In [35]:

```
import numpy as np
import matplotlib.pyplot as plt
import datetime as dt
import numpy.lib.recfunctions as nlr
%matplotlib inline
```

Firstly import csv and open the csv file.

In [36]:

```
import csv
file = open('public_layout.csv','r')
```

In [37]:

```
reader = csv.reader(file, delimiter=',')
```

In [38]:

```
fullcsv = list(reader)
```

Create an empty dictionary and for each variable whose absolute value of correlation coefficient with the heating space energy consumption is greater than 0.35 we use the dictionary to store the position of this variable and the correlation coefficient.

In [39]:

```
dic_1=dict()
print(dic_1)
for i in range(801):
    data = np.genfromtxt('recs2009_public.csv',delimiter=',',skip_header=1,use
    coef = np.corrcoef(data[:,0],data[:,1])
    if abs(coef[0][1])>=0.35:
        dic_1[i]=coef[0][1]
print(dic_1)
```

```
{}
{1: -0.44426586461766993, 2: -0.4190966807063039, 3: -0.467637225
09007732, 667: 0.41840045330589914, 38: 0.49731017738955707, 6:
0.57346926429617373, 7: -0.47280585981691542, 8: 0.57809758561548
819, 9: -0.441908303201280999, 11: -0.55489482644281862, 461: 0.36
566012307803447, 46: 0.51335383806870183, 40: 0.4987542249568328,
35: 0.40833383452217303, 430: -0.37091206303465069, 43: 0.4887697
5822655272, 705: 0.41145641648498477, 315: 0.37933839920084622}
```

Sort the dictionary according to their values of correlation coefficient.

In [40]:

```python
import operator
sortedDic=sorted(dic_1.items(), key=operator.itemgetter(1))
sortedDic
```

Out[40]:

```
[(11, -0.55489482644281862),
 (7, -0.47280585981691542),
 (3, -0.46763722509007732),
 (1, -0.44426586461766993),
 (9, -0.44190830320128099),
 (2, -0.4190966807063039),
 (430, -0.37091206303465069),
 (461, 0.36566012307803447),
 (315, 0.37933839920084622),
 (35, 0.40833383452217303),
 (705, 0.41145641648498477),
 (667, 0.41840045330589914),
 (43, 0.48876975822655272),
 (38, 0.49731017738955707),
 (40, 0.4987542249568328),
 (46, 0.51335383806870183),
 (6, 0.57346926429617373),
 (8, 0.57809758561548819)]
```

I will print the indices and correlation coefficient of variables we will use in our model besides materials.

In [42]:

```python
variables_chosen=[6, 315, 430, 705]
print(sortedDic[-2])
print(sortedDic[8])
print(sortedDic[6])
```

```
(6, 0.57346926429617373)
(315, 0.37933839920084622)
(430, -0.37091206303465069)
```

After selecting the most influential variables,we will construct the design matrix for these variables and variable of 'WALLTYPE' whcih is material for outer wall.