Air raids on the UK: understanding the effect of time on casualties*

How nighttime can influence the destructiveness of warfare

Dennis Netchitailo

December 2, 2024

With wars raging across the world, notably, between Russia and Ukraine, and aerial bombardment using aircraft or missiles a major cause of civilian casualties far from the frontlines, understanding, based on historical data, whether either nighttime or daytime attacks are more damaging would allow policymakers to better allocate resources when dealing with the treat of aerial attack. This study models the German bombing campaign of Britain during the Second World War by analyzing the data collected by the Home Security War Room from 1939 to 1945 that summarised air raid damage. I use a multiple linear regression model to estimate the impact of the time an attack occurred on the resulting casualties.

1 Introduction

On September 1st, 1939, the German armed forces crossed the Polish border. The United Kingdom declared war on Germany on September 3rd, 1939. The Ministry of Home Security was the British government department established in 1939 to direct the response to German air raids. Part of its responsibility was tracking data related to air raids. This data found more than 66 thousand people and 221 thousand people had been killed or injured by Germany's air campaign against the UK. Approximately 0.7% of the UK's 1939 population of 41.46 million would be killed or injured by air raids between 1939 to 1945. With 32 000 incidents, there were an average of 10 casualties per incident, including 2 killed and 8 injured. Yet, there is great variation. Some incidents resulted in hundreds of casualties. While many others resulted in zero casualties. Understanding what factors lead to such drastic differences, often ascribed to luck, is key.

*Code and data are available at: https://github.com/dennisnetchitailo/STA304P3.

The estimand is the number of casualties sustained during an attack. I use the Home Security data with a multiple linear regression model to estimate the difference in casualties during nighttime air raids compared to daytime air raids.

This paper's goal is to try to develop a model that predicts casualties based on the time of an air raid by controlling for multiple variables: country, military region, time, date, and the number of incidents occurring.

The analysis of this data can help countries facing war, especially as it relates to missile or airstrikes, better understand how to allocate resources to deal with casualties.

The paper is structured as follows: Section 2 discusses the data and data cleaning process, followed by the analysis of key variables and relationships in **?@sec-model**. **?@sec-results** then presents the results from the model, followed by a discussion of the implications of these findings in **?@sec-discussion**. **?@sec-discussion** also concludes with suggestions for future research and practical applications of the model.

2 Data

2.1 Overview

This study uses the (Cookson 2020) dataset to study the relationship between variables such as time to casualties.

I used the statistical programming language R (R Core Team 2023) for data cleaning and analysis. The R packages used were **usethis** for file paths (Wickham and RStudio 2023), `tidyverse` for data manipulation and visualization (Hadley Wickham et al. 2023), `here` for file path management (Müller 2023), , `arrow` for data storage and processing (Richardson et al. 2023), `testthat` for unit testing (Wickham 2023), `readr` for reading structured data (Wickham, Hester, et al. 2023), `dplyr` for data manipulation (Wickham, François, et al. 2023), , `ggplot2` for creating visualizations (Wickham, Chang, et al. 2023), `brms` for modeling (Bürkner 2023), and `modelsummary` for summarizing models and results (David Robinson 2023).

2.2 Measurement

The measurement process refers to the attempt to explain real world events through data, such as an individual act of warfare, as represented by the use of a self-propelled (missile) or gravity released explosive (bomb) to strike land or sea target, into numerical entries describing this event. Each entry in the dataset represents a bombing incident that took place in the UK during WW2.

Country: The country where an incident occurred. This includes England, Scotland, Wales, Northern Ireland, and the Channel Islands. **Civil Defense Region**: A region defined by the military for purposes of organization. The United Kingdom was divided into 13 regions. **Coordinates: Longitude and Latitude**: The location of an incident as recorded by the mercator projection.
Date: Year & Month: The year & month an incident occurred at a specific location. **Time** The time of day an incident occurred. Day is defined as being between 6AM and 6PM. Night is defined as being between 6PM and 6AM. **Casualties: Killed, Injured, Total**: The number of people killed or injured due to an incident. Total casualties is the sum of the number killed and injured in the same incident. **Lethality**: the proportion of casualties that were killed.

2.3 Data Cleaning

The two datasets that were used to make the single dataset that provided the data for the analysis were cleaned to ensure accuracy and consistency. Missing values were dealt with by removing them from the analysis. Variables were converted into logical data types. Unneeded columns were removed. The two cleaned datasets were unified into one.

2.4 Outcome variables

The outcome variable is `total_casualties`. This is the primary dependent variable that the model is designed to predict. It represents the number of people killed or injured by air raids. The model seeks to identify the variables affecting this number.

Figure 1 visualizes the distribution of casualties from 1939 to 1945 on a monthly basis in a histogram.

A large number of casualties occurred in the last months of 1940 and the first months of 1941. This coincides with "The Blitz", the name given to the most intense period of air raids on the UK, following the defeat of France in June 1940. The number of casualties rises month to month following the defeat of the Allies in western Europe, as Germany could concentrate its airforce on bombing the United Kingdom. Based on this graph, a seasonal trend is not observed.

Figure 2 reflects the events going on in Europe at the time. In 1939, Germany occupied Poland. In 1940, it turned its attention to the west, and coinciding with the invasion of France, Germany's mass air raids against the UK began. This continued into 1941. In 1941, Germany invaded the Soviet Union, redirecting Germany's resources to the east. The renewed attention in 1944 coincides with the allied invasion of Normandy. Finally, in 1945, Germany surrended in May.

Figure 3 reflects the events going on in Europe at the time. In 1939, Germany occupied Poland. In 1940, it turned its attention to the west, and coinciding with the invasion of

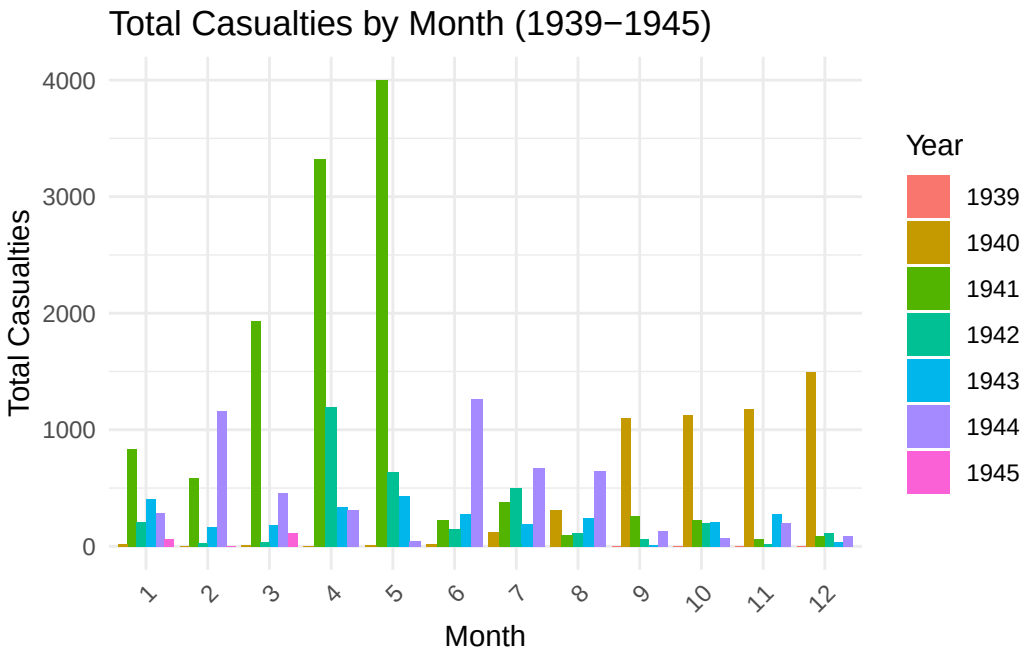

Figure 1: Distribution of casualties in the UK caused by air raids throughout the war.

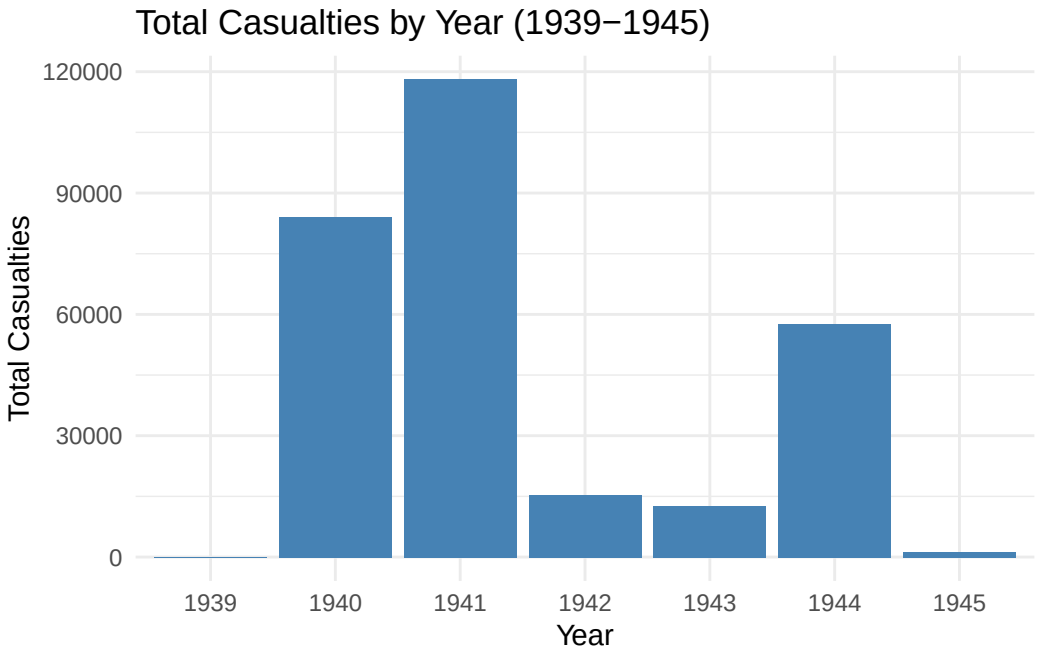

Figure 2: Distribution of casualties in the UK caused by air raids throughout the war.

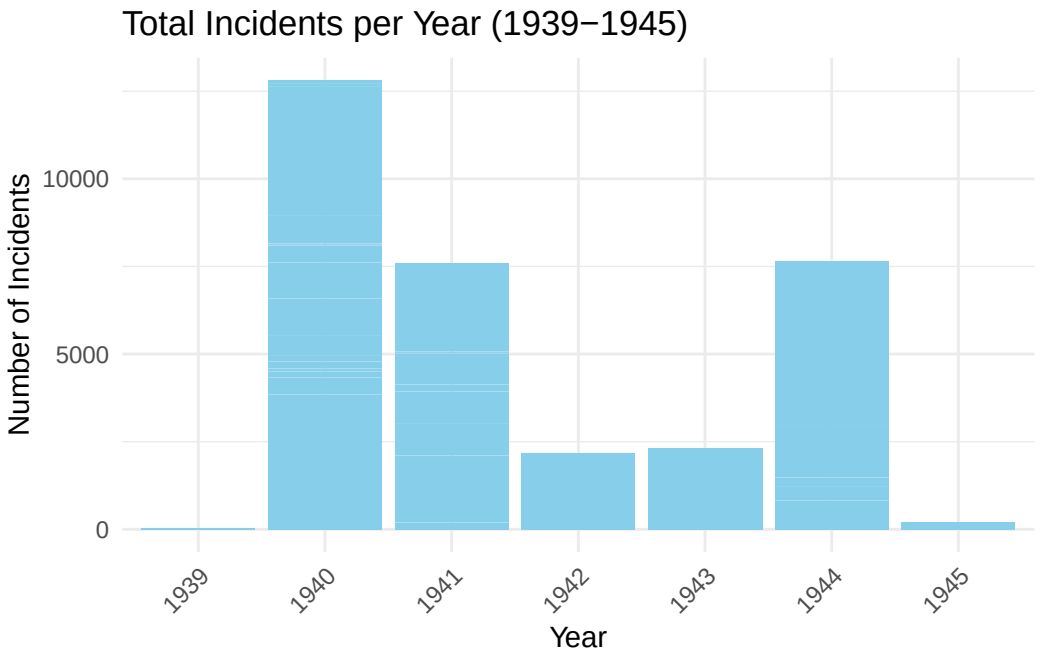

Figure 3: Distribution of casualties in the UK caused by air raids throughout the war.

France, Germany's mass air raids against the UK began. This continued into 1941. In 1941, Germany invaded the Soviet Union, redirecting Germany's resources to the east. The renewed attention in 1944 coincides with the allied invasion of Normandy. Finally, in 1945, Germany surrended in May. As Figure 2 showed, though there were more casualties in 1941 than in 1940, there were actually more incidents in 1940.

2.5 Predictor variables

The **predictor variables** are the factors that are assumed to influence the number of casualties.

1. **Time ('time_binary')**. This variable represents whether an incident occurred during daytime, defined as between 6AM and 6PM, or nighttime, defined as between 6PM and 6AM. The assumption here is that the response time of medical, military, and police services differs at night, given more people at rest (implying worse response time), but also less traffic (implying faster response time). This is the key variable being assessed.
2. **Country ('country')** is a factor, because each country's cultural, social, and economic conditions can influence how much damage an air raid has by influencing factors such as the the abundance of underground shelter (such as the London Underground) or the quality of building construction (making a structure more or less susceptible to collapse).

3. **Date ('year') and ('month')** is a factor because over time, technology and capabilities changed for both sides in the war. Equally, seasonal effects may have played a role. For instance, in wintertime, it is possible that there could have been cases of frostbite caused by the destruction of shelter and heating infrastructure.
4. **Civil Defense Region ('civil_defense_region')**: is a factor because different military regions differed in their capabilities to prevent and to respond to air raids, as well as in intensity of targeting by the German airforce.
5. **Casualties ('killed'), ('injured'), ('total_casualties'), and ('lethality_category')**: is a factor that assesses the number and types of casualties, as well as how deadly an air raid was.

3 Model

The goal of this build a model using R (R Core Team 2023) to predict the difference in casualties caused in an incident based on several predictor variables.

3.1 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

$$difference\_casualties_i = \beta_0 + s(year_i) + \beta_1 (month_i) + \beta_3 (time\_binary_i) + \beta_4 (lethality\_category_i) + \beta_5 (co$$

I ran the model in R (R Core Team 2023) using the `brsm` package of Bürkner (2023). I used the default priors from `brms`.

Where:

- $difference\_casualties_i$ is the number of casualties for the i-th incident.
- $beta_0$ is the intercept term (baseline value when all predictors are zero).
- $year_i$ is the year the incident occurred in.
- $month_i$ is the month the incident occurred in.
- $time\_binary_i$ is whether the incident occurred during daytime or nighttime.
- $lethality\_category_i$ is how deadly an air raid was.
- $country_i$ is the country an incident occurred in.

- $civil\_defense\_region_i$ is the military region an incident occurred in.
- $1|casualty\_group_i$ is a random intercept term, accounting for variability caused by grouping.
- $1|civil\_defense\_region_i$ is a random intercept term, accounting for variability caused by grouping.
- ϵ_i is the error term, assumed to be normally distributed with a mean of 0 and constant variance $\epsilon_i \sim N(0, \sigma^2)$

3.1.1 Model justification

The zero-inflated negative binomial distribution was chosen because of the nature of the data. First off, the response variable (total_casualties) is a count with non-negative integers. Secondly, due to the variance being larger than the mean, indicating overdispersion, the Poisson distribution would be innaprpriate, due to its assumption of the mean and variance being equal. The negative binomial distribution accounts for this overdispersion by introducing an additional parameter that models the variance separately from the mean, making it a more flexible choice for this data. Thirdly, relating to the large variance, but low mean, there are many zeros in the data. I.e.: many incidents had zero casualties. But a relatively small number of incidents have dozens, hundreds, or thousands of casualties. This is why the model was chosen to be a zero-inflated binomial distribution, to take into account the excess number of zeros.

Other models such as a linear regression were considered, but deemed not to be suited to the characteristics of this dataset: data being a count, over-dispersion, excess zeros.

4 Results

Our results are summarized in **?@tbl-modelresults**.

5 Discussion

5.1 First discussion point

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Appendix

A Idealized Survey

Purpose: better understand the human and economic impact.

Provide details regarding each (sub)question.

Q1 - What was the location of air raid?

(a) Country

(b) Civil Defense region

(c) Longitude and Latitude

(d) Municipality

Q2 - When did it occur?

(a) What was the date of the first enemy strike on this location?

(b) In what hour did this strike occur?

Q3 - What casualties are associated with this strike?

(a) How many people were killed?

(b) If causes of death are known, list them, and the number associated with each cause.

(c) How many people were injured?

(d) If types of injury are known, list them, and the number associated with each cause.

(e) If known, what are the ages of the victims?

(f) If known, what is the sex of the victims?

(g) If known, how many of the casualties were civilians?

(h) If known, how many of the casualties were military personnel?

Q4 - What was the economic impact?

(a) What objects was struck?

(b) If any, what damage was sustained by nearby objects?

B Additional data details

C Model details

C.1 Posterior predictive check

C.2 Diagnostics

References

Bürkner, Paul-Christian. 2023. *Brms: Bayesian Regression Models Using 'Stan'*. https://CRAN.R-project.org/package=brms.

Cookson, Alex. 2020. *Data about the Bombing of Britain in WW2*. https://github.com/tacookson/data/tree/master/britain-bombing-ww2.

David Robinson. 2023. *modelsummary: Summary Tables for Model Results*. https://CRAN.R-project.org/package=modelsummary.

Hadley Wickham, Romain François, Lionel Henry, and Karthik Ram. 2023. *tidyverse: Easily Install and Load the 'Tidyverse'*. R Foundation for Statistical Computing. https://CRAN.R-project.org/package=tidyverse.

Müller, Kirill. 2023. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, et al. 2023. *Arrow: Integration to 'Apache Arrow'*. https://CRAN.R-project.org/package=arrow.

Wickham, Hadley. 2023. *Testthat: Unit Testing for r*. https://cran.r-project.org/web/packages/testthat/index.html.

Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani, Dewey Dunnington, and Teun van den Brand. 2023. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. https://CRAN.R-project.org/package=ggplot2.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, Romain Francois, and Jennifer Bryan. 2023. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Wickham, Hadley, and RStudio. 2023. *Usethis: Automate Package and Project Setup*. https://CRAN.R-project.org/package=usethis.