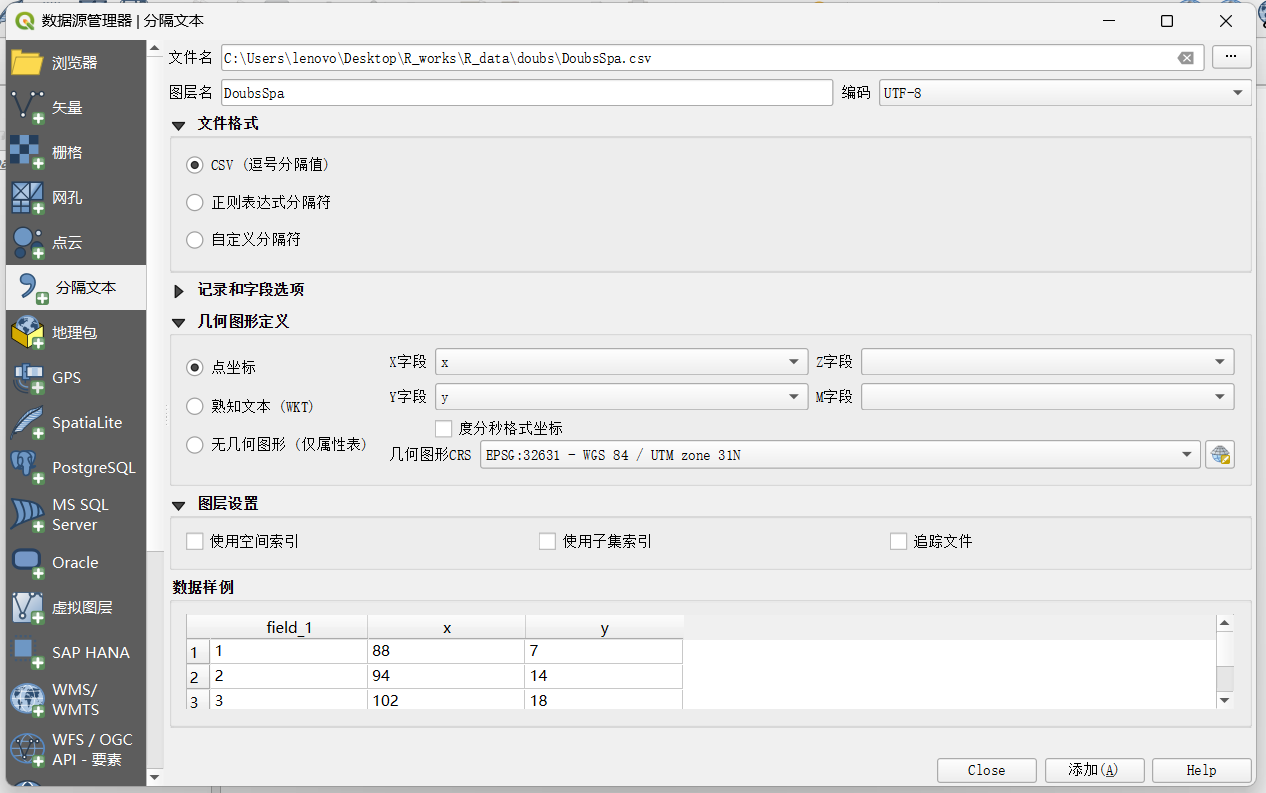
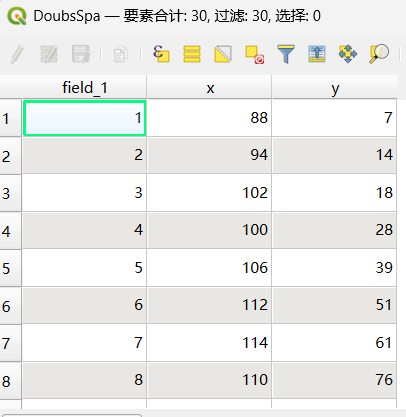
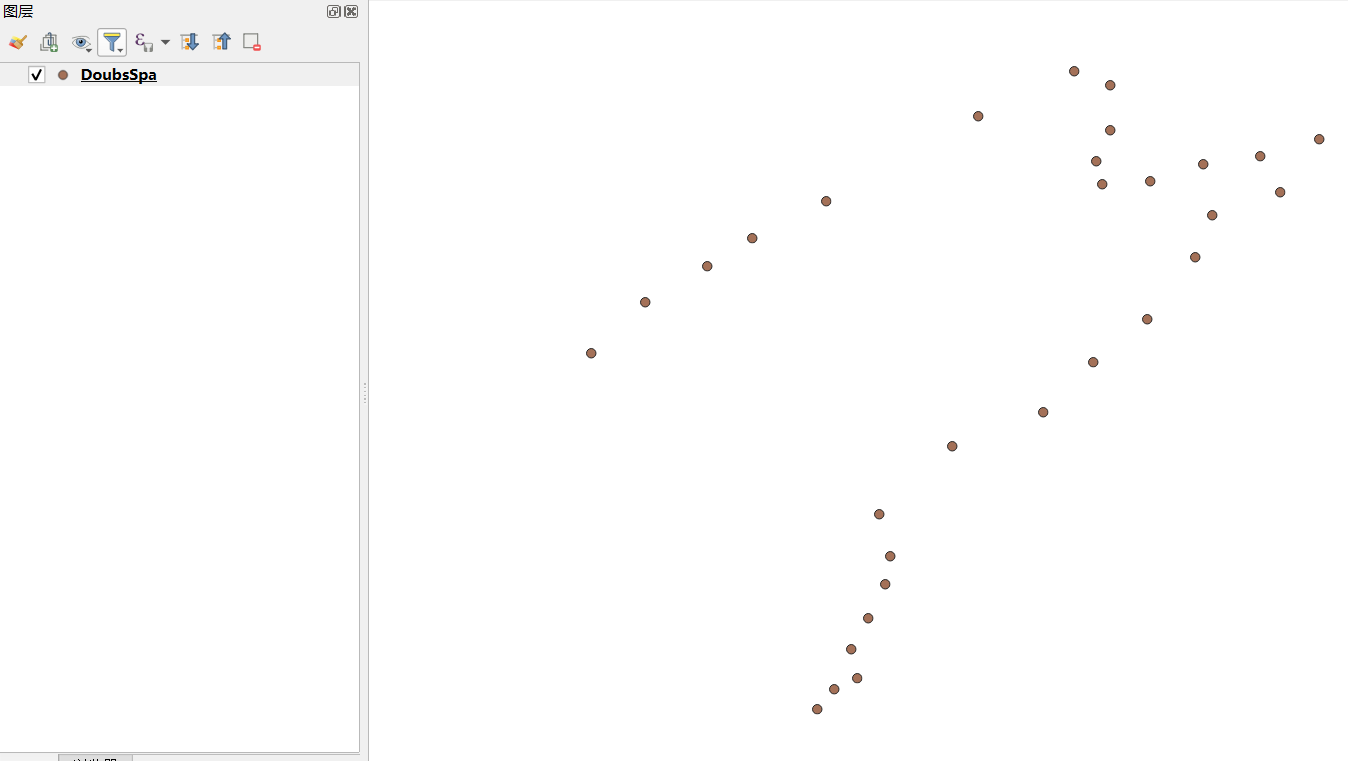
## **Q1：对于Doubs数据集，采样点坐标为千米，但没有参考系信息，只有通过QGIS确定各点地理坐标。请简述QGIS操作过程并给出对应点的地理坐标。**

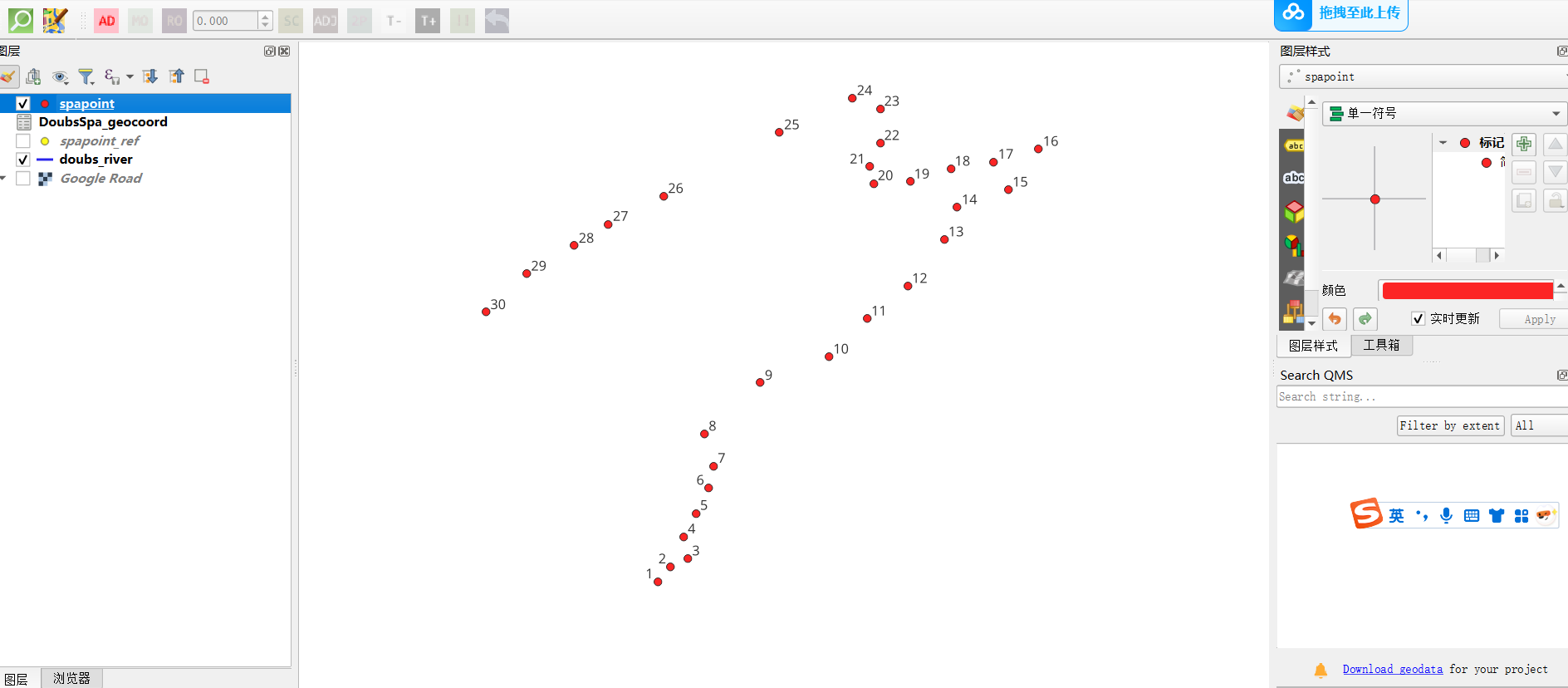
1. 打开QGIS 3.40.4，工具栏 > 图层 > 添加图层 > 添加分隔文本图层（csv文件） > 将包含Doubs数据集采样点信息的DoubsSpa.csv导入到图层中成为点数据。



导入的点数据及其属性表如下图所示：

1. 为点文件添加“单一标注”，标注每个采样点的id，方便后续创建新点文件时对应；工程 > 导出 > 导出地图为PDF格式文件，预备后续配准使用，若使用PDF文件可以避免图像在配准过程中畸变，且较少配准点即可达到较好的配准效果

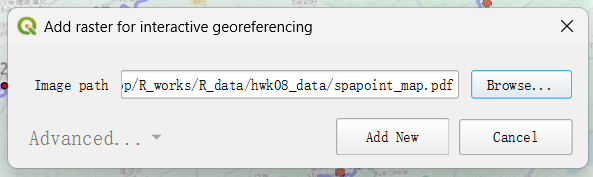


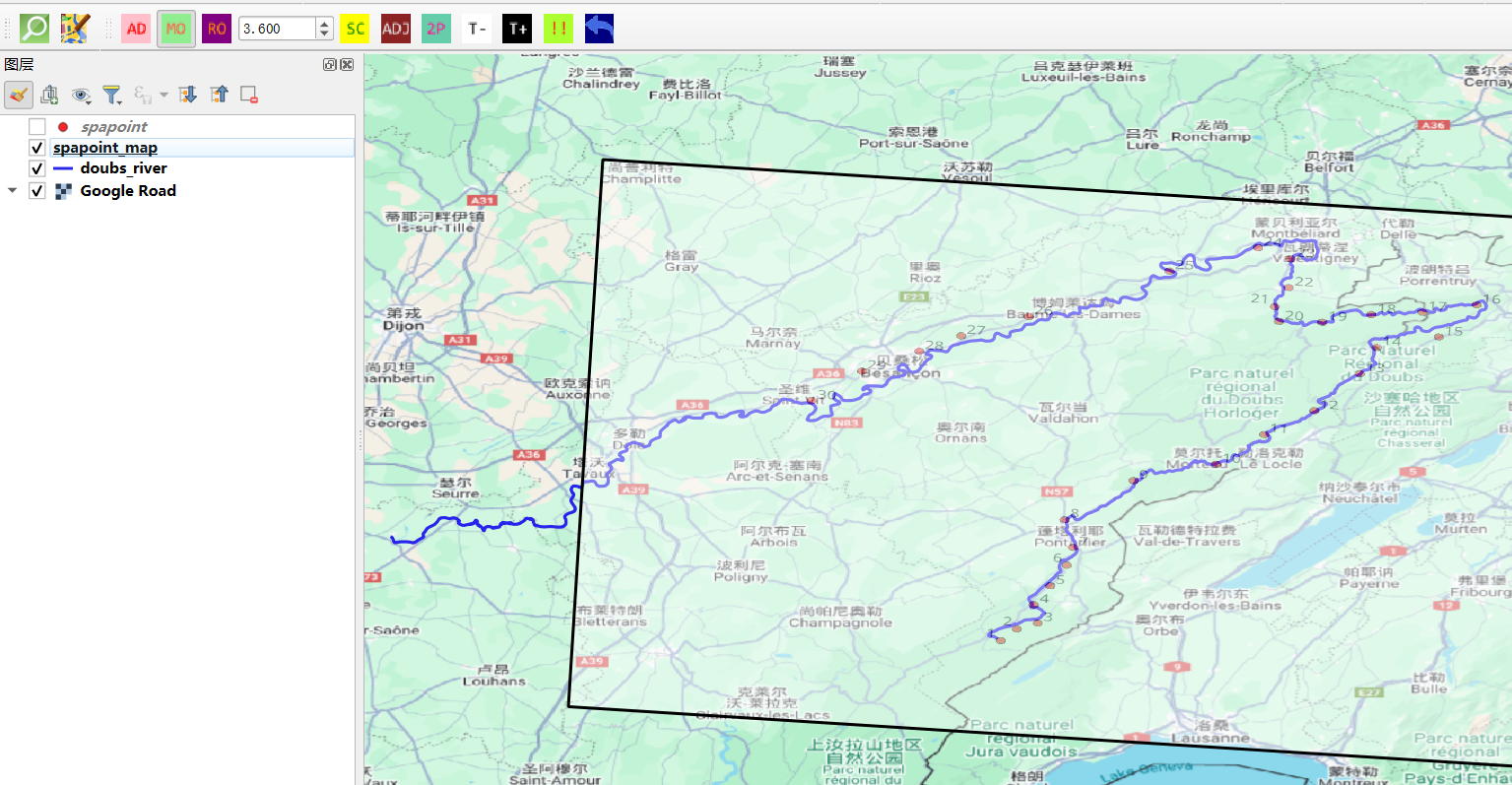


1. 安装Freehand Raster Georeference 插件

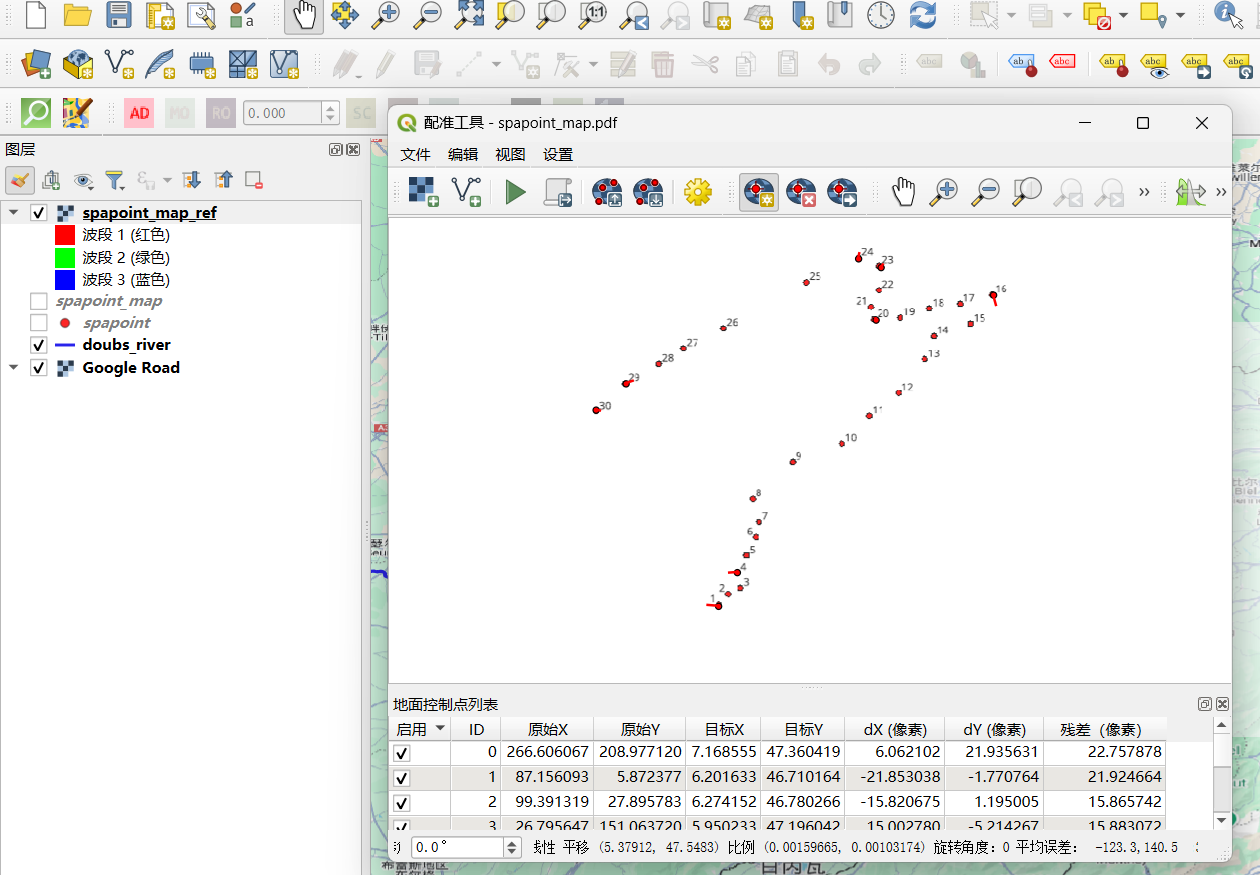


1. 将上次作业中下载的Doubs河数据导入到画布中，添加一个底图；然后使用Freehand raster导入，用MO、SC等工具移动缩放，调整前面保存的点图层pdf文件，使其与Doubs河数据基本匹配，便于添加控制点时对应。

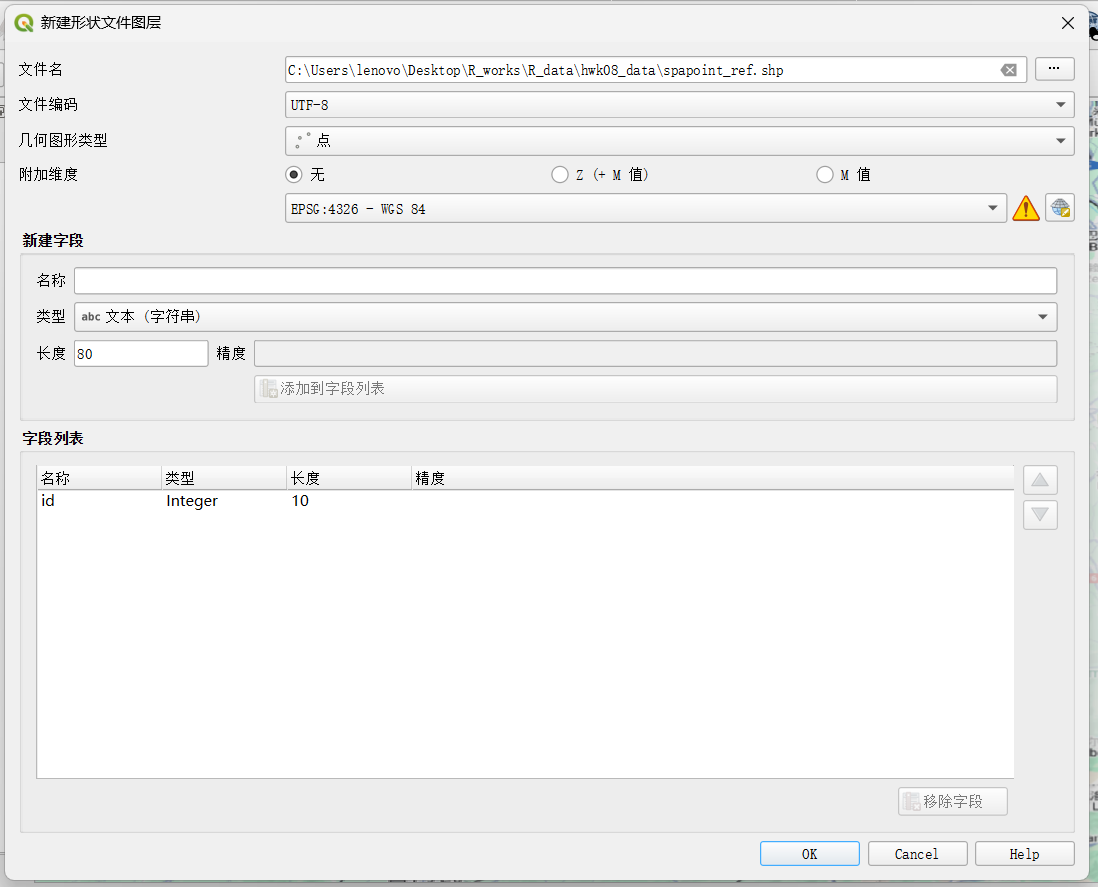


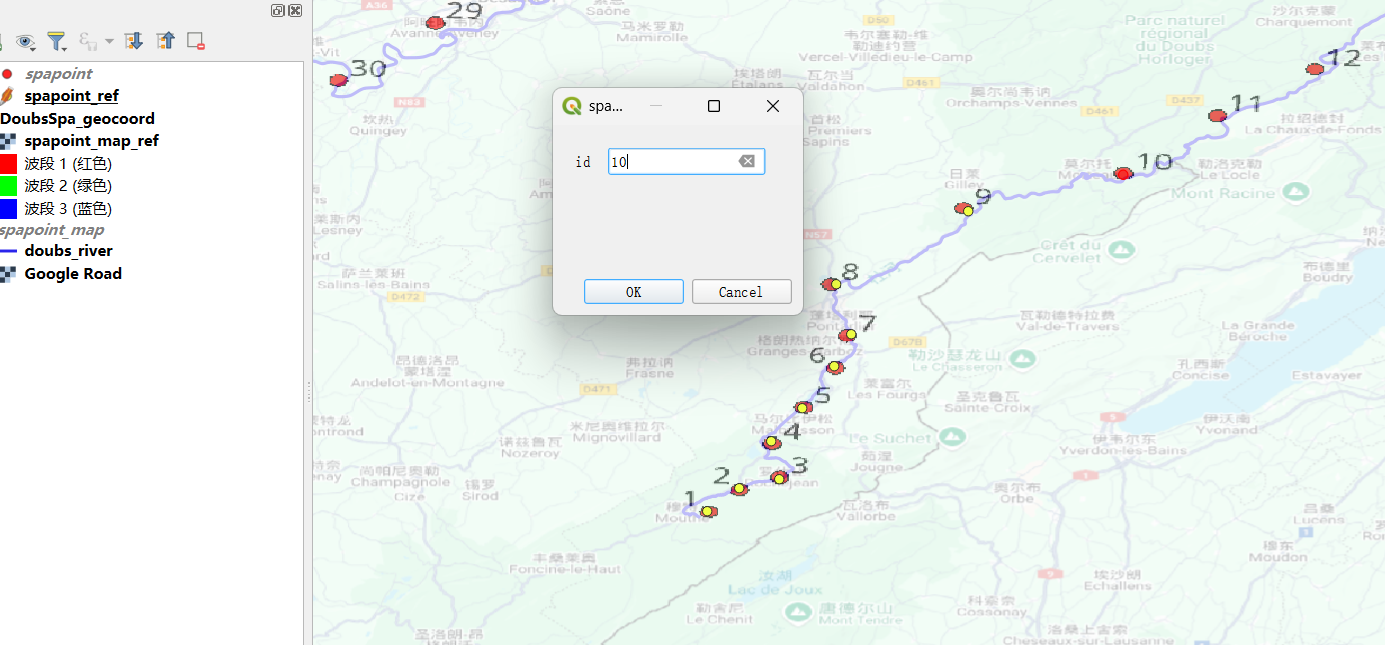


1. 根据Doubs河线数据，使用工具栏 > 图层 > 配准工具（Georeference）对pdf点文件进行配准，尽量使所有采样点位于河流上。

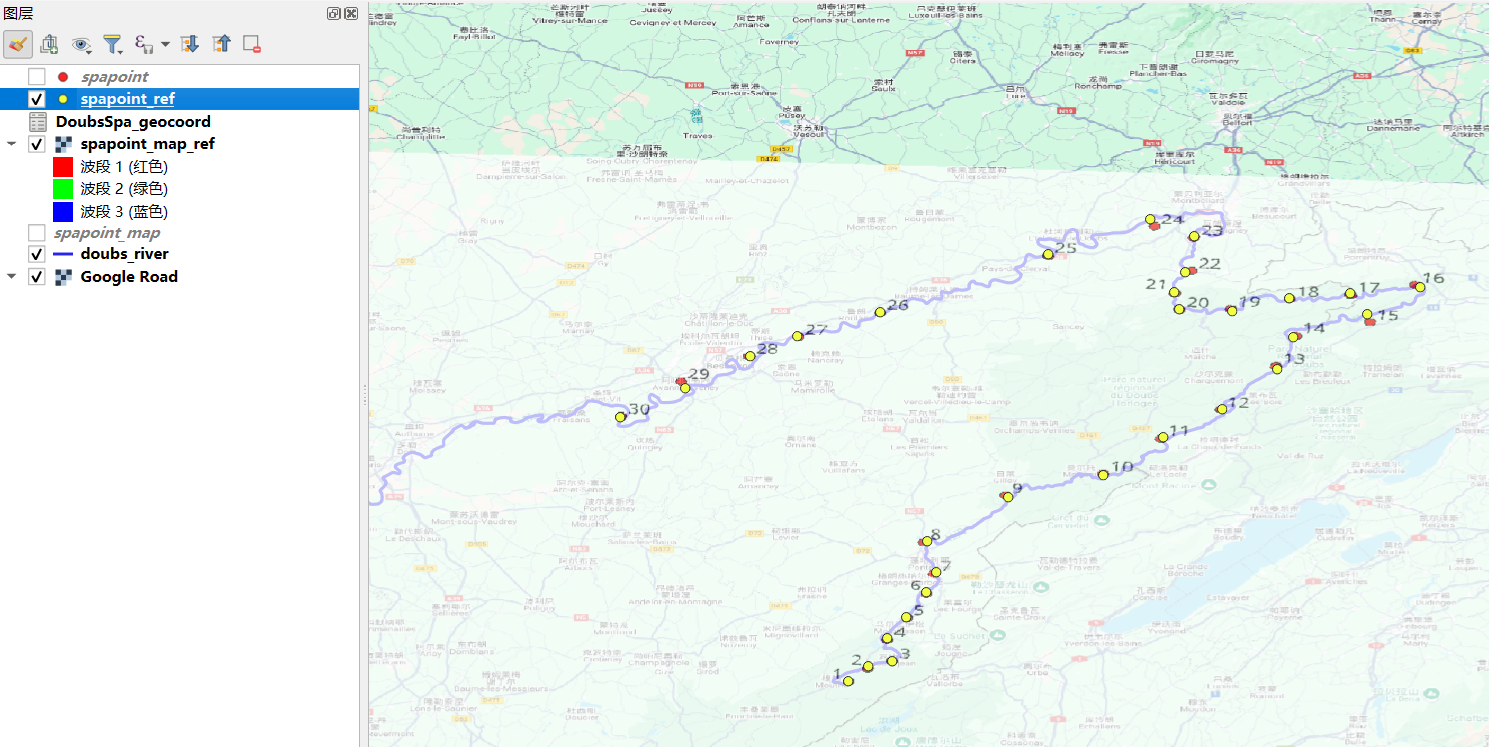


1. 然后在工程文档中创建一个新的点shp文件，根据配准后的栅格图层，将采样点矢量化，并对应标注各点的id，然后将矢量点数据的地理坐标保存为csv格式，即实现提取采样点的对应地理坐标。



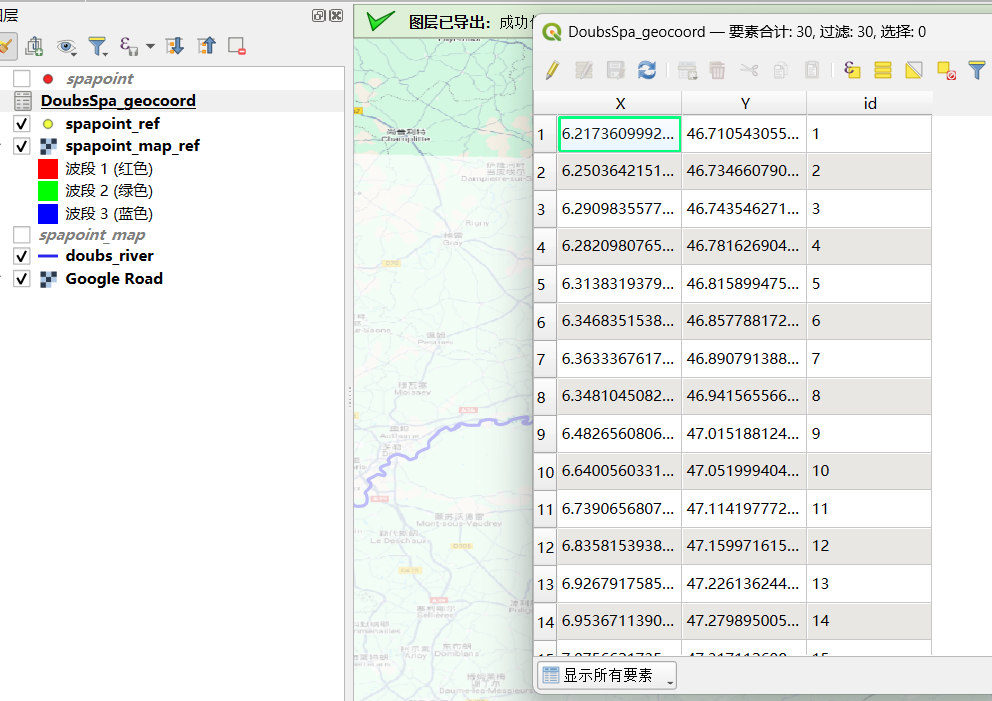


根据已配准栅格图层进行的采样点矢量化结果如下（黄色点）：

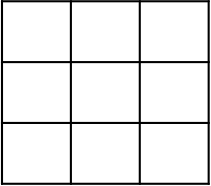
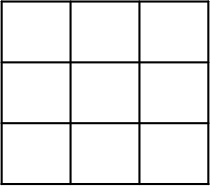




1. 最终获得的采样点地理坐标如下，保存在DoubsSpa\_geocoord.csv中



## **Q2：为确定是否存在空间自相关，需要确定空间邻近，包括Rook和queen相邻，请分别指出如下两图中蓝色位置的空间邻近，并写出各自的空间权重矩阵。**



第一张图为Rook邻近，仅考虑共边单元的相邻关系；

第二张图为Queen邻近，考虑共边和共顶点单元的相邻关系；

空间权重矩阵：

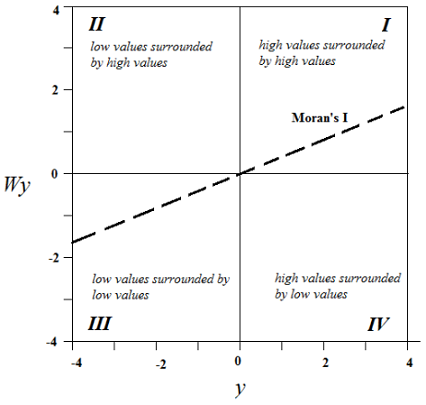
第一张图（rook邻近） ；

第二张图（queen邻近）

**Q3：在空间数据分析中，常见Moran‘s scatter plot和Lagged mean plot，请回答如下问题：**

1. **Q3-1：根据Moran’s scatter plot图，如何分辨空间自相关？**

如下方示意图，Moran散点图中，在拟合直线的辅助下进行观察，若散点多集中在一、三象限，且回归线斜率>0，则表明在地理空间内，高值附近存在较多高值（HH），低值附近存在较多低值（LL），因此存在正空间自相关；若散点集中在二、四象限，且回归线斜率<0，则表明低值附近存在较多高值（LH），高值附近存在较多低值（HL），判断为存在负的空间自相关。若散点在四个象限内散乱分布，或拟合曲线趋势接近于0，则判断为可能不存在空间自相关。

****

1. **Q3-2：两图都有一直线，各自的意思是什么？**

Lagged mean plot（滞后均值图）的直线是一条45°的固定对角线，即x=y线。这条对角线作为参考线，代表某一单元周边所有相邻单元的平均值，测量某一单元值与其相邻单元平均值的偏离程度，即用于初步判断该单元值与周围单元值的相似程度。当点刚好落在参考线上或集中于参考线附近时（即存在一定的线性趋势），偏离程度小，局部空间自相关程度高，若偏离参考线，则局部自相关程度低。

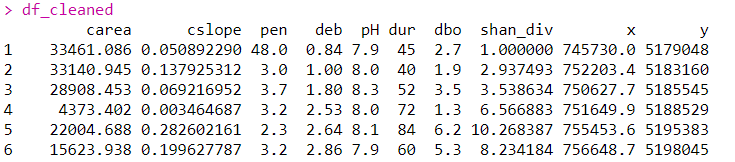
Moran’s scatter plot（Moran散点图）用于进一步判断空间自相关，其中的直线是拟合其中散点的线性趋势的回归线，其中斜率代表Moran’s I（即Moran指数），因此这条直线斜率的正负、绝对值代表空间自相关的方向及强度（在Q3-1中已回答）。

## **Q4：****通过探索性分析，结果表明doubs河流中鱼群的物种多样性存在一定程度的空间自相关，请简述如何将空间自相关纳入鱼群物种多样性-环境要素关系模型中。**

以doubs河流中的鱼群物种多样性数据为例：

构建模型时，target为各采样点鱼类的shannon多样性指数（shan\_div字段），而features则为前面已筛选过、排除了高度共线性（pearson相关系数 > 0.8，或其他自定义阈值）且与待预测target密切相关的环境变量，最终保留carea（面积）、cslope（坡度）、pen、deb、pH、dur、dbo这几个字段；

在此基础上，纳入空间自相关的方式有多种：

* 直接将各采样点的坐标即x，y作为独立的两个feature纳入到机器学习模型中，以表征各采样点间的空间关系，得到feature-target数据框如下（shan\_div为target）：用上述数据框中的所有feature变量和shan\_div构建简单线性回归模型，即初步实现将空间自相关纳入到鱼群物种多样性-环境要素关系的机器学习模型中，代码如下：
* 构建一定缓冲区范围内的空间格网，将鱼类丰度点数据分层后，计算各层上点到各格网的缓冲区距离（即已知点-预测点距离，buffer\_distance），然后将各层buffer\_distance数据作为新的features引入机器学习建模，评估并筛选较为重要的缓冲区距离变量。
* 还有许多其他方法，选择合适的空间邻近，构建空间权重矩阵，计算各邻近点目标变量均值（即空间滞后），以此作为特征（feature）纳入到机器学习模型中参与训练；通过空间聚类，为各点增加分类变量feature，例如河的“上游/中游/下游”，作为特征参与到机器学习模型中。