In [5]:

```
import pandas as pd
from sklearn import linear_model
import matplotlib.pyplot as plt
```

In [4]:

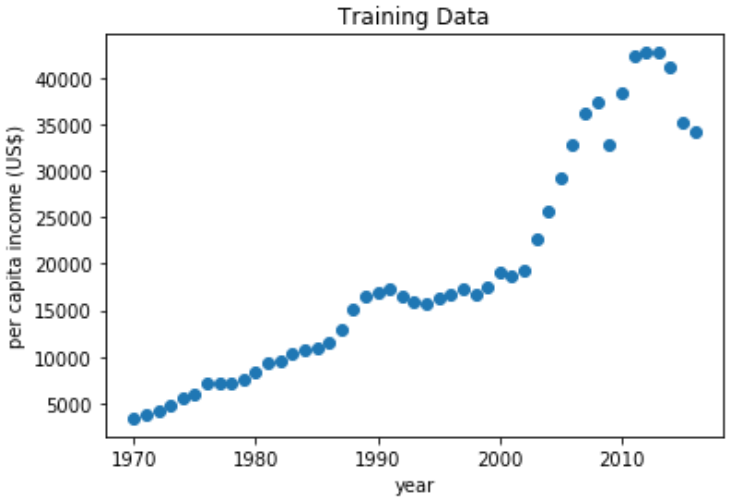
```
data = pd.read_csv("per_capita_income.csv")
data.head()
```

Out[4]:

	year	per capita income (US$)
0	1970	3399.299037
1	1971	3768.297935
2	1972	4251.175484
3	1973	4804.463248
4	1974	5576.514583

In [20]:

```
plt.scatter(data.year, data['per capita income (US$)'])
plt.xlabel("year")
plt.ylabel("per capita income (US$)")
plt.title("Training Data")
plt.savefig("training_data.png")
```



In [10]:

```
lin_reg_model = linear_model.LinearRegression()
```

In [11]:

```
lin_reg_model.fit(data[['year']], data['per capita income (US$)'])
```

Out[11]:

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

In [12]:

```
lin_reg_model.coef_
```

Out[12]:

```
array([828.46507522])
```

In [14]:

```
lin_reg_model.intercept_
```

Out[14]:

```
-1632210.7578554575
```

In [15]:

```
lin_reg_model.predict([[2020]])
```

Out[15]:

```
array([41288.69409442])
```

In [21]:

```
plt.scatter(data.year, data['per capita income (US$)'])
plt.xlabel("year")
plt.ylabel("per capita income (US$)")
plt.plot(data.year, lin_reg_model.predict(data[['year']]))
plt.title("Showing Regression Line")
plt.savefig("lin_reg.png")
```