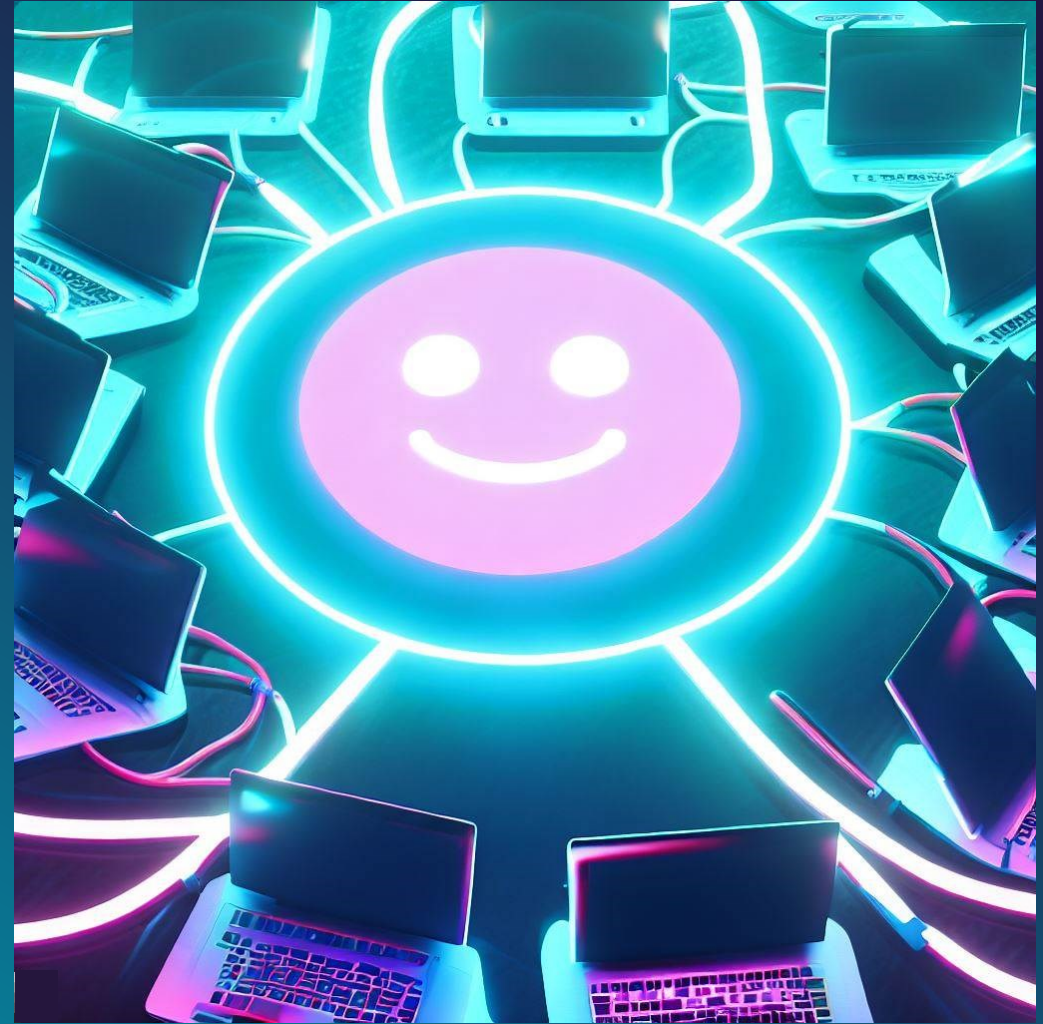


R Projects, Github, and R Markdown



What you're learning

R + RStudio

R Projects

R Markdown

Github

R + RStudio

Statistical programming language

Powerful, flexible tool for data processing, data analysis, and data visualization

Increasingly powerful tools for other programming tasks too

- E.g., formr allows you to generate and manage survey studies *in R*
- Shiny lets you build interactive web apps

100% free, available on PC, Mac, Linux, and the cloud

Valuable skill:

- More competitive for graduate applications
- Increasingly taught and used in PhD programs
- Increasingly used and sought after for industry jobs (e.g., data science/data analyst/UEX-related jobs)

What you're learning

R + RStudio

R Projects

R Markdown

Github

Organization

+

Sharing

Organization

Research generates a large volume of files:

- Raw data
- Processed data
- Data processing scripts
- Analysis scripts
- Analysis reports
- Figures
- Manuscript drafts
- Posters + Talk slides

Contains: 29 Files, 15 Folders

Contains: 68 Files, 16 Folders

Contains: 220 Files, 43 Folders

Contains: 274 Files, 25 Folders

Contains: 780 Files, 69 Folders

Contains: 1,857 Files, 33 Folders












Organization

Projects change over time, files change with them:

- Revisions
- Restarts
- Changes in direction

Computers get lost, stolen, break

- Hard drives fail *all the time*
- Redundant backups of all files is a *must*

 MVP ATT Analysis Script FLETCHCOR 01122016.R	1/12/2016 3:50 PM
 MVP ATT Analysis Script 101620155.R	12/23/2015 4:21 PM
 MVP ATT Analysis Script 101620155 Subsample Loop.R	11/10/2015 12:33 PM
 MVP ATT Analysis Script 101620155 STLT.R	10/17/2015 1:55 PM
 MVP ATT Analysis Script 12262015.R	12/26/2015 3:16 PM
 MVP ATT Analysis Script 12112015 ABS.R	12/26/2015 3:02 PM
 MVP ATT Analysis Script 11052015 FACE.R	11/5/2015 4:42 PM
 MVP ATT Analysis Script 09272015.R	10/9/2015 9:11 AM
 MVP ATT Analysis Script 09252015.R	9/27/2015 4:12 PM
 MVP ATT Analysis Script 04272016 MODELCOMP.R	4/27/2016 1:40 PM
 MVP ATT Analysis Script 01162016 MODELCOMP.R	2/28/2016 1:47 PM

Organization

Success in research *requires* strong organization

- Keep files together
- Keep track of files over time as they evolve
- Keep backups organized and up to date

Less likely to lose important files

Waste less time searching for critical files

Less likely to make mistakes

Sharing

Research is collaborative

You will work with:

- Mentors (graduate students, postdocs, PIs)
- Peers (other grad. Students)
- Local collaborators
- Collaborators at other universities
- International collaborators

Of my last 10 projects, only 1 has been me alone

Same is true in industry

Sharing

You need to:

- Make files available to collaborators
- Keep track of and integrate files as multiple people make changes to them
- Allow people to work simultaneously if needed

Not easy to do with just emailing files back and forth

- Things get lost
- Changes get overwritten
- Collaboration gets slow and frustrating

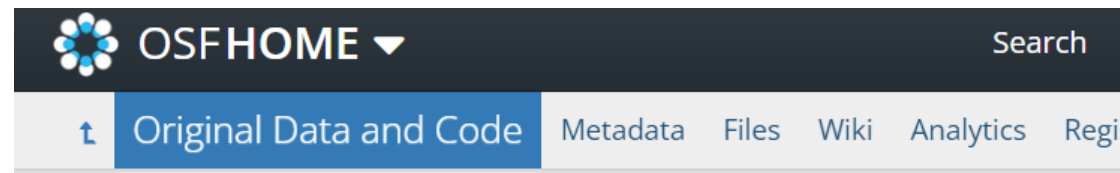
Sharing

Increasing expectation that you make *all* study materials publicly available on publication

[109] Data Falsificada (Part 1): "Clusterfake"

Posted on June 17, 2023 by Uri, Joe, & Leif

```
<c r="R3" i="1"/>  
<c r="S3" i="1"/>  
<c r="R70" i="1"/>  
<c r="R4" i="1"/>  
<c r="S4" i="1"/>  
<c r="R5" i="1"/>  
<c r="S5" i="1" s="1"/>  
<c r="R6" i="1"/>
```



Reducing Dishonesty - Replication(s) /

Original Data and Code

Sharing

Not enough to just make files available

- Need to *explain* your work to others

Collaboration and sharing facilitated by reports

- Word document, PDF, or PPT slides
- Briefly explain and visualize what you've been working on

Helps your colleagues understand the files you've shared with them

- Helps you pick up where you left off

Helps you write your paper

What you're learning

R + RStudio	—————→	Analyze and visualize data
R Projects	—————→	Organize files
R Markdown	—————→	Prepare reports using R code
Github	—————→	Share, backup, and collaborate

All integrate seamlessly with one another

R Projects

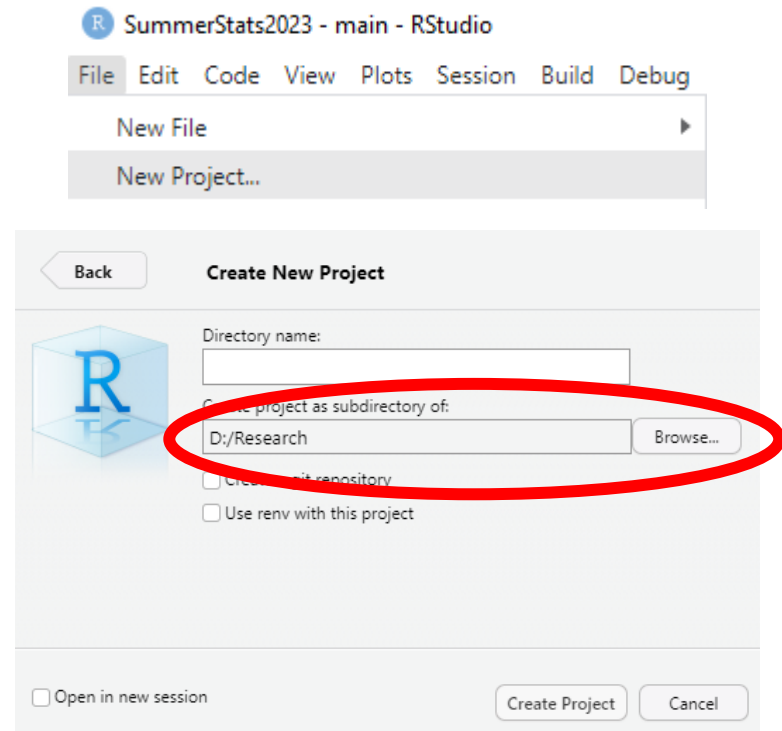
Built-in functionality to Rstudio


Creates a new folder on your computer *dedicated to a research project*

- Creates a .Rproj file that defines the project

Make sure you save this somewhere *safe and easy to find*

- Name the project something clear and memorable



PC > Data (D:) > Research > Example				Search Example	
Name	Date modified	Type	Size		
 Example.Rproj	7/2/2023 12:43 PM	R Project	1 KB		

R Projects

Can open the project through the .Rproj file

- Creates an R session dedicated to your project

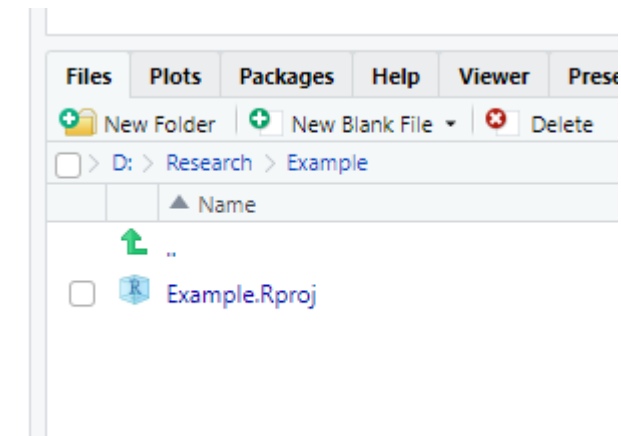
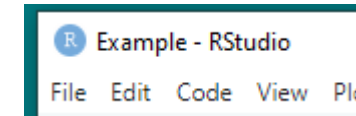
All files in your project folder will be accessible to R

Automatically sets your “working directory” to the project folder

- “Working directory” = where R looks for files
- Normally some unhelpful default location
 - Varies from computer to computer

Can share your project folder with collaborators

- All code will *just work*



```
> getwd()  
[1] "C:/Users/Dconr/Documents"
```

R Markdown

Integrate code, text, and visualization in one document

```
11
12 ## R Markdown Example
13
14 This is text that I'm writing for the class to see.
15
16 {r,echo=F}
17 #Here's some code that won't appear in the document:
18 x<-rnorm(100,0,1)
19 {r}
20
21
22 {r}
23 #This code will appear in the document
24 #And so will the graph
25 hist(x)
26
27
28 I can also combine code and text together. For
   example, the mean of my vector is  $M = \text{round}(\text{mean}(x), 2)$ 
```

R Markdown Example

This is text that I'm writing for the class to see.

```
#This code will appear in the document
#And so will the graph
hist(x)
```

Histogram of x

I can also combine code and text together. For example, the mean of my vector is $M = -0.14$

R Markdown

```
---
title: "Example R Markdown Document"
author: "Daniel Conroy-Beam"
date: "`r Sys.Date()`"
output: word_document
---
```

YAML header: sets the settings for the R Markdown document

```
11
12 ## R Markdown Example
13
14 This is text that I'm writing for the class to see.
15
16
17 ```{r,echo=F}
18 #Here's some code that won't appear in the document:
19 x<-rnorm(100,0,1)
20 ```
21
22 ```{r}
23 #This code will appear in the document
24 #And so will the graph
25 hist(x)
26 ```
27
28 I can also combine code and text together. For
example, the mean of my vector is *M* = `r
round(mean(x),2)`
```

Text: will “render” as ordinary text in your document

- Can format like normal: bold, italicize, hyperlink

Code:

- Code chunk: write R code like normal
 - Can optionally include in or hide from final document
 - Figures automatically output to document (but can hide them if you want)
- Inline R code: write R code *inside text*
 - Executes and displays the output mid-sentence

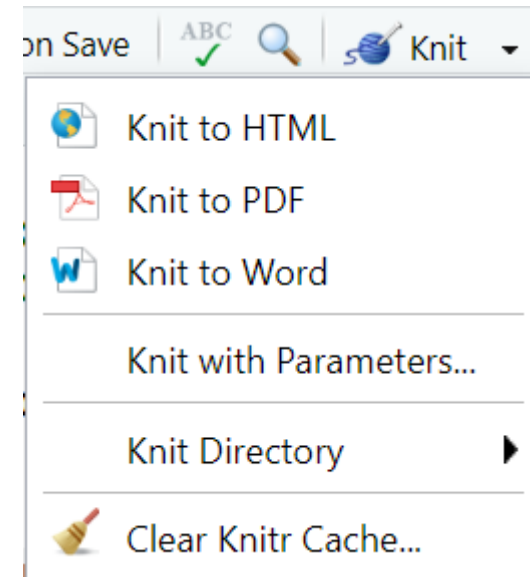
R Markdown

When finished, can “knit” your .Rmd file into document

- HTML file, word document, PDF, powerpoint slides, markdown file, interactive notebook

You can do your analysis/visualization and prepare your report *at the same time*

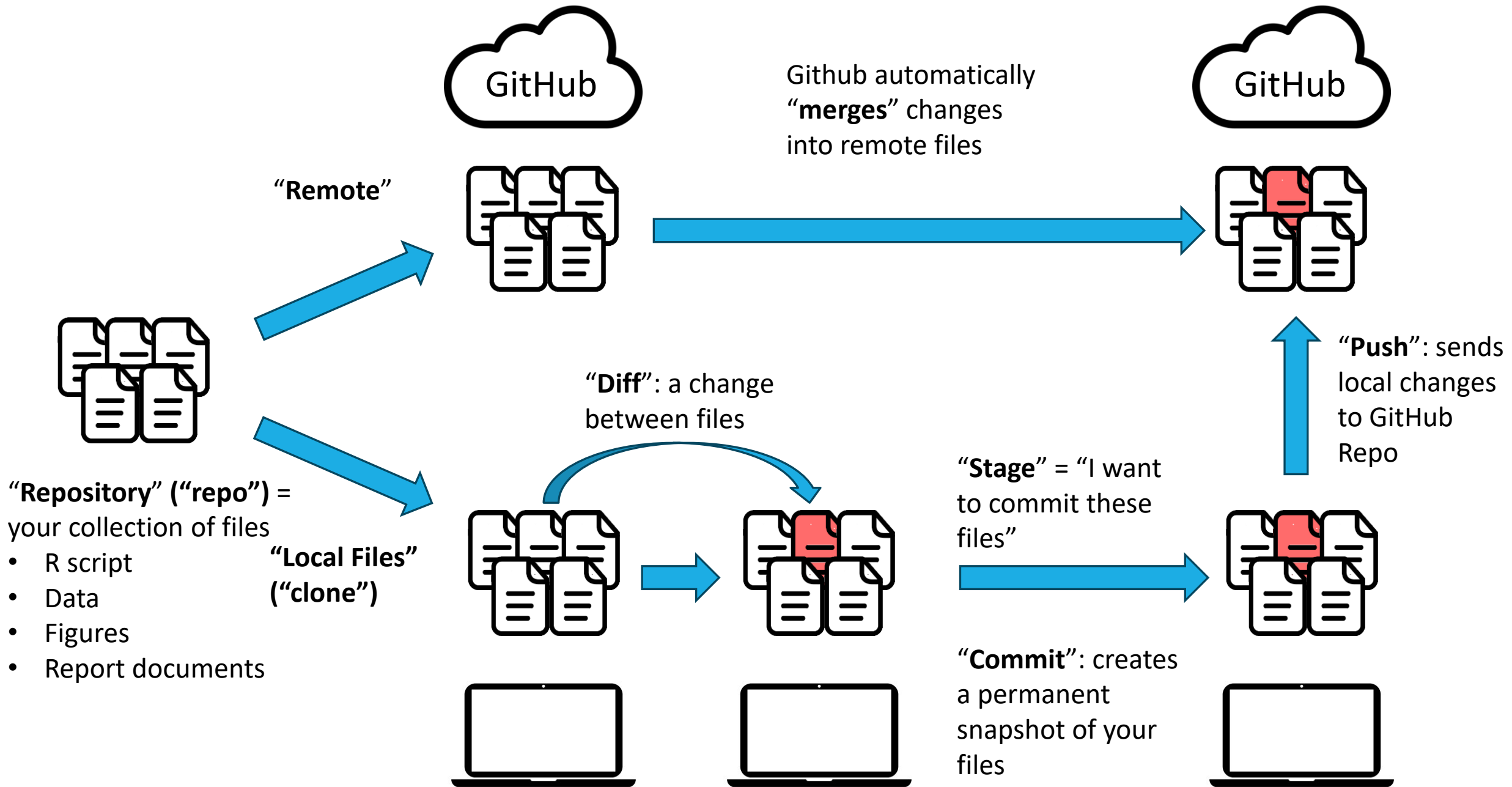
- Could write an entire research paper in R Markdown

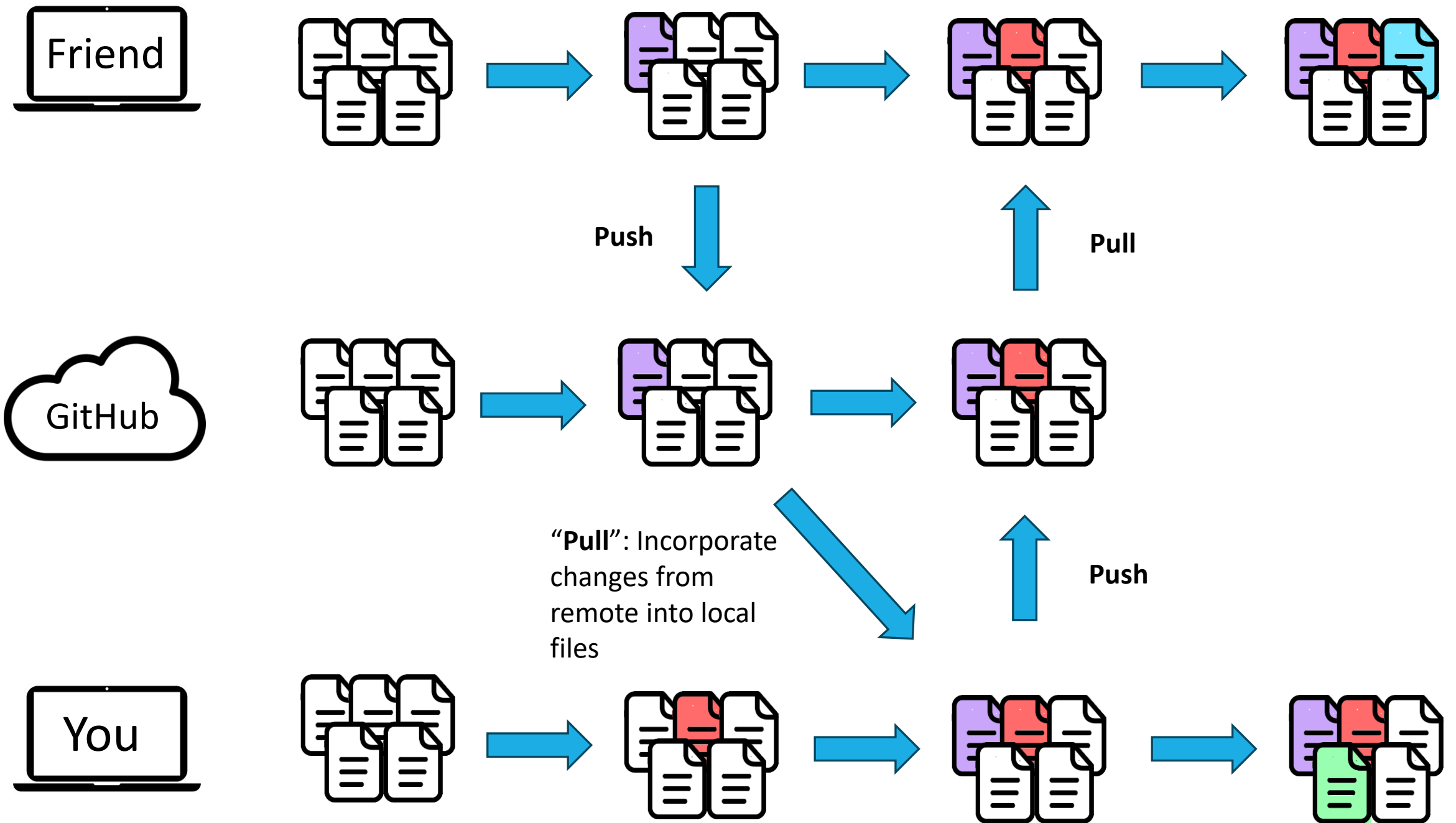


GitHub

Version control system

- Keeps a record of all changes made to a set of files
- Hosts files online for sharing with collaborators or others
- Automatically merges changes to files made across multiple people
- Can render documents for easy sharing of reports online





Using a Git commit is like using anchors and other protection when climbing. If you're crossing a dangerous rock face you want to make sure you've used protection to catch you if you fall. Commits play a similar role: if you make a mistake, you can't fall past the previous commit. Coding without commits is like free-climbing: you can travel much faster in the short-term, but in the long-term the chances of catastrophic failure are high! Like rock climbing protection, you want to be judicious in your use of commits. Committing too frequently will slow your progress; use more commits when you're in uncertain or dangerous territory. Commits are also helpful to others, because they show your journey, not just the destination.

--Hadley Wickham

Simplest Github and R Workflow

Stage and **commit** any time you've completed some work you don't want to lose

- Leave a commit message describing what you've done
- GitHub will create a unique ID for that snapshot of the repo
- You can **always return** to this commit later if you need to

Push once your files are ready to be shared with others

- Pushed files are publicly visible **forever**
- Don't push things you don't want others to see
- **Don't put sensitive information into your repository**

If you are working with someone else:

- Pull before you push
- Look at happygitwithr.com to see what to do with “merge conflicts”

Rendering Documents

Github can also render certain documents as webpages

- readme.md files will automatically be rendered on your repo page
- .html files can be rendered as an independent webpage through “GitHub Pages”

CMCL Summer Statistics Workshop

Instructor: Daniel Conroy-Beam

Email: dconroybeam@ucsb.edu

Office: Psych. East 3809

Tentative Workshop Schedule

Meeting	Topic	Materials
1: Tuesday, July 1	Introduction to R	Meeting 1 Materials
2: Thursday, July 3	R Projects, GitHub, R Markdown, and Data Cleaning	Meeting 2 Materials
3: Tuesday, July 8	Data Cleaning and Processing	Meeting 3 Materials
4: Thursday, July 10	Basic Data Analysis	Meeting 4 Materials

What you're learning

R + RStudio + R Projects + R Markdown + GitHub

Can:

- Process data
- Analyze data
- Create figures
- Share files
- Write reports (word, PDF, powerpoint, HTML)
- Host online
 - Backup files
 - Facilitate collaboration
 - Linkable website for sharing reports

Can do all of this in **one software: RStudio**

All 100% free

Marketable skills for academia and industry

Data Cleaning

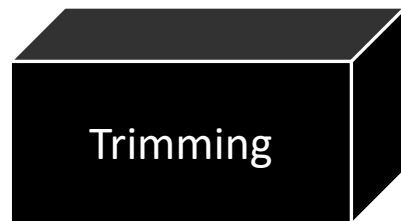


Tidy Data

Hadley Wickham
RStudio

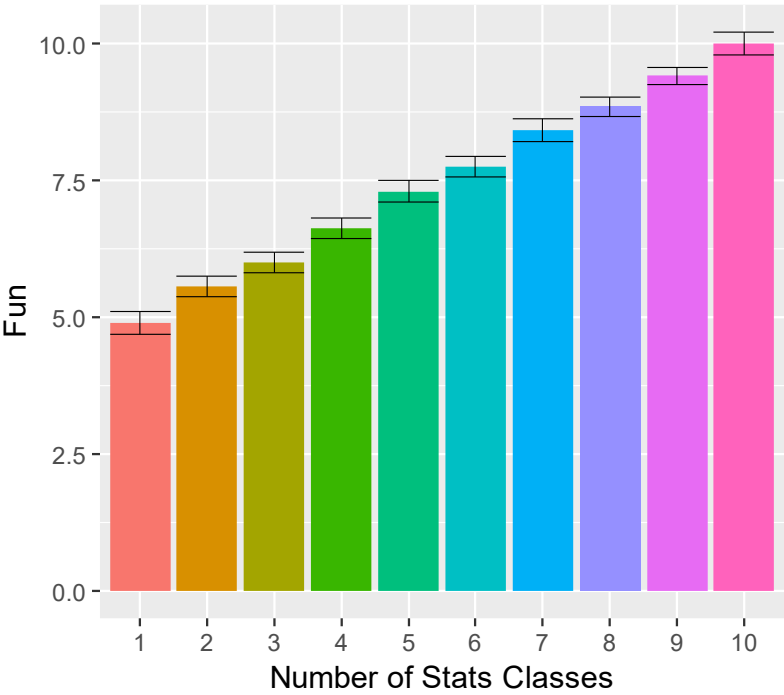
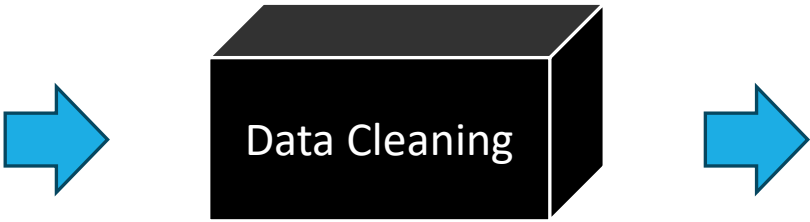
Martin Monkman

*The Data Preparation
Journey*
Finding Your Way With R

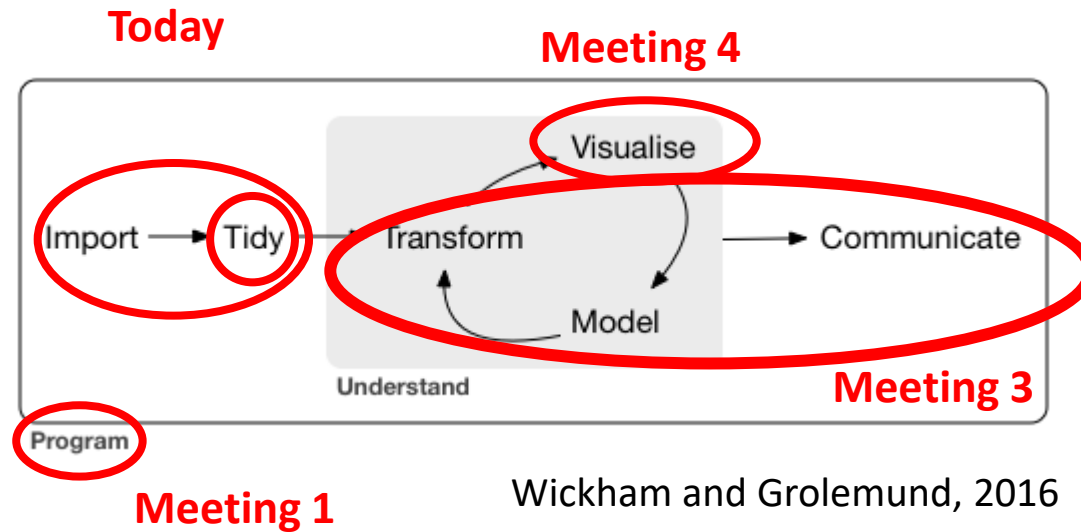


	var1	var2	var3	var4	var5	var6
1	5.082520	5.907829	7.073734	7.884836	6.324818	8.437262
2	5.443151	7.327779	6.050779	5.991775	5.954487	8.307324
3	4.024609	4.881766	2.734992	7.532117	8.664693	5.746925
4	4.480013	5.121075	5.930708	8.619031	7.977853	7.893341
5	3.973761	6.783843	6.567947	7.843699	6.626595	8.080891
6	3.582721	5.856451	4.185818	4.937832	8.730098	7.508317
7	5.335790	3.517303	5.621268	7.377966	8.066959	7.014396
8	5.191268	4.818285	5.429966	7.181727	6.207386	7.291512
9	6.403909	5.418646	5.696535	6.638591	7.451374	8.104118
10	5.131530	6.384721	5.296592	6.585722	8.171418	7.812140
11	5.765037	6.683692	4.739766	7.428553	6.292216	6.575642
12	5.361895	4.621512	7.468145	7.290700	9.264005	8.643813
13	4.421865	5.503943	7.108561	7.322792	6.871867	7.731232
14	5.194853	6.172803	5.393268	6.906846	9.319122	8.511639
15	4.935410	4.590636	6.746634	5.931916	5.377686	7.150316

Raw Data



Analysis



Tidy data:

- 1. Every variable has its own column
- 2. Every observation has its own row
- 3. Every type of data has its own dataframe

Every cell should tell you exactly one piece of information about exactly one observation of the world

Like families, tidy datasets are all alike but every messy dataset is messy in its own way.

Wickham, 2014

Data Cleaning Guidelines

Monkman (2023):

- Complete: ~represent the population as much as possible
- Consistent: be systematic
 - Name your variables according to a consistent system
 - Keep your variables in consistent units
 - Structure your data systematically
- Accurate: free of errors
 - Data entry errors
 - Data processing errors

Data Cleaning Guidelines

Use an automated process

- **Process nothing by hand**
- Hand processing leads to errors
- Does not produce receipts

Should yield:

- Raw data (**completely** unedited)
- Processed data
- Data processing script
- Ideally: codebook

Hongbo Yu

Moral emotions (e.g., guilt, gratitude), affective neuroscience, social psychology

[YES lab](#)



Week 3 Example Data

$N = 720$ Participants ($n = 360$ women)

Each participant paired with a sham partner

Learn one piece of information about partner. Either:

- Morally good: e.g., they volunteer their time at a soup kitchen
- Morally bad: e.g., they occasionally steal money from an elderly family member

Puzzle task:

- Must solve a puzzle in under 1 minute or partner receives electric shock
- Either participant or partner attempts puzzle
- Shock is either low, medium, or high voltage
- Puzzle is unsolvable (partner always gets shocked)

Week 3 Example Data

Measures:

Guilt:

- “How much guilt do you feel toward your partner?”
- 7-point Likert Scale
- Some missing data due to computer failure

Generosity:

- Participant asked how they would divide \$10 between themselves and their partner
- Two time points: (1) start of experiment and (2) after shock

Personality:

- Ten-Item Personality Inventory (TIPI)
- 10 7-point Likert scales
- Measures Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism
 - 2 items each
- Half of items reverse-coded

Ten-Item Personality Inventory-(TIPI)

Here are a number of personality traits that may or may not apply to you. Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement. You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other.

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

I see myself as:

1. ____ Extraverted, enthusiastic.
 2. ____ Critical, quarrelsome.
 3. ____ Dependable, self-disciplined.
 4. ____ Anxious, easily upset.
 5. ____ Open to new experiences, complex.
 6. ____ Reserved, quiet.
 7. ____ Sympathetic, warm.
 8. ____ Disorganized, careless.
 9. ____ Calm, emotionally stable.
 10. ____ Conventional, uncreative.
-

TIPI scale scoring (“R” denotes reverse-scored items):

Extraversion: 1, 6R; Agreeableness: 2R, 7; Conscientiousness: 3, 8R; Emotional Stability: 4R, 9;

Openness to Experiences: 5, 10R.

Predictions

People will feel more guilty and be more generous when:

- They caused the shock (vs. partner)
- The partner is morally good (vs. morally bad)
- The shock is stronger (vs. weaker)

Women will report more guilt and be more generous than men

Agreeable people and neurotic people will be more guilty and more generous