

Reshaping data



Getting and Cleaning Data

wide data

	A	B	C	D	E	F	G
1	ID	LastName	FirstName	Height_inches	Weight_lbs	Insulin	Glucose
2	1004	Smith	Jane	65	180	0.60	163
3	4587	Nayef	Mohammed	75	215	1.46	150
4	1727	Doe	Janice	62	124	0.72	177
5	6879	Jordan	Alex	77	160	1.23	205



long data

	A	B	C
1	ID	Variable	Value
2	1004	LastName	Smith
3	4587	LastName	Nayef
4	1727	LastName	Doe
5	6879	LastName	Jordan
6	1004	FirstName	Jane
7	4587	FirstName	Mohammed
8	1727	FirstName	Janice
9	6879	FirstName	Alex
10	1004	Height_inches	65
11	4587	Height_inches	75
12	1727	Height_inches	62
13	6879	Height_inches	77
14	1004	Weight_lbs	180
15	4587	Weight_lbs	215
16	1727	Weight_lbs	124
17	6879	Weight_lbs	160
18	1004	Insulin	0.60
19	4587	Insulin	1.46
20	1727	Insulin	0.72
21	6879	Insulin	1.23
22	1004	Glucose	163
23	4587	Glucose	150
24	1727	Glucose	177
25	6879	Glucose	205



wide data

	A	B	C	D	E	F	G
1	ID	LastName	FirstName	Height_inches	Weight_lbs	Insulin	Glucose
2	1004	Smith	Jane	65	180	0.60	163
3	4587	Nayef	Mohammed	75	215	1.46	150
4	1727	Doe	Janice	62	124	0.72	177
5	6879	Jordan	Alex	77	160	1.23	205

reshaping data



long data

	A	B	C
1	ID	Variable	Value
2	1004	LastName	Smith
3	4587	LastName	Nayef
4	1727	LastName	Doe
5	6879	LastName	Jordan
6	1004	FirstName	Jane
7	4587	FirstName	Mohammed
8	1727	FirstName	Janice
9	6879	FirstName	Alex
10	1004	Height_inches	65
11	4587	Height_inches	75
12	1727	Height_inches	62
13	6879	Height_inches	77
14	1004	Weight_lbs	180
15	4587	Weight_lbs	215
16	1727	Weight_lbs	124
17	6879	Weight_lbs	160
18	1004	Insulin	0.60
19	4587	Insulin	1.46
20	1727	Insulin	0.72
21	6879	Insulin	1.23
22	1004	Glucose	163
23	4587	Glucose	150
24	1727	Glucose	177
25	6879	Glucose	205

```
> head(airquality)
```

	ozone	solar.r	wind	temp	month	day
1	41	190	7.4	67	5	1
2	36	118	8.0	72	5	2
3	12	149	12.6	74	5	3
4	18	313	11.5	62	5	4
5	NA	NA	14.3	56	5	5
6	28	NA	14.9	66	5	6



```
## install the package  
install.packages('tidyr')
```

```
## load the package into R Session  
library(tidyr)
```



```
## use gather() to reshape from wide to long  
gathered <- gather(airquality)
```

```
## take a look at first few rows of long data  
head(gathered)
```



```
> head(airquality)
```

	ozone	solar.r	wind	temp	month	day
1	41	190	7.4	67	5	1
2	36	118	8.0	72	5	2
3	12	149	12.6	74	5	3
4	18	313	11.5	62	5	4
5	NA	NA	14.3	56	5	5
6	28	NA	14.9	66	5	6

gather(airquality)



```
> head(gathered)
```


	key	value
1	ozone	41
2	ozone	36
3	ozone	12
4	ozone	18
5	ozone	NA
6	ozone	28




```
> head(airquality)
```

	ozone	solar.r	wind	temp	month	day
1	41	190	7.4	67	5	1
2	36	118	8.0	72	5	2
3	12	149	12.6	74	5	3
4	18	313	11.5	62	5	4
5	NA	NA	14.3	56	5	5
6	28	NA	14.9	66	5	6

gather(airquality,
key = "variable",
value="val")



```
> head(gathered)
```

	variable	value
--	----------	-------


1	ozone	41
2	ozone	36
3	ozone	12
4	ozone	18
5	ozone	NA
6	ozone	28



```
> head(airquality)
```

