



<b>Module title</b>	<b>Handling and Visualising Data</b>
<b>Code</b>	D1
<b>Degree Programme</b>	Master of Science in Life Sciences
<b>Workload</b>	3 ECTS (90 student working hours) <ul style="list-style-type: none"> <li>- Asynchronous and synchronous distance learning, decentralized teaching: 32 h</li> <li>- Self-study: 58 h (20 h self-study before module starts)</li> </ul>
<b>Module Coordinator</b>	<b>Name:</b> Dr. Manuel Gil <b>Phone:</b> +41 (0)58 934 57 44 <b>Email:</b> <a href="mailto:manuel.gil@zhaw.ch">manuel.gil@zhaw.ch</a> <b>Address:</b> ZHAW Life Sciences und Facility Management, Schloss 1, 8820 Wädenswil
<b>Lecturers</b>	<ul style="list-style-type: none"> <li>• Dr. Manuel Gil, ZHAW</li> <li>• Dr. Simone Ulzega, ZHAW</li> </ul>
<b>Entry requirements</b>	<ol style="list-style-type: none"> <li>1. <b>Basic statistics</b> experience at the bachelor level is necessary, including: descriptive statistics, basics of probability theory, probability distributions, basic hypothesis testing, and correlation measures. Prior to the course, students will be provided with a detailed list of topics and corresponding references to learning materials.</li> <li>2. Students require some <b>experience with the software R</b>. Prior to the course (one month in advance) preparatory e-learning material will be provided as part of the self-study. Students are expected to work through the material before the course starts and will be evaluated with an entry test.</li> <li>3. The following open source <b>software has to be installed</b> on the students' notebooks: <ul style="list-style-type: none"> <li>• RStudio</li> <li>• Apache Open Office Base</li> </ul> Details (download and installation instructions) will be provided on Moodle prior to the course. </li> </ol>
<b>Learning outcomes and competences</b>	After completing the module, students will be able to: <ul style="list-style-type: none"> <li>• apply programming structures in R (variables, if-statement, loops, functions)</li> <li>• organise data, control data quality,</li> <li>• work with relational databases with graphical user interfaces (GUI),</li> <li>• understand the application of semantic web concepts (triple stores, ontologies) for biological data integration,</li> <li>• reformat, prepare and process data for further analysis,</li> <li>• import data (into statistics software),</li> <li>• handle missing data (imputation),</li> <li>• describe data, check skewness, outliers or unequal variance and quantify these phenomena,</li> <li>• use robust measures of location and scatter to protect from outliers,</li> <li>• understand the grammar of graphics (and apply it with ggplot2),</li> <li>• produce quick exploratory plots as well as publication quality plots of the data,</li> <li>• use different types of plots, adapted to the data type (independent or correlated data such as time series or spatial data, univariate and multivariate data),</li> </ul>

	<ul style="list-style-type: none"> <li>• weigh advantages and disadvantages of different plot types (e.g. what is hidden/glossed over in a particular plot, what is the minimal/maximal sensible sample size for a particular plot, what plot is suited to illustrate which type of relation, etc.),</li> <li>• produce “meaningful” plots, suited to visualize the answer to the research question (e.g. integrating regression lines into scatter plot) or to display the extracted information,</li> <li>• apply principles of good graph design.</li> </ul>
<p><b>Module contents</b></p>	<p><u>Introduction to R (<i>self-study with e-learning</i>)</u></p> <ul style="list-style-type: none"> <li>• Basic R (import/export of data, command line, basic plotting, basic commands)</li> <li>• Programming structures (variables, if-statement, loops, functions)</li> </ul> <p><u>Introduction to the topic “Handling and visualising data” (<i>lecture</i>)</u></p> <p><u>Organising data (<i>lectures and exercises</i>)</u></p> <ul style="list-style-type: none"> <li>• Flat files and redundant data</li> <li>• Relational databases (concepts and querying with a GUI)</li> <li>• Semantic Web technology (Triple, RDF, Ontologies)</li> <li>• Tidy data in R</li> <li>• Classifying and treating missing data</li> </ul> <p><u>Exploring and describing Data (<i>lectures and exercises</i>)</u></p> <ul style="list-style-type: none"> <li>• Measures of location and scatter</li> <li>• Skewness, outliers, unequal variance</li> </ul> <p><u>Visualising data (<i>lectures and exercises</i>)</u></p> <ul style="list-style-type: none"> <li>• Grammar of graphics</li> <li>• Plots in R with <i>ggplot2</i></li> <li>• Design characteristics of good plots</li> </ul> <p><u>Project work (<i>self-study</i>)</u></p> <ul style="list-style-type: none"> <li>• Apply and reinforce the material</li> </ul>
<p><b>Teaching / learning methods</b></p>	<p>The central teaching will consist of lectures, exercises and a group project. During the local coaching the students will continue/complete the work on the exercises and projects from the central teaching. Thus, the local coaching will supplement the central teaching and allow the students to interact personally with a coach to ask questions and obtain closer supervision. Local coaching can be timed flexibly, subject to taking place between the central teaching slots.</p> <p>The self-study will consist of e-learning units (in particular to prepare for the entry requirements), online tutorials, additional reading, and a project work. For the <i>Introduction to R</i> e-learning unit, beginners will require 15-25 hours to work carefully through the tutorial. About 10h are reserved for the completion of the <i>project work</i>, and 10h for exam preparation.</p>



<b>Assessment of learning outcome</b>	- Entry exam on preparatory self-study exercises (open book, 25%) - The final assessment of learning outcome contributes 75% and will either be a written exam, or a project work. This will be decided one month before module starts. Both cases are individual and open book.
<b>Format</b>	7-weeks
<b>Timing of the module</b>	For ZHAW and FHNW: Autumn semester, CW 38-44 For BFH and HES-SO: Spring semester, CW 8-14
<b>Venue</b>	online / decentralized teaching at respective school
<b>Bibliography</b>	<u>Pre-course work</u> Peter Kauf, R online course, provided on Moodle  <u>Course material</u> Wickham, Hadley, 2014. "Tidy data." <i>Journal of Statistical Software</i> 59.10: 1-23. Wickham, Hadley, 2010. "A layered grammar of graphics." <i>Journal of Computational and Graphical Statistics</i> 19.1: 3-28. Wickham, Hadley, 2016. <i>ggplot2: elegant graphics for data analysis</i> . Springer. Tufté, Edward, and P. Graves-Morris, 2014. "The visual display of quantitative information.," 1983."
<b>Language</b>	English
<b>Links to other modules</b>	This module is the basis for module D2 "Design and Analysis of Experiments" and module D3 "Modelling and Exploration of Multivariate Data".
<b>Comments</b>	Material treated during local teaching is relevant for the exam.
<b>Last Update</b>	03.04.2025