**Attacks on Healthcare Workers and Infrastructure in the Ongoing Ukraine War**

****

**Kevin Ryan**

**Unleashing Open Data with Python**

**JHU, Spring 2023**

**Prof. Jason Renn**

Table of Contents

Background 3

Research Question 4

Literature Review 4

Methodology for Data Exploration 7

Description of the Data 9

Exploratory Analysis 11

Healthcare Workers 11

Healthcare Facilities 12

Locations of Attacks on Healthcare Workers and Infrastructure 13

Perpetrators of Attacks 14

Weapons used in Attacks 15

Russian Personnel Losses 16

Russian Equipment Losses 17

Bivariate Regression 18

T-Test 20

Correlation 20

Implications of Findings 21

Conclusion 21

Citations 23

Code 24

# Background

One of the hallmarks of the Russia’s ongoing campaign to seize Ukrainian territory has been the deliberate targeting of non-combatants and civilian population centers. While collateral damage is often unavoidable, Russia’s actions in Ukraine represent a clear violation of the Law of Armed Conflict (LOAC). Specifically, Russian attacks on healthcare workers and infrastructure are a violation of Article 19 of the Geneva Conventions, which state that *“medical units… hospitals and mobile medical facilities, may in no circumstances be attacked*.7” After the collapse of the Soviet Union, Russia declared itself the legal successor of the USSR and, therefore, recognized itself as a party to all international treaties signed by the Soviets- including the Geneva Conventions.

These Ukrainian healthcare facilities have been carefully singled out, among other buildings within civilian areas, leading observers to the only logical conclusion- that they are specific targets of the Russian Ministry of Defense (MoD). On the road, civilian medical transports (ambulances) and protected casualty evacuation efforts are attacked on sight by elements of the Russian military. In the occupied towns and villages, Ukrainian healthcare workers are subject to assault, harassment, arrest, and even kidnapping by Russian forces.

Figure - Maternity Ward of a hospital in Mariupol that was targeted by Russian Forces on 09 March 22

As the invasion has thus far gone extremely poorly for Russia, the vast majority of their horrific crimes- the massacre at Bucha, widespread looting of homes, rape, child abduction at filtration camps3, and other atrocities have played out in front of the world lens, for all to see. During this ongoing episode, Russia has suffered extremely significant losses of personnel and equipment. While estimates vary, one data source reports a total of over 170,000 Russians killed in action (KIA) between the initial invasion and 26 March 2023. Equipment losses are also extremely high, leading Russia to fall back on antiquated vehicles and aircraft from Soviet stockpiles.

The lack of concern on the part of the Russians for the documentation and publication of their crimes, combined with the fact that they continue to deliberately conduct attacks against Ukrainian healthcare workers and infrastructure, raises the suspicion that such unlawful attacks might be conducted out of spite, and that a correlation may exist between losses and attacks. It is also possible that in the fog of war either side may accidentally strike an unlawful target.

# Research Question

***Do Russian personnel and equipment losses in the War in Ukraine drive Russian attacks on Ukrainian healthcare workers and infrastructure?***

The ongoing conflict between Russia and Ukraine has resulted in a humanitarian crisis, particularly in the healthcare sector. Russian military attacks on Ukrainian healthcare workers and infrastructure have not only caused significant damage to essential medical facilities but have also led to numerous civilian casualties. As mentioned, these attacks are in clear violation of the Article 19 of the Geneva Convention, which protects medical personnel and facilities during wartime.

Understanding the relationship between Russian battlefield losses and unlawful attacks on healthcare workers and infrastructure is crucial in gaining insights into the motivations behind these attacks and finding ways to prevent future violations. Data exploration is an essential step towards achieving this goal, as it can reveal patterns and trends that may not be immediately apparent. By analyzing relevant datasets and performing statistical analyses, researchers can gain a deeper understanding of the factors that contribute to these attacks and identify potential solutions to mitigate the harm caused to innocents. Therefore, data exploration is critical for advancing our knowledge of the conflict and ultimately promoting peace and stability in the region.

Through exploratory data analysis and visualization, I will examine the relationship between attacks on Ukrainian healthcare workers and infrastructure, including severity of the attack in terms of number of facilities damaged or destroyed, and also workers killed, kidnapped, arrested, injured, or assaulted. By comparing these events to Russian losses in personnel and equipment, I will determine if a relationship exists between the two.

As a potential hypothesis, I will test if Russian attacks on protected targets follow events of high battlefield losses, and if so attempt to quantify how long the typical time lag between losses and attack is. If the severity of the attack on healthcare workers and infrastructure is scaled to the severity of the loss, I will attempt to quantify that as well.

# Literature Review

The literature review consisted of both primary and secondary data collection, including peer-reviewed journal articles, news articles from respected publications, and other information from respected sources. Information from Mil-bloggers and from the popular Telegram channels which follow the conflict were not used due to the level of inherent bias present.

The literature review yielded 14 relevant sources, of which 5 were used. Themes used to perform this downselect included discussion of attacks on civilian targets as well as attacks on healthcare workers and infrastructure, as well as any material linking Russian forces to Article 19 violations.

One particular source proved to be uniquely valuable in researching this topic- *Ambulances Under Siege in Syria*, by C. Hayes Wong and Christine Yen-Ting Chen9. This 2018 paper explores “repeated air strikes on hospitals and ambulances, and the largest death toll of health workers in any recorded conflict.” It further discusses how Ambulances in Syria have been bombed, shot at, stolen, looted, and obstructed- significantly impeding their ability to safely evacuate the wounded and provide medical aid.

While attacks on civilian targets are typically the work of terror organizations rather than professional military units, Russia stands in stark contrast to the rest of the world in this area. The author’s analysis of 204 recorded attacks on ambulances in Syria between 2016 and 2017 revealed the following data:

|  |  |  |
| --- | --- | --- |
| Perpetrator | Number of Attacks | Percentage of Attacks |
| Syrian Regime | 123 | 60% |
| Russian Forces | 60 | 29% |
| Islamic Extremist Groups | 4 | 2% |
| Other Parties | 1 | 0% |
| Unknown | 16 | 8% |
| **TOTAL** | **204** | **100** |

Figure - Table of Attacks on Ambulances in Syria, 2016-2017

While fighting to bolster and preserve the regime of Bashar al Asad, Russian forces conducted dozens of attacks annually against ambulances, accounting for almost 30% of Article 19 violations of this type documented by the Syrian Network for Human Rights (SNHR). It is apparent from this data that while the world turned a blind eye to the conflict in Syria, Russian troops were targeting healthcare workers and infrastructure without consequence. In addition to numerous other examples from previous conflicts of a lack of professional standards and discipline within the Russian military, this particular data represents a direct link between the criminal actions of Russian Forces in Syria and in Ukraine. Many of the same combatants, especially in the early days of the invasion of Ukraine, had participated in the Syrian conflict.

Additionally, *Terrorist Attacks Against Healthcare Facilities: A Review* by Garrett A. Cavaliere, Reem Alfalasi, Gregory N. Jasani, Gregory R. Ciottone, and Benjamin J. Lawner proved to be a valuable source of information5. This research paper discussed how Healthcare facilities play an essential role in response to terrorist attacks, but are also “soft targets” due to their quick accessibility and limited security. The authors obtained data from the Global Terrorism Database (GTD) and conducted a search of terrorist attacks directed against hospitals and healthcare facilities between 1970 and 2018. Of particular relevance to this topic, the authors also found that the most common method of attack was bombing, followed by armed assaults.

|  |  |  |  |
| --- | --- | --- | --- |
| Attack Type | Bombing/ Explosion | 270 | 59% |
| Armed Assault | 77 | 17% |
| Hostage Taking (Kidnapping) | 38 | 8% |
| Assassination | 24 | 5% |
| Facility/ Infrastructure Attack | 16 | 4% |
| Unknown | 14 | 3% |
| Hostage Taking (Barricade Incident) | 11 | 2% |
| Unarmed Assault | 3 | 1% |
| Hijacking | 1 | 0% |
|  | **TOTAL** | **454** | **100** |

Figure - Worldwide Terrorist Attacks against Hospitals, 1970-2019

Another helpful research paper was *Foreign Fighters, Rebel Command Structure, and Civilian Targeting in Civil War*, by Austin C. Doctor and John D. Willingham4. This study contained an analysis of sixty-nine rebel groups active between 1989 and 2015. It asserts that foreign fighters are associated with greater levels of anti-civilian violence only when active in groups with centralized command structures. This was interesting, and contradictory to my own experience in the Iraq War- where loose cellular networks were more likely to attack civilian targets than groups that retained more centralized command structures, who more typically attacked government targets and security forces.

