Cost Difference of Gendered Lego Sets

Rachel Lawson, Raley Long, Amelia Tarno, Mattea Whitlow

Cost Difference of Gendered Lego Sets

Rachel Lawson, Raley Long, Amelia Tarno, Matea Whitlow

Proposal

In this project, we will be exploring the different prices and number of pieces per set of Lego's that have a specific gender as their target audience, looking at the difference between the binary genders male and female, and between gendered sets and sets that are deemed neutral. We will be making our hypotheses based on long-standing societal norms about toys aimed towards boys versus girls. Our main hypothesis is that sets whose target audience is girls will have, on average, a higher price than sets geared towards boys. A secondary hypothesis is that sets that have a specific gender audience will have higher prices than sets with a gender neutral audience. A tertiary hypothesis is that sets aimed at girls will have less pieces per set, on average, than sets that are neutral or aimed at boys.

Data

The data we will be using can be found here, and comes from

Outcome Variable

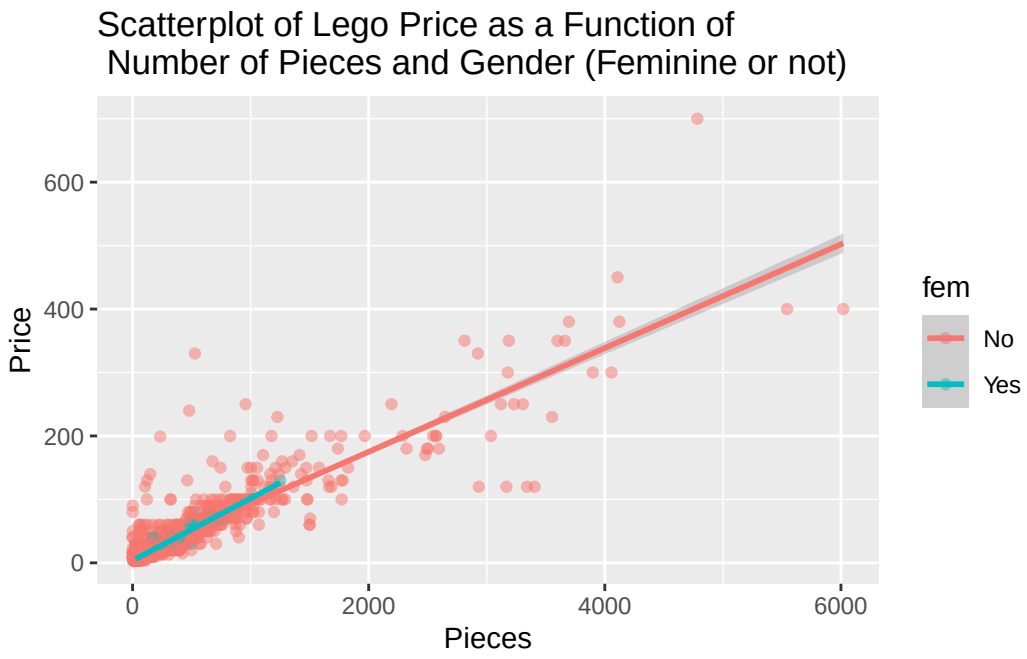
The response variable will be the price of Lego sets, measured in US dollars. This is a numeric variable with values from 1.99 to 699.99.

Explanatory Variables

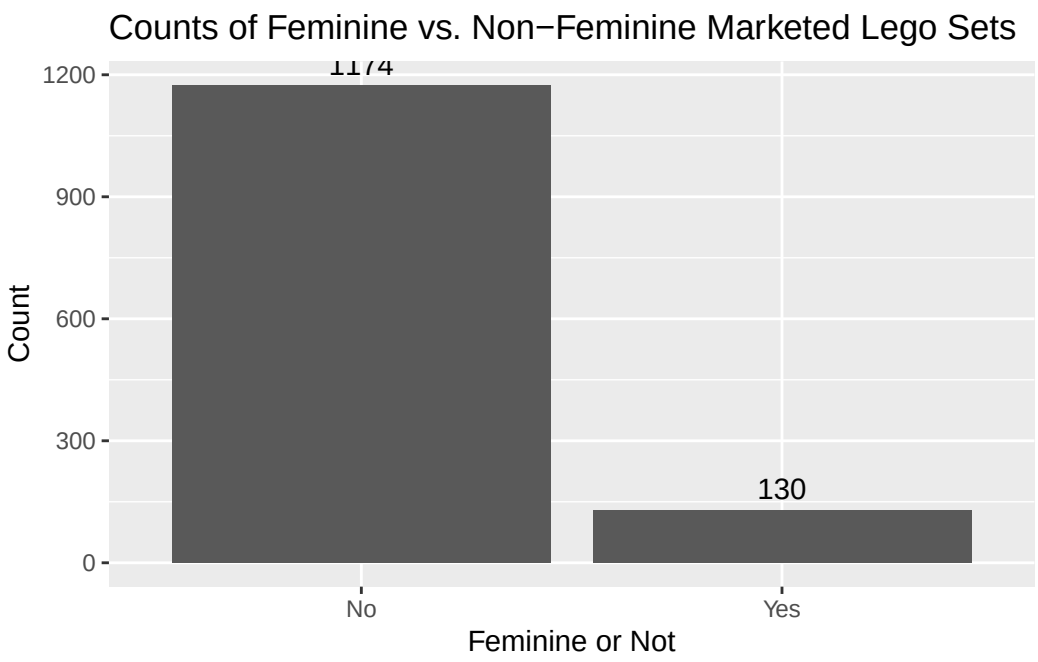
The explanatory variables will be pieces per set, a numerical variable measured in integers ranging from 1 to 6020, and filtered by gender which will be divided into two categorical variables taking a value of 0 or 1. One variable will be girl, for which an observation will be taking a value of 1 if it is for an audience of girls, or 0 if it is not. Another variable will be boys, for which an observation will take a value of 1 if it is for an audience of boys or 0 if it is not. This way, if neither variable is "on", the reference category will be gender neutral.

