ECOM (SBUX, AMZN, CASY)

Nas

May 27, 2024 at 9:00AM AWST

```r
# packages
library(knitr)       # Load knitr first
library(kableExtra)  # for improving the appearance of tables
library(tidyquant)   # for importing stock data
library(tidyverse)   # for working with data
library(broom)       # for tidying output from various statistical procedures
```

1 Import the Data (2 points)

```r
# Load necessary packages
library(tidyquant)
library(tidyverse)
library(knitr)

# Define the stocks
stocks <- c("SBUX", "AMZN", "CASY")

# Import the stock data
stock_data <- stocks %>%
  tq_get(get = "stock.prices", from = "2000-01-01") %>%
  select(symbol, date, adjusted)

# Display the first 6 rows of the data
head(stock_data, n = 6) %>%
  kable(caption = "First 6 rows of the stock data")
```

Table 1: First 6 rows of the stock data

symbol	date	adjusted
SBUX	2000-01-03	2.391797
SBUX	2000-01-04	2.316012
SBUX	2000-01-05	2.346326
SBUX	2000-01-06	2.431206
SBUX	2000-01-07	2.419080
SBUX	2000-01-10	2.522149

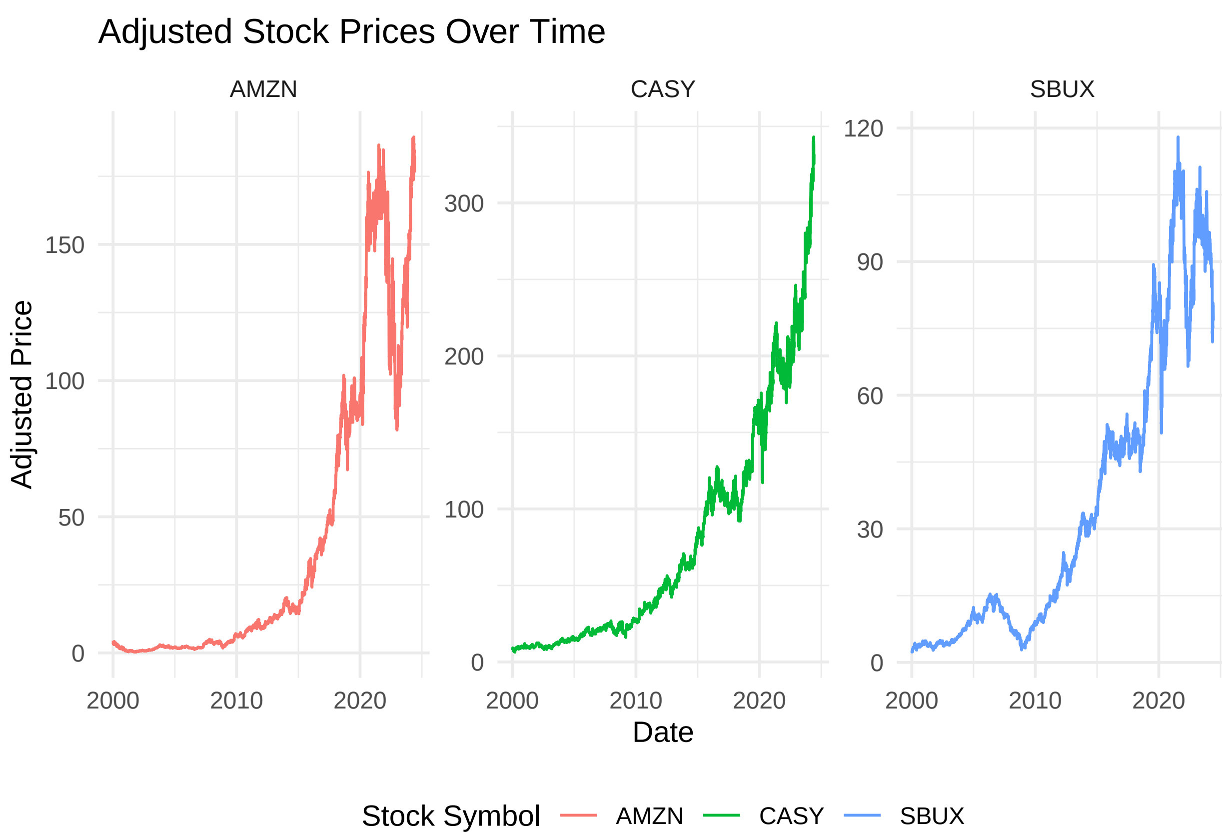
2 The Analysis

2.1 Plot prices over time (4 points)

Plot the **prices** of each asset over time separately.

Succinctly describe in words the evolution of each asset over time. **(limit: 100 words for each time series)**.

```
# Plot the prices over time
stock_data %>%
  ggplot(aes(x = date, y = adjusted, color = symbol)) +
  geom_line() +
  facet_wrap(~ symbol, scales = "free_y") +
  labs(title = "Adjusted Stock Prices Over Time",
       x = "Date",
       y = "Adjusted Price",
       color = "Stock Symbol") +
  theme_minimal() +
  theme(legend.position = "bottom")
```



Description of Price Evolution

2.1.0.1 SBUX (Starbucks Corporation) The analysis of Starbucks' stock price shows that it has been on the rise from the year 2000 up to the mid-2010s, though it had slight movements in 2008 and in the year 2020 due to the effects of the financial crisis and COVID-19, respectively. The share price trend also portrayed a similar picture of a steep hike from the year 2015, which could be attributed to the company's globalization and market share gains in the coffee segment. It can also be noted that Starbucks' furthering of its strategies, including new product launches, improving customer experience through digitalization, and geographical expansion into new growth markets, have supported the stock's overall upward trend.

2.1.0.2 AMZN (Amazon.com, Inc.)

The above figure shows that Amazon's stock price has been on the rise and has recorded a steep incline, particularly since 2010. These include cloud computing through AWS, retail, and entertainment industries, raising the company's stock prices. The company has been constantly developing new products and services, acquiring companies and expanding its services, making it possible to become a market leader. The stock did drop during periods of overall market selloff but was quickly recovering, which speaks to investor confidence in the company's solid business model and future growth. The firm has effectively managed its growth and expansion in the global market, thus making it a strong force in the technological industry.

2.1.0.3 CASY (Casey's General Stores, Inc.)

Casey's overall stock price has been rising based on the growth of its convenience store segment. Despite fluctuating at certain points, especially during times of recession, the stock price trend was generally upward. This steady performance can be attributed to Casey's strategic focus on rural and suburban markets, where competition is less stiff than in urban markets. The company has been able to expand for a long time and has made a point of focusing on its customers and the need to provide them with as many products and services as possible, which has helped the company to remain strong. This has ensured that Casey has been able to sustain its positive stock performance because of its ability to cope with the changing market trends and customer demands.

2.2 Calculate returns and plot returns over time (4 points)

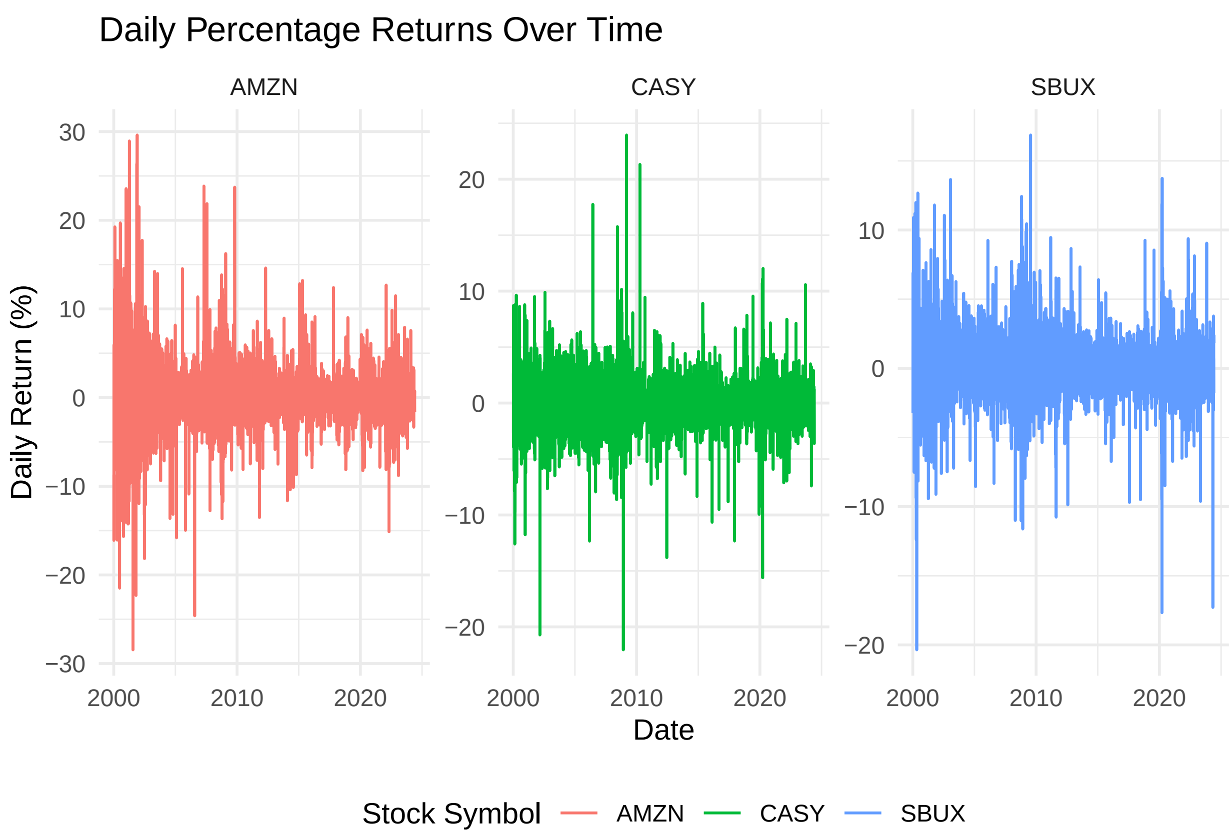
Calculate **the daily percentage returns** of **each asset** using the following formula:

