

---

## Data science & programming

### Talk to your data in python and R

---

**Sep 22 – Oct 8 2020, 6 sessions 9:30 – 1300**

All sessions will be online

Data Science and Prediction

Leibniz Centre for Tropical Marine Research (ZMT) GmbH

Fahrenheitstraße 6, 28359 Bremen, Germany

**Instructors:** Dr. Arjun Chennu and Dr. Fridolin Haag

**Course description:** This course is an interactive programming workshop adapted to be taught in digital learning settings online. The sessions will cover concepts of programming in general, provide introductions to python and R, data handling, simple data computations, selected analyses on dataframes and arrays, data visualization. Additionally, a session on literate programming, workflow management and project design for reproducible data science will be provided. The course will try to be as interactive as possible through an online collaborative programming environment, live video discussions and exercises.

**Target group:** The course is for those interested in developing skills in data analysis through programming. The course is open to PhD candidates in the 4D-REEF project, as well as any other interested participants at ZMT. A total of 25 students places are available.

**Credit:** Not applicable. This is a voluntary course with no university alignment.

**Preparation:** <https://forms.gle/n1qzMdGb1Yeh4kiT9> (finish before Sep 10)

**Schedule:**

<b>Sep 22</b>	09:30 - 12:30	<p><b>Talk to your computer in python</b></p> <p>What is programming? Is it for me? Getting started with python. Data types and operations. Variables and types. Exercises to explore tasks and language.</p>
<b>Sep 24</b>	09:30 - 12:30	<p><b>Talk to your data in R</b></p> <p>People with a basic knowledge of R who have already done some analysis and want a refresher. What is R and a very short history. What R does well and when not to R. Creating data objects in R. Extracting and indexing data objects in R. Structure of R functions.</p>
<b>Sep 29</b>	09:30 - 12:30	<p><b>Tackling dataframes with R's tidyverse</b></p> <p>Read in rectangular data (dataframes), save rectangular data. Perform basic data wrangling operations on a dataframe. Calculate new values and summaries of a dataframe. Apply functions within dataframes. Run a simple linear regression model on grouped data.</p>
<b>Oct 01</b>	09:30 - 12:30	<p><b>Talk to your data in python</b></p> <p>The data ecosystem in python. Intro to dataframes and statistics in python. Beyond rectangular data: arrays, images and text. Labeled N-dimensional arrays. Statistical and machine learning models. Examples and exercises.</p>
<b>Sep 06</b>	09:30 - 12:30	<p><b>Data visualization in python and R</b></p> <p>What is a good data visualization? Data visualization of dataframes: R &amp; python. The typology of plots: lines, bars, box, images, maps. Examples through R and python.</p>
<b>Oct 08</b>	09:30 - 12:30	<p><b>Reproducible and scalable workflows</b></p> <p>Going beyond functions and scripts. Literate programming as a digital but active lab book. Scalable workflow management. Towards reproducible and collaborative research. Project design considerations.</p>