

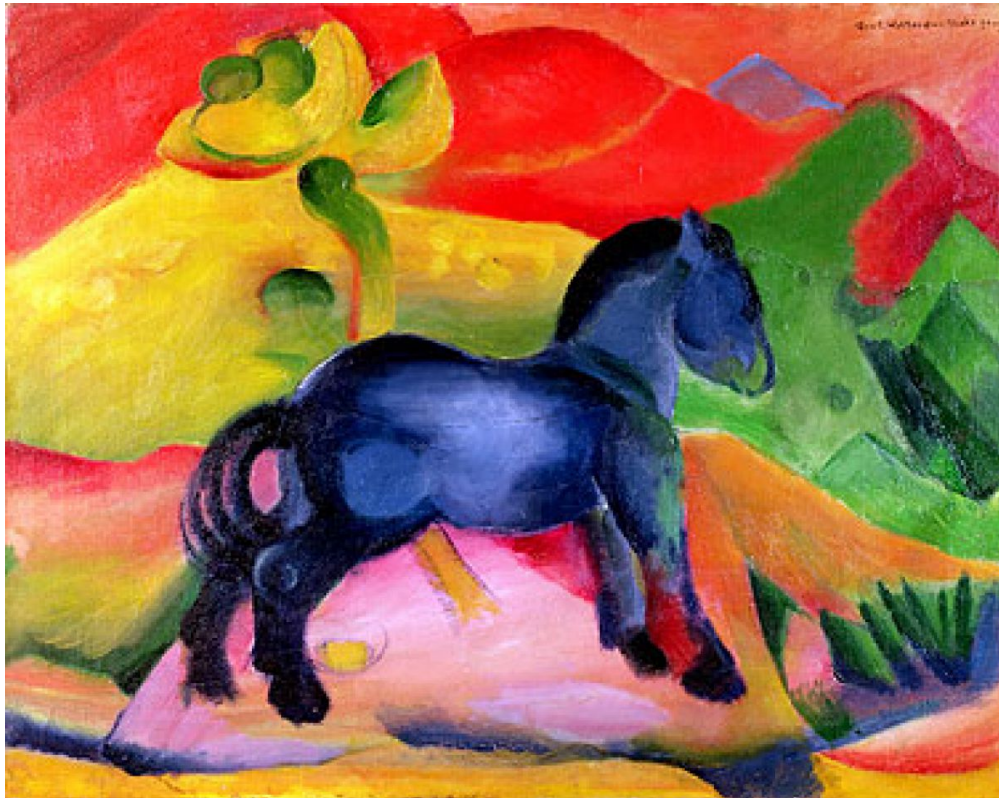
# Course overview

Introduction to data science (DSC 105) Fall 2023

## Table of Contents

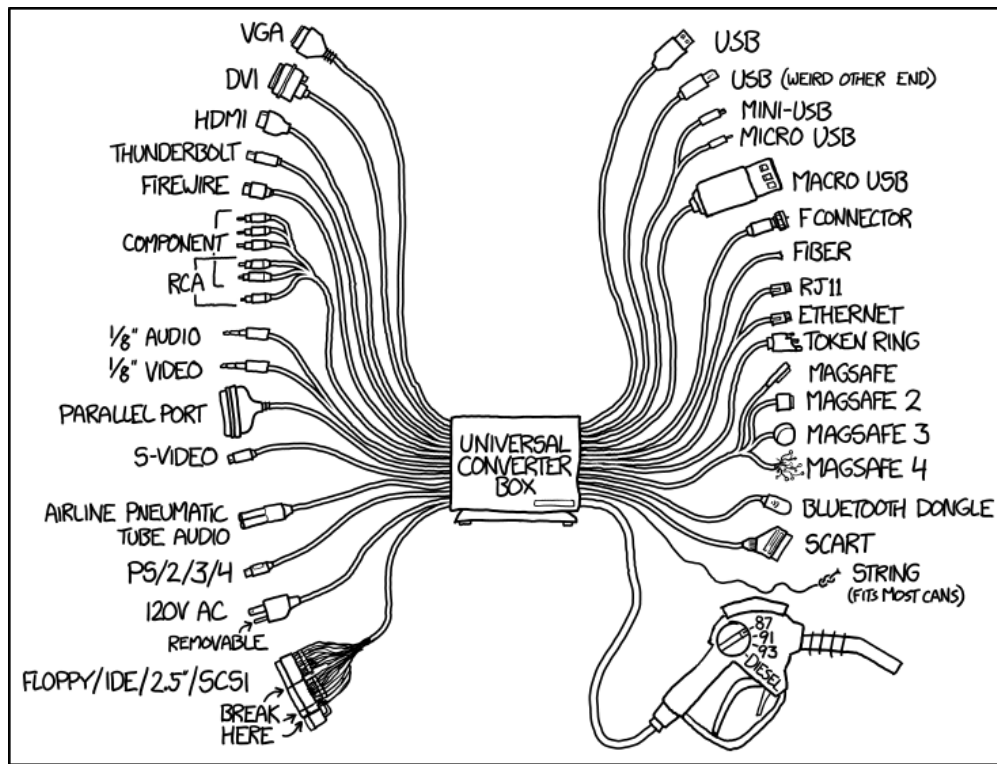
- [About me](#)
- [About you](#)
- [Course syllabus](#)
- [Home assignments](#)
- [Practice - first R program](#)
- [Next: DataCamp Workspace](#)

