

Conjunto de Problemas — Análise Exploratória - Soluções

James R. Hunter

24 de fevereiro de 2017

Economia de Combustível nos Carros nos EUA entre 1999 e 2008

Os dados ficam num conjunto de dados chamada `mpg` que faz parte de pacote `ggplot2`.

Carregar os dados usando o comando `data(mpg)`.

1. Crie um novo conjunto chamado `suvcompact` que só contém as classes `suv` e `subcompact`. Quantos elementos tem `suvcompact`?

```
suvcompact <- mpg %>%  
  filter(class == "suv" | class == "subcompact")  
elementos = nrow(suvcompact)  
paste("Elementos =", elementos)  
  
## [1] "Elementos = 97"
```

2. Qual é a economia de combustível nas estradas (`hwy`) para todos os carros combinados. Mostre a mediana e média. Também o IQR e o desvio padrão

```
## solução mais simples  
paste("Média =", mean(mpg$hwy), "Mediana =", median(mpg$hwy),  
      "IQR =", IQR(mpg$hwy), "Desvio Padrão =", sd(mpg$hwy))  
  
## [1] "Média = 23.4401709401709 Mediana = 24 IQR = 9 Desvio Padrão = 5.95464344116645"
```

3. As variáveis “`hwy`” e “`cty`” são expressos em “miles per gallon”. Converter esses variáveis em “km por litro”. Em média, quantos quilômetros por litro consegue um Volkswagen Passat na cidade (`cty`)?

- 1 galão = 3.78541 litros
- 1 milha = 1.60934 quilômetros

```
kmmilha <- 1.60934  
litgal <- 3.78541  
convertrate <- kmmilha/litgal  
kmpl <- mpg %>%  
  mutate(cty = cty * convertrate) %>%  
  mutate(hwy = hwy * convertrate)  
meanPassat <- kmpl %>%  
  filter(model == "passat")  
avgPassat <- mean(meanPassat$cty)  
paste("Passat cidade kmpl =", avgPassat)
```

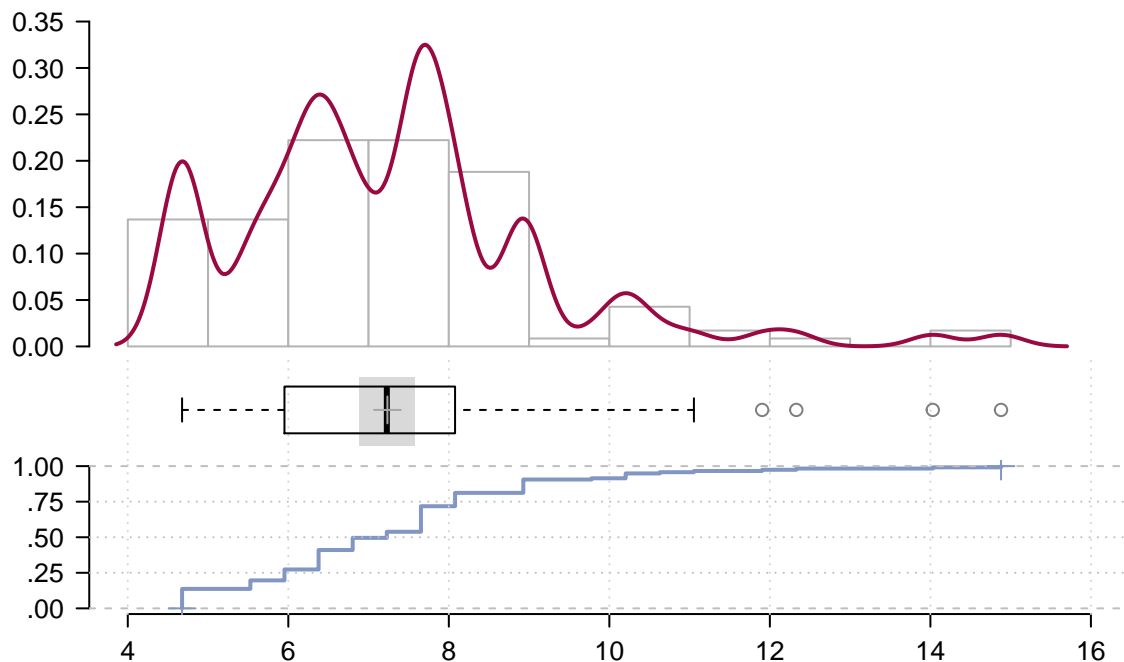
```
## [1] "Passat cidade kmpl = 7.89551009194324"
```

4. Usando `descTools::Desc` e `kilometros/litro`, mostre um resumo das estatísticas descritivas para `hwy` para carros de 1999 com o gráfico (`plotit = TRUE`). Também, imprime o `summary` de Base R.

```
Desc(kmpl$cty[kmpl$year == "1999"], plotit = TRUE)
```

```
## -----  
## kmpl$cty[kmpl$year == "1999"] (numeric)  
##  
##      length      n      NAs    unique      Os      mean      meanCI  
##      117      117       0       17       0  7.234696  6.887606  
##      100.0%    0.0%          0.0%          7.581786  
##  
##      .05      .10      .25    median      .75      .90      .95  
##  4.676571  4.676571  5.952000  7.227428  8.077714  8.928000  10.288457  
##  
##      range      sd      vcoef      mad      IQR      skew      kurt  
##  10.203428  1.895537  0.262006  1.260634  2.125714  1.178955  2.475057  
##  
## lowest : 4.676571 (16), 5.526857 (7), 5.952 (9), 6.377143 (16), 6.802286 (10)  
## highest: 11.053714, 11.904, 12.329143, 14.029714, 14.88
```

kmpl\$cty[kmpl\$year == "1999"] (numeric)

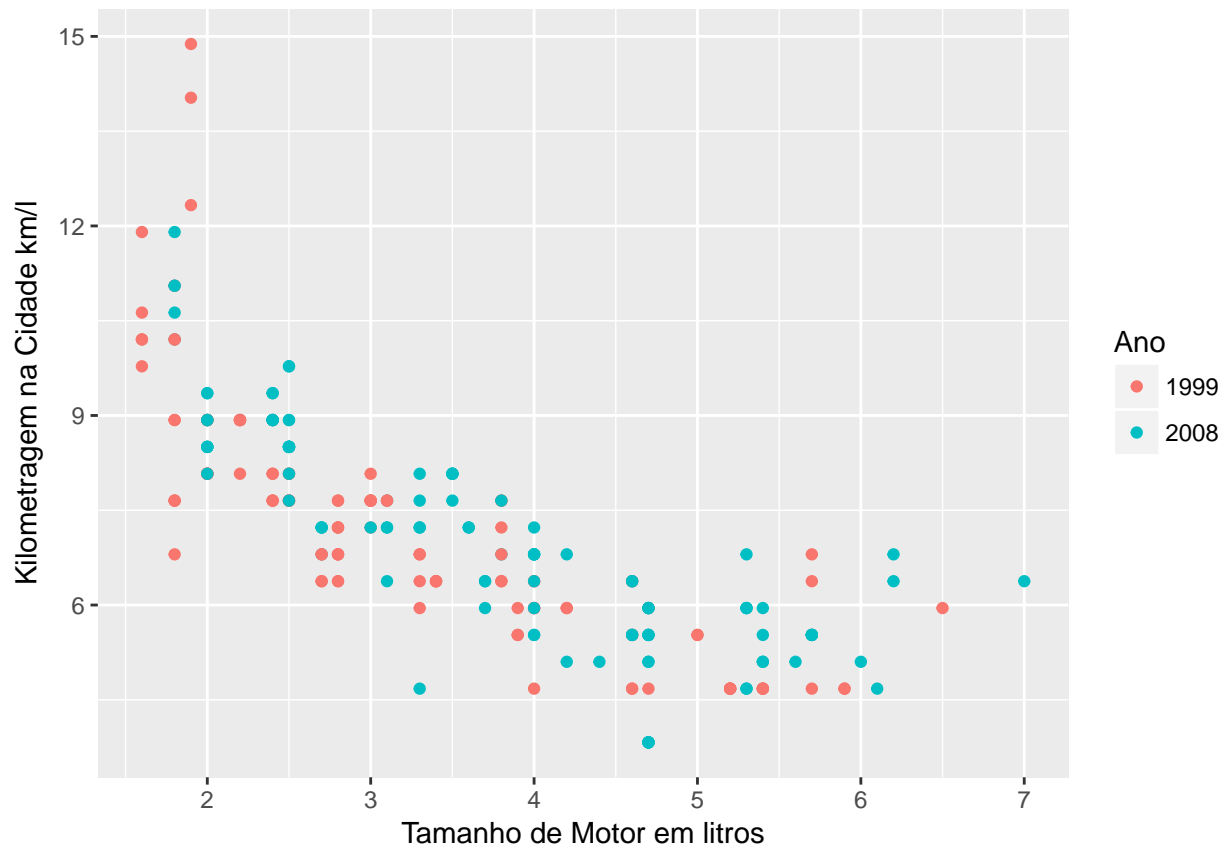


```
summary(kmpl$cty[kmpl$year == "1999"])
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##  4.677   5.952   7.227   7.235   8.078  14.880
```

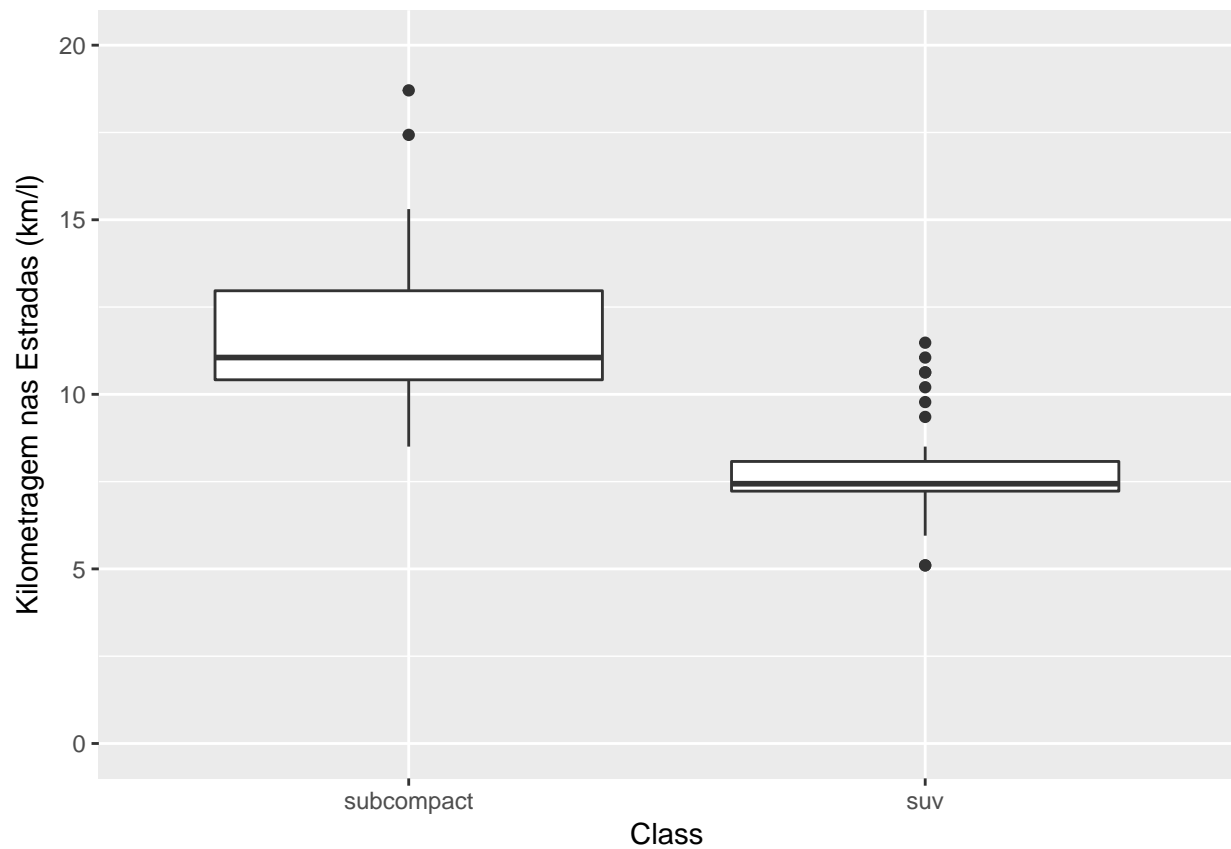
5. Faça um gráfico de dispersão que compara o kilometragem na cidade e o tamanho de motor (displ). Inclua rótulos para os eixos do gráfico e a legenda. Use cor para diferenciar entre carros de 1999 e 2008.

```
gr1 <- ggplot(data = kmp1, mapping = aes(x = displ, y = cty, colour = factor(year))) + geom_point()
gr1 <- gr1 + labs(x = "Tamanho de Motor em litros",
                  y = "Kilometragem na Cidade km/l",
                  colour = "Ano")
gr1
```



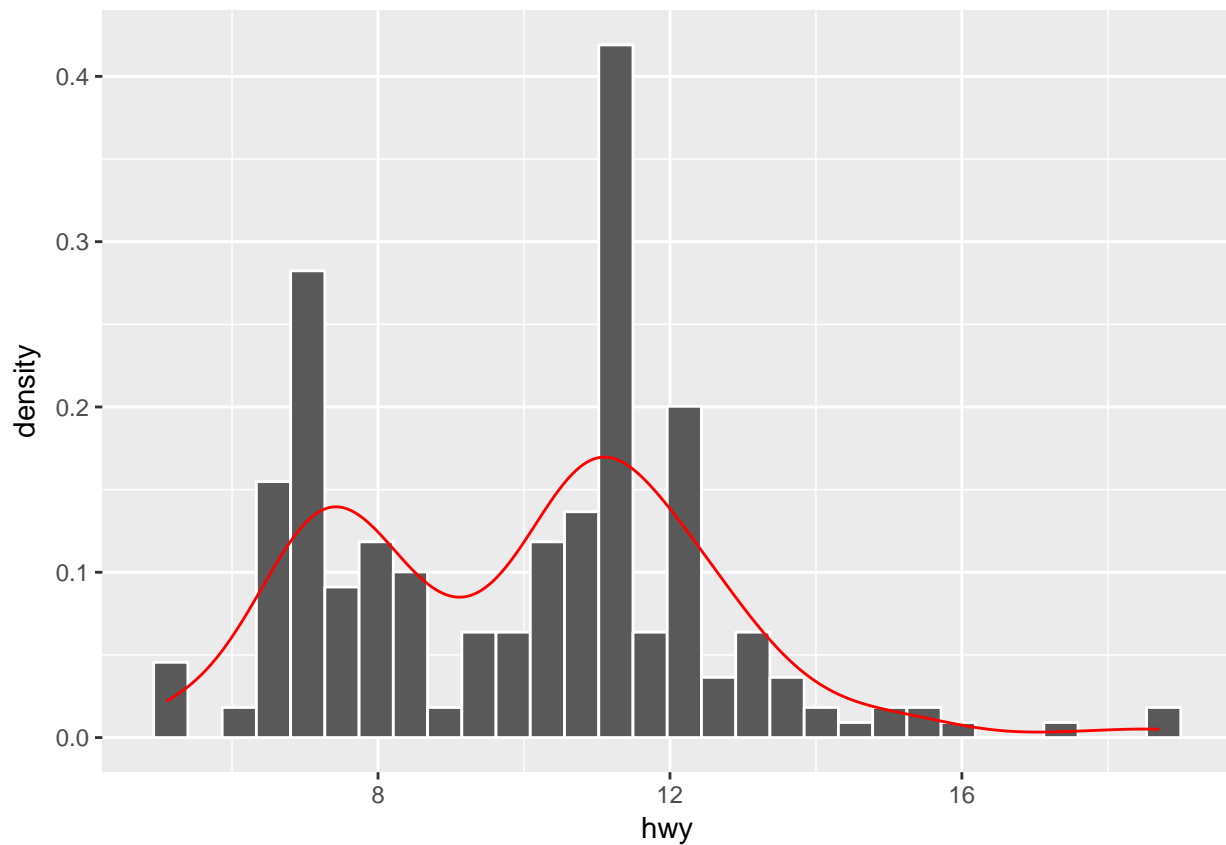
6. Desenhe um boxplot que compara a kilometragem nas estradas (hwy) para os carros de classe SUV contra os subcompactos.

```
gr2 <- kmp1 %>%
  filter(class == "suv" | class == "subcompact") %>%
  ggplot(mapping = aes(x = class, y = hwy)) + geom_boxplot() + ylim(0, 20)
gr2 <- gr2 + labs(x = "Class", y = "Kilometragem nas Estradas (km/l)")
gr2
```



7. Faça uma histograma de hwy com uma curva de densidade em vermelho sobreposta. Use o `geom_linha()` para fazer a curva.

```
gr3 <- ggplot(data = kmpl, mapping = aes(x = hwy, y = ..density..))
gr3 <- gr3 + geom_histogram(color = "white")
gr3 <- gr3 + geom_line(stat = "density", color = "red")
gr3
```



8. Finalmente, queremos ver um histograma e curva de densidade que mostra a diferença entre todas as classes de veículos na amostra. Use `facet_wrap()` com 4 fileiras.

```
gr4 <- gr3 + facet_wrap(~ class, nrow = 4)
gr4
```

