

# Wrangle

April 23, 2020



Gina Reynolds  
Photo Credit: Joshua Coleman

# Set up

**The following exercises are designed to help you practice your data wrangling "vocabulary". In the exercises, you'll always start from the full gapminder data and wrangle to a target using functions from {dplyr} (one of the packages made available in the tidyverse).**

The following exercises are designed to help you practice your data wrangling "vocabulary". In the exercises, you'll always start from the full gapminder data and wrangle to a target using functions from {dplyr} (one of the packages made available in the tidyverse).

## Load the following packages:

```
library(gapminder) # The data package  
library(tidyverse) # Contains the data manipulation tools
```

The following exercises are designed to help you practice your data wrangling "vocabulary". In the exercises, you'll always start from the full gapminder data and wrangle to a target using functions from {dplyr} (one of the packages made available in the tidyverse).

## Load the following packages:

```
library(gapminder) # The data package
library(tidyverse) # Contains the data manipulation tools
```

Make sure to that the number of observations and columns in your result match the target outcome. Let's recall the full dimensions of the gapminder data:

```
# dim gives number of rows, then number of columns
dim(gapminder)
```

```
[1] 1704    6
```

The following exercises are designed to help you practice your data wrangling "vocabulary". In the exercises, you'll always start from the full gapminder data and wrangle to a target using functions from {dplyr} (one of the packages made available in the tidyverse).

## Load the following packages:

```
library(gapminder) # The data package
library(tidyverse) # Contains the data manipulation tools
```

Make sure to that the number of observations and columns in your result match the target outcome. Let's recall the full dimensions of the gapminder data:

```
# dim gives number of rows, then number of columns
dim(gapminder)
```

```
[1] 1704    6
```

Some other features to have in mind:

- {gapminder} only reports data every 5 years, from 1952 to 2007
- only {142} countries are included in the data

# You might use some of the following functions from the {dplyr} package:

function	action
filter()	<i>keep rows (if true)</i>
select()	<i>keep variables (or drop them -var)</i>
mutate()	<i>create a new variable</i>
case_when()	<i>is used for "recoding" variable, often used with mutate()</i>
rename()	<i>renaming variables</i>
arrange()	<i>order rows based on a variable</i>
slice()	<i>*keep or drop rows based on row number</i>

# Target #1

```
# A tibble: 142 x 5
  country      year      pop gdpPercap      gdp
  <fct>      <int>    <int>    <dbl>    <dbl>
1 Afghanistan  2002  25268405      727.  18363410424.
2 Albania      2002   3508512     4604.  16153932130.
3 Algeria      2002  31287142     5288.  165447670333.
4 Angola       2002  10866106     2773.   30134833901.
5 Argentina    2002  38331121     8798.  337223430800.
6 Australia    2002  19546792    30688.  599847158654.
7 Austria      2002   8148312    32418.  264148781752.
8 Bahrain      2002    656397    23404.   15362026094.
9 Bangladesh   2002  135656790     1136.  154159077921.
10 Belgium     2002  10311970    30486.  314369518653.
# ... with 132 more rows
```



# Walk through #1

```
# A tibble: 1,704 x 6
```

	country	continent	year	lifeExp	pop	gdpPercap
	<fct>	<fct>	<int>	<dbl>	<int>	<dbl>
1	Afghanistan	Asia	1952	28.8	8425333	779.
2	Afghanistan	Asia	1957	30.3	9240934	821.
3	Afghanistan	Asia	1962	32.0	10267083	853.
4	Afghanistan	Asia	1967	34.0	11537966	836.
5	Afghanistan	Asia	1972	36.1	13079460	740.
6	Afghanistan	Asia	1977	38.4	14880372	786.
7	Afghanistan	Asia	1982	39.9	12881816	978.
8	Afghanistan	Asia	1987	40.8	13867957	852.
9	Afghanistan	Asia	1992	41.7	16317921	649.
10	Afghanistan	Asia	1997	41.8	22227415	635.

```
# ... with 1,694 more rows
```

# Walk through #1

```
# A tibble: 142 x 6
  country      continent  year lifeExp      pop gdpPercap
  <fct>        <fct>    <int>   <dbl>    <int>    <dbl>
1 Afghanistan Asia      2002   42.1  25268405    727.
2 Albania     Europe    2002   75.7   3508512   4604.
3 Algeria     Africa    2002   71.0  31287142   5288.
4 Angola      Africa    2002   41.0  10866106   2773.
5 Argentina   Americas  2002   74.3  38331121   8798.
6 Australia   Oceania    2002   80.4  19546792  30688.
7 Austria     Europe    2002   79.0   8148312  32418.
8 Bahrain     Asia      2002   74.8    656397  23404.
9 Bangladesh  Asia      2002   62.0 135656790   1136.
10 Belgium    Europe    2002   78.3  10311970  30486.
# ... with 132 more rows
```

# Walk through #1

```
# A tibble: 142 x 4
  country      year      pop gdpPercap
  <fct>      <int>    <int>    <dbl>
1 Afghanistan  2002  25268405      727.
2 Albania      2002   3508512     4604.
3 Algeria      2002  31287142     5288.
4 Angola       2002  10866106     2773.
5 Argentina    2002  38331121     8798.
6 Australia    2002  19546792    30688.
7 Austria      2002   8148312    32418.
8 Bahrain      2002    656397    23404.
9 Bangladesh   2002  135656790     1136.
10 Belgium     2002  10311970    30486.
# ... with 132 more rows
```

# Walk through #1

```
# A tibble: 142 x 5
  country      year      pop gdpPercap      gdp
  <fct>      <int>    <int>    <dbl>    <dbl>
1 Afghanistan  2002  25268405    727.  18363410424.
2 Albania      2002   3508512   4604.  16153932130.
3 Algeria      2002  31287142   5288.  165447670333.
4 Angola       2002  10866106   2773.  30134833901.
5 Argentina    2002  38331121   8798.  337223430800.
6 Australia    2002  19546792  30688.  599847158654.
7 Austria      2002   8148312  32418.  264148781752.
8 Bahrain      2002    656397  23404.  15362026094.
9 Bangladesh   2002  135656790   1136.  154159077921.
10 Belgium     2002  10311970  30486.  314369518653.
# ... with 132 more rows
```

# Target #2

```
# A tibble: 10 x 4
```

	country	year	pop	pop_million
	<fct>	<int>	<int>	<dbl>
1	China	2007	1318683096	1319.
2	India	2007	1110396331	1110.
3	United States	2007	301139947	301.
4	Indonesia	2007	223547000	224.
5	Brazil	2007	190010647	190.
6	Pakistan	2007	169270617	169.
7	Bangladesh	2007	150448339	150.
8	Nigeria	2007	135031164	135.
9	Japan	2007	127467972	127.
10	Mexico	2007	108700891	109.

# Walk through #2

```
# A tibble: 1,704 x 6
  country      continent  year lifeExp      pop gdpPercap
  <fct>        <fct>    <int>  <dbl>    <int>    <dbl>
1 Afghanistan Asia      1952   28.8  8425333    779.
2 Afghanistan Asia      1957   30.3  9240934    821.
3 Afghanistan Asia      1962   32.0 10267083    853.
4 Afghanistan Asia      1967   34.0 11537966    836.
5 Afghanistan Asia      1972   36.1 13079460    740.
6 Afghanistan Asia      1977   38.4 14880372    786.
7 Afghanistan Asia      1982   39.9 12881816    978.
8 Afghanistan Asia      1987   40.8 13867957    852.
9 Afghanistan Asia      1992   41.7 16317921    649.
10 Afghanistan Asia      1997   41.8 22227415    635.
# ... with 1,694 more rows
```

# Walk through #2

```
# A tibble: 142 x 6
  country      continent  year lifeExp      pop gdpPercap
  <fct>        <fct>    <int>   <dbl>    <int>    <dbl>
1 Afghanistan Asia      2007   43.8  31889923    975.
2 Albania     Europe    2007   76.4   3600523   5937.
3 Algeria     Africa    2007   72.3  33333216   6223.
4 Angola      Africa    2007   42.7  12420476   4797.
5 Argentina   Americas  2007   75.3  40301927  12779.
6 Australia   Oceania    2007   81.2  20434176  34435.
7 Austria     Europe    2007   79.8   8199783   36126.
8 Bahrain     Asia      2007   75.6   708573    29796.
9 Bangladesh  Asia      2007   64.1 150448339   1391.
10 Belgium    Europe    2007   79.4  10392226  33693.
# ... with 132 more rows
```

## Walk through #2

```
# A tibble: 142 x 3
  country      year      pop
  <fct>      <int>    <int>
1 Afghanistan  2007  31889923
2 Albania      2007   3600523
3 Algeria      2007  33333216
4 Angola       2007  12420476
5 Argentina    2007  40301927
6 Australia    2007  20434176
7 Austria      2007   8199783
8 Bahrain      2007    708573
9 Bangladesh   2007 150448339
10 Belgium     2007  10392226
# ... with 132 more rows
```



## Walk through #2

```
# A tibble: 142 x 4
  country      year      pop pop_million
  <fct>      <int>    <int>      <dbl>
1 Afghanistan  2007  31889923    31.9
2 Albania      2007   3600523     3.60
3 Algeria      2007  33333216    33.3
4 Angola       2007  12420476    12.4
5 Argentina    2007  40301927    40.3
6 Australia    2007  20434176    20.4
7 Austria      2007   8199783     8.20
8 Bahrain      2007    708573     0.709
9 Bangladesh   2007 150448339   150.
10 Belgium     2007  10392226    10.4
# ... with 132 more rows
```

## Walk through #2

```
# A tibble: 10 x 4
```

	country	year	pop	pop_million
	<fct>	<int>	<int>	<dbl>
1	Bangladesh	2007	150448339	150.
2	Brazil	2007	190010647	190.
3	China	2007	1318683096	1319.
4	India	2007	1110396331	1110.
5	Indonesia	2007	223547000	224.
6	Japan	2007	127467972	127.
7	Mexico	2007	108700891	109.
8	Nigeria	2007	135031164	135.
9	Pakistan	2007	169270617	169.
10	United States	2007	301139947	301.

# Walk through #2

```
# A tibble: 10 x 4
```

	country	year	pop	pop_million
	<fct>	<int>	<int>	<dbl>
1	China	2007	1318683096	1319.
2	India	2007	1110396331	1110.
3	United States	2007	301139947	301.
4	Indonesia	2007	223547000	224.
5	Brazil	2007	190010647	190.
6	Pakistan	2007	169270617	169.
7	Bangladesh	2007	150448339	150.
8	Nigeria	2007	135031164	135.
9	Japan	2007	127467972	127.
10	Mexico	2007	108700891	109.

# Target #3

```
# A tibble: 13 x 3
```

	country	lifeExp	year
	<fct>	<dbl>	<int>
1	Japan	82.6	2007
2	Hong Kong, China	82.2	2007
3	Iceland	81.8	2007
4	Switzerland	81.7	2007
5	Australia	81.2	2007
6	Spain	80.9	2007
7	Sweden	80.9	2007
8	Israel	80.7	2007
9	France	80.7	2007
10	Canada	80.7	2007
11	Italy	80.5	2007
12	New Zealand	80.2	2007
13	Norway	80.2	2007

# Walk through #3

```
# A tibble: 1,704 x 6
  country      continent  year lifeExp      pop gdpPercap
  <fct>        <fct>    <int>   <dbl>    <int>    <dbl>
1 Afghanistan Asia      1952    28.8   8425333    779.
2 Afghanistan Asia      1957    30.3   9240934    821.
3 Afghanistan Asia      1962    32.0  10267083    853.
4 Afghanistan Asia      1967    34.0  11537966    836.
5 Afghanistan Asia      1972    36.1  13079460    740.
6 Afghanistan Asia      1977    38.4  14880372    786.
7 Afghanistan Asia      1982    39.9  12881816    978.
8 Afghanistan Asia      1987    40.8  13867957    852.
9 Afghanistan Asia      1992    41.7  16317921    649.
10 Afghanistan Asia      1997    41.8  22227415    635.
# ... with 1,694 more rows
```

# Walk through #3

```
# A tibble: 142 x 6
  country      continent  year lifeExp      pop gdpPercap
  <fct>        <fct>    <int>   <dbl>    <int>    <dbl>
1 Afghanistan Asia      2007   43.8  31889923    975.
2 Albania     Europe    2007   76.4   3600523   5937.
3 Algeria     Africa    2007   72.3  33333216   6223.
4 Angola      Africa    2007   42.7  12420476   4797.
5 Argentina   Americas  2007   75.3  40301927  12779.
6 Australia   Oceania    2007   81.2  20434176  34435.
7 Austria     Europe    2007   79.8   8199783   36126.
8 Bahrain     Asia      2007   75.6   708573    29796.
9 Bangladesh  Asia      2007   64.1 150448339   1391.
10 Belgium    Europe    2007   79.4  10392226  33693.
# ... with 132 more rows
```

# Walk through #3

```
# A tibble: 142 x 3
  country    lifeExp  year
  <fct>      <dbl> <int>
1 Afghanistan  43.8  2007
2 Albania      76.4  2007
3 Algeria      72.3  2007
4 Angola       42.7  2007
5 Argentina    75.3  2007
6 Australia    81.2  2007
7 Austria      79.8  2007
8 Bahrain      75.6  2007
9 Bangladesh   64.1  2007
10 Belgium     79.4  2007
# ... with 132 more rows
```

# Walk through #3

```
# A tibble: 13 x 3
```

	country	lifeExp	year
	<fct>	<dbl>	<int>
1	Australia	81.2	2007
2	Canada	80.7	2007
3	France	80.7	2007
4	Hong Kong, China	82.2	2007
5	Iceland	81.8	2007
6	Israel	80.7	2007
7	Italy	80.5	2007
8	Japan	82.6	2007
9	New Zealand	80.2	2007
10	Norway	80.2	2007
11	Spain	80.9	2007
12	Sweden	80.9	2007
13	Switzerland	81.7	2007



# Walk through #3

```
# A tibble: 13 x 3
  country      lifeExp year
  <fct>        <dbl> <int>
1 Japan        82.6  2007
2 Hong Kong, China 82.2  2007
3 Iceland      81.8  2007
4 Switzerland   81.7  2007
5 Australia     81.2  2007
6 Spain        80.9  2007
7 Sweden       80.9  2007
8 Israel       80.7  2007
9 France       80.7  2007
10 Canada      80.7  2007
11 Italy       80.5  2007
12 New Zealand  80.2  2007
13 Norway      80.2  2007
```

**Congrats! You're done.**