HDDA Tutorial: Getting Started with R

Department of Econometrics and Business Statistics, Monash University

Tutorial 2

The aim of this week's tutorial is to do more preliminary data analysis using R. You will need to use the datasets *Beer.rds* and *comScore.rds* from which can be downloaded from Moodle.

Beer Data

1. Without using the `qplot` function, produce a histogram of the cost per 12 fl. oz. variable.
2. Without using the `qplot` function, produce boxplots of alcohol content. On the same plot there should be a separate boxplot for light beers and a separate boxplot for nonlight beers.
3. Produce a frequency table of beer rating
4. Produce a cross tab of beer rating against light/nonlight

comScore Data

The company comScore records the online behaviour of subscribers. Each observation of the dataset that you have been provided with is a unique visit to the wesite apple.com. Four variables are recorded: **Buy** indicates whether a purchase was made, **Sales** indicates the value of any purchase, **Duration** indicates how much time was spent on apple.com, while **PageViews** indicates how many pages were clicked on under the apple.com domain name in a single visit. An interesting marketing question is whether browsing behaviour (duration and page views) are associated with purchase behaviour (buy and sales).

1. Without using the `qplot` function, produce histograms of

- Sales
- Duration
- Page Views

2. Produce summary statistics for the comScore data
3. Without using the `qplot` function, produce a scatter plot of duration against page views

More Advanced

1. Create a new data frame that removes the two outliers using the `filter` function in `dplyr`.
2. Do the scatterplot again with the outliers removed.
3. Do the scatterplot where the points have a different colour if the observation corresponds to a buy and a different colour if it corresponds to no buy.