



**Faculty of Human Sciences, University of Cologne**

## **Workshop: “Introduction to Data Science with Python“ (advanced)**

**Dr. Arnim Bleier, Dr. Fabian Flöck, Dr Juhi Kulshrestha**

**14-16 October 2019, 9:30am to 5:30pm**

Data Science is the interdisciplinary science of the extraction of interpretable and useful knowledge from potentially large datasets. Due to the rapid surge of digital trace data (often as “Big Data”) in a wide range of application areas, Data Science is also increasingly utilized in the social sciences and humanities. In contrast to empirical social science, Data Science methods often serve purposes of exploration and inductive inference. In this course, we aim to provide an introduction into Data Science for practitioners. In particular, we want to impart basic understanding of the main methods and algorithms and understand how these can be deployed in practical application scenarios.

For that purpose, our schedule alternates between lecture sessions that present the theoretical and technical background of data analysis and practical sessions that allow participants to directly apply acquired knowledge with code in the Python programming language. The workshop will cover aspects of data visualization, and machine learning, using basic Python and key packages. Data used will cover a large array of sources, from “native Web” data such as Social Media data to more “traditional” survey data.

Participants will obtain extensive knowledge about typical data types and structures encountered when dealing with digital behavioral data, state-of-the-art data analysis methods and tools in Python, and they will learn how this approach differs from those typically encountered in survey-based or experimental research. This will enable them to identify benefits and pitfalls of these data types and methods in their field of interest and will thus allow them to select and appropriately apply data analysis and machine-learning methods for large datasets in their own research. The knowledge obtained in this course provides a starting point for participants to investigate specialized methods for their individual research projects.

Participants should be willing to study algorithmic approaches on abstract and applied levels. Some previous knowledge on (i) statistics as well as (ii) programming in Python, another programming language (like R, Java) or at least scripting language (Syntax-Code in SPSS, Stata) is very advantageous to follow the coursework - otherwise the learning curve will be quite steep.

To ensure a common starting level between participants, we expect attendants to familiarize themselves with the basic concepts of Python such as variables, lists, and loops via provided learning materials beforehand.

The workshop will include the following modules:

- Data visualization
- Machine learning – Intro
- Machine learning - Supervised methods
- Machine learning - Unsupervised methods

For those who are interested in the workshop but are not familiar enough with basic Python terms and applications or programming in Python to be able to follow this workshop seamlessly, an additional introductory workshop: "Introduction to Python" will be offered on 1-2 October. The modules of the introductory workshop will be adjusted by the instructors to the requirements of this Data Science course. For those who are inexperienced with Python or are beginners, we strongly recommend to also attend the introductory workshop in order to have a better starting point for participating in the Data Science Workshop.

**Date:**

14-16 October, 2019, 9:30 am to 5:30 pm

**Venue:**

Regional Computing Centre of the University of Cologne (RRZK)  
Course Room 4 (Basement, -1.02)  
Building 133  
Weyertal 121  
50931 Cologne  
Interactive map: <http://lageplan.uni-koeln.de/#!133>

**Registration:**

Participation is limited to 15 persons.

Therefore, we would like to ask you to register via e-mail to [Graduiertenschule-HF@uni-koeln.de](mailto:Graduiertenschule-HF@uni-koeln.de) by **Friday, 27th September 2019**.

Those who have already taken part in the survey of interests during the lecture period in summer semester 2019 will be given priority in the registration process.

**Lecturers:**

Dr. Arnim Bleier is a postdoctoral researcher at the Department Computational Social Science at GESIS. His research interests are in the field of Natural Language Processing and Computational Social Science. In collaboration with social scientists, he develops Bayesian models for the content, structure and dynamics of social phenomena.

Dr. Fabian Flöck studied communication sciences and sociology, and subsequently acquired a PhD in computer science. Specifically, he developed algorithmic methods to extract rich behavioral traces from Wikipedia editing data and studied them with data science methods. He is a postdoctoral researcher at the Computational Social Science department at GESIS and interested in collaborative content production, crowdsourcing and data visualization.

Dr. Juhi Kulshrestha is a postdoctoral researcher at the Computational Social Science Department at GESIS. She obtained her PhD at the Max Planck Institute for Software Systems (MPI-SWS) in 2017. Her research focuses on how users consume news and information on online social media and on evaluating the role played by automated retrieval algorithms, like search and recommendation systems, in shaping the users' news and information consumption.