Syllabus

Epi/Biostats Winter Institute - Introduction to R for Public Health Researchers 140.604.73

Class Website: http://jhudatascience.org/intro_to_R_class/
CoursePlus: https://courseplus.jhu.edu/core/index.cfm/go/syl:syl.public.view/coid/16733/

Zoom link will be emailed to students.

Day/Time: Jan 10 - 21: 8:30AM-11:50AM on Zoom

Instructors: Carrie Wright (cwrigh60@jhu.edu), Ava Hoffman (ava.hoffman@jhu.edu), and Candace Savonen (csavone1@jhu.edu)

TAs: Grant Schumock (gschumo1@jhmi.edu) and Qier Meng(qmeng11@jhmi.edu)

Communication will mainly occur through Slack and we will email you about how to connect to slack.

Overview: This course will provide **"hands-on" training** for learning how to analyze data in the R statistical software package. We will cover data input/output, data management and manipulation, and how to make useful and informative graphics

Course Format: Each class will consist of 2 or 3 hour-long modules: each module features a lecture and an R programming lab, where students apply the skills taught in the modules to real data.

By the end of the course, students should be comfortable:

- Reading data into R
- Recoding and manipulating data
- Using R add-on packages
- Making exploratory plots
- Performing basic statistical tests
- Understanding basic programming syntax
- Creating reproducible R documents

Tentative Schedule:

Day Overview

Time (EST)	Content
8:30am - 9:30am	Session 1
9:30am - 9:40am	Break
9:40am - 10:40am	Session 2
10:40am - 10:50am	Break
10:50am - 11:50am	Session 3

In recognition of Martin Luther King Jr. Day, there will be no class on Monday January 17th 2022. There is an assignment due before class to install software and we recommend completing HW 1 (uploading a screenshot showing that you finished the Dataquest module indicated below) before class, but all other assignments will be due January 26th. You are welcome to turn assignments in earlier if you wish.

Day	Module	Slides	Code	Resource	Cheatsheets/Guides
Day 0	Homework 1		Dataquest		
Day 1	Intro	HTML, PDF	R, Rmd		Day 1 Cheatsheet
	RStudio	HTML, PDF	R, Rmd	Lab, Key, Key HTML	
	Reproducibility	Slides, HTML, PDF	R, Rmd		Good scientific coding practices
Day 2	Basic R	HTML, PDF	R, Rmd	Lab, Key, Key HTML	Day 2 Cheatsheet
	Data IO	HTML, PDF	R, Rmd	Lab, Key, Key HTML	Debugging tips guide
Day 3	Subsetting Data in R	HTML, PDF	R, Rmd	Lab, Key, Key HTML	Day 3 Cheatsheet
	Homework 2		Rmd, HTML, Key, Key HTML		
Day 4	Data Summarization	HTML, PDF	R, Rmd	Lab, Key, Key HTML	Day 4 Cheatsheet
	Data Classes	HTML, PDF	R, Rmd	Lab, Key, Key HTML	
Day 5	Data Cleaning	HTML,PDF	R,Rmd	Lab, Key, Key HTML	Day 5 Cheatsheet
	Homework 3		Rmd, HTML, Key, Key HTML		
Day 6	Manipulating Data in R	HTML, PDF	R, Rmd	Lab, Key, Key HTML	Day 6 Cheatsheet
	Intro to Data Visualization	HTML, PDF	R, Rmd	Lab, Key, Key HTML	
Day 7	Data Visualization	HTML, PDF	R, Rmd	Lab, Key, Key HTML	Day 7 Cheatsheet
	Factors	HTML, PDF	R, Rmd	Lab, Key, Key HTML	
Day 8	Statistics	HTML, PDF	R, Rmd	Lab, Key, Key HTML	Day 8 Cheatsheet
	Project Guidelines	HTML, Rmd	Example RMD, Example HTML	Instructions	
Day 9	Functions	HTML, PDF	R, Rmd	Lab, Key, Key HTML	Day 9 Cheatsheet

Day	Module	Slides	Code	Resource	Cheatsheets/Guides
	Project Guidelines	HTML, Rmd	Example RMD, Example HTML	Instructions	

Grading

1. Attendance/Participation: 20% (Please let the instructors know if attendance will be difficult for you.)
2. Homework: 3 x 15%
3. Final "Project": 35%

All assignments are due Wednesday, Jan 26, 2022 at 11:59pm EST.

Note: only people taking the course for credit must turn in the assignments. However, we will evaluate all submitted assignments in case others would like feedback on their work.

Submitting Assignments

Submit each assignment to the designated Drop Box on CoursePlus.

You should complete the following:

1. Data Quest: Introduction to Programming in R
2. Homework 2 Problem Set:
   - Questions (Rmd), Questions (HTML)
   - Key (Rmd), Key (HTML)
3. Homework 3 Problem Set:
   - Questions (Rmd), Questions (HTML)
   - Key (Rmd), Key (HTML)
4. Final Project (see below)

Final Projects

This project should entail:

- reading in a dataset of your choosing
- doing some light data cleaning
- performing some data summarization
- creating a couple of visualizations
- doing some very light statistical analysis, like regression or t-tests

You may use one of these public datasets, a different public dataset, or your own data for the project. If using your own dataset, be sure any sensitive information is protected.

- https://www.data.gov/
- https://data.baltimorecity.gov/
- https://github.com/awesomedata/awesome-public-datasets

Example projects can be found with the source code: Rmd, and the output html here.

See the guidelines/instructions for final projects.

Code of Conduct

We would like to create an open, safe, welcoming, diverse, inclusive, intellectually stimulating, and hopefully fun class experience.

We strive to be a space in which individual differences are respected, so that each individual can reach their fullest potential.

Guidelines

- Demonstrating empathy and kindness toward other people
- Giving and gracefully accepting constructive feedback
- Avoiding making assumptions about others
- Being respectful of differing opinions, viewpoints, and experiences
- Take the time to consider how your speech and actions may influence others
- Actively seek to acknowledge and respect the boundaries of other members
- Demonstrating academic integrity
  - Students can work together, but assignments must show unique contributions from the student who turns in the assignment
  - Students can use online resources for help, but assignments must show unique contributions from the student who turns in the assignment - plagiarism is not allowed
  - If students are struggling, reach out to instructors or TAs as early as possible

This applies to emails, surveys, Slack, Zoom, office hours, meetings with other students, instructors, or TAs.

Please reach out to a TA or instructor if you witness or experience a violation of the class guidelines or other JHU codes of conduct.

- **JHU Student Code of Conduct:**
  https://studentaffairs.jhu.edu/policies-guidelines/student-code/

- **Hopkins School of Medicine:**
  https://www.hopkinsmedicine.org/research/resources/offices-policies/OPC/Research_Integrity/som_code_of_conduct_04302020.pdf

- **JHSPH Academic Ethics Code:**
  https://www.jhsph.edu/offices-and-services/student-affairs/resources/student-policies/_documents/academic-ethics-code.pdf

The University has developed avenues for reporting and for seeking help including:

JHU Sexual Assault Helpline, 410-516-7333 (confidential) Campus Safety and Security, 410-516-7777 University Sexual Assault Response and Prevention website Johns Hopkins Compliance Hotline, 844-SPEAK2US (844-733-2528) JHU Office of Institutional Equity 410-516-8075 (nonconfidential) Johns Hopkins Student Assistance Program (JHSAP), 443-287-7000 University Health Services - Mental Health (UHS-MS), 410-955-1892 The Faculty and Staff Assistance Program (FASAP), 443-997-7000

Module Details:

Day 0 (before the first session)

- Homework 1

Day 1

- Introduction
- RStudio
- Reproducible Research

Day 2

- Basic R: Variables/Objects in R
- Data Input/Output

Day 3

- Subsetting Data
- Discuss Homework 2 - work on Homework 2 in class if there is time

Day 4

- Summarization
- Data Classes

Day 5

- Data Cleaning
- Discuss Homework 3 - work on Homework 3 in class if there is time

Day 6

- Data Manipulation
- Data Visualization with Esquisse

Day 7

- Data Visualization Continued
- Factors

Day 8

- Statistics
- Discuss and work on Final Project

Day 9

- Functions
- Work on Final Project