**LECTURE / SESSION[1] PLAN**

| **Name Lecturer:**<br>Sofia Gil-Clavel | **Date session:**<br>8 | **Expected number of students:**<br>20 |
|---|---|---|
| Course title | R-Workshop | |
| Topic of lecture or session | Text Mining | |
| Situational factors (e.g. group size, prior knowledge, expected motivation) | The class lasts 3hrs.<br>The students are in different career stages, from PhD students to professors.<br>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.<br>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:<br>1. Explain the main purpose of text mining.<br>2. Quantify what a document is about by calculating the tf-idf.<br>3. Extract relationships between words using the n-grams.<br>4. Calculate the n-grams tf-idf.<br>5. Calculate the correlation between words.<br>6. Visualize the relations between the words using networks. | |
| Learning material (book, chapters, …) | The lecture is based on the book:<br>• 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.<br>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:<br>https://github.com/SofiaG1l/R_Course/tree/master/R4SocialScientists/Session8_TextMining |

| **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 | • Quick recap of previous session. | • The teacher uses the slides to remind the students what they learned the previous session:<br>  ○ How to open and clean textual data.<br>  ○ Transform textual data into a data frame.<br>  ○ Tidy text principles<br>  ○ Indexing<br>• The teacher opens RStudio and opens the already written script that will be used during the class. | • Before telling each element the students learned, the teacher waits some seconds for the students to fill out the information out loud.<br>• The students open and follow the steps that 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. |

[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

| | | | | |
|---|---|---|---|---|
| 10 | • Explain the main purpose of text mining. | • The teacher uses the slides to introduce the topic: Text Mining. | • 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. |
| 20 | • Quantify what a document is about by calculating the tf-idf. | • The teacher uses the slides to explain what tf-idf and how to calculate it using R. | • The students passively digest what the teacher is explaining.<br>• 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 | • Extract relationships between words using the n-grams. | • The teacher uses the slides to explain what an n-gram is and why they are useful in text mining. | • The students passively digest what the teacher is explaining.<br>• 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 | • Calculate the n-grams' tf-idf. | • The teacher uses the slides to remind students what a tf-idf is and also explains how interpret it. | • The students passively digest what the teacher is explaining. | • The teacher walks around the classroom to check on the students and provide feedback |

| | | | | |
|---|---|---|---|---|
| | | | • 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 | • Calculate the correlation between words. | • The teacher uses the slides to explain how to calculate the correlation between words and also explains how interpret it. | • The students passively digest what the teacher is explaining.<br>• 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 | • Visualize the relations between the words using networks. | • The teacher uses the slides to explain what a network is and why they are useful to visualize relations between words. | • The students passively digest what the teacher is explaining.<br>• 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 | • Quick overview of Natural | • The teacher uses the slides to explain what Natural Language Processing is and its differences with text mining. | • The students passively digest what the teacher is explaining.<br>• The students independently start running the R-script. This | • The teacher walks around the classroom to check on the students and provide feedback when something is not |

| | | | | |
|---|---|---|---|---|
| | Language Processing | • The teacher provides some code so the students can play around with some basic Natural Language Processing functions. | is based on what the teacher explained using the slides. | working on their computers. When the teacher detects a common error, then the teacher uses the whiteboard to clarify. |