GIS学习小组练习Part I

2020/12/15

纽约共享单车数据处理

* 数据来源：[【数据分享】纽约花旗共享单车轨迹数据（2013-2020）](https://mp.weixin.qq.com/s?__biz=MzA3NTk0MTU3MA==&mid=2247504214&idx=1&sn=45cabb2f8b7352b9b34800dabb3d9735&chksm=9f6a42fba81dcbed8af62fe2e0835c7a3c0b4eb861d2252dd58e510c659c98af7c0ec73467cb&scene=126&sessionid=1607993471&key=2f88c2a11d638eea38954844950eb3d4a0d2e474fe5aa2b1b058bb12d0ad38dce02a63d1d5c001db9df6c71199a5db630d077c74628b335e19e77b0b2db9d470e9fb8723377741fbbf59609e5e06611285a5dab60c19d6061fece81d66aa62c8f1688d66512bddc694bec941bbbc170e9b2d2d7af334cefb2b8590598959fe1c&ascene=1&uin=Mjc3MTE5MzQ2Mg%3D%3D&devicetype=Windows+10+x64&version=6300002f&lang=zh_CN&exportkey=AUZZSfMZfW8IaSqkkob7XEI%3D&pass_ticket=kODwkd39ekHcduyOZOoHVP1sKpNPlAe2MDfWDkUAfb0qUlQc9G2ydfqVHCpu0%2Fs4&wx_header=0)

下载2020年11月月数据

* 目标：生成GIS数据，保存为Shapefile格式。每一行为一条骑行记录，保留所有原始变量，增加两个新变量：（1）骑行起点和终点的直线连线，（2）骑行起点和终点的直线距离。
* 步骤
  + 根据起点坐标和终点坐标生成分别代表起点和中点的shapely Point。
  + 生成代表起点和终点之间直线连线的shapely LineString。
  + 改为适合计算（纽约的）距离的投影。
  + 计算起点和终点的直线距离。

（这只是完成任务的一种步骤，可以设计自己的步骤，比如直接从起点和终点的坐标生成直线连线。）

* 需要的包：pandas, geopandas. shapely.geometry。建议先读GeoPandas的手册中Data Structure的部分<https://geopandas.org/data_structures.html>。
* 也许会用到的资料
  + 原始数据地理信息存储为经度和纬度，没有说明使用了什么坐标系，可以默认为WGS84（代码EPSG:4326）。适合计算（纽约的）距离的投影可以使用UTM zone 18N （代码EPSG:26718）。代码查询可以参考<https://epsg.io>。
  + 根据坐标生成shapely Point，使用geopandas的points\_from\_xy，参考<https://geopandas.org/gallery/create_geopandas_from_pandas.html>。
  + 根据点（shapely Point）生成线（shapely LineString），使用shapely.geometry的LineString，参考<https://automating-gis-processes.github.io/site/notebooks/L1/geometric-objects.html>。
  + 生成GeoDataSeries或者GeoDataFrame，参考GeoPandas的手册。
  + 将GeoDataFrame和普通DataFrame横向合并到一起时，可以使用merge，但是一定要把GeoDataFrame放在前面，否则会丢失shapely geometry信息。
* 安装geopandas注意事项
  + 为了避免安装时和其他包产生冲突，建议建一个新环境安装，在这个新环境中完成数据处理。
  + 在Anaconda下使用jupyter notebook/lab，可能遇到这样的问题：在新环境中打开jupyter notebook却发现它依然只有默认（base）环境。如果出现这个问题，请到base环境中装nb\_conda\_kernels，然后在新环境安装ipykernel，应该就可以解决问题了。
  + 安装pandas时就自动安装了shapely，不需要再单独安装。

预期后续的练习（可以自行探索）

* 从网络获取纽约市边界地图作为base map，制作骑行站点的地图。
* 结合OpenStreet Map的道路信息，得到骑行实际的最短路径，计算距离。
* 结合OpenStreet Map的Points of Interest数据，探索骑行的目的地是什么类型的地方。

如果有兴趣自行探索的话，建议参考GeoPandas的手册和tutorial：<https://automating-gis-processes.github.io/site/index.html>。