Snow Day work with Gapminder

March 21, 2018

## We did the first part before Spring Break:

Look at <https://www.youtube.com/watch?v=jbkSRLYSojo>

<https://www.gapminder.org>

Play with using this for different datasets and countries

<https://www.gapminder.org/data/>

Download data for average systolic blood pressure for women and the GDP per capita (both from World bank and the alternative GCP per capita) and add it to the gapminder tibble.

gapminder

## # 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

# Determine if the current working directory has a data folder. If not, make one.

if(!file.exists("./data")) {dir.create("./data")}

# these are the urls you should have saved. If not, go to the gapminder website and find the URLs we discussed in class for alt\_GDP, women’s blood pressure and GDP per capita.

fileUrls <- c("https://docs.google.com/spreadsheet/pub?key=tSUr\_yZVbM6a3AGJEq\_Z2Pw&output=xlsx",  
 "https://docs.google.com/spreadsheet/pub?key=0ArfEDsV3bBwCdHBzUVVSMDlTX1ZCUnNJQ3ZFdkFXVFE&output=xlsx",  
 "https://docs.google.com/spreadsheet/pub?key=0AkBd6lyS3EmpdHo5S0J6ekhVOF9QaVhod05QSGV4T3c&output=xlsx")

# save the variable names into a vector  
var\_names <- c("alt\_GDP","blood\_press","GDP")

# Fill in the function get\_clean to download and read in the excel file from the url provided and then put the data in a column with the variable name specified in var\_name

get\_clean <- function(url\_in, var\_name){  
# you need to fill in this part

}  
# testing line – try it out with the first entry in fileUrls and var\_names to get:

out1 <- get\_clean(fileUrls[1],var\_names[1])  
head(out1)

## # A tibble: 6 x 3  
## country year alt\_GDP  
## <chr> <int> <dbl>  
## 1 Afghanistan 1970 1731.  
## 2 Afghanistan 1971 1748.  
## 3 Afghanistan 1972 2120.  
## 4 Afghanistan 1973 2119.  
## 5 Afghanistan 1974 2148.  
## 6 Afghanistan 1975 2263.

# Now use map2 to apply this function to the entire list of urls and names.

all\_data <- map2(fileUrls, var\_names, get\_clean)  
head(all\_data)

## [[1]]  
## # A tibble: 7,334 x 3  
## country year alt\_GDP  
## <chr> <int> <dbl>  
## 1 Afghanistan 1970 1731.  
## 2 Afghanistan 1971 1748.  
## 3 Afghanistan 1972 2120.  
## 4 Afghanistan 1973 2119.  
## 5 Afghanistan 1974 2148.  
## 6 Afghanistan 1975 2263.  
## 7 Afghanistan 1976 2270.  
## 8 Afghanistan 1977 2121.  
## 9 Afghanistan 1978 2205.  
## 10 Afghanistan 1979 2121.  
## # ... with 7,324 more rows  
##   
## [[2]]  
## # A tibble: 5,771 x 3  
## country year blood\_press  
## <chr> <int> <dbl>  
## 1 Afghanistan 1980 122.  
## 2 Afghanistan 1981 122.  
## 3 Afghanistan 1982 122.  
## 4 Afghanistan 1983 123.  
## 5 Afghanistan 1984 123.  
## 6 Afghanistan 1985 123.  
## 7 Afghanistan 1986 123.  
## 8 Afghanistan 1987 123.  
## 9 Afghanistan 1988 124.  
## 10 Afghanistan 1989 124.  
## # ... with 5,761 more rows  
##   
## [[3]]  
## # A tibble: 7,988 x 3  
## country year GDP  
## <chr> <int> <dbl>  
## 1 Albania 1980 1061.  
## 2 Albania 1981 1100.  
## 3 Albania 1982 1111.  
## 4 Albania 1983 1101.  
## 5 Albania 1984 1065.  
## 6 Albania 1985 1060.  
## 7 Albania 1986 1092.  
## 8 Albania 1987 1054.  
## 9 Albania 1988 1014.  
## 10 Albania 1989 1092.  
## # ... with 7,978 more rows

Now we will join the data we got to the existing gapminder dataset. You need to fill this code in to produce the desired result.

new\_gapminder <- gapminder  
for (i in seq\_along(all\_data)){  
 # add continent data to the data we downloaded

# you need to fill in this part  
 # now put the datasets together

# you need to fill in this part  
 }

new\_gapminder %>%  
 arrange(country, year)

## # A tibble: 1,704 x 9  
## country continent year lifeExp pop gdpPercap GDP blood\_press  
## <chr> <fct> <int> <dbl> <int> <dbl> <dbl> <dbl>  
## 1 Afghanistan Asia 1952 28.8 8.43e6 779. NA NA   
## 2 Afghanistan Asia 1957 30.3 9.24e6 821. NA NA   
## 3 Afghanistan Asia 1962 32.0 1.03e7 853. NA NA   
## 4 Afghanistan Asia 1967 34.0 1.15e7 836. NA NA   
## 5 Afghanistan Asia 1972 36.1 1.31e7 740. NA NA   
## 6 Afghanistan Asia 1977 38.4 1.49e7 786. NA NA   
## 7 Afghanistan Asia 1982 39.9 1.29e7 978. NA 122.  
## 8 Afghanistan Asia 1987 40.8 1.39e7 852. NA 123.  
## 9 Afghanistan Asia 1992 41.7 1.63e7 649. NA 124.  
## 10 Afghanistan Asia 1997 41.8 2.22e7 635. NA 124.  
## # ... with 1,694 more rows, and 1 more variable: alt\_GDP <dbl>