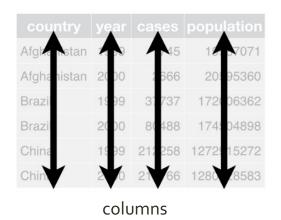
Data Science for Everyone – Data Wrangling – Tidy Data

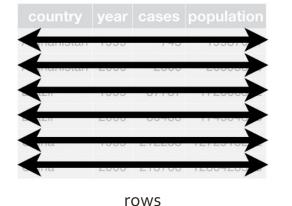
Dr. Ab Mosca (they/them)

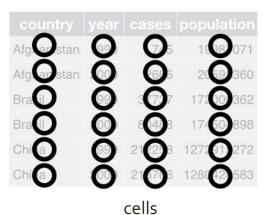
Plan for Today

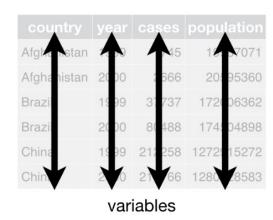
- Define Tidy data
- Clean messy data

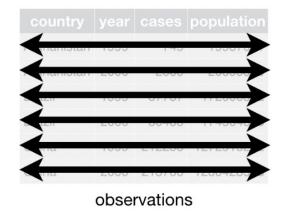
Reminder: Table Vocabulary

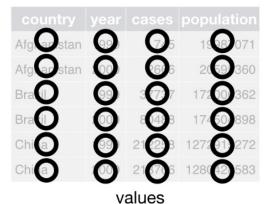






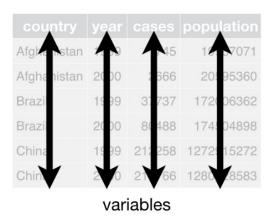


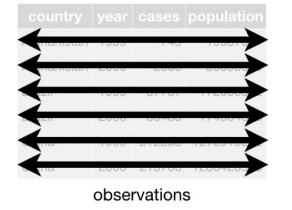


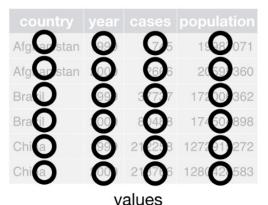


Reminder: Table Vocabulary

- When data is tidy, every column is a variable, every row is an observation, and every value has it's own cell
- We won't always get data in this format; sometimes data collectors record data in different ways







• When data is tidy, every column is a variable, every row is an observation, and every value has it's own cell

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

• When data is tidy, every column is a variable, every row is an observation, and every value has it's own cell

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

• When data is tidy, every column is a variable, every row is an observation, and every value has it's own cell

```
table3
#> # A tibble: 6 × 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
```

• When data is tidy, every column is a variable, every row is an observation, and every value has it's own cell

```
table3
#> # A tibble: 6 × 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
```

• When data is tidy, every column is a variable, every row is an observation, and every value has it's own cell

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

• When data is tidy, every column is a variable, every row is an observation, and every value has it's own cell

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

• When data is tidy, every column is a variable, every row is an observation, and every value has it's own cell

##		Republican	Independent	Democrat	the_date
##	1	16	47	85	2009-01-21
##	2	18	48	86	2009-01-26
##	3	17	45	84	2009-02-02
##	4	18	46	81	2009-02-09
##	5	17	46	82	2009-02-16
##	6	18	44	82	2009-02-23

• When data is tidy, every column is a variable, every row is an observation, and every value has it's own cell

	##		Republican	<u>Independent</u>	Democrat	the date
	##	1	16	47	85	2009-01-21
•	##	2	18	48	86	2009-01-26
	##	3	17	45	84	2009-02-02
	##	4	18	46	81	2009-02-09
	##	5	17	46	82	2009-02-16
	##	6	18	44	82	2009-02-23

• When data is tidy, every column is a variable, every row is an observation, and every value has it's own cell

• When data is tidy, every column is a variable, every row is an observation, and every value has it's own cell

```
## # A tibble: 3_x_3
                  `1999`
                          `2000`
##
     country
                   <int>
## * <chr>
                           <int>
   1 Afghanistan
                     745
                            2666
  2 Brazil
                           80488
                   37737
## 3 China
                  212258 213766
```

- Uniformity makes learning tools easier
- Most functions in the tidyverse are designed to work with tidy data (**tidy**verse, get it? ②)
 - Ex. ggplot, dplyr

- Uniformity makes learning tools easier
- Most functions in the tidyverse are designed to work with tidy data (**tidy**verse, get it? ①)
 - Ex. ggplot, dplyr

Can you calculate case rate with one line of code for table 3?

- Uniformity makes learning tools easier
- Most functions in the tidyverse are designed to work with tidy data (**tidy**verse, get it? ①)
 - Ex. ggplot, dplyr

What about table1?

```
table1
#> # A tibble: 6 × 4
     country
                year cases population
     <chr>
                 <dbl> <dbl>
                                   <dbl>
#> 1 Afghanistan
                         745
                  1999
                                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
```

```
table1 %>%

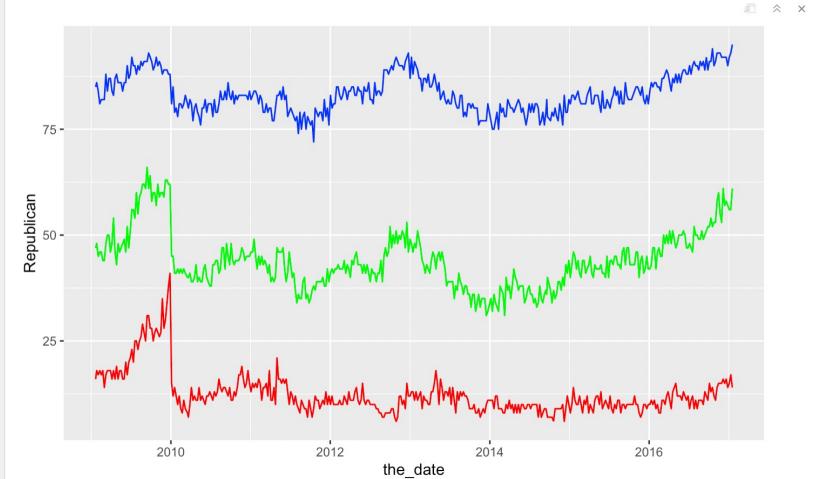
mutate(rate = cases / population * 10000)
```

What about table1?

```
table1
#> # A tibble: 6 × 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
               1999 212258 1272915272
#> 5 China
#> 6 China 2000 213766 1280428583
```

How would you plot a line chart for this data with a line for each party?

##		Republican	Independent	Democrat	the_date
##	1	16	47	85	2009-01-21
##	2	18	48	86	2009-01-26
##	3	17	45	84	2009-02-02
##	4	18	46	81	2009-02-09
##	5	17	46	82	2009-02-16
##	6	18	44	82	2009-02-23



How do we make data Tidy?

Tidy Data

How do we make data Tidy?

Re-write this table in a tidy format

##		Republican	Independent	Democrat	the_date
##	1	16	47	85	2009-01-21
##	2	18	48	86	2009-01-26
##	3	17	45	84	2009-02-02
##	4	18	46	81	2009-02-09
##	5	17	46	82	2009-02-16
##	6	18	44	82	2009-02-23

Tidy Da These values are approval ratings

How do we make data Tidy?

Re-write this table in a tidy format

```
Republican Independent Democrat
                                     the_date
##
## 1
                                85 2009-01-21
            16
                       47
## 2
            18
                       48
                                86 2009-01-26
## 3
            17
                       45
                                84 2009-02-02
## 4
            18
                       46
                                81 2009-02-09
                       46
## 5
            17
                                82 2009-02-16
## 6
            18
                       44
                                82 2009-02-23
## # A tibble: 4 x 3
##
     the_date
                              approval
                 party
     <date>
                 <chr>
                                  <int>
##
  1 2009-01-21 Republican
                                     16
  2 2009-01-21 Independent
                                     47
## 3 2009-01-21 Democrat
                                     85
## 4 2009-01-26 Republican
                                     18
```

How do we make data Tidy?

Re-write this table in a tidy format

How do we make data Tidy?

Re-write this table in a tidy format

Tidy Data

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

```
## # A tibble: 6 x 3
##
     country
                  year
                          cases
##
     <chr>
                          <int>
                  <chr>
   1 Afghanistan 1999
                            745
  2 Afghanistan 2000
                           2666
## 3 Brazil
                          37737
                  1999
## 4 Brazil
                  2000
                          80488
## 5 China
                  1999
                         212258
   6 China
                  2000
                         213766
```

What is similar in these two cases?

<int>

37737

year

<chr>

1999

2000

1999

2000

4 Brazil

5 China

6 China

745

<int>

2666

80488

cases

<int>

745

2666

37737

80488

212258

213766

```
## # A tibble: 3 x 3
     Republican Independent Democrat
##
                                      the date
                                                     ##
                                                           country
                                                                        `1999` `2000`
## 1
             16
                        47
                                    2009-01-21
                                                     ## * <chr>
                                 86 2009-01-26
##
             18
                        48
                                                         1 Afghanistan
## 3
                                 84 2009-02-02
                        45
             17
                                                     ## 2 Brazil
## 4
             18
                        46
                                 81 2009-02-09
## 5
             17
                        46
                                 82 2009-02-16
                                                     ## 3 China
                                                                        212258 213766
## 6
                        44
                                  82 2009-02-23
             18
                                                       ## # A tibble: 6 x 3
     A tibble: 4 \times 3
##
                                                       ##
                                                            country
     the_date
##
                                approval
                  party
                                                       ##
                                                             <chr>
                  <chr>
                                    <int>
##
     <date>
                                                       ## 1 Afghanistan 1999
     2009-01-21 Republican
                                       16
                                                       ## 2 Afghanistan 2000
   2 2009-01-21 Independent
                                       47
                                                          3 Brazil
   3 2009-01-21 Democrat
                                       85
##
```

18

4 2009-01-26 Republican

Sometimes column headers contain data

5 China

6 China

1999

2000

212258

213766

	##	Γ	Republican Ir	ndependent	Democrat	the date	##	#	A tibble: 3		
	##		16	47		2009-01-21	##		country	`1999`	`2000`
	##		18	48		2009-01-26	##	*	<chr></chr>	<int></int>	<int></int>
	##		17	45		2009-02-02	##	1	Afghanistan	745	2666
	##	4	18	46	81	2009-02-09	##	2	Brazil	37737	80488
	##	5	17	46	82	2009-02-16	##	3	China	212258	213766
	##	6	18	44	82	2009-02-23					
ı											
	##	#	A tibble:	4 x 3			#	##	# A tibble:	6 x 3	
	##		the_date	party	app	roval	#	##	country	year	cases
	##		<date></date>	<chr></chr>		<int></int>	#	##	<chr></chr>	<chr></chr>	<int></int>
	##	1	2009-01-21			16	#	##	1 Afghanista	n 1999	745
									2 Afghanista		2666
	##	2	2009-01-21	Independ	ent	47			-		
	##	3	2009-01-21	Democrat		85	#	##	3 Brazil	1999	37737
	##	4	2009-01-26	Republic	an	18	#	##	4 Brazil	2000	80488

Sometimes column headers contain data

• To correct this, we need to pivot our table

4 Brazil

6 China

5 China

80488

212258

213766

2000

1999

2000

					_					
##		Republican I	Independent [Democrat	the_date		#	A tibble: 3		```
##	1	16	47	85	2009-01-21	##		country	`1999`	`2000`
##		18	48		2009-01-26	##	*	<chr></chr>	<int></int>	<int></int>
##	3	17	45	84	2009-02-02	##	1	Afghanistan	745	2666
##	4	18	46	81	2009-02-09	##	2	Brazil	37737	80488
##	5	17	46	82	2009-02-16	##	3	China	212258	213766
##	6	18	44	82	2009-02-23					
##	#	A tibble:	4 x 3			#	##	# A tibble:	6 x 3	
##		the_date	party	app	roval	#	##	country	year	cases
##		- <date></date>	· <chr></chr>		<int></int>	#	##	<chr></chr>	<chr></chr>	<int></int>
##	1	2009-01-2	1 Republica	an	16	#	##	1 Afghanista	n 1999	745
			1 Independe		47	#	##	2 Afghanista	n 2000	2666
			1 Democrat	Tong 1860 To	85	#	##	3 Brazil	1999	37737

18

4 2009-01-26 Republican

- Sometimes column headers contain data
- To correct this, we need to pivot our table
- When we move column headers to a variable, we pivot_longer

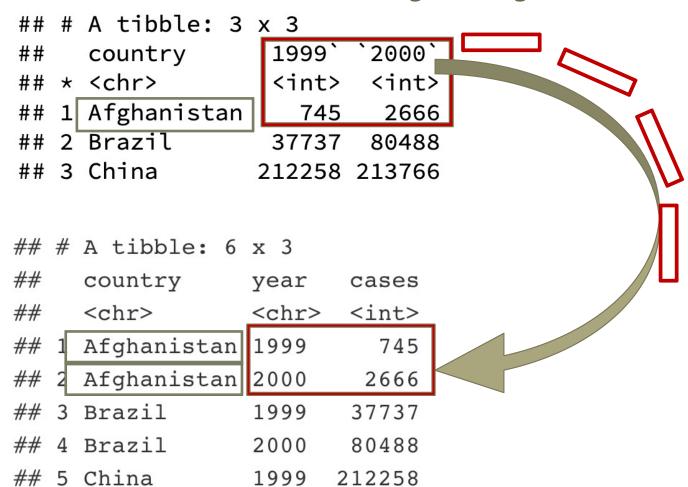
Τi	## ## ## ## ##	2 3 4	18 4 17 4 18 4	17 85 18 86 15 84 16 81	the_date	ate -21 -26 -02 -09	## ## ## ##	* 1 2	<chr> Afghanistan Brazil</chr>	`1999` <int> 745 37737</int>	`2000` <int> 2666 80488 213766</int>
	##	6	18	14 82	2009-02	-23					
	##	#	A tibble: 4 x 3				#	# :	# A tibble: 6	x 3	
	##		the_date party	арр	roval		#	#	country	year	cases
	##		<date> <chr></chr></date>		<int></int>		#	#	<chr></chr>	<chr></chr>	<int></int>
	##	1	2009-01-21 Republ	ican	16		#	#	1 Afghanistan	1999	745
	##	2	2009-01-21 Indepe	ndent	47		#	#	2 Afghanistan	2000	2666
	##	3	2009-01-21 Democr	at	85		#	#	3 Brazil	1999	37737
	##	4	2009-01-26 Republ	ican	18		#	#	4 Brazil	2000	80488

- •pivot_longer
 - Each observation gets its own row
 - Number of rows increases (table gets longer)

##		Republican	Independent	Democrat	the_u_
##	1	16	47	85	2009-01-21
##	2	18	48	86	2009-01-26
##	3	17	45	84	2009-02-02
##	4	18	46	81	2009-02-09
##	5	17	46	82	2009-02-16
##	6	18	44	82	2009-02-23

```
## # A tibble: 4 x 3
     the_date
                             approval
##
                party
##
     <date>
                <chr>
                                <int>
## 1 2009-01-21 Republican
                                   16
## 2 2009-01-21 Independent
                                   47
## 3 2009-01-21 Democrat
                                   85
## 4 2009-01-26 Republican
                                   18
```

- •pivot_longer
 - Each observation gets its own row
 - Number of rows increases (table gets longer)



How do we make data Tidy?

Re-write this table in a tidy format

```
## # A tibble: 6 x 4
##
                year type
    country
                                   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
```

How do we make data Tidy?

count

<int>

745

2666

37737

1999 population 172006362

Re-write this table in a tidy format

```
##
    country
                 year type
    <chr>
                <int> <chr>
  1 Afghanistan
                1999 cases
                1999 population 19987071
  2 Afghanistan
  3 Afghanistan
                 2000 cases
  4 Afghanistan
                 2000 population 20595360
  5 Brazil
                 1999 cases
```

A tibble: 6 x 4

6 Brazil

```
## # 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
```

How do we make data Tidy?

Re-write this table in a tidy format

```
# A tibble: 6 x 3
 name distance time
 <chr> <chr>
              <chr>
1 Ab 5k
               18:53
2 Ab 10k
               39:00
3 Kaden 5k
               19:37
4 Kaden 10k
               38:00
5 Kylee 5k
               17:50
6 Kylee 10k
               36:00
```

How do we make data Tidy?

Re-write this table in a tidy format

```
# A tibble: 6 x 3
 name distance time
 <chr> <chr>
               <chr>
                             # A tibble: 3 x 3
1 Ab
       5k
               18:53
                               name `5k` `10k`
2 Ab
       10k
                39:00
                                <chr> <chr> <chr>
3 Kaden 5k
               19:37
                             1 Ab 18:53 39:00
4 Kaden 10k
                38:00
                              2 Kaden 19:37 38:00
5 Kylee 5k
               17:50
                              3 Kylee 17:50 36:00
6 Kylee 10k
                36:00
```

What is similar in these two cases?

Tidy Data

```
## # A tibble: 6 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
```

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

```
# A tibble: 6 x 3
  name distance time
  <chr> <chr>
                 <chr>
1 Ab
        5k
                 18:53
2 Ab
        10k
                 39:00
3 Kaden 5k
                 19:37
                 38:00
4 Kaden 10k
5 Kylee 5k
                 17:50
6 Kylee 10k
                 36:00
```

Sometimes observations are split between rows

Tidy Data

```
## # A tibble: 6 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
```

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

```
# A tibble: 6 x 3
        distance time
  <chr> <chr>
                 <chr>
1 Ab
        5k
                 18:53
2 Ab
        10k
                 39:00
3 Kaden 5k
                 19:37
                 38:00
4 Kaden 10k
5 Kylee 5k
                 17:50
6 Kylee 10k
                 36:00
  # A tibble: 3 x 3
          `5k` `10k`
     <chr> <chr> <chr>
           18:53 39:00
  1 Ab
   2 Kaden 19:37 38:00
```

3 Kylee 17:50 36:00

- Sometimes observations are split between rows
- To correct this we need to pivot the table

```
## # A tibble: 6 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
```

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

```
# A tibble: 6 x 3
  name distance time
  <chr> <chr>
                 <chr>
1 Ab
        5k
                 18:53
2 Ab
        10k
                 39:00
3 Kaden 5k
                 19:37
                 38:00
4 Kaden 10k
5 Kylee 5k
                 17:50
6 Kylee 10k
                 36:00
  # A tibble: 3 x 3
          `5k` `10k`
     <chr> <chr> <chr>
```

18:53 39:00

2 Kaden 19:37 38:00

3 Kylee 17:50 36:00

1 Ab

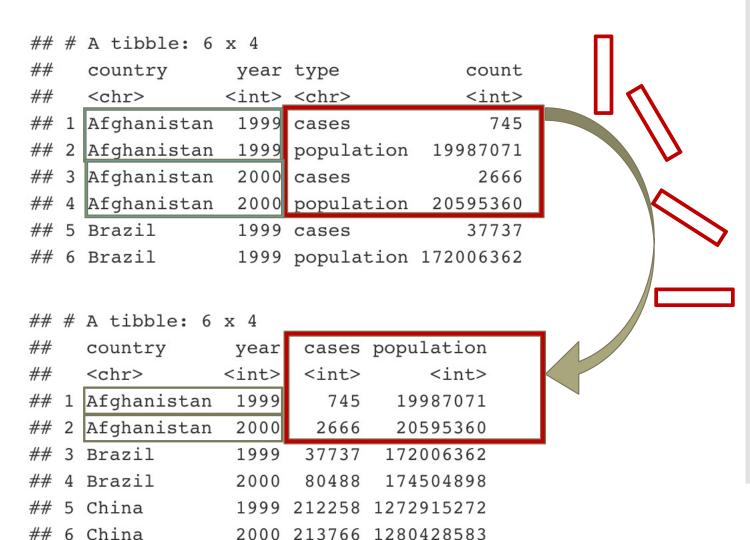
- Sometimes observations are split between rows
- To correct this we need to pivot the table
- When we move observations from multiple rows to one row, we pivot wider

```
## # A tibble: 6 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
```

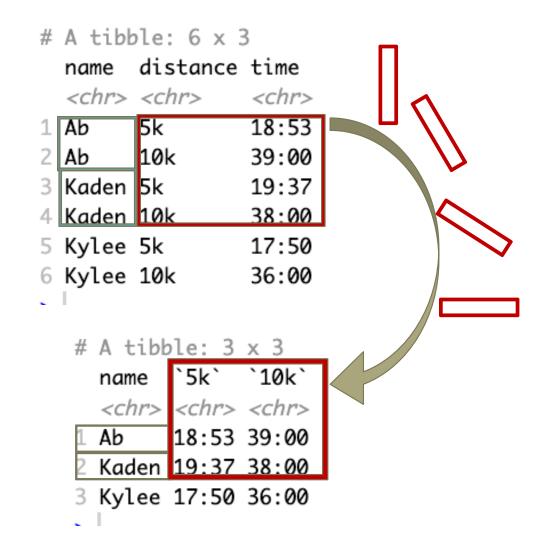
```
# A tibble: 6 x 4
##
     country
                        cases population
                  year
##
     <chr>
                 <int>
                         <int>
                                    <int>
   1 Afghanistan
                  1999
                           745
                 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

```
# A tibble: 6 x 3
       distance time
  <chr> <chr>
                 <chr>
1 Ab
        5k
                 18:53
2 Ab
        10k
                 39:00
3 Kaden 5k
                 19:37
                 38:00
4 Kaden 10k
5 Kylee 5k
                 17:50
                 36:00
6 Kylee 10k
```

- •pivot_wider
 - Each observation gets its own row
 - Number of rows decreases



- •pivot_wider
 - Each observation gets its own row
 - Number of rows decreases

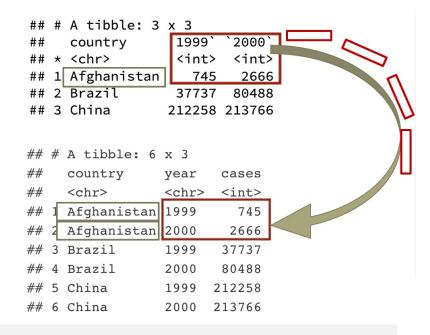




- •tidyr
 - R package that helps make data tidy
 - We will primarily use two functions:
 - •pivot longer()
 - •pivot wider()



•tidyr
•pivot_longer()
• wide → narrow





```
•tidyr
•pivot_longer()
• wide → narrow
```

```
## # A tibble: 3 x 3
     country
                  1999
                         2000
                  <int>
   * <chr>
                         <int>
   1 Afghanistan
                   745
                          2666
   2 Brazil
                  37737
                         80488
## 3 China
                 212258 213766
## # A tibble: 6 x 3
     country
                 year
                        cases
    <chr>
                 <chr>
                       <int>
## 1 Afghanistan 1999
                          745
  2 Afghanistan 2000
                         2666
## 3 Brazil
                 1999
                        37737
## 4 Brazil
                 2000
                        80488
## 5 China
                 1999
                       212258
## 6 China
                 2000 213766
```

- -country: pivot all columns except country
- names_to = "year": make a new column called year (into which we'll put the pivoted column names)
- values to = "cases": make another new column called cases (into which we'll put the pivoted values)



tidyrpivot_wider()narrow → wide

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



•tidyr
•pivot_wider()
•narrow → wide

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

- names_from = type: grab the values in the column called type (we'll pivot these values out to become the names of our new columns)
- values_from = count: grab the values in the column called count (we'll pivot these across their corresponding columns)

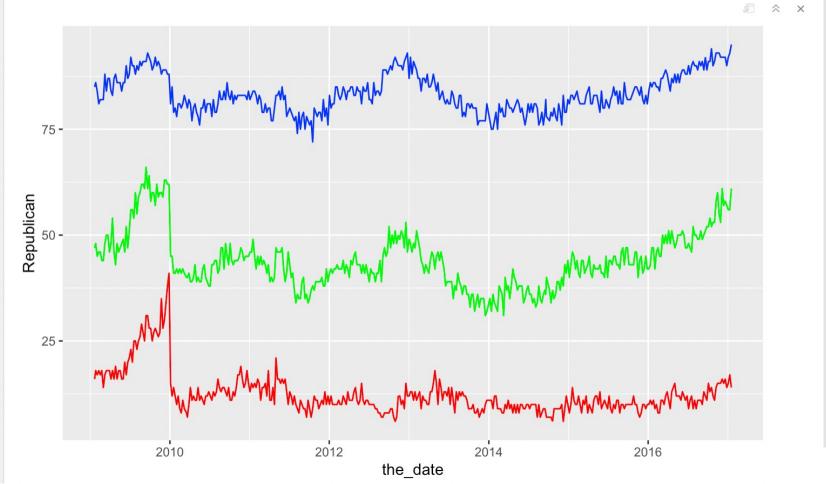
• We tend to use pivot longer () most often

Fill in the missing code below to pivot presapproval from wide form to long form.

```
Republican Independent Democrat the_date
##
                       85 2009-01-21
## 1
         16
                 47
## 2 18
             48
                       86 2009-01-26
## 3 17 45 84 2009-02-02
        18 46 81 2009-02-09
## 4
## 5 17
                46
                       82 2009-02-16
                44
## 6
         18
                       82 2009-02-23
```

```
Republican Independent Democrat
                              the_date
##
## 1
          16
                    47
                           85 2009-01-21
## 2
          18
                    48
                           86 2009-01-26
## 3
          17
                    45
                           84 2009-02-02
## 4
          18
                    46
                           81 2009-02-09
                    46
## 5
                           82 2009-02-16
          17
## 6
          18
                    44
                           82 2009-02-23
```

- -the date: pivot everything except the date
- names_to = "party": make a new column called party into which we'll put pivoted column names
- values_to = "approval": make a new column called approval into which we'll put pivoted values



```
40 → ## Easier plot
                                                                                                     € 
    ggplot(presapproval_tidy,
43
           aes(x = the\_date, y = approval, color = party)) +
44
      geom_line()
45 - ```
                                                                                                     75 -
                                                                                         party
      approval
oc
                                                                                             Democrat
                                                                                             Independent
                                                                                             Republican
        25 -
                                      2012
                     2010
                                                                        2016
                                                      2014
                                            the_date
```