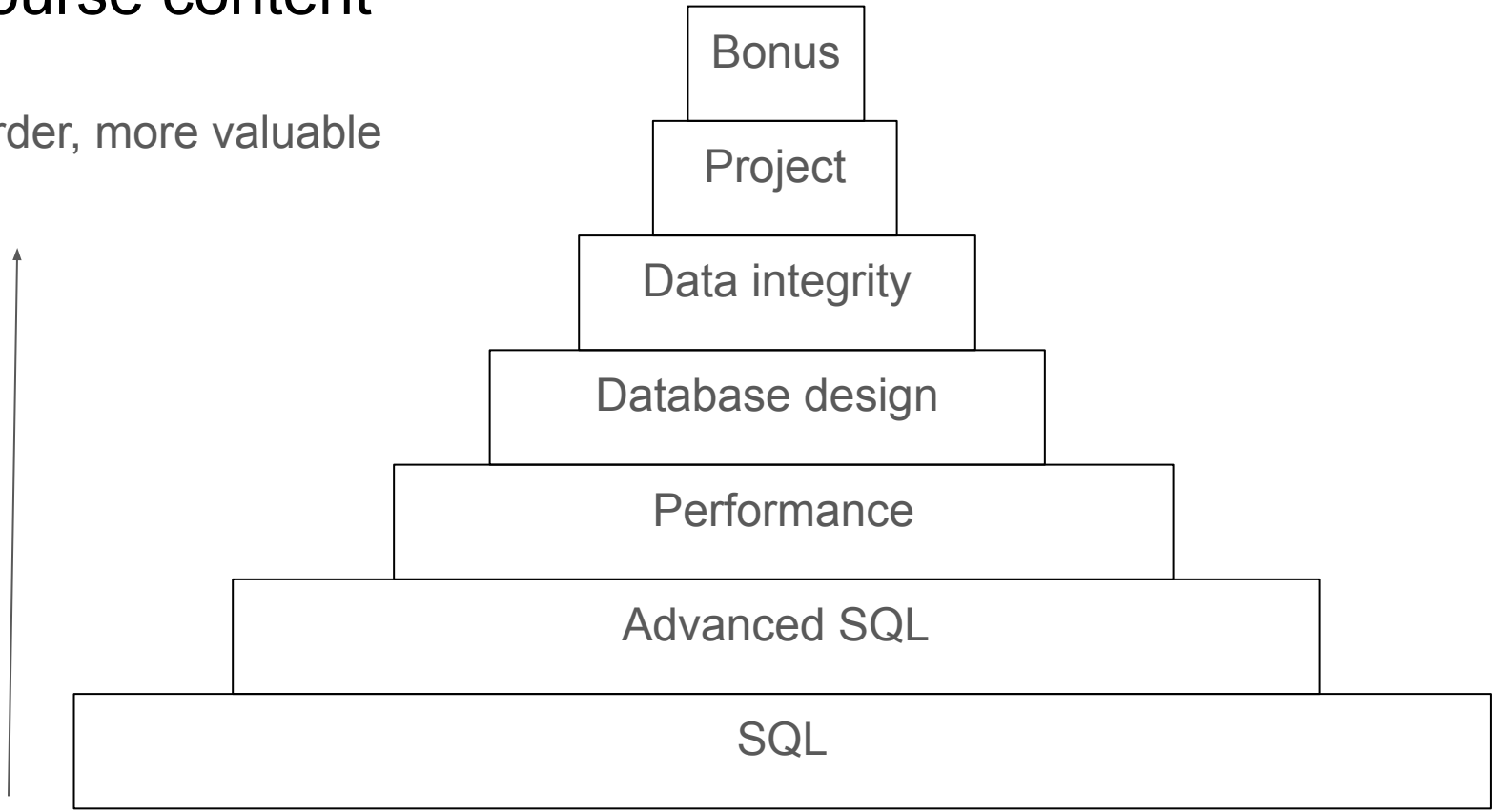
Databases for data analytics

https://github.com/evidencebp/databases-course/

Administration

Course content

Harder, more valuable



Guidelines

- Learning a language requires a lot of training
  - Acing the course is easier if you come prepared
  - See https://github.com/evidencebp/databases-course/ for many resources
- Ongoing project to practice and understand
- Understanding why is more important than how
- **This is a course about thinking.** There are many answers to the same question. You should find the one the fits your needs and explain why.
- רן בר זיק, "ללמוד MySql בעברית"

Grade structure

- Non-graded assignments
- Graded assignments (20%)
- Project (20%)
- Exam (60%)
- Bonus tasks - usually 2 points per task, up to 5 points

Course website

- https://github.com/evidencebp/databases-course

Content

- IMDB content
- Assignments
- Past exams and answers
- Enrichment questions and answers
- Lecture notes
- Presentations
- Content for the project
- Link to tutorials and relevant content

Assignments

- Appear already on course site.
- Content might change a bit (e.g., clarifying your questions)
- All tasks must be submitted
  - None graded tasks should let you evaluate your current skill and know that you are on schedule
- Any questions regarding the assignments should be directed to the teaching assistant
- Late submission will be possible only in extreme justified situation, and will lead to grade reduction.

Project - movies personal recommendation system

- You will build a personal recommendation system during the course
- The goal is to let you
  - Use SQL for your needs and not for question answering
  - Tackle and cope with real R&D problems
  - Build data science intuition
  - Get a portfolio project to present
- You can do the project in pairs.
  - You might be ask to present and justify the project
- In order to help you advance and get feedback, you will do sub-tasks during the semester and finish it afterwards

Bonus tasks

- Bonus tasks deal with advanced topics, helping to make the course better
  - Solving questions
  - Writing lecture notes
  - Etc.
- In general, one can obtain 2 points per task
- Maximum bonus is 5 points

Exams

- See site for past exam examples
- Note that this year the course is 4 points with additional content of
  - Performance (in the exam)
  - Variable schemas (json/xml)
- The exams will have **more difficult** bonus questions
  - So a minor mistake will not prevent you from getting 100
- The course is about **thinking**, not detail memorization
  - You will have a course cheat-sheet
  - Typos and similar mistakes are not important
- The course is about **thinking**
  - Many times there is no single correct answer. Your justification is the important part.

Use of AI

- You can use whatever you want
- However, it is your grade that will be reduced due to
  - AI mistakes
  - Not being able to stand behind your answers
- AI is like auxiliary wheels, you can start faster but you will pay in the long term.
- PLEASE - TRY TO SOLVE FIRST ALONE
  - Getting feedback from AI later is great
- Warning - last semester AI did not do well on the advanced SQL questions.

Additional help

- I’ll give personal time before each class. Please coordinate before to avoid few people at the same time.
- You can also coordinate time with the teaching assistant.
- You can ask question in th Moodle.
    - This is the place for questions in which others will be interested too.

Feedback

- Feedback is the key to improvement
- You feedback is valuable to me and I actively look for it.
- Please provide as much as possible