



RISIS Summer School on Data Science for Studying Science, Technology and Innovation

University of Strathclyde, Glasgow, 24th-28th June 2019

Programme

Location: University of Strathclyde, 199 Cathedral Street, Glasgow, G4 0QU, Room CW306

Monday, 24th June 2019

08:30	09:00	Registration
09:00	09:15	Welcome <i>Nigel Lockett (University of Strathclyde)</i>
09:15	09:30	1.1. Introduction <i>Abdullah Gok, Bernd Wurth (University of Strathclyde)</i>
09:30	10:30	1.2. Keynote <i>Phillipe Laredo (UPEM)</i> "Towards new positioning indicators: in between semantics, open data, computer power and theory-based approaches"
10:30	10:45	Break
10:45	12:30	1.3. R as a Data Science Tool <i>Kate Pyper (University of Strathclyde)</i> The purpose of this session is to introduce the R programming language and show you how this can be used for data science. The first part will focus on the basics. We will focus on using the R Studio IDE for developing code. We will initially look at the different variable types that R can handle and how these can be manipulated, before moving on to more complex concepts like loops and conditional statements.
12:30	13:30	Lunch
13:30	15:00	1.3. R as a Data Science Tool (cont'd) <i>Kate Pyper (University of Strathclyde)</i> The second part of the session will introduce the Tidyverse as a means for data manipulation, starting with some background into what Tidyverse actually does. The concept of Tidy data will be introduced, followed by the introduction of specific tools for data manipulation from the tidyr and dplyr packages.
15:00	15:15	Break
15:15	17:00	1.3. R as a Data Science Tool (cont'd) <i>Kate Pyper (University of Strathclyde)</i> The final part of the session will continue where the first left off, discussing relational data and how the Tidyverse can be used to combine multiple data sets in a variety of ways. The session will conclude by discussing elementary data visualisation using ggplot2.

The place of useful learning

The University of Strathclyde is a charitable body, registered in Scotland, number SC015263

17:00	19:00	Reception
-------	-------	------------------

Tuesday, 25th June 2019

09:00	10:45	2.1. Publication Data Analysis <i>Rodrigo Costas Comesaña, Martijn Visser (University of Leiden)</i> A general introduction to the publications data analysis, with special focus on understanding and interpret a wide range of publication- and citation-based indicators and statistics. Practical applications will be shown by engaging participants with basic calculations.
10:45	11:00	Break
11:00	12:30	2.1. Publication Data Analysis (cont'd) <i>Rodrigo Costas Comesaña, Martijn Visser (University of Leiden)</i>
12:30	13:30	Lunch
13:30	15:00	2.1. Publication Data Analysis (cont'd) <i>Rodrigo Costas Comesaña, Martijn Visser (University of Leiden)</i>
15:00	15:15	Break
15:15	17:00	2.2. Data Sources for Publications <i>Rodrigo Costas Comesaña, Martijn Visser (University of Leiden)</i> Major scientometric data sources, including Web of Science, Scopus, Dimensions, Microsoft Academics and Google Scholar, will be discussed, including their strengths and limitations. Practical applications will be shown by engaging participants with the data collection.
18:30	22:00	Summer School Dinner <i>Gandolfi Fish, 84-86 Albion St, Glasgow, G1 1NY</i>

Wednesday, 26th June 2019

09:00	10:45	3.1. Cleaning and analysing publication data <i>Abdullah Gok, Bernd Wurth (University of Strathclyde)</i> This session will demonstrate some of the widely used tools for the cleaning and analysis of publication data including VantagePoint, OpenRefine and R bibliometrix package.
10:45	11:00	Break
11:00	12:30	3.2. Citation Analysis and Vosviewer <i>Rodrigo Costas Comesaña, Martijn Visser (University of Leiden)</i> Analysis based on funding acknowledgments, mobility, open access and altmetrics will be presented. Additionally, a short presentation of VOSviewer, a software tool for constructing and visualizing bibliometric networks, will be presented.
12:30	13:30	Lunch
13:30	15:00	3.3. Network Data Science: Theoretical and formal background <i>Thomas Scherngell, Martina Neuländtner (AIT)</i>

This session will give an overview on the theoretical and conceptual background of Network Data Science in the context of STI studies, mainly inspired from a social network, but also from an economic geography perspective. In addition, a formal network definition will be provided, from which we derive basic descriptive measures for characterising real-world R&D collaboration networks. Spatial interaction models will be introduced as a specifically useful instrument for STI studies to explain R&D collaboratoin dynamics.

15:00	15:15	Break
-------	-------	--------------

15:15	17:00	3.4. Network Data Science: Network analysis in R and some basic examples I <i>Thomas Scherngell, Martina Neuländtner (AIT)</i>
-------	-------	--

This session will demonstrate the methods as introduced in the previous session by means of the R-Cran package i-graph, using R&D collaboration networks in Advanced Manufacturing Technology. This first part will focus on basic descriptive measures from Social Network Analysis (SNA) and network visualisation.

Thursday, 27th June 2019

09:00	10:45	4.1. Network analysis in R and some basic examples II & Introduction and set-up of group work <i>Thomas Scherngell, Martina Neuländtner (AIT)</i>
-------	-------	---

The second part will shift attention to the specification and estimation of spatial interaction models in R, demonstrating how to estimate factors influencing network dynamics, again applied to the AMT R&D collaboration network. Furthermore, the group work will be set up (data provision, objectives of the group works, etc.)

10:45	11:00	Break
-------	-------	--------------

11:00	12:30	4.2. Network Data Science: Group work descriptive analysis and visualisation <i>Thomas Scherngell, Martina Neuländtner (AIT)</i>
-------	-------	--

This session will be devoted to the group work, starting with descriptive analyses and visualisation

12:30	13:30	Lunch
-------	-------	--------------

13:30	15:00	4.3. Network Data Science: Group work inferential network analysis <i>Thomas Scherngell, Martina Neuländtner (AIT)</i>
-------	-------	--

This will continue with the group work, shifting attention to the estimation and interpretation of spatial interaction models

15:00	15:15	Break
-------	-------	--------------

15:15	17:00	4.4. Network Data Science: Presentation of group work <i>Thomas Scherngell, Martina Neuländtner (AIT)</i>
-------	-------	---

The results of the group works will be presented in short presentations (10min) by each group, with respective feedback and discussion after each presentation.

18:00	23:30	Social Event: Ceilidh <i>Sloans, 108 Argyle St, Glasgow, G2 8BG</i>
-------	-------	---

Friday, 28th June 2019

09:00	10:45	5.1. Text Mining <i>Yashar Moshfeghi (University of Strathclyde)</i> Introduction of the concepts behind a Text Mining Workflow This session will provide the basic workflow for text-mining in STI. Topics will include formatting and preparing data for text-mining (e.g. string manipulation, regex, tidy text format, stemming, tokenization, etc.) as well as the tools and techniques for analysis including topic modelling.
10:45	11:00	Break
11:00	12:30	5.2. Text Mining <i>Yashar Moshfeghi (University of Strathclyde)</i> The second part will shift attention to text mining process from a practical aspect where demonstrating how to perform data preparation (tokenization, stop-words removal, etc.) as well as topic modelling in R. The session will utilise R text-mining packages.
12:30	13:30	Lunch
13:30	15:00	5.3. Text Mining <i>Yashar Moshfeghi (University of Strathclyde)</i> A hands-on text mining session This session will be a hands-on session where participants get the chance to practice the text mining workflow on a STI data set provided to them.

If you require and further information, please contact Bernd (bernd.wurth@strath.ac.uk) or Abdullah (Abdullah.gok@strath.ac.uk).

Funding:

This Summer School is supported by the RISIS project (EU H2020 Grant Agreement n°824091) and Strathclyde Business School, University of Strathclyde.

