The world today generates a huge amount of data. Turning these data into meaningful and useful knowledge requires specialized tools and skills. On this course you will be introduced to the open-source software R, a leading data science tool, and learn to use it to turn data into applicable knowledge.
R has grown rapidly to become one of the most widely-used programming languages for statistical analyses in the social and economic data sciences. This is primarily due to its incredible versatility and prepackaged tools, which are at the cutting edge of data analysis technology. On top of that, R is also an ideal environment for socio-economic research, offering many powerful packages tailored to analyzing, visualizing, and documenting surveys and other types of data common in the social and economic sciences. This course provides a thorough introduction to R for modern social data science.

WHAT YOU WILL LEARN
You will learn to implement your entire research workflow from data input to results in RStudio. This includes learning to use R to:

- Import data from multiple sources (e.g., txt-files, Excel, Stata, or SPSS).
- Find and access socio-economic data online through APIs and web scraping (e.g. Eurostat, OECD, or Danmarks Statistik).
- Clean, prepare, and visualize data using tidyverse packages.
- Perform statistical analyses typically used in socio-economic research (multiple linear and logistic regression, principal components/factor analysis).
- Visualize and present the results of these statistical analyses.
- Integrate this entire workflow into a final report or presentation in a single dynamic document by using R Markdown.

In addition to the above-mentioned R packages, we will also introduce you to the RStudio environment and a number of R packages that are particularly useful for socio-economic research, such as estimatr (for robust statistical inference), ggplot2 (for visualization), psych (for factor analysis), and other API packages.

PARTICIPANTS
While R is a highly versatile programming language, this course takes a pragmatic approach by teaching participants to use existing packages to conduct sophisticated socio-economic research. This means that participants need no prior programming experience to attend the course. Rather, the course is tailored for participants who have some experience with socio-economic research and want to learn how to leverage R to streamline their workflow and implement cutting-edge methods.

The course is particularly suited for:

- Individuals with experience in working with Excel, Stata or SPSS who would like to switch to R
- Individuals with some R experience and an interest in learning more about RStudio, R Markdown, and the tidyverse

R (www.r-project.org) and RStudio (www.rstudio.com) are open source and available free of charge. Participants are expected to bring a laptop and will be instructed in how to install R.

COURSE DATES
17-21 August at the University of Copenhagen, South Campus

COURSE DIRECTORS
Friedolin Merhout, Assistant Professor, Department of Sociology, University of Copenhagen
Merlin Schaeffer, Associate Professor, Department of Sociology, University of Copenhagen

For more information and registration
TEACHING MATERIAL
This course is based on the book “R for Data Science” (2016) by Garrett Grolemund and Hadley Wickham which participants can access for free by visiting this website: https://r4ds.had.co.nz/

COURSE FEE
EUR 2,680 / DKK 19,900 excl. Danish VAT. Fee includes teaching, course materials and all meals during the course.

”I have over the years participated in numerous courses at different universities and venues. The service, faculty, catering and facilities at “Copenhagen Summer University” stand out as the best of them all.”

Kreesten Madsen, Senior Director, LEO Pharma, participant about Copenhagen Summer University 2019

For more information and registration
W W COPENHAGENSUMMERUNIVERSITY.DK  csu@adm.ku.dk  +45 35 33 34 23