

# EDA

2025-07-12

## iPhones from eBay

```
library(ggplot2)
library(tidyverse)
```

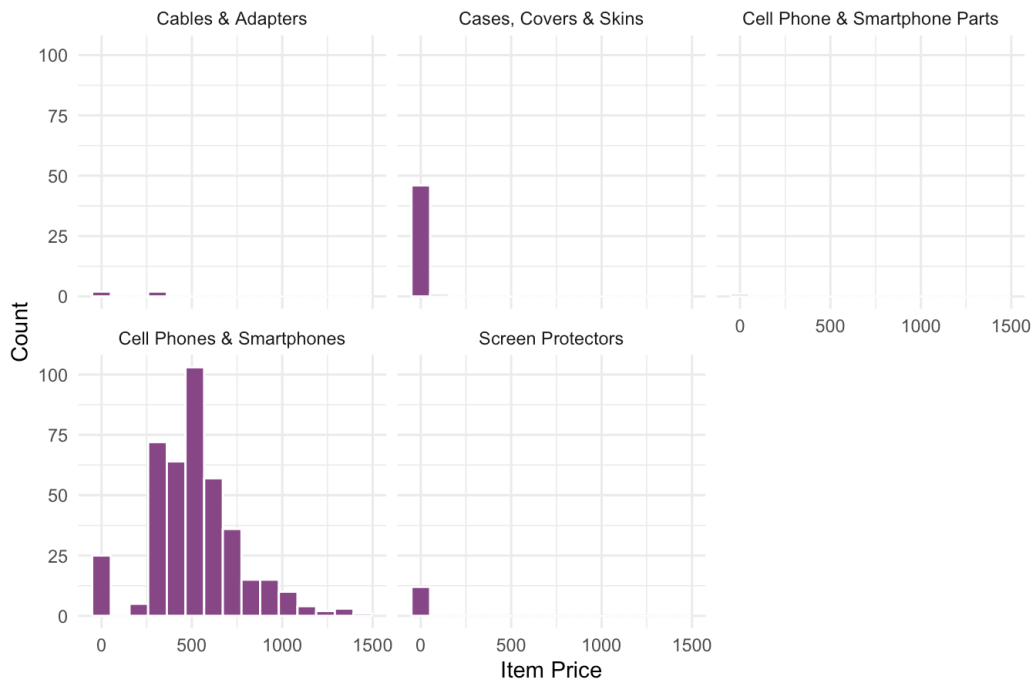
```
## — Attaching core tidyverse packages — tidyverse 2.0.0 —
## ✓ dplyr      1.1.4      ✓ readr      2.1.5
## ✓ forcats    1.0.0      ✓ stringr    1.5.1
## ✓ lubridate  1.9.4      ✓ tibble     3.2.1
## ✓ purrr      1.0.2      ✓ tidyr      1.3.1
## — Conflicts — tidyverse_conflicts() —
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()     masks stats::lag()
## ⓘ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
iphones = read.csv("data/iphone_cleaned.csv")
#head(iphones)
#str(iphones)
summary(iphones)
```

```
##      itemId      title      condition      conditionId
## Length:476      Length:476      Length:476      Min.   :1000
## Class :character Class :character Class :character 1st Qu.:1500
## Mode  :character Mode  :character Mode  :character Median :2020
##                                     Mean   :2049
##                                     3rd Qu.:3000
##                                     Max.   :7000
##
## price.value      price.currency      seller.username
## Min.   : 4.49      Length:476      Length:476
## 1st Qu.: 319.00      Class :character Class :character
## Median : 473.50      Mode  :character Mode  :character
## Mean   : 454.21
## 3rd Qu.: 609.99
## Max.   :1449.99
##
## seller.feedbackPercentage seller.feedbackScore itemLocation.country
## Min.   : 0.00      Min.   : 0      Length:476
## 1st Qu.: 97.88      1st Qu.: 4342      Class :character
## Median : 98.30      Median : 22977      Mode  :character
## Mean   : 94.82      Mean   : 66320
## 3rd Qu.: 98.80      3rd Qu.: 115563
## Max.   :100.00      Max.   :1044914
##
## marketingPrice.originalPrice.value marketingPrice.discountPercentage
## Min.   : 7.88      Min.   : 0.00
## 1st Qu.: 361.11      1st Qu.: 5.00
## Median : 799.00      Median :10.00
## Mean   : 691.54      Mean   :21.66
## 3rd Qu.: 999.00      3rd Qu.:43.75
## Max.   :1758.00      Max.   :67.00
## NA's   :331      NA's   :294
## shipping_cost      days_listed      category_id      category_name
## Min.   : 0.0000      Min.   : 0.0      Min.   : 9355      Length:476
## 1st Qu.: 0.0000      1st Qu.: 60.0      1st Qu.: 9355      Class :character
## Median : 0.0000      Median : 182.5      Median : 9355      Mode  :character
## Mean   : 0.2786      Mean   : 285.7      Mean   : 12710
## 3rd Qu.: 0.0000      3rd Qu.: 407.0      3rd Qu.: 9355
## Max.   :37.4000      Max.   :2015.0      Max.   :123422
## NA's   :17
## seller_item_count      model_number
## Min.   : 1.00      Min.   :14.00
## 1st Qu.: 2.00      1st Qu.:14.00
## Median : 6.00      Median :15.00
## Mean   :16.71      Mean   :15.08
## 3rd Qu.:24.00      3rd Qu.:16.00
## Max.   :58.00      Max.   :16.00
##
```

```
ggplot(iphones) +
  geom_histogram(aes(price.value), color="white", fill="orchid4", bins=15) +
  facet_wrap(~category_name) +
  theme_minimal() +
  labs(x = "Item Price", title = "Frequency of Items by Category - Search Term: iPhone 16", y = "Count")
```

## Frequency of Items by Category - Search Term: iPhone 16

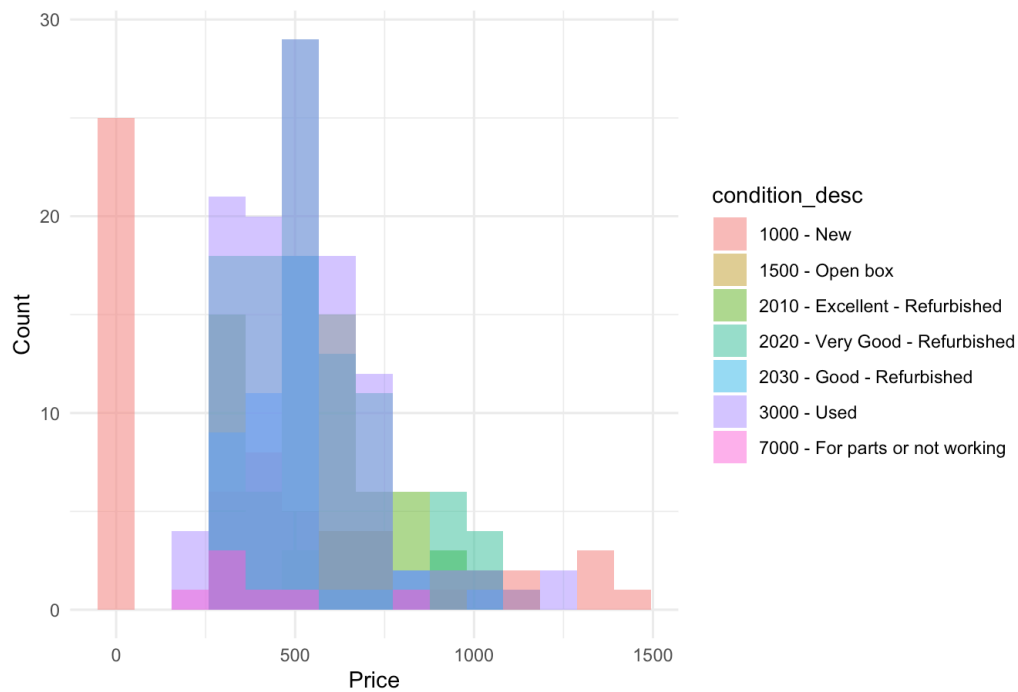


```
ggsave("images/histograms.png", bg="white")
```

```
## Saving 7 x 5 in image
```

```
iphones_only = iphones %>%  
  filter(category_id=="9355")  
  
#combine condition id and condition  
iphones_only$condition_desc = paste(iphones_only$conditionId, "-", iphones_only$condition)  
  
ggplot(iphones_only, aes(x = price.value, fill = condition_desc)) +  
  geom_histogram(alpha = 0.5, position = "identity", bins = 15) +  
  theme_minimal() +  
  labs(title = "Histogram of Phones by Condition", x = "Price", y = "Count")
```

# Histogram of Phones by Condition

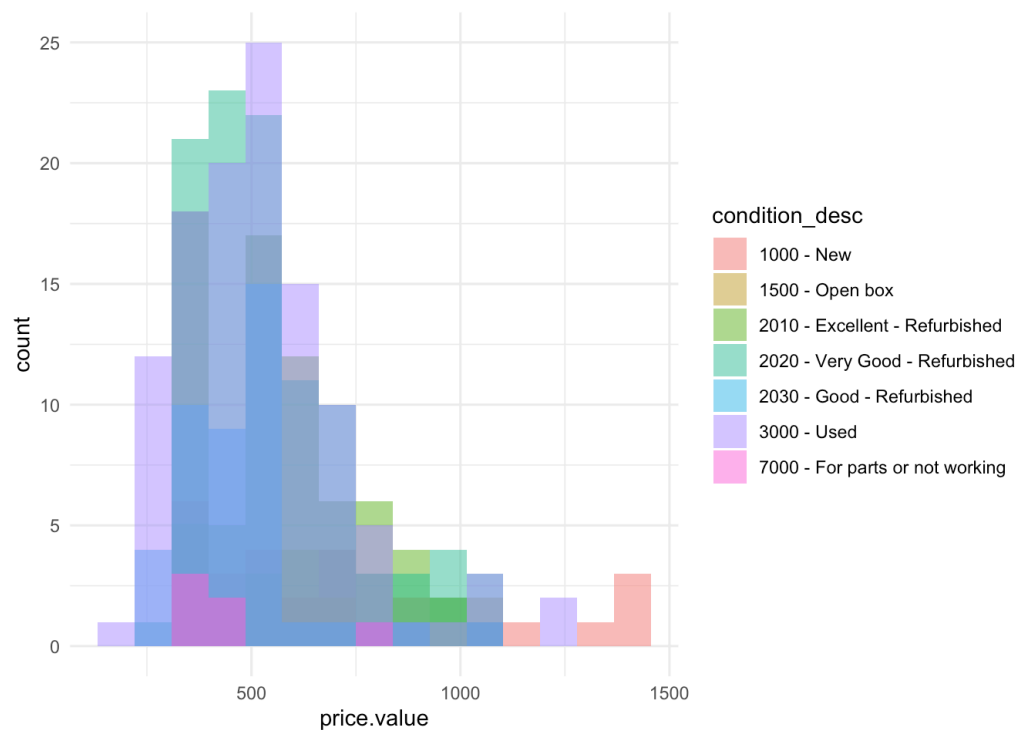


```
ggsave("images/hist_cond.png", bg="white")
```

```
## Saving 7 x 5 in image
```

```
iphones_only = iphones_only %>%
  filter(iphones_only$price.value > 200)

ggplot(iphones_only, aes(x = price.value, fill = condition_desc)) +
  geom_histogram(alpha = 0.5, position = "identity", bins = 15) +
  theme_minimal()
```



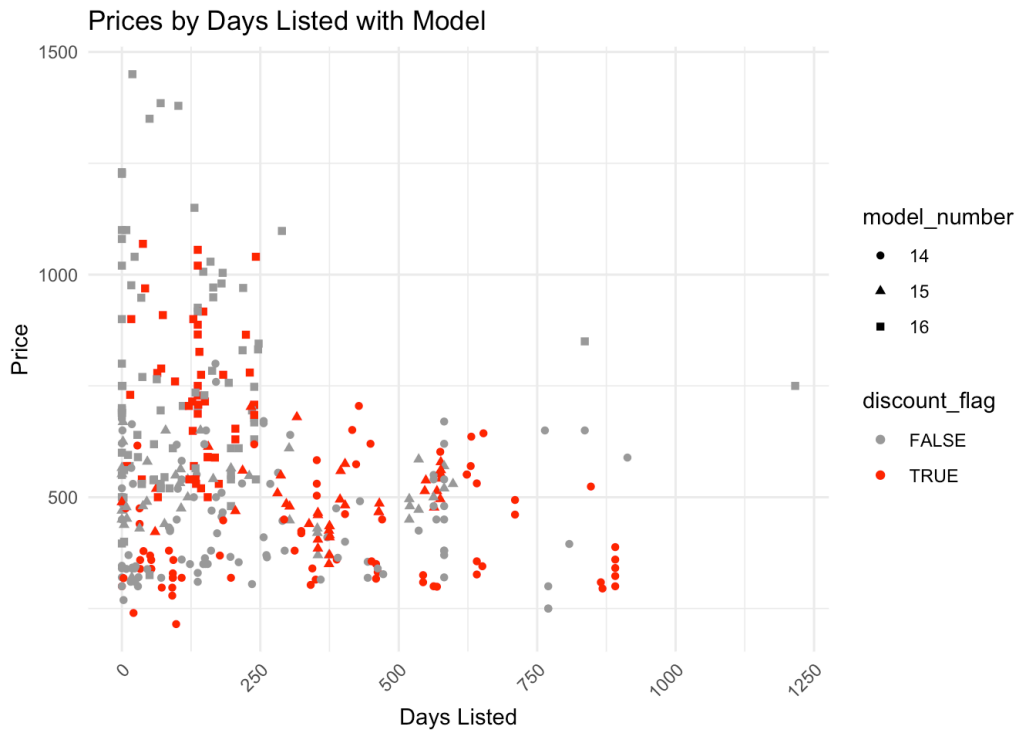
```

iphones_only$discount_flag = !is.na(iphones_only$marketingPrice.discountPercentage) &
                              iphones_only$marketingPrice.discountPercentage != ""

iphones_only$model_number = as.factor(iphones_only$model_number)

ggplot(iphones_only, aes(x = days_listed, y = price.value)) +
  geom_point(aes(color = discount_flag, shape = model_number)) +
  scale_color_manual(values = c("FALSE" = "grey60", "TRUE" = "red")) +
  theme_minimal() +
  labs(x = "Days Listed", y = "Price", title = "Prices by Days Listed with Model") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```



```

ggsave("images/items_by_days.png", bg="white")

```

```

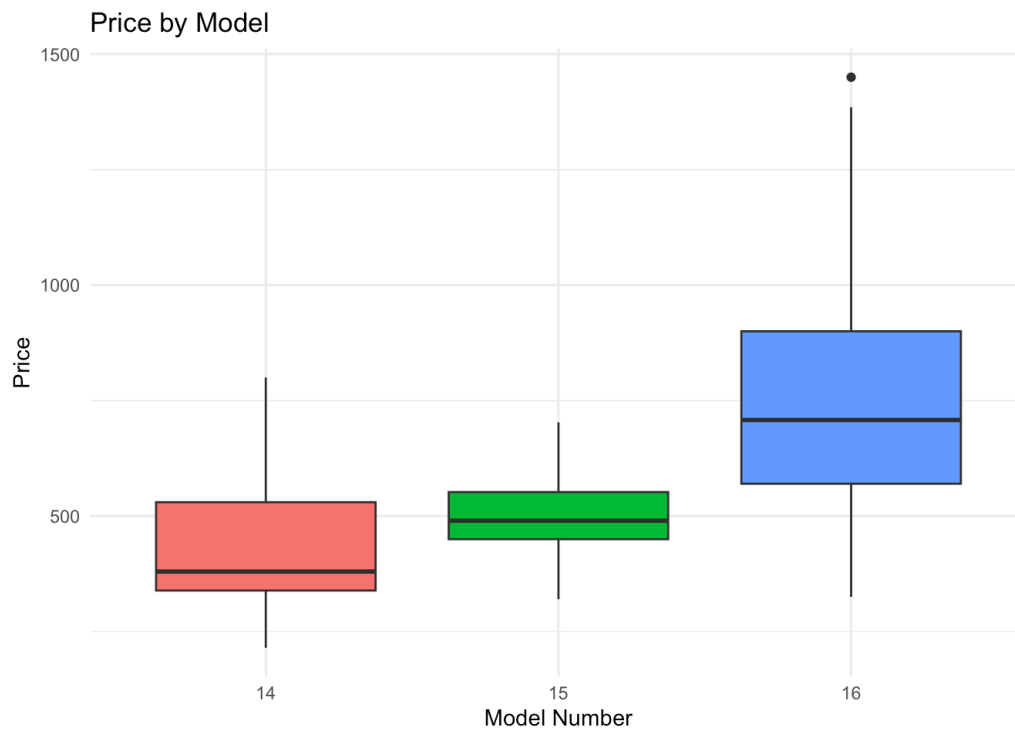
## Saving 7 x 5 in image

```

```

ggplot(iphones_only) +
  geom_boxplot(mapping = aes(x = model_number, y = price.value, fill = model_number)) +
  theme_minimal() +
  labs(x = "Model Number", y = "Price", title = "Price by Model") +
  theme(legend.position="none")

```

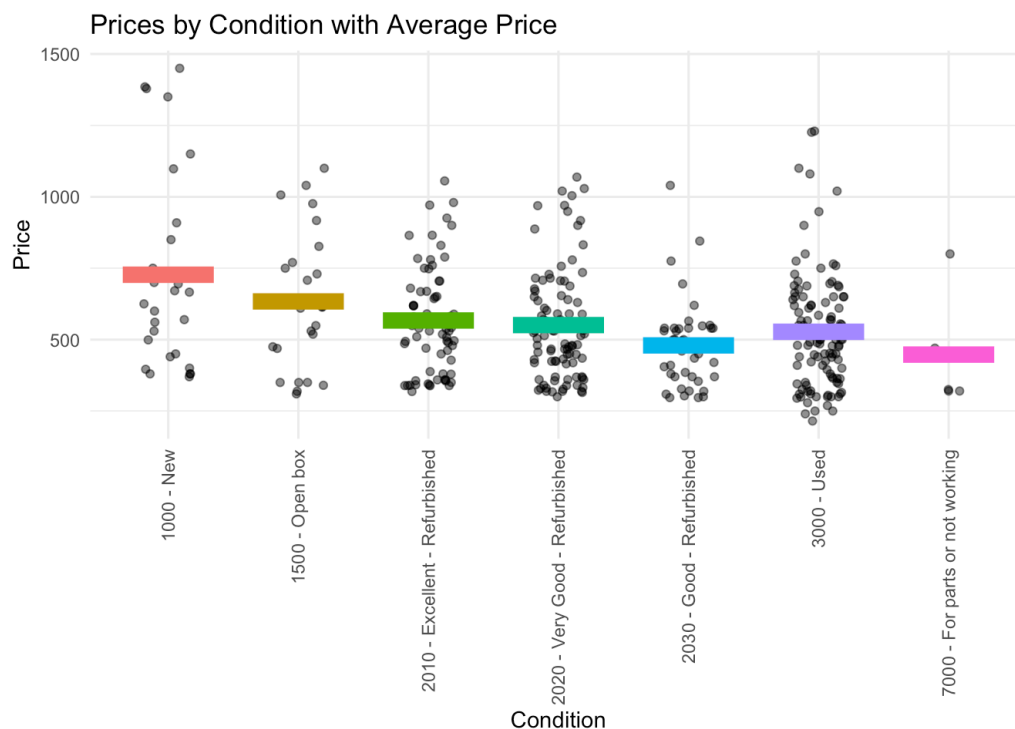


```
lmod = lm(price.value ~ condition + model_number, data = iphones_only)
#summary(lmod)

avg_prices <- iphones_only %>%
  group_by(condition_desc) %>%
  summarize(mean_price = mean(price.value, na.rm = TRUE))

ggplot(iphones_only, aes(x = condition_desc, y = price.value, color = condition_desc)) +
  geom_jitter(width = 0.2, alpha = 0.5, color = "black") + # individual points
  stat_summary(fun = mean, geom = "crossbar", width = 0.7, size = 1.5) + # bold average line
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  labs(x = "Condition", y = "Price", title = "Prices by Condition with Average Price") +
  theme(legend.position="none")
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



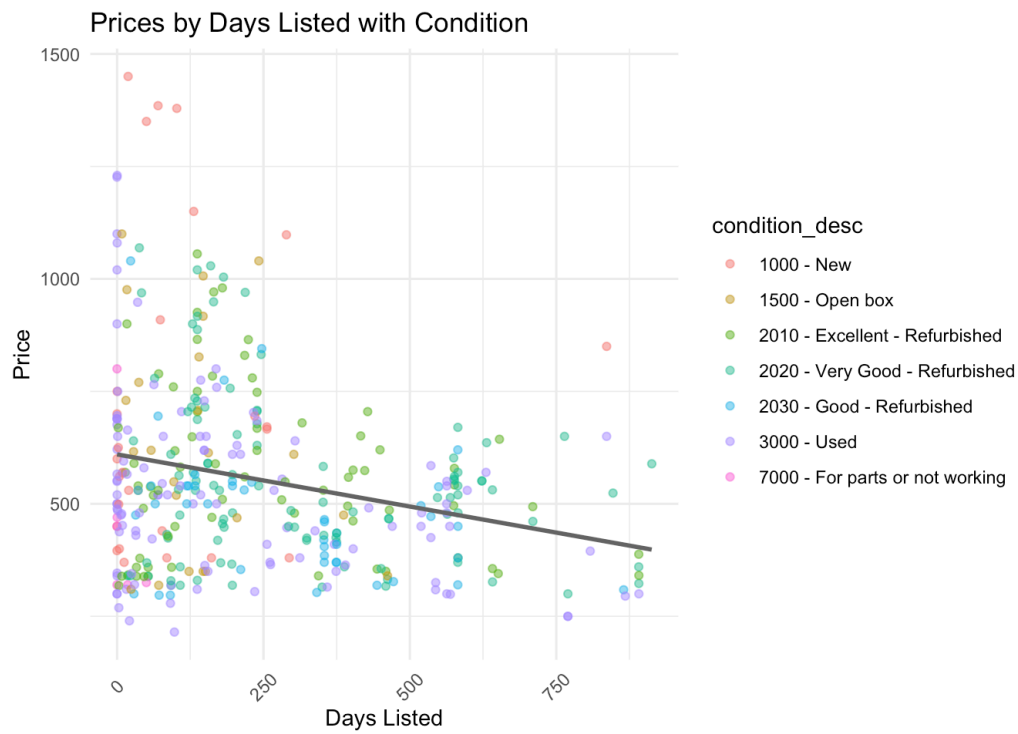
```
ggsave("images/price_by_cond.png", bg="white")
```

```
## Saving 7 x 5 in image
```

```
iphones_only = iphones_only %>%
  filter(iphones_only$days_listed < 1000)

ggplot(iphones_only, aes(x = days_listed, y = price.value, color = condition_desc)) +
  geom_jitter(width = 0.2, alpha = 0.5) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, vjust = 0.5, hjust = 0.5)) +
  labs(x = "Days Listed", y = "Price", title = "Prices by Days Listed with Condition") +
  geom_smooth(aes(group=1), method="lm", color="grey40", se=FALSE)
```

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



```
ggsave("images/prices_bydays.png", bg="white")
```

```
## Saving 7 x 5 in image
## `geom_smooth()` using formula = 'y ~ x'
```

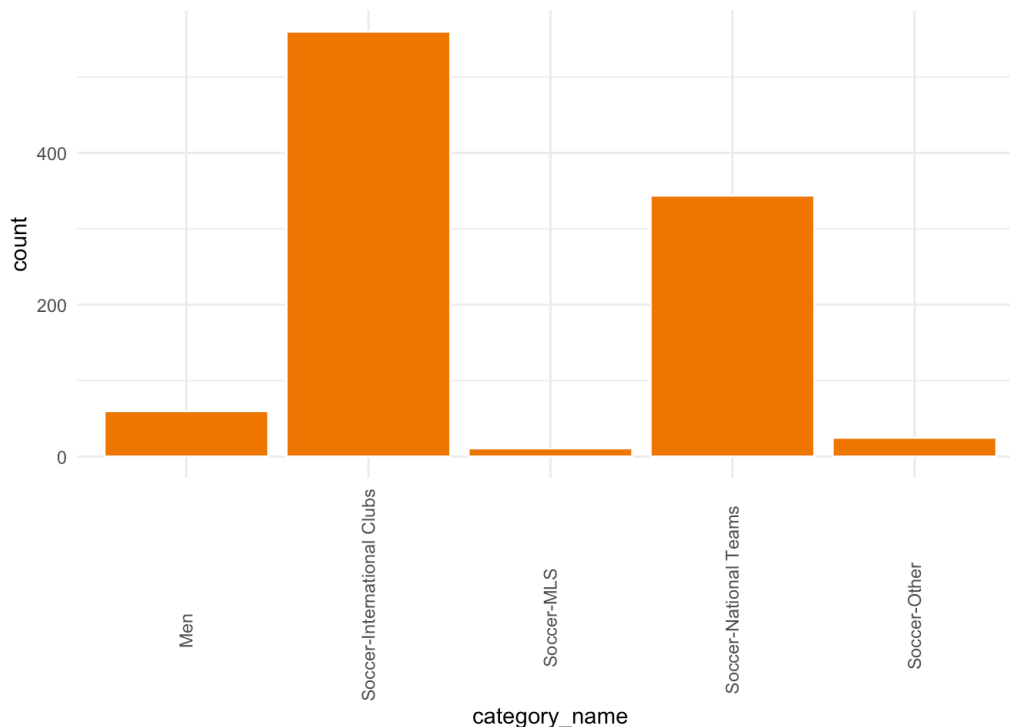
## Soccer Jerseys from eBay

```
jerseys = read.csv("data/soccer_jerseys.csv")
jerseys$condition_desc = paste(jerseys$conditionId, "-", jerseys$condition)
jerseys$condition = as.factor(jerseys$condition)
jerseys$itemLocation.country = as.factor(jerseys$itemLocation.country)
jerseys$category_id = as.factor(jerseys$category_id)
jerseys$club = as.factor(jerseys$club)
jerseys$country = as.factor(jerseys$country)
str(jerseys)
```



```
## 'data.frame': 1000 obs. of 21 variables:
## $ itemId : chr "v1|317069403907|615526711901" "v1|177245638563|476899094411" "v1|317059278518|615510384436" "v1|388511742986|655644751940" ...
## $ title : chr "Liverpool Home Jersey 25/26 ( Player Version)" "Tigres UANL Away Jersey 2023/24 Size S-XL" "Real Madrid Baseball Style Jersey Limited Edition" "Neymar Jr Santos 2012 Jersey - Retro Jersey - Mens Sizes" ...
## $ conditionId : int 1000 1000 1000 1000 1000 1000 1000 1000 1000 1000 ...
## $ condition : Factor w/ 6 levels "New","New with tags",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ price.value : num 45 18 52 65 35 ...
## $ price.currency : chr "USD" "USD" "USD" "USD" ...
## $ seller.username : chr "iri4147" "soccerelf" "iri4147" "shipezusa" ...
## $ seller.feedbackPercentage : num 97.7 100 97.7 97.1 100 95.6 95.8 100 97.7 100 ...
## $ seller.feedbackScore : int 47 10 47 84 10 166 1092 10 47 10 ...
## $ itemLocation.country : Factor w/ 33 levels "AE","AR","BD",...: 33 33 33 33 33 33 33 33 33 33 ...
...
## $ marketingPrice.originalPrice.value: num NA NA NA NA NA NA NA NA NA NA ...
## $ marketingPrice.discountPercentage : int NA NA NA NA NA NA NA NA NA NA ...
## $ shipping_cost : num 0 0 0 0 0 0 0 0 0 0 ...
## $ days_listed : int 3 7 6 122 7 103 1 7 4 7 ...
## $ category_id : Factor w/ 5 levels "2885","2887",...: 2 2 2 2 2 2 2 2 2 2 ...
## $ category_name : chr "Soccer-International Clubs" "Soccer-International Clubs" "Soccer-International Clubs" "Soccer-International Clubs" ...
## $ seller_item_count : int 25 8 25 11 8 8 15 8 25 8 ...
## $ club : Factor w/ 13 levels "", "Arsenal", "Barcelona",...: 9 1 13 1 1 3 3 1 3 1 ...
...
## $ country : Factor w/ 11 levels "", "Argentina",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ year : int 2026 2024 NA 2012 2026 2011 NA 2025 2026 2025 ...
## $ condition_desc : chr "1000 - New" "1000 - New" "1000 - New" "1000 - New" ...
```

```
ggplot(jerseys) +
  geom_bar(aes(category_name), fill = "darkorange2", color = "white") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 0.3))
```



```

jerseys$top_club = jerseys$club != ""
jerseys$top_country = jerseys$country != ""
jerseys$from_usa = jerseys$itemLocation.country == "US"

club_jerseys = jerseys %>%
  filter(category_id == 2887)
country_jerseys = jerseys %>%
  filter(category_id == 2891)

library(patchwork)
all_prices = c(club_jerseys$price.value, country_jerseys$price.value)
y_min = min(all_prices, na.rm = TRUE)
y_max = max(all_prices, na.rm = TRUE)
p1 = ggplot(club_jerseys) +
  geom_boxplot(aes(x = top_club, y = price.value, color = top_club)) +
  theme_minimal() +
  labs(title = "Price for Top Clubs (Club Jerseys Only)", x = "Top Clubs", y = "Price") +
  theme(legend.position = "none") +
  ylim(y_min, y_max)

p2 = ggplot(country_jerseys) +
  geom_boxplot(aes(x = top_country, y = price.value, color = top_country)) +
  theme_minimal() +
  labs(title = "Price for Top Countries (National Team Jerseys Only)", x = "Top Countries", y = "Price") +
  theme(legend.position = "none") +
  ylim(y_min, y_max)

combined_plot = p1 + p2

ggsave("images/top_clubs.png", combined_plot, bg="white", width = 14, height = 6, dpi = 300)

```

```

lmod_soccer = lm(price.value ~ condition + seller.feedbackPercentage + seller.feedbackScore + itemLocation.country +
  category_id + shipping_cost + days_listed + top_club + country + year, data = jerseys)
#summary(lmod_soccer)
cat("AIC:", AIC(lmod_soccer), "\n")

```

```
## AIC: 5989.908
```

```

png("images/qqplot.png")
QQ = plot(lmod_soccer, 2)

```

```

## Warning: not plotting observations with leverage one:
## 46, 219, 380, 500, 516, 545, 555, 559, 599

```

```
dev.off()
```

```

## quartz_off_screen
## 2

```

```

jerseys$log_price = log(jerseys$price.value) #add log-price column
jerseys_clean = na.omit(jerseys) #omit NA
lm_log = lm(log_price ~ condition + seller.feedbackPercentage +
  itemLocation.country + category_id + shipping_cost + top_club, data = jerseys_clean)
#summary(lm_log)
AIC(lm_log)

```

```
## [1] 49.64011
```

```
png("images/qqplot_log_price.png")
resids = resid(lm_log)
qqnorm(resids, main = "Q-Q Plot using Log Price")
qqline(resids, col = "darkblue", lwd = 1, lty = "dashed")
dev.off()
```

```
## quartz_off_screen
##                2
```

```
library(car)
```

```
## Loading required package: carData
```

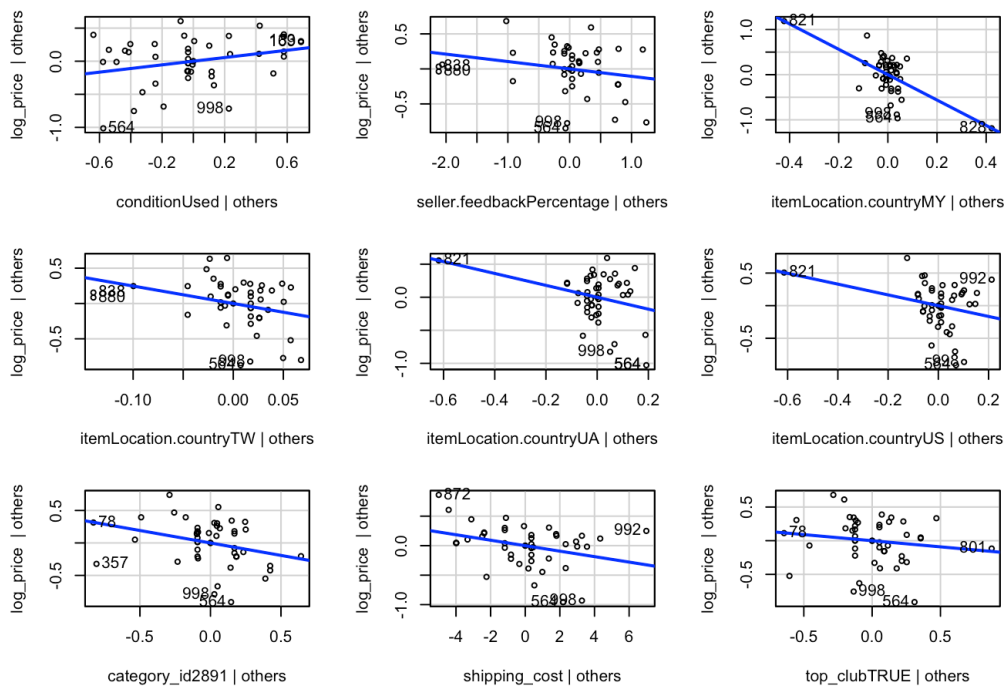
```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##      recode
```

```
## The following object is masked from 'package:purrr':
##
##      some
```

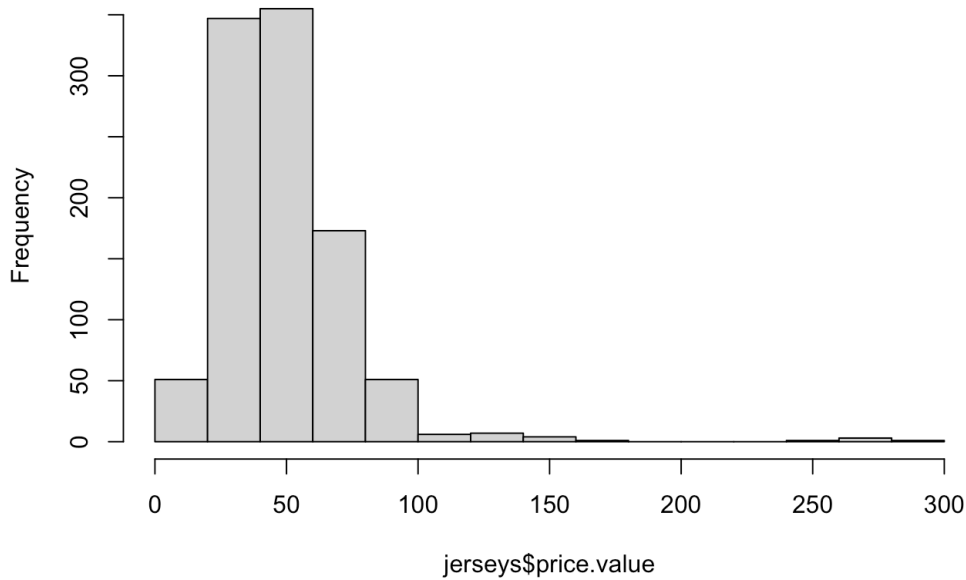
```
avPlots(lm_log)
```

### Added-Variable Plots



```
hist(jerseys$price.value)
```

## Histogram of jerseys\$price.value



```
#install.packages("gamlss")  
library(gamlss) #for other distributions
```

```
## Loading required package: splines
```

```
## Loading required package: gamlss.data
```

```
##  
## Attaching package: 'gamlss.data'
```

```
## The following object is masked from 'package:datasets':  
##  
## sleep
```

```
## Loading required package: gamlss.dist
```

```
## Loading required package: nlme
```

```
##  
## Attaching package: 'nlme'
```

```
## The following object is masked from 'package:dplyr':  
##  
## collapse
```

```
## Loading required package: parallel
```

```
## ***** GAMLSS Version 5.4-22 *****
```

```
## For more on GAMLSS look at https://www.gamlss.com/
```

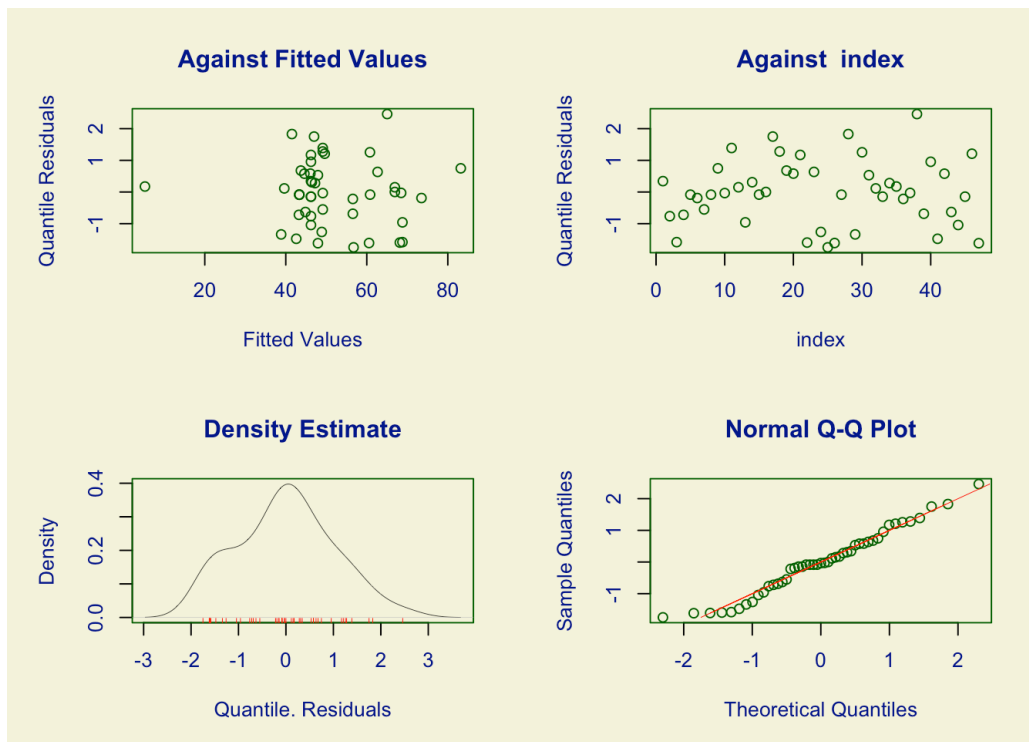
```
## Type gamlssNews() to see new features/changes/bug fixes.
```

```
lmod_soccer2 = gamlss(price.value ~ condition + seller.feedbackPercentage + itemLocation.country +
                      category_id + shipping_cost + top_club, data = jerseys_clean, family = "ST1")
```

```
## GAMLSS-RS iteration 1: Global Deviance = 387.46
## GAMLSS-RS iteration 2: Global Deviance = 382.2987
## GAMLSS-RS iteration 3: Global Deviance = 381.3021
## GAMLSS-RS iteration 4: Global Deviance = 379.8304
## GAMLSS-RS iteration 5: Global Deviance = 379.7584
## GAMLSS-RS iteration 6: Global Deviance = 382.5324
## GAMLSS-RS iteration 7: Global Deviance = 378.1871
## GAMLSS-RS iteration 8: Global Deviance = 378.355
## GAMLSS-RS iteration 9: Global Deviance = 377.6372
## GAMLSS-RS iteration 10: Global Deviance = 377.8513
## GAMLSS-RS iteration 11: Global Deviance = 383.089
## GAMLSS-RS iteration 12: Global Deviance = 381.5771
## GAMLSS-RS iteration 13: Global Deviance = 386.0966
## GAMLSS-RS iteration 14: Global Deviance = 377.7339
## GAMLSS-RS iteration 15: Global Deviance = 377.602
## GAMLSS-RS iteration 16: Global Deviance = 379.6855
## GAMLSS-RS iteration 17: Global Deviance = 377.8786
## GAMLSS-RS iteration 18: Global Deviance = 377.994
## GAMLSS-RS iteration 19: Global Deviance = 377.6332
## GAMLSS-RS iteration 20: Global Deviance = 377.588
```

```
## Warning in RS(): Algorithm RS has not yet converged
```

```
plot(lmod_soccer2)
```



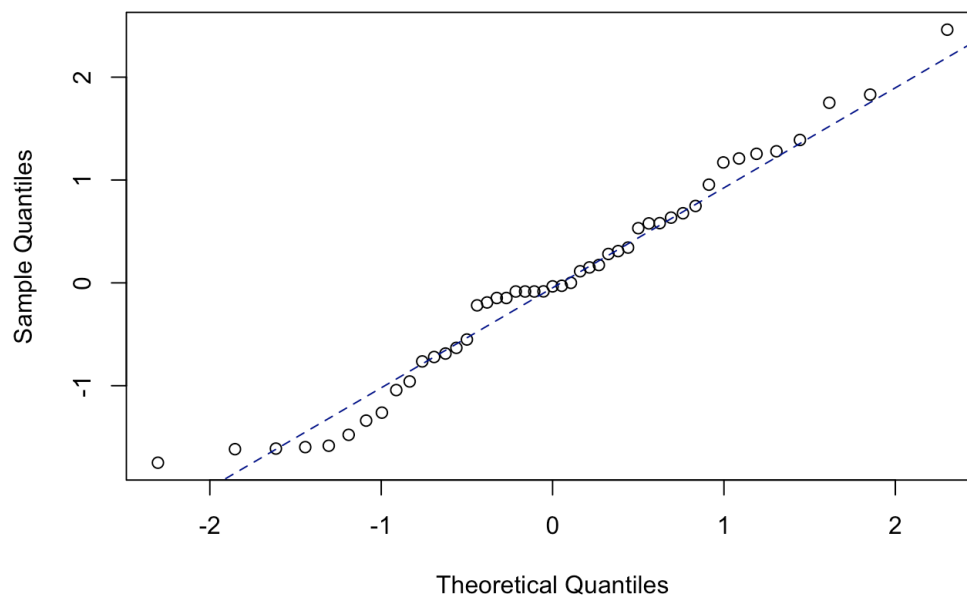
```
## *****
##      Summary of the Quantile Residuals
##              mean = -0.005812045
##              variance = 1.031057
##              coef. of skewness = 0.1385793
##              coef. of kurtosis = 2.433816
## Filliben correlation coefficient = 0.9888599
## *****
```

```
cat("\n", "AIC: ", lmod_soccer2$aic, "\n")
```

```
##  
## AIC: 403.588
```

```
resids = residuals(lmod_soccer2, type = "simple")  
qqnorm(resids, main = "Q-Q Plot using Student Skew t distribution with gamlss")  
qqline(resids, col = "darkblue", lwd = 1, lty = "dashed")
```

### Q-Q Plot using Student Skew t distribution with gamlss



## Microwaves from Amazon

```
microwaves = read.csv("data/microwaves.csv")  
microwaves$price_list_price = as.numeric(gsub("\\$", "", microwaves$price_list_price))
```

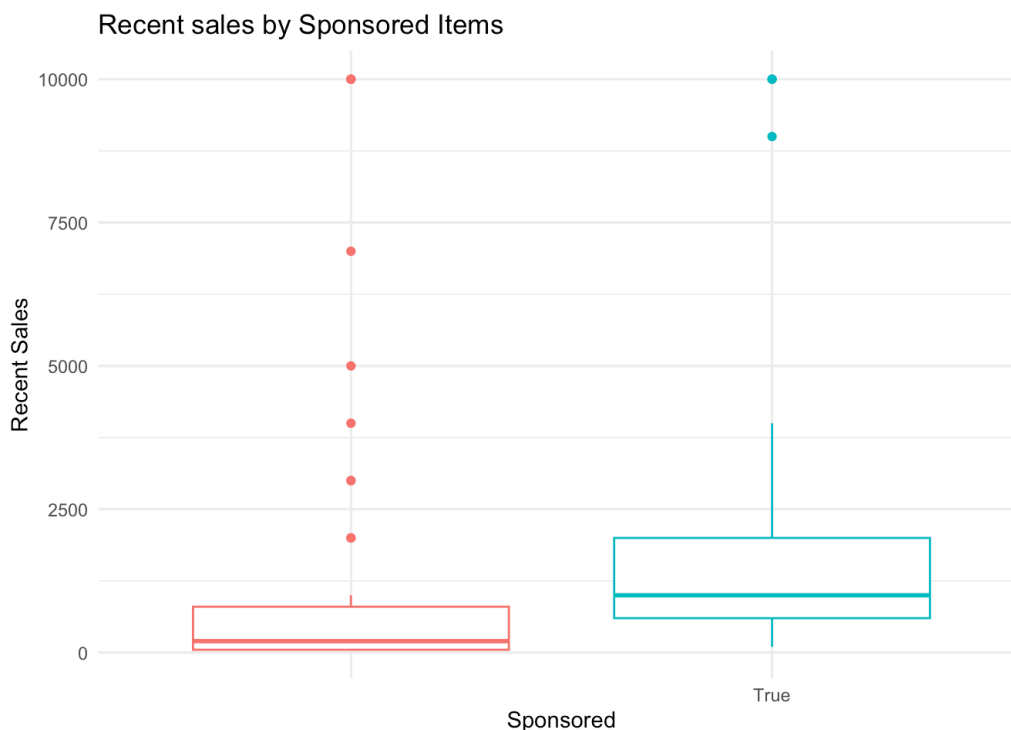
```
## Warning: NAs introduced by coercion
```

```
microwaves$price_name = as.factor(microwaves$price_name)  
microwaves$sponsored = as.factor(microwaves$sponsored)  
microwaves$is_small_business = as.factor(microwaves$is_small_business)  
microwaves$stock_info = as.factor(microwaves$stock_info)  
microwaves$is_prime = as.factor(microwaves$is_prime)  
str(microwaves)
```

```
## 'data.frame': 137 obs. of 15 variables:
## $ position : int 1 2 3 5 7 8 9 10 11 12 ...
## $ title : chr "Chefman Countertop Microwave Oven 0.7 Cu. Ft., 700 Watts with 10 Power Levels, 6 C
ooking Presets with One-Touch"| __truncated__ "Farberware Countertop Microwave 1000 Watts, 1.1 cu ft - Microwave
Oven With LED Lighting and Child Lock - Perfe"| __truncated__ "TOSHIBA EM131A5C-BS Countertop Microwave Ovens 1.2
Cu Ft, 12.4\" Removable Turntable Smart Humidity Sensor 12 A"| __truncated__ "5304522796 Microwave Door Latch Com
patible with Brands Electrolux, Frigidaire, Gibson, Kelvinator, Westinghouse"| __truncated__ ...
## $ asin : chr "B0DY95T5HB" "B01EIZSF6I" "B071WCB1T6" "B0F5WPTLZS" ...
## $ is_prime : Factor w/ 2 levels "False","True": 1 1 1 1 2 1 2 1 1 1 ...
## $ rating : num 4.3 4.4 4.4 5 4.4 4 4.5 4.3 NA 4.4 ...
## $ ratings_total : num 1396 23609 54371 1 11371 ...
## $ sponsored : Factor w/ 2 levels "", "True": 2 2 1 1 2 2 2 1 2 1 ...
## $ is_small_business: Factor w/ 2 levels "", "True": 1 1 1 1 1 1 1 1 1 1 ...
## $ coupon : chr "" "" "" "" "" ...
## $ cu_ft : num 0.7 1.1 1.2 NA 1.9 NA NA 0.7 NA 0.7 ...
## $ recent_sales_num : num 9000 4000 10000 50 400 800 100 10000 NA 10000 ...
## $ price_value : num 65 120 150 20 343 ...
## $ price_list_price : num 100 130 NA 22 NA ...
## $ price_name : Factor w/ 7 levels "", "Limited time deal",...: 2 3 1 3 1 3 1 1 3 1 ...
## $ stock_info : Factor w/ 19 levels "", "Only 1 left in stock - order soon.",...: 1 1 1 1 1 1 1 1 1 1 ...
```

```
ggplot(microwaves) +
  geom_boxplot(aes(x=sponsored, y = recent_sales_num, color = sponsored)) +
  theme_minimal() +
  theme(legend.position = "none") +
  labs(x="Sponsored", y = "Recent Sales", title = "Recent sales by Sponsored Items")
```

```
## Warning: Removed 39 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```



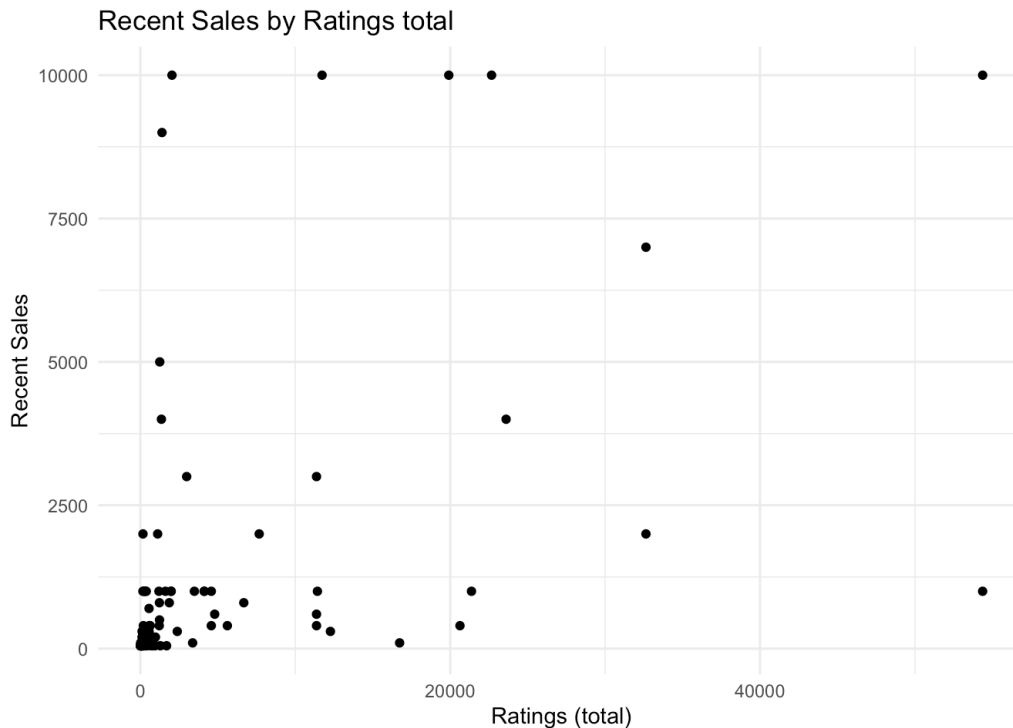
```
ggsave("images/recent_spon.png", bg="white")
```

```
## Saving 7 x 5 in image
```

```
## Warning: Removed 39 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```

```
ggplot(microwaves) +
  geom_point(aes(x=ratings_total, y=recent_sales_num)) +
  theme_minimal() +
  labs(x="Ratings (total)", y="Recent Sales", title="Recent Sales by Ratings total")
```

```
## Warning: Removed 39 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



```
ggsave("images/sales_by_rat.png", bg="white")
```

```
## Saving 7 x 5 in image
```

```
## Warning: Removed 39 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

## Lego from Amazon

```
lego = read.csv("data/lego.csv")
lego$price_list_price = as.numeric(gsub("\\$", "", lego$price_list_price))
```

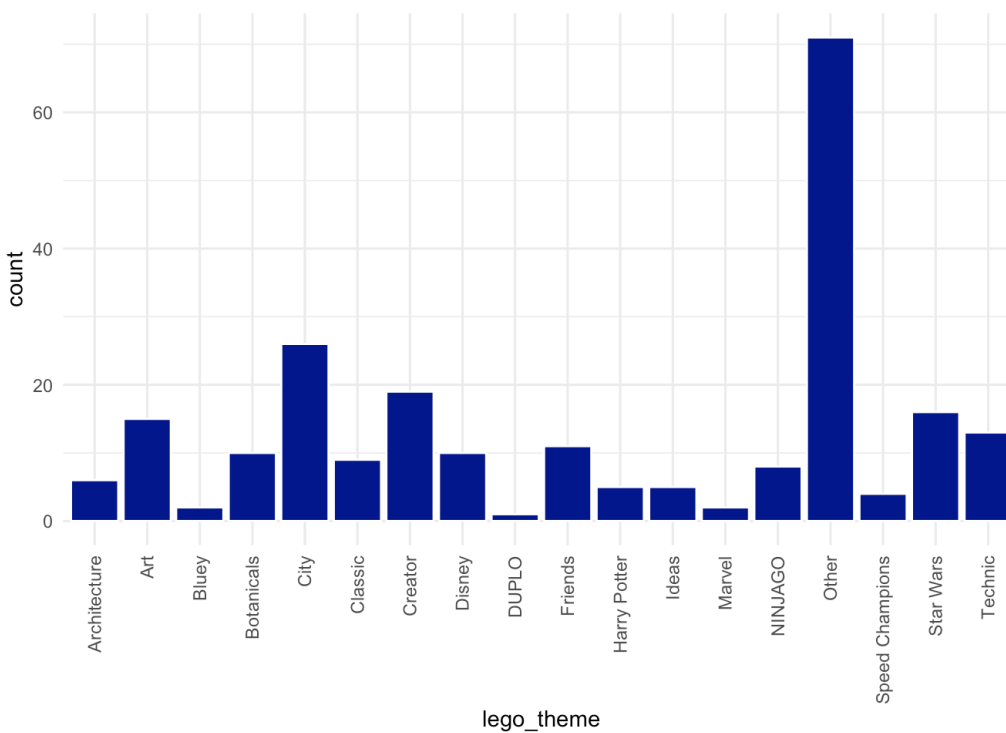
```
## Warning: NAs introduced by coercion
```

```
lego$price_name = as.factor(lego$price_name)
lego$sponsored = as.factor(lego$sponsored)
lego$stock_info = as.factor(lego$stock_info)
lego$is_prime = as.factor(lego$is_prime)
lego$age_rec = as.factor(lego$age_rec)
lego$lego_theme = as.factor(lego$lego_theme)
str(lego)
```



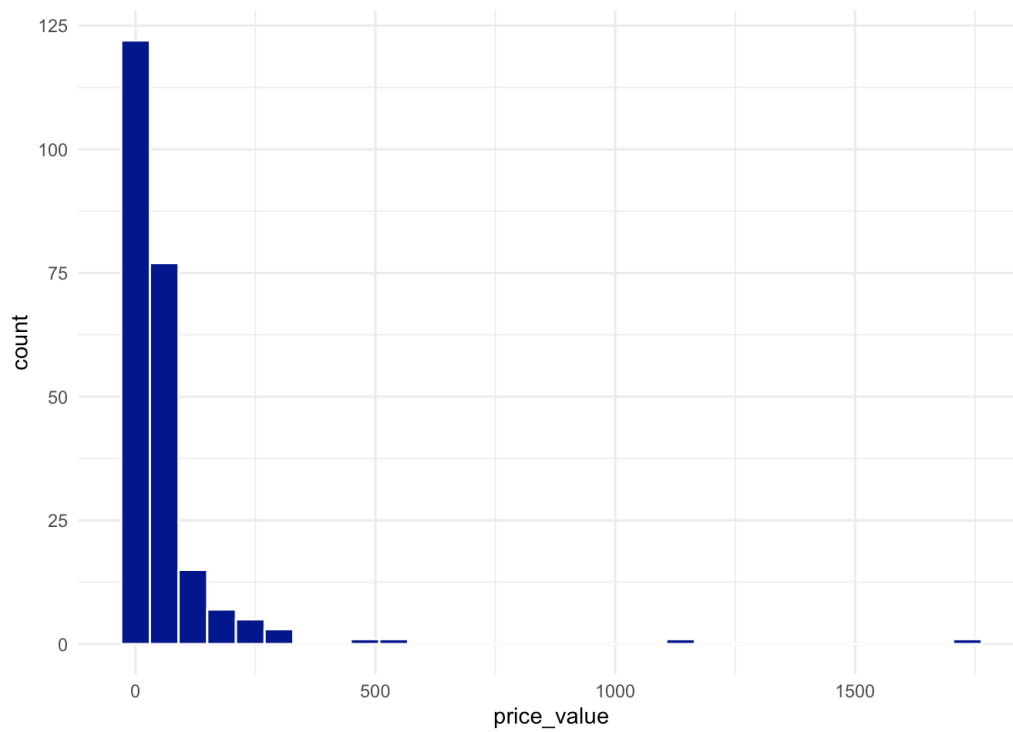
```
## 'data.frame':   233 obs. of  15 variables:
## $ position      : int   1 2 3 4 5 6 7 8 9 10 ...
## $ title         : chr   "LEGO Bluey: Bluey's Family House - Toddler Toy for 4 Year Old Boys and Girls - House Playset Includes Bluey and"| __truncated__ "LEGO I Disney Princess Beauty and The Beast Castle Building Set - Room Decor for Women and Men, Ages 18+ - Mode"| __truncated__ "LEGO Star Wars: The Phantom Menace Battle Droid with STAP - Building & Learning Toy for Boys & Girls, Ages 12+ "| __truncated__ "LEGO Harry Potter Book Nook: Hogwarts Express - Kids Toy - DIY Kit and Pretend Play Building Set for Boys and Girls"| __truncated__ ...
## $ asin          : chr   "B0DRW8L7HY" "B0DRW654DB" "B0DRW6VY43" "B0DRW7WRX3" ...
## $ is_prime      : Factor w/ 2 levels "False","True": 1 1 1 1 1 1 1 1 1 1 ...
## $ rating        : num   4.7 5 4.2 4.8 4.8 4.8 4.7 4.8 4.8 4.9 ...
## $ ratings_total  : num   51 2 11 26 441 ...
## $ sponsored     : Factor w/ 2 levels "", "True": 2 2 2 2 1 1 1 1 1 1 ...
## $ coupon        : chr   "" "" "" "" "" ...
## $ lego_theme     : Factor w/ 18 levels "Architecture",...: 3 8 17 11 15 6 15 11 1 16 ...
## $ age_rec       : Factor w/ 11 levels "3","4","5","6",...: NA 11 9 NA 11 NA NA NA 9 11 ...
## $ recent_sales_num: num   2000 NA 700 1000 10000 20000 NA 2000 6000 5000 ...
## $ price_value    : num   70 280 140 100 214 ...
## $ price_list_price: num   NA NA NA NA 230 ...
## $ price_name     : Factor w/ 11 levels "", "4K", "Blu-ray",...: 1 1 1 1 7 7 1 7 7 7 ...
## $ stock_info     : Factor w/ 8 levels "", "Only 1 left in stock - order soon.",...: 1 1 1 1 1 1 1 1 1 1 ...
```

```
ggplot(lego) +
  geom_bar(aes(x=lego_theme), fill = "darkblue", color = "white") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))
```



```
ggplot(lego) +
  geom_histogram(aes(x=price_value), color="white", fill = "darkblue") +
  theme_minimal()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



## Summary

```
library(kableExtra)
```

```
##  
## Attaching package: 'kableExtra'
```

```
## The following object is masked from 'package:dplyr':  
##  
## group_rows
```

```

mean_iphone = mean(iphones_only$price.value)
med_iphone = median(iphones_only$price.value)
len_iphone = nrow(iphones_only)
mean_soccer = mean(jerseys$price.value)
med_soccer = median(jerseys$price.value)
len_soccer = nrow(jerseys)
mean_microwaves = mean(microwaves$price_value)
med_microwaves = median(microwaves$price_value)
len_microwaves = nrow(microwaves)
mean_lego = mean(lego$price_value)
med_lego = median(lego$price_value)
len_lego = nrow(lego)

init_feat_iphone = c("Condition, Seller Feedback, Discount, Shipping Cost, Days Listed, Seller Count, Model No.")
init_feat_soccer = c("Condition, Seller Feedback, Discount, Shipping Cost, Days Listed, Seller Count, Club, Country, Year")
init_feat_micro = c("Product Ratings, Sponsored, Size, Recent Sales, Amt off List Price")
init_feat_lego = c("Product Ratings, Sponsored, Theme, Age Rec, Recent Sales, Amt off List Price")

summary_df <- data.frame(
  category = c("iPhones", "Soccer Jerseys", "Microwaves", "LEGO"),
  mean_price = c(mean_iphone, mean_soccer, mean_microwaves, mean_lego),
  median_price = c(med_iphone, med_soccer, med_microwaves, med_lego),
  n = c(len_iphone, len_soccer, len_microwaves, len_lego),
  features = c(init_feat_iphone, init_feat_soccer, init_feat_micro, init_feat_lego)
)
#summary_df

table = summary_df %>%
  kbl(
    caption = "Summary Statistics",
    col.names = c("Category", "Mean Price", "Median Price", "No. of Obs.", "Potential Features"),
    digits = 2
  ) %>%
  kable_styling(full_width = FALSE, position = "center")

#install.packages("webshot2")
library(webshot2)
save_kable(table, file = "images/summary_table.png")

```

```
## file:///private/var/folders/vg/bxzl7hss0gs1hqxsc9zj92l80000gn/T/Rtmp6LDk0C/summary_table402e30b01ab9.html screenshot completed
```

```
## save_kable will have the best result with magick installed.
```

```
table
```

## Summary Statistics

Category	Mean Price	Median Price	No. of Obs.	Potential Features
iPhones	554.60	520.00	385	Condition, Seller Feedback, Discount, Shipping Cost, Days Listed, Seller Count, Model No.
Soccer Jerseys	50.63	48.00	1000	Condition, Seller Feedback, Discount, Shipping Cost, Days Listed, Seller Count, Club, Country, Year
Microwaves	263.32	140.14	137	Product Ratings, Sponsored, Size, Recent Sales, Amt off List Price
LEGO	62.68	27.99	233	Product Ratings, Sponsored, Theme, Age Rec, Recent Sales, Amt off List Price