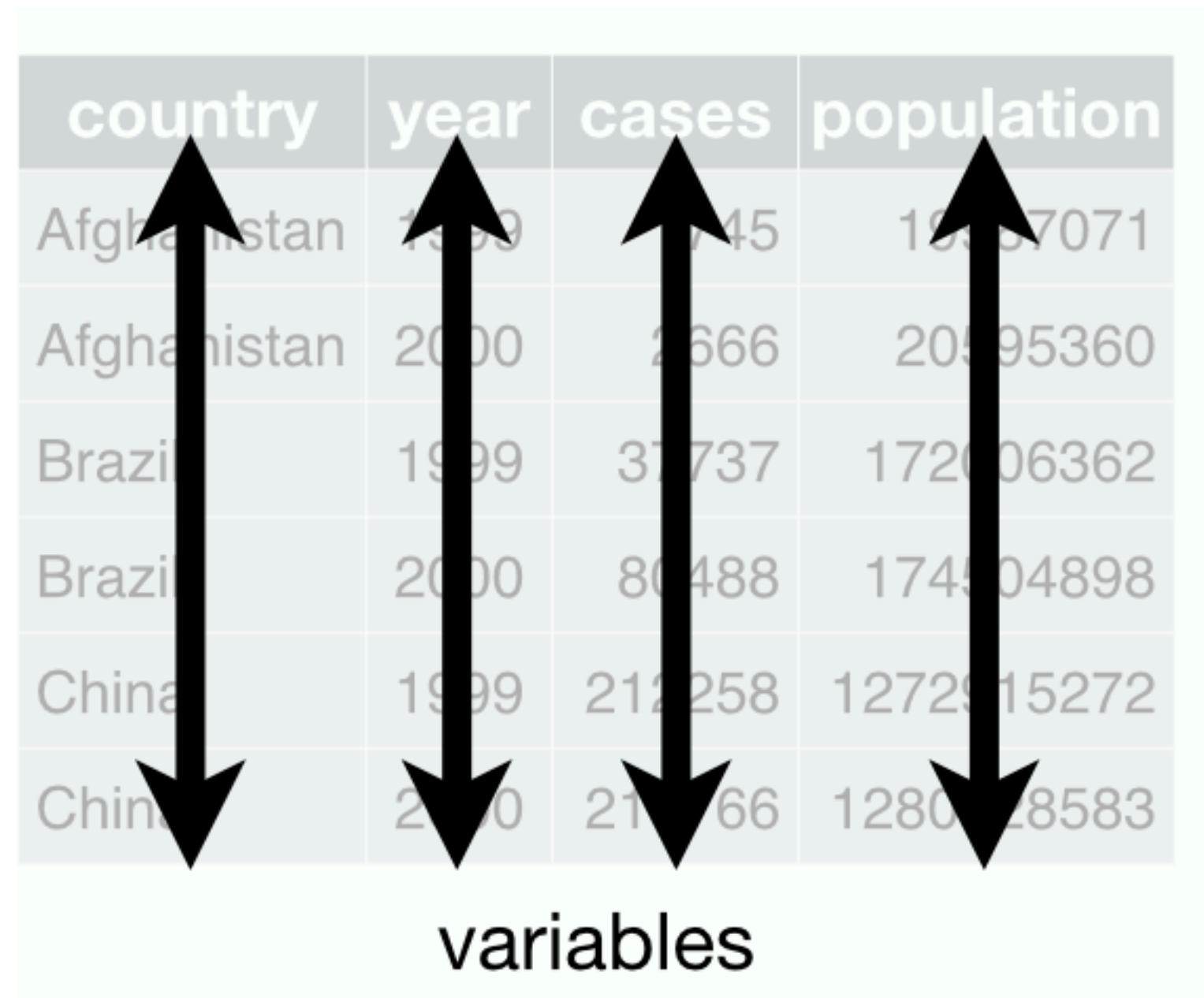


Tidy tables follow three rules:

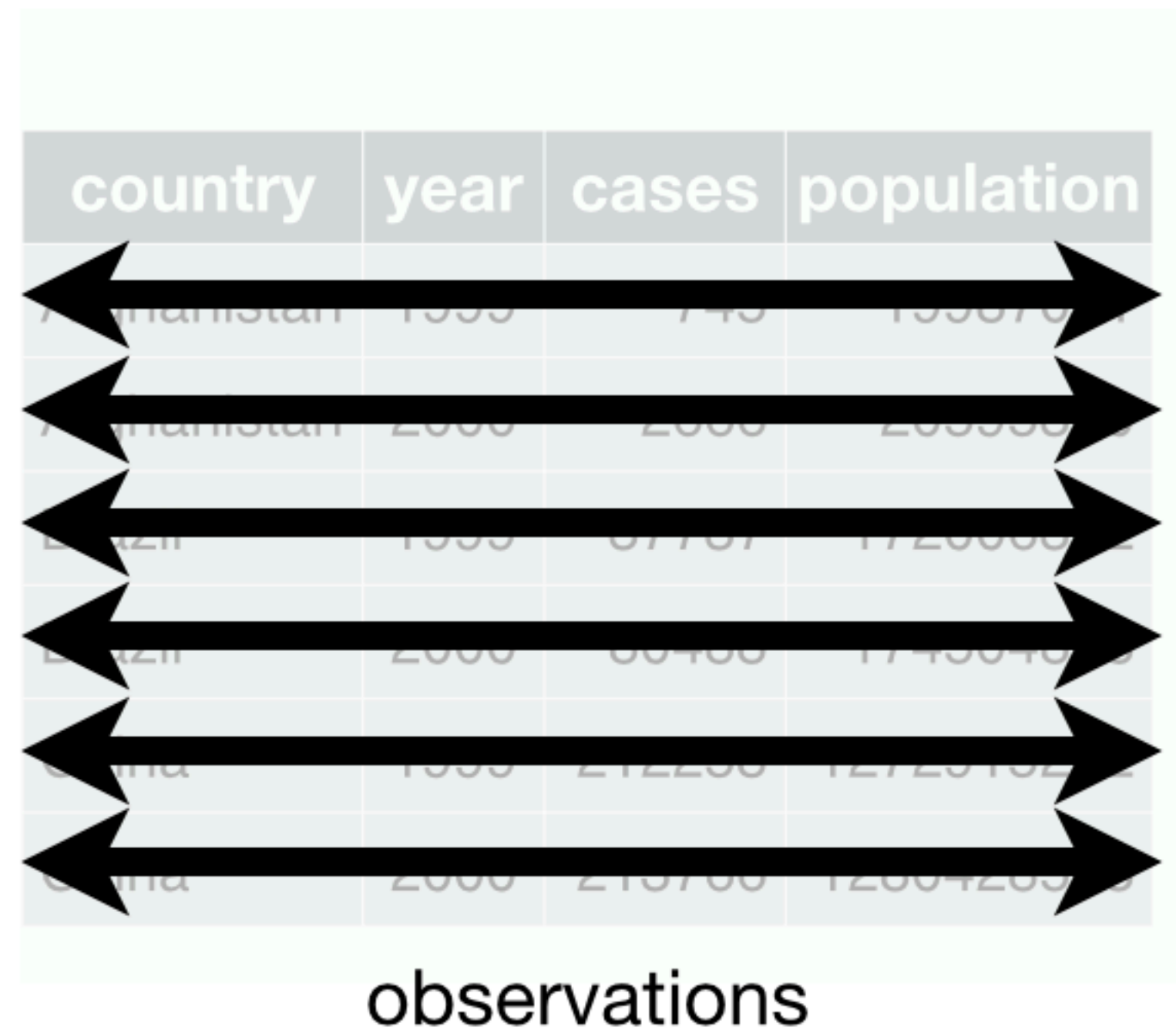
1. Each variable must have its own column
2. Each observation must have its own row
3. Each value must have its own cell



A diagram illustrating the first rule of tidy tables: each variable must have its own column. It shows a table with four columns: country, year, cases, and population. Four vertical double-headed arrows are placed over each column, indicating that each column represents a single variable across all observations.

country	year	cases	population
Afghanistan	1999	1745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

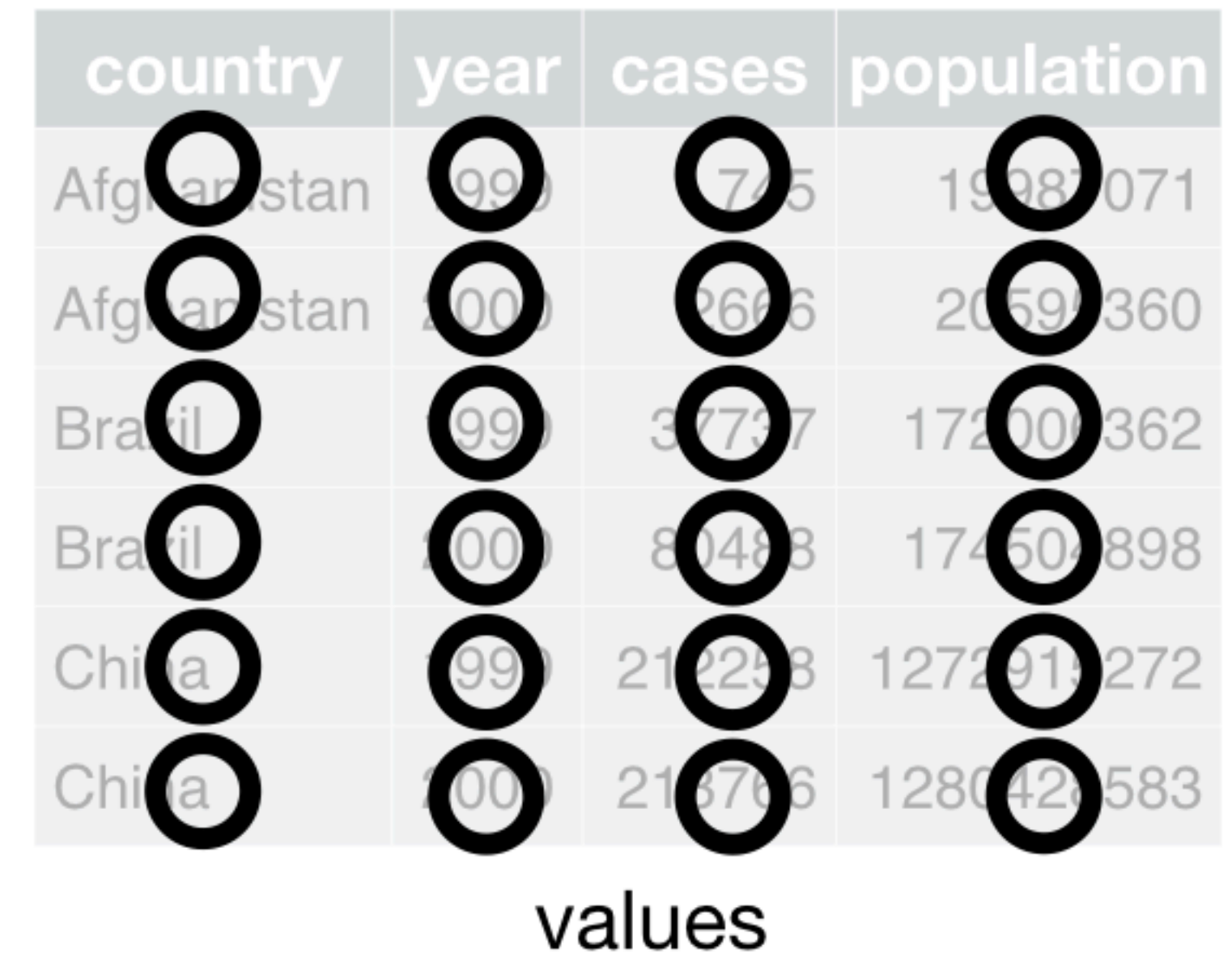
variables



A diagram illustrating the second rule of tidy tables: each observation must have its own row. It shows the same table as the first diagram. Five horizontal double-headed arrows are placed over each row, indicating that each row represents a single observation across all variables.

country	year	cases	population
Afghanistan	1999	1745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

observations



A diagram illustrating the third rule of tidy tables: each value must have its own cell. It shows the same table as the previous diagrams. Each individual value in the table is enclosed in a thick black circle, emphasizing that every data point is stored in a unique cell.

country	year	cases	population
Afghanistan	1999	1745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

values

Problem: Tables are usually not formatted for convenient analysis in R

Solution: “Tidy” up the table

“Wide”

```
table1
#> # A tibble: 6 x 4
#>   country      year  cases population
#>   <chr>      <int> <int>      <int>
#> 1 Afghanistan 1999     745  19987071
#> 2 Afghanistan 2000    2666  20595360
#> 3 Brazil       1999   37737  172006362
#> 4 Brazil       2000   80488  174504898
#> 5 China        1999  212258 1272915272
#> 6 China        2000  213766 1280428583
```

“Long”

```
table2
#> # A tibble: 12 x 4
#>   country      year type      count
#>   <chr>      <int> <chr>      <int>
#> 1 Afghanistan 1999 cases         745
#> 2 Afghanistan 1999 population 19987071
#> 3 Afghanistan 2000 cases         2666
#> 4 Afghanistan 2000 population 20595360
#> 5 Brazil       1999 cases         37737
#> 6 Brazil       1999 population 172006362
#> # ... with 6 more rows
```

“Gross”

```
table3
#> # A tibble: 6 x 3
#>   country      year rate
#> * <chr>      <int> <chr>
#> 1 Afghanistan 1999 745/19987071
#> 2 Afghanistan 2000 2666/20595360
#> 3 Brazil       1999 37737/172006362
#> 4 Brazil       2000 80488/174504898
#> 5 China        1999 212258/1272915272
#> 6 China        2000 213766/1280428583
```

“Split”

```
table4a # cases
#> # A tibble: 3 x 3
#>   country      `1999` `2000`
#> * <chr>      <int> <int>
#> 1 Afghanistan     745     2666
#> 2 Brazil          37737  80488
#> 3 China           212258 213766
table4b # population
#> # A tibble: 3 x 3
#>   country      `1999`      `2000`
#> * <chr>      <int>      <int>
#> 1 Afghanistan 19987071 20595360
#> 2 Brazil      172006362 174504898
#> 3 China       1272915272 1280428583
```