

NAME: LEGO Data

TYPE: Population and random sample

SIZE: Population has 1,304 observations, sample has 75 observations, and 15 variables

ARTICLE TITLE: Building your Multiple Linear Regression Model with LEGO Bricks

#### DESCRIPTIVE ABSTRACT:

We present an innovative activity that uses data about LEGO sets to help students self-discover multiple linear regression. Students are guided to predict the price of a LEGO set posted on Amazon.com (Amazon price) using LEGO characteristics such as the number of pieces, the theme (i.e., product line), and the general size of the pieces. By starting with graphical displays and simple linear regression, students are able to develop additive multiple linear regression models as well as interaction models to accomplish the task. We provide examples of student responses to the activity and suggestions for teachers based on our experiences.

#### SOURCES:

The population dataset includes LEGO sets from all themes produced between January 1, 2018 and September 11, 2020 (lego.population.csv). This population is restricted to 1,304 LEGO sets posted on Brickset (Brickset, n.d.), excluding LEGO mystery sets where the number of pieces were unknown. Data was scraped from Brickset and merged with data scraped from BrickInstructions.com (n.d.).

The sample dataset includes a subset of LEGO City, Friends, and Duplo sets produced between January 1, 2018 and September 11, 2020 (lego.sample.csv). This sample consists of 75 LEGO sets posted on Brickset (Brickset, n.d.).

#### VARIABLE DESCRIPTIONS:

Data are comma delimited. Variables are:

Item\_Number: Item number of the set scraped from brickset.com

Set\_Name: Name of the set scraped from brickset.com

Theme: Theme of the set (e.g., City, Duplo, Friends) scraped from brickset.com

Pieces: Number of pieces in the set from brickset.com.

Price: LEGO recommended price of the set from brickset.com.

Amazon\_Price: Amazon price of the set scraped from brickset.com

Year: The year the set was produced scraped from brickset.com

Ages: LEGO set age recommendation from brickset.com

Pages: Number of pages in the set instructions manual scraped from  
lego.brickinstructions.com

Minifigures: Number of minifigures (LEGO people) in the set scraped from brickset.com.  
LEGO sets with no minifigures have been coded as NA. NA's also represent missing data. This is due to how brickset.com reports their data.

Packaging: Type of packaging (Blister pack, Box, Box with handle, Foil pack, Plastic box, Polybag, Shrink-wrapped, Other) scraped from brickset.com

Weight: Weight of the LEGO set in kg and lbs scraped from brickset.com

Unique\_Pieces: Number of unique pieces in the set recorded in the instructions manual scraped from brickset.com

Availability: Type of availability of the set (e.g., LEGOLAND exclusive, Promotional, Retail) scraped from brickset.com

Size: General size of the interlocking bricks (Large = LEGO Duplo sets which include large brick pieces safe for children ages 1 to 5, Small = LEGO sets which include the traditional smaller brick pieces created for age groups 5 and older, e.g., City, Friends)

Disclaimers: LEGO® is a trademark of the LEGO Group of companies which does not sponsor, authorize or endorse this paper. Amazon is a trademark of the Amazon Services LLC which does not sponsor, authorize or endorse this paper.

#### REFERENCES:

BrickInstructions.com. (n.d.). Retrieved February 2, 2021 from [lego.brickinstructions.com](http://lego.brickinstructions.com)

Brickset. (n.d.). *BRICKSET: Your LEGO® set guide*. Retrieved February 2, 2021 from <https://brickset.com/>

#### SUBMITTED BY:

Anna D. Peterson  
Department of Statistics  
Iowa State University  
3115 Snedecor Hall  
Ames, IA 50011-1210  
[ericksad@iastate.edu](mailto:ericksad@iastate.edu)

Laura Ziegler  
Department of Statistics  
Iowa State University  
3409 Snedecor Hall  
Ames, IA 50011-1210  
[lziegler@iastate.edu](mailto:lziegler@iastate.edu)