Analysis of factors related to highway safety

by

M.S., Wenxiao Zhou, Haoxi Ma

University of Connecticut, 2019

Project Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Statistics

in the

Department of Statistics

Advisor: Haiying Wang

Winter 2019

Abstract

This dataset contains 385 cases for 3ST(three-way stop controlled) intersections. In order to analyze highway safety, we select six factors, which are Annual Average Daily Traffic(AADT) for major roads, Annual Average Daily Traffic(AADT) for minor roads, Lighting, Approach of Left-turn, Approach of Right-turn and Skew angle then focus on the relationship between these factors and whole crashes which is the sum of PDO, B+C and K+A. After mathematical methods, we make some conclusions and expansions.

Key words:

Traffic for major and minor roads, Lighting, Left-turn, Right-turn, skew angle, whole crashes

1. Introduction and Description of Data

I. Research questions

1.Indentify the quantitative and qualitative variables.
2.Analyze the relationship between six factors (Annual Average Daily Traffic(AADT) for major roads, Annual Average Daily Traffic(AADT) for minor roads, Lighting, Left-turn, Right-turn, Skew angle) and whole crashes which is the sum of PDO, B+C and K+A.
3.Judge which distribution mostly satisfies the quantitative variables and make some statistic inference.
4.Compare the 3ST intersections' whole crashes with the other two kinds of intersection and make some comment.

II. Variables descriptions

We can draw a scatter matrix for all factors:

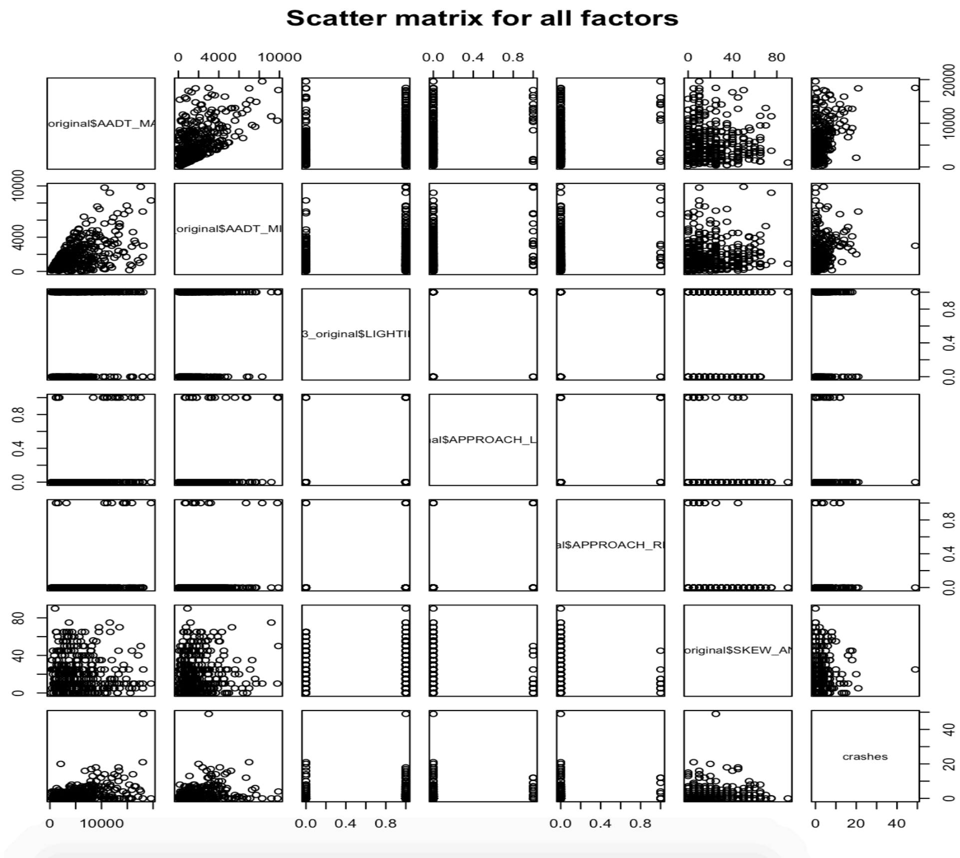

Figure 1.1 Scatter matrix for all factors

According to the Figure 1.1, the identities of all factors are obvious. Therefore,
Quantitative variables:
AADT for major roads(noted as major in the following), AADT for minor roads(noted as minor in the following), whole crashes, Skew angle.
Qualitative variables:
Lighting, Approach of Right-turn, Approach of Left-turn.

2. Exploratory Data Analysis

In this section, for the qualitative variables, we just concentrate on their correlation with major, minor and whole crashes. This makes sense and can help improve efficiency. Besides, we divided major and minor into four levels: “Highest”, “High”, “Low” and “Lowest”.

I. Lighting

We construct barplot for major vs lighting:

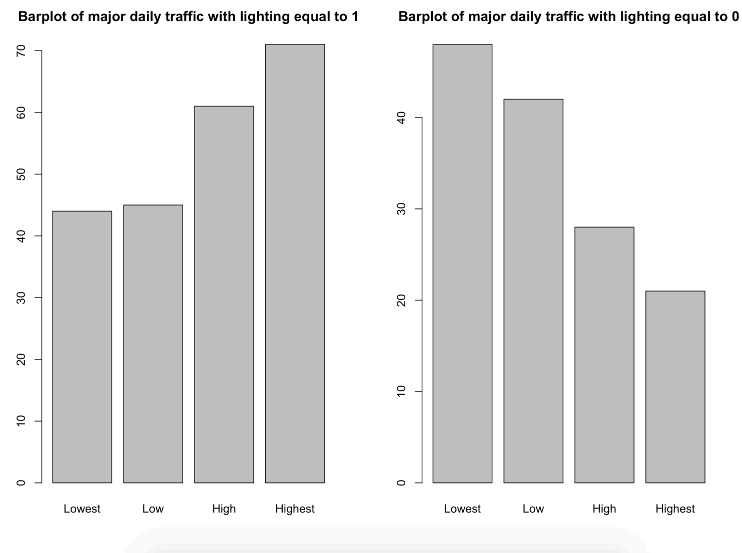

Figure 2.1.1 Barplot for major vs lighting

From the figure above, we can infer that major and lighting have a positive correlation. After that, we construct barplot for minor vs lighting:

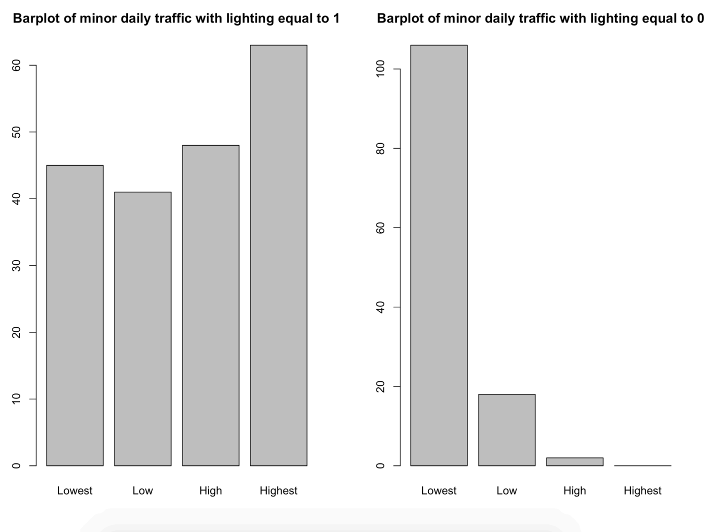

Figure 2.1.2 Barplot for minor vs lighting

From the figure, we can infer that minor and lighting have a positive correlation.

Then we draw overlaid histogram for lighting and whole crashes:

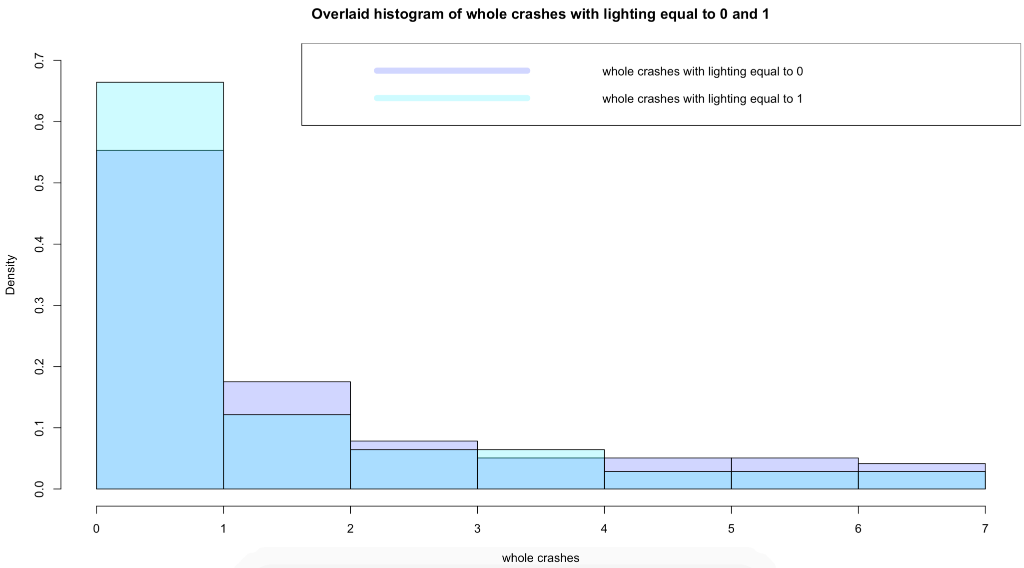

Figure 2.1.3 Overlaid histogram for lighting and whole crashes

According to the figure 2.1.3, whole crashes under two conditions of lighting have almost the same distribution so we can infer that lighting has no significant influence on whole crashes. And we construct a side-by-side boxplot to verify it:

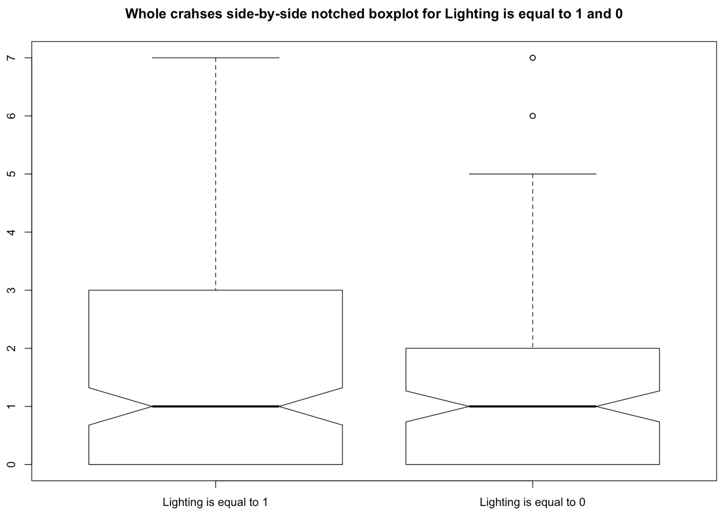

Figure 2.1.4 Whole crashes side-by-side boxplot for Lighting

Because two notches overlap, there is no significant difference between two medians. To sum up, Lighting has a positive correlation with major and minor but does not influence whole crashes.

