

---

# The Tragic Impact of Drone Strikes on Civilian Casualties

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



STATS 140XP

Lecture 2

Group 19

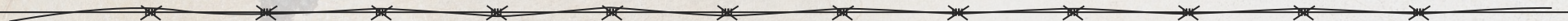
Final Team 19: Agastya Rao, Brandon Erickson,  
Ivy Le, Kuanyi Chen, Kylie Linden,  
Michael Gureghian, Zhiyuan Zhou





# Introduction

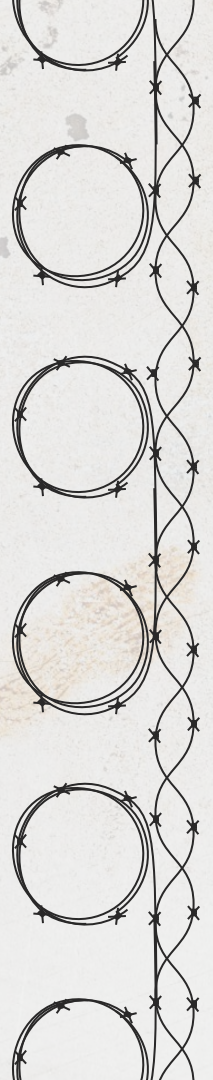
- U.S. military operations have utilized drone warfare in counterterrorism efforts across Afghanistan, Pakistan, Somalia, and Yemen
- Drone strikes are described as precision attacks to minimize collateral damage
- However, the impact, especially on the lives of civilians, remains a debate
- The *Drone Wars* dataset, containing data from The Bureau of Investigative Journalism, provides data on drone-enabled airstrikes, their geographic distribution, their casualty rate, and their confirmation status
- This project aims to explore the relationship between geographic location and the number of drone strike deaths
- Specifically, whether certain regions have higher civilian casualty rates and what factors, like population density or conflict intensity, impact these rates
- We will utilize statistical analysis and data visualization to provide a data-driven perspective on the impact of drone warfare





# “No boots on the ground”: the effectiveness of US drones against Al Qaeda in the Arabian Peninsula

- 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
- It discusses what allows terrorist organizations to function and highlights how U.S. drone strikes should intervene with their resources
- The study finds these methods ineffective and counterintuitive: the drone strikes only put innocent civilians in danger and increase hostility towards the U.S.
- We would like to build on their work by focusing on the civilian casualties
- We want to examine how different factors (geographic, presidential term, etc.) influence the number of civilian deaths
- Our research could provide a 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






# Research Questions

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.
- 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.

---



To what extent do presidential leadership, geographic location, and casualty demographics predict the estimated number of casualties?

- **Objective:** We want to identify key predictors of casualty estimates, which could help inform humanitarian efforts and policy-making related to war and crisis management.
  - **Key Variables:** (1) **Political Influence** (Presidency) (2) **Geographic Location** (Longitude and Latitude)  
(3) **Casualty Demographics** (Civilian and child fatality ratios) (4) **Injury Data** (Minimum reported injured)  
(5) **Reporting Variance** (Variance of total people killed of different drone strikes)
  - **Methodology:** Use **linear regression** to assess the significance and predictive power of each variable.
- 



# Variables of Interest

Based on what is sufficient for our methods, we conducted our research using the following variables

COUNTRY	01	Country - Afghanistan	08	Date (MM-DD-YYYY)
	02	Country - Pakistan	09	Latitude
	03	Country - Yemen	10	Longitude
	04	Country - Somalia	11	Minimum Reported Injured
PRESIDENCY	05	President - Bush	12	Ratio of Civilians to Total Killed
	06	President - Obama	13	Total People Killed Reporting Variance
	07	President - Trump	14	Ratio of Children to Total People Killed

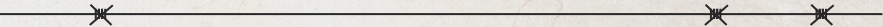


# Methods



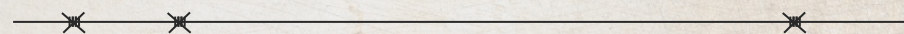

## I. *Geospatial heatmap*

Provide a geospatial heatmap by country, which clearly represents the **maximum number of people killed** in different regions. This heatmap will identify key regions with high civilian casualty rates.