	ozone	solar.r	wind	temp	month	day
1	41	190	7.4	67	5	1
2	36	118	8.0	72	5	2
3	12	149	12.6	74	5	3
4	18	313	11.5	62	5	4
5	NA	NA	14.3	56	5	5
6	28	NA	14.9	66	5	6

```
gather(airquality,  
key="variable",  
value="value",  
ozone, solar.r, wind, temp)
```



```
> head(gathered)
```

	month	day	variable	value
1	5	1	ozone	41
2	5	2	ozone	36
3	5	3	ozone	12
4	5	4	ozone	18
5	5	5	ozone	NA
6	5	6	ozone	28

ozone, solar.r, wind, temp:

The variables to move into the variable column.

The other variables keep their original column.



```
## Use spread() to reshape from long to wide  
spread_data <- spread(gathered, key=variable, value=value)
```

```
## Take a look at the spread data  
head(spread_data)
```

```
## Compare that back to the original  
head(airquality)
```



```
> head(spread_data)
```

	month	day	ozone	solar.r	temp	wind
1	5	1	41	190	67	7.4
2	5	2	36	118	72	8.0
3	5	3	12	149	74	12.6
4	5	4	18	313	62	11.5
5	5	5	NA	NA	56	14.3
6	5	6	28	NA	66	14.9

```
> head(airquality)
```

	ozone	solar.r	wind	temp	month	day
1	41	190	7.4	67	5	1
2	36	118	8.0	72	5	2
3	12	149	12.6	74	5	3
4	18	313	11.5	62	5	4
5	NA	NA	14.3	56	5	5
6	28	NA	14.9	66	5	6



```
## install the package  
install.packages( 'reshape2' )  
  
## load the package into R Session  
library(reshape2)
```



```
## puts each column name into the 'variable' column  
## puts corresponding variable's value in 'value' column  
melted <- melt(airquality)
```

```
## let's take a look at the top of the melted data frame  
head(melted)
```

```
## and at the bottom of that melted data frame  
tail(melted)
```



```
> head(airquality)
```

	ozone	solar.r	wind	temp	month	day
1	41	190	7.4	67	5	1
2	36	118	8.0	72	5	2
3	12	149	12.6	74	5	3
4	18	313	11.5	62	5	4
5	NA	NA	14.3	56	5	5
6	28	NA	14.9	66	5	6

melt(airquality)



```
> head(melted)
```

	variable	value
1	ozone	41
2	ozone	36
3	ozone	12
4	ozone	18
5	ozone	NA
6	ozone	28


```
## melt the data frame  
## specify each row using month and day  
melted <- melt(airquality, id.vars = c("month", "day"))  
  
## look at the first few rows of the melted data frame  
head(melted)
```




```
> head(airquality)
```

	ozone	solar.r	wind	temp	month	day
1	41	190	7.4	67	5	1
2	36	118	8.0	72	5	2
3	12	149	12.6	74	5	3
4	18	313	11.5	62	5	4
5	NA	NA	14.3	56	5	5
6	28	NA	14.9	66	5	6

`melt(airquality,
id.vars = c("month","day"))`



```
> head(melted)
```

	month	day	variable	value
1	5	1	ozone	41
2	5	2	ozone	36
3	5	3	ozone	12
4	5	4	ozone	18
5	5	5	ozone	NA
6	5	6	ozone	28



```
## to get our data back to its original form  
## specify which columns should be combined to use as identifiers  
## and which column should be used to specify the columns  
original <- dcast(melted, month + day ~ variable)  
  
head(original)  
  
head(airquality)
```



```
> head(original)
```

	month	day	ozone	solar.r	wind	temp
1	5	1	41	190	7.4	67
2	5	2	36	118	8.0	72
3	5	3	12	149	12.6	74
4	5	4	18	313	11.5	62
5	5	5	NA	NA	14.3	56
6	5	6	28	NA	14.9	66

```
> head(airquality)
```

	ozone	solar.r	wind	temp	month	day
1	41	190	7.4	67	5	1
2	36	118	8.0	72	5	2
3	12	149	12.6	74	5	3
4	18	313	11.5	62	5	4
5	NA	NA	14.3	56	5	5
6	28	NA	14.9	66	5	6

,



Summarizing: Reshaping data



Getting and Cleaning Data