

CALDISS Workshop Catalog, fall 2019

This catalog contains descriptions of the workshops that are currently available for demand in CALDISS. Make sure to check back as it is continuously updated!

When 5 requests for the same workshop is received, a workshop will be scheduled with the people interested.

Request a workshop by filling out the formula “[Request a workshop in CALDISS](#)”.

CALDISS Workshop Catalog, fall 2019

- Introduction to R/python

- Extended introduction in R/python

- Flipped classroom workshop series in R/python

- Working with Table Data in R/python

- Exploratory Data Analysis in R/python

- Text Mining and Natural Language Processing in R/python

- Introduction to Geodata with ArcGIS

- Introduction to STATA

- Extended introduction to STATA

- Working with Register Data in SAS/STATA

- Getting Data from the Web (Web Scraping with Web Tools)

- Web Scraping with R/python

- Interactive Cheat-sheet / Refresher Workshop (R, python, STATA)

- Introduction to Qualitative Data Analysis with NVivo

- Analyzing Data with NVivo

- Reporting and Collaborating on Your Research with LaTeX

Introduction to R/python

This workshop is available either in R or python

This 3-hour workshop is meant to introduce how to do data analysis in either R or python.

Both R and python are very powerful tools for data analysis and can be used for statistics, visualization, web scraping, machine learning and so on.

After this workshop you will have the basic knowledge of the language to jump into one of our many workshops offered through the CALDISS workshop catalog.

This workshop will cover:

- Working with the R/python language in either RStudio (for R) or Jupyter Notebook (for python)
- Creating variables and objects
- Importing and exporting data
- Creating simple visualizations

Documentation for python version can be found here:
https://git.its.aau.dk/CLAAUDIA/teach_data.social.sciences

Prerequisites

- None

Requirements

- Own laptop or MacBook.

Duration

- 3 hours

Instructor

- Kristian Gade Kjellmann

Extended introduction in R/python

This workshop is available either in R or python

This 3-hour workshop builds on the introductory workshop and focuses on the generic programming functions of R or python.

This workshop will cover:

- Basic programming skills in R/python (if statements and for loops)
- Creating your own basic functions
- Using functional programming to create re-usable code
- Applying functions on tabular data

Prerequisites

- Basic R/python (see our introductory workshop)

Requirements

- Own laptop or MacBook.

Duration

- 3 hours

Instructor

- Kristian Gade Kjellmann

Flipped classroom workshop series in R/python

This workshop series is available either in R or python

Showing up for 3-hour workshops not your thing? Prefer the comfort of your own office, couch, bed or perhaps pool? We got you covered.

In this series, you will learn either R or python - from basics to some advanced techniques - by completing a series of e-courses at DataCamp (<https://www.datacamp.com/>). 3 exercise workshops (1.5 hours each) will be planned throughout the course material, where you are expected to reach a certain point in the material before each workshop. At the workshops, you will be presented with exercises where you apply the methods and techniques learned.

Workshops can be scheduled to take place physically (at CALDISS) or virtually (using Adobe Connect - <https://www.aub.aau.dk/software-web/adobe-connect/>)

The workshop will by default cover a lot of the same content as the introductory and table data workshops but can be adapted to areas like text mining or exploratory data analysis, if desired.

Prerequisites

- None

Requirements

- Own laptop or MacBook.

Duration

- 9-12 hours (e-courses) + 4.5 hours (exercise workshops)

Instructor

- Kristian Gade Kjellmann

Working with Table Data in R/python

This workshop is available either in R or python

Building on a basic knowledge of python/R, this 2-day (6 hours) workshop teaches how to work with table data: from data-wrangling to doing statistics and basic visualizations.

Note: This workshop does not teach the mathematical and theoretical foundations of statistics but illustrates how to do statistics with the programming language.

This workshop will cover:

- Why python/R for data wrangling and statistics?
- Working with data frames: data structured in rows and columns
- Joining, merging and creating new variables with examples from Statistics Denmark
- Re-coding and sorting data
- Working with text variables (strings) and dates
- Doing descriptive statistics and tests of independence
- Working with regression models and visualizing results

Documentation for python version can be found here:

https://git.its.aau.dk/CLAAUDIA/teach_data.wrangling.statistics

Prerequisites

- Basic R/python (see our introductory workshop)
- Statistics (preferred)

Requirements

- Own laptop or MacBook.

Duration

- 2x3 hours

Instructors

- R: Kristian Gade Kjellmann
- python: Tobias Lindstrøm Jensen

Exploratory Data Analysis in R/python

This workshop is available either in R or python

Building on a basic knowledge of python/R, this 2-day (6 hours) workshop teaches you how to do exploratory data analysis in python/R. The purpose of exploratory data analysis (EDA) is to apply an inductive approach to data and gain insights from data without necessarily working from a pre-defined hypothesis.

In this 2-day workshop we will briefly cover the usage of python/R for tabular data (which is the focus of another CALDISS workshop).

Day one will cover various ways to explore larger datasets with a time dimension as well as visualize some of the results. We will be working with the Dataset from Stanford's Open Policing project: <https://openpolicing.stanford.edu/>.

Day two will go deeper, exploring unsupervised machine learning techniques (dimensionality reduction – PCA, NMF, T-SNE, UMAP and clustering – Kmeans, Hierarchical, (H)DBSCAN) for exploratory analysis.

Prerequisites

- Basic R/python (see our introductory workshop)
- Statistics (preferred)

Requirements

- Own laptop or MacBook.

Duration

- 2x3 hours

Instructors

- R: Daniel Hain
- python: Roman Jurowetski

Text Mining and Natural Language Processing in R/python

This workshop is available either in R or python

Building on a basic knowledge of R/python, this 2-day (6 hours) workshop will cover how to work with textual data in R/python. The workshop will cover everything from reading text data into R/python, various text mining techniques to basic Natural Language Processing (NLP) algorithms and tools.

This workshop will cover:

- Loading textual data into R/python
- Familiarity with essential text mining packages in R/python
- Building a corpus in R/python
- Pre-processing text data for analysis
- Exploring key features of texts (keywords, word associations)
- Discovering common themes in texts using topic modelling
- Analyzing data from the web and social media

Prerequisites

- Basic R/python (see our introductory workshop)

Requirements

- Own laptop or MacBook.

Duration

- 2x3 hours

Instructors

- R: Arnim Decker / Kristian Gade Kjellmann
- python: Roman Jurowetski

Introduction to Geodata with ArcGIS

This 2-day workshop introduces how to work with geodata in ArcGIS.

The workshop will focus on use-cases within history research but attendees from other areas are of course welcome.

It is possible to work with your own data during the workshop as well, if it is applicable.

This workshop will cover:

- Understanding and using the ArcGIS interface.
- Appropriate geodata and how to import it
- Various data sources for geographic data
- Creating various visualizations in ArcGIS (heatmaps, rate-maps etc.)
- Differences and use-cases with polygon-maps and point-maps

Prerequisites

- None

Requirements

- Own laptop or MacBook.

Duration

- 2x3 hours + follow-up

Instructor

- Rolf L. Lund

Introduction to STATA

Using STATA for your statistics course? Or do you just want to learn the program? Then join this introductory workshop.

STATA is a relatively simple and powerful statistics and econometrics software tool. This 3-hour workshop starts from the very basics and will give you the foundation needed to get started doing statistics with STATA.

This workshop will cover:

- Navigating the STATA interface
- Reading and inspecting data in STATA
- Using the command window
- Using do-files to easily re-do your steps and analysis
- Basic recoding and variable manipulation
- Descriptive statistics

Prerequisites

- None

Requirements

- Own laptop or MacBook.

Duration

- 3 hours

Instructors

- Andreas L. Jakobsen
- Kristian Gade Kjellmann

Extended introduction to STATA

This 3-hour workshop builds on the introductory workshop and teaches additional functions in STATA.

This workshop will cover:

- Additional recoding and variable manipulation features
- Good tips on using the do-file
- Additional commands for statistics, including (summary statistics, cross-tables and tests of independence)

Prerequisites

- Basic STATA (see our introductory STATA workshop)

Requirements

- Own laptop or MacBook.

Duration

- 3 hours

Instructors

- Andreas L. Jakobsen
- Kristian Gade Kjellmann

Working with Register Data in SAS/STATA

Statistics Denmark (<https://www.dst.dk/>) possess a variety of register data containing a lot of different information. Working with data from Statistics Denmark can present a challenge compared to typical data analysis workflows due to the peculiar structure and formats of the data.

This 3-hour workshop introduces good practices for getting started with working with register data - both what register data looks like and how to work with it in either SAS or STATA. The workshop contain exercises on simulated register data allowing you to familiarize yourself with the data before you have to work with the real thing.

This workshop will cover:

- Types and structures of register data
- Tips and tricks for combining register data in SAS or STATA
- How to use the documentation from Statistics Denmark
- How to handle frequently encountered problems when working with register data (string and date variables)

Prerequisites

- SAS or STATA

Requirements

- Own laptop or MacBook.

Duration

- 3 hours

Instructors

- SAS: Kristian Gade Kjellmann
- STATA: Rolf L. Lund

Getting Data from the Web (Web Scraping with Web Tools)

This 3-hour workshop is meant to introduce web scraping: how to get data from the web. The workshop will cover different types of scraping (using API's and getting raw website data) and will introduce various existing web tools for getting data from the web.

This workshop will cover:

- Various data sources from the web (API's, raw web data, social media data)
- Existing tools for extracting data from the web (Digital Methods Initiative)
- Good practices and considerations when working with the web

Prerequisites

- None

Requirements

- Own laptop or MacBook.

Duration

- 3 hours

Instructor

- Kristian Gade Kjellmann

Web Scraping with R/python

This workshop is available either in R or python

This 3-hour workshop teaches how to use programming tools to get data from the web - either through using API's or scraping raw data.

The workshop will introduce working with data from the web. It covers how programming languages interact with the web and how they can be used for automated collection (scrapers).

This workshop will cover:

- Various data sources from the web (API's, raw web data, social media data)
- Using programming language to collect data from the web (using python examples)
- Good practices and considerations when working with the web

Prerequisites

- Basic R/python (see our introductory workshop)

Requirements

- Own laptop or MacBook.

Duration

- 3 hours

Instructor

- Kristian Gade Kjellmann

Interactive Cheat-sheet / Refresher Workshop (R, python, STATA)

This 3-hour workshop is an on-demand Q&A going through a lot of the most basic and useful function in either R, python and STATA. Material is pre-planned but the workshop will primarily answer questions from the participants.

Prerequisites

- Basic R/python/STATA (see our introductory workshop)

Requirements

- Own laptop or MacBook.

Duration

- 3 hours

Instructors

- STATA: Andreas L. Jakobsen
- R/python: Kristian Gade Kjellmann

Introduction to Qualitative Data Analysis with NVivo

This introduction covers what kind of program NVivo is and its main uses. From there the interface is introduced, taking you through all the basic functions and explaining the terminology using an existing project. Finally, you will start your own project and try out the basic functions on a single text file.

The introduction can be supplemented with a more in-depth workshop on using NVivo.

Prerequisites

- None

Requirements

- Own laptop or MacBook.

Duration

- 3 hours

Instructor

- Kristian Gade Kjellmann

Analyzing Data with NVivo

This 1-day workshop will go in-depth in working with NVivo. From using the program's interface to making use of the many features for qualitative analysis. It will be possible to work with your own data at the workshop. If you do not have any data to work with a research case with data will be provided.

Contents of the workshop

This workshop is partitioned into two modules.

MODULE 1: Analyzing data with NVivo

Starting from a blank project, this module goes through the various local file and data formats that NVivo can import and work with. The module then focuses on how to make use of the basic and recurrent functions when analyzing such data and how they differ (the use of nodes, cases and attributes). The module also introduces the use of queries and visualization for gaining insights from your data.

MODULE 2: Advanced functions and NVivo as a collaborative platform

This module builds on an existing project and introduces how to use the external data function (internet, social media, NVivo Capture etc.). The module focuses on how to use the basic and recurrent functions with these external sources (nodes, cases and attributes) as well as how to gain insights using queries and visualizations.

The module goes on to discuss advanced project-structures when working with multiple data sources as well as using NVivo as a collaborative platform

Prerequisites

- Basic knowledge of NVivo (see our introductory NVivo workshop)

Requirements

- Own laptop or MacBook.

Duration

- 2x3 hours

Instructor

- Kristian Gade Kjellmann

Reporting and Collaborating on Your Research with LaTeX

LaTeX is one of the most common typesetting languages used in the academic world. This course introduces you to the LaTeX language to write papers, reports, and slides. To produce the content we will make use of Overleaf, an online resource that removes installation hassles. It also allows for multiple users in the same documents (goodbye overwritten files) and comments in your text and code. Finally we discuss how to produce graphs and LaTeX tables in R that can be imported directly into your LaTeX files.

This workshop will cover:

- Setting up your first LaTeX document using Overleaf
- Including tables, pictures, equations, and footnotes in your document
- Splitting a report in multiple files for simultaneous work on project
- Reference handling in LaTeX using Mendeley
- Using R and LaTeX for reproducibility
- [Optional] Presentation slides in LaTeX
- [Optional] Using Rmarkdown to write LaTeX documents

Prerequisites:

- None
- Participants unfamiliar with data analysis in R may look up alternatives to the statistical softwares they use instead, e.g. regout2 in Stata.

Requirements:

- Own laptop and an active Overleaf account. To create an Overleaf account, simply go to [Overleaf.com](https://overleaf.com) and sign up. If you are an AAU affiliate and sign in with your AAU account, you will get access to the premium version of

Duration:

- 2.5 hours
- If sufficiently many participants are interested it is possible to get an introduction to Rmarkdown. This allows you to write LaTeX documents directly from RStudio.

Instructor:

- Jesper Eriksen