## About me



- PhD: theoretical particle physics
- Industry: Knowledge management
- [Teaching data science](#) / [Coding w/AI](#)

## About you



1. Why are you here?
2. What would delight you?
3. What would disappoint you?
4. Where are you headed?

## Course syllabus

The screenshot shows the Canvas LMS interface for the course 'DSC 105 01 > Syllabus'. The left sidebar contains navigation links: Account, Dashboard, Courses (selected), Calendar, Inbox, History, and Help. The main content area is titled 'Course Syllabus' and includes a 'Jump to Today' link. Below the title is a 'Table of Contents' with links to 11 sections: 1. General Course Information, 2. Objectives, 3. Target audience, 4. Student Learning Outcomes, 5. Course requirements, 6. Grading, 7. Dates and class schedule, 8. Learning management system, 9. DataCamp, 10. GitHub, and 11. Lyon College Standard Policies (Fall 2023). The first section, '1. General Course Information', is expanded, showing details such as course title, number, meeting times, place, professor, office, phone, hours, and textbook.

2023-2024 - Fall Semester

Home

Assignments

Grades

Syllabus

CoursEval

Zoom

Immersive Reader

## Course Syllabus

Jump to Today

### Table of Contents

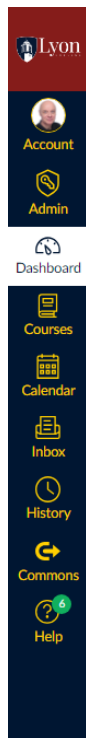
- 1. [General Course Information](#)
- 2. [Objectives](#)
- 3. [Target audience](#)
- 4. [Student Learning Outcomes](#)
- 5. [Course requirements](#)
- 6. [Grading](#)
- 7. [Dates and class schedule](#)
- 8. [Learning management system](#)
- 9. [DataCamp](#)
- 10. [GitHub](#)
- 11. [Lyon College Standard Policies \(Fall 2023\)](#)

### 1. General Course Information

- Course title: Introduction to data science
- Course number and section: DSC 105.01
- Meeting Times: Mon-Wed-Fri from 13:00-13:50 am
- Meeting place: Derby Science Building computer lab room 209
- Professor: Marcus Birkenkrahe
- Professor's Office: Derby Science Building 210
- Phone: (870) 307-7254 (office) / (501) 422-4725 (private)
- Office hours: Mon/Wed 2-3pm, Tue 4-4.45pm, Thu 9.30-10:45am
- Textbook: The Book of R (Part I, ch. 1-8) by T Davies ([NoStarch, 2016](#))  
- reserved copy available in the library

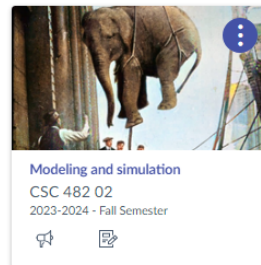
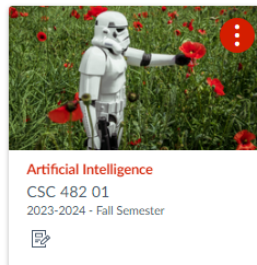
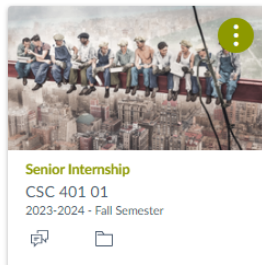
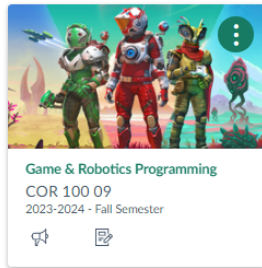
- [Syllabus on Canvas](#)
- General information & standard policies
- Course information (grading, schedule)

## Canvas LMS



## Dashboard

### Published Courses (8)



- All [grades](#) should be visible at all times
- Control your own [notifications](#) (email)
- Important course [links](#) on a page
- Add all your Canvas calendars to Google

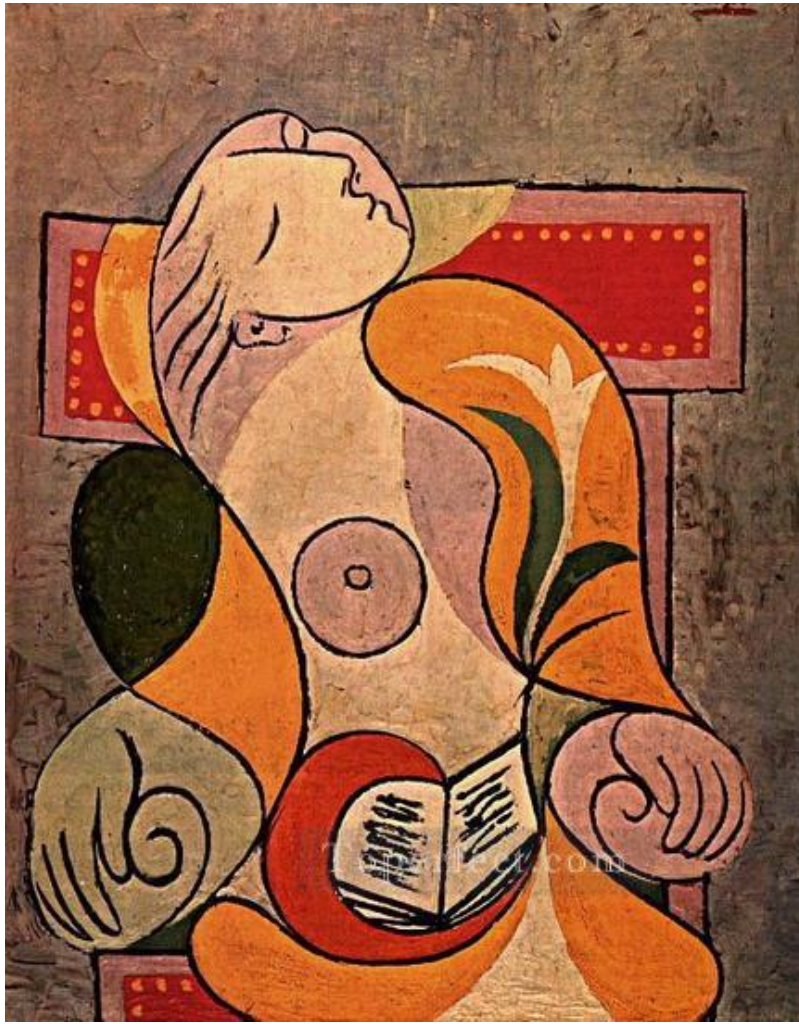
## Course topics



1. The R statistical programming language
2. Basics of data visualization with R
3. Software development methods
4. From R to Python ... and back again

