**Data Analytics in R Community**

**Instructors**

[Megha Joshi, PhD](https://meghapsimatrix.com/)

Email: [megha.j456@gmail.com](mailto:megha.j456@gmail.com)

Name

Email:

**Course Description**

This course is intended to sharpen your data analytic skills to prepare you for jobs and careers requiring such skills. We will cover fundamentals of data analytics in R particularly. We will also cover some translations between R and Structured Query Language (SQL). Some target job positions include data analyst, quantitative researcher, quantitative user researcher, and data scientist.

The course will involve one-hour lecture every week for 12 weeks. After the lecture each week, the we will provide a set intensive practice exercises that will help you learn concepts covered in class. Some of the exercises will be based on topics covered during the lecture and some will require you to think a bit creatively and maybe do some searching online. We will provide answers to these exercises one week after.

Additionally, we will attempt to provide some resources and help for job search. We will provide example resumes, example technical and behavioral interview questions, and descriptions of different careers paths.

**Expectations**

The more you put into this course, the more you will get out. This course is really for you to learn and improve your data analytic skills. We are not going to provide grades.

We will provide a list of recommended readings each week. Most of the readings will be from the [R for Data](https://r4ds.hadley.nz/) Science book with some other articles and book chapters. We encourage you to look through recommended readings to help guide you as you learn R.

We particularly emphasize that completing the practice exercises that we provide will help you really learn the concepts. Instructors will not be able to grade and provide detailed feedback on your work on these problems. We will, however, provide answers to these problems. We believe that people learn programming best not just by watching lectures but by actively practicing. Thus, we highly encourage you to complete the practice exercises. Feel free to work together in groups and learn from each other. We have created a Slack channel to form a community to learn R together. Please reach out to each other and post questions or resources. Also, try to apply what you learn here on your own projects.

**Schedule & Recommended Readings**

Class 1: Introduction to the Course and to R

* Recommended readings:
  + [R Programming for DS Chapter 2](https://bookdown.org/rdpeng/rprogdatascience/history-and-overview-of-r.html)
  + [R4DS Introduction](https://r4ds.hadley.nz/intro.html)
  + [R4DS Chapter 6](https://r4ds.hadley.nz/workflow-scripts.html)
  + [R4DS Chapter 28 - Quarto](https://r4ds.hadley.nz/quarto.html)
* Problem set 1

Class 2: Data Visualization with ggplot

* Recommended readings:
  + [R4DS Chapter 3](https://r4ds.had.co.nz/data-visualisation.html)
* Problem set 2

Class 3: Querying and Wrangling Data (select, mutate, filter, arrange)

* Recommended readings:
  + [R4DS Chapter 5 Sections 1-5](https://r4ds.had.co.nz/transform.html)
  + R Programming for DS Chapter 12
* Problem set 3

Class 4: Pivoting (pivot\_longer, pivot wider)

* Recommended readings:
  + [R4DS Chapter 12 Section 3](https://r4ds.had.co.nz/tidy-data.html#pivoting)
  + [Vignette on Pivoting](https://tidyr.tidyverse.org/articles/pivot.html)
* Problem set 4

Class 5: Working with Grouped Data (group\_by, summarize)

* Recommended readings:
  + [R4DS Chapter 5 Sections 6-7](https://r4ds.had.co.nz/transform.html#grouped-summaries-with-summarise)
* Problem set 5

Class 6: Relational Databases

* Recommended readings:
  + [R4DS Chapter 13](https://r4ds.had.co.nz/relational-data.html)
  + dbplyr [vignette](https://dbplyr.tidyverse.org/articles/translation-verb.html) on translation of dplyr to SQL
* Problem set 6

Class 7: String Manipulation

* Recommended readings:
  + [R4DS Chapter 14](https://r4ds.had.co.nz/strings.html)
* Problem set 7

Class 8: Working with Other Data Types in R (factors, vectors, lists, matrices, arrays, dates)

* Recommended readings:
  + [R4DS Chapter 15](https://r4ds.had.co.nz/factors.html)
  + [R4DS Chapter 16](https://r4ds.had.co.nz/dates-and-times.html)
* Problem set 8

Class 9: Functions

* Recommended readings:
  + [R4DS Chapter 19](https://r4ds.had.co.nz/functions.html)
* Problem set 9

Class 10: Iteration

* Recommended readings:
  + [R4DS Chapter 21 Iteration](https://r4ds.had.co.nz/iteration.html)
* Problem set 10

Class 11: Linear Regression

* Recommended readings:
* Problem set 11

Class 12: Gentle Introduction to Machine Learning

* Recommended readings:
* Problem set 12

**Further Recommended Readings**

Healy, K. (2018). *Data Visualization: A Practical Introduction.* Princeton University Press. Free version [here](https://socviz.co/).

Ismay, C., & Kim, A. Y. (2019). *Statistical inference via data science: A ModernDive into R and the tidyverse*. Chapman and Hall/CRC. Free version [here](https://moderndive.com/v2/index.html).

Peng, R. D. (2016). *R programming for data science* (pp. 86-181). Victoria, BC, Canada: Leanpub. Free version [here](https://bookdown.org/rdpeng/rprogdatascience/).

Wickham, H., Cetinkaya-Rundel, M. & Grolemund, G (2023). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data (2e) R for Data Science (2e).* O’Reilly Media, Inc. Free version [here](https://r4ds.hadley.nz/).