

## Análise de Dados Categorizados

### 1ª lista – 1º/2016

3. Alguns conjuntos de dados do livro *Categorical Data Analysis* (A. Agresti, 3rd ed., 2012) encontram-se na página <http://www.stat.ufl.edu/~aa/cda/data.html>. Represente graficamente os dados correspondentes aos seguintes itens:

(a) 2. Rajon Rondo assists data set for Exercise 5.3.

Os dados foram copiados da página indicada para o arquivo `dadosL1-3a.txt` (sem a primeira linha de rótulos, sem a linha em branco de separação e sem a última linha com “;”). Em seguida são apresentados o código em R e o gráfico resultante.

```
## Lista 1 - 3(a)
dados3a <- ftable(read.table("dadosL1-3a.txt"))
freqres <- prop.table(dados3a, margin = 1)
plot(attr(freqres,"row.vars")$V1, freqres[, 2], pch = 20,
      xlab = "Número de assistências", ylab = "Proporção de vitórias")
```

A função `ftable` lê os dados do arquivo e os transforma em uma tabela de contingências. A variável `V1` (linhas) representa os diferentes números de assistências observados nos dados (por exemplo, não ocorreram partidas com quatro assistências). A variável `V2` (colunas) indica se o resultado foi derrota (0) ou vitória (1).

A função `prop.table` transforma a tabela de contingências (`dados3a`) em uma tabela de frequências relativas condicionais em `V1` (`margin = 1`). O gráfico utiliza a segunda coluna de `freqres`, que corresponde à proporção de vitórias (`V2 = 1`).

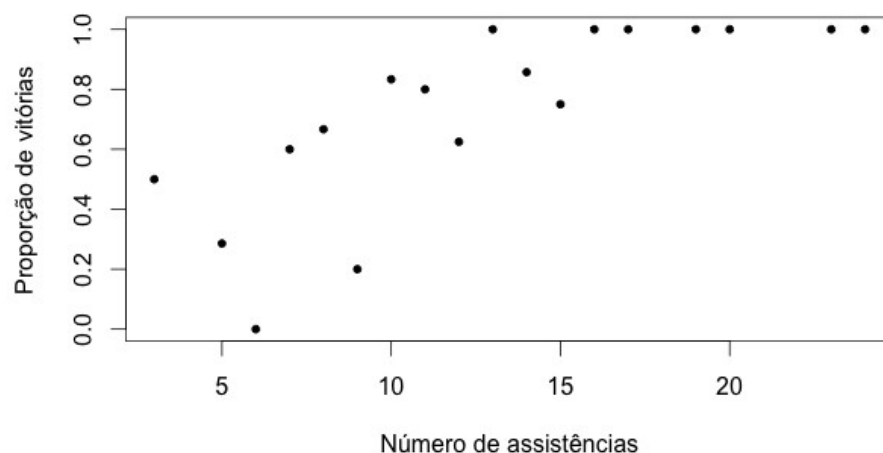


Figura 1. Proporção de vitórias do Boston Celtics *versus* número de assistências de Rajon Rondo.

Nota 1. Comente o gráfico da Fig. 1.

(b) 9. Data set for Exercise 6.28.

Os dados listados abaixo foram copiados para o arquivo dadosL1-3b.txt sem nenhuma modificação.

Gender	Residence	IQ	Status	Aspiration		
				High	Low	
Male	Rural	High	High	117	47	
			Low	54	87	
		Low	High	29	78	
			Low	31	262	
		Small	High	High	350	80
			Low	70	85	
	Large	High	High	71	120	
			Low	33	265	
		Low	High	151	31	
			Low	27	23	
		Low	High	30	27	
			Low	12	52	
Female	Rural	High	High	102	69	
			Low	52	119	
		Low	High	32	73	
			Low	28	349	
		Small	High	High	338	96
			Low	44	99	
	Large	High	High	76	107	
			Low	22	344	
		Low	High	148	35	
			Low	17	39	
		Low	High	21	47	
			Low	6	116	

Em seguida são apresentados o código em R e a figura resultante. Para cada combinação dos níveis das variáveis explicativas é apresentado o gráfico de barras das frequências relativas da variável resposta (Aspiration).

```
## Lista 1, 3(b)
library(lattice)
dados3b <- read.ftable("dadosL1-3b.txt")
freqasp <- c(dados3b) # Freq. de aspiration high e low empilhadas
gender <- gl(2, 1, 48, labels = c("Male", "Female"))
residence <- gl(3, 2, 48, labels = c("Rural", "Small urban", "Large urban"))
iq <- gl(2, 6, 48, labels = c("High IQ", "Low IQ"))
socstatus <- gl(2, 12, 48, labels = c("High", "Low"))
aspir <- gl(2, 24, 48, labels = c("High aspiration", "Low aspiration"))
# Var. resposta: Occupational aspirations (aspir)
tab3b <- xtabs(~ socstatus + iq + residence + gender + aspir) * freqasp
tab3bp <- prop.table(tab3b, margin = 1:4) * 100 # Freq. relativas
as.data.frame(tab3b)
barchart(tab3bp, xlab = "Percentage", ylab = "Socioeconomic status",
          stack = FALSE, layout = c(6, 2), scale = list(cex = 1.5),
          auto.key = list(space = "top", columns = 2))
```

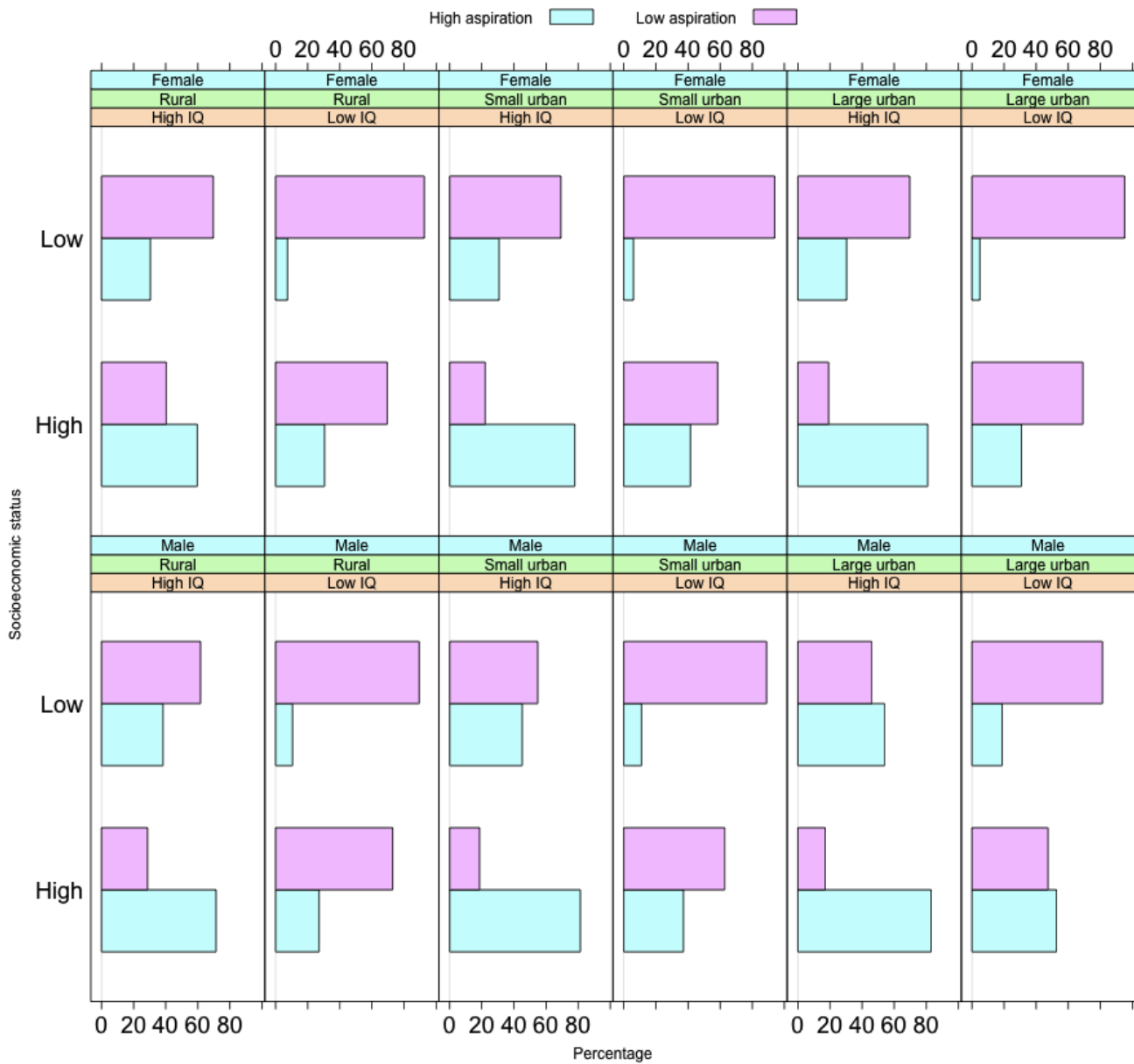


Figura 2. Gráfico de barras de *occupational aspirations* de acordo com diferentes características.

Nota 2. Comente o gráfico da Fig. 2.

Nota 3. Procure entender todas as linhas do código utilizado para obter o gráfico.

Nota 4. Represente a distribuição condicional da variável resposta (*Aspiration*) para cada uma das quatro variáveis explicativas.

Nota 5. Refaça o gráfico com o pacote SPSS, por exemplo.