

Assignment 10

This homework will help you to practice and self assess your knowledge and ability around the course concepts. Use the answer key to assess your work.

1. Load the **tidyverse** & **exampledata** packages.
2. Download the life expectancy data set from <https://data.worldbank.org/indicator/SP.DYN.LE00.IN>. This video shows how to download the data set: <https://youtu.be/oYnQgZUai5k>.
3. Extract the .csv data files from the zip folder. Open the 3 files and have a look.

Notice that there is meta data above the headers in ‘API_SP.DYN.LE00.IN_DS2_en_csv_v2.csv’? You’ll want to use this information to read in the data properly in the next question.

4. Read the following files into R and assign to the object name **data** and **meta** respectively:
 - API_SP.DYN.LE00.IN_DS2_en_csv_v2.csv
 - Metadata_Country_API_SP.DYN.LE00.IN_DS2_en_csv_v2.csv

Note there will be parse errors; this is likely ok (we investigate this more in question 6)

5. Join the two tables together so that every row of **data** has accompanying meta data from **meta**. Assign this to an object called **life**.
6. From the **life** table, drop the columns **Indicator Name** & **Indicator Code** as well as any other columns that were created by a bad read-in of the data (for me the columns X6 & X63 were created because of a mal-formed row(s) in the 2 .csv files). It is good practice to examine malformed columns and look at the values to make sure there isn’t a serious read problem. In this case the values are all NA indicating there isn’t a larger issue.
7. Use the code you wrote in question 6 and continue the piping/chaining (**%>%**) to make this data set tidy. Call the **gathered** columns **Year** & **Age**. Hint: Ask what columns are actually a single variable?
8. Use the code you wrote in question 7 and continue the piping/chaining (**%>%**) to make the **Year** & **Age** variables integer and numeric rather than character. Assign this to a variable/object called **clean_life**.
9. Use **clean_life** and make a plot to show the change of life expectancy over time. Facet by region. You may want to set the **alpha** level of the geom lower to help with overplotting. Select the geom that is most useful for showing change over time.
10. Add a smoothed trend line to each facet via **geom_smooth(color = 'darkred')**. Note that you likely used **group = `Country Name`** in a global **aes** rather than inside of an **aes** in **geom_line** in question 9. If so, you will want to move the **group = `Country Name`** inside of **geom_line**.
11. There appears to be a story in the ‘Sub-Saharan Africa’ region. What might be the reason for this?
12. Using the code from question ten look at the trends for **IncomeGroup** instead of **Region**. Does this tell a another story?