Intro to R

Data Services @ the Cushing/Whitney Medical Library

Yale UNIVERSITY LIBRARY
Harvey Cushing/John Hay Whitney Medical Library

Workshop content

1. Defining R
2. Running R
3. R Basics
4. Demo in RStudio

Defining R

Yale UNIVERSITY LIBRARY
Harvey Cushing/John Hay Whitney Medical Library

What is R?

- R is a scripting language and programming environment for data manipulation, calculations, and graphical display
- R is open source and free
- R includes:
  - Data handling
  - Basic arithmetic and comparison operators
  - Robust functions

R vs Python

	R	Python
Purpose	A scripting language that focuses on user friendly data analysis, statistics and graphical models	A C based object oriented language that e emphasizes productivity and code readability
User base	R has been used primarily in academics in research. • Statisticians and biostatisticians • Medical research • Data Scientists	Python is used by many as a “learning language” • Developers & programmers • Data Scientists • Researchers

https://www.guru99.com/r-vs-python.html

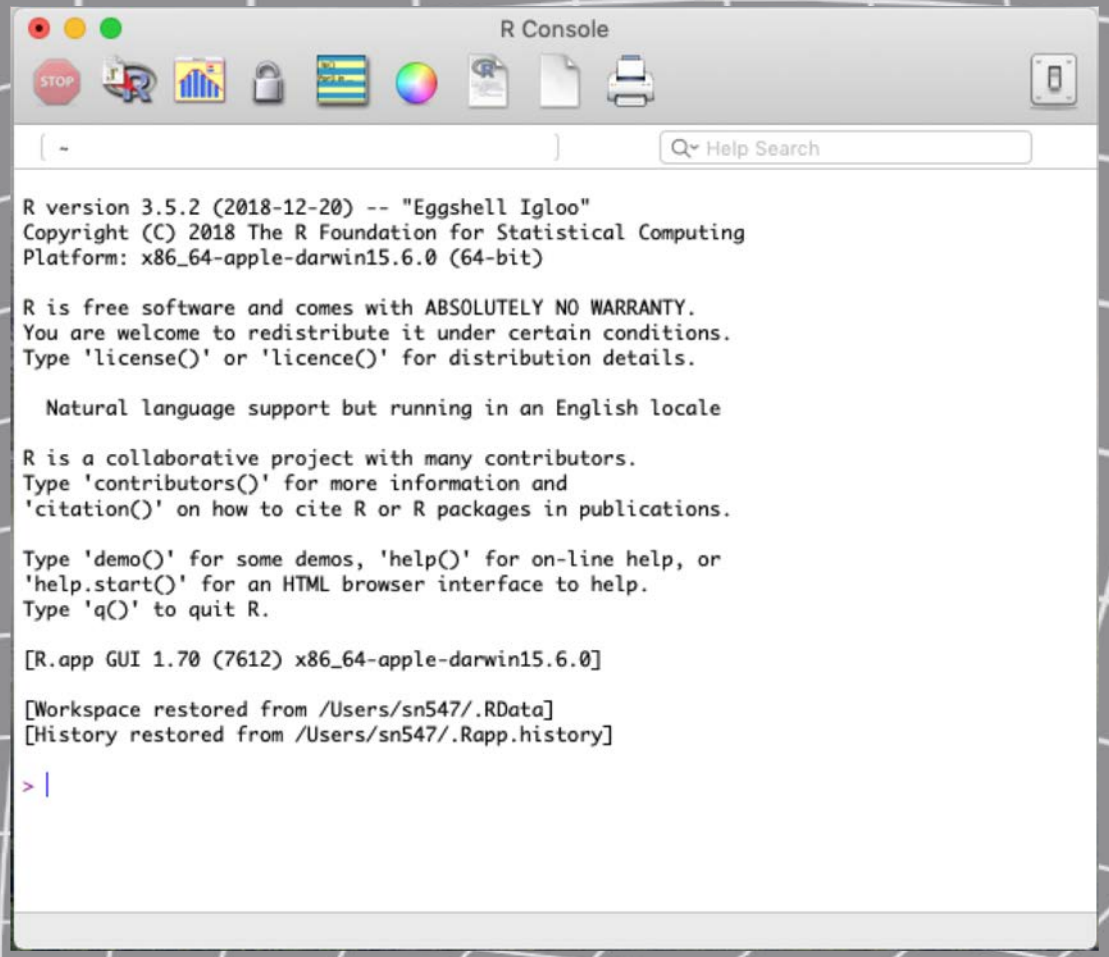
Running R

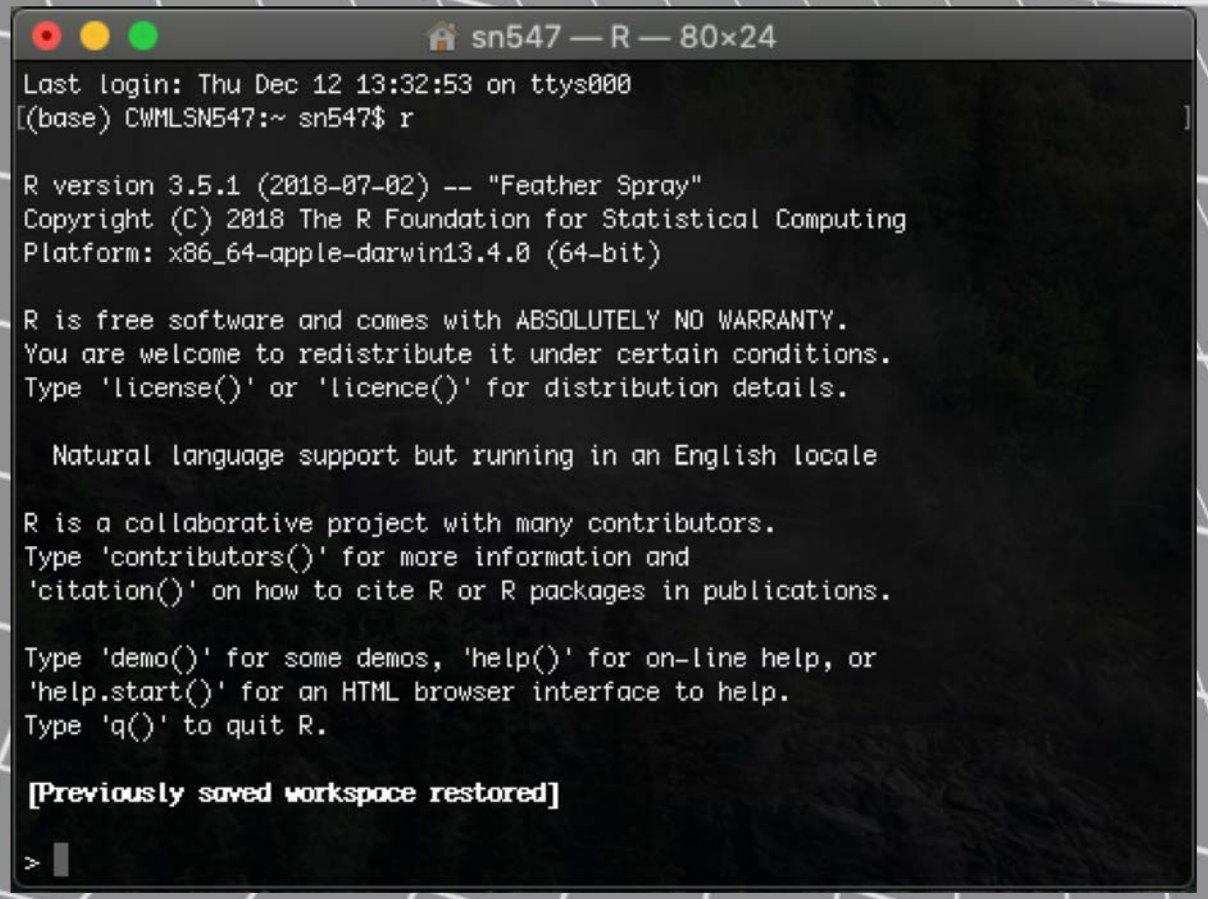
Yale UNIVERSITY LIBRARY
Harvey Cushing/John Hay Whitney Medical Library

Where can you run R?

- Terminal window
  - You can run R in HPC
- R Console window
- Jupyter Notebook
- RStudio
- Anaconda (RStudio)

R via command line interfaces





RStudio

- RStudio is a (more) graphical user interface for running the R language
- RStudio is not necessary for running the R language

RStudio interface

~/Desktop/calibration belt validation - RStudio

Go to file/function | Addins | calibration belt validation

Untitled1* | icuData_selected | file | icuData

Source on Save | Run | Source

```
1  install.packages("givitiR")
2  library(dplyr)
3  library(givitiR)
4  data("icuData")
5  icuData_selected <- select(icuData, outcome, probSaps)
6  file <- read.csv("stroke_dum_nomiss.csv")
7  file_selected <- filter(file, nomo_risk > 0 & nomo_risk < 1)
8  cb <- givitiCalibrationBelt(o = icuData$outcome, e = icuData$pr
9                              devel = "external")
10 cb2 <- givitiCalibrationBelt(o = file_selected$stroke, e = file
11 summary(file$nomo_risk)
12
13 plot(cb, main = "SAPSII calibration",
```

11:24 (Top Level) R Script

Console | Terminal

~/Desktop/calibration belt validation/

```
> cb <- givitiCalibrationBelt(o = icuData$outcome, e = icuData$probSa
ps,
+                             devel = "external")
> plot(cb, main = "SAPSII calibration",
+      xlab = "xlab",
+      ylab = "ylab")
$m
[1] 2

$p.value
[1] 6.269563e-10

>
```

Environment | History | Connections

Import Dataset | List

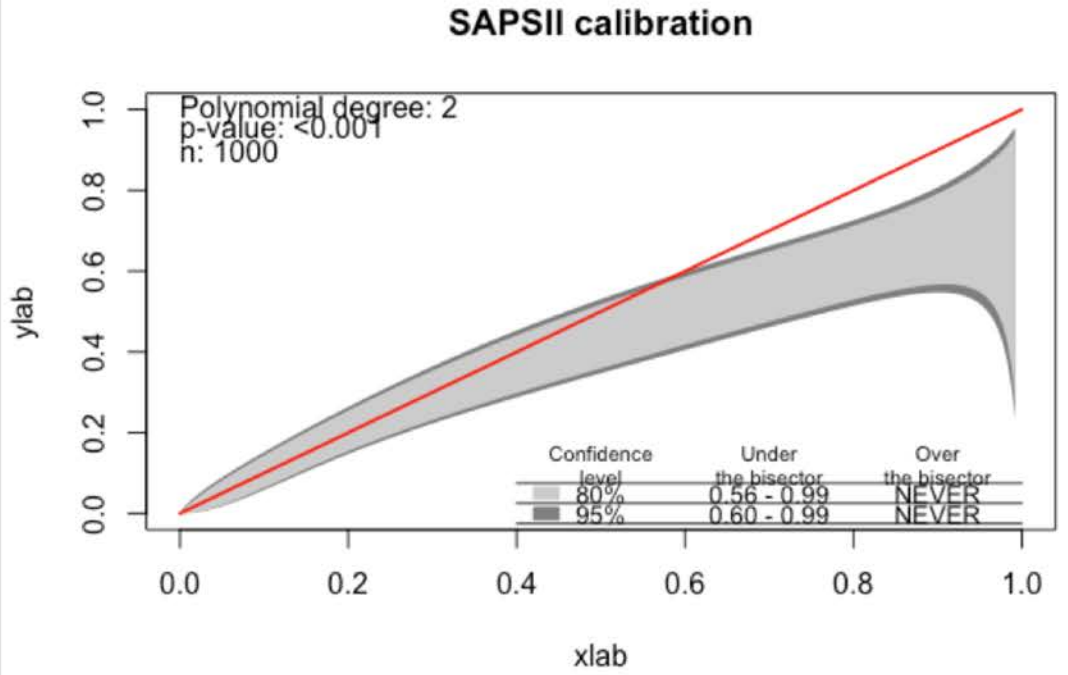
Global Environment

Data

cb	List of 10
file	1487662 obs. of 170 variables
file_selected	1460256 obs. of 170 variables
icuData	1000 obs. of 33 variables
icuData_selected	1000 obs. of 2 variables

Files | Plots | Packages | Help | Viewer

Zoom | Export | Publish



R basics

Yale UNIVERSITY LIBRARY
Harvey Cushing/John Hay Whitney Medical Library

Variables

- Variables are stored data within your R environment
- Stored data might include
  - Basic data types (string, numeric, or Booleans)
  - Data structures (lists and data frames)

Functions

- Set instructions perform specific tasks
- Are usually package specific (or are part of “base R”)
- Can be user defined (i.e., you can create your own functions)

```
#Here is an example of a function that would read a csv file into your R environment
> read.csv(“data_folder/data_file.csv”)
```

R Packages

Packages are bundles of additional R functions for specific purposes

Example packages:

- Bioconductor
- Seurat
- Tidyverse (dplyr, tidyr, stringr, ggplot2)
- Haven

```
# To install packages in R, use the following syntax
> install.packages("package_name")
```

Continued learning after class

LinkedIn Learning through Yale

- Google "Yale LinkedIn Learning" to sign in through with your netID

Recommended tutorials, texts, and sources for further information

- library.medicine.yale.edu/research-data/data-tools-software/about-r
- Find more classes on R at Yale: https://library.medicine.yale.edu/research-data/classes-materials

Finding R Help

Cushing/Whitney Medical Library

- Data Office Hours
- StatLab Consultants

On the web

- R Project Manuals and FAQs
- Stack Overflow
- Cheat sheets

```
# Find information on functions using the script:
> ?function_name
```

Questions before we move on to the demonstration?

medicaldata@yale.edu

Yale UNIVERSITY LIBRARY
Harvey Cushing/John Hay Whitney Medical Library

Demo Topics

- RStudio project initialization
- Using RStudio
- Data types (strings, Booleans, and numeric data)
- Data structures (lists and data frames)
- Data import and export
- Performing basic statistical functions
- Creating basic graphs