In [1]:

```
import pandas as pd
import numpy as np
# import scipy.stats as sps
# import tensorflow
import matplotlib.pyplot as plt
import seaborn as sns
# import sklearn as skl
# from sklearn import preprocessing
from sklearn.decomposition import PCA
# from sklearn.model_selection import train_test_split
# from keras.layers import Input, Dense
# from keras.models import Model
%matplotlib inline
```

In [2]:

```
%%time
fa_dir = '/Users/stevecoggeshall/Documents/Teaching/Fraud Analytics/2018 USC fra
ud class'
myvars = pd.read_csv(fa_dir + '/data/NY property/NY property vars 1 million zsca
le.csv', index_col=0)
```

```
CPU times: user 35.1 s, sys: 3.11 s, total: 38.2 s
Wall time: 40.4 s
```

In [3]:

```
def mem_usage(pandas_obj):
    if isinstance(pandas_obj,pd.DataFrame):
        usage_b = pandas_obj.memory_usage(deep=True).sum()
    else: # we assume if not a df it's a series
        usage_b = pandas_obj.memory_usage(deep=True)
    usage_mb = usage_b / 1024 ** 2 # convert bytes to megabytes
    return "{:03.2f} MB".format(usage_mb)
```

In [4]:

```
print(mem_usage(myvars))
```

```
488.00 MB
```

In [5]:

```
numrecords = len(myvars)
print(numrecords)
```

```
1048575
```

In [6]:

```
myvars.shape
```

Out[6]:

```
(1048575, 60)
```

In [7]:

```
%%time
mydata = (myvars - myvars.mean()) / myvars.std()
```

```
CPU times: user 1.42 s, sys: 1.86 s, total: 3.28 s
Wall time: 2.22 s
```

In [8]:

```
mydata.head(25)
```

Out[8]:

	fv_la_z3	vl_la_z3	vt_la_z3	fv_la_z5	vl_la_z5	vt_la_z5	fv_la_tc
RECORD							
1	-0.017442	-0.006894	-0.035573	0.039096	-0.001195	-0.017468	-0.003404
2	0.041219	-0.007313	-0.023622	0.031649	-0.017204	-0.035889	-0.042937
3	0.173077	-0.057336	-0.080118	0.173978	-0.059177	-0.083077	0.252584
4	-0.126544	-0.052837	-0.077530	-0.127203	-0.054533	-0.080394	-0.083387
5	-0.185336	-0.053291	-0.077321	-0.186301	-0.055001	-0.080177	-0.147934
6	0.092040	0.010679	-0.005715	0.048598	-0.001939	-0.017496	0.039271
7	-0.145741	-0.039019	0.018256	-0.134313	-0.034903	0.028601	-0.127953
8	-0.057213	-0.014697	-0.042009	-0.017558	-0.022163	-0.028557	-0.038114
9	-0.022529	-0.029363	-0.030958	-0.012399	-0.025098	-0.017080	-0.039074
10	0.032587	-0.001736	-0.029278	0.014435	-0.006640	-0.023259	0.040257

11	-0.185120	-0.057079	-0.074450	-0.186346	-0.059049	-0.078506	-0.146887
12	0.070893	-0.010213	-0.032185	0.049757	-0.014666	-0.026673	0.073688
13	0.021683	-0.009986	-0.026622	0.070190	0.003577	-0.012810	-0.065035
14	-0.036051	-0.028151	-0.042927	-0.024571	-0.019479	-0.031873	-0.079766
15	0.191803	0.002140	0.000678	0.085621	-0.018391	-0.019235	0.035982
16	-0.186832	-0.056908	-0.078343	-0.186813	-0.058392	-0.081706	-0.149652
17	-0.057750	-0.003240	-0.038228	-0.014818	0.002996	-0.021360	-0.038583
18	-0.127829	0.028800	0.060426	-0.140528	0.001892	0.011178	-0.117673
19	-0.127546	0.212890	0.069285	-0.105719	0.354861	0.162641	-0.107952
20	0.044302	-0.005226	-0.025818	0.025804	-0.003795	-0.012105	0.006627
21	-0.122235	0.054226	0.023287	-0.133842	0.045778	0.037200	-0.125405
22	0.071487	0.001660	-0.010291	0.032116	0.000741	0.003441	0.025216
23	-0.089114	-0.027748	-0.053597	-0.064053	-0.028246	-0.040597	-0.065956
24	-0.032166	-0.007253	-0.035733	0.007916	-0.006832	-0.012249	-0.016255
25	-0.181604	-0.057720	-0.076454	-0.182239	-0.059737	-0.080377	-0.129295

25 rows × 60 columns

In [9]:

```
mydata_transpose = mydata.transpose()
mydata_transpose.head()
```

Out[9]:

RECORD	1	2	3	4	5	6	7
fv_la_z3	-0.017442	0.041219	0.173077	-0.126544	-0.185336	0.092040	-0.145741
vl_la_z3	-0.006894	-0.007313	-0.057336	-0.052837	-0.053291	0.010679	-0.039019
vt_la_z3	-0.035573	-0.023622	-0.080118	-0.077530	-0.077321	-0.005715	0.018256
fv_la_z5	0.039096	0.031649	0.173978	-0.127203	-0.186301	0.048598	-0.134313
vl_la_z5	-0.001195	-0.017204	-0.059177	-0.054533	-0.055001	-0.001939	-0.034903

5 rows × 1048575 columns

In [10]:

```
mydata.describe()
```

Out[10]:

	fv_la_z3	vl_la_z3	vt_la_z3	fv_la_z5	vl_la_z5	vt_l
count	1.048575e+06	1.048575e+06	1.048575e+06	1.048575e+06	1.048575e+06	1.04
mean	3.060097e-17	-2.210921e-17	1.850339e-17	-6.476865e-17	1.034353e-17	-8.0 19
std	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.00
min	-1.878308e-01	-5.857795e-02	-8.083236e-02	-1.888085e-01	-6.045835e-02	-8.3 02
25%	-1.147560e-01	-4.308624e-02	-5.781444e-02	-1.101547e-01	-4.328888e-02	-5.9 02
50%	-2.695101e-02	-1.952813e-02	-3.772837e-02	-1.869088e-02	-1.809100e-02	-3.3 02
75%	4.210179e-02	-7.795125e-04	-1.631675e-02	3.964738e-02	-2.969563e-05	-9.3 03
max	4.698105e+02	7.180752e+02	6.741482e+02	4.722559e+02	7.411263e+02	6.99

8 rows × 60 columns

In [12]:

```
mydata_transpose.shape
```

Out[12]:

```
(60, 1048575)
```

In [13]:

```
%%time
covariance = mydata_transpose.dot(mydata)/(len(mydata)-1)
```

```
CPU times: user 264 ms, sys: 19.3 ms, total: 283 ms
Wall time: 321 ms
```

In [14]:

```
print(covariance)
```

```
              fv_la_z3  vl_la_z3  vt_la_z3  fv_la_z5  vl_la_z5  vt_l
a_z5  \
fv_la_z3      1.000000  0.661558  0.703109  0.984641  0.655264  0.69
7466
vl_la_z3      0.661558  1.000000  0.896382  0.658099  0.988189  0.89
8941
vt_la_z3      0.703109  0.896382  1.000000  0.690665  0.872938  0.97
3475
fv_la_z5      0.984641  0.658099  0.690665  1.000000  0.664654  0.70
6027
vl_la_z5      0.655264  0.988189  0.872938  0.664654  1.000000  0.90
3967
vt_la_z5      0.697466  0.898941  0.973475  0.706027  0.903967  1.00
0000
fv_la_tc      0.929804  0.468063  0.488307  0.928143  0.471315  0.49
8382
vl_la_tc      0.898276  0.603545  0.525884  0.902470  0.609779  0.54
1689
vt_la_tc      0.919974  0.449050  0.464041  0.920059  0.452272  0.47
2782
fv_la_bo      0.980145  0.578312  0.631494  0.975604  0.577150  0.63
4613
vl_la_bo      0.659617  0.965445  0.853572  0.660410  0.971275  0.87
1254
vt_la_bo      0.679287  0.840405  0.932289  0.676686  0.841361  0.93
9203
fv_la_none    0.981100  0.640188  0.673661  0.974729  0.640611  0.67
9224
vl_la_none    0.665680  0.963388  0.856543  0.667762  0.968534  0.87
1974
vt_la_none    0.698902  0.842603  0.927386  0.696030  0.839643  0.92
7041
fv_ba_z3      0.064839  0.101390  0.065028  0.059791  0.100733  0.06
4093
```

```
vl_ba_z3      0.058533  0.143621  0.092611  0.057661  0.146131  0.09
8137
vt_ba_z3      0.048991  0.118709  0.091145  0.048145  0.122055  0.09
8367
fv_ba_z5      0.076409  0.126589  0.081066  0.086353  0.130773  0.08
5693
vl_ba_z5      0.084968  0.207339  0.127010  0.087405  0.212945  0.13
4191
vt_ba_z5      0.080276  0.188089  0.126097  0.084135  0.193656  0.13
4476
fv_ba_tc      0.026074  0.041777  0.027894  0.025560  0.041855  0.02
7986
vl_ba_tc      0.013590  0.028128  0.017955  0.013579  0.028465  0.01
8827
vt_ba_tc      0.024286  0.043439  0.032982  0.024147  0.044149  0.03
4835
fv_ba_bo      0.055685  0.090210  0.057429  0.053227  0.090081  0.05
7393
vl_ba_bo      0.082890  0.202461  0.122705  0.082550  0.207192  0.12
8429
vt_ba_bo      0.056771  0.136790  0.093745  0.056717  0.140632  0.09
9397
fv_ba_none    0.051533  0.085357  0.053972  0.049337  0.085388  0.05
4131
vl_ba_none    0.049897  0.116467  0.073521  0.048807  0.117714  0.07
6756
vt_ba_none    0.046305  0.099666  0.073532  0.045801  0.101195  0.07
7677
fv_bv_z3      0.033118  0.045648  0.028469  0.028533  0.044024  0.02
7805
vl_bv_z3      0.022681  0.060382  0.039915  0.022040  0.059909  0.04
2630
vt_bv_z3      0.021965  0.057712  0.045098  0.021011  0.058099  0.04
9173
fv_bv_z5      0.038927  0.057070  0.035434  0.045084  0.058817  0.03
8294
vl_bv_z5      0.039655  0.102997  0.065177  0.041020  0.103964  0.06
9628
vt_bv_z5      0.039152  0.098321  0.068477  0.041300  0.099781  0.07
4146
fv_bv_tc      0.003745  0.007164  0.004595  0.003521  0.007042  0.00
4596
vl_bv_tc      0.002593  0.006306  0.004031  0.002491  0.006258  0.00
4203
vt_bv_tc      0.003478  0.007818  0.006014  0.003343  0.007855  0.00
6452
fv_bv_bo      0.025995  0.039945  0.024205  0.023713  0.038996  0.02
4348
vl_bv_bo      0.028627  0.074642  0.046359  0.028341  0.075214  0.04
9057
vt_bv_bo      0.025068  0.065778  0.047232  0.024704  0.066644  0.05
1003
fv_bv_none    0.021448  0.038052  0.022911  0.019959  0.037478  0.02
```

```
3352
vl_bv_none     0.017007  0.044621  0.029103  0.016460  0.044447  0.03
0714
vt_bv_none     0.018051  0.047336  0.036393  0.017509  0.047713  0.03
9503
fv_none_z3     0.177369  0.359698  0.237394  0.173296  0.364614  0.23
8436
vl_none_z3     0.129240  0.332888  0.228778  0.129624  0.343199  0.24
8764
vt_none_z3     0.058500  0.153041  0.135912  0.058131  0.160115  0.15
1174
fv_none_z5     0.177101  0.382789  0.245188  0.186077  0.393398  0.25
7086
vl_none_z5     0.220313  0.546208  0.330333  0.224899  0.560332  0.34
9769
vt_none_z5     0.147064  0.351955  0.239068  0.152544  0.360511  0.25
5764
fv_none_tc     0.171417  0.239643  0.169878  0.171277  0.240631  0.16
9436
vl_none_tc     0.164677  0.324901  0.211442  0.170984  0.331049  0.22
3794
vt_none_tc     0.124498  0.193677  0.154744  0.126082  0.197333  0.16
2478
fv_none_bo     0.172195  0.363638  0.233462  0.170773  0.370171  0.23
6271
vl_none_bo     0.204808  0.509954  0.318523  0.205671  0.525670  0.33
7546
vt_none_bo     0.113062  0.281339  0.204539  0.113341  0.291843  0.22
0264
fv_none_none   0.176130  0.304244  0.206157  0.172319  0.303796  0.20
3325
vl_none_none   0.171358  0.383553  0.245093  0.169557  0.388324  0.25
5891
vt_none_none   0.129092  0.248747  0.191687  0.127179  0.252119  0.19
9821

               fv_la_tc  vl_la_tc  vt_la_tc  fv_la_bo       ...
\
fv_la_z3       0.929804  0.898276  0.919974  0.980145       ...
vl_la_z3       0.468063  0.603545  0.449050  0.578312       ...
vt_la_z3       0.488307  0.525884  0.464041  0.631494       ...
fv_la_z5       0.928143  0.902470  0.920059  0.975604       ...
vl_la_z5       0.471315  0.609779  0.452272  0.577150       ...
vt_la_z5       0.498382  0.541689  0.472782  0.634613       ...
fv_la_tc       1.000000  0.936945  0.987323  0.936251       ...
vl_la_tc       0.936945  1.000000  0.948862  0.894860       ...
vt_la_tc       0.987323  0.948862  1.000000  0.937089       ...
fv_la_bo       0.936251  0.894860  0.937089  1.000000       ...
vl_la_bo       0.476802  0.610570  0.459600  0.596221       ...
vt_la_bo       0.476782  0.510678  0.454797  0.639534       ...
fv_la_none     0.942226  0.908246  0.927906  0.969087       ...
vl_la_none     0.507356  0.644955  0.485268  0.586057       ...
vt_la_none     0.543733  0.572528  0.515408  0.633022       ...
```

```
fv_ba_z3      0.052233  0.076318  0.045395  0.045743      ...
vl_ba_z3      0.040643  0.083448  0.039059  0.041583      ...
vt_ba_z3      0.035494  0.067250  0.036056  0.034754      ...
fv_ba_z5      0.064155  0.094082  0.056220  0.055799      ...
vl_ba_z5      0.057868  0.117482  0.054340  0.060265      ...
vt_ba_z5      0.058284  0.108910  0.055732  0.057474      ...
fv_ba_tc      0.029095  0.040504  0.025956  0.019423      ...
vl_ba_tc      0.013284  0.024728  0.012729  0.010000      ...
vt_ba_tc      0.027139  0.040423  0.026460  0.017990      ...
fv_ba_bo      0.047218  0.069076  0.040952  0.041490      ...
vl_ba_bo      0.056227  0.115945  0.052748  0.059306      ...
vt_ba_bo      0.040198  0.077828  0.039092  0.040913      ...
fv_ba_none    0.043315  0.064168  0.037581  0.038348      ...
vl_ba_none    0.037056  0.074024  0.035130  0.036094      ...
vt_ba_none    0.038489  0.066048  0.037535  0.033846      ...
fv_bv_z3      0.026459  0.036462  0.021625  0.022914      ...
vl_bv_z3      0.015340  0.033091  0.014970  0.015767      ...
vt_bv_z3      0.015364  0.031138  0.015984  0.015184      ...
fv_bv_z5      0.033211  0.046559  0.027324  0.028352      ...
vl_bv_z5      0.026689  0.056119  0.025414  0.027596      ...
vt_bv_z5      0.027833  0.054500  0.027145  0.027515      ...
fv_bv_tc      0.003814  0.006250  0.003361  0.002780      ...
vl_bv_tc      0.002216  0.005047  0.002233  0.001887      ...
vt_bv_tc      0.003513  0.006460  0.003639  0.002557      ...
fv_bv_bo      0.021910  0.031332  0.017919  0.019243      ...
vl_bv_bo      0.019171  0.041134  0.018214  0.020322      ...
vt_bv_bo      0.017496  0.035923  0.017541  0.017936      ...
fv_bv_none    0.015970  0.025391  0.013304  0.015763      ...
vl_bv_none    0.011866  0.025682  0.011496  0.012025      ...
vt_bv_none    0.013276  0.026792  0.013651  0.012919      ...
fv_none_z3    0.130290  0.220862  0.119799  0.125799      ...
vl_none_z3    0.087051  0.180291  0.086395  0.092120      ...
vt_none_z3    0.041259  0.079055  0.045113  0.041662      ...
fv_none_z5    0.131102  0.230262  0.120748  0.127454      ...
vl_none_z5    0.146547  0.298867  0.136588  0.156814      ...
vt_none_z5    0.103921  0.193878  0.098457  0.105560      ...
fv_none_tc    0.198375  0.246716  0.177645  0.125694      ...
vl_none_tc    0.163173  0.279493  0.155155  0.120661      ...
vt_none_tc    0.144650  0.190953  0.139715  0.090776      ...
fv_none_bo    0.125246  0.219473  0.115117  0.124240      ...
vl_none_bo    0.135400  0.280384  0.128730  0.146260      ...
vt_none_bo    0.076872  0.151366  0.076770  0.081220      ...
fv_none_none  0.153305  0.226163  0.140085  0.127359      ...
vl_none_none  0.130704  0.249337  0.123653  0.124241      ...
vt_none_none  0.111922  0.172787  0.108203  0.092851      ...

              vt_none_z5  fv_none_tc  vl_none_tc  vt_none_tc  fv_non
e_bo  \
fv_la_z3        0.147064    0.171417    0.164677    0.124498    0.17
2195
vl_la_z3        0.351955    0.239643    0.324901    0.193677    0.36
3638
vt_la_z3        0.239068    0.169878    0.211442    0.154744    0.23
```

```
3462
fv_la_z5        0.152544    0.171277    0.170984    0.126082    0.17
0773
vl_la_z5        0.360511    0.240631    0.331049    0.197333    0.37
0171
vt_la_z5        0.255764    0.169436    0.223794    0.162478    0.23
6271
fv_la_tc        0.103921    0.198375    0.163173    0.144650    0.12
5246
vl_la_tc        0.193878    0.246716    0.279493    0.190953    0.21
9473
vt_la_tc        0.098457    0.177645    0.155155    0.139715    0.11
5117
fv_la_bo        0.105560    0.125694    0.120661    0.090776    0.12
4240
vl_la_bo        0.347217    0.236449    0.322078    0.184392    0.37
2947
vt_la_bo        0.203134    0.144753    0.182989    0.120254    0.21
2548
fv_la_none      0.148545    0.212992    0.191548    0.152392    0.18
0246
vl_la_none      0.349898    0.309520    0.389226    0.238359    0.39
0907
vt_la_none      0.225827    0.260304    0.262614    0.205255    0.25
6394
fv_ba_z3        0.144124    0.197882    0.213340    0.158151    0.17
9203
vl_ba_z3        0.238215    0.155190    0.269535    0.217059    0.18
9976
vt_ba_z3        0.291853    0.143589    0.297789    0.311022    0.18
7771
fv_ba_z5        0.182741    0.215023    0.224653    0.175128    0.20
3812
vl_ba_z5        0.288181    0.174473    0.266425    0.192352    0.24
8031
vt_ba_z5        0.338106    0.193214    0.294584    0.255227    0.26
3227
fv_ba_tc        0.061399    0.109101    0.112994    0.096109    0.07
6923
vl_ba_tc        0.044006    0.059692    0.084362    0.072777    0.04
4232
vt_ba_tc        0.094764    0.123568    0.172215    0.179730    0.08
3849
fv_ba_bo        0.131524    0.178874    0.193723    0.142483    0.16
3019
vl_ba_bo        0.246973    0.171774    0.267966    0.188909    0.23
0272
vt_ba_bo        0.241793    0.136349    0.242186    0.226566    0.18
7100
fv_ba_none      0.121326    0.161376    0.178061    0.129997    0.14
7475
vl_ba_none      0.180218    0.158429    0.247699    0.180319    0.16
5744
```

```
vt_ba_none       0.212809     0.159505     0.265803     0.244817     0.16
4979
fv_bv_z3         0.086636     0.111210     0.124424     0.086996     0.09
7175
vl_bv_z3         0.124060     0.071579     0.129677     0.105616     0.08
5043
vt_bv_z3         0.170730     0.078595     0.166228     0.170161     0.10
1247
fv_bv_z5         0.108351     0.122879     0.131735     0.096566     0.11
0849
vl_bv_z5         0.184328     0.100594     0.165155     0.125466     0.13
2503
vt_bv_z5         0.231871     0.118934     0.191853     0.174334     0.16
0498
fv_bv_tc         0.013786     0.019088     0.021300     0.018481     0.01
4613
vl_bv_tc         0.012564     0.015210     0.023157     0.021908     0.01
0834
vt_bv_tc         0.022401     0.024057     0.037831     0.041379     0.01
6982
fv_bv_bo         0.075374     0.091176     0.104347     0.071244     0.08
1000
vl_bv_bo         0.119650     0.070202     0.117278     0.086804     0.09
2947
vt_bv_bo         0.158270     0.076727     0.149160     0.143291     0.10
3954
fv_bv_none       0.074086     0.069407     0.078850     0.057029     0.07
5091
vl_bv_none       0.091459     0.057641     0.100011     0.077605     0.06
6253
vt_bv_none       0.138276     0.069348     0.139726     0.137758     0.08
5778
fv_none_z3       0.871997     0.760034     0.609315     0.630714     0.98
7465
vl_none_z3       0.688792     0.336202     0.652159     0.643435     0.47
9836
vt_none_z3       0.636395     0.263236     0.502289     0.640302     0.39
8771
fv_none_z5       0.923738     0.710192     0.574710     0.588053     0.96
3880
vl_none_z5       0.795634     0.474670     0.611989     0.456116     0.72
6316
vt_none_z5       1.000000     0.566948     0.539684     0.584984     0.86
7003
fv_none_tc       0.566948     1.000000     0.802419     0.827169     0.72
7778
vl_none_tc       0.539684     0.802419     1.000000     0.877501     0.56
8595
vt_none_tc       0.584984     0.827169     0.877501     1.000000     0.58
6632
fv_none_bo       0.867003     0.727778     0.568595     0.586632     1.00
0000
vl_none_bo       0.740297     0.425269     0.682415     0.574441     0.64
```

```
9270
vt_none_bo       0.800650    0.406473    0.572758    0.676196    0.65
0577
fv_none_none     0.692869    0.859811    0.714075    0.672871    0.86
3318
vl_none_none     0.613231    0.615334    0.855667    0.667400    0.61
2909
vt_none_none     0.727611    0.643948    0.745097    0.797583    0.67
5812

              vl_none_bo  vt_none_bo  fv_none_none  vl_none_none  vt
_none_none
fv_la_z3         0.204808    0.113062      0.176130      0.171358
0.129092
vl_la_z3         0.509954    0.281339      0.304244      0.383553
0.248747
vt_la_z3         0.318523    0.204539      0.206157      0.245093
0.191687
fv_la_z5         0.205671    0.113341      0.172319      0.169557
0.127179
vl_la_z5         0.525670    0.291843      0.303796      0.388324
0.252119
vt_la_z5         0.337546    0.220264      0.203325      0.255891
0.199821
fv_la_tc         0.135400    0.076872      0.153305      0.130704
0.111922
vl_la_tc         0.280384    0.151366      0.226163      0.249337
0.172787
vt_la_tc         0.128730    0.076770      0.140085      0.123653
0.108203
fv_la_bo         0.146260    0.081220      0.127359      0.124241
0.092851
vl_la_bo         0.523810    0.279906      0.303220      0.383251
0.239193
vt_la_bo         0.289066    0.168485      0.179623      0.214809
0.151295
fv_la_none       0.204348    0.113486      0.206639      0.189286
0.149812
vl_la_none       0.517011    0.277704      0.377937      0.444452
0.294164
vt_la_none       0.305101    0.179987      0.289978      0.282159
0.231727
fv_ba_z3         0.165770    0.104816      0.220427      0.227081
0.168080
vl_ba_z3         0.320880    0.274317      0.197155      0.318228
0.279695
vt_ba_z3         0.390634    0.431346      0.179142      0.351810
0.402510
fv_ba_z5         0.200096    0.127471      0.225359      0.224477
0.173946
vl_ba_z5         0.365697    0.257226      0.217996      0.309654
0.241651
vt_ba_z5         0.393817    0.337406      0.235587      0.339720
```

```
0.319239
fv_ba_tc        0.067803    0.043023       0.093167      0.092010
0.069156
vl_ba_tc        0.057934    0.045497       0.052136      0.069412
0.055521
vt_ba_tc        0.122014    0.123694       0.096455      0.134359
0.136265
fv_ba_bo        0.149345    0.094771       0.199892      0.206840
0.152225
vl_ba_bo        0.346686    0.244831       0.220361      0.318323
0.244672
vt_ba_bo        0.326642    0.312636       0.171899      0.286788
0.293265
fv_ba_none      0.140535    0.087786       0.179299      0.189064
0.137004
vl_ba_none      0.245181    0.193207       0.201568      0.292433
0.232269
vt_ba_none      0.284429    0.289621       0.197274      0.311569
0.313640
fv_bv_z3        0.081493    0.054456       0.126413      0.135580
0.095401
vl_bv_z3        0.145653    0.132238       0.092577      0.154017
0.137805
vt_bv_z3        0.205400    0.234529       0.101083      0.198014
0.222879
fv_bv_z5        0.096836    0.066743       0.130453      0.134355
0.099804
vl_bv_z5        0.202138    0.161603       0.128305      0.194198
0.161570
vt_bv_z5        0.237880    0.232369       0.150221      0.225280
0.224694
fv_bv_tc        0.012040    0.008296       0.014962      0.014871
0.010857
vl_bv_tc        0.014228    0.012195       0.011209      0.015630
0.013185
vt_bv_tc        0.025886    0.028519       0.016483      0.025924
0.027885
fv_bv_bo        0.068695    0.045174       0.104677      0.114342
0.078832
vl_bv_bo        0.144789    0.112912       0.091091      0.139370
0.113448
vt_bv_bo        0.189592    0.198491       0.099119      0.177672
0.187828
fv_bv_none      0.067226    0.046169       0.078786      0.083807
0.060118
vl_bv_none      0.106544    0.093003       0.074539      0.118407
0.101177
vt_bv_none      0.165079    0.184726       0.089237      0.166099
0.180333
fv_none_z3      0.649121    0.652012       0.884135      0.641566
0.705884
vl_none_z3      0.903358    0.921200       0.421323      0.770132
0.831260
```

```
vt_none_z3      0.711115    0.941938      0.325008      0.593286
0.830450
fv_none_z5      0.673143    0.660666      0.818964      0.605231
0.661253
vl_none_z5      0.900224    0.638287      0.590669      0.708766
0.569478
vt_none_z5      0.740297    0.800650      0.692869      0.613231
0.727611
fv_none_tc      0.425269    0.406473      0.859811      0.615334
0.643948
vl_none_tc      0.682415    0.572758      0.714075      0.855667
0.745097
vt_none_tc      0.574441    0.676196      0.672871      0.667400
0.797583
fv_none_bo      0.649270    0.650577      0.863318      0.612909
0.675812
vl_none_bo      1.000000    0.835656      0.541784      0.810951
0.744611
vt_none_bo      0.835656    1.000000      0.511617      0.678138
0.877525
fv_none_none    0.541784    0.511617      1.000000      0.767331
0.772338
vl_none_none    0.810951    0.678138      0.767331      1.000000
0.851030
vt_none_none    0.744611    0.877525      0.772338      0.851030
1.000000

[60 rows x 60 columns]
```

In [15]:

```
# %%time
# eig_vals, eig_vecs = np.linalg.eig(covariance)
```

In [16]:

```
# print('Eigenvectors \n%s' %eig_vecs)
```

In [17]:

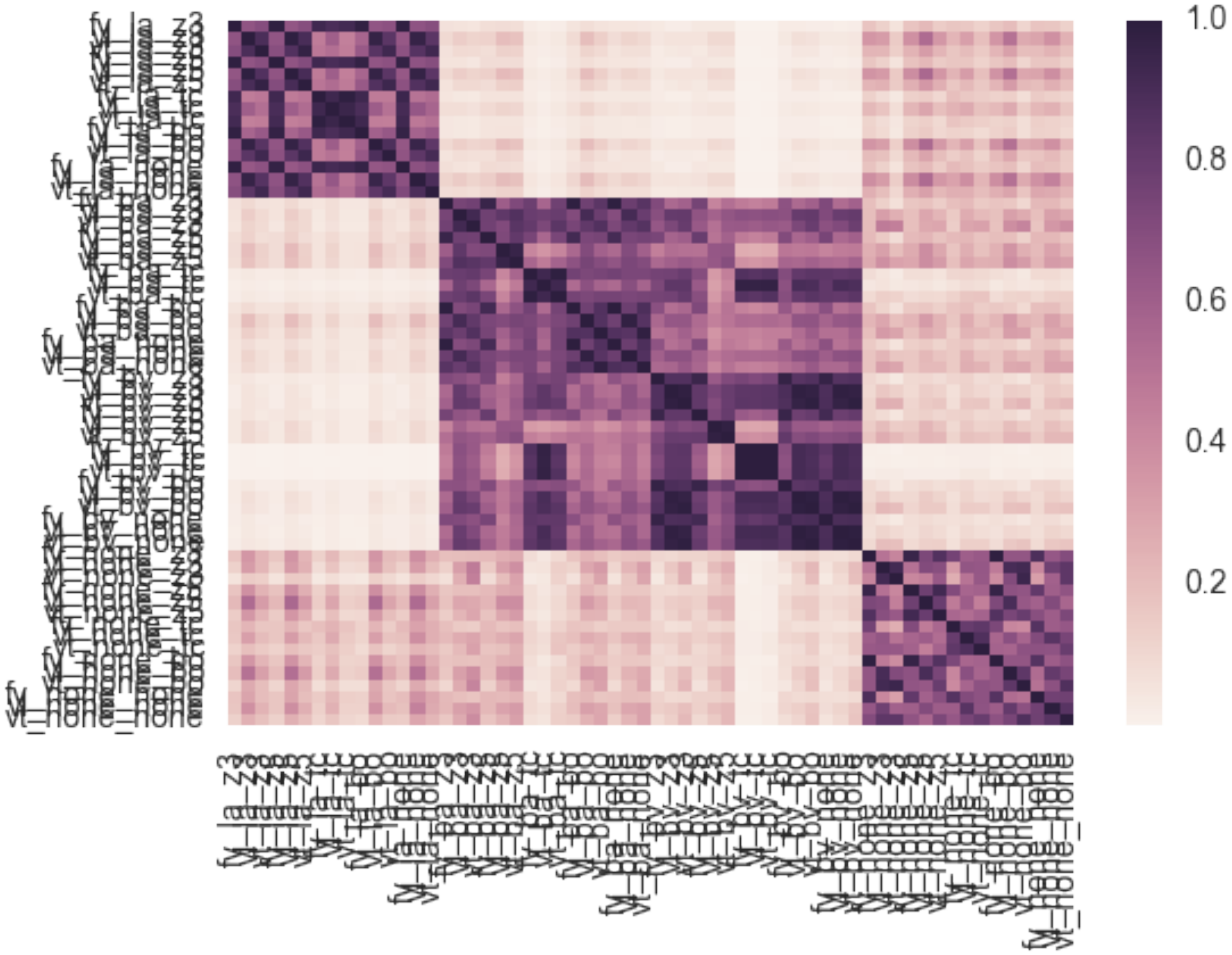
```
# print('Eigenvalues \n%s' %eig_vals)
```

In [18]:

```
sns.heatmap(covariance)
```

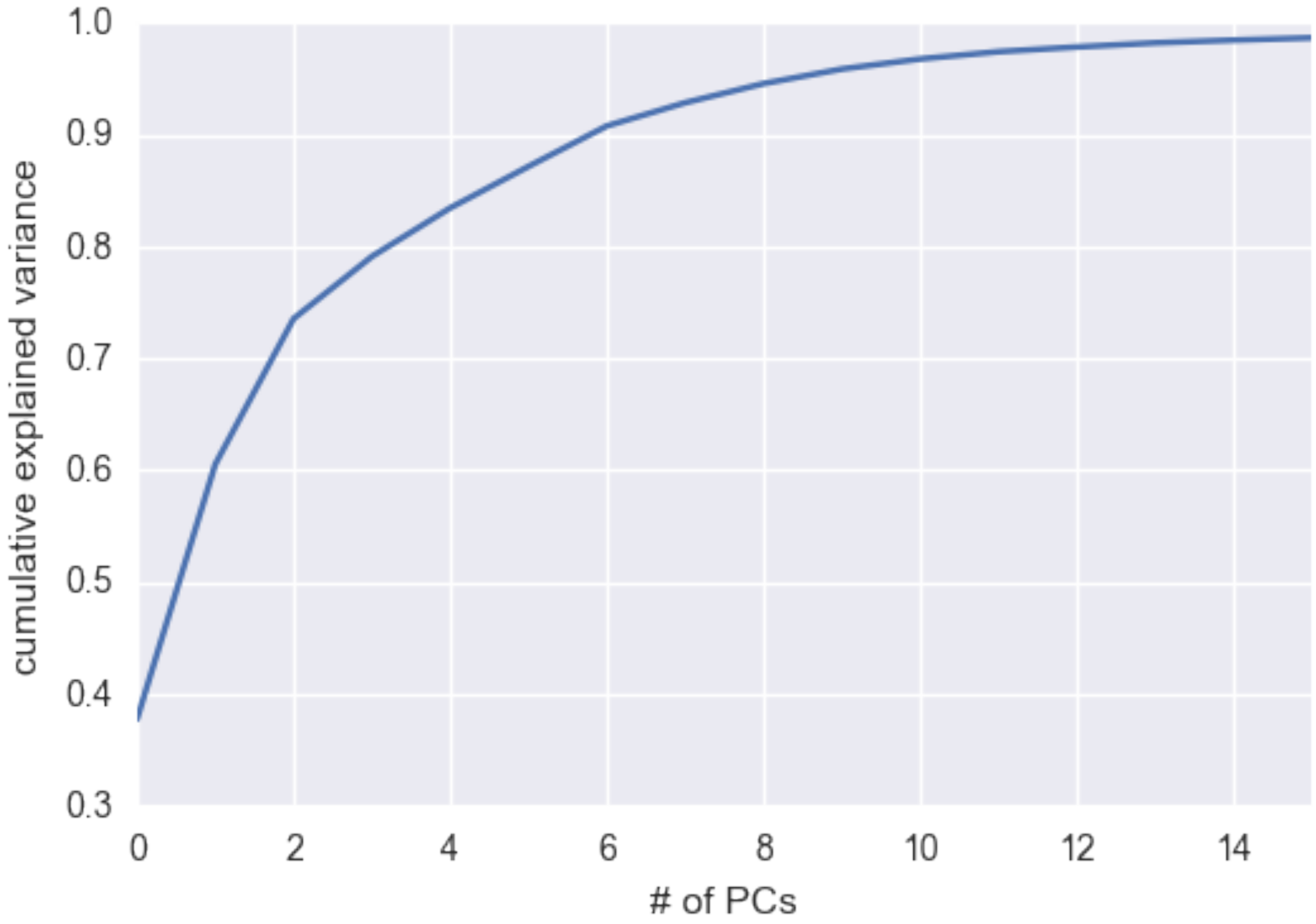
Out[18]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x131c809b0>
```



In [19]:

```
%%time
pca = PCA(n_components=50).fit(mydata)
plt.plot(np.cumsum(pca.explained_variance_ratio_))
plt.xlabel('# of PCs')
plt.ylabel('cumulative explained variance')
plt.xlim(xmax=15)
plt.ylim(ymax=1)
plt.show()
```



```
CPU times: user 8.39 s, sys: 2.78 s, total: 11.2 s
Wall time: 8.78 s
```

In [20]:

```
%%time
NPCs = 10
PCs = pd.DataFrame(PCA(n_components=NPCs).fit_transform(mydata))
```

```
CPU times: user 14.3 s, sys: 5.65 s, total: 19.9 s
Wall time: 13.3 s
```

In [21]:

```
PCs.shape
```

Out[21]:

```
(1048575, 10)
```

In [22]:

```
PCs.head()
```

Out[22]:

	0	1	2	3	4	5	6	7
0	-0.034264	-0.085701	0.071586	0.029461	0.019315	0.033291	0.010059	-0.01281
1	-0.077732	-0.081204	0.019688	0.036318	0.013258	-0.001528	0.035214	-0.03249
2	-0.266122	-0.057719	0.223981	-0.026517	0.132294	0.140012	0.171081	-0.16995
3	-0.314207	-0.293961	-0.059221	-0.030026	-0.164627	0.034447	0.135383	-0.10139
4	-0.125387	-0.358425	-0.204000	-0.045647	-0.201698	0.046472	0.039459	0.040546

In [23]:

```
%%time
PCs_zscale = (PCs - PCs.mean())/ PCs.std()
```

```
CPU times: user 529 ms, sys: 202 ms, total: 732 ms
Wall time: 480 ms
```

In [24]:

```
PCs_zscale.head()
```

Out[24]:

	0	1	2	3	4	5	6	7
0	-0.007197	-0.023181	0.025625	0.016139	0.011971	0.022319	0.006828	-0.0114
1	-0.016327	-0.021964	0.007048	0.019895	0.008217	-0.001024	0.023903	-0.029
2	-0.055897	-0.015612	0.080176	-0.014526	0.081994	0.093864	0.116126	-0.152
3	-0.065997	-0.079512	-0.021199	-0.016449	-0.102034	0.023094	0.091895	-0.090
4	-0.026337	-0.096948	-0.073024	-0.025006	-0.125009	0.031155	0.026784	0.036

In [25]:

```
PCs.describe()
```

Out[25]:

	0	1	2	3	4	5
count	1.048575e+06	1.048575e+06	1.048575e+06	1.048575e+06	1.048575e+06	1.0
mean	-3.875824e-16	5.998621e-17	4.646066e-17	2.754557e-17	4.269339e-17	-1.
std	4.760963e+00	3.697088e+00	2.793606e+00	1.825435e+00	1.613460e+00	1.4
min	-3.512390e-01	-1.369115e+03	-1.105806e+03	-8.284555e+02	-5.254451e+02	-4.
25%	-1.681646e-01	-2.223510e-01	-1.112228e-01	-2.895072e-02	-1.733281e-01	-1.
50%	-7.233993e-02	-1.148778e-01	-1.323046e-02	-2.949506e-03	-3.031621e-02	1.1
75%	-9.371534e-03	-9.002384e-03	7.212620e-02	2.789066e-02	6.037596e-02	2.4
max	3.362582e+03	2.091879e+03	1.223548e+03	9.599008e+02	7.627368e+02	8.2

In [26]:

```
PCs_zscale.describe()
```

Out[26]:

	0	1	2	3	4	5
count	1.048575e+06	1.048575e+06	1.048575e+06	1.048575e+06	1.048575e+06	1.0
mean	6.911923e-17	2.452717e-18	-3.321742e-17	-1.569507e-17	-1.031590e-17	-5.
std	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.0
min	-7.377478e-02	-3.703225e+02	-3.958347e+02	-4.538400e+02	-3.256635e+02	-2.
25%	-3.532156e-02	-6.014220e-02	-3.981333e-02	-1.585963e-02	-1.074263e-01	-7.
50%	-1.519439e-02	-3.107250e-02	-4.735977e-03	-1.615783e-03	-1.878956e-02	7.6
75%	-1.968412e-03	-2.434993e-03	2.581831e-02	1.527891e-02	3.742017e-02	1.6
max	7.062819e+02	5.658181e+02	4.379815e+02	5.258477e+02	4.727335e+02	5.5

In [27]:

```
Scores = pd.DataFrame(np.ones(numrecords), columns = ['s1'])
Scores['s2'] = np.ones(numrecords)
```

In [28]:

```
%%time
Scores['s1'] = PCs_zscale.abs().sum(axis=1)
PCs_zscale_sq = PCs_zscale **2
Scores['s2'] = PCs_zscale_sq.abs().sum(axis=1)
```

```
CPU times: user 1.25 s, sys: 91.1 ms, total: 1.34 s
Wall time: 557 ms
```

In [29]:

```
Scores.head(10)
```

Out[29]:

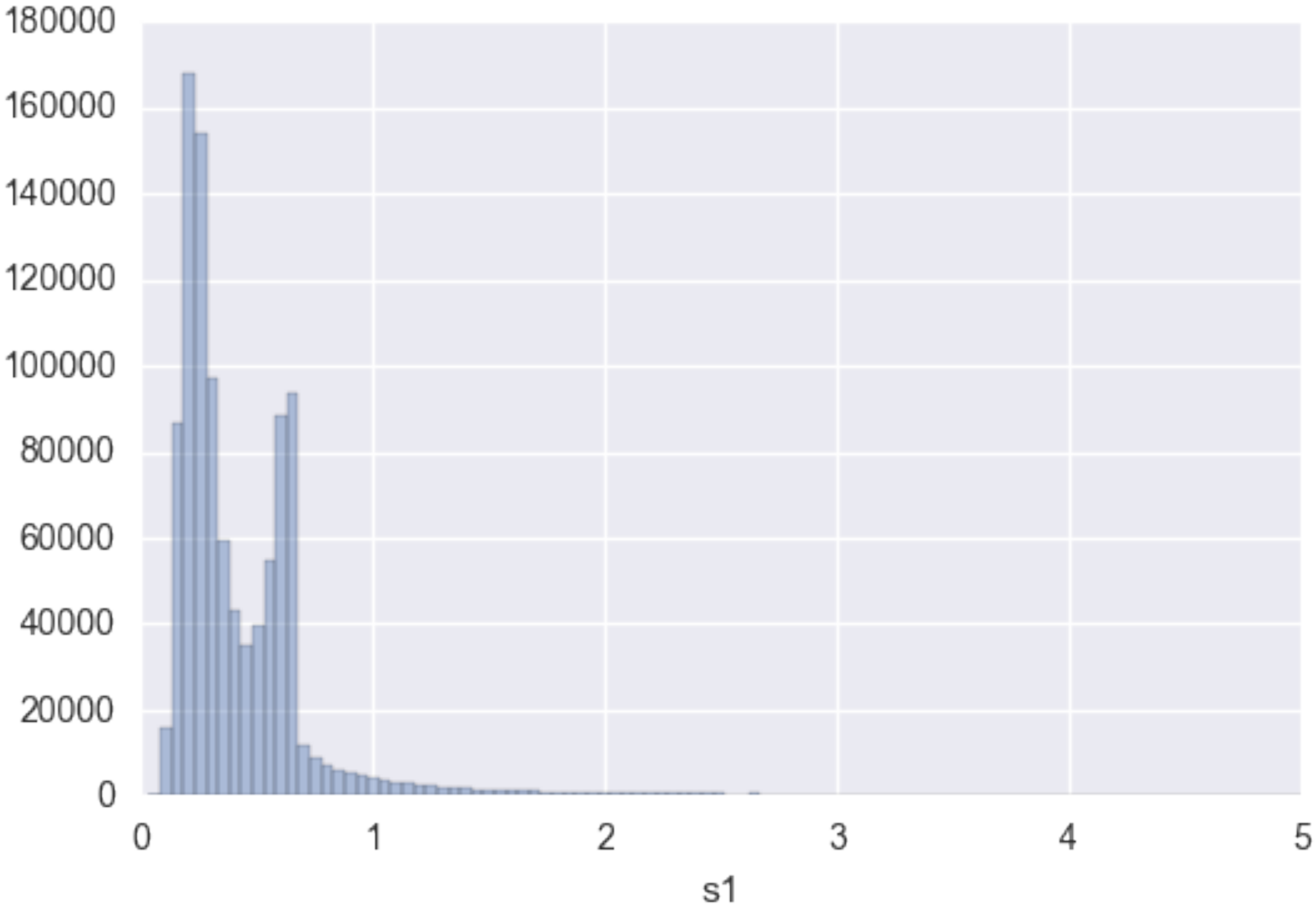
	s1	s2
0	0.214021	0.007214
1	0.166653	0.003586
2	0.623062	0.062306
3	0.555506	0.041120
4	0.586902	0.055246
5	0.256063	0.012991
6	0.672738	0.068247
7	0.173454	0.004553
8	0.157571	0.003098
9	0.171562	0.006851

In [30]:

```
xhigh = 5
sns.plt.xlim(0,xhigh)
temp = Scores[Scores['s1'] <= xhigh]
sns.distplot(temp['s1'], bins = 100, kde = False)
```

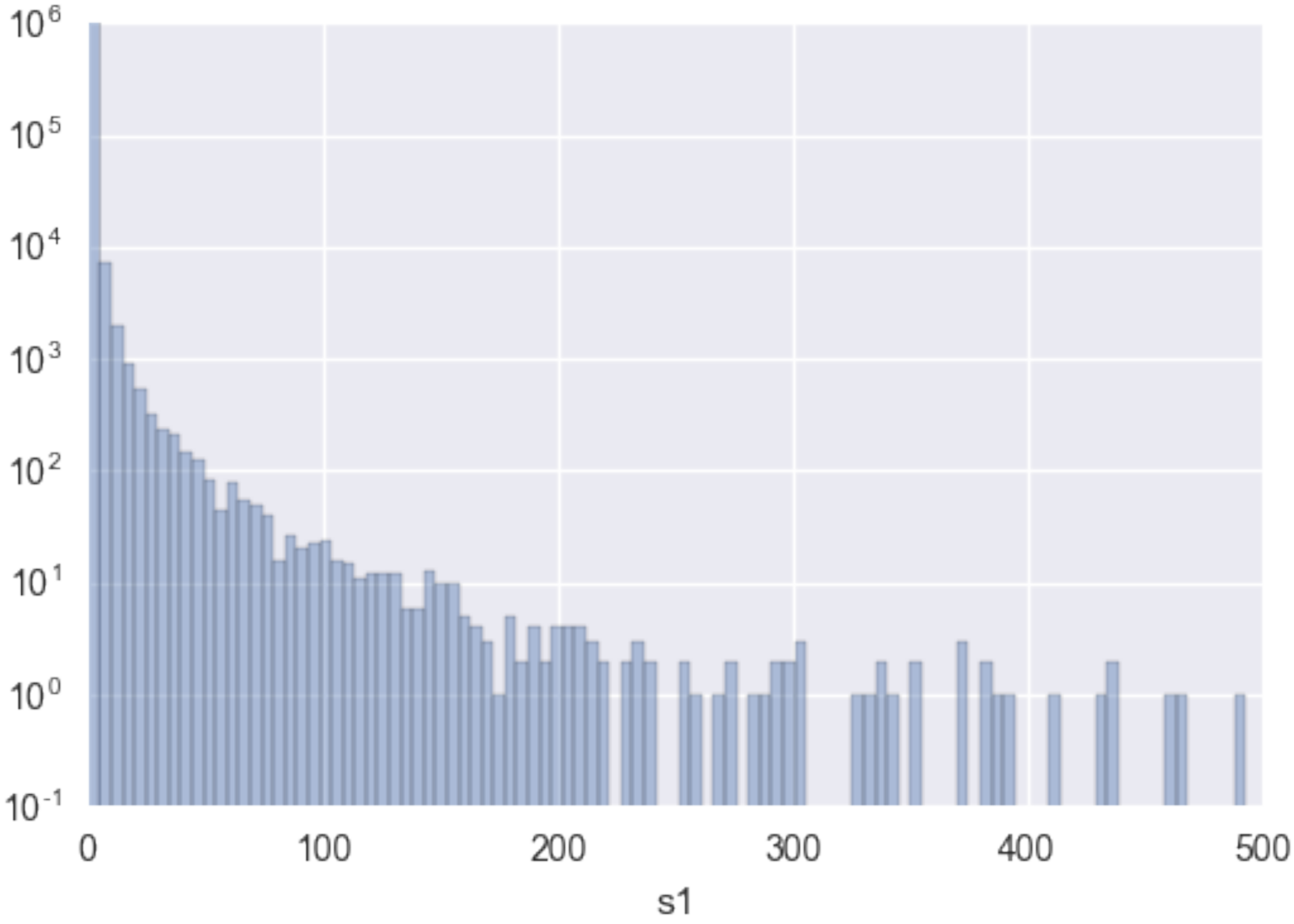
```
Out[30]:

<matplotlib.axes._subplots.AxesSubplot at 0x11b2114a8>
```


180000
160000
140000
120000
100000
80000
60000
40000
20000
0
0
1
2
3
4
5
s1

In [31]:

```
xhigh = 500
temp = Scores[Scores['s1'] <= xhigh]
sns.plt.xlim(0, xhigh)
sns.plt.ylim(.1, 10**6)
ax = sns.distplot(temp['s1'], bins = 100, kde=False)
ax.set_yscale('log')
plt.savefig('log.png')
```

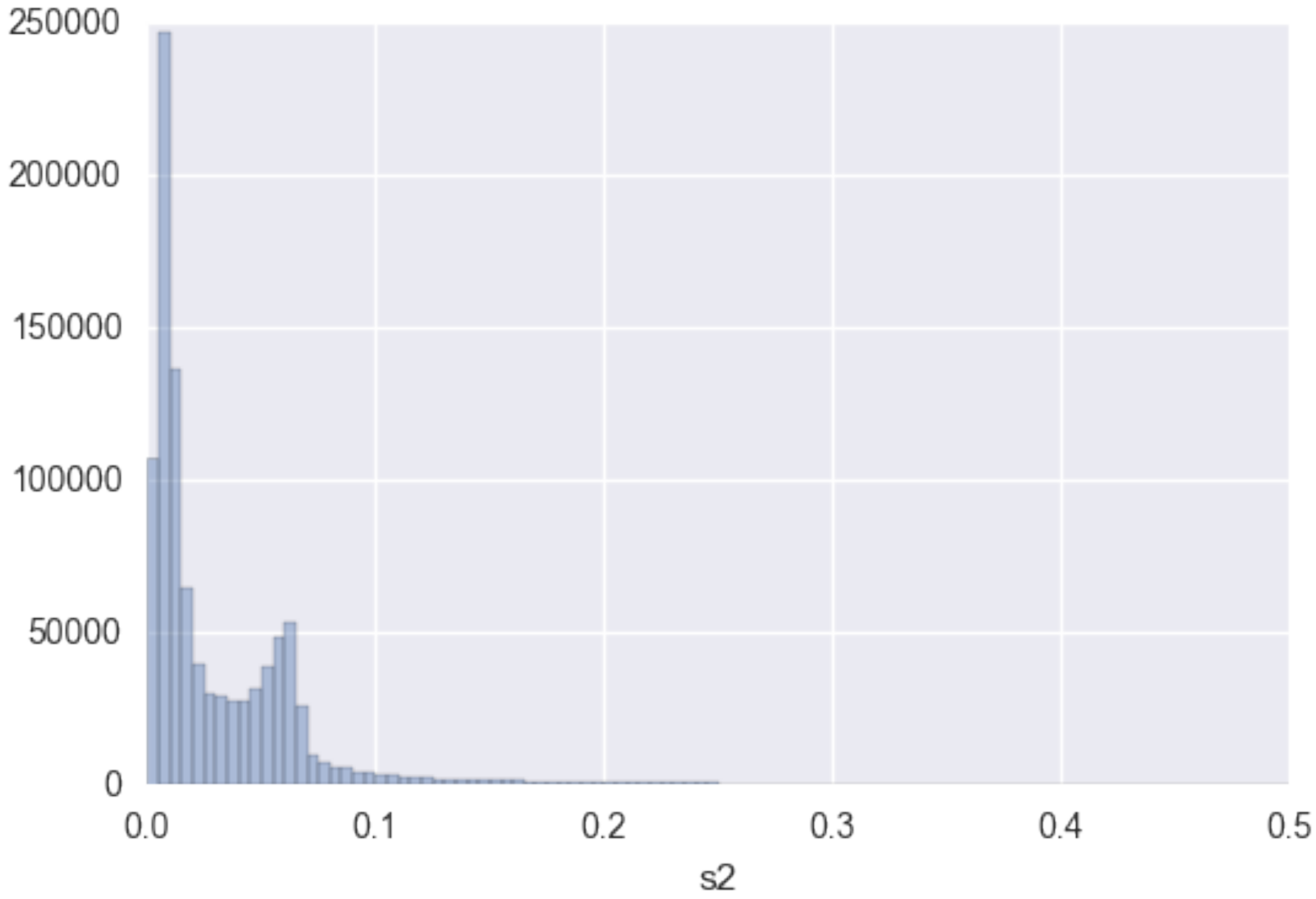


In [32]:

```
xhigh = .5
sns.plt.xlim(0,xhigh)
temp = Scores[Scores['s2'] <= xhigh]
sns.distplot(temp['s2'], bins = 100, kde = False)
```

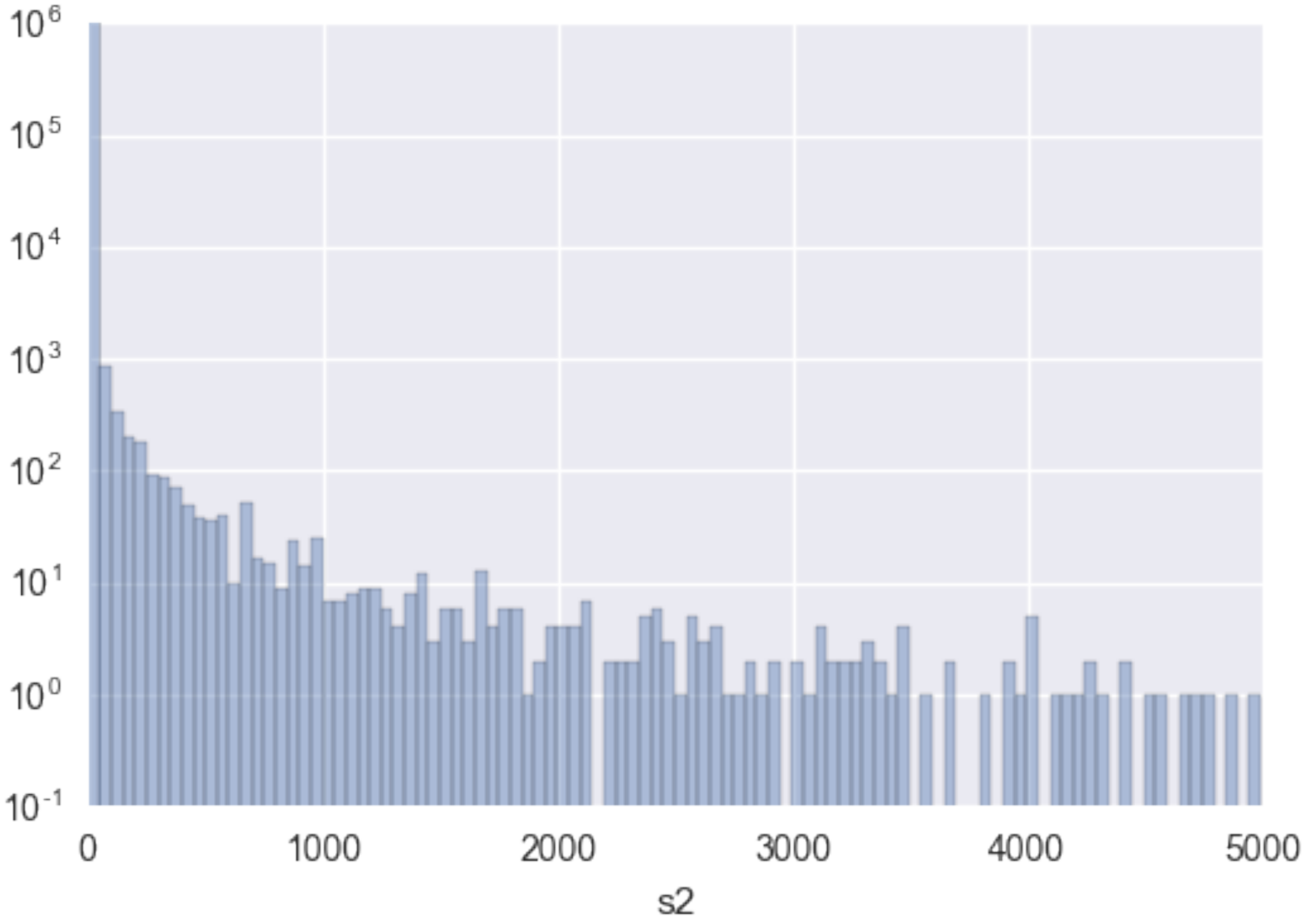
Out[32]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x11b10f278>
```



In [33]:

```
xhigh = 5000
temp = Scores[Scores['s2'] <= xhigh]
sns.plt.xlim(0, xhigh)
sns.plt.ylim(.1, 10**6)
ax = sns.distplot(temp['s2'], bins = 100, kde=False)
ax.set_yscale('log')
plt.savefig('log.png')
```



In [34]:

```
Scores['record'] = Scores.index + 1
```

In [35]:

```
Scores.head()
```

Out[35]:

	s1	s2	record
0	0.214021	0.007214	1
1	0.166653	0.003586	2
2	0.623062	0.062306	3
3	0.555506	0.041120	4
4	0.586902	0.055246	5

In [36]:

```
Scores.sort_values('s1').tail(12)
```

Out[36]:

	s1	s2	record
977470	1316.559485	3.603928e+05	977471
24585	1424.827559	3.637766e+05	24586
648674	1484.598645	3.929890e+05	648675
902255	1528.395442	4.163797e+05	902256
787891	1829.780582	5.960451e+05	787892
970080	1835.168673	6.397387e+05	970081
294060	1873.017687	4.426446e+05	294061
1046263	1910.958069	6.109203e+05	1046264
78803	2181.573331	7.841722e+05	78804
315452	2325.012043	6.381404e+05	315453
5392	2430.121225	1.028361e+06	5393
376242	2598.794418	1.020363e+06	376243

In [37]:

```
Scores.sort_values('s2').tail(12)
```

Out[37]:

	s1	s2	record
977470	1316.559485	3.603928e+05	977471
24585	1424.827559	3.637766e+05	24586
648674	1484.598645	3.929890e+05	648675
902255	1528.395442	4.163797e+05	902256
294060	1873.017687	4.426446e+05	294061
787891	1829.780582	5.960451e+05	787892
1046263	1910.958069	6.109203e+05	1046264
315452	2325.012043	6.381404e+05	315453
970080	1835.168673	6.397387e+05	970081
78803	2181.573331	7.841722e+05	78804
376242	2598.794418	1.020363e+06	376243
5392	2430.121225	1.028361e+06	5393

In [ ]: