**加速器数据预处理说明文档**

**代码1：获取数据**

1. 通过index.txt获取到所有数据的下载链接
2. 通过观察数据，发现数据为以字典为元素的列表，可以直接保存为json格式以便读取
3. 通过for循环遍历所有数据进行储存

**代码2：数据预处理**

1. 用户输入需要进行处理的文件名
2. 通过os.listdir()获取到所有数据文件名，判断用户输入数据文件是否在其中，若不存在，直接退出
3. 观察数据可知，数据中存在空文件，若文件过小（为空），直接进行删除并退出
4. 将每列的空值（缺失值）总和打印出来，供用户判断采取何种填充措施
5. 通过展示每个坐标轴的盒图，可以看出没个坐标轴的特殊点，供用户判断这些特殊点的实际意义
6. 因采样率为5Hz，可将行数除以5再除以60算出数据总时间，通过答题时长判断数据真实性
7. 通过各坐标轴方差大小，可判断手机是否手持（方差是否过小），若方差过小，给予提示
8. 通过前25%数据与后75%数据方差之差，可以判断用户答题过程中的状态变化，若差值过大，给予提示

注：每一步过程在代码中都有注释，可以结合代码进行阅读