Nov 11, 2024

Data Visualization
Week 6. Visualizing associations
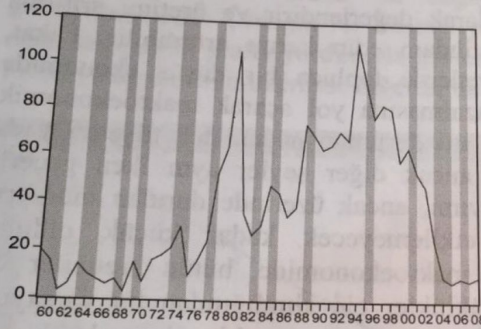
Reminder

Visualizing proportions:

- Barplot is always better

- Pie chart can be useful for easy proportions

Visualizing nested proportions:

- Mosaic plot

- Treemap

- Sankey diagram

İKTİSADİ BAKIŞ 4

Yandaki grafik bize Türkiye'de enflasyon oranının 1959-2010 döneminde izlediği seyri gösteriyor. Grafikte görülen taralı alanlar ise Fenerbahçe'nin şampiyon olduğu futbol

sezonlarını ifade ediyor. Grafiğe göre, Fenerbahçe'nin ligde iyi performans gösterdiği dönemlerde enflasyon oranı düşüyor, kötü performans gösterdiği dönemlerde ise, genellikle, enflasyon yükselme eğilimine giriyor. Sizce birkaç sezon üst üste Fenerbahçe'yi daha sezon başında şampiyon ilan ederek Türkiye ekonomisini enflasyon sorunundan kurtarmamız mümkün mü?

Correlation does not imply causation; however, it is a good reason to look for one!

Introduction

Determining the relationship between two variables is a common challenge in scientific studies and practical research. This relationship can be expressed along two dimensions:

- Direction (positive or negative)
- Strength (such as low, moderate, or high)

For example, one may want to measure the relationship between air temperatures and the amount of energy needed daily, or between per capita income and average life expectancy. The most basic tool used to visualize such relationships is the scatter plot. Additionally, correlograms and bubble charts can also be utilized.
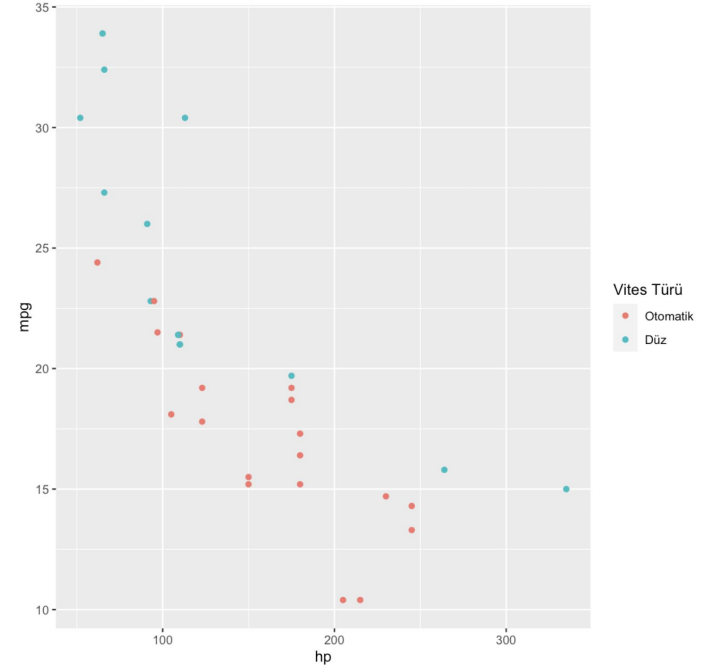
Scatterplot

- It is one of the fundamental tools used to explore the relationship between two variables.

- On a two-dimensional axis, the observed values of one variable are paired with the observed values of the other variable on the opposite axis.
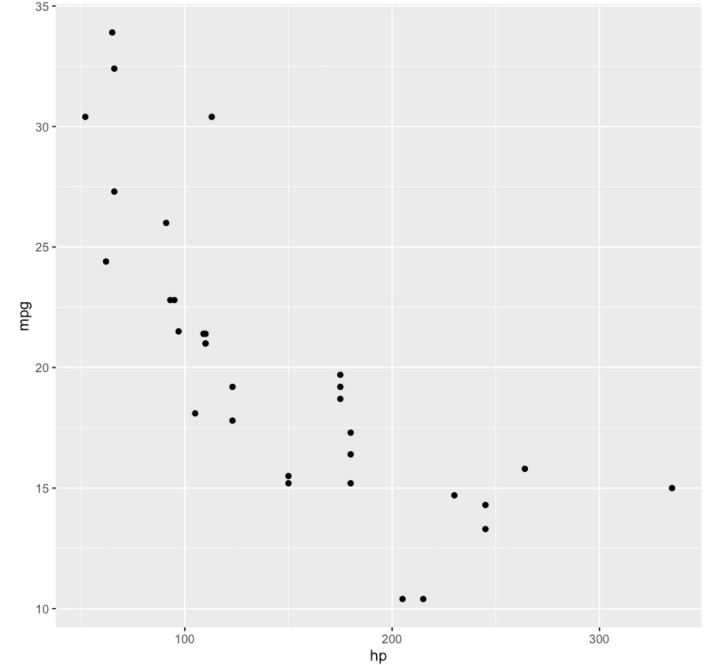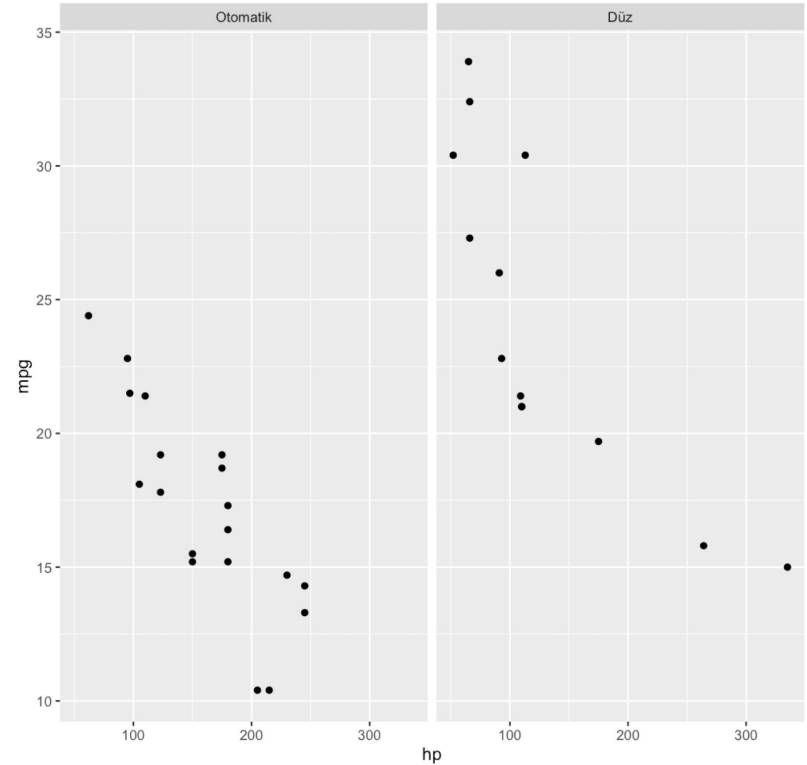
Scatterplot

By using color aesthetics in scatter plots, we can include a categorical variable in the visualization.

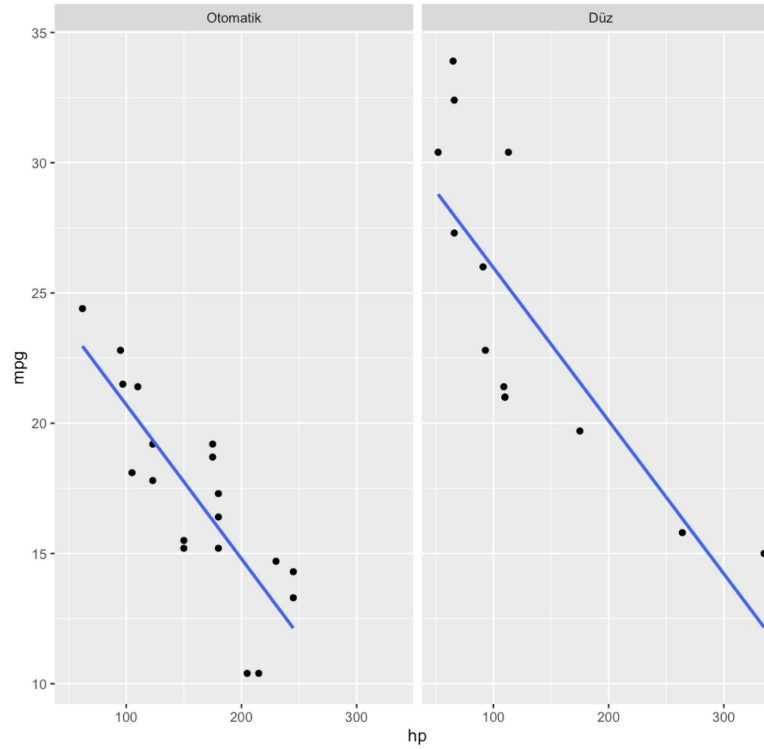**Could we use a different visualization strategy here?**

Scatterplot

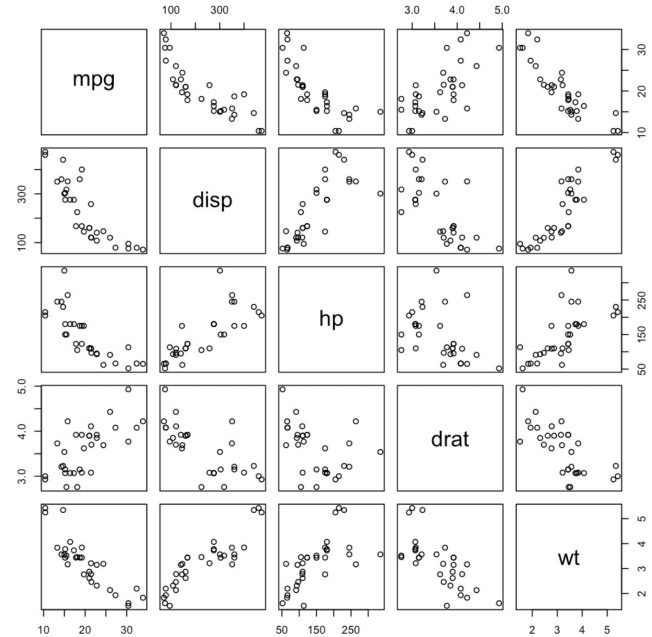By adding a regression line to scatter plots, we can reveal the trend of the observed values.
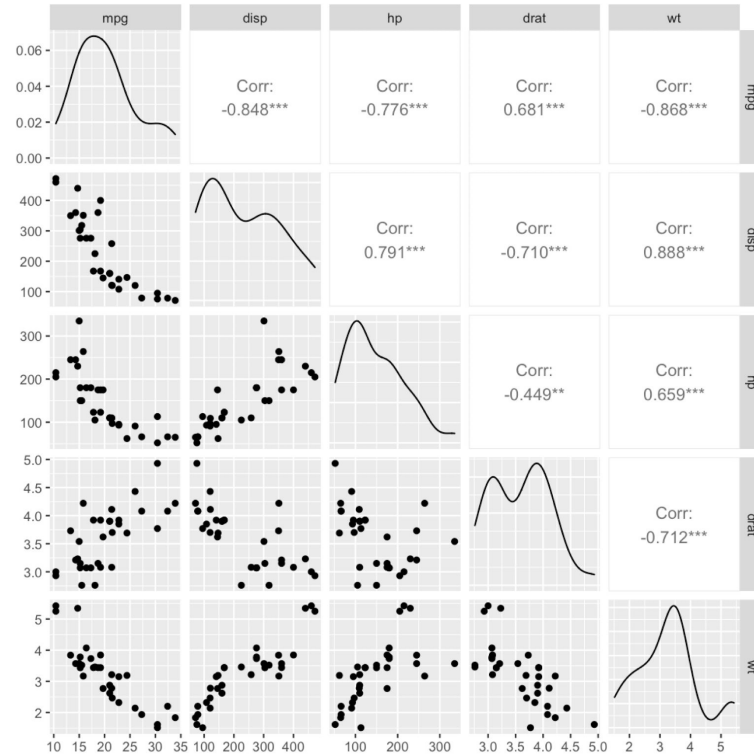
Scatterplot

Scatter matrix

Scatter matrix is used for visualizing multiple
pairs of variables.

Scatter matrix

Correlogram

As the number of variables increases, reading scatterplot matrices becomes more challenging. In such cases, instead of visualizing the relationship, it is more useful to measure the strength of the relationship and visualize these measurements. The easiest way to do this is by calculating the correlation coefficient (ρ).

- The correlation coefficient ranges between -1 and +1.
- A value of 0 indicates no relationship between the two variables, while values of -1 or +1 indicate a strong relationship.
- The sign of the correlation coefficient shows the direction of the relationship.
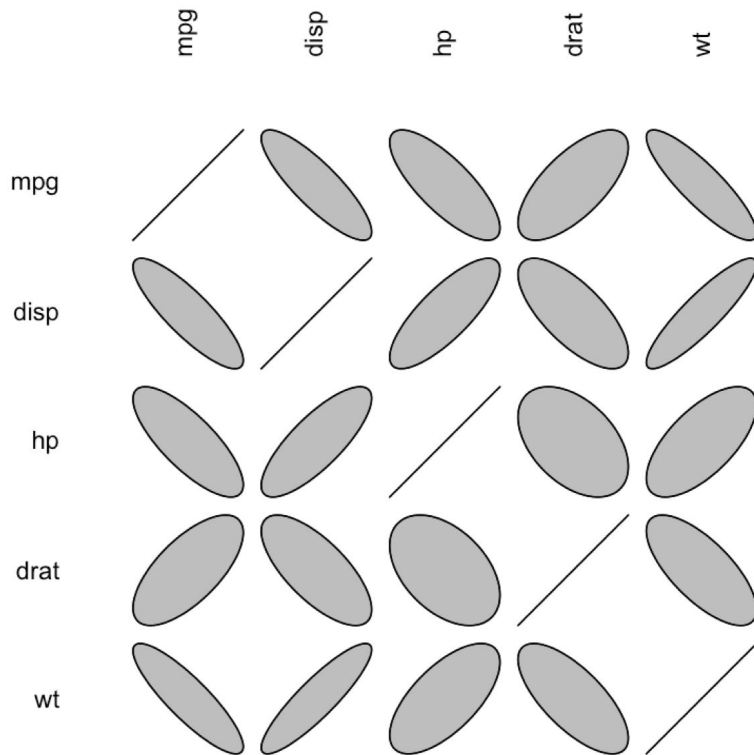- A correlogram is used to visualize correlation coefficients.

Correlogram
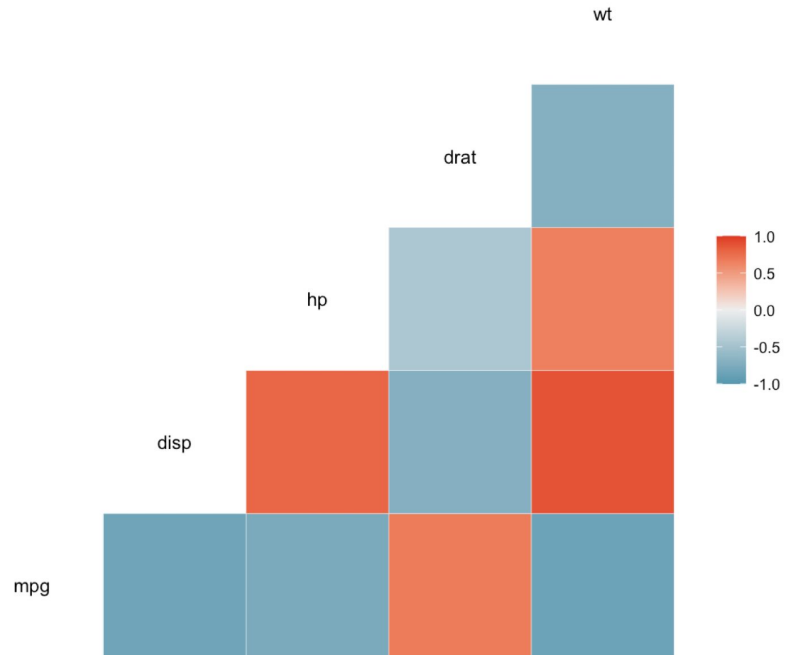
Correlogram

Here, oval shapes are positioned according to the direction of the correlation and take on different forms based on its strength.

As the correlation coefficient increases, the shapes become narrower; as it decreases, they widen and approach a circular form.

Reference

The notes and plots in the presentation are compiled from Claus O. Wilke's book, *Fundamentals of Data Visualization*.