

Análise estatística das variáveis utilizadas no Projeto Integrador do Grupo 1 da Digital House

Grupo 1 - Digital House

Introdução

Análise estatística e de correlação de algumas variáveis do Projeto Integrador do curso de Analista de Dados da Digital House. Variáveis estão agrupadas por estados e anos, começando em 2012 e terminando em 2021.

Variáveis usadas nesse trabalho:

- Vendas Diesel;
- Preço do barril do Petróleo;
- Vendas de gasolina pelas distribuidoras;
- Preço médio da Gasolina;
- Preço médio do Diesel;
- Vendas Etanol;
- Preço médio do Etanol;
- Salário-mínimo;
- Renda média;
- Inflação;
- Valor da cotação do dólar (dezembro).

Instalando Pacotes que serão usados nessa análise

```
pacotes <- c("tidyverse","stringr","readxl", "ggpubr", "Hmisc", "corrplot",  
            "PerformanceAnalytics", "gridExtra")  
if(sum(as.numeric(!pacotes %in% installed.packages())) != 0){  
  instalador <- pacotes[!pacotes %in% installed.packages()]  
  for(i in 1:length(instalador)) {  
    install.packages(instalador, dependencies = T)  
    break()  
  }  
  sapply(pacotes, require, character = T)  
} else {  
  sapply(pacotes, require, character = T)  
}
```

Carregando os datasets

```
# Organizei e limpei esse dataset no excel antes de importar para o R  
  
dados <- read_xlsx("tabela.xlsx", col_names = T)
```

```
# Dataset sobre a renda real por estado no Brasil do IBGE

renda <- read_xlsx("renda.xlsx", col_names = T)

# removendo a acentuação do nome dos estados do objeto "dados" para que o
# processo de junção ocorra de maneira adequada

dados$UF <- gsub("ã", "a", dados$UF)
dados$UF <- gsub("á", "a", dados$UF)
dados$UF <- gsub("í", "i", dados$UF)
dados$UF <- gsub("ô", "o", dados$UF)
```

Juntando a tabela de dados com a de renda para que eu possa começar a desenvolver a minha análise estatística sobre eles.

```
dados <- left_join(x = dados,
                  y = renda,
                  by = c("Ano" = "Ano",
                        "UF" = "Estado"))
```

Dando uma olhada nas primeiras 20 linhas e 5 primeiras colunas do dataset acima

Tive que mostrar apenas 5 colunas do dataset acima porque, como há muitas, elas não imprimiram corretamente no Rmarkdown

```
dados[1:20, 1:5] %>%
  knitr::kable(caption = "Variáveis usadas no Projeto Integrador do Grupo 1")
```

Table 1: Variáveis usadas no Projeto Integrador do Grupo 1

UF	Ano	Diesel_Vendas	Petroleo_barril_dolar	Gasolina_vendas_distr
Acre	2012	215.0000	0.00	119.0000
Acre	2013	157.3636	0.00	125.3300
Acre	2014	166.9902	0.00	138.1517
Acre	2015	160.4606	0.00	140.2588
Acre	2016	157.8521	0.00	136.8177
Acre	2017	151.1271	0.00	136.5926
Acre	2018	151.8963	0.00	133.8723
Acre	2019	155.4049	0.00	138.0014
Acre	2020	160.8718	0.00	136.4177
Acre	2021	157.8847	0.00	138.5681
Alagoas	2012	405.0000	111.40	364.0000
Alagoas	2013	401.5474	108.63	400.5180
Alagoas	2014	420.8324	105.35	441.6120
Alagoas	2015	403.3953	55.30	425.8727
Alagoas	2016	380.9068	40.08	453.9623
Alagoas	2017	356.0201	50.42	455.6607
Alagoas	2018	356.7090	69.79	417.7187
Alagoas	2019	352.0511	61.58	437.0712
Alagoas	2020	337.2551	42.34	423.8674
Alagoas	2021	367.0248	63.46	439.7723

Adicionando a região de cada estado ao dataset

```

dados$Regiao <- NA

dados$Regiao[dados$UF %in% c("Sao Paulo", "Minas Gerais", "Rio de Janeiro",
                             "Espírito Santo")] <- "Sudeste"

dados$Regiao[dados$UF %in% c("Rio Grande do Sul", "Santa Catarina",
                             "Parana")] <- "Sul"

dados$Regiao[dados$UF %in% c("Mato Grosso do Sul", "Mato Grosso", "Goias",
                             "Distrito Federal")] <- "Centro-Oeste"

dados$Regiao[dados$UF %in% c("Bahia", "Alagoas", "Sergipe", "Pernambuco",
                             "Paraiba", "Ceara", "Maranhao", "Piaui",
                             "Rio Grande do Norte")] <- "Nordeste"

dados$Regiao[dados$UF %in% c("Tocantins", "Rondonia", "Amazonas", "Acre",
                             "Para", "Roraima", "Amapa")] <- "Norte"

```

Parte 1 - Matriz de correlação com o p-value (nível de significância)

- Nível de significância: **0.05**
- H0 (Hipótese nula): Aumento do preço dos combustíveis não afeta a renda média dos brasileiros;
- H1 (Hipótese alternativa): Aumento do preço dos combustíveis afeta a renda média dos brasileiros.

```
correl <- rcorr(as.matrix(dados[,c(3:13)]))
```

Matriz - Coeficiente de correlação

```
round(correl$r,2)
```

```

##                Diesel_Vendas Petroleo_barril_dolar Gasolina_vendas_distr
## Diesel_Vendas                1.00                0.15                0.94
## Petroleo_barril_dolar        0.15                1.00                0.18
## Gasolina_vendas_distr        0.94                0.18                1.00
## Gasolina_preco_medio        -0.03               -0.15               -0.07
## Diesel_preco_medio          -0.08               -0.22               -0.12
## Etanol_vendas                0.87                0.17                0.84
## Etanol_preco_medio          -0.21               -0.15               -0.19
## Salario_minimo              0.00               -0.19               -0.03
## Inflacao                    0.02                0.02                0.00
## Dolar_valor_dez             0.01               -0.17               -0.03
## Renda                       0.30               -0.22                0.35
##
##                Gasolina_preco_medio Diesel_preco_medio Etanol_vendas
## Diesel_Vendas                -0.03               -0.08                0.87
## Petroleo_barril_dolar        -0.15               -0.22                0.17
## Gasolina_vendas_distr        -0.07               -0.12                0.84
## Gasolina_preco_medio         1.00                0.96                0.01
## Diesel_preco_medio           0.96                1.00                0.00
## Etanol_vendas                0.01                0.00                1.00
## Etanol_preco_medio           0.92                0.92               -0.17
## Salario_minimo              0.91                0.91                0.06

```

```
## Inflacao                0.14                0.13                0.00
## Dolar_valor_dez         0.85                0.85                0.06
## Renda                   0.36                0.34                0.32
##
##      Etanol_preco_medio Salario_minimo Inflacao
## Diesel_Vendas          -0.21             0.00      0.02
## Petroleo_barril_dolar  -0.15            -0.19      0.02
## Gasolina_vendas_distr  -0.19            -0.03      0.00
## Gasolina_preco_medio    0.92             0.91      0.14
## Diesel_preco_medio      0.92             0.91      0.13
## Etanol_vendas          -0.17             0.06      0.00
## Etanol_preco_medio      1.00             0.85      0.14
## Salario_minimo          0.85             1.00     -0.09
## Inflacao                0.14            -0.09      1.00
## Dolar_valor_dez         0.78             0.91      0.20
## Renda                   0.27             0.41     -0.05
##
##      Dolar_valor_dez Renda
## Diesel_Vendas        0.01  0.30
## Petroleo_barril_dolar -0.17 -0.22
## Gasolina_vendas_distr -0.03  0.35
## Gasolina_preco_medio  0.85  0.36
## Diesel_preco_medio    0.85  0.34
## Etanol_vendas         0.06  0.32
## Etanol_preco_medio    0.78  0.27
## Salario_minimo        0.91  0.41
## Inflacao              0.20 -0.05
## Dolar_valor_dez       1.00  0.37
## Renda                 0.37  1.00
```

Matriz - p-value

```
round(correl$P, 2)
```

```
##      Diesel_Vendas Petroleo_barril_dolar Gasolina_vendas_distr
## Diesel_Vendas      NA                0.02                0.00
## Petroleo_barril_dolar 0.02                NA                0.00
## Gasolina_vendas_distr 0.00                0.00                NA
## Gasolina_preco_medio  0.65                0.02                0.25
## Diesel_preco_medio    0.20                0.00                0.05
## Etanol_vendas         0.00                0.01                0.00
## Etanol_preco_medio    0.00                0.01                0.00
## Salario_minimo        0.97                0.00                0.66
## Inflacao              0.77                0.77                0.95
## Dolar_valor_dez       0.89                0.00                0.57
## Renda                 0.00                0.00                0.00
##
##      Gasolina_preco_medio Diesel_preco_medio Etanol_vendas
## Diesel_Vendas            0.65                0.20                0.00
## Petroleo_barril_dolar    0.02                0.00                0.01
## Gasolina_vendas_distr    0.25                0.05                0.00
## Gasolina_preco_medio      NA                0.00                0.89
## Diesel_preco_medio        0.00                NA                0.94
## Etanol_vendas            0.89                0.94                NA
## Etanol_preco_medio        0.00                0.00                0.01
## Salario_minimo           0.00                0.00                0.32
```

```
## Inflacao                0.02                0.03                0.94
## Dolar_valor_dez         0.00                0.00                0.32
## Renda                   0.00                0.00                0.00
##
## Etanol_preco_medio Salario_minimo Inflacao
## Diesel_Vendas        0.00                0.97                0.77
## Petroleo_barril_dolar 0.01                0.00                0.77
## Gasolina_vendas_distr 0.00                0.66                0.95
## Gasolina_preco_medio  0.00                0.00                0.02
## Diesel_preco_medio    0.00                0.00                0.03
## Etanol_vendas         0.01                0.32                0.94
## Etanol_preco_medio    NA                 0.00                0.02
## Salario_minimo        0.00                NA                 0.15
## Inflacao              0.02                0.15                NA
## Dolar_valor_dez       0.00                0.00                0.00
## Renda                 0.00                0.00                0.42
##
## Dolar_valor_dez Renda
## Diesel_Vendas   0.89  0.00
## Petroleo_barril_dolar 0.00  0.00
## Gasolina_vendas_distr 0.57  0.00
## Gasolina_preco_medio  0.00  0.00
## Diesel_preco_medio    0.00  0.00
## Etanol_vendas         0.32  0.00
## Etanol_preco_medio    0.00  0.00
## Salario_minimo        0.00  0.00
## Inflacao              0.00  0.42
## Dolar_valor_dez       NA    0.00
## Renda                 0.00  NA
```

Correlação e p-value das variáveis do teste de hipótese

- Preço médio do Diesel
- Preço médio da Gasolina
- Preço médio do Etanol
- Renda média

```
correl2 <- rcorr(as.matrix(dados[,c(6:7, 9, 13)]))
```

Matriz - Coeficiente de correlação

```
round(correl2$r,2) %>%
  knitr::kable(caption = "Matriz - Coeficiente de Correlação")
```

Table 2: Matriz - Coeficiente de Correlação

	Gasolina_preco_medio	Diesel_preco_medio	Etanol_preco_medio	Renda
Gasolina_preco_medio	1.00	0.96	0.92	0.36
Diesel_preco_medio	0.96	1.00	0.92	0.34
Etanol_preco_medio	0.92	0.92	1.00	0.27
Renda	0.36	0.34	0.27	1.00

Vemos que a correlação entre os preços dos combustíveis é bem alta, enquanto a correlação entre a renda e os preços dos combustíveis é bem baixa.

Matriz - p-value

```
correl2$P %>%  
  knitr::kable(caption = "Matriz - p-value")
```

Table 3: Matriz - p-value

	Gasolina_preco_medio	Diesel_preco_medio	Etanol_preco_medio	Renda
Gasolina_preco_medio	NA	0	0.0e+00	0.0e+00
Diesel_preco_medio	0	NA	0.0e+00	0.0e+00
Etanol_preco_medio	0	0	NA	5.5e-06
Renda	0	0	5.5e-06	NA

Podemos ver que o p-value foi bem menor que 0.001 para todas as variáveis, e isso significa que os resultados são altamente significantes. Ou seja, rejeitamos a hipótese nula, e isso nos diz que o aumento dos preços dos combustíveis afeta a renda média dos brasileiros.

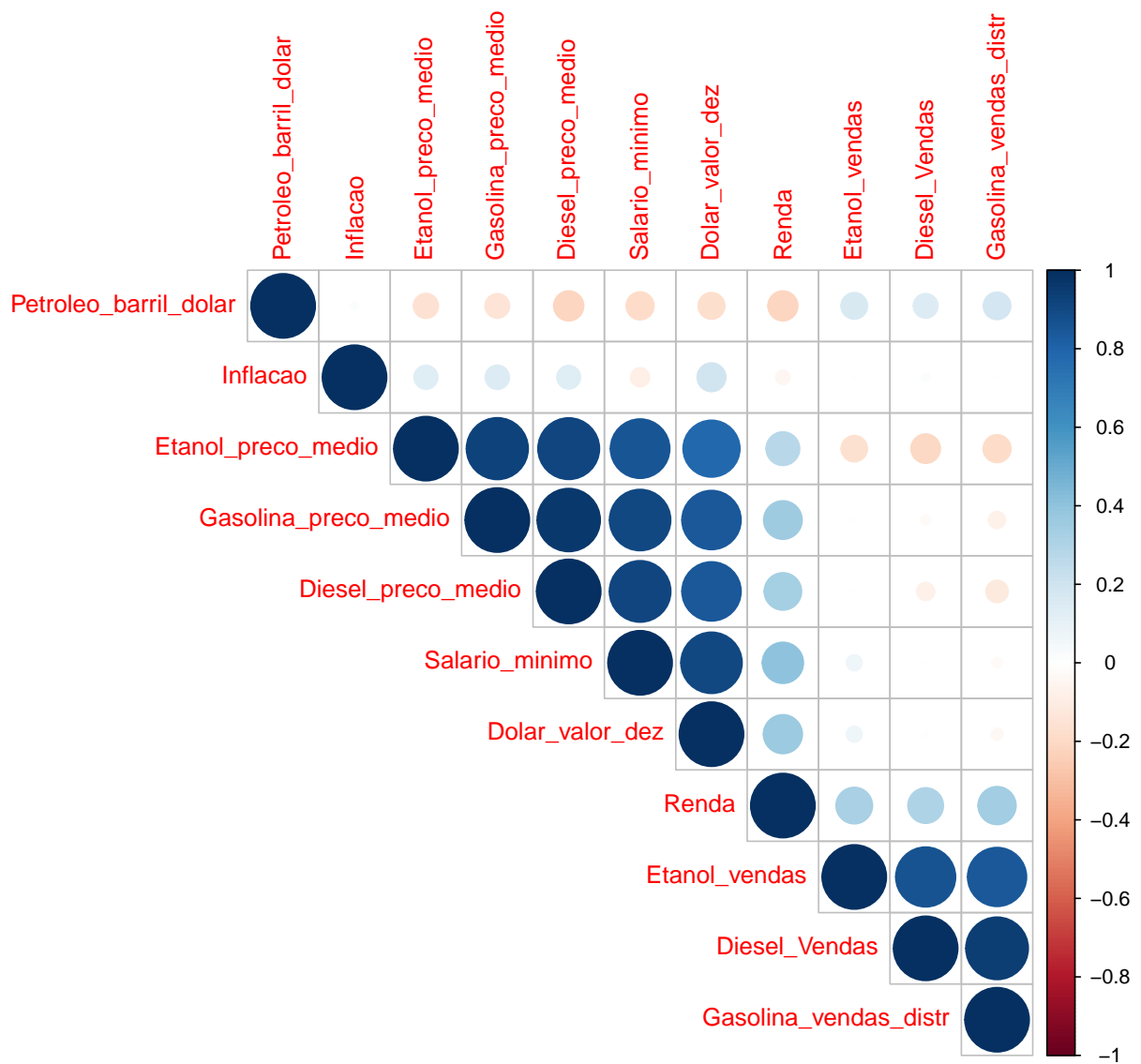
É interessante notar como cada variável se correlaciona de forma muito distinta umas com as outras e para se ter uma ideia melhor dessa correlação, vou criar alguns gráficos para visualizar esses comportamentos de maneira mais clara.

Parte 2 - Correlogramas

Correlograma de bolhas

Muito usado para visualizar de maneira gráfica como a correlação se dá com cada variável do dataset.

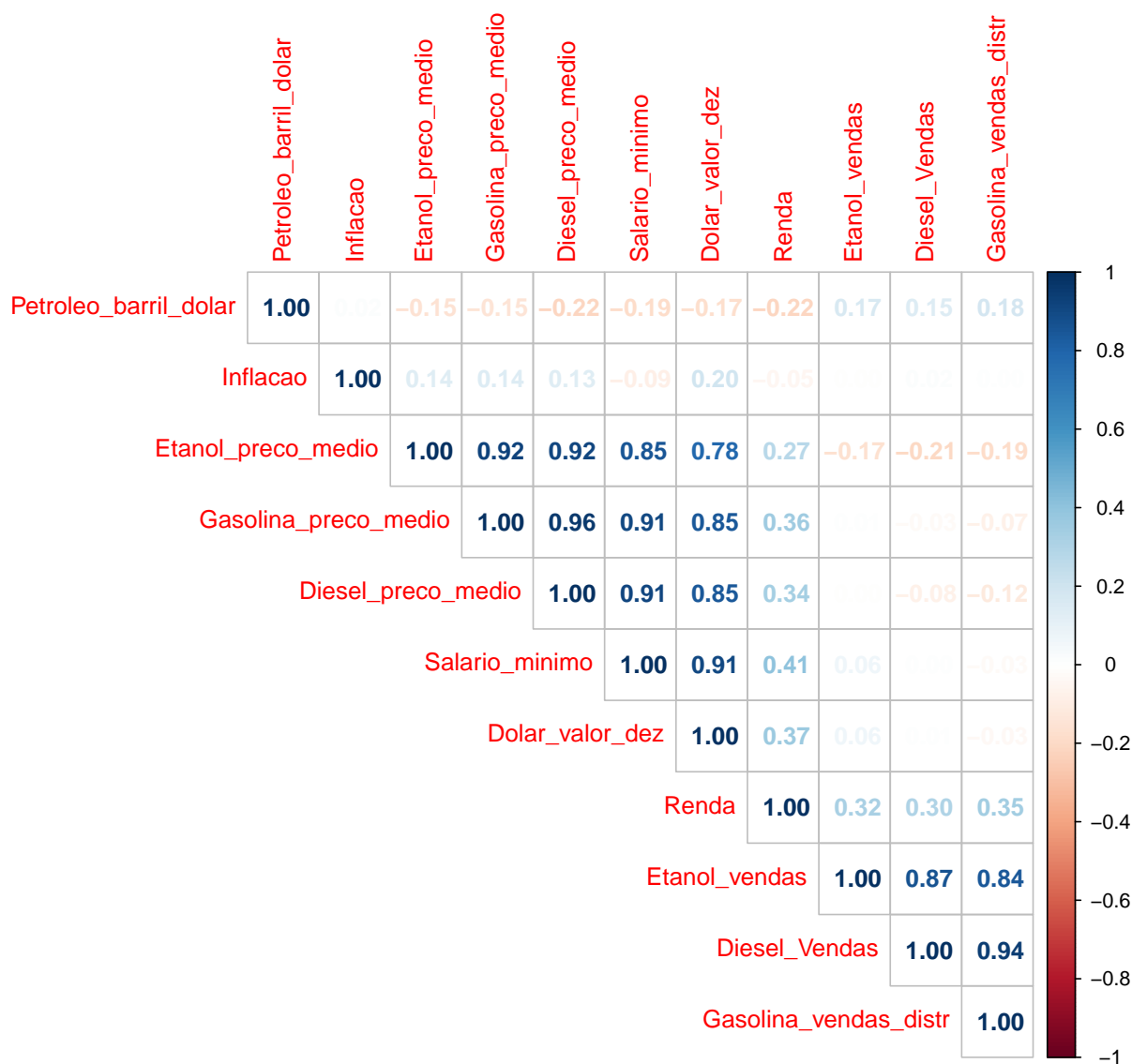
```
corrplot(correl$r, type = "upper", order = "hclust")
```



Correlograma com valores de correlação

Varição do correlograma acima com valores numéricos ao invés de gráficos

```
corrplot(correl$r, type = "upper", order = "hclust", method = "number")
```



Podemos ver que há uma correlação positiva forte entre:

- Preço médio do etanol e o preço médio da gasolina (0.92);
- Preço médio do etanol e o preço médio do diesel (0.92);
- Preço médio do etanol e o salário mínimo (0.85);
- Preço médio do etanol e o valor do dólar (0.78);
- Preço médio da gasolina e preço médio do diesel (0.96);
- Preço médio da gasolina e o valor do dólar (0.85);
- Preço médio do Diesel e o salário mínimo (0.91);
- Preço médio do Diesel e o valor do dólar (0.85);
- Salário mínimo e o valor do dólar (0.91);
- Venda de etanol e venda de diesel (0.87);
- Venda de etanol e venda de gasolina (0.84);
- Venda de diesel e venda de gasolina (0.94);

Essas correlações positivas fortes indicam que quando uma dessas variáveis aumenta, a outra aumenta também

A correlação negativa entre Inflação e Renda / Salário-mínimo foi um achado muito curioso, já que qualquer um pensaria que a correlação seria positiva já que teoricamente a renda e/ou salário-mínimo deveriam aumentar de acordo com o aumento da inflação, mas vimos que isso não acontece. **Entretanto é importante sempre deixar claro que correlação não significa causa!**

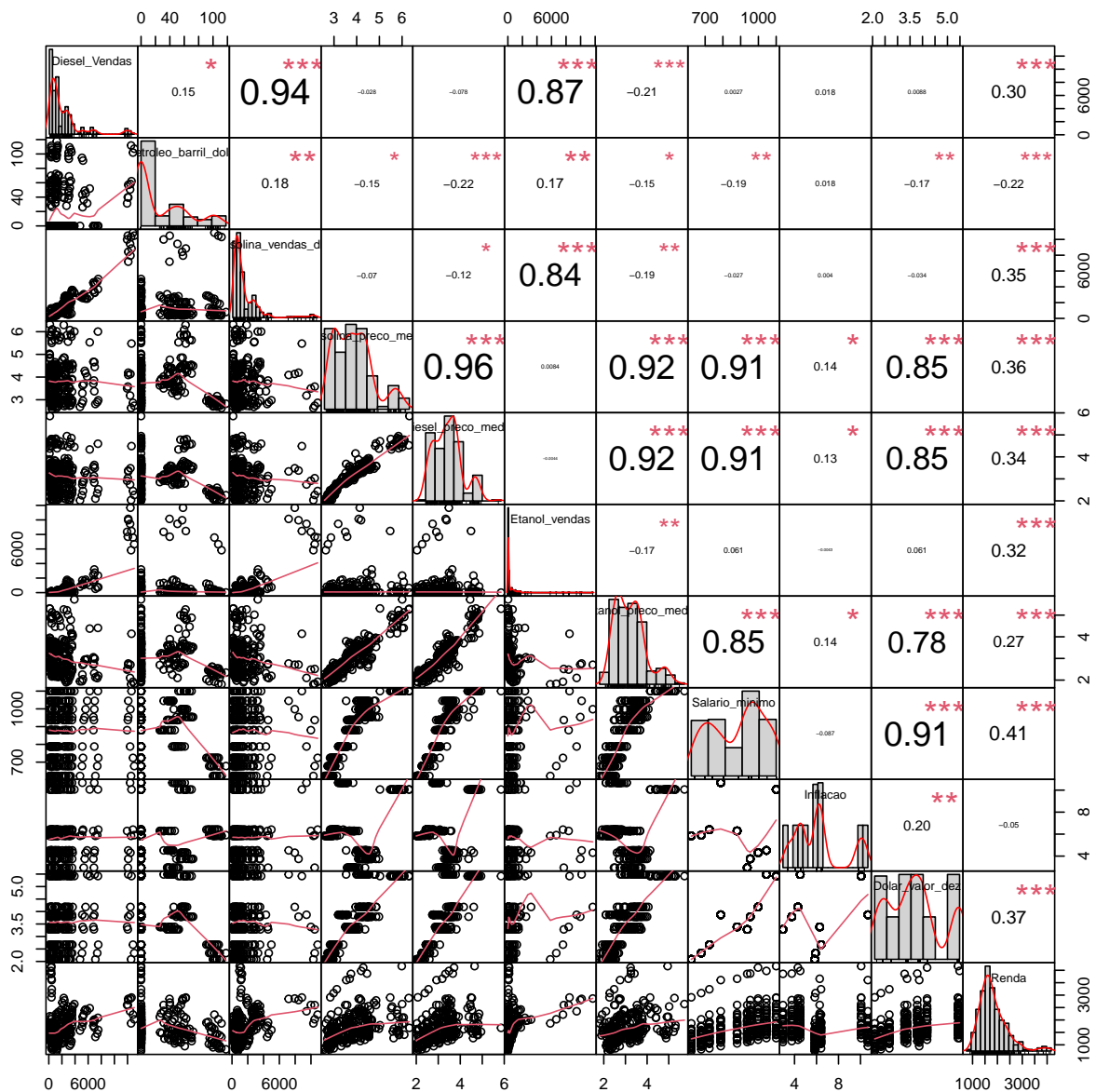
Plotarei um gráfico mais aprofundado de correlação para obter uma visão geral melhor das correlações mostradas acima:

Parte 3 - Gráfico de correlação

Gráfico estatístico muito interessante que mostra além do valor da correlação, o gráfico de dispersão entre cada variável do dataset, nível de significância, e histograma mostrando como os dados estão distribuídos no nosso conjunto de dados. Essa ferramenta é muito importante para se ter um overview sobre os nossos dados e como eles se relacionam e estão distribuídos.

Usando o método de Pearson

```
chart.Correlation(dados[,3:13], histogram = T, pch = 19, method = "pearson")
```



No gráfico acima:

- A distribuição (histograma) de cada variável é mostrada na diagonal;
- Na parte inferior da diagonal: gráficos de dispersão de cada correlação é mostrado para evidenciar como a correlação se comporta (positiva, neutra ou negativa);
- Na parte superior da diagonal: valor da correlação + o nível de significância mostrado como estrelas;
- Cada nível de significância é associado a um símbolo: p-values(0, 0.001, 0.01, 0.05, 0.1, 1) <=> símbolos (“”, “”, “”, “”, “”).

Parte 3.1 - Correlação das variáveis no qual verificamos que houve uma mudança de comportamento no dashboard

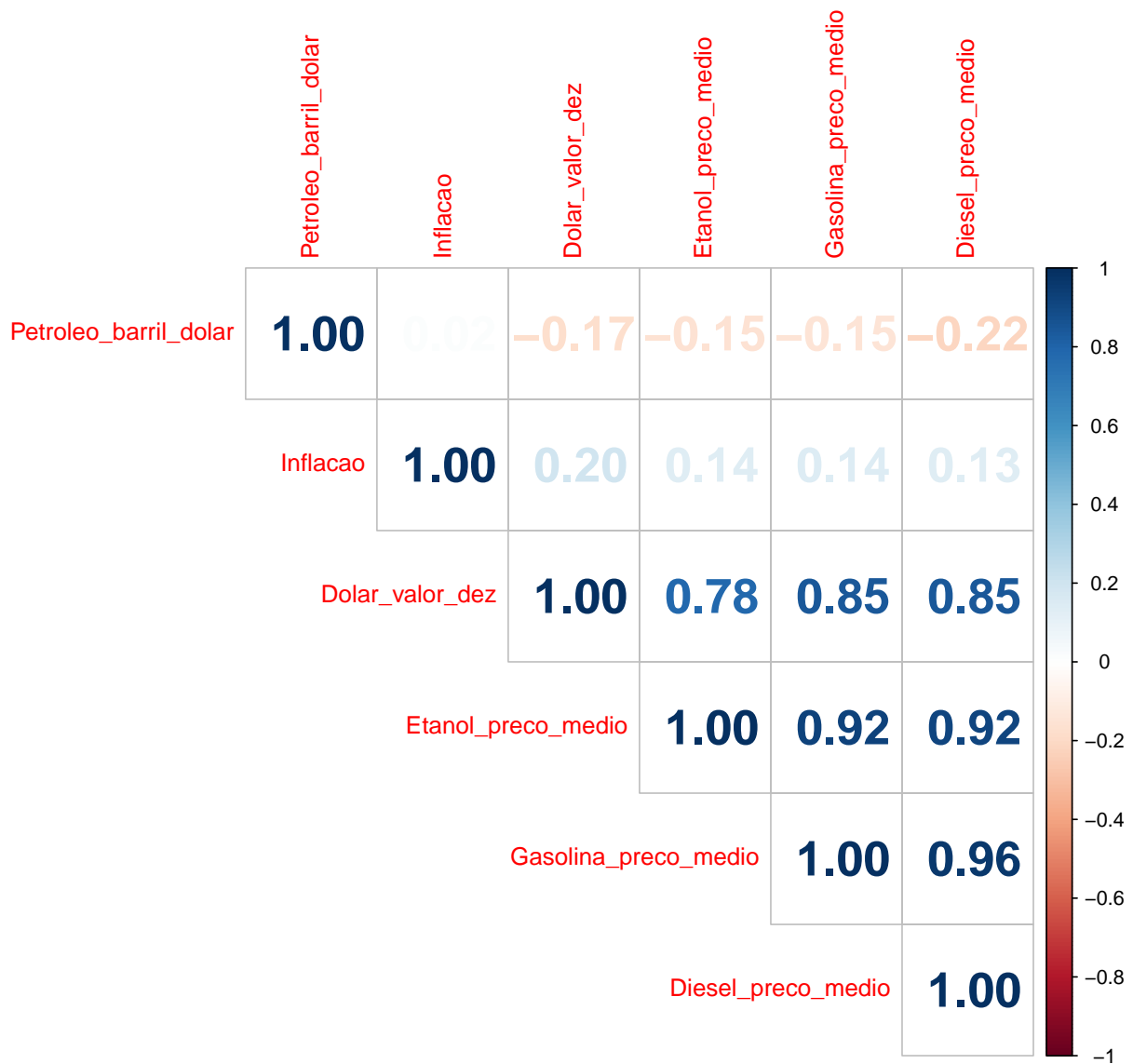
Separando o arquivo em 2 períodos diferentes pois verificamos no Dashboard que a correlação das seguintes variáveis mudou antes e depois de 2017:

- Valor do Dólar;
- Preço do Barril do Petróleo.
- Preço do Diesel;
- Preço da Gasolina;
- Preço do Etanol;
- Inflação.

```
dados_primeiro <- dados %>%  
  filter(Ano == c(2012:2016))  
  
dados_segundo <- dados %>%  
  filter(Ano == c(2017:2021))
```

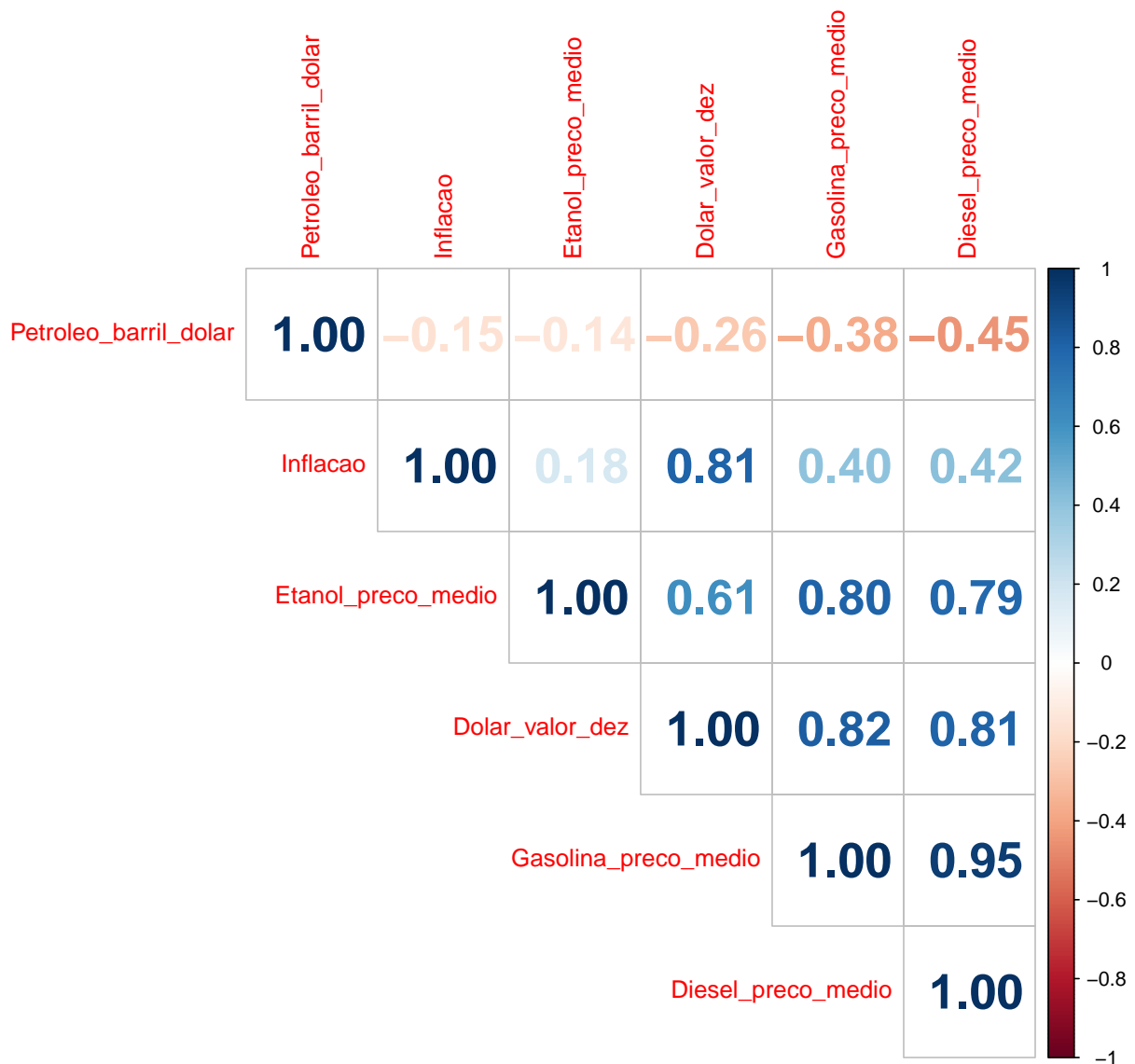
Análise de Correlação de todo o período

```
correl_total <- rcorr(as.matrix(dados[,c(4, 6:7, 9, 11, 12)]))  
  
corrplot(correl_total$r, type = "upper", order = "hclust", method = "number", number.cex = 2)
```



Análise de Correlação do período de 2012 até 2016

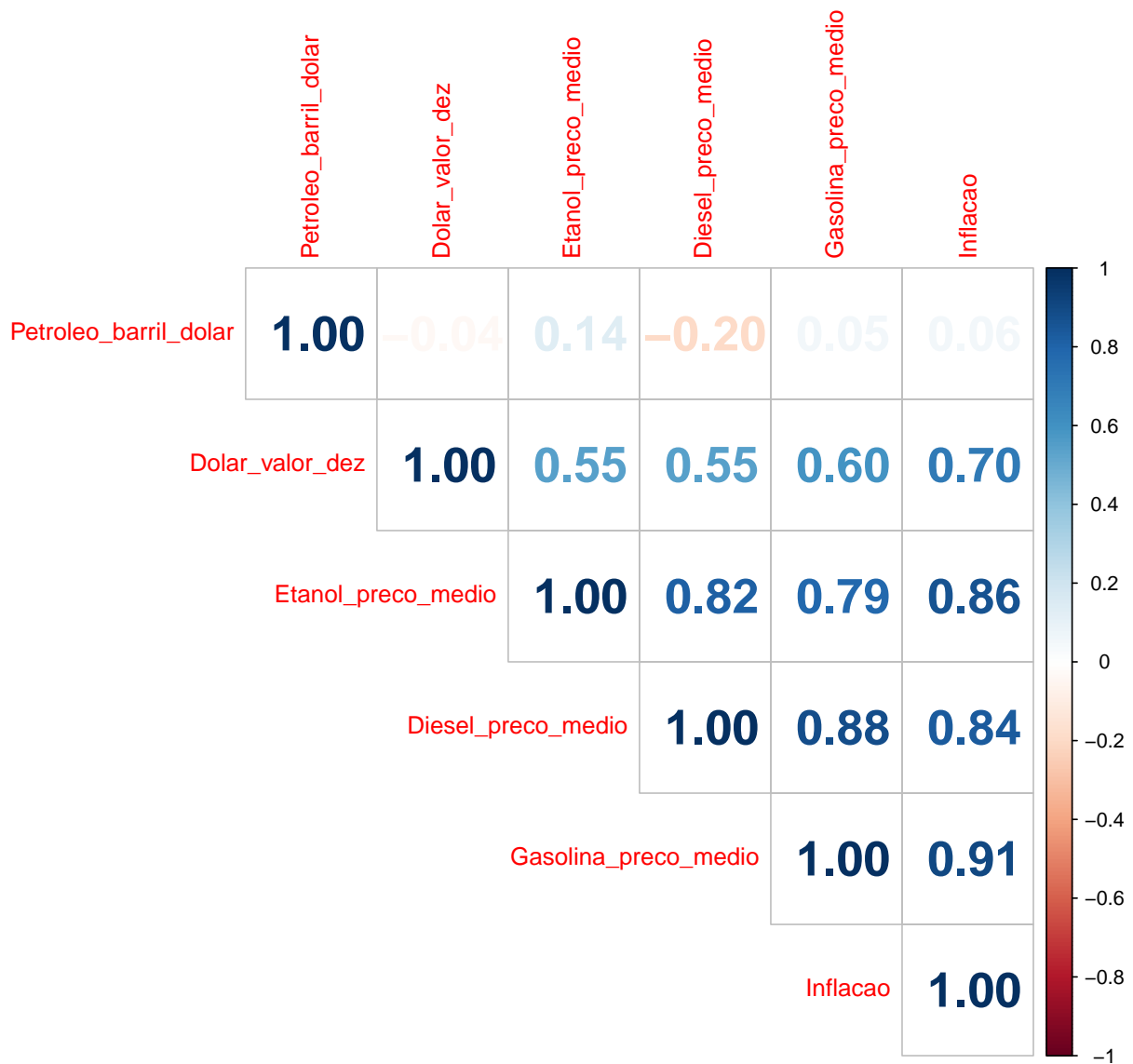
```
correl_primeiro <- rcorr(as.matrix(dados_primeiro[,c(4, 6:7, 9, 11, 12)]))
corrplot(correl_primeiro$r, type = "upper", order = "hclust", method = "number", number.cex = 2)
```



Vemos que a correlação de todas as variáveis selecionadas são muito baixas, principalmente a referente ao preço do barril de petróleo e valor do dólar.

Análise de Correlação do período de 2017 até 2021

```
correl_segundo <- rcorr(as.matrix(dados_segundo[,c(4, 6:7, 9, 11, 12)]))
corrplot(correl_segundo$r, type = "upper", order = "hclust", method = "number", number.cex = 2)
```



Aqui já vemos que a correlação para os preços dos combustíveis e inflação é bem forte. A correlação do preço do barril de petróleo e valor do dólar continuou muito fraca.

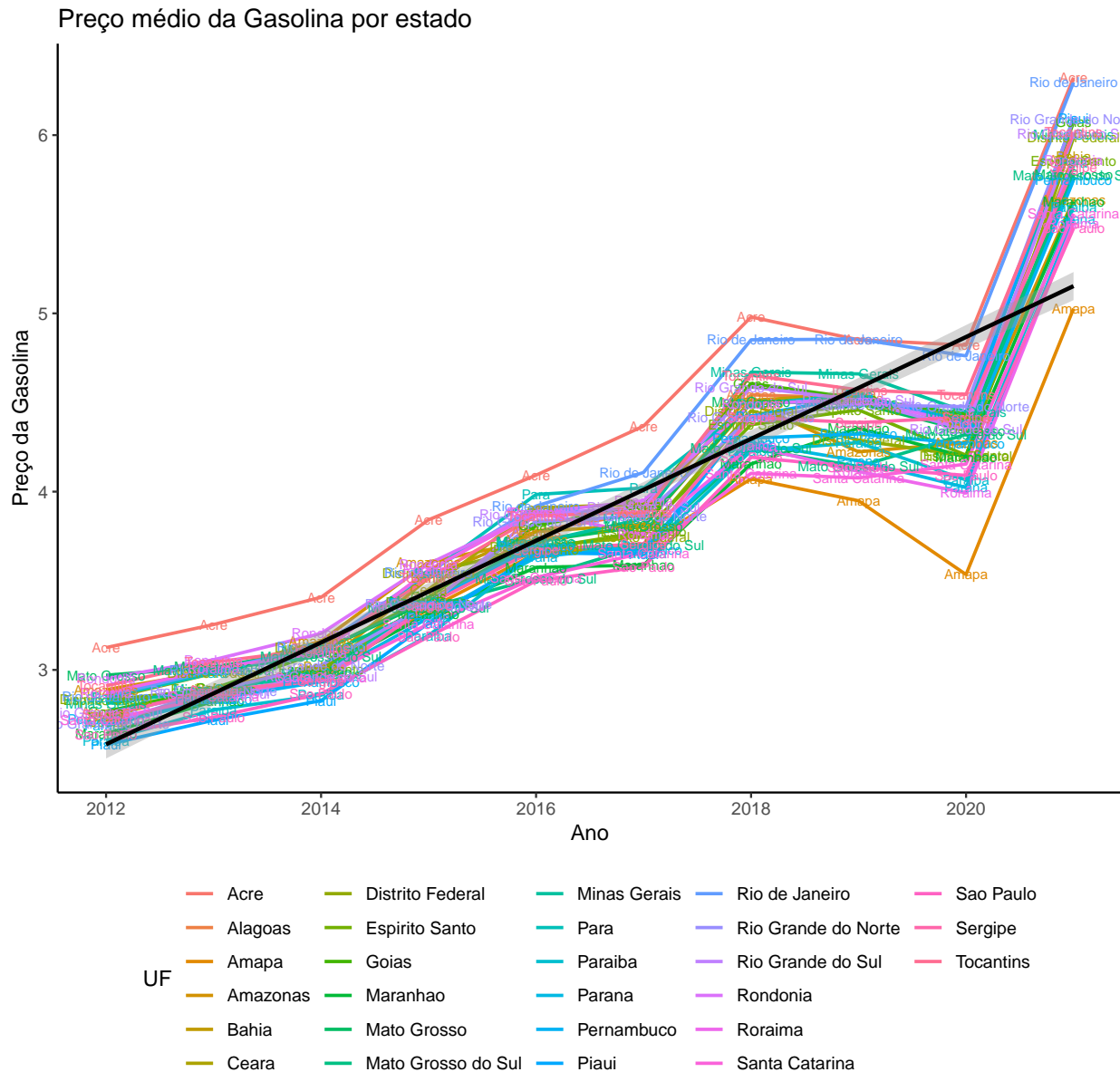
Parte 4 - Gráficos de linha

Criarei alguns gráficos de linha para analisar como as seguintes variáveis se comportaram durante a última década:

- Preço da Gasolina;
- Preço do Diesel;
- Preço do Etanol;
- Renda média.

Gráfico 1 - Gasolina

```
g1 <- ggplot(dados, aes(x = Ano, y = Gasolina_preco_medio, colour = UF)) +  
  geom_line(size = 0.8) +  
  labs(  
    title = "Preço médio da Gasolina por estado",  
    x = "Ano",  
    y = "Preço da Gasolina",  
    caption = "Fonte: ANP"  
  ) +  
  geom_text(label = dados$UF, size = 2.5, show.legend = FALSE) +  
  stat_smooth(method=lm, col='black', se = T) +  
  theme_classic()  
  
g1 + theme(legend.position = "bottom")
```



Fonte: ANP

Acre e Rio de Janeiro possuem a gasolina mais cara do país, e Amapá a mais barata.

Correlação do Preço da Gasolina por região

```
g2 <- dados %>%
  group_by(Regiao, Ano) %>%
  summarise(preco_gasol = mean(Gasolina_preco_medio))

g2g <- ggplot(g2, aes(x = Ano, y = preco_gasol, colour = Regiao)) +
  geom_line(size = 1) +
  labs(
    title = "Preço médio da Gasolina por região",
    x = "Ano",
    y = "Preço da Gasolina",
    caption = "Fonte: ANP"
```

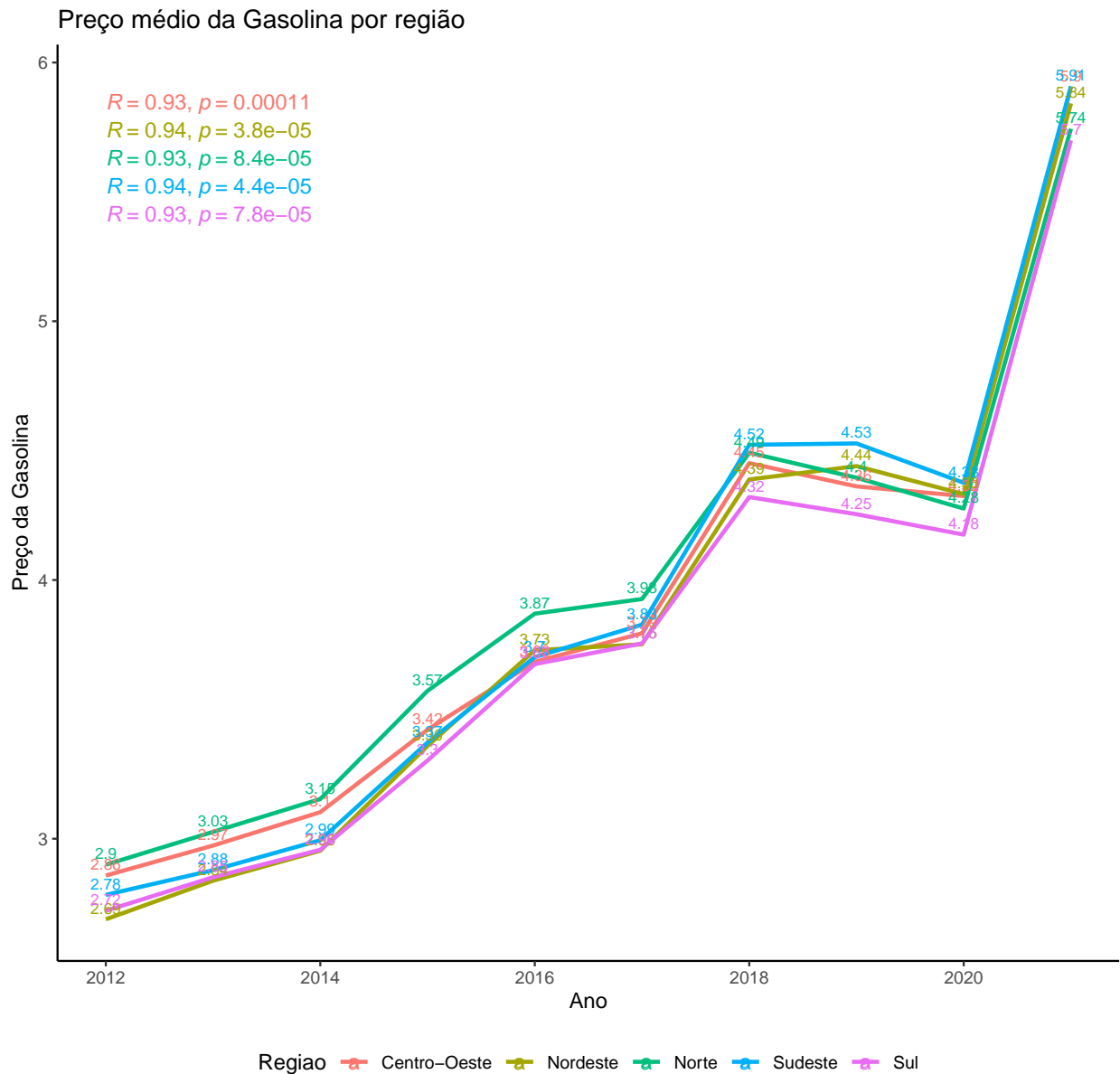


```

) +
  geom_text(label = round(g2$preco_gasol, 2), vjust = -0.5, size = 2.8) +
  stat_cor() +
  theme_classic()

g2g + theme(legend.position = "bottom")

```



Fonte: ANP

Região Sul na série histórica apresenta a gasolina mais barata, e a região sudeste ocupou nos últimos anos o posto de gasolina mais cara do país.

E a correlação bem alta nos mostra que o preço tem aumentado bastante de acordo com o passar dos anos.

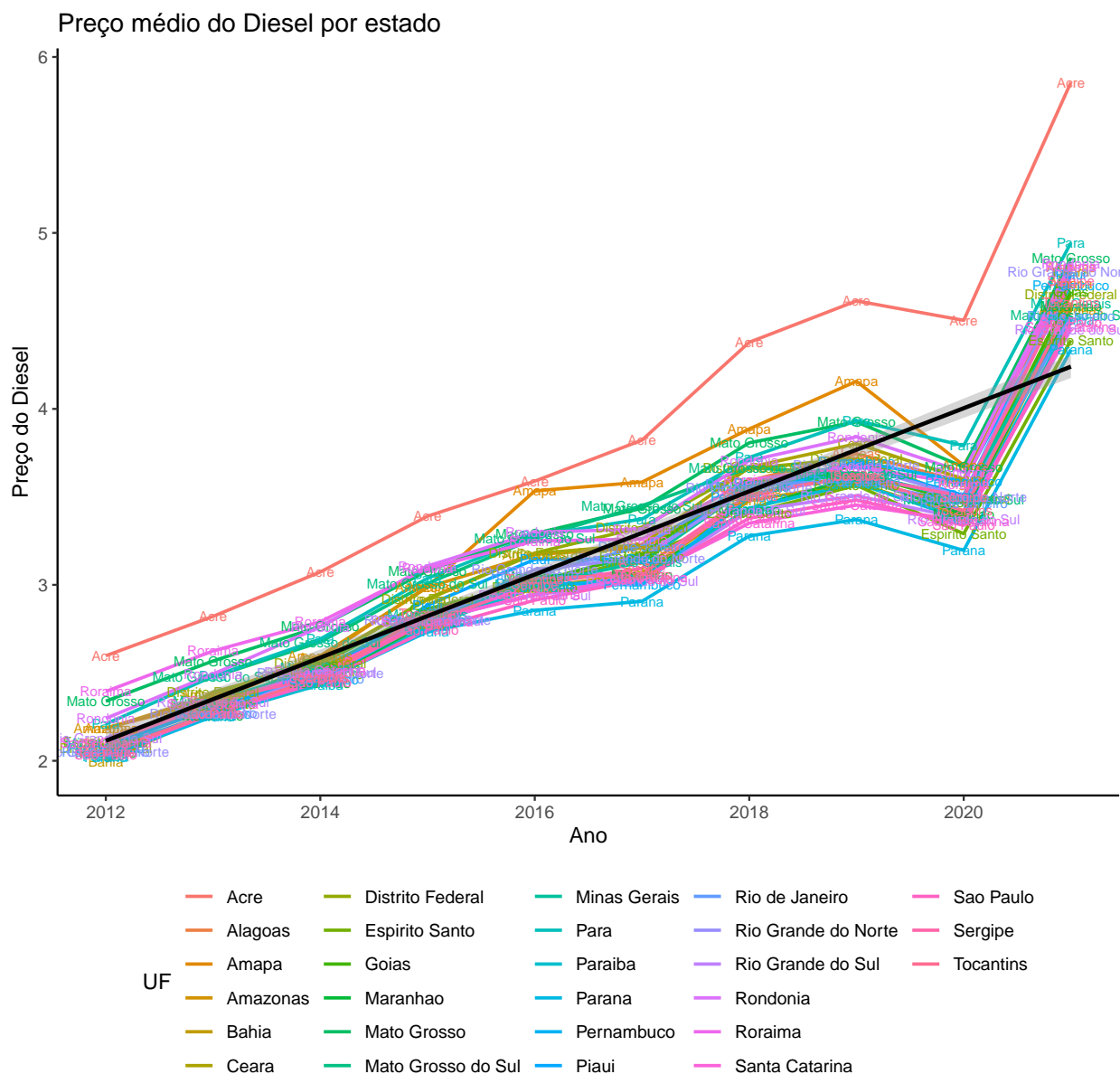
Gráfico 2 - Diesel

```

g3 <- ggplot(dados, aes(x = Ano, y = Diesel_preco_medio, colour = UF)) +
  geom_line(size = 0.8) +
  labs(
    title = "Preço médio do Diesel por estado",
    x = "Ano",
    y = "Preço do Diesel",
    caption = "Fonte: ANP"
  ) +
  geom_text(label = dados$UF, size = 2.5, show.legend = FALSE) +
  stat_smooth(method=lm, col='black', se = T) +
  theme_classic()

g3 + theme(legend.position = "bottom")

```



Fonte: ANP

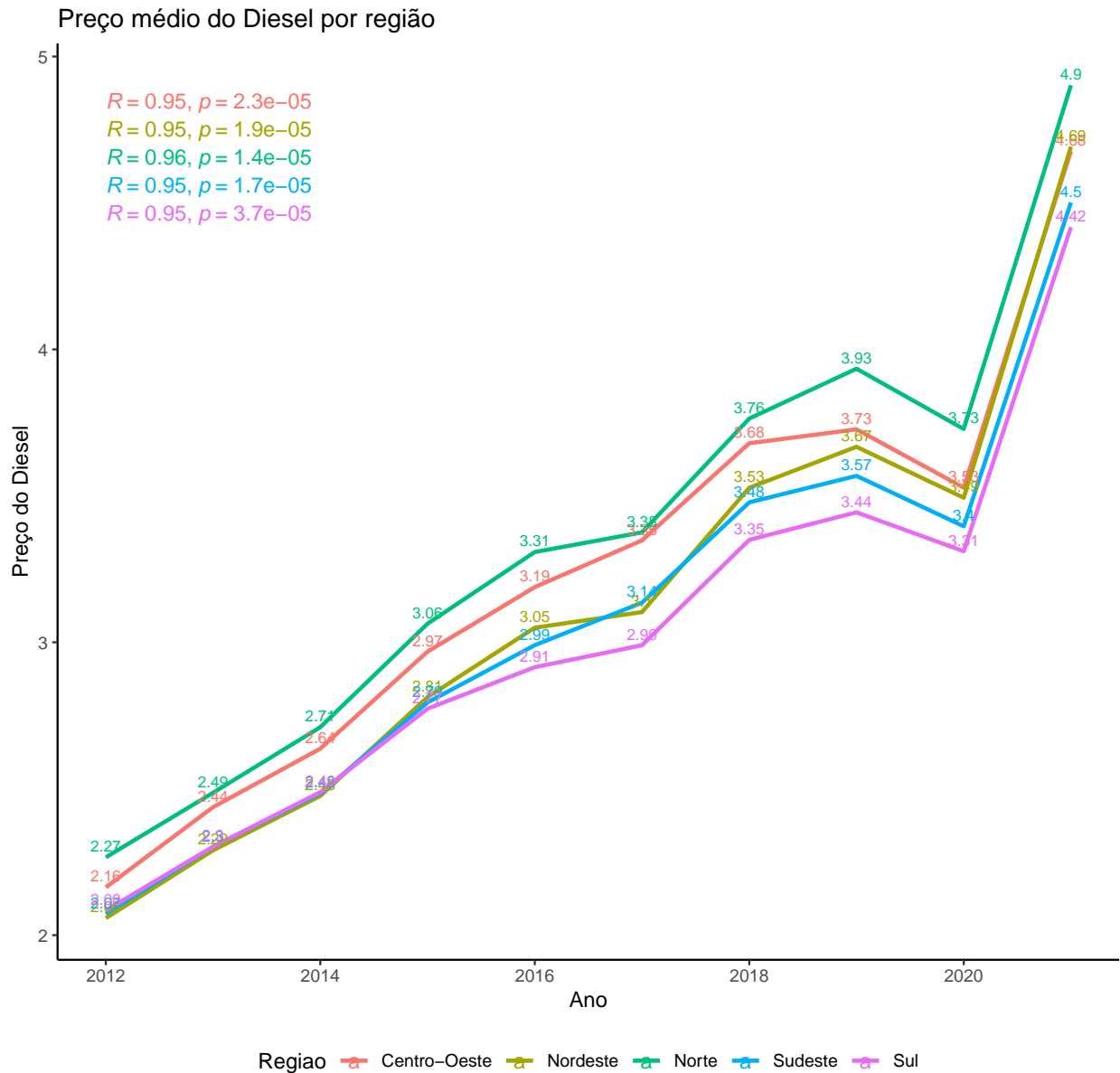
Acre possui o Diesel mais caro do País, e Paraná o mais barato.

Correlação do Preço do Diesel por região.

```
g4 <- dados %>%
  group_by(Regiao, Ano) %>%
  summarise(preco_diesel = mean(Diesel_preco_medio))

g4g <- ggplot(g4, aes(x = Ano, y = preco_diesel, colour = Regiao)) +
  geom_line(size = 1) +
  labs(
    title = "Preço médio do Diesel por região",
    x = "Ano",
    y = "Preço do Diesel",
    caption = "Fonte: ANP"
  ) +
  stat_cor() +
  geom_text(label = round(g4$preco_diesel, 2), show.legend = FALSE, vjust = -0.5, size = 2.8) +
  theme_classic()

g4g + theme(legend.position = "bottom")
```



Fonte: ANP

Norte apresenta o Diesel mais caro, e Sul o mais barato.

