```
In [40]:    import pandas as pd
            from sklearn.model_selection import train_test_split
            from sklearn.linear_model import LinearRegression
            from sklearn.metrics import r2_score
            import matplotlib.pyplot as plt
            %matplotlib inline
```
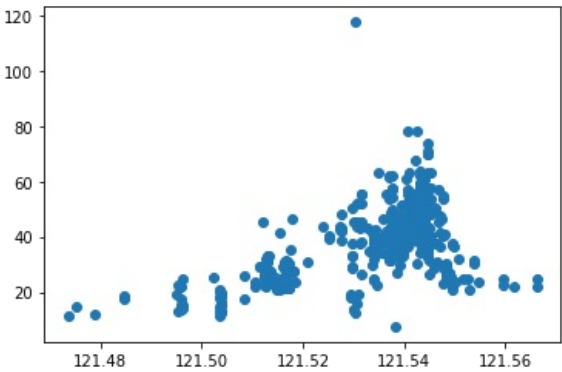
```
In [41]:    data = pd.read_csv('files/house_prices.csv')
            data.head()
```

Out[41]:

	Transaction	House age	Distance to MRT station	Number of convenience stores	Latitude	Longitude	House unit price
0	2012.917	32.0	84.87882	10	24.98298	121.54024	37.9
1	2012.917	19.5	306.59470	9	24.98034	121.53951	42.2
2	2013.583	13.3	561.98450	5	24.98746	121.54391	47.3
3	2013.500	13.3	561.98450	5	24.98746	121.54391	54.8
4	2012.833	5.0	390.56840	5	24.97937	121.54245	43.1

```
In [42]:    fig, ax = plt.subplots()

            ax.scatter(x=data['Longitude'], y=data['House unit price'])
```

Out[42]:   <matplotlib.collections.PathCollection at 0x7fdc900687f0>


```
In [43]:    data.corr()
```

Out[43]:

	Transaction	House age	Distance to MRT station	Number of convenience stores	Latitude	Longitude	House unit price
Transaction	1.000000	0.017549	0.060880	0.009635	0.035058	-0.041082	0.087491
House age	0.017549	1.000000	0.025622	0.049593	0.054420	-0.048520	-0.210567
Distance to MRT station	0.060880	0.025622	1.000000	-0.602519	-0.591067	-0.806317	-0.673613
Number of convenience stores	0.009635	0.049593	-0.602519	1.000000	0.444143	0.449099	0.571005
Latitude	0.035058	0.054420	-0.591067	0.444143	1.000000	0.412924	0.546307
Longitude	-0.041082	-0.048520	-0.806317	0.449099	0.412924	1.000000	0.523287
House unit price	0.087491	-0.210567	-0.673613	0.571005	0.546307	0.523287	1.000000

```
In [44]:    data.isnull().sum()
```

Out[44]:   Transaction   0
           House age   0
           Distance to MRT station   0
           Number of convenience stores   0
           Latitude   0
           Longitude   0
           House unit price   0
           dtype: int64

```
In [45]:    X = data.iloc[:,:-1]
```

```
y = data.iloc[:,-1]
```

In [46]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=0, test_size=.15)
```

In [47]:
```
lin = LinearRegression()
lin.fit(X_train, y_train)
```

Out[47]:
```
LinearRegression()
```

In [48]:
```
y_pred = lin.predict(X_test)
```

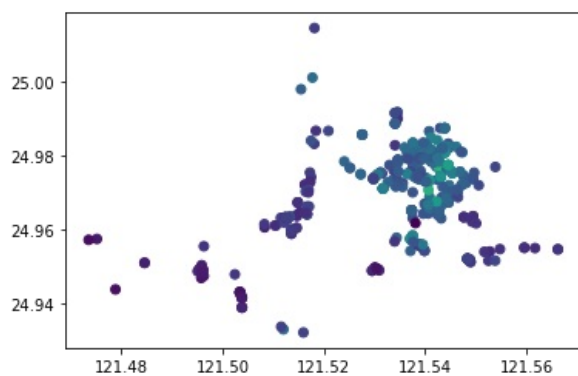In [49]:
```
r2_score(y_test, y_pred)
```

Out[49]:
```
0.680008800201434
```

In [50]:
```
lin.coef_
```

Out[50]:
```
array([ 4.67857744e+00, -2.58421926e-01, -4.28607244e-03,  1.21784177e+00,
        2.23558483e+02, -1.45002564e+00])
```

In [51]:
```
fig, ax = plt.subplots()

ax.scatter(x=X['Longitude'], y=X['Latitude'], c=data['House unit price'])
```

Out[51]:
```
<matplotlib.collections.PathCollection at 0x7fdc9008a850>
```


In [ ]: