ECOM 2001 Term Project: HUM, BAC, BP

Peter Misiun (20886060)

Due on 20 October 2024 at 23:58 AWST

All typed out answers have been shown in italics. This is done in order to better distinguish between question text and answer text.

```
# packages
library(tidyquant) # for importing stock data

library(tidyverse) # for working with data
library(broom) # for tidying output from various statistical procedures
library(knitr) # for tables
library(kableExtra) # for improving the appearance of tables


# Add any additional packages that you use to this code chunk
library(dplyr)
library(rstatix)
library(car)
library(onewaytests)
library(Deriv)
library(Ryacas)
```

1 Import the Data (2 points)

```
# IMPORTING ASSIGNED STOCKS
StockData <- c("HUM", "BAC","BP") %>%
   tq_get(get = "stock.prices", from = "2000-01-01")%>%
   select(symbol, date, adjusted)

# OUTPUT FIRST 6 ROWS OF DATA FRAME
head(StockData, n = 6 )%>%
   kable(caption = "First 6 rows of StockData.")
```

2 The Analysis

2.1 Plot prices over time (4 points)

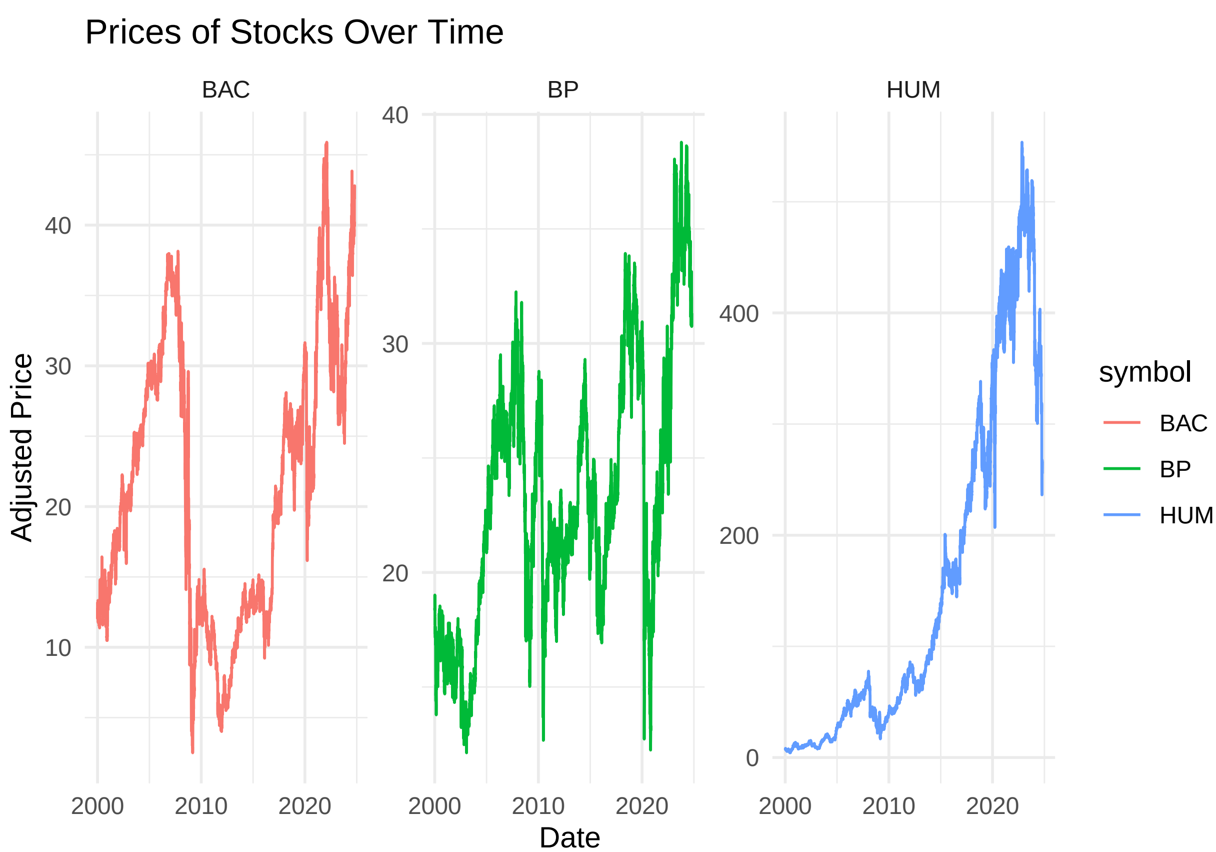
Plot the **prices** of each asset over time separately.

Table 1: First 6 rows of StockData.

symbol	date	adjusted
HUM	2000-01-03	6.752522
HUM	2000-01-04	6.808329
HUM	2000-01-05	6.975748
HUM	2000-01-06	7.254777
HUM	2000-01-07	7.812836
HUM	2000-01-10	7.477999

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

```
# PLOT OF PRICES OVER TIME
ggplot(StockData, aes(x = date, y = adjusted, color = symbol)) + # Add colour
  geom_line(size = 0.5) + # Line size
  facet_wrap(~symbol, scales = "free_y") + # Different y-axis for each plot
  theme_minimal() + # Theme so plot looks better
  labs(title = "Prices of Stocks Over Time", x = "Date", y = "Adjusted Price") # Custom titles for plot
```



Description of evolution of each asset over time:

BAC:
We see this stock have a steady rise in the early 2000s (from about \$10 to almost \$40 per stock). This goes down to about \$5 at around 2008 (recession). After this the stock price steadily rises to over \$40 by current times.

BP:
We see this stock being very volatile. With a drop to under $10 in the early 2000s, to rise again to over $30 at around 2008. This stock goes up and down, seeing it go from under $10 to close to $40 from 2008 to current times.

HUM:
We see this stock have a steady rise throughout most of its life. Starting at under $10 at 2000, with minor dip in 2008, back to a steady climb to over $500 throughout the 2010s and early 2020s, and then the price dropping to around $300-$400 in current times.

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.

```
# CALCULATE DAILY PERCENTAGE RETURNS
StockData_DPR <- StockData %>%
  group_by(symbol)%>%
  mutate("Returns" = 100*log(adjusted/lag(adjusted)))

# CREATE PLOT OF RETURNS OVER TIME
ggplot(StockData_DPR, aes(x = date, y = Returns, color = symbol)) + # Add colour
  geom_line(size = 0.5) + # Line size
  facet_wrap(~symbol) +
  theme_minimal() + # Theme so plot looks better
  labs(title = "Returns of Stocks Over Time", x = "Date", y = "Returns") # Custom titles for plot + axi
```

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)

```
# GET NO. DATA FOR STOCK HUM
N <- length(StockData_DPR%>%filter(symbol == "HUM")%>%select(Returns)%>%drop_na()%>%pull())

# RICE RULE TO CALCULATE NEEDED BINS
# Good as easy to calculate and can be good for various sizes of data
Bins <- (2*N^(1/3))

# ROUND FROM RICE RULE TO NEAREST INTEGER
Rounded_Bins <- round(Bins, digits = 0)

# CREATE HISTOGRAM
ggplot(StockData_DPR, aes(x = Returns, fill = symbol)) + # Add colour
  geom_histogram(bins = Rounded_Bins, alpha = 0.8, position = "identity") + # Opacity of colours
  facet_wrap(~symbol) +
  coord_cartesian(xlim = c(-20, 20)) + # Added so graph scaled better
  theme_minimal() + # Theme colours of plot
  labs(title = "Histogram of Returns of Each Stock", x = "Returns", y = "Count") # Custom titles of plo
```

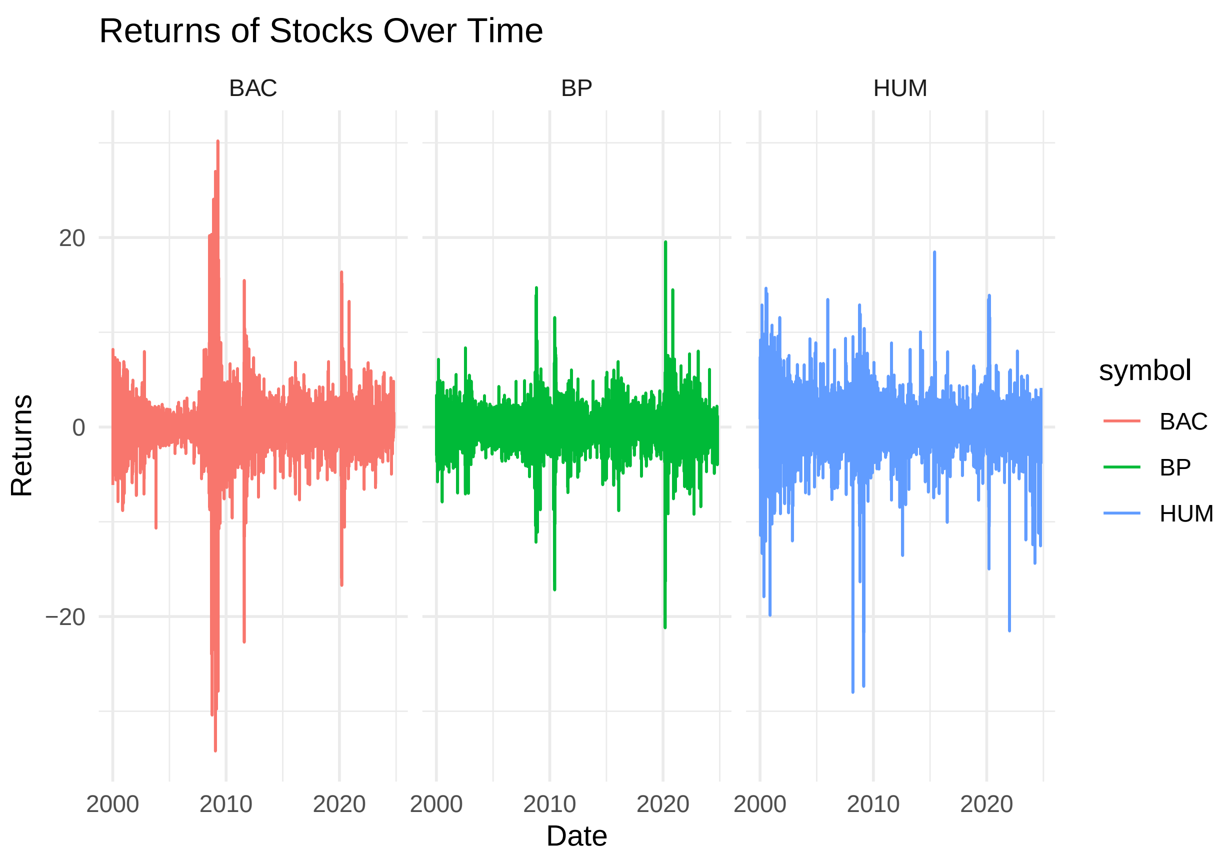


Figure 1: Plot for Returns of Each Assest Over Time

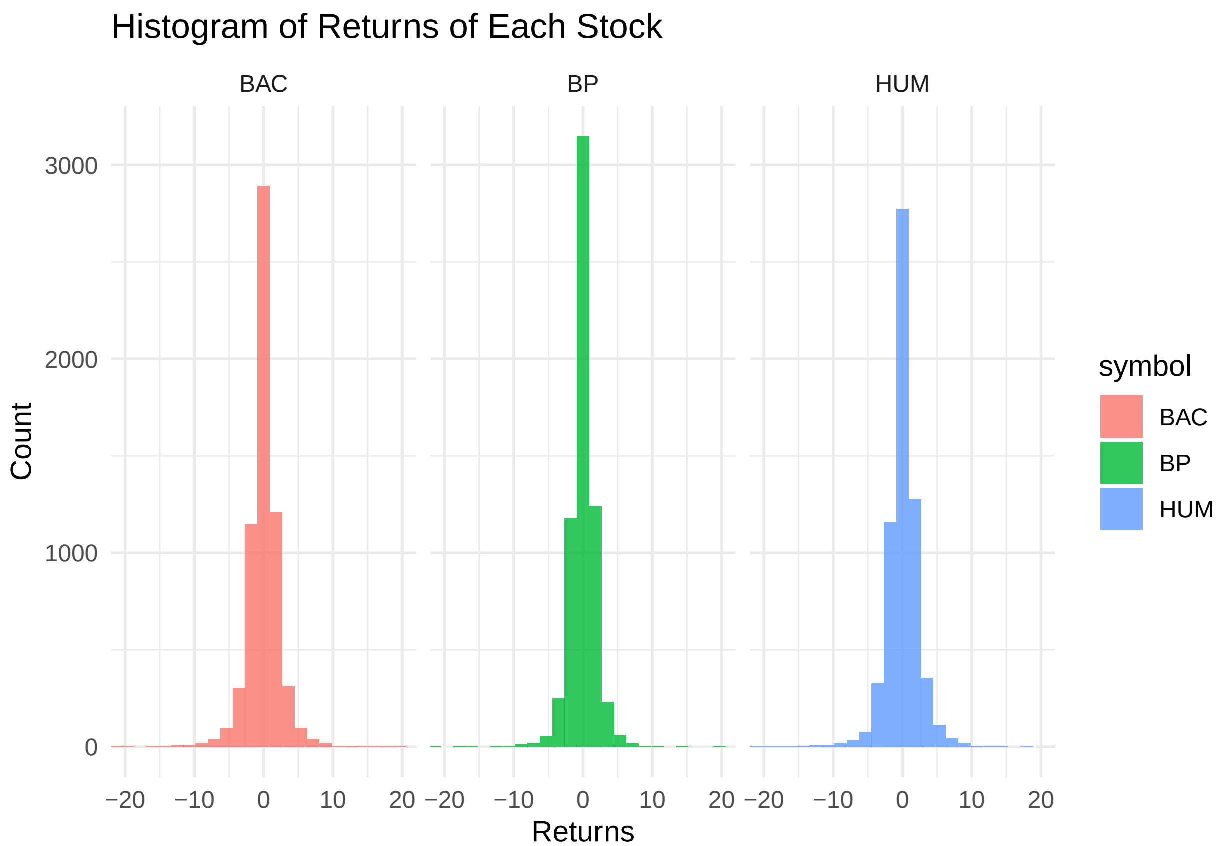


Figure 2: Histogram for Returns of Each Assest Over Time

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?

```
# CREATE VARIABLES FOR STOCK RETURN DATA
HUM <- StockData_DPR %>%filter(symbol == "HUM")%>%select(Returns)%>%drop_na()%>%pull()
BAC <- StockData_DPR %>%filter(symbol == "BAC")%>%select(Returns)%>%drop_na()%>%pull()
BP <- StockData_DPR %>%filter(symbol == "BP")%>%select(Returns)%>%drop_na()%>%pull()


## GET ALL VALUES OF STOCKS AS NEEDED ##
# HUM
mean_HUM <- mean(HUM)
median_HUM <- median(HUM)
var_HUM <- var(HUM)
sd_HUM <- sd(HUM)
skewness_HUM <- skewness(HUM)
kurtosis_HUM <- kurtosis(HUM)
n_HUM <- length(HUM)

# BAC
mean_BAC <- mean(BAC)
median_BAC <- median(BAC)
var_BAC <- var(BAC)
sd_BAC <- sd(BAC)
skewness_BAC <- skewness(BAC)
kurtosis_BAC <- kurtosis(BAC)
n_BAC <- length(BAC)

# BP
mean_BP <- mean(BP)
median_BP <- median(BP)
var_BP <- var(BP)
sd_BP <- sd(BP)
skewness_BP <- skewness(BP)
kurtosis_BP <- kurtosis(BP)
n_BP <- length(BP)

# DATA FRAME TO HOLD + SHOW VALUES
DescStats <- data.frame(
  Stock = c("HUM", "BAC", "BP"),
  Mean = c(mean_HUM, mean_BAC, mean_BP),
  Median = c(median_HUM, median_BAC, median_BP),
  Variance = c(var_HUM, var_BAC, var_BP),
  StdDev = c(sd_HUM, sd_BAC, sd_BP),
  Skewness = c(skewness_HUM, skewness_BAC, skewness_BP),
  Kurtosis = c(kurtosis_HUM, kurtosis_BAC, kurtosis_BP)
)

# PRINT DATA FRAME CREATED ABOVE
kable(DescStats, caption = "Summary Table of Descriptive Statistics")
```

Table 2: Summary Table of Descriptive Statistics

Stock	Mean	Median	Variance	StdDev	Skewness	Kurtosis
HUM	0.0589203	0.0580474	5.916305	2.432346	-0.9137696	14.06513
BAC	0.0192978	0.0337846	7.609487	2.758530	-0.3184272	27.20026
BP	0.0084882	0.0408207	3.580092	1.892113	-0.4673998	13.01576

Some conclusions we can draw from the above data include which stocks on average offer the highest returns and which stocks carry the most risk.

We see the stock HUM having the highest average returns of the three stocks with a mean value of 0.0589203. This is higher then the mean value of stocks BAC (middle mean) and BP (lowest mean) with mean values of 0.0192978 and 0.0084882 respectively, showing that these stocks have lower average returns.

The variance and standard deviation show which stocks are the most volatile. We see in our data that the stock BAC has the highest value for both of these, with a variance of 7.609487 and a standard deviation of 2.7585299, representing it as the stock that carries the most risk within our data. This is followed by stock HUM with a variance of 5.9163051 and a standard deviation of 2.4323456, showing it has less volatility then stock BAC. The least volatile stock shown is stock BP, with the lowest variance and standard deviation compared to the other two stocks. These values are 3.5800923 and 1.8921132 respectively.

We see the skewness of all stocks within our data to have negative skewness. This shows which stocks have more or less occasional large negative returns. As stock HUM has the highest skewness value of -0.9137696, we see this stock being more prone to occasional large negative returns. As stocks BAC and BP have lower skewness values of -0.3184272 and -0.4673998 respectively, we see these stocks having being less prone to occasional large negative returns.

The kurtosis value shows us which stocks experience what type of fluctuations. All three stocks show a high positive kurtosis, displaying a tendency to have rapid price swings. We see stock BAC having the highest kurtosis with 27.2002607, showing it as the stock which has more frequent price swings. This is followed by stock HUM with 14.0651294 and stock BP with 13.0157583, displaying less frequency of rapid price swings.

