In [62]:
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [63]:
```python
DATA_PATH = '../data/Environment_Temperature_change_E_All_Data_NOFLAG.csv'
```

In [64]:
```python
data_frame = pd.read_csv(DATA_PATH, encoding='cp1252')
print(data_frame.shape)
```
```
(9656, 66)
```

In [65]:
```python
no_rows, no_columns = data_frame.shape
no_years_columns = 2019 - 1961 + 1
no_not_years_columns = no_columns - no_years_columns
year_columns = data_frame.columns[no_columns - no_years_columns:]
```

In [66]:
```python
data_frame.iloc[:10, : 10]
```

Out[66]:

	Area Code	Area	Months Code	Months	Element Code	Element	Unit	Y1961	Y1962	Y1963
0	2	Afghanistan	7001	January	7271	Temperature change	°C	0.777	0.062	2.744
1	2	Afghanistan	7001	January	6078	Standard Deviation	°C	1.950	1.950	1.950
2	2	Afghanistan	7002	February	7271	Temperature change	°C	-1.743	2.465	3.919
3	2	Afghanistan	7002	February	6078	Standard Deviation	°C	2.597	2.597	2.597
4	2	Afghanistan	7003	March	7271	Temperature change	°C	0.516	1.336	0.403
5	2	Afghanistan	7003	March	6078	Standard Deviation	°C	1.512	1.512	1.512
6	2	Afghanistan	7004	April	7271	Temperature change	°C	-1.709	0.117	0.919
7	2	Afghanistan	7004	April	6078	Standard Deviation	°C	1.406	1.406	1.406
8	2	Afghanistan	7005	May	7271	Temperature change	°C	1.412	-0.092	-0.690
9	2	Afghanistan	7005	May	6078	Standard Deviation	°C	1.230	1.230	1.230

In [67]:
```python
data_frame.describe()
```

Out[67]:

	Area Code	Months Code	Element Code	Y1961	Y1962	Y1963	Y196
count	9656.000000	9656.000000	9656.00000	8287.000000	8322.000000	8294.000000	8252.00000
mean	821.806338	7009.882353	6674.50000	0.402433	0.315527	0.317393	0.26938
std	1781.072213	6.038255	596.53089	0.701567	0.713777	0.853133	0.7492
min	1.000000	7001.000000	6078.00000	-4.018000	-5.391000	-8.483000	-7.30900
25%	78.000000	7005.000000	6078.00000	0.057000	-0.033000	0.030250	-0.10250
50%	153.500000	7009.000000	6674.50000	0.366000	0.333000	0.355000	0.32600
75%	226.250000	7016.000000	7271.00000	0.676500	0.627000	0.647750	0.60900
max	5873.000000	7020.000000	7271.00000	5.771000	4.373000	4.666000	5.23300

8 rows × 62 columns

```python
In [68]:  data_frame.loc[[355]].isnull().sum().sum()
```

Out[68]: 59

```python
In [69]:  null_values_list = []
          for index in range(data_frame.shape[0]):
              if data_frame.loc[index].isnull().sum().sum() == no_years_columns:
                  null_values_list.append(index)
```
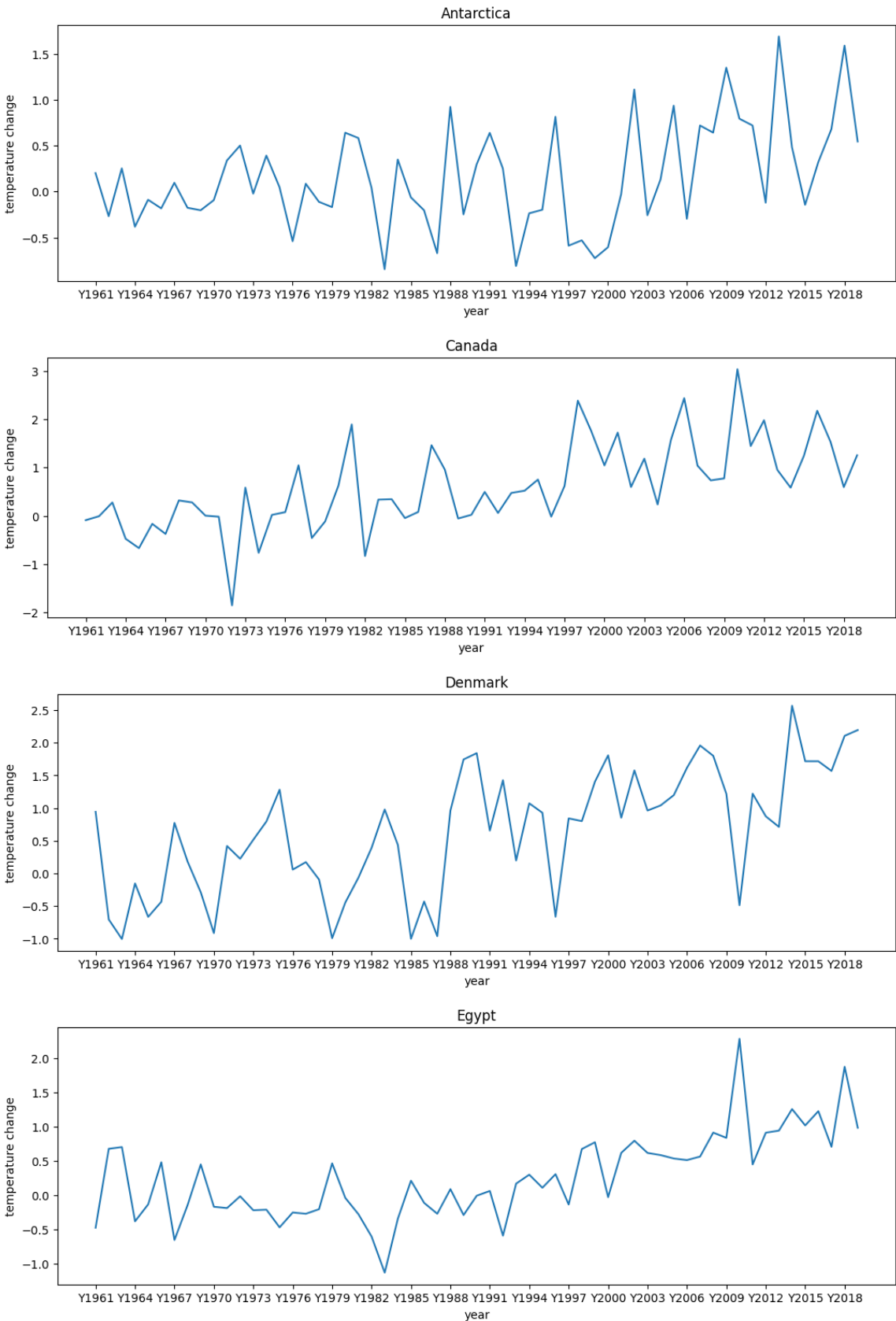
```python
In [70]:  data_frame.drop(null_values_list, inplace=True)
```

```python
In [71]:  no_rows, no_columns = data_frame.shape
```
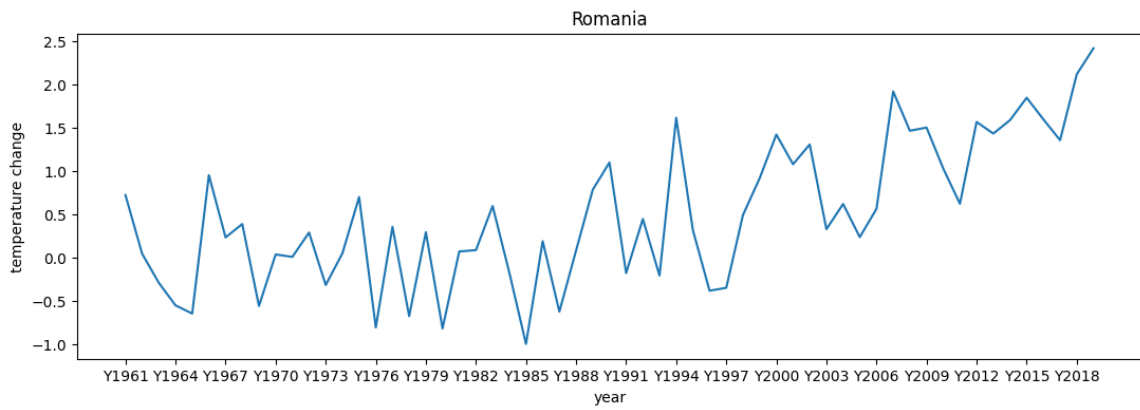
```python
In [72]:  for column in year_columns:
              data_frame[column] = data_frame[column].replace(np.nan, data_frame[column].m
```

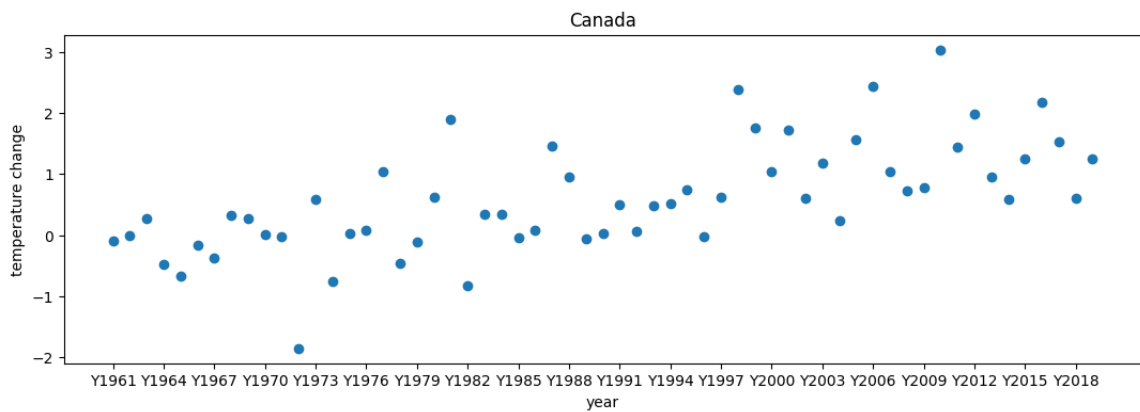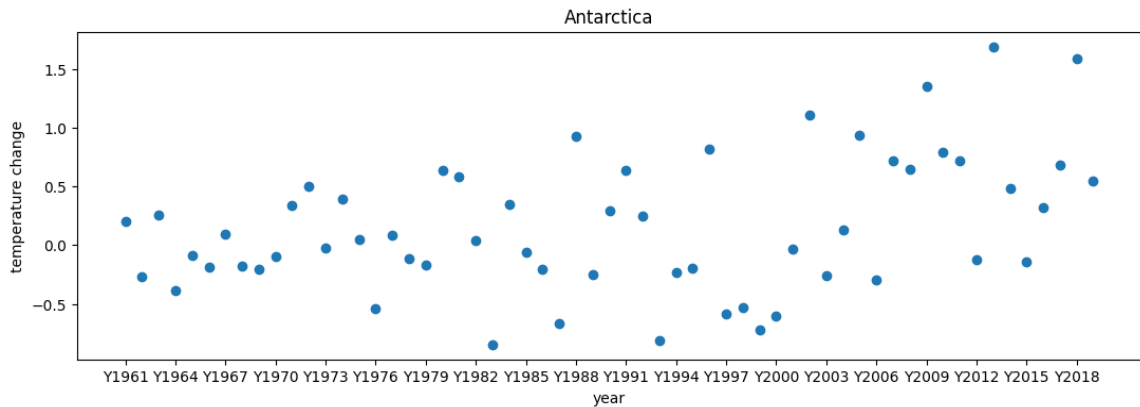# Am analizat datele referitoare la variatia temperaturii fata de temperaturile din perioada 1951-1980

```python
In [73]:  countries = ['Antarctica', 'Canada', 'Denmark', 'Egypt', 'Romania']
```
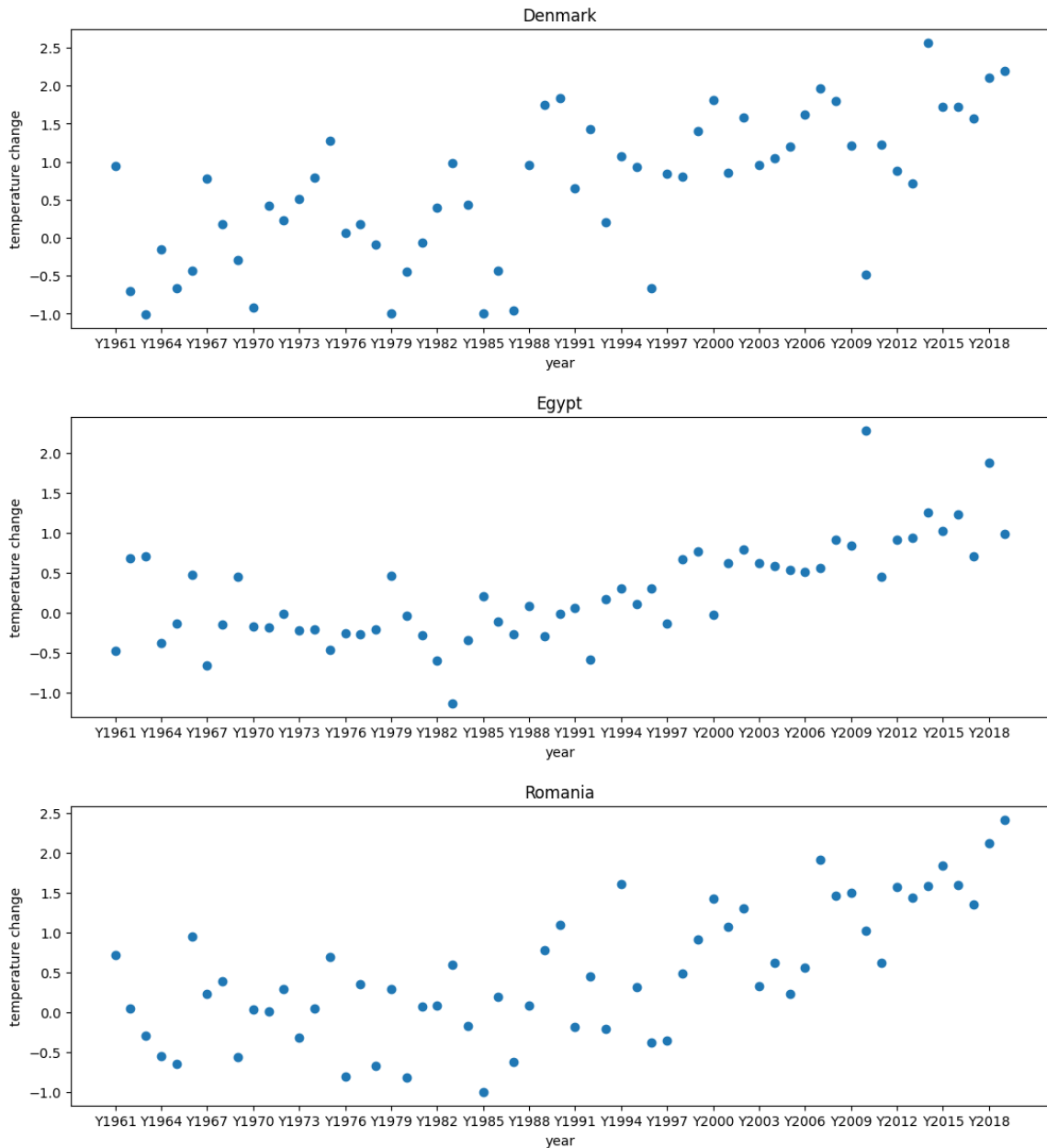
```python
In [74]:  for country in countries:
              plt.figure(figsize=[13,4])
              df = data_frame.loc[((data_frame.Area == country) & (data_frame.Element == '
              year_mean_temp_change = df[year_columns].mean()
              plt.plot(year_columns, year_mean_temp_change)
              plt.xlabel('year')
              plt.ylabel('temperature change')
              plt.xticks(year_columns[::3])
              plt.title(country)
```

Antarctica

Canada

Denmark

Egypt

Romania

```
In [75]:    for country in countries:
                plt.figure(figsize=[13,4])
                df = data_frame.loc[((data_frame.Area == country) & (data_frame.Element ==
                year_mean_temp_change = df[year_columns].mean()
                plt.scatter(year_columns, year_mean_temp_change)
                plt.xlabel('year')
                plt.ylabel('temperature change')
                plt.xticks(year_columns[::3])
                plt.title(country)
```


Antarctica

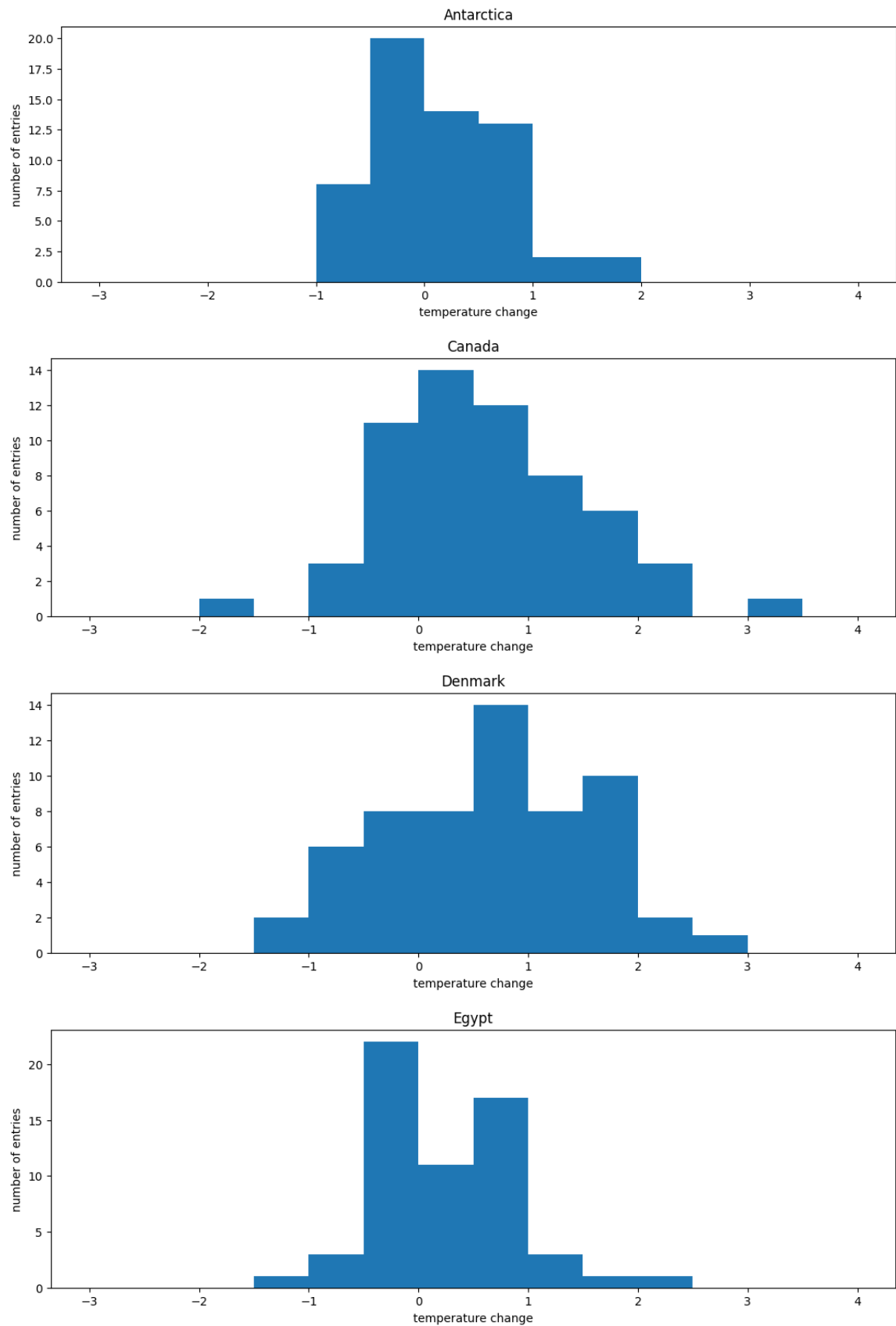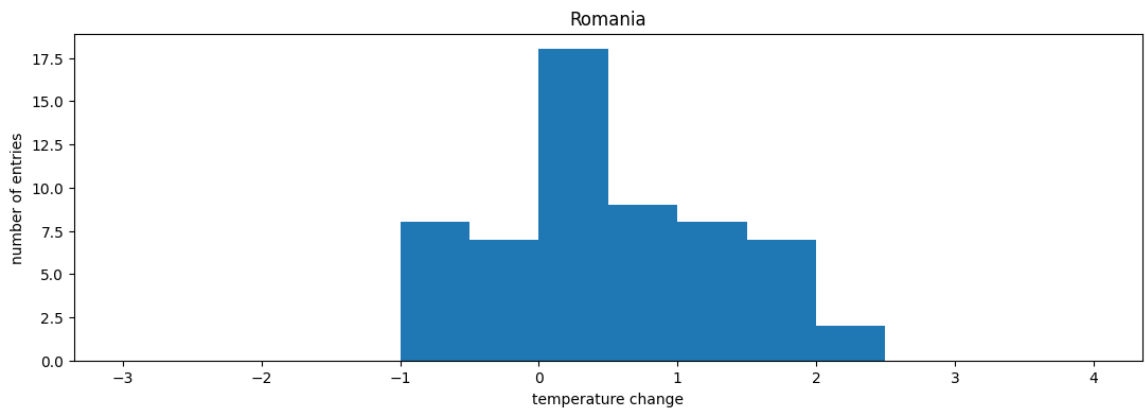Canada

Denmark

Egypt

Romania

Din scatterplot se observa ca se poate determina o functie cu proprietatea ca distanta de la fiecare punct la grafic este relativ constanta

```
In [76]:   for country in countries:
               plt.figure(figsize=[13,4])
               bin_range = (-3,4)
               bin_count = 14
               df = data_frame.loc[((data_frame.Area == country) & (data_frame.Element ==
               year_mean_temp_change = df[year_columns].mean()
               hist, bin_edges = np.histogram(year_mean_temp_change, bins=bin_count, range=

               plt.hist(bin_edges[:-1], bin_edges, weights=hist)
               plt.xlabel('temperature change')
               plt.ylabel('number of entries')
               plt.title(country)
```

Antarctica

Canada

Denmark

Egypt

Am decis sa adaugam media anuala a schimbarii de temperatura in histograma pentru a evidentia trendul ascendent, majoritatea valorilor fiind pozitive.