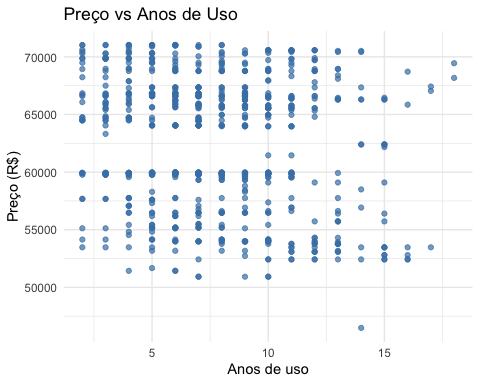
Exercicio1.R

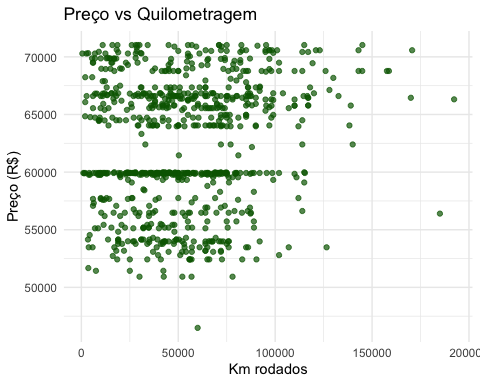
joaopaulodecker

2025-09-01

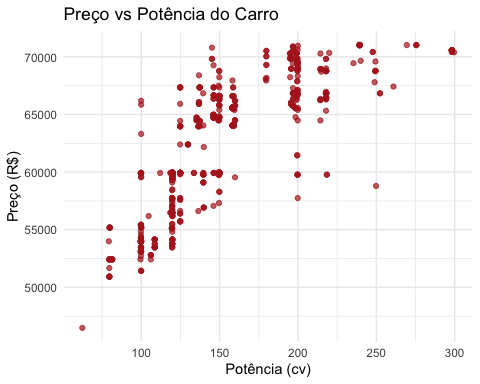
# 1 - Amostra  
base = read.csv2("exercicio.csv")  
set.seed(07052006)  
base1 = base[sample(nrow(base), 800),]  
  
  
  
  
  
# 2 - Gráfico de dispersão de cada váriavel relacionada com o preço  
library(ggplot2)  
  
# Gráfico 1: preco vs anos  
ggplot(base1, aes(x = anos, y = preco)) +  
 geom\_point(color = "steelblue", alpha = 0.7) +  
 labs(  
 title = "Preço vs Anos de Uso",  
 x = "Anos de uso",  
 y = "Preço (R$)"  
 ) +  
 theme\_minimal()



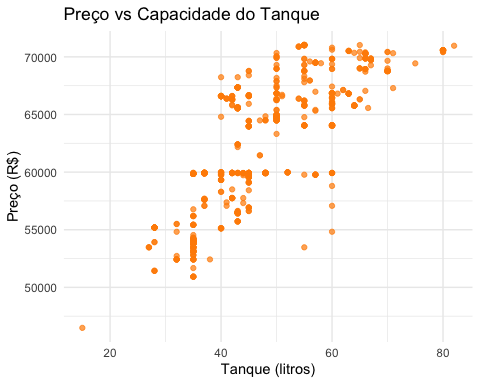
# Gráfico 2: preco vs km  
ggplot(base1, aes(x = km, y = preco)) +  
 geom\_point(color = "darkgreen", alpha = 0.7) +  
 labs(  
 title = "Preço vs Quilometragem",  
 x = "Km rodados",  
 y = "Preço (R$)"  
 ) +  
 theme\_minimal()



# Gráfico 3: preco vs potencia  
ggplot(base1, aes(x = potencia, y = preco)) +  
 geom\_point(color = "firebrick", alpha = 0.7) +  
 labs(  
 title = "Preço vs Potência do Carro",  
 x = "Potência (cv)",  
 y = "Preço (R$)"  
 ) +  
 theme\_minimal()



# Gráfico 4: preco vs tanque  
ggplot(base1, aes(x = tanque, y = preco)) +  
 geom\_point(color = "darkorange", alpha = 0.7) +  
 labs(  
 title = "Preço vs Capacidade do Tanque",  
 x = "Tanque (litros)",  
 y = "Preço (R$)"  
 ) +  
 theme\_minimal()



# 3 - Calculo do coeficiente de correlacao de cada variável quantitativa  
  
vars\_quant <- base1[, c("preco", "anos", "km", "potencia", "tanque")]  
correlacoes <- cor(vars\_quant, use = "complete.obs")  
correlacoes["preco", -1]

## anos km potencia tanque   
## -0.1427165 0.1460225 0.8076732 0.7540899

# 4 - Intervalo de confiança da correlação entre potência e preço   
  
r <- cor(base1$preco, base1$potencia, use = "complete.obs")  
n <- sum(complete.cases(base1$preco, base1$potencia))  
  
# Transformação de Fisher  
z <- 0.5 \* log((1 + r) / (1 - r))  
se <- 1 / sqrt(n - 3)  
  
# Intervalo de confiança em z  
z\_low <- z - 1.96 \* se  
z\_high <- z + 1.96 \* se  
  
# Voltar para escala de correlação  
r\_low <- (exp(2 \* z\_low) - 1) / (exp(2 \* z\_low) + 1)  
r\_high <- (exp(2 \* z\_high) - 1) / (exp(2 \* z\_high) + 1)  
  
# Resultado  
c("inferior" = r\_low, "superior" = r\_high)

## inferior superior   
## 0.7821456 0.8304941

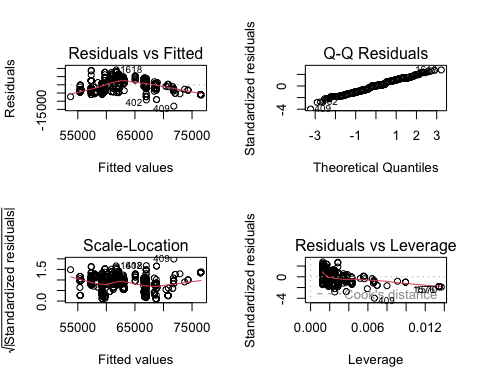
# 5 - Modelo de regressão da variável preço x potência  
modelo <- lm(preco ~ potencia, data = base1)  
summary(modelo)

##   
## Call:  
## lm(formula = preco ~ potencia, data = base1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -13017.3 -2318.8 171.6 2573.1 9087.4   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 47708.772 398.069 119.9 <2e-16 \*\*\*  
## potencia 96.511 2.494 38.7 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3287 on 798 degrees of freedom  
## Multiple R-squared: 0.6523, Adjusted R-squared: 0.6519   
## F-statistic: 1497 on 1 and 798 DF, p-value: < 2.2e-16

# a - o coeficiente de determinação e interprete o valor.  
summary(modelo)$r.squared

## [1] 0.6523361

# b - a análise de resíduos  
par(mfrow = c(2, 2))  
plot(modelo)



# c - teste de significancia do modelo  
summary(modelo)$coefficients

## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 47708.7720 398.068563 119.85064 0.000000e+00  
## potencia 96.5114 2.494143 38.69522 2.923956e-185

# d - interpretacao dos coeficientes do modelo  
coef(modelo)

## (Intercept) potencia   
## 47708.7720 96.5114

# 6 - Gráfico de dispersão preco x potencia diferenciando transmissão automática e manual  
library(ggplot2)  
  
ggplot(base1, aes(x = potencia, y = preco, color = transm)) +  
 geom\_point(alpha = 0.7, size = 2) +  
 labs(  
 title = "Preço vs Potência com Tipo de Transmissão",  
 x = "Potência (cv)",  
 y = "Preço (R$)",  
 color = "Transmissão"  
 ) +  
 theme\_minimal()