II. Approach of Left-turn

We construct barplot for major vs Approach of Left-turn:

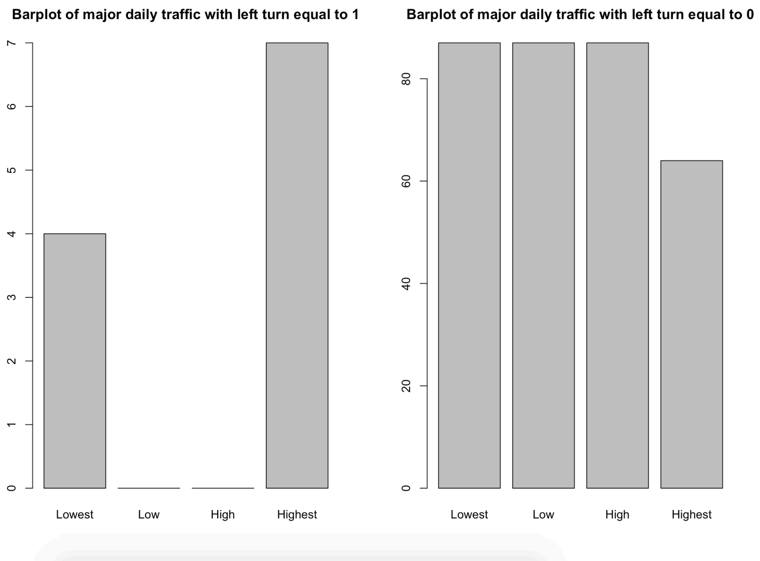

Figure 2.2.1 Barplot for major vs Approach of Left-turn

From the figure above, although we have a small sample size when Approach of Left-turn is 1, the comparison of the shape under two conditions can show that major and Approach of Left-turn have a positive correlation.
After that, we construct barplot for minor vs Approach of Left-turn:

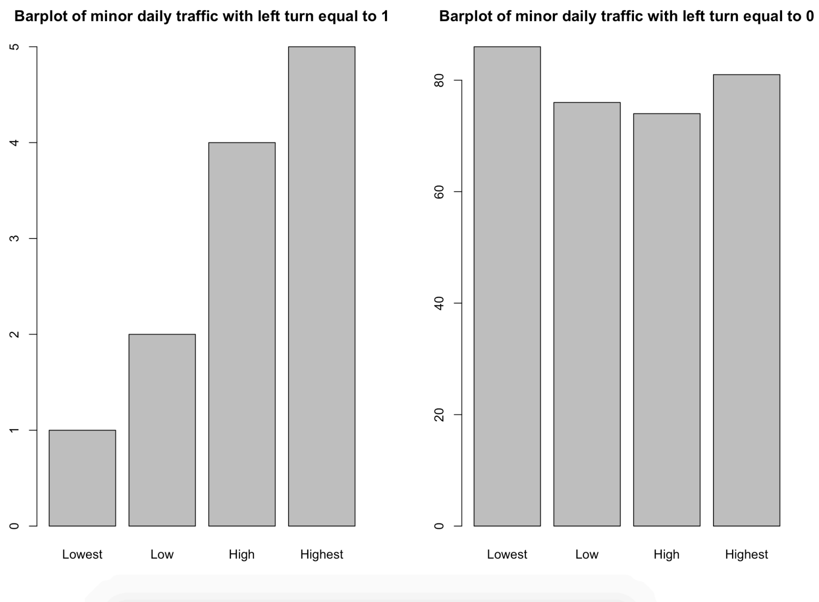

Figure 2.2.2 Barplot for minor vs Approach of Left-turn

From the figure, as mentioned before, we can infer that minor and Approach of Left-turn have a positive correlation.
Then we draw overlaid histogram for Approach of Left-turn and whole crashes:

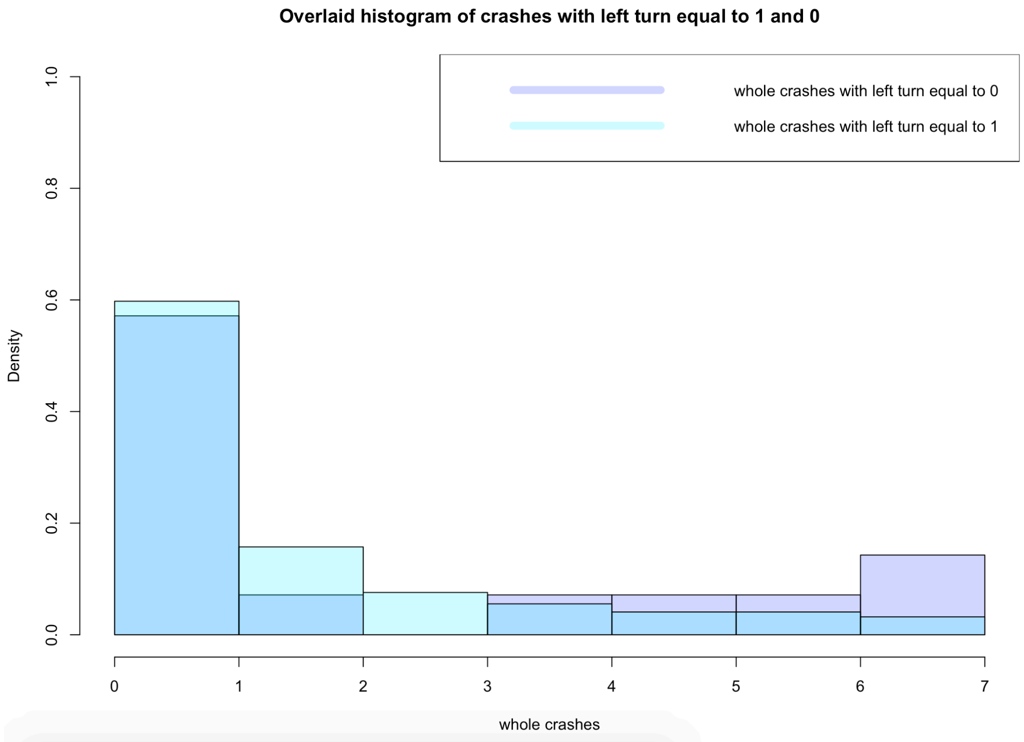

Figure 2.2.3 overlaid histogram for Approach of Left-turn and whole crashes

According to the figure 2.2.3, whole crashes under two conditions of Approach of Left-turn have almost the same distribution so we can infer that Approach of Left-turn has no significant influence on whole crashes. And we construct a side-by-side boxplot to verify it:

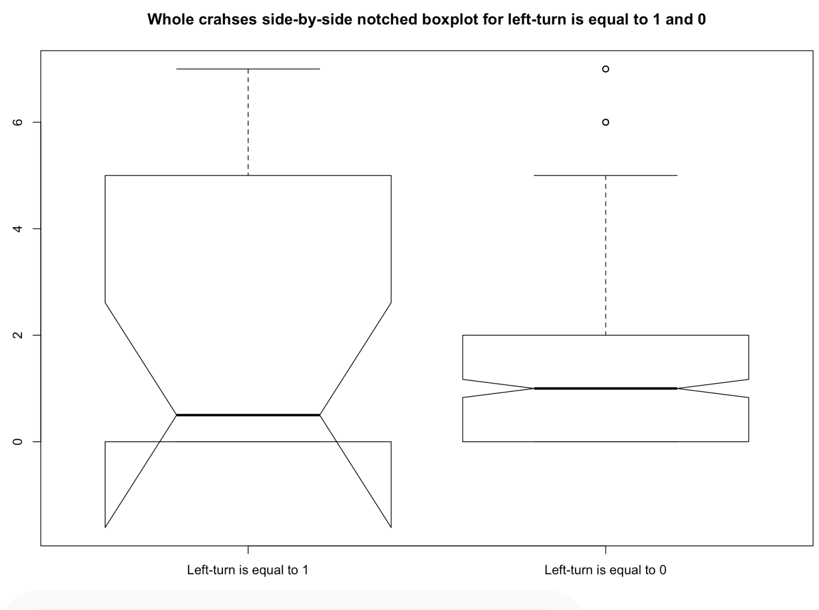

Figure 2.2.4 Whole crashes side-by-side boxplot for Approach of Left-turn

Because two notches overlap, there is no significant difference between two medians.

To sum up, Approach of Left-turn has a positive correlation with major and minor but does not influence whole crashes.

III. Approach of Right-turn

We construct barplot for major vs Approach of Right-turn:

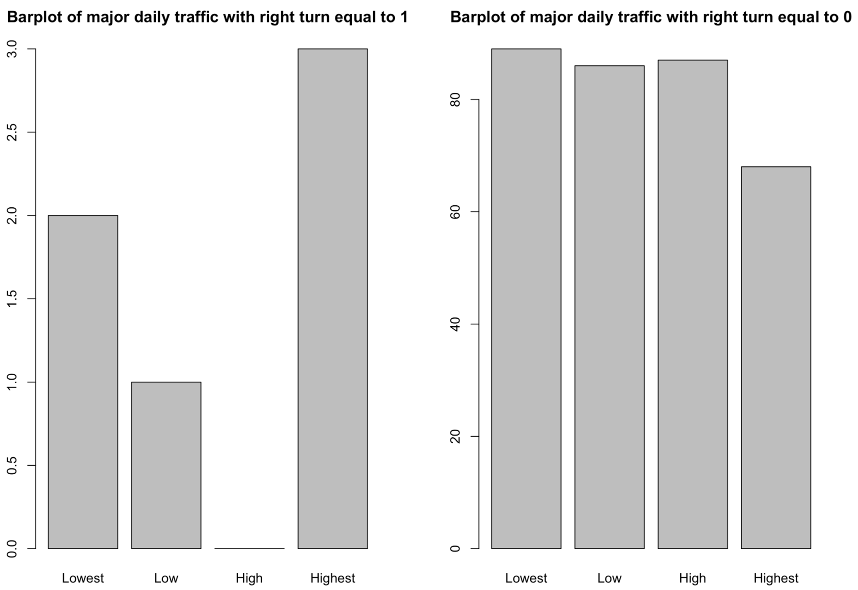

Figure 2.3.1 Barplot for major vs Approach of Right-turn

From the figure above, although we have a small sample size when Approach of Right-turn is 1, the comparison of the shape under two conditions can show that major and Approach of Right-turn may have a positive correlation.

After that, we construct barplot for minor vs Approach of Right-turn:

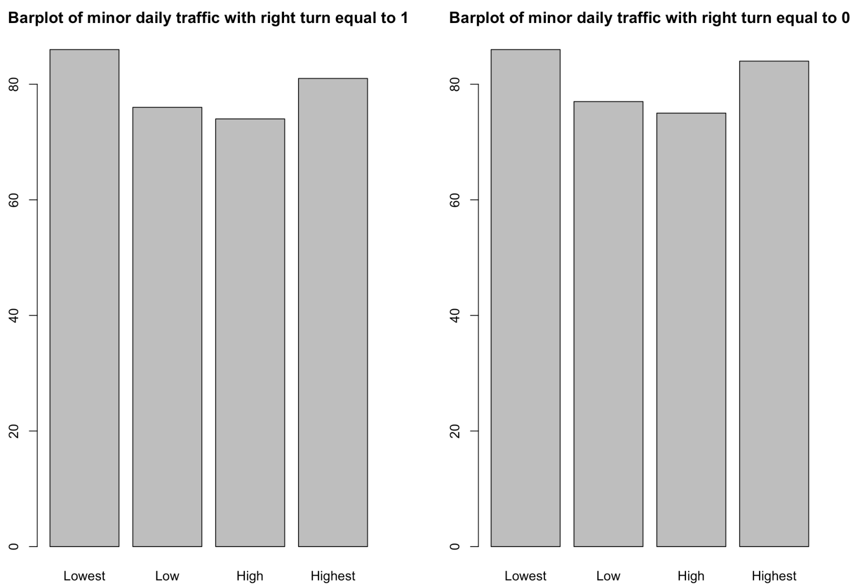

Figure 2.3.2 Barplot for minor vs Approach of Right-turn

From the figure, minor's distribution has almost the same underlying shape so we can infer that minor and Approach of Right-turn are unrelated.
Then we draw overlaid histogram for Approach of Right-turn and whole crashes:

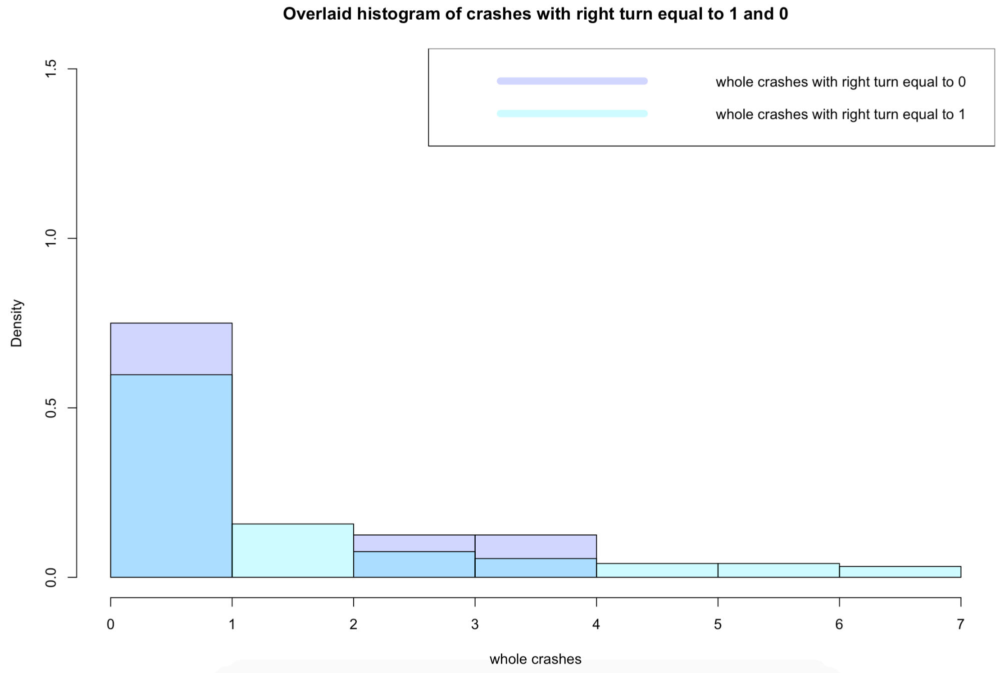

Figure 2.3.3 overlaid histogram for Approach of Right-turn and whole crashes

According to the figure 2.3.3, whole crashes under two conditions of Approach of Right-turn have almost the same distribution so we can infer that Approach of Right-turn has no significant influence on whole crashes
To sum up, Approach of Right-turn has a positive correlation with major and is unrelated to minor and whole crashes.

IV. Skew angle
We construct Skew angle side-by-side notched boxplot for major:

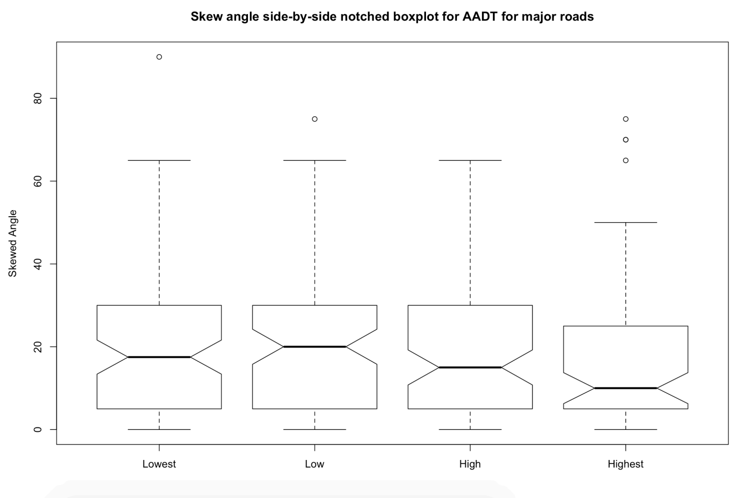

Figure 2.4.1 Skew angle side-by-side notched boxplot for major

According to the figure above, we can find that notches of “Lowest”, “Low” and “High” overlap so medians of theses levels are not significantly different. What’s more, whether the median of “Low” level is equal to the median of “Highest” is not clear. So we do Wilcoxon rank sum test to check it:

```
> wilcox.test(ST32$SKEW_ANGLE,ST34$SKEW_ANGLE,alternative = "two.sided")

        Wilcoxon rank sum test with continuity correction

data:  ST32$SKEW_ANGLE and ST34$SKEW_ANGLE
W = 3519, p-value = 0.13
alternative hypothesis: true location shift is not equal to 0
```

Figure 2.4.2 Output of Wilcoxon rank sum test

Because P-value is 0.13>0.05, we can not reject the null hypothesis. Then medians of these two levels are same at 5% significance. Therefore, we conclude that Skew angle and major are unrelated.

After that, we construct Skew angle side-by-side notched boxplot for minor:

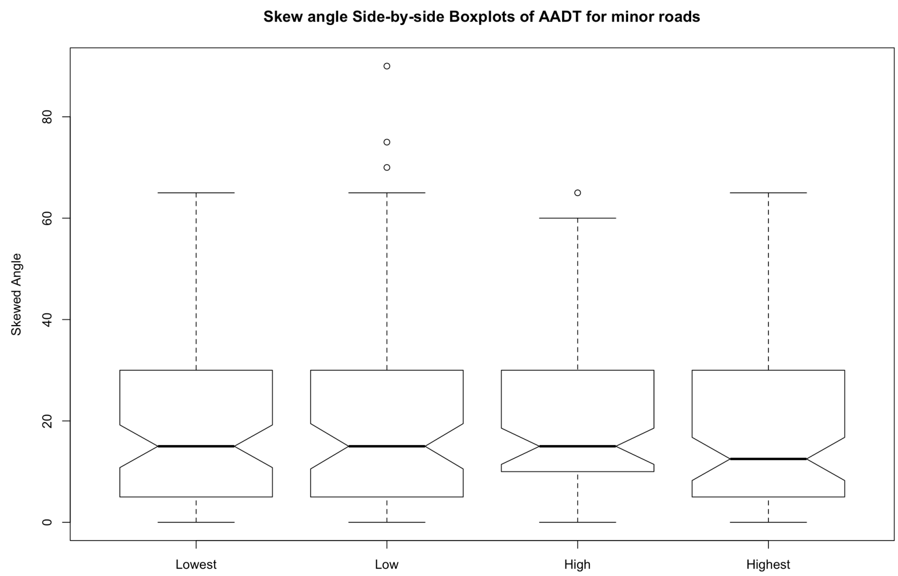

Figure 2.4.3 Skew angle side-by-side notched boxplot for minor

From the figure 2.4.3, we can find that all notches overlap so there is no significant difference between these pairs of medians. Therefore, we conclude that that Skew angle and minor are unrelated.

Then we construct overlaid histogram for Skew angle and whole crashes:

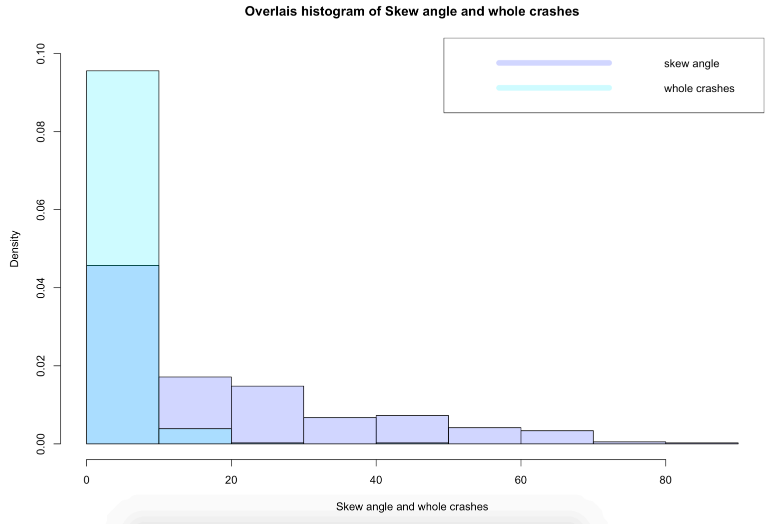

Figure 2.4.4 overlaid histogram for Skew angle and whole crashes

It seems that the correlation of these two factors is not strong.
For checking it, we construct scatter plot for Skew angle and whole crashes:

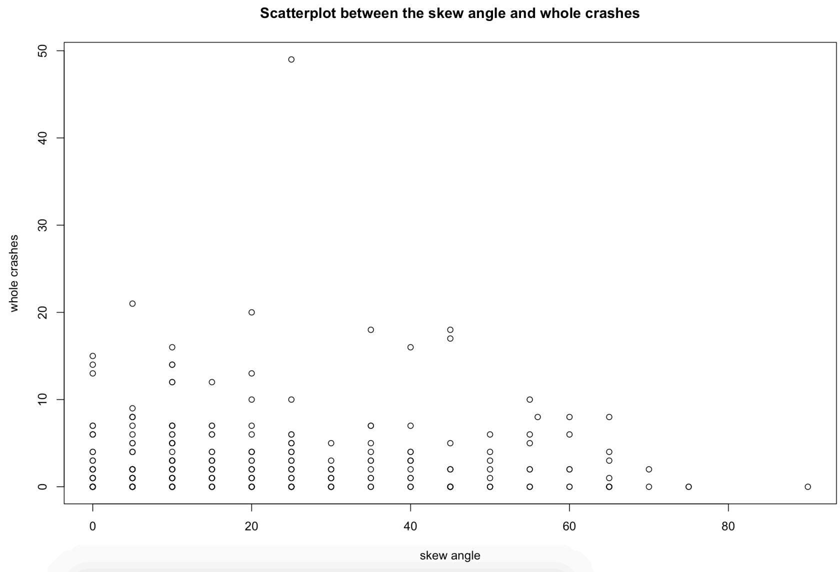

Figure 2.4.5 scatter plot for Skew angle and whole crashes

These graph comes to the same conclusion.
Above all, Skew angle is unrelated to major, minor and whole crashes.

V. Minor

First, we draw histograms for minor and whole crashes:

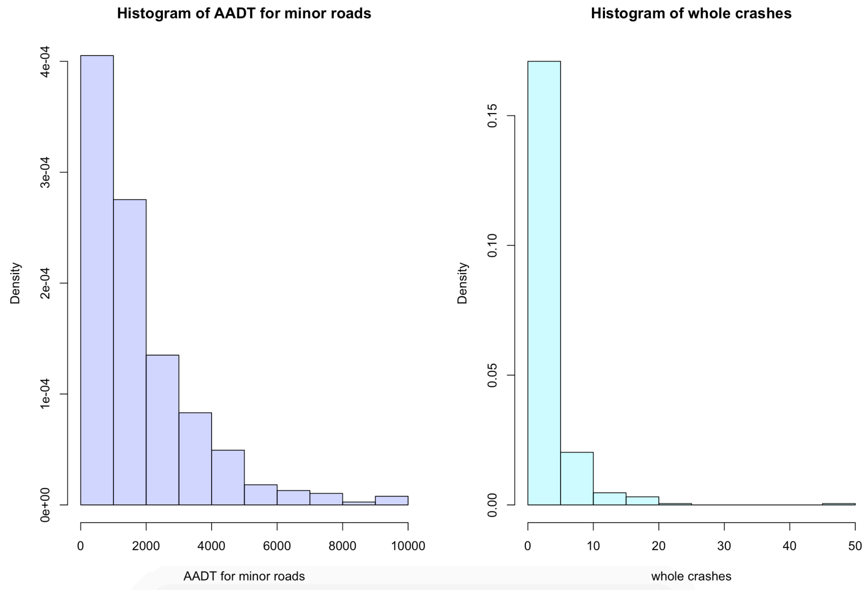

Figure 2.5.1 Histograms for minor and whole crashes

From the figure 2.5.1, we can infer that maybe these two factor has a correlation but this need to be verified. So we construct a whole crashes side-by-side notches boxplot for 4 levels of minor:

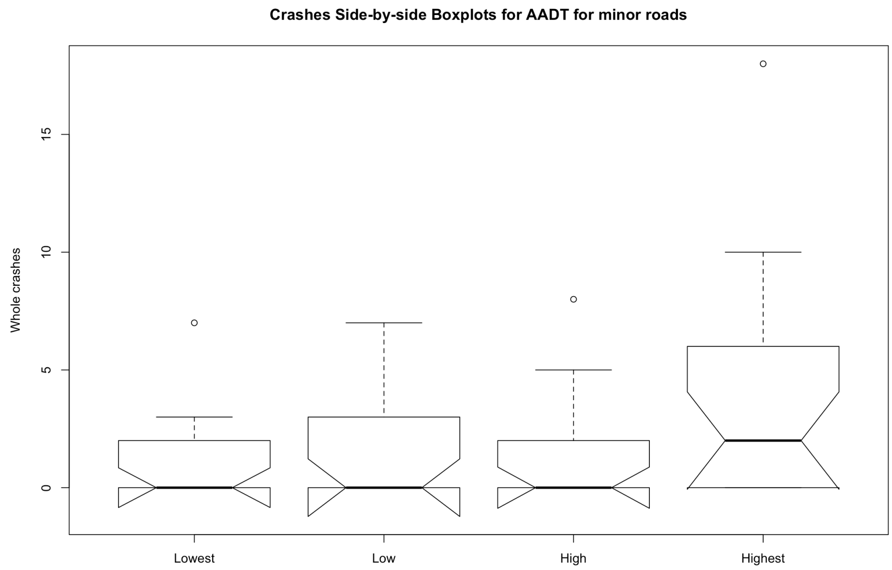

Figure 2.5.2 whole crashes side-by-side notches boxplot for 4 levels of minor

According to the figure above, we can find that all notches overlap so the corresponding medians are equal at 5% significance. Therefore, we think that minor does not have significant influence on whole crashes.

