

# Meet the toolkit: version control and collaboration

Data Science in a Box  
[datasciencebox.org](http://datasciencebox.org)



# Course toolkit

## Course operation

- Course website
- Moodle

## Doing data science

- Programming:
  - R
  - RStudio
  - tidyverse
  - R Markdown
- Version control and collaboration:
  - Git
  - GitHub



# Git and GitHub



# Git and GitHub

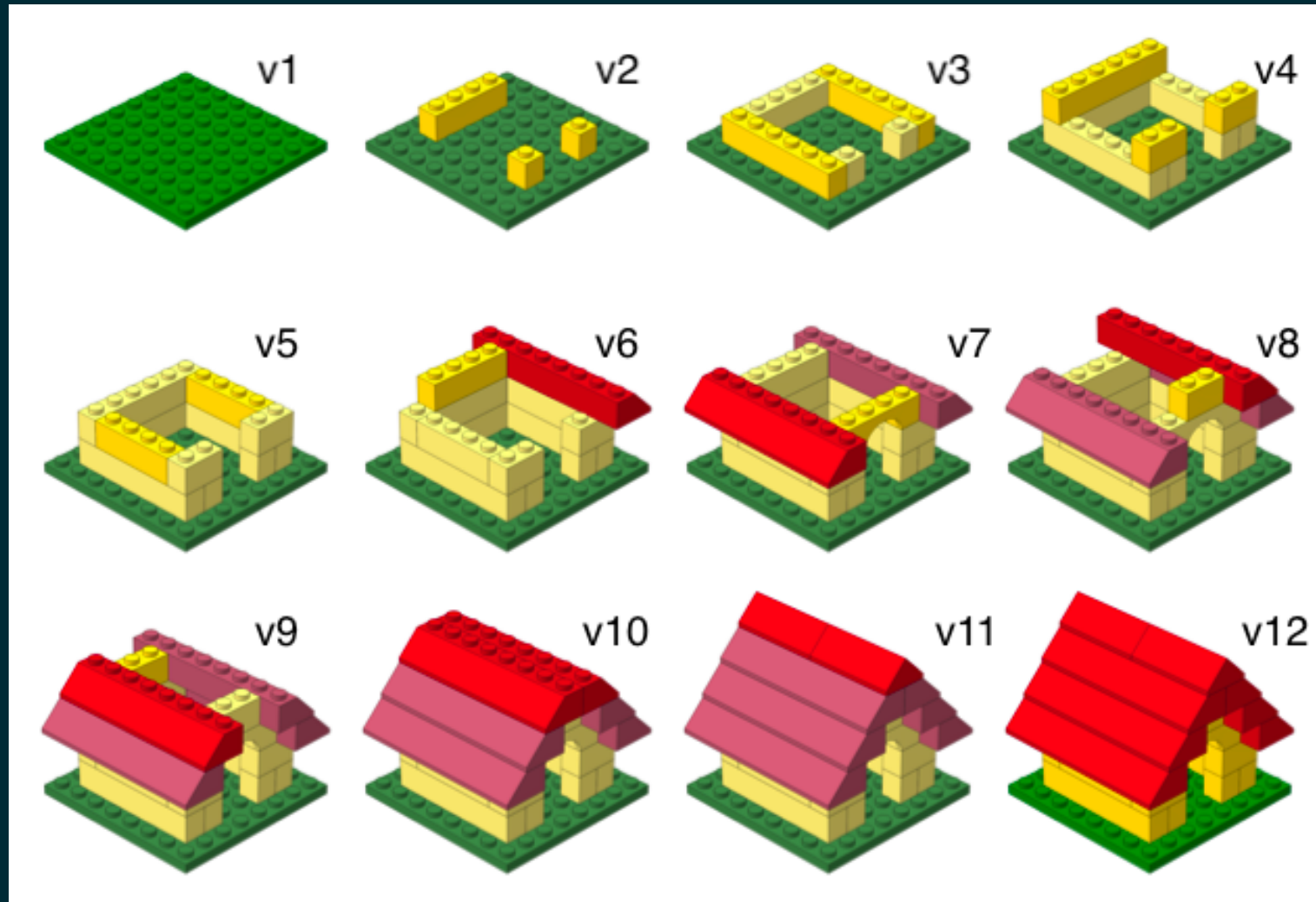


- Git is a version control system -- like “Track Changes” features from Microsoft Word, on steroids
- It's not the only version control system, but it's a very popular one

