获得数据以后，我删除了其中无用的的空文件，之后我从了两种不同的角度对数据进行了预处理。

第一种：

通过观察发现，一共有三种传感器，有的人的数据只有其中两种或一种，也有一部分人三种传感器的数据都被采集到了。于是我只保留了三种传感器都具备的数据，然后通过perprocessing.py对这些数据进行了处理。在代码中，我将数据中的缺失值用0代替，之后我通过公式（（MAX（三种传感器采集的数据数目）/5）/60）计算出了一个人的可能的答题时间，之后我保留了答题时间处于5min-50min的数据。

缺陷：操作过于繁琐，需要手动输入文件名，且需要人工判断时间是否处于5-50，无法处理大批量的数据，所获得的时间无法存储下来。

第二种：

我保留了除空文件以外的所有数据，之后将它们按照传感器种类\_健康状况分别存储在了6个文件夹中。然后我在perprocessing2.py中对每个文件夹中的数据进行了预处理。删除了所有数据中答题时间不在5min-50min这之间的数据文件。将所有的缺失数据用0代替，计算出了每个人被每种传感器记录的时间，以列表的形式打印了出来。

缺陷：计算出的时间只能按照不同的传感器计算，可能会出现同一个人的时间不同的问题。