How we are going to handle data.

“Water Quality”

1. pH ,Turbidity, Temp, Oxygen from 2010 - 2018
2. Download yearly data, and we will sum those later

( it will be collected at google drive folder)

Ex. ) pH\_2010\_2011.csv. / temp\_2011\_2012.csv..

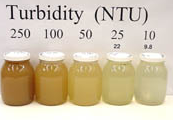
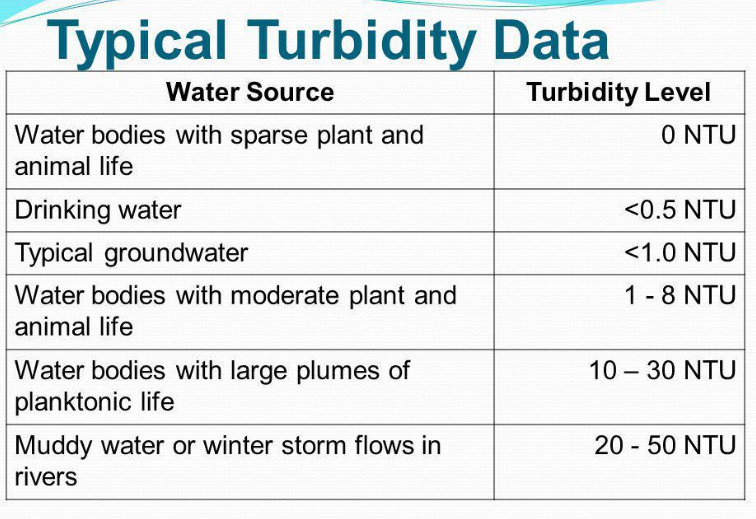
3. We cluster the data by Longitude and Latitude → we will round up the Longitude and Latitude data to the first decimal,

4. Find out the standard level of each parameters (pH, temp, turbidity, oxygen)

ex) Turbidity NTU Nephelometric Turbidity Units (**NTU**) this is Unit

And

|  |  |
| --- | --- |
| NTU |  |
| < 1 | Ideally drinkable water |
| < 5 | Drinkable |
|  |  |

Our Plan.on WQ.

Download WQ data by May 6 Sun.

May 7 -11 / Cluster data of first 2 years ( 2010 - 2012)

May 12-13 / see the clustered data. Do some analysis….

Question to ask TA.)

>>> how to handle the outliers. Like 8 sample pH data of 7~8 and 1 pH data of 10.5 from one location in a month