In [4]:

```
data.head(10)
```

Out[4]:

	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	AI
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	
5	1906	22.28	23.69	27.31	31.93	34.11	32.19	31.01	30.30	29.92	29.55	27.60	24.72	
6	1907	24.46	24.01	27.04	31.79	32.68	31.92	31.05	29.58	30.67	29.87	27.78	24.44	
7	1908	23.57	25.26	28.86	32.42	33.02	33.12	30.61	29.55	29.59	29.35	26.88	23.73	
8	1909	22.67	24.36	29.22	30.79	33.06	31.70	29.81	29.81	30.06	29.25	27.69	23.69	
9	1910	23.24	25.16	28.48	31.42	33.51	31.84	30.42	29.86	29.82	28.91	26.32	23.37	

In [6]:

```
x=data.drop('ANNUAL',axis=1)
y=data['ANNUAL']
print("shape of x",x.shape)
print("shape of y",y.shape)
```

```
shape of x (117, 17)
shape of y (117,)
```

In [12]:

```
len(data.axes[0])
```

Out[12]:

```
117
```

In [13]:

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=51)
print("shape of x",x.shape)
print("shape of y",y.shape)
```

```
shape of x (117, 17)
shape of y (117,)
```

In [14]:

```
print("shape of x_test",x_test.shape)
```

```
shape of x_test (24, 17)
```

In [19]:

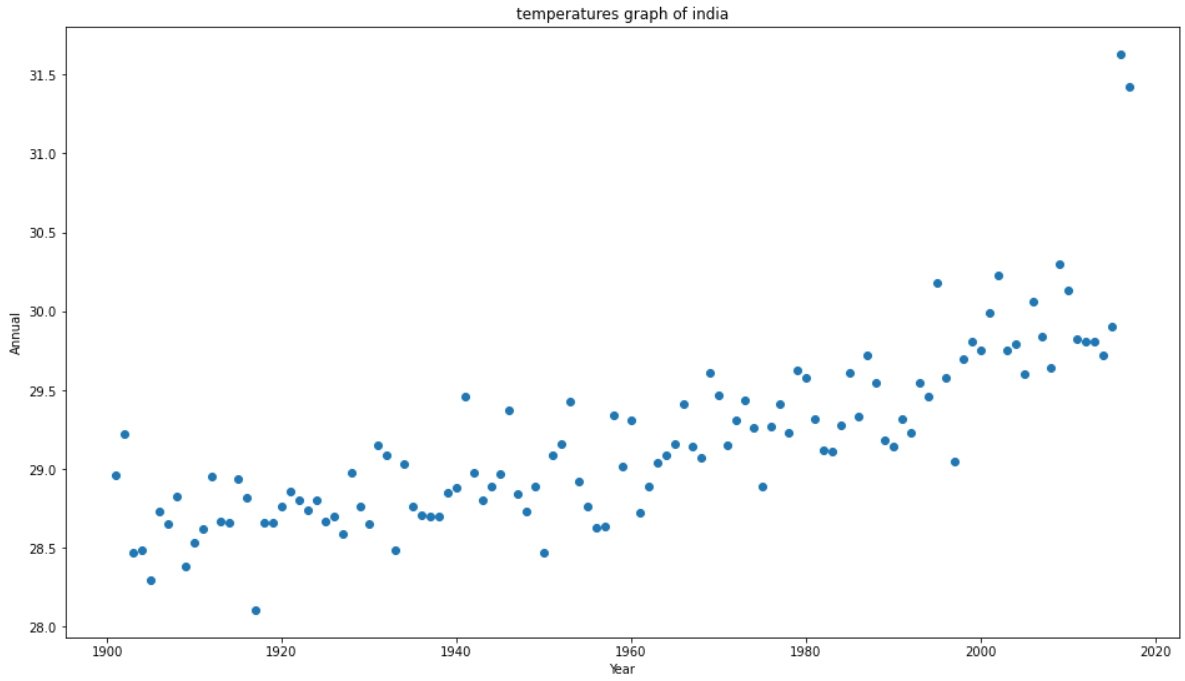
```
import matplotlib.pyplot as plt
```

In [26]:

```
x=data['YEAR']
y=data['ANNUAL']

plt.figure(figsize=(16,9))
plt.title("temperatures graph of india")
plt.xlabel('Year')
plt.ylabel('Annual')
plt.scatter(x,y)
```

Out[26]:

```
<matplotlib.collections.PathCollection at 0x24a7c13fdf0>
```



In [27]:

```
x=x.values
```

In [28]:

```
x
```

Out[28]:

```
array([1901, 1902, 1903, 1904, 1905, 1906, 1907, 1908, 1909, 1910, 1911,
       1912, 1913, 1914, 1915, 1916, 1917, 1918, 1919, 1920, 1921, 1922,
       1923, 1924, 1925, 1926, 1927, 1928, 1929, 1930, 1931, 1932, 1933,
       1934, 1935, 1936, 1937, 1938, 1939, 1940, 1941, 1942, 1943, 1944,
       1945, 1946, 1947, 1948, 1949, 1950, 1951, 1952, 1953, 1954, 1955,
       1956, 1957, 1958, 1959, 1960, 1961, 1962, 1963, 1964, 1965, 1966,
       1967, 1968, 1969, 1970, 1971, 1972, 1973, 1974, 1975, 1976, 1977,
       1978, 1979, 1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988,
       1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999,
       2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010,
       2011, 2012, 2013, 2014, 2015, 2016, 2017], dtype=int64)
```

In [29]:

```
x=x.reshape(117,1)
```

In [30]:

```
x
```

```
       [1998],
       [1999],
       [2000],
       [2001],
       [2002],
       [2003],
       [2004],
       [2005],
       [2006],
       [2007],
       [2008],
       [2009],
       [2010],
       [2011],
       [2012],
       [2013],
       [2014],
       [2015],
       [2016],
       [2017]], dtype=int64)
```

In [32]:

```
from sklearn.linear_model import LinearRegression
```

In [33]:

```
regressor=LinearRegression()
```

In [34]:

```
regressor.fit(x,y)
```

Out[34]:

```
LinearRegression()
```

In [35]:

```
regressor.coef_
```

Out[35]:

```
array([0.01312158])
```

In [36]:

```
regressor.intercept_
```

Out[36]:

```
3.4761897126187016
```

In [40]:

```
regressor.predict([[5439]])
```

Out[40]:

```
array([74.84447052])
```

In [41]:

```
predicted=regressor.predict(x)
```

In [43]:

```
predicted
```

Out[43]:

```
array([28.4203158 , 28.43343739, 28.44655897, 28.45968055, 28.47280213,
       28.48592371, 28.49904529, 28.51216687, 28.52528846, 28.53841004,
       28.55153162, 28.5646532 , 28.57777478, 28.59089636, 28.60401794,
       28.61713952, 28.63026111, 28.64338269, 28.65650427, 28.66962585,
       28.68274743, 28.69586901, 28.70899059, 28.72211218, 28.73523376,
       28.74835534, 28.76147692, 28.7745985 , 28.78772008, 28.80084166,
       28.81396324, 28.82708483, 28.84020641, 28.85332799, 28.86644957,
       28.87957115, 28.89269273, 28.90581431, 28.91893589, 28.93205748,
       28.94517906, 28.95830064, 28.97142222, 28.9845438 , 28.99766538,
       29.01078696, 29.02390855, 29.03703013, 29.05015171, 29.06327329,
       29.07639487, 29.08951645, 29.10263803, 29.11575961, 29.1288812 ,
       29.14200278, 29.15512436, 29.16824594, 29.18136752, 29.1944891 ,
       29.20761068, 29.22073227, 29.23385385, 29.24697543, 29.26009701,
       29.27321859, 29.28634017, 29.29946175, 29.31258333, 29.32570492,
       29.3388265 , 29.35194808, 29.36506966, 29.37819124, 29.39131282,
       29.4044344 , 29.41755599, 29.43067757, 29.44379915, 29.45692073,
       29.47004231, 29.48316389, 29.49628547, 29.50940705, 29.52252864,
       29.53565022, 29.5487718 , 29.56189338, 29.57501496, 29.58813654,
       29.60125812, 29.6143797 , 29.62750129, 29.64062287, 29.65374445,
       29.66686603, 29.67998761, 29.69310919, 29.70623077, 29.71935236,
       29.73247394, 29.74559552, 29.7587171 , 29.77183868, 29.78496026,
       29.79808184, 29.81120342, 29.82432501, 29.83744659, 29.85056817,
       29.86368975, 29.87681133, 29.88993291, 29.90305449, 29.91617608,
       29.92929766, 29.94241924])
```

In [44]:

```
#mean absolute error
y-predicted
```

Out[44]:

```
0      0.539684
1      0.786563
2      0.023441
3      0.030319
4     -0.172802
         ...
112   -0.079933
113   -0.183054
114   -0.016176
115    1.700702
116    1.477581
Name: ANNUAL, Length: 117, dtype: float64
```

In [46]:

```
np.mean(abs(y-predicted))
```

Out[46]:

```
0.22535284978630413
```

In [47]:

```
from sklearn.metrics import mean_absolute_error
mean_absolute_error(y,predicted)
```

Out[47]:

```
0.22535284978630413
```

In [48]:

```
mean_squared_error(y,predicted)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Input In [48], in <cell line: 1>()
----> 1 mean_squared_error(y,predicted)

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

In [49]:

```
from sklearn.metrics import mean_squared_error
mean_squared_error(y,predicted)
```

Out[49]:

```
0.10960795229110352
```

In [2]:

```
#R square metrics
from sklearn.metrics import r2_score
r2_score(y,predicted)
```

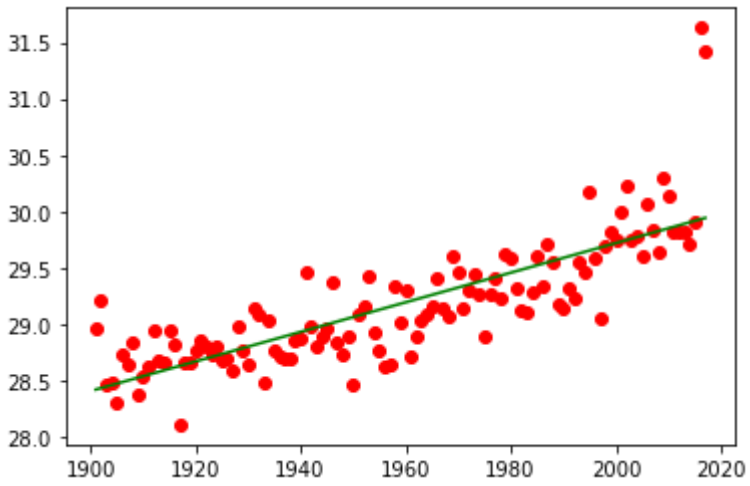
```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_11140/4220044763.py in <module>
      1 #R square metrics
      2 from sklearn.metrics import r2_score
----> 3 r2_score(y,predicted)

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

In [54]:

```
plt.scatter(x,y,label='actual',color='r')
plt.plot(x,predicted,label="predicted",color='g')
```

Out[54]:

```
[<matplotlib.lines.Line2D at 0x24a7c8fc5e0>]
```



In [55]:

```
regressor.predict([[5678]])
```

Out[55]:

```
array([77.98052846])
```

In [ ]:

```

```