Your Document Title

Varianza Italiana

Group members:

- Baldoni Chiara
- Pinna Paola
- Torella Marta

EDA

```r
data <- read.csv("FIN 36182-Industry_Portfolios.CSV")
head(data)
```

```
##     Date Cnsmr Manuf HiTec  Hlth Other
## 1 192607  5.43  2.73  1.83  1.77  2.13
## 2 192608  2.76  2.33  2.41  4.25  4.35
## 3 192609  2.16 -0.44  1.06  0.69  0.29
## 4 192610 -3.90 -2.42 -2.26 -0.57 -2.84
## 5 192611  3.70  2.50  3.07  5.42  2.11
## 6 192612  3.62  2.76  1.03  0.11  3.47
```

```r
sum(is.na(data))
```

```
## [1] 0
```

```r
anyDuplicated(data)
```

```
## [1] 0
```

```r
data[, 2:6] <- data[, 2:6] / 100 ## Convert returns to decimal form as asked
head(data)
```

```
##     Date   Cnsmr   Manuf   HiTec    Hlth   Other
## 1 192607  0.0543  0.0273  0.0183  0.0177  0.0213
## 2 192608  0.0276  0.0233  0.0241  0.0425  0.0435
## 3 192609  0.0216 -0.0044  0.0106  0.0069  0.0029
## 4 192610 -0.0390 -0.0242 -0.0226 -0.0057 -0.0284
## 5 192611  0.0370  0.0250  0.0307  0.0542  0.0211
## 6 192612  0.0362  0.0276  0.0103  0.0011  0.0347
```

Task 1

1. Report the arithmetic mean of the returns for each of the five industries over the entire sample

```r
monthly_means <- colMeans(data[, c("Cnsmr", "Manuf", "HiTec", "Hlth", "Other")])
annualized_means <- monthly_means * 12
annualized_means <- round(annualized_means, 4)
annualized_means
```

```
##  Cnsmr  Manuf  HiTec   Hlth  Other
## 0.1215 0.1152 0.1203 0.1288 0.1104
```

2. Report the standard deviation of the returns for each of the five industries over the entire sample

```r
monthly_std <- apply(data[, c("Cnsmr", "Manuf", "HiTec", "Hlth", "Other")], 2, sd)
annualized_std <- monthly_std * sqrt(12)
annualized_std <- round(annualized_std, 4)
annualized_std
```

```
##  Cnsmr  Manuf  HiTec   Hlth  Other
## 0.1828 0.1905 0.1935 0.1910 0.2211
```

3. Report the Sharpe ratio of each industry

```r
sharpe_ratios <- annualized_means / annualized_std
sharpe_ratios <- round(sharpe_ratios, 4)
sharpe_ratios
```

```
##  Cnsmr  Manuf  HiTec   Hlth  Other
## 0.6647 0.6047 0.6217 0.6743 0.4993
```

```r
results_table1 <- data.frame(
  Mean_Returns = annualized_means,
  Std_Returns = annualized_std,
  Sharpe_Ratio = sharpe_ratios
 )
results_table1
```

```
##       Mean_Returns Std_Returns Sharpe_Ratio
## Cnsmr       0.1215      0.1828       0.6647
## Manuf       0.1152      0.1905       0.6047
## HiTec       0.1203      0.1935       0.6217
## Hlth        0.1288      0.1910       0.6743
## Other       0.1104      0.2211       0.4993
```

4. Is there evidence that technology stocks have better risk-adjusted returns?

Based on the results, technology stocks (HiTec) have a Sharpe ratio of 0.6216, indicating better risk-adjusted returns compared to the manufacturing (0.6045) and other sectors (0.4994). However, they do not outperform the consumer (0.6646) and health industries (0.6746), which exhibit higher Sharpe ratios. Therefore, while technology stocks offer relatively good risk-adjusted returns, they are not the best performers overall, as both the consumer and health sectors deliver better risk-adjusted outcomes.

5. Provide a table (5×5) with the sample correlation between the returns of the five industries. Comment briefly.

```r
correlation_matrix <- cor(data[, c("Cnsmr", "Manuf", "HiTec", "Hlth", "Other")])
correlation_matrix <- round(correlation_matrix, 4)
correlation_matrix
```

```
##        Cnsmr  Manuf  HiTec   Hlth  Other
## Cnsmr 1.0000 0.8670 0.8164 0.7739 0.8716
## Manuf 0.8670 1.0000 0.7996 0.7427 0.8924
```

```
## HiTec 0.8164 0.7996 1.0000 0.7072 0.7932
## Hlth  0.7739 0.7427 0.7072 1.0000 0.7371
## Other 0.8716 0.8924 0.7932 0.7371 1.0000
```

```r
library(knitr)
kable(correlation_matrix, caption = "Correlation Matrix of Industry Returns", format = "html")
```

Correlation Matrix of Industry Returns

Cnsmr

Manuf

HiTec

Hlth

Other

Cnsmr

1.0000

0.8670

0.8164

0.7739

0.8716

Manuf

0.8670

1.0000

0.7996

0.7427

0.8924

HiTec

0.8164

0.7996

1.0000

0.7072

0.7932

Hlth

0.7739

0.7427

0.7072

1.0000

0.7371

Other

0.8716

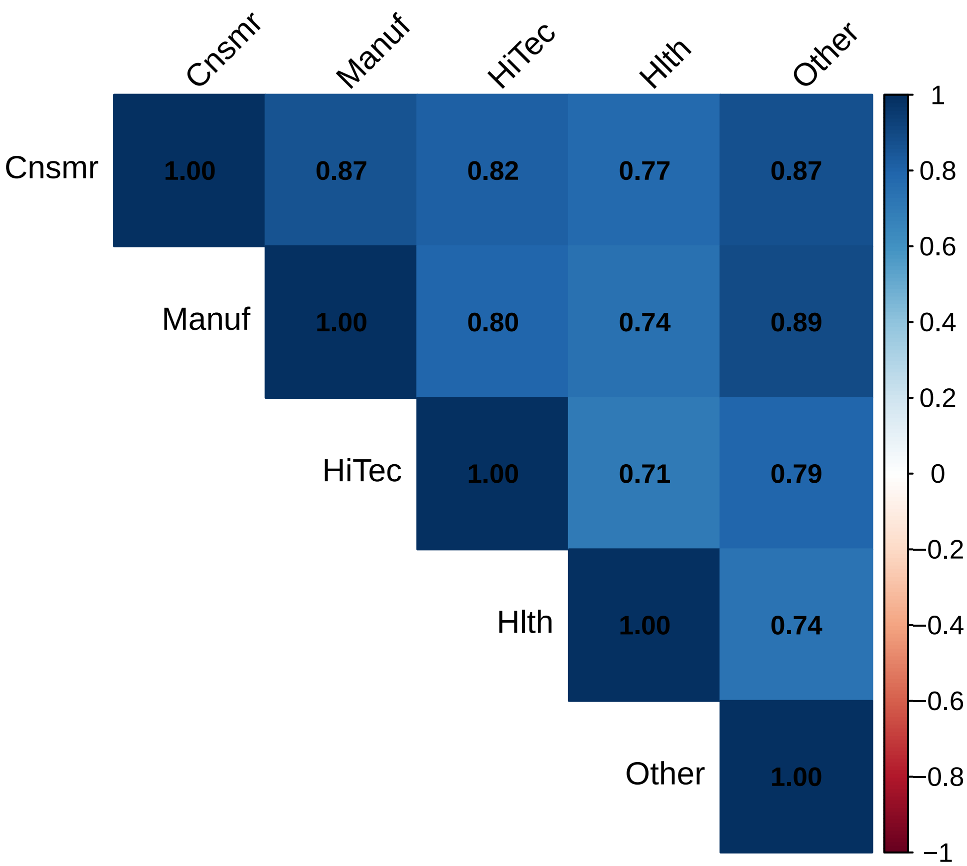
0.8924

0.7932

0.7371

1.0000

```r
library(corrplot)
corrplot(correlation_matrix, method = "color", type = "upper",
         tl.col = "black", tl.srt = 45,
         addCoef.col = "black",
         number.cex = 0.8,
         number.digits = 2)
```



The correlation matrix indicates that all five industries have positive correlations with each other, suggesting that their returns tend to move in the same direction. The strongest correlations are between Manufacturing and Other (0.8924) and between Manufacturing and Consumer (0.8670), indicating that these industries have highly synchronized return movements. On the other hand, the weakest correlations are between Technology and Health (0.7072), and between these two industries and the others, implying that Technology and Health exhibit slightly more independent return patterns. While the positive correlations across all industries suggest some degree of co-movement, the relatively lower correlations for Technology and Health indicate potential diversification benefits, though the absence of negative correlations limits the extent of diversification.

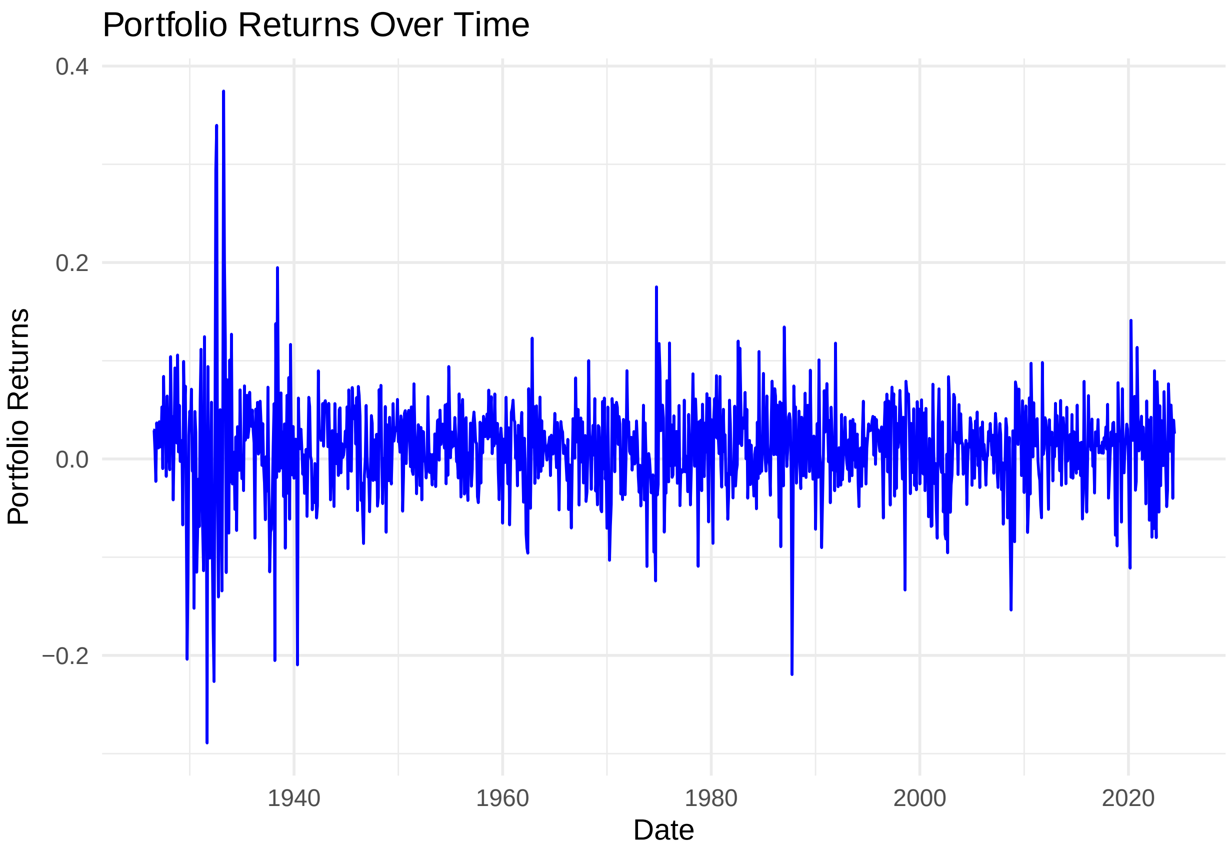
6. Construct a time series of the simple, non-cumulative returns of a portfolio where capital is allocated equally across the first four industries (excluding Other). Report the arithmetic mean, standard deviation and Sharpe. Comment briefly on the gains achieved by this diversified portfolio.

```
portfolio_returns <- rowMeans(data[, c("Cnsmr", "Manuf", "HiTec", "Hlth")])
data$Portfolio_Returns <- portfolio_returns
mean_portfolio <- mean(portfolio_returns)
sd_portfolio <- sd(portfolio_returns)

annualized_mean_portfolio <- mean_portfolio * 12
annualized_sd_portfolio <- sd_portfolio * sqrt(12)
sharpe_ratio_portfolio <- annualized_mean_portfolio / annualized_sd_portfolio

portfolio_results <- data.frame(
  Metric = c("Annualized Mean Return", "Annualized Standard Deviation", "Sharpe Ratio"),
  Value = round(c(annualized_mean_portfolio, annualized_sd_portfolio, sharpe_ratio_portfolio), 4)
)
print(portfolio_results)
```

```
##                          Metric  Value
## 1        Annualized Mean Return 0.1214
## 2 Annualized Standard Deviation 0.1734
## 3                  Sharpe Ratio 0.7004
```

The diversified portfolio, which allocates capital equally across the first four industries, achieves an annualized mean return of 12.14% and an annualized standard deviation of 17.34%, resulting in a Sharpe ratio of 0.7004. This Sharpe ratio indicates that the portfolio offers better risk-adjusted returns than some individual industries, such as Manufacturing (0.6045) and Other (0.4994). It also slightly outperforms Technology (0.6216), suggesting that diversification has improved the portfolio's performance relative to risk. However, the gains from diversification are not substantial, as the portfolio's Sharpe ratio is only marginally higher than that of Health (0.6746) and Consumer (0.6646). Overall, diversification provides some benefit, but the improvement in risk-adjusted returns is moderate.

```
library(ggplot2)
data$Date <- as.character(data$Date)
data$Date <- paste0(data$Date, "01")
data$Date <- as.Date(data$Date, format = "%Y%m%d")

plot_data <- data.frame(Date = data$Date, Portfolio_Returns = data$Portfolio_Returns)
ggplot(plot_data, aes(x = Date, y = Portfolio_Returns)) +
  geom_line(color = "blue") +
  labs(title = "Portfolio Returns Over Time",
       x = "Date",
       y = "Portfolio Returns") +
  theme_minimal()
```



The portfolio returns exhibit significant volatility, especially in the earlier periods before the 1950s, with large spikes in both positive and negative directions. This high volatility is particularly evident around historical market crashes, such as the Great Depression in the 1930s and the 2008 financial crisis. After World War II, the portfolio returns become more stable, with fewer extreme fluctuations, though volatility remains, particularly during financial crises. In the last two decades, there has been a noticeable increase in volatility, likely reflecting events like the dot-com bubble, the 2008 crisis, and the COVID-19 market disruptions. Overall, while the portfolio appears more stable post-1950, significant market events continue to cause sharp fluctuations.

Task 2

In this task you will treat the portfolio you computed in Task 1, point (6), as the market portfolio, denote its returns as Rm, and will estimate and interpret beta and alpha coefficients in the context of the CAPM.

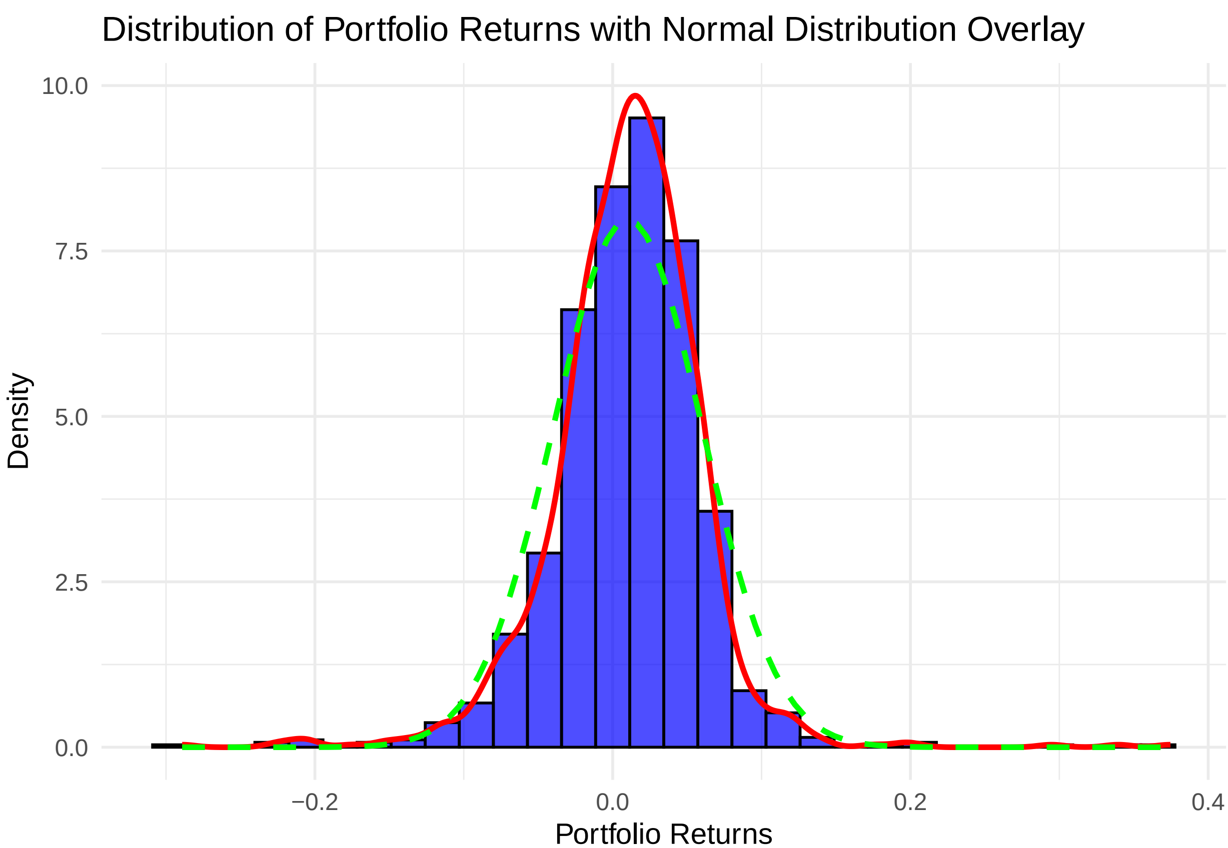
1. Compute the kurtosis and skeweness of Rm

```r
library(moments)

skewness_value <- skewness(data$Portfolio_Returns, na.rm = TRUE)
kurtosis_value <- kurtosis(data$Portfolio_Returns, na.rm = TRUE)

kurtosis_value <- round(kurtosis_value, 4)
skewness_value <- round(skewness_value, 4)

portfolio_results2 <- data.frame(
  Metric = c("Kurtosis", "Skewness"),
  Value = c(kurtosis_value, skewness_value)
```

```r
)
print(portfolio_results2)
```

```
##     Metric   Value
## 1 Kurtosis 10.2856
## 2 Skewness  0.0237
```

```r
library(ggplot2)
ggplot(data = data.frame(returns = data$Portfolio_Returns), aes(x = returns)) +
  geom_histogram(aes(y = ..density..), bins = 30, fill = "blue", color = "black", alpha = 0.7) +
  geom_density(color = "red", size = 1) +
  stat_function(fun = dnorm,
                args = list(mean = mean(portfolio_returns, na.rm = TRUE),
                            sd = sd(portfolio_returns, na.rm = TRUE)),
                color = "green", linetype = "dashed", size = 1) +
  labs(title = "Distribution of Portfolio Returns with Normal Distribution Overlay",
       x = "Portfolio Returns",
       y = "Density") +
  theme_minimal()
```



2. How do the values in (1) compare with the normal distribution?

The portfolio returns exhibit a kurtosis of 10.2856, which is significantly higher than the normal distribution's kurtosis of 3, indicating the presence of fat tails. This suggests that the portfolio is more likely to experience extreme returns, both positive and negative, compared to a normal distribution. In contrast, the skewness of 0.0237 is very close to zero, implying that the distribution of returns is symmetric, with no significant bias toward either side. Overall, while the portfolio returns are prone to extreme fluctuations, these events are equally likely to be positive or negative, reflecting a balance in the distribution's shape.

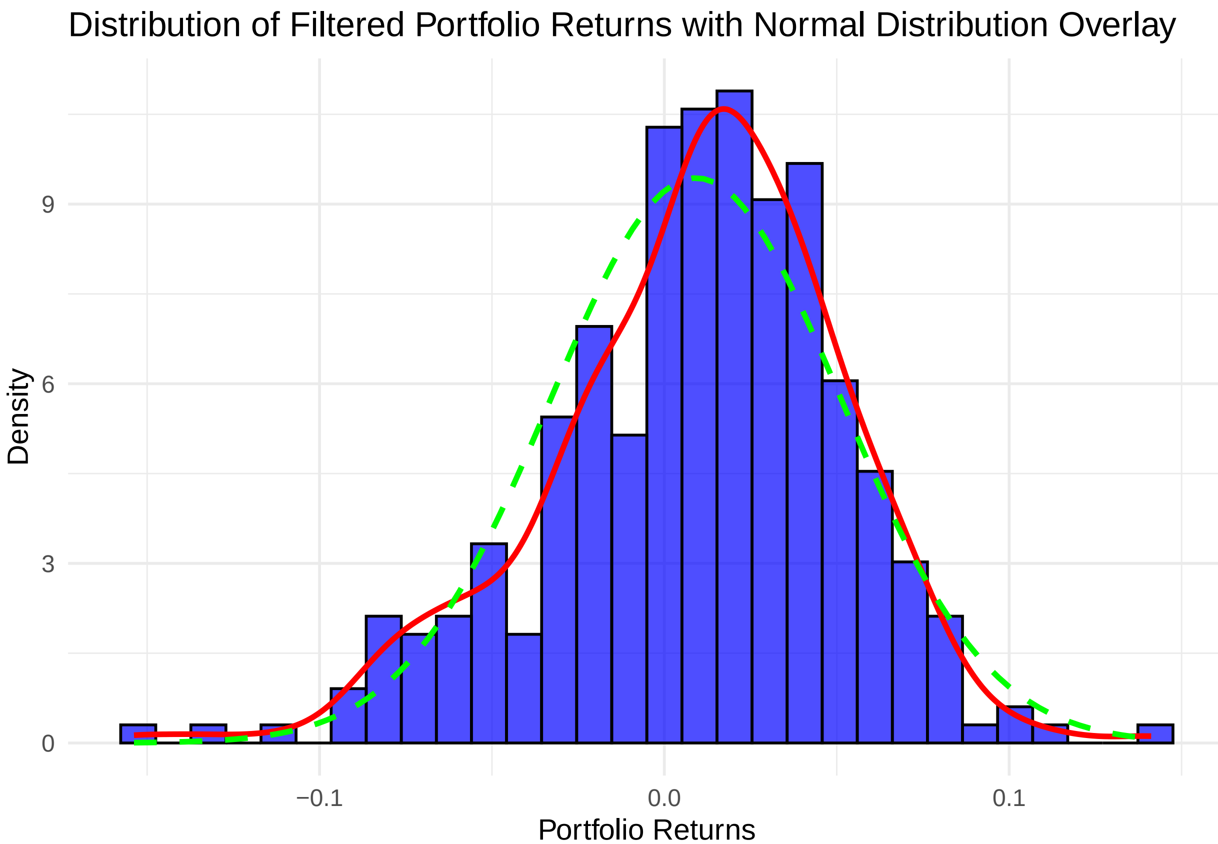
Moreover, the above plot shows the distribution of portfolio returns (blue histogram with red density curve) overlaid with a normal distribution (green dashed line). The portfolio returns exhibit fatter tails compared to the normal distribution, particularly on the left side, which is consistent with the high kurtosis observed earlier, indicating a higher probability of extreme returns. The distribution is also roughly symmetric, aligning with the nearly zero skewness, showing no significant bias toward either positive or negative returns. Additionally, the portfolio's return distribution has a slightly higher peak than the normal curve, suggesting a greater concentration of returns near the mean. Overall, the portfolio returns deviate from the normal distribution, with a greater chance of extreme events and a higher central concentration of values.

3. Repeat point (1), but eliminating the first 70 years of data (i.e. from 199706).

```r
data_filtered <- data[data$Date >= as.Date("1997-06-01"), ]
portfolio_returns_filtered <- data_filtered$Portfolio_Returns

skewness_filtered <- skewness(portfolio_returns_filtered, na.rm = TRUE)
kurtosis_filtered <- kurtosis(portfolio_returns_filtered, na.rm = TRUE)

skewness_filtered <- round(skewness_filtered, 4)
kurtosis_filtered <- round(kurtosis_filtered, 4)

portfolio_results3 <- data.frame(
  Metric = c("Kurtosis", "Skewness"),
  Value = c(kurtosis_filtered, skewness_filtered)
)
print(portfolio_results3)
```

```
##     Metric   Value
## 1 Kurtosis  3.7380
## 2 Skewness -0.5123
```

```r
ggplot(data = data.frame(returns = portfolio_returns_filtered), aes(x = returns)) +
  geom_histogram(aes(y = after_stat(density)), bins = 30, fill = "blue", color = "black", alpha = 0.7) +
  geom_density(color = "red", linewidth = 1) +
  stat_function(fun = dnorm,
                args = list(mean = mean(portfolio_returns_filtered, na.rm = TRUE),
                            sd = sd(portfolio_returns_filtered, na.rm = TRUE)),
                color = "green", linetype = "dashed", linewidth = 1) +
  labs(title = "Distribution of Filtered Portfolio Returns with Normal Distribution Overlay",
       x = "Portfolio Returns",
       y = "Density") +
  theme_minimal()
```



4. Compute and report the covariance of the first four industries with Rm

```r
cov_Cnsmr_Rm <- cov(data$Cnsmr, data$Portfolio_Returns, use = "complete.obs")
cov_Manuf_Rm <- cov(data$Manuf, data$Portfolio_Returns, use = "complete.obs")
cov_HiTec_Rm <- cov(data$HiTec, data$Portfolio_Returns, use = "complete.obs")
cov_Hlth_Rm <- cov(data$Hlth, data$Portfolio_Returns, use = "complete.obs")

cov_Cnsmr_Rm <- round(cov_Cnsmr_Rm, 4)
cov_Manuf_Rm <- round(cov_Manuf_Rm, 4)
cov_HiTec_Rm <- round(cov_HiTec_Rm, 4)
cov_Hlth_Rm <- round(cov_Hlth_Rm, 4)

covariances <- data.frame(
  Industry = c("Cnsmr", "Manuf", "HiTec", "Hlth"),
  Covariance_with_Rm = c(cov_Cnsmr_Rm, cov_Manuf_Rm, cov_HiTec_Rm, cov_Hlth_Rm)
)

print(covariances)
```

```
##   Industry Covariance_with_Rm
## 1    Cnsmr             0.0025
## 2    Manuf             0.0026
## 3    HiTec             0.0025
## 4     Hlth             0.0024
```

5. Use the results obtained so far to compute the beta values for the first four industries for the full sample

```
# calculate the variance of the portfolio returns (Rm)
var_Rm <- var(data$Portfolio_Returns, use = "complete.obs")
var_Rm <- round(var_Rm, 4)

# calculate the beta values for each industry
beta_Cnsmr <- cov_Cnsmr_Rm / var_Rm
beta_Manuf <- cov_Manuf_Rm / var_Rm
beta_HiTec <- cov_HiTec_Rm / var_Rm
beta_Hlth <- cov_Hlth_Rm / var_Rm

beta_Cnsmr <- round(beta_Cnsmr, 4)
beta_Manuf <- round(beta_Manuf, 4)
beta_HiTec <- round(beta_HiTec, 4)
beta_Hlth <- round(beta_Hlth, 4)

beta_values <- data.frame(
  Industry = c("Cnsmr", "Manuf", "HiTec", "Hlth"),
  Beta = c(beta_Cnsmr, beta_Manuf, beta_HiTec, beta_Hlth)
)
print(beta_values)
```

```
##   Industry Beta
## 1    Cnsmr 1.00
## 2    Manuf 1.04
## 3    HiTec 1.00
## 4     Hlth 0.96
```

