

# R for Literary Analysis

Tassie Gniady

Cyberinfrastructure for Digital Humanities Manager  
UITS Research Technology, Indiana University  
Pervasive Technology Institute

October 13, 2017

Gniady, Tassie. "R for Literary Analysis." 13 October 2017. Digital Tools & Visualization Methods for Humanists Workshop Series. Scholars' Commons, Wells Library, Indiana University, Bloomington. Retrieved from: <http://hdl.handle.net/2022/21745>



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY



# Digital Arts and Humanities Workshop Series – Fall 2017

Fridays @ noon -- Scholars Commons IQ-Wall

Date	Topic	Presenter
Aug. 25	Intro to Visualization	Michael Boyles
Sep. 1	Intro to Digital Humanities	Tassie Gniady
Sep. 8	Virtual Reality	Bill Sherman
Sep. 15	Intro to R	Tassie Gniady
Sep. 22	Advanced Media	Chris Eller
Sep. 29	Augmented Reality	Chauncey Frend
Oct. 13	R for Text	Tassie Gniady
Oct. 20	Network Graphs	David Kloster
Oct. 27	R for Twitter	Tassie Gniady
Nov. 3	3D Scanning & Printing	Jeff Rogers
Nov. 10	3D Photogrammetry	Tassie Gniady
Oct. 27	IQ-Tables & Touch-Enabled Software Workflows	David Reagan



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY



# R for Literary Analysis

Follow along in today's presentation for helpful links:

<https://iu.box.com/v/rforlit>



**RESEARCH  
TECHNOLOGIES**

---

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

---

INDIANA UNIVERSITY



# Karst Account

If you don't already have a Karst account, go ahead and request one right now.

<https://kb.iu.edu/d/bezu#account>



**RESEARCH  
TECHNOLOGIES**

---

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

---

INDIANA UNIVERSITY



# Methodology

- Interactive Shiny app to get a sense of how an algorithm works
- RNotebook in which every line of code is explained
- RScript, which is lightly annotated and where you can alter some variables (suggestions given) or enter your own corpus



**RESEARCH  
TECHNOLOGIES**

---

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

---

INDIANA UNIVERSITY



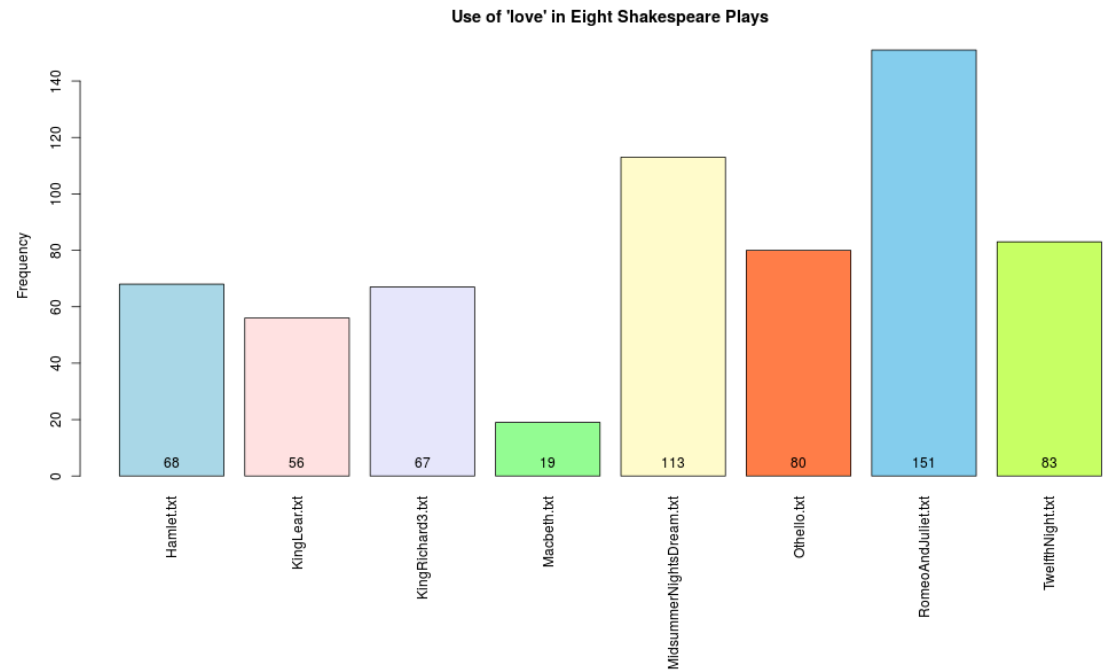
Shiny App: <https://cyberdh.shinyapps.io/TermFreq/>

## Frequency of Terms Across a Corpus

This interactive plot shows how frequent certain words appear in each text throughout a section of the Shakespeare Drama Corpus.

Choose a word:

love



RESEARCH  
TECHNOLOGIES

INDIANA UNIVERSITY  
University Information Technology Services



PERVASIVE TECHNOLOGY  
INSTITUTE

INDIANA UNIVERSITY



# RStudio

The screenshot displays the RStudio environment. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Tools, and Help. The main editor window shows a script named 'wordcloudPlainText.R' with the following code:

```
22 #to change the stopwords list, use other dictionaries available with the tm package
23 #Add early modern stopwords by u adding "myStopWords"
24 myStopWords <- scan("data/earlyModernStopword.txt", what="character", sep="\n")
25 corpus <- tm_map(corpus, removeWords, c(stopwords("english"), myStopWords))
26 corpus <- tm_map(corpus, removePunctuation)
27 corpus <- tm_map(corpus, stripWhitespace)
28 corpus <- tm_map(corpus, PlainTextDocument)
29
30
31 wordcloud(corpus, random.order=FALSE, scale=c(4,1), rot.per=0,
32           max.words=75, colors=brewer.pal(8, "Dark2"))
33
34
```

The Environment pane on the right shows the following values:

Variable	Value
corpus	Large VCorpus (4154 elements, 14.6 Mb)
myStopWords	chr [1:353] "a" "about" "above" "across" "after" "afterwards" "aga...
text_raw	chr [1:4154] " Who's there?" " Nay, answer me. Stand and unfold ya...

The Console pane shows the following output:

```
[Workspace loaded from ~/Text-Analysis/.RData]
> source("~/Text-Analysis/Rscripts/wordcloudPlainText.R")
Loading required package: RColorBrewer
Loading required package: NLP
Read 4154 items
Read 353 items
> source("~/Text-Analysis/Rscripts/wordcloudPlainText.R")
Read 4154 items
Read 353 items
>
```

The Plots pane on the right displays a word cloud with the following words:

little blood  
sword means comes hand  
art set great nature believe  
son heart hold laertes follow  
eyes death heaven life poor  
away father know hear words  
matter say let good look show  
sweet king lord time seen  
earth day think man queen  
long mother come love god old  
noble dear mad hamlets speak head  
leave make tell soul fair  
thoughts fear pray night dead true farewell  
england horatio world till  
denmark madness friends ophelia



