Course Title: Introduction to R
Course Number: G450
Credits: 1
Course Directors: Jared Lander
Course Prerequisites: None; Open to first year Cancer Biology PhD students; open to others as approved
Grading Policy: Pass/Fail
Software Need for the Course:

- Attendees will bring their own laptops with R installed as well as an appropriate IDE.
- During a preparatory session on September 9, 2026, students will meet with the teaching fellows to ensure that all necessary software and packages are downloaded and functioning. Instructions on required software and packages will be listed on the course GitHub repository, which can be accessed with your Github name. https://github.com/jaredlander/msk2026fall/issues
- Students will be asked to use Claude Code during the last session of class. Students may register for a Claude Pro Account ($20 a month) and then submit their receipts to Stacey Lara following the MSK reimbursement policy. GSK will reimburse for up to 3 months of Claude usage ($60 total). Note you must not enter PHI or sensitive data/code into Claude.

Course Structure
Five two hour training session per week for five weeks and five one-hour breakout sessions for five weeks.

Week 1

Introduction to the Basics

- The RStudio Interface
- Basic Math
- Assigning Variables
- Working Directories
- Relative Paths
- Reading Data
    - Read from text files with readr
    - Read from Excel files with readxl

Quarto/RMarkdown

- Markdown
    - Sections
    - Text Formatting
    - Lists
    - Links

- Integrating R into Markdown
    - Chunk Options
- Including Figures
- Output Formats
    - HTML
    - PDF
    - Word
- Presentations

Week 2

Data Manipulation with dplyr

- Understanding a tibble
- Use pipes for cleaner code
- Select columns with select
- Filter rows with filter
- Change and create columns with mutate
- Calculate summary statistics with summarize
    - Mean
    - Max
    - Minimum
- Group data for calculations with group_by

Creating Visualizations

- ggplot2 paradigm
- Aesthetics
- Scatter plots
    - Color Coding
    - Size
    - Shape
    - Opacity
- Small multiple plots
- Histograms
- Density Plots
- Combining Layers
- Themes

Week 3

Reading Data

- CSVs with readr
- Databases with DBI
- JSON with jsonlite

Iterating over Lists and Reshaping Data

- Basics of functional programming
- Mapping over a list
- Difference from lapply
- Consistent Data Types
- Mapping to different data types
    - character
    - numeric
    - data.frame
- Mapping functions with multiple arguments
- Convert from wide to long with pivot_longer
- Convert from long to wide with pivot_wider

Week 4

Basic Modeling

- Mean
- Standard deviation
- T-test
- Simple Linear Regression
- Multiple Regression

Combining Analyses

- Read in new dataset
- Basic exploratory plotting
- Manipulate data for modeling
- Fit multiple models and pick the best
- Run all processes and report results in a rendered PDF via Quarto

Week 5

Working in the Terminal

- Using the terminal in RStudio/Positron/VS Code
- Using your computer's built-in terminal
  - Windows: Windows Terminal
  - Mac: iTerm
- Navigating the file system
  - ls
  - cd
  - tree
- Creating Files and Directories
  - mkdir
  - touch
  - echo
- Viewing files
  - cat
  - head/tail
  - vim
- Environment Variables
  - Setting/Reading
  - Using in R

Setting Up Claude Code

- Installing Claude code
- Install gh cli
- Create ~/.Claude folder
- Setting up Claude code
  - Model
  - IDE
  - Statusline
- Ask a question in Claude
- Ask Claude to summarize existing project
- Ask it to create a CLAUDE.md file based on what it see in the project
- Plan Mode
- Skills
  - Ask Claude web to write a specific skill
  - Save skill to web
  - Download to computer and unzip
- MCP
  - Context7

    - Others
- Ask Claude to update skills to use context7
- Ask Claude to update claude.md to use skills and context7
- Plugins
    - Explain them
    - Add market place
    - Install plugin
- Claude code browser extension
    - Browse LinkedIn for me

Teaching Fellows

Teaching Fellows, drawn from senior GSK students and the postdoctoral community at MSK, are present in the course sessions. Their role is to act as an additional source of information/assistance, to help keep the discussion sessions moving.

Assignments and Methods for Assessing Student Achievement

Attendees will be given take-home assignments each week, due the following week that will be graded by MSK teaching assistants and graders and provided back to them. This course is graded on a pass/fail basis; class participation and assignments will form the basis of the grading.

Course Evaluation

Students are expected to complete surveys regarding the lectures and overall course via their student portal. This feedback will be used to evaluate the effectiveness and relevance of the topics and provide direction for the subsequent iterations of the course.

Academic Dishonesty, Plagiarism and Artificial Intelligence
The Policy can be found in the Student and Faculty Handbook linked on the GSK Website.

Workgroups
Students will be arranged into working groups for projects and assignments as follows; please be advised that although you can work in groups, each student should submit his/her/their own individual assignments.

Group 1

Xing, Kristi
Adelstein, Anna
Kapadia, Naviya

Group 2

Lin, William
Davis, Samantha
Kenum, Carson

Group 3

Zaroogian, Zachary
Deng, Yiran
White, Christian

Group 4

Xu, Shengze
Donahue, Rebecca
Molina , Samuel

Course Schedule
The course schedule can be found on the following pages and on Moodle.