Laptop Analysis

Augustus Chu

2025-02-28

Laptops Analysis

For this project, I will analyse an uncleaned dataset on Laptop Specs and their prices. The first section involves exploring, understanding then the cleaning the data for our purposes. The second section involves visualising the data. The main purpose of this project is to showcase what is possible with data analytics!

Exploratory Analysis + Data Cleaning

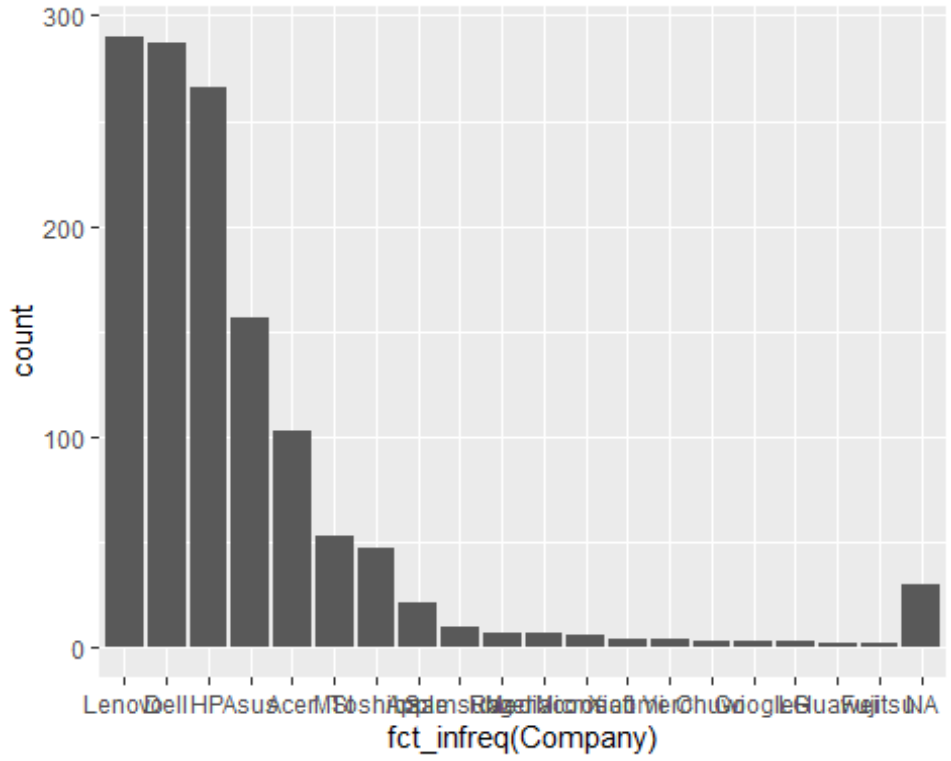
```
# Load packages
pacman::p_load(tidyverse, dplyr)

# Read data
laptop.csv <- read_csv("laptopData.csv")

## Rows: 1303 Columns: 12
## ── Column specification
──────────────────────────────────────────────────────
## Delimiter: ","
## chr (10): Company, TypeName, Inches, ScreenResolution, Cpu, Ram, Memory,
Gpu...
## dbl  (2): Unnamed: 0, Price
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this
message.

# Analyse the data closely
(laptop.csv)

## # A tibble: 1,303 × 12
##    `Unnamed: 0` Company TypeName  Inches ScreenResolution     Cpu   Ram
Memory
##           <dbl> <chr>   <chr>     <chr>  <chr>                <chr> <chr>
<chr>
##  1            0 Apple   Ultrabook 13.3   IPS Panel Retina Di… Inte… 8GB
128GB…
##  2            1 Apple   Ultrabook 13.3   1440x900             Inte… 8GB
128GB…
##  3            2 HP      Notebook  15.6   Full HD 1920x1080    Inte… 8GB
256GB…
##  4            3 Apple   Ultrabook 15.4   IPS Panel Retina Di… Inte… 16GB
512GB…
```

```
##  5            4 Apple   Ultrabook 13.3   IPS Panel Retina Di… Inte… 8GB
256GB…
##  6            5 Acer    Notebook  15.6   1366x768             AMD … 4GB
500GB…
##  7            6 Apple   Ultrabook 15.4   IPS Panel Retina Di… Inte… 16GB
256GB…
##  8            7 Apple   Ultrabook 13.3   1440x900             Inte… 8GB
256GB…
##  9            8 Asus    Ultrabook 14     Full HD 1920x1080    Inte… 16GB
512GB…
## 10            9 Acer    Ultrabook 14     IPS Panel Full HD 1… Inte… 8GB
256GB…
## # ℹ 1,293 more rows
## # ℹ 4 more variables: Gpu <chr>, OpSys <chr>, Weight <chr>, Price <dbl>
```

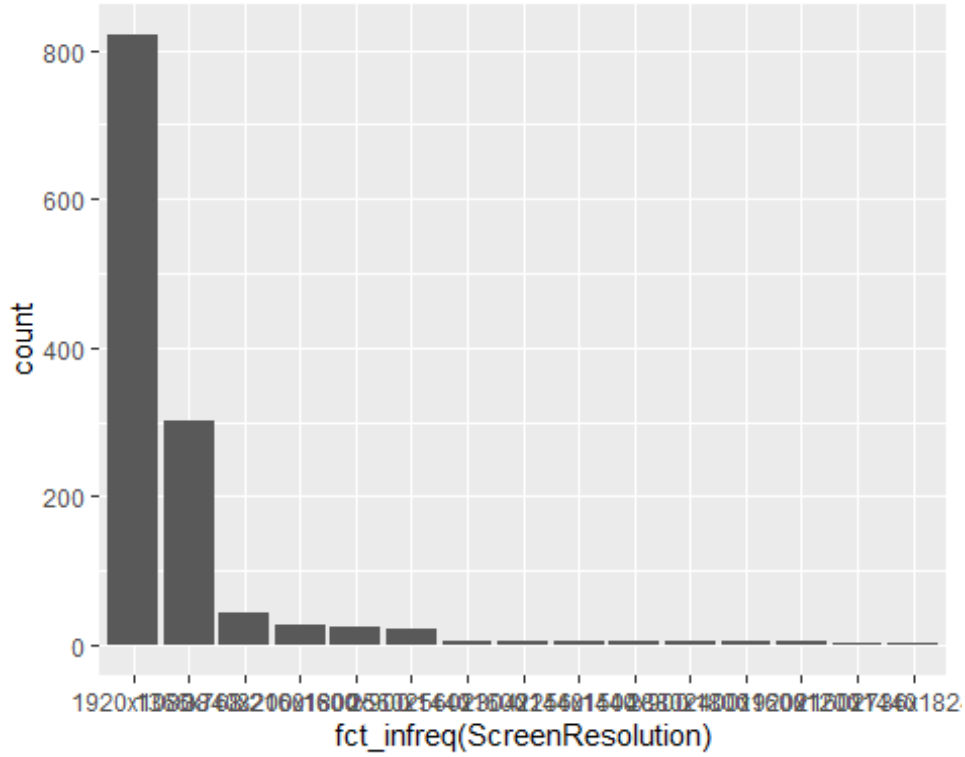
```
# We can change column types to help us for the analysis/cleaning process
# We should also select only the important columns to keep
laptop <- laptop.csv %>%
  select(-`Unnamed: 0`) %>%
  mutate(Company = as_factor(Company),
         TypeName = as_factor(TypeName),
         Inches = as.double(Inches),
         Ram = as_factor(Ram),
         OpSys = as_factor(OpSys))
```

```
## Warning: There was 1 warning in `mutate()`.
## ℹ In argument: `Inches = as.double(Inches)`.
## Caused by warning:
## ! NAs introduced by coercion
```

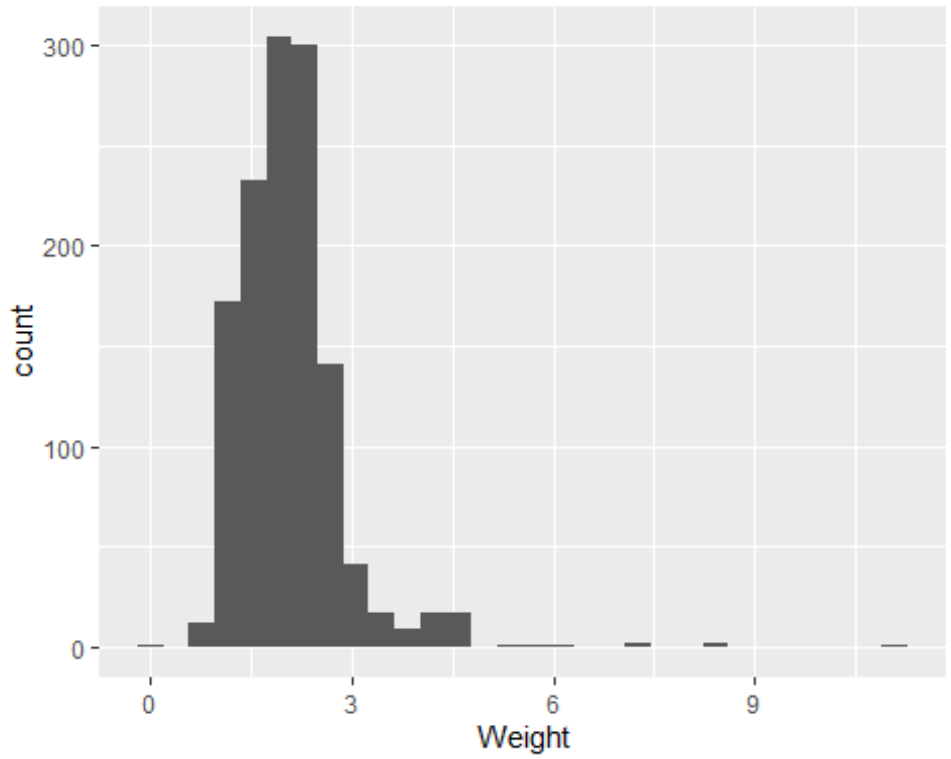
```
# Lets take a look at the different brands of laptops we have. The bar graph
shows us the most common brands are Lenovo, Dell, HP, Asus and Acer which all
have over 100 laptops, with the remaining brands all having less than 100
laptops.
ggplot(laptop, aes(fct_infreq(Company))) + geom_bar()
```




```
# We also see there are some NA values, lets remove them.

laptop <- drop_na(laptop)

# Now lets make some columns more usable, starting with the Screen Resolution
info
# We can separate the actual resolution from the description, and change to a
factor
laptop <- laptop %>%
  mutate(ScreenDescription = ScreenResolution,
    ScreenResolution = as_factor(str_split_i(ScreenResolution, " ", i=-1)))

# We can now see which size of screen is more common. There appears to be
some niche screen sizes, but the majority of laptops have a screen size of
1920x1080 or 1366x768
ggplot(laptop, aes(fct_infreq(ScreenResolution))) + geom_bar()
```




```
# Similarly, we can make the weight column more manageable
laptop <- laptop %>%
  mutate(Weight = round(as.numeric(str_sub(Weight, 1, -3)),2))

# We can now see what to expect for the weights of the laptops
ggplot(laptop, aes(Weight)) + geom_histogram()

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

## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_bin()`).
```

