In [22]:
```
import geopandas as gpd
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.pylab as pylab
import pylab as mpl  #导入中文字体，避免显示乱码
mpl.rcParams['font.sans-serif']=['SimHei']  #设置为黑体字
```

In [21]:
```
## 读取数据
selected_heights = gpd.read_file("outputs/selected_heights.geojson")
```

手肘法确定K值

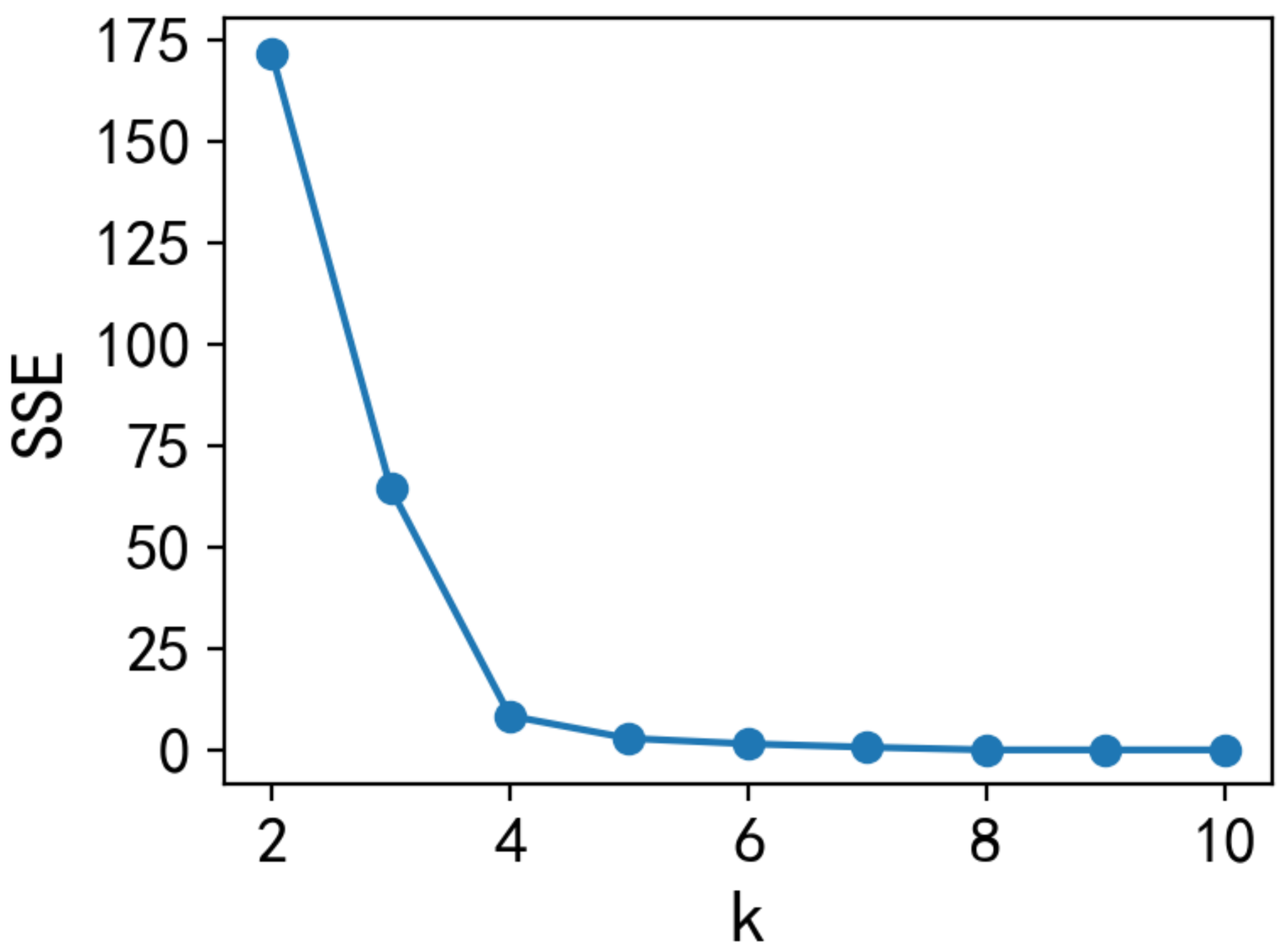
In [33]:
```
from sklearn.cluster import KMeans
d=[]
df = selected_heights['Floor'].values.reshape(-1,1)
for i in range(2,11):    #k取值1~10，做kmeans聚类，看不同k值对应的簇内误差平方和
    km=KMeans(n_clusters=i)
    km.fit(df)
    d.append(km.inertia_)  #inertia簇内误差平方和

#生成figure对象
plt.figure(figsize = (4,3), dpi = 200)
plt.plot(range(2,11),d,marker='o')
plt.xlabel('k',fontsize = 16)
plt.ylabel('SSE',fontsize = 16)
plt.xticks(fontsize = 14)
plt.yticks(fontsize = 14)

plt.show()
```

```
C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:6: ConvergenceWarning: Number of distinct clusters (8) found smaller than n_cluste
rs (9). Possibly due to duplicate points in X.

C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:6: ConvergenceWarning: Number of distinct clusters (8) found smaller than n_cluste
rs (10). Possibly due to duplicate points in X.
```



由上图可知，k的取值为4

k-means聚类

In [35]:
```
k = 4

# 构建模型，随机种子设为123
kmeans_model = KMeans(n_clusters = k,random_state=123)
fit_kmeans = kmeans_model.fit(df)  # 模型训练
# 查看聚类结果
kmeans_cc = kmeans_model.cluster_centers_  # 聚类中心
print('各类聚类中心为：\n',kmeans_cc)
kmeans_labels = kmeans_model.labels_  # 样本的类别标签
print('各样本的类别标签为：\n',kmeans_labels)
r1 = pd.Series(kmeans_model.labels_).value_counts()  # 统计不同类别样本的数目
print('最终每个类别的数目为：\n',r1)
```

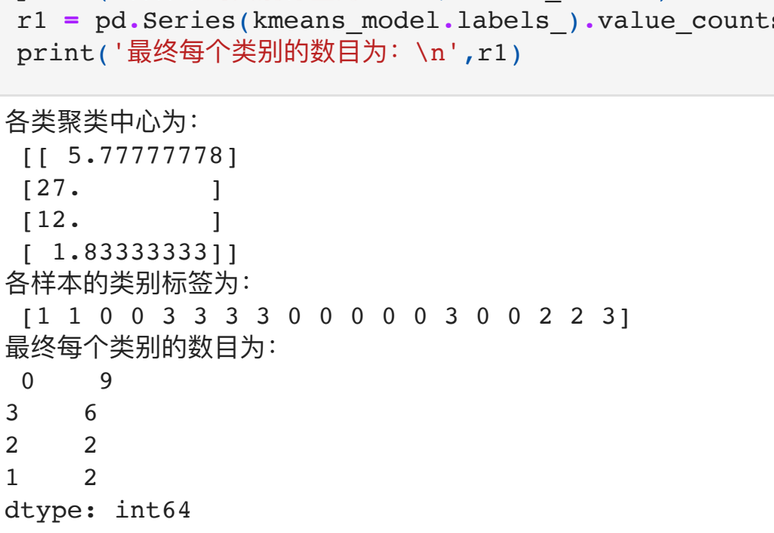
```
各类聚类中心为：
 [[ 5.77777778]
 [27.        ]
 [12.        ]
 [ 1.83333333]]
各样本的类别标签为：
 [1 1 0 0 3 3 3 3 0 0 0 0 0 3 0 0 2 2 3]
最终每个类别的数目为：
 0    9
3    6
2    2
1    2
dtype: int64
```

In [40]:
```
selected_heights['label'] = kmeans_labels
```

In [42]:
```
## 展示聚类结果
selected_heights.plot(column="label")
```

```
C:\ProgramData\Anaconda3\lib\site-packages\matplotlib\backends\backend_agg.py:211: RuntimeWarning: Glyph 8722 missing from current font.
  font.set_text(s, 0.0, flags=flags)
C:\ProgramData\Anaconda3\lib\site-packages\matplotlib\backends\backend_agg.py:180: RuntimeWarning: Glyph 8722 missing from current font.
  font.set_text(s, 0, flags=flags)
```

Out[42]:
```
<matplotlib.axes._subplots.AxesSubplot at 0xb124400>
```



In [ ]: