Introduction to DA1

Ágoston Reguly

Data Analysis 1: Exploration

2020

This course

1. This course introduces data collection and data wrangling (management), presentation and understanding of descriptive statistics and basics of visualization.

2. Cover classical statistics methods and their applications, such as data collection and sampling, generalization from the sample to the population and hypothesis testing.

Exploration

- Data Analysis 1 is about exploration
- Figuring out where it comes from, how it's structured, describing and understanding some key patterns
- Exploring data is a process

Exploration

- Data Analysis 1 is about exploration
- Figuring out where it comes from, how it's structured, describing and understanding some key patterns
- Exploring data is a process

START with idea

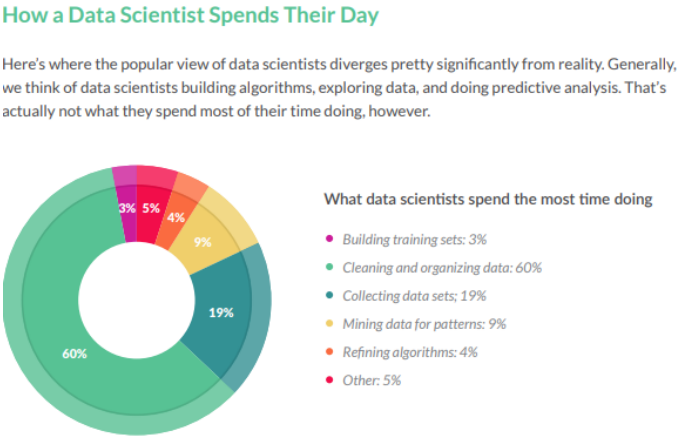
1. writing code –>
2. getting some result –>
3. interpreting that result –>
4. improved/altered idea –>
5. writing code –>
6. getting some result –>
7. interpreting that result –>
8. improved/altered idea –>

...

STOP if happy / run out of time

Data management is key task

- About 80% of data science tasks are composed of managing data, from understanding and altering features of the dataset and variables, to combining various datasets.



Data Analysis 1: Exploration - topics

1. Origins of data (data table, data quality, survey, sampling, ethics)
2. Preparing data for analysis (tidy data, source of variation, variable types, missing data, data cleaning)
3. Exploratory data analysis, Describing variables (probability, distributions, extreme values, summary stats), data visualization
4. Comparison and correlation (conditional probability, conditional distribution, conditional expectation, visual comparisons, correlation)
5. Generalizing from a dataset (repeated samples, confidence interval, standard error estimation via bootstrap and formula, external validity)
6. Testing hypotheses (null and alternative hypotheses, t-test, false positives / false negatives, p-value, testing multiple hypotheses)

Office hours

Feel free to write me any time!

- reguly_agoston@phd.ceu.edu
- *Office hours*: Monday 10:30-12:00 or by appointment at N13 220.
- Weekends - I am not checking my email.

Course Material - Data Analysis Textbook

Gábor Békés (CEU) and Gábor Kezdi (U. Michigan):

- *Data Analysis for Business, Economics and Policy* (It is almost done...)
- Mixing intro statistics and key ideas from data science with case studies
- Official website: `https://gabors-data-analysis.com/`
- Github repo for case studies
- Data also available - read the 'readme.rtf' files about availability and for ethical usage!

Part 1: Exploration is for DA1

- Six lectures - six chapters - handouts and read-only pdf
- Slides on moodle

Coding with R (2 credits / 12 weeks) is a complement to this course.

Quiz, Exam, Assignment

- Start-of-the-class Quizzes
  - Past lecture material
  - Simple question, close to practice question at the end of handout chapter
  - 5 quiz - consider best 4, each 2.5 points.
- Assignment in teams
  - Wait 3 slides...
- Closed book exam on 20th of October 5.30-7.00 pm CET
  - Textbook chapter 1-6
  - All sections unless otherwise noted
  - 75 minutes, short questions, like end of chapters
  - Practice questions will be provided to help preparation!

Grading policy

- 10% quiz, 20% group assignment, 70% exam
- To pass, students will need to get at least 50% of the overall grade AND at least 50% of the exam.
- Lectures - can not miss more than 2 - measured as quiz submitted.
  - In case of online participation: write me with the reason and in the next class you can do both quizzes.

Extra

- End of chapter - Data exercises
  - Submit any 3 to get bonus points
  - * Easy/quicker - 1%
  - ** Harder/longer - 2%
- Suggesting other useful resources/materials with short presentation (2 min in class)
  - Reference your resource/material on slack channel
  - Scientific article - 2%
  - Useful forum/community - 1%
- Max extra points in DA1 overall is 6%.

Assignments - teams

Team Assignment

- Form 3-member teams by next class.
- Teams must be (i) mixed gender and/or (ii) have at least 2 nationalities.
- Sign up for teams on moodle. You may propose your team's name! Write me in email.

Assignments - the task

The task

- Visit restaurants and collect prices on margheritha pizza and on a optional beverage served in 0.5 l volume. You should visit at least 20 places in inner Budapest and 20 places which delivers pizza to CEU Budapest Campus.
- Use any districts in Budapest from the "inner city". Inner city is defined as districts: 5,6,7,8,9, 13(Pest side) , 2, 11 (Buda side). Everything else is outer Budapest.
- With the collected prices, create a dataset with tidy table(s)
- For each restaurant, please register the address, and 2-3 key features of the shop (e.g. area in m^3, stars of the restaurant, size of the pizza, etc.)

Assignments - deadlines

Task 1. - Collect data (8%)

- Create a csv, xlsx or your favorite data container which is readable in R.
- Deadline 18 October Sunday 23.55.am. (upload to ceu-learning site)

Task 2: Create a report describing your data. (12%)

- Joint assignment with Coding 1.
- TO be specified after class 3.
- Deadline: 25 October Sunday 23.55 (upload to ceu-learning site)