In [1]:

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [9]:

```python
df1 = pd.read_csv("C:/Users/Jahnavi/Downloads/Unemployement/Unemployment in India.csv")
df1
```

Out[9]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area
0	Andhra Pradesh	31-05-2019	Monthly	3.65	11999139.0	43.24	Rural
1	Andhra Pradesh	30-06-2019	Monthly	3.05	11755881.0	42.05	Rural
2	Andhra Pradesh	31-07-2019	Monthly	3.75	12086707.0	43.50	Rural
3	Andhra Pradesh	31-08-2019	Monthly	3.32	12285693.0	43.97	Rural
4	Andhra Pradesh	30-09-2019	Monthly	5.17	12256762.0	44.68	Rural
...	...	...	...	...	...	...	...
763	NaN	NaN	NaN	NaN	NaN	NaN	NaN
764	NaN	NaN	NaN	NaN	NaN	NaN	NaN
765	NaN	NaN	NaN	NaN	NaN	NaN	NaN
766	NaN	NaN	NaN	NaN	NaN	NaN	NaN
767	NaN	NaN	NaN	NaN	NaN	NaN	NaN

768 rows × 7 columns

In [10]:

```python
df2 = pd.read_csv("C:/Users/Jahnavi/Downloads/Unemployement/Unemployment_Rate_upto_11_2020.csv")
df2
```

Out[10]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Region.1	longitude	latitude
0	Andhra Pradesh	31-01-2020	M	5.48	16635535	41.02	South	15.9129	79.740
1	Andhra Pradesh	29-02-2020	M	5.83	16545652	40.90	South	15.9129	79.740
2	Andhra Pradesh	31-03-2020	M	5.79	15881197	39.18	South	15.9129	79.740
3	Andhra Pradesh	30-04-2020	M	20.51	11336911	33.10	South	15.9129	79.740
4	Andhra Pradesh	31-05-2020	M	17.43	12988845	36.46	South	15.9129	79.740
...	...	...	...	...	...	...	...	...	...
262	West Bengal	30-06-2020	M	7.29	30726310	40.39	East	22.9868	87.855
263	West Bengal	31-07-2020	M	6.83	35372506	46.17	East	22.9868	87.855
264	West Bengal	31-08-2020	M	14.87	33298644	47.48	East	22.9868	87.855
265	West Bengal	30-09-2020	M	9.35	35707239	47.73	East	22.9868	87.855
266	West Bengal	31-10-2020	M	9.98	33962549	45.63	East	22.9868	87.855

267 rows × 9 columns

In [15]:

```python
df2 = df2.rename(columns={df2.columns[0]:'State',df2.columns[3]:'EUR',df2.columns[4]:'EE', df2.columns[5]:'ELPR',
df2.head()
```

Out[15]:

	State	Date	Frequency	EUR	EE	ELPR	Region	longitude	latitude
0	Andhra Pradesh	31-01-2020	M	5.48	16635535	41.02	South	15.9129	79.74
1	Andhra Pradesh	29-02-2020	M	5.83	16545652	40.90	South	15.9129	79.74

2	Andhra Pradesh	31-03-2020	M	5.79	15881197	39.18	South	15.9129	79.74
3	Andhra Pradesh	30-04-2020	M	20.51	11336911	33.10	South	15.9129	79.74
4	Andhra Pradesh	31-05-2020	M	17.43	12988845	36.46	South	15.9129	79.74

In [16]:

```
df1.describe()
```

Out[16]:

	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)
count	740.000000	7.400000e+02	740.000000
mean	11.787946	7.204460e+06	42.630122
std	10.721298	8.087988e+06	8.111094
min	0.000000	4.942000e+04	13.330000
25%	4.657500	1.190404e+06	38.062500
50%	8.350000	4.744178e+06	41.160000
75%	15.887500	1.127549e+07	45.505000
max	76.740000	4.577751e+07	72.570000

In [17]:

```
df2.describe()
```

Out[17]:

	EUR	EE	ELPR	longitude	latitude
count	267.000000	2.670000e+02	267.000000	267.000000	267.000000
mean	12.236929	1.396211e+07	41.681573	22.826048	80.532425
std	10.803283	1.336632e+07	7.845419	6.270731	5.831738
min	0.500000	1.175420e+05	16.770000	10.850500	71.192400
25%	4.845000	2.838930e+06	37.265000	18.112400	76.085600
50%	9.650000	9.732417e+06	40.390000	23.610200	79.019300
75%	16.755000	2.187869e+07	44.055000	27.278400	85.279900
max	75.850000	5.943376e+07	69.690000	33.778200	92.937600

In [19]:

```
df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 267 entries, 0 to 266
Data columns (total 9 columns):
 #   Column      Non-Null Count  Dtype  
---  ------      --------------  -----  
 0   State       267 non-null    object 
 1    Date       267 non-null    object 
 2    Frequency  267 non-null    object 
 3   EUR         267 non-null    float64
 4   EE          267 non-null    int64  
 5   ELPR        267 non-null    float64
 6   Region      267 non-null    object 
 7   longitude   267 non-null    float64
 8   latitude    267 non-null    float64
dtypes: float64(4), int64(1), object(4)
memory usage: 18.9+ KB
```

In [18]:

```
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 7 columns):
 #   Column                                    Non-Null Count  Dtype  
---  ------                                    --------------  -----  
 0   Region                                    740 non-null    object 
 1    Date                                     740 non-null    object 
 2    Frequency                                740 non-null    object 
 3    Estimated Unemployment Rate (%)          740 non-null    float64
 4    Estimated Employed                       740 non-null    float64
 5    Estimated Labour Participation Rate (%)  740 non-null    float64
 6   Area                                      740 non-null    object 
dtypes: float64(3), object(4)
```

```
memory usage: 42.1+ KB
```

In [20]:

```
df1.isnull().sum()
```

Out[20]:

```
Region                                      28
 Date                                       28
 Frequency                                  28
 Estimated Unemployment Rate (%)            28
 Estimated Employed                         28
 Estimated Labour Participation Rate (%)    28
Area                                        28
dtype: int64
```

In [21]:

```
df2.isnull().sum()
```

Out[21]:

```
State         0
 Date         0
 Frequency    0
EUR           0
EE            0
ELPR          0
Region        0
longitude     0
latitude      0
dtype: int64
```

In [22]:

```
df2["State"].unique()
```

Out[22]:

```
array(['Andhra Pradesh', 'Assam', 'Bihar', 'Chhattisgarh', 'Delhi', 'Goa',
       'Gujarat', 'Haryana', 'Himachal Pradesh', 'Jammu & Kashmir',
       'Jharkhand', 'Karnataka', 'Kerala', 'Madhya Pradesh',
       'Maharashtra', 'Meghalaya', 'Odisha', 'Puducherry', 'Punjab',
       'Rajasthan', 'Sikkim', 'Tamil Nadu', 'Telangana', 'Tripura',
       'Uttar Pradesh', 'Uttarakhand', 'West Bengal'], dtype=object)
```

In [23]:

```
df2["Region"].unique()
```

Out[23]:

```
array(['South', 'Northeast', 'East', 'West', 'North'], dtype=object)
```

In [24]:

```
df2.groupby("Region").size()
```

Out[24]:

```
Region
East         40
North        79
Northeast    38
South        60
West         50
dtype: int64
```

In [26]:

```
region_statistics = df2.groupby(['Region'])[['EUR','EE','ELPR']].mean().reset_index()
region_statistics = round(region_statistics,2)
region_statistics
```

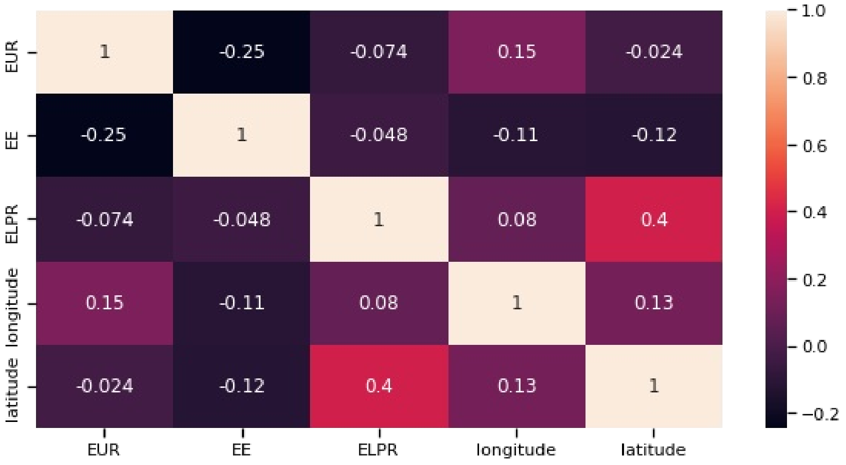
Out[26]:

	Region	EUR	EE	ELPR
0	East	13.92	19602366.90	40.11
1	North	15.89	13072487.92	38.70
2	Northeast	10.95	3617105.53	52.06
3	South	10.45	14040589.33	40.44
4	West	8.24	18623512.72	41.26

In [34]:

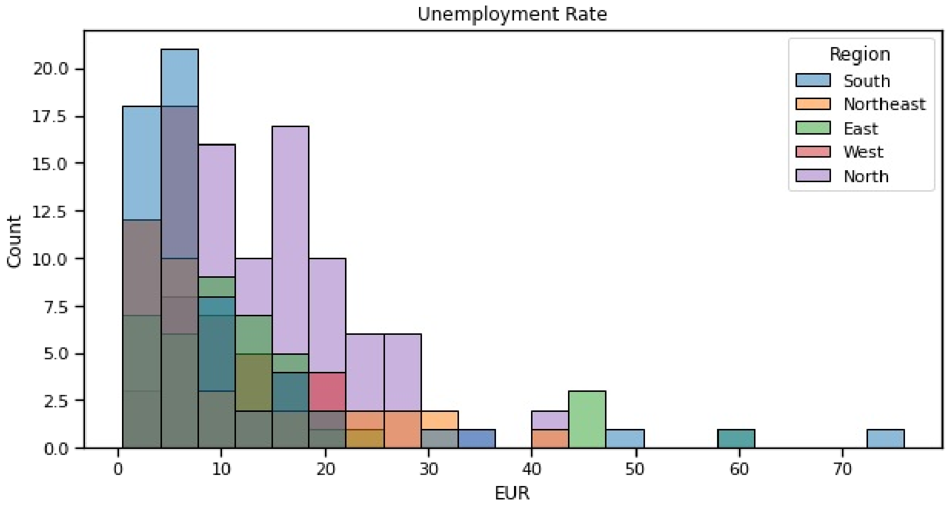
```
heat_maps = df2[['EUR','EE', 'ELPR','longitude', 'latitude']]
heat_maps = heat_maps.corr()

plt.figure(figsize=(10,5))
sns.set_context('notebook',font_scale=1)
sns.heatmap(heat_maps, annot=True);
```



In [35]:

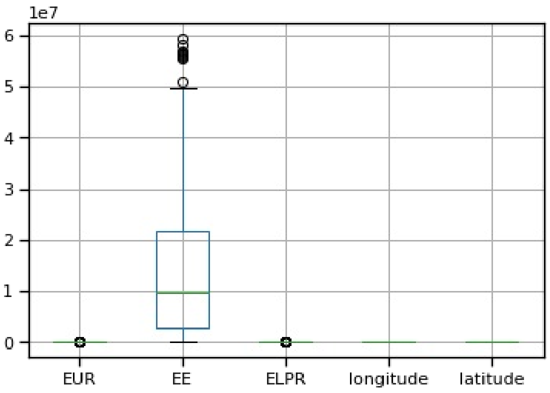
```
df2.columns= ["State","Date","Frequency","EUR","EE","ELPR","Region","longitude","latitude"]
plt.figure(figsize=(10, 5))
plt.title("Unemployment Rate")
sns.histplot(x="EUR", hue="Region", data=df2)
plt.show()
```



In [40]:

```
df2[["State","Date","Frequency","EUR","EE","ELPR","Region","longitude","latitude"]]
df2.boxplot()
```

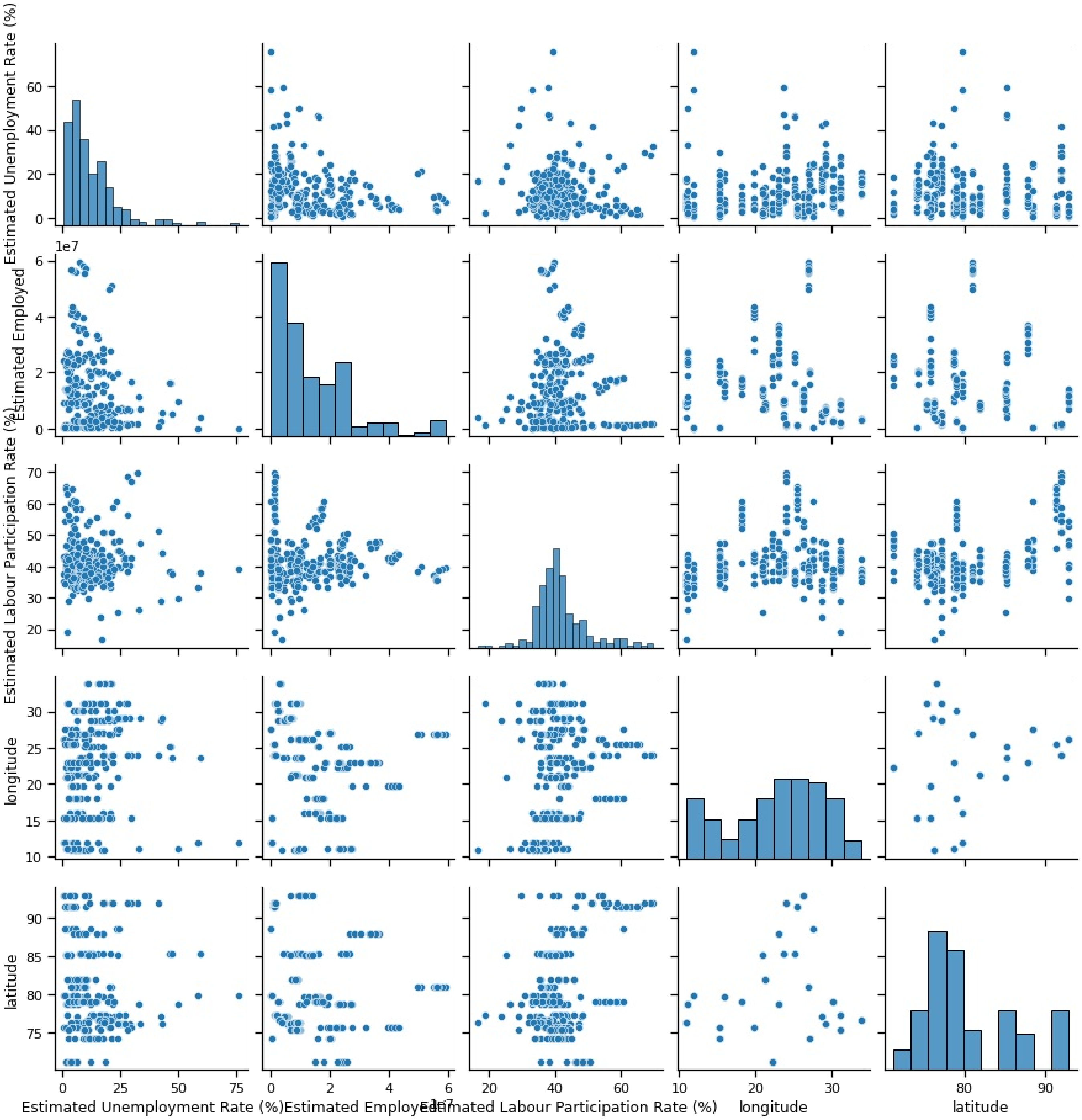
Out[40]: <AxesSubplot:>



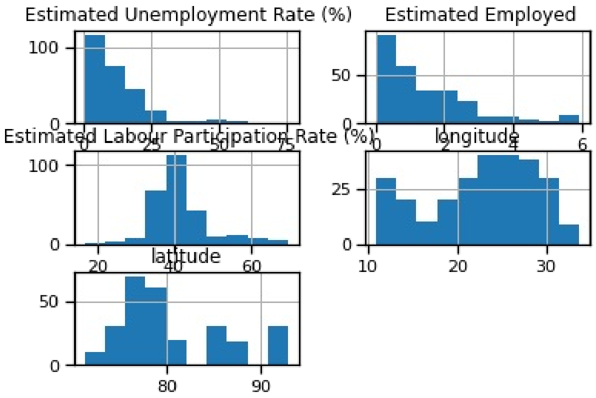
In [42]:

```
import seaborn as sns
sns.pairplot(df)
```

Out[42]:
```
<seaborn.axisgrid.PairGrid at 0x27575499670>
```



In [43]:
```
df.hist()
plt.show()
```



Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js