Witch Hunting: Disenchantment with Magic's Modern Format

DATA 824: Summer Class Project, 2020

Brendan Stevens

Department of Biostatistics and Data Science
University of Kansas, USA
July 23, 2020

Contents

Abstract 1
Introduction 2
  Primary Analysis Objectives 2
  Secondary Analysis Objectives 3
Materials and Methods 3
  Data Sources 3
  Statistical Analysis 4
Results 42
Primary Objective Results 44
Discussion and Conclusion 45
Appendix: R-code 47

List of Tables

1 Summary Statistics of the Data Part 1 5
2 Summary Statistics of the Data Part 2 6
3 Summary Statistics of the Data Part 3 6
4 Archetype vs Deck Price 35
5 Deck Count 1 Year vs Deck Price 35
6 Flooded Strand vs Deck Price 35
7 Polluted Delta vs Deck Price 36
8 Bloodstained Mire vs Deck Price 36
9 Wooded Foothills vs Deck Price 36
10 Windswept Heath vs Deck Price 36
11 Marsh Flats vs Deck Price 37
12 Scalding Tarn vs Deck Price 37
13 Verdant Catacombs vs Deck Price 37
14 Arid Mesa vs Deck Price 37
15 Misty Rainforest vs Deck Price 38
16 Prismatic Vista vs Deck Price 38
17 Total Fetchlands vs Deck Price 38
18 Full Model vs Deck Price Intercept and Coefficients 40
19 Reduced Model vs Deck Price Intercept and Coefficients 40

List of Figures

1 Archetype vs Deck Price . . . 7
2 Deck Count 1 Year vs Deck Price . . . 9
3 Deck Count 1 Year vs Deck Price Test for Normality . . . 10
4 Flooded Strand vs Deck Price . . . 11
5 Flooded Strand vs Deck Price Test for Normality . . . 12
6 Polluted Delta vs Deck Price . . . 13
7 Polluted Delta vs Deck Price Test for Normality . . . 14
8 Bloodstained Mire vs Deck Price . . . 15
9 Bloodstained Mire vs Deck Price Test for Normality . . . 16
10 Wooded Foothills vs Deck Price . . . 17
11 Wooded Foothills vs Deck Price Test for Normality . . . 18
12 Windswept Heath vs Deck Price . . . 19
13 Windswept Heath vs Deck Price Test for Normality . . . 20
14 Marsh Flats vs Deck Price . . . 21
15 Marsh Flats vs Deck Price Test for Normality . . . 22
16 Scalding Tarn vs Deck Price . . . 23
17 Scalding Tarn vs Deck Price Test for Normality . . . 24
18 Verdant Catacombs vs Deck Price . . . 25
19 Verdant Catacombs vs Deck Price Test for Normality . . . 26
20 Arid Mesa vs Deck Price . . . 27
21 Arid Mesa vs Deck Price Test for Normality . . . 28
22 Misty Rainforest vs Deck Price . . . 29
23 Misty Rainforest vs Deck Price Test for Normality . . . 30
24 Prismatic Vista vs Deck Price . . . 31
25 Prismatic Vista vs Deck Price Test for Normality . . . 32

26 Total Fetchlands vs Deck Price 33
27 Total Fetchlands vs Deck Price Test for Normality 34
28 Full Model vs Deck Price Test for Normality 39
29 Fitted vs Residual graphs for the phases of the model. 41
30 Fitted vs Residual graphs for the phases of the model. 42

Abstract

The popularity of the trading card game, “Magic: the Gathering” has grown tremendously in the last ten years, leading to an increase in demand without additional supply and an inevitable conversation about the cost barrier to entry of playing in the various formats offered by the game.

This study seeks to shed light on cost prediction in the Modern Format based on a subset of the cards played in Modern, the “Fetchlands”. The goal is to find the best model for predicting the cost of a deck for a player based upon as few variables as possible, while maintaining a high degree of accuracy. The analysis methods used in this study are linear regression and multiple regression.

Contrary to the what many in the playerbase view as accepted truth, the “Fetchlands” as a whole group of 11 cards (including Prismatic Vista) are not in totality statistically more significant predictors for the price of a deck than other cards in the format. There are, however, only five pieces of information necessary in order to predict over 70% of the variability in a deck’s cost.

Introduction

Magic: the Gathering is a collectable created by Richard Garfield first debuting in 1993. As it is and remains a *collectible* card game, the cost of cards varies - as in any economy - based upon supply and demand. With the release of Shards of Alara in 2008, the game saw an explosion of growth through the next decade. However, with this growth in playerbase came an increase in demand on older collectible product that was not printed at the new playerbase supply level needed. This demand has generated a price increase acrossed much of the product as well as generated a growing narrative in subsections of the community is that cards in general and specific formats (formats are ways to play the game) are too hard to get into due to their cost barriers. In particular, the cost of a specific subgroup of cards, nicknamed the "Fetchlands", has created an outcry of frustration, due to the fact that many players believe that the cards are "required to play the game". The statement that a card is "too expensive" is an often an emotional one based upon a player's desire to aquire the card, but not being able to justify the cost personally. Lest it be forgotton, Magic: the Gathering is a *collectible* card game founded around the concept of scarcity and rarity. It is not by nature a standalone card game like Uno or games played with a standard deck of 52 playing cards. Without this structure, it is entirely possible that Magic would not have found the financial success and therefore the funds necessary to continue making the game better every year. However, when the game becomes so cost prohibitive that the very scarcity that let it grow begins to stave off new players or disenfranchise long time players, the designers of the game must consider the possibility that some of the outcry is reasonable. In the pursuit of understanding cost prohibition and better arming the average player with the knowledge necessary to navigate the ocean of new cards, formats, and other players, the aim of this study is to help those interested to understand the subtleties of one of the most popular and longest running formats in the game, "Modern". For example, the claim that "Some colors (there are five colors in Magic) are more expensive to play than others". Or, "Aggro decks are cheap and Control decks are expensive". These are things that a player can use in order to make educated decisions when choosing decks to play. But better yet, what if estimating the price of a deck wasn't based on "conventional wisdom", and instead a quantifiable exercise which required nothing more than answering a few questions. The aim of this study is to do just that. To create a model that can reasonably predict the cost of purchasing an entire 60 card, Modern-format-ready deck with no starting collection. All the player would have to know is few choice pieces of information about what the deck contains in order to forcast the price.

A Quick Note: Other than "Basic Lands", a Deck may only contain anywhere between 1-4 copies of a uniquely named card. Fetchlands are not basic. This means that a deck may only have between 1-4 copies of a uniquely named Fetchland.

Primary Analysis Objectives

Determine whether there exists a good model to predict the price of a deck in modern. This will be done in part by determining if there are statistically significant predictor variables useful in the creation of a model that can predict the price of a Modern deck given certain information.

Secondary Analysis Objectives

Visualize the relationship between the number of Fetchlands in a deck and the Deck Price. Summarize the statistics of the decks in the format as a whole. Determine if specific Fetchlands are more price predictive on the cost of a deck than others. Determine if the number of decks of a specfic name correlates with the price of that deck. Determine if the number of decks of a specific "archetype" is dependent on the number of fetchlands required for that deck.

Materials and Methods

Data Sources

The data are available in .csv (comma separated value) format. They include 59 unique observations which were the top 59 most played decks in the format on the date that the data were gathered, June 14, 2020, from the popular article and data aggregation site for the game, "MTG Goldfish". The data include a number of variables. Here is a list and a basic explanation of what that variable is:

- "Deck Name", the deck's name, sometimes indicates the colors, othertimes indicates the iconic cards that the deck is built around
- "Archetype", the general method the deck uses to win the game
- "Format", the various ways in which the game can be played, though this analysis only considers the "Modern" format
- "Deck Count 1 Year", how many decks of that name were recorded by MTG Goldfish during the last calendar year
- "Percent of Meta", what percent of the decks are this deck
- "Deck Price", cost of buying each of the cards for the deck from TCGPlayer and CardKindgom, online retailers for singles
- "Flooded Strand", the White-Blue Fetchland
- "Polluted Delta", the Blue-Black Fetchland
- "Bloodstained Mire", the Black-Red Fetchland
- "Wooded Foothills", the Red-Green Fetchland
- "Windswept Heath", the White-Green Fetchland
- "Marsh Flats", the White-Black Fetchland
- "Scalding Tarn", the Blue-Red Fetchland
- "Verdant Catacombs", the Black-Green Fetchland
- "Arid Mesa", the White-Red Fetchland
- "Misty Rainforest", the Blue-Green Fetchland
- "Prismatic Vista", the basic land Fetchland
- "Total Fetchlands", total fetchlands in the average deck of that name
- "Total Fetch Price", total price of the average combination of fetchlands present in deck of that name
- "Deck Price No Fetch", total price of the average deck of that name without any fetchlands

- "Fetch Price Cost Proportion", total cost of average combination of fetchlands in deck of that name divided by average price of that deck including average combination of fetchlands

This data was chosen due to the relevance of the topic in the community at this time, the explosion in popularity of the game in the last ten years, and the personal interest of the individual writing this report and analyzing the data.

Statistical Analysis

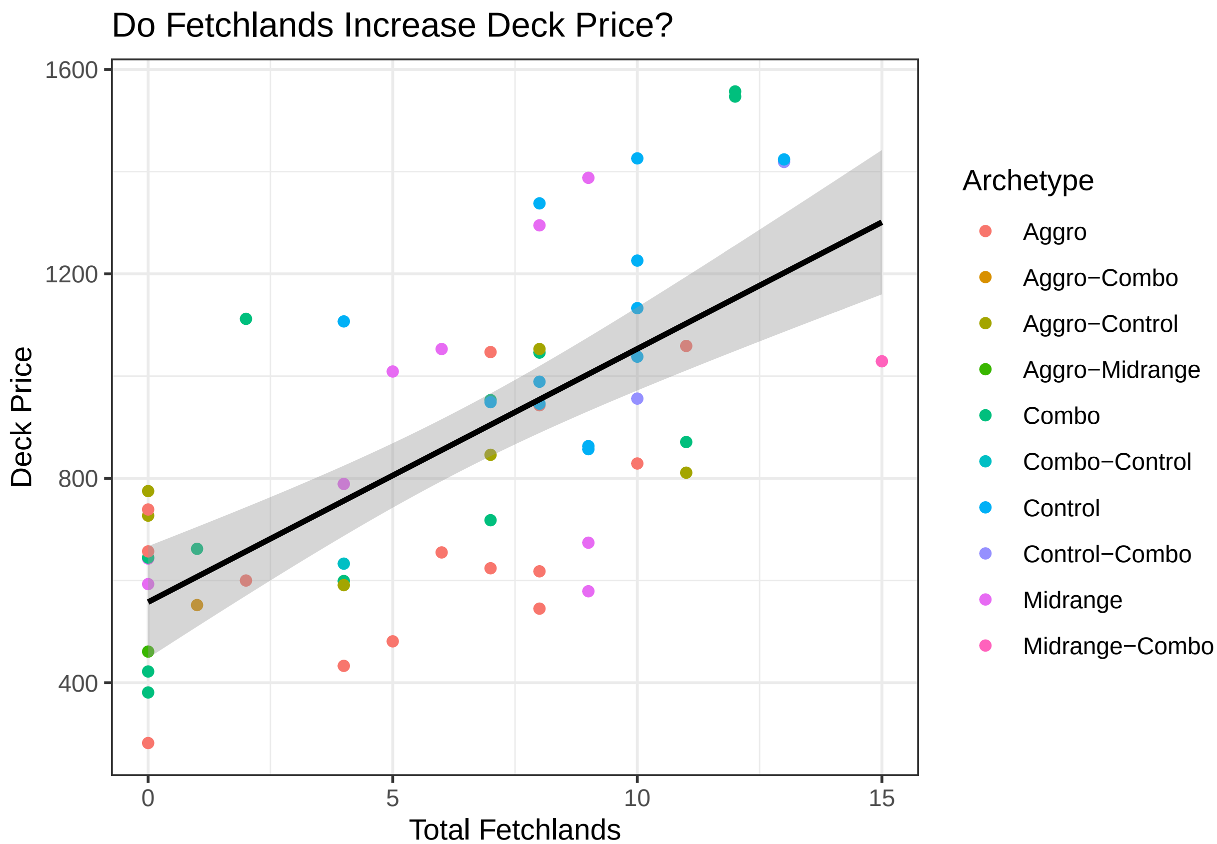
The analysis is done using the statistical software program R, version 3.6.3 (2020-02-29). The analysis is primarily focused on single and multiple linear regression. The data was provided in with some missing information. That information was what "Archetype" each deck was, and that data was solely researched and added by the writer of this report. The visualizations in the report begin with univariate regression in order to explore each predictor variable in isolation. After this, a multivariable regression model is fitted, and the process of removing "bad" predictor variables is conducted.

Model Assumptions

All inferences are conducted using $\alpha = 0.05$ unless stated otherwise. No adjustments for multiplicity are made as this is an exploratory analysis. Discrete variables are summarized with proportions and frequencies. Continuous variables are summarized using mean, median, standard deviation, and other common summary statistics.

Secondary Objective Analysis

To begin the analysis, consider the relationship of the Deck Prices of the top 59 most played decks in the format versus the number of Fetchlands present in said deck.



The graph almost speaks for itself here. There is absolutely a positive linear relationship between the number of Fetchlands in a deck and the Price of the deck. This is not trivial, because to put more Fetchlands into a deck means to remove other cards from the deck. This means that Fetchlands are, on average, more expensive than all of the other cards that people are running in Modern.

The following two tables display the summary statistics of the base data.

Table 1: Summary Statistics of the Data Part 1

	Deck Count 1 Year	Deck Price	Deck Price No Fetch	Fetch Price Cost Proportion	Percent of Meta	Total Fetch Price
Mean	52.25424	867.74576	587.13441	0.29008	1.47746	280.61136
Std.Dev	52.44919	308.92023	190.77150	0.19356	1.48643	218.39116
Min	7.00000	282.00000	282.00000	0.00000	0.20000	0.00000
Q1	14.00000	624.00000	448.20000	0.07949	0.40000	66.60000
Median	30.00000	846.00000	570.43000	0.33965	0.85000	288.69000
Q3	75.00000	1053.00000	725.56000	0.46132	2.13000	425.23000
Max	286.00000	1557.00000	1019.00000	0.59691	8.10000	771.22000
MAD	32.61720	306.89820	196.25176	0.21460	0.91921	260.44834
IQR	60.00000	421.50000	270.05500	0.35049	1.70000	345.61500
CV	1.00373	0.35600	0.32492	0.66728	1.00607	0.77827

	Deck Count 1 Year	Deck Price	Deck Price No Fetch	Fetch Price Cost Proportion	Percent of Meta	Total Fetch Price
Skewness	1.87818	0.41723	0.32140	-0.30501	1.88187	0.28879
SE.Skewness	0.31118	0.31118	0.31118	0.31118	0.31118	0.31118
Kurtosis	4.82934	-0.60284	-0.78006	-1.33949	4.82498	-0.93238
N.Valid	59.00000	59.00000	59.00000	59.00000	59.00000	59.00000
Pct.Valid	100.00000	100.00000	100.00000	100.00000	100.00000	100.00000

Table 2: Summary Statistics of the Data Part 2

	Bloodstained Mire	Flooded Strand	Marsh Flats	Polluted Delta	Windswept Heath	Wooded Foothills
Mean	0.55932	0.61017	0.13559	0.67797	0.74576	0.72881
Std.Dev	1.17841	1.38983	0.65542	1.41958	1.45736	1.27077
Min	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Q1	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Median	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Q3	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000
Max	4.00000	4.00000	4.00000	4.00000	4.00000	4.00000
MAD	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
IQR	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000
CV	2.10686	2.27777	4.83371	2.09387	1.95419	1.74361
Skewness	1.94232	1.91624	4.91937	1.74123	1.55471	1.49806
SE.Skewness	0.31118	0.31118	0.31118	0.31118	0.31118	0.31118
Kurtosis	2.50640	1.84590	23.71973	1.24127	0.63602	0.82890
N.Valid	59.00000	59.00000	59.00000	59.00000	59.00000	59.00000
Pct.Valid	100.00000	100.00000	100.00000	100.00000	100.00000	100.00000

Table 3: Summary Statistics of the Data Part 3

	Arid Mesa	Misty Rainforest	Prismatic Vista	Scalding Tarn	Total Fetchlands	Verdant Catacombs
Mean	0.22034	0.67797	0.33898	0.89831	6.25424	0.66102
Std.Dev	0.67128	1.35749	1.01047	1.57234	4.17130	1.38477
Min	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Q1	0.00000	0.00000	0.00000	0.00000	2.00000	0.00000
Median	0.00000	0.00000	0.00000	0.00000	7.00000	0.00000
Q3	0.00000	1.00000	0.00000	2.00000	9.00000	0.00000
Max	3.00000	4.00000	4.00000	4.00000	15.00000	4.00000
MAD	0.00000	0.00000	0.00000	0.00000	4.44780	0.00000
IQR	0.00000	0.50000	0.00000	1.50000	6.00000	0.00000

	Arid Mesa	Misty Rainforest	Prismatic Vista	Scalding Tarn	Total Fetchlands	Verdant Catacombs
CV	3.04657	2.00230	2.98087	1.75034	0.66696	2.09491
Skewness	3.08700	1.76316	2.84989	1.26364	-0.17647	1.83445
SE.Skewness	0.31118	0.31118	0.31118	0.31118	0.31118	0.31118
Kurtosis	8.74375	1.47740	6.83869	-0.23525	-1.07339	1.63806
N.Valid	59.00000	59.00000	59.00000	59.00000	59.00000	59.00000
Pct.Valid	100.00000	100.00000	100.00000	100.00000	100.00000	100.00000

Primary Objective Analysis

Before moving onto the 13 quantitative predictor variables, the sole categorical variable, Archetype is considered.

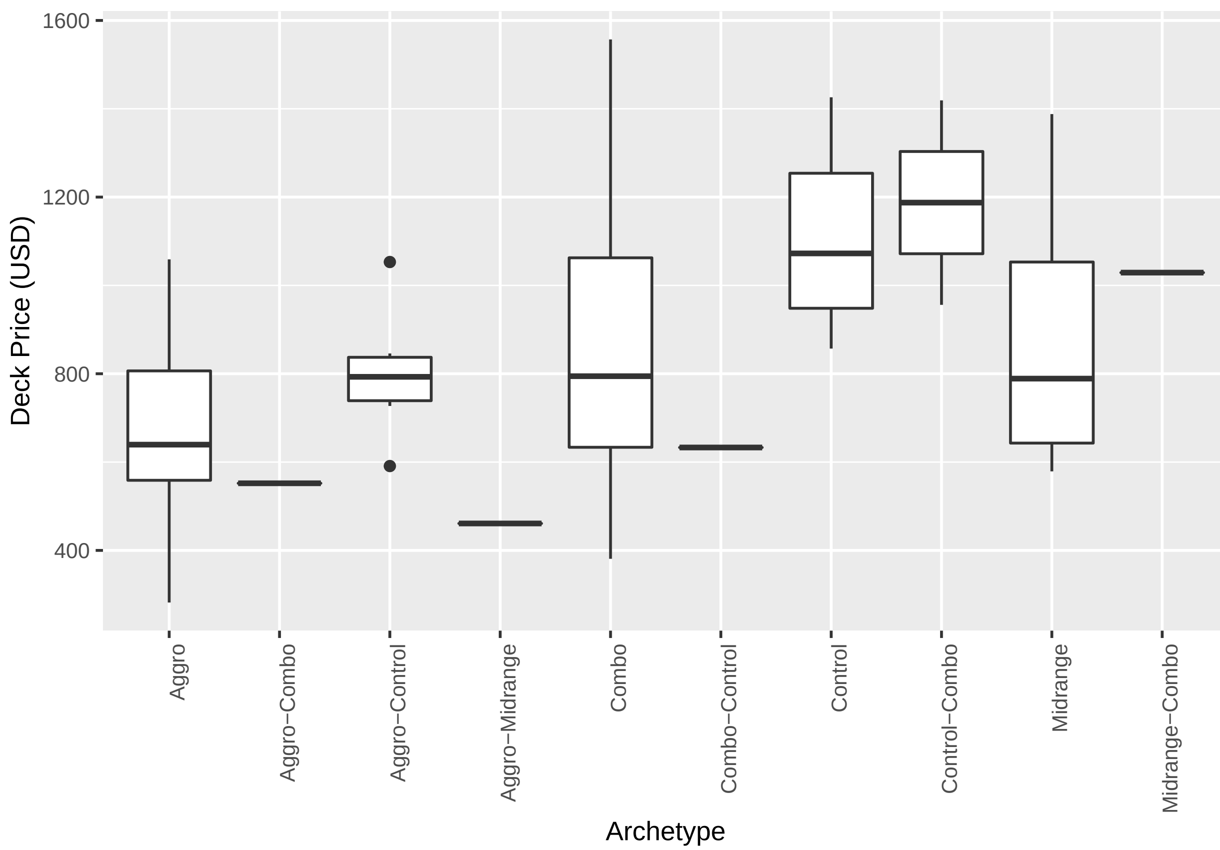


Figure 1: Archetype vs Deck Price

What can be seen in figure 1 is better understood by considering only the pure archetypes, or non-hybrid. The price of an "Aggro" deck is less than the other three, "Combo", "Control", and "Midrange". Combo and Midrange decks are about the same price, which makes sense to a long-time player of the game, because midrange decks have slowly replaced combo decks in the design philosophy of the game designers, but that's another conversation. Finally, the Control Archetype has the most expensive decks.

In fact, if all of the archetypes are considered, the most expensive decks are mainly the ones that have "Control" in their name, and the least expense? Aggro. No surprise to long time players of the game, but nonetheless, it may be interesting to see that the intuition of the playerbase on which decks are cheaper and which decks are more expensive is rather accurate.

What remains to be seen is if the intuition is right about the Fetchlands!

There are a total of 14 predictor quantitative variables to consider. The scatter plots of each of the single predictor variables should be analyzed in order to determine if there are any predictors that would aid in the creation of a strong predictive model for the Deck Price before moving on to multiple regression.

Each one of the models also needs to be tested for the regression requirements of normality, which will be done with the help of the Shapiro-Wilk tests and Q-Q Norm plots.

Special attention will be given to data that has promising Goodness of Fit indicator values, in our case limited to the R^2 value or Coefficient of Determination.

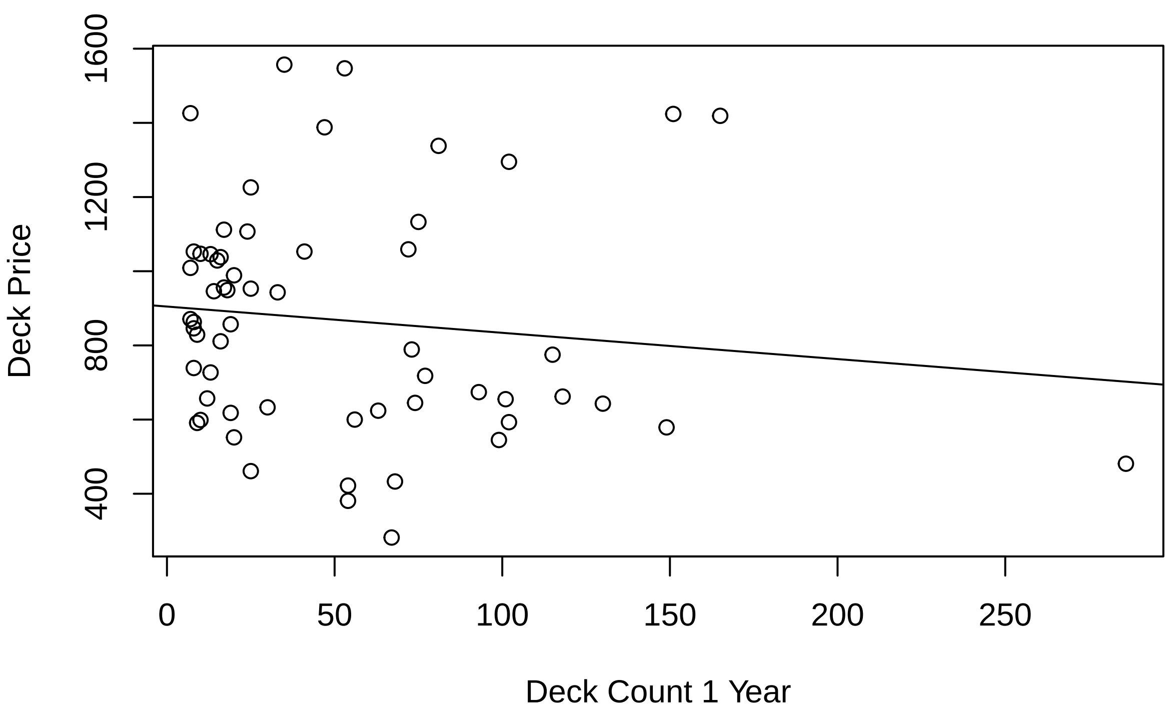


Figure 2: Deck Count 1 Year vs Deck Price

These Residuals appear to “hug” the line quite well, indicating normality. However, the Shapiro Wilk test gives a p-value of .034, which is less than .05, the normal cutoff. This should be kept in mind, as too many variables failing part of the normality test will affect the final model.

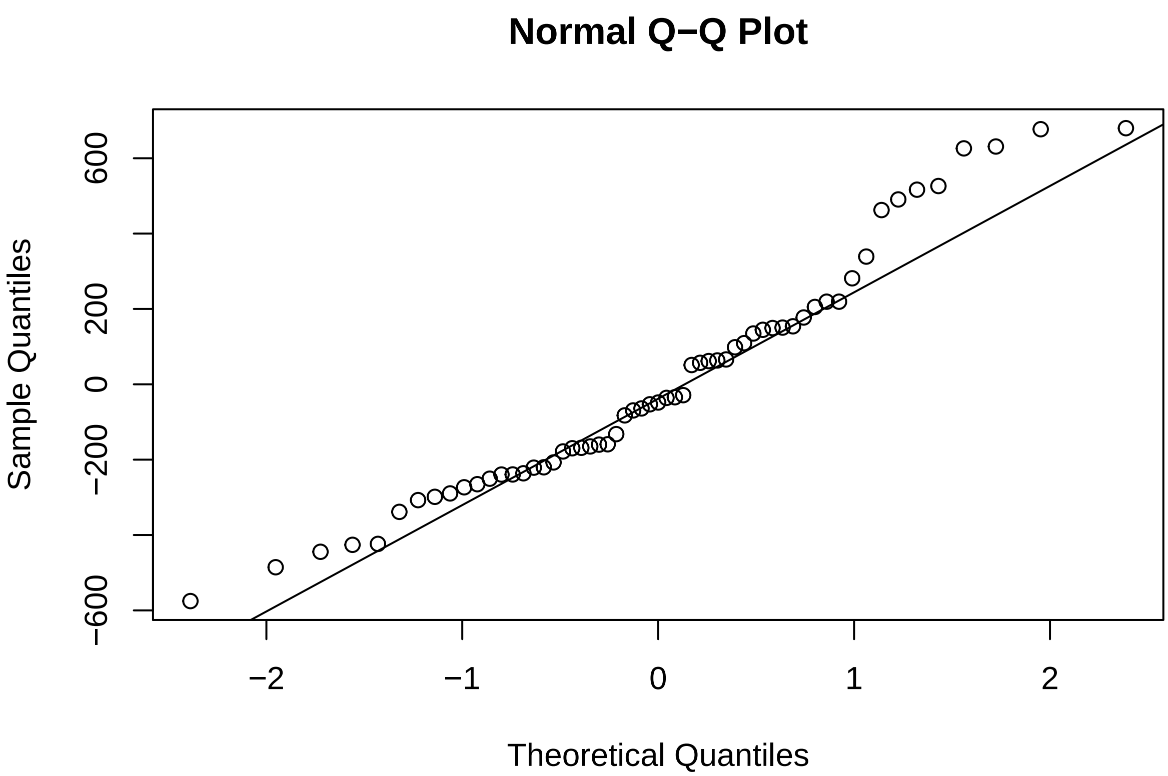


Figure 3: Deck Count 1 Year vs Deck Price Test for Normality

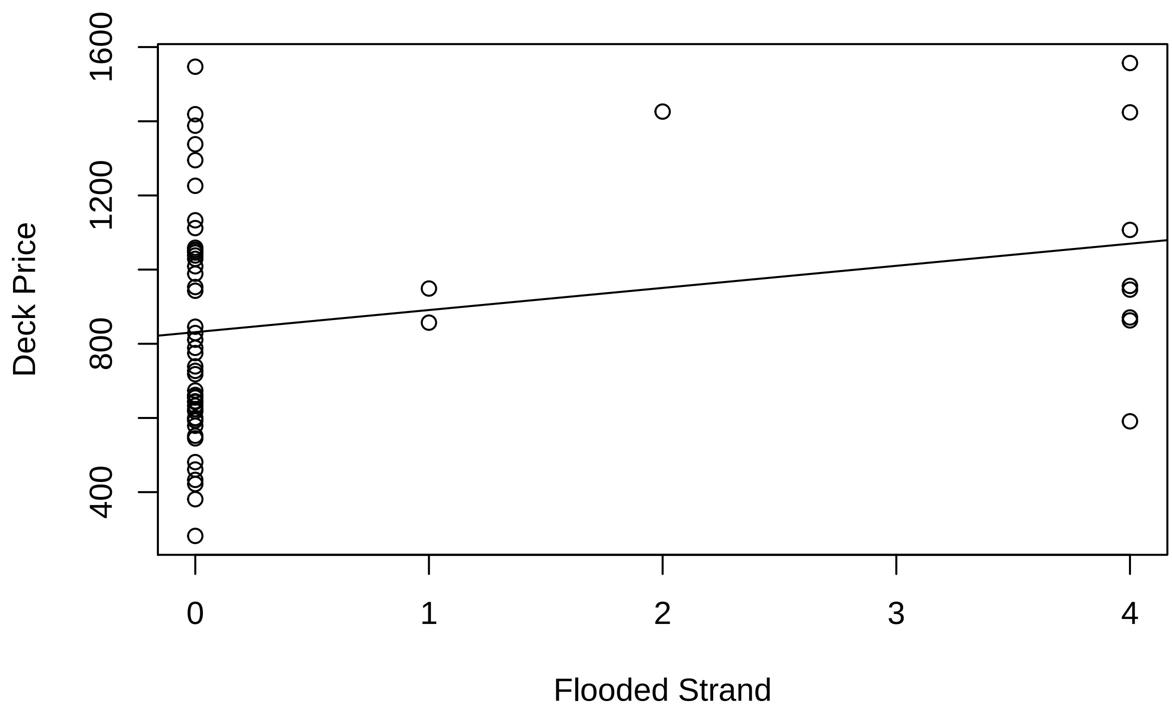


Figure 4: Flooded Strand vs Deck Price

These Residuals appear to “hug” the line quite well, indicating normality. The Shapiro Wilk test produces $p - value = .089$, which is more than our threshold of $\alpha = .05$, so we fail to reject normality. This data appears to be normal and regression analysis is appropriate.

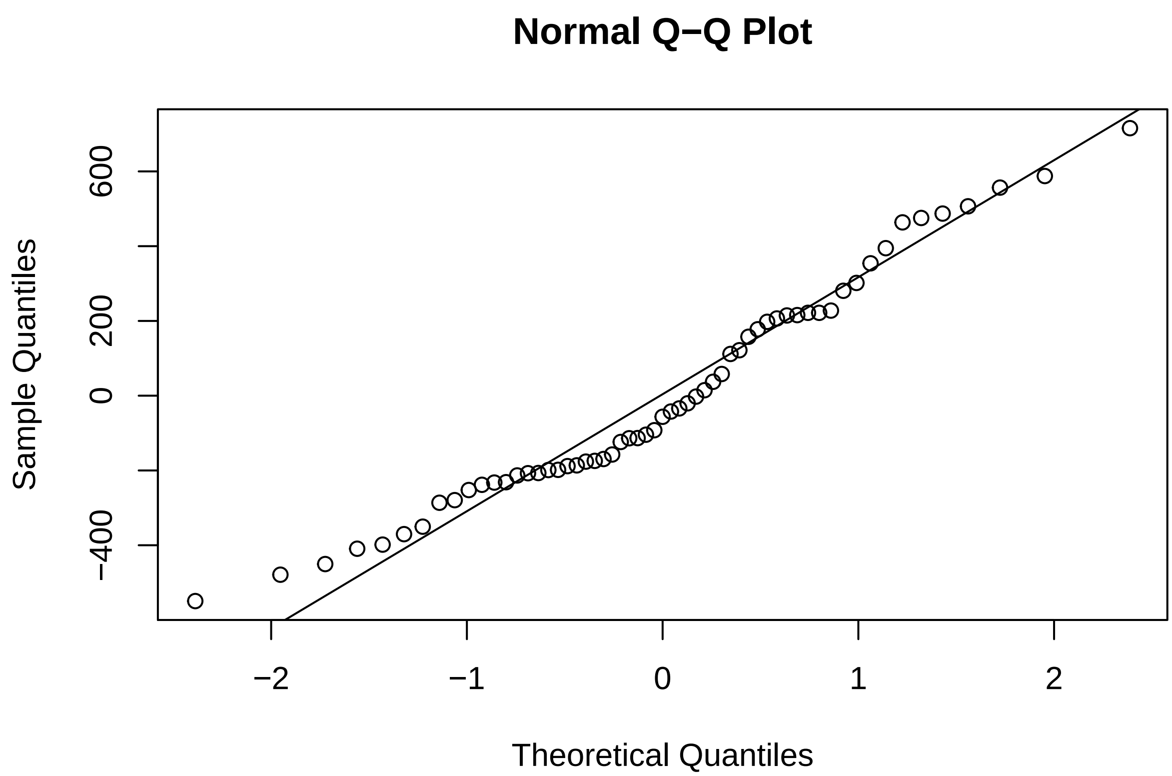


Figure 5: Flooded Strand vs Deck Price Test for Normality

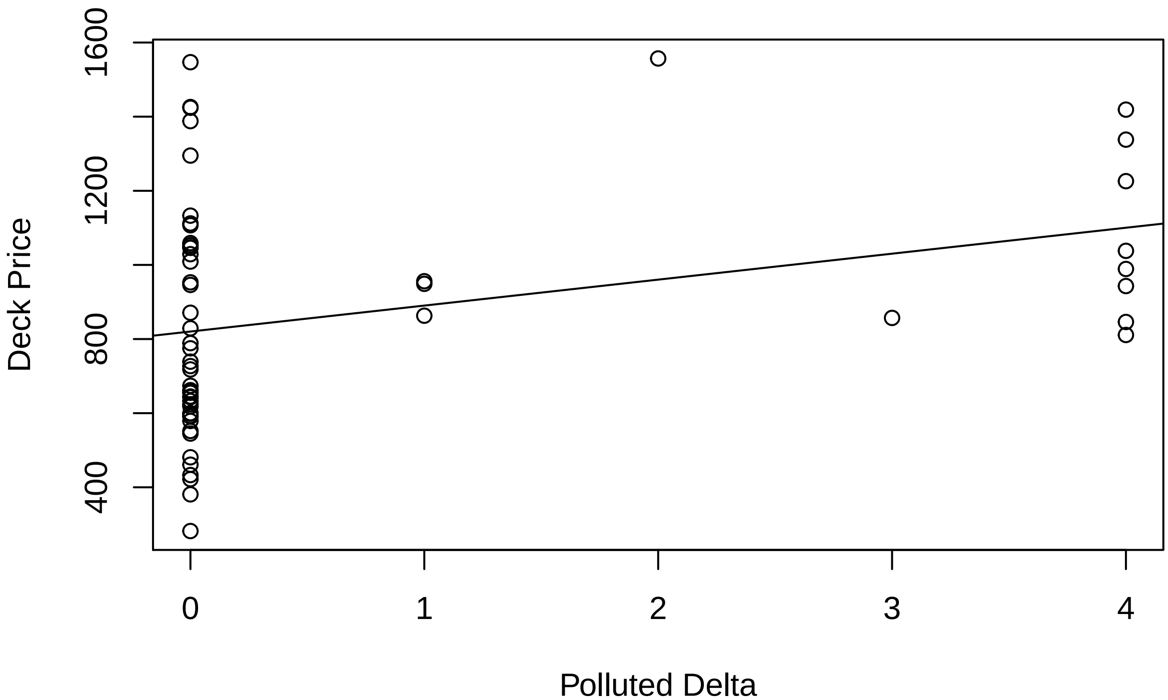


Figure 6: Polluted Delta vs Deck Price

These Residuals appear to “hug” the line quite well, indicating normality. The Shapiro Wilk test produces $p - value = .015$, which is less than our threshold of $\alpha = .05$, so we have another solo predictor value that would not be good for regression alone.

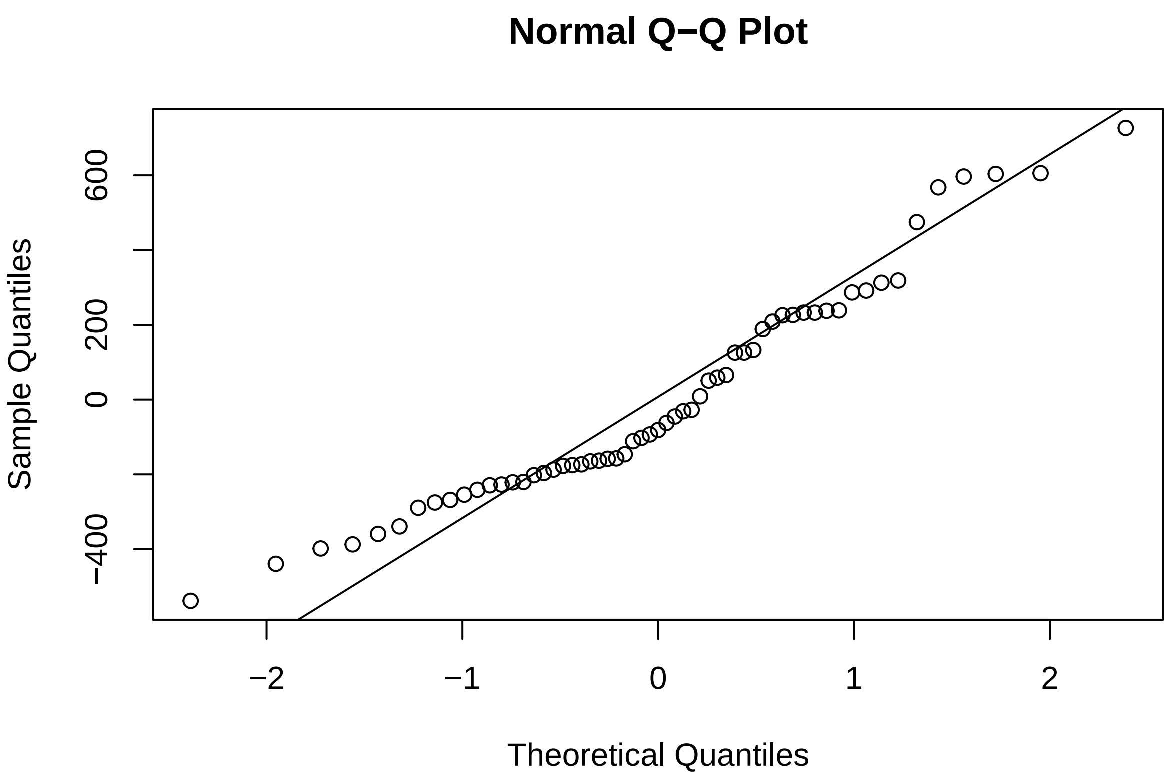


Figure 7: Polluted Delta vs Deck Price Test for Normality

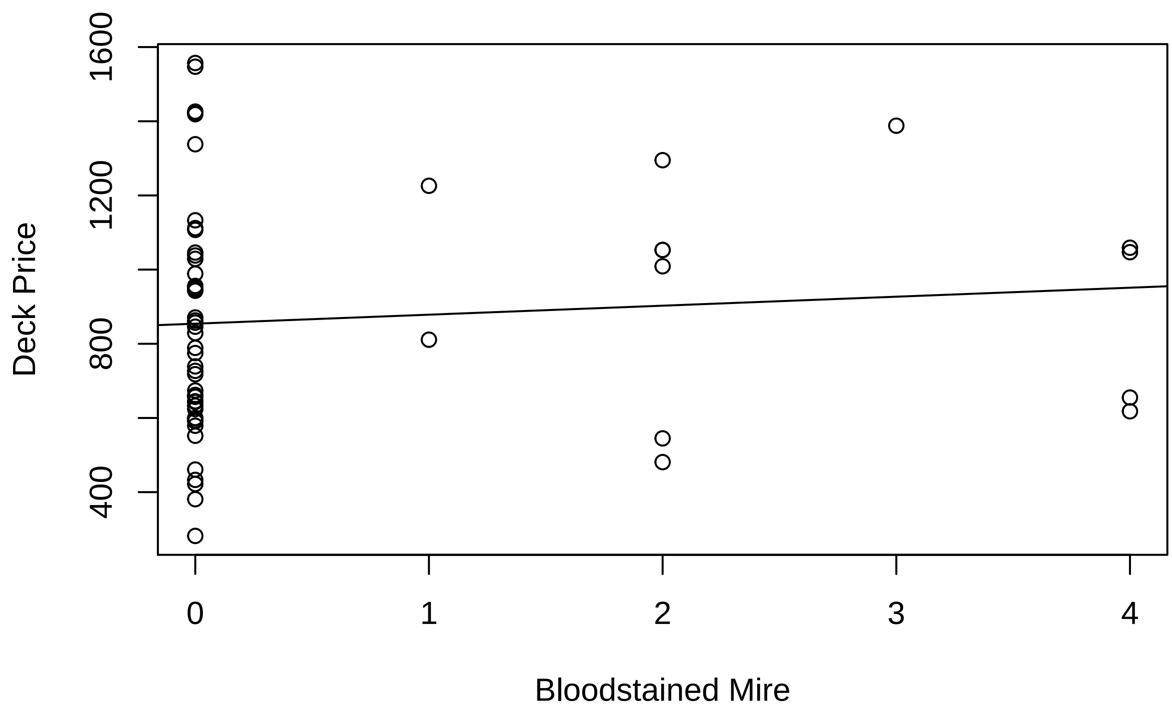


Figure 8: Bloodstained Mire vs Deck Price

These Residuals appear to “hug” the line quite well, indicating normality. The Shapiro Wilk test produces $p - value = .11$, which is more than our threshold of $\alpha = .05$, so we fail to reject normality. This data appears to be normal and regression analysis is appropriate.

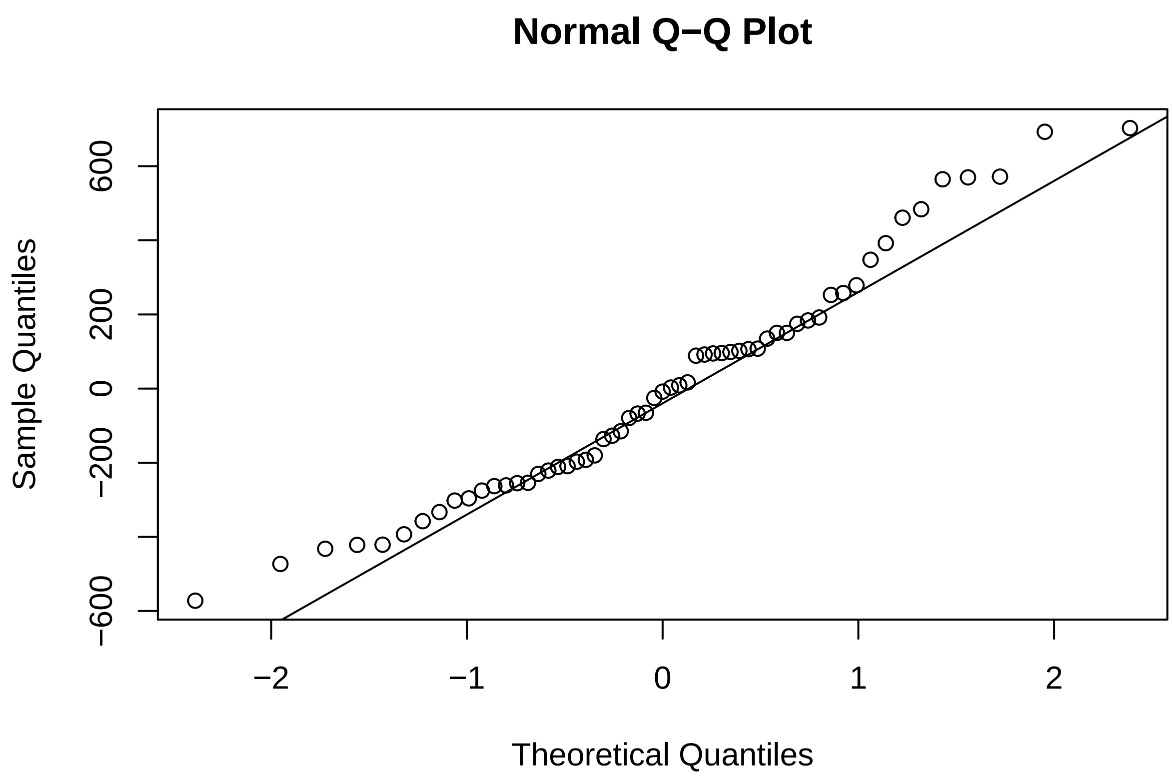


Figure 9: Bloodstained Mire vs Deck Price Test for Normality

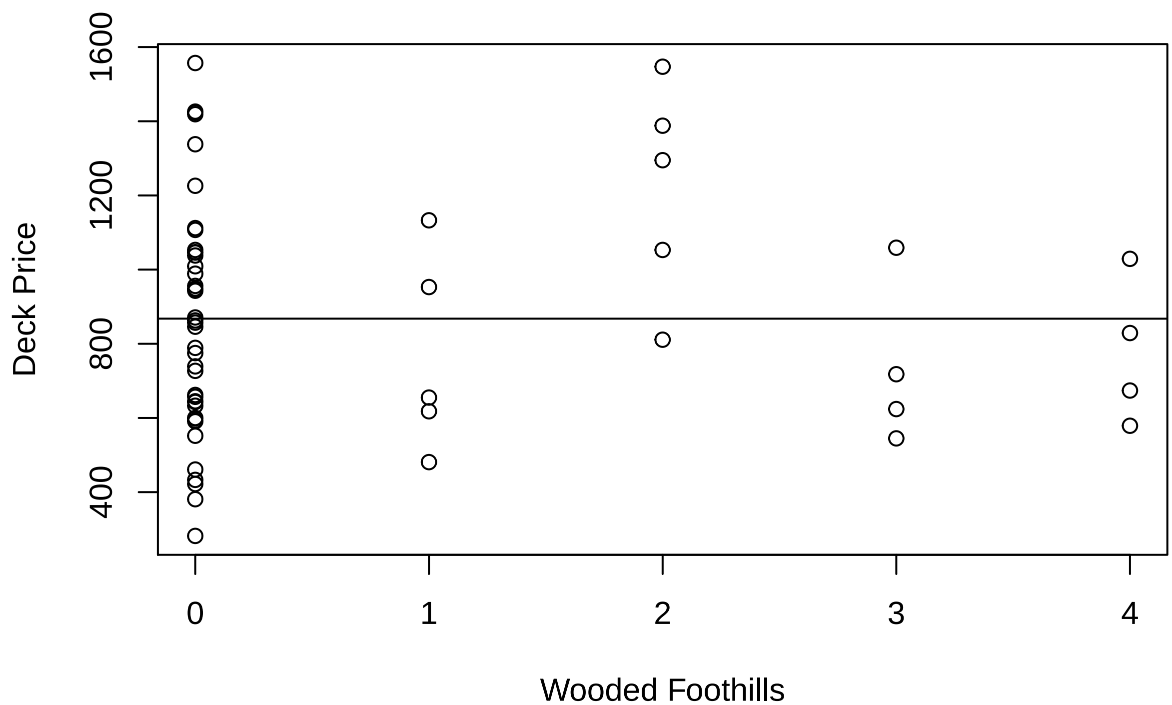


Figure 10: Wooded Foothills vs Deck Price

These Residuals appear to "hug" the line quite well, indicating normality. The Shapiro Wilk test produces $p-value = .091$, which is more than our threshold of $\alpha = .05$, so we fail to reject normality. This data appears to be normal and regression analysis is appropriate.

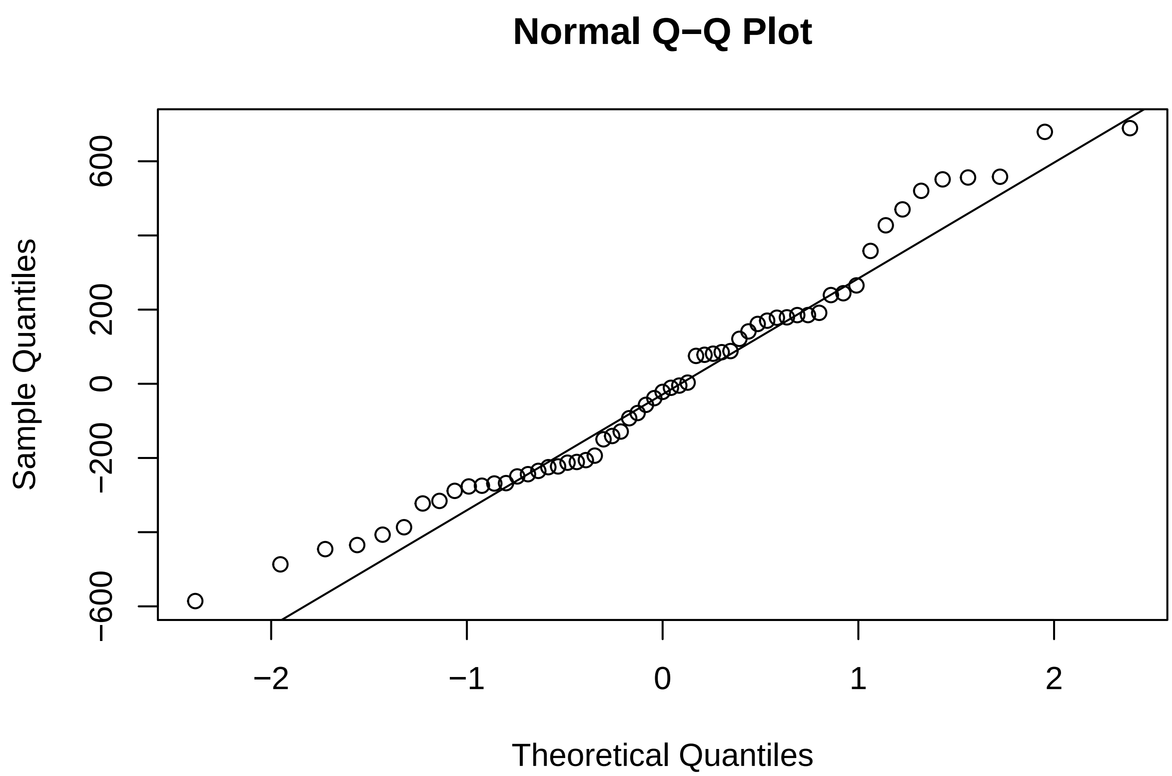


Figure 11: Wooded Foothills vs Deck Price Test for Normality

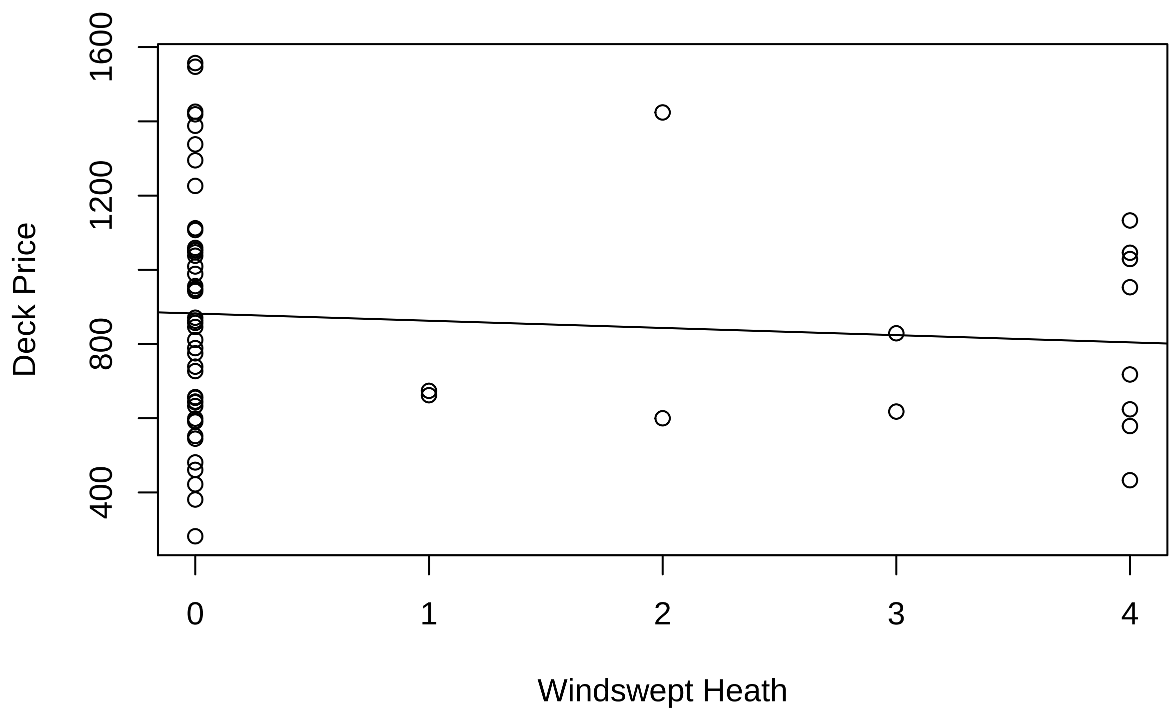


Figure 12: Windswept Heath vs Deck Price

These Residuals appear to "hug" the line quite well, indicating normality. The Shapiro Wilk test produces $p - value = .17$, which is more than our threshold of $\alpha = .05$, so we fail to reject normality. This data appears to be normal and regression analysis is appropriate.