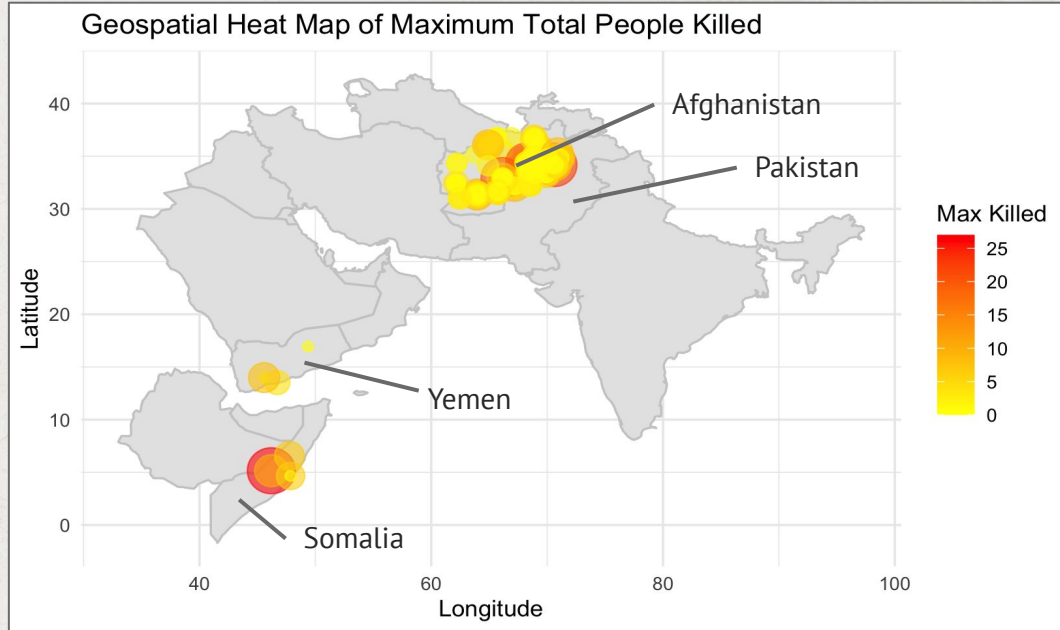
## II. *Linear regression for number of casualties*

Continuous numeric response variable will be derived by averaging minimum and maximum reported total casualties for each observation, representing total casualties for that observation. We will regress the total casualties on multiple predictor variables and find their significance in predicting casualties.



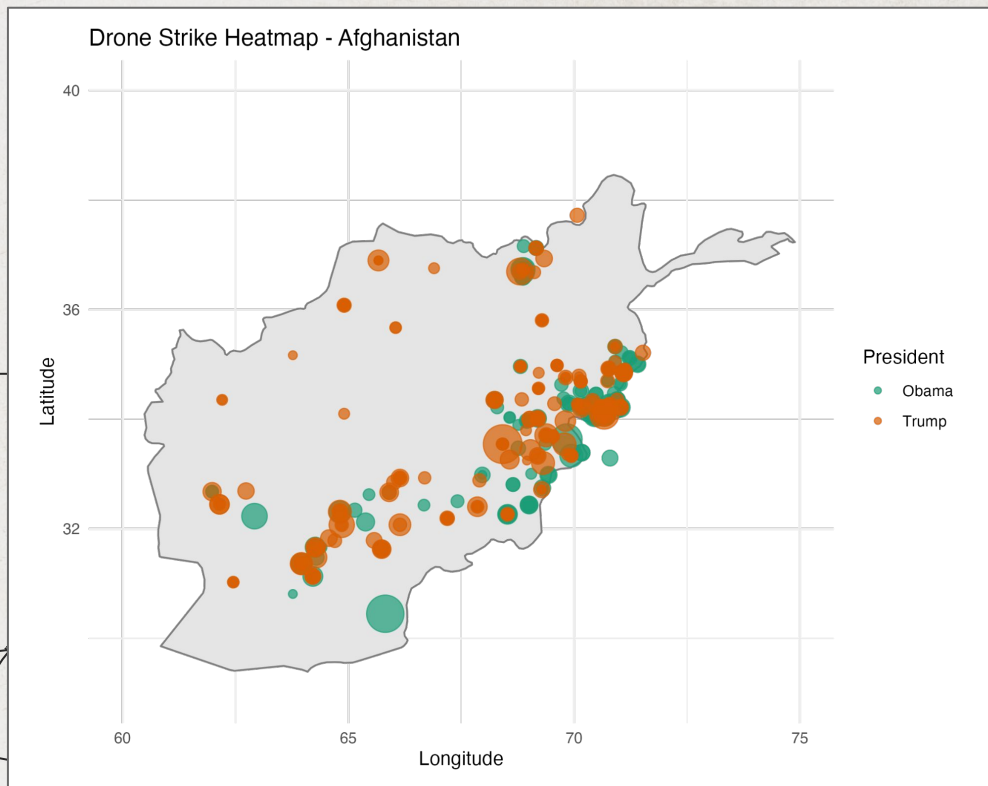


# Overall Map



- **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



- **Overall data:**

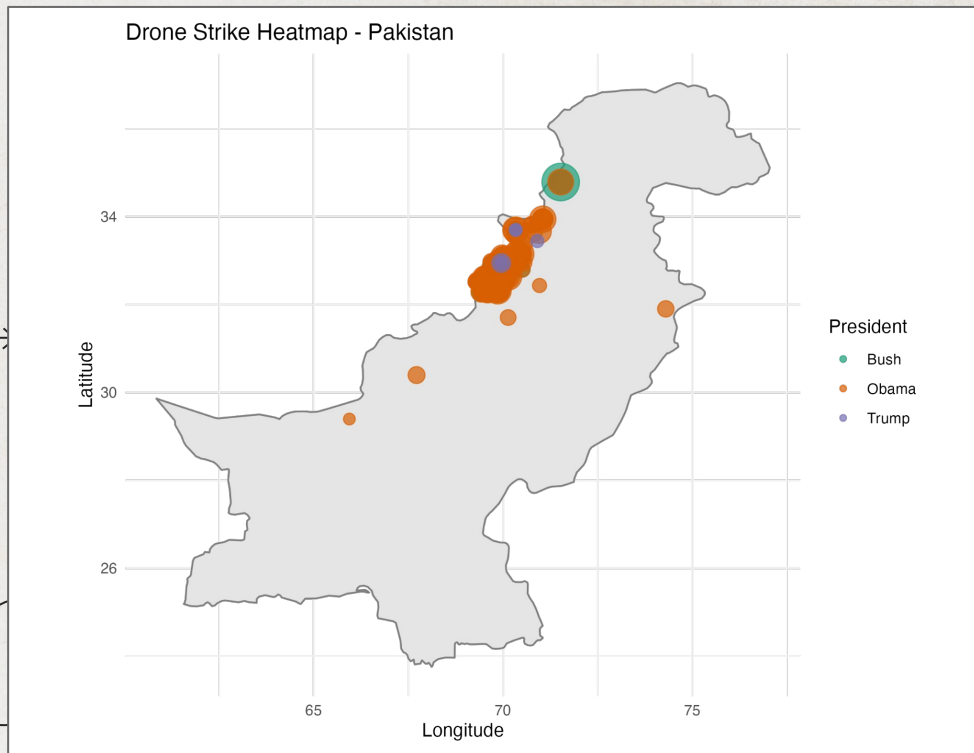
- Consistent high amounts of drone strikes, indicating high military activity
- Drone strikes are overall concentrated in the east side of Afghanistan (border with north-west Pakistan)

- **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
- These data indicate the ongoing conflict between Afghanistan and Pakistan



# Pakistan



- **Overall data:**

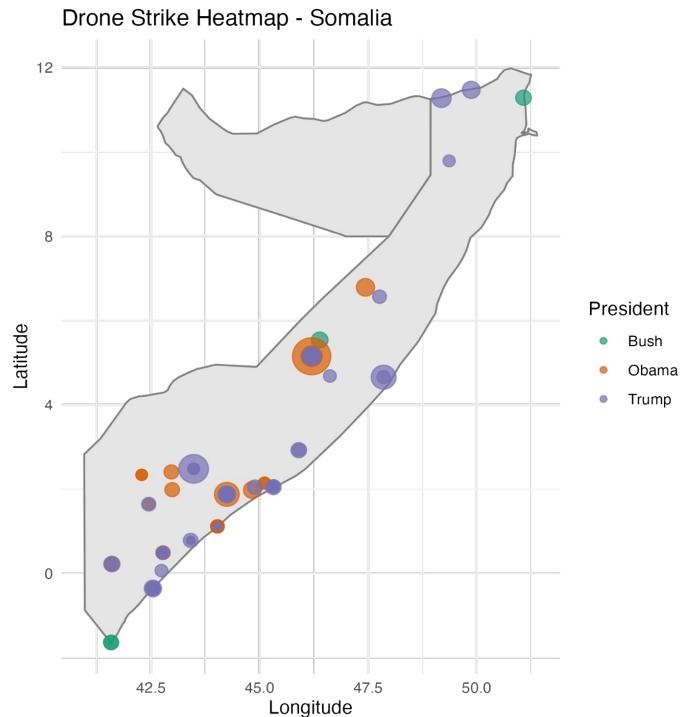
- Consistent high amounts and highly concentrated drone strikes, indicating high military activity
- Drone strikes are overall concentrated in the north-west of Pakistan (border with east Afghanistan)

- **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
- These data indicate the ongoing conflict between Afghanistan and Pakistan



# Somalia



- **Overall data:**

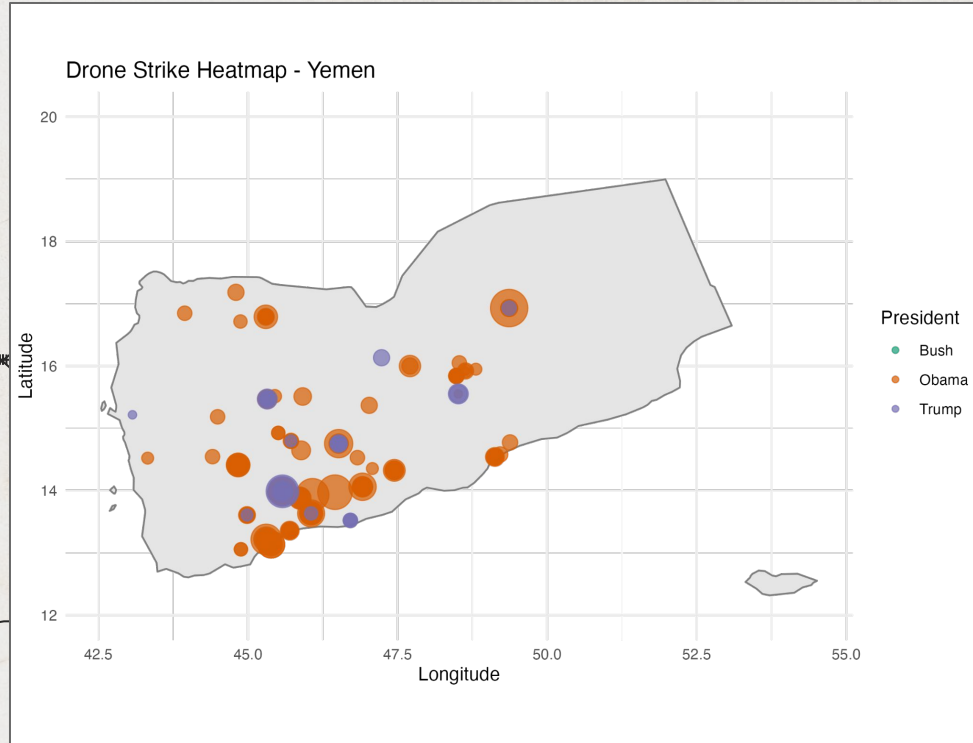
- Drone strikes are relatively sparse and overall scattered across the south and middle of Somalia

- **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
- These data indicate the ongoing civil war conflict within Somalia



# Yemen



- **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
- These data indicate the ongoing civil war conflict within Yemen



# Linear Regression

## I. 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.**
  - Using stepwise AIC with AIC=7715.54

## II. $R^2$ & P-Value

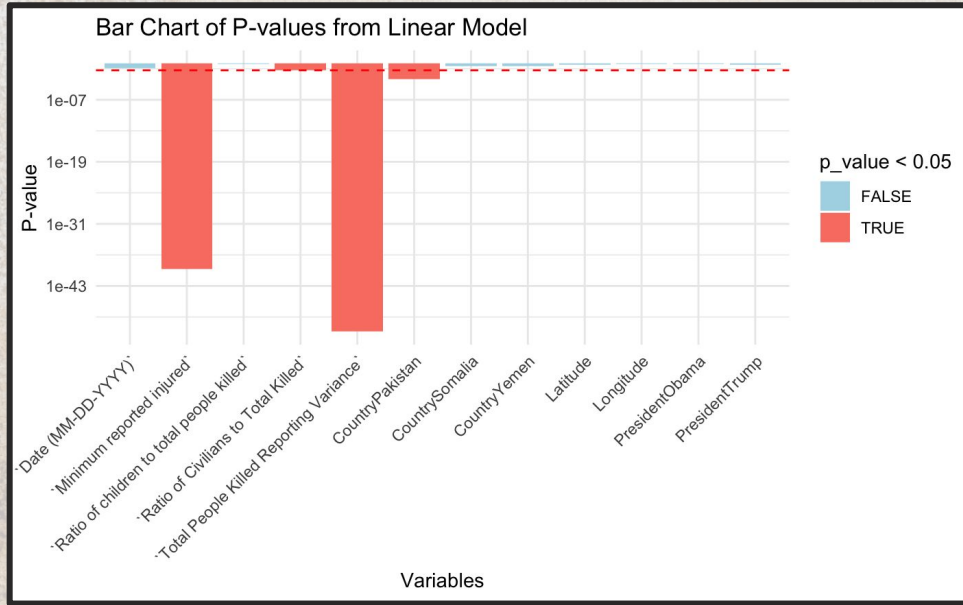
- **Multiple  $R^2$ : 0.2787**
- **Adjusted  $R^2$ : 0.2736**
  - With an  $R^2$  value of 0.27, this means only 27% of the variance in the data is explained by our linear regression model. However, considering the context of the variety of types of variables on different scales, this is a relatively good value.

**P-value:  $< 2.2e-16$**

- This is an extremely small p-value, meaning that our model is significant and we can make conclusions about our data using our model.



# Linear Regression - Significant Variables



BARPLOT OF P-VALUES SHOWING SIGNIFICANT VARIABLES

## Variable I

Minimum  
Reported Injured  
(0.9855377)

The minimum  
reported casualties

## Variable III

Total People Killed  
Reporting  
Variance  
(0.0087484)

Variance in the  
total casualties  
reported

## Variable II

Ratio of Civilians to  
Total Killed  
(1.5157317)

Civilian casualties  
compared to total  
casualties

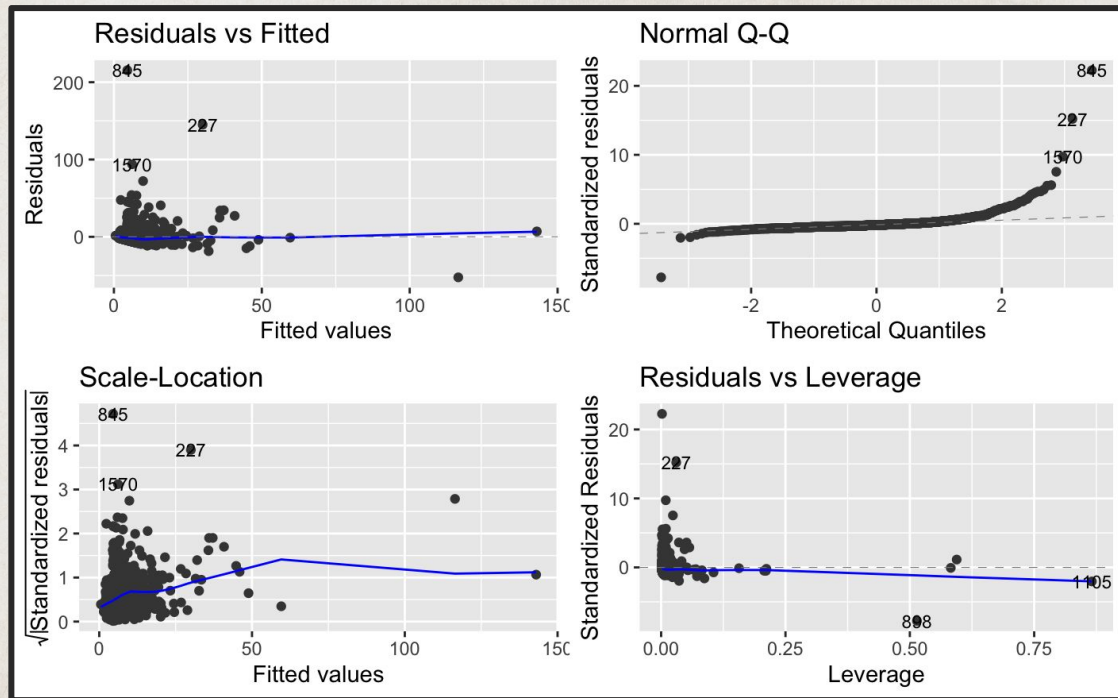
## Variable IV

Country:  
Pakistan  
(-3.3524959)

Country with a  
significant difference  
in casualties



# Diagnostics



DIAGNOSTIC PLOTS FOR LINEAR REGRESSION MODEL

## Validity

### Assumptions Correct

Based on our diagnostics, we can assume the four assumptions are met, especially normality. There is presence of some outliers.

## Explanation

### Plot Perspective

When one zooms into these plots, there is scatter around the line, meaning the assumptions are met.



# Summary of Results

---

## Findings

We found the most significant variables that impact casualty rates to be **(1) minimum reported injured, (2) ratio of civilians to total killed, (3) total people killed reporting variance, and (4) 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**.

---

## Impact

From our findings, we can determine that the drone strikes do have a significant impact on the lives of civilians. **With more drone strikes, there were higher casualties reported**. 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.



# Closing Statements

## Conclusions

### Limitations

- We only created one model to analyze the data
- We had a relatively small  $R^2$  value, and although it fit in context, it is still relatively weak
- There were a few missing values that may provide more clarity

### Advice

- With more time, we could try other models that may be better suited for our data
- Logistic regression may be an avenue to explore, as well as KNN
- The strike confirmation status by the U.S. could be an insightful variable to explore

### Results

- We did find geographic factors, such as country, latitude, and longitude, to have a significant impact on casualty rates
- There is a correlation between the number of casualties and drone strikes
- U.S. drone strikes are ineffective in minimizing collateral damage





# Citations

Bolland, Thomas, and Jan Andre Lee Ludvigsen. "'No Boots on the Ground': the Effectiveness of US Drones against Al Qaeda in the Arabian Peninsula." *Defense & Security Analysis*, vol. 34, no. 2, Mar. 2018, pp. 127–143., [doi:10.1080/14751798.2018.1478184](https://doi.org/10.1080/14751798.2018.1478184).

*Data / Critique*. DRONE WARS | Data. (n.d.). <https://dronewars.github.io/data/>



---

STATS 140XP  
*Lecture 2*  
*Group 19*

# Thank You

Please let us know if you have any questions.

