

# RISCO DE CRÉDITO

Apresentado por: Clevia Bento de Oliveira

# Introdução

Uma organização deseja prever quem são os possíveis inadimplentes no produto de empréstimos ao consumidor.

Eles têm dados sobre o comportamento histórico do cliente com base no que observaram. Portanto, quando adquirem novos clientes, eles desejam prever quem é mais arriscado e quem não é.



## Objetivo

Identificar qual perfil de clientes possui mais risco para a organização de serem inadimplentes.



## Metodologia

Será aplicado um teste de hipóteses Qui-quadrado para identificar associações entre a variável resposta e variáveis explicativas.

O conjunto de dados de treinamento servirá para identificar padrões de clientes inadimplentes.

Para esta análise o software utilizado será o RStudio.

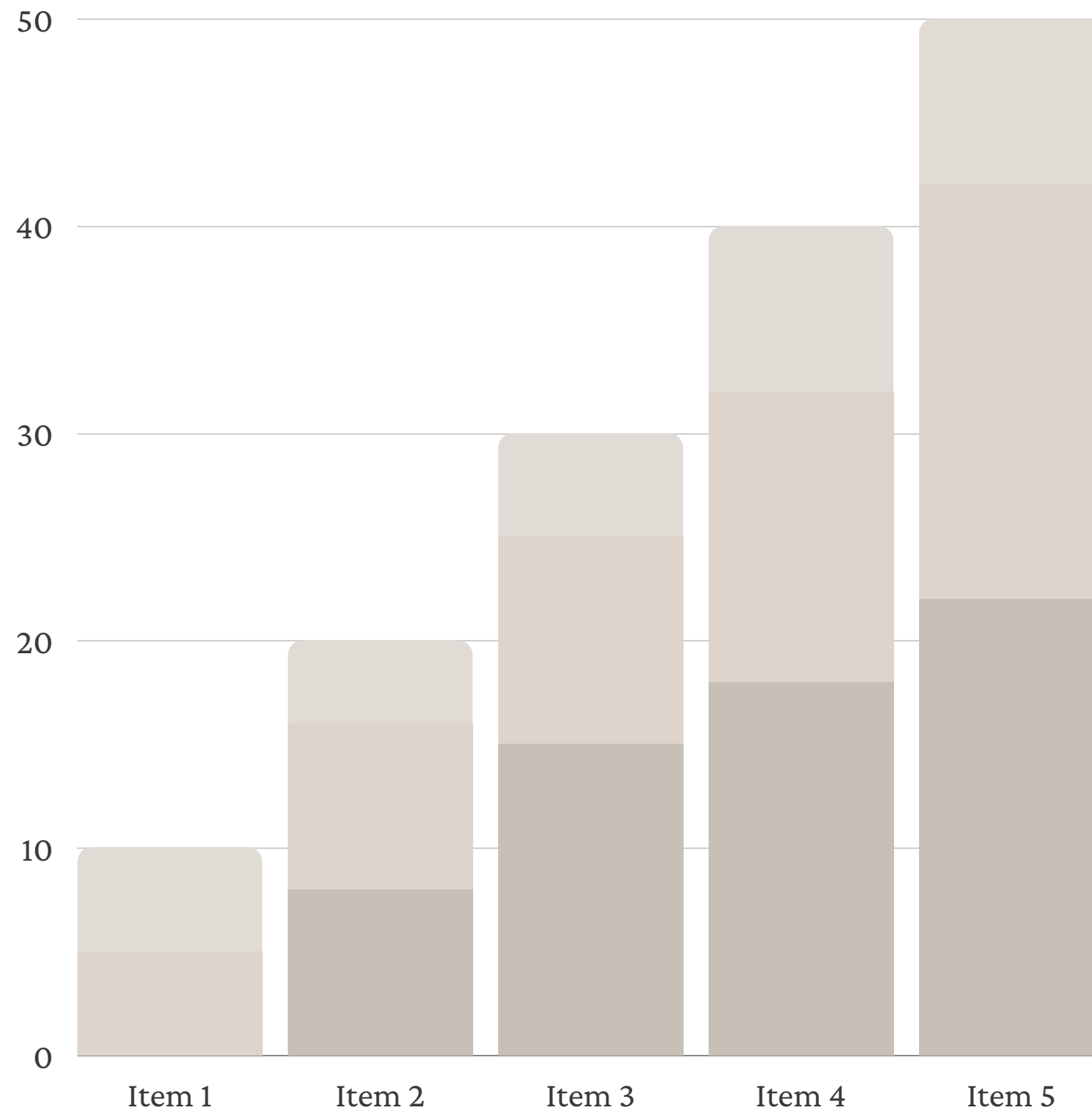
## O QUE É O TESTE QUI-QUADRADO?

O teste é utilizado para verificar se a frequência com que um determinado acontecimento é observado em uma amostra se desvia significativamente ou não da frequência com que ele é esperado.

## HIPÓTESES DO TESTE

Hipótese nula ( $H_0$ ) – frequências observadas = frequências esperadas. Não há associação entre os grupos (casualidade).

Hipótese alternativa ( $H_1$ ) – as frequências observadas  $\neq$  frequências esperadas. Os grupos estão associados.



# Estudo Descritivo das Variáveis

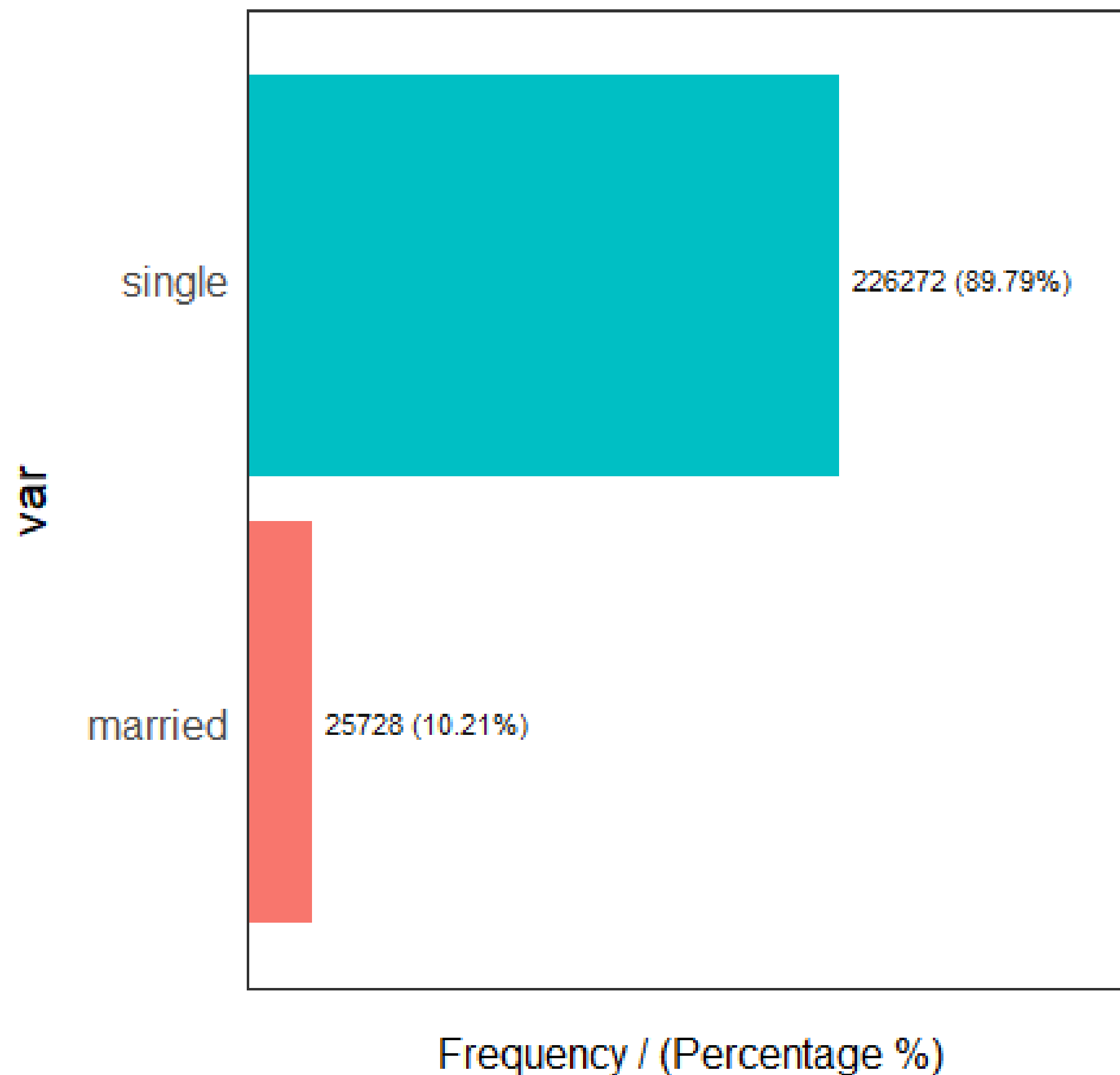
O banco de dados utilizado é composto de 13 variáveis, sendo elas:

<b>Id</b>	Se refere a cada cliente e tem um total de 252000 observações
<b>Income</b>	Se refere à renda dos clientes, possui o valor mínimo de 10310 e máximo de 9999938, e média 4997117
<b>Age</b>	Se refere às idades, varia de 21 a 79 anos com média de 50 anos.
<b>Experience</b>	Se refere a experiência. Com valor mínimo de 0 e máximo 20
<b>Married</b>	Se refere a estado civil, se divide em married(casado) e single(solteiro)
<b>House_ownership</b>	Referente possuir residência, se divide em casa, aluguel e não possuir casa nem aluguel.

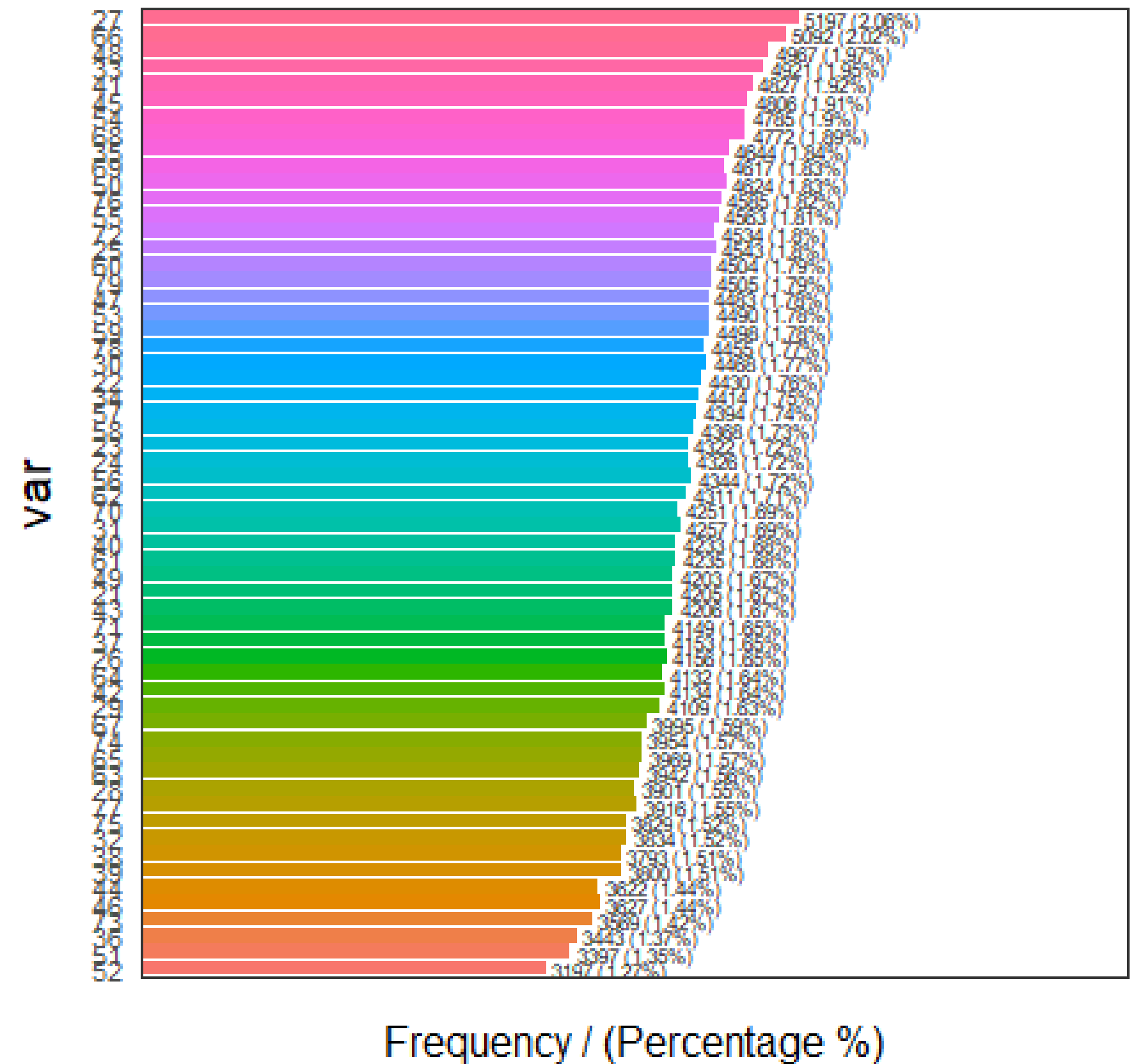
Características do Dataset	
<b>car_ownership</b>	Referente a possuir carro, se divide em sim ou não
<b>profession</b>	Referente a profissão dos clientes, com 51 profissões diferentes
<b>city</b>	Cidades em que moram os clientes, com 250 cidades diferentes
<b>state</b>	Estado em que os clientes moram. Possui 29 estados diferentes.
<b>current_job_years</b>	Referente a anos de trabalho. Com mínimo 0 e máximo 14, média 6.
<b>current_house_years</b>	Anos de casa, com variação de 10 a 14 e média 12
<b>risk_flag</b>	Se refere a bandeira de crédito. Está dividido entre 0 e 1, sendo 0 como adimplente e 1 como inadimplência.

# Frequência da variável Estado Civil

Podemos ver que 89,79%  
dos clientes são solteiros e  
19,21% são casados.

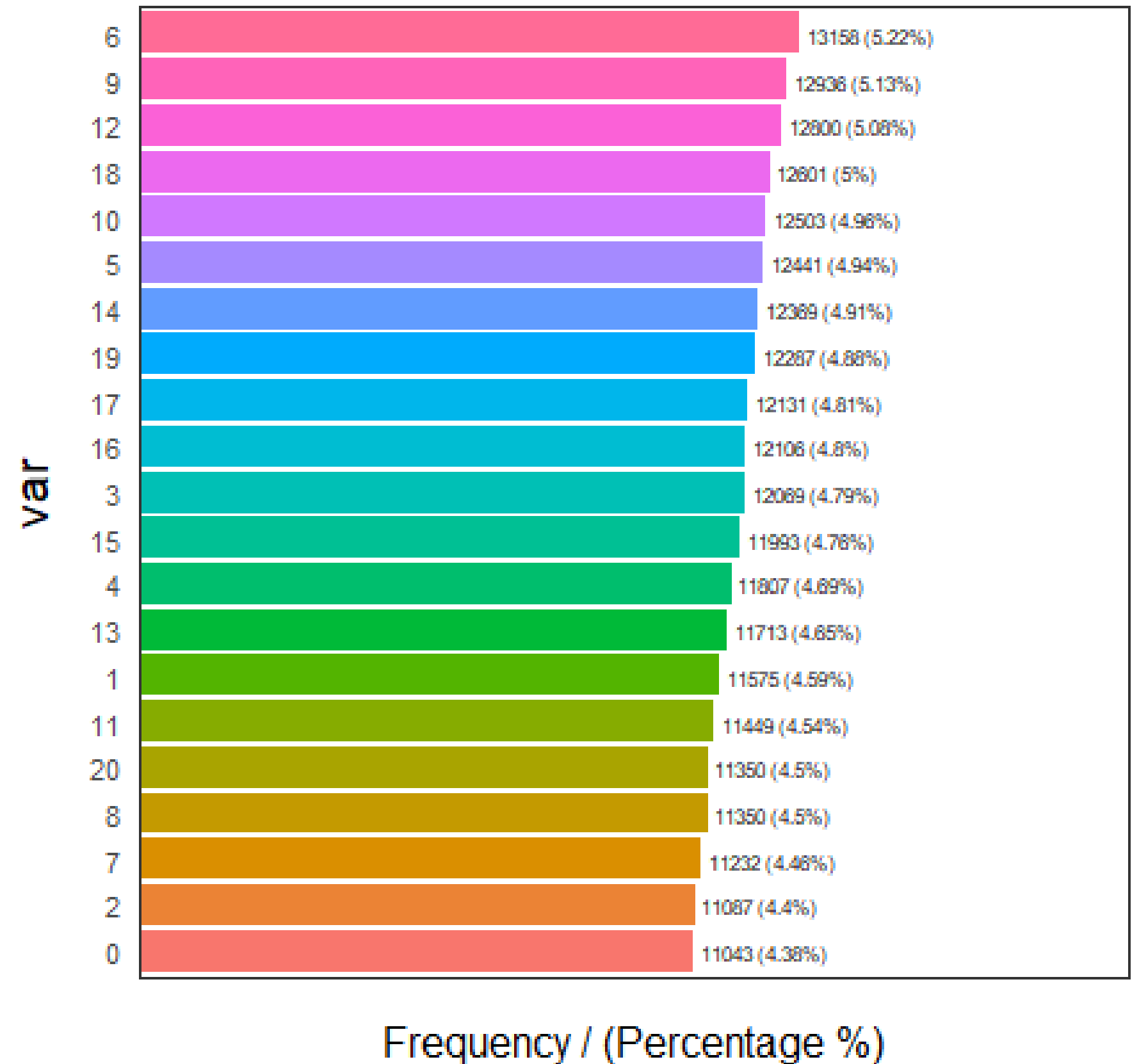


# Frequência da variável Idade



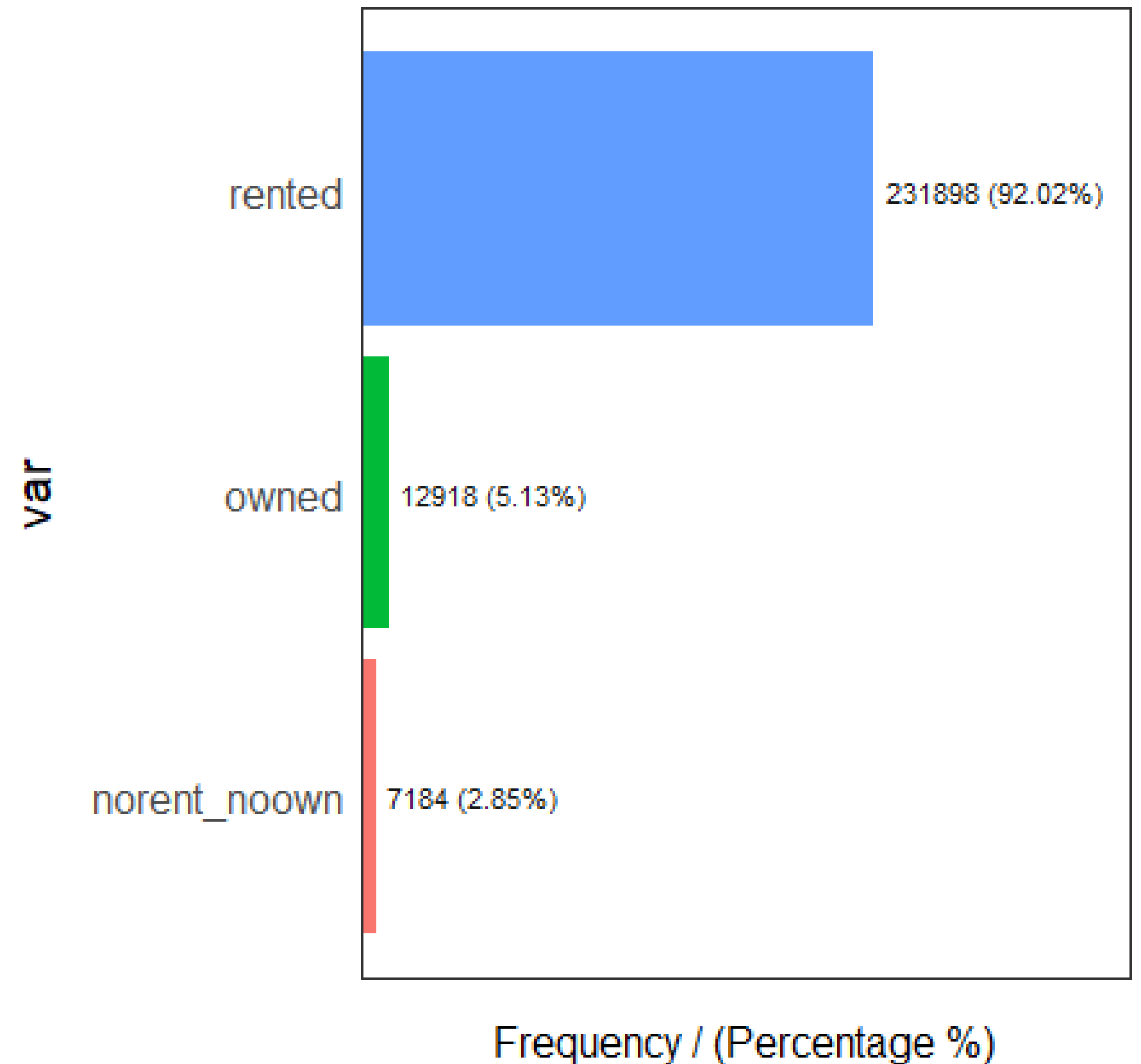


Frequência do  
tempo de  
experiência de  
cada cliente  
pertencente à  
organização.



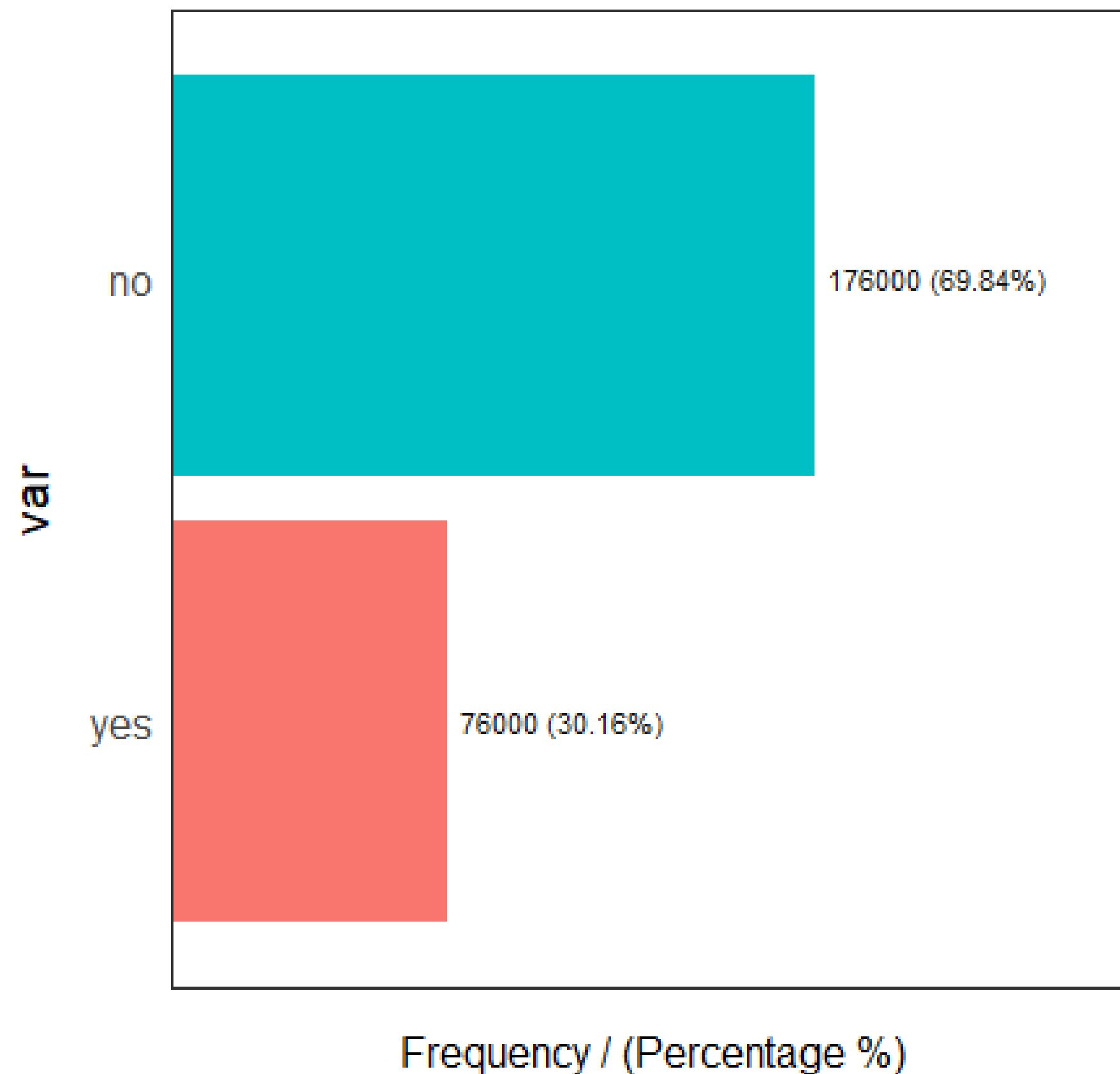
# Frequência dos clientes que possuem residência ou moram de aluguel

Podemos ver que aproximadamente  
92% dos clientes moram de aluguel



# Frequência de clientes que possuem carro ou não

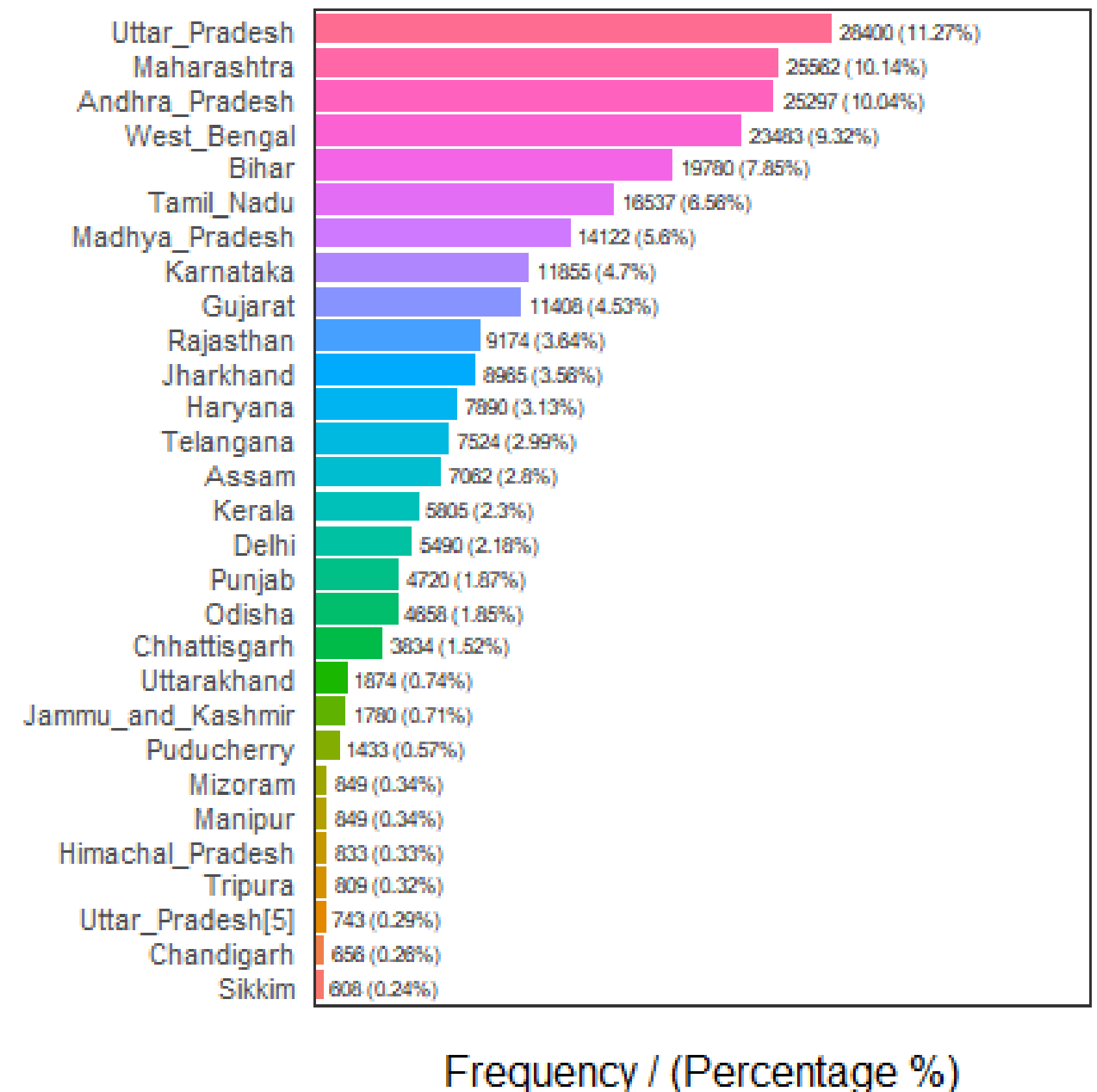
Vemos que  
aproximadamente 70% dos  
clientes não possuem carro



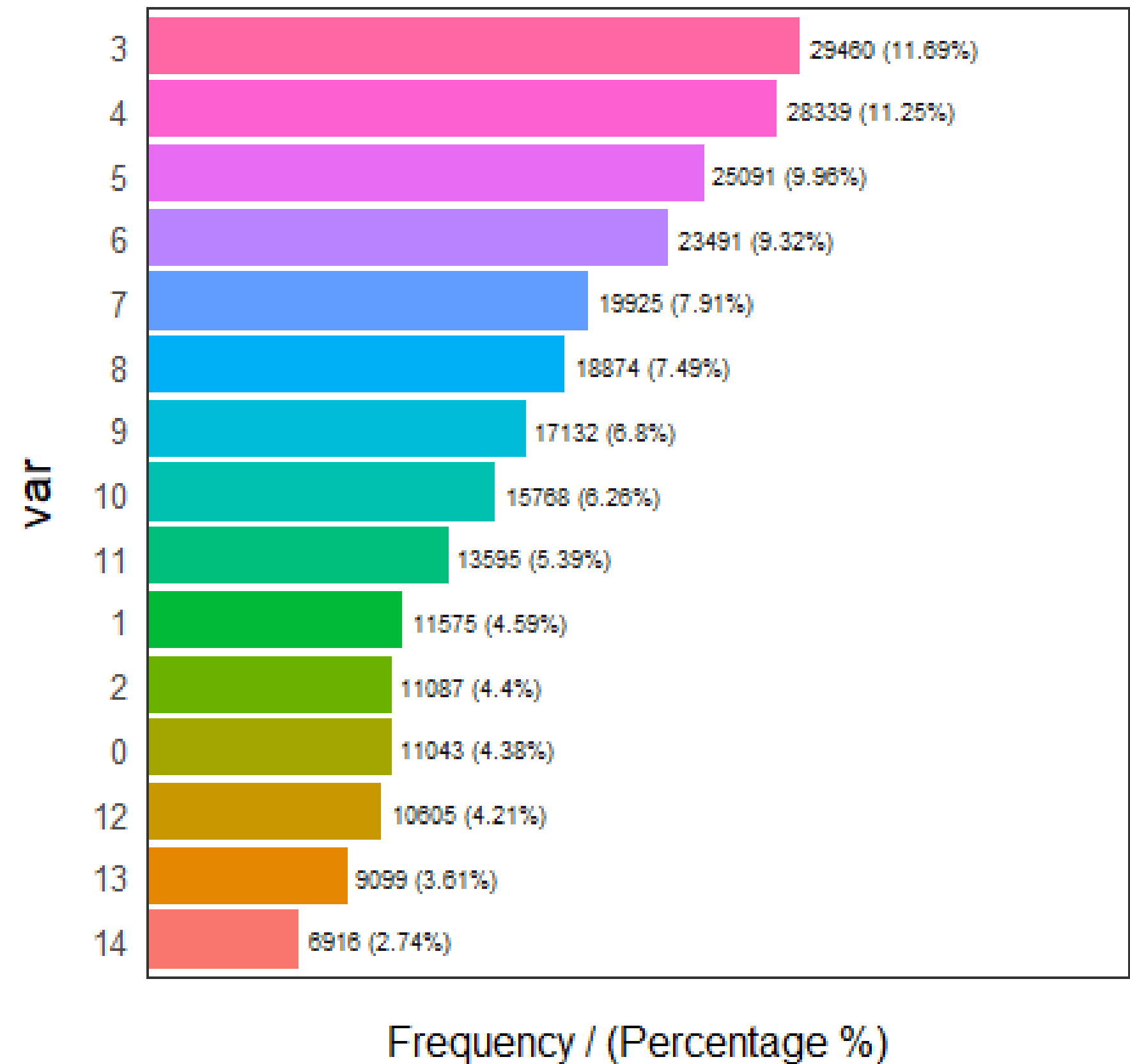
# Frequência de clientes por Estado.

Podemos ver que há mais clientes nos estados de Utta, Maharashtra e Andhra.

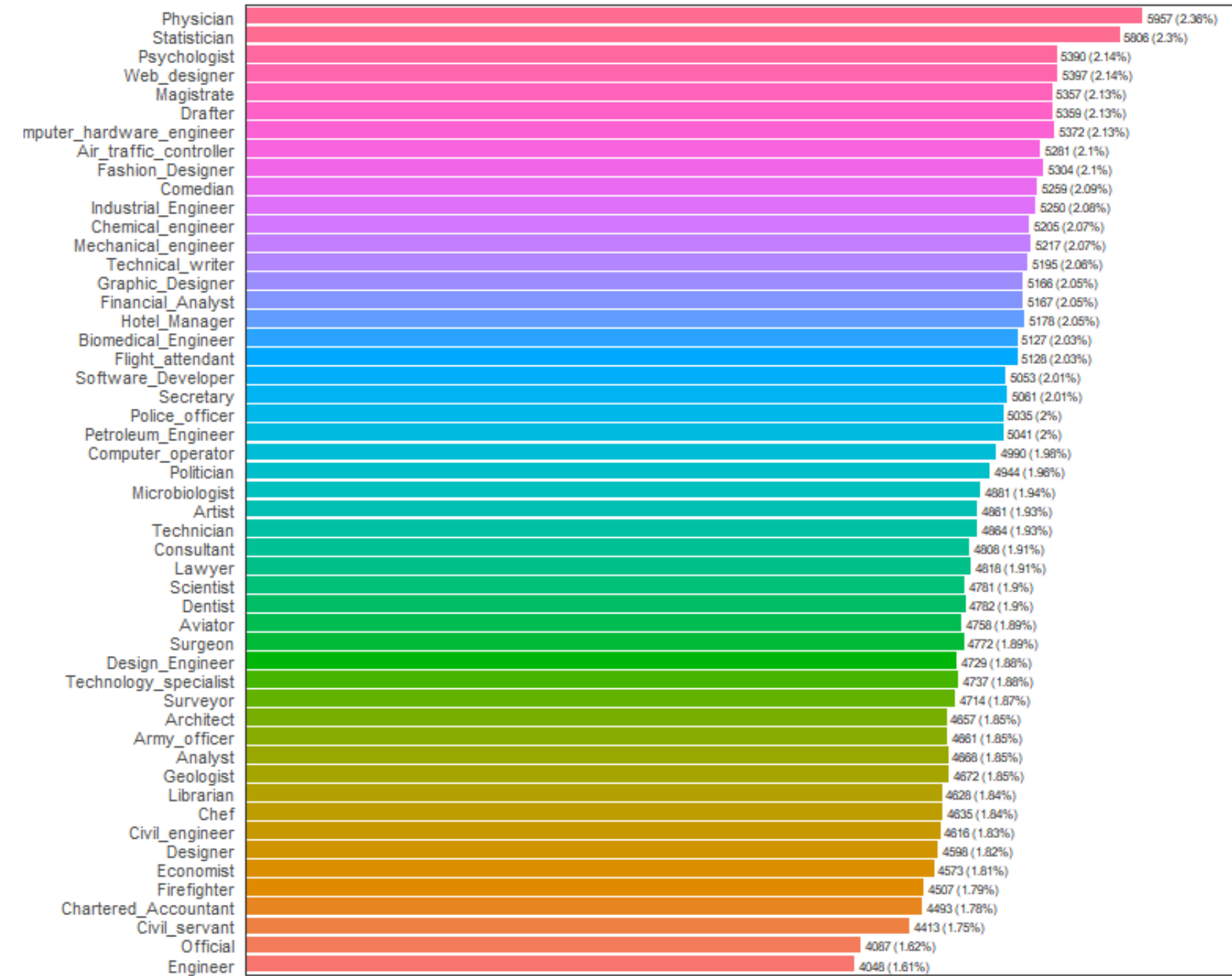
var



# Frequência dos anos de trabalho de cada cliente

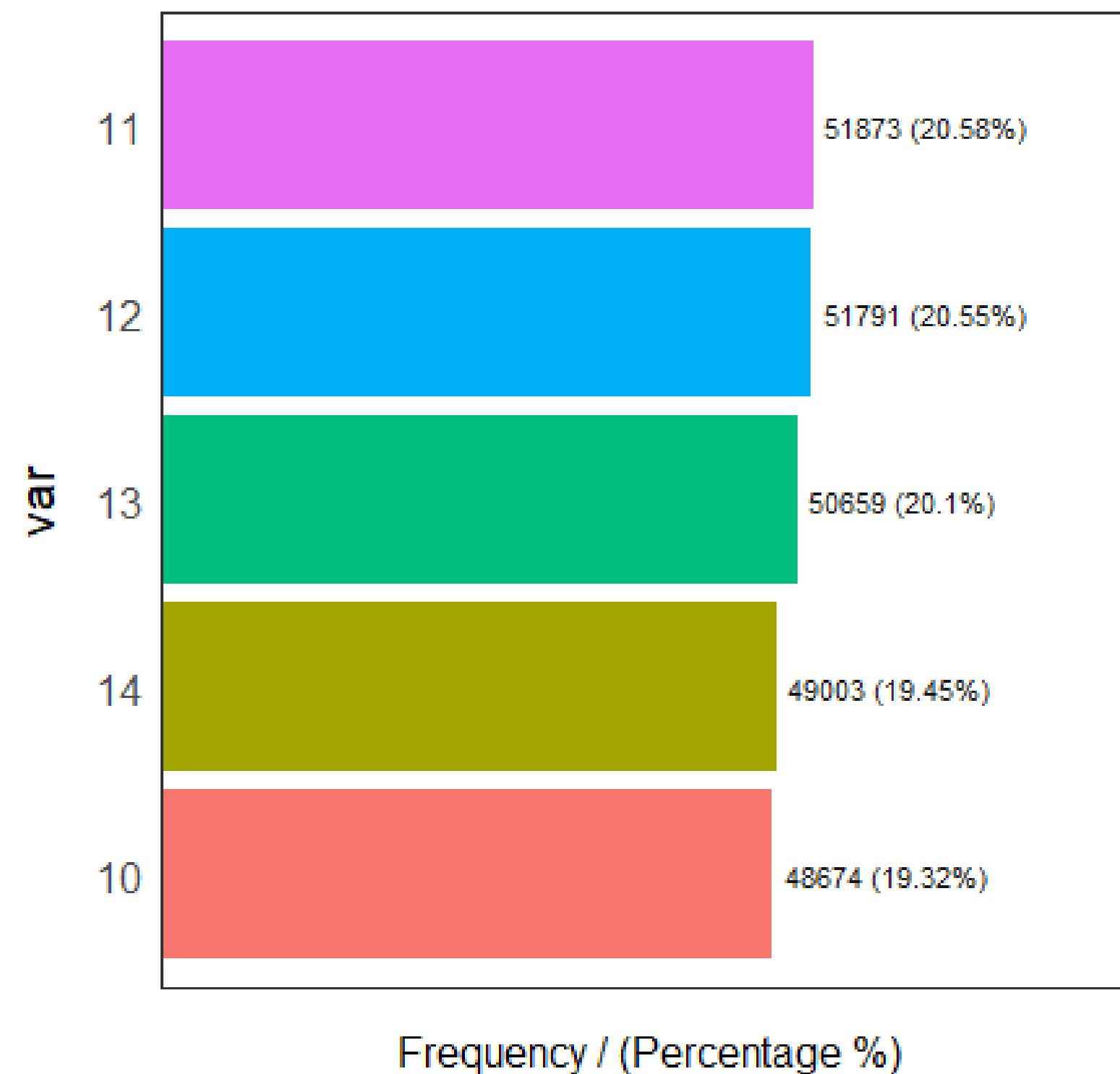


# Frequência de profissões dos clientes

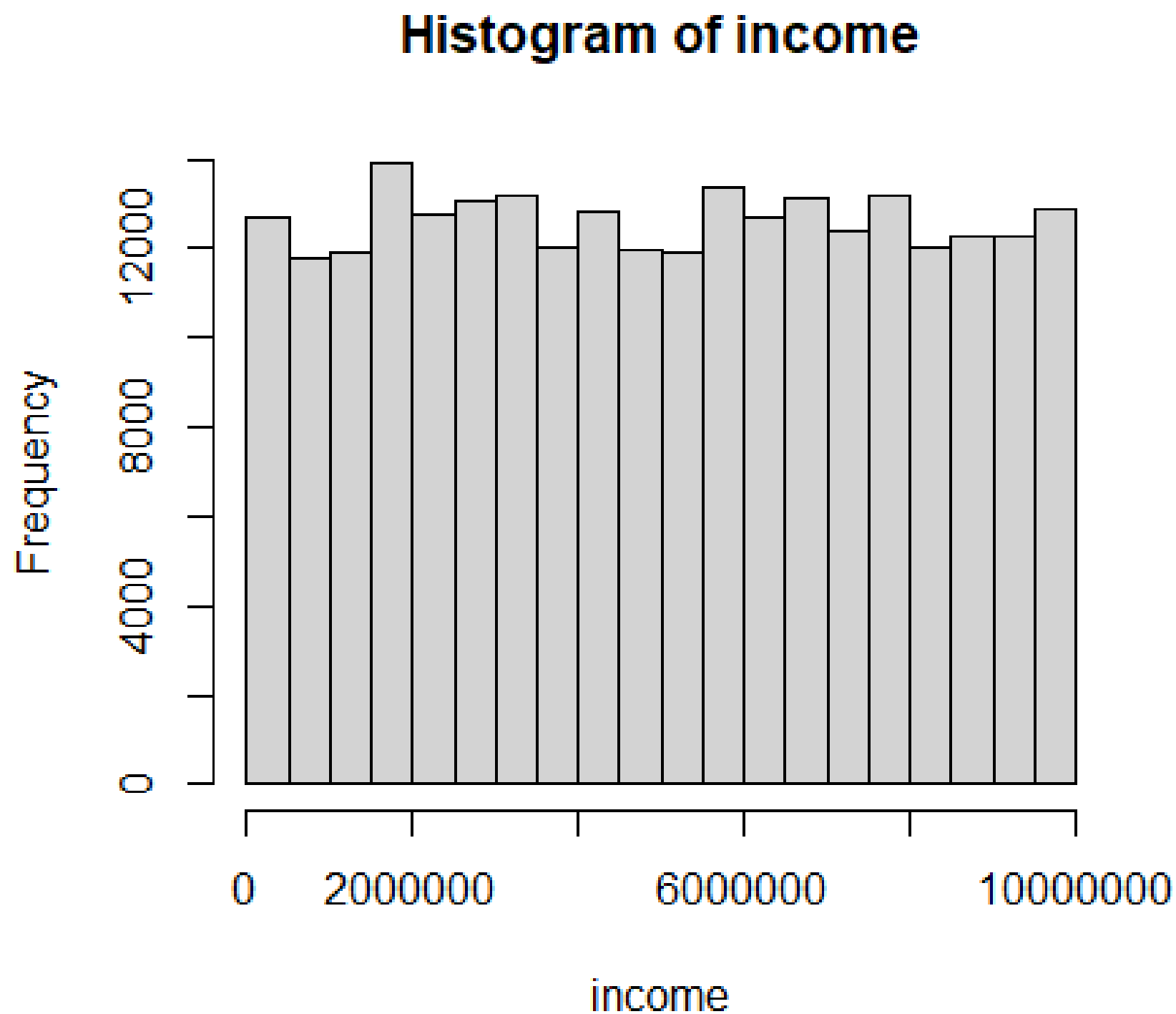


Frequency / (Percentage %)

# Frequência dos anos de casa.



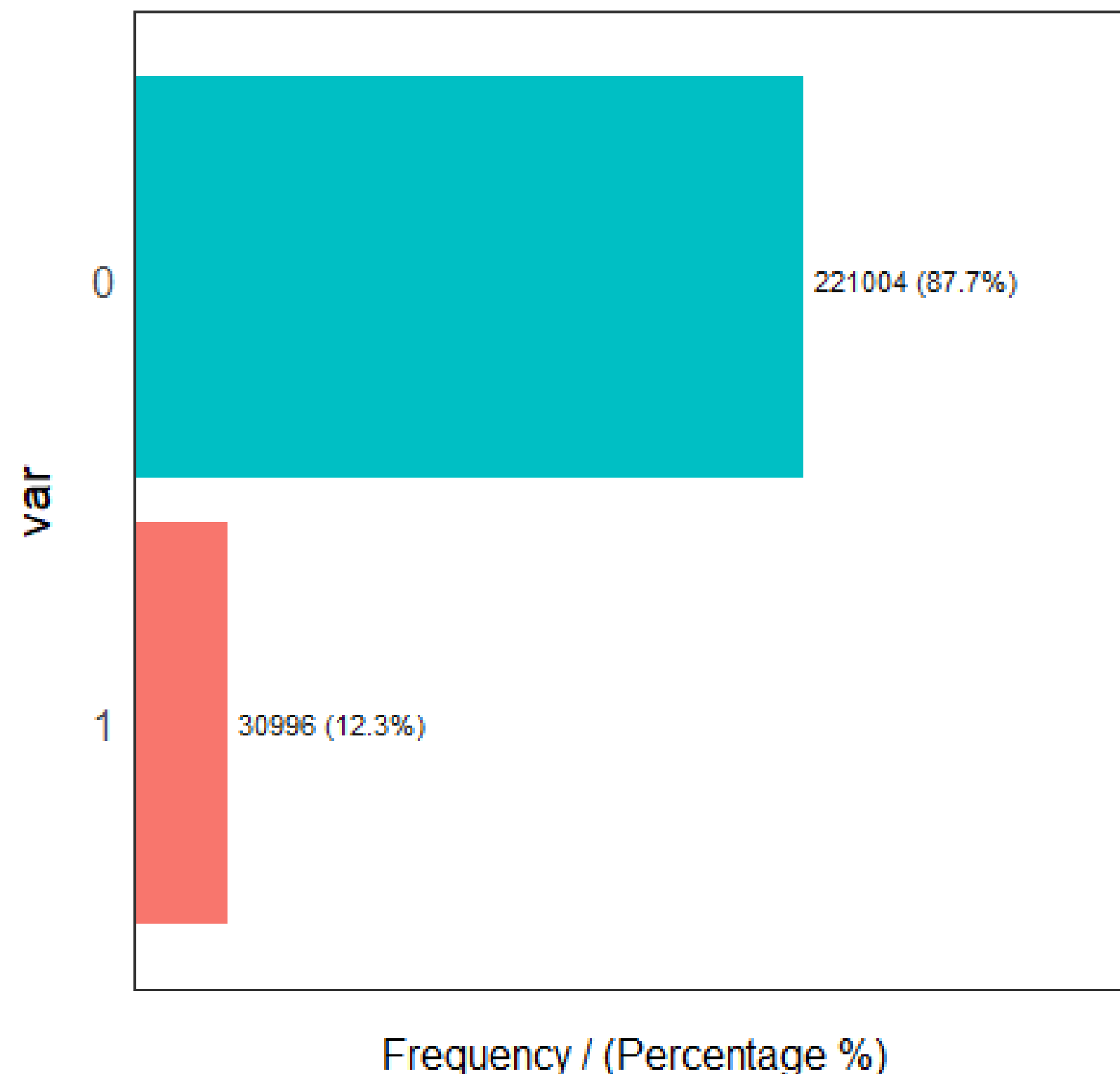
# Histograma dos dados de renda

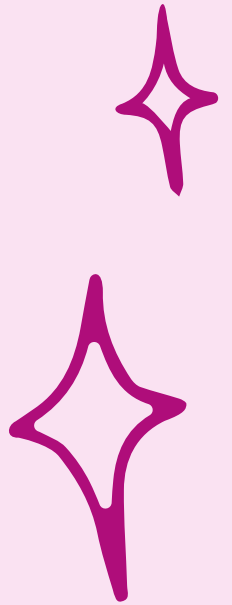




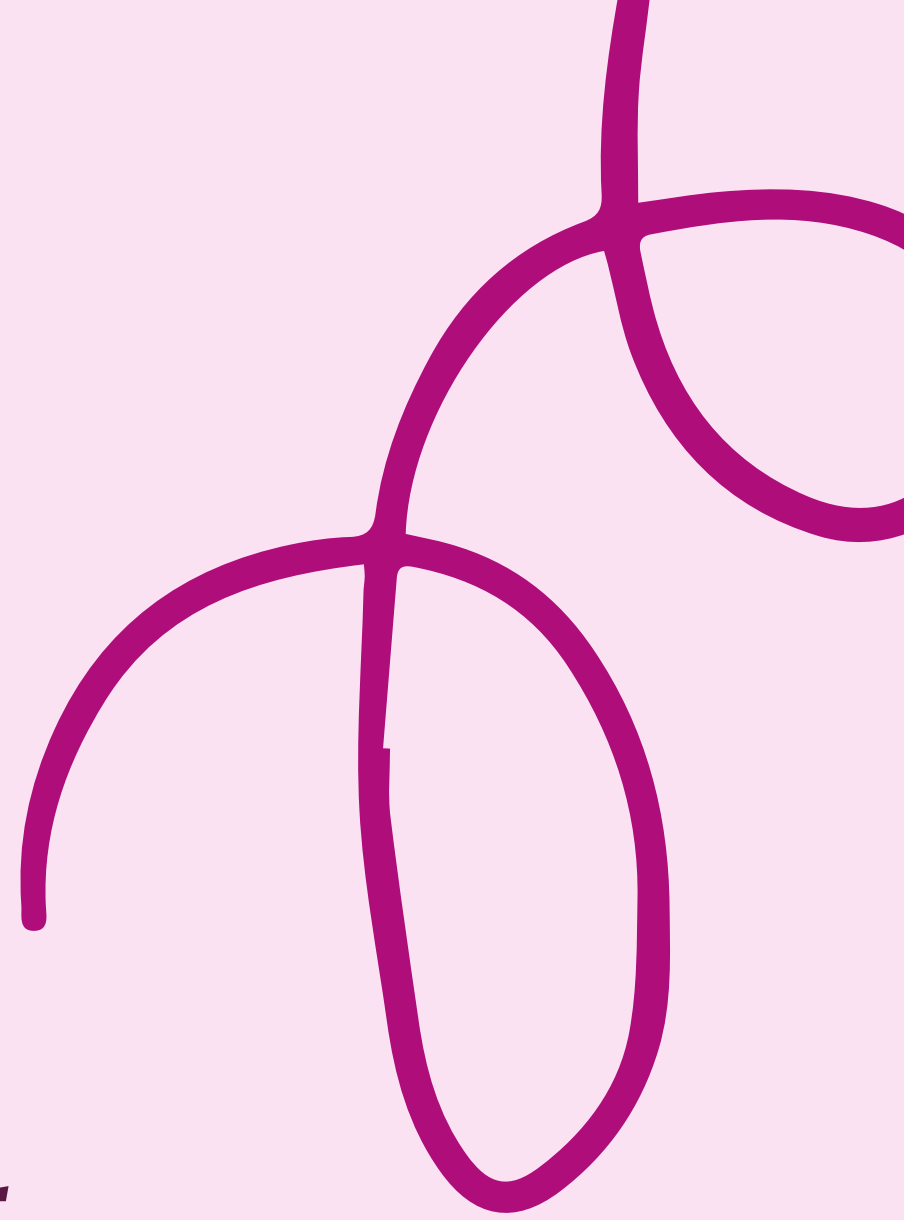
# Frequência dos clientes que são inadimplentes ou não

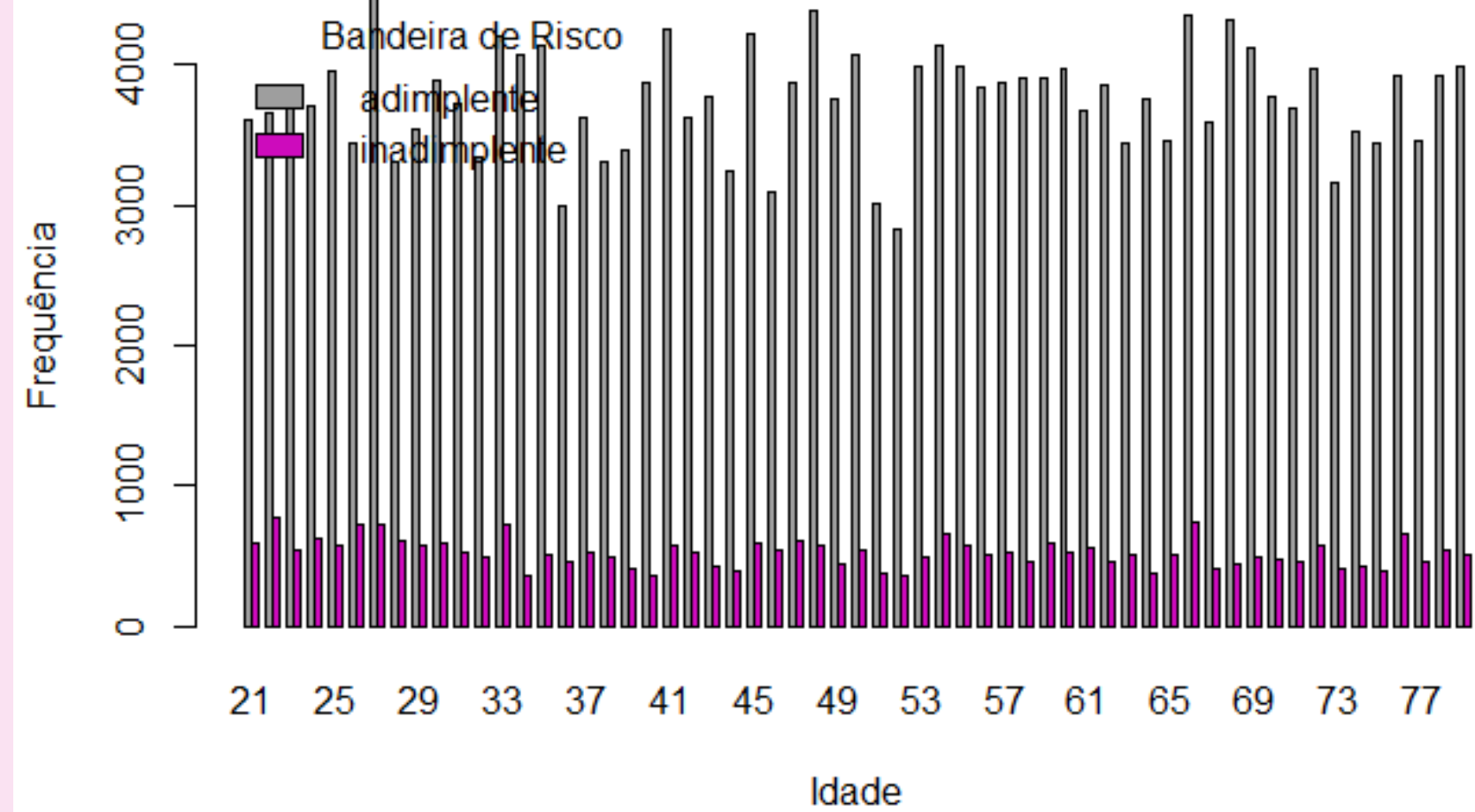
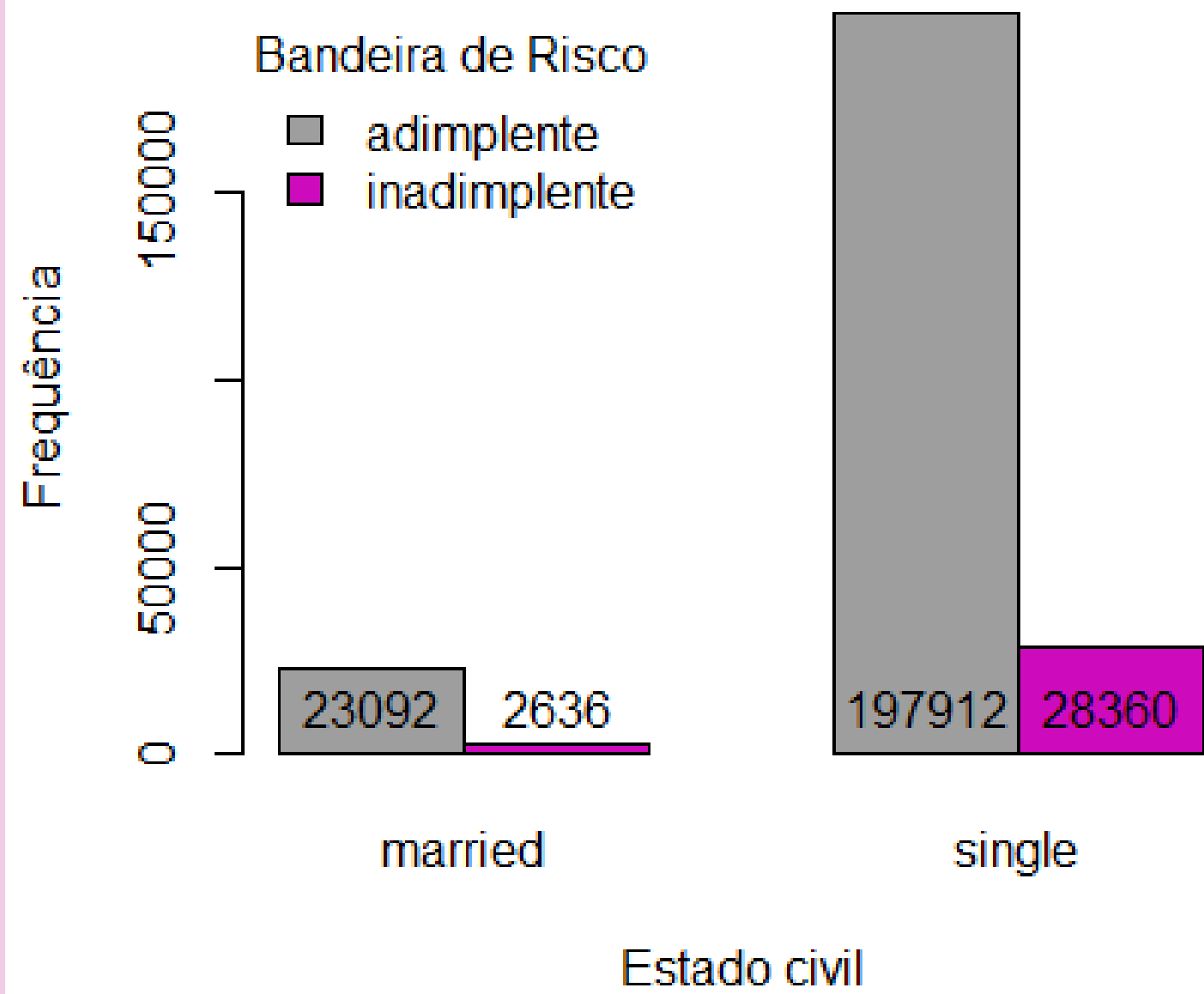
zero se refere aos clientes adimplentes, vemos que 87% aproximadamente são bons pagadores

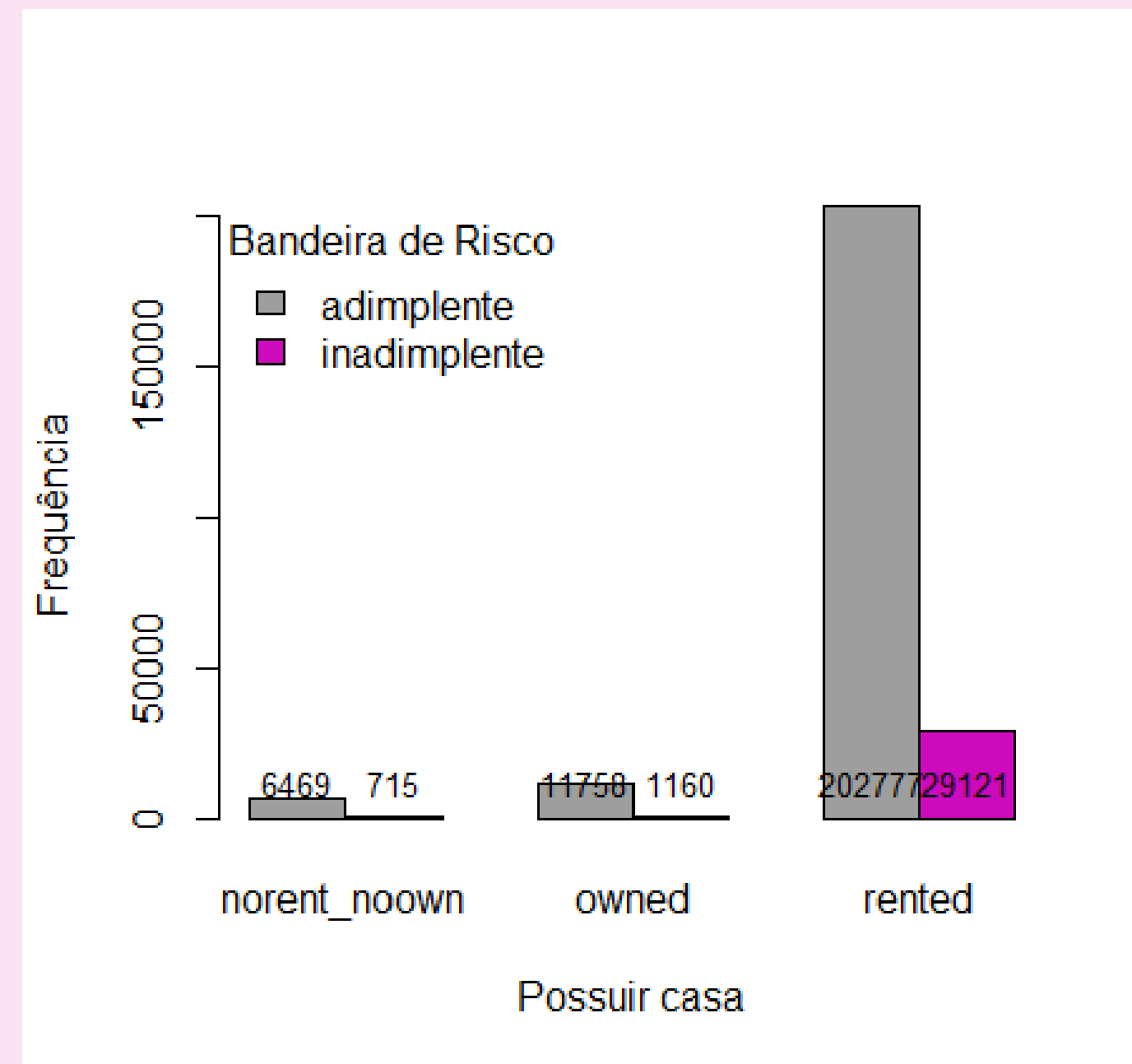
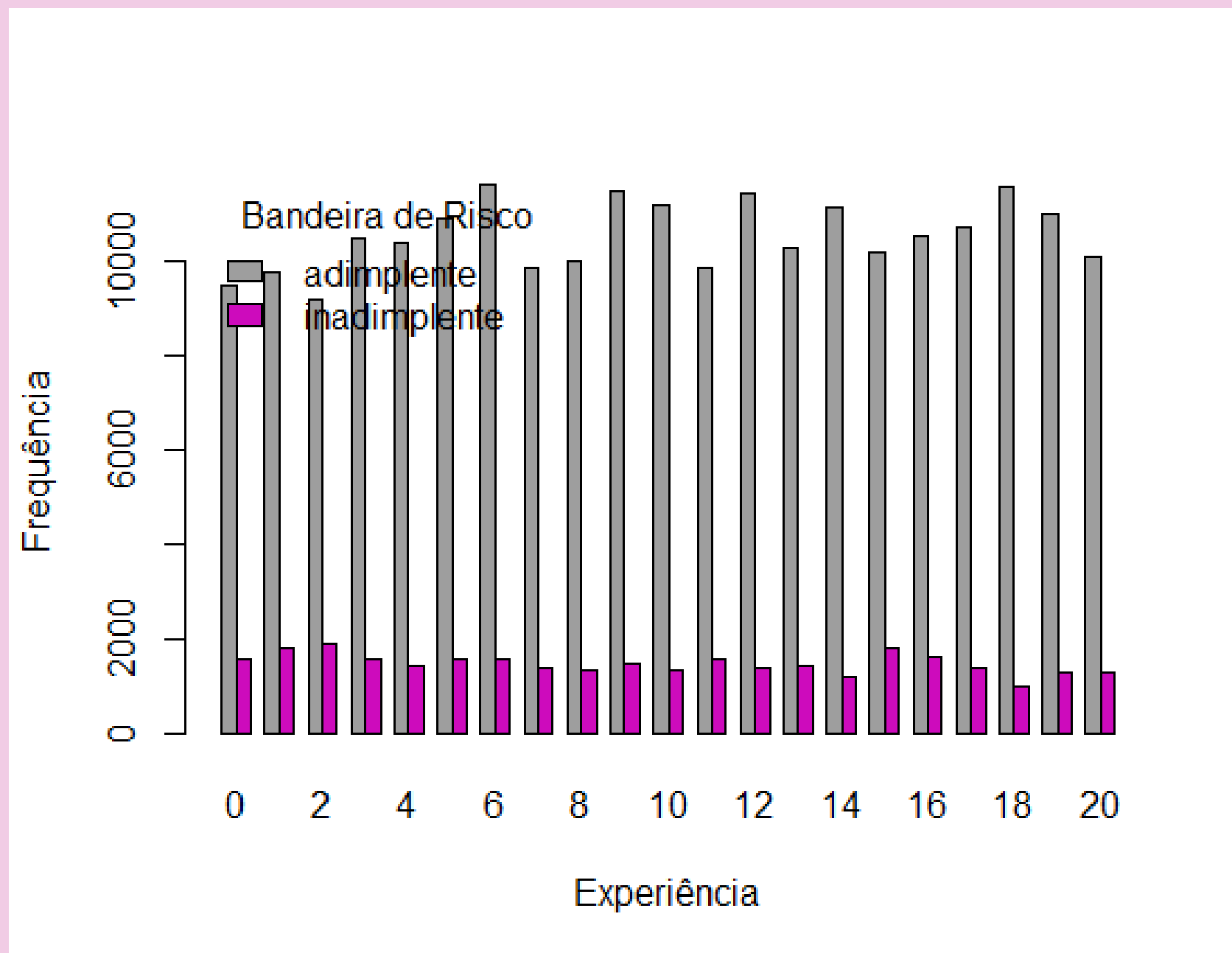


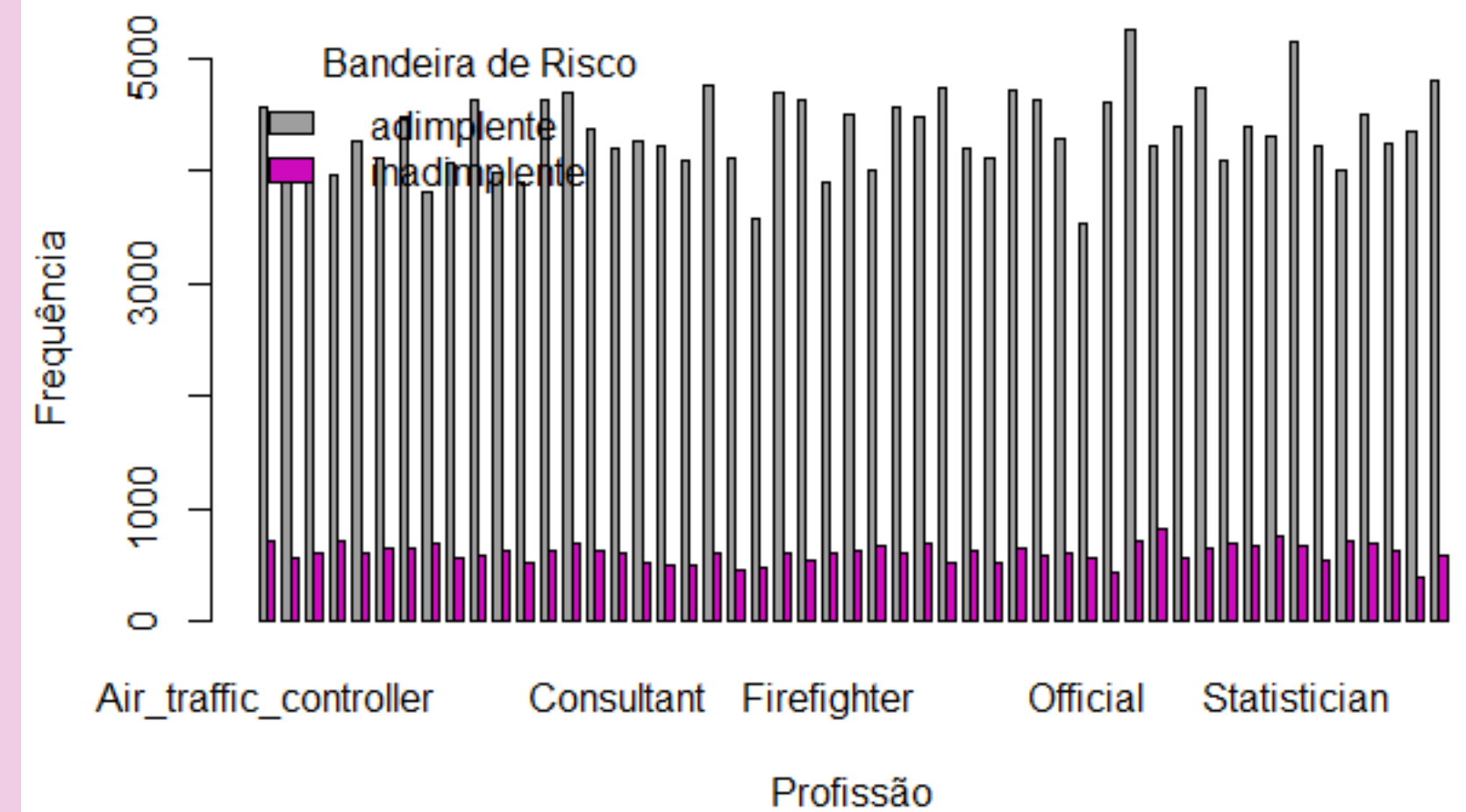
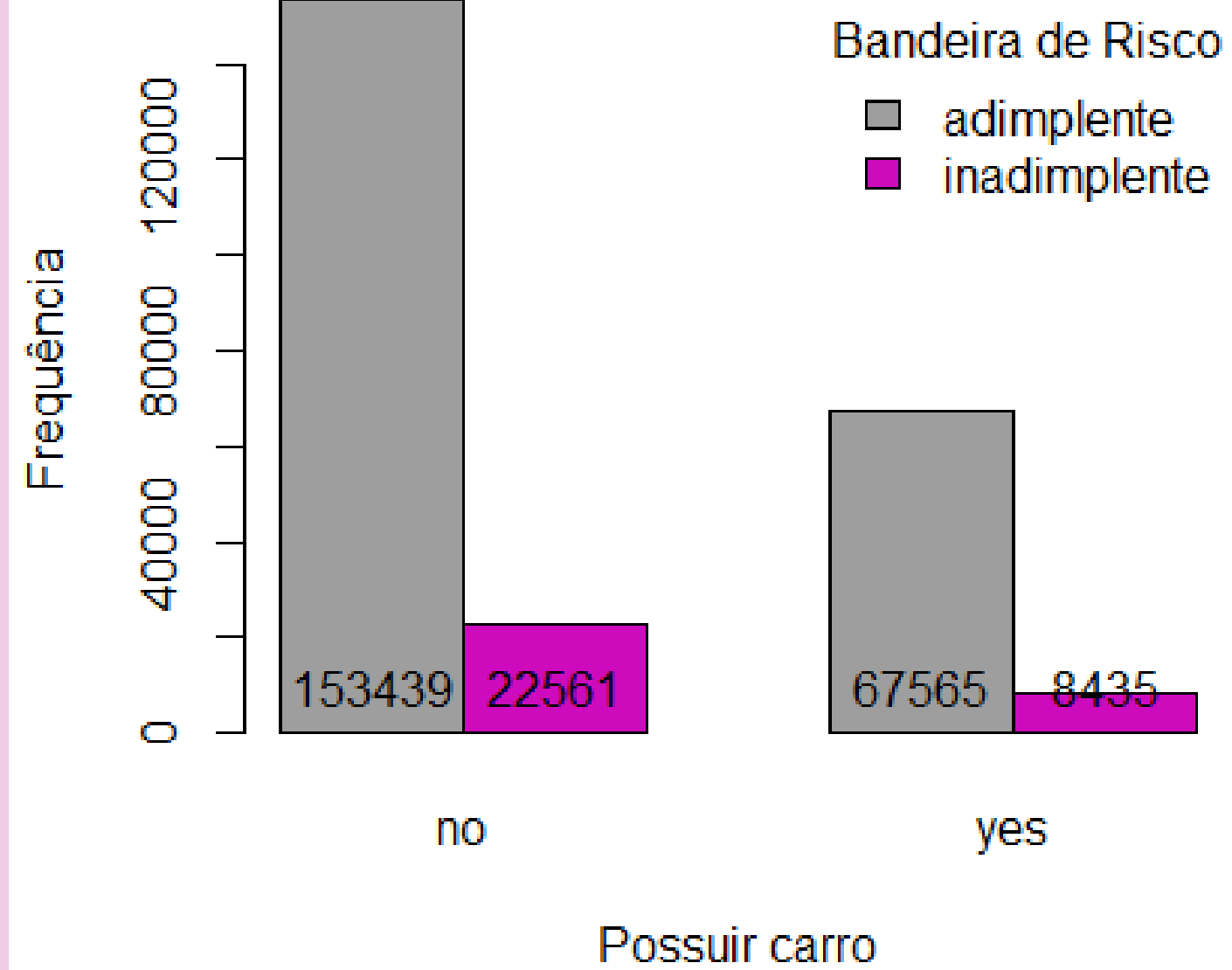


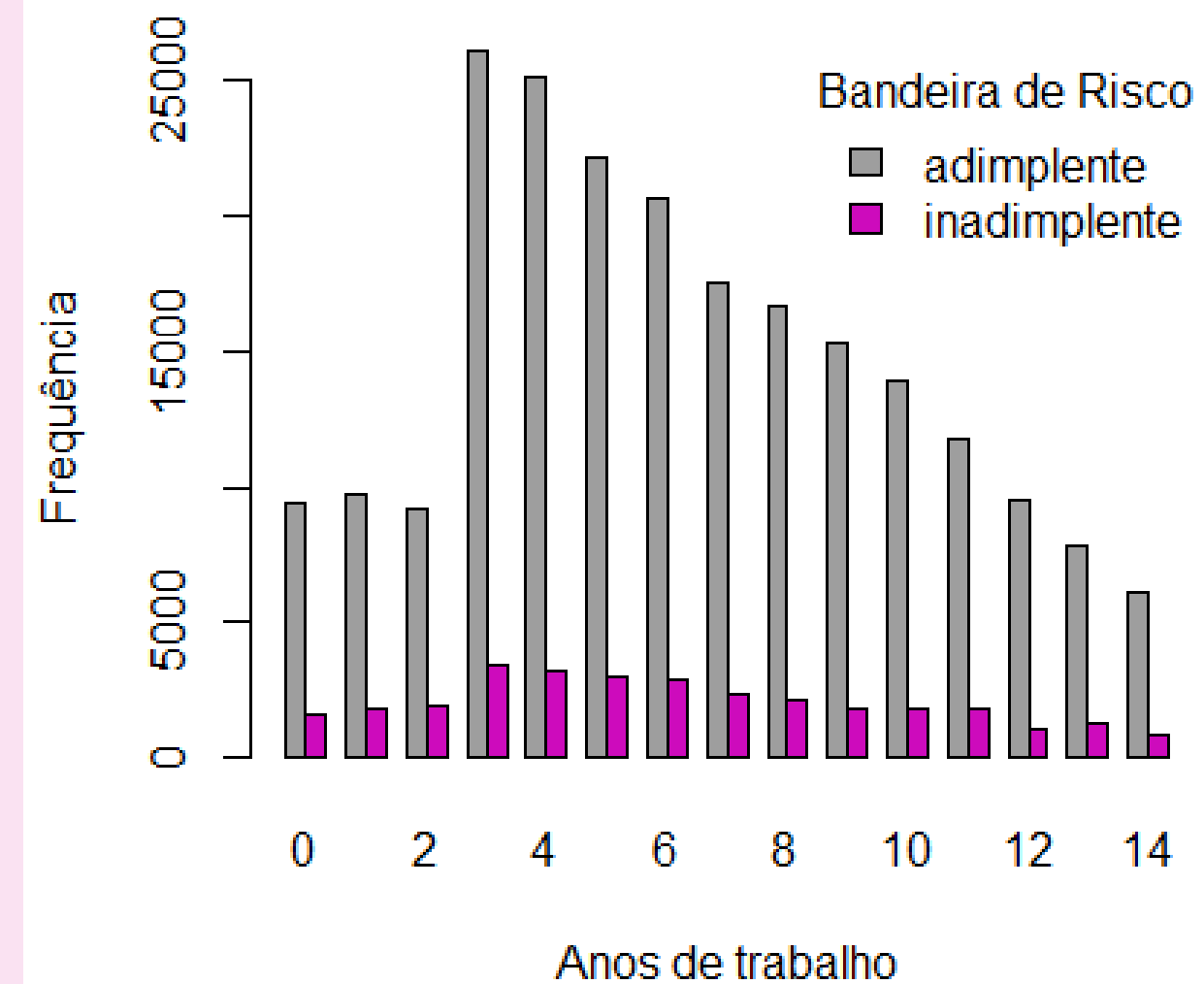
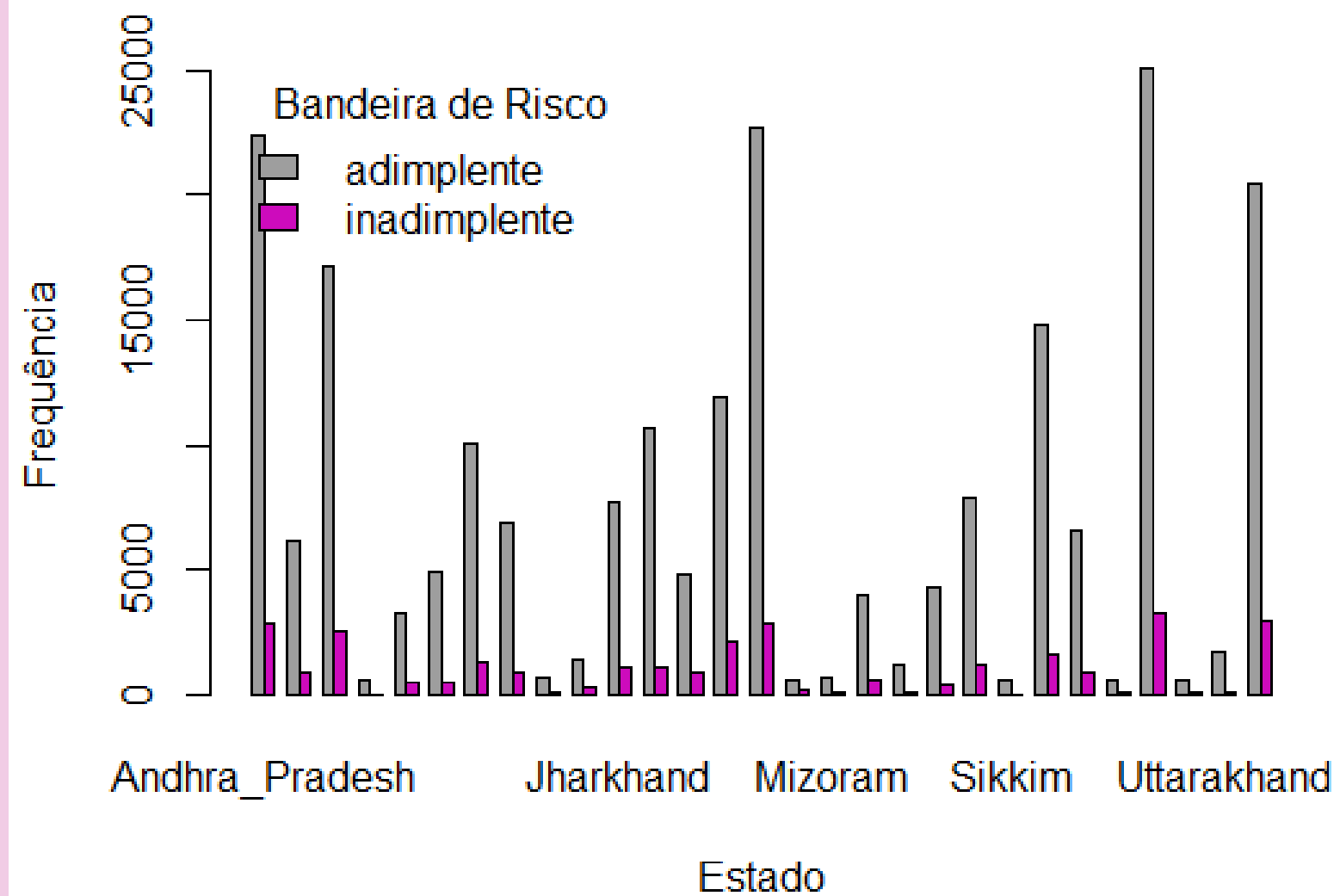
# FREQUÊNCIA DAS VARIÁVEIS EM RELAÇÃO A VARIÁVEL RESPOSTA

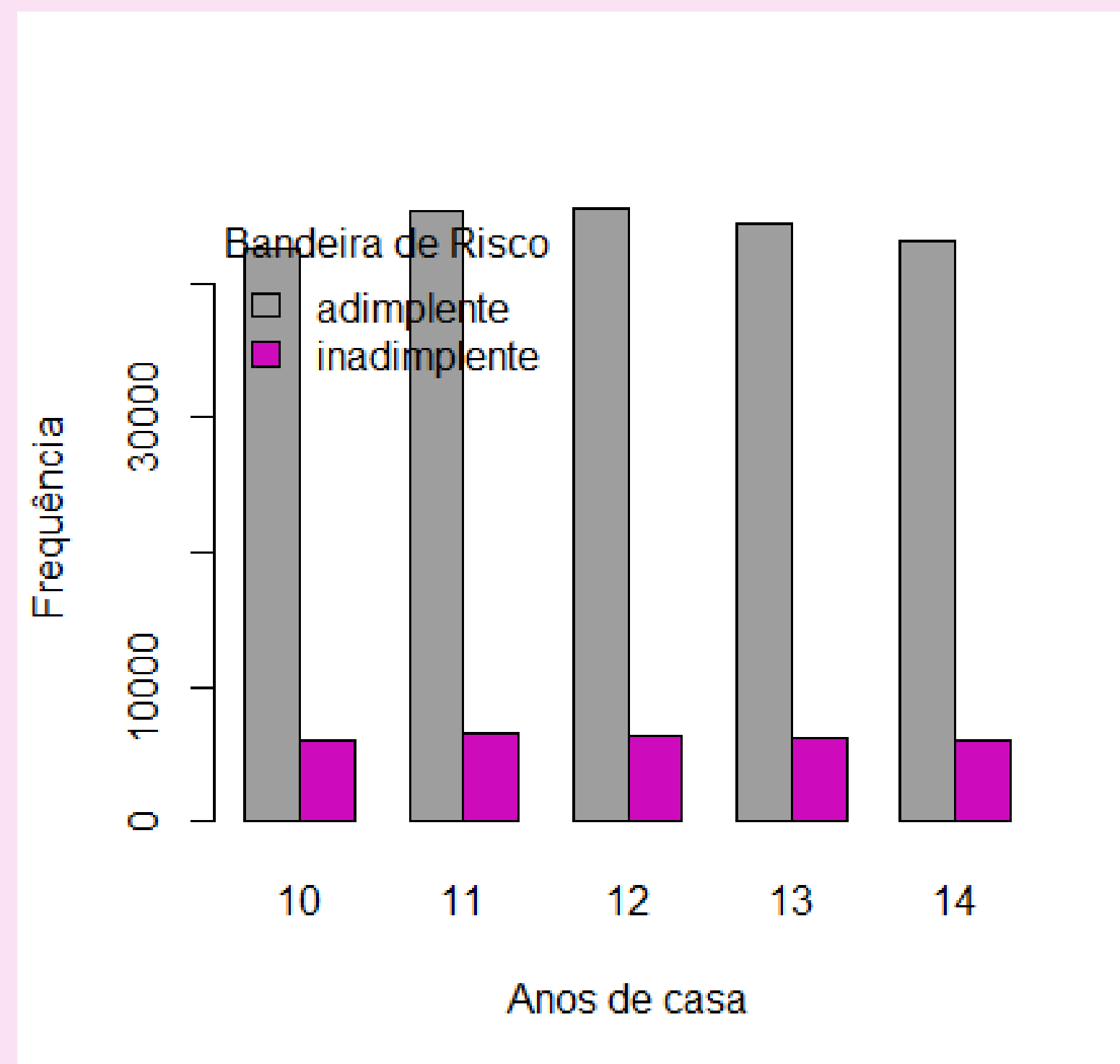












# TESTE QUI- QUADRADO

Nossa variável resposta é risk\_flag (bandeira de risco)



# Assumindo as Hipóteses como:



**H0:** as variáveis são independentes, ou seja, não influenciam na variável resposta.



**H1:** as variáveis são dependentes, ou seja, têm influência na variável resposta.

# Para estado civil o resultado foi:

$X\text{-squared} = 111.89, df = 1, p\text{-value} < 0.00000000000000000022$

Assim, rejeitamos  $H_0$  e concluimos que o estado civil influencia ao risco de crédito

# Para idade o resultado do teste foi:

$X\text{-squared} = 776.37, df = 58, p\text{-value} < 0.00000000000000000022$

Podemos concluir que a idade influencia no risco de crédito.

# Para a variável possuir casa ou aluguel temos:

$X\text{-squared} = 182.99, df = 2, p\text{-value} < 0.00000000000000000022$

Concluimos que morar de aluguel ou possuir moradia influencia no risco de crédito.

# Para a variável estado o resultado foi:

$X\text{-squared} = 733.4, df = 28, p\text{-value} < 0.00000000000000000022$

Podemos concluir que o estado em que o cliente mora influencia no risco de crédito.

# Para a variável anos de trabalho temos:

$X\text{-squared} = 733.4$ ,  $df = 28$ ,  $p\text{-value} < 0.00000000000000000022$

Concluimos que os anos de trabalho possuem influência no risco de crédito.

# Para a variável anos de casa temos:

$X\text{-squared} = 6.115, df = 4, p\text{-value} = 0.1907$

Podemos concluir que esta variável não influencia no risco de crédito.

# Para a variável experiência temos:

$X\text{-squared} = 915.32, df = 20, p\text{-value} < 0.00000000000000000022$

Podemos concluir que esta variável influencia no risco de crédito.



# Para a variável possuir carro temos:

$X\text{-squared} = 145.42, df = 1, p\text{-value} < 0.00000000000000000022$

Concluimos que possuir carro influencia no risco de crédito.

# Para a variável profissão temos:

$X\text{-squared} = 613.49, df = 50, p\text{-value} < 0.00000000000000000022$

Concluimos que profissão influencia no risco de crédito.

# Para a variável cidade o resultado foi:

$X\text{-squared} = 5223.1, df = 316, p\text{-value} < 0.00000000000000000022$

Podemos concluir que a cidade em que o cliente mora influencia no risco de crédito.

# Para a variável renda o resultado foi:

X-squared = 162006, df = 41916, p-value < 0.00000000000000000022

Podemos concluir que a renda de cada cliente influencia no risco de crédito

# CONCLUSÃO

Concluimos que pelo teste Qui-Quadrado as variáveis, estado civil, idade, possuir casa, estado em que mora, anos de trabalho, experiência, renda, possuir carro, profissão e cidade influenciam no risco de crédito.