HS & MS - Data Science II

Summer 2020

Instructor: Dr. Immanuel Williams
Email Address: **jamesijw23@gmail.com**
Website: https://jamesijw23.github.io/datascience_2/
Instagram: crd_statcalpoly
Office Location: **Zoom Link:** https://calpoly.zoom.us/j/440387952
Office Hours: 30 minutes before Class Starts
Classtime: MTWRF, 1:00 PM - 4:00 PM

Course Description

Overview: The ability to think with a data science mind set is becoming increasingly important for life in our data driven world. You will encounter data science throughout your life in various professional, public, and personal contexts. To prepare you for this lifelong journey, this course will present key data science concepts in R through examples from a variety of real, recent, and relevant examples. The goal of this course is to discuss topics and examples to **inspire** your intellectual curiosity and help you see data science as a way of thinking critically about the world.

Prerequisite: Experience with algebra and familiarity with computers is ideal.

Expected Learning Outcomes:

- Ability to joing and combine data frames in R
- Ability to manipulate data within R
- Ability to create a R-Shiny (web application) in R

Notes and Logistics

Notes will be posted at https://jamesijw23.github.io/datascience_2/ every morning before class. You have the option to have your video on during class (highly encourage) but it is imperative that you stay focus for the entirity of the class. There will be designated times for 5-10 minute breaks, however, you may leave at anytime if an emergency arises.

There will be videos that go over the content in a simplistic manner that are paramount to the work done in class. Be sure to watch videos before attending class. If issues or questions arise, please take note of them and ask during class.

Grading

There is no formal grade for this course, however, there will be daily homework assignments. This work is meant to reinforce the concepts learned in class as well as new concepts. There will be a **final project** where the students will create data visualizations within an R Shiny.

Concepts

Major Concepts	Minor Concepts
Concept 5r	Review I of DS 1: R and R files, Review II of DS 1: Extraction, Manipulation, Visualization (EMV), Review III of DS 1: Statistical Concepts
Concept 6	Description of Fornite dataframe, Bind Columns, Bind Rows, Inner Join, Left Join, Full Join
Concept 7	String Manipulations I - (str_trim, str_view_all, str_detect), String Manipulations II - (str_remove_all, str_replace_all, str_extract), String Manipulations III - (str_split, str_order, str_to_title)
Concept 8	Basics I: How R Shiny works, Basics II: Type of Widgets, Basics III: Display Information, Application I: Kable, Application II: Visualizations, Application III: Flexdashboard Part I, Application III: Flexdashboard Part II
Concept 9	Box Plot Faceted Graphs, Scatter Plot with 5 variables

The concepts may change based on the pace of the class.

How to send Emails:

Subject Line: Reading in Excel files

Dear **Professor Williams**:

I hope your day is going well. My name is **Jackie Robinson** and I am a student in your **Data 101 Course** ...

Thank you very much and have a great day!

Sincerely,
Jackie Robinson
Alpha Phi Alpha Fraternity President
Sherwood High School
Class of 2023

Strategies For Success:

- **Attend class every day**. The lectures notes and homework work together to give you a fuller understanding of course material. Participate actively in class and ask questions. Pay attention, take the examples seriously, take notes, and respond to the instructor's questions. Speak up and ask questions.
- **Review** notes before and after class. Plan a study cycle where you are both preparing for new material while also reviewing previous topics. Be mindful of connections between course topics and concepts.
- **Get to know your classmates**. Work together on problems and studying for exams. Practice speaking the language of data science and build confidence in your understanding of statistical concepts.
- **Be organized**, start homeworkd early, and keep up. Create a weekly schedule for this class and stick to it. Spend about 1-2 hours per week outside of class reviewing and studying this material. It is best to spend some time every day working on this class. Spend this time: preparing for class and answering practice problems.
- **Get lots of practice**. We will cover in class just a few examples relating to each topic. Do review problems from class, but also be sure to work additional exercises on your own. If the material is too easy ask me and I will be provide additional problems for you. Also, get in the habit of identifying and critiquing data science information you see in daily life.
- **Effective communication is essential.** Get plenty of practice writing and speaking about data science concepts and analysis. In this class, your answer will rarely just be a single number resulting from a mathematical calculation. You will be graded on the

quality of your writing. Choose your words carefully, read what you have written out loud, and revise for clarity.

- **Be responsible for your own learning.** I'm here to help you succeed in whatever ways I can, but ultimately you will get out of this course what you put in. Being responsible also includes spending quality time on the material outside of class, knowing and meeting course due dates, starting assignments early, asking questions, reviewing solutions and your graded assignments, etc. Take the course seriously and take pride in your work.
- **Have fun!** I hope you will find this course fun and interesting. I have tried to present real-world examples that are interesting and relevant. Allow yourself to enjoy these aspects of the course, and feel free to suggest other applications that appeal to you. Think about ways the course material applies to your other classes and interests.