Basic Statistics and Reporting in R with RStudio	SUMMER SCHOOL SPB_/L	The 4 <sup>th</sup> International IPSA – HSE Summer School for Methods of Political & Social Research  Course Syllabus
Instructor:	Course title:	Basic Statistics and Reporting in R with RStudio
### Strief course description (up to 100 brooks):  ### This course offers an introduction to popular statistical tests and project reporting in R and R Markdown. Participants will learn the techniques of data wrangling, parametric and non-parametric tests and linear regression in R. In addition, the course will introduce creating interactive completed a full-cycle analysis in R, from reading in the data, performing basic preprocessing, to performing statistical tests and reporting the results. Previous knowledge of statistical tests and reporting the results. Previous knowledge of statistical tests and any statistical software is not required but will be an asset.  #### Variable measurement data wrangling means comparison test of independence analysis of variance linear regression regression diagnostics assumptions testing reporting data visualization  ### Workshops' overview:  ### Workshops' overview:  ### Day 1 Variable measurement scales. Data classes in R. R and RStudio environment.  ### Day 2 Data wrangling: renaming, merging, splitting variables, subsetting a dataset. Drawing basic plots.  ### Day 3 Chi-square test of independence. Comparing group means.  ### Day 4 Linear regression. Diagnostics of linear regression.  ### Day 5 Reporting with R Markdown. Interactive reports and graphics.  ### At the online exam, participants will solve a data problem and prepare a report of it in R. The data problem will require using one of the tests covered by the course. To prepare for the exam, participants will get a homework task after each of the first four days, which will be discussed in the morning session the day after.  #### Course Requirements / Prior knowledge of basic statistical tests: Diez, D.M., Cetinkaya-Rundel, M., & C.D. Barr (2020) OpenIntro Statistics, 4th ed. URL: https://www.openintro.org/book/os/.	Instructor:	
This course offers an introduction to popular statistical tests and project reporting in R and R Markdown. Participants will learn the techniques of data wrangling; parametric and non-parametric tests and linear regression in R. In addition, the course will introduce creating interactive reports in RStudio. By the end of this course, participants will have completed a full-cycle analysis in R, from reading in the data, performing basic preprocessing, to performing statistical tests and reporting the results. Previous knowledge of statistical tests and any statistical software is not required but will be an asset.  Indicative concepts (up to 10):  Indicative concepts (up to 10):  Workshops' overview:  Variable measurement data wrangling measurement scales. Data classes in R. R and RStudio environment.  Day 1 Variable measurement scales. Data classes in R. R and RStudio environment.  Day 2 Data wrangling: renaming, merging, splitting variables, subsetting a dataset. Drawing basic plots.  Day 3 Chi-square test of independence. Comparing group means.  Day 4 Linear regression. Diagnostics of linear regression.  Day 5 Reporting with R Markdown. Interactive reports and graphics.  At the online exam, participants will solve a data problem and prepare a report of it in R. The data problem will require using one of the tests covered by the course. To prepare for the exam, participants will get a homework task after each of the first four days, which will be discussed in the morning session the day after.  Course Requirements / Prior knowledge of basic statistical testing). No prior experience with Stata, SAS, IBM SPSS or other statistical software will be an asset.  To revise the knowledge of statistical tests: Diez, D.M., Cetinkaya-Rundel, M., & C.D. Barr (2020) OpenIntro Statistics, 4th ed. URL: https://www.openintro.org/book/os/.	ECTS / academic hours	
reporting in R and R Markdown. Participants will learn the techniques of data wrangling, parametric and non-parametric tests and linear regression in R. In addition, the course will introduce creating interactive reports in RStudio. By the end of this course, participants will have completed a full-cycle analysis in R, from reading in the data, performing basic preprocessing, to performing statistical tests and reporting the results. Previous knowledge of statistical tests and any statistical software is not required but will be an asset.  Indicative concepts (up to 10):  Indicative concepts (up to 10):  Variable measurement data wrangling means comparison test of independence analysis of variance linear regression regression diagnostics assumptions testing reporting data visualization  Variable measurement scales. Data classes in R. R and RStudio environment.  Day 1 Variable measurement scales. Data classes in R. R and RStudio environment.  Day 2 Data wrangling: renaming, merging, splitting variables, subsetting a dataset. Drawing basic plots.  Day 3 Chi-square test of independence. Comparing group means.  Day 4 Linear regression. Diagnostics of linear regression.  Day 5 Reporting with R Markdown. Interactive reports and graphics.  Assessment techiques to receive graded certificate:  Prevent of it in R. The data problem will require using one of the tests covered by the course. To prepare for the exam, participants will get a homework task after each of the first four days, which will be discussed in the morning session the day after.  Course Requirements / Prior knowledge of basic statistical testing).  No prior experience with R or RStudio is required. Previous experience with Stata, SAS, IBM SPSS or other statistical software will be an asset.  To revise the knowledge of statistical tests: Diez, D.M., Cetinkaya-Rundel, M., & C.D. Barr (2020) Openintro Statistics, 4th ed. URL: https://www.openintro.org/book/os/.		·
data wrangling means comparison test of independence analysis of variance linear regression regression diagnostics assumptions testing reporting data visualization  Day 1 Variable measurement scales. Data classes in R. R and RStudio environment.  Day 2 Data wrangling: renaming, merging, splitting variables, subsetting a dataset. Drawing basic plots.  Day 3 Chi-square test of independence. Comparing group means. Day 4 Linear regression. Diagnostics of linear regression. Day 5 Reporting with R Markdown. Interactive reports and graphics.  Assessment techiques to receive graded certificate:  At the online exam, participants will solve a data problem and prepare a report of it in R. The data problem will require using one of the tests covered by the course. To prepare for the exam, participants will get a homework task after each of the first four days, which will be discussed in the morning session the day after.  Course Requirements / Prior knowledge of basic statistics is required. Previous experience with Stata, SAS, IBM SPSS or other statistical software will be an asset.  Essential readings:  (Reading list will be supplemented with other sources and distributed	• • •	reporting in R and R Markdown. Participants will learn the techniques of data wrangling, parametric and non-parametric tests and linear regression in R. In addition, the course will introduce creating interactive reports in RStudio. By the end of this course, participants will have completed a full-cycle analysis in R, from reading in the data, performing basic preprocessing, to performing statistical tests and reporting the results. Previous knowledge of statistical tests and any statistical software is not
analysis of variance linear regression regression diagnostics assumptions testing reporting data visualization  Workshops' overview:  Day 1 Variable measurement scales. Data classes in R. R and RStudio environment.  Day 2 Data wrangling: renaming, merging, splitting variables, subsetting a dataset. Drawing basic plots.  Day 3 Chi-square test of independence. Comparing group means.  Day 4 Linear regression. Diagnostics of linear regression.  Day 5 Reporting with R Markdown. Interactive reports and graphics.  Assessment techiques to receive graded certificate:  At the online exam, participants will solve a data problem and prepare a report of it in R. The data problem will require using one of the tests covered by the course. To prepare for the exam, participants will get a homework task after each of the first four days, which will be discussed in the morning session the day after.  Course Requirements / Prior knowledge of basic statistics is required (sampling, central limit theorem, null hypothesis statistical testing).  No prior experience with R or RStudio is required. Previous experience with Stata, SAS, IBM SPSS or other statistical software will be an asset.  Essential readings:  (Reading list will be supplemented with other sources and distributed	Indicative concepts (up to 10):	data wrangling
regression diagnostics assumptions testing reporting data visualization  Day 1 Variable measurement scales. Data classes in R. R and RStudio environment.  Day 2 Data wrangling: renaming, merging, splitting variables, subsetting a dataset. Drawing basic plots.  Day 3 Chi-square test of independence. Comparing group means.  Day 4 Linear regression. Diagnostics of linear regression.  Day 5 Reporting with R Markdown. Interactive reports and graphics.  At the online exam, participants will solve a data problem and prepare a report of it in R. The data problem will require using one of the tests covered by the course. To prepare for the exam, participants will get a homework task after each of the first four days, which will be discussed in the morning session the day after.  Course Requirements / Prerequisites  Course Requirements / Prerequisites  Course Requirements / Prerequisites  To revise the knowledge of statistical testing). No prior experience with R or RStudio is required. Previous experience with Stata, SAS, IBM SPSS or other statistical software will be an asset.  To revise the knowledge of statistical tests: Diez, D.M., Cetinkaya-Rundel, M., & C.D. Barr (2020) OpenIntro Statistics, 4th ed. URL: https://www.openintro.org/book/os/.		analysis of variance
assumptions testing reporting data visualization  Workshops' overview:  Day 1 Variable measurement scales. Data classes in R. R and RStudio environment.  Day 2 Data wrangling: renaming, merging, splitting variables, subsetting a dataset. Drawing basic plots.  Day 3 Chi-square test of independence. Comparing group means.  Day 4 Linear regression. Diagnostics of linear regression.  Day 5 Reporting with R Markdown. Interactive reports and graphics.  At the online exam, participants will solve a data problem and prepare a report of it in R. The data problem will require using one of the tests covered by the course. To prepare for the exam, participants will get a homework task after each of the first four days, which will be discussed in the morning session the day after.  Course Requirements /  Prerequisites  Course Requirements /  Prior knowledge of basic statistics is required (sampling, central limit theorem, null hypothesis statistical testing).  No prior experience with R or RStudio is required. Previous experience with Stata, SAS, IBM SPSS or other statistical software will be an asset.  Essential readings:  (Reading list will be supplemented with other sources and distributed  Assumptions  Day 1 Variable measurement scales. Data classes in R. R and RStudio environment.  Day 2 Data wrangling: renaming, merging, splitting variables, and RStudios, all string variables, subsect plots.  Previous experience with R or RStudio is required. Previous experience with Stata, SAS, IBM SPSS or other statistical software will be an asset.  To revise the knowledge of statistical tests: Diez, D.M., Cetinkaya-Rundel, M., & C.D. Barr (2020) OpenIntro Statistics, 4th ed. URL:  https://www.openintro.org/book/os/.		
reporting data visualization    Day 1		
Day 1   Variable measurement scales. Data classes in R. R and RStudio environment.		· · ·
Workshops' overview:  Day 1		1
Workshops' overview:    Day 2		
Day 2 Data wrangling: renaming, merging, splitting variables, subsetting a dataset. Drawing basic plots.  Day 3 Chi-square test of independence. Comparing group means.  Day 4 Linear regression. Diagnostics of linear regression.  Day 5 Reporting with R Markdown. Interactive reports and graphics.  At the online exam, participants will solve a data problem and prepare a report of it in R. The data problem will require using one of the tests covered by the course. To prepare for the exam, participants will get a homework task after each of the first four days, which will be discussed in the morning session the day after.  Course Requirements / Prior knowledge of basic statistics is required (sampling, central limit theorem, null hypothesis statistical testing).  No prior experience with R or RStudio is required. Previous experience with Stata, SAS, IBM SPSS or other statistical software will be an asset.  Essential readings: (Reading list will be supplemented with other sources and distributed https://www.openintro.org/book/os/.	Workshops' overview:	'
Day 3 Chi-square test of independence. Comparing group means.  Day 4 Linear regression. Diagnostics of linear regression.  Day 5 Reporting with R Markdown. Interactive reports and graphics.  At the online exam, participants will solve a data problem and prepare a report of it in R. The data problem will require using one of the tests covered by the course. To prepare for the exam, participants will get a homework task after each of the first four days, which will be discussed in the morning session the day after.  Course Requirements / Prior knowledge of basic statistics is required (sampling, central limit theorem, null hypothesis statistical testing).  No prior experience with R or RStudio is required. Previous experience with Stata, SAS, IBM SPSS or other statistical software will be an asset.  Essential readings: (Reading list will be supplemented with other sources and distributed  Day 4 Linear regression. Diagnostics of linear regression.  Day 5 Reporting with R Markdown. Interactive reports and graphics.  At the online exam, participants will solve a data problem and prepare a report of it in R. The data problem will require using one of the tests covered by the course. To prepare for the exam, participants will get a homework task after each of the first four days, which will be discussed in the morning session the day after.  Prior knowledge of basic statistics is required (sampling, central limit theorem, null hypothesis statistical testing).  No prior experience with R or RStudio is required. Previous experience with Stata, SAS, IBM SPSS or other statistical software will be an asset.  To revise the knowledge of statistical tests: Diez, D.M., Cetinkaya-Rundel, M., & C.D. Barr (2020) OpenIntro Statistics, 4th ed. URL: https://www.openintro.org/book/os/.		
Day 4 Linear regression. Diagnostics of linear regression.  Day 5 Reporting with R Markdown. Interactive reports and graphics.  At the online exam, participants will solve a data problem and prepare a report of it in R. The data problem will require using one of the tests covered by the course. To prepare for the exam, participants will get a homework task after each of the first four days, which will be discussed in the morning session the day after.  Course Requirements / Prior knowledge of basic statistics is required (sampling, central limit theorem, null hypothesis statistical testing).  No prior experience with R or RStudio is required. Previous experience with Stata, SAS, IBM SPSS or other statistical software will be an asset.  Essential readings: (Reading list will be supplemented with other sources and distributed  To revise the knowledge of statistics, 4th ed. URL: https://www.openintro.org/book/os/.		subsetting a dataset. Drawing basic plots.
Assessment techiques to receive graded certificate:  At the online exam, participants will solve a data problem and prepare a report of it in R. The data problem will require using one of the tests covered by the course. To prepare for the exam, participants will get a homework task after each of the first four days, which will be discussed in the morning session the day after.  Course Requirements / Prior knowledge of basic statistics is required (sampling, central limit theorem, null hypothesis statistical testing). No prior experience with R or RStudio is required. Previous experience with Stata, SAS, IBM SPSS or other statistical software will be an asset.  Essential readings: (Reading list will be supplemented with other sources and distributed  Day 5  Reporting with R Markdown. Interactive reports and graphics.  At the online exam, participants will solve a data problem and prepare a report of it in R. The data problem will require using one of the tests covered by the course. To prepare for the exam, participants will get a homework task after each of the first four days, which will be discussed in the morning session the day after.  Prior knowledge of basic statistics is required (sampling, central limit theorem, null hypothesis statistical testing).  No prior experience with R or RStudio is required. Previous experience with Stata, SAS, IBM SPSS or other statistical software will be an asset.  To revise the knowledge of statistical tests: Diez, D.M., Cetinkaya-Rundel, M., & C.D. Barr (2020) OpenIntro Statistics, 4th ed. URL: https://www.openintro.org/book/os/.		Day 3 Chi-square test of independence. Comparing group means.
Assessment techiques to receive graded certificate:  At the online exam, participants will solve a data problem and prepare a report of it in R. The data problem will require using one of the tests covered by the course. To prepare for the exam, participants will get a homework task after each of the first four days, which will be discussed in the morning session the day after.  Course Requirements / Prior knowledge of basic statistics is required (sampling, central limit theorem, null hypothesis statistical testing).  No prior experience with R or RStudio is required. Previous experience with Stata, SAS, IBM SPSS or other statistical software will be an asset.  Essential readings: (Reading list will be supplemented with other sources and distributed  At the online exam, participants will solve a data problem and prepare a report of it in R. The data problem will require using one of the tests covered by the course. To prepare for the exam, participants will get a homework task after each of the first four days, which will be discussed in the morning session the day after.  Prior knowledge of basic statistics is required (sampling, central limit theorem, null hypothesis statistical testing).  No prior experience with R or RStudio is required. Previous experience with Stata, SAS, IBM SPSS or other statistical software will be an asset.  To revise the knowledge of statistical tests: Diez, D.M., Cetinkaya-Rundel, M., & C.D. Barr (2020) OpenIntro Statistics, 4th ed. URL: https://www.openintro.org/book/os/.		Day 4 Linear regression. Diagnostics of linear regression.
report of it in R. The data problem will require using one of the tests covered by the course. To prepare for the exam, participants will get a homework task after each of the first four days, which will be discussed in the morning session the day after.  Course Requirements / Prior knowledge of basic statistics is required (sampling, central limit theorem, null hypothesis statistical testing). No prior experience with R or RStudio is required. Previous experience with Stata, SAS, IBM SPSS or other statistical software will be an asset.  Essential readings: (Reading list will be supplemented with other sources and distributed report of it in R. The data problem will require using one of the tests covered by the course. To prepare for the exam, participants will get a homework task after each of the exam, participants will get a homework task after each of the exam, participants will get a homework task after each of the exam, participants will get a homework task after each of the exam, participants will get a homework task after each of the exam, participants will get a homework task after each of the exam, participants will get a homework task after each of the exam, participants will get a homework task after each of the exam, participants will get a homework task after each of the exam, participants will get a homework task after each of the exam, participants will get a homework task after each of the exam, participants will get a homework task after each of the exam, participants will get a homework task after each of the exam, participants will get a homework task after each of the first four days, which will get a homework task after each of the first four days, which will get a homework task after each of the first four days, which will get a homework task after each of the first four days, which will be discussed in the morning session the day after.		Day 5 Reporting with R Markdown. Interactive reports and graphics.
covered by the course. To prepare for the exam, participants will get a homework task after each of the first four days, which will be discussed in the morning session the day after.  Course Requirements / Prior knowledge of basic statistics is required (sampling, central limit theorem, null hypothesis statistical testing). No prior experience with R or RStudio is required. Previous experience with Stata, SAS, IBM SPSS or other statistical software will be an asset.  Essential readings: (Reading list will be supplemented with other sources and distributed  Course Requirements / Prior knowledge of basic statistics is required (sampling, central limit theorem, null hypothesis statistical testing). No prior experience with R or RStudio is required. Previous experience with Stata, SAS, IBM SPSS or other statistical software will be an asset.  To revise the knowledge of statistical tests: Diez, D.M., Cetinkaya-Rundel, M., & C.D. Barr (2020) OpenIntro Statistics, 4th ed. URL: https://www.openintro.org/book/os/.	Assessment techiques to receive	At the online exam, participants will solve a data problem and prepare a
theorem, null hypothesis statistical testing).  No prior experience with R or RStudio is required. Previous experience with Stata, SAS, IBM SPSS or other statistical software will be an asset.  Essential readings:  (Reading list will be supplemented with other sources and distributed  To revise the knowledge of statistical tests: Diez, D.M., Cetinkaya-Rundel, M., & C.D. Barr (2020) OpenIntro Statistics, 4th ed. URL: https://www.openintro.org/book/os/.	graded certificate:	covered by the course. To prepare for the exam, participants will get a homework task after each of the first four days, which will be discussed in
No prior experience with R or RStudio is required. Previous experience with Stata, SAS, IBM SPSS or other statistical software will be an asset.  Essential readings: (Reading list will be supplemented with other sources and distributed  No prior experience with R or RStudio is required. Previous experience with Stata, SAS, IBM SPSS or other statistical software will be an asset.  To revise the knowledge of statistical tests: Diez, D.M., Cetinkaya-Rundel, M., & C.D. Barr (2020) OpenIntro Statistics, 4th ed. URL: https://www.openintro.org/book/os/.	•	, , , ,
with Stata, SAS, IBM SPSS or other statistical software will be an asset.  Essential readings: (Reading list will be supplemented with other sources and distributed with other sources and distributed with stata, SAS, IBM SPSS or other statistical software will be an asset.  To revise the knowledge of statistical tests: Diez, D.M., Cetinkaya-Rundel, M., & C.D. Barr (2020) OpenIntro Statistics, 4th ed. URL: https://www.openintro.org/book/os/.	Prerequisites	
<b>Essential readings:</b> (Reading list will be supplemented with other sources and distributed  To revise the knowledge of statistical tests: Diez, D.M., Cetinkaya-Rundel, M., & C.D. Barr (2020) OpenIntro Statistics, 4th ed. URL: https://www.openintro.org/book/os/.		, , ,
(Reading list will be supplemented with other sources and distributed M., & C.D. Barr (2020) OpenIntro Statistics, 4th ed. URL: https://www.openintro.org/book/os/.		
with other sources and distributed https://www.openintro.org/book/os/.	_	
during Day 1)  It is recommended to gain a general understanding of the R software and its working principles with any educational videos or textbooks released	during Day 1)	
2013 and later which are available at the library. For instance: Wright, C.,		
et al. (2021) Tidyverse Skills for Data Science in R		
https://leanpub.com/tidyverseskillsdatascience		
Contacts: ashirokanova@hse.ru	Contacts:	