Data Analysis Essay

Table of Contents

2. Data

1. **Generate a binary variable named "bombing" equal to 1 if the variable "attacktype1_txt" is equal to "Bombing/Explosion" and 0 otherwise.**

   I have used the ifelse function to generate a dummy binary variable:

   ```
   gtd$bombing <- ifelse(gtd$attacktype1_txt == "Bombing/Explosion",1,0)
   ```

2. **Generate a categorical variable named "attack.target".**

   I have used the ifelse function to generate a dummy categorical variable:

   ```
   gtd$attack.target <- ifelse(gtd$targtype1_txt == "Business", "1",
                        ifelse(gtd$targtype1_txt == "Government (General)", "2",
                        ifelse(gtd$targtype1_txt == "Military", "3",
                        ifelse(gtd$targtype1_txt == "Police", "4",
                        ifelse(gtd$targtype1_txt == "Private Citizens & Property", "5",
   ```

```
ifelse(gtd$targtype1_txt == is.na(gtd$targtype1_txt), "0","0"))))))
```

3. Generate a categorical variable named "cat.gdp".

As 'gdp' was originally a factor variable, I coerced it into a numeric variable.

I then used the ifelse function to generate a dummy categorical variable:

```
gtd$gdp <- as.numeric(gtd$gdp)

q.25 <- quantile(gtd$gdp, probs= c(0.25), na.rm = TRUE)
q.50 <- quantile(gtd$gdp, probs= c(0.50), na.rm = TRUE)
q.75 <- quantile(gtd$gdp, probs= c(0.75), na.rm = TRUE)

gtd$cat.gdp <- ifelse(gtd$gdp < q.25, "1",
                      ifelse (gtd$gdp>= q.25 & gtd$gdp < q.50, "2",
                      ifelse(gtd$gdp >= q.50 & gtd$gdp < q.75, "3",
                      ifelse(gtd$gdp >= q.75, "4", "0"))))
```

4. Generate a binary variable named "low.capacity" equal to 1 if "cat.gdp" = 1 and 0 otherwise.

I used ifelse to generate a dummy binary variable:

```
gtd$low.capacity <- ifelse(gtd$cat.gdp == "1", "1", "0")
```

3. Research Questions

a) On average, are bombings more or less lethal than other types of attack?

Null hypothesis – On average bombings are more lethal than other types of attack.

Alternative hypothesis – On average bombings are less lethal than other types of attack.

Calculations

```
tapply(gtd$nkill[gtd$nkill != "-999"], gtd$bombing[gtd$nkill != "-999"], mean)
rownames(bombings.moreorless) = c("Other types of attack", "Bombings")
```

```
Other types of attack          Bombings
             2.407631          2.223714
```

As shown by the mean casualties of each, bombings are less lethal than other types of attack, on average. However, the contrast between the two figures is not substantial as there is only a 0.17 unit difference.

b) On average, are suicide bombings more lethal than non-suicide bombings?

Null hypothesis - On average, suicide bombings are more lethal than non-suicide bombings.

Alternative hypothesis - On average, suicide bombings are less lethal than non-suicide bombings.

<u>Calculations:</u>

tapply(gtd$nkill[gtd$nkill != "-999" & gtd$attacktype1_txt =="Bombing/Explosion"], gtd$suicide[gtd$nkill != "-999" & gtd$attacktype1_txt == "Bombing/Explosion"], mean)

```
Other types of bombings        Suicide Bombings
               1.061333                8.277778
```

As shown by the mean casualties of each, suicide bombings are more lethal than non-suicide bombings on average.

Differences in lethality across various target types

Null hypothesis - There will be no difference in lethality across different target types.

Alternative hypothesis - There will be a difference in lethality across different target types.

<u>Calculations:</u>

leth.targ <- tapply(gtd$nkill[gtd$nkill != "-999"], gtd$attack.target[gtd$nkill != "-999"], mean)
rownames(leth.targ) = c("Others", "Business", "Government(General)", "Military", "Police", "Private Citizens & Property")

```
   Others        Business   Government(General)     Military        Police  Private Citizens & Property
1.7811159       0.4603175             0.8928571    3.9738220     3.1840000                    2.0763052
```

As shown by the mean casualties between target types, there is a difference in lethality across various intended targets, on average.

What factors influence the success or failure of an attack?

1. Is success more likely against some targets than others?

Null hypothesis - The likelihood of success does not vary across the type of target

Alternative hypothesis - The likelihood of success varies across the type of target

Calculations:

```
succ.targ <- tapply(gtd$success, gtd$attack.target, sum)
prop.table(succ.targ)
```

Others	Business	Government(General)	Military	Police	Private Citizens & Property
0.17402597	0.07012987	0.08441558	0.20519481	0.13636364	0.32987013

As shown by the proportions of success for each target type, the levels of successful terrorist attacks differ between intended targets.

Effect of low economic capacity on success or failure of an attack:

Null hypothesis – Countries with governments that have low levels of economic capacity have insufficient funds to combat terrorist activities.

Alternative hypothesis - Countries with governments that have low levels of economic capacity do not have insufficient funds to combat terrorist activities.

Calculations:

```
econcap.suc <- tapply(gtd$success[gtd$success == "1"],
gtd$low.capacity[gtd$success == "1"], sum)
rownames(econcap.suc) = c("High Economic Capacity", "Low Economic Capacity")
```

High Economic Capacity	Low Economic Capacity
618	83

As shown by the number of successful terrorist attacks in 2017 grouped by high and low economic capacity governments, countries with low economic capacity do not necessarily have insufficient funds to combat terrorist activities – as they suffered substantially less terrorist attacks than high capacity governments.

4. Descriptive Analysis

Reasoning for plots used:

I have chosen to mainly rely on bar plots for my descriptive analysis section as it is the best visual tool to facilitate comparisons between non-continuous data into nominal categories.

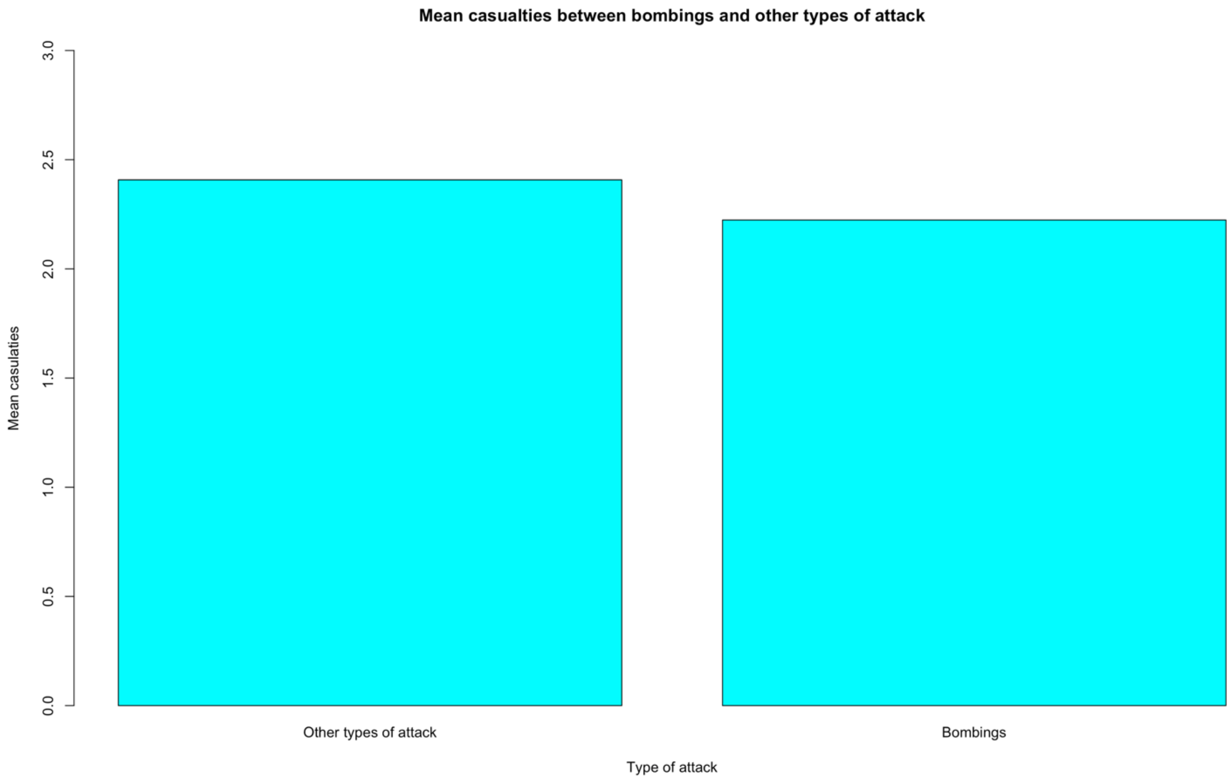
Relationship between bombings and other types of attack

Methodology:

```
bomb.vs.oth <- tapply(gtd$nkill[gtd$nkill != "-999"], gtd$bombing[gtd$nkill != "-
999"], mean)
```

```
       0        1
2.407631 2.223714
```

```
barplot(bomb.vs.oth,
    main = "Mean casualties between bombings and other types of attack",
    ylim = c(0.0, 3.0),
    ylab = "Mean casulaties",
    xlab = "Type of attack",
    col = "cyan",
    names = c("Other types of attack", "Bombings"))
```



Descriptive Analysis:

This chart provides for a comparative assessment of the respective levels of lethality between bombing, and other types of attack in 2017.

As shown above, the mean of casualties for bombing and other types of attack did not vary substantially – with only a difference of 0.17 units.

Thus, the data does not suggest varying levels of difference between bombings and other types of attack in 2017 **(Tested in Hypothesis Test 5.1)**

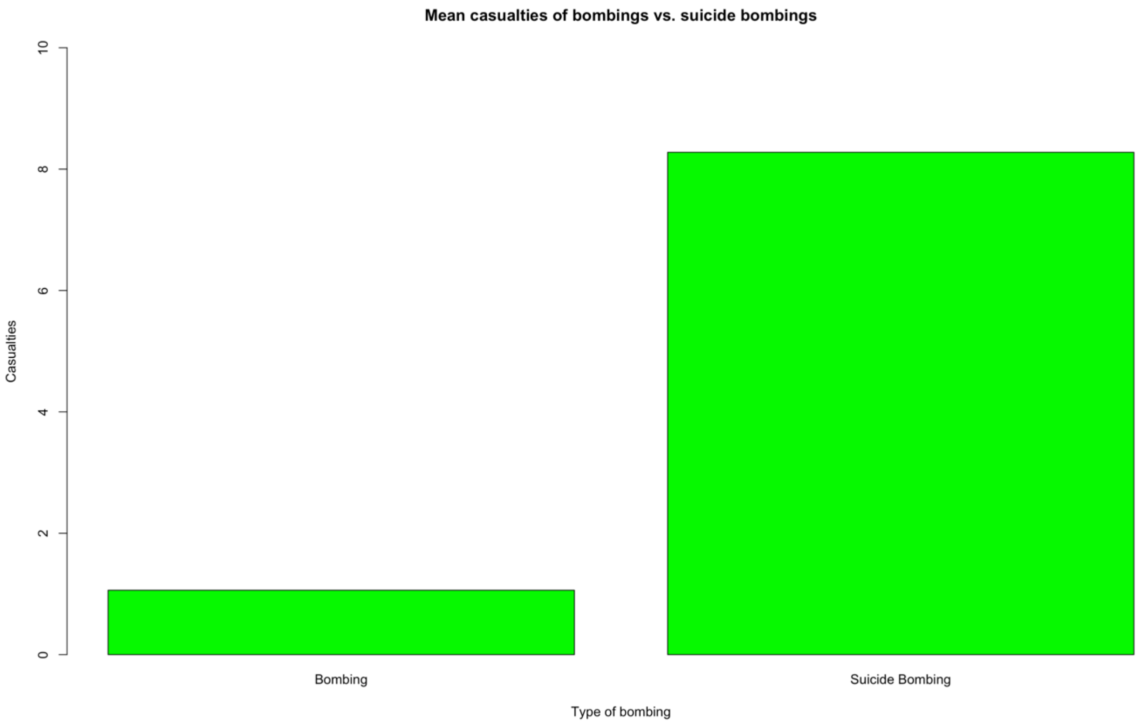
Relationship between lethality and type of bombing

<u>Methodology:</u>

```
mean.of.bombings <- tapply(gtd$nkill[gtd$nkill != "-999" & gtd$attacktype1_txt ==
"Bombing/Explosion"], gtd$suicide[gtd$nkill != "-999" & gtd$attacktype1_txt ==
"Bombing/Explosion"], mean)
```

```
       0        1
1.061333 8.277778
```

```
barplot(mean.of.bombings,
    main = "Mean casualties of bombings vs. suicide bombings",
    xlab = "Type of bombing",
    ylab = "Casualties",
    ylim = c(0,10),
    col = "green",
    names.arg = c("Bombing", "Suicide Bombing"))
```



Descriptive Analysis:

This chart provides for a comparative assessment of the respective levels of lethality between bombing, and suicide bombings attacks in 2017.

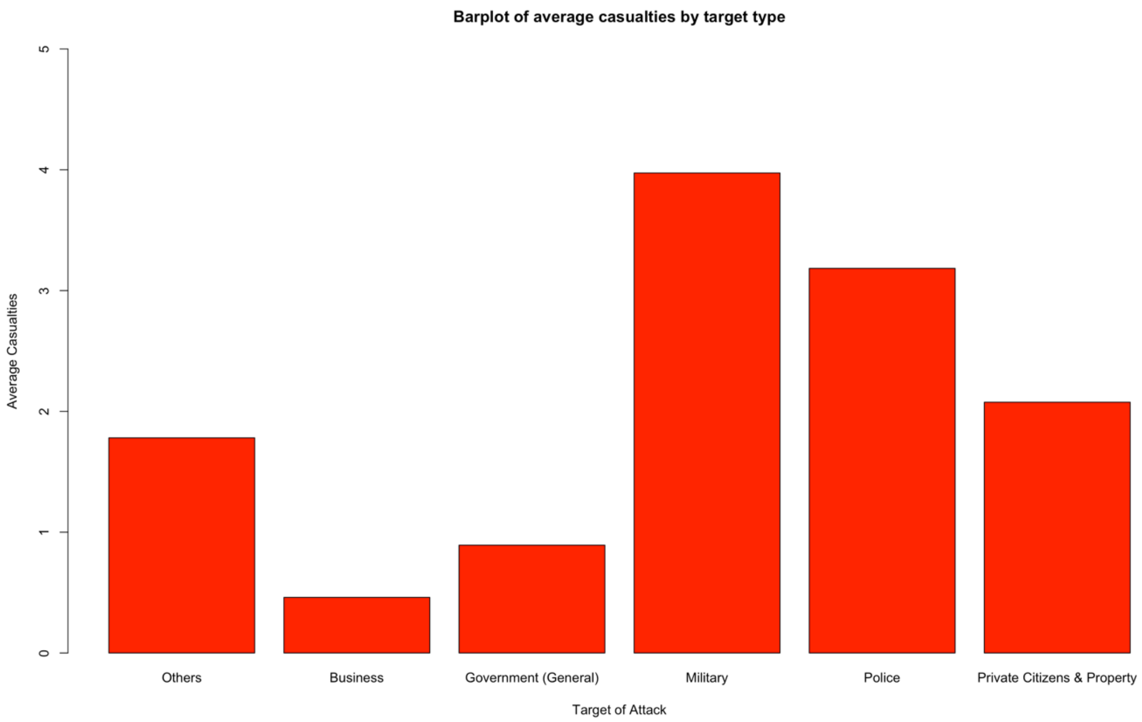
As shown above, the mean casualties for suicide bombing attacks substantially outnumbered mean bombing casualties, with the two constituting with a mean difference of 7.22 units.

Thus, the data suggests that suicide bombings are more lethal than non-suicide bombing attacks. **(Tested in Hypothesis Test 5.2)**

Relationship between lethality of terrorist attack and target type

Methodology:

```
mean.of.casualties <- tapply(gtd$nkill[gtd$nkill != "-999"], gtd$attack.target[gtd$nkill != "-999"], mean)
rownames(mean.of.casualties) = c("Others", "Business", "Government(General)", "Military", "Police", "Private Citizens & Property")
barplot(mean.of.casualties,
    main = "Barplot of average casualties by target type",
    xlab = "Target of Attack",
    ylab = "Average Casualties",
    ylim = c(0.0, 5.0),
    col = "red",
    names.arg = c("Others", "Business", "Government (General)", "Military", "Police",
    "Private Citizens & Property"))
```



Descriptive Analysis:

This chart illustrates the average levels of lethality of terrorist attacks between different target types, in 2017.

As shown above, intended targets displayed a broad range of casualty levels. While military targets peaked at an average of 3.97 fatalities due to terrorist attacks, business targets had the least number of average deaths at 0.46.

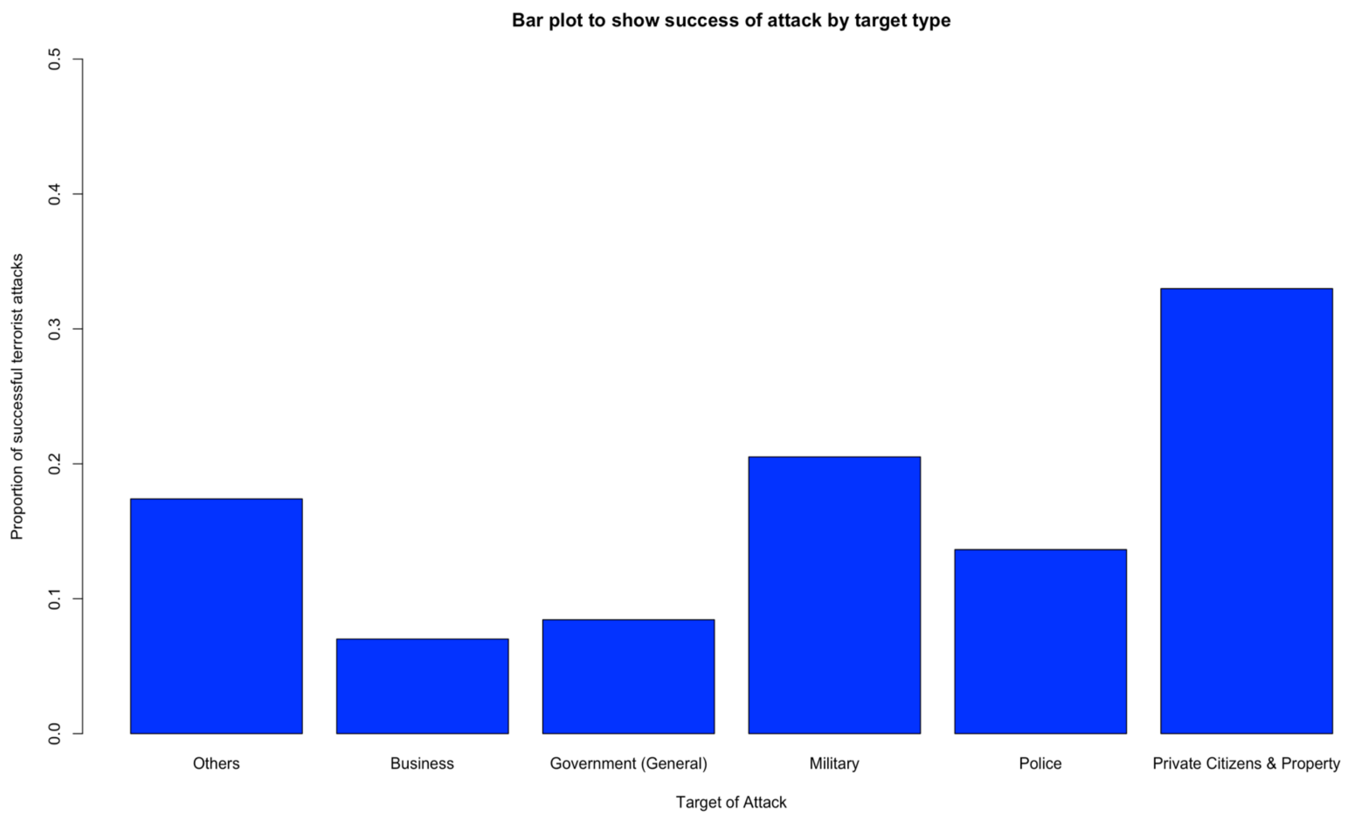
Thus, the data suggests varying levels of lethality between intended targets. **Tested in Hypothesis Test 5.3**

Relationship between success in terrorist attacks and target type

Methodology:

```
prop.cas <- prop.table(table(gtd$attack.target[gtd$success == "1"]))

barplot(prop.cas,
    main = "Bar plot to show success of attack by target type",
    xlab = "Target of Attack",
    ylab = "Proportion of successful terrorist attacks",
    ylim = c(0.0, 0.5),
    col = "blue",
    names.arg = c("Others", "Business", "Government (General)", "Military",
    "Police", "Private Citizens & Property"))
```



Descriptive Analysis:

This chart illustrates the various proportions of success for terrorist attacks in 2017 by target type.

As shown above, while ‘Private Citizens & Property’ constituted the highest proportion of successful attacks with 32.30%, ‘Business’ targets only made up 7.01%.
When combining the trends shown in this plot with the one illustrating lethality by target type, we can see that the ‘Business’ category made up the least number of casualties and successes.

Thus, the data suggests that the success of attack varies between intended targets. **Tested in Hypothesis Test 5.4**

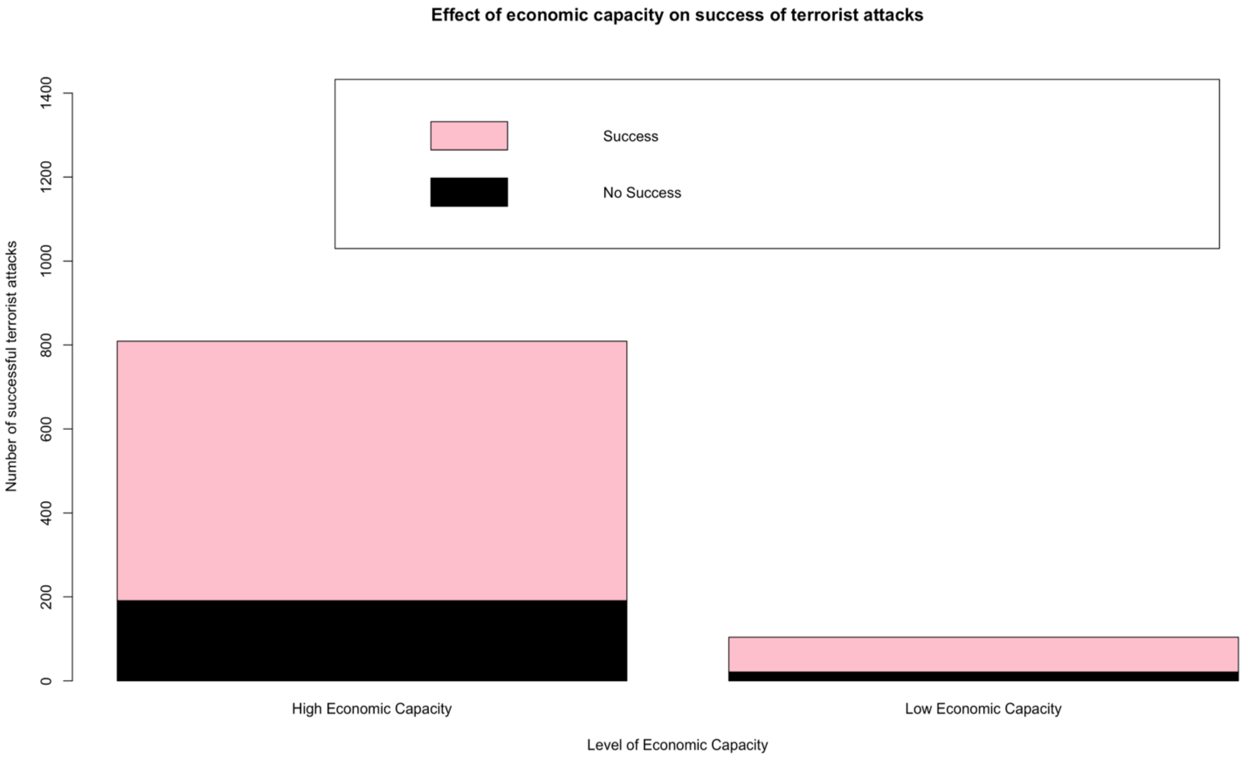
Relationship between success in terrorist attacks and economic capacity

Methodology:

```
lowecontbl <- table(gtd$success, gtd$low.capacity)
colnames(lowecontbl) = c("High Economic Capacity", "Low Economic Capacity")
rownames(lowecontbl) = c("No Success", "Success")
```

```
           High Economic Capacity Low Economic Capacity
No Success                    191                    21
Success                       618                    83
```

```
barplot(lowecontbl,
    main = "Effect of economic capacity on success of terrorist attacks",
    legend = T,
    ylim = c(0.0, 1500),
    col = c("black", "pink"),
    xlab = "Level of Economic Capacity",
    ylab = "Number of successful terrorist attacks")
```



Descriptive Analysis:

This chart showcases the different number of successful terrorist attacks for governments with 'high' and 'low' economic capacity, in 2017.

Prioritizing cases of low economic capacity, one can see that successful terrorist attacks outnumbered unsuccessful terrorist attacks, 83 to 21.

Thus, there is evidence to suggest that low economic capacity governments are less capable, for whichever reason, of preventing successful terrorist attacks from occurring.

Tested in Hypothesis Test 5.5

5. Hypothesis Testing

N.B! – The null and alternative hypotheses stated are based off of the information stipulated in Section 3

5.1. Test if bombings are more or less lethal than other types of attack.

Null hypothesis

The levels of lethality between bombings and other types of attack do not differ.

OR: $H_0 = \mu_1 = \mu_2$

Alternative hypothesis

The levels of lethality between bombings and other types of attack differ.

OR: $H_A = \mu_1 \neq \mu_2$

Welch Two-sample test:

t.test(gtd$nkill[gtd$nkill != "-999"] ~ gtd$bombing[gtd$nkill != "-999"], alt = "two.sided")

```
        Welch Two Sample t-test

data:  gtd$nkill[gtd$nkill != "-999"] by gtd$bombing[gtd$nkill != "-999"]
t = 0.35933, df = 839.66, p-value = 0.7194
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.8207082  1.1885419
sample estimates:
mean in group 0 mean in group 1 
       2.407631        2.223714 
```

At the 5% error level:

As the p value of 0.7194 is greater than the alpha ($p > 0.05$), the result is insignificant, and we cannot reject the null hypothesis at the 5% error level.

At the 10% error level:

As the p value of 0.7194 is greater than the alpha ($p > 0.10$), the result is insignificant, and we cannot reject the null hypothesis at the 10% error level.

At the 1% error level:

As the p value of 0.7194 is greater than the alpha ($p > 0.01$), the result is insignificant, and we cannot reject the null hypothesis at the 1% error level.

5.2. Test if suicide bombings are more lethal than non-suicide bombings.

Null hypothesis

Suicide bombings are less lethal than non-suicide bombings.

OR: $H_0 = \mu_1 < \mu_2$

Alternative hypothesis

Suicide bombings are more lethal than non-suicide bombings.

OR: $H_A = \mu_1 > \mu_2$

Welch's Two-sample test:

t.test(gtd$nkill[gtd$nkill != "-999" & gtd$attacktype1_txt == "Bombing/Explosion"] ~ gtd$suicide[gtd$nkill != "-999" & gtd$attacktype1_txt == "Bombing/Explosion"], conf = 0.95, var.eq = FALSE)

```
	Welch Two Sample t-test

data:  gtd$nkill[gtd$nkill != "-999" & gtd$attacktype1_txt == "Bombing/Explosion"] by gtd$suicide[gtd$nkill != "-999" & gtd$attacktype1_txt == "Bombing/Explosion"]
t = -3.1083, df = 71.618, p-value = 0.002698
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -11.844961  -2.587928
sample estimates:
mean in group 0 mean in group 1 
       1.061333        8.277778 
```

Here the p-value of 0.002698 the sum of two tails and will thus be halved to produce the p-value for this experiment, as the premise of the alternative hypothesis is that "suicide bombings are more lethal than non-suicide bombings.

```
> 0.002698/2
[1] 0.001349
```

At the 5% error level:

As the p value of 0.001349 is less than the alpha ($p < 0.05$), the result is significant, and we can reject the null hypothesis at the 5% error level.

At the 10% error level:

As the p value of 0.001349 is less than the alpha ($p < 0.10$), the result is significant, and we can reject the null hypothesis at the 10% error level.

At the 1% error level:

As the p value of 0.001349 is less than the alpha ($p < 0.01$), the result is significant, and we can reject the null hypothesis at the 1% error level.

5.3. Does the lethality of an attack differ based on the intended target?

Null hypothesis

The lethality of an attack does not differ based on the intended target.

OR: $H_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_0$

Alternative hypothesis

The lethality of an attack differs based on the intended target.

OR: $H_A = \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5 \neq \mu_0$

ANOVA Test

anova.test <- aov(formula = gtd$nkill[gtd$nkill != "-999"] ~ gtd$attack.target[gtd$nkill != "-999"])

```
Call:
   aov(formula = gtd$nkill[gtd$nkill != "-999"] ~ gtd$attack.target[gtd$nkill !=
    "-999"])

Terms:
                gtd$attack.target[gtd$nkill != "-999"] Residuals
Sum of Squares                                 1087.14  55586.71
Deg. of Freedom                                      5       939

Residual standard error: 7.694009
Estimated effects may be unbalanced
```

summary(anova.test)

```
                                          Df Sum Sq Mean Sq F value  Pr(>F)
gtd$attack.target[gtd$nkill != "-999"]    5   1087   217.4   3.673 0.00269 **
Residuals                               939  55587    59.2
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

At 5% error level:

As the p value of 0.00269 is less than the alpha ($p < 0.05$), the result is significant, and we can reject the null hypothesis at the 5% error level.

At 10% error level:

As the p value of 0.00269 is less than the alpha ($p < 0.10$), the result is significant, and we can reject the null hypothesis at the 10% error level.

At 1% error level:

As the p value of 0.00269 is less than the alpha ($p < 0.01$), the result is significant, and we can reject the null hypothesis at the 1% error level.

5.4. Does the success of an attack differ based on the intended target?

Null hypothesis

The likelihood of success does not vary across the target types.

OR: $H_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_0$

Alternative hypothesis

The likelihood of success varies across target type.

OR: $H_A = \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5 \neq \mu_0$

Chi-squared Test

```
targsuc <- table(gtd$attack.target, gtd$success)
colnames(targsuc) = c("No Success", "Success")
rownames(targsuc) = c("Others", "Business", "Government(General)", "Military", "Police",
"Private Citizens & Property")
```

```
                            No Success Success
Others                             105     134
Business                             9      54
Government(General)                 22      65
Military                            53     158
Police                              23     105
Private Citizens & Property         18     254
```

```
chisq.test(targsuc, correct = T)
```

```
	Pearson's Chi-squared test

data:  targsuc
X-squared = 105.68, df = 5, p-value < 2.2e-16
```

At 5% error level:

As the p value of '2.2e – 16' is less than the alpha ($p < 0.05$), the result is significant, and we can reject the null hypothesis at the 5% error level.

At 10% error level:

As the p value of '2.2e – 16' is less than the alpha ($p < 0.10$), the result is significant, and we can reject the null hypothesis at the 10% error level.

At 1% error level:

As the p value of '2.2e – 16' is less than the alpha ($p < 0.01$), the result is significant, and we can reject the null hypothesis at the 1% error level.

5.5. Does low economic capacity affect the success of terrorist attacks?

Null hypothesis

Governments with low levels of economic capacity have insufficient resources to effectively combat terrorist activities.

OR: $H_0 = H_1 < 0$

Alternative hypothesis

Governments with low levels of economic capacity have sufficient resources to effectively combat terrorist activities.

OR: $H_A = H_1 > 0$

Chi-squared Test

```
lowecontbl <- table(gtd$success, gtd$low.capacity)
colnames(lowecontbl) = c("High Economic Capacity", "Low Economic Capacity")
rownames(lowecontbl) = c("No Success", "Success")
```

```
           High Economic Capacity Low Economic Capacity
No Success                    191                    21
Success                       618                    83
```

```
chisq.test(lowecontbl)
```

```
        Pearson's Chi-squared test with Yates' continuity correction

data:  lowecontbl
X-squared = 0.4271, df = 1, p-value = 0.5134
```

At 5% error level:

As the p value of 0.5134 is more than the alpha ($p > 0.05$), the result is insignificant, and we cannot reject the null hypothesis at the 5% error level.

At 10% error level:

As the p value of 0.5134 is more than the alpha ($p > 0.10$), the result is insignificant, and we cannot reject the null hypothesis at the 10% error level.

At 1% error level:

As the p value of 0.5134 is more than the alpha ($p > 0.01$), the result is insignificant, and we cannot reject the null hypothesis at the 1% error level.

Conclusion and limitations

1. **Is the analysis above (Section 4 and 5) consistent with the expectations outlined in Section 3?**

   Conclusion for differences in lethality between bombing and non-bombing types of attack:

   With regards to this experiment, the null hypothesis affirmed that the levels of lethality did not differ between bombing and non-bombing attacks, while the alternative hypothesis stated that the degree of fatality varied between the two.

   As shown by the bar plot in Section 4 as well as the Welch two-sample test in Section 5, the results both support the null hypothesis and refute the alternative hypothesis. In the case of the former, the plot illustrated that mean casualties associated to bombing and non-bombing types reflected only a 0.17 unit difference. Furthermore, the Welch two-sample test supported the null hypothesis at a 10%, 5% and 1% error level.

   Thus, the analysis in Section 4 and 5 is consistent with the expectations outlined in Section 3.

Conclusion for differences in lethality between suicide-bombings and non-suicide bombings:

With regards to this experiment, the null hypothesis affirmed that suicide-bombings are less lethal than non-suicide bombings, while the alternative hypothesis stated that suicide bombings would have more fatalities than non-suicide bombings.

As shown by the bar plot in Section 4 as well as the Welch two-sample test in Section 5, the results both rejected the null hypothesis and supported the alternative hypothesis. In the case of the former, the plot illustrated that mean casualties associated to bombing and non-suicide bombing reflected a substantial 7.22 unit difference. Furthermore, the Welch two-sample test supported the null hypothesis at a 10%, 5% and 1% error level.

Thus, the analysis in Section 4 and 5 is consistent with the expectations outlined in Section 3.

Conclusion for the relationship between lethality and target types:

With regards to this experiment, the null hypothesis affirmed that there would be no difference in lethality between target types, while the alternative hypothesis stated that there would be difference between the casualties of intended targets.

As shown by the bar plot in Section 4 as well as the ANOVA Test carried out in Section 5, the results of both refute the null hypothesis and support the alternative hypothesis.

In the former, the average casualties associated to each target type reflected a range of 0.46 to 3.97 fatalities. Furthermore, the latter rejected the null hypothesis at a 10%, 5% and 1% error level.

Thus, the analysis in Section 4 and 5 is consistent with the expectations outlined in Section 3.

Conclusion for relationship between success of attack and target types:

With regards to this experiment, the null hypothesis affirmed that the likelihood of success did not differ based on the intended target, while the alternative hypothesis stated that success in attacks would vary depending of target type.

As shown by the bar plot in Section 4 as well as the Chi-squared test in Section 5, the results both refute the null hypothesis and support the alternative hypothesis.
In the former, the number of successes associated to each target type reflected a range of proportions from 0.07 to 0.33. Furthermore, the latter rejected the null hypothesis at a 10%, 5% and 1% error level.

Thus, the analysis in Section 4 and 5 is consistent with the expectations outlined in Section 3.

Conclusion for relationship between low economic capacity and success of attack:

With regards to this experiment, the null hypothesis affirmed that governments with low economic capacity have insufficient resources to effectively combat terrorist activities while the alternative hypothesis stated that low capacity countries have sufficient means for combatting terrorism.

As shown by the bar plot in Section 4 as well as the Chi-squared test in Section 5, the results both support the null hypothesis and refute the alternative hypothesis.

In the former, the plot illustrated that in low economic-capacity countries, successful terrorist attacks outnumbered unsuccessful attempts 83 to 21. Furthermore, the Chi-squared test supported the null hypothesis at a 10%, 5% and 1% error level.

Thus, the analysis in Section 4 and 5 is consistent to the expectations outlined in Section 3.

2. **Do you see any limitations with the current analysis? Please explain your answer.**

Perils of hypothesis testing

<u>Most null hypothesis are obviously false on *a priori* grounds:</u>

As illustrated by this essay, most null hypotheses were trivial strawmen that could easily be disproved without the need of time-consuming data collection.

E.g.: *"There are no differences in lethality between intended targets of terrorist attacks"*

In null hypotheses such as this one and others like it, there is no insight into the magnitude of the differences, and hardly advance the scope of science.

<u>The alpha level is arbitrary and without theoretical basis:</u>
Using a fixed α-level arbitrarily classifies results into meaningless categories like 'significant' and 'non-significant'. These terms do not relate to biological importance but only to arbitrary academic classification.

Problem with 'low economic capacity experiment'

<u>Misleading data:</u>

The gtd$low.capacity column does not provide an individual measure of low economic capacity for each government, but rather, lists every instance whereby a government targeted by a terrorist attack was one of high or low capacity – sometimes listing the same country twice or more.

Due to these inconsistencies, while the null and alternative hypothesis questioned whether governments of low capacity had sufficient or insufficient resources to combat terrorism **as a whole**, the statistical tool of the chi-squared test only investigated whether the average successful **instance** of a terrorist attack involving a government of low capacity had sufficient funds to combat terrorist attacks.

Thus, the results found in the descriptive data and hypothesis testing do not reflect the data.