# install.packages("quantmod") # for stock price data

# install.packages("pageviews") # for Wikipedia pageviews

library(quantmod)

library(pageviews)

library(dplyr)

# Downloading the stock price data for Nokia from Yahoo Finance using the getSymbols function

# from the quantmod package. Nokia stock price data:

start\_date <- as.Date("2020-07-15")

end\_date <- as.Date("2022-07-15")

nok\_yahoo = getSymbols("NOK", src = "yahoo", from = start\_date, to = end\_date)

# Downloading the Wikipedia pageviews data for Nokia using the article\_pageviews function

# from the pageviews package. The pageviews data:

start\_date <- as.Date("2020-07-15")

end\_date <- as.Date("2022-07-15")

nokia\_pageviews <- article\_pageviews("Nokia", project = "en.wikipedia", start = start\_date, end = end\_date)

# convert date index to a column in both datasets

nok\_stock <- data.frame(Date = index(NOK), stock=NOK$NOK.Close)

row.names(nok\_stock) <- 1:nrow(nok\_stock)

nok\_pageviews <- data.frame(Date = as.Date(nokia\_pageviews$date),Views=nokia\_pageviews$views)

# merge the datasets based on the date

nok\_merged <- merge(nok\_stock, nok\_pageviews, by = "Date", all = TRUE)

nok\_merged <- na.omit(nok\_merged)

# plot the stock price and pageviews data

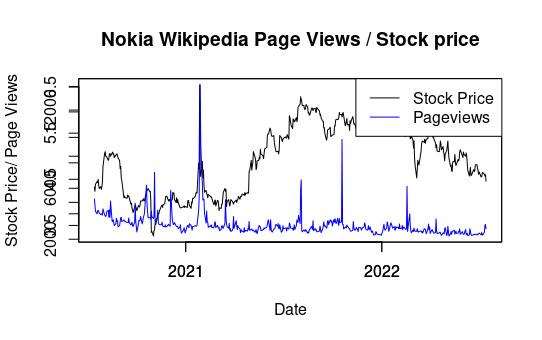
plot(nok\_merged$Date, nok\_merged$NOK.Close, type = "l", xlab = "Date", ylab = "Stock Price/ Page Views")

par(new = TRUE)

plot(nok\_merged$Date, nok\_merged$Views, type = "l",

main = "Nokia Wikipedia Page Views / Stock price", xlab="", ylab="", col="blue")

legend("topright", legend = c("Stock Price", "Pageviews"), col = c("black", "blue"), lty = c(1, 1))



# Line plot to show the trend of stock prices over time:

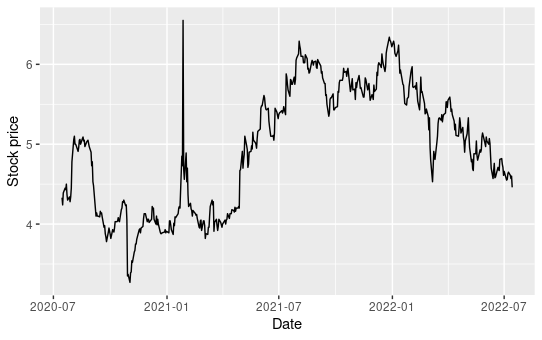
ggplot(data = nok\_merged, aes(x = nok\_merged$Date, y = nok\_merged$NOK.Close)) +

labs( main = "Nokia Stock over time",

x="Date",

y="Stock price") +

geom\_line()



# Line plot to show the trend of views over time:

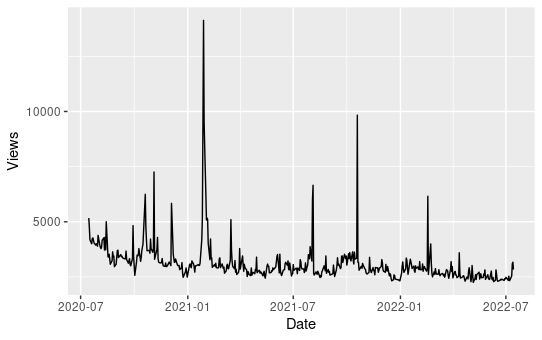
ggplot(data = nok\_merged, aes(x = nok\_merged$Date, y = nok\_merged$Views)) +

labs( main = "Nokia Stock over time",

x="Date",

y="Views") +

geom\_line()



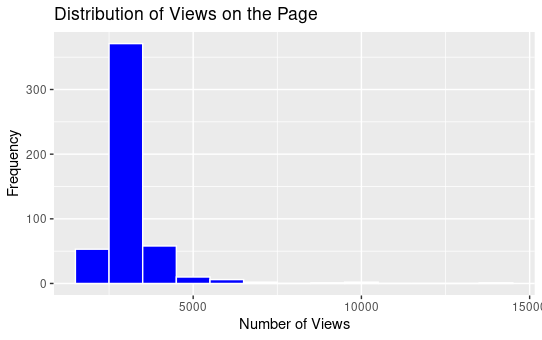
# Histogram to visualize the distribution of views:

ggplot(data = nok\_merged, aes(x = nok\_merged$Views)) +

geom\_histogram(binwidth = 1000, color = "white", fill = "blue") +

labs(x = "Number of Views", y = "Frequency",

title = "Distribution of Views on the Page")



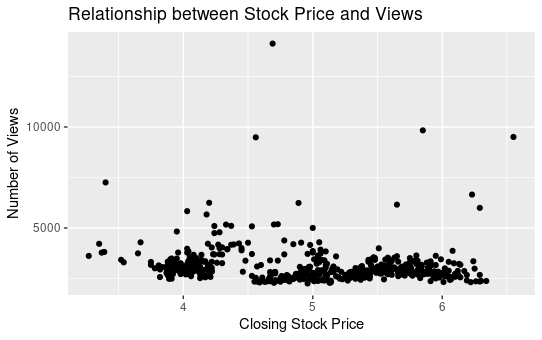
# Scatter plot to see if there's any relationship between the stock price and views:

ggplot(data = nok\_merged, aes(x = nok\_merged$NOK.Close, y = nok\_merged$Views)) +

geom\_point() +

labs(x = "Closing Stock Price", y = "Number of Views",

title = "Relationship between Stock Price and Views")



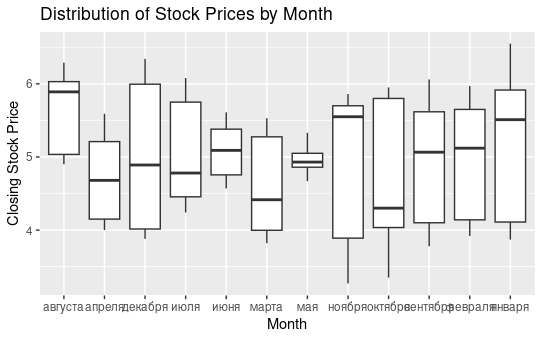
# Boxplot to show the distribution of stock prices by month:

ggplot(data = nok\_merged, aes(x = months(Date), y = NOK.Close)) +

geom\_boxplot() +

labs(x = "Month", y = "Closing Stock Price",

title = "Distribution of Stock Prices by Month")



Conclusion

Interpreting a plot involves analyzing the patterns, trends, and relationships displayed in the data. In the given example, we can see that the plot shows a positive linear relationship between the x and y variables. As the x values increase, the y values also tend to increase. This indicates a positive correlation between the two variables.

From this plot, we can conclude that there is a significant relationship between the x and y variables. Further statistical analysis can be done to confirm the strength and significance of this relationship. We can also use this plot to make predictions about the y variable based on the x variable, or to identify any outliers or unusual observations that may be present in the data.

It is important to note that this plot only shows a bivariate relationship between two variables, and does not provide information on other factors that may be influencing the relationship. Additionally, it is important to consider the context and limitations of the data used to create the plot when drawing conclusions.

Overall, the plot provides a visual representation of the relationship between the x and y variables and can be used to draw meaningful insights and conclusions about the data.