Advanced Data Analysis

Anna Shirokanova, HSE SPb, Fall 2020

For whom is this course?

- If you want a better structure and armamentarium to your data analysis practice
- If you are academically versed
- If you did not attend the Data Science or Programming Minor but you can do numbers

Learnings goals for the course

- To collect Web data using basic string patterns
- To understand and impute missing data
- To model binary outcome data
- To reduce data complexity based on object distance
- To improve visualization, data mgmt and coding patterns
- To understand the basics of Bayesian inference

Prior knowledge requirements

- Very good command of base data analysis courses (OLS, factor analysis)
- Tidyverse

Expected workload

- 1-2 modules, 48/180/228
- Starting offline but likely to move online soon
- One double class per week
- Class up to 30 participants, need a laptop

Key graded activities

- Two data analysis projects (one per module)
- Smaller tasks or projects every two weeks (6)
- Exam test (to check understanding)

Grading

Activity	%	Deadline
Web data, regular expressions	**10**	Mid-Sept
Data imputation	**10**	End of September
Logistic regression project	**20**	Mid-October
Distances, MDS	**10**	Early November
Cluster analysis project	**20**	End of November
R workshop, viz quiz, Bayes	**20**	Early December
Exam test	**10**	Mid-December

Working environment

- LMS (single info point)
- Zoom (online meetings)
- Dropbox for materials and projects
- GoogleForms for the bi-weekly viz quiz and feedback

Extras

- Regular expressions
- Custom themes in graphs
- Better visualizations
- Cleaner code
- Nerdic discussions

Weekly timetable

Week	Topic	Hours	Homework
1	Overview and Grammar of graphics	4	Rmd template
2-3	Web data, data imputation	8	Dataset
4-5	Binary logistic regression	8	Home project
6-9	Distances, MDS, PCA, cluster analysis	16	Data collection, home project
10-12	Data culture (reproducible science, Bayesian stats)	12	R Gems seminar, discussions
			Viz quiz and feedback (bi-weekly)