

case_study_2

April 25, 2017

1 Question 04

1.1 (i)

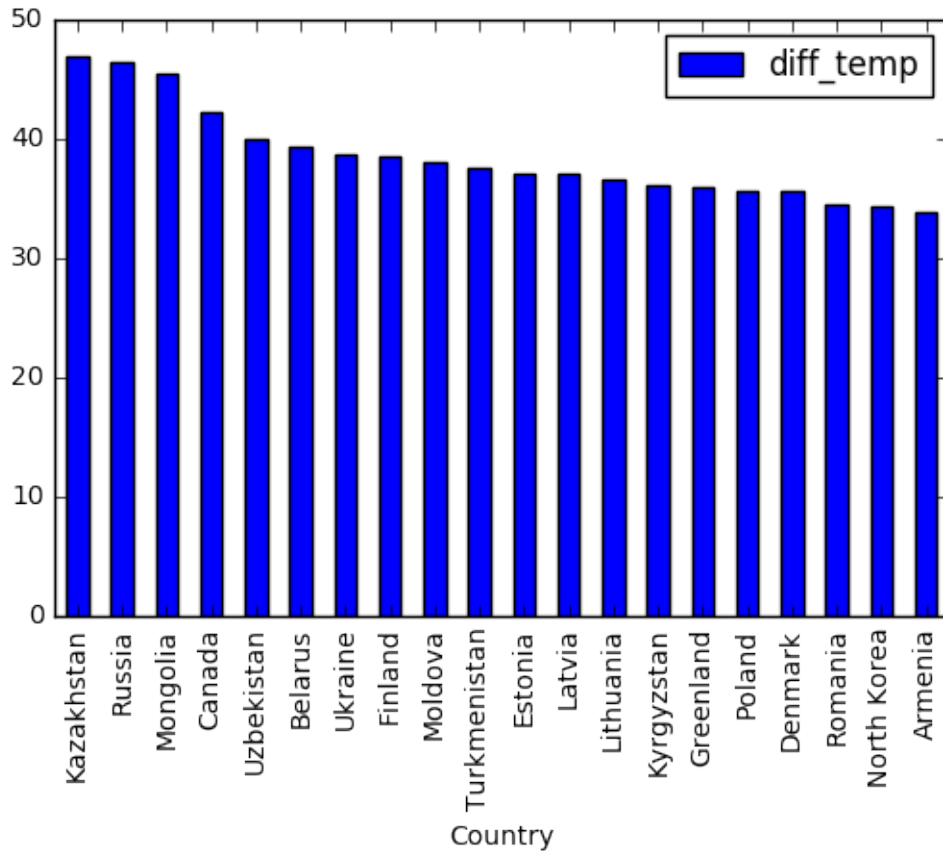
```
In [248]: import pandas
temp_countries = pandas.read_csv('./TEMP.csv')
temp_countries_gt_1900 = temp_countries[temp_countries['Date'] > '1900-01-01']
temp_countries_gt_1900_agg = temp_countries_gt_1900.groupby(['Country'])
temp_countries_gt_1900_agg['diff_temp'] = abs(temp_countries_gt_1900_agg['min_temp'] - temp_countries_gt_1900_agg['max_temp'])
temp_countries_gt_1900_agg = temp_countries_gt_1900_agg.drop(['min_temp', 'max_temp'])
temp_countries_gt_1900_agg_20_largest = temp_countries_gt_1900_agg.nlargest(20, 'diff_temp')
```

```
In [249]: temp_countries_gt_1900_agg_20_largest
```

```
Out[249]:
```

Country	diff_temp
Kazakhstan	46.971
Russia	46.454
Mongolia	45.426
Canada	42.240
Uzbekistan	39.886
Belarus	39.338
Ukraine	38.660
Finland	38.528
Moldova	38.012
Turkmenistan	37.573
Estonia	37.124
Latvia	36.985
Lithuania	36.635
Kyrgyzstan	36.084
Greenland	35.950
Poland	35.616
Denmark	35.547
Romania	34.464
North Korea	34.371
Armenia	33.856

```
In [250]: p = temp_countries_gt_1900_agg_20_largest.plot(y='diff_temp', kind='bar')
```



1.2 (ii)

```
In [251]: UStemp = temp_countries[(temp_countries['Date'] >= '1900-01-01') & (temp_
```

1.2.1 (ii) a)

```
In [252]: UStemp['TempInFahrenheit'] = 1.8 * UStemp['Monthly AverageTemp'] + 32
```

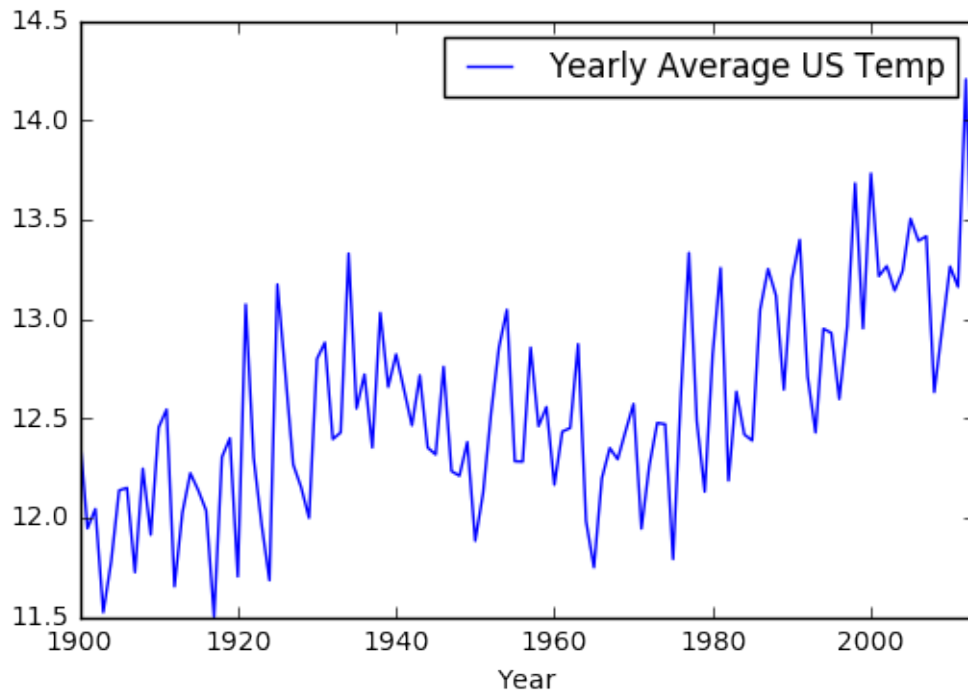
1.2.2 (ii) b)

```
In [253]: UStemp['Date'] = pandas.to_datetime(UStemp['Date'])
          UStemp['Year'] = UStemp['Date'].dt.year
```

```
In [254]: import numpy as np
          yearly_UStemp = UStemp.groupby(['Year'])['Monthly AverageTemp'].agg({"Yearly AverageTemp": lambda x: np.mean(x) * 12})
```

```
In [163]: import matplotlib.pyplot as plt

          %matplotlib inline
          p = yearly_UStemp.plot(x='Year', y='Yearly AverageTemp', label="Yearly AverageTemp")
```



1.2.3 (ii) c)

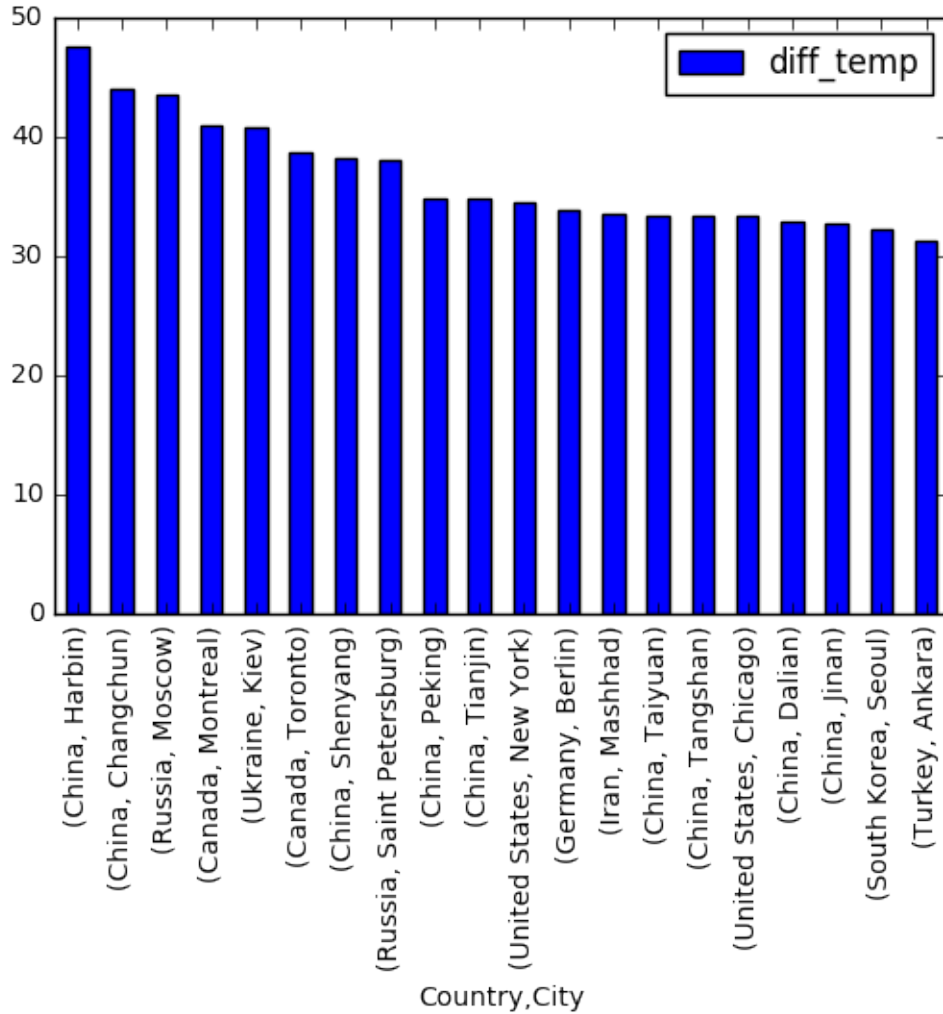
```
In [204]: yearly_UStemp_diff = abs(yearly_UStemp['Yearly AverageTemp'] - yearly_UStemp['Yearly AverageTemp'].shift(1))
maxindex = np.argmax(yearly_UStemp_diff)
y1 = yearly_UStemp.loc[maxindex-1]['Year'].astype(int)
y2 = yearly_UStemp.loc[maxindex]['Year'].astype(int)
max_diff = yearly_UStemp_diff[maxindex]
print("Max temp difference is between year %s and %s and is equal to %s."
```

Max temp difference is between year 1924 and 1925 and is equal to 1.48875.

1.3 (iii)

```
In [267]: import pandas
temp_cities = pandas.read_csv('./CityTemp.csv')
temp_cities_gt_1900 = temp_cities[temp_cities['Date'] > '1900-01-01']
temp_cities_gt_1900_agg = temp_cities_gt_1900.groupby(['Country', 'City'])
temp_cities_gt_1900_agg['diff_temp'] = abs(temp_cities_gt_1900_agg['min_temp'] - temp_cities_gt_1900_agg['max_temp'])
temp_cities_gt_1900_agg = temp_cities_gt_1900_agg.drop(['max_temp', 'min_temp'])
temp_cities_gt_1900_agg_20_largest = temp_cities_gt_1900_agg.nlargest(20, 'diff_temp')
```

```
In [268]: p = temp_cities_gt_1900_agg_20_largest.plot(y='diff_temp', kind='bar')
```



1.4 (iv)

The observation of the two graphs shows that they have approximately the same slope as we can see that both histograms have a maximum value around 480 and the minimum value around 350. Another remark is that the countries and cities that have the max average temperature are located in the same continent like China (in the country, City graph) and Kazakhstan (in the Country graph) are in Asia, and the lowest values are for Europe as Turkey and Arminia.