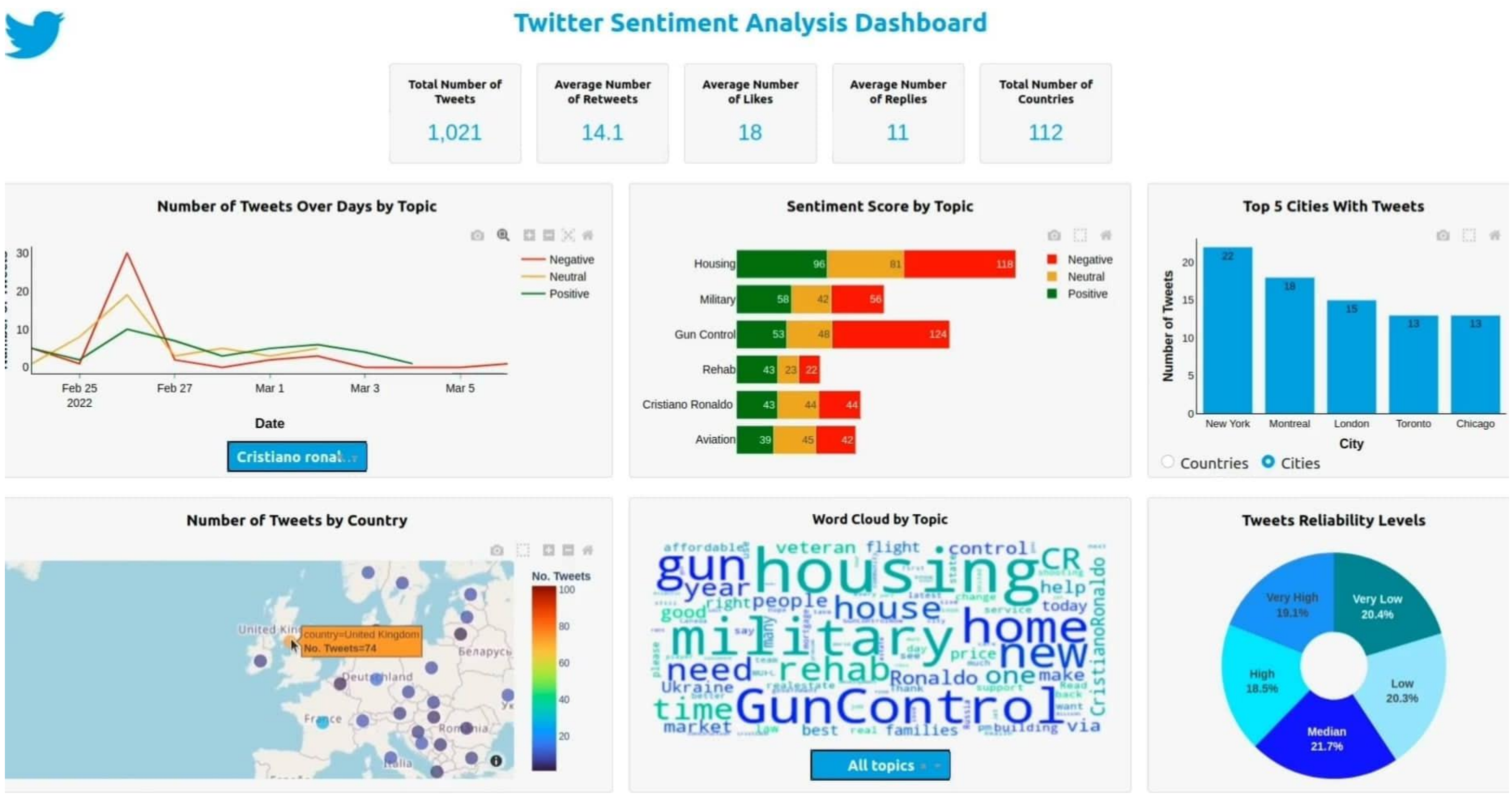
Brushstroke analysis

C. R. Johnson et al. (2008) IEEE Signal Processing Magazine, 25(4): 37-48

Importance of Computational Research Skills

Where can computational research be applied in the humanities?

- Sentiment Analysis and Social Media (useful in Political Science & Journalism)



Twitter Sentiment Analysis Dashboard

Total Number of Tweets: 1,021
Average Number of Retweets: 14.1
Average Number of Likes: 18
Average Number of Replies: 11
Total Number of Countries: 112

https://github.com/Mazen72/Twitter_Sentiment_Analysis_Dashboard

Importance of Computational Research Skills

Where can computational research be applied in the humanities?

Computational Research Skill	**Humanities Research Field**
Text Analysis	Ancient and Modern Languages
Data wrangling of census data	History and Political Science
Image processing	Art History and Film Studies
Spatial data analysis	Geography
Audio analysis	Music

Martyn Jessop (2004) Ubiquity, ACM

Importance of Computational Research Skills

What computational research skills are important?

This can be determined with a **needs analysis** using:

- **Surveys** and **Questionnaires** sent to humanities departments,
- **Literature reviews** focused on computing education research,
- **Interviews** with faculty and staff from humanities departments.

Summary

1. Computational research skills are essential for everyone.

2. Humanities students can greatly benefit from computational research skills by:

    - Exploring new research directions, and,

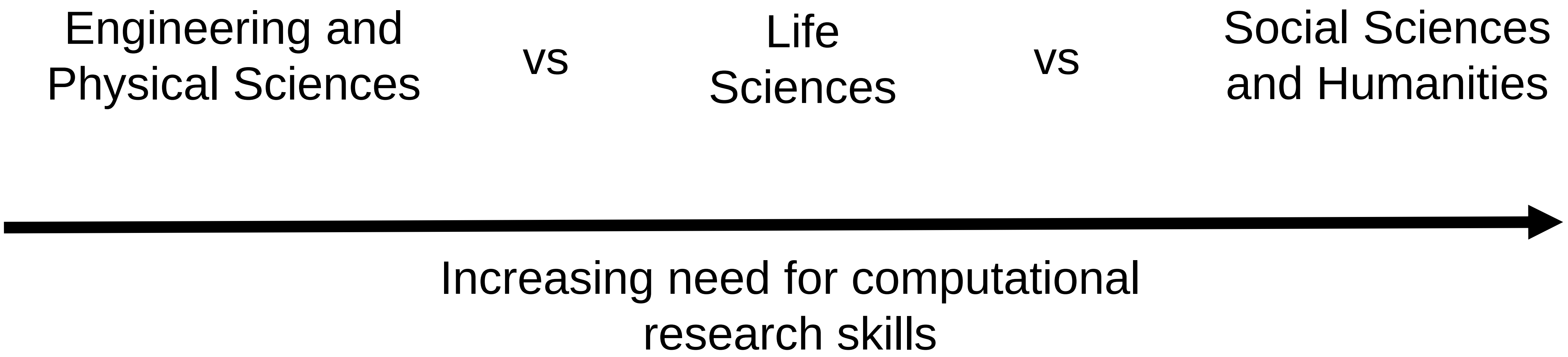
    - Increasing their employment options.

Martyn Jessop (2004) Ubiquity, ACM

Effective teaching of computational research skills

Effective teaching of computational research skills

Recognize the diversity of needs



Engineering and Physical Sciences

vs

Life Sciences

vs

Social Sciences and Humanities

Increasing need for computational research skills

Effective teaching of computational research skills

What activities should the instruction program focus on?

1. **Workshops** – where the most common computational tasks and the needs identified by graduate students are taught.

2. **Office hours** – supervised by an undergraduate/graduate student with a background in computational research.

3. **Mentoring** - pair humanities graduate students with computational science students.

4. **Bootcamps** – focused on specific data analysis tools/programs.

5. **Journal/Data clubs** – where new and relevant tools, techniques, and software packages will be demonstrated.

Effective teaching of computational research skills

A number of independent initiatives exist to train the general public in basic data and computational skills. These include:

THE CARPENTRIES

Goal: To teach foundational coding and data science skills to researchers worldwide.

DATA CARPENTRY

Fundamental data skills needed to conduct research.

library carpentry

Software and data skills for library and information-related communities.

software carpentry

Lab skills for research computing.

https://carpentries.org/

Effective teaching of computational research skills

A number of independent initiatives exist to train the general public in basic data and computational skills. These include:

CODE REFINERY

Goal: Teach all the essential tools which are usually skipped in academic education so everyone can make full use of software, computing, and data.

FORRT

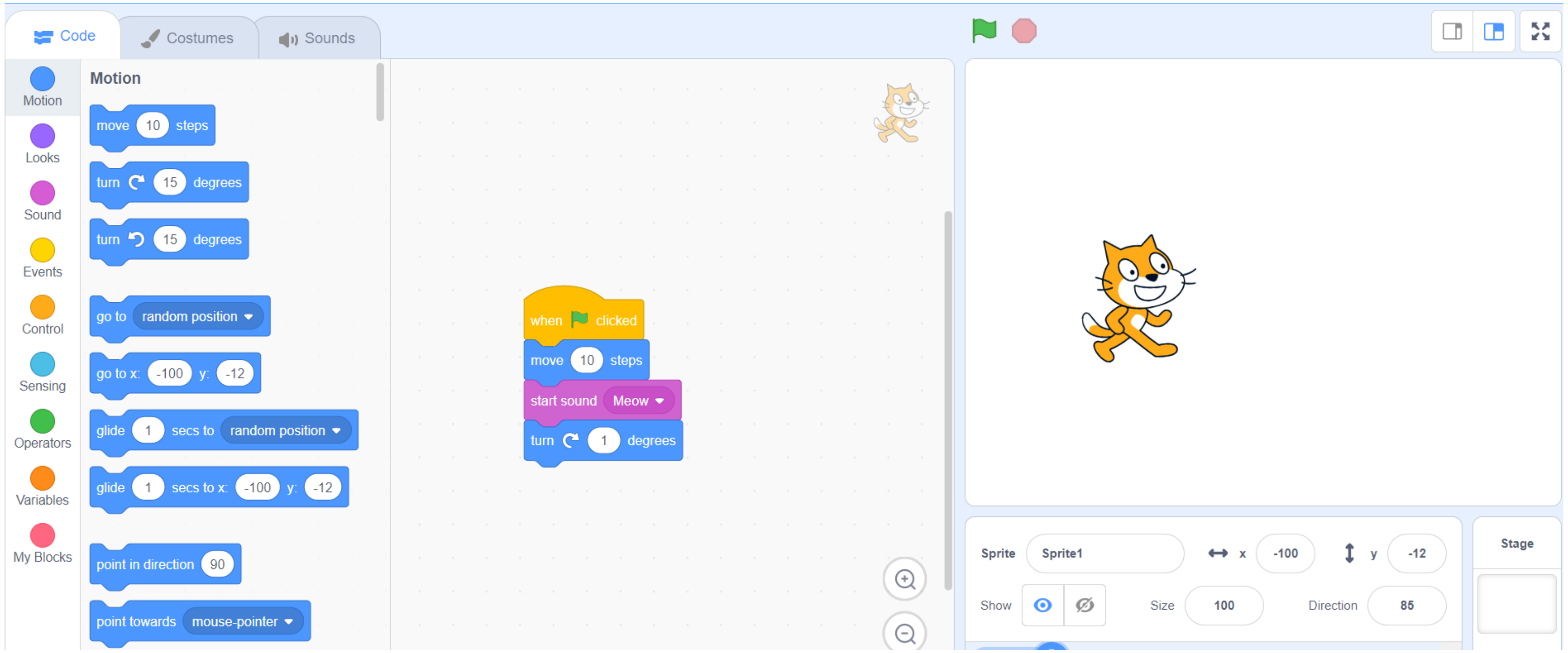
Framework for Open and Reproducible Research Training

https://coderefinery.org/ | https://forrt.org/

Effective teaching of computational research skills

What should novice coders learn? (**My personal opinion**)

- Scratch (block-based visual programming language) for complete novices.
- Help new coders learn the basics of programming, such as loops, conditionals, and variables.

