The Tragic Impact of Drone Strikes on Civilian Casualties

A Research Project on the Statistical Impact of U.S. Attacks

Agastya Rao, Brandon Erickson, Ivy Le, Kuanyi Chen, Kylie Linden,

Michael Gureghian, Zhiyuan Zhou

Stats 140 XP: Practice of Statistical Consulting

Prof. Vivian Lew

March 13, 2025

Contents

The Tragic Impact of Drone Strikes on Civilian Casualties

There have been many debates about the necessity, morality, and consequences of U.S. drone strikes, particularly in Afghanistan, Pakistan, Yemen, and Somalia. Many view it as a simple retaliation against prior terrorist attacks, but the reality is not so simple. We aimed to discover the true consequences of such attacks by analyzing the available data, such as current president, drone strike location, and the minimum reported injuries, and then relating this to the estimated civilian casualties. We researched how geographic location had an influence on casualties and if presidency, location, demographics, influenced the estimated number of casualties in conflict-related events. We used a geospatial heatmap for the first goal, and linear regression for the second–both providing useful insight into the issue at hand. Ultimately, we were able to conclude geographic factors, such as country, latitude, and longitude, to have a significant impact on casualty rates, and that presidency and demographics also have an impact. There is a correlation between the number of casualties and drone strikes and the attacks are ineffective in minimizing collateral damage. With more drone strikes, there were higher casualties reported and, because of this, we determine that the drone strikes do have consequences on civilian casualties and are not only detrimental to U.S. relations, but the lives of innocent bystanders.

Introduction

Drone warfare has played a significant role in U.S. military operations, particularly in counterterrorism efforts across Afghanistan, Pakistan, Somalia, and Yemen. While drone strikes are often justified as precision attacks to minimize collateral damage, the reality of their impact—especially on civilian populations—remains a subject of debate. The Drone Wars dataset, compiled using data from The Bureau of Investigative Journalism, provides a detailed record of drone-enabled airstrikes, offering an opportunity to analyze their geographic distribution, casualty rates, and the confirmation status of U.S. involvement.

This project aims to explore the relationship between geographic location and the number of drone strike deaths. Specifically, we seek to determine whether certain regions experience higher civilian casualty rates and to what extent factors such as population density or conflict intensity influence these patterns. By leveraging statistical analysis and data visualization, our research will provide a data-driven perspective on the humanitarian and operational implications of drone warfare.

We aim to answer two research questions within this report, the being how does the geographic location affect the number of drone strike deaths? Are certain regions more prone to higher civilian casualty rates? The geographic distribution of drone strikes is a critical factor in understanding their impact. Different

regions may exhibit varying casualty rates due to differences in population density, military activity, and local conflict dynamics. For example, densely populated areas might experience higher civilian casualties compared to remote regions, where strikes may be more targeted. Additionally, regional political stability, the presence of militant groups, and the availability of reliable reporting sources may also influence the recorded number of deaths. By examining the geographic aspect, we can assess whether certain regions bear a disproportionate burden of drone warfare and explore potential factors contributing to these disparities.

Our next research question asks to what extent do presidential leadership, geographic location, and casualty demographics predict the estimated number of casualties in conflict-related events? We want to identify key predictors of casualty estimates in conflict-related events, which could help inform humanitarian efforts and policy-making related to war and crisis management. We will analyze presidency, geographic location, casualty demographics, injury data, and reporting variance to determine the effect. Each of the aforementioned variables can heavily impact what is reported and not reported in regards to estimated casualties, and there is a potential relationship. Leadership plays a huge role in these strikes, and where the strikes are located can also determine how civilians are affected, such as hills, debris, etc., it will be insightful to research. We will access the true impact of the factors and if they have an influence on the tragedy these strikes pose.

Ultimately, this report hopes to shed light on the effectiveness, or rather ineffectiveness, of the U.S. drone strikes and the danger they pose to innocent civilians. It may enlighten readers as to how these attacks have the opposite effect, rather than reducing terrorism, it further increases support for such causes through human and financial support. The U.S. much seek different ways to resolve such conflicts, and we seek to build upon this belief with data exploration and methods that will soon be discussed.

Literature Review - "No boots on the ground": the effectiveness of US drones against Al Qaeda in the Arabian Peninsula by Thomas Bolland and Jan Andre Lee Ludvigsen

This piece of literature conducts a study on if drones are making a difference in stopping terrorism and goes in depth on the effectiveness of drone strikes. Their response variable was the ability of the terrorist organization to conduct attacks on the U.S. and Western-Europe, while the predictor variables are hierarchical structure, qualified human resources, and key material resources. They then discuss what allows a terrorist organization to function, such as financial resources and weapons, then highlight how the U.S. drone strikes should intervene with these resources if effective. However, the study finds these methods to be ineffective and counterintuitive. Not only do these drone strikes not eliminate what they hope, they also put innocent civilians in terrible danger and only increase hostility towards the U.S., leading to more terrorist attacks. The drone strikes in Yemen have eliminated some key resources, but, in return, bolster the support towards

operations with human and financial resources. Ultimately, the U.S. drone strikes are ineffective, dangerous, and ultimately do nothing but turn other countries and innocent civilians against the United States.

Bolland and Ludvigsen's work provides a useful framework for understanding the broader consequences of drone strikes, including the effectiveness of drone warfare and its blowbacks. We would like to build on their work by focusing on the civilian casualties. In particular, we want to examine how geographic and confirmation-status factors influence the number of civilian deaths. Geographic factors like population density and the presence of militant groups likely influence the casualty rate. Similarly, the confirmation of US involvement may affect reporting transparency, which usually leads to underreported civilian deaths. Both of these topics have not been analyzed systematically yet. By studying these, our research could provide a lucid picture of the long-term consequences of drone strikes and offer more insights into Bolland and Ludvigsen's work on the backfires of the strikes such as local resentment and radicalization.

Methodology

The data is in xlsx format and examines four countries: Afghanistan, Somalia, Yemen, Pakistan; each country has an individual sheet. On each respective sheet, each observation represents a reported strike, with the variables: strike ID, strike date, US president during that time, location (village, district, province), coordinates, type of attack, number of people killed, number of civilians killed, number of children killed. Each casualty rate is represented by two variables reporting minimum and maximum casualty estimates, effectively creating a casualty range for each observation. We will use the following two methods to explore our data.

Geospatial Heatmap

Provide a geospatial heatmap which clearly represents civilian casualty rates between regions. Construct this heatmap using regional data (district or province) linked with mean civilian casualty to total casualty ratio for each region. This heatmap will identify key regions with high civilian casualty rates, allowing further analysis to identify factors in these regions which correlate to their high casualty rates.

Linear Regression for Number of Casualties

Continuous numeric response variable will be derived by averaging minimum and maximum reported total casualties for each observation, representing total casualties for that observation. A binary variable will be constructed to mark observations with significant variance between minimum and maximum reported casualties, above a specified threshold. A linear regression will then be constructed onto the response

variable from this new binary variable and the existing variables in the dataset, applying variable selection to condense this model.

Variables

Response

1.) Civilian casualties: The goal of our project, to understand how civilian casualties are impacted by the influence of other variables. The death toll of civilians.

Predictor

1.) **Country:** The country being researched for civilian casualties: Afghanistan, Pakistan, Yemen, Somalia.

2.) **Presidency:** The current president at the time of the strikes: Bush, Obama, Trump.

3.) **Date (MM-DD-YYYY):** The date of the attack.

4.) **Latitude:** The latitude of the attack.

5.) **Longitude:** The longitude of the attack.

6.) **Minimum Reported Injured:** The minimum reported number of injuries from the attack.

7.) **Ratio of Civilians to Total Killed:** The proportion of civilians killed to total killed in the attack.

8.) **Total People Killed Reporting Variance:** The variance from different reportings on the total casualties.

9.) **Ratio of Children to Total People Killed:** The proportion of children killed in relation to total killed in the attack.

Potential Impact

Understanding the impact of drone strikes on civilian casualties is crucial for informing policy, military strategy, and public discourse. Our research hopes to provide a data-driven perspective on the unintended consequences of drone warfare, offering insights into how geographic factors and confirmation status influence casualty rates.

By analyzing geospatial trends and reporting biases, our findings could contribute to a more transparent discussion on the effectiveness and ethical considerations of drone operations. This research may also aid

humanitarian organizations and policymakers in assessing the risks associated with drone warfare and developing strategies to mitigate civilian harm.

Furthermore, our study can be used to improve military accountability, potentially influencing operational decisions and transparency standards in reporting drone strike casualties. We want to investigate the humanitarian effects of U.S. drone warfare in these regions. By shedding light on the underreported impact of unconfirmed drone strikes, we hope to provide a foundation for future investigations into the long-term consequences of drone warfare on affected communities.

Exploratory Data Analysis (EDA)

The dataset comprises 1,694 recorded drone strikes, each with 17 attributes detailing the circumstances surrounding the events, including identifiers (`Strike ID`), geographic details (`Country`, `Most Specific Location`, `Latitude`, `Longitude`), temporal context (`Date`), U.S. presidency in office (`President`), casualty estimates (`Minimum` and `Maximum` casualties including civilians and children), injury reports, and whether the strike was officially confirmed by the U.S. government (`US confirmed?`).

Missing Data

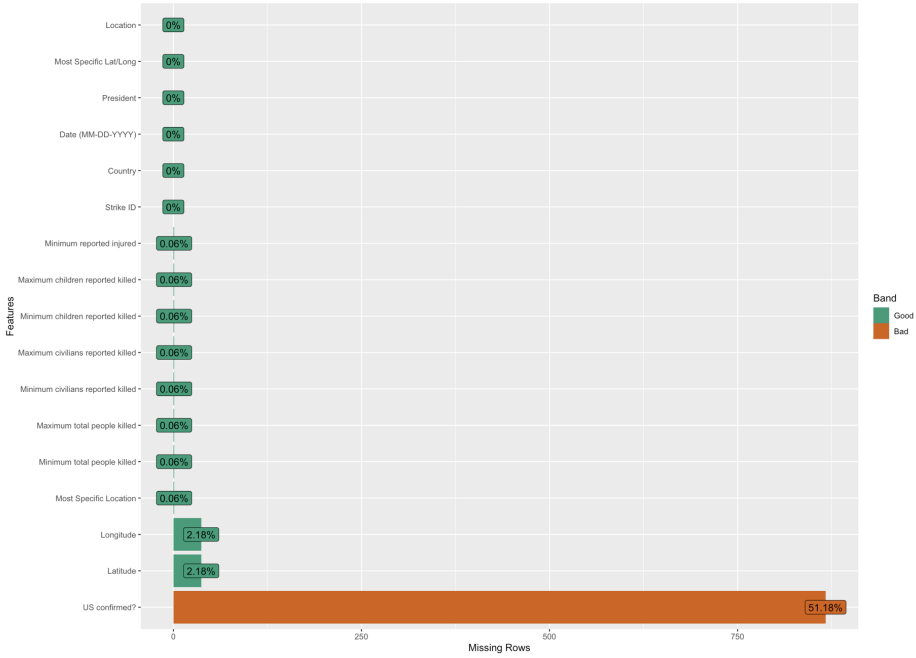

Figure 1: Missing Values

An examination of missing data revealed substantial missingness specifically within the `US confirmed?` variable, accounting for 867 missing observations. Given that confirmation status may itself reflect meaningful

reporting practices or data limitations, simplistic removal or imputation of these records could introduce bias. A deeper analysis of these missing values by presidency identified varying distribution, indicating potential patterns linked to administrative reporting behaviors. Specifically, strikes with missing confirmation status were distributed as follows:

- Bush administration: 59 strikes
- Obama administration: 664 strikes
- Trump administration: 144 strikes

These discrepancies highlight the importance of interpreting findings with an awareness of such administrative biases.

Numerical Variable Distribution

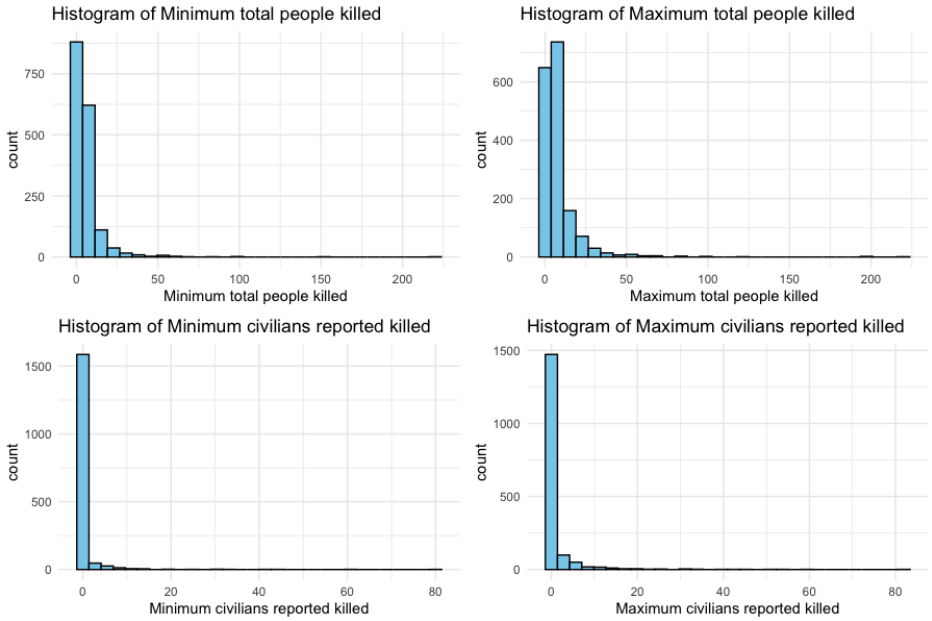

Figure 2: Variable Distribution

Histograms of key numerical variables revealed substantial right skewness. Variables such as `Minimum total people killed`, `Maximum total people killed`, `Minimum civilians reported killed`, and `Maximum civilians reported killed` predominantly featured low casualty counts, yet contained long right tails representing severe, high-casualty incidents. The pronounced skewness suggests future analyses, such as regression modeling, could benefit from transformations (e.g., logarithmic scaling) to normalize these distributions and mitigate the influence of extreme outliers.

Geographic Outlier Identification

A summary examination identified a clear geographic anomaly: a Longitude value reported as 9999, an obvious data-entry error. Addressing such anomalies through correction or exclusion is essential for accurate spatial analyses, particularly for subsequent geospatial mapping and heatmap visualizations.

Insights and Recommendations

The EDA has shown several important considerations. First of all, the data skewness demands careful handling of extreme values and suggests the potential need for data transformation or standardization to achieve robust and reliable analyses for the regression. Additionally, the significant amount of missing data regarding confirmation status highlights the importance of employing specialized analytic methods and extra background research, as straightforward exclusion or imputation could lead to misleading conclusions.

Furthermore, geographic data anomalies identified during the analysis requires exclusion to ensure accurate and meaningful spatial analyses. Finally, the distribution of drone strikes across different presidential administrations should inform interpretations, particularly when comparing administrative practices and outcomes.

Results and Interpretation

Geospatial Heatmaps

Overall Map

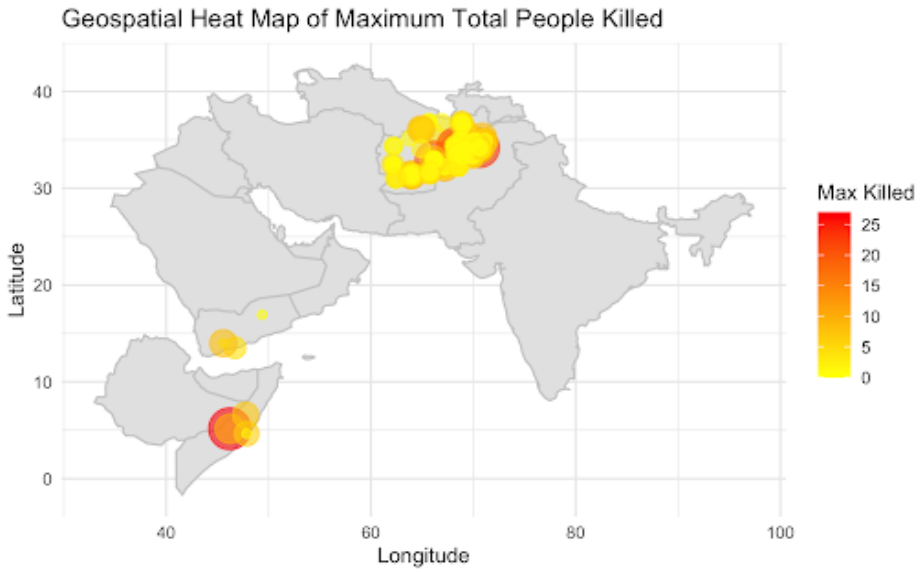

Figure 3: Strike Distribution on Map

Geographical Hotspots: High-casualty strikes concentrated in specific regions (Afghanistan-Pakistan border, Yemen, Somalia).

Operational Intensity: Maximum casualties per strike reaches 25, suggesting large-scale engagements.

Humanitarian and Ethical Concerns:

- High casualties raise concerns about civilian impact and potential backlash.
- There are ethical dilemmas in balancing military objectives with minimizing collateral damage.

Policy Implications: Informs officials to reduce humanitarian impact and emphasizes the need for transparency and accountability in drone strike reporting.

Afghanistan

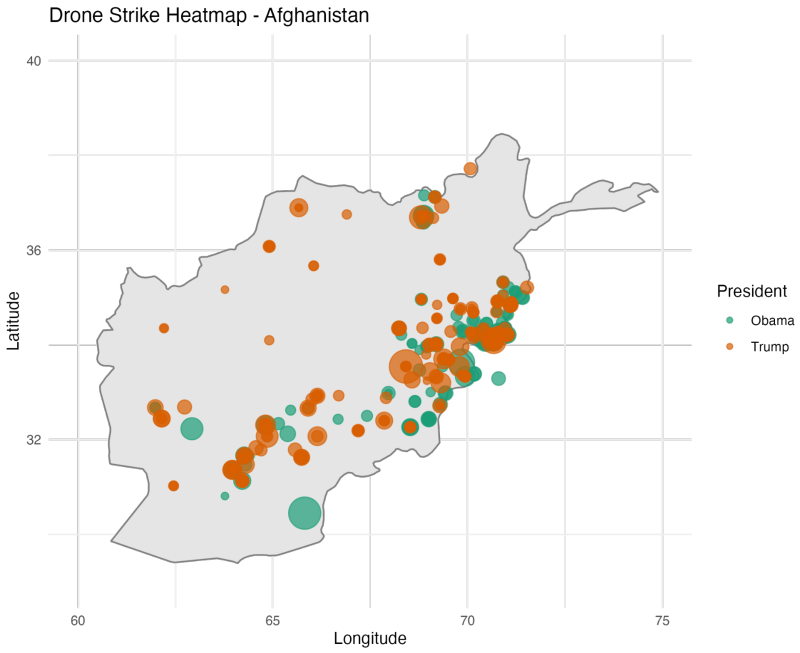

Figure 4: Afghanistan Strikes Map

Overall data:

- Consistent high amounts of drone strikes, indicating high military activity
- Drone strikes are overall concentrated in the east side of Afghanistan (border with north-west Pakistan)

Geographical Patterns:

- Obama in office: Drone strikes mainly in east and south-east Afghanistan

- Trump in office: Drone strikes area extends to west and north → indicating greater scale of military activity and US sanction in these areas
- Data indicates the ongoing conflict between Afghanistan and Pakistan

Pakistan

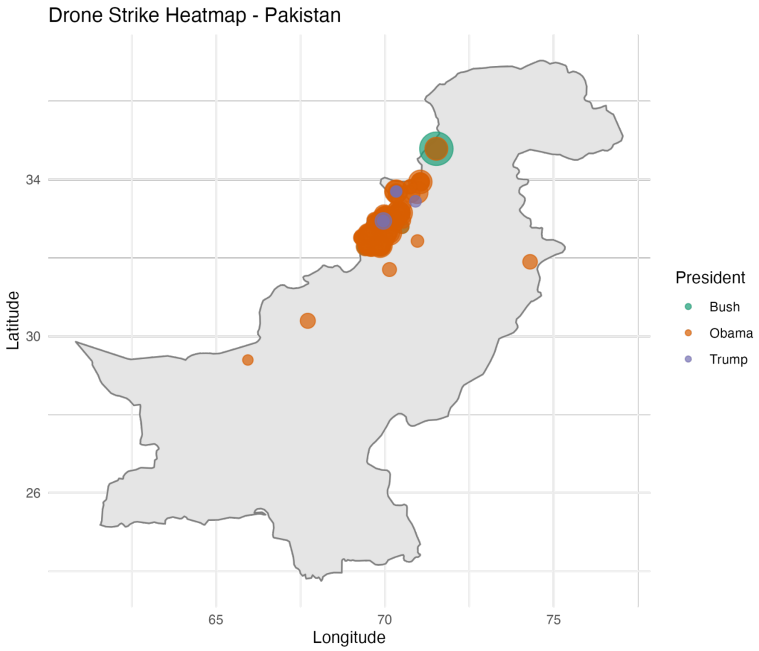

Figure 5: Pakistan Strikes Map

Overall data:

- Consistent high amounts and highly concentrated drone strikes, indicating high military activity
- Drone strikes are overall concentrated in the north-west of Pakistan (border with east Afghanistan)

Geographical Patterns:

- Bush in office: One major strike in upper northwest Pakistan
- Obama in office: High amounts of drone strikes in north-west Pakistan
- Trump in office: Less drone strikes → striking area shifts towards Afghanistan
- Data indicates the ongoing conflict between Afghanistan and Pakistan

Yemen

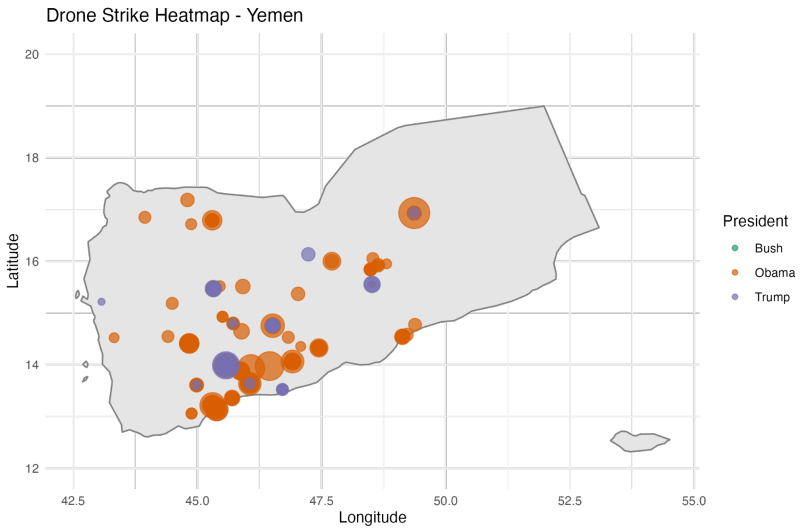

Figure 6: Yemen Strikes Map

Overall data:

- Drone strikes are relatively sparse and overall scattered across the south and middle of Somalia

Geographical Patterns:

- Bush in office: Little amounts of drone strikes
- Obama in office: Drone strikes focuses on middle and southern Somalia
- Trump in office: Drone strikes amounts increase and expand further north and south
- Data indicates the ongoing civil war conflict within Somalia

Somalia

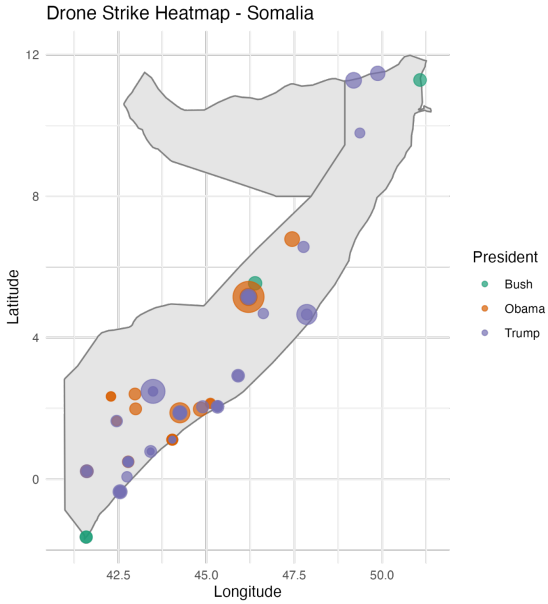

Figure 7: Somalia Strikes Map

Overall data:

- Consistent high amounts of drone strikes, indicating high military activity
- Drone strikes are overall concentrated in the southern and western parts of Yemen

Geographical Patterns:

- Bush in office: No strikes
- Obama in office: High amounts of drone strikes in middle, south, and west Yemen
- Trump in office: Less drone strikes, still focused on southern regions
- Data indicates the ongoing civil war conflict within Yemen

Linear Regression

Formula:

Estimated Casualties ~ Country + Date + President + Latitude + Longitude + Minimum reported injured + Ratio of Civilians to Total Killed + Ratio of children to Total Killed + Total People Killed Reporting Variance.

Used with a stepwise AIC with AIC=7715.54

R^2: 0.2736

We have a reported R^2 value of 0.2736, meaning that 27.36% of the variance in estimated casualties is explained by our linear regression model. This is not a necessarily high percentage, and requires some future insight, but it works fine in our context. We will use this to make conclusions with our data and research.

P-Value: < 2.2e-16

This is the P-Value of our model, and it is extremely small, which is a good sign. This means that our model is significant, and that we are able to make conclusions using our linear regression model. We can reject the standard null that our variables have no effect on estimated casualties and conclude that our predictors do have influence.

Significant Variables

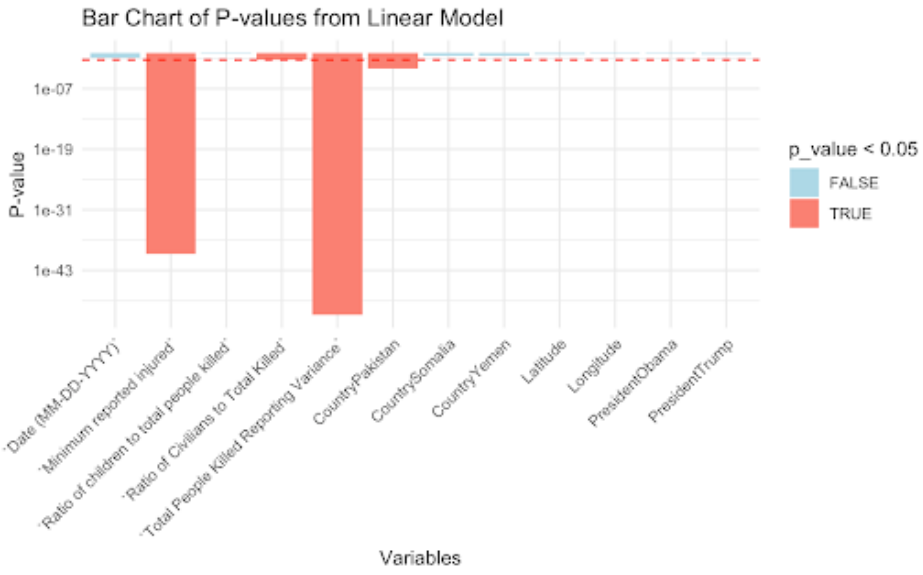

Figure 8: P-Values Plot

1.) Minimum Reported Injured (Coefficient: 0.9855377)

2.) Ratio of Civilians to Total Killed (Coefficient: 1.5157317)

3.) Total People Killed Reporting Variance (Coefficient: 0.0087484)

4.) Country: Pakistan (Coefficient: -3.3524959)

These four variables are the most influential on our linear regression model. The minimum reported

injured, ratio of civilian casualties to total casualties, and the reporting variance all have a positive-trending influence on the estimated casualties of an attack. These three play the biggest role in what determines the death toll and what is, at the end of the day, estimated. However, if these attacks took place in Pakistan, this actually lowers the estimated casualties, indicating that the current administration may have been more careful in this country, or the geographic location/terrain provided more cover for civilians. Either way, these four variables had the largest impact on the estimated casualties of an attack and provide the most insight into how the drone strikes are going wrong.

Diagnostics

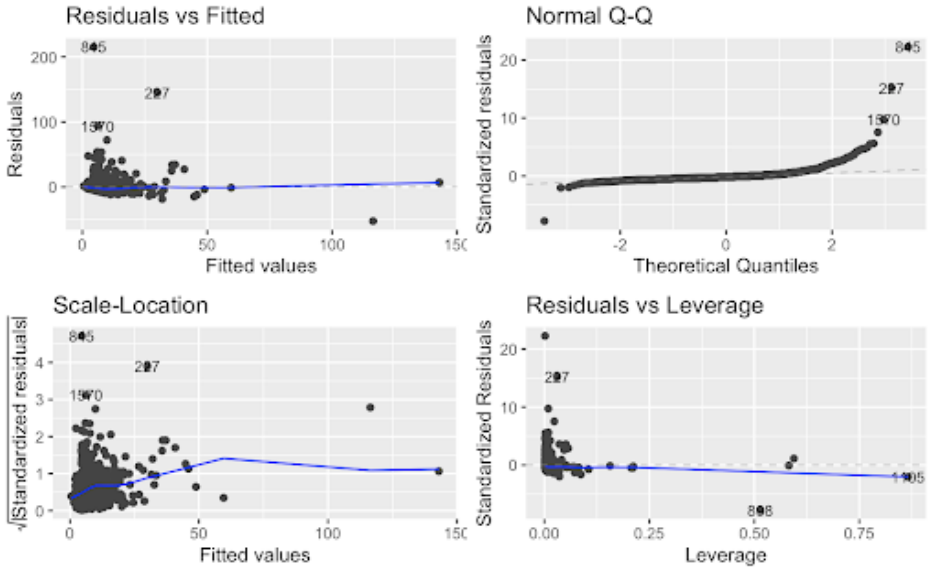

Figure 9: Diagnostics Plots

From the plot above, we can see that the four assumptions of a linear regression model–linearity, independence, constant variance, and normality–are all met. Most of the data fall on the diagonal line of the normal Q-Q plot, with some outliers trailing off at the end. The other plots also showcase the presence of outliers. Ultimately, upon zooming into our diagnostic plots, we can determine there is proper scatter about the line and that our model is valid. We are now able to make conclusions with our model, data, and research.

Discussion and Conclusion

U.S. drone strikes are reckless, dangerous, and, at the end of the day, completely counterintuitive to its goal. These drone strikes are pursued with the goal of reducing terrorism, but when they kill innocent civilians with every strike, they end up fostering more support for the terrorist organizations than ever

before. Our research and models highlight this fact–casualties are far too high and influenced by U.S. attacks to simply be ignored. The United States government must take action to reduce the death toll and cherish the innocent lives being taken, only then will their objective even begin to be fulfilled.

We found the most significant variables that impact casualty rates to be the minimum number of reported injuries, the ratio of civilians to total killed, the reporting variance in total deaths, and the country. From our overall heatmap, we can see that there is a very significant difference in the casualties in each country with the highest amount of casualties occurring in Afghanistan and Pakistan. Despite, Pakistan having a negative coefficient, they still had a high number of casualties. The four variables are the most useful in predicting the total number of deaths and give greater insight into the mistakes the U.S. continues to make. From our linear regression model, we were able to determine these variables and see that, despite only have an R^2 of 0.2736, our P-value $<$ 2.2e-16 allows us to determine that our model is significant and accurate in determining the estimated casualties. We were able to sue these two methods to make our final statements. From our findings, we can determine that the drone strikes do have a significant impact on the lives of civilians. With more US.S drone strikes, the reported casualties increased, and, because of this, we can conclude that these attacks do have significant consequences on the civilian death toll, leading to not only further support of terrorism, but the tragic destruction of families and innocent bystanders.

Before delving further into our final statements, we want to acknowledge our limitations and give some final advice. First of all, we only used one model, linear regression, to analyze our data. The geospatial heatmaps were more of a supplemental tool. With more time, we would recommend using K nearest neighbors (KNN), logistic regression, or even random forest to gain further insights into this dataset. These might fit the structure of the data better and perhaps even determine if other variables have more impact or high correlation with the other variables. By using only one model, we limited ourselves to whatever that model concludes, we could have used more flexibility. We also had a relatively small R^2 that, although it fit in our context, still could have been a lot better. Generally speaking, the variance was not explained much, which is something we advise others to improve on in the future. Lastly, we had some missing values that may have provided more clarity for us and our model. In the future, if one can account for this missing data, maybe using averages or locating the data, that can do nothing but improve the results. Lastly, we advise looking into the strike confirmation status by the United States. This would be an incredibly useful variable to explore and truly underscore the impact the U.S. has had on these countries. By taking into account our limitations and advice, we hope that future endeavors with this dataset or objective can be much more successful and improve upon our results.

To answer our research questions, we ultimately did find that geographic factors, such as country, latitude,

and longitude, have a significant impact on casualty rates. Alongside this, presidential leadership, geographic location, and casualty demographics can help predict the estimated number of casualties in conflict-related events. We hope both of these conclusions and answered questions can inform humanitarian efforts and help war-related policy makers in their crisis management, ultimately reducing innocent civilian deaths. There is a correlation between the number of casualties and drone strikes, and the United States must recognize this rather than just viewing it as another statistic. These are people with families, friends, and lives to live, and they must be treated as such. These drone strikes are ineffective in minimizing collateral damage, especially in certain geographic locations. The United States government has a lot of power–power that has destroyed innocent lives. We hope this report gave greater insight to the issue at hand and allows people to see the tragedy taking place thousands of miles away, only when this is acknowledged can change truly take place and lives can be saved.

References

Bolland, T., & Lee Ludvigsen, J. A. (2018). "No boots on the ground": The effectiveness of US drones against al qaeda in the arabian peninsula. *Defense & Security Analysis*, *34*(2), 127–143. https://doi.org/10.1080/14751798.2018.1478184

Drone Wars. (n.d.). *Data / critique. (DRONE WARS | data) [data set].* https://dronewars.github.io/data/.