The tidyr Library for Data Transformation

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
table1
```

```
## # A tibble: 6 x 4
##   country      year  cases population
##   <chr>       <dbl>  <dbl>      <dbl>
## 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
```

```
table2
```

```
## # A tibble: 12 x 4
##    country      year type            count
##    <chr>       <dbl> <chr>           <dbl>
##  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
##  7 Brazil       2000 cases           80488
##  8 Brazil       2000 population  174504898
##  9 China        1999 cases          212258
## 10 China        1999 population 1272915272
## 11 China        2000 cases          213766
## 12 China        2000 population 1280428583
```

```
table3
```

```
## # A tibble: 6 x 3
##   country      year rate
##   <chr>       <dbl> <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
```

```
table4a
```

```
## # A tibble: 3 x 3
##   country     `1999` `2000`
##   <chr>        <dbl>  <dbl>
## 1 Afghanistan    745   2666
## 2 Brazil       37737  80488
## 3 China       212258 213766
```

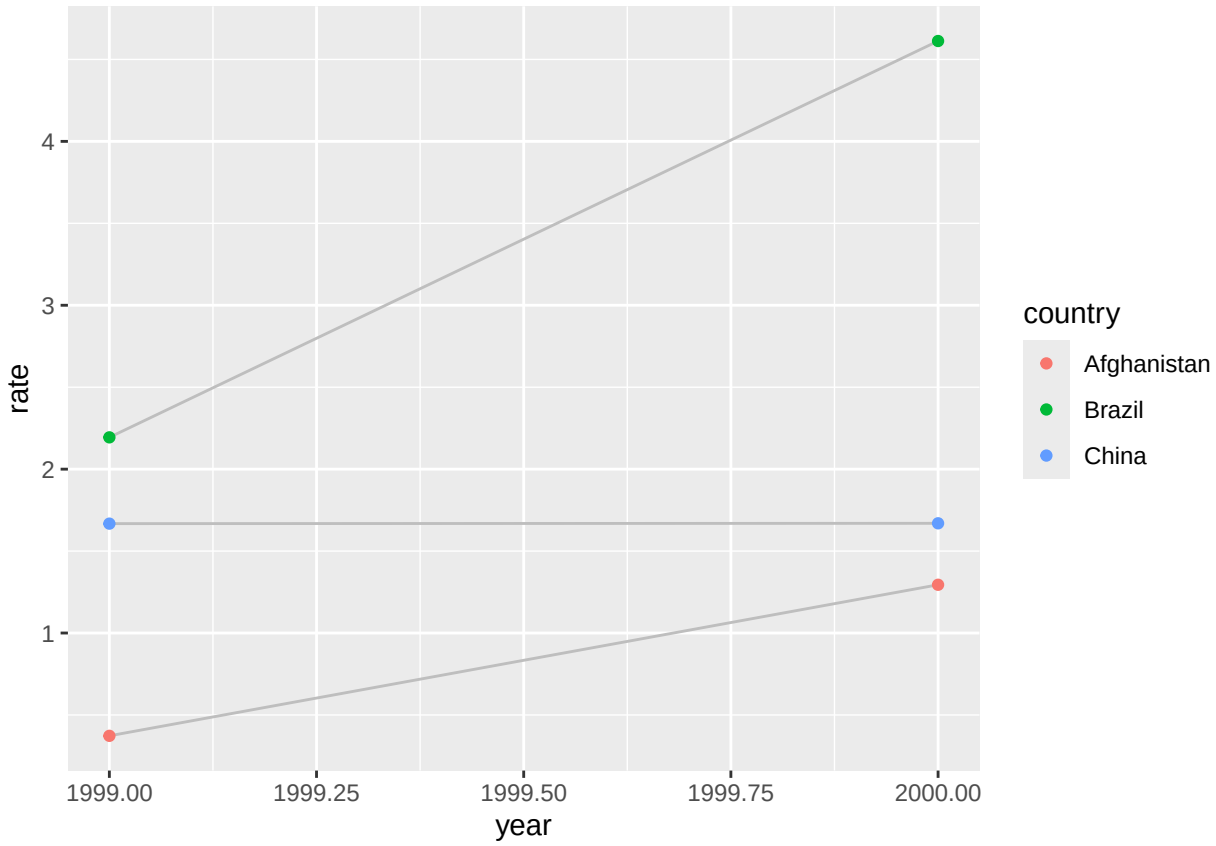
```
table1 %>% mutate(rate = (cases / population)*10000)
```

```
## # A tibble: 6 x 5
##   country      year  cases population  rate
##   <chr>       <dbl>  <dbl>      <dbl> <dbl>
## 1 Afghanistan  1999    745   19987071 0.373
## 2 Afghanistan  2000   2666   20595360 1.29
## 3 Brazil       1999  37737  172006362 2.19
## 4 Brazil       2000  80488  174504898 4.61
## 5 China        1999 212258 1272915272 1.67
## 6 China        2000 213766 1280428583 1.67
```

```
table1 %>% count(year, wt = cases)
```

```
## # A tibble: 2 x 2
##    year      n
##   <dbl>  <dbl>
## 1  1999 250740
## 2  2000 296920
```

```
ggplot(data = table1 %>% mutate(rate = (cases / population)*10000), aes(x = year, y = rate)) +
  geom_line(aes(group = country), color = 'gray')+
  geom_point(aes(color = country))
```




```
# Select year, country and population, and
# turn the long dataframe into a wide one
table1 %>% select(c(year, country, population))
```

```
## # A tibble: 6 x 3
##    year country     population
##   <dbl> <chr>            <dbl>
## 1  1999 Afghanistan   19987071
## 2  2000 Afghanistan   20595360
## 3  1999 Brazil       172006362
## 4  2000 Brazil       174504898
## 5  1999 China       1272915272
## 6  2000 China       1280428583
```

```
table1 %>% select(c(year, country, population)) %>% spread(key = country, value = population)
```

```
## # A tibble: 2 x 4
##    year Afghanistan    Brazil      China
##   <dbl>       <dbl>     <dbl>      <dbl>
## 1  1999    19987071 172006362 1272915272
## 2  2000    20595360 174504898 1280428583
```

```
table1 %>% mutate(rate = (cases/population)*10000) %>% select(c(year, country, rate)) %>% spread(key = c
```

```
## # A tibble: 2 x 4
##    year Afghanistan Brazil China
##   <dbl>       <dbl>  <dbl> <dbl>
## 1  1999       0.373   2.19  1.67
## 2  2000       1.29    4.61  1.67
```

```r
# Create a wide dataframe
table1.wide = table1 %>% filter(!is.na(cases)) %>% mutate(rate = (cases/population)*10000) %>% select(c
head(table1.wide)
```

```
## # A tibble: 2 x 4
##    year Afghanistan Brazil China
##   <dbl>       <dbl>  <dbl> <dbl>
## 1  1999       0.373   2.19  1.67
## 2  2000       1.29    4.61  1.67
```

```r
# Convert wide dataframe to a long one
head(table1.wide)
```

```
## # A tibble: 2 x 4
##    year Afghanistan Brazil China
##   <dbl>       <dbl>  <dbl> <dbl>
## 1  1999       0.373   2.19  1.67
## 2  2000       1.29    4.61  1.67
```

```r
table1.wide %>% gather(key = 'country', value = 'rate', -year)
```

```
## # A tibble: 6 x 3
##    year country      rate
##   <dbl> <chr>       <dbl>
## 1  1999 Afghanistan 0.373
## 2  2000 Afghanistan 1.29
## 3  1999 Brazil      2.19
## 4  2000 Brazil      4.61
## 5  1999 China       1.67
## 6  2000 China       1.67
```

```r
# Importing data
file = 'mtcars.csv'
carData = read.csv(file, header = TRUE, row.names = 1, stringsAsFactors = FALSE)
head(carData, 5)
```

```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
```

```r
# Exporting data
write_csv(table1.wide, file = 'tablewide.csv')
```