3 Hands On: Data Exploration

1 Summarization

Load the data set carIns final. It already has the imputation of missing values.

1. Using the package dplyr, answer the following questions:

(a) Obtain the number of cars by bodyStyle.

(b) Obtain the number of cars by bodyStyle and fuelType.

(c) Obtain the mean and the standard deviation of the attribute cityMpg by bodyStyle in ascending order.

(d) Also by bodyStyle, and for the attributes cityMpg and highwayMpg, obtain the mean, the standard deviation, the median and the inter-quartile range.

2 Visualization

2. Using the package ggplot2, create graphs that you find adequate to answer the following questions.

(e) Show the relationship between the attributes cityMpg and highwayMpg

(f) Show the distribution of cars by bodyStyle.

(g) Show the distribution of cars by price. Suggestion: create bins of width equal to 5000.

(h) Add the information of the density estimation to the previous graph.

(i) Check (visually) if it is plausible to consider that price follows a normal distribution.

(j) Show the distribution of price by make attribute. Suggestion: use boxplots and the function coord_flip().

(k) Show the distribution of price by nDoors attribute. Suggestion: use histograms.

(l) Show the distribution of price by bodyStyle and nDoors attributes. Suggestion: use histograms.

(m) Add the parameter scales="free_y" to the facet function in the previous graph.