

# Project

TaoZhang, Gefei Shen, Xuhao Wang, simon xiong

2024-11-10

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
##   filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
library(reshape2)
```

```
## Warning: package 'reshape2' was built under R version 4.3.3
```

```
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.3.3
```

```
## corrplot 0.92 loaded
```

```
data <- read.csv("user_behavior_dataset.csv")
```

Question 1: What is the preferred operating system for each age

# group?

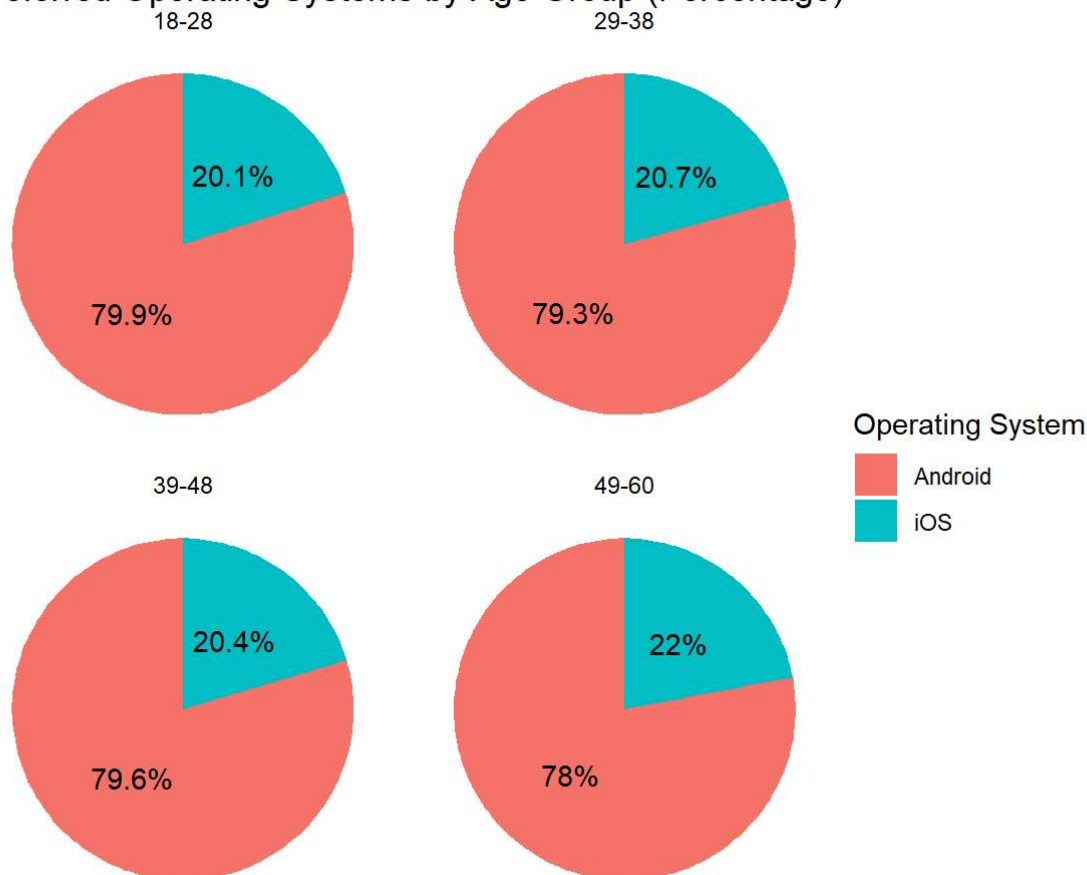
```
# Define age groups
data$Age_Group <- cut(data$Age, breaks = c(18, 28, 38, 48, 60, Inf),
                      labels = c("18-28", "29-38", "39-48", "49-60", "60+"), right = FALSE)

