```r
library(readr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(knitr)
library(ggplot2)
library(magrittr)
library(ggdendro)
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```r
library(tidyr)
```

```
##
## Attaching package: 'tidyr'
```

```
## The following object is masked from 'package:magrittr':
##
##     extract
```

```r
data <- read_csv("Country-data.csv")
```

```
## Rows: 167 Columns: 10
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): country
## dbl (9): child_mort, exports, health, imports, income, inflation, life_expec...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
dic <- read_csv("data-dictionary.csv")
```

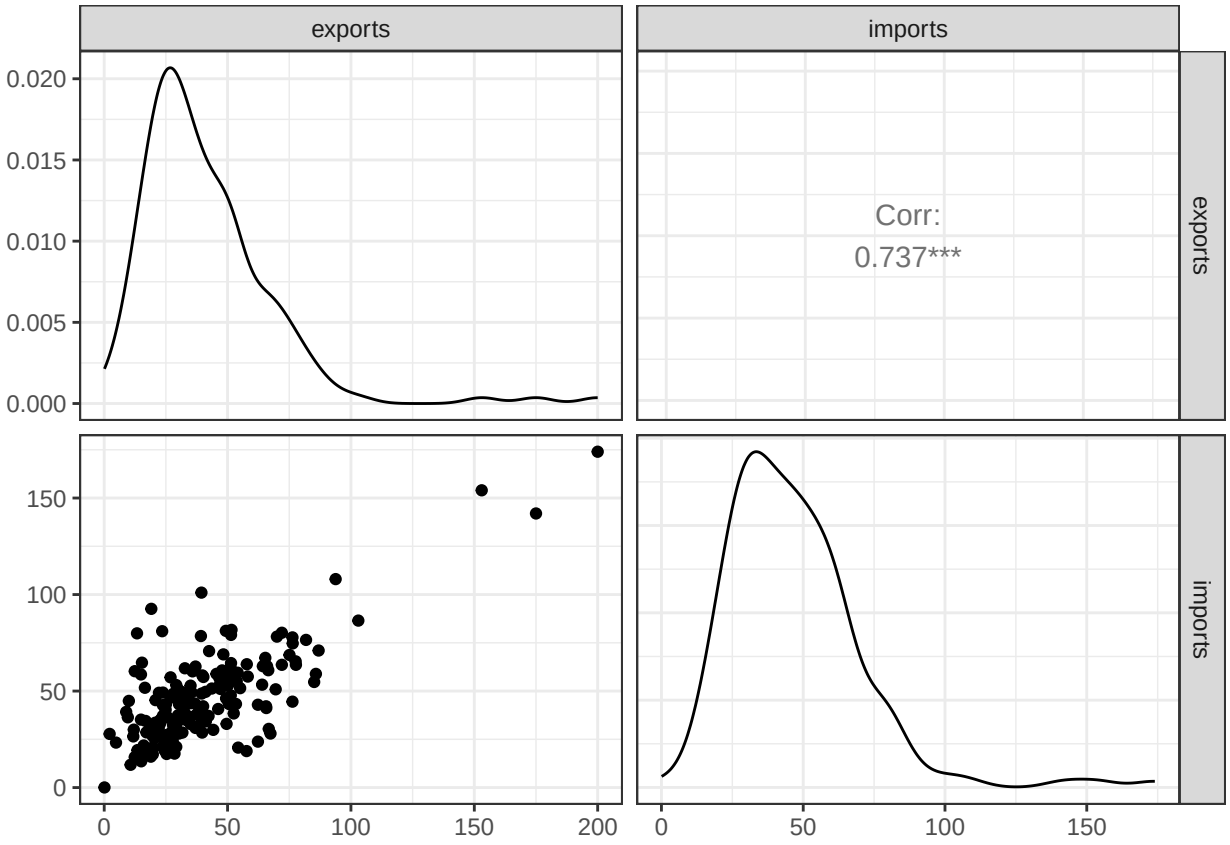
```
## Rows: 10 Columns: 2
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): Column Name, Description
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
kable(dic)
```

Column Name	Description
country	Name of the country
child_mort	Death of children under 5 years of age per 1000 live births
exports	Exports of goods and services per capita. Given as %age of the GDP per capita
health	Total health spending per capita. Given as %age of GDP per capita
imports	Imports of goods and services per capita. Given as %age of the GDP per capita
Income	Net income per person
Inflation	The measurement of the annual growth rate of the Total GDP
life_expec	The average number of years a new born child would live if the current mortality patterns are to remain the same
total_fer	The number of children that would be born to each woman if the current age-fertility rates remain the same.
gdpp	The GDP per capita. Calculated as the Total GDP divided by the total population.

