



Passo-a-passo

ETAPA 7. GRÁFICOS DE DISPERSÃO E DENSIDADE

Prof. Pedro Feliú

INTRODUÇÃO

Nesta etapa executaremos comandos referentes à criação de gráficos de dispersão e um gráfico de densidade. Iniciaremos com o gráfico de dispersão por meio da função **plot**. Utilizaremos novamente o banco de dados “mundo”. Importe ao Rstudio o banco de dados em Excel e iniciamos os procedimentos.

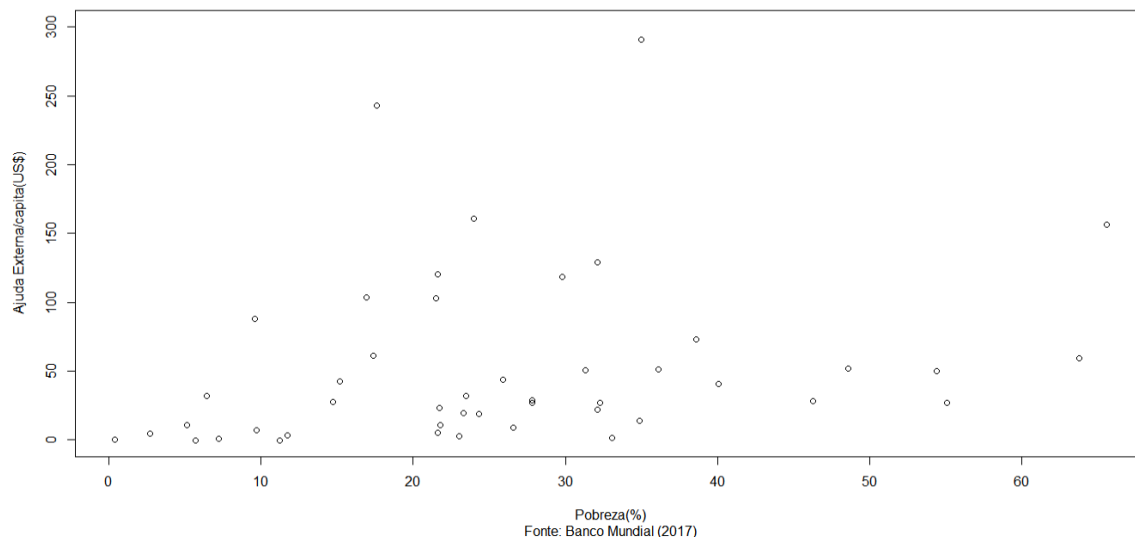
PASSO 1: Gráfico de Dispersão e Reta de Regressão

Digite os seguintes comandos:

```
attach(mundo)
```

```
plot(pobreza, `ajuda externa_capita`, main="Gráfico 10. Dispersão entre Pobreza e  
Recepção de Ajuda Externa no Mundo", xlab="Pobreza(%)", ylab="Ajuda  
Externa/capita(US$)", sub="Fonte: Banco Mundial (2017)", ylim=c(0,300))
```