6. Compute the beta values for the first four industries for the sample starting from 199706. Briefly comment on how results compare with those in point (5)

```
var_Rm_filtered <- var(portfolio_returns_filtered, use = "complete.obs")

beta_Cnsmr_filtered <- cov(data_filtered$Cnsmr, data_filtered$Portfolio_Returns, use = "complete.obs") /
beta_Manuf_filtered <- cov(data_filtered$Manuf, data_filtered$Portfolio_Returns, use = "complete.obs") /
beta_HiTec_filtered <- cov(data_filtered$HiTec, data_filtered$Portfolio_Returns, use = "complete.obs") /
beta_Hlth_filtered <- cov(data_filtered$Hlth, data_filtered$Portfolio_Returns, use = "complete.obs") / v

beta_Cnsmr_filtered <- round(beta_Cnsmr_filtered, 4)
beta_Manuf_filtered <- round(beta_Manuf_filtered, 4)
beta_HiTec_filtered <- round(beta_HiTec_filtered, 4)
beta_Hlth_filtered <- round(beta_Hlth_filtered, 4)

beta_values_filtered <- data.frame(
  Industry = c("Cnsmr", "Manuf", "HiTec", "Hlth"),
  Beta_Filtered = c(beta_Cnsmr_filtered, beta_Manuf_filtered, beta_HiTec_filtered, beta_Hlth_filtered)
)
print(beta_values_filtered)
```

```
##   Industry Beta_Filtered
## 1    Cnsmr        0.9236
## 2    Manuf        0.9533
```

```
## 3    HiTec        1.3305
## 4     Hlth        0.7926
```

```
comparison <- data.frame(
  Industry = c("Cnsmr", "Manuf", "HiTec", "Hlth"),
  Beta_Full_Sample = c(beta_Cnsmr, beta_Manuf, beta_HiTec, beta_Hlth),  # point 5
  Beta_Filtered = c(beta_Cnsmr_filtered, beta_Manuf_filtered, beta_HiTec_filtered, beta_Hlth_filtered)
)
print(comparison)
```

```
##   Industry Beta_Full_Sample Beta_Filtered
## 1    Cnsmr             1.00        0.9236
## 2    Manuf             1.04        0.9533
## 3    HiTec             1.00        1.3305
## 4     Hlth             0.96        0.7926
```

The comparison of beta values between the full sample and the post-June 1997 period shows some notable changes. For Consumer stocks, the beta decreases from 1.00 in the full sample to 0.9236 in the filtered period, indicating slightly less sensitivity to the market. Similarly, Manufacturing shows a reduction in beta from 1.04 to 0.9533, suggesting that its returns are less volatile relative to the market in the more recent period. In contrast, Technology stocks see a significant increase in beta from 1.00 to 1.3305, implying a stronger reaction to market movements in the post-1997 period. Lastly, Health stocks exhibit a lower beta, decreasing from 0.96 to 0.7926, indicating that they have become less sensitive to market fluctuations. Overall, the results suggest that Technology has become more volatile relative to the market, while the other industries show reduced sensitivity in the more recent period.

7. Assuming a risk-free rate of 5%, compute Jensen's alpha for each of the first four industries (on the full sample). Report the alpha in percentage terms. Briefly discuss your results

```
annualized_returns_industries <- annualized_means
risk_free_rate <- 0.05

alpha_Cnsmr <- (annualized_returns_industries["Cnsmr"] - (risk_free_rate + beta_Cnsmr * (annualized_mea
alpha_Manuf <- (annualized_returns_industries["Manuf"] - (risk_free_rate + beta_Manuf * (annualized_mea
alpha_HiTec <- (annualized_returns_industries["HiTec"] - (risk_free_rate + beta_HiTec * (annualized_mea
alpha_Hlth <- (annualized_returns_industries["Hlth"] - (risk_free_rate + beta_Hlth * (annualized_mean_p

jensens_alpha <- data.frame(
  Industry = c("Cnsmr", "Manuf", "HiTec", "Hlth"),
  Alpha = round(c(alpha_Cnsmr, alpha_Manuf, alpha_HiTec, alpha_Hlth), 4)
)
print(jensens_alpha)
```

```
##       Industry   Alpha
## Cnsmr    Cnsmr  0.0051
## Manuf    Manuf -0.9107
## HiTec    HiTec -0.1149
## Hlth      Hlth  1.0209
```

The Jensen's alpha results, reported in percentage terms, show varying performance across the four industries. Consumer stocks have an alpha of 0.0051%, indicating performance roughly in line with expectations under the CAPM model. Manufacturing stocks, on the other hand, have a negative alpha of -0.9107%, meaning they significantly underperformed relative to what the CAPM would predict. Similarly, Technology stocks exhibit a negative alpha of -0.1149%, suggesting a slight underperformance. In contrast, Health stocks stand out with a positive alpha of 1.0209%, indicating that this sector outperformed its expected return by over 1%, given its risk and the market conditions. Overall, the Health sector performed well above expectations, while Manufacturing and Technology stocks underperformed.

Task 3

Use the lm( ) function to run a few regressions:

1. Regress Rm(t) on an intercept and on Rm(t−1). Report estimates and t-statistics. Briefly interpret the results

```r
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
```

```r
data <- data %>%
  mutate(Rm_lag = lag(Portfolio_Returns, 1))
head(data)
```

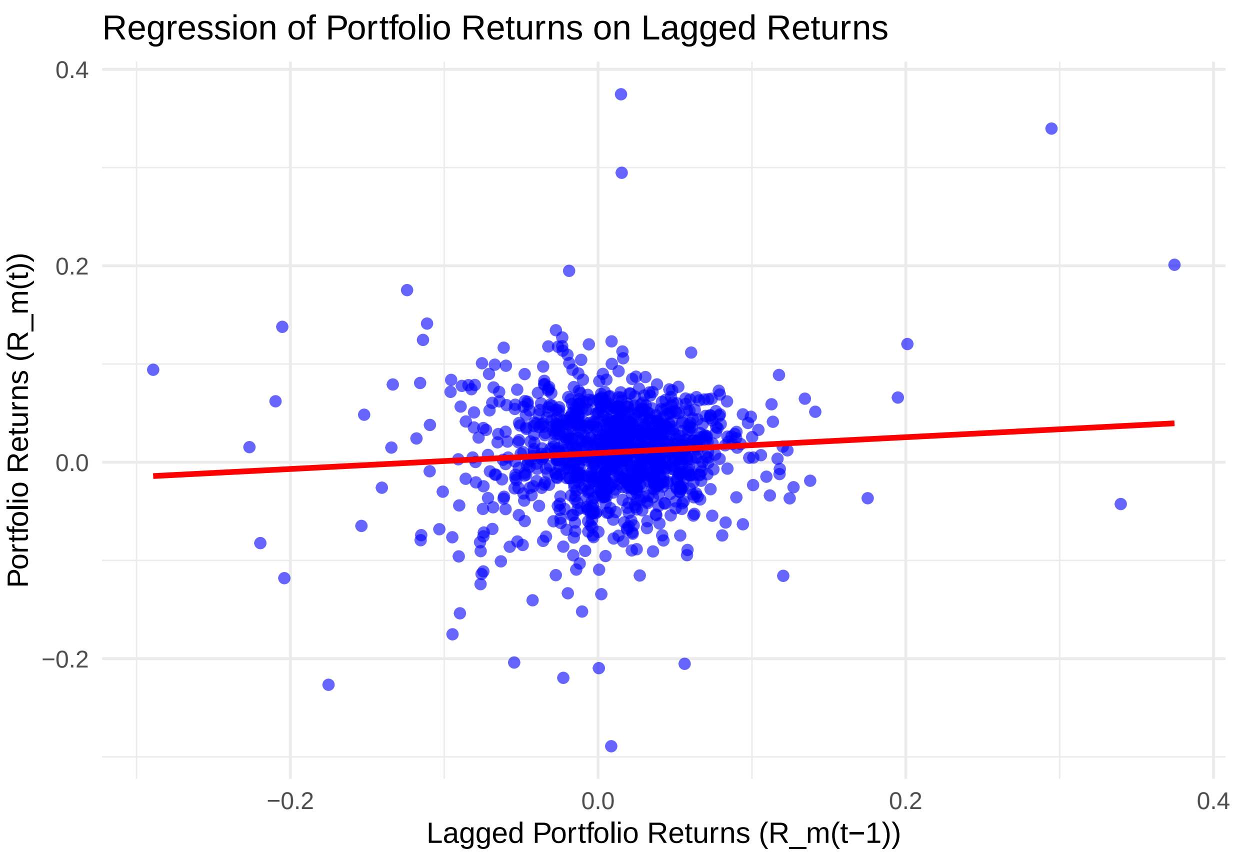
```
##         Date   Cnsmr   Manuf   HiTec    Hlth   Other Portfolio_Returns
## 1 1926-07-01  0.0543  0.0273  0.0183  0.0177  0.0213          0.029400
## 2 1926-08-01  0.0276  0.0233  0.0241  0.0425  0.0435          0.029375
## 3 1926-09-01  0.0216 -0.0044  0.0106  0.0069  0.0029          0.008675
## 4 1926-10-01 -0.0390 -0.0242 -0.0226 -0.0057 -0.0284         -0.022875
## 5 1926-11-01  0.0370  0.0250  0.0307  0.0542  0.0211          0.036725
## 6 1926-12-01  0.0362  0.0276  0.0103  0.0011  0.0347          0.018800
##      Rm_lag
## 1        NA
## 2  0.029400
## 3  0.029375
## 4  0.008675
## 5 -0.022875
## 6  0.036725
```

```r
model <- lm(Portfolio_Returns ~ Rm_lag, data = data)
summary(model)
```

```
##
## Call:
## lm(formula = Portfolio_Returns ~ Rm_lag, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29915 -0.02596  0.00175  0.02912  0.36413
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.009287   0.001486   6.249 5.76e-10 ***
## Rm_lag      0.080911   0.029101   2.780  0.00552 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04993 on 1173 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.006547,    Adjusted R-squared:  0.0057
## F-statistic:  7.73 on 1 and 1173 DF,  p-value: 0.005518
```

```r
library(ggplot2)
ggplot(data, aes(x = Rm_lag, y = Portfolio_Returns)) +
  geom_point(color = "blue", alpha = 0.6) +  # Scatter plot of returns
  geom_smooth(method = "lm", color = "red", se = FALSE) +  # Regression line without confidence interva
  labs(title = "Regression of Portfolio Returns on Lagged Returns",
       x = "Lagged Portfolio Returns (R_m(t-1))",
       y = "Portfolio Returns (R_m(t))") +
  theme_minimal()
```



The regression of current portfolio returns Rm(t) on lagged returns Rm(t-1) shows a small but significant positive relationship. The coefficient for the lagged returns is 0.0809, indicating that for every 1-unit increase in the lagged returns, the current returns increase by about 0.0809 units. This coefficient is statistically significant (p-value = 0.00552), meaning that the relationship between lagged and current returns is unlikely to be due to chance. However, the R-squared value is very low, at only 0.65%, suggesting that the lagged returns explain only a tiny fraction of the variation in current returns. Most of the variability in the portfolio returns is not captured by the lagged returns, indicating that other factors or random fluctuations have a much larger influence. The intercept is 0.009287 and is highly significant (p-value = 5.76 times 10^{-10}), suggesting that even when the lagged returns are zero, the portfolio tends to produce a small positive return of about 0.93%. While the lagged returns do show some predictive power, their overall effect on current returns is weak, as evidenced by the low explanatory power of the model.

