Data Mining & Discovery
INFO 523

Policies, etc.

My info

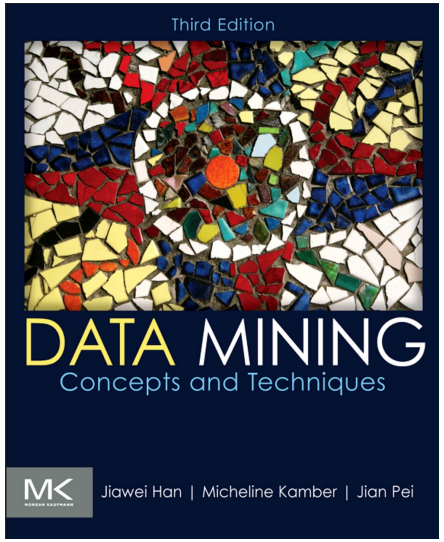
- Dr. Greg Chism, Harvill 420D.
- Office hours – M: 2-3pm
- Email me at least one day in advance for additional office hours (gchism@arizona.edu).

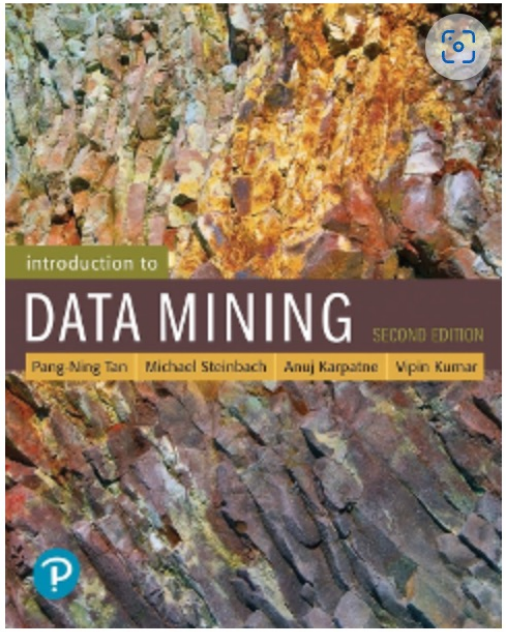
Course website & slack

- Website: datamineaz.org
- Slack: datamineaz.slack.com

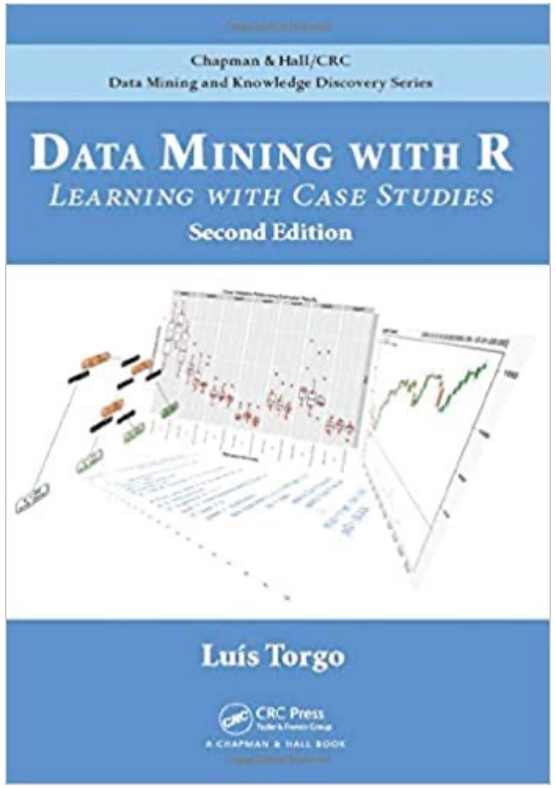
Books

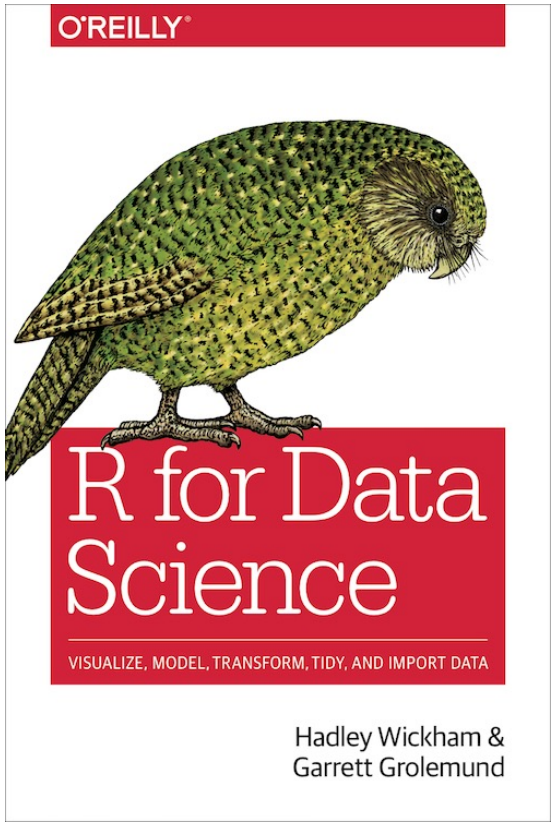
Unfortunately, there's not a single **free** book that includes all the relevant topics at once. I have selected several of them that are excellent at certain aspects. Books are posted in D2L (under **Start Here**) and I will be linking the relevant chapters to each week.





For conceptual discussions…





On the applied side of DM…

Technology requirements

- R and R Studio
  - Install R 3.4 or above
  - Install a recent version of R Studio
- Interested in using python? Please reach out!
- A working laptop or desktop computer
- A GitHub account (We will be using GitHub classroom from Week2)

Group work (optional)

- Please read carefully the group work policy in the syllabus
  - Teamwork means “no free riders”.
  - Teammates are expected to help each other.
- Find teammates using Slack
  - **Please introduce you to the class by posting the following info to the Slack forum “Self-Introduction”**

  - Name, home department, background in statistics, data mining, R, the day in a week you plan to work on this course, your work style, and your goals. Any information that help students find the teammate is welcome.

Do NOT submit the same assignment as your classmate…work on your own document. Academic misconduct = 0 in that homework.

Class schedule

- Weekly lecture (~2.5h): Basic R, git, R packages, plotting in R, basic stats, supervised/unsupervised techniques…
- Weekly homework and assigned readings
- Presentations (last week)
- You'll be improving over some data mining code (HW6 or 7).

Workload

- If you find yourself overwhelmed with the amount of work, please reach out as soon as you can!! I can adjust the number of questions in your HWs, extend deadlines, provide additional accommodations, etc.

- *From my end, I will respond your emails within a day, hold office hours each week (1hr), and give feedback within a week from the deadline.*
  - *Note however that I will not be emailing on weekends, nor will I expect you too either.*

Course organization

- Course website
  - Suggested readings
  - Lectures
  - **Homeworks (6-7)**
  - **Final project. See course schedule.**
- Schedule may shift depending on our progress
- Student-instructor communication
  - Use the office hours
  - Slack (Course questions, basically all communication)
  - gchism@arizona.edu (for absences and personal info)

Homework submissions

- After HW1, detailed instructions will be in the header of the HW AND GitHub repo.
- All HW submissions must be in RMD or **QMD**.
- I won't grade .R, .txt, etc. files.
- You can save data and variables in an R workspace (and include the workspace in your submission)
- Should be able to reproduce (render) everything from your RMD or QMD file and the files you provide.
- **I won't grade your assignment if I cannot knit it. Please make sure your RMD or QMD knit properly before submitting (You'll probably have access to results through github actions).**

How is this course actually structured?

- Each week, students are expected to approach the materials in the following order.

- 1. Go over the suggested readings
- 2. Review the lectures and slides
- 3. Ask questions (e.g., email the instructor, sign up for office hours, post in Slack)
- 4. Work on the HW (also #3)

Grading

- Homework/Exercises (70%): 6 R Exercises; 6–7 HWs
- Final project (30%)

Homework this week

1. Read the syllabus carefully and go over the course schedule on the course website.

2. Install R, RStudio (or use **posit cloud**), and git.

3. Create a GitHub account.

4. Introduce yourself on Slack.

5. **Go over HW1 (a "Quiz" linked in D2L and the course website about what data mining is…do your readings before)**

*If you disagree with any of the answers in the Quiz – please email me or schedule office hours! I'm open to discussion!!