Tools of Data Science (G7, W7, M7) Cheat Sheet

by [deleted] via cheatography.com/2754/cs/11059/

G7 is "Graphs of Numbers"

Stratified Sampling is also useful in G7

- **Bar Graph:** Analysis of size. ("Pareto Chart" is arranged bar garaph.)
- **Line Chart:** Time series analysis. ("Control Chart" is one of the line chart.)
- **Circle Graph:** Analysis of ratio
- **Histogram:** Analysis of distribution of 1 variable. (One of the Bar Graph)
- **Box Plot:** Analysis of distribution of 1 variable. (similar to stratified histogram)
- **Scatter Diagram:** Analysis of the relationship of two variables. Analysis of distribution of composed 2 variables. (To find outlier or to study small data set)
- **Heat Map:** Analysis of the relationship of two variables. Analysis of distribution of composed 2 variables. (To study big data set)

W7 is "Analysis of Words"

Most of W7 are Concept Analysis .

Affinity Diagram: Classification of idea. The method to collect Brains-torming

Cause-and-Effect Diagram: To collect reasons and results.

Tree Diagram: Similar to FMEA . The next step of Cause-and-Effect Diagram.

Relation Diagram: Main part of Systems Thinking .

Matrix Diagram: Applications are QFD , Multi Dimensional Scaling and AHP.

Arrow Diagram: Planning method.

Flow Chart: Analysis of the process

Why-why Analysis is a basis of concept analysis

W7 analyze the structure of phenomena as levels or networks. The idea to think the structure is also useful in G7 and M7.

M7 is "Mathematical Analysis"

- **Error Analysis:** Analysis of the quality of the data
- **Average & Standard Deviation:** Analysis of statistical value.
- **Testing of Difference of the Average:** Basic tool of Hypothesis Testing.
- **Regression Analysis:** Include of the analysis Correlation .
- **Principal Component Analysis:** In N7, called "Matrix Data Analysis".
- **Decision Tree:** Application of Stratified Sampling
- **Linear Programming:** To find the best in constraints.