# Tidy Data with R

`gather()`

- `gather()` does the reverse of `spread()`.
- `gather()` collects a set of column names and places them into a single key column.
- It also collects the cells of those columns and places them into a single value column.
- You can use `gather()` to tidy table4.

# Tidy Data with R

```
table4  # cases

## Source: local data frame [3 x 3]
##
##       country   1999   2000
## 1 Afghanistan    745   2666
## 2      Brazil  37737  80488
## 3       China 212258 213766
```

# Tidy Data with R

- To use `gather()`, pass it the name of a data frame to reshape.
- Then pass `gather()` a character string to use for the name of the key column that it will make, as well as a character string to use as the name of the value column that it will make.
- Finally, specify which columns `gather()` should collapse into the key value pair (here with integer notation).

# Tidy Data with R

```
gather(table4, "year", "cases", 2:3)
```

```
## Source: local data frame [6 x 3]
##
##       country year  cases
## 1 Afghanistan 1999    745
## 2      Brazil 1999  37737
## 3       China 1999 212258
## 4 Afghanistan 2000   2666
## 5      Brazil 2000  80488
## 6       China 2000 2137664
```

# Tidy Data with R

- `gather()` returns a copy of the data frame with the specified columns removed.
- To this data frame, `gather()` has added two new columns: a key column that contains the former column names of the removed columns, and a value column that contains the former values of the removed columns.

# Tidy Data with R

- `gather()` repeats each of the former column names (as well as each of the original columns) to maintain each combination of values that appeared in the original data set.
- `gather()` uses the first string that you supplied as the name of the new key column, and it uses the second string as the name of the new value column.

# Tidy Data with R

- Just like `spread()`, gather maintains each of the relationships in the original data set.
- `gather()` also maintains each of the observations in the original data set, organizing them in a tidy fashion.

# Tidy Data with R

```
table5  # population

## Source: local data frame [3 x 3]
##
##       country       1999       2000
## 1 Afghanistan   19987071   20595360
## 2      Brazil  172006362  174504898
## 3       China 1272915272 1280428583
```

# Tidy Data with R

```
gather(table5, "year", "population", 2:3)
```

```
## Source: local data frame [6 x 3]
##
##       country year population
## 1 Afghanistan 1999   19987071
## 2      Brazil 1999  172006362
## 3       China 1999 1272915272
## 4 Afghanistan 2000   20595360
## 5      Brazil 2000  174504898
## 6       China 2000 1280428583
```

# Tidy Data with R

- Here we identified the columns to collapse with a series of integers. 2:3 describes the second and third columns of the data frame.
- You can identify the same columns with each of the commands below.
- You can also identify columns by name with the notation introduced by the select function in dplyr

```
gather(table5, "year", "population", c(2, 3)
gather(table5, "year", "population", -1)
```