RESEARCH  
TECHNOLOGIES

INDIANA UNIVERSITY  
University Information Technology Services



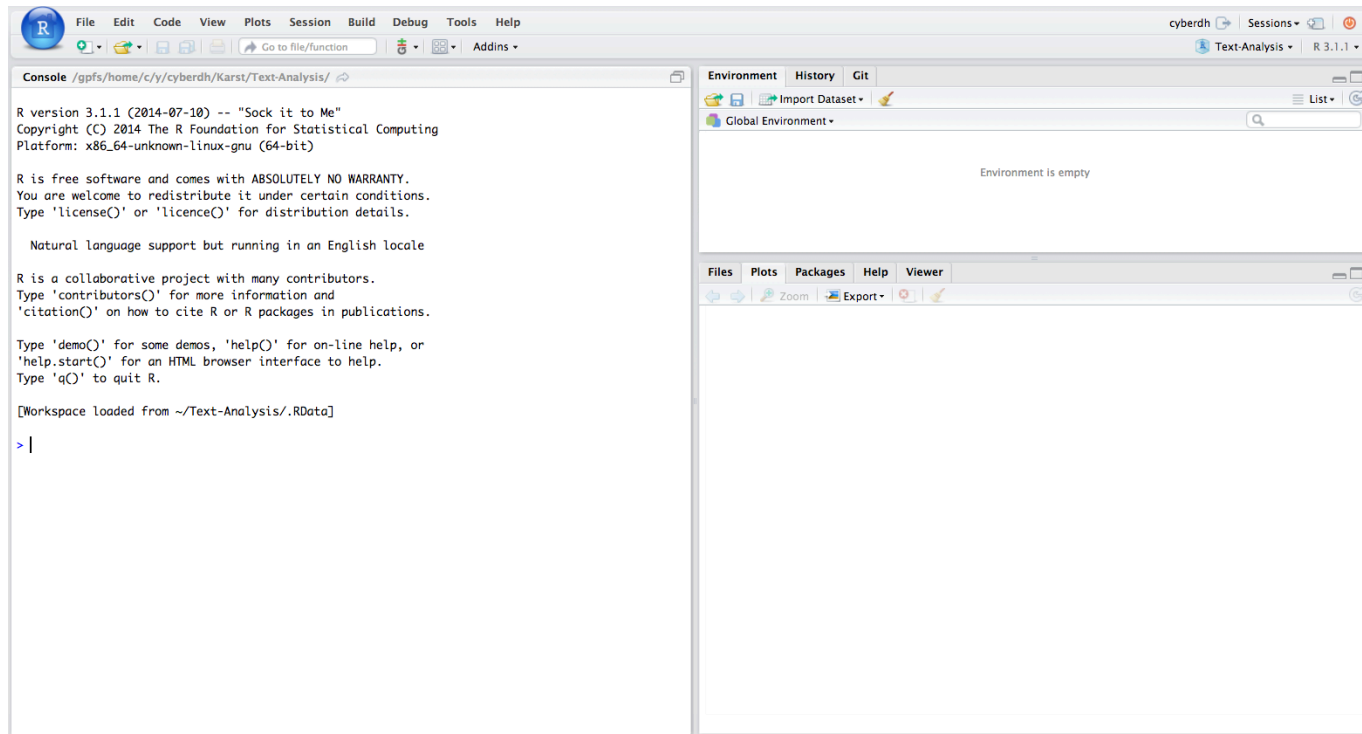
PERVASIVE TECHNOLOGY  
INSTITUTE

INDIANA UNIVERSITY



# RStudio on Karst

- Go to <https://rstudio.iu.edu>



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

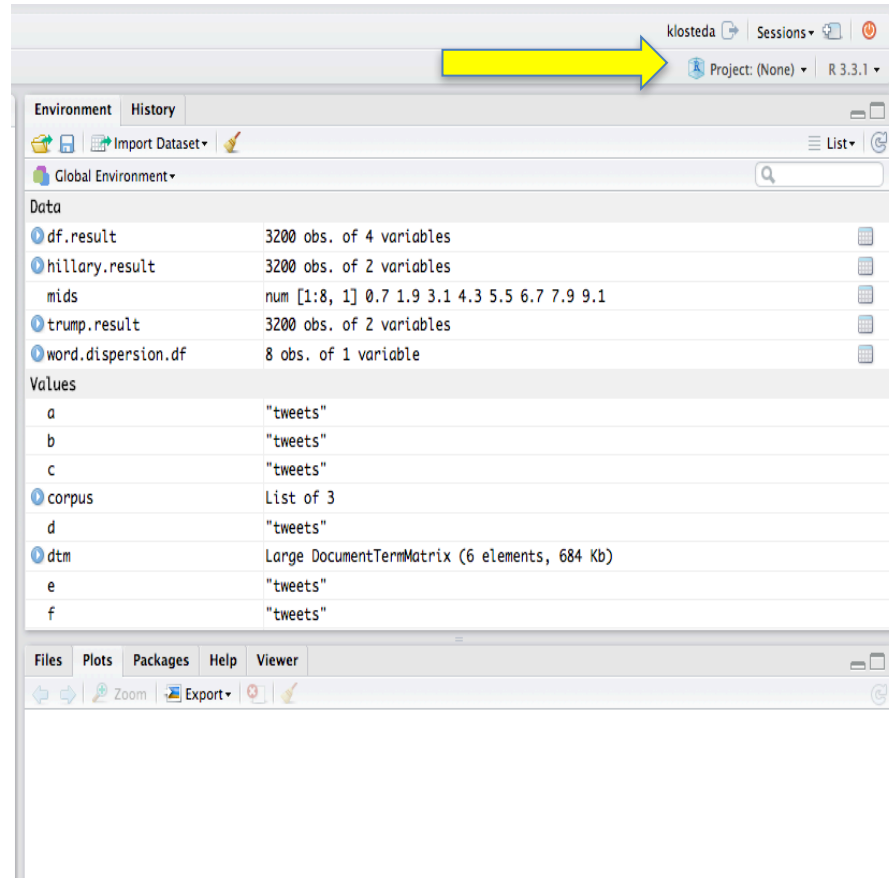
INDIANA UNIVERSITY





# Creating a Project in R Studio from GitHub

In R Studio, click on drop down arrow that says “Project: (None)”



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services

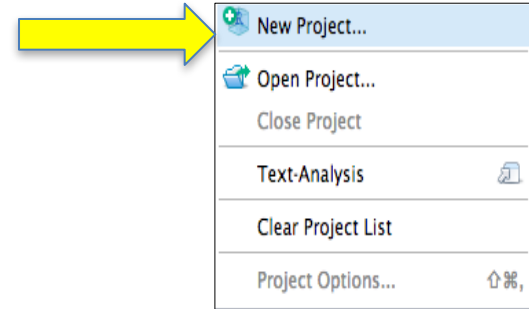


**PERVASIVE TECHNOLOGY  
INSTITUTE**

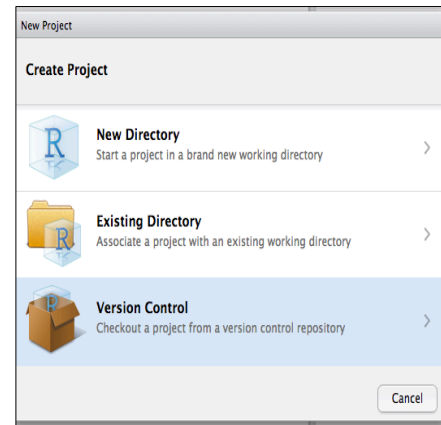
INDIANA UNIVERSITY



- Click on the first option "New Project..."



- Now choose "Version Control."



RESEARCH  
TECHNOLOGIES

INDIANA UNIVERSITY  
University Information Technology Services

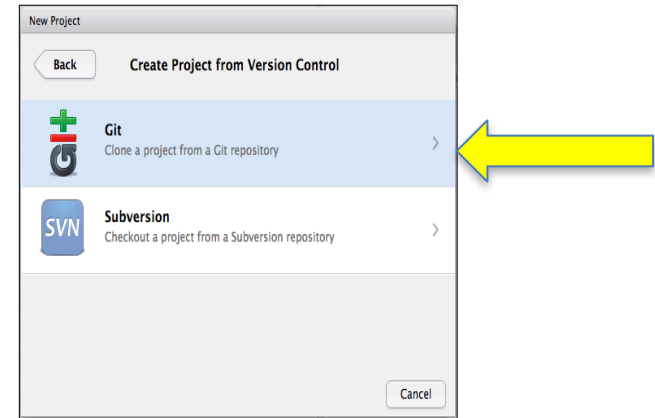


PERVASIVE TECHNOLOGY  
INSTITUTE

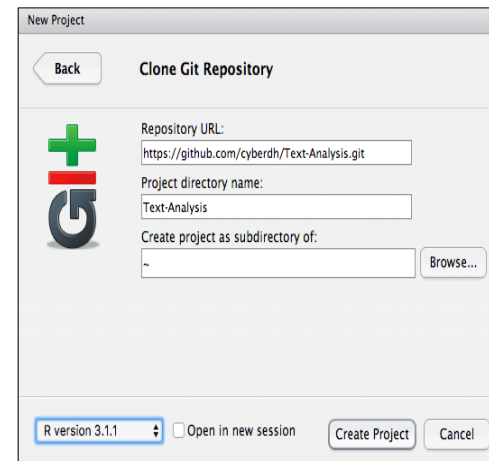
INDIANA UNIVERSITY



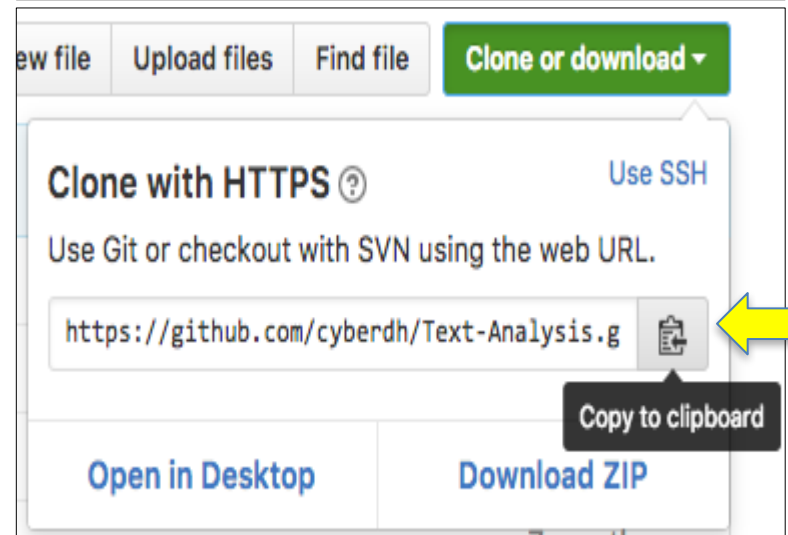
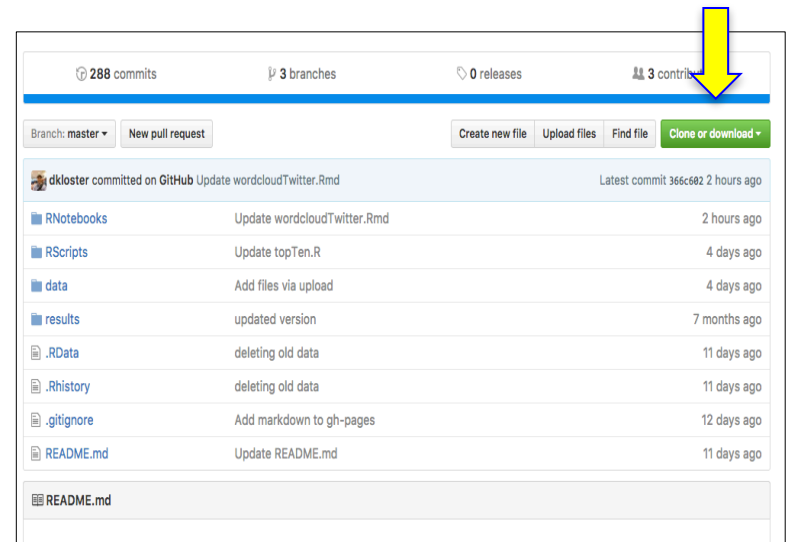
- Another box should appear that asks you to choose between “Git” or “Subversion (SVN).” Choose Git.



- This box should now appear asking for a “Repository URL” and directory name. For the URL we will go to the Cyber DH Text-Analysis GitHub repository at <https://github.com/cyberdh/Text-Analysis.git>



- The Cyber DH Text-Analysis repository on GitHub should look like this. Click on the green button labeled “Clone or download.”
- This dropdown menu should appear. Make sure it says “Clone with HTTPS” in the top left corner. Then click the button with the clipboard at which the arrow is pointing. This will save it to your clipboard. Now go back to the R Studio page.



RESEARCH  
TECHNOLOGIES

INDIANA UNIVERSITY  
University Information Technology Services

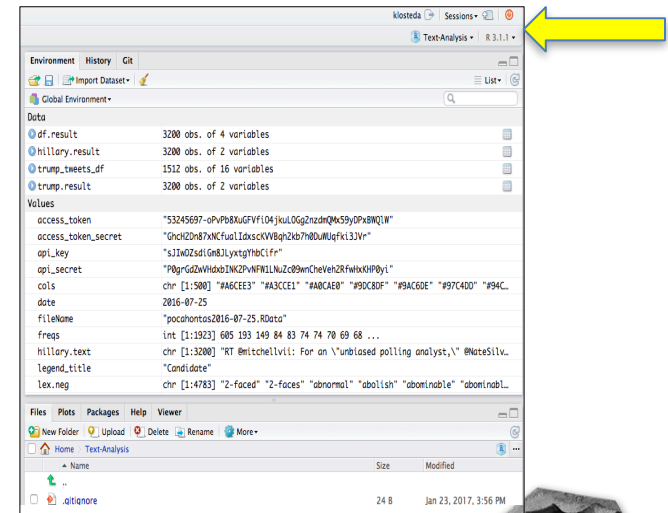
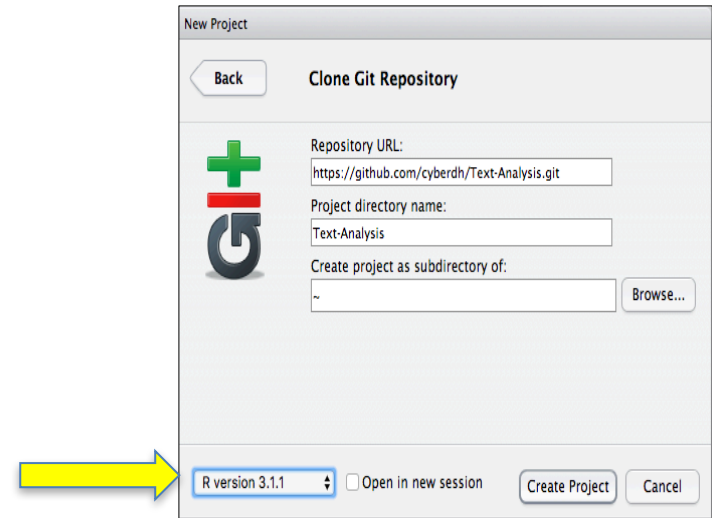


PERVASIVE TECHNOLOGY  
INSTITUTE

INDIANA UNIVERSITY



- Now paste the copied URL in the "Repository URL:" space. Next make sure the "Project directory name:" is **Text-Analysis** exactly including capitalization and dash. Otherwise you will need to change this in the script every time. Also, make sure you have R version 3.1.1 chosen in the lower left corner where the blue arrow is pointing. Then click the "Create Project" button.
- You should now see "Text-Analysis" listed in the project box and "R 3.1.1" in the box just to the right. You now have the Text-Analysis repository loaded on Karst.



RESEARCH  
TECHNOLOGIES

INDIANA UNIVERSITY  
University Information Technology Services



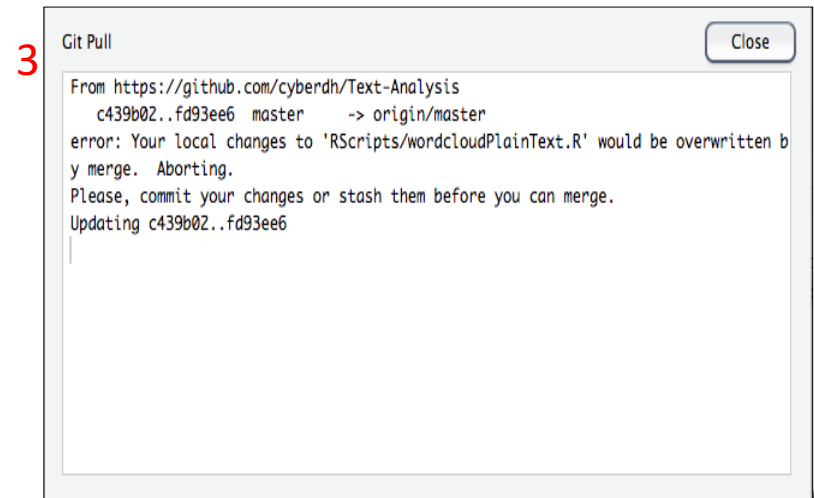
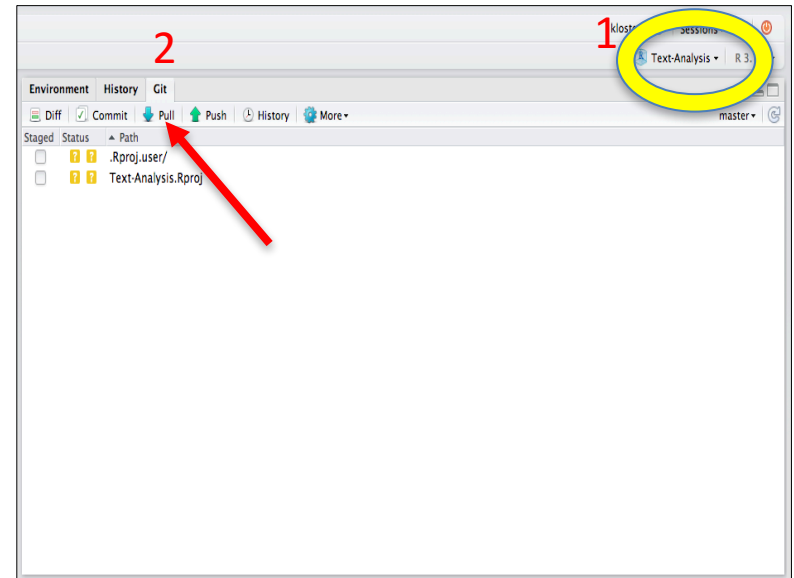
PERVASIVE TECHNOLOGY  
INSTITUTE

INDIANA UNIVERSITY



# Pulling updates from Cyber DH GitHub repository on R Studio (Karst or Desktop)

1. Make sure you are in the Text-Analysis project. It should say it in the upper right hand corner.
2. Next, make sure you are in the **Git** tab. Now, there should be a button with a blue arrow pointing down that says **Pull**. Click on that button.
3. As you make changes, you will have to “stash” them locally.



RESEARCH  
TECHNOLOGIES

INDIANA UNIVERSITY  
University Information Technology Services



PERVASIVE TECHNOLOGY  
INSTITUTE

INDIANA UNIVERSITY



# What Just Happened?

The screenshot shows the GitHub interface for the repository 'cyberdh / Text-Analysis'. At the top, there are navigation links for 'Code', 'Issues 0', 'Pull requests 0', 'Projects 0', 'Wiki', 'Pulse', 'Graphs', and 'Settings'. Below these, a description reads 'For all kinds of textual analysis: literary, social media, surveys...' with an 'Edit' button. A 'New Add topics' button is also present. The repository statistics show 337 commits, 2 branches, 0 releases, and 3 contributors. A 'Branch: master' dropdown and a 'New pull request' button are visible. Action buttons include 'Create new file', 'Upload files', 'Find file', and 'Clone or download'. The commit history shows a recent commit by 'tassieg' deleting 'termFrequency.R', followed by commits deleting files in 'RNotebooks' and 'RScripts', and updating 'data', '.RData', and '.Rhistory'.

Commit Message	Time Ago
Latest commit 80cc59e	2 days ago
Delete termFrequency.R	3 days ago
Delete cooccurrencePlainText.pdf	3 days ago
Delete termFrequency.R	2 days ago
deleting survey data	3 days ago
updating twitter data	22 days ago
updating twitter data	22 days ago
Add markdown to gh-pages	a month ago
Update README.md	a month ago

This repository includes the RNotebooks, RScripts, and the data needed to perform the analysis we'll do today.



RESEARCH  
TECHNOLOGIES

INDIANA UNIVERSITY  
University Information Technology Services



PERVASIVE TECHNOLOGY  
INSTITUTE

INDIANA UNIVERSITY



# Hands-On

- Now for wordclouds and top ten frequency graphs!



**RESEARCH  
TECHNOLOGIES**

---

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

---

INDIANA UNIVERSITY





# Star Trek: Next Generation



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY



# More Information

- Reference this presentation at: <https://iu.box.com/v/intror>
- If you'd like to make an appointment, please email [cyberdh@iu.edu](mailto:cyberdh@iu.edu).
- Also, check out our website: <https://www.indiana.edu/~cyberdh/> for interesting blog posts and written tutorials about R.

Contact:  
Tassie Gniady  
Cyberinfrastructure for Digital Humanities Manager  
[ctgniady@iu.edu](mailto:ctgniady@iu.edu)  
@tassietheg



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY

