In [2]:

```
#load csv dataset
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df=pd.read_csv("globallandtemperaturesbymajorcity.csv")
df
```

3	1849-04-01	26.140	1.387	Abidjan	Côte D'Ivoire	5.63N	3.23W
4	1849-05-01	25.427	1.200	Abidjan	Côte D'Ivoire	5.63N	3.23W
...	...	...	...	...	...	...	...
239172	2013-05-01	18.979	0.807	Xian	China	34.56N	108.97E
239173	2013-06-01	23.522	0.647	Xian	China	34.56N	108.97E
239174	2013-07-01	25.251	1.042	Xian	China	34.56N	108.97E
239175	2013-08-01	24.528	0.840	Xian	China	34.56N	108.97E
239176	2013-09-01	NaN	NaN	Xian	China	34.56N	108.97E

239177 rows × 7 columns

In [3]:

```
#checking the dataset information
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 239177 entries, 0 to 239176
Data columns (total 7 columns):
 #   Column                         Non-Null Count   Dtype  
---  ------                         --------------   -----  
 0   dt                             239177 non-null  object 
 1   averagetemperature             228175 non-null  float64
 2   averagetemperatureuncertainty  228175 non-null  float64
 3   city                           239177 non-null  object 
 4   country                        239177 non-null  object 
 5   latitude                       239177 non-null  object 
 6   longitude                      239177 non-null  object 
dtypes: float64(2), object(5)
memory usage: 12.8+ MB
```

In [4]:

```
#checking the first 100 raws
df.head(100)
```

Out[4]:

	dt	averagetemperature	averagetemperatureuncertainty	city	country	latitude	long
0	1849-01-01	26.704	1.435	Abidjan	Côte D'Ivoire	5.63N	3
1	1849-02-01	27.434	1.362	Abidjan	Côte D'Ivoire	5.63N	3
2	1849-03-01	28.101	1.612	Abidjan	Côte D'Ivoire	5.63N	3
3	1849-04-01	26.140	1.387	Abidjan	Côte D'Ivoire	5.63N	3
4	1849-05-01	25.427	1.200	Abidjan	Côte D'Ivoire	5.63N	3
...	...	...	...	...	...	...	
95	1856-12-01	NaN	NaN	Abidjan	Côte D'Ivoire	5.63N	3
96	1857-01-01	26.549	1.749	Abidjan	Côte D'Ivoire	5.63N	3
97	1857-02-01	NaN	NaN	Abidjan	Côte D'Ivoire	5.63N	3
98	1857-03-01	27.299	1.263	Abidjan	Côte D'Ivoire	5.63N	3
99	1857-04-01	26.069	1.206	Abidjan	Côte D'Ivoire	5.63N	3

100 rows × 7 columns

In [5]:

```
#checking the last 100 raws
df.tail(100)
```

Out[5]:

	dt	averagetemperature	averagetemperatureuncertainty	city	country	latitude	lor
239077	2005-06-01	24.538	1.157	Xian	China	34.56N	1
239078	2005-07-01	25.045	0.429	Xian	China	34.56N	1
239079	2005-08-01	21.882	0.306	Xian	China	34.56N	1
239080	2005-09-01	19.307	0.402	Xian	China	34.56N	1
239081	2005-10-01	11.386	0.331	Xian	China	34.56N	1
...	...	...	...	...	...	...	
239172	2013-05-01	18.979	0.807	Xian	China	34.56N	1
239173	2013-06-01	23.522	0.647	Xian	China	34.56N	1
239174	2013-07-01	25.251	1.042	Xian	China	34.56N	1
239175	2013-08-01	24.528	0.840	Xian	China	34.56N	1
239176	2013-09-01	NaN	NaN	Xian	China	34.56N	1

100 rows × 7 columns

In [6]:

```
df.describe()
```

Out[6]:

	averagetemperature	averagetemperatureuncertainty
count	228175.000000	228175.000000
mean	18.125969	0.969343
std	10.024800	0.979644
min	-26.772000	0.040000
25%	12.710000	0.340000
50%	20.428000	0.592000
75%	25.918000	1.320000
max	38.283000	14.037000

In [7]:

```
df.shape
```

Out[7]:

```
(239177, 7)
```

In [8]:

```
#checking the columns
df.columns
```

Out[8]:

```
Index(['dt', 'averagetemperature', 'averagetemperatureuncertainty', 'cit
y',
       'country', 'latitude', 'longitude'],
      dtype='object')
```

In [9]:

```
#summary of the dataset

df.describe(include="all")
```

Out[9]:

	dt	averagetemperature	averagetemperatureuncertainty	city	country	latitude
count	239177	228175.000000	228175.000000	239177	239177	239177
unique	3239	NaN	NaN	100	49	49
top	1983-12-01	NaN	NaN	Rome	India	31.35N
freq	100	NaN	NaN	3239	36582	13875
mean	NaN	18.125969	0.969343	NaN	NaN	NaN
std	NaN	10.024800	0.979644	NaN	NaN	NaN
min	NaN	-26.772000	0.040000	NaN	NaN	NaN
25%	NaN	12.710000	0.340000	NaN	NaN	NaN
50%	NaN	20.428000	0.592000	NaN	NaN	NaN
75%	NaN	25.918000	1.320000	NaN	NaN	NaN
max	NaN	38.283000	14.037000	NaN	NaN	NaN

In [10]:

```
#the correllation

df.corr()
```

```
C:\Users\User\AppData\Local\Temp\ipykernel_12168\1014361338.py:3: FutureWa
rning: The default value of numeric_only in DataFrame.corr is deprecated.
In a future version, it will default to False. Select only valid columns o
r specify the value of numeric_only to silence this warning.
  df.corr()
```

Out[10]:

	averagetemperature	averagetemperatureuncertainty
averagetemperature	1.00000	-0.19938
averagetemperatureuncertainty	-0.19938	1.00000

In [11]:

```
#obtain the averagetemperature column

df.averagetemperature
```

Out[11]:

```
0         26.704
1         27.434
2         28.101
3         26.140
4         25.427
           ...
239172    18.979
239173    23.522
239174    25.251
239175    24.528
239176       NaN
Name: averagetemperature, Length: 239177, dtype: float64
```

In [12]:

```
#obtain the averagetemperatureuncertainty column

df.averagetemperatureuncertainty
```

Out[12]:

```
0         1.435
1         1.362
2         1.612
3         1.387
4         1.200
          ...
239172    0.807
239173    0.647
239174    1.042
239175    0.840
239176      NaN
Name: averagetemperatureuncertainty, Length: 239177, dtype: float64
```

In [13]:

```
#checking empty cells

df.isnull()
```

Out[13]:

	dt	averagetemperature	averagetemperatureuncertainty	city	country	latitude	lo
0	False	False	False	False	False	False	
1	False	False	False	False	False	False	
2	False	False	False	False	False	False	
3	False	False	False	False	False	False	
4	False	False	False	False	False	False	
...	...	...	...	...	...	...	
239172	False	False	False	False	False	False	
239173	False	False	False	False	False	False	
239174	False	False	False	False	False	False	
239175	False	False	False	False	False	False	
239176	False	True	True	False	False	False	

239177 rows × 7 columns

In [14]:

```
#removing empty cells
df.dropna()
```

Out[14]:

	dt	averagetemperature	averagetemperatureuncertainty	city	country	latitude	longitude
0	1849-01-01	26.704	1.435	Abidjan	Côte D'Ivoire	5.63N	3.23W
1	1849-02-01	27.434	1.362	Abidjan	Côte D'Ivoire	5.63N	3.23W
2	1849-03-01	28.101	1.612	Abidjan	Côte D'Ivoire	5.63N	3.23W
3	1849-04-01	26.140	1.387	Abidjan	Côte D'Ivoire	5.63N	3.23W
4	1849-05-01	25.427	1.200	Abidjan	Côte D'Ivoire	5.63N	3.23W
...	...	...	...	...	...	...	...
239171	2013-04-01	12.563	1.823	Xian	China	34.56N	108.97E

In [16]:

```
#confirm empty cells

df.isnull()
df.count()
```

...

In [17]:

```
df.dropna()
df.count()
```

Out[17]:

```
dt                               239177
averagetemperature               228175
averagetemperatureuncertainty    228175
city                             239177
country                          239177
latitude                         239177
longitude                        239177
dtype: int64
```

In [18]:

```
#mean of the averagetemperature

mean_value=df["averagetemperature"].mean()
mean_value
```

Out[18]:

```
18.125968852854168
```

In [20]:

```
#checking the total empty cell in averagetemperature
df.isnull().averagetemperature.sum()
```

Out[20]:

```
11002
```

In [22]:

```
#removing empty cells in averagetemperature

df.dropna().averagetemperature
df.count()
```

Out[22]:

```
dt                               239177
averagetemperature               228175
averagetemperatureuncertainty    228175
city                             239177
country                          239177
latitude                         239177
longitude                        239177
dtype: int64
```

In [24]:

```
#mean of the averagetemperatureuncertainty

mean_value=df["averagetemperatureuncertainty"].mean()
mean_value
```

Out[24]:

```
0.9693434381505424
```

In [28]:

```
#removing empty cells in averagetemperature uncertainty

df.dropna().averagetemperatureuncertainty
df.count()
```

Out[28]:

```
dt                               239177
averagetemperature               228175
averagetemperatureuncertainty    228175
city                             239177
country                          239177
latitude                         239177
longitude                        239177
dtype: int64
```

In [30]:

```
#replacing empty cell with mean value
df["averagetemperature"].fillna(mean_value, inplace=True)
```

In [31]:

```
df["averagetemperatureuncertainty"].fillna(mean_value, inplace=True)
```

In [32]:

```
#confirm empty cells
df.isnull().sum()
```

Out[32]:

```
dt                               0
averagetemperature               0
averagetemperatureuncertainty    0
city                             0
country                          0
latitude                         0
longitude                        0
dtype: int64
```

In [35]:

```
#check duplicates

df.duplicated().sum()
```

Out[35]:

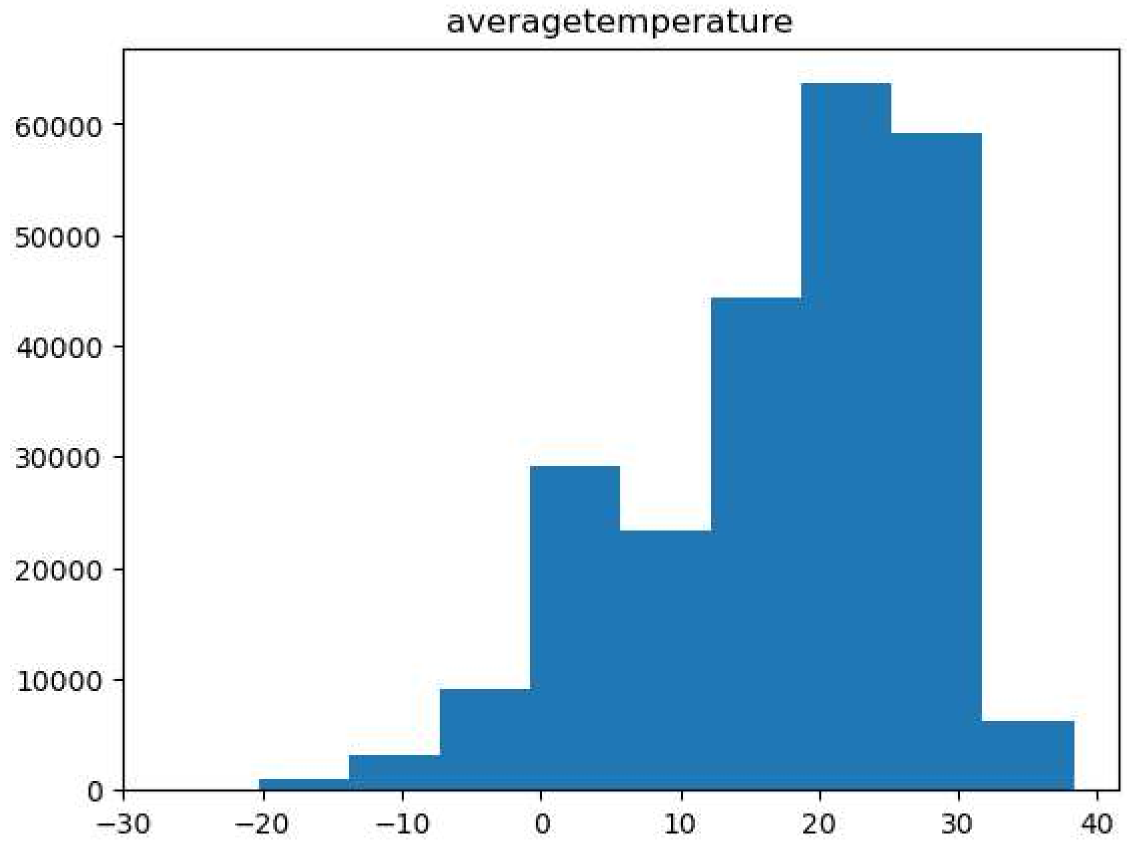
```
0
```

In [36]:

```
#visualizing using histogram
plt.title("averagetemperature")
plt.hist(df.averagetemperature)
plt.figure(figsize=(20,8))
```

Out[36]:

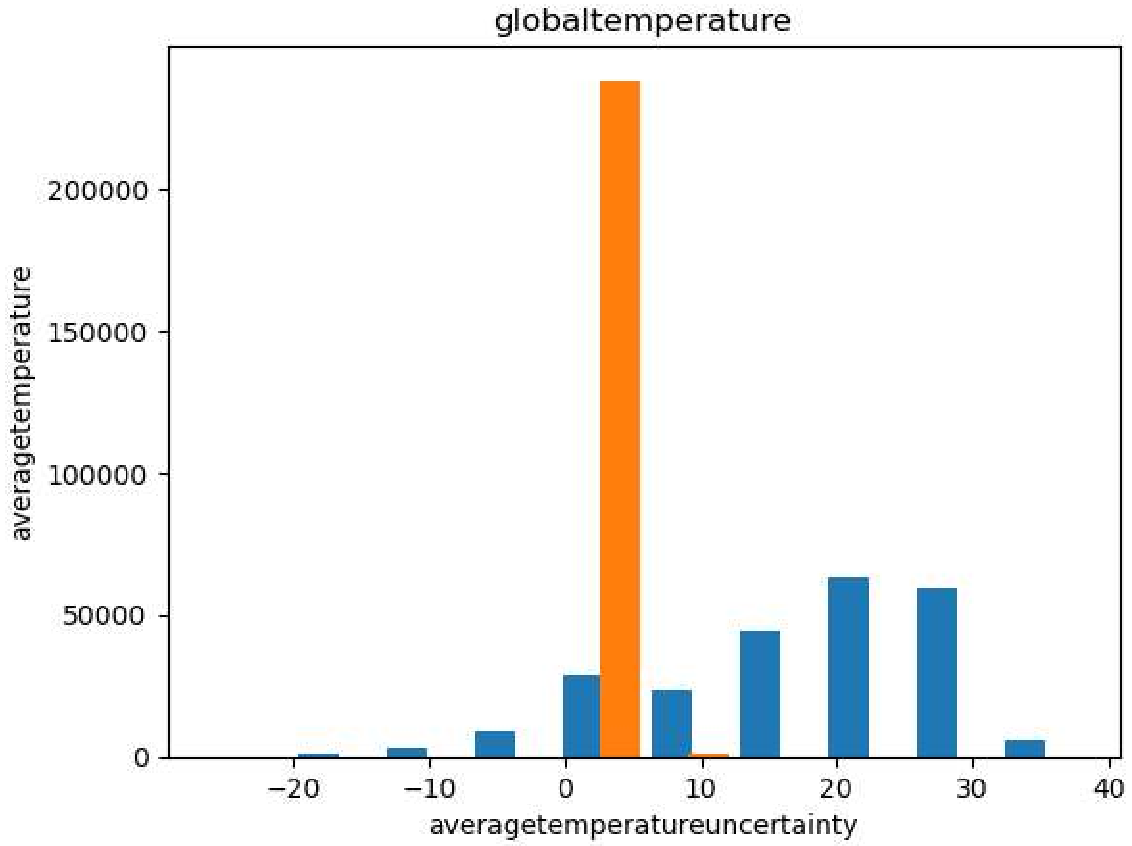
```
<Figure size 2000x800 with 0 Axes>
```



```
<Figure size 2000x800 with 0 Axes>
```

In [47]:

```
plt.hist(d,width=3, align="mid")
plt.title("globaltemperature")
plt.xlabel("averagetemperatureuncertainty")
plt.ylabel("averagetemperature")
plt.show()
```

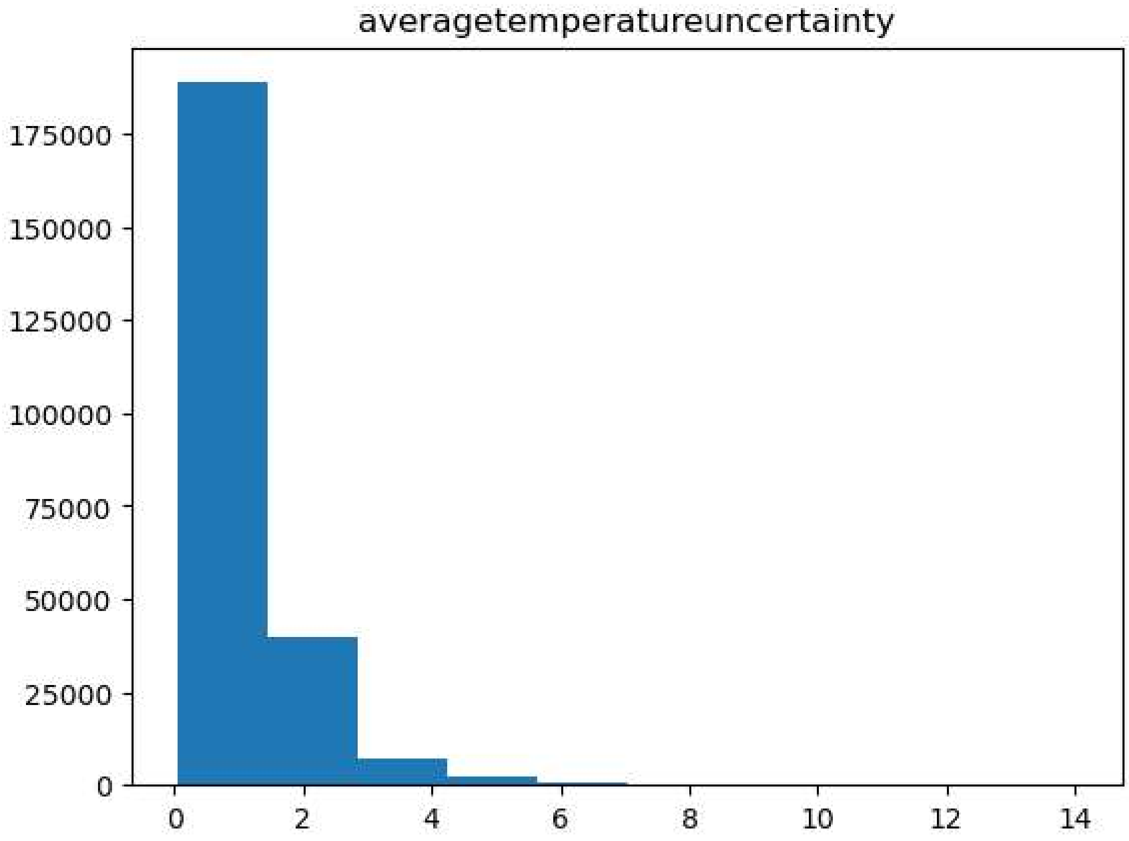


In [41]:

```
plt.title("averagetemperatureuncertainty")
plt.hist(df.averagetemperatureuncertainty)
plt.figure(figsize=(20,8))
```

Out[41]:

```
<Figure size 2000x800 with 0 Axes>
```



```
<Figure size 2000x800 with 0 Axes>
```

In [45]:

```
d=df[["averagetemperature","averagetemperatureuncertainty"]]
d
```

Out[45]:

```
        averagetemperature  averagetemperatureuncertainty
0                26.704000                       1.435000
1                27.434000                       1.362000
2                28.101000                       1.612000
3                26.140000                       1.387000
4                25.427000                       1.200000
...                    ...                            ...
239172           18.979000                       0.807000
239173           23.522000                       0.647000
239174           25.251000                       1.042000
239175           24.528000                       0.840000
239176            0.969343                       0.969343
```

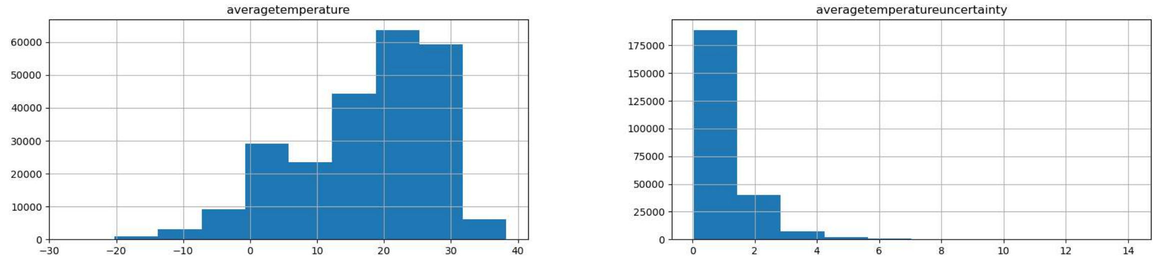
239177 rows × 2 columns

In [48]:

```
df.hist(figsize=(20,4))
```

Out[48]:

```
array([[<Axes: title={'center': 'averagetemperature'}>,
        <Axes: title={'center': 'averagetemperatureuncertainty'}>]],
      dtype=object)
```

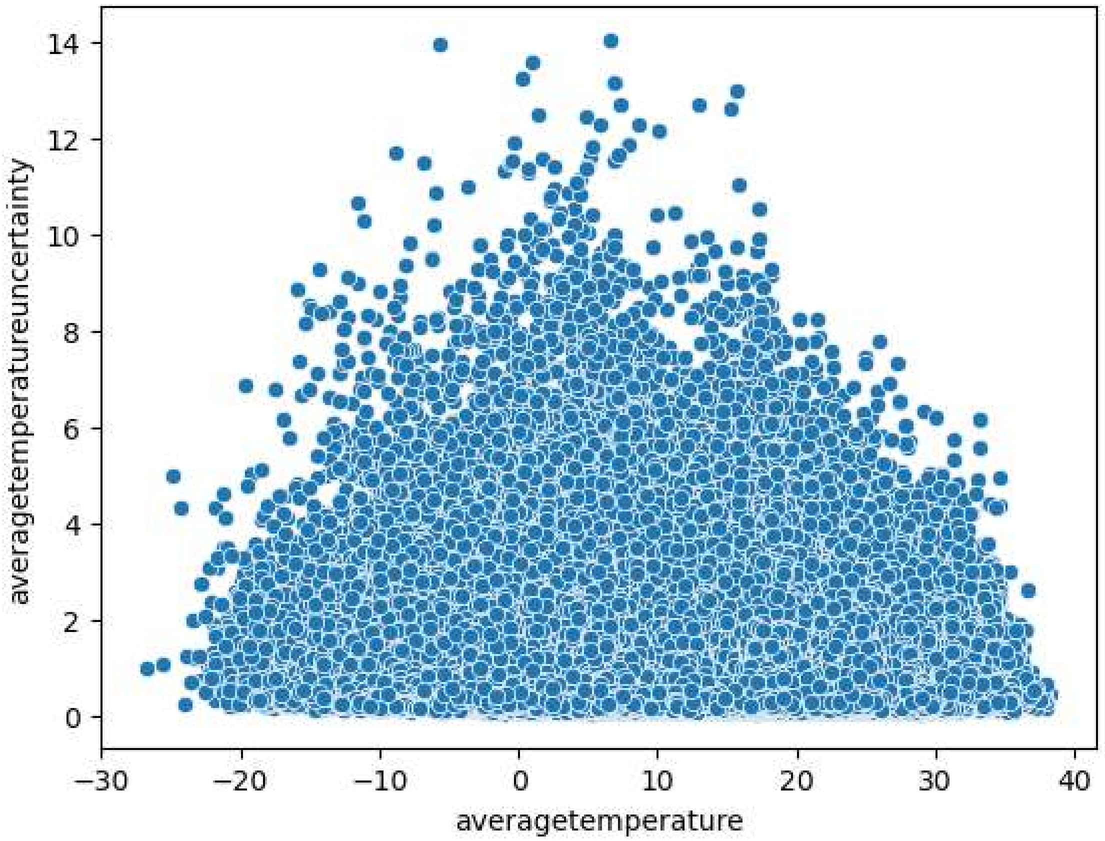


In [69]:

```
sns.scatterplot(x=df.averagetemperature, y=df.averagetemperatureuncertainty)
```

Out[69]:

```
<Axes: xlabel='averagetemperature', ylabel='averagetemperatureuncertaint
y'>
```

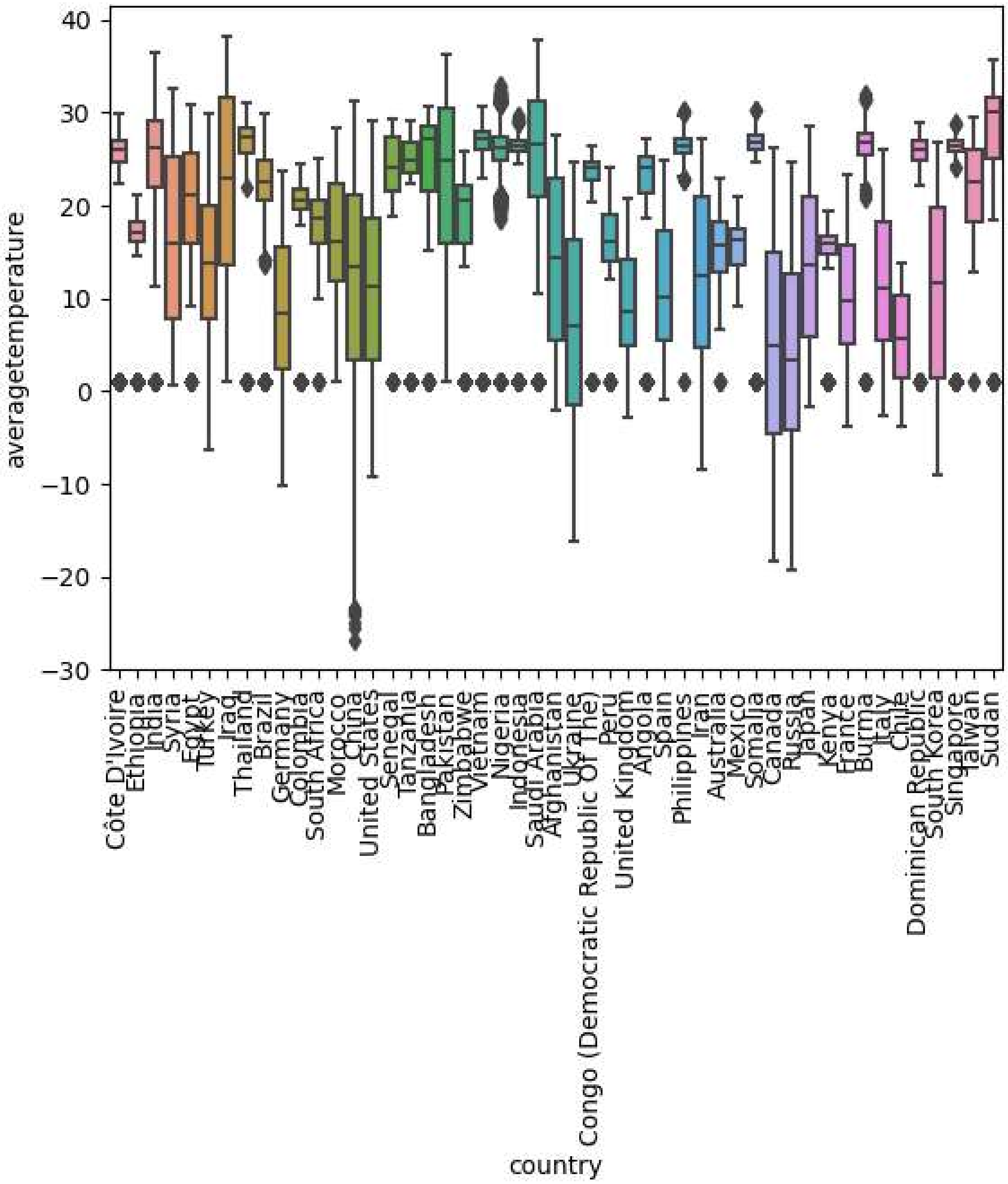


In [67]:

```
sns.boxplot(x=df.country, y=df.averagetemperature)
plt.xticks(rotation=90)
plt.figure(figsize=(20,8))
```

Out[67]:

```
<Figure size 2000x800 with 0 Axes>
```



```
<Figure size 2000x800 with 0 Axes>
```

In [ ]:

```

```

In [ ]: