

Appending and updating rows and columns

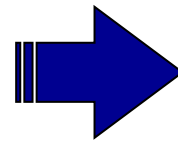
Appending rows into data.table

myData

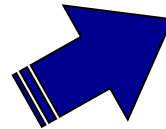
Customer	TransDate	Quantity	PurchAmount	Cost
149332	15.11.2005	1	199.95	107.00
172951	29.08.2008	1	199.95	108.00
120621	19.10.2007	1	99.95	49.00
149236	14.11.2005	1	39.95	18.95
149236	12.06.2007	1	79.95	35.00
...

New row

Customer	TransDate	Quantity	PurchAmount	Cost
199332	18.11.2011	1	99.95	47.00



Append the row
to the data.table
object



Customer	TransDate	Quantity	PurchAmount	Cost
149332	15.11.2005	1	199.95	107.00
172951	29.08.2008	1	199.95	108.00
120621	19.10.2007	1	99.95	49.00
149236	14.11.2005	1	39.95	18.95
149236	12.06.2007	1	79.95	35.00
199332	18.11.2011	1	99.95	47.00
...

```
rbind(myData, data.table(Customer=199332,
  TransDate=dmy("18.11.2011"),
  Quantity=1, PurchAmount=99.95,
  Cost=47.00) )
```

1
New row

2
Make sure the formats
are correctly set!

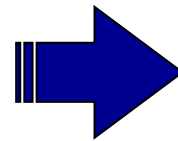
Appending rows into data.table

myData

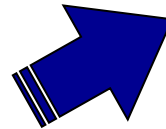
Customer	TransDate	Quantity	PurchAmount	Cost
149332	15.11.2005	1	199.95	107.00
172951	29.08.2008	1	199.95	108.00
120621	19.10.2007	1	99.95	49.00
149236	14.11.2005	1	39.95	18.95
149236	12.06.2007	1	79.95	35.00
...

New row

Customer	TransDate	Quantity	PurchAmount	Cost
199332	18.11.2011	1	99.95	47.00



Append the row
to the data.table
object



Customer	TransDate	Quantity	PurchAmount	Cost
149332	15.11.2005	1	199.95	107.00
172951	29.08.2008	1	199.95	108.00
120621	19.10.2007	1	99.95	49.00
149236	14.11.2005	1	39.95	18.95
149236	12.06.2007	1	79.95	35.00
199332	18.11.2011	1	99.95	47.00
...

```
rbind(myData, data.table(Customer=199332,
  TransDate=dmy("18.11.2011"),
  Quantity=1, PurchAmount=99.95,
  Cost=47.00) )
```

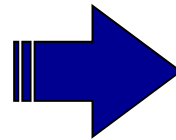
1
New row

2
Make sure the formats
are correctly set!

Adding columns to a data.table

Customer	TransDate	Quantity	PurchAmount	Cost
149332	15.11.2005	1	199.95	107.00
172951	29.08.2008	1	199.95	108.00
120621	19.10.2007	1	99.95	49.00
149236	14.11.2005	1	39.95	18.95
149236	12.06.2007	1	79.95	35.00
...

Add column
NewCol



Customer	TransDate	Quantity	PurchAmount	Cost	NewCol
149332	15.11.2005	1	199.95	107.00	107
172951	29.08.2008	1	199.95	108.00	108
120621	19.10.2007	1	99.95	49.00	49
149236	14.11.2005	1	39.95	18.95	19
149236	12.06.2007	1	79.95	35.00	35
...

New column ¹

Column name ²

R Base

```
myData$NewCol <-  
  round(myData$Cost)
```

Column value ³

data.table package

```
myData[, NewCol := round(Cost)]
```

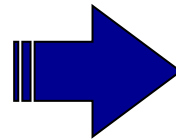
Column name ⁴

Column value ⁵

Adding columns to a data.table

Customer	TransDate	Quantity	PurchAmount	Cost
149332	15.11.2005	1	199.95	107.00
172951	29.08.2008	1	199.95	108.00
120621	19.10.2007	1	99.95	49.00
149236	14.11.2005	1	39.95	18.95
149236	12.06.2007	1	79.95	35.00
...

Add column
NewCol



Customer	TransDate	Quantity	PurchAmount	Cost	NewCol
149332	15.11.2005	1	199.95	107.00	107
172951	29.08.2008	1	199.95	108.00	108
120621	19.10.2007	1	99.95	49.00	49
149236	14.11.2005	1	39.95	18.95	19
149236	12.06.2007	1	79.95	35.00	35
...

New column ¹

Column name ²

R Base

```
myData$NewCol <-  
  round(myData$Cost)
```

Column value ³

data.table package

```
myData[, NewCol := round(Cost)]
```

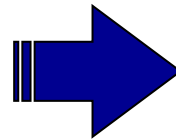
Column name ⁴

Column value ⁵

Adding columns to a data.table

Customer	TransDate	Quantity	PurchAmount	Cost
149332	15.11.2005	1	199.95	107.00
172951	29.08.2008	1	199.95	108.00
120621	19.10.2007	1	99.95	49.00
149236	14.11.2005	1	39.95	18.95
149236	12.06.2007	1	79.95	35.00
...

Add column
NewCol



Customer	TransDate	Quantity	PurchAmount	Cost	NewCol
149332	15.11.2005	1	199.95	107.00	107
172951	29.08.2008	1	199.95	108.00	108
120621	19.10.2007	1	99.95	49.00	49
149236	14.11.2005	1	39.95	18.95	19
149236	12.06.2007	1	79.95	35.00	35
...

New column ¹

Column name ²

R Base

```
myData$NewCol <-  
  round(myData$Cost)
```

Column value ³

data.table package

```
myData[, NewCol := round(Cost)]
```

Column name ⁴

Column value ⁵

R Basics: By default data.table changes are made by reference

This can result in a messy situation:

```
head(myData, n=3)
```

1
myData before adding a new column

Customer	TransDate	Quantity	PurchAmount	Cost
149332	15.11.2005	1	199.95	107.00
172951	29.08.2008	1	199.95	108.00
120621	19.10.2007	1	99.95	49.00

```
myData.modified <- myData
```

2
Duplicate myData

```
myData.modified[, NewCol :=round(Cost)]
```

3
Add a new column to myData_modified

```
head(myData, n=3)
```

Customer	TransDate	Quantity	PurchAmount	Cost	NewCol
149332	15.11.2005	1	199.95	107.00	107.0
172951	29.08.2008	1	199.95	108.00	108.0
120621	19.10.2007	1	99.95	49.00	49.0

➤ Use copy() to avoid this:

```
myData.modified <- copy(myData)
```

4
Changes are made to myData too

R Basics: By default data.table changes are made by reference

This can result in a messy situation:

```
head(myData, n=3)
```

1
myData before adding a new column

Customer	TransDate	Quantity	PurchAmount	Cost
149332	15.11.2005	1	199.95	107.00
172951	29.08.2008	1	199.95	108.00
120621	19.10.2007	1	99.95	49.00

```
myData.modified <- myData
```

2
Duplicate myData

```
myData.modified[, NewCol :=round(Cost)]
```

3
Add a new column to myData_modified

```
head(myData, n=3)
```

Customer	TransDate	Quantity	PurchAmount	Cost	NewCol
149332	15.11.2005	1	199.95	107.00	107.0
172951	29.08.2008	1	199.95	108.00	108.0
120621	19.10.2007	1	99.95	49.00	49.0

➤ Use copy() to avoid this:

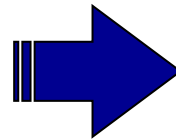
```
myData.modified <- copy(myData)
```

4
Changes are made to myData too

Remove a single column from a data.table

Customer	TransDate	Quantity	PurchAmount	Cost
149332	15.11.2005	1	199.95	107.00
172951	29.08.2008	1	199.95	108.00
120621	19.10.2007	1	99.95	49.00
149236	14.11.2005	1	39.95	18.95
149236	12.06.2007	1	79.95	35.00
...

Remove column
Quantity



Customer	TransDate	PurchAmount	Cost
149332	15.11.2005	199.95	107.00
172951	29.08.2008	199.95	108.00
120621	19.10.2007	99.95	49.00
149236	14.11.2005	39.95	18.95
149236	12.06.2007	79.95	35.00
...

R Base

```
myData$Quantity <- NULL
```

Assign NULL to the column ¹

data.table package

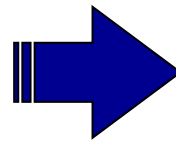
```
myData[, Quantity := NULL]
```

Changes made by reference! ²

Remove multiple columns at the same time

Customer	TransDate	Quantity	PurchAmount	Cost
149332	15.11.2005	1	199.95	107.00
172951	29.08.2008	1	199.95	108.00
120621	19.10.2007	1	99.95	49.00
149236	14.11.2005	1	39.95	18.95
149236	12.06.2007	1	79.95	35.00
...

Remove column
Quantity
and Cost



Customer	TransDate	PurchAmount
149332	15.11.2005	199.95
172951	29.08.2008	199.95
120621	19.10.2007	99.95
149236	14.11.2005	39.95
149236	12.06.2007	79.95
...

Use quotation marks ²

```
myData[, c("Quantity", "Cost") := NULL]
```

Vector ¹

R Basics: Change the name of a variable in a data.table

To change names of a data.table variable:

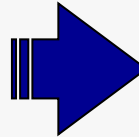
Old name ¹

New name ²

```
setnames(myData, "PurchAmount", "TransValue")
```

Customer	TransDate	Quantity	PurchAmount	Cost
149332	15.11.2005	1	199.95	107.00
172951	29.08.2008	1	199.95	108.00
120621	19.10.2007	1	99.95	49.00
149236	14.11.2005	1	39.95	18.95
149236	12.06.2007	1	79.95	35.00
...

Change variable
name
PurchAmount

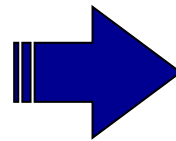


Customer	TransDate	Quantity	TransValue	Cost
149332	15.11.2005	1	199.95	107.00
172951	29.08.2008	1	199.95	108.00
120621	19.10.2007	1	99.95	49.00
149236	14.11.2005	1	39.95	18.95
149236	12.06.2007	1	79.95	35.00
...

Remove rows from a data.table

Customer	TransDate	Quantity	PurchAmount	Cost
149332	15.11.2005	1	199.95	107.00
172951	29.08.2008	1	199.95	108.00
120621	19.10.2007	1	99.95	49.00
149236	14.11.2005	1	39.95	18.95
149236	12.06.2007	1	79.95	35.00
...

Remove
first 3 rows



Customer	TransDate	Quantity	PurchAmount	Cost
149236	14.11.2005	1	39.95	18.95
149236	12.06.2007	1	79.95	35.00
...

Delete the first three
rows in the data table

1

```
myData[-(1:3),]
```

```
myData[(4:.N),]
```

Keep rows 4 to last

2