Data Translation Challenge

## Importing Libraries and Data

This section will import the necessary libraries used in the following code, as well as importing the data sets for this project.

#importing libraries  
library(rio)  
library(tidyverse)  
library(vtable)  
library(ggplot2)  
library(lubridate)  
library(forcats)  
library(scales)

#importing data  
sales <- import('sales\_data.Rdata')  
zip\_info <- import('zip\_info.csv')

## Cleaning Data

This section will provide an overview of the data, as well as clean certain variables that should be presented differently and aid with future calculations.

vtable(sales, lush= TRUE)

sales

| Name | Class | Values | Missing | Summary |
| --- | --- | --- | --- | --- |
| Product | character | '20in Monitor' '27in 4K Gaming Monitor' '27in FHD Monitor' '34in Ultrawide Monitor' 'AA Batteries (4-pack)' and 14 more | 0 | nuniq: 19 |
| Quantity | character | '1' '2' '3' '4' '5' and 4 more | 0 | nuniq: 9 |
| PriceEach | character | '109.99' '11.95' '11.99' '14.95' '149.99' and 18 more | 0 | nuniq: 23 |
| DateTime | POSIXct | Time: 2019-01-01 03:07:00 to 2020-01-01 05:13:00 | 0 | median: 2019-07-17 20:40:30, nuniq: 142395 |
| Date | Date | Time: 2019-01-01 to 2020-01-01 | 0 | median: 2019-07-17, nuniq: 366 |
| ZIP | character | '02215' '04101' '10001' '30301' '73301' and 5 more | 0 | nuniq: 10 |
| State | character | 'CA' 'GA' 'MA' 'ME' 'NY' and 3 more | 0 | nuniq: 8 |
| City | character | 'Atlanta' 'Austin' 'Boston' 'Dallas' 'Los Angeles' and 4 more | 0 | nuniq: 9 |

From looking at the vtable, certain variables are presented as characters/ strings when they should really be numerical/ values. Therefore, I will clean the data set so that Quantity, PriceEach, and ZIP are all numerical values.

sales <- sales %>%  
 mutate(Quantity = as.numeric(Quantity)) %>%  
 mutate(PriceEach = as.numeric(PriceEach)) %>%  
 mutate(ZIP = as.numeric(ZIP))

## Target Audience

I will create a story and a set of data visualizations, targeting the managers of Amazon’s technology products division. This data seems to represent a select 19 technology products ( as seen in the vtable), of which this data can provide an idea of how these products are selling and earning revenue. The goal will be to convince the managers that certain technology product categories, as well as specific products, are worth pursing and pushing to be sold.

## Data Exploration

#QuantitySold - represents the 7 broad categories of technology products, and demonstrates which category sells the most to these select urban cities  
QuantitySold <- sales %>%  
 select(Product, Quantity) %>%  
 group\_by(Product) %>%  
 summarize(NumberOfSales = sum(Quantity)) %>%  
 arrange(-NumberOfSales) %>%  
 mutate(Product = recode(Product,  
 "AAA Batteries (4-pack)" = "Batteries",  
 "AA Batteries (4-pack)" = "Batteries",  
 'USB-C Charging Cable' = 'Charging Cable',  
 'Lightning Charging Cable' = 'Charging Cable',  
 'Wired Headphones' = 'Headphones',  
 'Apple Airpods Headphones' = 'Headphones',  
 'Bose SoundSport Headphones' = 'Headphones',  
 '27in FHD Monitor' = 'Monitor',  
 '27in 4K Gaming Monitor' = 'Monitor',  
 '34in Ultrawide Monitor' = 'Monitor',  
 '20in Monitor' = 'Monitor',  
 "Google Phone" = "Phone",  
 'iPhone' = 'Phone',  
 'Vareebadd Phone' = 'Phone',  
 'Macbook Pro Laptop' = 'Laptop',  
 'ThinkPad Laptop' = 'Laptop',  
 .default = "Other Products")) %>%  
 group\_by(Product) %>%  
 summarize(Sales = sum(NumberOfSales))

#Prices - represents the 7 broad categories of technology products, and demonstrates the average prices of the products within each cateogry  
Prices <- sales %>%  
 select(Product, PriceEach) %>%  
 group\_by(Product) %>%  
 summarize(AvgPrice = mean(PriceEach)) %>%  
 mutate(Product = recode(Product,  
 "AAA Batteries (4-pack)" = "Batteries",  
 "AA Batteries (4-pack)" = "Batteries",  
 'USB-C Charging Cable' = 'Charging Cable',  
 'Lightning Charging Cable' = 'Charging Cable',  
 'Wired Headphones' = 'Headphones',  
 'Apple Airpods Headphones' = 'Headphones',  
 'Bose SoundSport Headphones' = 'Headphones',  
 '27in FHD Monitor' = 'Monitor',  
 '27in 4K Gaming Monitor' = 'Monitor',  
 '34in Ultrawide Monitor' = 'Monitor',  
 '20in Monitor' = 'Monitor',  
 "Google Phone" = "Phone",  
 'iPhone' = 'Phone',  
 'Vareebadd Phone' = 'Phone',  
 'Macbook Pro Laptop' = 'Laptop',  
 'ThinkPad Laptop' = 'Laptop',  
 .default = "Other Products")) %>%  
 group\_by(Product) %>%  
 summarize(AvgPrices = sum(AvgPrice)) %>%  
 arrange(-AvgPrices)

#TopFour - combines data from top 2 cateogries of QuantitySold and top 2 categories of the Prices data sets to show the average profit of each cateogry  
TopFour <- full\_join(QuantitySold, Prices, by = 'Product') %>%  
 filter(Product %in% c('Batteries', 'Headphones', 'Laptop', 'Phone')) %>%  
 mutate(profit = AvgPrices\*Sales) %>%  
 arrange(-profit)

#Phone - displays all data from the 'Phone' cateogry, specially looking at the quantity sold by date  
Phone <- sales %>%  
 filter(Product %in% c('Google Phone', 'iPhone', 'Vareebadd Phone')) %>%  
 group\_by(Date, Product, Quantity) %>%  
 summarize(sold = sum(Quantity))

#iPhone - displays all data of iPhones sold, specifically looking at the amount of iPhones sold to various urban cities  
iPhone <- sales %>%  
 filter(Product %in% c('iPhone')) %>%  
 group\_by(City) %>%  
 summarize(count = sum(Quantity)) %>%  
 arrange(-count)

## My Story

Dear Managers of Amazon’s Technology Product Division,

I want to highlight that on average, the best selling category of technology products are batteries and headphones (see fig. 1). This is important you to know because it can be critical to keep batteries and headphones in consistent stock throughout the year, as they tend to either sell rather often compared to other technology products.

Apart from knowing which types of products sell the most, it might be useful to know which products sell, on average, for the highest price. Hence, making these products the most profitable type of technology for Amazon to sell. The technology products selling for the highest prices on average are laptops and phones (see fig. 2). Laptops, by far, create the most revenue for selling a singular unit. This is important because prices can determine whether a consumer buys the product or not. If managers are seeing a small amount of laptops sold, could this be because of high prices of average? Or could this simply mean less customers want laptops over other technology product?

Given that batteries and headphones sell the most on average, and laptops and phones sell for the highest prices, you might want to know which product actually produces the highest revenue. By multiplying the average price by the sales for the year, I found the highest revenue is brought in from phones and laptops (see fig. 3). It is important to realize here that phones are sold substantially more the laptops, and that is why phones on average have produced more revenue for Amazon.

To dig further into the demographics of phones sold, I graphed the amount of phones sold throughout the 2019-2020 year by phone type (see fig. 4). I found that, given the data of select zip codes from major cities, iPhones clearly sold the most, then Google Phones, and finally Vareebadd Phones. Because the data is so consistent within these select zip codes, we might infer this is the same for many urban cities as well. To maximize revenue gain, managers might promote the sale of iPhones on Amazon’s website by partnering with Apple and so forth.

To see which cities to target, given the data provided, I graphed the number of iPhones sold in each Urban City (see fig. 5). Of the nine cities we have collected data from, the most iPhones were sold in San Fransicso and California. While these are not full representations of the cities total sales, it might be useful to know that targeting California residents to buy iPhones can be a start to maximizing revenue.

The main takeaway I want to provide to the Technology Product Division is that certain categories and certain products produce substantially more revenue for the company than other products. If the division’s goal were to expand revenue, instead of selling substantially more batteries, it will be more effective to sell more laptops and phones. A deeper dive into this theory looks at what types of phones to promote and where might the company start promoting them.

## Final Data Visualizations

Fig. 1: **Sales for Each Technology Product Category**

ggplot(QuantitySold, aes(x = Sales, y = reorder(Product, Sales))) +  
 geom\_col(fill = c('red', 'grey', 'red', 'grey', 'grey', 'grey', 'grey')) +  
 theme\_classic() +  
 xlab('Quantity Sold') +  
 ylab('Product Category') +  
 labs(title = 'Sales for Each Product Category', hujst = 0.5)

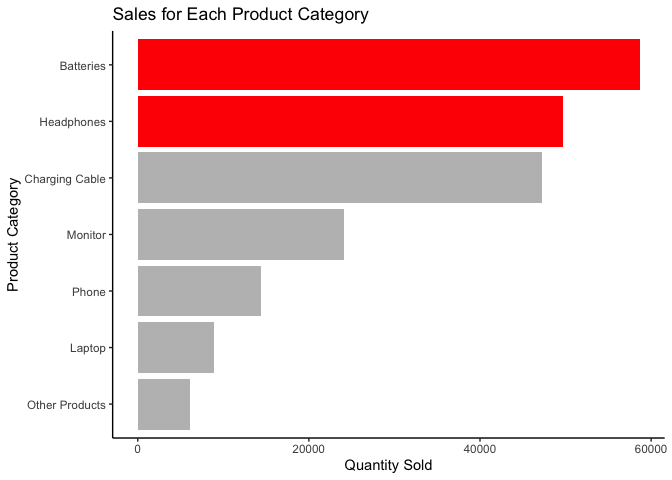


Fig. 2: **Average Prices for Each Product Category**

ggplot(Prices, aes(x = AvgPrices, y = reorder(Product, AvgPrices))) +  
 geom\_col(fill = c('red', 'red', 'grey', 'grey', 'grey', 'grey', 'grey')) +  
 theme\_classic() +  
 xlab('Average Prices') +  
 ylab('Product Category') +  
 labs(title = 'Average Prices for each Product Category', hujst = 0.5)

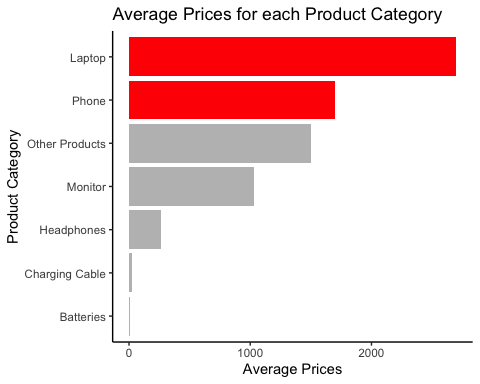


Fig. 3: **Average Revenue for Each Product Category**

ggplot(TopFour, aes(x = profit, y = reorder(Product, profit))) +  
 geom\_col(fill = c('red', 'red', 'grey', 'grey')) +  
 theme\_classic() +  
 xlab('Average Prices') +  
 ylab('Product Category') +  
 labs(title = 'Average Profit for Each Product Category', hujst = 0.5) +  
 scale\_x\_continuous(labels = label\_number(scale = 1/1000000, suffix = 'M'))

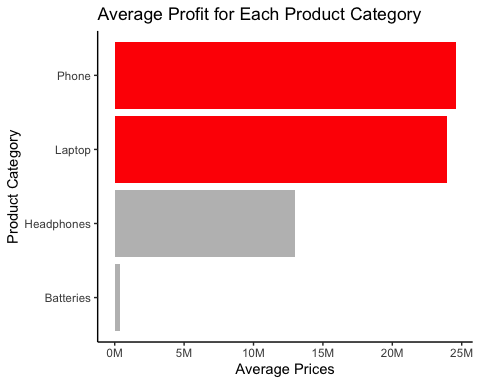


Fig. 4: **Types of Phones Sold from 2019- 202**

ggplot(Phone, aes(x = Date, y = sold, color = Product)) +  
 geom\_point() +  
 theme\_classic() +  
 xlab('Date (2019 - 2020)') +  
 ylab('Phones Sold') +  
 ggtitle(expression(underline("Phones Sold from 2019- 2020"))) +  
 theme(plot.title = element\_text(hjust = 0.5)) +  
 theme(legend.position = "top",  
 legend.title = element\_text(size = 9, face = 'bold'))

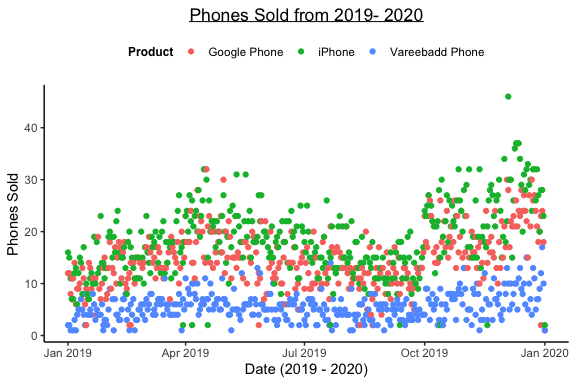


Fig. 5: **Number of iPhone’s Sold in Each Major City**

ggplot(iPhone, aes(x = count, y = reorder(City, count))) +  
 geom\_col(fill = c('red', 'red','grey', 'grey', 'grey', 'grey', 'grey', 'grey', 'grey')) +  
 theme\_classic() +  
 xlab('Number of iPhones Sold') +  
 ylab('City') +  
 labs(title = 'Number of iPhones Sold in Each Major City', hujst = 0.5)

