**1. Plotting many variables at once**

To visualize many variables, you'll need to use more advanced plot types. Here we'll look at pair plots, correlation heatmaps, and parallel coordinates plots.

**2. When should you use a pair plot?**

Pair plots work with up to about ten variables at once, and they show you the distribution of each variable, and the relationship between each pair of variables.

**3. pair plot all**

Here is a pair plot of LA home prices. There are four variables in the dataset, giving four rows of panels and four columns. Let's explore this piece by piece.

**4. pair diag disc**

Panels on the diagonal show distributions of variables. City is a categorical variable, so its distribution is represented as a bar plot of counts for each city.

**5. pair diag cts**

The other three variables - number of beds, price, and area - are continuous, so their distributions are represented by histograms.

**6. pair cts**

Panels off the diagonal show relationships between pairs of variables. When both variables are continuous, you see scatter plots of each pair of variables, and their correlation. For example, in the second column, fourth row, you see a scatter plot of number of beds versus area. In the fourth column, second row, you see that number of beds and area have a positive correlation of zero-point-seven-eight-two.

**7. pair combo**

When comparing a categorical variable to a continuous variable you get a box plot and a histogram of the continuous variable split by the categorical variable. For example, in the third column, first row, you see a box plot of prices for each city. In the first column, third row, you see a histogram of the same thing. The histogram is vertical so the positions match those in the city panel on the diagonal.

**8. pair plot all again**

Pair plots can be tremendously helpful for quickly exploring a new dataset. For the special case where you have many continuous variables, a close relative of the pair plot called a correlation heatmap is simpler, and scales to visualizing even more variables at once.

**9. pair corr**

The idea is that you draw a pair plot, only including the panels for correlations, but instead of showing numbers, you show a color.

**10. When should you use a correlation heatmap?**

Correlation heatmaps are designed to show relationships between pairs of continuous variables. They are compact, so you can easily compare tens of variables at once.

**11. corr heat**

Here's a correlation heatmap of a customer satisfaction survey for a Yellow Pages advertising product. Customers rated the importance of product features on a scale from 1 to 10. Where two features consistently received similar scores by customers, they are positively correlated, and colored in a more vibrant red. Those bright reds in the bottom left show that the price related product aspects all strongly correlate with each other.

1. 1 Rossi, Allenby, and McCulloch (2005). Bayesian Statistics & Marketing

**12. The United Nations dataset again**

Here's a plot of the United Nations country data again. It showed that you can use color on a scatter plot to display three variables at once. However, with four or more variables scatter plots can quickly become complicated to interpret.

**13. When should you use a parallel coordinates plot?**

Parallel coordinates plots provide a solution to this when you have lots of continuous variables, and want to understand their relationship or group them into clusters.

**14. A parallel coordinates plot**

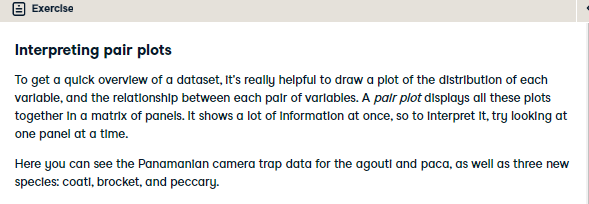
Here's a parallel coordinates plot of the three variables you saw before, plus the human development index score. Each line represents one country, and each continuous variable appears on the x-axis, just like a bar plot. To make each variable comparable, the y-axis simply ranges from the lowest value in the dataset for that variable to the highest value. Unfortunately, this plot is a mess because the world is complicated. Let's try splitting the dataset by continent.

**15. para coord by continent**

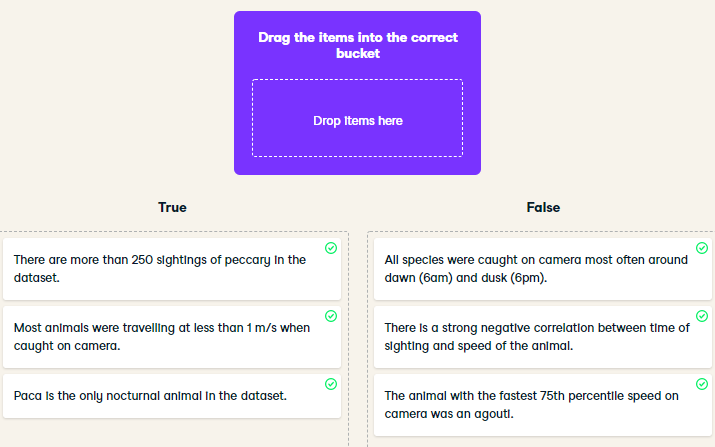
Now some patterns emerge. The South American countries are quite consistent. Their GNIs are low, and their human development index, life expectancy and schooling values are mostly between the median and the 75th percentile. In Europe, you see a wide range of GNIs, but the other metrics are all high. In Africa, the GNIs are all low, but the other metrics show a wide range. The wonderful thing about this plot is that more metrics can just be added to the x-axis. You can easily compare ten or twenty variables at once.

**16. Let's practice!**

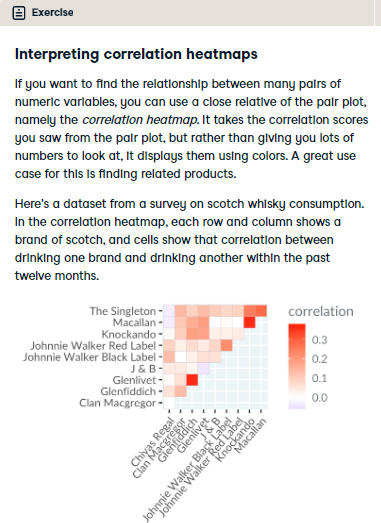
Let's visualize!

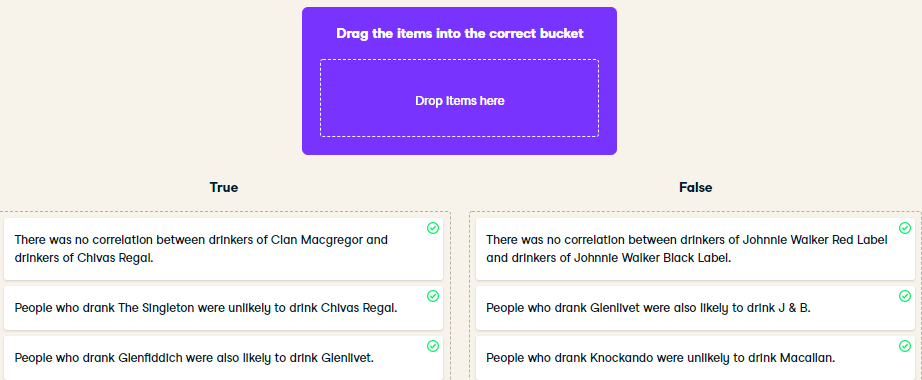




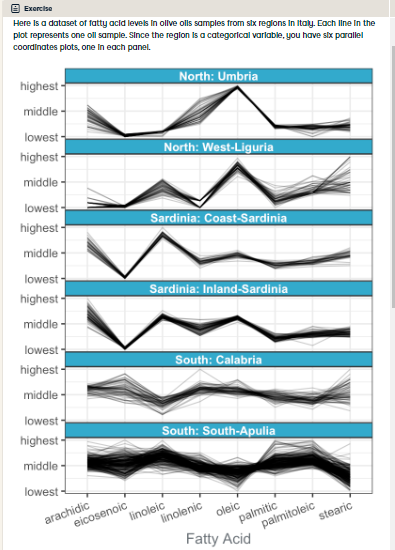


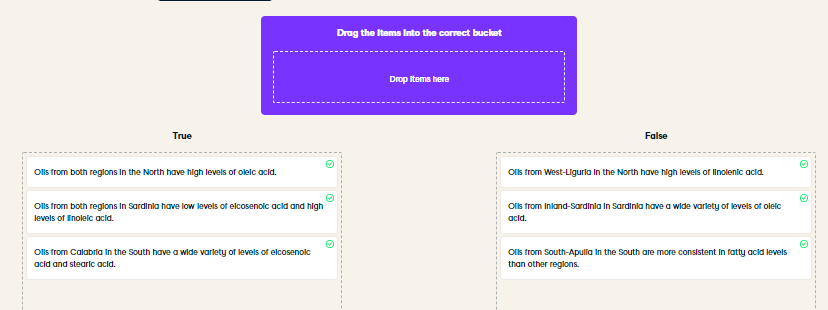
Perfect pairing! Pair plots are fantastic for getting quick insight into several variables at once.





Sláinte! Correlation heatmaps are a great way to visualize the correlation between many continuous variables at once.





Outstanding olive oil analysis! Parallel coordinates plots let you explore the relationship between many continuous variables at once.