

Statistical Trends of the Secondhand GPU Market (Q1-2023)

Computational Statistics & Probability: Final Project

Mark Gotesman

Spring, 2023

Abstract

The present paper analyzed data from eBay to reach conclusions about the state of the used GPU market in Q1 2023. The data ($N = \sim 10,000$) was collected and processed by the author for the sake of the paper. Research questions sought to ascertain basic features of the marketplace and what a buyer might consider when shopping for a used GPU. These included the difference in listed and sold postings to the site, changes in variance due to the age of the card, and market preference for a certain board partner. This paper begins an exploration of what could be even deeper insight into the marketplace arising from more sophisticated tools.

1. Introduction

1.1 GPUs & the Market of eBay

The graphics processing unit (GPU) is a well-known PC component that provides high-throughput parallel compute power due to its unique architecture relative to a CPU. The discrete (as opposed to integrated) GPU market is (at least as of the past decade) made up of a few primary segments (scientific compute, cryptocurrency mining, and PC gaming), and sees sales of ~ 40 million annually. The dominant vendor in the space is Nvidia, though AMD (and recently Intel) also have moderate market share¹.

GPUs are typically sold by “generations” (or “series”), within which there are different tiers; for example, in the Nvidia 10-series cards, one can find the 1060, 1070, 1080 etc. These tiers are (often) consistent throughout the different series so that consumers can have a reasonable expectation of what performance to expect in a purchase of a tier for that generation.

This large consumer base of a persistent resource naturally gives rise to a rich secondhand market. Many of these transactions take place on the online marketplace of eBay, which will be the focus of this paper. eBay’s rich database is publicly searchable, both for items currently listed and those sold (within ~ 3 months); data will be culled from both of these types.

1.2 Research Questions

When approaching such a broad subject of investigation many interesting economic and statistical questions can be raised. This paper will seek to address the following:

1. **Normality of Underlying Populations.** Are the sold and listed secondhand GPU markets normally distributed? If not, does there exist a reasonable explanation?

¹<https://www.tomshardware.com/news/nvidia-maintains-lead-as-sales-of-graphics-cards-hit-all-time-low-in-2022-jpr>

2. **Average of Listed vs. Sold.** Common wisdom says that sellers on eBay often price their items higher than the market selling price. Is this claim statistically valid for the eBay GPU market?
3. **Effects of Age on Distribution Variance.** Of course, the *mean* price of older generation models is expected to be lower than newer models - but what about variance? Do older series of cards tend to have a smaller variance, indicating a general trend of market *stability* due to their age?
4. **Significance of Board Partner.** Besides for the manufacturer of the GPU chip itself, licensing deals are made with *board partners* to design the printed circuit board (PCB) and cooling system for the GPU. Does the market show any significant preference for a certain board partner over another?

2. Data

2.1 Features of Interest

The data set will consist of 7 features/variables:

- **Posting.** Binary - either “listed” or “sold”.
- **Model.** The full canonical name of the GPU model (e.g., 1060). From this, its **Series** and **Tier** can typically be extracted, which was done in the data processing step and labeled as a separate feature for easier manipulations.
- **Board Partner.** As explained above, these are companies that partner with the manufacturer to design a PCB and present the card to the end consumer. Nvidia also produces their own in-house PCB design which are known as “founders edition” cards (and are here denoted as FOUNDER). This field was extracted from the scraped eBay listing title, so there are some NA values.
- **Total_Price.** The sum of the listing price + shipping fee (if charged).
- **Date_Sold.** Applicable if the item was sold; otherwise NA.

2.2 Data Collection Process

All data used in this project was collected by the author from [eBay’s US site](#).

As mentioned above, the GPUs of interest in this study are Nvidia cards, the largest vendor in the space. A *range* of models were chosen to give the most basis for answering the research questions at hand. A summary of the eBay search parameters is as follows:

- (3) Generations: 10-series, 20-series, and 30-series.
- (3) Tiers: 60, 70, and 80 tiers.
- (2) Posting: listed or sold.

This results in $3 \times 3 \times 2 = 18$ total permutations to search for. After a bit of trial and error, a reasonable search query was developed that strives to eliminate garbage data using the “-” modifier in the eBay search URL (Figure 1).

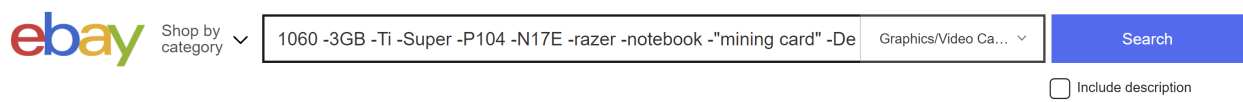


Figure 1: eBay Search With Excluded Keywords

A custom pipeline was created using the data mining tool [Octoparse](#), efficiently mining ~**10,000** entries of data across the 18 different search permutations (Figure 2).

This data was then imported to Google Sheets via [Google Cloud API](#) and features of interest were parsed (Figure 3).

From there, the processed data can be imported directly into R using the [openxlsx](#) library:

Home | Dashboard

Unified[Series][Tier][Posti...

1060 -3GB -Ti -Super -P104 -N17E -razer -notebook -"mining card" -Dell -HP -replace -miner -OEM -lapto...

Total 10 | Browse

Save Run

Sign in or register

Daily Deals

Brand Outlet

Help & Contact

Sell

Watchlist

My eBay

ebay

Shop by category

1060 -3GB -Ti -Super -P104 -N17E -razer -notebook -"mining card" -De

Graphics/Video Ca...

Search

Advanced

Include description

Category

All

< Computers/Tablets & Networking

< Computer Components & Parts

Graphics/Video Cards

Motherboards

Laptop Replacement Parts

Other Components & Parts

More

Business & Industrial

All Listings

Accepts Offers

Auction

Buy It Now

Condition

Shipping

Local

Best Match

188 results for 1060 -3GB -Ti -Su...

Save this search

Shipping to: 10033

Open box

Seller refurbished

Used

US Only

Clear All

NEW LISTING

EVGA GeForce GTX 1060 6GB Graphics Card

Pre-Owned · EVGA · 6 GB

\$83.99

Buy It Now

trasm5888 (7) 100%

Loop URLs

Go to Webpage

Pagination

Scroll Page

Loop Item

Extract

General

Options

Retry

Action

Go to Webpage

URL

https://www.ebay.com/sch/i.html?_sacat=27386&LH_BIN=1&LH_PrefLo

Load URLs in the loop

JSON

Timeout 120s before the next step

Apply

Data Preview

Total [100] line(s) of data. Preview the first [60] lin...

Preview more data

Data Fields	No.	URL	Keywords	Title	Price	Shipping_fee	Posting	Date_Sold	Actions
Page1	1	https://www.ebay.co...	1060 -3GB -Ti -Super ...	NEW LISTINGEVGA G...	83.99	18.95	Listed		
Extract	2	https://www.ebay.co...	1060 -3GB -Ti -Super ...	PNY GeForce GTX 10...	109.00	0	Listed		
	3	https://www.ebay.co...	1060 -3GB -Ti -Super ...	GEFORCE GTX 1060 6...	79.99	19.99	Listed		
	4	https://www.ebay.co...	1060 -3GB -Ti -Super ...	GIGABYTE GeForce G...	120.00	18.50	Listed		
	5	https://www.ebay.co...	1060 -3GB -Ti -Super ...	PNY GeForce GTX 10...	100.00		Listed		
	6	https://www.ebay.co...	1060 -3GB -Ti -Super ...	NEW LISTINGMSI Ga...	140.00	15.79	Listed		
	7	https://www.ebay.co...	1060 -3GB -Ti -Super ...	msi geforce gtx 1060...	200.00	26.45	Listed		
	8	https://www.ebay.co...	1060 -3GB -Ti -Super ...	NVIDIA GTX1060DE-Z...	125.00	0	Listed		
	9	https://www.ebay.co...	1060 -3GB -Ti -Super ...	Surface Book 2 15" 18...	349.94		Listed		
	10	https://www.ebay.co...	1060 -3GB -Ti -Super ...	EVGA NVIDIA Geforc...	99.99	18.50	Listed		
	11	https://www.ebay.co...	1060 -3GB -Ti -Super ...	NEW LISTINGEVGA N...	112.98	0	Listed		
	12	https://www.ebay.co...	1060 -3GB -Ti -Super ...	ZOTAC NVIDIA Gefor...	144.99	0	Listed		
	13	https://www.ebay.co...	1060 -3GB -Ti -Super ...	ASUS NVIDIA Geforc...	99.99	0	Listed		

Figure 2: Octoparse Task Creation Window

3	URL	Keywords	Title	Price	Shipping fee	Posting	Date_Sold		Posting	Series	Tier	Model	Board_Partner	Total_Price	Date_Sold
29	https://www.ebay.co.uk/itm/3080-G0B-Ti-Super-P-EVGA-GeForce-RTX-3080-FW3-ULTRA-GAMING-10-GB-Used	3080-G0B-Ti-Super-P-EVGA GeForce RTX 3080 FW3 ULTRA GAMING - 10 GB - Used	1,300.00	10.2	Listed		OK	Listed	30	80	3080	EVGA		1,310.20	NA
30	https://www.ebay.co.uk/itm/3080-G0B-Ti-Super-P-Nvidia RTX 3080 gty 3, 2 Founders Edition, 1 EVGA XC3 Ultra With Original Box	3080-G0B-Ti-Super-P Nvidia RTX 3080 gty 3, 2 Founders Edition, 1 EVGA XC3 Ultra With Original Box	1,250.00	35	Sold	2/26/2023	OK	Listed	30	80	3080	FOUNDER		1,285.00	2/26/2023
31	https://www.ebay.co.uk/itm/3080-G0B-Ti-Super-P-Gigabyte AORUS Xtreme GeForce RTX 2080 8GB	3080-G0B-Ti-Super-P Gigabyte AORUS Xtreme GeForce RTX 2080 8GB	1,192.00	49.35	Listed		OK	Listed	20	80	2080	GIGABYTE		1,241.35	NA
32	https://www.ebay.co.uk/itm/3060-G0B-Ti-Super-P msi ge76 raid rx 3060	3060-G0B-Ti-Super-P msi ge76 raid rx 3060	1,200.00	22.45	Listed		OK	Listed	30	60	3060	MSI		1,222.45	NA
33	https://www.ebay.co.uk/itm/2070-G0B-Ti-Super-P MSI GeForce RTX 2070 8GB GDDR6 Graphics Card (RTX2070GAMINGZG8)	2070-G0B-Ti-Super-P MSI GeForce RTX 2070 8GB GDDR6 Graphics Card (RTX2070GAMINGZG8)	1,200.00	12.55	Listed		OK	Listed	20	70	2070	MSI		1,212.55	NA
34	https://www.ebay.co.uk/itm/2070-G0B-Ti-Super-P Gigabyte GeForce RTX 2070 Gaming OC Graphics Card	2070-G0B-Ti-Super-P Gigabyte GeForce RTX 2070 Gaming OC Graphics Card	1,200.00	11.75	Listed		OK	Listed	20	70	2070	GIGABYTE		1,211.75	NA
35	https://www.ebay.co.uk/itm/3080-G0B-Ti-Super-P White ASUS ROG Strix GeForce RTX 3080 10 GB OC NON-LHR-GDDR6X,Still in a 1,999.00	3080-G0B-Ti-Super-P White ASUS ROG Strix GeForce RTX 3080 10 GB OC NON-LHR-GDDR6X,Still in a 1,999.00	0	0	Sold	2/12/2023	OK	Sold	30	80	3080	ASUS		1,999.00	2/12/2023
36	https://www.ebay.co.uk/itm/3080-G0B-Ti-Super-P White ASUS ROG Strix GeForce RTX 3080 10 GB OC NON-LHR-GDDR6X,Still in a 1,170.00	3080-G0B-Ti-Super-P White ASUS ROG Strix GeForce RTX 3080 10 GB OC NON-LHR-GDDR6X,Still in a 1,170.00	0	0	Sold	2/26/2023	OK	Sold	30	80	3080	ASUS		1,170.00	2/26/2023
37	https://www.ebay.co.uk/itm/3060-G0B-Ti-Super-P msi ge76 raid rx 3060	3060-G0B-Ti-Super-P msi ge76 raid rx 3060	1,150.00	19	Listed		OK	Listed	30	60	3060	MSI		1,169.00	NA
38	https://www.ebay.co.uk/itm/2080-G0B-Ti-Super-P NVIDIA GeForce RTX 2080 OC with 8GB Small form factor fits into any small ca	2080-G0B-Ti-Super-P NVIDIA GeForce RTX 2080 OC with 8GB Small form factor fits into any small ca	1,149.00	15	Listed		OK	Listed	20	80	2080	NA		1,164.00	NA
39	https://www.ebay.co.uk/itm/1080-G0B-Ti-Super-P ASUS ROG Strix NVIDIA GeForce GTX 1080 8GB Graphics Card	1080-G0B-Ti-Super-P ASUS ROG Strix NVIDIA GeForce GTX 1080 8GB Graphics Card	1,149.99	13.45	Listed		OK	Listed	10	80	1080	ASUS		1,163.44	NA
40	https://www.ebay.co.uk/itm/3070-G0B-Ti-Super-P EVGA 3090 XC3 ultra + RTX 3070 Founders Edition	3070-G0B-Ti-Super-P EVGA 3090 XC3 ultra + RTX 3070 Founders Edition	1,075.00	55.17	Sold	3/18/2023	OK	Sold	30	70	3070	EVGA		1,130.17	3/18/2023
41	https://www.ebay.co.uk/itm/3080-G0B-Ti-Super-P NVIDIA GeForce RTX 3080 Founders 10GB GDDR6X Graphics Card	3080-G0B-Ti-Super-P NVIDIA GeForce RTX 3080 Founders 10GB GDDR6X Graphics Card	1,125.00	0	Listed		OK	Sold	30	80	3080	FOUNDER		1,125.00	NA
42	https://www.ebay.co.uk/itm/3080-G0B-Ti-Super-P NVIDIA GeForce RTX 3080 Founders 10GB GDDR6X Graphics Card	3080-G0B-Ti-Super-P NVIDIA GeForce RTX 3080 Founders 10GB GDDR6X Graphics Card	1,125.00	0	Listed		OK	Listed	30	80	3080	FOUNDER		1,125.00	NA
43	https://www.ebay.co.uk/itm/3080-G0B-Ti-Super-P ASUS ROG Strix RTX 3080 White OC Edition 10GB Barley Used, Plastic Still in C	3080-G0B-Ti-Super-P ASUS ROG Strix RTX 3080 White OC Edition 10GB Barley Used, Plastic Still in C	1,099.99	19.99	Sold	2/18/2023	OK	Sold	30	80	3080	ASUS		1,119.98	2/18/2023
44	https://www.ebay.co.uk/itm/3080-G0B-Ti-Super-P GIGABYTE AORUS GeForce RTX 3080 Xtreme AIO Waterblock GPU	3080-G0B-Ti-Super-P GIGABYTE AORUS GeForce RTX 3080 Xtreme AIO Waterblock GPU	1,100.00	12.55	Listed		OK	Listed	30	80	3080	GIGABYTE		1,112.55	NA
45	https://www.ebay.co.uk/itm/3080-G0B-Ti-Super-P GIGABYTE AORUS GeForce RTX 3080 Xtreme AIO Waterblock GPU	3080-G0B-Ti-Super-P GIGABYTE AORUS GeForce RTX 3080 Xtreme AIO Waterblock GPU	1,100.00	12.55	Listed		OK	Listed	30	80	3080	GIGABYTE		1,112.55	NA
46	https://www.ebay.co.uk/itm/3080-G0B-Ti-Super-P MSI NVIDIA GeForce RTX 3080 10GB GDDR6X Graphics Card	3080-G0B-Ti-Super-P MSI NVIDIA GeForce RTX 3080 10GB GDDR6X Graphics Card	1,099.99	5	Listed		OK	Listed	30	80	3080	MSI		1,104.99	NA
47	https://www.ebay.co.uk/itm/3080-G0B-Ti-Super-P Zotac GAMING GeForce RTX 3080 Trinity OC LHR 10GB GDDR6X Graphics Ca	3080-G0B-Ti-Super-P Zotac GAMING GeForce RTX 3080 Trinity OC LHR 10GB GDDR6X Graphics Ca	1,104.99		Listed		OK	Listed	30	80	3080	ZOTAC		1,104.99	NA
48	https://www.ebay.co.uk/itm/3080-G0B-Ti-Super-P Zotac GAMING GeForce RTX 3080 Trinity OC LHR 10GB GDDR6X Graphics Ca	3080-G0B-Ti-Super-P Zotac GAMING GeForce RTX 3080 Trinity OC LHR 10GB GDDR6X Graphics Ca	1,104.99		Listed		OK	Listed	30	80	3080	ZOTAC		1,104.99	NA
49	https://www.ebay.co.uk/itm/3080-G0B-Ti-Super-P Asus RTX 3080 OC Edition 12GB TUF Gaming GeForce GDDR6X TUF-TX3080-4	3080-G0B-Ti-Super-P Asus RTX 3080 OC Edition 12GB TUF Gaming GeForce GDDR6X TUF-TX3080-4	1,099.99	0	Sold	2/8/2023	OK	Sold	30	80	3080	ASUS		1,099.99	2/8/2023
50	https://www.ebay.co.uk/itm/3080-G0B-Ti-Super-P GIGABYTE RTX 3080 VISION OC 10GB GDDR6X LHR Graphics Card GV-N3080-4	3080-G0B-Ti-Super-P GIGABYTE RTX 3080 VISION OC 10GB GDDR6X LHR Graphics Card GV-N3080-4	1,099.95	0	Sold	3/8/2023	OK	Sold	30	80	3080	GIGABYTE		1,099.95	3/8/2023
51	https://www.ebay.co.uk/itm/3080-G0B-Ti-Super-P ASUS NVIDIA GeForce RTX 3080 12GB GDDR6X Graphics Card ...	3080-G0B-Ti-Super-P ASUS NVIDIA GeForce RTX 3080 12GB GDDR6X Graphics Card ...	1,050.00	20.35	Sold	3/31/2023	OK	Sold	30	80	3080	ASUS		1,070.35	3/31/2023
52	https://www.ebay.co.uk/itm/3080-G0B-Ti-Super-P ASUS NVIDIA GeForce RTX 3080 12GB GDDR6X Graphics Card ...	3080-G0B-Ti-Super-P ASUS NVIDIA GeForce RTX 3080 12GB GDDR6X Graphics Card ...	1,050.00	20.35	Sold	3/31/2023	OK	Sold	30	80	3080	ASUS		1,070.35	3/31/2023
53	https://www.ebay.co.uk/itm/3060-G0B-Ti-Super-P EVGA GeForce RTX 3060 X GAMING 12GB GDDR6 Graphics Card (RTX-P5-36-1	3060-G0B-Ti-Super-P EVGA GeForce RTX 3060 X GAMING 12GB GDDR6 Graphics Card (RTX-P5-36-1	1,050.00	16.14	Listed		OK	Listed	30	60	3060	EVGA		1,066.14	NA
54	https://www.ebay.co.uk/itm/3080-G0B-Ti-Super-P ASUS ROG Strix GeForce RTX 3080 12GB GDDR6X OC (EVA Edition Graphics C	3080-G0B-Ti-Super-P ASUS ROG Strix GeForce RTX 3080 12GB GDDR6X OC (EVA Edition Graphics C	1,049.99	0	Sold	2/8/2023	OK	Listed	30	80	3080	ASUS		1,049.99	2/8/2023
55	https://www.ebay.co.uk/itm/2070-G0B-Ti-Super-P EVGA GeForce RTX 2070 8GB GDDR6 Graphics Card 08PA43173KR	2070-G0B-Ti-Super-P EVGA GeForce RTX 2070 8GB GDDR6 Graphics Card 08PA43173KR	1,000.00		Listed		OK	Listed	20	70	2070	EVGA		1,032.25	NA
56	https://www.ebay.co.uk/itm/3060-G0B-Ti-Super-P EVGA GeForce RTX 3060 X GAMING 12GB GDDR6 Graphics Card (RTX-P5-36-1	3060-G0B-Ti-Super-P EVGA GeForce RTX 3060 X GAMING 12GB GDDR6 Graphics Card (RTX-P5-36-1	1,000.00	25.45	Listed		OK	Listed	30	60	3060	EVGA		1,025.45	NA
57	https://www.ebay.co.uk/itm/3060-G0B-Ti-Super-P EVGA GeForce RTX 3060 X GAMING 12GB GDDR6 Graphics Card (RTX-P5-36-1	3060-G0B-Ti-Super-P EVGA GeForce RTX 3060 X GAMING 12GB GDDR6 Graphics Card (RTX-P5-36-1	1,000.00	25.45	Listed		OK	Listed	30	60	3060	EVGA		1,025.45	NA
+ ≡ Import Raw Processed															

Figure 3: Data Processing in Google Sheets

```
library(openxlsx)
dat = read.xlsx("https://docs.google.com/spreadsheets/d/1c2um4vqnE4ouyetGoRLnoV-Foz1DsGa4lCseMCJHYwk/exp
sheet = 3)
```

Data was formatted properly using R `as.()` functions:

```
dat$Posting = as.factor(dat$Posting)
dat$Series = as.factor(dat$Series)
dat$Tier = as.factor(dat$Tier)
dat$Model = as.factor(dat$Model)
dat$Board_Partner = as.factor(dat$Board_Partner)
dat$Date_Sold = as.Date(dat$Date_Sold, origin = "1899-12-30")
dat$Total_Price = as.numeric(dat$Total_Price)
```

2.3 Data Visualization & Summary Statistics

Quick looks at the data frame are as follows:

```
str(dat)
```

```
## 'data.frame': 9777 obs. of 7 variables:
## $ Posting : Factor w/ 2 levels "Listed","Sold": 1 1 1 1 1 1 1 1 1 1 ...
## $ Series : Factor w/ 3 levels "10","20","30": 3 3 3 3 2 3 3 3 3 3 ...
## $ Tier : Factor w/ 3 levels "60","70","80": 2 2 2 2 2 3 3 3 3 1 ...
## $ Model : Factor w/ 9 levels "1060","1070",...: 8 8 8 8 5 9 9 9 9 7 ...
## $ Board_Partner: Factor w/ 7 levels "ASUS","EVGA",...: 1 1 5 5 4 7 7 2 2 NA ...
## $ Total_Price : num 1999 1999 1921 1921 1800 ...
## $ Date_Sold : Date, format: NA NA ...
```

```
head(dat)
```

```
## Posting Series Tier Model Board_Partner Total_Price Date_Sold
## 1 Listed 30 70 3070 ASUS 1999.00 <NA>
## 2 Listed 30 70 3070 ASUS 1999.00 <NA>
## 3 Listed 30 70 3070 MSI 1921.45 <NA>
## 4 Listed 30 70 3070 MSI 1921.45 <NA>
## 5 Listed 20 70 2070 GIGABYTE 1800.00 <NA>
## 6 Listed 30 80 3080 ZOTAC 1750.00 <NA>
```

Summary statistics can also help give a basic sense of the distribution:

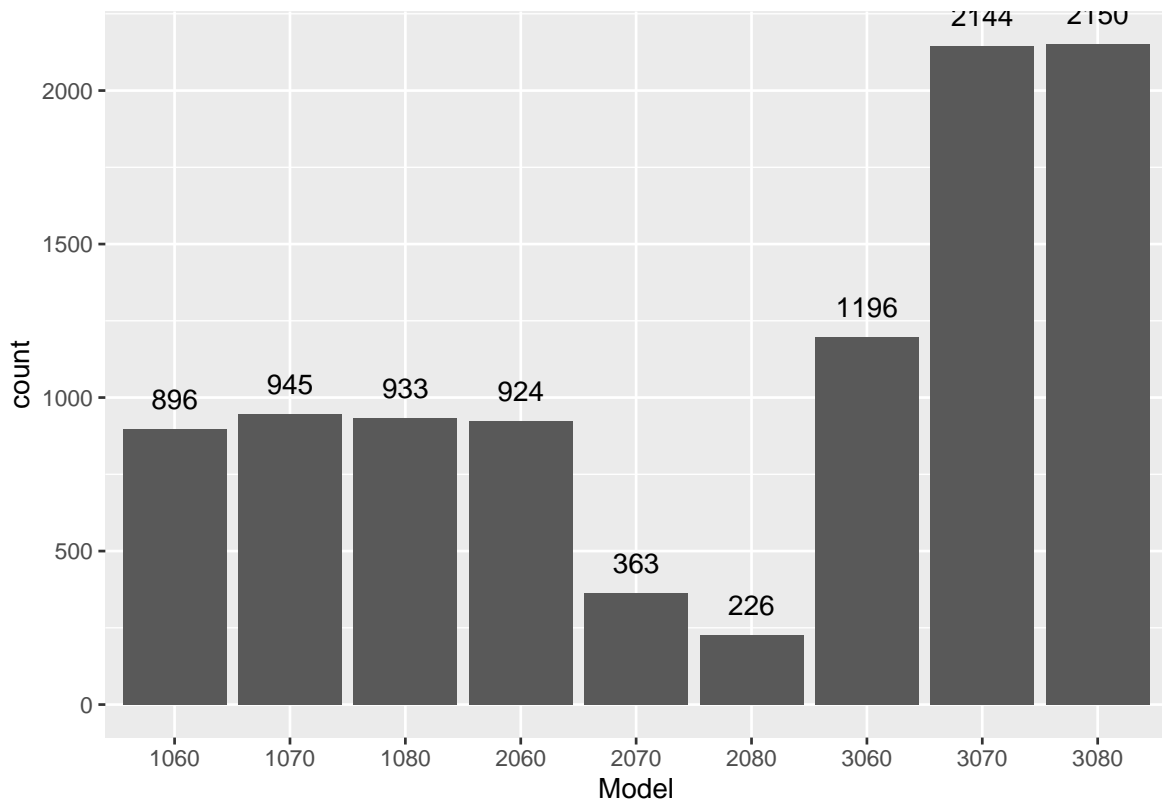
```
summary(dat)
```

```
## Posting Series Tier Model Board_Partner
## Listed:2221 10:2774 60:3016 3080 :2150 EVGA :2508
## Sold :7556 20:1513 70:3452 3070 :2144 MSI :1645
## 30:5490 80:3309 3060 :1196 ASUS :1602
## 1070 : 945 GIGABYTE:1386
## 1080 : 933 ZOTAC : 852
## 2060 : 924 (Other) :1149
## (Other):1485 NA's : 635
## Total_Price Date_Sold
## Min. : 3.0 Min. :2022-04-15
## 1st Qu.: 185.0 1st Qu.:2023-02-27
## Median : 332.6 Median :2023-03-24
## Mean : 351.2 Mean :2023-03-18
## 3rd Qu.: 480.0 3rd Qu.:2023-04-13
```

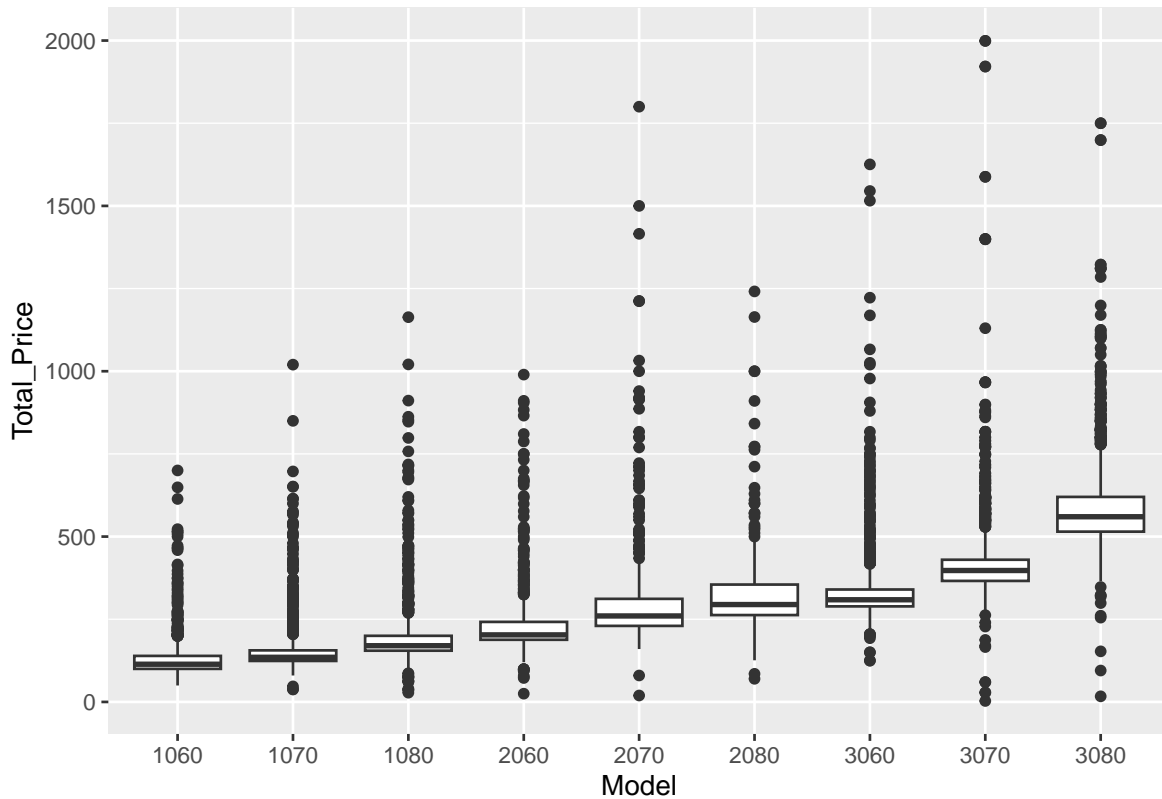
```
## Max.    :1999.0    Max.    :2023-04-28
##                NA's    :2221
```

ggplot2 can help visualize the observations. The most important features to note are the **model** of the card and the **total price**. The first figure to follow shows the count of observations for each model, and the second shows a box plot of the total price :

```
library(ggplot2)
plot = ggplot(dat, aes(Model))
plot + geom_bar() + geom_text(stat = "count", aes(label = after_stat(count)), vjust = -1)
```



```
plot = ggplot(dat, aes(Model, Total_Price))
plot + geom_boxplot()
```



3. Statistical Analysis

3.1 Normality of Underlying Populations (Q-Q Plots)

The best way to check the normality of the data is with a Q-Q plot. If the population is indeed normally distributed, the distance between quantiles in the sample should roughly match that of a normal distribution - plotted against each other a straight line should be produced. This will be performed both for the listed and the sold observations. Additionally, a histogram of the data overlaid with a normal curve for its mean and standard deviation will be plotted to give a visual sense of conformity to the normal.

```
normal_plot = function(distr, name, posting) {
  qqnorm(distr, main = paste(name, "-", posting))
  qqline(distr, main = name)

  hist(distr, prob = TRUE, breaks = 30, xlim = c(0, max(distr)), main = paste(name,
    "-", posting))
  x = seq(0, max(distr))
  f = dnorm(x, mean(distr), sd(distr))
  lines(x, f, col = "red", lwd = 2)
}

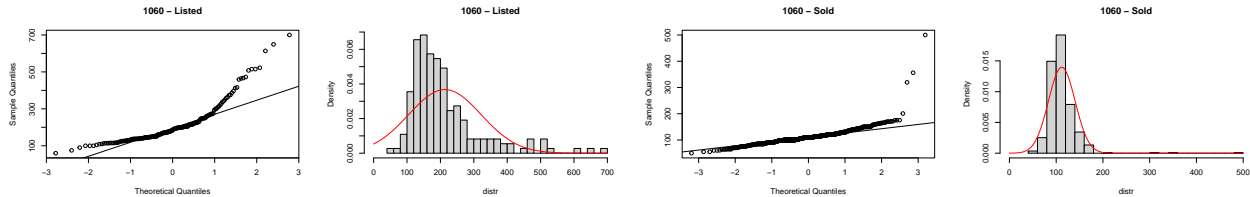
normality_check = function(model) {
  listed = dat[(dat$Model == model & dat$Posting == "Listed"), ]$Total_Price
  sold = dat[(dat$Model == model & dat$Posting == "Sold"), ]$Total_Price

  normal_plot(listed, model, "Listed")
  normal_plot(sold, model, "Sold")
}
```

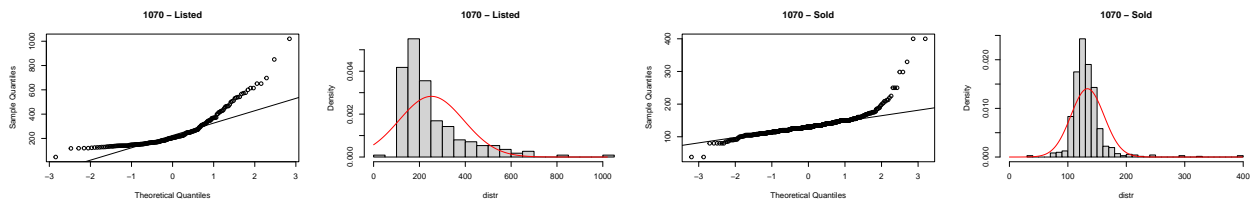
```
}
```

The 10-series will be used as an example.

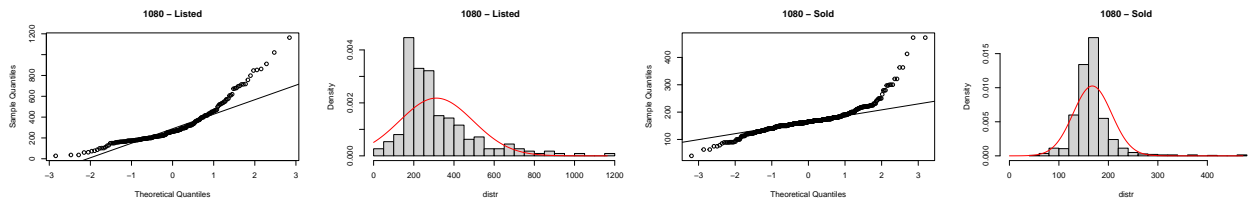
```
normality_check(1060)
```



```
normality_check(1070)
```



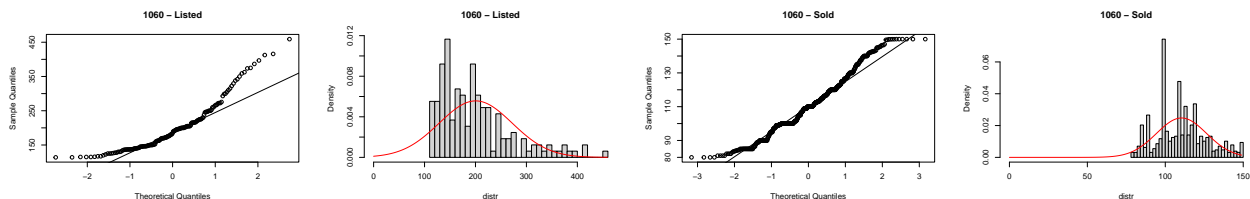
```
normality_check(1080)
```



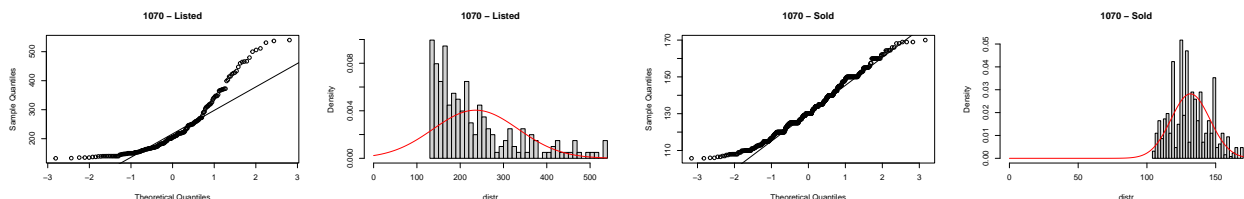
It appears from the plots that the sold data conforms much better to the assumption of normality. The clear trend that emerges when considering the two plots is the difference in variance – the sold listings are considerably more tightly clustered than the listed ones are, with a tail trailing off towards the higher values of price. This is quite logical: sellers may try to get an exceedingly ambitious price for their items, or post them too high and forget to ever change them, while items posted too low will be purchased off the market almost immediately. This results in a skew of listed items towards higher prices.

Though the sold items do not conform *perfectly* to the normal distribution, they are quite close. The issue seems to be a few outlier values that force the normal curve to be wider to accommodate those unexpected events. This indicates the existence of outliers and that truncating the data range on either end may result in a more accurate analysis (newly defined function not shown for brevity).

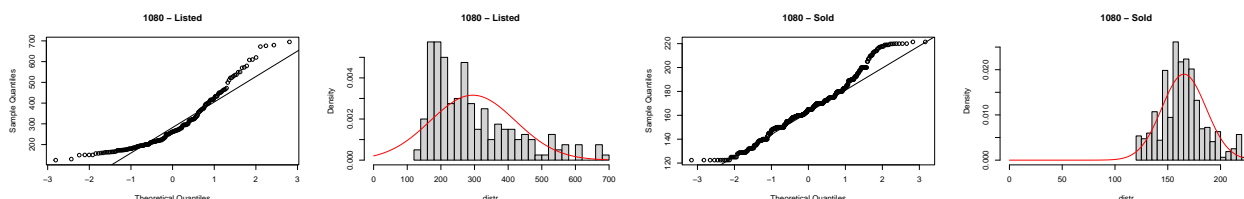
```
normality_check_outliers(1060, percentage = 0.05)
```



```
normality_check_outliers(1070, percentage = 0.05)
```



```
normality_check_outliers(1080, percentage = 0.05)
```



After cutting off the top and bottom 5%, the listed items do not seem to conform any better to the normal distribution on the Q-Q plot, while the sold items in many cases (see the 1070 especially) conform very well, which supports the hypothesis that the sold distributions are fundamentally normally distributed with a bit of noise due to outliers, while the listed populations are not.

3.2 Average of Listed vs. Sold (Z-Test)

Having speculated that the spread of the listed population seems to be much larger due to overly ambitious sellers, it would also be of note to determine if their means are statistically different from each other.

To compare means, the two-sample Z-Test is best used. Though the underlying populations may not be normal, there are enough samples to assume normality of the sample mean and standard deviation.

Our null hypothesis is that the means are equal, while the alternate is that they differ:

$$H_0 : \mu_1 - \mu_2 = 0 \quad H_A : \mu_1 - \mu_2 \neq 0$$

The test statistic is as follows:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \sim \mathcal{N}(0, 1)$$

The following will calculate this test statistic and compute the p-value for the test:

```
z_test = function(d1, d2) {
  x_bar1 = mean(d1)
  x_bar2 = mean(d2)

  s1 = sd(d1)
  s2 = sd(d2)

  n1 = length(d1)
  n2 = length(d2)

  Z = (x_bar1 - x_bar2)/(sqrt(s1^2/n1 + s2^2/n2))

  p = pnorm(-abs(Z), lower.tail = TRUE) + pnorm(abs(Z), lower.tail = FALSE)
  return(p)
```



```

}

listed = dat[(dat$Model == 1070 & dat$Posting == "Listed"), ]$Total_Price
sold = dat[(dat$Model == 1070 & dat$Posting == "Sold"), ]$Total_Price

print(z_test(listed, sold))

```

```
## [1] 4.41405e-36
```

The very low p-value suggests that we can reject H_0 with almost 100% certainty.

This process can be repeated after truncating the upper and lower 10% to exclude potential outliers:

```

listed = tr(listed, 0.1)
sold = tr(sold, 0.1)

print(z_test(listed, sold))

```

```
## [1] 9.627284e-61
```

Where the p-value is even more extreme. From this analysis, it becomes clear that prospective sellers would be wise to take an average of the sold items and not the listed ones to give their items the best chance to sell in a reasonable time window.

3.3 Effects of Age on Distribution Variance (F-Test)

Next, we will turn to consider the different generations of GPUs compared to each other. As noted in the introduction, GPU generations are released yearly or bi-yearly. This results in a natural lifecycle for the product, from bleeding edge to a value-budget option after years on the secondhand market. So of course, the mean of an older-generation tier (e.g. 1060) will be lower than the mean of a new-generation card of the same tier (e.g. 3060). But what about variance? Can it be shown that the dynamic of a newer release results in a market that is less settled?

As shown above, the **sold** listings are likely a better indicator of the actual transactions occurring in the market, so they will be used for this section. Further, after collecting the data, it became clear that the 20-series cards are distinctly underrepresented. The author's speculation on this is that in the 20-series, Nvidia released not only the 2060, 2070, and 2080, but also what they called "super" versions of the cards which were in-between each of the tiers for that generation, which splintered the consumer base and resulted in less volume of base cards on the market. Regardless of the reason, the 20-series will be overlooked and the comparison will be between the 10-series and 30-series, between all three tiers.

The relevant test to use when comparing the variance of two populations is the F-test. The test statistic and rejection ranges are as follows:

$$F = \frac{s_1^2}{s_2^2}$$

Since the test is two-sided, the rejection range has to be considered based on the choice for the "1" and the "2" distribution. Meaning, for a two-sided F test, the p-value should be low (indicating a significant difference in the ratio of the variances) either when F is very large or when F is very small. In a given ratio of variances from two distributions, these two cases are just inverses of each other - either the larger variance sits on top with the smaller on the bottom and F is large, or the inverse and F is small. So the rejection range is as follows:

$$\begin{aligned}
 F &> F_{n_2-1, \alpha/2}^{n_1-1} & \text{if} & & F > 1 \\
 F &< \frac{1}{F_{n_2-1, \alpha/2}^{n_1-1}} & \text{if} & & F < 1
 \end{aligned}$$

The following will calculate this test statistic and compute the p-value for the test. Due to the complication of the two-sided test outlined above, the p-value will be twice the area to the right of the critical value if $F > 1$, and to the left of the critical value if $F < 1$:

```
f_test = function(d1, d2) {
  s1 = sd(d1)
  s2 = sd(d2)

  n1 = length(d1)
  n2 = length(d2)

  F = (s1^2)/(s2^2)

  if (F > 1) {
    p = 2 * pf(F, n1 - 1, n2 - 1, lower.tail = FALSE)
  } else {
    p = 2 * pf(1/F, n1 - 1, n2 - 1, lower.tail = FALSE)
  }
  return(p)
}

print(f_test(dat[(dat$Model == 1060 & dat$Posting == "Sold"), ]$Total_Price, dat[(dat$Model ==
3060 & dat$Posting == "Sold"), ]$Total_Price))

## [1] 5.102431e-72

print(f_test(dat[(dat$Model == 1070 & dat$Posting == "Sold"), ]$Total_Price, dat[(dat$Model ==
3070 & dat$Posting == "Sold"), ]$Total_Price))

## [1] 2.829766e-180

print(f_test(dat[(dat$Model == 1080 & dat$Posting == "Sold"), ]$Total_Price, dat[(dat$Model ==
3080 & dat$Posting == "Sold"), ]$Total_Price))

## [1] 7.503043e-187
```

Since the p-values are so low, we can confidently reject H_0 and conclude that the variances indeed differ. Looking at the values of the sample standard deviations, it is clear that the 30-series GPUs have a much larger variance than the 10-series, which supports the theory that the newer market is significantly less settled due to the age of the cards and hence has a larger variance.

3.4 Significance of Board Partner (CIs)

The last question of interest surrounds the as-of-yet unexplored feature - the board partner of the GPUs. Can it be said that the expected price for a certain company is statistically higher than another?

Since there are 6 board partners, a reasonable way to visualize and interpret the comparison is to partition a GPU around all the board partners and generate a confidence interval for the mean of that subset.

We will select the 3070 as an example GPU due to its large number of observations and a relatively balanced count of observations for each board partner (see later).

First, we select the 3070s from the data set:

```
dat_model = dat[(dat$Model == 3070 & dat$Posting == "Sold"), ]
```

We will define a function that will return the confidence interval the mean of a sample:

```
mean_CI = function(distr, alpha) {
  x_bar = mean(distr)
  z = abs(qnorm(alpha))
  s = sd(distr)
  n = length(distr)

  lower = x_bar - z * s/sqrt(n)
  upper = x_bar + z * s/sqrt(n)
  return(c(lower, upper))
}
```

Then we can partition out the different board partners and create a new data frame (`intervals`) which we will use to plot:

```
partners = levels(dat_model$Board_Partner)

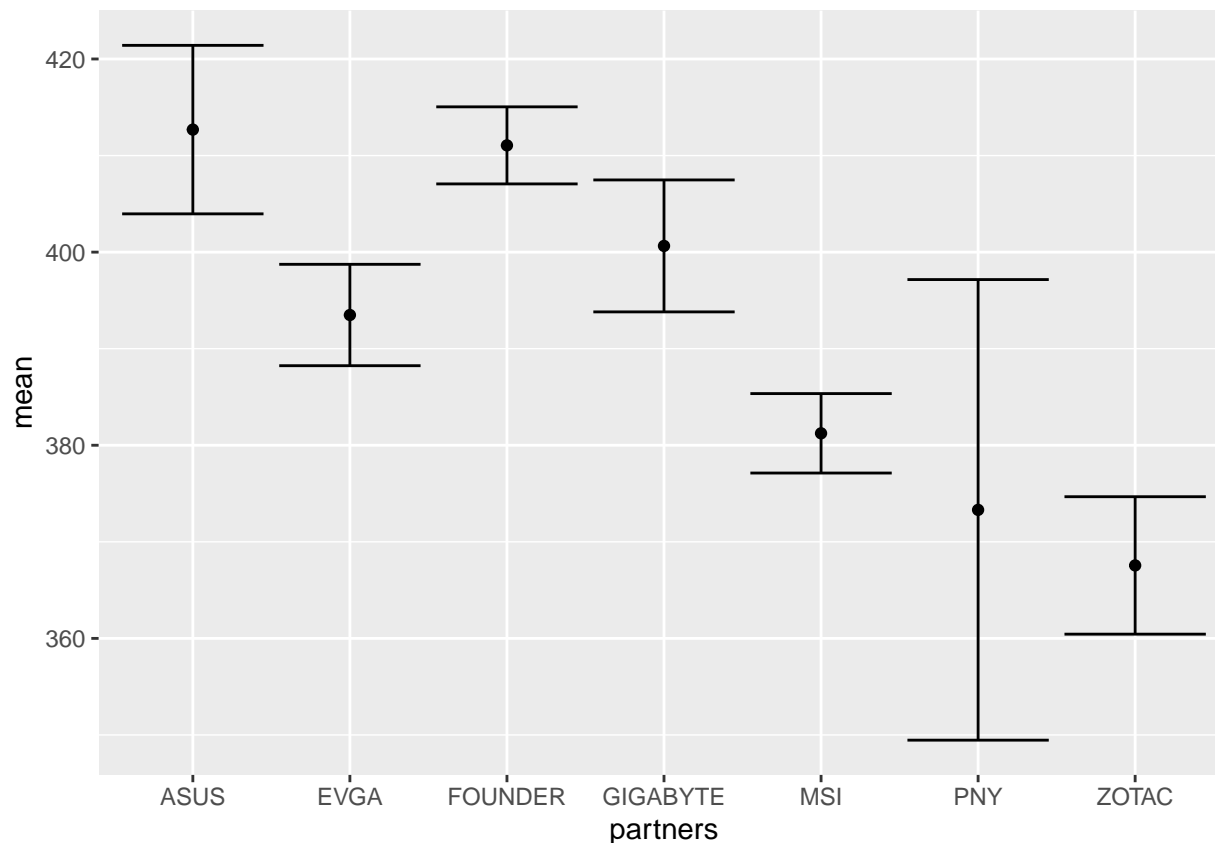
intervals = data.frame()
for (partner in partners) {
  distr = dat_model[dat_model$Board_Partner == partner, ]$Total_Price
  distr = na.omit(distr)
  CI = mean_CI(distr, 0.05)
  intervals[partner, "count"] = length(distr)
  intervals[partner, "mean"] = mean(distr)
  intervals[partner, "lower"] = CI[1]
  intervals[partner, "upper"] = CI[2]
}

print(intervals)
```

```
##          count      mean   lower   upper
## ASUS         322 412.6876 403.9627 421.4125
## EVGA          364 393.4879 388.2376 398.7383
## FOUNDER       226 411.0535 407.0631 415.0440
## GIGABYTE      254 400.6428 393.8104 407.4751
## MSI           282 381.2356 377.1223 385.3489
## PNY            44 373.3098 349.4616 397.1579
## ZOTAC         133 367.5556 360.4392 374.6720
```

Finally, we can plot the data:

```
ggplot(intervals, aes(partners, mean)) + geom_point() + geom_errorbar(aes(ymin = lower,
  ymax = upper))
```



Many of these intervals do not include the mean of the others and so they represent a 95% confident assertion that the true mean of, e.g. the ASUS 3070s, is different from the true mean of the ZOTAC 3070s. Typically, ZOTAC and PNY are seen as second-tier brands by consumers (or at least they have been in the past) and that seems to emerge from the data as well. However, caution should be exercised due to the possibility of skewed data due to the imbalance of observations for some of the partners more than others. For example, the count for PNY is only 44 and for ZOTAC only 133; this relatively less available data gives more room for error, especially for the PNY cards, as is reflected in the large confidence interval.

4. Conclusion

This paper has sought to synthesize a few interesting applications of basic statistical methods to a relatively unique application, providing insight into the secondhand GPU market on eBay as of Q1 2023. The process of data collection and analysis in R, though both relatively new experiences for the author, were incredibly instructive. In the process, it became clear that notwithstanding large *access* to data, proper collection and analysis of that data is critical to find useful results; hopefully, the analysis here can shed light on some of these insights.

The data set created for this project is intended to remain freely accessible in the form of a Google Sheet and can be found [here](#).