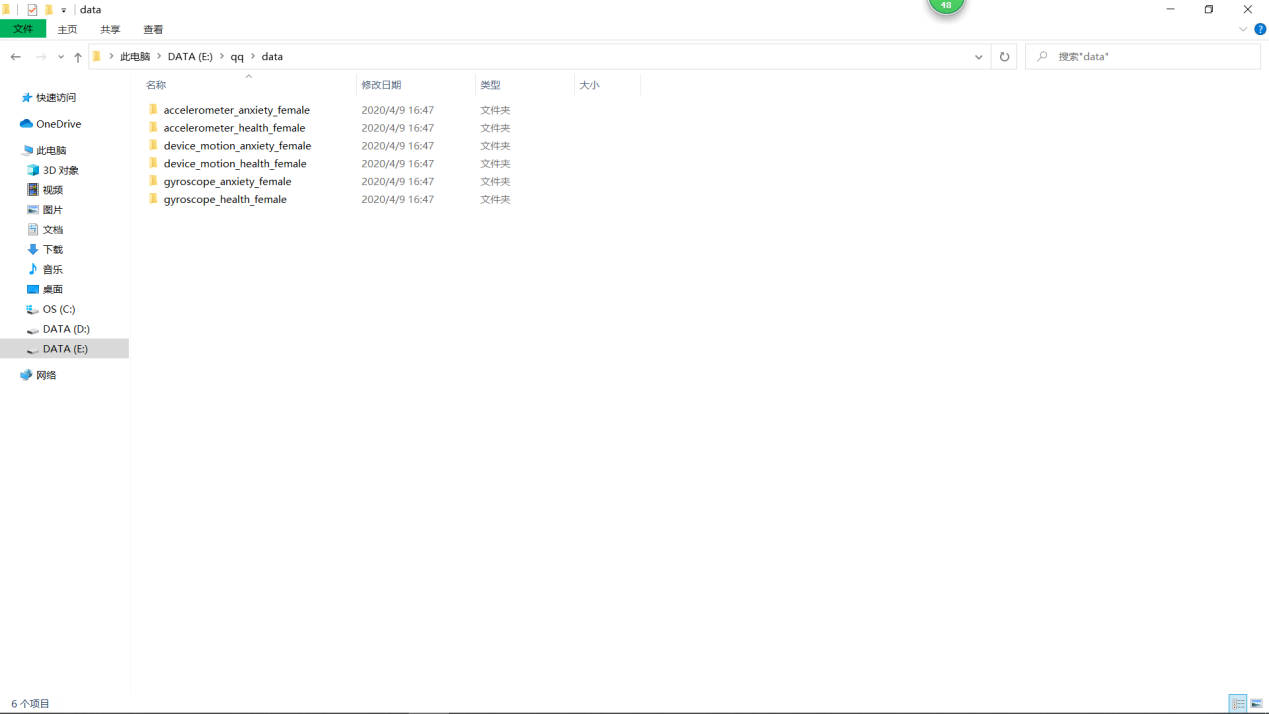
首先，将数据文件全部下载，并按照性别与是否患有焦虑症等分为6类，存在文件夹中。



随后，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。

通过Anaconda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7分钟，长于60分钟的数据进行删除处理。 代码中用数据长度除以60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：（仅仅为了显示折线图，未进行筛选）

import os

import pandas as pd

import numpy as np

from matplotlib import pyplot as plt

pathdir = os.listdir("E:\\qq\\data\\gyroscope\_anxiety\_female")

for s in pathdir:

newdir=os.path.join("E:\\qq\\data\\gyroscope\_anxiety\_female",s)

d=pd.read\_json(newdir)

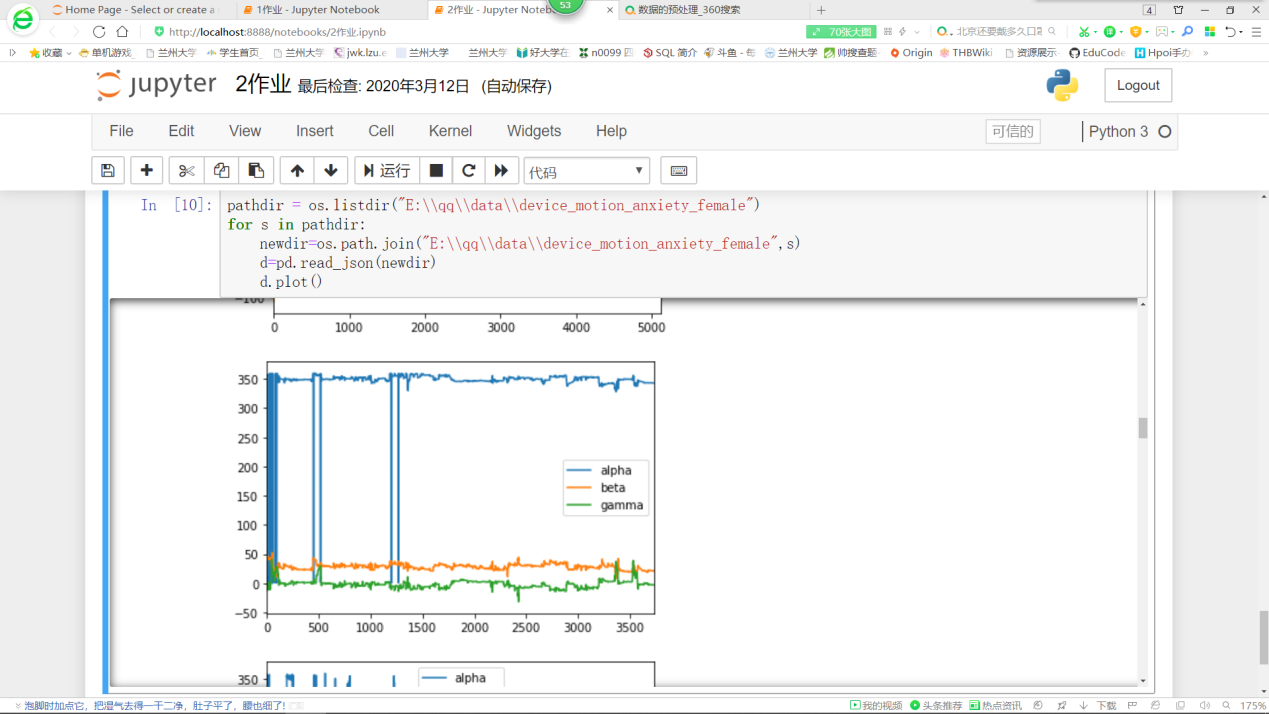
d.plot()

此外，部分数据的折现表呈现前期变化很大，在一次大的波动后趋于平稳，比如：



这类数据也应予以删除，可能是做题者一开始站着做题，手机有所晃动，但后面坐下了，或将手机放在了桌子上。

另外，一些数据明显只在x轴方向上有剧烈的波动，而在yz轴上几乎没有波动，如这种数据：



这种数据也应予以删除。

在代码部分，由于数据的异常情况较多且复杂，故采取对x轴部分取方差的方式来对合适的数据进行筛选。当在x上的方差小于0.01时，数据予以删除，故打印出方差大于0.01的文件列表。