DS Agenda

Agenda for CSC482/DSC205 Introduction to Advanced Data Science Spring 2022

README

This file contains the agenda overview (what I had planned), the objectives (what we managed to do) and (much of the) content of each taught session of the course. I want to avoid splitting the content up over many files - so that you have to navigate as little as possible (like a book)!

The companion file to this file, less structured and with the captain's log, is the notes.org file.

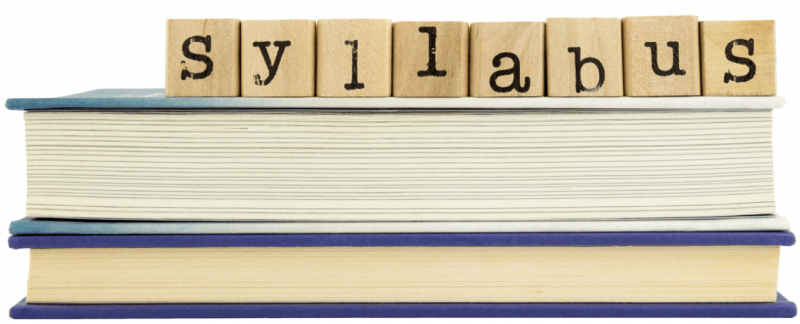
Course introduction - w1s1 (01/12/22)

Welcome

- Aspirations - changes spring 2022
- Ambitions - program 2021-2023
- Antagonization - new data science credo
- Syllabus - this course
- DataCamp assignments
- GNU Emacs Org-mode

  "After a course is launched, we don't consider it to be complete: the launch is just the start of data collection." Richie Cotton, DataCamp

Syllabus



- Syllabus in Schoology
- Syllabus in GitHub
- Schedule in GitHub

Aspirations (Changes in Spring 2022)

Cp. Good-bye fall 2021

FALL 2021	SPRINT 2021
Base R (stick shift) instead of "TidyVerse" (automatic)	Adding the "Tidyverse"
Use of interactive notebooks (literate programming!)	Intro to RStudio IDE and Emacs
Use GitHub as a code and materials repository	GitHub repo
Create lots of (ungraded) tests	Graded quizzes and tests
Use of DataCamp assignments	DataCamp assignments
Avoid mathematics as much as possible	No math
Reuse tests for the final exam	Reuse quizzes for final exam
Let students pick their own projects	No projects (only optional)

Ambitions (DS program 2021-2023)

CLASS	CODE	TERM	Topics
Data Science Tools and Methods	DSC 101	Fall 2021	R, Basic EDA, Base R
Introduction to Advanced Data Science	DSC 205	Spring 2022	R, Advanced EDA, Tidyverse, shell
Database Theory and Applications	CSC 330	Spring 2022	SQL, SQLite
Operating Systems	CSC 420	Spring 2022	Bash, awk, sed, regular expressions
Data Visualization	DSC 302	Fall 2022	D3, Processing, Javascript, Bokeh
Machine Learning	DSC 305	Spring 2023	Predictive algorithms, neural nets
Digital Humanities	CSC 105	Spring 2023	Data science applications

DataCamp

Intermediate R Conditionals and Control Flow Chapter	Team	Active	Jan 31, 15:00 CST	0	0	0% View
Intermediate R Loops Chapter	Team	Active	Feb 7, 15:00 CST	0	0	0% View
Intermediate R Functions Chapter	Team	Active	Feb 14, 15:00 CST	0	0	0% View
Intermediate R The apply family Chapter	Team	Active	Feb 21, 15:00 CST	0	0	0% View
Intermediate R Utilities Chapter	Team	Active	Feb 28, 15:00 CST	0	0	0% View
Introduction to the Tidyverse Data wrangling Chapter	Team	Active	Mar 7, 15:00 CST	0	0	0% View
Introduction to the Tidyverse Data visualization Chapter	Team	Active	Mar 14, 15:00 CDT	0	0	0% View
Introduction to the Tidyverse Grouping and summarizing Chapter	Team	Active	Mar 28, 15:00 CDT	0	0	0% View
Introduction to the Tidyverse Types of visualizations Chapter	Team	Active	Apr 4, 15:00 CDT	0	0	0% View
Exploratory Data Analysis in R Exploring Categorical Data Chapter	Team	Active	Apr 11, 15:00 CDT	0	0	0% View
Exploratory Data Analysis in R Exploring Numerical Data Chapter	Team	Active	Apr 20, 15:00 CDT	0	0	0% View
Exploratory Data Analysis in R Numerical Summaries Chapter	Team	Active	Apr 25, 15:00 CDT	0	0	0% View
Exploratory Data Analysis in R Case Study Chapter	Team	Active	May 2, 15:00 CDT	0	0	0% View

- Why are we using it?
- How are we using it?
- What will you have to do?

Antagonization

A new credo.

> “Getting it right is crucial when people’s lives are affected.” -Jonathan Steinhart

Figure 4: Lego fencing (Source: Unsplash)

What's next?

- See schedule:
  - install R / Emacs IDE - may do this together
  - Entry quiz (by Tue 18 Jan) - you should get > 50%
- Watch online lecture on "Systems" (to be published)
- Online followup notes (`notes.org` in GitHub)
- See you Friday 14-Jan online!
- Hopefully Wednesday 19-Jan in class!

Installing R / Windows PATH - w1s2 (01/14/22)

Overview

HOW	WHAT
Practice	Install R from CRAN
	Set PATH environment variable
	Test R in terminal and GUI
	Install GNU Emacs + ESS (FAQ)
	Set PATH environment variable
	Test R in Emacs
	Set .emacs init file
	Create Org file

HOW	WHAT
	Run R code blocks in an Org file

Objectives

- [X] Install R
- [X] Set PATH environment
- [X] Test R in terminal and GUI
- [ ] Install GNU Emacs
- [ ] Test R in Emacs

Installing and setting up GNU Emacs - w2s3 (01/19/22)

I'm back

Figure 6: "I'm back, baby."

Overview

HOW	WHAT
Review	Entry quiz
	Quiz 1 + feedback + discussion

HOW	WHAT
Practice	Install GNU Emacs + ESS (FAQ)
	Set PATH environment variable
	Test R in Emacs
	Set .emacs init file

Objectives

- [X] Install GNU Emacs + ESS
- [X] Set PATH environment to run R in Emacs
- [X] Test R in Emacs (however, see course FAQ)
- [ ] Configure Emacs

Next

- Create Emacs Org file
- Run R code blocks in an Org file
- DataCamp assignments beginning soon!

Understand Emacs Org-mode - w2s4 (01/21/22)

Overview

HOW	WHAT
Lecture/Demo	GNU Emacs Org-mode
Practice	GNU Emacs Tutorial (gh)
Homework	Set `emacs` init file
	Create `.org` file
	Run code in an `.org` file

Objectives

- [X] Understand what Org-mode is and what it's for
- [ ] Create an `.emacs` init file for GNU Emacs
- [ ] Create an Org file
- [ ] Run a code block in your Org file

Next

- Create Emacs Org file
- Run R code blocks in an Org file
- DataCamp assignments beginning soon

Customizing Emacs (init file) - w3s5 (01/24/22)

Overview

HOW	WHAT
Review	Quiz 2
Lecture/Demo	GNU Emacs Org-mode (Part 2)
	New: video playlist
Practice	GNU Emacs Tutorial cont'd (gh)
- Package manager	`M-x package-list-packages RET`
- Start R shell in Emacs	`M-x R` (R must be installed & in the `PATH`)
- Add init file	.emacs sample file (GitHub)
Assignment[1]	Set `emacs` init file
Assignment	Read 2022 Data trends and predictions
	Put your summary thoughts in an `.org` file
	Check the FAQ "How should you read?"

Objectives

- [X] Create an `.emacs` init file for GNU Emacs
- [ ] Create an Org file
- [ ] Run am R code block in your Org file

Reading assignment

- Read "2022 Data trends and predictions" (DataCamp, 2022).
- Prepare for discussion in class:
  - Which quantitative and which qualitative predictions were made?
  - What do you think how valid these predictions are?
  - Put your thoughts in an Org-mode file (filename = `YourName.org`)
  - Upload your submission to assignment/2022_predictions on GitHub

To identify yourself, use the `#+AUTHOR:` option. You can see how

this works from the options in the header of this `README.org` file.

There is no upper or lower limit on the number of words. The main point is to create a proper Org-mode file.

Next

- Create Org-mode file with R code in it and run it
- Org-mode assignment
- DataCamp assignments beginning soon (due Jan 31)

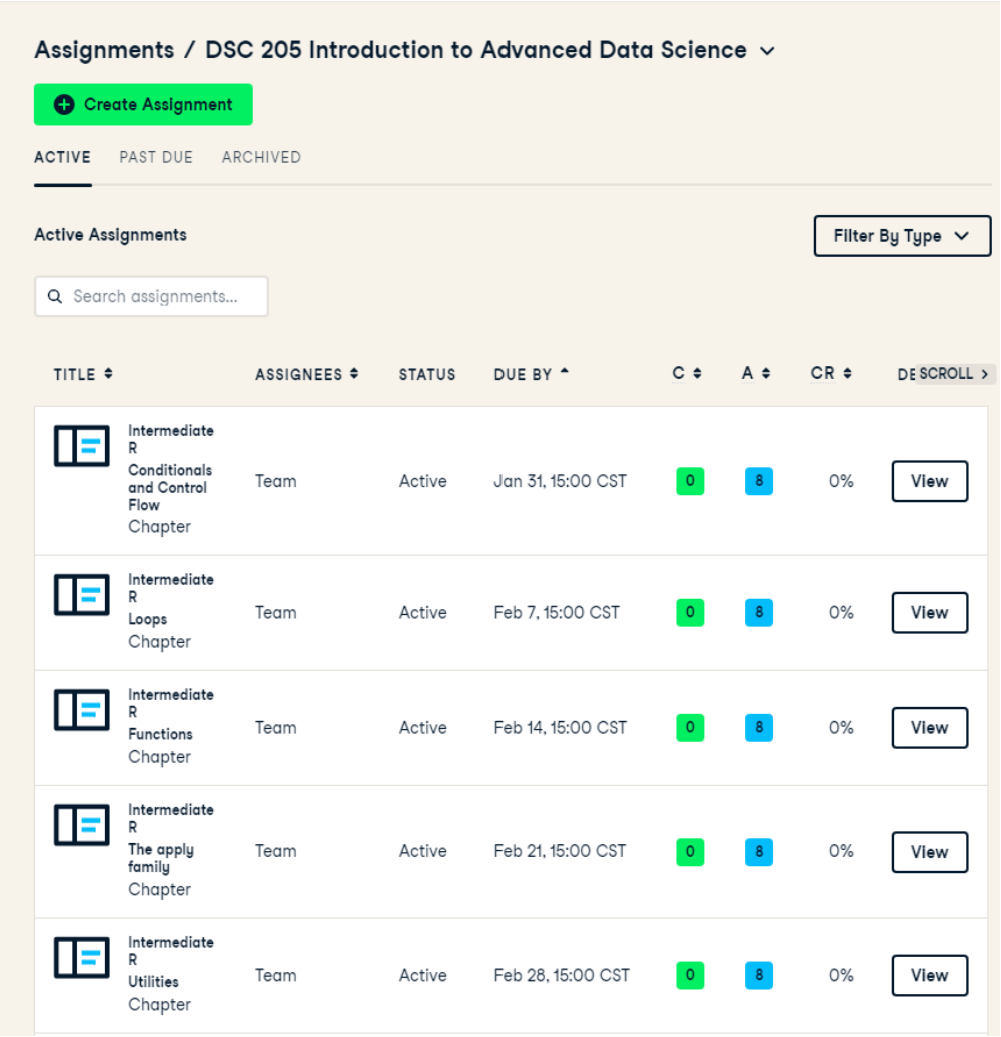


Figure 7: DataCamp assignments

Running code in Org-mode 1 - w3s6 (01/26/22)

Overview

HOW	WHAT	Link
Preview	DataCamp course "Intermediate R"	datacamp.com
Demo	Creating an Emacs Org-mode file with code and run it	README.org
Practice	Create Org-mode file with an R code block	

Objectives

- [X] Understand DataCamp assignment 1
- [X] Create an Org file
- [X] Run an R code block in your Org file

Next

- Submit Org-mode assignment in Schoology
- DataCamp assignments due Jan 31

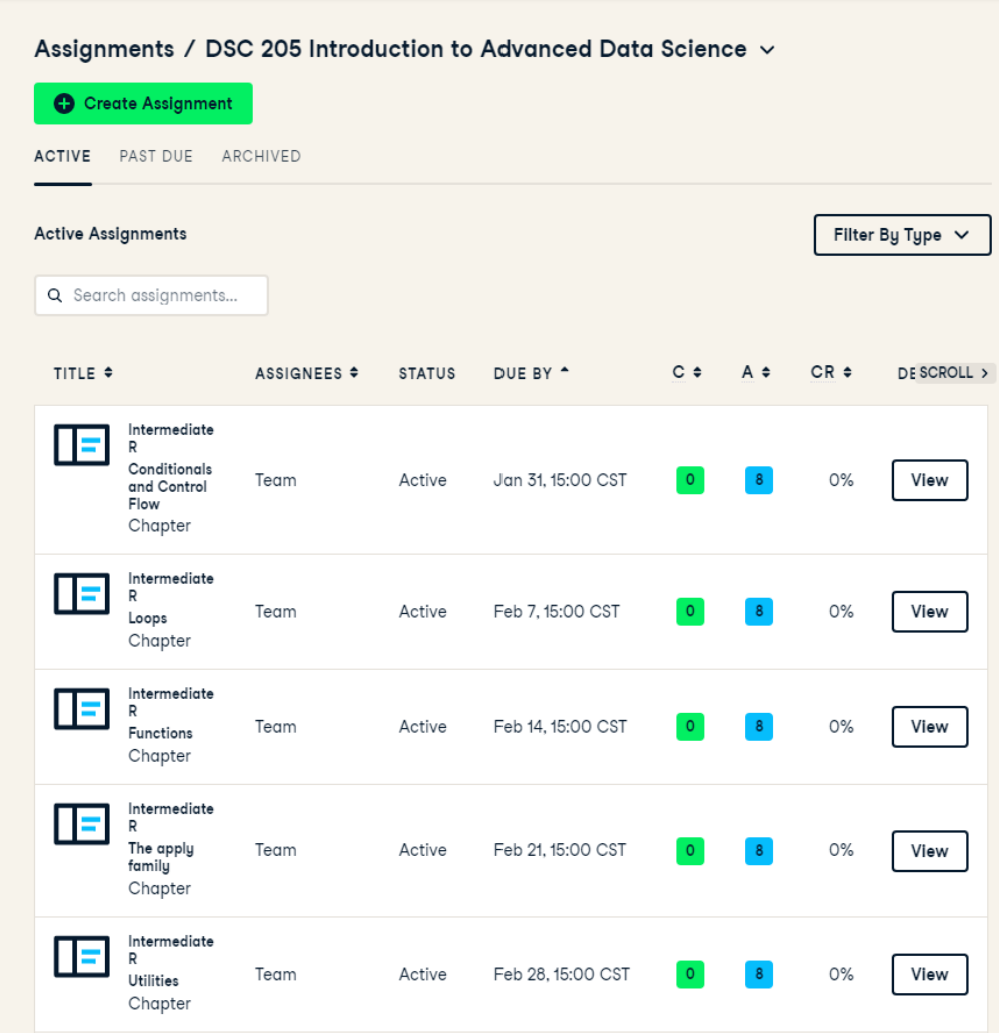


Figure 8: DataCamp assignments

Running code in Org-mode 2 - w3s7 (01/28/22)

1. We continue where we left it last Wednesday
2. Fixing the .emacs problem on Windows lab computers

3. Change of some deadlines - to finish basic Emacs training

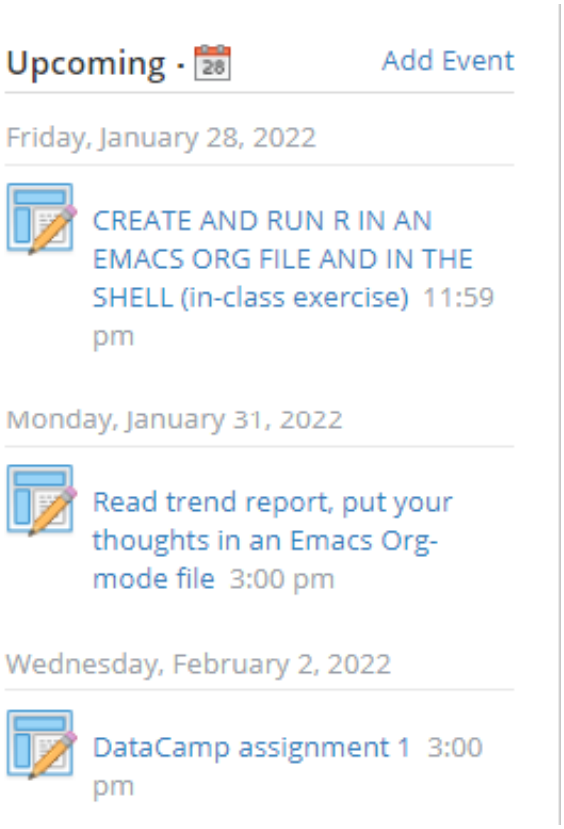


Figure 9: deadline changes in Schoology

4. Finish (expanded) Org-mode assignment
5. Submit results to Schoology.

Org-mode lab session - w4s8 (01/31/22)

Figure 10: Teaching Emacs on Dagobah

We will hold a special lab session tomorrow, Monday 31 January 3-3.50 PM, to sort out any issues related to Emacs and R. Bring your own PC to the session, or work on a lab desktop. I will spend the time going round to make sure that you can

- Install/ open / use the Emacs editor
- Create, run and tangle Org-mode files with R code
- Install / use the R programming language

- Understand the recent program assignments

  The necessary steps are also demonstrated in this tutorial video playlist.

We will continue with our regular program on Wednesday, 2nd February at 3 PM - a short quiz will be available before.

For those who know or can do all of this already: here's a second challenge (with solution) to practice while I sort others out.

What's next

- Deadline for 1st DataCamp assignment is looming (Wed 2 Feb 3pm)
- Scenario building for "Data Trends and Predictions 2022" report (assignment) - think about the 2 most important dimensions & watch this video about scenario planning
- Complete **quiz 3** including a **poll** on the prediction report before class
- Check out the webinar recording with DataCamp luminaries (panel)
- Use the breathing space to complete the Emacs tutorial (`C-h t`)

2022 Data Trends - w4s9 (02/02/22)

We meet today at 3-3.5- PM in the seminar room Lyon 106 - this room is directly adjacent to 104, our usual lab. We'll discuss the DataCamp 2022 trend report. The quiz will be available before end of the week. The planned first test (in class) will take place next Wednesday instead. (Schoology Update)

Overview

HOW	WHAT
Discussion	DataCamp 2022 report on Data Trends
Groupwork	Data science scenario planning (video)

Objectives

- [X] Understand the implications of the 2022 DataCamp trend report
- [X] Understand and apply the scenario planning technique

Next

- Quiz 3 - Conditionals and Control Workflow (DataCamp review)
- Test 1 (Friday 11 Feb 3 PM)
- Interactive R notebook - Writing functions

Studying with DataCamp - w5s10 (02/07/22)

Overview

HOW	WHAT
Review	Quiz 3 - Relational and logical operators
	How to study R with DataCamp
Preview	While and For Loops
Lecture	Writing functions in R
Test info	Test 1 on Friday 11 Feb 3.05-3.50 pm

Objectives

- [X] Review quiz 3 & how to study with DataCamp
- [X] Understand test conditions (Friday 11 Feb)
- [ ] Understand how to write functions in R (lecture)

Test 1 info

- Online in Schoology
- Entry quiz and Quiz 1-3 are not visible during the test
- The 10 hardest questions of entry quiy + quiz 1-3 (< 50%)
- 10 new questions
- Maximum time = 45 min

Next

- Interactive R notebook - loop problems
- Test 1 (Friday 11 Feb 3 PM)

NEXT w5s11 (02/09/22)

Overview

HOW	WHAT
Review	While and For loops
Lecture	Writing functions in R

Objectives

- [ ] Org-mode PROPERTY "shebang" stuff (meta data)
- [ ] Learning how to solve problems with loops (assignment)
- [ ] Understand how to write functions in R (lecture)

Next

- Test 1 (Friday 11 Feb 3-3.50 PM)
- Matthew Stewart, Stone Ward (Friday 18 Feb 3-3.50 PM)

References

Birkenkrahe (Jan 11, 2022). Interactive shell vs. interactive notebook (literate programming demo). URL: youtu.be/8HJGz3IYoHI.

Cotton (Oct 25, 2018). How DataCamp Handles Course Quality [blog]. URL: www.datacamp.com.

DataCamp (2022). 2022 Data trends and predictions. URL: datacamp.com.

ESS (n.d.). Emacs Speaks Statistics. URL: ess.r-project.org

Emacs Speaks Statistics (Mar 19, 2021). First Steps With Emacs [video]. URL: youtu.be/1YOrd7NCGkg.

GNU Emacs (n.d.). GNU Editor. URL: gnu.org/software/emacs/

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.r-project.org/.

System Crafters (Aug 1, 2021). Emacs Has a Built-in Pomodoro Timer?? [video]. URL: youtu.be/JbHE819kVGQ.

Footnotes:

[1] Submission of the assignment by Monday 24 January 3pm gives 10 extra credit points.

Author: Marcus Birkenkrahe
Created: 2022-02-08 Tue 20:56
Validate