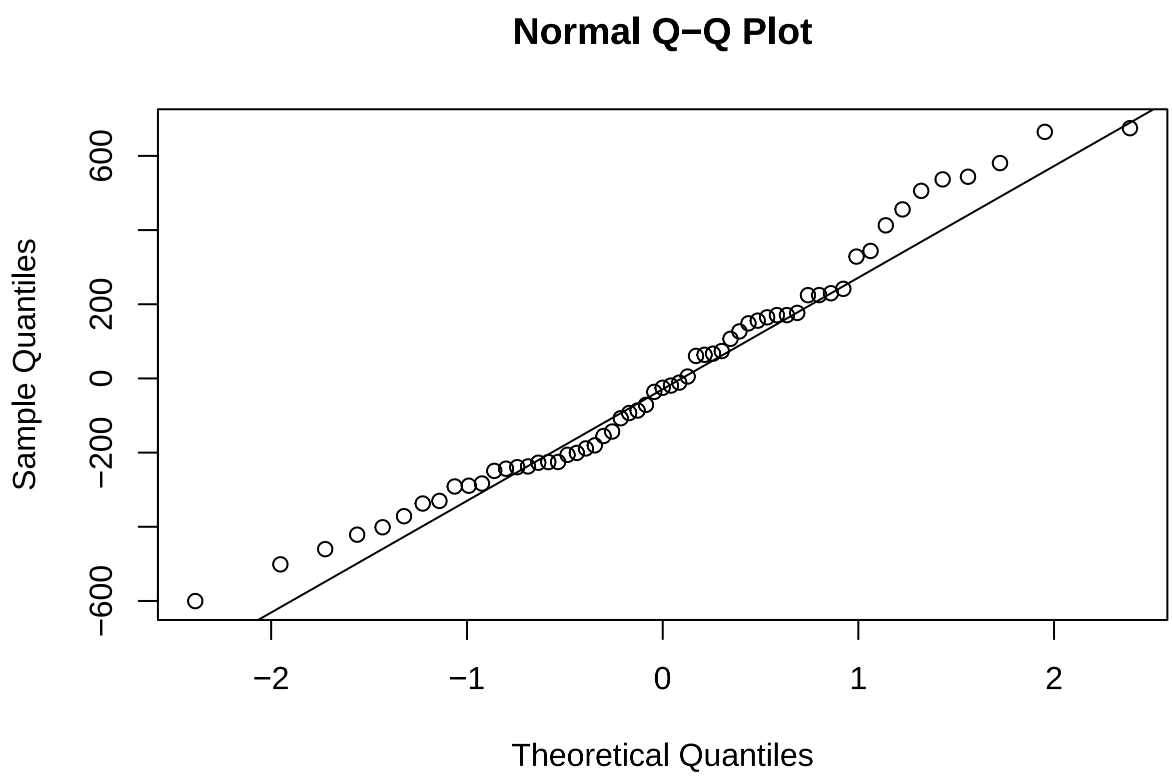


Figure 13: Windswept Heath vs Deck Price Test for Normality

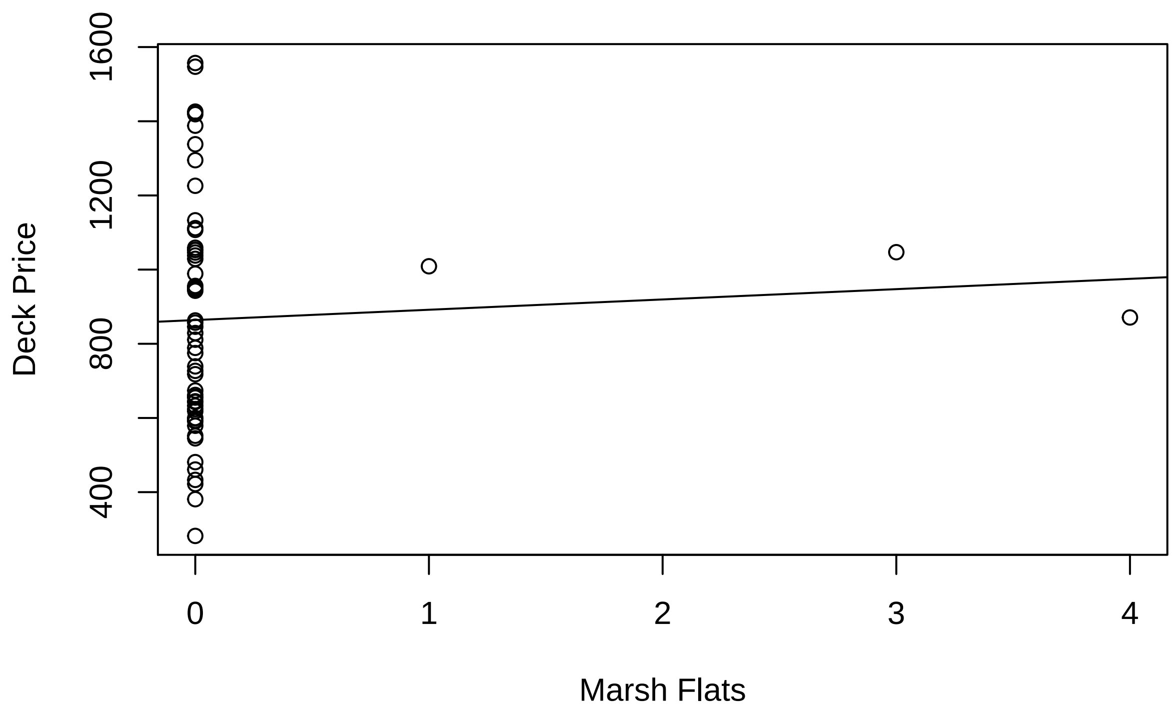


Figure 14: Marsh Flats vs Deck Price

These Residuals appear to "hug" the line quite well, indicating normality. The Shapiro Wilk test produces $p - value = .075$, which is more than our threshold of $\alpha = .05$, so we fail to reject normality. This data appears to be normal and regression analysis is appropriate.

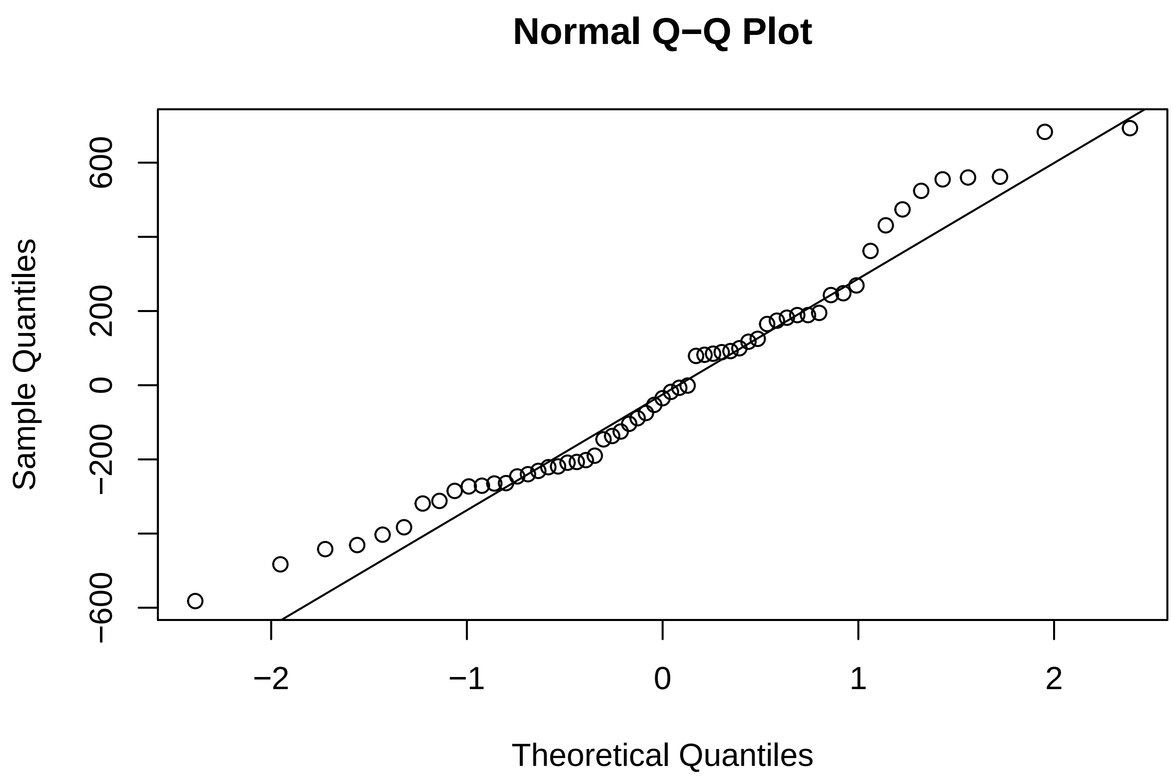


Figure 15: Marsh Flats vs Deck Price Test for Normality

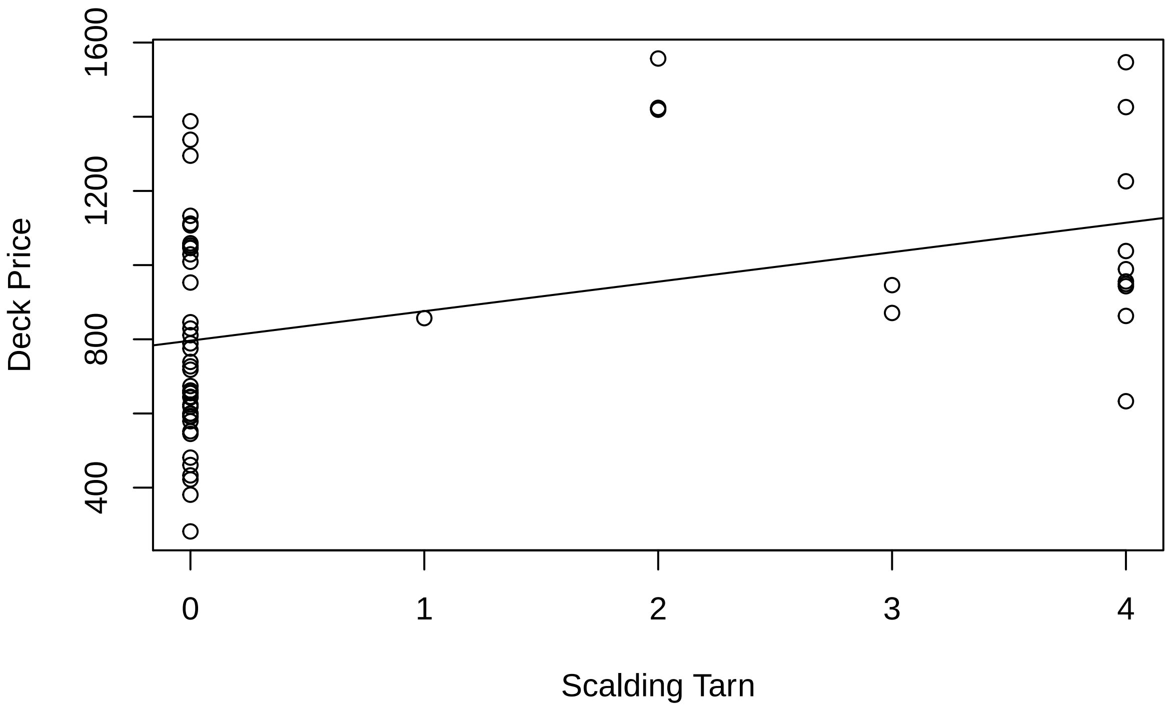


Figure 16: Scalding Tarn vs Deck Price

These Residuals appear to "hug" the line quite well, indicating normality. The Shapiro Wilk test produces $p - value = .011$, which is less than our threshold of $\alpha = .05$, so we have another solo predictor value that would not be good for regression alone.

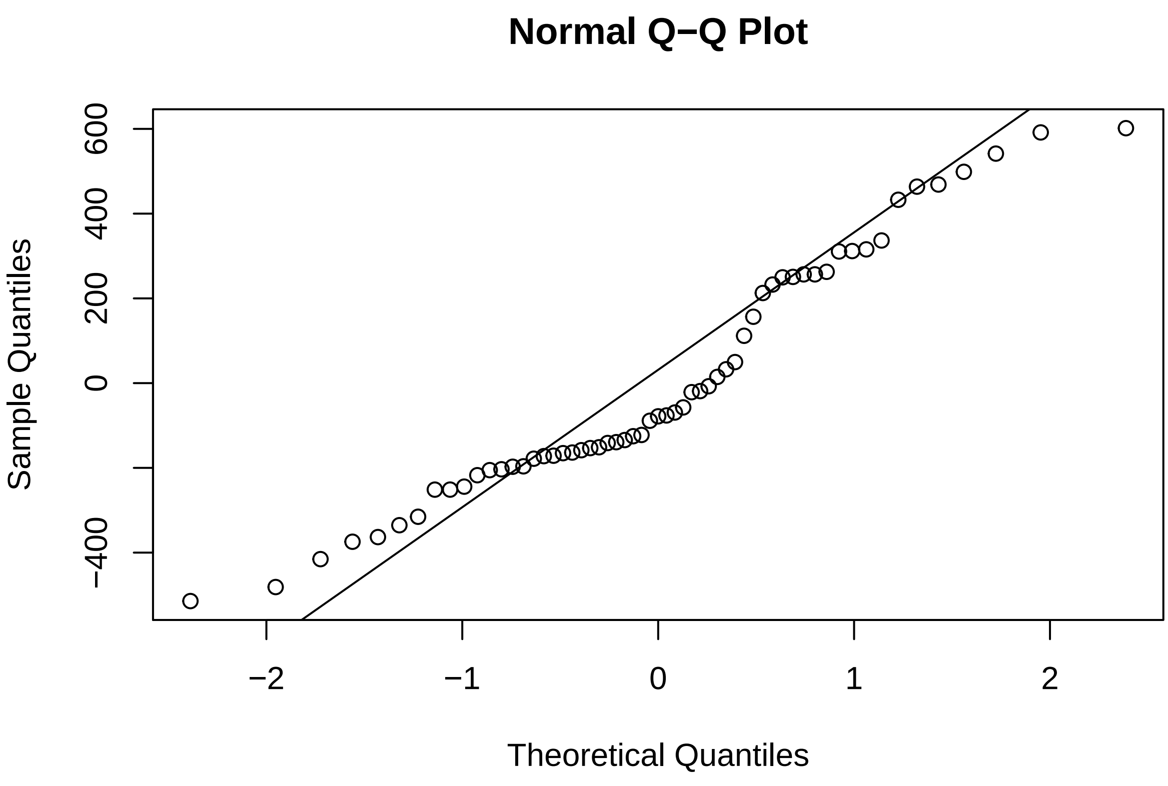


Figure 17: Scalding Tarn vs Deck Price Test for Normality

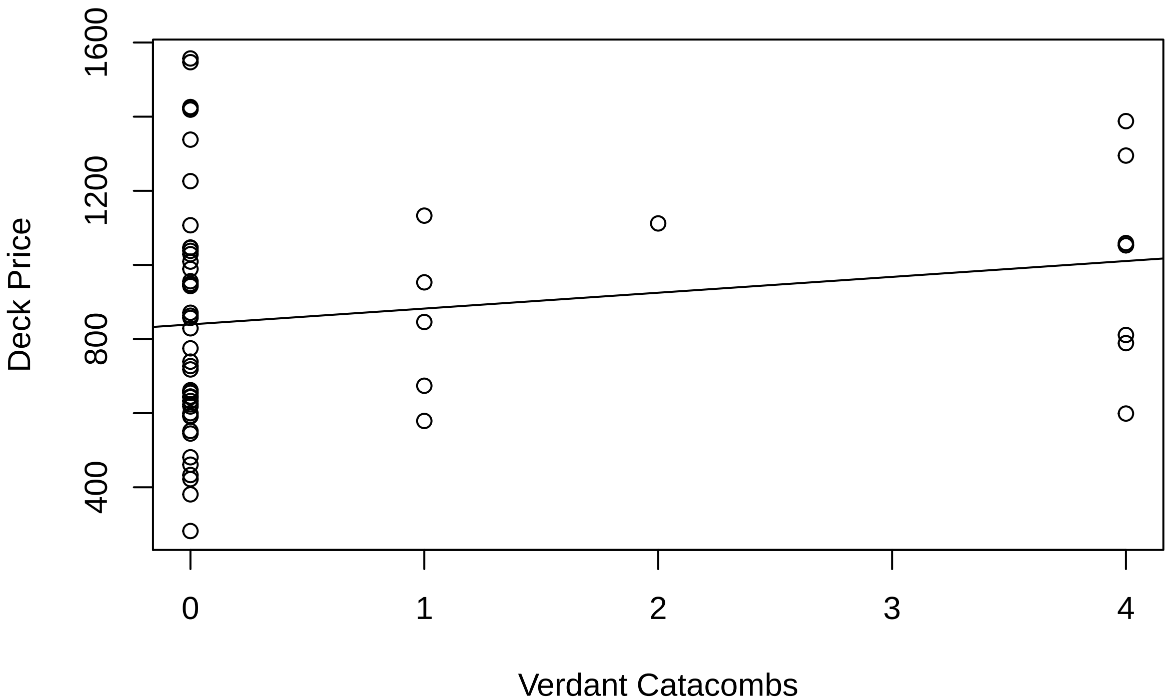


Figure 18: Verdant Catacombs vs Deck Price

These Residuals appear to "hug" the line quite well, indicating normality. The Shapiro Wilk test produces $p - value = .054$, which is barely more than our threshold of $\alpha = .05$, so we fail to reject normality. This data appears to be normal and regression analysis is "appropriate".

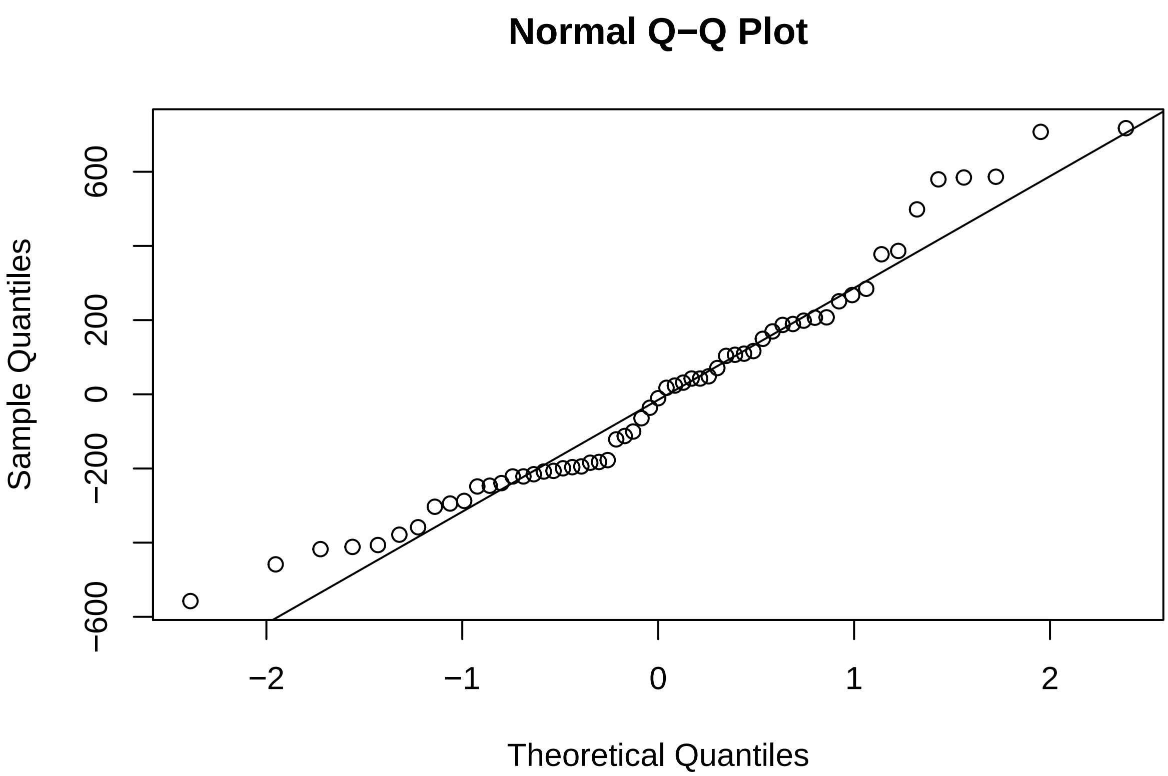


Figure 19: Verdant Catacombs vs Deck Price Test for Normality

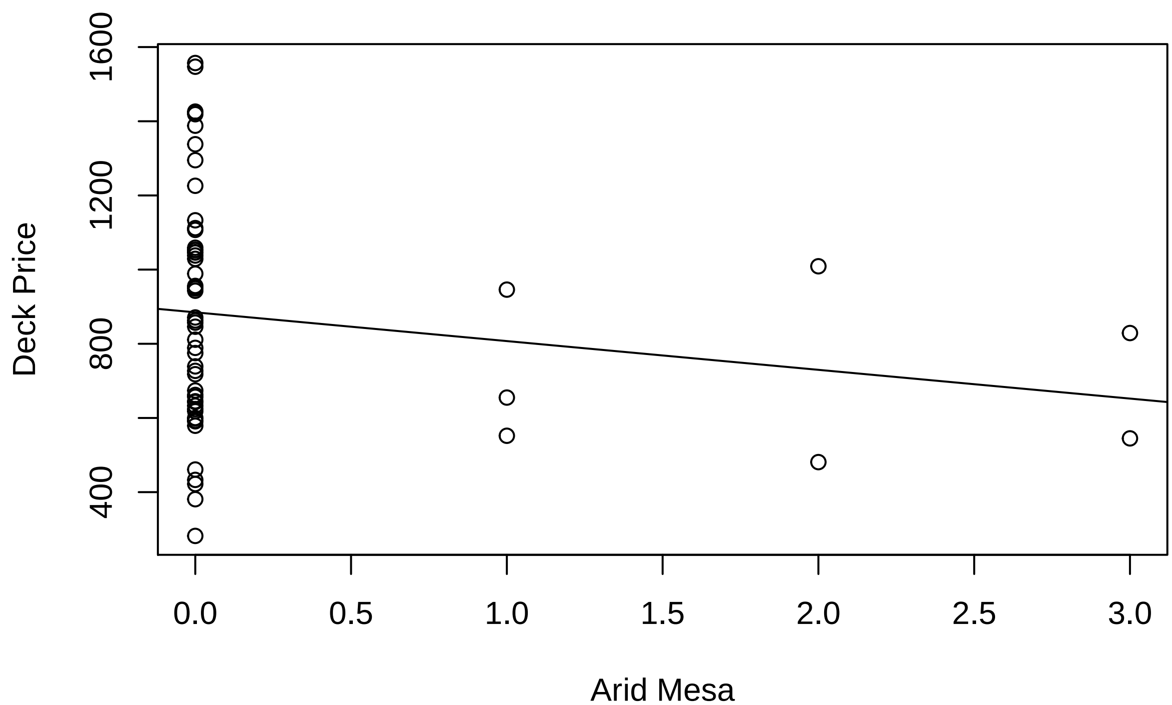


Figure 20: Arid Mesa vs Deck Price

These Residuals appear to "hug" the line quite well, indicating normality. The Shapiro Wilk test produces $p - value = .12$, which is more than our threshold of $\alpha = .05$, so we fail to reject normality. This data appears to be normal and regression analysis is appropriate.

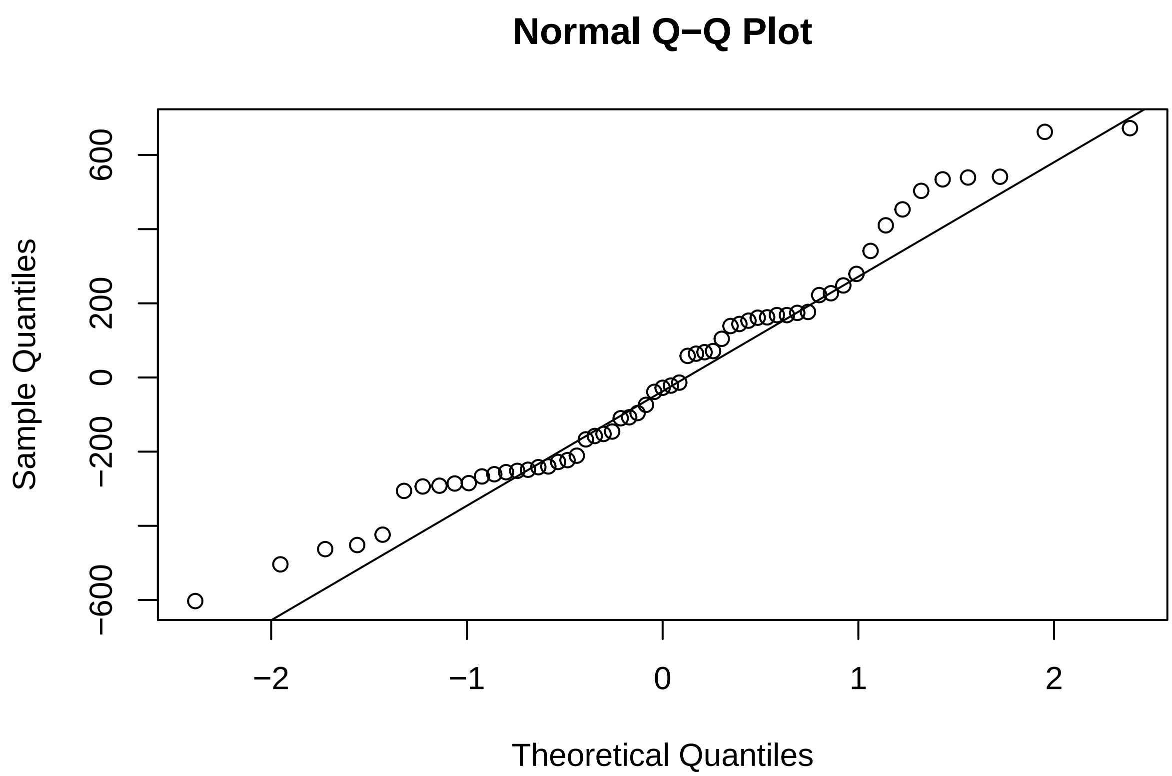


Figure 21: Arid Mesa vs Deck Price Test for Normality

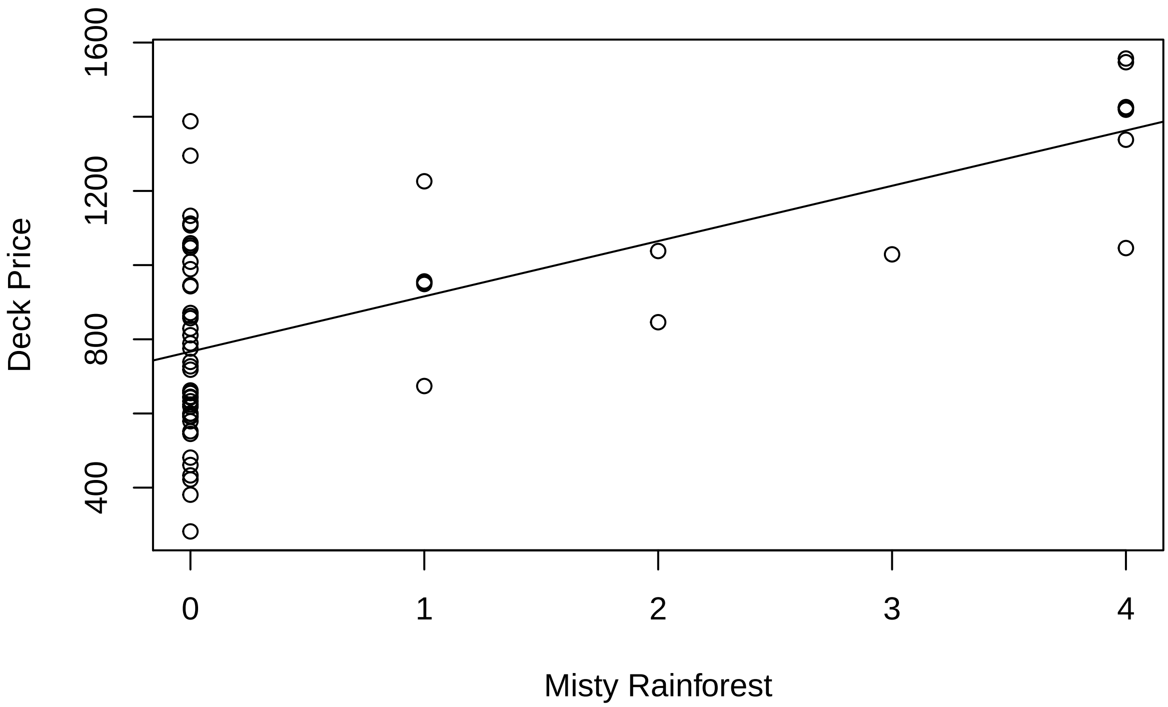


Figure 22: Misty Rainforest vs Deck Price

These Residuals appear to "hug" the line quite well, indicating normality. The Shapiro Wilk test produces $p-value = .54$, which is more than our threshold of $\alpha = .05$, so we fail to reject normality. This data appears to be normal and regression analysis is appropriate.

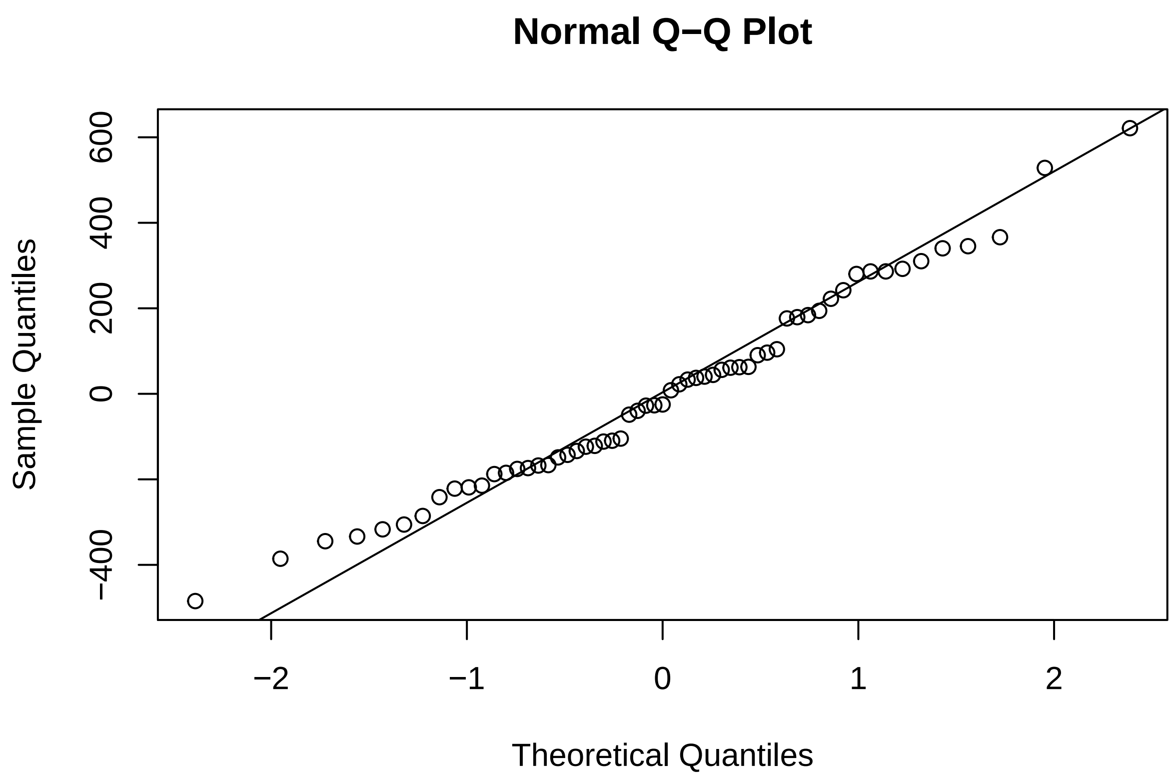


Figure 23: Misty Rainforest vs Deck Price Test for Normality

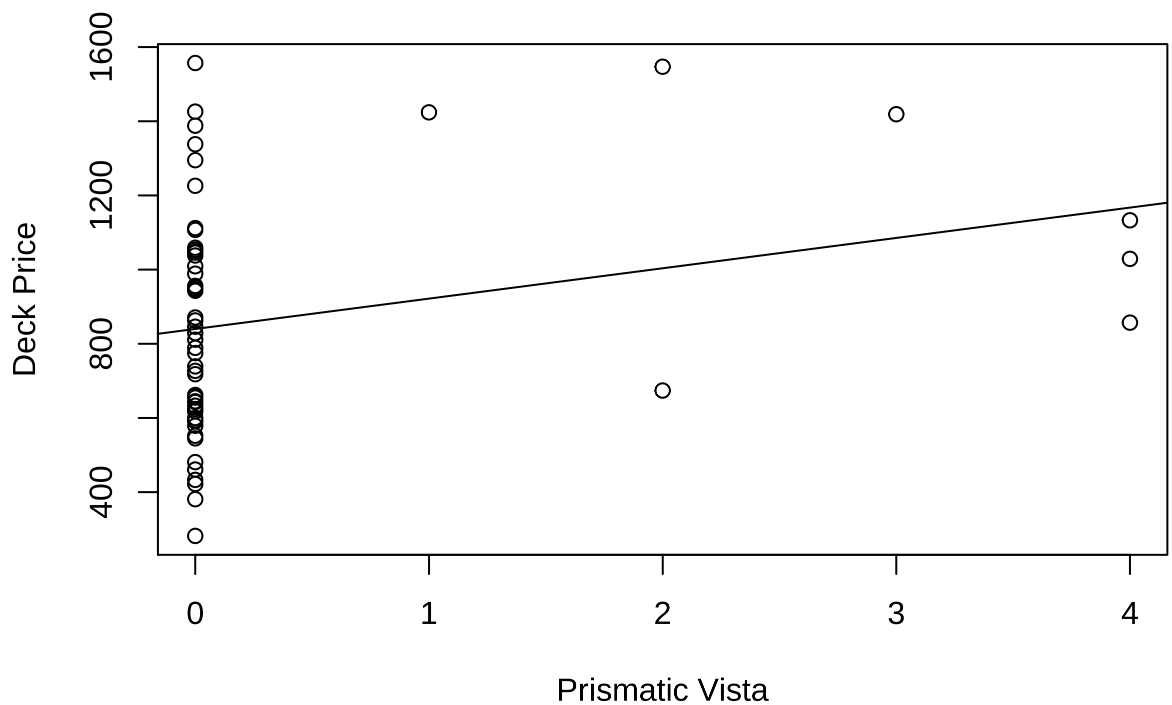


Figure 24: Prismatic Vista vs Deck Price

These Residuals appear to "hug" the line quite well, indicating normality. The Shapiro Wilk test produces $p - value = .11$, which is more than our threshold of $\alpha = .05$, so we fail to reject normality. This data appears to be normal and regression analysis is appropriate.

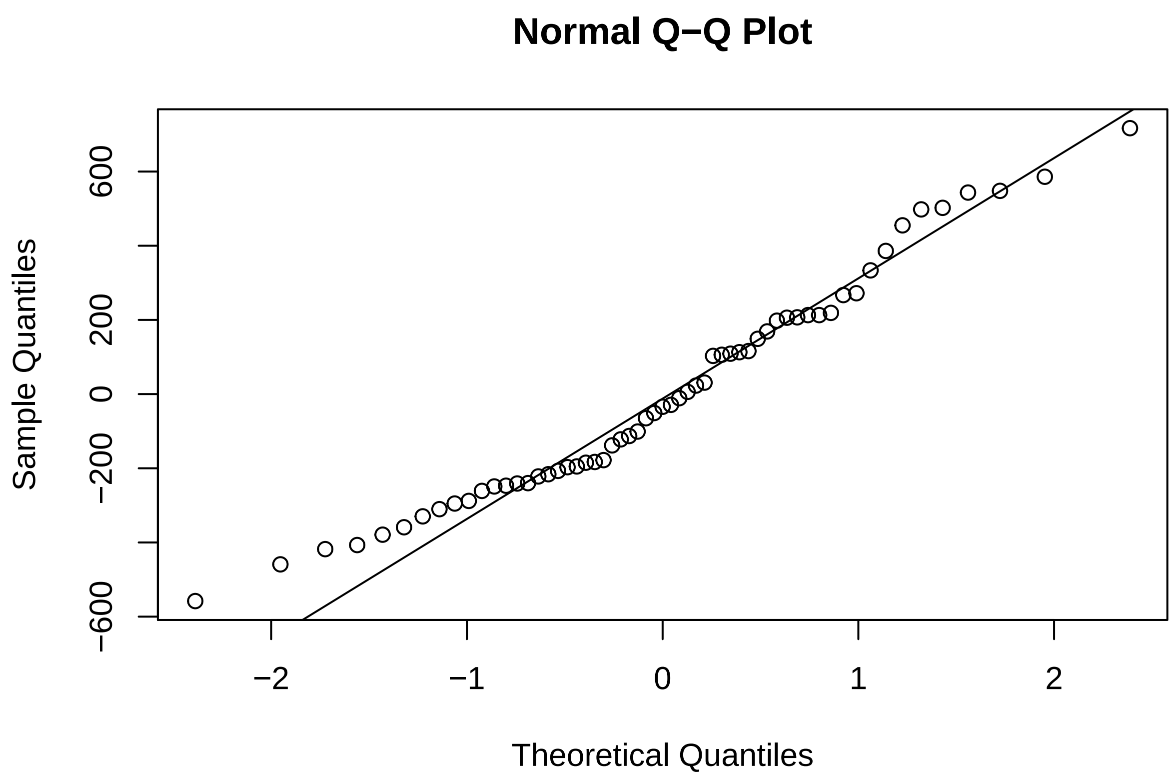


Figure 25: Prismatic Vista vs Deck Price Test for Normality

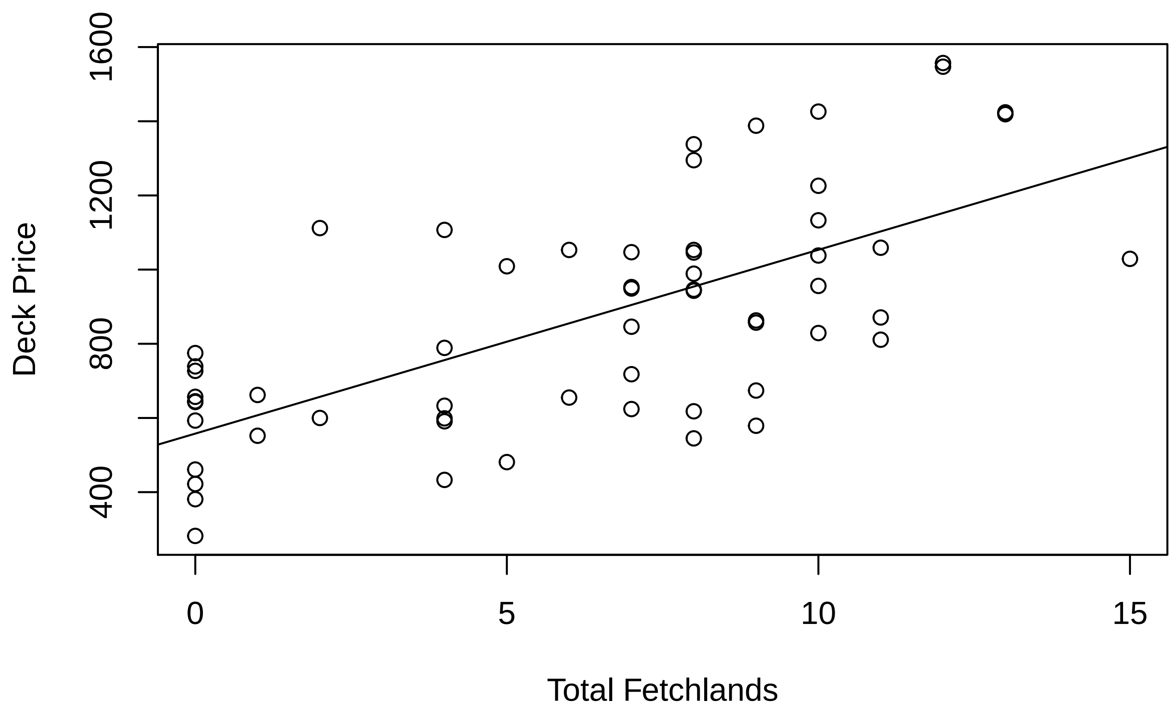


Figure 26: Total Fetchlands vs Deck Price

These Residuals appear to "hug" the line quite well, indicating normality. The Shapiro Wilk test produces $p - value = .2$, which is more than our threshold of $\alpha = .05$, so we fail to reject normality. This data appears to be normal and regression analysis is appropriate.

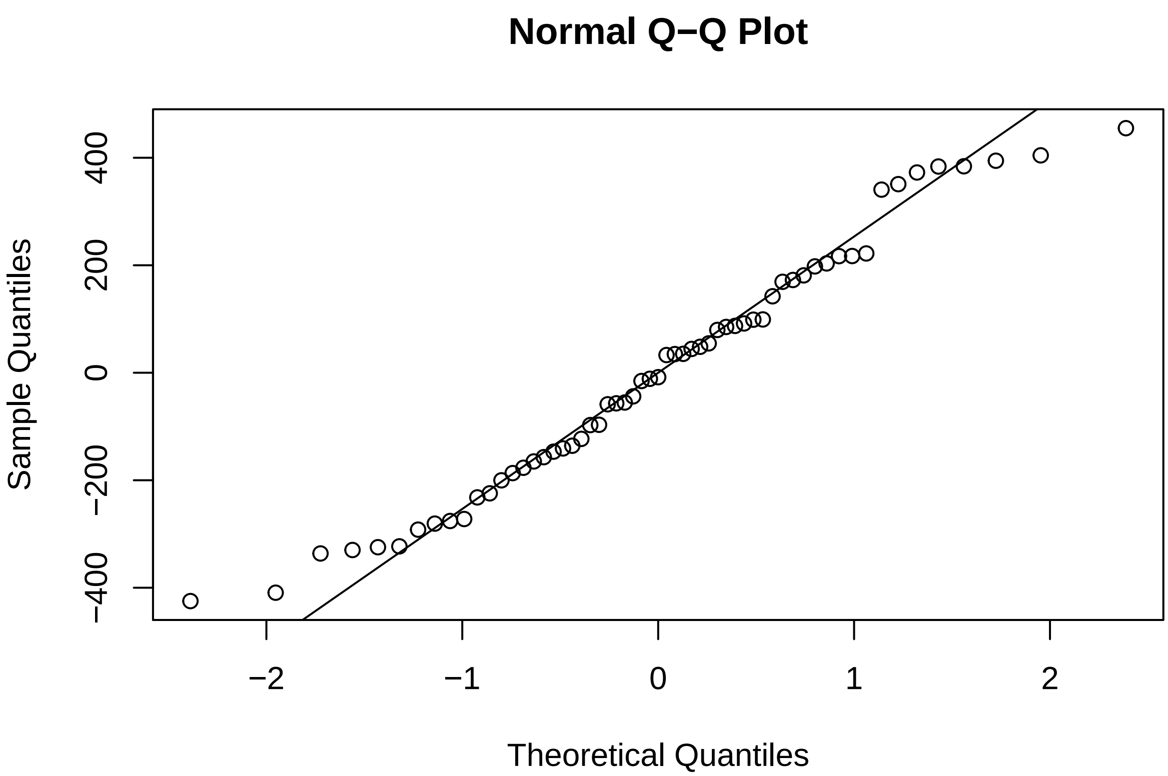


Figure 27: Total Fetchlands vs Deck Price Test for Normality

Below are the intercepts and coefficients of each of the 15 single predictor models. Again, special attention should be given to any R^2 values that larger, and any coefficients with p-values that are less than our threshold, $\alpha = .05$ (p-values are in the far right column in all of these tables). Beginning with Table 4.

Table 4: Archetype vs Deck Price

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	679.429	74.055	9.17463	0.00000
ArchetypeAggro-Combo	-127.429	286.814	-0.44429	0.65879
ArchetypeAggro-Control	121.071	135.206	0.89546	0.37492
ArchetypeAggro-Midrange	-218.429	286.814	-0.76157	0.44997
ArchetypeCombo	196.655	109.006	1.80407	0.07737
ArchetypeCombo-Control	-46.429	286.814	-0.16188	0.87207
ArchetypeControl	428.571	109.006	3.93162	0.00027
ArchetypeControl-Combo	508.071	209.460	2.42563	0.01901
ArchetypeMidrange	212.016	118.385	1.79090	0.07949
ArchetypeMidrange-Combo	349.571	286.814	1.21881	0.22875

The p-value for Combo and Control-Combo are both below the $\alpha = .05$ threshold, and are thus statistically significant values in the model. These dummy variables may be important to look for in the full model, later. It is also of note that the multiple R^2 value is .32, indicating a relationship between Archetype and Deck Price.

Table 5: Deck Count 1 Year vs Deck Price

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	904.75307	57.09517	15.84640	0.00000
`Deck Count 1 Year`	-0.70822	0.77448	-0.91445	0.36433

The p-value for Deck Count 1 Year is above the .05 threshold at .36. It would not be a good predictor variable on its own. $R^2 = .0145$, which is decidedly low, so again, no obvious linear relationship.

Table 6: Flooded Strand vs Deck Price

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	831.369	42.743	19.451	0.00000
`Flooded Strand`	59.617	28.362	2.102	0.03998

The p-value for Flooded Strand is below the threshold at .04. This means that there is a statistically significant relationship between the deck price and the number of Flooded Strand Fetchlands a deck has. It would not be a good predictor variable on its own, though, however, as $R^2 = .0719$

Table 7: Polluted Delta vs Deck Price

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	820.280	42.636	19.2390	0.00000
`Polluted Delta`	70.013	27.291	2.5654	0.01296

The p-value for Polluted Delta is below the threshold at .013. This means that there is a statistically significant relationship between the deck price and the number of Polluted Delta Fetchlands a deck has. It would not be a good predictor variable on its own, though, however, as $R^2 = .104$

Table 8: Bloodstained Mire vs Deck Price

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	854.181	44.785	19.07278	0.00000
`Bloodstained Mire`	24.253	34.574	0.70149	0.48585

The p-value for Bloodstained Mire is above the .05 threshold at .49. It would not be a good predictor variable on its own. $R^2 = .00856$, so no obvious linear relationship.

Table 9: Wooded Foothills vs Deck Price

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	867.76963	46.868	18.51535	0.00000
`Wooded Foothills`	-0.03275	32.199	-0.00102	0.99919

The p-value for Wooded Foothills is above the .05 threshold at 1. It would not be a good predictor variable on its own, and perhaps at all, considering how hight the p-value is. $R^2 = .00000001$, so no linear relationship.

Table 10: Windswept Heath vs Deck Price

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	882.250	45.461	19.40661	0.00000
`Windswept Heath`	-19.449	27.958	-0.69564	0.48948

The p-value for Windswept Heath is above the .05 threshold at .49. It would not be a good predictor variable on its own. $R^2 = .00842$, so no obvious linear relationship.

Table 11: Marsh Flats vs Deck Price

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	863.980	41.371	20.88372	0.00000
`Marsh Flats`	27.776	62.321	0.44568	0.65751

The p-value for Bloodstained Mire is above the .05 threshold at .66. It would not be a good predictor variable on its own. $R^2 = .00347$, so no obvious linear relationship.

Table 12: Scalding Tarn vs Deck Price

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	796.343	42.820	18.5976	0.00000
`Scalding Tarn`	79.486	23.799	3.3399	0.00148

The p-value for Scalding Tarn is below the threshold at .0015. This means that there is a statistically significant relationship between the deck price and the number of Scalding Tarn Fetchlands a deck has. It would not be a good predictor variable on its own, though, however, as $R^2 = .164$. It is, however, the strongest predictor so far.

Table 13: Verdant Catacombs vs Deck Price

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	839.486	44.192	18.9965	0.00000
`Verdant Catacombs`	42.752	29.001	1.4742	0.14593

The p-value for Verdant Catacombs is above the .05 threshold at .15. It would not be a good predictor variable on its own. $R^2 = .00367$, so no obvious linear relationship.

Table 14: Arid Mesa vs Deck Price

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	884.824	42.124	21.005	0.00000
`Arid Mesa`	-77.507	60.084	-1.290	0.20227

The p-value for Arid Mesa is above the .05 threshold at .20. It would not be a good predictor variable on its own. $R^2 = .0284$, so no obvious linear relationship.

Table 15: Misty Rainforest vs Deck Price

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	766.69	34.325	22.3362	0
`Misty Rainforest`	149.05	22.776	6.5443	0

The p-value for Misty Rainforest is below the threshold at .0000. There is a statistically significant relationship between the deck price and the number of Misty Fetchlands a deck has. It also appears to be a good predictor variable *on its own* as $R^2 = .429$.

Table 16: Prismatic Vista vs Deck Price

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	840.018	41.267	20.3556	0.00000
`Prismatic Vista`	81.798	39.017	2.0964	0.04049

The p-value for Prismatic Vista is below the threshold at .04. There is a statistically significant relationship between the deck price and the number of Prismatic Vista Fetchlands a deck has. It would not be a good predictor variable on its own, though, however, as $R^2 = .0716$.

Table 17: Total Fetchlands vs Deck Price

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	557.813	54.6578	10.2056	0
`Total Fetchlands`	49.556	7.2896	6.7981	0

The p-value for the Total number of Fetchlands a deck contains is below the threshold at .0000. There is a statistically significant relationship between the deck price and the number of Total Fetchlands a deck has. It also appears to be a good predictor variable *on its own* as $R^2 = .448$.

Now that each of the individual 14 individual predictor models has been assessed, it is time to consider a full model multiple linear regression analysis.

To begin, the full model would look as follows:

$$Y_i = b0 + b1X_{AgCom} + b2X_{AgCon} + b3X_{AgMid} + b4X_{Com}+$$

$$b5X_{ComCon} + b6X_{Con} + b7X_{ConCom} + b8X_{Mid} + b9X_{MidCom}+$$

$$b10X_{DC1Y} + b11X_{FS} + b12X_{PD} + b13X_{BM} + b14X_{WF}+$$

$$b15X_{WF} + b16X_{WH} + b17X_{MF} + b18X_{ST} + b19X_{VC}+$$

$$b20X_{AM} + b21X_{MR} + b22X_{PV} + b23X_{TotFetch}$$

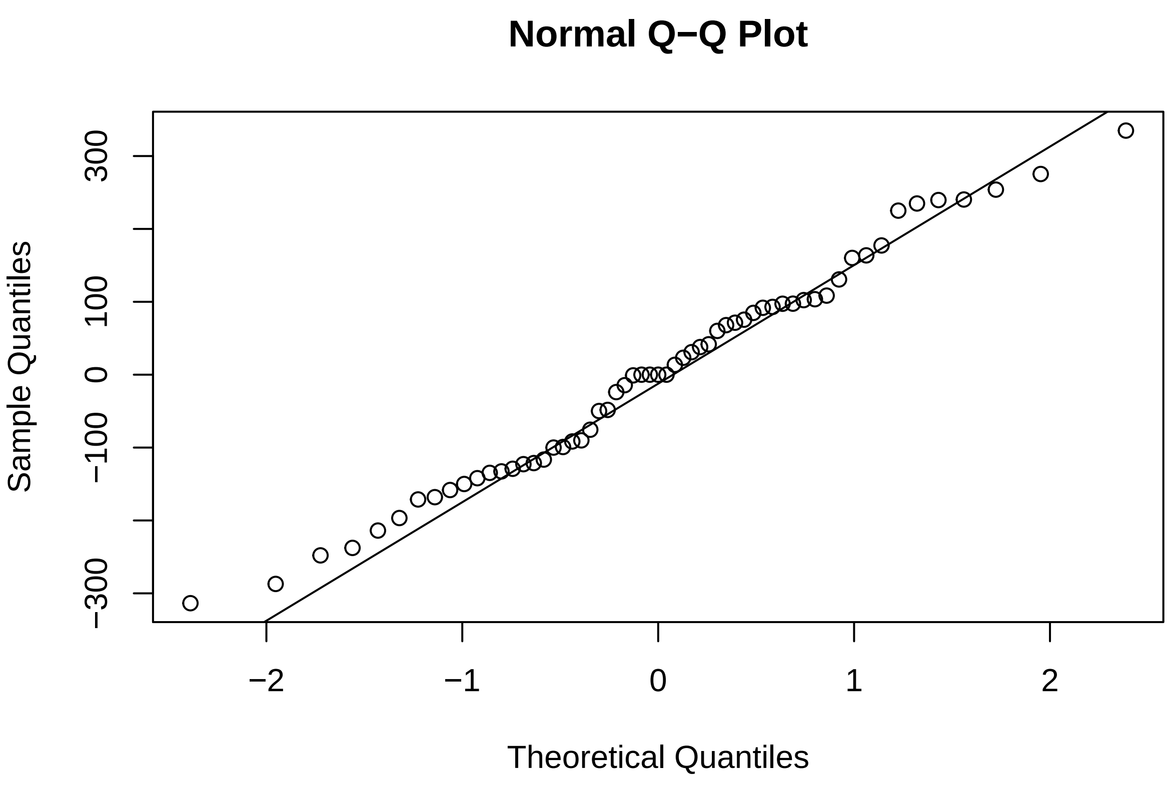


Figure 28: Full Model vs Deck Price Test for Normality

These Residuals appear to "hug" the line quite well, indicating normality for the full model. The Shapiro Wilk test produces $p - value = .68$, which is more than our threshold of $\alpha = .05$, so we fail to reject normality. This data appears to be normal and regression analysis is appropriate.

Table 18: Full Model vs Deck Price Intercept and Coefficients

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	581.42139	85.86562	6.77129	0.00000
ArchetypeAggro-Combo	-31.61419	206.21647	-0.15331	0.87899
ArchetypeAggro-Control	47.07363	114.30576	0.41182	0.68285
ArchetypeAggro-Midrange	-101.20621	205.75692	-0.49187	0.62572
ArchetypeCombo	78.73031	94.83445	0.83019	0.41176
ArchetypeCombo-Control	-63.32833	223.83662	-0.28292	0.77881
ArchetypeControl	207.62400	118.14419	1.75738	0.08713
ArchetypeControl-Combo	120.78536	192.44599	0.62763	0.53410
ArchetypeMidrange	138.31946	97.59988	1.41721	0.16479
ArchetypeMidrange-Combo	-156.87575	254.22090	-0.61708	0.54096
`Deck Count 1 Year`	-0.76861	0.57373	-1.33966	0.18853
`Flooded Strand`	15.26392	25.80897	0.59142	0.55784
`Polluted Delta`	2.34059	25.60426	0.09141	0.92766
`Bloodstained Mire`	64.29803	29.75649	2.16081	0.03726
`Wooded Foothills`	5.54349	30.81101	0.17992	0.85820
`Windswept Heath`	0.76733	24.75509	0.03100	0.97544
`Marsh Flats`	37.92028	45.59350	0.83170	0.41091
`Scalding Tarn`	34.49129	27.31763	1.26260	0.21463
`Verdant Catacombs`	65.01156	26.35622	2.46665	0.01839
`Arid Mesa`	17.56494	46.80967	0.37524	0.70962
`Misty Rainforest`	140.98602	23.45360	6.01127	0.00000
`Prismatic Vista`	41.94553	34.43235	1.21820	0.23086

The NA value in the last row indicates that the last predictor, Total Fetchlands, is linearly dependent on the other 13 variables, and can thus be dropped.

After this, the process of removing less effective predictors begins by "dropping off" the "bad predictors" by dropping the highest p-value (starting with Windswept Heath) and running the model again. Doing this process until the model contains only coefficents with p-values under the $\alpha = .05$ threshold.

Here is last table that includes only stastically significant values for the coefficents. Note that Scalding Tarn should be the next one to drop. However, doing so drops the R^2 value by .02, the largest loss yet. Also, the significance level of Scalding Tarn is very close to our .05 threshold at .0507. With it being this close, it should be kept in the model.

Table 19: Reduced Model vs Deck Price Intercept and Coefficients

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	548.1352	60.690	9.03176	0.00000
ArchetypeAggro-Combo	3.8648	193.384	0.01999	0.98414
ArchetypeAggro-Control	77.9053	99.863	0.78012	0.43940

	Estimate	Std. Error	t value	Pr(>|t|)
ArchetypeAggro-Midrange	-87.1352	193.384	-0.45058	0.65446
ArchetypeCombo	112.4114	86.320	1.30227	0.19945
ArchetypeCombo-Control	-89.2103	208.831	-0.42719	0.67128
ArchetypeControl	259.8950	94.688	2.74476	0.00867
ArchetypeControl-Combo	171.2029	160.580	1.06615	0.29204
ArchetypeMidrange	132.1598	90.003	1.46839	0.14896
ArchetypeMidrange-Combo	75.7381	204.030	0.37121	0.71222
`Bloodstained Mire`	70.6524	25.846	2.73359	0.00892
`Scalding Tarn`	43.5188	21.680	2.00734	0.05074
`Verdant Catacombs`	62.7462	22.283	2.81588	0.00720
`Misty Rainforest`	135.0423	21.393	6.31242	0.00000

Goodness of Fit Test

The goodness of fit analysis used include examining the residuals from the model, outlier detection, and the adjusted R-Squared values.

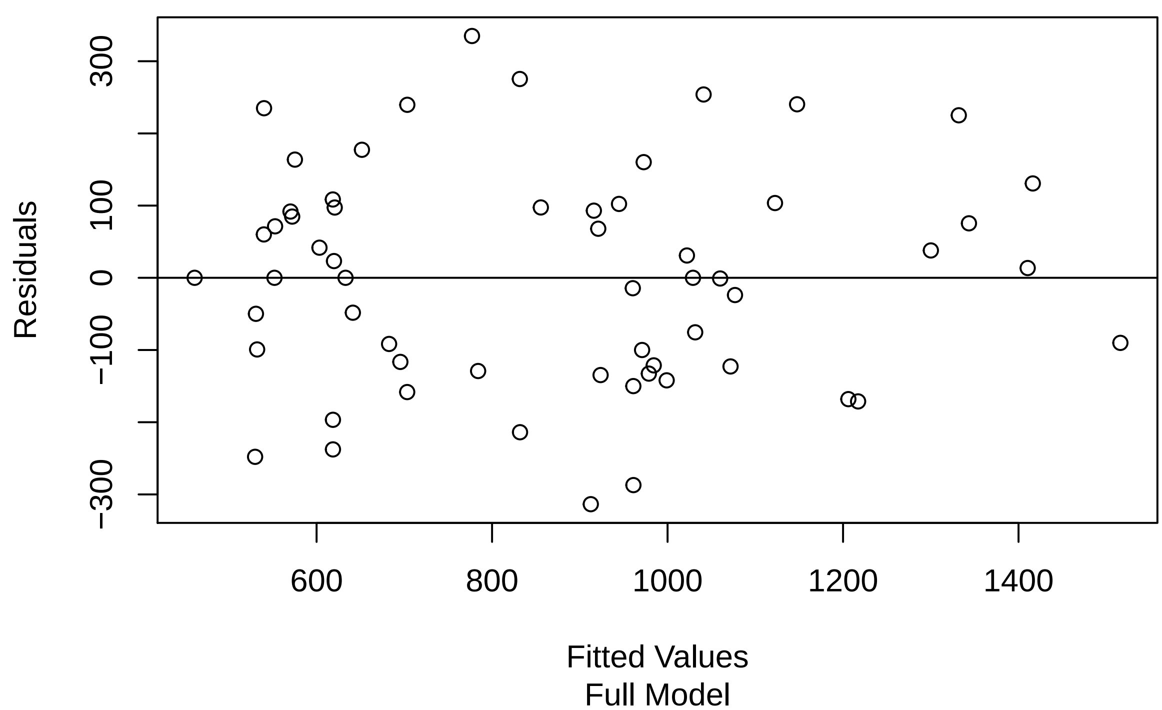


Figure 29: Fitted vs Residual graphs for the phases of the model.

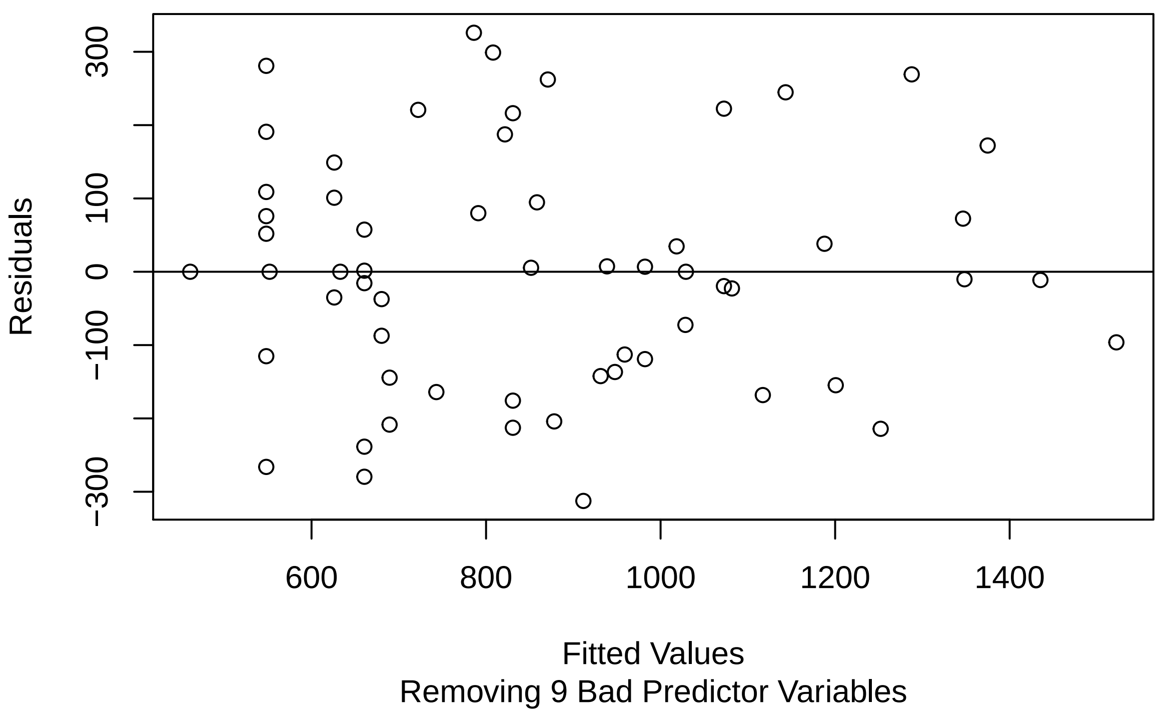


Figure 30: Fitted vs Residual graphs for the phases of the model.

The Fitted vs Residual plot for the original full model including all the predictor variables appeared to be realtively well-behaved. The homoscedasticity principle is not violated. The errors seem to be normally distributed and the $R^2 = .76$. There appear to be no outliers, so Cook's Distance is not necessary.

The Fitted vs Residual plot for the reduced version of the model also meet the requirements of regression analysis. The $R^2 = .726$

Results

For each of the variables, it is important to test if there is a linear association between the outcome variable, Deck Price, and the 13 predictor variables. This is done by using the t-test and applying the following hypothesis:

Null Hypothesis:: $H_0 : \beta_1 = 0$ Alternative Hypothesis:: $H_1 : \beta_1 \neq 0$

Stated another way, the Null Hypothesis is that there is no linear association between the Marketshare and the predictor variables, meaning that nothing can be predicted. The Alternative Hypothesis, on the other hand, states that there is a linear association between them.

Effect of Archetype on Deck Price With a significance level of $\alpha = 0.05$, multiple coefficients dictate success in rejecting the null hypothesis, H_0. This means that there is sufficient evidence that the Archetype that a person chooses to play is a predictor variable for the Price of their Deck. With an R-Squared value of only .032, this variable would only explain 32% of the variation in the model.

Effect of Deck Count 1 Year on Deck Price With a significance level of $\alpha = 0.05$, there is failure to reject null hypothesis, H_0. There is not sufficient evidence that the number of people playing a deck in a year can predict how much that deck costs. $R^2 = .0145$, the variable alone would only explain 1.45% of the variation in the model. Not an acceptable predictive model.

Effect of Flooded Strand on Deck Price With a significance level of $\alpha = 0.05$, reject the null hypothesis, H_0. There is sufficient evidence that the number of Flooded Strand Fetchlands in a deck can help predict how much that deck costs. $R^2 = .0719$, the variable alone would only explain 7.19% of the variation in the model. Not an acceptable predictive model alone.

Effect of Polluted Delta on Deck Price With a significance level of $\alpha = 0.05$, reject the null hypothesis, H_0. There is sufficient evidence that the number of Polluted Delta Fetchlands in a deck can help predict how much that deck costs. $R^2 = .104$, the variable alone would explain 10.4% of the variation in the model. Not an acceptable predictive model alone.

Effect of Bloodstained Mire on Deck Price With a significance level of $\alpha = 0.05$, fail to reject the null hypothesis, H_0. There is not sufficient evidence that the number of Bloodstained Mire Fetchlands in a deck can help predict how much that deck costs. $R^2 = .00856$, the variable alone would only explain .856% of the variation in the model. Terrible as a predictive model alone.

Effect of Wooded Foothills on Deck Price With a significance level of $\alpha = 0.05$, fail to reject the null hypothesis, H_0. There is not sufficient evidence that the number of Wooded Foothills Fetchlands in a deck can help predict how much that deck costs. $R^2 = .0000$, the variable alone would explain 0.00% of the variation in the model. This means the presence of a Wooded Foothills in a deck alone yields no predictive information.

Effect of Windswept Heath on Deck Price With a significance level of $\alpha = 0.05$, fail to reject the null hypothesis, H_0. There is not sufficient evidence that the number of Windswept Heath Fetchlands in a deck can help predict how much that deck costs. $R^2 = .00842$, the variable alone would only explain .842% of the variation in the model. Terrible as a predictive model alone.

Effect of Marsh Flats on Deck Price With a significance level of $\alpha = 0.05$, fail to reject the null hypothesis, H_0. There is not sufficient evidence that the number of Marsh Flats Fetchlands in a deck can help predict how much that deck costs. $R^2 = .00347$, the variable alone would only explain .347% of the variation in the model. Terrible as a predictive model alone.

Effect of Scalding Tarn on Deck Price With a significance level of $\alpha = 0.05$, reject the null hypothesis, H_0. There is sufficient evidence that the number of Scalding Tarn

Fetchlands in a deck can help predict how much that deck costs. $R^2 = .164$, the variable alone would only explain 16.4% of the variation in the model. Not a strong predictor alone, but worth noting.

Effect of Verdant Catacombs on Deck Price With a significance level of $\alpha = 0.05$, fail to reject the null hypothesis, H_0. There is not sufficient evidence that the number of Verdant Catacomb Fetchlands in a deck can help predict how much that deck costs. $R^2 = .0367$, the variable alone would only explain 3.67% of the variation in the model. Terrible as a predictive model alone. Oh how Jund has fallen.

Effect of Arid Mesa on Deck Price With a significance level of $\alpha = 0.05$, fail to reject the null hypothesis, H_0. There is not sufficient evidence that the number of Arid Mesa Fetchlands in a deck can help predict how much that deck costs. $R^2 = .0284$, the variable alone would only explain 2.84% of the variation in the model.

Effect of Misty Rainforest on Deck Price With a significance level of $\alpha = 0.05$, reject the null hypothesis, H_0. There is sufficient evidence that the number of Misty Rainforest Fetchlands in a deck can help predict how much that deck costs. $R^2 = .429$, the variable alone would explain *42.9%* of the variation in the model. This is powerful predictor variable on its own.

Effect of Prismatic Vista on Deck Price With a significance level of $\alpha = 0.05$, reject the null hypothesis, H_0. There is sufficient evidence that the number of Prismatic Vista Fetchlands in a deck can help predict how much that deck costs. $R^2 = .0716$, the variable alone would explain 7.16% of the variation in the model. Not great as a solo predictor model, but worth considering for the full model. Not bad for a card that just came out.

Effect of Total Fetchlands on Deck Price With a significance level of $\alpha = 0.05$, reject the null hypothesis, H_0. There is sufficient evidence that the number of Fetchlands in a deck can help predict how much that deck costs. $R^2 = .448$, the variable alone would explain *44.8%* of the variation in the model. The total number of Fetchlands in a deck is thus a powerful predictor variable on its own.

Primary Objective Results

Bottom line: The results from the t-test indicate that there is a linear association between Deck Price and the following: the **Archtype** you play and the number of **Bloodstained Mires**, **Scalding Tarns**, **Verdant Catacombs**, and **Misty Rainforests** your deck has in it. To find the most effective predictor variable, the coefficient of determination, R^2, is used. This value indicates what percentage of the variance in the outcome can be predicted by the model. The R-Squared value ranged from 0 to 1, with a higher value indicating a higher predictive association/responsibility for the model. For example, a linear model with an R^2 value of .92 can be said to be 92% responsible for the outcome of the real outcome, Y.

That said, the most influential predictor variables for the model in order are: 1. The number of Misty Rainforest 2. If the deck is Control or contains Bloodstained Mire or Verdant Catacombs 3. The number of Scalding Tarn

Discussion and Conclusion

The long and short of this study is basically that, yes, Fetchlands can predict the price of a deck, to a degree. It is beyond the scope of statistics and this study to discuss what should or should not happen moving forward with the game, but this information is a powerful tool for anyone looking to predict the cost of a deck, should that person decide to play in the Modern format.

The most statistically significant predictor on the cost of a deck, or said another way, "the thing that really predicts a deck's price the most accurately", is the presence of a Misty Rainforest. While the conventional wisdom is to talk about the price of Scalding Tarn, this study is not disproving or proving that cards are "too expensive" or not. That is an emotional topic, and one that statistics cannot answer.

This study is only concerned with the most predictive information. Using the model that will be introduced in a moment, a person interested in building a deck with 3 Misty Rainforests will have a statistically stronger chance of predicting the price of the deck that player wants to build than a player that builds a deck with 3 Scalding Tarn in it.

To find the most effective predictor variable, the coefficient of determination, R^2, is used. This value will indicates what percentage of the Deck Price variation can be attributed to this model. In this case, the R^2 value is .726. Or, said another way, 72.6% of the Deck Price variation can be attributed to *these* predictors.

So, without further delay, here is the model:

$$Y_i = 518.14 + 3.86X_{AgCom} + 77.912X_{AgCon} - 87.14X_{AgMid} + 112.41X_{Com} +$$

$$-89.21X_{ComCon} + 259.89X_{Con} + 171.20X_{ConCom} + 132.16X_{Mid} + 75.74X_{MidCom} +$$

$$70.65X_{BM} + 43.52X_{ST} + 62.75X_{VC} + 135.04X_{MR}$$

Let's look at an example.

Suppose that you are a Control player (I'm sorry). You choose a deck that includes 4 Flooded Strand and 2 Scalding Tarn. How much would the deck cost?

The first number in the equation is our baseline. On average, we can assume knowing nothing else, that a "could-be-anything" Modern deck is going to cost 518.14 USD. It will go up or down from there, based upon the decisions selected above. To start, Control was chosen as the deck Archetype. There are only 2 values that can be entered into Archetype variables in the equation, 0 or 1. 0 indicates that that archetype is not chosen. 1 indicated that it is chosen.

From here, we consider the Fetchland information. Based on this model and this study, the number of Flooded Strand in a player's deck does not help predict the price of that deck. So that information can be ignored. However, having 2 Scalding tarn in the deck does affect the model prediction.

The following the predicted price of the deck based upont the scenario:

$$Y_i = 518.14 + 3.86(0) + 77.912(0) - 87.14(0) + 112.41(0) +$$

$$-89.21(0) + 259.89(1) + 171.20(0) + 132.16(0) + 75.74(0) +$$

$$70.65(0) + 43.52(2) + 62.75(0) + 135.04(0)$$

Everytying being mulitplied by zero goes away, and the final predicted price, after adding up the 3 remaining terms, is $Y_1 = 865.07$ USD.

As we can see, the price of a Modern Format deck can be predicted using the model above.

But, are Fetchlands too expensive? Who knows. Maybe. That's an opinion. But they should reprint them. Cowards.

Appendix: R-code

```
rm(list = ls(all=TRUE))

my_data <- read.csv("Modern_Meta_6.21.2020.csv", header = T, sep = ',')

# Also we'll cut out the unnecessary attributes
# for our analysis

my_data <- my_data[, c(1:5, 9, 10, 12, 14, 16, 18,
                     20, 22, 24, 26, 28, 30, 32:35)]

colnames(my_data) <- c("Deck Name", "Archetype", "Format", "Deck Count 1 Year", "Percent of Me

# add the format "modern" to all entries in the format category
my_data$Format[my_data$Format == ''] <- "Modern"

# combine two observations that are
# the same deck that were accidentlly
# recorded separately.
# observation 53 combines with 27
# 9 decks added to observation 27
my_data[27, 4] <- my_data[27, 4] + 9
# observation 53 deleted
my_data <- my_data[-53, ]
# re index the rows
row.names(my_data) <- 1:59
# Check to see how many different archtypes
# were recorded
levels(my_data$Archetype)
# combining both "Aggro Control" Variants
my_data$Archetype[my_data$Archetype == "Aggro - Control"] <- "Aggro-Control"

# renaming the mis-archetyped "Aggro-Tempo"
# To "Aggro-Control"
my_data$Archetype[my_data$Archetype == "Aggro-Tempo"] <- "Aggro-Control"

# drop the unused level
my_data$Archetype <- droplevels(my_data$Archetype)

#checked to see if it worked
levels(my_data$Archetype)

# now add deck archtypes to the 5
# missing a deck archetype entry
# by researching deck
my_data[17,2] <- "Midrange"
```

```
my_data[31,2] <- "Combo"
my_data[38,2] <- "Control"
my_data[45,2] <- "Control"
my_data[59,2] <- "Control"

# See how many of each archetype
# was played this last year
summary(my_data$Archetype)
############### Initial data wrangling and cleaning

\subsubsection{Secondary Objective Analysis}

To begin the analysis, consider the relationship of the Deck Prices of the top 59 most played 


ggplot(data = my_data) +
  geom_point(aes(‘Total Fetchlands‘,
                 ‘Deck Price‘,
                 color = Archetype)) +
  ggtitle("Do Fetchlands Increase Deck Price?") +
  geom_smooth(method = "lm",
              color = 'black',
              data = my_data,
              aes(‘Total Fetchlands‘, ‘Deck Price‘)) +
  theme_bw()

# Interactive version below
#
# interactive_plot <- ggplot(data = my_data) +
#   geom_point_interactive(aes(‘Total Fetchlands‘,
#                              ‘Deck Price‘,
#                              color = Archetype,
#                              tooltip = ‘Deck Name‘)) +
#   ggtitle("Do Fetchlands Increase Deck Price?") +
#   geom_smooth(method = "lm",
#              color = 'black',
#              data = my_data,
#              aes(‘Total Fetchlands‘, ‘Deck Price‘)) +
#   theme_bw()
#
# ggiraph(code = print(interactive_plot))




kable(descr(my_data[, c(4:6,19:21)]), caption = "Summary Statistics of the Data Part 1")
```

```
kable(descr(my_data[,7:12]), caption = "Summary Statistics of the Data Part 2")

kable(descr(my_data[,13:18]), caption = "Summary Statistics of the Data Part 3")




# Regression models
# Plot of single predictor variable against outcome variable
# trendline
# summary
# tests for normality qqnorm, qqline, shapiro.test

############### Archetype, Linear Model, Test for Normality
m1 <- lm(‘Deck Price‘ ~ ‘Archetype‘, my_data)

ggplot(my_data, aes(x = ‘Archetype‘, y = ‘Deck Price‘)) +
  geom_boxplot() +
  xlab("Archetype") +
  theme(text = element_text(size=10),
        axis.text.x = element_text(angle=90, hjust=1)) +
  ylab("Deck Price (USD)")




############### Deck Count 1 Year, Linear Model, Test for Normality
m2 <- lm(‘Deck Price‘ ~ ‘Deck Count 1 Year‘, my_data)

with(my_data, plot(‘Deck Count 1 Year‘, ‘Deck Price‘))
abline(m2)


qqnorm(residuals(m2))
qqline(residuals(m2))



shapiro.test(residuals(m2))
summary(m2)
############### Deck Count 1 Year, Linear Model, Test for Normality




############### Flooded strand, Linear Model, Test for Normality
m3 <- lm(‘Deck Price‘ ~ ‘Flooded Strand‘, my_data)
```

```
with(my_data, plot(`Flooded Strand`, `Deck Price`))
abline(m3)


qqnorm(residuals(m3))
qqline(residuals(m3))


shapiro.test(residuals(m3))
summary(m3)
############### Flooded strand, Linear Model, Test for Normality




############### Polluted Delta, Linear Model, Test for Normality
m4 <- lm(`Deck Price` ~ `Polluted Delta`, my_data)

with(my_data, plot(`Polluted Delta`, `Deck Price`))
abline(m4)


qqnorm(residuals(m4))
qqline(residuals(m4))



shapiro.test(residuals(m4))
summary(m4)
############### Polluted Delta, Linear Model, Test for Normality




############### Bloodstained Mire, Linear Model, Test for Normality
m5 <- lm(`Deck Price` ~ `Bloodstained Mire`, my_data)

with(my_data, plot(`Bloodstained Mire`, `Deck Price`))
abline(m5)


qqnorm(residuals(m5))
qqline(residuals(m5))


shapiro.test(residuals(m5))
```

```
summary(m5)
############### Bloodstained Mire, Linear Model, Test for Normality




############### Wooded Foothills, Linear Model, Test for Normality
m6 <- lm(`Deck Price` ~ `Wooded Foothills`, my_data)

with(my_data, plot(`Wooded Foothills`, `Deck Price`))
abline(m6)


qqnorm(residuals(m6))
qqline(residuals(m6))


shapiro.test(residuals(m6))
summary(m6)
############### Wooded Foothills, Linear Model, Test for Normality




############### Windswept Heath, Linear Model, Test for Normality
m7 <- lm(`Deck Price` ~ `Windswept Heath`, my_data)

with(my_data, plot(`Windswept Heath`, `Deck Price`))
abline(m7)



qqnorm(residuals(m7))
qqline(residuals(m7))


shapiro.test(residuals(m7))
summary(m7)
############### Windswept Heath, Linear Model, Test for Normality




############### Marsh Flats, Linear Model, Test for Normality
m8 <- lm(`Deck Price` ~ `Marsh Flats`, my_data)

with(my_data, plot(`Marsh Flats`, `Deck Price`))
```

```
abline(m8)


qqnorm(residuals(m8))
qqline(residuals(m8))


shapiro.test(residuals(m8))
summary(m8)
############### Marsh Flats, Linear Model, Test for Normality




############### Scalding Tarn, Linear Model, Test for Normality
m9 <- lm(`Deck Price` ~ `Scalding Tarn`, my_data)

with(my_data, plot(`Scalding Tarn`, `Deck Price`))
abline(m9)


qqnorm(residuals(m9))
qqline(residuals(m9))


shapiro.test(residuals(m9))
summary(m9)
############### Scalding Tarn, Linear Model, Test for Normality




############### Verdant Catacombs, Linear Model, Test for Normality
m10 <- lm(`Deck Price` ~ `Verdant Catacombs`, my_data)

with(my_data, plot(`Verdant Catacombs`, `Deck Price`))
abline(m10)



qqnorm(residuals(m10))
qqline(residuals(m10))


shapiro.test(residuals(m10))
summary(m10)
############### Verdant Catacombs, Linear Model, Test for Normality
```

```
############### Arid Mesa, Linear Model, Test for Normality
m11 <- lm(‘Deck Price‘ ~ ‘Arid Mesa‘, my_data)

with(my_data, plot(‘Arid Mesa‘, ‘Deck Price‘))
abline(m11)


qqnorm(residuals(m11))
qqline(residuals(m11))


shapiro.test(residuals(m11))
summary(m11)
############### Arid Mesa, Linear Model, Test for Normality




############### Misty Rainforest, Linear Model, Test for Normality
m12 <- lm(‘Deck Price‘ ~ ‘Misty Rainforest‘, my_data)

with(my_data, plot(‘Misty Rainforest‘, ‘Deck Price‘))
abline(m12)


qqnorm(residuals(m12))
qqline(residuals(m12))


shapiro.test(residuals(m12))
summary(m12)
############### Misty Rainforest, Linear Model, Test for Normality




############### Prismatic Vista, Linear Model, Test for Normality
m13 <- lm(‘Deck Price‘ ~ ‘Prismatic Vista‘, my_data)

with(my_data, plot(‘Prismatic Vista‘, ‘Deck Price‘))
abline(m13)
```

```
qqnorm(residuals(m13))
qqline(residuals(m13))


shapiro.test(residuals(m13))
summary(m13)
############### Prismatic Vista, Linear Model, Test for Normality




############### Total Fetchlands, Linear Model, Test for Normality
m14 <- lm(‘Deck Price‘ ~ ‘Total Fetchlands‘, my_data)

with(my_data, plot(‘Total Fetchlands‘, ‘Deck Price‘))
abline(m14)


qqnorm(residuals(m14))
qqline(residuals(m14))


shapiro.test(residuals(m14))
summary(m14)
############### Total Fetchlands, Linear Model, Test for Normality




kable(summary(m1)$coefficients, caption = "Archetype vs Deck Price")
kable(summary(m2)$coefficients, caption = "Deck Count 1 Year vs Deck Price")
kable(summary(m3)$coefficients, caption = "Flooded Strand vs Deck Price")
kable(summary(m4)$coefficients, caption = "Polluted Delta vs Deck Price")
kable(summary(m5)$coefficients, caption = "Bloodstained Mire vs Deck Price")
kable(summary(m6)$coefficients, caption = "Wooded Foothills vs Deck Price")
kable(summary(m7)$coefficients, caption = "Windswept Heath vs Deck Price")
kable(summary(m8)$coefficients, caption = "Marsh Flats vs Deck Price")
kable(summary(m9)$coefficients, caption = "Scalding Tarn vs Deck Price")
kable(summary(m10)$coefficients, caption = "Verdant Catacombs vs Deck Price")
kable(summary(m11)$coefficients, caption = "Arid Mesa vs Deck Price")
kable(summary(m12)$coefficients, caption = "Misty Rainforest vs Deck Price")
kable(summary(m13)$coefficients, caption = "Prismatic Vista vs Deck Price")
kable(summary(m14)$coefficients, caption = "Total Fetchlands vs Deck Price")
```

```
############### Full Model, Linear Model, Test for Normality
full_model <- lm(‘Deck Price‘ ~ ‘Archetype‘ + ‘Deck Count 1 Year‘ +
                 ‘Flooded Strand‘ + ‘Polluted Delta‘ +
                 ‘Bloodstained Mire‘ + ‘Wooded Foothills‘ +
                 ‘Windswept Heath‘ + ‘Marsh Flats‘ +
                 ‘Scalding Tarn‘ + ‘Verdant Catacombs‘ +
                 ‘Arid Mesa‘ + ‘Misty Rainforest‘ +
                 ‘Prismatic Vista‘ + ‘Total Fetchlands‘,
               data = my_data)

qqnorm(residuals(full_model))
qqline(residuals(full_model))

shapiro.test(residuals(full_model))
summary(full_model)
############### Full Model, Linear Model, Test for Normality



############### Reducing Model
reduced_model_1 <- lm(‘Deck Price‘ ~ ‘Archetype‘ + ‘Deck Count 1 Year‘ +
                      ‘Flooded Strand‘ + ‘Polluted Delta‘ +
                      ‘Bloodstained Mire‘ + ‘Wooded Foothills‘ +
                      ‘Marsh Flats‘ +
                      ‘Scalding Tarn‘ + ‘Verdant Catacombs‘ +
                      ‘Arid Mesa‘ + ‘Misty Rainforest‘ +
                      ‘Prismatic Vista‘,
                    data = my_data)
# Windswept Heath

reduced_model_2 <- lm(‘Deck Price‘ ~ ‘Archetype‘ + ‘Deck Count 1 Year‘ +
                      ‘Flooded Strand‘ +
                      ‘Bloodstained Mire‘ + ‘Wooded Foothills‘ +
                      ‘Marsh Flats‘ +
                      ‘Scalding Tarn‘ + ‘Verdant Catacombs‘ +
                      ‘Arid Mesa‘ + ‘Misty Rainforest‘ +
                      ‘Prismatic Vista‘,
                    data = my_data)
# Polluted Delta

reduced_model_3 <- lm(‘Deck Price‘ ~ ‘Archetype‘ + ‘Deck Count 1 Year‘ +
                      ‘Flooded Strand‘ +
                      ‘Bloodstained Mire‘ +
                      ‘Marsh Flats‘ +
                      ‘Scalding Tarn‘ + ‘Verdant Catacombs‘ +
                      ‘Arid Mesa‘ + ‘Misty Rainforest‘ +
```

```
                     ‘Prismatic Vista‘,
                   data = my_data)
# Wooded Foothills

reduced_model_4 <- lm(‘Deck Price‘ ~ ‘Archetype‘ + ‘Deck Count 1 Year‘ +
                     ‘Flooded Strand‘ +
                     ‘Bloodstained Mire‘ +
                     ‘Marsh Flats‘ +
                     ‘Scalding Tarn‘ + ‘Verdant Catacombs‘ +
                     ‘Misty Rainforest‘ +
                     ‘Prismatic Vista‘,
                   data = my_data)
# Arid Mesa

reduced_model_5 <- lm(‘Deck Price‘ ~ ‘Archetype‘ + ‘Deck Count 1 Year‘ +
                     ‘Bloodstained Mire‘ +
                     ‘Marsh Flats‘ +
                     ‘Scalding Tarn‘ + ‘Verdant Catacombs‘ +
                     ‘Misty Rainforest‘ +
                     ‘Prismatic Vista‘,
                   data = my_data)
# Flooded Strand

reduced_model_6 <- lm(‘Deck Price‘ ~ ‘Archetype‘ + ‘Deck Count 1 Year‘ +
                     ‘Bloodstained Mire‘ +
                     ‘Scalding Tarn‘ + ‘Verdant Catacombs‘ +
                     ‘Misty Rainforest‘ +
                     ‘Prismatic Vista‘,
                   data = my_data)
# Marsh Flats

reduced_model_7 <- lm(‘Deck Price‘ ~ ‘Archetype‘ + ‘Deck Count 1 Year‘ +
                     ‘Bloodstained Mire‘ +
                     ‘Scalding Tarn‘ + ‘Verdant Catacombs‘ +
                     ‘Misty Rainforest‘,
                   data = my_data)
# Prismatic Vista

reduced_model_8 <- lm(‘Deck Price‘ ~ ‘Archetype‘ +
                     ‘Bloodstained Mire‘ +
                     ‘Scalding Tarn‘ + ‘Verdant Catacombs‘ +
                     ‘Misty Rainforest‘,
                   data = my_data)
# Deck Count 1 Year

reduced_model_9 <- lm(‘Deck Price‘ ~ ‘Archetype‘ +
                     ‘Bloodstained Mire‘ + ‘Verdant Catacombs‘ +
```

```
                      `Misty Rainforest`,
                    data = my_data)
# Scalding Tarn
############### Reducing Model




kable(summary(reduced_model_8)$coefficients, caption = "Reduced Model vs Deck Price Intercept a




plot(fitted(full_model), residuals(full_model),
     sub = "Full Model", xlab = "Fitted Values", ylab = "Residuals")
abline(h=0)

plot(fitted(reduced_model_8), residuals(reduced_model_8),
     sub = "Removing year", xlab = "Fitted Values", ylab = "Residuals")
abline(h=0)
# There appear to be no outliers!
```