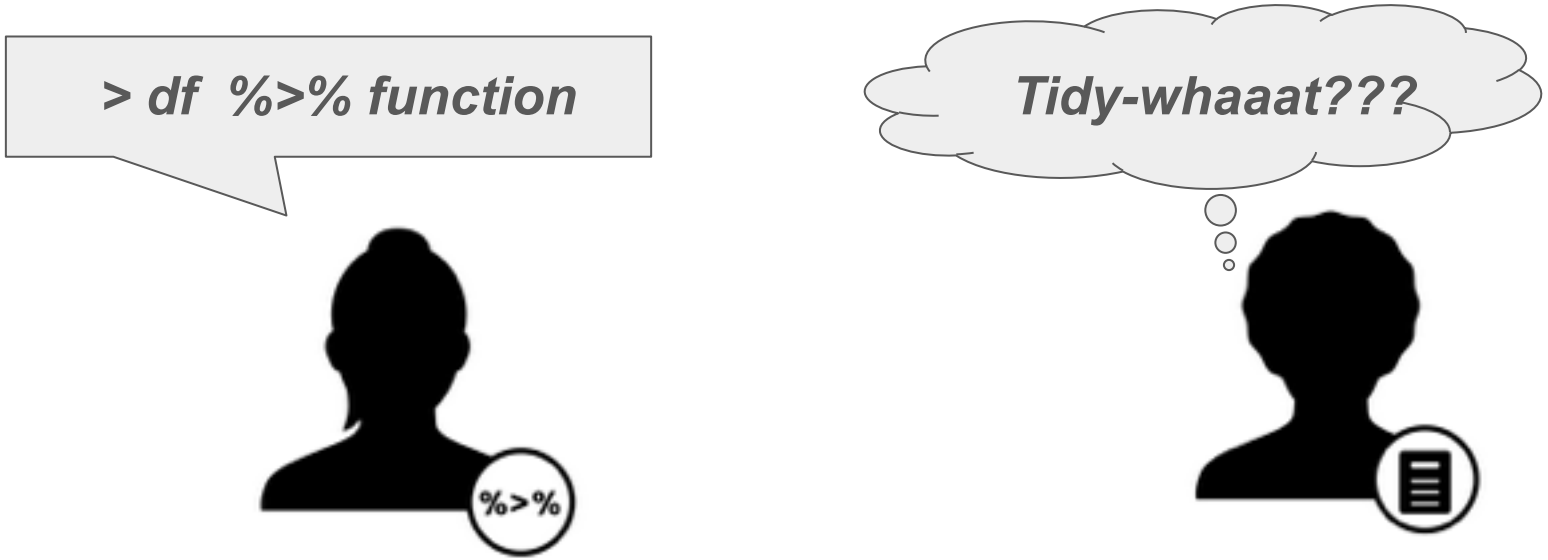
Integrating R & Google Drive for Collaborative Research

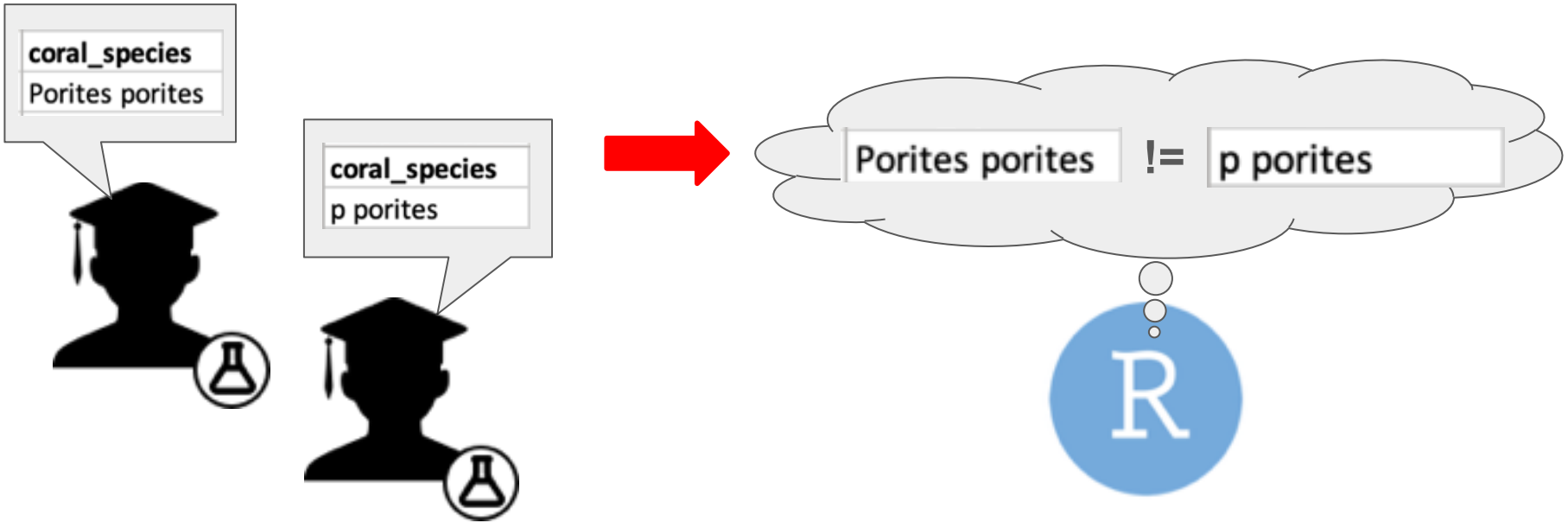
CAL POLY
Center for Coastal Marine Sciences

H. Rempel, K. Bodwin, B. Ruttenberg

Challenges: *Collaborators vary in knowledge of data management practices, R and/or Git literacy*



Challenges: *Consistent data entry*



Google Drive for Project Management

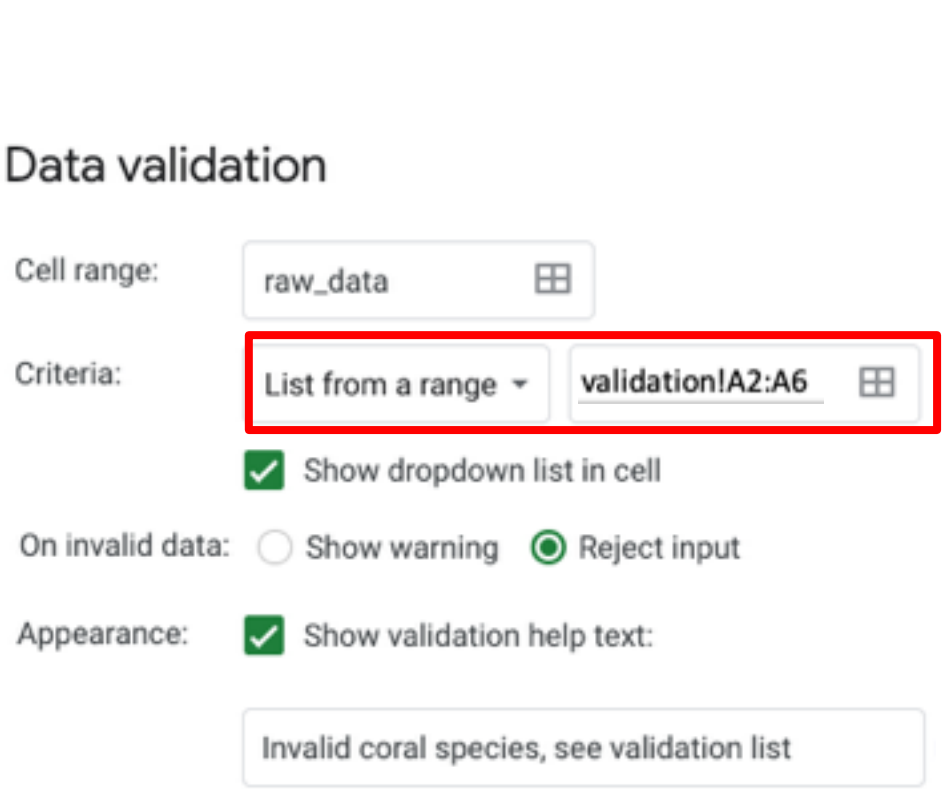
- **Free, familiar & user-friendly**

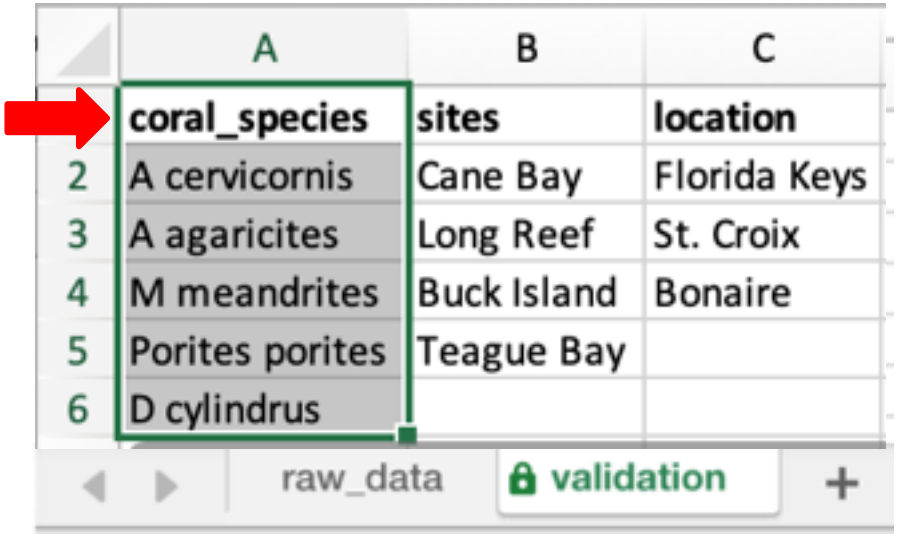
→ Not everyone is comfortable making the initial commit to Git

Google Drive for Project Management

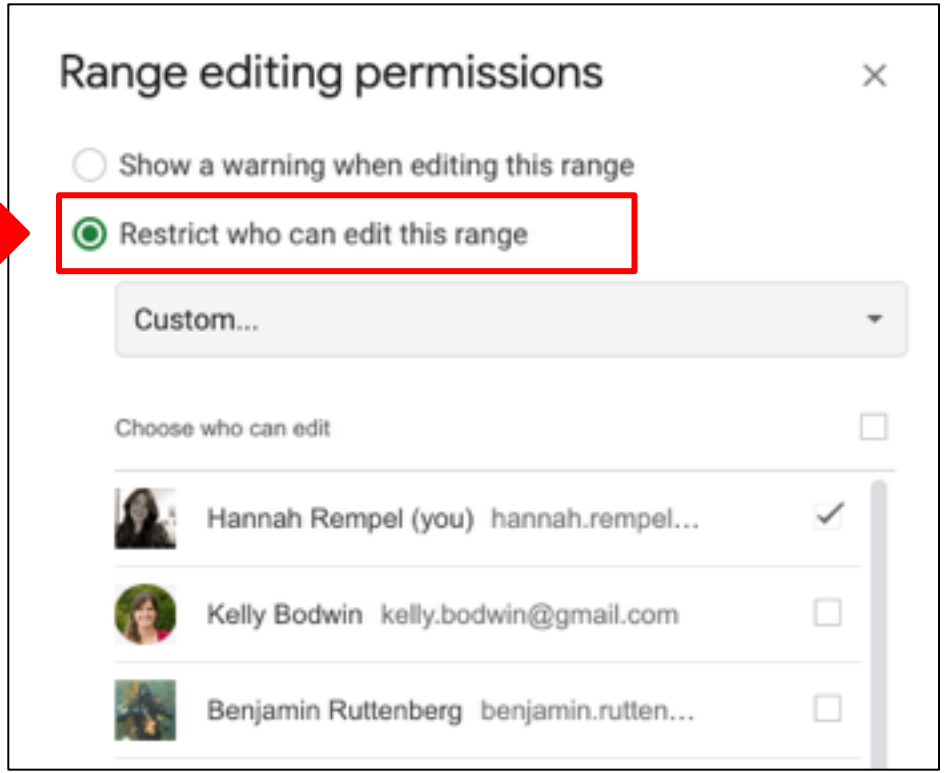
- Free, familiar & user-friendly, with version control
- **Version control, file backup & simultaneous editing**
- **Google sheets tools save time cleaning data in R**

Tools: Data validation– work smarter, not harder





Tools: Restricting User Access to Data Editing



Tools: Restricting User Access to Data Editing

A	B	C	D	E	F	G	H
Photo_analyst	**Surveyor**	**Site**	**Survey**	**Bite_scar**	**Bite_scar_age**	**Bite_type**	
rachael l	madelyn roycroft	cb nearshore	1	yes	new	spot	
peter vd b	ali wolman	bucksofr	2	no	old and healing	focus	
hannah r	hannah rempel	long reef		Ask_Hannah	old with algae	Ask_Hannah	
ethan m	maurice goodma	bucklagoon			old with some healing and some algae		
tara h	ben ruttenberg	cb offshore			other_add note		
sierra s		salt river			Ask_Hannah		
		carambola					
		north star					
		teague bay					
		bucksc					

than | OANN_Rachael | OANN_Savannah | Metadata | Validation_DoNotModify

Easy Drive to R workflow: *'googlesheets'* package

```r
# reading in data
my_google_sheet <- gs_key("1vfAgxo6ikg")

# viewing worksheets / data frames in sheet
gs_ws_ls(my_google_sheet)

# extracting worksheet / data frame
my_df <- as.data.frame( gs_read (ss = my_google_sheet,
                                 ws = "my_work_sheet",
                                 col_names=T))
```

Easy Documentation

Meta documentation of Caribbean Parrotfish 'Net Impacts' project
Project start: 2015
Project end: TBD, ongoing through 2019

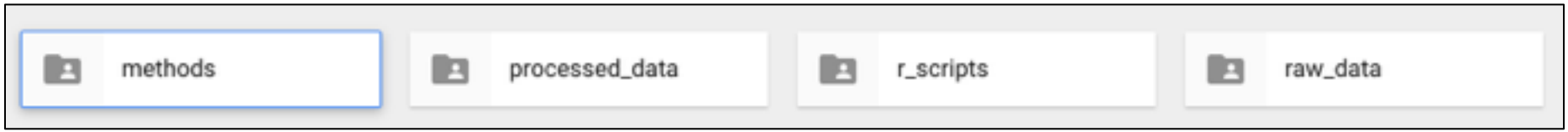
I. OVERVIEW
All project materials (raw data, methods, analyses and written work) can be found in the *Parrotfish Net Impacts* folder. This document details the organization structure for that folder..

II. PROJECT FILE ORGANIZATION

- *NOTES:*
  - Any folder containing **'raw'** in the title denotes raw data. **Do not modify raw data**, and data cleaning should be conducted in R, with a documented script.
  - Within this document, all folders are listed in blue italics, > denotes nested folders. All hyperlinks will take you to the linked Google Folder.
- *Parrotfish Net Impacts* contains the following subsections:
  - *Communicating Science,*
  - *Field Work Materials,*
  - *Manuscripts and Pubs,*

Easy Documentation

	A	B	C	D
1	**script**	**description**	**relevant data files**	**methods reference**
2	raw_coral_merge.rmd	*merging coral surveys from multiple islands and years*	coral_surveys	coral_survey_protocol.gdoc
3	prelim_coral_stats.rmd	*preliminary analysis of coral data normality and patterns*	coral_survey_merged_df	coral_survey_protocol.gdoc



Drive to Git Interface

- relevant_literature
- meeting_notes
- manuscripts_and_pubs
- funding_proposals
- field_work_materials
- data_collection_and_analyses
- .gitignore.txt

→ *You can still store scripts on GitHub*

Summary

- Data management is difficult– **especially with collaborators of varying data science literacy**
- Google Drive is **free, familiar and user-friendly**
  - Sheets has useful tools for upstream data management
  - *Data science doesn't need to be rocket science*

Questions?