

**Title:** LAB Group Project

**Author:** Benjy Ross, Joe Myers, and Shirley Zhao

**Date:** “12/10/2021”

## Introduction

Mario Kart Wii is a 2008 car racing game developed and published by the Japanese video game company Nintendo. This video game was specifically designed for play on the Wii home console, which creatively employs uniquely designed controllers that use motion sensors for much of the input to the console. Mario Kart Wii takes advantage of this by using the motion controls on the controller to simulate driving an actual car. For example, to steer one’s character in the video game, they must physically turn the remote to one side or the other. Alone, using the rectangular shaped Wii remote to do this turning may feel awkward. To assist with this, Nintendo released plastic wheel shaped covers for the Wii remotes to better emulate an actual steering wheel and to relieve the awkwardness of steering in the game.

eBay is an online auction site where anyone can sell goods. These auctions, like any other auctions, are bid on by others who have an account on the website. The price of these goods, therefore, is controlled by the bidding process, and the good sells to the person with the highest bid (Bajari & Hortagsu, 2003).

In recent years, Mario Kart Wii has made a resurgence into the gaming world as people want to revisit their nostalgia with the game they played when they were younger. However, Nintendo does not produce any new copies of Mario Kart Wii anymore, and therefore if one wants to own a copy of Mario Kart Wii they must turn to the second hand market, and eBay is an excellent and convenient choice.

Though many copies of Mario Kart Wii that sell on eBay don’t include any of the wheel covers, many in fact do, and in numbers. Since the max amount of players that can play Mario Kart Wii together at one time is four, the numerical range of owning remotes that actually have utility to the owner is 0 to 4. In this study, we wanted to see how including wheel covers influences the winning bidding price of Mario Kart Wii eBay auctions. With that being said, our hypothesis follows:

$H_O$ : The number of wheels does not affect the adjusted price or the adjusted price would be lower.

$H_A$ : As the number of wheels increases, the adjusted price would be higher.

The data used in this study is from the OpenIntro package for r. The data is Mario Kart Wii auctions sales on eBay during the month of October 2009. The variable Price refers to the final winning bid of that copy of Mario Kart Wii in US dollars, not including shipping, which has been excluded from this study as shipping increased the amount of variance in prices and did not reflect the actual contents of the auction sale. The variable Wheels refer to the amount of wheels included in the auction sale. Two data points were removed from the data, as they included the Wii console alongside Mario Kart Wii in the sale, and therefore did not accurately reflect the price of just the game and wheels. Sample population is  $n = 143$  before corrections,  $n = 141$  after.

## Data Exploration

```
## # A tibble: 143 x 12
##       id duration n_bids cond  start_pr ship_pr total_pr ship_sp seller_rate
##   <dbl>   <int>   <int> <fct>   <dbl>   <dbl>   <dbl> <fct>   <int>
## 1  1.50e11     3    20 new     0.99     4     51.6 standa~    1580
## 2  2.60e11     7    13 used     0.99    3.99    37.0 firstC~     365
## 3  3.20e11     3    16 new     0.99    3.5     45.5 firstC~     998
## 4  2.80e11     3    18 new     0.99     0     44 standa~       7
## 5  1.70e11     1    20 new     0.01     0     71 media      820
## 6  3.60e11     3    19 new     0.99     4     45 standa~   270144
## 7  1.20e11     1    13 used     0.01     0    37.0 standa~   7284
## 8  3.00e11     1    15 new     1       2.99    54.0 upsGro~   4858
## 9  2.00e11     3    29 used     0.99     4     47 priori~     27
## 10 3.30e11     7     8 used    20.0     4     50 firstC~    201
## # ... with 133 more rows, and 3 more variables: stock_photo <fct>,
## #   wheels <int>, title <fct>
```

The data frame is “Wii Mario Kart auctions from eBay”, and it comes from the course textbook, “OpenIntro: Introduction to Modern Statistics (2021) by Mine Çetinkaya-Rundel and Johanna Hardin, First Edition.” The sampling method is Total population sampling, which this data set includes all auctions for a full week in October 2009. The descriptions for each variables are shown below.

ID(sample population)

Duration(Description: Auction Length, in days; Numerical variable)

n-bids(Description: Number of bids; Numerical variable)

cond(Description: Game condition, either new or used; Categorical variable)

star\_pr(Description: Start price of the auction; Numerical variable)

ship\_pr(Description: Shipping price; Numerical variable)

total\_pr(Description: Total price, which equals the auction price plus the shipping price; Numerical variable)

ship\_sp(Description: Shipping speed or method; Categorical variable)

seller\_rate(Description: The seller’s rating on eBay; This is the number of positive ratings minus the number of negative ratings for the seller; Numerical variable)

stock\_photo(Description: Whether the auction feature photos was a stock photo or not.If the picture was used in many auctions, then it was called a stock photo; Categorical variable)

wheels(Description: Number of Wii wheels included in the auction. These are steering wheel attachments to make it seem as though you are actually driving in the game. When used with the controller, turning the wheel actually causes the character on screen to turn; Categorical variable, but also can be a numerical variable)

After listed all the variables from the the dataset, this research presents the data exploration with multiple types of graph in the followings, and for each graph, a short description is provided.

