

# Brick by Brick; A look at the Lego pricing model

Elisabeth Starr Snider

## Introduction:

This project attempts to analyse the factors that go into the pricing model used by the Lego company. It then aims to predict the price of Lego sets based on that factor analysis.

This data set was created in 2018 by scraping the Lego website for product information. It was downloaded from Kaggle.com where it is hosted.

## The Data Set:

Let's take a look at the actual data set.

```
legos<-read.csv("lego_sets.csv")
str(legos)
```

```
## 'data.frame':   12261 obs. of  14 variables:
## $ ages          : chr  "6-12" "6-12" "6-12" "12+" ...
## $ list_price     : num  30 20 13 100 80 ...
## $ num_reviews    : num  2 2 11 23 14 7 37 24 23 11 ...
## $ piece_count    : num  277 168 74 1032 744 ...
## $ play_star_rating : num  4 4 4.3 3.6 3.2 3.7 3.7 4.4 3.6 3.6 ...
## $ prod_desc      : chr  "Catapult into action and take back the eggs from the Piggy Trike!" "Laun
## $ prod_id        : num  75823 75822 75821 21030 21035 ...
## $ prod_long_desc  : chr  "Use the staircase catapult to launch Red into the air and race after the
## $ review_difficulty: chr  "Average" "Easy" "Easy" "Average" ...
## $ set_name       : chr  "Bird Island Egg Heist" "Piggy Plane Attack" "Piggy Car Escape" "United S
## $ star_rating     : num  4.5 5 4.3 4.6 4.6 4.9 4.2 4.7 4.7 4.8 ...
## $ theme_name     : chr  "Angry Birds" "Angry Birds" "Angry Birds" "Architecture" ...
## $ val_star_rating : num  4 4 4.1 4.3 4.1 4.4 4.1 4.3 4.1 4.5 ...
## $ country        : chr  "US" "US" "US" "US" ...
```

### Initial Observations:

There are 12,261 observations and 14 variables in this data set. Some of the variables include lengthy descriptions. This is significantly more data than is needed to answer questions about the Lego pricing model so the data set needs to be evaluated in terms of the most important variables.

The variables of most interest in understanding and predicting `list_price` are `piece_count`, `theme_name`, and `review_difficulty`. The data set includes markets from 21 countries, but this project will only include only observations from the US market in this analysis.

Additionally, creating a `price_per_piece` feature will contribute to clarity in evaluating the impact of licensing on price.

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.4      v dplyr  1.0.7
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   2.0.1      v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

lego_data<-legos%>%
  mutate(price_per_piece = list_price/piece_count)%>%
  filter(country == "US")%>%
  select(list_price, piece_count, price_per_piece, set_name, theme_name, review_difficulty)
```

## Vizualizations:

This project is exploring three factors as they relate to the price of lego sets. Those factors are licensing (theme\_name), level of complexity (review\_difficulty), and the number of pieces per set (piece\_count).

### First let's look at licensing:

In order to get this data ready for analysis some work needed to be done first. The MindStorm line of Lego sets is based in the idea of building robots using lego pieces. So some of the pieces included in these sets were tiny computers. This heavily skewed the price per piece data as some single MindStorm pieces cost over 100 dollars. These needed to be removed before an accurate analysis could take place.

```
lego_data%>%
  filter(price_per_piece < .4)%>%
  ggplot() +
  geom_point(mapping = aes(x = piece_count, y = price_per_piece), color = "yellow", size = .8) +
  facet_wrap(~ theme_name, nrow = 7)+
  theme_dark()+
  xlab("Number of Pieces")+
  ylab("Price Per Piece")
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Angry Birds' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Angry Birds' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Angry Birds' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DC Comics Super Heroes' in 'mbcsToSbcs': dot
## substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DC Comics Super Heroes' in 'mbcsToSbcs': dot
## substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DC Comics Super Heroes' in 'mbcsToSbcs': dot
## substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DIMENSIONS' in 'mbcsToSbcs': dot substituted for <e2>
```

```

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DIMENSIONS' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DIMENSIONS' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Disney' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Disney' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Disney' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ghostbusters' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ghostbusters' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ghostbusters' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Minecraft' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Minecraft' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Minecraft' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'NEXO KNIGHTS' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'NEXO KNIGHTS' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'NEXO KNIGHTS' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Star Wars' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Star Wars' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Star Wars' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'THE LEGO® NINJAGO® MOVIE' in 'mbcsToSbcs': dot
## substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'THE LEGO® NINJAGO® MOVIE' in 'mbcsToSbcs': dot
## substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'THE LEGO® NINJAGO® MOVIE' in 'mbcsToSbcs': dot
## substituted for <a2>

```

```

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Angry Birds' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Angry Birds' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Angry Birds' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DC Comics Super Heroes' in 'mbcsToSbcs': dot
## substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DC Comics Super Heroes' in 'mbcsToSbcs': dot
## substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DC Comics Super Heroes' in 'mbcsToSbcs': dot
## substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DIMENSIONS' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DIMENSIONS' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DIMENSIONS' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Disney' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Disney' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Disney' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ghostbusters' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ghostbusters' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ghostbusters' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Minecraft' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Minecraft' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Minecraft' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'NEXO KNIGHTS' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'NEXO KNIGHTS' in 'mbcsToSbcs': dot substituted for <84>

```

```

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'NEXO KNIGHTS' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Star Wars' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Star Wars' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Star Wars' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'THE LEGO® NINJAGO® MOVIE' in 'mbcsToSbcs': dot
## substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'THE LEGO® NINJAGO® MOVIE' in 'mbcsToSbcs': dot
## substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'THE LEGO® NINJAGO® MOVIE' in 'mbcsToSbcs': dot
## substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Angry Birds' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Angry Birds' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Angry Birds' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DC Comics Super Heroes' in 'mbcsToSbcs': dot
## substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DC Comics Super Heroes' in 'mbcsToSbcs': dot
## substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DC Comics Super Heroes' in 'mbcsToSbcs': dot
## substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DIMENSIONS' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DIMENSIONS' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DIMENSIONS' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Disney' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Disney' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Disney' in 'mbcsToSbcs': dot substituted for <a2>

```

```

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ghostbusters' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ghostbusters' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ghostbusters' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Minecraft' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Minecraft' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Minecraft' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'NEXO KNIGHTS' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'NEXO KNIGHTS' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'NEXO KNIGHTS' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Star Wars' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Star Wars' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Star Wars' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'THE LEGO® NINJAGO® MOVIE' in 'mbcsToSbcs': dot
## substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'THE LEGO® NINJAGO® MOVIE' in 'mbcsToSbcs': dot
## substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'THE LEGO® NINJAGO® MOVIE' in 'mbcsToSbcs': dot
## substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Angry Birds' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Angry Birds' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Angry Birds' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DC Comics Super Heroes' in 'mbcsToSbcs': dot
## substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DC Comics Super Heroes' in 'mbcsToSbcs': dot

```

```

## substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DC Comics Super Heroes' in 'mbcsToSbcs': dot
## substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DIMENSIONS' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DIMENSIONS' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DIMENSIONS' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Disney' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Disney' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Disney' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ghostbusters' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ghostbusters' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ghostbusters' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Minecraft' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Minecraft' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Minecraft' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'NEXO KNIGHTS' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'NEXO KNIGHTS' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'NEXO KNIGHTS' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Star Wars' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Star Wars' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Star Wars' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'THE LEGO® NINJAGO® MOVIE' in 'mbcsToSbcs': dot
## substituted for <e2>

```

```

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'THE LEGO® NINJAGO® MOVIE' in 'mbcsToSbcs': dot
## substituted for <84>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'THE LEGO® NINJAGO® MOVIE' in 'mbcsToSbcs': dot
## substituted for <a2>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Star Wars' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Star Wars' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Star Wars' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'THE LEGO® NINJAGO® MOVIE' in 'mbcsToSbcs': dot
## substituted for <e2>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'THE LEGO® NINJAGO® MOVIE' in 'mbcsToSbcs': dot
## substituted for <84>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'THE LEGO® NINJAGO® MOVIE' in 'mbcsToSbcs': dot
## substituted for <a2>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Minecraft' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Minecraft' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Minecraft' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'NEXO KNIGHTS' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'NEXO KNIGHTS' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'NEXO KNIGHTS' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ghostbusters' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ghostbusters' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ghostbusters' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DIMENSIONS' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DIMENSIONS' in 'mbcsToSbcs': dot substituted for <84>

```



```

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'DIMENSIONS' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Disney' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Disney' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Disney' in 'mbcsToSbcs': dot substituted for <a2>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x,
## x$y, : conversion failure on 'DC Comics Super Heroes' in 'mbcsToSbcs': dot
## substituted for <e2>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x,
## x$y, : conversion failure on 'DC Comics Super Heroes' in 'mbcsToSbcs': dot
## substituted for <84>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x,
## x$y, : conversion failure on 'DC Comics Super Heroes' in 'mbcsToSbcs': dot
## substituted for <a2>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Angry Birds' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Angry Birds' in 'mbcsToSbcs': dot substituted for <84>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Angry Birds' in 'mbcsToSbcs': dot substituted for <a2>

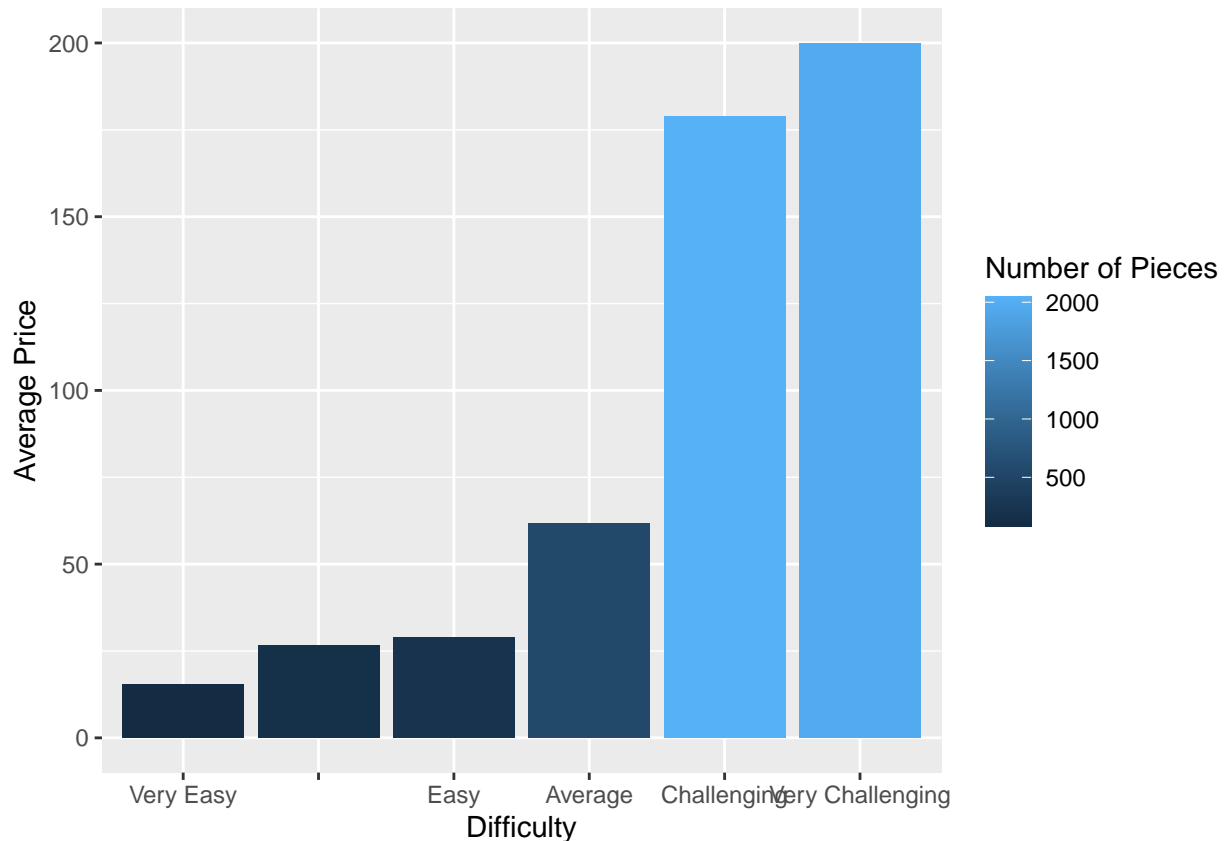
```



If licensing was a significant factor in the Lego pricing model, it would be clear because licensed lego sets, such as Ninjago, Minecraft, and Batman would have a higher price per piece. We do not see that at all. The average price per piece is pretty similar across the board when comparing licensed sets and Lego classic sets. Star Wars sets appear to be somewhat of an outlier in that there is a large spread of price per piece across all Star Wars sets.

Next, let's look at level of difficulty:

```
lego_data%>%
  group_by(review_difficulty)%>%
  summarise(avg_price = mean(list_price), avg_pieces = mean(piece_count))%>%
  ggplot() +
  geom_col(aes(fct_reorder(review_difficulty, avg_price), avg_price, fill = avg_pieces))+
  scale_fill_continuous(name = "Number of Pieces")+
  xlab("Difficulty")+
  ylab("Average Price")
```

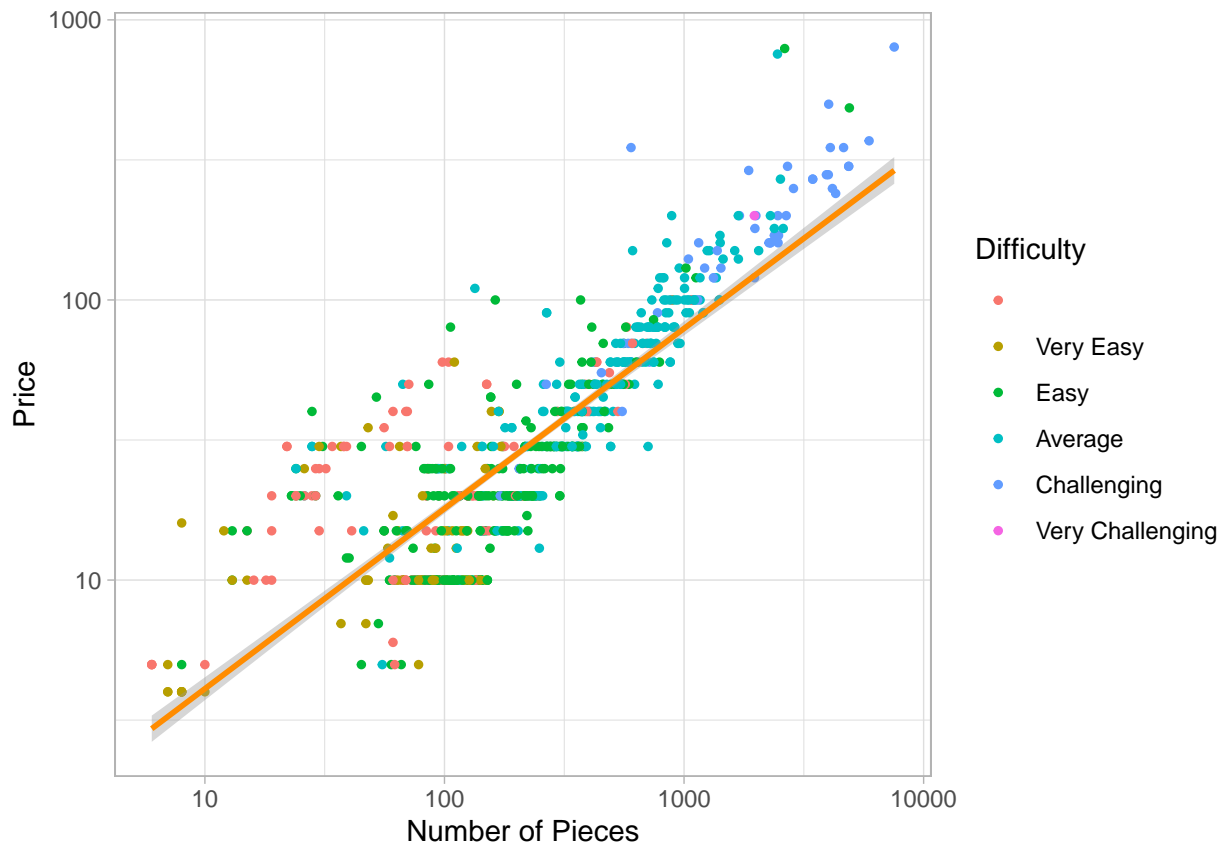


If level of difficulty was a significant factor in price, due to the cost of engineering a more complex set, the graph of the average price would increase as level of difficulty increases. In order to test this hypothesis, the summarise function was used to create the features for average number of pieces per difficulty level and average price per difficulty level. It is clear that average price increases as level of difficulty increases. However, when the number of pieces per difficulty level is taken into consideration it would appear that it is tricky to disentangle price as function of number of pieces and price as function of difficulty.

**So let's see what the relationship is between price and number of pieces:**

```
lego_data%>%
  filter(!price_per_piece > 2)%>%
  ggplot(aes(x=piece_count,y=list_price, color = fct_reorder(review_difficulty, piece_count)))+
  geom_point(size = 1)+
  theme_light()+
  stat_smooth(aes(x=piece_count,y=list_price),method="lm", color="dark orange")+
  scale_color_discrete(name = "Difficulty")+
  scale_x_log10()+
  scale_y_log10()+
  xlab("Number of Pieces")+
  ylab("Price")

## `geom_smooth()` using formula 'y ~ x'
```



This plot of price as function of the number of pieces in a set shows the strong linear relationship between the two variables. The color code by level of difficulty shows that while there is some correlation between price and difficulty, it is not the primary variable in the pricing model. It is more likely that complexity is a function of the number of pieces in a set.

## Linear Model and Prediction:

In exploring the pricing model for Legos, the question was raised “Are legos getting more expensive?” The first Lego sets were released to the US market in 1960. They had 100 pieces each and cost 5.65. When adjusting for inflation, that is about 52.95 in today’s dollars.

In order to better understand how pricing has changed over the years, a linear model was created, along with an observation with variables that matched those of the original Lego sets.

```
price_model<-lm(list_price~piece_count,lego_data)
summary(price_model)
```

```
##
## Call:
## lm(formula = list_price ~ piece_count, data = lego_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -148.56   -9.26   -5.38    4.28   553.47
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  11.114489   1.465318   7.585 9.07e-14 ***
```

```
## piece_count 0.085672 0.001749 48.980 < 2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 36.19 on 815 degrees of freedom
## Multiple R-squared: 0.7464, Adjusted R-squared: 0.7461
## F-statistic: 2399 on 1 and 815 DF, p-value: < 2.2e-16

single_obs<-data_frame(piece_count = 100, review_difficulty = "Average", theme_name = "Classic")

## Warning: `data_frame()` was deprecated in tibble 1.1.0.
## Please use `tibble()` instead.

This observation, which was created to be as close as possible to the parameters of the original Lego sets, was
then used to predict the cost of an original lego set if it was produced today.

predict(price_model,single_obs)

##          1
## 19.68174
```

According to this model, the cost of the first lego sets today would be 19.68, which is much less than 52.95! To check this model against actual data a person can go to Lego.com and search for a Classic Lego set with 100-200 pieces and the sets that it returns cost \$19.99, which means that this prediction is extremely accurate and a good match for real world data.

## Conclusions:

One of the most interesting insights gained from analyzing this data is that the cost of a Lego set is heavily dependent on how many pieces are included in the set. The average price of a Lego piece does not vary much when outliers (like MindStorm sets) are removed. So this is a reliable way to predict the price of a Lego set when the number of pieces is known.

Using our prediction model we were also able to conclude that the cost of Legos has actually decreased in the last 50 years, when accounting for inflation. Lego sets consisting of 100 pieces first retailed for around 50 (inflation adjusted) US dollars in 1960. But the same set today costs about 20 US dollars.

### Further Study:

#### Interesting questions to explore in future analysis:

Are more expensive Lego sets rated higher than less expensive Lego sets in reviews?

Do prices and/or licensing franchises vary widely between international markets?

Is there a difference in complexity and/or price between Lego sets marketed to girls and sets marketed to boys?

Are licensed Lego sets reviewed as highly as Lego originals?