

Session 4 Topic: Data Structures and Visualization

Based on main textbook chapter 4.

Topic Objectives

- Learn data structures at variable and dataset level.
- Understand the impact of data structure on downstream analysis (plots, tables, statistics, models,...etc).
- Learn to facet charts for visual comparison.

Hence, this topic focus on **intended learning outcomes 2 and 3**:



2. Apply selected predictive techniques to solve the business problem.



3. Explain the results of the selected predictive techniques in the context of the business problem.

Quiz

Ungraded. Check your understanding of this Session Content.
Use your real name (not nickname) in the quiz.

Discuss solution to Exercise 4.1

- Est. Duration: 40 mins
- Warm call first responders.
- Others can add on after the first response.

Team Activity: Gender Discrimination Lawsuit

Submit within 5 days. Presentation at start of next class.

- Odd-numbered Team: Analytics consultant for the female doctors.
- Even-numbered Team: Analytics consultant for the college.
- Individually, read the lawsuit documentation PDF and analyze the data.
- As a team, discuss and prepare your case using tables, statistics or/and visualization.
- May use Base R, data.table, ggplot2 or other R packages.
- Models not required for now.
- Include a conclusion slide that summarize the key points in support of your case.
- Submit (a) team PPTX (max 8 Slides) and (b) Rscript, in NTULearn class site > Teams
- This is a team submission. Only one set of files to be submitted per team.
- Some teams will be selected by instructor to present their case for their client.
 - Estimated only 1 or 2 teams from each side will be selected to present their case in next class.
- Target audience: Senior Mgt, Court Judge and Jury. i.e. not analytics expert/programmer.
- Find your best way to analyze and present the case to support your client.

Selected Team Presentation and Q&A at start of next class

- Est. Duration for entire activity: 60 mins
- Selected odd-numbered teams presents first, then selected even numbered teams. i.e. Plaintiffs before the defence.
- Presentation Duration: Est. 10 mins per team.
- After each presentation, opposing team (any) to ask at least 2 questions to the presenting team.
- Then open to any student in any team to ask question (if any).
- Good analysis, presentation, question and response will earn extra class participation points.
- Students may still earn class participation points for good questions or answer even if not presenting case.

Summary in terms of Intended Learning Outcomes

Impact of different data type on charts, tables, statistics. Use ggplot2, esp. faceting. Select and apply tables and charts to support your client in Gender Discrimination Lawsuit.



2. Apply selected predictive techniques to solve the business problem.



3. Explain the results of the selected predictive techniques in the context of the business problem.

Present your case for your client using selected tables, statistics or/and charts, regardless of which side.

Reflection on your Learning (5 mins)

Go

NTULearn Class Site > Journal

Post

Read the instructions and post your entry on this Session.

- Reply on the 3 questions as stated in the Journal Instructions.