**实验报告**

1. 算法描述：

1.1把原始数据按照列的方式排列成50\*23的形式

1.2把这个ndarray零均值化，即每一行减去这一行的均值

1.3算出协方差矩阵C=X\*X’

1.4求出来协方差矩阵C的特征值和特征向量，并取最大的两个特征值所对应的特征向量，即最后两个，组成ndarry，并且命名为P

1.5 Y=P\*X，此时Y就是降到2维之后的矩阵

1. 实验结果

2.1遇到的问题和解决方法：

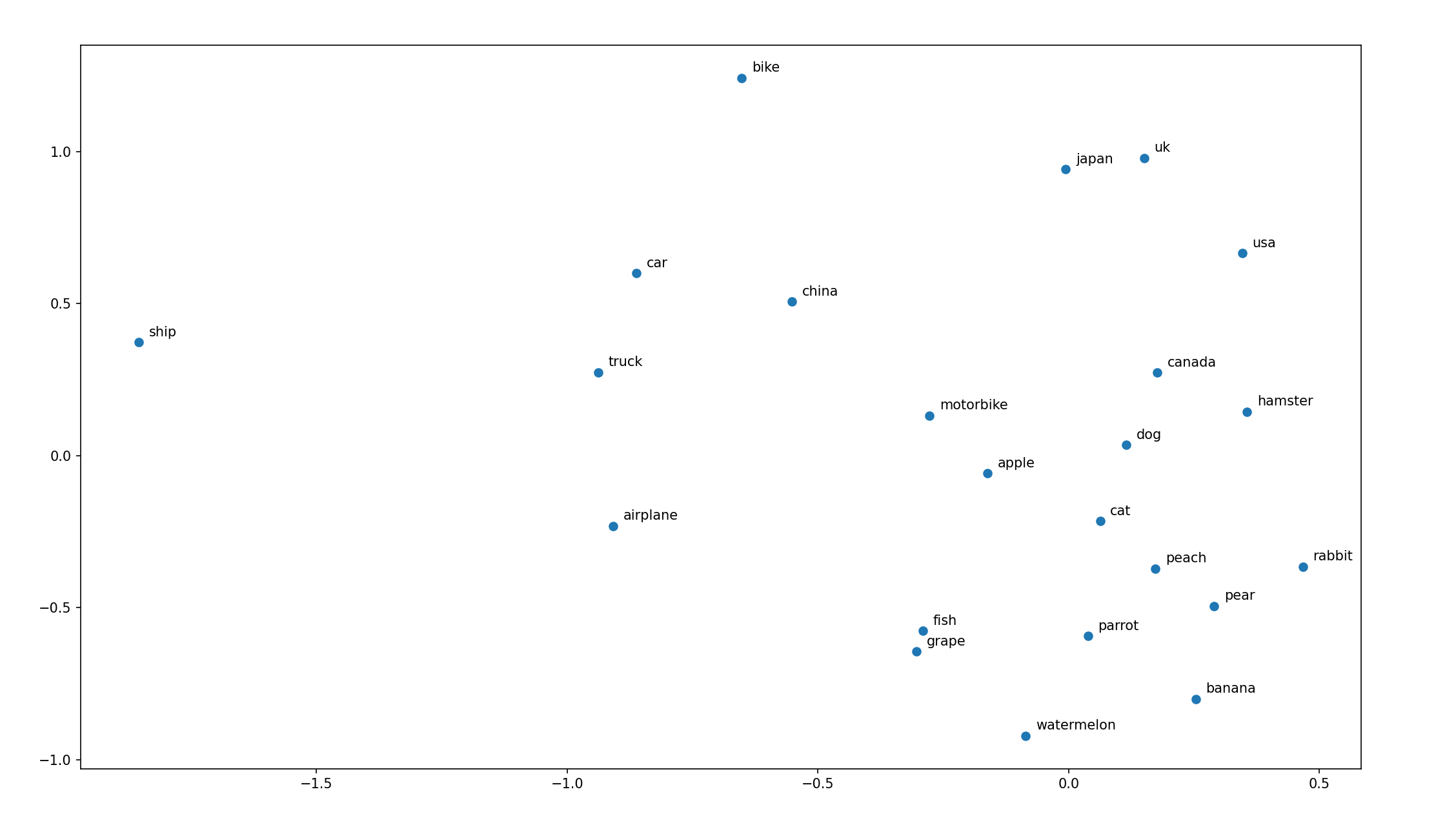
原因：

因为china没有作为index在dataframe中存在，而是作为了其中一个column  
文件标头名是附加的自定义名称，但是会发现，原来的标头名（列标签名）并没有被删除，此时可以使用header 参数来删除它。

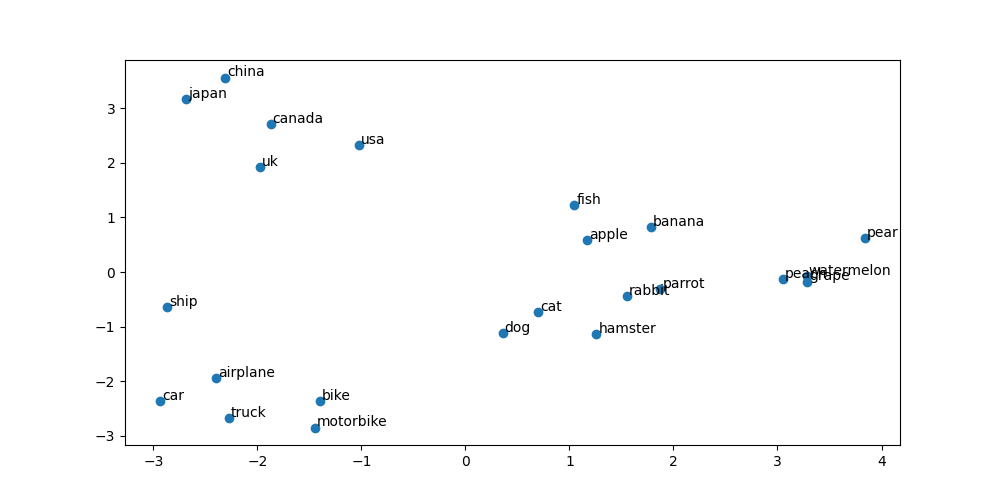
解决：  
让header=None》》没有标头了！。也可以显示设置标头为names，来设置自定义的Ecolumns

2.2实验结果

自己实现的PCA



调用sklern实现的PCA：



发现自己实现的PCA算法仍有许多不足，存在相似的几个单词相去甚远。但sklearn库中实现的就极为准确。