Jose Dixon
12.09.2016
Dr. Rahman

Assignment 4

Question 1

Source Code

```
import csv
from numpy import genfromtxt
from pandas import Series, DataFrame
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
data = pd.read_csv("weather.csv")
data ##1. Prints out the data as a summary.
data.iloc[364] ##2. Prints out the second to last row.
rangea = pd.read_csv('weather.csv', usecols=[1])
rtempa = rangea.max(skipna=True, numeric_only=True)
rtempa
rangeb = pd.read_csv('weather.csv', usecols=[3])
rtempb = rangeb.min(skipna=True, numeric_only=True)
rtempb
np.subtract(rtempa, rtempb) #3. The range of two temperatures in a dataset.
data1 = pd.read_csv('weather.csv', quoting=2, usecols=[1], sep=",") ##4. First histogram of
MaxTemperatureF.
data1.hist(bins=2)
plt.xlim([0, 150])
plt.ylim([0, 300])
plt.title("MaxTemperatureF")
plt.xlabel("Range of Time")
plt.ylabel("Temperature Range")
plt.legend()
plt.show()
data2 = pd.read_csv('weather.csv', quoting=2, usecols=[2], sep=",") ##4. Second histrogram of
MeanTemperatureF
data2.hist(bins=2)
plt.xlim([0, 150])
plt.ylim([0, 250])
plt.title("MeanTemperatureF")
plt.xlabel("Range of Time")
plt.ylabel("Temperature Range")
plt.legend()
plt.show()
rangeb = pd.read_csv('weather.csv', quoting=2, usecols=[20])
rtempc = rangeb.mean(skipna=True, numeric_only=True)
rtempc ##5.Group the DataFrame, and find the mean temperature of group.
ranged = pd.read_csv('weather.csv', usecols=[16])
ranged
rangee = pd.read_csv('weather.csv', usecols=[17])
```

rangee
np.subtract(ranged, rangee)
print "The mean temperature is more variable on days of rain than snow." #6

Output

#1 Display the data out as a summary.

Out[12]:

	EDT	Max TemperatureF	Mean TemperatureF	Min TemperatureF	Max Dew PointF	MeanDew PointF	Min DewpointF	Max Humidity	Mean Humidity	Min Humidity	...	Max VisibilityMiles	Mean Visibil
0	2012-3-10	56	40	24	24	20	16	74	50	26	...	10	10
1	2012-3-11	67	49	30	43	31	24	78	53	28	...	10	10
2	2012-3-12	71	62	53	59	55	43	90	76	61	...	10	10
3	2012-3-13	76	63	50	57	53	47	93	66	38	...	10	10
4	2012-3-14	80	62	44	58	52	43	93	68	42	...	10	10

#2. Prints out the second to last row.

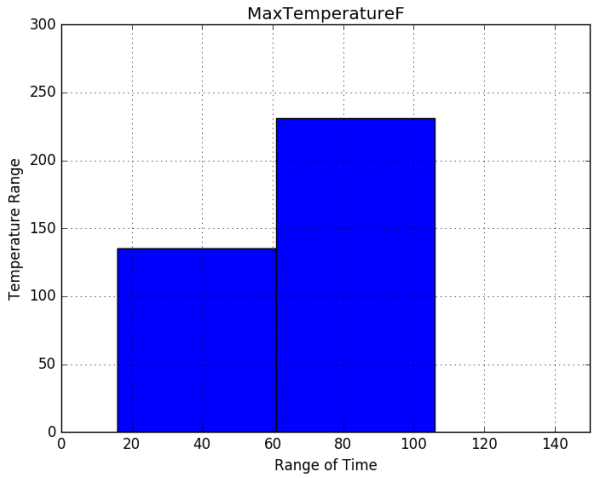
EDT 2013-3-9
Max TemperatureF 56
Mean TemperatureF 45
Min TemperatureF 33
Max Dew PointF 32
MeanDew PointF 29
Min DewpointF 23
Max Humidity 75
Mean Humidity 57
Min Humidity 38
Max Sea Level PressureIn 30.32
Mean Sea Level PressureIn 30.16
Min Sea Level PressureIn 29.96
Max VisibilityMiles 10
Mean VisibilityMiles 10
Min VisibilityMiles 10
Max Wind SpeedMPH 16
Mean Wind SpeedMPH 10
Max Gust SpeedMPH 24
PrecipitationIn T
CloudCover 2
Events NaN
WindDirDegrees 141
Name: 364, dtype: object

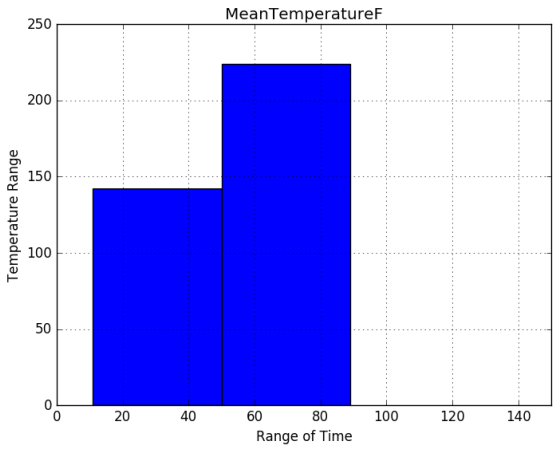
Max TemperatureF 106
dtype: int64

Min TemperatureF 1
dtype: int64

#3. The range of two temperatures.
Max TemperatureF 105
dtype: int64

#4. Histogram for the two columns





##5.Group the DataFrame, and find the mean temperature of group.

CloudCover 2.885246
dtype: float64

	Max Wind SpeedMPH
0	13
1	22
2	24
3	16
4	16
5	31
6	14
7	12
8	18
9	17
10	17
11	17
12	31
13	21
14	12
15	14
16	17

	Mean Wind SpeedMPH
0	6
1	7
2	14
3	5
4	6
5	10
6	5
7	5
8	8
9	6
10	5
11	6
12	4
13	6
14	5
15	5
16	9

	Max Wind SpeedMPH
0	7
1	15
2	10
3	11
4	10
5	21
6	9
7	7
8	10
9	11
10	12
11	11
12	27
13	15
14	7
15	9
16	8

Question 2

Source Code

```
import csv
from numpy import genfromtxt
from pandas import Series, DataFrame
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
data = pd.read_csv("student.csv")
data
```

```
rangea = pd.read_csv('student.csv', usecols=[2, 3, 4]) #1. What does the data tell you. Sum of data.
rtempa = rangea.sum(skipna=True, numeric_only=True)
rtempa
rtempb = rangea.mean(skipna=True, numeric_only =True) #1. What does the data tell you. Mean of data.
rtempb
print "The schools with the highest dropout rates are Peace and Education High School and Banner Academy South High School."
rtempc = rangea.max(skipna=True, numeric_only=True)
rtempc
#2. Schools with highest dropout rates.
print "Network of schools with the higher dropout rates is Alternative Schools. "
rtempg = data.iloc[9]
#3. Network of schools with highest dropout rates.
print "The schools with the highest passing rates are Devry High School, Northside Learning High School, and Northside Learning Prep High School."
rtempd = rangea.min(skipna=True, numeric_only=True)
rtempd #4. Schools with highest passing rates.
print "The school with the highest passing rate of students is Northside Prep High School."
rtempe = data.iloc[94]
rtempe #5. School with largest passing percentages.
print "The school with highest passing students and percentages is Northside Learning High School."
rtempf = data.iloc[93]
rtempf #6. School with passing percentage and high pass students.
```

Output

	School Name	Network	DO2010	DO2011	DO2012
0	ACE TECHNICAL CHARTER HS	Charter/Contract Schools	7.1	7.6	7.3
1	AIR FORCE HS	Southwest Side High School Network	4.9	5.0	0.7
2	ALCOTT HS	North-Northwest Side High School Network	0.0	4.8	1.8
3	AMUNDSEN HS	North-Northwest Side High School Network	9.1	6.9	6.0
4	ASPIRA CHTR - EARLY COLLEGE HS	Charter/Contract Schools	4.7	4.3	4.8
5	ASPIRA CHTR - RAMIREZ HS	Charter/Contract Schools	6.4	7.6	4.8
6	AUSTIN BUS & ENTRP HS	West Side High School Network	14.7	13.8	6.9
7	AUSTIN POLY HS	West Side High School Network	6.0	1.1	18.5
8	BANNER ACADEMY SOUTH HS	Alternative Schools	NaN	NaN	52.6
9	BANNER ACADEMY WEST HS	Alternative Schools	NaN	NaN	52.4

DO2010 1118.1
DO2011 1076.9
DO2012 1293.5
dtype: float64

DO2010 8.102174
DO2011 7.478472
DO2012 8.566225
dtype: float64

DO2010 46.8
DO2011 41.9
DO2012 52.6

dtype: float64

Network of schools with the higher dropout rates is Alternative Schools.

The schools with the highest passing rates are Devry High School, Northside Learning High School, and Northside Learning Prep High School.
DO2010 0.0
DO2011 0.1
DO2012 0.0
dtype: float64

The school with the highest passing rate of students is Northside Learning High School.

Out[26]:

School Name	NORTHSIDE PREP HS
Network	North-Northwest Side High School Network
DO2010	0.4
DO2011	0.1
DO2012	0.4

Name: 94, dtype: object

The school with highest passing students and percentages is Northside Learning High School.

School Name	NORTHSIDE LEARNING HS
Network	North-Northwest Side High School Network
DO2010	1
DO2011	1
DO2012	0

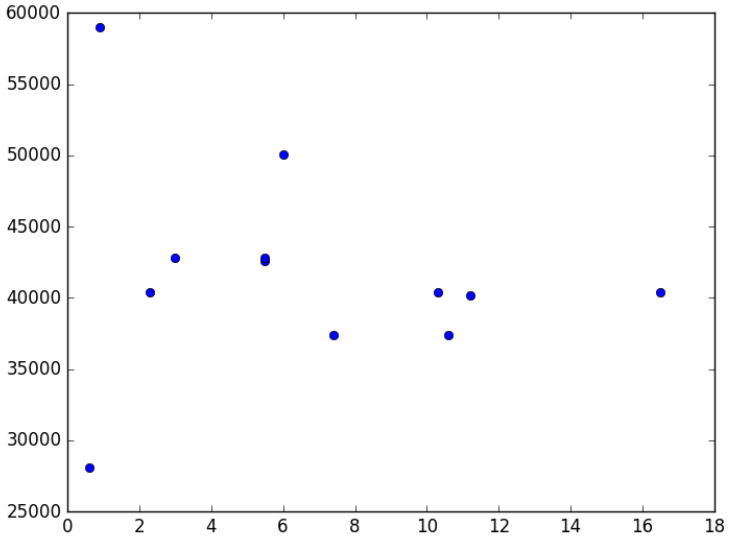
Name: 93, dtype: object

Question 3

Source Code

```
import scipy as sc
import matplotlib.pyplot as plt
import numpy as pd
school= ['amundsen', 'clemente', 'corliss', 'douglass', 'eric solorio academy hs', 'fenger', 'gage park', 'harlan', 'hirsch', 'hubbard', 'juarez', 'kelly']
DO2012=[6, 0.9, 2.3, 0.6, 3, 16.5, 10.6, 10.3, 11.2, 7.4, 5.5, 5.5]
median=[50065, 58987, 40394, 28059, 42809, 40394, 37367, 40394, 40176, 37367, 42575, 42809]
plt.plot(cor_data['DO2012'], cor_data['median'], 'bo')
plt.show()
regressionline = sc.stats.linregress(DO2012, median)
m = 0
b = 1
x = np.linspace(START, END, NUMBER_OF_POINTS)
plt.plot(x, m*x + b)
plt.show()
```

Output


60000
55000
50000
45000
40000
35000
30000
25000
0
2
4
6
8
10
12
14
16
18