HW3

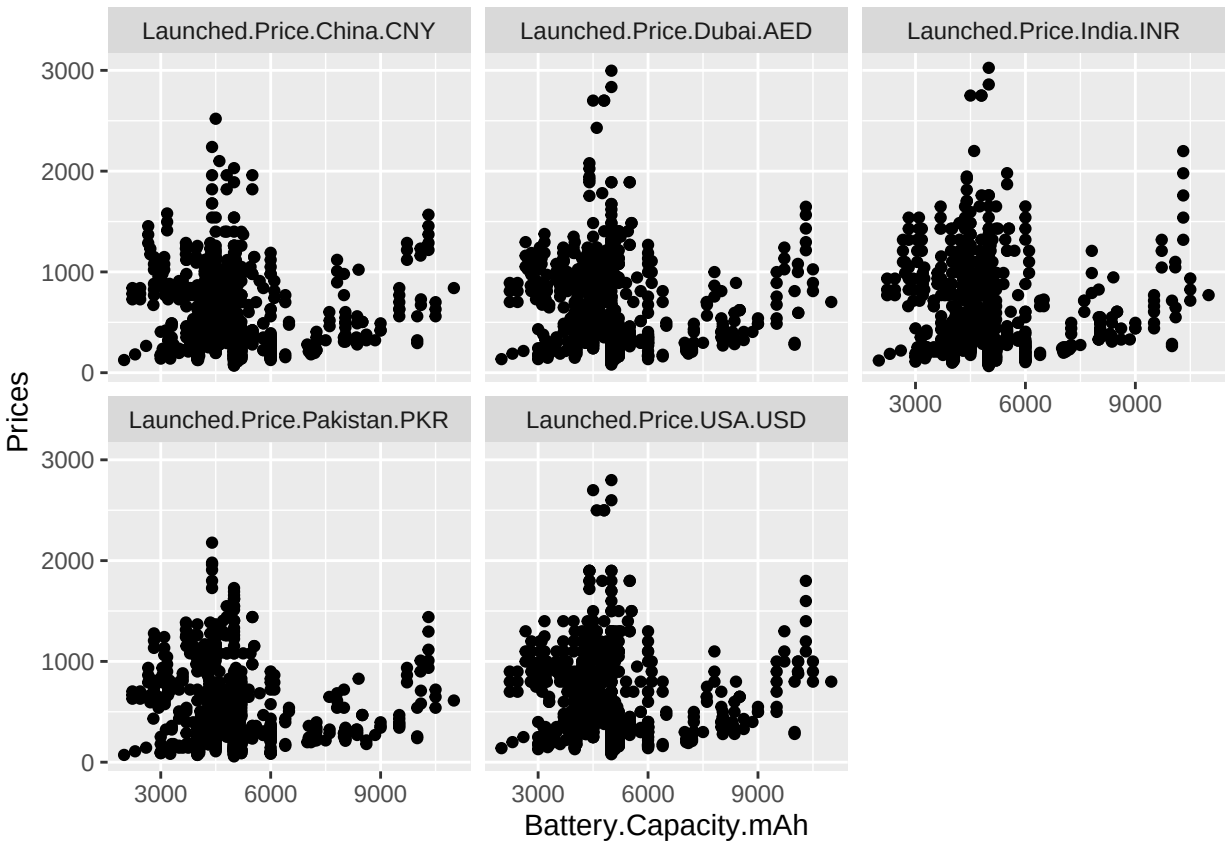
2025-03-15

Part 1

```
head(mobiles_dataset)
```

```
## # A tibble: 6 x 15
##   Company.Name Model.Name Mobile.Weight RAM   Front.Camera Back.Camera Processor
##   <chr>        <chr>      <chr>         <chr> <chr>        <chr>       <chr>
## 1 Apple        iPhone 16~ 174g          6GB   12MP         48MP        A17 Bion~
## 2 Apple        iPhone 16~ 174g          6GB   12MP         48MP        A17 Bion~
## 3 Apple        iPhone 16~ 174g          6GB   12MP         48MP        A17 Bion~
## 4 Apple        iPhone 16~ 203g          6GB   12MP         48MP        A17 Bion~
## 5 Apple        iPhone 16~ 203g          6GB   12MP         48MP        A17 Bion~
## 6 Apple        iPhone 16~ 203g          6GB   12MP         48MP        A17 Bion~
## # i 8 more variables: Battery.Capacity.mAh <dbl>, Screen.Size.inches <dbl>,
## #   Launched.Price.Pakistan.PKR <dbl>, Launched.Price.India.INR <dbl>,
## #   Launched.Price.China.CNY <dbl>, Launched.Price.USA.USD <dbl>,
## #   Launched.Price.Dubai.AED <dbl>, Launched.Year <dbl>
```

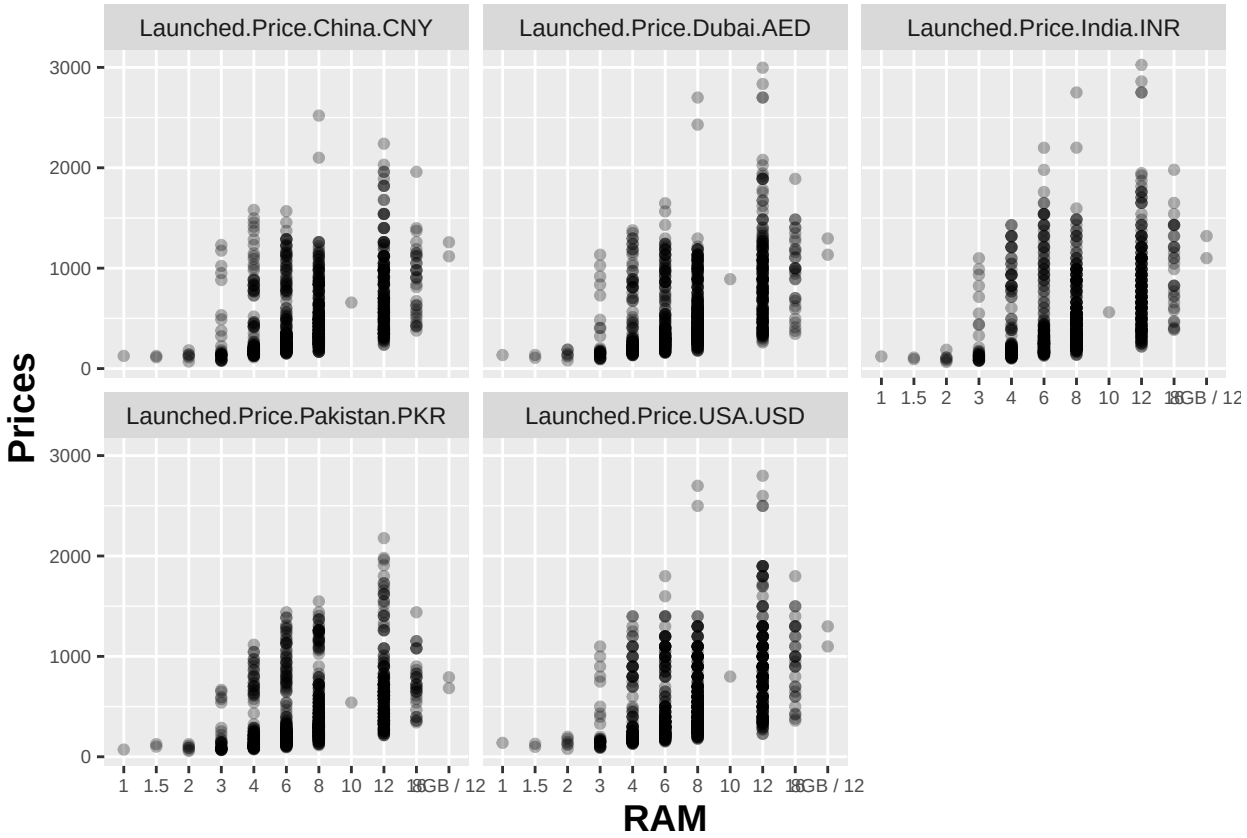
1.



No Correlation exists between the launch price and Battery Capacity

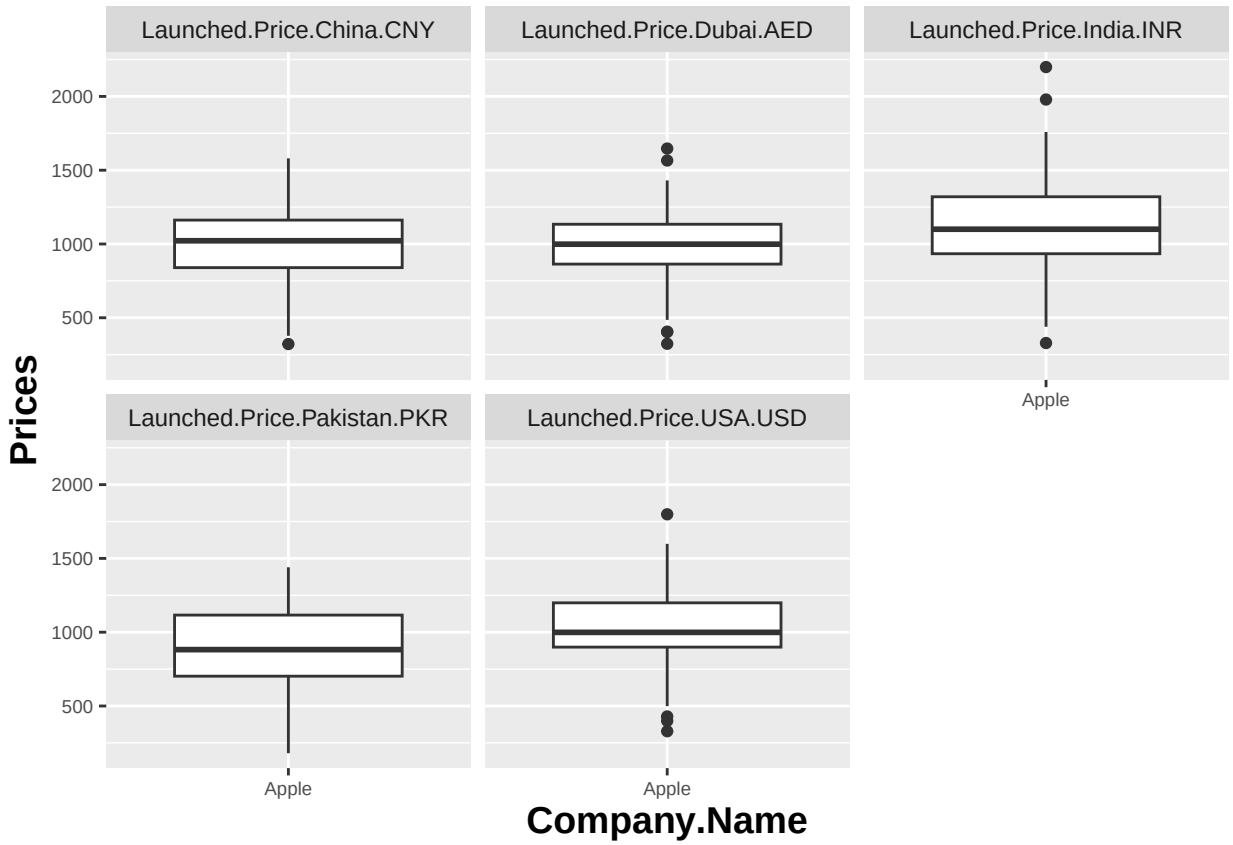
2.

```
mobile_dataset_longer$RAM = substr(mobile_dataset_longer$RAM,1,nchar(mobile_dataset_longer$RAM)-2)

ggplot(mobile_dataset_longer) +
  geom_point(aes(x = RAM, y = Prices), alpha = 0.3) +
  facet_wrap(~Countries) +
  theme(axis.text=element_text(size=7), axis.title=element_text(size=14,face="bold")) +
  scale_x_discrete(limits = c(1, 1.5, 2, 3, 4, 6, 8, 10, 12, 16, '8GB / 12'))
```



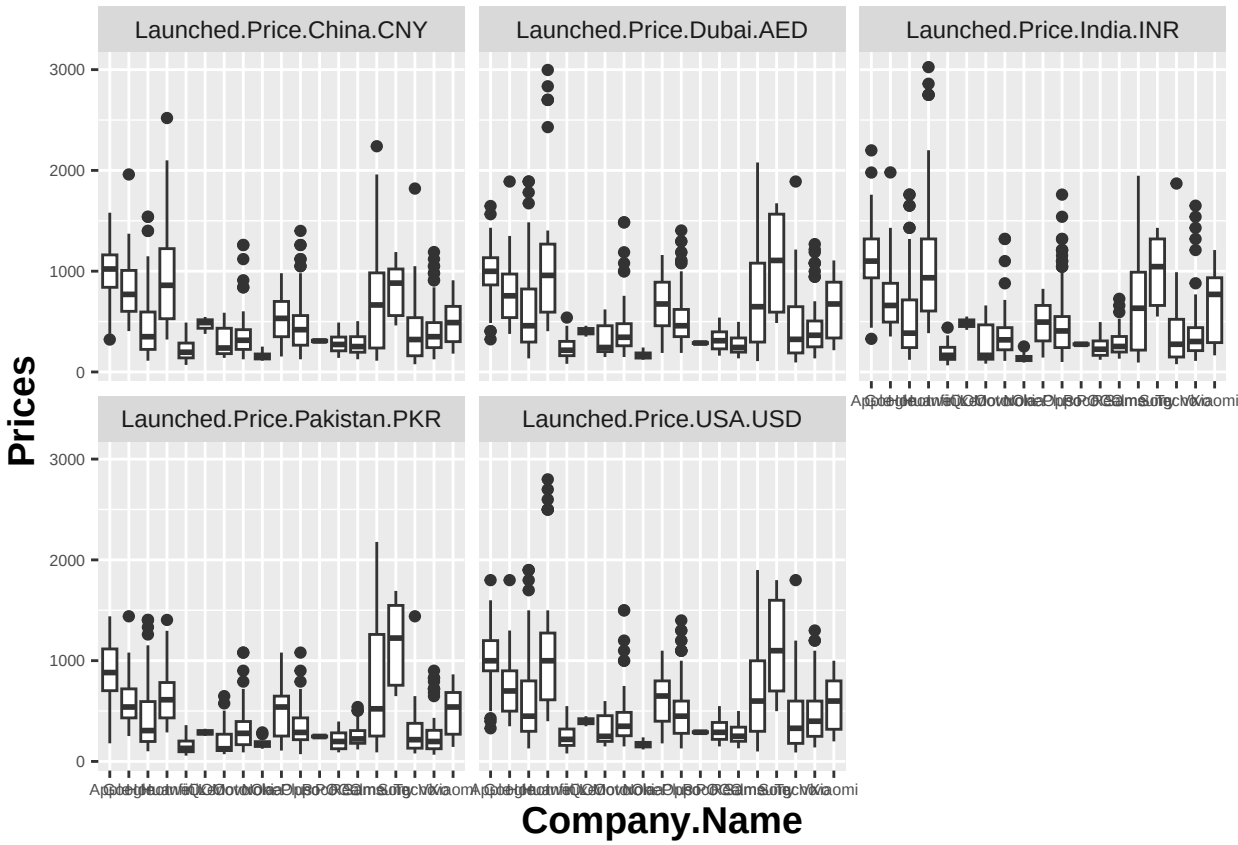
There is a slight positive correlation between RAM and Price as we can see from the graph.

```
mobile_dataset_longer_apple <- mobile_dataset_longer%>%
  filter(Company.Name == 'Apple')
ggplot(mobile_dataset_longer_apple) +
  geom_boxplot(aes(x = Company.Name, y = Prices)) +
  facet_wrap(~Countries) + theme(axis.text=element_text(size=7), axis.title=element_text(size=14,face="
```



We can see that Apple hase a bigger range in India and a smaller range in Pakistan and China. India has the highest markup, as both the maximum price and the highest median price belongs to Indian market.

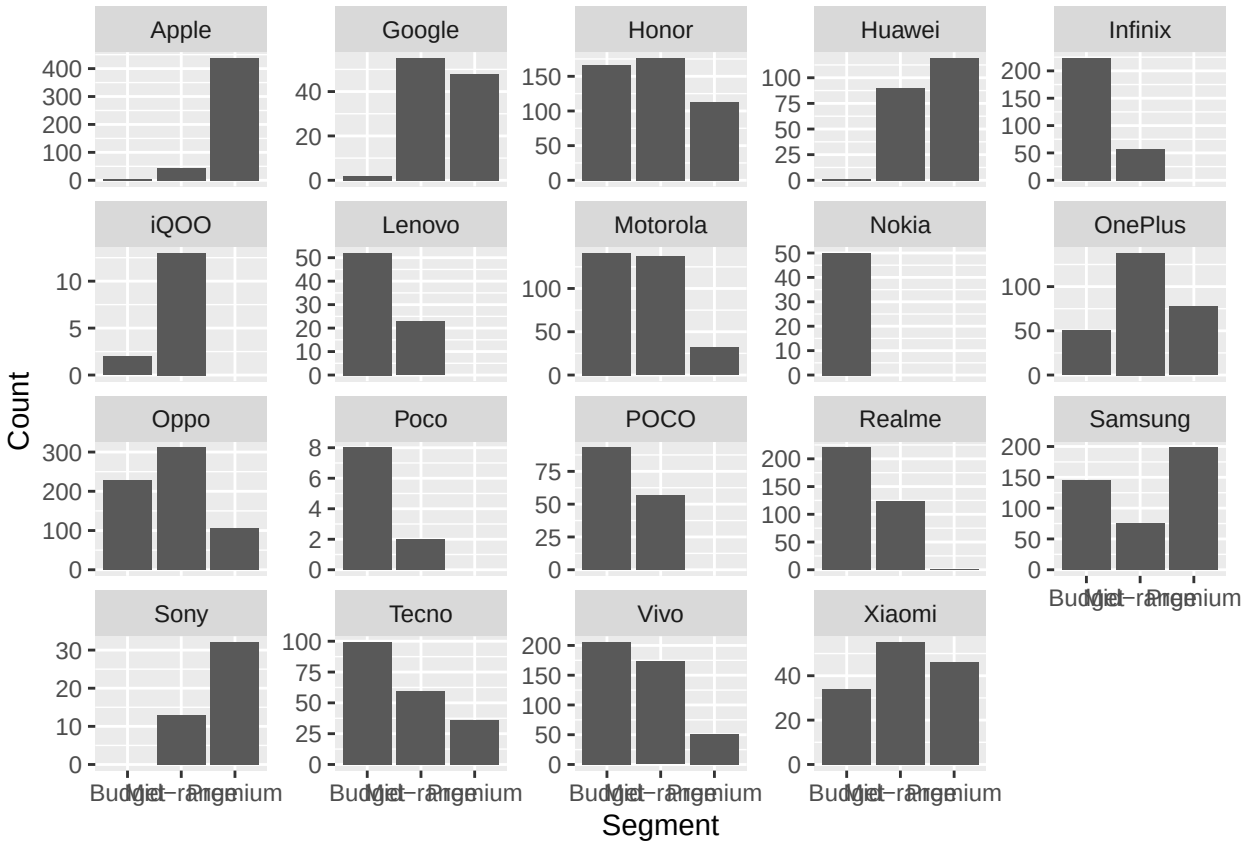
```
ggplot(mobile_dataset_longer) +
  geom_boxplot(aes(x = Company.Name, y = Prices)) +
  facet_wrap(~Countries) + theme(axis.text=element_text(size=7), axis.title=element_text(size=14,face="
  theme(axis.text=element_text(size=6))
```



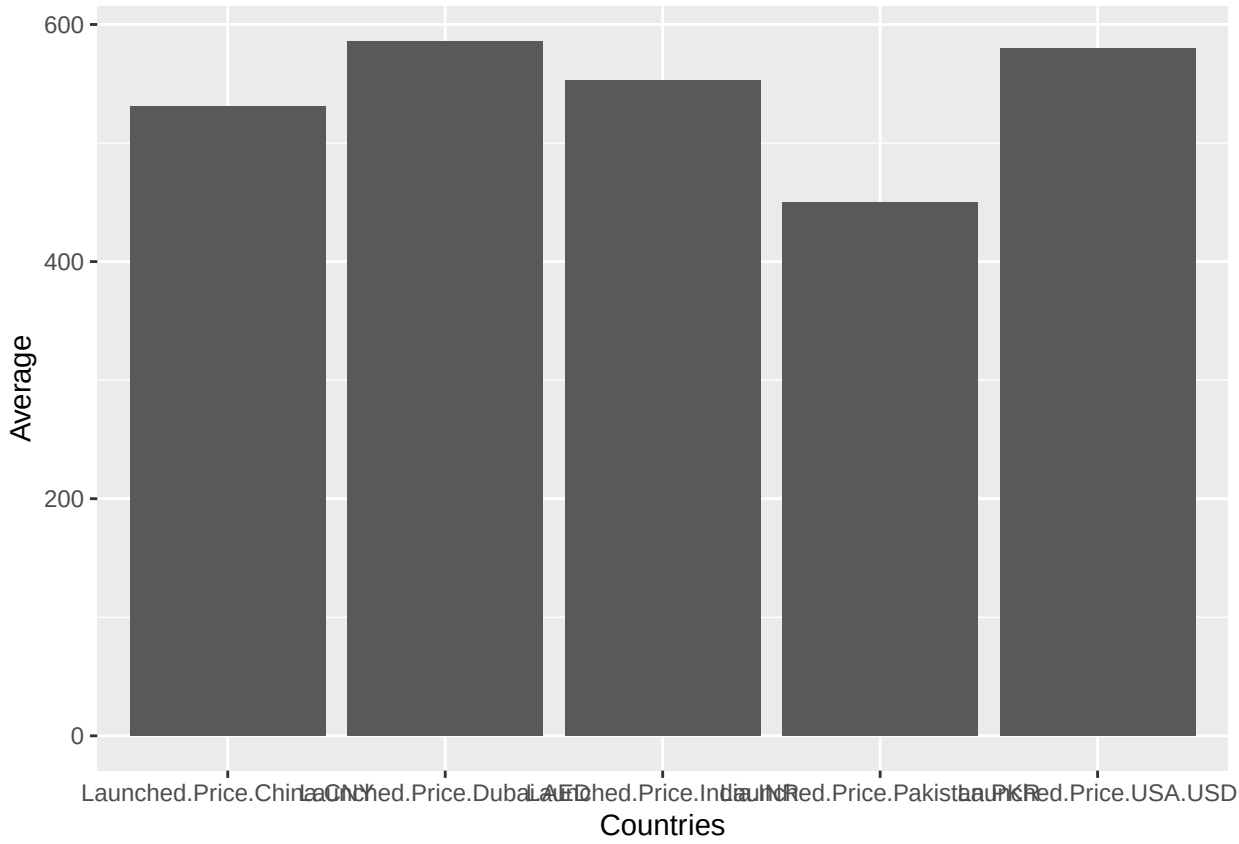
If we ignore Nokia and IQOO, as we have little data regarding them, the most stable price among all countries is supposedly Infinix and Poco

```
mobile_dataset_summary <- mobile_dataset_longer %>%
  mutate(Segment = case_when(
    Prices < 300 ~ "Budget",
    Prices >= 300 & Prices <= 700 ~ "Mid-range",
    Prices > 700 ~ "Premium"
  )) %>%
  group_by(Segment, Company.Name) %>%
  summarise(Count = n(), .groups = 'drop') %>%
  complete(Segment, Company.Name, fill = list(count = 0))
#Found this in the internet. Didn't want to ignore 0's
ggplot(mobile_dataset_summary) + geom_bar(aes(x=Segment, y=Count), stat = 'identity') + facet_wrap(~Comp
```

```
## Warning: Removed 8 rows containing missing values or values outside the scale range
## (`geom_bar()`).
```




```
# For Ex. 1
mobile_dataset_region <- mobile_dataset_longer %>%
  group_by(Countries) %>%
  summarise(Average = mean(Prices)) %>%
  arrange(Average)
#For Ex. 2
ggplot(mobile_dataset_region) + geom_bar(aes(x=Countries, y=Average), stat='identity')
```



Pakistan has the cheapest Phones

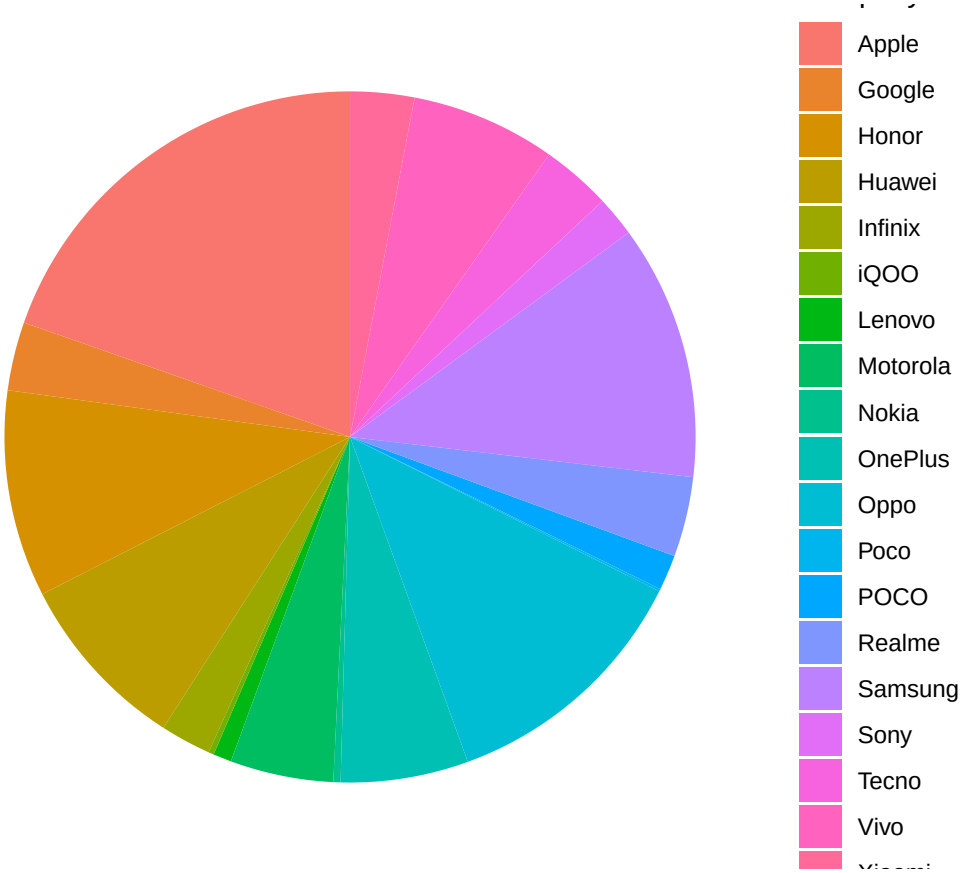
```
mobile_dataset_avg <- mobile_dataset_longer %>%
  group_by(Countries, Company.Name) %>%
  summarise(Average = mean(Prices), .groups = 'drop')

mobile_dataset_diff <- mobile_dataset_avg %>%
  group_by(Company.Name) %>%
  summarise(Difference = mean(Average) - min(Average)) %>%
  arrange(desc(Difference))
mobile_dataset_diff
```

```
## # A tibble: 19 x 2
##    Company.Name Difference
##    <chr>             <dbl>
##  1 Huawei            338.
##  2 Sony              214.
##  3 Google            142.
##  4 Vivo              133.
##  5 Oppo              127.
##  6 iQOO              121.
##  7 Tecno             120.
##  8 Apple             119.
##  9 Honor              98.4
## 10 OnePlus            77.3
## 11 Xiaomi             70.1
## 12 POCO               65.7
## 13 Motorola           59.4
```

```
## 14 Lenovo              59.0
## 15 Infinix             53.6
## 16 Samsung             45.8
## 17 Poco                34.1
## 18 Realme              21.7
## 19 Nokia               21.3
```

Ex.2 2)

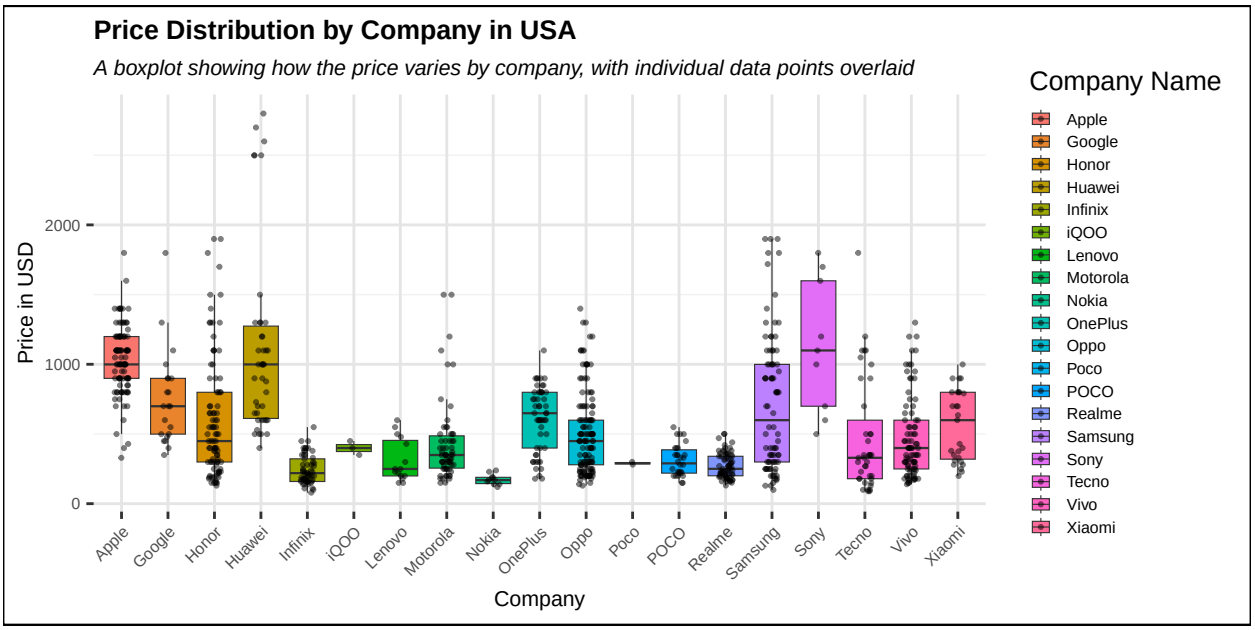
```
mobile_dataset_share <- mobile_dataset_longer %>%
  group_by(Company.Name) %>%
  summarise(Share = sum(Prices))

ggplot(mobile_dataset_share, aes(x = "", y = Share, fill = Company.Name)) + geom_bar(stat = "identity",
  coord_polar(theta = "y")  +
  theme_void()
```



