HDDA Tutorial: Introduction and R

Department of Econometrics and Business Statistics, Monash University

Tutorial 1

Motivation

1. Describe the role that preliminary data analysis plays in the overall statistical analysis of a dataset. Give examples to illustrate your answer.
2. Think of a problem from business or another discipline where the data are multivariate. How may the data be explained by a smaller number of unobserved variables.

Measurement

Think of an example of a non-metric variable and metric variable. For each variable answer the following:

1. Is the variable measured on a nominal, ordinal or ratio scale?
2. What would be a good summary or plot to use to get an idea about this data?

Introduction to R

1. Open R Studio on your workstation and load the pacakge ggplot2. Once you have familiarised yourself with R, install R, R studio and the add-on package ggplot2 on to your own laptop.
2. Load the dataset Beer.rds which can be downloaded from Moodle.
3. Produce a histogram of the price variable. Use the function `qplot` for this.
4. Do you identify any outliers in the data?
5. Produce a cross tab of beer rating against origin. Use the function `table` for this.