首先，将数据文件进行爬取，存在预设文件夹中。使用程序对每个文件夹中的每个json文件进行坐标个数的统计，将其中时间过长与过短的数据进行删除处理。

运行后，可以看到各个文件中坐标的组数，由于计数的频率是5HZ，故将组数除以5即可得到每个人完成问卷所花的时间，虽然每个人完成的时间各有不同，但过短与过长的时间显然不合常理，经过检查后，决定将做题时间短于5分钟，长于40分钟的数据进行删除处理。 代码中用数据长度除以40再除以5来实现。

经过初步处理后，利用plot函数，对每个数据进行折线图的绘制，由于xyz坐标表示手机在xyz方向上的加速度，因此，如果发现折线图的波动十分小，则说明做题者可能在做题时将手机放于桌子上，因此这类的数据应予以删除处理。使用的代码如下：

import os

import pandas as pd

import numpy as np

from matplotlib import pyplot as plt

dp = os.listdir("E:\\python\_src\\data\\gyroscope\_anxiety\_female")

for a in dp:

ndp=os.path.join("E:\\python\_src\\data\\gyroscope\_anxiety\_female",a)

d=pd.read\_json(ndp)

d.plot()

time = d.iloc[:,0].size // (5\*40)

if 5 < time < 40:

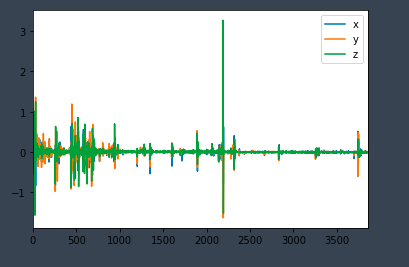
d = d.iloc[500:]

var = d['x'].var()

if var > 0.01:

print(ndp)

此外，部分数据的折线图呈现前期变化很大，在一次较大的波动后趋于平稳，比如：



正常的，可采取的折线图，比如：