2. Regress Rm(t) on an intercept and on (Pm(t−1)/Pm(t−13)−1). Report estimates and t-statistics. Briefly interpret the results

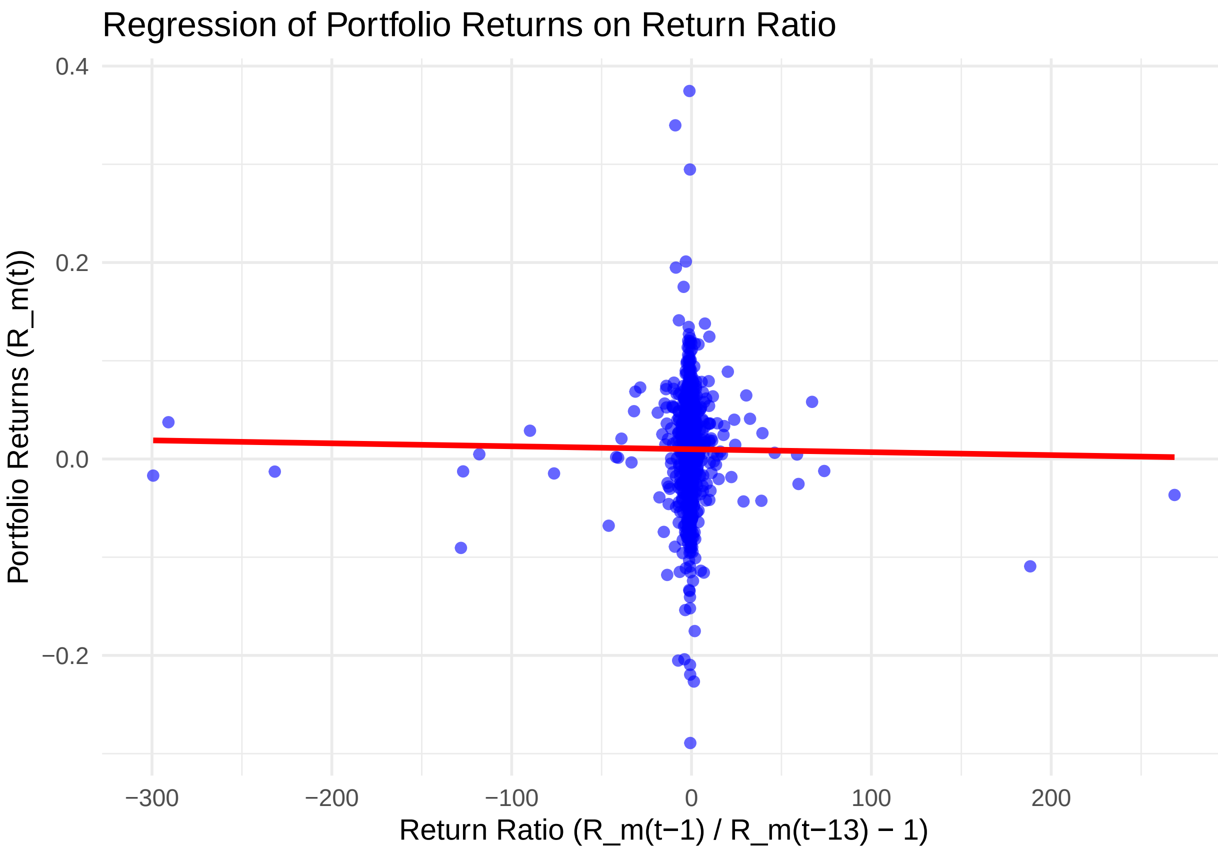
```
data <- data %>%
  mutate(
    Rm_lag13 = lag(Portfolio_Returns, 13),
    Return_Ratio = (Rm_lag / Rm_lag13) - 1
  )
head(data,20)
```

```
##          Date   Cnsmr   Manuf   HiTec    Hlth   Other Portfolio_Returns
## 1  1926-07-01  0.0543  0.0273  0.0183  0.0177  0.0213          0.029400
## 2  1926-08-01  0.0276  0.0233  0.0241  0.0425  0.0435          0.029375
## 3  1926-09-01  0.0216 -0.0044  0.0106  0.0069  0.0029          0.008675
## 4  1926-10-01 -0.0390 -0.0242 -0.0226 -0.0057 -0.0284         -0.022875
## 5  1926-11-01  0.0370  0.0250  0.0307  0.0542  0.0211          0.036725
## 6  1926-12-01  0.0362  0.0276  0.0103  0.0011  0.0347          0.018800
## 7  1927-01-01 -0.0119  0.0015  0.0046  0.0505  0.0150          0.011175
## 8  1927-02-01  0.0528  0.0400  0.0419  0.0171  0.0505          0.037950
## 9  1927-03-01  0.0164 -0.0143  0.0365  0.0101  0.0122          0.012175
## 10 1927-04-01  0.0352 -0.0113  0.0135  0.0274  0.0083          0.016200
## 11 1927-05-01  0.0609  0.0567  0.0528  0.0412  0.0654          0.052900
## 12 1927-06-01 -0.0195 -0.0303  0.0051  0.0054 -0.0215         -0.009825
## 13 1927-07-01  0.0869  0.0751  0.0759  0.0984  0.0604          0.084075
## 14 1927-08-01  0.0523  0.0197  0.0294  0.0028 -0.0103          0.026050
## 15 1927-09-01  0.0589  0.0490  0.0393  0.0565  0.0459          0.050925
## 16 1927-10-01 -0.0300 -0.0470 -0.0454  0.0513 -0.0382         -0.017775
## 17 1927-11-01  0.0774  0.0770  0.0652  0.0368  0.0450          0.064100
## 18 1927-12-01  0.0360  0.0220  0.0128 -0.0046  0.0196          0.016550
## 19 1928-01-01 -0.0121  0.0018 -0.0003  0.0269 -0.0101          0.004075
## 20 1928-02-01 -0.0153 -0.0111 -0.0038 -0.0138 -0.0202         -0.011000
##       Rm_lag  Rm_lag13 Return_Ratio
## 1         NA        NA           NA
## 2   0.029400        NA           NA
## 3   0.029375        NA           NA
## 4   0.008675        NA           NA
## 5  -0.022875        NA           NA
## 6   0.036725        NA           NA
## 7   0.018800        NA           NA
## 8   0.011175        NA           NA
## 9   0.037950        NA           NA
## 10  0.012175        NA           NA
## 11  0.016200        NA           NA
## 12  0.052900        NA           NA
## 13 -0.009825        NA           NA
## 14  0.084075  0.029400    1.8596939
## 15  0.026050  0.029375   -0.1131915
## 16  0.050925  0.008675    4.8703170
## 17 -0.017775 -0.022875   -0.2229508
## 18  0.064100  0.036725    0.7454050
## 19  0.016550  0.018800   -0.1196809
## 20  0.004075  0.011175   -0.6353468
```

```
model_2 <- lm(Portfolio_Returns ~ Return_Ratio, data = data)
summary(model_2)
```

```
##
## Call:
## lm(formula = Portfolio_Returns ~ Return_Ratio, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29913 -0.02583  0.00292  0.02881  0.36466
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.931e-03  1.477e-03   6.723 2.79e-11 ***
## Return_Ratio -3.002e-05  7.519e-05  -0.399     0.69
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05026 on 1161 degrees of freedom
##   (13 observations deleted due to missingness)
## Multiple R-squared:  0.0001372,  Adjusted R-squared:  -0.000724
## F-statistic: 0.1593 on 1 and 1161 DF,  p-value: 0.6898
```

```r
ggplot(data, aes(x = Return_Ratio, y = Portfolio_Returns)) +
  geom_point(color = "blue", alpha = 0.6) +
  geom_smooth(method = "lm", color = "red", se = FALSE) +
  labs(title = "Regression of Portfolio Returns on Return Ratio",
       x = "Return Ratio (R_m(t-1) / R_m(t-13) - 1)",
       y = "Portfolio Returns (R_m(t))") +
  theme_minimal()
```



The regression of portfolio returns Rm(t) on the return ratio (Rm(t−1)/Rm(t−13)−1) shows that the intercept is 0.99%, which is statistically significant with a t-value of 6.723 and a p-value of 0.69, indicating a positive baseline return when the return ratio is zero. However, the coefficient for the return ratio is -0.00003002, with a t-value of -0.399 and a p-value of 0.69, suggesting that the return ratio is not a significant predictor of current portfolio returns. The R-squared value is extremely low at 0.0001372, indicating that the return ratio explains only 0.01% of the variation in returns, and the adjusted R-squared is negative, further confirming the poor fit of the model. The F-statistic of 0.1593 and its p-value of 0.69 also indicate that the overall model is not statistically significant. In conclusion, the return ratio provides no meaningful explanatory power for current portfolio returns, and other factors likely drive the variation in returns.

3. Are the t-statistics reported by lm( ) in (2) reliable? Explain

Based on the results of your regression model where Rm(t) is regressed on (Rm(t−1)/Rm(t−13)−1), the t-statistics reported by lm() are likely not reliable. The model shows a very low R-squared value (0.0001372), indicating that the independent variable explains almost none of the variation in the dependent variable. Additionally, the coefficient for the return ratio is extremely close to zero and statistically insignificant (p-value = 0.69). This suggests that the model provides very little explanatory power. Furthermore, because you're dealing with time series data, there's a strong possibility of autocorrelation in the residuals, which violates one of the key assumptions of ordinary least squares (OLS) regression. Autocorrelation can lead to underestimated or overestimated standard errors, making the t-statistics unreliable. Therefore, without further diagnostic checks for autocorrelation or heteroscedasticity, the t-statistics reported by lm() should be treated with caution, and it's likely that they do not provide an accurate assessment of the significance of the coefficients.

4. Regress Rm(t) on an intercept and on abs(Rm(t)). Report estimates and t-statistics. Briefly interpret the results

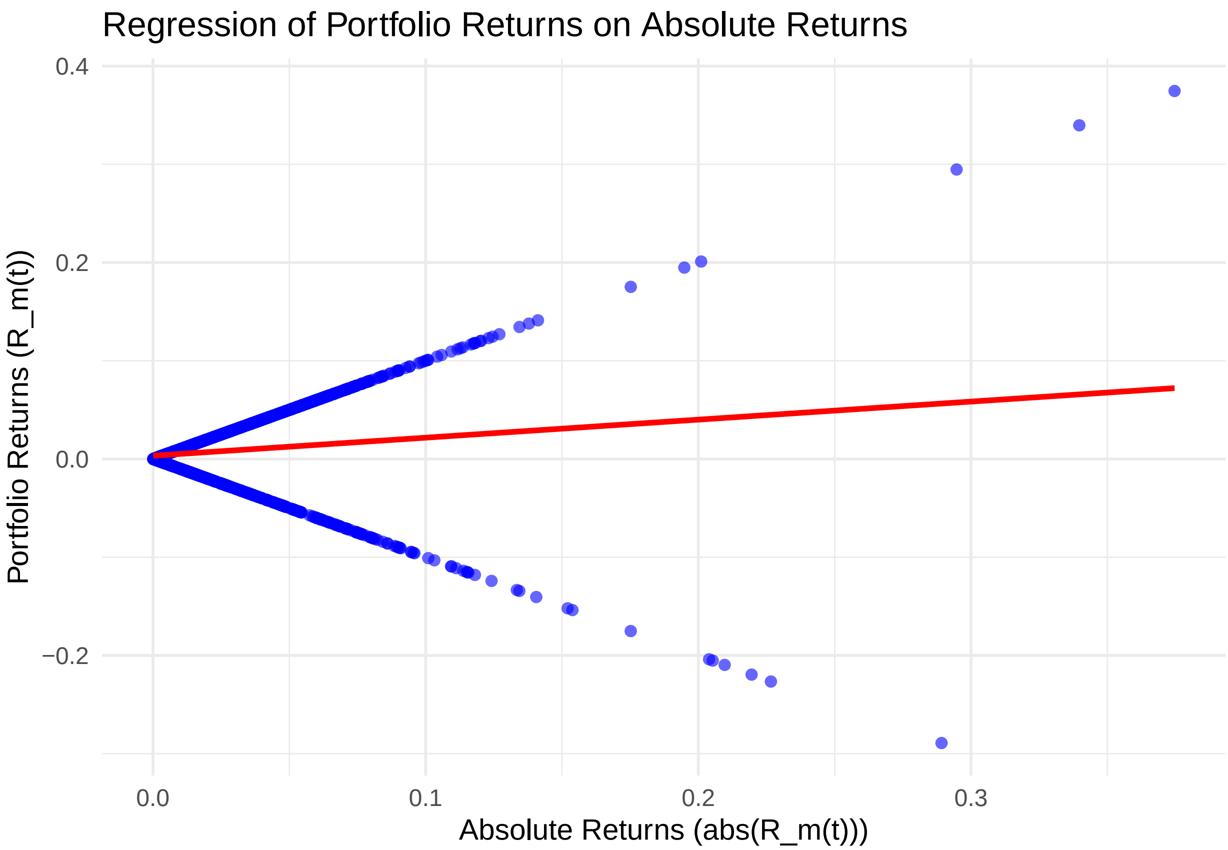
```
data <- data %>%
  mutate(abs_Rm = abs(Portfolio_Returns))
head(data)
```

```
##         Date   Cnsmr   Manuf   HiTec    Hlth   Other Portfolio_Returns
## 1 1926-07-01  0.0543  0.0273  0.0183  0.0177  0.0213          0.029400
## 2 1926-08-01  0.0276  0.0233  0.0241  0.0425  0.0435          0.029375
## 3 1926-09-01  0.0216 -0.0044  0.0106  0.0069  0.0029          0.008675
## 4 1926-10-01 -0.0390 -0.0242 -0.0226 -0.0057 -0.0284         -0.022875
## 5 1926-11-01  0.0370  0.0250  0.0307  0.0542  0.0211          0.036725
## 6 1926-12-01  0.0362  0.0276  0.0103  0.0011  0.0347          0.018800
##      Rm_lag Rm_lag13 Return_Ratio   abs_Rm
## 1        NA       NA           NA 0.029400
## 2  0.029400       NA           NA 0.029375
## 3  0.029375       NA           NA 0.008675
## 4  0.008675       NA           NA 0.022875
## 5 -0.022875       NA           NA 0.036725
## 6  0.036725       NA           NA 0.018800
```

```
model_3 <- lm(Portfolio_Returns ~ abs_Rm, data = data)
summary(model_3)
```

```
##
## Call:
## lm(formula = Portfolio_Returns ~ abs_Rm, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.34561 -0.02172  0.00733  0.02828  0.30249
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.003313   0.002105   1.573    0.116
## abs_Rm      0.183720   0.041245   4.454 9.22e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04966 on 1174 degrees of freedom
## Multiple R-squared:  0.01662,    Adjusted R-squared:  0.01578
## F-statistic: 19.84 on 1 and 1174 DF,  p-value: 9.218e-06
```

```
ggplot(data, aes(x = abs_Rm, y = Portfolio_Returns)) +
  geom_point(color = "blue", alpha = 0.6) +  # Scatter plot of returns
  geom_smooth(method = "lm", color = "red", se = FALSE) +  # Regression line without confidence interva
  labs(title = "Regression of Portfolio Returns on Absolute Returns",
       x = "Absolute Returns (abs(R_m(t)))",
       y = "Portfolio Returns (R_m(t))") +
  theme_minimal()
```



The regression of portfolio returns Rm(t) on the absolute value of returns abs(Rm(t)) reveals some notable findings. The intercept is estimated at 0.003313 (approximately 0.33%), but it is not statistically significant, with a t-value of 1.573 and a p-value of 0.116, suggesting that the expected return when the absolute value of returns is zero is not different from zero in a statistically significant way. However, the coefficient for abs(Rm(t)) is estimated at 0.18372, which is statistically significant with a t-value of 4.454 and a p-value of 9.22e-06. This implies that larger absolute returns (whether positive or negative) are associated with higher current returns, indicating a relationship between volatility (represented by absolute returns) and current portfolio returns. The R-squared value of 0.01662 is low, meaning that only about 1.66% of the variation in portfolio returns is explained by the absolute value of returns. While the relationship between volatility and returns is statistically significant, the explanatory power of the model is limited, suggesting other factors likely play a more substantial role in determining portfolio returns.

5. Repeat (1) on data from 199706. (Delete all data prior to 199706, then compute lagged returns). Comment briefly

```
data_filtered <- data %>%
  filter(Date >= as.Date("1997-06-01"))

data_filtered <- data_filtered %>%
  mutate(Rm_lag = lag(Portfolio_Returns, 1))

head(data_filtered)
```

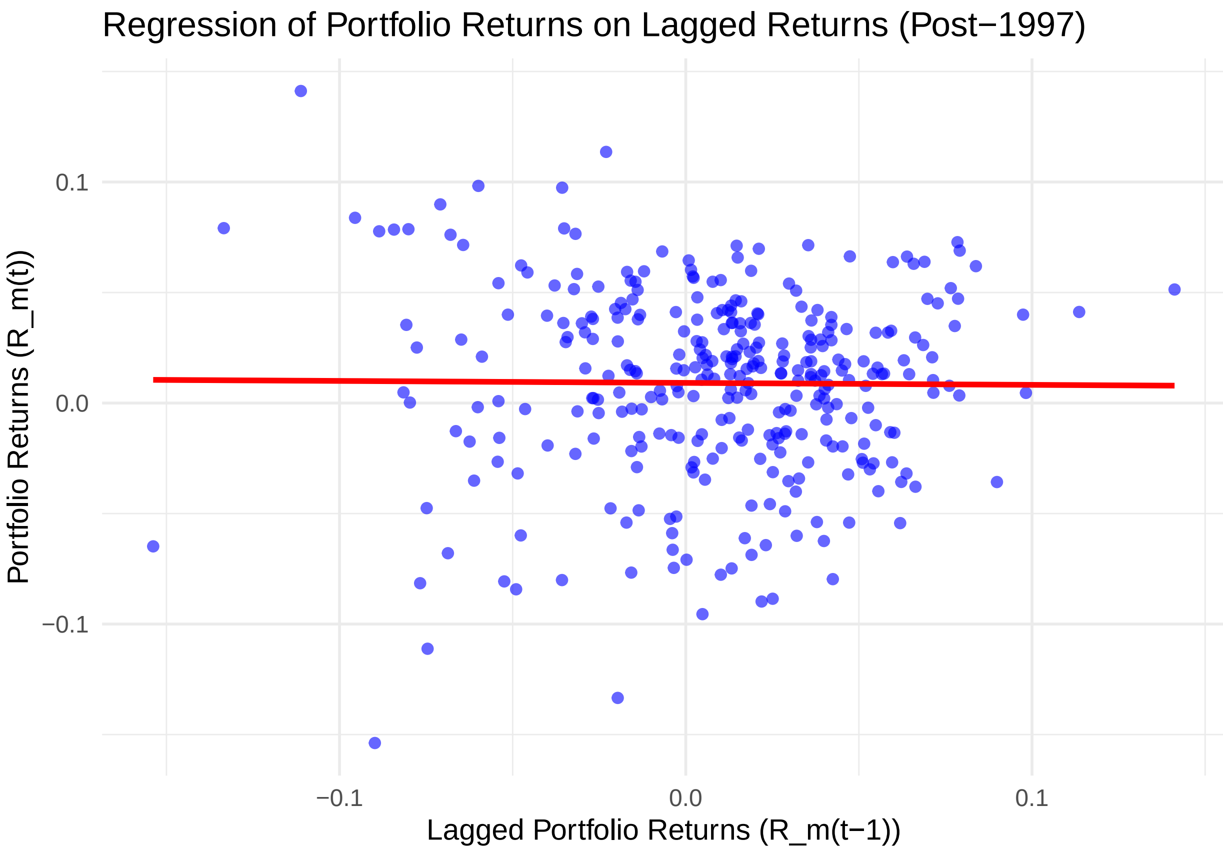
```
##         Date   Cnsmr   Manuf   HiTec    Hlth   Other Portfolio_Returns
## 1 1997-06-01  0.0412  0.0398  0.0198  0.0888  0.0525          0.047400
## 2 1997-07-01  0.0569  0.0666  0.1209  0.0210  0.0928          0.066350
## 3 1997-08-01 -0.0378 -0.0355 -0.0171 -0.0611 -0.0478         -0.037875
```

```
## 4 1997-09-01  0.0537  0.0426  0.0502  0.0662  0.0804          0.053175
## 5 1997-10-01 -0.0274 -0.0398 -0.0538  0.0010 -0.0276         -0.030000
## 6 1997-11-01  0.0491  0.0263  0.0407  0.0282  0.0297          0.036075
##      Rm_lag  Rm_lag13 Return_Ratio   abs_Rm
## 1        NA  0.032650   1.24042879 0.047400
## 2  0.047400 -0.006475  -8.32046332 0.066350
## 3  0.066350 -0.060125  -2.10353430 0.037875
## 4 -0.037875  0.031025  -2.22078969 0.053175
## 5  0.053175  0.057925  -0.08200259 0.030000
## 6 -0.030000  0.002350 -13.76595745 0.036075
```

```r
model_filtered <- lm(Portfolio_Returns ~ Rm_lag, data = data_filtered)
summary(model_filtered)
```

```
##
## Call:
## lm(formula = Portfolio_Returns ~ Rm_lag, data = data_filtered)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.163737 -0.024790  0.004623  0.028573  0.131073
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.009114   0.002407   3.787 0.000182 ***
## Rm_lag      -0.008882   0.055668  -0.160 0.873335
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04235 on 322 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  7.905e-05,  Adjusted R-squared:  -0.003026
## F-statistic: 0.02546 on 1 and 322 DF,  p-value: 0.8733
```

```r
ggplot(data_filtered, aes(x = Rm_lag, y = Portfolio_Returns)) +
  geom_point(color = "blue", alpha = 0.6) +
  geom_smooth(method = "lm", color = "red", se = FALSE) +
  labs(title = "Regression of Portfolio Returns on Lagged Returns (Post-1997)",
       x = "Lagged Portfolio Returns (R_m(t-1))",
       y = "Portfolio Returns (R_m(t))") +
  theme_minimal()
```



The regression of portfolio returns Rm(t) on lagged returns Rm(t-1), using data from June 1997 onward, shows that the intercept is significant, estimated at 0.91% with a p-value of 0.000182, indicating a positive expected return when lagged returns are zero. However, the coefficient for lagged returns is -0.008882, with a t-value of -0.160 and a p-value of 0.8733, suggesting that lagged returns have no statistically significant effect on current returns. The R-squared value is extremely low (0.000079), meaning that the model explains virtually none of the variation in returns, and the adjusted R-squared is negative, further confirming the model's poor fit. The overall model is not statistically significant, as indicated by the F-statistic of 0.02546 and its p-value of 0.8733. In conclusion, there is no meaningful relationship between past and current returns in this dataset.

6.Repeat (2) on data from 199706. (Delete all data prior to 199706, then compute lagged returns). Comment briefly

```
data_filtered <- data_filtered %>%
  mutate(
    Rm_lag13 = lag(Portfolio_Returns, 13),
    Return_Ratio = (Rm_lag / Rm_lag13) - 1
  )

head(data_filtered, 20)
```

```
##          Date   Cnsmr   Manuf   HiTec    Hlth   Other Portfolio_Returns
## 1  1997-06-01  0.0412  0.0398  0.0198  0.0888  0.0525          0.047400
## 2  1997-07-01  0.0569  0.0666  0.1209  0.0210  0.0928          0.066350
## 3  1997-08-01 -0.0378 -0.0355 -0.0171 -0.0611 -0.0478         -0.037875
## 4  1997-09-01  0.0537  0.0426  0.0502  0.0662  0.0804          0.053175
## 5  1997-10-01 -0.0274 -0.0398 -0.0538  0.0010 -0.0276         -0.030000
```

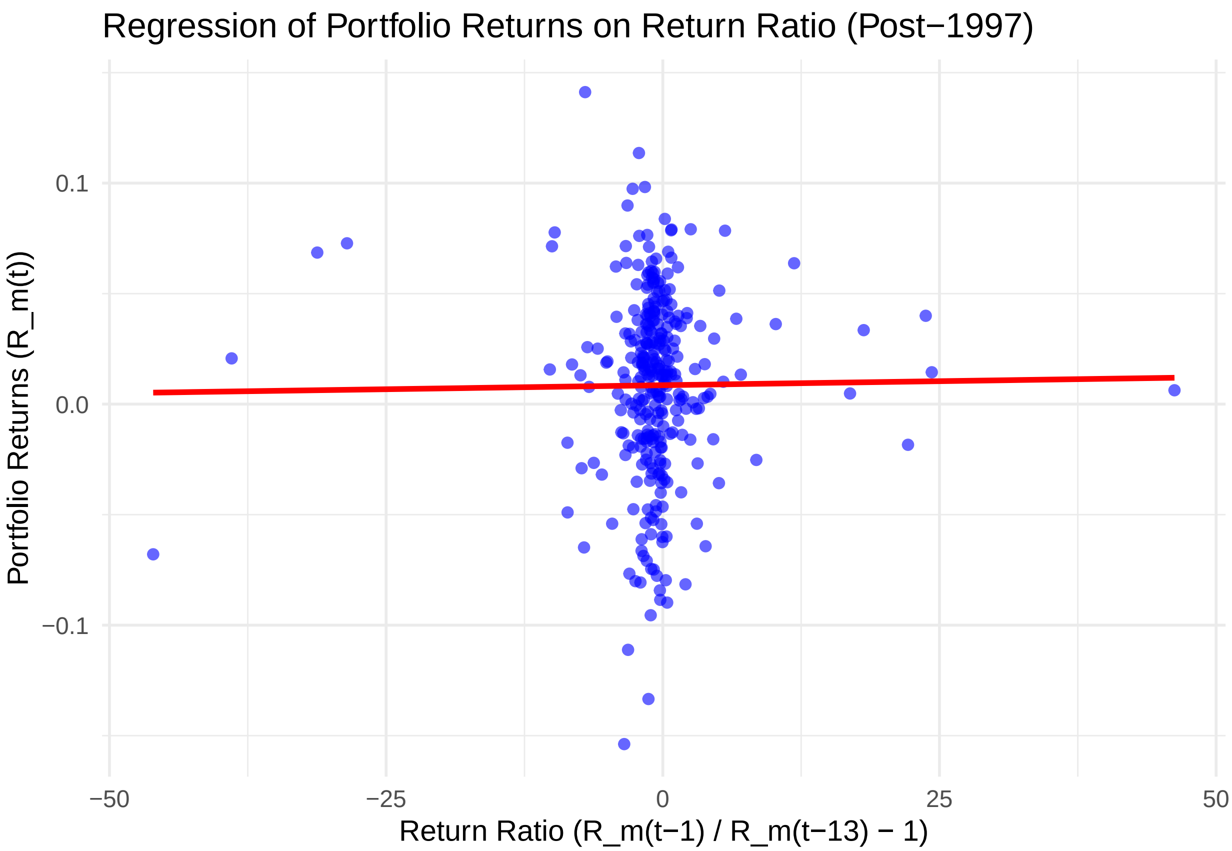
```
## 6  1997-11-01  0.0491  0.0263  0.0407  0.0282  0.0297          0.036075
## 7  1997-12-01  0.0211  0.0073 -0.0167  0.0353  0.0514          0.011750
## 8  1998-01-01 -0.0081 -0.0163  0.0520  0.0568 -0.0193          0.021100
## 9  1998-02-01  0.0821  0.0638  0.0782  0.0551  0.0825          0.069800
## 10 1998-03-01  0.0621  0.0510  0.0398  0.0357  0.0596          0.047150
## 11 1998-04-01 -0.0149  0.0125  0.0248  0.0191  0.0196          0.010375
## 12 1998-05-01  0.0243 -0.0265 -0.0544 -0.0250 -0.0307         -0.020400
## 13 1998-06-01  0.0414 -0.0029  0.0698  0.0617  0.0322          0.042500
## 14 1998-07-01 -0.0323 -0.0589  0.0152 -0.0026 -0.0186         -0.019650
## 15 1998-08-01 -0.1436 -0.1086 -0.1580 -0.1235 -0.2146         -0.133425
## 16 1998-09-01  0.0132  0.0474  0.1355  0.1204  0.0278          0.079125
## 17 1998-10-01  0.1086  0.0640  0.0645  0.0387  0.0902          0.068950
## 18 1998-11-01  0.0772  0.0248  0.0917  0.0619  0.0611          0.063900
## 19 1998-12-01  0.0491  0.0164  0.1495  0.0500  0.0345          0.066250
## 20 1999-01-01  0.0025 -0.0190  0.1329  0.0022  0.0095          0.029650
##       Rm_lag  Rm_lag13 Return_Ratio   abs_Rm
## 1          NA        NA           NA 0.047400
## 2   0.047400        NA           NA 0.066350
## 3   0.066350        NA           NA 0.037875
## 4  -0.037875        NA           NA 0.053175
## 5   0.053175        NA           NA 0.030000
## 6  -0.030000        NA           NA 0.036075
## 7   0.036075        NA           NA 0.011750
## 8   0.011750        NA           NA 0.021100
## 9   0.021100        NA           NA 0.069800
## 10  0.069800        NA           NA 0.047150
## 11  0.047150        NA           NA 0.010375
## 12  0.010375        NA           NA 0.020400
## 13 -0.020400        NA           NA 0.042500
## 14  0.042500  0.047400   -0.1033755 0.019650
## 15 -0.019650  0.066350   -1.2961567 0.133425
## 16 -0.133425 -0.037875    2.5227723 0.079125
## 17  0.079125  0.053175    0.4880113 0.068950
## 18  0.068950 -0.030000   -3.2983333 0.063900
## 19  0.063900  0.036075    0.7713098 0.066250
## 20  0.066250  0.011750    4.6382979 0.029650
```

```r
model_2 <- lm(Portfolio_Returns ~ Return_Ratio, data = data_filtered)
summary(model_2)
```

```
##
## Call:
## lm(formula = Portfolio_Returns ~ Return_Ratio, data = data_filtered)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.162139 -0.024194  0.004691  0.027735  0.133118
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.568e-03  2.424e-03   3.535  0.00047 ***
## Return_Ratio 7.279e-05  3.881e-04   0.188  0.85135
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.04252 on 310 degrees of freedom
##   (13 observations deleted due to missingness)
## Multiple R-squared:  0.0001135,  Adjusted R-squared:  -0.003112
## F-statistic: 0.03518 on 1 and 310 DF,  p-value: 0.8513
```

```
ggplot(data_filtered, aes(x = Return_Ratio, y = Portfolio_Returns)) +
  geom_point(color = "blue", alpha = 0.6) +
  geom_smooth(method = "lm", color = "red", se = FALSE) +
  labs(title = "Regression of Portfolio Returns on Return Ratio (Post-1997)",
       x = "Return Ratio (R_m(t-1) / R_m(t-13) - 1)",
       y = "Portfolio Returns (R_m(t))") +
  theme_minimal()
```



The regression of portfolio returns Rm(t) on the return ratio (R_m(t-1) / R_m(t-13) - 1), using data from June 1997 onward, shows that the intercept is significant, estimated at 0.008568 (about 0.86%) with a t-value of 3.535 and a p-value of 0.00047. However, the coefficient for the return ratio is 7.279e-05, with a t-value of 0.188 and a p-value of 0.85135, indicating that the return ratio is not a significant predictor of current portfolio returns. The R-squared is 0.0001135, meaning that the return ratio explains only a tiny fraction (0.01%) of the variance in the portfolio returns. The adjusted R-squared is negative, further highlighting the poor fit of the model. Overall, the return ratio does not appear to have any meaningful explanatory power for current returns in this dataset.

7. Repeat (4) on data from 199706. Comment briefly

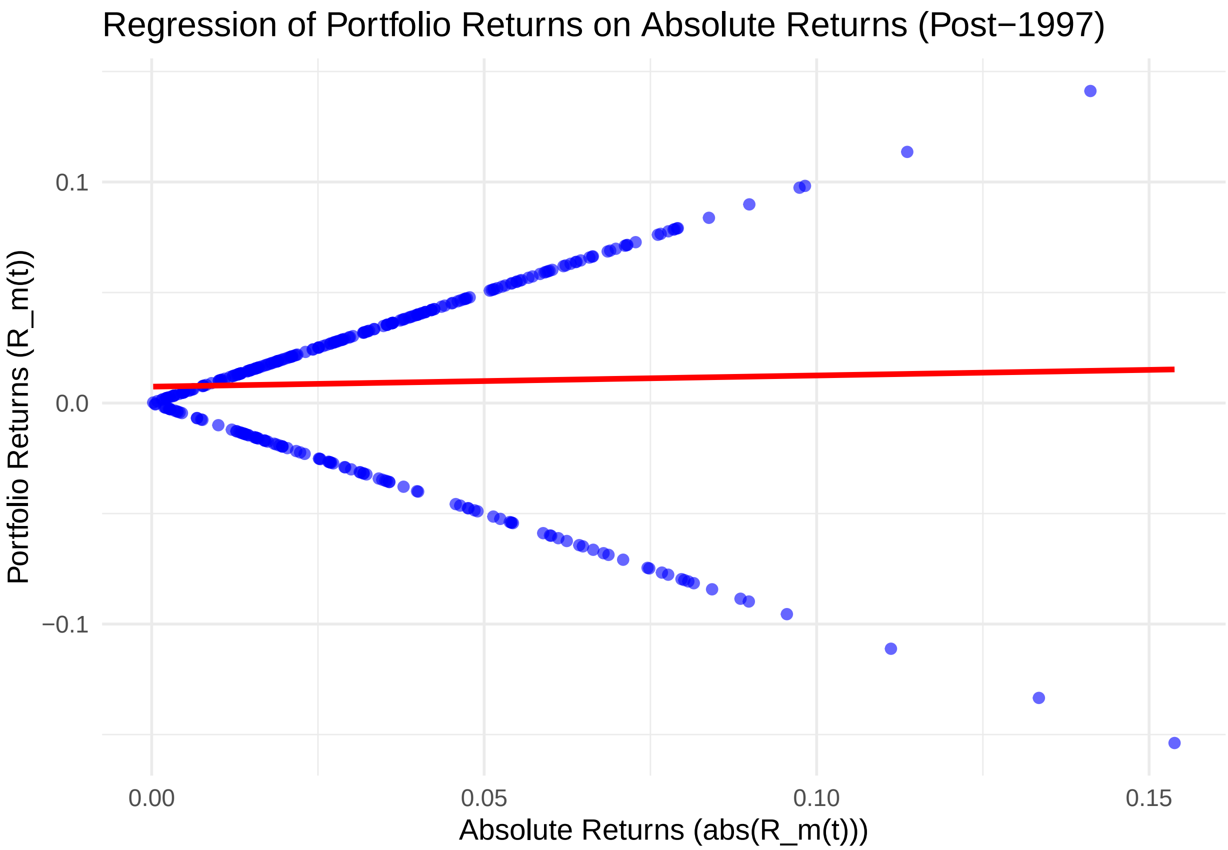
```
data_filtered <- data_filtered %>%
  mutate(abs_Rm = abs(Portfolio_Returns))
head(data_filtered)
```

```
##         Date   Cnsmr   Manuf   HiTec    Hlth   Other Portfolio_Returns
## 1 1997-06-01  0.0412  0.0398  0.0198  0.0888  0.0525          0.047400
## 2 1997-07-01  0.0569  0.0666  0.1209  0.0210  0.0928          0.066350
## 3 1997-08-01 -0.0378 -0.0355 -0.0171 -0.0611 -0.0478         -0.037875
## 4 1997-09-01  0.0537  0.0426  0.0502  0.0662  0.0804          0.053175
## 5 1997-10-01 -0.0274 -0.0398 -0.0538  0.0010 -0.0276         -0.030000
## 6 1997-11-01  0.0491  0.0263  0.0407  0.0282  0.0297          0.036075
##      Rm_lag Rm_lag13 Return_Ratio   abs_Rm
## 1        NA       NA           NA 0.047400
## 2  0.047400       NA           NA 0.066350
## 3  0.066350       NA           NA 0.037875
## 4 -0.037875       NA           NA 0.053175
## 5  0.053175       NA           NA 0.030000
## 6 -0.030000       NA           NA 0.036075
```

```r
model_3 <- lm(Portfolio_Returns ~ abs_Rm, data = data_filtered)
summary(model_3)
```

```
##
## Call:
## lm(formula = Portfolio_Returns ~ abs_Rm, data = data_filtered)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.169025 -0.023901  0.005339  0.028384  0.126616
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.007406   0.003891   1.903   0.0579 .
## abs_Rm      0.050668   0.090091   0.562   0.5742
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04232 on 323 degrees of freedom
## Multiple R-squared:  0.0009783,  Adjusted R-squared:  -0.002115
## F-statistic: 0.3163 on 1 and 323 DF,  p-value: 0.5742
```

```r
ggplot(data_filtered, aes(x = abs_Rm, y = Portfolio_Returns)) +
  geom_point(color = "blue", alpha = 0.6) +
  geom_smooth(method = "lm", color = "red", se = FALSE) +
  labs(title = "Regression of Portfolio Returns on Absolute Returns (Post-1997)",
       x = "Absolute Returns (abs(R_m(t)))",
       y = "Portfolio Returns (R_m(t))") +
  theme_minimal()
```



The regression of portfolio returns Rm(t) on the absolute value of returns absRm(t), using data from June 1997 onward, shows that neither the intercept nor the absolute returns are statistically significant. The intercept is estimated at 0.007406 with a t-value of 1.903 and a p-value of 0.0579, indicating weak statistical significance. The coefficient for absolute returns is 0.050668, with a t-value of 0.562 and a p-value of 0.5742, suggesting no significant relationship between absolute returns and current portfolio returns. The R-squared is 0.0009783, meaning that less than 0.1% of the variation in returns is explained by the model. The adjusted R-squared is negative, further confirming the poor fit of the model. Overall, the absolute returns do not provide meaningful explanatory power for current returns in this dataset.