Our GitHub Organization
&
Finding and Reusing Data

DATA 3101

Elizabeth Stregger & Doug Campbell

September 15, 2022

Mta Data3101 Organization on GitHub

https://github.com/MtADATA3101

- Includes course content

- Each of you will have a personal repository for your assignments. Over the term, this will build up to become your portfolio of work.

- We will give you feedback on your work and you can update it based on this feedback.

In-class activity: Part 2: Push to GitHub

- Create a new version control project in RStudio, linked to your personal repository
- Go to File – New File – R Markdown
- Give the file a human readable filename (can have spaces here)
- Add the text you created in Part 1 to the bottom of the bottom of the default text
- Save with a machine-readable filename (no spaces!)
- Go to Settings – Output Options – Advanced
- Check "Keep markdown source file"
- Knit
- In the Git pane, stage the new folder, commit, and push

Reusing data: Opportunities

- Reuse of data
  - Reduces respondent burden (for over-studied populations)
  - Enables data linkage and the creation of new datasets
  - Provides transparency
  - Over time, data sets collected from one disciplinary perspective can become of interest to another

- Think about:
  - Reinterpretations
  - New questions
  - New approaches

Reusing data: Challenges

- Finding relevant data
- Fit of the dataset to your new research questions
- Time to get to know the data, documentation, and code
- Lack of statistical skills
- Missing documentation
- Concerns about ethical reuse

Finding Data

- Data guide at Mount Allison University Libraries & Archives: https://libraryguides.mta.ca/data/find

- R3data.org: Registry of research data repositories https://www.re3data.org/

In-class activity on finding data

- Look for 1-2 datasets that interest you for reuse throughout this class. You will use these dataset to practice research data management skills.

- Add a link to the dataset(s) in GitHub. Explain why the dataset(s) interests you.