

Rug Plot

James Hunter

6 de março de 2017

Introdução

Na aula de 24 de fevereiro sobre Visualização, eu quis mostrar um “rug plot” como adjunto a uma histograma, mas tive dificuldades com ele. A “plotagem de tapete” enriquece uma histograma. Ela mostra onde caem todos os pontos de dados da variável sob consideração.

Aqui é o código que permite a construção de um rug plot com os mesmos dados que estávamos usando na aula.

Carregar Pacotes

```
suppressMessages(library(tidyverse))
suppressPackageStartupMessages(library(knitr))
options(scipen = 1000)
```

Carregar Dados

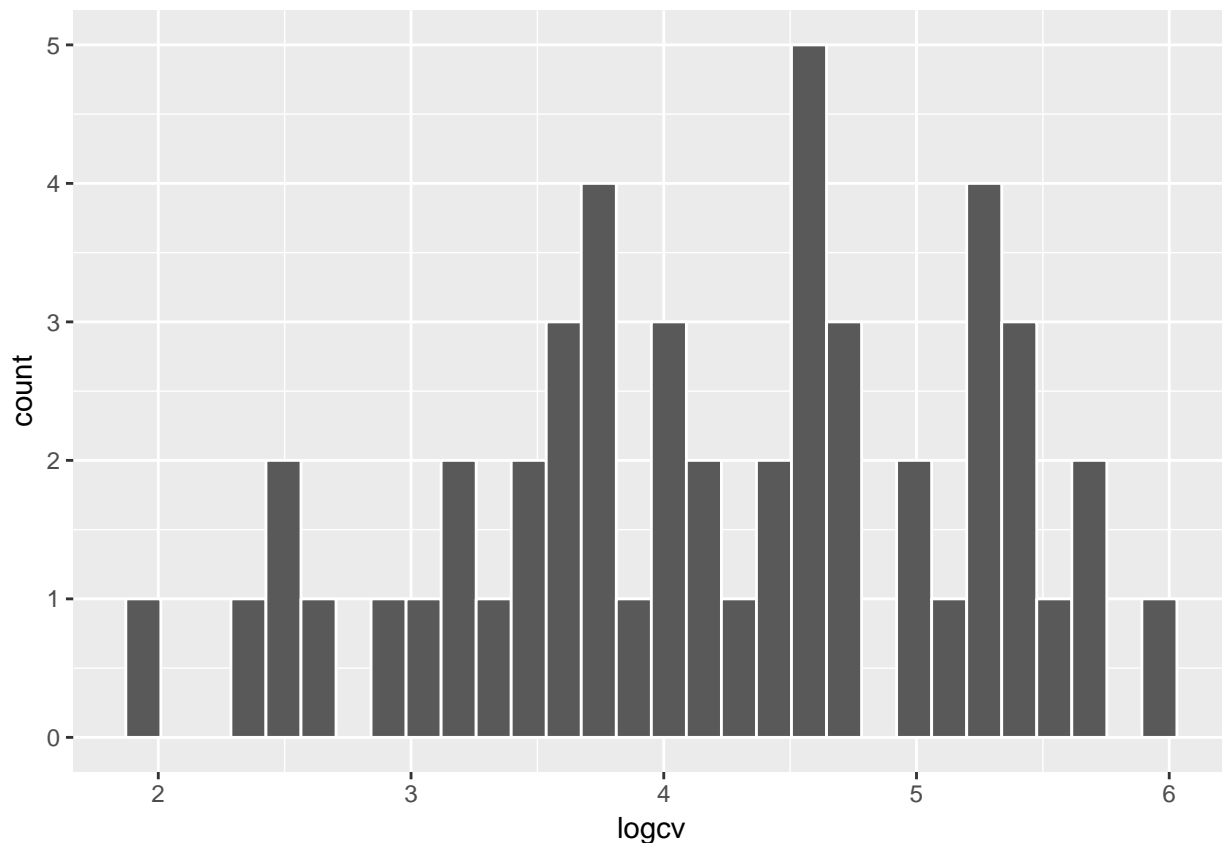
```
testes <- read_csv("pac_demo.csv") %>%
  mutate(logcv = log10(copias_cv)) %>%
  select(c(codepac, logcv, contagem_cd4, contagem_cd8))
```

Construir Histograma

A única diferença ao gráfico que usei na aula é que tirei as referências a densidades. `geom_rug` não brinca bem com densidade.

```
gr1 <- ggplot(data = testes, mapping = aes(x = logcv)) +
  geom_histogram(color = "white")
gr1

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

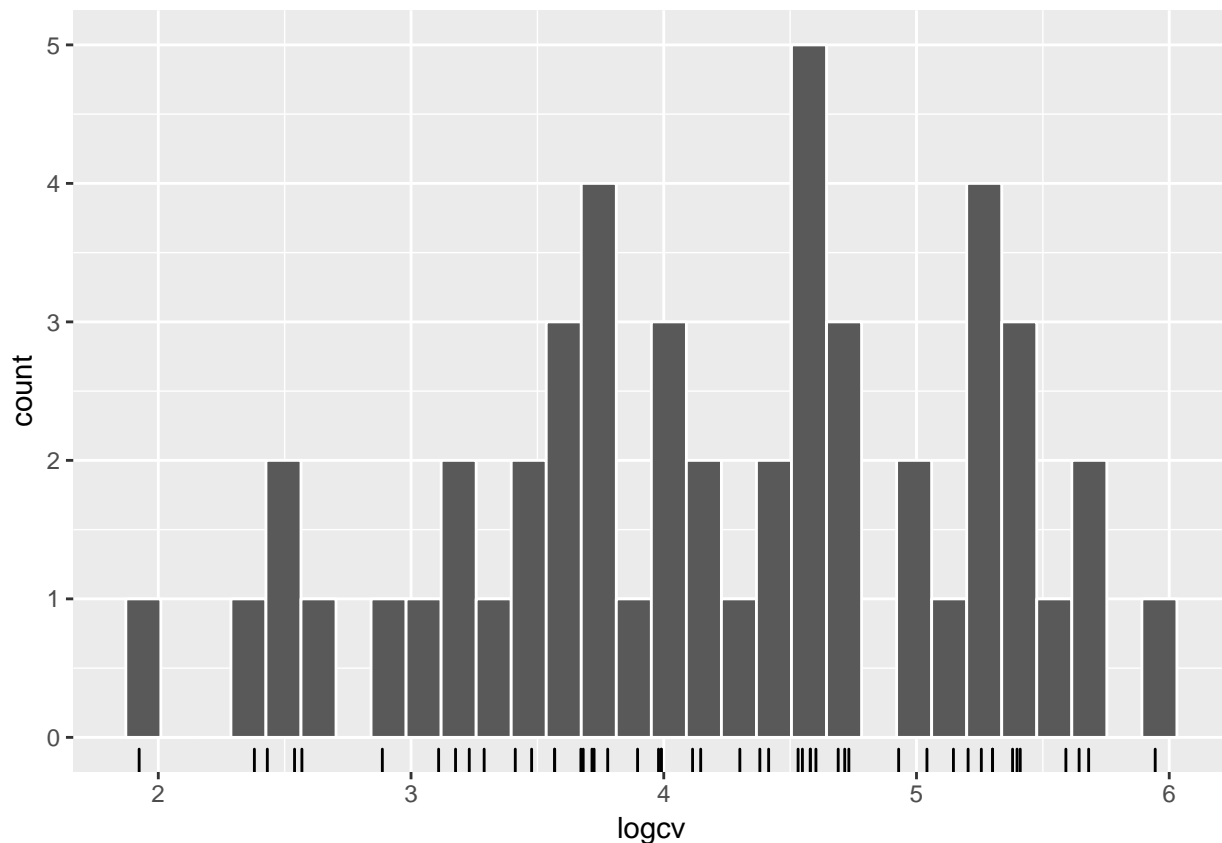


Anote que deixei a advertência (“warning”) visível aqui. Quando você faz uma histograma usando o valor padrão de 30 “bins”, o programa vai advertir você que talvez pode achar um valor melhor baseado não no número de bins mas na largura deles. É somente uma advertência. Não é um erro! O gráfico calculará e se mostrará sem problemas. Se você quer mexer mais com o gráfico, pode achar um valor melhor para os bins, mas não é necessário especialmente quando você está explorando os dados.

Colocando um Rug Plot no Gráfico

```
gr1 + geom_rug()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Esse foi simples. Só precisa tirar as referências a densidades. `geom_rug` não brinca bem com densidade.

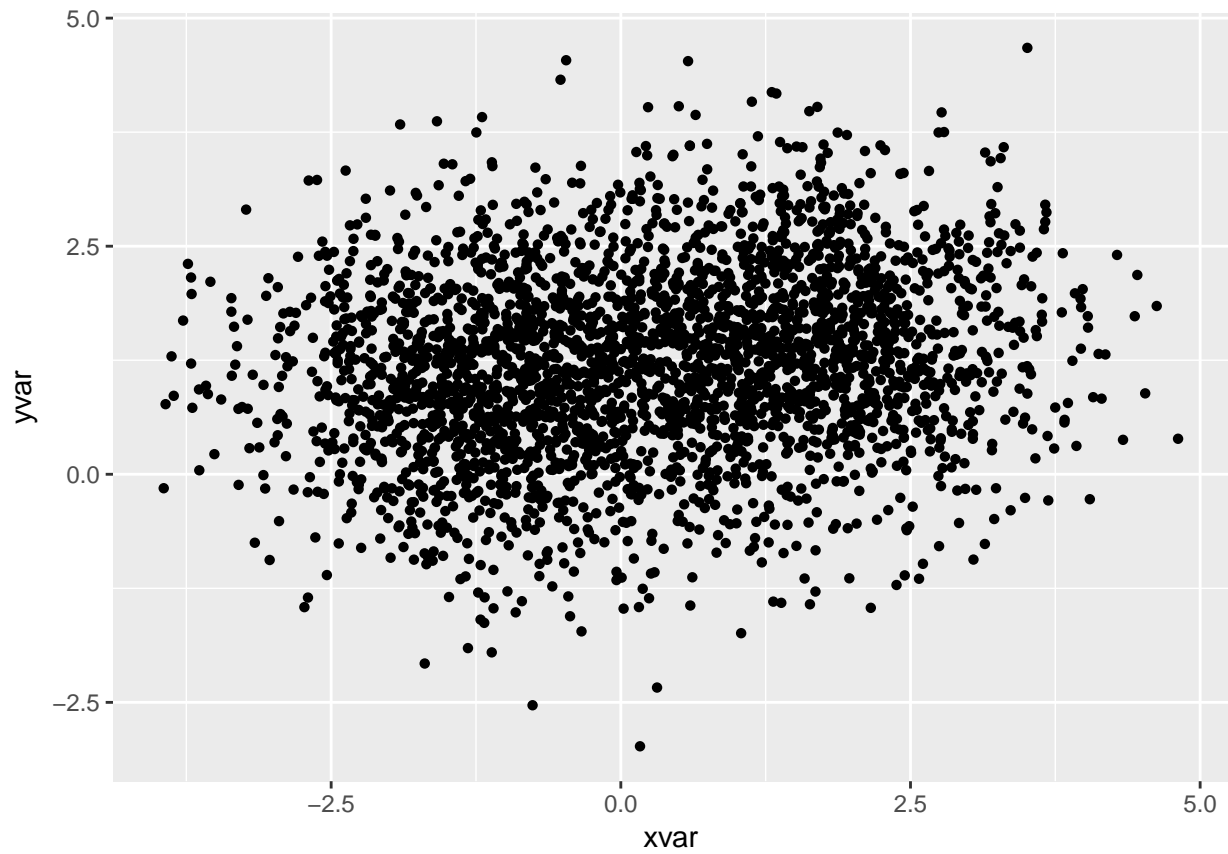
Rug Plots com Gráficos de Dispersão

Quando você tem muito pontos numa gráfico de dispersão, um rug plot pode ajudar com a visualização. Vamos usar dados inventados (grato ao blog R for Public Health).

```
set.seed(10005)
xvar <- c(rnorm(1500, mean = -1), rnorm(1500, mean = 1.5))
yvar <- c(rnorm(1500, mean = 1), rnorm(1500, mean = 1.5))
zvar <- as.factor(c(rep(1, 1500), rep(2, 1500)))
xy <- tibble(xvar, yvar, zvar)
```

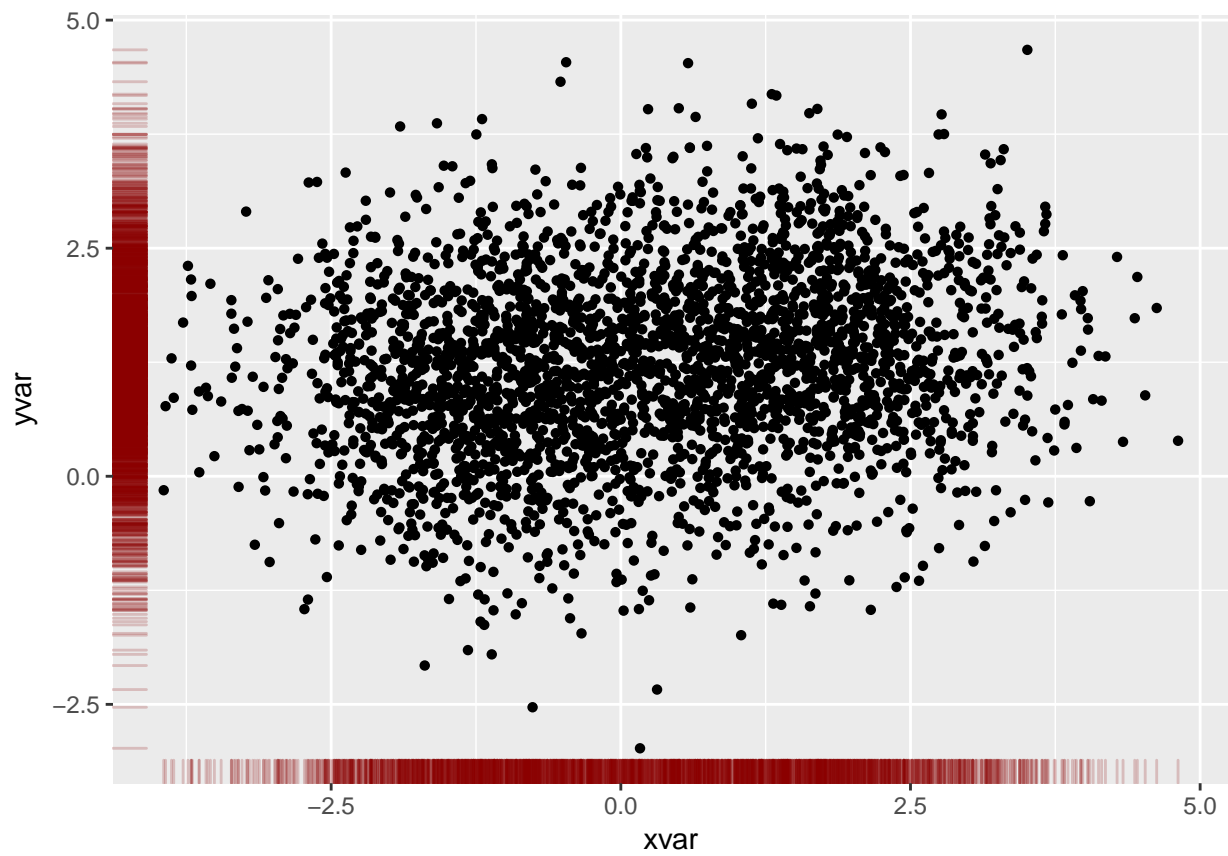
Pode criar o gráfico base de `xvar` x `yvar`:

```
gr2 <- ggplot(xy, aes(x = xvar, y = yvar)) + geom_point(shape = 16)
gr2
```



Visualização dos pontos é difícil. Um rug plot pode ajudar.

```
gr2 + geom_rug(color = "darkred", alpha = 0.2)
```



Onde os pontos são mais concentrados, os rugs parecem mais solidos, mas onde são mais dispersos, o vermelho fica mais leve (por causa do parâmetro `alpha`)