ADS1 exercises – Advanced pandas

1. (a) Attached is the file `UK_cities.txt`. Inspect it in a text editor. You'll see that the columns are separated by a variable number of blanks. Read the file with `pd.read_csv` using the keyword argument `sep='\s+'`. `\s+` is a so-called regular expression standing for one or more blanks. You'll notice that entries including spaces are protected by quotes.

   (b) Inspect basic statistical properties using the `describe` method. Also calculate and compare the median and average of the Population. Calculate Pearson's and Kendall's correlations.

   (c) Create a new column containing the population in thousands.

   (d) Use `groupby` to sum up the populations of the different Nation/Region.

   (e) Create new dataframes with the data for old cities (city status granted before 1888) and new cities. Sum up the populations of the two groups.

2. (a) Attached the file `energy_per_head.xls`. Read it and have a look. It is Worldbank data organised in rows. Transpose it. Write the first line containing country names into the header and remove the first two lines (not containing numerical data).

   (b) Remove NaN's for the Brazil and China entries. Plot them as function of the year with labels and legend. Note that years are in the index column. You can get these values as `dataframe.index`. The x-labels will look cluttered. This is because the index values are of type str. This can be changed by the operation

```
dataframe.index = dataframe.index.astype(int)
```