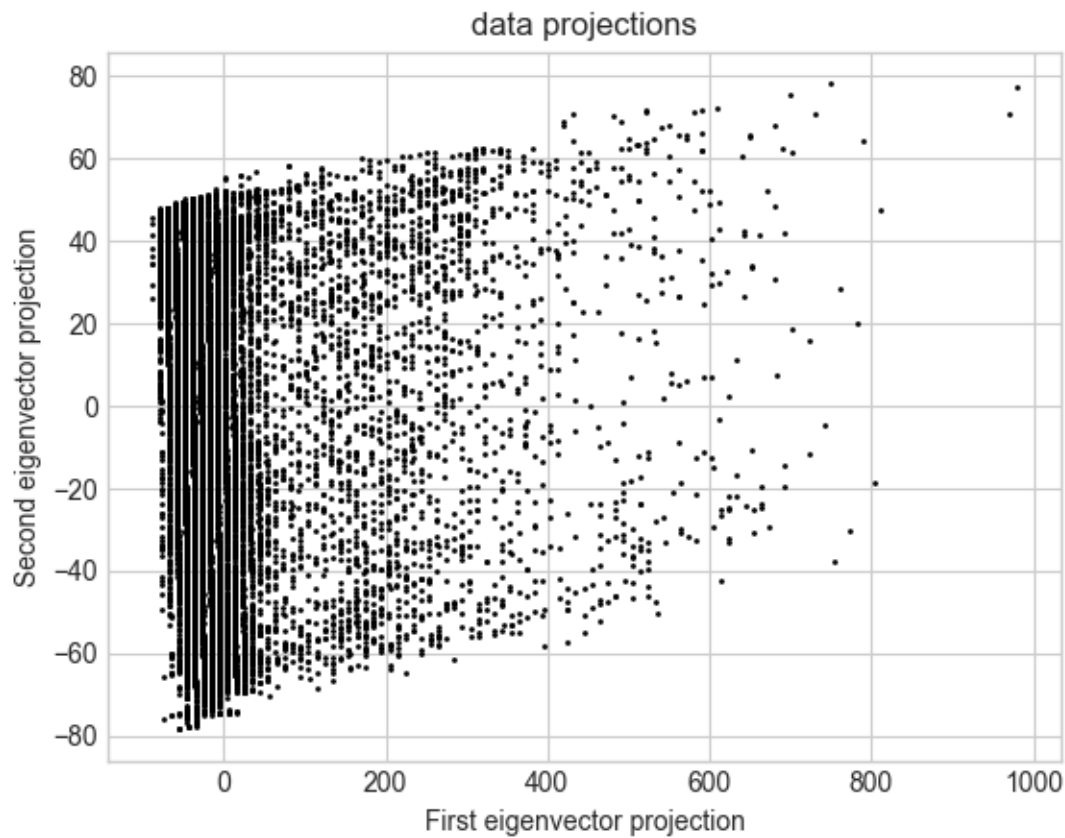Jae Park
Data Mining
HW2

$ python3 Assign2-part1.py energydata_complete.csv 0.975
6 dimensions required to capture α=0.975
MSE along 3rd eigenvector: 603.4087192828594

data projections

$ python3 Assign2-part2.py energydata_complete.csv > part2.out
d = 10
min: 8.537736462515939e-07
max: 179.99999914622634
range: 0.000001 - 179.999999
mean: 89.92643242376826
variance: 380.2098573590691

d = 100
min: 65.16541251029841
max: 116.10388113733993

range: 65.165413 - 116.103881
mean: 90.01013691664623
variance: 33.52851745364576

d = 1000
min: 82.53040768358306
max: 98.04784624731151
range: 82.530408 - 98.047846
mean: 90.01281611377772
variance: 3.2755354760172364

number of bins: 180

occurrence out of 100000 samples

Angle between two random half-diagonals of dimension 10

number of bins: 180

Angle between two random half-diagonals of dimension 100

number of bins: 180

Angle between two random half-diagonals of dimension 1000