IST 687 What is Data Science & R Overview

Corey Jackson

2021-01-13 19:27:50

Agenda for 2020-07-01

- A brief mention of data science
- Introductions
- Course overview
- Complete lab assignment 1
- Homework 1 Tips
- Next week's agenda

Welcome to IST 687!!

- ▶ Developed by faculty in the School of Information
- ▶ Taught by a team of instructors (Course designer: Jeff Saltz)

My Contact Information
Corey B. Jackson
Office Hours: 6-7pm ET Wednesday
Slack: via direct message or @Corey Jackson (Instructor)
Email: cjacks04@syr.edu

Data Science

What is Data Science?

Drawing useful conclusions from data using computation

Exploration

- ▶ Identifying patterns in information
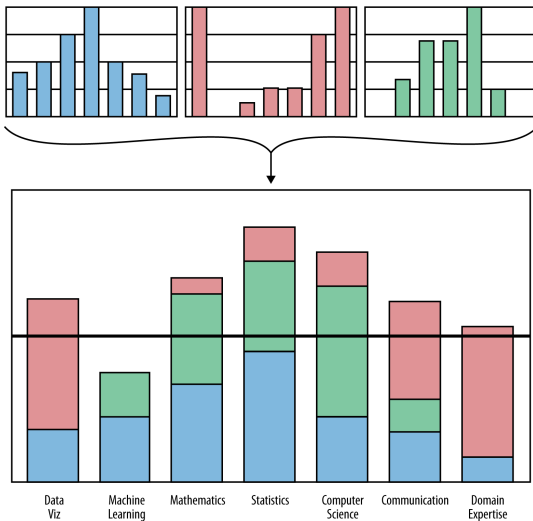- ▶ Uses visualization techniques

Prediction

- ▶ Making informed (or statistical) guess
- ▶ Uses machine learning and optimization

Inference

- ▶ Quantifying our degree of certainty
- ▶ Uses stochastics and statistical decision theory

What skills are needed?

No one person can be the perfect data scientist, so we need teams.

Brief Introductions

Course

Learning R

- ▶ Week 1: What is Data Science & R Overview
- ▶ Week 2: Using R to manipulate data
- ▶ Week 5: Connecting to external data sources

Start your download of the R Studio GUI. You'll also need to install R.

Exploration

- ▶ Week 3: Descriptive statistics and functions
- ▶ Week 6: Introduction to data visualization
- ▶ Week 7: Working with map data

Prediction

- ▶ Week 8: Modeling data
- ▶ Week 9: Association rule mining/support vector machines
- ▶ Week 10: Text mining

Inference

- ▶ Week 4: Inferential statistics

- Asynchronous lectures on the LMS should be viewed prior to live meetings
- Synchronous meetings with a focus on completing lab assignments
  - Intended to provide a space to ask questions, obtain immediate feedback, and practice coding
  - Lab Assignments are central to synchronous sessions and should be completed in-class
  - Each person will work with a peer (pair programming) in a breakout room to complete the weeks lab assignment.

Components of IST 687

- Assesments/Grading
  - Homework (30%)
  - Mid-term exam (30%)
  - Final Project (30%)
  - Class Participation (10%)

Required text: **Saltz, Jeffrey S and Jeffrey M. Stanton. Introduction to Data Science. SAGE Publications, 2016. (Free PDF / Amazon)**

Course Materials

- The Learning Management System (LMS)
- Your section's Syllabus

Course Collaboration/Communication

- All communucation will happen via SLACK. Join the workspace via this link: #jackson-687-winter21
- Programming exercises completed in R programming language using RStudio Graphical User Interface
- Compiling code for lab and homework submissions using RMarkdown.
- For weekly pair programming and group project assignments view this document: Group Assignments

Lab 1: Basic R coding (vectors, conditionals)

Goal: Write a piece of R code using logic and relational operators.

Steps : (1) Brainstorm a scenario that could could use to implement boolean logic. (2) Find the appropitate operator notations (3) Write three conditional statements that will evaluate the logical operators: AND,OR, and NOT

Lab 1 Demo

An example. . .

Sorting my sheep into pastures based on their fleece colors

```
sallysheep = "black"
mollysheep = "white"
billysheep = "black"

pasture1 = "black"
pasture2 = "brown"
```

The expression `sallysheep = "black"` assigns the value 'black' anytime sallysheep is referenced in the command line/code. '<-' can also be used to assign values.

Is the variable stored in sallysheep the same as the variable in pasture 1?

```
sallysheep == pasture1
```

```
## [1] TRUE
```

Here, == is a "relational operator" that means "exactly equal to."

Other relational operators

Relational Operators in R

Operator	Description
<	Less than
>	Greater than
<=	Less than or equal to
>=	Greater than or equal to
==	Equal to
!=	Not equal to

We could even add multiple expressions to and use a logical operator to carry out Boolean operations

The AND syntax in R is the ampersand (&) e.g., `x & y`

```
sallysheep == pasture1 & billysheep == pasture2
```

```
## [1] FALSE
```

Note: Logical AND (&) implies that both test comparisons need to be TRUE for the test statement to be TRUE. Thus, a test comparison of `sallysheep == pasture1 & billysheep == pasture2` would return FALSE. In this case, `billysheep == pasture2` is FALSE.

Other logical operators

Operator	Description
!	Logical NOT
&	Element-wise logical AND
&&	Logical AND
|	Element-wise logical OR
||	Logical OR

Next, we can print statements based on whether our test expression returns TRUE or FALSE

The basic syntax for an if-else statement in R is

```
if(test_expression)
  print("statement1") else
    print("statement2")
```

Evaluating our test expression and providing feedback

```
if(sallysheep == pasture1)
print("The fleece and pasture match") else
  print("The fleece and pasture do not match")

## [1] "The fleece and pasture match"
```

Evaluating our test expression and providing feedback

We can also evaluate mutiple test expressions

```
if(mollysheep == pasture1 & mollysheep == pasture2)
  print("Fleece and pasture colors match") else
    print("At least one fleece color does not match the pasture")

## [1] "At least one fleece color does not match the pasture"
```

Lab Instructions

- ▶ Work with your partner to complete the lab
- ▶ A link to the .rmd can be found in the syllabus under week 1 section
- ▶ Be sure to include the names of all group members in the YAML header
- ▶ *Need help???* ping me on Zoom

Scenarios to consider:

- ▶ Sorting groceries for pizza into cabinet or refrigerator
- ▶ Evaluating cars and sorting them into "Asian","European", or "American" brands
- ▶ Grading student assignments

Homework 1 Tips

Working with Vectors

HW 1 is due at 11:59 on 2020-07-06. Please post questions to SLACK

Next Week

Agenda for 2020-07-08

Asynchronous Session

- ▶ Submit Lab 1 and Homework 1 via the LMS by 11:59 on 2020-07-06
- ▶ Documents must be submitted as either a .docx, .pdf, or .html (**not nb.html**)
- ▶ Watch week 2 async videos and read Chapters 4-6 in Saltz and Stanton book
- ▶ Begin looking for datasets and consider the insights you want to obtain from the project

Synchronous Session

- ▶ More course information
- ▶ Overview of Week 1
- ▶ Lab 2: Data frames & sorting
- ▶ Meet with your project team