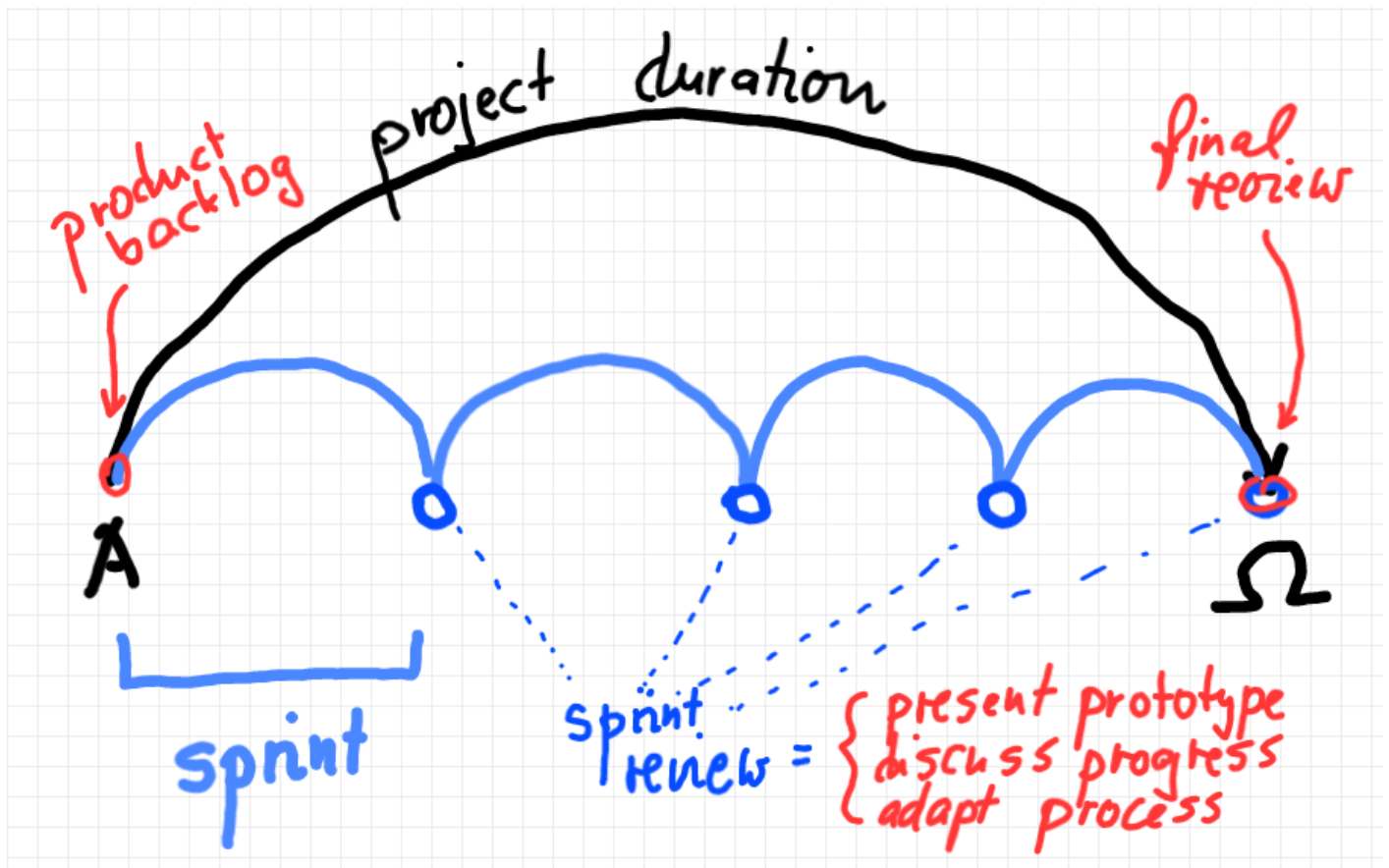
## **Video lectures**





- [Emacs + Org-mode + R](#) (Tutorial videos Spring '22)
- [Introduction to R: installation and shell](#)
- Vectors in R ([part 1](#), [part 2](#), [part 3](#))
- [Data frames, matrices, lists, factors in R](#)
- [Data frames in R](#)
- [Base R plotting](#)
- [Plotting with ggplot2](#)
- [Data import](#) with R
- [RStudio R Notebooks and literate programming](#)

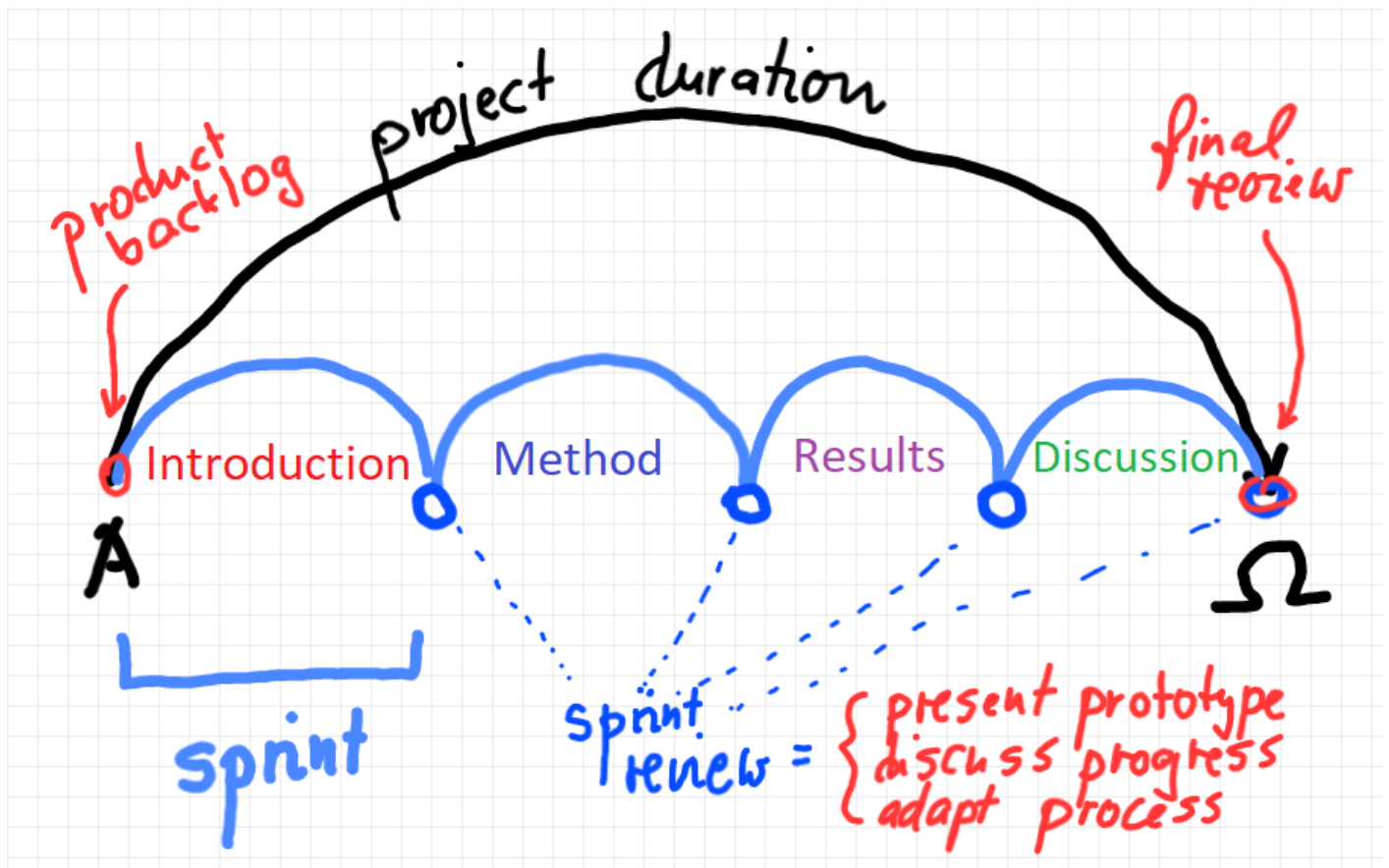
**Agile project (with "Scrum")**



- The project makes up 25% of your final grade for this course.
- What is a Scrum project? ([GitHub FAQ](#))
- Do you have examples for data science projects? (FAQ)
- Can you do a project as an absolute beginner? (FAQ)

**Note:** the first [sprint review](#) is in the week of Sep-11. Use it to present your initial results (see FAQ on [what to deliver](#), and [1st sprint review](#)).

## IMRaD and Scrum



- Introduction (research question - what you want to find out)
- Method (how you want to do it)
- Results (what you found out)
- Discussion (what it means)

([Video: Research Writing with IMRaD](#))

**Many project opportunities**




**DATA SCIENCE COMPETITIONS**

## Grow your data science skills by solving real-world problems

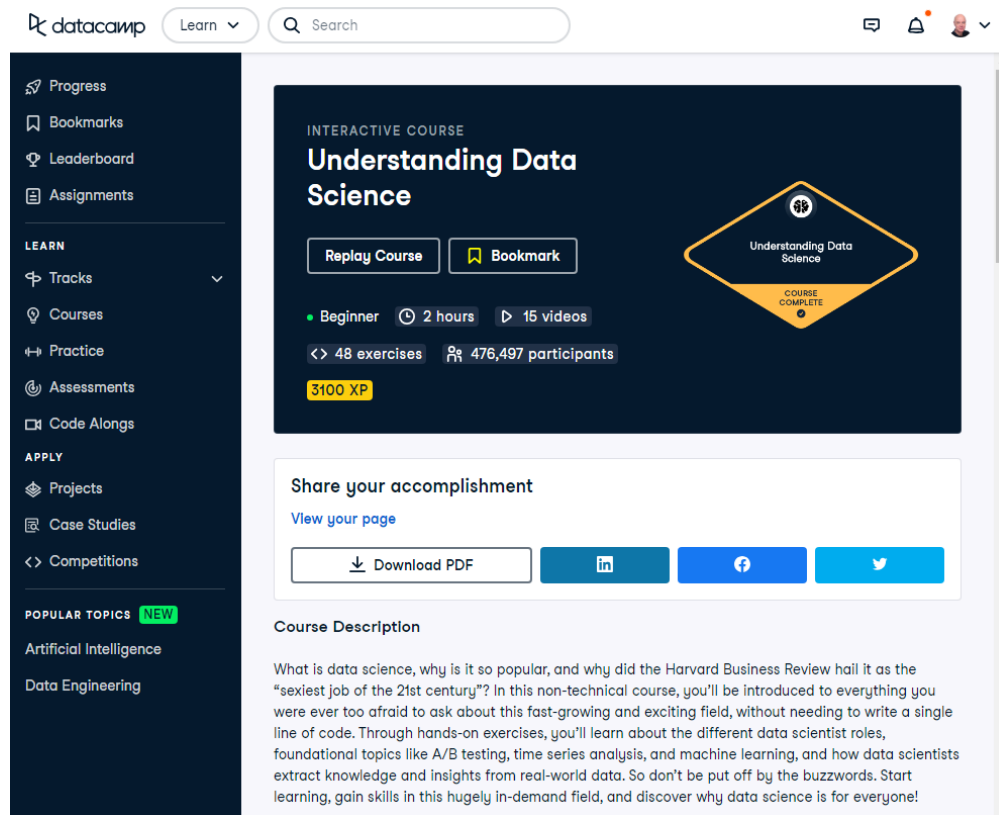
- 🏆 Apply your knowledge to real-life scenarios
- 🔍 Discover best problem-solving practices
- 📁 Build a portfolio of shareable data science work
- 💰 Gain recognition and earn cash rewards

**Join A Competition**



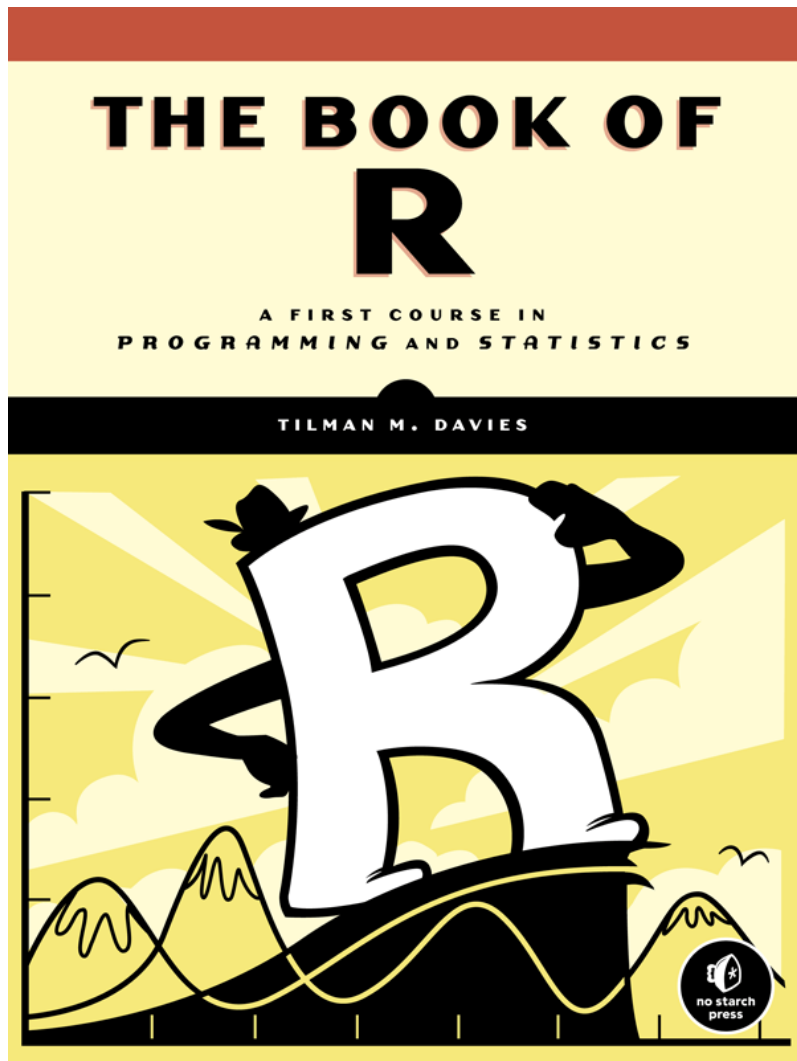
- Explore and document an R package
- Document an [extended analysis example](#)
- Explore a data set of your choice
- Complete a DataCamp competition
- See [DataCamp projects](#) for examples
- You can branch out: SQL, Python, Java etc.
- See GitHub issues for examples (e.g. [whale song](#))
- Double/triple up if you're in > 1 of my courses
- Use problems from other courses for your project, e.g. data collected by yourself, or data in economics, business, art etc.

## Introduction to DataCamp



- [DataCamp](#) is a data science learning platform
- Access for you is free (academic alliance)
- All assignments are DataCamp assignments
- Assignments are drawn from 4 courses
  1. Data science for everyone
  2. Introduction to R
  3. Data manipulation with data.table
  4. Python for R users
- Complete them on time to get full points
- Completed DataCamp courses can [support your resume](#)

**NEXT Introduction to the textbook**



- R is *FOSS* with focus on stats and graphics
- Davies' "[Book of R](#)" is extensive (832p.)
- You don't have to read along but it might help
- Many other tutorials and textbooks available
- The best short online tutorial: [Matloff's "fasteR"](#)
- Beware of ideologies (cp. Matloff's "[TidyverseSceptic](#)")