With respect to the Russian way of waging war, I found useful the 2011 paper published by the U.S. Army War College’s Strategic Studies Institute, *“The Russian Military And The Georgia War: Lessons And Implications,”* by Ariel Cohen and Robert E. Hamilton1. In this analysis, the authors examine the conflict in three different areas:

* The goals of war- which the authors claim were the annexation of Abkhazia, the weakening or toppling the Saakashvili regime, and the prevention of NATO enlargement in the Caucasus.
* Russia’s military performance- and need of significant reforms.
* NATO and EU weakness- in relation to security assistance to countries of  the former Soviet Union.

Attacks on civilian targets, including hospitals, were a major part of that conflict, perpetrated mostly by Russia’s Vostok Battalion comprised of Chechens. In light of the current conflict (and what we have learned about Putin’s objectives since the paper was written in 2018), I disagree with the authors that the goal was only to capture Abkhazia- it now seems clear that Putin would have happily conquered and annexed all of Georgia had that invasion not also gone badly for him.

Lastly, *Carrots, Sticks, and Insurgent Targeting of Civilians*, by Victor Asal, Brian J. Phillips, R. Karl Rethemeyer, Corina Simonelli, and Joseph K. Young was of particular interest in developing my method of analysis2. This paper attempted to use statistical analysis and model development to draw insights, but was flawed in its logic and structure. The authors use “Terrorism” as a dichotomous variable and define several binary variables such as “Carrot” and “Stick” to describe actions. Among the conclusions based on output from the models generated are *“The coefficient associated with carrot is negatively signed and statistically significant, suggesting that governments that use conciliatory tactics toward a rebel group are less likely to see terrorism from that group in the following year.”* This suggests support for a policy of appeasement would be effective in reducing terror attacks. While I only agree with very little of the material presented and conclusions drawn, it was very interesting to see how their analytical process and method was derived.

# Methodology for Data Exploration

As a methodology for data exploration, time plots are a valuable tool to visualize trends and patterns in data over time. In this study, I created time plots to compare Russian battlefield losses and unlawful attacks on Healthcare workers and infrastructure. The time plots were created by plotting the number of Russian battlefield losses and unlawful attacks on Healthcare workers and infrastructure over time, respectively, using a consistent time interval.

To ensure that the time plots were valid representations of the data, I performed data cleaning and preprocessing steps. These steps included checking for missing values, outliers, and inconsistencies in the data. Verbal explanations and date ranges in the “ISODate” field were removed, appropriate data classes were coerced to ensure proper analytic processing, and Ukrainian language event descriptions were translated into English using the GoogleTrans library. For the datasets pertaining to Russian personnel and equipment losses, a “diff” function was used to calculate daily total losses rather than cumulative total losses across the date range sampled. I also ensured that the time plots accurately reflected the underlying data by using appropriate data visualization techniques, such as line plots and histograms, and by choosing appropriate scales for the axes as well as legends.

I then visually compared the time plots of Russian battlefield losses and unlawful attacks on Healthcare workers and infrastructure in order to identify any potential relationships or correlations between the two variables. This allowed me to identify any trends or patterns in the data that may be of interest, and to evaluate my hypotheses.

Overall, the creation and comparison of time plots is a valid and valuable technique for exploring data and identifying trends and patterns over time. This can help to uncover hidden relationships between variables and inform further analysis and modeling efforts.

# Description of the Data

The primary data source for use in the analysis is from The Humanitarian Data Exchange (HDX) of The United Nations Office for the Coordination of Humanitarian Affairs (OCHA)8. It is described as follows:

*“This dataset includes any reported incident (referred to as attack) that affected heatlhcare in Ukraine between 24 February 2022 and 04 March 2023.”*

The dataset consists of 1 x XLSX file:

* 2022-2023 Ukraine Attacks on Health Care Incident Data.xlsx

This is a detailed dataset from an official source, though it only contains data regarding attacks on health care workers and facilities. It covers the complete period from the beginning of the invasion on 24 February 2022 to 04 March 2023, encompassing all 4 campaign phases launched by the Russians thus far. It contains data on 785 discrete attacks. Included within the dataset is a description of the event in the Ukrainian language, coordinates for the location of the event, and the yield in terms of effect type and severity of the harmful act. Additionally there is attribution to an actor and also the type of weapon used. A description of the numeric variables follows:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **eventSindID** | **categoryHealthFacilitiesDamagedDestroyed** | **Infrastructure: Hospital** | **Infrastructure: Health Transport** | **Infrastructure: Other** | **HealthWorkersAttack: Health Building** | **HealthWorkersAttack: No Information** | **HealthWorkersAttack: Everyday Activities** | **HealthWorkersAttack: Outside Health Facility** | **Number of Attacks on Health Facilities Reporting Destruction** | **Number of Attacks on Health Facilities Reporting Damaged** | **Health Workers Killed** | **Health Workers Kidnapped** | **Health Workers Arrested** | **Health Workers Injured** | **Health Workers Assaulted** |
| **count** | 785 | 785 | 785 | 785 | 785 | 785 | 785 | 785 | 785 | 785 | 785 | 785 | 785 | 785 | 785 | 785 |
| **mean** | 34662.56178 | 0.407643 | 0.365605 | 0.054777 | 0.243312 | 0.043312 | 0.008917 | 0.015287 | 0.035669 | 0.078981 | 0.365605 | 0.105732 | 0.078981 | 0.031847 | 0.086624 | 0.003822 |
| **std** | 1691.583912 | 0.49171 | 0.481906 | 0.22769 | 0.429356 | 0.203689 | 0.094069 | 0.122769 | 0.185581 | 0.317643 | 0.539358 | 0.599983 | 1.519055 | 0.721573 | 0.473951 | 0.107075 |
| **min** | 30995 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **25%** | 33364 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **50%** | 34717 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **75%** | 36069 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| **max** | 37668 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 5 | 5 | 12 | 42 | 20 | 7 | 3 |

Additional datasets regarding Russian personnel and equipment losses were gathered from Dr. Petro Ivanuik, a Lviv, Ukraine-based Data Scientist who shares data related to the conflict on Kaggle6. It is described as follows:

*“This is a dataset that describes Equipment Losses, Death Toll, Military Wounded, and Prisoner of War of Russians in 2022 Ukraine Russia War.”*

The dataset used consists of 2 x CSV files:

* russia\_losses\_equipment.csv
* russia\_losses\_personnel.csv

These contained a cumulative daily total representing the physical costs of the war on the Russian MoD. For personnel losses, the dataset contains information from 25 February 2022 to 26 March 2023, representing a total of 170,550 Russians Killed in Action (KIA). A description of the numeric variables follows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **day** | **personnel** | **POW** | **daily\_increase** |
| **count** | 395 | 395 | 62 | 394 |
| **mean** | 199 | 66318.78734 | 386.387097 | 425.761421 |
| **std** | 114.170924 | 44805.05097 | 131.440363 | 305.628221 |
| **min** | 2 | 2800 | 0 | 0 |
| **25%** | 100.5 | 31000 | 389 | 200 |
| **50%** | 199 | 52250 | 421 | 350 |
| **75%** | 297.5 | 97985 | 474.5 | 620 |
| **max** | 396 | 170550 | 496 | 3160 |

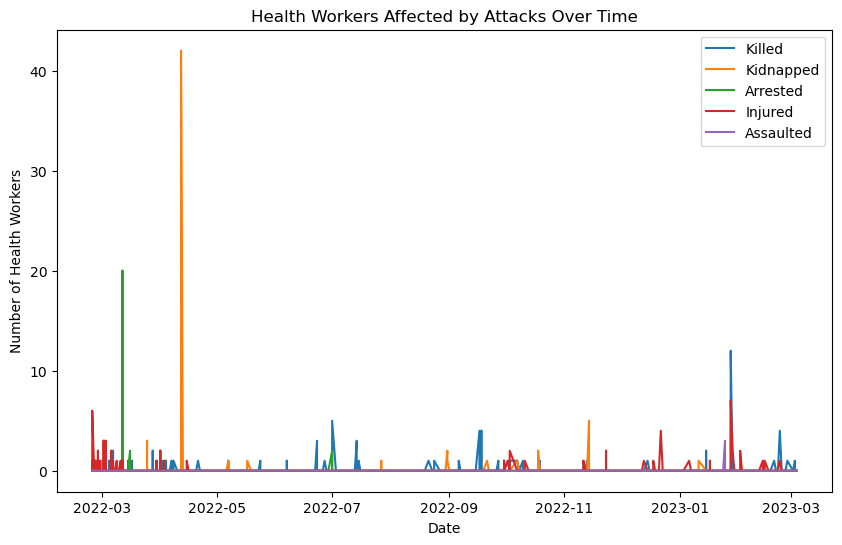
For the equipment losses, the dataset contains information from 25 February 2022 to 26 March 2023, representing various types of equipment lost or destroyed6. A description of the numeric variables follows:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **day** | **aircraft** | **helicopter** | **tank** | **APC** | **field artillery** | **MRL** | **military auto** | **fuel tank** | **drone** | **naval ship** | **anti-aircraft warfare** | **special equipment** | **mobile SRBM system** |
| **count** | 395 | 395 | 395 | 395 | 395 | 395 | 395 | 65 | 65 | 395 | 395 | 395 | 376 | 36 |
| **mean** | 199 | 232.921519 | 211.967089 | 2098.126582 | 4457.972152 | 1294.458228 | 301.726582 | 1047.507692 | 69.323077 | 1045.091139 | 13.840506 | 151.805063 | 121.255319 | 3.944444 |
| **std** | 114.170924 | 64.794571 | 61.818396 | 967.923865 | 1705.429199 | 743.553903 | 131.007173 | 466.16206 | 7.545917 | 682.146583 | 4.208691 | 69.647736 | 70.660833 | 0.333333 |
| **min** | 2 | 10 | 7 | 80 | 516 | 49 | 4 | 100 | 60 | 0 | 2 | 0 | 10 | 2 |
| **25%** | 100.5 | 210 | 175 | 1371.5 | 3372.5 | 677.5 | 207 | 600 | 60 | 537.5 | 13 | 95 | 55 | 4 |
| **50%** | 199 | 239 | 212 | 2136 | 4584 | 1259 | 311 | 1178 | 73 | 898 | 15 | 162 | 125 | 4 |
| **75%** | 297.5 | 281 | 264 | 2986 | 5960.5 | 1947.5 | 410 | 1437 | 76 | 1648.5 | 16 | 211 | 178 | 4 |
| **max** | 396 | 305 | 291 | 3595 | 6947 | 2631 | 522 | 1701 | 76 | 2216 | 18 | 277 | 282 | 4 |

# Exploratory Analysis

As described in the methodology section of this paper, the creation and comparison of time plots is a valid and valuable technique for exploring data and identifying trends and patterns over time. This can help to uncover hidden relationships between variables and inform further analysis and modeling efforts. Visualization of single variables is provided for perpetrator of attack and weapon used in the form of a pie chart. An interactive map depicting the locations of the attacks along with date, description, and destructive yield (quantified effects on healthcare workers and infrastructure ) was generated, and a sample view is provided.

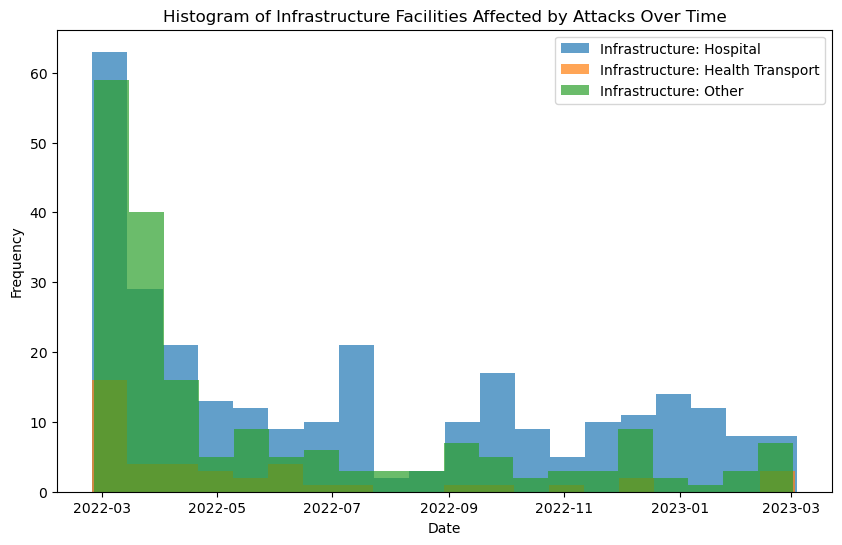
## Healthcare Workers

Plotting the crimes perpetrated against Healthcare workers over the time period analyzed paints a revealing picture. We see daily killing during the first four weeks following the invasion, followed by a lull in all types of harm. Category highs for kidnap and arrest also occur during this early period, though a review of the event description reveals that arbitrary categorization may have taken place regarding these two variables.

The 20 arrests on 12 March 2022 were the result of a regional hospital being taken-over by Russian Forces. Doctors and surrounding citizens were forced inside and prohibited to leave under threat of death. Russian Forces used the hospital as a military base and moved numerous weapons on to the grounds. The high for kidnapping was on 12 April 2022 when 42 doctors at a military hospital were held as prisoners of war by the Russians.

Surprisingly, the highest death toll waged against Healthcare workers during this period came at the hands of Ukrainian forces, on 28 January 2023. A hospital that had been taken over and used by Russian forces as a military hospital was hit by Ukrainian Forces rockets, killing 14 and injuring 24 hospital patients and medical staff. This was not the same location that was commandeered on 12 March 2022.

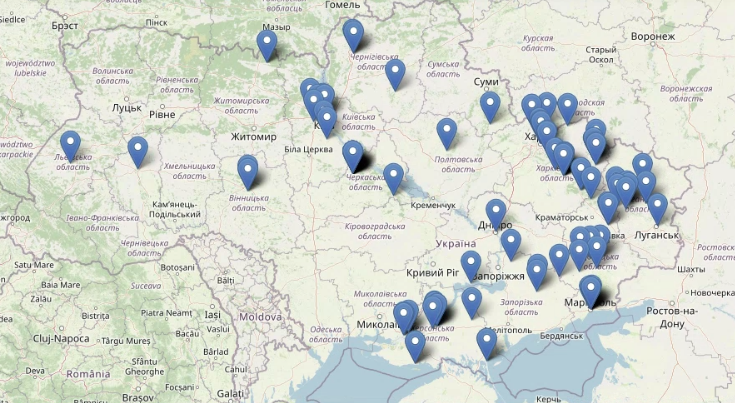
## Healthcare Facilities

Constructing a histogram of healthcare facilities affected by attacks roughly mirrors the effects on workers, though with a few surprises.

While one may assume that attacks made against Healthcare workers on the road in the course of their duties, and in their villages may occur at higher frequency than attacks on infrastructure facilities, that is not true. In the early phase of the conflict especially, attacks on Healthcare infrastructure were very high. Had those facilities been fully staffed, the early war death toll would have been much higher for healthcare workers. Many of the healthcare facilities early in the war were destroyed after critical patients evacuated to western Ukraine, and have not been reopened to date.

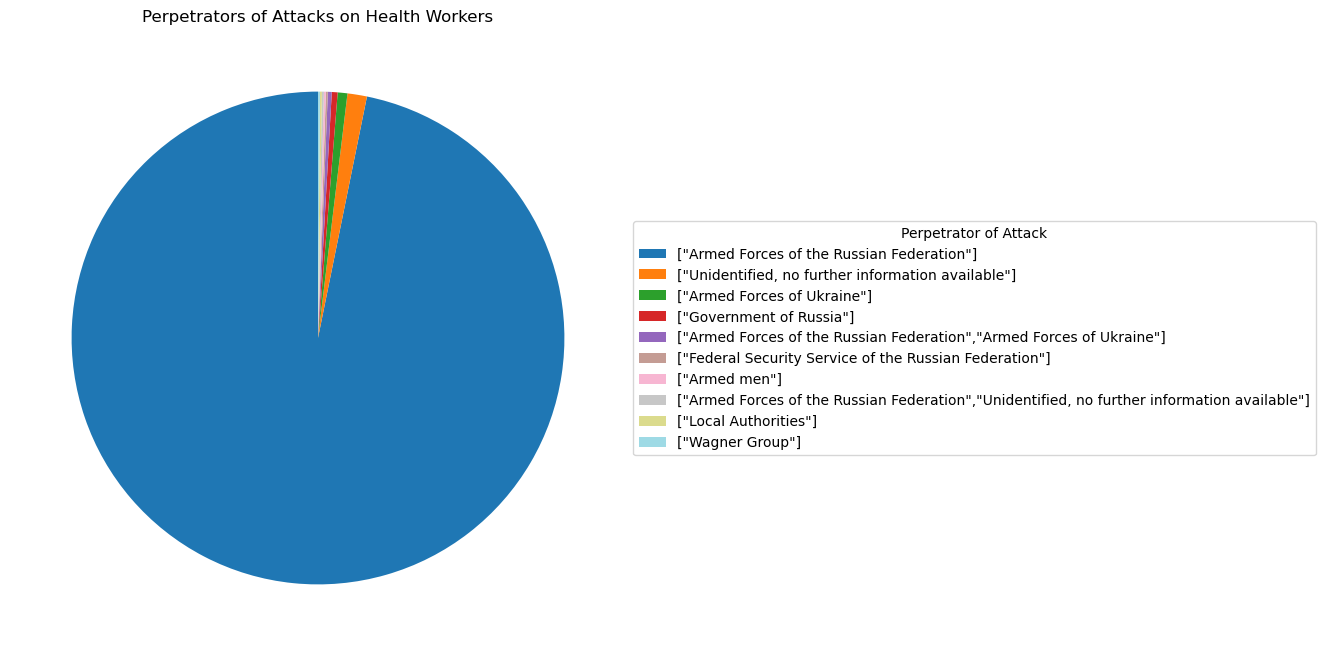
While the attacks on workers tend to level off and then spike, it is also interesting to note that attacks on facilities maintain two week averages in the double digits throughout most of the conflict, with only a lull in attacks against Hospitals in August of 2022. While some of the effect is due to the daily line chart totals for workers vs. histogram visualization for facilities, there are only 3 events resulting in double digit effects for healthcare workers, each in a different category.

## Locations of Attacks on Healthcare Workers and Infrastructure

A sample of the interactive visualization generated is depicted at left. Attacks on healthcare workers and infrastructure are plotted using their latitude and longitude coordinates contained within the dataset. As a tooltip (not pictured), mousing over a plot will reveal the date of the event, the translated, English language description of the event, and the yield in terms of quantitative effect (Example, 4 x Healthcare Workers Wounded, 2 x Healthcare Workers Wounded Killed).

This is a powerful tool for visualization, and shows the early events in the north and west eventually shifting to territories of the east in Donetsk and Lughansk Oblasts.

## Perpetrators of Attacks



While the primary perpetrator, by far, of attacks against Healthcare workers is the Armed Forces of the Russian Federation and associated entities, it may come as a surprise to note that at total of 5 attacks, representing just over half a percentage of attacks on healthcare workers were perpetrated by the Armed Forces of Ukraine. The Private Military Corporation, Wagner group, known for it’s brutality and lack of adherence to the laws of war, is attributed with only a single attack on healthcare workers.

**Perpetrator of Attack Total Percentage**

--------------------------------------------------------------------------------------------------------------------------------------

["Armed Forces of the Russian Federation"] 760 96.8%

["Unidentified, no further information available"] 10 1.3%

["Armed Forces of Ukraine"] 5 0.6%

["Government of Russia"] 3 0.4%

["Armed Forces of the Russian Federation","Armed Forces of Ukraine"] 2 0.3%

["Federal Security Service of the Russian Federation"] 1 0.1%

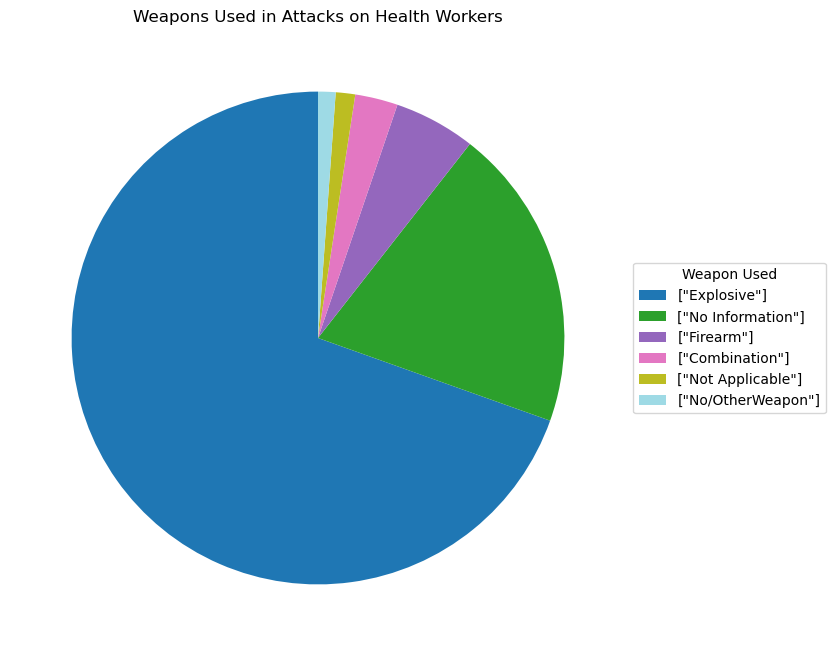
["Armed men"] 1 0.1%

["Armed Forces of the Russian Federation","Unidentified, nfia"] 1 0.1%

["Local Authorities"] 1 0.1%

["Wagner Group"] 1 0.1%

## Weapons used in Attacks



The vast majority of attacks on Health workers were carried out using explosives of some type- a category which includes missiles, rockets, artillery, and mortars, as well as land mines. This may be interpreted as a case of unintentional targeting, though in the early phase of the war Kaliber cruise missiles did specifically target major hospitals as well as other civilian infrastructure. In the case of firearms, it is highly likely that targeting was deliberate and intentional.

**Weapon Used Total Percentage**

--------------------------------------------------------------------------------------------------------------------------------------

["Explosive"] 546 69.6%

["No Information"] 156 19.9%

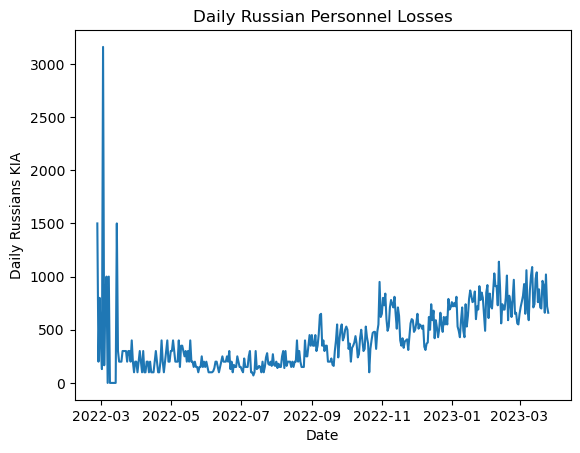
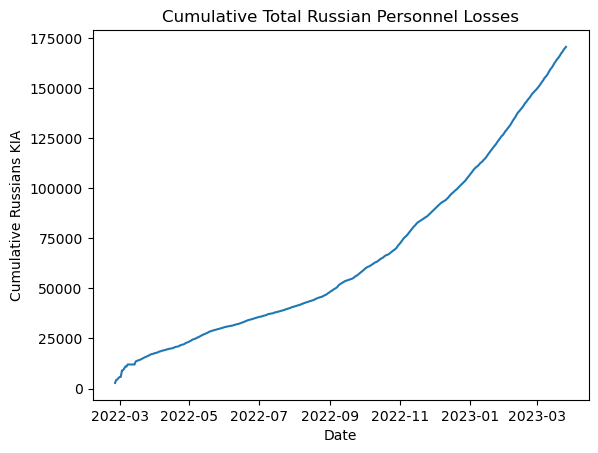
["Firearm"] 42 5.4%

