Analyzing Mass Shootings in the USA between 1982 and 2023.

An Exploratory Data Analysis.

Sandra Gedig, Mag.

Bachelor of Science Data Science

sandra.gedig@iu-study.org

Matriculation number: 92201974

DLBDSEDAV01 – Exploratory Data Analysis and Visualization

Task 1: Visually Exploring a Data Set

Supervisor: Prof. Dr. Visieu Lac

Study Code:

https://github.com/sgbyteninja/data_exploration_and_visualization_shootings_usa

Date: 02.09.2024

Table of Contents

List of Abbreviations I

List of Figures I

1. Introduction: Public Health and Mass Shootings 1

2. Investigation of Mass Shootings in the USA. Descriptive Statistics and Correlations. 2

3. Deeper Analysis of Mass Shootings 4

3.1. Increasing Prevalence of Mass Shootings 4

3.2. Social Demographics of Mass Shooters: Analysis of Race, Age and Gender 4

4. Qualitative Analysis of White Mass Shooters between 20 and 25 Years of Age 7

5. Conclusion 8

References 9

List of Abbreviations

Abbreviations	Meaning
FBI	Federal Bureau of Investigation

List of Figures

All charts are based on the Mother Jones mass shooting data, available at https://www.motherjones.com/politics/2012/12/mass-shootings-mother-jones-full-data/. All charts were created by the author using Python.

Figure 1: Correlation Matrix for Numerical Features of Mass Shootings Incidents 3

Figure 2: Mass Shootings in the USA per Year with Trend Analysis 4

Figure 3: Age Distribution of Mass Shooters in the USA 5

Figure 4: Age Distribution of Mass Shooters per Ethnicity Group 6

Figure 5: Distribution of Prior Signs of Mental Health Issues by Race 7

1. Introduction: Public Health and Mass Shootings

Mass shootings in the USA are one of the central interfaces between public health and safety and pose a major threat to citizens' sense of security, but more importantly are a threat to their lives. According to the FBI, a mass shooting is defined as the use of weapons in public with at least four victims without an organizational background and not for economical gain. This excludes for example gang related violence or terroristic attacks (Federal Bureau of Investigation, 2005, p. 8). As a result, the concept of mass shooting, as it is established in research, has a certain bias and only represents part of the gun violence in the USA. Furthermore, it is important to mention that this form of violence is primarily a US-American phenomenon. Compared to the rest of the developed world, the USA has the highest number of mass shootings per capita. Only the civil war in Yemen recorded more mass shootings than the USA between 1966 and 2012 (Fisher & Keller, 2017; Lankford, 2016).

This study is based on Mother Jones mass shooting data compiled in the wake of the July 2012 massacre at an Aurora movie theater (Follmann et al., 2012, 2024). This database attempts to record all such crimes committed in the United States, consistent with the FBI's definition of mass shootings.

In total, the database records 149 mass shootings in a period from 1982 to the end of 2023. Data would also be available for 2024, but as this year has not yet been completed, this data is not included in this study. In total, over 1149 people died and 1628 people were injured in these events. This demonstrates that, given the large number of deaths in the US due to gun violence, the mass shooting only represents a portion of gun violence in the US (Follmann et al., 2015).

In fact, many mass shooters exhibit a wide range of behaviors and psychiatric illnesses, such as depression, anxiety or personality disorders in general, in the years or months before they become perpetrators. At the same time, it is important to note that only a small proportion of people with mental disorders commit mass shootings (Follmann et al., 2012).

It is important to mention that the data set obviously only contains information about crimes that have already been committed. This leads to a retrospective view of events, which provides valuable insights, but does not allow any predictions or preventive measures to be directly derived.

To understand the problem, political and cultural circumstances must be considered in addition to the psychological and social aspects. The USA has an exceptionally high rate of firearms compared to other countries. The availability of weapons goes hand in hand with the possibility of gun violence. There are various approaches to minimize this risk. One of these is restricting the availability of firearms and implementing controls to ensure compliance.

The structure of the study is as follows: In the first step, some general aspects of the existing data set are described, like correlations, mode and descriptive aspects. In the second part, the data is visualized more deeply and finally from that analysis out, some qualitative analysis about white shooters between 20 and 25 years are presented. The reason for that choice will be explained further in the study.

All results of the analysis concerning the Mother Jones mass shooting data are available at https://github.com/sgbyteninja/data_exploration_and_visualization_shootings_usa. A cookie cutter project folder structure was installed in advance for better organization of the project. All Jupyter notebooks are stored in the notebooks folder. The data review and EDA can be found under data_preparation.ipynb, the analysis under mass_shooting.ipynb. The data is divided into raw (original data), interim (cleaned data).

2. Investigation of Mass Shootings in the USA. Descriptive Statistics and Correlations.

The data set analyzed contains 25 columns, of which only 5 are numerical. These numerical columns include the number of deaths, the number of injuries, the total number of victims, the age of the shooter and the year of the crime. It should be noted that the year of the crime was extracted from the date of the crime (Follmann et al., 2024).

The results of the correlation analysis are unsurprising. Only the number of injured, dead and the total number of victims correlate significantly with each other, which can be seen in Figure 1. This is as expected, since a higher total number of victims implies a higher number of fatal or non-fatal injuries, which is in a way tautological.

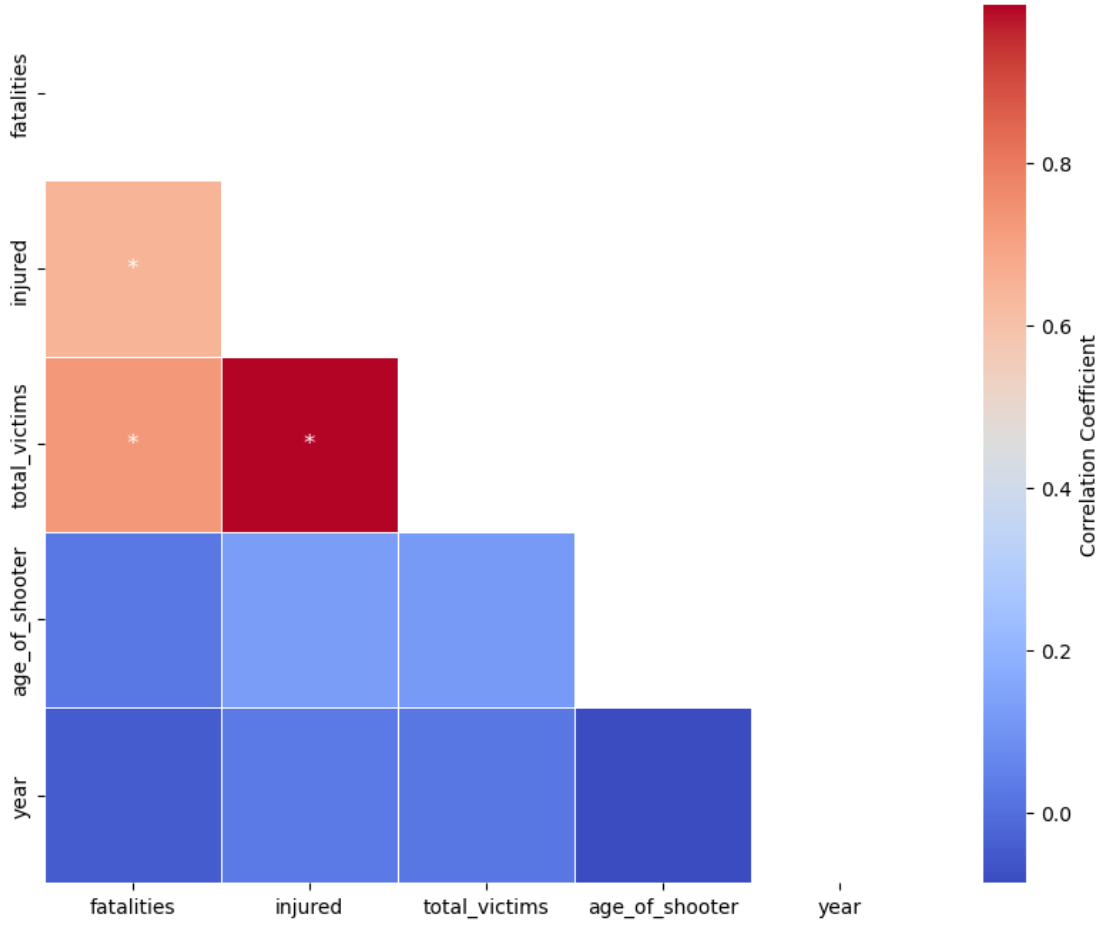
Interestingly, the analysis shows that the year of the mass shooting has no effect on the number of injuries, nor does the age of the perpetrator. This is surprising as one might assume that with the increase in gun control measures or legislation, mass shootings would be less fatal. There is also no linear influence of the age of the perpetrator on the number of victims, which suggests that aspects such as physical fitness or similar do not play a role in the effectiveness of the crimes.

However, it should be noted that correlations only depict linear relationships. Other relationships, such as an initially increasing and then decreasing relationship, cannot be captured in the correlation matrix. There is also the question of the average age of the perpetrators and the average lethality of a mass shooting.

The average perpetrator is 34 years old, with 96 percent being male, and only four percent identified as non-male. Their crimes claimed on average a total of 18 victims, of whom just 8 were fatally and 10 were not fatally injured. However, it should be noted here that the data set is based on the Federal Bureau of Investigation (FBI) definition of mass shootings, which states that a mass shooting is only classified as such if there are more than 4 victims (Federal Bureau of Investigation, 2005, p. 8; Follmann et al., 2012). This means that mass shootings that are stopped are not included in the dataset. For example, shooters can be prevented by armed citizens, police officers or, under certain circumstances, by malfunctioning weapons. The present data set therefore exhibits a sampling bias.

Figure 1

Correlation Matrix for Numerical Features of Mass Shootings Incidents



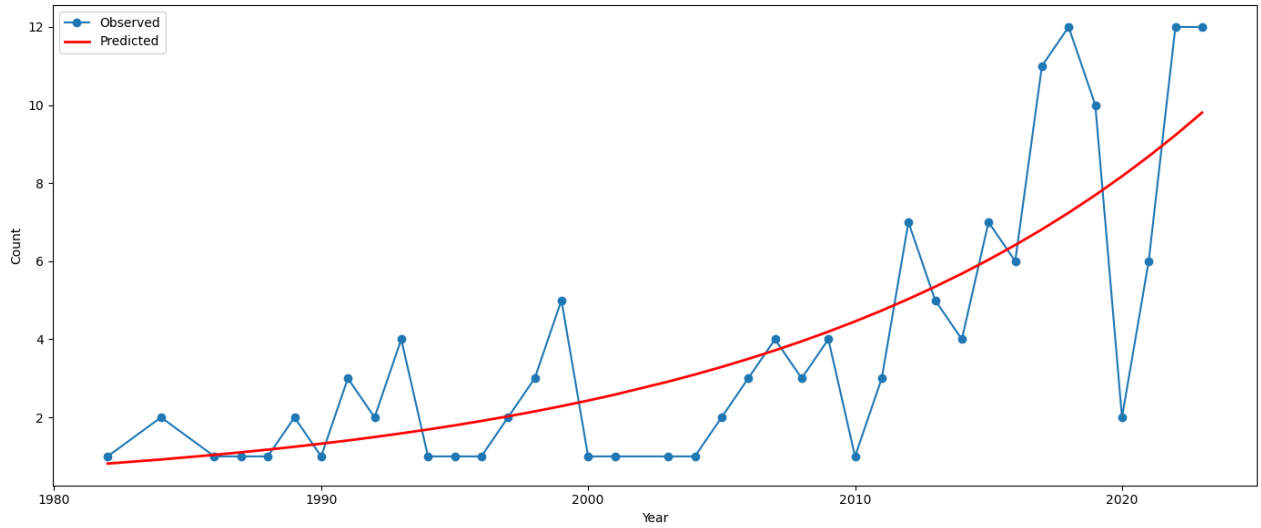
3. Deeper Analysis of Mass Shootings

3.1. Increasing Prevalence of Mass Shootings

In general, the number of mass shootings is escalating, as can be seen in Figure 2.

Figure 2

Mass Shootings in the USA per Year with Trend Analysis



This figure shows the number of shootings per year in the period between 1982 and 2023. What is clear visually can also be understood statistically.

A Poisson regression was calculated to illustrate the relationship between time and mass shooting events. A Poisson regression is equipped with the logarithmic link as standard, so that it is well able to depict exponential growth. Furthermore, Poisson regression is particularly suitable for count data, as there can be no negative numbers of shootings. There is a highly statistically significant relationship between the year and the number of shootings. To interpret this relationship in terms of ratios, the coefficient can be exponentiated: exp(0.0606) ≈ 1.0625. This means that the expected number of shootings increases by a factor of approximately 1.0625 for each additional year, which corresponds to an increase of approximately 6.25%. In other words, the number of shootings increases by just around 6% per year. The p-value is $p<0.001$ which indicates high significance.

3.2. Social Demographics of Mass Shooters: Analysis of Race, Age and Gender

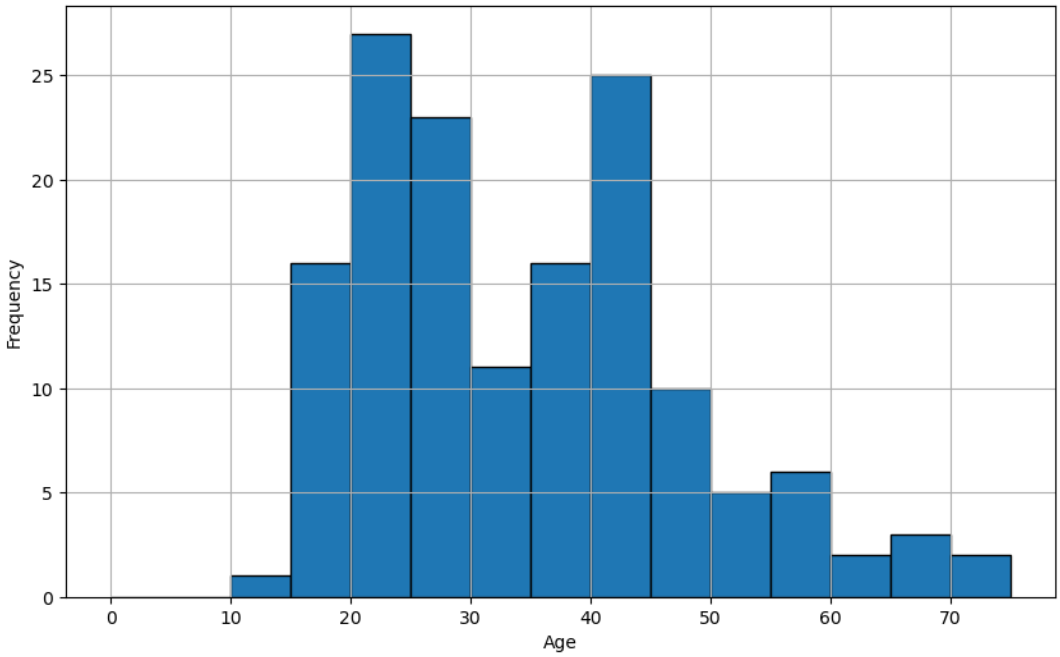
Even though the average age of the shooters has already been briefly discussed, it is important to show the distribution of the shooters. A meaningful interpretation is only possible with symmetrical and unimodal distributions. However, age is not usually distributed symmetrically, which is also shown by the age distribution of the shooters in Figure 3.

Here it can be seen that the age distribution of the shooters is bimodal, i.e. there are two modes. The first mode is between 20 and 25, and the second is between 40 and 45 years. This finding demonstrates that different age groups are represented differently and may require different

prevention approaches. The bimodal distribution may indicate different psychosocial factors or life circumstances that are particularly pronounced in these age groups. A further analysis of these factors could provide further insights into preventing and combating mass shootings.

Figure 3

Age Distribution of Mass Shooters in the USA



In general, a bimodal distribution often indicates that several underlying processes overlap, i.e. that different populations are present. This leads to the next question, namely the ethnicity of the shooters. It is important to note that ethnicity alone does not provide information about specific character traits. However, ethnicity and being a perpetrator can be confounded through other factors such as poverty, discrimination or place of residence. Since such confounding aspects are not included in the present data set, it is only possible to investigate whether there are differences in mass shootings by ethnicity.

In the USA, there are major differences between Afro-American and White citizens in several socio-demographic factors which has a deep effect on their lives (Correll et al., 2002; Demny, 2004, pp. 18–22). It can therefore be assumed that, at least on average, these divergent experiences can also influence the incidence and form of mass shootings.

Figure 4

Age Distribution of Mass Shooters per Ethnicity Group

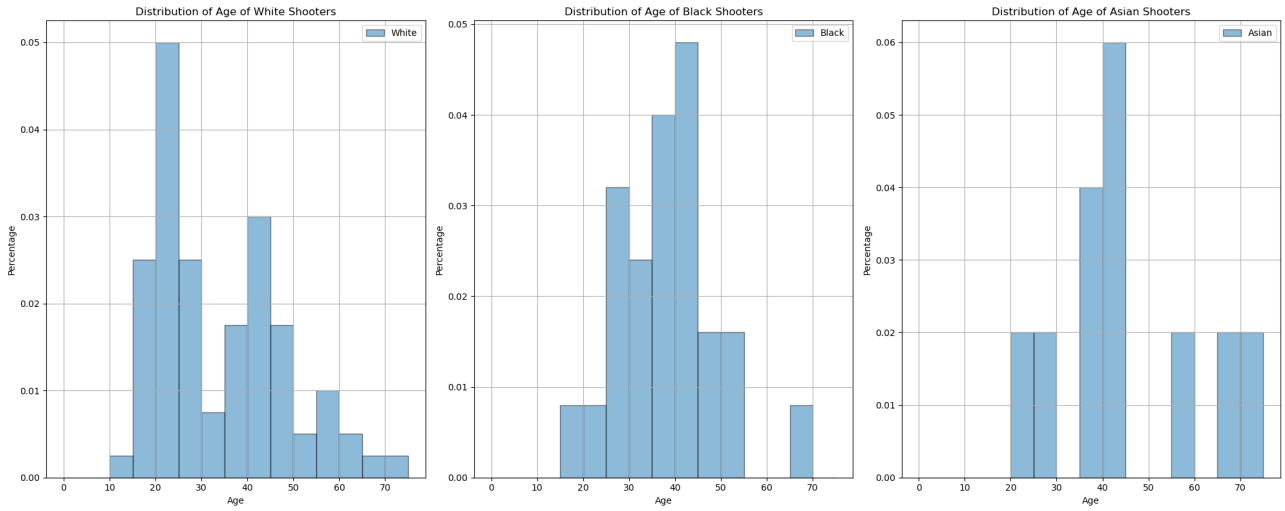


Figure 4 compares the age distribution of White, Afro-American and Asian mass shooters. All three groups have a peak between 40 and 45, yet the White mass shooters age distribution depicts an anomaly.

White mass shooters have a bimodal age distribution. Offenders between the ages of 20 and 25 are particularly common. Interestingly, this age cohort is completely absent among African American and Asian mass shooters. This suggests that white mass shooters between the ages of 20 and 25 are a special case that should be investigated further. To answer this question more precisely, a random sample of white mass shooters between the ages of 20 and 25 will be drawn and qualitatively analyzed to identify commonalities in chapter 4.

However, white, African American and Asian mass shooters have a mode between the ages of 40 and 45. This suggests that a risk factor for mass shootings is particularly pronounced at this age. One possible explanation could be that people aged 40 to 45 believe that their lives will not change fundamentally as they are in later adulthood. This is also an age at which mental illness often becomes entrenched. For example, on average, white shooters exhibit mental health issues 46 percent of the time before they become mass shooters, while in the 40-45 age group, that rate rises to 75 percent.

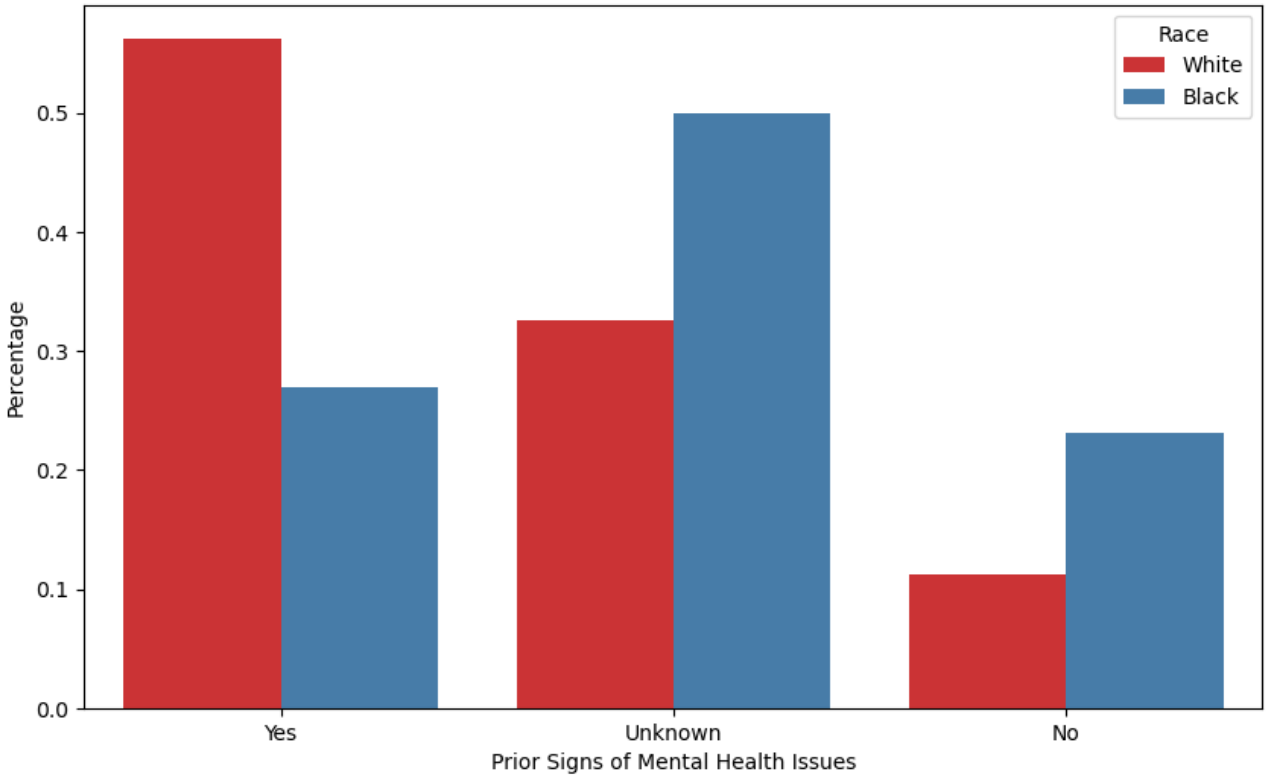
Finally, the question arises regarding the data on racial disparities in mental health, specifically whether there are differences in the prevalence of previous psychological issues among shooters of different racial backgrounds. The data indeed shows a significant disparity. White shooters exhibit more prior signs of mental health issues compared to African American shooters.

However, it is important to consider potential confounding factors (Neal, 2020, p. 4; Pearl & Mackenzie, 2018). It is likely that white individuals have better access to healthcare and insurance, which could result in a higher rate of diagnosed mental health conditions compared to African

Americans (Case & Deaton, 2021). Thus, these figures might offer limited information on the actual differences in mental health among shooters.

Figure 5

Distribution of Prior Signs of Mental Health Issues by Race



4. Qualitative Analysis of White Mass Shooters between 20 and 25 Years of Age

As mentioned before the white mass shooter between 20 and 25 is an anomaly, which demands further investigation. Therefore, a random sample of five shooters was selected and their motives were analyzed:

These shooters were: Robert "Bobby" E. Crimo III, who perpetrated the Highland Park parade shooting on July 4, 2022; Patrick Wood Crusius, who carried out the 2019 El Paso shooting; Dylann Roof who committed the Charleston Shooting and Jonathan Sapirman of the Indiana Mall Shooting in 2022 (AP, 2019; Guarino et al., 2022; Nelson, 2022; Sanchez & Payne Ed, 2016). All these shooters, who had information available, had racist motives. The 95 percent confidence interval for the proportion of white mass shooters having racist motives is (0.449, 1). For one of the random samples no publicly available information could be found. The age might be a confounding factor because in this life period the perpetrators radicalize themselves in online communities and combined with other troubling life situations the risk factors might escalate.

The violence of the white mass shooters analyzed here was directed against the Asian and Latino communities or exhibited general racist and anti-Semitic tendencies. This indicates that a right-wing extremist ideology often contributes significantly to the violence perpetrated by white shooters in this

age group, a factor that has not been adequately addressed in the debate. Such acts of violence committed by white perpetrators against minorities could also be understood as a form of terrorism (Schmid, 2011, pp. 99–148). Thus, one reason why white mass shooters are more prevalent might be a form of racism. It must be added that three samples are relatively small, yet this research opens further research endeavors that might be tested in the future.

5. Conclusion

Overall, the Mother Jones mass shooting data on shootings in the USA paints a dark picture. A time series analysis of the years 1982 to 2023 shows a worrying increase in mass shootings. Using a Poisson regression, an annual increase of 6.25 percent was demonstrated, which indicates exponential growth. This makes it even more important to develop effective solutions, but without jumping to false conclusions:

Although many gun rampage killers suffer from mental illnesses such as personality disorders or depression. However, it should be emphasized at this point that most people whose mental health is affected in this way do not become mass murderers. Thus, it is important to focus on preventive measures and not to falsely stigmatize mental illness (Follmann et al., 2012).

Unfortunately, the retrospective nature of the present data set does not allow any direct derivation of action plans or predictions. However, the analysis does provide valuable insights into the demographic characteristics of mass shooters. It was found that the age distribution of the perpetrators peaks between 20 and 25 years and between 40 and 45 years. Furthermore, the findings indicate that white perpetrators in particular exhibit a high level of racism in the younger age group.

This raises the question of the extent to which these radical political attitudes have influenced the decision to go on the rampage. This aspect urgently requires further investigation. Because if it turns out that the white mass shooters are pursuing a political agenda, these acts will have to be classified as terrorism. Consequently, the rise in these acts would also have to be assessed not as a crisis of mass shootings, but as a crisis of racist terror. This distinction is important because this is a political problem rather than a problem with psychological problems and therefore requires different measures. For example, possible measures to reduce mass shootings triggered by mental health problems could include the introduction of state-subsidized statutory health insurance and the associated closer treatment of these illnesses.

Other measures are needed to counter the racist radicalization of white men. In addition to political education programs, the promotion of intercultural dialogues and community projects, the monitoring of extremist content and the associated development of alternative positive narratives are all possible options.

However, all these measures to reduce mass shootings are based on research. This should develop a comprehensive understanding and provide a basis for effective practical solutions.

References

AP (2019, August 6). The Latest: Discrepancies between US, Mexico victim lists. *AP-News*. https://apnews.com/article/5fb2144947974631912d1efa889f932b

Case, A., & Deaton, A. (2021). *Deaths of despair and the future of capitalism*. Princeton University Press. https://www.jstor.org/stable/10.2307/j.ctv161f3f8

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The Police Officer's Dilemma: Using Ethnicity to Disambiguate Potentially Threatening Individuals. *Journal of Personality and Social Psychology*, *83*(6), 1314–1329. https://doi.org/10.1037//0022-3514.83.6.1314

Demny, O. (2004). *Die Wut des Panthers: Schwarzer Widerstand in den USA ; [die Geschichte der Black Panther Party]* (4., erg. Aufl.). Unrast.

Federal Bureau of Investigation. (2005). *Serial Murder: Multi-Disciplinary Perspectives for Investigators.*

Fisher, M., & Keller, J. (2017, November 7). Why Does the U.S. Have So Many Mass Shootings? Research Is Clear: Guns. *New York Times*. https://www.nytimes.com/2017/11/07/world/americas/mass-shootings-us-international.html

Follmann, M., Aronsen, G., & Pan, D. (2012). A Guide to Mass Shootings in America. *Mother Jones*. https://www.motherjones.com/politics/2012/07/mass-shootings-map/

Follmann, M., Aronsen, G., & Pan, D. (2024). *US Mass Shootings, 1982–2024: Data From Mother Jones' Investigation*. https://www.motherjones.com/politics/2012/12/mass-shootings-mother-jones-full-data/

Follmann, M., Lurie, J., & West, J. (2015, April 15). The True Cost of Gun Violence in America: The data the NRA doesn't want you to see. *Mother Jones*. https://www.motherjones.com/politics/2015/04/true-cost-of-gun-violence-in-america/

Guarino, M., Kornfield, Meryl, & Warrick, J. (2022, July 4). At least six dead in shooting at July Fourth parade outside Chicago: Massacre in Highland Park joins other recent mass shootings that have restarted emotional debate over gun control. *Washington Post*. https://www.washingtonpost.com/politics/2022/07/04/fatal-shooting-independence-day-parade/

Lankford, A. (2016). Public Mass Shooters and Firearms: A Cross-National Study of 171 Countries. *Violence and Victims*, *31*(2), 187–199. https://doi.org/10.1891/0886-6708.VV-D-15-00093

Neal, B. (2020). *Introduction to Causal Inference: from a Machine Learning Perspective* [s.p.]. https://www.bradyneal.com/Introduction_to_Causal_Inference-Dec17_2020-Neal.pdf

Nelson, S. (2022, December 21). Police: Greenwood mall shooter fascinated with German Reich, yet 'no clear motive' for attack. *Indianapolis Star*.

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect* (First trade paperback edition). Basic Books.

Sanchez, R., & Payne Ed (2016, December 16). Charleston church shooting: Who is Dylann Roof? *CNN*. https://edition.cnn.com/2015/06/19/us/charleston-church-shooting-suspect/

Schmid, A. P. (2011). The Definition of Terrorism. In A. P. Schmid (Ed.), *The Routledge Handbook of Terrorism Research* (pp. 39–157). Routledge.