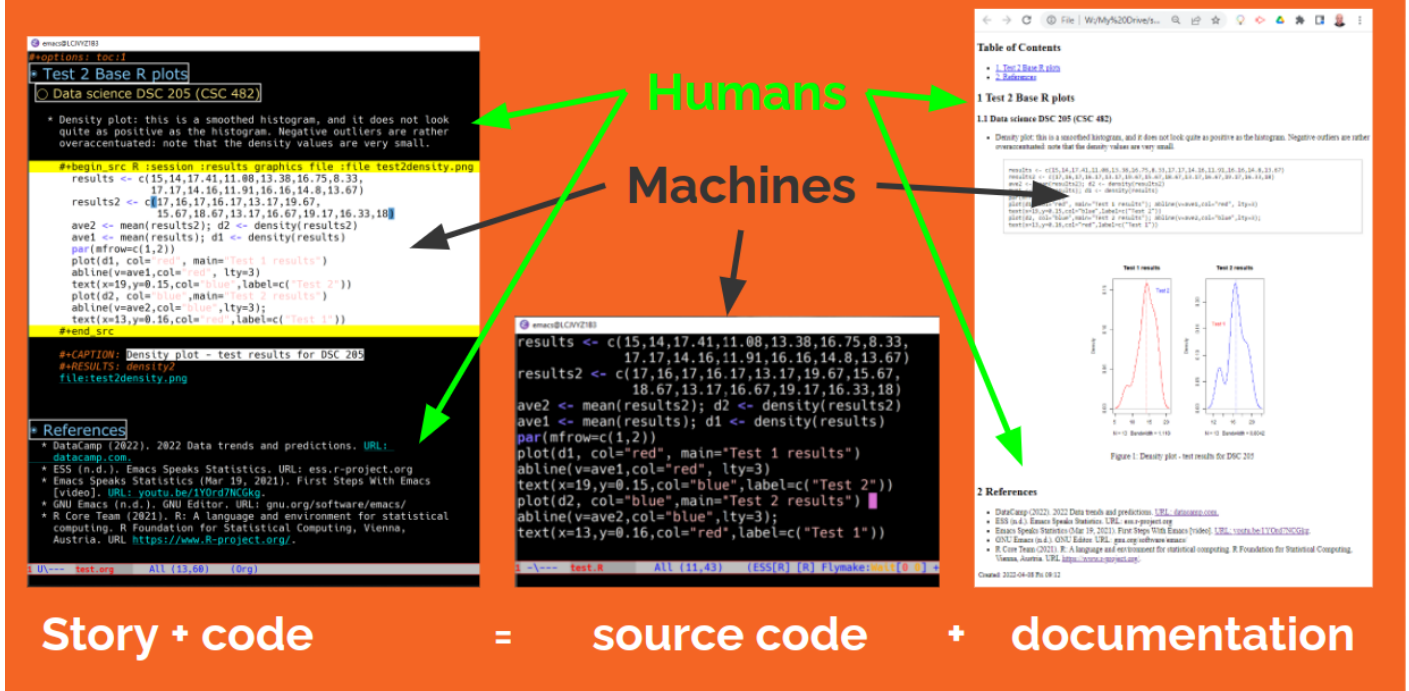
**Good-bye Emacs**



- Emacs: self-documenting, extensible *FOSS* text editor
- Process, file and package management (like an OS)
- *Literate programming* environment for 43 languages
- *IDE* for R programming and *REPL* for interactive coding

## Literate programming

# What is literate programming?





Source: "[Teaching data science with hacker tools](#)" (2022)

- Common practice among data scientists
- *Paradigm* behind interactive computing notebooks
- Useful when learning any programming language

Tests (multiple-choice)







14:18  
Time Remaining

Return


Submit




Account




Dashboard




Courses




Calendar




Inbox



History



Commons



Help

## Entry quiz

Entry quiz (**not graded**) to see what you already know (if anything) about data science! This course assumes no prior knowledge - the quiz only for me to find out what you already know, and for assessment purposes (you'll get this quiz again at the end). Don't worry if you cannot answer any of the questions - all of this will be taught in the course!

- Questions may have one or more than one correct answer.
- Partial credit is allowed.
- Questions are not timed.

11 point

**What is the purpose of data science?**

☐ Decision support

☐ Machine learning

☐ Data literacy

☐ Data visualization

21 point

**Which of these are skills that data scientists really need?**

☐ Programming skills

☐ Database management

☐ Math and statistics

☐ Domain knowledge

- Tests have to be completed online, are timed, and have a deadline; after the deadline, you can play them an unlimited number of times
- There will be a revision quiz on Canvas every week, consisting of 5-10 multiple choice, matching and true/false questions.
- A subset of the test questions will form the final exam (25% of your final grade) - we will practice in the last week before the exam.

## Home assignments



- Register with DataCamp now
- Read "[Adventures with R](#)" by Monday 28 August 1 PM ([Canvas](#))

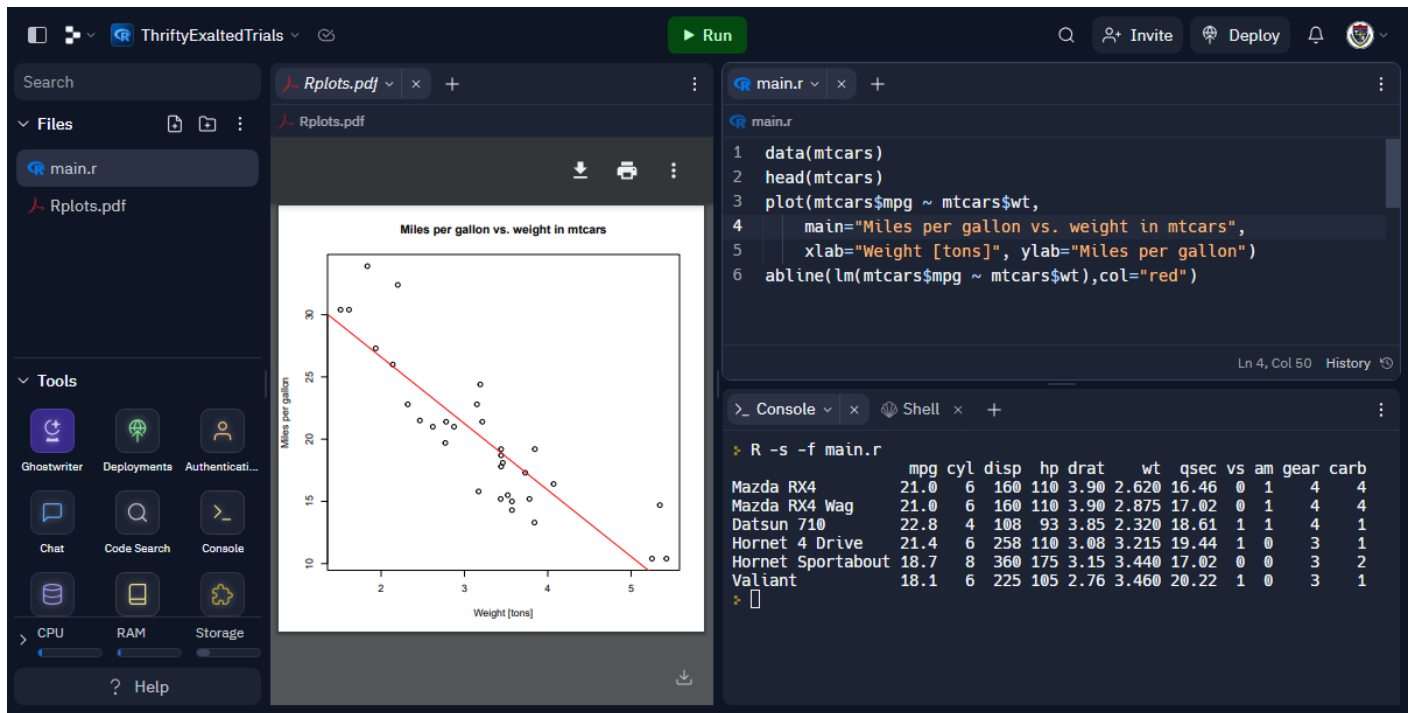
We're going to look at DataCamp workspace tomorrow - here is a good short summary video of the AI coding capabilities: <https://youtu.be/9XHop0xTTCM?si=bY4UsJmSrzcCYMUI> - for this particular tool: what're your thoughts on this? How is it going to change coding and data science? For the better or for the worse?

- Complete 1st DataCamp assignment by Sunday 11.59 PM

### [Introduction to data science](#)

- Data science definition
- Data science workflow
- Application to real-world problems
- Different professional data science roles


## Practice - first R program







- In your browser, go to [replit.com](https://replit.com)
- Register using your Lyon email
- Create a REPL using an R template

```
print("hello, world") data(mtcars) head(mtcars) plot(mtcars$mpg ~ mtcars$wt, main="Miles per gallon vs.
weight in mtcars", xlab="Weight [tons]", ylab="Miles per gallon") abline(lm(mtcars$mpg ~
mtcars$wt),col="red")
```

## Next: DataCamp Workspace


Workspace


Lyon College Data Science
Premium

Create workspace

Overview

Workspaces

Bookmarks

Databases

Templates


Datasets

Community


Code Alongs

### Workspaces


Search workspaces




**DataCamp Python Demo (problem)**


Marcus Birkenkrahe in Lyon College Data Science Fall 2023


0
...




**DataCamp Python demo (solution)**


Marcus Birkenkrahe in Lyon College Data Science Fall 2023


Aug 22 at ...
PRIVATE
0
...




**Untitled workspace**


Marcus Birkenkrahe in Lyon College Data Science Fall 2023


Aug 22 at ...
PRIVATE
0
...




**Unicorn Companies in Python (Org-mode)**


Marcus Birkenkrahe in Lyon College Data Science Fall 2023


Aug 19 at ...
PRIVATE
0
...




**[mod] Setting up the notebook environment**


Marcus Birkenkrahe in Lyon College Data Science Fall 2023

Aug 11 at ...
PRIVATE
0
...



**AI\_python**


Marcus Birkenkrahe in Lyon College Data Science Fall 2023

Aug 11 at ...
PRIVATE
0
...

FILTERS

Sort by

☒ Last opened
 ☐ Last modified
 ☐ Name
 ☐ Last created

Owned by

☐ Anyone
 ☒ You
 ☐ Others

- Simple example: <https://tinyurl.com/SimpleRexample>
- Turn mtcars into code font and view the text cell.
- Load the mtcars data set and run the code cell.
- Print the first few lines of mtcars.
- Open the help for mtcars.
- Plot miles-per-gallon vs. weight for all cars.
- Compute and draw a red trendline through the points.
- Code:

```
data(mtcars)
head(mtcars)
#help(mtcars)
mpg = mtcars$mpg
wt = mtcars$wt
plot(x=wt, y=mpg)
abline(lm(mpg~wt), col="red")
```

Author: Marcus Birkenkrahe

Created: 2023-09-20 Wed 12:33