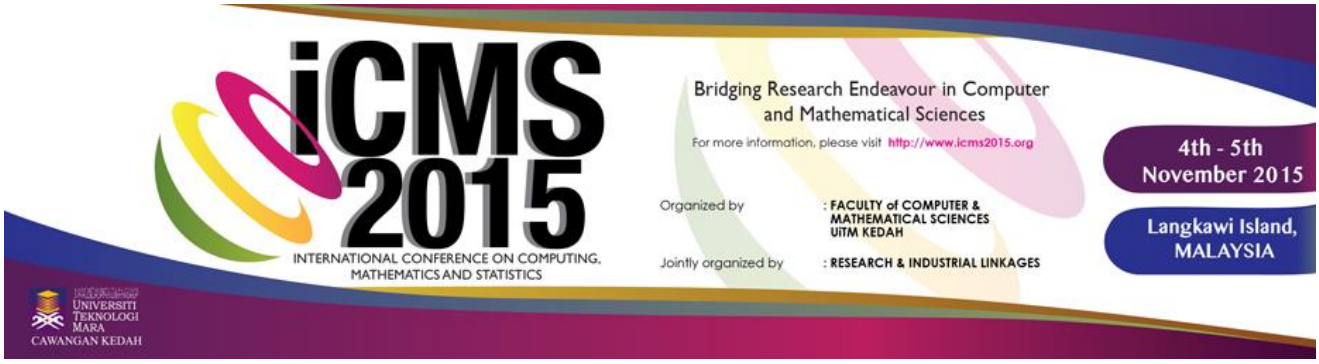

PRE-CONFERENCE WORKSHOP
"Introduction to R and Data Visualization"

Date & Location

Date : 2 - 3 November 2015
Time : 9 am - 5 pm
Venue : Melur 1, Langkawi Lagoon, Langkawi Island, MALAYSIA

Detailed Course Outline

1. Introduction to R Statistical Software 6
1.0 The beginning of R 6
1.0.1 What is R? 6
1.0.2 From where R comes? 6
1.0.3 Why R? 6
1.0.4 R – a top language 6
1.0.5 R's features 6
1.0.6 Who are the R users? 6
1.0.7 Starting learning R? 6
1.0.8 R-ide the Wave 6
1.0.9 What you really want to do? 6
1.1 R - Introducing the R Console 6
1.1.1 R Quick installation 6
1.1.2 R Console 6
1.1.3 Working directory 7
1.1.4 Always useful… 7
1.1.5 Sometimes useful… 7
1.1.6 Environment…view, load & save 7
1.1.7 Environment… customize 7
1.2 R - Installation, Packages, CRAN, Components 7
1.2.1 R - Environment 7
1.2.2 R - Packages 7
1.2.3 R - Contributed packages 7.168+ 7

1.2.4 R - Most commonly used packages 7
1.2.5 R – managing the package 7
1.2.6 R – CRAN 7
1.2.7 R – CRAN Task Views 7
1.3 Graphical User Interfaces: R Console, R Studio, R Commander, R resources and online community 7
1.3.1 GUI: R Console, R Studio, R Commander 8
1.3.2 R Studio – features 1 8
1.3.3 R Studio – features 2 8
1.3.4 R Studio – Installation 8
1.3.5 R Studio – Version 8
1.3.6 R Studio – IDE 8
1.3.7 R Commander – Installation 8
1.3.8 R Commander – other 8
1.3.9 R resources 8
1.3.10 R online community 8
1.3.11 R – Who can we ask? 8
1.3.12 R – But before asking… 8
1.3.13 R – What to keep in mind when asked a question online community? 8
1.3.14 R code style guide 8
1.3.15 R - What are the limits? 8
1.3.16 Your imagination! 9
1.3.17 Millions of data! 9
1.3.18 R lives! 9
2. Databases 9
2.1 Data manipulation 9
2.1.1 Operators & functions 9
2.1.2 Operator syntax 9
2.1.3 Functions 9
2.1.4 +Inf, -Inf, NaN 9
2.1.5 R objects 9
2.1.6 R - factor object 9
2.1.7 Data structures 9
2.1.8 Data structures – atomic vector 9
2.1.9 Data structures – matrix 9
2.1.10 Data structures – array 10
2.1.11 Data structures – data frame 10
2.1.12 Data structures – list 10
2.1.13 Data structures – table 10
2.1.14 R – object attributes 10
2.1.15 Viewing data series available in packages 10
2.1.16 Data selection and manipulation I 10
2.1.17 Data selection and manipulation II 10
2.1.18 Data reshaping 10
2.1.19 Online resources 10
2.1.20 R – Terminology (R vs. RDBMS*) 10
2.1.21 R – data manipulation 10

2.1.22 selection 10
2.1.23 sorting 10
2.1.24 concatenation / merging data frames 10
2.1.25 finding and removing duplicate records 11
2.1.26 levels identification 11
2.1.27 renaming levels 11
2.1.28 changing the order of levels of a factor 11
2.1.29 adding and removing variables (columns) from a data frame 11
2.1.30 reordering the variables in a data frame 11
2.1.31 renaming of variables 11
2.1.32 R – How we work with databases? 11
2.1.33 Packages…Packages… R ←→ RDBMS 11
2.1.34 Connection via ODBC 11
2.1.35 Windows – ODBC Data Source Administrator 11
2.1.36 RODBC - connection and metadata functions 11
2.1.37 RODBC - database and table metadata functions 11
2.1.38 RODBC – R code 11
2.1.39 RODBC - direct connection method 11
2.1.40 RODBC - querying and deleting data 12
2.1.41 RODBC - diagnosing errors 12
2.1.42 RODBC - saving data 12
2.1.43 RODBC - stored procedures 12
2.2 Queries 12
2.2.1 Packages… 12
2.2.2 Opening an existing dataset 12
2.2.3 sqldf Package 12
2.2.4 sqldf Package - histogram 12
2.2.5 sqldf Package – group by 12
2.3 Using SQL within R 12
2.3.1 Column & Row names 12
2.3.2 Subsetting columns 12
2.3.3 All columns 12
2.3.4 Only one column 12
2.3.5 Case sensitivity 13
2.3.6 R extensions 13
2.3.7 Other column selections 13
2.3.8 Subsetting rows - conditions 13
2.3.9 Subsetting rows - with 13
2.3.10 Logical operators 13
2.3.11 First few 13
2.3.12 Row names versus numbers 13
2.3.13 NULL, Not NULL, Is NULL 13
2.3.14 No missing values 13
2.3.15 Quotes 13
2.4 Data aggregation 13
2.4.1 transpose 13
2.4.2 aggregate 13

2.4.3 with vs. without aggregation 13
2.4.4 {reshape} package 14
2.5 Matching 14
2.5.1 match() 14
3. Data Visualization & Graphics Environments 14
3.1 Base graphics (Scatterplot, Box-and-whiskers plot, Histogram) 14
3.1.1 Exploratory graphs ~ final graphs 14
3.1.2 Base graphics 14
3.1.3 Density plots 14
3.1.4 Dot plots 14
3.1.5 Bar charts 14
3.1.6 Line charts 14
3.1.7 Pie charts 14
3.1.8 Boxplots 14
3.1.9 Scatter plots 14
3.2 Lattice 15
3.2.1 {lattice} package 15
3.2.2 Customizing Lattice Graphs 15
3.2.3 densityplot(), bwplot(), xyplot(), cloud(), dotplot(), splom() 15
3.3 ggplot2 15
3.3.1 qplot() 15
3.3.2 Customizing ggplot2 Graphs 15
3.3.3 Kernel density plots 15
3.3.4 Scatterplot 15
3.3.5 Regressions 15
3.3.6 Boxplot 15
3.4 Interactive graphics in R 15
3.5 Reproducibility 15
4. Regression Analysis with R 16
4.1. Why R for regression analysis? 16
4.2. Before you start a regression analysis 16
4.3. Regression analysis objectives 16
4.4. When to use Regression Analysis? 16
4.5. Types of regression analysis 16
4.6. Linear regression models 16
4.6.1. Estimating β – OLS Method 16
4.6.2. lm function 16
4.6.3. Application 1. (NRG_Data.csv) 16
4.6.3.1. Preliminary data visualization 16
4.6.3.2. Interpret the regression results 16
4.6.3.3. Extract elements from a linear model fit 16
4.6.3.4. Estimate parameters for the regression 16
4.6.3.5. Confidence interval for the estimated parameters? 17
4.6.3.6. Plot residuals. 17
4.6.3.7. Prediction 17
4.6.4. Application 2 (HBS_youth.csv) 17
4.7. Generalized Linear Models (GLM). Logistic Regression 17

4.7.1. Binomial Logistic Regression. Models 17
4.7.1.1. Odds Ratio 17
4.7.1.2. Maximum Likelihood Estimation 17
4.7.1.3. Logistic regression implementation in R 17
4.7.2. Application 3. (EUSILC.csv) - fitting a binary logistic regression model 17
4.7.2.1. Predict the probability of poor vs. rich people, as a function of some characteristics of population 17
4.7.2.2. Interpreting the results of logistic regression model 17
4.7.2.3. Predicted probabilities 17
4.7.3. Multinomial logistic regression 17
4.7.4. Application 4. (Census.csv) - fitting a binary logistic regression model 18
4.7.4.1. Predict the probability of high social class vs. other social categories, as a function of some characteristics of population 18
4.7.4.2. Interpreting multinomial regression results 18
4.7.4.3. Predicted probabilities 18

All the materials: slides, examples (R scripts), documents and papers will be shared to participants during the workshop.

Authors