

Lista 03 - MQ

Martha Gaudencio

2026-02-01

```
#MQ101 - Lista 03
#Nome: Martha Gaudencio da Silva
#Data: 04/11/2025
#Descrição: Gráficos e Visualização de Dados
```

```
set.seed(101)
library(tidyverse)
```

```
## — Attaching core tidyverse packages — tidyverse 2.0.0 —
## ✓ dplyr      1.1.4    ✓ readr      2.1.6
## ✓ forcats    1.0.1    ✓ stringr    1.6.0
## ✓ ggplot2     4.0.1    ✓ tibble     3.3.1
## ✓ lubridate  1.9.4    ✓ tidyr      1.3.2
## ✓ purrr      1.2.1
## — Conflicts — tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()     masks stats::lag()
## ⓘ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to be
come errors
```

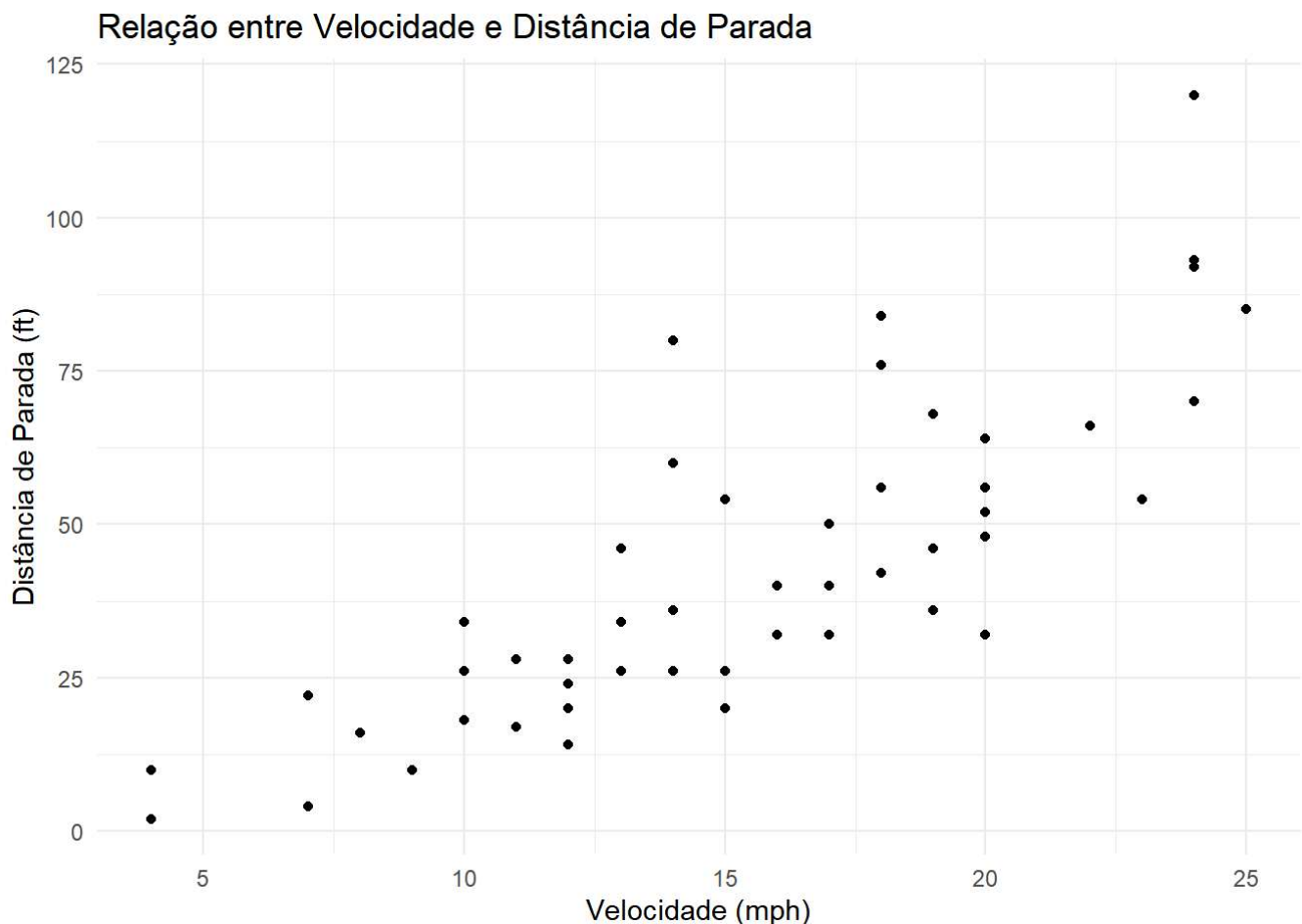
```
library(ggplot2)
```

```
#1 - Tabela vs. gráficos
```

```
data(cars)
head(cars, 10)
```

```
##      speed dist
## 1         4     2
## 2         4    10
## 3         7     4
## 4         7    22
## 5         8    16
## 6         9    10
## 7        10    18
## 8        10    26
## 9        10    34
## 10       11    17
```

```
ggplot(cars, aes(x = speed, y = dist)) +
  geom_point() +
  labs(x = "Velocidade (mph)",
       y = "Distância de Parada (ft)",
       title = "Relação entre Velocidade e Distância de Parada") +
  theme_minimal()
```



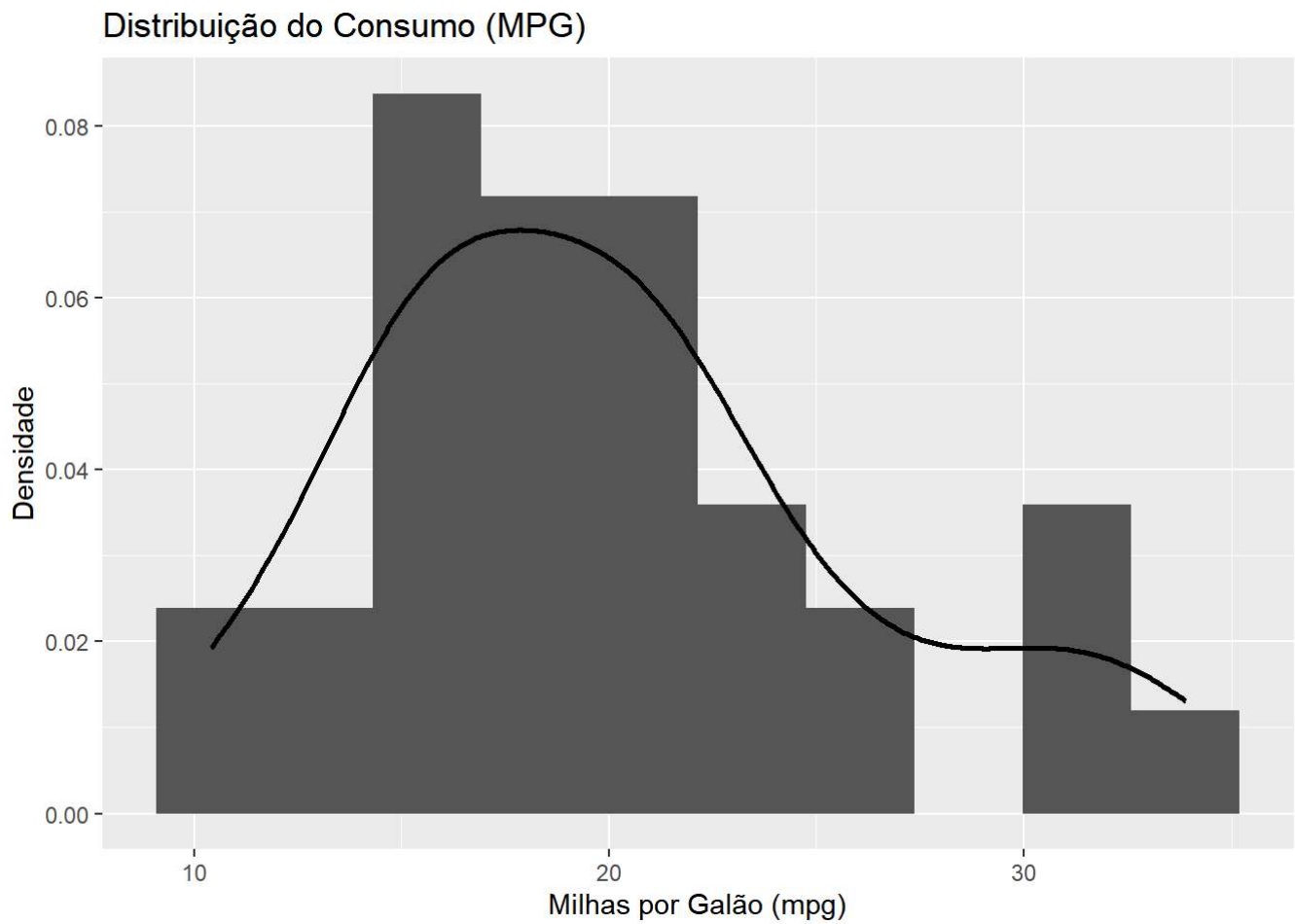
#O gráfico mostra que há uma tendência de crescimento da distância de parada relacionada ao aumento da velocidade, em uma correlação positiva.

#2 - Distribuições univariadas e grupos

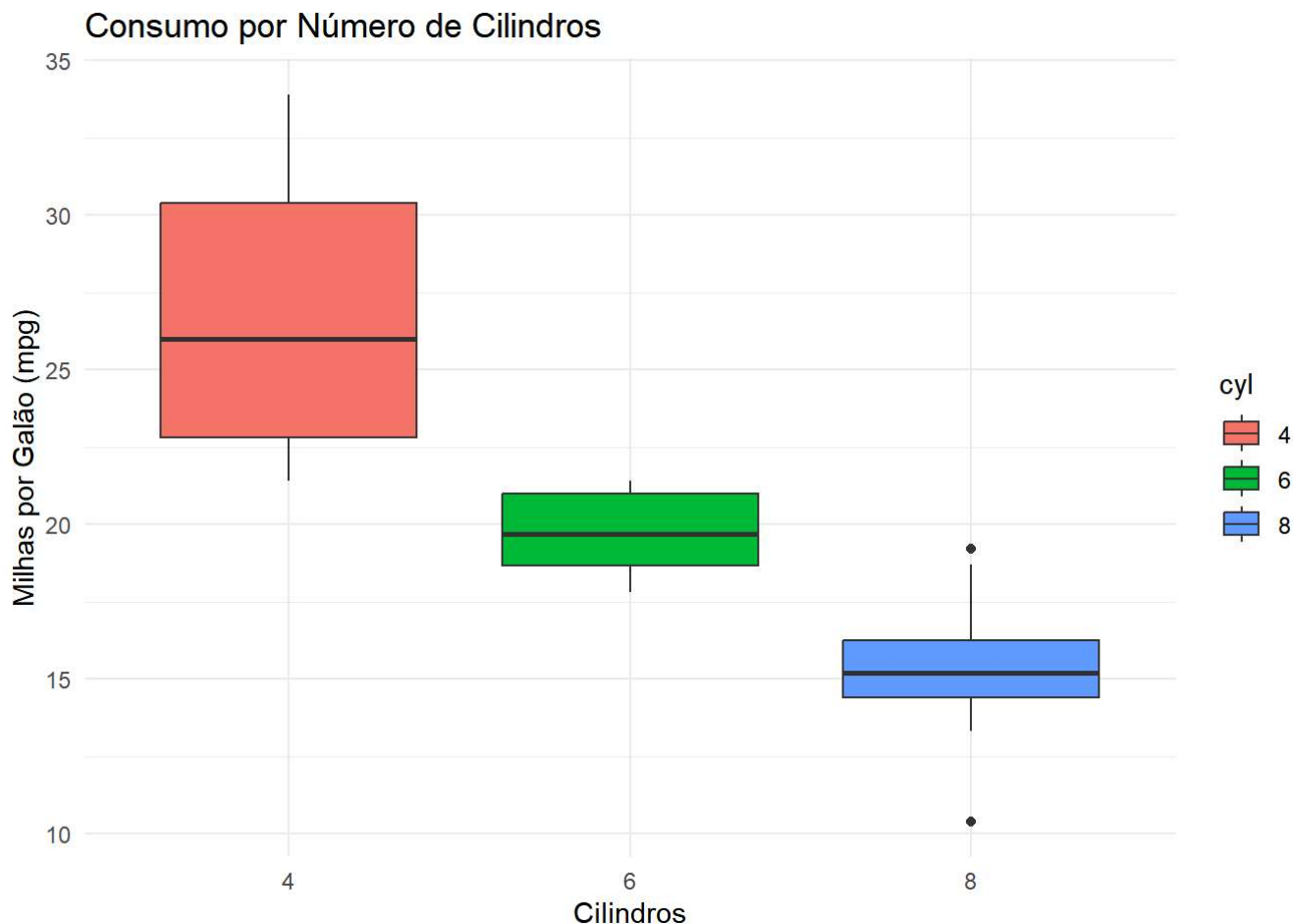
```
data(mtcars)
mtcars <- mtcars |>
  mutate(cyl = as.factor(cyl))

ggplot(mtcars, aes(x = mpg)) +
  geom_histogram(aes(y = after_stat(density)), bins = 10) +
  geom_density(size = 1) +
  labs(title = "Distribuição do Consumo (MPG)",
       x = "Milhas por Galão (mpg)", y = "Densidade")
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once per session.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



```
ggplot(mtcars, aes(x = cyl, y = mpg, fill = cyl)) +  
  geom_boxplot() +  
  labs(title = "Consumo por Número de Cilindros",  
        x = "Cilindros", y = "Milhas por Galão (mpg)") +  
  theme_minimal()
```



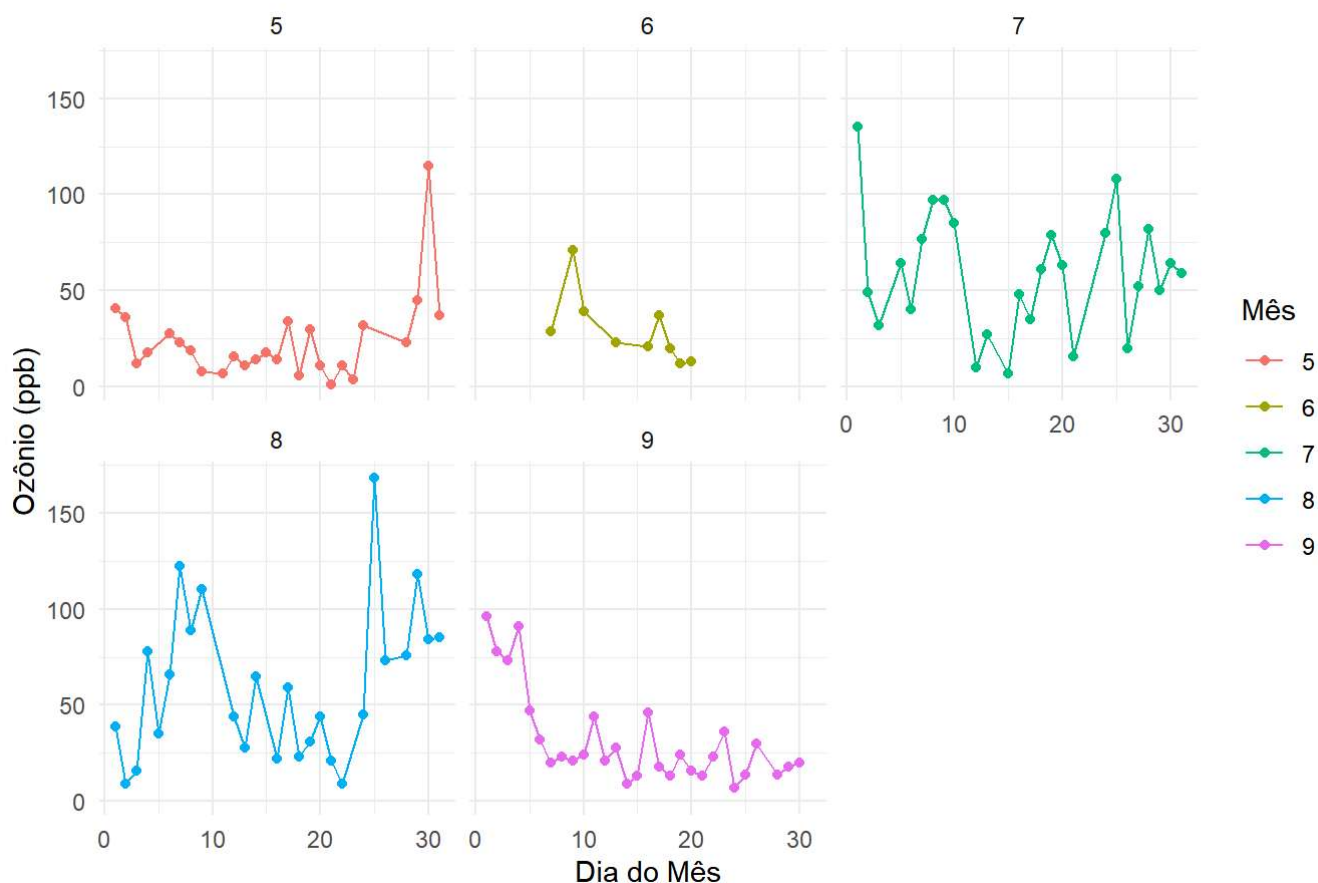
#0 histograma mostra que o maior consumo dos carros está em torno de 20mpg,
 #e o gráfico tem uma cauda à direita. Já o boxplot mostra que quanto mais
 #cilindros tem o motor do automóvel menos milhas por galão ele faz, revelando
 #um padrão monotônico decrescente.

#3 - Série temporal simples

```
aq <- airquality |>
  as_tibble() |>
  drop_na(Ozone) |>
  mutate(Month = factor(Month),
         Day = as.integer(Day))

ggplot(aq, aes(x = Day, y = Ozone, group = Month, color = Month)) +
  geom_line() + # Cria a linha conectando os dias
  geom_point() + # Adiciona os pontos de medição
  facet_wrap(~Month) +
  labs(title = "Níveis de Ozônio por Dia (Nova York, 1973)",
       x = "Dia do Mês",
       y = "Ozônio (ppb)",
       color = "Mês") +
  theme_minimal()
```

Níveis de Ozônio por Dia (Nova York, 1973)



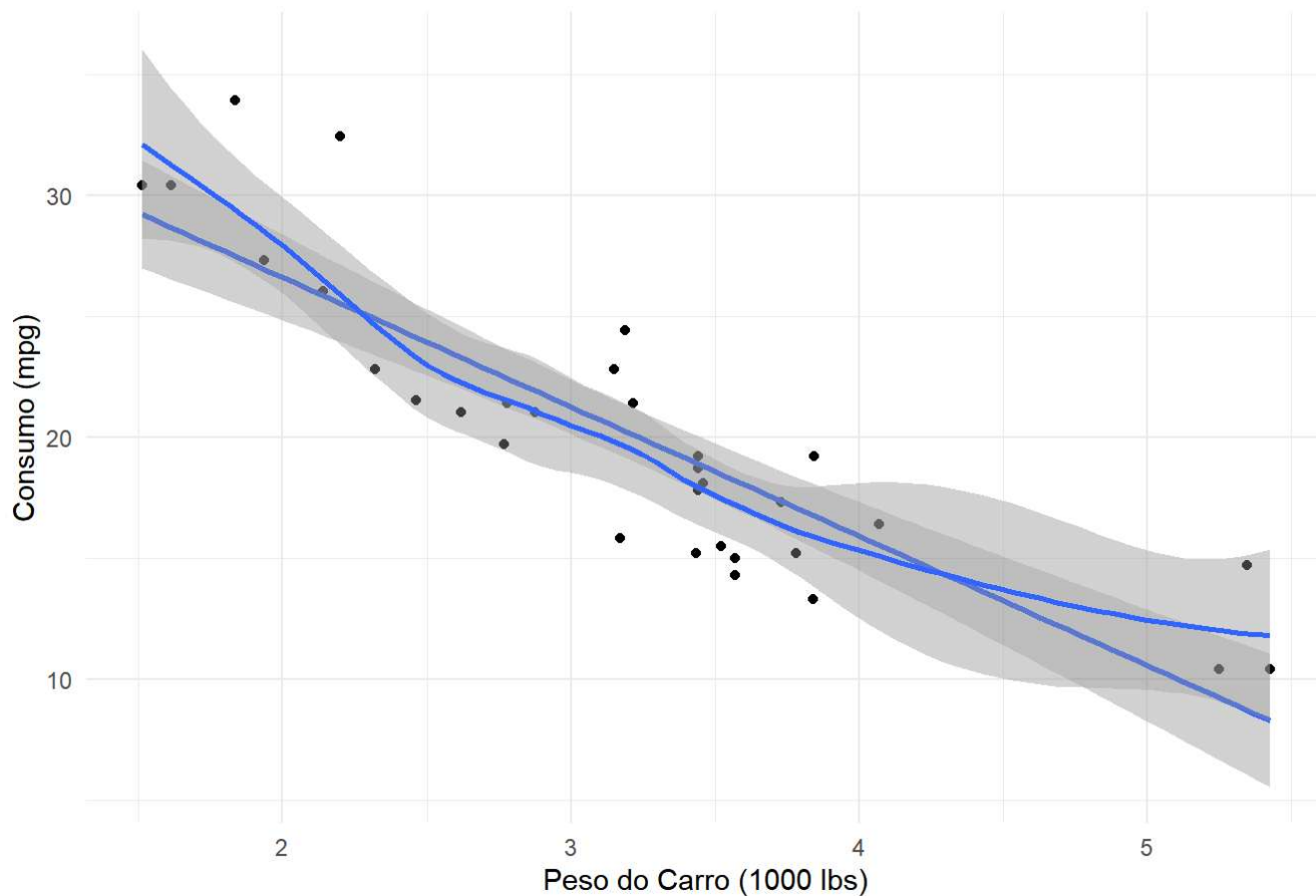
#Os gráficos mostram que os níveis de ozônio são maiores em julho e agosto,
#nos meses de verão em Nova Iorque. Os níveis variam ao longo dos dias dos meses,
#não havendo uma tendência de linearidade.

#4 Relações bivariadas

```
ggplot(mtcars, aes(x = wt, y = mpg)) +
  geom_point() +
  # Linha Reta (LM - Linear Model) em azul
  geom_smooth(method = "lm") +
  # Linha Curva (LOESS - Suavizada) em vermelho
  geom_smooth(method = "loess") +
  labs(title = "Relação Peso vs Consumo (mtcars)",
        x = "Peso do Carro (1000 lbs)",
        y = "Consumo (mpg)") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```

Relação Peso vs Consumo (mtcars)



#0 gráfico com as linhas lm e loess mostra que ainda que haja diferenças entre as duas, uma vez que uma é linear e outra é flexível, ambas evidenciam que quanto mais pesado o veículo mais combustível ele gasta.

