

# Desafio Boticário

Reuben Solomon Katz T22.1

22/11/2019

## Introdução

O boticário é uma das maiores empresas cosméticas no Brasil. Possuindo mais de 4000 lojas no mundo inteiro. O Boticário faz parte de uma empresa denominada Grupo Boticário, o qual possui 5 grandes marcas:

- O Boticário;
- Eudora;
- Quem disse, Berenice?;
- The Beauty Box;
- Vult Cosméticos;

Esse grupo propôs um desafio de análise de dados, para assim encontrar informações relevantes acerca de suas vendas.

## Solução proposta

### Encontrando Variáveis Úteis

```
myData <- read.csv(  
  file="/home/groot/Estatistica_ITA/Desafio_Boticario/DESAFIO_ITA.csv",  
  header=TRUE,  
  sep=";"  
)  
  
myData$DIA = ymd(myData$DIA) ##transformar para o formato dia  
  
###colocando os dados por mês e por semana  
myData$MES <- as.Date(cut(myData$DIA,  
  breaks = "month"))  
myData$SEMANA <- as.Date(cut(myData$DIA,  
  breaks = "week",  
  start.on.monday = FALSE)) # a semana irá começar na segunda  
  
###pegar dados por região  
myDataNordeste <- myData[myData$REGIAO == "Região Nordeste",]  
myDataSudeste <- myData[myData$REGIAO == "Região Sudeste",]  
myDataSul <- myData[myData$REGIAO == "Região Sul",]  
  
myDataNordeste_0 <- myDataNordeste[myDataNordeste$COD_CATEGORIA == "0",]  
myDataNordeste_1 <- myDataNordeste[myDataNordeste$COD_CATEGORIA == "1",]  
myDataNordeste_2 <- myDataNordeste[myDataNordeste$COD_CATEGORIA == "2",]  
myDataSudeste_0 <- myDataSudeste[myDataSudeste$COD_CATEGORIA == "0",]
```

```

myDataSudeste_1 <- myDataSudeste[myDataSudeste$COD_CATEGORIA == "1",]
myDataSudeste_2 <- myDataSudeste[myDataSudeste$COD_CATEGORIA == "2",]

myDataSul_0 <- myDataSudeste[myDataSul$COD_CATEGORIA == "0",]
myDataSul_1 <- myDataSudeste[myDataSul$COD_CATEGORIA == "1",]
myDataSul_2 <- myDataSudeste[myDataSul$COD_CATEGORIA == "2",]

totalNordeste = sum(myDataNordeste$VOLUME)

totalSudeste = sum(myDataSudeste$VOLUME)

totalSul = sum(myDataSul$VOLUME)

totalNordeste_0 = sum(myDataNordeste_0$VOLUME)
totalNordeste_1 = sum(myDataNordeste_1$VOLUME)
totalNordeste_2 = sum(myDataNordeste_2$VOLUME)

totalSudeste_0 = sum(myDataSudeste_0$VOLUME)
totalSudeste_1 = sum(myDataSudeste_1$VOLUME)
totalSudeste_2 = sum(myDataSudeste_2$VOLUME)

totalSul_0 = sum(myDataSul_0$VOLUME)
totalSul_1 = sum(myDataSul_1$VOLUME)
totalSul_2 = sum(myDataSul_2$VOLUME)

estados = levels(droplevels(unique(myData$REGIAO.1)))

totalEstados = length(estados)

lojas = sort(unique(myData$COD_LOJA))

totalLojas = length(lojas)

materiais <- sort(unique(myData$COD_MATERIAL))

tamanhoMateriais <- length(materiais)

categorias <- sort(unique(myData$COD_CATEGORIA))

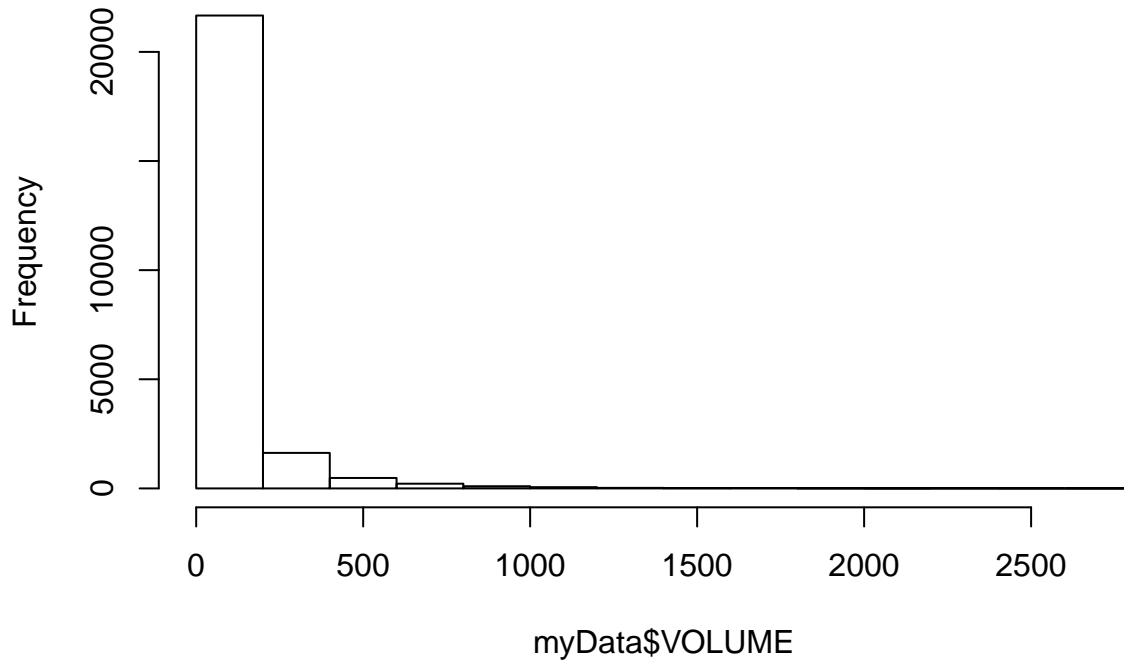
tamanhoCategorias <- length(categorias)

```

## Análise da distribuição da frequência

```
hist(myData$VOLUME)
```

**Histogram of myData\$VOLUME**



## Analisando o Resultado temporal dos produtos vendidos por região

```
#Separar por mês, região, produto

##Analisar cada em um nível regional e mensal

ggplot(data = myDataNordeste,
       aes(MES, VOLUME)) +
  stat_summary(fun.y = sum, # somar as observações mensais
              geom = "bar") +
  scale_x_date(
    labels = date_format("%Y-%m"),
    breaks = "1 month") + # colocar um "eixo-x" customizado

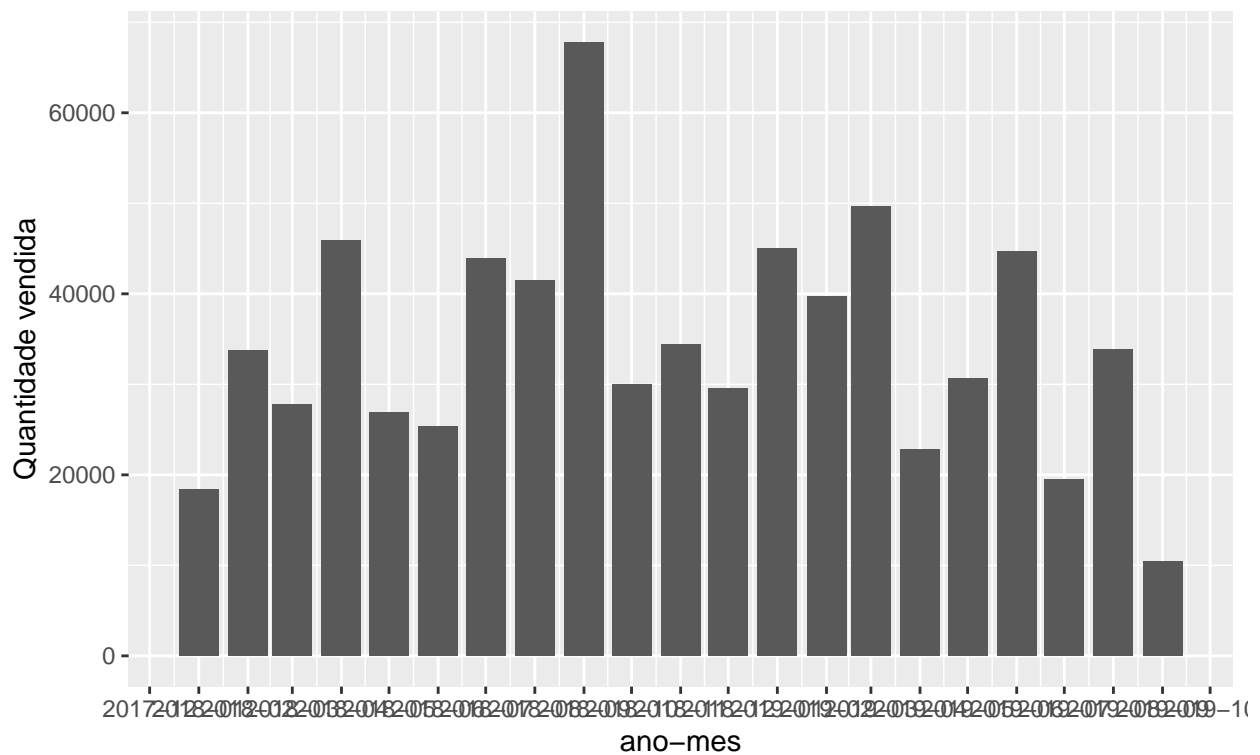
  labs( title="Quantidade total de produtos vendidos",
        subtitle="Região Nordeste"
  ) +

  xlab("ano-mes") +

  ylab("Quantidade vendida")
```

## Quantidade total de produtos vendidos

### Região Nordeste

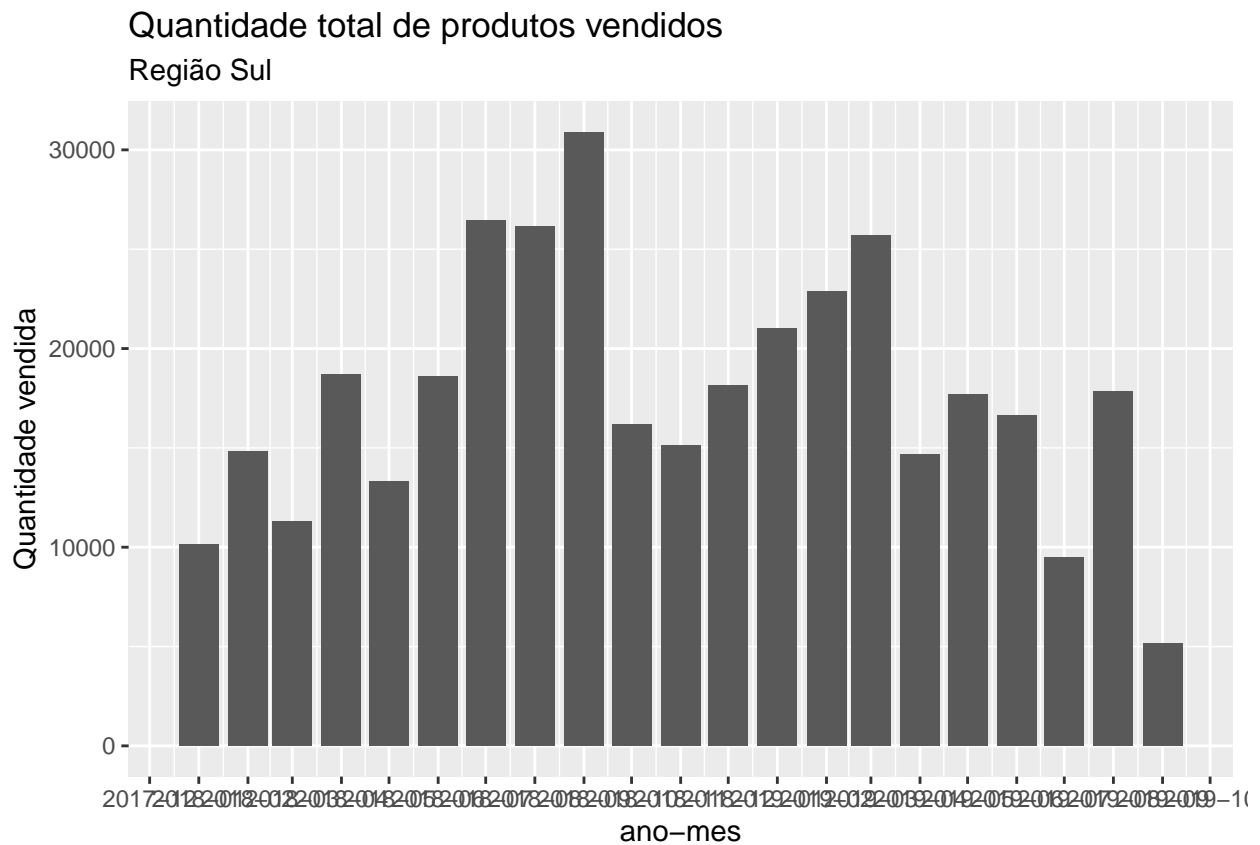


```
ggplot(data = myDataSul,
  aes(MES, VOLUME)) +
  stat_summary(fun.y = sum, # somar as observações mensais
    geom = "bar") +
  scale_x_date(
    labels = date_format("%Y-%m"),
    breaks = "1 month") + # colocar um "eixo-x" customizado

  labs( title="Quantidade total de produtos vendidos",
    subtitle="Região Sul"
  ) +

  xlab("ano-mes") +

  ylab("Quantidade vendida")
```



```
ggplot(data = myDataSudeste,
  aes(MES, VOLUME)) +
  stat_summary(fun.y = sum, # somar as observações mensais
    geom = "bar") +
  scale_x_date(
    labels = date_format("%Y-%m"),
    breaks = "1 month") + # colocar um "eixo-x" customizado

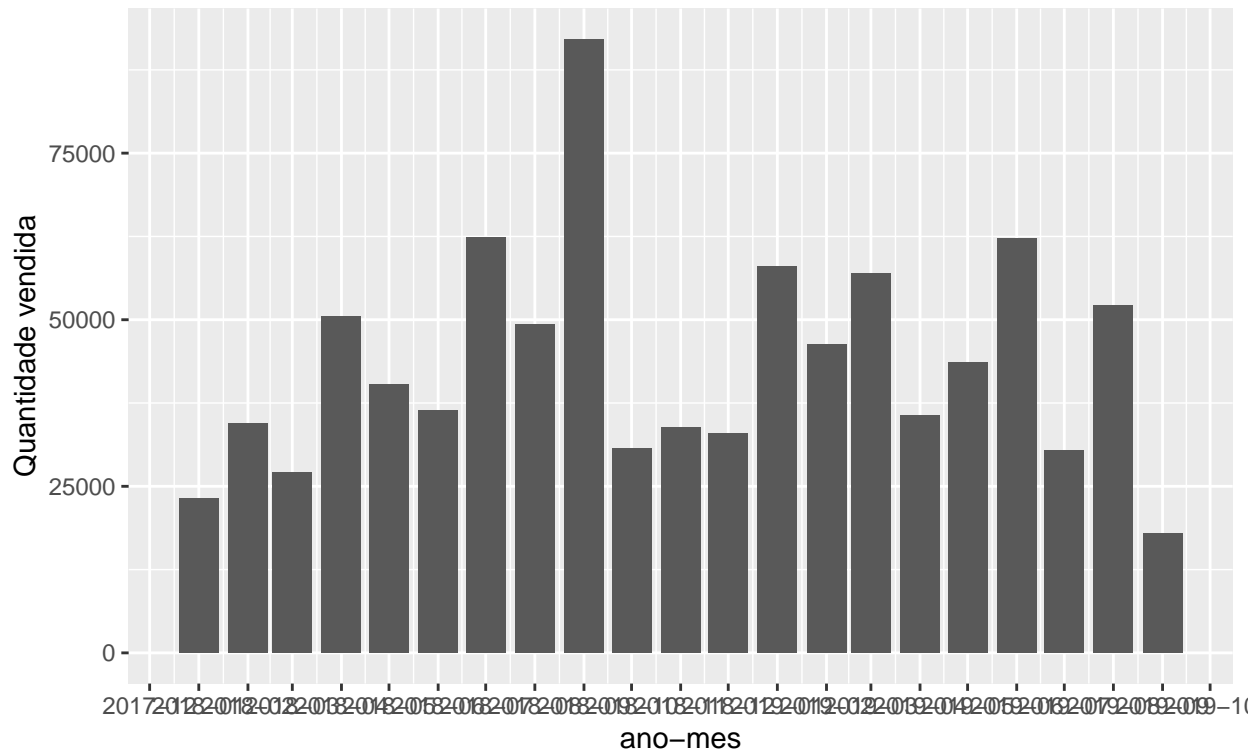
  labs( title="Quantidade total de produtos vendidos",
    subtitle="Região Sudeste"
  ) +

  xlab("ano-mes") +

  ylab("Quantidade vendida")
```

## Quantidade total de produtos vendidos

Região Sudeste



```
quantidadeProdutoPorEstados = c(4*totalEstados)

i = 1

for(estado in estados)
{
  dadosEstado = myData[myData$"REGIAO.1" == estado,]
  quantidadeProdutoPorEstados[i] = sum(dadosEstado$VOLUME)
  quantidadeProdutoPorEstados[i+1] = sum(dadosEstado[dadosEstado$COD_CATEGORIA == "0",]$VOLUME)
  quantidadeProdutoPorEstados[i+2] = sum(dadosEstado[dadosEstado$COD_CATEGORIA == "1",]$VOLUME)
  quantidadeProdutoPorEstados[i+3] = sum(dadosEstado[dadosEstado$COD_CATEGORIA == "2",]$VOLUME)
  i = i+4
}

tabelaVolumeEstadoPorProduto <- matrix(quantidadeProdutoPorEstados,ncol=4,byrow=TRUE)
colnames(tabelaVolumeEstadoPorProduto) <- c("Total", "Produto 0", "Produto 1", "Produto 2")
rownames(tabelaVolumeEstadoPorProduto) <- estados
tabelaVolumeEstadoPorProduto <- as.table(tabelaVolumeEstadoPorProduto)
kable(tabelaVolumeEstadoPorProduto)
```

	Total	Produto 0	Produto 1	Produto 2
CE	127677	57303	177	70197
MG	44499	26013	396	18090
PE	342051	89196	690	252165
PI	33870	14505	90	19275
PR	186171	54540	9372	122259
RJ	437721	212976	1974	222771

	Total	Produto 0	Produto 1	Produto 2
RS	12864	4350	303	8211
SC	171858	51180	4680	115998
SE	218295	88098	393	129804
SP	434418	178170	2376	253872

Perceba que estes histogramas são formas de mostrar quais produtos tem mais sucesso por região. Isso é uma forma de descobrir a **melhor forma distribuir o estoque** de produtos. Além disso, é uma forma de fazer previsão temporal de quanto será vendido. Perceba que no último mês de cada histograma, houve uma **queda brusca** em relação ao ano passado. Isso, para fim de análise, será considerado como **erro do dataset**. Apesar de o grupo boticário ter comprado a Vult por volta dessa época, a queda brusca não pode ser justificada apenas por esse fator.

