Data Science in Economics

ELEVATING MANDATORY UNDERGRADUATE EDUCATION

Alexander Rieber
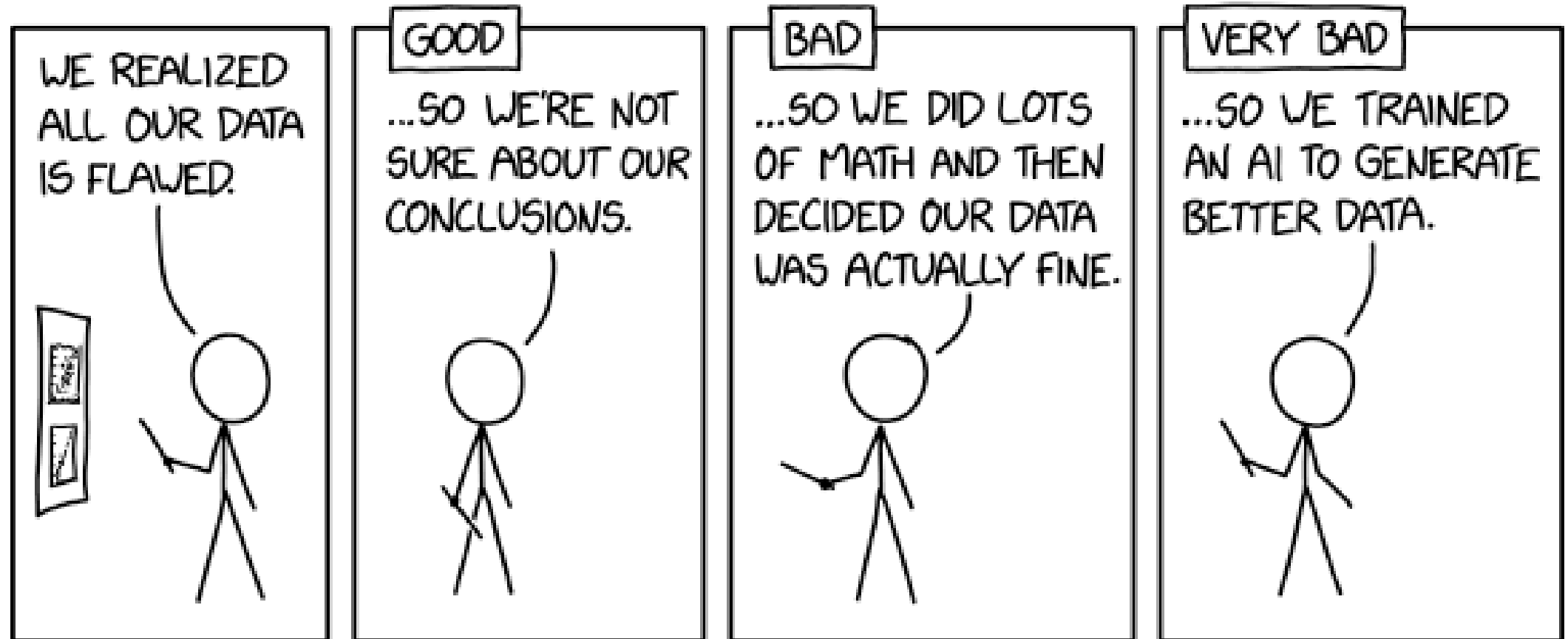
 AlexRieber

 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

The course is divided into two parts:

The Lecture Phase: Lecture videos for the theoretical background

Winter Term

Calendar Week	Topic	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	

Lecture + Case Study (handwritten annotation above "Topic")

STRUCTURE OF THE COURSE

The course is divided into two parts:

The Lecture Phase:
Dedicated RTutor Problem Sets for each semester

Winter Term

Calendar Week	Topic	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	

(handwritten annotations: "Lecture + Case Study" above Topic column; "RTutor Problem Sets" in Tasks column)

STRUCTURE OF THE COURSE

The course is divided into two parts:

The Project Phase: Students must complete three projects.

Winter Term

Calendar Week	Topic	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	

(handwritten annotations: Lecture + Case Study; RTutor Problem Sets; 3 Projects + 2 Review Reports)

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

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

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

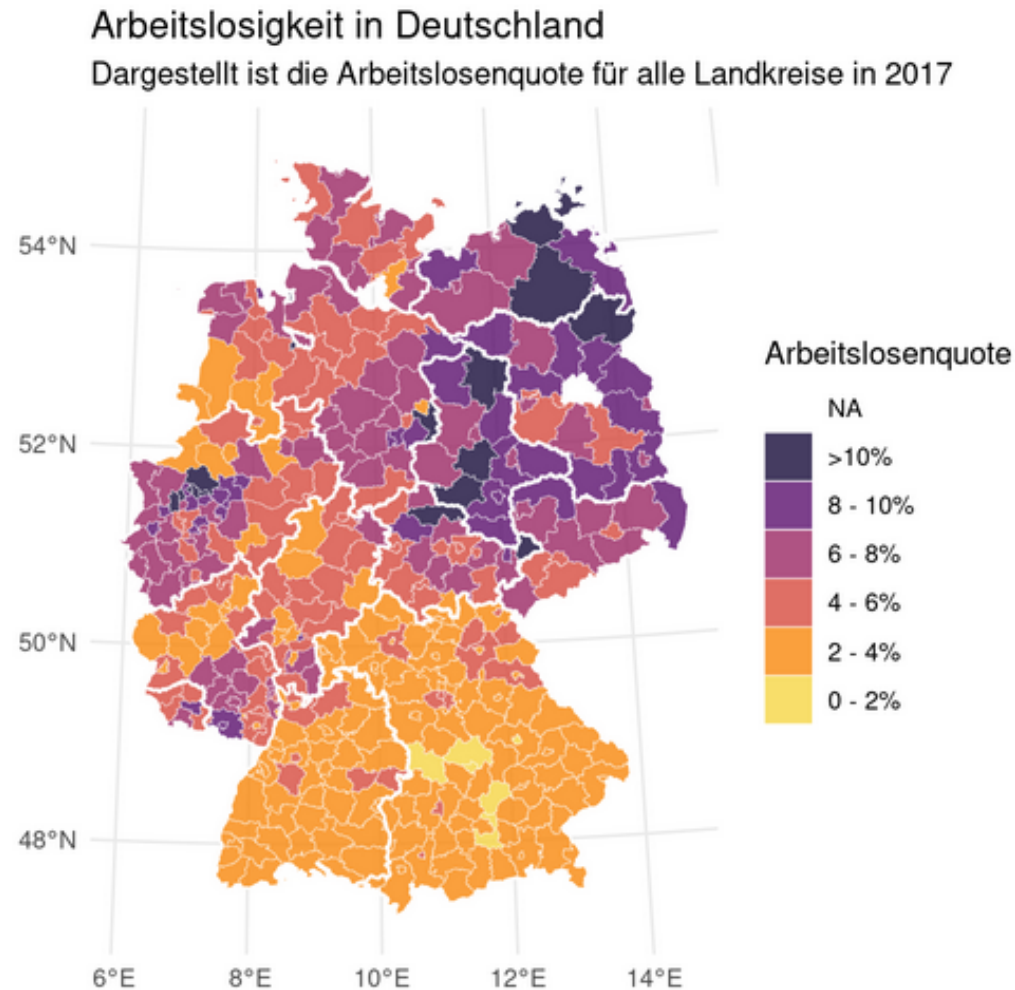Source: Case-Study in the project course Data Science, 2023/2024.

Bundesland	Arbeitslosenquote		
	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: Case-Study in the project course Data Science, 2023/2024.

Arbeitslosenquote in Deutschland
Eine Beobachtung repräsentiert einen Landkreis
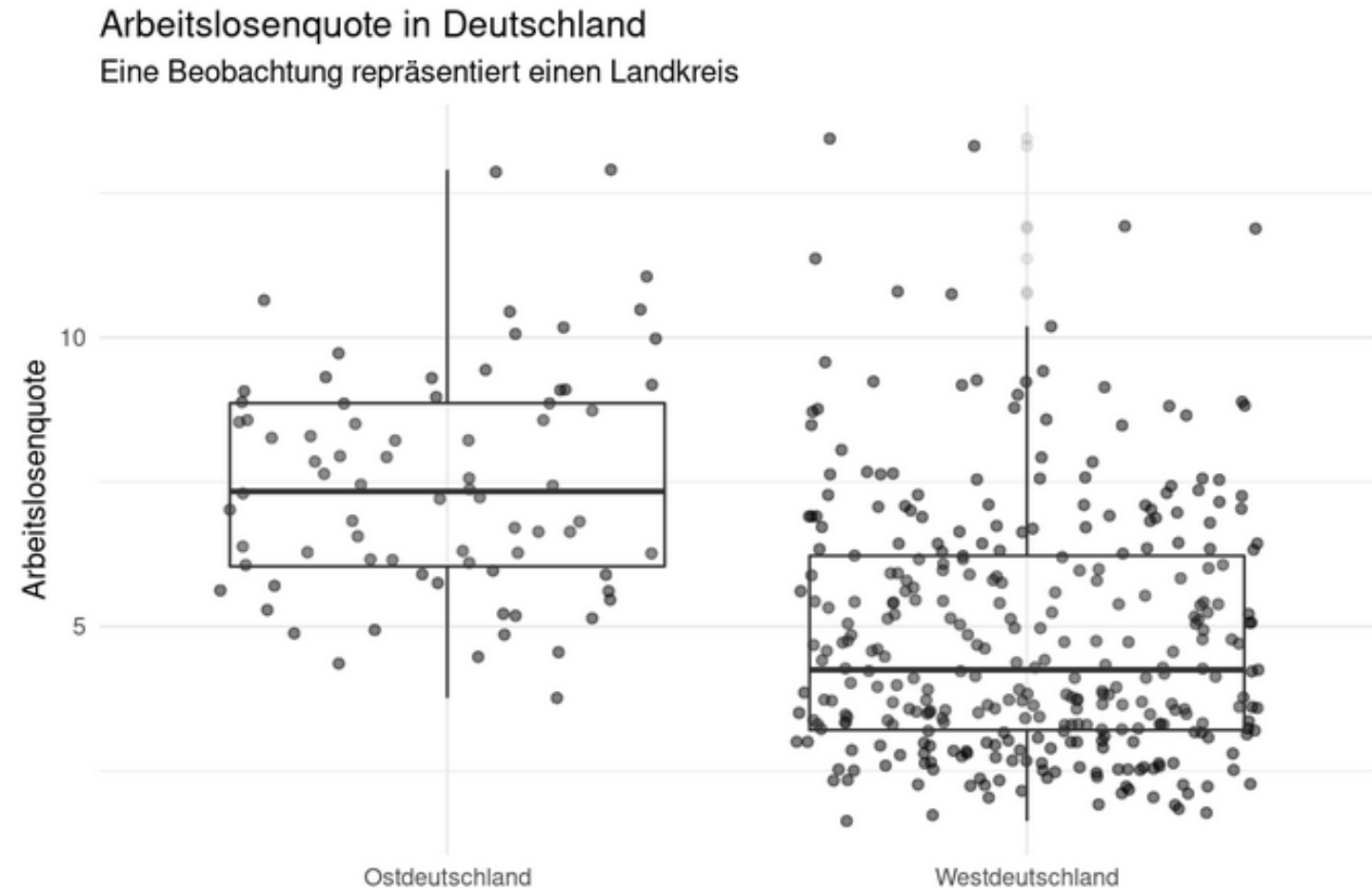
Quelle: Daten der Agentur für Arbeit aus dem Jahr 2017

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

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

Arbeitslosigkeit in Deutschland
Dargestellt ist die Arbeitslosenquote für alle Landkreise in 2017

Arbeitslosenquote

NA
>10%
8 - 10%
6 - 8%
4 - 6%
2 - 4%
0 - 2%

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

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

Arbeitslosigkeit mit Interaktionstermen

	Alo-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

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

Source: RTutor Problem Sets, 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: [RTutor Problem Sets](#), 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)

  - Questions become more complex over the two semesters (descriptive statistics $\rightarrow$ estimation of causal effects)
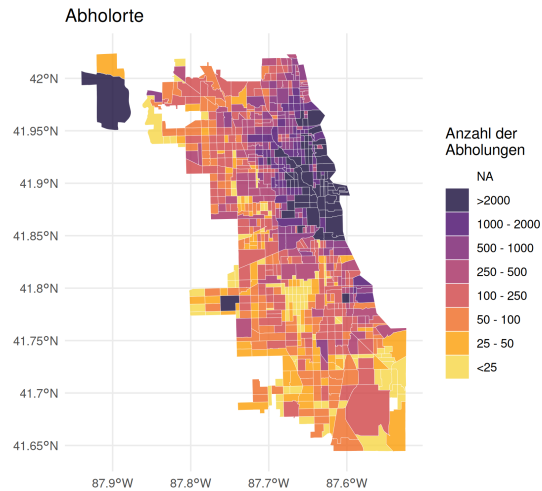
Uber driver in Chicago
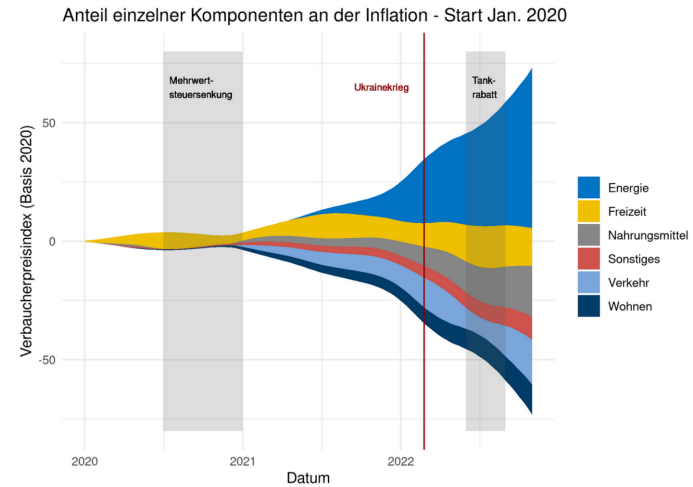
Inflation in Europe
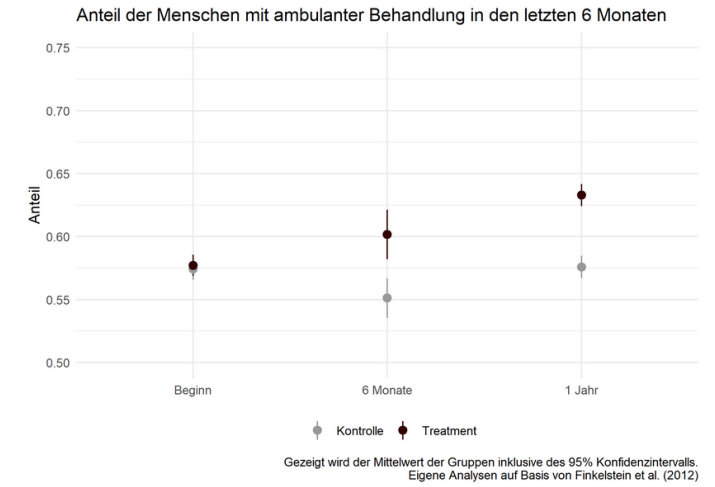
Oregon Health Insurance Experiment

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.

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**
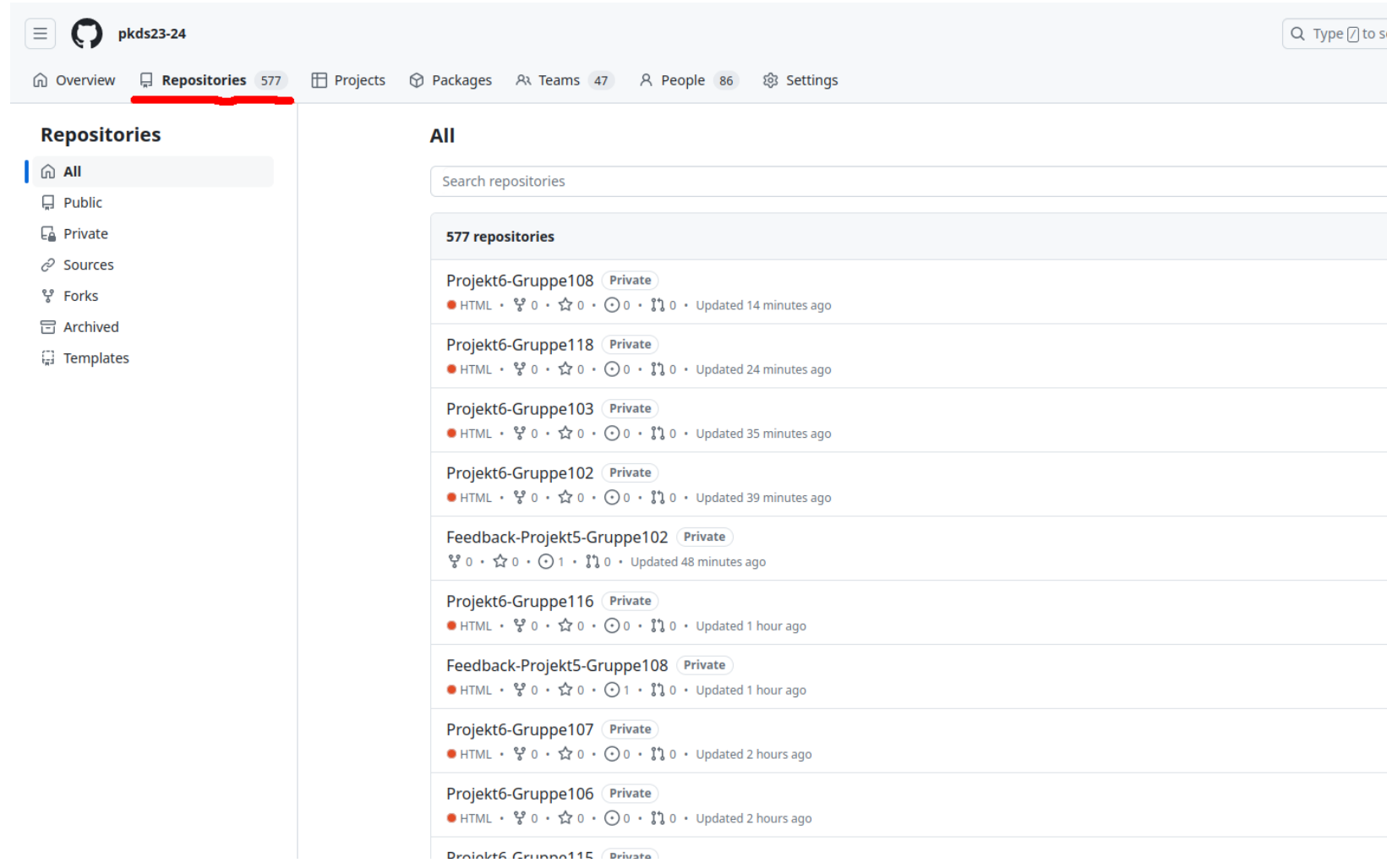
  ✚ What if a team member doesn't contribute? (Free-Rider Problem)

  ✚ **Solution**: Provide clear steps and resources for resolving such issues.

THE PROJECT PHASE: PROJECT DISTRIBUTION

How to do project distribution and maintainance?

Use the `ghclass` Package to manage hundreds of repos with R!

How to verify project Rmds?

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

```
on:
  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
      run: |
        R -e "install.packages('remotes')"

    - name: Install missing R packages
      run: |
        Rscript install_packages.R

    - name: Check Rmd Rendering
      run:  Rscript -e "rmarkdown::render('Name_of_Project.Rmd')"
```

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

+ **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 RTutor
    + Students do projects with RMarkdown or Quarto
    + 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