Ex. 3.

```
ggplot(mobiles_dataset, aes(x=Company.Name, y=Launched.Price.USA.USD, fill = Company.Name)) + geom_boxpl
  labs(
    title = "Price Distribution by Company in USA",
    subtitle = "A boxplot showing how the price varies by company, with individual data points overlaid
    x = "Company",
    y = "Price in USD",
    fill = "Company Name") +
  theme(
  axis.text.x = element_text(angle = 45, hjust = 1, size=6),
```

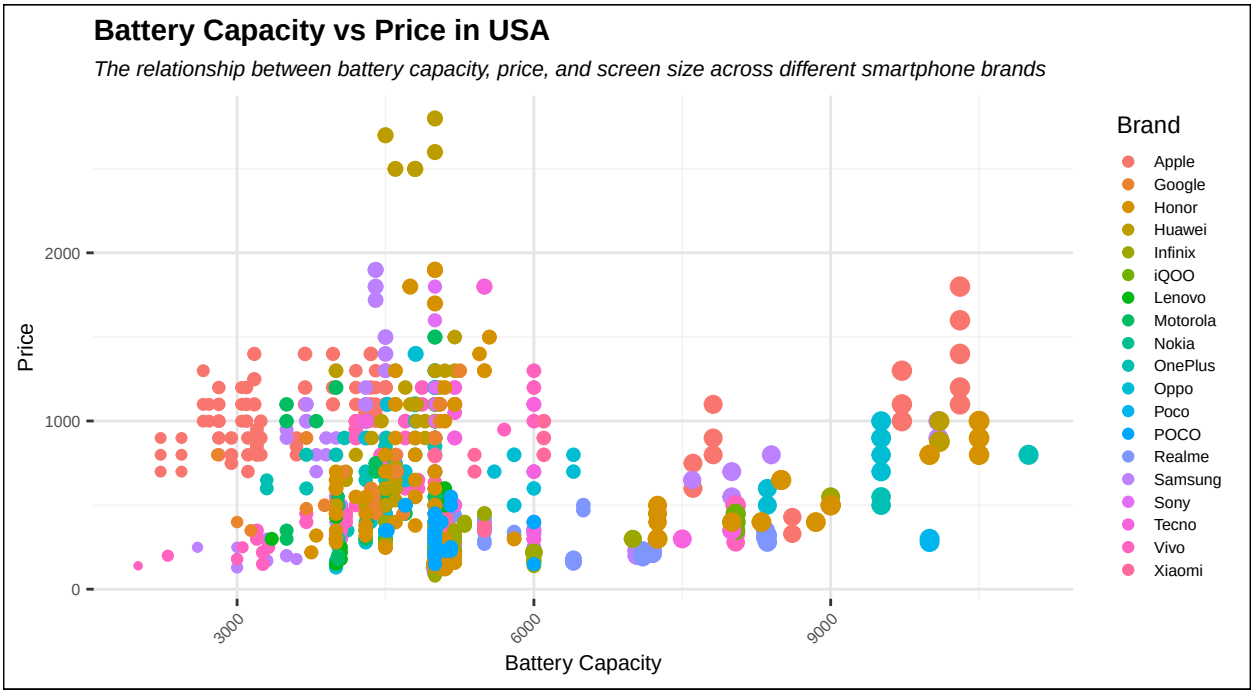
```
  axis.text.y = element_text(size=6),
  plot.background = element_rect(fill = "white", color = 'black'),
  panel.background = element_rect(fill = "white"),
  panel.grid.major = element_line(color = "gray90"),
  panel.grid.minor = element_line(color = "gray95"),
  plot.title = element_text(face = "bold", size = 10),
  plot.subtitle = element_text(size = 8, face='italic'),
  axis.title = element_text(size = 8),
  axis.text = element_text(size = 8),
  legend.title = element_text(size = 10),
  legend.text = element_text(size = 6),
  legend.key.size = unit(0.3, "cm"),
  ) + coord_fixed(0.003)
```



2.

```
ggplot(mobiles_dataset, aes(x=Battery.Capacity.mAh, y=Launched.Price.USA.USD, color = Company.Name, size
  labs(
    title = "Battery Capacity vs Price in USA",
    subtitle = "The relationship between battery capacity, price, and screen size across different smar
    x = "Battery Capacity",
    y = "Price",
    color = "Brand")+
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, size=6),
    axis.text.y = element_text(size=6),
    plot.background = element_rect(fill = "white", color = 'black'),
    panel.background = element_rect(fill = "white"),
    panel.grid.major = element_line(color = "gray90"),
    panel.grid.minor = element_line(color = "gray95"),
    plot.title = element_text(face = "bold", size = 11),
    plot.subtitle = element_text(size = 8, face='italic'),
    axis.title = element_text(size = 8),
    axis.text = element_text(size = 8),
```

```
    legend.title = element_text(size = 9),
    legend.text = element_text(size = 6),
    legend.key.size = unit(0.3, "cm"),
  ) + guides(size='none') +
    coord_fixed(1.7)+
    scale_size_continuous(range = c(1, 3))
```

