Print

Summary of Exploratory Data Analysis Tools

In the introduction to this lesson, you learned a definition of exploratory data analysis, or EDA: Exploratory data analysis is the process of using numerical summaries and visualizations to explore your data and identify potential relationships between variables.

You also learned that exploratory data analysis is an investigative process, used to get to know your data and to understand what you might be able to learn from them.

In exploratory data analysis, and in data analysis in general, the methods you use depend on the nature of your data, the number of variables you have, and what you hope to learn from your data.

You analyze your data one variable at a time, then you analyze your data two variables at a time, and then you analyze your data more than two variables at a time. You apply the techniques that make sense, based on your data, but also based on what you learn through your exploratory journey.

In the first lesson of this module, you learned a variety of core tools for graphing data.

For continuous data, you learned about histograms, box plots, run charts, scatterplots, and scatterplot matrices

For categorical data, you learned about bar charts and mosaic plots.

You also learned how to create tabular summaries of your data, and how to use dynamically linked graphs to explore potential relationships between variables.

In this lesson, you learned additional tools for exploratory data analysis. You learned how to add colors, markers and legends, stratify your analysis with the local data filter, update a graph with different variables using the Column Switcher, use overlay variables to add multiple variables or levels of a variable to the same graph, create trellis plots to create a matrix of graphs, and add frequency and size variables to your graphs and analyses.

You also learned some powerful tools for graphing multidimensional or multivariate data.

You learned about sorted bar charts and Pareto plots, for identifying the top issues or opportunities, packed bar charts, for graphing one categorical variable with many levels, tree maps, for graphing two categorical variables with many levels, heat maps, for graphing values of a variable as colors, bubble plots, for adding motion and animation to your graph, and geographic maps, for visualizing spatial data.

This is, by no means, a complete list of tools for exploratory data analysis. But this is a powerful set of tools that can help you gain insights from your data.

For a more comprehensive set of exploratory tools, and for additional details about applying the tools you learned in this lesson, see the Read About It for this module.

Statistical Thinking for Industrial Problem Solving

Close