

## 6. Assignment NMST 539

Frantisek Helebrand

April 2022

### 1 Correlation between crimes

In this assignment we will be analyzing the data from the US, specifically, we will be focusing on the crime data from different US states in 1985. We have observations from 50 different US states with 11 variables. We will use only 7 of them (number of murders, rapes, robberies, assaults, burglaries, larcenies and auto thefts) for principal analysis. Other variables in the data set are not useful for our analysis (land area, population and division do not have an impact on crimes).

The correlation between crimes can be seen in the figure 1 and 2.

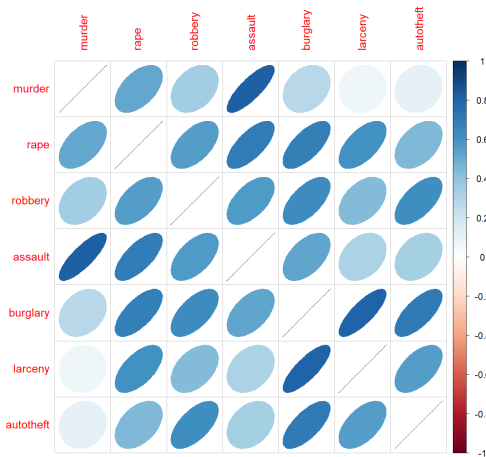


Figure 1: Correlation between crimes.

We can observe that all variables are positively correlated. The largest correlation can be observed between assault and murder and also between larceny and burglary. All correlation makes sense.

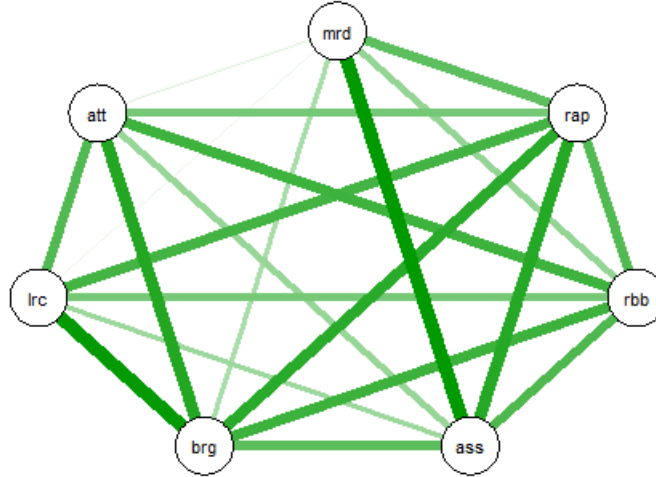


Figure 2: Correlation between crimes.

## 2 Principal component analysis

By using principal components analysis we obtain the following table which shows standard deviation, proportion of variability and cumulative proportion of variability.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
St. dev.	2.0191	1.1965	0.7945	0.5832	0.4984	0.3737	0.3636
Prop. of Var.	0.5824	0.2045	0.0902	0.0486	0.0355	0.0199	0.0189
Cumul. Prop.	0.5824	0.7869	0.8771	0.9257	0.9612	0.9811	1.0000

We can see that for an explanation of 90,2 % of variability it is enough to use only the first 4 components. If we would like to explain 95 % of variability we will have to use the first 5 components. The second power of standard deviation can be seen in the figure 3

From the figures 4 and 5 we can see clusters for data. We have used the first three components. It can be said that in the South prevail assaults and murders, in the west, there are more lancers and burglaries and in the Northeast prevail robberies and autothefts.

Overall we can say that the South region is dangerous in comparison with others.

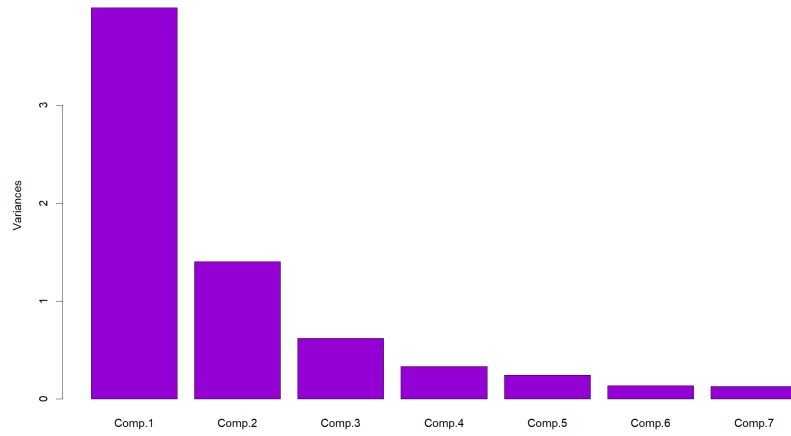


Figure 3: The second power of standard deviation.

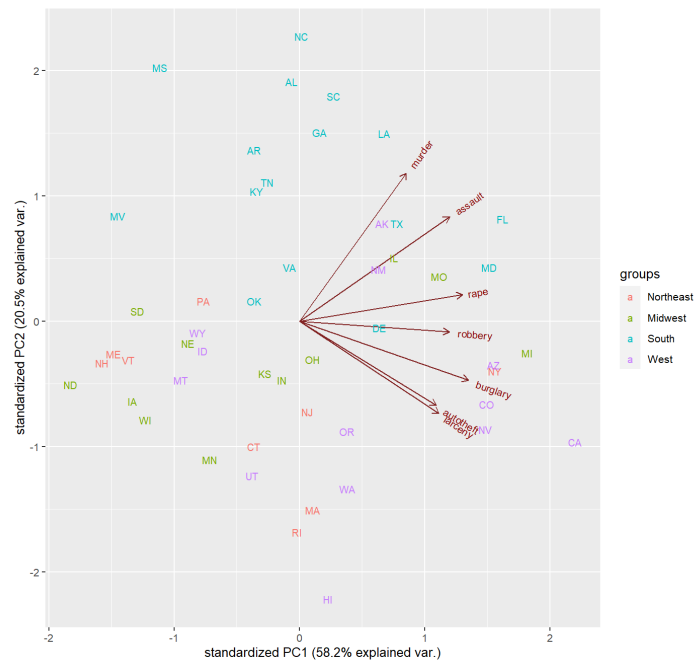


Figure 4: PCA - first and second component.

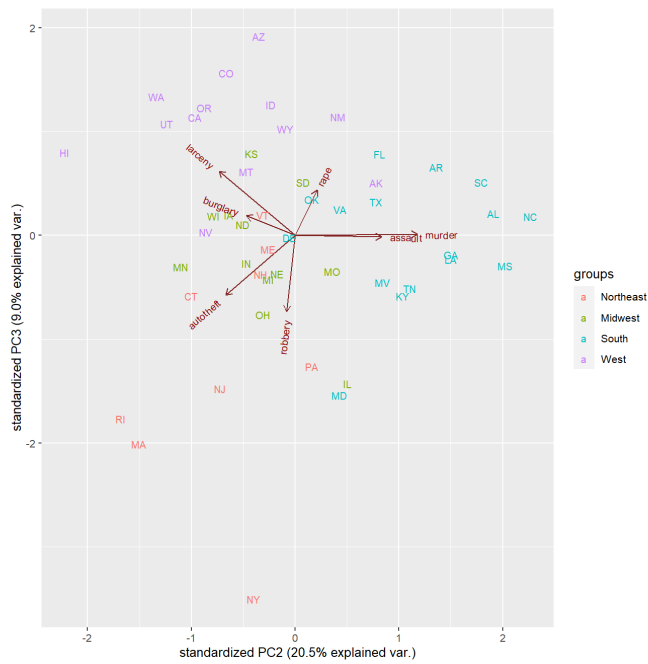


Figure 5: PCA - third and second component.