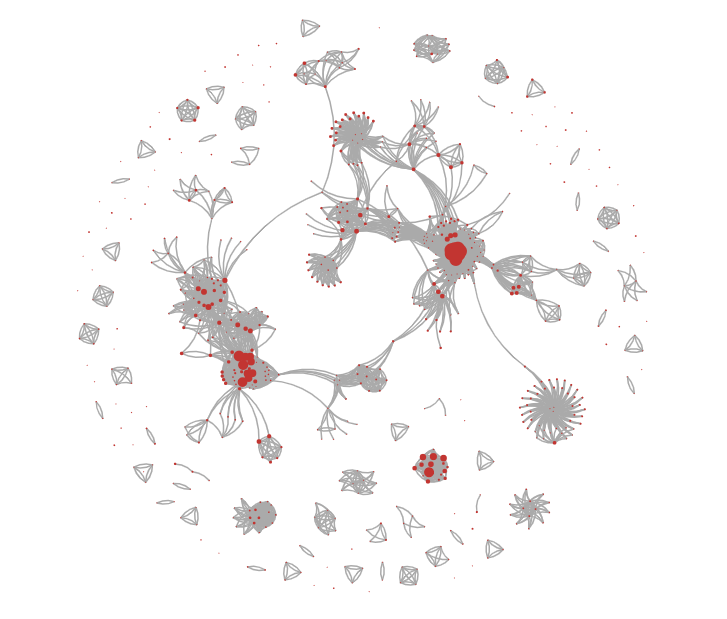
**上一期遗留问题：**

在上一期的网络图中，由于没有关注强度较低的节点和联系，导致在网络中出现了许多孤立的点。为解决该问题，应当更新数据处理的算法，设定阈值，将强度较低的节点和联系不显示在最终图形中。

当数据量较大时，每个节点上的标签（用作显示被引作者的姓名）没有显示的必要性。因此，进入pyecharts库文件夹，修改LabelOpts类源代码，使得网络图不显示作者名，以便更好地查看引文形态。



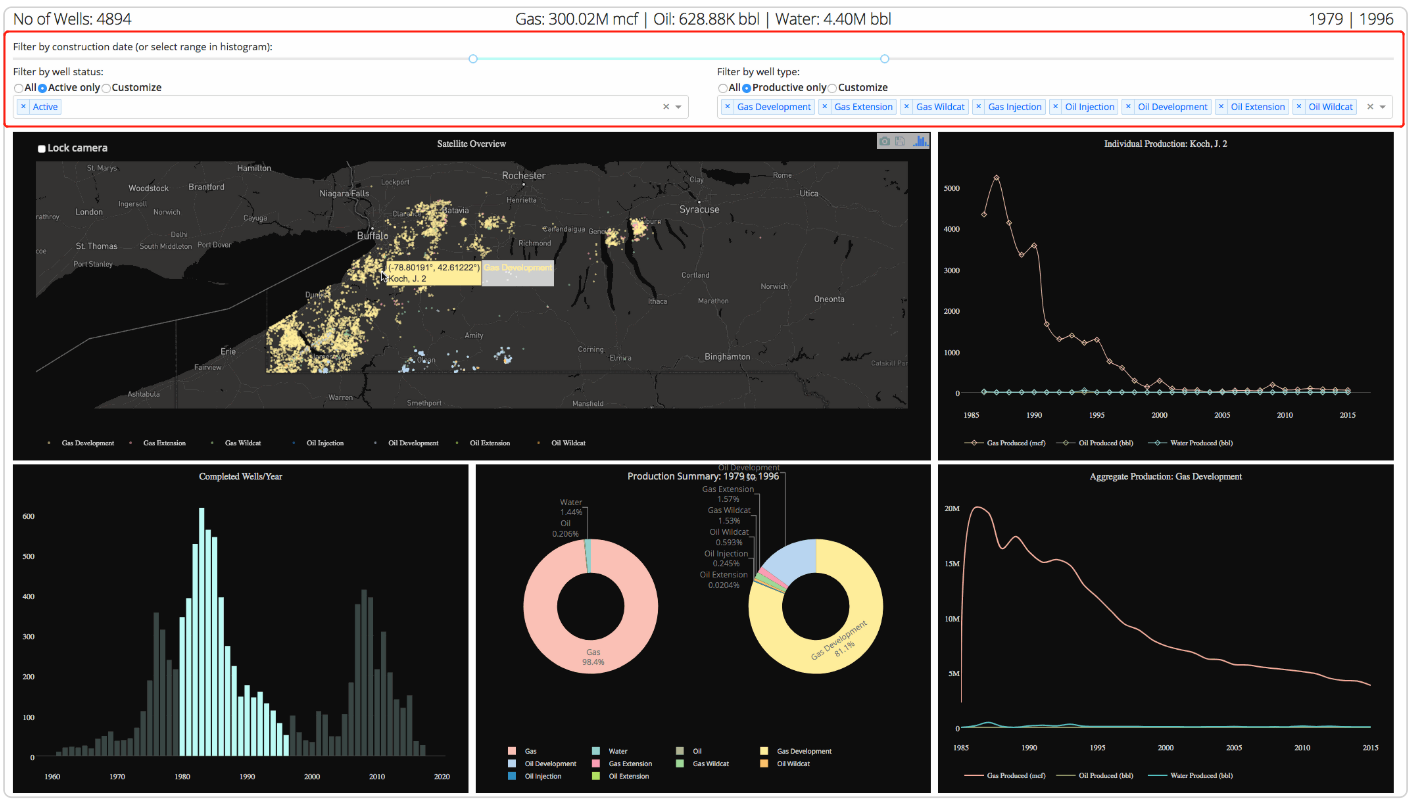
可以看到，更新数据处理过程后，通过筛去低强度的点，网络图中的孤立点大大减少了。

**小结**

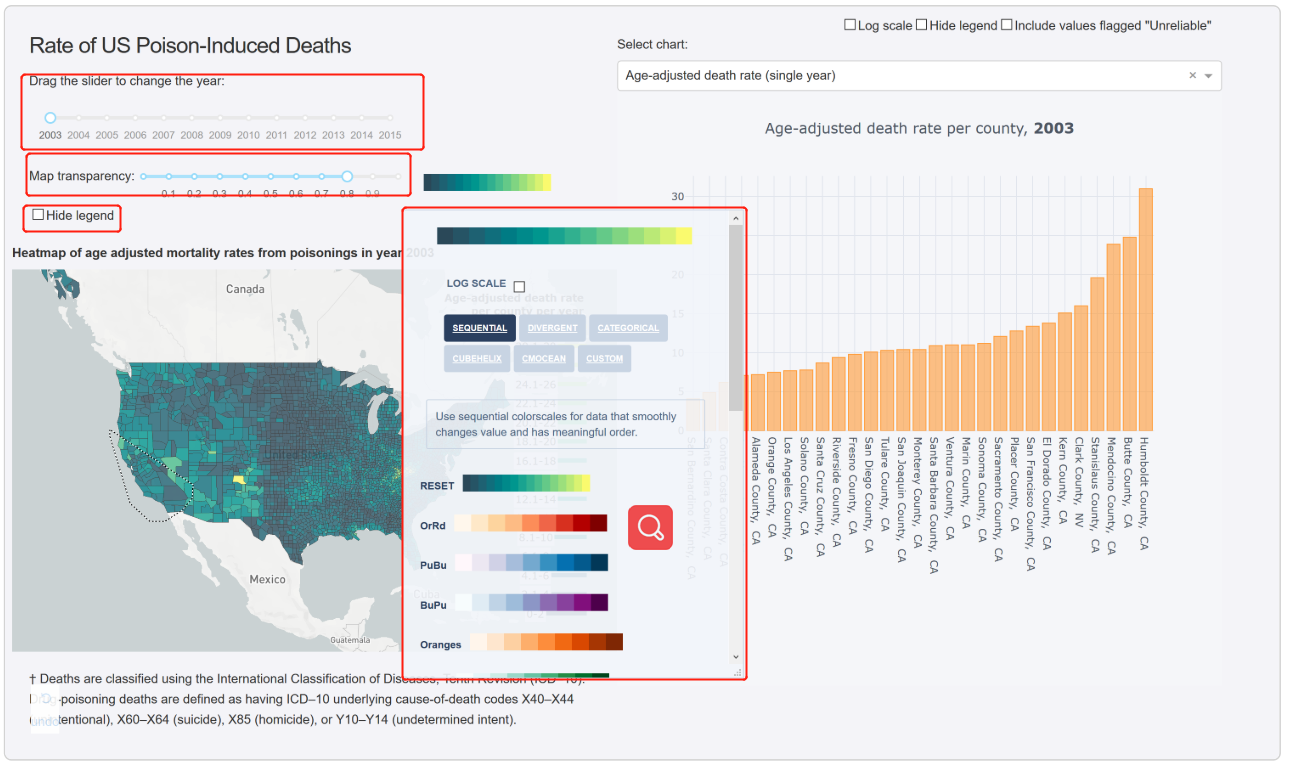
已将之前的散点图、热力图、关系图封装为函数，留待后用。

在使用了plotly\_express、pyecharts、cufflinks等python数据可视化库后，我发现它们都是基于Web的交互视图库，但并非那种能将文档、表格、视图整合在一起的交互式数据分析报告框架。这些库在互动性、定制性、表现力上都有较大的局限：尽管它们可以用少量代码快速生成较高质量的图表，然而一旦涉及到类似引文网络图这种不在内置图表中的可视化需求，我们就势必需要定位和大量修改库的源码。既然如此，不如直接使用开发框架定制数据分析报告。

Dash是用于搭建响应式Web应用的Python开源库，可以轻松构建包含交互元素的复杂应用。由Dash开发的app可以在浏览器中使用，因而它天生就具有跨平台的适应性，通过URL地址即可在Linux、Mac、Windows和移动端等平台使用。



示例图：更多的交互选项



示例图：用户自主选择展示颜色、透明度、数据范围

接下来的计划是参考Cite Space软件的功能，使用Dash框架来开发web app，重点是可视化的互动性、美观性和独特性。

**Cite Space软件学习笔记：**

CiteSpace软件系统是由美籍华人学者陈超美博士开发的一款主要用于计量和分析科学文献数据的信息可视化软件, 它主要基于共引分析理论(co-citation)和寻径网络算法(path Finder) 等,对特定领域文献( 集合) 进行计量，可以用来绘制科学和技术领域发展的知识图谱, 直观地展现科学知识领域的信息全景, 识别某一科学领域中的关键文献、热点研究和前沿方向。

科学知识图谱是以知识域( knowledge domain) 为对象,显示科学知识的发展进程与结构关系的一种图像。它具有“图”和“谱”的双重性质与特征: 既是可视化的知识图形,又是序列化的知识谱系,显示了知识单元或知识群之间网络、结构、互动、交叉、演化或衍生等诸多隐含的复杂关系,而这些复杂的知识关系正孕育着新的知识的产生。

科学知识图谱的绘制主要包括引文分析、共词分析和共被引分析。共被引分析是当两篇及以上文献同时被另一篇文献引用时, 就构成了共被引关系, 用于识别高价值文献、主要作者等。引文分析和共词分析是利用数学算法及计量方法对文献关键词、主题词等进行数据统计和聚类分析, 以获得某领域的热点主题和前沿趋势。

一般说来,Cite Space知识图谱的合格满意标准主要是: 数据完整、程序正确、图谱美观、解读合理，因此其主要流程可以概括为数据采集、数据处理、参数功能选择、可视化和解读。

Cite Space图谱示例：

