Final Data Set Project for Political Science 811 and 812

Karla Flores[1]

[1] University of Wisconsin-Madison

Author Note

Political Science Ph.D. Student at the University of Wisconsin-Madison

Abstract

Inter-State wars can be noted to have decreased since the end of World War II, but despite this decrease this does not mean that they have ended all together. Due to this observation, I sought to use a data set that included interstate wars. For the purpose of this descriptive analysis paper, I am using inter-state data set by the Correlates of War Project- version 4. The dataset is provided by Meredith Reid Sarkees and Frank Wayman. In particular, it focusses on wars that took place from 1816-2007. Unfortuanltey, this dataset holds more categorical varibales as opposed to numerical varaibles, and given this it may be difficult to create an inferal data analysis. Nevertheless, in being an International Relations student I wanted to use this data set to apply different statistical conceptual methods we have discussed in class along with use certain application of functions we have learned in the 811 course as well.

Keywords: Inter-State Wars, Mean, Standard Deviation, Summary Statistics, and Graphs

Final Data Set Project for Political Science 811 and 812

Statistical Methods Discussed in Class

```
## [1] TRUE
```

```
## [1] 95195.55
```

Two functions were used above. The first involved ensuring that the variable I planned to use was in fact numerical and second involved finding the mean of said variable. In this case, the battle death variable is one of the few numerical variables in this data set that I will focus on. As we may be able to note by our variable that accounts for the battle deaths that have occurred across all interstate wars, the average number of deaths appears to be 9,5196 people. And as we have learned in our statistics course, the mean of this variable refers to the sum of the total number of deaths for each war that is then divided by the total number of wars.

```
## [1] 2000
```

The number above corresponds to the median of battle deaths. As for the median of our battle death variable, this process entails placing all of the number of deaths in numerical order from smallest to largest and locating the center number. In theory, this would locate the number in the middle. With the help of R, we are able to learn that the median number of battle deaths is 2,000.

```
## [1] 501199.7
```

One of the last concepts we discussed early into the course was that of standard deviation. In the case of our command above, it helps illustrate the dispersion of this variable in relation to the mean.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
##       -9      400     2000    95196    10000 7500000
```

While I showed different parts of this summary previously, I wanted to take the time to illustrate that this is one of the things we learned in our course as well. Both Stata and R offer us a function that is able to create a summary table for certain variables we choose. This is what the summary table looks like for the battle deaths and includes some of the concepts discussed further in depth previously.

Similarly, I have accounted for the summary tables of two variables that provide the years of when wars begun and ended to further demonstrate my understanding of function application.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1823    1900    1939    1931    1969    2003
```

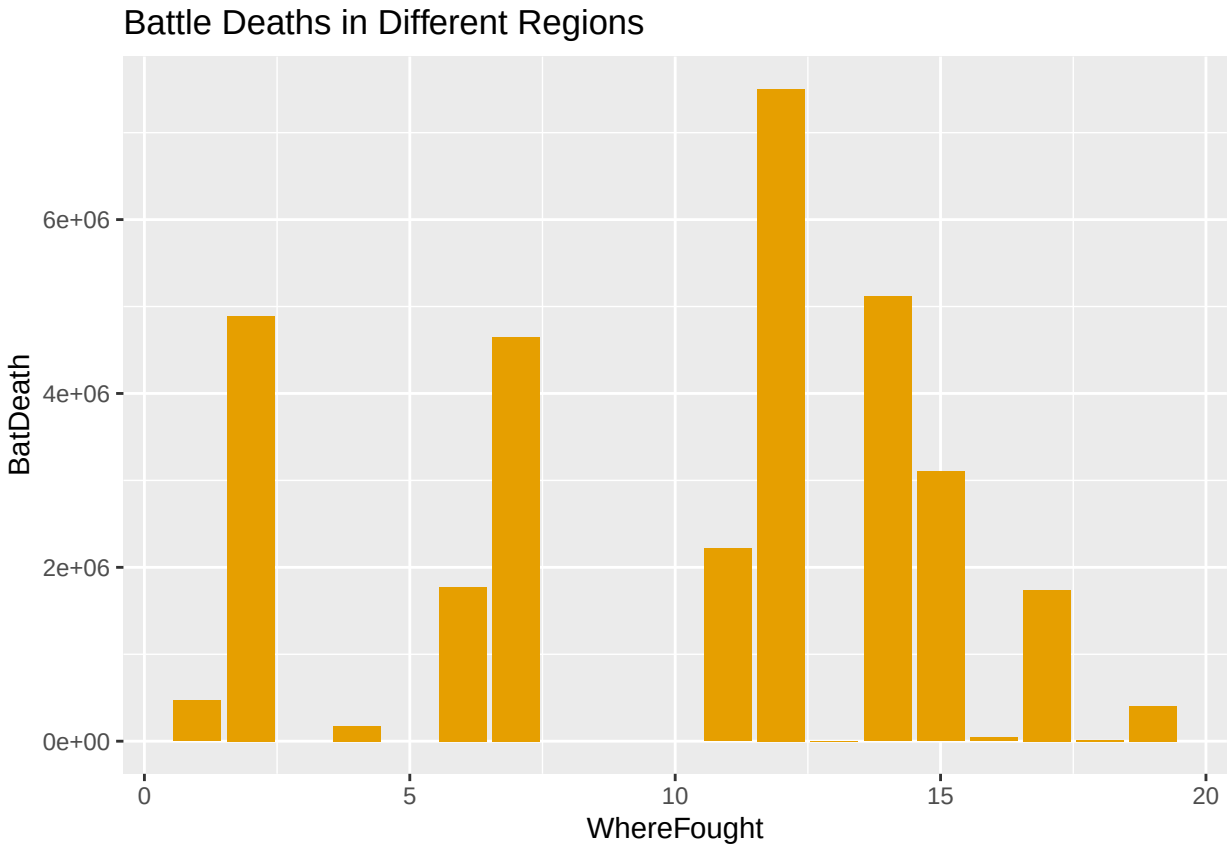
The table above corresponds to the summary of the years in which wars begun. The average year in which wars begun as is provided by the mean is 1931. And the median number is 1939.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1823    1900    1941    1932    1973    2003
```

The table above corresponds to the summary of the years wars ended. And as we can tell, the average year for this variable was 1932 and the median is 1941. I chose to help provide this summary given that this includes another variable that I will use later on.

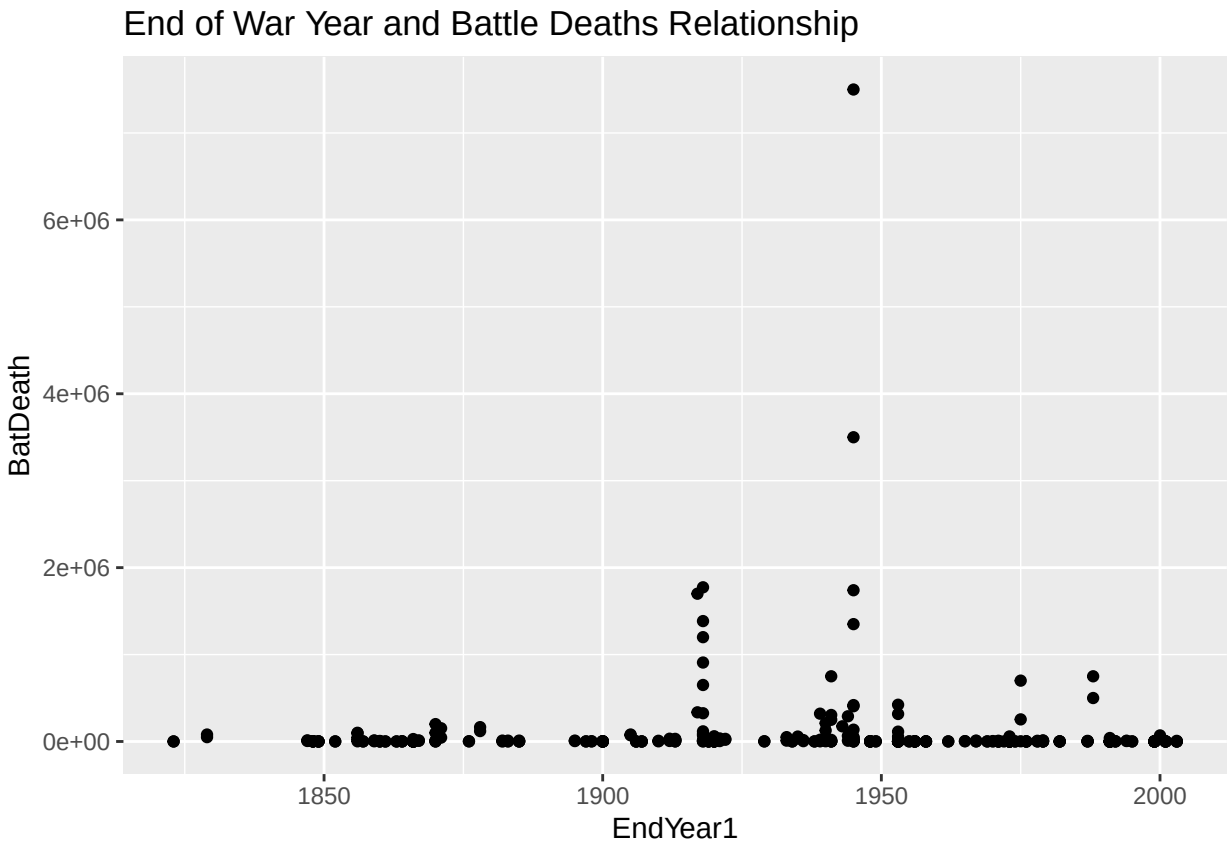
Next, I will attempt to illustrate some relationships and findings through the use of graphs.

Statistical Figures/Graphs Discussed throughout the Course



As the graph title indicate, this is a figure illustrating the relationship between battle deaths and where in the world these deaths were seen to occur according to the region in which wars were fought. I was interested to see where most wars were fought, but also how that may have corresponded to the deaths in said regions. According to the code book each of the numbers for where of the wars were fought in correspondence to the region are the following: 1 = Western Hemisphere, 2 = Europe, 4 = Africa, 6 = Middle East, 7 = Asia, 9 = Oceania. Through this bar graph in particular we are able to observe that the most battle deaths were seen to occur within wars were fought in the Oceania region. This may be due to many World War II battles taking place in this region. Nevertheless, it is interesting to see that the Western Hemisphere does not

Next, I sought to see and attempt to understand if any relationship existed between the year wars ended and the number of battle deaths.



