STAT 480 HW06

Donghan Liu, Donghan2

I firstly use pig to join the data together and store the data into "joindata.txt", and then, I used the joindata as basis to process the command line (pig –x local), and store the results in different files.

Some code are refered from the course notes

#Exercise 1

usaf	wban	year	temperature	usaf	wban	station
28060	99999	1911	-122	28060	99999	UNKNOWN87
28060	99999	1911	-94	28060	99999	UNKNOWN87
28060	99999	1911	-189	28060	99999	UNKNOWN87
28060	99999	1911	-83	28060	99999	UNKNOWN87
28060	99999	1911	-128	28060	99999	UNKNOWN87

Based on the above table, which is the joint data with Data1910s.txt and StationCodes.txt, the first column and fifth column represent the same meaning, usaf identifier. Plus, the second column and sixth column mean the same thing, which is wban identifier. The third column is years and fourth means observed temperature (Celsius unit *10) and the last column is that station that corresponding to each year and temperature.

#Exercise 2

Station	Max	Min	Mean Temperature	Trusted Obs

```
OULU                 317 -361  21.387204874333587   6565
KUOPIO               322 -372  30.47103775026746    6543
VYBORG               333 -356  43.11222780569514    6567
GRIMSEY              378 -300  17.387429766104795   7653
KAJAANI              283 -367  19.472436604189635   1814
KUUSAMO              300 -411  -6.049479166666667   6528
RUSSARO              294 -233  56.46740785866585    6566
JYVASKYLA            317 -356  30.927328244274808   3275
SODANKYLA            283 -478  -12.093059936908517  3170
SORTAVALA            322 -361  31.70109689213894    3282
UNKNOWN87            300 -406  -15.876810489403873  6559
VAASA AIRPORT        300 -283  34.290509612450414   3277
HELSINKI-MALMI       272 -300  45.529841656516446   3284
TAMPERE/PIRKKALA     328 -300  43.53942161339422    6570
MARIEHAMN/ALAND ISL  278 -239  52.979274611398964   3281
```

The observed temperature is in Celsius unit. The stations and their corresponding maximum temperature (Celsius*10), minmum temperature (Celsius*10), average temperature (Celsius*10) and trusted observations are listed above.

#Exercise 3

Year	Max	Min
1910	250	-167
1911	294	-194
1912	283	-233
1913	283	-194
1914	283	-217
1915	256	-228

The maximum mean temperature station is Russaro, the above table is the table for each year (1910-1915), max temperature (Celsius *10) and min temperature (Celsius*10) in the station of Russaro.