

**Humanitarian Department AGH**  
**2015/2016 Academic Year**

**Academic sphere:** sociology

**1. Subject title:** Data Mining and media analysis

**2. Subject duration:** 30 hrs.

**3. Lecturer:** Prof. Bohdan Yuskiv, PhD ([yuskivbm@gmail.com](mailto:yuskivbm@gmail.com))

**4. Classes form:** lecturing + laptop-based laboratory works

**5. Objective:**

introducing students to data mining analysis of media

**6. Subject requirements:**

- knowledge of the subject-related basics (statistics and related methods),
- PC user knowlegde,
- no more than 20 persons per group,
- Wi-Fi Internet access is required.

**7. Subject results:**

- students will get knowledge in data mining analysis, characterizing basics of data processing, describing the data exploration as well as specific data mining methods;
- students will use new methods in the R language to perform deep data analysis in the media;
- students will be able to obtain additional information from the data using scientific approach.

**8. Subject contents with specific classes:**

CLASSES:

1. Introducing the Data Mining practice (2 hrs.)
2. Data Import and Export (2 hrs.)
3. Data exploration methods (2 hrs.)
4. Data and results visualization (2 hrs.)
5. Data clusters and factors analysis algorithms (2 hrs.)
6. Bagging and boosting methods (2 hrs.)
7. Time series. The SARIMA model prognosis (3 hrs.)

LABORATORIAN WORKS (based on the media-knowledge examples)

1. Introduction of R basics as a statistics calculation environment. Program installation (3 hrs.)
2. Data import and export in R (2 hrs.)
3. Pre-analysis data exploration in R (2 hrs.)
4. The R visualization (2 hrs.)
5. Data clusters and factors analysis in R (2 hrs.)
6. Bagging and boosting methods in R (2 hrs.)
7. Time series. The SARIMA model prognosis in R (2 hrs.)

## PROGRAMS / SUBJECT TOOLS:

- R
- RStudio
- R packages

**9. Literature****9.1. Recommended reading**

1. Crawley Michael J. *The R Book*, John Wiley & Sons Ltd 2007.
2. Ledolter Johannes *Data mining and business analytics with R*, University of Iowa, Wiley 2013.
3. Maindonald John, Braun W. John *Data Analysis and Graphics. Using R – an Example-Based Approach*, Cambridge University Press 2003.
4. Paradis Emmanuel *R for Beginners*, Institut des Sciences de l' Evolution 2005.
5. *Statystyczna analiza danych z wykorzystaniem programu R*, pod red. nauk. M.Walesiaka, E.Gatnara, Wydawnictwo Naukowe PWN, Warszawa 2013.
6. Torgo Luis *Data Mining with R: learning with case studies*, LIACC-FEP, University of Porto 2003.
7. Yanchang Zhao *R and Data Mining: Examples and Case Studies*, Elsevier 2012.

**9.2. Additional literature**

8. Adler Joseph *R in a Nutshell*, O'Reilly 2010.
9. Coghlan Avril *A Little Book of R For Multivariate Analysis*, [https://github.com/avrilcoghlan/LittleBookofRMultivariateAnalysis/raw/master/\\_build/latex/MultivariateAnalysis.pdf](https://github.com/avrilcoghlan/LittleBookofRMultivariateAnalysis/raw/master/_build/latex/MultivariateAnalysis.pdf).
10. Gatnar E. *Podejście wielomodelowe w zagadnieniach dyskryminacji i regresji*, PWN, Warszawa 2008.
11. Kabacoff Robert I. *R in Action. Data analysis and graphics with R*, Manning Publications Co 2011 (Кабачков Роберт И. *R в действии. Анализ и визуализация данных в программе R*, ДМК Пресс, Москва 2014).
12. Larose D.T. *Metody i modele eksploracji danych*, PWN, Warszawa 2008.
13. Pałka Dariusz, Zaskórski Piotr, *Data mining w procesach decyzyjnych*, Zeszyty Naukowe Warszawskiej Wyższej Szkoły Informatyki, nr 7, Warszawa 2012, s. 143-161.
14. Stanisław A. *Przystępny kurs statystyki. t III Analizy wielowymiarowe*, StatSoft, Kraków 2005.
15. StatSoft: *Internetowy Podręcznik Statystyki. Techniki zgłębiania danych (data mining)*, <http://www.statsoft.pl/textbook/glosfra.html>.
16. Venables W. N., Smith D. M. and the R Core Team *An Introduction to R. A Programming Environment for Data Analysis and Graphics*, 2015.
17. Vries Andrie de, Meys Joris *R For Dummies*, John Wiley & Sons, Ltd 2012.