Analyzing the number of online shopping activity using (R)

Installing Packages

To run the install.packages command remove the '-' from the r code chunks

- ```{r warning = FALSE, message=FALSE}
- install.packages("tidyverse")
- install.packages("ggthemes")
- ```

Functions part of Tidyverse ggplot2, tibble, tidyr, readr, purrr, dplyr, stringr, forcats

Calling the Functions

```
library(tidyverse)
library(dplyr)
library(forcats)
library(ggplot2)
library(ggthemes)
```

Importing Data

```
getwd()
```

```
## [1] "D:/Capstone Project - (Online Shopping 2)"
```

```
shoppers <- read.csv("online-shoppers.csv")
agegroup <- read.csv("online-shoppers-by-age-group.csv")
valuebypurchase <- read.csv("online-shoppers-by-value-of-purchase.csv")
mobilephoneshopping <- read.csv("reasons-for-shopping-via-mobile-phone-in-sin
gapore-2015.csv")
top10ecommercesite <- read.csv("top10ecommercesite.csv")
smartphonepenrate <- read.csv("smartphone-penetration-rate.csv")
```

Looking into the Datasets

```
summary(shoppers)
```

```
##       year        percentage
##  Min.   :2007   Min.   :35.00
##  1st Qu.:2010   1st Qu.:41.50
##  Median :2012   Median :51.00
##  Mean   :2012   Mean   :48.75
##  3rd Qu.:2015   3rd Qu.:54.00
##  Max.   :2018   Max.   :62.00
```

```
summary(agegroup)
```

```
##       year               age_group    percentage
##  Min.   :2007   15 to 24 years:12   Min.   : 1.0
##  1st Qu.:2010   25 to 34 years:12   1st Qu.:24.5
##  Median :2014   35 to 49 years:12   Median :36.0
##  Mean   :2013   50 to 59 years:12   Mean   :41.1
##  3rd Qu.:2016   60 and above  :12   3rd Qu.:58.0
##  Max.   :2018   7 to 14 years : 5   Max.   :95.0
##                 less than 7   : 5
```

```
summary(valuebypurchase)
```

```
##       year           value_of_purchase   percentage
##  Min.   :2007   $100 to $199  :12      Min.   : 5.00
##  1st Qu.:2010   $1000 to $1999:12      1st Qu.:11.00
##  Median :2012   $200 to $499  :12      Median :16.00
##  Mean   :2012   $2000 or more :12      Mean   :16.69
##  3rd Qu.:2015   $500 to $999  :12      3rd Qu.:21.00
##  Max.   :2018   less than $100:12      Max.   :37.00
```

```
summary(mobilephoneshopping)
```

```
##                                                                  x
##  It is a good way to pass the time                                :1
##  It is convenient                                                 :1
##  Like it because I can do it on-the-go                            :1
##  My friends are all doing it                                      :1
##  There are more and more apps that make it easy to shop digitally:1
##
##        y
##  Min.   :16.50
##  1st Qu.:30.40
##  Median :34.80
##  Mean   :37.28
##  3rd Qu.:47.40
##  Max.   :57.30
```

```
summary(top10ecommercesite)
```

```
##  Top.10.e.commerce.sites.in.Singapore.2019   percentage
##  Courts Singapore:1                        Min.   :0.390
##  eBay            :1                        1st Qu.:0.550
```

```
##  EZBuy           :1                          Median :1.010
##  Forty two       :1                          Mean   :2.432
##  Lazada          :1                          3rd Qu.:3.433
##  Love Bonito     :1                          Max.   :7.800
##  (Other)         :4
```

```
summary(smartphonepenrate)
```

```
##  Smartphone.penetration.rate..Share.of.the.population.in.Singapore.from.20
17.to.2023.
##  2017 :1

##  2018 :1

##  2019 :1

##  2020 :1

##  2021*:1

##  2022*:1

##  2023*:1

##        X
##  Min.   :75.0
##  1st Qu.:77.0
##  Median :78.0
##  Mean   :78.0
##  3rd Qu.:79.5
##  Max.   :80.0
##
```

Filter and sort the years in sequence

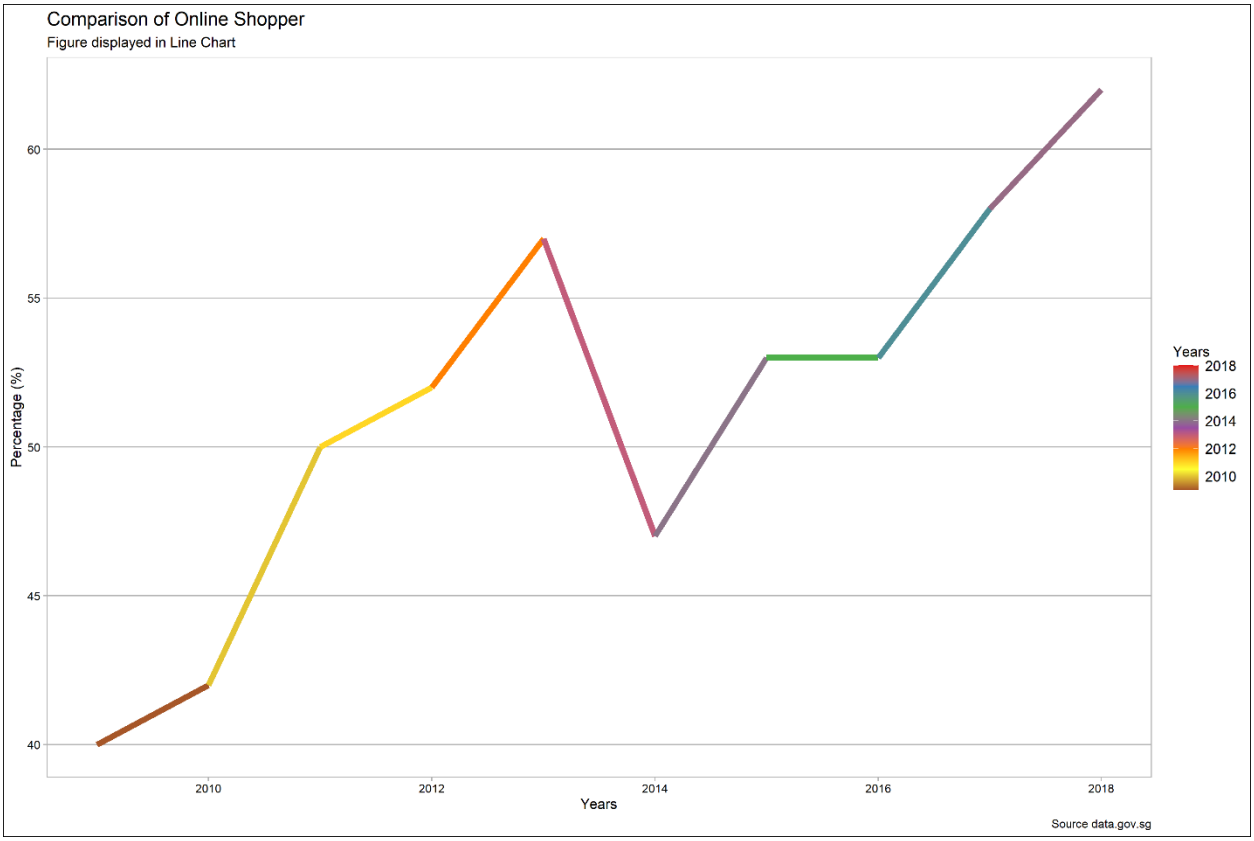
- shoppers %>%
- filter(year >= 2009L & year <= 2018L %>%

Comparison of Online Shoppers

As we can see from this line chart, the total number of online shoppers has increased significantly from 2010 to 2018 from around 35% to 65%, a whooping 30% jump. It shows that generally, over the years, online shopping has become more popular probably due to the increasing convenience of doing so. However, the increasing trend is affected by a

sudden and sharp drop between 2013 to 2014. It dropped from about 57% to 47%.Despite the drop, popularity of online shopping continue on an upward climb from 2014 onwards to 67% in 2018.

```
  shoppers %>%
  filter(year >= 2009L & year <= 2018L) %>%
  ggplot() +
  aes(x = year, y = percentage, colour = year) +
  geom_line(size = 2L) +
  scale_color_distiller(palette = "Set1") +
  labs(x = "Years", y = "Percentage (%)", title = "Comparison of Online Shopp
er", subtitle = "Figure displayed in Line Chart", caption = "Source data.gov.
sg", color = "Years") +
  ggthemes::theme_calc()
```

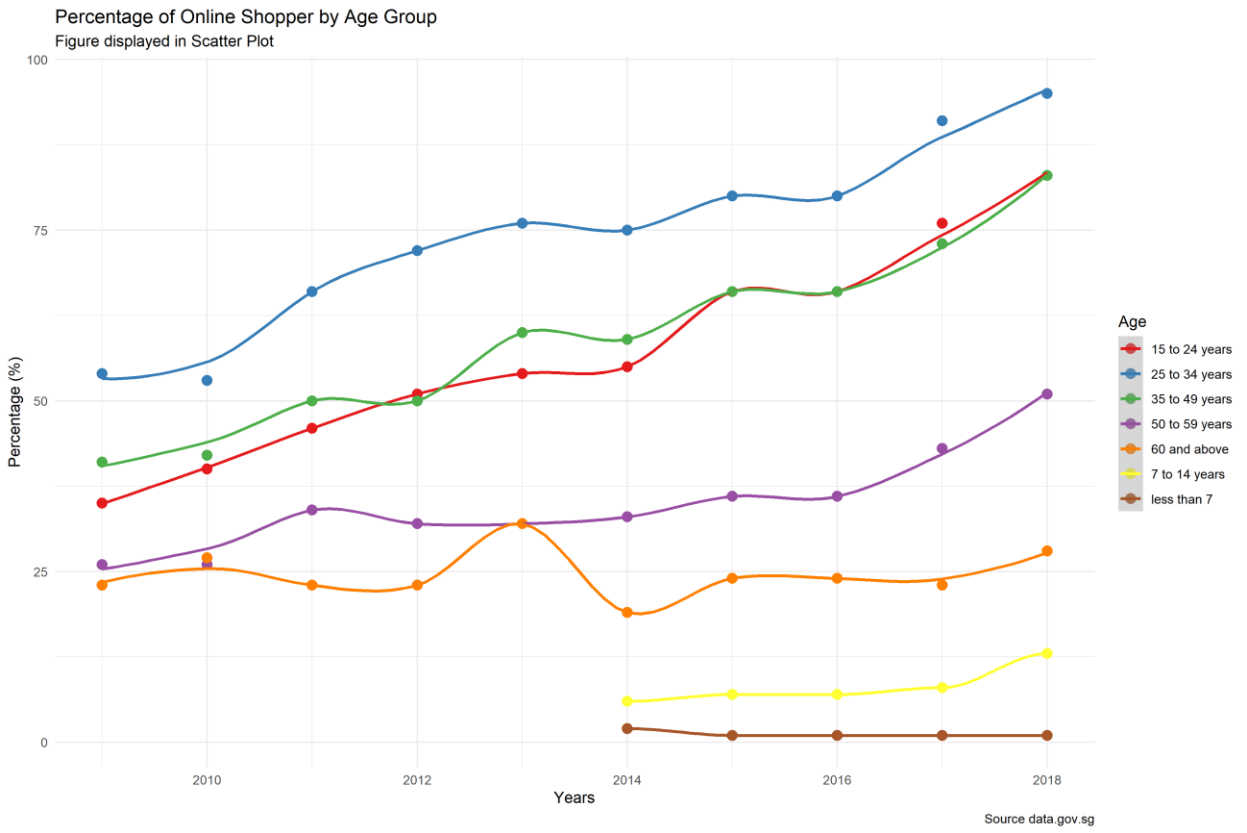


Percentage of Online Shopper by Age Group

Generally we can see that as the population gets older, the rate of increase in online shopping gets lower. From the data, it is evident that that the percentage of online shoppers for younger age group of 15 to 24, 25 to 35 and 35 to 49 increased very substantially from about 30+ through 40% in 2010 to more than 80% in 2018. Such is about a huge 50% jump. Online shoppers of age 50 to 59 also increased but the increase

was more gradual, standing at a mere 20%. Online shoppers of age 60 and above fluctuate over the years between 20% to 30% and no marked increase or decrease is seen.

```
  agegroup %>%
    filter(year >= 2009L & year <= 2018L) %>%
    ggplot() +
    aes(x = year, y = percentage, colour = age_group, group = age_group) +
    geom_point(size = 3L) +
    geom_smooth(span = 0.5) +
    scale_color_brewer(palette = "Set1") +
    labs(x = "Years", y = "Percentage (%)", title = "Percentage of Online Sho
pper by Age Group", subtitle = "Figure displayed in Scatter Plot", caption =
"Source data.gov.sg", color = "Age") +
    theme_minimal()
```



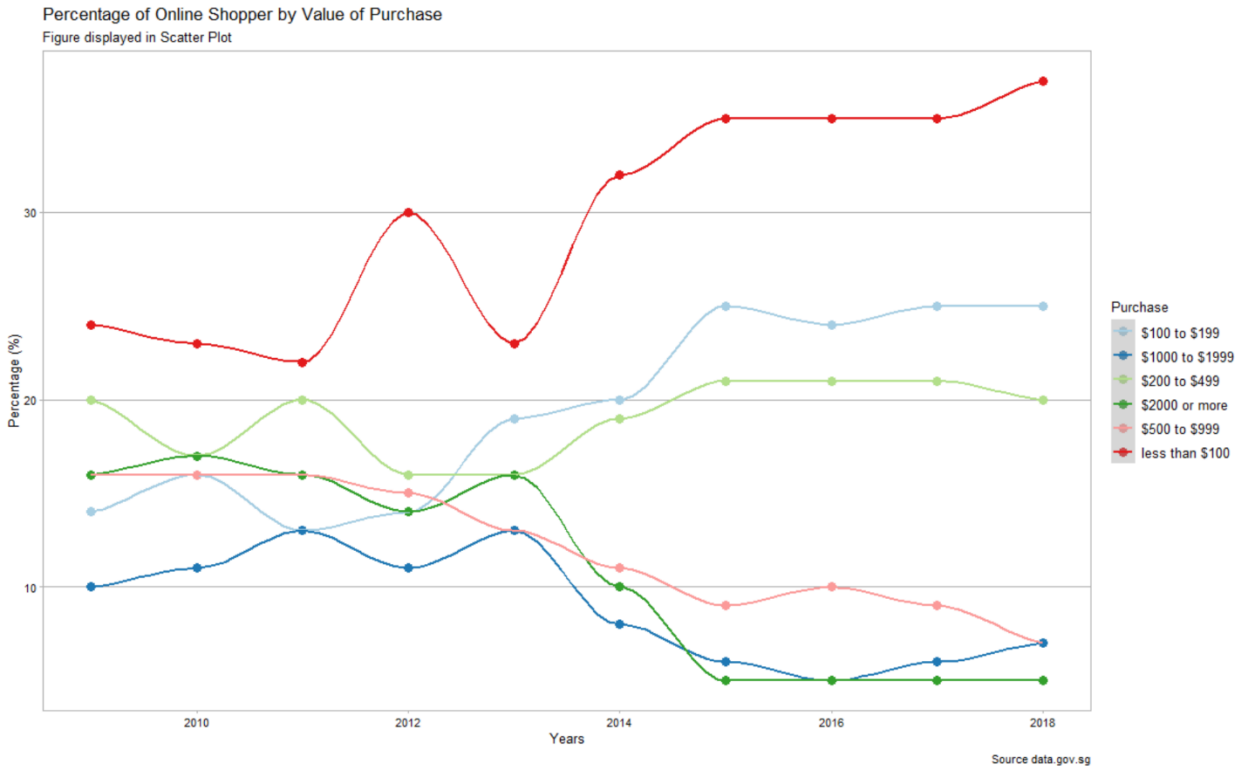
Percentage of Online Shopper by Value of Purchase

- geom_smooth using method 'loess' and formula ' y ~ x '

This graph is more interesting. We see a divergent trend in percentage of online shopping by value of purchase. That's over the years, more people are buying inexpensive items online and less people are making higher value purchases online. Although gradual, we do

see an increasing trend in online purchases of items that are less than $100, between $100 to $199 and between $200 to $499.But we notice a significant decrease in online purchases of higher value items that cost more than $500. You see that the pink, dark blue and dark green color line all 3 representing higher value purchases, decreasing over the years.Purchases of items more than $2000 decreased from 17% in 2010 to 5% in 2018.Purchases of items from $500 to $999 decreased from 12% to 7%. To sum up this graph, we can see that people are more willing to buy cheaper items online and probably are more reserved in making higher value purchases online.

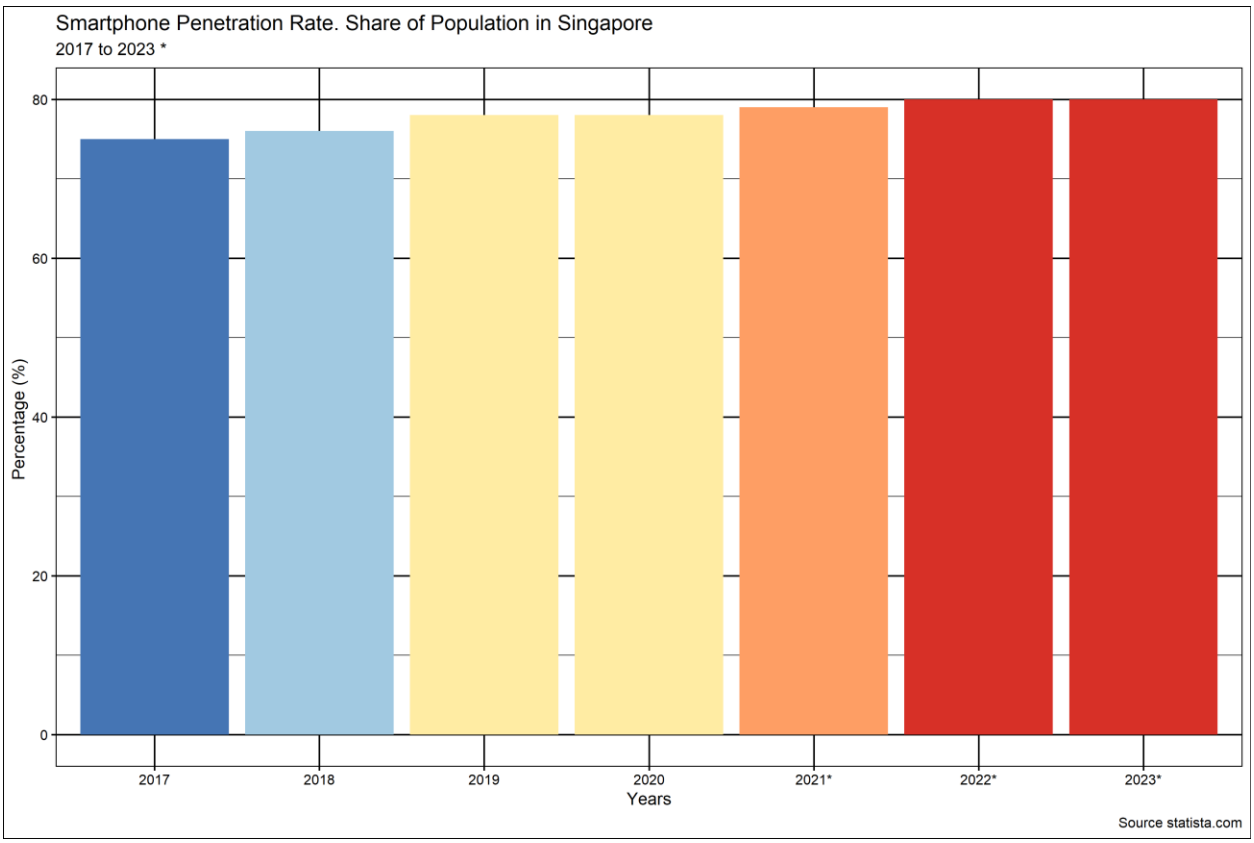
```
valuebypurchase %>%
  filter(year >= 2009L & year <= 2018L) %>%
  ggplot() +
  aes(x = year, y = percentage, colour = value_of_purchase, group = value_of_
purchase) +
  geom_point(size = 3L) +
  geom_smooth(span = 0.35) +
  scale_color_brewer(palette = "Paired") +
  labs(x = "Years", y = "Percentage (%)", title = "Percentage of Online Shopp
er by Value of Purchase", subtitle = "Figure displayed in Scatter Plot ", cap
tion = "Source data.gov.sg", color = "Purchase") +
  ggthemes::theme_calc()
```



Smartphone Penetration rate. Share of population in Singapore

From 2017 to 2020 currently, smartphone penetration rate has steadily increased by 3% and is expected to continually increase to 80% by 2023. We are confident of an increasing trend of smartphone penetration rate

```
ggplot(smartphonepenrate) +
  aes(x = Smartphone.penetration.rate..Share.of.the.population.in.Singapore.from.2017.to.2023., fill = X, weight = X) +
  geom_bar() +
  scale_fill_distiller(palette = "RdYlBu") +
  labs(x = "Years", y = "Percentage (%) ", title = "Smartphone Penetration Rate. Share of Population in Singapore", subtitle = "2017 to 2023 *", caption = "Source statista.com") +
  ggthemes::theme_foundation() +
  theme(legend.position = "none")
```



Convenience as reason why Singaporean turn to online shopping:

As we can see from the graph, the top 2 reasons that Singaporeans turn to online shopping is because they can do it on the go and the fact that it is convenient. Being able to do online on the go is largely due to the increase in smartphone penetration rate and this shows that convenience is a very important reason why Singaporeans turned to online shopping.

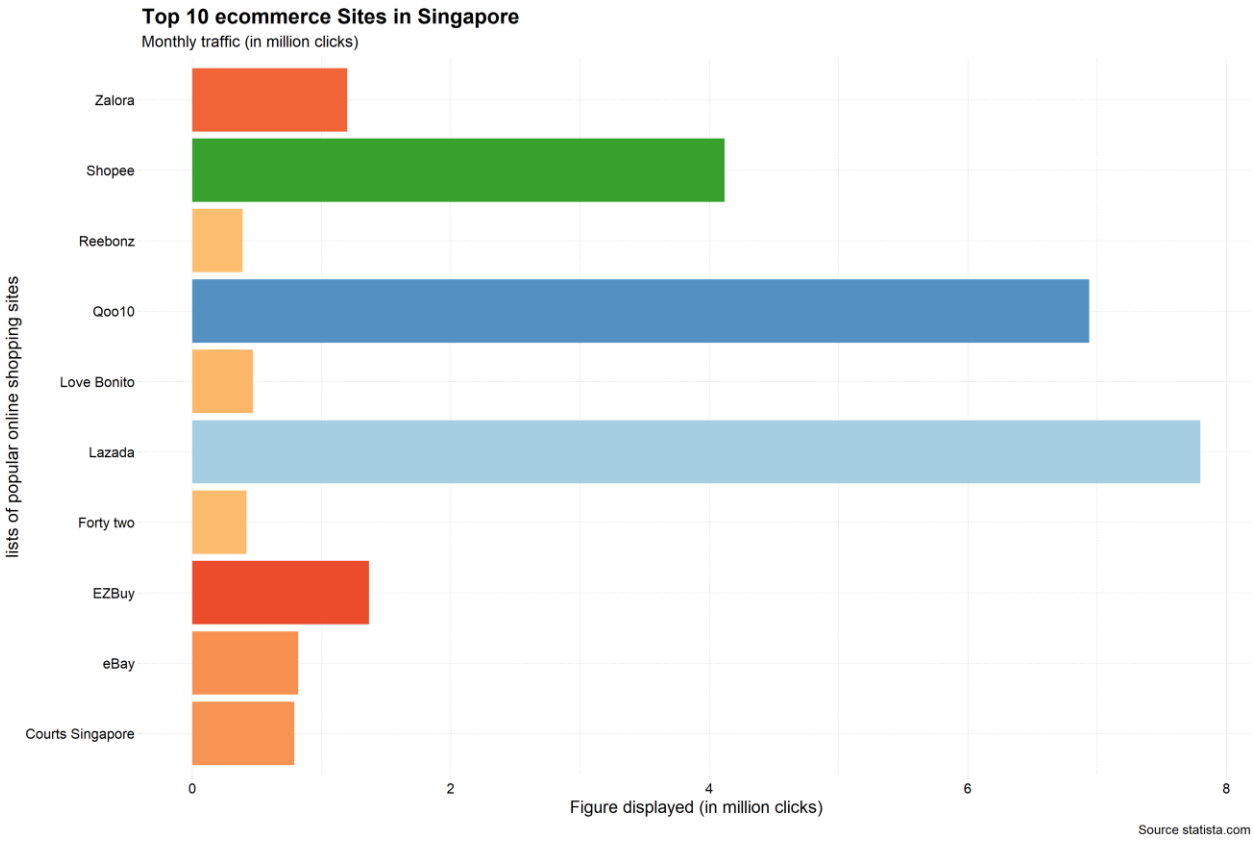
```
ggplot(mobilephoneshopping, aes(x=x, y=y)) +
  geom_point( color="orange", size=7, alpha=0.6) +
  geom_segment( aes(x=x, xend=x, y=0, yend=y), size=1.5, color="darkblue") +
  coord_flip() +
  ggthemes::theme_economist_white() +
  xlab("Source statista.com") +
  ylab("Percentage (%)") +
  ggtitle("Reason Why Singaporean Prefer Shopping Online")
```

Top 10 e-commerce sites in Singapore, by monthly traffic (in million clicks)

As we can see from the graph, there are so many different platforms that Singaporeans can choose to do online shopping. With the most popular one being Lazada Qoo10 and Shopee. This points out to how the large amount of choices consumers have access to contributed to the popularity of online shopping.

```
ggplot(top10ecommercesite) +
  aes(x = Top.10.e.commerce.sites.in.Singapore.2019, fill = percentage, weigh
t = percentage) +
  geom_bar() +
  scale_fill_distiller(palette = "Paired") +
  labs(x = "lists of popular online shopping sites", y = "Figure displayed (i
```

```
n million clicks)", title = "Top 10 ecommerce Sites in Singapore", subtitle =
 "Monthly traffic (in million clicks)", caption = "Source statista.com") +
  coord_flip() +
  ggthemes::theme_pander() +
  theme(legend.position = "none")
```



Conclusion

The reasons we have identified as our motivation to do this projects can be summarized to be the greater convenience of online shopping, greater variety of products online and the future online shopping has. These reasons require some statistics and data to prove and we aim to collect the following data to do so. To give a brief overview of the online shopping scene in Singapore, we aim to collect data related to the general trend of online shopping.

For reason 1, greater convenience of online shopping, getting data on smartphone penetration rate and reasons why singaporeans prefer online shopping would be helpful. For reason 2, greater variety of products, it would be good if we can get data on the variety of products one can access online and also the online platforms available to singaporeans.