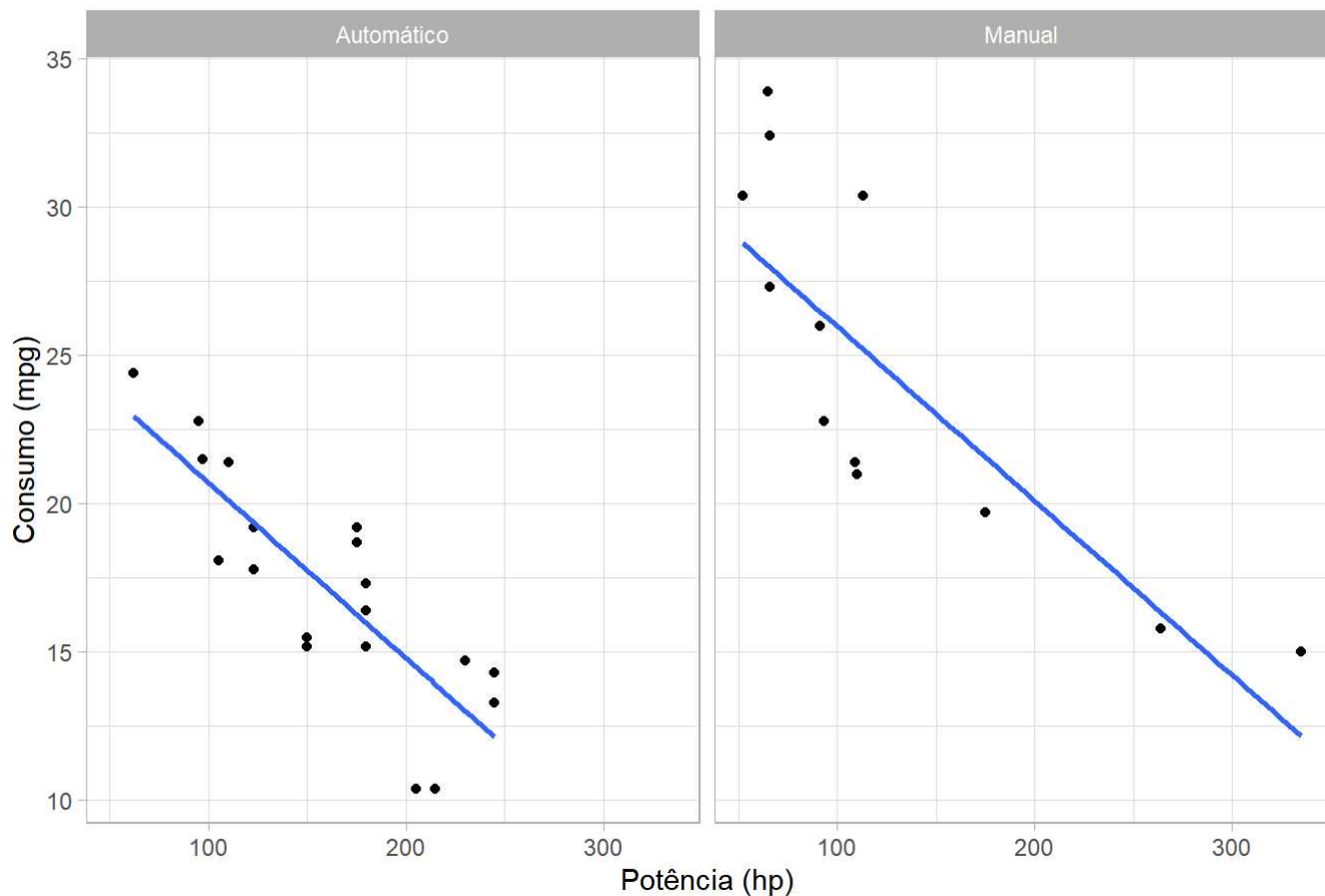
#5 - Facetas

```
mtcars <- mtcars |>
  mutate(am = factor(am, labels = c("Automático", "Manual")))

ggplot(mtcars, aes(x = hp, y = mpg)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  facet_wrap(~ am) + # Aqui acontece a mágica de separar os quadros
  labs(title = "Potência (hp) vs Consumo (mpg) por Transmissão",
       x = "Potência (hp)",
       y = "Consumo (mpg)") +
  theme_light()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Potência (hp) vs Consumo (mpg) por Transmissão



#Os gráficos mostram que os carros manuais atingiam maior potência, sendo o único tipo a alcançar 300hp. Também são mais econômicos. Ainda assim, os dois gráficos têm tendência de correlação negativa.

#6 - Simulação I - Correlação controlada

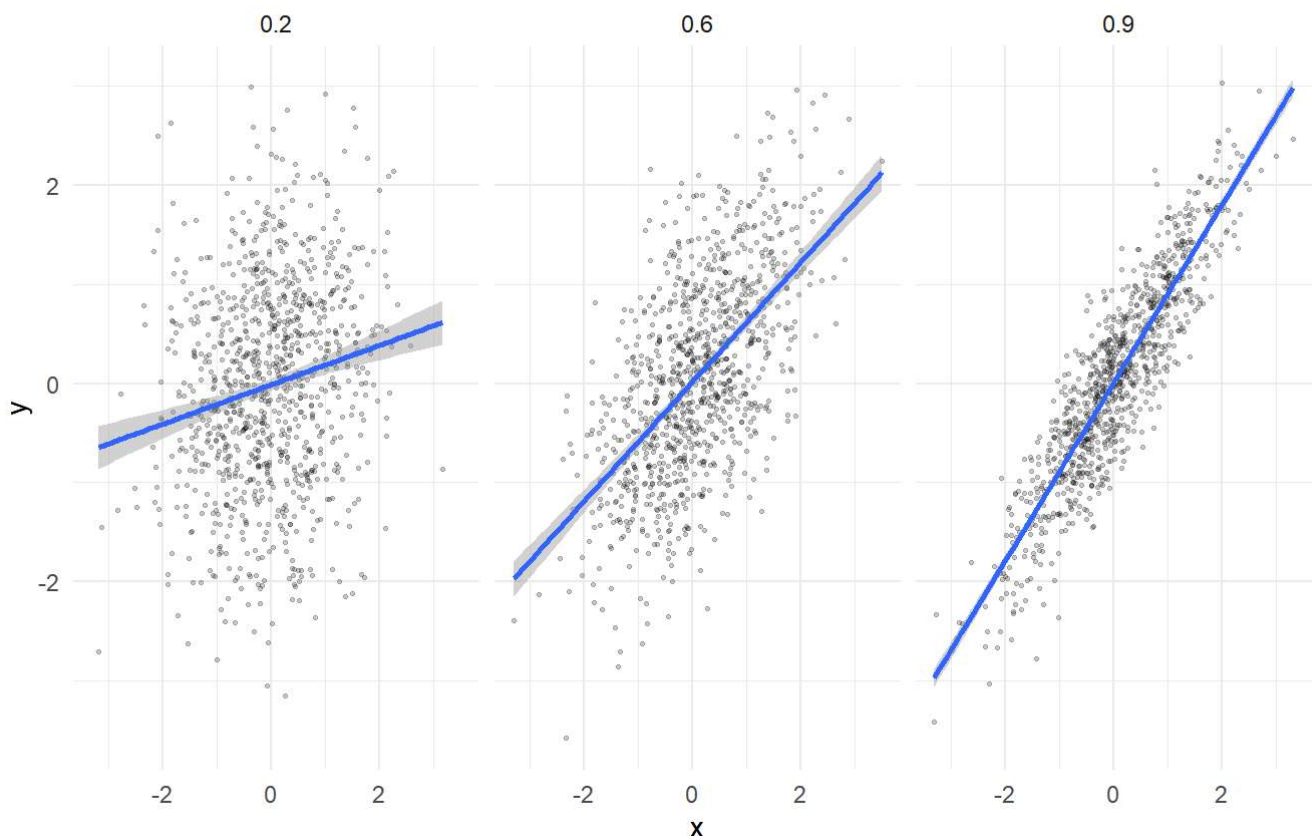
```
n <- 1000; rhos <- c(0.2, 0.6, 0.9)
sim <- purrr::map_dfr(rhos, \(rho) {
  x <- rnorm(n); e <- rnorm(n)
  y <- rho*x + sqrt(1 - rho^2)*e
  tibble(rho = rho, x = x, y = y)
})

ggplot(sim, aes(x = x, y = y)) +
  geom_point(alpha = 0.2, size = 0.5) +
  geom_smooth(method = "lm") +
  facet_wrap(~ rho) +
  labs(title = "Simulação: O efeito do Rho (ρ) na dispersão",
       subtitle = "0.2 é espalhado, 0.9 é quase uma linha") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Simulação: O efeito do Rho (ρ) na dispersão

0.2 é espalhado, 0.9 é quase uma linha



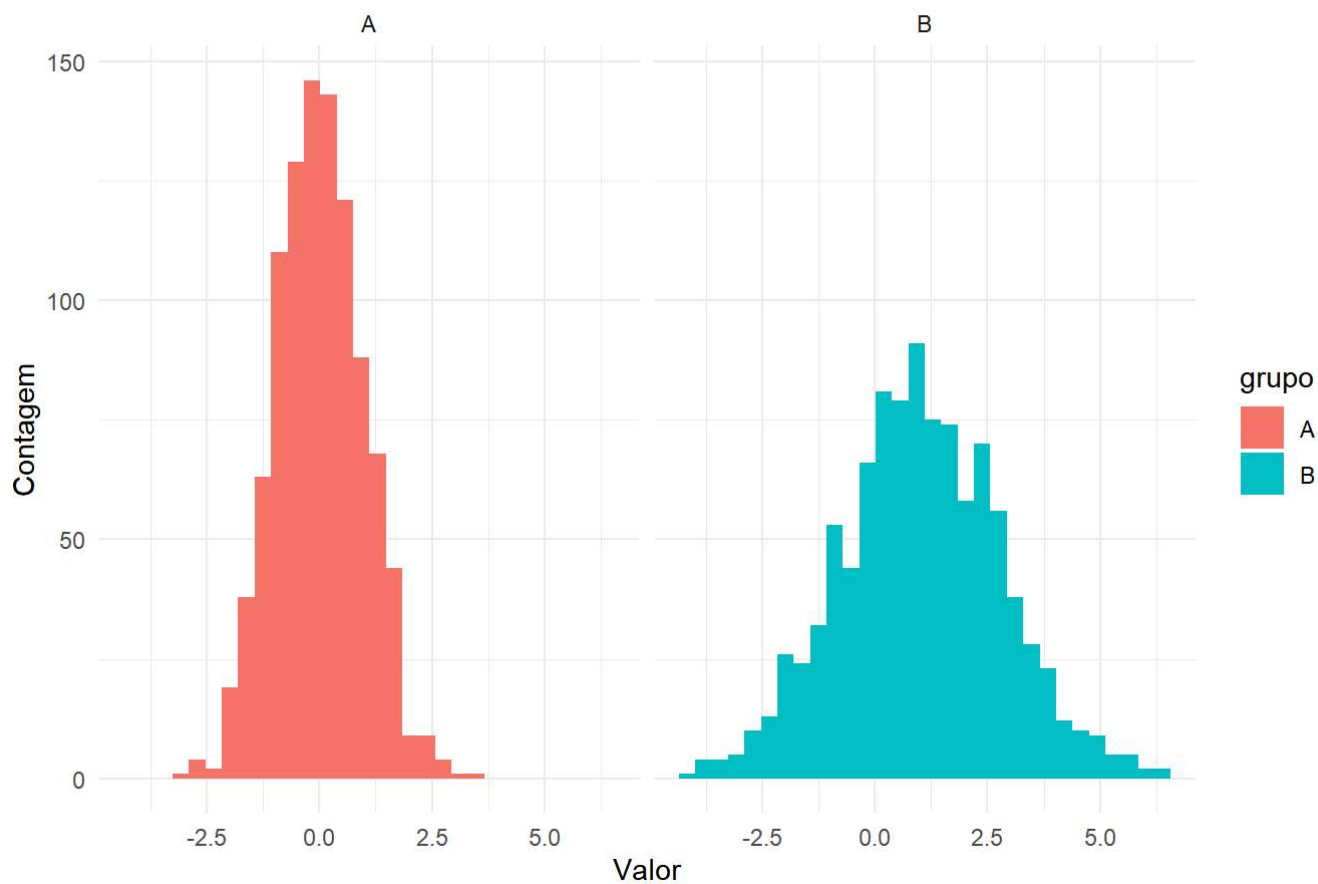
*#Os gráficos mostram a Correlação de Pearson, variando de quando ela é mais
#fraca para mais forte: conforme o 'p' aumenta, mais se tem uma correlação
#positiva forte.*

#7 Simulação II - Diferenças entre grupos

```
n <- 1000
df_grupos <- tibble(
  grupo = rep(c("A","B"), each = n),
  valor = c(rnorm(n, 0, 1), rnorm(n, 1, 1.8))
)

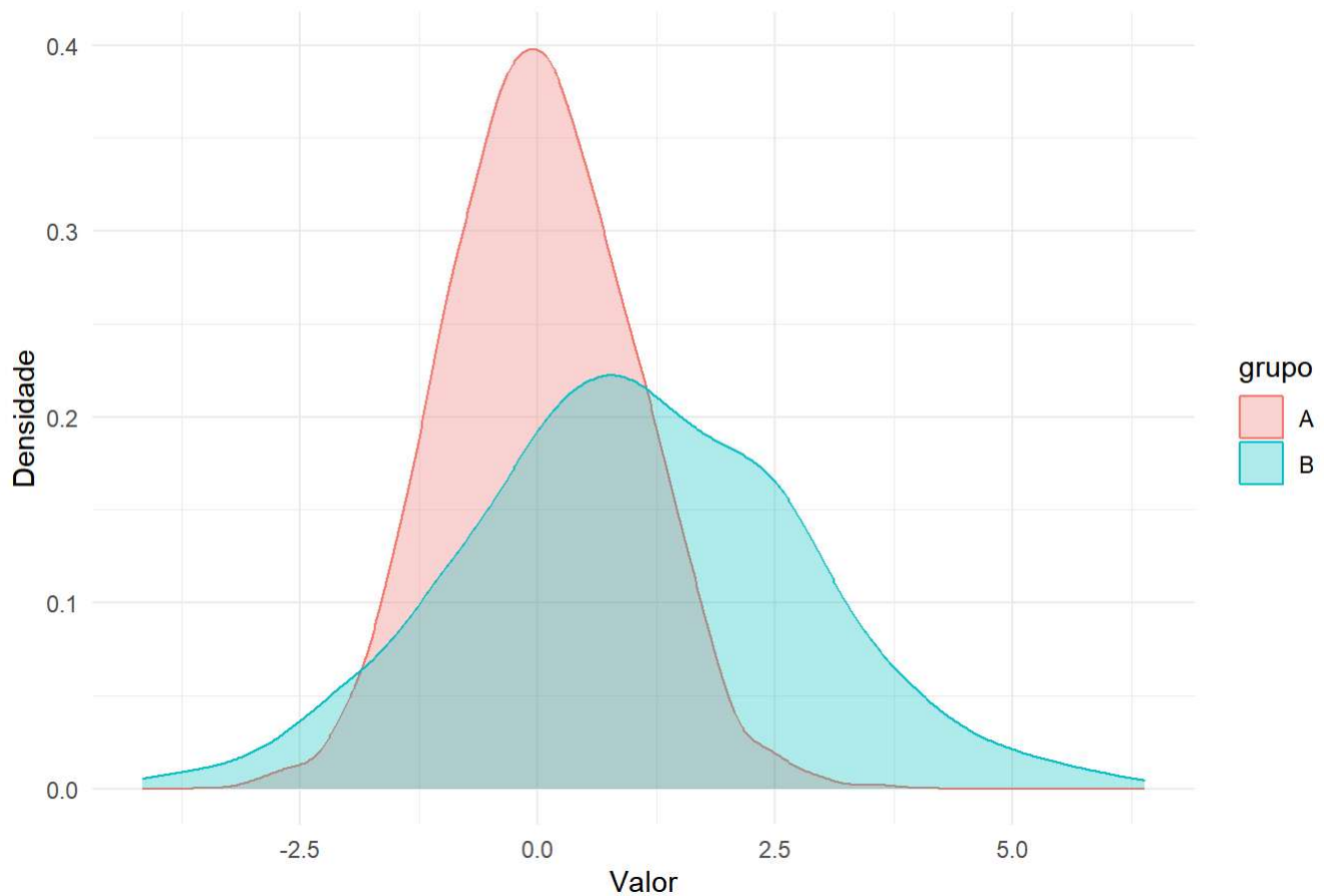
ggplot(df_grupos, aes(x = valor, fill = grupo)) +
  geom_histogram(bins = 30) +
  facet_wrap(~grupo) +
  labs(title = "Histogramas: Frequência de Valores", x = "Valor", y =
    "Contagem") +
  theme_minimal()
```


Histogramas: Frequência de Valores



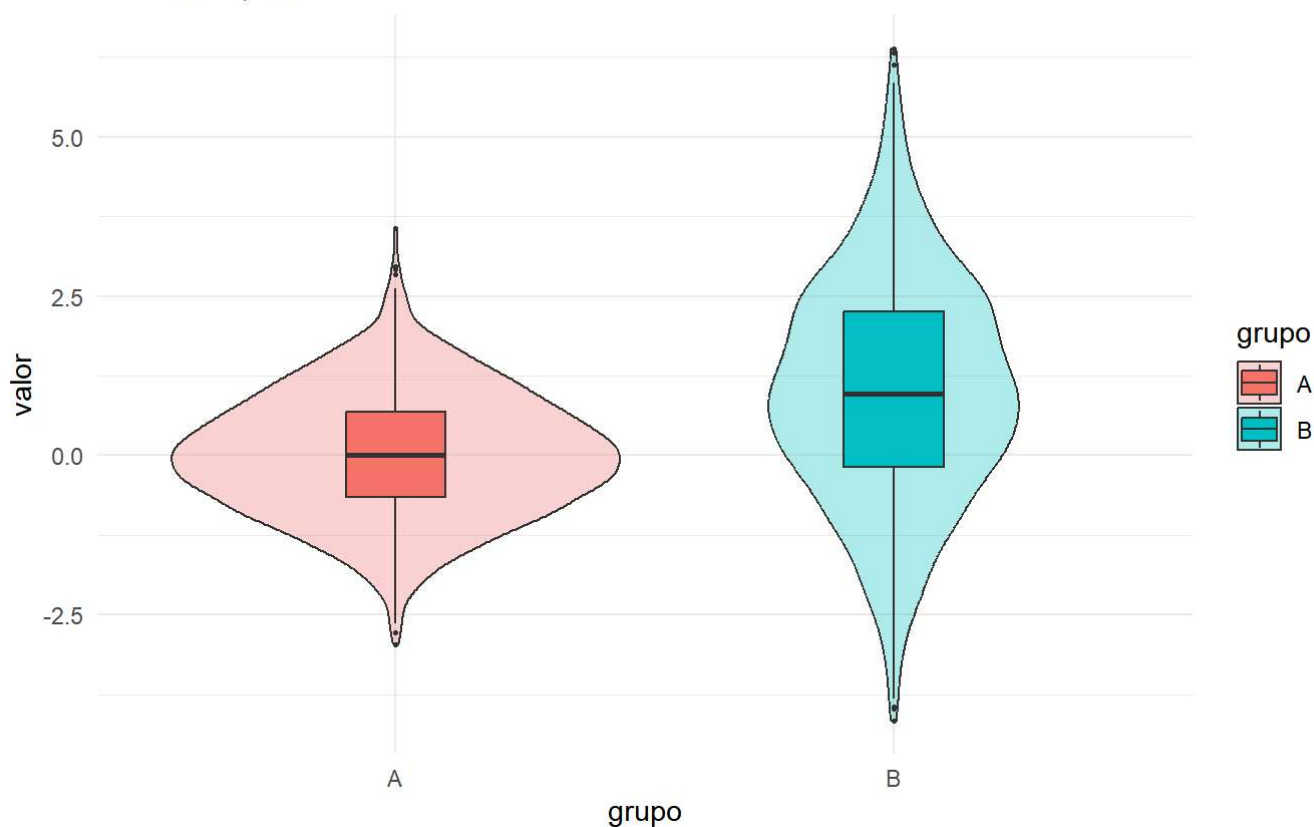
```
ggplot(df_grupos, aes(x = valor, color = grupo, fill = grupo)) +  
  geom_density(alpha = 0.3) +  
  labs(title = "Densidade: Comparação da Forma (A vs B)", x = "Valor", y =  
    "Densidade") +  
  theme_minimal()
```

Densidade: Comparação da Forma (A vs B)



```
ggplot(df_grupos, aes(x = grupo, y = valor, fill = grupo)) +  
  geom_violin(alpha = 0.3) +  
  geom_boxplot(width = 0.2, outlier.size = 0.5) +  
  labs(title = "Distribuição: Grupo A vs Grupo B", subtitle = "Violin Plot  
e Boxplot") +  
  theme_minimal()
```

Distribuição: Grupo A vs Grupo B

Violin Plot
e Boxplot

#8 - Educação e saúde

```
library(readr)
educ <- read.csv(file.choose())
glimpse(educ)
```

```
## Rows: 10,000
## Columns: 13
## $ id          <chr> "P00001", "P00002", "P00003", "P00004", "P00005", "P...
## $ sexo        <chr> "M", "F", "F", "F", "F", "F", "F", "F", "F", "F", "M...
## $ escolaridade <chr> "Medio", "Medio", "Medio", "Medio", "Medio", "Medio"...
## $ anos_estudo  <int> 10, 9, 9, 10, 11, 11, 10, 8, 15, 11, 6, 7, 5, 4, 10,...
## $ rede_escolar <chr> "privada", "pública", "pública", "pública", "pública...
## $ municipio    <chr> "Santos", "Dourados", "Várzea Grande", "Curitiba", "...
## $ UF           <chr> "SP", "MS", "MT", "PR", "RJ", "SP", "SC", "SP", "SP"...
## $ idade        <int> 39, 30, 31, 44, 26, 51, 47, 48, 25, 53, 23, 27, 35, ...
## $ faltas_esc   <int> 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 2, 0, 0, 0...
## $ tempo_estudo_h <dbl> 2.3, 6.6, 8.8, 4.2, 3.7, 4.9, 0.2, 1.9, 14.4, 3.3, 3...
## $ pressao_sistolica <int> 110, 109, 121, 131, 134, 124, 139, 124, 132, 134, 13...
## $ diagnostico  <chr> "sem", "sem", "sem", "sem", "sem", "HAS", "sem", "se...
## $ plano_saude  <chr> "privado", "privado", "nenhum", "nenhum", "SUS", "pr..."
```

#Boxplot

#Variáveis escolhidas: anos de estudo - educação e tipo de plano - saúde.

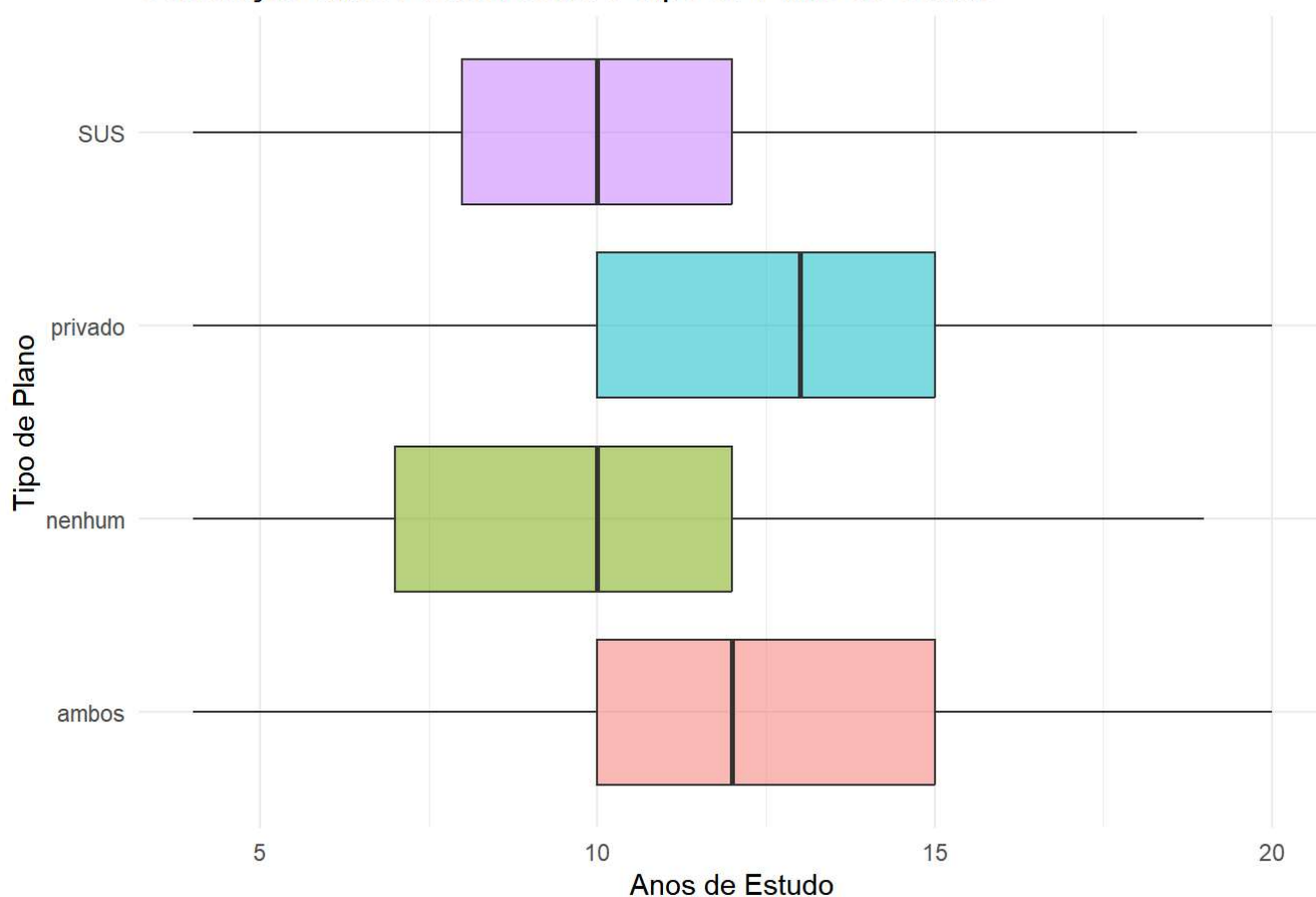
#O gráfico mostrou que menos anos de estudos apresentam mais 'nenhum' como

#plano de saúde e pessoas com mais anos de estudo tendem a usar mais tanto

#o sus e a rede privada como a rede privada.

```
ggplot(educ, aes(x = anos_estudo, y = plano_saude, fill = plano_saude)) +
  geom_boxplot(alpha = 0.5, outlier.shape = NA) +
  labs(title = "Correlação entre Escolaridade e Tipo de Plano de Saúde",
        x = "Anos de Estudo",
        y = "Tipo de Plano") +
  theme_minimal() +
  theme(legend.position = "none")
```

Correlação entre Escolaridade e Tipo de Plano de Saúde

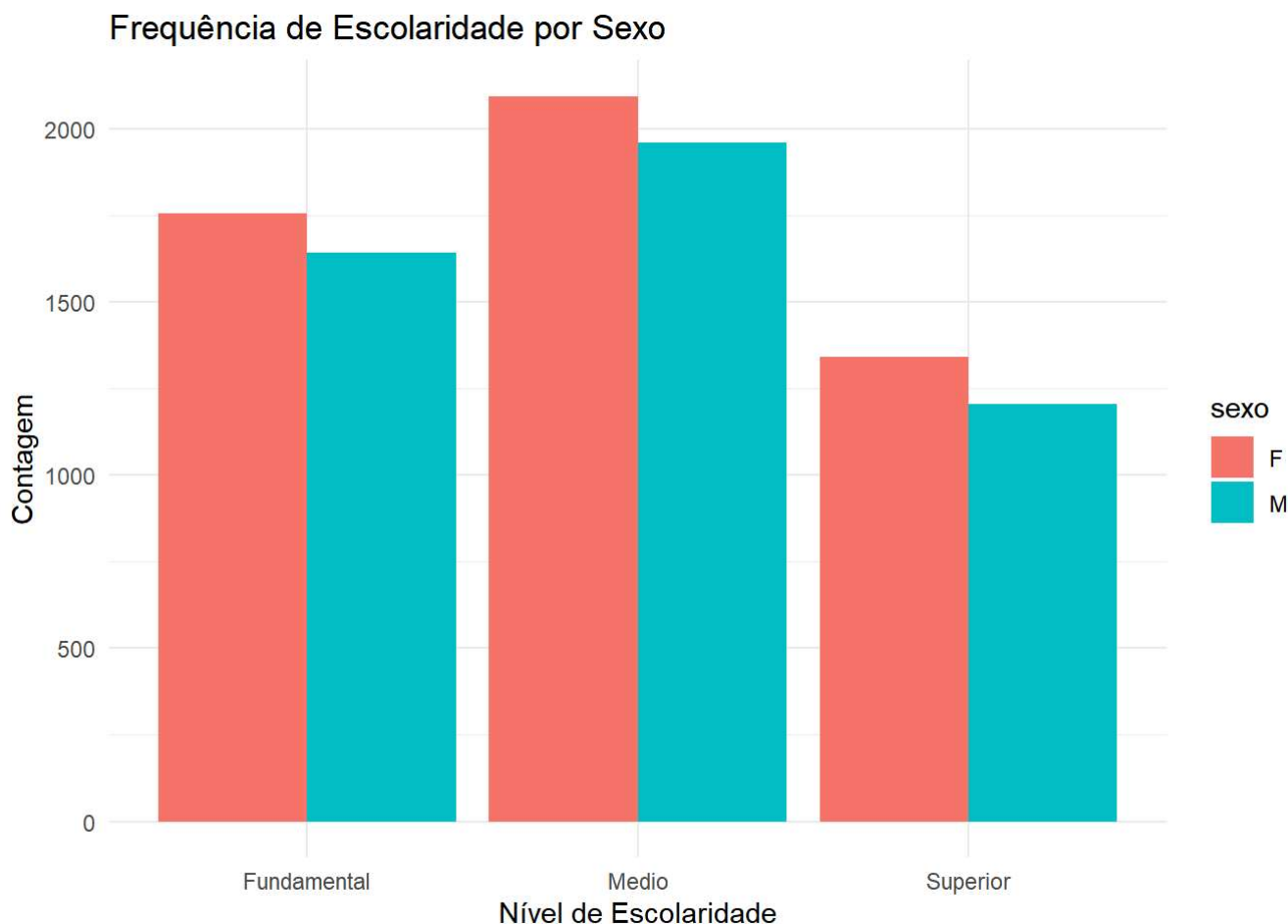


#Histograma

#variáveis: nível de escolaridade e sexo

#esse gráfico apontou que mulheres tendem a ter mais escolaridade que os homens.

```
ggplot(educ, aes(x = escolaridade, fill = sexo)) +
  geom_bar(position = "dodge") +
  labs(title = "Frequência de Escolaridade por Sexo",
        x = "Nível de Escolaridade",
        y = "Contagem") +
  theme_minimal()
```



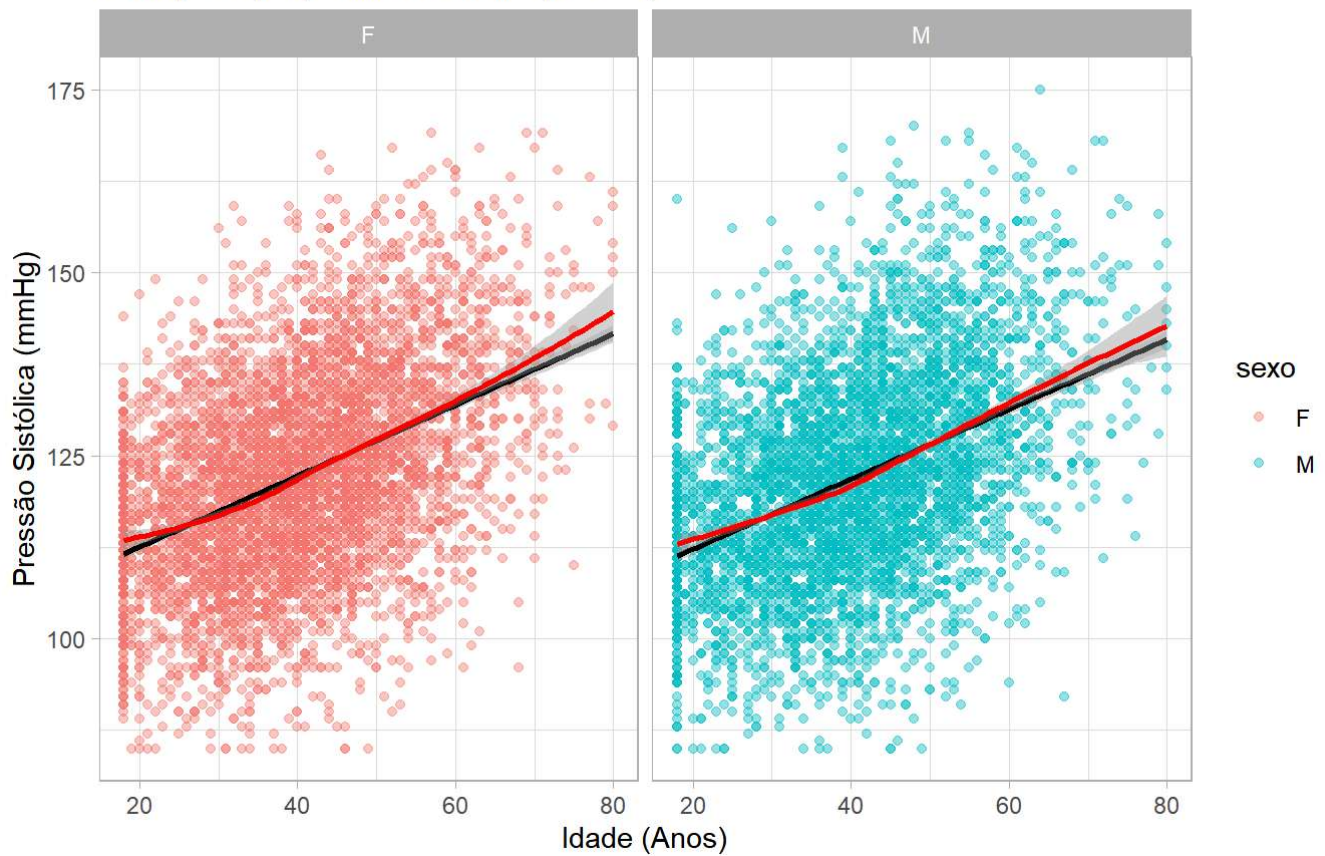
#9 - Figura Final

```
ggplot(educ, aes(x = idade, y = pressao_sistolica, color = sexo)) +
  geom_point(alpha = 0.4) + # Os pontinhos
  geom_smooth(method = "lm", color = "black") + # A linha RETA (Tendência geral)
  geom_smooth(method = "loess", color = "red") + # A linha CURVA (Tendência real)
  facet_wrap(~sexo) + # Um gráfico pra cada sexo
  labs(title = "Relação entre Idade e Pressão Sistólica",
        subtitle = "Linha preta (LM) e linha vermelha (LOESS)",
        x = "Idade (Anos)",
        y = "Pressão Sistólica (mmHg)") +
  theme_light()
```

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'`
```

Relação entre Idade e Pressão Sistólica

Linha preta (LM) e linha vermelha (LOESS)



#A figura final mostra que há uma correlação positiva entre idade e pressão:
#conforme a idade aumenta, a pressão tende a aumentar. Ademais, os gráficos
#evidenciam uma informação interessante: as mulheres apresentam pressão maior
#que a dos homens pelas informações apresentadas.