UNIVERSITY OF OREGON

Exploring Data with R

EDLD 610 – 4 Credits – CRN 37117

University of Oregon, Educational Methodology, Policy, and Leadership (EMPL)

Spring 2017 Term Syllabus

Meeting Days/Times: TR 10:00 – 11:50
Location: Hedco 146

1. General

INSTRUCTOR: .....Daniel Anderson, PhD, IES Post-Doctoral Fellow, Center on Teaching and Learning
PHONE .....(541) 346-3317
E-MAIL .....daniela@uoregon.edu (preferred contact)
OFFICE HOURS ....TBA during first class session, following discussion with students

2. Course Overview

R is a freely available and open-source computing environment for statistical programming and data visualization that has tremendous advantages for research and dissemination of empirical findings. As an open-source program, the code for all functions is available. This means the user does not have to blindly trust that the software is working as intended. Rather, users can look “beneath the hood” a bit, if they so choose, to see exactly how the analysis is being conducted. Countless packages have been developed for R (over 10,000 are now on CRAN), which are all also freely available, helping make complex analyses within specific contexts more feasible. Further, R lends itself to reproducible research – meaning that others can view all the procedures taken in any analysis and reproduce the results exactly. Reproducible research facilitates transparency and development of research communities, with interested users sharing analytic procedures and outcomes. Readers of reports also do not have to guess at any intermediary steps taken during the analysis, if the source code is available. Finally, R is immensely flexible. Simon Blomberg, an evolutionary biologist and R programmer, once replied to an R user who was asking "if" a specific process was possible with R by stating that "This is R. There is no if. Only how.”

The overall purpose of this course is to provide students with a basic foundation in R. We will focus primarily on three components: data processing (munging/manipulating/wrangling), data visualization, and reproducible workflows. This course is oriented around the philosophy of *tidy data* and, as such, primarily relies on tools within the *tidyverse* for manipulating and visualizing data. The *tidyverse* is a suite of packages developed by RStudio, generally led by Hadley Wickham, which are all optimized for tidy data. The focus of this course is on working with R, as opposed to the specifics of any given analysis. Students should, therefore, expect to acquire a basic understanding of the foundational practices of exploring data within the R framework. As such, students are expected to have a basic understanding of statistics (e.g., regression) prior to taking the course, given that statistical models will be used for illustrative examples throughout, but high-level statistical knowledge (e.g., multilevel modeling or structural equation modeling) is not a prerequisite.

Moving to R from other statistical software generally requires a fundamental shift in the way we think about and interact with data. Aside from this shift in thinking, there is also a substantial amount of code to learn, which can be both frustrating and intimidating. The primary goal for this course is to make this shift less intimidating and the learning curve less steep. When people learn to read, they generally start by learning their alphabet, then learning

sounds and blends before moving to decoding. As students begin to gain fluency (with many sight words) they no longer have to focus on decoding and can begin to comprehend the text at deeper levels – eventually shifting from learning to read to reading to learn. An essentially equivalent process takes place when learning R (or learning to program in general), and the "decoding" phase can be slow and frustrating. But once you start to build fluency you will find that the possibilities are endless, and you will never want to go back to point-and-click interfaces.

Finally, this class is predicated on the philosophy that the only way to truly learn R and become proficient with it is to dive in and practice, practice, practice. You cannot build fluency by just watching. Until you reach at least a basic level of fluency, R will mostly be frustrating. As such, this class will include built-in labs every Thursday that will vary in length (i.e., from 45 minutes to the full class period). Lectures will also be interactive, with the instructor asking students to follow along at times. Weekly homework will also be assigned, which is intended to be relatively brief but get you more practice to help build fluency. All homework assignments can be completed independently or in small groups, and are scored on a completion basis. Feedback will be provided on your code, including ideas to make it better (e.g., clearer, more efficient). All homework must be reproducible (Note: The concept of reproducible research will be discussed fully). The course also includes a term project that requires students to complete the three main components of the course described above with a "real" dataset. This can, again, be completed independently or in small groups. Students are encouraged to find data that they are actually interested in analyzing and using as part of their program of research. Students who do not have access to data should contact the instructor as soon as possible.

