

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	IQ-Tables & Touch-Enabled Software Workflows	David Reagan
Nov. 3	3D Scanning & Printing	Jeff Rogers
Nov. 10	3D Photogrammetry	Tassie Gniady
Dec. 1	R for Twitter	Tassie Gniady



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services



**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY



Intro to R for Humanists

Tassie Gniady

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

Gniady, Tassie. "Introduction to R for Humanists.". Digital Tools & Visualization Methods for Humanists Workshop Series. Scholars' Commons, Wells Library, Indiana University, Bloomington. Retrieved from: <http://hdl.handle.net/2022/21715>



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services



**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY



Follow along in today's presentation for helpful links:

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



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services



**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY



The Basics

- R is an open source language
- It uses <- instead of = to assign variables
- It begins counting vector positions at 1, not 0 like many other languages



**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://one.iu.edu/task/iu/account-creation>



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services



**PERVASIVE TECHNOLOGY
INSTITUTE**

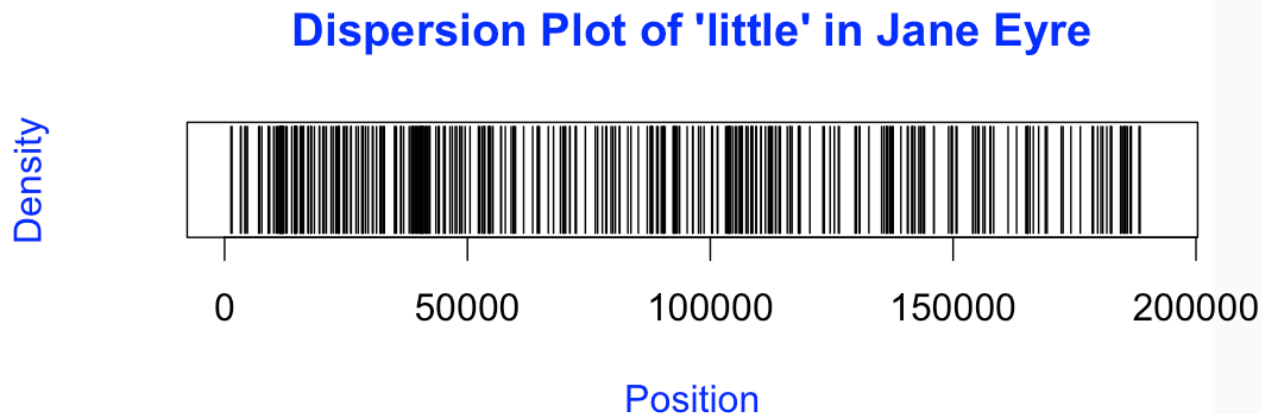
INDIANA UNIVERSITY



What is it good for?

Developed for statisticians, R is very powerful. While we won't get to these examples today, here are some of the outputs that you can generate:

- Dispersion plots



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services



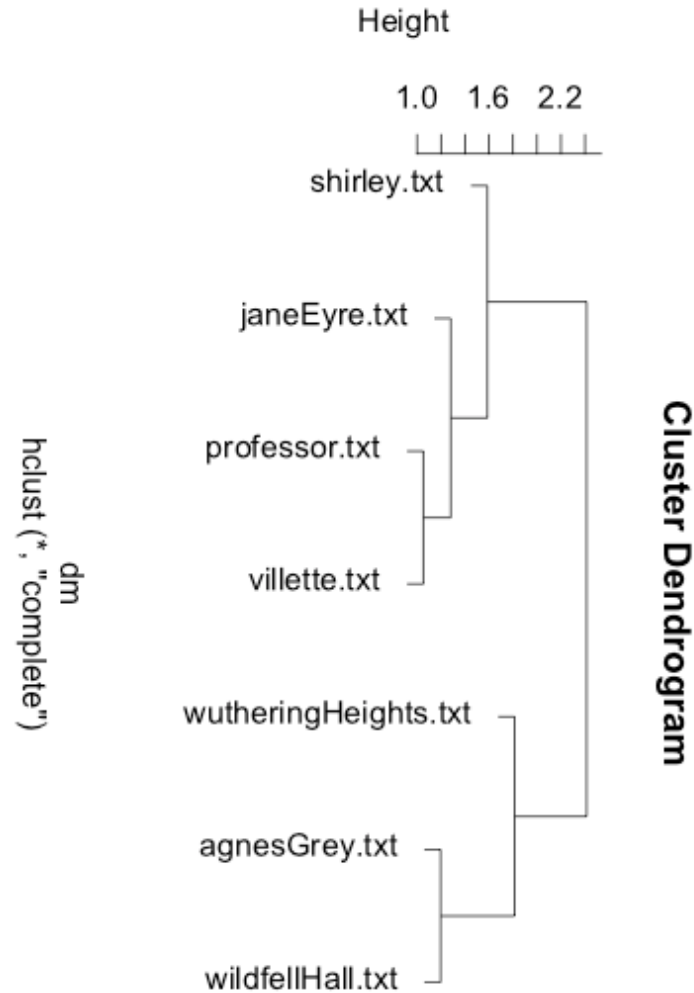
**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY



More Uses

Dendrograms:



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services

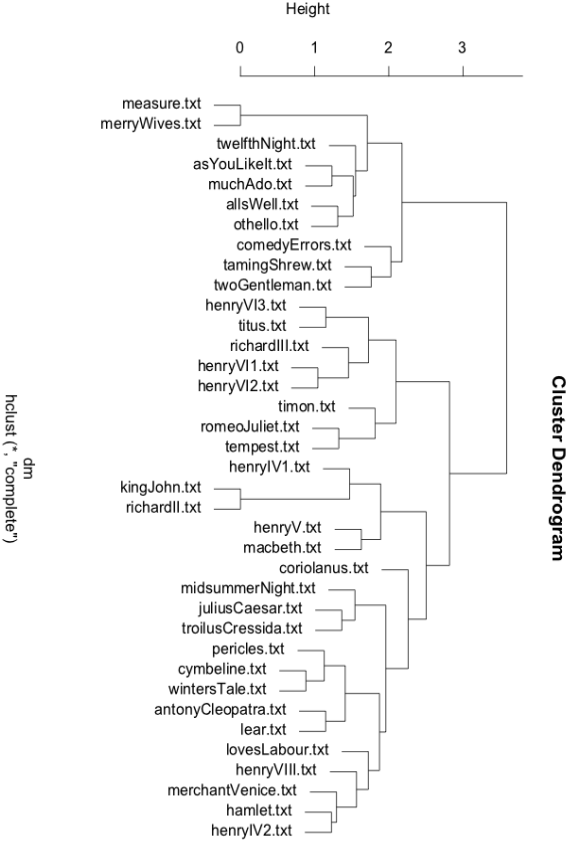


**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY



Shakespeare Dendrogram



R, Me Matey!

Let's get started!

<http://try.codeschool.com/levels/1/challenges/1>



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services



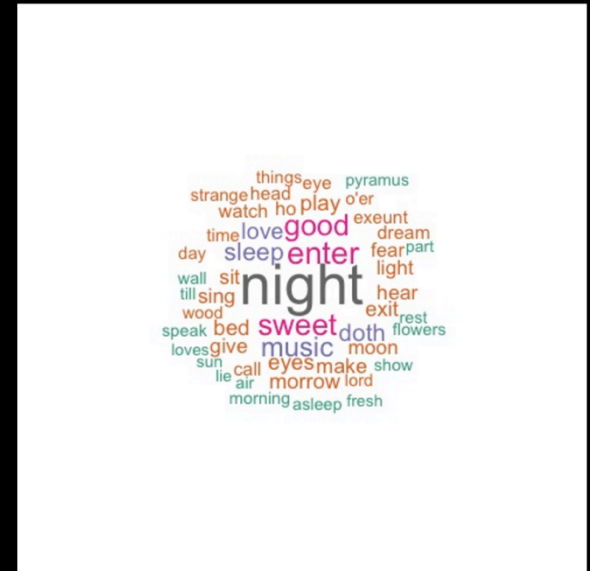
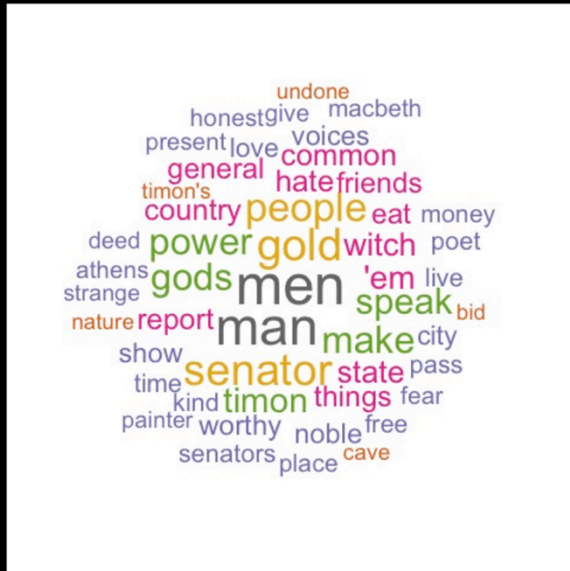
**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY



Why use wordclouds?

Topic Modeling (LDA)



RESEARCH
TECHNOLOGIES

INDIANA UNIVERSITY
University Information Technology Services



PERVASIVE TECHNOLOGY
INSTITUTE

INDIANA UNIVERSITY



Which witch?

henryIV.txt
henryV2.txt
henryV.txt
henryVIII.txt
henryVI1.txt
henryVI2.txt
henryVI3.txt
juliusCaesar.txt
kingJohn.txt
lear.txt
lovesLabour.txt
macbeth.txt
measure.txt
merchantVenice.txt
merryWives.txt
midsummerNight.txt
muchAdo.txt
othello.txt
pericles.txt
richardII.txt
richardIII.txt
romeoJuliet.txt
tamingShrew.txt
tempest.txt
timon.txt
titus.txt
troilusCressida.txt
twelfthNight.txt
twoGentleman.txt
wintersTale.txt

4	I amazed ran from her as a	witch: And, I think, if my breast had not	comedyErrors
5	JS OF SYRACUSE Avaunt, thou	witch! Come, Dromio, let us go. DROMIO OF	comedyErrors
6	to stay here still and turn	witch. ANTIPHOLUS OF SYRACUSE I will not s	comedyErrors
7	ruet manner'd; such a holy	witch That he enchants societies into him;	cymbeline.tx
8	strike, No fairy takes, nor	witch hath power to charm, So hallow'd and	hamlet.txt
9	strike, No fairy takes, nor	witch hath power to charm, So hallow'd ar	hamletCharac
10	nd wind a fiery Pegasus And	witch the world with noble horsemanship. F	henryIV1.txt
11	c Who, like a foul and ugly	witch, doth limp So tediously away. The po	henryV.txt
12	I draw on thee, thou art a	witch, And straightway give thy soul to hi	henryVI1.txt
13	here I am, nor what I do; A	witch, by fear, not force, like Hannibal,	henryVI1.txt
14	ny treachery. Pucelle, that	witch, that damned sorceress, Hath wrought	henryVI1.txt
15	rgery Jourdain, the cunning	witch, With Roger Bolingbroke, the conjure	henryVI2.txt
16	nor gives gold to bring the	witch: Gold cannot come amiss, were she a	henryVI2.txt
17	the place of execution: The	witch in Smithfield shall be burn'd to ash	henryVI2.txt
18	soul inconstancy, To sit and	witch me, as Ascanius did When he to maddi	henryVI2.txt
19	d in burning Troy! Am I not	witch'd like her? or thou not false like	henryVI2.txt
20	body in gay ornaments, And	witch sweet ladies with my words and looks	henryVI3.txt
21	n plight, And, aroint thee,	witch, aroint thee! KENT How fares your gr	lear.txt
22	. Enter three Witches First	Witch When shall we three meet again In th	macbeth.txt



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 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 console window shows the following output:

```

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[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 Environment pane shows the following variables:

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 Plots pane 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 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



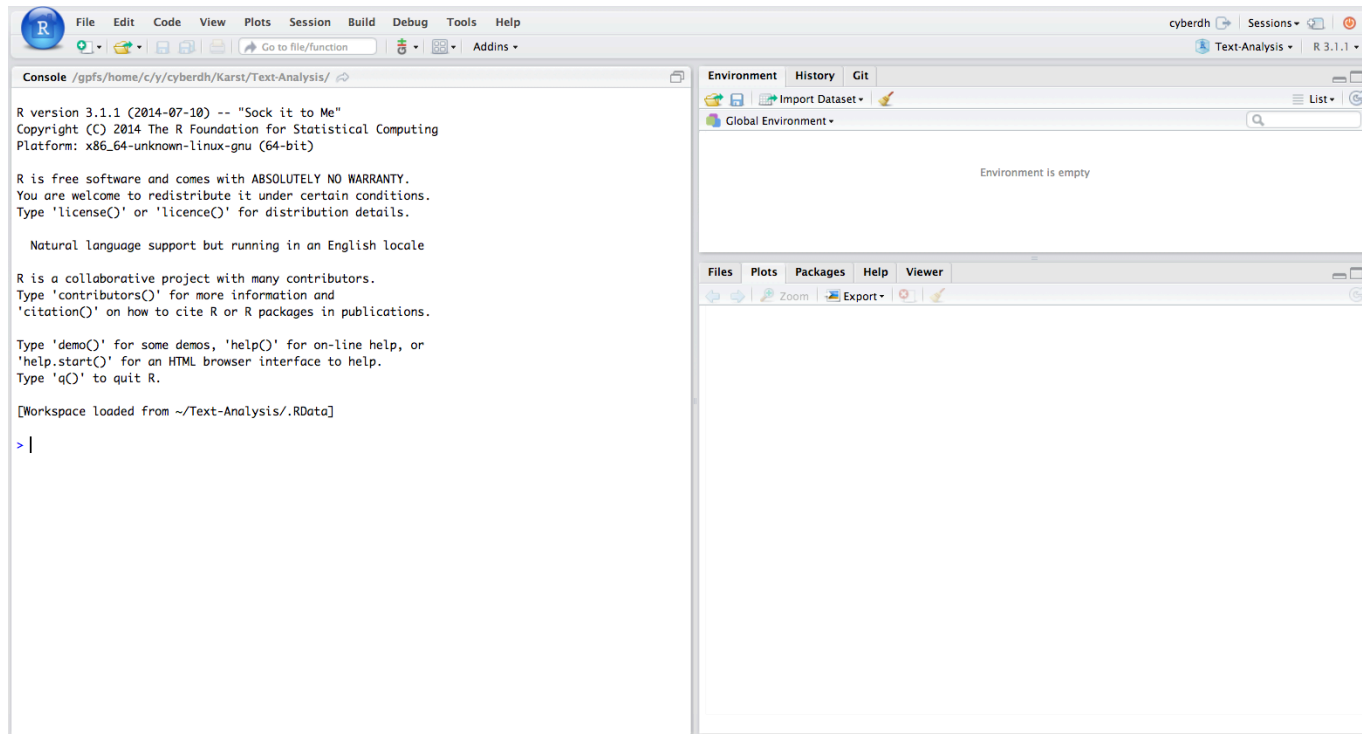
PERVASIVE TECHNOLOGY
INSTITUTE

INDIANA UNIVERSITY



RStudio on Karst

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



RESEARCH
TECHNOLOGIES

INDIANA UNIVERSITY
University Information Technology Services



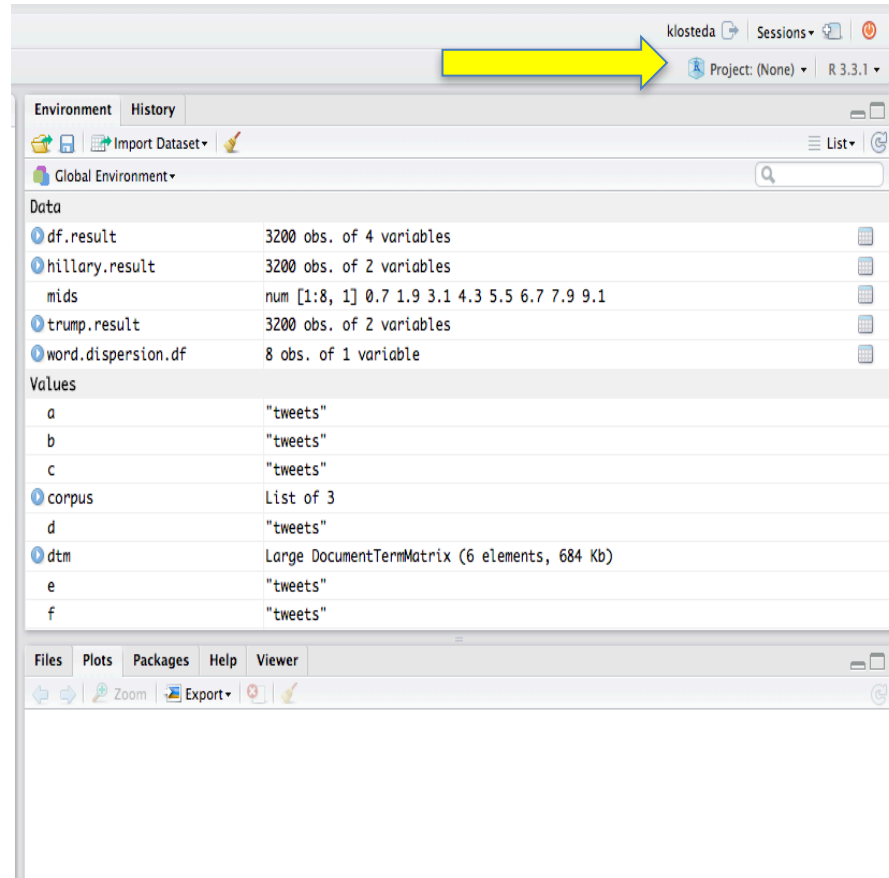
PERVASIVE TECHNOLOGY
INSTITUTE

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

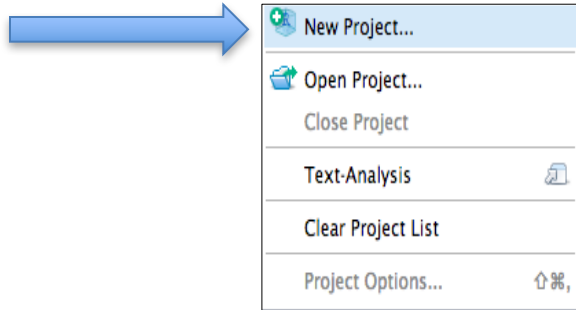


**PERVASIVE TECHNOLOGY
INSTITUTE**

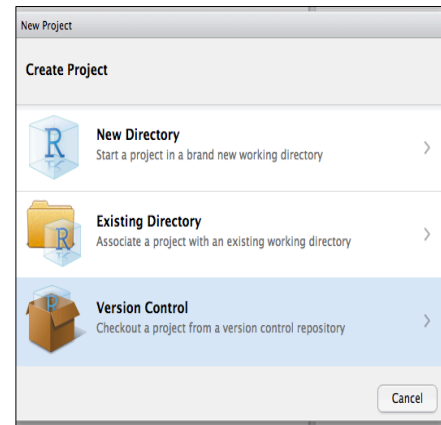
INDIANA UNIVERSITY



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



- A box should now appear asking you to choose between "New Directory," "Existing Directory," and "Version Control." Choose "Version Control."



RESEARCH
TECHNOLOGIES

INDIANA UNIVERSITY
University Information Technology Services

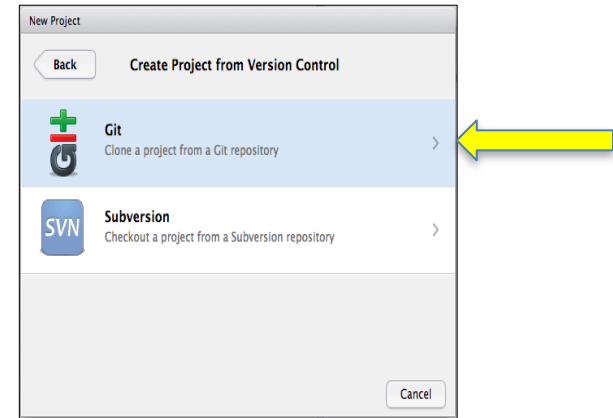


PERVASIVE TECHNOLOGY
INSTITUTE

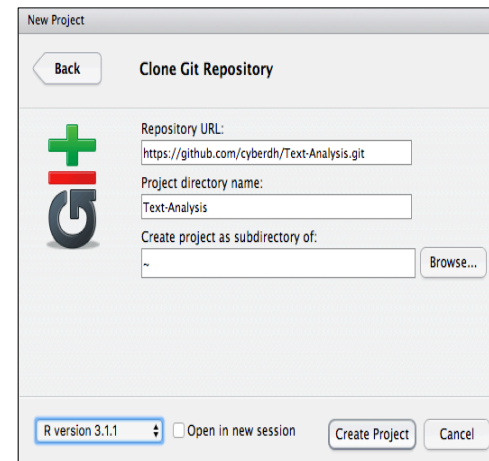
INDIANA UNIVERSITY



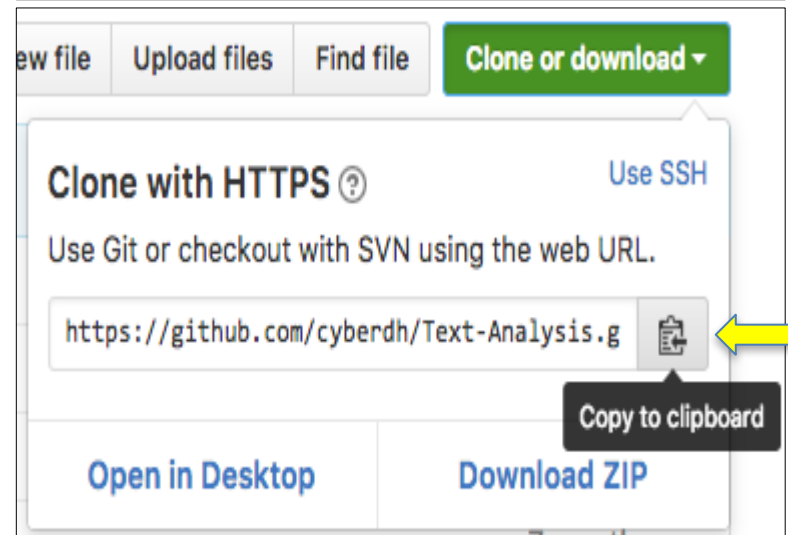
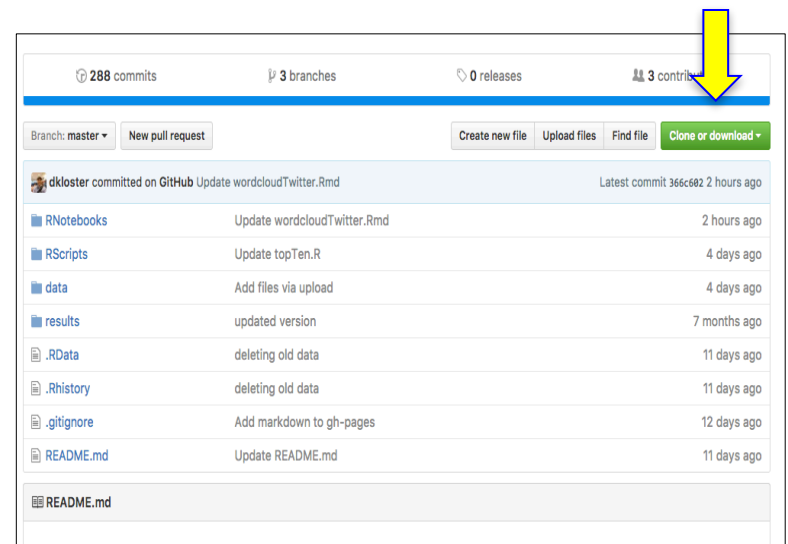
- Another box should appear that asks you to choose between “Git” or “Subversion (SVN).” Choose Git.



- 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**

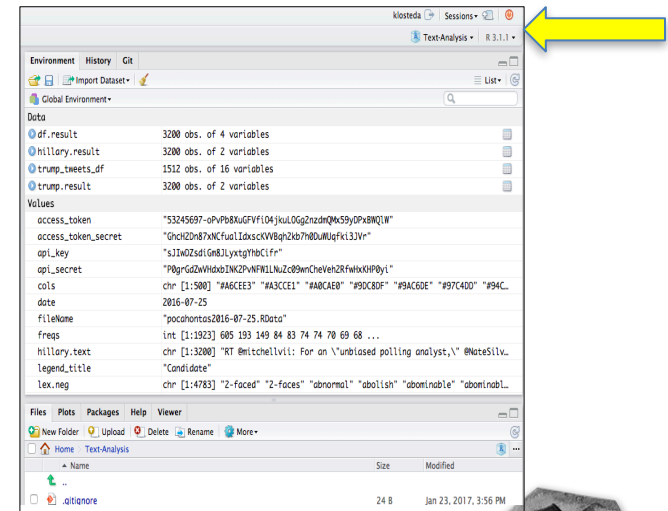
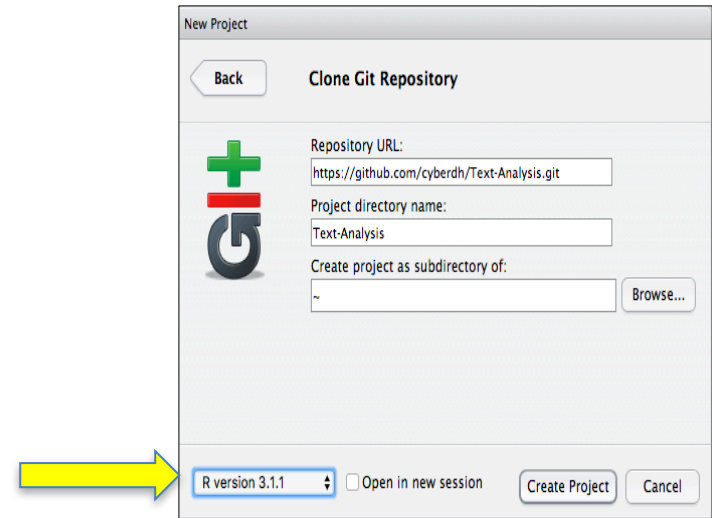
INDIANA UNIVERSITY
University Information Technology Services



**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



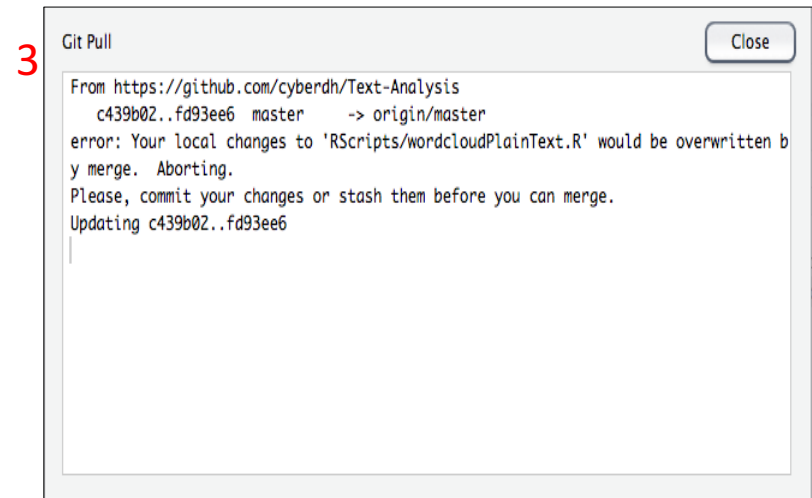
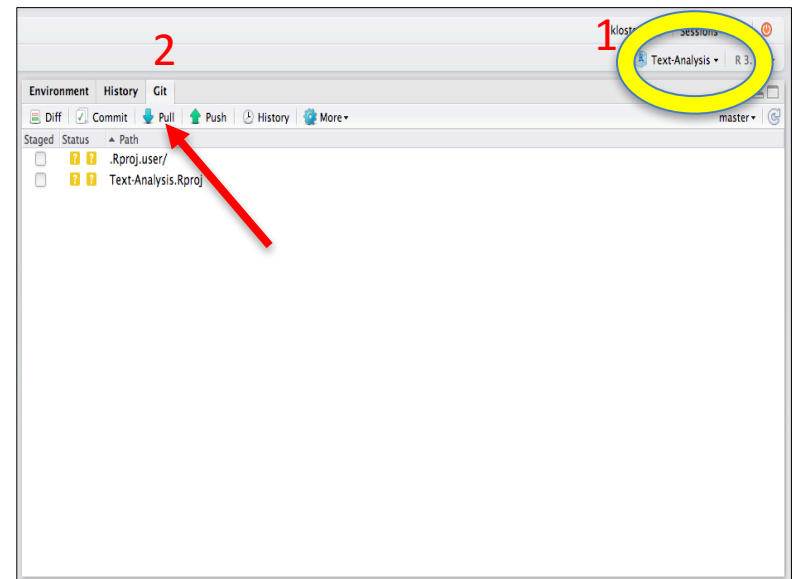
PERVASIVE TECHNOLOGY
INSTITUTE

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



TweetSets

- George Washington has made tweetsets available:
<https://tweetsets.library.gwu.edu/datasets>
- OSoME, a collection of tweets and algorithms collected here at IU:
<https://truthy.indiana.edu/tools/>
- TAGS to collect your own tweets from today forward:
<https://tags.hawksey.info/>



**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>
- Get to know the Text Analysis repo through our GitHub Page [Text Analysis Toolkit](#)
- If you'd like to make an appointment, please email 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
@tassietheg



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services



**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY

