



UNIVERSIDADE DE SÃO PAULO  
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE DE RIBEIRÃO PRETO  
BACHARELADO EM FINANÇAS E NEGÓCIOS

PEDRO SARTORI DIAS DOS REIS

TRABALHO DO CURSO DE FINANÇAS I

RIBEIRÃO PRETO  
2024

## SUMÁRIO

1	ANÁLISE DE ATIVOS .....	2
2	PERÍODO DE ANÁLISE .....	2
3	ATIVO LIVRE DE RISCO .....	3
4	ESTATÍSTICAS DESCRITIVAS .....	4
5	APRESENTAÇÃO GRÁFICA .....	5
6	ANÁLISE DE CORRELAÇÃO .....	7
7	INCLUSÃO DO ATIVO LIVRE DE RISCO .....	11
8	FONTEIRA EFICIENTE .....	13
	REFERÊNCIAS .....	18
	ANEXO 1 – CÓDIGO R COMPLETO .....	19

## 1 ANÁLISE DE ATIVOS

Os ativos selecionados para esta análise foram Itaúsa, Petrobras, Porto Seguro, Taesa e Weg. A escolha desses ativos se baseou na diversidade setorial e na tendência de alta observada ao longo do período analisado.

## 2 PERÍODO DE ANÁLISE

O período de análise abrange de 8 de maio de 2020 a 8 de junho de 2024. Este período foi escolhido para capturar uma ampla gama de movimentos de mercado e diluir o impacto de eventos que possam ter impactado os ativos de forma atípica. A frequência dos dados utilizada foi diária, proporcionando uma maior quantidade de observações. No entanto, todas as taxas de retorno foram calculadas com base na taxa mensal acumulada utilizando-se a fórmula:

$$\frac{\text{Fechamento atual}}{\text{Fechamento de um mês atrás}} - 1$$

```
start_date = "2020-05-08"
end_date = "2024-06-08"

# Definir o período de análise
start_date <- as.Date(start_date)
end_date <- as.Date(end_date)

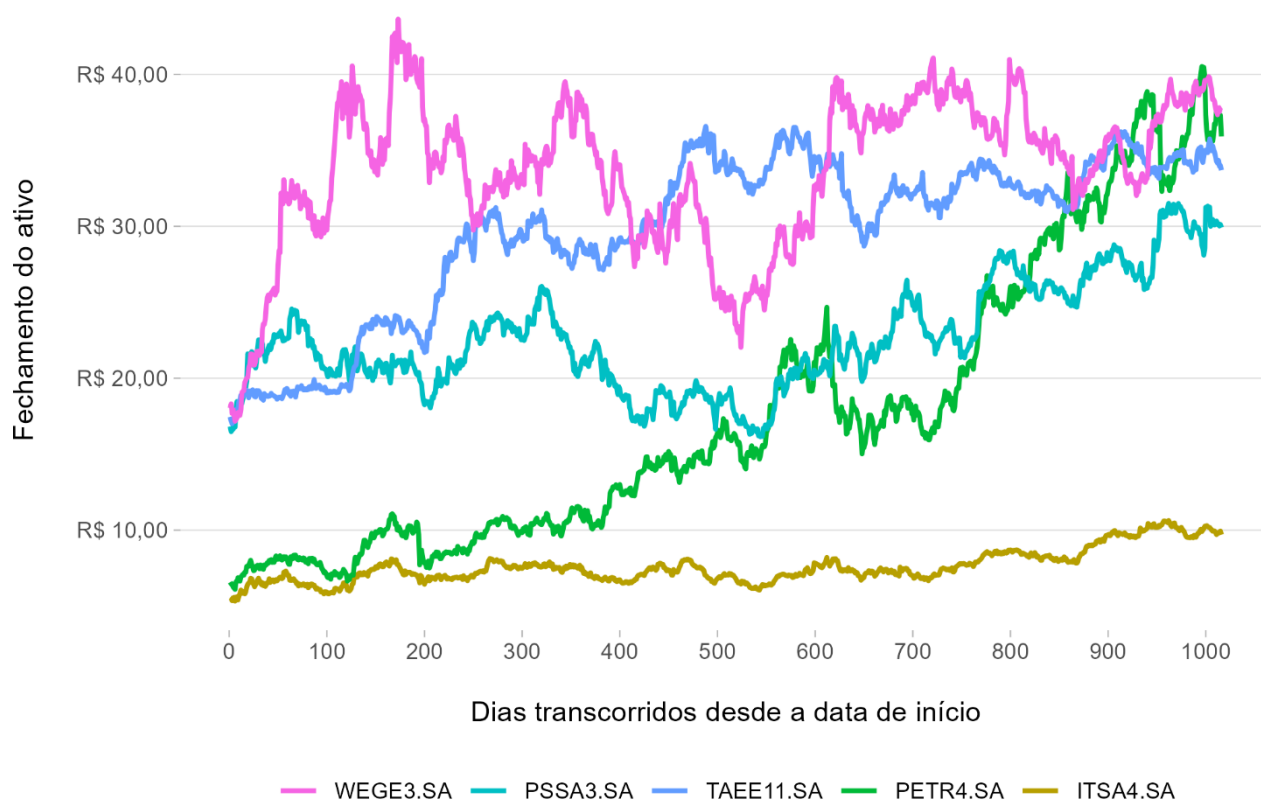
# Coletar dados históricos dos ativos
tickers <- c("^BVSP", "WEGE3.SA", "PSSA3.SA", "TAEE11.SA", "PETR4.SA", "ITSA4.SA")
getSymbols(tickers, src = "yahoo", from = start_date, to = end_date)
tickers <- c("BVSP", "WEGE3.SA", "PSSA3.SA", "TAEE11.SA", "PETR4.SA", "ITSA4.SA")

# Combinar os dados em um único dataframe
prices <- do.call(merge, lapply(tickers, function(ticker) Ad(get(ticker))))
colnames(prices) <- tickers
prices = as.data.frame(prices)

# Calcular os retornos mensais
returns <- na.omit(ROC(prices, type = "discrete", n=30))
returns1 = returns
returns1$index = 1:nrow(returns1)
returnsPivot = pivot_longer(returns1, 1:(ncol(returns1)-1), names_to = 'Ativo', values_to = 'retorno')
```

Na Figura 1 é possível observar as cotações dos ativos ao longo do período analisado, nota-se que todos os ativos possuem uma tendência geral de alta.

Figura 1 - Dados históricos dos ativos



### 3 ATIVO LIVRE DE RISCO

Para o ativo sem risco, utilizou-se a Selic, onde foram utilizados os dados históricos obtidos no site do Banco Central. Como a taxa fornecida é diária ela foi convertida para mensal utilizando-se a seguinte equação:

$$Taxa\ mensa\ = (1 + Taxa\ di\ria)^{22} - 1$$

Onde o n mero 22   o n mero m dio de dias  teis em um m s.

```

# Incorporando a Selic
selic = read.csv('selic.csv', sep = ';')
returnsRf = returns
returnsRf$data = rownames(returnsRf)
returnsRf = merge(returnsRf, selic, by='data')
returnsRf = returnsRf %>% rename(selic = valor)

# Convertendo para taxa mensal
selic = returnsRf$selic
selic = as.numeric(gsub(',', '.', selic))
selic = (selic/100 + 1)^22 - 1

```

#### 4 ESTATÍSTICAS DESCRITIVAS

Foram calculadas as estatísticas descritivas dos retornos dos ativos selecionados. A análise incluiu a média, a variância e o desvio padrão, que fornecem informações sobre o retorno médio e a volatilidade dos ativos. Adicionalmente, foram incluídas a mediana e o intervalo interquartil, que são medidas mais resistentes a outliers.

Inicialmente, as estatísticas foram calculadas sem considerar o ativo livre de risco. Além disso, as mesmas estatísticas descritivas foram calculadas para o Ibovespa, permitindo que os retornos dos ativos selecionados fossem comparados com o desempenho deste importante índice de mercado.

```

# Calculando as estatísticas descritivas
df_stats = returnsPivot %>%
  group_by(Ativo) %>%
  reframe(
    'Média' = sprintf("%.3f", mean(retorno)),
    'Desvio Padrão' = sprintf("%.3f", sd(retorno)),
    'Variância' = sprintf("%.3f", var(retorno)),
    'Mediana' = sprintf("%.3f", median(retorno)),
    'Intervalo interquartil' = sprintf("%.3f", IQR(retorno))
  )

```

Na Tabela 1, podemos observar as estatísticas descritivas de cada um dos ativos analisados. O ativo com maior retorno médio é a Petrobras, apresentando um retorno de 6,1% a.m. Em seguida,

temos a Weg com um retorno médio de 3,0% a.m. No entanto, tanto a Petrobras quanto a Weg exibem a maior variância entre os ativos, ambas com 1,8% a.m. Isso indica que, apesar do alto retorno médio, esses ativos também estão sujeitos a maiores flutuações nos preços.

Adicionalmente, a análise revela que a Petrobras possui um intervalo interquartil maior que o da Weg. Isso sugere que, enquanto a Weg apresenta *outliers* que podem estar influenciando sua variância.

Além disso, é notável que Porto Seguro, Itaúsa e Taesa, embora apresentem retornos médios relativamente baixos, também exibem baixa variância. Esse comportamento sugere que esses ativos são menos voláteis, proporcionando retornos mais estáveis ao longo do período analisado.

Também é possível observar que todos os ativos tiveram retorno médio superior ao Ibovespa, indicando um desempenho relativamente melhor no período considerado.

Tabela 1 – Estatísticas descritivas dos retornos sem o ativo livre de risco

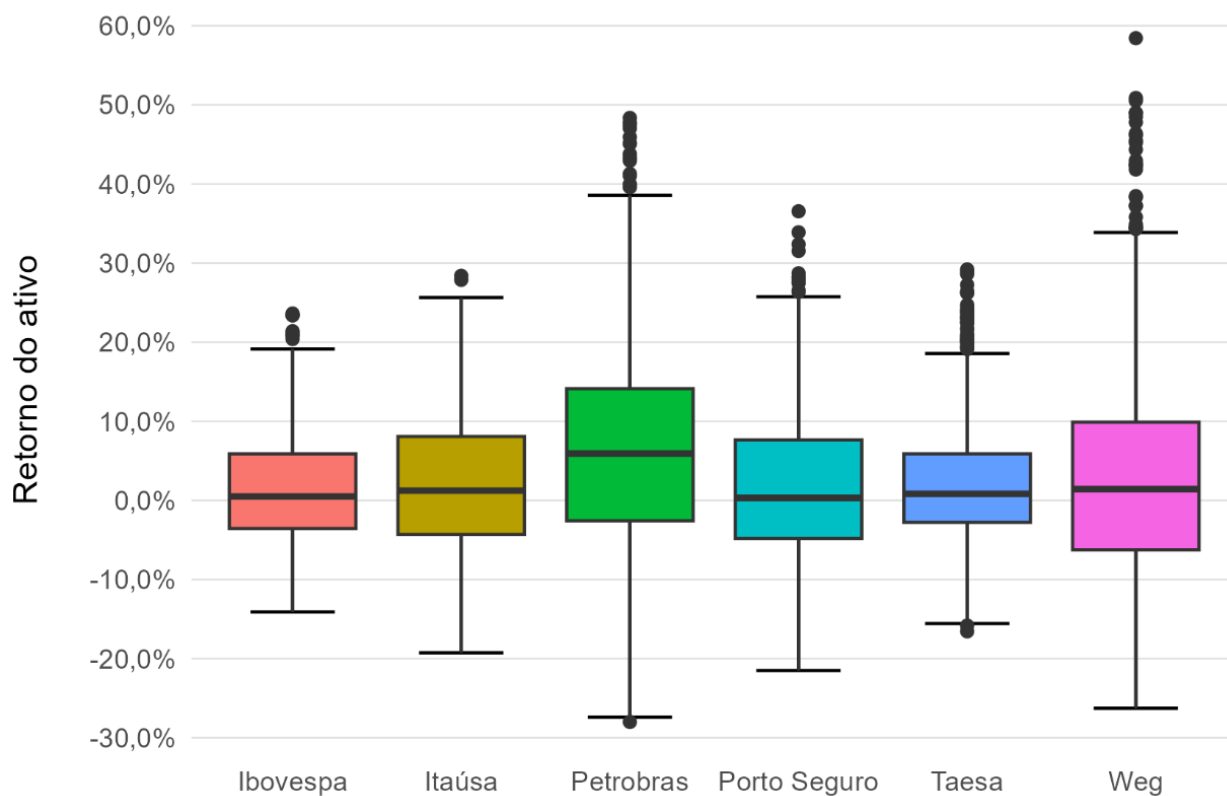
Ativo	Média	Desvio Padrão	Variância	Mediana	Intervalo interquartil
<b>BVSP</b>	0.013	0.069	0.005	0.005	0.094
<b>ITSA4.SA</b>	0.019	0.086	0.007	0.012	0.124
<b>PETR4.SA</b>	0.061	0.133	0.018	0.059	0.167
<b>PSSA3.SA</b>	0.018	0.099	0.010	0.003	0.124
<b>TAEE11.SA</b>	0.022	0.073	0.005	0.008	0.087
<b>WEGE3.SA</b>	0.030	0.133	0.018	0.014	0.161

## 5 APRESENTAÇÃO GRÁFICA

Neste capítulo, serão apresentadas e discutidas as visualizações das estatísticas descritivas dos ativos selecionados. Foram utilizados *boxplots* e histogramas para ilustrar a evolução dos preços, a variabilidade dos retornos e a distribuição das frequências ao longo do período analisado.

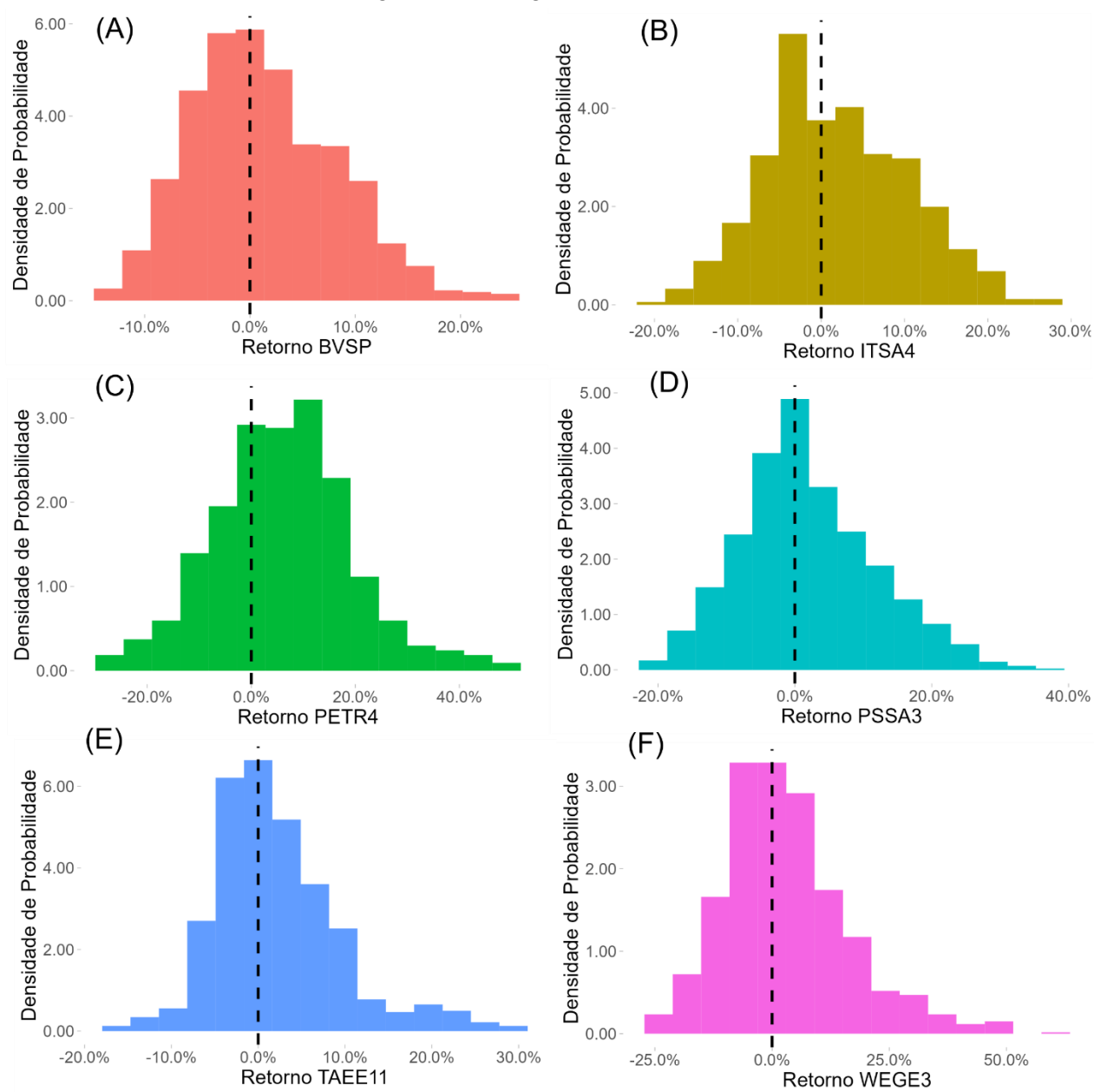
Na Figura 2, são apresentados os *boxplots* dos retornos dos ativos. Observa-se que Petrobras e Weg possuem as maiores medianas e também a maior dispersão entre os ativos analisados, enquanto Itaúsa, Taesa e Porto Seguro apresentam medianas e dispersões menores. A teoria apresentada no capítulo anterior é confirmada: a variância da Weg foi significativamente influenciada por *outliers*, como evidenciado pela quantidade e magnitude dos *outliers* presentes na distribuição dos retornos da Weg.

Figura 2 - Boxplots dos retornos dos ativos



Na Figura 3, os histogramas dos retornos mostram que todos os ativos seguem uma distribuição aproximadamente normal, mas com uma leve assimetria positiva, evidenciada por uma cauda levemente alongada à direita. Essa característica sugere que, embora os retornos geralmente se concentrem em torno da média, há uma tendência de ocorrerem retornos mais elevados com maior frequência do que o esperado em uma distribuição perfeitamente normal. Isso é confirmado pelos *boxplots* apresentados anteriormente, onde há muito mais *outliers* positivos influenciando os retornos dos ativos.

Figura 3 - Histogramas dos retornos



(A): Ibovespa; (B): Itaúsa; (C): Petrobrás; (D): Porto Seguro; (E): Taesa; (F): Weg

## 6 ANÁLISE DE CORRELAÇÃO

Para a análise de correlação entre os retornos dos ativos selecionados, foram calculados os coeficientes de correlação entre cada par de ativos e gerados gráficos para sua visualização. Os resultados dessas correlações serão apresentados e discutidos neste capítulo.



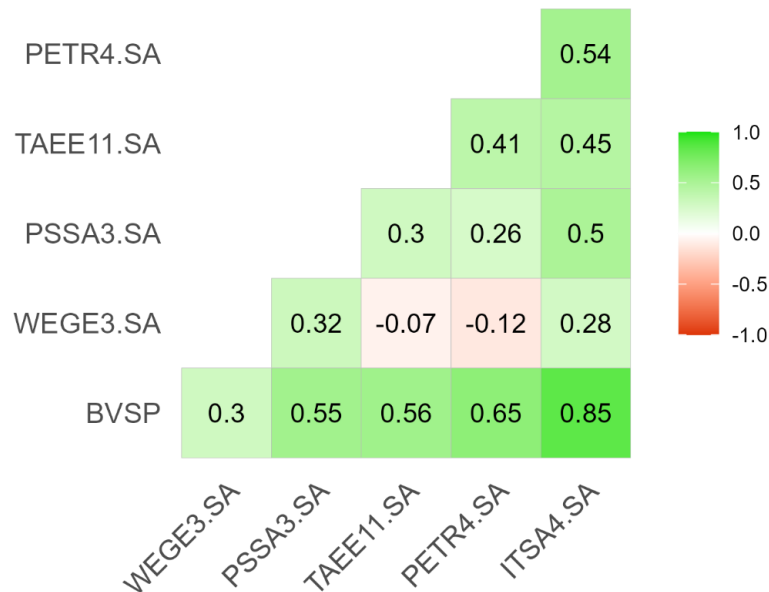
```
# Fazendo as correlações entre os retornos
ggcorrplot(
  cor(returns),
  type = 'lower',
  lab = TRUE,
  colors = c('#de3507', '#ffffff', '#15e30e')
) +
  theme(
    legend.title = element_blank(),
    panel.grid = element_blank()
  )
```

Ao observar a Figura 4, é possível notar que a maioria dos ativos apresenta correlações positivas inferiores a 0,5 entre si. Destaca-se que a Weg exibe correlações fracas e negativas com a Taesa e a Petrobrás, sugerindo que, embora a Weg e a Petrobrás tenham os maiores retornos médios, elas não variam em conjunto.

Essas baixas correlações entre os ativos podem ser atribuídas ao fato de pertencerem a setores diferentes da bolsa de valores. Além disso, a Itaúsa demonstra correlações mais fortes e positivas com a maioria dos ativos, possivelmente devido ao seu papel como holding, que detém participações em diversas empresas, resultando em uma maior correlação com diferentes setores da economia.

Também é observável que a maioria dos ativos possui correlações superiores a 0,5 com o Ibovespa, indicando uma relação mais estreita com o mercado como um todo. Nesse contexto, o Itaúsa se destaca por apresentar a maior correlação com o Ibovespa (0,85), refletindo sua influência significativa sobre o desempenho do mercado acionário brasileiro. Por outro lado, a Weg exibe a menor correlação com o Ibovespa, sugerindo uma menor dependência em relação às tendências gerais do mercado.

Figura 4 - Correlações entre os retornos dos ativos

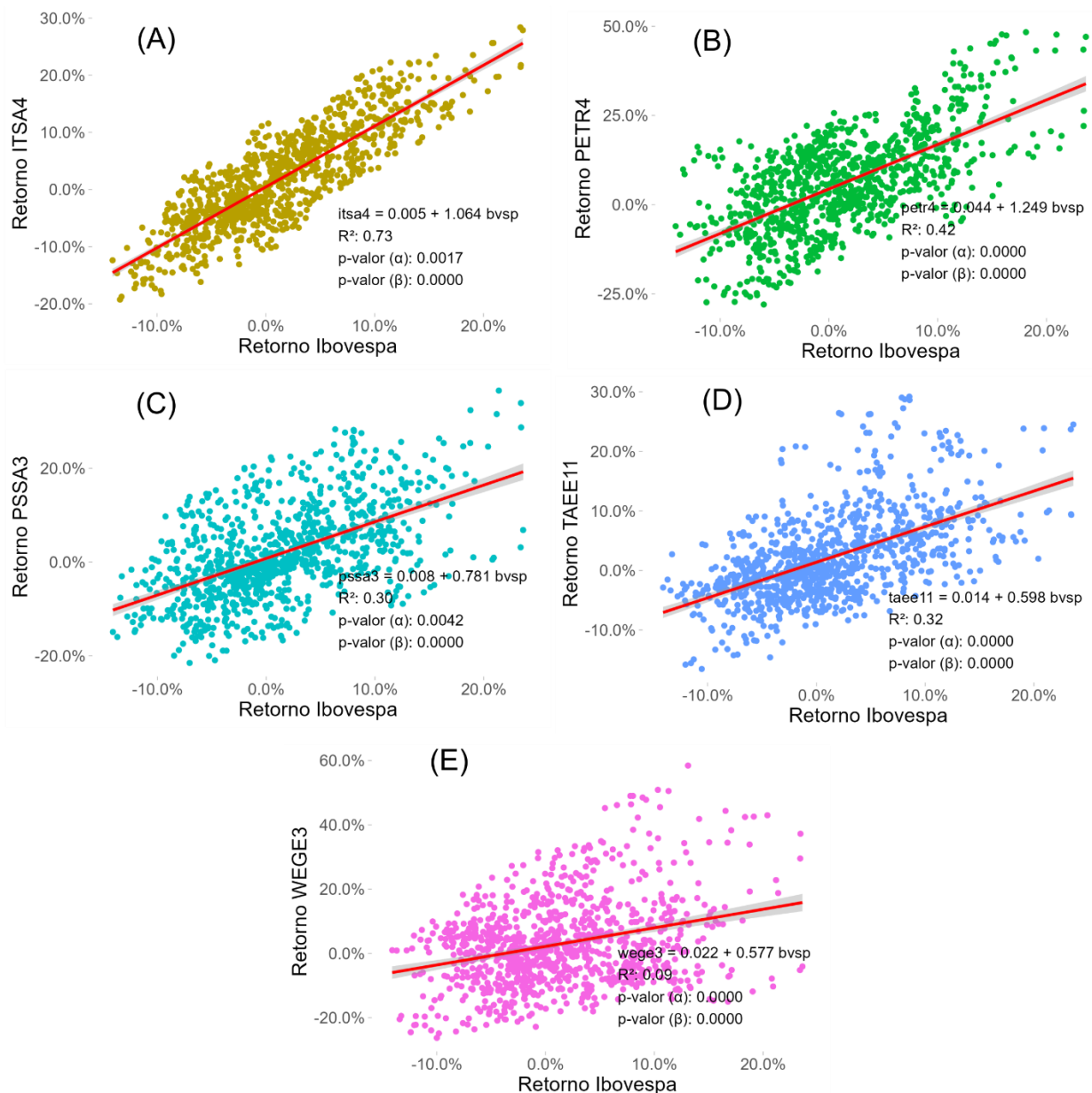


Para entender melhor a relação de cada ativo com o Ibovespa foram feitas regressões lineares entre o Ibovespa e os ativos, onde o Ibovespa foi usado como variável independente e os ativos foram usados como variável dependente:

```
# Regressões
colors = c('#f564e3', '#00bfc4', '#619cfe', '#00ba38', '#b79f00')
for(i in 2:length(tickers)) {
  retornoX = na.omit(returns[, 'BVSP'])
  posicaoX = max(retornoX) - 0.45 * (max(retornoX) - min(retornoX))
  model = lm(paste0(tickers[i], '~BVSP'), data=returns)
  sumModel = summary(model)
  legend = sprintf(
    "%s = %.3f + %.3f %s\nR²: %.2f\nnp-valor (α): %.4f\nnp-valor (β): %.4f",
    tolower(gsub('\\.SA', '', tickers[i])), model$coefficients[1],
    model$coefficients[2], 'bvsp', sumModel$r.squared, sumModel$coefficients[1,4],
    sumModel$coefficients[2,4])
  plot = ggplot() +
    theme_light() +
    geom_point(aes(y=returns[, tickers[i]], x=returns[, tickers[1]]), color=colors[i-1]) +
    theme(
      panel.grid = element_blank(),
      panel.border = element_blank(),
      text = element_text(size=15)
    ) +
    ylab(paste0('Retorno ', gsub('\\.SA', '', tickers[i]))) +
    xlab('Retorno Ibovespa') +
    geom_smooth(aes(y=returns[, tickers[i]], x=returns[, tickers[1]]), method='lm', color='red') +
    scale_y_continuous(label = function(x) sprintf("%.1f%%", x*100)) +
    scale_x_continuous(label = function(x) sprintf("%.1f%%", x*100)) +
    annotate("text", x=posicaoX, y=-0.10, label=legend, size=4, hjust=0)
  print(plot)
  ggsave(paste0('img/reg_', tolower(gsub('\\.SA', '', tickers[i])), '.png'))
}
```

Na Figura 5, observamos que todas as regressões apresentaram p-valores inferiores ao nível de significância de 0,001 tanto para o parâmetro  $\alpha$  quanto para o parâmetro  $\beta$ . Destaca-se que a regressão do Itaúsa exibe um  $R^2$  de 0,73, sugerindo que o Ibovespa é um índice confiável para prever os movimentos do Itaúsa, ressaltando a forte relação entre eles. Por outro lado, a Weg registrou um  $R^2$  de apenas 0,09, indicando que o Ibovespa não é um preditor eficaz para os movimentos da Weg. Além disso, observam-se indícios de heterocedasticidade na regressão da Weg com o Ibovespa, o que sugere que a variância dos erros não é constante ao longo do tempo.

