Final Project Distribution

The folder in which this document resides contains all the deliverables for the INFO 480-Intro to Data science Final Project.  
  
The Dataset consists of “Military Casualties Final Dataset.xlsx”. The data contains records of all record military deaths from 2001-2012. A table containing encoded values for the country code is included as well. The “Combined Occupation Codes.xls” is a breakdown of the occupation code. These were used along with the text document, “casualtyquery.txt”, to form the desired data set. Some more manipulation was completed in Excel to get the date in the Date data type as well as the occupation code, using a LIKE state in SQL. Another way the date field was converted to an actual date was using formulas within Excel to parsing the different components - DATE(year, month, day).  
  
The .twbx file contains the presented visuals of the data. To view this data, download the Tableau Reader Software and open the file within the program. The software can be found at: http://www.tableausoftware.com/products/reader

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

The original data set we started with can be found at <http://catalog.data.gov/dataset/us-active-duty-military-deaths-since-1-jan-2001> . It is a detailed listing of all U.S. Military Active Duty deaths since 1/1/2001 giving branch of service, age at death, military occupational code, location of death, and casualty category. The data set was released by the Department of Defense. We’ve made many adjustments to the data to optimize it for our analysis in Tableau.

One adjustment was to determine the occupation codes details. We originally focused on obtaining the branch, an officer or an enlisted personnel, career field, jobs and specialty. Unfortunately, the different branches used different sets of codes – one had all numbers while another used letters and numbers. No two branch used the same codes. Additionally, there was no one location/website which contained all the codes. Due to the time consuming nature, the focus shifted to determining what was in the dataset. By focusing on the actual listing, we would not be obtaining needless data. However, this too proved time-consuming. An example is there are many types of pilots and required capturing what the pilot flew - an F-16, F-17, or numerous other types. After some discussion, we pulled back on the details and aim to capture the branch, officer or enlisted, career field and jobs.

Our initial interest in this data was sparked by the statistic that there were more military deaths domestically than abroad. Upon further investigation, it became clear that there were also a surprisingly high percentage of “self-inflicted” deaths. The number of “self-inflicted” deaths was well over half the number of “killed in action” deaths. The distribution of deaths across various ages was interesting as well, since half of the deaths are of soldiers age 25 and under.

Although it was not addressed here, we think these statistics, along with other information within this dataset, could be useful when examined alongside statistics about PTSD (post traumatic stress disorder) among military personnel. Some meaningful correlations may be drawn from such a comparison in regards to assessing how we treat our troops from a mental health perspective.

I certify that:

This paper/project/exam is entirely my own work. I have not quoted the words of any other person from a printed source or a website without indicating what has been quoted and providing an appropriate citation.

I have not submitted this paper/project to satisfy the requirements of any other course.

Signature Christopher Cottitta

Date June 14, 2013

Signature Xan Christopherson

Date June 14th, 2013

Signature Pheng Tang

Date June 14th, 2013