

PL 9239 Assignment 1: LEGO Sets (Resit)

Please write up all your answers in an R script.

- This R script has to contain all code. The R script should be running from top to bottom in one go once the working directory is adapted to the respective machine.
- Any answers that you are adding in your own words should also go into the document. So you could for example alternate between the R script and substantive answers that you are providing.
- Please make sure that you structure your script well so that it is easy to read.

In the scenario for this labsheet, you are working on data for the company LEGO. It is your task to provide a report with an overview over the data and the first analysis of some key questions. To generate the data, I scraped the website from <https://www.lego.co.uk>. If you are lacking some contextual knowledge, feel free to browse around there to get some ideas.

1 Reading Data and Project Management (1 point)

- Set the working directory correctly to the folder path where your .r file and your data is located. Read the data. And make sure that your .r script ultimately runs in one go and produces all results without returning error messages.

2 Describing Data (1 point)

Describe one ordinal variable from the data set. Think about good measures that can help express the key characteristics of the variable:

- Can you provide an overview over the data?
- How could you offer an idea of a typical value in the data, e.g. a category that is chosen very often?
- Can you say something about the distribution of the frequencies in each category?

Show how you analyse the data using R code. Write 3-4 sentences to describe each variable.

3 Selecting from Data (2 points)

How would you select the following data? Show how you analyse the data with the R code.

- From the variable `piece_count`, select those observations that have more than 30 pieces, but less than 400.
- From the variable `vip_points`, select those observations that have less than 75 vip points or more than 920.
- Get the piece counts for all sets from the theme 'MINDSTORMS'.
- Subset the data so that it only contains sets with more than 300 pieces.

4 Visualise Data

Piece Count Data (1 point)

- Select the data for the Technic theme sets.
- Visualise the variable `piece_count` of this subset of the data with a plot type of your choice.

Show how you generate the figure using R code.

Boxplot (1 point)

- Visualise one continuous variable of your choice with a boxplot.
- Describe the visualisation in 3-4 sentences and report all statistical information that you can retrieve from a boxplot.

5 Data Analysis

Now analyse the data. To answer the questions you can use any tool you want, from looking at data in the spreadsheet, tabulating it, plotting it etc.. In all you do, it needs to be evident for the markers how you come to your conclusions, so document all the code that you are writing for your analysis.

Price per Piece (2 points)

- Make a plot with the number of pieces in a set on the x-axis and the price of a set in the y-axis. What do you conclude?
- Calculate the price per piece and analyse the most characteristic values (min, max, median, mean, standard deviation). What do you conclude?
- Now take a closer look at all sets that have a price per piece larger than 11GBP. Also consider all values that have a price per piece smaller than 0.06 GBP. Tabulating the two resulting data vectors, what do you conclude?

Price of Superheroes (2 points)

- Generate four subsets for the LEGO sets with Batman, Spider-Man, Harry Potter and Marvel.
- Calculate the average price in each of the subsets.
- What do you conclude?
- Now generate a boxplot with the price per piece for the sets with the theme Batman, Spider-Man, Harry Potter and Marvel. What do you conclude?