COURSE OVERVIEW

(Data Science Tools and Methods)

MARCUS BIRKENKRAHE

Created: 2021-08-19 Do 10:50

TABLE OF CONTENTS

- What're you going to learn today?
- Who am I?
    - Science
    - Industry
    - Teaching
    - Pleasure
- What are your expectations?
- Which topics will we cover?
    - Introduction to data science
    - Introduction to R programming
    - Visualization using R
    - FasteR approach
    - Schedule (see Syllabus)
- How will we do it?
    - Classroom sessions
    - Lecture scripts with exercises (GitHub)
    - Reading suggestions
    - Video lectures (YouTube)
    - Online assignments (DataCamp)
    - Team EDA project
    - Agile project management
    - Tests and final exam
    - Podcasts and feeds
    - Summary of course activities
- What do you have to do to pass?

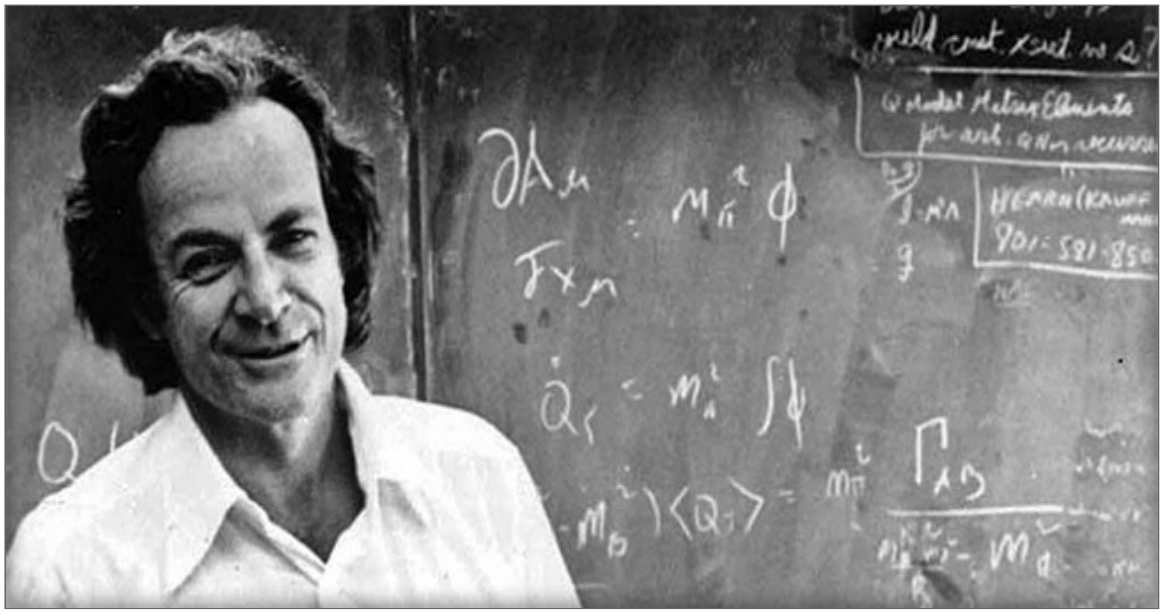
    - DataCamp assignments (> 50%)
    - Team project (> 50%)
        - What is a team project?
        - Do you have project examples?
        - Can I do a project as an absolute beginner?
    - Final exam (> 50%)
- What's next?
    - In the course
    - Your challenges
- Any questions?

WHAT'RE YOU GOING TO LEARN TODAY?

- Who is your lecturer?
- Who are you and what do you want?
- Which topics will we cover?
- How will we do it?
- What do you have to do to pass?
- What's next?

WHO AM I?

SCIENCE



- Development of WWW
- PhD theoretical particle physics
- 60 research publications
- Assoc. Ed. Int. J. of Data Science
- Ed. Board Int. J. of Big Data Mgmt.
- Scientific member **d-cube@Berlin**

INDUSTRY

- Executive at Accenture & Shell
- Coach and consultant
- Certified psychotherapist
- Startup mentor

TEACHING

- Business informatics @**HWR Berlin**
- Visiting professor of data science @Lyon
- Adviser for **CPU @LA**
- Internship supervision

PLEASURE

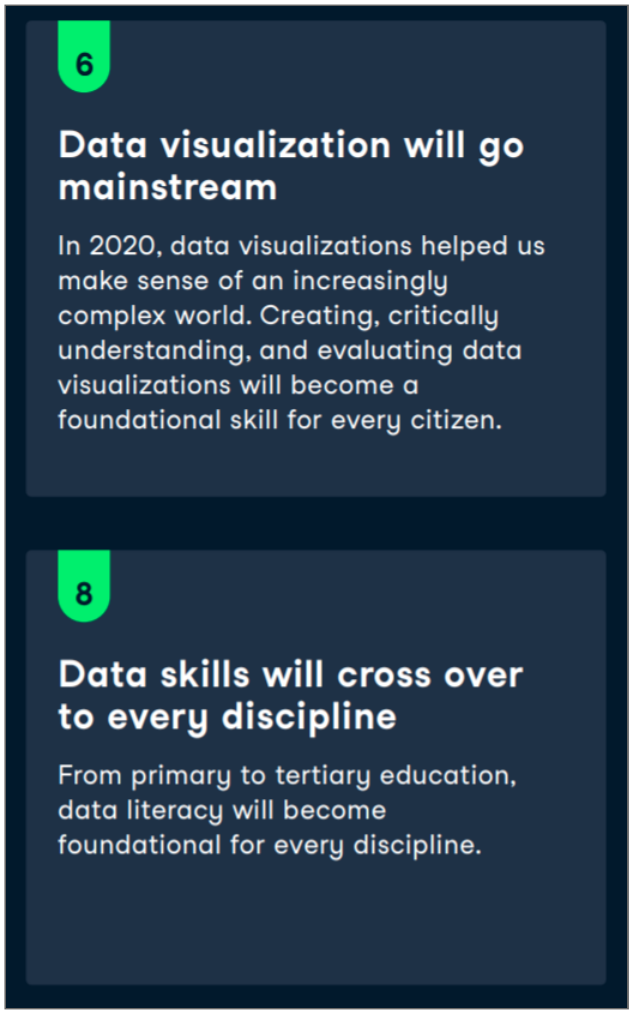
- Playing: **Assassin's Creed Valhalla** (2020)
- Reading: **Waugh, Sword of Honour** (1952-1961)
- Watching: **The Middle** (2009-2018)

WHAT ARE YOUR EXPECTATIONS?

- What do you want to learn here?
- What would you like to avoid?
- What did you take away from another course?
- What did you really not like in another course?

WHICH TOPICS WILL WE COVER?

INTRODUCTION TO DATA SCIENCE



Source: **datacamp.com**

INTRODUCTION TO R PROGRAMMING

Source: **RStudio**

VISUALIZATION USING R

Source: **Thomas Lin Pedersen**

FASTER APPROACH

- Focus on data exploration (EDA)
- Stay close to base R
- Use real data sets
- Compute interactively
- Prepare for DSC201 (ML)

SCHEDULE (SEE SYLLABUS)

No	Date	Lectures[1]	DataCamp[2]	Tests/Quiz[3]	fasteR[4]
1	17-Aug	Overview	Introduction to R:		Overview and Getting Started
2	19-Aug	On the R Shell	- Intro to basics	Test 1	Installing R and first steps
3	24-Aug	Vectors in R	- Vectors	Test 2	More on Vectors
4	26-Aug	Data frames in R	- Matrices	Test 3	On to Data Frames!
5	31-Aug	Factors in R	- Factors	Test 4	R Factor Class
6	2-Sep	apply functions	- Data		The tapply Function
7	7-Sep	Cleaning data	- Lists		Data Cleaning
8	9-Sep	Lists in R	Intermediate R:	Test 7	R List Class
9	14-Sep	Nile exploration	- Conditionals	Test 8	Another Look at the Nile Data
10	16-Sep	Visualization	- Loops	Test 9	Introduction to Base R Graphics
11	21-Sep	Base R graphics	- Functions	Test 10	More on Base Graphics
12	23-Sep	Writing functions	- The apply family	Test 11	Writing Your Own Functions
13	28-Sep	Iteration I	- Utilities	Test 12	for Loops
14	30-Sep	Fibonacci series		Test 13	Functions with Blocks
15	5-Oct	Literate Programming	Data Visualization:	Test 14	IDEs: Text Editing, Saving, Executing
16	7-Oct	Conditions	- Base R graphics	Test 15	If, Else, Ifelse: Conditions
17	12-Oct	EDA example I	- Different plot types	Test 16	Do Pro Athletes Keep Fit?
18	14-Oct	Linear regressio			Regression Analysis, I
19	19-Oct	Object-orientati			isses
20	21-Oct	EDA example II	- Plot customization	Test 19	Baseball Player Analysis (cont'd)
21	26-Oct	Packages		Test 20	R Packages, CRAN, Etc.
22	28-Oct	Grammar of Graphics	Regression in R:	Test 21	A First Look at ggplot2
23	2-Nov	Functional Programmi	- Simple linear regression	Test 22	Should You Use Functional Programmin
24	4-Nov	Text mining I	- Predictions and models	Test 23	Simple Text Processing I
25	9-Nov	Text mining II	- Assessing model fit	Test 24	Simple Text Processing II
26	11-Nov	Linear regression II			r Regression Analysis, II
27	16-Nov	Dates and times			ng with the R Date Class
28	18-Nov	Coding style	Text mining (Bag-of-Words)	Test 27	Tips on R Coding Style and Strategy
29	23-Nov	Logistic regression	- Basics of Bag-of-Words	Test 28	The Logistic Model
30	25-Nov	Version control	- Word clouds and visuals	Test 29	Files and Directories
31	30-Nov	Iteration II	- Text mining techniques	Test 30	R while loops
32	2-Dec	Summary and outlook	- Text mining case study		Summary and Outlook
33	TBD			Final exam	

BASICS

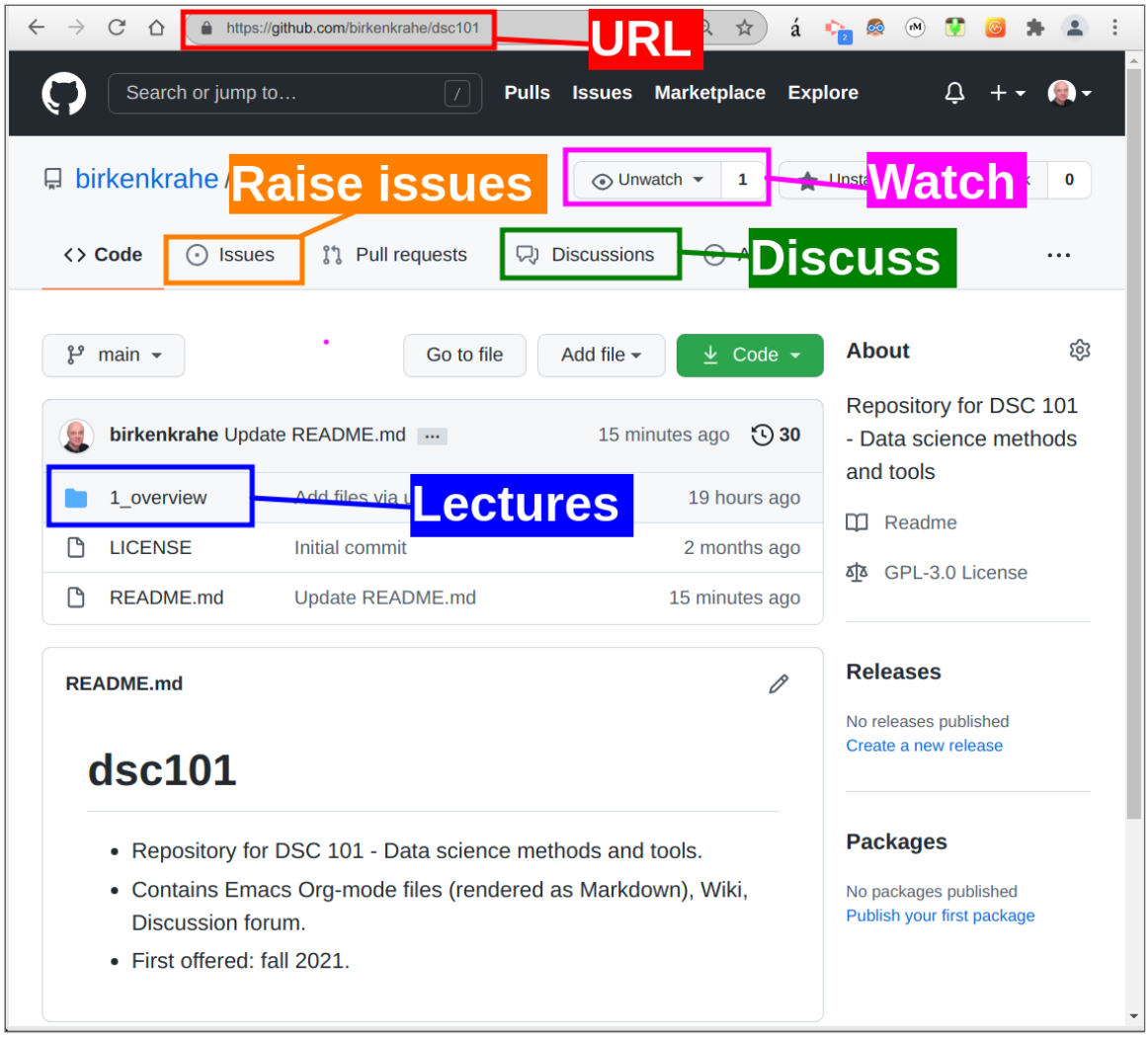
VISUALIZATION

APPLICATIONS

HOW WILL WE DO IT?

CLASSROOM SESSIONS

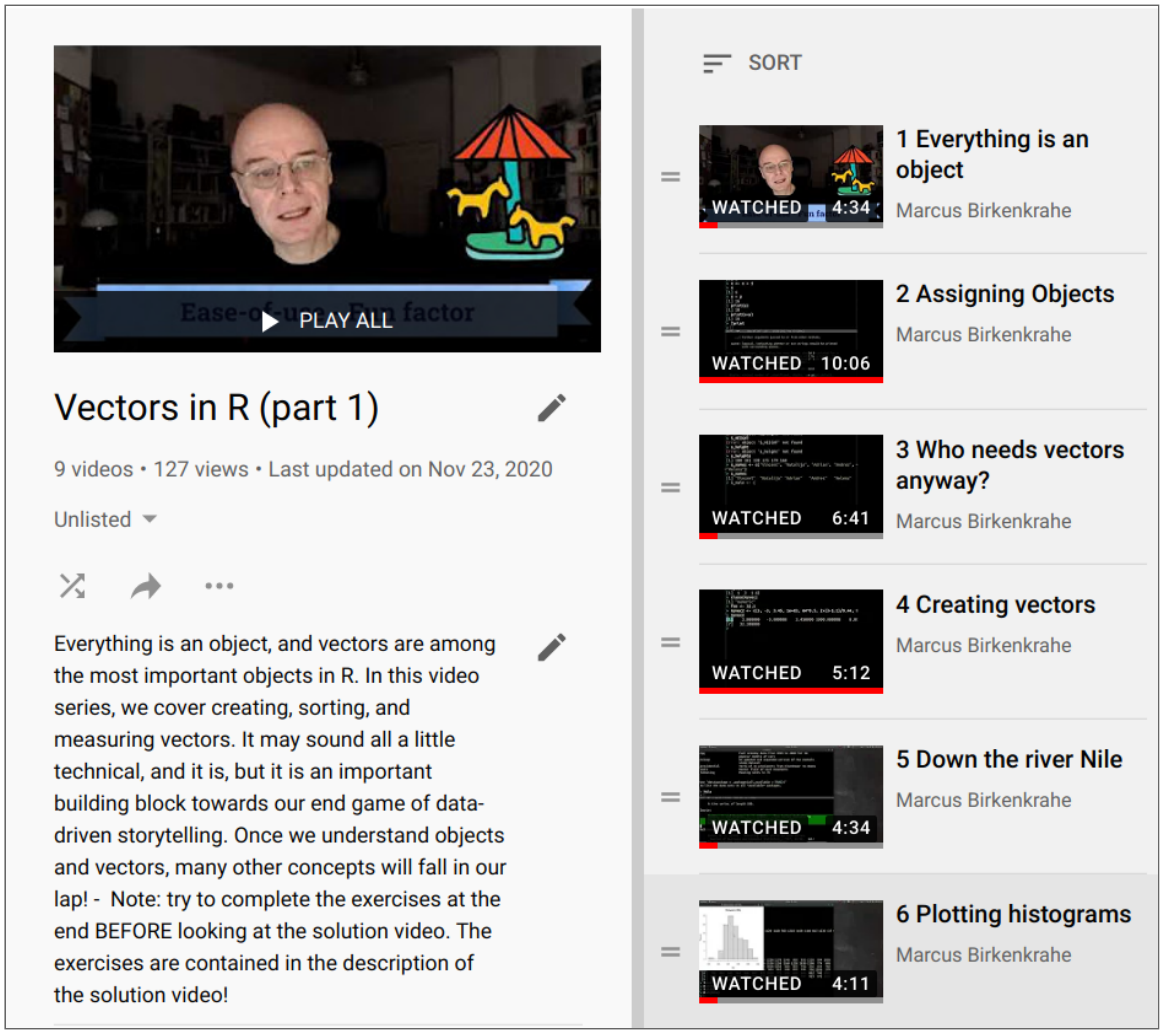
LECTURE SCRIPTS WITH EXERCISES (GITHUB)



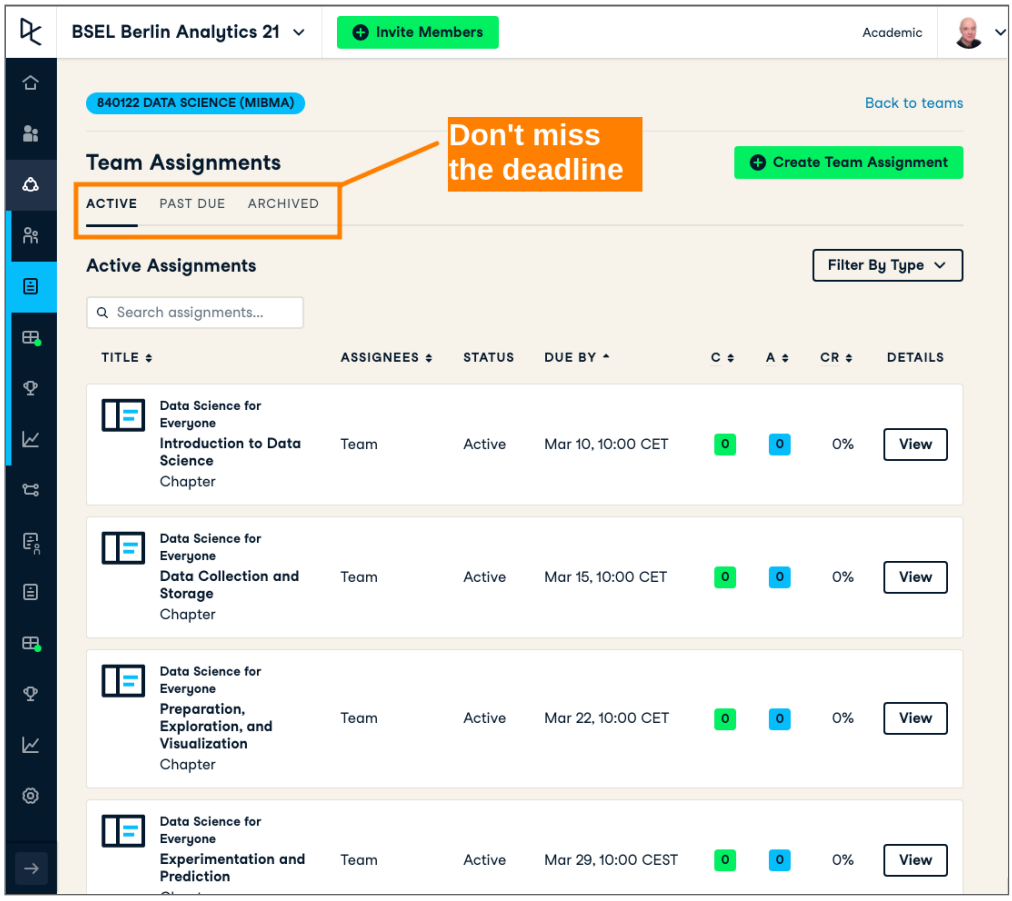
READING SUGGESTIONS

- Matloff: **fasteR: Fast Lane to Learning R!** (2021)
- Matloff: **The Art of R Programming** (2011)

VIDEO LECTURES (YOUTUBE)

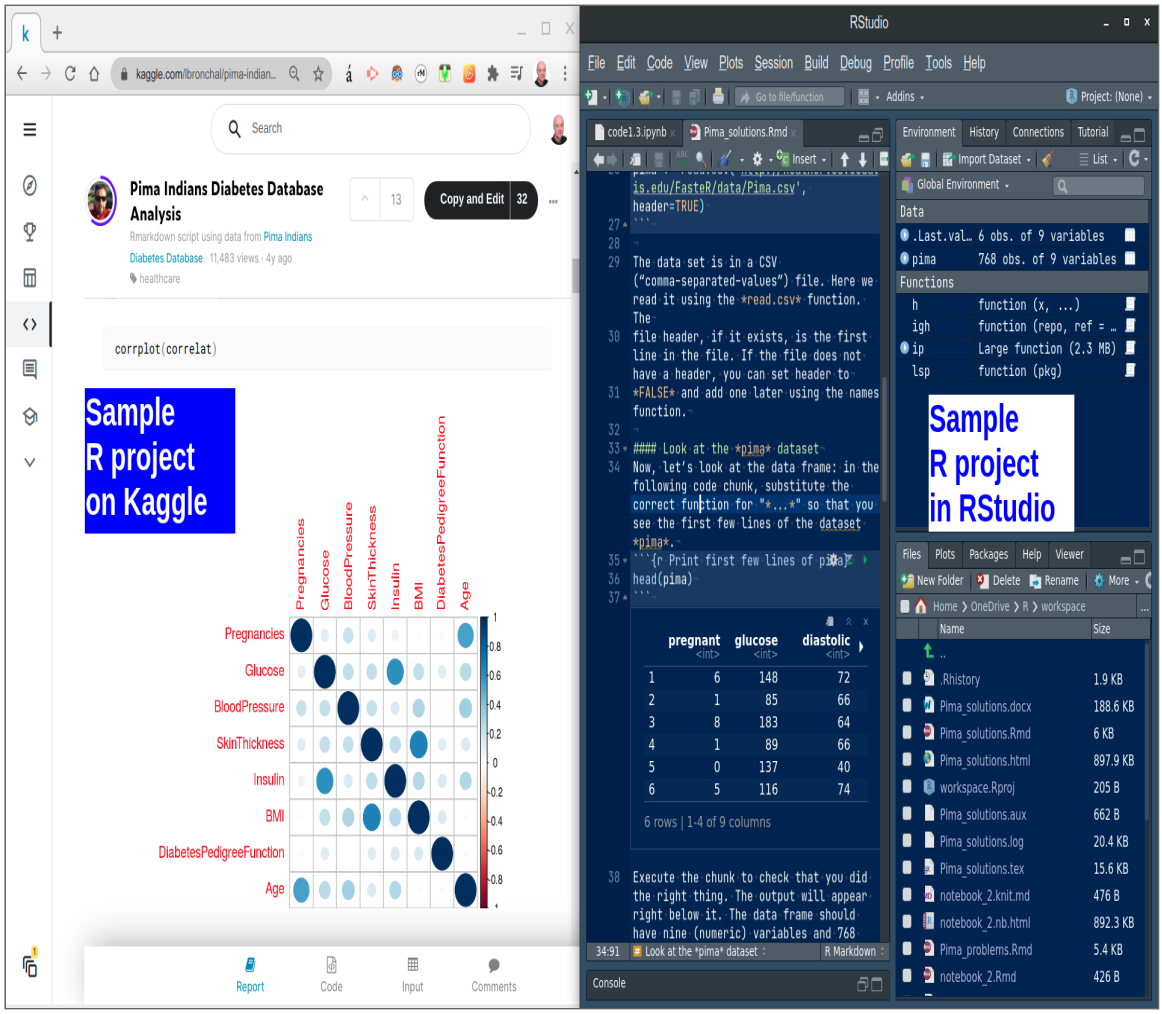


ONLINE ASSIGNMENTS (DATACAMP)



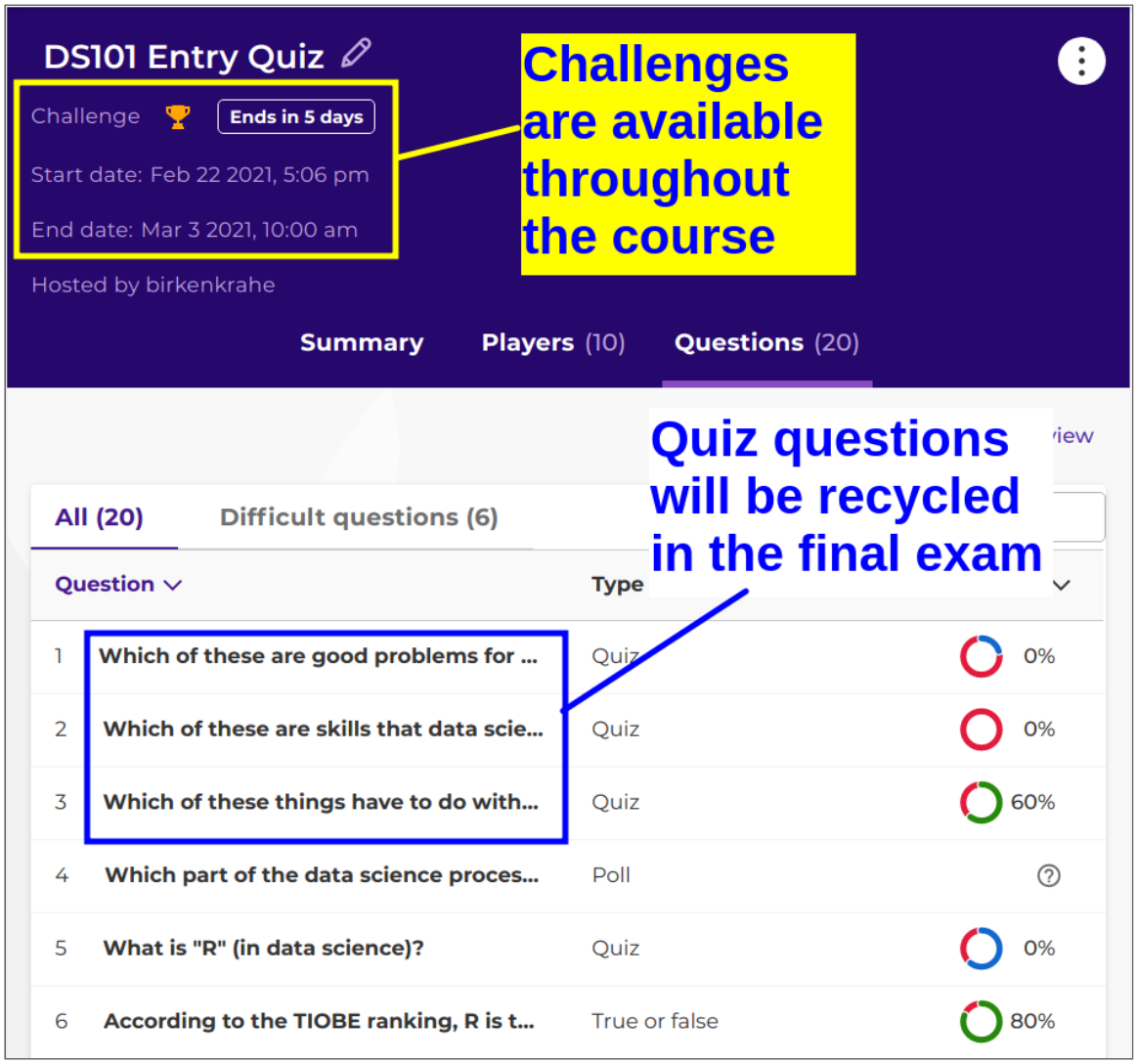
- **Register at DataCamp** today!

TEAM EDA PROJECT

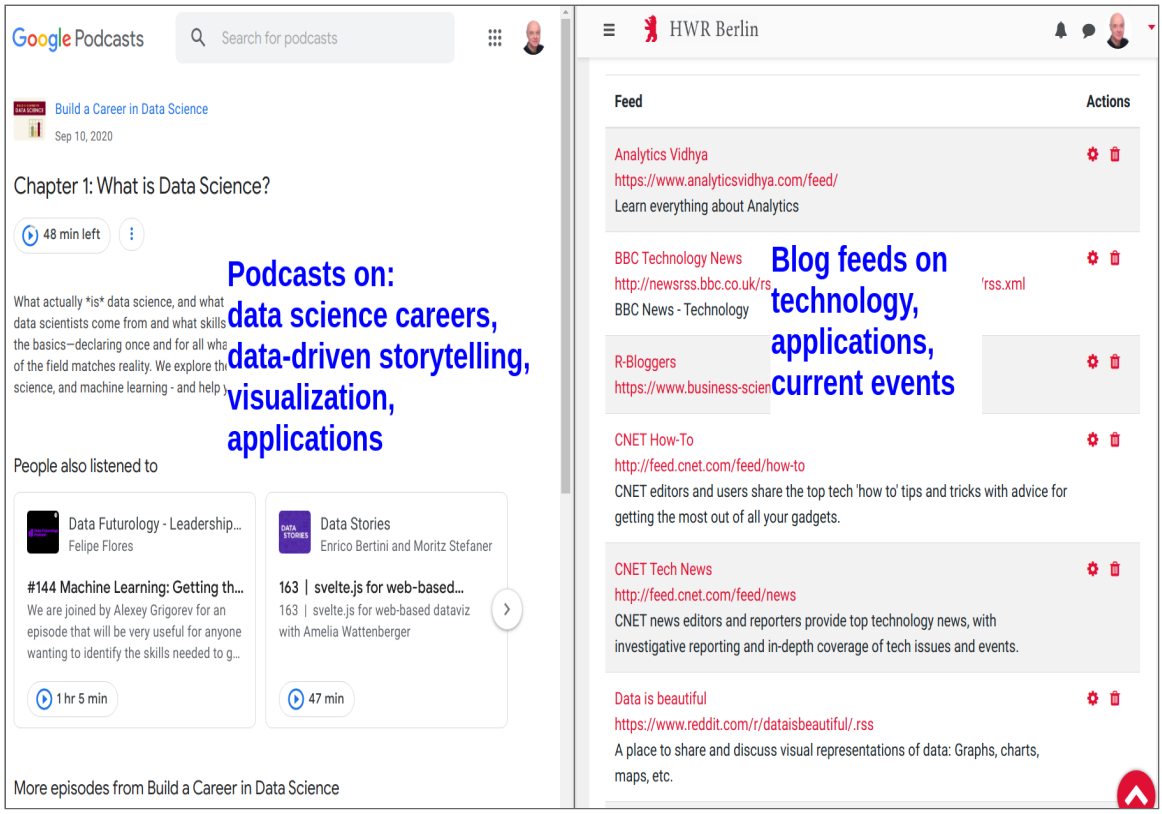


AGILE PROJECT MANAGEMENT

TESTS AND FINAL EXAM



PODCASTS AND FEEDS



Shared via GitHub / Schoology

SUMMARY OF COURSE ACTIVITIES

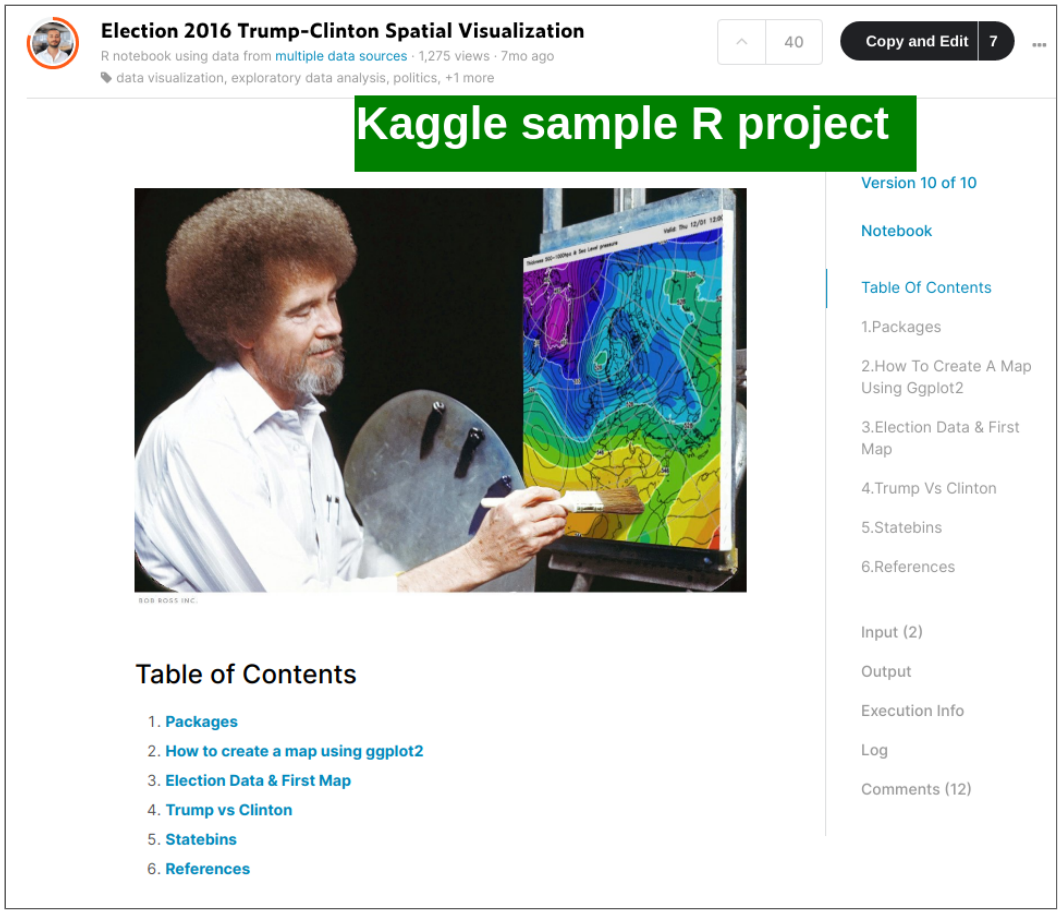
- Twice weekly classroom meetings
- Lecture scripts (GitHub)
- Reading assignments (Online)
- Video lectures (YouTube)
- **Online assignments** (DataCamp)
- **Team EDA projects** (Sprints)
- **Tests and final exam**
- Podcasts and feeds

WHAT DO YOU HAVE TO DO TO PASS?

DATACAMP ASSIGNMENTS (> 50%)

Complete at least 8 of 15 assignments

TEAM PROJECT (> 50%)



Present on Nov 30 or Dec 2

WHAT IS A TEAM PROJECT?

- Description of the dataset
- Introduction of the problem statement
- Description of the methods used
- Visualization of the data (plots!)
- Analysis of the plots
- Limitations of own analysis
- References

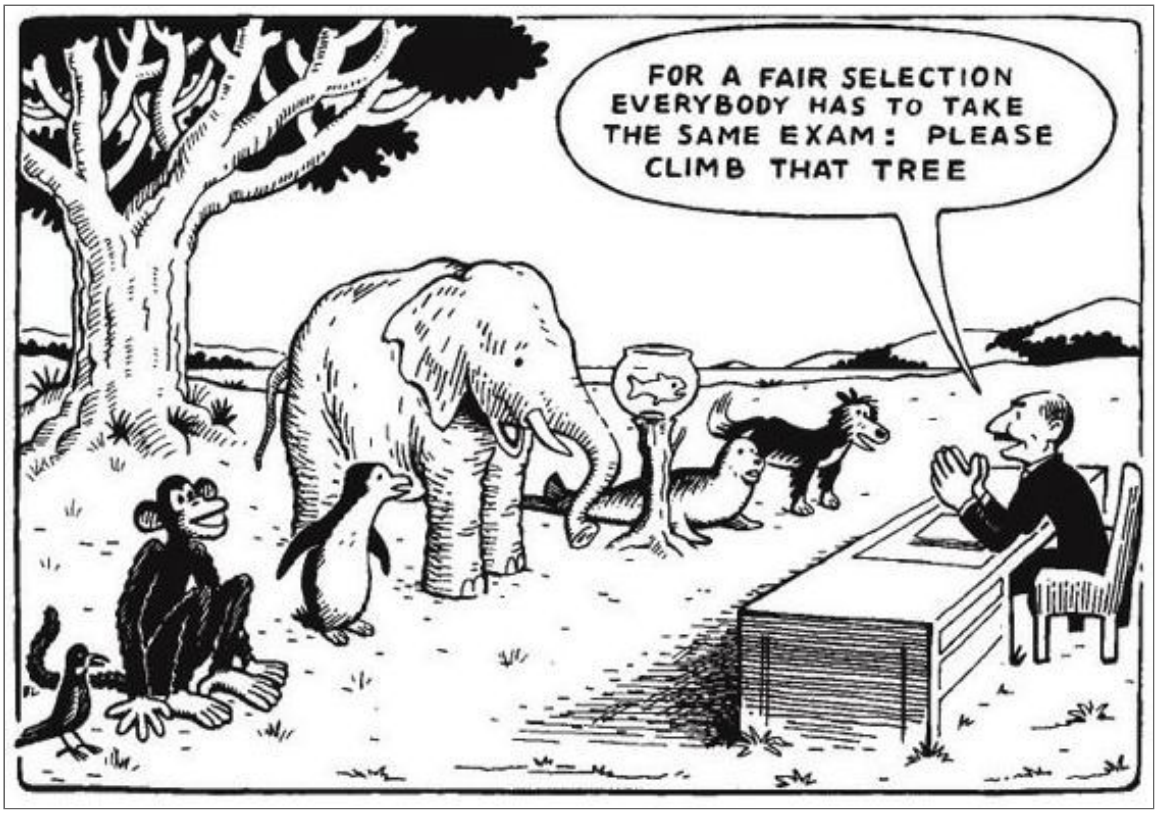
DO YOU HAVE PROJECT EXAMPLES?

- Examples on Kaggle (**example**)
- Examples on data science blogs (**example**)
- Translate from Python to R (**example**)
- Extend someone else's EDA (**example**)
- Document an R package (**example**)
- Use your own data (**example**)

CAN I DO A PROJECT AS AN ABSOLUTE BEGINNER?

- Keep It Simply Scientific (**IMRaD**)
- Look at examples (e.g. in my **bookmarks**)
- Create data set (e.g. your productivity)
- Researchers are beginners

FINAL EXAM (> 50%)



Final exam: date TBD

WHAT'S NEXT?

IN THE COURSE

- Intro to Data science (Lecture)
- Intro to DataCamp (Practice)
- Intro to GitHub (Productivity)
- Intro to R (Language)

YOUR CHALLENGES

What?	When?
Register at DataCamp	Today
Register at GitHub	Today
Complete test challenge	Aug 24
Complete DataCamp assignment	Aug 24
Set up team project (2-3 ppl)	Sep 2
Check FAQs x 2 in GitHub	n.d.
Ask questions (class/GitHub)	n.d.

**) do this every week until December*

ANY QUESTIONS?

A copy of this presentation is available.