OSOS 2020 Participant guide

September 12th and 13th, 2020 9am to 4pm Central Standard Time (CST)

Organized by the TAMU Ecology & Evolutionary Biology Program

Emergency Contacts: Nick Jacobsen, Coordinator (njacobsen@tamu.edu); Giri Athrey, Outreach Chair (giri.athrey@tamu.edu).

Table of Contents

1. General notes about the event

Thank you for registering for the 2020 edition of the Open Source Open Science workshop. The entire workshop will be delivered live via Zoom. This document gives an overview of the program, as well as a guide to preparing your computer to participate and maximize learning at the workshop. Two sections are outlined below: the "**Before workshop**" and "**During workshop**" resources. It is critical that you read this guide carefully as you prepare for the workshop.

2. Schedule at a Glance

Session	Time	Topic	Instructor
Session 1, Saturday	9:00-10:30am	Intro to R	Dr. Danielle Walkup
	10:30-10:45am	Bio Break	
	10:45-12:00pm	Intro to R continued	Dr. Danielle Walkup
BREAK	12:00-1:00pm	Lunch	
Session 2, Saturday	1:00-2:30 pm	Analysis & Visualization in R	Dr. Carl Hjelmen
	2:30-2:45pm	Bio Break	
	2:45-4:00pm	Analysis & Visualization in R	Dr. Carl Hjelmen
Session 3, Sunday	9-10:30am	Data exploration in R	Dr. Michelle Mellers
	10:30-10:45am	Bio Break	
	10:45-12:00pm	Data exploration in R	Dr. Michelle Mellers
BREAK	12:00-1:00pm	Lunch	
Session 4, Sunday	1:00-2:30 pm	Data analysis in Python	Mr. Matthew Hielsberg
	2:30-2:45pm	Bio Break	
	2:45-4:00pm	Data analysis in Python	Mr. Matthew Hielsberg

3. Preparing your computer before attending OSOS

All the sessions are designed to be practical, hands-on activities. Therefore, you need to bring your own laptop or desktop computer. All the workshop sessions will utilize open source tools (R, Python) and hence, have cross platform compatibility. Due to the virtual workshop format, our ability to provide hands-on assistance is limited. Therefore, we are providing separate instructions for **Beginner** and **Intermediate** level participants. We also suggest that having two screens may be optimal for following the Zoom stream while working on the data examples. Alternatively two devices - one for streaming and one for working may work equally well.

💡 **Beginner:** Little or no prior experience with R or Python
Intermediate: Some prior experience with R or Python, including use of Integrated Development Environment (IDE) or Python notebooks (eg. Jupyter).

3.1 Preparing for R sessions

Cloud-Based Beginner Level

RStudio is a well developed and powerful development environment (IDE) for R. IDEs are software that allow us to work on code in an interactive environment with code editing, console, file management, and visualization capabilities. Typically R and RStudio are installed locally on your computer (see Intermediate instructions if you want to do so). To avoid platform dependent issues that may arise in this virtual workshop, we **strongly recommend** that you use the cloud version of RStudio - called RStudio Cloud. To use this service, you will need to create a free account and have your login information ready before the workshop.

1. Go to RStudio Cloud Free and sign up using your email ID, login through your Google account or Github account.

2. Once you create your account and log in, please carefully read the Guide to understand the layout and functionality.

3. Instructions are provided in **Section 4.2** on how to upload your data files to RStudio Cloud

4. During the workshop you will be installing and loading packages needed for the analyses and visualization.

5. OPTIONAL: The free version of RStudio Cloud gives you 15 hours per month with 1GB RAM, and 1CPU. If you think you will use more than 15 hours of RStudio during the workshop, please consider the R and RStudio local installation (free) or upgrade to an academic version of RStudio Cloud ($$).

Intermediate Level

Download R: Go to the website (www.r-project.org) and select the link "download R" and you will be sent to a site with a list of CRAN Mirrors. Select the mirror nearest to you and you will be sent to the download page. Select the download link that is appropriate for your computer platform and follow the instructions on the following web pages. After you download the appropriate package, make sure to install R on your computer, by unpacking the download archive. Once installed, clicking the 'R' logo will open the console.

Download RStudio: RStudio is an Integrated Development Environment (IDE) that pack- ages file managing, code editing, and visualizer built in with the console. Go to the website (www.rstudio.com/products/rstudio/download/) and download RStudio Desktop with the Open Source License for free. Be sure to select the download link that matches your computer platform and install the program after download.

During the workshop you will be installing and loading packages needed for the analyses and visualization.

3.2 Preparing for the Python session

We will use Jupyter Notebooks for the Python session. The Juypter project is an open source project to support data science education and project management. Jupyter Notebook is a web-based technology, meaning that all interaction with the note-books occurs within your web browser (e.g. Firefox, Chrome, Safari, IE). A Jupyter Notebook server is used to serve the notebook content. The server can be run locally on your machine by installing it directly from source or through some other distribution (e.g. Anaconda), or the server can be run in the cloud via a service (e.g. Google Colab).

Cloud-Based Beginner Level

For beginners, we **strongly recommend** the use of Google Colab, as it requires the least effort to get up and running, and it supports all the features we wish to use during the session. Google Colab does not require any installation, and you can view notebooks without a Google account. However, Google Colab does require that you have a Google account in order to run any code within the notebooks as it was designed to work together with Google Drive. To load the data files during the session you will need to be logged in to your Google Account, and we will discuss how to connect Colab to your Google Drive.

1. If you do not have a Google account then you may sign up for a free account at http://www.google.com (Click "Sign In" in the top right corner, then click "Create Account"). The Google account should give you free access to both Google Colab and Google Drive.

2. To access Google Colab go to https://colab.research.google.com/ and log in with your Google Account (Click the "Sign In" button in the top right corner). This should load the "Welcome to Colaboratory" notebook and show the open dialog where you can access additional notebooks on Google Drive, GitHub or your local system.
The open dialog can be accessed again through File > Open notebook. We encourage participants to try this out prior to attending the session, however we will briefly describe some of the Colab interface at the beginning of the session.

3. Upload the Python Notebooks (.pynb) and the data files (.csv) that you can find in the downloads section (4.1) to your Google Drive.

4. Within the Colab open dialog select the "Google Drive" tab at the top of the dialog. This will display any notebooks you have in the Colab Notebooks folder in your Google Drive account. The Colab Notebooks folder is created for you automatically the first time you save a notebook in Google Colab (or when you upload a notebook). Select the notebook from the list to open it.

5. From within Google Drive in your browser you can navigate to an .ipynb file, and use right- click to select the "Open with ¿" option. From the menu that pops up select "Google Colaboratory" to open the notebook in Google Colab directly.

6. Thorough documentation on the installation can be found here and also inside the Python session download (Section 4.1)

Intermediate Level

For more advanced users who would prefer a local install of Jupyter Notebooks, we recommend using the Anaconda based installation documented on the Jupyter site here. However, be advised that this current workshop tutorial and command has been developed for Google Colab, and will not work with a local Jupyter notebook without adjustment. Therefore we recommend you follow this workshop with Google Colab and try out local installations of Jupyter for your use at a later time. Thorough documentation on the installation can be found here and also inside the Python session download (Section 4.1)

4. Resources for attending the workshop

Zoom call details

Sep 12, 2020 09:00 AM Central Time (CST)

Sep 13, 2020 09:00 AM Central Time (CST)

Join Zoom Meeting

https://tamu.zoom.us/j/96912351490?pwd=TTRqcG9GQ0g3OEJzR01mNDlBZHNQQT09

Meeting ID: 969 1235 1490

Passcode: 684086

One tap mobile

+13462487799,,96912351490# US (Houston)

+12532158782,,96912351490# US (Tacoma)

Dial by your location

+1 346 248 7799 US (Houston)

+1 253 215 8782 US (Tacoma)

+1 669 900 6833 US (San Jose)

+1 929 436 2866 US (New York)

+1 301 715 8592 US (Germantown)

+1 312 626 6799 US (Chicago)

877 853 5247 US Toll-free

888 788 0099 US Toll-free

Meeting ID: 969 1235 1490

Find your local number: https://tamu.zoom.us/u/abNQazGQBr

4.1 Session Specific Data and Code Downloads

Please download the documentation, data-files and code for each of the sessions before joining the session. Session participants will be led through the installation of necessary packages at the beginning of the respective sessions.

Session Specific downloads

Aa Session	☰ Download Link
Session 1, Saturday AM	Session 1 Files
Session2, Saturday PM	Session 2 Files
Session 3, Sunday AM	Session 3 Files
Session 4, Sunday PM	Session 4 Files

4.2 Loading data into RStudio Cloud

As RStudio cloud works on a remote server, you will need to upload your data and code files into the environment to access and work with them. Here are the steps to upload your files into RStudio Cloud.

1. Locate the 'Files' window on the RStudio interface
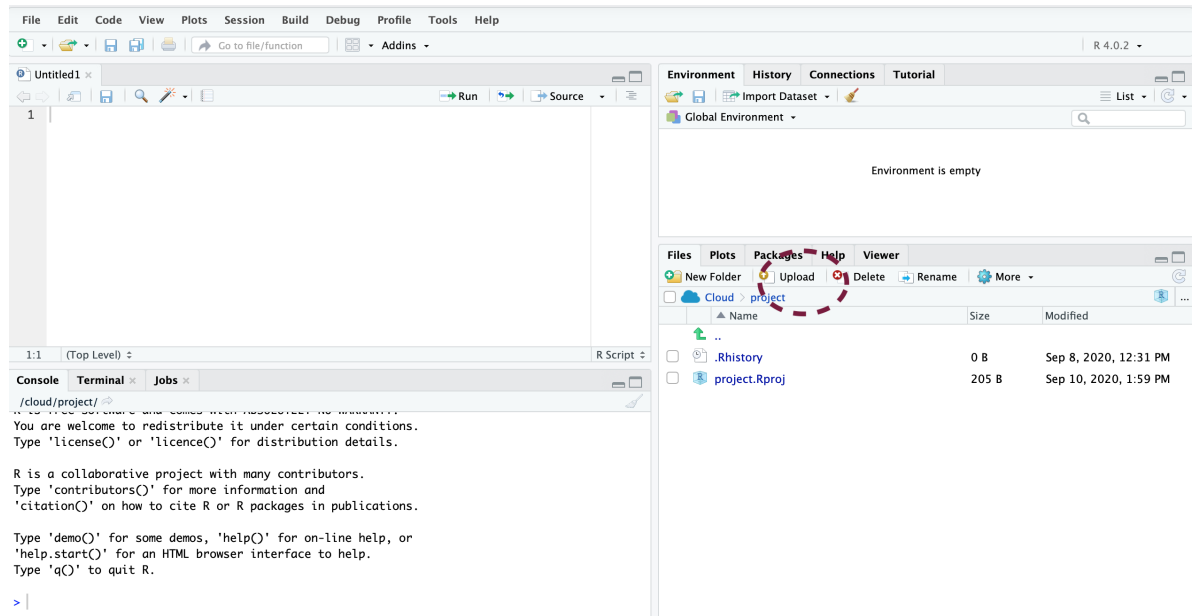
2. Select the 'Upload' option and find the downloaded .zip folder for the session you are in.
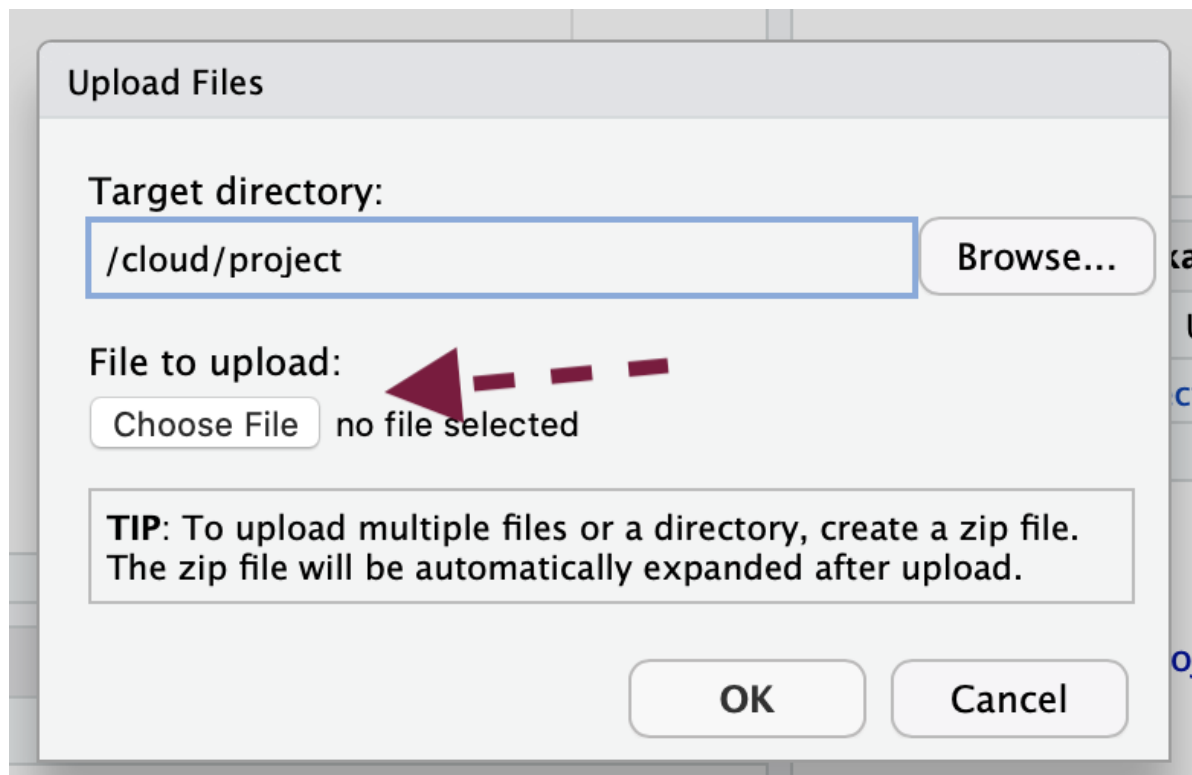
3. Select the .R script or R Markdown (.Rmd) file and double-click to open it up in our environment.

OSOS-2020.pdf	666.1 KB	Sep 10, 2020, 2:05 PM
OSOS-2020.Rmd	38.5 KB	Sep 10, 2020, 2:05 PM
OSOS-2020.tex	99.1 KB	Sep 10, 2020, 2:05 PM
rmdFigs		
Rscripts		

5. Seeking help when you hit trouble

We expect that all participants **will** encounter some roadblocks during this workshop, and broadly speaking, while attempting to learn new data science skills with R or Python. This is normal, and we ask you to stick with the task. In this online format, we expect that some issues may seem more frustrating, but we ask you to stay patient, and wait to be helped. To help everyone stay on track, we are using two approaches.

1. **SLACK CHANNEL:** Using a dedicated Slack Channel to seek help. If you are not a Slack user, please register and download the Slack app and join the 'osos2020' channel. **Click here to join this Slack channel. IMPORTANT:** Please join the Slack channel before the workshop as it will take a few minutes to get this set up. You may ask questions at any time and one of the instructors or helpers will try to answer your question. **The chat function in Zoom will be disabled, so all questions are to be posted on Slack.**

2. **BREAKOUT ROOMS:** The instructors will pause the sessions at frequent intervals to open up 'Break Out' rooms. You will be randomly assigned to a breakout room where you can work with your random group to obtain peer-to-peer assistance or use the 'Ask for Help' button to call for help. The instructor or one of the volunteer helpers will join your breakout room to help.

We have a large number of participants registered and to ensure that the session moves smoothly, please pay attention to the most common sources of errors in coding. The most common issues to expect in these workshops are:

1. Spelling and capitalization errors in code

2. Extra spaces or characters in code

3. missing quotes, parentheses, or square brackets

If none of these explain your errors, please seek help by asking for help.

Non-technical questions may be directed to contacts listed at the top of the document.

6. Code of Conduct

1. This is an open source workshop sponsored by Texas A&M University. We welcome all learners, no matter where you are from, your background, your training or expertise level. However, we ask you to be respectful to all the instructors and co-participants. We reserve the right to expel any individual whose behavior runs afoul of common standards of decency.

2. You are free to use the material provided in this workshop for your own education or for teaching purposes. Please contact the EEB coordinator if you have questions about reuse of the data or resources provided in this workshop.