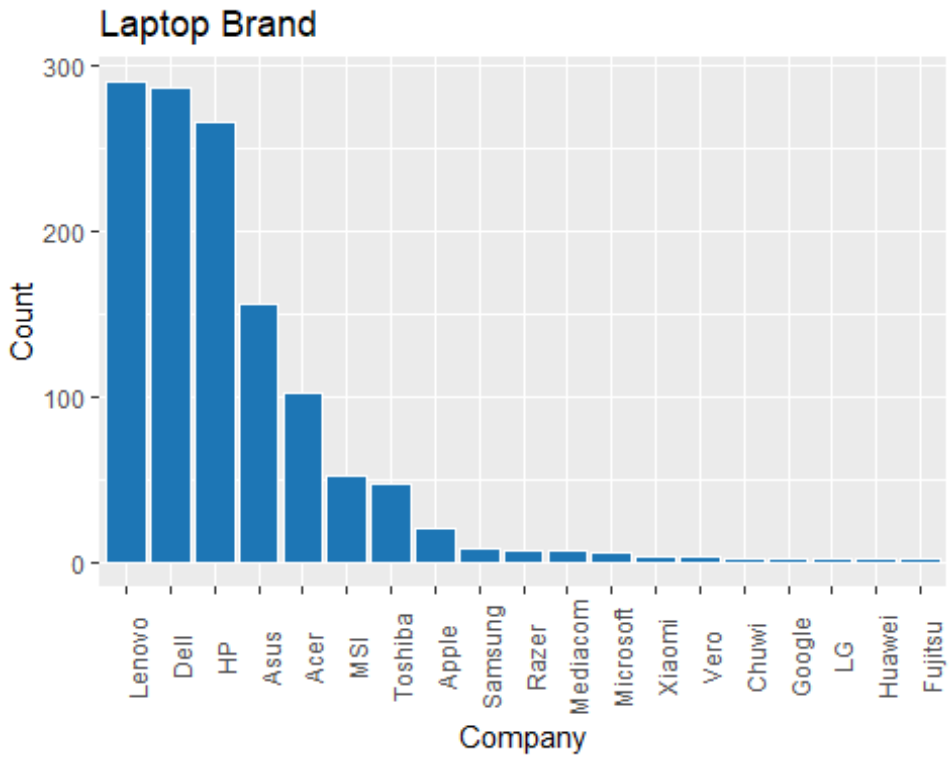



```
# The Price is currently in Indian Rupees, lets convert it to NZ Dollar,
using a rate of 100Rupees : 2.04NZD
laptop <- laptop %>%
  mutate(PriceNZD = Price * 0.0204)

# Finally, lets select all the columns that have the most useful information
laptop <- laptop %>%
  select(Company, TypeName, Inches, ScreenResolution, Cpu, Ram, Memory, Gpu,
OpSys, Weight, PriceNZD)
```

Visualisations

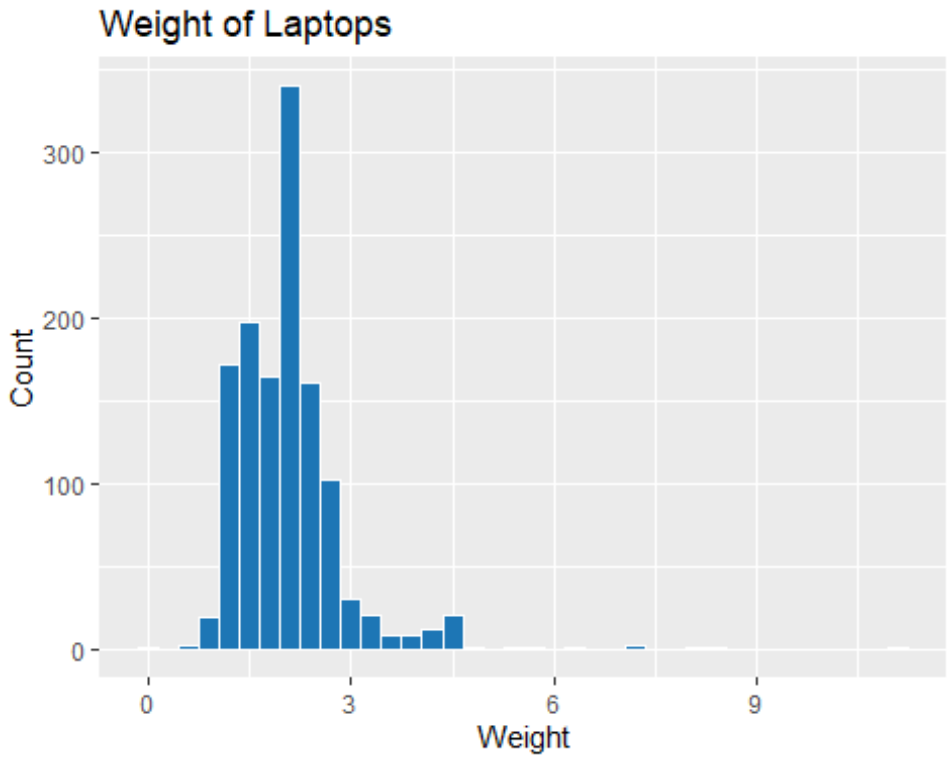
Now that we have done our initial analysis, lets upgrade the visualisations of what we found

```
# Firstly, we look at the different brands of laptops
ggplot(laptop, aes(fct_infreq(Company)),) + geom_bar(fill  = "#1D76B5", color
= "white") + theme(legend.position = "bottom", axis.text.x =
element_text(angle = 90)) + labs(x = "Company", y = "Count", title = "Laptop
Brand")
```

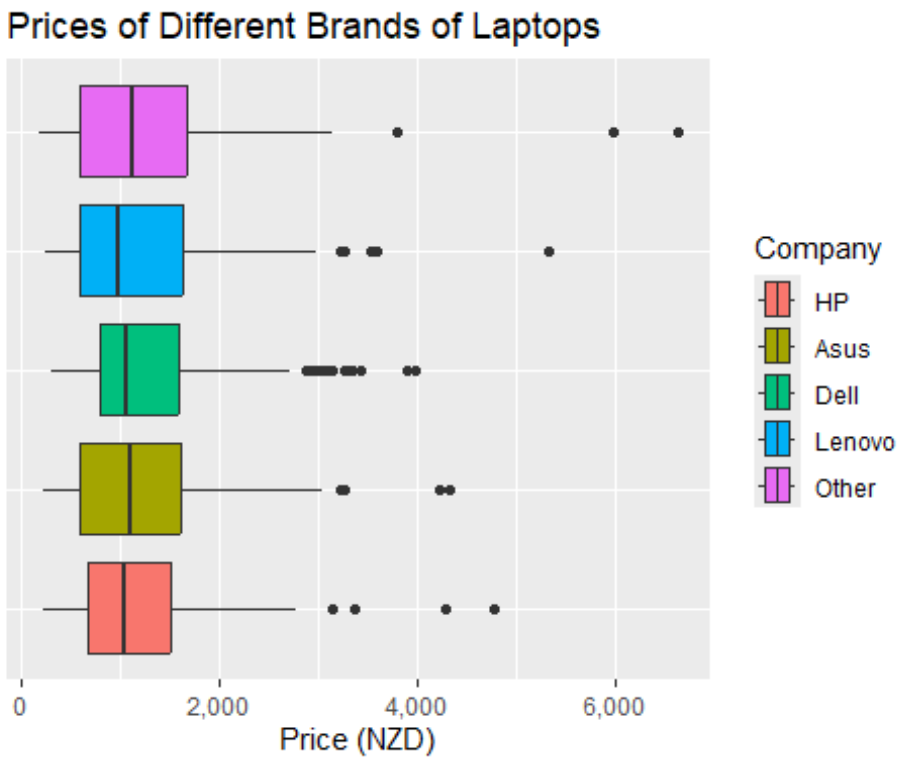



```
# Next, we look at the different weights of our laptops
ggplot(laptop, aes(Weight)) + geom_histogram(binwidth = 0.3, fill  =
"#1D76B5",
    color = "white") + labs(x = "Weight", y = "Count", title = "Weight of
Laptops")
```

```
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_bin()`).
```

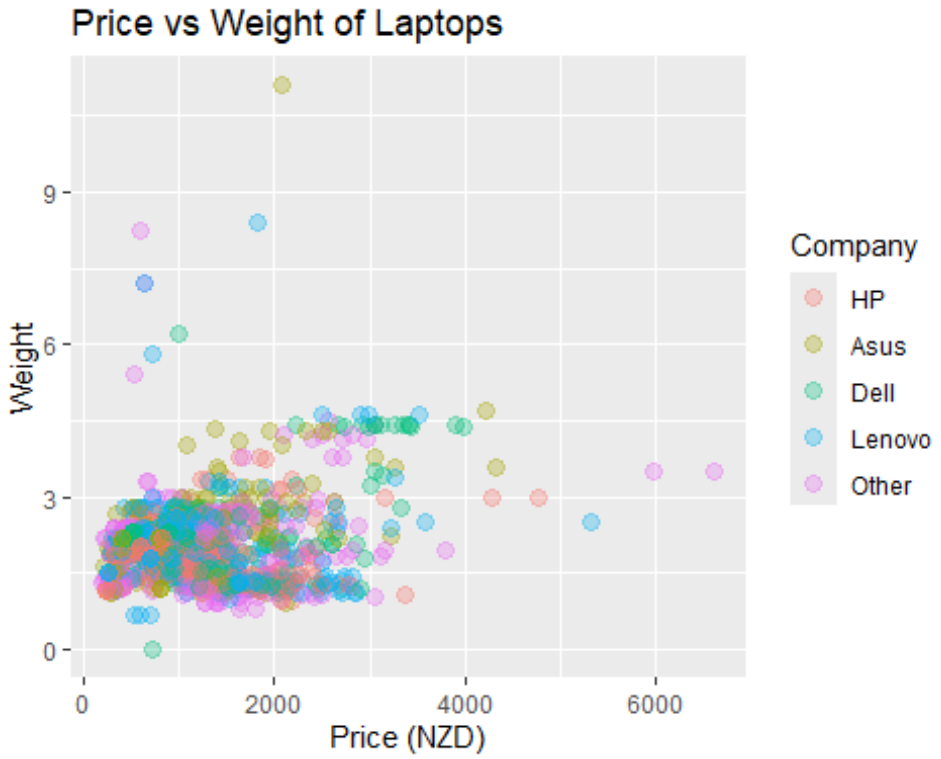



```
# Perhaps we are interested in the prices of laptops for each company. Lets
just focus on the 4 most common brands, and group the remaining ones. For
this, we must group the laptop brands with low frequencies.

laptop <- laptop %>%
  mutate(CompanyGrouped = fct_lump(Company, n=4))

ggplot(laptop, aes(x = PriceNZD, y = CompanyGrouped, fill = CompanyGrouped))
+ geom_boxplot() + scale_x_continuous(labels = scales::comma) + labs(x =
"Price (NZD)", y = "", title = "Prices of Different Brands of Laptops") +
guides(y = "none") +  scale_fill_discrete(name = "Company")
```




```
# We can also find out which laptop is the cheapest or most expensive

laptop[which.min(laptop$PriceNZD),]

## # A tibble: 1 × 12
##   Company TypeName Inches ScreenResolution Cpu   Ram   Memory Gpu   OpSys
Weight
##   <fct>   <fct>     <dbl> <fct>            <chr> <fct> <chr>  <chr> <fct>
<dbl>
## 1 Acer    Netbook    11.6 1366x768         Inte… 2GB   32GB … Inte… Chro…
1.3
## # ℹ 2 more variables: PriceNZD <dbl>, CompanyGrouped <fct>

laptop[which.max(laptop$PriceNZD),]

## # A tibble: 1 × 12
##   Company TypeName Inches ScreenResolution Cpu   Ram   Memory Gpu   OpSys
Weight
##   <fct>   <fct>     <dbl> <fct>            <chr> <fct> <chr>  <chr> <fct>
<dbl>
## 1 Razer   Gaming     17.3 3840x2160        Inte… 32GB  1TB S… Nvid… Wind…
3.49
## # ℹ 2 more variables: PriceNZD <dbl>, CompanyGrouped <fct>

# Are larger laptops typically more expensive than smaller laptops? We can
also compare different brands too
ggplot(laptop, aes(y = Weight, x = PriceNZD, col = CompanyGrouped)) +
```

```
geom_point(alpha = 0.3, size = 3) + labs(title = "Price vs Weight of
Laptops", x = "Price (NZD)") + guides(color = guide_legend(title =
"Company"))

## Warning: Removed 1 row containing missing values or values outside the
scale range
## (`geom_point()`).
```



Doing this type of analysis can help you find trends, or key points about the data which may bring valuable direction for making positive changes!