

Welcome to a GCDI Workshop!

Introduction to Qualitative Data and Text Analysis in R

Yuxiao Luo

Support Fellow: Connor French



GC Digital Initiatives

THE
GRADUATE
CENTER
CITY UNIVERSITY
OF NEW YORK

What is GCDI?

GCDI offers different types of ***support for digital scholarship*** such as evening **workshops**, **one-on-one consultations** for faculty and students, **working groups** based around common tools or data sources, special **events** such as our annual Digital Showcase and Sound Series, and **online resources**.

Whether you have participated in our workshops before or your idea of the perfect software is a paperback edition, there is something for you! Our offerings are open to scholars at all levels of digital experience and at all stages of graduate research. Whether you are digitally driven, curious, or defiant, we are prepared to help.



GC Digital Initiatives

OK, so how can I get involved with GCDI?

- Join the **GCDI Group** on the CUNY Academic Commons for updates on how to be involved. **[cuny.is/group-gcdi](https://cuny.academiccommons.org/group-gcdi)**
- Check the Event Calendar for upcoming **events** and **workshops**. **cuny.is/workshops**
- Request a **one-on-one consultation**. Faculty: **cuny.is/gcdfconsults**
- Join the **R Users Group (RUG)** at **cuny.is/rug**
- Join the **GIS Working Group** at **cuny.is/group-gis-working-group**
- Join the **Data Visualization Group** at : **cuny.is/dvg**
- Join the **Python Users Group (PUG)** at **cuny.is/pug**

How else can I keep in touch with GCDI?

- Follow GCDI on Twitter: **@cunygcdi**
- Follow the GC Digital Fellows on Twitter: **@digital_fellows**
- Follow the **#digitalGC** hashtag on Twitter
- Follow the GC Digital Fellows blog, Tagging the Tower:
digitalfellows.commonsgc.cuny.edu
- Contact the fellows at **digitalfellows.commonsgc.cuny.edu/contact-us/**
- Drop-in to the Digital Scholarship Lab **Open House** hosted once per semester, or attend PUG, Events, or Workshops!



Introduction to Qualitative Data and Text Analysis in R

Oct 7th 2020



GC Digital Initiatives

Plan

- Introduction
- Installation
- Use RQDA for qualitative coding
- Navigate through GUI
- Textual Analysis

What is RQDA

- “RDQA” is a R package for Qualitative Data Analysis, who works on Windows and Mac OSX platforms.
- “RQDA” only supports plain text formatted data, i.e., “.txt”
- It seamlessly integrates with R
- It includes standard computer-aided qualitative data analysis features
 - statistical analysis on the coding is possible
 - data manipulation and analysis can be easily extended by writing R functions (ex., doing quantitative textual analysis)

What are the features of RQDA

Through the GUI

Qualitative Coding

- Character-level coding using codes and file Editing after coding
- Memos of documents, codes, coding, project, files and more
- Retrieval of coding as well as Conditional retrieval

Organization

- Organize codes into code categories (for theory building)
- Organize files into file categories

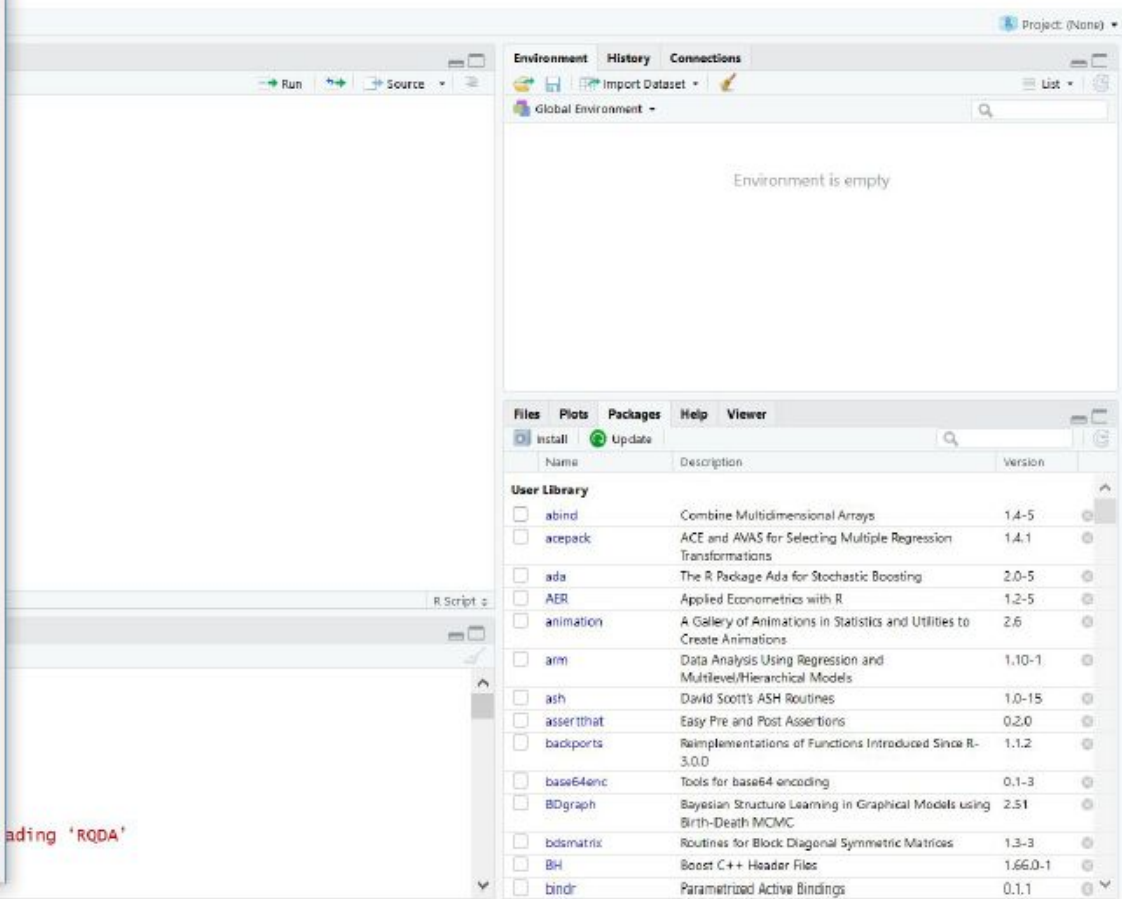
Search

- Search files by keywords with highlighting of the keyword in the open file
- Search information about a selected case on the Internet via a pop-up menu

What are the features of RQDA

Through R functions

- Manipulate the temporary deletions (list, undo temporary deletion, or permanently erase them)
- Calculate the relation between two codings, given the coding indexes
- Give a summary of coding and inter-code relationships
-
- Use the command to see more functions: `help(package="RQDA")`



Installing R and RStudio

- Instructions for installing R and RStudio:
<https://rstudio-education.github.io/hopr/starting.html>

Another instruction (same thing, option 2):

<https://github.com/GC-DRI/install/blob/master/sections/R.md>

Please install R 3.6.3 from the historical version (2020-02-29)

- RQDA is not available in R version > 4.0.0 (Most recent is R 4.0.2, Sep, 2020)

The image shows the RStudio desktop environment with several panels and annotations:

- Source Editor (Top Left):** Labeled "WRITE CODE HERE" with a red arrow pointing to the "Untitled1" script editor.
- Environment Panel (Top Right):** Labeled "LISTS PREVIOUSLY ENTERED COMMANDS" (blue text) and "LISTS R OBJECTS" (red text). It shows the "Global Environment" and "Environment is empty".
- Files Panel (Bottom Left):** Labeled "FIND FILES TO LOAD" with a red arrow pointing to the "Files" tab.
- Console (Bottom Left):** Labeled "RUN CODE HERE" with a red arrow pointing to the prompt ">". It displays the R startup message.
- Plots Panel (Bottom Right):** Labeled "PLOTS APPEAR HERE" (blue text) with a blue arrow pointing to the "Plots" tab.
- Packages Panel (Bottom Right):** Labeled "LOAD ADD-ON R PACKAGES" (green text) with a green arrow pointing to the "Packages" tab.
- Help Panel (Bottom Right):** Labeled "R COMMAND HELP PAGES" (orange text) with an orange arrow pointing to the "Help" tab.
- Viewer Panel (Bottom Right):** Labeled "R COMMAND HELP PAGES" (orange text) with an orange arrow pointing to the "Viewer" tab.

Console 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.

>
```

Installing RQDA (Windows)

Dropbox folder: <https://www.dropbox.com/sh/mzsrcfh15o9vy6d/AAD5VyY3sLYJpl7MSimsf3eRa?dl=0>

Before: `install.packages("RQDA", dependencies = TRUE)` # doesn't work anymore

Now: follow my shared instruction R file in the folder "RQDA setup.r"

1. Check R version: R version 3.6.3 (2020-02-29); RStudio version doesn't matter; You can adjust R from "Tools/Global Options" in RStudio setting if you have more than two R version in one computer.
2. Install Rtools35 (don't install the newest Rtools40):
<https://cran.r-project.org/bin/windows/Rtools/history.html>

After installing Rtools35, run the command below:

```
writeLines('PATH="%${RTOOLS35_HOME}\\usr\\bin;%${PATH}%', con = "~/.Renvron")
```

3. Go to SourceForge.net to download GTK+, use the link below (or download from folder)
https://sourceforge.net/projects/gladewin32/files/gtk%2B-win32-runtime/2.10.11/gtk-2.10.11-win32-1.exe/download?use_mirror=versaweb&modtime=1175123376&big_mirror=0

Installing RQDA (Windows)

4. The following 3 packages can be downloaded from CRAN directly:

```
install.packages(c("RGtk2", "cairoDevice", "DBI"))
```

5. The following 3 packages were delisted from CRAN, please go to shared folder to copy-paste the 4 folders to your personal R library (use command `".libPaths()"` in Rstudio console to see the library location, ex., "C:\Users\Henry\Documents\R\win-library\3.6")

- a. gWidgets
- b. gWidgetsRGtk2
- c. RQDA
- d. RSQLite

RQDA package dependencies:

- DBI (<https://cran.r-project.org/web/packages/DBI/index.html>)
- RSQLite (<https://cran.r-project.org/web/packages/RSQLite/index.html>) doesn't support R 4.0+
- gWidgets (<https://cran.r-project.org/web/packages/gWidgets/index.html>) delisted 2020 Sep
- gWidgetsRGtk2 (<https://cran.r-project.org/web/packages/gWidgetsRGtk2/index.html>) delisted 2020

Installing RQDA (Mac)

Before: `install.packages("RQDA", dependencies = TRUE)` # doesn't work anymore

Now: follow my shared instruction R file in the folder "RQDA setup.r"

1. Check R version: R version 3.6.3 (2020-02-29); RStudio version doesn't matter; You can adjust R from "Tools/Global Options" in RStudio setting if you have more than two R version in one computer.
2. Rtools is only for Windows, Mac doesn't need it.
3. Go to SourceForge.net to download GTK+ for Mac osx, use the link below:
<https://sourceforge.net/projects/gtk-osx/>
4. The following 3 packages can be downloaded from CRAN directly:
`install.packages(c("RGtk2", "cairoDevice", "DBI"))`

Installing RQDA (Mac)

5. The following 3 packages were delisted from CRAN, please go to shared folder to copy-paste the 4 folders to your personal R library (Use command “.libPaths()” in Rstudio console to see the library location, i.e. “desktop/Macintosh HD/Library/Frameworks/R.framework/Resources/library”)

- a. gWidgets
- b. gWidgetsRGtk2
- c. RQDA
- d. RSQLite

Official sites of these packages:

- DBI (<https://cran.r-project.org/web/packages/DBI/index.html>)
- RSQLite (<https://cran.r-project.org/web/packages/RSQLite/index.html>) doesn't support R 4.0+
- gWidgets (<https://cran.r-project.org/web/packages/gWidgets/index.html>) delisted 2020 Sep
- gWidgetsRGtk2 (<https://cran.r-project.org/web/packages/gWidgetsRGtk2/index.html>) delisted 2020 Sep

Introduction to RQDA

- Introduction to RQDA Slides:
https://www.dropbox.com/s/cw3j5w6ly9rv49s/Intro_RQDA.pdf?dl=0
- Official website: <https://rqda.r-forge.r-project.org/>
- RQDA review:
 1. <https://www.predictiveanalyticstoday.com/rqda/>
 2. <https://www.r-bloggers.com/2019/04/a-short-r-package-review-rqda/#:~:text=I'd%20say%20RQDA%20is,qualitative%20project%20with%20multiple%20transcripts.>
 3. https://rpubs.com/dastan/rqda_english

Text Analysis

1. Text Data Preprocessing (“tm”, “tidyverse”)
2. Word Cloud (“wordcloud”)
3. Topic Modeling (“topicmodels”)
4. Network Analysis (“igraph”)

More topics to be explored in the future: sentiment analysis, part of speech tagging, word embeddings, word2vec, etc...

Useful Manual Book for Text Analysis

1. Text Analysis in R: <https://m-clark.github.io/text-analysis-with-R/intro.html>
2. Tidy Text Mining with R: <https://www.tidytextmining.com/>
3. Paper: Geva, Hilah, Gal Oestreicher-Singer, and Maytal Saar-Tsechansky. "Using Retweets when Shape our Online Persona: A Topic Modeling Approach." *MISQ* 2019.

Connect with the R community!



cuny.is/rug



cuny.is/dvg

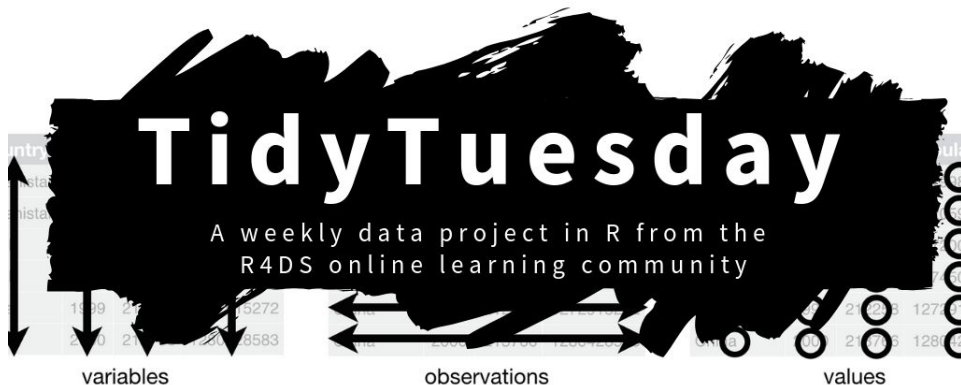


#rstats

#tidyverse

#r4ds

#rspatial



<https://github.com/rfordatascience/tidytuesday>



<https://rladies.org/>



<https://www.rfordatasci.com/>

Learn more!

R Studio Primers

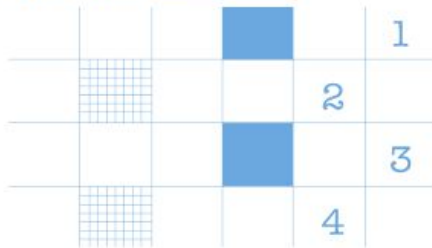
Learn data science basics with the interactive tutorials below.

The Basics



Start here to learn the skills that you will rely on in every analysis (and every primer that follows): how to inspect, visualize, subset, and transform your data, as well as how to run code.

Work with Data



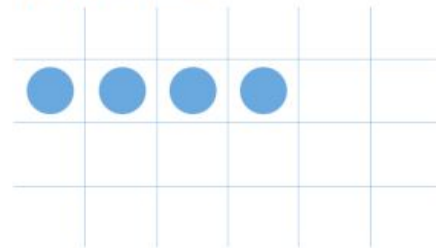
Learn the most important data handling skills in R: how to extract values from a table, subset tables, calculate summary statistics, and derive new variables.

Visualize Data



Learn how to use ggplot2 to make any type of plot with your data. Then learn the best ways to visualize patterns within values and relationships between variables.

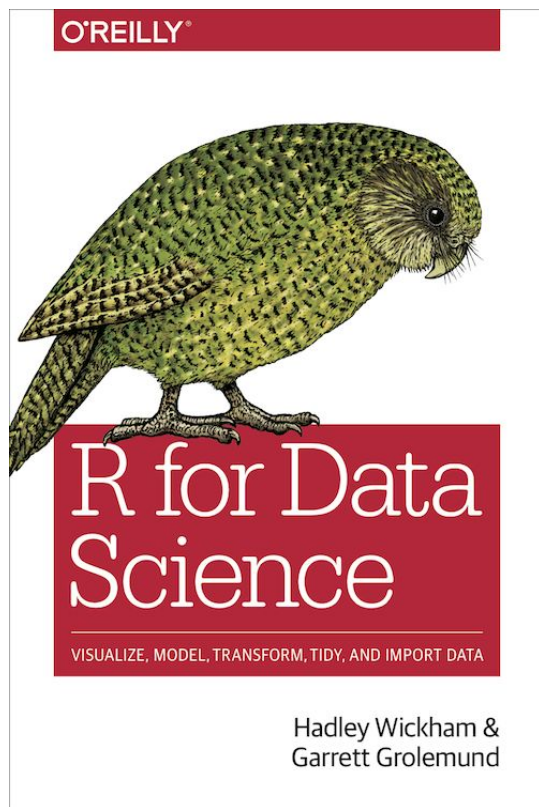
Tidy Your Data



Unlock the tidyverse by learning how to make and use tidy data, the data format designed for R.

<https://rstudio.cloud/learn/primers>

Learn more!



<https://r4ds.had.co.nz/>

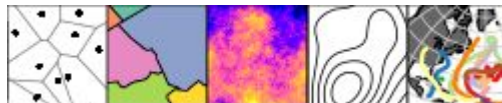


<http://cuny.is/gcdi>



<https://community.rstudio.com/>

r-spatial



<https://www.r-spatial.org/>



<https://stackoverflow.com/>

What next?

Before you leave, **please fill out a workshop evaluation** so we can improve our programming! **cuny.is/gcdievals2020**

- Need more support about what you just learned? Request follow-up individual consultations at **cuny.is/gcdfconsults** (students) or **cuny.is/gcfacultyconsults** (faculty)
- Join the GCDI Group on the CUNY Academic Commons for all GCDI-related updates!
cuny.is/group-gcdi

Thank you for attending and being involved! The *#DigitalGC* is you!