



UNIVERSITI
TEKNOLOGI
MARA

WHAT'S WHAT FSKM



2756-7729

APRIL 2026
VOLUME 1

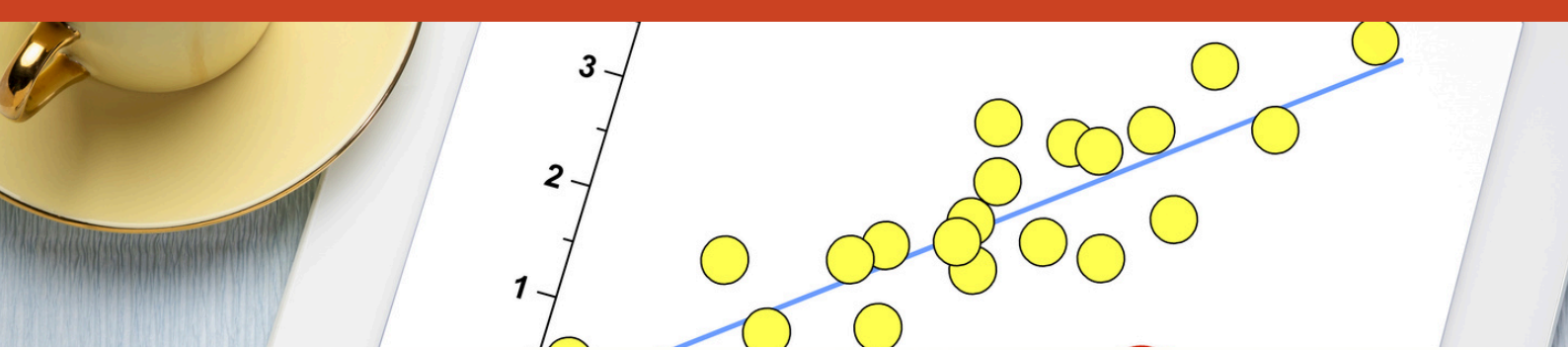
LEVERAGING ONLINE
REVIEWS FOR SERVICE
IMPROVEMENT:
SENTIMENT ANALYSIS
OF SUKI-YA PAVILION

TIME SERIES
FORECASTING USING
THE PROPHET MODEL

MEDIA SOSIAL DAN
KESIHATAN MENTAL
PELAJAR

SMALL CHANGES, BIG IMPACT:
THE ROLE OF THRESHOLD
SELECTION IN EXTREME RISK
ASSESSMENT

VISUALIZING CORRELATION
STRUCTURES IN DATA USING R:
METHODS AND APPLICATIONS



VISUALIZING CORRELATION STRUCTURES IN DATA USING R: METHODS AND APPLICATIONS

Nurul Nisa' Khairol Azmi

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM), Cawangan Negeri Sembilan, Kampus Seremban, 70300, Negeri Sembilan Darul Khusus, Malaysia

nurulnisa@uitm.edu.my

EXECUTIVE SUMMARY

Correlation analysis is commonly used to examine relationships among variables in multivariate datasets. However, traditional correlation matrices presented as numerical tables often limit interpretability, especially when dealing with large numbers of variables. This article demonstrates how correlation plots can enhance data visualization and support exploratory data analysis using the R programming language. Several visualization approaches, including heatmaps, hierarchical clustering correlation plots, and scatterplot matrices are illustrated. The paper highlights interpretation strategies, common pitfalls and practical considerations when applying correlation visualization in research and teaching. The proposed workflow aims to assist researchers, analysts and educators in transforming correlation analysis into a more intuitive and insightful visual process

INTRODUCTION

Data visualization plays a crucial role in exploratory data analysis (EDA), allowing researchers to identify trends, clusters, redundancies, and potential multicollinearity before model development (Castellano-Escuder, et al. 2021). Among various visualization methods, correlation plots provide a compact and intuitive representation of pairwise relationships (Koo et al., 2018). With the rapid growth of open-source statistical computing tools, the R programming language offers several packages that support flexible and informative correlation visualization. This article presents a practical guide for constructing and interpreting correlation plots in R.

CORRELATION VISUALIZATION IN R USING SPECIALIZED PACKAGES

R has become one of the most widely used platforms for statistical computing due to its flexibility, reproducibility, and strong visualization capabilities. Several packages have been developed specifically for correlation visualization, including `corrplot`, `ggcorrplot`, `GGally`, and `PerformanceAnalytics`.

CORRELATION ANALYSIS OF THE MTCARS DATASET USING R

The `mtcars` dataset, which contains technical specifications and fuel consumption measurements for 32 automobiles, was used to demonstrate correlation techniques. All variables are numerical, hence the dataset suitable for correlation analysis.

Load Data and Required Packages

```
# Install packages (run once if needed)
install.packages(c("corrplot", "GGally", "ggcorrplot", "PerformanceAnalytics"))

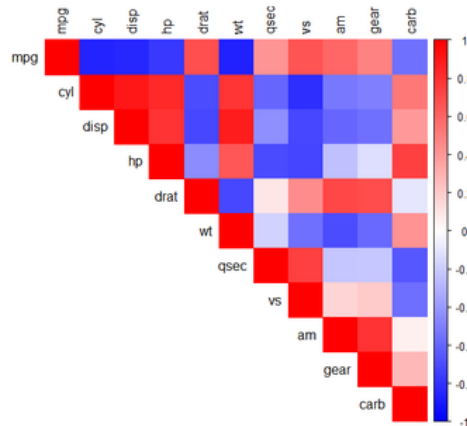
# Load libraries
library(corrplot)
library(GGally)
library(ggcorrplot)

# Compute correlation matrix
cor_matrix <- cor(mtcars)
```

Heatmap Correlation Plot

```
corrplot(cor_matrix, method = "color", type = "upper",  
col = colorRampPalette(c("blue", "white", "red"))(200), tl.col = "black")
```

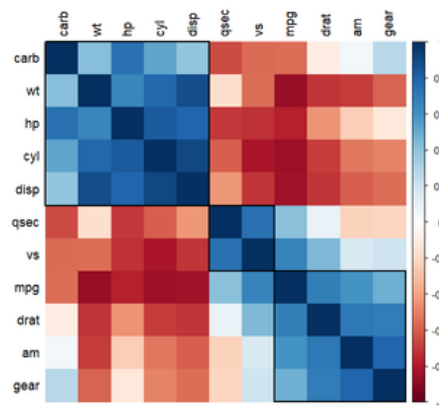
Heatmap visualization emphasizes correlation strength using color gradients. Positive correlations appear in warm tones, whereas negative correlations are displayed in cooler tones, facilitating quick identification of association patterns.



Hierarchical Clustering Correlation Plot

Hierarchical clustering reorders variables based on similarity, grouping highly correlated variables together. This helps reveal hidden structures and clusters that may guide variable selection or dimensionality reduction.

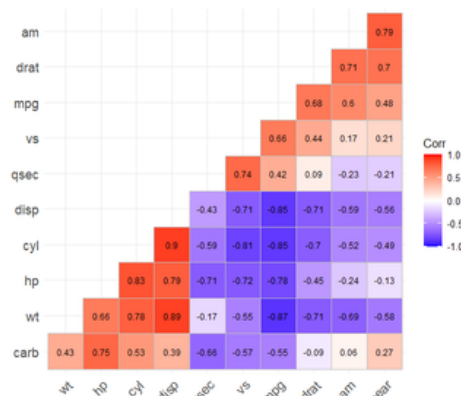
```
corrplot(cor_matrix, method = "color", order = "hclust", addrect = 3, tl.col = "black")
```



Publication-Style Correlation Plot (ggcorrplot)

The ggcorrplot package provides enhanced aesthetics and coefficient labels, improving readability and suitability for publication figures.

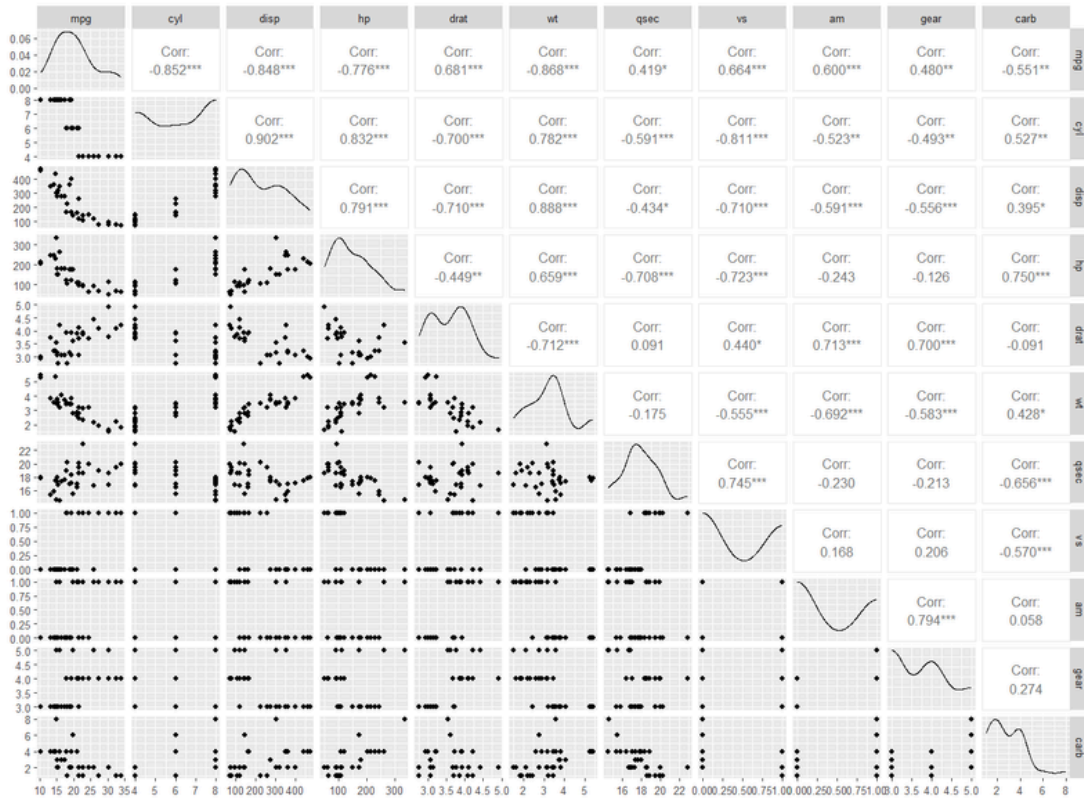
```
ggcorrplot(cor_matrix, hc.order = TRUE, type = "lower", lab = TRUE, lab_size = 3,  
colors = c("blue", "white", "red"))
```



Correlation with Significance Testing

Scatterplot matrices complement correlation plots by displaying pairwise scatterplots, marginal distributions, and correlation coefficients simultaneously. This visualization helps identify linearity, outliers, and distributional characteristics.

```
ggpairs(mtcars)
```



CONCLUSION

Correlation plots transform numerical association measures into intuitive visual representations that enhance exploratory data analysis. Through R-based visualization techniques, researchers and educators can better understand variable relationships, identify structural patterns and support more informed modelling decisions. As data complexity continues to grow, integrating correlation visualization into analytical workflows provides an effective strategy for improving statistical interpretation and communication.

REFERENCES

- [1] Castellano-Escuder, P., González-Domínguez, R., Carmona, F., Andrés-Lacueva, C., & Sánchez-Pla, Á. (2021). POMASHiny: A user-friendly web-based workflow for metabolomics and proteomics data analysis. *Plos Computational Biology*, 17(7), e1009148. <https://doi.org/10.1371/journal.pcbi.1009148>
- [2] Koo, H., Chun, Y., & Griffith, D. A. (2018). Integrating spatial data analysis functionalities in a GIS environment: Spatial Analysis using ArcGIS Engine and R (SAAR). *Transactions in Gis*, 22(3), 721-736. <https://doi.org/10.1111/tgis.12452>