Data Challenge 1

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

2/20/2017

```
Confrontations <- read.csv('/Users/shilpasure/Documents/Grad School/Columbia/Spring 2017/QMSS-GR5069/we

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.

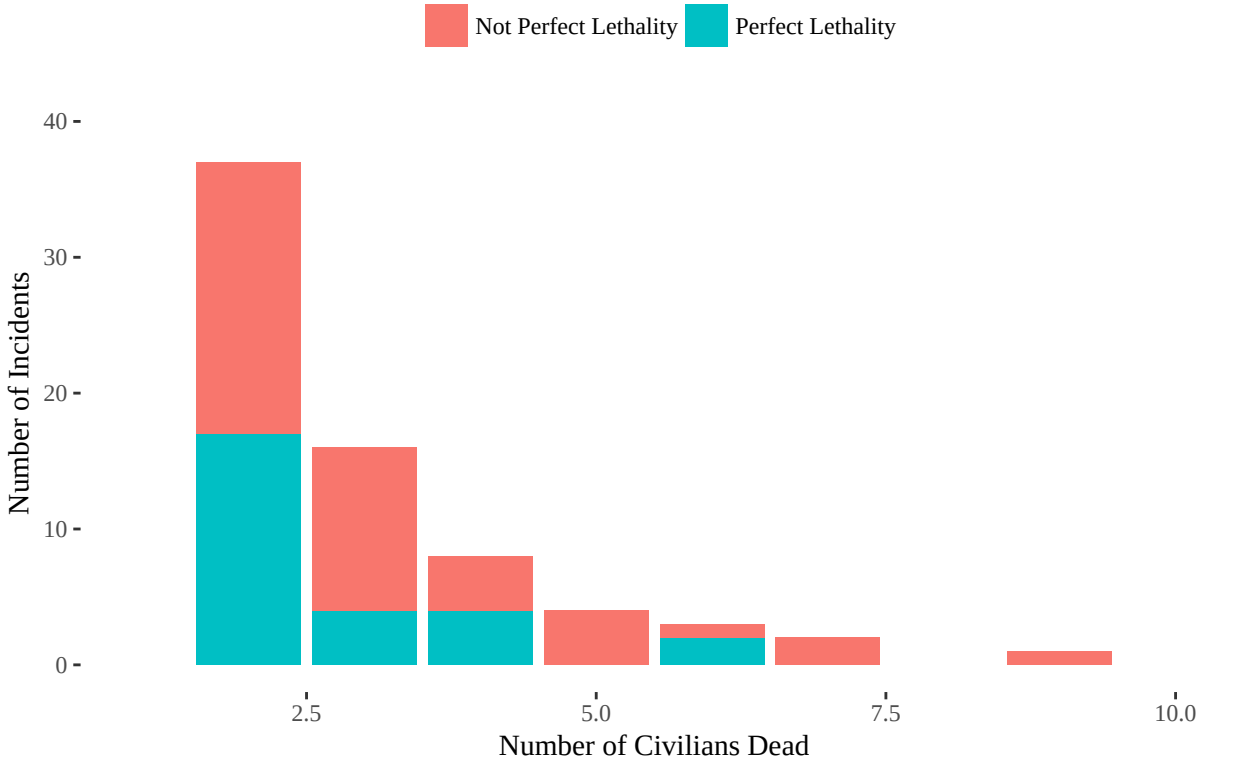
If we consider the information we do have, we can consider the percentage of civilians killed in incidents of "perfect lethality" across all confrontations (with state, municipal, or federal forces):

```
perfect.lethality <- ifelse(Confrontations$total.people.dead!=0 & Confrontations$total.people.wounded==0
civ <- sum(Confrontations$civilian.dead[perfect.lethality==1])
pct.dead <- sum(civ)/sum(Confrontations$civilian.dead)
pct.dead
```

```
## [1] 0.393617
```

The percentage of civilians killed in incidences of "perfect lethality" across all confrontations is 39.3%, which is off by 46.8% to the 86.1% figure claimed for confrontations with federal forces. We do not have the information necessary to know which confrontations involve federal forces, so we cannot replicate the 86.1% number with the data available. The graph below shows our calculation as a visualization:

```
perfect.lethality <- ifelse(Confrontations$total.people.dead!=0 & Confrontations$total.people.wounded ==
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 instan
  ggthemes::theme_tufte() +
  theme(legend.position = "top", plot.title = element_text(hjust = 0.5, size = 24))
```

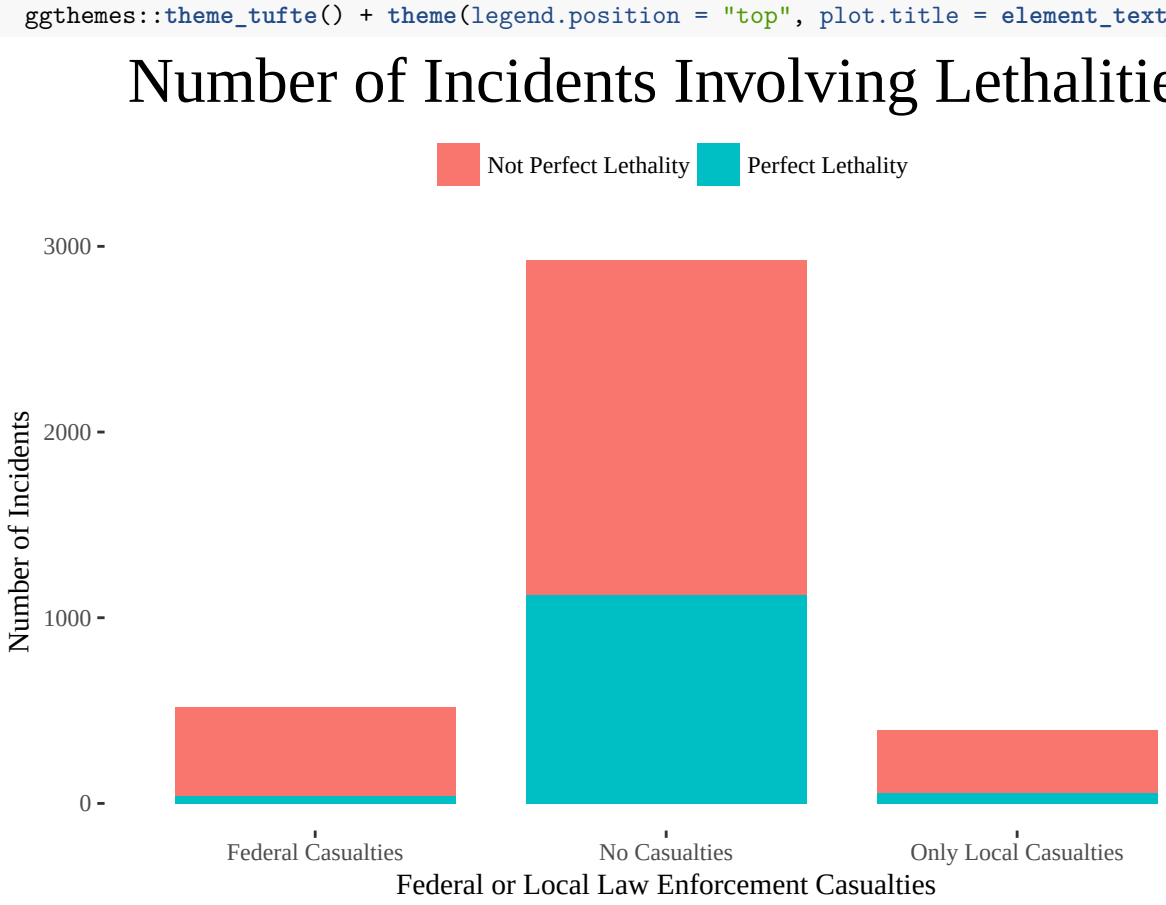


As we can see, the proportion of dead civilians killed in instances of perfect lethality (represented in blue) appears to be much less than 86.1%.

We also considered that maybe we could extract which cases involved federal forces by looking at confrontations in which federal forces (Army, Navy, Federal Police, or AFI) were wounded or killed. However, this does not include cases federal forces were involved but not wounded or killed. As it turns out, there are very few cases that involve wounded or killed federal forces, as we can see from the visualization below:

```
Confrontations$class <- NA_character_
Confrontations$class[which(with(Confrontations, military.dead + navy.dead + federal.police.dead + afi.d
                           military.wounded + navy.wounded + federal.police.wounded + afi.wounded
                         != 0))] <- "Federal Casualties"
Confrontations$class[which(with(Confrontations, military.dead + navy.dead + federal.police.dead + afi.d
                           military.wounded + navy.wounded + federal.police.wounded + afi.wounded == 0
                         state.police.dead + ministerial.police.dead +
                           municipal.police.dead + public.prosecutor.dead +
                           state.police.wounded + ministerial.police.wounded +
                           municipal.police.wounded + public.prosecutor.wounded != 0))] <- "Only Local
Confrontations$class[which(with(Confrontations, military.dead + navy.dead + federal.police.dead + afi.d
                           military.wounded + navy.wounded + federal.police.wounded + afi.wounded +
                           state.police.dead + ministerial.police.dead +
                           municipal.police.dead + public.prosecutor.dead +
                           state.police.wounded + ministerial.police.wounded +
                           municipal.police.wounded + public.prosecutor.wounded == 0))] <- "No Casualt
ggplot(Confrontations, aes(class, fill=perfect.lethality)) + geom_bar(width = .8) +
  xlab("Federal or Local Law Enforcement Casualties") + ylab("Number of Incidents") +
  ggtitle("Number of Incidents Involving Lethalities") +
  scale_fill_discrete("") +
```

```
ggthemes::theme_tufte() + theme(legend.position = "top", plot.title = element_text(hjust = 0.5, size =
```



Because the number of cases in which federal forces are wounded or killed are so low, we assume that there are several cases in which federal forces are involved but not wounded or killed. Based on the data available, however, we cannot isolate those cases. Trying to calculate the percentage of civilians killed in instances of perfect lethality when a member of a federal force was wounded or killed (as a proxy for instances involving federal forces) gives us a figure of 1.1%, which is very far off of 86.1%:

```
federal <- ifelse(Confrontations$military.dead!=0 | Confrontations$navy.dead!=0 | Confrontations$federal

perfect.lethality <- ifelse(Confrontations$total.people.dead!=0 & Confrontations$total.people.wounded==0

civ_2 <- sum(Confrontations$civilian.dead[perfect.lethality==1 & federal==1])


pct_dead_2 <- sum(civ_2)/sum(Confrontations$civilian.dead)
pct_dead_2
```

```
## [1] 0.0106383
```

We cannot use just the cases in which federal forces were wounded or killed. Without an indicator for which armed forces were involved in each confrontation, we cannot replicate the 86.1% figure for the percentage of dead civilians in confrontations with federal armed forces who were killed in instances of "perfect lethality".

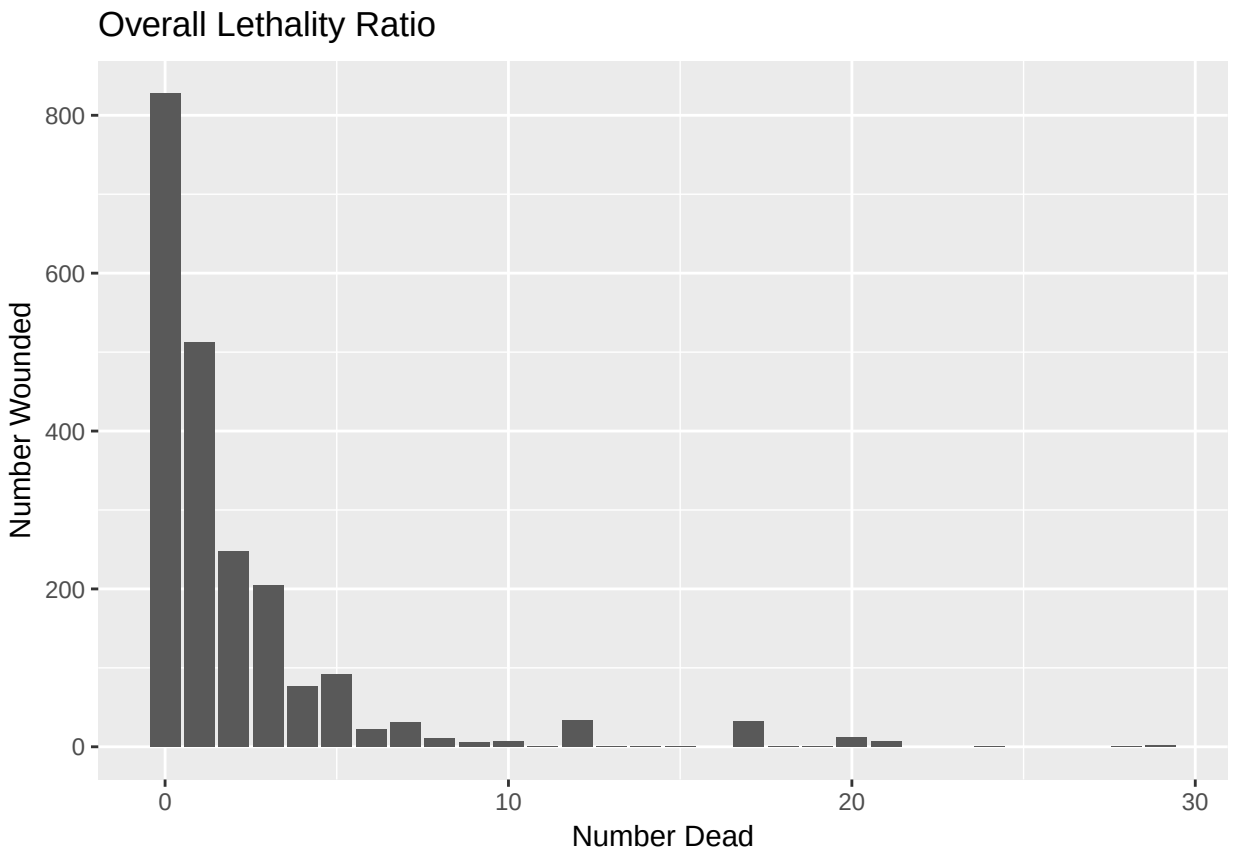
Now considering the lethality index, we could not exactly replicate the 2.6 number, but our calculation came close (2.3). We are assuming here that the calculation of "lethality ratio" does not include armed forces. The lethality ratio that we are using is (number of criminals dead + number of civilians dead)/(number of criminals wounded + number of civilians wounded).

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

```
## [1] 2.309972
```

Visualizing the lethality ratio in a plot:

```
## Graph of this lethality ratio
ggplot(data=Confrontations, mapping=aes(x = civilian.dead + organized.crime.dead, y = organized.crime.w
```



With the lethality ratio, we also cannot replicate the lethality ratios for Federal Police, Army, and Navy without knowing specifically if the Federal Police, Army, or Navy were involved in a confrontation. The data we have is missing this piece of information.

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 - there is no accounting for if those who were killed or wounded were meant to be targeted in the confrontation. Additionally, it does not show if those who were killed were killed by armed forces or by members of organized crime (or another party).

- 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.

Another slightly better way to gauge a lethality ratio would be to look at the number of killed divided by number of wounded separately for organized crime casualties and for civilian casualties. When we do that, we see that the federal and local forces are highly efficient at killing organized crime members (lethality ratio over 3), but are also more reserved when it comes to civilians (lethality ratio of 0.57). Besides, we do not know from this dataset how the civilians are getting harmed in incidences of violence. Perhaps, the smaller lethality ratio for civilians represents that the federal and local forces are doing a good job protecting civilians from getting killed by organized crime members.

```
# Better lethality ratios subdivided by organized crime and civilians

# Lethality ratio for organized crime casualties
sum(Confrontations$organized.crime.dead) / sum(Confrontations$organized.crime.wounded)
```

```
## [1] 3.078751
```

```
# Lethality ratio for civilian casualties
sum(Confrontations$civilian.dead) / sum(Confrontations$civilian.wounded)
```

```
## [1] 0.575804
```

- 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 lethality of the armed forces specifically.