

# akulas\_HW3

January 23, 2019

## 1 HomeWork 3

Adam Kulas

watiam: akulas

student number: 20302000

References: <https://archive.ics.uci.edu/ml/machine-learning-databases/communities/>

```
In [1]: import pandas as pd
import numpy as np
from numpy import linalg as LA
from IPython.display import display, HTML
from sklearn.decomposition import PCA
from sklearn.pipeline import make_pipeline
import matplotlib.pyplot as plt

pd.options.display.max_rows = None
pd.options.display.max_columns = None

In [2]: data = pd.read_csv('data/communities.data', header=None, na_values='?')
with open('data/communities.names') as f:
    columns = [line.split(' ')[1] for line in f.readlines() if line.startswith('@attrib')]
data.columns = columns
print(data.shape)
display(data.head(15))
```

(1994, 128)

	state	county	community	communityname	fold	population	\
0	8	NaN	NaN	Lakewoodcity	1	0.19	
1	53	NaN	NaN	Tukwilacity	1	0.00	
2	24	NaN	NaN	Aberdeentown	1	0.00	
3	34	5.0	81440.0	Willingborotownship	1	0.04	
4	42	95.0	6096.0	Bethlehemtownship	1	0.01	
5	6	NaN	NaN	SouthPasadenacity	1	0.02	
6	44	7.0	41500.0	Lincolntown	1	0.01	
7	6	NaN	NaN	Selmacity	1	0.01	
8	21	NaN	NaN	Hendersoncity	1	0.03	

9	29	NaN	NaN	Claytoncity	1	0.01
10	6	NaN	NaN	DalyCitycity	1	0.13
11	36	NaN	NaN	RockvilleCentrevillage	1	0.02
12	25	21.0	44105.0	Needhamtown	1	0.03
13	55	87.0	30075.0	GrandChutetown	1	0.01
14	6	NaN	NaN	DanaPointcity	1	0.04

	householdsize	racepctblack	racePctWhite	racePctAsian	racePctHisp	\
0	0.33	0.02	0.90	0.12	0.17	
1	0.16	0.12	0.74	0.45	0.07	
2	0.42	0.49	0.56	0.17	0.04	
3	0.77	1.00	0.08	0.12	0.10	
4	0.55	0.02	0.95	0.09	0.05	
5	0.28	0.06	0.54	1.00	0.25	
6	0.39	0.00	0.98	0.06	0.02	
7	0.74	0.03	0.46	0.20	1.00	
8	0.34	0.20	0.84	0.02	0.00	
9	0.40	0.06	0.87	0.30	0.03	
10	0.71	0.15	0.07	1.00	0.41	
11	0.46	0.08	0.91	0.07	0.10	
12	0.47	0.01	0.96	0.13	0.02	
13	0.44	0.00	0.98	0.04	0.01	
14	0.36	0.01	0.85	0.14	0.26	

	agePct12t21	agePct12t29	agePct16t24	agePct65up	numbUrban	pctUrban	\
0	0.34	0.47	0.29	0.32	0.20	1.00	
1	0.26	0.59	0.35	0.27	0.02	1.00	
2	0.39	0.47	0.28	0.32	0.00	0.00	
3	0.51	0.50	0.34	0.21	0.06	1.00	
4	0.38	0.38	0.23	0.36	0.02	0.90	
5	0.31	0.48	0.27	0.37	0.04	1.00	
6	0.30	0.37	0.23	0.60	0.02	0.81	
7	0.52	0.55	0.36	0.35	0.00	0.00	
8	0.38	0.45	0.28	0.48	0.04	1.00	
9	0.90	0.82	0.80	0.39	0.02	1.00	
10	0.40	0.52	0.35	0.33	0.15	1.00	
11	0.34	0.36	0.22	0.57	0.04	1.00	
12	0.29	0.32	0.20	0.52	0.04	1.00	
13	0.35	0.53	0.32	0.23	0.02	0.77	
14	0.32	0.46	0.30	0.31	0.05	1.00	

	medIncome	pctWWage	pctWFarmSelf	pctWInvInc	pctWSocSec	pctWPubAsst	\
0	0.37	0.72	0.34	0.60	0.29	0.15	
1	0.31	0.72	0.11	0.45	0.25	0.29	
2	0.30	0.58	0.19	0.39	0.38	0.40	
3	0.58	0.89	0.21	0.43	0.36	0.20	
4	0.50	0.72	0.16	0.68	0.44	0.11	
5	0.52	0.68	0.20	0.61	0.28	0.15	

6	0.42	0.50	0.23	0.68	0.61	0.21
7	0.16	0.44	1.00	0.23	0.53	0.97
8	0.17	0.47	0.36	0.34	0.55	0.48
9	0.54	0.59	0.22	0.86	0.42	0.02
10	0.49	0.71	0.16	0.47	0.36	0.40
11	0.72	0.53	0.23	0.74	0.63	0.17
12	0.80	0.55	0.18	0.87	0.51	0.07
13	0.46	0.77	0.41	0.73	0.28	0.10
14	0.71	0.67	0.42	0.55	0.25	0.14

	pctWRetire	medFamInc	perCapInc	whitePerCap	blackPerCap	indianPerCap	\
0	0.43	0.39	0.40	0.39	0.32	0.27	
1	0.39	0.29	0.37	0.38	0.33	0.16	
2	0.84	0.28	0.27	0.29	0.27	0.07	
3	0.82	0.51	0.36	0.40	0.39	0.16	
4	0.71	0.46	0.43	0.41	0.28	0.00	
5	0.25	0.62	0.72	0.76	0.77	0.28	
6	0.54	0.43	0.47	0.44	0.40	0.24	
7	0.41	0.15	0.10	0.12	0.08	0.17	
8	0.43	0.21	0.23	0.23	0.19	0.10	
9	0.31	0.85	0.89	0.94	0.11	0.09	
10	0.45	0.47	0.33	0.37	0.40	0.22	
11	0.59	0.80	0.72	0.71	0.36	0.70	
12	0.55	0.82	0.78	0.77	0.62	0.14	
13	0.22	0.47	0.41	0.40	0.38	0.32	
14	0.36	0.70	0.79	0.82	0.29	0.37	

	AsianPerCap	OtherPerCap	HispPerCap	NumUnderPov	PctPopUnderPov	\
0	0.27	0.36	0.41	0.08	0.19	
1	0.30	0.22	0.35	0.01	0.24	
2	0.29	0.28	0.39	0.01	0.27	
3	0.25	0.36	0.44	0.01	0.10	
4	0.74	0.51	0.48	0.00	0.06	
5	0.52	0.48	0.60	0.01	0.12	
6	0.86	0.24	0.36	0.01	0.11	
7	0.27	0.18	0.21	0.03	0.64	
8	0.26	0.29	0.22	0.04	0.45	
9	0.33	0.17	0.80	0.00	0.11	
10	0.32	0.34	0.38	0.06	0.18	
11	0.43	0.35	0.41	0.01	0.07	
12	0.54	0.17	0.46	0.01	0.07	
13	0.09	0.13	0.75	0.00	0.07	
14	0.53	0.28	0.38	0.02	0.18	

	PctLess9thGrade	PctNotHSGrad	PctBSorMore	PctUnemployed	PctEmploy	\
0	0.10	0.18	0.48	0.27	0.68	
1	0.14	0.24	0.30	0.27	0.73	
2	0.27	0.43	0.19	0.36	0.58	

3	0.09	0.25	0.31	0.33	0.71
4	0.25	0.30	0.33	0.12	0.65
5	0.13	0.12	0.80	0.10	0.65
6	0.29	0.41	0.36	0.28	0.54
7	0.96	0.82	0.12	1.00	0.26
8	0.52	0.59	0.17	0.55	0.43
9	0.04	0.03	1.00	0.11	0.44
10	0.33	0.36	0.38	0.32	0.62
11	0.18	0.19	0.72	0.18	0.51
12	0.07	0.07	0.89	0.24	0.55
13	0.20	0.22	0.42	0.13	0.76
14	0.11	0.14	0.58	0.19	0.68

	PctEmplManu	PctEmplProfServ	PctOccupManu	PctOccupMgmtProf	\
0	0.23	0.41	0.25	0.52	
1	0.57	0.15	0.42	0.36	
2	0.32	0.29	0.49	0.32	
3	0.36	0.45	0.37	0.39	
4	0.67	0.38	0.42	0.46	
5	0.19	0.77	0.06	0.91	
6	0.44	0.53	0.33	0.49	
7	0.43	0.34	0.71	0.18	
8	0.59	0.36	0.64	0.29	
9	0.20	1.00	0.02	0.96	
10	0.20	0.29	0.35	0.30	
11	0.18	0.67	0.13	0.79	
12	0.29	0.73	0.10	0.88	
13	0.67	0.24	0.42	0.47	
14	0.28	0.35	0.15	0.61	

	MalePctDivorce	MalePctNevMarr	FemalePctDiv	TotalPctDiv	PersPerFam	\
0	0.68	0.40	0.75	0.75	0.35	
1	1.00	0.63	0.91	1.00	0.29	
2	0.63	0.41	0.71	0.70	0.45	
3	0.34	0.45	0.49	0.44	0.75	
4	0.22	0.27	0.20	0.21	0.51	
5	0.49	0.57	0.61	0.58	0.44	
6	0.25	0.34	0.28	0.28	0.42	
7	0.38	0.47	0.59	0.52	0.78	
8	0.62	0.26	0.66	0.67	0.37	
9	0.30	0.85	0.39	0.36	0.31	
10	0.41	0.55	0.46	0.46	0.85	
11	0.18	0.40	0.24	0.22	0.59	
12	0.13	0.38	0.20	0.18	0.50	
13	0.37	0.39	0.27	0.32	0.47	
14	0.60	0.50	0.68	0.66	0.35	

PctFam2Par PctKids2Par PctYoungKids2Par PctTeen2Par \

0	0.55	0.59	0.61	0.56
1	0.43	0.47	0.60	0.39
2	0.42	0.44	0.43	0.43
3	0.65	0.54	0.83	0.65
4	0.91	0.91	0.89	0.85
5	0.62	0.69	0.87	0.53
6	0.77	0.81	0.79	0.74
7	0.45	0.43	0.34	0.34
8	0.51	0.55	0.58	0.47
9	0.65	0.73	0.78	0.67
10	0.62	0.56	0.68	0.62
11	0.85	0.87	0.89	0.78
12	0.87	0.90	0.87	0.73
13	0.88	0.92	0.88	0.82
14	0.66	0.68	0.87	0.59

	PctWorkMomYoungKids	PctWorkMom	NumIlleg	PctIlleg	NumImmig	\
0	0.74	0.76	0.04	0.14	0.03	
1	0.46	0.53	0.00	0.24	0.01	
2	0.71	0.67	0.01	0.46	0.00	
3	0.85	0.86	0.03	0.33	0.02	
4	0.40	0.60	0.00	0.06	0.00	
5	0.30	0.43	0.00	0.11	0.04	
6	0.57	0.62	0.00	0.13	0.01	
7	0.29	0.27	0.02	0.50	0.02	
8	0.65	0.64	0.02	0.29	0.00	
9	0.72	0.71	0.00	0.07	0.01	
10	0.83	0.81	0.05	0.30	0.28	
11	0.31	0.39	0.01	0.14	0.01	
12	0.33	0.48	0.00	0.03	0.01	
13	0.56	0.60	0.00	0.08	0.00	
14	0.30	0.30	0.01	0.10	0.03	

	PctImmigRecent	PctImmigRec5	PctImmigRec8	PctImmigRec10	PctRecentImmig	\
0	0.24	0.27	0.37	0.39	0.07	
1	0.52	0.62	0.64	0.63	0.25	
2	0.07	0.06	0.15	0.19	0.02	
3	0.11	0.20	0.30	0.31	0.05	
4	0.03	0.07	0.20	0.27	0.01	
5	0.30	0.35	0.43	0.47	0.50	
6	0.00	0.02	0.02	0.10	0.00	
7	0.50	0.59	0.65	0.59	0.69	
8	0.12	0.09	0.07	0.13	0.00	
9	0.41	0.44	0.52	0.48	0.22	
10	0.36	0.43	0.52	0.57	1.00	
11	0.11	0.17	0.21	0.27	0.05	
12	0.25	0.28	0.26	0.25	0.13	
13	0.08	0.06	0.25	0.22	0.01	

14	0.69	0.74	0.70	0.69	0.65
----	------	------	------	------	------

	PctRecImmig5	PctRecImmig8	PctRecImmig10	PctSpeakEnglOnly	\
0	0.07	0.08	0.08	0.89	
1	0.27	0.25	0.23	0.84	
2	0.02	0.04	0.05	0.88	
3	0.08	0.11	0.11	0.81	
4	0.02	0.04	0.05	0.88	
5	0.50	0.56	0.57	0.45	
6	0.01	0.01	0.03	0.73	
7	0.72	0.71	0.60	0.12	
8	0.00	0.00	0.00	0.99	
9	0.21	0.22	0.19	0.85	
10	1.00	1.00	1.00	0.00	
11	0.07	0.08	0.10	0.82	
12	0.13	0.11	0.10	0.87	
13	0.00	0.02	0.01	0.96	
14	0.61	0.53	0.47	0.74	

	PctNotSpeakEnglWell	PctLargHouseFam	PctLargHouseOccup	PersPerOccupHous	\
0	0.06	0.14	0.13	0.33	
1	0.10	0.16	0.10	0.17	
2	0.04	0.20	0.20	0.46	
3	0.08	0.56	0.62	0.85	
4	0.05	0.16	0.19	0.59	
5	0.28	0.25	0.19	0.29	
6	0.05	0.12	0.13	0.42	
7	0.93	0.74	0.75	0.80	
8	0.01	0.12	0.12	0.35	
9	0.03	0.09	0.06	0.15	
10	0.58	0.88	0.82	0.77	
11	0.08	0.34	0.31	0.50	
12	0.05	0.18	0.19	0.50	
13	0.05	0.14	0.14	0.45	
14	0.32	0.26	0.22	0.39	

	PersPerOwnOccHous	PersPerRentOccHous	PctPersOwnOccup	PctPersDenseHous	\
0	0.39	0.28	0.55	0.09	
1	0.29	0.17	0.26	0.20	
2	0.52	0.43	0.42	0.15	
3	0.77	1.00	0.94	0.12	
4	0.60	0.37	0.89	0.02	
5	0.53	0.18	0.39	0.26	
6	0.54	0.24	0.65	0.03	
7	0.68	0.92	0.39	0.89	
8	0.38	0.33	0.50	0.10	
9	0.34	0.05	0.48	0.03	
10	0.79	0.72	0.49	0.91	

11	0.64	0.20	0.72	0.09
12	0.55	0.25	0.83	0.02
13	0.63	0.19	0.65	0.04
14	0.37	0.45	0.47	0.21

	PctHousLess3BR	MedNumBR	HousVacant	PctHousOccup	PctHousOwnOcc	\
0	0.51	0.5	0.21	0.71	0.52	
1	0.82	0.0	0.02	0.79	0.24	
2	0.51	0.5	0.01	0.86	0.41	
3	0.01	0.5	0.01	0.97	0.96	
4	0.19	0.5	0.01	0.89	0.87	
5	0.73	0.0	0.02	0.84	0.30	
6	0.46	0.5	0.01	0.89	0.57	
7	0.66	0.0	0.01	0.91	0.46	
8	0.64	0.0	0.04	0.72	0.49	
9	0.58	0.0	0.02	0.72	0.38	
10	0.63	0.0	0.06	0.87	0.49	
11	0.41	0.5	0.02	0.89	0.62	
12	0.25	0.5	0.01	0.94	0.78	
13	0.45	0.5	0.01	0.92	0.55	
14	0.54	0.5	0.10	0.43	0.50	

	PctVacantBoarded	PctVacMore6Mos	MedYrHousBuilt	PctHousNoPhone	\
0	0.05	0.26	0.65	0.14	
1	0.02	0.25	0.65	0.16	
2	0.29	0.30	0.52	0.47	
3	0.60	0.47	0.52	0.11	
4	0.04	0.55	0.73	0.05	
5	0.16	0.28	0.25	0.02	
6	0.09	0.49	0.38	0.05	
7	0.22	0.37	0.60	0.28	
8	0.05	0.49	0.50	0.57	
9	0.07	0.47	0.04	0.01	
10	0.10	0.20	0.46	0.05	
11	0.07	0.43	0.06	0.02	
12	0.00	0.26	0.29	0.01	
13	0.05	0.21	0.85	0.08	
14	0.00	0.19	0.79	0.07	

	PctWOFullPlumb	OwnOccLowQuart	OwnOccMedVal	OwnOccHiQuart	RentLowQ	\
0	0.06	0.22	0.19	0.18	0.36	
1	0.00	0.21	0.20	0.21	0.42	
2	0.45	0.18	0.17	0.16	0.27	
3	0.11	0.24	0.21	0.19	0.75	
4	0.14	0.31	0.31	0.30	0.40	
5	0.05	0.94	1.00	1.00	0.67	
6	0.05	0.37	0.38	0.39	0.26	
7	0.23	0.15	0.13	0.13	0.21	

8	0.22	0.07	0.07	0.08	0.14
9	0.00	0.63	0.71	0.79	0.44
10	0.28	0.75	0.74	0.70	0.73
11	0.09	0.74	0.76	0.79	0.45
12	0.15	0.71	0.69	0.68	0.60
13	0.07	0.19	0.18	0.18	0.35
14	0.07	0.88	0.93	0.97	0.96

	RentMedian	RentHighQ	MedRent	MedRentPctHousInc	MedOwnCostPctInc	\
0	0.35	0.38	0.34	0.38	0.46	
1	0.38	0.40	0.37	0.29	0.32	
2	0.29	0.27	0.31	0.48	0.39	
3	0.70	0.77	0.89	0.63	0.51	
4	0.36	0.38	0.38	0.22	0.51	
5	0.63	0.68	0.62	0.47	0.59	
6	0.35	0.42	0.35	0.46	0.44	
7	0.24	0.25	0.24	0.64	0.59	
8	0.17	0.16	0.15	0.38	0.13	
9	0.42	0.47	0.41	0.23	0.27	
10	0.67	0.82	0.67	0.58	0.67	
11	0.54	0.65	0.49	0.39	0.51	
12	0.82	1.00	0.87	0.40	0.39	
13	0.32	0.32	0.29	0.16	0.35	
14	0.91	1.00	0.87	0.55	0.81	

	MedOwnCostPctIncNoMtg	NumInShelters	NumStreet	PctForeignBorn	\
0	0.25	0.04	0.00	0.12	
1	0.18	0.00	0.00	0.21	
2	0.28	0.00	0.00	0.14	
3	0.47	0.00	0.00	0.19	
4	0.21	0.00	0.00	0.11	
5	0.11	0.00	0.00	0.70	
6	0.31	0.00	0.00	0.15	
7	0.28	0.00	0.00	0.59	
8	0.36	0.01	0.00	0.01	
9	0.28	0.00	0.00	0.22	
10	0.19	0.07	0.14	1.00	
11	0.68	0.01	0.00	0.20	
12	0.46	0.00	0.00	0.22	
13	0.26	0.01	0.00	0.03	
14	0.21	0.00	0.11	0.40	

	PctBornSameState	PctSameHouse85	PctSameCity85	PctSameState85	\
0	0.42	0.50	0.51	0.64	
1	0.50	0.34	0.60	0.52	
2	0.49	0.54	0.67	0.56	
3	0.30	0.73	0.64	0.65	
4	0.72	0.64	0.61	0.53	



5	0.42	0.49	0.73	0.64
6	0.81	0.77	0.91	0.84
7	0.58	0.52	0.79	0.78
8	0.78	0.48	0.79	0.75
9	0.42	0.34	0.23	0.09
10	0.35	0.53	0.39	0.62
11	0.88	0.91	0.90	0.93
12	0.66	0.82	0.66	0.73
13	0.86	0.43	0.47	0.69
14	0.42	0.31	0.51	0.53

	LemasSwornFT	LemasSwFTPerPop	LemasSwFTFieldOps	LemasSwFTFieldPerPop	\
0	0.03	0.13	0.96	0.17	
1	NaN	NaN	NaN	NaN	
2	NaN	NaN	NaN	NaN	
3	NaN	NaN	NaN	NaN	
4	NaN	NaN	NaN	NaN	
5	NaN	NaN	NaN	NaN	
6	NaN	NaN	NaN	NaN	
7	NaN	NaN	NaN	NaN	
8	NaN	NaN	NaN	NaN	
9	NaN	NaN	NaN	NaN	
10	NaN	NaN	NaN	NaN	
11	NaN	NaN	NaN	NaN	
12	NaN	NaN	NaN	NaN	
13	NaN	NaN	NaN	NaN	
14	NaN	NaN	NaN	NaN	

	LemasTotalReq	LemasTotReqPerPop	PolicReqPerOffic	PolicPerPop	\
0	0.06	0.18	0.44	0.13	
1	NaN	NaN	NaN	NaN	
2	NaN	NaN	NaN	NaN	
3	NaN	NaN	NaN	NaN	
4	NaN	NaN	NaN	NaN	
5	NaN	NaN	NaN	NaN	
6	NaN	NaN	NaN	NaN	
7	NaN	NaN	NaN	NaN	
8	NaN	NaN	NaN	NaN	
9	NaN	NaN	NaN	NaN	
10	NaN	NaN	NaN	NaN	
11	NaN	NaN	NaN	NaN	
12	NaN	NaN	NaN	NaN	
13	NaN	NaN	NaN	NaN	
14	NaN	NaN	NaN	NaN	

	RacialMatchCommPol	PctPolicWhite	PctPolicBlack	PctPolicHisp	\
0	0.94	0.93	0.03	0.07	
1	NaN	NaN	NaN	NaN	

2	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN
5	NaN	NaN	NaN	NaN
6	NaN	NaN	NaN	NaN
7	NaN	NaN	NaN	NaN
8	NaN	NaN	NaN	NaN
9	NaN	NaN	NaN	NaN
10	NaN	NaN	NaN	NaN
11	NaN	NaN	NaN	NaN
12	NaN	NaN	NaN	NaN
13	NaN	NaN	NaN	NaN
14	NaN	NaN	NaN	NaN

	PctPolicAsian	PctPolicMinor	OfficAssgnDrugUnits	NumKindsDrugsSeiz	\
0	0.1	0.07	0.02	0.57	
1	NaN	NaN	NaN	NaN	
2	NaN	NaN	NaN	NaN	
3	NaN	NaN	NaN	NaN	
4	NaN	NaN	NaN	NaN	
5	NaN	NaN	NaN	NaN	
6	NaN	NaN	NaN	NaN	
7	NaN	NaN	NaN	NaN	
8	NaN	NaN	NaN	NaN	
9	NaN	NaN	NaN	NaN	
10	NaN	NaN	NaN	NaN	
11	NaN	NaN	NaN	NaN	
12	NaN	NaN	NaN	NaN	
13	NaN	NaN	NaN	NaN	
14	NaN	NaN	NaN	NaN	

	PolicAveOTWorked	LandArea	PopDens	PctUsePubTrans	PolicCars	\
0	0.29	0.12	0.26	0.20	0.06	
1	NaN	0.02	0.12	0.45	NaN	
2	NaN	0.01	0.21	0.02	NaN	
3	NaN	0.02	0.39	0.28	NaN	
4	NaN	0.04	0.09	0.02	NaN	
5	NaN	0.01	0.58	0.10	NaN	
6	NaN	0.05	0.08	0.06	NaN	
7	NaN	0.01	0.33	0.00	NaN	
8	NaN	0.04	0.17	0.04	NaN	
9	NaN	0.00	0.47	0.11	NaN	
10	NaN	0.02	1.00	1.00	NaN	
11	NaN	0.01	0.63	1.00	NaN	
12	NaN	0.03	0.18	0.59	NaN	
13	NaN	0.08	0.04	0.00	NaN	
14	NaN	0.02	0.40	0.15	NaN	

	PolicOperBudg	LemasPctPolicOnPatr	LemasGangUnitDeploy	\
0	0.04	0.9	0.5	
1	NaN	NaN	NaN	
2	NaN	NaN	NaN	
3	NaN	NaN	NaN	
4	NaN	NaN	NaN	
5	NaN	NaN	NaN	
6	NaN	NaN	NaN	
7	NaN	NaN	NaN	
8	NaN	NaN	NaN	
9	NaN	NaN	NaN	
10	NaN	NaN	NaN	
11	NaN	NaN	NaN	
12	NaN	NaN	NaN	
13	NaN	NaN	NaN	
14	NaN	NaN	NaN	

	LemasPctOfficDrugUn	PolicBudgPerPop	ViolentCrimesPerPop
0	0.32	0.14	0.20
1	0.00	NaN	0.67
2	0.00	NaN	0.43
3	0.00	NaN	0.12
4	0.00	NaN	0.03
5	0.00	NaN	0.14
6	0.00	NaN	0.03
7	0.00	NaN	0.55
8	0.00	NaN	0.53
9	0.00	NaN	0.15
10	0.00	NaN	0.24
11	0.00	NaN	0.08
12	0.00	NaN	0.06
13	0.00	NaN	0.09
14	0.00	NaN	0.21

## 1.1 Clean Data

Some if the data is missing in the dataset. This missing data is filled with the mean value of the feature.

```
In [3]: data_filled = data
        data_filled.iloc[:,4:] = data.iloc[:,4:].apply(lambda x: x.fillna(x.mean()),axis=0)
        data_filled.columns = columns
        display(data_filled.head(30))
```

	state	county	community	communityname	fold	population	\
0	8	NaN	NaN	Lakewoodcity	1	0.19	
1	53	NaN	NaN	Tukwilacity	1	0.00	

2	24	NaN	NaN	Aberdeentown	1	0.00
3	34	5.0	81440.0	Willingborotownship	1	0.04
4	42	95.0	6096.0	Bethlehemtownship	1	0.01
5	6	NaN	NaN	SouthPasadenacity	1	0.02
6	44	7.0	41500.0	Lincolntown	1	0.01
7	6	NaN	NaN	Selmacity	1	0.01
8	21	NaN	NaN	Hendersoncity	1	0.03
9	29	NaN	NaN	Claytoncity	1	0.01
10	6	NaN	NaN	DalyCitycity	1	0.13
11	36	NaN	NaN	RockvilleCentrevillage	1	0.02
12	25	21.0	44105.0	Needhamtown	1	0.03
13	55	87.0	30075.0	GrandChutetown	1	0.01
14	6	NaN	NaN	DanaPointcity	1	0.04
15	19	187.0	91370.0	FortDodgecity	1	0.03
16	36	1.0	1000.0	Albanycity	1	0.15
17	34	27.0	17650.0	Denvilletownship	1	0.01
18	18	NaN	NaN	Valparaisocity	1	0.02
19	42	129.0	66376.0	Rostravertownship	1	0.00
20	6	NaN	NaN	Modestocity	1	0.25
21	12	31.0	NaN	Jacksonvillecity	1	1.00
22	41	NaN	NaN	KlamathFallscity	1	0.01
23	19	193.0	93926.0	SiouxCitycity	1	0.11
24	6	NaN	NaN	Delanocity	1	0.02
25	8	NaN	NaN	Goldencity	1	0.00
26	6	NaN	NaN	Gardenacity	1	0.06
27	39	29.0	61798.0	Perrytownship	1	0.01
28	54	NaN	NaN	Beckleycity	1	0.01
29	9	7.0	22490.0	EastHamptontown	1	0.00

	householdsize	racepctblack	racePctWhite	racePctAsian	racePctHispanic	\
0	0.33	0.02	0.90	0.12	0.17	
1	0.16	0.12	0.74	0.45	0.07	
2	0.42	0.49	0.56	0.17	0.04	
3	0.77	1.00	0.08	0.12	0.10	
4	0.55	0.02	0.95	0.09	0.05	
5	0.28	0.06	0.54	1.00	0.25	
6	0.39	0.00	0.98	0.06	0.02	
7	0.74	0.03	0.46	0.20	1.00	
8	0.34	0.20	0.84	0.02	0.00	
9	0.40	0.06	0.87	0.30	0.03	
10	0.71	0.15	0.07	1.00	0.41	
11	0.46	0.08	0.91	0.07	0.10	
12	0.47	0.01	0.96	0.13	0.02	
13	0.44	0.00	0.98	0.04	0.01	
14	0.36	0.01	0.85	0.14	0.26	
15	0.34	0.06	0.93	0.03	0.03	
16	0.31	0.40	0.63	0.14	0.06	
17	0.53	0.01	0.94	0.20	0.03	

18	0.47	0.01	0.97	0.07	0.02
19	0.41	0.05	0.96	0.01	0.01
20	0.54	0.05	0.71	0.48	0.30
21	0.42	0.47	0.59	0.12	0.05
22	0.34	0.02	0.87	0.07	0.11
23	0.43	0.04	0.89	0.09	0.06
24	0.96	0.05	0.00	1.00	1.00
25	0.33	0.02	0.91	0.16	0.09
26	0.49	0.46	0.00	1.00	0.43
27	0.37	0.01	0.99	0.02	0.01
28	0.27	0.43	0.64	0.08	0.01
29	0.50	0.01	0.98	0.02	0.02

	agePct12t21	agePct12t29	agePct16t24	agePct65up	numbUrban	pctUrban	\
0	0.34	0.47	0.29	0.32	0.20	1.00	
1	0.26	0.59	0.35	0.27	0.02	1.00	
2	0.39	0.47	0.28	0.32	0.00	0.00	
3	0.51	0.50	0.34	0.21	0.06	1.00	
4	0.38	0.38	0.23	0.36	0.02	0.90	
5	0.31	0.48	0.27	0.37	0.04	1.00	
6	0.30	0.37	0.23	0.60	0.02	0.81	
7	0.52	0.55	0.36	0.35	0.00	0.00	
8	0.38	0.45	0.28	0.48	0.04	1.00	
9	0.90	0.82	0.80	0.39	0.02	1.00	
10	0.40	0.52	0.35	0.33	0.15	1.00	
11	0.34	0.36	0.22	0.57	0.04	1.00	
12	0.29	0.32	0.20	0.52	0.04	1.00	
13	0.35	0.53	0.32	0.23	0.02	0.77	
14	0.32	0.46	0.30	0.31	0.05	1.00	
15	0.39	0.41	0.28	0.58	0.00	0.00	
16	0.58	0.72	0.65	0.47	0.16	1.00	
17	0.34	0.39	0.27	0.36	0.02	0.76	
18	0.70	0.67	0.63	0.37	0.00	0.00	
19	0.37	0.37	0.24	0.55	0.01	0.58	
20	0.42	0.48	0.28	0.32	0.26	1.00	
21	0.41	0.53	0.34	0.33	1.00	0.99	
22	0.49	0.56	0.43	0.47	0.00	0.00	
23	0.45	0.48	0.31	0.46	0.13	1.00	
24	0.54	0.58	0.39	0.33	0.00	0.00	
25	0.55	0.63	0.53	0.31	0.02	1.00	
26	0.35	0.50	0.32	0.34	0.08	1.00	
27	0.35	0.38	0.23	0.59	0.00	0.00	
28	0.36	0.33	0.22	0.74	0.00	0.00	
29	0.31	0.40	0.23	0.25	0.00	0.00	

	medIncome	pctWWage	pctWFarmSelf	pctWInvInc	pctWSocSec	pctWPubAsst	\
0	0.37	0.72	0.34	0.60	0.29	0.15	
1	0.31	0.72	0.11	0.45	0.25	0.29	

2	0.30	0.58	0.19	0.39	0.38	0.40
3	0.58	0.89	0.21	0.43	0.36	0.20
4	0.50	0.72	0.16	0.68	0.44	0.11
5	0.52	0.68	0.20	0.61	0.28	0.15
6	0.42	0.50	0.23	0.68	0.61	0.21
7	0.16	0.44	1.00	0.23	0.53	0.97
8	0.17	0.47	0.36	0.34	0.55	0.48
9	0.54	0.59	0.22	0.86	0.42	0.02
10	0.49	0.71	0.16	0.47	0.36	0.40
11	0.72	0.53	0.23	0.74	0.63	0.17
12	0.80	0.55	0.18	0.87	0.51	0.07
13	0.46	0.77	0.41	0.73	0.28	0.10
14	0.71	0.67	0.42	0.55	0.25	0.14
15	0.18	0.42	0.81	0.49	0.62	0.37
16	0.22	0.52	0.10	0.51	0.48	0.39
17	0.79	0.77	0.13	0.77	0.44	0.15
18	0.33	0.56	0.28	0.62	0.43	0.21
19	0.23	0.34	0.33	0.51	0.70	0.36
20	0.33	0.55	0.37	0.37	0.39	0.64
21	0.28	0.62	0.16	0.36	0.40	0.30
22	0.13	0.40	0.26	0.42	0.52	0.41
23	0.22	0.52	0.44	0.49	0.56	0.41
24	0.16	0.61	0.41	0.14	0.49	0.92
25	0.29	0.64	0.60	0.57	0.30	0.15
26	0.35	0.68	0.16	0.36	0.36	0.38
27	0.23	0.38	0.39	0.51	0.64	0.35
28	0.12	0.10	0.11	0.39	0.84	0.48
29	0.60	0.83	0.25	0.63	0.26	0.12

	pctWRetire	medFamInc	perCapInc	whitePerCap	blackPerCap	indianPerCap	\
0	0.43	0.39	0.40	0.39	0.32	0.27	
1	0.39	0.29	0.37	0.38	0.33	0.16	
2	0.84	0.28	0.27	0.29	0.27	0.07	
3	0.82	0.51	0.36	0.40	0.39	0.16	
4	0.71	0.46	0.43	0.41	0.28	0.00	
5	0.25	0.62	0.72	0.76	0.77	0.28	
6	0.54	0.43	0.47	0.44	0.40	0.24	
7	0.41	0.15	0.10	0.12	0.08	0.17	
8	0.43	0.21	0.23	0.23	0.19	0.10	
9	0.31	0.85	0.89	0.94	0.11	0.09	
10	0.45	0.47	0.33	0.37	0.40	0.22	
11	0.59	0.80	0.72	0.71	0.36	0.70	
12	0.55	0.82	0.78	0.77	0.62	0.14	
13	0.22	0.47	0.41	0.40	0.38	0.32	
14	0.36	0.70	0.79	0.82	0.29	0.37	
15	0.41	0.21	0.22	0.21	0.16	0.12	
16	0.51	0.30	0.29	0.34	0.23	0.13	
17	0.56	0.75	0.69	0.67	0.38	0.00	

18	0.40	0.41	0.34	0.32	0.22	0.37
19	0.87	0.24	0.22	0.21	0.24	0.15
20	0.44	0.32	0.29	0.32	0.23	0.17
21	0.45	0.29	0.30	0.35	0.21	0.20
22	0.41	0.17	0.20	0.20	0.13	0.08
23	0.39	0.25	0.25	0.25	0.16	0.07
24	0.28	0.12	0.08	0.20	0.14	0.11
25	0.41	0.32	0.34	0.34	0.27	0.15
26	0.34	0.34	0.32	0.32	0.38	0.17
27	0.63	0.24	0.23	0.22	0.23	0.25
28	0.82	0.21	0.25	0.27	0.16	0.28
29	0.31	0.58	0.48	0.46	0.41	0.08

	AsianPerCap	OtherPerCap	HispPerCap	NumUnderPov	PctPopUnderPov	\
0	0.27	0.36	0.41	0.08	0.19	
1	0.30	0.22	0.35	0.01	0.24	
2	0.29	0.28	0.39	0.01	0.27	
3	0.25	0.36	0.44	0.01	0.10	
4	0.74	0.51	0.48	0.00	0.06	
5	0.52	0.48	0.60	0.01	0.12	
6	0.86	0.24	0.36	0.01	0.11	
7	0.27	0.18	0.21	0.03	0.64	
8	0.26	0.29	0.22	0.04	0.45	
9	0.33	0.17	0.80	0.00	0.11	
10	0.32	0.34	0.38	0.06	0.18	
11	0.43	0.35	0.41	0.01	0.07	
12	0.54	0.17	0.46	0.01	0.07	
13	0.09	0.13	0.75	0.00	0.07	
14	0.53	0.28	0.38	0.02	0.18	
15	0.46	0.33	0.28	0.03	0.33	
16	0.20	0.20	0.30	0.15	0.48	
17	0.51	1.00	0.62	0.00	0.04	
18	0.69	0.72	0.35	0.02	0.22	
19	0.00	0.00	0.31	0.01	0.28	
20	0.17	0.24	0.30	0.19	0.34	
21	0.33	0.28	0.40	0.76	0.33	
22	0.18	0.17	0.19	0.03	0.58	
23	0.20	0.19	0.22	0.10	0.36	
24	0.17	0.18	0.21	0.05	0.63	
25	0.15	0.30	0.35	0.01	0.30	
26	0.37	0.28	0.33	0.04	0.25	
27	0.48	0.20	0.51	0.02	0.32	
28	1.00	0.00	0.86	0.03	0.57	
29	0.39	0.38	0.41	0.00	0.01	

	PctLess9thGrade	PctNotHSGrad	PctBSorMore	PctUnemployed	PctEmploy	\
0	0.10	0.18	0.48	0.27	0.68	
1	0.14	0.24	0.30	0.27	0.73	

2	0.27	0.43	0.19	0.36	0.58
3	0.09	0.25	0.31	0.33	0.71
4	0.25	0.30	0.33	0.12	0.65
5	0.13	0.12	0.80	0.10	0.65
6	0.29	0.41	0.36	0.28	0.54
7	0.96	0.82	0.12	1.00	0.26
8	0.52	0.59	0.17	0.55	0.43
9	0.04	0.03	1.00	0.11	0.44
10	0.33	0.36	0.38	0.32	0.62
11	0.18	0.19	0.72	0.18	0.51
12	0.07	0.07	0.89	0.24	0.55
13	0.20	0.22	0.42	0.13	0.76
14	0.11	0.14	0.58	0.19	0.68
15	0.30	0.37	0.24	0.32	0.42
16	0.30	0.38	0.47	0.37	0.44
17	0.10	0.18	0.58	0.11	0.68
18	0.16	0.24	0.45	0.19	0.48
19	0.27	0.38	0.19	0.72	0.25
20	0.34	0.43	0.24	0.59	0.43
21	0.25	0.39	0.28	0.30	0.59
22	0.31	0.42	0.24	0.63	0.32
23	0.29	0.37	0.27	0.33	0.53
24	1.00	1.00	0.10	1.00	0.22
25	0.11	0.18	0.54	0.30	0.56
26	0.35	0.46	0.25	0.35	0.57
27	0.30	0.45	0.16	0.34	0.37
28	0.46	0.49	0.30	0.54	0.13
29	0.18	0.27	0.44	0.22	0.81

	PctEmplManu	PctEmplProfServ	PctOccupManu	PctOccupMgmtProf	\
0	0.23	0.41	0.25	0.52	
1	0.57	0.15	0.42	0.36	
2	0.32	0.29	0.49	0.32	
3	0.36	0.45	0.37	0.39	
4	0.67	0.38	0.42	0.46	
5	0.19	0.77	0.06	0.91	
6	0.44	0.53	0.33	0.49	
7	0.43	0.34	0.71	0.18	
8	0.59	0.36	0.64	0.29	
9	0.20	1.00	0.02	0.96	
10	0.20	0.29	0.35	0.30	
11	0.18	0.67	0.13	0.79	
12	0.29	0.73	0.10	0.88	
13	0.67	0.24	0.42	0.47	
14	0.28	0.35	0.15	0.61	
15	0.38	0.48	0.48	0.36	
16	0.08	0.73	0.21	0.53	
17	0.42	0.37	0.21	0.66	



18	0.36	0.77	0.25	0.56
19	0.39	0.35	0.49	0.34
20	0.38	0.41	0.43	0.36
21	0.18	0.33	0.35	0.38
22	0.45	0.44	0.56	0.34
23	0.43	0.47	0.44	0.33
24	0.09	0.26	0.32	0.16
25	0.33	0.53	0.31	0.58
26	0.62	0.23	0.42	0.32
27	0.78	0.35	0.71	0.25
28	0.05	0.77	0.18	0.57
29	0.52	0.47	0.27	0.57

	MalePctDivorce	MalePctNevMarr	FemalePctDiv	TotalPctDiv	PersPerFam	\
0	0.68	0.40	0.75	0.75	0.35	
1	1.00	0.63	0.91	1.00	0.29	
2	0.63	0.41	0.71	0.70	0.45	
3	0.34	0.45	0.49	0.44	0.75	
4	0.22	0.27	0.20	0.21	0.51	
5	0.49	0.57	0.61	0.58	0.44	
6	0.25	0.34	0.28	0.28	0.42	
7	0.38	0.47	0.59	0.52	0.78	
8	0.62	0.26	0.66	0.67	0.37	
9	0.30	0.85	0.39	0.36	0.31	
10	0.41	0.55	0.46	0.46	0.85	
11	0.18	0.40	0.24	0.22	0.59	
12	0.13	0.38	0.20	0.18	0.50	
13	0.37	0.39	0.27	0.32	0.47	
14	0.60	0.50	0.68	0.66	0.35	
15	0.50	0.34	0.45	0.49	0.38	
16	0.50	0.94	0.45	0.49	0.36	
17	0.25	0.40	0.25	0.26	0.52	
18	0.38	0.58	0.43	0.43	0.41	
19	0.31	0.29	0.27	0.30	0.41	
20	0.60	0.35	0.68	0.67	0.60	
21	0.69	0.41	0.71	0.73	0.47	
22	0.72	0.49	0.69	0.72	0.40	
23	0.48	0.39	0.48	0.50	0.49	
24	0.28	0.49	0.36	0.33	1.00	
25	0.60	0.73	0.66	0.64	0.28	
26	0.55	0.59	0.65	0.62	0.58	
27	0.51	0.25	0.47	0.50	0.40	
28	0.61	0.33	0.58	0.63	0.35	
29	0.41	0.35	0.38	0.40	0.47	

	PctFam2Par	PctKids2Par	PctYoungKids2Par	PctTeen2Par	\
0	0.55	0.59	0.61	0.56	
1	0.43	0.47	0.60	0.39	

2	0.42	0.44	0.43	0.43
3	0.65	0.54	0.83	0.65
4	0.91	0.91	0.89	0.85
5	0.62	0.69	0.87	0.53
6	0.77	0.81	0.79	0.74
7	0.45	0.43	0.34	0.34
8	0.51	0.55	0.58	0.47
9	0.65	0.73	0.78	0.67
10	0.62	0.56	0.68	0.62
11	0.85	0.87	0.89	0.78
12	0.87	0.90	0.87	0.73
13	0.88	0.92	0.88	0.82
14	0.66	0.68	0.87	0.59
15	0.55	0.60	0.61	0.52
16	0.20	0.25	0.25	0.19
17	0.88	0.88	0.97	0.74
18	0.66	0.73	0.72	0.63
19	0.76	0.74	0.83	0.89
20	0.60	0.61	0.64	0.61
21	0.53	0.51	0.59	0.51
22	0.49	0.53	0.47	0.68
23	0.59	0.61	0.51	0.66
24	0.54	0.46	0.42	0.60
25	0.58	0.60	0.75	0.59
26	0.47	0.47	0.61	0.46
27	0.65	0.68	0.76	0.59
28	0.41	0.44	0.42	0.55
29	0.83	0.83	0.79	0.81

	PctWorkMomYoungKids	PctWorkMom	NumIlleg	PctIlleg	NumImmig	\
0	0.74	0.76	0.04	0.14	0.03	
1	0.46	0.53	0.00	0.24	0.01	
2	0.71	0.67	0.01	0.46	0.00	
3	0.85	0.86	0.03	0.33	0.02	
4	0.40	0.60	0.00	0.06	0.00	
5	0.30	0.43	0.00	0.11	0.04	
6	0.57	0.62	0.00	0.13	0.01	
7	0.29	0.27	0.02	0.50	0.02	
8	0.65	0.64	0.02	0.29	0.00	
9	0.72	0.71	0.00	0.07	0.01	
10	0.83	0.81	0.05	0.30	0.28	
11	0.31	0.39	0.01	0.14	0.01	
12	0.33	0.48	0.00	0.03	0.01	
13	0.56	0.60	0.00	0.08	0.00	
14	0.30	0.30	0.01	0.10	0.03	
15	0.64	0.71	0.01	0.19	0.00	
16	0.61	0.65	0.13	0.73	0.05	
17	0.51	0.48	0.00	0.04	0.01	

18	0.53	0.60	0.01	0.11	0.00
19	0.45	0.31	0.00	0.06	0.00
20	0.30	0.30	0.08	0.22	0.14
21	0.55	0.56	0.49	0.35	0.16
22	0.37	0.48	0.01	0.20	0.00
23	0.68	0.71	0.05	0.26	0.02
24	0.47	0.47	0.03	0.43	0.06
25	0.65	0.67	0.00	0.07	0.00
26	0.56	0.49	0.05	0.49	0.10
27	0.27	0.42	0.00	0.06	0.00
28	0.23	0.40	0.01	0.29	0.00
29	0.70	0.82	0.00	0.04	0.00

	PctImmigRecent	PctImmigRec5	PctImmigRec8	PctImmigRec10	PctRecentImmig	\
0	0.24	0.27	0.37	0.39	0.07	
1	0.52	0.62	0.64	0.63	0.25	
2	0.07	0.06	0.15	0.19	0.02	
3	0.11	0.20	0.30	0.31	0.05	
4	0.03	0.07	0.20	0.27	0.01	
5	0.30	0.35	0.43	0.47	0.50	
6	0.00	0.02	0.02	0.10	0.00	
7	0.50	0.59	0.65	0.59	0.69	
8	0.12	0.09	0.07	0.13	0.00	
9	0.41	0.44	0.52	0.48	0.22	
10	0.36	0.43	0.52	0.57	1.00	
11	0.11	0.17	0.21	0.27	0.05	
12	0.25	0.28	0.26	0.25	0.13	
13	0.08	0.06	0.25	0.22	0.01	
14	0.69	0.74	0.70	0.69	0.65	
15	0.05	0.04	0.03	0.13	0.00	
16	0.39	0.39	0.46	0.46	0.21	
17	0.15	0.26	0.32	0.30	0.09	
18	0.40	0.35	0.37	0.39	0.08	
19	0.00	0.00	0.00	0.00	0.00	
20	0.31	0.45	0.57	0.64	0.28	
21	0.27	0.34	0.40	0.43	0.07	
22	0.48	0.55	0.59	0.56	0.14	
23	0.49	0.59	0.62	0.64	0.11	
24	0.36	0.42	0.53	0.60	1.00	
25	0.79	0.77	0.69	0.61	0.29	
26	0.39	0.47	0.59	0.65	0.80	
27	0.18	0.13	0.11	0.09	0.02	
28	0.10	0.10	0.13	0.14	0.01	
29	0.22	0.16	0.19	0.16	0.04	

	PctRecImmig5	PctRecImmig8	PctRecImmig10	PctSpeakEnglOnly	\
0	0.07	0.08	0.08	0.89	
1	0.27	0.25	0.23	0.84	

2	0.02	0.04	0.05	0.88
3	0.08	0.11	0.11	0.81
4	0.02	0.04	0.05	0.88
5	0.50	0.56	0.57	0.45
6	0.01	0.01	0.03	0.73
7	0.72	0.71	0.60	0.12
8	0.00	0.00	0.00	0.99
9	0.21	0.22	0.19	0.85
10	1.00	1.00	1.00	0.00
11	0.07	0.08	0.10	0.82
12	0.13	0.11	0.10	0.87
13	0.00	0.02	0.01	0.96
14	0.61	0.53	0.47	0.74
15	0.00	0.00	0.01	0.95
16	0.19	0.20	0.18	0.81
17	0.14	0.15	0.13	0.82
18	0.06	0.06	0.06	0.95
19	0.00	0.00	0.00	0.94
20	0.36	0.41	0.42	0.64
21	0.07	0.08	0.08	0.90
22	0.14	0.13	0.12	0.90
23	0.11	0.11	0.10	0.92
24	1.00	1.00	1.00	0.00
25	0.25	0.20	0.17	0.88
26	0.84	0.96	0.97	0.26
27	0.01	0.01	0.01	0.95
28	0.01	0.01	0.01	0.94
29	0.02	0.03	0.02	0.94

	PctNotSpeakEnglWell	PctLargHouseFam	PctLargHouseOccup	PersPerOccupHous	\
0	0.06	0.14	0.13	0.33	
1	0.10	0.16	0.10	0.17	
2	0.04	0.20	0.20	0.46	
3	0.08	0.56	0.62	0.85	
4	0.05	0.16	0.19	0.59	
5	0.28	0.25	0.19	0.29	
6	0.05	0.12	0.13	0.42	
7	0.93	0.74	0.75	0.80	
8	0.01	0.12	0.12	0.35	
9	0.03	0.09	0.06	0.15	
10	0.58	0.88	0.82	0.77	
11	0.08	0.34	0.31	0.50	
12	0.05	0.18	0.19	0.50	
13	0.05	0.14	0.14	0.45	
14	0.32	0.26	0.22	0.39	
15	0.03	0.14	0.12	0.32	
16	0.08	0.22	0.13	0.21	
17	0.04	0.20	0.21	0.57	

18	0.01	0.12	0.11	0.33
19	0.02	0.10	0.12	0.46
20	0.35	0.43	0.39	0.56
21	0.04	0.23	0.20	0.42
22	0.10	0.23	0.18	0.32
23	0.09	0.27	0.24	0.43
24	1.00	1.00	1.00	1.00
25	0.03	0.08	0.08	0.24
26	0.67	0.46	0.40	0.51
27	0.01	0.13	0.13	0.38
28	0.03	0.16	0.13	0.26
29	0.02	0.14	0.15	0.51

	PersPerOwnOccHous	PersPerRentOccHous	PctPersOwnOccup	PctPersDenseHous	\
0	0.39	0.28	0.55	0.09	
1	0.29	0.17	0.26	0.20	
2	0.52	0.43	0.42	0.15	
3	0.77	1.00	0.94	0.12	
4	0.60	0.37	0.89	0.02	
5	0.53	0.18	0.39	0.26	
6	0.54	0.24	0.65	0.03	
7	0.68	0.92	0.39	0.89	
8	0.38	0.33	0.50	0.10	
9	0.34	0.05	0.48	0.03	
10	0.79	0.72	0.49	0.91	
11	0.64	0.20	0.72	0.09	
12	0.55	0.25	0.83	0.02	
13	0.63	0.19	0.65	0.04	
14	0.37	0.45	0.47	0.21	
15	0.39	0.22	0.62	0.04	
16	0.33	0.22	0.24	0.10	
17	0.58	0.34	0.89	0.04	
18	0.49	0.19	0.54	0.06	
19	0.45	0.41	0.83	0.04	
20	0.54	0.60	0.47	0.39	
21	0.46	0.40	0.55	0.20	
22	0.37	0.32	0.38	0.18	
23	0.48	0.35	0.64	0.12	
24	1.00	1.00	0.39	1.00	
25	0.30	0.23	0.49	0.09	
26	0.48	0.56	0.30	0.81	
27	0.41	0.32	0.66	0.03	
28	0.26	0.30	0.53	0.07	
29	0.55	0.29	0.81	0.04	

	PctHousLess3BR	MedNumBR	HousVacant	PctHousOccup	PctHousOwnOcc	\
0	0.51	0.5	0.21	0.71	0.52	
1	0.82	0.0	0.02	0.79	0.24	

2	0.51	0.5	0.01	0.86	0.41
3	0.01	0.5	0.01	0.97	0.96
4	0.19	0.5	0.01	0.89	0.87
5	0.73	0.0	0.02	0.84	0.30
6	0.46	0.5	0.01	0.89	0.57
7	0.66	0.0	0.01	0.91	0.46
8	0.64	0.0	0.04	0.72	0.49
9	0.58	0.0	0.02	0.72	0.38
10	0.63	0.0	0.06	0.87	0.49
11	0.41	0.5	0.02	0.89	0.62
12	0.25	0.5	0.01	0.94	0.78
13	0.45	0.5	0.01	0.92	0.55
14	0.54	0.5	0.10	0.43	0.50
15	0.53	0.5	0.04	0.76	0.58
16	0.60	0.0	0.22	0.64	0.22
17	0.27	0.5	0.01	0.88	0.86
18	0.55	0.5	0.01	0.89	0.46
19	0.42	0.5	0.01	0.75	0.83
20	0.48	0.5	0.15	0.83	0.49
21	0.49	0.5	1.00	0.60	0.54
22	0.71	0.0	0.03	0.68	0.38
23	0.56	0.0	0.09	0.80	0.61
24	0.63	0.0	0.01	0.87	0.42
25	0.61	0.0	0.02	0.70	0.47
26	0.79	0.0	0.05	0.83	0.33
27	0.40	0.5	0.02	0.82	0.64
28	0.53	0.5	0.06	0.50	0.54
29	0.38	0.5	0.02	0.57	0.77

	PctVacantBoarded	PctVacMore6Mos	MedYrHousBuilt	PctHousNoPhone	\
0	0.05	0.26	0.65	0.14	
1	0.02	0.25	0.65	0.16	
2	0.29	0.30	0.52	0.47	
3	0.60	0.47	0.52	0.11	
4	0.04	0.55	0.73	0.05	
5	0.16	0.28	0.25	0.02	
6	0.09	0.49	0.38	0.05	
7	0.22	0.37	0.60	0.28	
8	0.05	0.49	0.50	0.57	
9	0.07	0.47	0.04	0.01	
10	0.10	0.20	0.46	0.05	
11	0.07	0.43	0.06	0.02	
12	0.00	0.26	0.29	0.01	
13	0.05	0.21	0.85	0.08	
14	0.00	0.19	0.79	0.07	
15	0.41	0.47	0.25	0.24	
16	0.79	0.60	0.00	0.40	
17	0.25	0.54	0.33	0.03	

18	0.27	0.43	0.50	0.12
19	0.31	0.84	0.44	0.05
20	0.09	0.16	0.73	0.14
21	0.52	0.44	0.63	0.37
22	0.33	0.40	0.23	0.54
23	0.45	0.43	0.19	0.34
24	0.15	0.29	0.54	0.52
25	0.10	0.34	0.65	0.19
26	0.08	0.12	0.42	0.15
27	0.30	0.61	0.27	0.33
28	0.30	0.85	0.29	0.48
29	0.04	0.74	0.48	0.07

	PctWOFullPlumb	OwnOccLowQuart	OwnOccMedVal	OwnOccHiQuart	RentLowQ	\
0	0.06	0.22	0.19	0.18	0.36	
1	0.00	0.21	0.20	0.21	0.42	
2	0.45	0.18	0.17	0.16	0.27	
3	0.11	0.24	0.21	0.19	0.75	
4	0.14	0.31	0.31	0.30	0.40	
5	0.05	0.94	1.00	1.00	0.67	
6	0.05	0.37	0.38	0.39	0.26	
7	0.23	0.15	0.13	0.13	0.21	
8	0.22	0.07	0.07	0.08	0.14	
9	0.00	0.63	0.71	0.79	0.44	
10	0.28	0.75	0.74	0.70	0.73	
11	0.09	0.74	0.76	0.79	0.45	
12	0.15	0.71	0.69	0.68	0.60	
13	0.07	0.19	0.18	0.18	0.35	
14	0.07	0.88	0.93	0.97	0.96	
15	0.44	0.04	0.04	0.05	0.10	
16	0.26	0.23	0.23	0.23	0.31	
17	0.08	0.50	0.49	0.50	0.80	
18	0.05	0.14	0.14	0.14	0.27	
19	0.18	0.08	0.09	0.10	0.06	
20	0.28	0.31	0.31	0.31	0.40	
21	0.23	0.11	0.11	0.13	0.25	
22	0.29	0.06	0.06	0.08	0.14	
23	0.25	0.05	0.04	0.05	0.14	
24	1.00	0.14	0.12	0.10	0.18	
25	0.18	0.22	0.20	0.21	0.32	
26	0.34	0.54	0.52	0.50	0.60	
27	0.32	0.07	0.06	0.07	0.13	
28	0.22	0.07	0.07	0.09	0.10	
29	0.23	0.43	0.42	0.39	0.43	

	RentMedian	RentHighQ	MedRent	MedRentPctHousInc	MedOwnCostPctInc	\
0	0.35	0.38	0.34	0.38	0.46	
1	0.38	0.40	0.37	0.29	0.32	

2	0.29	0.27	0.31	0.48	0.39
3	0.70	0.77	0.89	0.63	0.51
4	0.36	0.38	0.38	0.22	0.51
5	0.63	0.68	0.62	0.47	0.59
6	0.35	0.42	0.35	0.46	0.44
7	0.24	0.25	0.24	0.64	0.59
8	0.17	0.16	0.15	0.38	0.13
9	0.42	0.47	0.41	0.23	0.27
10	0.67	0.82	0.67	0.58	0.67
11	0.54	0.65	0.49	0.39	0.51
12	0.82	1.00	0.87	0.40	0.39
13	0.32	0.32	0.29	0.16	0.35
14	0.91	1.00	0.87	0.55	0.81
15	0.13	0.14	0.13	0.40	0.20
16	0.32	0.36	0.33	0.52	0.35
17	0.76	0.96	0.80	0.69	0.52
18	0.29	0.30	0.30	0.45	0.22
19	0.12	0.15	0.20	0.19	0.20
20	0.40	0.45	0.40	0.65	0.59
21	0.28	0.32	0.30	0.45	0.41
22	0.17	0.18	0.16	0.54	0.38
23	0.17	0.18	0.17	0.44	0.20
24	0.22	0.23	0.22	0.50	0.55
25	0.31	0.33	0.28	0.55	0.59
26	0.57	0.63	0.56	0.62	0.48
27	0.14	0.13	0.17	0.42	0.27
28	0.14	0.16	0.15	0.62	0.36
29	0.49	0.60	0.54	0.28	0.63

	MedOwnCostPctIncNoMtg	NumInShelters	NumStreet	PctForeignBorn	\
0	0.25	0.04	0.00	0.12	
1	0.18	0.00	0.00	0.21	
2	0.28	0.00	0.00	0.14	
3	0.47	0.00	0.00	0.19	
4	0.21	0.00	0.00	0.11	
5	0.11	0.00	0.00	0.70	
6	0.31	0.00	0.00	0.15	
7	0.28	0.00	0.00	0.59	
8	0.36	0.01	0.00	0.01	
9	0.28	0.00	0.00	0.22	
10	0.19	0.07	0.14	1.00	
11	0.68	0.01	0.00	0.20	
12	0.46	0.00	0.00	0.22	
13	0.26	0.01	0.00	0.03	
14	0.21	0.00	0.11	0.40	
15	0.35	0.01	0.00	0.03	
16	0.40	0.23	0.02	0.23	
17	0.65	0.00	0.00	0.26	



18	0.28	0.00	0.00	0.08
19	0.28	0.00	0.00	0.02
20	0.21	0.12	0.09	0.38
21	0.32	0.39	0.36	0.10
22	0.42	0.05	0.00	0.12
23	0.43	0.05	0.01	0.09
24	0.21	0.00	0.00	1.00
25	0.39	0.00	0.00	0.15
26	0.11	0.00	0.00	0.87
27	0.26	0.00	0.00	0.04
28	0.26	0.01	0.00	0.05
29	0.35	0.00	0.00	0.07

	PctBornSameState	PctSameHouse85	PctSameCity85	PctSameState85	\
0	0.42	0.50	0.51	0.64	
1	0.50	0.34	0.60	0.52	
2	0.49	0.54	0.67	0.56	
3	0.30	0.73	0.64	0.65	
4	0.72	0.64	0.61	0.53	
5	0.42	0.49	0.73	0.64	
6	0.81	0.77	0.91	0.84	
7	0.58	0.52	0.79	0.78	
8	0.78	0.48	0.79	0.75	
9	0.42	0.34	0.23	0.09	
10	0.35	0.53	0.39	0.62	
11	0.88	0.91	0.90	0.93	
12	0.66	0.82	0.66	0.73	
13	0.86	0.43	0.47	0.69	
14	0.42	0.31	0.51	0.53	
15	0.85	0.60	0.74	0.78	
16	0.80	0.41	0.53	0.80	
17	0.64	0.80	0.74	0.73	
18	0.58	0.40	0.49	0.45	
19	0.98	0.91	0.78	0.93	
20	0.58	0.31	0.48	0.73	
21	0.44	0.45	0.60	0.48	
22	0.41	0.37	0.51	0.45	
23	0.72	0.58	0.72	0.64	
24	0.43	0.54	0.70	0.74	
25	0.34	0.31	0.36	0.44	
26	0.39	0.57	0.82	0.72	
27	0.82	0.63	0.82	0.82	
28	0.82	0.68	0.79	0.78	
29	0.71	0.68	0.64	0.78	

	LemasSwornFT	LemasSwFTPerPop	LemasSwFTFieldOps	LemasSwFTFieldPerPop	\
0	0.030000	0.130000	0.960000	0.170000	
1	0.069655	0.217461	0.924734	0.246332	

2	0.069655	0.217461	0.924734	0.246332
3	0.069655	0.217461	0.924734	0.246332
4	0.069655	0.217461	0.924734	0.246332
5	0.069655	0.217461	0.924734	0.246332
6	0.069655	0.217461	0.924734	0.246332
7	0.069655	0.217461	0.924734	0.246332
8	0.069655	0.217461	0.924734	0.246332
9	0.069655	0.217461	0.924734	0.246332
10	0.069655	0.217461	0.924734	0.246332
11	0.069655	0.217461	0.924734	0.246332
12	0.069655	0.217461	0.924734	0.246332
13	0.069655	0.217461	0.924734	0.246332
14	0.069655	0.217461	0.924734	0.246332
15	0.069655	0.217461	0.924734	0.246332
16	0.060000	0.300000	0.930000	0.380000
17	0.069655	0.217461	0.924734	0.246332
18	0.069655	0.217461	0.924734	0.246332
19	0.069655	0.217461	0.924734	0.246332
20	0.030000	0.080000	0.960000	0.120000
21	0.250000	0.160000	0.750000	0.190000
22	0.069655	0.217461	0.924734	0.246332
23	0.010000	0.110000	0.980000	0.140000
24	0.069655	0.217461	0.924734	0.246332
25	0.069655	0.217461	0.924734	0.246332
26	0.069655	0.217461	0.924734	0.246332
27	0.069655	0.217461	0.924734	0.246332
28	0.069655	0.217461	0.924734	0.246332
29	0.069655	0.217461	0.924734	0.246332

	LemasTotalReq	LemasTotReqPerPop	PolicReqPerOffic	PolicPerPop \
0	0.060000	0.180000	0.440000	0.130000
1	0.097994	0.215204	0.343636	0.217492
2	0.097994	0.215204	0.343636	0.217492
3	0.097994	0.215204	0.343636	0.217492
4	0.097994	0.215204	0.343636	0.217492
5	0.097994	0.215204	0.343636	0.217492
6	0.097994	0.215204	0.343636	0.217492
7	0.097994	0.215204	0.343636	0.217492
8	0.097994	0.215204	0.343636	0.217492
9	0.097994	0.215204	0.343636	0.217492
10	0.097994	0.215204	0.343636	0.217492
11	0.097994	0.215204	0.343636	0.217492
12	0.097994	0.215204	0.343636	0.217492
13	0.097994	0.215204	0.343636	0.217492
14	0.097994	0.215204	0.343636	0.217492
15	0.097994	0.215204	0.343636	0.217492
16	0.050000	0.210000	0.230000	0.300000
17	0.097994	0.215204	0.343636	0.217492

18	0.097994	0.215204	0.343636	0.217492
19	0.097994	0.215204	0.343636	0.217492
20	0.050000	0.110000	0.350000	0.080000
21	0.350000	0.190000	0.380000	0.160000
22	0.097994	0.215204	0.343636	0.217492
23	0.030000	0.140000	0.370000	0.110000
24	0.097994	0.215204	0.343636	0.217492
25	0.097994	0.215204	0.343636	0.217492
26	0.097994	0.215204	0.343636	0.217492
27	0.097994	0.215204	0.343636	0.217492
28	0.097994	0.215204	0.343636	0.217492
29	0.097994	0.215204	0.343636	0.217492

	RacialMatchCommPol	PctPolicWhite	PctPolicBlack	PctPolicHispanic	\
0	0.940000	0.930000	0.03000	0.070000	
1	0.689404	0.726959	0.22047	0.134859	
2	0.689404	0.726959	0.22047	0.134859	
3	0.689404	0.726959	0.22047	0.134859	
4	0.689404	0.726959	0.22047	0.134859	
5	0.689404	0.726959	0.22047	0.134859	
6	0.689404	0.726959	0.22047	0.134859	
7	0.689404	0.726959	0.22047	0.134859	
8	0.689404	0.726959	0.22047	0.134859	
9	0.689404	0.726959	0.22047	0.134859	
10	0.689404	0.726959	0.22047	0.134859	
11	0.689404	0.726959	0.22047	0.134859	
12	0.689404	0.726959	0.22047	0.134859	
13	0.689404	0.726959	0.22047	0.134859	
14	0.689404	0.726959	0.22047	0.134859	
15	0.689404	0.726959	0.22047	0.134859	
16	0.610000	0.890000	0.15000	0.010000	
17	0.689404	0.726959	0.22047	0.134859	
18	0.689404	0.726959	0.22047	0.134859	
19	0.689404	0.726959	0.22047	0.134859	
20	0.800000	0.820000	0.04000	0.190000	
21	0.820000	0.700000	0.45000	0.030000	
22	0.689404	0.726959	0.22047	0.134859	
23	0.840000	0.960000	0.00000	0.000000	
24	0.689404	0.726959	0.22047	0.134859	
25	0.689404	0.726959	0.22047	0.134859	
26	0.689404	0.726959	0.22047	0.134859	
27	0.689404	0.726959	0.22047	0.134859	
28	0.689404	0.726959	0.22047	0.134859	
29	0.689404	0.726959	0.22047	0.134859	

	PctPolicAsian	PctPolicMinor	OfficAssgnDrugUnits	NumKindsDrugsSeiz	\
0	0.100000	0.070000	0.020000	0.57000	
1	0.114859	0.259185	0.075549	0.55605	

2	0.114859	0.259185	0.075549	0.55605
3	0.114859	0.259185	0.075549	0.55605
4	0.114859	0.259185	0.075549	0.55605
5	0.114859	0.259185	0.075549	0.55605
6	0.114859	0.259185	0.075549	0.55605
7	0.114859	0.259185	0.075549	0.55605
8	0.114859	0.259185	0.075549	0.55605
9	0.114859	0.259185	0.075549	0.55605
10	0.114859	0.259185	0.075549	0.55605
11	0.114859	0.259185	0.075549	0.55605
12	0.114859	0.259185	0.075549	0.55605
13	0.114859	0.259185	0.075549	0.55605
14	0.114859	0.259185	0.075549	0.55605
15	0.114859	0.259185	0.075549	0.55605
16	0.060000	0.120000	0.100000	0.64000
17	0.114859	0.259185	0.075549	0.55605
18	0.114859	0.259185	0.075549	0.55605
19	0.114859	0.259185	0.075549	0.55605
20	0.190000	0.180000	0.050000	0.57000
21	0.050000	0.330000	0.130000	0.57000
22	0.114859	0.259185	0.075549	0.55605
23	0.000000	0.000000	0.020000	0.86000
24	0.114859	0.259185	0.075549	0.55605
25	0.114859	0.259185	0.075549	0.55605
26	0.114859	0.259185	0.075549	0.55605
27	0.114859	0.259185	0.075549	0.55605
28	0.114859	0.259185	0.075549	0.55605
29	0.114859	0.259185	0.075549	0.55605

	PoliceAveOTWorked	LandArea	PopDens	PctUsePubTrans	PoliceCars	\
0	0.290000	0.12	0.26	0.20	0.060000	
1	0.305987	0.02	0.12	0.45	0.163103	
2	0.305987	0.01	0.21	0.02	0.163103	
3	0.305987	0.02	0.39	0.28	0.163103	
4	0.305987	0.04	0.09	0.02	0.163103	
5	0.305987	0.01	0.58	0.10	0.163103	
6	0.305987	0.05	0.08	0.06	0.163103	
7	0.305987	0.01	0.33	0.00	0.163103	
8	0.305987	0.04	0.17	0.04	0.163103	
9	0.305987	0.00	0.47	0.11	0.163103	
10	0.305987	0.02	1.00	1.00	0.163103	
11	0.305987	0.01	0.63	1.00	0.163103	
12	0.305987	0.03	0.18	0.59	0.163103	
13	0.305987	0.08	0.04	0.00	0.163103	
14	0.305987	0.02	0.40	0.15	0.163103	
15	0.305987	0.04	0.15	0.04	0.163103	
16	0.220000	0.06	0.39	0.84	0.060000	
17	0.305987	0.03	0.09	0.21	0.163103	

18	0.305987	0.03	0.20	0.07	0.163103
19	0.305987	0.09	0.03	0.05	0.163103
20	0.360000	0.09	0.46	0.05	0.090000
21	1.000000	1.00	0.07	0.15	1.000000
22	0.305987	0.05	0.09	0.06	0.163103
23	0.290000	0.16	0.12	0.07	0.040000
24	0.305987	0.02	0.22	0.02	0.163103
25	0.305987	0.02	0.15	0.16	0.163103
26	0.305987	0.01	0.79	0.21	0.163103
27	0.305987	0.04	0.09	0.00	0.163103
28	0.305987	0.02	0.17	0.00	0.163103
29	0.305987	0.10	0.02	0.08	0.163103

	PolicOperBudg	LemasPctPolicOnPatr	LemasGangUnitDeploy	\
0	0.040000	0.900000	0.500000	
1	0.076708	0.698589	0.440439	
2	0.076708	0.698589	0.440439	
3	0.076708	0.698589	0.440439	
4	0.076708	0.698589	0.440439	
5	0.076708	0.698589	0.440439	
6	0.076708	0.698589	0.440439	
7	0.076708	0.698589	0.440439	
8	0.076708	0.698589	0.440439	
9	0.076708	0.698589	0.440439	
10	0.076708	0.698589	0.440439	
11	0.076708	0.698589	0.440439	
12	0.076708	0.698589	0.440439	
13	0.076708	0.698589	0.440439	
14	0.076708	0.698589	0.440439	
15	0.076708	0.698589	0.440439	
16	0.060000	0.910000	0.500000	
17	0.076708	0.698589	0.440439	
18	0.076708	0.698589	0.440439	
19	0.076708	0.698589	0.440439	
20	0.050000	0.880000	0.500000	
21	0.350000	0.730000	0.000000	
22	0.076708	0.698589	0.440439	
23	0.010000	0.810000	1.000000	
24	0.076708	0.698589	0.440439	
25	0.076708	0.698589	0.440439	
26	0.076708	0.698589	0.440439	
27	0.076708	0.698589	0.440439	
28	0.076708	0.698589	0.440439	
29	0.076708	0.698589	0.440439	

	LemasPctOfficDrugUn	PolicBudgPerPop	ViolentCrimesPerPop
0	0.32	0.140000	0.20
1	0.00	0.195078	0.67

2	0.00	0.195078	0.43
3	0.00	0.195078	0.12
4	0.00	0.195078	0.03
5	0.00	0.195078	0.14
6	0.00	0.195078	0.03
7	0.00	0.195078	0.55
8	0.00	0.195078	0.53
9	0.00	0.195078	0.15
10	0.00	0.195078	0.24
11	0.00	0.195078	0.08
12	0.00	0.195078	0.06
13	0.00	0.195078	0.09
14	0.00	0.195078	0.21
15	0.00	0.195078	0.30
16	0.88	0.260000	0.49
17	0.00	0.195078	0.07
18	0.00	0.195078	0.15
19	0.00	0.195078	0.03
20	0.76	0.130000	0.34
21	0.31	0.210000	0.69
22	0.00	0.195078	0.21
23	0.56	0.090000	0.63
24	0.00	0.195078	0.31
25	0.00	0.195078	0.12
26	0.00	0.195078	0.84
27	0.00	0.195078	0.10
28	0.00	0.195078	0.49
29	0.00	0.195078	0.02

## 1.2 Perform PCA

The data used for the PCA starts from column 4 onwards.

This decision was made because the county state or community cannot use mean to fill NaN values. Alternatively the mode could be used to replace the NaN values or the sample could have been dropped

```
In [4]: # Create a PCA instance: pca
        pca = PCA()

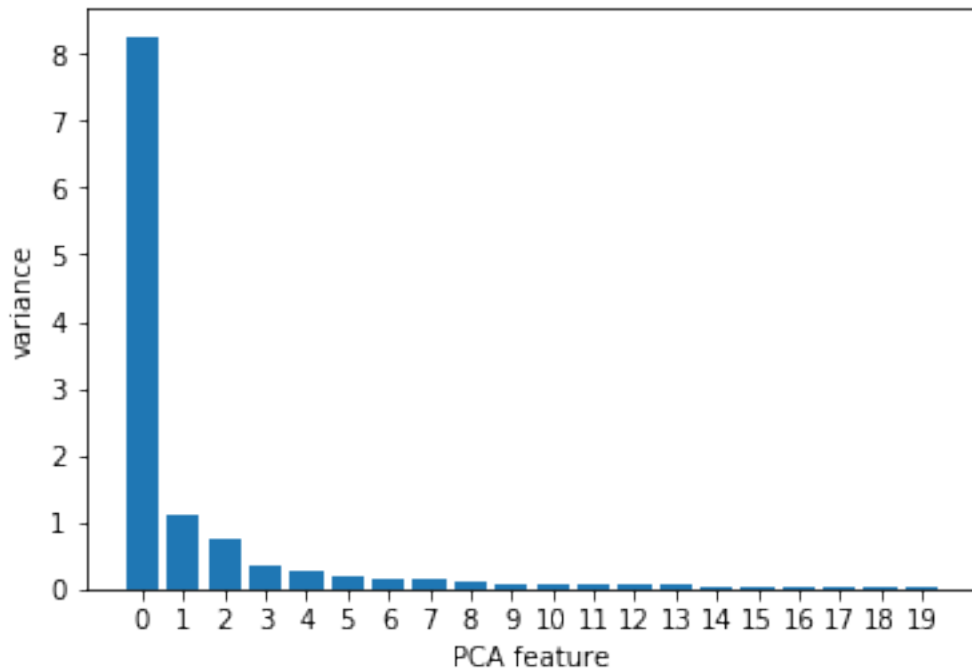
        # Fit the data
        pca.fit(data_filled.iloc[:,4:])

        # Plot the explained variances
        n = 20
        features = range(pca.n_components_)
        plt.bar(features[:n], pca.explained_variance_[:n])
        plt.xlabel('PCA feature')
```

```

plt.ylabel('variance')
plt.xticks(features[:n])
plt.show()
eigen_values = pd.Series(pca.explained_variance_)[:n]
display(eigen_values)

```



```

0      8.261395
1      1.089839
2      0.760844
3      0.333962
4      0.287441
5      0.186957
6      0.161923
7      0.133371
8      0.111869
9      0.088794
10     0.076732
11     0.057319
12     0.055713
13     0.053844
14     0.048041
15     0.045104
16     0.039366
17     0.033291
18     0.029954

```

```
19      0.028349
dtype: float64
```

```
In [5]: # Determine how many features must be kept to retain 95% of the variance
total_variance = np.sum(eigen_values)
num_pca_features = []
for num_components in range(20):
    ratio = np.sum(eigen_values[:num_components]) / total_variance
    num_pca_features.append(ratio)
    print('number of PCs: ', num_components, ' Variance Ratio: ', ratio)
```

```
number of PCs: 0  Variance Ratio: 0.0
number of PCs: 1  Variance Ratio: 0.6951633856327811
number of PCs: 2  Variance Ratio: 0.7868689827183053
number of PCs: 3  Variance Ratio: 0.8508909735025484
number of PCs: 4  Variance Ratio: 0.8789925212379436
number of PCs: 5  Variance Ratio: 0.9031794951857989
number of PCs: 6  Variance Ratio: 0.9189111428813157
number of PCs: 7  Variance Ratio: 0.932536311317395
number of PCs: 8  Variance Ratio: 0.9437589290582613
number of PCs: 9  Variance Ratio: 0.9531722386249255
number of PCs: 10 Variance Ratio: 0.9606438938352004
number of PCs: 11 Variance Ratio: 0.9671005844367883
number of PCs: 12 Variance Ratio: 0.9719237310352683
number of PCs: 13 Variance Ratio: 0.9766117595932257
number of PCs: 14 Variance Ratio: 0.9811425240972542
number of PCs: 15 Variance Ratio: 0.9851849662681417
number of PCs: 16 Variance Ratio: 0.9889802763341629
number of PCs: 17 Variance Ratio: 0.9922927319284601
number of PCs: 18 Variance Ratio: 0.9950940675282125
number of PCs: 19 Variance Ratio: 0.9976145618995417
```

We can see the top 20 pca features in the above table and scree plot figure. To me there is not a clear point where the dimensions could be cut off, however there are a couple techniques we could use to make the decision. - Looking at the scree plot, the variance begins to drop more drastically starting at PCA feature 3. Therefore we could arbitrarily choose PCA 0, 1, and 2 to represent the data - The second way would be to sum the value of the eigen\_values, and choose enough features to retain some ratio of the total variance. We can see that if we keep the top 9 highest variance features we will retain over 95% of the variance

### 1.3 Dimnsionality Reduction

This is what the sample would look like if the first two components are used and plotted on a 2d scatter plot

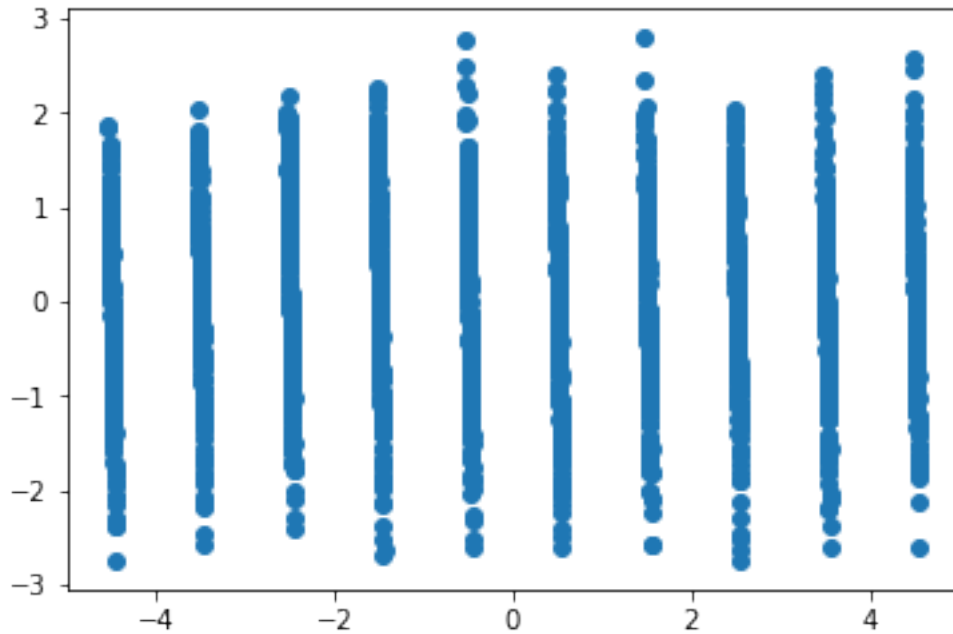
```
In [6]: # dimensionality reduction
pca = PCA(n_components=2)
```



```

# Fit the data
pca.fit(data_filled.iloc[:,4:])
transformed = pca.transform(data_filled.iloc[:,4:])
xs = transformed[:,0]
ys = transformed[:,1]
plt.scatter(xs, ys)
plt.show()

```



## 1.4 Manual Calculation of PCA and eigenvalues

For completeness the eigenvalues and eigenvectors were computed using only numpy. We can see that the values closely resemble sklearn's returned results. Sklearn uses SVD in the background to compute the PCA.

```

In [7]: X = data_filled.iloc[:,4:]
n_samples = X.shape[0]
# We center the data and compute the sample covariance matrix.
X -= np.mean(X, axis=0)
cov_matrix = np.dot(X.T, X) / n_samples

eigenValues, eigenVectors = LA.eig(cov_matrix)

idx = eigenValues.argsort()[::-1]
eigenValues = pd.Series(eigenValues[idx])

```

```
eigenVectors = pd.DataFrame(eigenVectors[:,idx])
display(eigenValues[:20])
```

0	8.257252
1	1.089292
2	0.760463
3	0.333794
4	0.287296
5	0.186863
6	0.161842
7	0.133304
8	0.111813
9	0.088749
10	0.076694
11	0.057290
12	0.055685
13	0.053817
14	0.048017
15	0.045081
16	0.039346
17	0.033275
18	0.029939
19	0.028335

dtype: float64