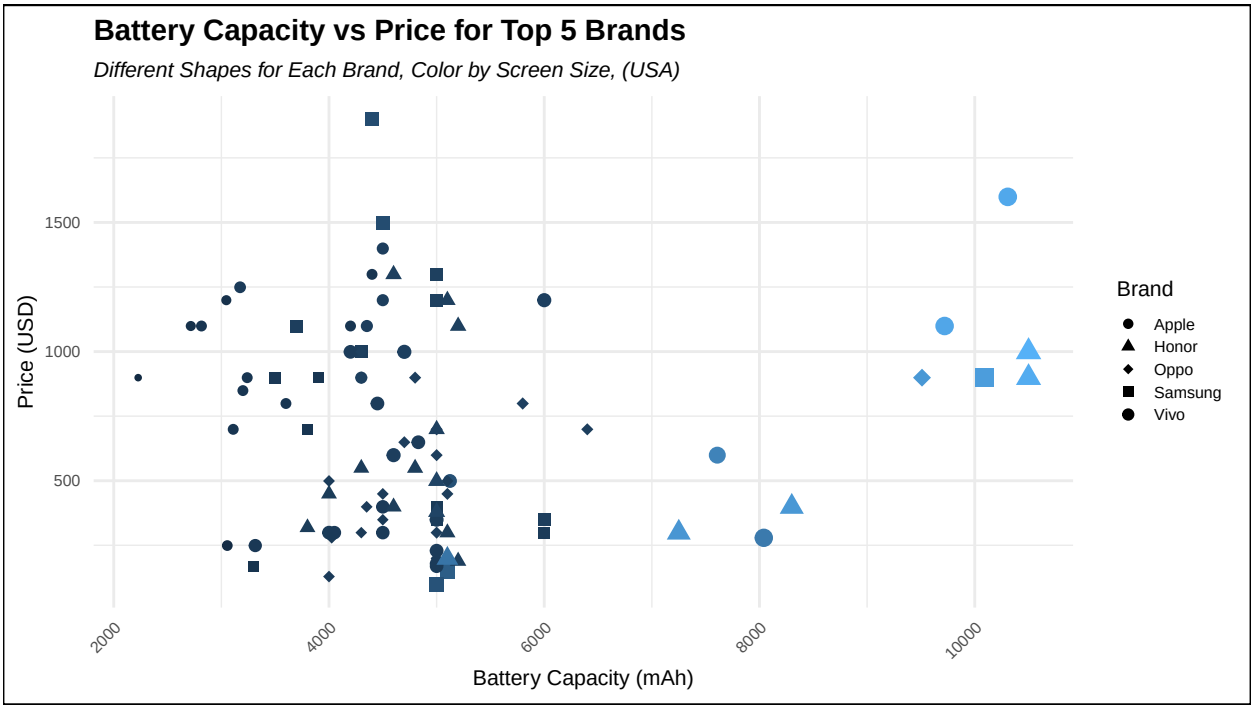


3.

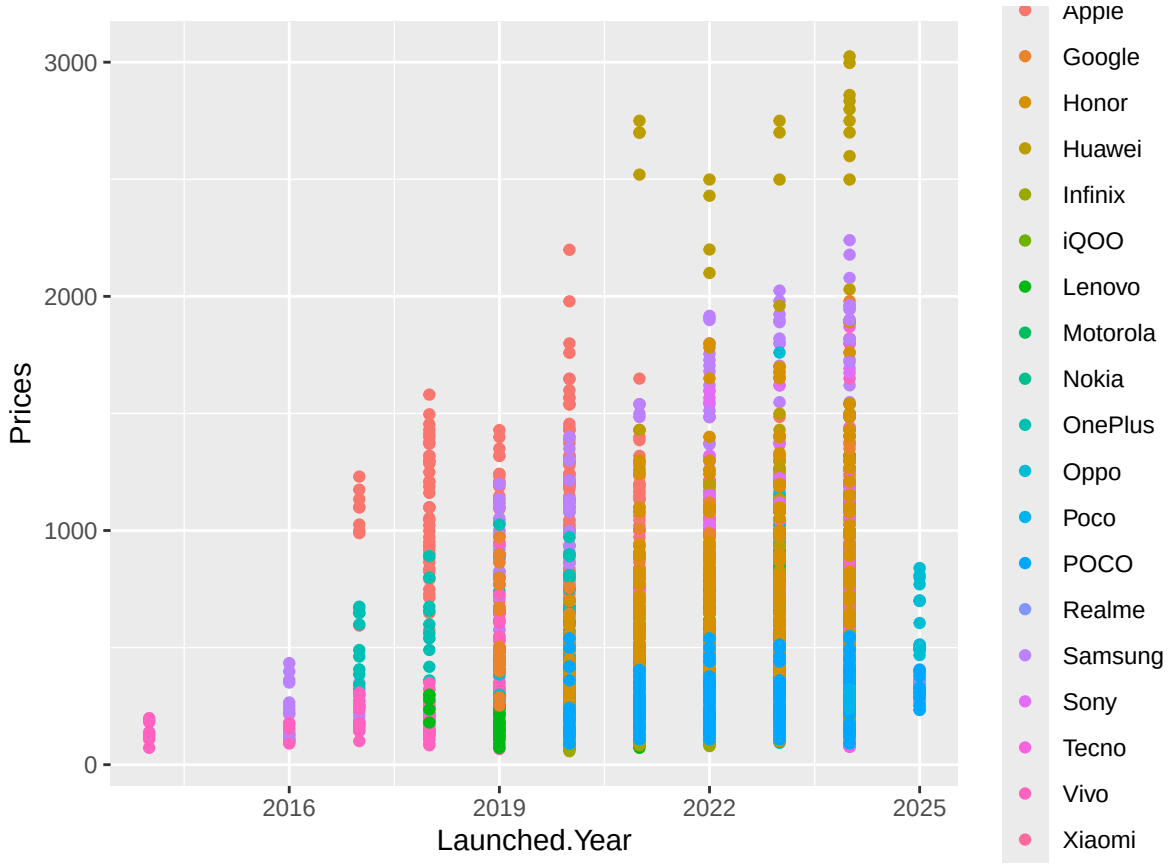
```
mobiles_dataset_top <- mobiles_dataset %>%
  filter(Company.Name == c('Apple', 'Honor', 'Oppo', 'Samsung', 'Vivo'))
mobiles_dataset_top
```

```
## # A tibble: 99 x 15
##    Company.Name Model.Name          Mobile.Weight RAM   Front.Camera Back.Camera
##    <chr>        <chr>               <chr>         <chr> <chr>        <chr>
##  1 Apple        iPhone 16 128GB     174g          6GB   12MP         48MP
##  2 Apple        iPhone 16 Plus 512~ 203g          6GB   12MP         48MP
##  3 Apple        iPhone 16 Pro Max ~ 221g          8GB   12MP / 4K    48MP + 12MP
##  4 Apple        iPhone 15 Plus 128~ 203g          6GB   12MP         48MP
##  5 Apple        iPhone 15 Pro 512GB 206g          8GB   12MP / 4K    48MP + 12MP
##  6 Apple        iPhone 14 256GB     172g          6GB   12MP         12MP + 12MP
##  7 Apple        iPhone 14 Pro 128GB 206g          6GB   12MP / 4K    48MP + 12MP
##  8 Apple        iPhone 14 Pro Max ~ 221g          8GB   12MP / 4K    48MP + 12MP
##  9 Apple        iPhone 13 256GB     174g          4GB   12MP         12MP + 12MP
## 10 Apple        iPhone 13 Pro Max ~ 238g          6GB   12MP / 4K    12MP + 12M~
## # i 89 more rows
## # i 9 more variables: Processor <chr>, Battery.Capacity.mAh <dbl>,
## #   Screen.Size.inches <dbl>, Launched.Price.Pakistan.PKR <dbl>,
## #   Launched.Price.India.INR <dbl>, Launched.Price.China.CNY <dbl>,
## #   Launched.Price.USA.USD <dbl>, Launched.Price.Dubai.AED <dbl>,
## #   Launched.Year <dbl>
```

```r
ggplot(mobiles_dataset_top, aes(x=Battery.Capacity.mAh, y=Launched.Price.USA.USD, shape = Company.Name,
  labs(
    title = "Battery Capacity vs Price for Top 5 Brands",
    subtitle = "Different Shapes for Each Brand, Color by Screen Size, (USA)",
    x = "Battery Capacity (mAh)",
    y = "Price (USD)",
    shape = "Brand")+
  theme_minimal() +
    theme(
    axis.text.x = element_text(angle = 45, hjust = 1, size=6),
    axis.text.y = element_text(size=6),
    plot.background = element_rect(fill = "white", color = 'black'),
    plot.title = element_text(face = "bold", size = 11),
    plot.subtitle = element_text(size = 8, face='italic'),
    axis.title = element_text(size = 8),
    axis.text = element_text(size = 8),
    legend.title = element_text(size = 8),
    legend.text = element_text(size = 6),
    legend.key.size = unit(0.3, "cm"),
  ) + guides(size='none', color='none') +
    coord_fixed(2.4) +
    scale_size_continuous(range = c(1, 3)) +
    scale_shape_manual(values = c(16, 17, 18, 15, 19))
```




```r
ggplot(mobile_dataset_longer, aes(x=Launched.Year, y=Prices, color = Company.Name)) + geom_point()
```

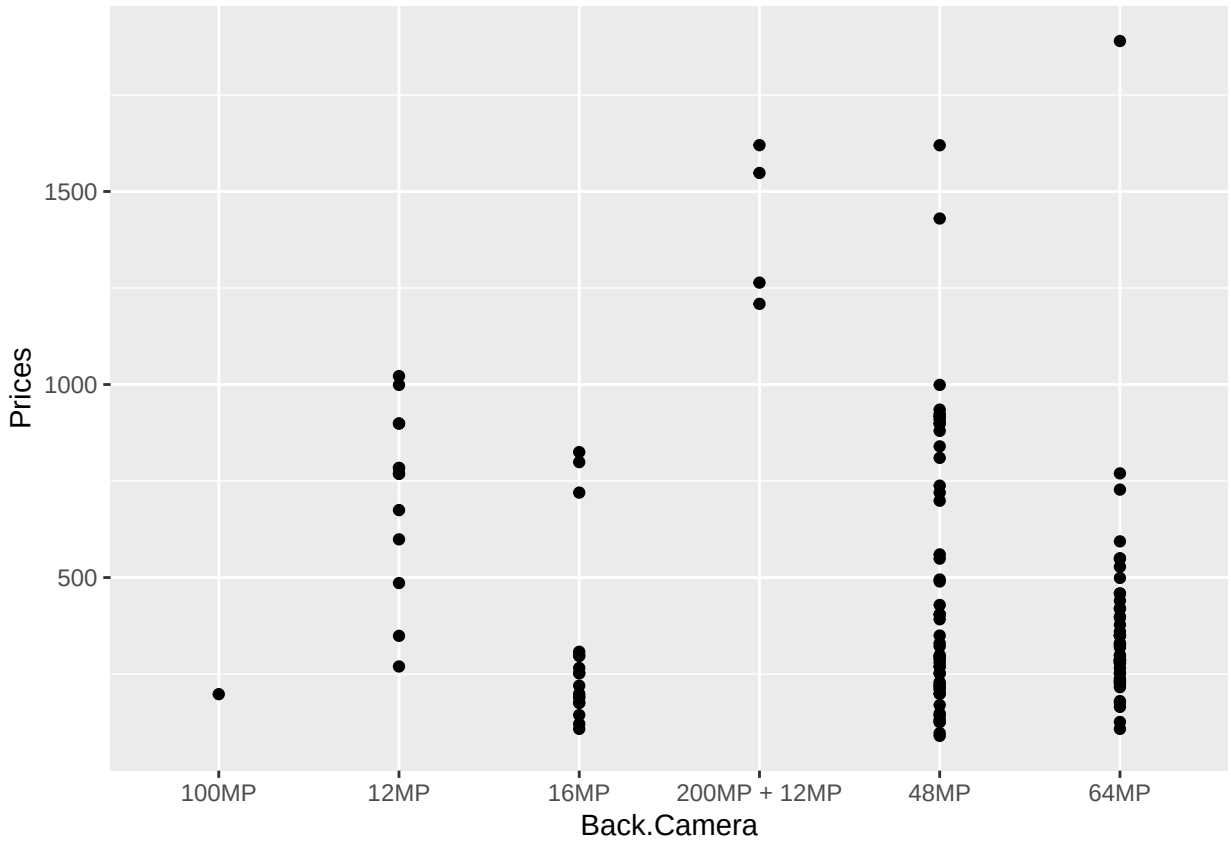


We have a clear positive correlation of Launched year and Prices, meaning the new the phone, the more expensive it is; Except 2025 where I suppose the luxurious brands haven't yet released any new phones, hence we can only see the cheaper phones elling companies like POCO (and Poco :)) and Oppo.

```
mobiles_dataset_backcamera <- mobile_dataset_longer %>%
  filter(Back.Camera == c('48MP', '12MP', '16MP', '64MP', '100MP', "200MP + 12MP"))
```

```
## Warning: There was 1 warning in `filter()`.
## i In argument: `Back.Camera == c("48MP", "12MP", "16MP", "64MP", "100MP",
##   "200MP + 12MP")`.
## Caused by warning in `Back.Camera == c("48MP", "12MP", "16MP", "64MP", "100MP", "200MP + 12MP")`:
## ! longer object length is not a multiple of shorter object length
```

```
ggplot(mobiles_dataset_backcamera, aes(x=Back.Camera, y=Prices)) + geom_point()
```



Weirdly there is a 64MP phone which is 7 times more expensive than the 100MP phone, hence there is no clear correlation there.