



Workshops at WSAD Summer School

TABLE OF CONTENTS

Workshop: Introduction to R for beginners.....	2
Workshop: Going Deeper into R: Data Processing and Basic Statistical Modeling	2
Workshop: Spatial Modeling of Socio-economic Phenomena.....	3
Workshop: Text analysis for the social sciences	3
Workshop: Introduction to Social Network Analysis.....	4
Workshop: Introduction to the Statistical Analysis of Social Networks	5
Workshop: Introduction to the Statistical Analysis of Dynamic Social Networks.....	6
Workshop: Data visualization techniques. Theory and Practice	6
Workshop: Data processing with SPSS.....	7
Workshop: Discovering Patterns in Your Data – Generalized linear mixed models and other techniques with R.....	8
Workshop: Wiggles and curves: The art of data exploration	8
Workshop: How to write clearly	9
Workshop: Structural Equation Models	9





I. WORKSHOP: INTRODUCTION TO R FOR BEGINNERS

Instructor: Artur Suchwałko (QuantUp)

Prerequisites: computer literacy, understanding of descriptive statistics (like mean or standard deviation)

Statistical system R (<http://www.r-project.org/>) is a powerful and free tool. It is applied virtually in all fields of science and business. You will get very practical skills. After the workshop you will be able to analyse single-handedly your own data using R: reading them, processing, and writing back the results.

We will work hands-on with computers. Basically no lectures.

Specifically, you will learn:

- Working with R: how to deal with it, help system, further information
- Using RStudio
- Basics of R language
- Reading data in different formats
- Summaries and descriptive statistics
- Basics of plotting
- Writing data and results

II. WORKSHOP: GOING DEEPER INTO R: DATA PROCESSING AND BASIC STATISTICAL MODELING

Instructor: Artur Suchwałko (QuantUp)

Prerequisites: Completed "Introduction to R for Beginners" or having the equivalent knowledge, basic understanding of linear regression.

During this workshop you will get important practical skills: processing data for further analysis and facilitating data analysis with help of R scripts. You will understand how to work effectively with data in R and how to build first linear regression models. These altogether will build up a strong basis for using R in your work and learning more of things you need in future. We will work hands-on with computers. Basically no lectures.

Specifically, you will learn:

- Basics of data wrangling: merging, subsetting, transformation, and aggregation
- Types and properties of objects in R
- Vectorization and indexing





- Working with results of statistical procedures
- Linear regression
- Basics of writing R scripts

III. WORKSHOP: SPATIAL MODELING OF SOCIO-ECONOMIC PHENOMENA

Instructor: Katarzyna Kopczewska (University of Warsaw)

Prerequisites: Basic knowledge of R software (data import, operations on data, simple plotting)

Main goal of this course is to introduce students with the methods and applications of spatial modelling in social sciences. Using spatial statistics and econometrics together with mapping allows for:

- visualization of the modeled data
- observing and measuring spatial clusters, hot-spots etc.
- observing spatial regimes (heterogeneity over space; clubs)
- improving quality of estimators (unbiased and consistent) in regression
- controlling for spatial diffusion process, spatial interactions and spatial correlation

Course will focus on:

- short introduction to spatial packages in R software, on-line collections of maps and data
- visualization on map of socio-economic phenomena to catch spatial patterns (spatial autocorrelation, spatial heterogeneity)
- spatial statistics to detect spatial clustering, spatial regimes and autocorrelation
- spatial weights matrix – building neighborhood relations
- spatial dependence models – basic econometric modelling, including spatial relations

IV. WORKSHOP: TEXT ANALYSIS FOR THE SOCIAL SCIENCES

Instructor: Roel Popping (University of Groningen)

Length: 12h

Text analysis is a data collection technique just like many others. Usually there are no direct responses to a (research) question like agree, agree a little bit till disagree with an argument. The answers are found in a piece of text, where the investigator should try to find them. In general these are not answers to direct questions (like open-ended questions in a survey), but the texts contain opinions or facts on a specific issue that are relevant for an investigator. This results especially in data





that allow describing certain developments in time, as development of democracy based on information in editorials in newspapers. It is also possible to describe differences between groups, view of women versus men on certain issues.

The workshop is meant as an introduction into quantitative text analysis. In there several qualitative elements will receive attention.

Generally social scientists use for their data collection some kind of survey research. Here the data are collected from individual respondents. However, there are several other methods to collect data. Text analysis is one of these. In such studies data might be sampled from individual respondents (via diaries, letters, etc.), but more often the data come from articles in newspapers, documents from a government, minutes of meetings, discussion groups, and so on. These articles usually are assumed to represent the opinions that are hold by a broad public. At the end of the workshop the participants should have an idea of how to perform a simple text analysis study and must be able to evaluate more complex studies.

In the workshop a short overview is given of the actual state of the art of text analysis mainly within research in sociology, political science, and communication studies. This overview is based on the literature. After being confronted with the traditional content analysis or instrumental thematic text analysis where concepts or themes are considered from the perspective of the investigator (Holsti, Krippendorff, and Weber) the student will learn about the representational form of analysis, where the concepts are considered from the point of view of the sender of the message. The instrumental and representational approach are not only applied to the thematic analysis (with focus on frequency of occurrence and co-occurrences of concepts or themes), but also to the semantic (having attention for relations between concepts: subject – verb — object) and network (text is transposed into networks) analyses. Here ambiguity as found in texts becomes relevant. Finally attention will be given to reliability and validity and some new developments will be mentioned as well as ways to find texts. The participants will “work” on some research questions (differences between leaders based on speeches), in doing so two different computer programs will be used.

V. WORKSHOP: INTRODUCTION TO SOCIAL NETWORK ANALYSIS

Instructor: Michał Bojanowski (University of Warsaw)

Prerequisites: Taking advantage of all the features of the course will require knowledge of R on the level of "Going deeper into R...". R-unrelated topics should be accessible to everybody.

Social Network Analysis (SNA) is an approach to study groups of actors, be they individuals or organizations, through the analysis of relations between them that combine into complex networks. Relations linking actors can come in various forms, such as kinship, friendship, collaboration,





information seeking, authority, but also co-membership in associations or having something else "in common". General goal of SNA is to understand the functioning of a group of actors by studying the pattern in which they are connected to one another.

The workshop will introduce SNA and demonstrate various basic techniques using R. In particular, the following topics will be covered:

1. Types of research problems addressable with SNA
2. Social network data collection techniques
3. Network data representations
4. Overview of various descriptive measures characterizing (a) actor's position in a network, (b) groups of actors, and (c) network as a whole.

Items (3) and (4) will be illustrated using R.

VI. WORKSHOP: INTRODUCTION TO THE STATISTICAL ANALYSIS OF SOCIAL NETWORKS

Instructors: Christoph Stadtfeld (University of Groningen), Per Block (University of Oxford), Zsófia Boda (University of Oxford), András Vörös (University of Oxford)

Prerequisites: Participation in the course "Introduction to Social Network Analysis" in week 1 or basic knowledge in social network analysis and R. A computer with the latest version of R and the software packages igraph, Statnet and RSiena

Social network studies can be designed in many different ways and so the number of social network analysis methods is immense. In this course, participants will get an overview of state-of-the-art methods for the statistical analysis of social networks and acquire practical skills in two well-established methods for the analysis of complete social network data (as opposed to ego-centered networks). First, exponential random graph models (ERGMs) and their estimation with the software package Statnet will be introduced. ERGMs can be used to analyze cross-sectional network data that was collected at one point in time. Second, stochastic actor-oriented models (SAOMs) will be introduced and how they can be estimated with the software package RSiena. SAOMs can be used to model longitudinal network data that was collected repeatedly over time. A number of social network research questions will be formulated and tested empirically on data sets from social network school studies. Exemplary research questions are: Do friends of friends tend to be friends (transitivity)? Do school children rather have friendship relations with other children of the same gender (homophily)? Do those with many friends attract even more friends over time (preferential attachment)? Students can also bring and work on their own data sets.



VII. WORKSHOP: INTRODUCTION TO THE STATISTICAL ANALYSIS OF DYNAMIC SOCIAL NETWORKS

Instructors: Christoph Stadtfeld (University of Groningen), Per Block (University of Oxford), Zsófia Boda (University of Oxford), András Vörös (University of Oxford)

Prerequisites: Participation in the course “Introduction to Social Network Analysis in week 1” or basic knowledge in social network analysis and R. Participation in the course “Introduction to the Statistical Analysis of Social Networks” A computer with the latest version of R and the software package igraph and RSiena

In this course, participants will deepen their understanding of stochastic actor-oriented models (SAOMs) and the RSiena software. The course builds upon the course “Introduction to the Statistical Analysis of Social Networks”. SAOMs can be used to model longitudinal network data that was collected repeatedly over time (longitudinal social network data). RSiena is a program for the statistical analysis of network data, with the focus on social networks. Networks here are understood as entire (complete) networks, not as personal (ego-centered) networks: it is assumed that a set of nodes (social actors) is given, and all ties (links) between these nodes are known — except perhaps for a moderate amount of missing data. In particular, the course focuses on the evolution of social networks, the co-evolution of social networks and individual behavior variables, the co-evolution of multiple one-mode networks and the co-evolution of one-mode and two-mode networks. All these topics will be illustrated in practicals that make use of empirical social network data. Examples of these applied topics are ethnic homophily in schools, the co-evolution of smoking and friendship and the co-evolution of gossip and friendship. Students can also bring and work on their own data sets.

VIII. WORKSHOP: DATA VISUALIZATION TECHNIQUES. THEORY AND PRACTICE

Instructor: Przemysław Biecek (University of Warsaw)

Prerequisites: Basic knowledge about R (it is enough to know how to read data and plot anything). Basic knowledge about statistics (what is mean/median/quartile, fractions, linear regression)

The goal is to improve participants’ data visualization skills.

During lectures following topics will be discussed: major points in history of statistical graphic, how we perceive elements of the chart, how different data characteristics can be presented graphically and why some ways are better than others, why bad charts are bad, how to create a chart in R with ggplot2 package.





During practice sessions participants will work in groups. Each group will work on a data visualization of some relations between variables from an international survey (most likely from PIAAC study). Groups will start their projects with paper and pencil in order to create a prototype / brainstorm ideas. Then they will work with R to create a data based version of proposed prototypes. And finally they will tune the visualization in Inkscape.

IX. WORKSHOP: DATA PROCESSING WITH SPSS

Instructors: Tomasz Jerzyński (University of Warsaw), Marcin Zieliński (University of Warsaw)

Prerequisites: none

The main objective of the course is learning to work effectively with data sets with SPSS. The course covers the basic issues related to the operations on the data files, data processing, diagnostic and description. The idea is that each WSAD participant will be able to concentrate on the undertaken problem without thinking about “how it was done”.

The basic criterion for effective working with SPSS is using syntax.

Saving steps of analysis (syntax commands) aligns the structure of our work. Syntax helps the description and interpretation of results.

Editing text commands allows to quickly define analyzes of a large number of variables and define fast and reproducible analyzes of high complexity.

Topics:

1. Using Help F1
2. File operation and working directory.
 - Data files. Import external data. Selections and filtering. Merging.
 - Data-set description.
 - Output files. Export to html, txt, doc, xls.
 - Syntax files. Pasting. Running. Auto saving commands.
3. Data description. Viewing of meta-data. Missing values. Display command (variable label list). Displaying labels or/and variable names or values in outputs. Descriptive statistics. Summaries. Examines.
4. Tables. Distributions. Cross-tabulations. Special tables.
5. Recode and compute commands. Transformation functions. Quantiles.
6. Basic analysis. Chi square. Conditional means. Analysis of variance. Regression.
7. Charts. SPSS native. Export of output values and using external graphics.





X. WORKSHOP: DISCOVERING PATTERNS IN YOUR DATA – GENERALIZED LINEAR MIXED MODELS AND OTHER TECHNIQUES WITH R

Instructor: Richard Gonzalez (University of Michigan)

Prerequisites: An introductory statistics course is a must. You should review the material on hypothesis testing, measures of location, measures of spread, and exploratory data analysis. All examples and homework will be prepared in R, so the willingness to learn R is necessary in order to complete the problem sets.

I will review the generalized linear mixed models. I will also review modern data analytic methods including classification and regression trees (CART), repeated measures, mixture models, dyadic analysis and visualization techniques. The statistics package R will be used throughout the course.

I present material in a way that gives new intuition about what these techniques do and how these techniques can be used in research settings — students who claim solid knowledge of ANOVA and/or regression still report learning quite a bit from the course.

Data analysis is an exploratory process where the end product is a description of what happened in the study. In this course we will learn simple procedures to uncover what the “data are telling us.” We will also review best practices for writing and documenting code, data management, version control and collaboration.

XI. WORKSHOP: WIGGLES AND CURVES: THE ART OF DATA EXPLORATION

Instructor: Warren Thorngate (Carleton University)

Science is art using numbers. Alas, much of the art has been lost in the social sciences to the mechanics of traditional statistical practices. By focusing on statistically significant differences among central tendencies of aggregates, these practices can obscure more subtle features of social science data that might offer useful insights about the people who generated the data. The purpose of this workshop is to learn some simple statistical techniques that explore data subtleties — techniques that deviate from traditional statistical practices in three ways. First, the techniques do not estimate how well samples of data might generalize to populations. Instead, the techniques indicate how well predictions generalize to samples of data. Second, the techniques do not require data to be aggregated across people before they are analysed. Instead, they allow data to be analysed individually before they are aggregated. Third, the indicators generated by these techniques are remarkably easy to calculate and to interpret.

People wishing to take this workshop should know enough about traditional statistical practices to be skeptical of them; a course or two in omni-mega-multivariate statistical analyses using commercial





software such as SPSS® will do. Participants will practice the alternative techniques using data from the European Social Survey and the International Social Survey.

XII. WORKSHOP: HOW TO WRITE CLEARLY

Instructor: Warren Thorngate (Carleton University)

Most social science literature is boring and obscure. It need not be. Most social science is fascinating. Sadly, however, most social scientists never learn to write about it well. The purpose of this workshop is to teach participants some of the techniques for writing clearly about social science. We will discuss the psychology of writing, and what distinguishes good writing from bad. We will examine published examples of good and bad writing, and show how to improve some of the bad stuff. Participants will also write short paragraphs in class, some of which will be shown anonymously to participants for class critique. By the end of the workshop, each participant will write a brief research report. People wishing to take this workshop should know enough English to write it badly or better. They should not be shy about making grammatical, stylistic or rhetorical errors. A good sense of humour would also be helpful.

XIII. WORKSHOP: STRUCTURAL EQUATION MODELS

Instructor: Roman Konarski (PBS, Gdańsk University)

Prerequisites: A course on intermediate statistics and/or practical experience with regression and factor analysis.

Structural equation model (SEM) represents a general statistical approach to the evaluation of theoretical models fit to empirical data. SEM with latent variables embodies simultaneous equations w multiple exogenous and endogenous variables (path models), along with measurement error models (confirmatory factor analysis). Thus SEM represents a synthesis of methods developed in econometrics and psychometrics.

This course introduces methods and applications of SEM with the *Mplus* software. The course will provide exposure to the fundamentals and extensions of SEM, with an emphasis on applications in applied research settings. The course covers such topics as data requirements, *Mplus* syntax, specification and identification of models with observed and latent variables, multigroup models, models for longitudinal and clustered data, and model estimation, testing and reporting. The course will be of interest to those who plan to utilize the general structural equation model with latent variables or its specializations. Familiarity with elementary matrix algebra will be useful, though not essential, for understanding *Mplus* syntax.

