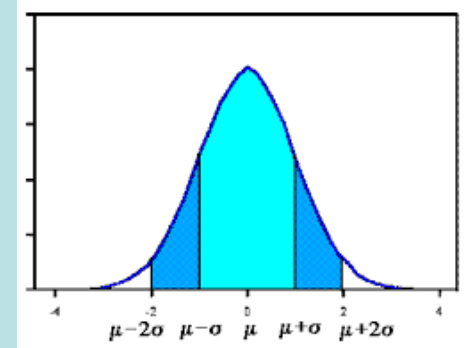




Spring 2020

Biometry

BIOL 3090



This course provides an introduction to the theory and practice of statistics and focuses on hypothesis testing, experimental design, and the interpretation of statistical results. Course lectures cover summary statistics, normality diagnostics, Z-scores, t-tests, correlation, regression, ANOVA, and ANCOVA. Practical assignments using R software and examples drawn from the biological sciences will augment instruction on statistical principles and methods. Pre-requisite: BIOL 2052.

---

Mon & Weds, 13:30 - 14:45, AC 203 (HLC)

Undergraduate, 3 Credits,

Instructor: David Hyrenbach ([khyrenbach@hpu.edu](mailto:khyrenbach@hpu.edu))

# What are Statistics ?

**My Definition:** The study of the collection, organization and interpretation of data.

They are involved in all stages of scientific research, from the planning of data collection (design of surveys and experiments) to the reporting and interpretation of results.

# Why use Statistics ?

To plan efficient sampling of patterns.

To rigorously quantify and compare observations.

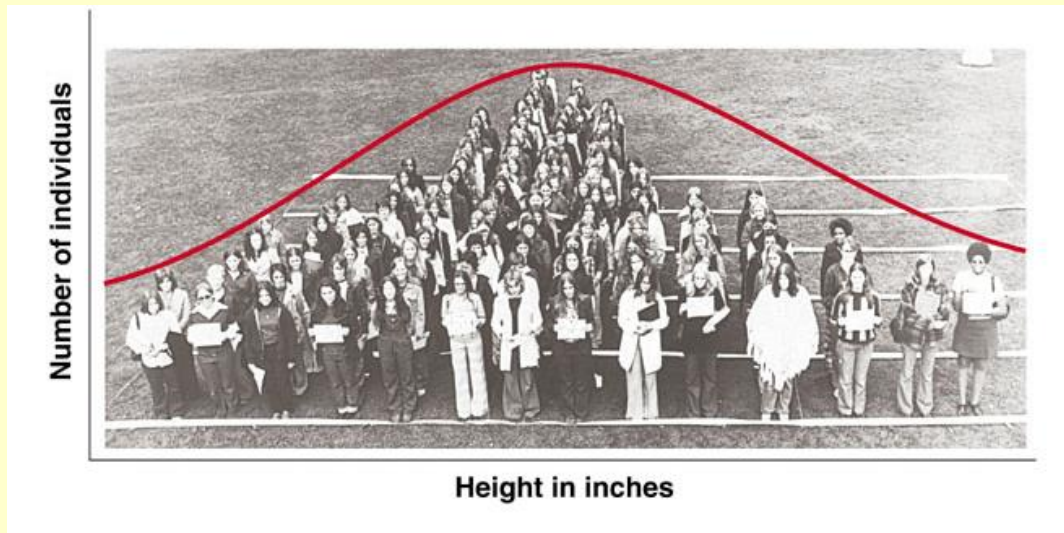
To develop models (and understanding) of patterns.



“Data don’t make any sense,  
we will have to resort to statistics.”

# Using Statistics to Summarize Data

Facilitate characterization of a single population and the comparison of multiple populations



Describe "shape" of frequency distributions:  
(range, mean, standard deviation, median)



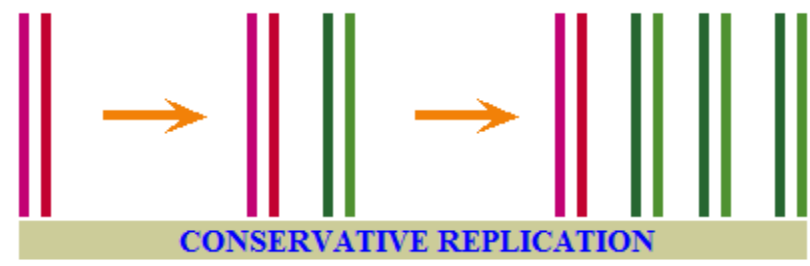
Compare populations  
(using statistical tests)

# Using Statistics to Test Predictions

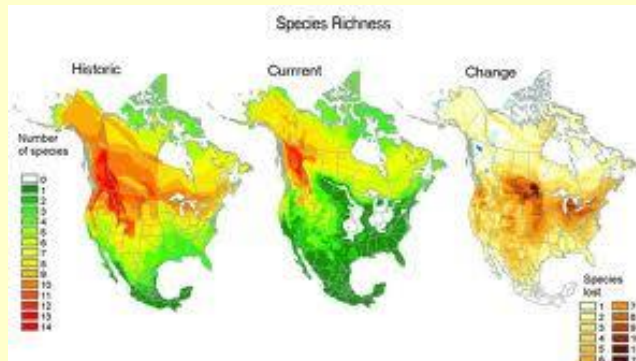
- Simple systems: Two Models of DNA Replication

SEMI-CONSERVATIVE

CONSERVATIVE



- Complex systems: models of species distributions



# Statistical Framework for this Class

## Fisherian Approach:

Makes assumptions about the underlying data distributions.

Uses only the data to estimate parameters.

Calculates probability that the pattern found in the sample occurred "by chance" (sampling).



Sir Ronald Aylmer Fisher  
(1890 - 1962)

Analysis of variance  
Fisher's exact test  
F-distribution



# Example

Tasked with comparing nest density of ground-foraging ant species in two adjacent habitats: agricultural field / forest.



Delineate Two Study Areas

Sample Randomly (lat / long)

Standardize Survey Effort (1 m x 1 m)



# Data Collection



Field#1: 6 nests / m<sup>2</sup>

Field#2: 10 nests / m<sup>2</sup>

Field#3: 12 nests / m<sup>2</sup>

Field#4: 12 nests / m<sup>2</sup>



Forest#1: 6 nests / m<sup>2</sup>

Forest#2: 9 nests / m<sup>2</sup>

Forest#3: 4 nests / m<sup>2</sup>

Forest#4: 6 nests / m<sup>2</sup>

Forest#5: 7 nests / m<sup>2</sup>

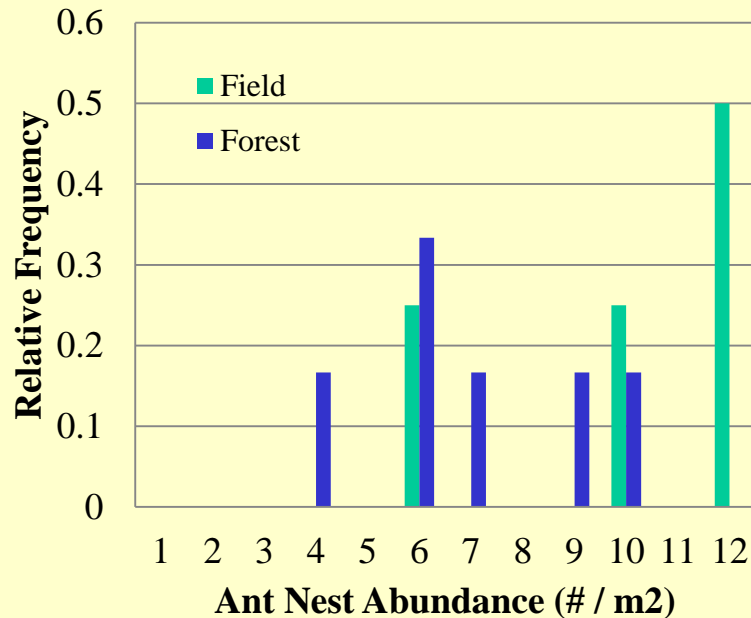
Forest#6: 10 nests / m<sup>2</sup>



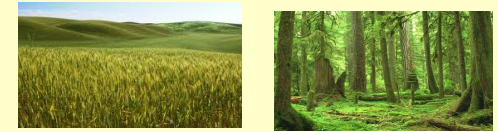
# Data Summarization

What is the simplest way to summarize these observations ?

