Predicting Characteristics of Mass Shooting

Deekshitha Balaji

ID:98701275

CIS 481 Computational Learning: Theory and Practice (Winter 2019)

Instructor: Luis E. Ortiz

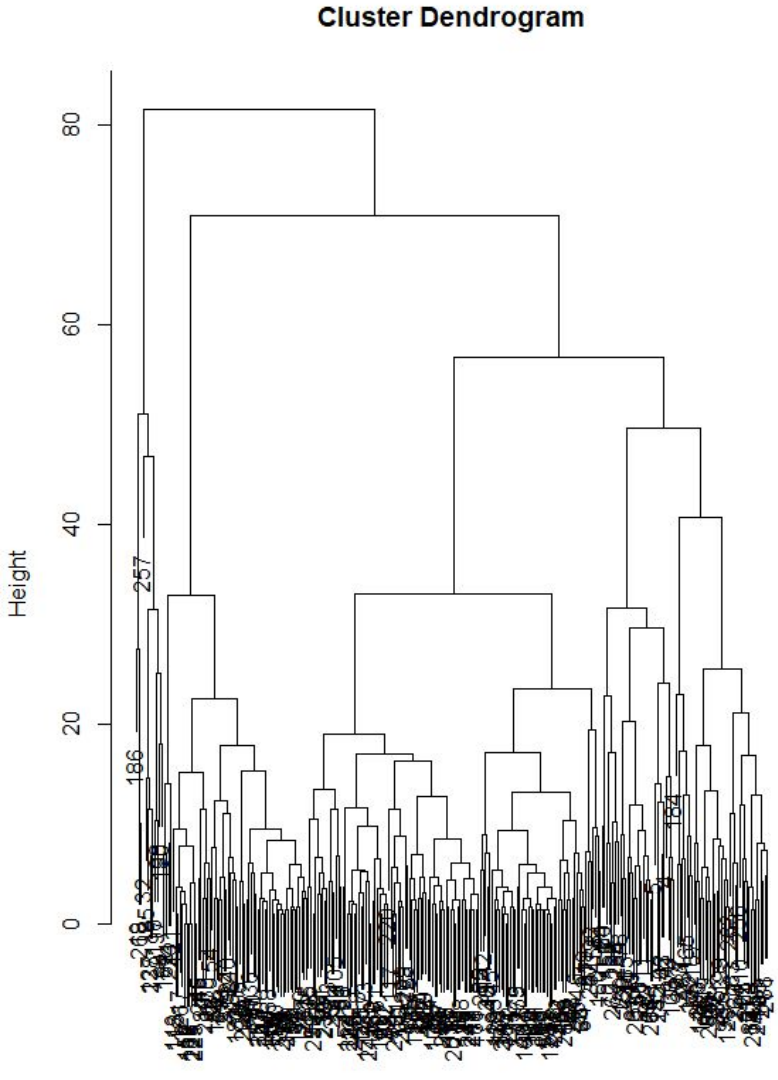
April 24, 2020

Project Summary

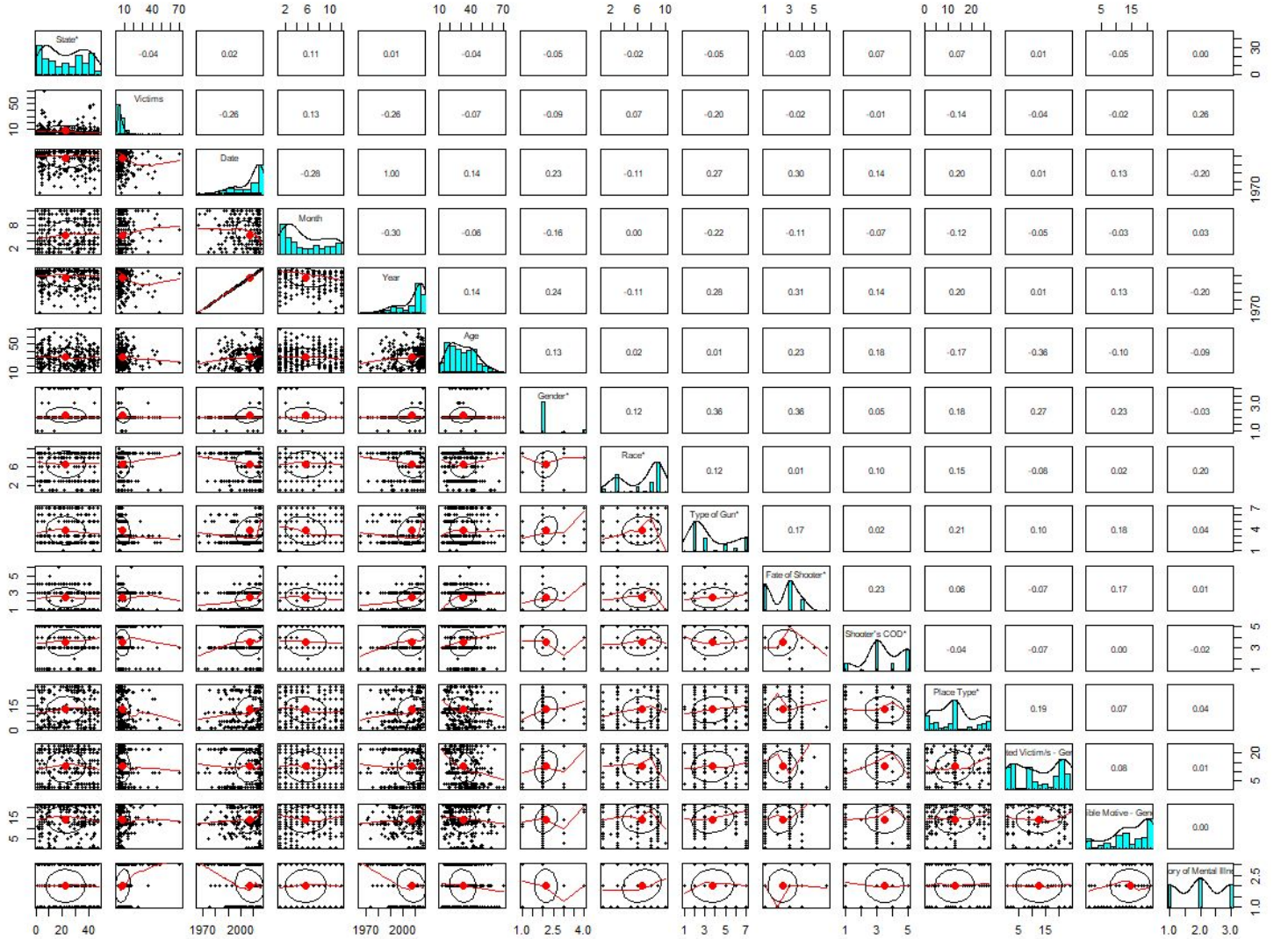
Problem: Can common characteristics of mass shootings in the future be accurately predicted using the data presented?

Dataset: Data retrieved from Stanford University's database.

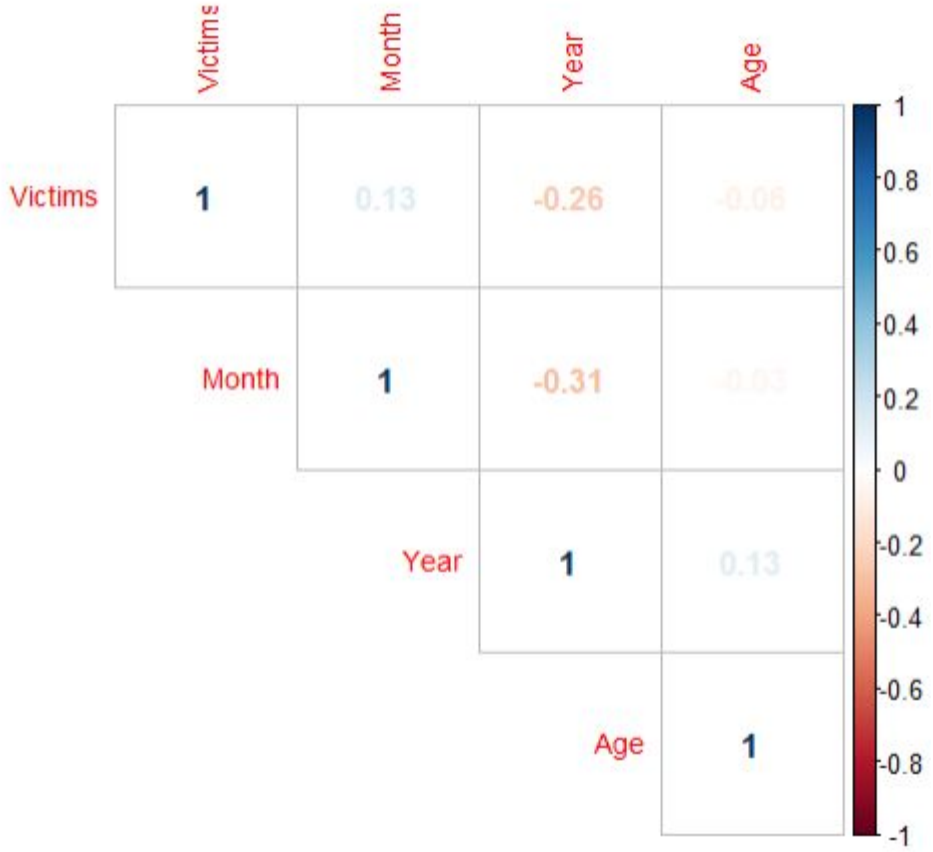
Idea: Be able to evaluate cluster analysis and linear regression models to predict the characteristics of future shooters and shootings

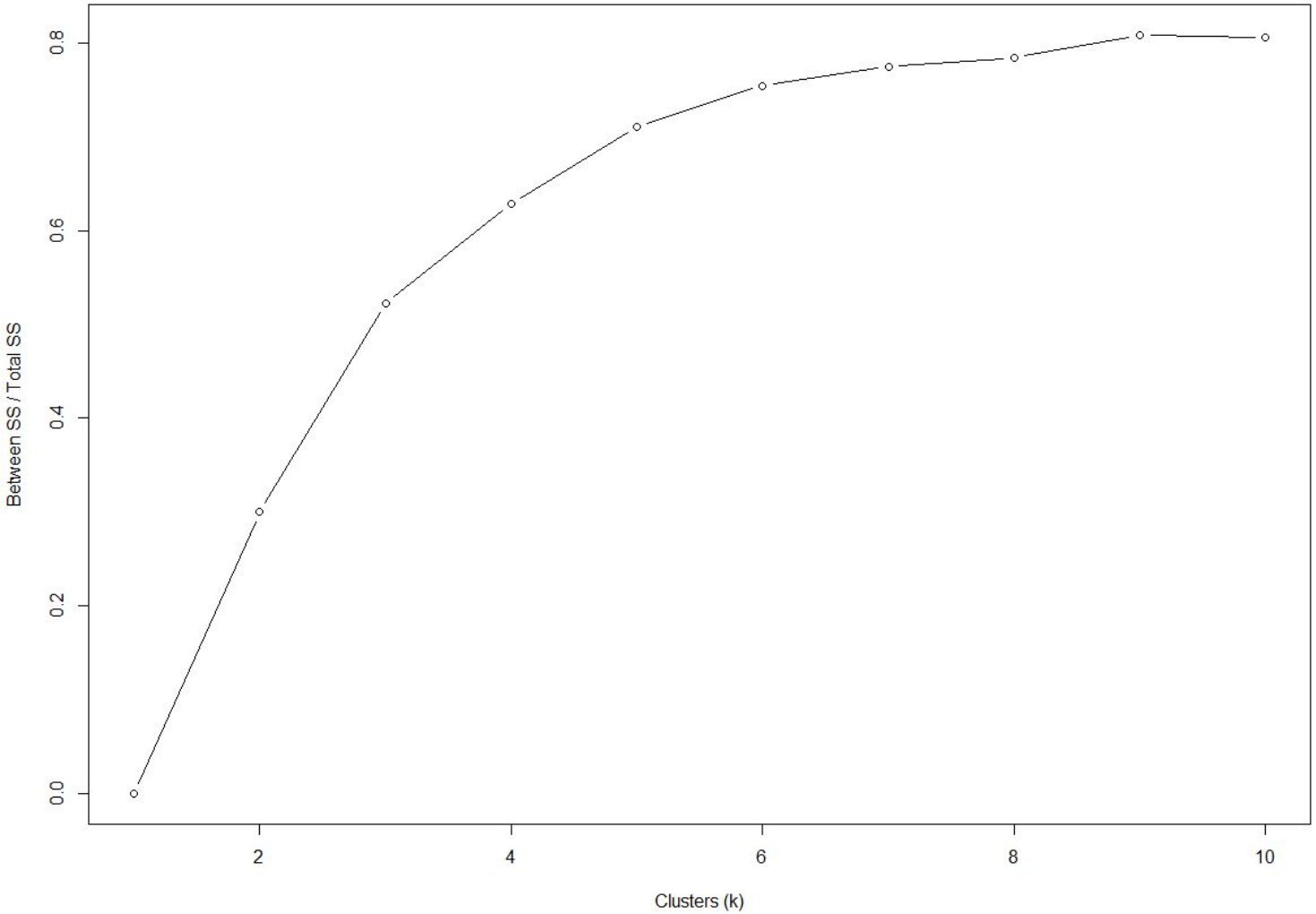

Cluster Dendrogram
Height
0
20
40
60
80
257
186

Correlation of all Variables

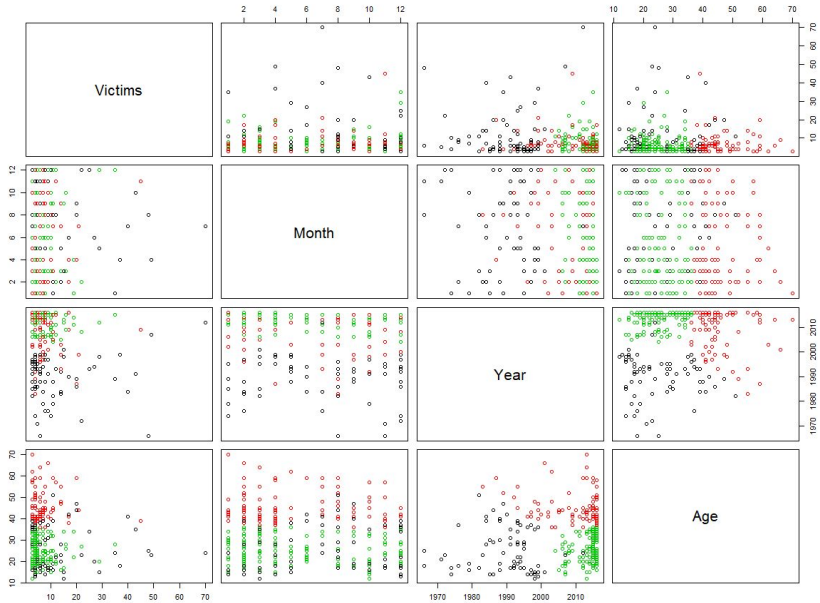


Correlation of Continuous Variables

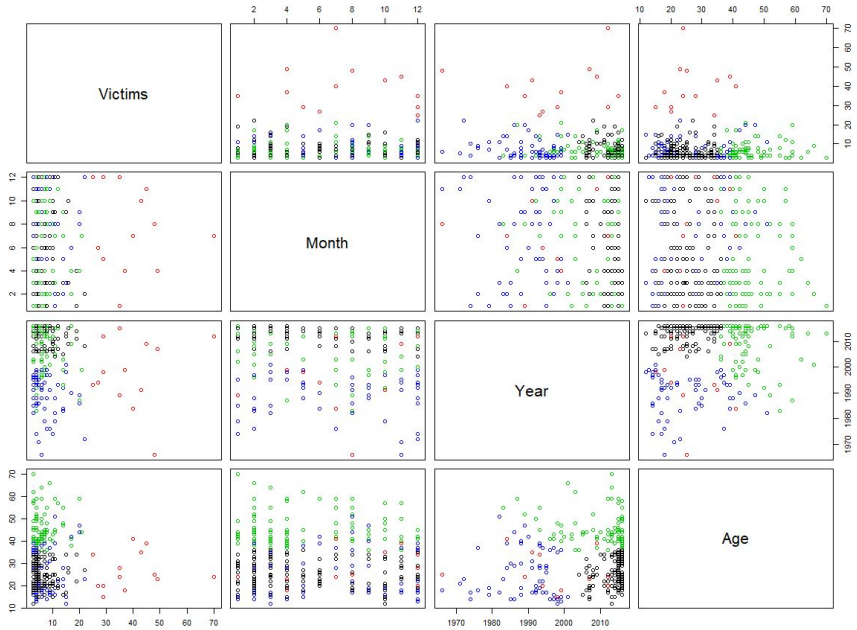



Between SS / Total SS
0.0
0.2
0.4
0.6
0.8
2
4
6
8
10
Clusters (k)

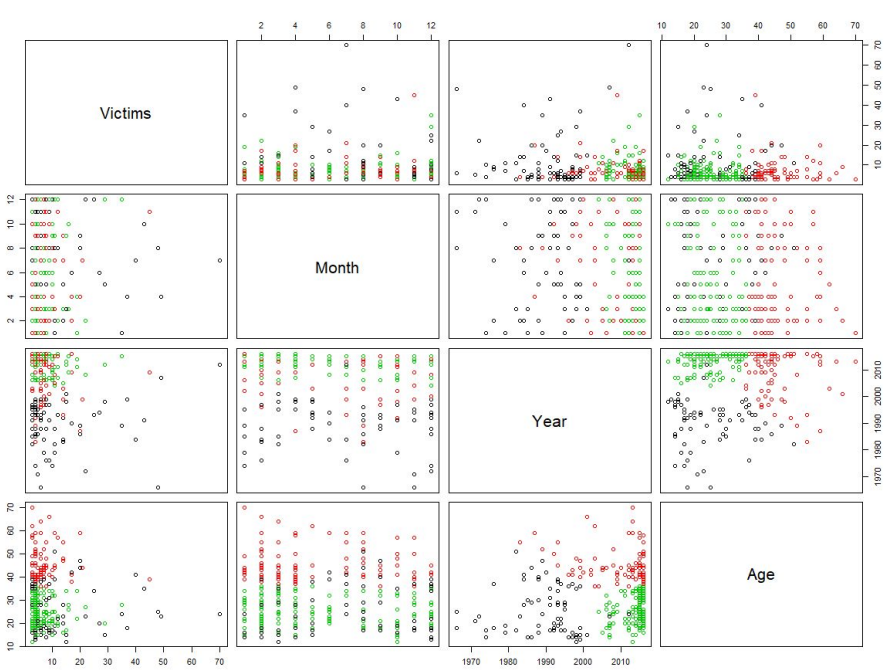
3 clusters

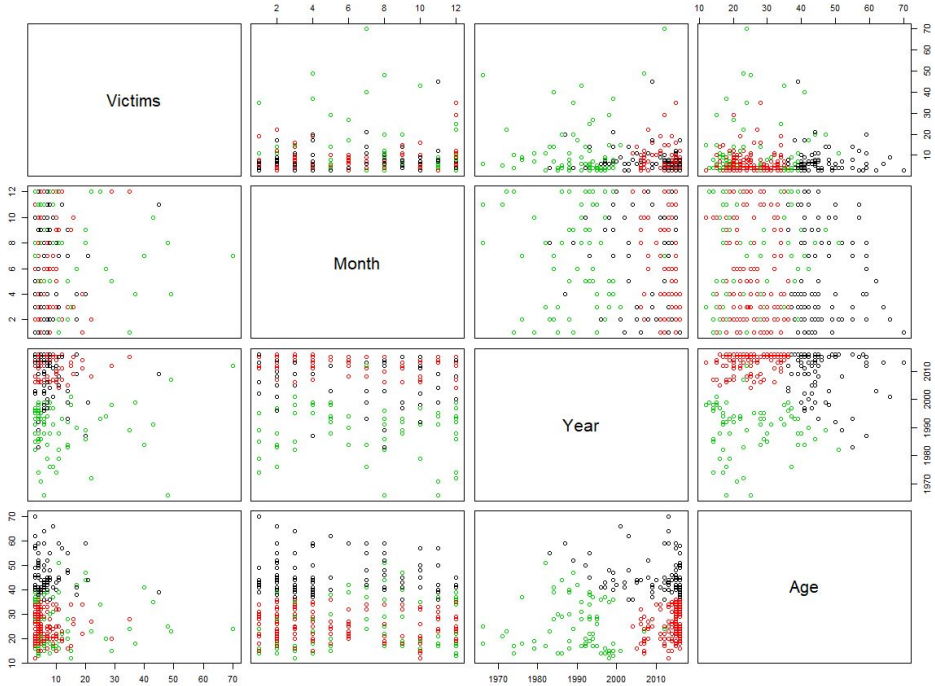


4 clusters



Test with 3 clusters



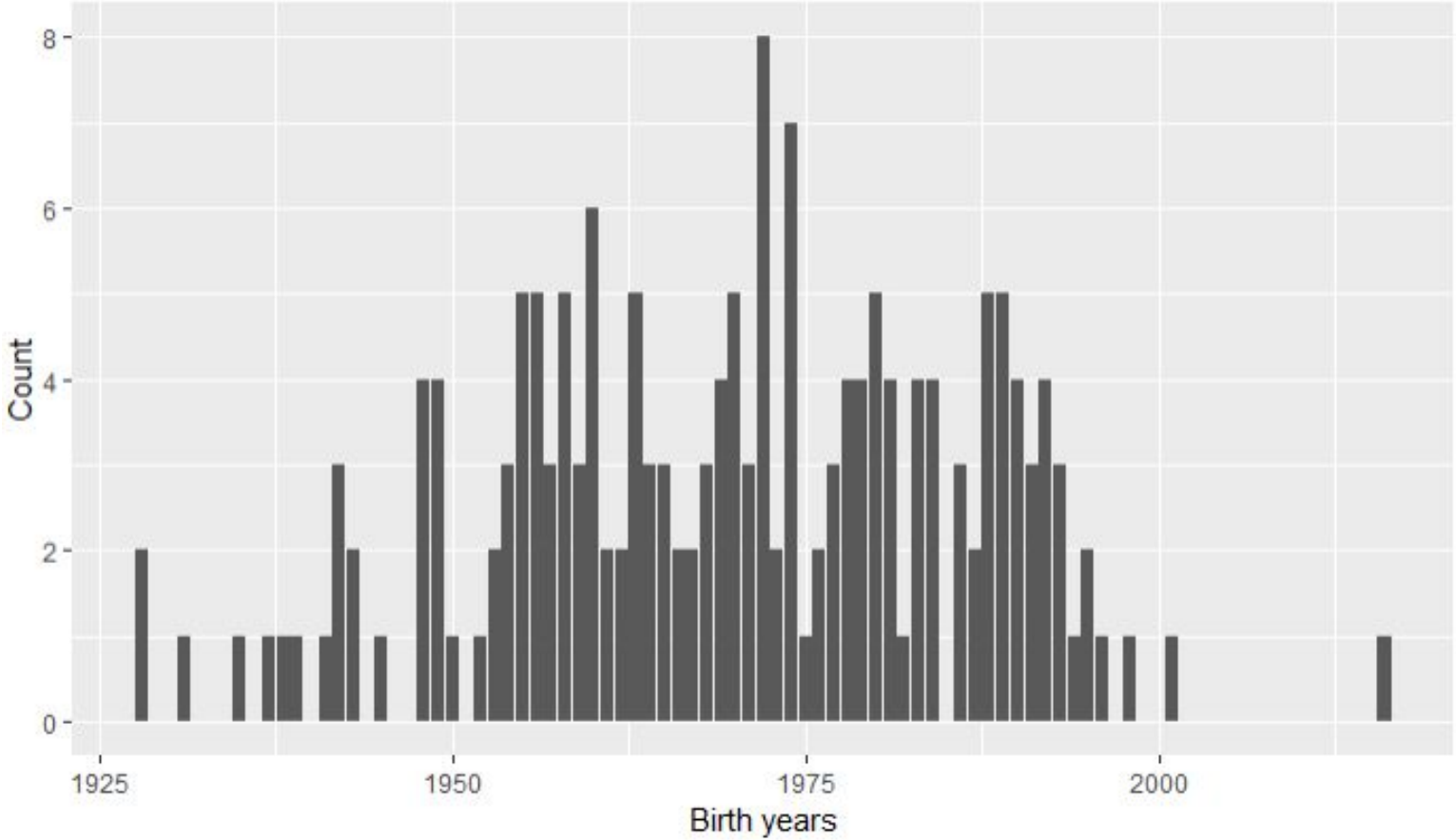


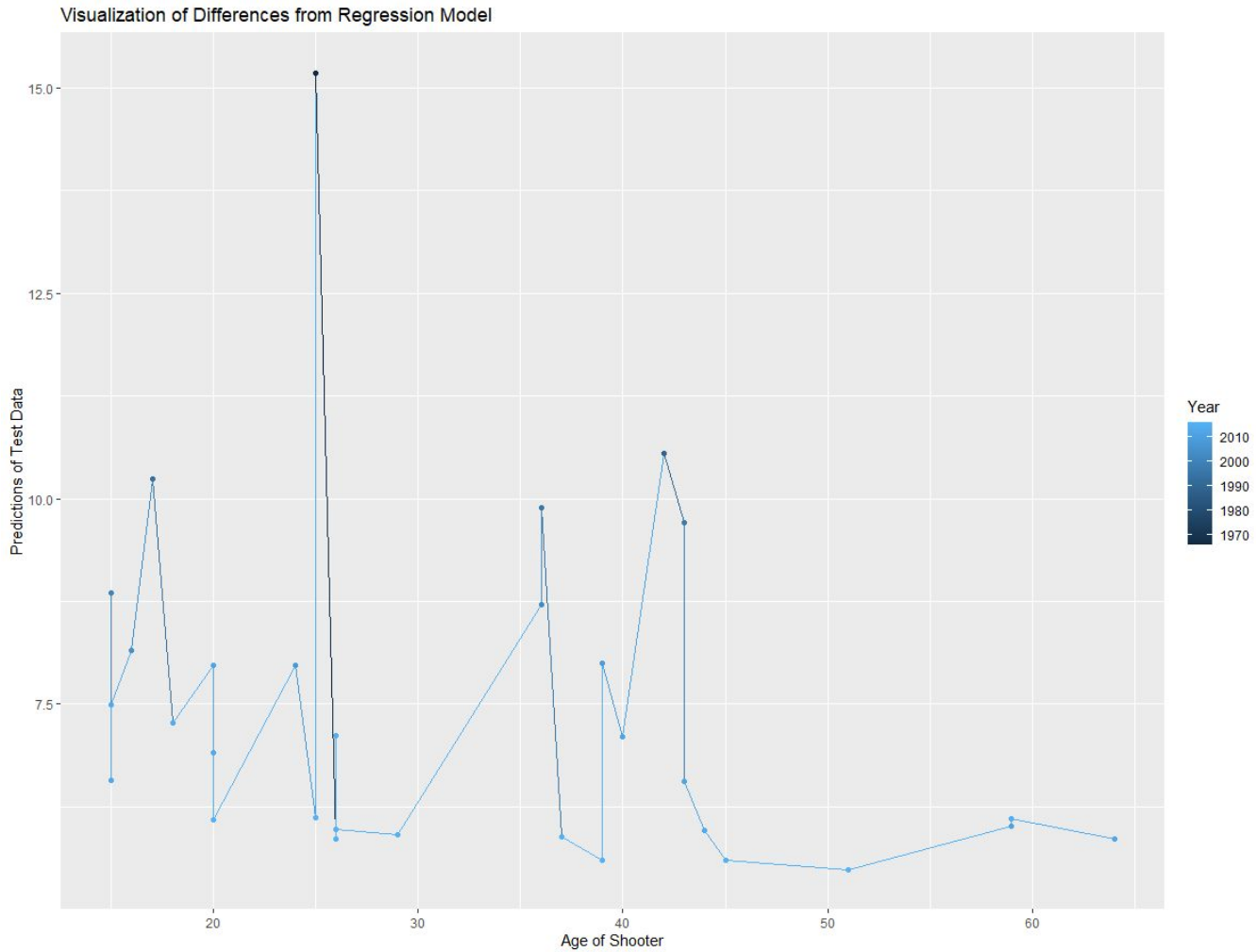
cluster 1			
victims	month	year	age
22	2	2008	27
9	11	1991	31
3	2	2015	19
3	2	2015	27
3	2	2015	18
3	3	2015	45
3	1	2013	70
4	11	1994	37
3	12	1993	39
12	2	2011	20
3	**2**	**2015**	**27**

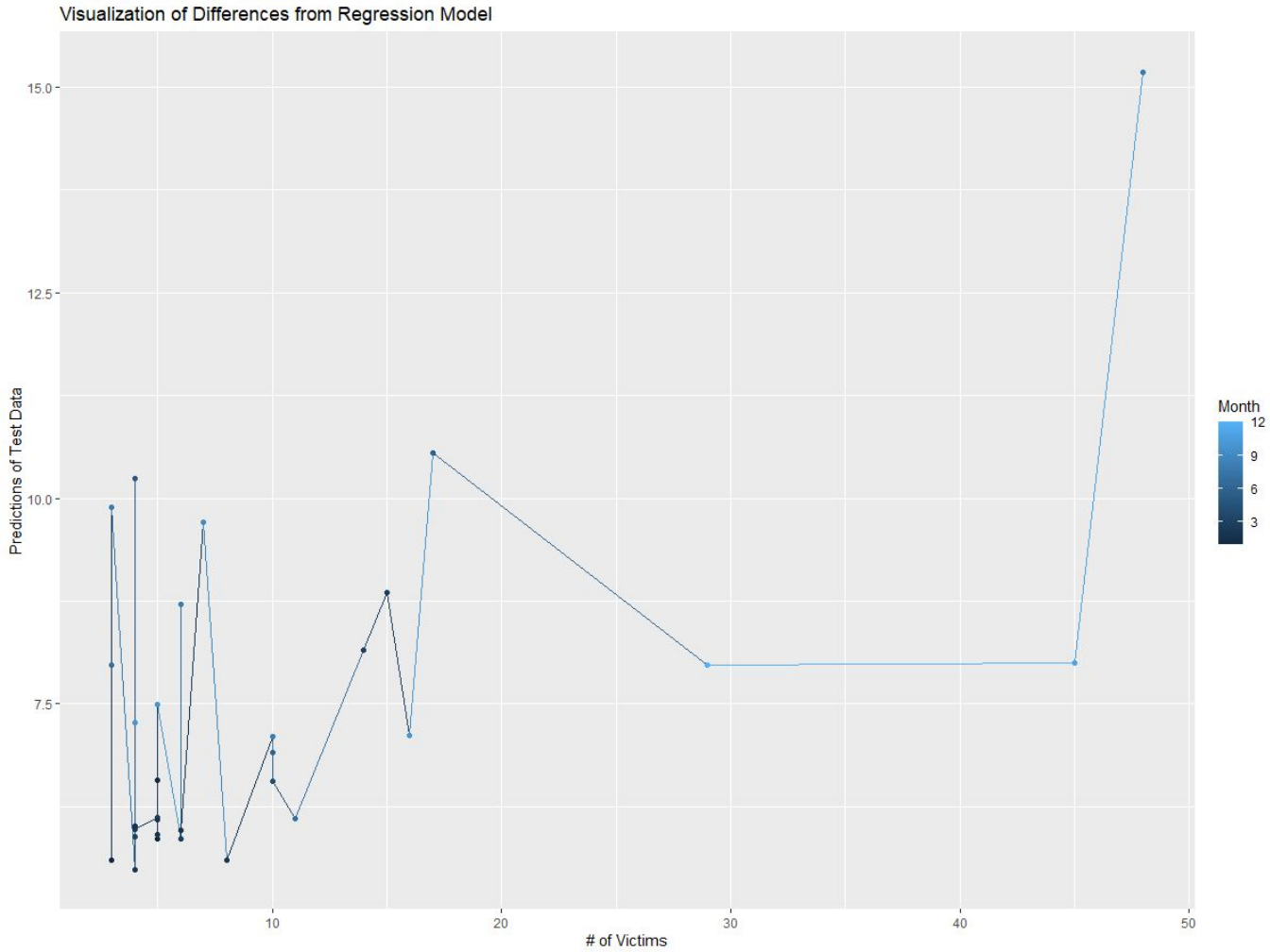
cluster 2			
victims	month	year	age
10	8	2010	34
4	8	1989	52
30	2	2016	51
30	7	2015	32
29	9	1993	29
4	2	1996	14
6	5	2012	40
7	10	2007	20
49	4	2007	23
3	10	2015	33
29	11	1998	15
6	10	1998	48
7	9	2006	44
30	3	2015	22
4	8	1983	55
30	2	2015	30
7	3	2015	22
30	**8**	**2015**	**22**

cluster 3			
victims	month	year	age
4	10	1992	50
4	4	2015	23
4	2	2016	26
4	4	2015	25
11	12	2004	23
4	**4**	**2015**	**23**

Figure 1. Distribution of birth years of the shooters




Visualization of Differences from Regression Model
Predictions of Test Data
15.0
12.5
10.0
7.5
20
30
40
50
60
Age of Shooter
Year
2010
2000
1990
1980
1970


Visualization of Differences from Regression Model
Predictions of Test Data
15.0
12.5
10.0
7.5
10
20
30
40
50
# of Victims
Month
12
9
6
3