DACSS 603 Final Project Chloe Morgado

Chloe Morgado

2024-12-01

The Influence of War Duration on Outcomes: A Quantitative Analysis of Interstate Wars

Introduction

The dynamics of interstate wars have long been a focus of academic inquiry, particularly regarding the factors that influence their outcomes. Among the many variables shaping conflict resolution, war duration stands out as a critical yet understudied determinant. The length of a war often reflects the complexity of the conflict, the resources available to participants, and the level of international involvement, all of which can significantly shape its resolution. Understanding how the duration of a war correlates with its eventual outcome - whether it results in victory, defeat, or stalemate, it offers valuable insight into the strategic and geopolitical dimensions of warfare.

Existing research has suggested that prolonged wars may lead to inconclusive outcomes due to resource exhaustion, shifts in political will, or international mediation efforts. Conversely, shorter wars are often associated with decisive outcomes, as they tend to involve well-planned and quickly executed campaigns. However, empirical evidence exploring these patterns remains limited, particularity when considering a comprehensive dataset spanning diverse conflicts across time and regions.

This final project examines the relationship between war duration and outcomes using the Interstate Wars dataset, which is a complied dataset of interstate conflicts. By analyzing the duration and outcomes of these wars, this final project examines the research question of;

Research Question and Hypothesis

What is the relationship between the duration of a way and its eventual outcome?

The hypothesis is as follows:

The duration of a war significantly influences its outcome, with shorter wars being more likely to result in decisive outcomes (victory or defeat), while longer wars are more likely to result in inconclusive outcomes (stalemates or negotiated settlements).

H0: There is no significant relationship between the duration of a war and its outcome.

H1: There is a significant relationship between the duration of a war and its outcome.

Through statistical and visual analyses, I am to uncover the patterns and trends that contribute to deeper understanding of how the passage of time influences the resolution of conflicts. The findings from this final project have both theoretical and practical implications. They can inform military strategy by highlighting the importance of time as a factor in conflict resolution and provide policymakers with insights for designing effective intervention strategies. Ultimately, this research contributes to the broader discourse on the factors that shape the trajectory and resolution of interstate conflicts, which offers a new perspective on the role of duration in determining war outcomes.

Descriptive Statistics

I ended up using the Correlates of War: Interstate Wars dataset for this final project, which was found on Kaggle. This goal of the project behind the creation of this dataset was to systemically record and analyze international conflicts. The data was collected through historical records, governmental archives, and scholarly publications.This dataset encompasses comprehensive information on interstate wars between 1816 and 2007, which clearly details countries involved, conflict dates, and associated fatalities.

This dataset is available to the public on Kaggle and can be accessed at the following link: https://www.kaggle.com/datasets/umichigan/interstate-wars/data

I was having a hard time finding ones that interested me, especially when it came to relating quantitative analysis to political science. Geraldine Santoso shared this dataset with me because her and I do similar research to political violence. The key variables for the dataset that I looked at contained:

Outcome: This categorical variable indicated the war's result (e.g., victory, defeat, stalemate).

Duration: I calculated the difference between the war's end and start years, which represented the conflict's length.

Initiation: This is the binary variable denoting whether a state initiated the conflict.

Participant Count: This is what I used to show the number of states that were involved in the war.

The main findings from the summary statistics is that the average duration is approximately 1.25 years, with conflicts ranging from less than a year to a maximum of 10 years. The

dataset includes various outcomes, so there is a notable distribution across victory, defeat, and stalemate.

Code of the Interstate War Dataset

```
library(tidygraph)
```

```
Attaching package: 'tidygraph'

The following object is masked from 'package:stats':

    filter
```

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.3     v tidyr     1.3.1
v purrr     1.0.2

-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks tidygraph::filter(), stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

```
library(dplyr)
```

```
wars <- read.csv("/Users/chloemorgado/Downloads/interstate.csv")

head(wars)
```

```
  war_id            war_name war_type state_code               state_name side
1      1  Franco-Spanish War        1        230                    Spain    2
2      1  Franco-Spanish War        1        220                   France    1
3      4 First Russo-Turkish        1        640           Ottoman Empire    2
```

```
4      4 First Russo-Turkish        1        365                  Russia    1
5      7    Mexican-American        1         70                  Mexico    2
6      7    Mexican-American        1          2 United States of America    1
  start_year1 start_month1 start_day1 end_year1 end_month1 end_day1 start_year2
1        1823            4          7      1823         11       13          -8
2        1823            4          7      1823         11       13          -8
3        1828            4         26      1829          9       14          -8
4        1828            4         26      1829          9       14          -8
5        1846            4         25      1847          9       14          -8
6        1846            4         25      1847          9       14          -8
  start_month2 start_day2 end_year2 end_month2 end_day2 previous_war initiation
1           -8         -8        -8         -8       -8          503          2
2           -8         -8        -8         -8       -8          503          1
3           -8         -8        -8         -8       -8          506          2
4           -8         -8        -8         -8       -8          506          1
5           -8         -8        -8         -8       -8           -8          2
6           -8         -8        -8         -8       -8           -8          1
  combat_location combat_fatalities outcome next_war
1               2               600       2       -8
2               2               400       1       -8
3              11             80000       2       -8
4              11             50000       1       -8
5               1              6000       2       -8
6               1             13283       1       -8
```

```
colSums(is.na(wars))
```

```
          war_id          war_name          war_type        state_code
               0                 0                 0                 0
      state_name              side       start_year1      start_month1
               0                 0                 0                 0
      start_day1         end_year1        end_month1          end_day1
               0                 0                 0                 0
     start_year2      start_month2        start_day2        end_year2
               0                 0                 0                 0
      end_month2          end_day2      previous_war        initiation
               0                 0                 0                 0
  combat_location combat_fatalities           outcome          next_war
               0                 0                 0                 0
```

```r
wars <- wars %>%
  mutate(duration = end_year1 - start_year1)

wars$start_year <- as.numeric(wars$start_year1)
wars$end_year <- as.numeric(wars$end_year1)
```

```r
summary(wars)
```

```
     war_id        war_name            war_type   state_code
 Min.   :  1.0   Length:337         Min.   :1   Min.   :  2.0
 1st Qu.: 82.0   Class :character   1st Qu.:1   1st Qu.:220.0
 Median :139.0   Mode  :character   Median :1   Median :355.0
 Mean   :126.9                      Mean   :1   Mean   :419.9
 3rd Qu.:172.0                      3rd Qu.:1   3rd Qu.:652.0
 Max.   :227.0                      Max.   :1   Max.   :920.0
  state_name             side        start_year1    start_month1
 Length:337         Min.   :1.000   Min.   :1823   Min.   : 1.000
 Class :character   1st Qu.:1.000   1st Qu.:1900   1st Qu.: 3.000
 Mode  :character   Median :1.000   Median :1939   Median : 6.000
                    Mean   :1.418   Mean   :1931   Mean   : 6.249
                    3rd Qu.:2.000   3rd Qu.:1969   3rd Qu.: 9.000
                    Max.   :2.000   Max.   :2003   Max.   :12.000
   start_day1      end_year1      end_month1        end_day1
 Min.   : 1.00   Min.   :1823   Min.   : 1.000   Min.   : 1.00
 1st Qu.: 7.00   1st Qu.:1900   1st Qu.: 4.000   1st Qu.:10.00
 Median :16.00   Median :1941   Median : 7.000   Median :14.00
 Mean   :15.35   Mean   :1932   Mean   : 6.623   Mean   :15.04
 3rd Qu.:23.00   3rd Qu.:1973   3rd Qu.: 9.000   3rd Qu.:23.00
 Max.   :31.00   Max.   :2003   Max.   :12.000   Max.   :31.00
  start_year2      start_month2      start_day2       end_year2
 Min.   :  -8.0   Min.   :-8.000   Min.   :-8.000   Min.   :  -8.0
 1st Qu.:  -8.0   1st Qu.:-8.000   1st Qu.:-8.000   1st Qu.:  -8.0
 Median :  -8.0   Median :-8.000   Median :-8.000   Median :  -8.0
 Mean   : 100.3   Mean   :-7.169   Mean   :-6.647   Mean   : 100.3
 3rd Qu.:  -8.0   3rd Qu.:-8.000   3rd Qu.:-8.000   3rd Qu.:  -8.0
 Max.   :1974.0   Max.   :10.000   Max.   :25.000   Max.   :1974.0
   end_month2        end_day2      previous_war         initiation
 Min.   :-8.000   Min.   :-8.000   Length:337         Min.   :1.000
 1st Qu.:-8.000   1st Qu.:-8.000   Class :character   1st Qu.:1.000
 Median :-8.000   Median :-8.000   Mode  :character   Median :2.000
 Mean   :-7.184   Mean   :-6.433                      Mean   :1.677
```

