

Aufgabe QualiDaten

Dominik Peter

3 11 2016

Lösung zu Aufgabe QualiDaten

Laden Packages und Daten

```
library(data.table)
library(MASS)

df <- data.table(Painter = rownames painters), painters)
head(df, 5)
```

##	Painter	Composition	Drawing	Colour	Expression	School
## 1:	Da Udine	10	8	16	3	A
## 2:	Da Vinci	15	16	4	14	A
## 3:	Del Piombo	8	13	16	7	A
## 4:	Del Sarto	12	16	9	8	A
## 5:	Fr. Penni	0	15	8	0	A

Aufgabe 1

Bestimmen Sie die Häufigkeitsverteilung der Variablen

```
N <- table(df$Composition)
cbind(N)
```

##	N
## 0	1
## 4	3
## 5	1
## 6	3
## 8	6
## 9	1
## 10	6
## 11	2
## 12	4
## 13	5
## 14	3
## 15	14
## 16	2
## 17	1
## 18	2

Aufgabe 2

Bestimmen Sie mit R diejenige Schule mit den meisten Malern

```
# Anzahl nach Schule
count.school <- df[, .N, by=School]

# Schule mit max. Anzahl
count.school[N == max(N)]
```

```
##      School  N
## 1:         A 10
## 2:         D 10
```

Die Schulen **A** und **D** haben jeweils **10** Maler

Aufgabe 3

Bestimmen Sie die relative Häufigkeitsverteilung der Variablen Composition

```
# relative Häufigkeitsverteilung
rh <- prop.table(table(df$Composition))
cbind(rh)
```

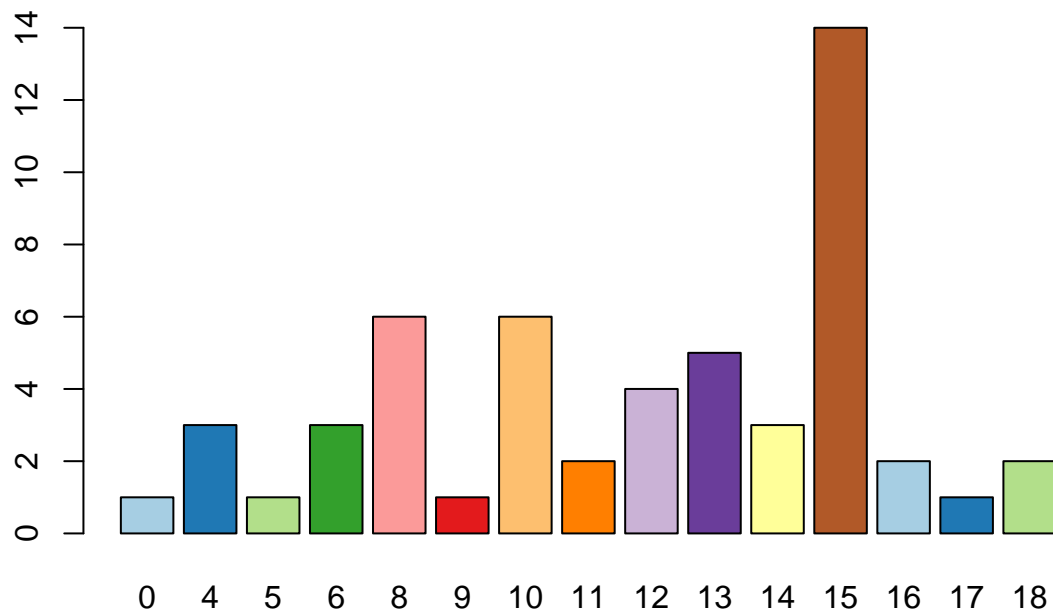
```
##      rh
## 0 0.0185
## 4 0.0556
## 5 0.0185
## 6 0.0556
## 8 0.1111
## 9 0.0185
## 10 0.1111
## 11 0.0370
## 12 0.0741
## 13 0.0926
## 14 0.0556
## 15 0.2593
## 16 0.0370
## 17 0.0185
## 18 0.0370
```

Aufgabe 4

Stellen Sie die Häufigkeitsverteilung der Variablen Composition mit einem Balkendiagramm dar

```
# Barplot der Häufigkeitsverteilung Composition
tbl <- table(df$Composition)
colors <- RColorBrewer::brewer.pal(nrow(tbl), "Paired")

barplot(tbl, col = colors)
```

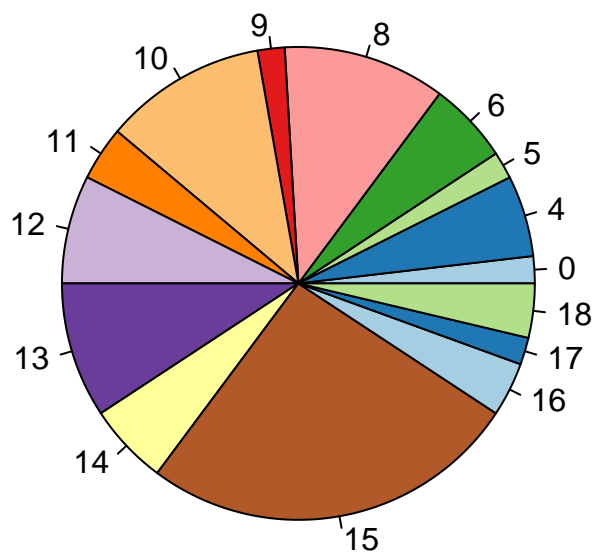


Aufgabe 5

Stellen Sie die Häufigkeitsverteilung der Variablen Composition mit einem Kuchendiagramm dar

```
# Barplot der Häufigkeitsverteilung Composition
tbl <- table(df$Composition)
colors <- RColorBrewer::brewer.pal(nrow(tbl), "Paired")

pie(tbl, col = colors, radius = 1)
```



Aufgabe 6

Finden Sie die Schule mit dem höchsten Wert der Variable Composition

```
# Schule
df[Composition == max(Composition), .(School, Composition)]

##      School Composition
## 1:      E           18
## 2:      G           18
```

Schule **E** und **G** haben Werte mit **Composition** mit dem Wert **18**

Aufgabe 7

Bestimmen Sie den Anteil aller Maler, deren Color-Wert mindestens 14 beträgt

```
# Variable Color: Kontrolle als TRUE and FALSE
check <- df[, ColorCheck := Colour >= 14]
head(check, 5)

##      Painter Composition Drawing Colour Expression School ColorCheck
## 1:  Da Udine           10      8     16           3      A      TRUE
## 2:  Da Vinci           15     16      4          14      A     FALSE
## 3: Del Piombo           8     13     16           7      A      TRUE
## 4: Del Sarto           12     16      9           8      A     FALSE
## 5: Fr. Penni           0     15      8           0      A     FALSE

# Anzahl Maler mit Color >= 14
sum.ColorCheck <- sum(check$ColorCheck)

# Relativer Anteil Maler mit Color >= 14
mean.ColorCheck <- mean(check$ColorCheck)

paste0("Anzahl Maler >= 14: ", sum.ColorCheck)

## [1] "Anzahl Maler >= 14: 20"

paste0("Relative Anzahl Maler >= 14: ", round(mean.ColorCheck * 100, 2), "%")

## [1] "Relative Anzahl Maler >= 14: 37.04%"
```

20 Maler haben Colour >= 14. Das sind **37.04** % aller Maler in den Beispieldaten