先定义Y1 Y2

Part1 如何确定16个X

Part2 通过Y1 Y2回归从16个X里面筛选若干个（建立了X到Y1Y2的模型）筛选两次

Part3 是筛选过的X随时间的变化（建立X到year的模型）

Part4 只针对Y2 各州的不同 去分析已筛选的X+4个其他变量的相似和相异

为了更好地描述每个州的energy profile，尤其是其中的清洁可再生能源的使用情况，我们从data中取出两个我们感兴趣的项目，TETCB 和 CRTCB 作为参考。将TETCB定义为Y1，CRTCB定义为Y2。

为了研究Y1Y2的变化，根据part1A对各州能源概况的描述如能源种类以及能源消耗部门，我们列出了16个我们觉得可能相关的变量。

16个X的表格

通过作X和Y1Y2对时间的图像，我们发现Y1Y2以及16个X对时间大部分存在着广义线性关系，因此我们考虑使用线性回归模型。为了筛选出16个X中对Y1Y2有着显著相关性的X，我们在Rstudio中运用向前向后的方法得到了各个州的若干个X。各个州的不同的X就描述了各个州的energy profile。（筛选次数为两次）

各个州的X 和函数（对Y1Y2的都有）

我们发现CA州的筛选所得的X与CA州的Y2相关性不显著，通过SIS

SIS part

为了描述各个州的energy profile如何随时间变化，利用各州筛选所得的X与时间作线性回归。

各个州的X随时间的变化情况

看图说话

AZ

CA

NM

TX

为了研究各个州清洁可再生能源的使用情况，我们将单独研究Y2与之前所筛选出来的与之有着显著相关性的X

各个州的CRTCB随时间的变化情况

看图说话

AZ

CA

NM

TX

各个州Y2的X的表格

对X进行相同相异分析

各个州地理天气工业人口数据

对这些再进行相同相异分析

结论：