HW2: Explore Gapminder and use dplyr

Carleena Ortega

27/09/2019

Exercise 1

1.1 Filter

Use `filter()` to subset the `gapminder` data to three countries of your choice in the 1970's.

```
filtered <- gapminder %>%
  arrange(year) %>%
  filter(year > 1969, year < 1980, country == "Canada" | country == "Mexico"
    | country == "Brazil") %>%
  arrange(country)
```

1.2 Pipe Operator

Use the pipe operator %>% to select "country" and "gdpPercap" from your filtered dataset in 1.1.

```
filtered %>%
  select(country, gdpPercap)
```

```
## # A tibble: 6 x 2
##   country gdpPercap
##   <fct>       <dbl>
## 1 Brazil      4986.
## 2 Brazil      6660.
## 3 Canada     18971.
## 4 Canada     22091.
## 5 Mexico      6809.
## 6 Mexico      7675.
```

1.3 Drop in Life Expectancy @@

Filter gapminder to all entries that have experienced a drop in life expectancy. Be sure to include a new variable that's the increase in life expectancy in your tibble. Hint: you might find the lag() or diff() functions useful

```
gapminder %>%
  group_by(country) %>%
  arrange(country,year) %>%
  mutate(d_lifeExp=lifeExp-first(lifeExp)) %>%
  filter(d_lifeExp < 0)
```

```
## # A tibble: 10 x 7
## # Groups:   country [6]
##    country   continent  year lifeExp      pop gdpPercap d_lifeExp
##    <fct>     <fct>     <int>   <dbl>    <int>     <dbl>     <dbl>
##  1 Botswana  Africa     2002    46.6  1630347    11004.    -0.988
##  2 Cambodia  Asia       1977    31.2  6978607      525.    -8.20
##  3 Rwanda    Africa     1992    23.6  7290203      737.   -16.4
##  4 Rwanda    Africa     1997    36.1  7212583      590.    -3.91
##  5 Swaziland Africa     2007    39.6  1133066     4513.    -1.79
##  6 Zambia    Africa     1997    40.2  9417789     1071.    -1.80
##  7 Zambia    Africa     2002    39.2 10595811     1072.    -2.84
##  8 Zimbabwe  Africa     1997    46.8 11404948      792.    -1.64
##  9 Zimbabwe  Africa     2002    40.0 11926563      672.    -8.46
## 10 Zimbabwe  Africa     2007    43.5 12311143      470.    -4.96
```

1.4

Choose one of the following:

```
Filter gapminder so that it shows the max GDP per capita experienced by each country. Hint: you might f
```

OR

```
Filter gapminder to contain six rows: the rows with the three largest GDP per capita, and the rows with
```

```
gapminder %>%
  group_by(country) %>%
  arrange(country,gdpPercap) %>%
  mutate(M_gdpPercap=max(gdpPercap))
```

```
## # A tibble: 1,704 x 7
## # Groups:   country [142]
##    country     continent  year lifeExp      pop gdpPercap M_gdpPercap
##    <fct>       <fct>     <int>   <dbl>    <int>     <dbl>       <dbl>
##  1 Afghanistan Asia       1997    41.8 22227415      635.        978.
##  2 Afghanistan Asia       1992    41.7 16317921      649.        978.
##  3 Afghanistan Asia       2002    42.1 25268405      727.        978.
##  4 Afghanistan Asia       1972    36.1 13079460      740.        978.
##  5 Afghanistan Asia       1952    28.8  8425333      779.        978.
##  6 Afghanistan Asia       1977    38.4 14880372      786.        978.
##  7 Afghanistan Asia       1957    30.3  9240934      821.        978.
##  8 Afghanistan Asia       1967    34.0 11537966      836.        978.
##  9 Afghanistan Asia       1987    40.8 13867957      852.        978.
## 10 Afghanistan Asia       1962    32.0 10267083      853.        978.
## # ... with 1,694 more rows
```

1.5

Produce a scatterplot of Canada's life expectancy vs. GDP per capita using ggplot2, without defining a new variable. That is, after filtering the gapminder data set, pipe it directly into the ggplot() function. Ensure GDP per capita is on a log scale.

```
#gapminder %>%
 # filter(country == "Canada") %>%
 # ggplot(aes(gdpPercap,lifeExp)) +
  #geom_bar()
```

Exercise 2

Pick one categorical variable and one quantitative variable to explore. Answer the following questions in whichever way you think is appropriate, using dplyr:

```
What are possible values (or range, whichever is appropriate) of each variable?
What values are typical? What's the spread? What's the distribution? Etc., tailored to the variable at h
Feel free to use summary stats, tables, figures.
```

Exercise 3

Make two plots that have some value to them. That is, plots that someone might actually consider making for an analysis. Just don't make the same plots we made in class – feel free to use a data set from the datasets R package if you wish.

```
A scatterplot of two quantitative variables.
One other plot besides a scatterplot.
```

You don't have to use all the data in every plot! It's fine to filter down to one country or a small handful of countries.

Bonus

Bonus 1

For people who want to take things further.

Evaluate this code and describe the result. Presumably the analyst's intent was to get the data for Rwanda and Afghanistan. Did they succeed? Why or why not? If not, what is the correct way to do this?

Bonus 2

Present numerical tables in a more attractive form using knitr::kable() for small tibbles (say, up to 10 rows), and DT::datatable() for larger tibbles.