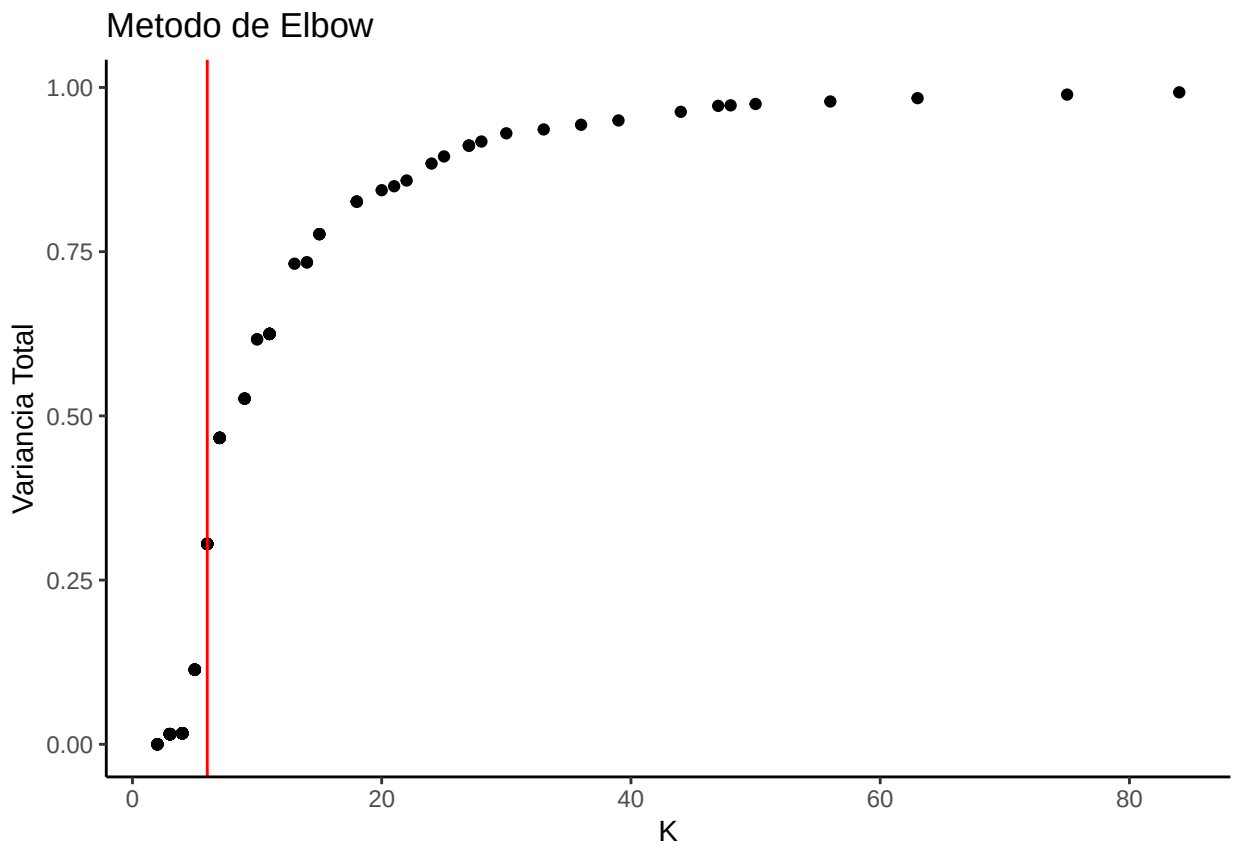
Dividir os países de acordo com o quanto eles conversam com outros países: importação e exportação

```
ggpairs(data %>% select(exports, imports), progress = F) +
  theme_bw()
```

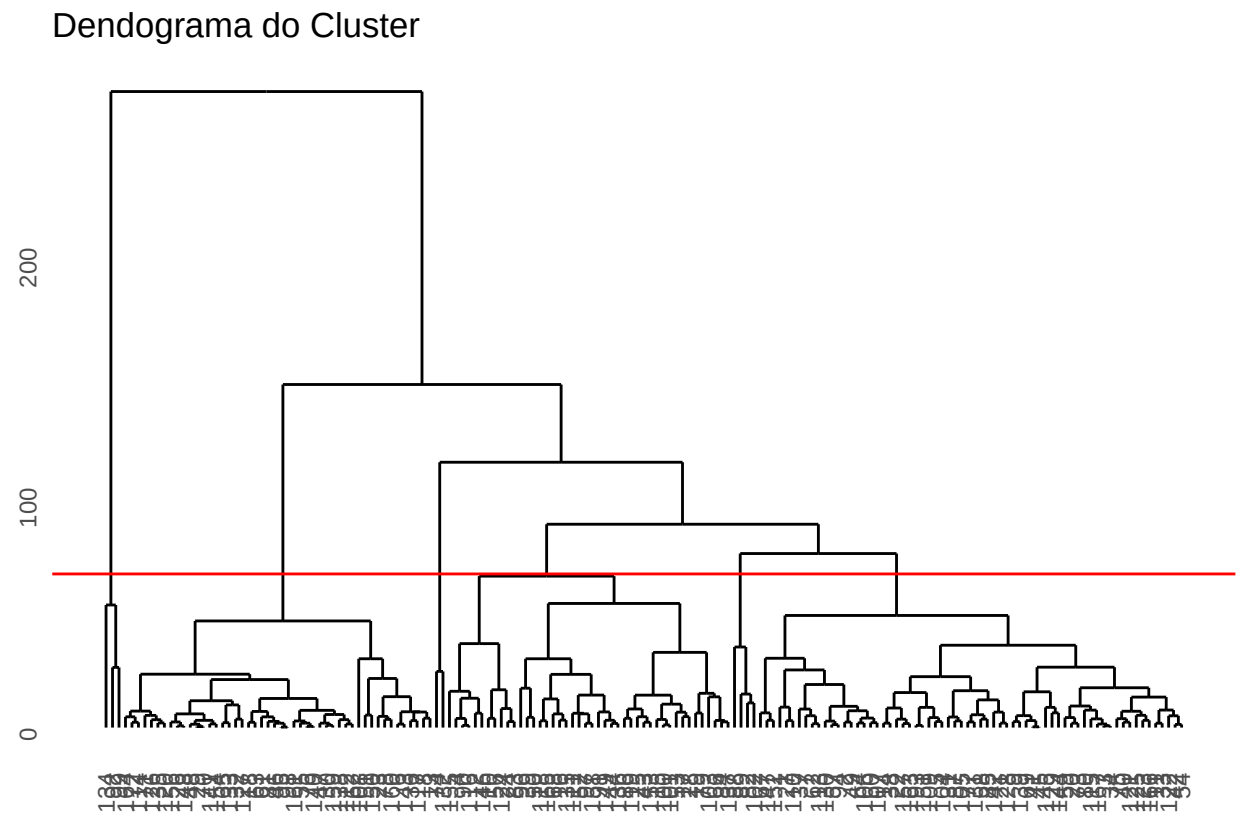


```r
IE <- data %>% select(exports, imports)
IE <- unique(IE) # transformando numa lista
cluster <- hclust(dist(IE), method = "complete")
```

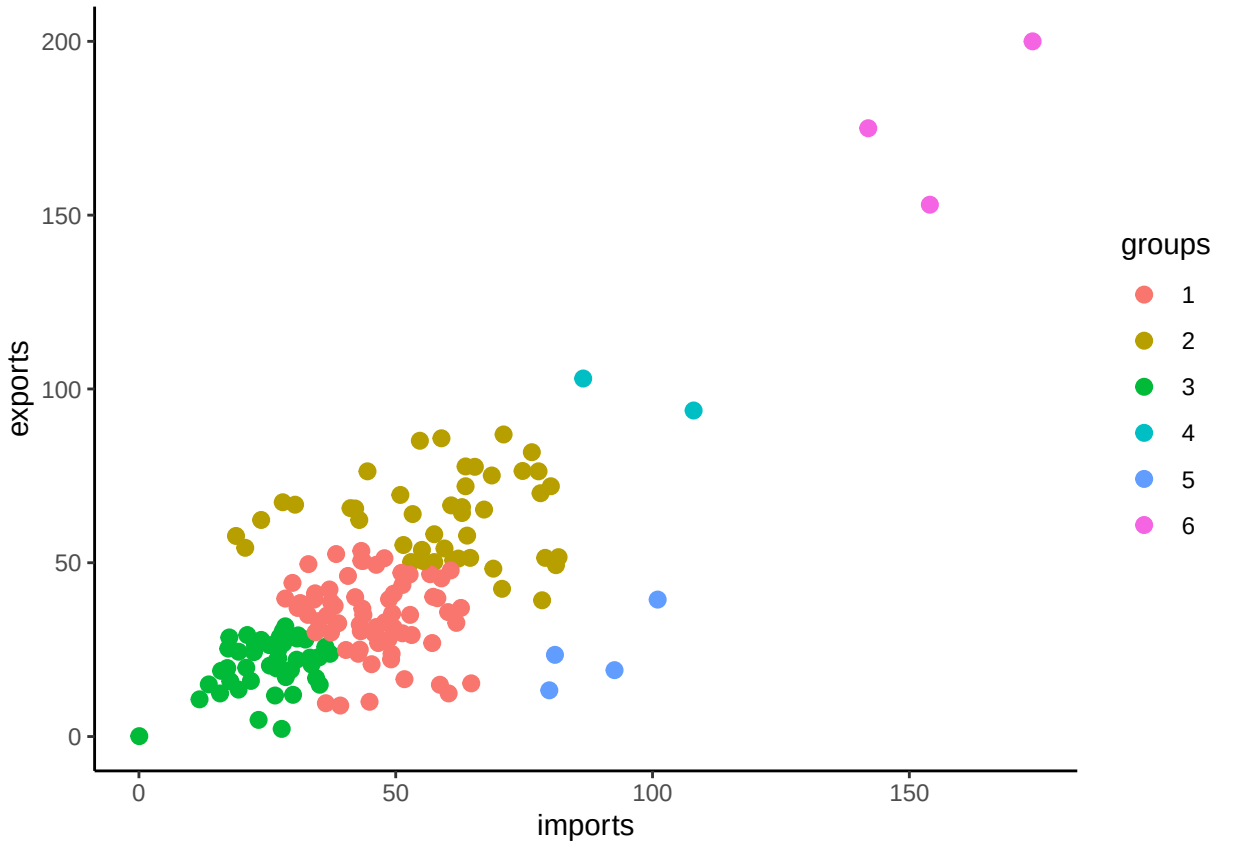
```r
h <- seq(5, 150, by = 1) # definindo o espa
N <- nrow(IE)


totalVar <- rep(0, length(h))
K <- numeric(length(h))
Y <- with(IE, cbind(imports, exports))

for(i in seq_along(h)){
  groups <- factor(cutree(cluster, h = h[i]))
  K[i] <- length(levels(groups))
  s.manova <- summary(manova(Y ~ groups), tol = 0)

  B <- s.manova$SS$groups
  W <- s.manova$SS$Residuals

  totalVar[i] <- det(B)/(det(B+W))
}

data.frame(totalVar, K) %>%
  ggplot() +
  geom_point(aes(x = K, y = totalVar)) +
  geom_vline(xintercept = 6, col = "red") +
  theme_classic() +
  labs(title = "Metodo de Elbow",
       y = "Variancia Total")
```




```
ggdendrogram(cluster) +
  geom_hline(yintercept = h[which.max(K==6)], col = "red") +
  labs(title = "Dendograma do Cluster")
```




```
groups <- cutree(cluster, h = h[which.max(K==6)])
data.frame(IE, groups) %>%
  ggplot(aes(x = imports, y = exports, col = as.factor(groups))) +
  geom_point(size = 2.5) + theme_classic() +
  labs(col = "groups")
```



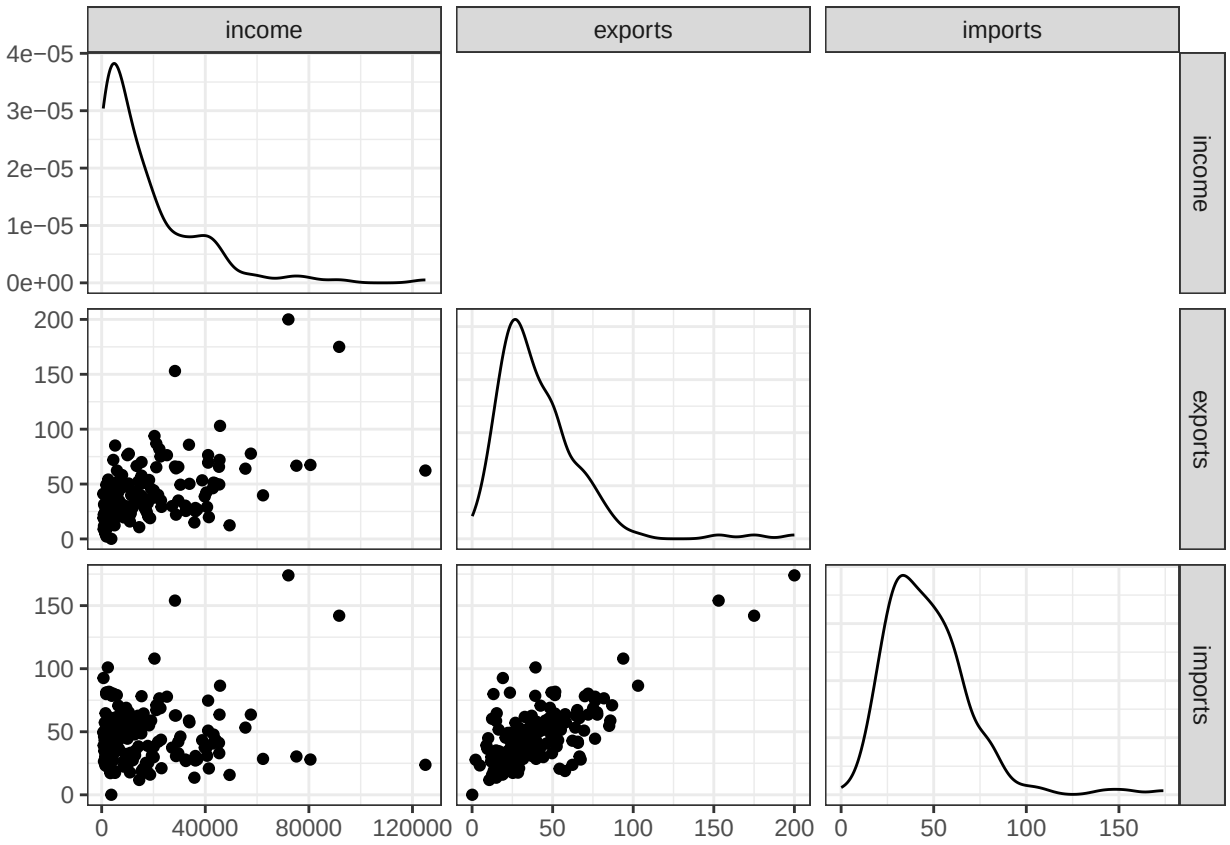
Brasil ta no terceiro grupo, junto com Argentin, Australia, Canada, China, Colombia, India, França (Todos os países do BRICS), Estados Unidos, Venezuela.

```
data.frame(data, groups) %>%
  filter(groups == 4)
```

```
##      country child_mort exports health imports income inflation life_expec
## 1    Ireland        4.2   103.0   9.19    86.5  45700     -3.22       80.4
## 2 Seychelles       14.4    93.8   3.40   108.0  20400     -4.21       73.4
##   total_fer  gdpp groups
## 1      2.05 48700      4
## 2      2.17 10800      4
```

Tentativa de modela (nao deu certo por causa da normalidade e independencia dos residuos)

```
ggpairs(data %>% select(income, exports, imports), upper = "blank") +
  theme_bw()
```




```
fit1 <- lm(income ~ exports + imports, data = (data.frame(data, groups) %>% filter(groups == 3)))
summary(fit1)
```

```
##
## Call:
## lm(formula = income ~ exports + imports, data = (data.frame(data,
##     groups) %>% filter(groups == 3)))
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -17439  -7898  -3480   2132  36412
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  13238.0     7247.7   1.827  0.07441 .
## exports        826.7      275.5   3.001  0.00438 **
## imports       -664.6      269.4  -2.467  0.01748 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12530 on 45 degrees of freedom
## Multiple R-squared:  0.1966, Adjusted R-squared:  0.1609
## F-statistic: 5.506 on 2 and 45 DF,  p-value: 0.007258
```

```r
anova(fit1)
```

```
## Analysis of Variance Table
##
## Response: income
##           Df     Sum Sq   Mean Sq F value  Pr(>F)
## exports    1  772822234 772822234  4.9251 0.03155 *
## imports    1  955224763 955224763  6.0875 0.01748 *
## Residuals 45 7061168765 156914861
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```