Gráfico 10. Dispersão entre Pobreza(%) e Ajuda Externa per capita no Mundo

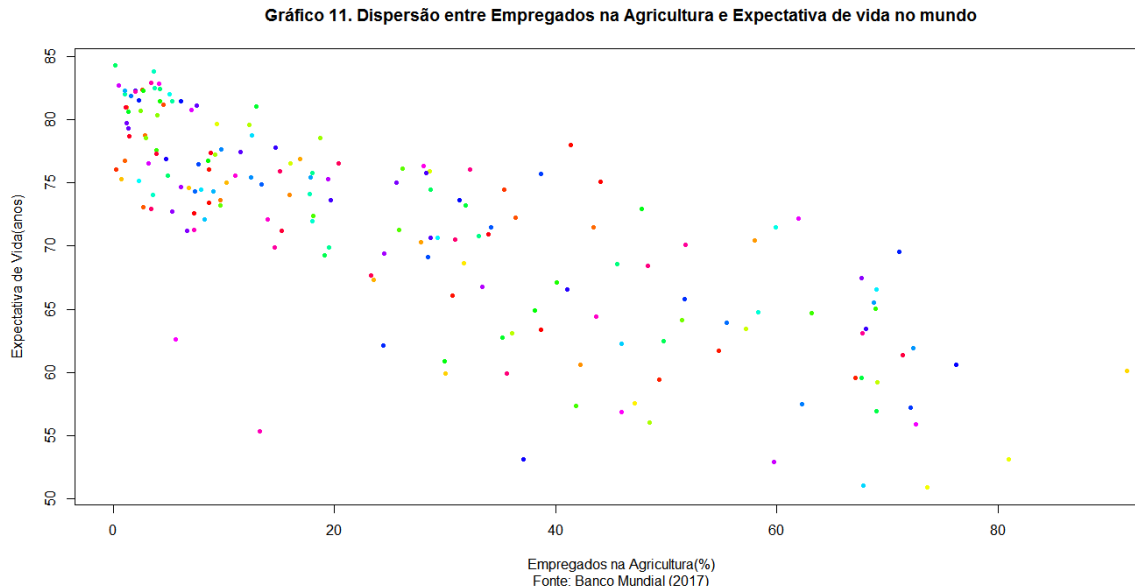


No gráfico 10 acima, cada ponto representa um país do mundo, em dois eixos (bidimensional) representados pela porcentagem de população pobre (horizontal) e a ajuda externa per capita recebida pelos países (vertical). Os gráficos de dispersão (ou diagramas de dispersão) auxiliam na observação do padrão de relacionamento entre duas variáveis quantitativas contínuas (que podem ser discretas tb). No gráfico 10 acima percebemos a ausência de um padrão claro entre pobreza e o recebimento de ajuda externa, aparentando não haver relação entre as mesmas. Há países com elevada porcentagem de população na pobreza, quase 50%, e baixo recebimento de ajuda externa (abaixo de 50 dólares per capita ao ano), assim como o seu oposto, países com baixa pobreza (10%) e elevado recebimento de ajuda externa (quase 100 dólares per capita). O dado permite concluir que não necessariamente os países mais pobres são os que recebem mais ajuda externa, rejeitando a hipótese do altruísmo ou erradicação da pobreza enquanto motivador principal para um país enviar ajuda a outro. Como aponta a literatura de relações internacionais, a ajuda externa está fortemente relacionada à compromissos políticos e interesses estratégicos, corroborando o gráfico 10.

Estatística I

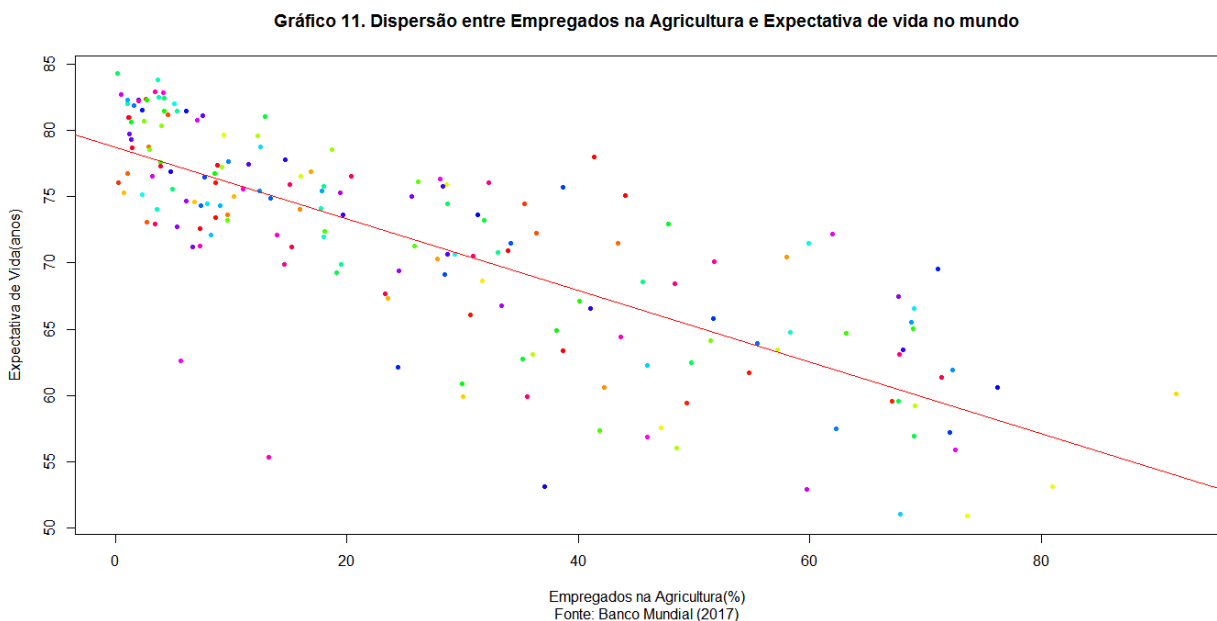
Façamos agora um gráfico com duas variáveis que se relacionam: porcentagem da população empregada na agricultura e a expectativa de vida.

```
plot(empregados_agricultura, expectativa_vida, main="Gráfico 11. Dispersão entre Empregados na Agricultura e Expectativa de vida no mundo", xlab="Empregados na Agricultura(%)", ylab="Expectativa de Vida (anos)", sub="Fonte: Banco Mundial (2017)", col=rainbow(190), pch=20)
```



No gráfico 11, ao contrário do gráfico 10, percebemos a presença de um padrão na relação entre as variáveis: quanto maior a proporção de pessoas empregadas na agricultura, menor a expectativa de vida, revelando uma relação negativa entre as mesmas. Note que preenchemos os pontos com a função **pch**. Vamos incluir agora uma reta de regressão neste gráfico para facilitar a interpretação do mesmo:

```
abline(lm(expectativa_vida ~ empregados_agricultura), col="red")
```



Estatística I

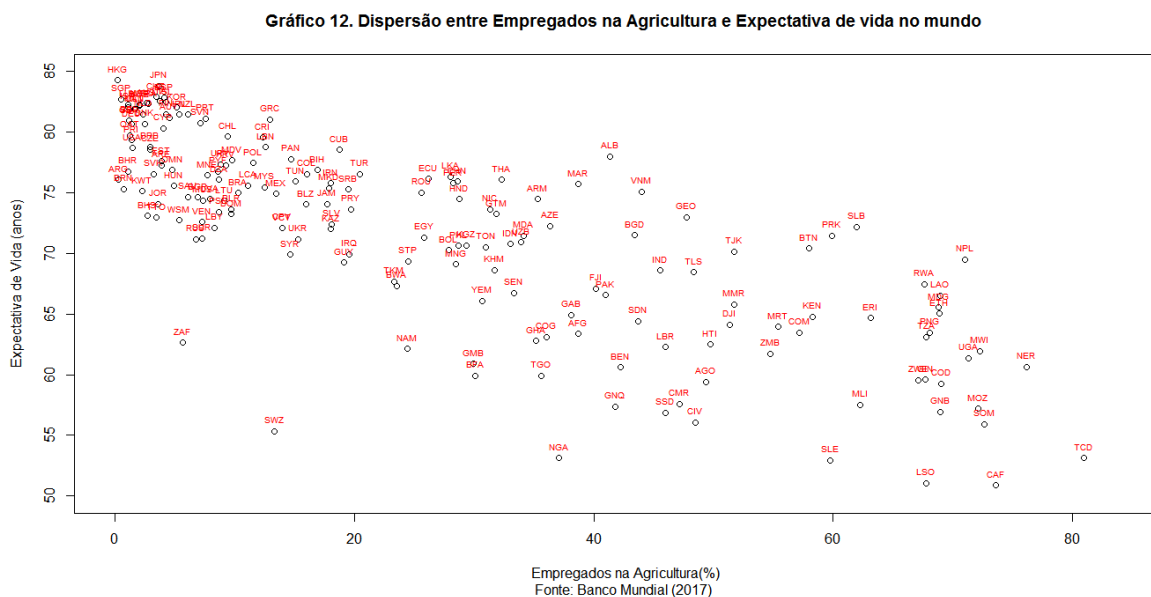
Percebemos que a reta indica a relação negativa entre as variáveis (mais empregadas na agricultura significa menos expectativa de vida). Quanto mais próximos os pontos estiverem da reta, mais forte a relação entre as variáveis. Muito provavelmente o trabalho no campo é mais exigente, perigoso e está associado à menor oferta de bens públicos de saúde, podendo impactar na expectativa de vida da população do campo.

Passo 2: Rótulos e Legendas

Neste passo vamos apenas inserir no gráfico de dispersão os rótulos dos dados e depois uma legenda. Iniciamos pelo rótulo, incluindo um comando para editar os eixos:

```
plot(empregados_agricultura, expectativa_vida, main="Gráfico 12. Dispersão entre Empregados na Agricultura e Expectativa de vida no mundo", xlab="Empregados na Agricultura(%)", ylab="Expectativa de Vida (anos)", sub="Fonte: Banco Mundial (2017)", xlim=c(0,85), ylim=c(50,85))
```

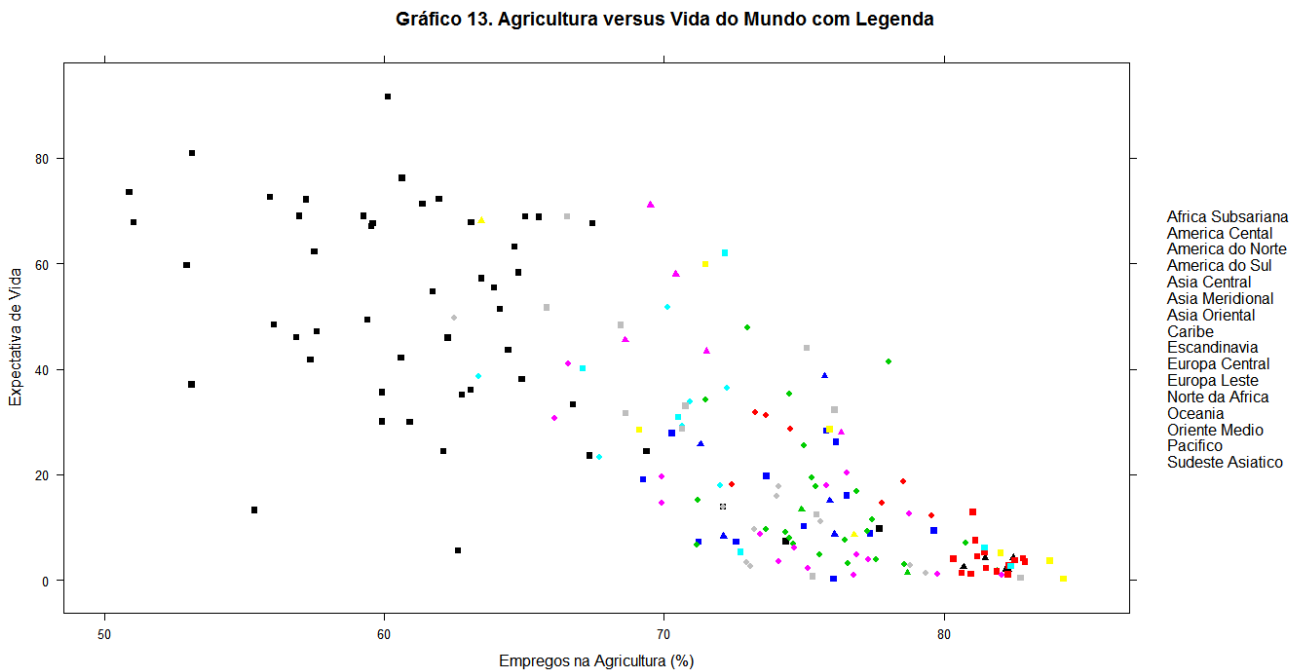
```
text(empregados_agricultura, expectativa_vida, labels=id, cex=0.6, pos=3, col="red")
```



No primeiro comando acima incluímos uma novidade as funções **xlim** e **ylim**, cuja serventia é delimitar a escala dos eixos horizontal e vertical, respectivamente. No exemplo acima, utilizamos no eixo horizontal uma escala que vai de 0 a 80, enquanto no eixo vertical de 50 a 85. O comando **text**, em seguida, inclui texto no gráfico, neste caso os rótulos de cada ponto (a sigla dos países com a função **labels=id**). O comando **cex** determina o tamanho do texto e **pos** a sua posição no gráfico, (1- abaixo; 2-à esquerda; 3-acima e 4-à direita). Passemos agora para o comando referente à criação de uma legenda **xyplot**. Para utilizar essa função, será necessário instalar e carregar o pacote **lattice**. Segue o comando:

Estatística I

```
require(lattice)
library(lattice)
xyplot(empregados_agricultura ~ expectativa_vida, groups=micro_regiao,
data=mundo, xlab="Empregos na Agricultura (%)", ylab="Expectativa de Vida",
main="Gráfico 13. Agricultura versus Vida do Mundo com Legenda", jitter.x=F,
auto.key=list(space="right", columns=1),
par.settings=list(superpose.symbol=list(pch=15:17, col=1:16, cex=0.8)))
```



No comando acima, há elementos novos a serem compreendidos:

groups – estabelece o subgrupo de plotagem, nesse caso as regiões do mundo

data – indica ao R qual o objeto tratado

jitter – função utilizada para visualização melhor de diagrama com muitos pontos sobrepostos, no nosso caso não foi necessário, mas é uma função muito útil.

auto.key=list(space="right", columns=1) – posiciona a legenda à direita do gráfico com apenas uma coluna

par.settings=list(superpose.symbol=list – determina a forma da plotagem (quadrado, ponto e triângulo, sempre codificados por números no R – tentem modificar esses números para testar o comando).

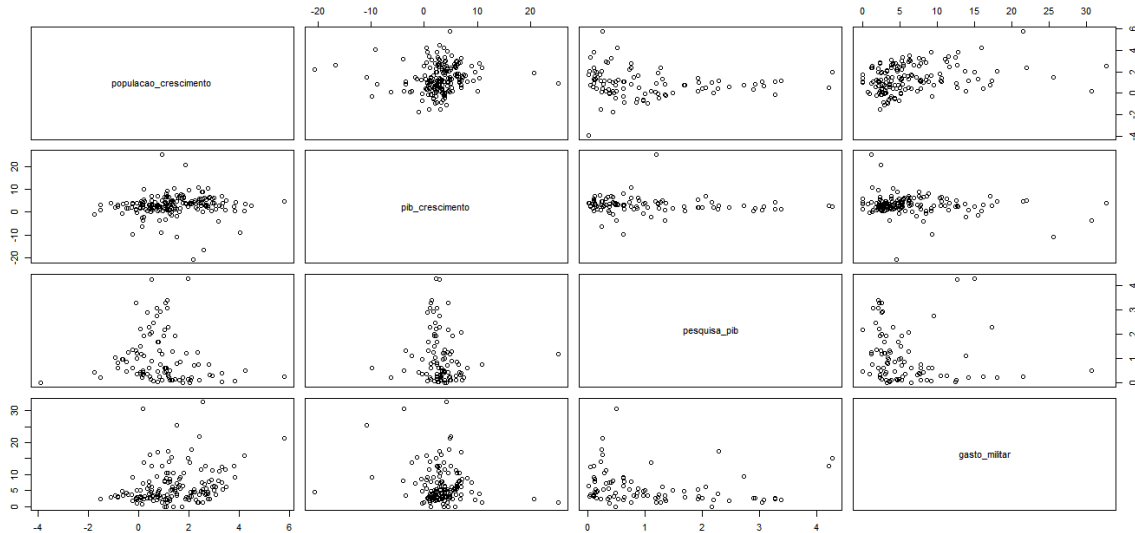
PASSO 3: Matriz de Dispersão

Vamos realizar nesse passo uma matriz de dispersão por meio da função **pairs**, representação que permite visualizar a relação entre mais de duas variáveis no mesmo gráfico. Segue o comando:

```
pairs(~populacao_crescimento+pib_crescimento+pesquisa_pib+gasto_militar,
main="Gráfico 14. Matriz da Dispersão entre, Crescimento Populacional,
Crescimento Econômico, Gastos Militares e Gastos em Pesquisa")
```

Estatística I

Gráfico 14. Matriz da Dispersão entre, Crescimento Populacional, Crescimento Econômico, Gastos Militares e Gastos em Pesquisa

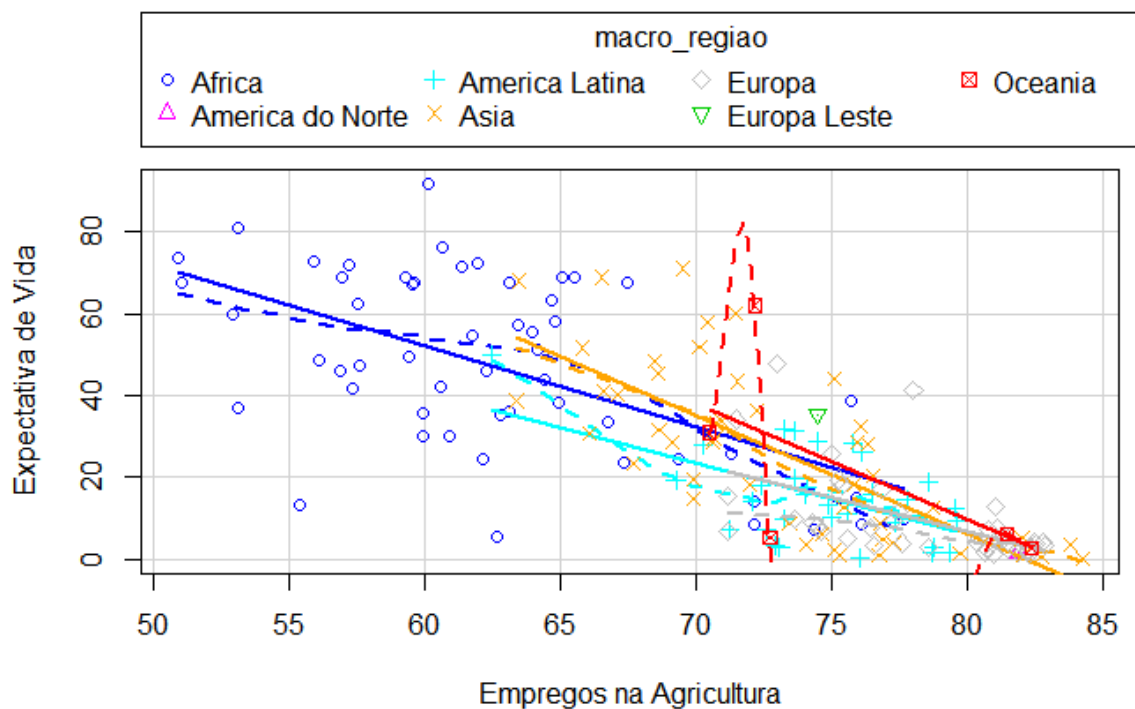


A interpretação da matriz acima é igual à análise individual. Vamos agora utilizar o pacote **car** para realizar uma matriz de dispersão com a função **scatterplotMatrix**.

```
install.packages("car")
```

```
library(car)
```

```
scatterplot(empregados_agricultura ~ expectativa_vida | macro_regiao,  
data=mundo, xlab="Empregos na Agricultura", ylab="Expectativa de Vida",  
main="Gráfico 16")
```



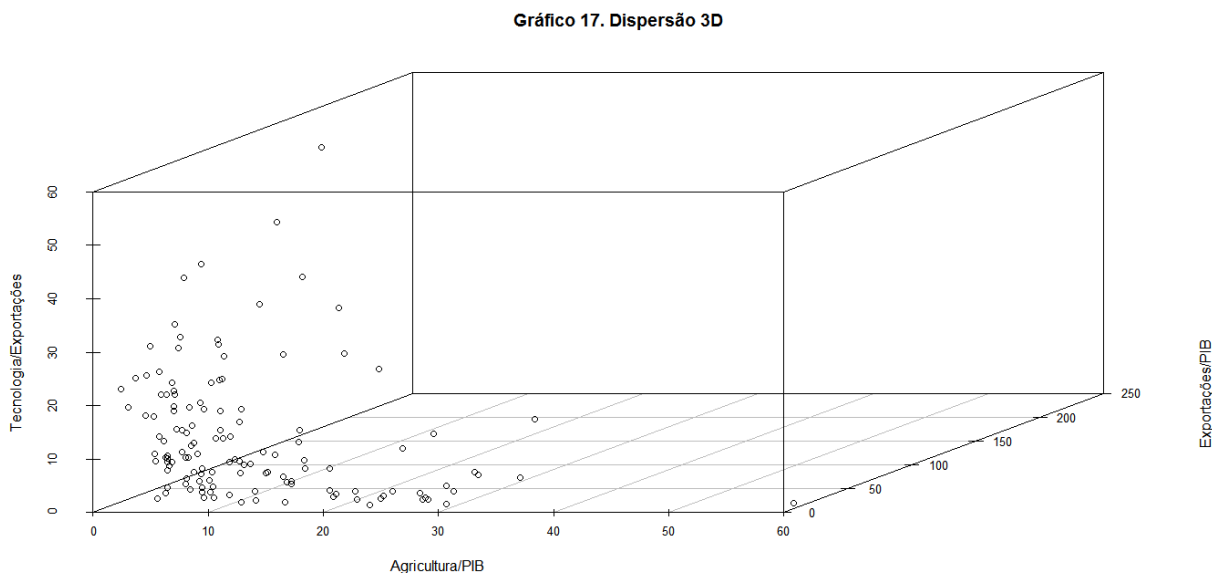
Antes de passar para o gráfico de densidade, vamos realizar mais um gráfico de dispersão, mas desta vez com três variáveis (3D).

Departamento de Relações Internacionais
Universidade Federal da Paraíba

Estadística I

Instale e carregue o pacote **scatterplot3d**. Vamos ao comando:

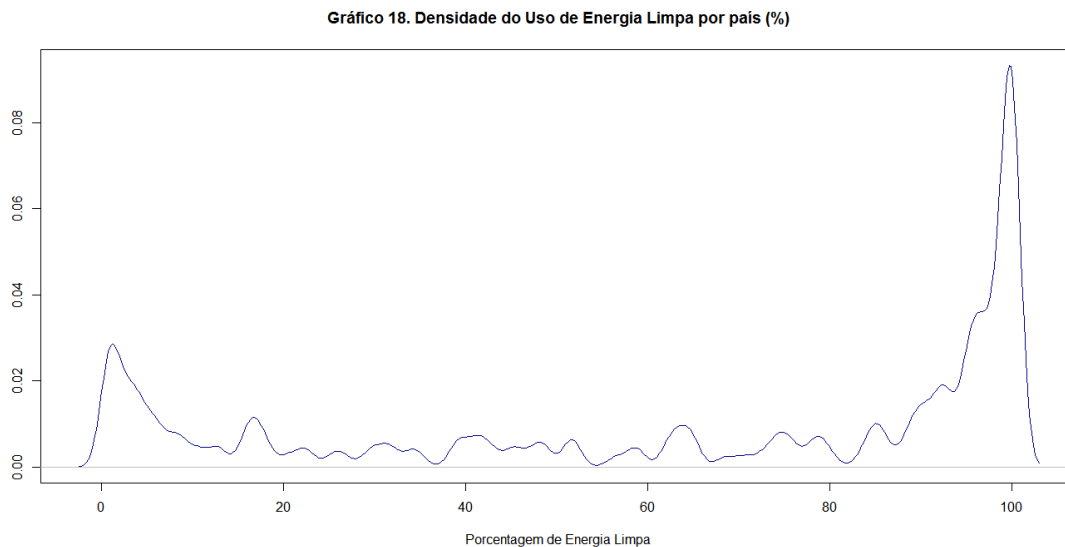
```
install.packages("scatterplot3d")  
library(scatterplot3d)  
scatterplot3d(`agricultura_%_pib`, `exportacoes_%_pib`, tecnologia_exportacoes,  
main="Gráfico 17. Dispersão 3D", xlab="Agricultura/PIB",  
ylab="Exportações/PIB", zlab="Tecnologia/Exportações")
```



PASSO 4: Gráfico de Densidade

Neste passo construiremos um gráfico de densidade. Digite o seguinte comando:

```
plot(density(`energia_limpa_%pop`, bw=1, na.rm=TRUE), col="darkblue",  
xlab="Porcentagem de Energia Limpa", main="Gráfico 18. Densidade do Uso de  
Energia Limpa por país (%)", ylab="Densidade de Kernell")
```



Estatística I

No gráfico 18 acima percebemos uma elevada densidade nas porcentagens 0 e 100, especialmente 100, indicando uma maior incidência de países nesta proporção de uso de energia limpa. Há uma distinção clara, ou seja, ou o país usa energia limpa ou não usa, já que as porcentagens intermediárias no eixo horizontal apresentam densidades pequenas. Modifique os valores da função **bw**, que determina a escala do eixo horizontal, para ver o comportamento do gráfico.