M21-506: R Primer (Summer 2017)
General Course Information

Revised: 6/25/2017

Instructor	Yun Ju Sung, yunju@wustl.edu
	RJ Waken, rjwaken@wustl.edu
Teaching Assistant	Yuxuan (Daisy) Jin, yuxuanjin@wustl.edu
Dates	July 5 through July 12
Times	Morning (9 am – 12 noon) and Afternoon (1:30 pm – 4 pm)
Location	Biostatistics Computer Lab, 5th Floor of Becker Medical Library

Grading

Assignments (due 11:59 pm before next class)	50%
Quiz	10%
Midterm Exam (1 pm – 4 pm on July 7)	20%
Final Exam (1 pm – 4 pm on July 12)	20%
Final Grade (+/- letter grades)	

Course Description

R is a free software environment for statistical computing and graphics (http://www.r-project.org) and offers rich statistical and graphical tools to handle large data sets. The 2-credit course offers intensive hands-on summer training in R over 6 full weekdays. The goal is to provide students with an opportunity to gain skills in data analysis and graphics using R. It is designed for students who are new to R but have had some basic experience working with computers. This course meets the prerequisite for two summer courses M21-515 Fundamentals of Genetic Epidemiology and M21-550 Introduction to Bioinformatics.

Learning Objectives

Students who completes this course will be able to

1. Understand basic data types in R
2. Perform basic exploratory data analysis in R
3. Create high quality graphics using R
4. Perform hypothesis testing and regression analysis

Before the 1st Day of Class

1. Complete pre-survey: don't spend more than 30 minutes. I would like to know your prior experience with R.

2. Complete the 7 chapters from tryR (http://tryr.codeschool.com/). This will provide online tutorials in the R basics.

3. Refresh your memory on statistics and hypothesis testing. Use your previous Statistics textbook if you have. You can also choose one of two sources below.

    OpenIntro Statistics
    - Book: https://www.openintro.org/stat/textbook.php?stat_book=os

- Videos https://www.youtube.com/channel/UCPAHrheP8gcpml0tSVhKOWg

Khan Academy
  - Hypothesis Testing: https://www.khanacademy.org/math/statistics-probability/significance-tests-one-sample
  - Other topics: https://www.khanacademy.org/math/statistics-probability?t=classes

4. Choose one R book that you like the best. There are many excellent books, as shown in http://www.introductoryr.co.uk/R_Resources_for_Beginners.html. I like the following 4 books:
   a. *R for Beginners* by Emmanuel Paradis
   b. *Using R for Introductory Statistics* by John Verzani
   c. *The R Book* by Michael J. Crawley (2nd Edition)
   d. *Data Analysis and Graphics Using R: An Example-Based Approach* by John Maindonald and W. Hohn Braun

   All these books are freely available (in a pdf format) either from the webpage above or simply searching through google. Let me know if you need help choosing a book.

5. The following YouTube videos may be useful:
   - **R Tutorial** series by tutorial (https://www.youtube.com/watch?v=ZoPJGmpYJzw&list=PL69A9CCD816A5F3A5&index=1)
   - **Statistics with R** series by Christoph Scherber (https://www.youtube.com/watch?v=Xh6Rex3ARjc)
   - **Statistics with R** series by Courtney Brown (https://www.youtube.com/watch?v=2-kw1MlOS1U)

Course Syllabus

Day	Morning (9 am to 12 noon)		Afternoon (1:30 pm to 4 pm)	
	Lecture/Lab	Instructor	Lecture/Lab	Instructor
1 7/5/W	• Introduction to R • R Studio • Functions and arguments	Sung Waken	• Data in R • Data Input and Output	Waken
2 7/6/Th	• R Objects • Data Frame and List • Operators and Indexing	Waken	• Univariate Data • Programming with R (I) • Flow Control	Sung
3 7/7/F	• Graphics • Exploratory Data	Waken	**Midterm Exam** **1 pm – 4 pm**	
4 7/10/M	• Statistical Inference • Hypothesis Testing	Sung	• Bivariate Data • Regression Analysis	Sung
5 7/11/T	• Programming with R (II)	Waken	• Writing Functions	Sung
6 7/12/W	• Real examples using R • Advanced topics	Waken	**Final Exam** **1 pm – 4 pm**	

Note that the schedule and topics are subject to change.

More on Grading

Assignments: The goal of assignments is to develop and demonstrate mastery of R concepts and tools. They will start out relatively simple at the beginning and become more challenging as the course progresses. You will begin the assignments in class and continue working on them outside of class if necessary.

Late work: To be fair to all students and to the instructor who is responsible for grading, **no late work** will be accepted after due date. To avoid losing data and to help you develop a sound workflow for data analysis and management, you are required to keep copies of all work, save your work **frequently**, and store back-up copies in **multiple locations**. We will not consider technology excuses such as lost, missing or stolen data.

Quiz: Quiz will be close-note, close-book, and close-R. It will be based on lecture and assignments.

Exams: Exams will be open-note, open-book, and open-R. It will test the skills you have learned during the course.

In-class participation and attendance: Lecture and lab will function together. Attendance is required. Because this is a hand-on course, true make-up sessions are often not possible. We expect you to attend and **actively** participate in all class meetings by voicing your opinions and asking questions. Students who must miss a class due to ill health, a death in the family, or out-of-town trip should inform Dr. Sung prior to the class session. If you miss a class, you are responsible for obtaining notes and information from the instructor; consulting with the instructor and/or a TA as necessary to gain an understanding of the material covered; and catching up on your work as needed.

Installing and Using R on Your Personal Computer

We recommend that you install R on your personal computer for use outside of class. R can be downloaded for free from: http://r-project.org/. You may use your computer or one of the lab computers during the R computer labs. For help with installing R, see the video tutorials on "How to Install R for Mac" and "How to Install R for Windows" on YouTube. We require that all students use the same default user interface that is included as part of the R installation. For this year, we plan to use RStudio as it contains more sophisticated features that can help with using R.

Laptops and Cell Phones:

You may bring laptops and phones to class but the sound must be muted. You may use your laptop for class activities such as computer assignments and note taking, but not for unrelated activities such as email, Facebook, or web browsing, as these activities are very distracting to the other students and the instructors. You may not use your phone or text message during class.

Academic Integrity

The academic integrity policy of Washington University in St. Louis states: "Effective learning, teaching and research all depend upon the ability of members of the academic community to trust one another and to trust the integrity of work that is submitted for academic credit or conducted in the wider arena of scholarly research. Such an atmosphere of mutual trust fosters the free exchange of ideas and enables all members of the community to achieve their highest potential. In all academic work, the ideas and contributions of others must be appropriately acknowledged, and work that is presented as original must be, in fact, original. Faculty, students, and administrative staff all share the responsibility of ensuring the honesty and fairness of the intellectual environment at Washington University."