

```
In [29]: import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import scipy.stats as stats
from sklearn.decomposition import PCA
```

Load The Data Set

```
In [79]: data = pd.read_excel('marriage.xlsx', header=5)
data.head()
```

Out[79]:

	Unnamed: 0	2019	2018	2017	2016	2015	2014	2013	2012
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	Alabama	6.697687	6.760408	7.047340	7.147821	7.351544	7.806776	7.817785	8.2
2	Alaska	6.512245	6.683952	6.914078	7.103441	7.407588	7.508836	7.293928	7.2
3	Arizona	5.302995	5.534434	5.834867	5.930541	5.922469	5.780449	5.401091	5.6
4	Arkansas	8.377284	8.863156	9.456845	9.860962	10.040279	10.112026	9.751052	10.9

5 rows × 24 columns



In [80]: data

Out[80]:

	Unnamed: 0	2019	2018	2017	2016	2015	2014	2013
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	Alabama	6.697687	6.760408	7.047340	7.147821	7.351544	7.806776	7.81771
2	Alaska	6.512245	6.683952	6.914078	7.103441	7.407588	7.508836	7.2939
3	Arizona	5.302995	5.534434	5.834867	5.930541	5.922469	5.780449	5.4010
4	Arkansas	8.377284	8.863156	9.456845	9.860962	10.040279	10.112026	9.7510
5	California	5.723191	6.035132	6.278250	6.463590	6.184957	6.441492	6.4604
6	Colorado	7.273297	7.585728	7.333845	7.425443	6.791807	7.061603	6.4526
7	Connecticut	5.048401	5.278133	5.553784	5.617858	5.292009	5.368845	5.0210
8	Delaware	4.951919	5.237957	5.528417	5.613062	5.712872	6.022783	6.5719
9	District of Columbia	7.773302	7.835377	8.239526	8.149214	8.220425	11.821343	10.7912
10	Florida	7.070065	7.332063	7.806895	8.125967	8.234362	7.301404	7.0096
11	Georgia	6.038471	6.391479	6.870975	6.753976	6.200379	---	---
12	Hawaii	14.172891	15.263736	15.346702	15.555557	15.940173	17.702656	16.2942
13	Idaho	7.389770	7.810362	7.796997	8.077759	8.151402	8.366045	8.178
14	Illinois	5.162794	5.478499	5.982023	5.800000	5.880873	6.169054	5.3940
15	Indiana	6.176270	6.554662	6.854694	6.935419	6.856525	7.08004	6.6074
16	Iowa	5.403684	5.737696	6.176028	6.149566	6.255004	6.863899	7.3909
17	Kansas	5.341683	5.366984	5.970225	6.210941	5.936515	6.101884	5.984
18	Kentucky	6.265454	6.774234	7.174146	7.379354	7.172506	6.941724	7.2786
19	Louisiana	5.094870	5.118908	5.592472	6.106373	6.777108	6.892953	6.3708
20	Maine	7.053203	7.390145	7.585857	7.640376	7.610612	7.745346	8.310
21	Maryland	5.592919	5.867393	6.268158	6.308374	6.187233	6.461575	6.8237
22	Massachusetts	5.019222	6.269931	5.749423	5.777052	5.497009	5.60485	5.5014
23	Michigan	5.183913	5.678220	5.887088	5.935155	5.984938	5.759103	5.8122
24	Minnesota	5.074090	5.304589	5.611836	5.617984	5.578555	5.895727	5.9982
25	Mississippi	6.018516	6.282877	6.720619	7.013691	6.993874	6.930345	6.7156
26	Missouri	5.969276	6.478791	6.594551	6.851633	6.800000	6.725225	6.4498
27	Montana	7.886577	7.713416	7.963880	7.840617	7.962639	7.857723	7.3987
28	Nebraska	5.498323	5.976878	6.313813	6.465784	6.378053	6.415084	6.2948
29	Nevada	25.894792	26.734186	28.556333	28.392297	31.017920	31.85095	32.2801
30	New Hampshire	6.637440	6.932762	7.030857	6.977101	6.938182	7.187147	6.9016
31	New Jersey	5.187009	5.367783	5.450804	5.654780	5.606042	5.403116	5.1308
32	New Mexico	5.981413	6.449279	5.925568	6.366124	6.171860	8.080277	7.2555
33	New York	7.217085	7.106515	7.344656	7.451752	7.056096	6.689784	6.8901
34	North Carolina	6.181491	6.433883	6.846406	6.967624	6.959910	6.892221	6.5100
35	North Dakota	5.423443	5.682319	5.761240	5.955522	6.172326	6.300356	6.3409
36	Ohio	5.329324	5.617035	5.821192	5.956154	5.905580	5.751515	5.7089

	Unnamed: 0	2019	2018	2017	2016	2015	2014	2013
37	Oklahoma	6.336918	6.371924	6.815296	6.678882	7.418945	7.075203	7.08331
38	Oregon	5.991839	6.322075	6.660993	6.859226	6.887356	6.834097	6.33471
39	Pennsylvania	5.351512	5.545691	5.724399	5.778683	5.684592	5.84287	5.43221
40	Rhode Island	6.149934	6.297083	6.767399	6.710361	6.353321	6.669996	6.21291
41	South Carolina	6.288172	6.585398	7.034316	6.632979	7.498347	7.624033	7.14201
42	South Dakota	6.144741	6.535674	6.742819	7.247063	7.214005	7.073578	7.00571
43	Tennessee	7.488753	7.951096	8.234385	8.614092	8.502342	8.409229	8.44781
44	Texas	4.867898	6.121488	7.096127	7.077625	7.214466	6.851589	6.97631
45	Utah	8.058122	8.355306	8.720650	8.990183	8.075986	7.317947	7.50361
46	Vermont	7.726098	7.943490	7.908193	8.299792	8.124056	8.698261	9.15531
47	Virginia	6.128508	6.398452	6.783573	6.988628	6.980323	6.744421	6.65771
48	Washington	5.749260	6.023947	6.246909	6.235593	6.219361	6.956141	7.1101
49	West Virginia	5.978862	6.067010	6.311620	6.354097	6.574923	6.669095	6.64401
50	Wisconsin	5.037240	5.430056	5.634561	5.616134	5.611351	5.691296	5.21911
51	Wyoming	7.008098	7.051652	7.125657	7.079407	7.341663	7.658952	7.54981
52	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
53	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
54	--- Data not available.	NaN	NaN	NaN	NaN	NaN	NaN	NaN
55	1 Marriage data includes nonlicensed marriages...	NaN	NaN	NaN	NaN	NaN	NaN	NaN
56	Note: Rate for 2015 for Missouri and for 2016 ...	NaN	NaN	NaN	NaN	NaN	NaN	NaN
57	Source: CDC/NCHS, National Vital Statistics Sy...	NaN	NaN	NaN	NaN	NaN	NaN	NaN

58 rows × 24 columns

Cleaning Process

```
In [31]: data.drop([0,52,53,54,55,56,57],axis=0,inplace=True)  
data
```

Out[31]:

	Unnamed: 0	2019	2018	2017	2016	2015	2014	2013
1	Alabama	6.697687	6.760408	7.047340	7.147821	7.351544	7.806776	7.817761
2	Alaska	6.512245	6.683952	6.914078	7.103441	7.407588	7.508836	7.293951
3	Arizona	5.302995	5.534434	5.834867	5.930541	5.922469	5.780449	5.401091
4	Arkansas	8.377284	8.863156	9.456845	9.860962	10.040279	10.112026	9.751091
5	California	5.723191	6.035132	6.278250	6.463590	6.184957	6.441492	6.460491
6	Colorado	7.273297	7.585728	7.333845	7.425443	6.791807	7.061603	6.452691
7	Connecticut	5.048401	5.278133	5.553784	5.617858	5.292009	5.368845	5.021091
8	Delaware	4.951919	5.237957	5.528417	5.613062	5.712872	6.022783	6.571951
9	District of Columbia	7.773302	7.835377	8.239526	8.149214	8.220425	11.821343	10.791291
10	Florida	7.070065	7.332063	7.806895	8.125967	8.234362	7.301404	7.009691
11	Georgia	6.038471	6.391479	6.870975	6.753976	6.200379	---	---
12	Hawaii	14.172891	15.263736	15.346702	15.555557	15.940173	17.702656	16.294291
13	Idaho	7.389770	7.810362	7.796997	8.077759	8.151402	8.366045	8.178951
14	Illinois	5.162794	5.478499	5.982023	5.800000	5.880873	6.169054	5.394091
15	Indiana	6.176270	6.554662	6.854694	6.935419	6.856525	7.080004	6.607491
16	Iowa	5.403684	5.737696	6.176028	6.149566	6.255004	6.863899	7.390951
17	Kansas	5.341683	5.366984	5.970225	6.210941	5.936515	6.101884	5.984951
18	Kentucky	6.265454	6.774234	7.174146	7.379354	7.172506	6.941724	7.278691
19	Louisiana	5.094870	5.118908	5.592472	6.106373	6.777108	6.892953	6.370891
20	Maine	7.053203	7.390145	7.585857	7.640376	7.610612	7.745346	8.310951
21	Maryland	5.592919	5.867393	6.268158	6.308374	6.187233	6.461575	6.823791
22	Massachusetts	5.019222	6.269931	5.749423	5.777052	5.497009	5.60485	5.501491
23	Michigan	5.183913	5.678220	5.887088	5.935155	5.984938	5.759103	5.812291
24	Minnesota	5.074090	5.304589	5.611836	5.617984	5.578555	5.895727	5.998291
25	Mississippi	6.018516	6.282877	6.720619	7.013691	6.993874	6.930345	6.715691
26	Missouri	5.969276	6.478791	6.594551	6.851633	6.800000	6.725225	6.449891
27	Montana	7.886577	7.713416	7.963880	7.840617	7.962639	7.857723	7.398791
28	Nebraska	5.498323	5.976878	6.313813	6.465784	6.378053	6.415084	6.294891
29	Nevada	25.894792	26.734186	28.556333	28.392297	31.017920	31.85095	32.280191
30	New Hampshire	6.637440	6.932762	7.030857	6.977101	6.938182	7.187147	6.901691
31	New Jersey	5.187009	5.367783	5.450804	5.654780	5.606042	5.403116	5.130891
32	New Mexico	5.981413	6.449279	5.925568	6.366124	6.171860	8.080277	7.255591
33	New York	7.217085	7.106515	7.344656	7.451752	7.056096	6.689784	6.890191
34	North Carolina	6.181491	6.433883	6.846406	6.967624	6.959910	6.892221	6.510091
35	North Dakota	5.423443	5.682319	5.761240	5.955522	6.172326	6.300356	6.340991
36	Ohio	5.329324	5.617035	5.821192	5.956154	5.905580	5.751515	5.708991
37	Oklahoma	6.336918	6.371924	6.815296	6.678882	7.418945	7.075203	7.083391

	Unnamed: 0	2019	2018	2017	2016	2015	2014	2013
38	Oregon	5.991839	6.322075	6.660993	6.859226	6.887356	6.834097	6.33471
39	Pennsylvania	5.351512	5.545691	5.724399	5.778683	5.684592	5.84287	5.43229
40	Rhode Island	6.149934	6.297083	6.767399	6.710361	6.353321	6.669996	6.21290
41	South Carolina	6.288172	6.585398	7.034316	6.632979	7.498347	7.624033	7.14203
42	South Dakota	6.144741	6.535674	6.742819	7.247063	7.214005	7.073578	7.00571
43	Tennessee	7.488753	7.951096	8.234385	8.614092	8.502342	8.409229	8.44784
44	Texas	4.867898	6.121488	7.096127	7.077625	7.214466	6.851589	6.97631
45	Utah	8.058122	8.355306	8.720650	8.990183	8.075986	7.317947	7.50361
46	Vermont	7.726098	7.943490	7.908193	8.299792	8.124056	8.698261	9.15531
47	Virginia	6.128508	6.398452	6.783573	6.988628	6.980323	6.744421	6.65771
48	Washington	5.749260	6.023947	6.246909	6.235593	6.219361	6.956141	7.11031
49	West Virginia	5.978862	6.067010	6.311620	6.354097	6.574923	6.669095	6.64401
50	Wisconsin	5.037240	5.430056	5.634561	5.616134	5.611351	5.691296	5.21911
51	Wyoming	7.008098	7.051652	7.125657	7.079407	7.341663	7.658952	7.54981

51 rows × 24 columns

In [32]: data = data.replace('---', np.nan)
data

41	South Carolina	6.288172	6.585398	7.034316	6.632979	7.498347	7.624033	7.14203
42	South Dakota	6.144741	6.535674	6.742819	7.247063	7.214005	7.073578	7.00571
43	Tennessee	7.488753	7.951096	8.234385	8.614092	8.502342	8.409229	8.44784
44	Texas	4.867898	6.121488	7.096127	7.077625	7.214466	6.851589	6.97631
45	Utah	8.058122	8.355306	8.720650	8.990183	8.075986	7.317947	7.50361
46	Vermont	7.726098	7.943490	7.908193	8.299792	8.124056	8.698261	9.15531
47	Virginia	6.128508	6.398452	6.783573	6.988628	6.980323	6.744421	6.65771
48	Washington	5.749260	6.023947	6.246909	6.235593	6.219361	6.956141	7.11031
49	West Virginia	5.978862	6.067010	6.311620	6.354097	6.574923	6.669095	6.64401
50	Wisconsin	5.037240	5.430056	5.634561	5.616134	5.611351	5.691296	5.21911
51	Wyoming	7.008098	7.051652	7.125657	7.079407	7.341663	7.658952	7.54981

51 rows × 24 columns

```
In [33]: data.fillna(data.median(numeric_only=True),inplace=True) #get only numeric columns  
data
```


Out[33]:

	Unnamed: 0	2019	2018	2017	2016	2015	2014	2013
1	Alabama	6.697687	6.760408	7.047340	7.147821	7.351544	7.806776	7.81771
2	Alaska	6.512245	6.683952	6.914078	7.103441	7.407588	7.508836	7.2939
3	Arizona	5.302995	5.534434	5.834867	5.930541	5.922469	5.780449	5.4010
4	Arkansas	8.377284	8.863156	9.456845	9.860962	10.040279	10.112026	9.7510
5	California	5.723191	6.035132	6.278250	6.463590	6.184957	6.441492	6.4604
6	Colorado	7.273297	7.585728	7.333845	7.425443	6.791807	7.061603	6.4526
7	Connecticut	5.048401	5.278133	5.553784	5.617858	5.292009	5.368845	5.0210
8	Delaware	4.951919	5.237957	5.528417	5.613062	5.712872	6.022783	6.5719
9	District of Columbia	7.773302	7.835377	8.239526	8.149214	8.220425	11.821343	10.7912
10	Florida	7.070065	7.332063	7.806895	8.125967	8.234362	7.301404	7.0096
11	Georgia	6.038471	6.391479	6.870975	6.753976	6.200379	6.892587	6.7697
12	Hawaii	14.172891	15.263736	15.346702	15.555557	15.940173	17.702656	16.2942
13	Idaho	7.389770	7.810362	7.796997	8.077759	8.151402	8.366045	8.1785
14	Illinois	5.162794	5.478499	5.982023	5.800000	5.880873	6.169054	5.3940
15	Indiana	6.176270	6.554662	6.854694	6.935419	6.856525	7.080040	6.6074
16	Iowa	5.403684	5.737696	6.176028	6.149566	6.255004	6.863899	7.3909
17	Kansas	5.341683	5.366984	5.970225	6.210941	5.936515	6.101884	5.9845
18	Kentucky	6.265454	6.774234	7.174146	7.379354	7.172506	6.941724	7.2786
19	Louisiana	5.094870	5.118908	5.592472	6.106373	6.777108	6.892953	6.3708
20	Maine	7.053203	7.390145	7.585857	7.640376	7.610612	7.745346	8.3106
21	Maryland	5.592919	5.867393	6.268158	6.308374	6.187233	6.461575	6.8237
22	Massachusetts	5.019222	6.269931	5.749423	5.777052	5.497009	5.604850	5.5014
23	Michigan	5.183913	5.678220	5.887088	5.935155	5.984938	5.759103	5.8122
24	Minnesota	5.074090	5.304589	5.611836	5.617984	5.578555	5.895727	5.9982
25	Mississippi	6.018516	6.282877	6.720619	7.013691	6.993874	6.930345	6.7156
26	Missouri	5.969276	6.478791	6.594551	6.851633	6.800000	6.725225	6.4498
27	Montana	7.886577	7.713416	7.963880	7.840617	7.962639	7.857723	7.3987
28	Nebraska	5.498323	5.976878	6.313813	6.465784	6.378053	6.415084	6.2948
29	Nevada	25.894792	26.734186	28.556333	28.392297	31.017920	31.850950	32.2801
30	New Hampshire	6.637440	6.932762	7.030857	6.977101	6.938182	7.187147	6.9016
31	New Jersey	5.187009	5.367783	5.450804	5.654780	5.606042	5.403116	5.1308
32	New Mexico	5.981413	6.449279	5.925568	6.366124	6.171860	8.080277	7.2555
33	New York	7.217085	7.106515	7.344656	7.451752	7.056096	6.689784	6.8901
34	North Carolina	6.181491	6.433883	6.846406	6.967624	6.959910	6.892221	6.5100
35	North Dakota	5.423443	5.682319	5.761240	5.955522	6.172326	6.300356	6.3409
36	Ohio	5.329324	5.617035	5.821192	5.956154	5.905580	5.751515	5.7089
37	Oklahoma	6.336918	6.371924	6.815296	6.678882	7.418945	7.075203	7.0833

	Unnamed: 0	2019	2018	2017	2016	2015	2014	2013
38	Oregon	5.991839	6.322075	6.660993	6.859226	6.887356	6.834097	6.33471
39	Pennsylvania	5.351512	5.545691	5.724399	5.778683	5.684592	5.842870	5.43221
40	Rhode Island	6.149934	6.297083	6.767399	6.710361	6.353321	6.669996	6.21290
41	South Carolina	6.288172	6.585398	7.034316	6.632979	7.498347	7.624033	7.14201
42	South Dakota	6.144741	6.535674	6.742819	7.247063	7.214005	7.073578	7.00571
43	Tennessee	7.488753	7.951096	8.234385	8.614092	8.502342	8.409229	8.44781
44	Texas	4.867898	6.121488	7.096127	7.077625	7.214466	6.851589	6.97631
45	Utah	8.058122	8.355306	8.720650	8.990183	8.075986	7.317947	7.50361
46	Vermont	7.726098	7.943490	7.908193	8.299792	8.124056	8.698261	9.15531
47	Virginia	6.128508	6.398452	6.783573	6.988628	6.980323	6.744421	6.65771
48	Washington	5.749260	6.023947	6.246909	6.235593	6.219361	6.956141	7.11031
49	West Virginia	5.978862	6.067010	6.311620	6.354097	6.574923	6.669095	6.64401
50	Wisconsin	5.037240	5.430056	5.634561	5.616134	5.611351	5.691296	5.21911
51	Wyoming	7.008098	7.051652	7.125657	7.079407	7.341663	7.658952	7.54981

51 rows × 24 columns

Analysis

In [34]: *#Extract to Matrix*

```
yearM = data.columns[1:].to_numpy().astype(float)
stateM = data.iloc[:,0]
M = data.iloc[:,1:].to_numpy()
M
```

Out[34]: array([[6.69768732, 6.76040755, 7.0473401 , ..., 10.8 ,
 9.8 , 10.6],
 [6.51224463, 6.68395174, 6.91407755, ..., 8.6 ,
 9. , 10.2],
 [5.30299502, 5.5344338 , 5.83486667, ..., 8.2 ,
 8.8 , 10.],
 ...,
 [5.97886223, 6.06700956, 6.31162035, ..., 7.5 ,
 6.1 , 7.2],
 [5.03724044, 5.43005603, 5.63456057, ..., 6.7 ,
 7. , 7.9],
 [7.00809836, 7.05165153, 7.12565703, ..., 9.9 ,
 10.6 , 10.7]])

In [35]: *#Make Some Plots*

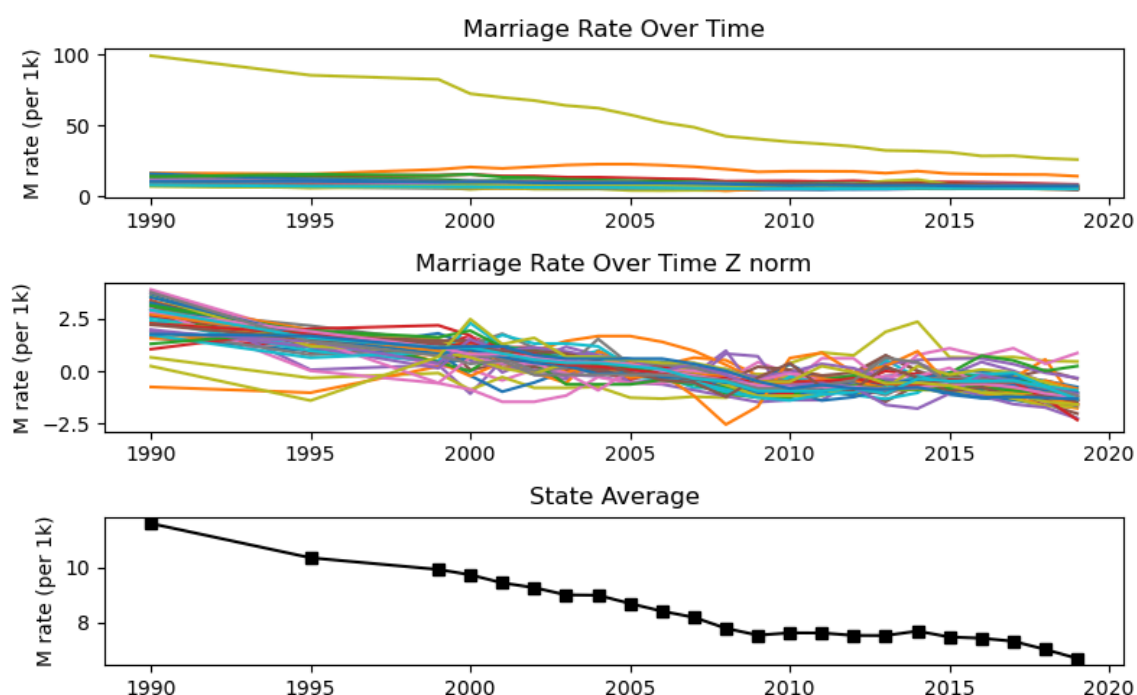
```
fig,ax = plt.subplots(3,1,figsize=(8,5))

ax[0].plot(yearM,M.T)
ax[0].set_title('Marriage Rate Over Time')
ax[0].set_ylabel('M rate (per 1k)')

ax[1].plot(yearM,stats.zscore(M.T))
ax[1].set_title('Marriage Rate Over Time Z norm')
ax[1].set_ylabel('M rate (per 1k)')

ax[2].plot(yearM,np.mean(M,axis=0),'ks-')
ax[2].set_title('State Average')
ax[2].set_ylabel('M rate (per 1k)')

plt.tight_layout()
plt.show()
```



In [36]: `meanM = np.mean(M,axis=1)`
`sinM = np.argsort(meanM)`

In [37]: `meanM`

Out[37]: `array([8.66214617, 7.96626378, 6.52203674, 11.84615666, 6.26900344,`
 `7.5445386 , 5.65130664, 6.00604285, 6.78393251, 8.32088568,`
 `6.99206978, 18.3119983 , 9.74656206, 6.2116219 , 7.32891628,`
 `6.84681698, 6.74838136, 8.42548302, 7.27189114, 7.94070218,`
 `6.66562802, 5.93560441, 6.10611667, 5.99048123, 6.35980892,`
 `7.06388375, 7.59233254, 6.83664216, 50.32289676, 7.50891741,`
 `5.63914628, 6.96652681, 7.09808813, 7.15180648, 6.52765902,`
 `6.45172763, 7.2795886 , 7.08653654, 5.7460886 , 6.82439382,`
 `8.45240286, 7.9766797 , 10.57164091, 7.71328474, 9.19659999,`
 `8.96327011, 7.75572571, 6.62789308, 7.03476565, 5.93216411,`
 `8.57892664])`

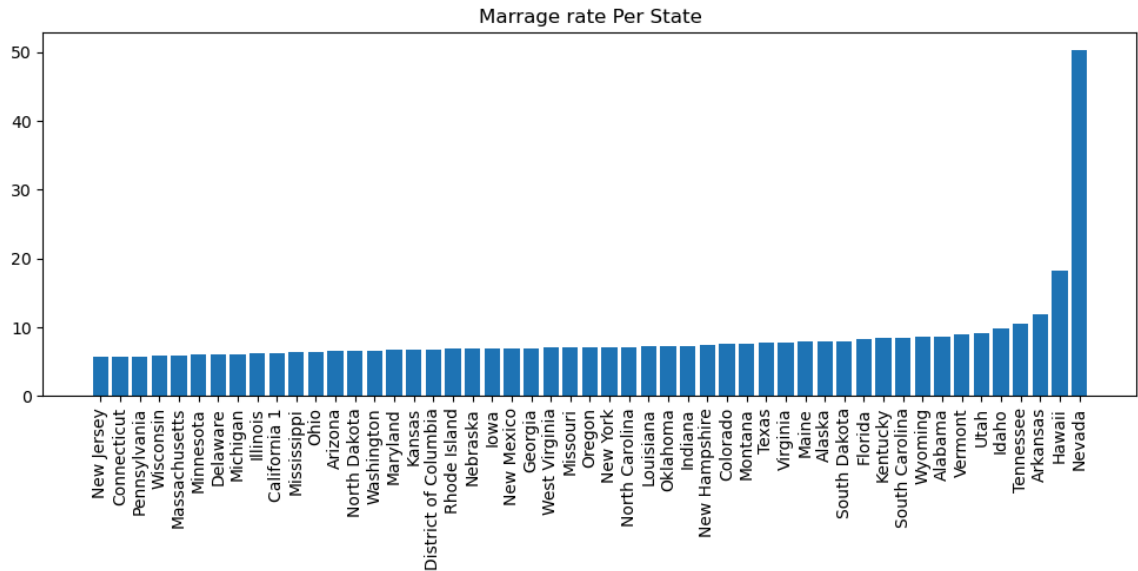
```
In [38]: stateM
```

```
Out[38]: 1           Alabama
2           Alaska
3           Arizona
4           Arkansas
5      California 1
6           Colorado
7           Connecticut
8           Delaware
9      District of Columbia
10          Florida
11          Georgia
12          Hawaii
13          Idaho
14          Illinois
15          Indiana
16          Iowa
17          Kansas
18          Kentucky
19          Louisiana
20          Maine
21          Maryland
22      Massachusetts
23          Michigan
24          Minnesota
25          Mississippi
26          Missouri
27          Montana
28          Nebraska
29          Nevada
30      New Hampshire
31          New Jersey
32          New Mexico
33          New York
34      North Carolina
35          North Dakota
36          Ohio
37          Oklahoma
38          Oregon
39          Pennsylvania
40          Rhode Island
41      South Carolina
42          South Dakota
43          Tennessee
44          Texas
45          Utah
46          Vermont
47          Virginia
48          Washington
49      West Virginia
50          Wisconsin
51          Wyoming
Name: Unnamed: 0, dtype: object
```

```
In [39]: fig=plt.figure(figsize=(12,4))
plt.bar(stateM.iloc[sinM],meanM[sinM])
plt.xticks(rotation=90)
plt.title('Marrage rate Per State')
plt.show()
```

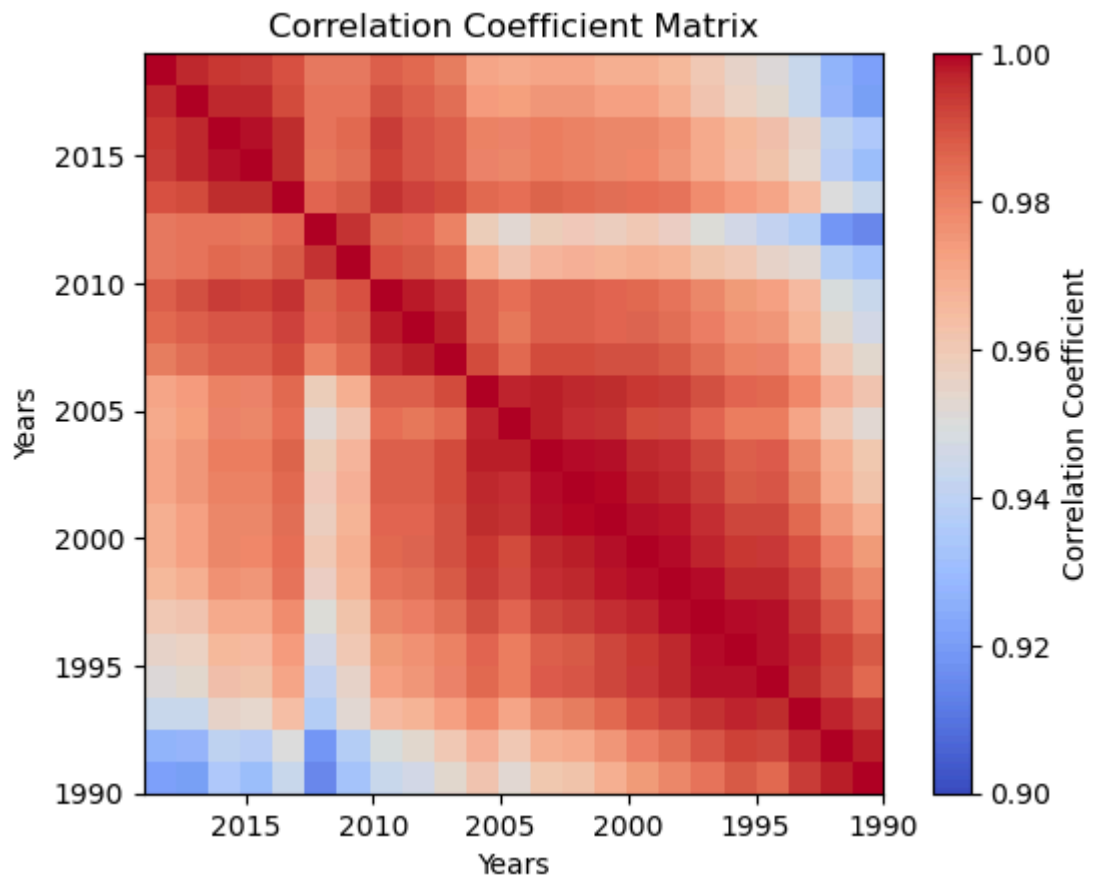
#Q

#is Navada a non representative data Or Error ?

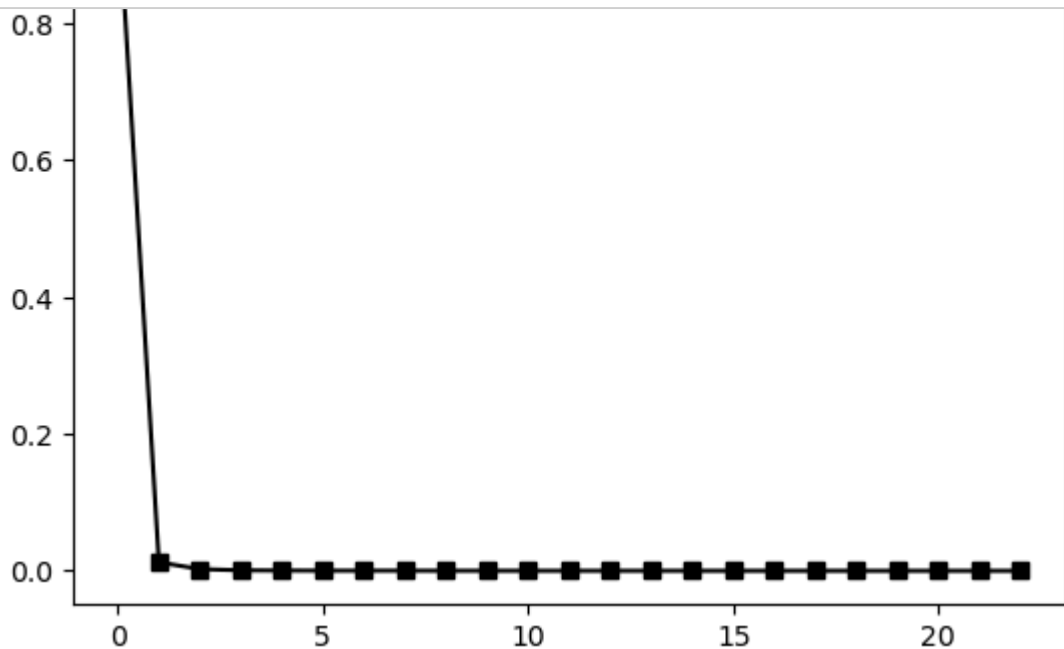


```
In [40]: #Correlation Matrix
corr_matrix = np.corrcoef(M.T)

# Plotting the correlation matrix
plt.imshow(corr_matrix, vmin=0.9, vmax=1, extent=[yearM[0], yearM[-1], yearM[0], yearM[-1]])
plt.colorbar(label='Correlation Coefficient')
plt.xlabel('Years')
plt.ylabel('Years')
plt.title('Correlation Coefficient Matrix')
plt.show()
```



```
In [41]: # PCA
pca = PCA().fit(M)
plt.plot(pca.explained_variance_ratio_, 'ks-')
plt.show()
```



Divorce Data Analysis

```
In [43]: dataD = pd.read_excel('divorce.xlsx', header=5)
dataD.head()
```

Out[43]:

	Unnamed: 0	2019	2018	2017	2016	2015	2014	2013	2012	2011
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	Alabama	3.572984	3.714296	3.660498	3.77871	3.85369	3.785022	3.68329	3.6	4
2	Alaska	3.638874	3.734551	3.580722	3.946655	4.138499	3.951233	4.461784	4.5	4
3	Arizona	2.915899	2.954273	3.479199	3.352007	3.578466	3.89201	3.92915	4.3	3
4	Arkansas	4.017159	4.066925	3.719694	3.926381	4.842516	4.845318	5.011872	5.3	5

5 rows × 24 columns

```
In [44]: dataD.drop([0,52,53,54,55,56,57],axis=0,inplace=True)  
dataD
```


Out[44]:

	Unnamed: 0	2019	2018	2017	2016	2015	2014	2013	2012
1	Alabama	3.572984	3.714296	3.660498	3.77871	3.85369	3.785022	3.68329	3.6
2	Alaska	3.638874	3.734551	3.580722	3.946655	4.138499	3.951233	4.461784	4.5
3	Arizona	2.915899	2.954273	3.479199	3.352007	3.578466	3.89201	3.92915	4.3
4	Arkansas	4.017159	4.066925	3.719694	3.926381	4.842516	4.845318	5.011872	5.3
5	California	---	---	---	---	---	---	---	--
6	Colorado	3.291	3.2736	3.186287	3.566256	3.725048	3.866602	4.073558	4.3
7	Connecticut	2.690386	2.853612	2.870533	3.168503	3.128197	2.577657	2.761618	2.7
8	Delaware	2.915491	2.800952	2.964845	3.079622	3.093239	3.256685	3.383206	3.5
9	District of Columbia	2.422958	2.521158	2.477045	2.739404	2.832372	2.57705	2.838584	2.9
10	Florida	3.489101	3.617673	3.612112	3.865675	3.956535	3.970182	4.110652	4.2
11	Georgia	2.287278	2.502691	3.496277	---	---	---	---	--
12	Hawaii	---	---	---	---	---	---	---	--
13	Idaho	3.857722	3.838199	3.884812	4.034127	4.12525	4.248488	4.495899	4.7
14	Illinois	1.293579	1.506623	1.870798	2.2	2.196579	2.213332	2.270509	2.4
15	Indiana	---	---	---	---	---	---	---	--
16	Iowa	2.258904	2.232153	1.963944	1.254668	1.211307	1.50525	1.865768	2.2
17	Kansas	2.32862	2.309802	2.379234	2.692543	2.82109	2.967954	2.972055	3.4
18	Kentucky	3.433107	3.54422	3.726604	3.788167	3.702296	3.849137	4.162633	4.1
19	Louisiana	2.399117	1.736703	1.923219	2.042863	2.772589	2.250479	2.170806	--
20	Maine	3.033004	3.197091	3.226272	3.428518	3.428048	3.581715	3.96973	3.9
21	Maryland	2.661901	2.400575	2.542721	2.686968	2.595564	2.541828	2.533559	2.8
22	Massachusetts	1.463184	2.116008	2.165801	2.290151	2.552977	2.680787	2.626694	2.7
23	Michigan	2.328661	2.844962	2.848937	2.908151	3.045681	2.997817	3.259926	3.3
24	Minnesota	---	---	---	---	---	---	---	--
25	Mississippi	2.625541	2.71787	2.900372	3.20471	3.37028	3.371321	3.587181	4
26	Missouri	2.890462	3.040912	3.138448	3.281963	3.177357	3.278092	3.42214	3.7
27	Montana	2.997816	3.033027	3.109968	3.104976	3.425145	3.447707	3.388612	3.9
28	Nebraska	2.663864	2.949305	3.024359	3.136149	3.177952	3.054473	3.278538	3.4
29	Nevada	4.170243	4.376824	4.459582	4.294473	4.636707	5.330212	5.121614	5.5
30	New Hampshire	3.072712	3.089664	3.11291	3.37505	3.328554	3.543077	3.726598	3.6
31	New Jersey	2.519086	2.746809	2.600813	2.668465	2.817924	2.778755	2.785263	2.8
32	New Mexico	---	---	---	---	3.288078	3.551544	3.437417	3
33	New York	2.89253	2.780085	2.676303	2.681045	2.707495	2.843986	2.735467	2.9
34	North Carolina	3.13327	3.075324	3.109968	3.248319	3.082307	3.401963	3.429203	3.7
35	North Dakota	2.481425	2.557636	2.486123	2.589874	2.839111	2.83036	2.943075	3.1
36	Ohio	2.768733	2.905956	2.922647	3.029867	3.07136	3.174787	3.294843	3.4
37	Oklahoma	3.91082	3.766853	4.070861	4.369755	4.374973	4.538878	4.497778	4.8

	Unnamed: 0	2019	2018	2017	2016	2015	2014	2013	2012
38	Oregon	3.191285	3.368639	3.381066	3.357058	3.422953	3.425235	3.619024	3.8
39	Pennsylvania	2.576553	2.608874	2.559596	2.639894	2.624487	2.662817	2.679782	2.8
40	Rhode Island	2.740331	2.913039	2.88872	2.841657	2.993473	2.84124	3.145949	3.2
41	South Carolina	2.587248	2.526491	2.617841	2.520198	2.77361	2.925619	3.18838	3.2
42	South Dakota	2.622479	2.567343	2.694138	2.784666	2.623275	2.782548	2.89983	3
43	Tennessee	3.48827	3.511221	3.546465	3.764136	3.694833	3.79717	4.132249	4.2
44	Texas	2.115818	2.582621	2.249599	2.6	2.589199	2.700045	2.889763	3
45	Utah	3.500982	3.774313	3.440224	3.587749	3.556838	3.078254	3.095966	3.3
46	Vermont	2.823768	3.081595	2.921478	3.098012	3.078068	3.460152	3.549144	3.5
47	Virginia	2.901522	3.07971	3.035766	3.373948	3.300969	3.481623	3.558905	3.7
48	Washington	2.844557	3.298481	3.406951	3.483672	3.383098	3.649351	3.821467	3.9
49	West Virginia	3.578947	3.254456	3.478798	3.7573	3.980201	4.1998	4.587166	4.7
50	Wisconsin	2.33064	2.457527	2.433102	2.597986	2.596625	2.656853	2.775866	2.9
51	Wyoming	3.799509	3.754303	3.961575	4.203238	4.149413	4.603246	4.259789	4.4

51 rows × 24 columns

```
In [45]: dataD = dataD.replace('---',np.nan)
dataD.fillna(dataD.median(numeric_only=True),inplace=True) #get only numeric
dataD
```

Out[45]:

	Unnamed: 0	2019	2018	2017	2016	2015	2014	2013	2012
1	Alabama	3.572984	3.714296	3.660498	3.778710	3.853690	3.785022	3.683290	3.660498
2	Alaska	3.638874	3.734551	3.580722	3.946655	4.138499	3.951233	4.461784	4.500000
3	Arizona	2.915899	2.954273	3.479199	3.352007	3.578466	3.892010	3.929150	4.300000
4	Arkansas	4.017159	4.066925	3.719694	3.926381	4.842516	4.845318	5.011872	5.000000
5	California	2.867509	2.951789	3.030063	3.168503	3.177654	3.324706	3.405376	3.500000
6	Colorado	3.291000	3.273600	3.186287	3.566256	3.725048	3.866602	4.073558	4.300000
7	Connecticut	2.690386	2.853612	2.870533	3.168503	3.128197	2.577657	2.761618	2.700000
8	Delaware	2.915491	2.800952	2.964845	3.079622	3.093239	3.256685	3.383206	3.500000
9	District of Columbia	2.422958	2.521158	2.477045	2.739404	2.832372	2.577050	2.838584	2.900000
10	Florida	3.489101	3.617673	3.612112	3.865675	3.956535	3.970182	4.110652	4.200000
11	Georgia	2.287278	2.502691	3.496277	3.168503	3.177654	3.324706	3.405376	3.500000
12	Hawaii	2.867509	2.951789	3.030063	3.168503	3.177654	3.324706	3.405376	3.500000
13	Idaho	3.857722	3.838199	3.884812	4.034127	4.125250	4.248488	4.495899	4.700000
14	Illinois	1.293579	1.506623	1.870798	2.200000	2.196579	2.213332	2.270509	2.400000
15	Indiana	2.867509	2.951789	3.030063	3.168503	3.177654	3.324706	3.405376	3.500000
16	Iowa	2.258904	2.232153	1.963944	1.254668	1.211307	1.505250	1.865768	2.200000
17	Kansas	2.328620	2.309802	2.379234	2.692543	2.821090	2.967954	2.972055	3.400000
18	Kentucky	3.433107	3.544220	3.726604	3.788167	3.702296	3.849137	4.162633	4.700000
19	Louisiana	2.399117	1.736703	1.923219	2.042863	2.772589	2.250479	2.170806	3.500000
20	Maine	3.033004	3.197091	3.226272	3.428518	3.428048	3.581715	3.969730	3.900000
21	Maryland	2.661901	2.400575	2.542721	2.686968	2.595564	2.541828	2.533559	2.800000
22	Massachusetts	1.463184	2.116008	2.165801	2.290151	2.552977	2.680787	2.626694	2.700000
23	Michigan	2.328661	2.844962	2.848937	2.908151	3.045681	2.997817	3.259926	3.300000
24	Minnesota	2.867509	2.951789	3.030063	3.168503	3.177654	3.324706	3.405376	3.500000
25	Mississippi	2.625541	2.717870	2.900372	3.204710	3.370280	3.371321	3.587181	4.000000
26	Missouri	2.890462	3.040912	3.138448	3.281963	3.177357	3.278092	3.422140	3.700000
27	Montana	2.997816	3.033027	3.109968	3.104976	3.425145	3.447707	3.388612	3.900000
28	Nebraska	2.663864	2.949305	3.024359	3.136149	3.177952	3.054473	3.278538	3.400000
29	Nevada	4.170243	4.376824	4.459582	4.294473	4.636707	5.330212	5.121614	5.500000
30	New Hampshire	3.072712	3.089664	3.112910	3.375050	3.328554	3.543077	3.726598	3.600000
31	New Jersey	2.519086	2.746809	2.600813	2.668465	2.817924	2.778755	2.785263	2.800000
32	New Mexico	2.867509	2.951789	3.030063	3.168503	3.288078	3.551544	3.437417	3.600000
33	New York	2.892530	2.780085	2.676303	2.681045	2.707495	2.843986	2.735467	2.900000
34	North Carolina	3.133270	3.075324	3.109968	3.248319	3.082307	3.401963	3.429203	3.700000
35	North Dakota	2.481425	2.557636	2.486123	2.589874	2.839111	2.830360	2.943075	3.700000
36	Ohio	2.768733	2.905956	2.922647	3.029867	3.071360	3.174787	3.294843	3.400000
37	Oklahoma	3.910820	3.766853	4.070861	4.369755	4.374973	4.538878	4.497778	4.800000

	Unnamed: 0	2019	2018	2017	2016	2015	2014	2013	2012
38	Oregon	3.191285	3.368639	3.381066	3.357058	3.422953	3.425235	3.619024	3.8
39	Pennsylvania	2.576553	2.608874	2.559596	2.639894	2.624487	2.662817	2.679782	2.8
40	Rhode Island	2.740331	2.913039	2.888720	2.841657	2.993473	2.841240	3.145949	3.2
41	South Carolina	2.587248	2.526491	2.617841	2.520198	2.773610	2.925619	3.188380	3.2
42	South Dakota	2.622479	2.567343	2.694138	2.784666	2.623275	2.782548	2.899830	3.0
43	Tennessee	3.488270	3.511221	3.546465	3.764136	3.694833	3.797170	4.132249	4.2
44	Texas	2.115818	2.582621	2.249599	2.600000	2.589199	2.700045	2.889763	3.0
45	Utah	3.500982	3.774313	3.440224	3.587749	3.556838	3.078254	3.095966	3.3
46	Vermont	2.823768	3.081595	2.921478	3.098012	3.078068	3.460152	3.549144	3.8
47	Virginia	2.901522	3.079710	3.035766	3.373948	3.300969	3.481623	3.558905	3.7
48	Washington	2.844557	3.298481	3.406951	3.483672	3.383098	3.649351	3.821467	3.9
49	West Virginia	3.578947	3.254456	3.478798	3.757300	3.980201	4.199800	4.587166	4.7
50	Wisconsin	2.330640	2.457527	2.433102	2.597986	2.596625	2.656853	2.775866	2.9
51	Wyoming	3.799509	3.754303	3.961575	4.203238	4.149413	4.603246	4.259789	4.4

51 rows × 24 columns

```
In [46]: yearD = dataD.columns[1:].to_numpy().astype(float)
         stateD = dataD.iloc[:,0]
         D = dataD.iloc[:,1:].to_numpy()
         D
```

```
Out[46]: array([[3.57298368, 3.71429606, 3.66049766, ..., 5.7, 6., ...],
               [6.1, ...],
               [3.63887389, 3.73455124, 3.58072169, ..., 5., 5., ...],
               [5.5, ...],
               [2.9158985, 2.95427298, 3.47919906, ..., 4.6, 6.2, ...],
               [6.9, ...],
               ...,
               [3.57894749, 3.25445556, 3.47879817, ..., 4.9, 5.2, ...],
               [5.3, ...],
               [2.33064042, 2.45752694, 2.43310178, ..., 3.2, 3.4, ...],
               [3.6, ...],
               [3.79950895, 3.75430343, 3.96157531, ..., 5.7, 6.6, ...],
               [6.6, ...]])
```

In [47]: *#Make Some Plots*

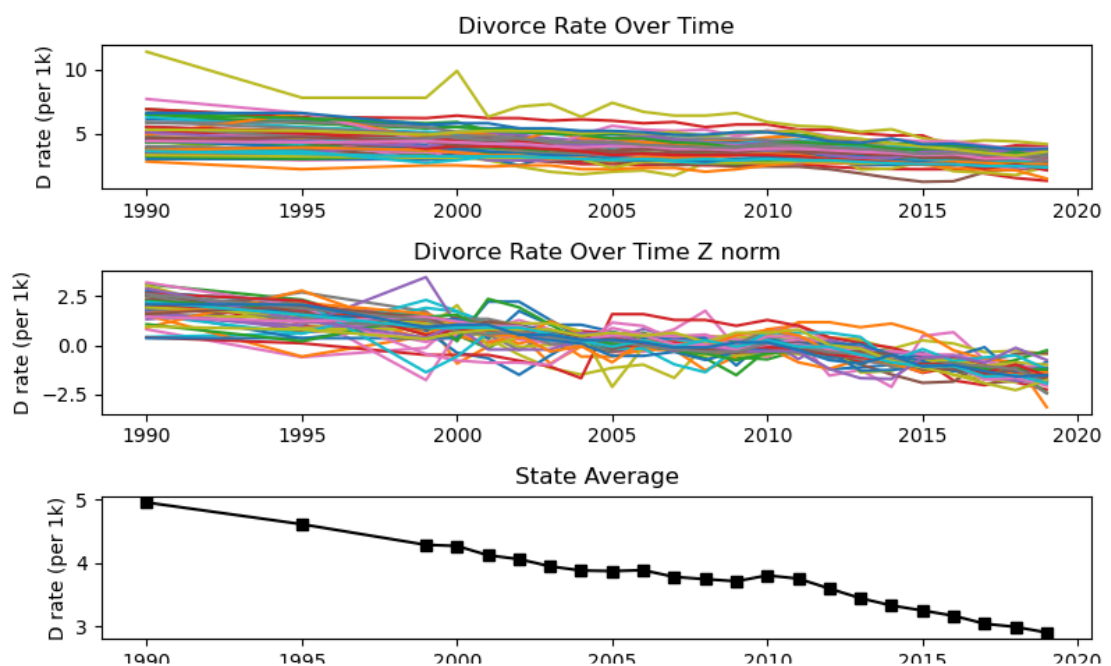
```
fig,ax = plt.subplots(3,1,figsize=(8,5))

ax[0].plot(yearD,D.T)
ax[0].set_title('Divorce Rate Over Time')
ax[0].set_ylabel('D rate (per 1k)')

ax[1].plot(yearD,stats.zscore(D.T))
ax[1].set_title('Divorce Rate Over Time Z norm')
ax[1].set_ylabel('D rate (per 1k)')

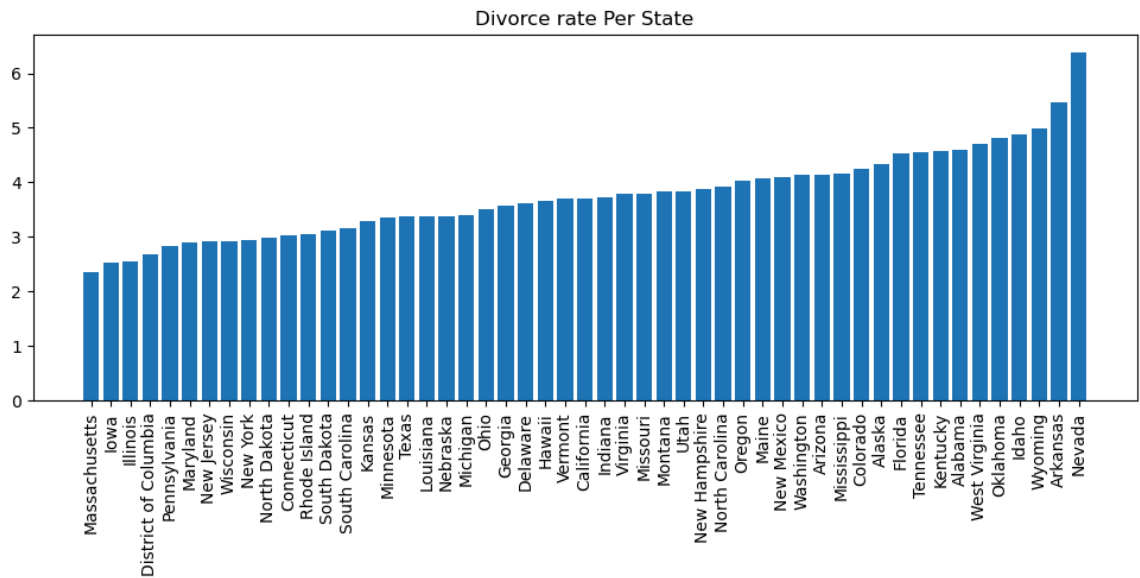
ax[2].plot(yearD,np.mean(D,axis=0),'ks-')
ax[2].set_title('State Average')
ax[2].set_ylabel('D rate (per 1k)')

plt.tight_layout()
plt.show()
```



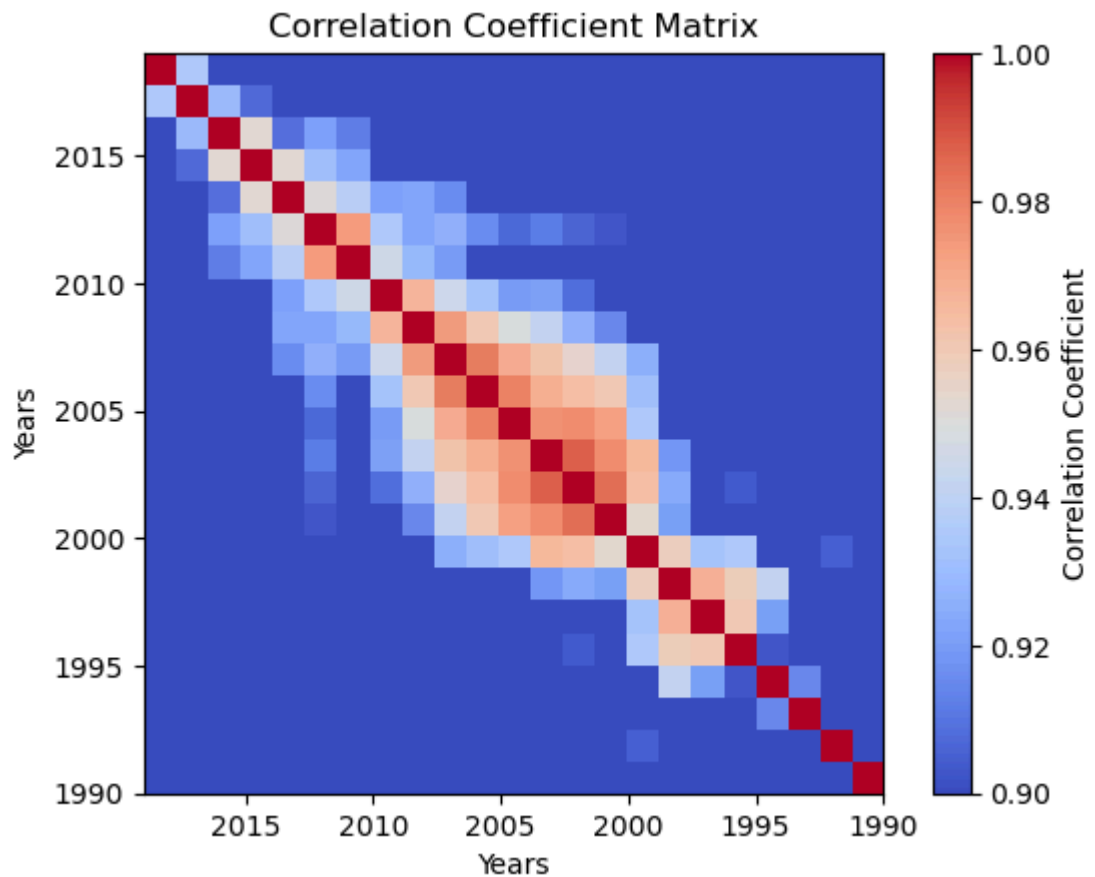
```
In [48]: meanD = np.mean(D,axis=1)
sinD = np.argsort(meanD)

fig=plt.figure(figsize=(12,4))
plt.bar(stateD.iloc[sinD],meanD[sinD])
plt.xticks(rotation=90)
plt.title('Divorce rate Per State')
plt.show()
```

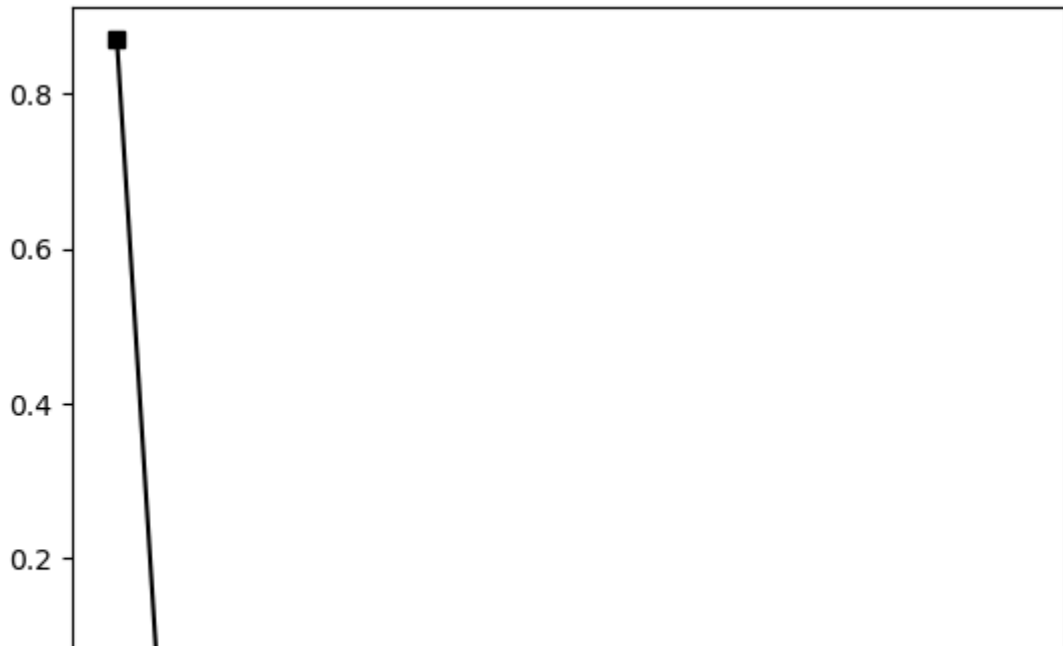


```
In [49]: #Correlation Matrix
corr_matrix = np.corrcoef(D.T)

# Plotting the correlation matrix
plt.imshow(corr_matrix, vmin=0.9, vmax=1, extent=[yearD[0], yearD[-1], yearD[0], yearD[-1]])
plt.colorbar(label='Correlation Coefficient')
plt.xlabel('Years')
plt.ylabel('Years')
plt.title('Correlation Coefficient Matrix')
plt.show()
```




```
In [50]: # PCA
pca = PCA().fit(D)
plt.plot(pca.explained_variance_ratio_, 'ks-')
plt.show()
```



Comparison

```
In [51]: #Years
print(np.sum(yearD-yearM))

0.0
```

```
In [52]: print(stateM.equals(stateD))

False
```

```
In [54]: statenames = pd.concat([stateM, stateD], axis=1)
print(statenames)
```

	Unnamed: 0	Unnamed: 0
1	Alabama	Alabama
2	Alaska	Alaska
3	Arizona	Arizona
4	Arkansas	Arkansas
5	California 1	California
6	Colorado	Colorado
7	Connecticut	Connecticut
8	Delaware	Delaware
9	District of Columbia	District of Columbia
10	Florida	Florida
11	Georgia	Georgia
12	Hawaii	Hawaii
13	Idaho	Idaho
14	Illinois	Illinois
15	Indiana	Indiana
16	Iowa	Iowa
17	Kansas	Kansas
18	Kentucky	Kentucky
19	Louisiana	Louisiana
20	Maine	Maine
21	Maryland	Maryland
22	Massachusetts	Massachusetts
23	Michigan	Michigan
24	Minnesota	Minnesota
25	Mississippi	Mississippi
26	Missouri	Missouri
27	Montana	Montana
28	Nebraska	Nebraska
29	Nevada	Nevada
30	New Hampshire	New Hampshire
31	New Jersey	New Jersey
32	New Mexico	New Mexico
33	New York	New York
34	North Carolina	North Carolina
35	North Dakota	North Dakota
36	Ohio	Ohio
37	Oklahoma	Oklahoma
38	Oregon	Oregon
39	Pennsylvania	Pennsylvania
40	Rhode Island	Rhode Island
41	South Carolina	South Carolina
42	South Dakota	South Dakota
43	Tennessee	Tennessee
44	Texas	Texas
45	Utah	Utah
46	Vermont	Vermont
47	Virginia	Virginia
48	Washington	Washington
49	West Virginia	West Virginia
50	Wisconsin	Wisconsin
51	Wyoming	Wyoming

In [56]: *#find difference*

```
np.where(statenames.iloc[:,0] != statenames.iloc[:,1])
```

Out[56]: (array([4], dtype=int64),)

It just a type error. Do not worry about that

Inferencial Statistics

```
In [78]: #correlate M and D over time per state

#Bonferroni Correlated Threshold

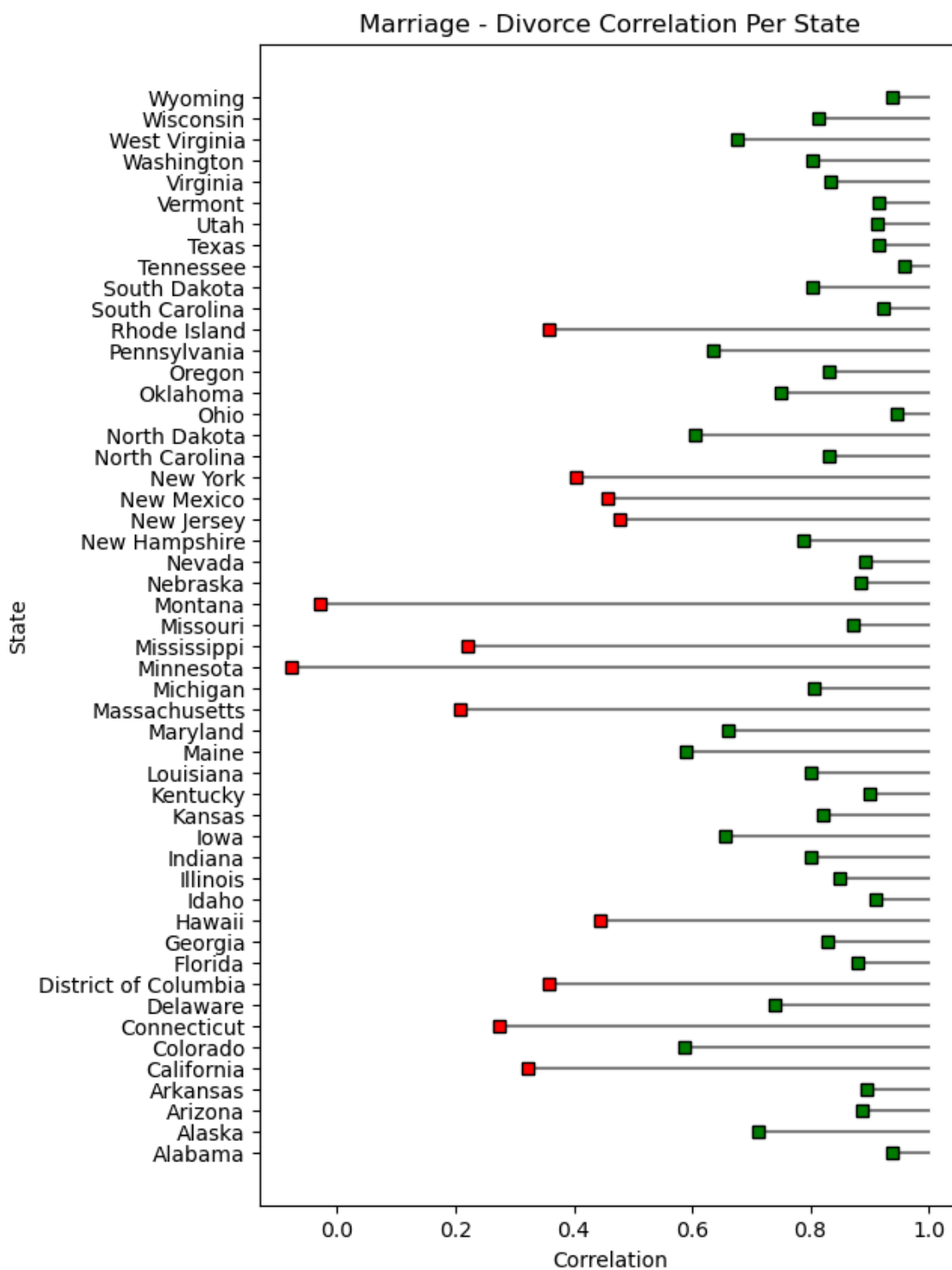
pvalT = 0.5/51

fig = plt.figure(figsize=(6,10))
color = 'rg'

for si in range(len(stateM)):
    #Compute Correlation
    r,p = stats.pearsonr(M[si,:],D[si,:])

    #plot
    plt.plot([r,1],[si,si], '-',color=[.5,.5,.5])
    plt.plot(r,si, 'ks',markerfacecolor=color[bool(p<pvalT)])

plt.ylabel('State')
plt.xlabel('Correlation')
plt.title('Marriage - Divorce Correlation Per State')
plt.yticks(range(len(stateM)),labels=stateD)
plt.show()
```



By Lahiru Sadakelum