

Universidade Federal de São Carlos
Departamento de Economia
ACIEPE: Estatística para Data Science
Profa. Dra. Andreza A. Palma

Homework 3 - Entrega: 18/06/2020

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
- 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*. A organização do seu arquivo também será avaliada.

Questões

- 1) Execute `ggplot(data=mpg)`. Explique o que este comando faz.
- 2) 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*).
- 3) Faça um gráfico de dispersão entre as variáveis *hwy* e *cyl*. Mude a cor do tema. Interprete os resultados.
- 4) Considerando o banco de dados utilizado no exercício anterior, escolha duas variáveis para analisar via diagrama de dispersão. Comente brevemente seus achados.
- 5) 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.
- 6) Usando o arquivo de dados *mpg* ainda, vamos agora treinar o uso do recurso *facet*.
 - a) 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.

- b) 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.
- c) Explique o que ocorre se você usar color=cyl para classificar por cilindros (cyl) e facet_grid(. ~ cyl)

7) 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)
```

```
homework3 = read_excel("homework3.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) O arquivo Economist1.pdf disponível junto com a lista de exercícios apresenta um gráfico usando 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.