

An

Doriedson

09-11-2024

```
require(ggplot2)
```

```
## Carregando pacotes exigidos: ggplot2
```

```
source('arquivo1.R')
```

- calorias (kcal)

```
dados = read.csv(Fonte)
```

```
attach(dados)
```

```
dsel = cbind(dados[, c(1:6)], dados$dessert)
```

```
colnames(dsel)
```

```
## [1] "title"      "rating"      "calories"    "protein"
## [5] "fat"        "sodium"      "dados$dessert"
```

```
colnames(dsel) <- c("titulo", "nota", "calorias", "proteina", "gordura", "sodio", "sobremesa")
```

```
dsel <- na.omit(dsel)
```

```
attach(dsel)
```

```
head(dsel); dim(dsel)
```

```
##                               titulo  nota calorias proteina gordura
## 1      Lentil, Apple, and Turkey Wrap 2.500      426      30      7
## 2 Boudin Blanc Terrine with Red Onion Confit 4.375      403      18     23
## 3      Potato and Fennel Soup Hodge 3.750      165       6      7
## 5      Spinach Noodle Casserole 3.125      547      20     32
## 6              The Best Blts 4.375      948      19     79
## 9      Korean Marinated Beef 4.375      170       7     10
##  sodio sobremesa
## 1    559         0
## 2   1439         0
## 3    165         0
## 5    452         0
## 6   1042         0
## 9   1272         0
```

```
## [1] 15864      7
```

```
#plot(gordura)

#boxplot(nota~sobremesa, col='navyblue')

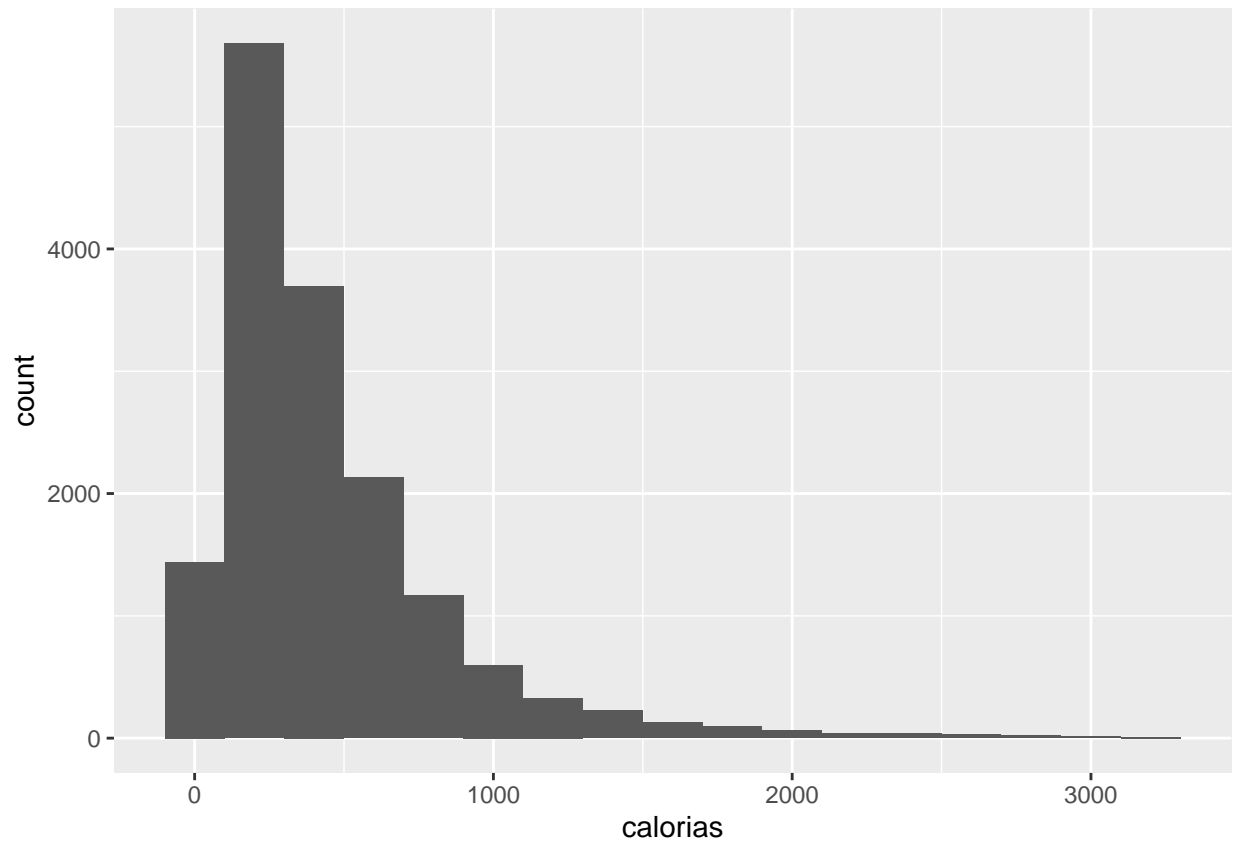
quantile(dsel$calorias, seq(0, 1, .01))
```

| | | | | | | |
|----|---------|---------|---------|---------|---------|---------|
| ## | 0% | 1% | 2% | 3% | 4% | 5% |
| ## | 0.00 | 19.00 | 29.26 | 43.00 | 54.00 | 62.00 |
| ## | 6% | 7% | 8% | 9% | 10% | 11% |
| ## | 73.00 | 83.00 | 91.00 | 100.00 | 108.00 | 116.00 |
| ## | 12% | 13% | 14% | 15% | 16% | 17% |
| ## | 122.00 | 129.00 | 136.00 | 143.00 | 149.00 | 155.00 |
| ## | 18% | 19% | 20% | 21% | 22% | 23% |
| ## | 161.00 | 167.00 | 172.00 | 178.00 | 183.00 | 189.00 |
| ## | 24% | 25% | 26% | 27% | 28% | 29% |
| ## | 194.00 | 199.00 | 204.00 | 209.00 | 214.00 | 219.00 |
| ## | 30% | 31% | 32% | 33% | 34% | 35% |
| ## | 224.00 | 229.00 | 233.00 | 237.00 | 242.42 | 247.05 |
| ## | 36% | 37% | 38% | 39% | 40% | 41% |
| ## | 253.00 | 258.00 | 263.00 | 269.00 | 274.00 | 279.00 |
| ## | 42% | 43% | 44% | 45% | 46% | 47% |
| ## | 284.00 | 290.00 | 296.00 | 301.00 | 306.00 | 311.61 |
| ## | 48% | 49% | 50% | 51% | 52% | 53% |
| ## | 318.00 | 325.00 | 333.00 | 340.00 | 348.00 | 357.00 |
| ## | 54% | 55% | 56% | 57% | 58% | 59% |
| ## | 365.00 | 374.00 | 384.00 | 394.00 | 403.00 | 412.00 |
| ## | 60% | 61% | 62% | 63% | 64% | 65% |
| ## | 421.00 | 429.00 | 438.00 | 448.00 | 457.00 | 468.00 |
| ## | 66% | 67% | 68% | 69% | 70% | 71% |
| ## | 477.00 | 488.00 | 499.00 | 508.00 | 519.00 | 532.00 |
| ## | 72% | 73% | 74% | 75% | 76% | 77% |
| ## | 544.00 | 558.00 | 572.00 | 587.25 | 601.00 | 616.00 |
| ## | 78% | 79% | 80% | 81% | 82% | 83% |
| ## | 632.00 | 650.00 | 667.00 | 687.03 | 708.66 | 729.00 |
| ## | 84% | 85% | 86% | 87% | 88% | 89% |
| ## | 753.00 | 778.55 | 805.00 | 833.00 | 865.00 | 901.00 |
| ## | 90% | 91% | 92% | 93% | 94% | 95% |
| ## | 942.00 | 992.33 | 1050.00 | 1116.00 | 1209.00 | 1319.85 |
| ## | 96% | 97% | 98% | 99% | 100% | |
| ## | 1452.48 | 1682.11 | 2066.96 | 3257.00 | 3011 | 1218.00 |

Observando o percentil 99 tem-se que 99% dos valores são menores ou iguais a 3257 kcal. Tomarei como limite das observações de caloria, pois é improvável/inviável que receitas ultrapassem de maneira tão acentuada esse valor; receitas com milhões de kcal são claramente erros nessa base de dados.

```
dsel <- subset(dsel, dsel$calorias <= quantile(dsel$calorias, .99)[[1]])

ggplot(dsel, aes(x=calorias)) + geom_histogram(binwidth = 200)
```

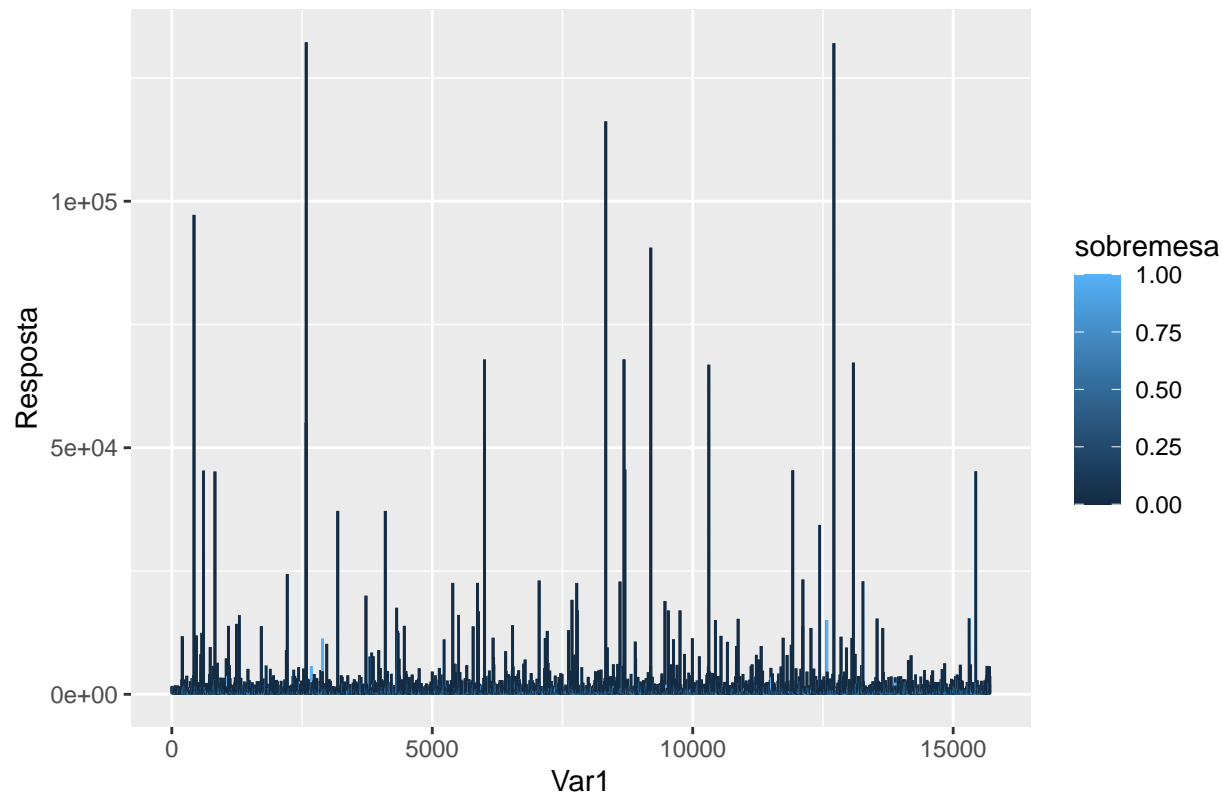


```
# Assumindo que 'var1' é uma variável numérica que pode ser interpretada como tempo

g1 = ggplot(dsel, aes(y = sodio, x = seq(1:15706), color = sobremesa)) +
  geom_line() +
  #theme_dark() +
  labs(x = "Var1", y = "Resposta",
       title = "Evolução da Resposta em Relação a Var1")

g1
```

Evolução da Resposta em Relação a Var1



```
#g1 + scale_color_manual(breaks=c(0:1), palette = "YlGn")#values=c("red", "green"))
```

```
#g1 + scale_fill_brewer(palette = "YlGn")
```

```
glm(sobremesa~calorias+proteina+gordura+sodio, family = binomial())
```

```
## Warning: glm.fit: probabilidades ajustadas numericamente 0 ou 1 ocorreu
```

```
##
```

```
## Call: glm(formula = sobremesa ~ calorias + proteina + gordura + sodio,
```

```
## family = binomial())
```

```
##
```

```
## Coefficients:
```

```
## (Intercept) calorias proteina gordura sodio
```

```
## -1.4194366 0.0008484 -0.0019997 -0.0078761 -0.0004182
```

```
##
```

```
## Degrees of Freedom: 15863 Total (i.e. Null); 15859 Residual
```

```
## Null Deviance: 15250
```

```
## Residual Deviance: 14440 AIC: 14450
```

Referências

<https://rpubs.com/mpfoley73/527573>

<https://www.kaggle.com/code/rtatman/regression-challenge-day-1>

[https://www.kaggle.com/code/rtatman/datasets-for-regression-analysis#Poisson-regression-\(predicting-a-count-value\)](https://www.kaggle.com/code/rtatman/datasets-for-regression-analysis#Poisson-regression-(predicting-a-count-value))

<https://www.kaggle.com/datasets/hugodarwood/epirecipes?resource=download>