

o#+TITLE:COURSE OVERVIEW



## What is an "Advanced intro to data science"?

- Learn R programming (functions, conditions, loops, utilities)
- Improve performance (`data.table`)
- Explorative Data Analysis (EDA) with the Tidyverse
- Transcend R programming (command line data science)

## What will you do in this course?

- Mixture of DataCamp lessons and current topics
- Last course in spring'22 was too "tidy"

## How will you be evaluated?

- All course requirements have deadlines
- Late submissions will be penalized (loss of points)



Figure 1: Map of Middle Earth

WEEK	DATE	TOPICS & ASSIGNMENTS	TESTS
1	Jan 11,13	Calling functions	
2	Jan 18,20	Intermed R: Conditionals	Test 1
3	Jan 23,25,27	Intermed R: Loops	Test 2
4	Jan 30, Feb 1,3	Intermed R: Functions	Test 3
5	Feb 6,8,10	Writing functions	
6	Feb 13,15,17	Intermediate R: apply	Test 4
7	Feb 20,22,24	Intermed R: Utilities	Test 5
8	Mar 1,3	Intro to Tidyverse	Test 6
9	Mar 6,8,10	Natural language processing	
10	Mar 13,15,17	EDA in R: Categorical Data	Test 7
11	Mar 27,19,31	EDA in R: Numerical Data	Test 8
12	Apr 3,5	EDA in R: Numerical Summaries	Test 9
13	Apr 10,12,14	EDA in R: Spam Detection	
14	Apr 17,19,21	Machine learning	Test 10
15	Apr 24,26,28	Command line data science	
16	May 1, 3	Project presentations	

Figure 2: Syllabus, Canvas ([lyon.instructure.com](http://lyon.instructure.com)) or GitHub ([github.com/birkenkrahe/ds2](https://github.com/birkenkrahe/ds2))

REQUIREMENT	UNITS	PPU	TOTAL	% of TOTAL
Final exam	1	100	100	20.
Home assignments	10	10	100	20.
Class assignments	10	10	100	20.
Project sprint reviews	5	20	100	20.
Multiple-choice tests	10	10	100	20.
TOTAL			500	100.

Figure 3: Source: syllabus, Canvas ([lyon.instructure.com](https://lyon.instructure.com)) or GitHub ([github.com/birkenkrahe/ml](https://github.com/birkenkrahe/ml))

- Final exam will be sourced by term test questions
- The project topic can come from any of the course sub-topics
- The project deliverable is a working **literate** program

### Which tools are you going to use?

- DataCamp courses (10 weekly home assignments)
- GitHub repository (all course materials except tests)
- GNU Emacs + ESS + R (literate programming environment)
- Canvas (learning management system)

### How can you register at DataCamp?

- You find the invitation link to the group for Spring 23 in Canvas.
- You will automatically be subscribed to the Digital Humanities team



Figure 4: Unsplash, workshop



	Intermediate R Conditionals and Control Flow Chapter	Team	Active	Jan 20, 13:00 CST
	Intermediate R Loops Chapter	Team	Active	Jan 27, 13:00 CST

Figure 5: DataCamp assignments for January 2023

- If you are in more than one course, I will add you later manually
- These accounts will be valid until July 8, 2023 only

## When is the first assignment due?



- The first DataCamp home assignment is due on January 20. For late submissions, you lose 1 point per day (out of 10 possible points)
- The first in-class assignment is due on January 20. For late submissions, you lose 1 point per day (out of 10 possible points)
- We'll write the first weekly multiple-choice test on January 20, covering all of the material that was covered until then.

## What should we do as a project?

- For your term project, **pick an R package** to explore and explain
- The exploration should include **examples and tests**
- Your final deliverable should be a **literate program** (Org-mode)
- The final presentation should include **exercises** for the audience
- We can only accommodate **four (4) projects** altogether
- You need to deliver: a proposal, a literature review, an abstract





Which R packages are suitable for student projects?



There are many R packages that could be suitable for student projects, depending on the subject area and the goals of the project. Some packages that might be particularly useful for student projects include:

- ``ggplot2``: a popular package for creating data visualizations
- ``dplyr``: a package for data manipulation and preparation
- ``tidyr``: a package for cleaning and reshaping data
- ``lme4``: a package for fitting linear mixed-effects models
- ``caret``: a package for training and evaluating machine learning models
- ``shiny``: a package for building interactive web applications with R

These are just a few examples, and there are many other packages that could be useful for student projects as well. It's always a good idea to explore the available packages and see what might be useful for your specific project.

Figure 6: R packages suitable for student projects by ChatGPT

What are you looking forward to?



- Developing lectures and exercises away from DataCamp
- Learning more about R packages through your projects
- Having fun with R programming and real applications

Next



Figure 7: Our next topic is "Argument matching"