

Capital

Dataset

Dataset consists of two features. The numeric variable *balance* and binary variable *gender*.

```
capital <- read.table('capital.csv', header = TRUE, sep = ';')
attach(capital)
head(capital)
```

```
##    balance gender
## 1     1018      1
## 2     1021      1
## 3     1081      1
## 4       300      1
## 5       769      2
## 6       486      1
```

```
head(is.na.data.frame(capital))
```

```
##      balance gender
## [1,]   FALSE  FALSE
## [2,]   FALSE  FALSE
## [3,]   FALSE  FALSE
## [4,]   FALSE  FALSE
## [5,]   FALSE  FALSE
## [6,]   FALSE  FALSE
```

Relative frequency

We will explore *balance* in relation to *gender*. Particularly, we are interested in frequency tables, bar and pie plots. Frequency table is defined as:

$$\text{Relative } \hat{\text{Frequency}} = \frac{\text{Frequency}}{\text{Sample Size}} \quad (1)$$

Get possible values of *gender*

```
factor(capital$gender)
```

```
##    [1] 1 1 1 1 2 1 1 1 1 2 1 1 1 1 2 1 1 1 1 2 1 2 1 1 1 1 2 1 2 1 1 1 2 1
##   [38] 2 1 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 2 2 1 2 1 1 1 1 1
##   [75] 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 2 1 1 1 2 1 2 1 1 2 2 2 1 1 2 1 1 1 1 1
##  [112] 1 1 2 1 1 1 2 2 1 2 1 1 1 1 2 1 1 2 1 2 1 1 1 1 1 1 1 2 1 2 2 1 1 1 2 1 1
##  [149] 1 2 2 1 1 1 2 2 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 2 2 1 2 1 2 1 2 1
```

```
gender.freq = table(capital$gender)
gender.relfreq = gender.freq / nrow(capital)
```

| Category | gender:relfreq |
|----------|----------------|
| 1 | 0.78 |
| 2 | 0.22 |