Figura 5 - Regressões lineares entre os ativos e o Ibovespa



(A): Ibovespa; (B): Itaúsa; (C): Petrobrás; (D): Porto Seguro; (E): Taesa; (F): Weg

## 7 INCLUSÃO DO ATIVO LIVRE DE RISCO

Na análise, foi incluída a Selic como ativo livre de risco, utilizando seu histórico obtido no site do Banco Central do Brasil. Em seguida, os retornos mensais dos ativos foram ajustados subtraindo-se a taxa de retorno da Selic. Com base nesse ajuste, foram calculadas as novas estatísticas para verificar se houve alterações significativas nos retornos dos ativos.

```
# Incorporando a Selic
selic = read.csv('selic.csv', sep = ';')
returnsRf = returns
returnsRf$data = rownames(returnsRf)
returnsRf = merge(returnsRf, selic, by='data')
returnsRf = returnsRf %>% rename(selic = valor)

# Convertendo para taxa mensal
selic = returnsRf$selic
selic = as.numeric(gsub(',', '.', selic))
selic = (selic/100 + 1)^22 - 1

returnsRf$selic = selic

# Subtraindo a selic dos retornos
returnsRf = returns - returnsRf$selic

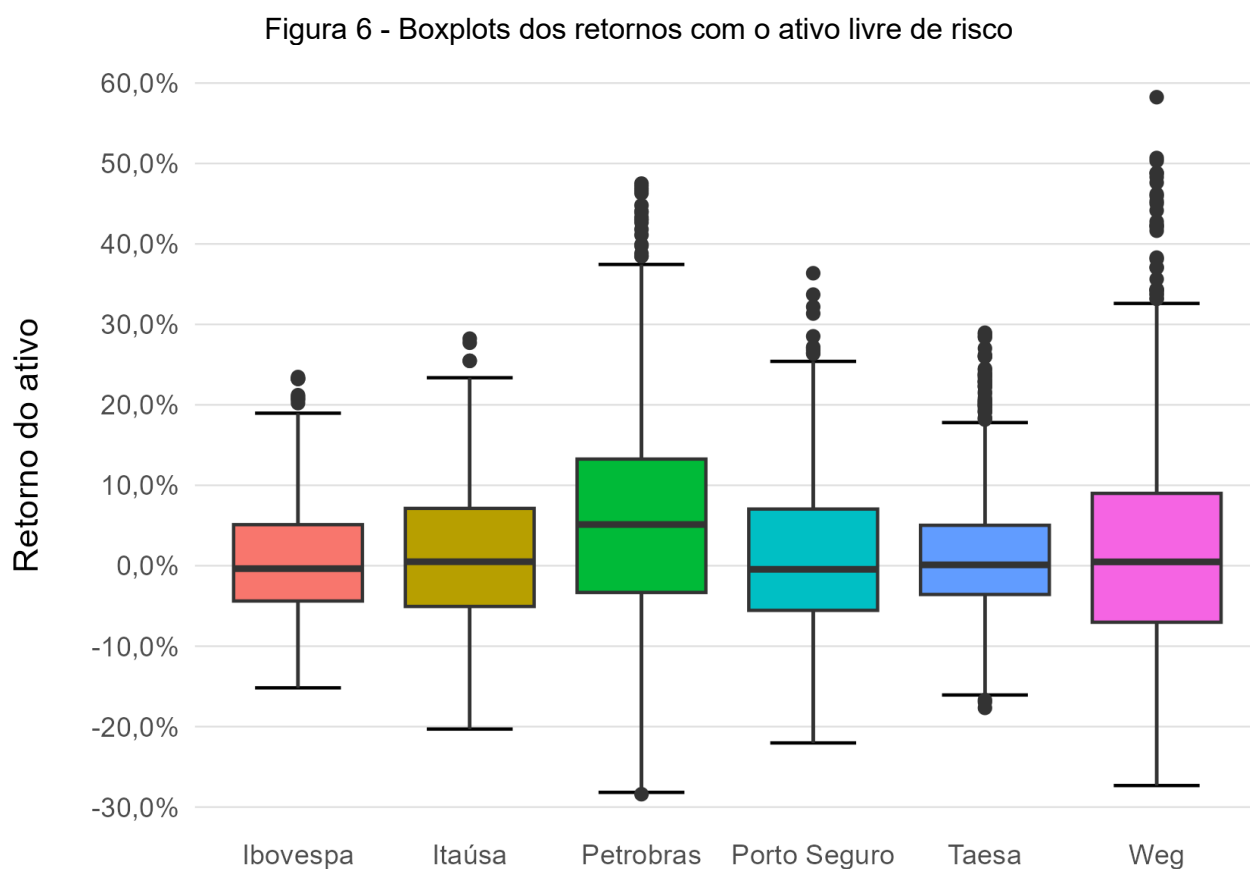
# Refazendo as estatísticas com os retornos subtraídos da selic
returns1 = returnsRf
returns1$index = 1:nrow(returns1)
returnsPivot = pivot_longer(
  returns1, 1:(ncol(returns1)-1),
  names_to = 'Ativo', values_to = 'retorno')
```

Ao observar a Tabela 2, percebe-se que as médias dos retornos dos ativos diminuíram e tanto o Ibovespa quanto a Porto Seguro tiveram medianas negativas. No entanto, a Petrobrás e a Weg ainda mantêm os maiores retornos médios e as maiores variâncias entre os ativos analisados.

Tabela 2 – Estatísticas descritivas dos retornos com o ativo livre de risco

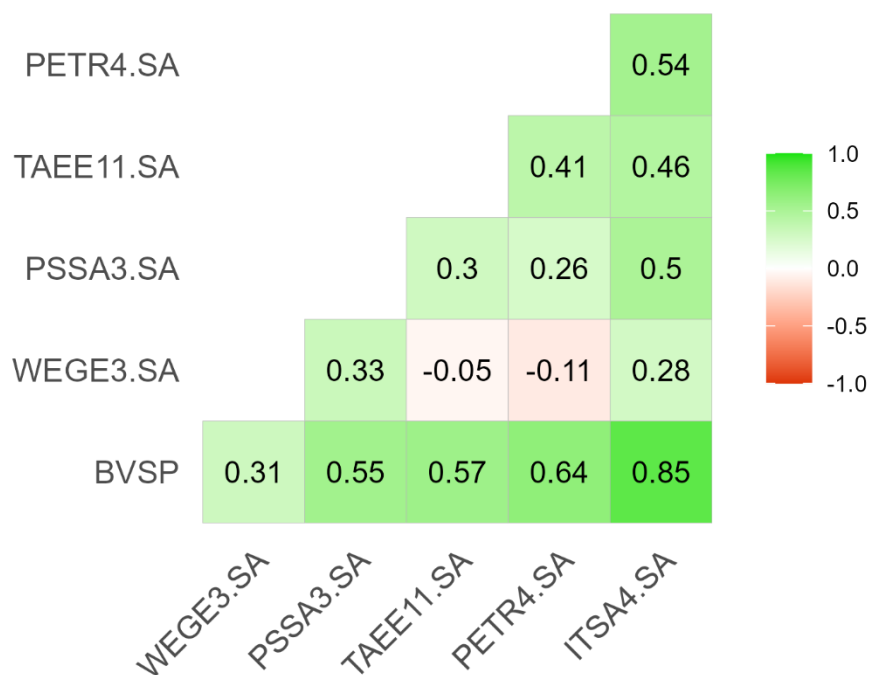
Ativo	Média	Desvio Padrão	Variância	Mediana	Intervalo interquartil
BVSP	0.006	0.070	0.005	-0.004	0.095
ITSA4.SA	0.011	0.086	0.007	0.005	0.122
PETR4.SA	0.053	0.133	0.018	0.051	0.166
PSSA3.SA	0.011	0.099	0.010	-0.004	0.126
TAEE11.SA	0.014	0.074	0.006	0.001	0.086
WEGE3.SA	0.022	0.134	0.018	0.005	0.160

De fato, ao observar a Figura 6, percebe-se que houve apenas um deslocamento vertical nas distribuições dos retornos, mantendo a Petrobrás e a Weg com as maiores medianas e intervalos interquartis entre os ativos analisados.



Ao observar a Figura 7, também é perceptível que não houve grandes alterações nas correlações, uma vez que os ativos mantiveram praticamente as mesmas correlações entre si.

Figura 7 - Correlações entre os retornos com o ativo livre de risco



## 8 FONTEIRA EFICIENTE

Foi realizada uma simulação com 20 mil carteiras, cada uma com pesos aleatórios para os ativos. Para cada carteira, calculou-se o retorno esperado, a variância e o índice de Sharpe. Dentre as carteiras analisadas, foram selecionadas duas ótimas: aquela com o maior índice de Sharpe e aquela com a menor variância. O código e os resultados dessa análise são apresentados abaixo.

```
# Removendo o bvsp
returns = returnsRf[,2:ncol(returnsRf)]

# Calculando a fronteira eficiente

# covariâncias e correlações
cov_returns = cov(returns)
cor_returns = cor(returns)

# números de ativos e carteiras
num_ativos = ncol(returns)
num_carteiras = 20000
```

```

# criando o dataframe dos dados
dados = data.frame()

# criando o dataframe dos pesos
dfPesos = data.frame()
for(i in 1:num_carteiras) {
  pesos = runif(num_ativos)
  pesos = pesos / sum(pesos)

  # retorno esperado
  retornoEsperado = as.numeric(pesos %*% as.matrix(colMeans(returns)))

  # volatilidade
  volatilidade = as.numeric(sqrt(t(pesos) %*% (cov_returns %*% pesos)))

  # Índice de Sharpe
  sharpe = retornoEsperado / volatilidade

  # Criando dataframe com os dados
  dados_i = data.frame(
    RetornoEsperado = retornoEsperado,
    Volatilidade = volatilidade,
    Sharpe = sharpe
  )

  dados = if(nrow(dados) == 0) dados_i else rbind(dados, dados_i)

  # Concatenando os pesos
  dfPesos_i = as.data.frame(matrix(pesos, ncol = length(pesos)))
  colnames(dfPesos_i) = colnames(returns)
  dfPesos = if(nrow(dfPesos) == 0) dfPesos_i else rbind(dfPesos, dfPesos_i)
}
dados = cbind(dados, dfPesos)

```

```

# Carteira de maior sharpe
carteiraOtima = dados %>% filter(Sharpe == max(Sharpe))

# Carteira de menor variância
carteiraMenorVariancia = dados %>% filter(Volatilidade == min(Volatilidade))

selicEsperada = mean(selic)

# linha da carteira livre de risco
dfLinha = data.frame(
  RetornoEsperado = c(selicEsperada, carteiraOtima$RetornoEsperado),
  Volatilidade = c(0, carteiraOtima$Volatilidade)
)

# Fazendo uma regressão para montar a linha
modellinha = lm('RetornoEsperado~Volatilidade', data=dfLinha)

# Montando a linha
linha = data.frame(Volatilidade=runif(100, -0.01, 0.12))
retornoLinha = predict(modellinha, newdata = linha)
linha$RetornoEsperado = retornoLinha

```

```

# Montando as carteiras dos ativos
carteirasAtivos = data.frame()
for(i in 1:ncol(returns)) {
  pesos = rep(0, ncol(returns))
  pesos[i] = 1

  # retorno esperado
  retornoEsperado = as.numeric(pesos %*% as.matrix(colMeans(returns)))

  # volatilidade
  volatilidade = as.numeric(sqrt(t(pesos) %*% (cov_returns %*% pesos)))

  # Índice de Sharpe
  sharpe = retornoEsperado / volatilidade

  # Criando dataframe com os dados
  dados_i = data.frame(
    Ativo = gsub('\\.SA', '', colnames(returns)[i]),
    RetornoEsperado = retornoEsperado,
    Volatilidade = volatilidade,
    Sharpe = sharpe
  )

  carteirasAtivos = if(nrow(carteirasAtivos) == 0) dados_i
  else rbind(carteirasAtivos, dados_i)
}

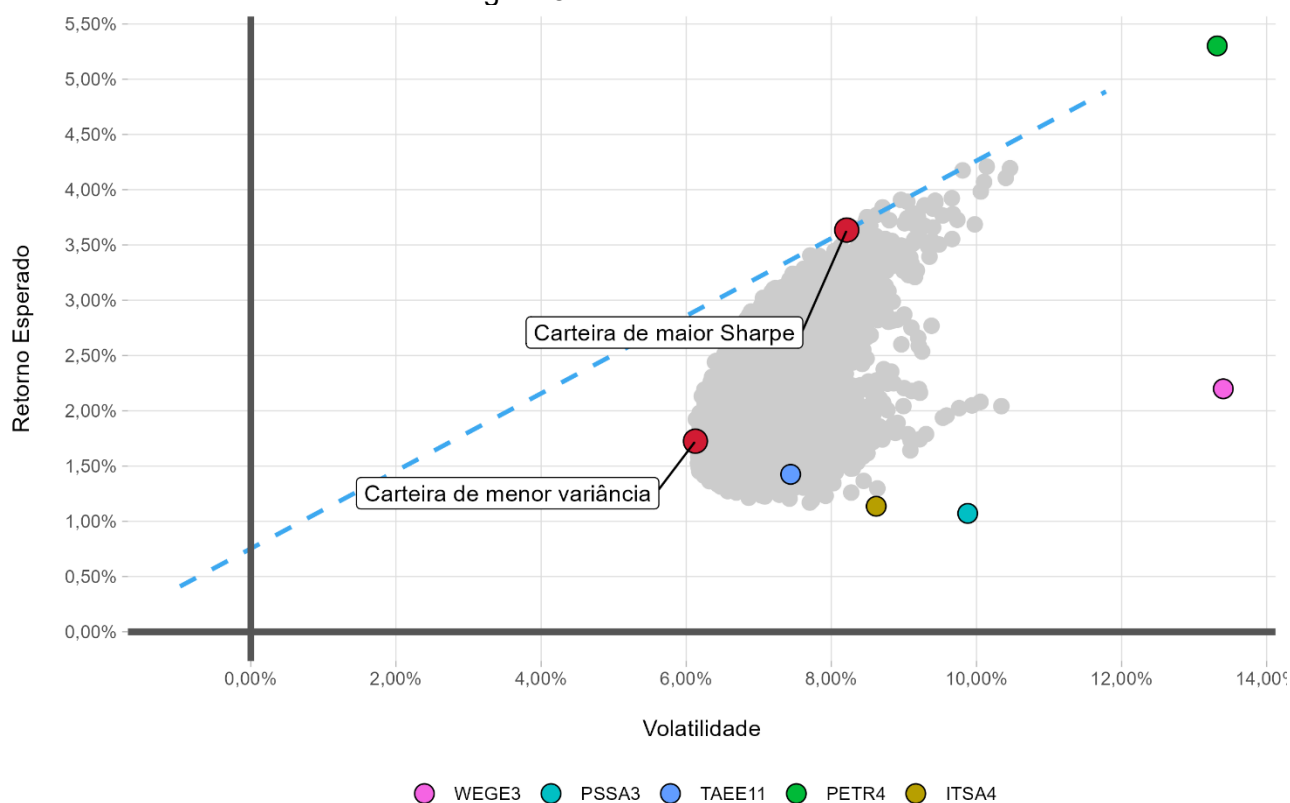
```

Na Figura 8, é possível visualizar a Fronteira Eficiente. Destaca-se a linha pontilhada azul, cujo intercepto representa a média da Selic no período analisado (0,75% a.m.), tangenciando a curva da Fronteira Eficiente e passando pelo ponto de maior índice de Sharpe. As carteiras de maior Sharpe e menor variância estão representadas em vermelho, enquanto as demais carteiras utilizadas na simulação aparecem em cinza.

Os pontos coloridos simbolizam as carteiras com 100% de alocação em um único ativo e 0% nos demais. Nota-se que os ativos Porto Seguro, Itaúsa e Taesa possuem retornos esperados abaixo do retorno esperado da carteira de menor variância. Isso sugere que alocar 100% da carteira nesses ativos é ineficiente, pois oferecem menos retorno que a carteira de menor variância e apresentam maior volatilidade.



Figura 8 - Fronteira eficiente



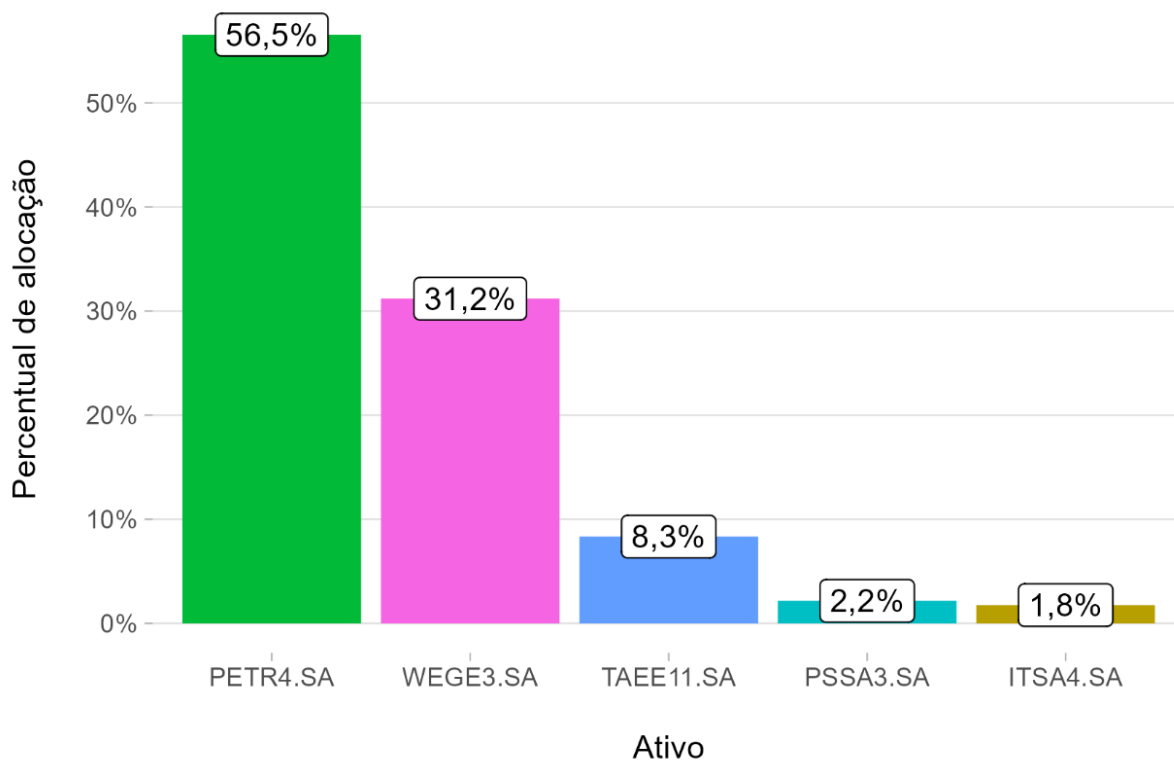
A tabela 3 apresenta os dados das duas carteiras analisadas em termos de retorno esperado, volatilidade e índice de Sharpe. Observamos que a carteira com maior índice de Sharpe oferece mais que o dobro do retorno em comparação com a carteira de menor variância. No entanto, essa maior rentabilidade vem acompanhada de uma volatilidade 1,39 vezes superior.

Tabela 3 – Características das carteiras

	Retorno Esperado	Volatilidade	Sharpe
<b>Carteira de maior Sharpe</b>	3,79%	8,53%	0,4447
<b>Carteira de menor variância</b>	1,65%	6,11%	0,2699

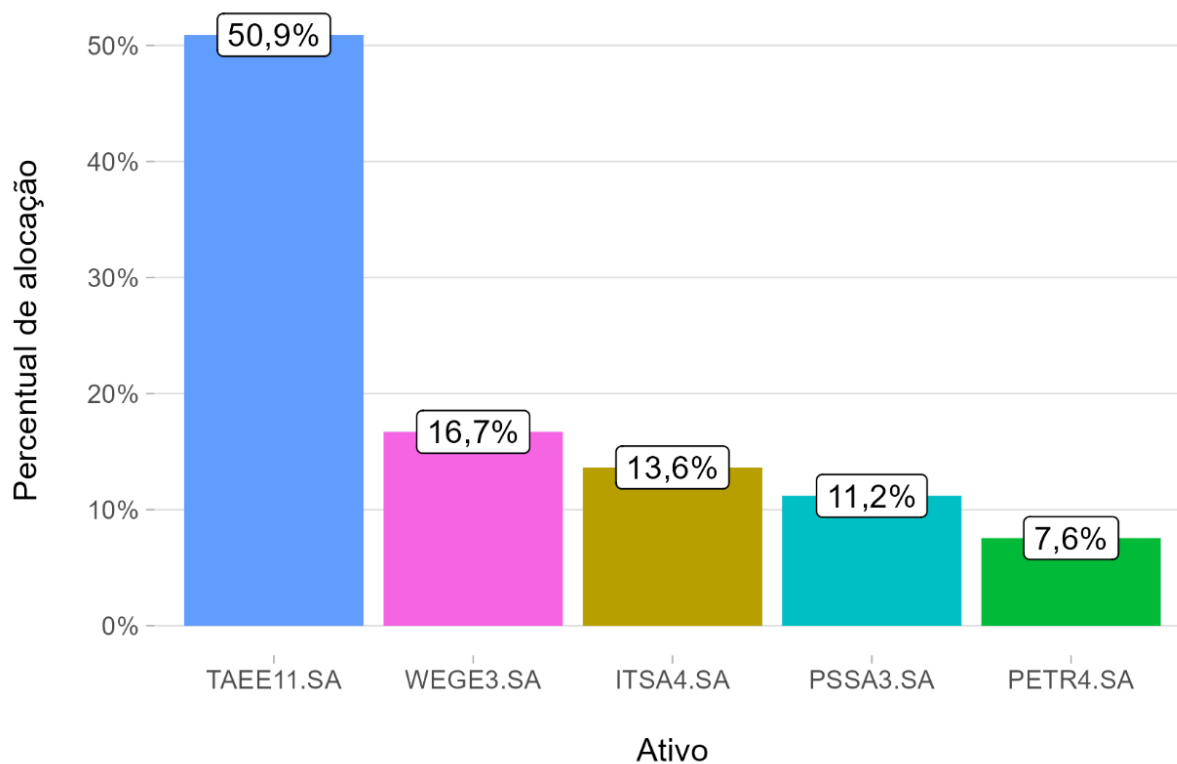
Na Figura 9, é possível observar o percentual de cada ativo na carteira de maior Sharpe. Nota-se uma alocação significativa em Petrobrás, com 56,5%, e Weg, com 31,2%. Essa distribuição provavelmente contribuiu para a volatilidade da carteira, já que estas são as empresas com maior variância.

Figura 9 - Alocação da carteira com maior Sharpe



Na Figura 10, observa-se a alocação de ativos na carteira de menor variância. Nota-se que 50,9% dos recursos estão alocados na Taesa, a empresa com a menor variância, o que contribuiu para a baixa volatilidade da carteira.

Figura 10 - Alocação da carteira com menor variância



## REFERÊNCIAS

BANCO CENTRAL DO BRASIL. Taxa de Juros – Selic. Disponível em:  
<<https://dadosabertos.bcb.gov.br/dataset/11-taxa-de-juros---selic>>. Acesso em: 13 de junho de 2024.

## ANEXO 1 – CÓDIGO R COMPLETO

```
# Link do GitHub: https://github.com/pedrosdr/Fronteira_Eficiente
setwd('Diretório de Análise')

library(quantmod)
library(tidyverse)
library(flextable)
library(ggcorrplot)
library(ggrepel)

start_date = "2020-05-08"
end_date = "2024-06-08"

# Definir o período de análise
start_date <- as.Date(start_date)
end_date <- as.Date(end_date)

# Coletar dados históricos dos ativos
tickers <- c("^BVSP", "WEGE3.SA", "PSSA3.SA", "TAEE11.SA", "PETR4.SA", "ITSA4.SA")
getSymbols(tickers, src = "yahoo", from = start_date, to = end_date)
tickers <- c("BVSP", "WEGE3.SA", "PSSA3.SA", "TAEE11.SA", "PETR4.SA", "ITSA4.SA")

# Combinar os dados em um único dataframe
prices <- do.call(merge, lapply(tickers, function(ticker) Ad(get(ticker))))
colnames(prices) <- tickers
prices = as.data.frame(prices)

# Printar dados Históricos
pricesHistoricos = prices
pricesHistoricos$index = 1:nrow(pricesHistoricos)
dadosHistoricosJuntos = pivot_longer(pricesHistoricos, 1:(ncol(pricesHistoricos)-1), names_to =
'Ativos', values_to = 'Fechamento')
ggplot() +
  theme_light() +
  geom_line(data=dadosHistoricosJuntos %>% filter(Ativos != 'BVSP'), aes(x=index, y=Fechamento,
color=Ativos), linewidth=0.9) +
```

```

theme(
  panel.grid.minor = element_blank(),
  panel.grid.major.x = element_blank(),
  panel.border = element_blank(),
  axis.title.y = element_text(margin = margin(r=15)),
  axis.title.x = element_text(margin = margin(t=15)),
  legend.title = element_blank(),
  legend.position = 'bottom'
) +
scale_y_continuous(labels = function(x) gsub("\\.", ',', sprintf("R$ %.2f", x))) +
scale_x_continuous(n.breaks = 10) +
ylab('Fechamento do ativo') +
xlab('Dias transcorridos desde a data de início') +
scale_color_manual(
  breaks=c("WEGE3.SA", "PSSA3.SA", "TAE11.SA", "PETR4.SA", "ITSA4.SA"),
  values=c(c('#f564e3', '#00bfc4', '#619cff', '#00ba38', '#b79f00'))
)

ggsave('img/historico.png')

# Calcular os retornos mensais
returns <- na.omit(ROC(prices, type = "discrete", n=30))
returns1 = returns
returns1$index = 1:nrow(returns1)
returnsPivot = pivot_longer(returns1, 1:(ncol(returns1)-1), names_to = 'Ativo', values_to = 'retorno')

# Printando os retornos mensais
ggplot() +
  theme_light() +
  geom_line(data=returnsPivot, aes(x=index, y=retorno, color=Ativo), linewidth=0.9) +
  theme(
    panel.grid.minor = element_blank(),
    panel.grid.major.x = element_blank(),
    panel.border = element_blank(),
    axis.title.y = element_text(margin = margin(r=15)),
    axis.title.x = element_text(margin = margin(t=15)),
    legend.title = element_blank(),
    legend.position = 'right'
  )

```

```

) +
scale_y_continuous(labels = function(x) gsub('\\.', ',', sprintf("%.2f%%", 100*x))) +
scale_x_continuous(n.breaks = 10) +
ylab('Retorno mensal do ativo') +
xlab('Dias transcorridos desde a data de início') +
scale_color_discrete(
  name='Empresa',
  labels=c("Ibovespa", "Itaúsa", "Petrobras", "Porto Seguro", "Taesa", "Weg"))
ggsave('img/retornos.png')

# Calculando as estatísticas descritivas
df_stats = returnsPivot %>%
  group_by(Ativo) %>%
  reframe(
    'Média' = sprintf("%.3f", mean(retorno)),
    'Desvio Padrão' = sprintf("%.3f", sd(retorno)),
    'Variância' = sprintf("%.3f", var(retorno)),
    'Mediana' = sprintf("%.3f", median(retorno)),
    'Intervalo interquartil' = sprintf("%.3f", IQR(retorno))
  )
flextable(df_stats)

# Fazendo o boxplot dos ativos
ggplot(returnsPivot) +
  theme_light() +
  stat_boxplot( aes(x=Ativo, y=retorno), geom='errorbar', linetype=1, width=0.5) +
  geom_boxplot(aes(x=Ativo, y=retorno, fill=Ativo)) +
  theme(
    panel.grid.minor = element_blank(),
    panel.grid.major.x = element_blank(),
    panel.border = element_blank(),
    axis.title.y = element_text(margin = margin(r=15)),
    axis.title.x = element_blank(),
    legend.title = element_blank(),
    legend.position = 'none',
    axis.ticks = element_blank()
  ) +
  scale_y_continuous(labels = function(x) gsub('\\.', ',', sprintf("%.1f%%", 100*x)), n.breaks = 10) +

```

```

scale_x_discrete(labels = c("Ibovespa", "Itaúsa", "Petrobras", "Porto Seguro", "Taesa", "Weg")) +
ylab('Retorno do ativo')
ggsave('img/boxplot_retornos.png')

```

```

# Fazendo as correlações entre os retornos

```

```

ggcorrplot(
  cor(returns),
  type = 'lower',
  lab = TRUE,
  colors = c('#de3507', '#ffffff', '#15e30e')
) +
theme(
  legend.title = element_blank(),
  panel.grid = element_blank()
)
ggsave('img/cor.png')

```

```

# Distribuições dos retornos

```

```

colors = c('#f8766d', '#f564e3', '#00bfc4', '#619cff', '#00ba38', '#b79f00')
for(i in 1:length(tickers)) {

```

```

  plot = ggplot() +
    theme_light() +
    geom_histogram(aes(x=returns[,tickers[i]], y=after_stat(density)), fill=colors[i], bins=15) +
    theme(
      panel.grid = element_blank(),
      panel.border = element_blank(),
      text = element_text(size=15)
    ) +
    geom_vline(xintercept = 0, linetype='dashed', linewidth=1) +
    xlab(paste0('Retorno ', gsub('\.SA', '', tickers[i]))) +
    ylab('Densidade de Probabilidade') +
    scale_y_continuous(label = function(x) sprintf("%.2f", x)) +
    scale_x_continuous(label = function(x) sprintf("%.1f%%", x*100))
    print(plot)
    ggsave(paste0('img/hist_', tolower(gsub('\.SA', '', tickers[i])), '.png'))
}

```

```

# Regressões
colors = c('#f564e3', '#00bfc4', '#619cff', '#00ba38', '#b79f00')
for(i in 2:length(tickers)) {
  retornoX = na.omit(returns[, 'BVSP'])
  posicaoX = max(retornoX) - 0.45 * (max(retornoX) - min(retornoX))

  model = lm(paste0(tickers[i], '~BVSP'), data=returns)
  sumModel = summary(model)
  legend = sprintf(
    "%s = %.3f + %.3f %s\nR²: %.2f\np-valor (α): %.4f\np-valor (β): %.4f",
    tolower(gsub("\\.SA", "", tickers[i])),
    model$coefficients[1],
    model$coefficients[2],
    'bvsp',
    sumModel$r.squared,
    sumModel$coefficients[1,4],
    sumModel$coefficients[2,4])

  plot = ggplot() +
    theme_light() +
    geom_point(aes(y=returns[, tickers[i]], x=returns[, tickers[1]]), color=colors[i-1]) +
    theme(
      panel.grid = element_blank(),
      panel.border = element_blank(),
      text = element_text(size=15)
    ) +
    ylab(paste0('Retorno ', gsub("\\.SA", "", tickers[i]))) +
    xlab('Retorno Ibovespa') +
    geom_smooth(aes(y=returns[, tickers[i]], x=returns[, tickers[1]]), method='lm', color='red') +
    scale_y_continuous(label = function(x) sprintf("%.1f%%", x*100)) +
    scale_x_continuous(label = function(x) sprintf("%.1f%%", x*100)) +
    annotate("text", x=posicaoX, y=-0.10, label=legend, size=4, hjust=0)
  print(plot)
  ggsave(paste0('img/reg_', tolower(gsub("\\.SA", "", tickers[i])), '.png'))
}

```



```

# model = lm('WEGE3.SA~BVSP',data=returns)
# plot(model)

# Hipóteses de Retornos
# Assumimos que os retornos esperados, variâncias e correlações dos ativos se mantêm estáveis
durante o período analisado.

# Incorporando a Selic
selic = read.csv('selic.csv', sep = ';')
returnsRf = returns
returnsRf$data = rownames(returnsRf)
returnsRf = merge(returnsRf, selic, by='data')
returnsRf = returnsRf %>% rename(selic = valor)

# Convertendo para taxa mensal
selic = returnsRf$selic
selic = as.numeric(gsub(',', '.', selic))
selic = (selic/100 + 1)^22 - 1

returnsRf$selic = selic

# Subtraindo a selic dos retornos
returnsRf = returns - returnsRf$selic

# Refazendo as estatísticas com os retornos subtraídos da selic
returns1 = returnsRf
returns1$index = 1:nrow(returns1)
returnsPivot = pivot_longer(
  returns1, 1:(ncol(returns1)-1), names_to = 'Ativo', values_to = 'retorno')

# Printando os retornos mensais
ggplot() +
  theme_light() +
  geom_line(data=returnsPivot, aes(x=index, y=retorno, color=Ativo), linewidth=0.9) +
  theme(

```

```

panel.grid.minor = element_blank(),
panel.grid.major.x = element_blank(),
panel.border = element_blank(),
axis.title.y = element_text(margin = margin(r=15)),
axis.title.x = element_text(margin = margin(t=15)),
legend.title = element_blank(),
legend.position = 'bottom'
) +
scale_y_continuous(labels = function(x) gsub('\\.', ',', sprintf("%.2f%%", 100*x))) +
scale_x_continuous(n.breaks = 10) +
ylab('Retorno mensal do ativo') +
xlab('Dias transcorridos desde a data de início') +
scale_color_discrete(
  name='Empresa',
  labels=c("Ibovespa", "Itaúsa", "Petrobras", "Porto Seguro", "Taesa", "Weg"))
ggsave('img/retornosRf.png')

```

```

df_stats = returnsPivot %>%
  group_by(Ativo) %>%
  reframe(
    'Média' = sprintf("%.3f", mean(retorno)),
    'Desvio Padrão' = sprintf("%.3f", sd(retorno)),
    'Variância' = sprintf("%.3f", var(retorno)),
    'Mediana' = sprintf("%.3f", median(retorno)),
    'Intervalo interquartil' = sprintf("%.3f", IQR(retorno))
  )
flextable(df_stats)

```

# Fazendo o boxplot dos ativos

```

ggplot(returnsPivot) +
  theme_light() +
  stat_boxplot(aes(x=Ativo, y=retorno), geom='errorbar', linetype=1, width=0.5) +
  geom_boxplot(aes(x=Ativo, y=retorno, fill=Ativo)) +
  theme(
    panel.grid.minor = element_blank(),
    panel.grid.major.x = element_blank(),
    panel.border = element_blank(),
    axis.title.y = element_text(margin = margin(r=15)),

```

```

axis.title.x = element_blank(),
legend.title = element_blank(),
legend.position = 'none',
axis.ticks = element_blank()
) +
scale_y_continuous(labels = function(x) gsub('\\.', ',', sprintf("%.1f%%", 100*x)), n.breaks = 10) +
scale_x_discrete(labels = c("Ibovespa", "Itaúsa", "Petrobras", "Porto Seguro", "Taesa", "Weg")) +
ylab('Retorno do ativo')
ggsave('img/boxplot_retornosRf.png')

```

# Fazendo as correlações entre os retornos

```

ggcorrplot(
  cor(returnsRf),
  type = 'lower',
  lab = TRUE,
  colors = c('#de3507', '#ffffff', '#15e30e')
) +
theme(
  legend.title = element_blank(),
  panel.grid = element_blank()
)
ggsave('img/corRf.png')

```

# Distribuições dos retornos

```

colors = c('#f8766d', '#f564e3', '#00bfc4', '#619cff', '#00ba38', '#b79f00')
for(i in 1:length(tickers)) {

plot = ggplot() +
  theme_light() +
  geom_histogram(aes(x=returnsRf[,tickers[i]], y=after_stat(density)), fill=colors[i], bins=15) +
  theme(
    panel.grid = element_blank(),
    panel.border = element_blank(),
    text = element_text(size=15)
  ) +
  geom_vline(xintercept = 0, linetype='dashed', linewidth=1) +
  xlab(paste0('Retorno ', gsub('\\.SA', '', tickers[i]))) +
  ylab('Densidade de Probabilidade') +

```

```

    scale_y_continuous(label = function(x) sprintf("%.2f", x)) +
    scale_x_continuous(label = function(x) sprintf("%.1f%%", x*100))
print(plot)
ggsave(paste0('img/histRf_', tolower(gsub("\\.SA", "", tickers[i])), '.png'))

}

# Regressões
colors = c('#f564e3', '#00bfc4', '#619cff', '#00ba38', '#b79f00')
for(i in 2:length(tickers)) {
  retornoX = na.omit(returnsRf[, 'BVSP'])
  posicaoX = max(retornoX) - 0.3 * (max(retornoX) - min(retornoX))

  model = lm(paste0(tickers[i], '~BVSP'), data=returnsRf)
  sumModel = summary(model)
  legend = sprintf(
    "%s = %.3f + %.3f %s\nR²: %.2f\np-valor (α): %.4f\np-valor (β): %.4f",
    tolower(gsub("\\.SA", "", tickers[i])),
    model$coefficients[1],
    model$coefficients[2],
    'bvsp',
    sumModel$r.squared,
    sumModel$coefficients[1,4],
    sumModel$coefficients[2,4])

  plot = ggplot() +
    theme_light() +
    geom_point(aes(y=returnsRf[, tickers[i]], x=returnsRf[, tickers[1]]), color=colors[i-1]) +
    theme(
      panel.grid = element_blank(),
      panel.border = element_blank(),
      text = element_text(size=15)
    ) +
    ylab(paste0('Retorno ', gsub("\\.SA", "", tickers[i]))) +
    xlab('Retorno Ibovespa') +
    geom_smooth(aes(y=returnsRf[, tickers[i]], x=returnsRf[, tickers[1]]), method='lm', color='red') +
    scale_y_continuous(label = function(x) sprintf("%.1f%%", x*100)) +
    scale_x_continuous(label = function(x) sprintf("%.1f%%", x*100)) +

```

```

    annotate("text", x=posicaoX, y=-0.10, label=legend, size=3, hjust=0)
print(plot)
ggsave(paste0('img/regRf_',tolower(gsub("\\.SA", "", tickers[i])),'.png'))
}

```

```
# Removendo o bvsp
```

```
returns = returnsRf[,2:ncol(returnsRf)]
```

```
# Calculando a fronteira eficiente
```

```
# covariâncias e correlações
```

```
cov_returns = cov(returns)
```

```
cor_returns = cor(returns)
```

```
# números de ativos e carteiras
```

```
num_ativos = ncol(returns)
```

```
num_carteiras = 10000
```

```
# criando o dataframe dos dados
```

```
dados = data.frame()
```

```
# criando o dataframe dos pesos
```

```
dfPesos = data.frame()
```

```
for(i in 1:num_carteiras) {
```

```
  pesos = runif(num_ativos)
```

```
  pesos = pesos / sum(pesos)
```

```
# retorno esperado
```

```
retornoEsperado = as.numeric(pesos %*% as.matrix(colMeans(returns)))
```

```
# volatilidade
```

```
volatilidade = as.numeric(sqrt(t(pesos) %*% (cov_returns %*% pesos)))
```

```
# Índice de Sharpe
```

```
sharpe = retornoEsperado / volatilidade
```

```
# Criando dataframe com os dados
```

```

dados_i = data.frame(
  RetornoEsperado = retornoEsperado,
  Volatilidade = volatilidade,
  Sharpe = sharpe
)

dados = if(nrow(dados) == 0) dados_i else rbind(dados, dados_i)

# Concatenando os pesos
dfPesos_i = as.data.frame(matrix(pesos, ncol = length(pesos)))
colnames(dfPesos_i) = colnames(returns)
dfPesos = if(nrow(dfPesos) == 0) dfPesos_i else rbind(dfPesos, dfPesos_i)
}
dados = cbind(dados, dfPesos)

# Carteira de maior sharpe
carteiraOtima = dados %>% filter(Sharpe == max(Sharpe))

# Carteira de menor variância
carteiraMenorVariancia = dados %>% filter(Volatilidade == min(Volatilidade))

# Montando as carteiras dos ativos
carteirasAtivos = data.frame()
for(i in 1:ncol(returns)) {
  pesos = rep(0, ncol(returns))
  pesos[i] = 1

  # retorno esperado
  retornoEsperado = as.numeric(pesos %*% as.matrix(colMeans(returns)))

  # volatilidade
  volatilidade = as.numeric(sqrt(t(pesos) %*% (cov_returns %*% pesos)))

  # Índice de Sharpe
  sharpe = retornoEsperado / volatilidade

  # Criando dataframe com os dados

```

```

dados_i = data.frame(
  Ativo = gsub("\\.SA'", "", colnames(returns)[i]),
  RetornoEsperado = retornoEsperado,
  Volatilidade = volatilidade,
  Sharpe = sharpe
)

carteirasAtivos = if(nrow(carteirasAtivos) == 0) dados_i else rbind(carteirasAtivos, dados_i)
}

# Calculando a selic Esperada
selicEsperada = mean(selic)

# linha da carteira livre de risco
dfLinha = data.frame(
  RetornoEsperado = c(selicEsperada, carteiraOtima$RetornoEsperado),
  Volatilidade = c(0, carteiraOtima$Volatilidade)
)

# Fazendo uma regressão para montar a linha
modelLinha = lm('RetornoEsperado~Volatilidade', data=dfLinha)

# Montando a linha
linha = data.frame(Volatilidade=runif(100, -0.01, 0.12))
retornoLinha = predict(modelLinha, newdata = linha)
linha$RetornoEsperado = retornoLinha

# Plotando a Fronteira eficiente
ggplot() +
  theme_light() +
  geom_point(data=dados, aes(x=Volatilidade, y=RetornoEsperado), color='#cccccc', size=3) +

  geom_line(data=linha, aes(x=Volatilidade, y=RetornoEsperado), linewidth=1, linetype='dashed',
color='#3ca8f0') +

  geom_point(data = carteiraOtima, aes(x=Volatilidade, y=RetornoEsperado), size=5, shape=21,
fill='#d11b33') +

```

```
geom_label_repel(data = carteiraOtima, aes(x=Volatilidade, y=RetornoEsperado), label='Carteira
de maior Sharpe', box.padding = 4) +
```

```
geom_point(data = carteiraMenorVariancia, aes(x=Volatilidade, y=RetornoEsperado), size=5,
shape=21, fill='#d11b33') +
```

```
geom_label_repel(data = carteiraMenorVariancia, aes(x=Volatilidade, y=RetornoEsperado),
label='Carteira de menor variância', box.padding = 5) +
```

```
geom_point(data = carteirasAtivos, aes(x=Volatilidade, y=RetornoEsperado, fill=Ativo), size=4,
shape=21) +
```

```
geom_vline(xintercept = 0, linewidth=1.5, color="#555555") +
```

```
geom_hline(yintercept = 0, linewidth=1.5, color="#555555") +
```

```
ylab('Retorno Esperado') +
```

```
xlab('Volatilidade') +
```

```
scale_x_continuous(n.breaks = 7, labels = function(x) gsub('\.', '', sprintf("%.2f%%", x*100))) +
```

```
scale_y_continuous(n.breaks = 15, labels = function(x) gsub('\.', '', sprintf("%.2f%%", x*100))) +
```

```
theme(
```

```
  panel.grid.minor = element_blank(),
```

```
  panel.border = element_blank(),
```

```
  axis.title.y = element_text(margin = margin(r=15)),
```

```
  axis.title.x = element_text(margin = margin(t=15)),
```

```
  legend.title = element_blank(),
```

```
  legend.position = 'bottom'
```

```
) +
```

```
scale_fill_manual(
```

```
  breaks=c("WEGE3", "PSSA3", "TAE11", "PETR4", "ITSA4" ),
```

```
  values=c(c('#f564e3', '#00bfc4', '#619cfe', '#00ba38', '#b79f00'))
```

```
)
```

```
ggsave('img/fronteira.png')
```

```
# Gráfico de donut da carteira ótima
```

```
donutOtima = pivot_longer(carreiraOtima, 4:ncol(carreiraOtima), values_to = 'Percentual', names_to
= 'Ativo')
```



```

ggplot(donutOtima) +
  theme_light() +

  scale_y_continuous(n.breaks = 7, labels = function(x) sprintf("%.0f%%", x*100)) +

  geom_col(aes(x=reorder(Ativo, -Percentual), y=Percentual, fill=Ativo)) +
  geom_label(aes(x=reorder(Ativo, -Percentual), y=Percentual, label=gsub("\\.", '', sprintf("%.1f%%",
Percentual*100)))) +

  ylab('Percentual de alocação') +
  xlab('Ativo') +

  theme(
    panel.grid.minor = element_blank(),
    panel.grid.major.x = element_blank(),
    panel.border = element_blank(),
    axis.title.y = element_text(margin = margin(r=15)),
    axis.title.x = element_text(margin = margin(t=15)),
    legend.position = 'none'
  ) +
  scale_fill_manual(
    breaks=c("WEGE3.SA", "PSSA3.SA", "TAE11.SA", "PETR4.SA", "ITSA4.SA" ),
    values=c('#f564e3', '#00bfc4', '#619cff', '#00ba38', '#b79f00'))
  )
ggsave('img/perc_otima.png')

# Gráfico de donut da carteira de menor Variância
donutMenorVariância = pivot_longer(carreiraMenorVariância, 4:ncol(carreiraMenorVariância),
values_to = 'Percentual', names_to = 'Ativo')

ggplot(donutMenorVariância) +
  theme_light() +

  scale_y_continuous(n.breaks = 7, labels = function(x) sprintf("%.0f%%", x*100)) +

  geom_col(aes(x=reorder(Ativo, -Percentual), y=Percentual, fill=Ativo)) +

```

```
geom_label(aes(x=reorder(Ativo, -Percentual), y=Percentual, label=gsub("\\.", ' ', sprintf("%.1f%%",
Percentual*100)))) +
```

```
ylab('Percentual de alocação') +
```

```
xlab('Ativo') +
```

```
theme(
```

```
  panel.grid.minor = element_blank(),
```

```
  panel.grid.major.x = element_blank(),
```

```
  panel.border = element_blank(),
```

```
  axis.title.y = element_text(margin = margin(r=15)),
```

```
  axis.title.x = element_text(margin = margin(t=15)),
```

```
  legend.position = 'none'
```

```
) +
```

```
scale_fill_manual(
```

```
  breaks=c("WEGE3.SA", "PSSA3.SA", "TAE11.SA", "PETR4.SA", "ITSA4.SA" ),
```

```
  values=c(c('#f564e3', '#00bfc4', '#619cfe', '#00ba38', '#b79f00'))
```

```
)
```

```
ggsave('img/perc_menor_variancia.png')
```

```
# Juntando as carteiras
```

```
carteirasOtimas = rbind(carteiraOtima[1:3], carteiraMenorVariancia[1:3])
```

```
flextable(carteirasOtimas)
```