

Foundations of R Software

Lecture 43

Data Frames: Some More Operations

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

Data Frames

An example data frame `painters` is available in the library MASS (here only an excerpt of a data set):

```
> library(MASS)
```

```
> painters
```

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

Here, the names of the painters serve as row identifications, i.e., every row is assigned to the name of the corresponding painter.

Data Frames

R Console

```
> library(MASS)
> painters
```

	Composition	Drawing	Colour	Expression	School
Da Udine	10	8	16	3	A
Da Vinci	15	16	4	14	A
Del Piombo	8	13	16	7	A
Del Sarto	12	16	9	8	A
Fr. Penni	0	15	8	0	A
Guilio Romano	15	16	4	14	A
	.	*	.	.	*
	.	*	.	.	*
	.	*	.	.	*
	.	*	.	.	*
Rubens	18	13	17	17	G
Teniers	15	12	13	6	G
Van Dyck	15	10	17	13	G
Bourdon	10	8	8	4	H
Le Brun	16	16	8	16	H

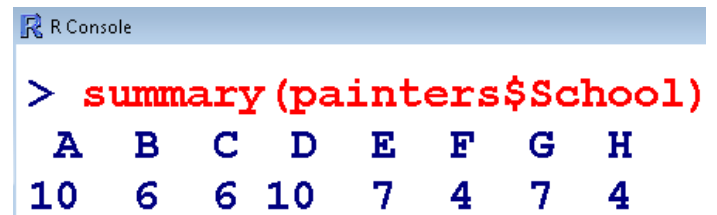
Data Frames

The **summary** function for a categorical variable returns a detailed frequency table:

summary is a generic function used to produce result summaries of the results of various model fitting functions.

```
> summary(painters$School)
```

A	B	C	D	E	F	G	H
10	6	6	10	7	4	7	4



```
R Console  
> summary(painters$School)  
  A  B  C  D  E  F  G  H  
10  6  6 10  7  4  7  4
```

Data Frames

❑ Attaching a data frame

With a command `attach()` over the data frame, the variables can be referenced directly by name.

It can address the names of a data frame directly, without the prefix dollar sign operator, e.g. `painters$`.

Example

```
> attach(painters)
```

Variable names are

- `Composition`,
- `Drawing`,
- `Colour`,
- `Expression`,
- `School`

Data Frames

> summary(School) # Character variable

A	B	C	D	E	F	G	H
10	6	6	10	7	4	7	4

```
R Console
> attach(painters)
> summary(School)
  A  B  C  D  E  F  G  H
10  6  6 10  7  4  7  4
```

> summary(Composition) # Numeric variable

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	8.25	12.50	11.56	15.00	18.00

```
R Console
> summary(Composition)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.00   8.25   12.50   11.56   15.00   18.00
```

Data Frames

- ❑ The command `detach()` recovers the default setting and then we have to use `painters$` again.

```
> detach(painters)
```

```
> summary(School)
```

```
Error in summary(School) : Object "School" not found
```

R Console

```
> detach(painters)
```

```
> summary(School)
```

```
Error in summary(School) : object 'School' not found
```

Data Frames

Subsets of a data frame can be obtained with `subset()` or with the second equivalent command:

```
> subset painters, School=='F')
```

(# == means logical equal sign)

	Composition	Drawing	Colour	Expression	School
Durer	8	10	10	8	F
Holbein	9	10	16	13	F
Pourbus	4	15	6	6	F
VanLeyden	8	6	6	4	F

Data Frames

Similar outcome can be also obtained from

```
> painters[ painters[["School"]] == "F", ]
```

	Composition	Drawing	Colour	Expression	School
Durer	8	10	10	8	F
Holbein	9	10	16	13	F
Pourbus	4	15	6	6	F
VanLeyden	8	6	6	4	F

```
R Console
> painters[ painters[["School"]] == "F", ]
      Composition Drawing Colour Expression School
Durer           8      10     10           8      F
Holbein          9      10     16          13      F
Pourbus          4      15      6           6      F
Van Leyden       8       6      6           4      F
```

Data Frames

Subsets of a data frame can be obtained with `subset()` or with the second equivalent command:

```
> subset painters, Composition <= 6)
```

```
R Console
> subset(painters, Composition <= 6)
      Composition Drawing Colour Expression School
Fr. Penni         0      15      8          0      A
Perugino          4      12     10          4      A
Bassano           6       8     17          0      D
Bellini           4       6     14          0      D
Murillo           6       8     15          4      D
Palma Vecchio     5       6     16          0      D
Caravaggio        6       6     16          0      E
Pourbus           4      15      6          6      F
> █
```

Data Frames

❑ Uninteresting columns can be eliminated.

```
> subset painters, School=="F", select=c(-3,-5))
```

	Composition	Drawing	Expression
Durer	8	10	8
Holbein	9	10	13
Pourbus	4	15	6
Van Leyden	8	6	4

The third and the fifth column (Colour and School) are not shown.

```
R Console
> subset(painters, School=="F", select=c(-3,-5))
      Composition Drawing Expression
Durer           8      10          8
Holbein          9      10         13
Pourbus          4      15          6
Van Leyden       8       6          4
> █
```

Data Frames

- ❑ The command `split` partitions the data set by values of a specific variable. This should preferably be a factor variable.

Example: Following command splits `painters` with respect to `School` (A,B,C,... categories)

```
> splitted = split(painters, painters$School)
```

Data Frames

> splitted

\$A

	Composition	Drawing	Colour	Expression	School
Da Udine	10	8	16	3	A
Da Vinci	15	16	4	14	A
Del Piombo	8	13	16	7	A
Del Sarto	12	16	9	8	A
Fr. Penni	0	15	8	0	A
Guilio Romano	15	16	4	14	A
Michelangelo	8	17	4	8	A
Perino del Vaga	15	16	7	6	A
Perugino	4	12	10	4	A
Raphael	17	18	12	18	A

\$B

	Composition	Drawing	Colour	Expression	School
F. Zucarro	10	13	8	8	B
Fr. Salviata	13	15	8	8	B
Parmigiano	10	15	6	6	B
Primaticcio	15	14	7	10	B
T. Zucarro	13	14	10	9	B
Volterra	12	15	5	8	B

Contd...

Data Frames

\$C

	Composition	Drawing	Colour	Expression	School
Barocci	14	15	6	10	C
Cortona	16	14	12	6	C
Josepin	10	10	6	2	C
L. Jordaens	13	12	9	6	C
Testa	11	15	0	6	C
Vanius	15	15	12	13	C

\$D

	Composition	Drawing	Colour	Expression	School
Bassano	6	8	17	0	D
Bellini	4	6	14	0	D
Giorgione	8	9	18	4	D
Murillo	6	8	15	4	D
Palma Giovane	12	9	14	6	D
Palma Vecchio	5	6	16	0	D
Pordenone	8	14	17	5	D
Tintoretto	15	14	16	4	D
Titian	12	15	18	6	D
Veronese	15	10	16	3	D

Contd...

Data Frames

\$E

	Composition	Drawing	Colour	Expression	School
Albani	14	14	10	6	E
Caravaggio	6	6	16	0	E
Corregio	13	13	15	12	E
Domenichino	15	17	9	17	E
Guercino	18	10	10	4	E
Lanfranco	14	13	10	5	E
The Carraci	15	17	13	13	E

\$F

	Composition	Drawing	Colour	Expression	School
Durer	8	10	10	8	F
Holbein	9	10	16	13	F
Pourbus	4	15	6	6	F
Van Leyden	8	6	6	4	F

Contd...

Data Frames

\$G

	Composition	Drawing	Colour	Expression	School
Diepenbeck	11	10	14	6	G
J. Jordaens	10	8	16	6	G
Otho Venius	13	14	10	10	G
Rembrandt	15	6	17	12	G
Rubens	18	13	17	17	G
Teniers	15	12	13	6	G
Van Dyck	15	10	17	13	G

\$H

	Composition	Drawing	Colour	Expression	School
Bourdon	10	8	8	4	H
Le Brun	16	16	8	16	H
Le Suer	15	15	4	15	H
Poussin	15	17	6	15	H

Remark: If the data set is not attached, we have to use
painters\$School.

Data Frames

R Console

```
> splitted = split painters, painters$School
> splitted
```

```
$A
      Composition Drawing Colour Expression School
Da Udine           10         8      16          3      A
Da Vinci            15        16       4         14      A
Del Piombo          8         13      16          7      A
Del Sarto           12        16       9          8      A
Fr. Penni           0         15       8          0      A
Guilio Romano       15        16       4         14      A
Michelangelo        8         17       4          8      A
Perino del Vaga     15        16       7          6      A
Perugino            4         12      10          4      A
Raphael            17        18      12         18      A
```

```
$B
      Composition Drawing Colour Expression School
F. Zucarro          10        13       8          8      B
Fr. Salviata         13        15       8          8      B
Parmigiano           10        15       6          6      B
Primaticcio          15        14       7         10      B
T. Zucarro           13        14      10          9      B
Volterra            12        15       5          8      B
```

```
$C
      Composition Drawing Colour Expression School
Barocci              14        15       6         10      C
Cortona              16        14      12          6      C
Josepin              10        10       6          2      C
L. Jordaens          13        12       9          6      C
Testa                11        15       0          6      C
Vanius               15        15      12         13      C
```

```
$D
      Composition Drawing Colour Expression School
Bassano              6         8      17          0      D
Bellini              4         6      14          0      D
Giorgione            8         9      18          4      D
Murillo              6         8      15          4      D
Palma Giovane        12         9      14          6      D
Palma Vecchio        5         6      16          0      D
Pordenone            8        14      17          5      D
```

R Console

```
$E
```

```
      Composition Drawing Colour Expression School
Albani               14        14      10          6      E
Caravaggio           6         6      16          0      E
Corregio             13        13      15         12      E
Domenichino          15        17       9         17      E
Guercino             18        10      10          4      E
Lanfranco            14        13      10          5      E
The Carraci          15        17      13         13      E
```

```
$F
```

```
      Composition Drawing Colour Expression School
Durer                8         10      10          8      F
Holbein              9         10      16         13      F
Pourbus              4         15       6          6      F
Van Leyden           8         6       6          4      F
```

```
$G
```

```
      Composition Drawing Colour Expression School
Diepenbeck           11        10      14          6      G
J. Jordaens          10         8      16          6      G
Otho Venius          13        14      10         10      G
Rembrandt            15         6      17         12      G
Rubens               18        13      17         17      G
Teniers              15        12      13          6      G
Van Dyck             15        10      17         13      G
```

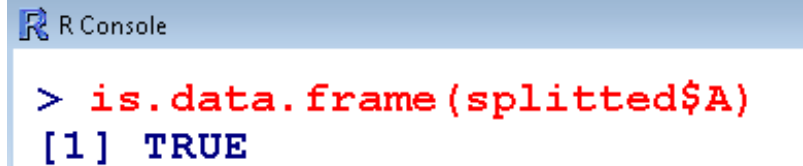
```
$H
```

```
      Composition Drawing Colour Expression School
Bourdon              10         8       8          4      H
Le Brun              16        16       8         16      H
Le Suer              15        15       4         15      H
Poussin              15        17       6         15      H
```

Data Frames

The objects `splitted$A` to `splitted$H` are themselves data frames:

```
> is.data.frame(splitted$A)
[1] TRUE
```

A screenshot of an R console window. The title bar is light blue and contains the R logo and the text "R Console". The console area has a white background and shows the command `> is.data.frame(splitted$A)` in red text, followed by the output `[1] TRUE` in blue text.

```
> is.data.frame(splitted$A)
[1] TRUE
```