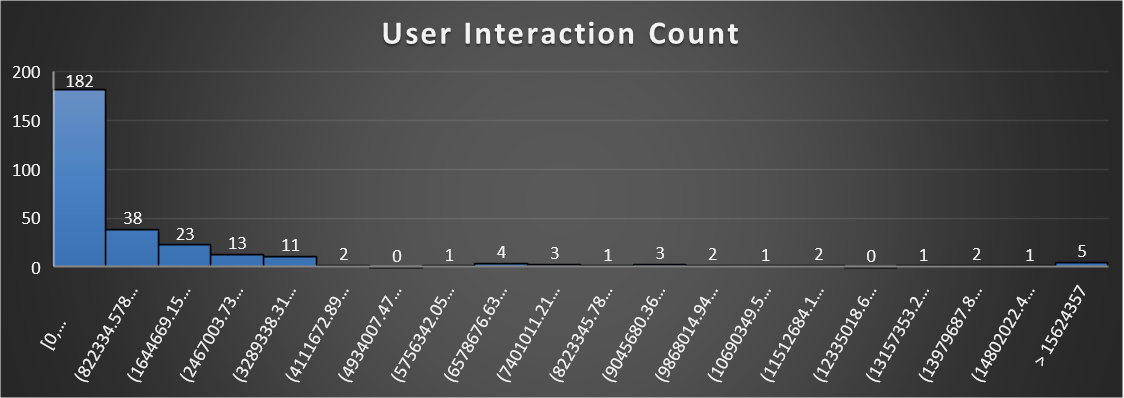
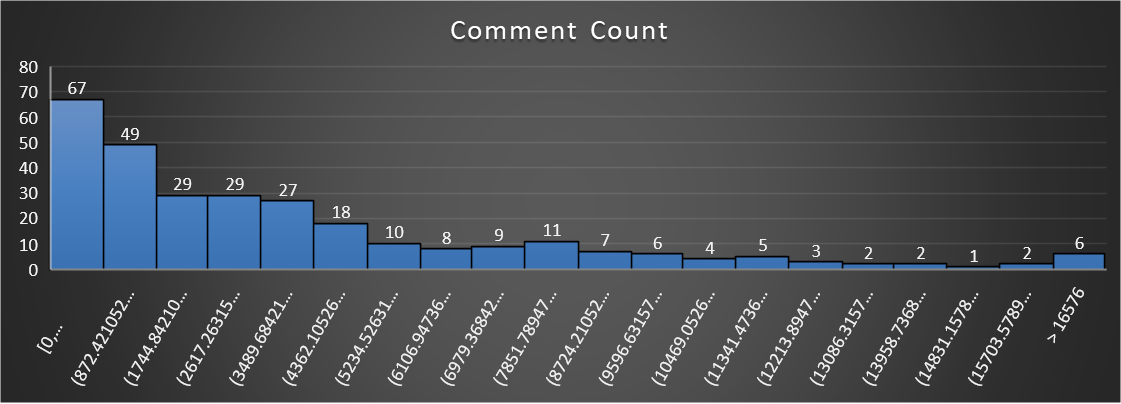
# Bilibili城市视频播放量弹幕数量可视化

通过上次的Bilibili城市视频信息爬取，可以获得全国地级市及以上城市的信息。这里从网上搜索到全国共295个地级市，由于城市数量有点多，为了提高速度，每个城市仅爬取搜索结果第一页的20个视频。从结果上看，第一页20个的视频的播放数量和弹幕数量还是很大的，足以反应Bilibili用户的情况。

从分布上看，如下分别是交互数量和弹幕数量的直方图。





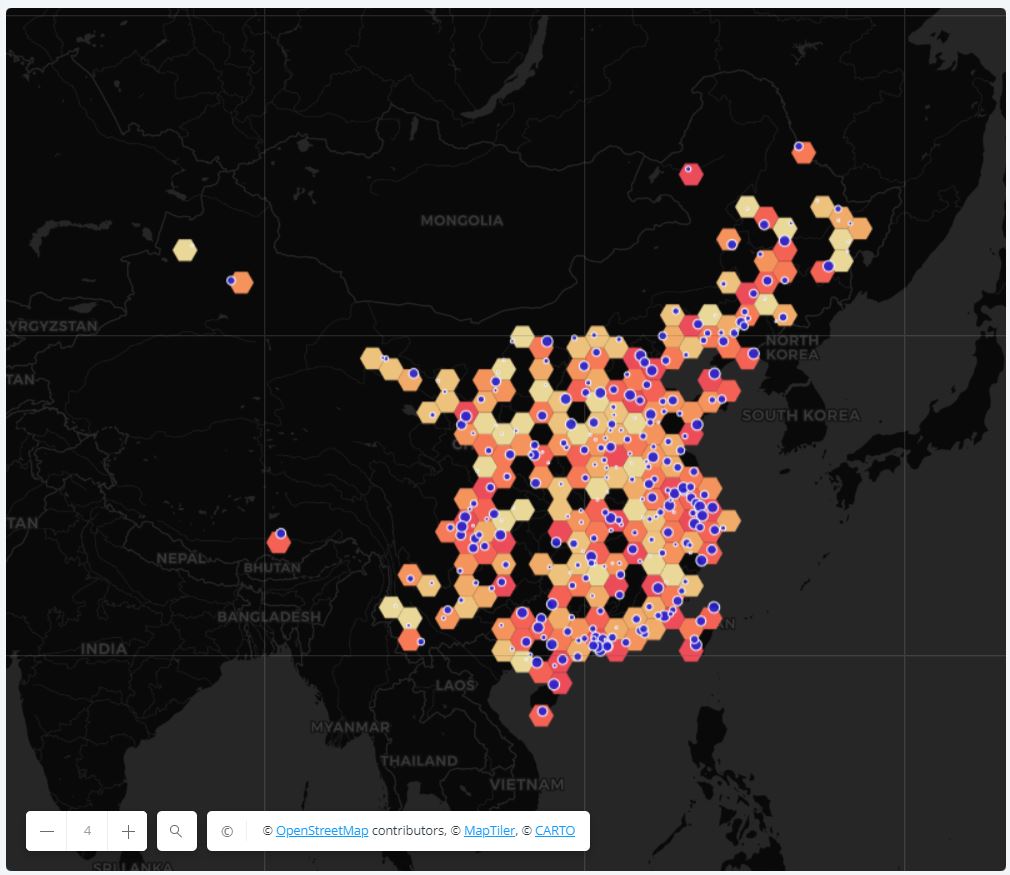
从上面两张图可以发现，大部分城市的交互数量和弹幕数量分布是一个常见的下降的分布，除了尾端的特别大的少数几个城市，从数据中可以知道这几个城市是：

交互数量：衡水、青岛、海口、武汉、香港。

弹幕数量：重庆、牡丹江、眉山、长沙、白银、台北。

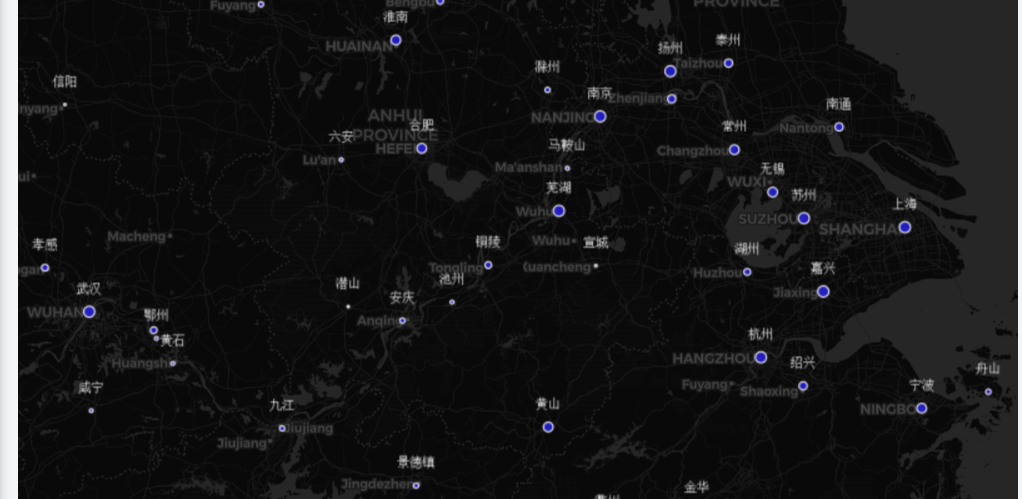
比较有意思的是香港、武汉。香港视频的播放数量如此大应该是源于去年的一系列运动，而武汉则是今年的新冠肺炎的事件。而弹幕数量中，重庆、长沙、台北的弹幕数量很高，说明这几个城市在Bilibili用户中具有较高的人气，是几个比较好的旅游城市。

为了将上述信息进行地图展示，需要获得每个城市的地理位置，即经纬度。通过调研百度地图提供了这样的API，可以依靠城市名称获得其经纬度，代码见附件的bilibili-comment-process.ipynb文件。



上图是将两个数量信息共同显示的效果，底层的六边形热度图展示了视频播放数量信息，上层的蓝色点的大小则表示了弹幕数量信息。可以大概地观察到全国的热点城市分为了6个区域：北京、江苏、台湾、广州、湖南湖北、四川。这几个区域也是目前发展较为好几个区域，人口众多，自然对于Bilibili的使用也就更多。

如上图所示，其中可以看到江苏省附近城市特别密集，所以打算对该区域仔细看一下，如下图，显示了城市名。



从江苏省一连串的城市，可以发现基本构成了一个链状的分布，从上海开始，一直到达武汉。这一串在Bilibili上评论较多的城市都是长江一代的城市，也是适合人口密集居住的城市。

从上面的地图可视化可知，作为现在用户数量很多的视频网站Bilibili，它的信息从一定程度上反映了社会热点事件，例如香港和武汉的情况；同时，在某种程度上，它也可以反映城市发展的情况，发展较好的城市人口将会很多，从而为视频播放量和弹幕数量提供基础。在以上的调研中仅仅使用弹幕数量作为信息，实际上如果采取自然语言处理的技术理解弹幕文本，然后再进行分析，也许可以获得更加有意思的结果，例如城市评价等。