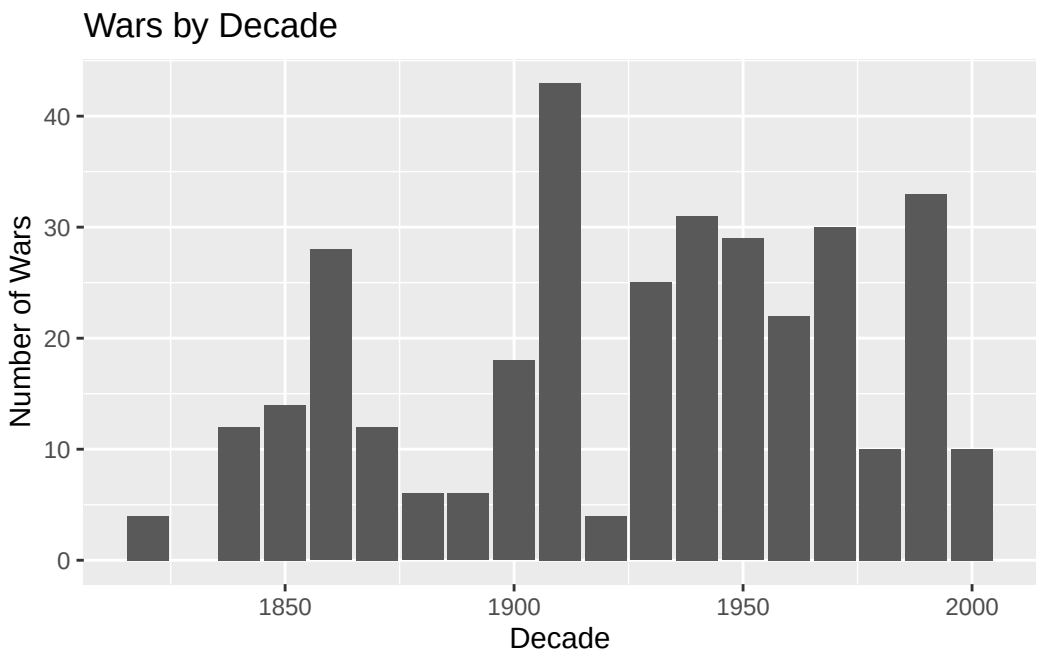
```
 3rd Qu.:-8.000   3rd Qu.:-8.000                      3rd Qu.:2.000
 Max.   :10.000   Max.   :31.000                      Max.   :2.000
 combat_location  combat_fatalities    outcome         next_war
 Min.   : 1.000   Min.   :     -9   Min.   :1.000   Min.   : -8.00
 1st Qu.: 2.000   1st Qu.:    400   1st Qu.:1.000   1st Qu.: -8.00
 Median : 6.000   Median :   2000   Median :2.000   Median : -8.00
 Mean   : 4.926   Mean   :  95196   Mean   :2.092   Mean   : 47.11
 3rd Qu.: 7.000   3rd Qu.:  10000   3rd Qu.:2.000   3rd Qu.: -8.00
 Max.   :19.000   Max.   :7500000   Max.   :8.000   Max.   :877.00
    duration        start_year      end_year
 Min.   : 0.000   Min.   :1823   Min.   :1823
 1st Qu.: 0.000   1st Qu.:1900   1st Qu.:1900
 Median : 0.000   Median :1939   Median :1941
 Mean   : 1.255   Mean   :1931   Mean   :1932
 3rd Qu.: 2.000   3rd Qu.:1969   3rd Qu.:1973
 Max.   :10.000   Max.   :2003   Max.   :2003
```

```r
wars %>%
  mutate(decade = floor(start_year1 / 10) * 10) %>%
  group_by(decade) %>%
  summarize(count = n())
```

```
# A tibble: 18 x 2
   decade count
    <dbl> <int>
 1   1820     4
 2   1840    12
 3   1850    14
 4   1860    28
 5   1870    12
 6   1880     6
 7   1890     6
 8   1900    18
 9   1910    43
10   1920     4
11   1930    25
12   1940    31
13   1950    29
14   1960    22
15   1970    30
16   1980    10
17   1990    33
```

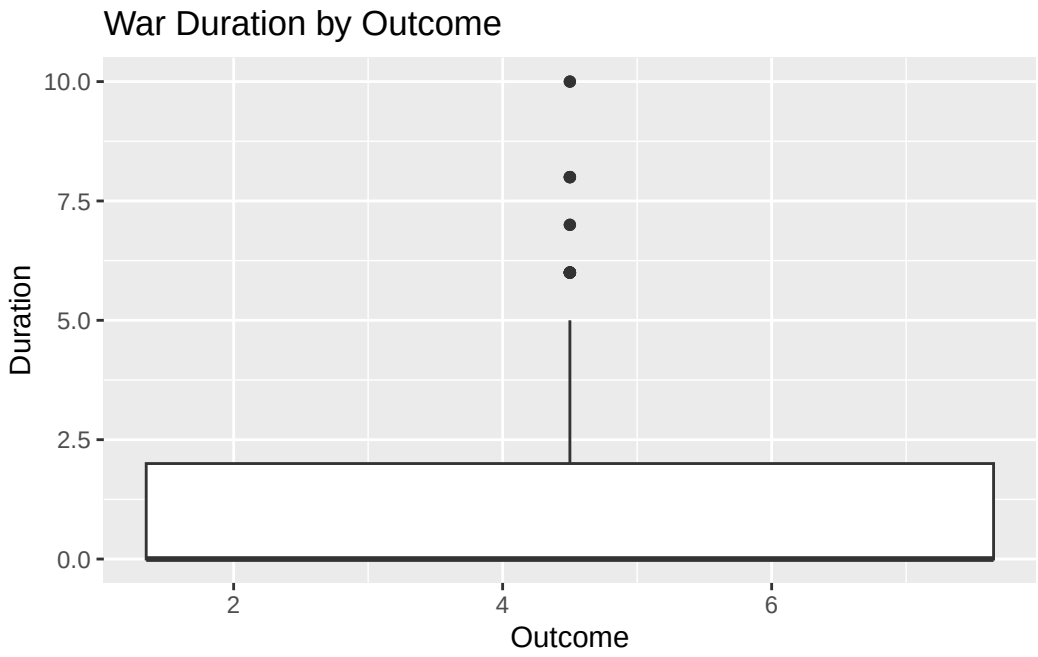
```
18    2000     10
```

```
wars %>%
  mutate(decade = floor(start_year / 10) * 10) %>%
  ggplot(aes(x = decade)) +
  geom_bar() +
  labs(title = "Wars by Decade", x = "Decade", y = "Number of Wars")
```




```
ggplot(wars, aes(x = outcome, y = duration)) +
  geom_boxplot() +
  labs(title = "War Duration by Outcome", x = "Outcome", y = "Duration")
```

```
Warning: Continuous x aesthetic
i did you forget `aes(group = ...)`?
```



Hypothesis Test

The results from the hypothesis test provide evidence that war duration varies significantly depending on the outcome of the conflict. Shorter wars were associated with decisive outcomes such as victory or defeat and longer wars tend to lead to inconclusive outcomes, such as stalemates or negotiated settlements.

The statistically significant p-value (p=0.0159) indicates that at least one outcome category has a mean duration that differs from the others. However, the hypothesis testing supports H1 (the alternative hypothesis) that war duration significantly influences outcomes. These findings highlight the importance of war duration as a determinant of conflict resolution and provide evidence for the hypothesis that time plays a critical role in shaping the trajectory and outcome of wars.

```
anova_result <- aov(duration ~ outcome, data = wars)
summary(anova_result)
```

```
             Df Sum Sq Mean Sq F value Pr(>F)
outcome       1   20.6  20.598    5.87 0.0159 *
Residuals   335 1175.5   3.509
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
library(caret)
```

```
Loading required package: lattice
```

```

Attaching package: 'caret'
```

```
The following object is masked from 'package:purrr':

    lift
```

```
library(randomForest)
```

```
randomForest 4.7-1.2
```

```
Type rfNews() to see new features/changes/bug fixes.
```

```

Attaching package: 'randomForest'
```

```
The following object is masked from 'package:dplyr':

    combine
```

```
The following object is masked from 'package:ggplot2':

    margin
```

```
library(xgboost)
```

```

Attaching package: 'xgboost'
```

```
The following object is masked from 'package:dplyr':

    slice
```

```
The following object is masked from 'package:tidygraph':

    slice
```

```r
wars$outcome <- as.factor(wars$outcome)
wars_clean <- wars %>%
  mutate(participant_count = str_count(state_name, ";") + 1)

set.seed(123)
train_index <- createDataPartition(wars_clean$outcome, p = 0.8, list = FALSE)
```

```
Warning in createDataPartition(wars_clean$outcome, p = 0.8, list = FALSE): Some
classes have a single record ( 8 ) and these will be selected for the sample
```

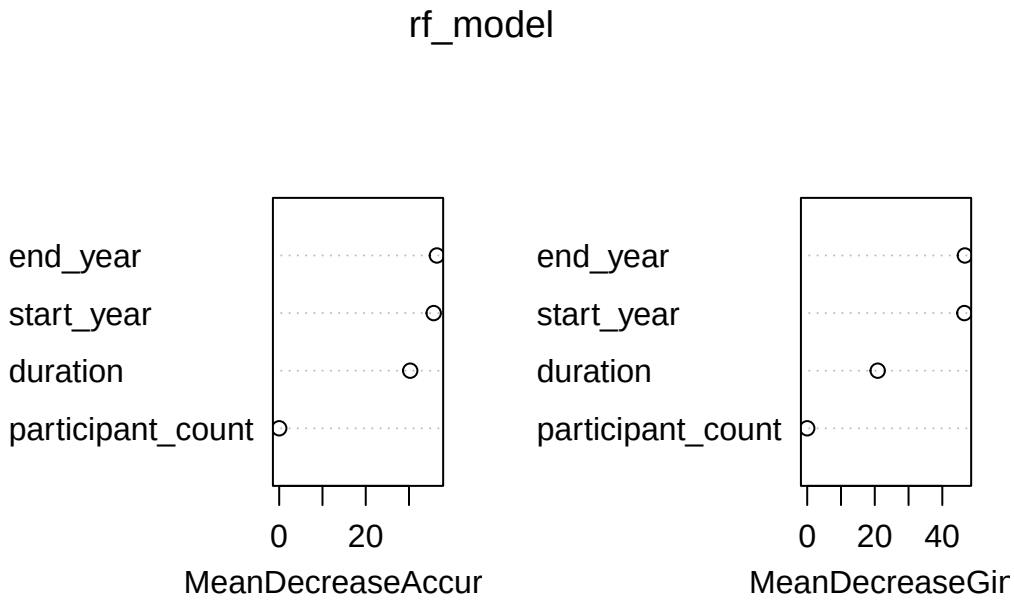
```r
train_data <- wars_clean[train_index, ]
test_data <- wars_clean[-train_index, ]
```

```r
rf_model <- randomForest(outcome ~ start_year + end_year + participant_count + duration,
                         data = train_data, importance = TRUE)

rf_predictions <- predict(rf_model, test_data)

confusionMatrix(rf_predictions, test_data$outcome)
```

```
Confusion Matrix and Statistics

          Reference
Prediction  1  2  3  4  6  8
         1 17 16  0  0  1  0
         2 13  6  0  0  0  0
         3  0  0  0  0  0  0
         4  1  1  0  5  0  0
         6  0  0  0  0  5  0
         8  0  0  0  0  0  0

Overall Statistics

               Accuracy : 0.5077
                 95% CI : (0.3807, 0.634)
    No Information Rate : 0.4769
    P-Value [Acc > NIR] : 0.3543
```

```
                  Kappa : 0.2207

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 1 Class: 2 Class: 3 Class: 4 Class: 6 Class: 8
Sensitivity            0.5484  0.26087       NA  1.00000  0.83333       NA
Specificity            0.5000  0.69048        1  0.96667  1.00000        1
Pos Pred Value         0.5000  0.31579       NA  0.71429  1.00000       NA
Neg Pred Value         0.5484  0.63043       NA  1.00000  0.98333       NA
Prevalence             0.4769  0.35385        0  0.07692  0.09231        0
Detection Rate         0.2615  0.09231        0  0.07692  0.07692        0
Detection Prevalence   0.5231  0.29231        0  0.10769  0.07692        0
Balanced Accuracy      0.5242  0.47567       NA  0.98333  0.91667       NA
```

```r
varImpPlot(rf_model)
```




```r
train_matrix <- model.matrix(outcome ~ . - 1, data = train_data)
test_matrix <- model.matrix(outcome ~ . - 1, data = test_data)
```

```r
train_label <- as.numeric(train_data$outcome) - 1
test_label <- as.numeric(test_data$outcome) - 1

common_features <- intersect(colnames(train_matrix), colnames(test_matrix))
train_matrix <- train_matrix[, common_features]
test_matrix <- test_matrix[, common_features]

xgb_train <- xgb.DMatrix(data = train_matrix, label = train_label)
xgb_test <- xgb.DMatrix(data = test_matrix, label = test_label)

xgb_model <- xgboost(
  data = xgb_train,
  max_depth = 6,
  eta = 0.3,
  nrounds = 100,
  objective = "multi:softmax",
  num_class = length(unique(train_label))
)
```

```
[1] train-mlogloss:1.108181
[2] train-mlogloss:0.787978
[3] train-mlogloss:0.585637
[4] train-mlogloss:0.445300
[5] train-mlogloss:0.344781
[6] train-mlogloss:0.271050
[7] train-mlogloss:0.217163
[8] train-mlogloss:0.175527
[9] train-mlogloss:0.143182
[10]    train-mlogloss:0.117929
[11]    train-mlogloss:0.098756
[12]    train-mlogloss:0.084242
[13]    train-mlogloss:0.072627
[14]    train-mlogloss:0.063540
[15]    train-mlogloss:0.055604
[16]    train-mlogloss:0.049727
[17]    train-mlogloss:0.044456
[18]    train-mlogloss:0.040475
[19]    train-mlogloss:0.036995
[20]    train-mlogloss:0.034369
[21]    train-mlogloss:0.032200
[22]    train-mlogloss:0.030385
[23]    train-mlogloss:0.028803
```

```
[24]    train-mlogloss:0.027476
[25]    train-mlogloss:0.026434
[26]    train-mlogloss:0.025508
[27]    train-mlogloss:0.024820
[28]    train-mlogloss:0.024010
[29]    train-mlogloss:0.023395
[30]    train-mlogloss:0.022725
[31]    train-mlogloss:0.022159
[32]    train-mlogloss:0.021688
[33]    train-mlogloss:0.021311
[34]    train-mlogloss:0.020839
[35]    train-mlogloss:0.020483
[36]    train-mlogloss:0.020110
[37]    train-mlogloss:0.019789
[38]    train-mlogloss:0.019527
[39]    train-mlogloss:0.019184
[40]    train-mlogloss:0.018946
[41]    train-mlogloss:0.018649
[42]    train-mlogloss:0.018362
[43]    train-mlogloss:0.018116
[44]    train-mlogloss:0.017873
[45]    train-mlogloss:0.017682
[46]    train-mlogloss:0.017484
[47]    train-mlogloss:0.017314
[48]    train-mlogloss:0.017188
[49]    train-mlogloss:0.017016
[50]    train-mlogloss:0.016912
[51]    train-mlogloss:0.016745
[52]    train-mlogloss:0.016630
[53]    train-mlogloss:0.016479
[54]    train-mlogloss:0.016339
[55]    train-mlogloss:0.016239
[56]    train-mlogloss:0.016126
[57]    train-mlogloss:0.016006
[58]    train-mlogloss:0.015917
[59]    train-mlogloss:0.015790
[60]    train-mlogloss:0.015663
[61]    train-mlogloss:0.015568
[62]    train-mlogloss:0.015483
[63]    train-mlogloss:0.015371
[64]    train-mlogloss:0.015310
[65]    train-mlogloss:0.015223
[66]    train-mlogloss:0.015159
```

```
[67]    train-mlogloss:0.015078
[68]    train-mlogloss:0.015029
[69]    train-mlogloss:0.014961
[70]    train-mlogloss:0.014893
[71]    train-mlogloss:0.014849
[72]    train-mlogloss:0.014785
[73]    train-mlogloss:0.014724
[74]    train-mlogloss:0.014652
[75]    train-mlogloss:0.014593
[76]    train-mlogloss:0.014537
[77]    train-mlogloss:0.014464
[78]    train-mlogloss:0.014410
[79]    train-mlogloss:0.014348
[80]    train-mlogloss:0.014294
[81]    train-mlogloss:0.014240
[82]    train-mlogloss:0.014189
[83]    train-mlogloss:0.014140
[84]    train-mlogloss:0.014091
[85]    train-mlogloss:0.014038
[86]    train-mlogloss:0.013989
[87]    train-mlogloss:0.013944
[88]    train-mlogloss:0.013900
[89]    train-mlogloss:0.013823
[90]    train-mlogloss:0.013767
[91]    train-mlogloss:0.013737
[92]    train-mlogloss:0.013661
[93]    train-mlogloss:0.013613
[94]    train-mlogloss:0.013565
[95]    train-mlogloss:0.013522
[96]    train-mlogloss:0.013479
[97]    train-mlogloss:0.013450
[98]    train-mlogloss:0.013397
[99]    train-mlogloss:0.013370
[100]   train-mlogloss:0.013344
```

```
xgb_predictions <- predict(xgb_model, xgb_test)

confusionMatrix(as.factor(xgb_predictions), as.factor(test_label))
```

```
Confusion Matrix and Statistics

          Reference
```

```
Prediction  0  1  3  4
         0 30  0  0  0
         1  1 23  0  1
         3  0  0  5  0
         4  0  0  0  5

Overall Statistics

               Accuracy : 0.9692
                 95% CI : (0.8932, 0.9963)
    No Information Rate : 0.4769
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9512

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 0 Class: 1 Class: 3 Class: 4
Sensitivity            0.9677   1.0000  1.00000  0.83333
Specificity            1.0000   0.9524  1.00000  1.00000
Pos Pred Value         1.0000   0.9200  1.00000  1.00000
Neg Pred Value         0.9714   1.0000  1.00000  0.98333
Prevalence             0.4769   0.3538  0.07692  0.09231
Detection Rate         0.4615   0.3538  0.07692  0.07692
Detection Prevalence   0.4615   0.3846  0.07692  0.07692
Balanced Accuracy      0.9839   0.9762  1.00000  0.91667
```

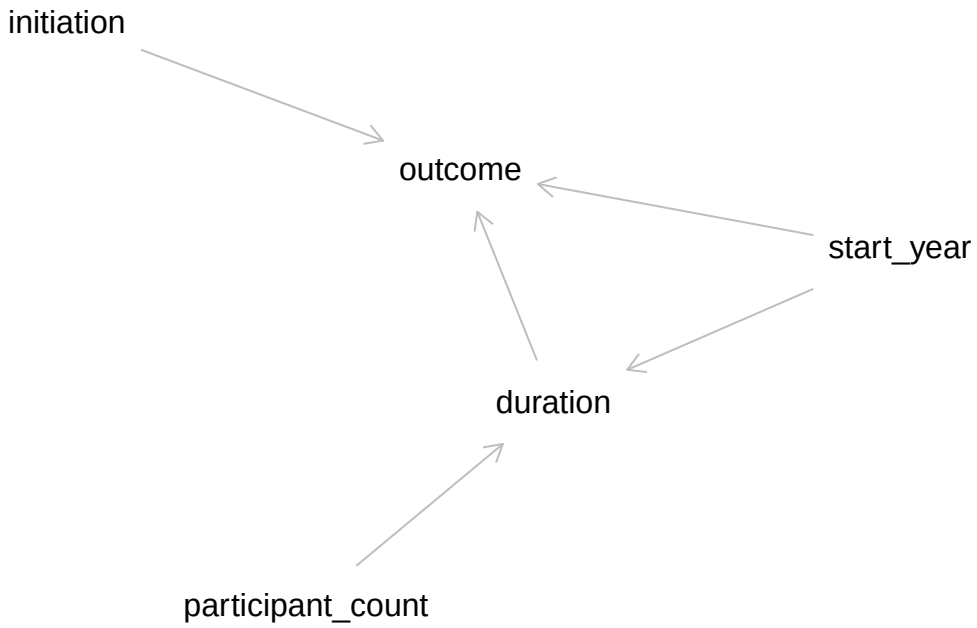
```
library(dagitty)
```

```
Attaching package: 'dagitty'

The following object is masked from 'package:tidygraph':

    convert
```

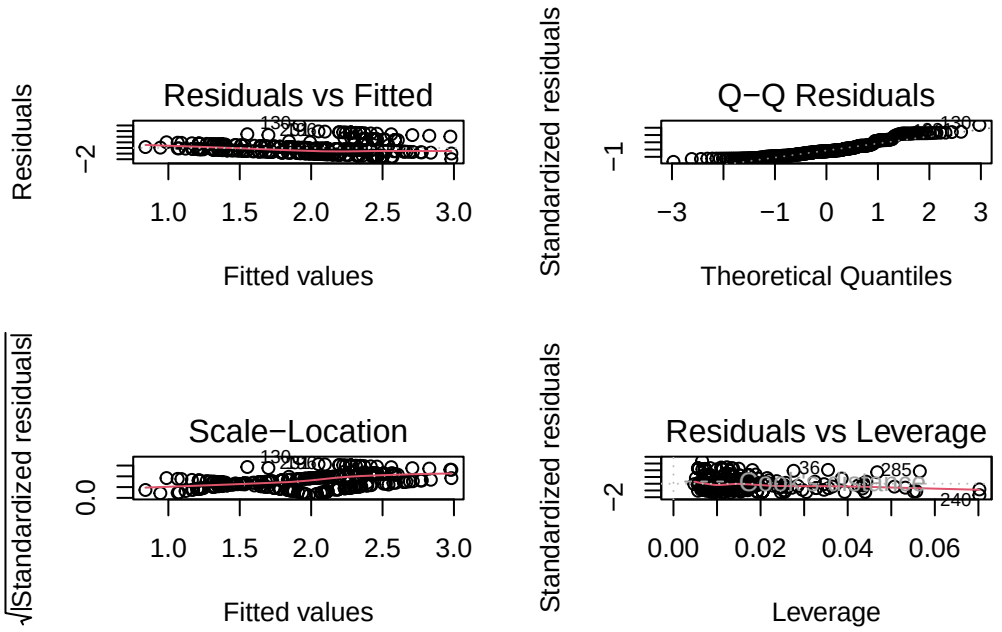
```
dag <- dagitty("
dag {
  participant_count -> duration
```

```
  duration -> outcome
  initiation -> outcome
  start_year -> duration
  start_year -> outcome
}")

plot(dag)
```

```
Plot coordinates for graph not supplied! Generating coordinates, see ?coordinates for how to
```




```
wars$outcome_numeric <- as.numeric(wars$outcome)

warmodel <- lm(outcome_numeric ~ duration + combat_location + initiation + start_year1, data

par(mfrow = c(2, 2))
plot(warmodel)
```



R-Squared Value

```
summary(warmodel)$r.squared
```

```
[1] 0.1230247
```

The R-Squared value of 0.123 suggests that only 12.3% of the variance in the dependent variable is explained by the model. This value suggests that a large proportion of variance remains unexplained, it is important to note that social science phenomena, such as war outcomes, are often influenced by a multitude of complex and unobserved factors. Despite the low R-Squared value from running the code, the statistical significance of the predictor variables demonstrates that war duration and other factors have an effect on outcomes.

