PSY6422 - Project

ID: 220251464

2023-05-16

STARBUCKS & COSTA COFFEE

Background & Research Question

Europe has been deemed to be the world's largest coffee consumer (per 2017-2021), accounting for 32% of global consumption in 2021 (cbi.eu, 2022). The two biggest coffee shop chain brands in Europe are **Costa Coffee** and **Starbucks** (Worldatlas.com/Economics, 2018). Therefore, many may be interested in the investigation of core differences of Costa versus Starbucks coffee.

For this reason, I wanted to explore the caffeine content and size difference in a selection of 'standard' drinks. Demonstrating the differences in what one would receive if they ordered the same drink from each brand. For the purpose of narrowing down the field of interest and ease of visualisation, I chose to focus on these two aspects.

The research Question becomes as follows: "How do the caffeine content and drink volume, for an equal size, differ between Costa Coffee and Starbucks for a selection of 'standard' coffee drinks in Europe and the UK?"

Data Origins

The raw data was obtained via the Starbucks and Costa websites, publicly accessible:

Starbucks. (2023, February). *Core Beverage Nutritional and Allergen Information (pdf)*. Starbucks Nutrition and Allergen Information. link
Costa. (2023, May). *Costa in-store products: allergen information guide (pdf)*. Costa Nutrition and allergens info. link

Fourteen 'core' coffee drink types were selected, each with existing as a Starbucks and Costa version. These are all on their general menus and the selection allows for a non-crowded visualization of 'core' types of coffee as well as equal comparison. Caffeine content (in milligrams (mg)) and drink volume (in milliliters (mL)) of these coffee drinks were inserted into an excel datasheet. This sheet makes up the raw data.
The data comprises the reported values as of May 2023 for Europe and UK in medium size (or standard if no medium exists) takeaway drinks.

Project Organization

The /data folder contains data required for the project, and /images contains images required for the project, as well as visualization outputs.

A codebook describing all labels and abbrevations used in this project for data, variables, functions etc. within this project is located at /codebook.xlsx.

Data Preparation

Loading packages

The project utilized the renv package to retain package versions, safeguarding it from potential updates to packages in the future.

Package versions used in this project are listed within the file /renv.lock

```
#Load packages with renv
install.packages("renv")
library(renv)
renv::restore()

#Import packages
library(tidyverse)
library(gapminder)
library(ggplot2)
library(png)
library(RCurl)
library(grid)
library(dplyr)
library(scales)
library(showtext)
library(here)
library(readxl)
library(knitr)
library(kableExtra)
library(readr)
```

Importing data

```r
#Specify image relative paths for logos and legend icons
Starbucks_image <- readPNG((here::here("images", "starbucks_logo.png")),
                          native = TRUE)
Costa_image <- readPNG((here::here("images", "costa_logo.png")),
                       native = TRUE)

starbucks_cup <- readPNG((here::here("images", "starbucks_cup.png")),
                         native = TRUE)
costa_cup <- readPNG((here::here("images", "costa_cup.png")),
                     native = TRUE)
```

```r
#Read raw data
rawdata <- read_excel(here::here("data","DAAVDATA.xlsx"))
kable(rawdata, format = "markdown")
```

Type of coffee	Volume (mL) Starbucks	Volume (mL) Costa	Caffeine (mg) Starbucks	Caffeine (mg) Costa
Decaf Coffee	455	382	2	2
Single-shot Espresso	25	30	33	100
Frappuccino / Frappé	455	499	33	100
Cold Brew	455	345	50	210
Cortado	170	180	66	141
Double-shot Espresso	50	60	66	200
Latte	455	364	66	200
Iced Latte	455	473	66	200
Flat White	227	300	66	241
Filter / Brewed Coffee	455	382	136	256
Mocha	455	332	66	325
Cappucino	455	362	66	325
Triple-shot Espresso	75	90	99	325
Americano	455	340	99	325

Cleaning the data

```r
#Create clean dataframe
cleandata <- data.frame(
  Type.of.Coffee = c(rep(rawdata$'Type of coffee', 2)),
  Brand = rep(c("Starbucks" , "Costa"), 14),
  Caffeine = c(rawdata$`Caffeine (mg) Starbucks`,
               rawdata$`Caffeine (mg) Costa`),
  Volume = c(rawdata$`Volume (mL) Starbucks`, rawdata$`Volume (mL) Costa`))

#Define the desired orders
order_list <- c(1, 1, 2, 2, 3, 3, 4, 4, 5, 5, 6, 6, 7, 7, 8, 8, 9, 9, 10, 10,
```

```
                11, 11, 12, 12, 13, 13, 14, 14)
order_list2 <- c(1, 15, 2, 16, 3, 17, 4, 18, 5, 19, 6, 20, 7, 21, 8, 22, 9, 23,
                10, 24, 11, 25, 12, 26, 13, 27, 14, 28)

#Define a function to reorder a vector based on an order list
reorder_vec <- function(vec, order_list) {
  vec[order_list]}

#Reorder the columns using the reorder_vec function
cleandata <- cleandata %>%
  mutate(Type.of.Coffee = reorder_vec(Type.of.Coffee, order_list),
         Caffeine = reorder_vec(Caffeine, order_list2),
         Volume = reorder_vec(Volume, order_list2))
```

```
#Specify order of Coffee
cleandata$Type.of.Coffee <- factor(cleandata$Type.of.Coffee,
                           levels = unique(cleandata$Type.of.Coffee),
                           ordered = TRUE)

#(see visualisation section for final dataframe)
```

Visualisation

```
#Define data to plot
ggp <- ggplot(cleandata, aes(fill=Brand, y=Caffeine, x=Type.of.Coffee))

#Display clean and processed data
kable(cleandata, format = "markdown")
```

Type.of.Coffee	Brand	Caffeine	Volume
Decaf Coffee	Starbucks	2	455
Decaf Coffee	Costa	2	382
Single-shot Espresso	Starbucks	33	25
Single-shot Espresso	Costa	100	30
Frappuccino / Frappé	Starbucks	33	455
Frappuccino / Frappé	Costa	100	499
Cold Brew	Starbucks	50	455
Cold Brew	Costa	210	345
Cortado	Starbucks	66	170
Cortado	Costa	141	180
Double-shot Espresso	Starbucks	66	50
Double-shot Espresso	Costa	200	60
Latte	Starbucks	66	455
Latte	Costa	200	364
Iced Latte	Starbucks	66	455
Iced Latte	Costa	200	473
Flat White	Starbucks	66	227
Flat White	Costa	241	300
Filter / Brewed Coffee	Starbucks	136	455
Filter / Brewed Coffee	Costa	256	382

Type.of.Coffee	Brand	Caffeine	Volume
Mocha	Starbucks	66	455
Mocha	Costa	325	332
Cappucino	Starbucks	66	455
Cappucino	Costa	325	362
Triple-shot Espresso	Starbucks	99	75
Triple-shot Espresso	Costa	325	90
Americano	Starbucks	99	455
Americano	Costa	325	340

Colours

Colours of bars in the visualisation correspond to iconic colours of the brands, for the Caffeine content, increasing glance value for those familiar with brands. This is important as those most likely to be interested in the data are those familiar with the brands.
Additionally, for better clarity, the bars corresponding to the drink Volume, are coloured coffee-like colours, a light contrast with the Caffeine bars is ensured by slightly fading them and brands are differentiated by the two different hues of brown.

(see Design section)

Custom Text

Using the showtext package, a personalized text was imported and incorporated into the visualization. The chosen font closely resembles the fonts employed by the two brands, thereby enhancing the professional appearance and overall attractiveness of the visualization.

```
font_add_google(name = "Source Sans Pro", family = "Source Sans Pro")
#Load new custom font for showtext package

showtext_auto()
#Automatically use showtext for plot
```

Design

The logos and icons imported earlier are used in the legends for the graph. They provide a visual appeal of immediate recognition for the logos, which correspond to the bar colours, and the coffee cup icons link in with the overall theme of the investigation: Coffee. These coffee cups are coloured correspondingly with the bars which represent each brands' drink volume.
The brighter logo-colours where chosen to represent Caffeine to draw more attention to this statistical representation, as compared to the coffee-colours representing the drink volume, as seemed fitting for the liquid measurement. The coffee-colours were made slightly transparent, making them lighter to provide a better contrast,

The decision to use grouped barcharts with two y-axis was based on the fact that this would best visualise the contrast between the two brands for both the caffeine content *and* the drink size difference. Especially, due to the fact that a small drink can still have a lot of caffeine and vice versa. Decaf coffee was especially included to reflect this and for the same reason, it was placed first on the x-axis as a point of comparison.
The rest of the data was ordered on the basis of least to most caffeine for *both* of the Brands, for each type of coffee. Meaning, although Costa Cold Brew had a higher caffeine content than Costa Cortado, Starbucks Cold Brew did not have a higher caffeine content than Starbucks Cortado, so, Cold brew is put first on the x-axis. The volume data is dependent on the type of drink for each brand and is therefore ordered based on

the caffeine data. Descriptive labels were included to be able to gain an understanding from the visualisation if it were to be presented on its own.

```r
#Plot graph and customize various plot aesthetics
graph <- ggp +
#adding volume barplot:
  geom_bar(aes(y = Volume / 1,
          fill = ifelse(Brand == "Starbucks", "C", "D")), #seperate bar colours
          position = "dodge", stat="identity",    #specifying a grouped barplot
          width = 0.8, alpha=.5,              #altering bar width and transparancy,
          show.legend = FALSE) +                  #hiding default legend
  #adding caffeine barplot, specifying colours
  geom_bar(aes(fill = ifelse(Brand == "Starbucks", "A", "B")),
          position = "dodge", stat = "identity",   #making it a grouped barplot
          width = 0.5,                #narrowing bar width to better see plot bars
          show.legend = FALSE) +                  #hiding default legend
  #naming left y-axis, changing axis breaks, making space for legends
  #and removing gap between graph and axis:
  scale_y_continuous(name = "Caffeine (mg)",
                  breaks = seq(0, 500, by = 50),
                  limits = c(0, 670), expand =c(0,0),
  #adding secondary y-axis:
    sec.axis = sec_axis(~.*1, name="Volume (mL)",
              breaks = seq(0, 500, by = 50))) +        #specyfing axis breaks
  ggtitle("Costa VS Starbucks: Caffeine & Size") +        #creating the title
  labs(x = "Type of Coffee*",                         #labelling the x-axis
      subtitle = "Europe & UK, 2023",                  #creating the subtitle
      caption = "*Takeaway, Standard/Medium Size") +      #and a x-axis caption
  #specifying colours for the different bars:
  scale_fill_manual(values = c("A" = c("#1b703f"), "B" = c("#B91345"),
                          "C" = c("#63330b"), "D" = c("#996633"))) +
  theme(aspect.ratio = 3/5,                   #setting the proportions of the plot
      panel.grid.major = element_blank(),             #removing the plot grid
      panel.grid.minor = element_blank(),
      panel.background = element_blank(),         #removing the plot background
      plot.margin = unit(c(1,1,2,1), "cm"),    #changing plot margin around it
    #changing colour + thickness of axis lines:
      axis.line = element_line(colour = "black", linewidth = 1),
    #adjusting x-axis break labels: direction, alignment and axis distance:
      axis.text.x = element_text(angle = 50, vjust = 0.5,
                              hjust = 1, margin = margin(t = -30)),
    #adjusting x-axis label title: distance and text size:
      axis.title.x = element_text(margin = margin(t = 70), size = 12),
    #adjusting left y-axis label: distance:
      axis.title.y = element_text(margin = margin(r = 15)),
    #adjusting right y-axis label: distance:
      axis.title.y.right = element_text(margin = unit
                                    (c(0, 0, 0, 5), 'mm')),
    #altering plot title aesthetics:
      plot.title = element_text(margin = margin(b = 6),
                  hjust = 0.5, size = 14, face = "bold"),
    #changing subtitle aesthetics:
      plot.subtitle = element_text(margin = margin (b = 20),
                      size = 11, hjust = 0.5),
```

```r
    #altering the captions aesthetics:
        plot.caption = element_text(margin = margin (t = 10),
                        color = "#6e6e6e", face = "italic",
                        hjust = 0.5)) +
#
#creating my own legend:
  #adding Starbucks logo + coordinates:
   annotation_raster(Starbucks_image, xmin = 0.73, xmax = 1.39,
                                      ymin = 521, ymax = 573) +
  #adding Costa logo + coordinates:
   annotation_raster(Costa_image, xmin = 0.491, xmax = 1.63,
                                   ymin = 578, ymax = 630) +
  #adding cup icon for Starbucks Volume + coordinates:
   annotation_raster(starbucks_cup, xmin = 13.5, xmax = 14.5,
                                     ymin = 515, ymax = 570) +
  #adding cup icon for Costa Volume + coordinates:
   annotation_raster(costa_cup, xmin = 13.5, xmax = 14.5,
                                ymin = 575, ymax = 630) +
  #adding legend text on the left:
   annotate("text", x = 1.7:1.7, y = c(550, 605),
            label = c("Starbucks", "Costa"), hjust = 0) +
  #adding legend text on the right:
   annotate("text", x = 13.5:13.5, y = c(543, 603),
            label = c("Starbucks", "Costa"), hjust = 1) +
  #adding legend title on the left:
   annotate("text", x = 0.8, y = 659, label = "Brand",
            hjust = 0, fontface = "bold") +
  #adding legend title on the right:
   annotate("text", x = 14.2, y = 658, label = "Brand Volume",
            hjust = 1, fontface = "bold")
```
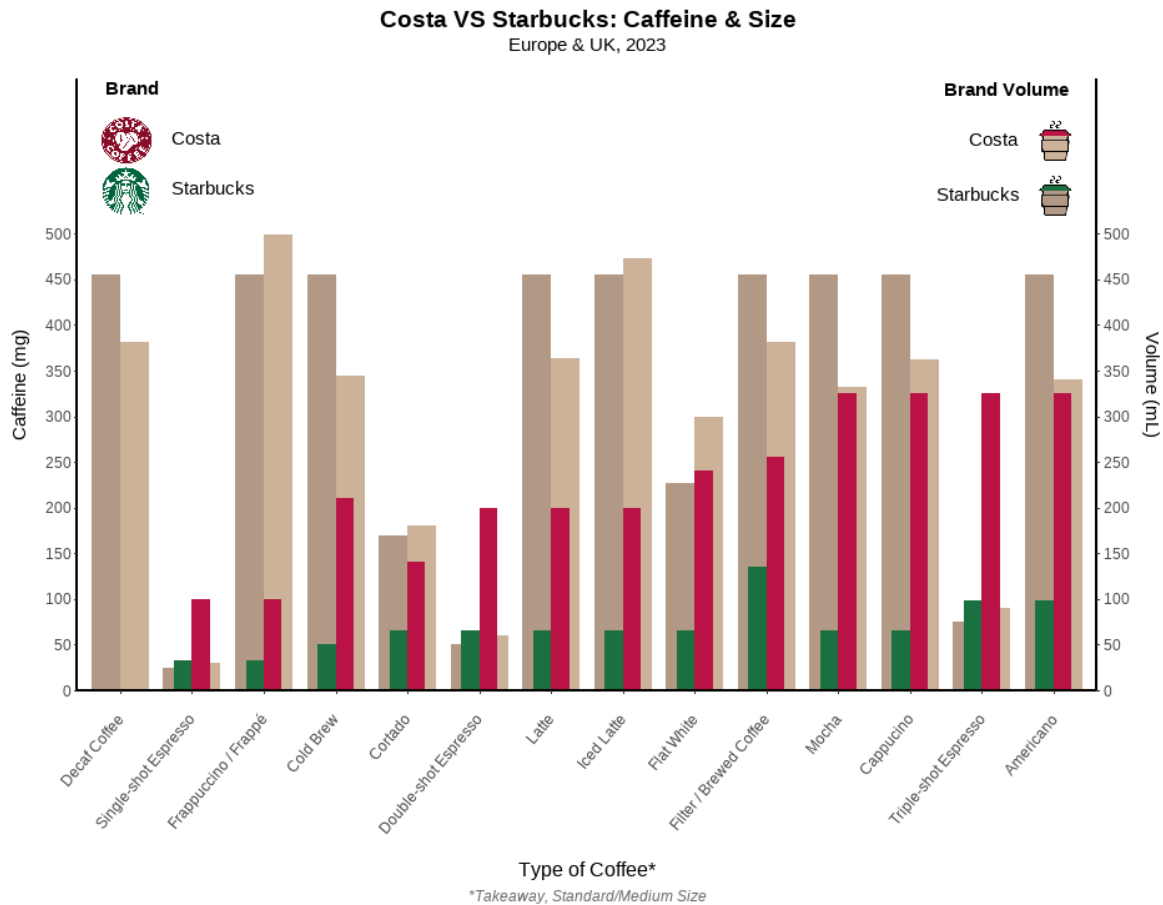
Saving Visualisation

```r
#Save plot to images folder
windows(width = 1000, height = 800) #Open windows graphics device
print(graph)
dev.print(file = here("images", "StarbucksVSCosta_Graph.png"), device = png,
          width = 1000, height = 800)
```

Result

Costa VS Starbucks: Caffeine & Size
Europe & UK, 2023

Interpretation & Future Direction

It appears that Costa coffees have more caffeine in them in comparison to Starbucks coffees, as this was shown for all drink types.

Additionally, there does not seem to be a big difference in drink size (volume) when contrasting the two brands. However, Starbucks does seem to have slightly larger drinks across the majority of the types of coffees.

Thus, although you may get slightly more coffee in your drink at Starbucks it seemingly has less than half the amount of caffeine as compared to a Costa coffee overall.

Furthermore, the graph informs on the relative differences between the coffees, such as the indication that there is just as much caffeine in an Americano as there is in a Triple-Shot Espresso.

Further aspects to investigate could be price differences as well as other nutritional information such as sugar content. This would best suit additional visualisations, for clarity, as for example sugar in grams would be quite alot higher than Caffeine milligrams and would not accurately demonstrate the individual differences.

Summary

I have improved my knowledge of R and utilization of R functions, R Markdown, and Github. The script serves as a testament to my newfound understanding of employing R methods for data manipulation, importing, cleaning, and visualization.

Additionally, I have improved the reproducibility of my work by using Renv for an automated package management system.

Moreover, the project showcases my proficiency in creating customized visualizations featuring multiple bar plots, axes, legends, and annotations, all aimed at effectively conveying information. As a result of this experience, I have also gained the ability to leverage R Markdown to produce dynamic reports and seamlessly integrate code, visualizations, and text.