THE COLLECTIVE

February 28-March 2, 2017 | Knoxville, TN

Cooking Up Data Visualizations with R

Convenor(s):

Alison Blaine, Librarian, NCSU Libraries

Erica Hayes, NCSU Libraries, NCSU Libraries

Brief Description of Program:

Are you a librarian who uses Excel? Well, we would love to introduce you to R! R is a free statistical computing and graphics language that can be an extremely useful tool for librarians working with digital data. Using R, you can gather, clean, analyze, and visualize data, as well as create publication-quality graphics. This workshop will provide you with hands-on practice in the basics of R and recipes for cooking up attractive data visualizations. Beginners are welcome!

Tags: visualization, coding, software, instruction

Estimated Session Attendance (Thursday): ~40

Estimated Session Attendance (Friday): ~20

NOTES FROM SESSION:

You will need to download the following software on your laptop to participate:

***R* -** http://archive.linux.duke.edu/cran/

***R Studio* -** https://www.rstudio.com/products/rstudio/download/#download

Goals: 1. Teach you to make graphs with R
2. Provide code recipes for reuse
3. Provide you with raw materials to take home and teach your own R workshops

Activity: Make line/bar graphs, make subsets of data, save graphs, make graphs pretty for presentation/publication
Tidy data: one variable per column, one observation per row
R: base coding language
R studio: most popular graphical user interface for R
R packages: bits of code you can pull into R for your own uses
Packages we will be using today:

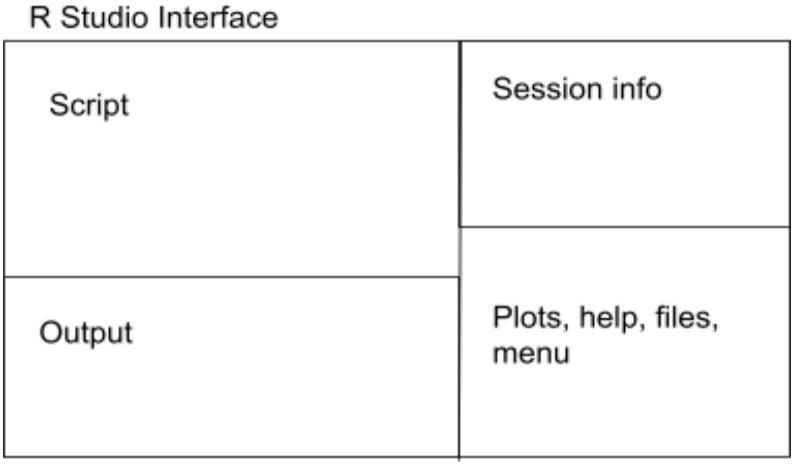
-readr: read in data

-dplyr: for data manipulation. We will use filter(). Cheat sheet: goo.gl/unnidA

-ggplot2: graphing library (2d images, not 3d). We will use ggplot(), geom_line(), geom_bar(), ggsave(). Cheat sheet: https://goo.gl/b2MJJK

-forcats: will help order categorical data. We will use fct_reorder()

Tidyverse: contains many common useful packages



Everyone logging in to R for Activity

Tips:

- Add comments in code using # (will not run with code)
- <- is a common assignment operator (shortcut to use label instead of file name)
- () is always part of a function
- Red text in output does not mean incorrect!
- ?head() will load help and information on the named function (i.e.: “head”)

Set working directory under Files > More > Set as Working Directory

Section 1

- Step 1: Install Packages (only needs to be done once)
- Step 2: Load package libraries (needs to be done every R session)
- Step 3: Load the data set
- Step 4: Use head/tail to look at first/last 6 rows of data
- Step 5: Create a line chart of costs of collection types over time (use “zoom” to view)
- Step 6: Create a line chart showing number of users
- Step 7 & 8: Create subplots (takes previous graph and puts each different line in its own graph, can make things easier to read) using facet_wrap()

Section 2

- Step 9: Create a new variable to make a new chart that looks at one variable

- Step 10: Make a plot of new variable (journal)
- Step 11: use theme_classic() to make it prettier (removes gray squares) (there are many other themes available, will use “fill in” feature to suggest themes)
- Step 12: Add a title using labs()
- Step 13: Add a legend title using the labs() command “color=”
- Step 14: Fix x axis title
- Step 15: Save chart to a variable named “journals_chart”
- Step 16: Save variable named in previous step to an image file, will show up in your working directory

Section 3

- Step 17: Create a bar chart
- Step 18: Use coord_flip() to change from vertical to horizontal bars
- Step 19: Order the bars from greatest to least by setting the category as “ordered,” choose to order based on cost, and then graphing the reordered variable
- Step 20: Add title, subtitle, and axis labels using labs()