EEE356 - Data Analytics (R)
Week 1: Course Introduction and Scope

ADANA ALPARSLAN TÜRKEŞ
SCIENCE AND TECHNOLOGY UNIVERSITY

Dr Kasım Zor

Department of Electrical and Electronic Engineering

Spring 2024

Outline

Outline

Course Instructor

Dr Kasım ZOR (Assistant Professor)
Electrical and Electronic Engineer, PhD

Research Interests

- Electric Load Forecasting, Energy Analytics and Informatics, Machine Learning, Distributed Generation, and Electrical Energy and Power Systems

Contact Information

- Office: Room 213, 2nd Floor, M1 Building
- E-mail: kzor@atu.edu.tr
- Web: www.kasimzor.com.tr

Laboratory Assistant

Mr. Ömer Can TOLUN (Research Assistant)
Electrical and Electronic Engineer, PhD Candidate

Research Interests

- Control and Command Systems, Robotics, Artificial Intelligence, and Autonomous Vehicles

Contact Information

- Office: Control Laboratory, M3 Building
- E-mail: octolun@atu.edu.tr

Course Information

Course Title	Code	Semester	T+L (Hours)	Credits	ECTS
Data Analytics	EEE356	6	3+2	4	6

Table 1: Table of Course Information

- Prerequisites: None
- Language: English
- Level: Bachelor
- Type: Elective

Course Assessment and Evaluation

Assessment Type	Quantity	Weight
Applied Midterm Examination	1	40%
Applied Final Examination	1	60%

Table 2: Table of Course Assessment and Evaluation

	Course Type	Allowed Rate	Allowed Hours
Absentee Rate	Main Course	30%	14
	Laboratory	20%	6

Table 3: Table of Absentee Rate

Laboratory Schedule

	Lab Contents
W1	Introduction to the Laboratory
W2	Introduction to R Programming Language
W3	Basics of R Programming Language
W4	Data Structures
W5	Control Structures
W6	Functions
W7	Data Wrangling - Part 1
W8	Data Wrangling - Part 2
W10	Data Wrangling - Part 3
W11	Data Visualisation - Part 1
W12	Data Visualisation - Part 2
W13	Exploratory Data Analysis
W14	Approaches to Missing Data
W15	Interactive Lab for FAQ by Students

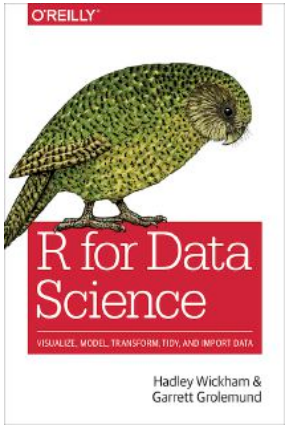
Objectives and Learning Outcomes

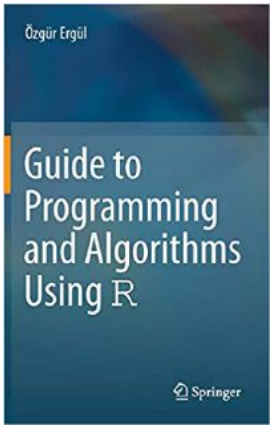
This course aims to gain students insight and required skills related to data analytics containing R Programming, data wrangling, data visualisation, exploratory data analysis, and approaches to missing data.

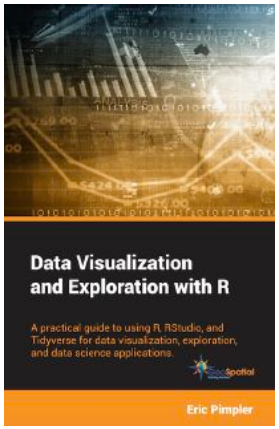
- Gaining insight about the term 'Data Analytics'
- Ability to use R programming language
- Possessing skills related to data analytics containing
  - Data Wrangling,
  - Data Visualisation,
  - Exploratory Data Analysis,
  - Approaches to Missing Data.

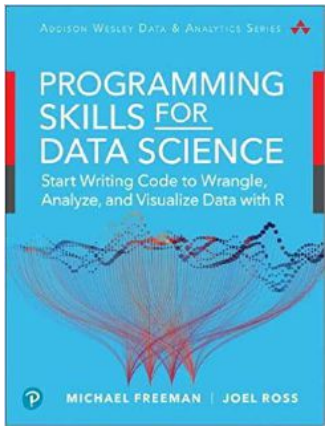
Recommended Sources

Textbooks [1, 2, 3, 4]

