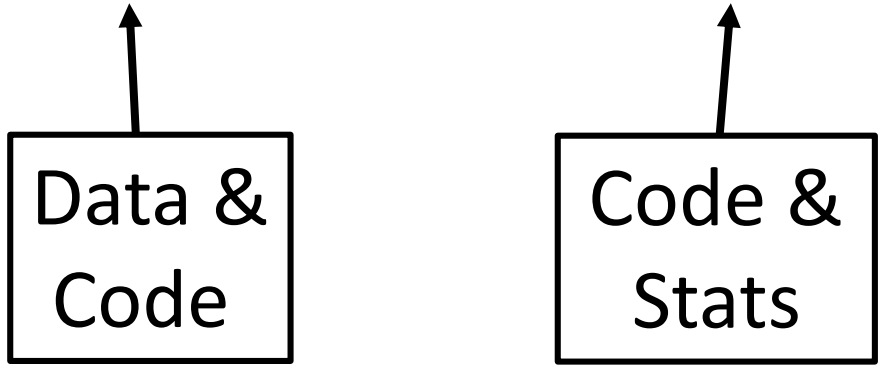
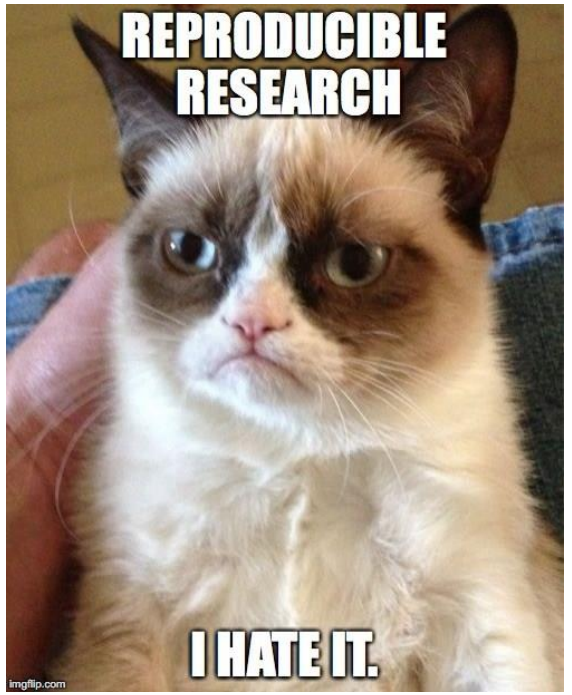
Reproducible Research

Reproducible = Accessible + Readable

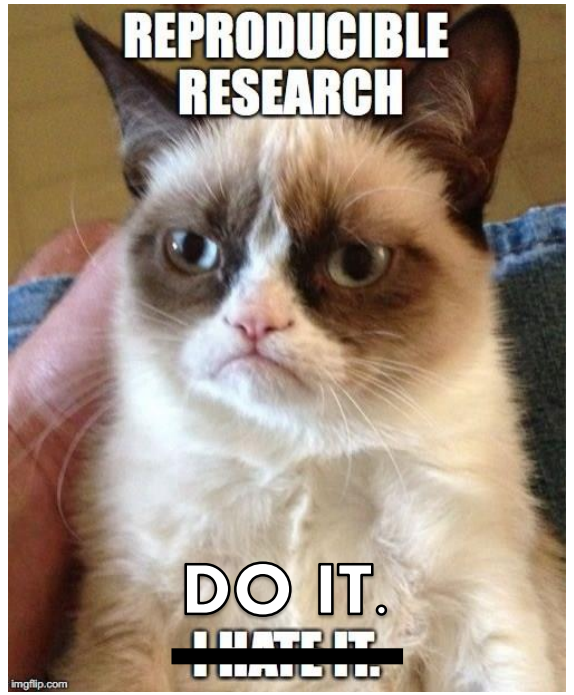


Why make your research reproducible?



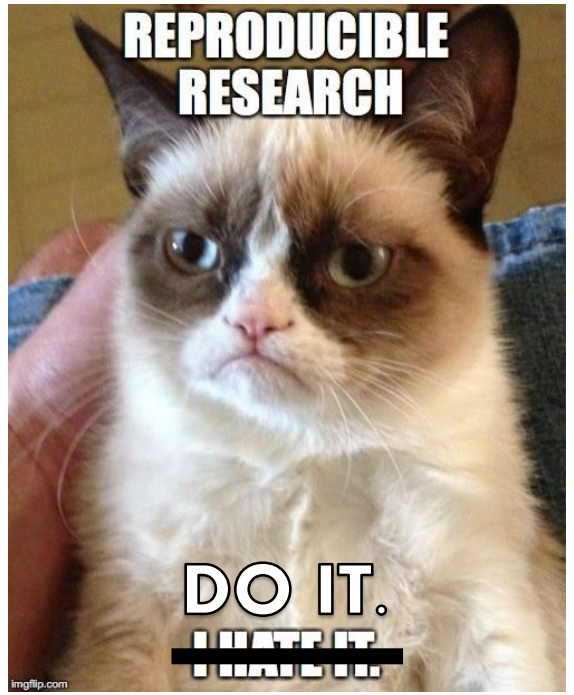
Why make your research reproducible?

- Remember what you did last week, month, year…



Why make your research reproducible?

- Remember what you did last week, month, year…
- Share progress with advisor



Why make your research reproducible?

- Remember what you did last week, month, year…
- Share progress with advisor
- Scientific progress – no need to reinvent the wheel
  - Record statistical decisions, analysis pipelines
  - Share progress and products with the lab

Why make your research reproducible?

- Build collaborations and increase the impact of your science
  - Shared code – examples: Ross Cunning, Hollie Putnam, Murat Eren

jrcunning **Ross** **Cunning**

Follow

Coral reef conservation research @SheddAquarium. Bioinformatics, stats, modeling, data viz, open science, Rstats. Formerly @CoralReefFuturesLab, @GatesCoralLab

Chicago, IL

hputnam **Hollie** **Putnam**

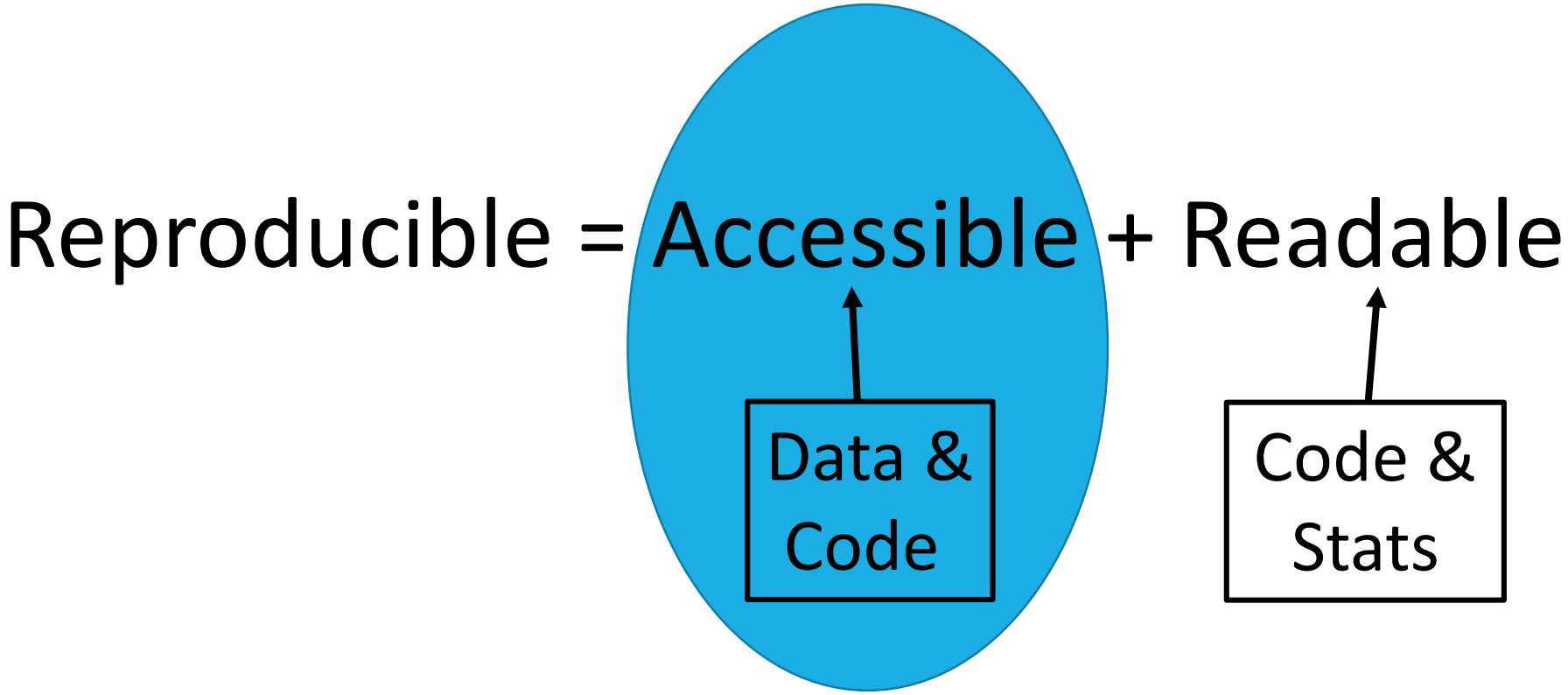
Follow

Rhode Island hputnam@uri.edu

Meren Lab

Meren Lab

University of Chicago http://merenlab.org meren@uchicago.edu


Reproducible =
Accessible
+ Readable
Data &
Code
Code &
Stats

Accessibility: Tools and Strategies

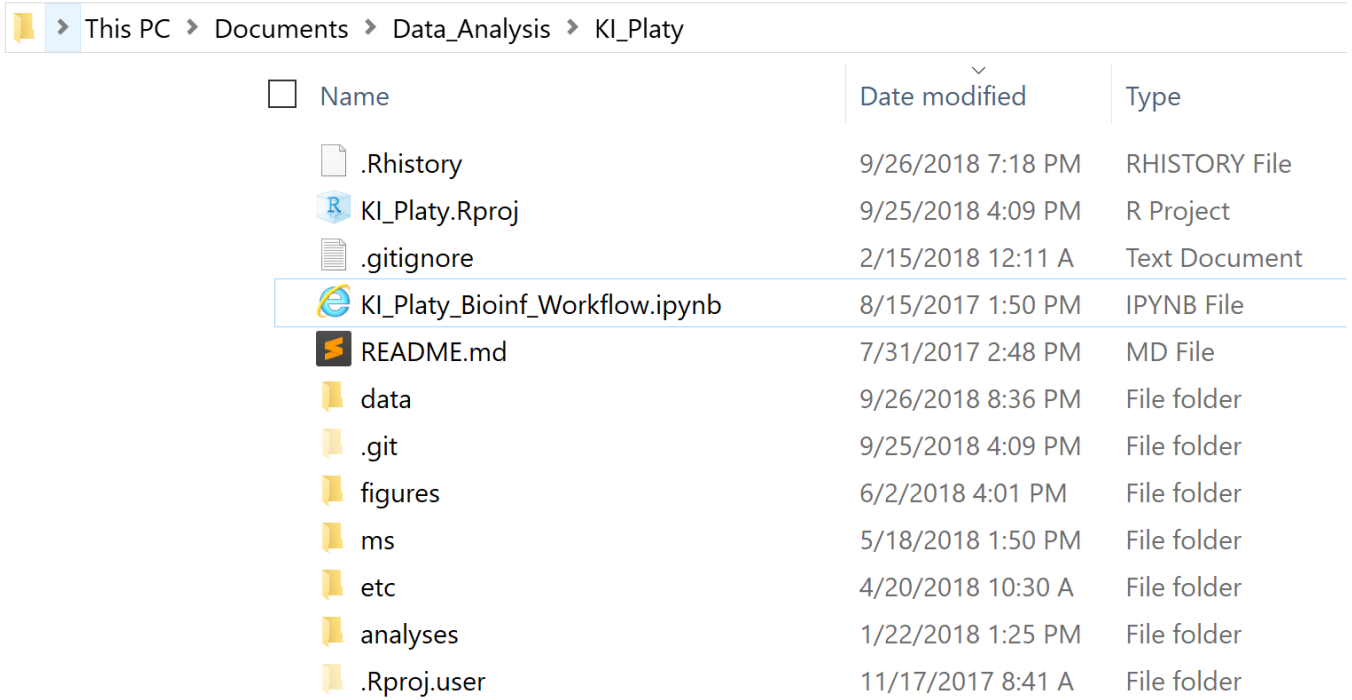
1. Consistent organization strategies

Accessibility: Tools and Strategies

1. Consistent organization strategies
   - Can you find that one figure?

Accessibility: Tools and Strategies

1. Consistent organization strategies
   - Can you find that one figure?
   - File structure and naming



Accessibility: Tools and Strategies

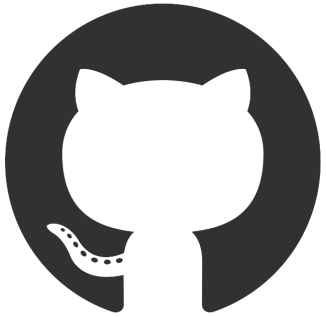
1. Consistent organization strategies
   - Can you find that one figure?
   - File structure and naming
2. Data accessibility
   - Extremely specific README files

Accessibility: Tools and Strategies

1. Consistent organization strategies
   - Can you find that one figure?
   - File structure and naming
2. Data accessibility
   - Extremely specific README files
   - Preserve raw data set, process data and make changes in R

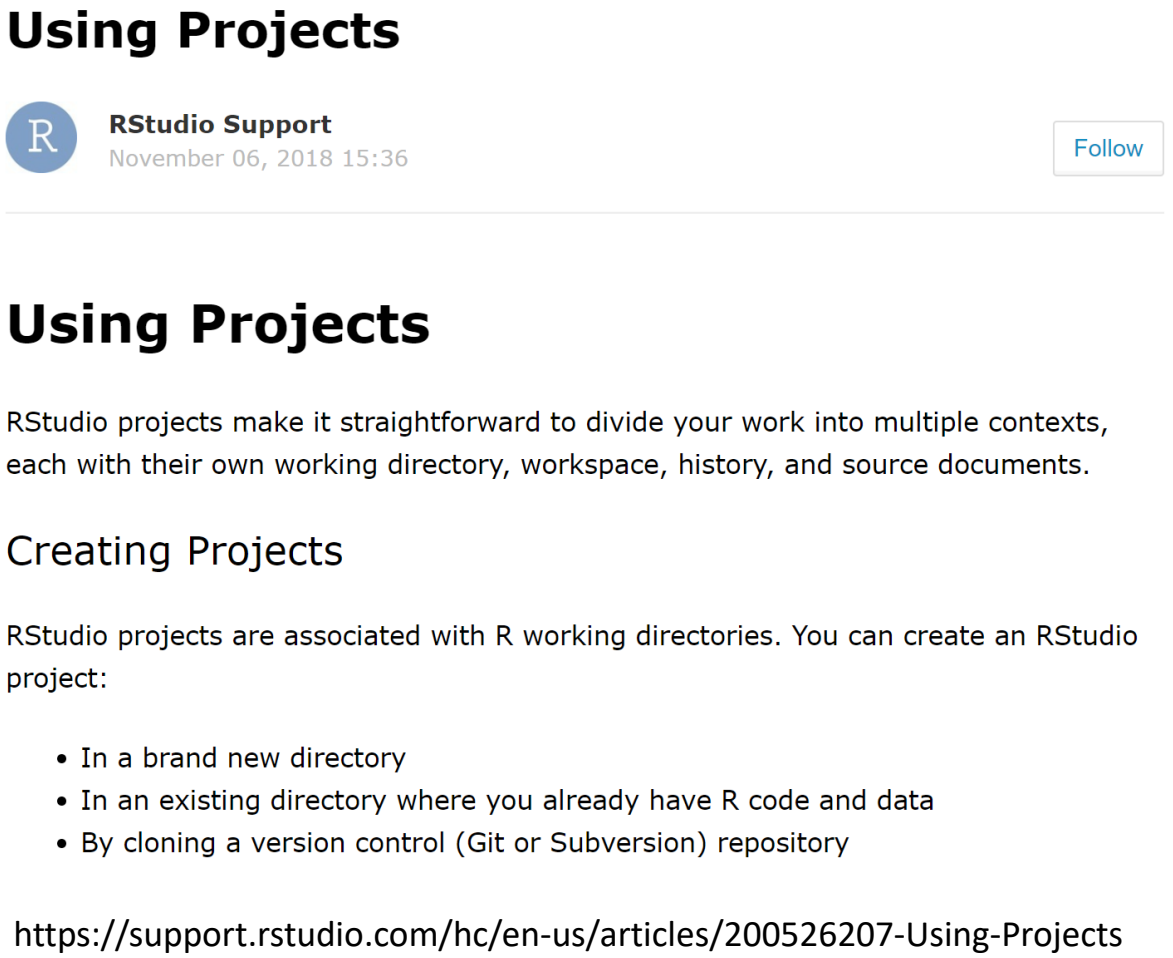
Accessibility: Tools and Strategies

1. Consistent organization strategies
   - Can you find that one figure?
   - File structure and naming
2. Data accessibility
   - Extremely specific README files
   - Preserve raw data set, process data and make changes in R
   - Make data available online: GitHub, Gen Bank, publish online, etc.



Accessibility: Tools and Strategies

3. R Projects
   - Pick up where you left off



Using Projects

RStudio Support
November 06, 2018 15:36

Follow

Using Projects

RStudio projects make it straightforward to divide your work into multiple contexts, each with their own working directory, workspace, history, and source documents.

Creating Projects

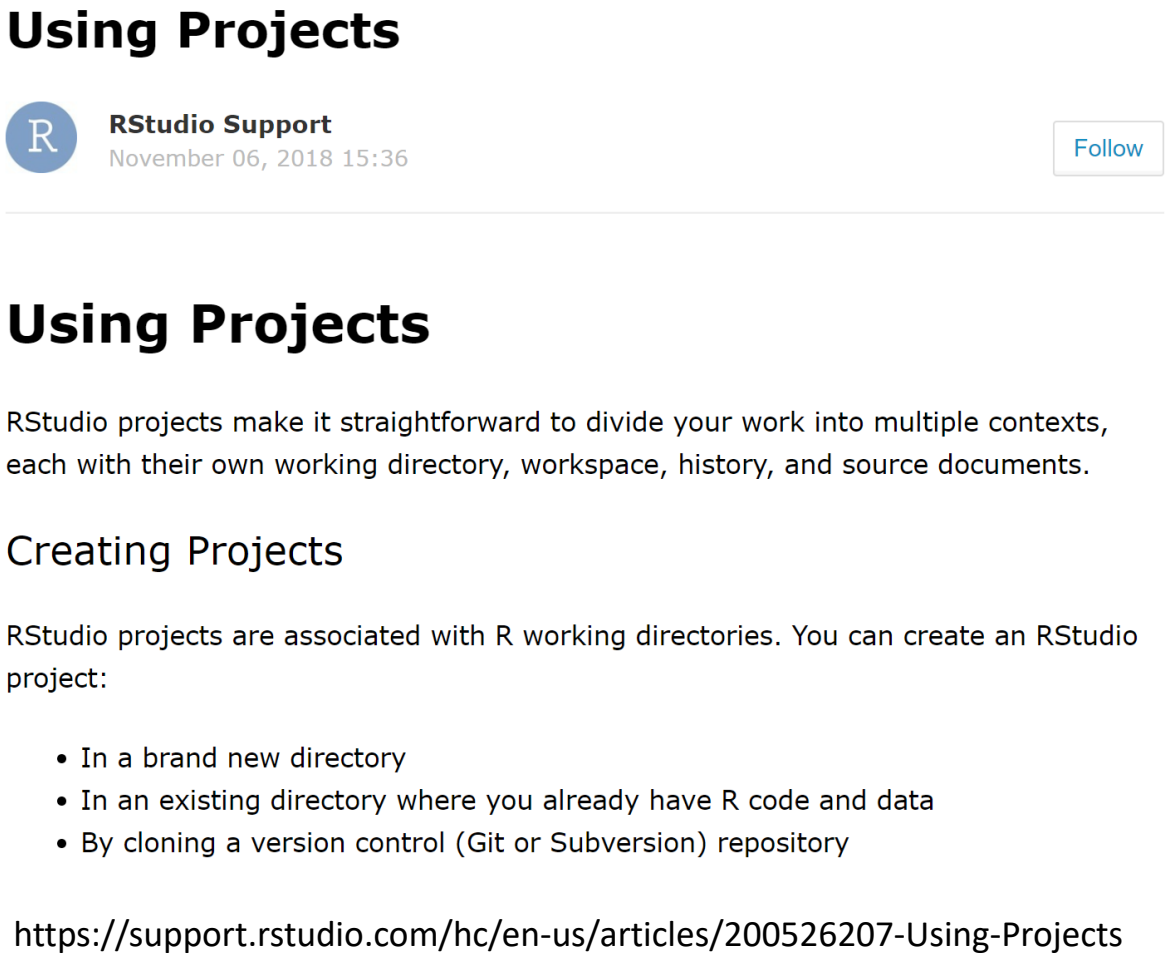
RStudio projects are associated with R working directories. You can create an RStudio project:

