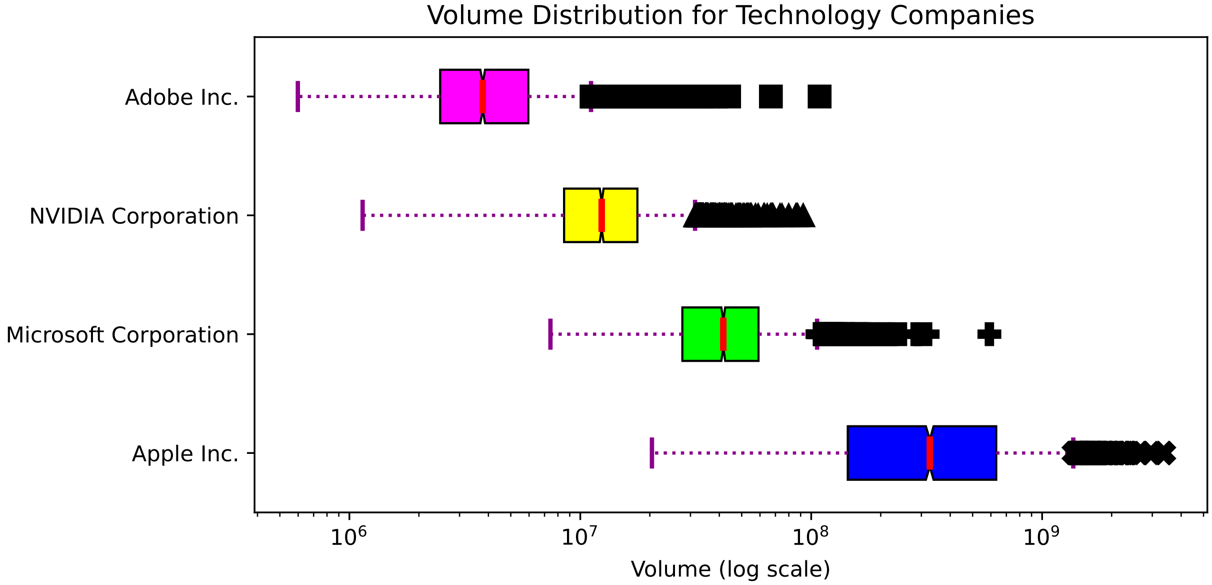
Data Visualization Problems

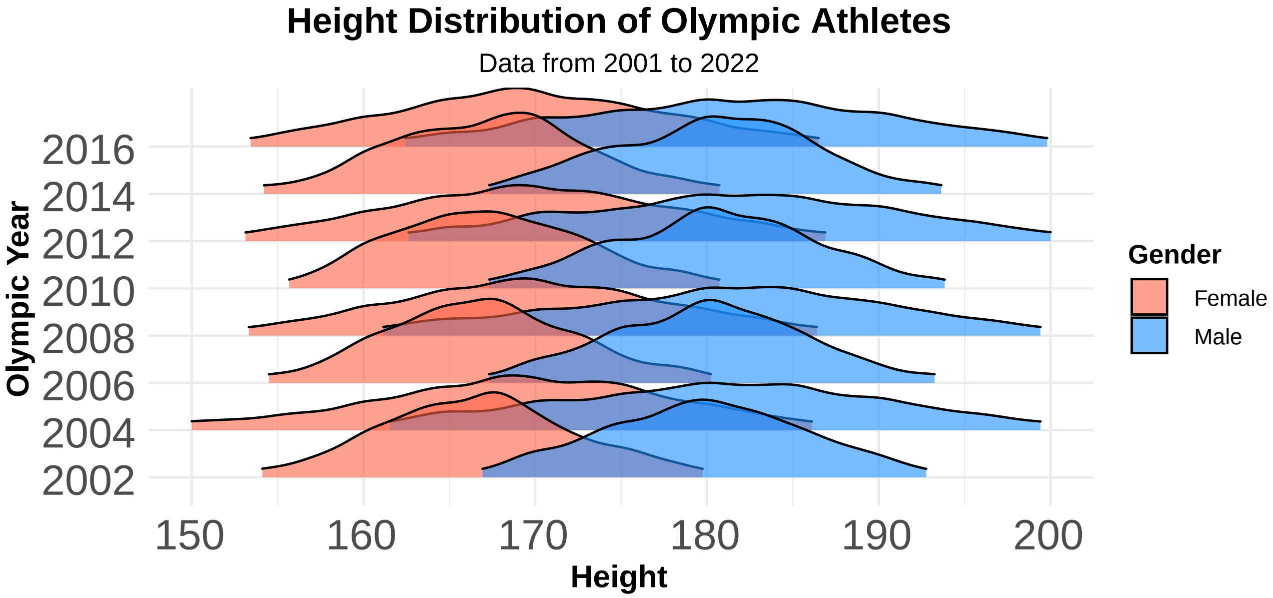
Problem 1 (Python)

Replicate the boxplot below that visualizes the volume distribution for Technology Companies using the data from StockIT folder.



Problem 2 (R)

Replicate the plot below obtained from olimpic dataset. Obtain similar plot for Weight column and plot below. What is the difference?



Source: athlete_events.csv

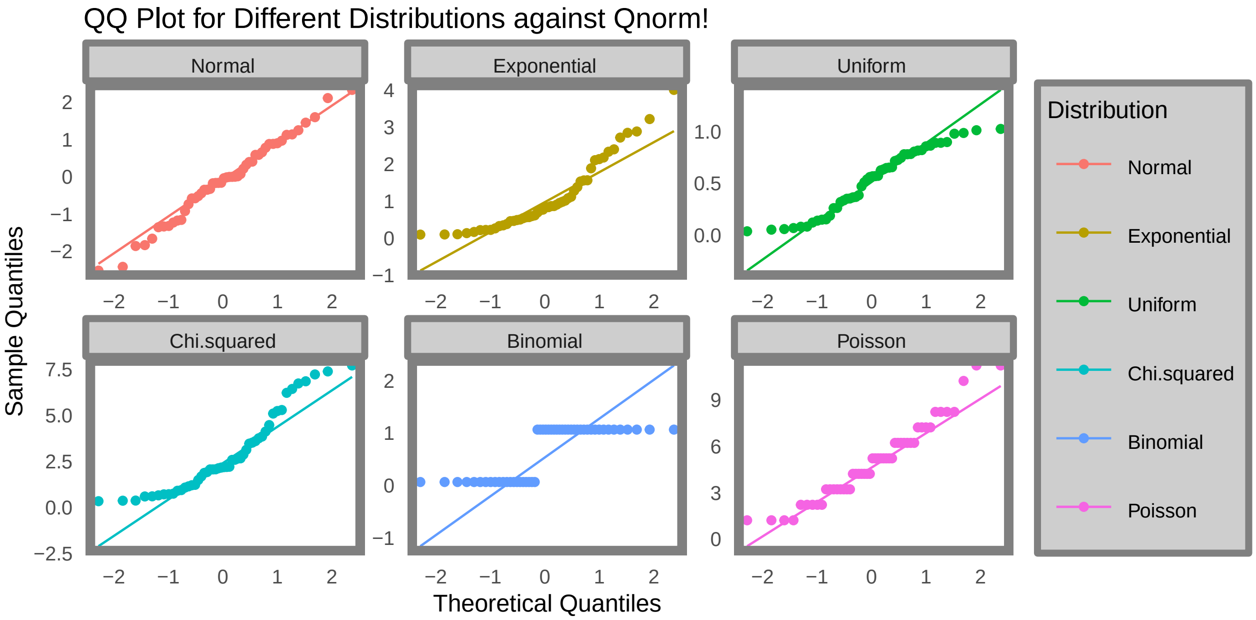
Problem 3 (R)

Given below dataset, replicate the plot.Plot similar one against all qunif, qext, qpois, qchisq, qbinom. Give summary and explanation. What is the difference between those distributions, and which one of them is different, and why?

```r
library(ggplot2)

# Set a seed for reproducibility
set.seed(2013)

# Generate random data from different distributions
n <- 50 # Sample size
normal_data <- rnorm(n)
exponential_data <- rexp(n, rate = 1)
uniform_data <- runif(n)
chi_squared_data <- rchisq(n, df = 3)
binomial_data <- rbinom(n, size = 1, prob = 0.5)  # Binomial distribution
poisson_data <- rpois(n, lambda = 5)  # Poisson distribution
```



Problem 4 (R)

Analyze and differentiate between light-tailed and heavy-tailed distributions both visually and statistically.

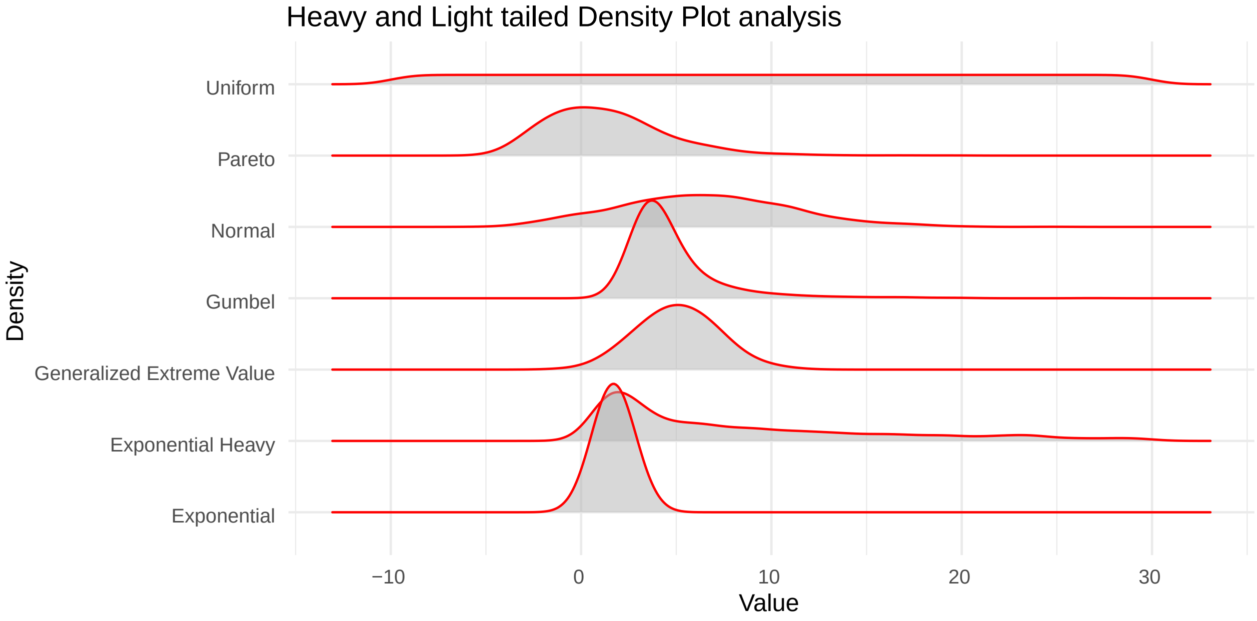
1. **Generate random samples** from any light-tailed and heavy-tailed distributions of your choice, for instance, Normal, Exponential, and Pareto, Generalized Extreme Value. You may choose your own parameters for the distribution. Compare at least 6 distributions here (its up to your choice) to make the comparison more reasonable.
2. **Analyze each distribution** statistically and explain your findings. You may want to look at the mean, variance, kurtosis, skewness, or any other statistical feature that can help differentiate between a light-tailed and a heavy-tailed distribution. Also, you can explain the difference using parameter comparison for the same distribution.
3. **Visualize the generated distributions.** Plot the ECDF (can also plot CDF, PDF to make point) for each of these distributions. Visually explain the difference between light-tailed and heavy-tailed distributions. Plot all in one.
4. **Histogram Analysis.** Plot the histogram of each of these distributions, observe and explain the histogram plots for these distributions.
5. **Exponential Distribution Comparison** - As an example, generate and visualize exponential distributions using different values of alpha. Compare the distributions and explain your findings. Which one are light tailed and heavy tailed? Why that happens?

6. Summarize your findings

Write a brief conclusion on insights obtained from the histograms, PDFs, and statistical analysis, highlighting the key differences between heavy-tailed and light-tailed distributions. (5 points)

Interesting plot that could be useful.

The following is an example of vizualion that you can replicate to make your point.



Problem 5(R)

Generate a dataset that should include the following attributes:

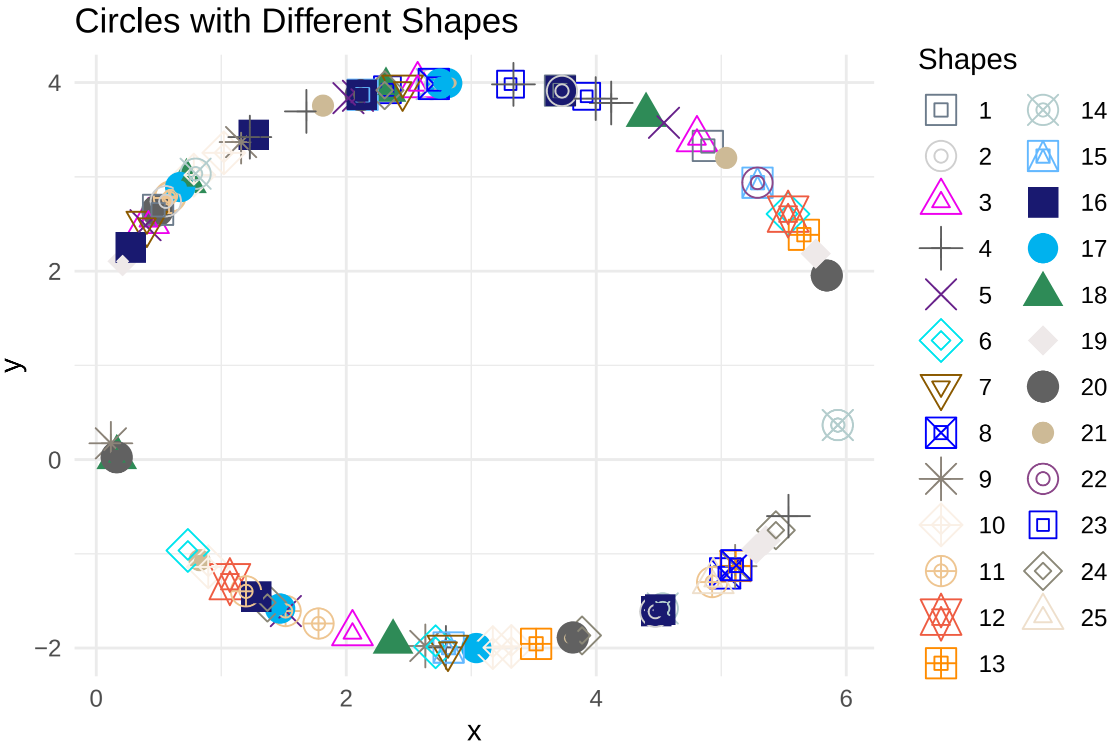
- It should have `x`, `y` values which should mimic the circle. The formula of a circle which is given by:

$$(x-3)^2 + (y-1)^2 = 9$$

- You must also generate shape feature.
- Make sure to cover all circle sides
- `N` number of datapoints = 100
- `M` number of classes = 25

Initialize your seed to ensure reproducible results. Global seed value is `2023`.

Replicate the plot



Problem 6 (R):

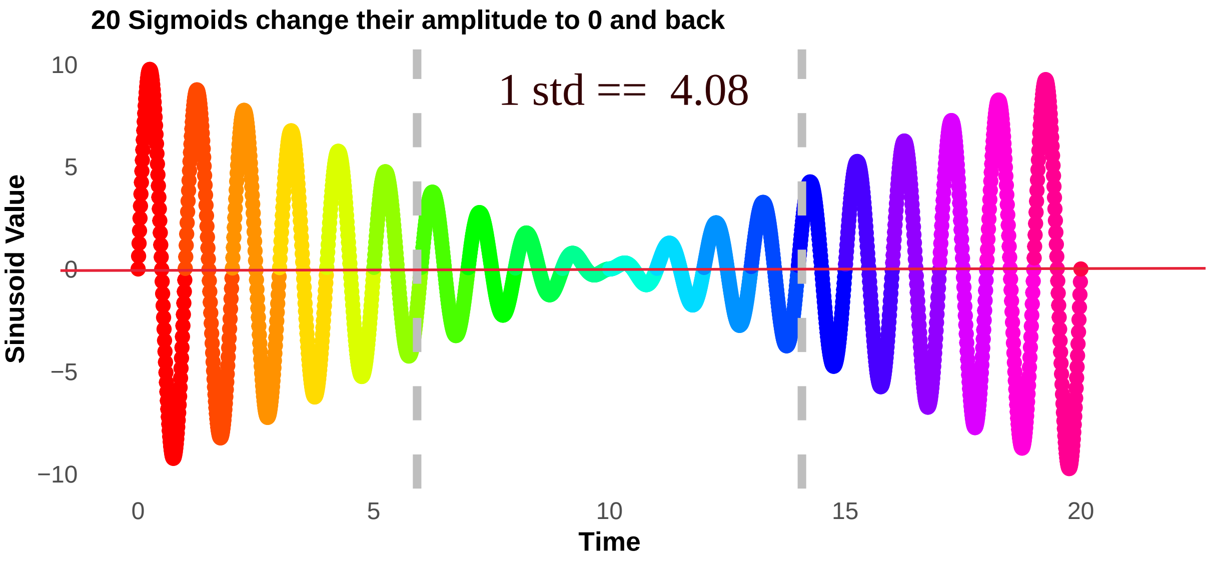
Data Generation for Sinusoidal Plot Replication

Generate a Dataset that corresponds to a sinusoidal function shown below. The Dataset should include the following attributes:

Replicating Visualization using ggplot2

The final plot should match with bottom one.

Note: There is no randomness involved in generating the dataset or creating the plot. Use the provided range of values for x (0 to 20), a total of 2000 data points, and the sinusoidal function to replicate the plot accurately.



Problem 7 (Python):

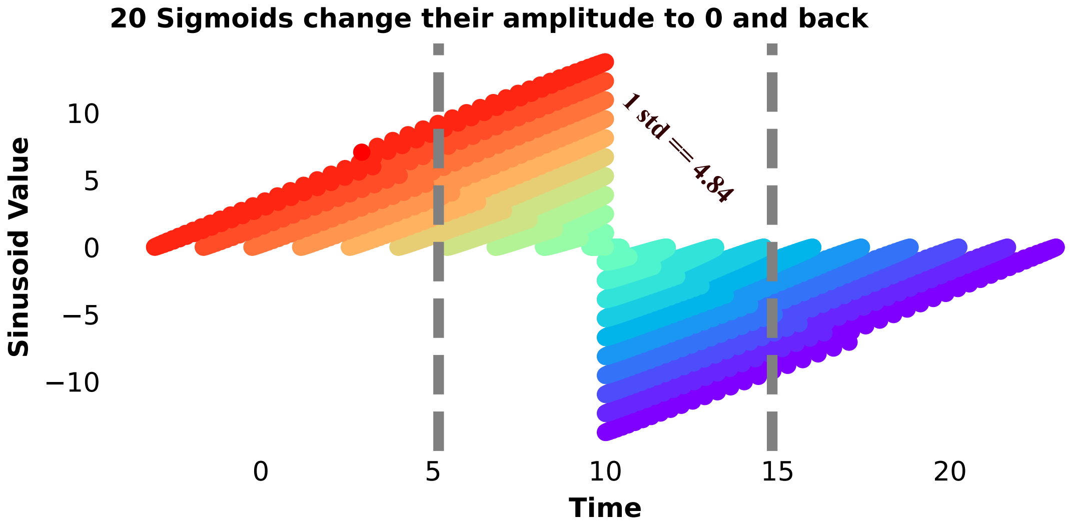
Data Generation for Sinusoidal Plot Replication in Python

Generate a dataset that corresponds to a sinusoidal function in Python.

Replicating Visualization in Python

The final plot should match with bottom one.

Note: There is no randomness involved in generating the dataset or creating the plot. Use the provided range of values for `x` (0 to 20), a total of 2000 data points, and the sinusoidal function to replicate the plot accurately.

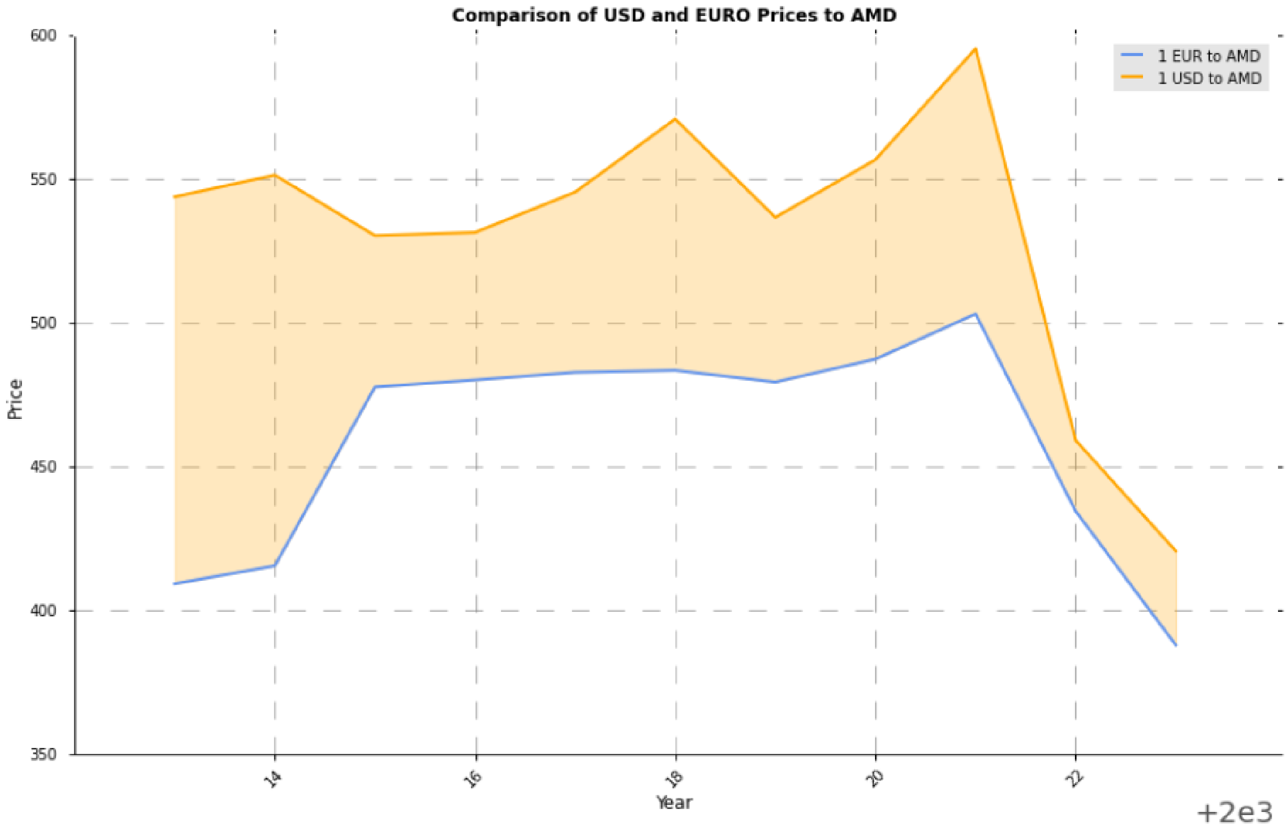


Problem 8 (Python)

The following task is an analysis of the AMD currency rate in exchange with USD and EURO from 2013 to 2023.

Visualisation

Reproduce the plot below

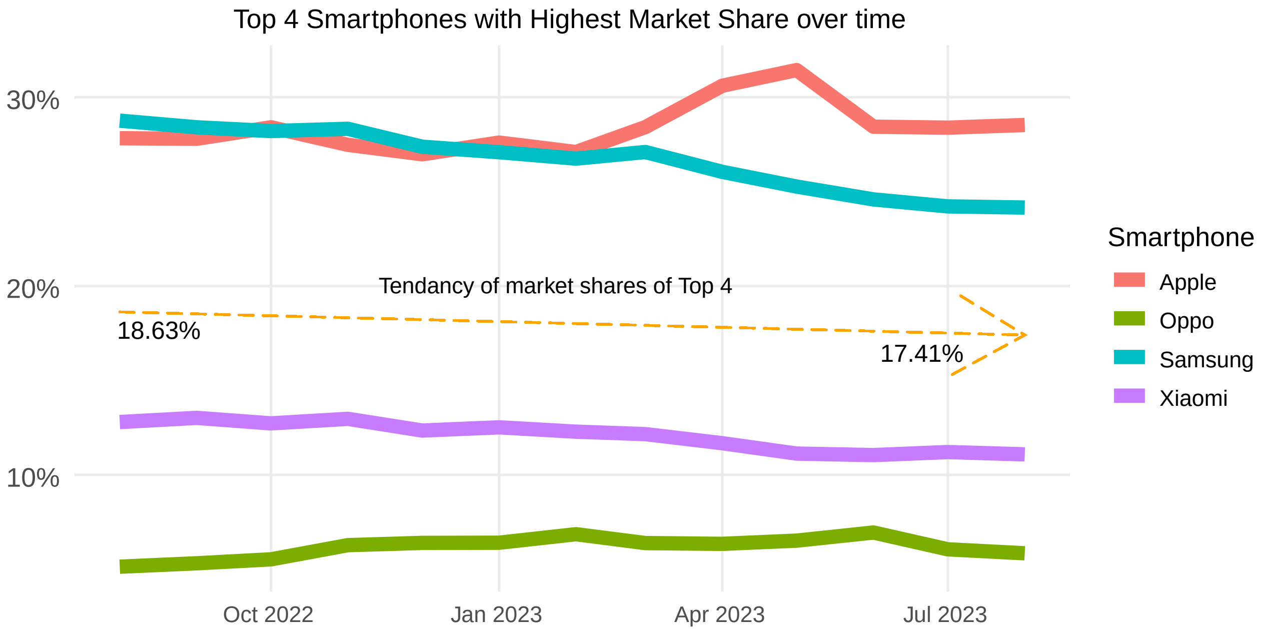


Problem 9 (R)

The dataset here is `MobileSmartphones.csv`.

```
df = read.csv("MobileSmartphones.csv")
```

Replicate the plot below


Top 4 Smartphones with Highest Market Share over time
Oct 2022
Jan 2023
Apr 2023
Jul 2023
Tendancy of market shares of Top 4
Smartphone
Apple
Oppo
Samsung
Xiaomi