3. Student Learning Outcomes

Upon completion of this course, the successful student will:

1. Understand the foundations of the R environment, including different data structures (vectors, matrices, lists, data frames) and data types (integer, double, character, logical).
2. Be able to read various data into R from various sources (e.g., .csv, .sav, .txt, etc.)
3. Be able to produce dynamic and reproducible documents with R Markdown in both pdf and html format.
4. Understand the principles of *tidy data*, and when it is and is not useful to have your data in a tidy format.
5. Understand the tools for manipulating data into a tidy format and be able to apply these tools to reshape relatively complex datasets into a tidy format.
6. Understand and be able to apply the grammar of graphics, as implemented through the *ggplot2* package, to tidy data for both exploratory and model-based plotting (exploratory plotting emphasized).

Role of discussion sections, laboratories, or other specialized learning opportunities in meeting these educational objectives:

Throughout the course, the second weekly scheduled class (i.e., Thursday sessions) will include a dedicated lab. These will generally include the instructor providing a "challenge" for students to complete. The instructor will be present for all lab sessions and provide guided feedback to students as they work through the challenges. The purpose of the labs is to provide students with more opportunity to practice and gain fluency, while being supported by the instructor and having questions addressed.

4. Textbooks and Reading Materials

We will use a variety of resources for the course, including help websites likes http://stackoverflow.com. Additional articles and relevant readings may also be assigned throughout the term.

Required Course Books (no purchases required)

Wickham, H., and Grolemund, G. (2017). *R for Data Science*. Sebastopol, CA: O'Reilly. Freely available online at http://r4ds.had.co.nz

Wickham, H. (2016). *ggplot2.* New York, NY: Springer (*select chapters provided by the instructor*)

Other readings (in order of appearance)

R-Markdown

- Lesson 1: http://rmarkdown.rstudio.com/lesson-1.html
- Lesson 2: http://rmarkdown.rstudio.com/lesson-2.html
- Lesson 3: http://rmarkdown.rstudio.com/lesson-3.html
- Lesson 4: http://rmarkdown.rstudio.com/lesson-4.html

Broman – *Knitr in a knutshell*: http://kbroman.org/knitr_knutshell/

Wickham, H. (2014). Tidy Data. *Journal of Statistical Software. 59*(10), 1-23.

Grolemund – Tidy data: http://garrettgman.github.io/tidying/

Bryan – Regular expressions: http://stat545.com/block022_regular-expression.html

Other resources (possibly referred to but not used explicitly)

- Kabacoff, R, I. (2011). *R in Action: Data analysis and graphics with R*. Shelter Island, NY: Manning Publications.
- Peng, R. D. (2015). *R Programming for Data Science*. Victoria, BC: Lean Publishing.
- Wickham, H. (2015). *Advanced R*. Available online at http://adv-r.had.co.nz
- Wickham, H. (2016). *ggplot2 (*Second Edition*)*. New York, NY: Springer.
- Xie, Y (2015). *Dynamic Documents with R and knitr* (Second Edition). Boca Raton, FL: CRC Press

5. Weekly Schedule of Topics and Assignments

Day	Theme	Readings (before class)	Homework	
			Assigned	Due
4/4	Intro to R: Why you should be excited	Kabacoff: Chapter 1 r4ds: Chapters 1 & 2 http://adv-r.had.co.nz/Style.html	Final Proj	
4/6	Very basics of R ***LAB:** Intro to R Markdown*	R Markdown, Lessons 1-4 r4ds: Chapters 4 & 30	HW 1	
4/11	More on R Markdown, plus introduction to data manipulations w/dplyr	Broman r4ds: Chapter 5		
4/13	Importing data ***LAB:** import data & manipulation*	r4ds: Chapter 11	HW 2	HW 1
4/18	Intro to data visualization w/*ggplot2*	ggplot2 book: Chapter 1 r4ds: Chapter 3		
4/20	Best practices in data visualization ***LAB:** ggplot*	ggplot2 book: Chapter 3	HW 2	HW 3
4/25	Intro to tidy data	Wickham (2014) r4ds: Chapter 12		
4/27	More on tidy data ***LAB:** tidying basic data*	r4ds: Chapters 9-11	HW 3	HW 4
5/2	Linking tidy data & data visualization	r4ds: Chapter 7		Proj Outline
5/4	Tidy data summaries/transformations ***LAB:** plotting summaries*	Grolemund: Tidy data	HW 4	HW 5

