**Section 2\_Data Wrangling Challenge**

Make observations about what your tables/figures show and about the process.

We’ll be using United Nations data about countries, life expectancy and GDP. This can be found in the github site under section 2.

If you want to do something comparable but different, i.e. swap one quantitative variable for another, be my guest! If you are feeling inspired and curious, then we’re doing this right. Go for it.

Relax about the following things:

* Tidying/reshaping is NOT your assignment. Many of your tables will be awkwardly shaped in the report. That’s OK.
* Table beauty is not a big deal. Simply printing to “screen” is fine
* For all things, graphical and tabular, if you’re dissatisfied with a result, discuss the problem, what you’ve tried and move on.

**The Challenge, if you are willing to accept, is the following …**

Get the maximum and minimum of GDP per capita for all continents.

Look at the spread of GDP per capita within the continents.

Compute a trimmed mean of life expectancy for different years. Or a weighted mean, weighting by population. Just try something other than the plain vanilla mean.

How is life expectancy changing over time on different continents?

Report the absolute and/or relative abundance of countries with low life expectancy over time by continent: Compute some measure of worldwide life expectancy – you decide – a mean or median or some other quantile or perhaps your current age. Then determine how many countries on each continent have a life expectancy less than this benchmark, for each year.

Find countries with interesting stories. This is open-ended and, therefore, hard. Promising but unsuccessful attempts are encouraged. This will generate interesting questions to follow up on in class.

Make up your own! Between the dplyr coverage in class and the list above, I think you get the idea.

Regex’ing for the win:

How many countries have the letters ‘i’ and ‘a in their names?

How many countries ***end*** with ‘ia’?

Which countries have punctuation in their country name? Hint – try “[[:punc:]]”

Which countries are multi-word countries? Example: “Congo, Dem. Rep”

For those countries with a space in them, like “South Africa,” replace the space with an underscore “\_”