Programming Toolkit for Data Science

Programming Toolkit: Version Control and Collaboration

Git and GitHub

The following content is based on Mine Çetinkaya-Rundel's excellent book Data Science in a Box

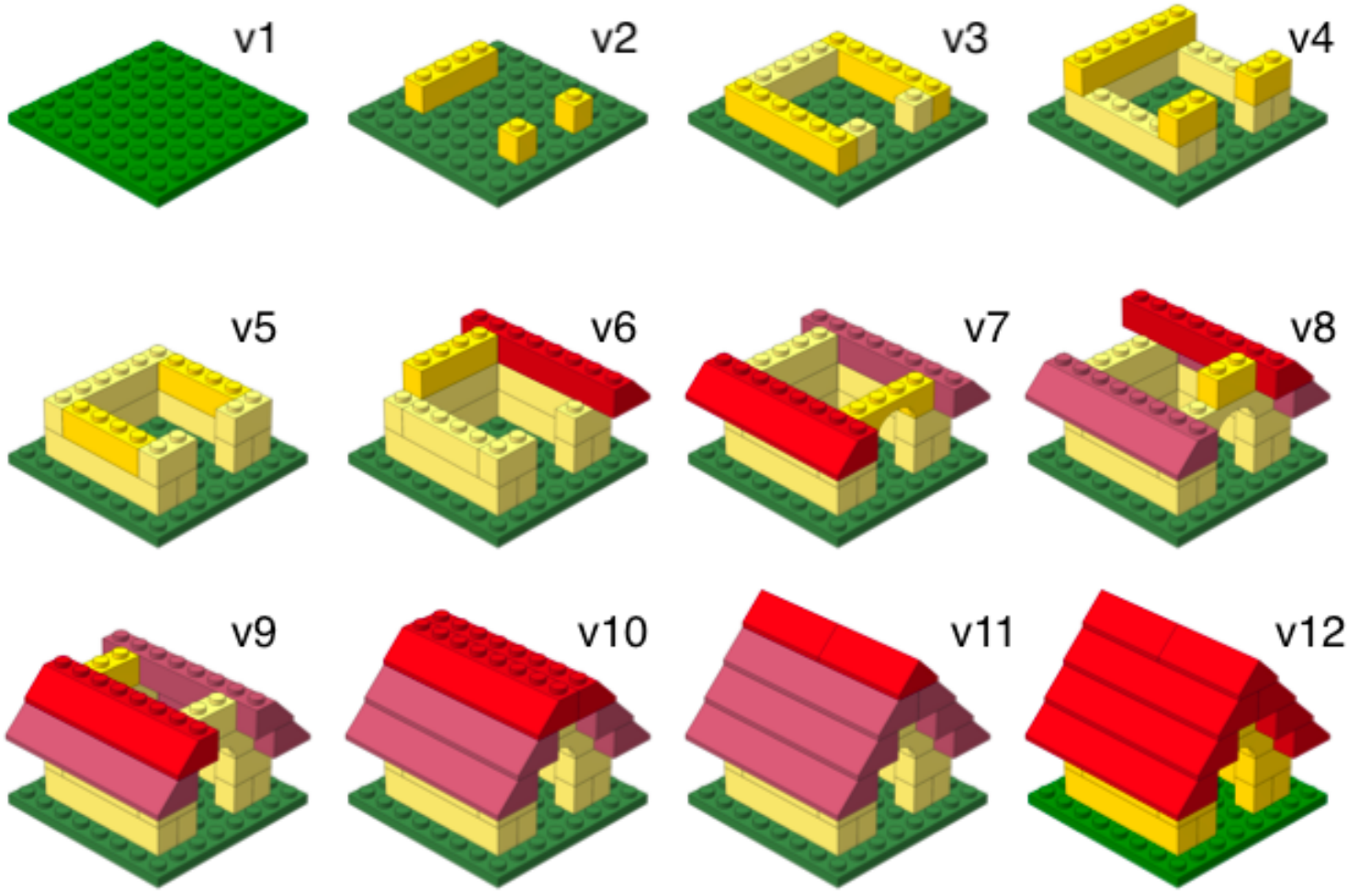
Git and GitHub

- Git is a version control system -- like "Track Changes" features from Microsoft Word, on steroids
- It's not the only version control system, but it's a very popular one

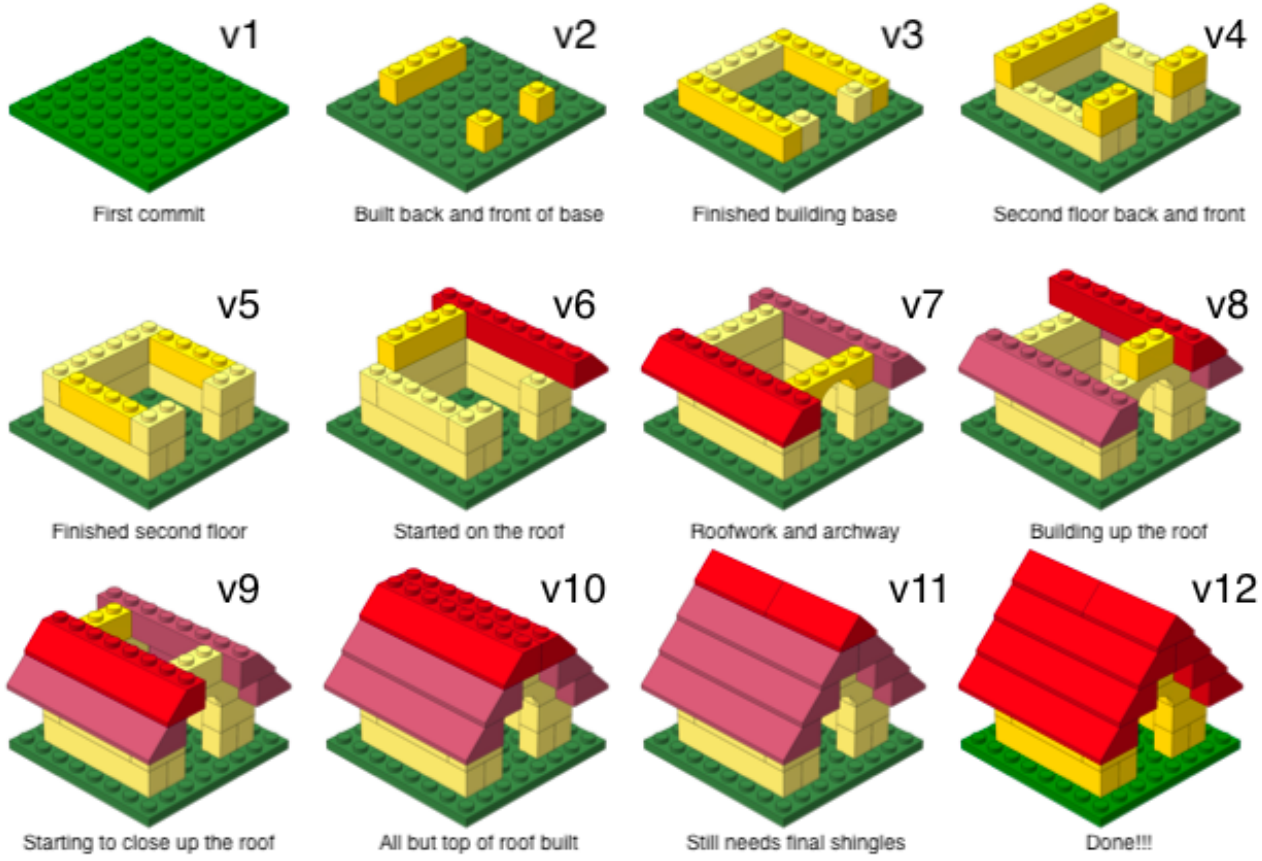
- GitHub is the home for your Git-based projects on the internet -- like DropBox but much, much better

- We will use GitHub as a platform for web hosting and collaboration (and as our course management system!)

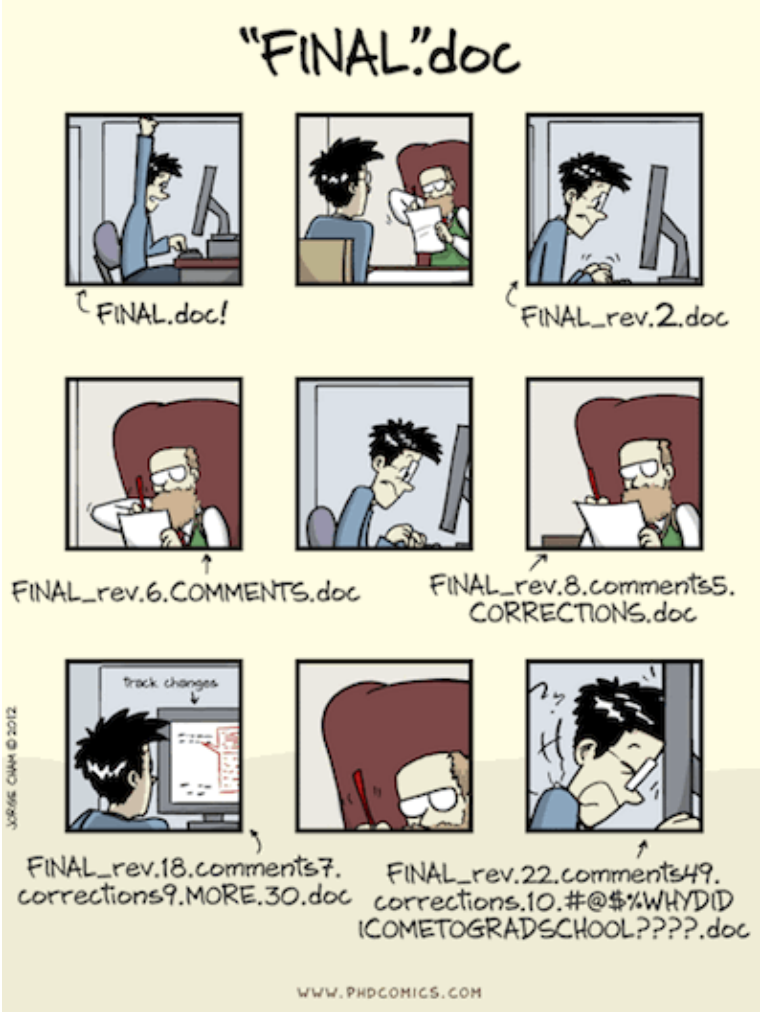
Versioning



Versioning
with human readable messages



Why do we need version control?



How will we use Git and GitHub?

How will we use Git and GitHub?

How will we use Git and GitHub?

How will we use Git and GitHub?

Git and GitHub tips

- There are a lot of git commands but 99% of the time you will use git to `add`, `commit`, `push`, and `pull`.
- We will be doing Git things and interfacing with GitHub through RStudio and GitHub Desktop.
- There is a great resource for working with git and R: happygitwithr.com.

Tour: Git and GitHub

- Create a GitHub account
- Download GitHub Desktop
- Verify your GitHub email
- Adjust your GitHub settings for a more pleasant GitHub experience
    - Settings > Update name and photo