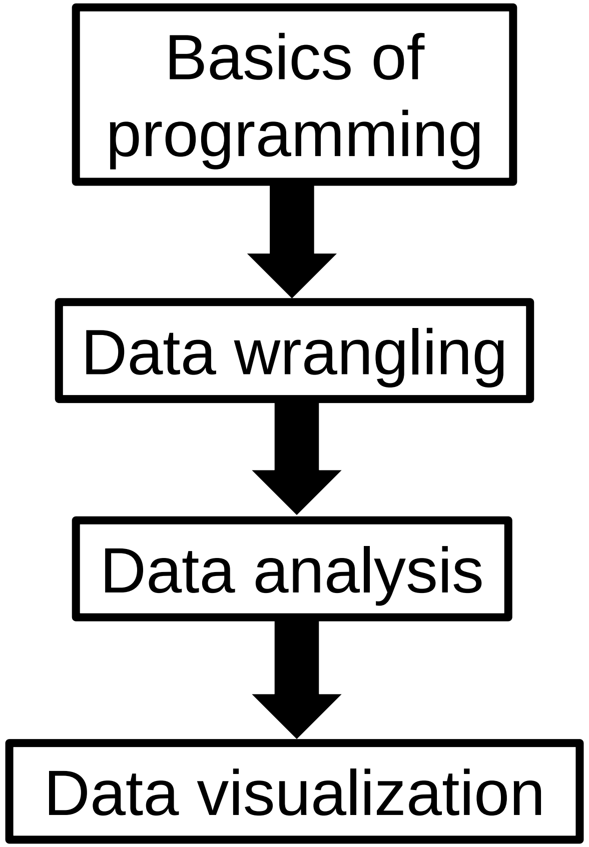


```
when 🏁 clicked
move 10 steps
start sound Meow ▾
turn ↻ 1 degrees
```

Scratch: https://scratch.mit.edu/

Effective teaching of computational research skills

What should novice/advanced beginners learn? (**My personal opinion**)



