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。

In order to describe the energy profile of each state, especially the usage of cleaner, renewable energy sources, we take 3 variables, TETCB, NUETB and RETCB, we interested in from all 605 variables as the reference. We add up the data of NUETB and RETCB, coded as CRTCB. We define TETCB as Y1 and CRTCB as Y2.

TETCB Total energy consumption

NUETB Electricity produced from nuclear power

RETCB Renewable energy total comsumption

CRTCB Clean and renewable energy total comsumption

CRTCB=NUETB+RETCB

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。

In order to study Y1 and Y2, we listed 16 variables below that we thought might be relevant according to the energy profile of each state in Part1\_A

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。（筛选次数为多次）

We plot X, Y1, Y2 against time respectively, discovering that most of them have general linear relationship. Consequently, we decide to apply the general linear regression model. In order to pick out the variables that are significant correlated with Y1 and Y2 respectively among 16 variables, we used a statistical method named **step forward-backward**. The result are shown below.

各个州的X 和函数（对Y1Y2的都有）

特别地，我们发现CA州的筛选所得的原有X与CA州的Y2相关性不显著，故通过SIS

We find that all 16 variables are not significant correlated with Y2, so we take SIS method to solve it.

SIS part

找到了上述图标的中的变量。

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。

In order to characterize how the energy profile of each of the four states has evolved from 1960 – 2009, we regress time and the X we selected.

各个州的X随时间的变化情况

通过Figure 6-9我们可以描述各个州的energy profile是如何随时间变化的

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

From Figure 6-9, we can see how the energy profile of the four states has evolved from 1960-2009.

In order to clarify the usage of cleaner, renewable energy sources, we use the model we developed above, choosing Y2 to observe.

各个州的CRTCB随时间的变化情况

As this image shown, Califonia triumphs and has much larger CRTCB than other states. Taxes and Arizona are in the middle, while New Mexico is left behind. There clearly distribute as three groups. In order to evaluate the simularities and differences, we introduce several X and other possible influencial factors.

各个州Y2的X的表格

对X进行相同相异分析

结论：

AZ州和NM州的与CRTCB显著性相关的X都是石油在交通部门的消耗量和支出，TX则是天然气在住宅部门的支出，而CA最为不同，与Y2显著性相关的是煤油型喷气燃料总消耗量.

From the chart, we notice that All petroleum products consumed by the transportation sector is highly correlated with CRTCB in both Arizona and New Mexico, while in Texas, Natural gas expenditures in the residential sector (including supplemental gaseous fuels) influence the CRTCB most. But in California, it's unusual that the Kerosene-type jet fuel total consumption is highly correlated to the CRTCB, which is not included in the primary 16 putative variables.

各个州地理天气工业人口排名（数据base）

CA州和TX州的CRTCB都非常高，对应地我们发现CA州和TX州的人口众多，工业发达，在气候因素中，CA州和TX州的降水都较NM州和AZ州较多。

NM州和AZ州的CRTCB都较低，其中NM州的CRTCB非常低，对应地我们发现NM州和AZ州的人口稀少，工业不发达，在气候因素中，NM州和AZ州的降水较CA州和TX州较少。

For other possible influencial factors, we notice that both the population and industrical develepment in Califonia and Texas are greater than other states, and the precipitation of them is more than others too.

On the contrary, the population and industrical develepment in New Mexico and Arizona are a lot less than other states and the precipitation of them is less than others too.

These may indicate that the precipitation is helpful for cleaner and renewable energy like hydroelectricity, and the industrical develpment and largh number of people will help to improve the tecnique for efficiently utlizing cleaner and renewable energy.