Data Challenge #1

Mario Saraiva - mhs2195

Contents

GR5069 - TOPICS IN APPLIED DATA SCIENCE FOR SOCIAL SCIENTISTS - Spring 2018

February 17, 2018

```
##About this Project
# File-Name:      Data_Challenge_1_MarioSaraiva.Rmd
# Version:        R 3.4.1
# Date:           Feb 14, 2018
# Author:         MS
# Purpose:        Verify the lethality index with reported figures reported by
# the NY Times in 2016 that said: " `86.1% of dead civilians` who presumably
# participated in confrontations with federal armed forces were killed in events
# of "perfect lethality" where there were only dead and no wounded. [...] Mexico
# has the terrible situation of having `lethality indices of 2.6`. The lethality
# index of the Federal Police is `2.6 dead for every wounded`, the Navy's reaches
# 17.3 dead for every wounded, and the Army's is 9.1 dead for every wounded."
# (Ahmed & Schmitt, 2016)
# Input Files:    1. AllViolenceData_171220.csv
# Output Files:   Data_Challenge_1_MarioSaraiva.html
# Data Output:    None
# Previous files: None
# Dependencies:   None
# Required by:    None
# Status:         IN PROGRESS
# Machine:        Mac laptop
```

```
# Throughout this document the following items were addressed:
#
# 1. Can you replicate the 86.1% number? the overall lethality ratio? the ratios for the Federal Police
#
#     a. Provide a visualization that presents this information neatly.
#
#     b. Please show the exact computations you used to calculate them.
#
#
# 2. Is this the right metric to look at? Why or why not?
#
# 3. What is the "lethality index" showing explicitly? What is it not showing? What is the definition a
#
# 4. With the same available data, can you think of an alternative way to capture the same construct? I
#
# 5. What additional information would you need to better understand the data?
```

```
#
# 6. What additional information could help you better capture the construct behind the "lethality index
```

Lethality Rate in Mexico

Stats given in the NY Times article:

1. `86.1% of dead civilians` in confrontations with **federal armed forces**.
2. `Mexico's lethality indices of 2.6`
3. `2.6 dead for every wounded` ratio for the federal POLICE.
4. `17.3 dead for every wounded` ratio for the NAVY.
5. `9.1 dead for every wounded` ratio for the ARMY.

The NY Times article did not discuss how these numbers were calculated. I explored different ways they could have calculated to arrive at these numbers. To no one's surprise, my results were different than the NY times results.

```
#The proportion for the aggregate civilians lethality rate was the following

dta_gov_nongov <- dta %>%
  mutate(armforces_death = navy_dead + military_dead + federal_police_dead + state_police_dead + municip
         armforces_wounded = navy_wounded + military_wounded + federal_police_wounded + state_police_wou
         nonfed_wounded = civilian_wounded + organized_crime_wounded + afi_wounded + public_prosecutor_w
         nonfed_dead = civilian_dead + organized_crime_dead + afi_dead + public_prosecutor_dead)
```

```
#Presence of Armed forces and the impact on civilians.

##The following will count the total number of dead non-gov subjects given the presence of Armed Forces

indexes <- dta_gov_nongov %>%
  filter(armforces_wounded != 0 | armforces_death != 0 ) %>% #to make sure the armed forces were involv
  summarize(count = sum(nonfed_dead))
colnames(indexes) <- "All.Non.AF_dead|A.F."
#The following will count the total number of wounded non-gov subjects given the presence of Armed Forc

indexes$All.NonAF.wound <- as.numeric(dta_gov_nongov %>%
  filter(armforces_wounded != 0 | armforces_death != 0 ) %>%
  summarize(count = sum(nonfed_wounded)))
#colnames(indexes$All.NonAF.wound) <- "All.Non.AF_wounded|A.F."

indexes$All.index <- as.numeric(indexes$`All.Non.AF_dead|A.F.` / indexes$All.NonAF.wound)
#colnames(indexes$All.index) <- "Armed Forces Lethality index "

count_all_total <- as.numeric(dta_gov_nongov %>%
   summarize(count = sum(armforces_wounded)))

indexes$All.Prop <- as.numeric(indexes$`All.Non.AF_dead|A.F.`/count_all_total)
#colnames(indexes$All.Prop) <- "non AF dead / AF dead"
```

```
#Perfect Lethality cases
indexes$Perfect.L.dead <- as.numeric(dta_gov_nongov %>%
     filter(perfect_lethality != 0 & civilian_dead != 0) %>%
   summarize(count = sum(nonfed_dead)))
```

```
indexes$Perfect.L..Index <- as.numeric(indexes$Perfect.L.dead / (indexes$`All.Non.AF_dead|A.F.` + indexe
```

```
#Lethality ratios.

##Similarly, this section will calculate the 'lethality index' (Num. of civilians dead / num. of civili

###This index has a clear shortcoming for it does not consider any armed forces officers that were dead

#Federal Police
##1 Aggregate for non A.F.
indexes$FP.nonAF.dead <- as.numeric(dta_gov_nongov %>%
   filter(federal_police_dead != 0 | federal_police_wounded != 0) %>%
   summarize(count = sum(nonfed_dead)))
##2
indexes$FP.nonAF.wounded <- as.numeric(dta_gov_nongov %>%
   filter(federal_police_dead != 0 | federal_police_wounded != 0) %>%
   summarize(count = sum(nonfed_wounded)))
##3
indexes$FP.nonAF.Index <- as.numeric(indexes$FP.nonAF.dead / (indexes$FP.nonAF.dead + indexes$FP.nonAF.
#colnames(indexes$FP.Index) <- "Lethality index | Federal Police "

##4 - Aggregates for Federal Police officers
indexes$FP.only.dead <- as.numeric(dta_gov_nongov %>%
  filter(federal_police_dead != 0 | federal_police_wounded != 0) %>%
   summarize(count = sum(federal_police_dead)))
##5
indexes$FP.only.wounded <- as.numeric(dta_gov_nongov %>%
  filter(federal_police_dead != 0 | federal_police_wounded != 0) %>%
   summarize(count = sum(federal_police_wounded)) )
##6
indexes$FP.only.Index <- as.numeric(indexes$FP.only.dead / (indexes$`All.Non.AF_dead|A.F.` + indexes$Al
#colnames(indexes$FP.only.Index) <- "FP only dead/wounded "

##7 Ratio of total non AF dead / total FP dead
indexes$FP.prop.of.dead <- as.numeric(indexes$`All.Non.AF_dead|A.F.` / indexes$FP.only.wounded)

display_index <- t(format(round(indexes, 2), nsmall = 2))
display_index <- as.data.frame(display_index)
#colnames(display_index)[1] <- "Results"
```

```
#Navy
##1 Aggregate for non A.F. | Navy presence
indexes$NV.nonAF.dead <- as.numeric(dta_gov_nongov %>%
   filter(navy_dead != 0 | navy_wounded != 0) %>%
   summarize(count = sum(nonfed_dead)))
##2
indexes$NV.nonAF.wounded <- as.numeric(dta_gov_nongov %>%
    filter(navy_dead != 0 | navy_wounded != 0) %>%
   summarize(count = sum(nonfed_wounded)))
##3
indexes$NV.nonAF.Index <- as.numeric(indexes$NV.nonAF.dead / (indexes$NV.nonAF.dead + indexes$NV.nonAF.
#colnames(indexes$FP.Index) <- "Lethality index | Navy "

##4 - Aggregates for Navy officers
```

```
indexes$NV.only.dead <- as.numeric(dta_gov_nongov %>%
  filter(navy_dead != 0 | navy_wounded != 0) %>%
   summarize(count = sum(navy_dead)))
##5
indexes$NV.only.wounded <- as.numeric(dta_gov_nongov %>%
  filter(navy_dead != 0 | navy_wounded != 0) %>%
   summarize(count = sum(navy_wounded)) )
##6
indexes$NV.only.Index <- as.numeric(indexes$NV.only.dead / (indexes$`All.Non.AF_dead|A.F.` + indexes$Al
#colnames(indexes$FP.only.Index) <- "FP only dead/wounded "

##7 Ratio of total non AF dead / total NV dead
indexes$NV.prop.of.dead <- as.numeric(indexes$`All.Non.AF_dead|A.F.` / indexes$NV.only.wounded)

#Army
##No data on a number of army officers wounded or killed.

#display_index <- t(format(round(indexes, 2), nsmall = 2))
#colnames(display_index)[1] <- "Results"
```

According to the NY times `86.1% of dead civilians` in confrontations with **federal police** . So I tried replicating this result through the following, where nonarmed forces can be interpreted as civilians:

$$\frac{TotalNum.of nonArmedForcessubjectsdead|ArmedForcespresence}{TotalNum.of nonArmedForcessubjects(wounded + dead)|ArmedForcespresence}$$
$$= \frac{1259}{1259 + 368} * 100 = 76.90\%$$

See the next section for an explanation of the computations used. According to my calculations, given the data, the ratio of $\frac{dead}{wounded}$ for all violent events is 0.53. For every wounded, during violent events, there was at least 1 dead.

86.1% of dead civilians in confrontations with federal armed forces. Mexico's lethality indices of 2.6 2.6 dead for every wounded ratio for the federal POLICE. 17.3 dead for every wounded ratio for the NAVY. 9.1 dead for every wounded ratio for the ARMY.

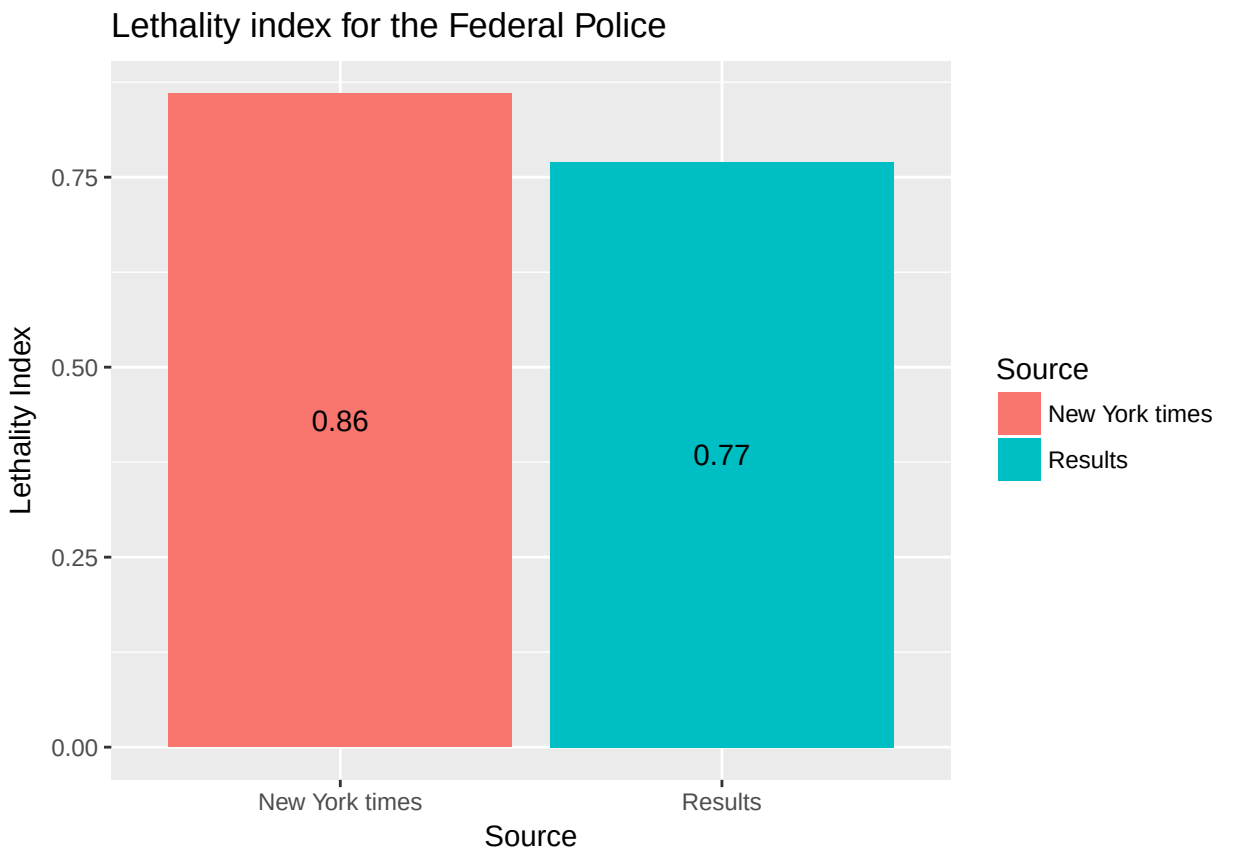
```
nytimes <- 0.86
nytimes <- as.data.frame(nytimes)
nytimes$All.index <- 2.6
nytimes$FP.prop.dead <- 2.6
nytimes$NV.prop.dead <- 17.3
nytimes$Army.prop.dead <- 9.1

compare_indexes <- rbind(nytimes, c(indexes$FP.nonAF.Index, indexes$All.index, indexes$FP.prop.of.dead,
row.names(compare_indexes)[1] <- "New York times"
row.names(compare_indexes)[2] <- "Results"
compare_indexes$Source <- c("New York times", "Results")

compare_indexes$nytimes <- x2(compare_indexes$nytimes)

ggplot(compare_indexes) +
  geom_col(aes(Source, nytimes, fill = Source)) +
  ggtitle("Lethality index for the Federal Police") +
```

```
  labs(x = "Source", y = "Lethality Index") +
  geom_text(aes(Source, nytimes, label = nytimes), position = position_stack(vjust = 0.5))
```



Discussion of Alertnative Calculations

Using the data available, I defined the lethality index can be defined as the ratio of the total number of civillians[1] dead given the presence of the armed forces[2] over the total number of civilians wounded given the presence of the armed forces. $\frac{Num.of dead}{Num.of wounded}$ per confrontation or event. But this is confusing. For example, the overall lethality of government armed forces should be:

$$\frac{TotalNum.of nonArmedForcessubjectsdead|ArmedForcespresence}{TotalNum.of nonArmedForcessubjectswounded|ArmedForcespresence}$$

.

So given the presence of at least one government agency what is the proportion of dead and wounded people during the violent event that is not considered part of the armed forces? Confrontations vary a lot and are complex. For instance, battling in the sea is very different than on the ground, so many wounded in the sea does not have the same implications as wounded in land confrontations. Therefore, we should not over-stretch this ratio as an indicator of the brutality of federal armed forces. This ratio also does not address brutality of armed forces given confrontation nor does it take into account the guns used; 100 wounded using a heavy automatic weaponry is not the same as 100 wounded with a knife or physical aggression.

Alternative calculations measures what happened given the presence of a certain group (e.g. Federal Police) engaging in a confrontation. In other words, it measures the ratio of Federal Police officers engaged in a confrontation that died in comparison to other officers that were wounded. Then we can compare this ratio for Federal Police and Civilians. Although, that still does not shows us much.

Computations

Through seven computations I investigated the lethality index for Mexico, the Federal Police, Navy, and Army, as well as, the dead of civilians per death of an agency's officer. I will use the calculations performed for the Navy as an example of these seven calculations. All other calculations follow a similar pattern.

I developed alternatives attempting to overcome the limitations imposed by the data. These can be seen in my calculations.

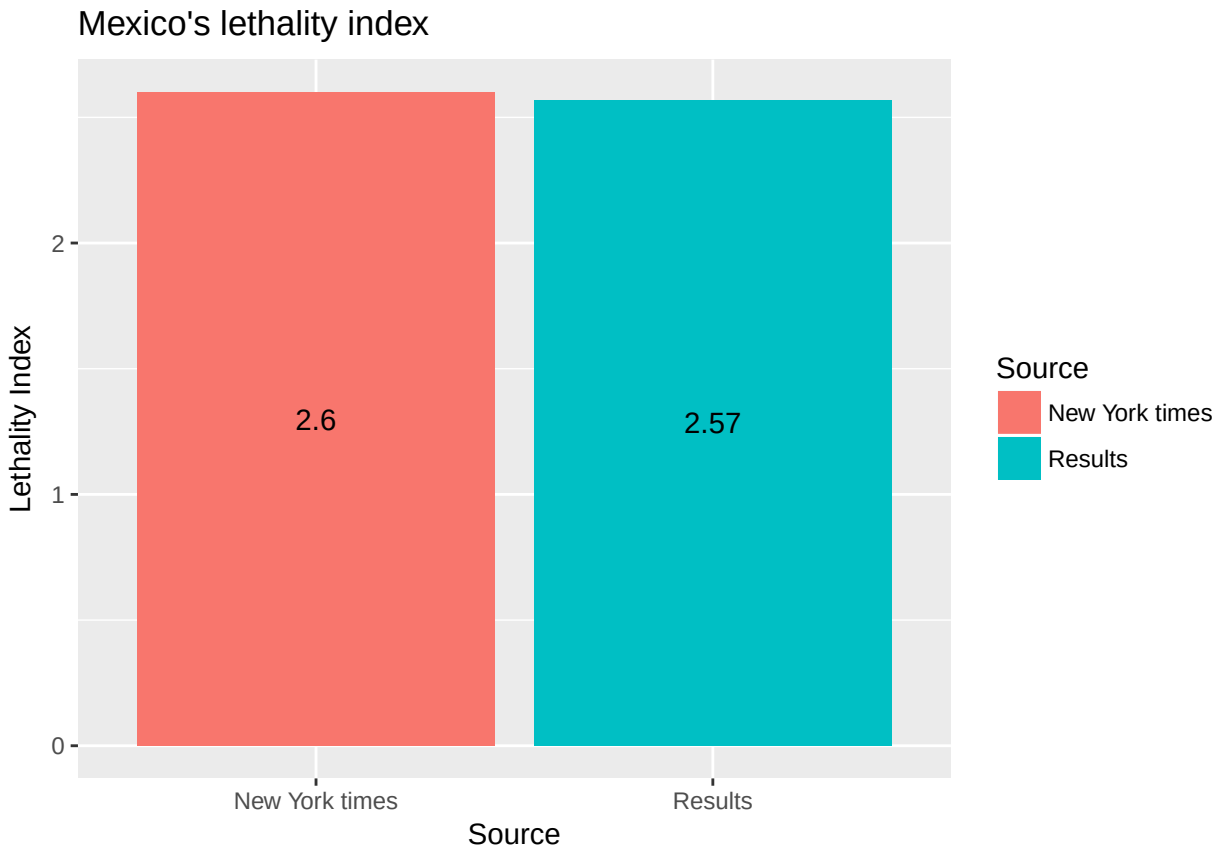
Calculations:

1. First, I calculated the aggregate the number of non-Armed Forces dead given the presence of the Navy in the violent event. I did not feel comfortable using the binary variables provided indicating the presence of an armed force or not. To produce conservative estimates, I used only observations that had either one dead from the navy or one wounded.
2. Similarly, I calculated the aggregate number of nonarmed Forces wounded.
3. Then I calculated the proportion of total Non-Armed forces dead over total Non Armed forces wounded. This is a lethality index capturing conflicts with the presence of the Navy.
4. Forth, I calculated the aggregate number of Navy officers dead in violent events.
5. Like in step 4, I calculated the aggregate number of Navy officers wounded in violent events.
6. At this stage, I calculated the number of nonarmed forces dead given the presence of the Navy in the violent event over the total amount of Non-Armed Forces people in the event.
7. Lastly, I calculated the ratio of the Total number of Non-Armed Forces dead over the total number of navy officers dead. This gives us the ratio of dead civilians per death in the navy.

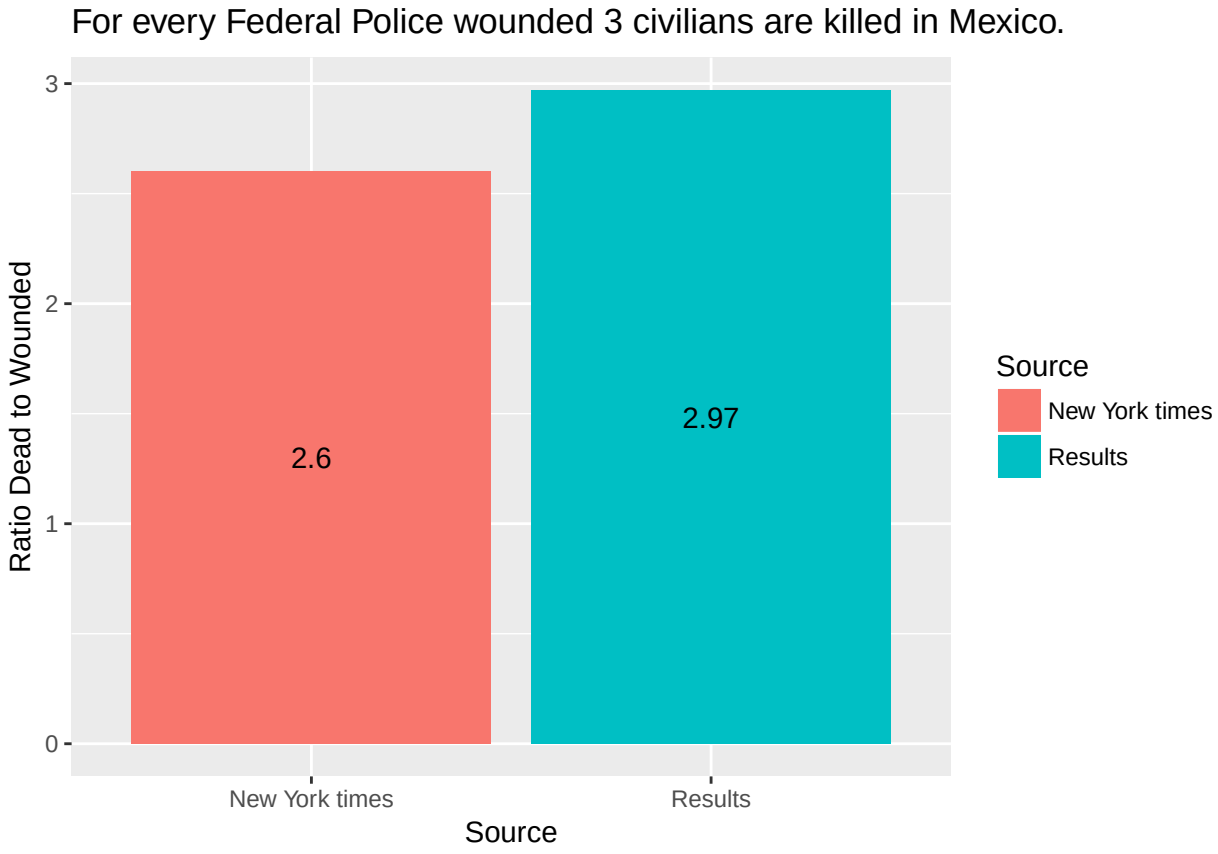
Results

Data visualizations

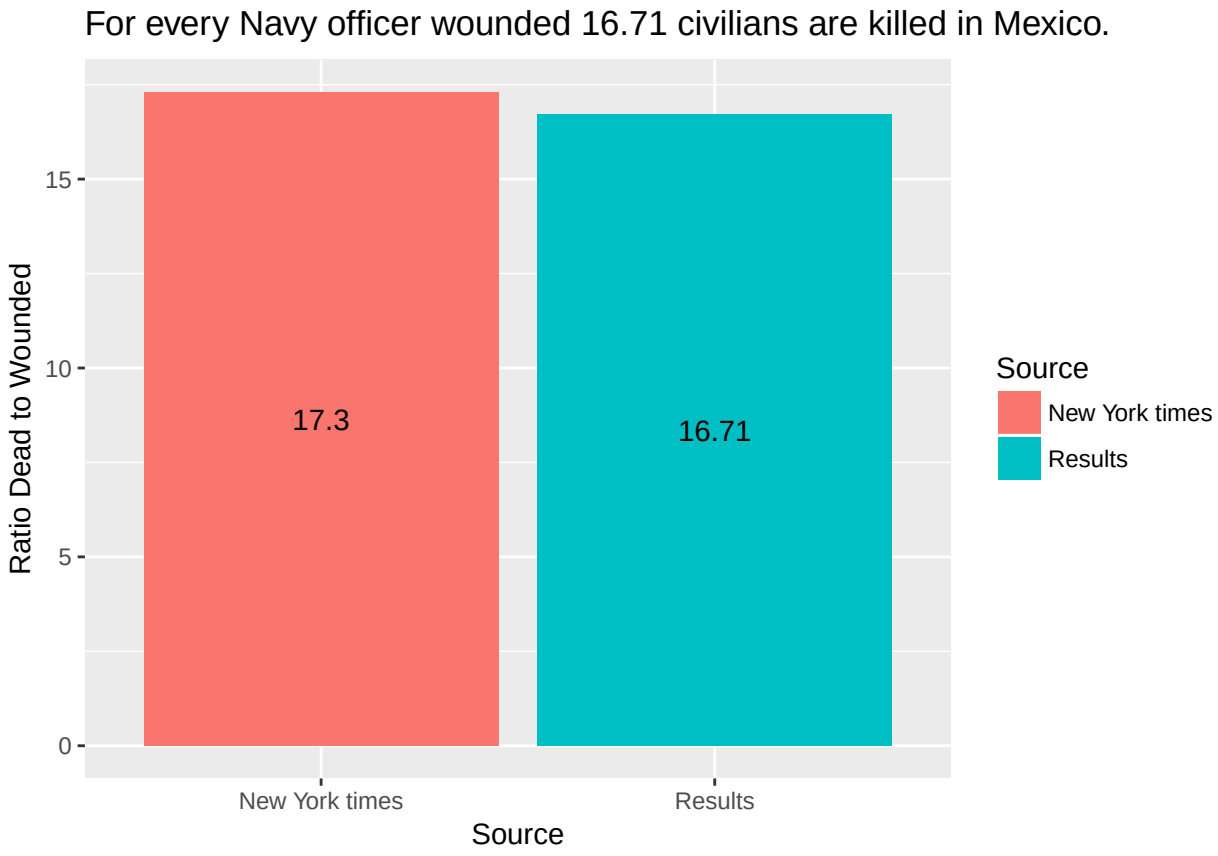
```
compare_indexes$All.index <- x2(compare_indexes$All.index)
compare_indexes$FP.prop.dead <- x2(compare_indexes$FP.prop.dead)
compare_indexes$NV.prop.dead <- x2(compare_indexes$NV.prop.dead)
compare_indexes$Army.prop.dead <- x2(compare_indexes$Army.prop.dead)
ggplot(compare_indexes) +
  geom_col(aes(Source, All.index, fill = Source)) +
  ggtitle("Mexico's lethality index") +
  labs(x = "Source", y = "Lethality Index") +
  geom_text(aes(Source, All.index, label = All.index), position = position_stack(vjust = 0.5))
```



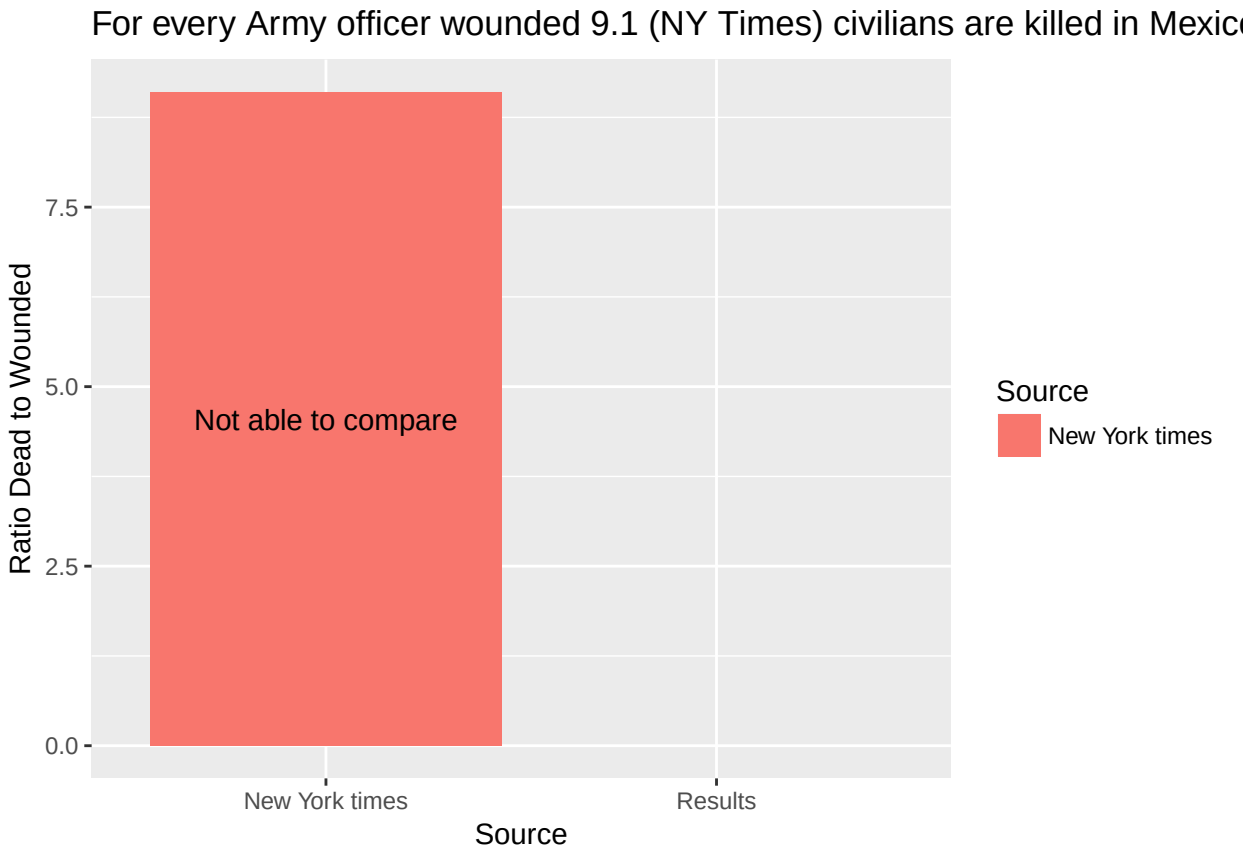

```
ggplot(compare_indexes) +
  geom_col(aes(Source, FP.prop.dead, fill = Source)) +
  ggtitle("For every Federal Police wounded 3 civilians are killed in Mexico.") +
  labs(x = "Source", y = "Ratio Dead to Wounded") +
  geom_text(aes(Source, FP.prop.dead, label = FP.prop.dead), position = position_stack(vjust = 0.5))
```




```
ggplot(compare_indexes) +
  geom_col(aes(Source, NV.prop.dead, fill = Source)) +
  ggtitle("For every Navy officer wounded 16.71 civilians are killed in Mexico.") +
  labs(x = "Source", y = "Ratio Dead to Wounded") +
  geom_text(aes(Source, NV.prop.dead, label = NV.prop.dead), position = position_stack(vjust = 0.5))
```




```
ggplot(compare_indexes) +
  geom_col(aes(Source, Army.prop.dead, fill = Source)) +
  ggtitle("For every Army officer wounded 9.1 (NY Times) civilians are killed in Mexico.") +
  labs(x = "Source", y = "Ratio Dead to Wounded") +
  geom_text(aes(Source, Army.prop.dead, label = "Not able to compare"), position = position_stack(vjust
```




```
# ggplot(df_rates) +
#    geom_col(aes(x = V2, y = V1, fill = V2)) +
#    labs(x = "Type of people", y = "Proportion") +
#    ggtitle("Proportion of Lethality (Num. deaths | wounded = 0) / (Total deaths) by people type")
#     ggsave(file = "plot1.png")
```

Summary Table

```
indexes$results <- "results"
display_index <- dcast(melt(indexes, id.vars = "results"), variable ~ results)
display_index$results <-  x2(display_index$results)
```

Variables	Results	New York Ti
Perfect Lethality Index	9.15%	
Number of Non Armed Forces dead	8608	
Number of Non Armed Forces wounded	3350	
Lethality index for Mexico	**2.57**	**2.60**
Dead-to-Wounded Ratio of all Non Armed forces	0.53	
Number of non AF dead given Federal Police	1259	
Number of non AF wounded given Federal Police	378	
Lethality Index for Federal Police	**77%**	**86%**
Number of Federal Police Officers dead	872	
Number of Federal Police Officers wounded	2902	
Federal Police officers dead-to-wound ratio	2.97	
Non Armed Forces Dead-to-Wound Ratio given Federal Police	**2.97**	**2.6**
Number of non Armed Forces dead given Navy	800	

Variables	Results	New York Ti
Number of non Armed Forces wounded given Navy	69	
Lethality Index for Navy	92%	
Number of Navy officers dead	153	
Number of Navy officers wounded	515	
Navy officers dead-to-wound ratio	16.71	
Non Armed Forces Dead-to-Wound Ratio given Navy	**16.71**	**17.3**
Army	**Not enough data to compare**	**9.1**

Discussion about the Lethality Index

There were some significant issues when trying to calculate the ratios:

1. I was unable to calculate the ratio for Army combatants given the lack of data available.
2. We are unable to see the events in which 100% dead but zero wounded in any category nor can we see the people that were "saved" or ran away and etc...
3. It is not easy to calculate the Navy ratio without a comprehensive mapping of the location of each confrontation in order to properly count the events in which the Navy could have participated.
4. Missingness of data: (I am not sure if that was intentional or impossible to calculate given the data we have). The lethality index variables in the dataset are mostly blank (see plot below). One potential explanation is that when creating the index (by applying the ratio previously discussed) we have many zero values which prevent the index from being more useful.

Variable	Count of missing values
state	20
state_abbr	20
municipality	20
organized_crime_lethality	19744
army_lethality	31463
navy_lethality	33787
federal_police_lethality	32680
organized_crime_NewIndex	19744
army_NewIndex	31463
navy_NewIndex	33787

Conclusion: Improving the index

Why killed-to-wounded ratios don't measure lethality

The lack of information on the total number of participants in each armed engagement is not clear. In other words, the denominator is unknown and "it is therefore very hard to estimate dead/wounded rates per event. It is important to make a distinction between the combat efficiency of armed forces - one index could be # of shots on target/shots fired and the brutality of the armed forces. It is very difficult to measure the"brutality" of, for example, a federal police officer for it is difficult to get accurate data on the number of wounded and dead and attribute the blame to the officer per se.

García-Ponce & Lajous (2016) drives the point home when they explain that "although the use of killed-to-wounded ratios has been suggested as an effective metric to track specific massacres (Coupland and Meddings, 1999), it is a statistic more frequently used to analyze medical attention during armed conflicts than to measure combat efficiency, the severity of civil war, or the use of force."

Furthermore, the data does not allow us to understand the context of the confrontation, for example, we are unsure how many people were involved or the type of weapons used. The calculations measure the consequences for a certain group (e.g. Federal Police) of engaging in a confrontation event. In other words, it measures the ratio of Federal Police officers engaged in a confrontation that died in comparison to other officers that were wounded. Since violent events can have the cooperation of multiple armed force agencies, to identify if an agency was involved in the violent event we set ($FederalPolicewounded! = 0$). So it becomes hard to say who killed/wounded who.

Lastly, I believe that to get a comprehensive picture on the brutality issue, there have to be qualitative assessments in order to capture the context of different conflicts. This is also very difficult given the high risk of violent conflicts, so most observers belong to either the armed forces or criminal organizations, and civilians are usually not comfortable to accurately report on what they saw as they fear any potential backlash. It is possible that, if the violent event was caught on video, to use the recorded images to make better assessments of the brutality and combat efficiency of the armed forces in Mexico.

Endnotes

[1] Civilians, in this case, are all that are not part of the Armed Forces, so it includes the organized crime and government employees in other roles.

[2] Armed forces include all government agencies used to keep the law and order such as the Federal Police, the Army, the Navy, among others.

References

García-Ponce, O., & Lajous, A. (2016, October 21). Challenges of Measuring an Armed Group's Lethality: Insights from the Mexican Drug War. Retrieved February 16, 2018, from https://politicalviolenceataglance.org/2016/10/21/challenges-of-measuring-an-armed-groups-lethality-insights-from-the-mexican-drug-war/