Week	Theme	Readings (before class)	Homework	
			Assigned	Due
5/9	Midterm review			
5/11	Where to next? ***LAB:*** *Putting it all together*		HW 5	HW 6
5/16	Working with strings	r4ds: Chapter 14		
5/18	Brief intro: Regular Expressions **LAB**: Working with strings	Bryan	HW 6	HW 7
5/23	Introduction to lists and the *purrr* package	r4ds: Chapters 20		
5/25	***LAB:*** *Fitting many models/plots/etc.*	r4ds: Chapter 25	HW 7	HW 8
5/30	Factors and Dates	r4ds: Chapters 15-16		
6/1	**Final Project Presentations**		HW 8	Pres
6/6	Advanced R Markdown	r4ds: Chapters 27, 29, & 30		
6/8	**Final Project Presentations**			Pres
6/12-16		**FINAL PROJECT DUE**		

6. Grading Components and Criteria

Student performance will be determined by a weighted combination of homework (scored on a completion), participation in lab activities, and a final project including a brief in-class presentation of the project. Each of these is described below.

Homework (10 points each, 80 points total)
A total of 8 homework assignments are scheduled throughout the course – corresponding to roughly weekly homework. These are intended to provide students with more practice and opportunities for feedback from the instructor. Below is an outline of the eight homework assignments.

1. *Creating an R Markdown Document*
    a. All works produced throughout the term, including all homework and the final project, will be required to be produced as a dynamic, fully reproducible document. As such, the first homework provides students an opportunity to produce a very basic R Markdown document. Students will be asked to email the instructor both pdf and html versions of the homework.

2. *Manipulating data*
    a. This homework will require students to read at least two different types of datasets into R and conduct basic manipulations (e.g., conditional selection of rows, remove specific columns, merge datasets)

3. *Data visualization*
    a. Data visualizations will be provided and students will be asked to reproduce them.

4. Tidy data
    a. Students will tidy a dataset, as well as discuss why example datasets are or are not tidy (and in the latter case, explain what would need to happen to make the data tidy).

5. Linking tidy data and *ggplot*
    a. Students will tidy a messy dataset and replicate example data visualizations. Importantly, the example data visualizations are either not possible or very difficult to reproduce without the data first being in a tidy format.

6. A full example
    a. Students will (a) load a dataset, (b) tidy it, (c) produce basic summary statistics, and (d) produce data visualizations of both the raw data, as well as summary data. All of this (as with all assignments) must be done through a reproducible workflow with R Markdown.
7. Strings
    a. Students will tidy a messy dataset, but the tidying will require manipulation of string variables. Students will also produce specific string manipulations.
8. Dates
    a. Students will parse multiple types of date data into useable formats, including moving from strings to dates. Students will also concatenate multiple variables into a single date variable.

Final Project (100 points)
The final project in this class requires students to use a "real world" dataset. This project will be discussed the first week of class. Ideally, students would work with a dataset that includes variables they are interested in using beyond just this class; however, if students do not have access to data, the instructor will provide one. Students who do not have access to data should plan to meet with the instructor as soon as possible to discuss their research interests so a dataset can be provided that – as closely as possible – aligns with his or her interests. Students are allowed to work individually or in groups of up to three people. While students are encouraged to work together, so they can bounce ideas off each other when they get stuck, there should be clear evidence of the contribution of each member to the final project, such that each member could reasonably reproduce the document on his or her own. An outline of the project is due the beginning of Week 5, while the final project is due finals week.

The purpose of the final project is to allow students an opportunity to demonstrate all the skills they have learned throughout the course. The final project **must** (a) be a reproducible and dynamic document produced with R Markdown, and include references to the extant literature; (b) demonstrate moving data from its raw "messy" format to a tidy data format through the R Markdown syntax, but not in the final pdf; (c) include at least <u>two</u> exploratory data visualizations, and (d) include at least summary statistics of the data in tables, although fitted models of any sort are an added bonus (not literally, there are not extra points for fitting a model). The points for the final project are broken down as follows:

- Adherence to each criterion listed above: 50 points (50%)
- Literature review (including citations): 10 points (10%)
- Two data visualizations: 20 points (10 points each)
- Production of <u>at least</u> one table (of summary statistics or model results): 10 points (10%)
- General style and grammar: 10 points (10%)

Final Project Presentation (30 points)
During weeks 9 and 10, students will be expected to present on their final project, which is expected to still be in progress. These presentations are expected to be informal. Students will be encouraged to discuss their process and the struggles they encountered along the way. Learning R is a difficult task, and we should all take solace knowing that others are struggling along with us. The final presentation should be equal parts "journey" and substantive findings/conclusions/results. Students are expected to present for approximately 10 minutes each, although the time may change depending on the enrollment of the class.

Labs (40 points; 5 points each)
The labs represent an integral part of this course, and as such are as important as the lectures themselves. Thus, participation in the labs contributes to your grade. If you are absent from a lab, please contact the instructor before the lab. Students may be excused from labs and given credit, provided they complete the lab as additional homework.

Grades will be determined, as follows:

Lower point range	Grade	Upper point range
$\geq$ 97%	A+	
$\geq$ 93%	A	$<$ 97%
$\geq$ 90%	A-	$<$ 93%
$\geq$ 87%	B+	$<$ 90%
$\geq$ 83%	B	$<$ 87%
$\geq$ 80%	B-	$<$ 83%
$\geq$ 77%	C+	$<$ 80%
$\geq$ 73%	C	$<$ 77%
$<$ 70%	C-	

7. Graduate/Undergraduate differentiation (if applicable)
N/A

8. Role of the Graduate Employee (GE)
N/A

9. Student Engagement Inventory.
Graduate: 1 credit hour = 40 hours of student engagement (4 credit hours = 160 hours of student engagement)

Educational activity	**Hours student engaged**	**Explanatory comments (if any):**
Course attendance	40	20 meetings, at 2 hours per meeting
Assigned readings	40	Weekly readings are assigned, and are expected to take roughly as long to complete as the in-seat time per week.
Project	40	Final project, as described above
Homeworks	40	A total of 8 homework assignments are expected to take students, on average, 5 hours to complete
Total hours:	160	

10. Attendance and Absence Guidelines
Attendance at all class and discussion groups is expected. See below for Absence Policy and Makeup Assignments.

Students must contact the instructor in case of illness or emergencies that preclude attending class sessions. Messages can be left on the instructor's voice mail or e-mail at any time of the day or night, prior to class. If no prior arrangements have been made before class time, the absence will be unexcused. Excused absences will involve make-up assignments, with make-up assignment procedures to be discussed in class on the first day.

If you are unable to complete an assignment due to a personal and/or family emergency, you should contact your instructor or discussion leader as soon as possible. On a case-by-case basis, the instructor will determine whether the emergency qualifies as an excused absence.

11. Expected Classroom Behavior
Students are expected to participate in classroom activities. If use of digital devices or engagement in other non-class activities during class for purposes not regarding classroom activities is critically necessary, the student should engage in these activities while on break, or check with the instructor to arrange for a 10-minute break for unusual circumstances, but only if critically necessary.

12. Graduate Employees (GE) in graduate level courses
N/A

13. Diversity, Equity and Inclusion
It is the policy of the University of Oregon to support and value equity and diversity and to provide inclusive learning environments for all students. To do so requires that we:

- respect the dignity and essential worth of all individuals.
- promote a culture of respect throughout the University community.
- respect the privacy, property, and freedom of others.
- reject bigotry, discrimination, violence, or intimidation of any kind.
- practice personal and academic integrity and expect it from others.
- promote the diversity of opinions, ideas and backgrounds which is the lifeblood of the university.

In this course, class discussions, projects/activities and assignments will challenge students to think critically about and be sensitive to the influence, and intersections, of race, ethnicity, nationality, language, religion, gender, socioeconomic background, physical and cognitive ability, sexual orientation, and other cultural identities and experiences. Students will be encouraged to develop or expand their respect and understanding of such differences.