["Combination"] 22 2.8%

["Not Applicable"] 10 1.3%

["No/Other Weapon"] 9 1.1%

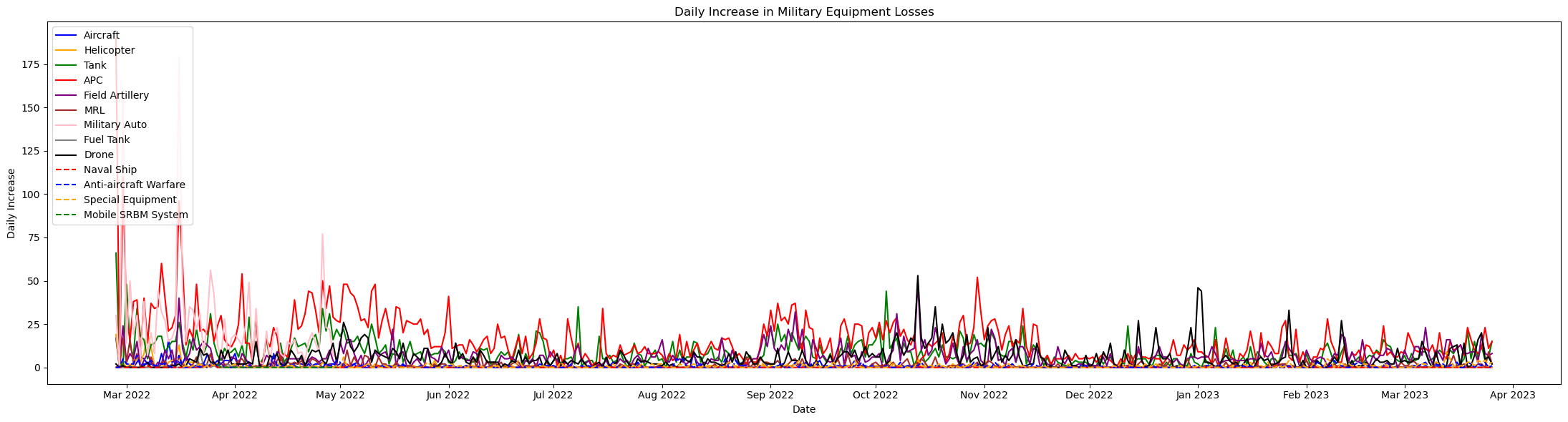
## Russian Personnel Losses



Transitioning to exploration of the datasets concerned with Russian battlefield losses, we can immediately see that the day-to-day trend of Russians Killed in Action (KIA) does not conform to the attacks on healthcare workers. After losing men in the thousands during the initial invasion, daily KIA totals have steadily climbed to reach an average of over 500 KIA per day for the past 90 days of the sampled date range. Viewing the cumulative total for Russian KIA, we can see a steep trend for the first month of the invasion that eventually becomes more gradual throughout the summer of 2022 during the Kherson Counteroffensive. Beginning in September of 2022 with the dramatically successful Kharkiv Counteroffensive, we see the curve become more steep as rate and scale improves for the Ukrainian Forces.

The hypothesis that attacks on Ukrainian healthcare workers and infrastructure is driven by Russian battlefield losses is therefore disproven. It even appears striking that increased operational tempo of the battlefield did not generate more Article 19 violations, considering that the Russian war machine currently relies on poorly trained conscripts and PMC Wagner forces, the vast majority of whom come from Russian jails and prisons.

## Russian Equipment Losses



Another interesting aspect of the data revealed by this exploration exercise, specifically with respect to equipment losses, is that in the year 2023, equipment losses of all types have been relatively low despite the high Russian KIA rate. This is due to a lack of available equipment to field, combined with high KIA rates of the conscript soldiers. We see high losses during the initial invasion, as we have come to expect in examining the other datasets, but after that we see a large loss of Armored Personnel Carriers (APCs) before and during the Kharkiv Counteroffensive, and then relatively low rates of equipment loss. The one exception to this is the high rate of drone interception and destruction that accompanied the January drone attacks on the Ukrainian power grid.

## Bivariate Regression

Adjusting for the variance in level of fidelity between the primary dataset- containing attacks on healthcare workers and infrastructure- and the supporting dataset- containing Russian personnel losses- a daily total was created containing wrapped healthcare workers killed in attacks. Using this as an index, a simple bivariate regression was performed to describe the relationships between Russian deaths and deaths resulting from attacks on healthcare workers, Russian deaths dependent variable. The results of that regression are below:

OLS Regression Results

========================================================================

Dep. Variable: daily\_increase R-squared: 0.011

Model: OLS Adj. R-squared: 0.008

Method: Least Squares F-statistic: 3.171

Date: Fri, 05 May 2023 Prob (F-statistic): 0.0761

Time: 19:33:42 Log-Likelihood: -1969.9

No. Observations: 275 AIC: 3944.

Df Residuals: 273 BIC: 3951.

Df Model: 1

Covariance Type: nonrobust

========================================================================

coef std err t P>|t| [0.025 0.975]

-------------------------------------------------------------------------------

const 405.9708 19.733 20.573 0.000 367.122 444.820

totaldeaths 33.7563 18.957 1.781 0.076 -3.563 71.076

========================================================================

Omnibus: 209.427 Durbin-Watson: 1.167

Prob(Omnibus): 0.000 Jarque-Bera (JB): 4411.915

Skew: 2.800 Prob(JB): 0.00

Kurtosis: 21.806 Cond. No. 1.35

========================================================================

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

These results provide some insight into the relationship between the dependent variable "daily\_increase," a daily measure of Russian KIA, and the independent variable and "totaldeaths," a daily measure of healthcare workers killed in attacks.

Examining R-squared, the R-squared value of 0.011 indicates that only a small portion (1.1%) of the variation in the daily\_increase can be explained by the totaldeaths variable. This means that **the totaldeaths variable has limited predictive power over daily\_increase.**

For Adjusted R-squared, The adjusted R-squared value of 0.008 takes into account the degrees of freedom and penalizes the model for the number of predictors. It provides a slightly lower value than the R-squared and suggests that the totaldeaths variable may not significantly improve the model's fit.

For the F-statistic, the value of 3.171 and the corresponding p-value of 0.0761 indicate that **the overall regression model is not statistically significant at the conventional significance level of 0.05. This suggests that the relationship between totaldeaths and daily\_increase may not be significant.**

The coefficients reveal the constant term to be 405.9708, indicating the expected value of daily\_increase (Russian KIA) when the totaldeaths (Healthcare workers killed) variable is zero. **The coefficient for the totaldeaths variable is 33.7563, suggesting that, on average, an increase of one unit in totaldeaths is associated with an increase of 33.7563 units in daily\_increase. However, since the p-value for this coefficient is 0.076, it is not statistically significant at the conventional significance level.**

Interpreting the p-values associated with the coefficients can provide a measure of their statistical significance, though in this case they are also unfortunately not significant. The p-value for the totaldeaths coefficient is 0.076, which is greater than 0.05. **This suggests that there is insufficient evidence to reject the null hypothesis that the coefficient is equal to zero.**

The Omnibus and Jarque-Bera tests reveal that the residuals in this case deviate significantly from a normal distribution.

The Durbin-Watson statistic of 1.167 measures the *autocorrelation* of the residuals. A value close to 2 suggests no autocorrelation. In this case, the value is less than 2, indicating the presence of positive autocorrelation.

The condition number of 1.35 suggests that there may be multicollinearity in the model.

Overall, looking at the results of this regression, the relationship between Healthcare workers killed and daily Russian KIA appears to not be statistically significant or meaningful. The low R-squared and the lack of significance in the coefficients suggest that there may be other variables that have a stronger influence on the Russian KIA.

## T-Test

A T-Test was peformed between the two variables, and the results are shown below:

T-test result:

t-statistic: 21.902439984650314

p-value: 7.403475411434446e-77

T-tests provide information about the statistical significance of the difference between the "totaldeaths" and "daily\_increase" variables. The t-statistic of 21.902439984650314 suggests a significant difference between the means of "totaldeaths" and "daily\_increase" variables, due ton one being consistently very low (healthcare workers killed) and the other showing more variation, but being consistently quite high.

The p-value shows us the probability of observing such a significant difference (or more extreme) between the groups if there were no true difference in the population. In this case, the p-value is 7.403475411434446e-77, which is an extremely small value close to zero. This means that the difference is highly unlikely to have occurred by chance alone, and that the means are very different (as we know). Therefore, we can conclude that there is a statistically significant difference between the average values of "totaldeaths" and "daily\_increase".

## Correlation

A correlation was calculated between daily\_increase (Russian KIA) and two combined variables representing total workers effected and total infrastructure effected. Total workers effected was calculated as ‘Health Workers Killed' + 'Health Workers Kidnapped' + Health Workers Arrested + Health Workers Injured' +'Health Workers Assaulted’. Total infrastructure effected was calculated as 'Infrastructure: Hospital'+'Infrastructure: Health Transport'+'Infrastructure: Other'. The results are below:

Correlation between daily increase and total workers effected: -0.007055612123944261

Correlation between daily increase and total infrastructure effected: -0.04895629684547883

Correlation between daily increase and total workers affected of -0.007 suggests a very weak negative correlation between daily increase and the total number of workers affected. **This means that there is almost no linear relationship between these two variables.** In other words, as the number of workers affected increases, the daily increase does not consistently decrease or increase.

Correlation between daily increase and total infrastructure affected of -0.049 suggests a weak negative correlation between daily increase and the total number of infrastructure affected. **This indicates a slight tendency for the daily increase to be lower when there is a higher number of infrastructure affected, but the relationship is weak.**

**In both cases, the correlations are close to zero, indicating that there is no strong linear relationship between daily increase and the total number of workers or infrastructure affected.**

# Implications of Findings

Based on the interpretation and analysis contained within this report, it is apparent that Russian personnel and equipment losses in the War in Ukraine do not drive Russian attacks on Ukrainian healthcare workers and infrastructure.

Further, the hypothesis that Russian attacks on protected targets follow events of high battlefield losses is not supported by statistical evidence in the datasets used. No attempt can be made to quantify how long the typical time lag between losses and attack is. The severity of the attack on healthcare workers and infrastructure does not appear to be scaled to the severity of the loss, as the relationship is weak between Russian KIA and Healthcare workers killed.

Some limitations of the analysis were that the level of fidelity between the datasets was not symmetrical. This frustrated fusion and enrichment of the primary dataset. Ideally, in order to perform such analysis at a high level, the same fidelity of data would be present for all types of attacks recorded. This information exists, and is recorded in real time by battlefield operating systems and sensors, but is not available open source for use in academic pursuits.

Some additional relevant variables might be meteorological data, as this has a significant effect on battlefield operations through restricted mobility and preclusion of air operations. This would be simple to obtain and enrich the primary dataset with, and is what I would do next as a future avenue of inquiry if I were to continue to develop this research topic.

# Conclusion

While it is disheartening that I did not find a relationship between these two variables, it is important to note that my study was limited to the available data and may not capture the full complexity of the situation. The absence of a significant correlation suggests that factors other than Russian KIA may be driving the number of attacks on healthcare workers.

On the bright side, this exercise has sparked my interest in continued research and investigation into the dynamics of the conflict and its impact on various aspects of society, including healthcare. Protecting healthcare workers and ensuring their safety is crucial for maintaining the well-being of the population and providing essential medical services in conflict zones during times of strife.

I express my hope that the conflict in Ukraine will be resolved soon, and that peace and stability will be restored. It is my sincere wish that the full territorial sovereignty of Ukraine is restored, allowing for the rebuilding of the affected regions and the establishment of a safe and stable environment for all individuals, including healthcare workers.

Слава Україні, Героям Слава, Смерть Ворогу!

# Citations

|  |  |
| --- | --- |
| 1 | Ariel Cohen Dr. and Robert E. Hamilton Colonel, *The Russian Military and the Georgia War: Lessons and Implications* ( US Army War College Press, 2011),https://press.armywarcollege.edu/monographs/576 |
| 2 | Asal, V., Phillips, B. J., Rethemeyer, R. K., Simonelli, C., & Young, J. K. (2019). Carrots, Sticks, and Insurgent Targeting of Civilians. Journal of Conflict Resolution, 63(7), 1710–1735. https://doi.org/10.1177/0022002718789748 |
| 3 | Cookman, Liz. "The Kids Aren’t Alright: Kyiv says more than 16,000 Ukrainian children have been taken to Russia." Foreign Policy, 17 APR 23. https://foreignpolicy.com/2023/04/17/russia-ukraine-children-kidnapping-war-crimes-icc/ |
| 4 | Doctor, Austin & Willingham, John. (2020). Foreign Fighters, Rebel Command Structure, and Civilian Targeting in Civil War. Terrorism and Political Violence. 34. 1-19. 10.1080/09546553.2020.1763320. |
| 5 | Garrett A. Cavaliere, Reem Alfalasi, Gregory N. Jasani, Gregory R. Ciottone, and Benjamin J. Lawner. Terrorist Attacks Against Healthcare Facilities: A Review. Health Security. Oct 2021.546-550.http://doi.org/10.1089/hs.2021.0004 |
| 6 | Ivanuik, Petro. "2022 Ukraine Russian War." Kaggle, 2023, https://www.kaggle.com/datasets/piterfm/2022-ukraine-russian-war. |
| 7 | "The Geneva Convention of 12 August 1949" (PDF, Library of Congress). LOC.gov. |
| 8 | "UCDP Data for Ukraine." Humanitarian Data Exchange, United Nations Office for the Coordination of Humanitarian Affairs, 2023, https://data.humdata.org/dataset/ucdp-data-for-ukraine. |
| 9 | Wong CH, Chen CY-T. Ambulances under siege in Syria. BMJ Glob Health2018;3:e001003. doi:10.1136/bmjgh-2018-001003 |

# Code

# # Kevin Ryan, JHU Spring '23

# # Unleashing Open Data with Python

# # Final Project - Attacks on Healthcare Workers and Infrstructure in Ukraine

# In[ ]:

import pandas as pd

# read the Excel file into a pandas dataframe

df = pd.read\_excel("2022-2023-ukraine-attacks-on-health-care-incident-data.xlsx")

# display the first 5 rows of the dataframe to verify it was read in correctly

print(df.head())

# ### Translate Event Descriptions from Ukrainian to English

# In[ ]:

from googletrans import Translator

# Create a translator object

translator = Translator()

# Define a function to translate from Ukrainian to English

def translate\_text(text):

return translator.translate(text, src='uk', dest='en').text

# Apply the function adn make a new column

df['eventDescriptionEnglishLang'] = df['eventDescriptionUkranianLang'].apply(translate\_text)

# Save the translated dataframe to a new CSV file

df.to\_csv("translated\_data.csv", index=False)

# ### Read in the CSV and Setup

# In[87]:

import matplotlib.pyplot as plt

df = pd.read\_csv("translated\_data1.csv")

# ### Force Numeric and Date columns to read properly

# In[88]:

# Convert the selected columns to numeric

numeric\_cols = ['categoryHealthFacilitiesDamagedDestroyed', 'Infrastructure: Hospital',

'Infrastructure: Health Transport', 'Infrastructure: Other',

'HealthWorkersAttack: Health Building', 'HealthWorkersAttack: No Information',

'HealthWorkersAttack: Everyday Activities', 'HealthWorkersAttack: Outside Health Facility']

df[numeric\_cols] = df[numeric\_cols].apply(pd.to\_numeric, errors='coerce')

numeric\_cols = ['Number of Attacks on Health Facilities Reporting Destruction',

'Number of Attacks on Health Facilities Reporting Damaged',

'Health Workers Killed', 'Health Workers Kidnapped',

'Health Workers Arrested', 'Health Workers Injured',

'Health Workers Assaulted']

df[numeric\_cols] = df[numeric\_cols].apply(pd.to\_numeric, errors='coerce')

# In[89]:

# Convert ISODate to date format, drop bad columns (Ukrainian text explanation instead of date)

df = df.apply(lambda col: pd.to\_datetime(col, errors='coerce') if col.name == 'ISODate' else col)

df = df.dropna(subset=['ISODate'])

# Reset the index

df = df.reset\_index(drop=True)

# In[90]:

# Describe df

df.describe()

# ### Plot types of attack on workers over time

# In[20]:

# Plot line graph

fig, ax = plt.subplots(figsize=(10, 6))

ax.plot(df['ISODate'], df['Health Workers Killed'], label='Killed')

ax.plot(df['ISODate'], df['Health Workers Kidnapped'], label='Kidnapped')

ax.plot(df['ISODate'], df['Health Workers Arrested'], label='Arrested')

ax.plot(df['ISODate'], df['Health Workers Injured'], label='Injured')

ax.plot(df['ISODate'], df['Health Workers Assaulted'], label='Assaulted')

# Fix graph

ax.set\_xlabel('Date')

ax.set\_ylabel('Number of Health Workers')

ax.set\_title('Health Workers Affected by Attacks Over Time')

ax.legend()

plt.show()

# ### Plot infrastructure facilities affected by attacks over time

# In[21]:

from matplotlib.dates import MonthLocator, DateFormatter

# Create plot

fig, ax = plt.subplots(figsize=(10, 6))

ax.plot(df['ISODate'], df['Infrastructure: Hospital'], label='Hospital')

ax.plot(df['ISODate'], df['Infrastructure: Health Transport'], label='Health Transport')

ax.plot(df['ISODate'], df['Infrastructure: Other'], label='Other')

# Label graph

ax.set\_xlabel('Date')

ax.set\_ylabel('Number of Infrastructure Facilities Affected')

ax.set\_title('Infrastructure Facilities Affected by Attacks Over Time')

ax.legend()

# Format x-axis to display only the months so it isn't crowded

months = MonthLocator()

date\_format = DateFormatter('%b %Y')

ax.xaxis.set\_major\_locator(months)

ax.xaxis.set\_major\_formatter(date\_format)

plt.xticks(rotation=45, ha='right')

plt.show()

#LOL

# In[23]:

# Define bins for the histogram

bins = 20

# Create histogram for each facility type

fig, ax = plt.subplots(figsize=(10, 6))

for facility in ['Infrastructure: Hospital', 'Infrastructure: Health Transport', 'Infrastructure: Other']:

ax.hist(df[df[facility] > 0]['ISODate'], bins=bins, alpha=0.7, label=facility)

# Label graph

ax.set\_xlabel('Date')

ax.set\_ylabel('Frequency')

ax.set\_title('Histogram of Infrastructure Facilities Affected by Attacks Over Time')

ax.legend()

plt.show()

# ### Plot attacks on a Map

# In[94]:

import folium

# Read in the data and drop any rows with missing latitude or longitude data

df = df.dropna(subset=['latitude2', 'longitude2'])

# Create a map and center it

m = folium.Map(location=[df['latitude2'].median(), df['longitude2'].median()], zoom\_start=5)

# Loop

for col in ['Health Workers Killed', 'Health Workers Kidnapped', 'Health Workers Arrested', 'Health Workers Injured', 'Health Workers Assaulted']:

df\_filtered = df[df[col] > 0]

for i, row in df\_filtered.iterrows():

popup\_text = f"{row['eventDescription']} ({col}: {row[col]})"

folium.Marker([row['latitude2'], row['longitude2']], popup=popup\_text).add\_to(m)

# Show the map

m

# ### Plot Pie Charts for Perpetrators of attack, Weapons used to attack

# In[37]:

### FIX

# Get the counts of each type of perpetrator

counts = df['Perpetrator of Attack'].value\_counts()

# Create the pie chart

fig, ax = plt.subplots(figsize=(8, 8))

wedges, \_, labels = ax.pie(counts.values, labels=counts.index, autopct='%1.1f%%', startangle=90)

# title

ax.set\_title("Perpetrators of Attacks on Health Workers")

# legend

ax.legend(wedges, labels, title="Perpetrator of Attack", loc="center left", bbox\_to\_anchor=(1, 0, 0.5, 1))

# Show the chart

plt.show()

# In[38]:

### Fix

# Get the counts of each type of weapon used

counts = df['Weapon Used'].value\_counts()

# Create the pie chart

fig, ax = plt.subplots(figsize=(8, 8))

wedges, \_, labels = ax.pie(counts.values, labels=counts.index, autopct='%1.1f%%', startangle=90, labeldistance=1.1)

# title

ax.set\_title("Weapons Used in Attacks on Health Workers")

# legend

ax.legend(wedges, labels, title="Weapon Used", loc="center left", bbox\_to\_anchor=(1, 0, 0.5, 1))

# Show the chart

plt.show()

# In[49]:

# Read in Russian Personnel losses

df1 = pd.read\_csv("Russia\_losses\_personnel.csv")

# Drop the non-useful column

df1 = df1.drop(columns=["personnel\*"])

# Fix that it is a cumulative total

df1["daily\_increase"] = df1["personnel"].diff()

# Force the date to be an ISODate

df1["date"] = pd.to\_datetime(df1["date"])

print(df1)

# Summary statistics for df

print('Summary statistics for df:')

print(df.describe())

# Summary statistics for df1

print('\nSummary statistics for df1:')

print(df1.describe())

# In[44]:

# Create a new column for total workers effected

df['Total workers effected'] = df['Health Workers Killed'] + df['Health Workers Kidnapped'] + df['Health Workers Arrested'] + df['Health Workers Injured'] + df['Health Workers Assaulted']

# Create a new column for total infrastructure effected

df['Total infrastructure effected'] = df[['Infrastructure: Hospital', 'Infrastructure: Health Transport', 'Infrastructure: Other']].sum(axis=1)

# Merge the daily increase column from df1 into df based on the date column

df = pd.merge(df, df1[['date', 'daily\_increase']], how='left', left\_on='ISODate', right\_on='date')

# Calculate the correlation between daily increase in personnel losses and total workers/infrastructure effected

corr\_workers = df['daily\_increase'].corr(df['Total workers effected'])

corr\_infra = df['daily\_increase'].corr(df['Total infrastructure effected'])

print(f"Correlation between daily increase and total workers effected: {corr\_workers}")

print(f"Correlation between daily increase and total infrastructure effected: {corr\_infra}")

# ### A weak, negative correlation exists between Russian invaders killed in action (KIA) and Attacks on Healthcare workers (-0.007055612123944261) and Attacks on Healthcare Infrastructure (-0.04895629684547883).

#

# In[51]:

# Plot daily\_increase over date

plt.plot(df1['date'], df1['daily\_increase'])

plt.xlabel('Date')

plt.ylabel('Daily Russians KIA')

plt.title('Daily Russian Personnel Losses')

plt.show()

# In[53]:

# Plot cumulative increase over date

plt.plot(df1['date'], df1['personnel'])

plt.xlabel('Date')

plt.ylabel('Cumulative Russians KIA')

plt.title('Cumulative Total Russian Personnel Losses')

plt.show()

# In[101]:

import pandas as pd

df2 = pd.read\_csv('russia\_losses\_equipment.csv')

# In[102]:

# Select the relevant columns

cumulative\_columns = ['aircraft', 'helicopter', 'tank', 'APC', 'field artillery', 'MRL', 'military auto', 'fuel tank', 'drone', 'naval ship', 'anti-aircraft warfare', 'special equipment', 'mobile SRBM system']

# Compute the daily increase for each column to fix cumulative

for col in cumulative\_columns:

df2[f'{col}\_daily\_increase'] = df2[col].diff()

print(df2)

# In[62]:

import matplotlib.dates as mdates

fig, ax = plt.subplots(figsize=(22,6))

plt.plot(df2['date'], df2['aircraft\_daily\_increase'], color='blue', label='Aircraft')

plt.plot(df2['date'], df2['helicopter\_daily\_increase'], color='orange', label='Helicopter')

plt.plot(df2['date'], df2['tank\_daily\_increase'], color='green', label='Tank')

plt.plot(df2['date'], df2['APC\_daily\_increase'], color='red', label='APC')

plt.plot(df2['date'], df2['field artillery\_daily\_increase'], color='purple', label='Field Artillery')

plt.plot(df2['date'], df2['MRL\_daily\_increase'], color='brown', label='MRL')

plt.plot(df2['date'], df2['military auto\_daily\_increase'], color='pink', label='Military Auto')

plt.plot(df2['date'], df2['fuel tank\_daily\_increase'], color='gray', label='Fuel Tank')

plt.plot(df2['date'], df2['drone\_daily\_increase'], color='black', label='Drone')

plt.plot(df2['date'], df2['naval ship\_daily\_increase'], color='red', linestyle='dashed', label='Naval Ship')

plt.plot(df2['date'], df2['anti-aircraft warfare\_daily\_increase'], color='blue', linestyle='dashed', label='Anti-aircraft Warfare')

plt.plot(df2['date'], df2['special equipment\_daily\_increase'], color='orange', linestyle='dashed', label='Special Equipment')

plt.plot(df2['date'], df2['mobile SRBM system\_daily\_increase'], color='green', linestyle='dashed', label='Mobile SRBM System')

plt.legend(loc='upper left')

plt.xlabel('Date')

plt.ylabel('Daily Increase')

plt.title('Daily Increase in Military Equipment Losses')

# Format x-axis labels to show only the month

ax.xaxis.set\_major\_locator(mdates.MonthLocator(interval=1))

ax.xaxis.set\_major\_formatter(mdates.DateFormatter('%b %Y'))

plt.tight\_layout()

plt.show()

# In[68]:

# Convert 'date' column in df2

df2['date'] = pd.to\_datetime(df2['date'])

# Merge the three dataframes on 'date' column

df3 = pd.merge(df, df1, on='date')

df3 = pd.merge(df3, df2, on='date')

# Set the index to 'date'

df3.set\_index('date', inplace=True)

# In[75]:

print(df3.shape)

print(df3.head)

# In[77]:

# Add a column to df3 called "Total\_Attacks\_Workers" that is the sum of the daily totals of the columns you specified

df3['Total\_Attacks\_Workers'] = df[['Health Workers Killed', 'Health Workers Kidnapped', 'Health Workers Arrested', 'Health Workers Injured', 'Health Workers Assaulted']].sum(axis=1)

# In[86]:

import numpy as np

import statsmodels.api as sm

df3 = df3.dropna() # drop rows with NaN values

X = sm.add\_constant(df3['personnel']) # add a constant term

y = df3['total\_attacks\_workers']

model = sm.OLS(y, X).fit() # fit the OLS model

print(model.summary()) # print the model summary

# In[91]:

df1.describe()

# In[92]:

df2.describe()

# In[93]:

df2.iloc[:, :15].describe()

# In[96]:

# Get the counts of each type of perpetrator

counts = df['Perpetrator of Attack'].value\_counts()

# Define color map

colors = plt.cm.tab20(np.linspace(0, 1, len(counts.index)))

# Create the pie chart

fig, ax = plt.subplots(figsize=(8, 8))

wedges, \_ = ax.pie(counts.values, colors=colors, startangle=90)

# title

ax.set\_title("Perpetrators of Attacks on Health Workers")

# legend

ax.legend(wedges, counts.index, title="Perpetrator of Attack", loc="center left", bbox\_to\_anchor=(1, 0, 0.5, 1))

# Show the chart

plt.show()

# In[97]:

# Get the counts of each type of weapon used

counts = df['Weapon Used'].value\_counts()

# Define color map

colors = plt.cm.tab20(np.linspace(0, 1, len(counts.index)))

# Create the pie chart

fig, ax = plt.subplots(figsize=(8, 8))

wedges, \_ = ax.pie(counts.values, colors=colors, startangle=90, labeldistance=1.1)

# title

ax.set\_title("Weapons Used in Attacks on Health Workers")

# legend

ax.legend(wedges, counts.index, title="Weapon Used", loc="center left", bbox\_to\_anchor=(1, 0, 0.5, 1))

# Show the chart

plt.show()

# In[98]:

from tabulate import tabulate

# Get the counts and percentages of each type of perpetrator

counts = df['Perpetrator of Attack'].value\_counts()

percentages = counts / counts.sum() \* 100

# Create a table with the counts and percentages

table = []

for i, index in enumerate(counts.index):

table.append([index, counts[i], f'{percentages[i]:.1f}%'])

# Print the table

print(tabulate(table, headers=['Perpetrator of Attack', 'Total', 'Percentage']))

# In[99]:

from tabulate import tabulate

# Get the counts and percentages of each type of weapon used

counts = df['Weapon Used'].value\_counts()

percentages = counts / counts.sum() \* 100

# Create a table with the counts and percentages

table = []

for i, index in enumerate(counts.index):

table.append([index, counts[i], f'{percentages[i]:.1f}%'])

# Print the table

print(tabulate(table, headers=['Weapon Used', 'Total', 'Percentage']))

# In[108]:

import matplotlib.dates as mdates

# It never reads right

df2['date'] = pd.to\_datetime(df2['date'])

# Clippity doooohh daaaaaahhhh

df2['APC\_daily\_increase'] = df2['APC\_daily\_increase'].clip(lower=0)

# The longest

fig, ax = plt.subplots(figsize=(22,6))

plt.plot(df2['date'], df2['aircraft\_daily\_increase'], color='blue', label='Aircraft')

plt.plot(df2['date'], df2['helicopter\_daily\_increase'], color='orange', label='Helicopter')

plt.plot(df2['date'], df2['tank\_daily\_increase'], color='green', label='Tank')

plt.plot(df2['date'], df2['APC\_daily\_increase'], color='red', label='APC')

plt.plot(df2['date'], df2['field artillery\_daily\_increase'], color='purple', label='Field Artillery')

plt.plot(df2['date'], df2['MRL\_daily\_increase'], color='brown', label='MRL')

plt.plot(df2['date'], df2['military auto\_daily\_increase'], color='pink', label='Military Auto')

plt.plot(df2['date'], df2['fuel tank\_daily\_increase'], color='gray', label='Fuel Tank')

plt.plot(df2['date'], df2['drone\_daily\_increase'], color='black', label='Drone')

plt.plot(df2['date'], df2['naval ship\_daily\_increase'], color='red', linestyle='dashed', label='Naval Ship')

plt.plot(df2['date'], df2['anti-aircraft warfare\_daily\_increase'], color='blue', linestyle='dashed', label='Anti-aircraft Warfare')

plt.plot(df2['date'], df2['special equipment\_daily\_increase'], color='orange', linestyle='dashed', label='Special Equipment')

plt.plot(df2['date'], df2['mobile SRBM system\_daily\_increase'], color='green', linestyle='dashed', label='Mobile SRBM System')

plt.legend(loc='upper left')

plt.xlabel('Date')

plt.ylabel('Daily Increase')

plt.title('Daily Increase in Military Equipment Losses')

# Format x-axis labels to show only the month

ax.xaxis.set\_major\_locator(mdates.MonthLocator(interval=1))

ax.xaxis.set\_major\_formatter(mdates.DateFormatter('%b %Y'))

plt.tight\_layout()

plt.show()

# In[109]:

import folium

dff = df.dropna(subset=['latitude2', 'longitude2', 'ISODate'])

# Define the color scale

color\_scale = ['blue', 'green', 'orange', 'red']

# Create a map and center it

m = folium.Map(location=[dff['latitude2'].median(), dff['longitude2'].median()], zoom\_start=5)

# Loop

for col in ['Health Workers Killed', 'Health Workers Kidnapped', 'Health Workers Arrested', 'Health Workers Injured', 'Health Workers Assaulted']:

dff\_filtered = dff[dff[col] > 0]

for i, row in dff\_filtered.iterrows():

popup\_text = f"{row['eventDescription']} ({col}: {row[col]})"

folium.Marker([row['latitude2'], row['longitude2']], popup=popup\_text, icon=folium.Icon(color=color\_scale[int(row['ISODate'][-4:])%4])).add\_to(m)

# Show the map

m

# In[112]:

## I struck out here, but I was trying to apply a color scale

## across time to the map plots to depict the shift eastward in combat operations

import pandas as pd

import folium

import matplotlib.pyplot as plt

import matplotlib.colors as mcolors

# Read in data and drop any rows with missing latitude or longitude

df = df.dropna(subset=['latitude2', 'longitude2'])

# Create a map and center it

m = folium.Map(location=[df['latitude2'].median(), df['longitude2'].median()], zoom\_start=5)

# Define the starting and ending dates

start\_date = pd.Timestamp('2022-02-01')

end\_date = pd.Timestamp('2023-03-31')

# Convert Timestamp objects to numeric values

start\_value = start\_date.timestamp()

end\_value = end\_date.timestamp()

# Loop

for col in ['Health Workers Killed', 'Health Workers Kidnapped', 'Health Workers Arrested', 'Health Workers Injured', 'Health Workers Assaulted']:

df\_filtered = df[df[col] > 0]

for i, row in df\_filtered.iterrows():

popup\_text = f"{row['eventDescription']} ({col}: {row[col]})"

time\_value = row['ISODate'].timestamp()

color = mcolors.rgb2hex(plt.cm.RdYlBu((time\_value - start\_value) / (end\_value - start\_value)))

folium.Marker([row['latitude2'], row['longitude2']], popup=popup\_text, icon=folium.Icon(color=color)).add\_to(m)

# Show the map

m

# In[116]:

from datetime import datetime

# Read the CSV file

df\_math = pd.read\_csv("HCK math.csv")

# Make it read correctly by coercing date

df\_math["groupid"] = pd.to\_datetime(df\_math["groupid"])

from scipy.stats import ttest\_ind

# Selecting the relevant columns from df1 and df\_math

df1\_selected = df1[["date", "daily\_increase"]]

df\_math\_selected = df\_math[["groupid", "totaldeaths"]]

# Merging the selected dataframes on "groupid"

merged\_df = pd.merge(df1\_selected, df\_math\_selected, left\_on="date", right\_on="groupid")

# Dropping rows with NaNs

merged\_df = merged\_df.dropna()

# Perform the T-test

result = ttest\_ind(merged\_df["daily\_increase"], merged\_df["totaldeaths"])

# Print the result

print("T-test result:")

print("t-statistic:", result.statistic)

print("p-value:", result.pvalue)

# In[117]:

import statsmodels.api as sm

# Define the variables

X = merged\_df['daily\_increase']

y = merged\_df['totaldeaths']

# Adding a constant to the IV

X = sm.add\_constant(X)

# Creating and fitting the model

model = sm.OLS(y, X)

results = model.fit()

# Printing the results

print(results.summary())

# In[118]:

# Selecting the relevant columns from df1 and df

df1\_selected = df1["daily\_increase"]

df\_selected = df[["categoryHealthFacilitiesDamagedDestroyed", "Infrastructure: Hospital", "Infrastructure: Health Transport", "Infrastructure: Other", "HealthWorkersAttack: Health Building", "HealthWorkersAttack: No Information", "HealthWorkersAttack: Everyday Activities", "HealthWorkersAttack: Outside Health Facility", "Number of Attacks on Health Facilities Reporting Destruction", "Number of Attacks on Health Facilities Reporting Damaged", "Health Workers Killed", "Health Workers Kidnapped", "Health Workers Arrested", "Health Workers Injured", "Health Workers Assaulted"]]

##### Does not work because the indexes can't line up

# Something about adding a constant might help?

df\_selected = sm.add\_constant(df\_selected)

# Fit the multivariate regression model

model = sm.OLS(df1\_selected, df\_selected)

results = model.fit()

# Print the summary

print(results.summary())

# In[120]:

# Define the variables

X = merged\_df['totaldeaths']

y = merged\_df['daily\_increase']

# Adding a constant to the independent variable

X = sm.add\_constant(X)

# Creating and fitting the model

model = sm.OLS(y, X)

results = model.fit()

# Print the results

print(results.summary())

# In[ ]: