Pandas Hands-on

February 14, 2017

1 Let's play with Pandas!

1.1 Simple Analysis of the Cantareira reservoir hydroclimatic data

1. Import pandas, numpy and matplotlib.pyplot
2. Create a dataframe from the file "DataCantareira.csv".
3. Print only the data from 2005 to 2010
4. Clean your file by droping the NaN
5. What is the minimum daily accumulated rainfall? Does it seems a resonable observed value? Remove every events (row) with a negative rainfall from the dataframe. hint: use boolean indexing
6. When the cantareira reservoir experienced its lowest level? Hint:
   - – select the volume column
   - – Find the lowest value
   - – Perform boolean indexing and select the corresponding index
7. Make a bar plot of the average annual precipitation.
8. Create a new column with the value of the reservoir volume in m^3. Knowing that the maximum capacity of the reservoir is approximatively 1000 billions of liter.

   Hint:
   - – Conversion: volume (m^3) = (volume (%)/100) * 10^9
   - – Create a function which return the volume in m^3.
   - – Use the apply method
9. In average, which month of the year the reservoir have the highest volume. Similarly, which month of the year there is the highest accumulation of rainfall? (Is there a delay?) Hint:
   - use groupby

```
In [ ]:
```