

Exercise: Reshaping

Day 2, Part B

```
> library(reshape2)
```

1. Load the Nigeria health metrics data set ('data/nigeria_healthmap.csv').

```
> main_dir <- "C:/Users/ngraetz/Documents/repos/r_training_penn/" # CHANGE TO YOUR LOCAL COPY L
> data <- read.csv(paste0(main_dir, "data/nigeria_healthmap.csv"), stringsAsFactors = F)
```

- a. Drop the location_id, geography, indicator_type, and units columns and keep only rows with data for 2000 and 2010.

```
> data[, c("location_id", "geography", "indicator_type", "units")] <- NULL
> data <- data[data$year == 2000 | data$year == 2010, ]
```

- b. Reshape this data long such that the data in the estimate, ci_lb, and ci_ub columns is combined into one column, e.g.:

```
> data <- melt(data, id.vars = c("location_name", "year", "indicator"))
> head(data)
```

	location_name	year	indicator	variable	value
1	Nigeria	2000	Antenatal care, 1 visit (ANC1)	(%) estimate	73
2	Nigeria	2010	Antenatal care, 1 visit (ANC1)	(%) estimate	68
3	Abia	2000	Antenatal care, 1 visit (ANC1)	(%) estimate	94
4	Abia	2010	Antenatal care, 1 visit (ANC1)	(%) estimate	93
5	Adamawa	2000	Antenatal care, 1 visit (ANC1)	(%) estimate	74
6	Adamawa	2010	Antenatal care, 1 visit (ANC1)	(%) estimate	75

- c. Reshape this data wide so that there are separate columns for the estimate, lower bound, and upper bound in each year, e.g.:

```
> data <- dcast(data, location_name + indicator ~ variable + year, value.var = "value")
> head(data)
```

	location_name	indicator	estimate_2000	estimate_2010	ci_lb_2000	ci_lb_2010	ci_ub_2000	ci_ub_2010
1	Abia	Antenatal care, 1 visit (ANC1) (%)	94	93	86	89	97	95
2	Abia	Antenatal care, 4 visits (ANC4) (%)	92	88	83	82	97	91
3	Abia	BCG immunization coverage (%)	84	93	64	86	93	97
4	Diphtheria-pertussis-tetanus, three doses (DPT3) immunization coverage (%)		66	80	42	66	84	90
5	Exclusive breastfeeding (EBF) (%)		7	15	2	8	18	28
6	Household ownership of at least one insecticide-treated net (ITN) (%)							

6	0	8	0	3	0	22
---	---	---	---	---	---	----

- d. Create a new column called `increase` which is `TRUE` if the estimate for 2010 is higher than for 2000 and `FALSE` otherwise.

```
> data$increase <- data$estimate_2010 > data$estimate_2000
```

- e. Remove all objects in your work space.

```
> rm(list = ls())
```

Bonus:

2. Still using the original data set, `nigeria_healthmap.csv`:

- a. Find and read the help docs for `reshape` from the `stats` library.

```
> ?reshape
```

- b. Use `reshape()` rather than `dcast()` to create the output from 1c above, starting with the original nigeria health metrics data.

```
> data <- read.csv(paste0(main_dir, "data/nigeria_healthmap.csv"), stringsAsFactors = F)
> data[, c("location_id", "geography", "indicator_type", "units")] <- NULL
> data <- data[data$year == 2000 | data$year == 2010, ]
> data <- reshape(data, direction = "wide", idvar = c("location_name", "indicator"), v.names =
+   "ci_lb", "ci_ub", timevar = "year", sep = "_")
> head(data)
```

	location_name	indicator	estimate_2000	ci_lb_2000
1	Nigeria Antenatal care, 1 visit (ANC1) (%)		73	71
15	Abia Antenatal care, 1 visit (ANC1) (%)		94	86
29	Adamawa Antenatal care, 1 visit (ANC1) (%)		74	63
43	Akwa Ibom Antenatal care, 1 visit (ANC1) (%)		79	71
57	Anambra Antenatal care, 1 visit (ANC1) (%)		96	94
71	Bauchi Antenatal care, 1 visit (ANC1) (%)		43	35

```

ci_ub_2000 estimate_2010 ci_lb_2010 ci_ub_2010
1          74           68          67          69
15         97           93          89          95
29         82           75          69          81
43         86           74          68          78
57         98           92          89          94
71         52           48          42          53

```