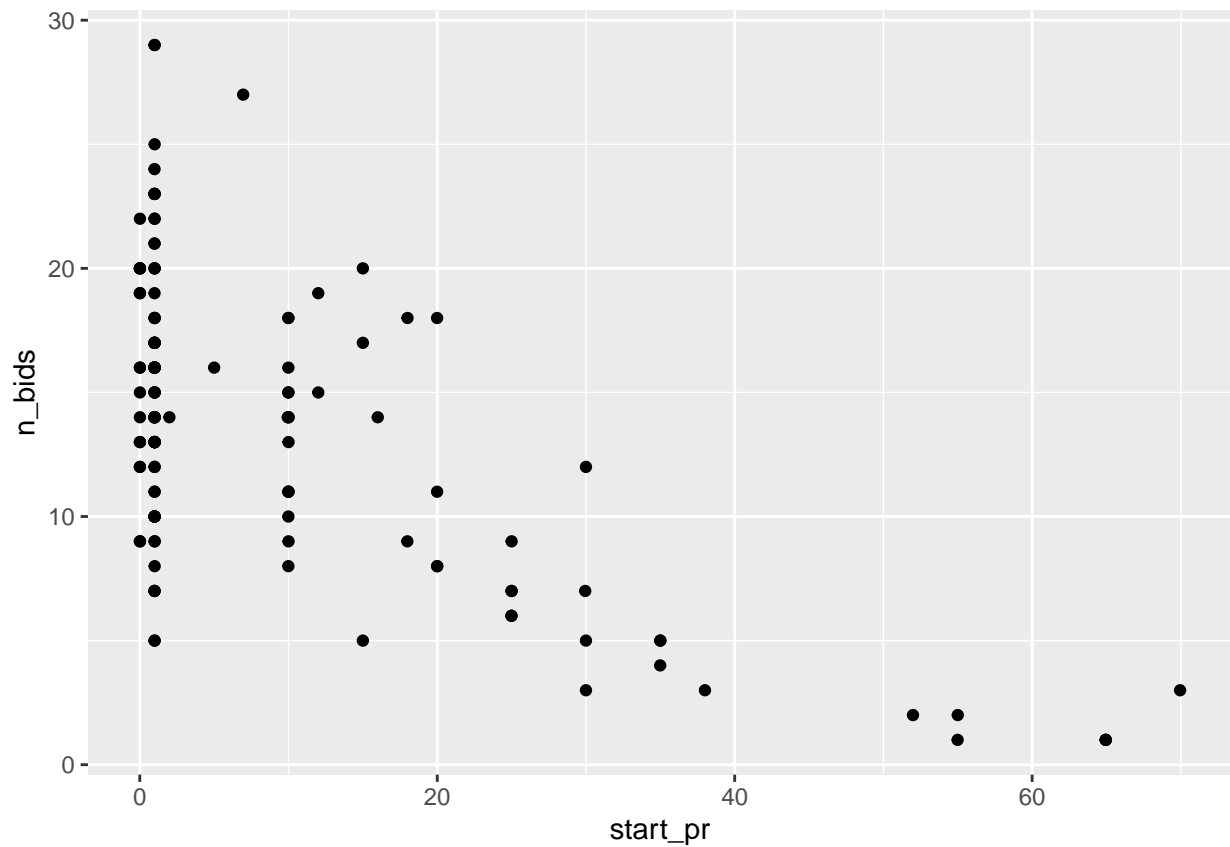
First, this research needed to filter out the “NA”s to be able to properly synthesize the data into visualization graphs.

```
##       duration      n_bids      start_pr      ship_pr      total_pr
## duration  1.00000000 -0.12102397  0.13472133  0.26650133 -0.041235447
## n_bids    -0.12102397  1.00000000 -0.63310360  0.02633711  0.127729402
## start_pr   0.13472133 -0.63310360  1.00000000  0.02999144  0.073406030
## ship_pr    0.26650133  0.02633711  0.02999144  1.00000000  0.542508581
## total_pr  -0.04123545  0.12772940  0.07340603  0.54250858  1.000000000
```

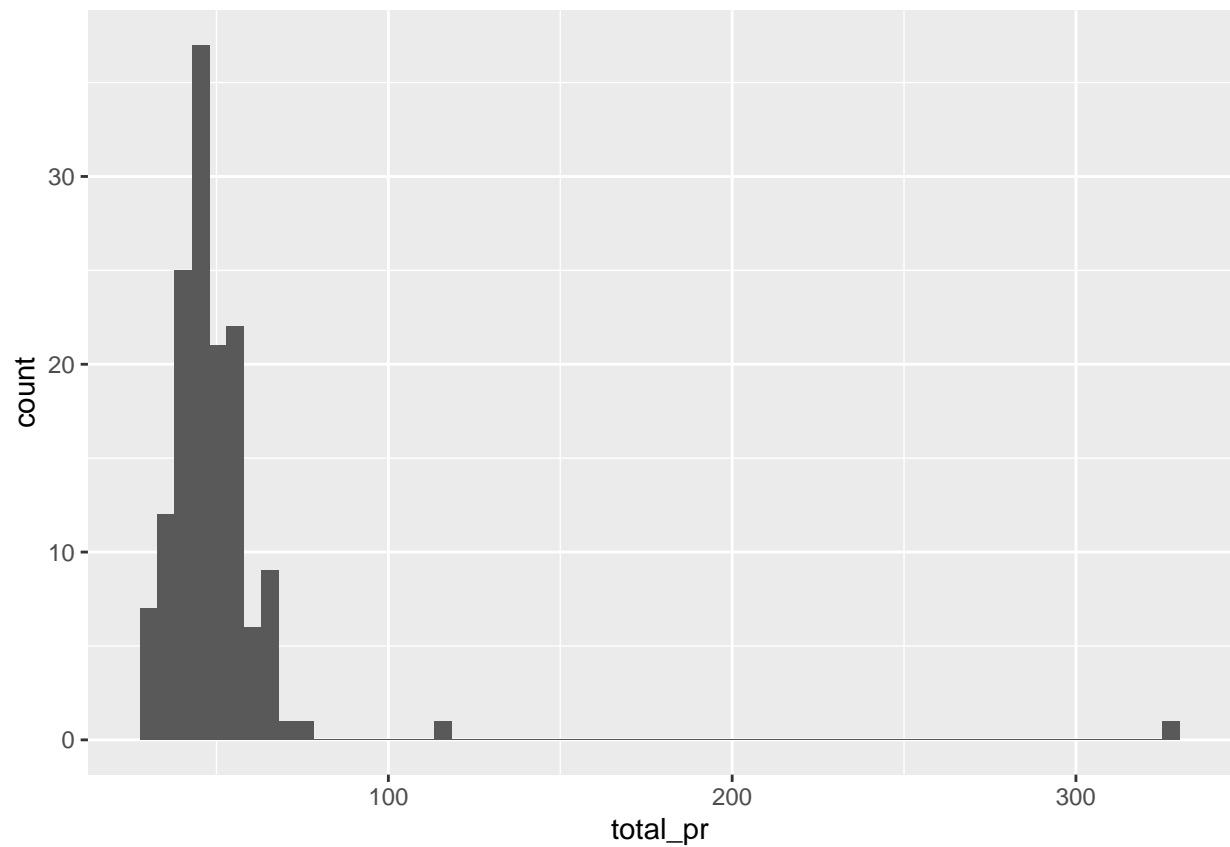
```

## seller_rate -0.15483297 -0.10615650 0.28132109 -0.02329238 0.009434813
## wheels      -0.29947345 -0.07961859 0.16368556 0.05228926 0.329983751
##            seller_rate      wheels
## duration    -0.154832970 -0.29947345
## n_bids       -0.106156497 -0.07961859
## start_pr     0.281321092 0.16368556
## ship_pr      -0.023292381 0.05228926
## total_pr     0.009434813 0.32998375
## seller_rate  1.000000000 -0.14576974
## wheels       -0.145769743 1.000000000

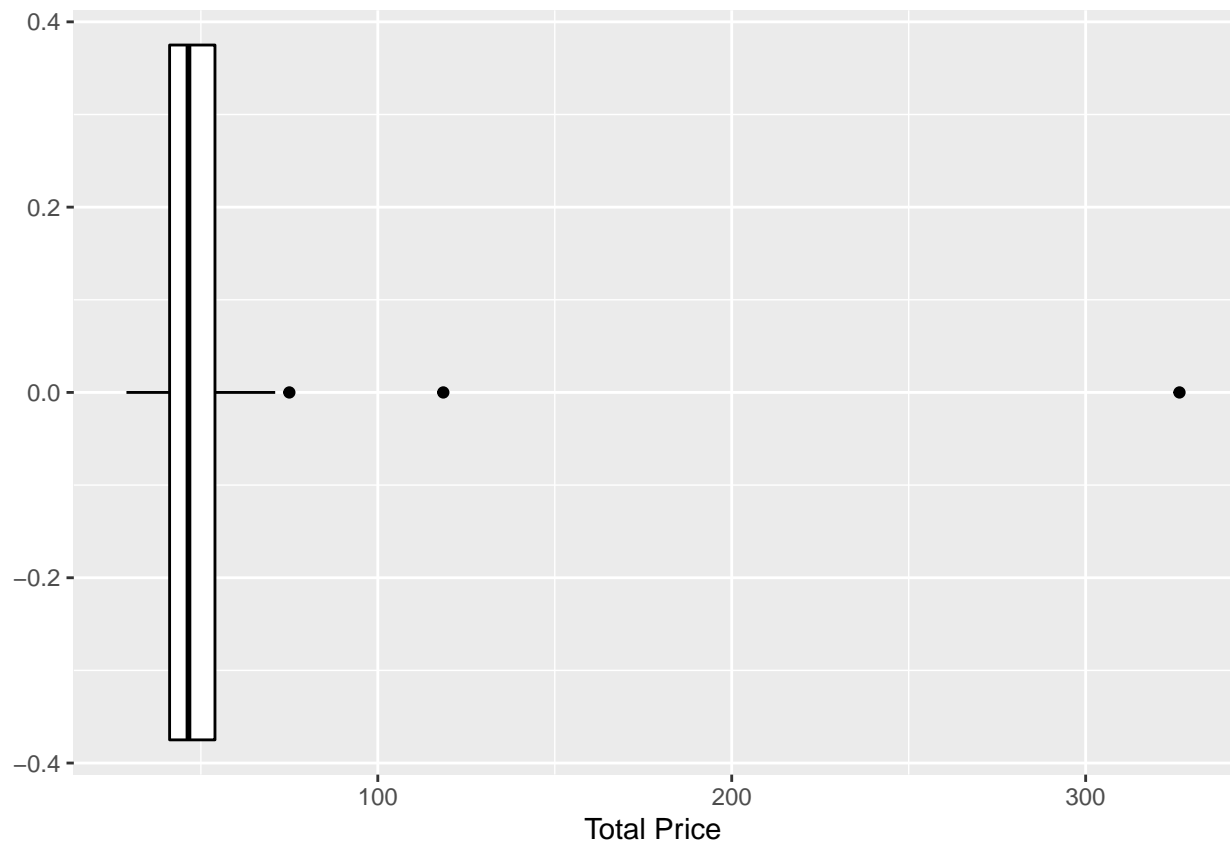
```



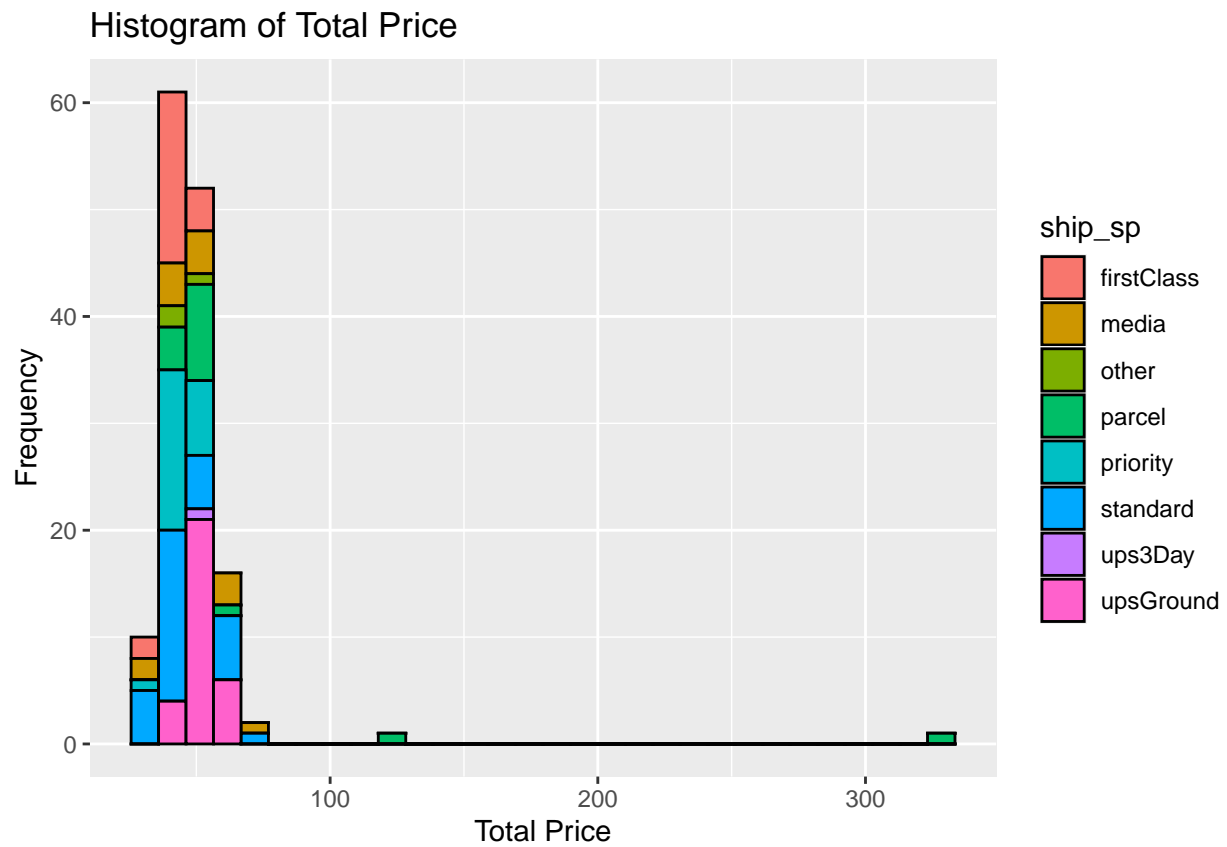
This scatter plot shows the relationships between the start price and the number of bids. This relationship will not be explored further since they did not appear to be a strong corollary relationship between the two variables.



The above histogram shows the distribution of total price of the karts. Also shows two outliers in this dataset. In the further exploration, a new variable called adjusted price will be introduced and replace total price.



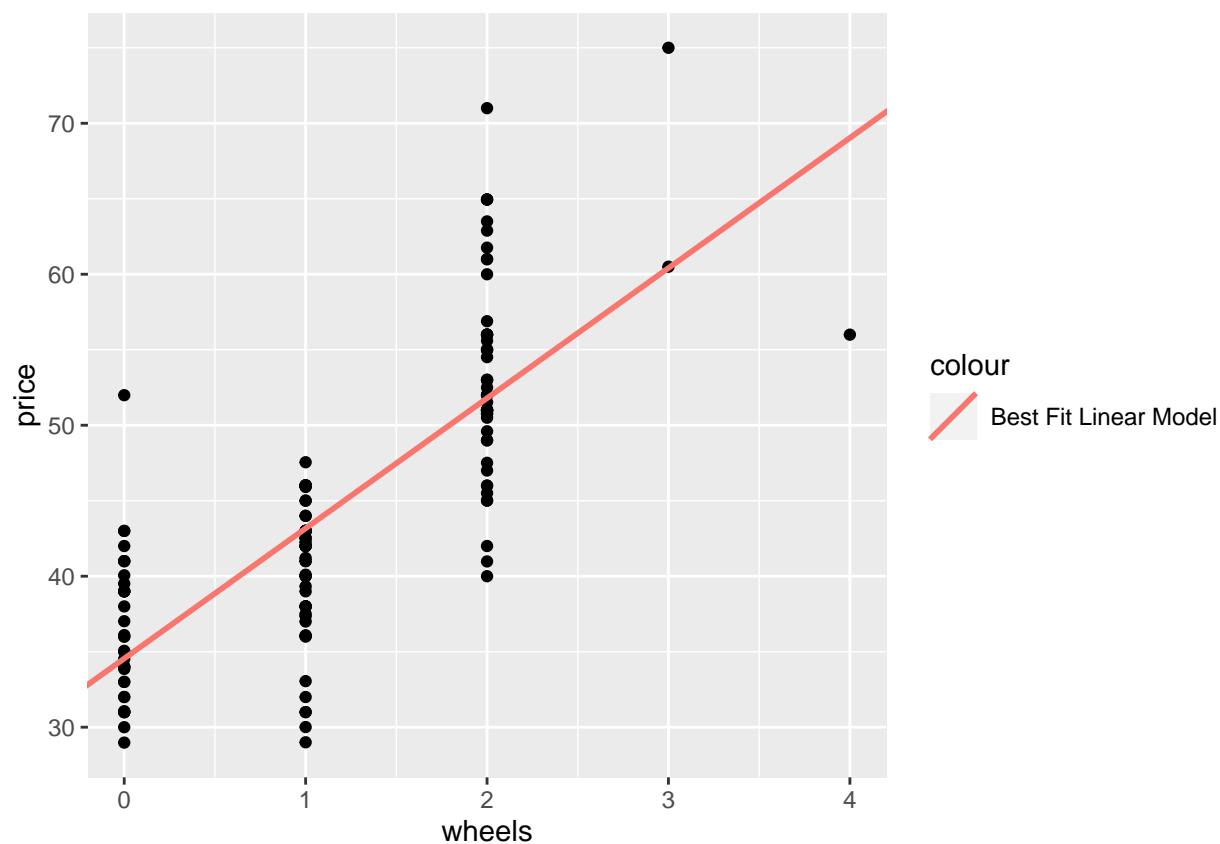
This graph shows another way to visualize the distribution of the total price. Also, it shows the median, 1st quarter, and 3rd quarter positions of the entire dataset.



This graph shows the total price based on different shipping methods. In the future discussions of the data, the factor of the shipping and shipping price will be excluded.

To better find the correlation between the wheels and the prices, this research introduced a new variable called “adjusted price”:

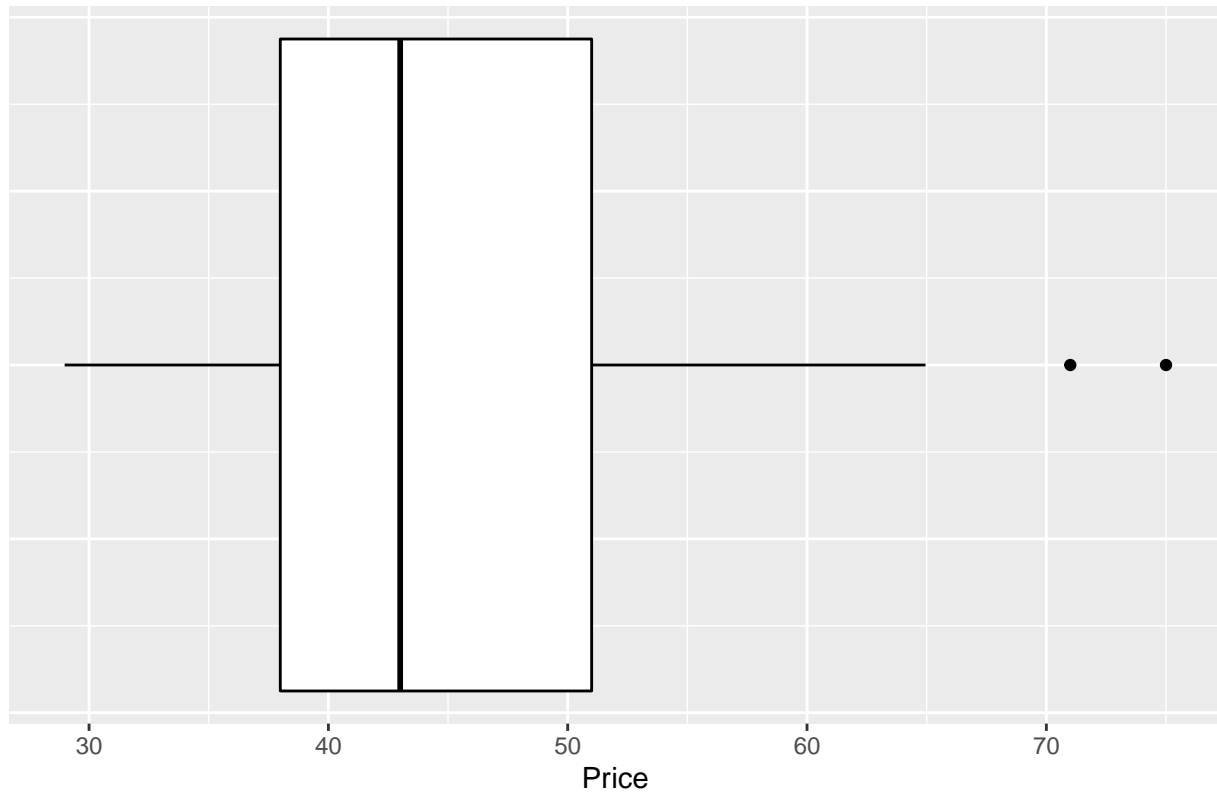
Adjusted price = Total price - Shipping Price



The above graph shows the liner regression of number of wheels and the adjusted price.



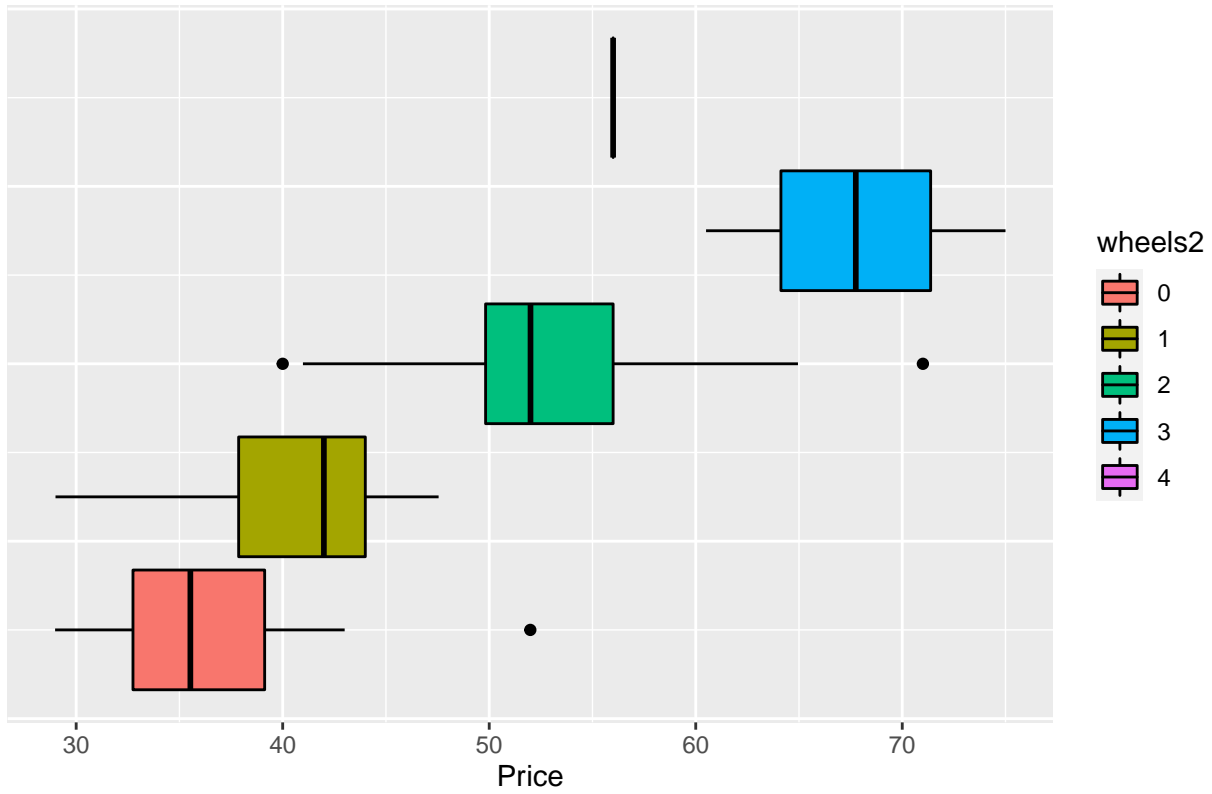
Boxplot of Adjusted Price

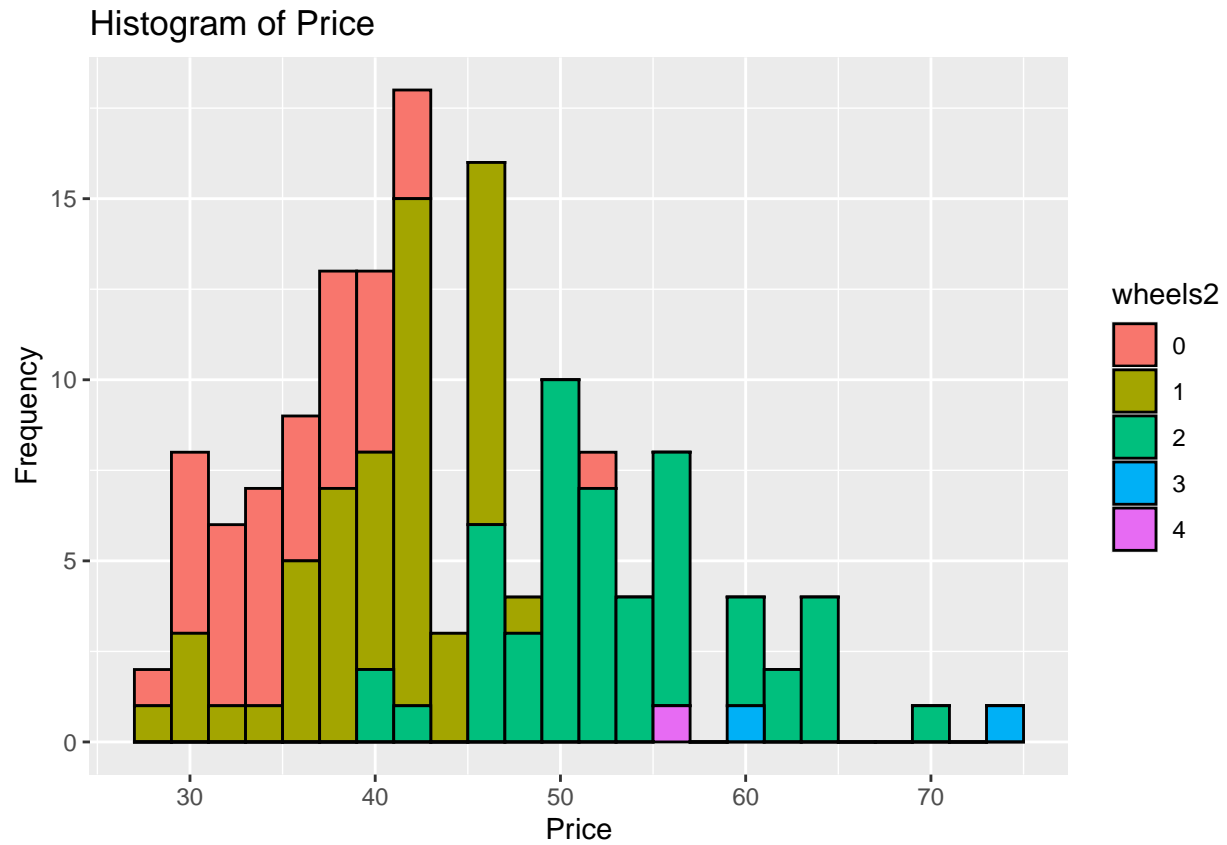


```
##   wheels mean_price
## 1     0      36.28
## 2     1      40.55
## 3     2      53.23
## 4     3      67.75
## 5     4      56.00
```

```
##   Wheels mean_price
## 1   zero      36.28
## 2   one      40.55
## 3   two      53.23
## 4 three      67.75
## 5   four      56.00
```

Boxplot of Price





The above visualizations represent the relationships between price and number of wheels. The boxplot and histogram compare the adjusted price of the Mario Kart game sold based on the number of wheels.

## Methods

### Hypothesis Formation

The Hypothesis in words:

$H_0$ : The number of wheels does not affect the adjusted price or the adjusted price would be lower.

$H_A$ : As the number of wheels increases, the adjusted price would be higher.

```
##
## Call:
## lm(formula = price ~ wheels, data = mkprice)
##
## Coefficients:
## (Intercept)      wheels
##      34.548      8.621
##
## Call:
## lm(formula = price ~ wheels, data = mkprice)
##
## Coefficients:
## (Intercept)      wheels
##      34.548      8.621
##
## Call:
## lm(formula = price ~ wheels, data = mkprice)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.169  -3.548  -0.270   2.861  19.210
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  34.5478    0.8594   40.20  <2e-16 ***
## wheels       8.6211    0.6034   14.29  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.03 on 139 degrees of freedom
## Multiple R-squared:  0.5949, Adjusted R-squared:  0.592
## F-statistic: 204.1 on 1 and 139 DF,  p-value: < 2.2e-16
SE = 0.6034 T statistic = 0.6024 P-Value = 2 x 10^-16
```

$$Y_{price} = b_0 + b_{1wheels}$$

In this scenario,  $b_0 = 34.548$  and  $b_1 = 8.6211$

This liner regression suggests that the initial price is 34.548, with one unit increase of wheels, the final price will be increased by 8.6211.

The Hypothesis with Math Symbols

$H_O$ :  $\beta_1 = 0$ , there is no linear relationship between number of wheels and the Mario karts' price

$H_A$ :  $\beta_1 > 0$ , there is some linear relationship between number of wheels and the Mario karts' price

## Applying Statistical Method

For hypothesis 1, the study will use linear regression testing to examine whether the relations between the two variables are strong or weak, and determine if it's possible to reject the null hypothesis.

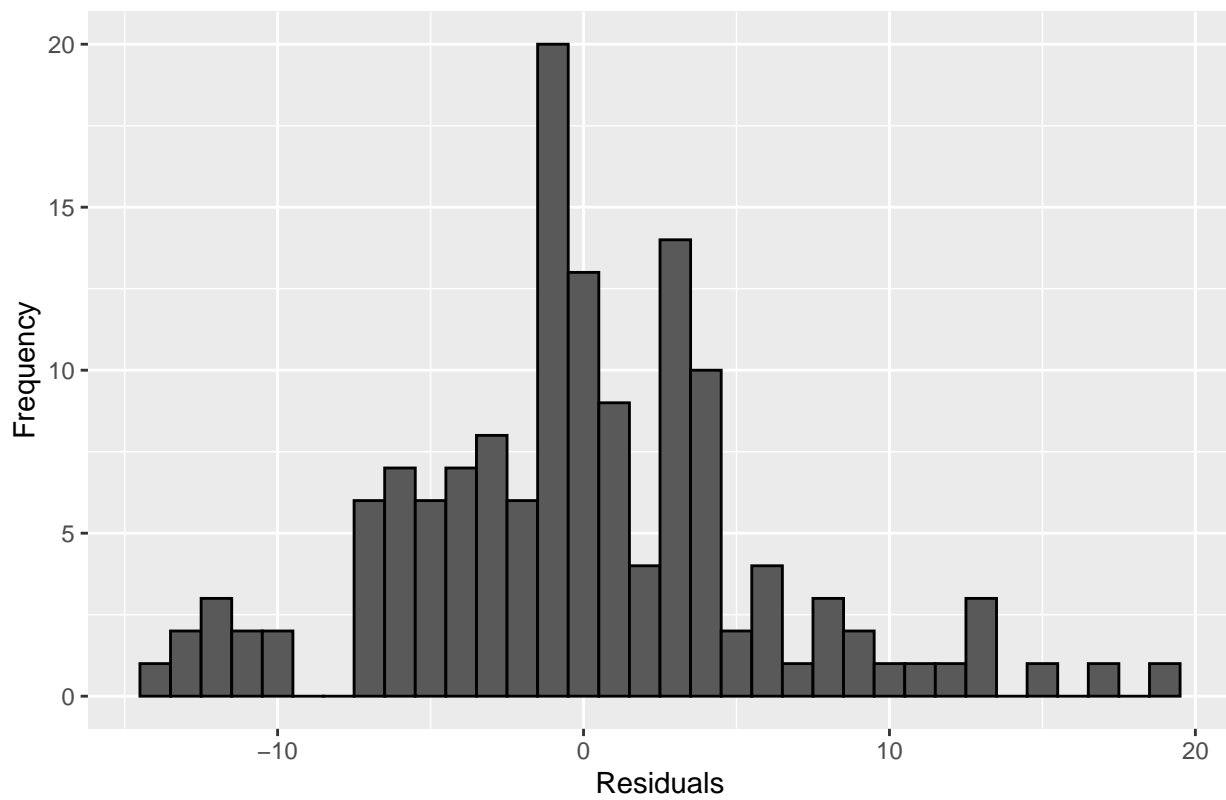
The explanatory variable here is number of wheels, and the response is price. The hypothesis is that the number of wheels has an effect on the overall price. These variables were selected because there is a strong positive correlation between wheels and price and wheels is the explanatory variable because it explains why the price increases.

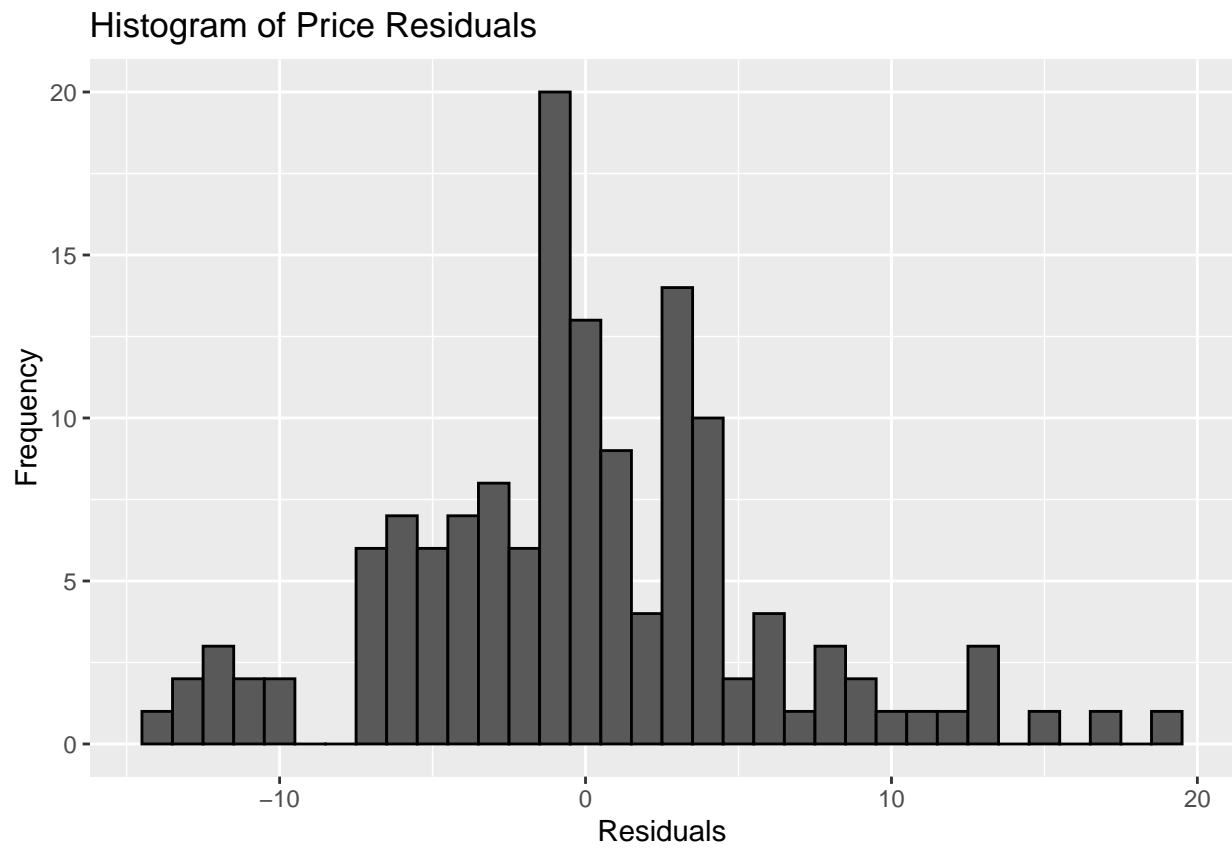
```
##
## Call:
## lm(formula = price ~ wheels, data = mkprice)
##
## Coefficients:
## (Intercept)      wheels
##      34.548        8.621
##
## Call:
## lm(formula = price ~ wheels, data = mkprice)
##
## Coefficients:
## (Intercept)      wheels
##      34.548        8.621
##
## Call:
## lm(formula = price ~ wheels, data = mkprice)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.169  -3.548  -0.270   2.861  19.210
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  34.5478    0.8594   40.20  <2e-16 ***
## wheels        8.6211    0.6034   14.29  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.03 on 139 degrees of freedom
## Multiple R-squared:  0.5949, Adjusted R-squared:  0.592
## F-statistic: 204.1 on 1 and 139 DF, p-value: < 2.2e-16
SE = 0.6034 T statistic = 0.6024 P-Value = 2 x 10^-16
95% confidence interval
CI = b +/- t* x SE
## [1] 1.977178
8.6211 - (1.9772 x 0.6034) = 7.4281 8.6211 + (1.9772 x 0.6034) = 9.8141
CI = (7.4281, 9.8141)
```

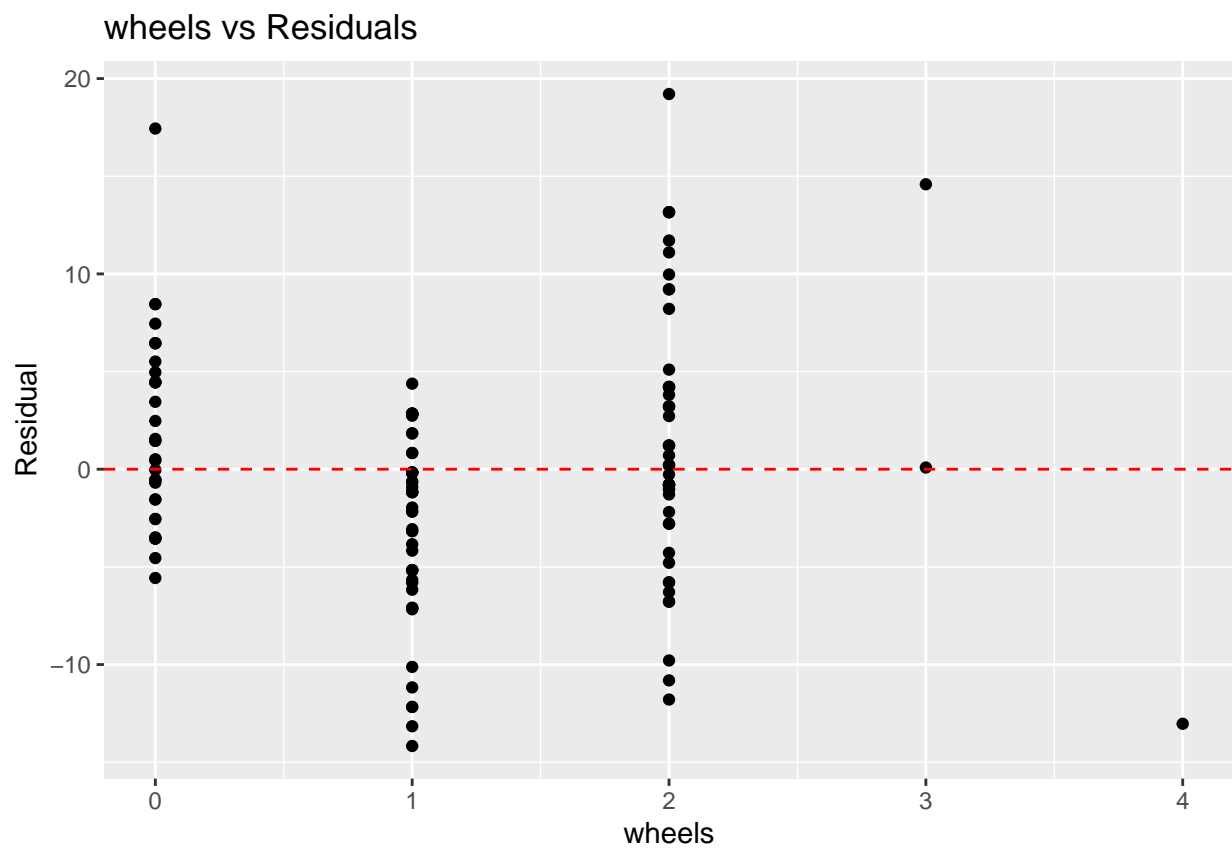
We are 95% certain that the true value of our slope is within the interval above, which does not include the null value of 0, which tracks with our P-Value of near zero, meaning our data is unlikely to have occurred by chance. We can reject the null hypothesis.

Our  $R^2$  value is also 0.59, meaning that 59% of our data is covered by the linear regression model, which is over 50%, meaning the linear regression model is a useful model in interpreting this data.

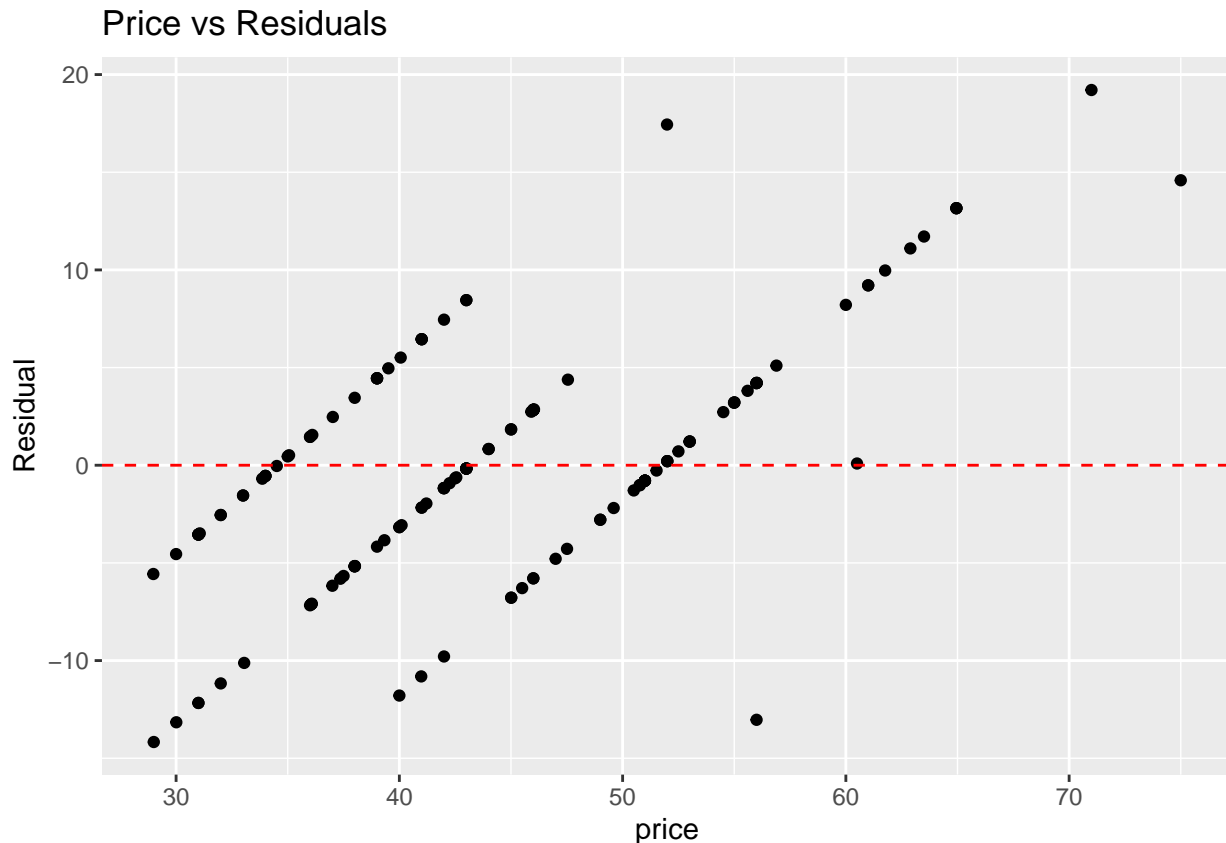
**Histogram of Wheels Residuals**











## Results and Discussions

The results are, unsurprisingly, that when one wants to get more from an eBay seller, one has to pay more. A P-value of  $2 \times 10^{-16}$  and 95% confidence interval of (7.4281, 9.8141) strongly agree that there is a strong linear relationship between price and number of wheels. Using the linear regression hypothesis testing to check the validity of the linear regression equation. The linear regression equation was  $Y_{price} = 8.6211 * b_{1wheels} + 34.5478$ . The study was able to complete a one sided hypothesis test and reject the null hypothesis that there is no positive linear relationship between the two variables. The R-sqaure value of 0.59 also shows that a majority of the data is encompassed by the linear regression model, meaning that the model is useful in interpreting the data and predicting values for the data. Although the results of the hypothesis test are as expected in the broad sense that we would expect people to pay more for receiving more wheels, the strength of the linear regression model is still interesting because it was computed a linear regression equation using R to match the hypothesis.

## Conclusions

Reviewing the evidence of the results coming about from the methods of this study, it is indisputable, and frankly intuitive, that the more wheels included in an eBay auction for a copy of Mario Kart Wii is more expensive it will be. What was surprising, however, was the extent of expense that each wheel put onto the auction. The wheels aren't necessary to play the game in any capacity, as mentioned before, but rather was included in the game to more accurately simulate steering an actual car or revealing slight discomfort when using the bare Wii remote to play the game. The wheels are plastic and constructed fairly cheaply as well. With that being said, it should be expected that each wheel would be fairly cheap, but that is not what the data showed with a mean wheel price of \$8.6211. Further research, either in the Psychological, Economic, or marketing field could examine why people may have been willing to shell out a significant amount of money for something not incredibly significant to the performance for the game, but rather something that simply completes the vibe of playing it.

## References

Bajari, P., & Hortaçsu, A. (2003). The Winner's Curse, Reserve Prices, and Endogenous Entry: Empirical Insights from eBay Auctions. *The RAND Journal of Economics*, 34(2), 329–355. <https://doi.org/10.2307/1593721>