LECTURE / SESSION[1] PLAN

Name Lecturer: Sofia Gil-Clavel	**Date session:** 7	**Expected number of students:** 20
Course title	R-Workshop	
Topic of lecture or session	Tidyverse	
Situational factors (e.g. group size, prior knowledge, expected motivation)	The class lasts 3hrs. The students are in different career stages, from PhD students to professors. The students have different backgrounds, from qualitative researchers that have never used R to quantitative researchers that want to learn new topics or move out of SPSS/STATA. The students use different Operating Systems (Windows, Mac, or Linux).	
Intended learning outcomes of this session	At the end of the session the students will be able to: 1. Open textual data and turn it into a data frame. 2. Deal with text that has the wrong encoding. 3. Use the package stringr for basic text handling. 4. Understand the tidy data principles for textual data. 5. Understand and perform text indexing, i.e. stemming and stop word removal. 6. Use the packages tidyverse and tidytext to extract and count word frequencies. 7. Use ggplot2 to visualize textual data.	
Learning material (book, chapters, ...)	The lecture is based on the book: • Robinson, Julia Silge and David. *Welcome to Text Mining with R	Text Mining with R*. Accessed March 12, 2025. https://www.tidytextmining.com/.	

[1] A session means a teaching and learning session for university students in the bachelor or master. It can be a lecture, a seminar or a specific type of educational meeting with a group of students that is relevant for the discipline where the lecturer can demonstrate a whole range of teaching skills. It means that in one session each of the six didactic elements appears at least once. Your lesson plan should show that you are using a powerful learning environment and that students are activated.

Media, equipment, tools	The students use their own laptops. The teacher needs access to a projector and a whiteboard.
Preparation for students	The students have access to the slides and codes before the class: https://github.com/SofiaG1l/R_Course/tree/master/R4SocialScientists/Session7_TextAsData

Time[2] (min.)	**Didactic element (goal)[3] and topic[4]**	**What the teacher does[5]** (teacher activity)	**What students do** (learner activity)	**Evaluation[6]** (feedback/assessment)
20	• Open textual data and turn it into a data frame.	• The teacher uses the slides to introduce the data that will be used during the class. • The teacher opens RStudio and opens the already written script that will be used during the class. • The teacher uses the slides to explain the logic behind the code to open textual data and turn it into a data frame.	• The students passively digest what the teacher is explaining. • The students follow the steps in their computers. • The students independently start filling out the missing parts of the R-script. This is based on what the teacher explained using the slides.	• The teacher walks around the classroom to check on the students and provide feedback when something is not working on their computers. When the teacher detects a common error, then the teacher uses the whiteboard to clarify.
30	• Deal with text that has the	• The teacher uses the slides to explain what encoding is and the	• The students passively digest what the teacher is explaining.	• The teacher walks around the classroom to check on the students and provide feedback

[2] Indicate the planned duration in minutes
[3] State the number(s) of the relevant ILO at each didactic element
[4] Only key words
[5] Specify the type of activity and write down also the questions that you prepared to ask in the session
[6] Specify the type of evaluation, i.e. the way in which you assess if the objective(s) has/have been achieved

	wrong encoding.	consequences of opening data with the wrong encoding. • The teacher explains how to set up RStudio to use the right encoding (utf-8) to open and save data. • The teacher also introduces some functions to deal with wrong encoding.	• The students follow the steps in their computers. • The students independently start filling out the missing parts of the R-script. This is based on what the teacher explained using the slides.	when something is not working on their computers. When the teacher detects a common error, then the teacher uses the whiteboard to clarify.
20	• Use the package stringr for basic text handling.	• The teacher uses the slides to explain why we use the package stringr instead of the basic R functions.	• The students passively digest what the teacher is explaining. • The students follow an exercise where they have to apply three different functions and explain out loud what they do and in what contexts they would use them.	• The teacher walks around the classroom to check on the students and provide feedback when something is not working on their computers. When the teacher detects a common error, then the teacher uses the whiteboard to clarify.
10	• Understand the tidy data principles for textual data.	• The teacher uses the slides to the tidy data principles for textual data.	• The students passively digest what the teacher is explaining.	• The students will use these concepts during the workshop. So, the teacher will detect when a student confuses them. Based on these confusions the teacher will be able to correct the student.
40	• Understand and perform text indexing,	• The teacher uses the slides to explain what text indexing is and when to perform it.	• The students passively digest what the teacher is explaining.	• The teacher walks around the classroom to check on the students and provide feedback

	i.e. stemming and stop word removal.		• The students independently start filling out the missing parts of the R-script. This is based on what the teacher explained using the slides.	when something is not working on their computers. When the teacher detects a common error, then the teacher uses the whiteboard to clarify.
40	• Use the packages tidyverse and tidytext to extract and count word frequencies.	• The teacher uses her R-script and the white board to explain the different steps the students need to follow to transform the text and count its words frequencies.	• The students passively digest what the teacher is explaining. • The students independently start filling out the missing parts of the R-script. This is based on what the teacher explained using the slides.	• The teacher walks around the classroom to check on the students and provide feedback when something is not working on their computers. When the teacher detects a common error, then the teacher uses the whiteboard to clarify.
20	• Use ggplot to visualize textual data.	• The teacher uses her R-script and the white board to explain the different steps the students need to follow to visualize the words frequencies. • This is done using a basic ggplot2 geometries and also using the package ggwordcloud to create a wordcloud.	• The students passively digest what the teacher is explaining. • The students independently start filling out the missing parts of the R-script. This is based on what the teacher explained using the slides.	• The teacher walks around the classroom to check on the students and provide feedback when something is not working on their computers. When the teacher detects a common error, then the teacher uses the whiteboard to clarify.