

SATVI Computational Course

Session Guide

SATVI Computational Group

05 MAR 2024

Table of contents

| | |
|--|-----------|
| Preface | 4 |
| 1 Introduction | 5 |
| 1.1 Instructor contacts | 5 |
| 2 Syllabus | 6 |
| 2.1 Module 1: Intro to R and MaRcus Training Course | 6 |
| 2.1.1 Session 1: Intro to R and swirl | 6 |
| 2.1.2 Session 2: MaRcus Training Course lesson 1 | 6 |
| 2.1.3 Session 3: MaRcus Training Course lesson 2 | 7 |
| 2.1.4 Session 4: MaRcus Training Course lesson 3 | 7 |
| 2.1.5 Session 5: MaRcus Training Course lesson 5 | 7 |
| 2.1.6 Session 6: MaRcus Training Course lesson 6 | 8 |
| 2.1.7 Session 7: MaRcus Training Course lesson 7 | 8 |
| 2.1.8 Session 8: Exporting data from R | 8 |
| 2.2 Module 2: Quarto, GitHub, and GUIs | 9 |
| 2.2.1 Intro to Quarto | 9 |
| 2.2.2 Intro to GitHub | 9 |
| 2.2.3 Intro to VS Code | 9 |
| 2.3 Module 3: Statistics | 10 |
| 2.3.1 Basic statistical tests | 10 |
| 2.3.2 Correlations | 10 |
| 2.4 Module 4: Commonly needed analyses for Immunology | 10 |
| 2.4.1 Heatmaps | 10 |
| 2.4.2 Dimensionality reduction | 11 |
| 2.4.3 Receiver operating characteristic (ROC) curves | 11 |
| 2.4.4 Background subtraction | 11 |
| 2.4.5 Basic flow cytometry analysis | 12 |
| 2.4.6 Automatic gating | 12 |
| 2.5 Module 5: Other coding languages | 12 |
| 2.5.1 Intro to Python | 12 |
| 3 Installations | 13 |
| 3.1 R | 13 |
| 3.2 R Studio | 13 |

| | | |
|----------|--|-----------|
| 3.3 | GitHub Desktop | 13 |
| 3.4 | Visual Studio Code (VS Code) | 14 |
| 4 | swirl | 15 |
| 4.1 | Install swirl | 15 |
| 4.2 | Initialize swirl | 15 |
| 4.3 | Install an interactive course | 16 |
| 4.4 | Run swirl | 16 |
| 4.5 | Exit swirl | 16 |
| 4.6 | Interactive commands | 17 |
| 4.7 | Homework | 17 |
| 4.8 | FAQ | 17 |
| 5 | MaRcus R Training | 19 |
| 5.1 | Content access | 19 |
| 5.2 | Homework | 19 |
| 5.3 | FAQ | 20 |
| 6 | Session Recordings | 23 |
| 6.1 | 2024 Session Recordings | 23 |
| 6.1.1 | Session 1: Intro to R and swirl | 23 |
| 6.1.2 | Session 2: MaRcus Training Course lesson 1 | 23 |
| 6.1.3 | Session 3: MaRcus Training Course lesson 2 | 23 |
| 6.1.4 | Session 4: MaRcus Training Course lesson 3 | 23 |
| 6.1.5 | Session 5: MaRcus Training Course lesson 4 | 24 |
| 6.1.6 | Session 6: MaRcus Training Course lesson 5 | 24 |
| 7 | Summary | 25 |
| | References | 26 |

Preface

This is a session guide book for the SATVI Computational Course.

This is a version-controlled living document that will be updated as needed as the course progresses. All changes are tracked using git.

1 Introduction

Welcome to the SATVI Computational Course! This course is designed to strengthen fundamental coding skills for SATVI trainees and staff. The curriculum will take you through the basics of R, using the terminal, creating and using git controlled projects, as well as more advanced data analysis methods commonly used at SATVI.

All lessons will be stored on the SATVI GitHub under the repository SATVI_ComputationalCourse.

To access all relevant course content, navigate to https://github.com/SATVILab/SATVI_ComputationalCourse.

A static webpage version of the course is also available at https://satvilab.github.io/SATVI_ComputationalCourse.

Your instructors are SATVI members with experience in each topic. For session-specific questions, please contact the relevant instructor:

1.1 Instructor contacts

Carly Young-Baile: `carly.young-bailie@uct.ac.za`

Monika Looney: `monika.looney@uct.ac.za`

Miguel Rodo: `miguel.rodo@uct.ac.za`

Simon Mendelsohn: `simon.mendelsohn@uct.ac.za`

Munyaradzi Musvosvi: `munyaradzi.musvosvi@uct.ac.za`

Denis Awany: `denis.awany@uct.ac.za`

The full curriculum can be found on the “Syllabus” page.

Happy coding!

2 Syllabus

This page serves as a syllabus for the SATVI Computational Course. Details for each session can be found on their dedicated page.

2.1 Module 1: Intro to R and MaRcus Training Course

2.1.1 Session 1: Intro to R and swirl

Topic: Introduction to R language and environments, RStudio, and swirl self-teaching tools.

Instructor: Monika Looney: monika.looney@uct.ac.za

Date: 05 MAR 2024

Time: 10h30 - 11h30

Location: Lekgotla 4A and 4B

Homework: Complete swirl “R Programming” interactive learning sessions at own pace.

2.1.2 Session 2: MaRcus Training Course lesson 1

Topic: Importing data into R environment and basic visualizations with ggplot2

Instructor: Carly Young-Baile: carly.young-bailie@uct.ac.za

Date: 19 MAR 2024

Time: 11h00 - 12h00

Location: Lekgotla 4A and 4B

Homework: See assignment from https://haswal.github.io/MaRcus/01_session.html

2.1.3 Session 3: MaRcus Training Course lesson 2

Topic: Creating histograms and statistical summaries; combining and exporting plots

Instructor: Carly Young-Baile: carly.young-bailie@uct.ac.za

Date: 26 MAR 2024

Time: 11h00 - 12h00

Location: Lekgotla 4A and 4B

Homework: See assignment from https://haswal.github.io/MaRcus/02_session.html

2.1.4 Session 4: MaRcus Training Course lesson 3

Topic: Basic data transformation using dplyr

Instructor: Carly Young-Baile: carly.young-bailie@uct.ac.za

Date: 02 APR 2024

Time: 11h00 - 12h00

Location: Lekgotla 4A and 4B

Homework: See assignment from https://haswal.github.io/MaRcus/03_session.html

Note - MaRcus Training Course lesson 4 was skipped as it covers R Markdown which will be replaced by a session on Quarto later.

2.1.5 Session 5: MaRcus Training Course lesson 5

Topic: Continuation of data transformation using dplyr and data wrangling

Instructor: Carly Young-Baile: carly.young-bailie@uct.ac.za

Date: 09 APR 2024

Time: 11h00 - 12h00

Location: Lekgotla 4A and 4B

Homework: See assignment from https://haswal.github.io/MaRcus/05_session.html

2.1.6 Session 6: MaRcus Training Course lesson 6

Topic: Clean up data using tidyr

Instructor: Carly Young-Baile: carly.young-bailie@uct.ac.za

Date: 30 APR 2024

Time: 11h00 - 12h00

Location: Lekgotla 4A and 4B

Homework: See assignment from https://haswal.github.io/MaRcus/06_session.html

2.1.7 Session 7: MaRcus Training Course lesson 7

Topic: Manipulating strings with stringr and intro to regular expressions

Instructor: Carly Young-Baile: carly.young-bailie@uct.ac.za

Date: 07 MAY 2024

Time: 11h00 - 12h00

Location: Lekgotla 4A and 4B

Homework: See assignment from https://haswal.github.io/MaRcus/07_session.html

2.1.8 Session 8: Exporting data from R

Topic: Exporting data and plots from R in different formats including csv, pdf, and jpeg

Instructor: Monika Looney: monika.looney@uct.ac.za

Date: 21 MAY 2024

Time: 11h00 - 12h00

Location: Lekgotla 4A and 4B

Homework: Export data frame to csv and save plots as pdf and jpeg;

2.2 Module 2: Quarto, GitHub, and GUIs

2.2.1 Intro to Quarto

Topic: Intro to technical publishing using Quarto

Instructor: Monika Looney: monika.looney@uct.ac.za

Date: 04 JUN 2024

Time: 11h00 - 12h00

Location: Lekgotla 4A and 4B

Homework: Initialize a Quarto project for your own study; Make GitHub account, access SATVILab GitHub, and download GitHub Desktop

2.2.2 Intro to GitHub

Topic: Intro to version control using Git and GitHub

Instructor: Monika Looney: monika.looney@uct.ac.za

Date: 18 JUN 2024

Time: 11h00 - 12h00

Location: Lekgotla 4A and 4B

Homework: Set up a version controlled project; Download VS Code

2.2.3 Intro to VS Code

Topic: Intro to VS Code as an alternative GUI to RStudio and git-aware terminals

Instructor: Monika Looney: monika.looney@uct.ac.za

Date: 02 JUL 2024

Time: 11h00 - 12h00

Location: Lekgotla 4A and 4B

Homework: Write a script in the VS Code GUI

2.3 Module 3: Statistics

2.3.1 Basic statistical tests

Topic: Computing commonly needed statistics and confidence intervals in R

Instructor: Miguel Rodo: miguel.rod@uct.ac.za

Date: 16 JUL 2024

Time: 11h00 - 12h00

Location: Lekgotla 4A and 4B

Homework:

2.3.2 Correlations

Topic: Computing correlation metrics in R

Instructor: Miguel Rodo: miguel.rod@uct.ac.za

Date: 30 JUL 2024

Time: 11h00 - 12h00

Location: Lekgotla 4A and 4B

Homework:

2.4 Module 4: Commonly needed analyses for Immunology

2.4.1 Heatmaps

Topic: Plotting and manipulating heatmaps

Instructor: Monika Looney: monika.looney@uct.ac.za

Date: 27 AUG 2024

Time: 11h00 - 12h00

Location: Lekgotla 4A and 4B

Homework:

2.4.2 Dimensionality reduction

Topic: Understanding and conducting dimensionality reduction using PCA and UMAP

Instructor: Monika Looney: monika.looney@uct.ac.za

Date: 03 SEP 2024

Time: 11h00 - 12h00

Location: Lekgotla 4A and 4B

Homework:

2.4.3 Receiver operating characteristic (ROC) curves

Topic: Understanding and computing ROC curves

Instructor: Simon Mendelsohn: simon.mendelsohn@uct.ac.za

Date: 17 SEP 2024

Time: 11h00 - 12h00

Location: Lekgotla 4A and 4B

Homework:

2.4.4 Background subtraction

Topic: Learning how to apply a function for background subtraction

Instructor: Miguel Rodo: miguel.rodo@uct.ac.za

Date: 01 OCT 2024

Time: 11h00 - 12h00

Location: Lekgotla 4A and 4B

Homework:

2.4.5 Basic flow cytometry analysis

Topic: Plotting background subtracted frequencies and MFIs from flow cytometry data

Instructor: Munyaradzi Musvosvi: munyaradzi.musvosvi@uct.ac.za

Date: 15 OCT 2024

Time: 11h00 - 12h00

Location: Lekgotla 4A and 4B

Homework:

2.4.6 Automatic gating

Topic: Generating inputs for and carrying out automated gating for flow cytometry data

Instructor: Munyaradzi Musvosvi: munyaradzi.musvosvi@uct.ac.za

Date: 29 OCT 2024

Time: 11h00 - 12h00

Location: Lekgotla 4A and 4B

Homework:

2.5 Module 5: Other coding languages

2.5.1 Intro to Python

Topic: Basics of using python and applications for computational immunology

Instructor: Denis Awany: denis.awany@uct.ac.za

Date: 12 NOV 2024

Time: 11h00 - 12h00

Location: Lekgotla 4A and 4B

Homework:

3 Installations

This document provides installation guides for basic programming tools.

3.1 R

R is a commonly used coding language for computational biologists and immunologists. Many software packages and analysis pipelines depend on R. R is also a computational environment used for computing and generating graphics.

To install R for Windows or Mac, follow the instructions provided by The Comprehensive R Archive Network (CRAN) found here: <https://cran.r-project.org/>

It is recommended to download the precompiled binary distribution appropriate for your machine.

To learn more about R, read the following introduction provided by CRAN: <https://www.r-project.org/about.html>

3.2 R Studio

RStudio is an integrated development environment (IDE) based on R. It provides a user-friendly option for building code and can incorporate multiple languages including python, which is also commonly used by computational immunologists.

To download and install RStudio Desktop, follow this link and the provided instructions: <https://posit.co/download/rstudio-desktop/#download>

3.3 GitHub Desktop

GitHub Desktop is a desktop application that interfaces with version-controlled code, GitHub, and other Git services. It provides a user friendly GUI where you can review changes made to code and perform Git commands. It is open source and free to use.

First sign up for a GitHub account at <https://github.com>

Now download and install GitHub Desktop, follow this link and the provided instructions: <https://desktop.github.com>. Sign in with your GitHub account login.

3.4 Visual Studio Code (VS Code)

VS Code is a text and code editor commonly used by developers. It can be used as an alternative for RStudio and supports multiple coding languages and various extensions for debugging and version control.

To download and install VS Code, follow this link and the provided instructions: <https://code.visualstudio.com/download>

4 swirl

swirl (<https://swirlstats.com/>) is an interactive R package that helps you self-teach the basics of R. It is run from directly from the R console.

This session guide follows the instructions provided by swirl. Visit the following link to access the full tutorial: <https://swirlstats.com/students.html>

You can also find the full swirl course tutorial on GitHub at https://github.com/swirldev/swirl_courses

4.1 Install swirl

swirl requires R 3.1.0 or later installed on your computer. It is also recommended that you have RStudio installed which will provide a user-friendly environment to work with.

For instructions on how to install R and RStudio, visit the Installations session guide page.

Once you have downloaded R and RStudio, perform the following steps:

1. Open RStudio.
2. In the RStudio console, type the following where you see the command prompt `>` :

```
install.packages("swirl")
```

4.2 Initialize swirl

Whenever you want to run swirl, you must load and initialize the package.

1. In the console, type the following:

```
library("swirl")  
swirl()
```

2. Follow any prompts that come up in the console. i.e. if swirl asks "What shall I call you?"

4.3 Install an interactive course

The first time you initialize swirl, you will need to install a course.

For the SATVI Computational Course, we recommend that those who are new to coding start with “R Programming”. This course will cover the basics of programming in R.

There are many courses to choose from, so those who are more advanced may opt for an intermediate or advanced course to work through in their own time. A repository with all available swirl courses can be found here: https://github.com/swirldev/swirl_courses#swirl-courses.

There is also an expansive swirl Network that expands further on open source interactive R lessons. You can access the Network and associated courses or become a swirl course author here: <https://swirlstats.com/scn/>

To install a course that is not part of the swirl course repository, type the following into the console:

```
?InstallCourses
```

4.4 Run swirl

For now, we will assume that we are starting with the basics and have chosen to install the “R Programming” course.

To run the interactive lessons:

Select a new lesson. The R Programming course offers 14 different short interactive lessons. Go through each one in order as the information from earlier lessons is required in later lessons.

4.5 Exit swirl

If at any time you need to exit a swirl lesson before it is complete, simply press the Esc key.

If you need to exit from a prompt, exit and save your work by typing: `bye()`

4.6 Interactive commands

While you are working in swirl, you may find that you want to skip a section that you are already comfortable with, or to work more on the current topic outside of an interactive session.

Below are some helpful commands for getting the most out of your swirl sessions:

From the R prompt (`>`):

To skip the current question: `skip()`

To experiment with R on your own without swirl interaction: `play()`

To re-initiate swirl interaction after playing: `nxt()`

To exit and save: `bye()`

To return to swirl's main menu: `main()`

To display these command options: `info()`

If you see a swirl output followed by ... press Enter to continue.

4.7 Homework

As beginners, regular practice is critical! It is recommended that you go through one or two lessons daily to improve and retain these fundamentals.

Over the next week, in your own time, complete the 14 short interactive lessons from the “R Programming” swirl course.

4.8 FAQ

Q1: Can functions learned in swirl be applied when writing my own R scripts?

A: Absolutely! The functions that you use in swirl are all base R functions that can be used

Q2: If I need to use an R package, do I need to install the package each time I start a new session?

A: Nope! Once a package is installed, you do not have to re-install when you open a new R session.

5 MaRcus R Training

The Marcus R Training program was developed by Hasse Walum of Emory University. The program will cover the following:

1. Importing data
2. Basic data visualization
3. Exporting and saving plots
4. Data transformation
5. R Markdown basics
6. Summarizing data
7. String manipulation and data joining

Rather than reinventing what is covered in the Marcus R Training program, we have been granted permission to use the materials for our SATVI Computational Course.

Over the next 6 weeks, we will refer to the Marcus R Training materials for our sessions.

5.1 Content access

The course and all associated resources are available at:

<https://haswal.github.io/MaRcus/index.html>

5.2 Homework

Please refer to the MaRcus R Training program session guides to access your homework assignments.

5.3 FAQ

5.3.0.1 Session 1

Q1: What are the best ways to set your working directory?

A: There are a few ways to do this:

1. If you are using Mac, you can navigate to the directory you would like to work in using the Finder.
2. You can also set the working directory using point and click in RStudio. To do so, navigate to the directory you want to set as the working directory, then click on the "Session" menu in the top right corner of the RStudio window, and select "Set Working Directory to Source File Location".
3. A note about setting working directories in scripts. It is good practice to avoid using relative paths in scripts, as they can be difficult to manage and can lead to errors. Instead, use absolute paths or the `setwd()` function to set the working directory.

Q2: When generating a plot using `ggplot2`, does the name used in the script for the row or column we want to plot have to match the `col` or `rowname` of the associated dataframe exactly?

A: Yes. The names must match exactly because R searches the dataframe for `col` or `rownames` as specified in the `facet_wrap()` or `facet_grid()` functions.

Q3: What is the difference between `facet_wrap()` and `facet_grid()`?

A: Both are options that can be applied to `ggplot2`. `facet_wrap()` wraps a 1d sequence of panels around a variable, while `facet_grid()` creates a grid of panels based on two or more variables.

Q4: When should I specify `aes` globally vs. locally?

A: In general, specify `aes` in mapping (global) so that the specifications are applied to all panels. Use `aes` in the `geom` layer (local) to specify aesthetics that vary across panels.

Q5: What are HEX codes?

A: HEX codes are unique alphanumeric codes assigned to specific colors. They can be used to specify colors in a consistent and reproducible way in your code.

Q6: What are your recommendations for using Chat GPT for help with coding?

A: Chat GPT is a quickly growing tool used by coders. It can be very helpful for designing /

5.3.0.2 Session 2

Q1: What is the difference between top and bottom windows in R Studio?

A: It can help to think of this as an analogy: In R Studio, the top left (script) is your recipe.

Q2: Can you plot confidence intervals automatically using `geom_errorbar` or do you have to calculate them separately first?

A: Confidence intervals should be calculated separately.

5.3.0.3 Session 3

Q1: How can you save the contents of the R console when I finish a session?

A1: You can save the contents of the base R console using the `'sink()'` function. Here you will

For example:

```
sink("output/console_content.txt")
```

Run code of your choice

```
sink()
```

A2: If using RStudio, you can do this via point and click. Navigate to "History" in the top right

Q2: How does `'filter()'` work?

A: The `'filter()'` function from the `'dplyr'` package is used to subset data frames based on specific

```
filter(.data, condition)
```

Here `.data` is any data frame in your environment that you want to filter. Condition needs to be a

```
filtered.data <- filter(original.data, original.data$frequency > 0.05)
```

filtered.data have rows with frequency > 0.05 removed.

Q3: What is the difference between a function and an operator?

A: A function is a chunk of code that is designed to perform a specific task. They typically
Alternatively, an operator is a simple symbol that is used to perform arithmetic, logical, or

Q4: Why does the 'is.na()' function work if the NA in my data frame is uppercase? Isn't it
case-specific?

A: Though most things in R are case specific, is.na() isn't actually looking for the specific

6 Session Recordings

On this page you will find links to the Microsoft Teams recordings for each session. You will only be able to access these recordings if you already have institutional access via Teams. If you would like to access the recordings, but do not have access, please email the course instructors.

6.1 2024 Session Recordings

6.1.1 Session 1: Intro to R and swirl

05 MAR 2024

https://uctcloud-my.sharepoint.com/:v:/r/personal/yngcar003_myuct_ac_za/Documents/Recordings/Introduct20240305_104252-Meeting%20Recording.mp4?csf=1&web=1&e=un87KY&nav=eyJyZWZlcuJhbEluZm8iO3sic

6.1.2 Session 2: MaRcus Training Course lesson 1

19 MAR 2024

https://uctcloud-my.sharepoint.com/personal/yngcar003_myuct_ac_za/_layouts/15/stream.aspx?id=%2Fper

6.1.3 Session 3: MaRcus Training Course lesson 2

26 MAR 2024

https://uctcloud-my.sharepoint.com/:v:/r/personal/yngcar003_myuct_ac_za/Documents/Recordings/SATVI20240326_110027-Meeting%20Recording.mp4?csf=1&web=1&e=2emgFj&nav=eyJyZWZlcuJhbEluZm8iO3sic

6.1.4 Session 4: MaRcus Training Course lesson 3

02 APR 2024

https://uctcloud-my.sharepoint.com/:v:/g/personal/yngcar003_myuct_ac_za/EQ_o7vugBPBJnsZl-sgSLJ8BtAegS-8CMl4fgM-YhSamYg?referrer=Teams.TEAMS-ELECTRON&referrerScenario=MeetingChiclet

6.1.5 Session 5: MaRcus Training Course lesson 4

09 APR 2024

https://uctcloud-my.sharepoint.com/:v:/g/personal/lmymon001_myuct_ac_za/EVgaOigWy4RHvEdDmGx87hOgS9ex5xkIvYw?referrer=Teams.TEAMS-ELECTRON&referrerScenario=MeetingChicletGetLink.view.view

6.1.6 Session 6: MaRcus Training Course lesson 5

30 APR 2024

https://uctcloud-my.sharepoint.com/:v:/g/personal/yngcar003_myuct_ac_za/Ec-xT42msztHmyoRuBJMonELELECTRON&referrerScenario=MeetingChicletGetLink.view.view

7 Summary

In summary, this book has no content whatsoever.

References