## Analizando um Resultado de Produto Vendidos Por Região, Loja e Tipo de Produto

```
##Analisar de maneira regional, mensal e por produto

###Nordeste

myDataNordeste_0 <- myDataNordeste[myDataNordeste$COD_CATEGORIA == "0",]
myDataNordeste_1 <- myDataNordeste[myDataNordeste$COD_CATEGORIA == "1",]
myDataNordeste_2 <- myDataNordeste[myDataNordeste$COD_CATEGORIA == "2",]

ggplot(data = myDataNordeste_0,
  aes(MES, VOLUME)) +
  stat_summary(fun.y = sum, # somar as observações mensais
    geom = "bar") +
  scale_x_date(
    labels = date_format("%Y-%m"),
    breaks = "1 month") + # colocar um "eixo-x" customizado

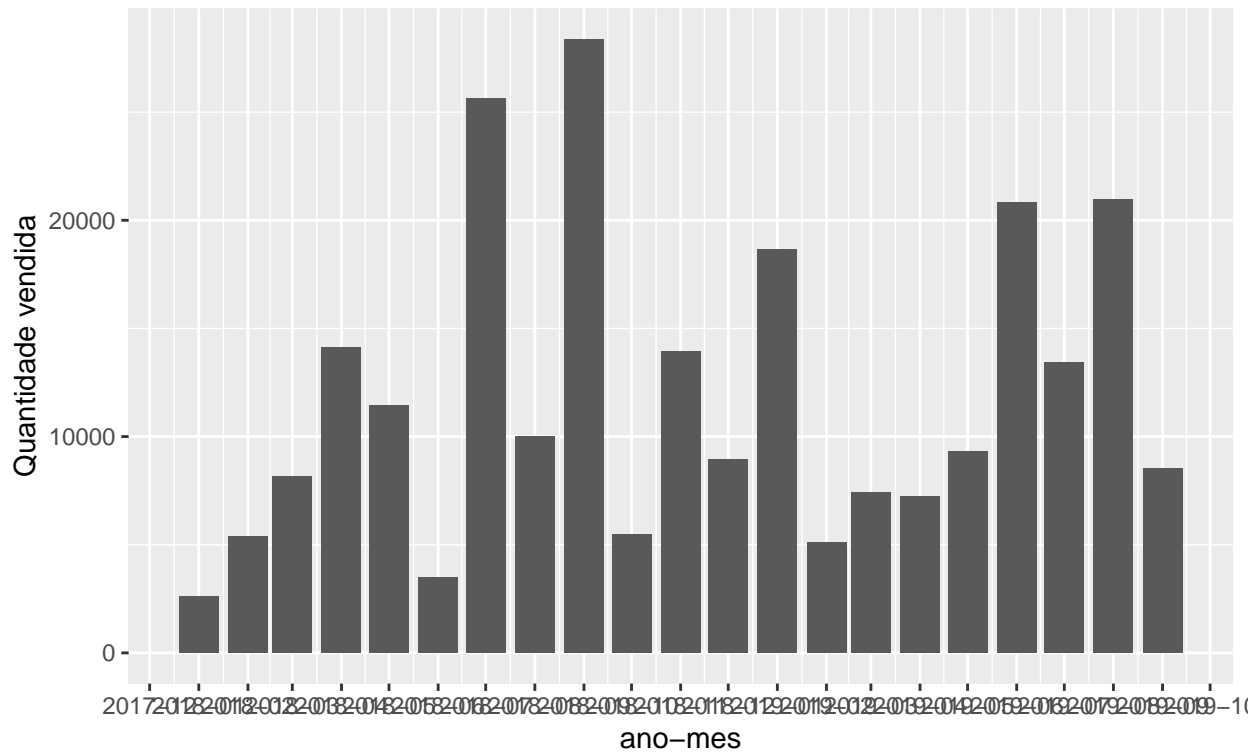
  labs( title="Quantidade vendidos produto 0",
    subtitle="Região Nordeste"
  ) +

  xlab("ano-mes") +

  ylab("Quantidade vendida")
```

## Quantidade vendidos produto 0

Região Nordeste

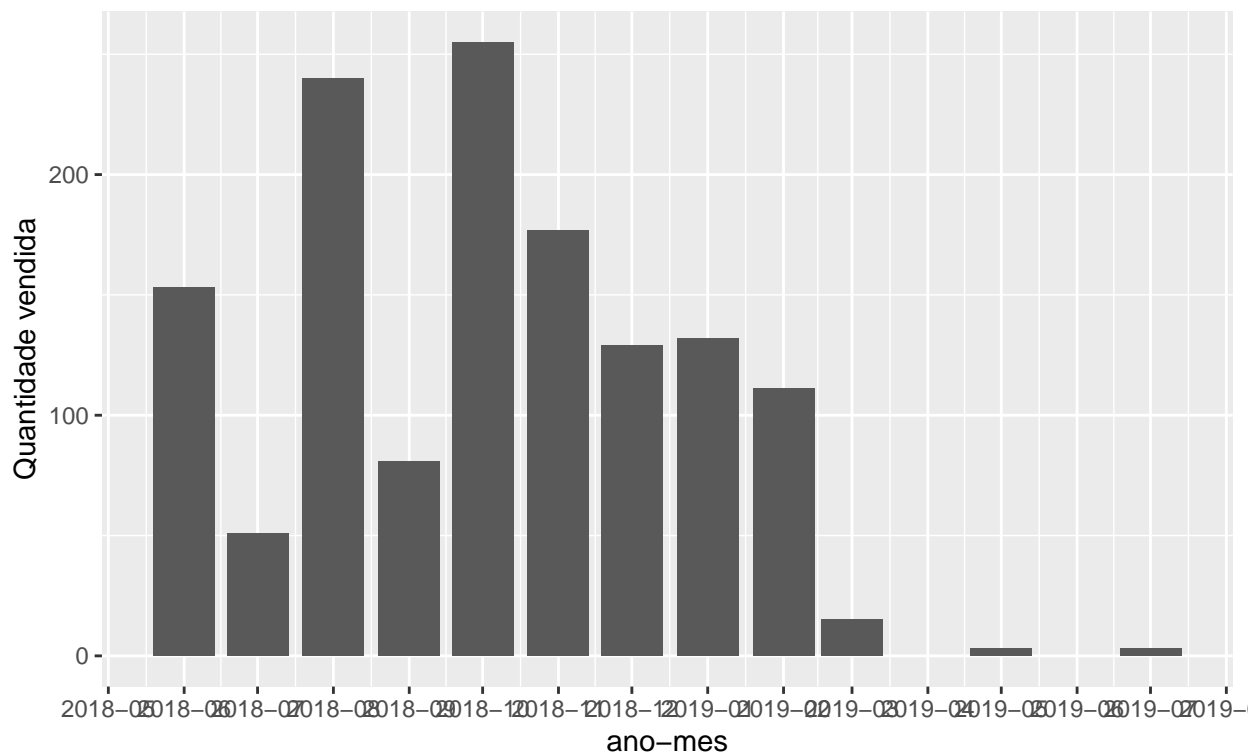


```
ggplot(data = myDataNordeste_1,  
  aes(MES, VOLUME)) +  
  stat_summary(fun.y = sum, # somar as observações mensais  
    geom = "bar") +  
  scale_x_date(  
    labels = date_format("%Y-%m"),  
    breaks = "1 month") + # colocar um "eixo-x" customizado  
  
  labs( title="Quantidade vendidos produto 1",  
    subtitle="Região Nordeste"  
  ) +  
  
  xlab("ano-mes") +  
  
  ylab("Quantidade vendida")
```



## Quantidade vendidos produto 1

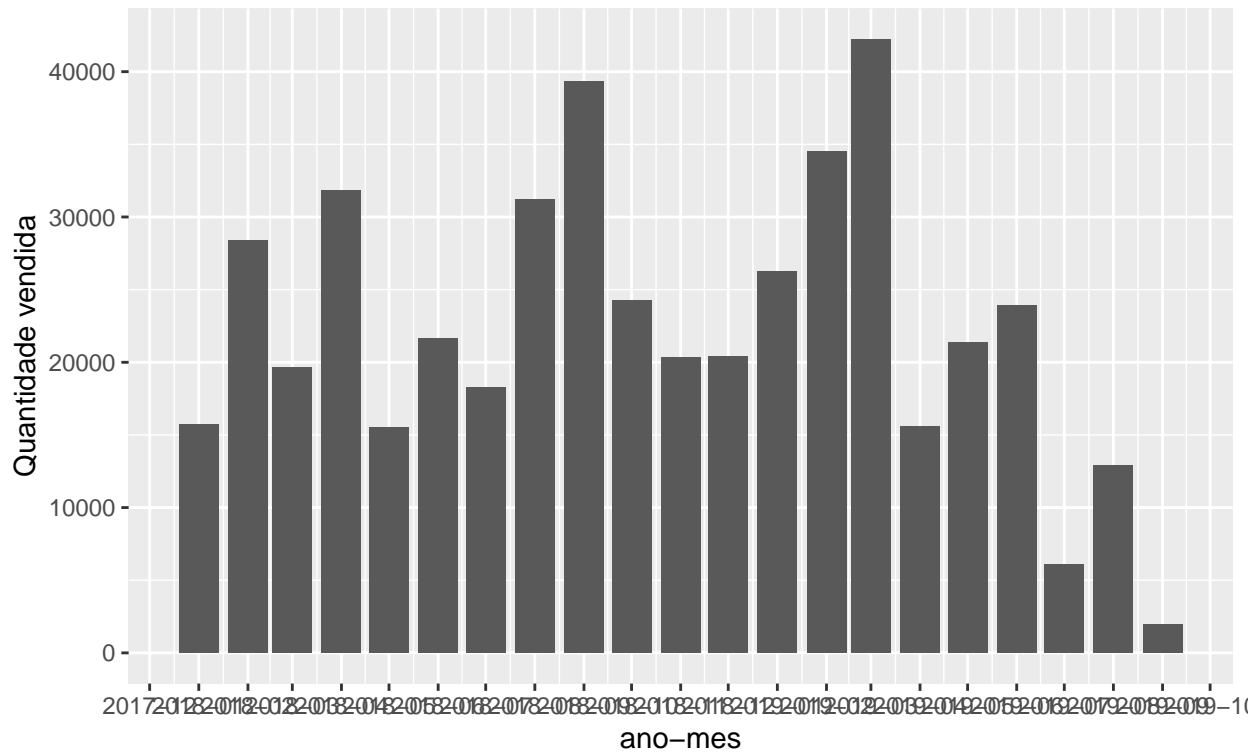
Região Nordeste



```
ggplot(data = myDataNordeste_2,  
  aes(MES, VOLUME)) +  
  stat_summary(fun.y = sum, # somar as observações mensais  
    geom = "bar") +  
  scale_x_date(  
    labels = date_format("%Y-%m"),  
    breaks = "1 month") + # colocar um "eixo-x" customizado  
  
  labs( title="Quantidade vendidos produto 2",  
    subtitle="Região Nordeste"  
  ) +  
  
  xlab("ano-mes") +  
  
  ylab("Quantidade vendida")
```

## Quantidade vendidos produto 2

Região Nordeste



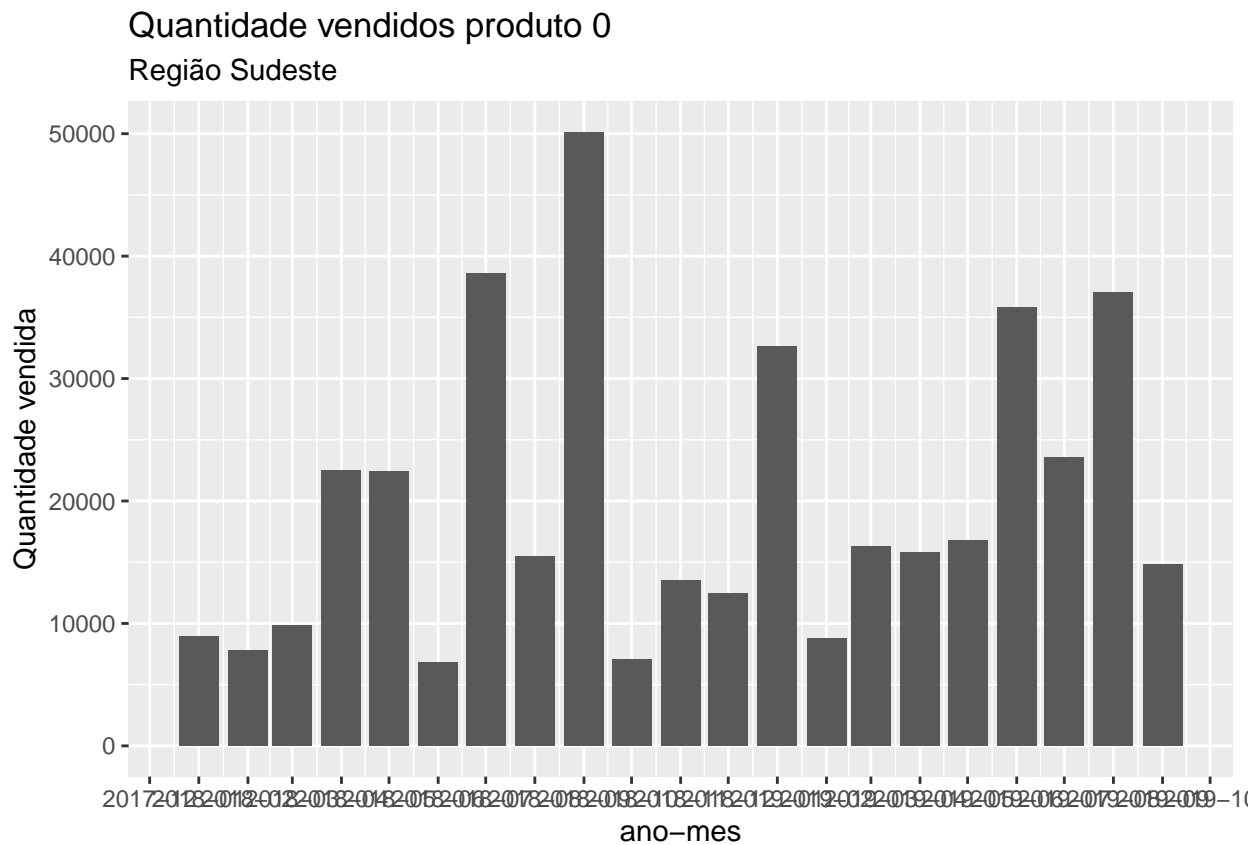
###Sudeste

```
myDataSudeste_0 <- myDataSudeste[myDataSudeste$COD_CATEGORIA == "0",]
```

```
myDataSudeste_1 <- myDataSudeste[myDataSudeste$COD_CATEGORIA == "1",]
```

```
myDataSudeste_2 <- myDataSudeste[myDataSudeste$COD_CATEGORIA == "2",]
```

```
ggplot(data = myDataSudeste_0,  
  aes(MES, VOLUME)) +  
  stat_summary(fun.y = sum, # somar as observações mensais  
    geom = "bar") +  
  scale_x_date(  
    labels = date_format("%Y-%m"),  
    breaks = "1 month") + # colocar um "eixo-x" customizado  
  
  labs( title="Quantidade vendidos produto 0",  
    subtitle="Região Sudeste"  
  ) +  
  
  xlab("ano-mes") +  
  
  ylab("Quantidade vendida")
```

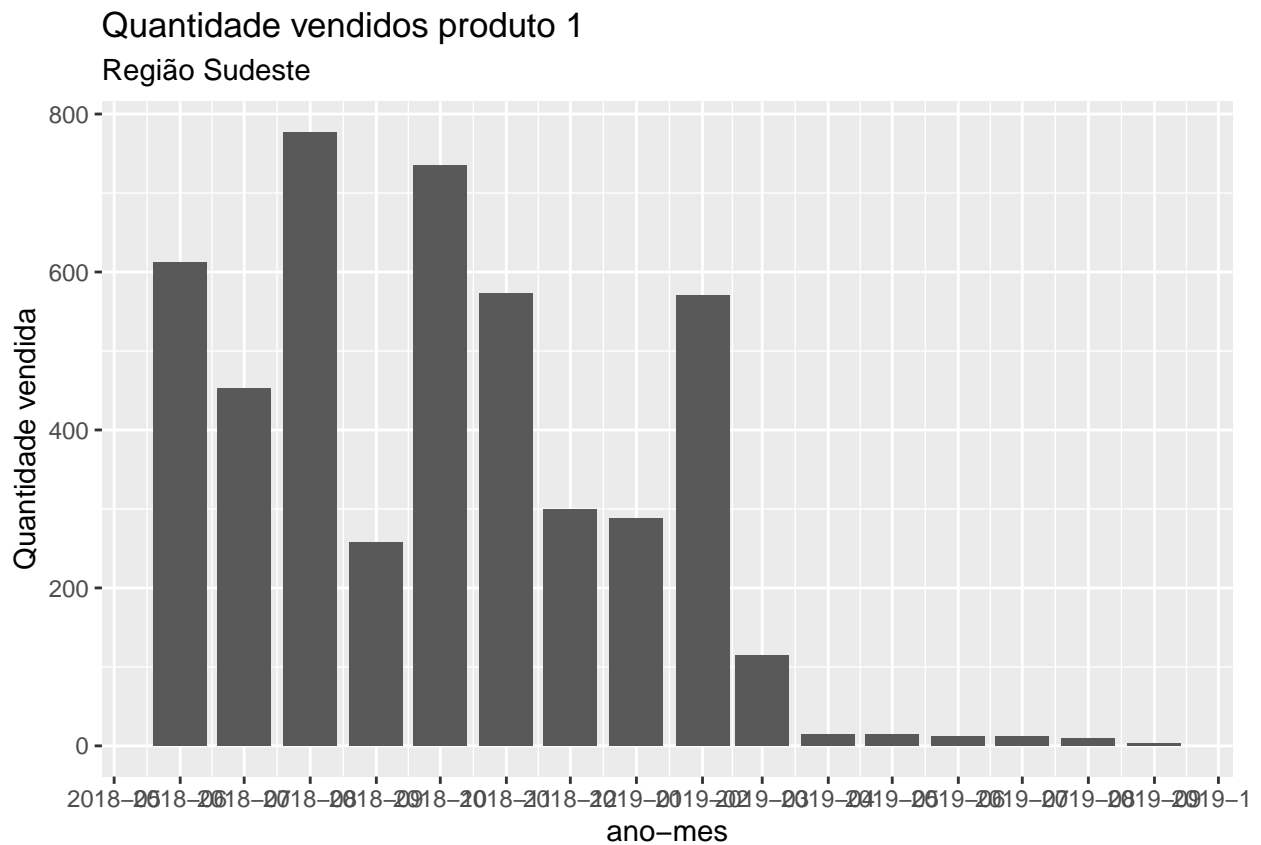


```
ggplot(data = myDataSudeste_1,
  aes(MES, VOLUME)) +
  stat_summary(fun.y = sum, # somar as observações mensais
    geom = "bar") +
  scale_x_date(
    labels = date_format("%Y-%m"),
    breaks = "1 month") + # colocar um "eixo-x" customizado

  labs( title="Quantidade vendidos produto 1",
    subtitle="Região Sudeste"
  ) +

  xlab("ano-mes") +

  ylab("Quantidade vendida")
```

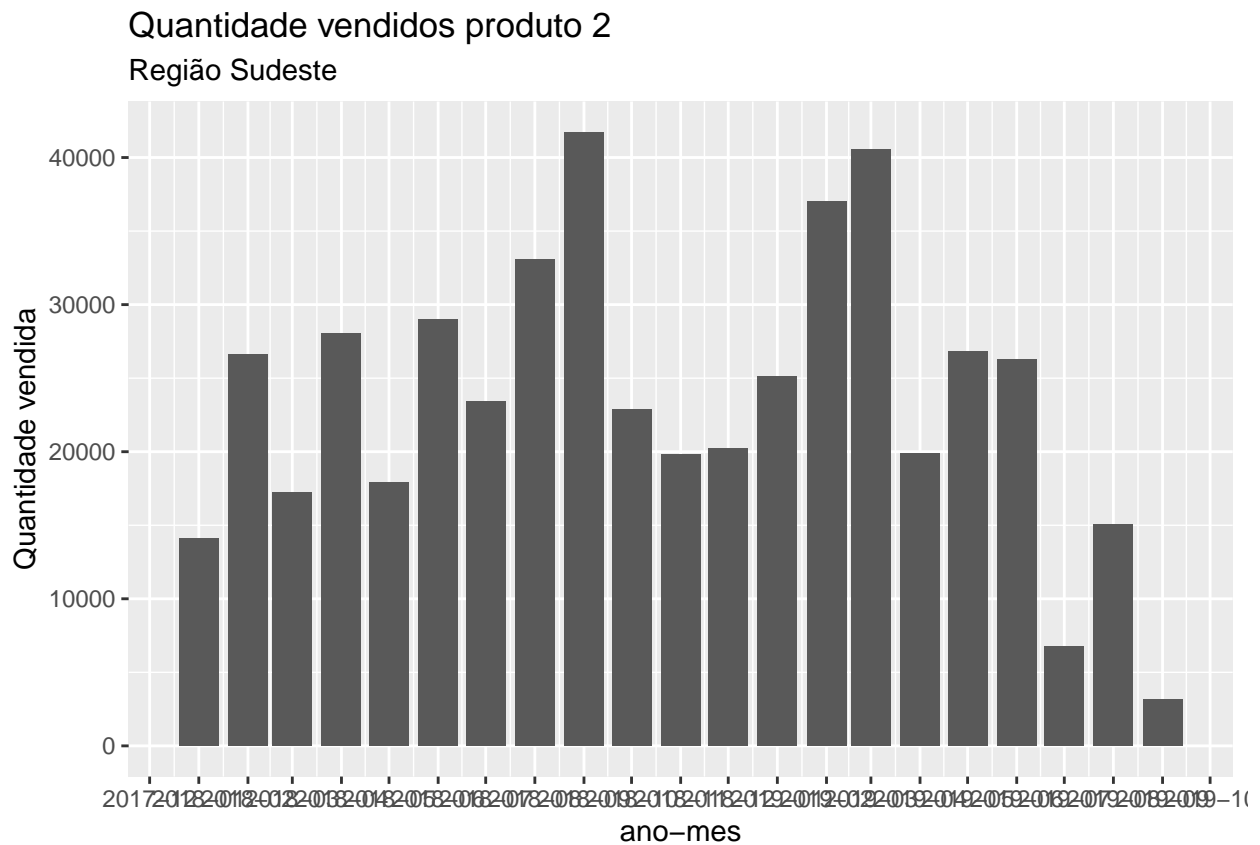


```
ggplot(data = myDataSudeste_2,
  aes(MES, VOLUME)) +
  stat_summary(fun.y = sum, # somar as observações mensais
    geom = "bar") +
  scale_x_date(
    labels = date_format("%Y-%m"),
    breaks = "1 month") + # colocar um "eixo-x" customizado

  labs( title="Quantidade vendidos produto 2",
    subtitle="Região Sudeste"
  ) +

  xlab("ano-mes") +

  ylab("Quantidade vendida")
```



```
###Sul

myDataSul_0 <- myDataSudeste[myDataSul$COD_CATEGORIA == "0",]
myDataSul_1 <- myDataSudeste[myDataSul$COD_CATEGORIA == "1",]
myDataSul_2 <- myDataSudeste[myDataSul$COD_CATEGORIA == "2",]

ggplot(data = myDataSul_0,
       aes(MES, VOLUME)) +
  stat_summary(fun.y = sum, # somar as observações mensais
              geom = "bar") +
  scale_x_date(
    labels = date_format("%Y-%m"),
    breaks = "1 month") + # colocar um "eixo-x" customizado

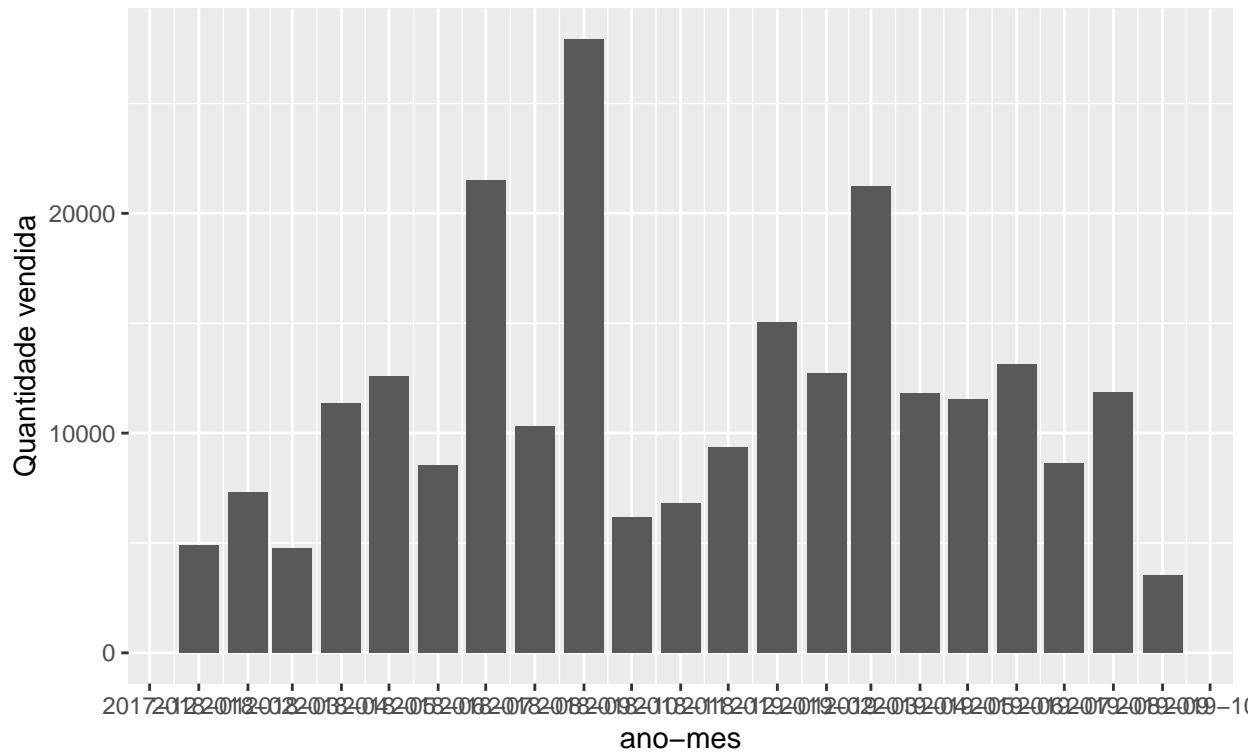
  labs( title="Quantidade vendidos produto 0",
        subtitle="Região Sul"
  ) +

  xlab("ano-mes") +

  ylab("Quantidade vendida")
```

# Quantidade vendidos produto 0

## Região Sul



```
ggplot(data = myDataSul_1,
  aes(MES, VOLUME)) +
  stat_summary(fun.y = sum, # somar as observações mensais
    geom = "bar") +
  scale_x_date(
    labels = date_format("%Y-%m"),
    breaks = "1 month") + # colocar um "eixo-x" customizado

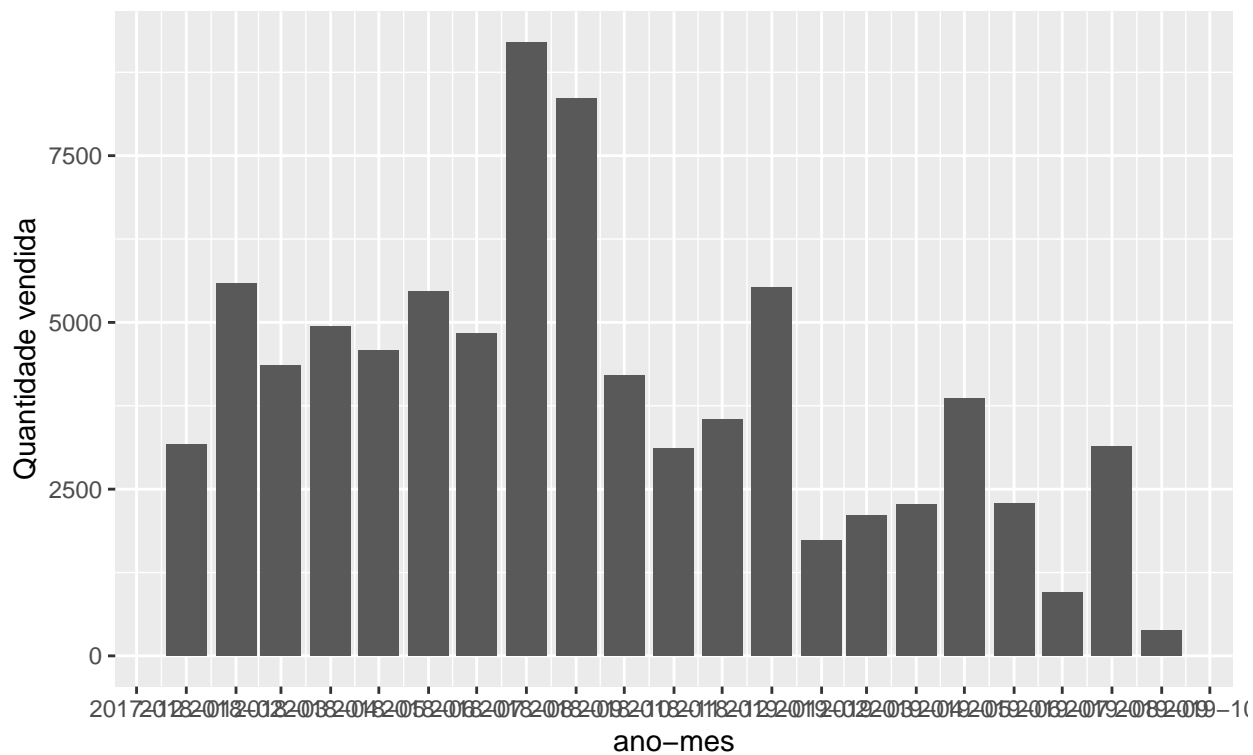
  labs( title="Quantidade vendidos produto 1",
    subtitle="Região Sul"
  ) +

  xlab("ano-mes") +

  ylab("Quantidade vendida")
```

## Quantidade vendidos produto 1

Região Sul

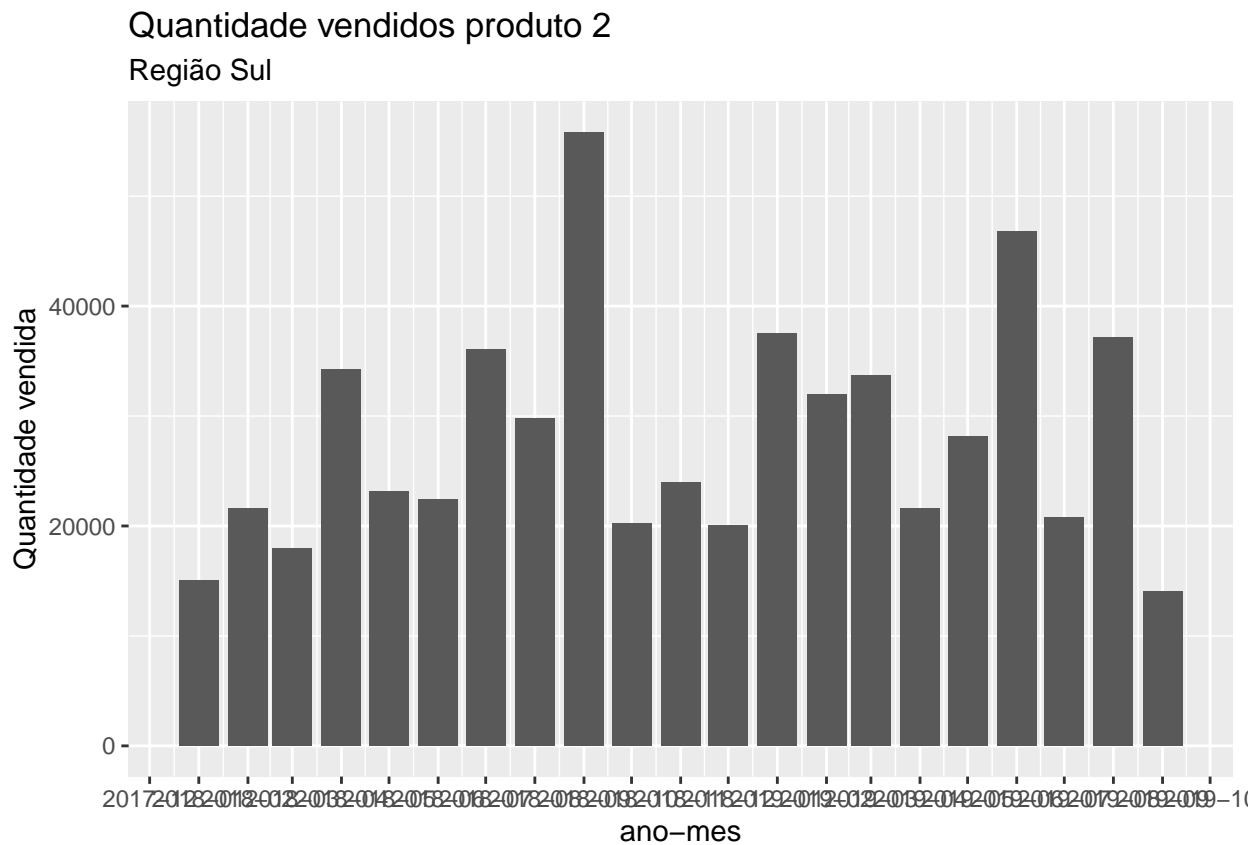


```
ggplot(data = myDataSul_2,
  aes(MES, VOLUME)) +
  stat_summary(fun.y = sum, # somar as observações mensais
    geom = "bar") +
  scale_x_date(
    labels = date_format("%Y-%m"),
    breaks = "1 month") + # colocar um "eixo-x" customizado

  labs( title="Quantidade vendidos produto 2",
    subtitle="Região Sul"
  ) +

  xlab("ano-mes") +

  ylab("Quantidade vendida")
```



Assim como demonstrando no tópico anterior, vemos aqui a melhor forma de distribuir cada tipo diferente de produto por região.

#### Análise por produto, loja e região

```
#Separar por loja, região e produto

##Analisar as lojas que vendem mais cada produto

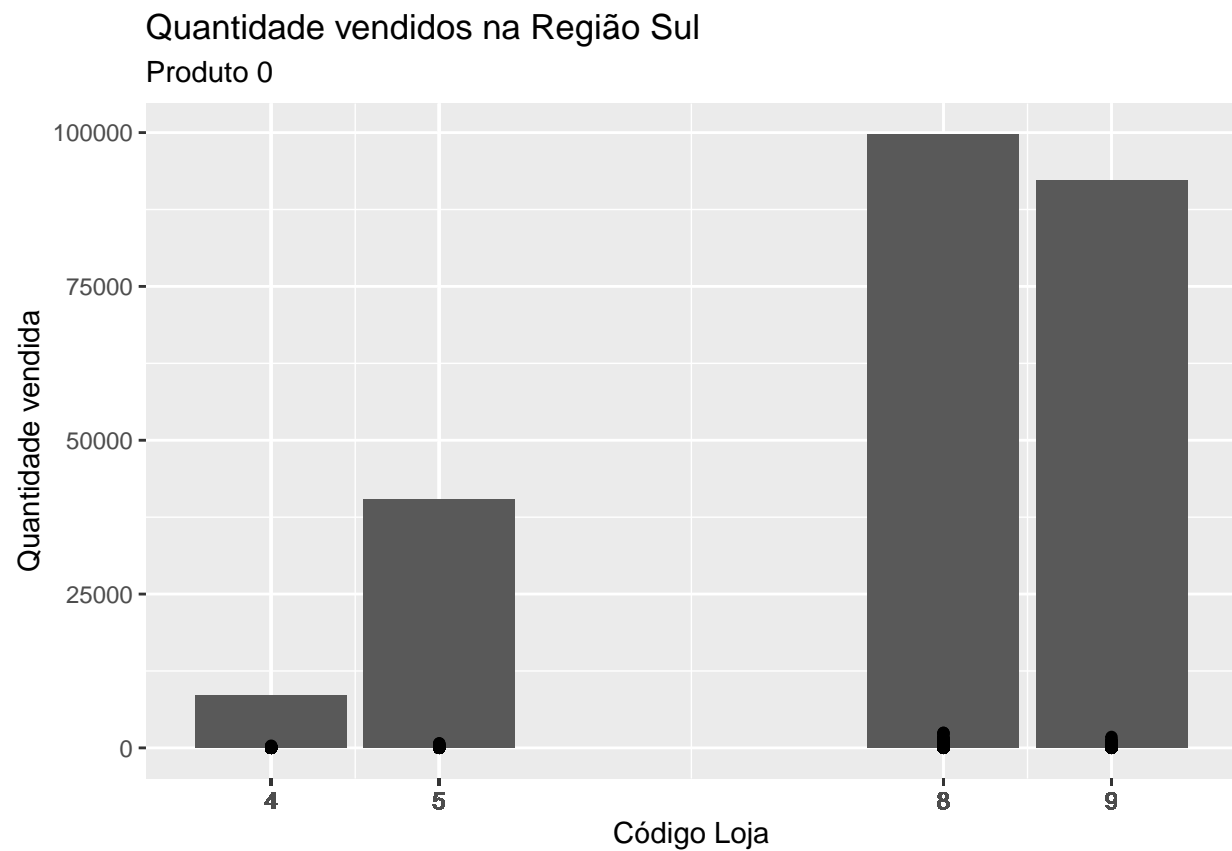
###Sul

ggplot(data = myDataSul_0,
  aes(x=COD_LOJA, y=VOLUME)) +
  stat_summary(fun.y = sum, # somar as observações para cada produto
    geom = "bar") +
  labs( title="Quantidade vendidos na Região Sul",
    subtitle="Produto 0"
  ) +

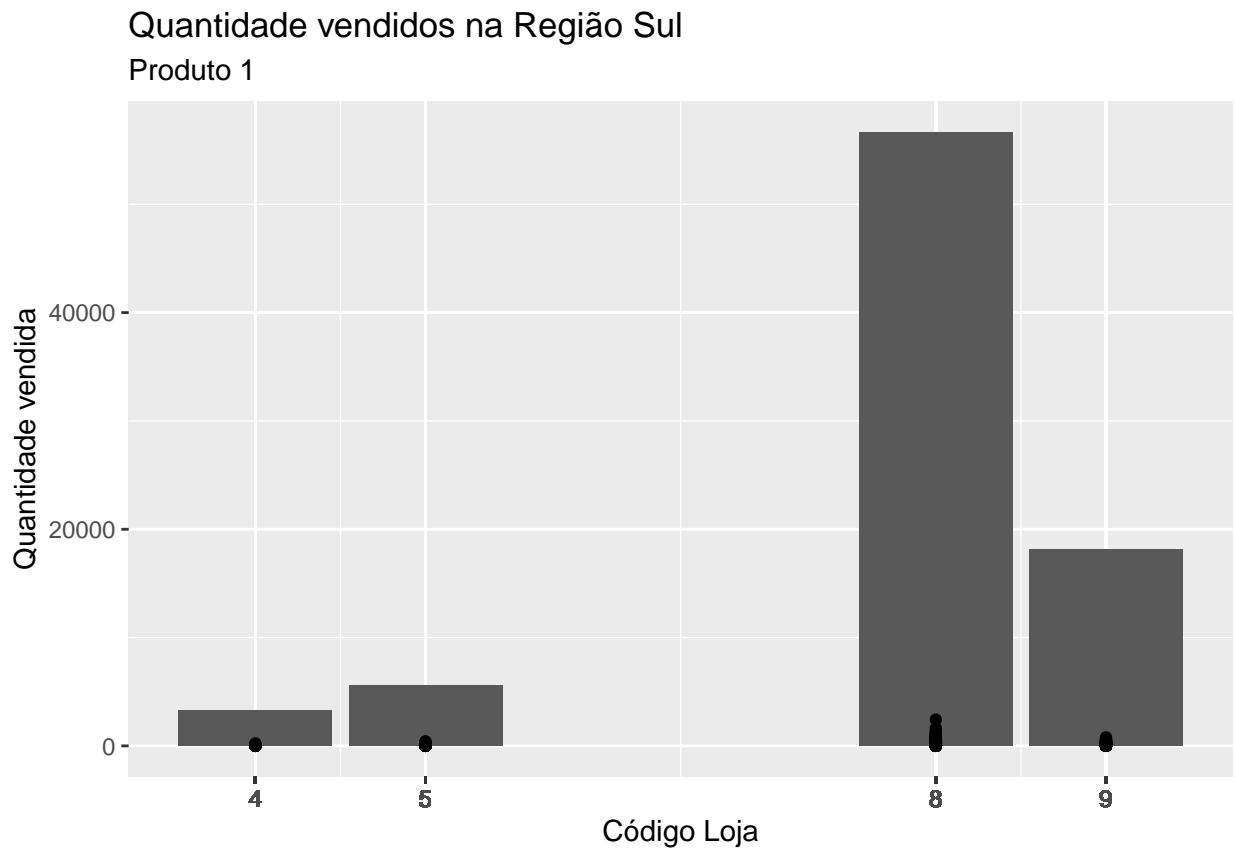
  ylab("Quantidade vendida") +

  scale_x_continuous("Código Loja", labels = as.character(myDataSul_0$COD_LOJA), breaks = myDataSul_0$COD_LOJA) +
  geom_point()
```





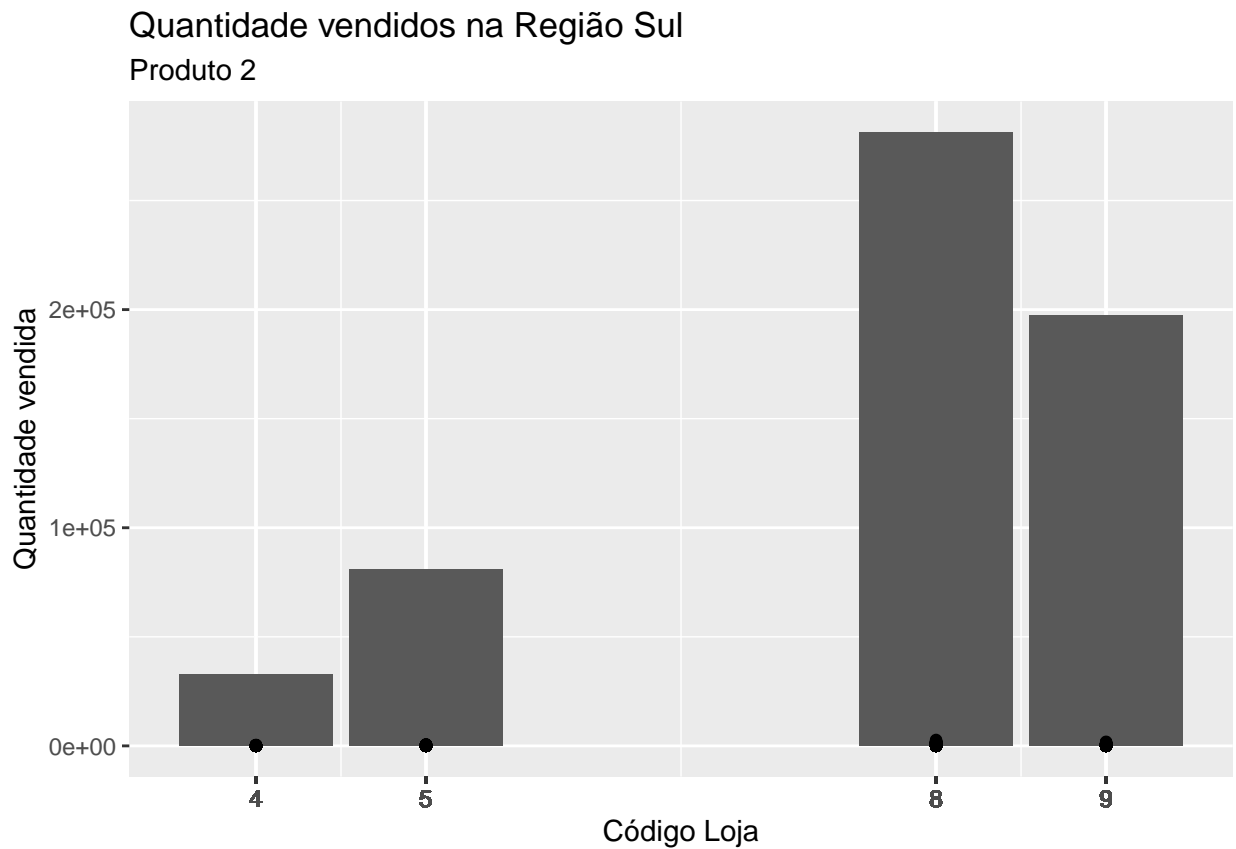
```
ggplot(data = myDataSul_1,  
  aes(x=COD_LOJA, y=VOLUME)) +  
  stat_summary(fun.y = sum, # somar as observações para cada produto  
    geom = "bar") +  
  labs( title="Quantidade vendidos na Região Sul",  
    subtitle="Produto 1"  
  ) +  
  
  ylab("Quantidade vendida") +  
  
  scale_x_continuous("Código Loja", labels = as.character(myDataSul_1$COD_LOJA), breaks = myDataSul_1$COD_LOJA) +  
  geom_point()
```



```
ggplot(data = myDataSul_2,
  aes(x=COD_LOJA, y=VOLUME)) +
  stat_summary(fun.y = sum, # somar as observações para cada produto
    geom = "bar") +
  labs( title="Quantidade vendidos na Região Sul",
    subtitle="Produto 2"
  ) +

  ylab("Quantidade vendida") +

  scale_x_continuous("Código Loja", labels = as.character(myDataSul_2$COD_LOJA), breaks = myDataSul_2$COD_LOJA) +
  geom_point()
```



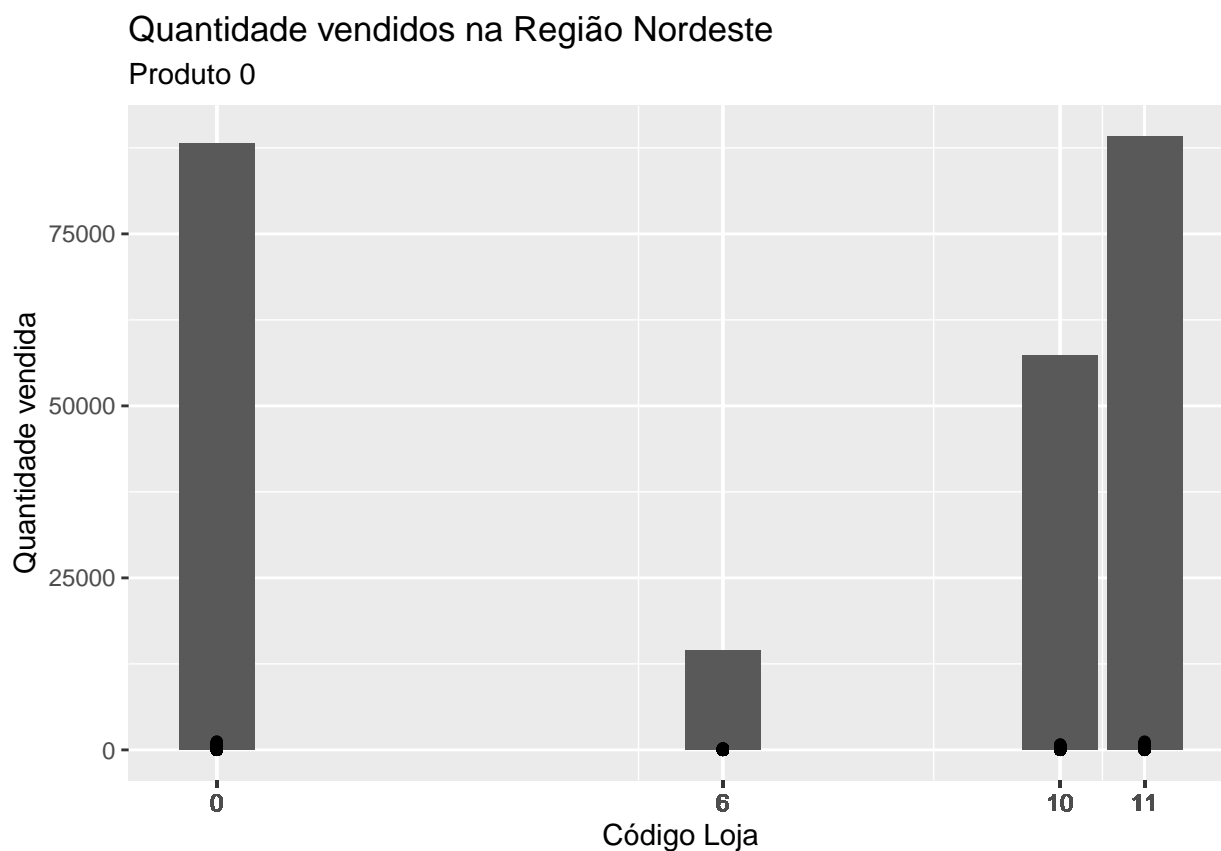
###Nordeste

```
ggplot(data = myDataNordeste_0,
  aes(x=COD_LOJA, y=VOLUME)) +
  stat_summary(fun.y = sum, # somar as observações para cada produto
    geom = "bar") +
  labs( title="Quantidade vendidos na Região Nordeste",
    subtitle="Produto 0"
  ) +

  ylab("Quantidade vendida") +

  scale_x_continuous("Código Loja", labels = as.character(myDataNordeste_0$COD_LOJA), breaks = myDataNo

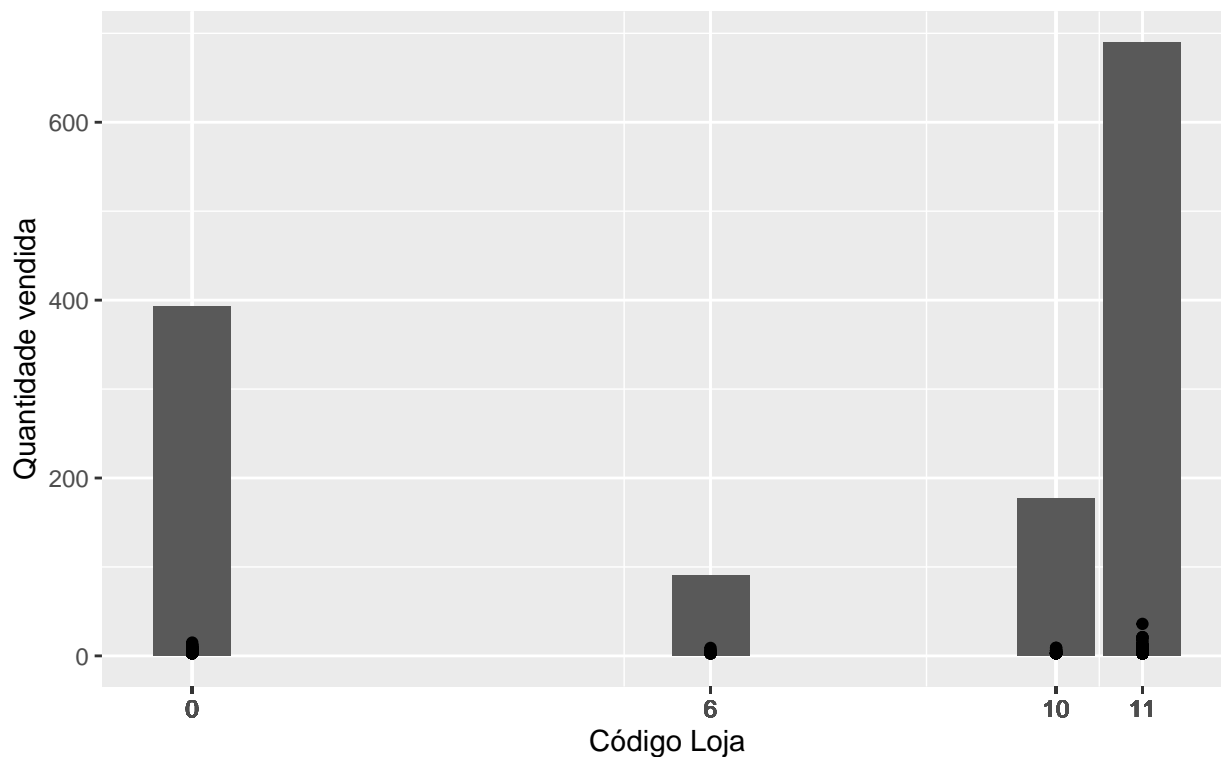
  geom_point()
```



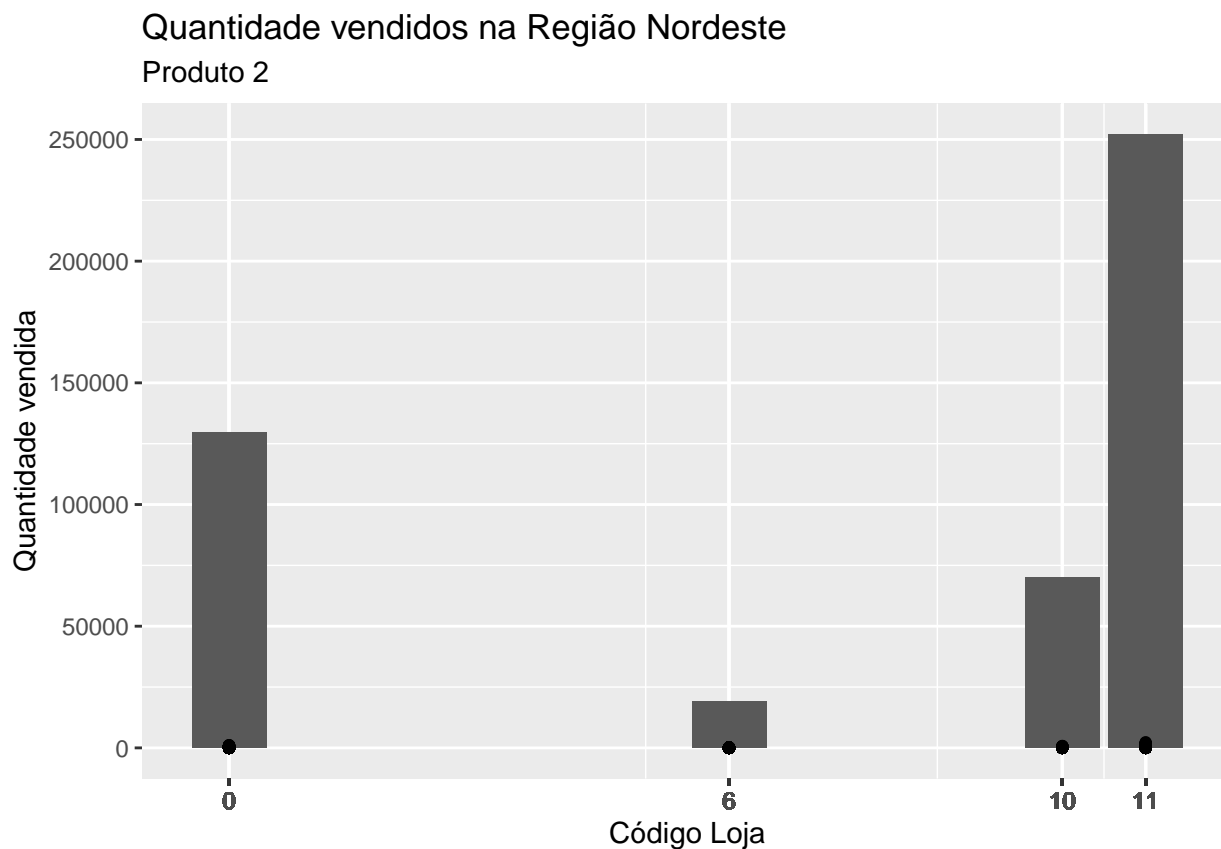
```
ggplot(data = myDataNordeste_1,  
  aes(x=COD_LOJA, y=VOLUME)) +  
  stat_summary(fun.y = sum, # somar as observações para cada produto  
    geom = "bar") +  
  labs( title="Quantidade vendidos na Região Nordeste",  
    subtitle="Produto 1"  
  ) +  
  
  ylab("Quantidade vendida") +  
  
  scale_x_continuous("Código Loja", labels = as.character(myDataNordeste_1$COD_LOJA), breaks = myDataNo  
  geom_point()
```

## Quantidade vendidos na Região Nordeste

Produto 1



```
ggplot(data = myDataNordeste_2,  
  aes(x=COD_LOJA, y=VOLUME)) +  
  stat_summary(fun.y = sum, # somar as observações para cada produto  
    geom = "bar") +  
  labs( title="Quantidade vendidos na Região Nordeste",  
    subtitle="Produto 2"  
  ) +  
  
  ylab("Quantidade vendida") +  
  
  scale_x_continuous("Código Loja", labels = as.character(myDataNordeste_2$COD_LOJA), breaks = myDataNo  
  geom_point()
```

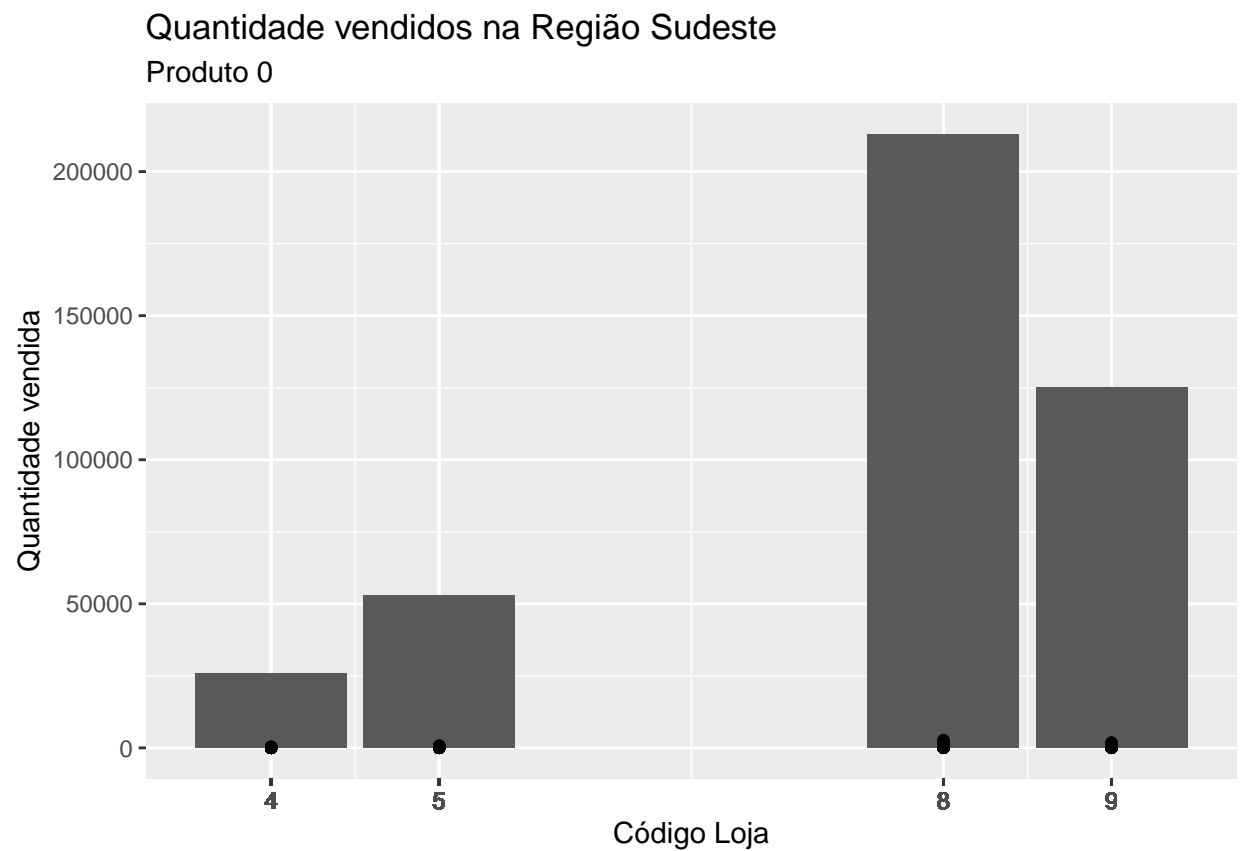


```
###Sudeste
```

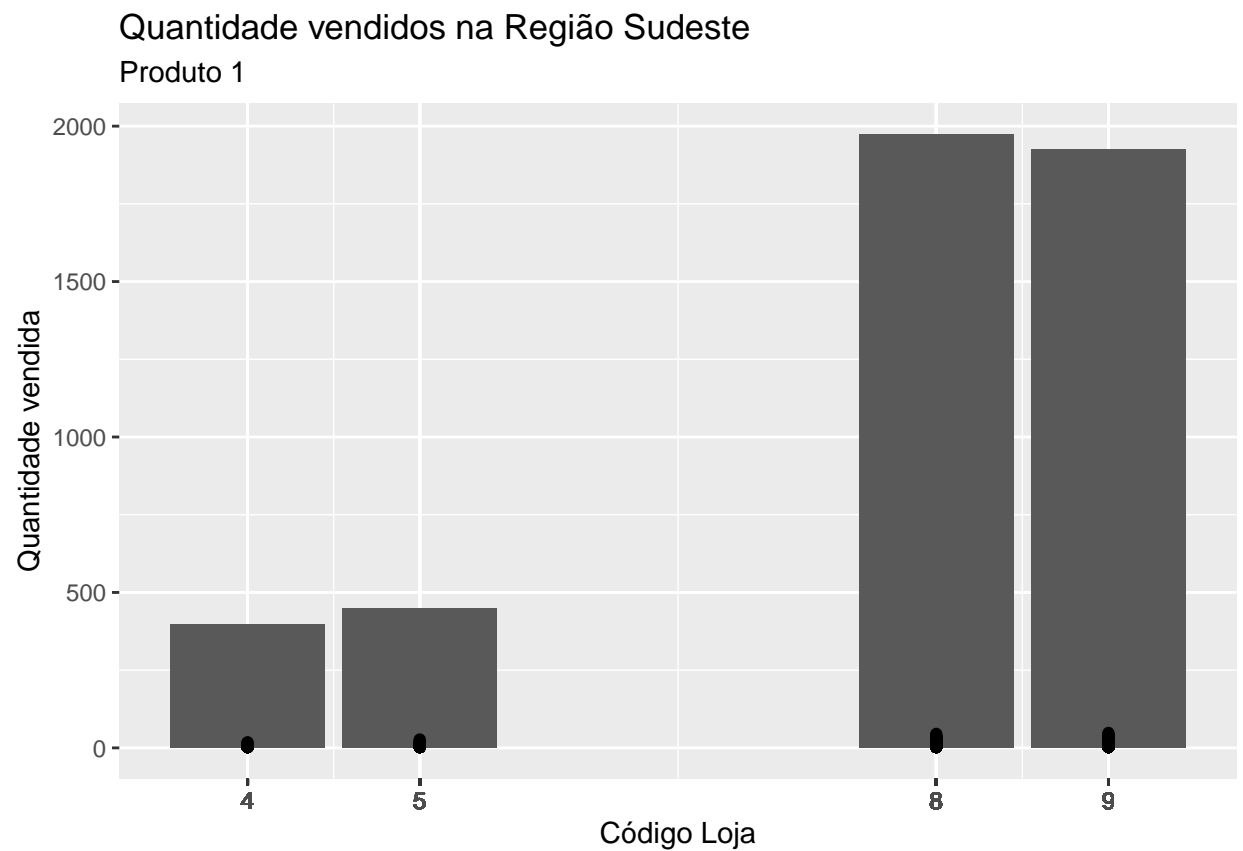
```
ggplot(data = myDataSudeste_0,
  aes(x=COD_LOJA, y=VOLUME)) +
  stat_summary(fun.y = sum, # somar as observações para cada produto
    geom = "bar") +
  labs( title="Quantidade vendidos na Região Sudeste",
    subtitle="Produto 0"
  ) +

  ylab("Quantidade vendida") +

  scale_x_continuous("Código Loja", labels = as.character(myDataSudeste_0$COD_LOJA), breaks = myDataSudeste_0$COD_LOJA) +
  geom_point()
```



```
ggplot(data = myDataSudeste_1,  
  aes(x=COD_LOJA, y=VOLUME)) +  
  stat_summary(fun.y = sum, # somar as observações para cada produto  
    geom = "bar") +  
  labs( title="Quantidade vendidos na Região Sudeste",  
    subtitle="Produto 1"  
  ) +  
  
  ylab("Quantidade vendida") +  
  
  scale_x_continuous("Código Loja", labels = as.character(myDataSudeste_1$COD_LOJA), breaks = myDataSudeste_1$COD_LOJA) +  
  geom_point()
```

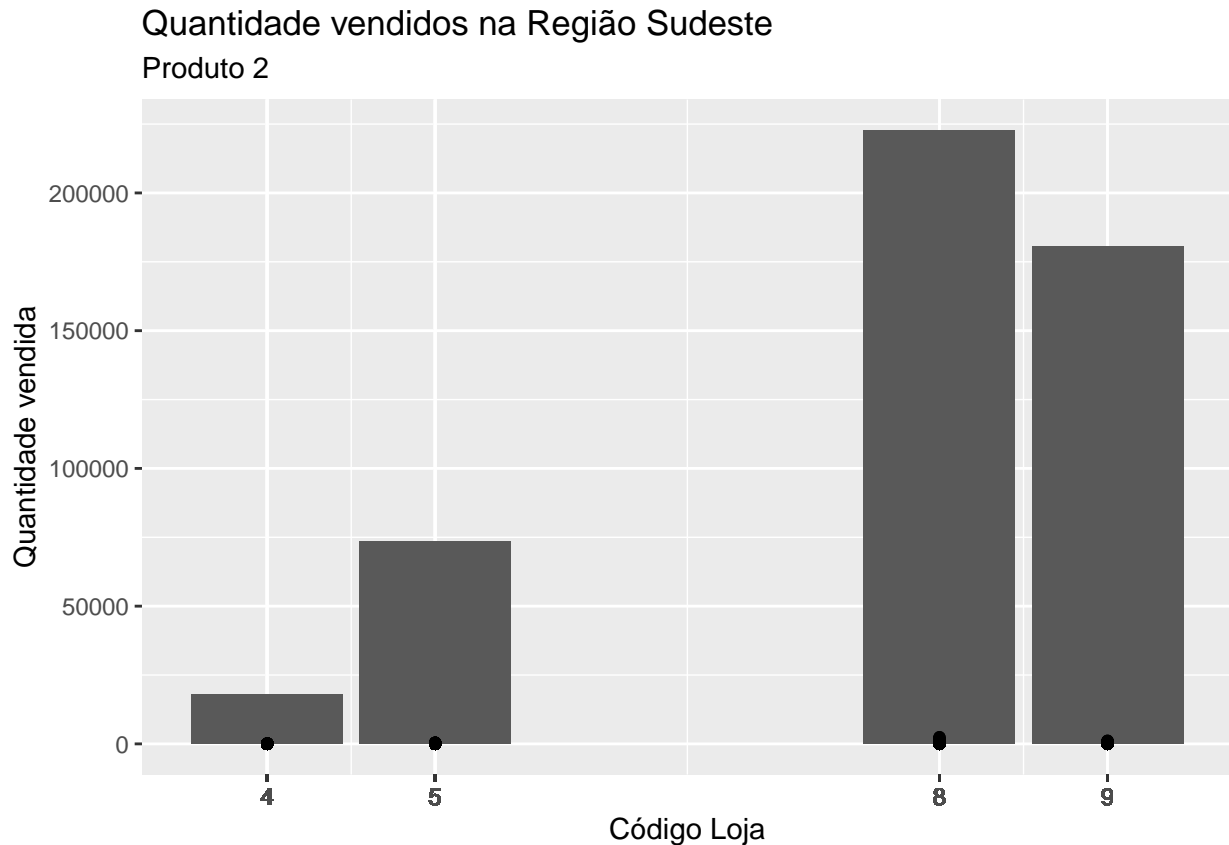


```
ggplot(data = myDataSudeste_2,
  aes(x=COD_LOJA, y=VOLUME)) +
  stat_summary(fun.y = sum, # somar as observações para cada produto
    geom = "bar") +
  labs( title="Quantidade vendidos na Região Sudeste",
    subtitle="Produto 2"
  ) +

  ylab("Quantidade vendida") +

  scale_x_continuous("Código Loja", labels = as.character(myDataSudeste_2$COD_LOJA), breaks = myDataSudeste_2$COD_LOJA) +
  geom_point()
```





Esses dados mostram quais são as melhores lojas de cada região. Perceba que esse dado mostra quais são **as lojas que podemos utilizar de modelo para aplicar em outras lojas** (caso viável), além de também mostrar a forma de distribuição do estoque dentro da região.

### Análise por Loja e Produto

```
#Separar por loja, região e produto

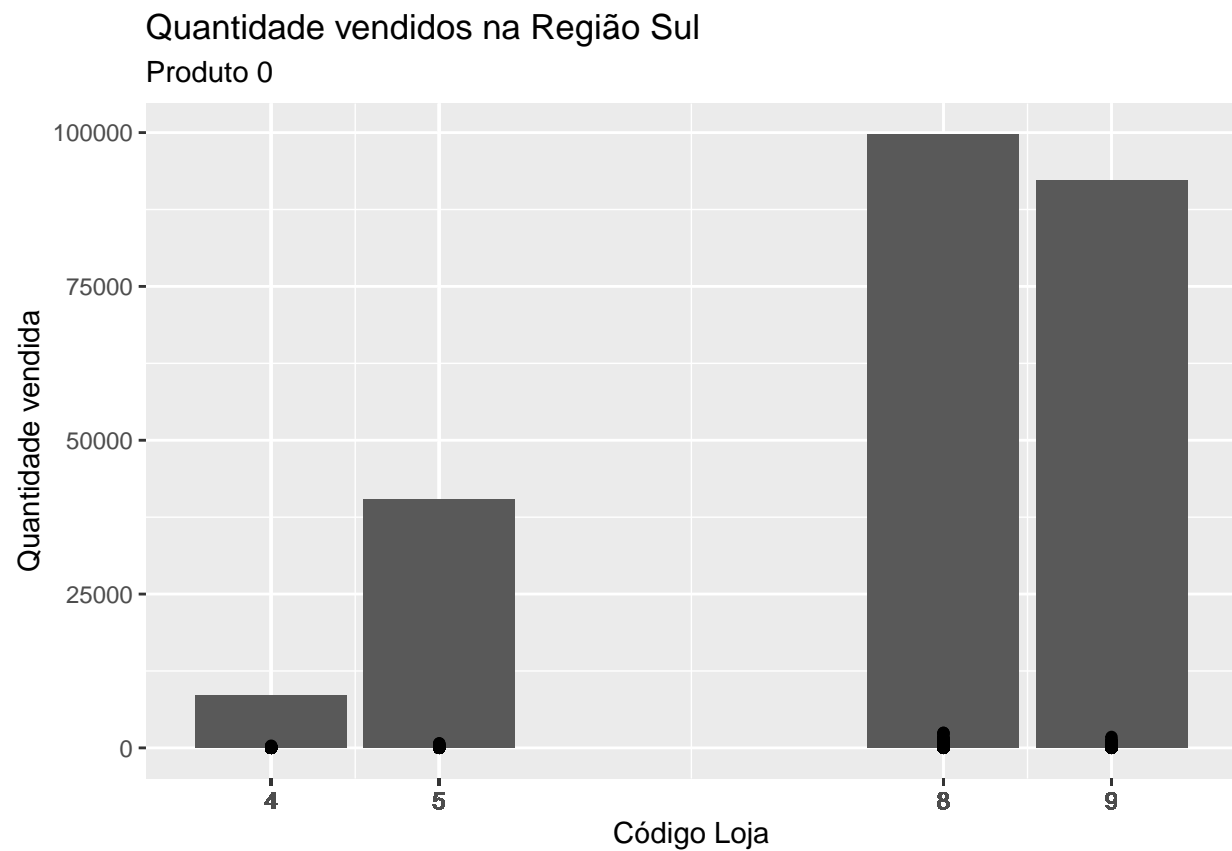
##Analisar as lojas que vendem mais cada produto

###Sul

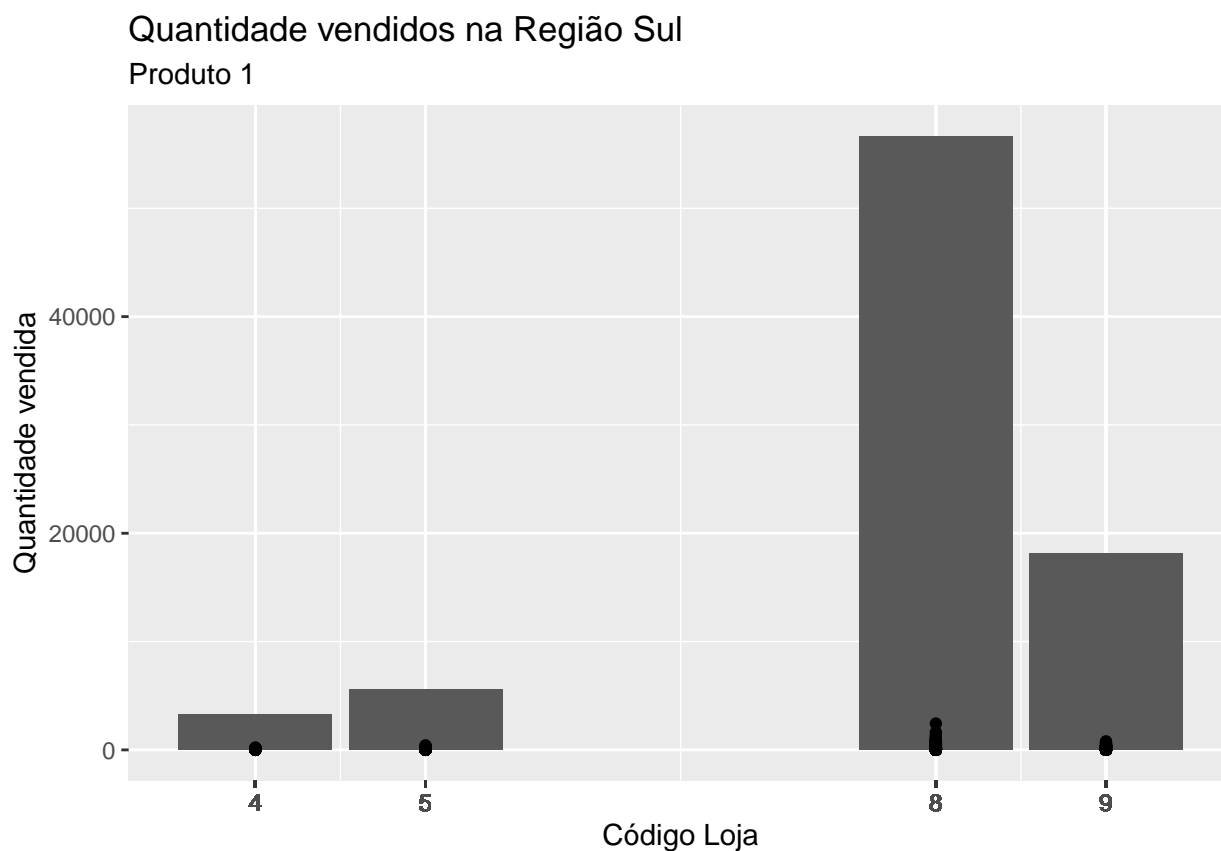
ggplot(data = myDataSul_0,
  aes(x=COD_LOJA, y=VOLUME)) +
  stat_summary(fun.y = sum, # somar as observações para cada produto
    geom = "bar") +
  labs( title="Quantidade vendidos na Região Sul",
    subtitle="Produto 0"
  ) +

  ylab("Quantidade vendida") +

  scale_x_continuous("Código Loja", labels = as.character(myDataSul_0$COD_LOJA), breaks = myDataSul_0$COD_LOJA) +
  geom_point()
```



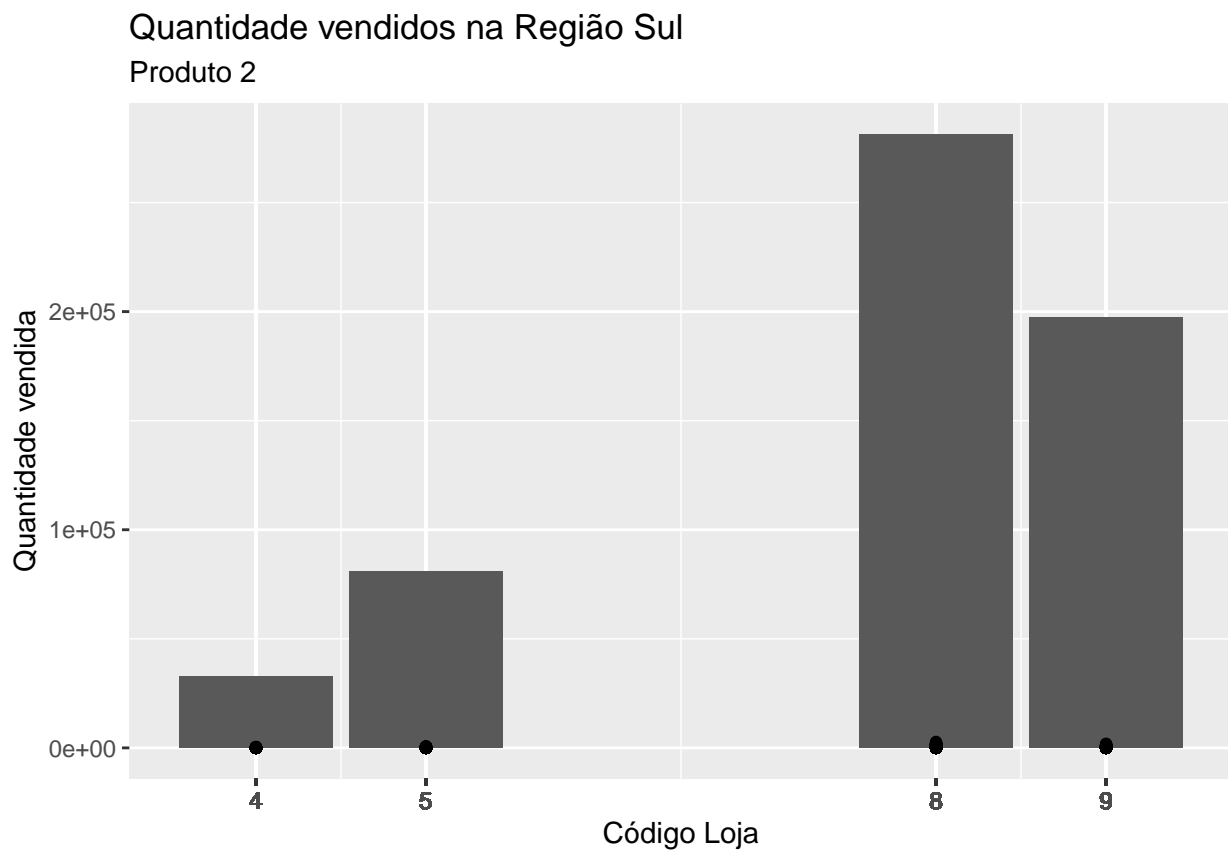
```
ggplot(data = myDataSul_1,  
  aes(x=COD_LOJA, y=VOLUME)) +  
  stat_summary(fun.y = sum, # somar as observações para cada produto  
    geom = "bar") +  
  labs( title="Quantidade vendidos na Região Sul",  
    subtitle="Produto 1"  
  ) +  
  
  ylab("Quantidade vendida") +  
  
  scale_x_continuous("Código Loja", labels = as.character(myDataSul_1$COD_LOJA), breaks = myDataSul_1$COD_LOJA) +  
  geom_point()
```



```
ggplot(data = myDataSul_2,
  aes(x=COD_LOJA, y=VOLUME)) +
  stat_summary(fun.y = sum, # somar as observações para cada produto
    geom = "bar") +
  labs( title="Quantidade vendidos na Região Sul",
    subtitle="Produto 2"
  ) +

  ylab("Quantidade vendida") +

  scale_x_continuous("Código Loja", labels = as.character(myDataSul_2$COD_LOJA), breaks = myDataSul_2$COD_LOJA) +
  geom_point()
```



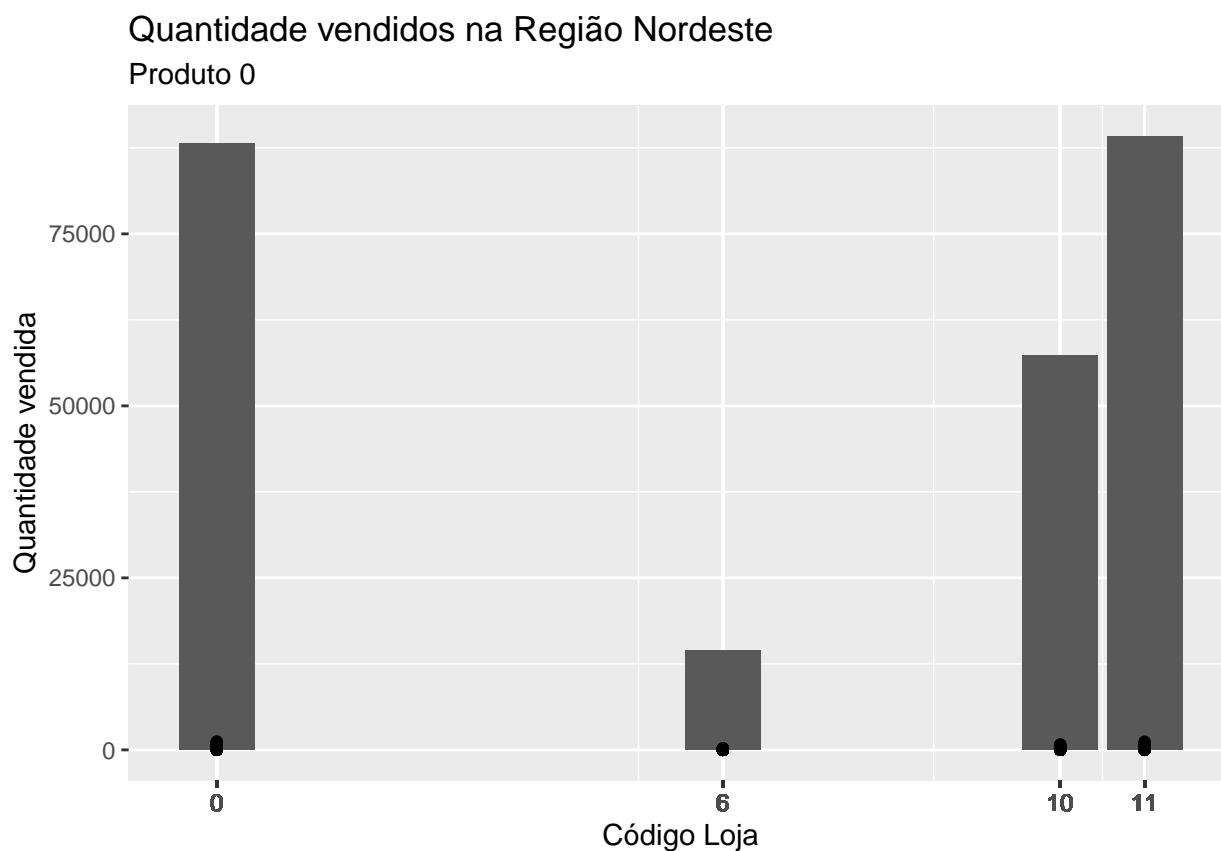
###Nordeste

```
ggplot(data = myDataNordeste_0,
  aes(x=COD_LOJA, y=VOLUME)) +
  stat_summary(fun.y = sum, # somar as observações para cada produto
    geom = "bar") +
  labs( title="Quantidade vendidos na Região Nordeste",
    subtitle="Produto 0"
  ) +

  ylab("Quantidade vendida") +

  scale_x_continuous("Código Loja", labels = as.character(myDataNordeste_0$COD_LOJA), breaks = myDataNo

  geom_point()
```



```
ggplot(data = myDataNordeste_1,
  aes(x=COD_LOJA, y=VOLUME)) +
  stat_summary(fun.y = sum, # somar as observações para cada produto
    geom = "bar") +
  labs( title="Quantidade vendidos na Região Nordeste",
    subtitle="Produto 1"
  ) +

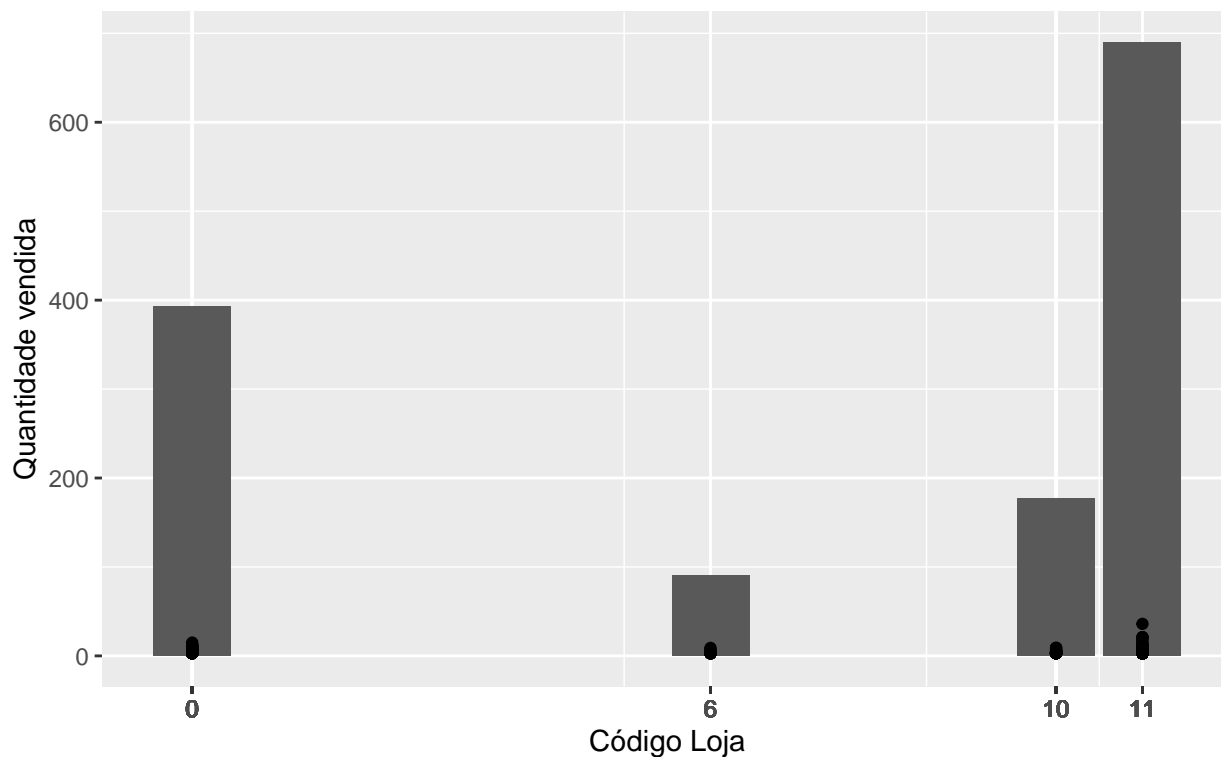
  ylab("Quantidade vendida") +

  scale_x_continuous("Código Loja", labels = as.character(myDataNordeste_1$COD_LOJA), breaks = myDataNo

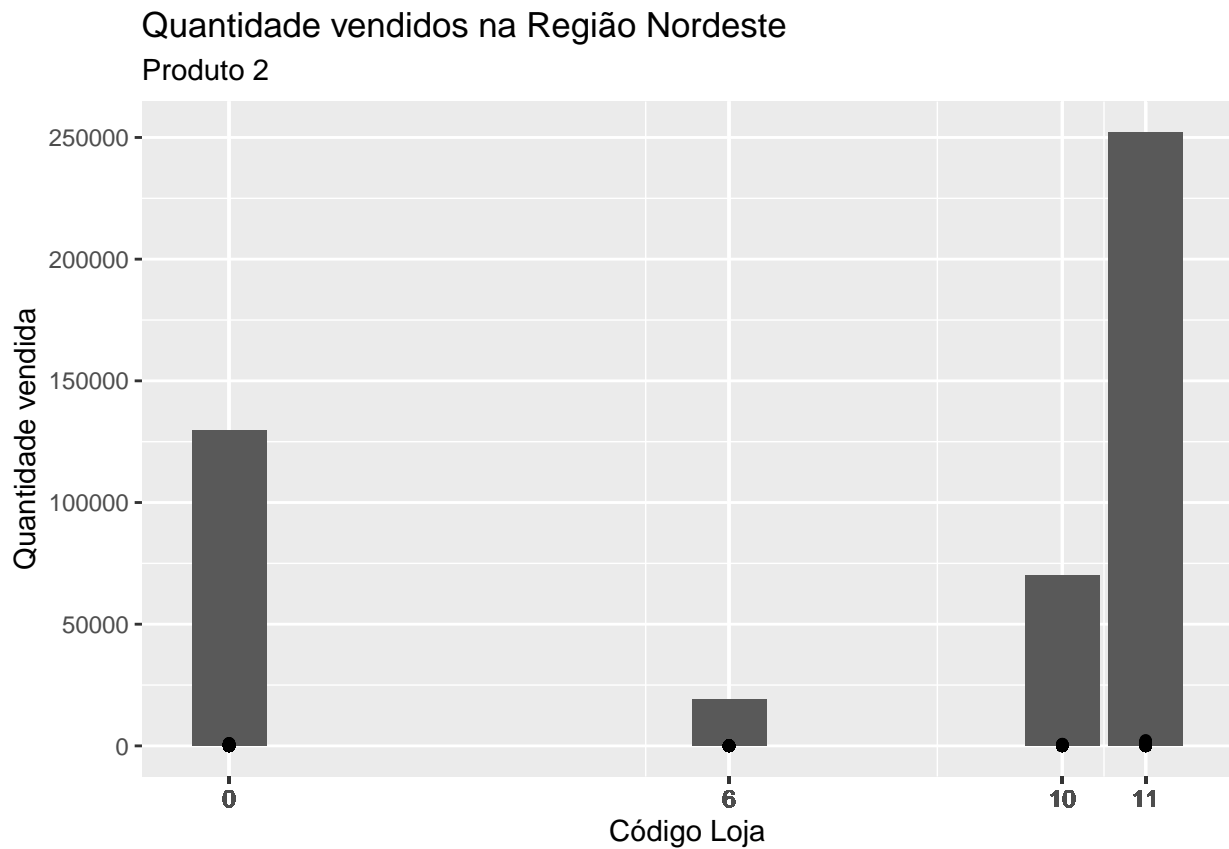
  geom_point()
```

## Quantidade vendidos na Região Nordeste

Produto 1



```
ggplot(data = myDataNordeste_2,  
  aes(x=COD_LOJA, y=VOLUME)) +  
  stat_summary(fun.y = sum, # somar as observações para cada produto  
    geom = "bar") +  
  labs( title="Quantidade vendidos na Região Nordeste",  
    subtitle="Produto 2"  
  ) +  
  
  ylab("Quantidade vendida") +  
  
  scale_x_continuous("Código Loja", labels = as.character(myDataNordeste_2$COD_LOJA), breaks = myDataNo  
  geom_point()
```

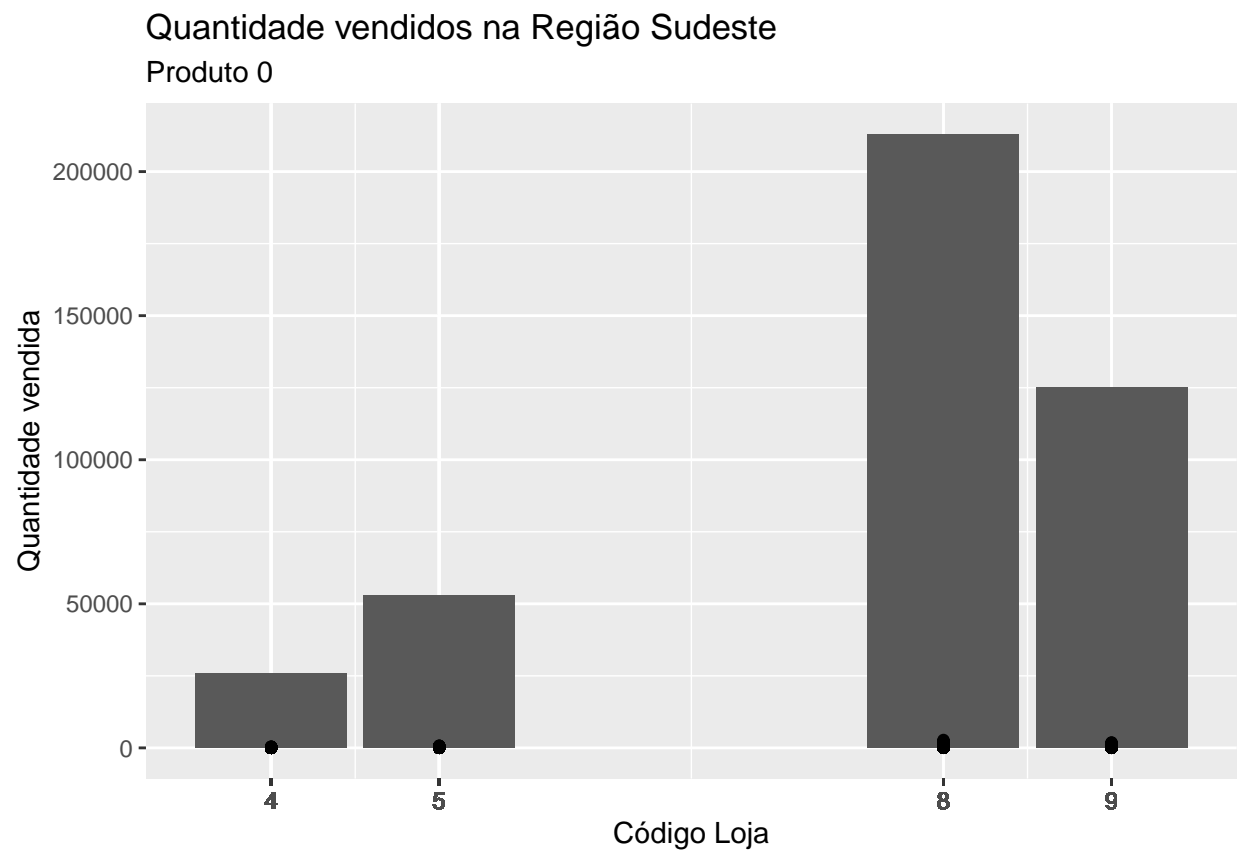


```
###Sudeste
```

```
ggplot(data = myDataSudeste_0,
  aes(x=COD_LOJA, y=VOLUME)) +
  stat_summary(fun.y = sum, # somar as observações para cada produto
    geom = "bar") +
  labs( title="Quantidade vendidos na Região Sudeste",
    subtitle="Produto 0"
  ) +

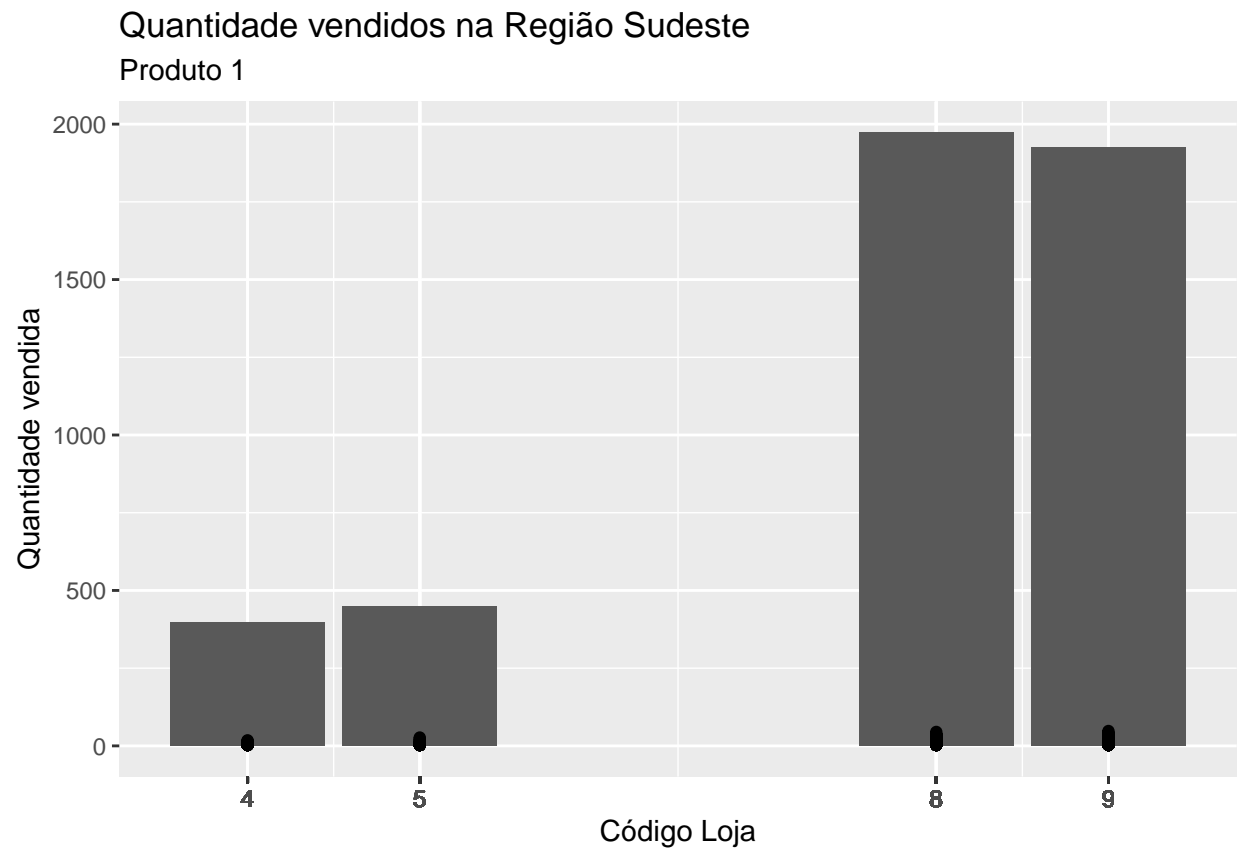
  ylab("Quantidade vendida") +

  scale_x_continuous("Código Loja", labels = as.character(myDataSudeste_0$COD_LOJA), breaks = myDataSudeste_0$COD_LOJA) +
  geom_point()
```



```
ggplot(data = myDataSudeste_1,  
  aes(x=COD_LOJA, y=VOLUME)) +  
  stat_summary(fun.y = sum, # somar as observações para cada produto  
    geom = "bar") +  
  labs( title="Quantidade vendidos na Região Sudeste",  
    subtitle="Produto 1"  
  ) +  
  
  ylab("Quantidade vendida") +  
  
  scale_x_continuous("Código Loja", labels = as.character(myDataSudeste_1$COD_LOJA), breaks = myDataSudeste_1$COD_LOJA) +  
  geom_point()
```

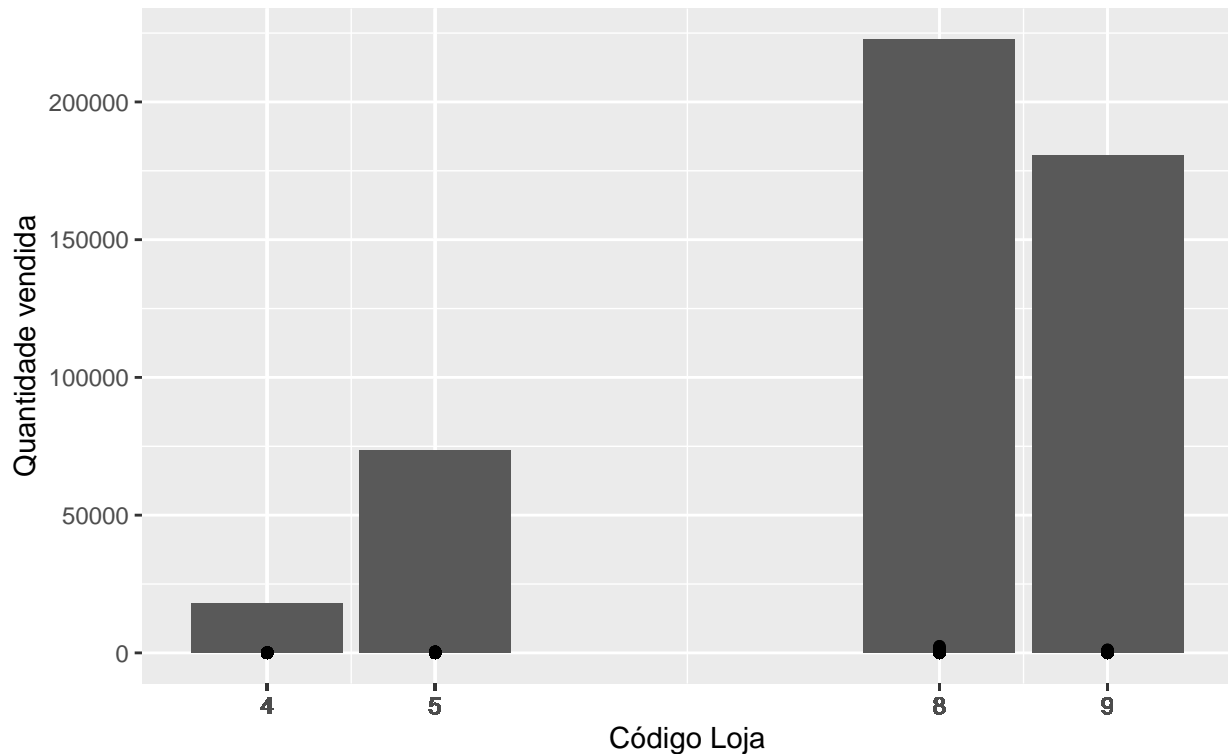




```
ggplot(data = myDataSudeste_2,  
  aes(x=COD_LOJA, y=VOLUME)) +  
  stat_summary(fun.y = sum, # somar as observações para cada produto  
    geom = "bar") +  
  labs( title="Quantidade vendidos na Região Sudeste",  
    subtitle="Produto 2"  
  ) +  
  
  ylab("Quantidade vendida") +  
  
  scale_x_continuous("Código Loja", labels = as.character(myDataSudeste_2$COD_LOJA), breaks = myDataSudeste_2$COD_LOJA) +  
  geom_point()
```

## Quantidade vendidos na Região Sudeste

### Produto 2



```
lojas = sort(unique(myData$COD_LOJA))

totalLojas = length(lojas)

quantidadeProdutoPorLoja = c(4*totalEstados)

i = 1

for(loja in lojas)
{
  dadosLoja = myData[myData$COD_LOJA == loja,]
  quantidadeProdutoPorLoja[i] = sum(dadosLoja$VOLUME)
  quantidadeProdutoPorLoja[i+1] = sum(dadosLoja[dadosLoja$COD_CATEGORIA == 0,]$VOLUME)
  quantidadeProdutoPorLoja[i+2] = sum(dadosLoja[dadosLoja$COD_CATEGORIA == 1,]$VOLUME)
  quantidadeProdutoPorLoja[i+3] = sum(dadosLoja[dadosLoja$COD_CATEGORIA == 2,]$VOLUME)
  i = i+4
}

tabelaVolumeLojaPorProduto <- matrix(quantidadeProdutoPorLoja,ncol=4,byrow=TRUE)
colnames(tabelaVolumeLojaPorProduto) <- c("Total", "Produto 0", "Produto 1", "Produto 2")
rownames(tabelaVolumeLojaPorProduto) <- lojas
tabelaVolumeLojaPorProduto <- as.table(tabelaVolumeLojaPorProduto)
kable(tabelaVolumeLojaPorProduto)
```

	Total	Produto 0	Produto 1	Produto 2
0	218295	88098	393	129804
1	76425	24354	2643	49428

	Total	Produto 0	Produto 1	Produto 2
2	95433	26826	2037	66570
3	186171	54540	9372	122259
4	44499	26013	396	18090
5	126930	53046	450	73434
6	33870	14505	90	19275
7	12864	4350	303	8211
8	437721	212976	1974	222771
9	307488	125124	1926	180438
10	127677	57303	177	70197
11	342051	89196	690	252165

Perceba que esse histograma mostra **as melhores lojas** na venda de determinados produtos. Isso mostra quais lojas devem ser analisadas para entender consumidores da loja ou até forma de venda para assim, caso possível, replicar o modelo em outras lojas.

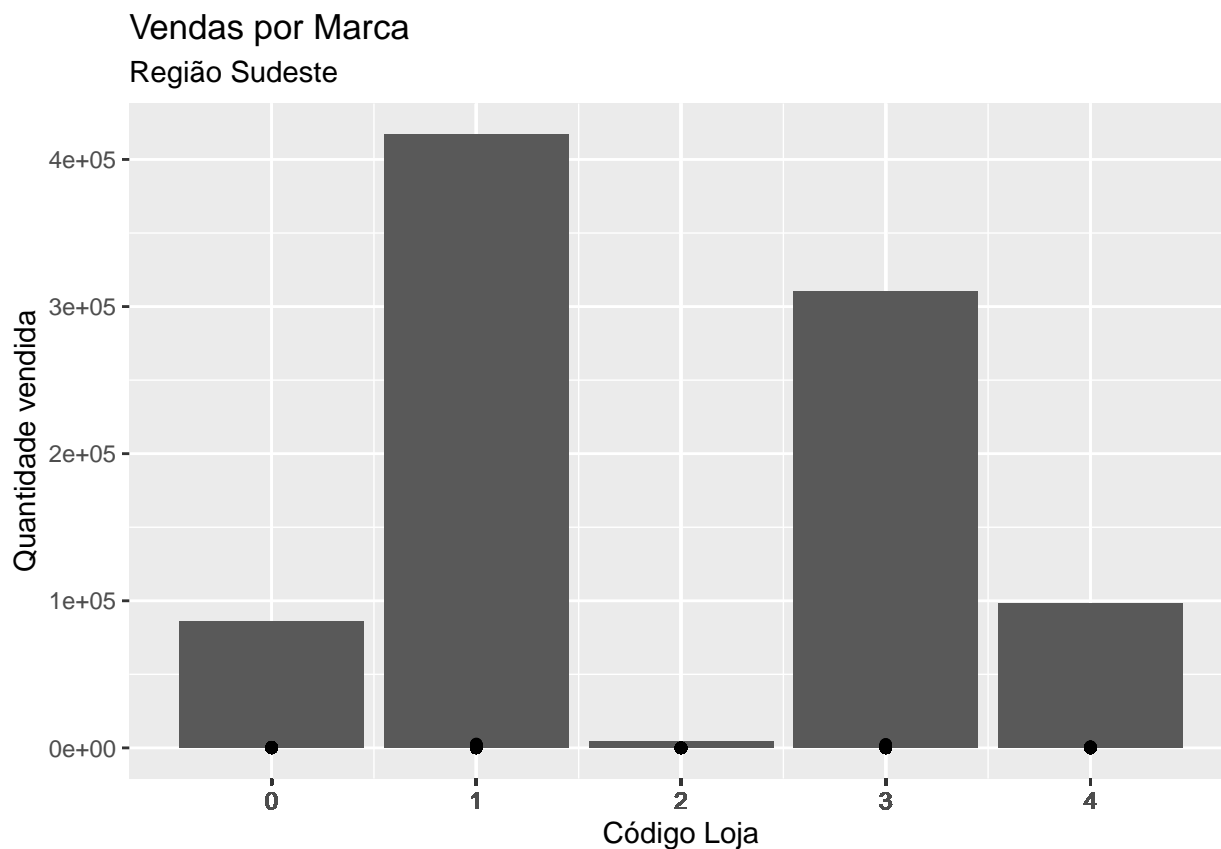
## Análise do produto vendido por região

```
##Analisar marca de cada produto mais vendido por região

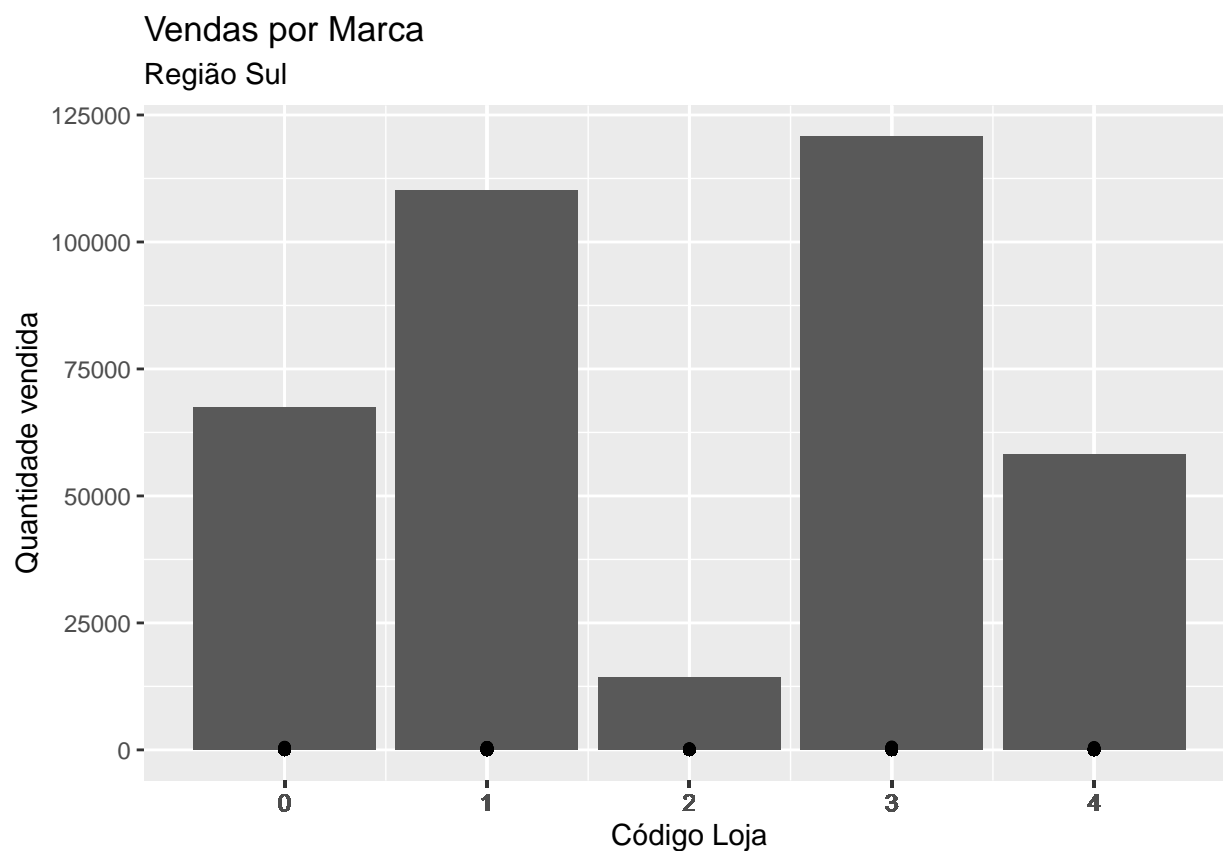
ggplot(data = myDataSudeste,
       aes(x=COD_MARCA, y=VOLUME)) +
  stat_summary(fun.y = sum, # somar as observações para cada produto
              geom = "bar") +
  labs( title="Vendas por Marca",
        subtitle="Região Sudeste"
  ) +

  ylab("Quantidade vendida") +

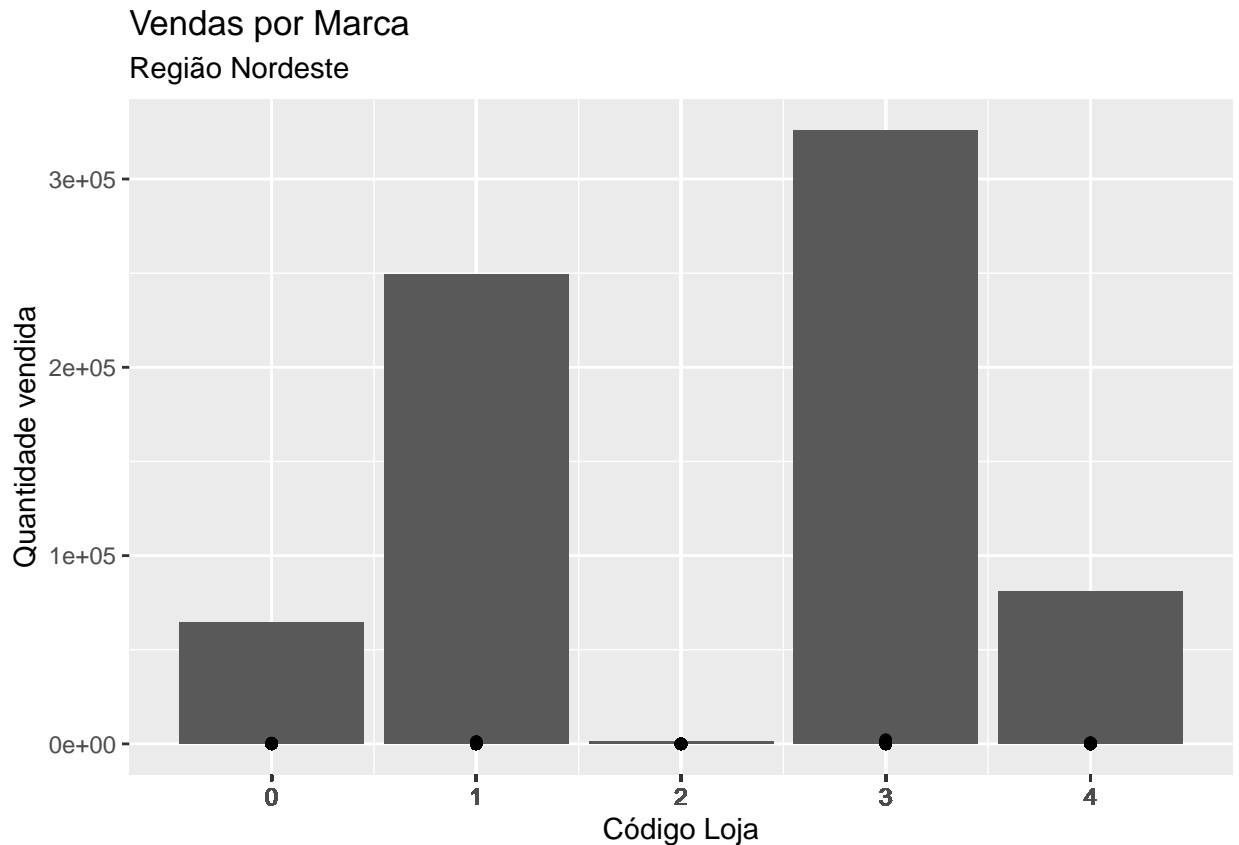
  scale_x_continuous("Código Loja", labels = as.character(myDataSudeste$COD_MARCA), breaks = myDataSudeste$COD_MARCA) +
  geom_point()
```



```
ggplot(data = myDataSul,  
  aes(x=COD_MARCA, y=VOLUME)) +  
  stat_summary(fun.y = sum, # somar as observações para cada produto  
    geom = "bar") +  
  labs( title="Vendas por Marca",  
    subtitle="Região Sul"  
  ) +  
  
  ylab("Quantidade vendida") +  
  
  scale_x_continuous("Código Loja", labels = as.character(myDataSul$COD_MARCA), breaks = myDataSul$COD_MARCA) +  
  geom_point()
```



```
ggplot(data = myDataNordeste,  
  aes(x=COD_MARCA, y=VOLUME)) +  
  stat_summary(fun.y = sum, # somar as observações para cada produto  
    geom = "bar") +  
  labs( title="Vendas por Marca",  
    subtitle="Região Nordeste"  
  ) +  
  
  ylab("Quantidade vendida") +  
  
  scale_x_continuous("Código Loja", labels = as.character(myDataNordeste$COD_MARCA), breaks = myDataNordeste$COD_MARCA) +  
  geom_point()
```



```
tabelaVolumeRegiaoPorProduto = matrix(c(totalNordeste, totalNordeste_0, totalNordeste_1, totalNordeste_2),
colnames(tabelaVolumeRegiaoPorProduto) <- c("Total", "Produto 0", "Produto 1", "Produto 2")
rownames(tabelaVolumeRegiaoPorProduto) <- c("Nordeste", "Sudeste", "Sul")
tabelaVolumeRegiaoPorProduto <- as.table(tabelaVolumeRegiaoPorProduto)
kable(tabelaVolumeRegiaoPorProduto)
```

	Total	Produto 0	Produto 1	Produto 2
Nordeste	721893	249102	1350	471441
Sudeste	916638	417159	4746	494733
Sul	370893	240948	83619	592071

Esses histogramas mostram a **melhor forma de distribuir o estoque por região**.

## Análise dos dados por loja e estado

```
quantidadeVendidaEstadoLoja <- c(length= totalEstados*totalLojas)

i <- 1

for(loja in lojas)
{
  dadosLoja <- myData[myData$COD_LOJA == loja,]

  for(estado in estados)
```

```

{
  dadosLojaEstado <- dadosLoja[myData$"REGIAO.1" == estado,]

  quantidadeVendidaEstadoLoja[i] <- sum(dadosLojaEstado$VOLUME,na.rm=T)
  i = i+1
}
}

tabelaVolumeLojaPorEstado <- matrix(quantidadeVendidaEstadoLoja,ncol=totalLojas,byrow=TRUE)
colnames(tabelaVolumeLojaPorEstado) <- lojas
rownames(tabelaVolumeLojaPorEstado) <- estados
tabelaVolumeLojaPorEstado <- as.table(tabelaVolumeLojaPorEstado)
kable(tabelaVolumeLojaPorEstado)

```

	0	1	2	3	4	5	6	7	8	9	10	11
CE	0	0	101022	0	0	0	0	0	0	117273	0	0
MG	29700	0	0	0	0	0	0	46725	0	0	36555	0
PE	0	0	0	0	0	58878	0	0	64329	0	0	0
PI	0	0	0	121842	0	0	22977	0	0	0	0	0
PR	0	21522	0	0	44409	0	0	0	0	0	0	82521
RJ	0	0	17388	0	0	0	0	0	0	16482	0	0
RS	6579	0	0	0	0	0	0	6285	0	0	117024	0
SC	0	0	0	0	0	320697	0	0	113523	0	0	0
SE	0	0	0	193965	0	0	60327	0	0	0	0	0
SP	0	67350	0	0	86640	0	0	0	0	0	0	255411

Perceba que esses dados mostram uma forma de inferir quais são as lojas que estão melhores em venda para cada estado. Isso mostra que, após análise das melhores lojas, podemos verificar a **melhor forma de implementar um modelo de vendas por estado**.

## Análise da quantidade vendida por materiais

```

quantidadeVendidaMateriais <- vector(length=tamanhoMateriais)

i = 1

for(material in materiais)
{
  dadosMateriais <- myData[myData$COD_MATERIAL==material,]

  quantidadeVendidaMateriais[i] <- sum(dadosMateriais$VOLUME)

  i = i + 1
}

tabelaVolumeMateriais <- matrix(quantidadeVendidaMateriais,ncol=1,byrow=TRUE)
colnames(tabelaVolumeMateriais) <- "Volume"
rownames(tabelaVolumeMateriais) <- materiais

```

```
tabelaVolumeMateriais <- as.table(tabelaVolumeMateriais)
kable(tabelaVolumeMateriais)
```

	Volume
0	757323
1	237183
2	218136
3	20451
4	776331

Essa tabela mostra uma forma de analisar quais são os materiais mais vendidos. Sendo assim, esses dados mostram as melhores formas de decidir quais são os **materiais que são melhores para serem investidos ou comprados**.

## Análise da relação Entre Vendas de Material por Categoria

```
quantidadeVendidaMaterialCategoria <- vector(length = tamanhoCategorias*tamanhoMateriais)

i <- 1

for(material in materiais)
{
  dadosMaterial <- myData[myData$COD_MATERIAL == material,]

  for(categoria in categorias)
  {
    dadosMaterialCategoria <- dadosMaterial[myData$COD_CATEGORIA == categoria,]

    quantidadeVendidaMaterialCategoria[i] <- sum(dadosMaterialCategoria$VOLUME,na.rm=T)
    i = i+1
  }
}

tabelaVolumeMaterialCategoria <- matrix(quantidadeVendidaMaterialCategoria,ncol=tamanhoMateriais,byrow=
colnames(tabelaVolumeMaterialCategoria) <- materiais
rownames(tabelaVolumeMaterialCategoria) <- categorias
tabelaVolumeMaterialCategoria <- as.table(tabelaVolumeMaterialCategoria)
kable(tabelaVolumeMaterialCategoria)
```

	0	1	2	3	4
0	231270	18144	507909	69906	5817
1	161460	64623	4668	148845	7251
2	543	12657	225951	28224	522156

Essa tabela mostra a relação entre os produtos vendidos e o material. Os resultados dispostos mostram quais são os **materiais mais adequados para produzir um determinado produto**.