

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press, pp. 24-27.

2.2 Visualizing single random variables (continued)

- 独立したファイルとして、画像を保存する方法
 - 右クリックで“copy as”か“save as”を選択すると、いくつかのフォーマット(png, pdf, jpeg, PostScript)のうちから選択できる。
 - これは、関数としての `png()`, `pdf()`, `jpeg()`, `postscript()`でも同じ。
- e.g. ①: `データ・セット ratings`の語の頻度 (Frequency) を棒グラフで表したい。
 1. `jpeg("barplot.jpeg", width=400,height=420)`
 2. `truehist(ratings$Frequency, xlab="log word frequency", col="grey")`
 3. `dev.off()`
 - *単位: png と jpeg ファイルはピクセル、pdf と postscript はインチ
 - *`dev.off()`がないと、画像が表示されないので注意
- 棒グラフに曲線を追加し、連続性を表現する方法
 - histogram の問題点: 連続性の欠如
 - `density()`「確率密度関数」を用いて滑らかなヒストグラムを作る。
- e.g. ②: `データ・セット lexdec`の反応速度 (RT) の分布を滑らかな曲線で表したい。
 1. 2つのグラフを重ねて表示するための工夫:
 - ◇ `h=hist(lexdec$RT, freq=FALSE, plot=FALSE)`
 - ◇ `d=density(lexdec$RT)`
 2. x軸、y軸それぞれの範囲を適切に定義する必要がある:
 - ◇ `xlimit=range(h$breaks,d$x)`
 - ◇ `ylim=range(0,h$density, d$y)`
 3. `hist(lexdec$RT, freq=FALSE, xlim=xlimit, ylim=ylim, main="",xlab="log RT", ylab="", col="lightgrey", border="darkgrey", breaks=seq(5.8,7.6,by=0.1))`
 4. 表示したヒストグラムに曲線を追加する:
 - ◇ `lines(dx, dy)`
 - ◇ `lines(d)`
- cf. 通常の作図関数 `plot()`でも描ける。

```
> hist(lexdec$RT, freq=FALSE, plot=FALSE)
```

```
$breaks
```

```
[1] 5.8 6.0 6.2 6.4 6.6 6.8 7.0 7.2 7.4 7.6
```

```
$counts
```

```
[1] 34 339 636 385 150 74 33 4 4
```

```
$intensities
```

```
[1] 0.10247135 1.02169982 1.91681736 1.16033755 0.45207957 0.22302592
```

```
[7] 0.09945750 0.01205546 0.01205546
```

```
$density
```

```
[1] 0.10247135 1.02169982 1.91681736 1.16033755 0.45207957 0.22302592
```

```
[7] 0.09945750 0.01205546 0.01205546
```

```
$mids
```

```
[1] 5.9 6.1 6.3 6.5 6.7 6.9 7.1 7.3 7.5
```

```
$xname
```

```
[1] "lexdec$RT"
```

```
$equidist
```

```
[1] TRUE
```

```
> density(lexdec$RT)
```

x	y
Min. :5.697	Min. :6.218e-05
1st Qu.:6.203	1st Qu.:1.330e-02
Median :6.708	Median :1.531e-01
Mean :6.708	Mean :4.942e-01
3rd Qu.:7.213	3rd Qu.:7.705e-01
Max. :7.719	Max. :2.049e+00

The estimated density values.
These will be non-negative,
but can be zero.

The n coordinates of the points
where the density is estimated.

```
> range(h$breaks,d$x)
```

```
[1] 5.697477 7.718780
```

```
> range(0,h$density, d$y)
```

```
[1] 0.000000 2.048774
```