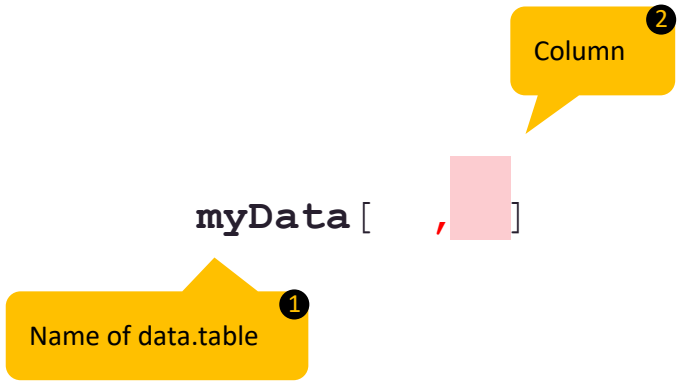
Selecting columns

There are multiple ways of selecting columns

1. Select a single column
2. Select multiple columns
3. Combine operations to select by rows and columns

Enter your selection commands in the column placeholder

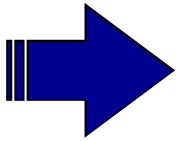
Customer	TransDate	Quantity	PurchAmount	Cost
149332	15.11.2005	1	199.95	107.00
172951	29.08.2008	1	199.95	108.00
120621	19.10.2007	1	99.95	49.00
149236	14.11.2005	1	39.95	18.95
149236	12.06.2007	1	79.95	35.00
...	...	...	...	...



Select a single column by column name / number (1/2)

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
...	...	...	...	...

Select column `TransDate`



TransDate
15.11.2005
29.08.2008
19.10.2007
14.11.2005
12.06.2007
...

1 Returns a data.table

2 **list**, not a **vector**

3 Column name

```
myData[, list(TransDate)]
```

4 No quotation marks

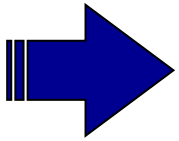
```
myData[, 2]
```

5 Column number

Select a single column by column name / number (2/2)

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
...	...	...	...	...

Select column `TransDate` and return it as a vector



```
c(15.11.2005, 29.08.2008,
  19.10.2007, 14.11.2005,
  12.06.2007, ...)
```

Returns a vector 1

R Base

```
myData$TransDate
```

Column name 2

data.table package

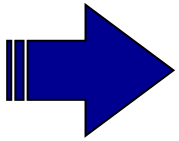
```
myData[, TransDate]
```

Column name 3

Select a single column by column name / number (2/2)

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
...	...	...	...	...

Select column `TransDate` and return it as a vector



```
c(15.11.2005, 29.08.2008,
  19.10.2007, 14.11.2005,
  12.06.2007, ...)
```

Returns a vector 1

R Base

```
myData$TransDate
```

Column name 2

data.table package

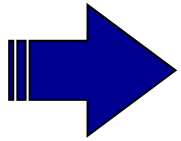
```
myData[, TransDate]
```

Column name 3

Select multiple columns by column name / number

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
...	...	...	...	...

Select columns `Customer`, `TransDate`, and `PurchAmount`



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
...	...	...

1 Returns a data.table

2 List of variable names

```
myData[, list(Customer, TransDate, PurchAmount)]
```

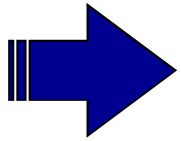
```
myData[, c(1:2, 4)]
```

3 Vector with variable numbers

Select multiple columns by column name / number

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
...	...	...	...	...

Select columns `Customer`, `TransDate`, and `PurchAmount`



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
...	...	...

1 Returns a data.table

2 List of variable names

```
myData[, list(Customer, TransDate, PurchAmount)]
```

```
myData[, c(1:2, 4)]
```

3 Vector with variable numbers

R Basics: Find out the column names

Get all column names from your data.table:

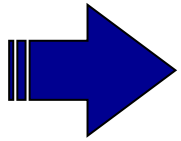
```
> names(myData)
```

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	…
…	…	…	…	…	…

Combine operations to select by rows and columns

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
...	...	...	...	...

Select column `TransDate` and `Cost` for entries where `PurchAmount` > 100



TransDate	Cost
15.11.2005	107.00
29.08.2008	108.00
...	...

```
myData[PurchAmount > 100, list(TransDate, Cost)]
```

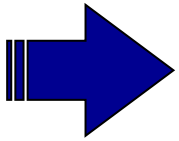
```
subset(myData, PurchAmount > 100, select = c(TransDate, Cost))
```

Alternatively, use `subset()`

Combine operations to select by rows and columns

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
...	...	...	...	...

Select column `TransDate` and `Cost` for entries where `PurchAmount` > 100



TransDate	Cost
15.11.2005	107.00
29.08.2008	108.00
...	...

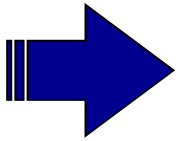
```
myData[PurchAmount > 100, list(TransDate, Cost)]
subset(myData, PurchAmount > 100, select = c(TransDate, Cost))
```

Alternatively, use `subset()`

Combine operations to select by rows and columns

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
...	...	...	...	...

Select column `TransDate` and `Cost` for entries where `PurchAmount` > 100



TransDate	Cost
15.11.2005	107.00
29.08.2008	108.00
...	...

```
myData[PurchAmount > 100, list(TransDate, Cost)]
```

```
subset(myData, PurchAmount > 100, select = c(TransDate, Cost))
```

Alternatively, use `subset()`