



Probability

March 2, 2022

The Dataset

This dataset contains the LEGO Parts/Sets/Colors and Inventories of every official LEGO set in the Rebrickable database. These files are current as of July 2017. If you need it to be more recent data, you can use Rebrickable's API which provides up to date data, and additional features. Include all of your code for each problem in your report.

1. How many observations are in the lego dataset?
2. How many variables are in the lego dataset?
3. What proportion of lego sets are 'star wars' themed?
4. Compare the average price of 'Easy' themed lego sets with the average price of 'Challenging' datasets.
5. Visualize the distribution of prices for 'star wars' themed lego sets. Write a description for the distribution that you create

Solution:

```
1
2 # load the Dataset
3 lego <- read.csv('/data/datasets/lego.csv')
4
5 # Question 1/2
6 length(lego$Price) # 12226 observations (rows)
7 length(lego) # 14 variables (columns)
8
9 # Question 3
10 sum(lego$theme=="Star Wars")/length(lego$theme)
11
12 # Question 4
13 easy <- lego$Price[lego$difficulty=="Easy"]
14 hard <- lego$Price[lego$difficulty=="Challenging"]
15 mean(easy)
16 mean(hard)
17
18 # Question 5
19 hist(lego$Price[lego$theme=="Star Wars"])
```