Data Science in Economics

ELEVATING MANDATORY UNDERGRADUATE EDUCATION

Alexander Rieber

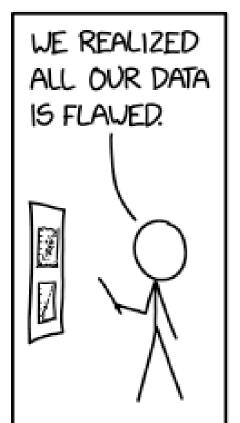
♠ AlexRieber

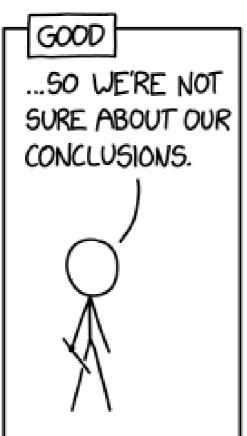
alexander.rieber@uni-ulm.de

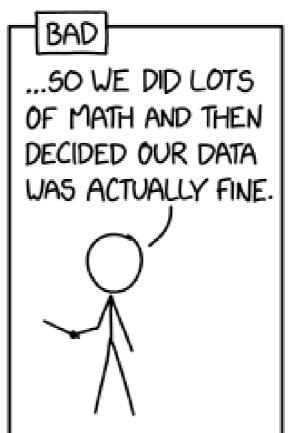
July 9th, 2024

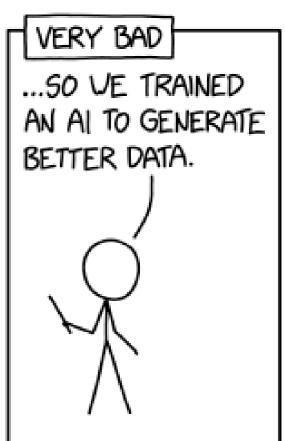


Why do Bachelor in Economics students need Data Science and Causal Inference?









MOTIVATION



How can Data Science be effectively integrated into the curriculum?

At Ulm University we include:

- **★** Mandatory course in the 3rd and 4th semesters in Data Science and Causal Inference
- + Flipped classroom approach with ample opportunity to illustrate theoretical content in lectures using a case study
- ♣ Modern, project-based teaching where groups independently develop solutions to various problems in business administration and economics
- ♣ Install a variety of feedback mechanisms
- Use of R for student work and course management in GitHub

DIFFICULTIES IN A MANDATORY COURSE



- ★ Many students with varying levels of knowledge/interest
- ♣ Project-based teaching is unfamiliar to students until then
 - **★** A lot of work during the semester may be perceived as burdensome
 - ★ Many deadlines, which means students must be constantly active
 - **★** Exercise sheets due during the first part of the semester (lecture phase)
 - **◆** Submission of projects and review reports during the second part of the semester (project phase)
- → Flipped classroom and the implemented feedback mechanisms rely on active student participation

STRUCTURE OF THE COURSE

Salzburg 2024

The course is divided into two parts:

The Lecture Phase: Lecture videos for the theoretical background

Calendar Week	Topic Leadure + Case Show	Tasks
Week 42	Week 1: Introduction to R, RStudio and Git(Hub)	Problem Set 1 (Deadline 23.10.2023)
Week 43	Week 2: Data Wrangling	Problem Set 2 (Deadline 31.10.2023)
Week 44	Week 3: Data Visualization	Problem Set 3 (Deadline 06.11.2023)
Week 45	Week 4: Sample vs. Population	
Week 46	Week 5: Projekt 1	Deadline 1st project (23.11.2023)
24.11.2023	Mock exam + Kick-off Project 2	Mock exam (24.11.2023)
20.12.2023	Presentation of the sample solution for the 2nd project	Deadline 2nd project (12.12.2023) Deadline Review Report for 2nd project (19.12.2023)
10.01.2024	Kick-off Project 3	
10.07.2024	Presentation of the sample solution for the 3rd project	Deadline 3rd project (30.01.2024) Deadline Review Report for 3rd project (07.02.2024)
06.03.2024	Multiple-Choice Exam	

STRUCTURE OF THE COURSE



The course is divided into two parts:

The Lecture Phase:
Dedicated RTutor Problem
Sets for each semester

Topic Lecture + Case Show	Tasks
Week 1: Introduction to R, RStudio and Git(Hub)	Problem Set 1 (Deadline 23.10.2023)
Week 2: Data Wrangling	Problem Set 2 (Deadline 31.10.2023)
Week 3: Data Visualization	Problem Set 3 (Deadline 06.11.2023)
Week 4: Sample vs. Population	RTutor Problem Sets
Week 5: Projekt 1	Deadline 1st project (23.11.2023)
Mock exam + Kick-off Project 2	Mock exam (24.11.2023)
Presentation of the sample solution for the 2nd project	Deadline 2nd project (12.12.2023) Deadline Review Report for 2nd project (19.12.2023)
Kick-off Project 3	
Presentation of the sample solution for the 3rd project	Deadline 3rd project (30.01.2024) Deadline Review Report for 3rd project (07.02.2024)
	Week 1: Introduction to R, RStudio and Git(Hub) Week 2: Data Wrangling Week 3: Data Visualization Week 4: Sample vs. Population Week 5: Projekt 1 Mock exam + Kick-off Project 2 Presentation of the sample solution for the 2nd project Kick-off Project 3 Presentation of the sample solution for

STRUCTURE OF THE COURSE

Salzburg 2024

The course is divided into two parts:

The Project Phase: Students must complete three projects.

Calendar Week	Topic Leasure + Case Shak	Tasks /
Week 42	Week 1: Introduction to R, RStudio and Git(Hub)	Problem Set 1 (Deadline 23.10.2023)
Week 43	Week 2: Data Wrangling	Problem Set 2 (Deadline 31.10.2023)
Week 44	Week 3: Data Visualization	Problem Set 3 (Deadline 06.11.2023)
Week 45	Week 4: Sample vs. Population	RTutor Problem Sets
Week 46	Week 5: Projekt 1	Deadline 1st project

Week 46	Week 5: Projekt 1	Deadline 1st project (23.11.2023)
24.11.2023	Mock exam + Kick-off Project 2	Mock exam (24.11.2023)
20.12.2023	Presentation of the sample solution for the 2nd project	Deadline 2nd project (12.12.2023) Deadline Review Report for 2nd project (19.12.2023)
10.01.2024	Kick-off Project 3	
10.07.2024	Presentation of the sample solution for the 3rd project	Deadline 3rd project (30.01.2024) Deadline Review Report for 3rd project (07.02.2024)
06.03.2024	Multiple-Choice Exam 3 Projects	s + 2 Review Reports



The Lecture Phase: A collaboratively developed case study gives students an idea of how to use data!

Source: <u>Case-Study in the</u> <u>project course Data Science</u>, 2023/2024.

🔼 Bu	undesagentur für Arbeit	t					I I					I I					1
	atistik																1
Bestar	nd an Arbeitslosen	nach Rech	ntskreise	en - Insç	jesamt		1										1
Länder,	Regierungsbezirke, Kreis	e und Gemei	nden in De	utschland			I I	zurück zu	m Inhalt			i i					1
	monat: Jahreszahlen 201						I I	Zui doit Zu	III ii ii iiais			į					1
			-llelhaziahar w	ee Ala und Ala	- 113 durch din /		l Inheit betreut u	and dochalls im	Beabtekroie C	COR III ale arb	nitelan aartibit	k-some im ex	OR II) Eranbair	nach Rock	stel-raisan sind	debor ab 1an	
Selt u	dem 1. Januar 2017 werden die sog.	. "Aufstocker" (Para	allelbezienei vi	n Ma nua Ma	•				Recntskreis 3	GB III ais aru	eitsios gezariii.	(Zuvor: im 59	B II). Ergeunis	se hach reun	tskreisen sinu		
			1			chtskreise SC	GB III und SGB II							kreis SGB III			
	Bundesland	ins-	ausgewählte Altersgruppen dar. (Spalte 1)			darunter			Alters gruppen Spalte 6)		ins			Altersgruppen Spalte 11)		darunte	
1	Regierungsbezirk	gesamt	unter 20	unter 25	50 Jahre	55 Jahre	Ausländer	unter 20	unter 25	50 Jahre	55 Jahre	gesamt	unter 20	unter 25	50 Jahre	55 Jahre	Auslände
	Kreis		Jahre	Jahre	und älter	und älter		Jahre	Jahre	und älter	und älter	ļ	Jahre	Jahre	und älter	und älter	
Schlüssel	Gemeinde	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Deutschland	2.532.837	47.840	230.694	839.545	529.002	655.249	14.022	71.133	136.088	73.241	855.431	13.090	94.069	335.978	243.027	152.820
01	Schleswig-Holstein	92.434	2.072	9.703	30.138	18.788	18.688	437	2.297	3.471	1.906	30.989	656	4.104	11.953	8.284	3.327
01001000	Flensburg, Stadt	4.512	108	566	1.120	654	983	18	141	137	70	1.309	30	196	381	249	194
01002000	Kiel, Landeshauptstadt	12.345	220	1.100	3.306	1.842	3.103	54	294	559	272	2.704	51	391	747	491	351
01003000	Lübeck, Hansestadt	9.692	180	836	2.902	1.667	2.050	41	204	391	213	2.429	37	291	846	576	305
01004000	Neumünster, Stadt	3.836	105	440	1.134	685	724	10	45	155	85	982	28	177	298	195	113
UIW-W	reculturator, same	0.000	100		1.107	-		100	72	100	-	5502	2.0		2.00	100	110
01051	Dithmarschen	4.628	147	634	1.402	826	791	27	141	115	65	1.508	46	238	621	443	155
01051001	Albersdorf	150	5	23	43	23	24	0	4	4	2	39	2	9	15	11	4
01051002	Arkebek	5	0	1	1	1	2					2	0	0	1	1	
01051003	Averlak	15	0	3	4	4	1		0			8	0	1	2	2	0
01051004	Bargenstedt	17	1	2	5	2	0	0	0			5		0	1	0	0
01051005	Barkenholm	3			0	0	0					2			0	0	
01051006	Sact	21	0	2	7	4	2	0	0			8		1	3	2	1
01051008	Bergewöhrden	1			1												
01051010	Brickeln	5	0	0	4	4						3	0	0	2	2	
01051011	Brunsbüttel, Stadt	483	14	71	141	80	108	2	19	14	6	111	3	22	50	39	8
01051012	Buchholz	24	0	3	6	5	1		0	-	_	10	0	2	2	1	0
01051013	Büsum Büsum	181	2	20	73	45	38	0	7	7	5	94	1	11	46	29	16
01051014	Büsumer Deichhausen	12 25	1	1	8	4	3	U	1	0	0	6 5	0	1	4	3 1	2
01051015	Bursoh Burra (Dithmarschen)	161	3	20	52	4 35	13		1	0 2	1	42	2	7	20	16	2
01051016	Burg (Dithmarschen)	101	3	20	3	2	3		0	2	1	42	2	1	1	16	0
01051017	Buserwurth Dellstedt	26	1	3	8	4	1		v	0	0	6	0	2	2	1	1
01051019	Delve	15	1	2	5	3	1			1	1	7	1	2	3	2	1
01051020	Diekhusen-Fahrstedt	17	1	2	5	2	0			-	•	10	0	1	2	2	0
01051022	Dingen	12	-	1	6	3	1			1		6	-	0	3	2	"
01051023	Dörpling	18	0	2	6	3	0			0		8		0	2	2	
01051024	Eddelak	44	1	6	10	7	2		1			16	0	2	5	3	0
01051026	Eggstedt	14	1	1	5	4	1					7	1	1	4	3	
01051027	Elpersbüttel	20	0	2	5	3	4		1	0	0	9	0	1	2	2	2
01051028	Epenwöhrden	17	0	2	6	2	3	0	1	0	0	8	0	1	4	1	0
01051030	Fedderingen	6	0	1	2	1						3	0	1	1	1	
01051032	Frestedt	5	0	1	1	0						4		0	1	0	
01051033	Friedrichsgabekoog	3		0	3	3	1			1	1	2		0	2	2	0
01051034	Friedrichskoog	69	1	7	32	17	4		0			37	0	5	20	13	2
01051035	Gaustiero	7	1	1	4	2	1					2			2	1	
01051036	Glusing	4	0	0	3	2						0		0			



The Lecture Phase: A collaboratively developed case study gives students an idea of how to use data!

Source: <u>Case-Study in the</u> <u>project course Data Science</u>, 2023/2024.

Arbeitslosenquote

Bundesland	Mittelwert	Std.	Median
Bayern	3.04	0.77	3.00
Baden-Württemberg	3.31	0.64	3.33
Hessen	5.01	1.35	5.06
Rheinland-Pfalz	5.31	1.47	5.26
Saarland	5.85	1.73	5.32
Niedersachsen	6.16	1.77	6.01
Thüringen	6.30	1.80	5.70
Schleswig-Holstein	6.36	1.00	6.82
Sachsen	6.65	1.08	6.31
Nordrhein-Westfalen	7.15	2.43	6.88
Brandenburg	7.99	1.95	8.28
Mecklenburg-Vorpommern	8.45	1.58	8.00
Bremen	8.95	2.54	8.95
Sachsen-Anhalt	9.03	1.59	8.91

Bitte beachten:

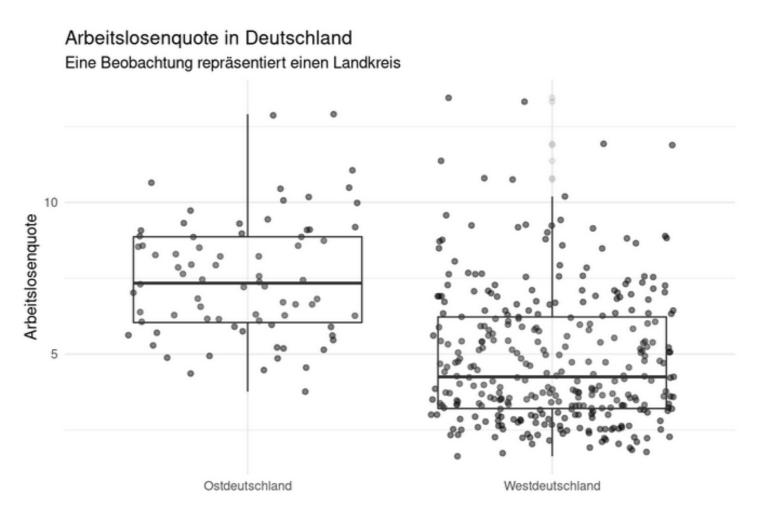
Wir haben keine Informationen zu Berlin und Hamburg, weshalb sie nicht in der Tabelle aufgeführt wurden.

¹ Die ostdeutschen Bundesländer sind grau hinterlegt.



The Lecture Phase: A collaboratively developed case study gives students an idea of how to use data!

Source: <u>Case-Study in the</u> <u>project course Data Science</u>, 2023/2024.

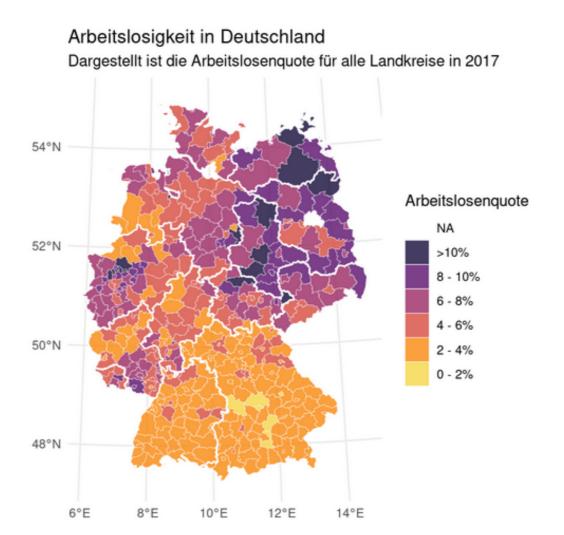


Quelle: Daten der Agentur für Arbeit aus dem Jahr 2017



The Lecture Phase: A collaboratively developed case study gives students an idea of how to use data!

Source: <u>Case-Study in the</u> <u>project course Data Science</u>, 2023/2024.





The Lecture Phase: A collaboratively developed case study gives students an idea of how to use data!

Source: <u>Case-Study in the</u> <u>project course Data Science</u>, 2023/2024.

Arbeitslosigkeit mit Interaktionstermen

	Ato-quote					
	(1)	(2)	(3)	(4)		
Konstante	3.31	3.23	3.23	6.11		
	[2.95, 3.67]	[2.86, 3.59]	[2.85, 3.60]	[4.75, 7.46]		
Anteil der Schulden	0.23	0.24	0.24	0.06		
	[0.19, 0.27]	[0.20, 0.28]	[0.20, 0.28]	[-0.08, 0.21]		
Ostdeutschland	1.36	2.88				
	[0.84, 1.88]	[1.43, 4.33]				
Anteil der Schulden:Ostdeutschland		-0.18				
		[-0.34, -0.02]				
N	396	396	321	75		
Adj. R ²	0.42	0.43	0.41	0.00		

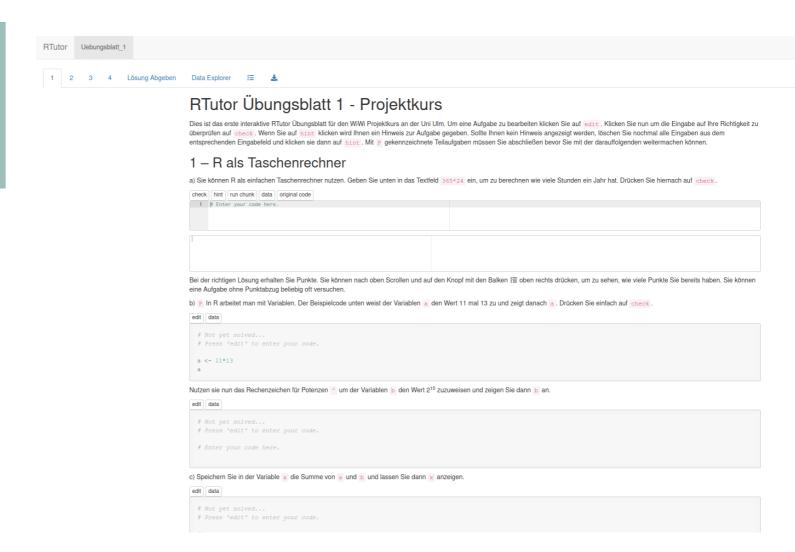
Alo-quote

THE LECTURE PHASE: RTUTOR PROBLEM SETS ONLINE



The Lecture Phase: Six RTutor Problem Sets show students how to work with RMarkdown and teach them Causal Inference.

Source: <u>RTutor Problem Sets</u>, 2023/2024.

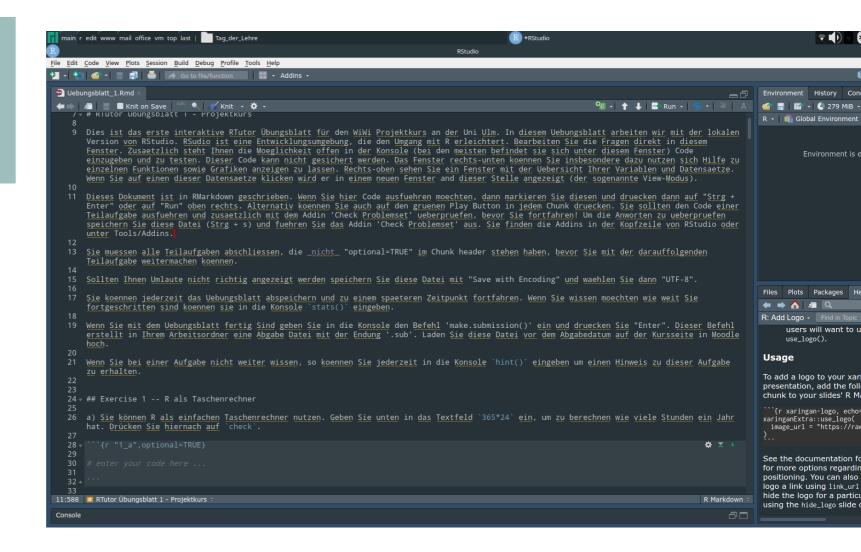


THE LECTURE PHASE: RTUTOR PROBLEM SETS OFFLINE



The Lecture Phase: Six RTutor Problem Sets show students how to work with RMarkdown and teach them Causal Inference.

Source: <u>RTutor Problem Sets</u>, 2023/2024.



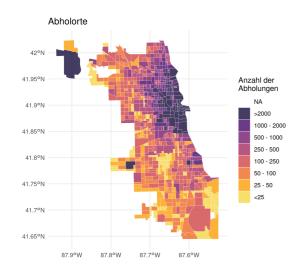
THE PROJECT PHASE

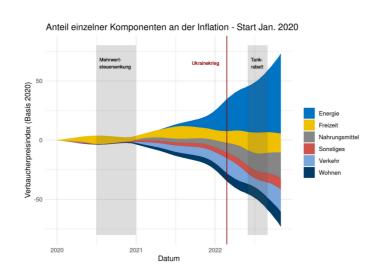


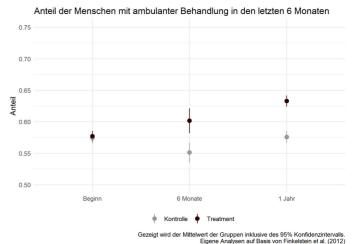
- ◆ Teams of 3 students
- ★ The first project of each semester is done together with the lecturer
 - **★** This means code, description, and interpretation are developed in the lecture together with the students
- **★** For the **second** and **third project** of each semester: Independent project work!
- **★** A total of 6 projects, with the difficulty of the projects constantly increasing:
 - ★ Increasing demands on econometric skills
 - **★** Increasing demands on technical skills (e.g., visualization)
 - lacktriangle Questions become more complex over the two semesters (descriptive statistics \rightarrow estimation of causal effects)

THE PROJECT PHASE: SAMPLE PROJECTS









Uber driver in Chicago

<u>Inflation in Europe</u>

Oregon Health Insurance Experiment

THE PROJECT PHASE: PROJECT DESIGN



What is important in project design?

- **★** A clear structure with clear and unambiguous questions
 - **★** Even small ambiguities lead to a flood of emails/discussion posts.
 - **★** Transfer questions can be asked at the end of the project. This allows for more degrees of freedom in the answers.
- **★** It has been helpful to develop the first project of each semester with the students.
 - **◆** Students learn the structure and expectations for responding to a project.

THE PROJECT PHASE: PROJECT WORK



How is project work going? Working as a Team?

Provide a Collaboration Tool!

- **★** I suggest to use Git and GitHub
 - **★** Ideal for teamwork and version control (also a must have skill in industry).
 - **◆** Students easily learn GitHub Desktop.
 - **◆** Introduce GitHub at the start of the course to address technical issues before the project phase.

Establish an Escalation Mechanism

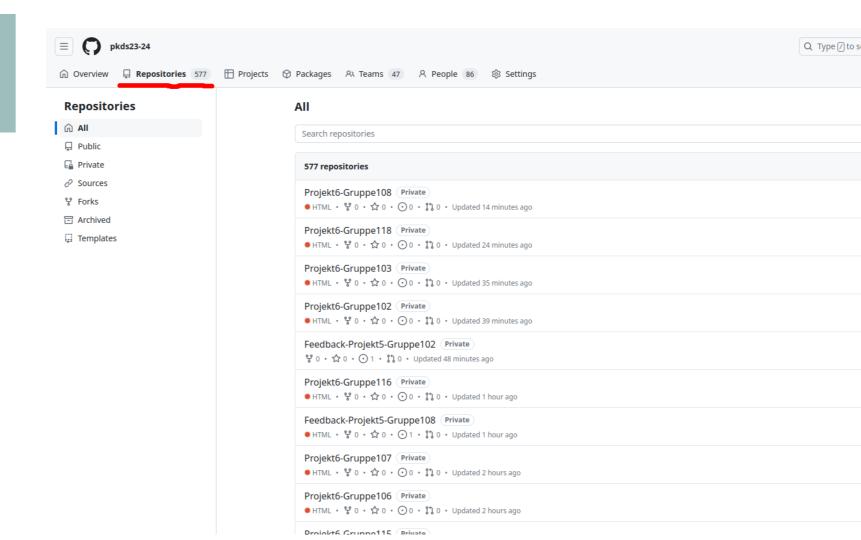
- **+** Addressing Challenges in Group Work
 - What if a team member doesn't contribute? (Free-Rider Problem)
 - **Solution**: Provide clear steps and resources for resolving such issues.

THE PROJECT PHASE: PROJECT DISTRIBUTION



How to do project distribution and maintainance?

Use the <u>ghclass Package</u> to manage hundreds of repos with R!





THE PROJECT PHASE: CONTINUOUS INTEGRATION FOR PROJECT VERIFICATION

How to verify project Rmds?

Use **GitHub Actions** and automatically knit the project RMarkdown file on upload.

```
push:
   branches:
   - main
    paths:
    - 'Name_of_Project.Rmd'
name: Check Repo
jobs:
  check-rmd:
    runs-on: ubuntu-latest
    container:
      image: rocker/tidyverse:latest
    steps:
   - name: Checkout
     uses: actions/checkout@main
   - name: Install remotes package
        R -e "install.packages('remotes')"
   - name: Install missing R packages
        Rscript install_packages.R
    - name: Check Rmd Rendering
      run: Rscript -e "rmarkdown::render('Name_of_Project.Rmd')"
```

THE PROJECT PHASE: PROJECT PRESENTATIONS



Should students present their work?

Yes, I think they have to train their presentation skills!

+ Format: Screencasts only, max 5 minutes

Content: Brief project summary, focus on results

Benefits

★ Fairness: Anonymized groups, unbiased evaluation

Efficiency: Saves time for students and lecturers

Reviewability: Easier to rewatch and understand

+ Detailed Feedback: Higher learning gains

THE PROJECT PHASE: REVIEW REPORTS



- **Requirement**: Each student reviews a randomly assigned project
- **★ Learning**: Develops critical evaluation and constructive feedback skills
- + Feedback:
 - **★** Peer review by students
 - ♣ Anonymous tutor feedback
 - **★** Lecturer feedback on screencasts
- **★** Maintenance and distribution: Again, the ghclass Package is fantastic to distribute group projects to (randomly selected) peers

CONCLUSION



Project-based teaching in a mandatory course is possible and can be a lot of fun!

If you are going to try this, you may want to consider the following in advance:

- **◆** What can I expect from my target audience?
 - **★** The difficulty level of the projects must be challenging but not overwhelming
- **◆** Clear communication about the number, scope, and duration of the projects is essential
 - **★** Students may still complain about too much work
- **★** Feedback is extremely important!
- ♣ Project work must be designed to align incentives
 - Avoid the free rider problem
 - **★** Show escalation mechanisms
- My suggestion: Use R and Git!
 - **★** Students could learn R with <u>RTutor</u>
 - Students do projects with <u>RMarkdown</u> or <u>Quarto</u>
 - **★** Manage your GitHub repositories with ghclass

R PACKAGES



Rundel C, Cetinkaya-Rundel M (2024). ghclass: Tools for Managing Classes on GitHub. R package version 0.3.0, https://github.com/rundel/ghclass.

Kranz S (2020). RTutor: Interactive R Problem Sets. https://github.com/skranz/RTutor