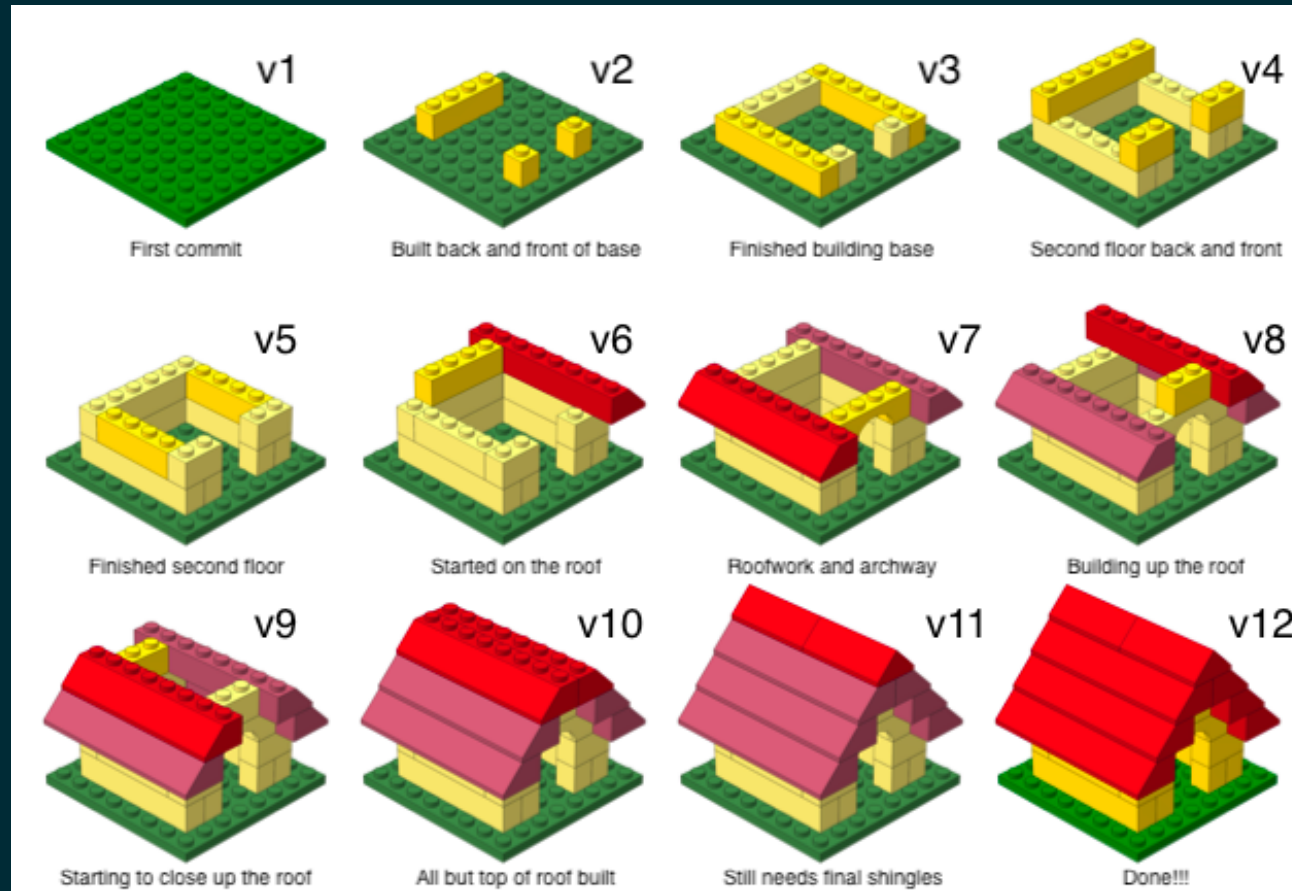
- GitHub is the home for your Git-based projects on the internet -- like DropBox but much, much better
- We will use GitHub as a platform for web hosting and collaboration (and as our course management system!)



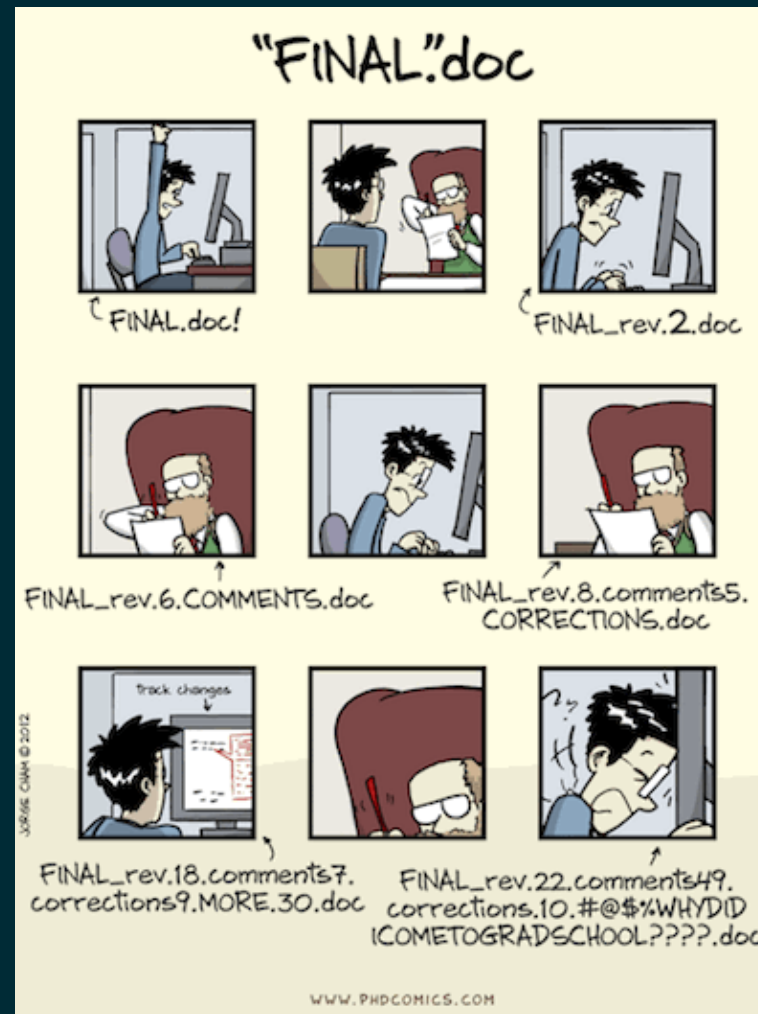
# Versioning



# 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 millions of git commands -- ok, that's an exaggeration, but there are a lot of them -- and very few people know them all. 99% of the time you will use git to add, commit, push, and pull.



# Git and GitHub tips

- There are millions of git commands -- ok, that's an exaggeration, but there are a lot of them -- and very few people know them all. 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, but if you google for help you might come across methods for doing these things in the command line -- skip that and move on to the next resource unless you feel comfortable trying it out.



# Git and GitHub tips

- There are millions of git commands -- ok, that's an exaggeration, but there are a lot of them -- and very few people know them all. 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, but if you google for help you might come across methods for doing these things in the command line -- skip that and move on to the next resource unless you feel comfortable trying it out.
- There is a great resource for working with git and R: [happygitwithr.com](http://happygitwithr.com). Some of the content in there is beyond the scope of this course, but it's a good place to look for help.



# Tour: Git and GitHub

- Create a GitHub account
- Verify your GitHub email
- Adjust your GitHub settings for a more pleasant GitHub experience
  - Settings > Emails > Uncheck "Keep my email address private"
  - Settings > Emails > Update name and photo

*Later...*

*Work with R, RStudio, Git, and GitHub together!<sup>†</sup>*

<sup>†</sup> Just like a real data scientist!



## Your turn: AE 02 - Bechdel + R Markdown

- The Bechdel test asks whether a work of fiction features at least two women who talk to each other about something other than a man, and there must be two women named characters.
- Go to RStudio Cloud and start the assignment AE 02 - Bechdel + R Markdown.
- Open and knit the R Markdown document `bechdel.Rmd`, review the document, and fill in the blanks.

