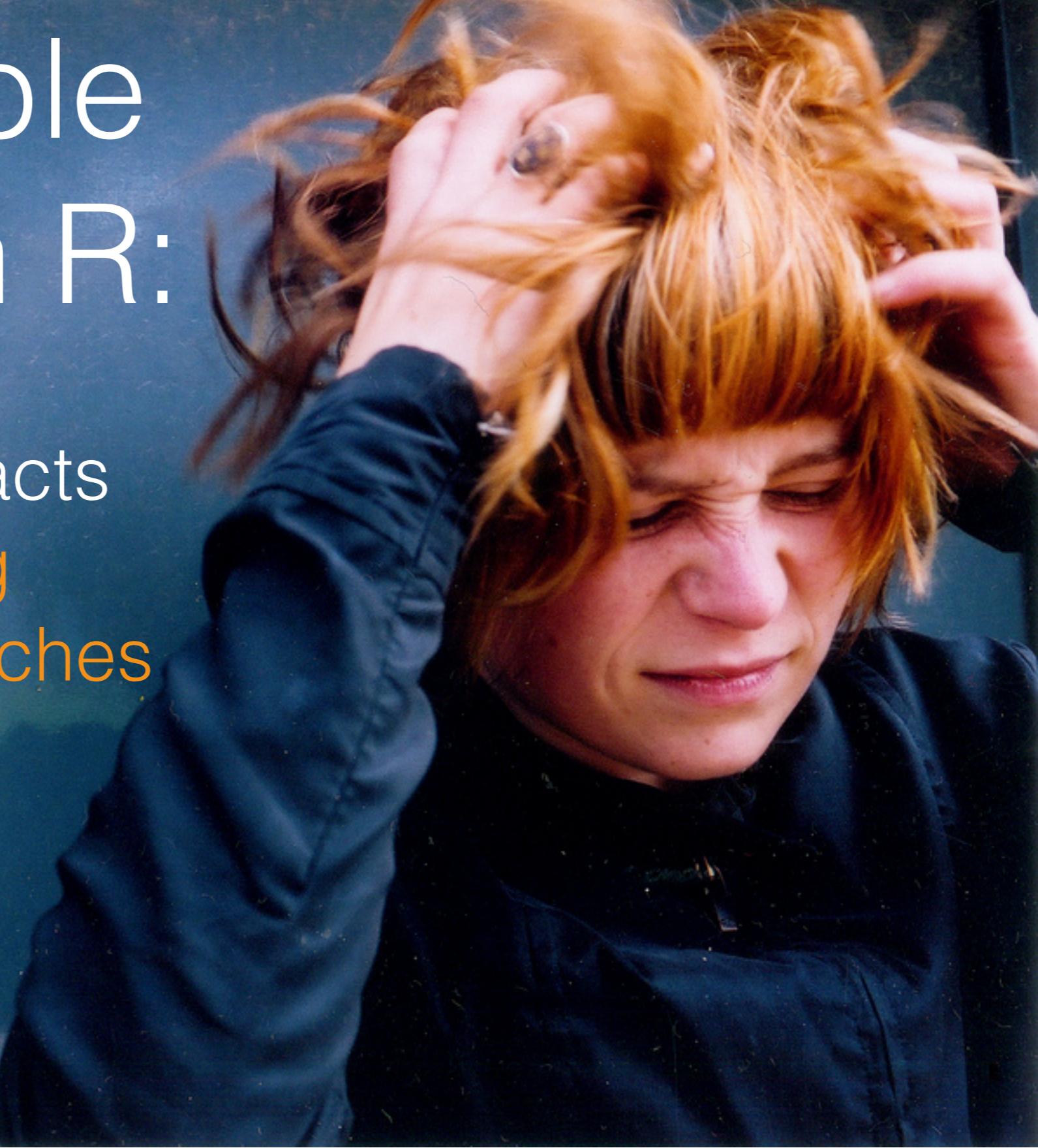


# Reproducible Research in R:

Increasing  
your research impacts  
while decreasing  
your research headaches



**Althea A. ArchMiller**

Postdoctoral Researcher  
Dept of Fisheries, Wildlife, & Cons Bio



@aaarchmiller

arche049@umn.edu

# Tip 1:

## Use **RStudio** to create a reproducible workflow



**Vince Buffalo**  
@vsbuffalo



 Follow

Managing your projects in a reproducible fashion doesn't just make your science reproducible, it makes your life easier.

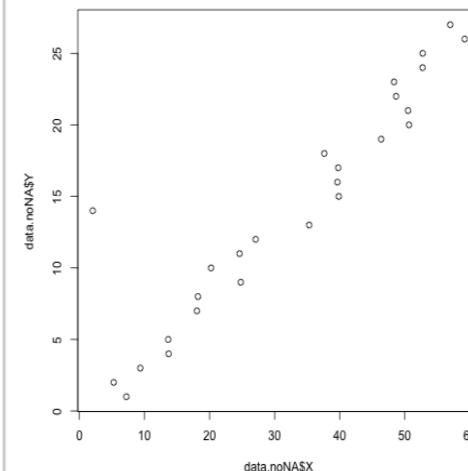
# R Code Window

```
RStudio File Edit Object Type  
3 #' File to make cleaned  
4 #' data file from raw data  
5 #' file.  
6 #'  
7 #' **Read CSV**  
8 rawdata <-  
9     read.csv("data/raw.csv")  
10 head(rawdata,n = 3)  
11  
  
> #' **Read CSV**  
> rawdata <-  
read.csv("data/raw.csv")  
> head(rawdata,n = 3)  
      X Y  
1 7.214992 1  
2 5.266567 2  
3 9.359408 3  
>
```

# Environment

**Data**  
▶ rawdata  
**Values**  
seqdata

Files | Plots



# R Results Console

# Files, Plots, etc

# Benefits of RStudio

## “Portable” Projects

```
*myProject.Rproj
  └── data/
      raw.csv
      cleaned.R
  └── programs/
      a_processing_data.R
      b_analysis.R
      c_plotting_results.R
  └── output/
      a_processing_data.html
      b_analysis.html
      c_plotting_results.htm
  └── figures/
      outliers.png
```

## Relative Pathways

```
3 #' File to make cleaned
4 #' data file from raw data
5 #' file.
6 #
7 #' **Read CSV**
8 rawdata <- *read.csv("data/raw.csv")
9 head(rawdata,n = 3)
10
11 #' **Clean up**
12 #
13 #
14 #' 1. Remove NAs
15 #' 2. Remove outliers
16 # Remove missing values in X
17 column
18 data.noNA <- rawdata[!
19             !is.na(rawdata$X),]
20 #' Plot data to check for
21 outliers
22 #+ outlierCheck
23 plot(x = data.noNA$X, y =
24       data.noNA$Y)
```

No more broken links when collaborating!

~~rawdata <- read.csv("users/Althea\_ArchMiller/Documents/RProjects/Reproducibility/data/rawdata.csv")~~

# Tip 2:

## Bundle R code, results, text, & figures with **knitr**



**Matt Upson**  
@m\_a\_upson

Submitted my first fully reproducible paper for review using **#Rstats**, **#LaTeX**, and **#knitr**. Surprisingly painless with editorial manager.

```

3 #' File to make cleaned
4 #' data file from raw data
5 #' file.
6 #
7 #' **Read CSV**
8 rawdata <-
9     read.csv("data/raw.csv")
10 head(rawdata,n = 3)
11
12 #' **Clean up**
13 #
14 #' 1. Remove NAs
15 #' 2. Remove outliers
16 # Remove missing values in X
17 column
18 data.noNA <- rawdata[!
19         is.na(rawdata$X),]
20 # Plot data to check for
21 outliers
22 #+ outlierCheck
23 plot(x = data.noNA$X, y =
24         data.noNA$Y)
25 #' Remove outliers

```

# Example knitr() Output

```

rawdata <- read.csv("data/raw.csv")
head(rawdata,n = 3)

##          X  Y
## 1 7.214992 1
## 2 5.266567 2
## 3 9.359408 3

```

## Clean up

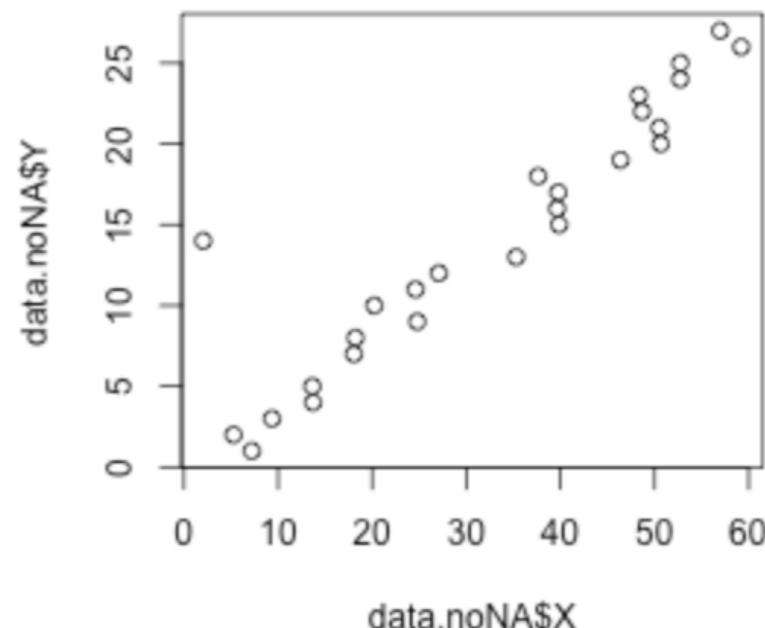
1. Remove NAs
2. Remove outliers

*# Remove missing values in X column*

```
data.noNA <- rawdata[! is.na(rawdata$X),]
```

Plot data to check for outliers

```
plot(x = data.noNA$X, y = data.noNA$Y)
```



# Tip 3: Distribute (and backup) your work with GitHub



 **Erin M. Kingsley**  
@erinmkingsley

#MLA16 #S736 Project will be published open  
access and created by open peer review and  
openly editing on GitHub.

# How GitHub Works

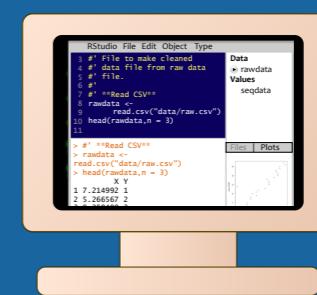
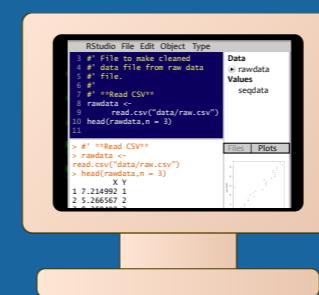
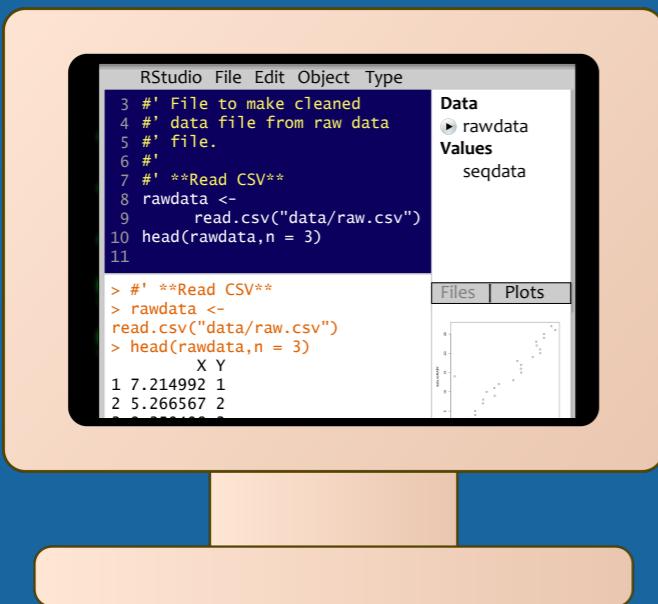


Main Version on  
GitHub Website



RStudio Project on  
*Your Computer*

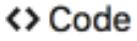
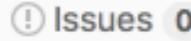
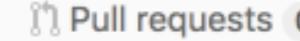
Public or Private  
Collaborators



# Example GitHub Repository

This repository Search Pull requests Issues Gist + 

 [aaarchmiller / reproducibility\\_with\\_R](#)   

 [Code](#)  [Issues 0](#)  [Pull requests 0](#)  [Projects 0](#)  [Wiki](#)  [Pulse](#)  [Graphs](#)  [Settings](#)

Presentation slides and example RStudio Project to demonstrate tips for making your research reproducible with R — [Edit](#)

 [11 commits](#)  [1 branch](#)  [0 releases](#)  [1 contributor](#)  [MIT](#)

Branch: [master](#) [New pull request](#) [Create new file](#) [Upload files](#) [Find file](#) [Clone or download](#)

 [aaarchmiller presentation files, updated with second artboard](#) Latest commit [e9c4e4a](#) 2 hours ago

 <a href="#">data</a>	processing data example	3 days ago
 <a href="#">output</a>	documentation code (eg knitr)	2 hours ago
 <a href="#">programs</a>	documentation code (eg knitr)	2 hours ago
 <a href="#">slides</a>	presentation files, updated with second artboard	2 hours ago
 <a href="#">.gitignore</a>	create .RProj file and raw data file	3 days ago
 <a href="#">LICENSE</a>	Initial commit	3 days ago
 <a href="#">README.md</a>	Initial commit	3 days ago
 <a href="#">reproducibility_with_R.Rproj</a>	create .RProj file and raw data file	3 days ago

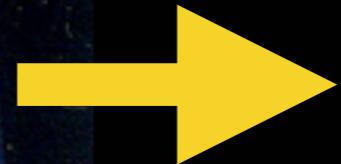
 [README.md](#)



**RStudio, knitr, &  
GitHub** can help  
you go from this



RStudio, knitr, &  
**GitHub** can help  
you go from this



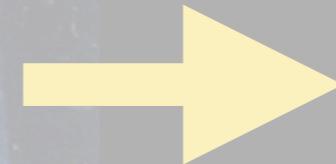
To this

Carmela Nava



Slides and data examples available at:  
[www.github.com/aaarchmiller/reproducibility\\_with\\_R](https://www.github.com/aaarchmiller/reproducibility_with_R)

**RStudio, knitr, &  
GitHub** can help  
you go from this



To this