

What Results in Death?

Analysis of Social Conflict in Africa, 1990–2011

Matthew Boyas

STAT 222: MA Capstone Midterm Report
March 21, 2014





Data Sources

- Social Conflict in Africa Database (SCAD)
 - Cullen Hendrix and Idean Salehyan
 - Hosted by Climate Change and African Political Stability (CCAPS) at the Robert S. Strauss Center for International Security and Law at the University of Texas at Austin
- Correlates of War Project (COW)
 - National Material Capabilities
 - World Religions
- Polity IV Project
 - Measures democracy/autocracy for government regime type



Research Questions

1. What differentiates an episode of social conflict that results in deaths from an episode of social conflict that does not result in deaths?
2. Is there a way to predict the number of deaths that will result from an episode of social conflict?

Figure 1: Number of Death and No Death Conflicts by Year

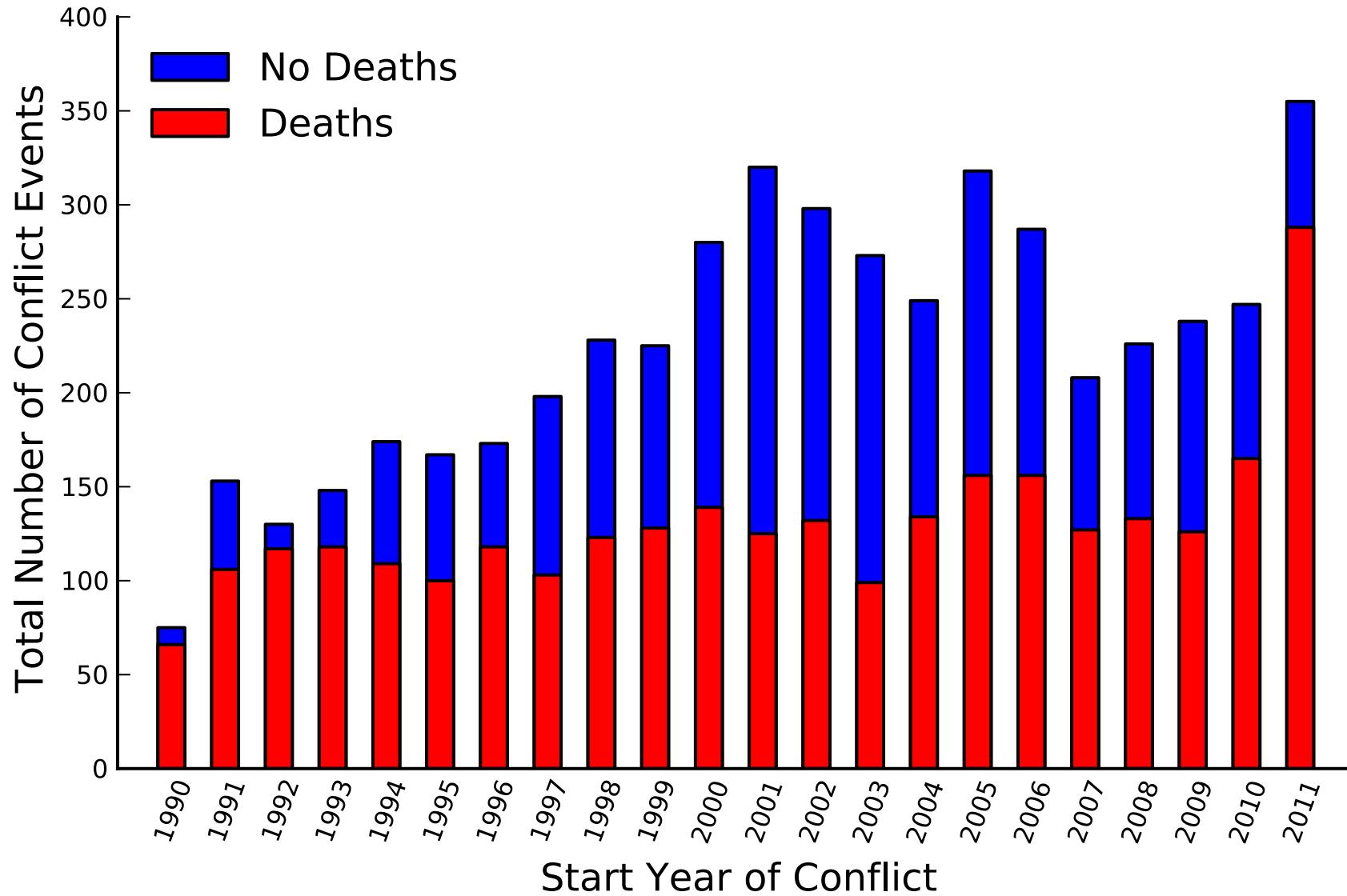
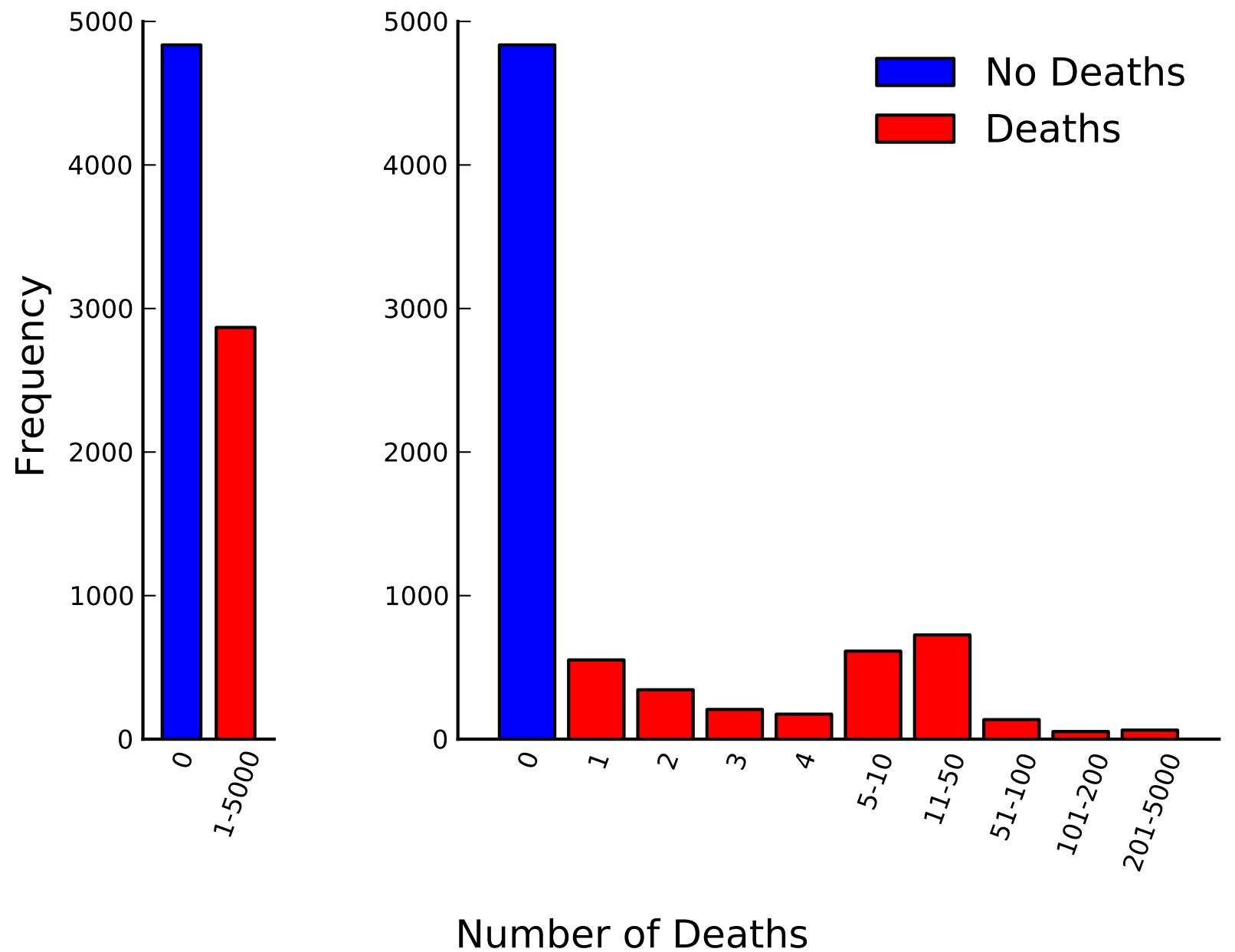


Table 1: Top 10 Most Violent Conflicts

	Deaths	Country	Start	End	Duration (days)
1	5000	Democratic Republic of the Congo	1999-09-01	1999-12-31	122
2	3500	Democratic Republic of the Congo	1993-03-20	1993-08-31	165
3	3000	Ghana	1994-02-04	1994-02-14	11
4	3000	Nigeria	1998-01-27	1998-09-15	232
5	2000	Rwanda	1995-04-24	1995-04-24	1
6	1800	Nigeria	1992-05-15	1992-05-20	6
7	1500	Democratic Republic of the Congo	2009-03-30	2009-06-30	93
8	1500	Liberia	1996-04-06	1996-05-27	52
9	1400	Democratic Republic of the Congo	2002-04-30	2002-06-11	43
10	1132	South Africa	1990-08-12	1990-10-01	51

Figure 2: Distribution of the Number of Deaths



Question #1

What differentiates an episode of social conflict that results in deaths from an episode of social conflict that does not result in deaths?



Table 2: Number of Death/No Death Conflicts by Dominant Religion

	Islam	Christianity	Animist	Hinduism
No Deaths	2339 (61.78%)	2333 (62.75%)	159 (81.96%)	5 (83.33%)
Deaths	1447 (38.22%)	1385 (37.25%)	35 (18.04%)	1 (16.67%)
Total	3786 (100.0%)	3718 (100.0%)	194 (100.0%)	6 (100.0%)

Table 3: Number of Death/No Death Conflicts by Regime Type

	Strong Autocracy	Weak Autocracy	Middle Ground	Weak Democracy	Strong Democracy
No Deaths	798 (76.36%)	1401 (69.84%)	74 (77.08%)	1010 (57.29%)	928 (65.58%)
Deaths	247 (23.64%)	605 (30.16%)	22 (22.92%)	753 (42.71%)	487 (34.42%)
Total	1045 (100.0%)	2006 (100.0%)	96 (100.0%)	1763 (100.0%)	1415 (100.0%)

Figure 3: Deaths by Conflict Size

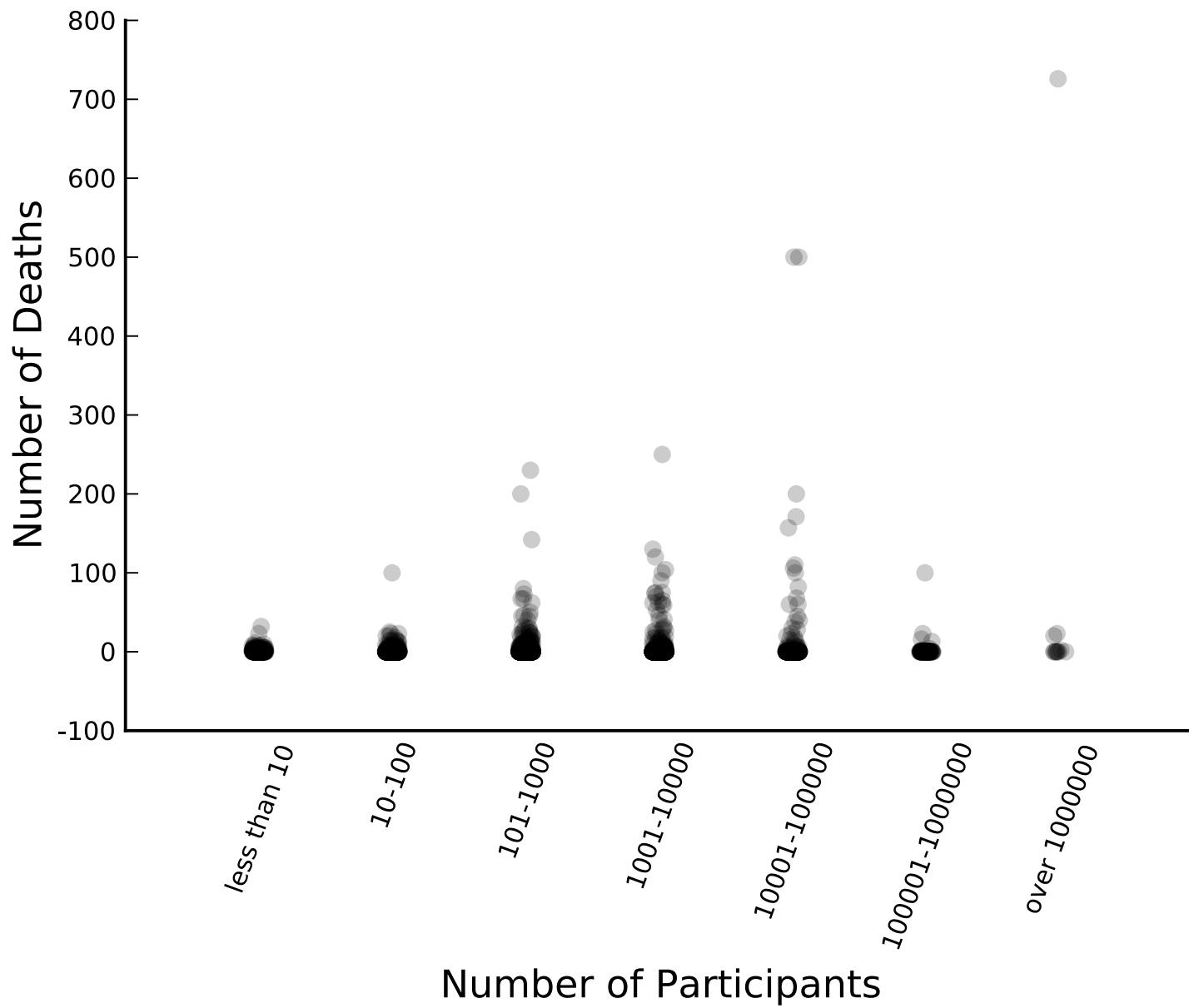


Figure 4: Deaths by Conflict Duration

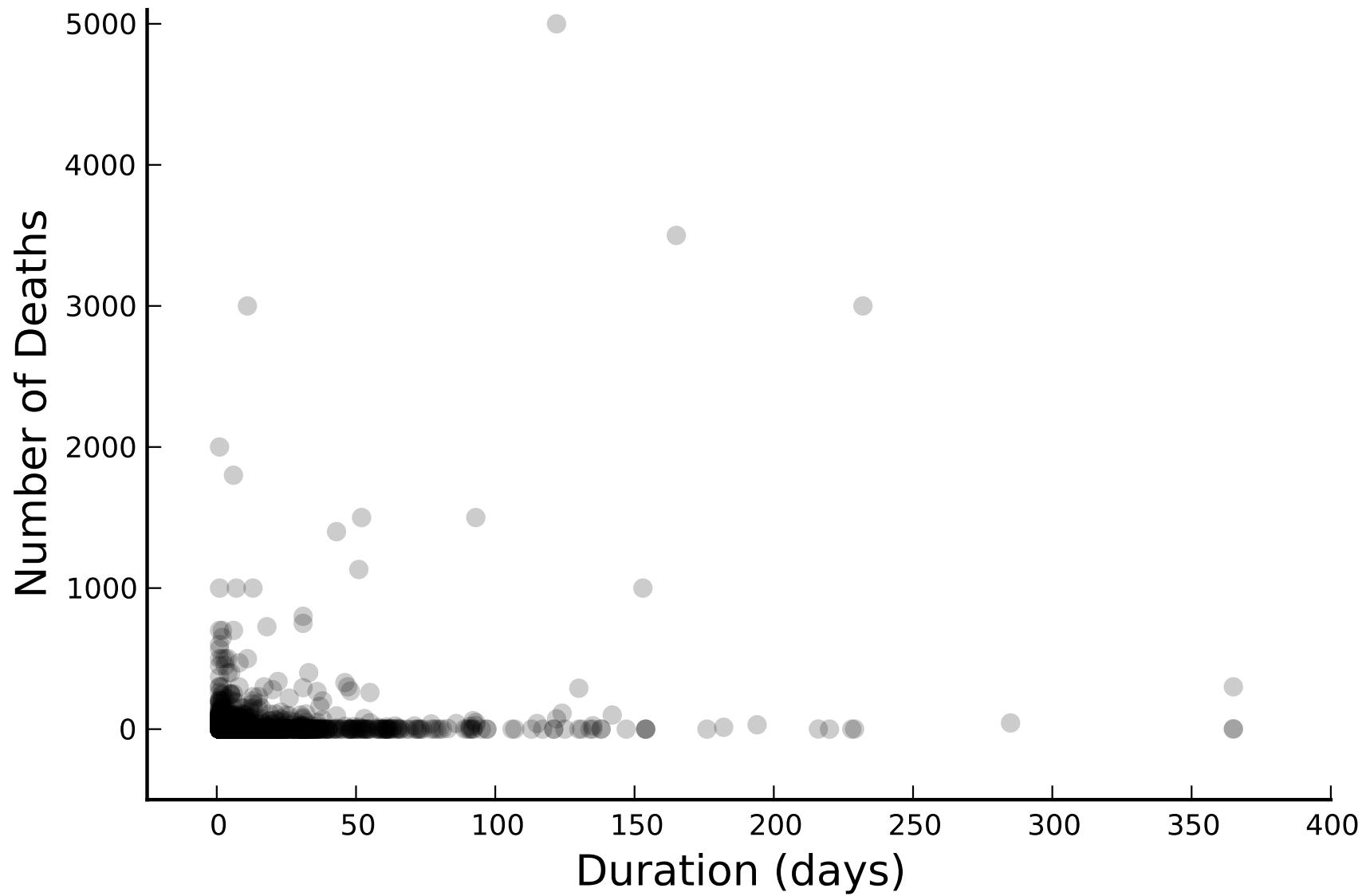


Figure 5: Map of Conflict Events

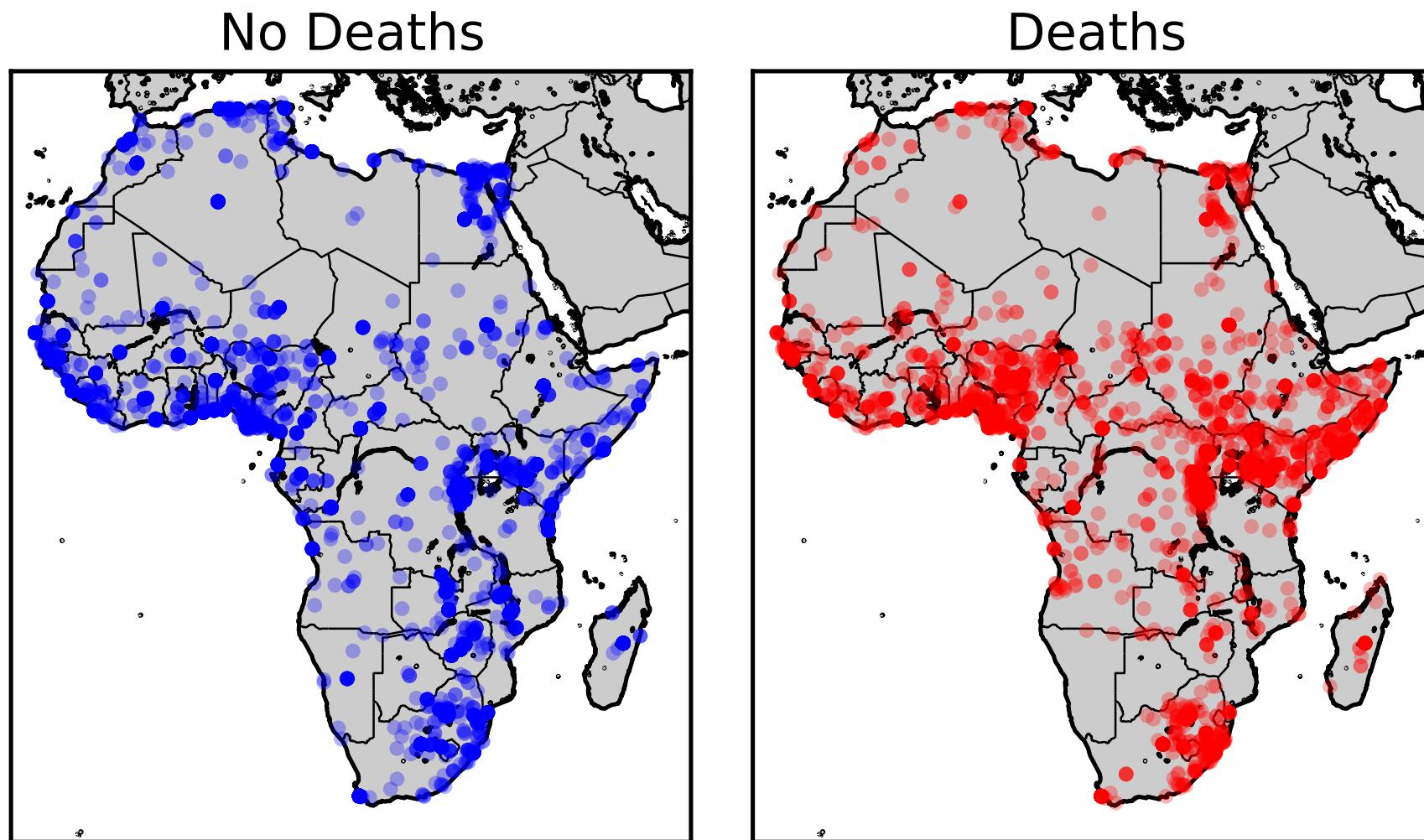
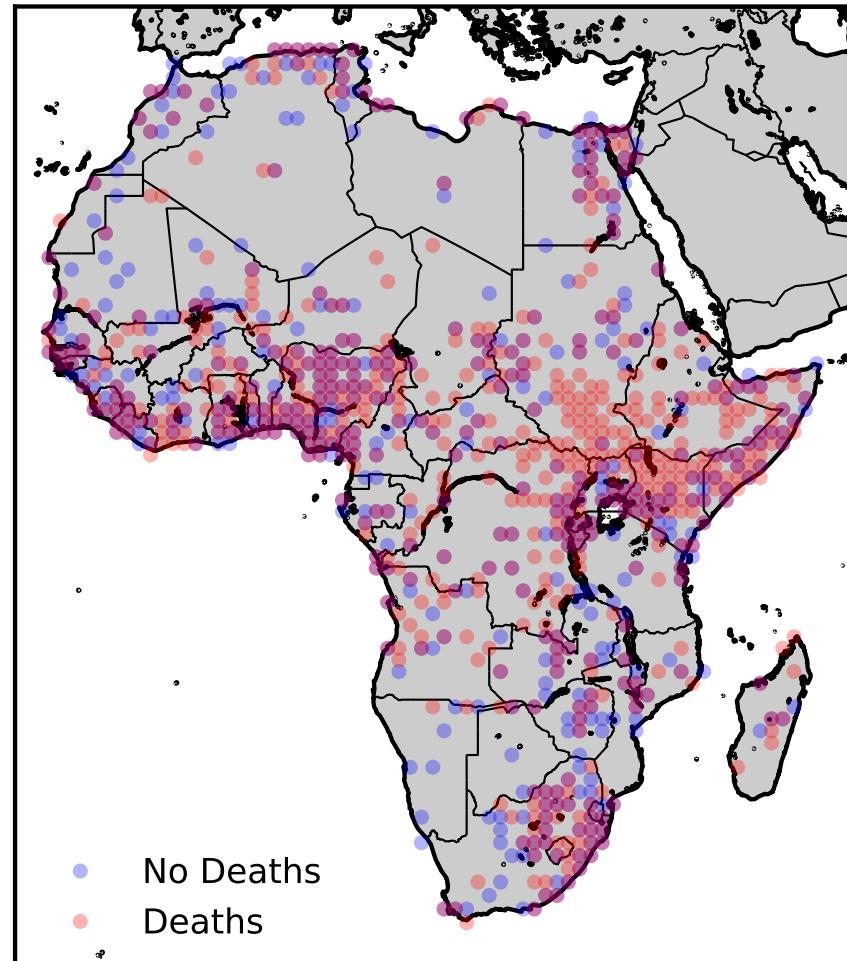


Figure 6: Map of Superimposed Deaths and No Deaths



Question #2

Is there a way to predict the number of deaths that will result from an episode of social conflict?





Binary Prediction

- Predict a binary, death/no-death indicator
- Separate dataset into train (70%) and test (30%)
- Logistic regression
 - Location
 - Event type
 - Central government target
 - Primary issue
 - National capability score
- Compare to KNN



Prediction Accuracy

Table 4: Logistic Regression Prediction Accuracy

	Predicted as No Deaths	Predicted as Deaths
No Deaths	957 (89.36%)	114 (10.64%)
Deaths	170 (30.41%)	389 (69.59%)

Table 5: KNN Prediction Accuracy (K=10)

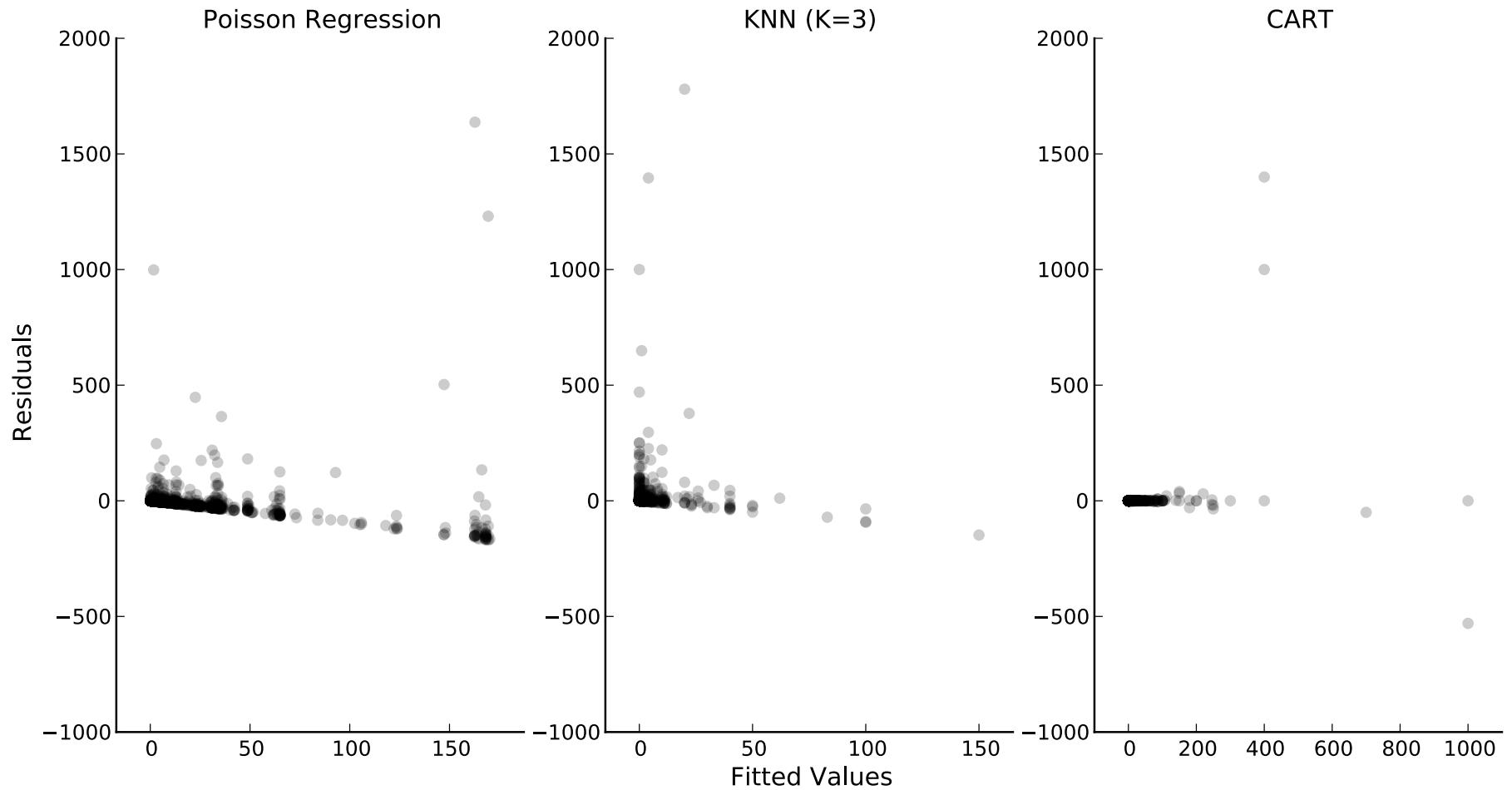
	Predicted as No Deaths	Predicted as Deaths
No Deaths	990 (92.44%)	81 (7.56%)
Deaths	225 (40.25%)	334 (59.75%)



Number of Deaths Prediction

- Predict the actual number of deaths
- Use same train/test sets from before
- Poisson regression
 - Similar model selection process as logistic
 - Same set of variables minimizes prediction error
- Compare to KNN and CART Decision Tree

Figure 7: Prediction Accuracy (Fitted Values vs. Residuals)





Acknowledgements & Questions

- Victoria Stodden
- Christine Ho
- Ryan Lovett

Matt Boyas
Stat 222 Midterm Report—Deviations from Proposal

This document is intended to accompany the included work and serves to explain some of the deviations from my original research proposal.

(1)

It made sense to me to lump together my first and third research questions, for as I began investigating question #1, I was also testing many of the variables that I wanted to look at in question #3.

(2)

Some of the titles of plots and tables have been altered or completely changed because I discovered that the titles no longer really worked once I had an actual figure/table at which to look.

(3)

The originally-proposed Figures 3 & 4 were changed to scatterplots. I was not sure how to calculate the original proportions for conflicts with zero deaths, and the scatterplots seemed to convey a very similar idea.

(4)

I chose not to include the regression summary due to the fact that three of the five variables are categorical, split into 7-10 binaries. As a result, there are a lot of variables listed with names that do not mean all that much, so I plan on just summarizing the regression summary in the text of my paper and perhaps including the entire summary output as an appendix.

(5)

Many of the proposed figures/tables relating to the regression were somewhat hypothetical, and as my modeling methods changed, so did some of the plots. For example, I chose not to include normal Q-Q plots due to not running linear regression. I also did not include fitted value vs. residual plots for the training dataset in favor of focusing on clear plots/tables regarding prediction accuracy (I did not want to have too many similar looking figures/tables in my paper for fear that overall meaning and flow of the paper would be affected.)

(6)

The figure to show prediction accuracy from the number of deaths model was modified to be a fitted value vs. residual plot for the three modeling methods that I attempted.