We also want to know the relationship between major and minor and we draw overlaid histogram for these two factors:

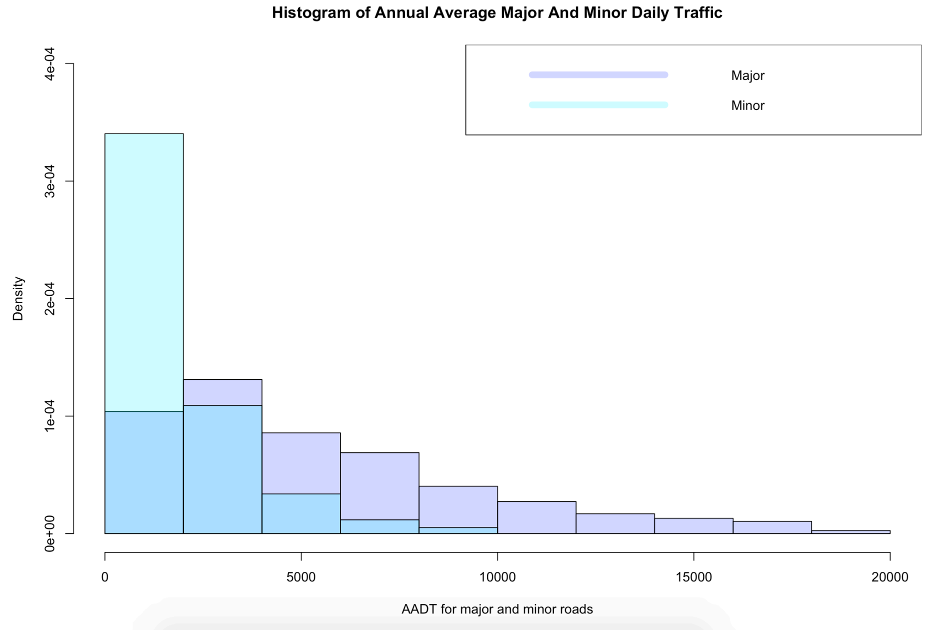

Figure 2.5.3 overlaid histogram for major and minor

From the figure 2.5.3, we can infer that maybe these two factor has a positive correlation. In order to check it, we construct a scatter plot for major and minor:

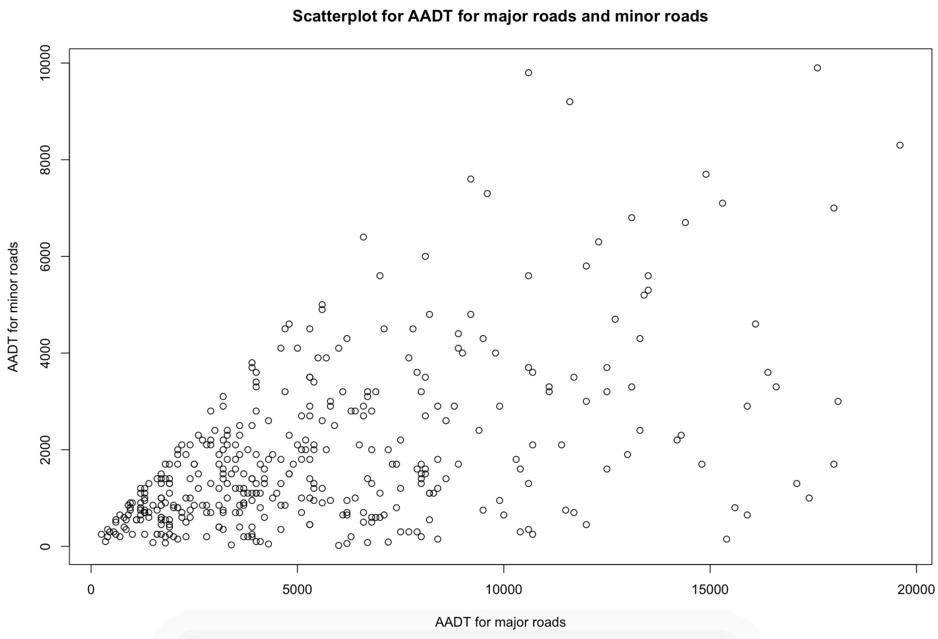

Figure 2.5.4 scatter plot for major and minor

According to the figure above, we can find it obvious that major and minor has a positive correlation.

To sum up, minor is positively related to major and unrelated to whole crashes.

VI. Major

First, we draw histograms for major and whole crashes:

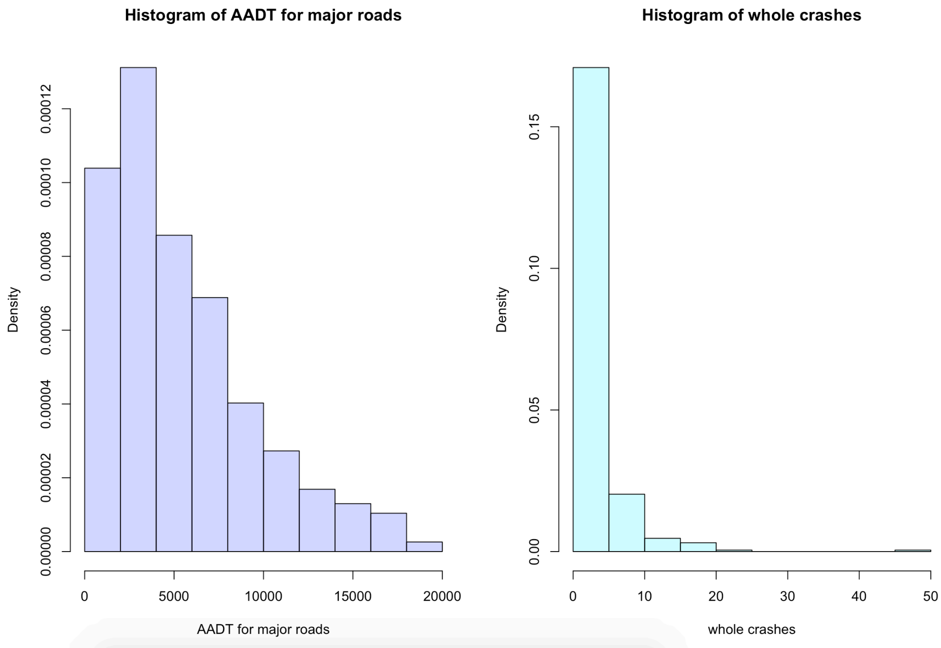

Figure 2.6.1 Histograms for major and whole crashes

From the figure 2.6.1, we can infer that maybe these two factor has a correlation but this need to be verified. So we construct a whole crashes side-by-side notches boxplot for 4 levels of major:

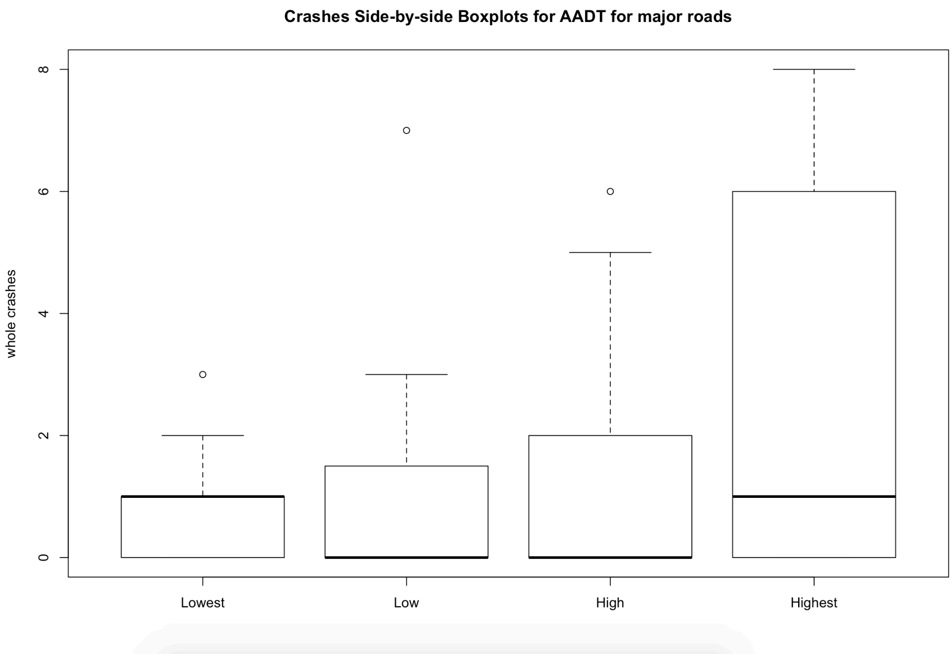

Figure 2.6.2 whole crashes side-by-side notches boxplot for 4 levels of major

According to the figure above, we can find that the corresponding medians may be not equal which suggests that whole crashes are positively correlated to major. To check it, we draw a scatter plot for major and whole crashes:

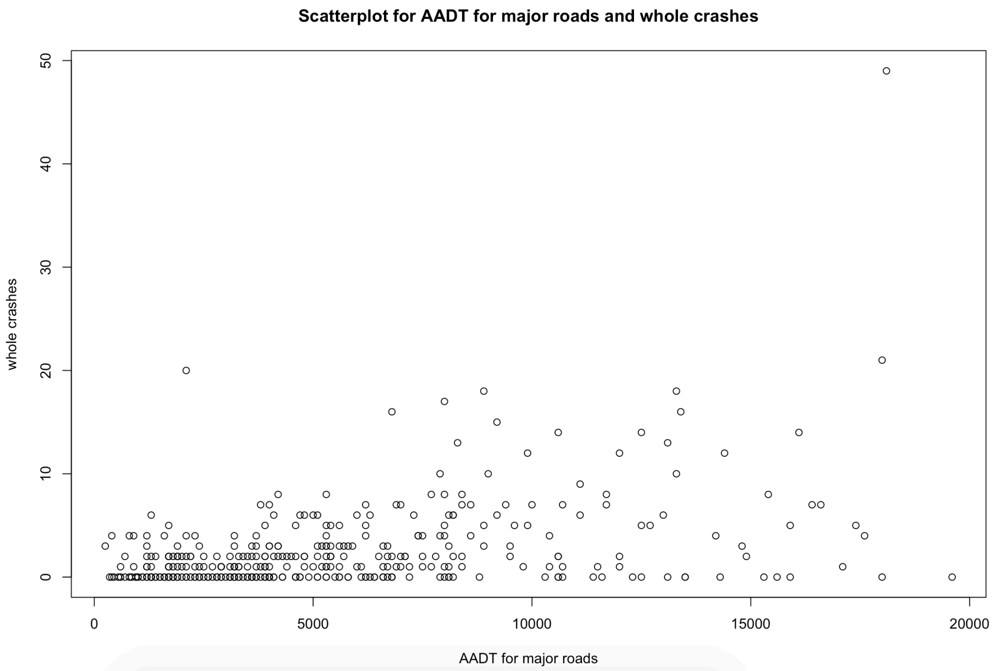

Figure 2.6.3 scatter plot for major and whole crashes

According to the figure above, we can find it obvious that major and whole crashes has a positive correlation.

VII. Analysis of whole factors together

After knowing each factors' effect on whole crashes, we put all factors together and draw some conclusion. This is helpful to determine how the intersection between factors will influence the whole crashes

Because we have known that minor is not related to the whole crashes but positively correlated to major. Besides, major and whole crashes have a positive correlation. So it is reasonable to just consider "High" major with "High" minor and "Low" major with "Low" minor.

What's more, we do not consider Skew angle because it is unrelated to major, minor and whole crashes.

Then we give some notation: "High" major with "High" minor noted as 1, Lighting=1 noted as 1, Left-turn=1 noted as1, Right-turn=1 noted as 1. Under this condition, "High" major with "High" minor and Lighting=1 and Left-turn=1 noted and Right-turn=1 is noted as 1111.

Then we draw a boxplot:

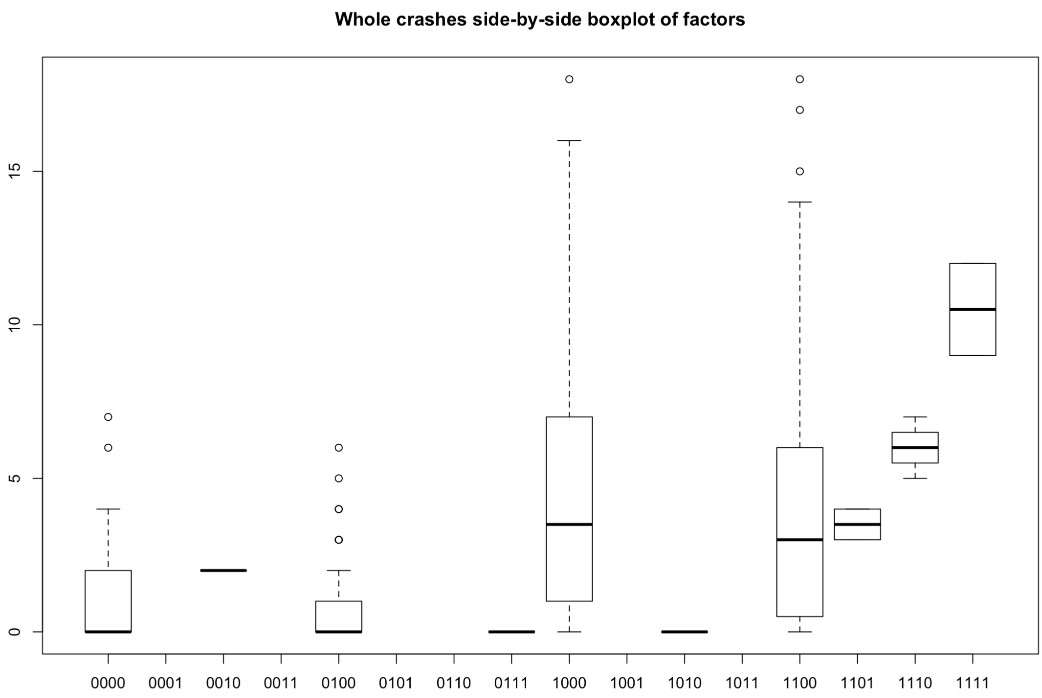

Figure 2.7.1 whole crashes side-by-side boxplot for factors

According to figure 2.7.1, it is obviously that if we want to have the lowest whole crashes, we should to choose "Low" major.

Noted that 1010 is also very small, this means that there exist intersected factors which influence the whole crashes.

3. Statistical Inference

In this section, we make statistical inferences among the quantitative variables.

I. Major and Minor

First, we construct a histogram for AADT for major roads and calculate the mean and variance: $\bar{X} = 5162.869, S^2 = 12232872$. Since the variance is much more larger than mean, we construct a gamma distribution for AADT for major roads.

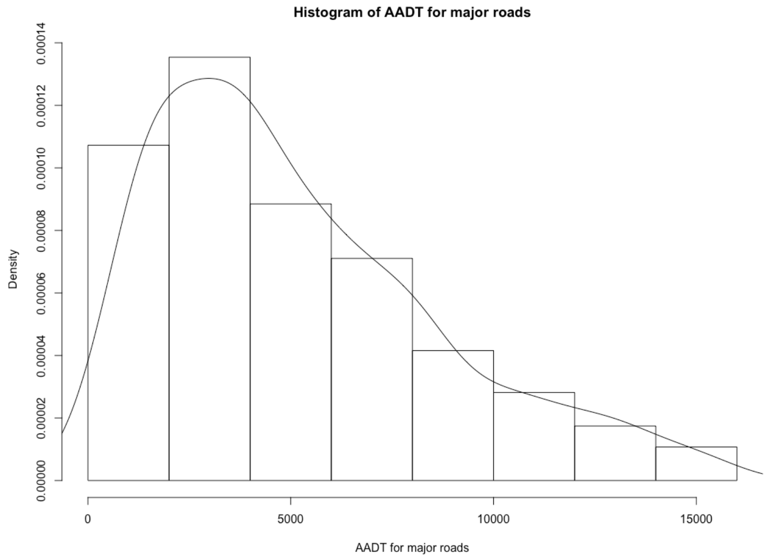

Figure 3.1.1 Histogram of AADT for major roads

Using the moment estimated method: $\alpha' = \frac{\bar{X}^2}{s^2} = 2.18$, $\beta' = \frac{s^2}{\bar{X}} = 2369$, we can fit the QQ-plot for AADT for major roads, Gamma distribution fits the data for AADT for major roads well.

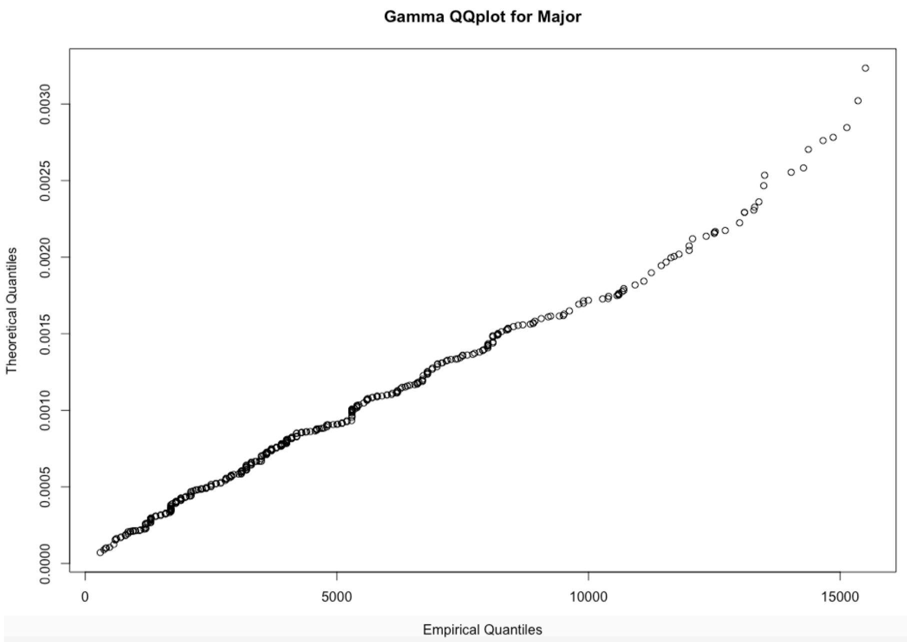

Figure 3.1.2 Gamma QQ-plot for AADT for major roads

Similarly, we can construct a histogram for AADT for minor roads and calculate the mean and variance: $\overline{X} = 1585.191, S^2 = 1429238$. Since the variance is much more larger than mean, we construct a gamma distribution for AADT for minr roads.

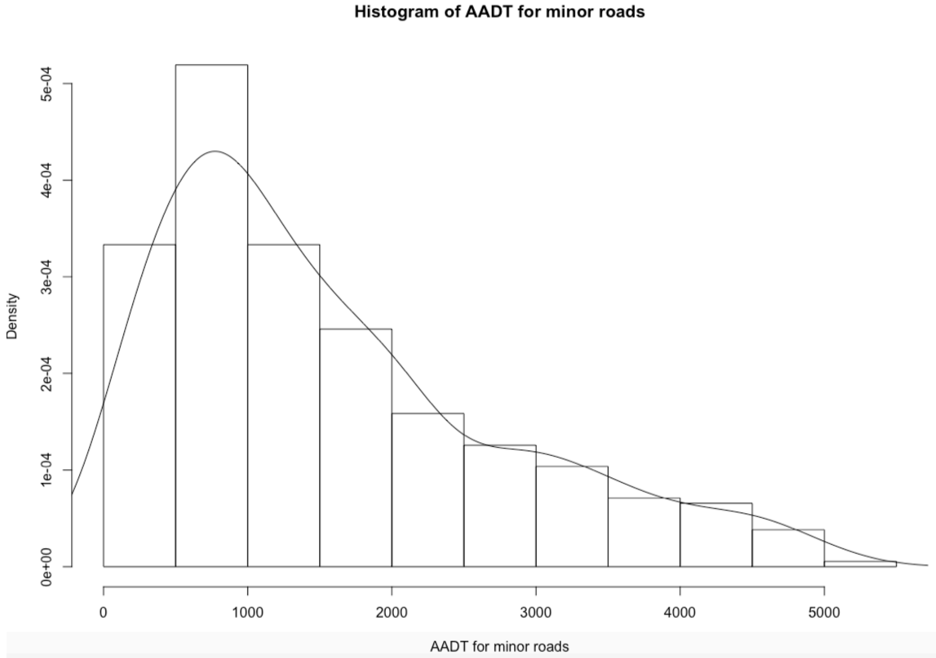

Figure 3.1.3 Histogram of AADT for minor roads

Using the moment estimated method: $\alpha' = \frac{\bar{X}^2}{s^2} = 1.78\,, \beta' = \frac{s^2}{\bar{X}} = 901.7$, we can fit the QQ-plot for AADT for minor roads, Gamma distribution fits the data for AADT for minor roads well.

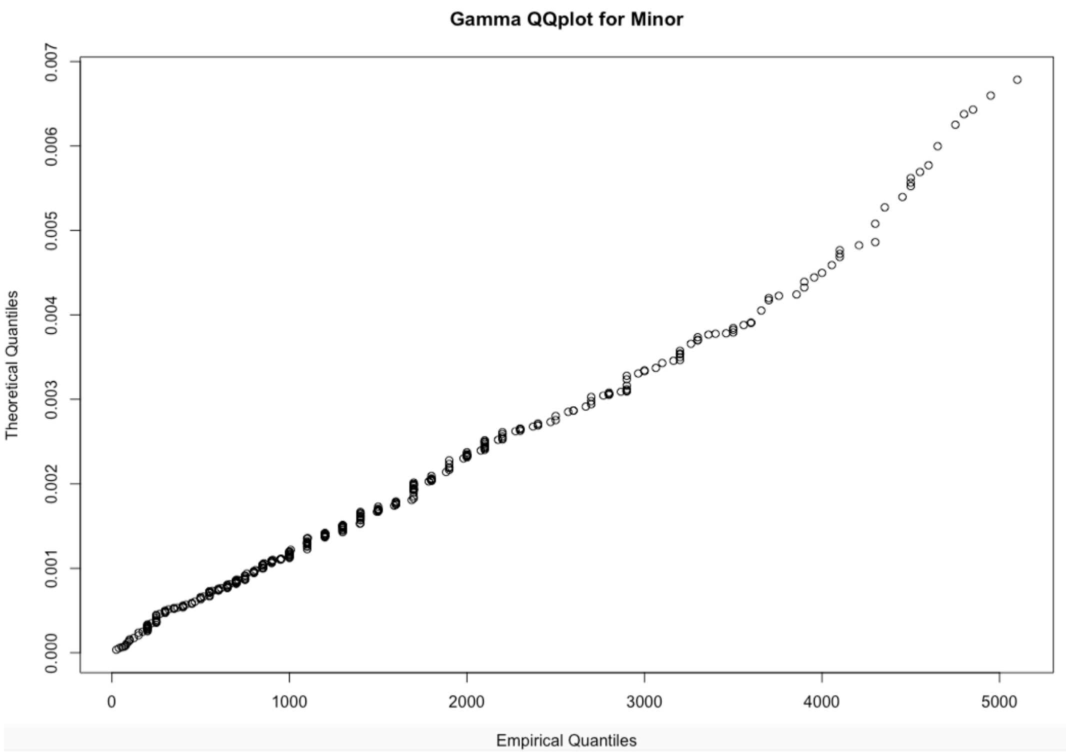

Figure 3.1.4 Gamma QQ-plot for AADT for minor roads

II. Comparation among different intersections

We can draw the boxplots of whole crashes for three kinds of intersections. It is clear that the whole crashes for 4SG(four-way signalized) is significantly different from the other two categories. So, we need to make a comparation of 3ST and 4ST later.

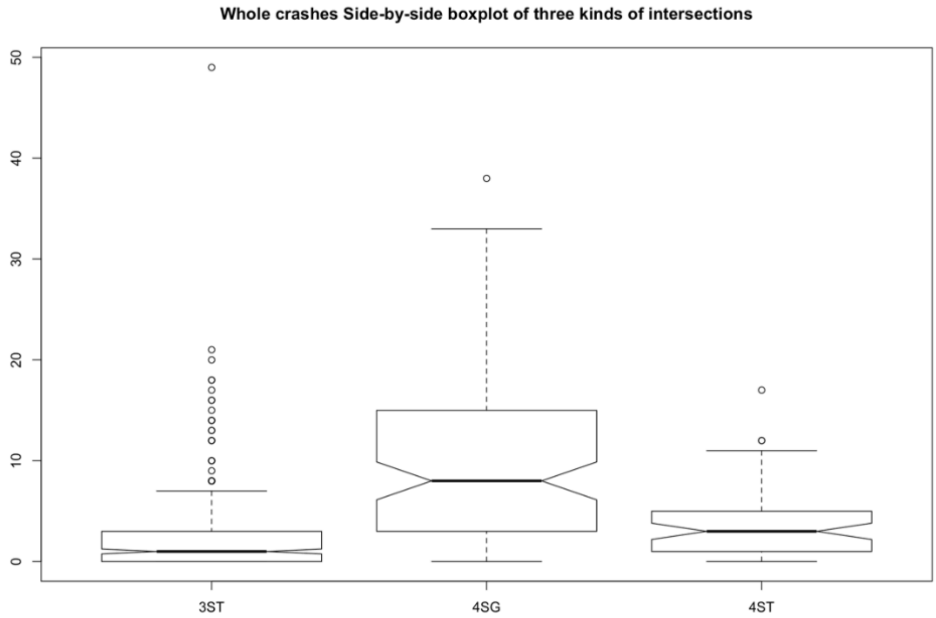

Figure 3.1.5 Side-by-side boxplots of whole crashes for three kinds of intersections

Similarly, we can construct side-by-side boxplots of AADT for major roads for three intersections. It is clear that the whole crashes for 4SG(four-way signalized) is significantly different from the other two categories. So, we need to make a comparation of 3ST and 4ST later.

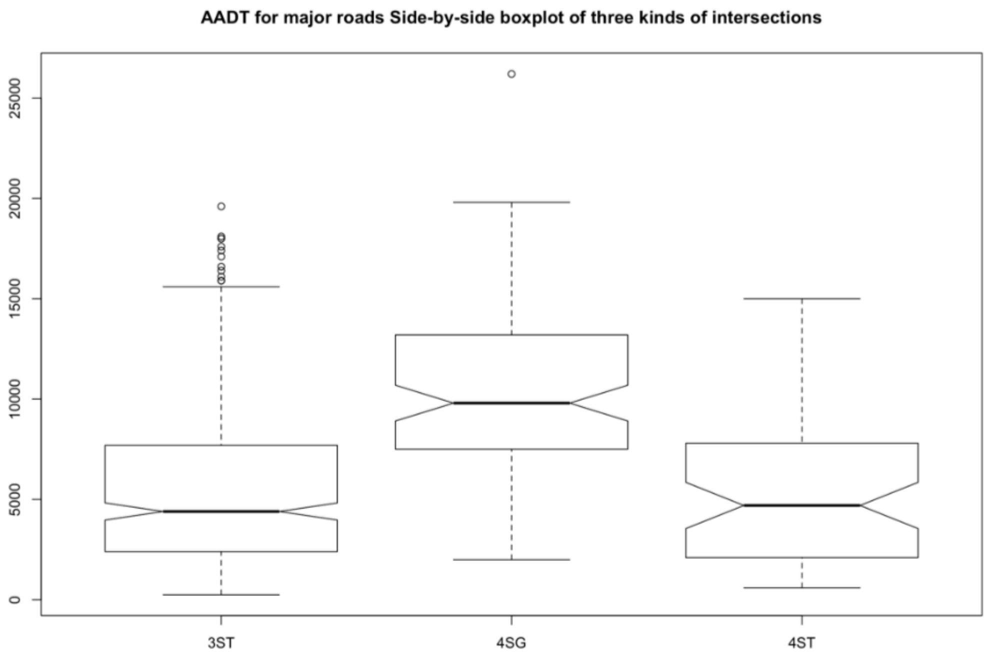

Figure 3.1.6 Side-by-side boxplots of AADT for major roads for three kinds of intersections

We use Jackknife procedure to test the means of AADT for major roads in 3ST and 4ST and the means of AADT for the whole crashes in 3ST and 4ST.
According to the R:

```
> cat("Q=",Q,"\n")
Q= 2.194318
```

Figure 3.1.7 output of Jackknife for whole crashes

According to the figure above, we reject the null hypothesis and consider that the whole crashes in 3ST and 4ST are different.

```
> cat("Q=",Q,"\n")
Q= -0.469956
```

Figure 3.1.7 output of Jackknife for major

According to the figure above, we cannot reject the null hypothesis and consider that the mean of the mean of AADT for major roads in 3ST and 4ST are different.

If we use the bootstrap algorithm to obtain the results, we can get a 95% confidence interval for the difference between the means of two groups:

```
> mdiff[1000*0.025];mdiff[1000*0.975]
[1] -751.963
[1] 1142.121
```

Figure 3.1.8 output of bootstrap for major

From the figure, we know that 95% confidence interval for the difference between the means of two groups is [-751.963,1142.121]. The CI includes 0, so we conclude that the there is no difference of means for AADT for major roads between 3ST and 4ST groups.
If we use the non-parametric methods to test the differences of median of AADT for major roads between 3ST and 4ST, we cannot reject the null hypothesis and conclude there is no difference of medians for two groups:

```
> wilcox.test(ST3_original$AADT_MAJOR,ST4_original$AADT_MAJOR,alternative = "two.sided")

	Wilcoxon rank sum test with continuity correction

data:  ST3_original$AADT_MAJOR and ST4_original$AADT_MAJOR
W = 11774, p-value = 0.9731
alternative hypothesis: true location shift is not equal to 0
```

Figure 3.1.9 output of Wilcoxon rank sum test for major

Similarly, when we test the differences of median of whole crashes between 3ST and 4ST, we reject the null hypothesis and conclude there is difference of medians for two groups:

```
> wilcox.test(ST3_original$PDO+ST3_original$KA+ST3_original$BC,
+             ST4_original$PDO+ST4_original$KA+ST4_original$BC,alternative="two.sided")

        Wilcoxon rank sum test with continuity correction

data:  ST3_original$PDO + ST3_original$KA + ST3_original$BC and ST4_original$PDO + ST4_original$KA + ST4_original$BC
W = 8793.5, p-value = 0.001151
alternative hypothesis: true location shift is not equal to 0
```

Figure 3.1.10 output of Wilcoxon rank sum test for crashes

4. Discussion of Results and Concluding Remarks

We make two tables for the conclusion in the Section II:

Factors	influence on whole crashes
major	significant influence
minor	no significant influence
Lighting	no significant influence
Approach of Left-turn	no significant influence
Approach of Right-turn	no significant influence
Skew Angle	no significant influence

Table 4.1 Factors' influence on whole crashes

Factors	relationship with major	relationship with minor
major	/	positive correlation
minor	positive correlation	/
Lighting	positive correlation	positive correlation
Approach of Left-turn	positive correlation	positive correlation
Approach of Right-turn	positive correlation	no correlation
Skew Angle	no correlation	no correlation

Table 4.2 Factors' relationship with major and

From these two tables, we can know that only major itself has a significant influence on whole crashes. Besides, major is positively related to minor, Lighting, Approach of Left-turn, Approach of Right-turn. This make sense:

When major roads have more traffics, minor roads will have more traffics too.

The more traffic on major roads means the more importance of the intersections. So, there is more likely a lighting on the intersections.

Similarly do the Approach of Left-turn and Approach of Right-turn.

Skew angle always depends on the location so it is reasonable to has no correlation with major.

Therefore, the highway's safety is mainly corresponding to its Annual Average Daily Traffic for major roads. We can also use ANOVA table to check the conclusion:

```
> summary(model)
                                Df Sum Sq Mean Sq F value Pr(>F)    
major_original                   1   1291  1290.6  85.725 <2e-16 ***
minor_original                   1     19    19.1   1.271  0.260    
ST3_original$LIGHTING            1      6     6.4   0.427  0.514    
ST3_original$APPROACH_LEFTTURN   1     17    17.0   1.127  0.289    
ST3_original$APPROACH_RIGHTTURN  1      9     9.5   0.628  0.429    
ST3_original$SKEW_ANGLE          1      1     0.7   0.046  0.830    
Residuals                      378   5691    15.1                   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 4.1 ANOVA table for whole model

From the figure 4.1, we find it comes to the same conclusion as we mentioned before. Besides, from the figure 2.7.1, we know that some intersection of factors will also influence the highway safety. So for 3ST intersections, if we can not reduce the AADT for major roads, it will be more safer to have a approach of left-turn.
What's more, from Section III, 3ST intersection and 4ST intersection has the same mean and median of AADT for major roads at 5% significance while they have different mean and median of whole crashes at 5% significance. This means for different kinds of highway intersections, their safety may depend on different factors with different degrees.

5. Appendix

R code for this project:

```
setwd("/Users/mahaoxi/Desktop/5505/project")
ST3_original<-read.csv("3ST.csv",header = TRUE, sep = ",")
major_original<-ST3_original$AADT_MAJOR
minor_original<-ST3_original$AADT_MINOR

#########make a whole comparsion fisrt
crashes<-ST3_original$PDO+ST3_original$KA+ST3_original$BC
pairs(~ST3_original$AADT_MAJOR+ST3_original$AADT_MINOR+ST3_original$LIGHTING+ST3_original$APPROACH_LEFTTURN+
        ST3_original$APPROACH_RIGHTTURN+ST3_original$SKEW_ANGLE+crashes,
      data=ST3_original,main="Scatter matrix for all factors")

#########analyze angle itself:
#delete the outliers of angle
boxplot(ST3_original$SKEW_ANGLE,main="Boxplot of skew angle");box4<-boxplot(ST3_original$SKEW_ANGLE,plot=FALSE)
range(box4$out);range(ST3_original$SKEW_ANGLE)
ST3<-ST3_original[which(ST3_original$SKEW_ANGLE<70),]
angle<-ST3$SKEW_ANGLE

#construct the histogram of the angle
hist(angle,freq = FALSE)
lines(density(angle))

#judge the distribution of the angle
shapiro.test(angle)#not normal
gam<-rgamma(380,1.296,15.36)
data_quantile<-quantile(angle,ppoints(380))
gam_quantile<-quantile(gam,ppoints(380))
plot(data_quantile,gam_quantile,xlab="Empirical Quantiles",ylab="Theoretical Quantiles",
     main="Gamma QQplot for angle")

#########analyze major and other quantity factors:
#delete major's outliers:
boxplot(major_original,main="Boxplot of AADT for major roads ");box1<-boxplot(major_original,plot=FALSE)
range(box1$out);range(major_original)
ST3<-ST3_original[which(250<=ST3_original$AADT_MAJOR & ST3_original$AADT_MAJOR<15900),]


#classify the major data:
major<-ST3$AADT_MAJOR
quantile(major)
range(major)
major_lowest<-major[which(250<major&major<2400)];major_low<-major[which(2400<major&major<4200)]
major_high<-major[which(4200<major&major<7100)];major_highest<-major[which(7100<major&major<15600)]

####analyze relationship between major and lighting:
#bar chart for major when lighting is 1:
st_light1<-ST3[which(ST3$LIGHTING==1),]
major_light1<-st_light1$AADT_MAJOR
major_lowest_light1<-major_light1[which(250<major_light1&major_light1<2400)]
major_low_light1<-major_light1[which(2400<major_light1&major_light1<4200)]
major_high_light1<-major_light1[which(4200<major_light1&major_light1<7100)]
major_highest_light1<-major_light1[which(7100<major_light1&major_light1<15600)]
x<-c(length(major_lowest_light1),length(major_low_light1),length(major_high_light1),length(major_highest_light1))
names<-c("Lowest","Low","High","Highest")
par(mfrow=c(1,2))
barplot(x,names.arg = names,main="Barplot of major daily traffic with lighting equal to 1")

#bar chart for major when lightening is 0:
st_light0<-ST3[which(ST3$LIGHTING==0),]
major_light0<-st_light0$AADT_MAJOR
major_lowest_light0<-major_light0[which(250<major_light0&major_light0<2400)]
major_low_light0<-major_light0[which(2400<major_light0&major_light0<4200)]
major_high_light0<-major_light0[which(4200<major_light0&major_light0<7100)]
major_highest_light0<-major_light0[which(7100<major_light0&major_light0<15600)]
x<-c(length(major_lowest_light0),length(major_low_light0),length(major_high_light0),length(major_highest_light0))
names<-c("Lowest","Low","High","Highest")
barplot(x,names.arg = names,main="Barplot of major daily traffic with lighting equal to 0")
```

```
#in order to verify the relationship between lighting and crashes:
####delete the outliers of crashes:
k<-ST3_original$PDO+ST3_original$KA+ST3_original$BC
par(mfrow=c(1,1))
boxplot(k,main="Boxplot of sum of the crashes");box2<-boxplot(k,plot=FALSE)
range(box2$out);range(k)
ST3<-ST3_original[which((ST3_original$PDO+ST3_original$KA+ST3_original$BC)<8),]

#lighting=1
st31<-ST3[which(ST3$LIGHTING==1),]
x<-st31$PDO+st31$KA+st31$BC
hist(x,main="Overlaid histogram of whole crashes with lighting equal to 0 and 1",freq=FALSE,
      col = rgb(0,0,1,0.2),xlab="whole crashes",ylim=c(0,0.7))
#lighting=0
st32<-ST3[which(ST3$LIGHTING==0),]
y<-st32$PDO+st32$KA+st32$BC
hist(y,add=TRUE,col = rgb(0,1,1,0.2),freq=FALSE)
legend("topright",c('whole crashes with lighting equal to 0','whole crashes with lighting equal to 1'),
        col=c(rgb(0,0,1,0.2),rgb(0,1,1,0.2)),lwd=8,text.width=3)
boxplot(x,y,notch=TRUE,names=c("Lighting is equal to 1","Lighting is equal to 0")
         ,main="Whole crahses side-by-side notched boxplot for Lighting is equal to 1 and 0")



####analyze relationship between major and left turn:
#bar chart for major when left is 1:
st_left1<-ST3[which(ST3$APPROACH_LEFTTURN==1),]
major_left1<-st_left1$AADT_MAJOR
major_lowest_left1<-major_left1[which(250<major_left1&major_left1<2400)]
major_low_left1<-major_left1[which(2400<major_left1&major_left1<4200)]
major_high_left1<-major_left1[which(4200<major_left1&major_left1<7100)]
major_highest_left1<-major_left1[which(7100<major_left1&major_left1<15600)]
x<-c(length(major_lowest_left1),length(major_low_left1),length(major_high_left1),length(major_highest_left1))
names<-c("Lowest","Low","High","Highest")
par(mfrow=c(1,2))
barplot(x,names.arg = names,main="Barplot of major daily traffic with left turn equal to 1")


#bar chart for major when left is 0:
st_left0<-ST3[which(ST3$APPROACH_LEFTTURN==0),]
major_left0<-st_left0$AADT_MAJOR
major_lowest_left0<-major_left0[which(250<major_left0&major_left0<2400)]
major_low_left0<-major_left0[which(2400<major_left0&major_left0<4200)]
major_high_left0<-major_left0[which(4200<major_left0&major_left0<7100)]
major_highest_left0<-major_left0[which(7100<major_left0&major_left0<15600)]
x<-c(length(major_lowest_left0),length(major_low_left0),length(major_high_left0),length(major_highest_left0))
names<-c("Lowest","Low","High","Highest")
barplot(x,names.arg = names,main="Barplot of major daily traffic with left turn equal to 0")

#in order to verify the relationship between left turn and crashes:
#left=1
st31<-ST3[which(ST3$APPROACH_LEFTTURN==1),]
x<-st31$PDO+st31$KA+st31$BC
par(mfrow=c(1,1))
hist(x,main="Overlaid histogram of crashes with left turn equal to 1 and 0",col=rgb(0,0,1,0.2),freq = FALSE,
      ylim=c(0,1),xlab="whole crashes")
#right=0
st32<-ST3[which(ST3$APPROACH_LEFTTURN==0),]
y<-st32$PDO+st32$KA+st32$BC
hist(y,add=TRUE,col = rgb(0,1,1,0.2),breaks=5,freq = FALSE)
legend("topright",c('whole crashes with left turn equal to 0','whole crashes with left turn equal to 1'),
        col=c(rgb(0,0,1,0.2),rgb(0,1,1,0.2)),lwd=8,text.width=2)
boxplot(x,y,notch=TRUE,names=c("Left-turn is equal to 1","Left-turn is equal to 0")
         ,main="Whole crahses side-by-side notched boxplot for left-turn is equal to 1 and 0")


####analyze relationship between major and right turn:
#bar chart for major when right turn is 1:
st_right1<-ST3[which(ST3$APPROACH_RIGHTTURN==1),]
major_right1<-st_right1$AADT_MAJOR
major_lowest_right1<-major_right1[which(250<major_right1&major_right1<2400)]
major_low_right1<-major_right1[which(2400<major_right1&major_right1<4200)]
major_high_right1<-major_right1[which(4200<major_right1&major_right1<7100)]
major_highest_right1<-major_right1[which(7100<major_right1&major_right1<15600)]
x<-c(length(major_lowest_right1),length(major_low_right1),length(major_high_right1),length(major_highest_right1))
names<-c("Lowest","Low","High","Highest")
par(mfrow=c(1,2))
barplot(x,names.arg = names,main="Barplot of major daily traffic with right turn equal to 1")
```

```
#bar chart for major when right turn is 0:
st_right0<-ST3[which(ST3$APPROACH_RIGHTTURN==0),]
major_right0<-st_right0$AADT_MAJOR
major_lowest_right0<-major_right0[which(250<major_right0&major_right0<2400)]
major_low_right0<-major_right0[which(2400<major_right0&major_right0<4200)]
major_high_right0<-major_right0[which(4200<major_right0&major_right0<7100)]
major_highest_right0<-major_right0[which(7100<major_right0&major_right0<15600)]
x<-c(length(major_lowest_right0),length(major_low_right0),length(major_high_right0),length(major_highest_right0))
names<-c("Lowest","Low","High","Highest")
barplot(x,names.arg = names,main="Barplot of major daily traffic with right turn equal to 0")

#in order to verify the relationship between right turn and crashes:
#right=1
st31<-ST3[which(ST3$APPROACH_RIGHTTURN==1),]
x<-st31$PDO+st31$KA+st31$BC
par(mfrow=c(1,1))
hist(x,main="Overlaid histogram of crashes with right turn equal to 1 and 0",col=rgb(0,0,1,0.2),freq = FALSE,
     ylim=c(0,1.5),breaks=4,xlim=c(0,7),xlab="whole crashes")
#right=0
st32<-ST3[which(ST3$APPROACH_LEFTTURN==0),]
y<-st32$PDO+st32$KA+st32$BC
hist(y,add=TRUE,col = rgb(0,1,1,0.2),breaks=6,freq = FALSE)
legend("topright",c('whole crashes with right turn equal to 0','whole crashes with right turn equal to 1'),
       col=c(rgb(0,0,1,0.2),rgb(0,1,1,0.2)),lwd=8,text.width=2)
boxplot(x,y,notch=TRUE,names=c("Right-turn is equal to 1","Right-turn is equal to 0")
        ,main="Whole crahses side-by-side notched boxplot for Right-turn is equal to 1 and 0")

####analyze relationship between major and angle:
ST31<-ST3[which(ST3$AADT_MAJOR<2400),]
ST32<-ST3[which(2400<ST3$AADT_MAJOR&ST3$AADT_MAJOR<4200),]
ST33<-ST3[which(4200<ST3$AADT_MAJOR&ST3$AADT_MAJOR<7100),]
ST34<-ST3[which(7100<ST3$AADT_MAJOR&ST3$AADT_MAJOR<15600),]
boxplot(ST31$SKEW_ANGLE,ST32$SKEW_ANGLE,ST33$SKEW_ANGLE,ST34$SKEW_ANGLE,names=c("Lowest","Low","High","Highest"),
        ylab="Skewed Angle",notch = TRUE,
        main="Skew angle side-by-side notched boxplot for AADT for major roads")
wilcox.test(ST32$SKEW_ANGLE,ST34$SKEW_ANGLE,alternative = "two.sided")

#judge the relationship between the angle and crashes
angle_original<-ST3_original$SKEW_ANGLE;cras_original<-ST3_original$PDO+ST3_original$BC+ST3_original$KA;
plot(angle_original,cras_original,main="Scatterplot between the skew angle and whole crashes",
     xlab="skew angle",ylab="whole crashes")
hist(angle_original,freq = FALSE,main="Overlais histogram of Skew angle and whole crashes",
     col=rgb(0,0,1,0.2),xlab="Skew angle and whole crashes",ylim=c(0,0.1),xlim=c(0,90))
hist(cras_original,add=TRUE,freq=FALSE,col=rgb(0,1,1,0.2),breaks=5)
legend("topright",c('skew angle','whole crashes'),
       col=c(rgb(0,0,1,0.2),rgb(0,1,1,0.2)),lwd=8,text.width=10)

#########analyze minor and other quantity factors:
#delete minor's outliers:
par(mfrow=c(1,1))
boxplot(minor_original,main="Boxplot of AADT for minor roads");box3<-boxplot(minor_original,plot=FALSE)
range(box3$out);range(minor_original)
ST3<-ST3_original[which(ST3_original$AADT_MINOR<5300),]

#classify the minor data:
minor<-ST3$AADT_MINOR
quantile(minor)
range(minor)
minor_lowest<-minor[which(20<minor&minor<700)];minor_low<-minor[which(700<minor&minor<1300)]
minor_high<-minor[which(1300<minor&minor<2200)];minor_highest<-minor[which(2200<minor&minor<5200)]

####analyze relationship between minor and lighting:
#bar chart for minor when lighting is 1:
st_light1<-ST3[which(ST3$LIGHTING==1),]
minor_light1<-st_light1$AADT_MINOR
minor_lowest_light1<-minor_light1[which(20<minor_light1&minor_light1<700)]
minor_low_light1<-minor_light1[which(700<minor_light1&minor_light1<1300)]
minor_high_light1<-minor_light1[which(1300<minor_light1&minor_light1<2200)]
minor_highest_light1<-minor_light1[which(2200<minor_light1&minor_light1<5200)]
x<-c(length(minor_lowest_light1),length(minor_low_light1),length(minor_high_light1),length(minor_highest_light1))
names<-c("Lowest","Low","High","Highest")
par(mfrow=c(1,2))
barplot(x,names.arg = names,main="Barplot of minor daily traffic with lighting equal to 1")
```

```r
#bar chart for minor when lightening is 0:
st_ligth0<-ST3[which(ST3$LIGHTING==0),]
minor_light0<-st_light0$AADT_MINOR
minor_lowest_light0<-minor_light0[which(250<minor_light0&minor_light0<2400)]
minor_low_light0<-minor_light0[which(2400<minor_light0&minor_light0<4200)]
minor_high_light0<-minor_light0[which(4200<minor_light0&minor_light0<7100)]
minor_highest_light0<-minor_light0[which(7100<minor_light0&minor_light0<15600)]
x<-c(length(minor_lowest_light0),length(minor_low_light0),length(minor_high_light0),length(minor_highest_light0)
names<-c("Lowest","Low","High","Highest")
barplot(x,names.arg = names,main="Barplot of minor daily traffic with lighting equal to 0")

####analyze relationship between minor and left turn:
#bar chart for minor when left is 1:
st_left1<-ST3[which(ST3$APPROACH_LEFTTURN==1),]
minor_left1<-st_left1$AADT_MINOR
minor_lowest_left1<-minor_left1[which(20<minor_left1&minor_left1<700)]
minor_low_left1<-minor_left1[which(700<minor_left1&minor_left1<1300)]
minor_high_left1<-minor_left1[which(1300<minor_left1&minor_left1<2200)]
minor_highest_left1<-minor_left1[which(2200<minor_left1&minor_left1<5200)]
x<-c(length(minor_lowest_left1),length(minor_low_left1),length(minor_high_left1),length(minor_highest_left1))
names<-c("Lowest","Low","High","Highest")
par(mfrow=c(1,2))
barplot(x,names.arg = names,main="Barplot of minor daily traffic with left turn equal to 1")

#bar chart for minor when left is 0:
st_left0<-ST3[which(ST3$APPROACH_LEFTTURN==0),]
minor_left0<-st_left0$AADT_MINOR
minor_lowest_left0<-minor_left0[which(20<minor_left0&minor_left0<700)]
minor_low_left0<-minor_left0[which(700<minor_left0&minor_left0<1300)]
minor_high_left0<-minor_left0[which(1300<minor_left0&minor_left0<2200)]
minor_highest_left0<-minor_left0[which(2200<minor_left0&minor_left0<5200)]
x<-c(length(minor_lowest_left0),length(minor_low_left0),length(minor_high_left0),length(minor_highest_left0))
names<-c("Lowest","Low","High","Highest")
barplot(x,names.arg = names,main="Barplot of minor daily traffic with left turn equal to 0")


####analyze relationship between minor and right turn:
#bar chart for major when right turn is 1:
st_right1<-ST3[which(ST3$APPROACH_RIGHTTURN==1),]
minor_right1<-st_right1$AADT_MINOR
minor_lowest_right1<-minor_right1[which(20<minor_right1&minor_right1<700)]
minor_low_right1<-minor_right1[which(700<minor_right1&minor_right1<1300)]
minor_high_right1<-minor_right1[which(1300<minor_right1&minor_right1<2200)]
minor_highest_right1<-minor_right1[which(2200<minor_right1&minor_right1<5200)]
x<-c(length(minor_lowest_right1),length(minor_low_right1),length(minr_high_right1),length(minor_highest_right1))
names<-c("Lowest","Low","High","Highest")
par(mfrow=c(1,2))
barplot(x,names.arg = names,main="Barplot of minor daily traffic with right turn equal to 1")

