

Articles Over Time  
data$PublishDate <- as.Date(data$PublishDate)  
ggplot(data, aes(x=PublishDate)) +  
 geom\_histogram(bins=50, fill='orange', color='black') +  
 theme\_minimal() +  
 ggtitle("Number of Articles Over Time") +  
 xlab("Publish Date") +  
 ylab("Number of Articles")  
  
# 5. Correlation between Social Media Engagements  
social\_data <- data %>%  
 select(Facebook, GooglePlus, LinkedIn) %>%  
 filter(Facebook >= 0 & GooglePlus >= 0 & LinkedIn >= 0) # Remove invalid values  
  
cor\_matrix <- cor(social\_data)  
corrplot(cor\_matrix, method="circle")  
  
# 6. Linear Regression: Predict Facebook Shares based on Sentiment Scores  
# Remove rows with missing or invalid Facebook data  
data\_clean <- data %>%  
 filter(Facebook >= 0)  
  
# Fit the model  
model <- lm(Facebook ~ SentimentTitle + SentimentHeadline, data=data\_clean)  
summary(model)  
  
# Visualize the regression  
ggplot(data\_clean, aes(x=SentimentTitle, y=Facebook)) +  
 geom\_point() +  
 geom\_smooth(method="lm", col="red") +  
 theme\_minimal() +  
 ggtitle("Linear Regression: Facebook Shares vs SentimentTitle") +  
 xlab("Sentiment Title") +  
 ylab("Facebook Shares")  
  
ggplot(data\_clean, aes(x=SentimentHeadline, y=Facebook)) +  
 geom\_point() +  
 geom\_smooth(method="lm", col="blue") +  
 theme\_minimal() +  
 ggtitle("Linear Regression: Facebook Shares vs SentimentHeadline") +  
 xlab("Sentiment Headline") +  
 ylab("Facebook Shares")

library.packages <- c("ggplot2", "dplyr", "lubridate", "corrplot")  
  
# Load necessary libraries  
library(ggplot2)  
library(dplyr)  
library(lubridate)  
library(corrplot)  
  
# Load the dataset  
data <- read.csv("/content/News\_Final.csv")  
  
# 1. Missing Data Analysis  
missing\_data <- sapply(data, function(x) sum(is.na(x)))  
print(missing\_data)  
  
# 2. Distribution of Sentiment Scores (Title & Headline)  
ggplot(data, aes(x=SentimentTitle)) +  
 geom\_histogram(bins=30, fill='blue', color='black') +  
 theme\_minimal() +  
 ggtitle("Distribution of Sentiment Scores (Title)")  
  
ggplot(data, aes(x=SentimentHeadline)) +  
 geom\_histogram(bins=30, fill='green', color='black') +  
 theme\_minimal() +  
 ggtitle("Distribution of Sentiment Scores (Headline)")  
  
# 3. Number of Articles per Source  
data %>%  
 count(Source) %>%  
 arrange(desc(n)) %>%  
 ggplot(aes(x=reorder(Source, -n), y=n)) +  
 geom\_bar(stat="identity", fill='purple') +  
 coord\_flip() +  
 theme\_minimal() +  
 ggtitle("Number of Articles per Source") +  
 xlab("Source") +  
 ylab("Number of Articles")



Output