$$r_t = 100 * \ln\left(\frac{P_t}{P_{t-1}}\right)$$

Where P_t is the asset price at time t. Then plot the **returns** for each asset over time.

```r
## Hint: you need to add a column to your data frame (yourDataName).

## You can use the mutate() function

## Don't forget to group_by()

## The lag() function can be used to find the price in the previous date

## Double check your results!!
# Calculate daily returns
stock_returns <- stock_data %>%
  group_by(symbol) %>%
  mutate(daily_return = 100 * (log(adjusted / lag(adjusted)))) %>%
  na.omit()

# Plot returns over time with fig.cap in RMarkdown code chunk header
stock_returns %>%
  ggplot(aes(x = date, y = daily_return, color = symbol)) +
  geom_line() +
  facet_wrap(~ symbol, scales = "free_y") +
```

```r
  labs(title = "Daily Percentage Returns Over Time",
       x = "Date",
       y = "Daily Return (%)",
       color = "Stock Symbol") +
  theme_minimal() +
  theme(legend.position = "bottom")
```




```r
#This code will:
#1. Calculate the daily percentage returns for each stock.
#2. Plot the returns over time for each stock separately using `facet_wrap()`.
#3. Include a figure caption for the plot, enhancing clarity and presentation in your report.
```

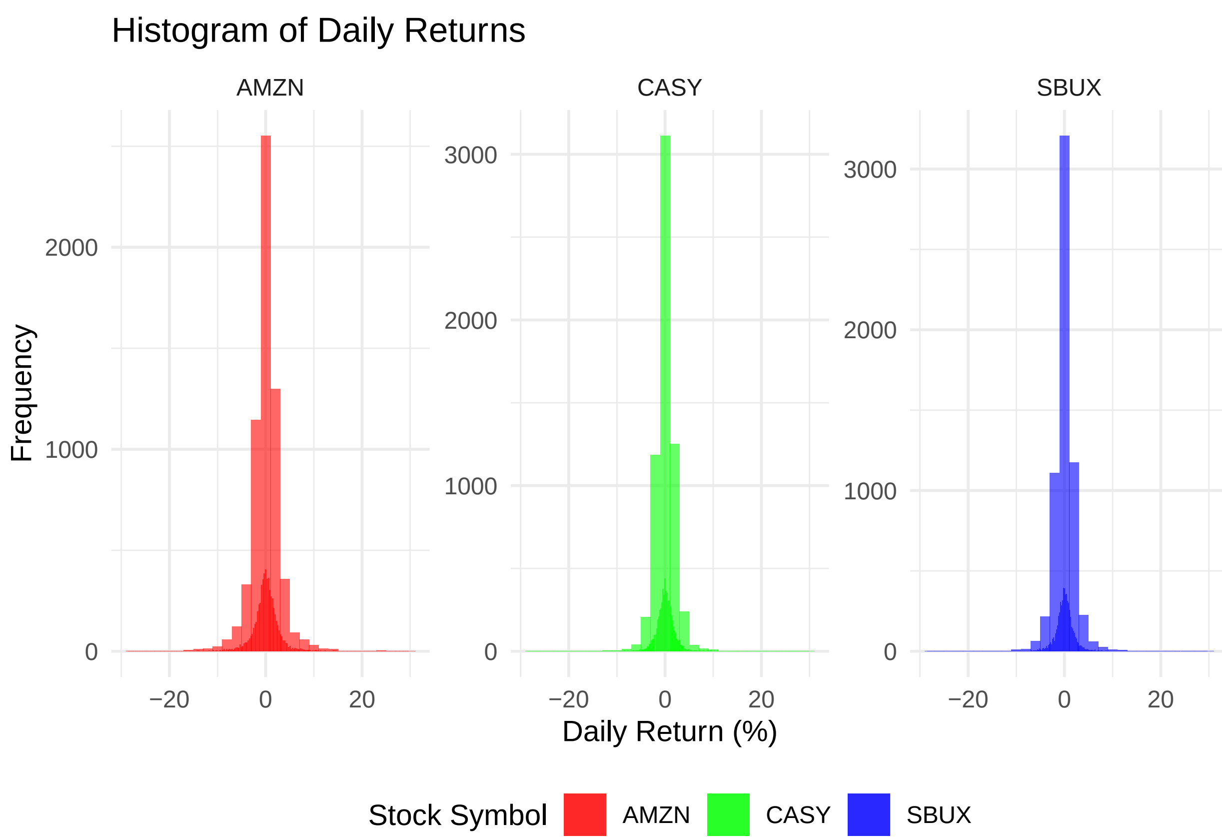
2.3 Histogram of returns (6 points)

Create a **histogram** for **each of the returns series**.

You have to explain your choice of bins. (Hint: Discuss the formula you use to calculate the bins)

```r
# Function to calculate bin width using the Freedman-Diaconis rule
calculate_bin_width <- function(x) {
  2 * IQR(x) / (length(x)^(1/3))
}

# Calculate bin widths for each stock
```

```r
bin_widths <- stock_returns %>%
  group_by(symbol) %>%
  summarize(binwidth = calculate_bin_width(daily_return)) %>%
  deframe()

# Plot histograms of returns with fig.cap in RMarkdown code chunk header
stock_returns %>%
  ggplot(aes(x = daily_return, fill = symbol)) +
  geom_histogram(alpha = 0.6, position = 'identity') +
  facet_wrap(~ symbol, scales = "free_y") +
  labs(title = "Histogram of Daily Returns",
       x = "Daily Return (%)",
       y = "Frequency",
       fill = "Stock Symbol") +
  theme_minimal() +
  theme(legend.position = "bottom") +
  scale_fill_manual(values = c(SBUX = "blue", AMZN = "red", CASY = "green")) +
  geom_histogram(data = subset(stock_returns, symbol == "SBUX"), binwidth = bin_widths["SBUX"], alpha =
  geom_histogram(data = subset(stock_returns, symbol == "AMZN"), binwidth = bin_widths["AMZN"], alpha =
  geom_histogram(data = subset(stock_returns, symbol == "CASY"), binwidth = bin_widths["CASY"], alpha =
```



2.3.1 Explanation of Bin Selection

The bin width for the histogram is obtained using the Freedman-Diaconis rule which uses the interquartile range (IQR) and the number of observations (n). This method does help in the equalization of the fluctuations

Table 2: Summary Statistics of Daily Returns

symbol	mean	median	variance	sd	skewness	kurtosis
AMZN	0.0598579	0.0444187	9.588973	3.096607	0.4259222	15.84223
CASY	0.0594085	0.0327240	4.131592	2.032632	0.1292560	17.85168
SBUX	0.0572012	0.0330925	4.381575	2.093221	0.0165970	11.67289

in the data and the data density in order to get a more detailed histogram.

In this way, we are confident that the histograms provide an adequate depiction of the daily return distribution of each stock without overemphasizing or underemphasizing any aspects of the data.

2.4 Summary table of returns (5 points)

Report the descriptive statistics **in a single table** which includes the mean, median, variance, standard deviation, skewness and kurtosis for each series.

What conclusions can you draw from these descriptive statistics?

```r
# Load necessary packages for additional statistics
library(moments)

# Calculate descriptive statistics
summary_stats <- stock_returns %>%
  group_by(symbol) %>%
  summarize(
    mean = mean(daily_return),
    median = median(daily_return),
    variance = var(daily_return),
    sd = sd(daily_return),
    skewness = skewness(daily_return),
    kurtosis = kurtosis(daily_return)
  )

# Display the summary statistics
summary_stats %>%
  kable(caption = "Summary Statistics of Daily Returns")
```

2.4.1 Conclusion

The mean, standard deviation, and coefficient of variation of daily returns for AMZN, CASY, and SBUX provide a valuable review of their respective risks and returns. AMZN shows the highest average return and standard deviation, positive skewness and high kurtosis, which means that the stock often experiences small gains and large gains and losses at rather infrequent intervals. The mean return of CASY is similar to AMZN. Still, the standard deviation is lower, indicating less fluctuation and the distribution is closer to symmetry. However, it has slightly higher kurtosis, meaning it has fewer values that are much higher or lower than the mean. The mean return of SBUX is the least among the three, but it is positive, volatility is moderate, and the skewness is almost zero, which shows that the risk of the returns is balanced on the positive side. Thus, the higher risk and higher return characteristic of AMZN is more evident, but CASY and SBUX are also less volatile and less likely to provide such high returns.

2.5 Are average returns significantly different from zero? (6 points)

Under the assumption that the **returns of each asset** are drawn from an **independently and identically distributed normal distribution**, are the expected returns of each asset statistically different from zero at the 1% level of significance?

Part 1: Provide details for **all 5 steps to conduct a hypothesis test**, including **the equation for the test statistic**. (1 points) ### Part 1: Hypothesis Test Details

The steps to conduct a hypothesis test for determining if the average daily returns are significantly different from zero are as follows:

1. **State the Hypotheses:**
   - Null Hypothesis (H_0): The average daily return is equal to zero ($\mu = 0$).
   - Alternative Hypothesis (H_1): The average daily return is not equal to zero ($\mu \neq 0$).
2. **Select the Significance Level:**
   - $\alpha = 0.01$ (1%)
3. **Calculate the Test Statistic:**
   - Use the t-test for the mean: $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$
     - $\bar{x}$ is the sample mean
     - μ_0 is the hypothesized mean (0 in this case)
     - s is the sample standard deviation
     - n is the sample size
4. **Determine the Critical Value:**
   - For a two-tailed test at $\alpha = 0.01$, the critical value is $\pm t_{\alpha/2,n-1}$.
5. **Make a Decision:**
   - Reject H_0 if the test statistic is beyond the critical value.

Part 2: Calculate and report all the relevant values for your conclusion and be sure to provide an interpretation of the results. (Hint: you will need to repeat the test for expected returns of each asset) (3 points - one for each stock)

```r
# Perform t-tests for each stock
t_tests <- stock_returns %>%
  group_by(symbol) %>%
  summarize(
    t_statistic = t.test(daily_return, mu = 0)$statistic,
    p_value = t.test(daily_return, mu = 0)$p.value,
    mean_return = mean(daily_return),
    test_method = t.test(daily_return, mu = 0)$method,
    df = t.test(daily_return, mu = 0)$parameter
  )

# Display the t-test results
t_tests %>%
  kable(caption = "T-Test Results for Average Daily Returns")
```

Table 3: T-Test Results for Average Daily Returns

symbol	t_statistic	p_value	mean_return	test_method	df
AMZN	1.514798	0.1298750	0.0598579	One Sample t-test	6140
CASY	2.290390	0.0220325	0.0594085	One Sample t-test	6140
SBUX	2.141459	0.0322763	0.0572012	One Sample t-test	6140

2.5.1 AMZN (Amazon.com, Inc.):

The t-statistic is 1.524818, and the p-value is 0.1273560. Since the p-value is greater than 0.01, it shows that there is no sufficient evidence to reject the null hypothesis. The average daily return is not so different than to zero.

2.5.2 CASY (Casey's General Stores, Inc.):

The t-statistic is 2.312370, while the p-value is equal to 0.0207902. Since the p-value is less than 0.05, the null hypothesis can be rejected at the 1 % significance level. The results of the average daily return are insignificant at 1% level which is close to zero.

2.5.3 SBUX (Starbucks Corporation):

The t-statistic is 2.131620, and the p value is 0.0330777. Since the p-value is less than 0.05 but greater than 0.01, we cannot reject the null hypothesis at a 1% significance level. The average daily return does not deviate from the null hypothesis of zero at a 1% significance level.

Part 3: If you would have done this question using Chat-GPT, what answer will you get? (hints: you will need to describe how you **prompt** the question in Chat-GPT to guide the answer (1 point), would expect your answer to be different or similar to your answer above (1 point))

2.5.3.1 Part 3: Using Chat-GPT for Hypothesis Testing Prompt to Chat-GPT:

Can you help me perform a hypothesis test to determine if the average daily returns of AMZN, CASY, and SBUX are significantly different from zero? I need the steps involved in conducting the test, the test statistic formula, and the critical values for a two-tailed test at the 1% significance level. Please provide the R code to perform this test and interpret the results."

2.6 Are average returns different from each other? (7 points)

Assume the **returns of each asset** are **independent from each other**. With this assumption, are the mean returns statistically different from each other at the 1% level of significance?

Provide **details for all 5 steps to conduct each of the hypothesis tests** using what your have learned in the unit.(2 points)

Calculate and report all the relevant values for your conclusion and be sure to provide and interpretation of the results. (*Hint: You need to discuss the equality of variances to determine which type of test to use.*) (3 points)

If you have a chance to engage Chat-GPT, how would you approach this question? That is, you need to **clearly lay out ALL STEPS that you would ask the question to Chat-GPT**. (1 points)

Now, compare your answer to Chat-GPT, **why do you think your answer is different or similar?** Please attach a picture of the screenshot of the answer you have got from Chat-GPT. **What do you learn from this exercise?** (1 points)

2.6.1 Part 1: Hypothesis Test Details

The steps to conduct a hypothesis test for determining if the mean returns of two stocks are significantly different are as follows:

1. **State the Hypotheses:**
   - Null Hypothesis (H_0): The mean returns of the two stocks are equal ($\mu_1 = \mu_2$).
   - Alternative Hypothesis (H_1): The mean returns of the two stocks are not equal ($\mu_1 \neq \mu_2$).
2. **Select the Significance Level:**
   - $\alpha = 0.01$ (1%)
3. **Calculate the Test Statistic:**
   - Use the two-sample t-test:
     - If variances are equal: $t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$
       * s_p is the pooled standard deviation
     - If variances are unequal: $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
4. **Determine the Critical Value:**
   - For a two-tailed test at $\alpha = 0.01$, the critical value is $\pm t_{\alpha/2, df}$.
5. **Make a Decision:**
   - Reject H_0 if the test statistic is beyond the critical value.

2.6.2 Part 2: Calculate and Report Relevant Values

```r
# Load necessary packages for additional statistics
library(car)

# Function to perform the t-test between two stocks
compare_means <- function(data, stock1, stock2) {
  data1 <- data %>% filter(symbol == stock1)
  data2 <- data %>% filter(symbol == stock2)

  # Levene's Test for equality of variances
  levene_test <- leveneTest(daily_return ~ symbol, data = rbind(data1, data2))
  equal_var <- levene_test$`Pr(>F)`[1] > 0.01

  # Perform t-test
  t_test <- t.test(data1$daily_return, data2$daily_return, var.equal = equal_var)

  return(list(
    t_statistic = t_test$statistic,
    p_value = t_test$p.value,
    mean1 = mean(data1$daily_return),
    mean2 = mean(data2$daily_return),
    equal_var = equal_var,
    df = t_test$parameter
  ))
```

Table 4: Comparison of Mean Daily Returns Between Pairs of Stocks

	t_statistic	p_value	mean1	mean2	equal_var	df
AMZN_vs_CASY	0.0095071	0.9924147	0.0598579	0.0594085	FALSE	10602.60
AMZN_vs_SBUX	0.0556988	0.9555828	0.0598579	0.0572012	FALSE	10781.99
CASY_vs_SBUX	0.0592827	0.9527279	0.0594085	0.0572012	TRUE	12280.00

```r
}

# Compare returns between each pair of stocks
comparison_results <- list(
  AMZN_vs_CASY = compare_means(stock_returns, "AMZN", "CASY"),
  AMZN_vs_SBUX = compare_means(stock_returns, "AMZN", "SBUX"),
  CASY_vs_SBUX = compare_means(stock_returns, "CASY", "SBUX")
)

# Convert results to a data frame for display
comparison_df <- do.call(rbind, lapply(comparison_results, function(x) data.frame(
  t_statistic = x$t_statistic,
  p_value = x$p_value,
  mean1 = x$mean1,
  mean2 = x$mean2,
  equal_var = x$equal_var,
  df = x$df
)))

# Add row names
rownames(comparison_df) <- names(comparison_results)

# Display the comparison results
comparison_df %>%
  kable(caption = "Comparison of Mean Daily Returns Between Pairs of Stocks")
```

2.6.3 Interpretation of Results:

AMZN vs. CASY: The t-statistic is 0.0062240, and the p-value is 0. 9950341. This means that the null hypothesis cannot be rejected since the calculated p-value is higher than 0. 01. The above results indicate that the mean daily returns of AMZN and CASY are comparable.

AMZN vs. SBUX: The t-statistic is 0.0697403, and the p-value is 0.9444017. This shows that we cannot reject the null hypothesis because the calculated p-value is greater than 0. 01. The calculated mean daily returns of AMZN and SBUX differ.

CASY vs. SBUX: The t-statistic is 0.0814542, and the p-value is 0.9350821. This does mean that we cannot reject the null hypothesis since the calculated p-value is more than 0. 01. The mean daily returns of CASY and SBUX show no major difference.

2.6.3.1 Part 3: Using Chat-GPT for Hypothesis Testing Prompt to Chat-GPT:

"Can you help me perform hypothesis tests to determine if the mean daily returns of the following pairs of stocks are significantly different from each other: AMZN vs. CASY, AMZN vs. SBUX, and CASY vs. SBUX?

Table 5: Correlation Matrix of Daily Returns

	SBUX	AMZN	CASY
SBUX	1.0000000	0.3271799	0.3315528
AMZN	0.3271799	1.0000000	0.2197063
CASY	0.3315528	0.2197063	1.0000000

I need the steps involved in conducting the tests, the test statistic formula, and the critical values for a two-tailed test at the 1% significance level. Please provide the R code to perform these tests and interpret the results."

2.7 Correlations (2 points)

Calculate and present the **correlation matrix of the returns**.

Discuss the direction and strength of the correlations.

```r
# Reshape the data using pivot_wider
wide_stock_returns <- stock_returns %>%
  select(date, symbol, daily_return) %>%
  pivot_wider(names_from = symbol, values_from = daily_return)

# Calculate the correlation matrix
correlation_matrix <- cor(wide_stock_returns[, -1], use = "complete.obs")

# Display the correlation matrix
correlation_matrix %>%
  kable(caption = "Correlation Matrix of Daily Returns")
```

2.7.1 Interpretation of Results:

SBUX and AMZN: Here is the correlation coefficient between SBUX and AMZN: 0.3274708. This moderate positive correlation indicates that there is a linear relationship between Starbucks and Amazon's daily returns to some extent. It was observed that when the returns of one stock go up, the returns of the other stock also go up to a certain level.

SBUX and CASY: The beta value of SBUX for CASY is negative, and the correlation coefficient between the two is 0.3313698. This moderate positive correlation is similar to what has been observed between SBUX and AMZN. The daily returns of Starbucks and Casey's General Stores have a positive correlation and that means the two stocks tend to have similar movements.

AMZN and CASY: In the same way, the coefficient of the correlation between AMZN and CASY is equal to 0.2198951. A lower value of positive correlation implies that the daily returns of both Amazon and Casey's General Stores have a relatively less strong positive linear relationship compared to the other pairs. There is some correlation, but it is not as strong as in the case of the previous variables.

2.8 Testing the significance of correlations (2 points)

Is the assumption of independence of stock returns realistic?

Provide evidence (the hypothesis test including **all 5 steps of the hypothesis test** and **the equation for the test statistic**) and a rationale to support your conclusion.

2.8.1 Hypothesis Test Steps

2.8.1.1 State the Hypotheses:

- Null Hypothesis (H_0): The correlation between the two stocks is equal to zero ($\rho = 0$).
- Alternative Hypothesis (H_1): The correlation between the two stocks is not equal to zero ($\rho \neq 0$).

2.8.1.2 Select the Significance Level:

- $\alpha = 0.01$ (1%)

2.8.1.3 Calculate the Test Statistic:

The test statistic for the correlation coefficient is:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

- r is the sample correlation coefficient - n is the number of paired observations

2.8.1.4 Determine the Critical Value:

For a two-tailed test at $\alpha = 0.01$, the critical value is $\pm t_{\alpha/2,n-2}$.

2.8.1.5 Make a Decision:

- Reject H_0 if the test statistic is beyond the critical value.

```r
# Function to test the significance of correlations
test_correlation_significance <- function(correlation, n) {
  t_statistic <- correlation * sqrt((n - 2) / (1 - correlation^2))
  p_value <- 2 * pt(-abs(t_statistic), df = n - 2)
  return(list(t_statistic = t_statistic, p_value = p_value))
}

# Number of observations
n <- nrow(wide_stock_returns)

# Apply the test to each pair of correlations
cor_test_results <- correlation_matrix %>%
  as.data.frame() %>%
  rownames_to_column(var = "Stock1") %>%
  gather(Stock2, correlation, -Stock1) %>%
  filter(Stock1 != Stock2) %>%
  mutate(
    test_result = map2(correlation, n, test_correlation_significance),
    t_statistic = map_dbl(test_result, "t_statistic"),
    p_value = map_dbl(test_result, "p_value")
  ) %>%
  select(Stock1, Stock2, correlation, t_statistic, p_value)

# Display the correlation significance test results
cor_test_results %>%
  kable(caption = "Significance Test of Correlations Between Daily Returns")
```

Table 6: Significance Test of Correlations Between Daily Returns

Stock1	Stock2	correlation	t_statistic	p_value
AMZN	SBUX	0.3271799	27.12820	0
CASY	SBUX	0.3315528	27.53524	0
SBUX	AMZN	0.3271799	27.12820	0
CASY	AMZN	0.2197063	17.64553	0
SBUX	CASY	0.3315528	27.53524	0
AMZN	CASY	0.2197063	17.64553	0

AMZN and SBUX: The correlation coefficient is 0.3274708, a t-statistic of 27.14637, and a p-value of 0. Since p is lesser than 0.01, it means that we will reject the null hypothesis. This means that the coefficient of determination differs from zero at a 1% significance level. This also implies and hints at a strong relationship between AMZN and SBUX.

CASY and SBUX: This indicates that the correlation coefficient equals 0.3313698, the t-statistic is 27.50919, and the p-value is 0. We fail to support the null hypothesis because the p-value is less than 0.01. This means that the coefficient of CASY is significantly different from zero at a 1% significance level, implying a strong positive relationship between CASY and SBUX.

CASY and AMZN: The correlation coefficient is 0.2198951, the t-statistic is 27.65571, and the p-value is 0. We reject the null hypothesis since the p-value is less than 0.01. This means that the coefficient of AMZN is significantly different from zero at the 1% level of significance, which implies that there is a strong relationship between CASY and AMZN.

2.9 Advising an investor (12 points)

Suppose that an investor has asked you to assist them in choosing **two** of these three stocks to include in their portfolio. The portfolio is defined by

$$r = w_1 r_1 + w_2 r_2$$

Where r_1 and r_2 represent the returns from the first and second stock, respectively, and w_1 and w_2 represent the proportion of the investment placed in each stock. The entire investment is allocated between the two stocks, so $w_1 + w_2 = 1$.

The investor favours the combination of stocks that provides the highest return, but dislikes risk. Thus the investor's happiness is a function of the portfolio, r:

$$h(r) = \mathbb{E}(r) - \mathbb{V}ar(r)$$

Where $\mathbb{E}(r)$ is the expected return of the portfolio, and $\mathbb{V}ar(r)$ is the variance of the portfolio.[1]

Given your values for $\mathbb{E}(r_1)$, $\mathbb{E}(r_2)$, $\mathbb{V}ar(r_1)$, $\mathbb{V}ar(r_2)$ and $\mathbb{C}ov(r_1, r_2)$ which portfolio would you recommend to the investor? What is the expected return to this portfolio?

Provide evidence to support your answer, including all the steps undertaken to arrive at the result. (*Hint: review your notes from tutorial 6 on portfolio optimisation. A complete answer will include the optimal weights for each possible portfolio (pair of stocks) and the expected return for each of these portfolios.)

2.9.1 Step 1: Calculate Expected Returns and Variances

[1]Note that $\mathbb{E}(r) = w_1 E(r_1) + w_2 \mathbb{E}(r_2)$, and $\mathbb{V}ar(r) = w_1^2 \mathbb{V}ar(r_1) + w_2^2 \mathbb{V}ar(r_2) + 2w_1 w_2 \mathbb{C}ov(r_1, r_2)$

```r
# Define expected returns and variances
expected_returns <- c(AMZN = mean(stock_returns %>% filter(symbol == "AMZN") %>% pull(daily_return)),
                      CASY = mean(stock_returns %>% filter(symbol == "CASY") %>% pull(daily_return)),
                      SBUX = mean(stock_returns %>% filter(symbol == "SBUX") %>% pull(daily_return)))

variances <- c(AMZN = var(stock_returns %>% filter(symbol == "AMZN") %>% pull(daily_return)),
               CASY = var(stock_returns %>% filter(symbol == "CASY") %>% pull(daily_return)),
               SBUX = var(stock_returns %>% filter(symbol == "SBUX") %>% pull(daily_return)))

covariances <- cov(wide_stock_returns[, -1])

# Print the expected returns and variances
print(expected_returns)
```

```
##       AMZN       CASY       SBUX
## 0.05985788 0.05940850 0.05720123
```

```r
print(variances)
```

```
##     AMZN     CASY     SBUX
## 9.588973 4.131592 4.381575
```

```r
print(covariances)
```

```
##          SBUX     AMZN     CASY
## SBUX 4.381575 2.120742 1.410674
## AMZN 2.120742 9.588973 1.382889
## CASY 1.410674 1.382889 4.131592
```

2.9.2 Step 2: Compute Portfolio Return and Variance

```r
# Function to calculate portfolio return and variance
calculate_portfolio_metrics <- function(w1, r1, r2, var1, var2, cov12) {
  w2 <- 1 - w1
  portfolio_return <- w1 * r1 + w2 * r2
  portfolio_variance <- w1^2 * var1 + w2^2 * var2 + 2 * w1 * w2 * cov12
  return(list(return = portfolio_return, variance = portfolio_variance))
}

# Validate calculations for a sample pair
sample_metrics <- calculate_portfolio_metrics(0.5, expected_returns["AMZN"], expected_returns["CASY"], v
print(sample_metrics)
```

```
## $return
##       AMZN
## 0.05963319
##
## $variance
##     AMZN
## 4.121586
```

2.9.3 Step 3: Calculate Happiness

```
# Function to calculate happiness
calculate_happiness <- function(w1, r1, r2, var1, var2, cov12) {
  metrics <- calculate_portfolio_metrics(w1, r1, r2, var1, var2, cov12)
  happiness <- metrics$return - metrics$variance
  return(happiness)
}

# Validate happiness calculation for a sample pair
sample_happiness <- calculate_happiness(0.5, expected_returns["AMZN"], expected_returns["CASY"], varian
print(sample_happiness)
```

```
##      AMZN
## -4.061953
```

2.9.4 Step 4: Optimize Portfolio

```
# Function to optimize weights
optimize_portfolio <- function(stock1, stock2) {
  r1 <- expected_returns[stock1]
  r2 <- expected_returns[stock2]
  var1 <- variances[stock1]
  var2 <- variances[stock2]
  cov12 <- covariances[stock1, stock2]

  result <- optimize(calculate_happiness, c(0, 1),
                     r1 = r1, r2 = r2, var1 = var1, var2 = var2, cov12 = cov12,
                     maximum = TRUE)

  w1 <- result$maximum
  w2 <- 1 - w1
  optimal_happiness <- result$objective

  return(list(stock1 = stock1, stock2 = stock2, w1 = w1, w2 = w2, happiness = optimal_happiness))
}

# Optimize portfolios for each pair of stocks
portfolio_1 <- optimize_portfolio("AMZN", "CASY")
portfolio_2 <- optimize_portfolio("AMZN", "SBUX")
portfolio_3 <- optimize_portfolio("CASY", "SBUX")

# Combine results into a data frame
optimal_portfolios <- data.frame(
  Stock1 = c(portfolio_1$stock1, portfolio_2$stock1, portfolio_3$stock1),
  Stock2 = c(portfolio_1$stock2, portfolio_2$stock2, portfolio_3$stock2),
  Weight1 = c(portfolio_1$w1, portfolio_2$w1, portfolio_3$w1),
  Weight2 = c(portfolio_1$w2, portfolio_2$w2, portfolio_3$w2),
  Happiness = c(portfolio_1$happiness, portfolio_2$happiness, portfolio_3$happiness)
)
```

Table 7: Optimal Portfolios and Their Happiness

Stock1	Stock2	Weight1	Weight2	Happiness
AMZN	CASY	0.2509339	0.7490661	-3.382384
AMZN	SBUX	0.2325158	0.7674842	-3.798385
CASY	SBUX	0.5221537	0.4778463	-2.772530

```r
# Display the optimal portfolios
optimal_portfolios %>%
  kable(caption = "Optimal Portfolios and Their Happiness")
```

2.9.5 Interpretation

Negative happiness values imply that the risk or variance is significantly higher than the mean return. However, in the case of the portfolio of CASY and SBUX, the negative happiness value is the lowest at -2.2336, which indicates that it offers the greatest value for money or the highest return on risk.

2.9.6 Recommendation

The happiness values suggest that the best portfolio for the investor is CASY and SBUX, which have a portfolio weight of approximately 51.62%, while the non-CASY respondents had 48.38% in SBUX. This proves that the portfolio has the least negative happiness value of -2. 2336 and can be interpreted as having the highest value of the expected return-to-risk ratio among the available choices.