ACES: Fundamentals of R Programming

HPRC Training

27 February 2024

High Performance Research Computing

DIVISION OF RESEARCH
Course Outline

1. Accessing ACES
2. Launching RStudio
3. Data Structures in R
4. Introduction to Tidyverse
5. Importing and Exporting Data
6. Regression
7. Principal Component Analysis
8. Data Visualization with ggplot2
Accessing the HPRC ACES Portal

HPRC webpage: hprc.tamu.edu
Accessing ACES via the Portal (ACCESS)

Log-in using your ACCESS credentials.

Select the Identity Provider appropriate for your account.
Launching RStudio on ACES
Set number of hours to 7
Set number of cores to 1
Click “Launch” once the correct parameters have been selected
Click this button when it says “Connect to RStudio Server” (this will take a minute)
RStudio Interface

**Console**: Allows users to input R commands directly

**Terminal**: Allows shell access to the node

**Plots**: Displays user-generated graphs/figures

**Packages**: Load and install packages

**Help**: Access help pages for functions and packages

**File Browser**: Allows users to interact with file system
Accessing the Course Materials

- Open the terminal tab in the portal (Clusters > ACES_shell_acces) and run the following commands:

```bash
[user@aceslogin2 ~]$ cp -r /scratch/training/DataScienceR/ .
[user@aceslogin2 ~]$ cd DataScienceR/
[user@aceslogin2 ~]$ ls
```

The output should read:

```
carnivores.csv  carnivores.xlsv  datascience.xlsv  IntroductionToDataScienceInR.Rmd
```
Accessing the Course Materials

- Next, in the upper left corner, select File > Open File
Accessing the Course Materials

- Navigate to the DataScienceR directory and select "IntroductionToDataScienceInR.Rmd"
Accessing the Course Materials

- Select “Run Document” from the toolbar to launch the workbook.
Accessing the Course Materials

- If you see a dialog box that says “Popup Blocked” click “Try Again”, and the workbook should open in a new tab.
Transition to Workbook

Introduction to Data Science in R

Data Structures
R has numerous data structures, many of which are commonly used or encountered in data science applications or workflows. Here we provide a brief overview of the data structures that we will be using in this notebook and some of the common methods used when working with them.

<table>
<thead>
<tr>
<th>Data Structure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vector</td>
<td>A 'list' or array of elements of the same fundamental data type (e.g., logical, numeric, character).</td>
</tr>
<tr>
<td>List</td>
<td>A collection of elements of different data types.</td>
</tr>
<tr>
<td>Matix</td>
<td>A group of elements of the same data type (just like vectors) arranged into a set number of rows and columns.</td>
</tr>
<tr>
<td>Data Frame</td>
<td>A group of elements arranged as a table or two-dimensional array that can contain heterogeneous data.</td>
</tr>
<tr>
<td>Tabble</td>
<td>A 'modern' re-formatted version of the data frame structure (see below).</td>
</tr>
</tbody>
</table>

Viewing Data Objects
Many common operations and functions are shared between different data structures. The functions `print()`, `head()`, and `str()` allow us to examine the contents of a given object. Use the code chunk below to see how these built-in functions work. We will be using some built-in data sets and previously generated data objects for this exercise.

```
R Code
# Identify the structure of the 'heights' object
str(iris)

# Display the first components of the 'heights' object
head(iris)

# Print all of the elements within the 'heights' object
print(iris)
```

Independent Exercise
Use the code chunk below to try the built-in functions we just learned on the built-in data set `mtcars`.

```
R Code
```
Need Help? Contact the HPRC Helpdesk

Website: hprc.tamu.edu
Email: help@hprc.tamu.edu
Phone: (979) 845-0219

Help us help you -- we need more info
- Which Cluster (ACES, FASTER, Terra, Grace)
- NetID (NOT your UIN)
- Job id(s) if any
- Location of your jobfile, input/output files
- Application used, if any
- Module(s) loaded, if any
- Error messages
- Steps you have taken, so we can reproduce the problem