Twin Cities Stat Chat

Matthew Beckman (2/26/2020)

What's all the excitement about?

- Data analysis is often collaborative.
  - team contributing code to the same project, even same document
  - collaboration may be asynchronous
  - an analyst may decide to continue work left behind by someone else
  - Each is simplified with VC or "source code management"
- Version control records changes to a file or a set of files
  - record the entire history/evolution a project;
  - manage contributions of multiple collaborators;
  - revisit history of the project
  - revert back to a specific version of any file if needed;
  - draft changes without modifying the main file until you feel ready to commit to them.
- Computational reproducibility
  - accuracy of your code & methods
  - allows others the opportunity to make use of your work
    - replicate
    - extend and modify for a new purpose.
  - bolster integrity of the work
  - preserve a complete and transparent record of all actions in a project from raw source data through final product.
- Professional caliber workflow

Steps for implementation as a learning objective

- First exposure to Git
- Configure RStudio & Git (trivial with RStudio Cloud)
- Deploying assignment Repos
  - student view
  - instructor view
  - individual & group assignments

Assessment

- completion task for intro table
- XC at first–low stakes incentive
- let VC rescue submission errors (wrong document; etc)
- submit project as a repo (or gh-pages)
- instructions in the repo (just a URL in LMS)
- individual & group repo assignments
- "count" commits (e.g., hwk repo end of term)
- assessment tasks about workflow & tools