

KUHeS CRM - Introduction to R

Session 2 - Practical

Marc Henrion

13 September 2021

Session 2 - Practical

1. Using the `iris` dataset (type `?iris` to get more information about this dataset) that comes pre-loaded with R, produce the following figures:
 - Produce histograms for each of `Sepal.Length`, `Sepal.Width`, `Petal.Length`, `Petal.Width`.
 - Produce a bar plot for `Species`.
 - Produce box and whisker plots for each of the 4 continuous variables. Put them all on a single, multi-panel figure.
 - Repeat for just `Sepal.Length` using a violin plot, stratifying by `Species`.
 - Produce a single graph (not multi-panel) that has histograms for `Sepal.Length` for each of the 3 flower species.
 - There are 4 continuous variables. This means there are 6 possible pairs of these. For each such pair, produce a scatter plot of one variable against the other and highlight the different flower species by using a different colour for each species.
 - For one of these 6 scatter plots: estimate the bivariate probability density and add density contour lines to the figure (also try adding a heatmap if you can).
2. Install the package `nycflights13`, then load it. This has data on flights that took off in the US during 2013. There are 5 data tables:
 - `airlines`, data on airlines
 - `airports`, data on airports
 - `planes`, data on planes
 - `weather`, hourly weather data at NYC airports for 2013
 - `flights`, data on flights leaving NYC airports during 2013
 - Compute the average delay by destination, then join the airports data frame to get the longitude and latitude of delays. Plot this (if you are using `ggplot2`, then the functions `borders()` and `coord_quickmap()` can be useful for a nicer figure).
 - Construct data frames giving average delay per wind speed / temperature / precipitation / visibility. Produce scatter plots of each of these against delay and add an average trend line.
3. Using the `iris` dataset, fit a linear regression model for `Petal.Width` (response) against `Sepal.Width` and `Sepal.Length`. What is your conclusion?
4. Using the `iris` dataset, test the null hypothesis that the distributions for `Sepal.Length` in the `setosa` and `virginica` flowers are the same.