SPAN 497 What R We Doing?

An Introduction to R for Data Processing and Management

Summer 2021

Monday, Wednesday, & Friday 9-11am

Instructor:	Dr. Katrina Connell	**Schedule:**	MFW 9-11am
Zoom Link:	Click Here	**Office Hours:**	Email Me
Email:	kzc501@psu.edu		
Course Site:	Canvas.psu.edu	**SharePoint Site:**	Click Here

Course Description and Objectives

This course examines the basics of R and its practical uses for data processing and graphing. In real analyses, our data comes to us from a variety of sources, in varied formats and rarely in the proper format to conduct statistical analyses. Through class and exercises, you will be prepared to work with real-world data by giving you the theoretical understanding of and practical skills for managing large datasets taking advantage of powerful tools built into R for data handling and processing. The course will begin by introducing R, R studio, the R Markdown file, and R Projects as tools for conducting organized and reproducible analyses of language related data. After covering the basics of how R works, this course will cover topics like good data storage and analysis practices, variable types, data transformations, relational data, various variable coding methods (e.g., contrast coding, dummy coding, etc.), centering variables, and data visualization using primarily the Tidyverse package in R. The focus of this course is on the R software and data processing, not statistical analysis. It is ideal for those who are R beginners or current R users who would like to have a more solid foundation in R and data science basics before moving on to instruction in statistics.

Required Materials

1. Main textbook - PDF available here

   Wickham, H., & Grolemund, G. (2016). *R for data science: import, tidy, transform, visualize, and model data*. " O'Reilly Media, Inc.".

2. Additional readings (provided in OneDrive)

Course Requirements

Software Skills: You will be required to download & learn the basics of **Microsoft Excel** (data organization software) and ***R/RStudio*** (data organization and analysis software), which are used to manage and analyze experimental data. The course will include instruction on each of these programs.

Assignments (70%)
Assignments for the class will be due almost daily in the class. This class is a skills class, and you must practice these skills to learn them and retain them. **Assignments will be at noon the day they are due**. This will allow you to ask questions in class and modify your assignments before turning them in. These assignments will mainly take the form of markdown files that you will complete and turn in, and may occasionally take other forms such as Excel spreadsheets or activities asking you to generate your own markdown. These assignments will be found in OneDrive and submitted via Canvas.

Quizzes (10%)
Occasionally throughout the class there will be quizzes in Canvas to serve as an understanding check for content knowledge. **Quizzes will be due before class the day they are due.** These will be posted with a due date in Canvas and are listed on the syllabus.

Final Analysis (20%)
Your final analysis will be a culmination of all of the skills and statistical knowledge you have acquired this semester. You will be provided with a description of a study and the dataset from that study. You will then use the code we have used for R over the course of the semester to conduct an analysis based on the requirements listed in the assignment sheet. You will save your analysis as a markdown file and upload the analysis. Additionally, based on the guidelines in the handout, you will write up a short summary of your analysis and what you found, interpreting the results of each of your analyses. You may work on this assignment with others in the class, but your work should be entirely your own. **You may NOT send lines of code to each other.** Please keep in mind that helping too much on an assignment like this actually HURTS that person in that they don't learn it for themselves. This assignment will be found in OneDrive and submitted via Canvas.

Late assignment policy: Late assignments will lose 10% of the points earned for every day that it is late. If you will be missing a day of class, you must tell the instructor ahead of time and turn in any assignments for the day in advance, except in cases of emergencies. **This class will be recorded automatically in Zoom, so if you miss a class, you will be able to catch up.** You can access these recordings in Canvas in the Zoom tab under Cloud Recordings.

Academic Misconduct

Academic misconduct will not be tolerated. Strict procedures for reporting plagiarism in written assignments or cheating during an in-class exam will be enforced. If either is discovered, you will receive a zero for that portion of your final grade and the incident will be reported to University authorities. If such academic misconduct occurs a second time, you will receive an F in the course and the incident will again be reported to University authorities.

Grading

A 95.0-100%	A- 90.0-94.9%	
B+ 87.0-89.9%	B 83.3-86.9%	B- 80.0-83.2%
C+ 75.0-79.9%	C 70.0-74.9%	D 60.0-69.9%
	59.9%-0 F	

Class Schedule

Subject to change

Week	Day	Date	Topics		Homework Due	Readings Due
1	1	10-May	Intro	rStudio & Scripts vs rMarkdown	Beginning R Complete Before Class	
	2	12-May	Data Visualization with ggplot2		Upload markdown from Day 1 Rmarkdown Assignment	Ch. 1 2, 4, 21, 24 Data Camp Videos on Canvas
	3	14-May	Data Visualization	Variable Types	Start Visualize Assignment	The Layered Grammar of Graphics.pdf. Data Camp Logical Operators Video on Canvas
2	4	17-May	Using Excel (for the first & last time)	Data Transformations with dplyr	Visualize Assignment, Variable Types 1 Canvas Quiz	Data Wrangling Videos on Canvas
	5	19-May	Data Transformations with dplyr		Variable Types 2 Canvas Quiz Excel Practice	Data Wrangling Videos on Canvas Ch. 3
	6	21-May	Data Transformations with dplyr		Data Transformations Assignment Data Transformations Canvas Quiz	Data Camp Pipe Reading first 1/2
3	7	24-May	Data Frames	Tibbles	Pipe Assignment CheatSheet Assignment Part 1	Ch. 6,7, Rprojects.pdf Data Camp Data Frames Videos on Canvas
	8	26-May	R Projects	Importing Data	Data Transformations Assignment	Ch. 8 Data Camp Import Data Video on Canvas
	9	28-May	Tidy Data with tidyr		Data Frames Canvas Quiz, R Project Assignment	Ch. 9 Data Camp Tidy Data Video on Canvas
4	10	31-May	Tidy Data with tidyr		Tidy Data Canvas Quiz	TBA
	11	2-Jun	Relational Data with dplyr		Tidy Data Assignment	Ch. 10
	12	4-Jun	ifelse Statements	Factors with Forcats	Relational Data Assignment	Ch. 12. Wrangling Cat Data.pdf Data Camp Forcats Video

Week	Day	Date	Topics		Homework Due	Readings Due
5	13	7-Jun	Analysis Introduction	TBD	CheatSheet Assignment Part 2 Strings Assignment	Ch. 11
	14	9-Jun	Graphing to Communicate Results		Factors Assignment	Ch. 22
	15	11-Jun	Introduction to Simple/Multiple Linear Regression		Regression Canvas Quiz	Ch. 5, 18
6	16	14-Jun	Centering Variables	Coding Variables	Analysis Part 1	Review Ch. 12
	17	16-Jun	Interpreting Regression Results			Barr, 2013
	18	18-Jun	Interpreting Regression Results	Project Day	Analysis Part 2	TBA
	June 25th		Final Analysis Due			