The low R-Squared value highlights the complexity of modeling war outcomes. Many unobserved factors—such as geopolitical alliances, economic pressures, or military strategies—likely contribute to the unexplained variance. Future research could enhance the predictive power of the model by incorporating additional variables that capture these dimensions. However, the results from this final project could still provide valuable insights, particularly regarding the role of war duration in shaping outcomes.

```r
table(wars$outcome_binary)
```

```
< table of extent 0 >
```

```r
summary(wars)
```

```
     war_id        war_name            war_type   state_code
 Min.   :  1.0   Length:337         Min.   :1   Min.   :  2.0
 1st Qu.: 82.0   Class :character   1st Qu.:1   1st Qu.:220.0
 Median :139.0   Mode  :character   Median :1   Median :355.0
 Mean   :126.9                      Mean   :1   Mean   :419.9
 3rd Qu.:172.0                      3rd Qu.:1   3rd Qu.:652.0
 Max.   :227.0                      Max.   :1   Max.   :920.0
  state_name             side        start_year1    start_month1
 Length:337         Min.   :1.000   Min.   :1823   Min.   : 1.000
 Class :character   1st Qu.:1.000   1st Qu.:1900   1st Qu.: 3.000
 Mode  :character   Median :1.000   Median :1939   Median : 6.000
                    Mean   :1.418   Mean   :1931   Mean   : 6.249
                    3rd Qu.:2.000   3rd Qu.:1969   3rd Qu.: 9.000
                    Max.   :2.000   Max.   :2003   Max.   :12.000
   start_day1      end_year1      end_month1        end_day1
 Min.   : 1.00   Min.   :1823   Min.   : 1.000   Min.   : 1.00
 1st Qu.: 7.00   1st Qu.:1900   1st Qu.: 4.000   1st Qu.:10.00
 Median :16.00   Median :1941   Median : 7.000   Median :14.00
 Mean   :15.35   Mean   :1932   Mean   : 6.623   Mean   :15.04
 3rd Qu.:23.00   3rd Qu.:1973   3rd Qu.: 9.000   3rd Qu.:23.00
 Max.   :31.00   Max.   :2003   Max.   :12.000   Max.   :31.00
  start_year2      start_month2      start_day2       end_year2
 Min.   :  -8.0   Min.   :-8.000   Min.   :-8.000   Min.   :  -8.0
 1st Qu.:  -8.0   1st Qu.:-8.000   1st Qu.:-8.000   1st Qu.:  -8.0
 Median :  -8.0   Median :-8.000   Median :-8.000   Median :  -8.0
 Mean   : 100.3   Mean   :-7.169   Mean   :-6.647   Mean   : 100.3
 3rd Qu.:  -8.0   3rd Qu.:-8.000   3rd Qu.:-8.000   3rd Qu.:  -8.0
 Max.   :1974.0   Max.   :10.000   Max.   :25.000   Max.   :1974.0
   end_month2        end_day2      previous_war         initiation
 Min.   :-8.000   Min.   :-8.000   Length:337         Min.   :1.000
 1st Qu.:-8.000   1st Qu.:-8.000   Class :character   1st Qu.:1.000
 Median :-8.000   Median :-8.000   Mode  :character   Median :2.000
 Mean   :-7.184   Mean   :-6.433                      Mean   :1.677
 3rd Qu.:-8.000   3rd Qu.:-8.000                      3rd Qu.:2.000
 Max.   :10.000   Max.   :31.000                      Max.   :2.000
```

```
 combat_location  combat_fatalities outcome    next_war         duration
 Min.   : 1.000   Min.   :     -9   1:155   Min.   : -8.00   Min.   : 0.000
 1st Qu.: 2.000   1st Qu.:    400   2:119   1st Qu.: -8.00   1st Qu.: 0.000
 Median : 6.000   Median :   2000   3:  4   Median : -8.00   Median : 0.000
 Mean   : 4.926   Mean   :  95196   4: 28   Mean   : 47.11   Mean   : 1.255
 3rd Qu.: 7.000   3rd Qu.:  10000   6: 30   3rd Qu.: -8.00   3rd Qu.: 2.000
 Max.   :19.000   Max.   :7500000   8:  1   Max.   :877.00   Max.   :10.000
   start_year      end_year    outcome_numeric
 Min.   :1823   Min.   :1823   Min.   :1.000
 1st Qu.:1900   1st Qu.:1900   1st Qu.:1.000
 Median :1939   Median :1941   Median :2.000
 Mean   :1931   Mean   :1932   Mean   :1.997
 3rd Qu.:1969   3rd Qu.:1973   3rd Qu.:2.000
 Max.   :2003   Max.   :2003   Max.   :6.000
```

References

Aliyev, H. (2017). Pro-regime militias and Civil War Duration. Terrorism and Political Violence, 32(3), 630–650. https://doi.org/10.1080/09546553.2017.1393415

Cirillo, P., & Taleb, N. N. (2016). On the statistical properties and tail risk of violent conflicts. Physica A: Statistical Mechanics and Its Applications, 452, 29–45. https://doi.org/10.1016/j.physa.2016.01.050

Civitello, L. (2018). The price war. University of Illinois Press. https://doi.org/10.5406/illinois/9780252041082.0

Clauset, A. (2018). Trends and fluctuations in the severity of Interstate Wars. Science Advances, 4(2). https://doi.org/10.1126/sciadv.aao3580

de Rouen, K. R., & Sobek, D. (2004). The dynamics of Civil War duration and outcome. Journal of Peace Research, 41(3), 303–320. https://doi.org/10.1177/0022343304043771

Dudley, R. (2023). Turning the tables: Military intervention and the onset of negotiations in Civil War. Journal of Conflict Resolution, 68(6), 1139–1167. https://doi.org/10.1177/00220027231190915

Li, W., Bradshaw, A. E., Clary, C. B., & Cranmer, S. J. (2017). A three-degree horizon of peace in the Military Alliance Network. Science Advances, 3(3). https://doi.org/10.1126/sciadv.1601895

Reiter, D. (2018). Unifying the study of the causes and duration of wars. Polity, 50(2), 168–177. https://doi.org/10.1086/696288

Richardson, L. F. (1944). The distribution of wars in time. Journal of the Royal Statistical Society, 107(3/4), 242. https://doi.org/10.2307/2981216

Shirkey, Z. C. (2012). When and how many: The effects of third party joining on casualties and duration in Interstate Wars. Journal of Peace Research, 49(2), 321–334. https://doi.org/10.1177/0022343311431597

Weisiger, A. (2016). Learning from the battlefield: Information, domestic politics, and Interstate War Duration. International Organization, 70(2), 347–375. https://doi.org/10.1017/s0020818316000059