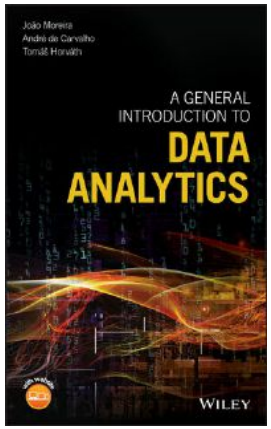


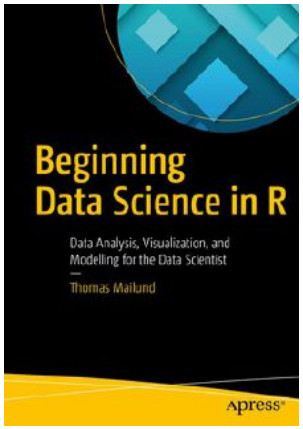


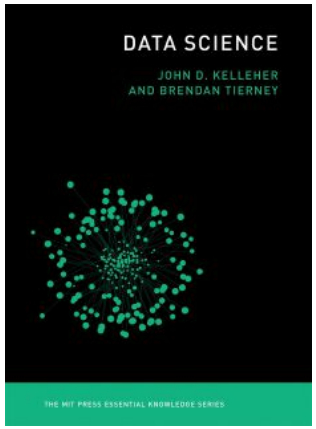




Additional Resources [5, 6, 7]







Outline

Course Contents – Week 2

Introduction to Data Analytics [8]

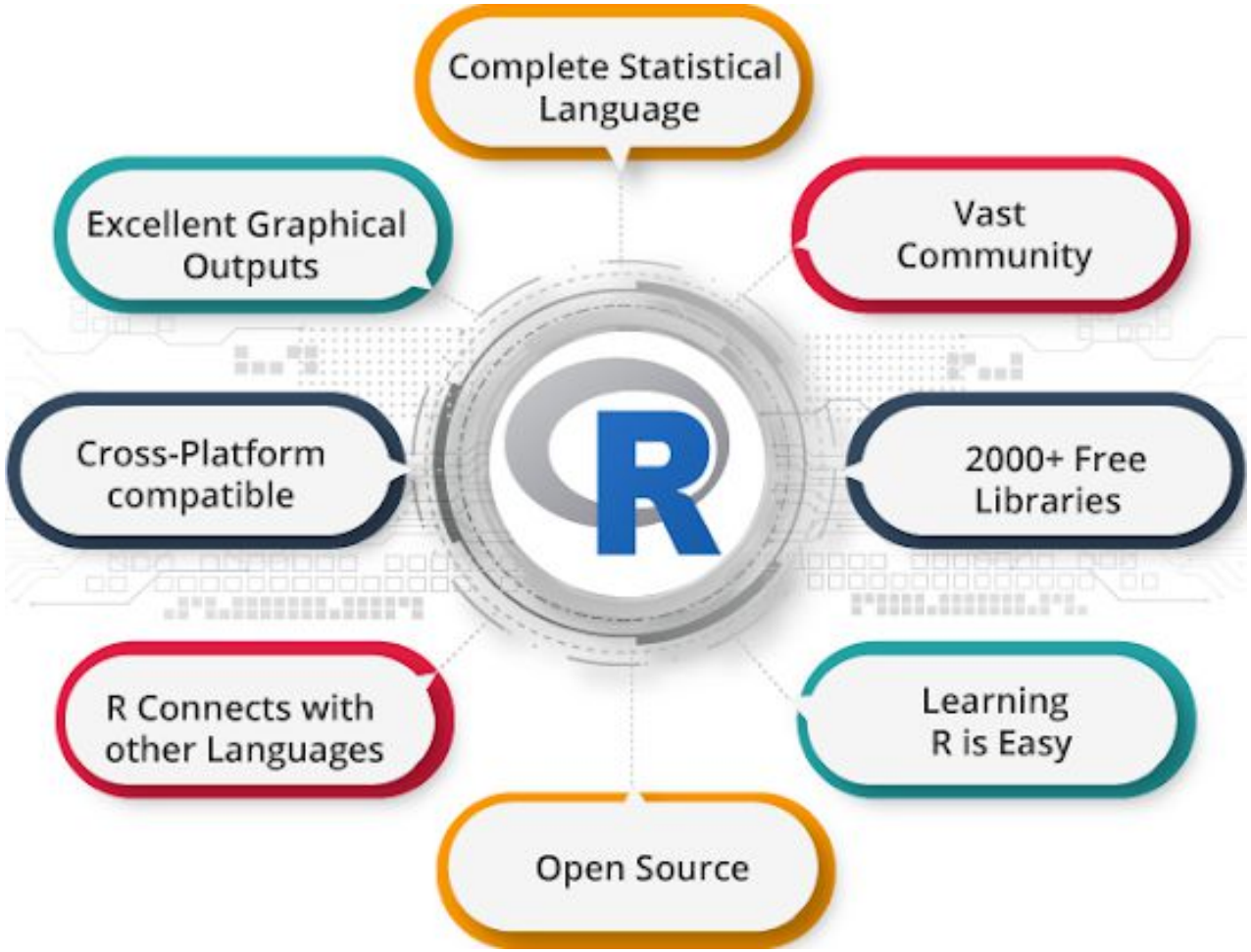
	Descriptive	Predictive	Prescriptive
	What HAS happened?	What COULD happen?	What SHOULD happen?
What the user needs to DO	• Increase asset reliability • Reduce labor and inventory costs	• Predict infrastructure failures • Forecast facilities space demands	• Increase asset utilization • Optimize resource schedules
What the user needs to KNOW	• The number and types of asset failures • Why maintenance costs are high • The value of the materials inventory	• How to anticipate failures for specific asset types • When to consolidate underutilized facilities • How to determine costs to improve service levels	• How to increase asset production • Where to optimally route service technicians • Which strategic facilities plan provides the highest long-term utilization
How analytics gets ANSWERS	• Standard reporting - What happened? • Query/drill down - Where exactly is the problem? • Ad hoc reporting - How many, how often, where?	• Predictive modeling - What will happen next? • Forecasting - What if these trends continue? • Simulation - What could happen? • Alerts - What actions are needed?	• Optimization - What is the best possible outcome? • Random variable optimization - What is the best outcome given the variability in specified areas?
What makes this analysis POSSIBLE	• Alerts, reports, dashboards, business intelligence	• Predictive models, forecasts, statistical analysis, scoring	• Business rules, organization models, comparisons, optimization