Exploratory Visualizations

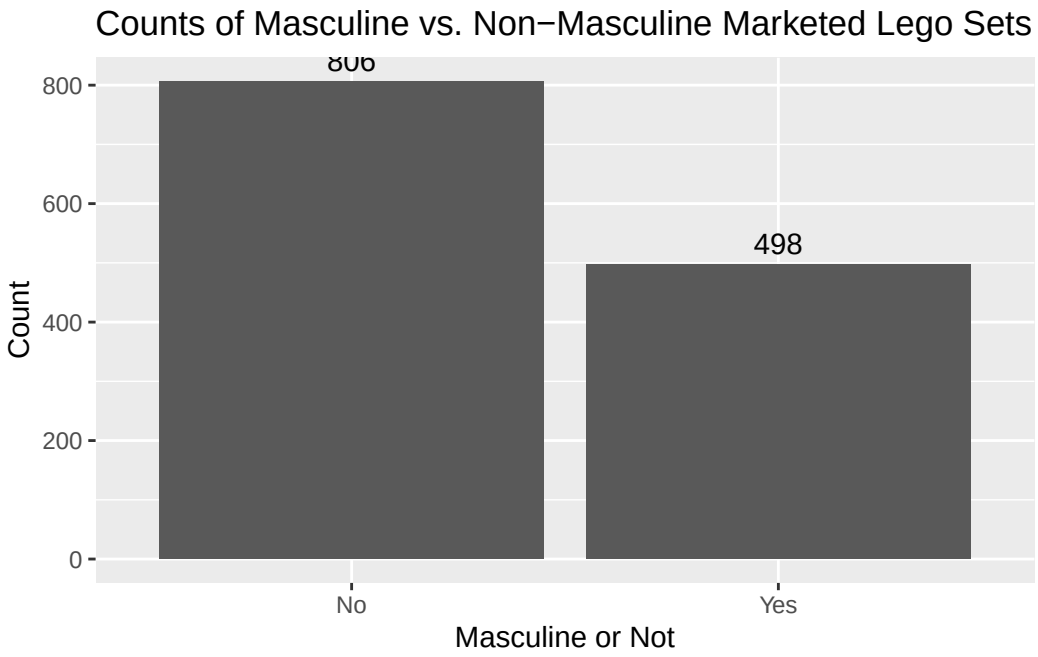
```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```



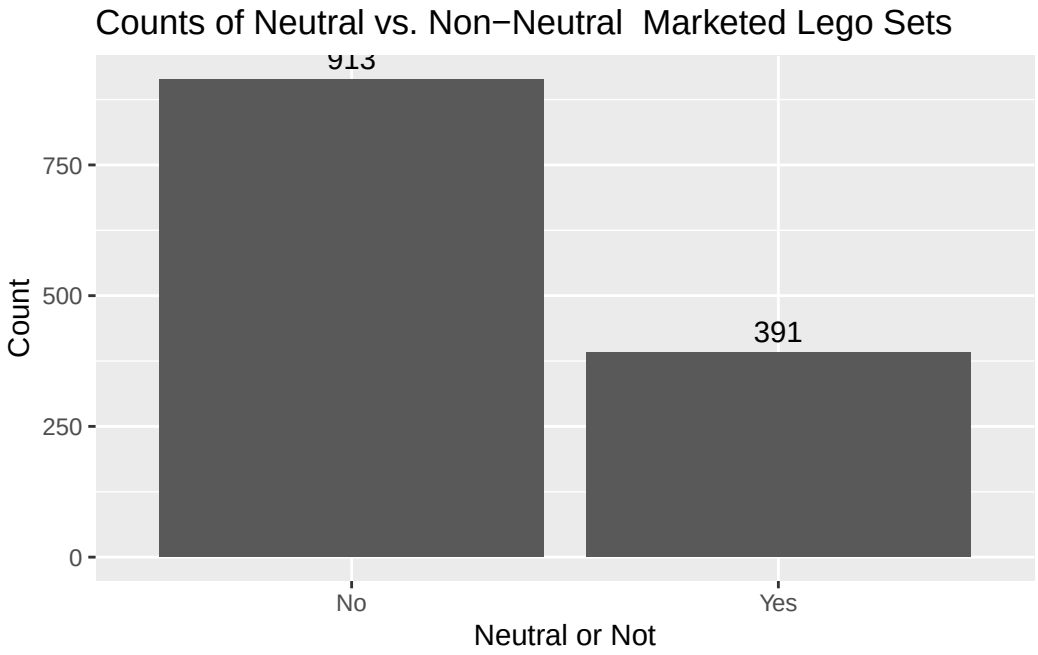
```
ggplot(lego_clean, aes(x = fem)) + geom_bar() + geom_text(stat = 'count', aes(label=after_
```



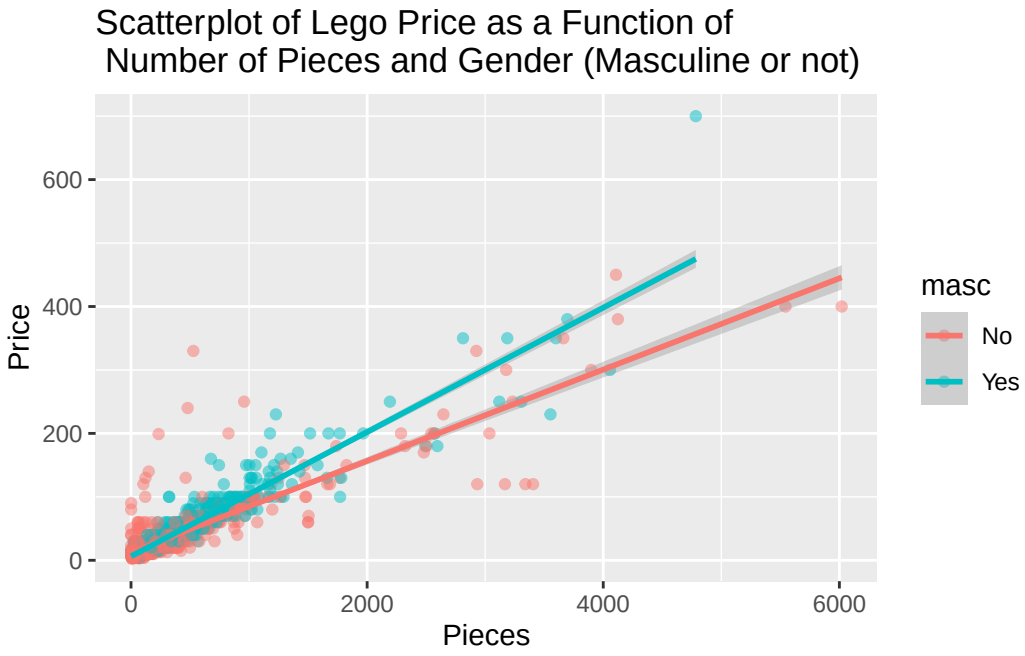

```
ggplot(lego_clean, aes(x = masc)) + geom_bar() + geom_text(stat = 'count', aes(label=after
```



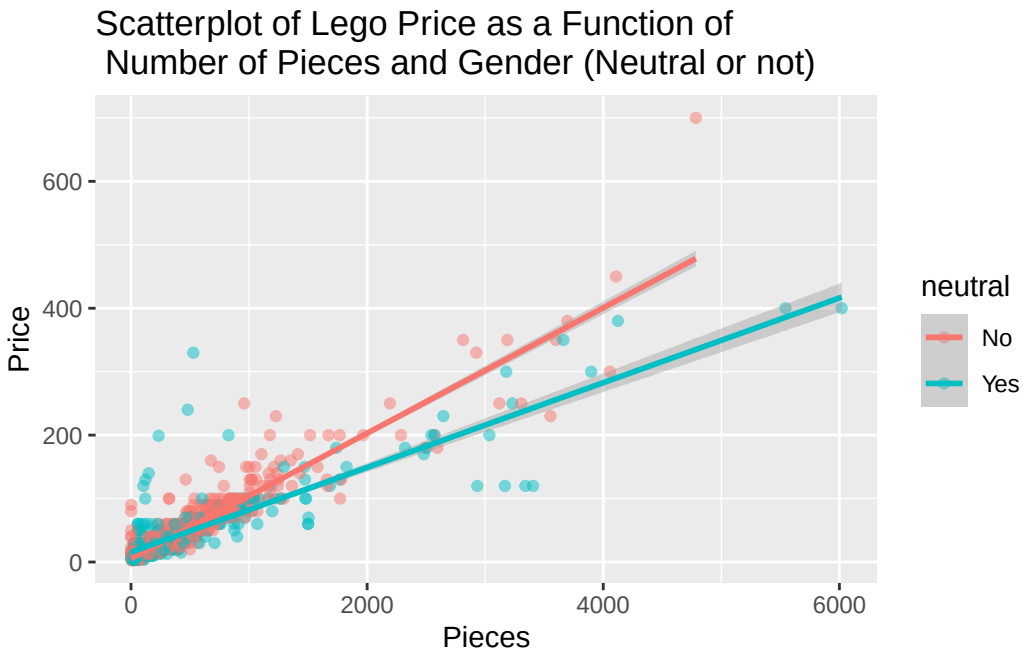

```
ggplot(lego_clean, aes(x = neutral)) + geom_bar() + geom_text(stat = 'count', aes(label=af
```



```
price_piece_fem <- lm(Price ~ Pieces + masc, lego_clean)
ggplot(lego_clean, aes( x = Pieces, y = Price, color = masc)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "lm") +
  labs(title = "Scatterplot of Lego Price as a Function of \n Number of Pieces and Gender
```




```
price_piece_fem <- lm(Price ~ Pieces + neutral, lego_clean)
ggplot(lego_clean, aes( x = Pieces, y = Price, color = neutral)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "lm") +
  labs(title = "Scatterplot of Lego Price as a Function of \n Number of Pieces and Gender
```


Scatterplot of Lego Price as a Function of
Number of Pieces and Gender (Neutral or not)
Price
600
400
200
0
0
2000
4000
6000
Pieces
neutral
No
Yes