

# Padronização da ESTAT

**Consultores Responsáveis:**

Tepe

Estatiano Desconhecido 1

Estatiano Desconhecido 2

Estatiano Desconhecido 2

...

Estatiano Desconhecido n

**Requerente:**

ESTAT

Brasília, 27 de agosto de 2024.



# Sumário

	Página
1 Introdução . . . . .	4
2 Formatações gerais . . . . .	5
2.1 Fontes . . . . .	5
2.2 Margens e disposições . . . . .	5
2.3 Números . . . . .	5
3 Cores . . . . .	7
3.1 Códigos relacionados . . . . .	7
4 Modelo de relatório . . . . .	9
4.1 Introdução . . . . .	9
4.2 Metodologia . . . . .	9
4.3 Análises . . . . .	9
4.3.1 Introdução das análises . . . . .	9
4.3.2 Separação das seções . . . . .	9
4.3.3 Geral . . . . .	10
4.3.4 Descritiva . . . . .	10
4.3.5 Testes de hipóteses . . . . .	10
4.4 Conclusão . . . . .	11
5 Padronização de Gráficos . . . . .	12
5.1 Introdução . . . . .	12
5.2 Contextualização . . . . .	12
5.3 Sobre tabelas e frequências . . . . .	14
5.4 Sobre ordenação nos gráficos . . . . .	14
5.5 Gráfico de Colunas/Barras . . . . .	14
5.5.1 Colunas univariado com frequências . . . . .	14
5.5.2 Colunas bivariado com frequências . . . . .	16
5.5.3 Observações . . . . .	17
5.6 Gráfico de Setor . . . . .	17
5.7 Boxplot . . . . .	18
5.7.1 Univariado . . . . .	18
5.7.2 Bivariado . . . . .	19
5.8 Histograma . . . . .	20
5.8.1 Univariado . . . . .	20
5.8.2 Univariado com porcentagem . . . . .	21
5.8.3 Bivariado com facet grid . . . . .	22
5.8.4 Bivariado com facet wrap . . . . .	23
5.9 Gráfico de Dispersão . . . . .	24

5.9.1	Bivariado . . . . .	24
5.9.2	Multivariado . . . . .	25
5.10	QQ Plot . . . . .	26
5.11	Gráfico de Linhas . . . . .	27
5.11.1	Univariado . . . . .	28
5.11.2	Bivariado . . . . .	29
5.11.3	Observação . . . . .	29
6	Padronização de $\text{\LaTeX}$ . . . . .	30
6.1	Tabelas . . . . .	30
6.1.1	Formatações gerais . . . . .	30
6.1.2	Tabela simples . . . . .	30
6.1.3	Tabela de dupla entrada . . . . .	31
6.2	Quadros . . . . .	32
6.2.1	Quadro de Medidas Resumo . . . . .	32
6.2.2	Quadros para testes . . . . .	35
6.2.3	Quadro de intervalos de confiança . . . . .	36
6.2.4	Quadro de coeficientes . . . . .	36
6.2.5	Observação . . . . .	36
6.3	Hipóteses dos testes . . . . .	36
6.4	Sequência dos gráficos e tabelas/quadros . . . . .	37
6.5	Casas decimais . . . . .	37
6.6	Dicas . . . . .	38
6.6.1	Comando <b>label{ }</b> . . . . .	38
6.6.2	Demais tabelas e quadros . . . . .	38
6.6.3	Aspas duplas . . . . .	38

# 1 Introdução

Este documento tem como objetivo padronizar modelos de projetos, fontes, cores, formatações, nomenclaturas, códigos e demais produções da **ESTAT**.

Também visa aprimorar as produções da **ESTAT**, facilitando sua execução e melhorando a qualidade das mesmas.

A qualquer momento tópicos que não foram contemplados podem ser adicionados a este documento por qualquer membro efetivo da **ESTAT** que sentir necessidade.

A alteração ou exclusão de quaisquer tópicos deste documento devem ser autorizadas ou aprovadas pelo membro ou diretor responsável pela manutenção do documento. Caso não exista clareza sobre o responsável, deve se contata a diretoria executiva vigente.

## 2 Formatações gerais

### 2.1 Fontes

A fonte padrão a ser utilizada em relatórios deve ser **Arial** de tamanho **12 pontos**. Para textos alternativos, como códigos de programação, devem utilizar **Fira code** ou alguma fonte **Mono espaçada**.

### 2.2 Margens e disposições

A folha deve sempre ter o tamanho A4 e as margens seguem a **Tabela 1** a seguir;

Tabela 1: Margens de folhas **A4**

Margem	Valor
Cima	3,00 cm
Baixo	2,00 cm
Esquerda	3,00 cm
Direita	2,00 cm

### 2.3 Números

Todos os números devem seguir os seguintes padrões:

- Casas decimais separadas por virgula;
- Milhares devem estar separados por ponto;
- Números inteiros não devem ter virgula;
- Números reais devem ter duas casas decimais de precisão;
- O arredondamento deve ser feito para cima;
- *p-valores* devem ter 3 casas decimais de precisão;
- Caso o *p-valor* seja menor que 0,001 deve-se representar por <0,001.

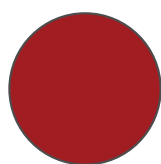
Existem também alguns comandos que facilitam a padronização dos números e medidas.

Tabela 2: Formatos e comandos para números gerais

Comando	Formatação	Exemplo	Resultado
<code>\num{}</code>	XX.XXX,XX	<code>\num{12345.78923}</code>	12.345,79
<code>\inteiro{}</code>	XX.XXX	<code>\inteiro{12345.78923}</code>	12.345
<code>\pvalor{}</code>	X,XXX	<code>\pvalor{0.78923}</code>	0,789
<code>\pvalor{}</code>	X,XXX	<code>\pvalor{0.00023}</code>	<0,001

### 3 Cores

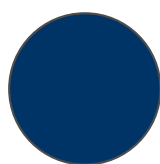
Os gráficos devem dar preferência às cores auxiliares, sugeridas no manual da logo da **ESTAT** Consultoria, já que essas cores ajudam a definir a identidade visual da empresa e as cores selecionadas possuem afinidade com a marca **ESTAT**. A paleta de cores a seguir possui os códigos no padrão hexadecimal e RGB. A seguir a paleta de cores e seus respectivos códigos:



**Cor Principal**

HTML #A11D21

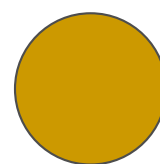
RGB (161, 29, 33)



**Cor Secundaria**

HTML #003366

RGB (0, 51, 102)














**Cor Terciaria**

HTML #CC9900

RGB (240, 153, 0)

Caso haja necessidade de mais cores use a **Tabela 3** abaixo.

Tabela 3: Tabela de cores completa

Cor	hex	RBG
	#A11D21	(161, 29, 33)
	#003366	( 0, 51, 102)
	#CC9900	(204, 153, 0)
	#663333	(102, 51, 51)
	#FF6600	(255, 102, 0)
	#CC9966	(204, 153, 102)
	#999966	(153, 153, 102)
	#006606	( 0, 102, 6)
	#008091	( 0, 128, 145)
	#041835	( 4, 24, 53)
	#666666	(102, 102, 102)

Caso a **Tabela 3** não seja suficiente ou necessite de alteração, contate o responsável por esse documento ou a **DIREX**.

#### 3.1 Códigos relacionados

No R temos a disponibilidade do vetor de cores a seguir.

```
estat_colors <- c(
  "#A11D21", "#003366", "#CC9900",
  "#663333", "#FF6600", "#CC9966",
  "#999966", "#006606", "#008091",
  "#041835", "#666666" )
```

Além das cores, todos os gráficos devem seguir o tema da **ESTAT**, que estabelece fundos, cores dos eixos, grades e vários outros aspectos visuais dos gráficos. A definição do tema está a seguir:

```
theme_estat <- function(...) {
  theme <- ggplot2::theme_bw() +
    ggplot2::theme(
      axis.title.y = ggplot2::element_text(colour = "black", size = 12),
      axis.title.x = ggplot2::element_text(colour = "black", size = 12),
      axis.text = ggplot2::element_text(colour = "black", size = 9.5),
      ,
      panel.border = ggplot2::element_blank(),
      axis.line = ggplot2::element_line(colour = "black"),
      legend.position = "top",
      ...
    )

  return(
    list(
      theme,
      scale_fill_manual(values = estat_colors),
      scale_colour_manual(values = estat_colors)
    )
  )
}
```

Pra inserir o tema no gráfico, basta colocar ao final do código:

```
ggplot(data) +
  ... +
  theme_estat()
```



## 4 Modelo de relatório

### 4.1 Introdução

A introdução deve abordar sobre o **objetivo do projeto**, o **banco de dados** e sobre o **software utilizado**.

No primeiro parágrafo fale sobre o objetivo, quais serão as análises que serão feitas (descritiva, correlação, multi, ...), mas só falar o que foi feito. Colocar o **nível de significância** usados nas análises (não esquecer de também acrescentar na metodologia). Pode começar assim: o seguinte projeto tem como objetivo ... Lembrando que não precisa falar de variável por variável.

No segundo, fale sobre o banco de dados, como foi feito a coleta (se foi pelo cliente ou por nós, se a amostra é boa, se é probabilística ... ). Fale sobre as variáveis que foram coletadas, e o que achar importante sobre o tópico.

No último parágrafo, fale qual o software utilizado e a versão dele.

### 4.2 Metodologia

Colocar um texto introdutório sobre a metodologia.

*"Este relatório é composto por técnicas estatísticas que serão descritas a seguir de acordo com o que foi utilizado em tal estudo".*

As metodologias estarão em um documento padrão elaborado pela projetos.

### 4.3 Análises

#### 4.3.1 Introdução das análises

Deve-se introduzir uma análise fazendo uma descritiva básica e superficial sobre as variáveis que vão ser descritas. Dizer o tipo de variável, quais suas categorias/quais suas variações, botar o que a sigla significa por exemplo.

Há uma flexibilidade com análises descritivas simples (como por exemplo, distribuição de sexo ou escolaridade), na qual não há muito o que introduzir e por isso, nesses casos não é obrigatório.

#### 4.3.2 Separação das seções

Não se deve separar as análises do relatório por sua natureza, ou seja, não se deve separar as análises em: univariada, bivariada, teste de Hipóteses... Separaremos por variáveis, para incluir por exemplo a descritiva e o próprio teste de hipótese na mesma seção.

No caso em que serão feitas comparações bivariadas, serão feitas as univariadas de cada uma e em seguida a comparação entre as duas, depois se continua com o restante das outras análises. Se houver uma “*variável-pivot*” que será analisada conjuntamente com muitas outras, faça a descritiva dessa variável e em seguida, dentro das seções das outras variáveis, faça a univariada e bivariada seguidamente.

#### 4.3.3 Geral

Não se deve mencionar as variáveis com letra inicial maiúscula no meio do texto, a menos de grupos/categorias com nomes (grupo Amor, grupo Bola, grupo Casa...). Nestes se faz importante pra ajudar a diferenciar no meio do texto.

**Observação:** Use moderadamente a palavra “variável” antes de mencionar uma variável, nem sempre será necessário.

#### 4.3.4 Descritiva

**Observação:** isso é mais um dica do que uma regra da padronização de fato

Uma boa sugestão é trabalhar com tópico frasal nos textos para evitar que ocorram parágrafos formados por um período só. Seria esse padrão de se introduzir uma ideia em uma frase e depois explicar sobre ela.

*“Avaliando em cada grupo, percebe-se por meio da Figura 10 que os grupos se comportam de maneira muito semelhante.” (Introdução) “O Quadro 5 mostra que a maioria dos valores das medidas são iguais ou bastante próximos, revelando que em ambos os grupos pelo menos 50% dos pacientes tiveram dor superior a 7 na escala numérica e 25% foram maior do que 8.” (Explicação)*

*“Em comparação com a pré-intervenção, a distribuição da escala de dor foi bastante puxada para valores menores. Agora, pessoas que possuem 0 na escala de dor não são mais casos extremos, além de que mais da metade foi inferior a 5, diferentemente do momento pré-intervenção, em que 75% dos pacientes tiveram dores superiores a 5 na escala.”*

#### 4.3.5 Testes de hipóteses

Não é necessário apresentar as hipóteses de testes dos pressupostos, como normalidade ou homocedasticidade, basta mencionar em texto que os pressupostos do teste foram atendidos (quando forem) ou não, e seguir com a apresentação das hipóteses do teste realizado.

Deve-se também, deixar claro quais são as hipóteses em texto, exemplo: “[...] As hipóteses testadas são:”(seguir com padronização apresentada nesta seção), uma vez

que o cliente pode não conhecer a notação que é usada para apresentação das hipóteses, assim busca-se deixar claro que as hipóteses estão contidas nas chaves.

Por fim, não esqueça de explicitar o nível de significância usado quando esse for mencionado.

## 4.4 Conclusão

A conclusão deve ser um resumo sucinto avaliando o objetivo do cliente com o projeto. Por exemplo, no projeto *2020012 do Marcos* imaginamos que o cliente queira saber qual o impacto de um tratamento nas lesões dos pacientes, e a ideia seria fazer uma conclusão que apresenta exatamente seus objetivos: **parágrafos** explicando em quais tratamentos a intervenção foi eficaz ou se piorou a dor dos pacientes, sem abordar diretamente os dados. **Aqui um exemplo de conclusão nesse estilo (página 20)**

## 5 Padronização de Gráficos

### 5.1 Introdução

O `ggplot2` é um pacote do software R desenvolvido para a criação de gráficos. É baseado em um bem fundamentado sistema teórico e apesar de ter sido criado há pouco tempo, já é amplamente utilizado por conta de sua versatilidade e bela interface gráfica.

O pacote é uma combinação de outros dois sistemas de construção de gráficos, o `base` e o `lattice`. Neste sistema o usuário pode incrementar o gráfico após a sua construção (sistema `base`) adicionando argumentos um a um. Entretanto, alguns aspectos estéticos são feitos automaticamente, sem nenhuma interferência do usuário (sistema `lattice`).

Como em qualquer outro pacote, o `ggplot2` também apresenta limitações. Porém, para muitos propósitos, o `ggplot2` retorna o melhor resultado em relação ao tempo investido, sempre auxiliando na visualização dos dados com sua excelente interface gráfica.

### 5.2 Contextualização

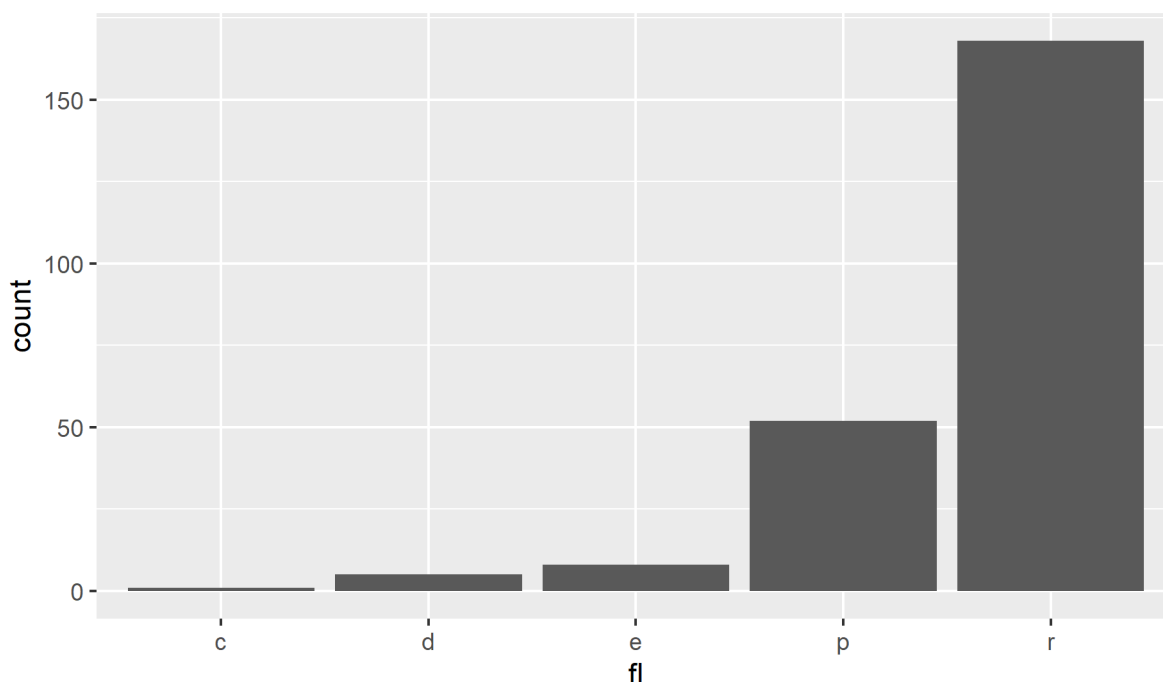
No desenvolvimento do gráfico, deve-se levar em consideração as propriedades dos dados. As variáveis são classificadas em qualitativas e quantitativas e suas características devem estar corretamente evidenciadas no gráfico. No sistema `base`, um gráfico de barras seria feito da seguinte forma:

```
barplot(table(mpg$fl))
```

Para realizar este mesmo gráfico no `ggplot2`, primeiro deve-se instalar e, em seguida, carregar o pacote. Então, no comando `ggplot( )`, especifica-se o banco de dados. Em seguida, com o comando `geom_bar( )` e argumento `aes( )`, determina-se qual variável se deseja plotar e como as barras serão desenhadas. Ou seja,

```
install.packages("ggplot2")  
library(ggplot2)  
ggplot(mpg) + geom_bar(aes(x = fl))
```

Figura 1: Gráfico de Barras feito com `ggplot()` e `geom_bar()`



Antes de abordar aspectos mais específicos, definir a terminologia utilizada no sistema será útil:

- O conjunto de dados e o que se deseja visualizar. Tal conjunto consiste em variáveis que são armazenadas em colunas de um data frame;
- *Geoms* são os objetos geométricos utilizados para representar os dados, como barras, linhas e pontos;
- *Aesthetics*, ou aspectos estéticos, são as propriedades visuais dos *geoms*, ou seja, posição de x e y, cor das linhas, formato dos pontos, etc;
- Há *mappings* para o que é passado em *aesthetics*;
- *Scales* controla o mapeamento dos valores em relação ao espaço estético;
- *Guides* para mostrar ao usuário como mapear as propriedades visuais de volta para o espaço de dados inicial.

Os comandos do gráficos são adicionados pelo símbolo `+`. Vale ressaltar que não há intercâmbio de funções e argumentos entre os sistemas gráficos. O comando utilizado para adicionar título a um gráfico feito no base não é o mesmo que o utilizado no `ggplot2`.

Para salvar um gráfico, basta executar a função `ggsave`, inserindo como parâmetro o caminho para onde será salvo o arquivo. Ao fazer isso, salve o arquivo no formato PDF.

Para dar continuidade, o banco de dados utilizado para desenvolver este documento de padronização será o banco contido no pacote `ggplot2`, *mpg*.

### 5.3 Sobre tabelas e frequências

As tabelas devem ser usadas caso seja pedido pelo cliente, ou em casos com quantidade considerável de grupos na variável. Em casos comuns, as tabelas deverão ser incorporadas aos gráficos (barras ou coluna) que apresentarão ambas as frequências: absoluta e relativa, no formato `X(Y%)`.

*Observação:* Essa regra é apenas para as tabelas de frequências, nos caso dos **quadros**, não há nada a pontuar, sempre que for necessário (seja por complementar um **boxplot** ou por apresentar os resultados de um teste) deve ser usado normalmente.

Ressalta-se que no exemplo do "gráfico de colunas bivariado com frequências" abaixo, as categorias no eixo x, somam cada uma, 100%, ou seja, nesse caso estamos comparando cada categoria da variável usadas na cores (fill) em função de cada categoria do eixo x. Esse deve ser o padrão utilizado, a menos que exista alguma demanda do gerente de projetos ou cliente.

