University of St.Gallen

Course and Examination Fact Sheet: Autumn Semester 2024

9,272: Big Data Analytics

ECTS credits: 4

Overview examination/s

(binding regulations see below)
decentral - Presentation, Analog, Group work group grade (40%)
Examination time: Term time
decentral - Written work, Digital, Group work group grade (60%)
Examination time: Term time

Attached courses

Timetable -- Language -- Lecturer
9,272,1.00 Big Data Analytics -- English -- Heiniger Sandro

Course information

Course prerequisites

- 'A Brief Introduction to Programming with R' in the integration week (or an equivalent basic R programming course).

Learning objectives

- Students will know the concept of Big Data in the context of empirical economic research.
- Students will understand the technical challenges of Big Data Analytics and how to practically deal with them.
- Students will know how to apply the relevant R packages and programming practices to effectively and efficiently handle large data sets.

Course content

Short summary
This course introduces students to the concept of Big Data in the context of empirical economic research. Students learn about the computational constraints underlying Big Data Analytics and how to handle them in the statistical computing environment R (local and in the cloud). Revisiting basic statistical/econometric concepts, we look at each step of dealing with large data sets in empirical economic research (storage/import, transformation, visualization, aggregation).

Description
The increasing size of datasets in empirical economic research (both in number of observations and number of variables) offers new opportunities and poses new challenges for economists. 'Big Data' is discussed as the new 'most valuable' resource in highly developed economies, driving the development of new products and services in various industries. Extracting knowledge from large data sets is increasingly seen as a strategic asset for firms, governments, and NGOs. Successfully navigating the data-driven economy presupposes understanding the technologies and methods used to gain insights from Big Data.

This course introduces students to the basic concepts of Big Data Analytics to gain insights from large and complex data sets. Thereby, the focus of the course is on the practical application of econometrics/machine learning, given large/complex datasets. The course does NOT (or only to a very limeted degree) introduce basic econometric/machine learning concepts/models. It is, therefore, crucial that students taking this course are already equiped with solid knowledge in statistics/econometrics (and basic knowledge in machine learning). The course combines conceptual/theoretical material with the practical application of the
concepts with the open source programming language R. Thereby, students will acquire the basic skillset of analysing large data sets both locally and in the cloud. The practical applications of the learned techniques are focused on empirical research in economics and the social sciences.

The first part of the course covers the basics of computation (in an applied econometrics context). Students learn about the physical constraints of standard computers used for data analytics and learn how to identify bottle-necks in data analysis tasks and how to identify them within the R environment. Students then learn how to handle the identified computational constraints with R (and related tools such as Keras and Spark), first locally and then in the cloud. Thereby, the course covers each step of the data pipeline in economic research (storage/import, transformation, visualization, aggregation, model estimation).

Course structure and indications of the learning and teaching design

Students will apply the relevant R packages and programming practices during the course to effectively and efficiently handle large data sets. After guided application of concepts in R, students are required to apply concepts in R independently in groups as well as present their approach/strategy and results in class. The course proceeds as follows:

Part I: The Basics

- Introduction: Big Data, Data Economy, Course Overview Walkowiak (2016): Chapter 1
- Computation and Memory in Applied Econometrics
- Advanced R Programming (Concepts/Applied) Wickham (2019): Chapters 2, 3, 17, 23, 24.

Part II: Local Big Data Analytics

- Import, Cleaning, and Transformation of Big Data (Applied) Walkowiak (2016): Chapter 3: p. 74-118.
- Aggregation and Visualization (Applied: data tables, ggplot) Walkowiak (2016): Chapter 3: p. 118-127; Wickham et al. (2015); Schwabish (2014).
- Data Storage, Databases Interaction with R Walkowiak (2016): Chapter 5.

Part III: Advanced Topics

- Cloud Computing: Introduction/Overview (Concepts)
- Machine Learning and GPUs
- Distributed Systems, MapReduce/Hadoop with R (Concepts/Applied)Walkowiak (2016): Chapter 4.
- Applied Econometrics with Spark

Other references will be mentioned for each relevant chapter separately.

Course literature

Main textbooks

- Matter, Ulrich (2023): Big Data Analytics: A Guide to Data Science Practitioners Making the Transition to Big Data, CRC Press .
- Walkowiak, Simon (2016): Big Data Analytics with R. Birmingham, UK: Packt Publishing.
- Wickham, Hadley (2019): Advanced R. Second Edition, CRC Press, FL: Boca Raton.

Journal articles and additional books

- Wickham, Hadley and Dianne Cook and Heike Hofmann (2015): Visualizing statistical models: Removing the blindfold. *Statistical Analysis and Data Mining: The ASA Data Science Journal.* 8(4):203-225.
- Schwabish, Jonathan A. (2014): An Economist's Guide to Visualizing Data. *Journal of Economic Perspectives.* 28(1):209-234.

Additional course information

--

Examination information

Examination sub part/s

1. Examination sub part (1/2)

Examination modalities

Examination type	Presentation
Responsible for organisation	decentral
Examination form	Oral examination
Examination mode	Analog
Time of examination	Term time
Examination execution	Asynchronous
Examination location	On Campus
Grading type	Group work group grade
Weighting	40%
Duration	--

Examination languages

Question language: English
Answer language: English

Remark

Presentation of analytics project. Teams of 3-4.

Examination-aid rule

Free aids provision

Basically, students are free to choose aids. Any restrictions are defined by the faculty members in charge of the examination under supplementary aids.

Supplementary aids

Each group must prepare presentations independently without undue outside help, and used sources and aids must be properly cited.

---

2. Examination sub part (2/2)

Examination modalities

Examination type	Written work
Responsible for organisation	decentral
Examination form	Written work
Examination mode	Digital
Time of examination	Term time
Examination execution	Asynchronous
Examination location	Off Campus
Grading type	Group work group grade
Weighting	60%
Duration	--

Examination languages

Question language: English
Answer language: English

Remark

Take-home exercise-set. Groups of 3-4.

Examination-aid rule

Free aids provision

Basically, students are free to choose aids. Any restrictions are defined by the faculty members in charge of the examination under supplementary aids.

Supplementary aids

All the aids that violate the following principles are not allowed:

1. Written assignments must be accomplished independently by each group without undue outside help; used sources and aids must be properly cited.

2. During group work, all members must contribute reasonably. Without such contribution, a person may neither be named as a co-author nor ask for inclusion in the author list.

3. Undue aid may not be offered to others. This includes asking members from other groups to help write the codes and provide solutions.

4. Open source codes available on various web resources should be properly cited and minimally used. Proactive posting requests of the ready solution on web resources are not allowed.

Examination content

- 60% - Take-home exercises as part of the students project, solved in teams of 3-4 students: Technical report on the project.
  Conceptual questions related to the application.
- 40% - Students' project presentation (same teams of 3-4 students): Own application of concepts in R, approach/strategy
  and results presented in class.

Examination relevant literature

Main textbooks

- Matter, Ulrich (2023): Big Data Analytics: A Guide to Data Science Practitioners Making the Transition to Big Data, CRC Press .
- Walkowiak, Simon (2016): Big Data Analytics with R. Birmingham, UK: Packt Publishing.
- Wickham, Hadley (2019): Advanced R. Second Edition, CRC Press, FL: Boca Raton.

Journal articles and additional books

- Wickham, Hadley and Dianne Cook and Heike Hofmann (2015): Visualizing statistical models: Removing the blindfold. *Statistical Analysis and Data Mining: The ASA Data Science Journal.* 8(4):203-225.
- Schwabish, Jonathan A. (2014): An Economist's Guide to Visualizing Data. *Journal of Economic Perspectives.* 28(1):209-234.

University of St.Gallen

Please note

Please note that only this fact sheet and the examination schedule published at the time of bidding are binding and takes precedence over other information, such as information on StudyNet (Canvas), on lecturers' websites and information in lectures etc.
Any references and links to third-party content within the fact sheet are only of a supplementary, informative nature and lie outside the area of responsibility of the University of St.Gallen.

Documents and materials are only relevant for central examinations if they are available by the end of the lecture period (CW51) at the latest. In the case of centrally organised mid-term examinations, the documents and materials up to CW 42 are relevant for testing.

Binding nature of the fact sheets:

- Course information as well as examination date (organised centrally/decentrally) and form of examination: from bidding start in CW 34 (Thursday, 22nd August 2024);
- Examination information (supplementary aids, examination contents, examination literature) for decentralised examinations: in CW 12 (Monday, 18 March 2024);
- Examination information (supplementary aids, examination contents, examination literature) for centrally organised mid-term examinations: in CW 42 (Monday, 14 October 2024);
- Examination information (regulations on aids, examination contents, examination literature) for centrally organised examinations: two weeks before ending with de-registration period in CW 45 (Monday, 04 November 2024).