Maintaining an inclusive classroom environment where all students feel able to talk about their cultural identities and experiences, ideas, beliefs, and values will not only be my responsibility, but the responsibility of each class member as well. Behavior that disregards or diminishes another student will not be permitted for any reason. This means that no racist, ableist, transphobic, xenophobic, chauvinistic or otherwise derogatory comments will be allowed. It also means that students must pay attention and listen respectfully to each other's comments.

14. Documented Disability
Appropriate accommodations will be provided for students with documented disabilities. If you have a documented disability and require accommodation, arrange to meet with the course instructor within the first two weeks of the term. The documentation of your disability must come in writing from the Accessible Education Center in the Office of Academic Advising and Student Services. Disabilities may include (but are not limited to) neurological impairment, orthopedic impairment, traumatic brain injury, visual impairment, chronic medical conditions, emotional/psychological disabilities, hearing impairment, and learning disabilities. For more information on the Accessible Education Center, please see http://aec.uoregon.edu.

15. Mandatory Reporting
UO employees, including faculty, staff, and GE's, are mandatory reporters of child abuse when the employee has "reasonable cause to believe any child with whom the employee comes in contact has suffered abuse or that any person with whom the employee comes in contact has abused a child." UO employees, including faculty, staff, and GE, also are mandatory reporters of prohibited discrimination when the employee obtains "credible evidence that any form of prohibited discrimination by or against students, faculty or staff is occurring." "Prohibited discrimination" includes discrimination, and discriminatory harassment, including sexual harassment and sexual assault. This statement is to advise you that that your disclosure of information about child abuse or prohibited discrimination to a UO employee may trigger the UO employee's duty to report that information to the designated authorities. Please refer to the following links for detailed information about mandatory reporting: http://around.uoregon.edu/mandatoryreporting.

16. Academic Misconduct Policy
All students are subject to the regulations stipulated in the UO Student Conduct Code http://conduct.uoregon.edu). This code represents a compilation of important regulations, policies, and procedures pertaining to student life. It is intended to inform students of their rights and responsibilities during their association with this institution, and to provide general guidance for enforcing those regulations and policies essential to the educational and research missions of the University.

17. Conflict Resolution
Several options, both informal and formal, are available to resolve conflicts for students who believe they have been subjected to or have witnessed bias, unfairness, or other improper treatment. It is important to exhaust the administrative remedies available to you, including discussing the conflict with the specific individual, or contacting the Department Head. Within the College of Education, you can contact the Associate Dean for Academic Affairs (Brigid Flannery, 346-2496, brigidf@uoregon.edu) or the Assistant Dean for Equity and Inclusion (Krista Chronister, 346-2415, kmg@uoregon.edu).

Outside the College, you can contact:

- UO Bias Response Team: 346-3216 http://bias.uoregon.edu/whatbrt.htm
- Conflict Resolution Services: 346-3216 http://studentlife.uoregon.edu/support
- Affirmative Action and Equal Opportunity: 346-3123 http://aaeo.uoregon.edu/

18. Grievance Policy
A student or group of students of the College of Education may appeal decisions or actions pertaining to admissions, programs, evaluation of performance and program retention and completion. Students who decide to file a grievance should follow the student grievance procedure, or alternative ways to file a grievance outlined in the Student Grievance Policy (https://education.uoregon.edu/academics/student-grievance or enter search: student grievance).

19. In Case of Inclement Weather
In the event the University operates on a curtailed schedule or closes, UO media relations will notify the Eugene-Springfield area radio and television stations as quickly as possible. In addition, a notice regarding the university's schedule will be posted on the UO main home page (in the "News" section) at http://www.uoregon.edu. Additional information is available at http://hr.uoregon.edu/policy/weather.html.

If an individual class must be canceled due to inclement weather, illness, or other reason, a notice will be posted on Canvas or via email. During periods of inclement weather, please check Canvas and your email rather than contact department personnel. Due to unsafe travel conditions, departmental staff may be limited and unable to handle the volume of calls from you and others.

20. Course Incomplete Policy
Students are expected to be familiar with university policy regarding grades of "incomplete" and the time line for completion. For details on the policy and procedures regarding incompletes, please see: https://education.uoregon.edu/academics/incompletes-courses.