An analysis on violence enhancing factors

Presentation & Analysis by: Elias Kalach Salame, Marco Buratti, Marec Reifert and Roberto Semeraro

10.11.20

Agenda

1. Introduction to our analysis and recap

1.1. Questions & addressed Stakeholders

- How did violence develop over the years?
- Does the impact of the analyzed factors differ between states?
- Have the violence rates been affected by a change in gun law?
- Is Violence affected by any *exogenous* feature and if yes, which feature has the highest impact?
- Can you expect a change in gun regulation affected by the number of violent incidents?

- US Government (Legislative branch)
  - Federal level (e.g. Congress)
  - State level (e.g. Governors)
- Law enforcement
  - E.g. police
- Judicial branch
  - E.g. Supreme Court
- News & Media

1.2. Description of the analyzed dataset

- Dataset contains data on 51 US states collected between 1977-1999
- Consisting of 1173 observations on 13 variables (see next slide for the explanation)
- Source: https://vincentarelbundock.github.io/Rdatasets/csv/AER/Guns.csv (accessed on 21.10.2020)

Explanation and format of observed variables

- State (fact): factor indicating observed state
- Year (fact): factor indicating observed year
- Violent (num): violent crime rate (incidents per 100.000 members of population)
- Murder (num): murder rate (incidents per 100.000 members of population)
- Robbery (num): robbery rate (incidents per 100.000 members of population)
- Prisoners (num): incarceration rate in the state in the previous year (sentenced prisoners per 100,000 residents)
- Cauc (num): percentage of Caucasian state population (age: 10 – 64)
- Afam (num): percentage of African-American state population (age: 10 – 64)
- Male (num): percentage of male state population (age: 10-29)
- Population (num): population of state in millions
- Income (num): real per capita personal income (in US-Dollar)
- Density (num): population per square mile of land area divided by 1000
- Law (fact): factor indicating if state has a shall carry law in effect in observed year

1.3. Preliminary work on the dataset

- In general:
  - Removed features „Murder“ and „Robbery“ from the dataset as we focused on the impact on violence and had concerns about potential intersections
- For the PCA & Cluster Analysis:
  - Created sub-datasets focusing on whether a gun law had been in place or not
    - Furthermore, it enables us to standardize data by excluding the binary feature „law“
  - Reduced the number of analyzed data entries to facilitate computation and to avoid altering the interpretation of the feature “law” as a binary feature
- For the Linear Regression & Classification:
  - Changed the feature “Violent” to logarithmic values in order to avoid skewness towards larger values

1.4. Limitations

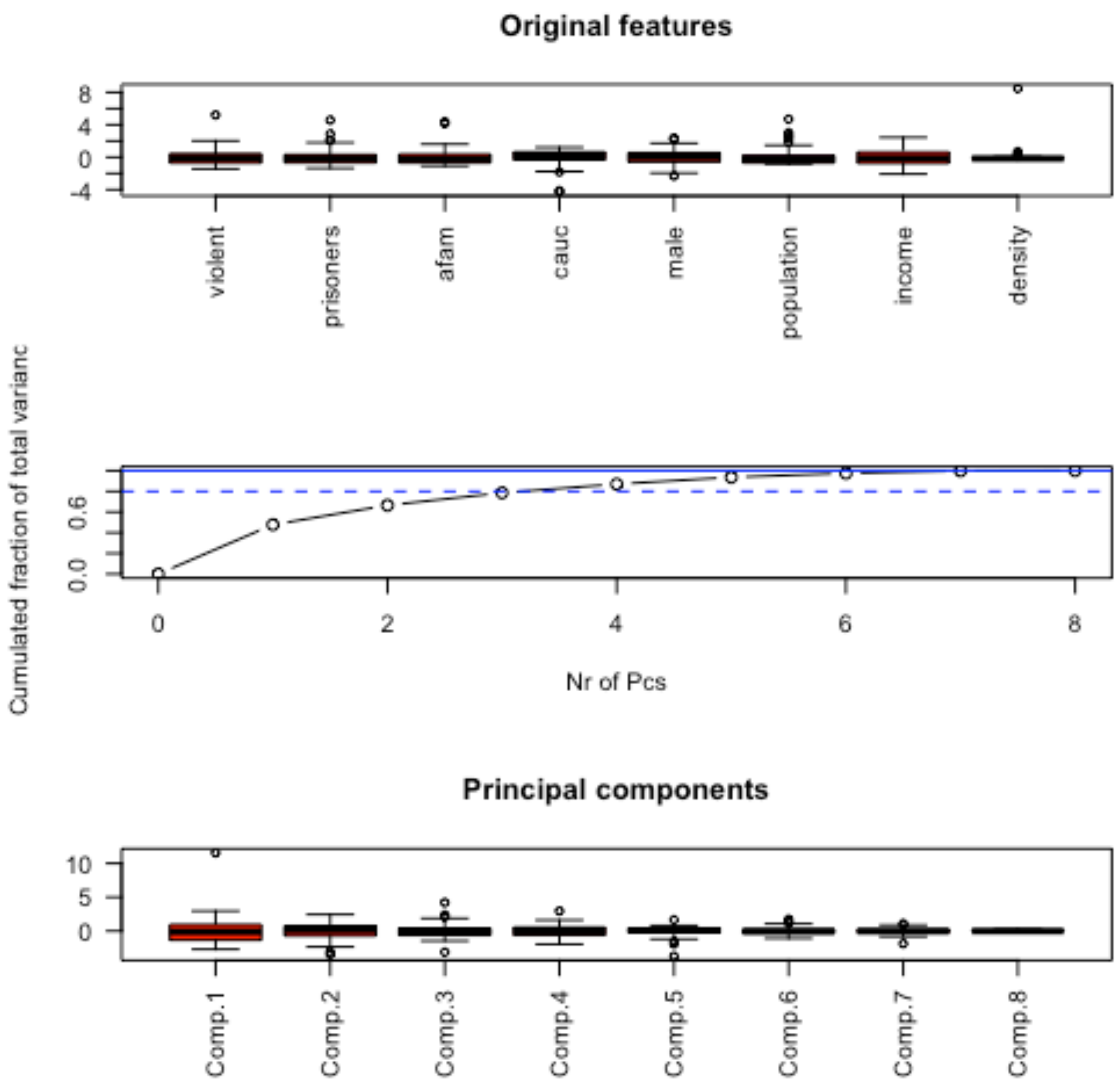
- Data was collected by a third party
- Outdated dataset
- Lack of information on features, such as: “murder”, ”robbery” or “law”
- Our own preliminary work on the dataset

1.5 Recap

PCA

- Plot 2 shows that 3 features cover around 80% of the total variance
- Plot 3 shows that starting from the fourth component the number of outliers are higher than in first three components

Figure 1



PCA

- Representation of chosen main loading vectors
- Vector 1: Afam, prisoners and violence are highly correlated, it amounts to 47% of the proportion
- Vector 2 shows income, population and Caucasian are highly correlated

Figure 2

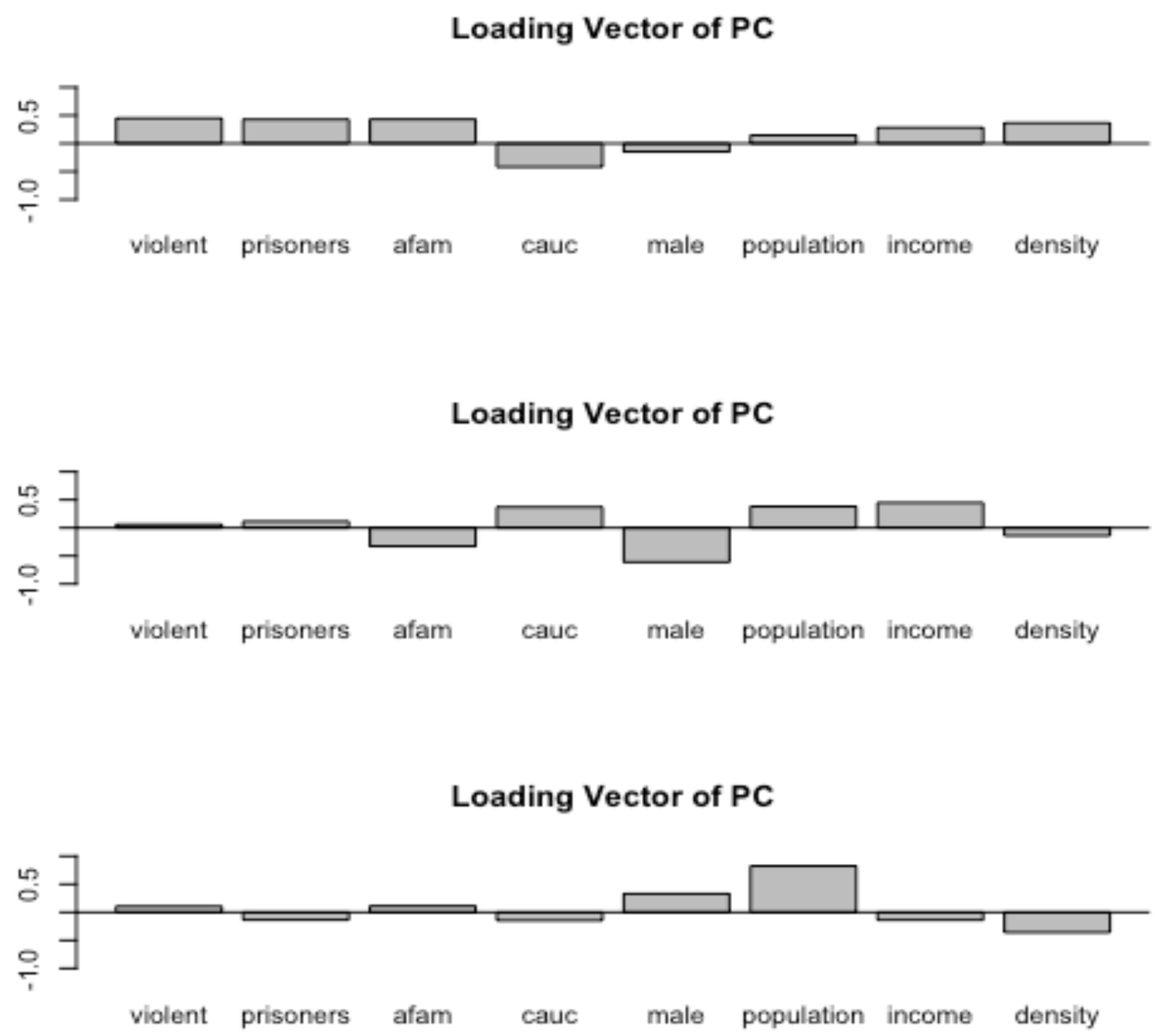


Figure 3

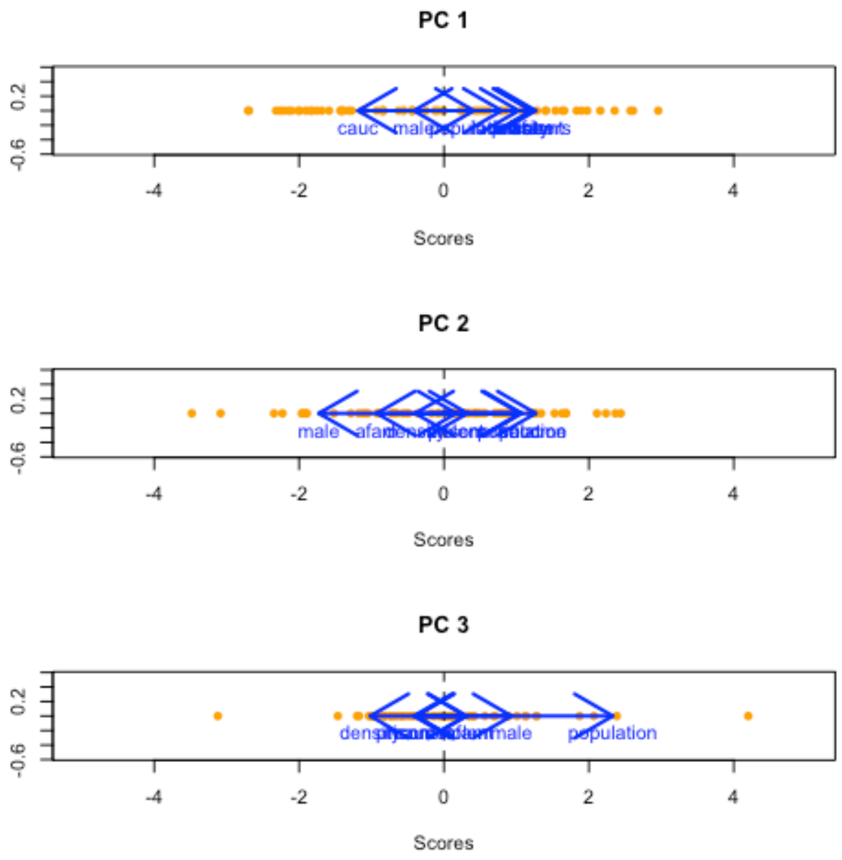
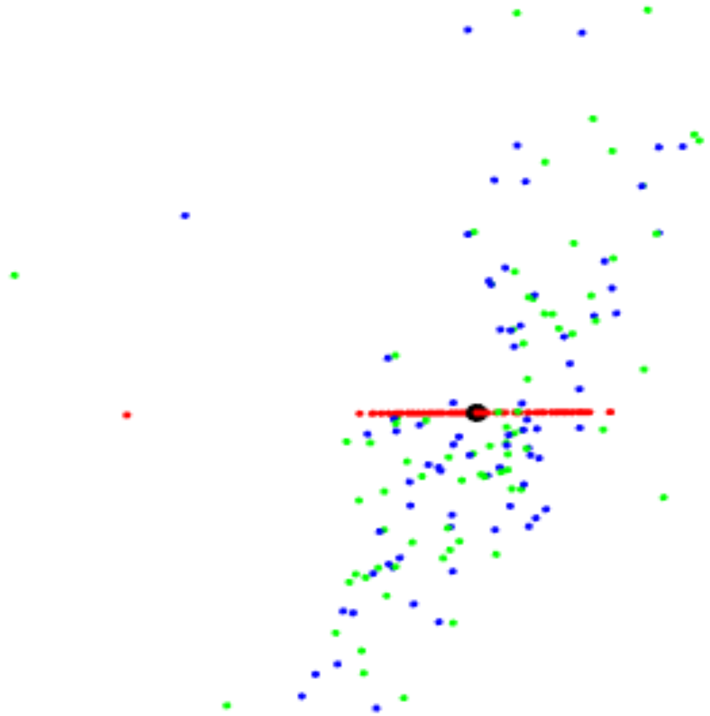


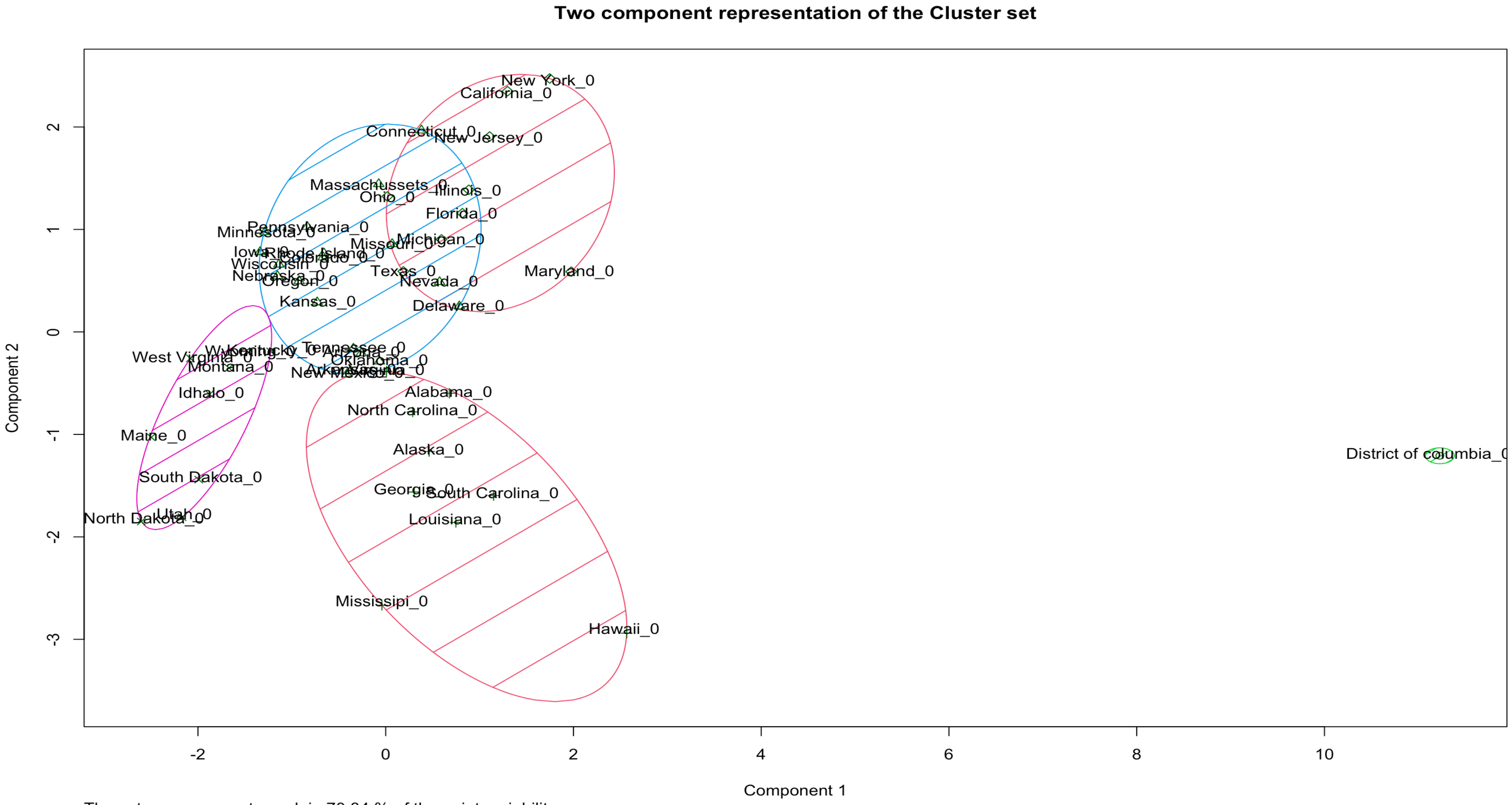
Figure 4



Each PCA shows the state-score

2D representation of states

Figure 5

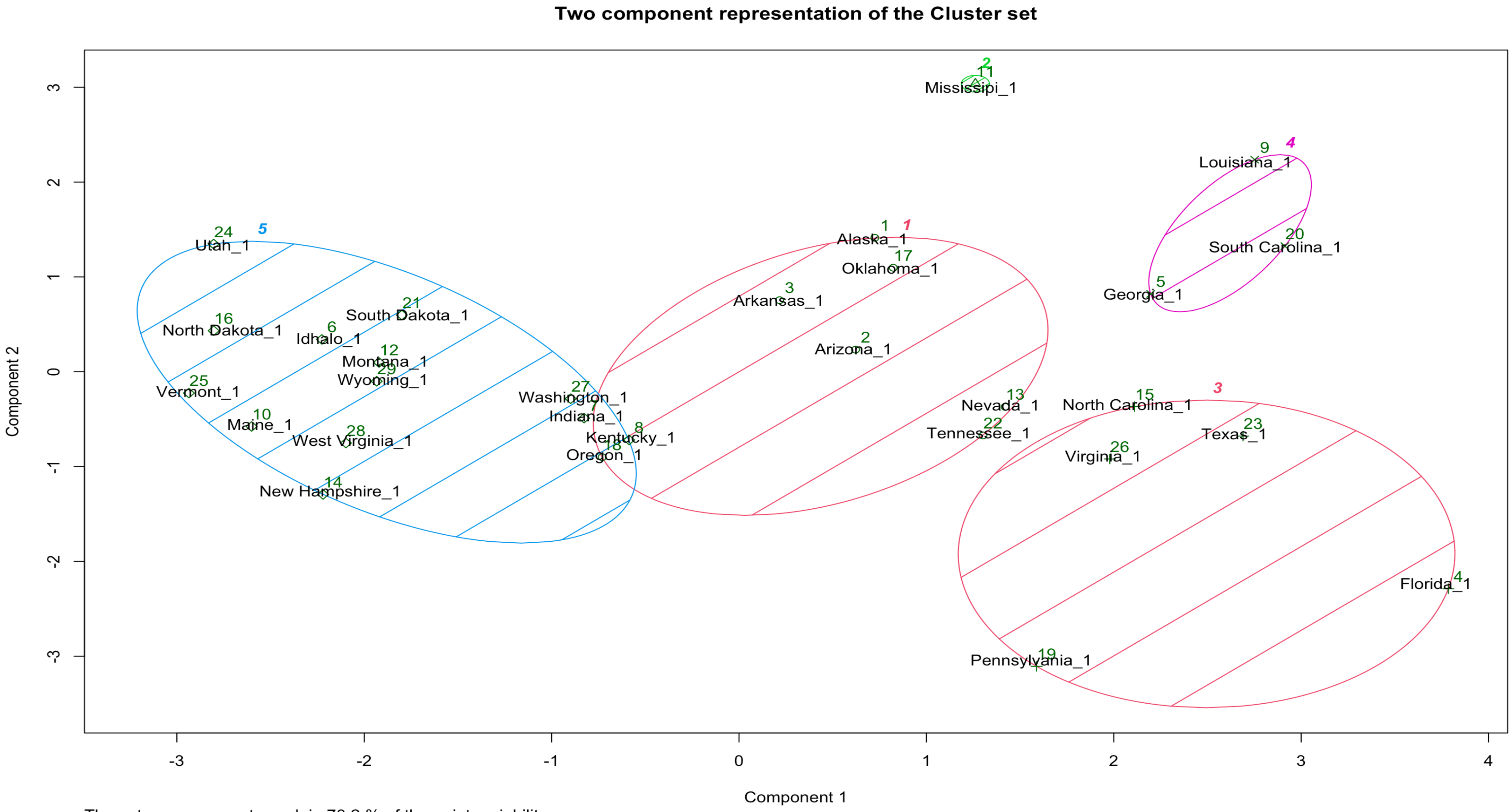


These two components explain 70.84 % of the point variability.

10.11.20

Representation of states in clusters without an effective gun-law

Figure 6

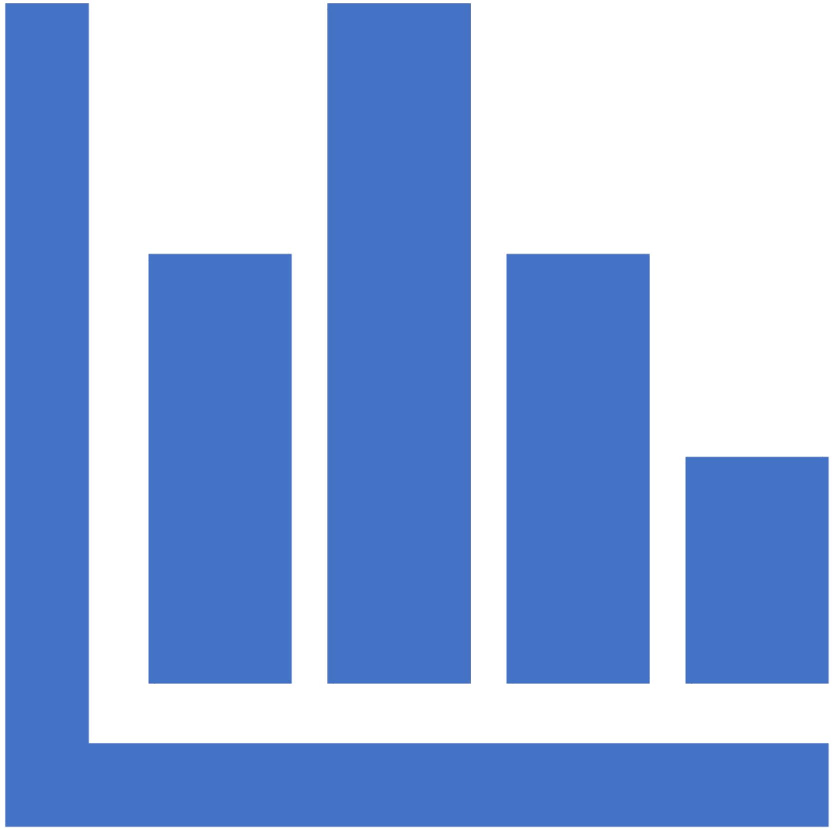


These two components explain 70.2 % of the point variability.

10.11.20

Representation of states in clusters with an effective gun-law

2. Analysis and main results

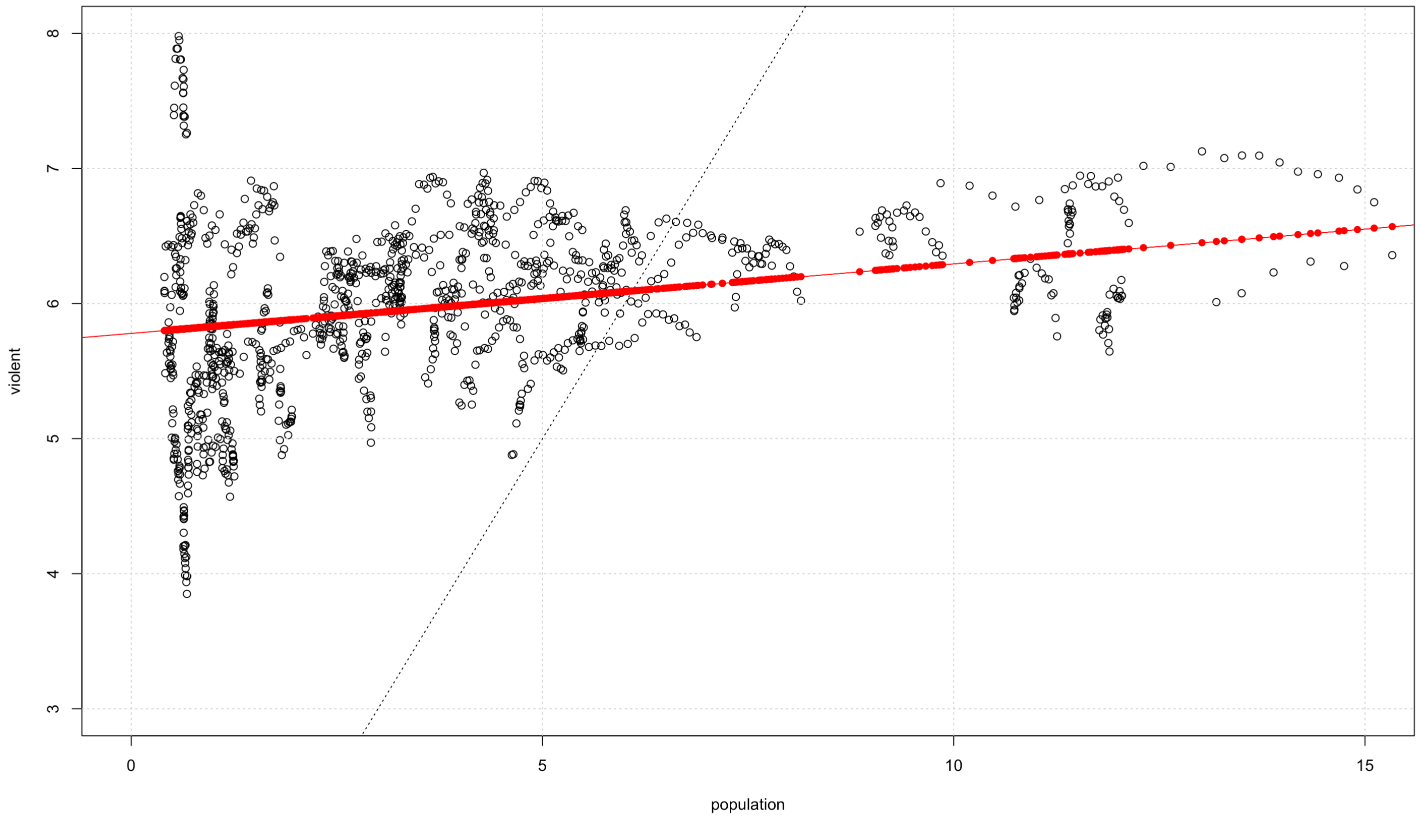


2.1 Simple and Multiple Linear Regression

Simple Linear Regression - Population

- The plot shows the best approximation model for the regressor “Population” and the response “Violent”

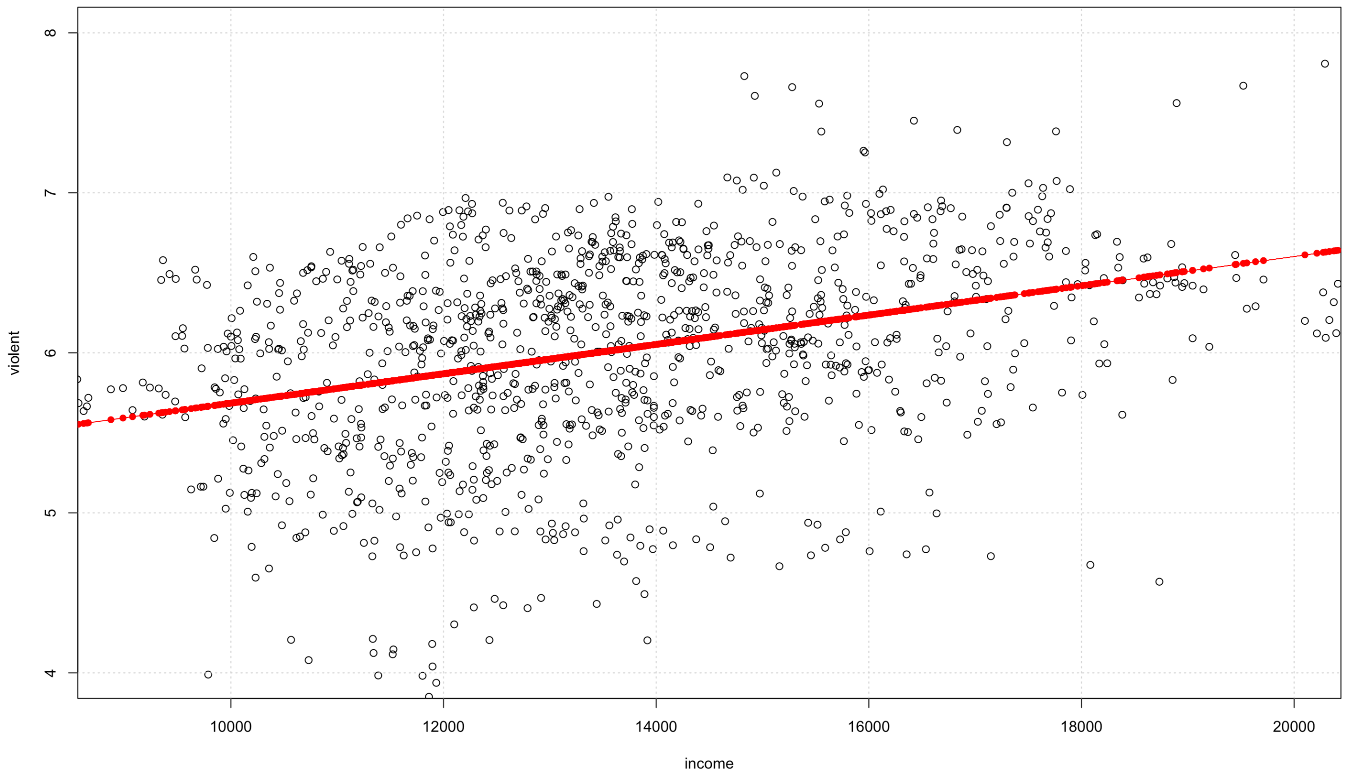
Figure 8



Simple Linear Regression - Income

- The plot shows the best approximation model for the regressor “Income” and the response “Violent”

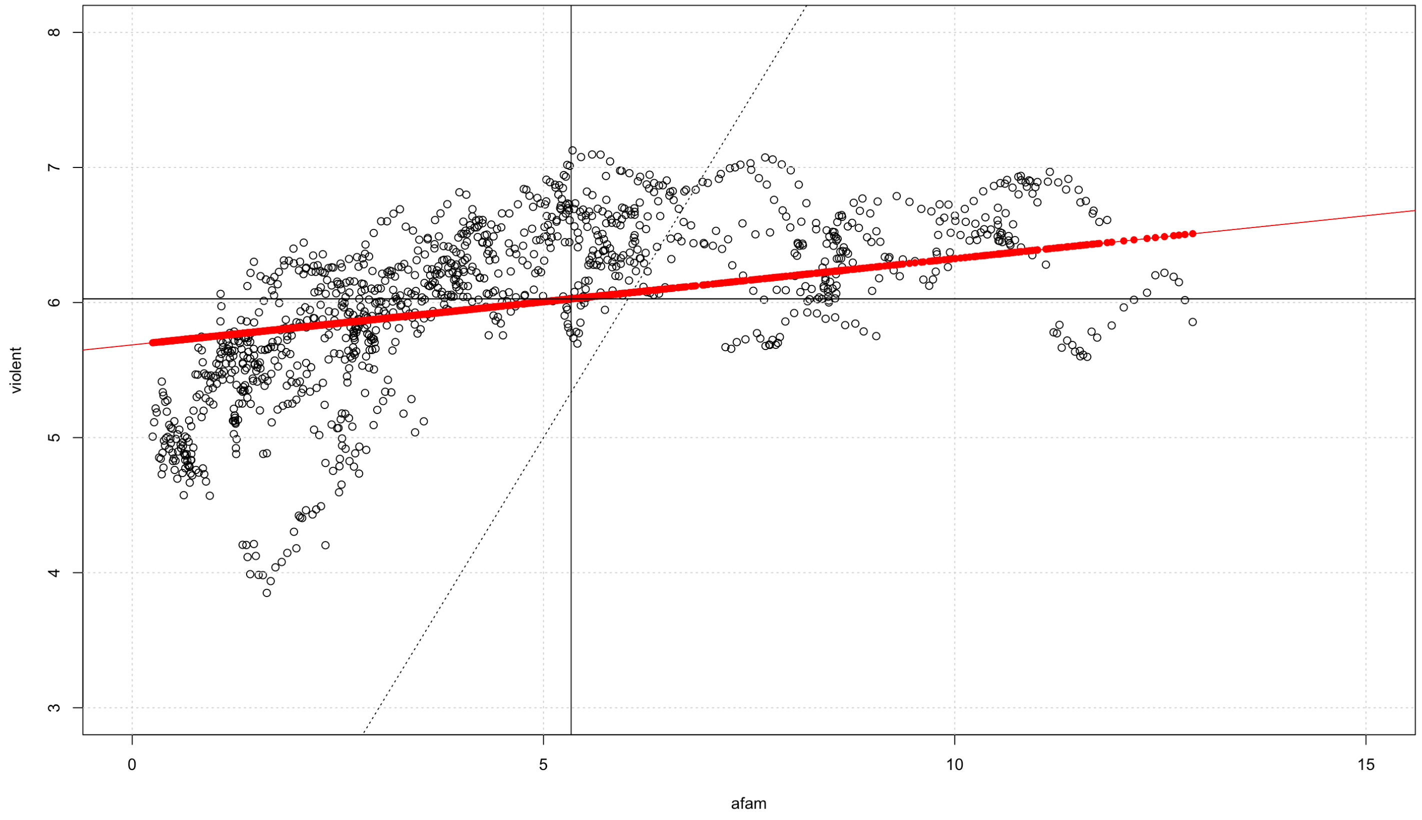
Figure 9



Simple Linear Regression - Afam

- The plot shows the best approximation model for the regressor “Afam” and the response “Violent”

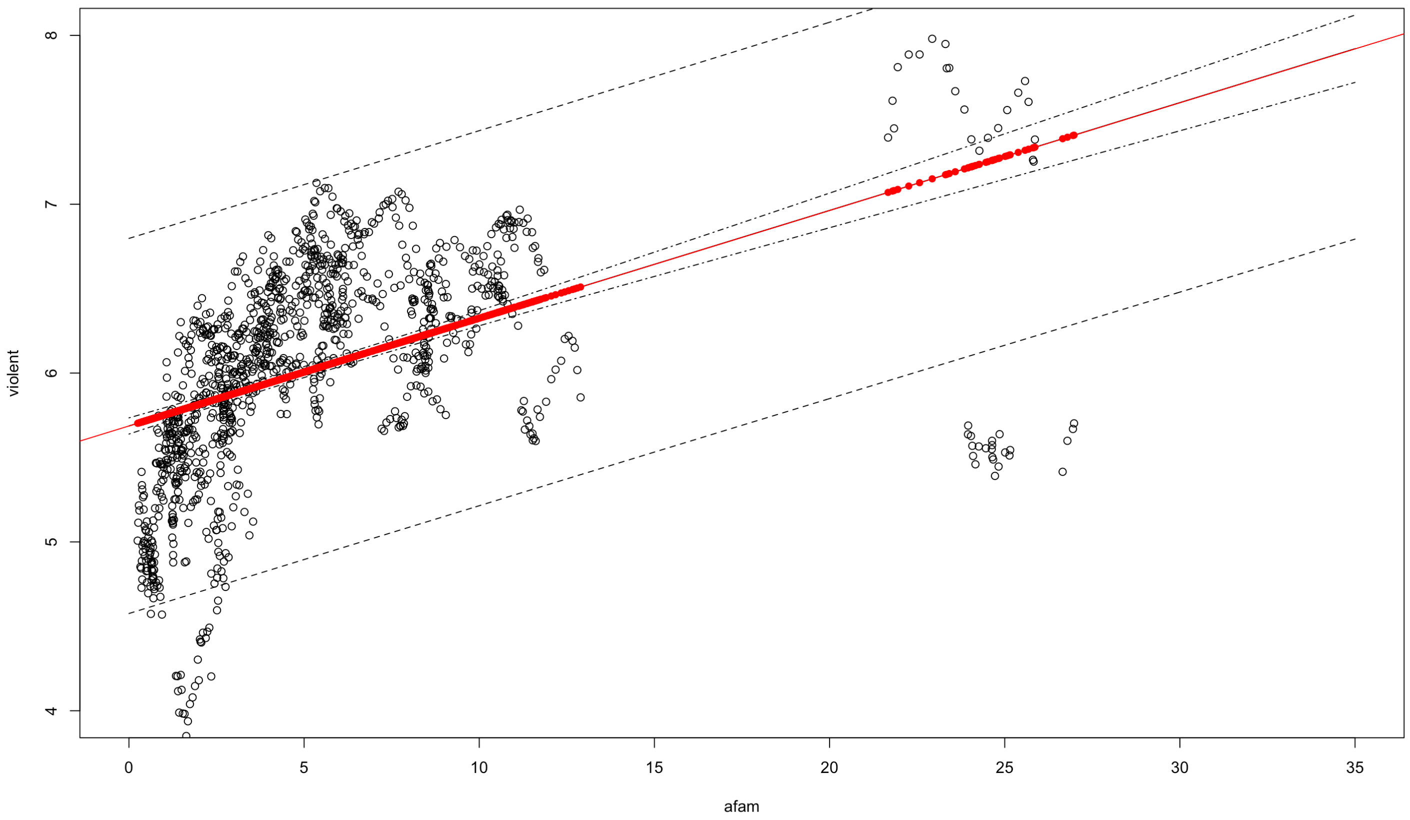
Figure 10



Simple Linear Regression - Afam

- The plot shows the best approximation model for the regressor “Afam” and the response “Violent”
- Additionally, showing confidence intervals and prediction intervals

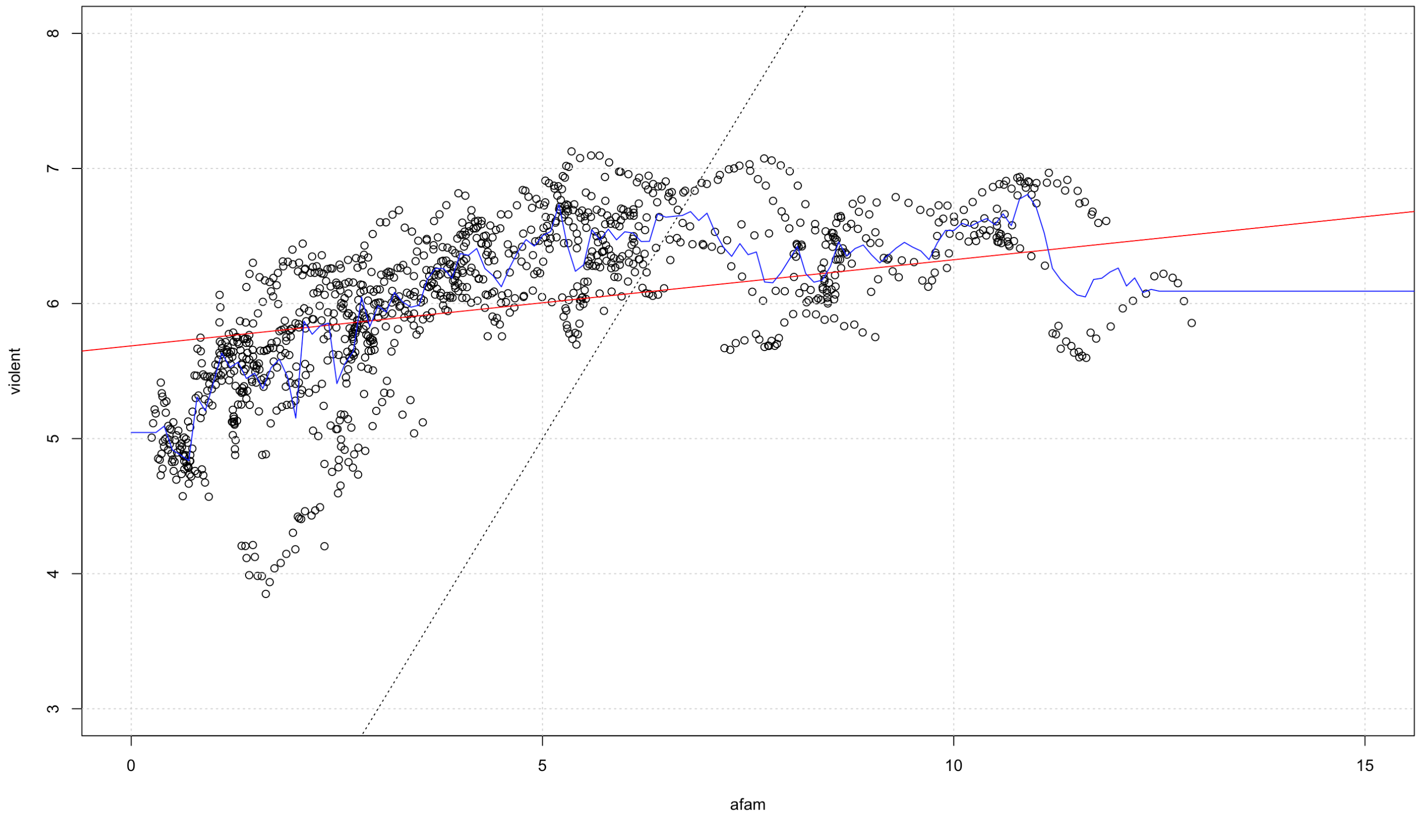
Figure 11



Simple Linear Regression - Afam

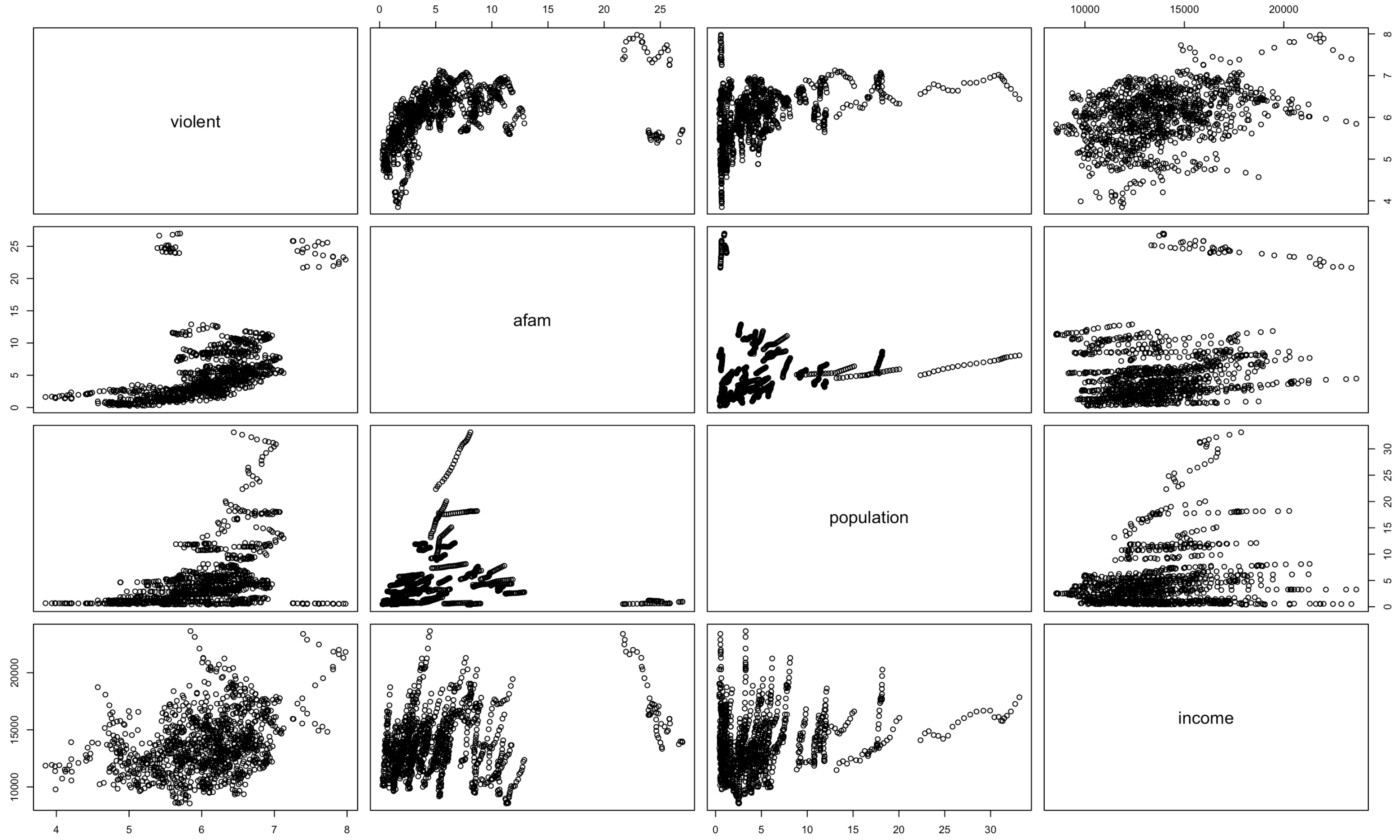
- The plot shows the best approximation model for the regressor “Afam” and the response “Violent”
- Blue line was obtained using the K-Nearest-Neighbor method

Figure 12



Multiple Linear Regression

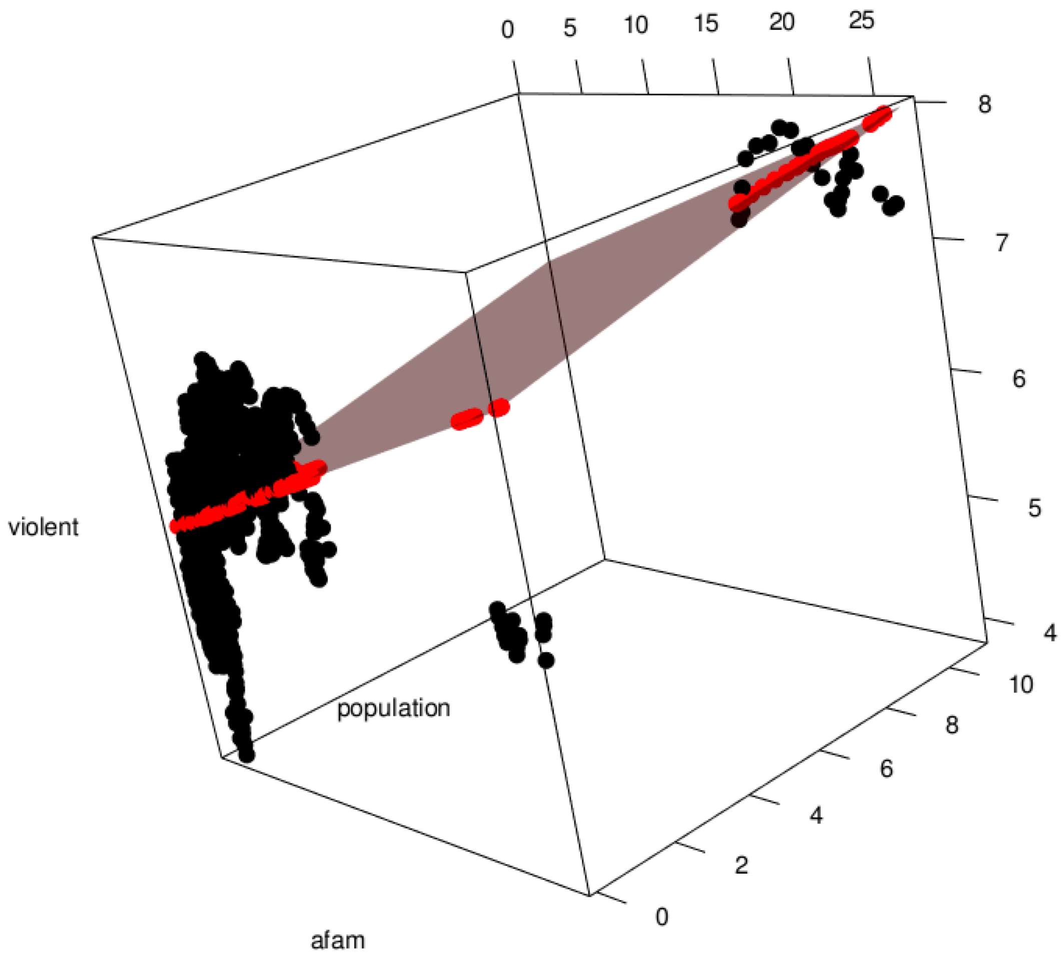
Figure 13



3D depiction Multiple Regression

- The plot depicts the relationship between "Afam" and "Population" as regressors and our response variable "Violent"

Figure 14



2.2 Prediction & relevance

Prediction Model

- F-Statistics: 294,5
  - Denying the H_o-hypothesis
- Adj. R^2: 55,6%
- P-Value: < 2,2e-16
- Equation derived from mixed selection
- Est. coefficients are significant

1. The fundamental equation:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 * x_1 + \hat{\beta}_2 * x_2 + \hat{\beta}_3 * x_3 + \hat{\beta}_4 * x_4 + \hat{\beta}_5 * x_5$$

2. Fundamental equation adjusted to our model:

$$\hat{y} = 3{,}28 + 0{,}024 * x_1 + 0{,}04 * x_2 + 0{,}002 * x_3 + 0{,}23 * x_4 - 0{,}36 * x_5$$

- $\hat{\beta}_0$(intercept) = *3,28*
- $\hat{\beta}_1$(afam) = *0,024*
- $\hat{\beta}_2$(population) = *0,04*
- $\hat{\beta}_3$(prisoners) = *0,002*
- $\hat{\beta}_4$(income) = *0,23*
- $\hat{\beta}_5$(law) = *-0,36*
  - In reference to factor 0 (i.e.: No Law)

3. Predicted violence rate (log) if a gun-law is in place (c.p.):

$$\hat{y} = 3{,}28 + 0{,}024 * 0 + 0{,}04 * 0 + 0{,}002 * 0 + 0{,}23 * 0 - 0{,}36 * 1$$
$$\hat{y} = \boldsymbol{2{,}92}$$

Most relevant features

- In order to compare the independent variables we had to center all variables
- Features are taken from our best model (see slide before)

1. Gun-Law (Yes/No): -0,36
2. Population: 0,04
3. Variety of ethnicity: 0,02
4. Prisoners: 0,002
5. Income: 0,00001

Figure 15

Logistic Regression

- The model indicates a negative relation between Violence and gun-law-regulation
- Est. coefficient of the regressor, log(violence),: -1,09
    - P-Value: <2e-16
    - Significant

3. Conclusion

Which answers can be obtained from our results?

- Violence rates differ greatly among the states but remain the same over time
- Violence occurs more in states with higher population, more prisoners, higher percentage of different ethnicities and higher personal income
- Yes, there are exogenous features which impact violence
  - These features are: Regulation (Gun-Law), Variety of ethnicity, population, prisoners and income
- If the violence rate (log value) is low, it indicates that there is most likely an effective gun-law in place

Thank you for your attention!