
Course Syllabus

EPPS6324: Data Management for Social Science Research Spring 2020

Times and Locations

- Wednesdays 4:00-6:45 pm in GR 3.402 B

Instructor

- **Michael Tiefelsdorf**, GR 3.204, (972) 883-4954, tiefelsdorf@utd.edu.
For course related emails please start the subject line with **EPPS6324**.
- Office hours: Monday 2:00 – 3:45 pm in my Office at GR3.204. By appointment or when my office door is open.


Teaching Assistant

- Yalin Yang, GR3.414, Yalin.Yang@UTDallas.edu
For course related email please start the subject line with **EPPS6324**.
- Office hours: TBA, GR3.204


The Homepage



- See the link to EPPS6324 for the Spring 2020 semester on *eLearning*.
-

Course Pre-requisites, Co-requisites, and/or other Restrictions

- Working knowledge of personal computers is required. Experience of working with  either at the command prompt or with scripts is preferred.
 - You must have a Comet Card, Student ID and password to use the School's lab computers and access eLearning
-

Course Description



This course provides a solid foundation of how to import, handle, manipulate, organize, explore and visualize a broad range of data sources. Furthermore, it introduces to automating many data preparation and analysis tasks. This course solely focuses on the extremely powerful and broadly used statistical data analysis system , rather than being spread thinly over several software environments.

 being this powerful implies that it has a somewhat steep learning curve; however, current advances have made it broadly accessible to the professional and academic community. Among many data analysis tasks  can be used for example for predictive analytics, decision support systems, standard and advance statistical data analysis including Bayesian statistics, time series analysis, social network analysis, geo-spatial information analysis and mapping, as well as standard and advanced machine learning algorithms.


Beyond data management this course also follows the *paradigm that research must be reproducible*, so that peers can review and validate it. Therefore, an objective of this course is to document all steps of the data preparation, the employed analysis procedure and its results.

Student Learning Objectives/Outcomes


Upon completing this class, students will:


- Understand the most commonly used data types and formats in statistical data analytics;
- Have capabilities to import data into the  statistical software environment;
- Obtain solid skills of preparing data for statistical data analyses;
- Have the ability to automate data analysis tasks;
- Learn to perform *reproducible research* by building a  package.

Required Textbooks and Materials

Textbooks: Required: Wickham H. & Golemund G., 2016:  for Data Science. Import, Transform, visualize, and Model Data. O'Reilly. (available at <https://r4ds.had.co.nz>. Its markup source is available at <https://github.com/hadley/r4ds>)

Optional: Kabacoff R.L., 2015.  in Action. Data Analysis and Graphics with . Manning Publications, 2nd edition (Available online at UTD's library)

Optional: Lander J.P., 2014.  for Everyone. Advanced Analytics and Graphics. Addison Wesley (available online at UTD's library)

Optional: Wickham H., 2015.  Packages. Organize, Test, Document, and Share your Code. O'Reilly (available at <http://r-pkgs.had.co.nz>)

Supplements: Lecture notes, tutorial presentations and additional reading resources on special topics augment the study material. They will be posted on *eLearning*.

Software: The software for this course is freely available at www.r-project.org , or <https://mran.microsoft.com/> and <https://www.rstudio.com/> .

Tentative Schedule:

The topic of each week should have been studied before the lecture and prepare the course participants for the weekly *eLearning* quizzes. Quiz01 to Quiz09 are based on <https://r4ds.had.co.nz>.

Date	Topic	Tasks
Jan-15	Introduction	
Jan-22	R Markdown	Sample Quiz00



Jan-29	Presentations: Data Visualization	Quiz01
Feb-05	Presentations: Data Transformation	Quiz02
Feb-12	Presentations: Exploratory Data Analysis	Quiz03 & Lab01 out
Feb-19	Presentations: Tibbles, Data Import & Tidy Data	Quiz04
Feb-26	Presentations: Relational Data & Strings	Quiz05
Mar-04	Presentations: Factors & Dates and Times	Quiz06 & Lab02 out
Mar-11	Presentations: Pipes & Functions	Quiz07
Mar-18	Spring Break	
Mar-25	Presentations: Vectors & Iterations	Quiz08 & Lab03 out
Apr-01	Presentations: Graphics for Communication & Markdown Formats	Quiz09
Apr-08	Building R Packages (Chapter 11: Creating a Package in Kabacoff)	Quiz10
Apr-15	Mining Census Data (Handout)	Lab04 out
Apr-22	Project Preparations I	
Apr-29	Project Preparations II	
TBA	Project presentations are held during the final exam time slot.	

List of 22 Student Presentations (see sections at <https://r4ds.had.co.nz>)

Date	Sections	Presenter
Jan-22	26.0	Tiefelsdorf
Jan-22	27.0 – 27.7	Tiefelsdorf
	Tutorial Presentation Set I	
Jan-29	3.0 – 3.5	N.N.
Jan-29	3.6 – 3.10	N.N.
Jan-29	4.0 – 4.4	Tiefelsdorf
Feb-05	5.0 – 5.5	N.N.
Feb-05	5.6 – 5.7	N.N.
Feb-05	6.0 – 6.3	Tiefelsdorf
Feb-12	7.0 – 7.4	N.N.
Feb-12	7.5 – 7.8	N.N.
Feb-12	8.0 – 8.5	Tiefelsdorf
Feb-12	9.0	Tiefelsdorf
Feb-19	10.0 – 10.5	N.N.
Feb-19	11.0 – 11.6	N.N.
Feb-19	12.0 – 12.7	N.N.
Feb-26	13.0 – 13.7	N.N.
Feb-26	14.0 – 14.7	N.N.
	Tutorial Presentation Set II	
Mar-04	15.0 – 15.5	N.N.
Mar-04	16.0 – 16.5	N.N.
Mar-04	17.0 – 17.1	Tiefelsdorf
Mar-11	18.0 – 18.4	N.N.
Mar-11	19.0 – 19.7	N.N.
Mar-25	20.0 – 20.7	N.N.
Mar-25	21.0 – 21.5	N.N.
Mar-25	21.6 – 21.9	N.N.
Apr-01	28.0 – 28.3	N.N.
Apr-01	28.4	N.N.

Apr-01	28.5 – 28.8	N.N.
Apr-01	29.0 – 29.10	N.N.

Tentative Grading Scheme

Tasks	Points
<u>Tutorial Presentation from Set I:</u> 25-35 minutes presentation & markdown handout	15 pts
<u>Tutorial Presentation from Set II:</u> 25-35 minutes presentation & markdown handout	15 pts
<u>Quizzes:</u> 10 quizzes with 2 pts each. By default, 2 pts are assigned to the weakest quiz or one missed quiz.	20 pts
<u>Labs:</u> 4 Labs, each ± 5 pts.	20 pts
<u>Project:</u> Design of an  package with at least [a] one user-defined function, [b] a dataset and [c] a vignette. Presentation and discussion of the  package during the finals time slot. Group work for larger packages with up to two participants is possible. Each group member must explicitly identify their unique contribution.	30 pts

Tentative Grading Scale

Points	Grade
90 to 100	A
85 to 89	A-
80 to 84	B+
75 to 79	B
70 to 74	B-
65 to 69	C+
60 to 64	C
less than 60	F

Course & Instructor's Policies

Make-up exam/Late assignment policy: A make-up exam will only be given in extenuating circumstances.

Participants will usually have 14 days to complete a lab. A late lab will lead to a deduction of its grade. A late lab can no longer be accepted once its solution has been posted and discussed.

Plagiarism: The university's rules of plagiarism will be strictly enforced. While you are encouraged to discuss the course material and labs with other course participants to enhance your understanding of the course material, the labs must be answered individually unless teamwork is explicitly requested by the instructor.

Class attendance: Class attendance is expected. Students, who consistently miss classes, are on their own and will be reported to the program director.

Decorum during class:

- Turn cell phones off during class and no text messaging.
 - No course unrelated computer uses during the lectures.
 - Respect the lab's no food and drink policies.
 - Respectful interaction among all participants.
-

Additional information relating to University policy on "Religious Holy Days", "Grade Appeals", "Disability Services", "Student Conduct" etc. can be found at: <http://go.utdallas.edu/syllabus-policies>

These descriptions and timelines are subject to change at the discretion of the instructor.