Basics of programming

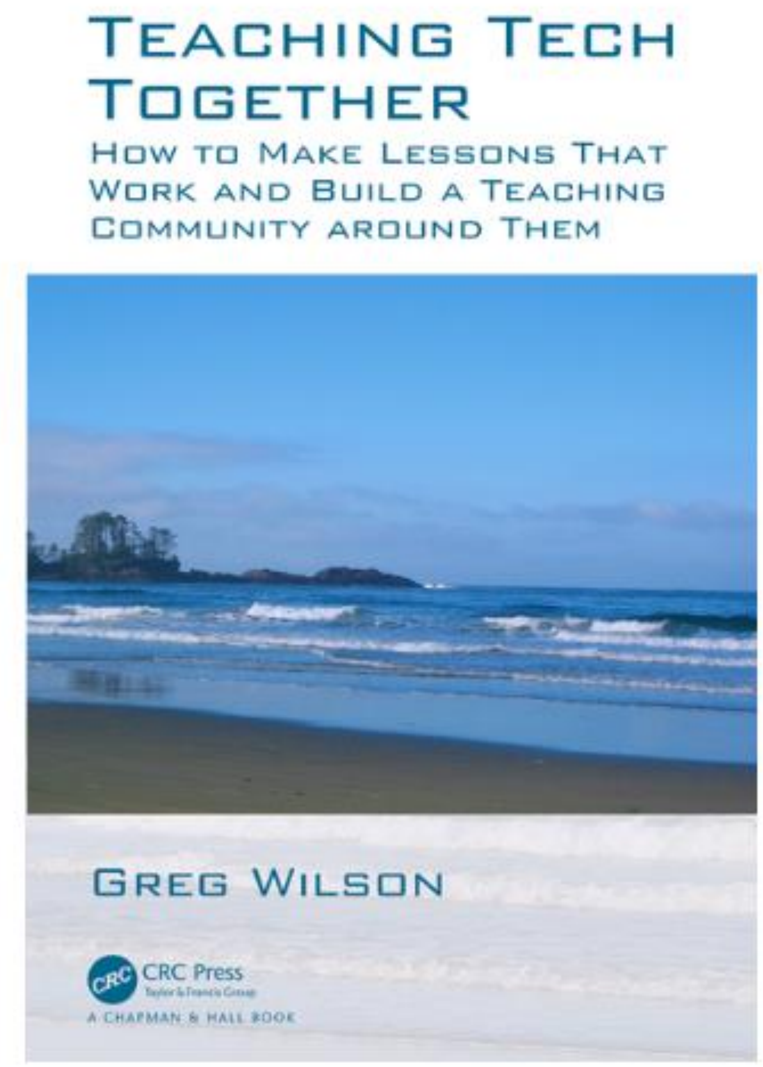
Data wrangling

Data analysis

Data visualization

Effective teaching of computational research skills

Key points for teaching novice coders



TEACHING TECH TOGETHER

HOW TO MAKE LESSONS THAT WORK AND BUILD A TEACHING COMMUNITY AROUND THEM

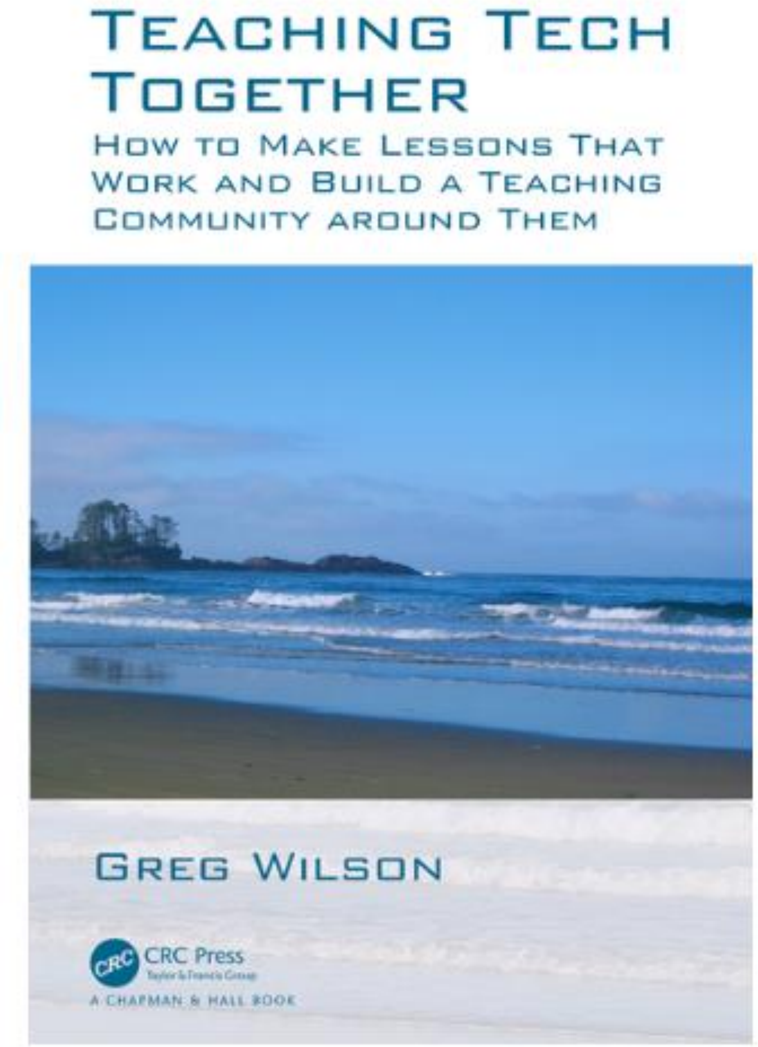
GREG WILSON

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

- Use live coding.
- Embrace coding errors as an opportunity to teach debugging.
- Use “worked examples” and pace learning to avoid cognitive overload.
- Encourage working in pairs and group coding.
- Promote daily practice.

Effective teaching of computational research skills

Key points for teaching novice coders



- Learn student personas and adapt lessons to suit each student.
- Give effective feedback that motivates.
- Create learning goals for students.
- Use real-life tasks relevant to student research.

Sample Lesson and Problem

Sample Lesson and Problem

Data Manipulation in R

dplyr is a grammar of data manipulation

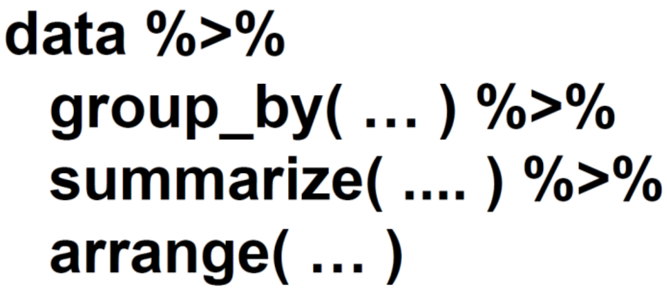
“pipe” function for combining a sequence of operations.

