

Introduction

BIOSTAT 213: Introduction to Computing in R

What is R?

What is statistical software?

- ▶ statistical software is software for conducting **statistical analysis**
- ▶ popular examples include: **R**, Stata, SAS, and IBM SPSS
- ▶ each software program often has its own **programming language**
- ▶ the language allows you to **write instructions** to your computer

What is R?

- ▶ R is a popular **statistical software** program
- ▶ the term also refers to the **programming language**

What is R?

- ▶ R is **open source**, meaning that the computer code behind it is **publicly available** online
- ▶ in contrast, for example, Apple does not publicly release the code behind macOS or iOS
- ▶ development (planning and improvement, etc) of R does not occur within any one private company, but rather within a **community of individual contributors**, from a variety of organizations
- ▶ there is no official technical support, but there are various **online forums and other resources**

Why use R?

- ▶ it is extremely **powerful**
 - ▶ unlike most other statistical software, it supports using **multiple data objects** and various types of objects (more on this during orientation and throughout the quarter)
 - ▶ it supports a **wide array of methods**, in large part because there are numerous (more than 12,000!) add-on **packages**
- ▶ it is extremely **popular** in academia as well as in industry
 - ▶ there are numerous **books and other resources** for using R
 - ▶ it is a **marketable skill!**
- ▶ it is **free of cost**

What is RStudio?

- ▶ RStudio is an **add-on interface** (a set of “windows”) for R
- ▶ RStudio includes numerous features including:
 - ▶ an interface (“window”) for **writing and editing code**
 - ▶ an interface for **browsing files** on the computer
 - ▶ an interface to R’s built-in documentation, with clickable links
 - ▶ an interface for browsing and viewing **data objects**
- ▶ we’ll discuss these features in greater depth on the first day of class

Installing R and RStudio

Installing R

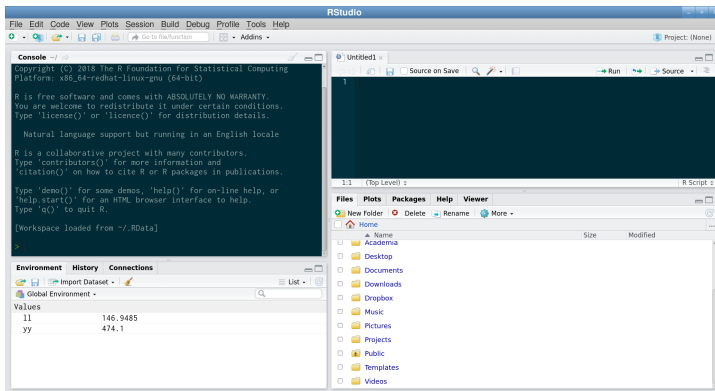
- ▶ if your computer's operating system is **macOS** (Apple) or **Windows** (Microsoft), visit cran.r-project.org/ and select the link for your operating system
 - ▶ for **macOS**, select the **latest release** (at the top of the page) and double-click the downloaded file
 - ▶ for **Windows**, select **base**, select the link at the top of the page, and double-click the downloaded file
- ▶ if your computer's operating system is **Linux**, the easiest way to install R should be via your package manager

Installing RStudio

- ▶ visit www.rstudio.com/products/rstudio/download/
- ▶ select the download link for the RStudio Desktop Open Source License
- ▶ select the download link for your operating system
- ▶ double-click the downloaded file

Verifying successful installation

- ▶ open **RStudio** (double-click or click ...)
- ▶ you should see something like this:



Installing R and RStudio

- ▶ please feel free to email me if you have trouble installing R or RStudio:
yea-hung.chen@ucsf.edu

Syllabus

Introduction

**BIOSTAT 213:
Introduction to
Computing in R**

What is R?

Installing R and
RStudio

Syllabus

Course objectives

Course schedule

Class format

Assignments

Where and when

Course instructors

Course objectives

Course objectives

*This is not a **statistics** class. This is not a **statistics** class. This is not a **statistics** class. This is not a **statistics** class. This is not a **statistics** class. This is not a **statistics** class. This is not a **statistics** class. This is not a **statistics** class. This is not a **statistics** class. This is not a **statistics** class.*

Introduction

BIOSTAT 213:
Introduction to
Computing in R

What is R?

Installing R and
RStudio

Syllabus

Course objectives

Course schedule

Class format

Assignments

Where and when

Course instructors

Course objectives

This is not an epidemiology class. This is not an epidemiology class.
This is not an epidemiology class. This is not an epidemiology class.
This is not an epidemiology class. This is not an epidemiology class.
This is not an epidemiology class. This is not an epidemiology class.
This is not an epidemiology class. This is not an epidemiology class.

Introduction

BIOSTAT 213:
Introduction to
Computing in R

What is R?

Installing R and RStudio

Syllabus

Course objectives

Course schedule

Class format

Assignments

Where and when

Course instructors

Course objectives

After successful completion of the course, you should be able to:

- ▶ assign objects in R and identify R object types
- ▶ save R work
- ▶ import data files into R
- ▶ conduct basic statistical analysis in R
- ▶ [manipulate data objects](#) in R
- ▶ write functions in R and [conditionally or iteratively apply R code](#) and functions
- ▶ generate reports using [R Markdown](#)

In addition, after completing the course, you should be able to undertake other UCSF courses that require prior experience with R.

Course schedule

Module 1: fundamentals (Wednesday, January 16, 2019)

- ▶ introduction to R and RStudio
- ▶ mathematical operators and functions
- ▶ object assignment
- ▶ numeric objects
- ▶ character objects
- ▶ scripts
- ▶ working directories and images

Module 2: univariate analysis (Wednesday, January 23, 2019)

- ▶ reading CSV files
- ▶ data frames
- ▶ factors
- ▶ means, standard deviations, and quantiles
- ▶ histograms and stem-and-leaf plots
- ▶ tables and `prop.table()`
- ▶ univariate hypothesis tests

Introduction

BIOSTAT 213:
Introduction to
Computing in R

What is R?

Installing R and
RStudio

Syllabus

Course objectives

Course schedule

Class format

Assignments

Where and when

Course instructors

Module 3: bivariate and multivariate analysis (Wednesday, January 30, 2019)

- ▶ scatter plots and `loess()`
- ▶ box plots
- ▶ introduction to *ggplot2*
- ▶ bivariate tables and `prop.table()`
- ▶ bivariate hypothesis tests
- ▶ linear regression

Module 4: data manipulation (Wednesday, February 6, 2019)

- ▶ logical objects
- ▶ missing values
- ▶ indexing
- ▶ `subset()`
- ▶ `ifelse()`
- ▶ date objects

Introduction

BIOSTAT 213:
Introduction to
Computing in R

What is R?

Installing R and
RStudio

Syllabus

Course objectives

Course schedule

Class format

Assignments

Where and when

Course instructors

Module 5: advanced data manipulation (Wednesday, February 13, 2019)

- ▶ reading other data files
- ▶ `rbind()` and `cbind()`
- ▶ merging data frames
- ▶ reshaping data and *tidyr*
- ▶ *dplyr*

Introduction

BIOSTAT 213:
Introduction to
Computing in R

What is R?

Installing R and
RStudio

Syllabus

Course objectives

Course schedule

Class format

Assignments

Where and when

Course instructors

Module 6: programming (Wednesday, February 20, 2019)

- ▶ writing functions
- ▶ `if()`
- ▶ `for()`
- ▶ the apply functions
- ▶ lists
- ▶ regular expressions

Introduction

BIOSTAT 213:
Introduction to
Computing in R

What is R?

Installing R and
RStudio

Syllabus

Course objectives

Course schedule

Class format

Assignments

Where and when

Course instructors

Module 7: report generation (Wednesday, February 27, 2019)

- R Markdown

Introduction

BIOSTAT 213:
Introduction to
Computing in R

What is R?

Installing R and
RStudio

Syllabus

Course objectives

Course schedule

Class format

Assignments

Where and when

Course instructors

Class format

Class format

- ▶ roughly 1.5 hours of lecture, interspersed with **hands-on exercises**
- ▶ roughly 1 hour of **additional hands-on exercises**

Introduction

BIOSTAT 213:
Introduction to
Computing in R

What is R?

Installing R and
RStudio

Syllabus

Course objectives

Course schedule

Class format

Assignments

Where and when

Course instructors

Assignments

Introduction

BIOSTAT 213:
Introduction to
Computing in R

What is R?

Installing R and
RStudio

Syllabus

Course objectives

Course schedule

Class format

Assignments

Where and when

Course instructors

Assignments

- ▶ 3 homework assignments

Introduction

BIOSTAT 213:
Introduction to
Computing in R

What is R?

Installing R and
RStudio

Syllabus

Course objectives

Course schedule

Class format

Assignments

Where and when

Course instructors

Where and when

Where and when

- ▶ Wednesdays, 1–3:30 pm
- ▶ Mission Hall 1407

Introduction

BIOSTAT 213:
Introduction to
Computing in R

What is R?

Installing R and
RStudio

Syllabus

Course objectives

Course schedule

Class format

Assignments

Where and when

Course instructors

Course instructors

- ▶ R user for roughly 18 years
- ▶ research projects:
 - ▶ studying [risk behaviors for HIV transmission](#) (with the San Francisco Department of Public Health)
 - ▶ [SEARCH](#), a randomized community trial of the test-and-treat strategy (Division of HIV, Department of Medicine, UCSF)

Adrienne Epstein, MS

- ▶ R user for roughly 2 years
- ▶ student in UCSF's [Epidemiology and Translational Science](#) doctoral program

Introduction

BIOSTAT 213:
Introduction to
Computing in R

What is R?

Installing R and RStudio

Syllabus

Course objectives

Course schedule

Class format

Assignments

Where and when

Course instructors