In [1]:

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import seaborn as sns
```

In [2]:

```
pwd
```

Out[2]:

```
'C:\\Users\\admin'
```

In [3]:

```
file_path=r"C:\Users\admin\Desktop\Pratham\temperatures.csv"
data=pd.read_csv(file_path)
```

In [4]:

```
data.describe()
```

Out[4]:

	YEAR	JAN	FEB	MAR	APR	MAY	JUN
count	117.000000	117.000000	117.000000	117.000000	117.000000	117.000000	117.000000
mean	1959.000000	23.687436	25.597863	29.085983	31.975812	33.565299	32.774274
std	33.919021	0.834588	1.150757	1.068451	0.889478	0.724905	0.633132
min	1901.000000	22.000000	22.830000	26.680000	30.010000	31.930000	31.100000
25%	1930.000000	23.100000	24.780000	28.370000	31.460000	33.110000	32.340000
50%	1959.000000	23.680000	25.480000	29.040000	31.950000	33.510000	32.730000
75%	1988.000000	24.180000	26.310000	29.610000	32.420000	34.030000	33.180000
max	2017.000000	26.940000	29.720000	32.620000	35.380000	35.840000	34.480000

In [5]:

```
data.head()
```

Out[5]:

	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	A
0	1901	22.40	24.14	29.07	31.91	33.41	33.18	31.21	30.39	30.47	29.97	27.31	24.49	
1	1902	24.93	26.58	29.77	31.78	33.73	32.91	30.92	30.73	29.80	29.12	26.31	24.04	
2	1903	23.44	25.03	27.83	31.39	32.91	33.00	31.34	29.98	29.85	29.04	26.08	23.65	
3	1904	22.50	24.73	28.21	32.02	32.64	32.07	30.36	30.09	30.04	29.20	26.36	23.63	
4	1905	22.00	22.83	26.68	30.01	33.32	33.25	31.44	30.68	30.12	30.67	27.52	23.82	

In [6]:

```
data.tail()
```

Out[6]:

	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
112	2013	24.56	26.59	30.62	32.66	34.46	32.44	31.07	30.76	31.04	30.27	27.83	25.37
113	2014	23.83	25.97	28.95	32.74	33.77	34.15	31.85	31.32	30.68	30.29	28.05	25.08
114	2015	24.58	26.89	29.07	31.87	34.09	32.48	31.88	31.52	31.55	31.04	28.10	25.67
115	2016	26.94	29.72	32.62	35.38	35.72	34.03	31.64	31.79	31.66	31.98	30.11	28.01
116	2017	26.45	29.46	31.60	34.95	35.84	33.82	31.88	31.72	32.22	32.29	29.60	27.18

In [7]:

```
type(data)
```

Out[7]:

```
pandas.core.frame.DataFrame
```

In [8]:

```
data.shape
```

Out[8]:

```
(117, 18)
```

In [9]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 117 entries, 0 to 116
Data columns (total 18 columns):
YEAR       117 non-null int64
JAN        117 non-null float64
FEB        117 non-null float64
MAR        117 non-null float64
APR        117 non-null float64
MAY        117 non-null float64
JUN        117 non-null float64
JUL        117 non-null float64
AUG        117 non-null float64
SEP        117 non-null float64
OCT        117 non-null float64
NOV        117 non-null float64
DEC        117 non-null float64
ANNUAL     117 non-null float64
JAN-FEB    117 non-null float64
MAR-MAY    117 non-null float64
JUN-SEP    117 non-null float64
OCT-DEC    117 non-null float64
dtypes: float64(17), int64(1)
memory usage: 16.5 KB
```

In [10]:

```
count=(data["JAN"]==22).sum()
```

In [11]:

```
count=(data["JAN"]==22).sum()
print(count)
```

```
1
```

In [12]:

```
column =data
count=column[column==0].count()
print(count)
```

```
YEAR       0
JAN        0
FEB        0
MAR        0
APR        0
MAY        0
JUN        0
JUL        0
AUG        0
SEP        0
OCT        0
NOV        0
DEC        0
ANNUAL     0
JAN-FEB    0
MAR-MAY    0
JUN-SEP    0
OCT-DEC    0
dtype: int64
```

In [13]:

```
data.isnull().sum()
```

Out[13]:

```
YEAR       0
JAN        0
FEB        0
MAR        0
APR        0
MAY        0
JUN        0
JUL        0
AUG        0
SEP        0
OCT        0
NOV        0
DEC        0
ANNUAL     0
JAN-FEB    0
MAR-MAY    0
JUN-SEP    0
OCT-DEC    0
dtype: int64
```

In [14]:

```
data.isnull().head()
```

Out[14]:

	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	AN
0	False	False	False	False	False	False	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	False	False	False	False	False	

In [15]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 117 entries, 0 to 116
Data columns (total 18 columns):
YEAR       117 non-null int64
JAN        117 non-null float64
FEB        117 non-null float64
MAR        117 non-null float64
APR        117 non-null float64
MAY        117 non-null float64
JUN        117 non-null float64
JUL        117 non-null float64
AUG        117 non-null float64
SEP        117 non-null float64
OCT        117 non-null float64
NOV        117 non-null float64
DEC        117 non-null float64
ANNUAL     117 non-null float64
JAN-FEB    117 non-null float64
MAR-MAY    117 non-null float64
JUN-SEP    117 non-null float64
OCT-DEC    117 non-null float64
dtypes: float64(17), int64(1)
memory usage: 16.5 KB
```

In [16]:

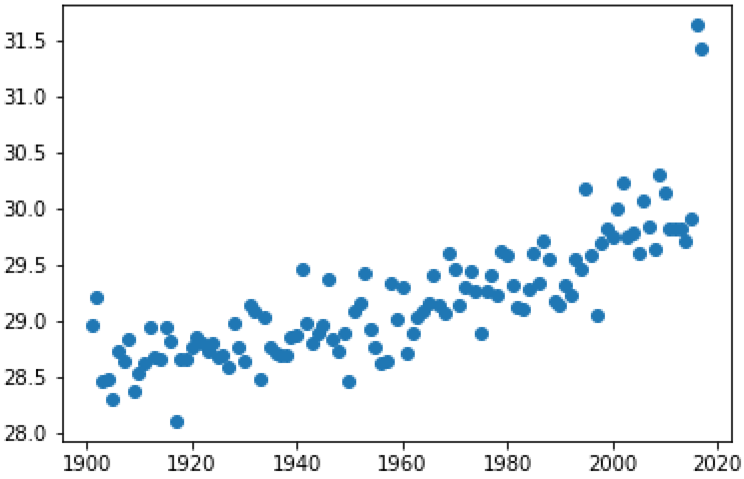
```
# x=data.iloc[:,1:6]
# y=data.iloc[:,-1:]
x=data["YEAR"]
y=data["ANNUAL"]
```

In [17]:

```
plt.plot(x,y,'o')
```

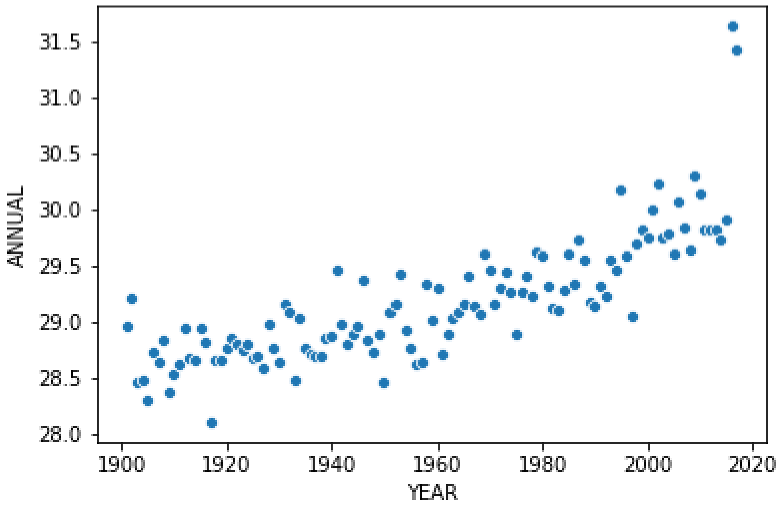
Out[17]:

```
[<matplotlib.lines.Line2D at 0x2622f628358>]
```



```
In [18]: sns.scatterplot(x=x,y=y)

Out[18]: <matplotlib.axes._subplots.AxesSubplot at 0x2622f5de860>
```



```
In [19]: type(x)

Out[19]: pandas.core.series.Series
```

```
In [20]: x.shape

Out[20]: (117,)
```

```
In [21]: x=x.x.values

         ---------------------------------------------------------------------------
         -
         AttributeError                            Traceback (most recent call las
         t)
         <ipython-input-21-b99b9b95d1cd> in <module>
         ----> 1 x=x.x.values

         C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\generic.py in __get
         attr__(self, name)
            5065             if self._info_axis._can_hold_identifiers_and_holds_nam
         e(name):
            5066                 return self[name]
         -> 5067             return object.__getattribute__(self, name)
            5068 
            5069     def __setattr__(self, name, value):

         AttributeError: 'Series' object has no attribute 'x'
```

```
In [22]: x=x.values
```

```
In [23]: x=x.reshape(117,1)
```

```
In [24]: x.shape

Out[24]: (117, 1)
```

In [25]:

```
type(x)
```

Out[25]:

```
numpy.ndarray
```

In [26]:

```
x_train,x_test,y_train,y_test =train_test_split(x,y,test_size=0.25)
print(f"x Training dataset: {x_train.shape}")
print(f"y Training dataset: {y_train.shape}")
print(f"x test dataset: {x_test.shape}")
print(f"y test dataset: {y_test.shape}")
```

```
x Training dataset: (87, 1)
y Training dataset: (87,)
x test dataset: (30, 1)
y test dataset: (30,)
```

In [27]:

```
model = LinearRegression()
```

In [28]:

```
model.fit(x_train,y_train)
```

Out[28]:

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
         normalize=False)
```

In [29]:

```
model.coef_ #w
```

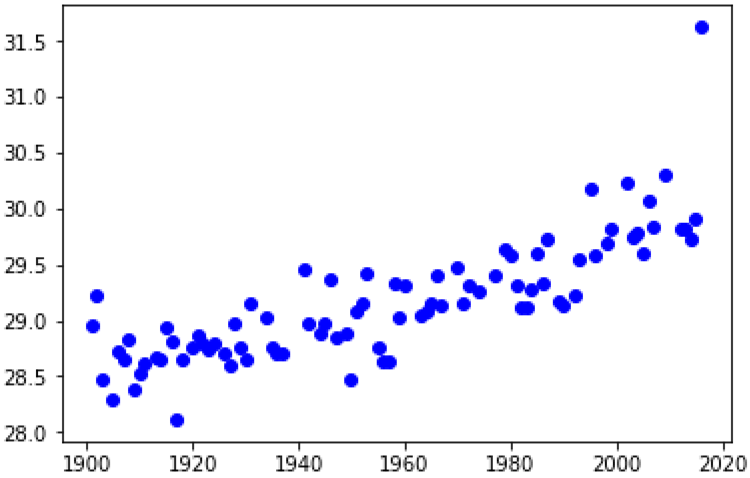
Out[29]:

```
array([0.01271217])
```

In [30]:
```
plt.scatter(x_train, y_train, color='blue')
plt.plot(x_test, y_pred, color='red', linewidth=3)
plt.title("Temperature vs Year")
plt.xlabel("Year")
plt.ylabel("Temperature")
plt.show()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-30-2def755fd8e0> in <module>
      1 plt.scatter(x_train, y_train, color='blue')
----> 2 plt.plot(x_test, y_pred, color='red', linewidth=3)
      3 plt.title("Temperature vs Year")
      4 plt.xlabel("Year")
      5 plt.ylabel("Temperature")

NameError: name 'y_pred' is not defined
```



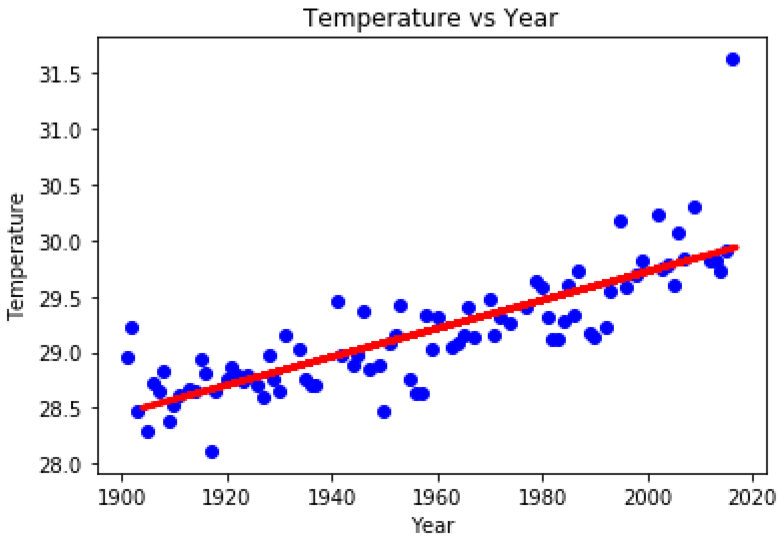
In [31]:
```
y_pred = model.predict(x_test)
```

In [32]:
```
y_pred.shape
```

Out[32]:
```
(30,)
```

In [33]:

```
plt.scatter(x_train, y_train, color='blue')
plt.plot(x_test, y_pred, color='red', linewidth=3)
plt.title("Temperature vs Year")
plt.xlabel("Year")
plt.ylabel("Temperature")
plt.show()
```



In [34]:

```
sns.regplot(data=df,x=x_train,y=y_train,)
```

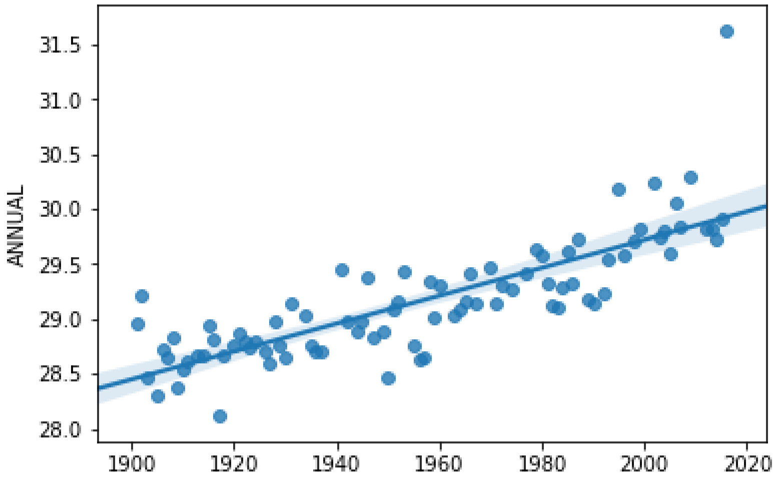
```
---------------------------------------------------------------------------
-
NameError                                 Traceback (most recent call las
t)
<ipython-input-34-9042aa249480> in <module>
----> 1 sns.regplot(data=df,x=x_train,y=y_train,)

NameError: name 'df' is not defined
```

In [35]:

Out[35]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2622fd0aac8>
```



In [36]:

```
from sklearn.metrics import mean_absolute_error,mean_squared_error,r2_score
print(f"MSE:  {mean_squared_error(y_test,y_pred)}")
print(f"MAE:  {mean_absolute_error(y_test,y_pred)}")
print(f"R-Sqaure :  {r2_score(y_test,y_pred)}")
```

```
MSE:  0.14443462476212654
MAE:  0.2644742598743015
R-Sqaure :  0.5979633222180765
```

In [ ]: