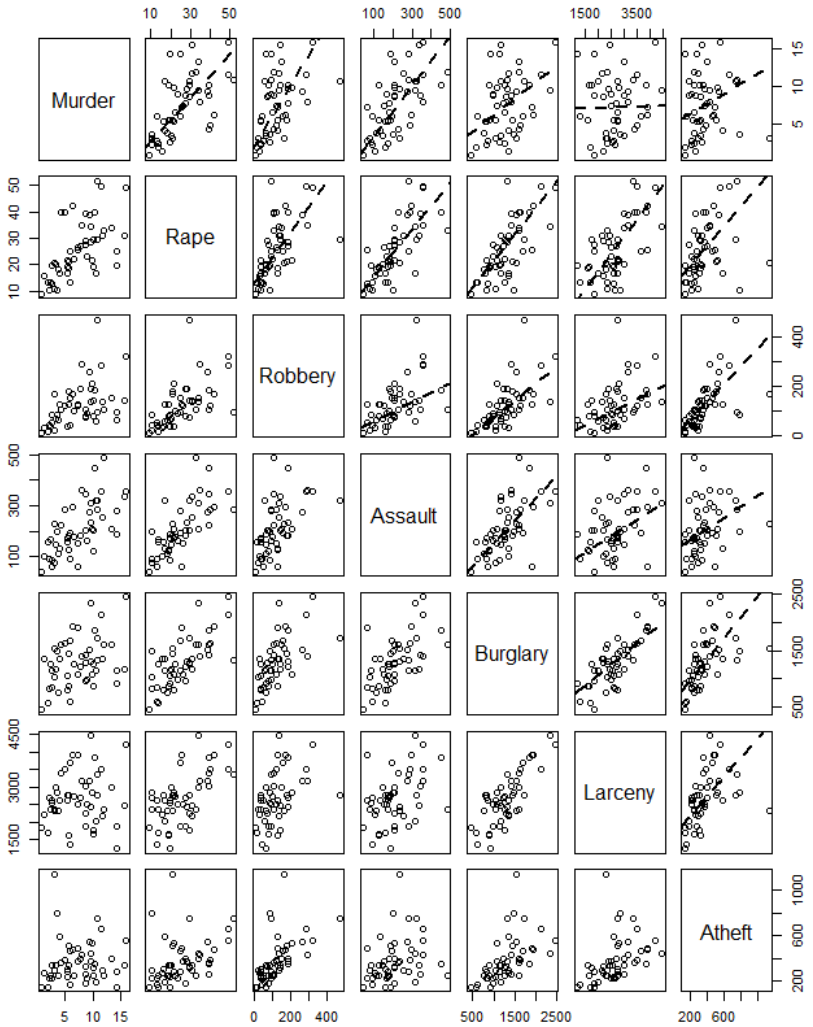
Kyle Beitz

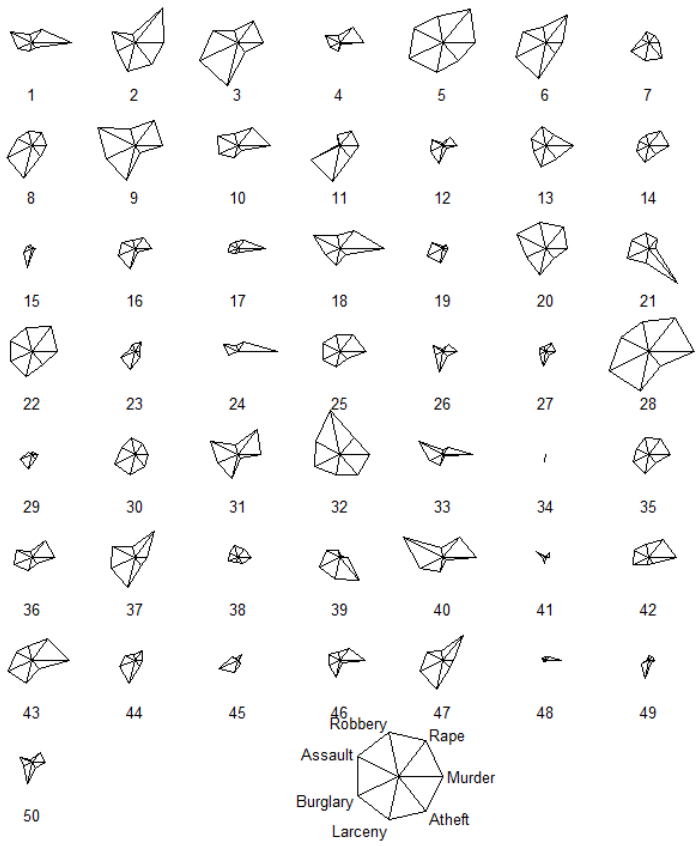
Stats 4300

Homework #1

- 1.1:
  a) After reviewing the data and scatterplot (below), Rape and Assault appear to have the strongest linear relationship because the data appears to follow the linear regression line the best.

	Murder	Rape	Robbery	Assault	Burglary	Larceny	Atheft
Murder	1.00	0.60	0.48	0.65	0.39	0.10	0.07
Rape	0.60	1.00	0.59	0.74	0.71	0.61	0.35
Robbery	0.48	0.59	1.00	0.56	0.64	0.45	0.59
Assault	0.65	0.74	0.56	1.00	0.62	0.40	0.28
Burglary	0.39	0.71	0.64	0.62	1.00	0.79	0.56
Larceny	0.10	0.61	0.45	0.40	0.79	1.00	0.44
Atheft	0.07	0.35	0.59	0.28	0.56	0.44	1.00





b) After reviewing the star plot, variables 15,19,21,29,33,34,41,48 and 49 seem to be different from the others. They are either very small or very distorted in certain directions of the star.

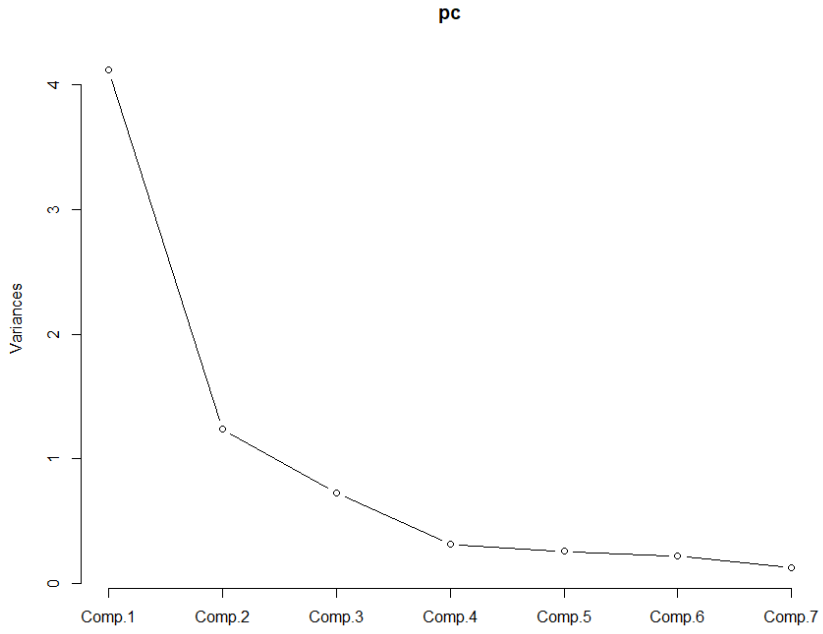
- 1.2:

a) In the image below you can see the eigenvalues, proportion of Variance and loadings. Researchers might believe that component 1 represents overall crime rate because each of the variances are closely related and are all positive. Component 2 could be interpreted to represent property crime versus violent crime because robbery, burglary, larceny and auto theft are all close to or below zero.

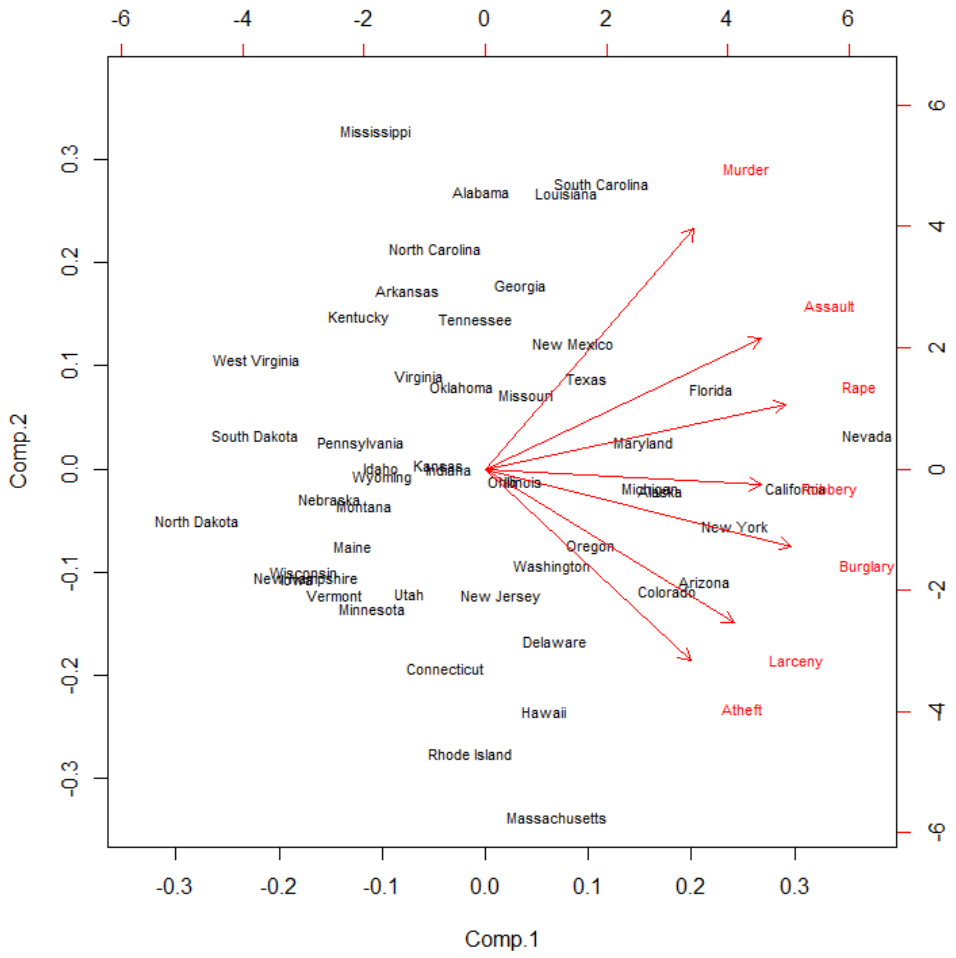
```
Importance of components:
                          Comp.1    Comp.2    Comp.3     Comp.4     Comp.5
Standard deviation     2.0285363 1.1129788 0.8519487 0.56252293 0.50791186
Proportion of Variance 0.5878514 0.1769603 0.1036881 0.04520458 0.03685349
Cumulative Proportion  0.5878514 0.7648116 0.8684997 0.91370429 0.95055778
                           Comp.6     Comp.7
Standard deviation     0.47121064 0.35221592
Proportion of Variance 0.03171992 0.01772229
Cumulative Proportion  0.98227771 1.00000000

Loadings:
         Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7
Murder    0.300  0.629  0.178  0.232  0.538  0.259  0.268
Rape      0.432  0.169 -0.244         0.188 -0.773 -0.296
Robbery   0.397  [redacted]  0.496  0.558 -0.520 -0.114
Assault   0.397  0.344        -0.630 -0.507  0.172  0.192
Burglary  0.440 -0.203 -0.210         0.101  0.536 -0.648
Larceny   0.357 -0.402 -0.539  0.235                0.602
Atheft    0.295 -0.502  0.568 -0.419  0.370         0.147
```

b) From the image above we can see that components 1 and 2 both have eigenvalues greater than one. For the proportion of variance, we can determine that the first two components account for almost 76% of the variance in the original variables. In the skree diagram below we can see that the first two components also have variances higher than one, which means we would want to keep those two principal components. Overall, we can determine that we can use components 1 and 2 without little loss of information.



c) From the image below we can see that South Dakota and Wyoming have relatively low crime rates while Colorado appears to have high auto theft and larceny crime rates.


Mississippi
Murder
Alabama
South Carolina
Louisiana
North Carolina
Arkansas
Georgia
Assault
Kentucky
Tennessee
New Mexico
West Virginia
Virginia
Texas
Oklahoma
Missouri
Florida
Rape
South Dakota
Pennsylvania
Maryland
Nevada
Idaho
Kansas
Wyoming
Indiana
Ohio
Illinois
Michigan
Alaska
California
Robbery
Nebraska
Montana
North Dakota
New York
Maine
Oregon
Burglary
Wisconsin
Washington
Arizona
Vermont
Utah
New Jersey
Colorado
Minnesota
Delaware
Larceny
Connecticut
Hawaii
Atheft
Rhode Island
Massachusetts
Comp.2
Comp.1
-6
-4
-2
0
2
4
6
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3