

# HIGH PERFORMANCE RESEARCH COMPUTING

## ACES: Fundamentals of R Programming

HPRC Training

27 February 2024



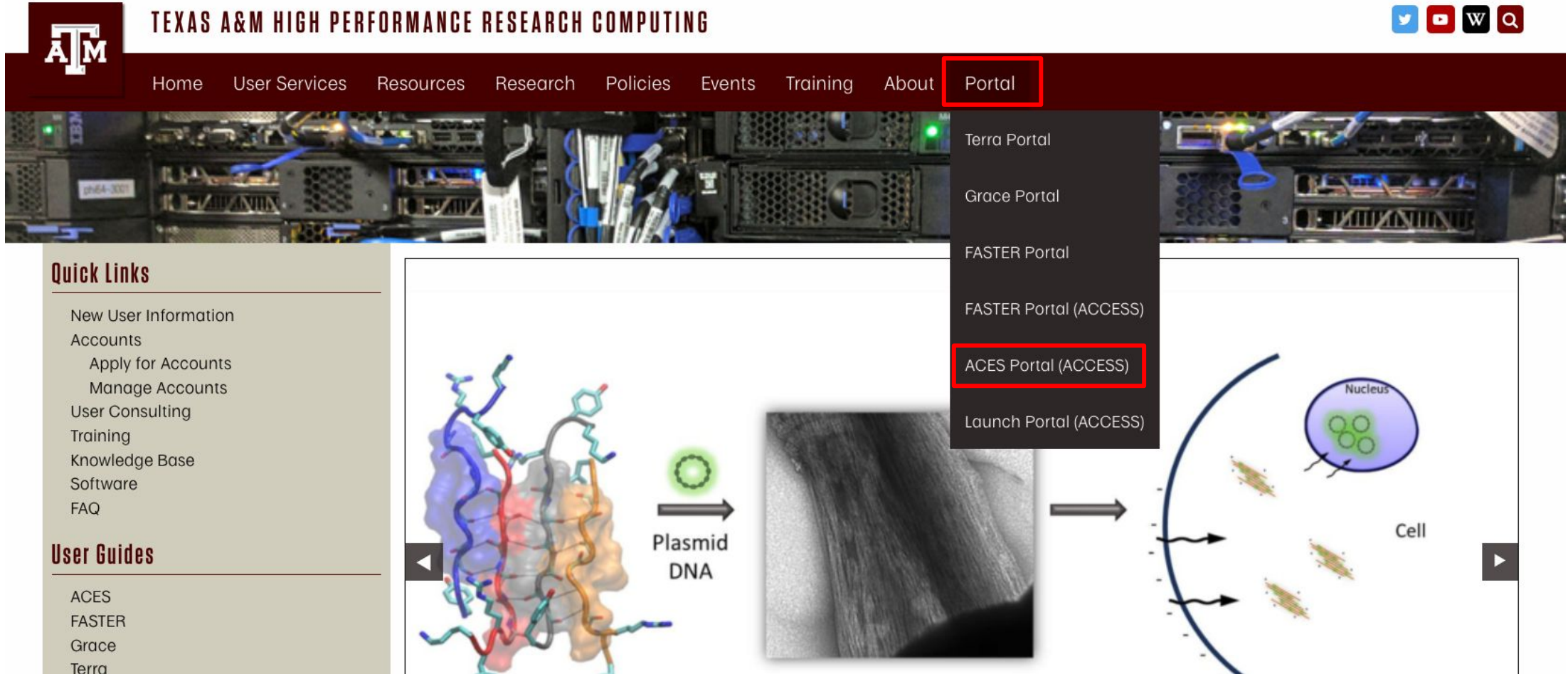
High Performance  
Research Computing  
DIVISION OF RESEARCH



# 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



The screenshot shows the Texas A&M High Performance Research Computing (HPRC) website. The top navigation bar includes links for Home, User Services, Resources, Research, Policies, Events, Training, About, and Portal. The Portal link is highlighted with a red box. A dropdown menu is open from the Portal link, listing several options: Terra Portal, Grace Portal, FASTER Portal, FASTER Portal (ACCESS), ACES Portal (ACCESS) (highlighted with a red box), and Launch Portal (ACCESS). Below the navigation bar, there is a banner image of server racks. On the left side, there are sections for Quick Links and User Guides. The Quick Links section includes: New User Information, Accounts (Apply for Accounts, Manage Accounts), User Consulting, Training, Knowledge Base, Software, and FAQ. The User Guides section includes: ACES, FASTER, Grace, and Terra. In the center, there is a diagram illustrating a biological process. It shows a 3D molecular model of a protein structure on the left, with a green circular plasmid DNA molecule being introduced. An arrow labeled 'Plasmid DNA' points to a grayscale micrograph of a cell. Another arrow points to a diagram of a cell with a nucleus and several green circular plasmids inside, labeled 'Cell'.

HPRC webpage: [hprc.tamu.edu](http://hprc.tamu.edu)

# Accessing ACES via the Portal (ACCESS)

Log-in using your ACCESS credentials.

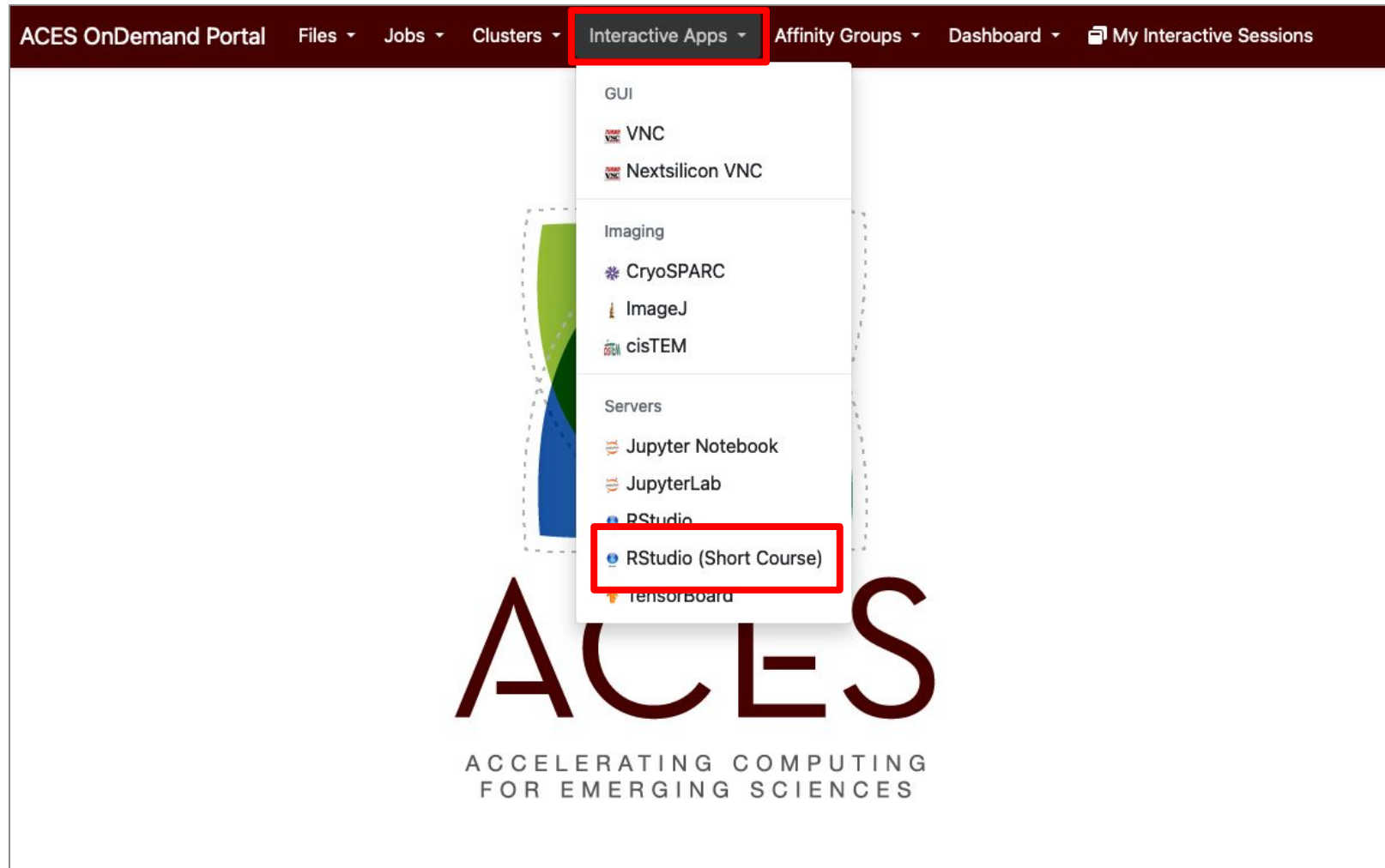
The screenshot shows the ACCESS portal interface. At the top left is the ACCESS logo, and at the top right is the 'Powered By CILogon' logo. A teal banner at the top reads 'Consent to Attribute Release'. Below this, a white box contains the text: 'TAMU FASTER ACCESS OOD requests access to the following information. If you do not approve this request, do not proceed.' followed by a bulleted list: 'Your CILogon user identifier', 'Your name', 'Your email address', and 'Your username and affiliation from your identity provider'. Below the list is a teal header 'Select an Identity Provider' and a dropdown menu showing 'ACCESS CI (XSEDE)'. A red box highlights this dropdown. Below the dropdown is a 'Remember this selection' checkbox and a 'Log On' button. At the bottom of the white box, it says 'By selecting "Log On", you agree to the [privacy policy](#).' At the very bottom of the page, there is a footer with links for 'FAQs', 'help@cilogon.org', and 'using the CILogon Service'.

The screenshot shows the ACCESS portal login page. At the top left is the ACCESS logo, and at the top right is the CILogon logo. The main heading is 'Login to CILogon'. Below this are two input fields: 'ACCESS Username' and 'ACCESS Password'. A checkbox labeled 'Don't Remember Login' is below the password field. A teal 'Login' button is at the bottom. To the right of the input fields, there is a teal box with the CILogon logo and the text 'CILogon facilitates secure access to CyberInfrastructure (CI)'. Below this, there are several links: 'If you had an XSEDE account, please enter your XSEDE username and password for ACCESS login', 'Register for an ACCESS Account', 'Forgot your password?', and 'Need Help?'. At the bottom of the page, there is a link 'Click Here for Assistance'.

This is a close-up of the 'Select an Identity Provider' dropdown menu. The dropdown is open, showing the selected option 'ACCESS CI (XSEDE)' with a small arrow pointing left and a question mark icon to its right. The entire dropdown area is enclosed in a red rectangular box.

Select the Identity Provider appropriate for your account.

# Launching RStudio on ACES



**Interactive Apps**

GUI

- VNC
- Nextsilicon VNC

Imaging

- CryoSPARC
- ImageJ
- cisTEM

Servers

- Jupyter Notebook
- JupyterLab
- RStudio**

**Interactive Apps [Sandbox]**

- Imaging
- CryoSPARC 4

**RStudio version: 2023.06.1-524**

This app will launch RStudio using the R\_tamu software module on an ACES compute node.

You can install your own R packages directly within RStudio.

R version

4.2.2

Number of hours (max 168)

7

Number of cores (max 96)

1

Total GB memory (max 480)

12

Email

Email address must be provided if the checkbox for email notification is checked (see below).

I would like to receive an email when the session starts

Slurm account (optional)

This field is needed only if you want to use a different account other than your default account. Leave it blank if you don't know what to provide.

**Launch**

\* The RStudio session data for this session can be accessed under the [data root directory](#).

Set number of hours to 7

Set number of cores to 1

Click "Launch" once the correct parameters have been selected

ACES OnDemand Portal Files Jobs Clusters Interactive Apps Dashboard My Interactive Sessions </> Develop Help

Session was successfully deleted. x

Home / My Interactive Sessions

**Interactive Apps**

- GUI
- VNC
- Nextsilicon VNC
- Imaging
- CryoSPARC
- ImageJ
- cisTEM
- Servers
- Jupyter Notebook
- JupyterLab
- RStudio

**Interactive Apps [Sandbox]**

- Imaging
- CryoSPARC 4

**RStudio (5427)** 1 node | 1 core | Running

Host: ac073 Delete

Created at: 2023-09-19 13:41:50 CDT

Time Remaining: 6 hours and 57 minutes

Session ID: 3afecef4-42f0-4596-9114-0a82c23caea1

Connect to RStudio Server

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

**Files:** File Browser: Allows users to interact with file system

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

**Packages:** Load and install packages

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



# Accessing the Course Materials

- Open the terminal tab in the portal (Clusters > ACES\_shell\_acces) and run the following commands:

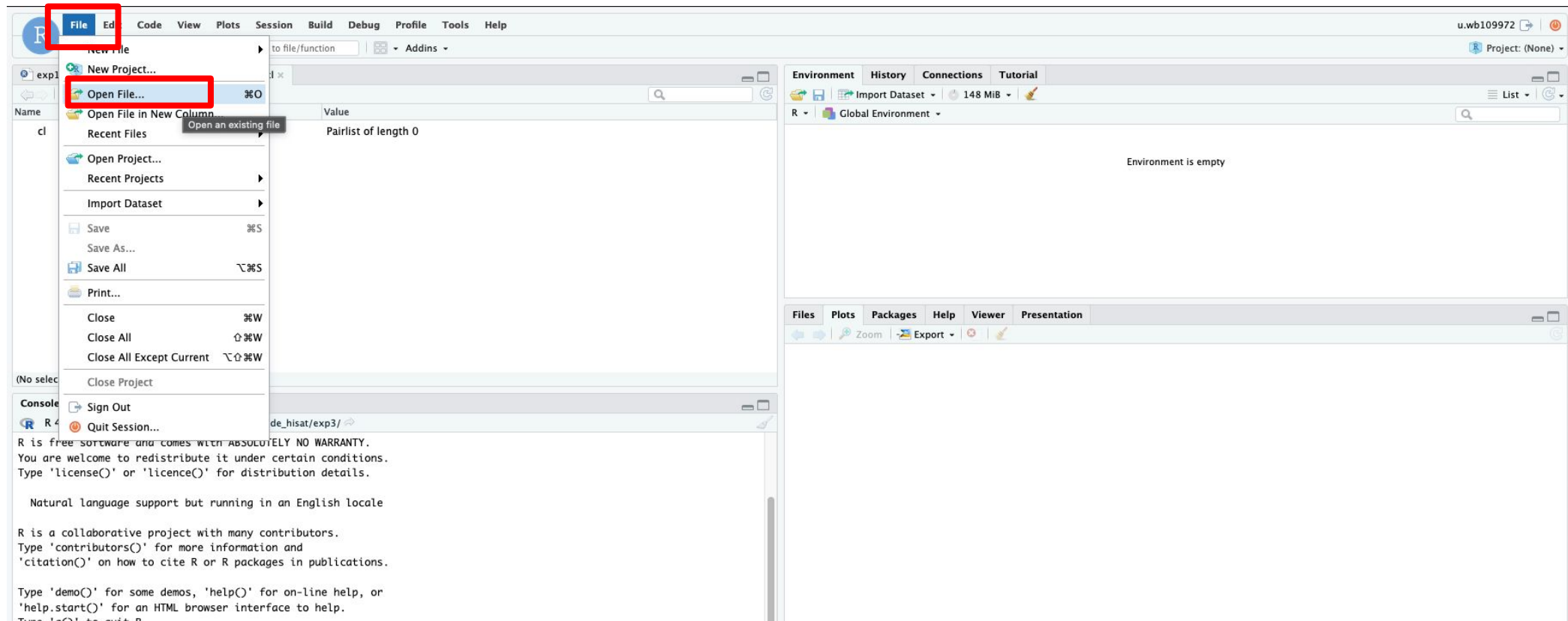
```
[user@aceslogin2 ~]$ cp -r /scratch/training/DataScienceR/ .  
[user@aceslogin2 ~]$ cd DataScienceR/  
[user@aceslogin2 ~]$ ls
```

The output should read:

```
carnivores.csv  carnivores.xlsx  datascience.xlsx  IntroductionToDataScienceInR.Rmd
```

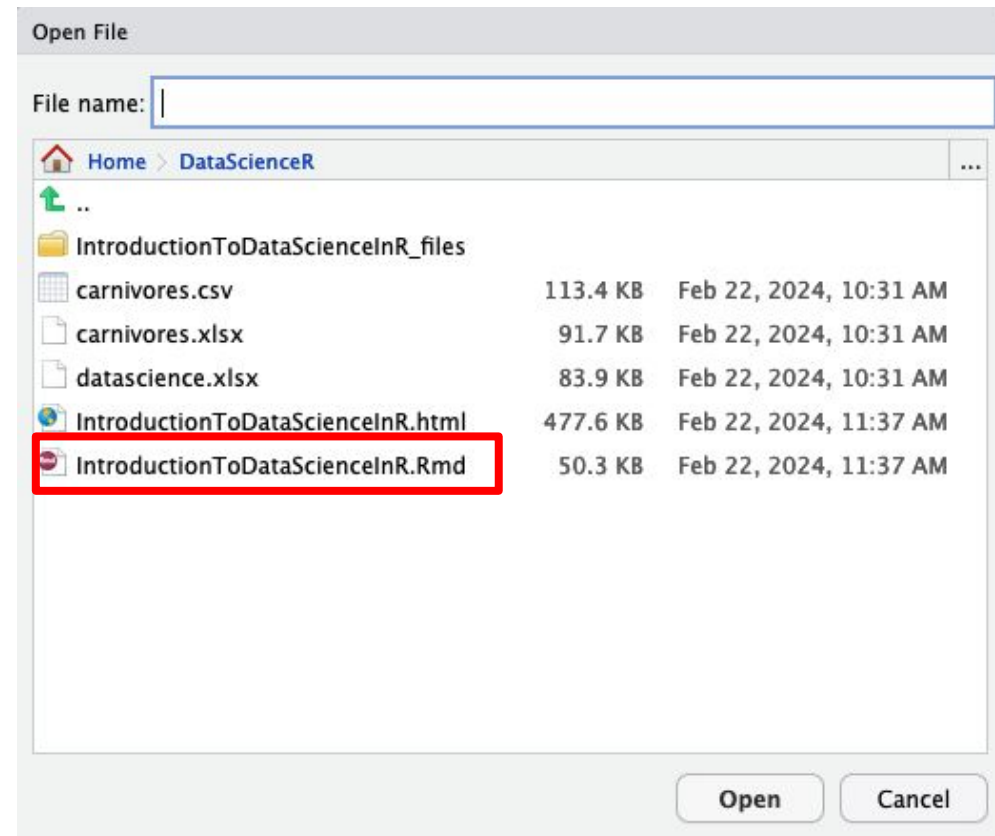
# Accessing the Course Materials

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



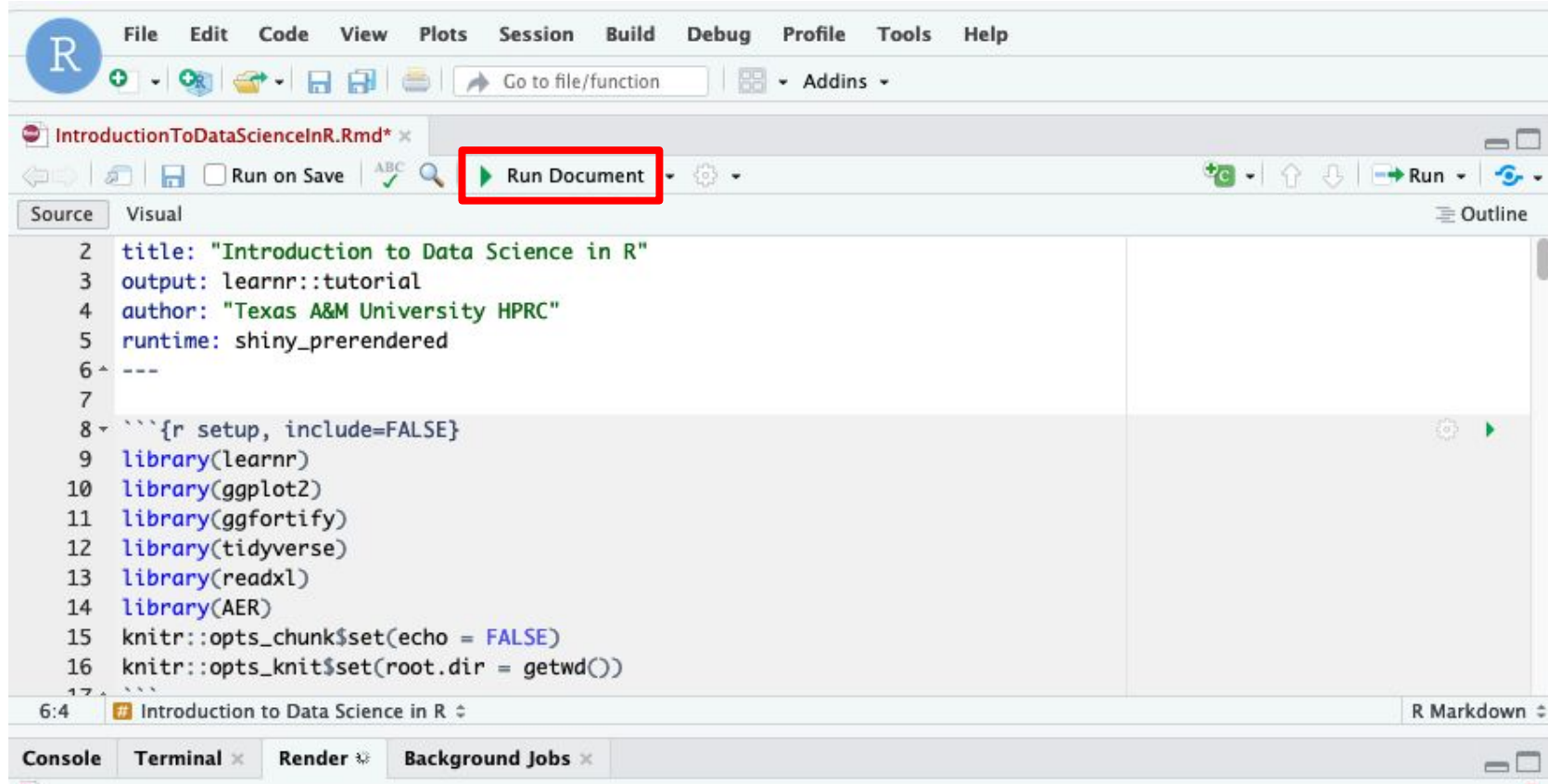
# Accessing the Course Materials

- Navigate to the DataScienceR directory and select “IntroductionToDataScienceInR.Rmd”



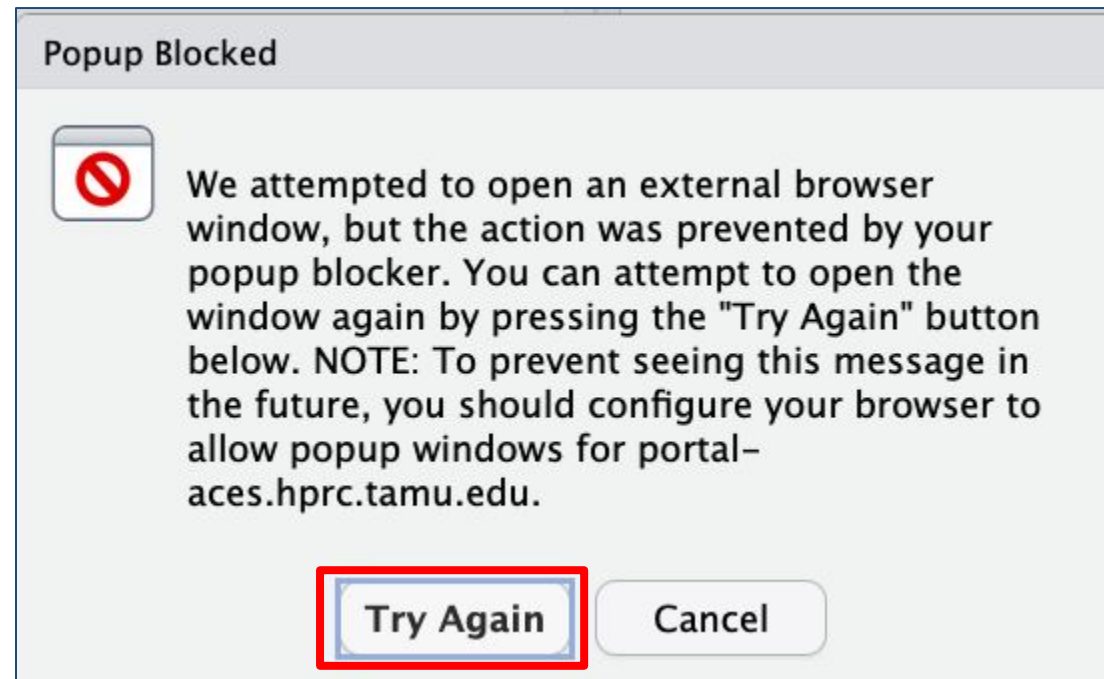
# Accessing the Course Materials

- 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

dplyr

Importing and Exporting Data

Regression

Principal Component Analysis

Data Visualization with ggplot2

Texas A&M University HPRC

Start Over

## 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.

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

## 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  
```

```
1 # Identify the structure of the 'heights' object
2 str(iris)
3
4 # Display the first components of the 'heights' object
5 head(iris)
6
7 # Print all of the elements within the 'heights' object
8 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   
```

```
1
```

# Need Help? Contact the HPRC Helpdesk

Website: [hprc.tamu.edu](http://hprc.tamu.edu)

Email: [help@hprc.tamu.edu](mailto: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