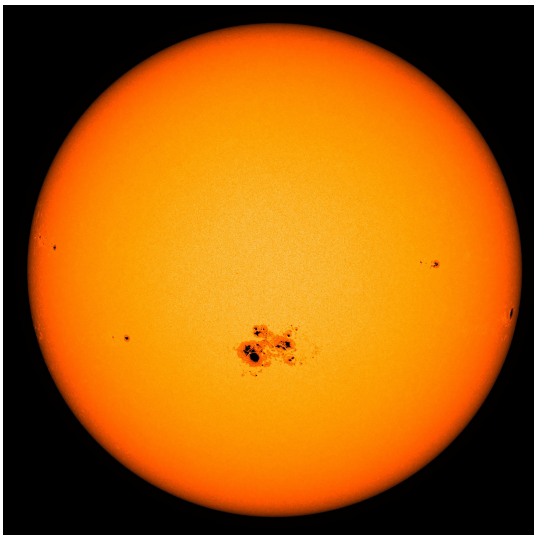
ADS1 exercises – Pandas 3

1. The files `gdp_region.csv` and `gdp_region_tr.csv` contain information on GDP (inflation corrected) of regions of the world. The first version has columns containing data on a region-by-region basis. The latter (transposed) has columns containing data for a year.

   (a) Read in `gdp_region_tr.csv` and create pie charts of regional GDP in 1980 and 2020. Use the standard pyplot function `plt.pie()` for this. (The pandas dataframe method is more complicated to use in this case). Label the pie charts with the regions and show the years in the titles.

   (b) Now, read in `gdp_region.csv`. Calculate a new column summing up the world wide GDP. Use `.iloc[]` for this. Plot the world GDP and GDP of one or more regions of your choice over time. This time the dataframe method is more convenient.

2. Sunspots appear in groups. A sunspot index is calculated by counting the number of sunspots and adding ten for every group visible from Earth. The sun displays a sunspot cycle of on average 11 years. The Sun has a rotation period of about a month. The daily sunspot index has been recorded since 1874.



   The file `sunspots_catala.csv` contains the sunspot indeces recorded for the northern and the southern hemisphere. Read the file into a pandas dataframe. Instead of commas a variable number of spaces is used as separator. Use the `sep='\s+'` keyword argument. `\s+` is a so-called regular expression which stands for a variable number of spaces.

   The date column contains strings. Before you do anything else convert them into the `pandas datetime` type.

   (a) Create a new column summing up the northern and southern sunspot numbers. Create new columns with the rolling mean over 365 days for all three indices. Plot the moving averages using the dataframe `.plot()` method.

(b) Extract data for the year 2014 into a new dataframe. Apply a rolling mean smoothing over five days and plot.

Note: When applying the rolling mean you will get an error message. The reason is that both dataframes are still linked. The extracted dataframe just a named part of the full dataframe. To unlink them use the pandas `.copy()` method, e.g.

```
df = df.copy()
```