

Best practices for uploading to GitHub

Andrew Ba Tran

This is from the seventh chapter of learn.r-journalism.com.

Just some suggestions on how to keep clean GitHub repositories.

Remember that there are different types of audiences for your data and methodology.

Provide summarized data

While it's excellent to include the scripts that detail the cleaning and wrangling process it took to turn raw data into the polished set you've published, there is a large audience of people who just want to download and play with the finalized data.

Include a folder in the repo that you can point them to so they don't have to dig through your methodology to reproduce the summarized data.

.gitignore

Use `.gitignore` to exclude certain files from being uploaded to GitHub. Such as:

- Files larger than 100 mb
 - GitHub will refuse to upload the data
- Files with your keys or passwords
- Extraneous files like your R history

You can borrow this `.gitignore` file for inspiration.

Include readmes and data dictionaries

Let people know what they're dealing with.

Be as specific as possible, including where you got the data from.

- BuzzFeed is a good model for how they index their story links and repos as a table

Please don't create monster data repos

To learn more about using git and GitHub with R

- <http://happygitwithr.com/>