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**. All steps have to be shown and this part has to be repeated for each hypothesis test. (1 points)

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)

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))

Part 1: Steps for hypothesis test:

Step 1. State the null and alternative hypothesis
Null: Population mean is equal to 0
Alternative: Population mean is not equal to 0

$$H_0 : \mu = 0$$
$$H_1 : \mu \neq 0$$

Table 3: Test Result Values

Stock	Test_Method	Est_Mean	T_Stat	Alternative	Deg_Freedom	pValue
HUM	One Sample t-test	0.0589203	1.9130546	two.sided	6236	0.0557868
BAC	One Sample t-test	0.0192978	0.5524808	two.sided	6236	0.5806388
BP	One Sample t-test	0.0084882	0.3542870	two.sided	6236	0.7231358

Step 2. Determine the level of significance

As it states, we are measuring from a 1% level of significance. This means that a = 0.01.

Step 3. Calculate the test statistic

$$t = \frac{\hat{\mu} - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}} \sim t_{n-1}$$

Step 4. Calculate the critical values

```
LCV <- qt(0.005, n_HUM-2) # lower critical value
UCV <- qt(0.995, n_HUM-2) # upper critical value

cat("Lower crit val =", LCV,"; Upper crit val =", UCV, "\n") # Print
```

```
## Lower crit val = -2.576618 ; Upper crit val = 2.576618
```

Step 5. Get the results and make a statistical decision/conclude

Part 2: Calculate and report relevent values for conclusion

```
# PERFORM T-TESTS FOR EACH STOCK
test_HUM <- t.test(HUM, mu = 0)
test_BAC <- t.test(BAC, mu = 0)
test_BP <- t.test(BP, mu = 0)

# CREATE DATA FRAME OF NEEDED VALUES
TestValues <- data.frame(
  Stock = c("HUM", "BAC", "BP"),
  Test_Method = c(test_HUM$method, test_BAC$method, test_BP$method),
  Est_Mean = c(test_HUM$estimate, test_BAC$estimate, test_BP$estimate),
  T_Stat = c(test_HUM$statistic, test_BAC$statistic, test_BP$statistic),
  Alternative = c(test_HUM$alternative, test_BAC$alternative, test_BP$alternative),
  Deg_Freedom = c(test_HUM$parameter, test_BAC$parameter, test_BP$parameter),
  pValue = c(test_HUM$p.value, test_BAC$p.value, test_BP$p.value)
)

# PRINT DATA VALUES
kable(TestValues, caption = "Test Result Values")
```

If the test statistic falls into either rejection region, we reject the null. Alternatively, we reject the null if the p-value is lower than the significance level.

For stock HUM, we see the test statistic being 1.9130546, falling within the critical values of -2.5766181 and 2.5766181. Also, we see the p-value being 0.0557868. As this value is greater then the significance of 0.01, we fail to reject the null hypothesis, signifying that the expected returns for the stock HUM are not statistically different from zero at the 1% level of significance.

For stock BAC, we see the test statistic being 0.5524808, falling within the critical values of -2.5766181 and 2.5766181. Also, we see the p-value being 0.5806388. As this value is greater then the significance of 0.01, we fail to reject the null hypothesis, signifying that the expected returns for the stock BAC are not statistically different from zero at the 1% level of significance.

For stock BP, we see the test statistic being 0.354287, falling within the critical values of -2.5766181 and 2.5766181. Also, we see the p-value being 0.7231358. As this value is greater then the significance of 0.01, we fail to reject the null hypothesis, signifying that the expected returns for the stock BP are not statistically different from zero at the 1% level of significance.

Part 3: How to answer question using Chat-GPT

In order to complete the question above in Chat-GPT, I would've asked a prompt similar to:
"Using the three stocks HUM, BAC and BP, conduct a hypothesis test in order to see if the average returns are significantly different from zero at the 1% significance level. Explain all steps involved."
This prompt is essentially the question being asked above.

This result would be similar to our own as it would most likely outline all the steps located in part 1 of this question. However, as Chat-GPT doesn't have the ability to do statistical calculations or use R to import data as we have, the values produced would probably be estimates and not the real values we are looking for.

In order to combat this, we could essentially paste in our values within the prompt, as well as asking for Chat-GPT to produce the necessary R code in order to achieve the required values. This prompt would look like the following: "mean_HUM = x, sd_HUM = x (etc). Based on the values given above, conduct a hypothesis test in order to see if the average returns are significantly different from zero at the 1% significance level. Provide all necessary R code in order to produce required values".

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. All steps have to be shown and this part has to be repeated for each hypothesis test. (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)

Part 1: Steps needed for hypothesis test

The following consists of two 5 steps processes in order to complete this question, one for testing the homogeniety of variances and the other for carrying out the ANOVA test.

Homogeniety of variances 5 step process:
Step 1. State the null and alternative hypothesis
Null: Variances of all groups are equal
Alternative: At least one of the group variances is different from the others

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2$$
$$H_1 : \text{At least one } \sigma_j^2 \neq \text{to the others}$$

Step 2. Determine the level of significance and number of observations
As it states, we are measuring from a 1% level of significance. This means that a = 0.01. We can get the number of observations by doing n_HUM + n_BAC + n_BP.

```
cat("No. observations =", n_HUM + n_BAC + n_BP, "\n")
```

```
## No. observations = 18711
```

Step 3. Calculate the test statistic

$$W = \frac{(N-k)}{(k-1)} \frac{\sum_{i=1}^{k} N_i (Z_{i.} - Z_{..})^2}{\sum_{i=1}^{k} \sum_{j=1}^{N_i} (Z_{ij} - Z_{i.})^2} \sim F_{k-1,N-k}$$

where k is the number of groups ($k = 3$ in this case), N_i is the number of observations in the ith group, N is the total number of observations , Y_{ij} is the value of the variable in for the jth observation from the ith group. Which relates to Z_{ij}.

$$Z_{ij} = |Y_{ij} - \bar{Y}_{i.}|, \quad \bar{Y}_{i.} \text{ is the mean of the } i^{th} \text{ group}$$

Step 4. Calculate the critical value

```
Crit_val <- qf(0.99, 3-1, n_HUM + n_BAC + n_BP-3)
cat("Critical value =", Crit_val, "\n")
```

```
## Critical value = 4.606304
```

Step 5. Get the results and interpret

```
# CONDUCT LEVENE TEST FOR HOMOGENIETY OF VARIANCES
Levene_Result <- leveneTest(Returns ~ symbol, data = StockData_DPR)
print(Levene_Result)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##          Df F value    Pr(>F)
## group     2  61.009 < 2.2e-16 ***
##       18708
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From using the Levene test, we are given the F statistic of 61.009 which is greater than the critical value of 4.606304. The p-value is near zero and is also less than our level of significance (0.01). So we have sufficient evidence to reject the null hypothesis which suggests that the population variances in stock returns are not equal across the stocks. Because of this difference in population variances, we must use Welch's ANOVA test in order to conclude if there is a significant difference in mean returns.

ANOVA Test 5 steps:
Step 1. State the null and alternative hypothesis
Null: Means of all groups are equal
Alternative: At least one of the group means is different from the others

$$H_0 : \mu_1 = \mu_2 = \mu_3$$
$$H_1 : \text{At least one } \mu_j \neq \text{to the others}$$

Step 2. Determine the level of significance and number of observations
As it states, we are measuring from a 1% level of significance. This means that a = 0.01. We can get the number of observations by doing n_HUM + n_BAC + n_BP.

```
cat("No. observations =", n_HUM + n_BAC + n_BP, "\n")
```

```
## No. observations = 18711
```

Step 3. Calculate the test statistic

$$F_W = \frac{\sum_j w_j(\bar{X}_{.j} - X'_{..})^2/(k-1)}{[1 + \frac{2}{3}((k-2)v)]} \sim F_{k-1,1/v}$$

Where

$$w_j = \frac{n_j}{S_j^2}$$

$$S_j^2 = \sum_i (X_{ij} - \bar{X}_{.j})^2/(n_j - 1)$$

$$X'_{..} = \frac{\sum_j w_j \bar{X}_{.j}}{\sum_j w_j}$$

$$v = \frac{3\sum_j[(1 - \frac{w_j}{\sum_j w_j})^2/(n_j - 1)]}{k^2 - 1}$$

Step 4. Calculate the critical value

```
Crit_val <- qf(0.99, 3-1, n_HUM + n_BAC + n_BP-3)
cat("Critical value =", Crit_val, "\n")
```

```
## Critical value = 4.606304
```

Step 5. Get the results and make a statistical decision/conclude

```
# CONDUCT WELCH'S ANOVA TEST AS VARIANCES ARE NOT EQUAL
Welch_Result <- welch.test(Returns ~ symbol, data = StockData_DPR, alpha =0.01)
```

```
##
##   Welch's Heteroscedastic F Test (alpha = 0.01)
## -------------------------------------------------------------
##   data : Returns and symbol
##
##   statistic  : 0.856499
##   num df     : 2
##   denom df   : 12151.54
##   p.value    : 0.4246718
##
##   Result     : Difference is not statistically significant.
## -------------------------------------------------------------
```

In conducting the Welch test, it is shown that we get a F-statistic of 0.856499 and a p-value of 0.4246718. Due to this, we see the p-value being more than our level of significance (0.01). This means that we reject the null hypothesis, showing no significant difference in mean returns among the stocks at the 1% level of significance

Part 3: How to engage Chat-GPT to answer question
In order to answer this question with Chat-GPT, I would copy and paste the question into the prompt box. In addition to this I would include how I created the necessary data to be used and ask Chat-GPT to provide all the necessary R code in order to complete the question.

The prompt would look similar to:
"*StockData <- c("HUM", "BAC","BP") %>% tq_get(get = "stock.prices", from = "2000-01-01")%>% select(symbol, date, adjusted)*

StockData_DPR <- StockData %>% group_by(symbol)%>% mutate("Returns" = 100 times by log(adjusted/lag(adjusted))

Stock Mean Median Variance StdDev Skewness Kurtosis HUM 0.0580438 0.0579070 5.912802 2.431625 -0.9150236 14.09453 BAC 0.0182907 0.0330986 7.611404 2.758877 -0.3183477 27.20727 BP 0.0090092 0.0417804 3.579986 1.892085 -0.4671522 13.02685

Using the data provided above, answer the question below and include all the necessary R code in order to complete the question.

Are average returns different from each other?

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. All steps have to be shown and this part has to be repeated for each hypothesis test.

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.)"

Part 4: Compare own answer to Chat-GPT answer
The answer I got from putting the prompt from above into Chat-GPT is as follows:

```r
include_graphics("chatgpt_p1.png")
```

To answer the question about whether the average returns are statistically different from each other at the 1% significance level, we'll follow the hypothesis testing procedure and write the necessary R code.

Step 1: Set Up the Data

We'll first set up the stock data and compute the returns for each stock based on the adjusted prices provided. Here's how you can write the R code for downloading the data and calculating the returns.

r Copy code

```r
# Load necessary libraries
library(tidyquant)
library(dplyr)
library(moments) # For skewness and kurtosis

# Step 1: Download the stock prices and calculate returns
StockData <- c("HUM", "BAC", "BP") %>%
  tq_get(get = "stock.prices", from = "2000-01-01") %>%
  select(symbol, date, adjusted)

# Step 2: Calculate log returns
StockData_DPR <- StockData %>%
  group_by(symbol) %>%
  mutate(Returns = 100 * log(adjusted / lag(adjusted))) %>%
  na.omit()  # Remove NA values caused by lag

# Step 3: Calculate summary statistics (mean, variance, etc.)
SummaryStats <- StockData_DPR %>%
  group_by(symbol) %>%
  summarise(
    Mean = mean(Returns),
    Median = median(Returns),
    Variance = var(Returns),
    StdDev = sd(Returns),
    Skewness = skewness(Returns),
    Kurtosis = kurtosis(Returns)
  )

