

Introduction to Data Science for Psychology Majors

PSY 490

Fall, 2024

Professor: Gaurav Suri, Ph. D.

suri@sfsu.edu

Classes are in person

Location: Burk Hall 236

Class Time: MW 2 – 3.15 pm

Welcome! I am looking forward to working with you this semester.

What profession did Harvard call the Sexiest Job of the 21st Century? That's right... the data scientist. Ah yes, the ever-mysterious data scientist. So, what exactly is data science?

Data science is an inter-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from many structural and unstructured data. Data science is related to data mining, machine learning and big data.

The objective of this course is to introduce psychology majors to data science. If you complete this course you will have a clear understanding of the basic principles of data science. Some of you may choose to use your experience in this class as a stepping stone towards a career in data science.

Prerequisites

You must have completed a college-level statistics course to participate in this class (PSY 371 or equivalent). In addition, you should be mentally prepared that this is a challenging class that requires 8 to 10 hours of work per week. Please do not take this class just because it suits your schedule or meets an Area-1 requirement. Wanting to learn data science, and being a self-starting problem solvers are core attributes to succeed in this class.

Computing Requirements

This course will be conducted in the programming language R. R is a programming language and **free** software environment for statistical computing and graphics supported by the R Foundation for Statistical Computing. The R language is widely used among statisticians and data miners for developing statistical software and data analysis. Once you learn R, learning other languages (e.g. Python) is relatively easier.

Important Note: To participate in this course, you must have access to a computer on which you can download R.

The main software we will be using in this class is R and RStudio. R is a statistical language. RStudio is an integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management.

Here is a guide to install R and RStudio.

<https://rstudio-education.github.io/hopr/starting.html>

Our class will mostly be conducted in RStudio. PLEASE DOWNLOAD R AND R-STUDIO BEFORE THE FIRST DAY OF CLASS

Course Description

The course will be hands-on and require you to learn by doing. Our lectures will be in person (I do not expect to have Zoom recordings). The teaching will be example heavy, and heavily reliant on working in RStudio.

Course Objectives

By the end of this course, you should be able to answer the following questions:

1. How to import any data into R?
2. How to ‘clean’ your data to make it understandable to yourself and others?
3. How to transform your data to enable the analysis that you are most interested in?
4. How to visualize data? (Visualization is an essential aspect of data science).
5. How to build models to understand the data?
6. How to communicate your analysis to other people?

Course Organization & Office Hours

The class will meet on Monday, Wednesday, 2-3.15pm, Burk Hall 236.

Office Hours

My office hours will be as follows:

Mondays: 3.30-5.30 pm, in EP 306

Please prepare specific questions before an in-person meeting. I will also be happy to answer questions via email or by appointment. If you need an appointment, please send me an email at suri@sfsu.edu.

Course Topics

There are three main components to this course:

- 1) Data Sub-setting and Visualization,
- 2) Introduction to Programming, and
- 3) Introduction to Machine Learning

Materials

The main textbook for 'Data Sub-setting and Visualization' will be the following on-line text:

R for Data Science

By Garrett Golemund & Hadley Wickham

It can be found here:

<https://r4ds.had.co.nz/index.html>

We will not be following the book exactly, but it will guide a substantial chunk of the material we will cover. We will also be referring to other materials that will be provided as needed.

A good tutorial for data visualization:

<http://r-statistics.co/Complete-Ggplot2-Tutorial-Part1-With-R-Code.html>

Here is the text for the Basic Programming module:

<https://rstudio-education.github.io/hopr/index.html>

I will distribute materials for the machine learning section of the course as we come to that portion of the course.

Academic Policies

Policy on Cheating

All campus rules for plagiarism, cheating, and incomplete grades will be strictly enforced. Definitions and consequences may be found in the SFSU Bulletin Please familiarize yourself with these policies as you are expected to abide by them.

Deadlines

Familiarize yourselves with the *deadlines* for withdrawing from class without a "W" and withdrawing for a serious and compelling reason. ("Serious and compelling" is not your call. The Bulletin explains that you must receive approval from the instructor, department chair, and school dean.)

