

Introduction to Statistics and Data Science using *eStat*

Chapter 2 Visualization of Qualitative Data

2.4 Word Cloud

Jung Jin Lee

Professor of Soongsil University, Korea

Visiting Professor of ADA University, Azerbaijan

2.4 Word Cloud

- **Word cloud is useful for quickly perceiving the most prominent words to determine its relative prominence.**
 - ⇒ **to visualize the topical content of political speeches.**
 - ⇒ **website navigation aids to determine hyperlink to items associated with the word in social software.**
 - ⇒ **keyword in the cloud is sometimes used as a search engine marketing term that are relevant to a specific website.**
- **There are many algorithms to generate a word cloud.**
 - ⇒ **eStat adopted the algorithm of d3 open software.**
 - ⇒ **some evolving approaches by applying word co-occurrences**

2.4 Word Cloud

[Example 2.4.1] Generate a word cloud on the following description on data science and analyze important words.

The development of these technologies has created massive data, simply called 'Big Data', that were unimaginable in the past. Typical examples of the big data include data from Google's search records, social media messages by mobile phones, web logs by internet connections, and telephone records of global telecom companies. The big data are expected to grow and increase exponentially in the future and the hyper-forecasting is also expected to be possible. The success or failure of each individual, group, company and even country would depend on how to utilize the big data efficiently.

The analysis of the big data that emerged this century is so enormous and diverse in the amount of data that can not be fully utilized just by traditional statistical approaches. For the analysis and utilization of the big data, theories of statistics, computer science, mathematics, management or related disciplines must also be applied simultaneously. Data Science is a new area of study in which statistics, mathematics, computer science and other disciplines are fused to analyze and utilize the big data that emerged this century.



Thank you