Business value →

Outline

Course Contents – Week 3

Introduction to R Programming Language [9]



Outline

Course Contents – Week 4

Data Structures

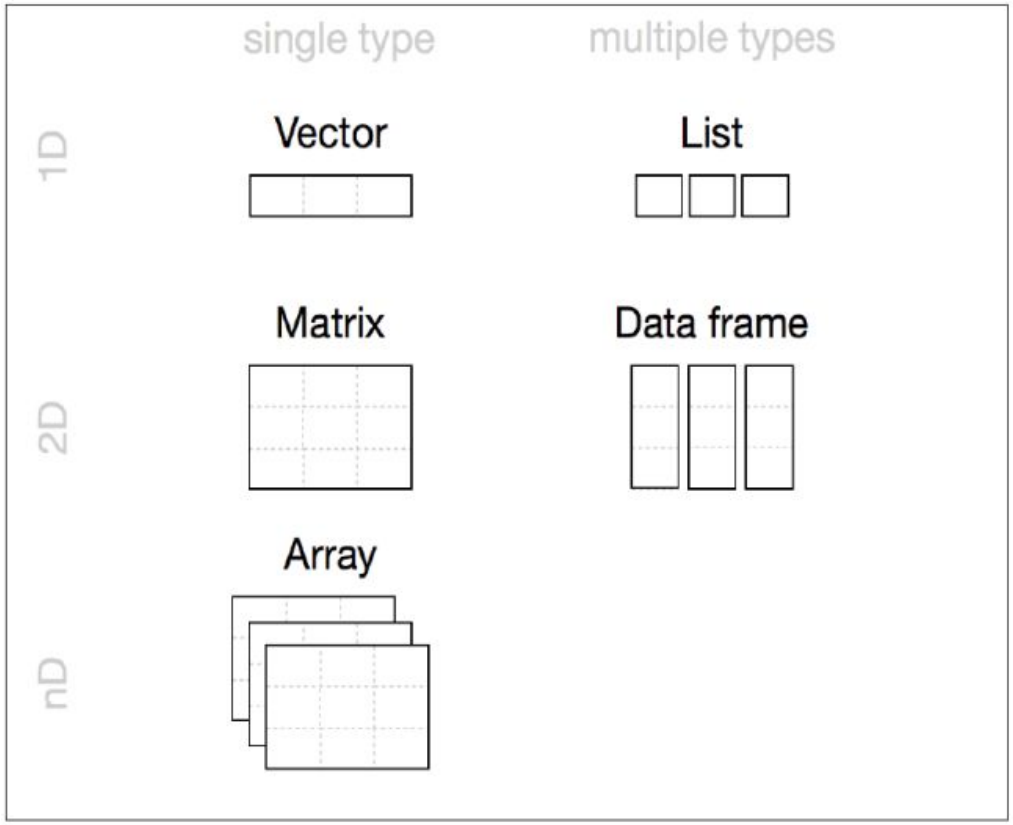


Figure 1: Common Data Structures in R [10]

