

Universidade Federal de São Carlos  
Departamento de Economia  
Análise e visualização de dados do coronavírus usando R  
Profa. Dra. Andreza A. Palma

**Homework 2 (ggplot) - Entrega: 15/10/2020**

**Instruções Gerais**

- Você deve criar um novo documento RMarkdown e responder as questões usando este documento no R.
- Você deverá entregar no AVA/Moodle o seu arquivo .RMD e o seu arquivo .HTML
- Para corrigir, eu irei analisar o arquivo .RMD, gerando o HTML. Portanto, teste antes de enviar se tudo está funcionando adequadamente.
- Não receberei as atividades via e-mail. Apenas via AVA/Moodle.
- Note que para cada exercício que exija um código no R, vc deve utilizar um *chunk*.

**Questões**

- 1) Utilizando o conjunto de dados *USArrests* descrito na aula, construa um histograma para cada uma das variáveis contida nesse banco de dados. Use uma cor diferente para cada histograma. Construa também *box-plots* para cada uma das variáveis, com as mesmas cores usadas no histograma.
- 2) Escolha duas variáveis do banco de dados para fazer um diagrama de dispersão (*scatterplot*), usando `geom_point`. Acrescente uma linha relacionando as duas variáveis para ajudar na interpretação. Descreva o fenômeno que você quer analisar com esse diagrama.
- 3) Utilizando a função `data()` do R, escolha um conjunto de dados de sua preferência e crie 4 gráficos, a sua escolha, usando as variáveis deste banco de dados. Faça uma breve interpretação de seus achados.
- 4) Considerando o banco de dados utilizado no exercício anterior, escolha duas variáveis para analisar via diagrama de dispersão. Comente brevemente seus achados.
- 5) Utilize o comando `?mpg` e descreva as variáveis contidas neste banco de dados. Faça uma análise preliminar deste banco de dados usando os comando *head*, *tail*, etc (como fizemos para *USArrests*).
- 6) Faça um gráfico de dispersão entre as variáveis *hwy* e *cyl*. Mude a cor do tema. Interprete os resultados.

7) Escolha duas variáveis do banco de dados e construa histograma e box-plot para elas. Personalize os gráficos usando cores diferentes do *default*. Não se esqueça de dar nomes para os eixos, bem como um título para os gráficos. Organize os gráficos em subplots usando `grid.arrange()`. Note que teremos 2 histogramas e 2 boxplots, totalizando 4 gráficos que vc deve plotar em duas colunas.

8) Usando o arquivo de dados mpg ainda, vamos agora treinar o uso do recurso *facet*.

1. Faça gráficos de dispersão entre *displ* e *hwy* separados para cada classe de veículos. Use `color = class` para diferenciar por cor e `facet_wrap( ~ class, nrow=2)` para obter os subplots.
2. Agora vamos criar subplots com base em duas variáveis: número de cilindros (*cyl*) e tipo de direção (*drv*). Diferencie por cor usando `color=drv` e utilize `facet_grid(drv ~ cyl)`. Algumas facetas ficarão vazias por não possuírem combinações de *drv* e *cyl*.
3. Explique o que ocorre se você usar `color=cyl` para classificar por cilindros (*cyl*) e `facet_grid(. ~ cyl)`

9) **Corrupção x desenvolvimento humano** Para realizar este exercício, usaremos um banco de dados disponibilizado na página do nosso curso (*homework3.xls*). Para carregar este conjunto de dados no R, execute os seguintes comandos:

```
library(readxl)
```

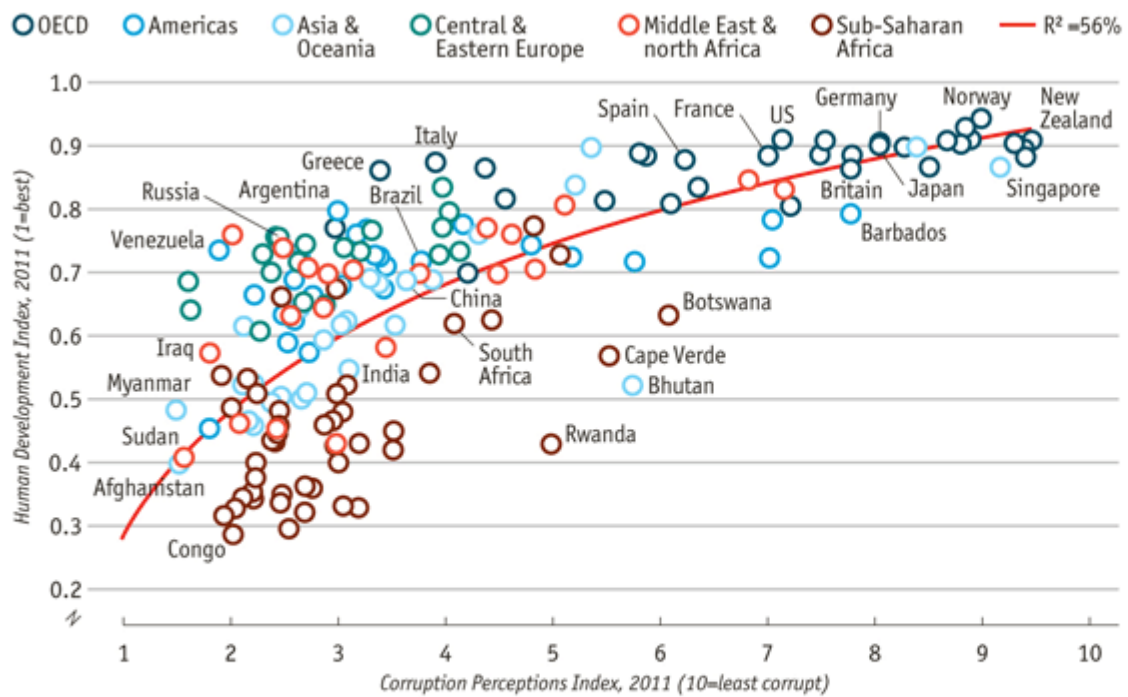
```
homework2 = read_excel("homework2.xlsx")
```

Lembre-se de salvar o arquivo na pasta usada como diretório do R.

Os dados dizem respeito basicamente ao índice de desenvolvimento humano (HDI) e percepção da corrupção (CPI)

- a Construa um diagrama de dispersão entre as variáveis CPI (eixo x) e HDI (eixo y). Coloque a cor vermelha nos pontos.
- b Agora diferencie a cor dos pontos por região.
- c Faça um gráfico com pontos maiores (`size =3`).
- d Faça um gráfico em que o tamanho dos pontos dependa da variável HDI.
- e Inclua nomes nos eixos x e y, um título para o seu gráfico bem como legendas que julgar necessárias.
- f **Desafio** O gráfico abaixo foi publicado na revista *The Economist* e utiliza os dados contidos no arquivo que estamos usando. Escreva um código no R usando `ggplot` que replica tão próximo quanto possível o gráfico em questão.

## Corruption and human development



Sources: Transparency International; UN Human Development Report