- In a brand new directory
- In an existing directory where you already have R code and data
- By cloning a version control (Git or Subversion) repository

https://support.rstudio.com/hc/en-us/articles/200526207-Using-Projects

Accessibility: Tools and Strategies

3. R Projects
   - Pick up where you left off
   - Avoid `setwd()` in every script



Using Projects

RStudio Support
November 06, 2018 15:36

Follow

Using Projects

RStudio projects make it straightforward to divide your work into multiple contexts, each with their own working directory, workspace, history, and source documents.

Creating Projects

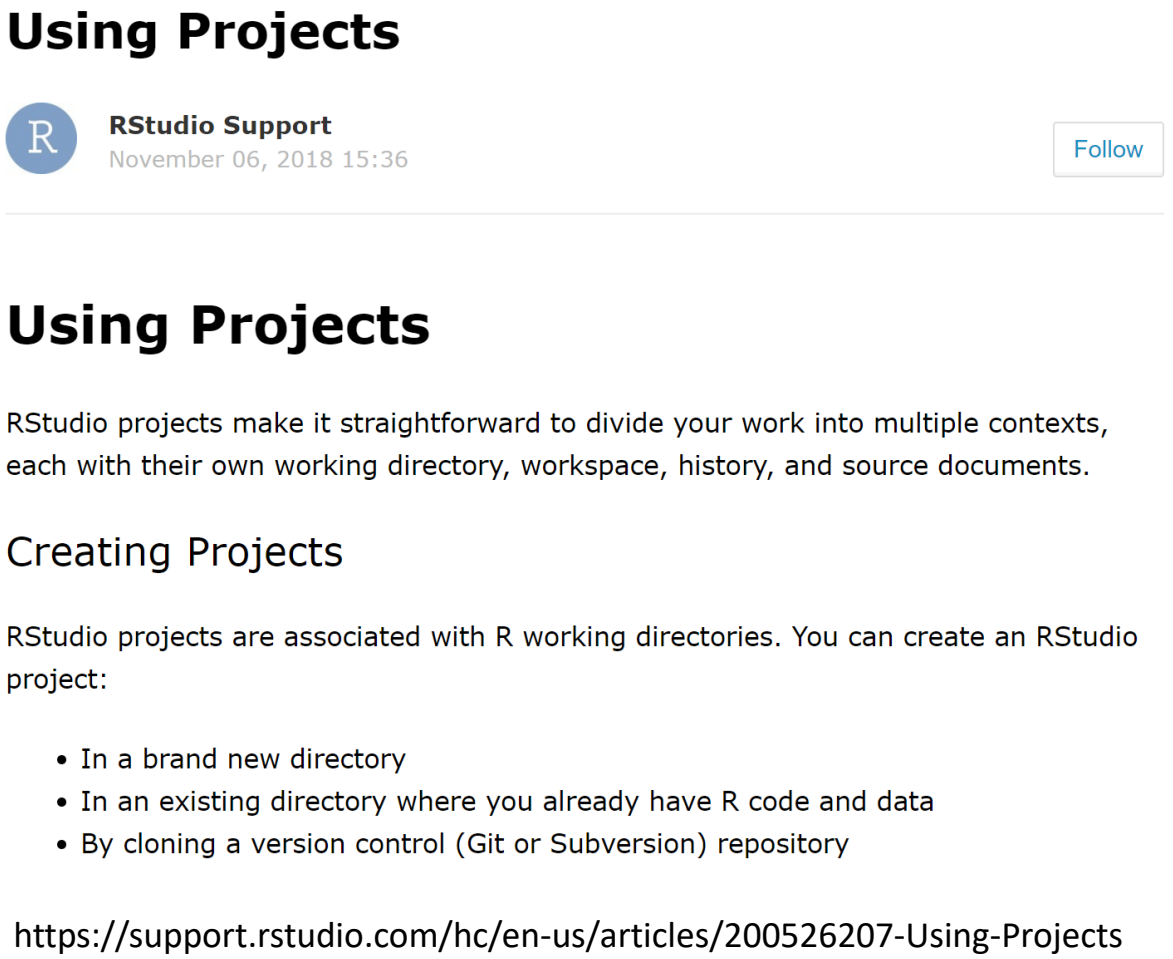
RStudio projects are associated with R working directories. You can create an RStudio project:

- In a brand new directory
- In an existing directory where you already have R code and data
- By cloning a version control (Git or Subversion) repository

https://support.rstudio.com/hc/en-us/articles/200526207-Using-Projects

Accessibility: Tools and Strategies

3. R Projects
   - Pick up where you left off
   - Avoid `setwd()` in every script
   - Easier for collaboration



Using Projects

RStudio Support
November 06, 2018 15:36

Follow

Using Projects

RStudio projects make it straightforward to divide your work into multiple contexts, each with their own working directory, workspace, history, and source documents.

Creating Projects

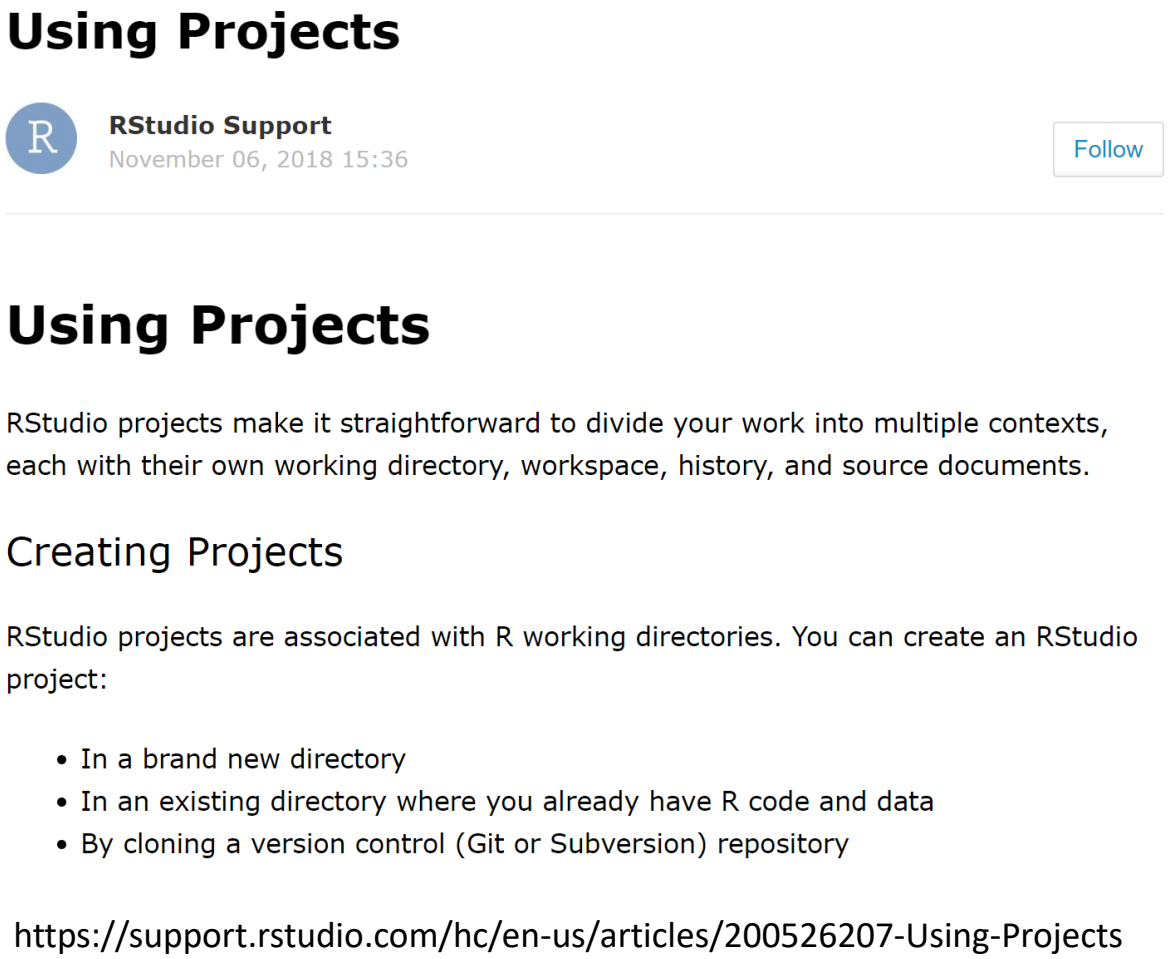
RStudio projects are associated with R working directories. You can create an RStudio project:

- In a brand new directory
- In an existing directory where you already have R code and data
- By cloning a version control (Git or Subversion) repository

https://support.rstudio.com/hc/en-us/articles/200526207-Using-Projects

Accessibility: Tools and Strategies

3. R Projects
   - Pick up where you left off
   - Avoid `setwd()` in every script
   - Easier for collaboration
   - Start at beginning of project, but if you haven't, it's not too late!



Using Projects

RStudio Support
November 06, 2018 15:36

Follow

Using Projects

RStudio projects make it straightforward to divide your work into multiple contexts, each with their own working directory, workspace, history, and source documents.

Creating Projects

RStudio projects are associated with R working directories. You can create an RStudio project:

- In a brand new directory
- In an existing directory where you already have R code and data
- By cloning a version control (Git or Subversion) repository

https://support.rstudio.com/hc/en-us/articles/200526207-Using-Projects

Accessibility: Tools and Strategies

4. GitHub – log steps as you move forward

Accessibility: Tools and Strategies

4. GitHub – log steps as you move forward
   - Yes, it’s worth it

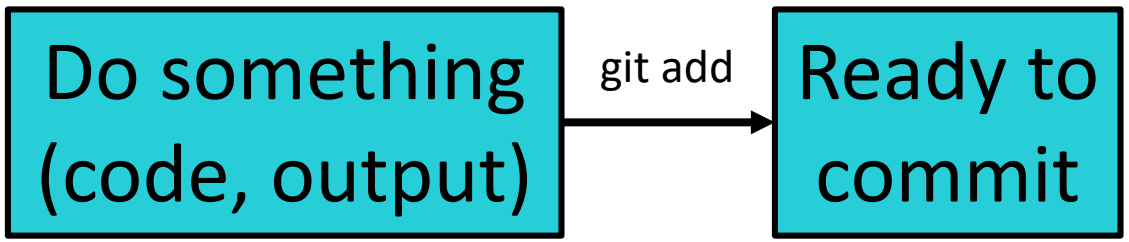
Accessibility: Tools and Strategies

4. GitHub – log steps as you move forward
   - Yes, it’s worth it
   - No, people won’t judge your messy code

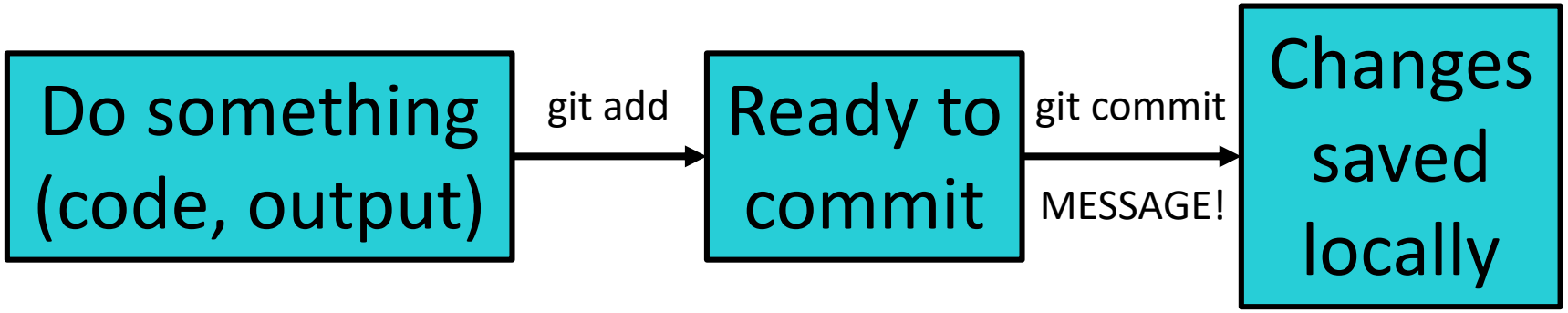
Accessibility: Tools and Strategies

4. GitHub – log steps as you move forward
   - Yes, it’s worth it
   - No, people won’t judge your messy code
   - Good, because it forces you to track progress and steps

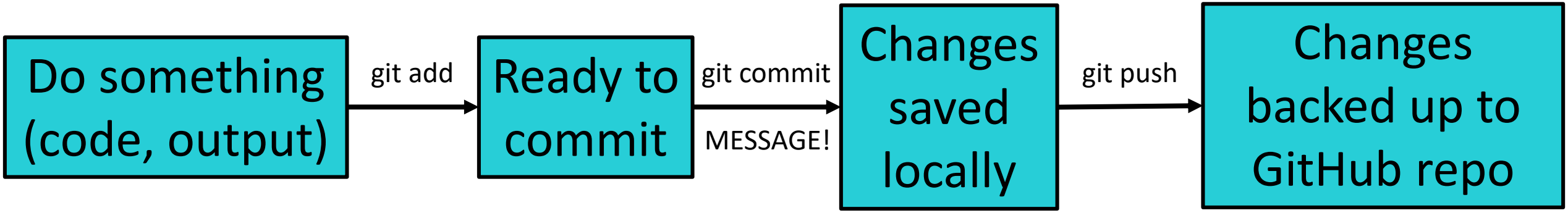
GitHub - Overview



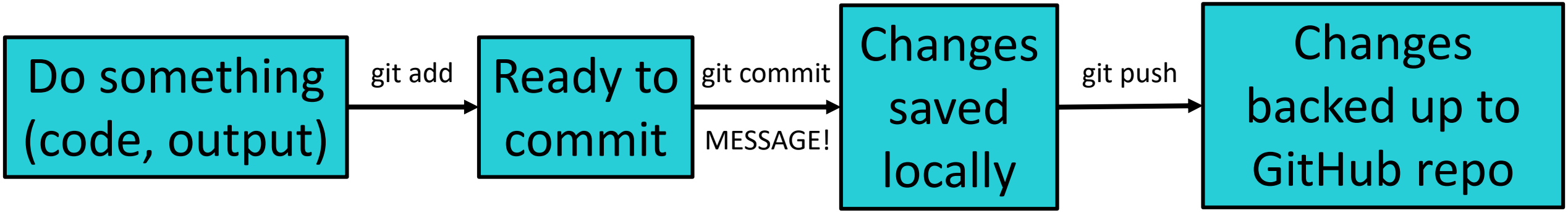
GitHub - Overview



GitHub - Overview



GitHub - Overview



Benefits:

- Track what you've done
- Online backup in case of computer failure
- Preps code to share with colleagues

GitHub

- My strategy – separate script for each task
  - If complicated, build a map!

GitHub

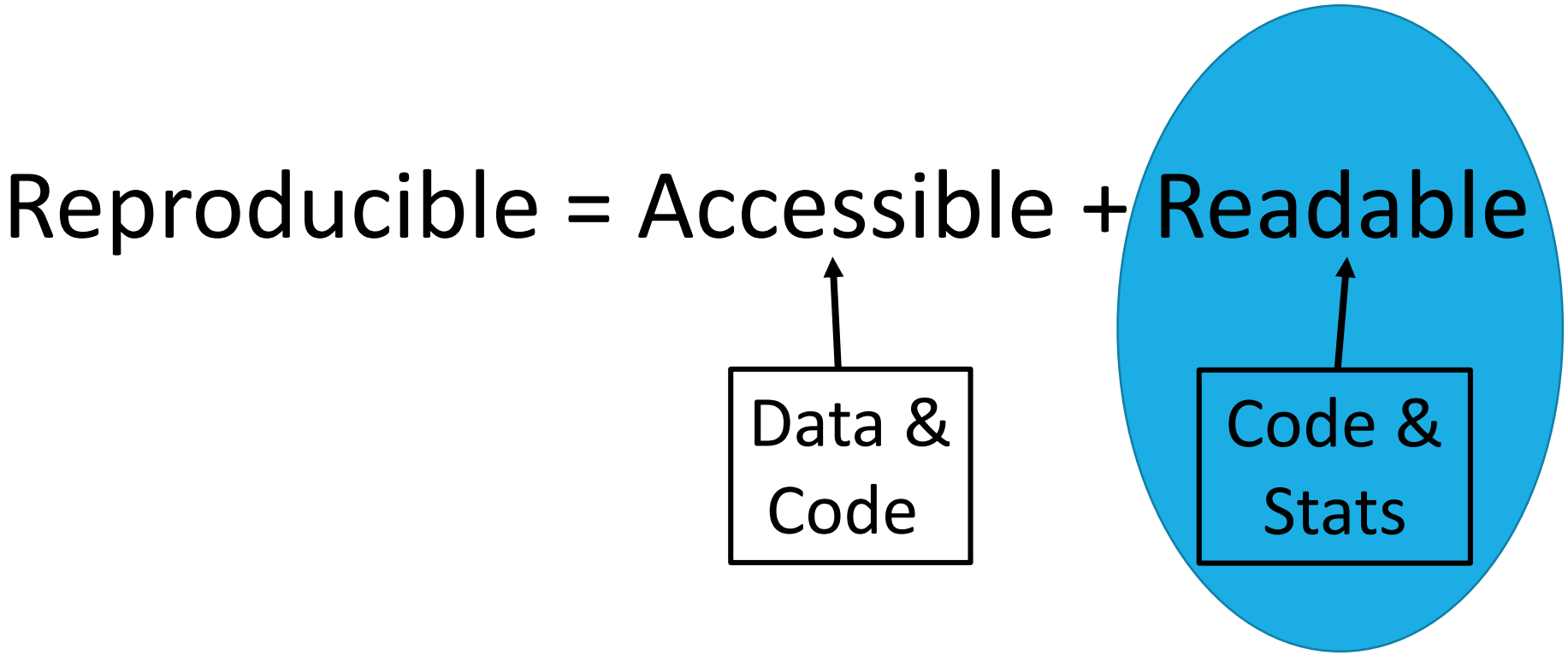
- My strategy – separate script for each task
  - If complicated, build a map!
- Commit all the time

GitHub

- My strategy – separate script for each task
  - If complicated, build a map!
- Commit all the time
- Make commit messages specific

GitHub

- My strategy – separate script for each task
  - If complicated, build a map!
- Commit all the time
- Make commit messages specific
- What about when code gets messy?


Reproducible = Accessible + Readable
Data & Code
Code & Stats

Readability: Tools and Strategies

- Commenting throughout code

Readability: Tools and Strategies

- Commenting throughout code
- Divide code into chunks or sections ⟶ R Markdown!

R Markdown

- Great for preliminary analyses & figures, as well as statistical analyses

R Markdown

- Great for preliminary analyses & figures, as well as statistical analyses
- Keep track of early data exploration

R Markdown

- Great for preliminary analyses & figures, as well as statistical analyses
- Keep track of early data exploration
- Helpful for thinking, exploring, returning to data

R Markdown

- Great for preliminary analyses & figures, as well as statistical analyses
- Keep track of early data exploration
- Helpful for thinking, exploring, returning to data
- Easiest way to share your progress with Julia (+ coauthors)

R Markdown

- Great for preliminary analyses & figures, as well as statistical analyses
- Keep track of early data exploration
- Helpful for thinking, exploring, returning to data
- Easiest way to share your progress with Julia (+ coauthors)

Now for a few examples…

A few resources…

- R Markdown Basics https://rmarkdown.rstudio.com/authoring_basics.html
- Baum Lab – Gitting Started – R Markdown Tips https://github.com/baumlab/Gitting-started/blob/master/ms-rmarkdown-tips.md
- Reproducible Research using R Markdown https://www.r-bloggers.com/reproducible-research-write-your-clinical-chemistry-paper-using-r-markdown/
- Using R Projects https://support.rstudio.com/hc/en-us/articles/200526207-Using-Projects
- Pimp my RMD: a few tips for R Markdown https://holtzy.github.io/Pimp-my-rmd/
- Example Repo with RMarkdown data analysis file https://github.com/hputnam/Moorea_Sym
- New Data Visualization Book http://serialmentor.com/dataviz/
- Organizing Multiple Plots on a Page https://cran.r-project.org/web/packages/egg/vignettes/Ecosystem.html?utm_content=buffercef33&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer