**Steps 2024-7-01**

对Million dollar plant数据里面所有厂的county，找到那个county 100km以内所有的county。也就是最后行成如下数据：

county | neighbor\_county | distance

06001 | 06002 | 50

06001 | 06003 | 70

06001 | 06004 | 80

...

这个exercise应该用我之前给你的county shapefile就能完成，距离的话定义为county centroid之间的距离。

先用qgis算出polygon的centroid，然后再county两两间用R算centroid之间距离。算出距离后，再限制到距离是<100km的pair就好了

我们这一步主要是想看一下，如果outcome是以厂的county为中心，计算一个area内的biodiversity outcome，结果会不会有不同

**Memo**

First, I used QGIS to generate the centroids for each county. However, there is a problem: one county might have several centroids due to municipalities, independent cities, or census-designated places. For example:

A map of the united states

Description automatically generated

This is not a significant problem since we want to find the distance between centroids. In conclusion, we have 224 distinct counties with 280 distinct centroids. We loop through the 280 centroids with 3824 centroids from mdp\_county\_centroid.gpkg.

Then, we filter out the data by distance less than 100 km. One point worth noting is that if we filter by `county != neighbor\_county`, we only have 222 distinct counties. This means we lose 2 counties, which are 04013 (Maricopa County) and 06073 (San Diego County). The nearest counties are farther than 100 km.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

I tested the data in several ways, and it might not be a problem with the data processing procedure. I would assume it might be because they are very large counties and are in desert areas.