我的可视化课题是“古典钢琴曲的数据分析”。选定这个题目后，我就在想两个问题：一是我最后展示的数据能否引起人们的兴趣，第二，对音乐进行数据处理真的能起到什么作用吗？当时我无法给出结论，然而在完成了整个课题后，我心中已经有了答案。

为了获得大量的名作midi供我筛选和分析，我自学制作了一个python脚本，花了几个小时从钢琴谱网站([www.gangqinpu.com](http://www.gangqinpu.com))上批量下载了mid文件，分别存放在以各个作者命名的文件夹里。随后，面对几千个midi文件，我兴奋地开始编写midi分析程序，在一开始我想用脚本语言写，但由于python不熟悉，还是选用的c++语言。

在网上查了midi的编码资料后，我感觉它不是太复杂，只要用二进制读取，位数、格式什么的都是固定的。然而在我编写的过程中，还是遇到了许多困难，例如动态数值的读取、midi信息的存储、c++如何进行批量文件操作等等。为此我又通过实例mid的二进制分析，见识了各种特殊情况，感叹要想bug free还是不容易呢。十天之后，我的程序完成度接近了100%，它不但可以按音轨、时间顺序输出所有音符（含拍号、调号、音高、时值），还可以统计每秒音符数、音高标准差、力度标准差、不稳定音程使用率、和声跨度这五个跟乐曲有关的量。

接下来我对下载的midi文件进行筛选，删除了大量非古典钢琴作曲家后，又删除了作品数量太少(<=10)的作曲家，经过层层筛选最后剩下7位古典钢琴作曲家。他们是：巴赫-P.E.Bach(1685-1750), 莫扎特-W.A.Mozart(1756-1791), 贝多芬-L.van.beethoven(1770-1827), 肖邦-F.F.chopin(1810-1849), 李斯特-F.Liszt(1811-1886), 勃拉姆斯-J.Brahms(1833-1897), 拉赫马尼若夫-S.V.Rachmaninoff(1873-1943)。

我小心翼翼的将这七位的文件夹复制到了另一个地方，作为备份。随后，经过了我程序的处理，他们由几百个mid文件变成了仅仅7个txt文件。此时的我迫不及待地逐个查看，感觉还是很满意的。此后，我用excel做了图表，又由于时间的放宽，我利用html和JavaScript做了一组网页。

这张图是钢琴家作品各项指标按比例折算后的数据，各项最高值为100%，给我们一个大体的、直观的感受，似乎晚辈们比先辈们的五维更大(作曲家以时间顺序排序，请看前面的出生年月)，对于图的分析，请看接下来的分量图：

从这幅图我们可以看出，由时代的变化，钢琴作曲家作品的速度、难度也在逐渐加强。在从16世纪开始的巴洛克时期，钢琴曲庄严、肃穆，巴赫所创的复调进行是当时主流。在巴赫逝世后，复调音乐后继无人，维也纳古典乐派时期到来。人们开始注重作品感情的融入，此间的突出的钢琴家有莫扎特、贝多芬等大师。到了约1820年，浪漫主义音乐开始萌芽，更加注重人的精神世界和主观感情，作品感情丰富，有时猛烈，有时抒情，更加为大众所接受。上图中，前文没提到的几位钢琴家均出自浪漫主义时期。

音符标准差是反映作品音域和变化程度的一个量。各位作曲家之间没有较明显的差距，但从时间顺序上看，晚期作品的音符标准差较大，这也反映了古典音乐从庄严肃穆到浪漫主义的转变。

由于网上下载的midi文件制作水平参差不齐，很多时候都是大批音符拥有同一个力度，所以这个量不能作为评价的依据，只能起简单的参考作用。

不稳定音程指的是用调内1、3、5音以外的音组成的音程关系。早期的古典音乐追求音乐的和谐悦耳，相比必定协和的稳定音程，作曲家不会大量使用有可能会产生不协和感的不稳定音程。随着古典音乐的发展，追求个性、抒发情感的作品越来越多，当人们想要表达不同于愉悦、舒畅的感情时，他们的作品里会大量的出现不协和音程，有的听着外张，有的则有一种紧缩感。这也丰富了音乐的表现形式，使得怪异的现代主义音乐能够被世人所接受。

关于最后一个属性最大跨度：由于我是按音轨整理的音符，而有的midi里存在某一处左手和右手弹的音符在同一音轨（一般是写在不同的音轨上）的情况，经我的程序统计，就会出现两个八度以上的大跨度。而在现实中估计只有姚明的手能有两个八度。因此我的程序把特别大的跨度忽视了，但同时这份结果也不真实了，只能参考着看看。

关于我的数据能不能引起别人的兴趣，我只能说不会有太大的作用。课题完成后，我再看着这些数据做成的图表，并没有感觉到强烈的想去感受古典音乐的愿望。音乐的主要感受器还是耳朵，而不是眼睛。不懂欣赏的人，看到这些数据，会有一时的冲动，随便找点推荐的古典音乐听一听，其目的不在古典音乐本身，而是在打发无聊的时间。真正懂得欣赏的人，听到了音乐，就仿佛看到了一大堆的数据。所以说，引起别人对古典音乐的兴趣，还是要让音乐在他的心灵留下印象，用数据还是太单薄了。

对音乐作品进行数据分析，还需要找到更加精确地反映作品特征的量，以这几种基础的统计量，是不能够解析作品与作品、 风格与风格之间的联系和区别的。