# Calculate the count and percentage of each operating system within each age group
os_counts <- data %>%
  group_by(Age_Group, Operating.System) %>%
  summarise(Count = n()) %>%
  group_by(Age_Group) %>%
  mutate(Percentage = Count / sum(Count) * 100) %>%
  ungroup()
```

```
## `summarise()` has grouped output by 'Age_Group'. You can override using the
## `.groups` argument.
```

```
# Create pie chart with percentage labels for each age group
ggplot(os_counts, aes(x = "", y = Percentage, fill = Operating.System)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y") +
  facet_wrap(~ Age_Group) +
  labs(title = "Preferred Operating Systems by Age Group (Percentage)",
       fill = "Operating System") +
  theme_void() +
  theme(legend.position = "right") +
  geom_text(aes(label = paste0(round(Percentage, 1), "%")),
           position = position_stack(vjust = 0.5))
```

## Preferred Operating Systems by Age Group (Percentage)



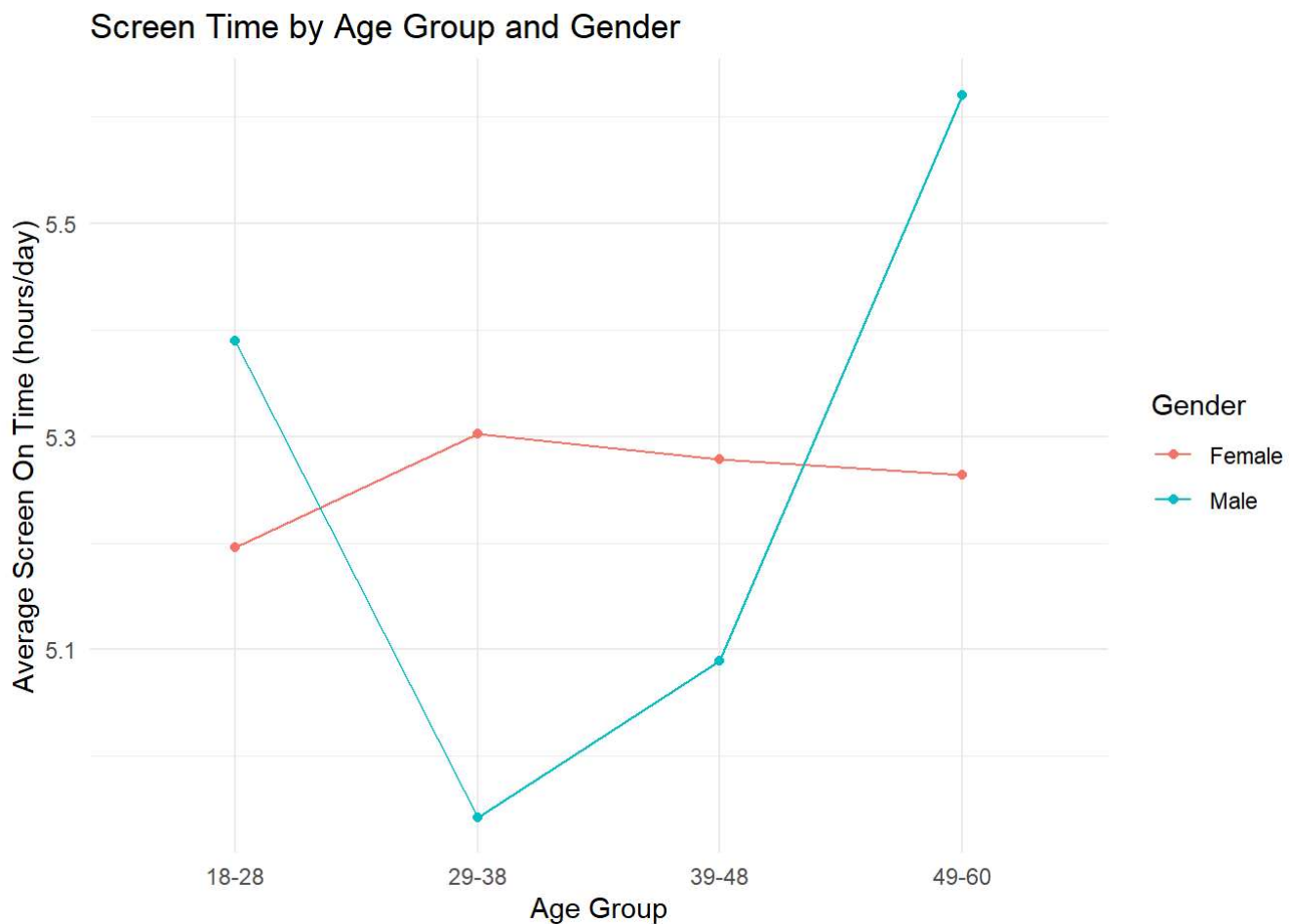
This chart shows the preference for operating systems (Android and iOS) by age group, revealing the proportion of users in each age group who use the different systems. As can be seen from the chart, Android dominates across all age groups, with the percentage of users ranging between 78% and 80%, demonstrating dominance. In contrast, iOS has a lower percentage of users, holding steady at around 20 per cent.

In terms of age groups, although iOS has seen a small rise in the older 49-60 age group (to 22 per cent), overall this change has not altered the dominance of Android across all age groups. There is a general preference for Android across all age groups, with Android particularly favoured in the younger to middle age group (18-48).

## Question 2: How does screen time vary by different age groups?

```
screen_time <- data %>%
  group_by(Age_Group, Gender) %>%
  summarise(Average_Screen_Time = mean(Screen.On.Time..hours.day., na.rm = TRUE), .groups = "drop")

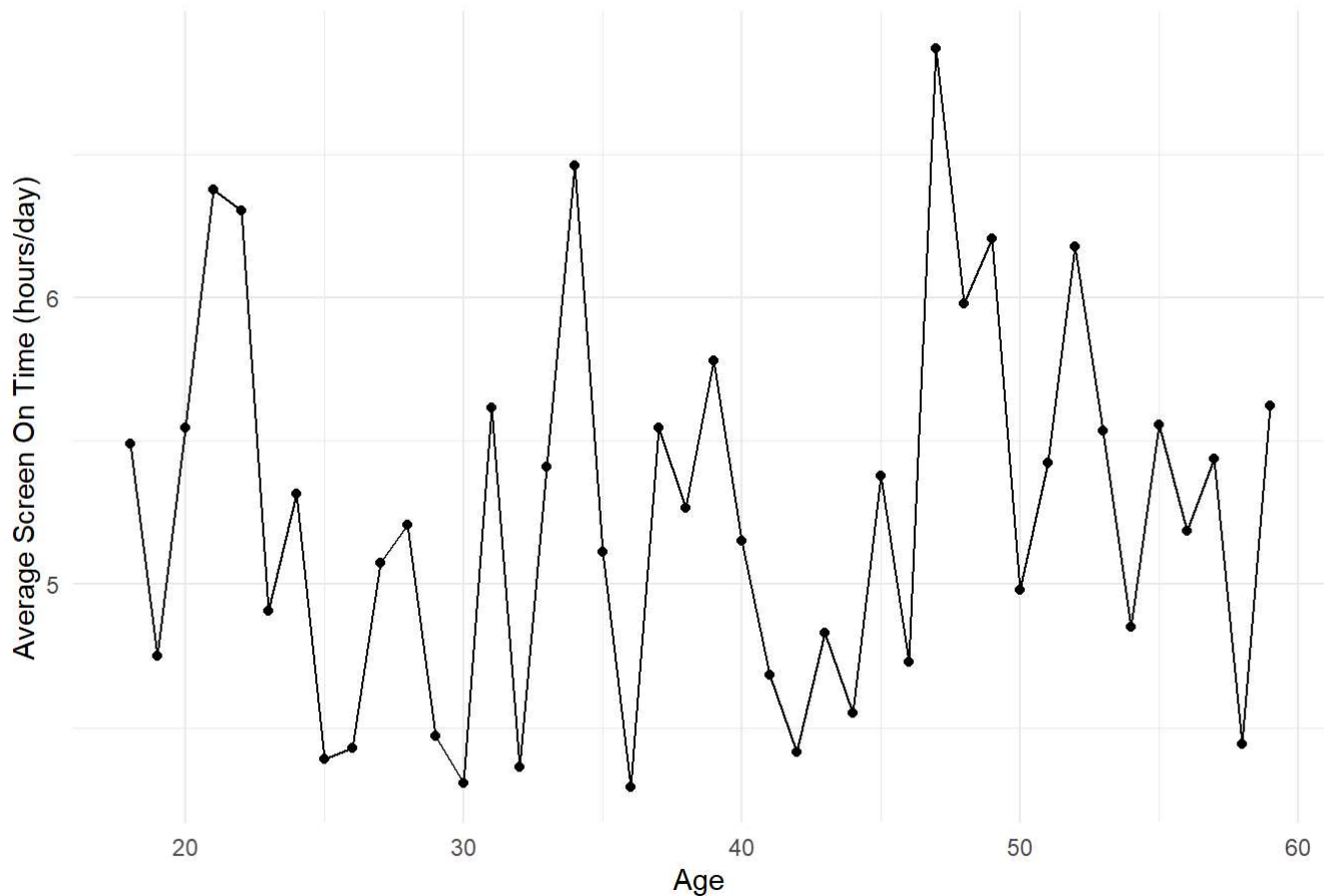
ggplot(screen_time, aes(x = Age_Group, y = Average_Screen_Time, color = Gender, group = Gender))
+
  geom_line() +
  geom_point() +
  labs(title = "Screen Time by Age Group and Gender",
       x = "Age Group",
       y = "Average Screen On Time (hours/day)") +
  theme_minimal()
```



```
screen_time_individual <- data %>%
  group_by(Age) %>%
  summarise(Average_Screen_On_Time = mean(Screen.On.Time..hours.day., na.rm = TRUE))

# Plot the average screen time by age
ggplot(screen_time_individual, aes(x = Age, y = Average_Screen_On_Time)) +
  geom_line() +
  geom_point() +
  labs(title = "Average Screen On Time by Age",
       x = "Age",
       y = "Average Screen On Time (hours/day)") +
  theme_minimal()
```

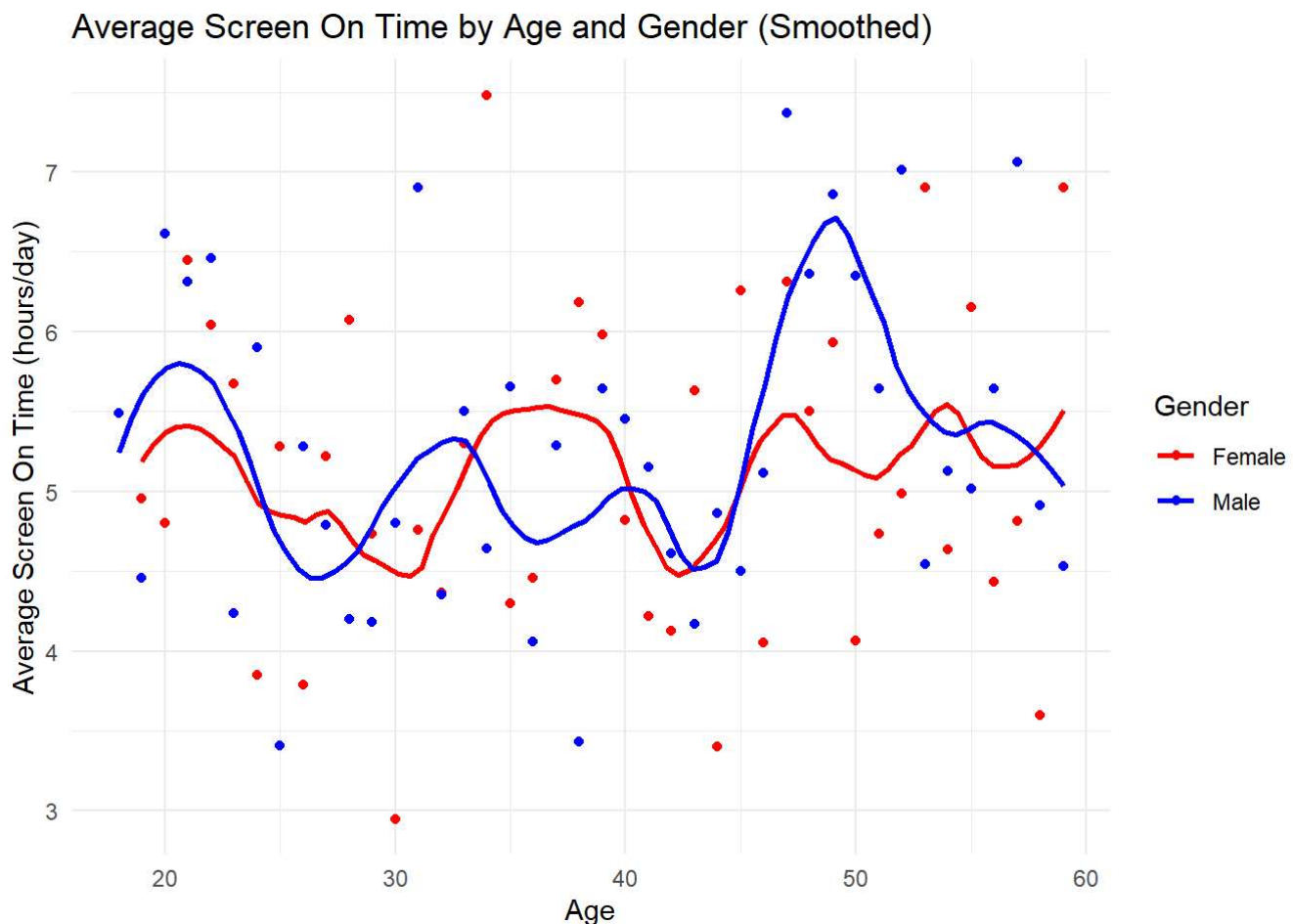
## Average Screen On Time by Age



```
screen_time_individual <- data %>%
  group_by(Age, Gender) %>%
  summarise(Average_Screen_On_Time = mean(Screen.On.Time..hours.day., na.rm = TRUE), .groups =
"drop")

# Plot the average screen time by age, with a smoothed curve
ggplot(screen_time_individual, aes(x = Age, y = Average_Screen_On_Time, color = Gender, group =
Gender)) +
  geom_point() + # Points for each age
  geom_smooth(method = "loess", span = 0.3, se = FALSE) + # Smooth Line with LOESS
  scale_color_manual(values = c("Male" = "blue", "Female" = "red")) +
  labs(title = "Average Screen On Time by Age and Gender (Smoothed)",
    x = "Age",
    y = "Average Screen On Time (hours/day)") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



In this plot, the x-axis represents the age of the user, and the y-axis shows the average screen usage time (in hours per day). The red and blue dots represent individual observations of screen usage time for females and males at different ages, respectively. Each dot indicates the average screen time for a specific gender at a particular age. In this plot, we use a smooth curve to display the overall trend to give audience a better and more intuitive comparison. We can observe a peak in screen usage time for males around 50 years old, while for females, there is a peak around 35 years old. Otherwise, the average screen usage time is similar for both genders.

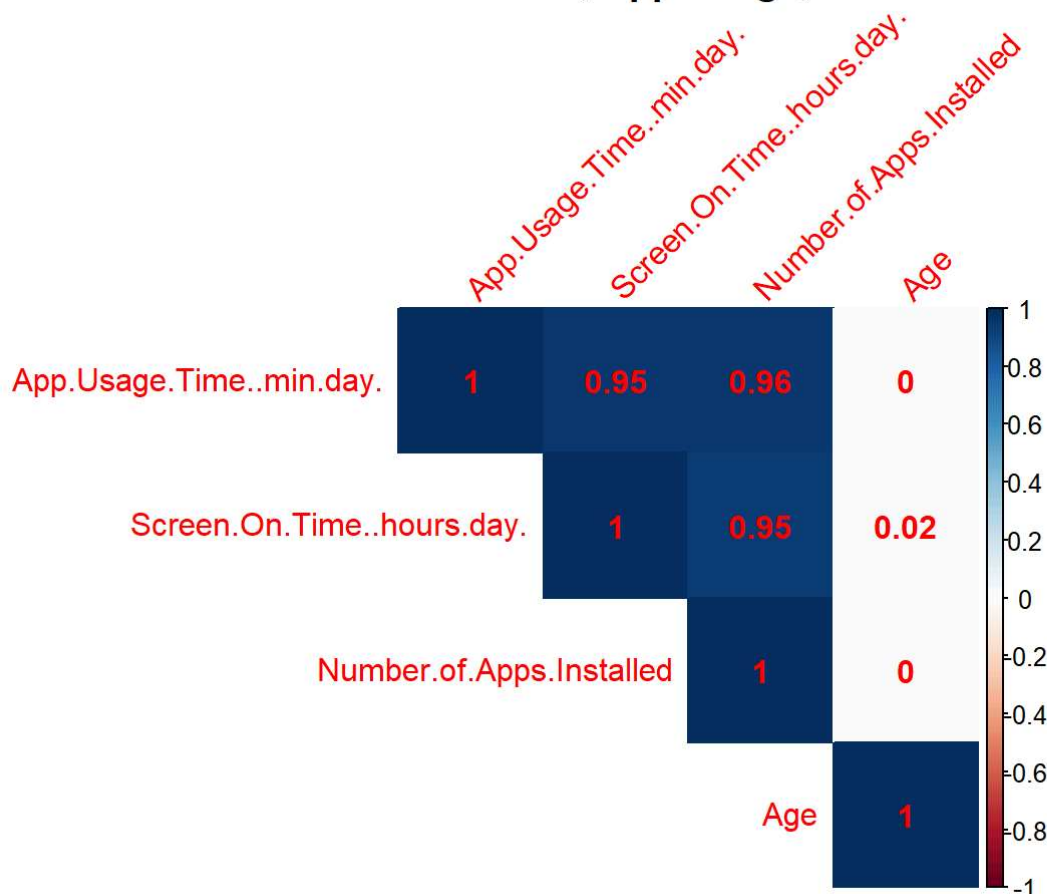
### Question 3: Does the amount of time spent on the phone daily correlate with the number of apps in that person's phone?

```
correlation_data <- data %>%
  select(App.Usage.Time..min.day., Screen.On.Time..hours.day., Number.of.Apps.Installed, Age)

correlation_matrix <- cor(correlation_data, use = "complete.obs")

corrplot(correlation_matrix, method = "color", type = "upper", tl.srt = 45,
  title = "Correlation Matrix for Screen Time, App Usage, and Number of Apps Installed",
  addCoef.col = "red")
```

## Correlation Matrix for Screen Time, App Usage, and Number of Apps Installed



Based on question 2, we want to further explore the relationship between screen using time and number of apps. This correlation matrix reveals strong interconnections between app usage time, screen-on time, and the number of apps installed. Specifically, app usage time has a very high positive correlation with both screen-on time and the number of apps installed, indicating that users who install more apps tend to spend more time using them and consequently keep their screens on longer. However, age shows almost no correlation with these variables, I think this is due to the fact that the data set was not sampled randomly, but the average of all ages was asked.