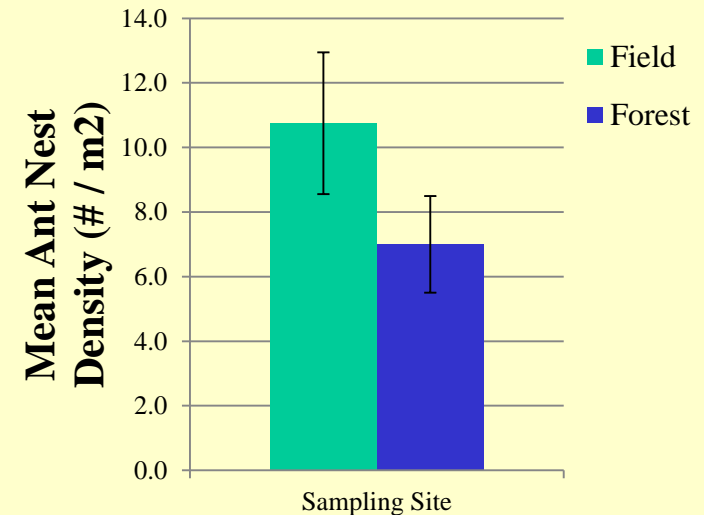
**Ant Abundance Comparison**



Frequency distribution of ground foraging ant nests at two adjacent sampling sites: an agricultural field and a forest.



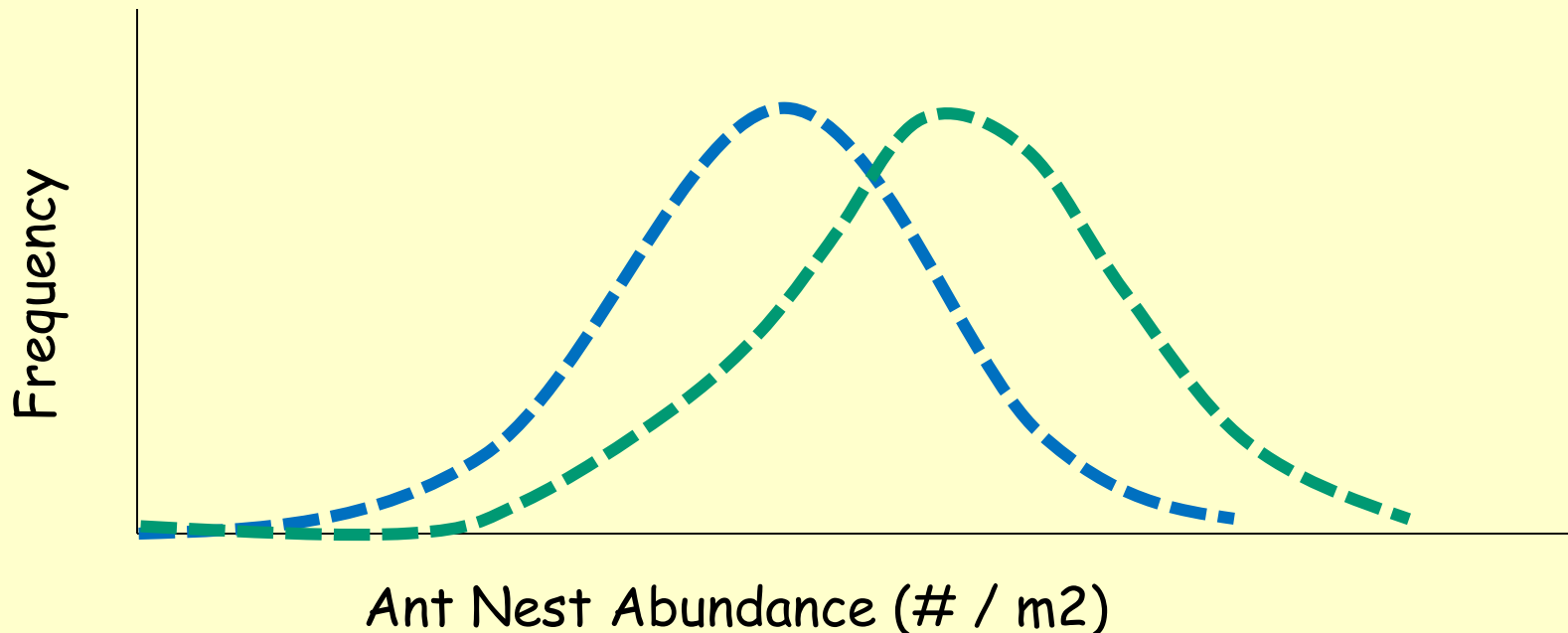
RANGE	6 - 12	4 - 10
MEAN	10.2	7.0
ST. DEV.	2.8	2.2



# Using Parametric Statistics

Based on the assumption that these data were sampled from a specific distribution - a Normal Distribution.

Assume that the parameters describing the frequency distribution of the sample (mean =  $\bar{x}$  ; s.d. =  $s$ ) are unbiased estimators of population parameters (mean =  $\mu$  ; s.d. =  $\sigma$ ).

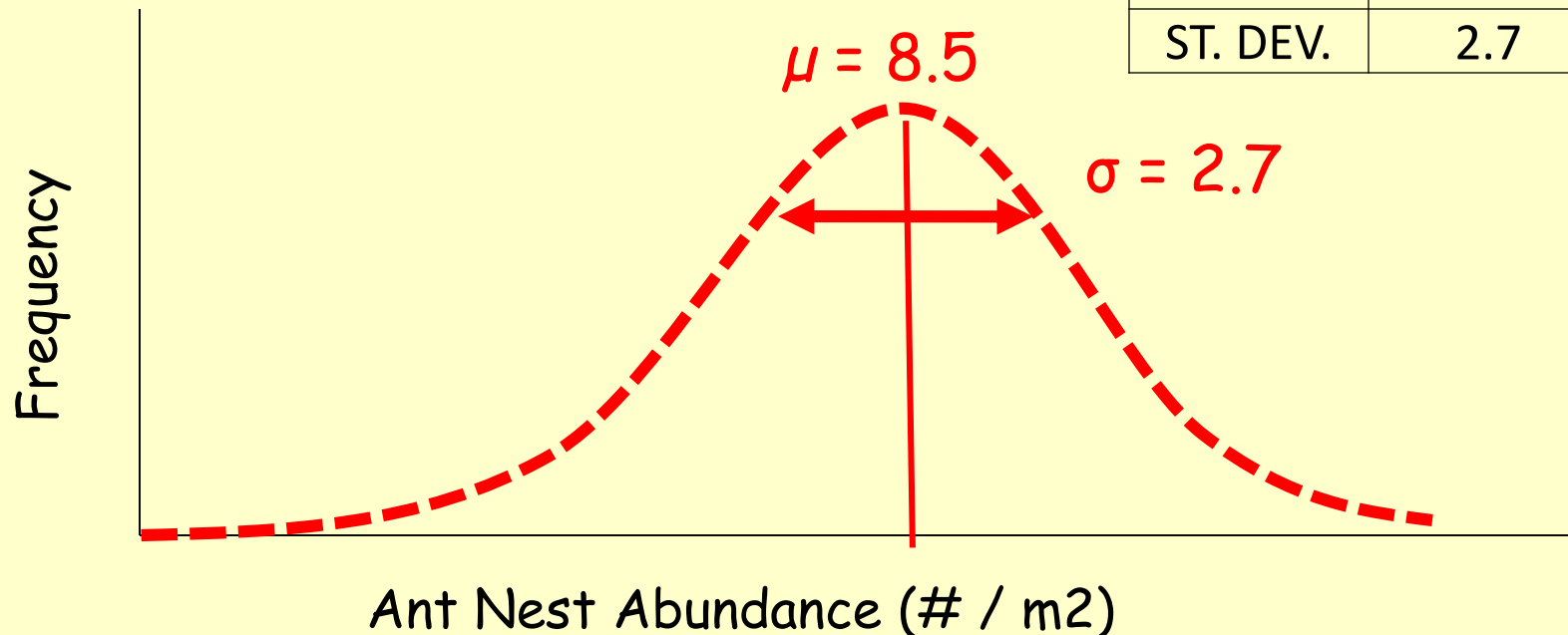


# Using Parametric Statistics

## Specifying the Null Hypothesis

**Null Hypothesis:** There is no difference... No Pattern.  
(In other words: both populations are really the same)

RANGE	4 - 12
MEAN	8.5
ST. DEV.	2.7

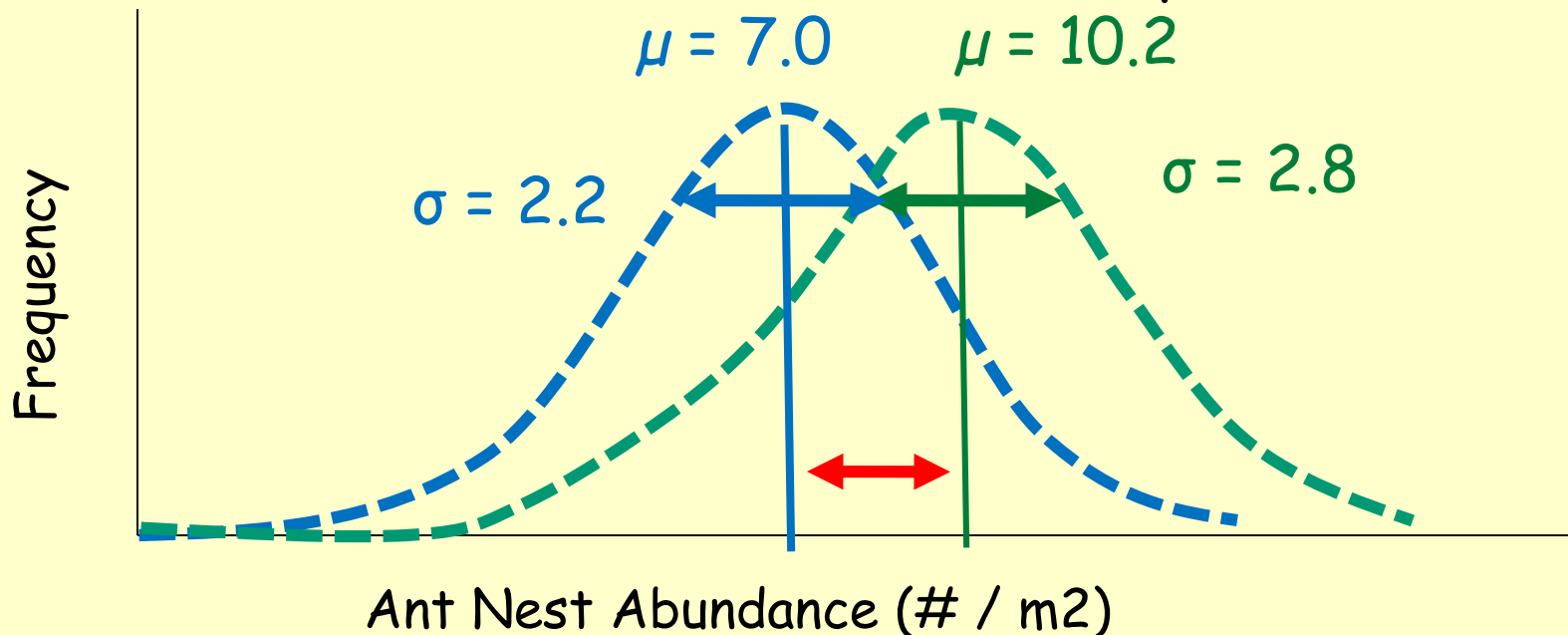


# Using Parametric Statistics

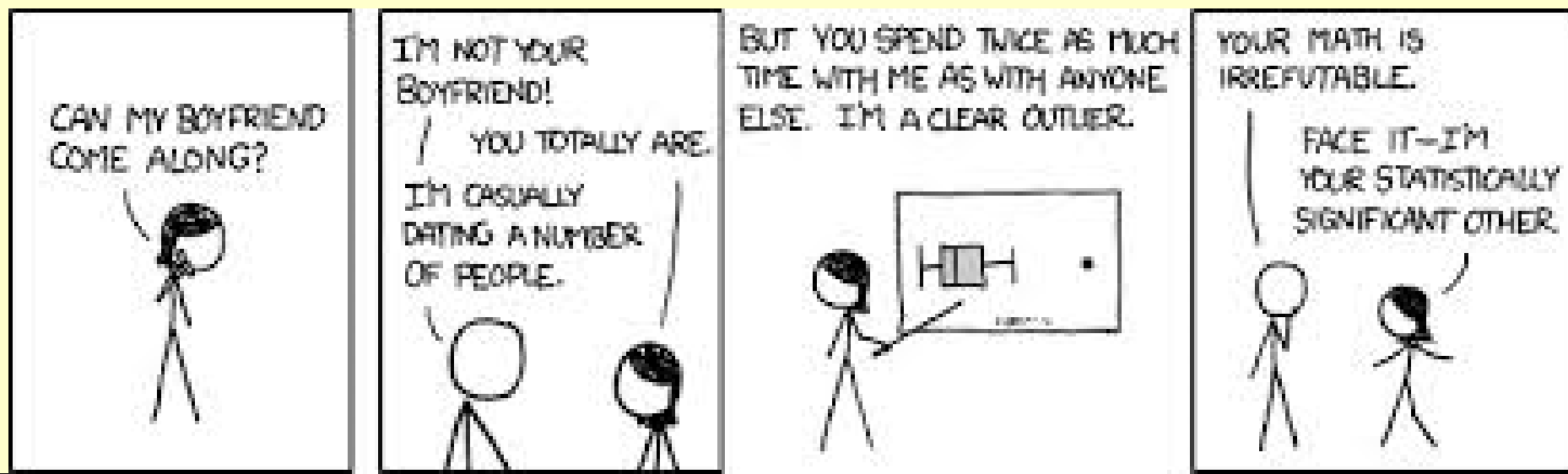
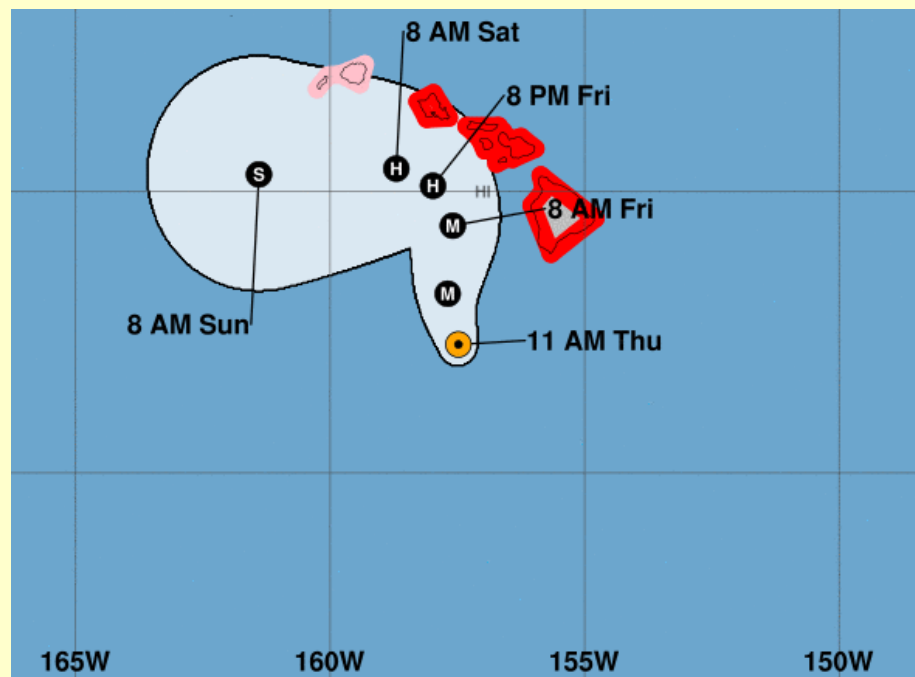
Once we have made the assumption of normality, we can proceed with the statistical comparison of the two sites:

## Specifying the test statistic

- Compare means (central tendency)
- Relate to their S.D.s (variability)



# Exposed to Statistics in Every-Day Life





# Learning to Use Statistics Correctly

Critical  
evaluation  
of the  
scientific  
literature

Elegant and  
rigorous  
study  
design and  
analysis



"I can prove it or disprove it! What do you want me to do?"

# Big Data:

A Revolution That is Transforming How We Live, Work, and Think



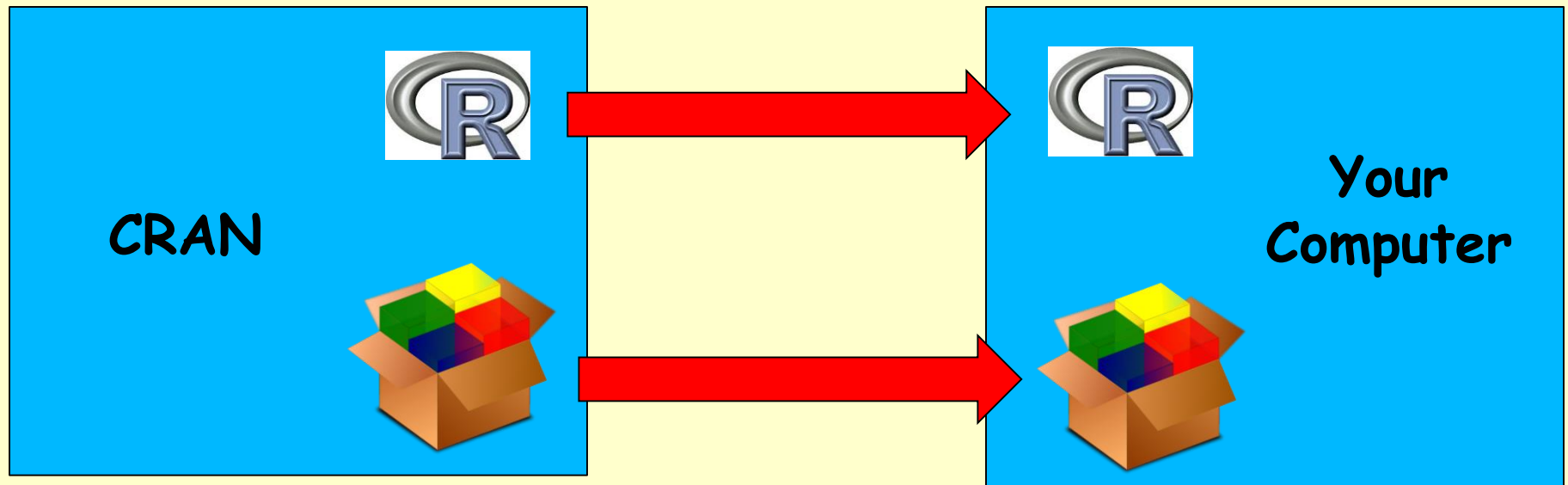
[www.ted.com/talks/susan\\_etlinger\\_what\\_do\\_we\\_do\\_with\\_all\\_this\\_big\\_data](http://www.ted.com/talks/susan_etlinger_what_do_we_do_with_all_this_big_data)

# Why R ?


Powerful, Versatile, Dynamic, Free !!!



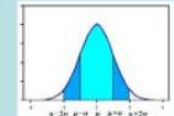
## The R Architecture



# Course Web-Site



Spring 2020  
Biometry  
BIOL 3090



Theoretical foundations and practical application of statistics to the synthesis, representation, and analysis of data. Through homework sets, quizzes and the use computer software applications, students will learn and apply a variety of statistical tests: contingency tables, t-tests, correlation, regression and analysis of variance. These applications will be presented in the context of experimental design and hypothesis testing.

Mon & Weds, 13:00 - 14:15, AC 203 (HLC)

---

Undergraduate, 3 Credits,  
Instructor: David Hyrenbach (khyrenbach@hpu.edu)

## BIOL 3090 Biometry

*Last Updated January 8, 2020*

Please report any problems here:  
[khyrenbach \(at\) hpu \(dot\) edu](mailto:khyrenbach@hpu.edu)

### Syllabus

(Last Updated: Jan 8)

### Quiz Keys

Quiz#1

Instructor	
Dr. David Hyrenbach <i>Offices:</i> AC 249, HLC EMSB, Oceanic Institute <i>Phone:</i> (808) 236-3563	<i>Office Hours:</i> HLC: Weds; 9:00- 12:00  OI: Tu & Th; by appointment

Lectures (pdfs)

Readings  
Assigned  
Extra

Homework Keys

Quiz Keys

[http://www.pelagicos.net/classes\\_biometry\\_sp20.htm](http://www.pelagicos.net/classes_biometry_sp20.htm)

# Course Objectives

The main focus of this course is to provide students with the background of knowledge and the tools necessary to select, perform and interpret statistical analyses of biological data.

Another major focus of this course is to review the philosophical underpinnings of the scientific method and how hypothesis testing is implemented.

Achievement of the learning outcomes will be assessed using assignments, quizzes, an individual project and in-class participation.



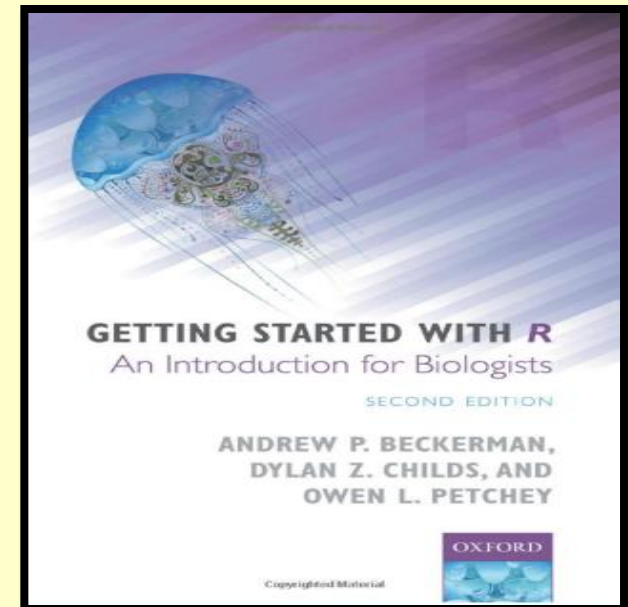
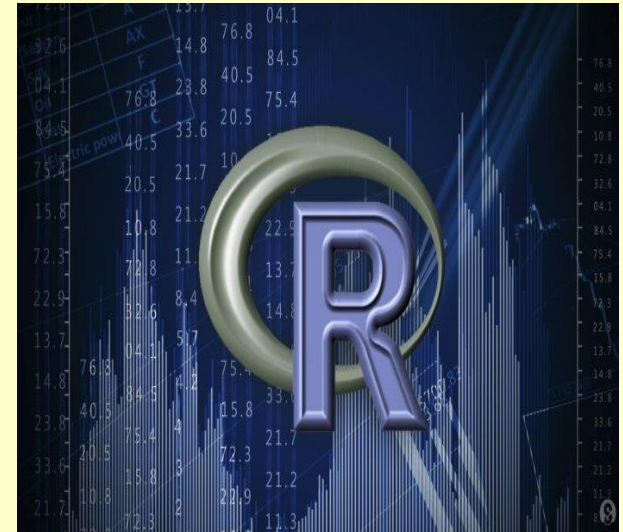
# Practical Course Objectives

A key goal of this course is to provide students with the tools necessary to perform statistical tests and data analyses ...

during this course

and

beyond this course



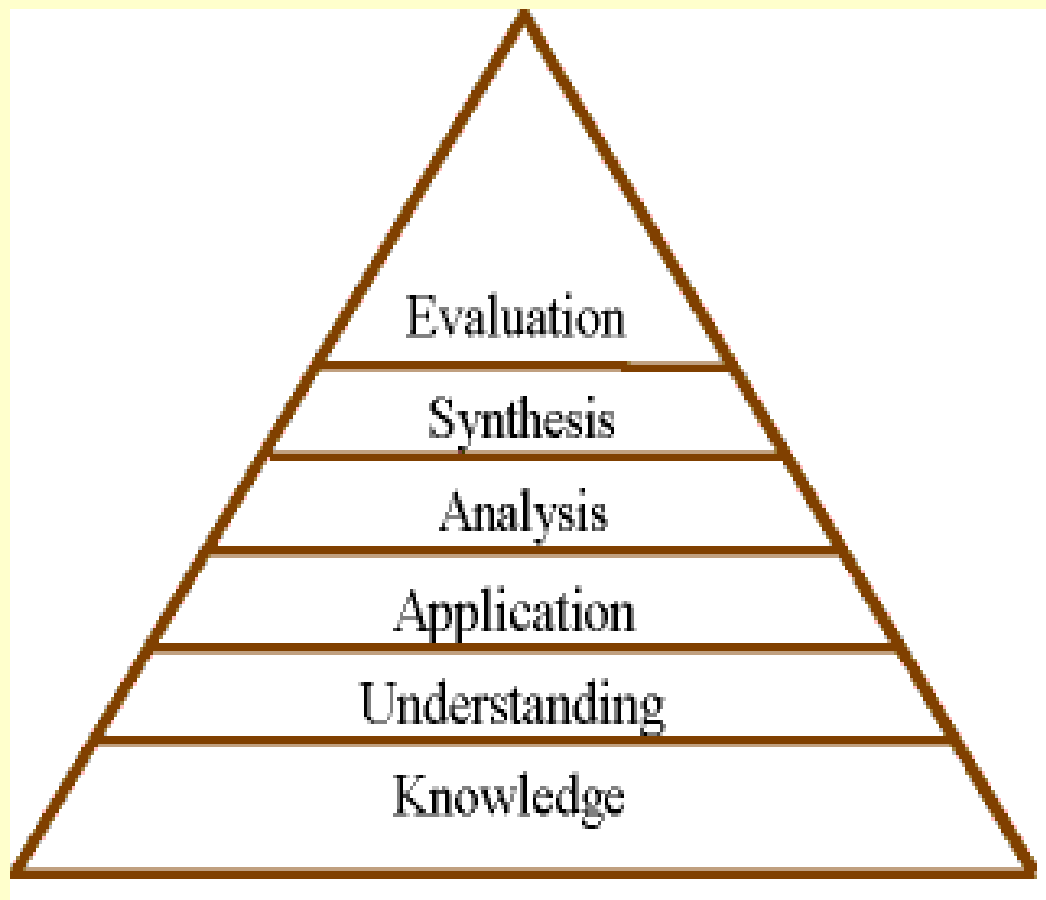
# Pragmatic Course Objectives

In addition to establishing a foundation of statistical knowledge and methods, I hope students will take away 3 ideas from this course:

- (1) Statistics is an applied field with a wide range of practical applications.
- (2) You do not have to be a math guru to learn from real, interesting data.
- (3) Data are messy, and statistical tools are imperfect. But, when you understand the strengths and weaknesses of these tools, you can use them to learn about the real world.

# Course Structure

*Bloom's Taxonomy: (Bloom, 1956)  
describes six levels of cognitive domains*



***Evaluation:***

appraise, argue, evaluate

***Synthesis:***

arrange, develop, formulate

***Analysis:***

analyze, compare, contrast

***Application:***

apply, employ, practice

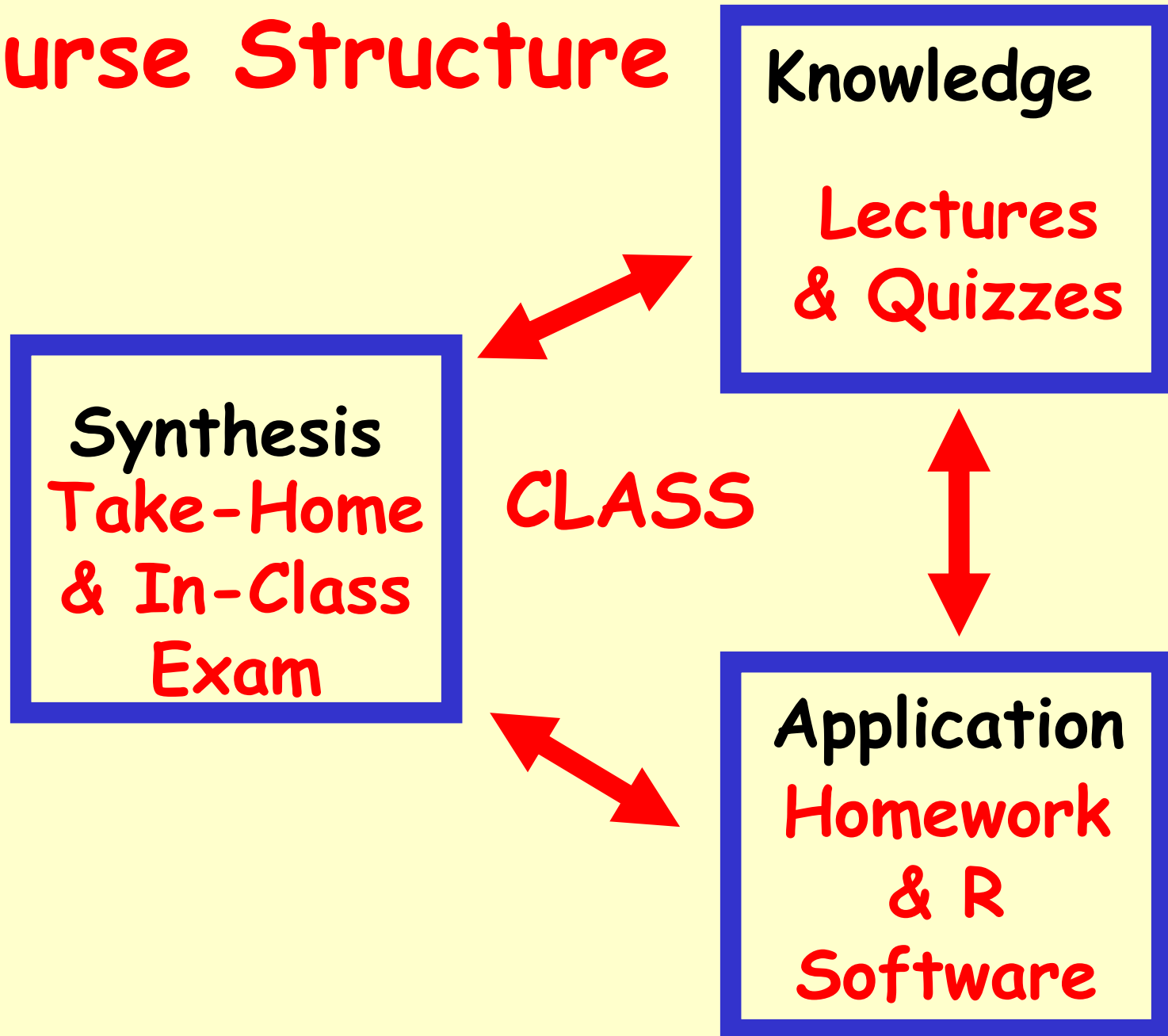
***Understanding:***

describe, discuss, explain

***Knowledge:***

define, label, list

# Course Structure



# Meeting Times / Places

➤ Meetings: M & W 13:30 - 14:45, AC 203, HLC

➤ Office Hours:

@ HLC, AC 249:

- Weds
- 9:00 - 12:00

@ EMSB, OI

- Tues & Thurs
- By appointment





# Required Readings

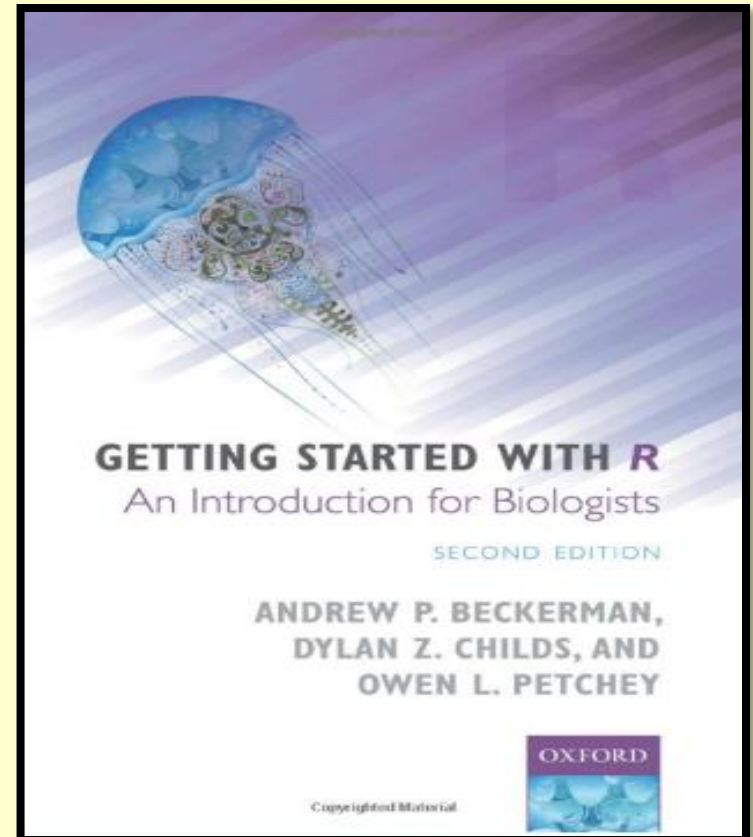
## ➤ Texts:

Getting Started with R: An Introduction for Biologists  
(2nd Edition)

by Andrew P. Beckerman,  
Dylan Z. Childs,  
Owen L. Petchey

**ISBN-13:** 978-0198787846

**ISBN-10:** 978019878784

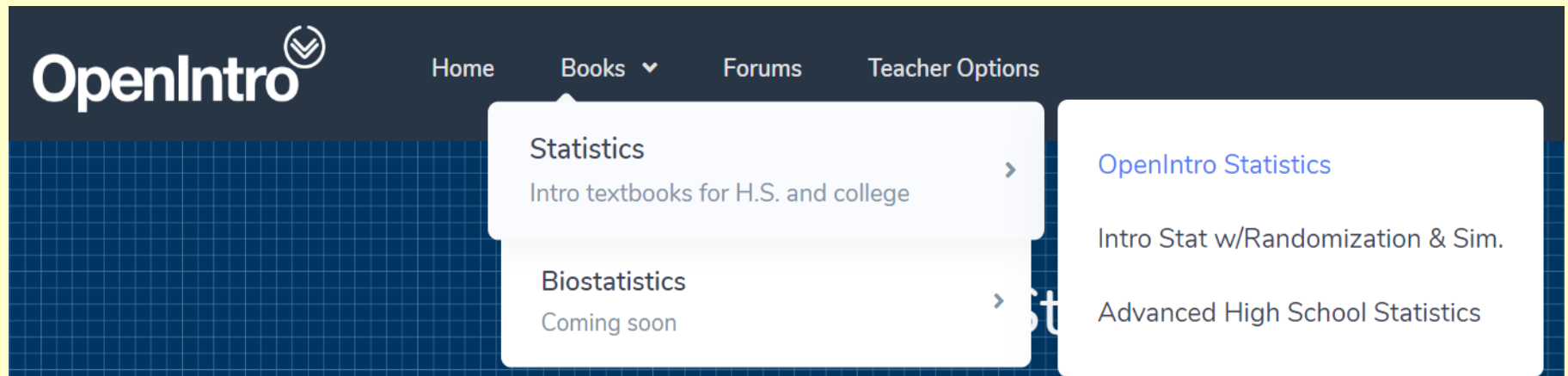


# Other Resources

## Online Text Web-Site:

You will also use a reference book, available freely online:

<https://www.openintro.org/book/os/>



## Scientific Articles: 1 - 2 each week

# Download Online Textbook and Materials

## [www.openintro.org/book/os/](http://www.openintro.org/book/os/)

### Getting Started

FREE -- OpenIntro Statistics PDF

If you want to skip the optional contribution, set the price to \$0

#### OpenIntro Statistics

Includes 1st, 2nd, 3rd, and 4th Editions



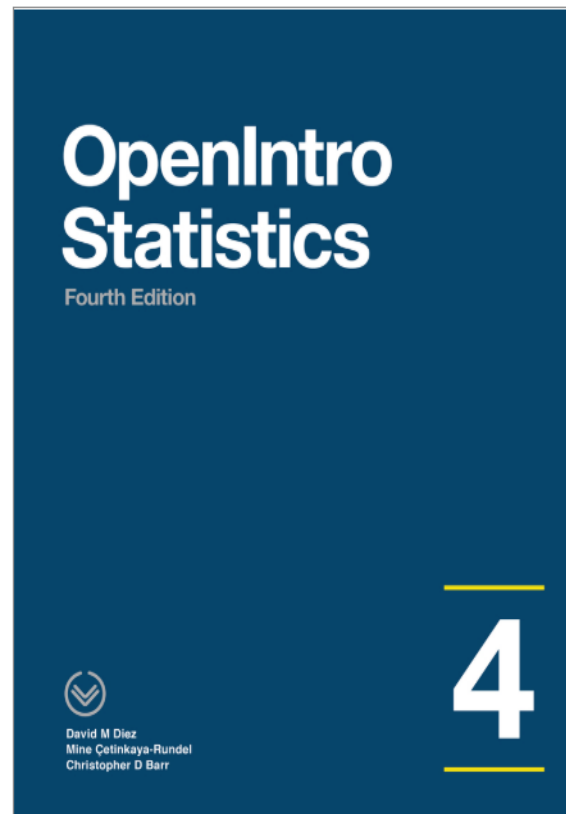
[David Diez](#), [Mine Cetinkaya-Rundel](#), and  
[Christopher Barr](#)

Complete foundation for Statistics, also serving  
as a foundation for Data Science.

Revenue supports OpenIntro (US-based  
nonprofit), e.g. we use funds to fund free desk  
copies to teachers interested in using OpenIntro  
Statistics in the classroom. To make a tax-  
deductible donation (US buyers), use  
[openintro.org/donate](http://openintro.org/donate).

More resources:...

[Read Free Sample](#)



This book is 100% complete  
COMPLETED ON 2019-11-12

You own this book! [View it in your Library](#)

Free!

\$15.00

MINIMUM PRICE

SUGGESTED PRICE

YOU PAY

\$15.00

AUTHORS EARN

\$12.00

Packages [Details](#)

☐ The Book

☒ The Book + Tablet-Friendly PDF of the Book

YOU PAY (US\$)

\$15.00

EU customers: Price excludes VAT.  
VAT is added during checkout.

[Add Ebook to Cart](#)

[Add to Wish List](#)

# Download Online Textbook and Materials

[www.openintro.org/book/os/](http://www.openintro.org/book/os/)

## Getting Started

\$20 -- B&W paperback

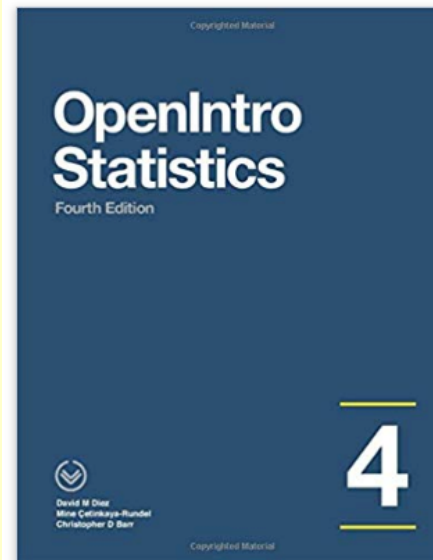
Available on Amazon and in select bookstores

### OpenIntro Statistics: Fourth Edition

by David Diez (Author), Mine Çetinkaya-Rundel (Author), Christopher Barr (Author)

★★★★☆ 13 ratings

Look inside ↓



ISBN-13: 978-1943450077

ISBN-10: 1943450072

[Why is ISBN important?](#) ↓

**Paperback**

**\$18.39**

**Other Sellers**

from \$18.39

Buy new

FREE delivery: **Tuesday**

Order within 7 hrs 20 mins [Details](#)

© MARYETA - Kailua 96734

**In Stock.**

Ships from and sold by Amazon.com.

✓prime

**More Buying Choices**

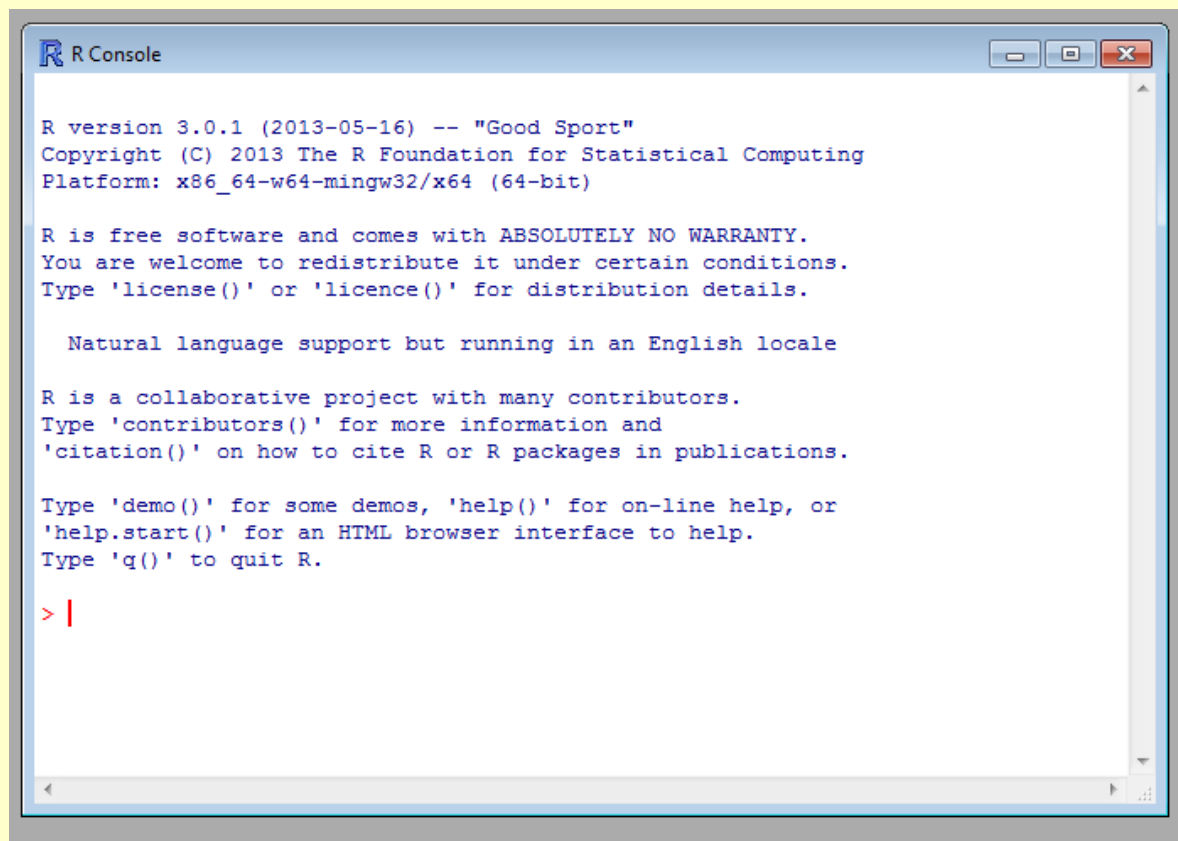
1 New from \$18.39 | 3 Used from \$19.99

# R Software

<http://www.r-project.org/>

## Getting Started:

- R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. **To download R, please choose your preferred CRAN mirror.**
- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.



```
R Console

R version 3.0.1 (2013-05-16) -- "Good Sport"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

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.

> |
```

R console allows the use of interactive Markdown script commands and the creation / loading and running of scripts.

Minimal use of menu-driven GUIs



# Other Resources

## R Studio Freeware:



Windows and Mac OS X versions of R come with simple programming editors, but I strongly recommend using the RStudio interactive development environment (IDE) for command-line use of R.

RStudio supports R Markdown documents and incorporates a powerful editor and easy-to-use file management tools.

To download, visit the RStudio web site at:  
<https://www.rstudio.com/products/rstudio/>  
for details, including extensive documentation.

# Other Resources

## R Commander Package:



R Commander is a basic graphical user interface (GUI) for R, which provides a series of menus that allow users to run many statistics and create graphics without typing code.

While more advanced features of R are not accessible through R Commander, you can use it for the majority of your statistics.

We will use R Commander to get the hang of the R language and to learn how to perform many analyses.

# Examinations

## ➤ QUIZZES:

- 9 Quizzes, Drop lowest score, Keep 8

## ➤ FINALS:

- In class
- Take-Home



# Grades

➤ Final grades will be determined as follows:

Quizzes - 8 40%

R Homework Sets - 5: 25%

In-Class Final: 10%

R Take-Home Final: 20%

Participation: 5%

(come prepared to class, seek help)

---

100 %

# Participation

Students will be assessed on their effort and commitment to learning. This includes class attendance, readiness, and - if necessary - seeking help in class or at office hours.

Students will review readings and homework keys before coming to class to become familiar with the material.

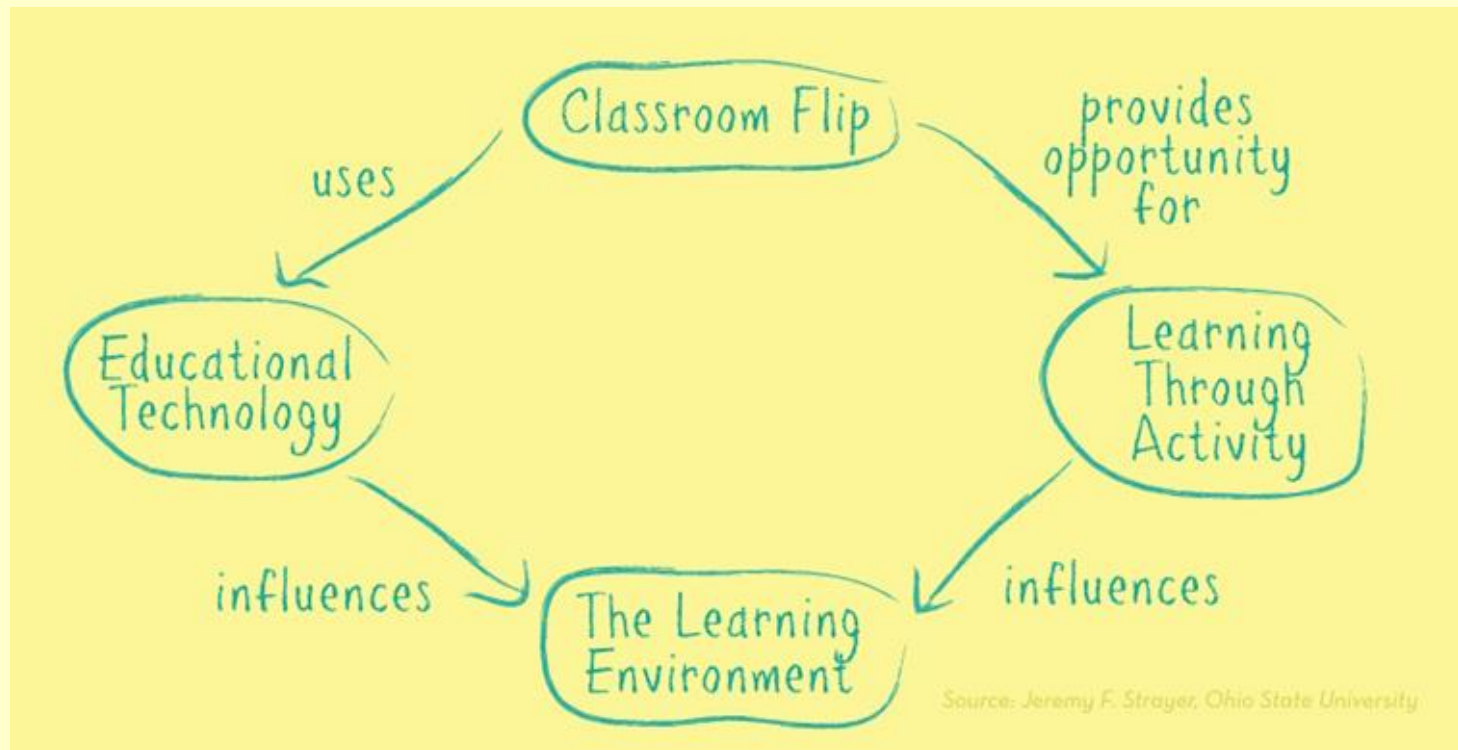
The instructor will organize group activities and will randomly select volunteers to explain concepts / problems from homework to the class.

Students will be evaluated qualitatively (+ / -).

# Flipped Teaching

Form of blended learning in which students review lectures online and work on problem sets in class.

This approach allows the instructors to spend more time interacting with students instead of lecturing.





# Class Policies

Attendance - Attendance is mandatory. There will be no make-up quizzes / 5-minute papers except in the case of documented medical necessity.

Coming to class late - Tardiness disturbs others. If you must come late or leave early discuss the need with me and try to make as small a disturbance as possible by sitting close to the door.

Cell phones are *never* allowed in class; turn them off (make them silent) before entering the room.

Laptops are allowed to take notes / view the lecture pdfs. This is a privilege which will be revoked if laptops are used for non-class activities (e.g., email / facebook).

# Academic Integrity

It is academically dishonest to try to pass off someone else's intellectual work as your own, or to help someone else to do so. Thus, there are no circumstances under which including someone else's writing or results in your papers or assignments is permissible.

Plagiarism will result in a zero on the assignment, and issuance of an academic dishonesty report to the University's Office of Academic Affairs. Serious cases of academic dishonesty will lead to an "F" in the course and may lead to expulsion from the University. Students are expected to comply with HPU's Academic Integrity

(<http://www.hpu.edu/StudentServices/AcademicIntegrity/index.html>)

# Your Tasks for This Week

Read textbook chapters and article pdfs

Download software (R and R Studio)

Install and run software (R and R Studio)



# PC Installation Instructions

Download R from <http://cran.us.r-project.org/> (click on "Download R for Windows" > "base" > "Download R 2.x.x for Windows")

Install R.

Leave all default settings in the installation options.

Download RStudio from:  
<http://rstudio.org/download/desktop> and install it.

Leave all default settings in the installation options.

Open RStudio.

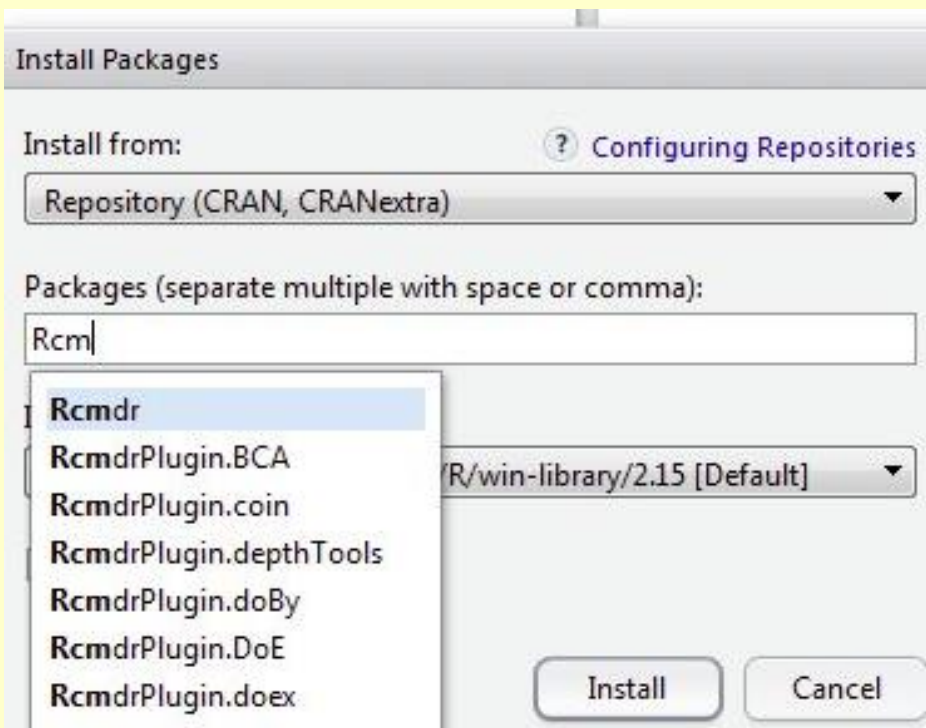
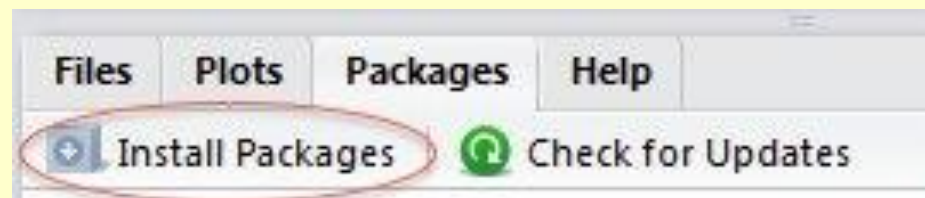


# PC Installation Instructions

Go to "Packages" tab and click on "Install Packages".

The first time you do this you will be prompted to choose a CRAN mirror. Choose the location closest to you ("USA CA 1" or "USA CA 2", which are housed at UC Berkeley and UCLA, respectively).

R will download all necessary files from the server you select.





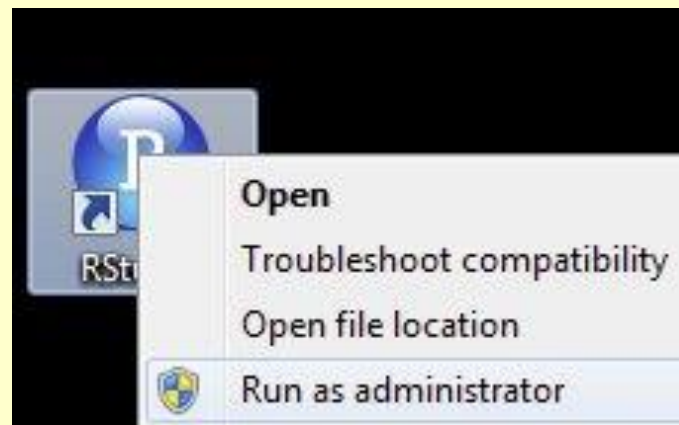
# PC Installation Instructions

Typing "Rcmdr" and the package will appear in the list.

Ensure that "Install dependencies" is checked - to install all the needed supporting packages - , and click "Install".

Wait while all the parts of the R Commander package are installed.

**NOTE:** If you get permission errors while installing packages, close R Studio and reopen it with administrator privileges.



Once you've installed R Commander, you won't have to go through all those steps again !



R

# Mac Installation Instructions

Download R from <http://cran.us.r-project.org/>  
(click on "Download R for MacOS X" > "R-2.x.x.pkg  
(select latest version)") and Install R.

Download RStudio from:  
<http://rstudio.org/download/desktop>.

Install RStudio by dragging the application icon to your  
Applications folder.

Download Tcl/Tk from:  
<http://cran.r-project.org/bin/macosx/tools/>  
(click on **tcltk-8.x.x-x11.dmg**; OS X needs this to run  
R Commander.)

Install Tcl/Tk.

# R

# Mac Installation Instructions

Go to your Applications folder and find a folder named Utilities.

Verify you have a program named "X11".



If you do not, go to:  
<http://xquartz.macosforge.org/>

and download and install the latest version of XQuartz.

R

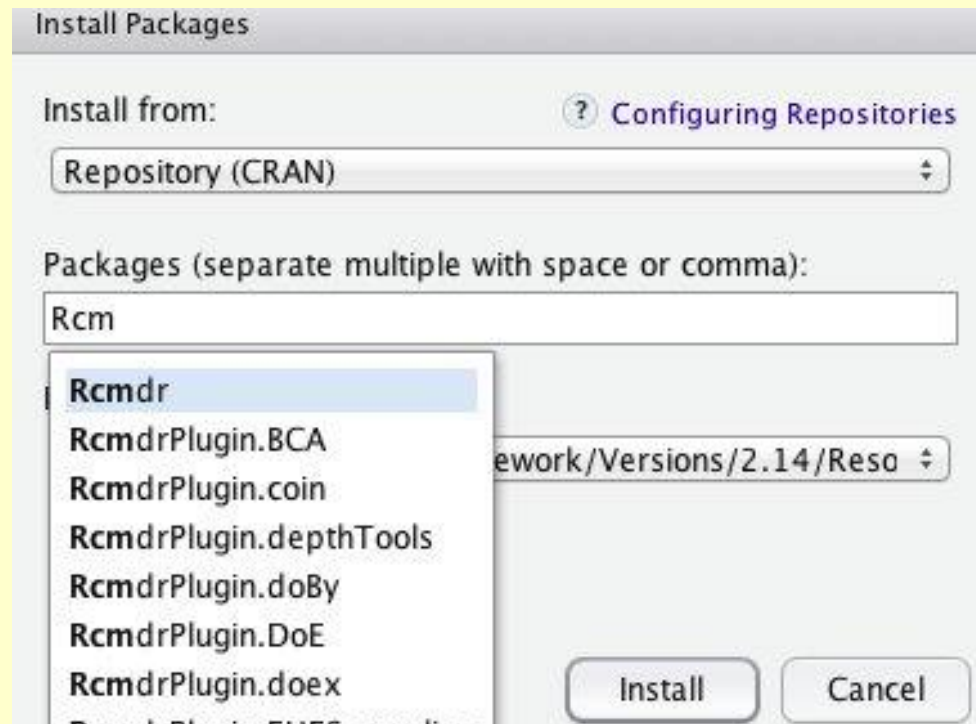
# Mac Installation Instructions

Open RStudio.

Go to "Packages" tab and click on "Install Packages".

The first time you do this you will be prompted to choose a CRAN mirror. R will download all files from the server you select.

Choose closest location ("USA CA 1" or "USA CA 2", housed at UC Berkeley and UCLA).



R

# Mac Installation Instructions

Start typing "Rcmdr" until you see it appear in a list. Select the first option (or finish typing Rcmdr), ensure that "Install dependencies" is checked, and click "Install".

Wait while all the parts of the R Commander package are installed.

Open R Commander in Windows and OS X

**NOTE:** Once you've installed R Commander, you won't have to go through all those steps again !