### 5.4 Sobre ordenação nos gráficos

Fazer ordenação decrescente das **variáveis ordinais** (nos gráficos, quanto mais à direita menor é a ordenação; nas tabelas, quanto mais pra baixo maior é a ordenação) e nas **qualitativas nominais**, ordenar por qual é a categoria mais frequente seguindo a mesma ideia para gráficos e tabelas.

### 5.5 Gráfico de Colunas/Barras

#### 5.5.1 Colunas univariado com frequências

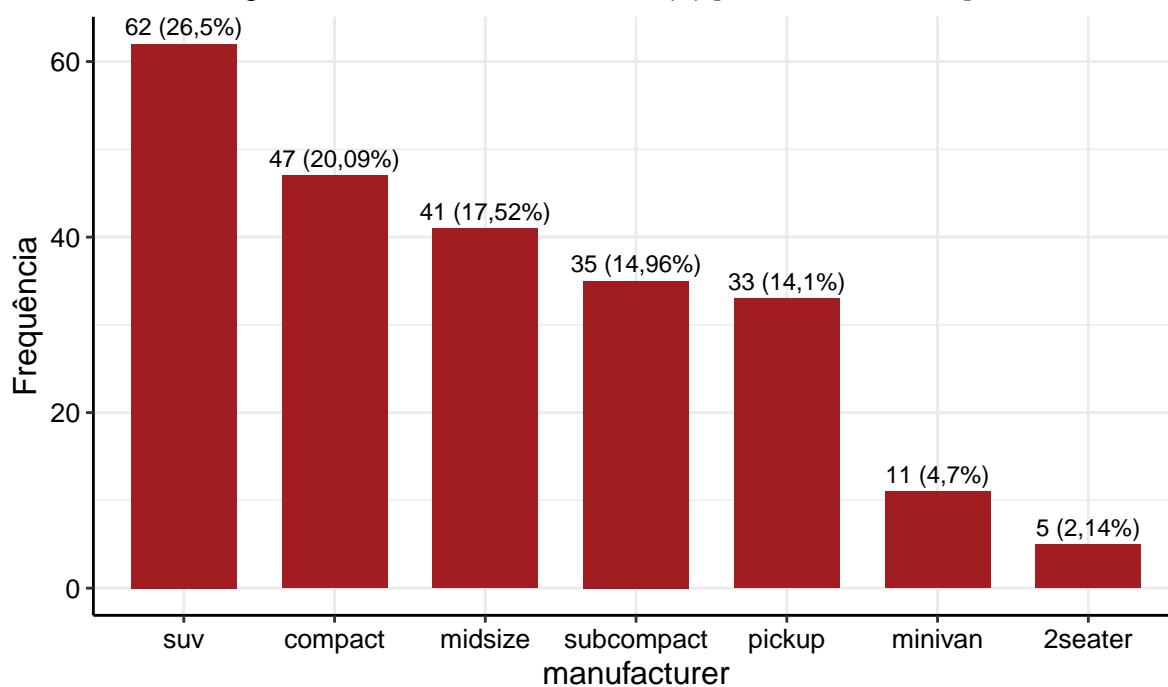
```
classes <- mpg %>%  
  filter(!is.na(class)) %>%  
  count(class) %>%  
  mutate(  
    freq = n,  
    relative_freq = round((freq / sum(freq)) * 100, 1),  
    freq = gsub("\\.", ",", relative_freq) %>% paste("%", sep = ""),  
    label = str_c(n, " (", freq, ")") %>% str_squish()  
  )  
  
ggplot(classes) +
```

```

aes(x = fct_reorder(class, n, .desc=T), y = n, label = label) +
geom_bar(stat = "identity", fill = "#A11D21", width = 0.7) +
geom_text(
  position = position_dodge(width = .9),
  vjust = -0.5, #hjust = .5,
  size = 3
) +
labs(x = "manufacturer", y = "Frequência") +
theme_estat()
ggsave("colunas-uni-freq.pdf", width = 158, height = 93, units = "mm"
)

```

Figura 2: Gráfico de colunas da(o) [nome da variável]



### 5.5.2 Colunas bivariado com frequências

```
trans_drv <- mpg %>%
  mutate(trans = case_when(
    trans %>% str_detect("auto") ~ "auto",
    trans %>% str_detect("manual") ~ "manual"
  )) %>%
  group_by(trans, drv) %>%
  summarise(freq = n()) %>%
  mutate(
    freq_relativa = round(freq / sum(freq) * 100, 1)
  )

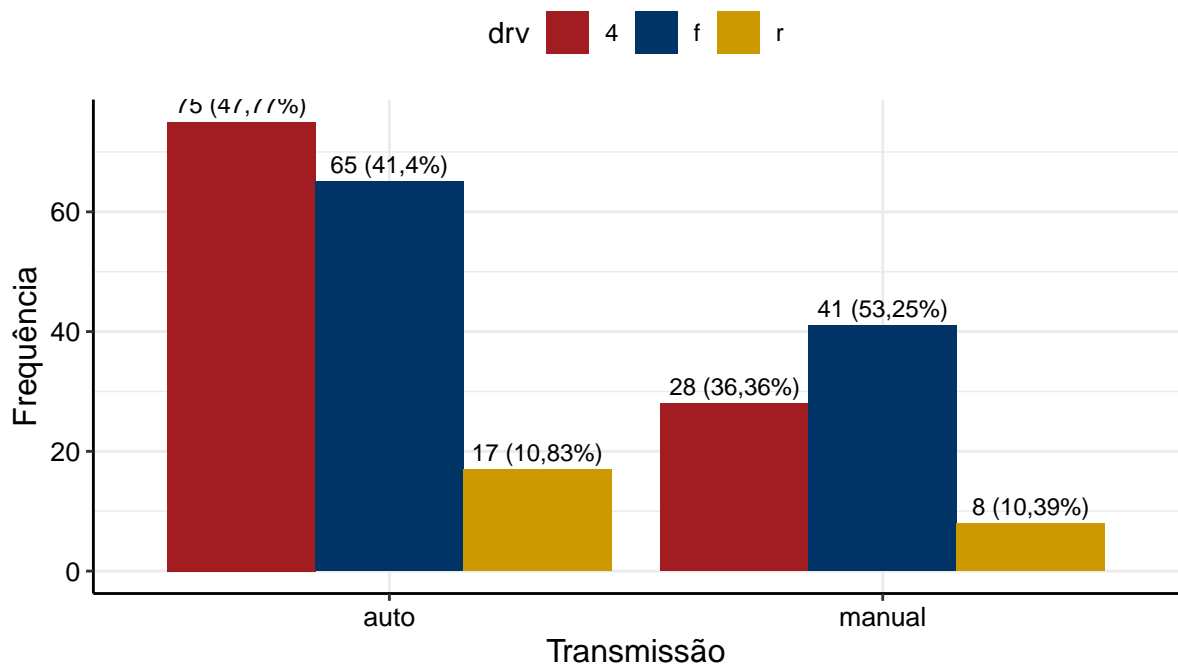
porcentagens <- str_c(trans_drv$freq_relativa, "%") %>% str_replace(
  "\\.", ",")

legendas <- str_squish(str_c(trans_drv$freq, " (", porcentagens, ")")
  )

ggplot(trans_drv) +
  aes(
    x = fct_reorder(trans, freq, .desc = T), y = freq,
    fill = drv, label = legendas
  ) +
  geom_col(position = position_dodge2(preserve = "single", padding =
    0)) +
  geom_text(
    position = position_dodge(width = .9),
    vjust = -0.5, hjust = 0.5,
    size = 3
  ) +
  labs(x = "Transmissão", y = "Frequência") +
  theme_estat()
ggsave("colunas-bi-freq.pdf", width = 158, height = 93, units = "mm")
```



Figura 3: Gráfico de colunas da(o) [nome da variável] pela(o) [nome da variável]



### 5.5.3 Observações

Para transformar/fazer gráficos de barras, basta adicionar a linha de código: `coord_flip()`, no final do código do gráfico de colunas. Faça gráficos de barras apenas quando necessário e utilize *barras* no lugar de *colunas* no título.

Os parâmetros, *vjust* e *hjust* na função `geom_text` devem ser ajustados. São os parâmetros que modificam as posições dos *labels* de frequências.

Gráficos tanto de barras quanto colunas sem frequências deverão ser usados nos casos em que a tabela se faz necessária.

Exemplos das observações no código *Padronização.R*.

## 5.6 Gráfico de Setor

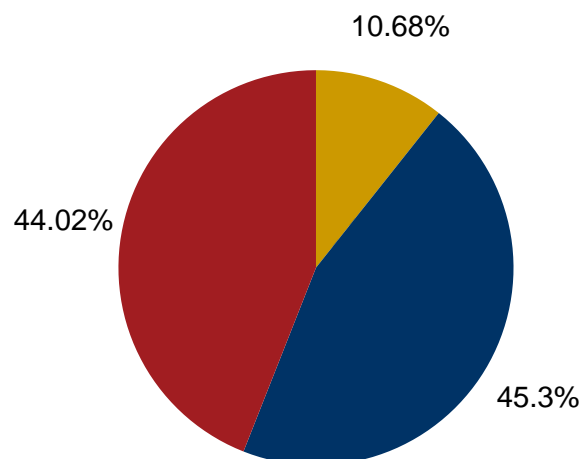
```
contagem <- mpg %>%
  group_by(drv) %>%
  summarise(Freq = n()) %>%
  mutate(Prop = round(100*(Freq/sum(Freq)), 1)) %>%
  arrange(desc(drv)) %>%
  mutate(posicao = cumsum(Prop) - 0.5*Prop)

ggplot(contagem) +
  aes(x = factor(""), y = Prop, fill = factor(drv)) +
  geom_bar(width = 1, stat = "identity") +
```

```
coord_polar(theta = "y") +
geom_text(
  aes(x = 1.8, y = posicao, label = paste0(Prop, "%")),
  color = "black"
) +
theme_void() +
theme(legend.position = "top") +
scale_fill_manual(values = estat_colors, name = 'DRV')
ggsave("setor.pdf", width = 158, height = 93, units = "mm")
```

Figura 4: Gráfico de setores da(o) [nome da variável]

DRV 4 f r

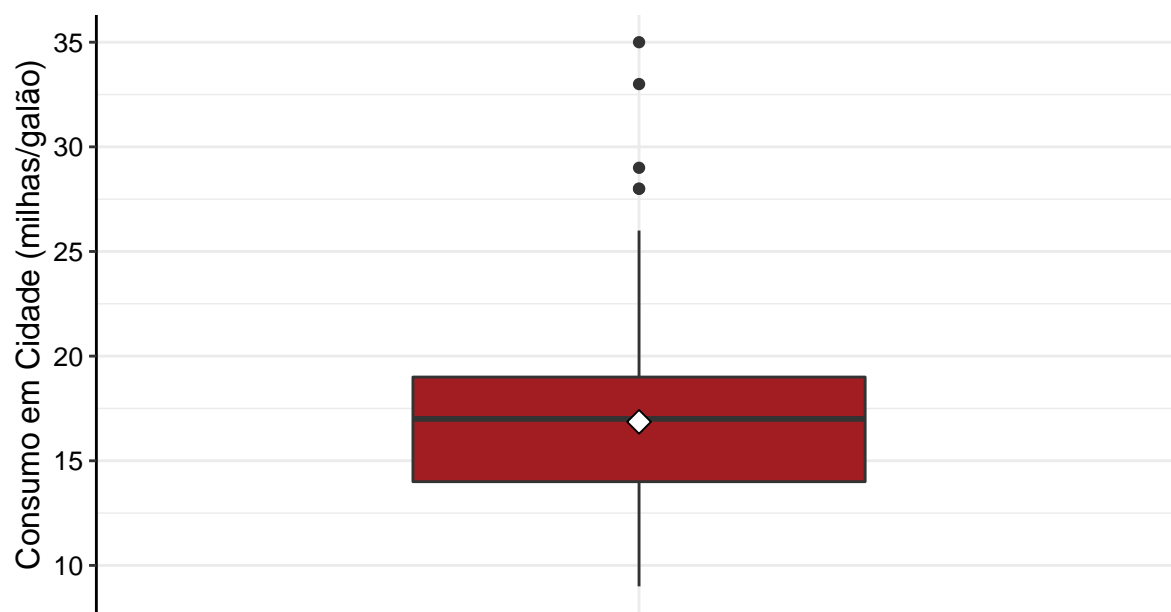


## 5.7 Boxplot

### 5.7.1 Univariado

```
ggplot(mpg) +
  aes(x=factor(""), y=cty) +
  geom_boxplot(fill=c("#A11D21"), width = 0.5) +
  guides(fill=FALSE) +
  stat_summary(fun="mean", geom="point", shape=23, size=3, fill="
    white")+
  labs(x="", y="Consumo em Cidade (milhas/galão)") +
  theme_estat()
```

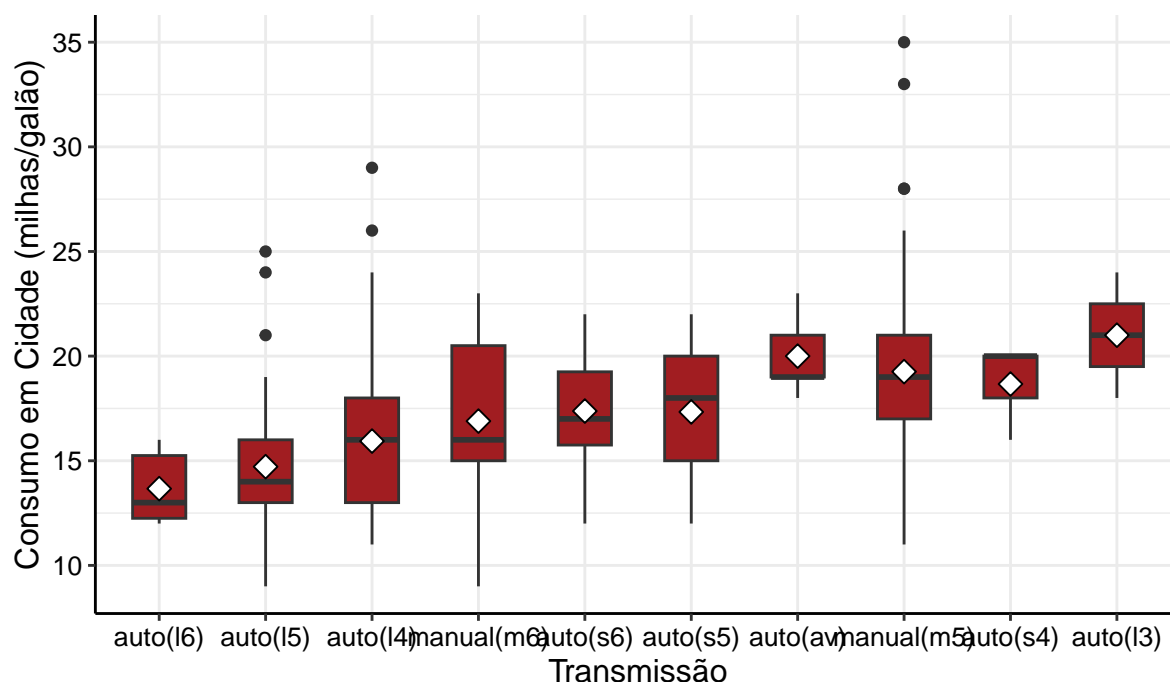
Figura 5: Boxplot do(a) [nome da variável]



### 5.7.2 Bivariado

```
ggplot(mpg) +
  aes(x = reorder(trans, cty, FUN = median), y = cty) +
  geom_boxplot(fill = c("#A11D21"), width = 0.5) +
  stat_summary(
    fun = "mean", geom = "point", shape = 23, size = 3, fill = "white"
  ) +
  labs(x = "Transmissão", y = "Consumo em Cidade (milhas/galão)") +
  theme_estat()
ggsave("box_bi.pdf", width = 158, height = 93, units = "mm")
```

Figura 6: Boxplot do(a) [nome da variável] pelo(a) [nome da variável]



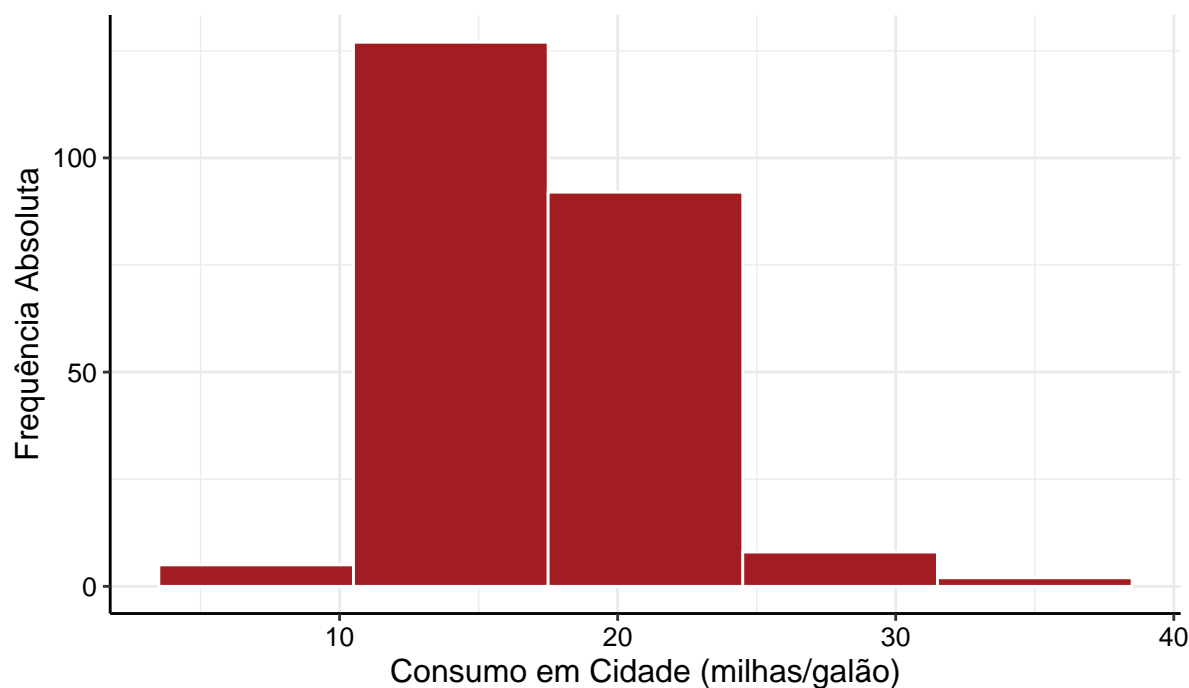
**Observação:** As variáveis devem estar ordenadas em ordem crescente de *mediana* para facilitar a visualização. Caso outra deseje evidenciar outra estatística sinta-se livre para ordenar por ela, por exemplo, a *média*.

## 5.8 Histograma

### 5.8.1 Univariado

```
ggplot(mpg) +
  aes(x = cty) +
  geom_histogram(colour = "white", fill = "#A1D21", binwidth = 7) +
  labs(x = "Consumo em Cidade (milhas/galão)", y = "Frequência
  Absoluta") +
  theme_estat()
ggsave("hist_uni.pdf", width = 158, height = 93, units = "mm")
```

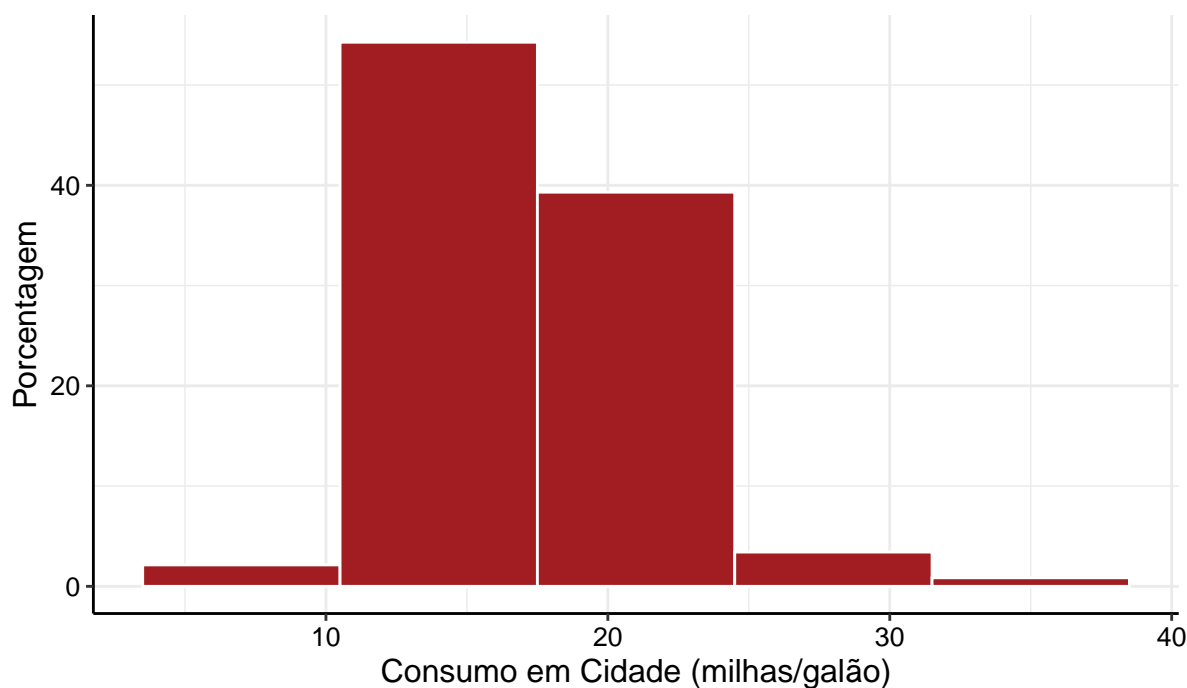
Figura 7: Histograma da(o) [nome da variável]



### 5.8.2 Univariado com porcentagem

```
ggplot(mpg) +
  aes(x = cty) +
  geom_histogram(
    aes(y = 100 * after_stat(count) / sum(after_stat(count))),
    colour = "white",
    fill = "#A11D21",
    binwidth = 7
  ) +
  labs(x = "Consumo em Cidade (milhas/galão)", y = "Porcentagem") +
  theme_estat()
ggsave("hist_uni_porc.pdf", width = 158, height = 93, units = "mm")
```

Figura 8: Histograma da(o) [nome da variável]

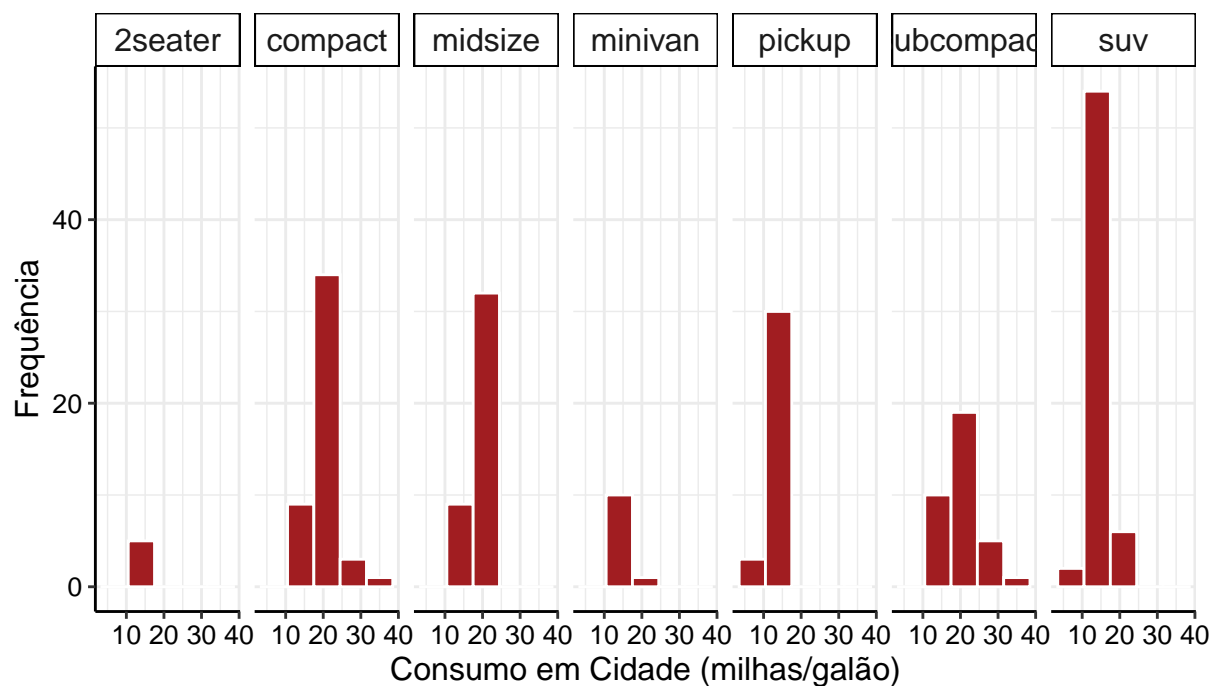


### 5.8.3 Bivariado com facet grid

```
ggplot(mpg) +
  aes(x = cty) +
  geom_histogram(colour = "white", fill = "#A11D21", binwidth = 7) +
  facet_grid(. ~ class) +
  labs(x = "Consumo em Cidade (milhas/galão)", y = "Frequência") +
  theme_estat(
    strip.text = element_text(size = 12),
    strip.background = element_rect(colour = "black", fill = "white")
  )

ggsave("charts/hist_grid.pdf", width = 200, height = 93, units = "mm")
```

Figura 9: Histograma da(o) [nome da variável] por [nome da variável]

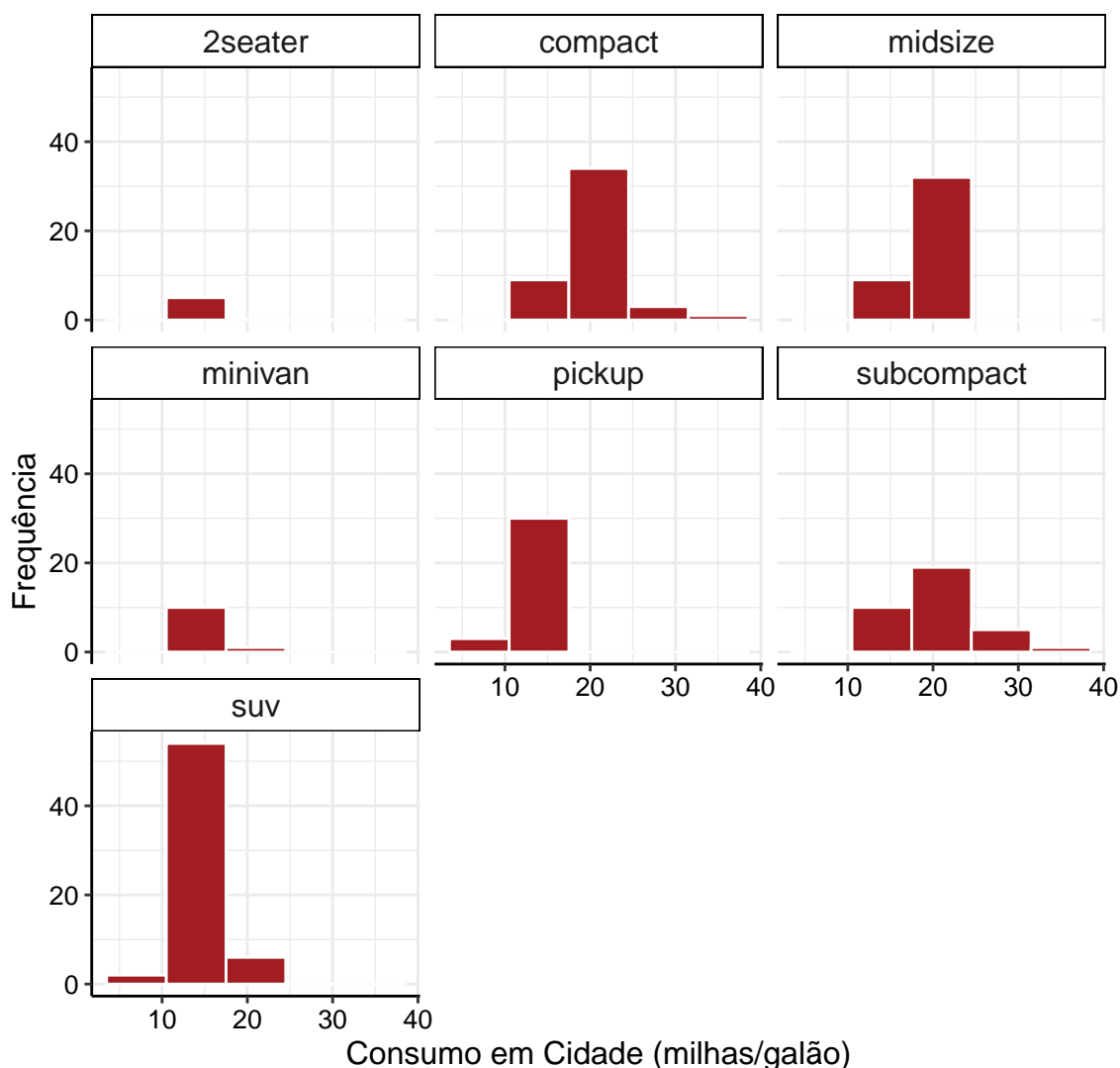


#### 5.8.4 Bivariado com facet wrap

```
ggplot(mpg) +
  aes(x=cty) +
  geom_histogram(colour="white", fill="#A11D21", binwidth=7)+
  facet_wrap(class ~ .) +
  labs(x="Consumo em Cidade (milhas/galão)", y="Frequência") +
  theme_estat(
    strip.text = element_text(size=12),
    strip.background = element_rect(colour="black", fill="white")
  )

ggsave("hist_wrap.pdf", width = 160, height = 150, units = "mm")
```

Figura 10: Histograma da(o) [nome da variável] por [nome da variável]



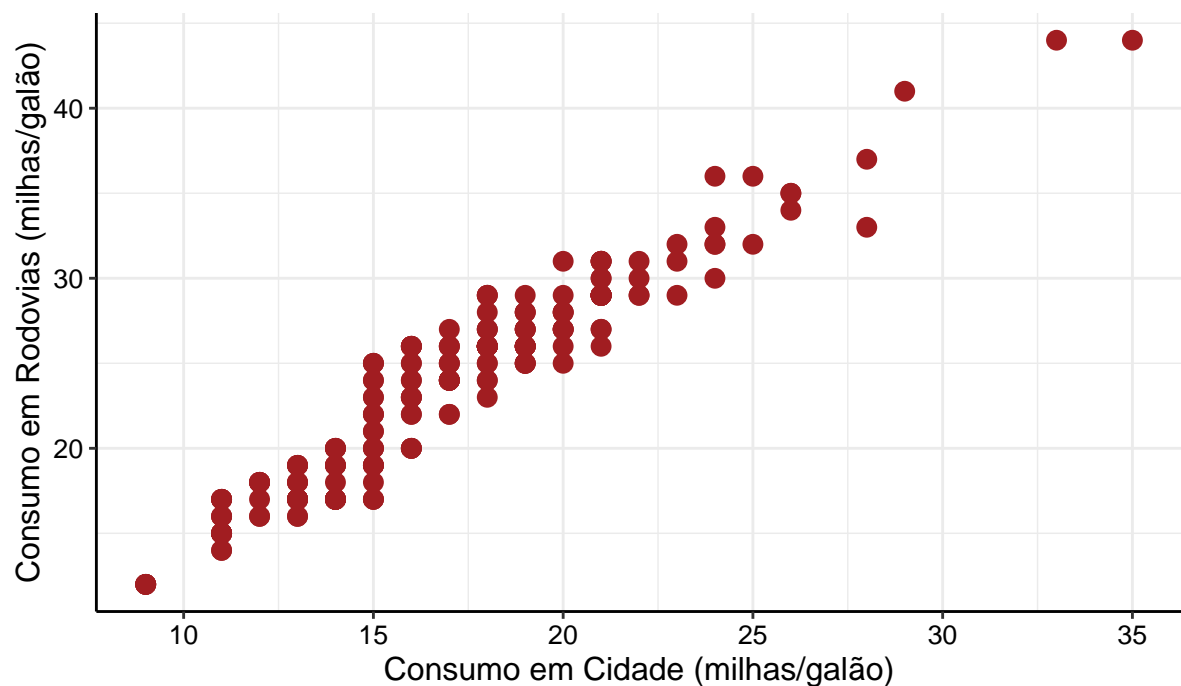
## 5.9 Gráfico de Dispersão

### 5.9.1 Bivariado

```
ggplot(mpg) +
  aes(x = cty, y = hwy) +
  geom_point(colour = "#A11D21", size = 3) +
  labs(
    x = "Consumo em Cidade (milhas/galão)",
    y = "Consumo em Rodovias (milhas/galão)"
  ) +
  theme_estat()
ggsave("disp_uni.pdf", width = 158, height = 93, units = "mm")
```



Figura 11: Gráfico de dispersão da(o) [nome da variável 1] pela [nome da variável 2]

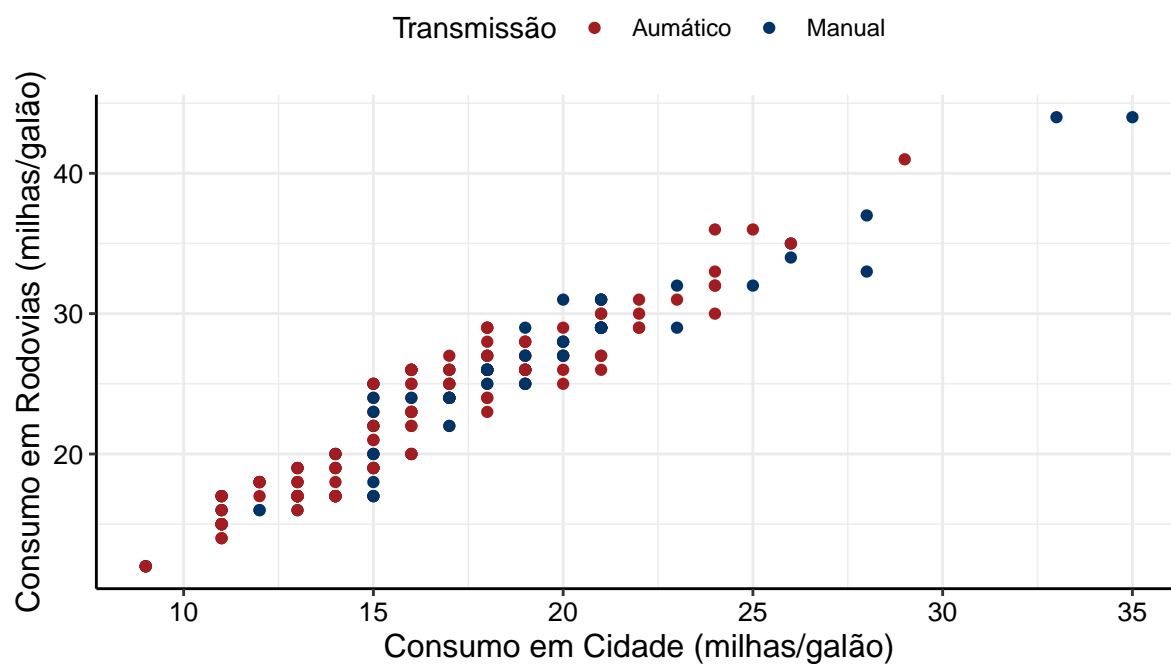


Em casos de muito pontos sobrepostos consultar o código *Padronização.R*, há duas alternativas: usar o *geom\_jitter* ou alterar o *alpha* em *geom\_point*.

### 5.9.2 Multivariado

```
mpg %>%
  mutate(Transmissão = case_when(
    trans %>% str_detect("auto") ~ "Automático",
    trans %>% str_detect("manual") ~ "Manual"
  )) %>%
  ggplot() +
  aes(x=cty, y=hwy) +
  geom_point(aes(colour=Transmissão)) +
  labs(
    x="Consumo em Cidade (milhas/galão)", y="Consumo em Rodovias (
      milhas/galão)"
  ) +
  theme_estat()
ggsave("disp_bi.pdf", width = 158, height = 93, units = "mm")
```

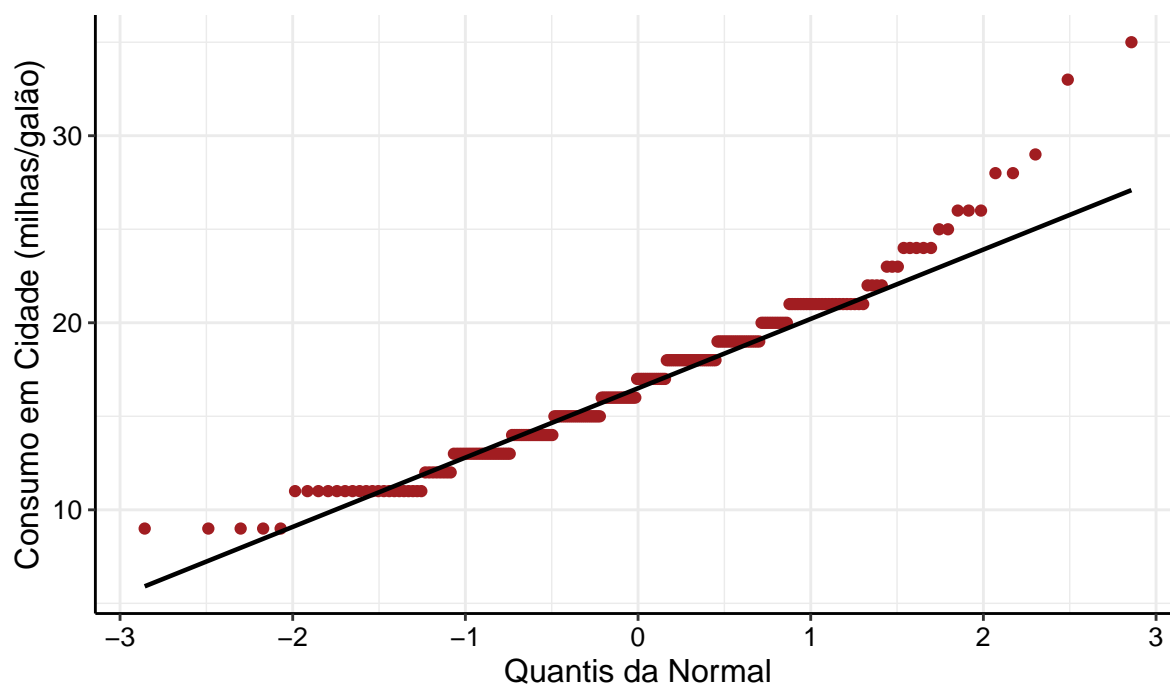
Figura 12: Gráfico de dispersão da(o) [nome da variável 1] pela [nome da variável 2] e [nome da variável 3]



## 5.10 QQ Plot

```
ggplot(mpg) +
  aes(sample = cty) +
  stat_qq(colour = "#A11D21") +
  stat_qq_line(size = 0.8) +
  labs(
    x = "Quantis da Normal",
    y = "Consumo em Cidade (milhas/galão)"
  ) +
  theme_estat()
ggsave("charts/qq_plot.pdf", width = 158, height = 93, units = "mm")
```

Figura 13: Gráfico de probabilidade normal do(a) [nome da variável]



## 5.11 Gráfico de Linhas

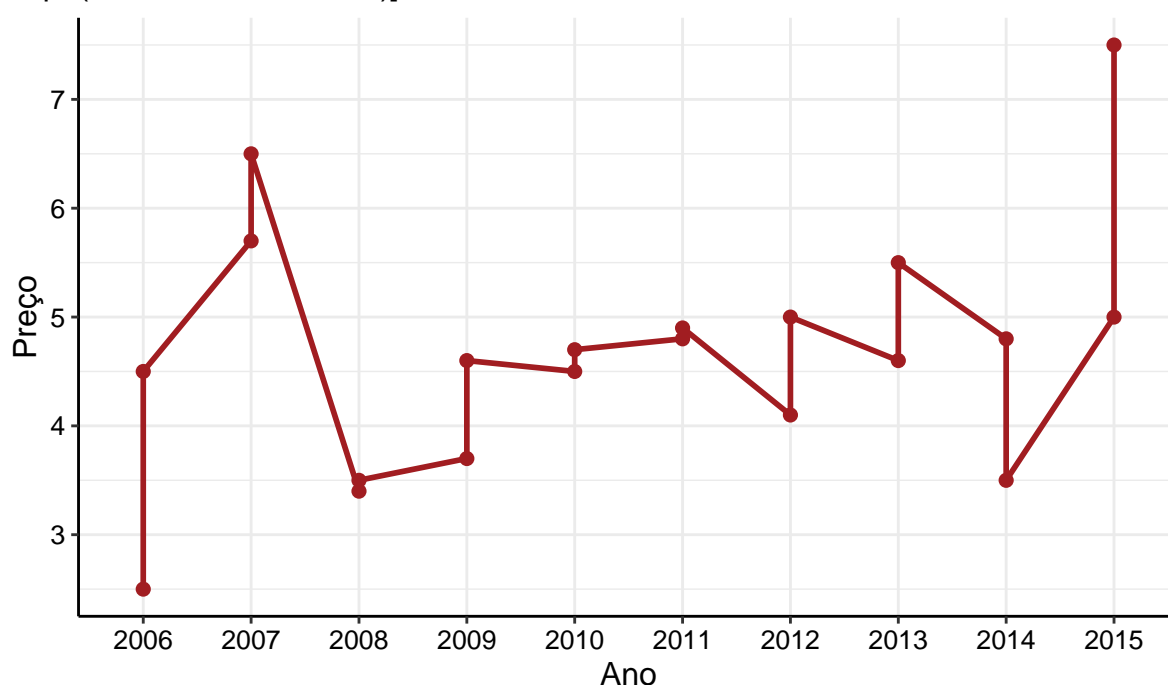
Considerando o seguinte banco de dados:

```
dados <- tibble(
  ano = c(
    "2006", "2007", "2008", "2009", "2010", "2011", "2012", "2013", "
    2014",
    "2015", "2006", "2007", "2008", "2009", "2010", "2011", "2012", "
    2013",
    "2014", "2015"
  ),
  preco = c(
    2.5, 5.7, 3.4, 3.7, 4.5, 4.8, 4.1, 4.6, 4.8,
    5, 4.5, 6.5, 3.5, 4.6, 4.7, 4.9, 5, 5.5, 3.5,
    7.5
  ),
  produto = c(
    "a", "a", "a", "a", "a", "a", "a", "a", "b",
    "b", "b", "b", "b", "b", "b", "b", "b", "b",
    "b", "b"
  )
)
```

### 5.11.1 Univariado

```
ggplot(dados) +  
  aes(x=ano, y=preco, group=1) +  
  geom_line(size=1, colour="#A11D21") + geom_point(colour="#A11D21",  
    size=2) +  
  labs(x="Ano", y="Preço") +  
  theme_estat()  
ggsave("series_uni.pdf", width = 158, height = 93, units = "mm")
```

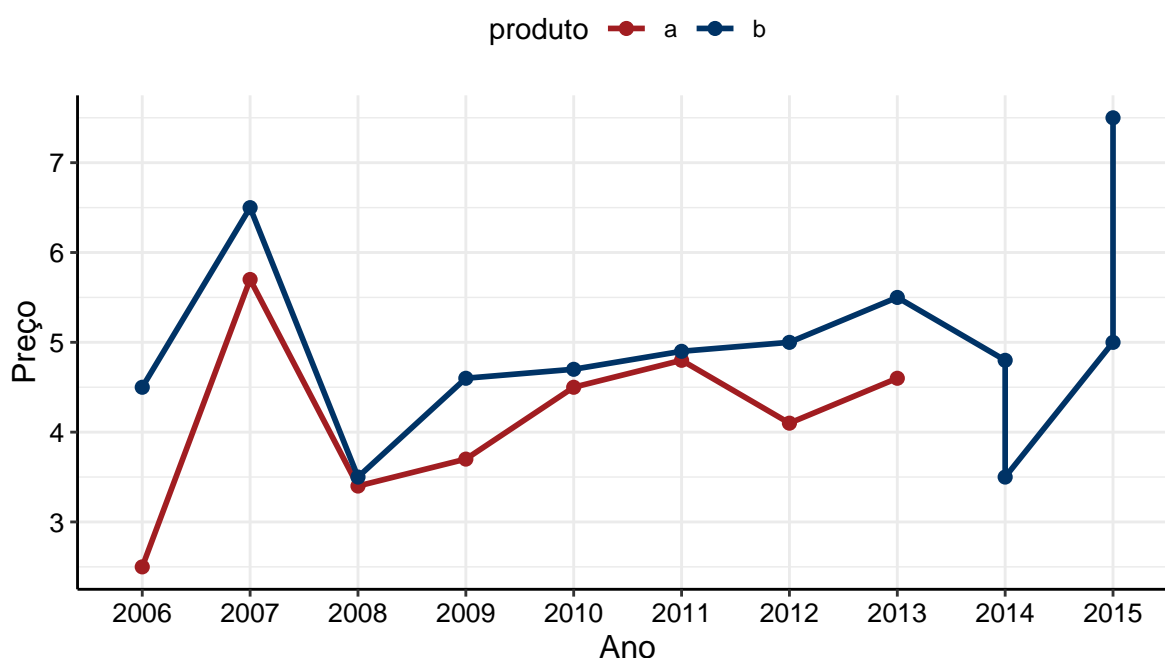
Figura 14: Gráfico de linhas da [nome da variável 1] com o decorrer dos [unidade de tempo(anos,dias,meses,etc)]



### 5.11.2 Bivariado

```
ggplot(dados) +  
  aes(x = ano, y = preco, group = produto, colour = produto) +  
  geom_line(size = 1) +  
  geom_point(size = 2) +  
  labs(x = "Ano", y = "Preço") +  
  theme_estat()  
ggsave("series_grupo.pdf", width = 158, height = 93, units = "mm")
```

Figura 15: Gráfico de linhas da [nome da variável 1] e [nome da variável 2] com o decorrer dos [unidade de tempo(anos,dias,meses,etc)]



### 5.11.3 Observação

Caso a variável possua o formato de mês, dia ou data, recomenda-se que o seguinte comando seja usado para ordenar o eixo x:

```
scale_x_date()  
scale_y_date()  
scale_x_datetime()  
scale_y_datetime()
```

## 6 Padronização de $\text{\LaTeX}$

### 6.1 Tabelas

Nesta seção estão contidas as padronizações para tabelas em geral.

#### 6.1.1 Formatações gerais

- **Título:** Deve conter o nome da variável em questão e a variável que separa as categorias.
- **Coluna:** Nomes simples das variáveis presentes na coluna. Idealmente em negrito (`\textbf{ }`)
- **Label:** Idealmente deve seguir a forma *tab:[identificador da tabela]* aonde o *[identificador da tabela]* deve seguir os padrões de nomenclatura de código.

Tabela 4: Alinhamentos de Tabelas Simples

Tipo de entrada	Alinhamento
Texto	Esquerda
Numero	Direita
Intervalo	Centro
Titulo	O mesmo da variável

#### 6.1.2 Tabela simples

- Sobre o **Total**: Quando a tabela apresentar o Total da coluna, essa linha deve ser a última da tabela, separada por uma linha (`\midrule`) o "nome" da linha deve estar em negrito (`\textbf{Total}`);
- Sobre as linhas de divisória: as tabelas não devem conter linhas verticais separando todas as colunas. Para melhor visualizarmos as categorias da tabela, podemos colocar uma linha ali. Não se deve colocar linhas verticais nas extremidades, a menos que seja requisitado pelo requerente;
- Números: devem seguir as formatações gerais de números.

**Exemplos:**

Tabela 5: Frequências da variável [nome da variável]

Nome da variável	Frequência	Porcentagem
Categoria 1	15	8,93%
Categoria 2	439	41,69%
Categoria 3	520	49,38%
<b>Total</b>	1053	100,00%

A Tabela **Tabela 5** expressa valores relativos e absolutos.

Tabela 6: Frequência de classes da variável [nome da variável]

nome da variável	Frequência	Porcentagem
Intervalo 1	43	4,08%
Intervalo 2	96	9,11%
Intervalo 3	225	21,36%
Intervalo 4	210	19,94%
Intervalo 5	205	19,46%
Intervalo 6	156	14,81%
Intervalo 7	69	6,55%
Intervalo 8	27	2,56%
Intervalo 9	13	1,23%
Intervalo 10	9	0,09%
<b>Total</b>	1053	100,00%

### 6.1.3 Tabela de dupla entrada

- **Linhas Verticais:** Deverão haver linhas verticais separando a **Variável resposta** e o **Total**.
- **multicolumn:** Caso usar o comando `\multicolumn{ }{ }` fique atento às linhas verticais! muitas vezes é necessário colocar uma linha vertical no mesmo comando, como no exemplo abaixo foi usado `\multicolumn{2}{c|}`

**Exemplo:**

Tabela 7: Frequências da [variável resposta] por [variável explicativa]

Variável Resposta	Variável Explicativa		Total
	Categoria $b_1$	Categoria $b_2$	
Categoria $a_1$	75	19	94
Categoria $a_2$	300	139	439
Categoria $a_3$	220	300	520
<b>Total</b>	595	458	1053

## 6.2 Quadros

Nesta seção estão contidas as padronizações para os quadros de medidas resumo e quadros para testes

### 6.2.1 Quadro de Medidas Resumo

O quadro resumo pode ser criado no *R* com a seguinte função.

```
print_quadro_resumo <- function(data, var_name, title="Medidas resumo
  da(o) [nome da variável]", label="quad:quadro_resumo1")
{
  var_name <- substitute(var_name)
  data <- data %>%
    summarize(`Média` = round(mean(!sym(var_name)),2),
              `Desvio Padrão` = round(sd(!sym(var_name)),2),
              `Variância` = round(var(!sym(var_name)),2),
              `Mínimo` = round(min(!sym(var_name)),2),
              `1º Quartil` = round(quantile(!sym(var_name), probs =
                .25),2),
              `Mediana` = round(quantile(!sym(var_name), probs = .5)
                ,2),
              `3º Quartil` = round(quantile(!sym(var_name), probs =
                .75),2),
              `Máximo` = round(max(!sym(var_name)),2)) %>%
    t() %>%
    as.data.frame() %>%
    rownames_to_column()

  latex <- str_c("\\begin{quadro}[H]
\\t\\caption{", title, "}
\\t\\centering
\\t\\begin{adjustbox}{max width=\\textwidth}
\\t\\begin{tabular}{", sep="")

  col_count <- ncol(data)
  row_count <- nrow(data)
  latex <- str_c(latex, "| l |\\n", sep=" ")
  for (i in seq(2, col_count))
  {
    numCount <- data[i, -c(1)] %>%
      as.numeric() %>%
      {floor(log10(.)) + 1} %>%
```



```

max()
latex <- str_c(latex, "\t\t\tS[table-format = ", numCount, ".2]\n", sep="")
}

latex <- str_c(latex, "\t\t\t|}\n\t\t\ttoprule\n\t\t\t", sep="")
if (col_count > 2)
{
  for (i in seq(1,col_count))
  {
    if (i == 1)
      latex <- str_c(latex, "\\textbf{Estatística}", sep="")
    else
      latex <- str_c(latex, " \\textbf{", data[1, i], "}", sep="")

    if (i < col_count)
      latex <- str_c(latex, "&", sep=" ")
    else
      latex <- str_c(latex, "\\\\\\\\n", sep=" ")
  }
}
else
{
  latex <- str_c(latex, "\\textbf{Estatística} & \\textbf{Valor}\\\\\\n", sep="")
}

latex <- str_c(latex, "\t\t\t\\midrule\n", sep="")

if (col_count > 2)
  starting_number <- 2
else
  starting_number <- 1

for (i in seq(starting_number, row_count))
{
  latex <- str_c(latex, "\t\t\t", str_flatten(t(data[i,]), collapse = " & "), " \\\\\\\\n")
}
latex <- str_c(latex, "\t\t\t\\bottomrule

```

```

\t\\end{tabular}
\t\\label{"", label, ""}
\t\\end{adjustbox}
\\end{quadro}", sep="")

writeLines(latex)
}

```

Para apresentar as medidas resumo de uma variável com mais de 1 categoria, será utilizado o seguinte quadro:

Quadro 1: Medidas resumo da(o) [nome da variável]

Estatística	2seater	compact	midsize	minivan	pickup	subcompact	suv
Média	6,16	2,33	2,92	3,39	4,42	2,66	4,46
Desvio Padrão	0,53	0,45	0,72	0,45	0,83	1,10	1,07
Variância	0,28	0,20	0,52	0,20	0,69	1,22	1,14
Mínimo	5,70	1,80	1,80	2,40	2,70	1,60	2,50
1º Quartil	5,70	2,00	2,40	3,30	3,90	1,90	4,00
Mediana	6,20	2,20	2,80	3,30	4,70	2,20	4,65
3º Quartil	6,20	2,80	3,50	3,80	4,70	3,25	5,30
Máximo	7,00	3,30	5,30	4,00	5,90	5,40	6,50

O **Quadro 1** foi gerado com o seguinte código.

```

mpg %>%
  group_by(class) %>% # caso mais de uma categoria
  print_quadro_resumo(var_name = displ)

```

Para os casos em que apenas uma categoria será analisada o código pode ser readaptado para:

```

mpg %>%
  print_quadro_resumo(var_name = "displ")

```

Gerando o **Quadro 2**.

Quadro 2: Medidas resumo da(o) [nome da variável]

Estatística	Valor
Média	3,47
Desvio Padrão	1,29
Variância	1,67
Mínimo	1,60
1º Quartil	2,40
Mediana	3,30
3º Quartil	4,60
Máximo	7,00

1. **Observação:** pode-se mudar o nome e a `\label{}` do quadro como parâmetros da função;
2. **Observação:** Não se esqueça de mudar o nome e a `\label{}`;
3. **Observação:** É possível a adição do coeficiente de variação e outras medidas ao quadro.
4. **Observação:** O alinhamento das colunas, por padrão está à direita com o indicador S que já padroniza todos os números.

### 6.2.2 Quadros para testes

Para apresentar resultados de testes de hipóteses, devem ser construídos quadros, em que essas informações estarão descritas. Para todo teste, sempre deve-se apresentar as variáveis envolvidas, o valor do p-valor e a interpretação (resultado) deste.

O título do quadro deve seguir o exemplo "P-valor do teste de associação (Teste Exato de Fisher) entre as variáveis X e Y". Abaixo, segue dois quadros para servir de exemplo:

Quadro 3: P-valor do teste de associação (Teste Exato de Fisher) entre as variáveis X e Y

Grupo	P-valor	Decisão do teste
X Y	<0,001	Rejeita $H_0$

Quadro 4: P-valor do teste de normalidade (Teste de Shapiro-Wilk) de X

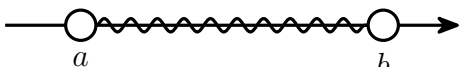
Variável	P-valor	Decisão do teste
X	0,175	Não rejeita $H_0$

Quadro 5: P-valor dos testes de normalidade (Shapiro-Wilk) e comparação de médias (teste T de comparação de médias) de X por Y

Variáveis	Normalidade	Decisão do teste	Comparação de médias	Decisão do teste
X - Categoria 1 de Y X - Categoria 2 de Y	0,070	Não rejeita $H_0$	0,240	Não Rejeita $H_0$

### 6.2.3 Quadro de intervalos de confiança

Os quadros para os intervalos de confiança devem mostrar o nível de confiança do intervalo, o parâmetro que está sendo estimado e os limites inferior e superior do intervalo.

Quadro 6: Intervalo de confiança para [parâmetro]	
Parâmetro	Intervalo de Confiança (95%)
Média da Variável X	

### 6.2.4 Quadro de coeficientes

Ainda não há padronização, recomenda-se por enquanto citar o coeficiente em texto ao explicá-lo, mas lembre-se de usar a simbologia correta. Por exemplo,  $\rho = 0,87$ .

### 6.2.5 Observação

No caso de quadros grandes, que ultrapassem a margem, usar o comando `\resizebox{\textwidth}{!}{} e dentro dos {} colocar o \begin{tabular} e \end{tabular}.`

## 6.3 Hipóteses dos testes

Sempre que um teste for realizado, suas hipóteses devem ser apresentadas. Abaixo, segue um exemplo de conjunto de hipóteses.

$$\left\{ \begin{array}{l} H_0 : \text{A diferença entre homens e mulheres segue uma distribuição simétrica em torno de zero} \\ H_1 : \text{A diferença entre homens e mulheres não segue uma distribuição simétrica em torno de zero} \end{array} \right.$$

Não é necessário apresentar as hipóteses de testes dos pressupostos, como normalidade ou homoscedasticidade. Basta mencionar em texto que os pressupostos do teste foram atendidos (quando forem) ou não, e seguir com a apresentação das hipóteses do teste realizado. Deve-se também, deixar claro quais são as hipóteses em texto, exemplo: “[...] As hipóteses testadas são:”(seguir com padronização apresentada nesta seção), uma vez que o cliente pode não conhecer a notação que é usada

para apresentação das hipóteses, assim busca-se deixar claro que as hipóteses estão contidas nas chaves.

## 6.4 Sequência dos gráficos e tabelas/quadros

Sempre que for necessária a presença de gráfico(s), tabela(s) e/o quadro(s), eles deverão seguir uma ordem específica. É importante, também, lembrar de sempre ter um texto introdutório sobre o que será feito e as análises daquela subseção, a ordem deve ser a seguinte:

1. Texto introdutório
2. Gráfico
3. Tabela
4. Quadro
5. Análise

## 6.5 Casas decimais

Sempre que forem completar tabelas, quadros, gráficos, e fazerem as análises, o número de casas decimais deverá seguir o seguinte padrão:

- Frequência absoluta : sempre que possível, ser o número absoluto;
- Frequência relativa : **duas** casas decimais. Exemplo: 5,65%, 89,90%;
- Medidas resumo:
  - Média : **duas** casas decimais;
  - Mediana : **duas** casas decimais;
  - Quartis : **duas** casas decimais;
  - Desvio padrão : **duas** casas decimais;
  - Máximo e mínimo : **duas** casas decimais;
- P-valor: **três** casas decimais.;
  - Se o p-valor for muito pequeno, não coloque 0. Use  $<0,001$ ;
- Estatística do teste: se necessário, utilize **duas** casas decimais.

### Observações:

- Separador de casas decimais é **vírgula** (3,45);
- Separador de milhar é **ponto** (3.234,93).

## 6.6 Dicas

### 6.6.1 Comando `label{ }`

O site a seguir explica muito bem as utilizações do comando `label{ }` e também fornece exemplos:

[https://pt.wikibooks.org/wiki/Latex/Refer%C3%A2ncias\\_internas](https://pt.wikibooks.org/wiki/Latex/Refer%C3%A2ncias_internas)

### 6.6.2 Demais tabelas e quadros

Para as tabelas e quadros que não se encontrem nesta padronização, pode-se construir manualmente uma tabela, utilizando o site a seguir:

<https://www.tablesgenerator.com/>

**IMPORTANTE:** Se gerar alguma tabela pelo <https://www.tablesgenerator.com/> lembre-se de substituir os `\hline` pelos comandos `\toprule`, `\midrule` e `\bottomrule` para as linhas horizontais da tabela.

### 6.6.3 Aspas duplas

O método padrão do  $\text{\LaTeX}$  para colocar algo entre aspas é colocando o seu texto entre `` `` e `' '` da seguinte forma `` `exemplo' '` que tem resultado “exemplo”.