While I must take into account the limitations of my graph given that it does not account for the trajectory of the deaths from when the war begun in its entirety, we are still able to notice some interesting observations. In particular, we are able to observe what years resulted in greater battle deaths. As this scatter plot helps illustrate, we are able to see the number of deaths increased in the 1910-1920 years and the 1940's. This observation in particular makes sense once we take into account world history, which would help indicate that during those two time periods World War I and World War II took place. This would also allow and help explain why there may have been higher death battles during the time frame- especially for World War II which resulted in one of the highest numbers of moralities caused by a war.

Conclusion

As I discussed previously, my goal in this paper was to demonstrate a general understanding of both statistical methods discussed in our course along with illustrate an application and use of certain functions discussed in the 811 course. It is my understanding

as well that this data set continues to be built upon by both the Correlates of War project, but also by other scholars- namely, Eric Min at the University of California, Los Angeles. As this project continues to advance and hopefully, as more numerical values become accessible, this may create an avenue for even more inferential research questions. Given the limitations of this final project as well, I chose to focus on some of the most basic concepts discussed in class in order to help demonstrate my own conceptualization and understanding of the information taught in both classes- 811 and 812. I found it interesting to be able to create certain graphs to understand if any relationship (at all) existed within certain variables and whether the observations held any significance when taking into account history. Moving forward, I would like to look into this data with greater depth and perhaps even work on merging two data sets. As for now, here are some of my findings and general understandings.