Linguistic Data Analysis using R - Syllabus FS2024
!under construction -- subject to change!

Instructor: Sandra Auderset (Office B163) sandra.auderset@unibe.ch
Time and Place: Tuesdays 10:15-12:00 in F002

Evaluation:

- project proposal: 10%
- draft of paper and code: 40%
- final paper with code and incorporated feedback: 50%

Readings, homework and attendance policies:

- Readings are mandatory and are assigned for the next class (e.g., the reading assigned on Feb 27 needs to be completed before class on March 5).
- I don't take attendance, but since the class is hands-on, I recommend consistent attendance. If you miss a class, it is your responsibility to catch up.
- Homework is not mandatory and not graded, but also recommended. If you turn them in on time, I will give you feedback and comments.
- Homework is due before class on the respective due date (check the assignment sheet). They should be submitted on ILIAS.

Class	**Content**	**Weekly readings**	**Due before next class**
20.02.	Class organization; Why data analysis for linguistics?; What is a database?; Intro to R and the tidyverse	Winter 2020: Sections 0.2 + 0.3 R4DS 2023: Introduction OHLDM 2021: Chapter 1.1 (1.2 optional)	Homework 1, **R and RStudio installed**
27.02.	Discussion of homework; Getting started with R data formats & types	Winter 2020: Sections 1.1-1.11 R4DS 2023: 2 Workflow: basics; 7 Data import	--
05.03.	Getting started with the tidyverse: reading in data, intro to pipes; basics of data wrangling	Winter 2020: Sections 1.13-1.15; 2.1, 2.2, 2.5 Tidyverse Style Guide: 1, 2, 4	Homework 2
12.03.	Data wrangling: understanding 'tidy' and 'messy' data; cleaning and transforming data; How to get help	R4DS 2023: 5 Data tidying, 12 Logical vectors, 13 Numbers, 14 Strings, 16 Factors, 18 Missing values	--
19.03.	Data wrangling: summarizing and aggregating data; joining data sets; Intro to descriptive statistics	Winter 2020: Chapter 3 R4DS 2023: 19 Joins	Homework 3
26.03.	Intro to data visualization; Getting started with ggplot2; Exploratory data analysis; Discussion of final project guidelines	Winter 2020: Chapter 3 Exercises R4DS 2023: 9 Layers, 10 Exploratory data analysis	--
Easter break			

09.04.	Data visualization and plotting II; labels, palettes, and other customization	R4DS 2023: 11 Communication GGPLOT2 book: 3 Individual geoms, 8 Annotations, 11 Colour scales and legends	Homework 4
16.04.	Making maps with ggmaps and maps	Kahle & Wickham 2013 GGPLOT2 book: 6 maps	**get a GitHub account, install GitHub Desktop**
23.04.	Project management; File versioning with GitHub and GitHub Desktop; Rmarkdown	OHLDM 2021: Chapter 8 (min. Sections 2.3, 2.4, 2.5) look around in R Markdown: The Definitive Guide Intro to Github Desktop	--
30.04.	**project proposal due** Cluster analysis I	DataNovia: Cluster analysis datarundown: Types of clustering	Homework 5
07.05.	Cluster analysis II; Final project guidelines discussion; Correlation plots and measures (if possible)	Norvik et al. 2022: Uralic typology in the light of a new comprehensive data set https://doi.org/10.1075/jul.00002.nor	--
14.05.	Use case 1: Uralic Typological database UraTyp	Yang 2022: The phonetic tone change *high > rising (pdf on ILIAS)	
21.05.	Use case 2: Phonetic tone change in Ngwi	OHLDM 2021: Chapter 4 and 5	prepare presentation
28.05.	Data (base) management; Ethics of Open science and open data; Project presentations and feedback; discussion	--	--
12.07.	**draft version of paper and code due**		
23.08.	**final version of paper and code due**		