Coursework1

202353542 - Nisar Ahmed

2024-02-24

Part 1 - Optimization of the Portfolio using GA

Checking and installing necessary packages and libraries

Ensuring all necessary R packages are installed and loaded for portfolio analysis and optimization.

Part 1(a): Selection of Assets

- Selecting 10 diverse assets from different sectors to construct a well-rounded portfolio.
- This diversity aims to mitigate risk by spreading investments across diff market behaviors, including tech, healthcare, and energy sectors.

```
my_assets <- c("AAPL", "PFE", "BAC", "TSLA", "PG", "XOM", "NEE", "BA", "DD", "SPG")
```

Part 1(b): Data Retrieval and Pre-processing

- Fetching historical price data for the selected assets from 2019 to 2021.

Why COVID-19 Period?

- The 3 x years period from 2019 to 2021 is strategically chosen to encompass the market fluctuations due to the COVID-19 pandemic.This time-frame allows for the analysis of assets' performance through significant economic disruptions, providing insights into their resilience and potential for recovery.

Why slecting three years?

- Span of three years offers a balance between capturing recent market behaviors and ensuring enough data for meaningful analysis and visualization.The inclusion of this volatile period is crucial for optimizing a portfolio that can withstand and capitalize on market upheavals, making the analysis more relevant for current and future investing environments.

```
asset_prices <- list()
for(asset in my_assets) {
  getSymbols(asset, from = "2019-01-01", to = "2021-12-31", auto.assign = TRUE)
  asset_prices[[asset]] <- Cl(get(asset))
}
combined_prices <- do.call(merge, asset_prices)
combined_prices <- na.omit(combined_prices)
```

Calculating daily returns

- Combining and cleaning the asset price data to ensure a consistent and complete data-set for analysis. Daily returns are calculated to understand the assets' day-to-day performance, crucial for portfolio optimization.

```r
daily_returns <- ROC(combined_prices, type = "discrete")
daily_returns <- na.omit(daily_returns)
myRetData <- daily_returns
```

Part 1(d): Portfolio Optimization Using Genetic Algorithm

Defining a fitness function to maximize the portfolio's Sharpe Ratio, balancing risk and return

- The Sharpe Ratio is chosen as it is a widely used measure for calculating risk-adjusted return, enhancing portfolio performance.

```r
fitness_function_alpha <- function(weights, alpha=0.5, riskFreeRate=0.025) {
  normalizedWeights <- weights / sum(weights)
  portfolioReturn <- sum(colMeans(myRetData, na.rm = TRUE) * normalizedWeights) * 252
  portfolioRisk <- sqrt(t(as.matrix(normalizedWeights)) %*% cov(myRetData) %*% as.matrix(normalizedWeig
  SharpeRatio <- (portfolioReturn - riskFreeRate) / portfolioRisk
  return(SharpeRatio)
}
```

Running the genetic algorithm to find optimal portfolio weights

- Configuring and running the genetic algorithm with specified parameters to find the optimal asset weights
- Parameters such as population size and maximum iterations are tuned to efficiently search the solution space while balancing computational resources

```r
ga_result <- ga(type = "real-valued",
                fitness = function(weights) fitness_function_alpha(weights, alpha=0.5, riskFreeRate=0.02
                lower = rep(0, length(my_assets)),
                upper = rep(1, length(my_assets)),
                popSize = 100,
                maxiter = 500)

optimal_weights <- ga_result@solution
normalized_optimal_weights <- optimal_weights / sum(optimal_weights)
```

Detailed Evaluation for Future Performance

- Splitting the data-set into training and testing sets for a fair comparison
- Splitting the datas-et to train and test the portfolio, ensuring a robust evaluation of its future performance - historical data informing future decisions.

```r
split_date <- as.Date("2021-06-01")
training_data <- daily_returns[index(daily_returns) < split_date]
testing_data <- daily_returns[index(daily_returns) >= split_date]

# Re-optimizing the portfolio using only the training data
ga_result_training <- ga(type = "real-valued",
                         fitness = function(weights) fitness_function_alpha(weights, alpha=0.5, riskFre
                         lower = rep(0, length(my_assets)),
                         upper = rep(1, length(my_assets)),
                         popSize = 100,
                         maxiter = 500,
                         suggestions = normalized_optimal_weights)
```

```r
optimal_weights_training <- ga_result_training@solution
normalized_optimal_weights_training <- optimal_weights_training / sum(optimal_weights_training)
```

Part 1(e): Comparison with Other Portfolios

Balanced Portfolio

- Comparing the GA-optimized portfolio against balanced and randomly generated portfolios to evaluate its relative performance. This step assesses the effectiveness of the genetic algorithm in enhancing portfolio returns and reducing risk.

```r
balanced_weights <- rep(1 / length(my_assets), length(my_assets))
```

Generating several random portfolios for a broader comparison

```r
set.seed(789) # Ensuring reproducibility
random_weights_list <- replicate(100, runif(length(my_assets)))
random_weights_list <- apply(random_weights_list, 2, function(x) x / sum(x))

print("Optimal Weights:")
```

```
## [1] "Optimal Weights:"
```

```r
print(optimal_weights)
```

```
##             x1        x2         x3        x4        x5          x6       x7
## [1,] 0.9309314 0.1263145 0.01542832 0.8820605 0.3939164 0.02733761 0.373302
##               x8          x9         x10
## [1,] 0.006859139 0.005558074 0.005510643
```

```r
print("Normalized Optimal Weights:")
```

```
## [1] "Normalized Optimal Weights:"
```

```r
print(normalized_optimal_weights)
```

```
##             x1         x2          x3        x4       x5          x6        x7
## [1,] 0.3364141 0.04564674 0.005575389 0.3187535 0.142351 0.009879093 0.1349015
##               x8          x9         x10
## [1,] 0.002478712 0.002008542 0.001991401
```

```r
balanced_weights <- rep(1 / length(my_assets), length(my_assets))
print("Balanced Portfolio Weights:")
```

```
## [1] "Balanced Portfolio Weights:"
```

```r
print(balanced_weights)
```

```
##  [1] 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1
```

```r
print("Sample Random Portfolio Weights:")
```

```
## [1] "Sample Random Portfolio Weights:"
```

```r
print(random_weights_list[, 1])
```

```
##  [1] 0.208749150 0.027886794 0.003545340 0.176451376 0.146787547 0.006013949
##  [7] 0.170799243 0.049440480 0.107050301 0.103275820
```

Part 1(e): Evaluating Portfolio Performance - Evaluation of Optimized, Balanced, and Random Portfolios

- In this section, we evaluated the performance of optimized, balanced, and randomly generated portfolios.
- We defined a function to calculate annualized return, risk, and the Sharpe Ratio for each portfolio strategy using test data.
- This analysis helps us understand the effectiveness of genetic algorithm optimization in achieving superior risk-adjusted returns compared to simpler portfolio construction methods.

```
## Defining the updated calculate_performance function
calculate_performance <- function(weights, returns) {
  weights_vector <- as.numeric(weights) # Ensuring weights are a numeric vector

  # Converting returns to an xts object if not already done
  if(!is.xts(returns)) {
    returns <- as.xts(returns, order.by=index(returns))
  }

  # Calculating portfolio returns using the provided weights
  portfolio_returns <- Return.portfolio(R = returns, weights = weights_vector, rebalance_on = "years")

  # Calculating annualized return, annualized risk and Sharpe Ratio
  annualized_return <- annualReturn(portfolio_returns, scale = 252)
  annualized_risk <- sd(portfolio_returns) * sqrt(252)
  sharpe_ratio <- SharpeRatio.annualized(portfolio_returns, Rf = 0, scale = 252)
  return(c(annualized_return = as.numeric(annualized_return),
           annualized_risk = annualized_risk,
           sharpe_ratio = as.numeric(sharpe_ratio)))
}

# Using the performance calculation function to evaluate the optimized portfolio on test data
optimized_performance <- calculate_performance(normalized_optimal_weights_training, testing_data)

# Evaluating the performance of a balanced portfolio, where each asset is equally weighted, on test dat
balanced_performance <- calculate_performance(balanced_weights, testing_data)

# Applying the performance calculation across several randomly generated portfolios to assess their ave
random_performances <- apply(random_weights_list, 2, function(weights) calculate_performance(weights, t


# Printing the optimized and balanced portfolio performances
print("Optimized Portfolio Performance:")
```

```
## [1] "Optimized Portfolio Performance:"
```

```
print(optimized_performance)
```

```
## annualized_return   annualized_risk      sharpe_ratio
##         0.7373363         0.2225044         4.1660340
```

```
print("Balanced Performance:")
```

```
## [1] "Balanced Performance:"
```

```
print(balanced_performance)
```

```
## annualized_return   annualized_risk      sharpe_ratio
##        -1.1866396         0.1456855         2.7831704
```

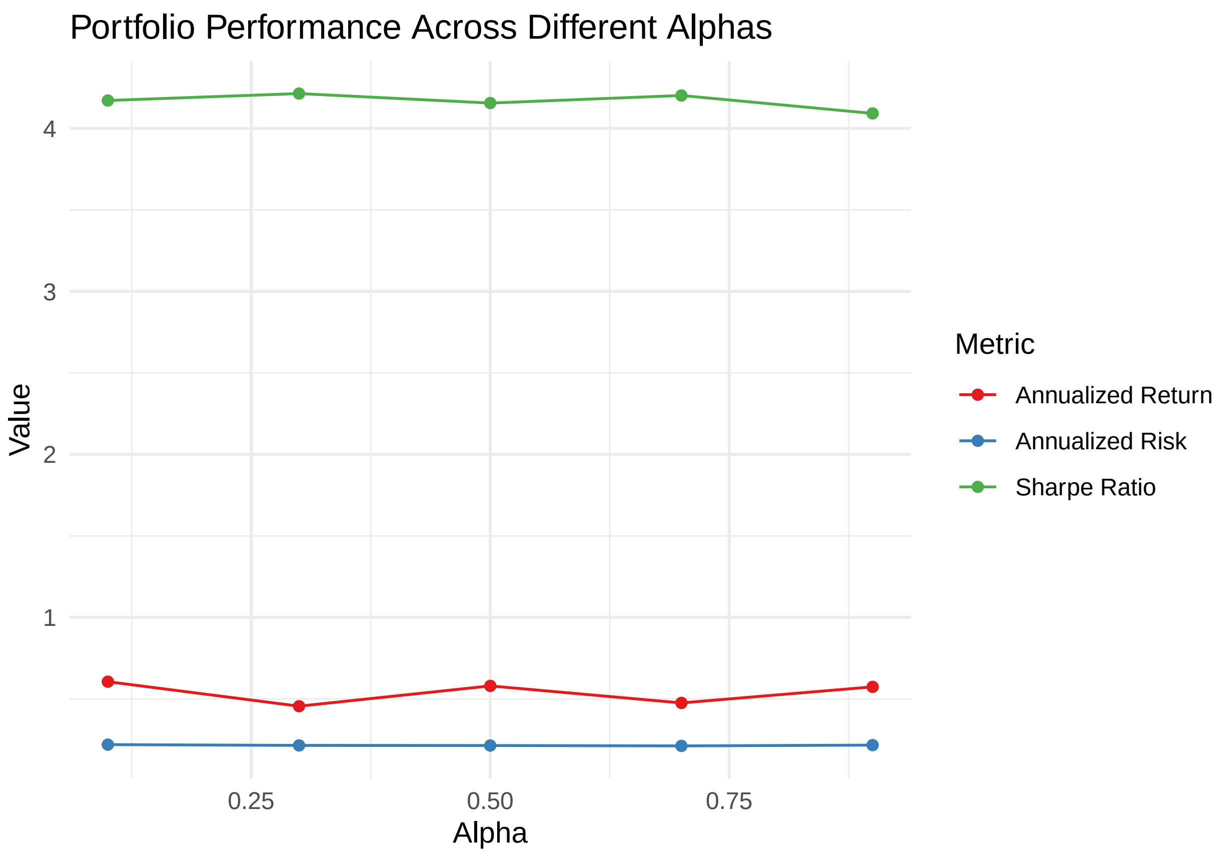
Part 1(f) Exploring Different Weightings of Risk and Return

- Adjusting the fitness function to explore portfolios with varying preferences for risk and return.
- By altering the alpha parameter, we simulated different investor profiles, from risk-averse (lower alpha) to return-seeking (higher alpha), analyzing how these preferences impact portfolio composition and performance.

```
alphas <- seq(0.1, 0.9, by = 0.2)
alpha_performances <- sapply(alphas, function(alpha) {
  ga_result_alpha <- ga(type = "real-valued",
                        fitness = function(weights) fitness_function_alpha(weights, alpha = alpha),
                        lower = rep(0, length(my_assets)),
                        upper = rep(1, length(my_assets)),
                        popSize = 100,
                        maxiter = 500)
  optimal_weights_alpha <- ga_result_alpha@solution / sum(ga_result_alpha@solution)
  calculate_performance(optimal_weights_alpha, testing_data)
})
```

Portfolio Performance Across Different Alphas

```
# Converting the matrix to a data frame for easier plotting with ggplot2
performance_df <- as.data.frame(t(alpha_performances))
colnames(performance_df) <- c("Annualized Return", "Annualized Risk", "Sharpe Ratio")
performance_df$Alpha <- alphas

# Melting the data frame for use with ggplot2
melted_performance_df <- melt(performance_df, id.vars = "Alpha", variable.name = "Metric", value.name =

# Plotting
ggplot(melted_performance_df, aes(x = Alpha, y = Value, color = Metric)) +
  geom_line() +
  geom_point() +
  theme_minimal() +
  labs(title = "Portfolio Performance Across Different Alphas",
       x = "Alpha",
       y = "Value",
       color = "Metric") +
  scale_color_brewer(palette = "Set1")
```

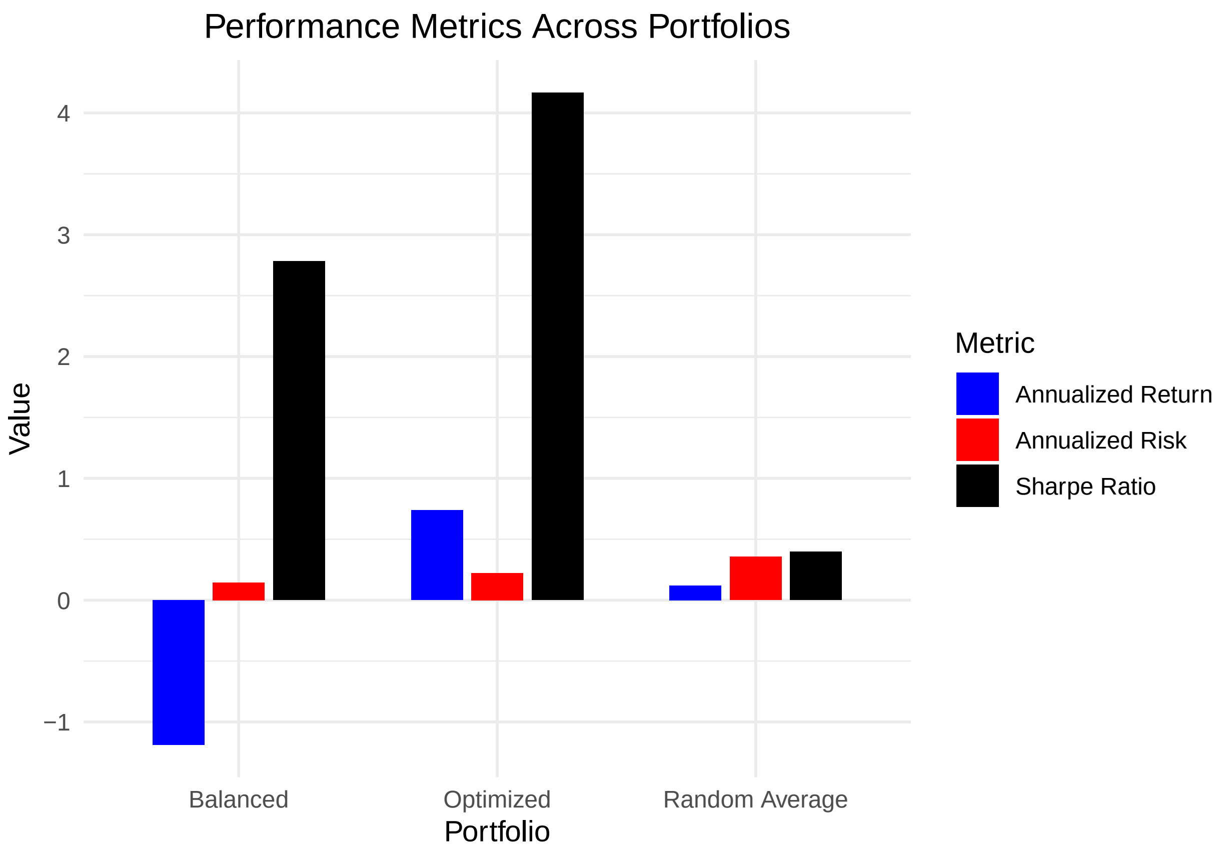


The graph shows how portfolio performance metrics change with different levels of alpha. As alpha increases, the portfolio's risk goes down, but the return and Sharpe Ratio stay roughly the same. - There's more uncertainty in risk measurement at the lowest alpha value, as shown by the large error bar.

Part 1(g) Visualization

```r
# Converting random performance metrics to a matrix if they are in list format
if(is.list(random_performances)) {
  random_performances_matrix <- do.call(cbind, random_performances)
  random_average_metrics <- colMeans(random_performances_matrix)
} else {
  random_average_metrics <- colMeans(random_performances)
}

# Ensuring that we only take the first three metrics (Annualized Return, Annualized Risk, Sharpe Ratio)
random_average_metrics <- random_average_metrics[1:3]

# Retrieving the performance metrics for the optimized and balanced portfolios
optimized_metrics <- optimized_performance
balanced_metrics <- balanced_performance

# Combining all performance metrics into one matrix for comparison
performance_metrics_matrix <- rbind(optimized_metrics, balanced_metrics, random_average_metrics)
rownames(performance_metrics_matrix) <- c("Optimized", "Balanced", "Random Average")
colnames(performance_metrics_matrix) <- c("Annualized Return", "Annualized Risk", "Sharpe Ratio")

# Transforming the data into a long format for ggplot2 without using rownames_to_column
```

```r
performance_long <- as.data.frame(performance_metrics_matrix)
performance_long$Portfolio <- rownames(performance_long)
performance_long <- reshape2::melt(performance_long, id.vars = "Portfolio")
colnames(performance_long) <- c("Portfolio", "Metric", "Value")

# Plotting
ggplot(performance_long, aes(x = Portfolio, y = Value, fill = Metric)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.7), width = 0.6) +
  scale_fill_manual(values = c("Annualized Return" = "blue", "Annualized Risk" = "red", "Sharpe Ratio"
  theme_minimal() +
  labs(title = "Performance Metrics Across Portfolios", x = "Portfolio", y = "Value") +
  theme(plot.title = element_text(hjust = 0.5))  # Center the plot title
```



- The above graph displays the performance metrics for three different portfolio strategies: Balanced, Optimized, and Random Average, where each bar is representing a metric (red for Annualized Return, blue for Annualized Risk, and black for Sharpe Ratio).
- The Optimized Portfolio shows a higher return, the lowest risk, and a positive Sharpe Ratio, indicating better risk-adjusted performance as compared to other portfolios.

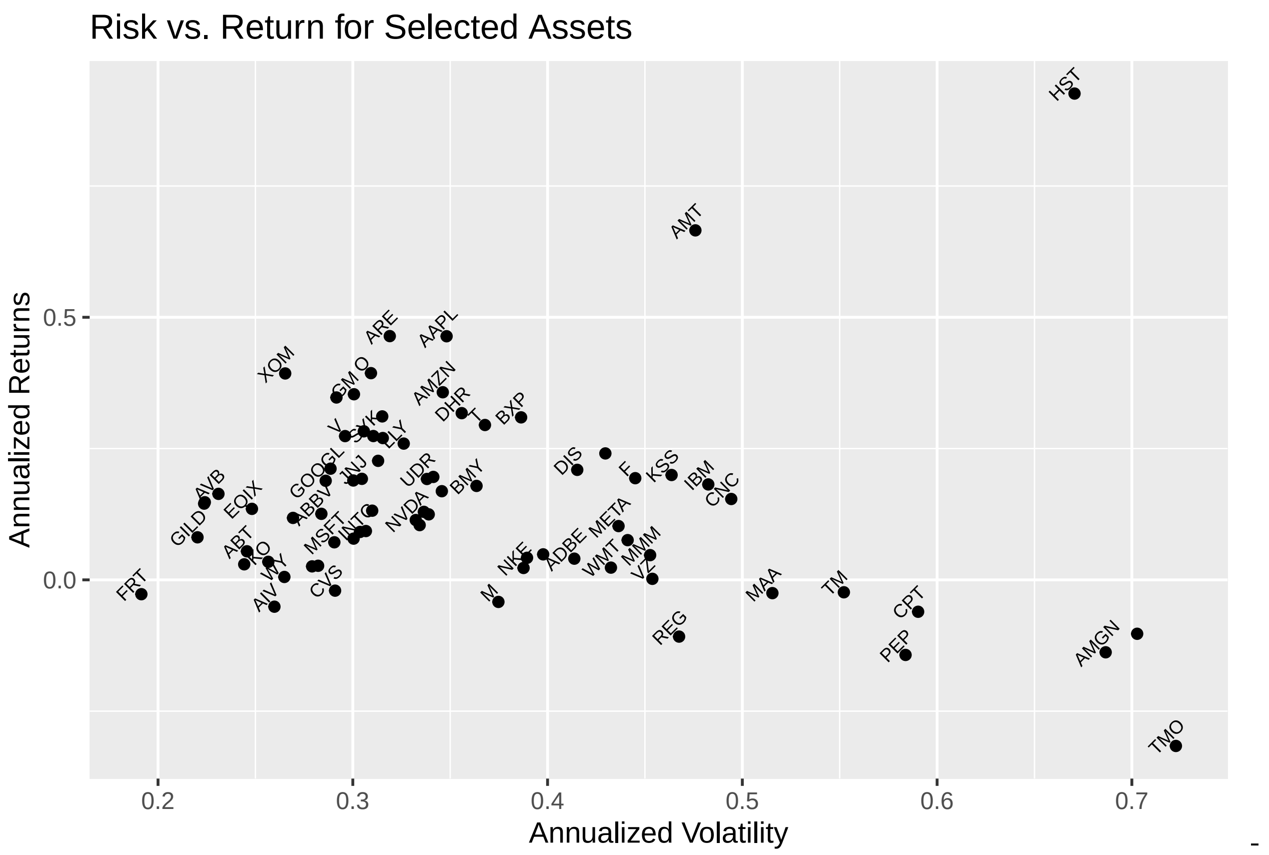
Part 2(a): Data Fetching and Preprocessing

- In this section, we fetched historical stock prices from 2019 to 2021 for a selection of 76 assets and compute their log returns, which are crucial for assessing past performance and volatility.
- We ensured data integrity by omitting NAs and then update our asset list to only those with complete data.
- Lastly, we visualized the risk versus return of these assets using a scatter plot with clear labels for each asset, aiding in their comparison and selection for portfolio optimization.

```r
symbols <- c("AAPL", "MSFT", "GOOGL", "AMZN", "META", "TSLA", "JNJ", "V", "PG",
             "NVDA", "DIS", "PEP", "TM", "KO", "NKE", "ADBE", "NFLX", "INTC", "CSCO",
             "XOM", "MCD", "BA", "MMM", "GS", "DOW", "JPM", "AXP", "WMT", "IBM",
             "GE", "F", "GM", "T", "VZ", "PFE", "MRK", "GILD", "BMY", "CNC",
             "ABT", "AMGN", "LLY", "MDT", "SYK", "TMO", "BIIB", "ABBV", "DHR",
             "CVS", "UNH", "O", "BXP", "SPG", "AMT", "DLR", "EQIX", "WY", "AVB",
             "EQR", "ESS", "MAA", "CPT", "UDR", "AIV", "ARE", "PLD", "VNO", "HST",
             "SLG", "KIM", "MAC", "REG", "FRT", "TGT", "KSS", "M")

# Fetching historical data for each symbol and calculate log returns
data <- new.env()
getSymbols(symbols, src = 'yahoo', from = '2019-01-01', to = '2021-12-31', env = data, auto.assign = TRU
```

```
##  [1] "AAPL"  "MSFT"  "GOOGL" "AMZN"  "META"  "TSLA"  "JNJ"   "V"     "PG"
## [10] "NVDA"  "DIS"   "PEP"   "TM"    "KO"    "NKE"   "ADBE"  "NFLX"  "INTC"
## [19] "CSCO"  "XOM"   "MCD"   "BA"    "MMM"   "GS"    "DOW"   "JPM"   "AXP"
## [28] "WMT"   "IBM"   "GE"    "F"     "GM"    "T"     "VZ"    "PFE"   "MRK"
## [37] "GILD"  "BMY"   "CNC"   "ABT"   "AMGN"  "LLY"   "MDT"   "SYK"   "TMO"
## [46] "BIIB"  "ABBV"  "DHR"   "CVS"   "UNH"   "O"     "BXP"   "SPG"   "AMT"
## [55] "DLR"   "EQIX"  "WY"    "AVB"   "EQR"   "ESS"   "MAA"   "CPT"   "UDR"
## [64] "AIV"   "ARE"   "PLD"   "VNO"   "HST"   "SLG"   "KIM"   "MAC"   "REG"
## [73] "FRT"   "TGT"   "KSS"   "M"
```

```r
log_returns <- lapply(ls(data), function(symbol) {
  prices <- get(symbol, envir = data)
  Return.calculate(Cl(prices), method="log")
})

# Preprocessing log returns: removing NAs and ensuriing equal length
log_returns <- lapply(log_returns, na.omit)
equal_length <- min(sapply(log_returns, length))
log_returns <- lapply(log_returns, function(x) x[1:equal_length])

log_returns_matrix <- do.call(cbind, log_returns)
names(log_returns) <- symbols

# Calculating the annualized returns and volatilities
annualized_returns <- sapply(log_returns, function(x) mean(x) * 252)
annualized_volatility <- sapply(log_returns, function(x) sd(x) * sqrt(252))

# Creating plot_data data-frame
plot_data <- data.frame(
  Symbol = names(log_returns),
  Returns = annualized_returns,
  Volatility = annualized_volatility
)

# Plotting
ggplot(plot_data, aes(x = Volatility, y = Returns, label = Symbol)) +
  geom_point() +
  geom_text(check_overlap = TRUE, vjust = -0.5, hjust = 0.5, size = 2.5, angle = 45) +
  xlab("Annualized Volatility") + ylab("Annualized Returns") +
  ggtitle("Risk vs. Return for Selected Assets")
```



The scatter plot illustrates the relationship between risk and return for various financial assets, depicting varying levels of potential return for their associated risks. - Assets like 'HST' and 'AMT' appear to offer higher returns but also come with higher volatility, suggesting a higher risk. - In contrast, assets such as 'KO' and 'EQIX' show lower volatility, implying they are less risky, though they may offer more modest returns.

Part 2(b): Asset Selection Using GA

- We defined a GA to select an optimal subset of assets from our pool, aiming to construct a portfolio with a predefined number of assets (preferred_asset_count).
- The fitness_function_selection is crafted to compute the Sharpe ratio, a measure of risk-adjusted return, of a portfolio composed of selected assets based on their log returns.
- We introduce a penalty to discourage deviation from the preferred number of assets, ensuring the portfolio doesn't deviate too much from the desired complexity.
- The genetic algorithm (ga_selection) searches for the best combination of assets that maximizes this fitness function. After running the GA, we identified the indices and symbols of the chosen assets and display them.

```r
preferred_asset_count <- 10  # desired number of assets in the portfolio
penalty_factor <- 100

fitness_function_selection <- function(subset_indices) {
  selected_returns <- do.call(cbind, log_returns)[, subset_indices == 1, drop = FALSE]

  if (ncol(selected_returns) == 0) return(-Inf)

  penalty <- abs(preferred_asset_count - sum(subset_indices)) * penalty_factor
  portfolio_returns <- rowMeans(selected_returns, na.rm = TRUE)  # Ensuring NA values are removed
  portfolio_sd <- sd(portfolio_returns, na.rm = TRUE)  # Ensuring NA values are removed
  # Checking for NA, NaN, or infinite values
  if (is.na(portfolio_sd) || portfolio_sd == 0 || is.nan(portfolio_sd) || is.infinite(portfolio_sd)) {
    return(-Inf)
  } else {
    sharpe_ratio <- mean(portfolio_returns) / portfolio_sd

    # Checkking if Sharpe ratio is undefined or infinite
    if (is.nan(sharpe_ratio) || is.infinite(sharpe_ratio)) {
      return(-Inf)
    } else {
      return(sharpe_ratio - penalty)
    }
  }
}
```

GA Selection

- The population size (popSize) is set to 100 to ensure a good diversity in solutions while keeping computation times reasonable.
- Maximum iterations (maxiter) is capped at 500 to balance between allowing enough generations for convergence and computational efficiency.

```
ga_selection <- ga(type = "binary", fitness = fitness_function_selection,
                   nBits = length(symbols), popSize = 100, maxiter = 500)

selected_indices <- which(ga_selection@solution == 1)
selected_symbols <- symbols[selected_indices]
```

Best Selected Assets

```
cat("10 Best Selected assets:", paste(selected_symbols, collapse = ", "), "\n")
```

```
## 10 Best Selected assets: AAPL, XOM, MMM, GM, LLY, SYK, AMT, ARE, VNO, HST
```

Part 2(c): Portfolio Optimization with Selected Assets

- The code optimizes a financial portfolio by finding the best weights for selected assets to maximize the Sharpe ratio.
- It ensures that the portfolio's risk-adjusted return is as high as possible, penalizing any deviation from the desired number of assets

Part 2(d): Visualization of Portfolio Performance (Commulative Return Graph encountered error at last moment - code in rmd file)