Data Challenge 1

Meredith Meeks, Hao Liu, Shilpa Sure, Kai Huo, Steve Andriyishen

2/20/2017

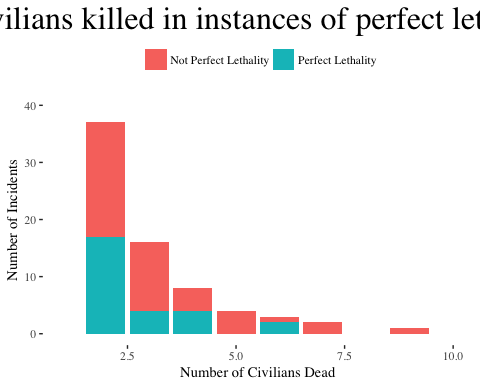
Confrontations <- read.csv('/Users/meredithmeeks/Desktop/ConfrontationsData\_170209.csv', header = TRUE, stringsAsFactors = FALSE)  
  
library(tidyverse)  
library(lubridate)  
library(readr)  
library(ggplot2)

# Question 1: Can you replicate the 86.1% number? the overall lethality ratio? the ratios for the Federal Police, Navy and Army?

* Provide a visualization that presents this information neatly.
* Please show the exact computations you used to calculate them (most likely than not, you'll need to do some additional munging in the data to get there)
* If you could not replicate them, please show why and the difference relative to your own computations (also, include a neat graph that summarizes this)
* Be very explicit: What are you assuming to generate these computations?

No, we cannot replicate this 86.1% number, due to the fact that we do not know in which cases federal forces participated, apart from when they were wounded or killed. In order to generate these computations, we have to make some assumptions. We are assuming that there is no indicator of which armed force is involved in each confrontation. We do not know whether a confrontation is involved with Federal Police, Navy, or Army. We also assume that the calculation of “lethality ratio” does not include armed forces. Lethality ratio= (number of criminals dead + number of civilians dead)/(number of criminals wounded + number of civilians wounded).

perfect.lethality <- ifelse(Confrontations$total.people.dead!=0 & Confrontations$total.people.wounded == 0, "Perfect Lethality" , "Not Perfect Lethality")  
ggplot(Confrontations, aes(civilian.dead, fill = perfect.lethality)) +   
 geom\_bar() + xlim(1, 10) + ylim(0, 40) +  
 scale\_fill\_discrete("", labels = c("Not Perfect Lethality", "Perfect Lethality")) +  
 xlab("Number of Civilians Dead") + ylab("Number of Incidents") + ggtitle("Civilians killed in instances of perfect lethality") +  
 ggthemes::theme\_tufte() +   
 theme(legend.position = "top", plot.title = element\_text(hjust = 0.5, size = 24))



# Question 2: Now you know the data more intimately. Think a little bit more about it, and answer the following questions:

* Is this the right metric to look at? Why or why not?

The 86.1% number is the percentage of civilians killed in instances of “perfect lethality.” This tells us that there are very few instances in which civilians are wounded but not killed. This is an imperfect metric because it does not give us a sense of if the number of civilians killed is small or large compared to total deaths or criminal deaths. It is only looking within civilian deaths, so does not give us a sense of scale about how often civilians are killed overall. It also excludes cases where civilians lived - it may be that the majority of confrontations do not involve any civilians. Also, we do not know if civilians were killed by criminals or by armed forces. The ratio of killed to wounded (the lethality index) is not a good statistic to look at because it doesn’t give a full picture of the situation. It does not consider the people who left the confrontation alive, and so is not an accurate measure of true lethality.

* What is the "lethality index" showing explicitly? What is it not showing? What is the definition assuming?

The lethality index is showing the ratio of killed to wounded civilians or criminals. By definition, the lethality index is not accounting for the instances in which civilians or criminals were not harmed in any way (e.g. peaceful arrests). The definition is assuming that there are no alternatives to being wounded or killed, such as non-confrontational engagements. It is also assuming that civilian deaths or injuries are equivalent to organized crime deaths or injuries.

* With the same available data, can you think of an alternative way to capture the same construct? Is it "better"?

A slightly better way to measure “lethality” would be to look at the ratio of instances in which civilians were killed or wounded against all confrontations, and same for organized crime. However, these ratios would still not account for the cases in which the situation was resolved peacefully, as we simply do not have data on those instances.

* What additional information would you need to better understand the data? What additional information could help you better capture the construct behind the "lethality index"?

We would be able to better able to understand the data and the results obtained by the NY Times if we had a variable indicating which law enforcement party was involved (state, municipal, or federal armed forces). We could also obtain richer insights if we knew how many people were involved in the incident but were not wounded or killed (for example, if they remained alive and unwounded or were arrested without violence). Also, it would be helpful to know if civilians and/or criminals were killed by armed forces or by other criminals, if we want to know about the specific lethality of the armed forces.

## Overall lethality ratio  
(sum(Confrontations$organized.crime.dead) + sum(Confrontations$civilian.dead)) /   
 (sum(Confrontations$organized.crime.wounded) + sum(Confrontations$civilian.wounded))

## [1] 2.309972

## Graph of this lethality ratio  
ggplot(data=Confrontations, mapping=aes(x = civilian.dead + organized.crime.dead, y = organized.crime.wounded + civilian.wounded)) + geom\_bar(stat = "identity", position = "stack") + ggtitle("Overall Lethality Ratio") + labs(x = "Number Dead", y = "Number Wounded")

