UNIVERSITY OF MANNHEIM

School of Social Sciences
Chair of Political Science, International Relations

Complexities in analyzing conflicts: Data wrangling and data management in R

(Methods of International Relations, 6 ECTS)

Fall 2019

Instructor:	Cosima Meyer	**Time:**	Wednesdays, 8:30 – 11:45
			Fridays, 8:30 – 11:45
E-mail:	cosima.meyer@uni-mannheim.de	**Room:**	A5,6 C-108

Student hours: Upon request. Please contact me via e-mail (including a short description of your problem or question).

Course description: Quantitative research in conflict studies is of high societal interest but still remains a challenging task. In this course, students will tackle common problems that arise during data management processes in conflict research by being introduced to R. Using standard datasets in conflict studies, students will learn how to master R and submit a first research project with R as part of their final submission. The goal of this course is to provide students with guidelines for good data management practices and visualization.
To give you a good overview of important datasets, we will also present and discuss parts of the existing data basis.

The course is taught in English.

Requirements: Download and install R and R Studio.

Important Deadlines:

1) Complete "Intro to R" on DataCamp ..... 05.11.2019 (ILIAS)
2) Research question ..... 08.11.2019 (ILIAS)
3) Homework ..... Respective week (ILIAS)
4) Presentation of course projects ..... 29.11.2019 (in class)
5) Handout for data set ..... 30.11.2019 (ILIAS)
6) Term paper ..... 10.01.2020 (ILIAS)

Submissions (1), (2), (3), (5) and (6) must be handed in electronically via ILIAS at 23:59 (the latest) on the respective day.
Giving a brief presentation of the course project (4) is expected in-class.

Late submission policy: Late submissions will not be accepted. If there are any (severe) problems in meeting the deadlines, I expect you to discuss the issues with me *in good time beforehand*.

Plagiarism: The University's minimum penalty for plagiarism is failing the course. Cheating or plagiarism can lead to expulsion from the University of Mannheim. Your written assignments will be checked for plagiarism.

Grading policy: The examination regulation of the University of Mannheim differentiates between coursework ("Studienleistung" or "SL") and examination ("Prüfungsleistung" or "PL"). Coursework is graded as 'passed' or 'failed'. It is defined as a required preliminary coursework and thus necessary for the admission to an examination. *Please keep in mind: without successfully passing the coursework, there is no admission to the final examination.* The final examination is assessed at the end of the semester and is graded according to the grading scheme of the University of Mannheim (for further information please see your examination regulations, in particular §12(2), §13(a), §17; http://bit.ly/2hNyxTS).

1. Complete "Intro to R" on DataCamp (pass/fail)
2. Research question (pass/fail)
3. Homework (pass/fail): You need to pass 2 of 3 homeworks to pass the coursework section *Homework*.
4. Presentation of course projects (pass/fail)
5. Handout for data set (pass/fail)
6. Term paper (graded)

You find all essential information on general requirements in the Appendix in the section "Assessment criteria".

Class attendance and participation: The learning process of this class is based on in-class discussion and participation. Attendance and careful preparation of the course material is therefore highly recommended. This includes coming to class on time.

Laptop and phone policy: In order to ensure an active participation and to keep your attention on the important things (our class), please avoid distracting yourself and others through electronic devices. For further insights on the consequences of multitasking, I recommend the study by Bellur, Nowak, and Hull (2015) (`https://bit.ly/2GnyTf2`). They found that in-class multitasking leads to significantly lower performance.

Interesting links: There is a number of great websites and podcasts that provide additional interesting information. See for instance R-bloggers, MZES Methods Bites, R Graph Gallery, Towards Data Science, FiveThirtyEight, and Bad Hessian.
As a thematic input, I recommend to follow the news regularly and strongly encourage you to refer to them during the class discussions. Besides the newspapers, I also recommend to follow the political science blogs "Monkey Cage", "Political Violence @ a Glance", Christopher Blattman's blog, the talks at "The McMillan Report", or the app "The Economist Espresso" (gives you a daily morning briefing). For further input, have a look at the Chair's homepage.

DataCamp: We will use DataCamp as an online learning platform throughout this class. For this, I acquired you a six-months free premium access that you will use for this course but are also encouraged to use it beyond the course for your own personal learning development. You can easily take courses in R, Python and SQL through a combination of short expert videos and hands-on-the-keyboard exercises. There are more than 100 courses by expert instructors on topics such as importing data, data visualization or machine learning. Some courses will be required to pass the coursework but I also selected several additional classes that fit to the respective sessions (highlighted with ✎ in the syllabus).

Schedule

Week 0		**Preparation**	*Complete the course "Intro to R" on DataCamp by November 5, 2019.*
Week 1	Nov 6	**Introduction to the seminar and organizational issues**	
		Intro to R I	
	Nov 8	**Intro to R II**	*Hand in your research question.*
		Intro to R III	
Week 2	Nov 13	**Data wrangling and merging I**	*Hand in your homework.*
		Data wrangling and merging II	

*There is **no** class on November 15.*

Week 3	Nov 20	**Data wrangling and merging III**	*Hand in your homework.*
		Data wrangling and merging IV	
	Nov 22	**Data visualization I**	
		Data visualization II	
	Nov 26	**LaTeX workshop (optional)**	*The workshop will take place in room A231 (A5, 6) at 12:00-13:30.*
Week 4	Nov 27	**Hackathon I**	*Hand in your homework.*
		Hackathon II	
	Nov 29	**Presentation of course projects**	*Present your course project Submit a handout on your dataset*
		Outlook and semester-wrap up	

Course outline:

The readings consist of book chapters. All *required readings* can be accessed on ILIAS.

I ask you to submit your assignments (❗) within the deadline and to prepare the required readings (📖) carefully. CheatSheets (💡) and additional input on DataCamp (📝) are optional. All the resources should help you if you need further input or want to expand your knowledge.

Week 0: Preparation

Before starting with the course, I want you to already get some first insights into R so that we can depart from a similar level.

❗ DataCamp Course: Intro to R

- ☐ Sign up on DataCamp.
- ☐ Complete the class "Intro to R".

Week 1: What is R and what can we do with it?

Session 1: Introduction to the seminar and organizational issues (06.11.2019)

This session will introduce you to the seminar in general, give a brief outline of the course, and address organizational issues. We will use the paper by Hultman, Kathman, and Shannon (2014) as a working example that guides us through this course.

📖 Required readings

- ☐ Hultman, Lisa, Kathman, Jacob D., and Shannon, Megan (2014). "Beyond keeping peace: United Nations effectiveness in the midst of fighting." *American Political Science Review* 108(4): 737–753

Session 2: Intro to R I (06.11.2019)

We will then have a short introduction to R and R Studio and meet RMarkdown – an efficient way of typesetting text and code in R.

📖 Required readings

- ☐ Healy, Kieran (2018). *Data visualization: A practical introduction.* Princeton University Press, Chapter 2.1-2.4

- ☐ Imai, Kosuke (2018). *Quantitative social science: An introduction.* Princeton University Press, Chapter 1.3

Optional readings

- Wickham, Hadley and Grolemund, Garrett (2016). *R for data science: Import, tidy, transform, visualize, and model data.* O'Reilly Media, Inc., Chapter 2, 6, 9, 21, 23, 24

Additional input on DataCamp

- ☐ Introduction to the Tidyverse
- ☐ Data Scientist With R (Career Track)
- ☐ Tidyverse Fundamentals (Skill Track)
- ☐ Intermediate Tidyverse Toolbox (Skill Track)

CheatSheet

- ☐ R Studio
- ☐ R Markdown
- ☐ R Markdown Reference Guide
- ☐ Tidyverse for Beginners

Session 3: Intro to R II (08.11.2019):

Now that we have met R, we dedicate this session to the learn more of R's logic and basics – objects, vectors, and data frames.

Additional input on DataCamp

- ☐ Introduction to R (for a recap)

CheatSheet

- ☐ Brief Introduction to Language Elements and Control Structures
- ☐ Base R

Research question

- ☐ Please hand in your research question by today.

Session 4: Intro to R III (08.11.2019):

In this session, we will learn how to import different data formats in R and have a first (descriptive) look at the data.

Additional input on DataCamp

- ☐ Importing and Cleaning Data With R
- ☐ Intermediate R
- ☐ Importing & Cleaning Data With R (Skill Track)

CheatSheet

- ☐ Data Import

Week 2: Data wrangling and merging

Session 5: Data wrangling and merging I (13.11.2019)

Now that we successfully loaded all data and had some first looks at the descriptive statistics, we will learn how to clean (more or less) messy data in R.

Additional input on DataCamp

- ☐ Cleaning Data in R

CheatSheet

- ☐ Data Wrangling With `dplyr` and `tidyr`

Homework

- ☐ Homework 1 is due today.

Session 6: Data wrangling and merging II (13.11.2019)

We then proceed and learn how to manipulate data in R to adjust it to our (format) needs before we learn how to merge the data in the next session.

Additional input on DataCamp

- ☐ Data Manipulation With R
- ☐ Data Manipulation with dplyr in R

CheatSheet

- ☐ Data Wrangling With `dplyr` and `tidyr`

No class on November 15, 2019.

Week 3: Data wrangling and merging & data visualization

Session 7: Data wrangling and merging III (20.11.2019)

Merging data can easily become one of the most complex parts of data processing. In this session, we will first learn the basics of merging in R before we look at more tricky parts in the next session of today's course.

Additional input on DataCamp

- ☐ Joining Data in R With `data.table`

CheatSheet

- ☐ Data Wrangling With `dplyr` and `tidyr`

Homework

- ☐ Homework 2 is due today.

Session 8: Data wrangling and merging IV (20.11.2019)

We now apply our merging knowledge to tackle more tricky merging tasks.

Additional input on DataCamp

- ☐ Wrangling and Visualizing Musical Data

Session 9: Data visualization I (22.11.2019)

In the third part of our seminar, we delve into the magic world of `ggplot2` – a great way of plotting your results in R. In the first session, we will learn the basics of `ggplot2`.

Required readings:

- ☐ Healy, Kieran (2018). *Data visualization: A practical introduction.* Princeton University Press, Chapter 3.1-3.7

Additional input on DataCamp

- ☐ Data Visualization With R
- ☐ Data Visualization With `ggplot2` (Part 1)
- ☐ Data Visualization With `ggplot2` (Part 2)

- ☐ Data Visualization With `ggplot2` (Part 3)
- ☐ Data Visualization With R (Skill Track)

CheatSheet

- ☐ `ggplot2`

Session 10: Data visualization II (22.11.2019)

In the second session, we will then produce more advanced figures with `ggplot2`. This Medium blogpost gives you an additional idea how to produce Tableau-like graphs in R.

Required readings

- ☐ Healy, Kieran (2018). *Data visualization: A practical introduction.* Princeton University Press, Chapter 8.1-8.4

Optional readings

- Healy, Kieran (2018). *Data visualization: A practical introduction.* Princeton University Press, Chapter 7

Additional input on DataCamp

- ☐ Interactive Data Visualization With `plotly` in R
- ☐ Data Visualization With R (Skill Track)

CheatSheet

- ☐ `plotly`

LaTeX workshop (optional, 26.11.2019)

For those of you who are interested, there will be a LaTeX workshop offered by the Data and Methods Unit at the MZES. This LaTeX workshop includes a quick introduction, hands-on practices and a template for your term papers and theses. The aim is to provide you with sufficient knowledge to write (future) term papers with the template and to cope with common problems in LaTeX. However, there is no need to write your term paper with LaTeX. I just experienced that using LaTeX is way more efficient than using Word (or any other text processing program).
The workshop is part of the Social Science Data Lab at the MZES and is co-organized with Dennis Hammerschmidt.

💡 CheatSheet

- ☐ LaTeX

Please bring a charged laptop.

The session will take place in room A231 (A5, 6) at 12:00-13:30.

Week 4: Hackathon, project presentations and wrap-up

Session 11 + 12: Hackathon I + II (27.11.2019)

During the Hackathon, we dedicate two double-sessions entirely to your course projects. It allows you to work concentrated on your projects, consult me individually for help and address open questions as well as structure your final projects.

❗ Homework

- ☐ Homework 3 is due today.

Session 13: Presentation of course projects (29.11.2019)

During this session, we will set up a mini conference setting. You will have the chance to present your course projects and receive constructive feedback from your peers that should help you when preparing your final paper.

Session 14: Outlook and wrap-up: What else can we do in R? (29.11.2019)

To wrap up the seminar, I will give you a some insights what other cool things that you can do in and with R.

We will also have a final discussion and deal with potential challenges of your term papers. If you have any further questions or would like to discuss something particular, please feel free to contact me beforehand.

Assessment criteria

Coursework

1. Complete "Intro to R" on DataCamp

Before we delve into R, I want you to familiarize yourself with first basics in R. The courses offered by DataCamp provide an incredibly good way in doing so.

- Sign up on DataCamp
- Complete the course "Intro to R" **before November 6, 2019**
- You will receive a certificate once you completed the class. Upload the certificate on ILIAS as a proof for completion.

2. Research question

To match your interests for the course projects, I ask you to submit your research questions. I will then use your research questions to group you according to your interests. Collaboration is key and group works are meant to build these synergistic learning environments where you can learn and teach from and with others. While everyone has to submit her/his individual final paper (project), you should tackle the basic (data) problems in your group together and then use your output to answer your individual research question.
I will also provide you with feedback on the feasibility and plausibility of your research question.

Deadline: November 8, 2019 (23:59)

Submission: On ILIAS

Should be included:

- Analytical research question
- (You may also add a few explanatory sentences.)

3. Homework

I will assign homeworks to you via DataCamp. The homework should help you to strengthen your previously learned knowledge. To pass this coursework, **you need to pass 2 out of 3** homeworks.

Deadline: Respective week

Submission: You will receive a certificate once you completed the class on DataCamp. Upload the certificate on ILIAS as a proof for completion.

4. Presentation of course projects

The presentations give you the chance to present your projects and to receive constructive feedback from your peers.
The data presentations should not be longer than 15 minutes (**maximum**) with slides (PowerPoint, markdown (html, PDF), LaTeX (PDF), or Prezi) and a handout for the rest of the class.

Deadline: 29.11.2019 (in class)

Formal requirements: 15 minutes presentation with slides

Submission: Presentation (in-class)

Should be included:

- What is your group's over-arching research question/thematic focus?
- What are your research questions?
- Which data did you use?
- What is the data about?
- What is the unit of analysis and the time frame?
- How was the data collected?
- What does the data tell you about your research question(s)? (Come up with some descriptive statistics, some visualization, ...)
- Where can we access the data? (Reference, format)
- Brief concluding discussion
- Potential problems

5. Handout

To provide your peers a good idea of different data sets, I ask you to also submit a handout on the data set you used for your course project.

Deadline: 30.11.2019 (23:59)

Formal requirements: 1-page handout on the data set

Submission: As PDF on ILIAS

Should be included:

- What is the data about?

- What is the unit of analysis and the time frame?
- How was the data collected?
- Where can we access the data? (Reference, format)

Final examination

Term paper

The deadline for the term paper is January, 10 2020 (23:59; upload on ILIAS as PDF).
For the purpose of consistency, please use a *coherent citation style* (see for example the APSA citation guidelines (uploaded on ILIAS) or Mannheim's "Arbeitshilfen-Reader" (also uploaded on ILIAS; unfortunately only in German)).

General requirements

1. Formal requirements

    - Cover page: university, chair, semester, course type and title, name of the instructor, title of the term paper, date; name, contact information, program and semester of the student; indicate word count
    - Table of contents: complete structure of the final paper including page numbers and informative headlines
    - Length: 5 pages in total (max. 1,500 words, will be checked); 1,500 words include only the content (*no* cover page, table of content, references, or appendices)
    - Signed statutory declaration (both in English and in German, as uploaded on ILIAS)
    - Layout: 1.5 line spacing, font size 11, Times New Roman, pagination of the text
    - Orthography and grammar
    - Formatted as a PDF or html
    - **Submit your datafiles and Rmarkdown file as well**

2. Scientific standards

    - Appropriate use and formal correctness of references, e.g. for example according to APSA or Mannheim's citations style (as uploaded on ILIAS)
    - Independent research
    - List of references: coherent citation style, e.g. for example according to APSA or Mannheim's citations style (as uploaded on ILIAS).

Content

1. Introduction

    - Relevance
    - Analytical research question

- Overview of the paper's structure

2. Theoretical part: Theory and Hypotheses
   - Briefly: Theoretical argument and argumentative structure
   - Generate one hypothesis
3. Empirical part
   3.1 Research Design
      - Discussion of sample and time span/case selection
      - Discussion operationalization
   3.2 Discussion
      - Testing hypotheses using data visualization
      - Describe and interpret your results
4. Conclusion
   - Summary and critical discussion
   - Answering the research question