Outline

Course Contents – Week 5

Control Structures

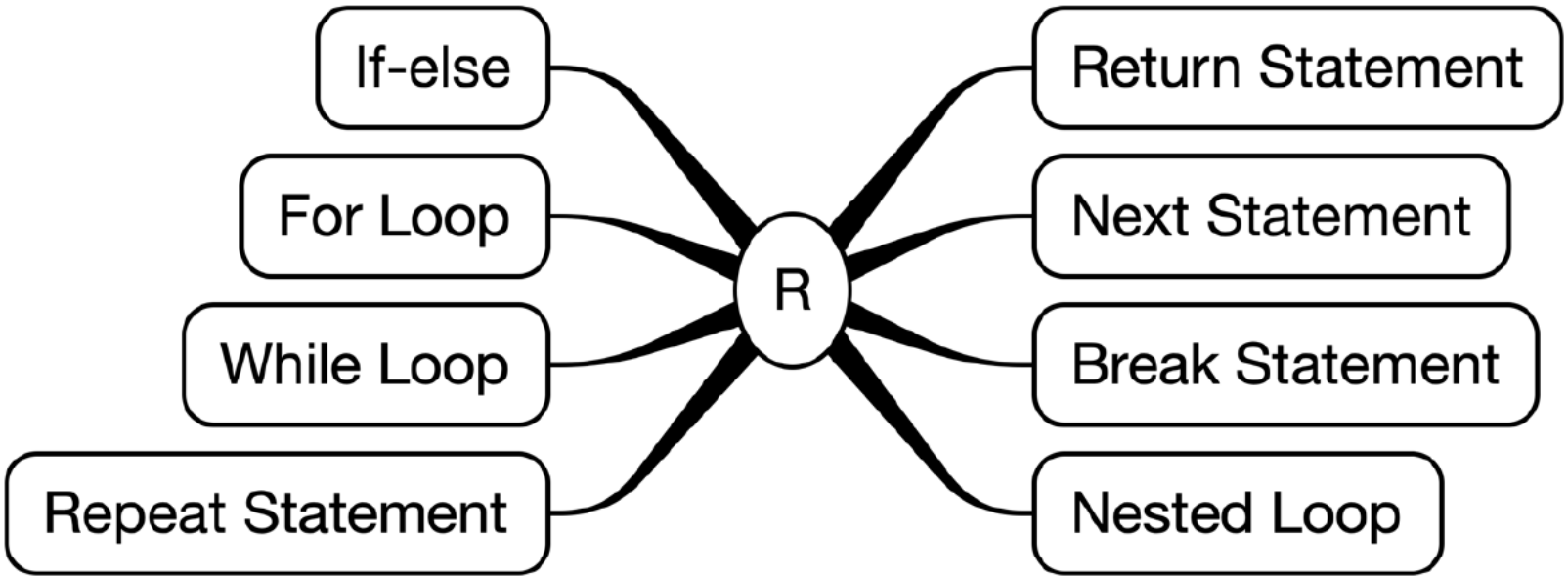


Figure 2: Illustration of Control Structures in R

Outline

Course Contents – Week 6

Functions

Functions

```
function_name <- function(var){
    Do something
    return(new_variable)
}
```

Example

```
square <- function(x){
    squared <- x*x
    return(squared)
}
```

Figure 3: Illustration of Functions in R [11]

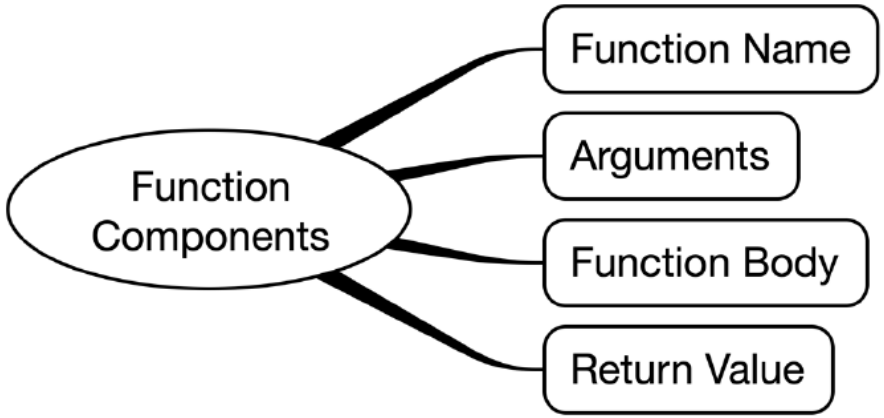


Figure 4: Function Components in R

Outline

Course Contents – Week 7 and 8

Data Wrangling

Data wrangling, is the process of importing, cleaning, and transforming raw data into actionable information for analysis [12].

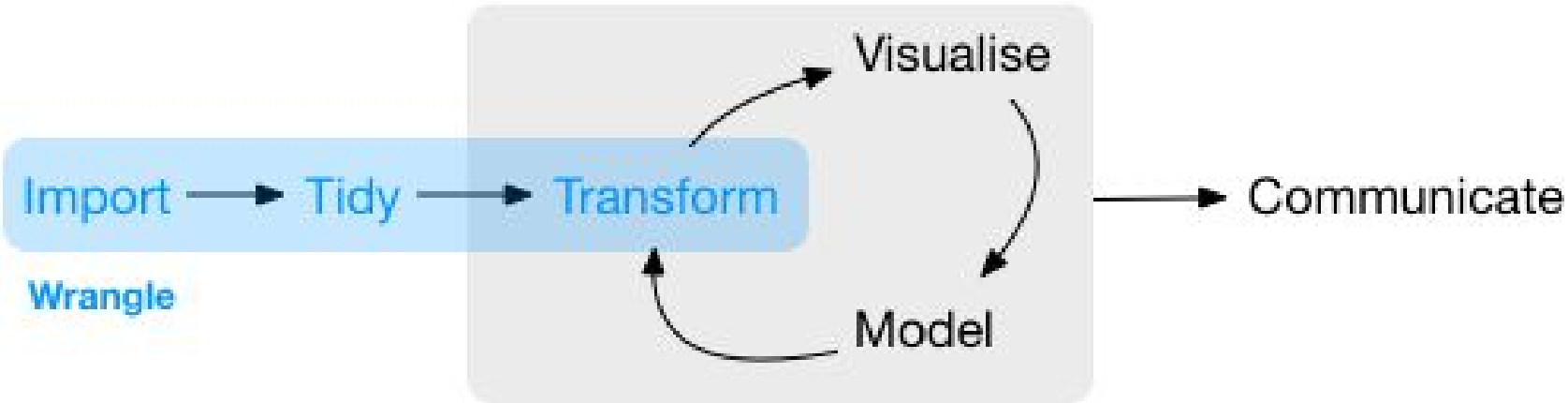


Figure 5: Demonstration of Data Wrangling Process [1]

Outline

Course Contents – Week 9

Applied Midterm Examination

An example of midterm exam and its solutions will be shared with students before the exam.

#	Difficulty	Minutes	Pts	Scope
Q1	Very Easy	5	10	W1–W3
Q2	Easy	10	15	W1–W4
Q3	Moderate	30	25	W5–W7
Q4	Hard	45	50	W5–W8
Total		90	100	W1–W8

Table 4: Assessment of Midterm Examination

Outline

Course Contents – Week 10

Tidyverse Packages for Data Wrangling

dplyr is a grammar of data manipulation, while the goal of tidyr is to help you create tidy data [13, 14].

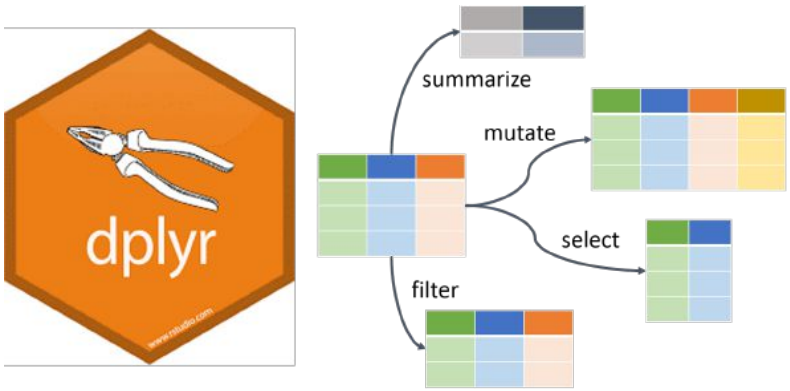


Figure 6: dplyr Package [15]

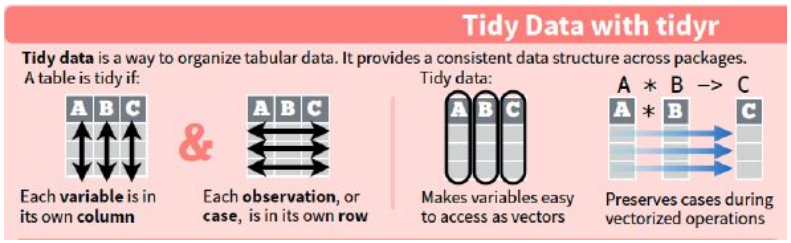


Figure 7: tidyr Package [16]

Outline

Course Contents – Week 11 and 12

Tidyverse Package for Data Visualisation

Via ggplot2, any graph can be built from the same components: a data set, a coordinate system, and geoms–visual marks that represent data points [17].

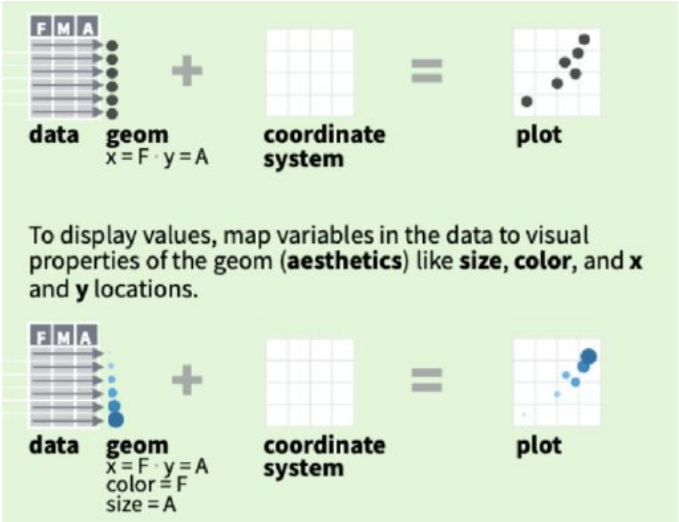


Figure 8: ggplot2 Package [17]

Outline

Course Contents – Week 13

Exploratory Data Analysis (EDA)

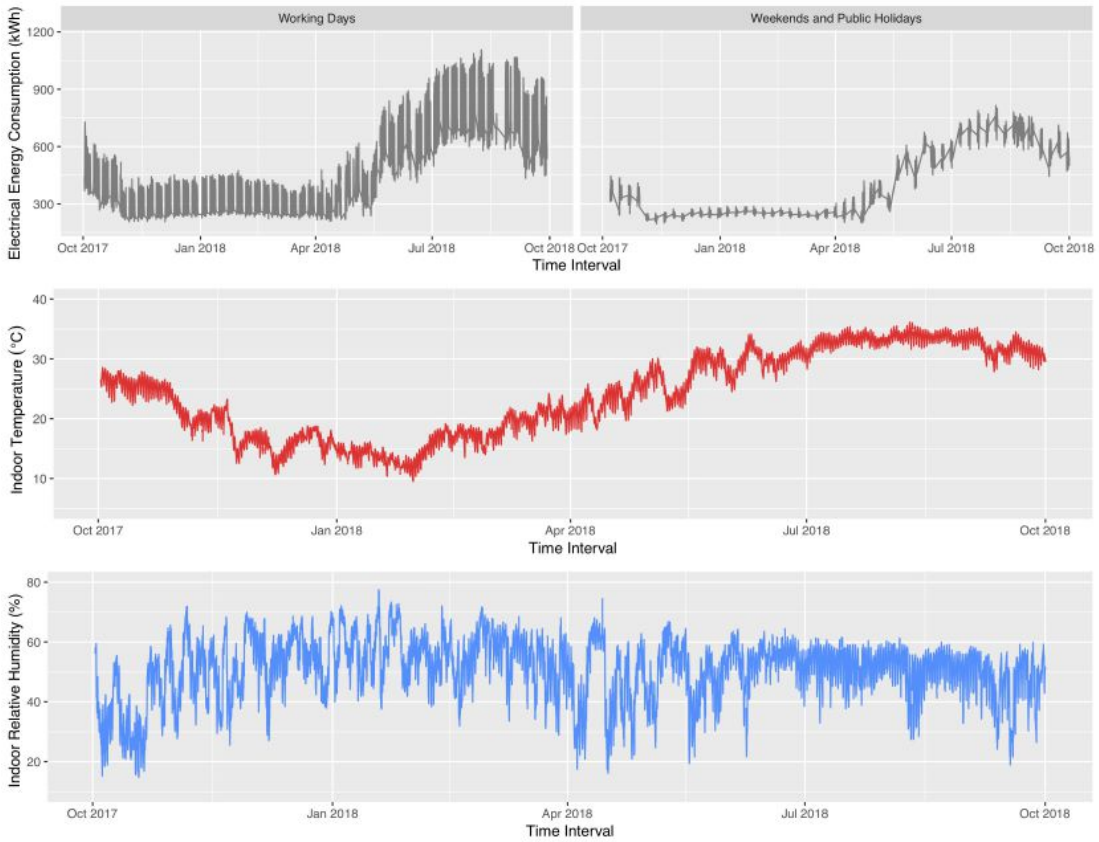


Figure 9: An Output of EDA [18]

Outline

Course Contents – Week 14

Approaches to Missing Data

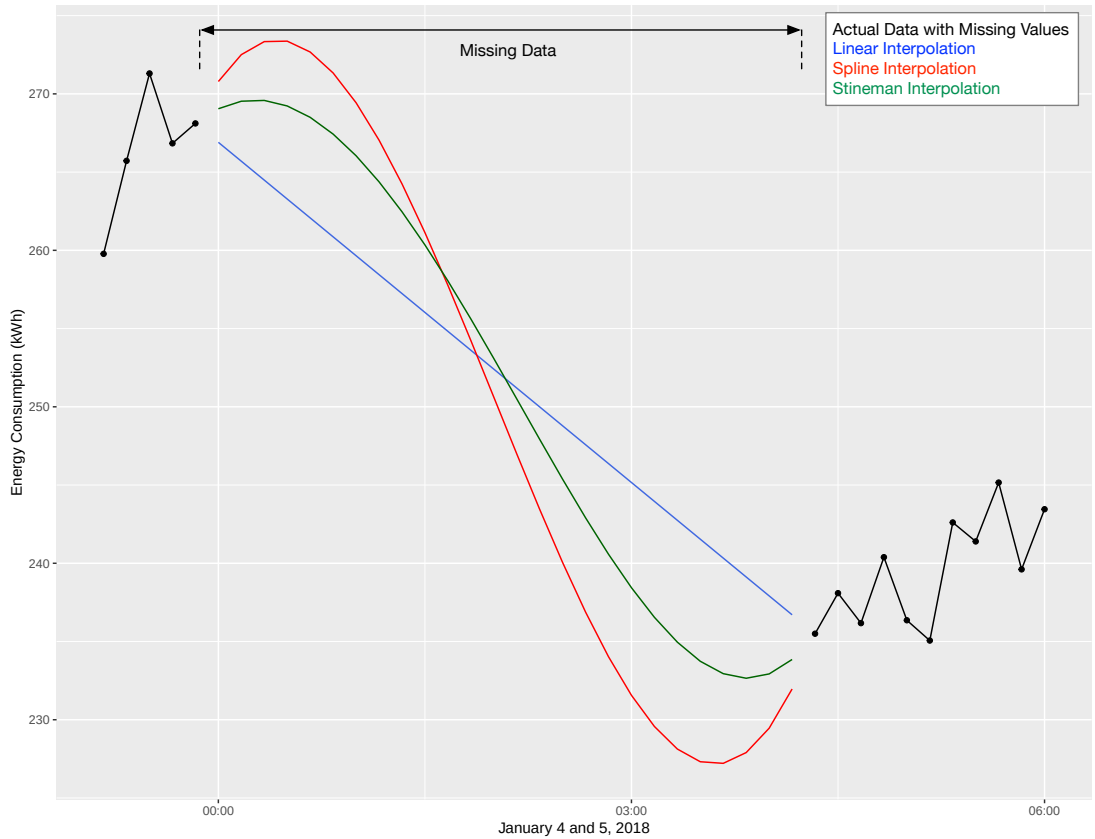
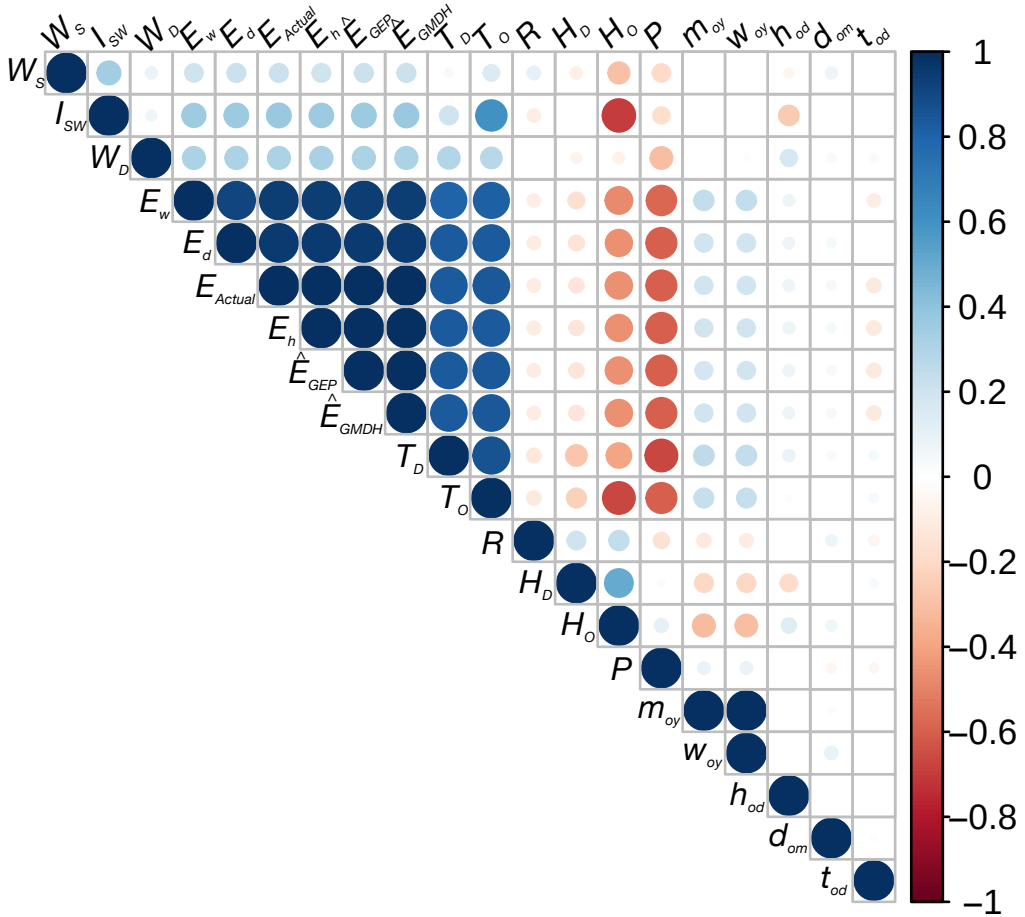


Figure 10: An Example of Missing Data Imputation [18]

Outline

Course Contents – Week 15

Case Study: Correlation Map (Blank p-values < 0.01) [19]



Outline

Course Contents – Week 16

Applied Final Examination

A computer-based applied exam will be carried out for the final exam.

- A CSV file containing a data set will be provided.
- The followings will be applied to the data set:
  - Data wrangling including missing data imputation,
  - Data visualisation,
  - Exploratory data analysis,
  - Case study (similar to Week 15).
- Students will act in accordance with the instructions given by the instructor throughout the applied exam.
- Obtained results will be reported in according to the instructions and delivered to the instructor by storing them in a USB memory.

References I

[1] Hadley Wickham and Garrett Grolemund. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media, 2017. URL https://r4ds.had.co.nz.

[2] Özgür Ergül. *Guide to Programming and Algorithms Using R*. Springer London, 2013. doi: 10.1007/978-1-4471-5328-3. URL https://doi.org/10.1007/978-1-4471-5328-3.

[3] Eric Pimpler. *Data Visualization and Exploration with R: A practical guide to using R, RStudio, and Tidyverse for data visualization, exploration, and data science applications*. Geospatial Training Services, 2017.

[4] Michael Freeman and Joel Ross. *Programming Skills for Data Science: Start Writing Code to Wrangle, Analyze, and Visualize Data with R*. Addison-Wesley, 2019. ISBN 978-0-13-513310-1.

[5] Joao Mendes Moreira, Andre C. P. L. F. de Carvalho, and Tomas Horvath. *A General Introduction to Data Analytics*. John Wiley & Sons, 2019. ISBN 978-1-119-29625-6.

[6] Thomas Mailund. *Beginning Data Science in R*. Apress, 2017. doi: 10.1007/978-1-4842-2671-1. URL https://doi.org/10.1007/978-1-4842-2671-1.

[7] John D. Kelleher and Brendan Tierney. *Data Science*. The MIT Press, 2018. ISBN 978-0-262-53543-4.

[8] IBM Watson. Descriptive, predictive, and prescriptive analytics, 2015. URL https://gemba.nl/wp-content/uploads/2015/10/watsonbusinessvalue.png.

[9] Meera Kumar. Benefits of r programming, 10th Nov, 2018. URL https://miro.medium.com/max/1280/0*xhbfmMlL68YQ3Jk7.png.

References II

[10] Fan Ting Wei. Common data structures in r, 2019. URL https://i0.wp.com/nusbasdata.files.wordpress.com/2018/01/datastructures.png?ssl=1&w=450&zoom=2.

[11] Mhairi McNeill. Base r cheat sheet, May, 2016. URL https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=2ahUKEwim0cSLuuDlAhXQAxAIHUPZAXMQFjAAegQIABAC&url=https%3A%2F%2Frstudio.com%2Fwp-content%2Fuploads%2F2016%2F05%2Fbase-r.pdf&usg=AOvVaw2N09W19RLkf2S9pwnVCCmr.

[12] Vasileios Tsakalos. Data wrangling : Transforming (3/3), 2nd Aug, 2017. URL https://www.r-bloggers.com/data-wrangling-transforming-33/.

[13] Hadley Wickham, Romain François, Lionel Henry, and Kirill Müller. dplyr, 2019. URL https://dplyr.tidyverse.org.

[14] Hadley Wickham and Lionel Henry. tidyr, 2019. URL https://tidyr.tidyverse.org.

[15] Jeff Griesemer. Data manipulation in r with dplyr, 5th Sep, 2019. URL https://miro.medium.com/max/1840/1*NXRsFH_12sfj79W-P4qI0Q.png.

[16] Roshan Talimi. tidyr, 2019. URL http://talimi.se/wp-content/uploads/2017/07/Sk\unhbox\voidb@x\bgroup\accent127a\penalty\@M\hskip\z@skip\egrouprmavbild-2017-07-21-kl.-19.10.36.png.

[17] Hadley Wickham, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, and Hiroaki Yutani. ggplot2, 2019. URL https://ggplot2.tidyverse.org.

References III

[18] Kasım Zor. *Research and Application of Real-Time Short-Term Electrical Energy Consumption Forecasting Using Artificial Intelligence Based Techniques*. PhD thesis, Department of Electrical and Electronics Engineering, Institute of Natural and Applied Sciences, Çukurova University, Adana, Turkey, 6th Sep, 2019.

[19] Kasım Zor, Özgür Çelik, Oğuzhan Timur, and Ahmet Teke. Short-term building electrical energy consumption forecasting by employing gene expression programming and gmdh networks. *Energies*, 13(5), 2020. ISSN 1996-1073. doi: 10.3390/en13051102. URL https://www.mdpi.com/1996-1073/13/5/1102.