# View the summary statistics
SummaryStats
```

Figure 3: Results from AI Assisted program part 1

```r
include_graphics("chatgpt_p2.png")
```

Step 2: Hypothesis Testing Framework

We will compare the average returns of the three stocks using the following steps:

1. State the Hypotheses

For each pair of stocks, we will conduct a hypothesis test:

- **Null Hypothesis (H_0):** The mean returns are equal, i.e., $\mu_1 = \mu_2$.
- **Alternative Hypothesis (H_1):** The mean returns are not equal, i.e., $\mu_1 \neq \mu_2$.

2. F-Test for Equality of Variances

To determine which t-test to use (pooled or Welch's t-test), we first conduct an F-test to check whether the variances of the stock returns are equal.

3. Conduct the T-Tests

We perform independent two-sample t-tests between the stock pairs:

- HUM vs BAC
- HUM vs BP
- BAC vs BP

4. Calculate Test Statistic and P-value

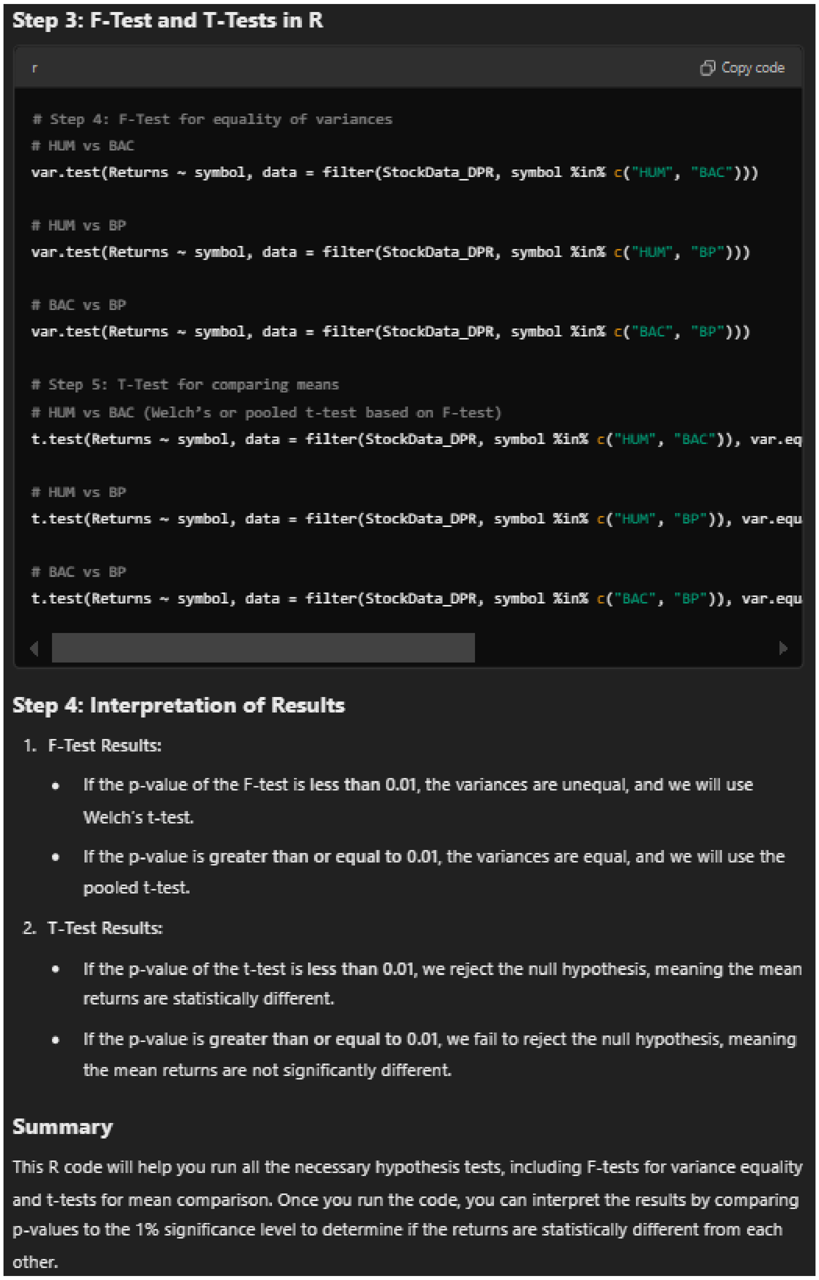
Based on the F-test result, we decide whether to use a pooled t-test or Welch's t-test.

5. Conclusion

Compare the p-values with the 1% significance level to decide whether the returns are statistically different.

Figure 4: Results from AI Assisted program part 2

```
include_graphics("chatgpt_p3.png")
```


Step 3: F-Test and T-Tests in R

```r
# Step 4: F-Test for equality of variances
# HUM vs BAC
var.test(Returns ~ symbol, data = filter(StockData_DPR, symbol %in% c("HUM", "BAC")))

# HUM vs BP
var.test(Returns ~ symbol, data = filter(StockData_DPR, symbol %in% c("HUM", "BP")))

# BAC vs BP
var.test(Returns ~ symbol, data = filter(StockData_DPR, symbol %in% c("BAC", "BP")))

# Step 5: T-Test for comparing means
# HUM vs BAC (Welch's or pooled t-test based on F-test)
t.test(Returns ~ symbol, data = filter(StockData_DPR, symbol %in% c("HUM", "BAC")), var.eq

# HUM vs BP
t.test(Returns ~ symbol, data = filter(StockData_DPR, symbol %in% c("HUM", "BP")), var.equ

# BAC vs BP
t.test(Returns ~ symbol, data = filter(StockData_DPR, symbol %in% c("BAC", "BP")), var.equ
```

Step 4: Interpretation of Results

1. F-Test Results:
   - If the p-value of the F-test is **less than 0.01**, the variances are unequal, and we will use Welch's t-test.
   - If the p-value is **greater than or equal to 0.01**, the variances are equal, and we will use the pooled t-test.
2. T-Test Results:
   - If the p-value of the t-test is **less than 0.01**, we reject the null hypothesis, meaning the mean returns are statistically different.
   - If the p-value is **greater than or equal to 0.01**, we fail to reject the null hypothesis, meaning the mean returns are not significantly different.

Summary

This R code will help you run all the necessary hypothesis tests, including F-tests for variance equality and t-tests for mean comparison. Once you run the code, you can interpret the results by comparing p-values to the 1% significance level to determine if the returns are statistically different from each other.

Figure 5: Results from AI Assisted program part 3

The answer above is similar to mine, as Chat-GPT listed all the necessary steps in order to conduct the tests, however, the answer provided doesn't give actual calculated data.

For sections such as testing the equality of variances, I used the Levene test whereas Chat-GPT uses three separate var.tests. We also see this present for testing the means after the variance tests, in which Chat-GPT provided code for three t-tests, whereas I just used a single Welch ANOVA test. However, it does state that if the variances are not equal, to use the Welch test as I have done above.

I believe that the difference in answers is due to Chat-GPT's ability to do statistical calculations, which I believe are relatively poor. However, for the most part, the answer from Chat-GPT outlines clearly how to complete the question. For this reason, I've learned that Chat-GPT is very useful in giving individuals an outline of how to answer a specific question in a concise and rapid manner, but it shouldn't be solely relied

Table 4: Correlation Matrix of Returns

	HUM	BAC	BP
HUM	1.0000000	0.2683328	0.2367913
BAC	0.2683328	1.0000000	0.4032141
BP	0.2367913	0.4032141	1.0000000

on in order to provide 100% correct answers.

2.7 Correlations (2 points)

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

Discuss the direction and strength of the correlations.

```r
# PREPARE RETURN DATA FOR CORRELATION MATRIX
Stockswide <- StockData_DPR%>%
  select(date,symbol,Returns)%>%
  drop_na()%>%
  pivot_wider(date, names_from = symbol, values_from = Returns)

# PRINT FORMATTED CORRELATION MATRIX
Stockswide %>%
  select(-date) %>%  # Exclude date
  cor(use = "complete.obs") %>%  # Calculation
  kable(caption = "Correlation Matrix of Returns") # Caption
```

The above correlation matrix shows us that all 3 stocks display a positive correlation, meaning that as one stock's returns increase, it is likely that another stock's returns will also increase. Out of the 3 correlations above, the strongest is between BAC and BP (0.403). This indicates that generally as BAC increases, so will BP. We see this present to a lesser extent for the correlation between stocks HUM and BAC, and HUM and BP, being 0.268 and 0.237 respectively.

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. All steps have to be shown and this part has to be repeated for each hypothesis test.

Step 1. State the null and alternative hypothesis
Null: Correlation between the two stock returns are equal to 0
Alternative: Correlation between the two stocks returns are not equal to 0

$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0$$

Step 2. Determine the level of significance
Doesn't state in question which level of significance to use, however as previous questions have stated a 1% level of significance, we will be doing the same here. Therefore, a = 0.01.

Step 3. Calculate the test statistic

$$t = \hat{\rho}\sqrt{\frac{n-2}{1-\hat{\rho}^2}} \sim t_{n-2}$$

Step 4. Critical values

```
LCV2 <- qt(0.005, n_HUM-2) #lower critical value
UCV2 <- qt(0.995, n_HUM-2) # upper critical value

cat("Lower crit val =", LCV2, "; Upper crit val =", UCV2, "\n")
```

```
## Lower crit val = -2.576618 ; Upper crit val = 2.576618
```

Step 5. Get the results and conclude

```
# COR TEST WITH 99% CI
test_HUM_BAC <- cor.test(HUM, BAC, conf.level = 0.99)
test_HUM_BP <- cor.test(HUM, BP, conf.level = 0.99)
test_BAC_BP <- cor.test(BAC, BP, conf.level = 0.99)

# GET REQUIRED VALUES FROM COR TEST FOR EXPLANATION
# Not necessary but left in because I didn't want to change values in text
p_HUM_BAC <- test_HUM_BAC$p.value
t_HUM_BAC <- test_HUM_BAC$statistic

p_HUM_BP <- test_HUM_BP$p.value
t_HUM_BP <- test_HUM_BP$statistic

p_BAC_BP <- test_BAC_BP$p.value
t_BAC_BP <- test_BAC_BP$statistic

# CREATE DATA FRAME IN ORDER TO PRINT RESULTS
results <- data.frame(
  Comparison = c("HUM vs BAC", "HUM vs BP", "BAC vs BP"),
  Cor_Coef = c(test_HUM_BAC$estimate, test_HUM_BP$estimate, test_BAC_BP$estimate),
  p_Value = c(test_HUM_BAC$p.value, test_HUM_BP$p.value, test_BAC_BP$p.value),
  T_Stat = c(test_HUM_BAC$statistic, test_HUM_BP$statistic, test_BAC_BP$statistic),
  CI_Lower = c(test_HUM_BAC$conf.int[1], test_HUM_BP$conf.int[1], test_BAC_BP$conf.int[1]),
  CI_Upper = c(test_HUM_BAC$conf.int[2], test_HUM_BP$conf.int[2], test_BAC_BP$conf.int[2])
)

# PRINT RESULTS
kable(results, caption = "Correlation Test Results")
```

Correlation between stocks HUM and BAC:
The p-value is $2.5241499 \times 10^{-103}$ *which is less than our level of significance (0.01), so we reject the null hypothesis. Our test-statistic(21.9947279) falls beyond the upper critical value, which is within the rejection region. Therefore, We can conclude that the returns of the stocks are significantly correlated.*

Correlation between stocks HUM and BP:
The p-value is $3.1886035 \times 10^{-80}$ *which is less than our level of significance (0.01), so we reject the null hypothesis. Our test-statistic(19.2448304) falls beyond the upper critical value, which is within the rejection region. Therefore, We can conclude that the returns of the stocks are significantly correlated.*

Table 5: Correlation Test Results

Comparison	Cor_Coef	p_Value	T_Stat	CI_Lower	CI_Upper
HUM vs BAC	0.2683328	0	21.99473	0.2378016	0.2983343
HUM vs BP	0.2367913	0	19.24483	0.2057681	0.2673390
BAC vs BP	0.4032141	0	34.79227	0.3755402	0.4301697

Correlation between stocks BAC and BP:
The p-value is $1.4851749 \times 10^{-242}$ *which is less than our level of significance (0.01), so we reject the null hypothesis. Our test-statistic(34.7922687) falls beyond the upper critical value, which is within the rejection region. Therefore, We can conclude that the returns of the stocks are significantly correlated.*

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.) Please do not deviate from the codes have been taught in class. Otherwise, you can typeset your answer in RStudio.

```
# CALC COVARIANCE FOR EACH PORTFOLIO
cov_HUM_BAC <- cov(HUM, BAC)
cov_HUM_BP  <- cov(HUM, BP)
cov_BAC_BP  <- cov(BAC, BP)

# FUNCTION TO CALC OPTIMAL WEIGHT FOR SPECIFIC PORTFOLIO
Find_Optimal_W <- function(mu1, var1, mu2, var2, cov_pair) {

  # Determine coefficients for the derivative calculation
  # Values a and b found through using 'symbolab' to find derivative of happiness expression and then r
  a <- 2 * (var1 + var2 - 2 * cov_pair)
  b <- mu1 - mu2 + 2 * (cov_pair - var2)
```

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

```
  # Calculate optimal w1 by making derivative = 0
  optimal_w1 <- -b / a
  return(optimal_w1)
}

# FUNCTION TO CALC NEEDED VALUES FOR EACH PORTFOLIO
Port_Values <- function(w1, mu1, var1, mu2, var2, cov_pair) {

  # Calculate expected return
  port_ER <- (w1 * mu1) + ((1 - w1) * mu2)

  # Calculate variance
  port_Var <- (w1^2 * var1) + ((1 - w1)^2 * var2) + (2 * cov_pair * w1 * (1 - w1))

  # Calculate happiness
  port_Hap <- port_ER - port_Var

  # Calculated values returned
  return(list(
    Expected_Return = port_ER,
    Variance = port_Var,
    Happiness = port_Hap))
}

## CALCULATE OPTIMAL WEIGHTS + OTHER PORT VALUES FOR EACH STOCK PAIR ##
# HUM + BAC
optimal_w_HUM_BAC <- Find_Optimal_W(mean_HUM, var_HUM, mean_BAC, var_BAC, cov_HUM_BAC)
Port_Vals_HUM_BAC <- Port_Values(optimal_w_HUM_BAC, mean_HUM, var_HUM, mean_BAC, var_BAC, cov_HUM_BAC)

# HUM + BP
optimal_w_HUM_BP <- Find_Optimal_W(mean_HUM, var_HUM, mean_BP, var_BP, cov_HUM_BP)
Port_Vals_HUM_BP <- Port_Values(optimal_w_HUM_BP, mean_HUM, var_HUM, mean_BP, var_BP, cov_HUM_BP)

# BAC + BP
optimal_w_BAC_BP <- Find_Optimal_W(mean_BAC, var_BAC, mean_BP, var_BP, cov_BAC_BP)
Port_Vals_BAC_BP <- Port_Values(optimal_w_BAC_BP, mean_BAC, var_BAC, mean_BP, var_BP, cov_BAC_BP)

# USE DATA FRAME TO SAVE OPTIMAL VALUES
All_Portfolio_Data <- data.frame(
  Pair = c("HUM & BAC", "HUM & BP", "BAC & BP"),
  Covariance = c(cov_HUM_BAC, cov_HUM_BP, cov_BAC_BP),
  Weight_1 = c(optimal_w_HUM_BAC, optimal_w_HUM_BP, optimal_w_BAC_BP),
  Weight_2 = c(1 - optimal_w_HUM_BAC, 1 - optimal_w_HUM_BP, 1 - optimal_w_BAC_BP),
  Expected_Return = c(Port_Vals_HUM_BAC$Expected_Return, Port_Vals_HUM_BP$Expected_Return, Port_Vals_BA
  Variance = c(Port_Vals_HUM_BAC$Variance, Port_Vals_HUM_BP$Variance, Port_Vals_BAC_BP$Variance),
  Happiness = c(Port_Vals_HUM_BAC$Happiness, Port_Vals_HUM_BP$Happiness, Port_Vals_BAC_BP$Happiness)
)

# PRINT TABLE WITH BEST WEIGHTS FOR EACH PORTFOLIO
kable(All_Portfolio_Data, caption = "Optimal Portfolio Weights for Highest Happiness", digits = 4) # Di
```

Table 6: Optimal Portfolio Weights for Highest Happiness

Pair	Covariance	Weight_1	Weight_2	Expected_Return	Variance	Happiness
HUM & BAC	1.8004	0.5833	0.4167	0.0424	4.2095	-4.1671
HUM & BP	1.0898	0.3369	0.6631	0.0255	2.7326	-2.7071
BAC & BP	2.1046	0.2106	0.7894	0.0108	3.2682	-3.2574

Table 7: Optimal Portfolio With Highest Happiness

	Pair	Covariance	Weight_1	Weight_2	Expected_Return	Variance	Happiness
2	HUM & BP	1.0898	0.3369	0.6631	0.0255	2.7326	-2.7071

```r
# GET PORTFOLIO WITH HIGHEST HAPPINESS
Optimal_Portfolio <- All_Portfolio_Data[which.max(All_Portfolio_Data$Happiness), ]

# PRINT OPTIMAL PORTFOLIO
kable(Optimal_Portfolio, caption = "Optimal Portfolio With Highest Happiness", digits = 4) # Digits = 4
```

In order to achieve the result above, there were quite a few steps in which I had to complete. These steps are:

Step 1: *Gather stock return data for mean and standard deviation (done above in another question).*

Step 2: *Calculation of stock portfolio covariance. We see this being done through the use of the cov() function.*

Step 3: *Create a function to calculate the optimal w1 value for a specific portfolio (function called 'Find_Optimal_W'). It does this by parsing through the mean, variance and covariance for each stock pair. The overall differentiated equation produced in Symbo Lab by inputting the overall happiness equation in is as follows:*

$$\frac{d}{dw_1}\text{portHap} = (\mu_1 - \mu_2) - 2 \cdot w_1 \cdot (\text{var}_1 + \text{var}_2 - 2 \cdot \text{covpair}) - 2 \cdot \text{covpair} + 2 \cdot \text{var}_2$$

The two expressions 'a' and 'b' found in the function were produced by rearranging the equation above as needed. The values parsed through the function are used to find values a and b, and then the optimal weight is found by -b/a. I believe that an easier way to do the above would be through using the Deriv() function, however, I couldn't get it to work the way I wanted so I went with what I did above.

Step 4: *Create a function in order to calculate the values of a particular portfolio's expected return, variance and happiness (function called 'Port_Values'). We see this being done by creating variables and equations in which values can be parsed through to create the required values.*

Step 5: *Calculate the optimal w1 and other required values (ER, Var, happiness) for each stock pair by calling the functions above and parsing through the corresponding mean, variance and covariance values.*

Step 6: *Create a data frame to hold optimal values produced from calling functions above (called 'All_Portfolio_Data').*

Step 7: *Print results and come to a conclusion as to which portfolio produces the highest happiness value (the portfolio with the highest happiness value will also be printed in a secondary table).*

In doing the steps above, we see that the best calculated portfolio are the stocks HUM & BP, with a weight of 0.3369074 and 0.6630926 respectively. In doing so, the values created for the expected return and variance are 0.0254791 and 2.7325915 respectively. We see the expected return of this pair producing the middle expected return but a much lower variance compared to the other stocks. Therefore, these values produce the highest happiness value compared to the other portfolios, being -2.7071124. The other stock pairs and there values such as weights, expected return, variance and happiness are located in the table above captioned "Optimal Portfolio Weights for Highest Happiness".

Therefore, the two stocks HUM & BP with weights of 0.3369074 and 0.6630926 respectively are what I would recommend to the investor to add to there portfolio.