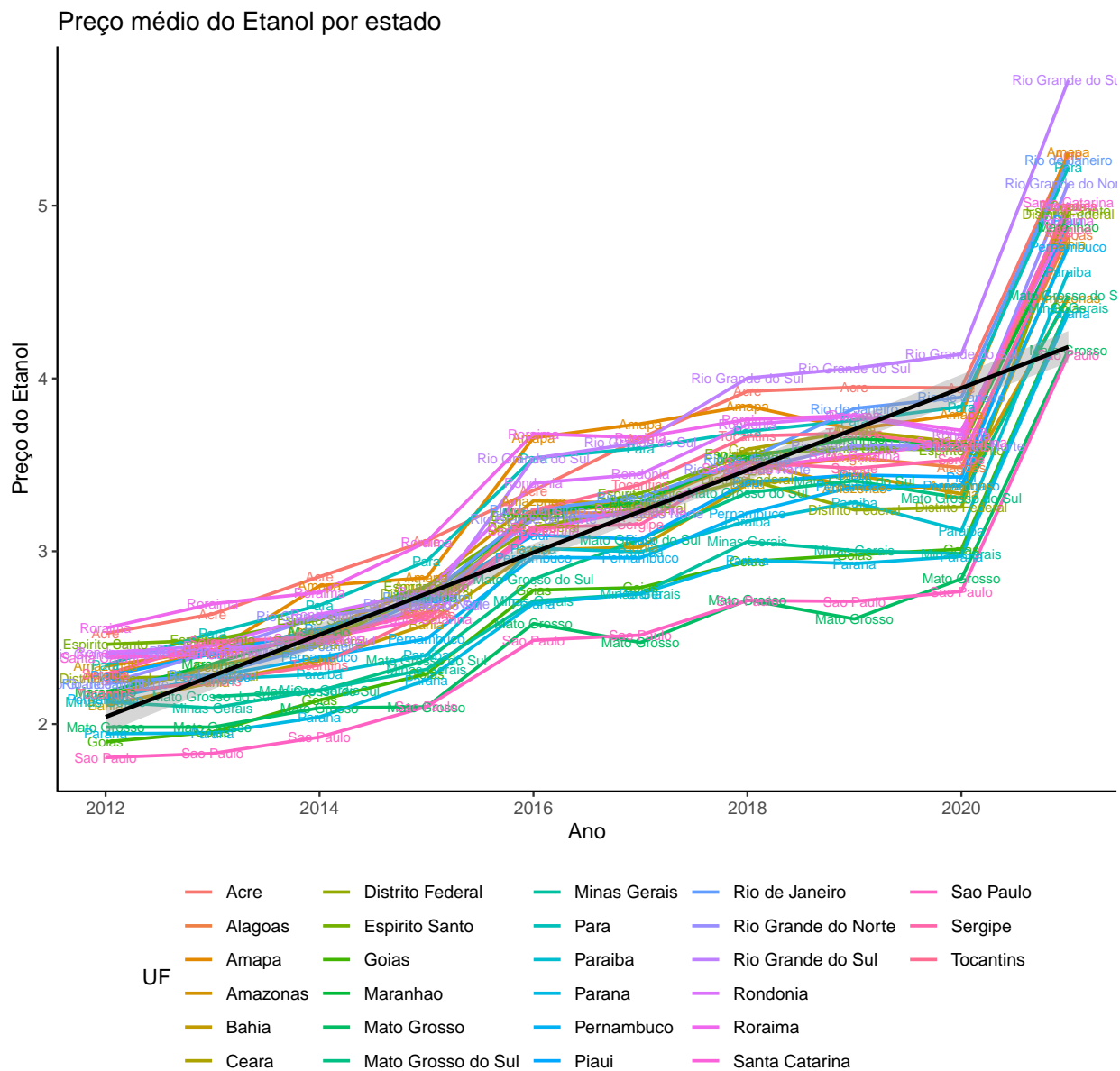
Outra vez a correlação altíssima nos mostra como preço tem aumentado significativamente durante a última década.

Gráfico 3 - Etanol

```
g5 <- ggplot(dados, aes(x = Ano, y = Etanol_preco_medio, colour = UF)) +
  geom_line(size = 0.8) +
  labs(
    title = "Preço médio do Etanol por estado",
    x = "Ano",
    y = "Preço do Etanol",
    caption = "Fonte: ANP"
  ) +
```

```
geom_text(label = dados$UF, size = 2.5, show.legend = FALSE) +
stat_smooth(method=lm, col='black', se = T) +
theme_classic()

g5 + theme(legend.position = "bottom")
```



Fonte: ANP

Rio Grande do Sul possui o etanol mais caro do País, e São Paulo o mais barato.

É interessante notar como a variabilidade do preço do etanol é muito maior entre os estados do que com o preço da gasolina e do diesel.

Correlação do Preço do Etanol por região

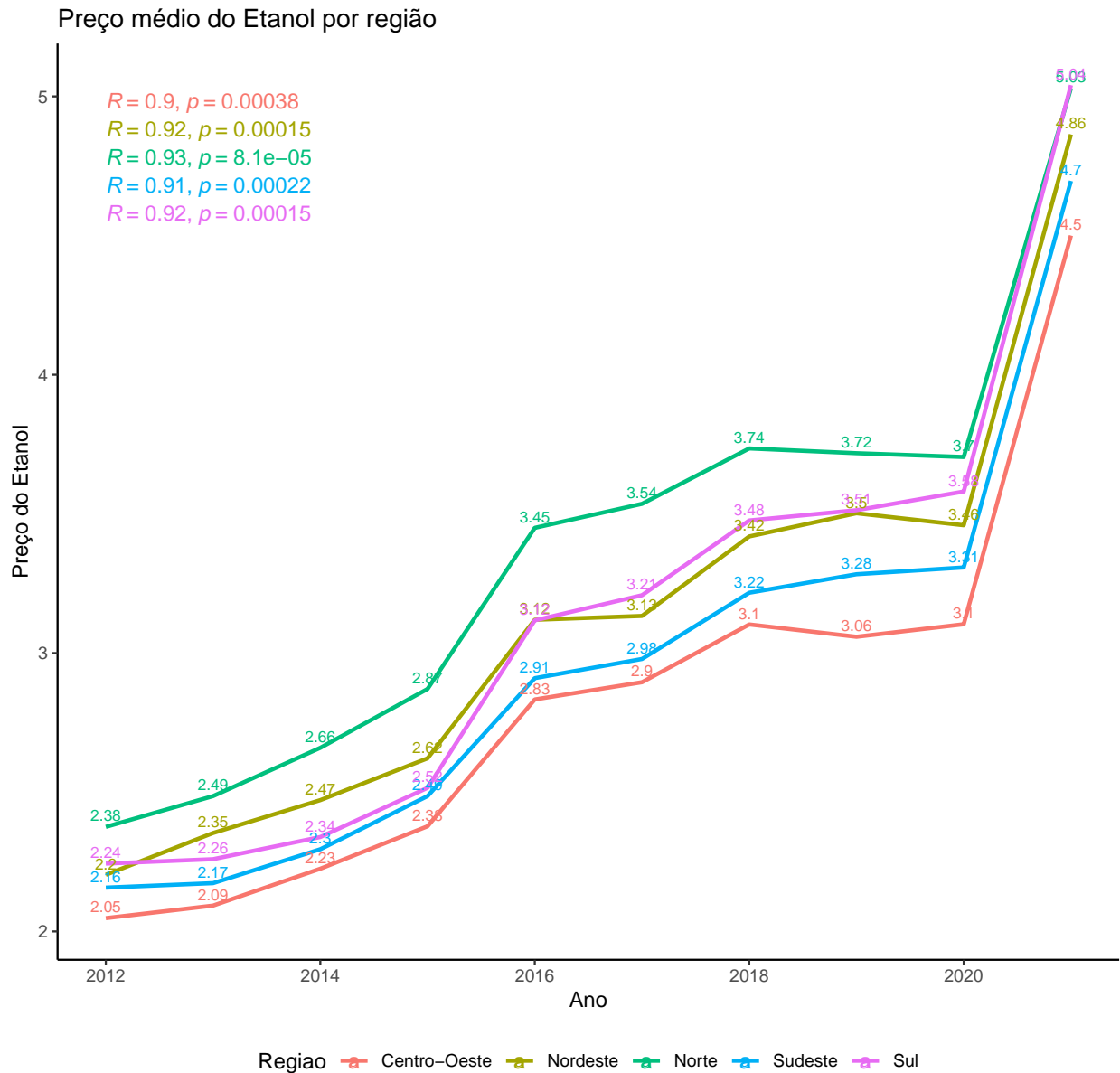
```

g6 <- dados %>%
  group_by(Regiao, Ano) %>%
  summarise(preco_etanol = mean(Etanol_preco_medio))

g6g <- ggplot(g6, aes(x = Ano, y = preco_etanol, colour = Regiao)) +
  geom_line(size = 1) +
  labs(
    title = "Preço médio do Etanol por região",
    x = "Ano",
    y = "Preço do Etanol",
    caption = "Fonte: ANP"
  ) +
  stat_cor() +
  geom_text(label = round(g6$preco_etanol, 2), vjust = -0.5, size = 2.8) +
  theme_classic()

g6g + theme(legend.position = "bottom")

```



Fonte: ANP

Norte e Sul com os preços do etanol mais elevados do país atualmente, e Centro-Oeste o mais barato.

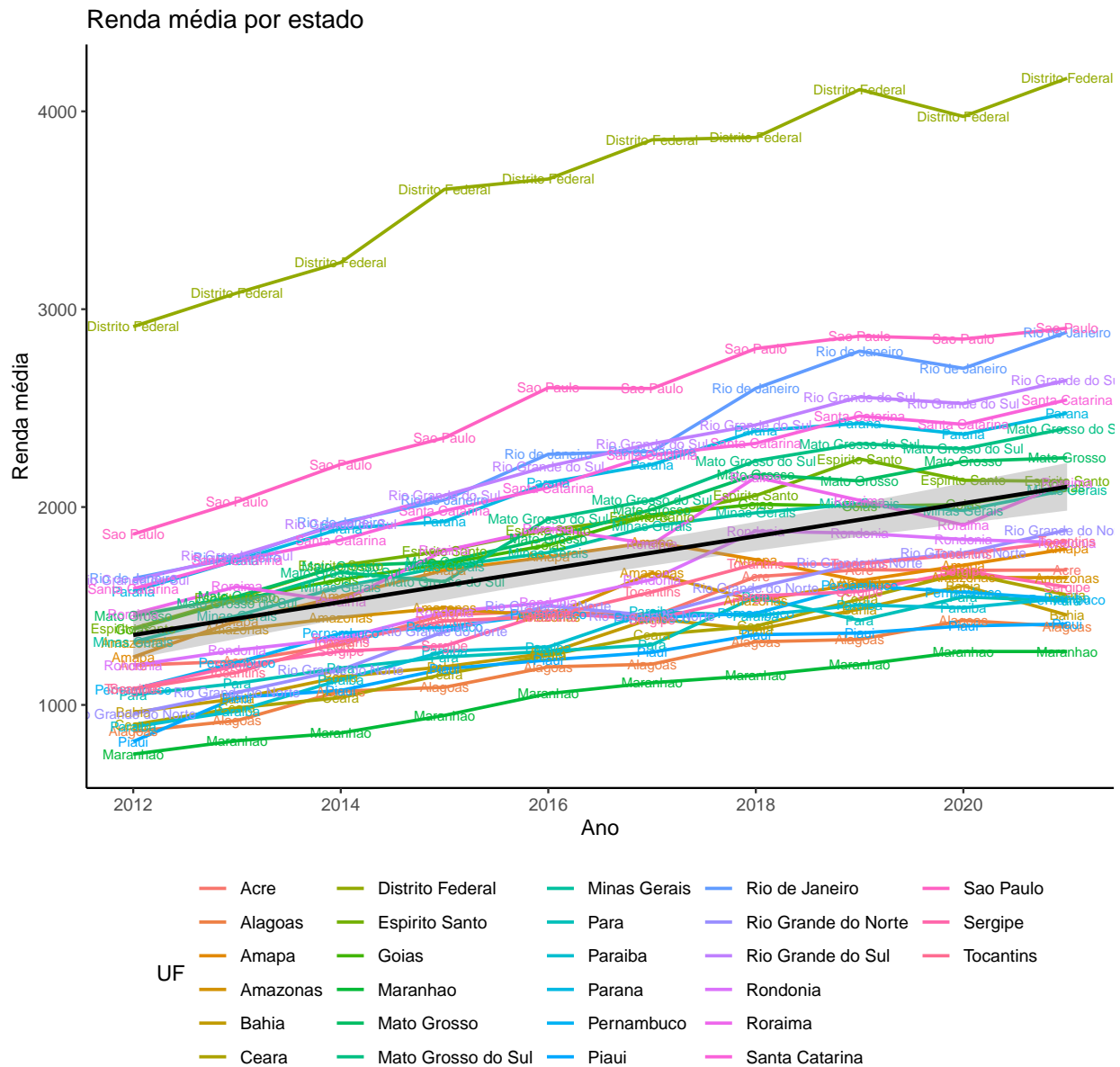
Valores de correlação são ligeiramente mais baixos do que os do Diesel e da Gasolina, e isso é facilmente explicado devido a maior variabilidade do preço entre os estados/regiões, conforme mencionado anteriormente.

Gráfico 4 - Renda

```
g7 <- ggplot(dados, aes(x = Ano, y = Renda, colour = UF)) +
  geom_line(size = 0.8) +
  labs(
    title = "Renda média por estado",
    x = "Ano",
    y = "Renda média",
    caption = "Fonte: IBGE"
  ) +
```

```
geom_text(label = dados$UF, size = 2.5, show.legend = FALSE) +
stat_smooth(method=lm, col='black', se = T) +
theme_classic()

g7 + theme(legend.position = "bottom")
```



Fonte: IBGE

Brasília, como capital federal, se comporta como um outlier clássico no nosso gráfico acima, e o Maranhão é historicamente o estado mais pobre (com menor renda) do Brasil.

Assim como no caso do Etanol, é interessante notar como a variabilidade da renda é muito grande entre os estados. Isso nos mostra o quão desigual é a distribuição de renda por estado no país.

Correlação da Renda por região

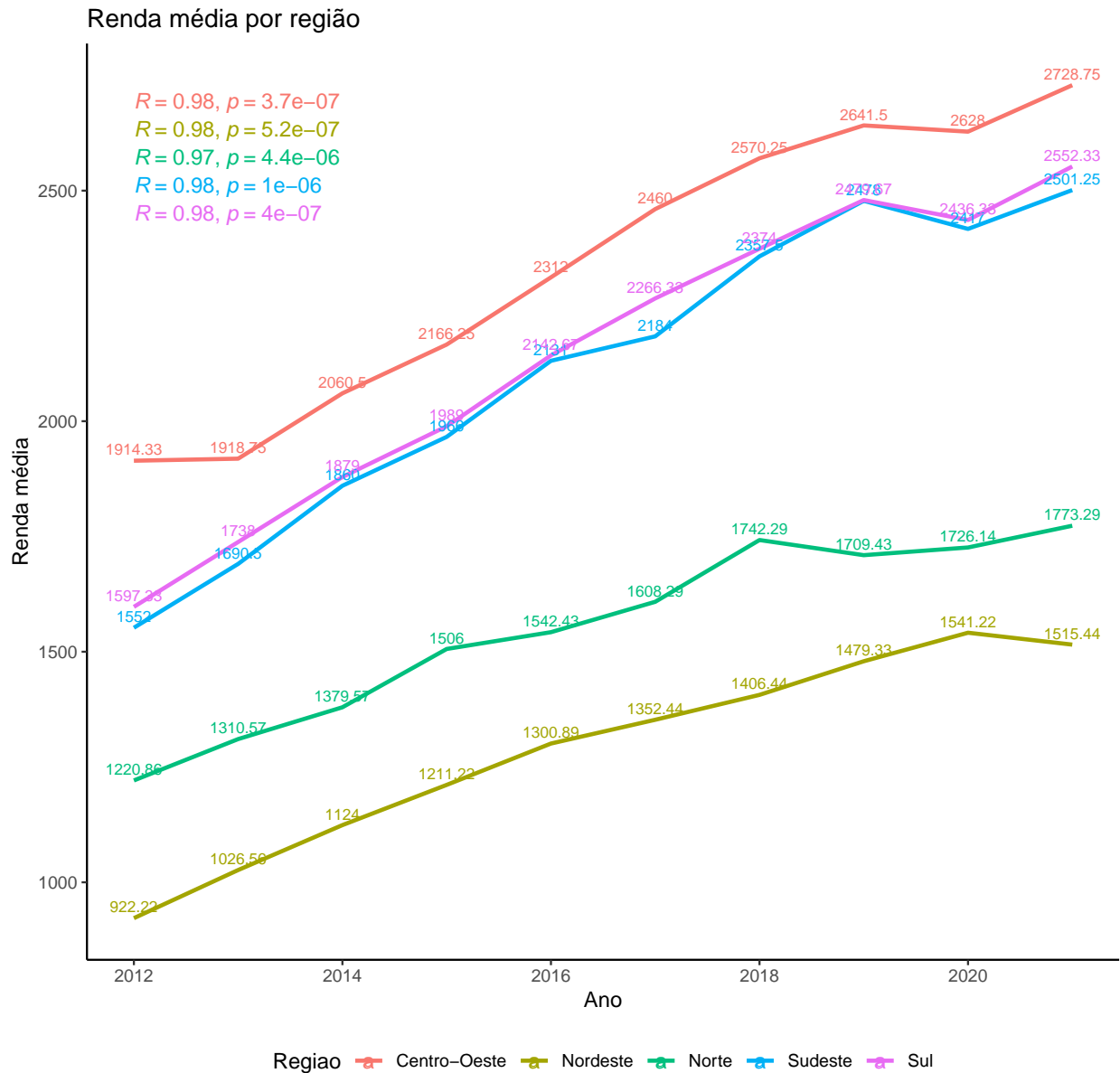

```

g8 <- dados %>%
  group_by(Regiao, Ano) %>%
  summarise(Renda = mean(Renda))

g8g <- ggplot(g8, aes(x = Ano, y = Renda, colour = Regiao)) +
  geom_line(size = 1) +
  labs(
    title = "Renda média por região",
    x = "Ano",
    y = "Renda média",
    caption = "Fonte: IBGE"
  ) +
  stat_cor() +
  geom_text(label = round(g8$Renda, 2), vjust = -0.5, size = 2.8) +
  theme_classic()

g8g + theme(legend.position = "bottom")

```



Fonte: IBGE

Vemos claramente no gráfico acima que existem “2 Brasis” no quesito renda, dividido entre Centro-sul e Norte-Nordeste, e isso mais uma vez nos mostra o quão grande é a desigualdade em distribuição de renda dentro do próprio país.

Brasília por ser um outlier pode estar puxando a média da região Centro-Oeste para cima. **Pensando nisso, resolvi comparar a média e a mediana do Centro-Oeste para ver se o efeito outlier de Brasília influencia substancialmente essa média da região.**

Calculando a Média e a Mediana da região Centro-Oeste

Calculando a média de renda da região Centro-Oeste de 2012 até 2021

```
CO_media <- g8 %>%
  filter(Regiao == "Centro-Oeste") %>%
  group_by(Regiao) %>%
  summarise(media = mean(Renda))
```

```
# Calculando a mediana de renda da região Centro-Oeste de 2012 até 2021
```

```
CO_mediana <- g8 %>%  
  filter(Regiao == "Centro-Oeste") %>%  
  group_by(Regiao) %>%  
  summarise(mediana = median(Renda))
```

```
cat(paste0("MÉDIA da renda do Centro-Oeste: R$", round(CO_media$media)))
```

```
## MÉDIA da renda do Centro-Oeste: R$2340
```

```
cat(paste0("MEDIANA da renda do Centro-Oeste: R$", round(CO_mediana$mediana)))
```

```
## MEDIANA da renda do Centro-Oeste: R$2386
```

Como podemos ver acima, Brasília não influencia substancialmente a média do Centro-Oeste já que a média da renda apresenta um valor ligeiramente menor que a mediana.

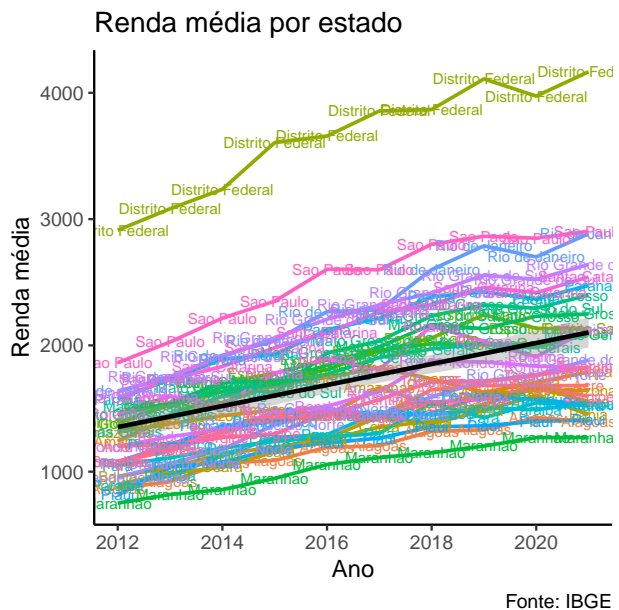
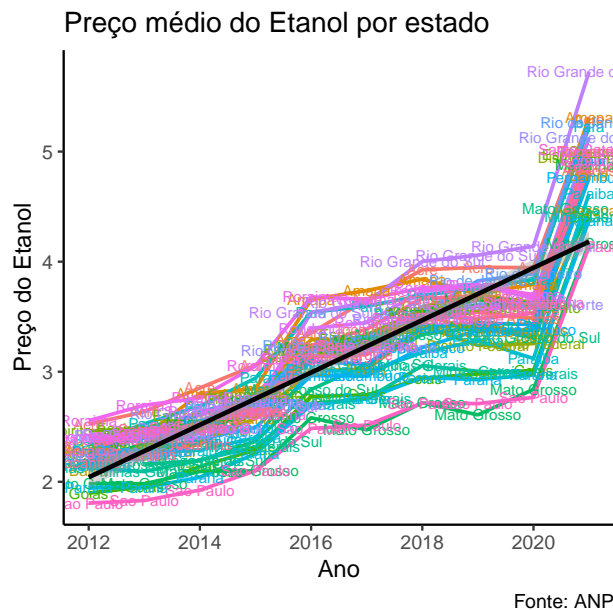
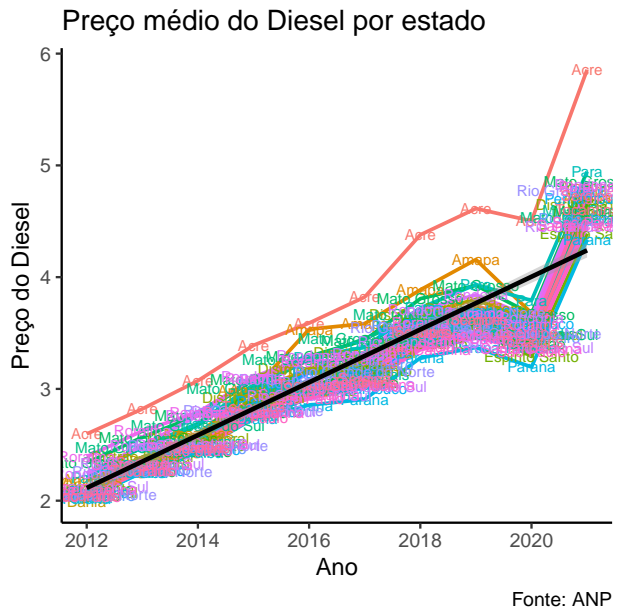
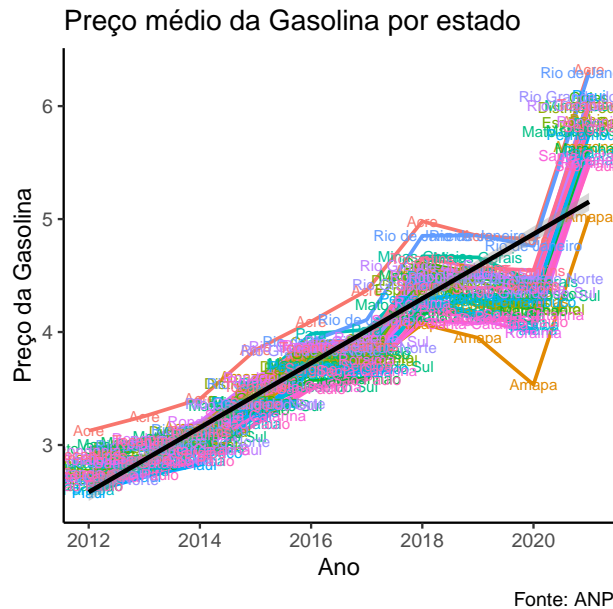
Parte 5 - Comparando lado a lado as variáveis a nível estadual

Gráficos de linha a nível estadual

```
# removendo a legenda dos gráficos acima para a visualização  
# da comparação entre eles fique melhor
```

```
c1 <- g1 + theme(legend.position = "none")  
c3 <- g3 + theme(legend.position = "none")  
c5 <- g5 + theme(legend.position = "none")  
c7 <- g7 + theme(legend.position = "none")
```

```
grid.arrange(c1, c3, c5, c7, ncol = 2)
```



Podemos ver que os gráficos referentes a combustíveis apresentam uma linha de tendência positiva com uma inclinação muito parecida, enquanto o gráfico sobre renda apresenta uma linha de tendência bem menos inclinada no sentido positivo.

Vamos ver esse comportamento através de outra perspectiva agora.

Gráficos de dispersão a nível estadual

```
d1 <- ggplot(dados, aes(x = Ano, y = Gasolina_preco_medio)) +
  geom_point(color = "blue") +
  labs(
    title = "Preço da Gasolina por estado",
    x = "Ano",
    y = "Preço da Gasolina",
    caption = "Fonte: ANP"
```

```

) +
geom_smooth(method = "lm", color = "red") +
stat_cor() +
theme_classic()

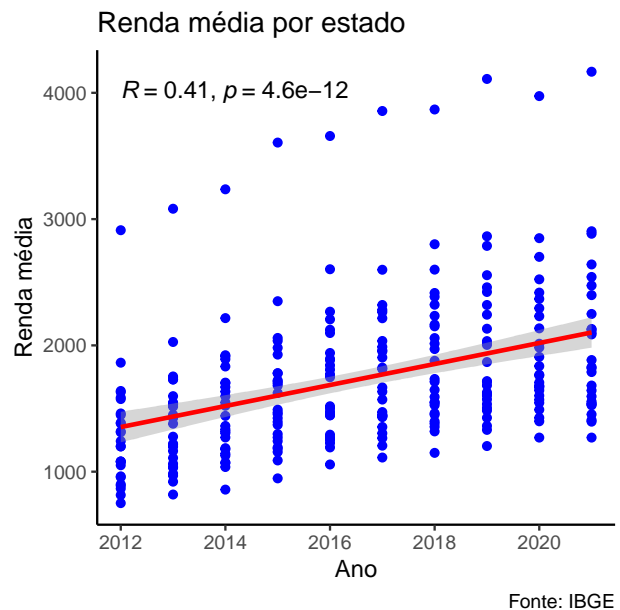
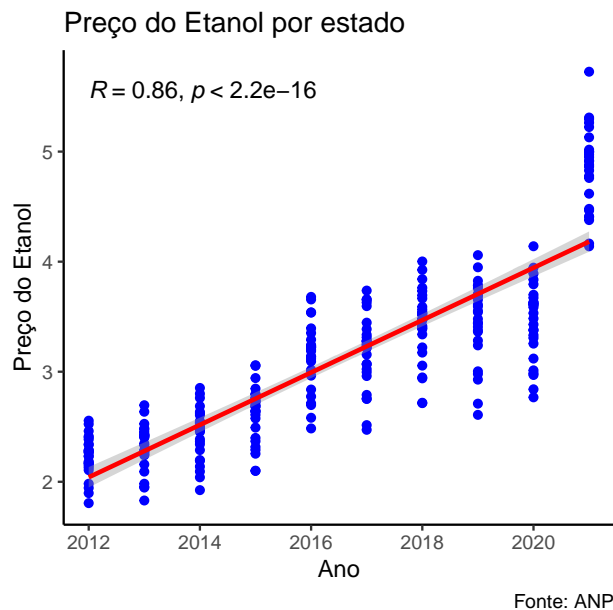
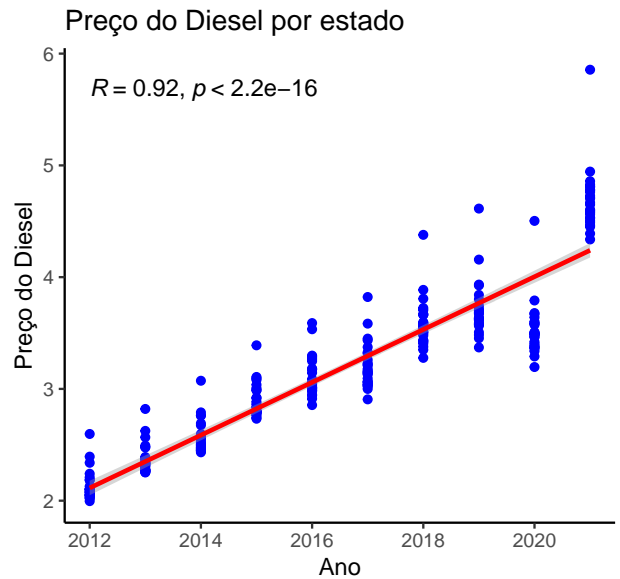
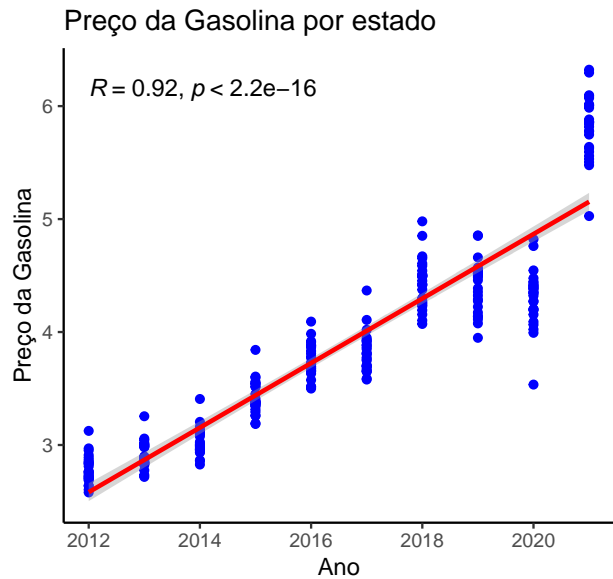
d2 <- ggplot(dados, aes(x = Ano, y = Diesel_preco_medio)) +
geom_point(color = "blue") +
labs(
  title = "Preço do Diesel por estado",
  x = "Ano",
  y = "Preço do Diesel",
  caption = "Fonte: ANP"
) +
geom_smooth(method = "lm", color = "red") +
stat_cor() +
theme_classic()

d3 <- ggplot(dados, aes(x = Ano, y = Etanol_preco_medio)) +
geom_point(color = "blue") +
labs(
  title = "Preço do Etanol por estado",
  x = "Ano",
  y = "Preço do Etanol",
  caption = "Fonte: ANP"
) +
geom_smooth(method = "lm", color = "red") +
stat_cor() +
theme_classic()

d4 <- ggplot(dados, aes(x = Ano, y = Renda)) +
geom_point(color = "blue") +
labs(
  title = "Renda média por estado",
  x = "Ano",
  y = "Renda média",
  caption = "Fonte: IBGE"
) +
geom_smooth(method = "lm", color = "red") +
stat_cor() +
theme_classic()

grid.arrange(d1, d2, d3, d4, ncol = 2)

```



Vemos que a correlação nos gráficos de dispersão dos combustíveis é muito maior que no de renda. Olhando os 4 gráficos de linha e os de dispersão lado a lado, e comparando o coeficiente de correlação de cada um, podemos afirmar o seguinte:

O preço dos combustíveis apresentou uma tendência de ter os seus respectivos valores aumentados em uma taxa muito mais rápida/elevada do que a renda média dos brasileiros. Isso quer dizer que na última década houve uma queda no poder de compra do brasileiro em relação aos combustíveis estudados no nosso Projeto Integrador, que foram a Gasolina, Etanol e Diesel.

Tendo em mente que o preço dos combustíveis aumenta em uma taxa muito mais rápida que o salário médio dos brasileiros, resolvemos *calcular qual deveria ser o salário médio do Brasil para o ano de 2021 para equiparar o aumento do preço dos combustíveis na última década.*

Parte 6 - Calculando a Renda Média do Brasil em 2021 para equiparar o aumento do preço médio dos combustíveis entre 2012 e 2021

Analisaremos qual deveria ser o salário médio brasileiro em 2021 para manter o poder de compra que as pessoas tinham em 2012, já que verificamos que houve uma perda no poder de compra dos brasileiros em relação aos combustíveis citados nesse projeto nos últimos 10 anos.

Calculando a média de preço de todos os combustíveis em 2012 e 2021

```
pm_comb <- dados %>%
  filter(Ano == 2012 | Ano == 2021) %>%
  group_by(Ano) %>%
  summarise(Media_g = mean(Gasolina_preco_medio),
            Media_d = mean(Diesel_preco_medio),
            Media_e = mean(Etanol_preco_medio),
            Media_preco = (Media_g + Media_d + Media_e) / 3)

pm_comb %>%
  select(1, 5) %>%
  knitr::kable(caption = "Preço médio de todos os combustíveis para
                      o ano de 2012 e 2021")
```

Table 4: Preço médio de todos os combustíveis para o ano de 2012 e 2021

Ano	Media_preco
2012	2.381372
2021	5.117653

Agora vamos calcular a taxa de crescimento do preço de todos os combustíveis entre 2012 e 2021

```
pm_taxa <- round(((pm_comb[2, 5] - pm_comb[1, 5]) / pm_comb[1, 5]) * 100, 2)

cat(paste0("Taxa de crescimento do preço dos combustíveis no Brasil entre 2012 e 2021: ", pm_taxa, "%"))

## Taxa de crescimento do preço dos combustíveis no Brasil entre 2012 e 2021: 114.9%
```

Vemos que a taxa de crescimento no preço dos combustíveis no Brasil foi de incríveis 114,9% na última década.

Calculando a média de renda de todos os estados brasileiros em 2012 e 2021

```
rm_estados <- renda %>%
  filter(Ano == 2012 | Ano == 2021) %>%
  group_by(Ano) %>%
  summarise(renda_media = mean(Renda))

rm_estados %>%
  knitr::kable(caption = "Renda média de todos os estados para o ano
                      de 2012 e 2021")
```

Table 5: Renda média de todos os estados para o ano de 2012 e 2021

Ano	renda_media
2012	1295.407
2021	2023.296

Agora vamos calcular a taxa de crescimento da renda média dos brasileiros na última década

```
rm_taxa <- round(((rm_estados[2, 2] - rm_estados[1, 2]) / rm_estados[1, 2]) * 100, 2)
cat(paste0("Taxa de crescimento da renda média no Brasil entre 2012 e 2021: ", rm_taxa, "%"))
```

```
## Taxa de crescimento da renda média no Brasil entre 2012 e 2021: 56.19%
```

Vemos que a taxa de crescimento na renda média do brasileiro foi de apenas 56,19% na última década. **Ou seja, cresceu em uma taxa 2x mais lenta que o preço dos combustíveis. Para ser mais preciso, o preço dos combustíveis cresceu 58,71% a mais em 10 anos.**

Tendo essas informações, agora é possível calcular qual deveria ser a renda média do brasileiro para o ano de 2021 para equiparar o aumento que os combustíveis tiveram na última década.

```
rm_estados[3, 1] <- 2021
rm_estados[3, 2] <- (rm_estados[2, 2] + (rm_estados[2, 2] * 0.5871))
rm_estados %>% knitr::kable(caption = "Salário médio brasileiro em 2021
ajustado pelo aumento do preço dos combustíveis")
```

Table 6: Salário médio brasileiro em 2021 ajustado pelo aumento do preço dos combustíveis

Ano	renda_media
2012	1295.407
2021	2023.296
2021	3211.174

Como visto acima, a renda média do Brasileiro deveria ser **R\$ 3211,174** ao invés de **R\$ 2023,296** em 2021 para manter o poder de compra no qual as pessoas tinham no ano de 2012. Esse achado é outra confirmação da queda do poder de compra do Brasileiro em relação ao preço dos combustíveis (Diesel, Gasolina e Etanol) nos últimos dez anos.