www.tidyverse.org

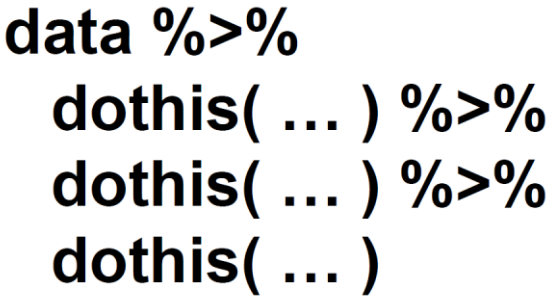
Sample Lesson and Problem

Data Manipulation in R

1. To demonstrate the role of the pipe operator

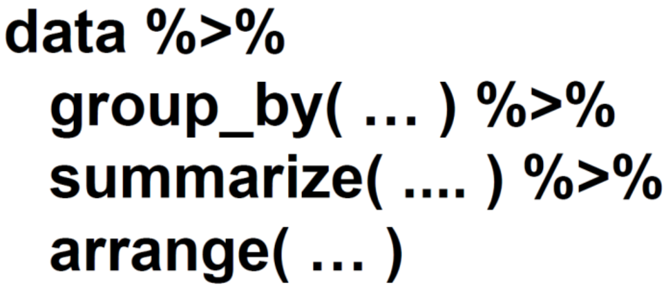


```
data %>%
  group_by( … ) %>%
  summarize( .... ) %>%
  arrange( … )
```

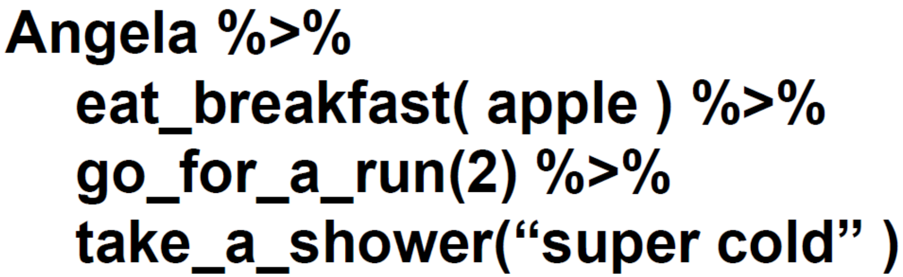


```
data %>%
  dothis( … ) %>%
  dothis( … ) %>%
  dothis( … )
```

2. To demonstrate the sequence/order of operations



```
data %>%
  group_by( … ) %>%
  summarize( .... ) %>%
  arrange( … )
```



```
Angela %>%
    eat_breakfast( apple ) %>%
    go_for_a_run(2) %>%
    take_a_shower(“super cold” )
```

Sample Lesson and Problem

Data Visualization in R

system for declaratively
creating graphics

Parson's Problem

A. ggplot(data = …, aes(x = …, y = …)) +
    geom_point()

B. geom_point() +
    ggplot(data = …, aes(x = …, y = …))

C. (data = …) +
    ggplot(aes(x = …, y = …)) +
    geom_point()

www.tidyverse.org

Challenges associated with teaching computational research skills

Challenges associated with teaching computational research skills

Barriers to overcome include:

- Attitudes towards coding.
- The variety of needs (even within humanities):
  - Example: Psychology vs Literature.
- Over-reliance on no-code commercially available tools and a lack of willingness to change.

Challenges associated with teaching computational research skills

What students should avoid:

- Learning too much at once without mastering the basics.
- Continuously working on tutorials but not applying knowledge to personal projects.
- Copy-and-pasting code rather than understanding why the code was used.
- Learning multiple tools/programs simultaneously (such as Python, R, and Javascript).

Conclusions

Conclusions

- Computational thinking is for everyone.
- Computational research skills can be very helpful for humanities students.
- Various initiatives such as the Carpentries and CodeRefinery have devised effective blueprints for teaching.
- Novices should consider starting with visual block-based programming.
- There is a need to teach students effective learning practices and overcome the barriers to coding.
- Teaching should be adapted to different student abilities and should include motivational feedback.

Questions