
THE UNIVERSITY OF CHICAGO

# R Programming Workshop

Narayanan Venkataraman
narayanan@uchicago.edu 312.721.9944

# Workshop Objectives

- Get Comfortable with R
  - For Your Classes
  - Individual Projects
- Learn to Use Tools Built For The R Environment
- Learn to Think like a Programmer
- Apply Performance and Optimization Methods
- Sources of Information
- Learn to Collaborate on Data Projects

# Today's Objectives

- Discussion of R Features
    - And How They Influence Coding Decisions
- Managing Workflows with Rstudio
- Installing Packages
- Finding Help
- Learning to Write Pseudocode and Outlines
- Course Project Outline

# Intros

# My Background

- Energy Modeling, Pricing, Consumer Web and Restaurant Industry Analytics
- Product Manager, M&A, Technical Program Mgr
- MBA, Marketing (Babson College)
- Bachelors in Engineering (Univ. of Mumbai)

# Why  ?

# Features

- In-Memory Object Storage
- Support for Statistical Packages
- Functional Programming
- Growing Integration Across Data Analysis Value Chain

# Workflow

# Working With RStudio

- Home Folder
- Rstudio Project
- Enclosing folders
  - Code
  - Data
  - Output

# Installing Packages From CRAN

- install.packages("devtools")
- Use Rstudio GUI: ggplot2
- Use command line: dplyr
- Your choice: swirl

# Installing Packages From Github

- install_github("ramnathv/slidify")
- devtools::install_github("ramnathv/slidify")
- library(devtools)
  - install_github("ramnathv/slidifyLibraries")

# Working With Packages

- library(dplyr)
- require(ggplot2)
- (.packages())
- search()
  - searchpaths()
- detach("package:ggplot2", unload=TRUE)

# Getting Help

- ? – Exact match
- help
  - help(pnorm)
  - help(stats)
- library(help = “stats”)
- ?? – Fuzzy match

# Vignettes

- ??vignettes
- vignette(all = FALSE) # All *ATTACHED* Packages
- vignette(all = TRUE) # All *INSTALLED* Packages
- vignette(package=“grid”)
  - vignette(“plotexample”)
- browseVignettes()
- browseVignettes(package="httr”)

# Resources

- Stackoverflow - http://stackoverflow.com/search?q=rstats
- Reddit
  - /r/rstats
  - /r/RStudio
- R Bloggers - http://www.r-bloggers.com/

# Pseudocode

Adapted from materials by Damian Gordon

# Pseudocode

- The general structure of all programs is:

```
PROGRAM <ProgramName>:

<Do stuff>

END.
```

# Sequence

# Sequence

- When we write programs, we assume that the computer executes the program starting at the beginning and working its way to the end.
- This is a basic assumption of all algorithm design.
- We call this SEQUENCE.

# Sequence

- In Pseudo code it looks like this:

```
Statement1;

Statement2;

Statement3;

Statement4;

Statement5;
```

# Sequence

- Or as a program:

```
PROGRAM MonteCarloSimulation:

 Set N = 10,000;

 RAND = Generate N Random Numbers;

 Plot(RAND);

END.
```

# Selection

# Selection

- What if we want to make a choice, for example, we only want ODD Numbers?
- We call this SELECTION.

# Selection

- So, we could state this as:

```
IF (number modulo 2 = 0)
    THEN discard number;
    ELSE store number;
ENDIF;
```

# Selection

- Or, in general:

```
IF (<CONDITION>)
    THEN <Statements>;
    ELSE <Statements>;
ENDIF;
```

# Iteration

# Iteration

- What if we need to tell the computer to keep doing something until some condition occurs?
- Let's say we wish to calculate the mean stock price for the S&P 500.
- We need a loop, or ITERATION.

# Iteration

- So, we could state this as:

```
SET SUM = 0

WHILE (N NOT EQUAL TO 500)

    SUM = SUM + Nth STOCK PRICE

     N = N + 1

ENDWHILE;

AVG = SUM / 500;

END;
```

# Iteration

- Or, in general:

```
WHILE (<CONDITION>)

    DO <Statements>;

ENDWHILE;
```

# Practice

- Write the pseudocode to solve the following:

1. Given demographic details (age, sex, address, salary levels), plot relevant statistics.

2. Given a list of 10 stocks, plot the opening and closing prices over 1 month.

# Building Blocks

# Data Structures (Session 2)

- Data Structures
  - Vectors
  - Matrices and Arrays
  - DataFrames
  - Factors
  - data.table

# Control Structures (Session 3)

- Control Statements
  - *if - else*
  - *switch*
  - *ifelse*
- Loops
  - *for*
  - *while*
  - *nested* (AVOID!)
- Arithmetic and Boolean Operations
  - Compound Statements

# General Operations

- Reading and Writing Data
- Group Manipulation (via *apply* family of functions)
- Data Reshaping
- Modeling
- Visualization
- Documentation

# Assignments

# Assignment 1 - Due 2/5

Write Pseudocode to solve this problem

- Use a random number generator to generate 1000 random numbers. Create three different maximization functions using these numbers as inputs, and plot the maximum of the three values for each input.

# Course Project

http://priceonomics.com/jobs/puzzle/

# Discussion Board

https://www.reddit.com/r/uchicagorprogramming