

Exercício 1

Matheus Silvano Pereira

2025-04-01

Variável	Tipo	Subtipo
experiencia	Qualitativa	Ordinal
emprego	Qualitativa	Nominal
cargo	Qualitativa	Nominal
salario_USD	Quantitativa	Contínua
trab_remoto	Qualitativa	Ordinal

```
# Carregar pacotes necessários
library(dplyr)

##
## Anexando pacote: 'dplyr'

## Os seguintes objetos são mascarados por 'package:stats':
##
##   filter, lag

## Os seguintes objetos são mascarados por 'package:base':
##
##   intersect, setdiff, setequal, union

# 2
# Definir semente para reprodutibilidade
set.seed(19092005)

# Carregar a base de dados
base <- read.csv("dados_salarios.csv", sep = ";")

# Amostragem aleatória de 350 observações
base1 <- base %>% sample_n(350)

# 3
# Recodificar os valores da variável trab_reamoto
```

```
base1$trab_remoto <- factor(base1$trab_remoto,
                           levels = c(0, 50, 100),
                           labels = c("Não", "Parcial", "Sim"))

# 4
# Carregar pacotes necessários
library(ggplot2)

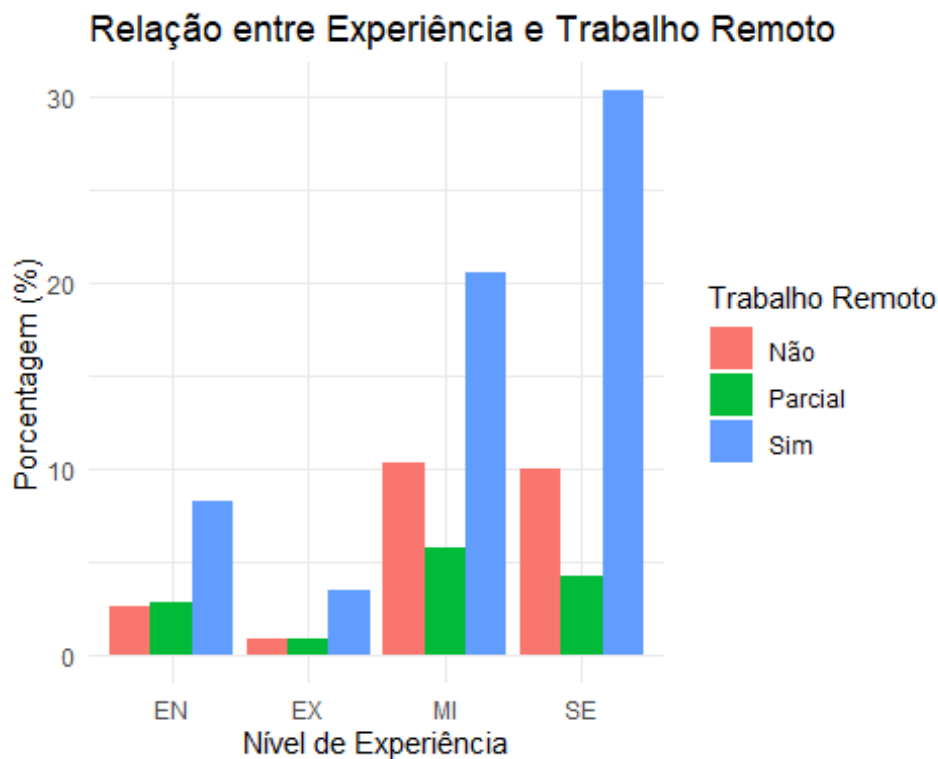
# Criar gráfico de barras com proporção em porcentagem
ggplot(base1, aes(x = experiencia)) +
  geom_bar(aes(y = (..count..) / sum(..count..) * 100), fill = "steelblue") +
  labs(title = "Distribuição de Experiência dos Profissionais",
       x = "Nível de Experiência",
       y = "Porcentagem (%)") +
  theme_minimal()

## Warning: The dot-dot notation (`..count..`) was deprecated in ggplot2
## 3.4.0.
## i Please use `after_stat(count)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



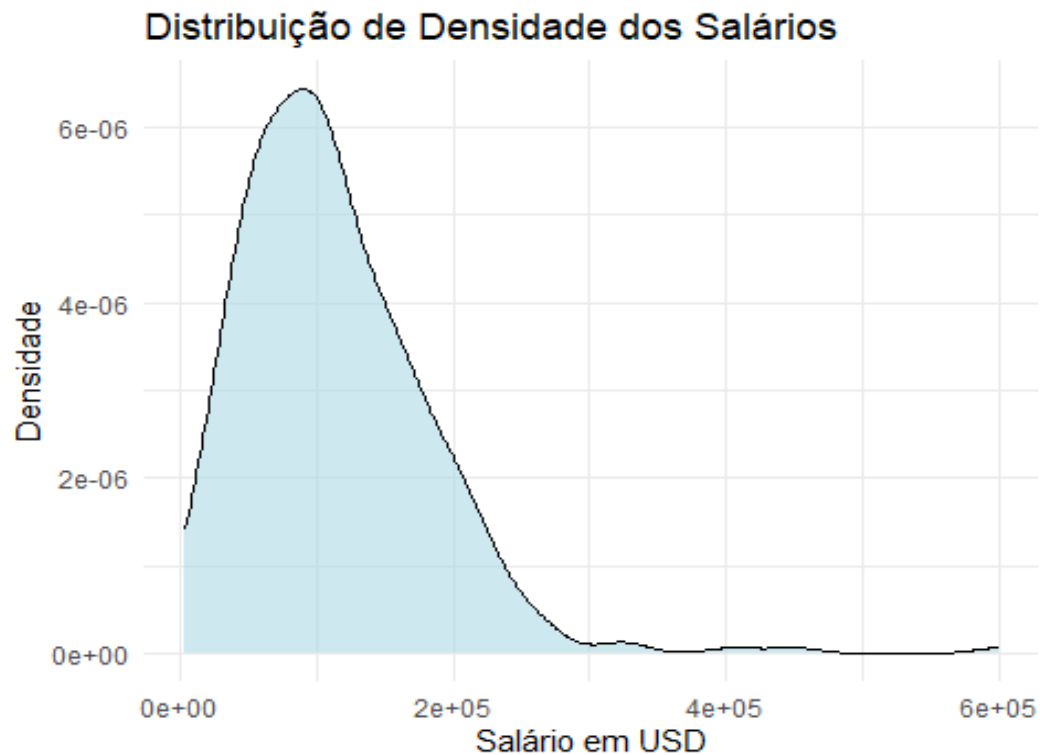
O gráfico mostra que a maioria dos profissionais de Data Science está nos níveis Sênior (SE) e Pleno (MI), sugerindo que o setor exige experiência. Há poucos profissionais nos níveis Executivo (EX) e Júnior (EN), o que pode indicar dificuldades para iniciantes e barreiras para alcançar cargos de liderança.

```
# 5
ggplot(base1, aes(x = experiencia, fill = trab_remoto)) +
  geom_bar(aes(y = (..count..) / sum(..count..) * 100), position = "dodge") +
  labs(title = "Relação entre Experiência e Trabalho Remoto",
       x = "Nível de Experiência",
       y = "Porcentagem (%)",
       fill = "Trabalho Remoto") +
  theme_minimal()
```



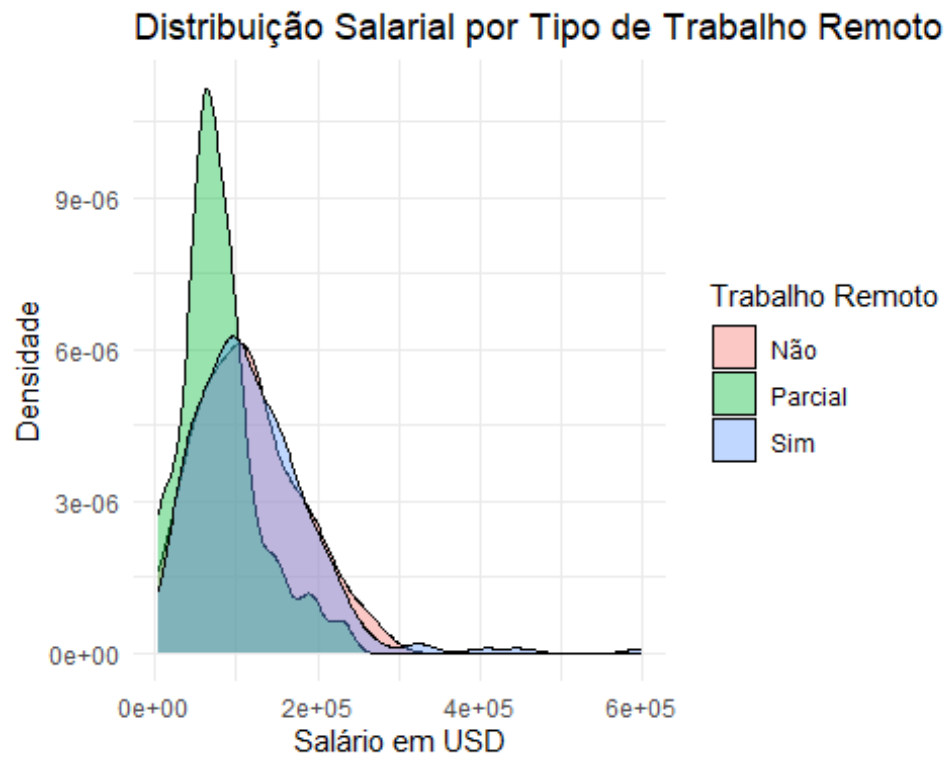
Profissionais mais experientes (SE e MI) trabalham mais remotamente, enquanto iniciantes (EN) e executivos (EX) têm menor adesão ao trabalho remoto. Isso sugere que a flexibilidade aumenta com a experiência, enquanto iniciantes podem ter mais trabalho presencial para aprendizado e integração.

```
# 6
ggplot(base1, aes(x = salario_USD)) +
  geom_density(fill = "lightblue", alpha = 0.6) +
  labs(title = "Distribuição de Densidade dos Salários",
       x = "Salário em USD",
       y = "Densidade") +
  theme_minimal()
```



A formação de um pico de densidade entre 0 e 200.000 USD sugere que a maioria dos trabalhadores na base de dados recebe salários dentro dessa faixa. Isso pode refletir uma concentração de salários medianos, com uma menor frequência de salários muito baixos ou muito altos.

```
# 7
ggplot(base1, aes(x = salario_USD, fill = trab_remoto)) +
  geom_density(alpha = 0.4) +
  labs(title = "Distribuição Salarial por Tipo de Trabalho Remoto",
       x = "Salário em USD",
       y = "Densidade",
       fill = "Trabalho Remoto") +
  theme_minimal()
```



O gráfico mostra que profissionais em regime híbrido (parcial) apresentam um pico de densidade salarial em uma faixa menor, mas sua distribuição se estende mais que a dos trabalhadores presenciais e remotos, indicando que há casos de salários mais elevados nesse grupo.