# Syllabus

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

Class Website: http://jhudatascience.org/intro_to_r/
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 student 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 |

### Day 1

- Introduction
- RStudio
- Reproducible Research

### Day 2

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

### Day 3

- Subsetting Data
- Homework 2

### Day 4

- Summarization
- Data Classes

### Day 5

- Data Cleaning

### Day 6

- Data Manipulation
- Homework 3

### Day 7

- Data Visualization

### Day 8

- Statistics
- Work on Final Project

### Day 9

- Functions
- Good code practices
- Work on Final Project

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 |
|---|---|---|---|---|
| Day 0 | Homework 1 | | Dataquest | |
| Day 1 | Intro | HTML, PDF | R, Rmd | |
| | RStudio | HTML, PDF | R, Rmd | Lab, Key, Key HTML |
| | Reproduciblility | HTML, PDF | R, Rmd | |
| Day 2 | Basic R | HTML, PDF | R, Rmd | Lab, Key, Key HTML |
| | Data IO | HTML, PDF | R, Rmd | Lab, Key, Key HTML |

| Day | Module | Slides | Code | Resource |
|---|---|---|---|---|
| Day 3 | Subsetting Data in R | HTML, PDF | R, Rmd | Lab, Key, Key HTML |
| | Homework 2 | | Rmd, HTML, Key, Key HTML | |
| Day 4 | Data Summarization | HTML, PDF | R, Rmd | Lab, Key, Key HTML |
| | Data Classes | HTML, PDF | R, Rmd | Lab, Key, Key HTML |
| Day 5 | Data Cleaning | HTML,PDF | R,Rmd | Lab, Key, Key HTML |
| Day 6 | Manipulating Data in R | HTML, PDF | R, Rmd | Lab, Key, Key HTML |
| | Homework 3 | | Rmd, HTML, Key, Key HTML | |
| Day 7 | Data Visualization | HTML, PDF | R, Rmd | Lab, Key, Key HTML |
| Day 8 | Statistics | HTML, PDF | R, Rmd | Lab, Key, Key HTML |
| | Project Guidelines | HTML, Rmd | Example RMD, Example HTML | Instructions |
| Day 9 | Functions | HTML, PDF | R, Rmd | |
| | 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 and visualization, and
- ☐ doing some very light statistical analysis, like regression or chi-squared tests.

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

See the guidelines/instructions for final projects.