

# CS100 Tutorial 13

Weather data analysis and Visualization

# Preliminary

- Download data file [weather\\_data.xlsx](#) from piazza, collected from [http://tianqi.2345.com/wea\\_history/71146.htm](http://tianqi.2345.com/wea_history/71146.htm)
- The file has Pudong's weather data from July-December

	A	B	C	D	E	F
1	日期	最高气温	最低气温	天气	风向风力	空气质量指数
2	2018-12-02星期日	20℃	14℃	小雨~多云	东南风3级	46优
3	2018-12-03星期一	18℃	14℃	小雨	东南风2级	71良
4	2018-12-04星期二	16℃	10℃	小雨~阴	西北风4级	45优
5	2018-12-05星期三	14℃	11℃	小雨	东北风4级	51良
6	2018-12-06星期四	11℃	6℃	中雨~小雨	东北风4级	26优
7	2018-12-07星期五	8℃	3℃	阴~小雨	东北风4级	34优
8	2018-12-08星期六	6℃	2℃	小雨~雨夹雪	西北风4级	38优

# Preliminary

- Install **NumPy**, **Pandas**, **Matplotlib**
- In order to read .xlsx format file, you might need to install **xlrd**
- Check whether everything is correct by print the **first five columns**

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

data =
pd.read_excel('weather_data.xlsx')
data.head()
```

# Task 1: Data process

## Divide

1. the column “日期” into two columns “日期” and “星期”
2. the column “天气” into two columns “最好天气” and “最坏天气” in the order of  
晴>多云>阴>阵雨>雷阵雨>小雨>中雨>大雨>暴雨>雨夹雪>中雪
3. the column “风向风力” into three columns “风向” and “最低风速” and “最高风速”
4. the column “空气质量指数” into two columns “空气质量” and “空气指数”
5. remove ‘C°’ in the “气温” columns

# Task 1: Data process

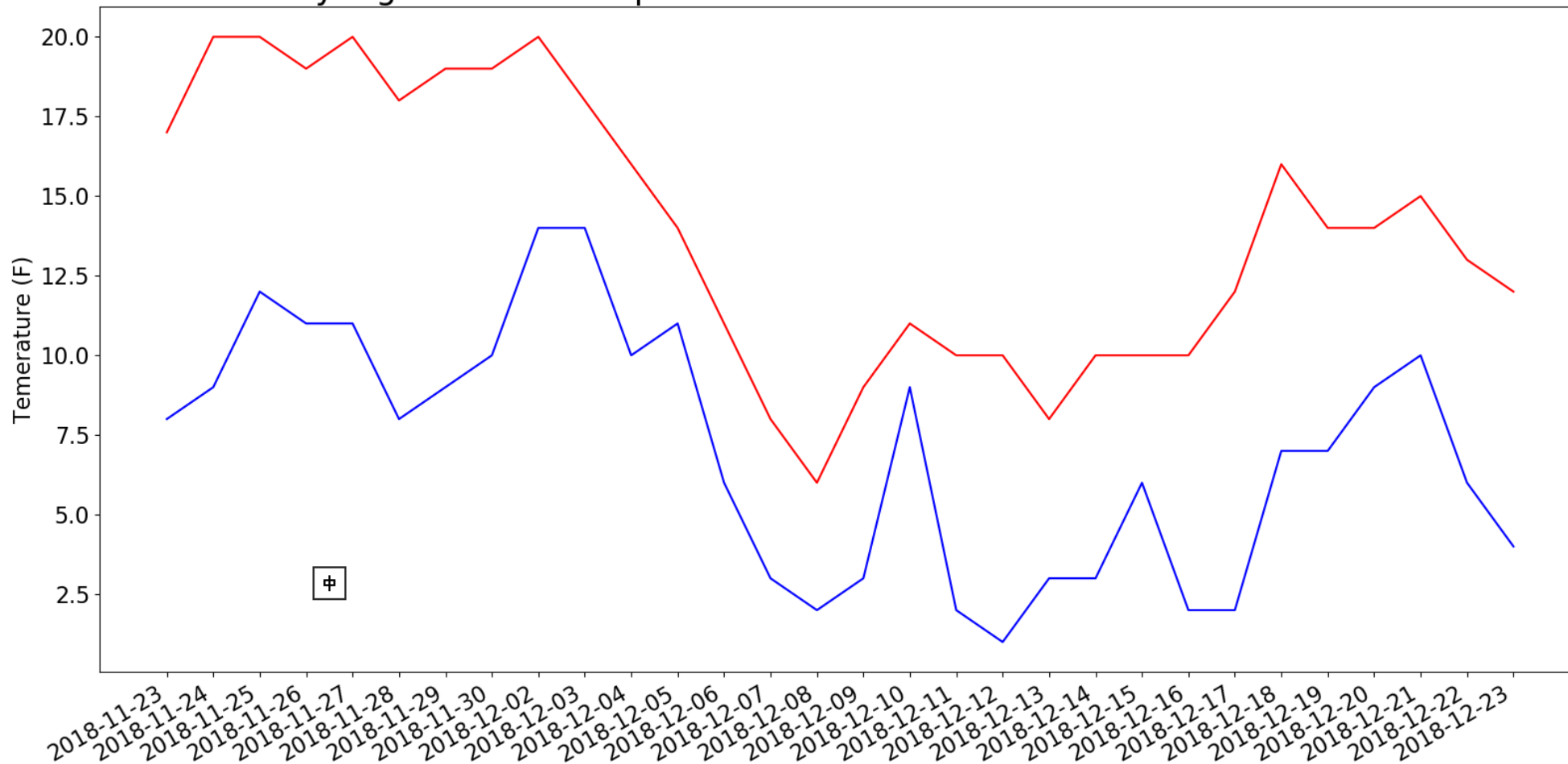
Result:

	日期	最高气温	最低气温	星期	最好天气	最坏天气	风向	最低风速	最高风速	空气指数	空气质量
0	2018-12-02	20	14	星期日	多云	小雨	东南风	3	3	46	优
1	2018-12-03	18	14	星期一	小雨	小雨	东南风	2	2	71	良
2	2018-12-04	16	10	星期二	阴	小雨	西北风	4	4	45	优
3	2018-12-05	14	11	星期三	小雨	小雨	东北风	4	4	51	良
4	2018-12-06	11	6	星期四	小雨	中雨	东北风	4	4	26	优
5	2018-12-07	8	3	星期五	阴	小雨	东北风	4	4	34	优
6	2018-12-08	6	2	星期六	小雨	雨夹雪	西北风	4	4	38	优
7	2018-12-09	9	3	星期日	小雨	小雨	西北风	3	3	40	优
8	2018-12-10	11	9	星期一	小雨	中雨	东北风	4	4	20	优
9	2018-12-11	10	2	星期二	多云	小雨	西北风	4	4	25	优
10	2018-12-12	10	1	星期三	多云	多云	西北风	4	4	92	良
11	2018-12-13	8	3	星期四	多云	多云	西北风	3	3	90	良

# Task 2: Two Lines Plot

Draw the lines of “最高气温” and “最低气温” for latest 30 days

Daily high and low temperatures between 2018-11-23 and 2018-12-23



`fig.autofmt_xdate()`  
# format label of  
x-axis

## Task 3: Two subplots

Draw two plots, one is the plot in Task 2 and the another is the line of difference in temperature in the latest 30 days

Daily high and low temperatures between 2018-11-23 and 2018-12-23