Learning Adjustments

Please contact Dr. Suri right away (first week of class) if you have a documented difference in learning ability or if you have some other special needs that may require adjustments for you to learn/understand the material. Students with disabilities should work with Disability Programs and Resource Center (415-338-2472) and the instructor to arrange appropriate support. Visit this office on the web at <http://www.sfsu.edu/~dprc/>. The DPRC is located in the SSB-110.

Mandatory Title IX Reporting

SFSU policy is that professors (and employees) must report any allegations or instances of sexual violence that they hear about or witness. For more information, about Title IX rights and protections at SFSU, please visit: <http://titleix.sfsu.edu/>

Class Requirements and Activities

Attendance

This is a fast-moving course in which our discussion in class will have a lot to do with understanding of the material and your ability to do the HWs. Therefore, you must plan on attending all class sessions. Missing classes WILL affect your grade.

Homework

There will be 7 homework assignments during the semester, given out approximately every two weeks. Each HW will be worth 100 points. HWs will generally be due one week after they are released, but please consult each assignment for the relevant due date. 10%/day will be deducted for late HWs.

You may work collaboratively on the homework, but please write your own solutions. Many of the HWs will require you to submit R-code.

Some of the homeworks will take some effort. Start early and keep at it!

Grading

Item	Weight
Homework	70% (10% Each)
Final	30%

I will grade on a curve.

If you're struggling with the HWs and fighting to stay afloat, don't forget that everyone struggles with something. Come and talk to me, and hopefully, we'll find a way forward.

Key Topics

There are three key topics we will cover in this course: 1) Data Sub-setting and Visualization, 2) Introduction to Programming, and 3) Introduction to Machine Learning.

The pace of these topics will depend on how fast the class wants (or can) go. I'm not interested in creating an environment where only a few people learn, and everyone else is lost. We will take our time if we need to, but we won't dilly-dally when we don't.

Here is the rough schedule I'm expecting. I will update this schedule as we proceed (color-coded for three topics).

#	Week of...	Chapters/Topic	HWs
1	8/26	Syllabus, R Overview	
2	9/2	1-4	HW 1 Released
3	9/9	4, 5	
4	9/16	6, 7	HW 2 Released
5	9/23	7, 8	
6	9/30	9, 10	HW 3 Released
7	10/7	Basic programming with R – variables and data structures	
8	10/14	Basic programming with R - conditionals	HW 4 Released
9	10/21	Basic programming with R - Loops	HW 5 Released
10	10/28	Putting it together: Writing end-to-end programs	
11	11/4	Basic programming with R - Consolidation	HW 6 Released
12	11/11	Regression - 1	

13	11/18	Regression - 2	HW 7 Released
14	11/25	Supervised Learning - 1	
15	12/2	Supervised Learning - 2	Final Project Released
16	12/9	Review: Putting it together	
17	Final	Take Home	

Data Science Certificate for Psychology Majors

The Psychology department offers a Certificate in Data Science for Psychology Majors. Obtaining this certificate will allow students to showcase their data science skills to future employers.

To obtain this certificate, students are required to complete the following classes:

Course Prefix	Course number	Title	Prerequisite	Units
PSY	171	Quantitative Reasoning in Psychology	None	3
PSY	200	General Psychology	None	3
PSY	371	Psychological Statistics	PSY 171 and PSY 200 or equivalents	3
PSY	571	Intermediate Psychological Statistics	PSY 371	3
PSY	490	Introduction to Data Science for Psychology	PSY 371	3
CSC	101	Introduction to Computing	None	3
			Total Units	18

Once you have completed these courses, please bring your transcript to the Psychology Department (EP, 301) to pick up your certificate!

Calendar Highlights

Monday, August 26, 2024

Fall '24 First Day of Instruction

Fall '24 Payment Due Date for All New International Students Including new graduate students

Fall '24 Adding Classes | Permission numbers ARE NOT required to add most classes Begins

Monday, September 2, 2024

Fall '24 Labor Day; No Classes; Office Closed

Fall '24 Adding Classes | Permission numbers ARE NOT required to add most classes Ends

Tuesday, September 3, 2024

Fall '24 Adding Classes | Permission numbers ARE Required Begins

Sunday, September 15, 2024

Fall '24 Last Day for Faculty Drops

Monday, September 16, 2024

Fall '24 Adding Classes | Permission numbers ARE Required Ends

Fall '24 Last Day to Drop/Withdraw Classes without a W grade

Fall '24 Audit Deadline

Tuesday, September 17, 2024

Fall '24 Withdrawal from Classes or University For Serious and Compelling Reasons Begins

Monday, October 21, 2024

Last Day to Register to Vote

Tuesday, November 5, 2024

General Election

Monday, November 11, 2024

Fall '24 Veterans Day; No Classes; Office Closed

Fall '24 Withdrawal from Classes or University For Serious and Compelling Reasons Ends

Tuesday, November 12, 2024

Fall '24 Withdrawal from Classes or University By Exception for Documented Serious and Compelling Reasons Begins

Monday, November 25, 2024

Fall '24 Recess; No Classes; Campus Open

Tuesday, November 26, 2024

Fall '24 Recess; No Classes; Campus Open

Wednesday, November 27, 2024

Fall '24 Recess; No Classes; Campus Open

Thursday, November 28, 2024

Fall '24 Recess; No Classes; Campus Closed (Thanksgiving)

Friday, November 29, 2024

Fall '24 Recess; No Classes; Campus Closed

Fall '24 Indigenous Peoples Day; No Classes; Office Closed

Saturday, November 30, 2024

Fall '24 Recess; No Classes; Campus Closed

Friday, December 6, 2024

Fall '24 Grading Option Deadline (This includes Cr/No CR or Letter Grade)
Friday, December 13, 2024
Fall '24 Last Day of Classes
Fall '24 Withdrawal from Classes or University By Exception for Documented Serious
and Compelling Reasons Ends

Psychology Department Statement on Inclusion

We VALUE Respect, Civility, and Inclusion

The Psychology Department is committed to providing a multi-cultural academic setting and climate that is respectful, inclusive, civil, and harassment-free for individuals regardless of ability, age, ethnicity, gender identity or expression, indigenous heritage, immigration status, nationality, physical appearance, political affiliation, race, religion or spirituality (or lack thereof), sexual orientation, social class, socioeconomic status, or other cultural, demographic, and social. Examples of language or behaviors in the academic setting that are not respectful include but are not limited to:

- Discriminatory statements that are overt (e.g., hostility) or subtle (e.g., microaggressions)
- Posting discriminatory materials on campus or campus-related forums (iLearn, Zoom)
- Behaviors or statements that interfere with the learning of others
- Marginalizing individuals based on their social or demographic identity, or perceived identity

identification*.

We ASK that all members of our department, including undergraduate students, graduate students, faculty, staff, and guests, treat each other with respect and civility. We strive for this in every aspect of our interactions including, for example, in language (e.g., pronouns), verbal and written communication (e.g., thoughtful and active listening), and behaviors in the virtual or real classroom, research laboratory, as well as in any other setting that holds a department function.

- Inappropriate, unwelcome, or non-consensual comments, behaviors, photography or recording
 - o Disability Programs and Resource Center (<https://access.sfsu.edu/>)
 - o Diversity, Student Equity, and Interfaith Programs (<https://equity.sfsu.edu/node/16>)
 - o Mental health at Counseling & Psychological Services (<https://psyservs.sfsu.edu/>)
 - o Physical health at the Health Center (<https://health.sfsu.edu/>)
- Advocating for or encouraging any of the above

We encourage ACCOUNTABILITY and seeking SUPPORT

- To receive support for
 - o Asian American and Pacific Islander Retention and Education (<https://aspire.sfsu.edu/home>)
 - o Sexualized violence at the Safe Place (<https://psyservs.sfsu.edu/content/safe-place>)
 - o Undocumented students at the Dream Resource Center (<https://drc.sfsu.edu>)
- To contact
 - o Professors (<https://psychology.sfsu.edu/>)
 - o The University Police Department (<https://upd.sfsu.edu/>)
- To report
 - o Harassment or discrimination based on sex, gender, or sexual orientation (<https://titleix.sfsu.edu/content/reporting>)
 - o A bias incident (<https://equity.sfsu.edu/content/bias-incident-education-team-sf-state>)
- To learn about the University Code of Conduct
 - o Undergraduate: (<https://conduct.sfsu.edu/standards>)
 - o Graduate: (<http://bulletin.sfsu.edu/graduate-education/academic-policies-procedures/>)

rev 4/21/2021

*We alphabetized identities to avoid implying a preference for a particular identity.