#bar chart for minor when right turn is 0:
st_right0<-ST3[which(ST3$APPROACH_RIGHTTURN==0),]
minor_right0<-st_right0$AADT_MINOR
minor_lowest_right0<-minor_right0[which(20<minor_right0&minor_right0<700)]
minor_low_right0<-minor_right0[which(700<minor_right0&minor_right0<1300)]
minor_high_right0<-minor_right0[which(1300<minor_right0&minor_right0<2200)]
minor_highest_right0<-minor_right0[which(2200<minor_right0&minor_right0<5200)]
x<-c(length(minor_lowest_right0),length(minor_low_right0),length(minor_high_right0),length(minor_highest_right0))
names<-c("Lowest","Low","High","Highest")
barplot(x,names.arg = names,main="Barplot of minor daily traffic with right turn equal to 0")

####analyze relationship between minor and angle:
par(mfrow=c(1,1))
ST31<-ST3[which(ST3$AADT_MINOR<700),]
ST32<-ST3[which(700<ST3$AADT_MINOR&ST3$AADT_MINOR<1300),]
ST33<-ST3[which(1300<ST3$AADT_MINOR&ST3$AADT_MINOR<2200),]
ST34<-ST3[which(2200<ST3$AADT_MINOR&ST3$AADT_MINOR<5200),]
boxplot(ST31$SKEW_ANGLE,ST32$SKEW_ANGLE,ST33$SKEW_ANGLE,ST34$SKEW_ANGLE,names=c("Lowest","Low","High","Highest"),
        ylab="Skewed Angle",notch = TRUE,main="Skew angle Side-by-side Boxplots of AADT for minor roads")


#########analyze major and minor:
hist(major_original,col=rgb(0,0,1,0.2),xlab="AADT for major and minor roads",
     main="Histogram of Annual Average Major And Minor Daily Traffic",freq = FALSE,ylim=c(0,0.0004))
hist(minor_original,col=rgb(0,1,1,0.2),add=TRUE,breaks = 4,freq=FALSE)
legend("topright",c('Major','Minor'),
       col=c(rgb(0,0,1,0.2),rgb(0,1,1,0.2)),lwd=8,text.width=4000)
boxplot(major_original,minor_original,main="Side-by-side Boxplots for AADT for major roads and minor roads",
        notch=TRUE,
        names=c("Major Roads","Minor Roads"))
plot(major_original,minor_original,main="Scatterplot for AADT for major roads and minor roads",
     xlab="AADT for major roads",ylab="AADT for minor roads")
```

```r
#########analyze relationship between minor and crashes:
crashes_original<-ST3_original$PDO+ST3_original$KA+ST3_original$BC
par(mfrow=c(1,2))
hist(minor_original,col=rgb(0,0,1,0.2),xlab="AADT for minor roads",
     main="Histogram of AADT for minor roads",freq = FALSE,ylim=c(0,0.0004))
hist(crashes_original,col=rgb(0,1,1,0.2),freq=FALSE,xlab="whole crashes",main="Histogram of whole crashes")

par(mfrow=c(1,1))
ST3<-ST3_original[which(ST3_original$AADT_MINOR<5300),]
ST31<-ST3$PDO+ST3$KA+ST3$BC;boxplot(ST31);box5<-boxplot(ST31,plot=FALSE)
ST33<-ST3[which(ST33$PDO+ST33$KA+ST33$BC<8),]
ST331<-ST33[which(ST33$AADT_MINOR<700),]
ST332<-ST33[which(700<ST33$AADT_MINOR&ST33$AADT_MINOR<1300),]
ST333<-ST33[which(1300<ST33$AADT_MINOR&ST33$AADT_MINOR<2200),]
ST334<-ST33[which(2200<ST33$AADT_MINOR&ST33$AADT_MINOR<5200),]
boxplot(ST331$PDO+ST331$KA+ST331$BC,
        ST332$PDO+ST332$KA+ST332$BC,ST333$PDO+ST333$KA+ST333$BC,
        ST334$PDO+ST334$KA+ST334$BC,names=c("Lowest","Low","High","Highest"),
        ylab="Whole crashes",notch = TRUE,main="Crashes Side-by-side Boxplots for AADT for minor roads")
plot(minor_original,crashes_original)

#########analyze relationship between major and crashes:
par(mfrow=c(1,2))
crashes_original<-ST3_original$PDO+ST3_original$KA+ST3_original$BC
hist(major_original,col=rgb(0,0,1,0.2),xlab="AADT for major roads",
     main="Histogram of AADT for major roads",freq = FALSE)
hist(crashes_original,col=rgb(0,1,1,0.2),freq=FALSE,xlab="whole crashes",main="Histogram of whole crashes")

par(mfrow=c(1,1))
ST3<-ST3_original[which(ST3_original$AADT_MAJOR<15900),]
ST31<-ST3$PDO+ST3$KA+ST3$BC;boxplot(ST31);box5<-boxplot(ST31,plot=FALSE)
ST33<-ST3[which(ST33$PDO+ST33$KA+ST33$BC<8),]
ST331<-ST33[which(ST33$AADT_MAJOR<2400),]
ST332<-ST33[which(2400<ST33$AADT_MAJOR&ST33$AADT_MAJOR<4200),]
ST333<-ST33[which(4200<ST33$AADT_MAJOR&ST33$AADT_MAJOR<7100),]
ST334<-ST33[which(7100<ST33$AADT_MAJOR&ST33$AADT_MAJOR<15600),]
boxplot(ST331$PDO+ST331$KA+ST331$BC,
        ST332$PDO+ST332$KA+ST332$BC,ST333$PDO+ST333$KA+ST333$BC,
        ST334$PDO+ST334$KA+ST334$BC,names=c("Lowest","Low","High","Highest"),
        ylab="whole crashes",main="Crashes Side-by-side Boxplots for AADT for major roads")
plot(major_original,crashes_original,xlab="AADT for major roads",
     ylab="whole crashes",main="Scatterplot for AADT for major roads and whole crashes")

#########analyze major itself:
#construct the histogram of the major
par(mfrow=c(1,1))
hist(major,freq = FALSE,main="Histogram of AADT for major roads",xlab="AADT for major roads",ylab="Density")
lines(density(major))

#judge the distribution of the major
shapiro.test(major)#not normal
library(vcd)
gam<-rgamma(373,2.18,2369)
data_quantile<-quantile(major,ppoints(385))
gam_quantile<-quantile(gam,ppoints(385))
plot(data_quantile,gam_quantile,xlab="Empirical Quantiles",ylab="Theoretical Quantiles",
     main="Gamma QQplot for Major")

#########analyze minor itself:
#construct the histogram of the minor
hist(minor,freq = FALSE,main="Histogram of AADT for minor roads",xlab="AADT for minor roads",ylab="Density")
lines(density(minor))

#judge the distribution of the minor
shapiro.test(minor)#not normal
gam<-rgamma(366,1.76,901.7)
data_quantile<-quantile(minor,ppoints(366))
gam_quantile<-quantile(gam,ppoints(366))
plot(data_quantile,gam_quantile,xlab="Empirical Quantiles",ylab="Theoretical Quantiles",
     main="Gamma QQplot for Minor")

#########analyze whole crashes itself:
par(mfrow=c(1,1))
hist(ST3_original$PDO+ST3_original$KA+ST3_original$BC,
     freq = FALSE,main="Histogram of whole crashes",xlab="Whole crashes",ylab="Density",ylim = c(0,0.3))
lines(density(ST3_original$PDO+ST3_original$KA+ST3_original$BC))
```

```r
#########analyze the whole model
Z0000<-ST3_original[which(ST3_original$AADT_MAJOR<4200&ST3_original$AADT_MINOR<1300
                          &ST3_original$LIGHTING==0&ST3_original
                          $APPROACH_LEFTTURN==0&ST3_original$APPROACH_RIGHTTURN==0),]
Z0001<-ST3_original[which(ST3_original$AADT_MAJOR<4200&ST3_original$AADT_MINOR<1300
                          &ST3_original$LIGHTING==0&ST3_original
                          $APPROACH_LEFTTURN==0&ST3_original$APPROACH_RIGHTTURN==1),]
Z0010<-ST3_original[which(ST3_original$AADT_MAJOR<4200&ST3_original$AADT_MINOR<1300
                          &ST3_original$LIGHTING==0&ST3_original
                          $APPROACH_LEFTTURN==1&ST3_original$APPROACH_RIGHTTURN==0),]
Z0011<-ST3_original[which(ST3_original$AADT_MAJOR<4200&ST3_original$AADT_MINOR<1300
                          &ST3_original$LIGHTING==0&ST3_original
                          $APPROACH_LEFTTURN==1&ST3_original$APPROACH_RIGHTTURN==1),]
Z0100<-ST3_original[which(ST3_original$AADT_MAJOR<4200&ST3_original$AADT_MINOR<1300
                          &ST3_original$LIGHTING==1&ST3_original
                          $APPROACH_LEFTTURN==0&ST3_original$APPROACH_RIGHTTURN==0),]
Z0101<-ST3_original[which(ST3_original$AADT_MAJOR<4200&ST3_original$AADT_MINOR<1300
                          &ST3_original$LIGHTING==1&ST3_original
                          $APPROACH_LEFTTURN==0&ST3_original$APPROACH_RIGHTTURN==1),]
Z0110<-ST3_original[which(ST3_original$AADT_MAJOR<4200&ST3_original$AADT_MINOR<1300
                          &ST3_original$LIGHTING==1&ST3_original
                          $APPROACH_LEFTTURN==1&ST3_original$APPROACH_RIGHTTURN==0),]
Z0111<-ST3_original[which(ST3_original$AADT_MAJOR<4200&ST3_original$AADT_MINOR<1300
                          &ST3_original$LIGHTING==1&ST3_original
                          $APPROACH_LEFTTURN==1&ST3_original$APPROACH_RIGHTTURN==1),]
Z1000<-ST3_original[which(4200<ST3_original$AADT_MAJOR&ST3_original$AADT_MAJOR<15600
                          &1300<ST3_original$AADT_MINOR&ST3_original$AADT_MINOR<5200
                          &ST3_original$LIGHTING==0&ST3_original
                          $APPROACH_LEFTTURN==0&ST3_original$APPROACH_RIGHTTURN==0),]
Z1001<-ST3_original[which(4200<ST3_original$AADT_MAJOR&ST3_original$AADT_MAJOR<15600
                          &1300<ST3_original$AADT_MINOR&ST3_original$AADT_MINOR<5200
                          &ST3_original$LIGHTING==0&ST3_original
                          $APPROACH_LEFTTURN==0&ST3_original$APPROACH_RIGHTTURN==1),]
Z1010<-ST3_original[which(4200<ST3_original$AADT_MAJOR&ST3_original$AADT_MAJOR<15600
                          &1300<ST3_original$AADT_MINOR&ST3_original$AADT_MINOR<5200
                          &ST3_original$LIGHTING==0&ST3_original
                          $APPROACH_LEFTTURN==1&ST3_original$APPROACH_RIGHTTURN==0),]
```


$APPROACH_LEFTTURN==1&ST3_original$APPROACH_RIGHTTURN==0),]


```r
Z1011<-ST3_original[which(4200<ST3_original$AADT_MAJOR&ST3_original$AADT_MAJOR<15600
                          &1300<ST3_original$AADT_MINOR&ST3_original$AADT_MINOR<5200
                          &ST3_original$LIGHTING==0&ST3_original
                          $APPROACH_LEFTTURN==1&ST3_original$APPROACH_RIGHTTURN==1),]
Z1100<-ST3_original[which(4200<ST3_original$AADT_MAJOR&ST3_original$AADT_MAJOR<15600
                          &1300<ST3_original$AADT_MINOR&ST3_original$AADT_MINOR<5200
                          &ST3_original$LIGHTING==1&ST3_original
                          $APPROACH_LEFTTURN==0&ST3_original$APPROACH_RIGHTTURN==0),]
Z1101<-ST3_original[which(4200<ST3_original$AADT_MAJOR&ST3_original$AADT_MAJOR<15600
                          &1300<ST3_original$AADT_MINOR&ST3_original$AADT_MINOR<5200
                          &ST3_original$LIGHTING==1&ST3_original
                          $APPROACH_LEFTTURN==0&ST3_original$APPROACH_RIGHTTURN==1),]
Z1110<-ST3_original[which(4200<ST3_original$AADT_MAJOR&ST3_original$AADT_MAJOR<15600
                          &1300<ST3_original$AADT_MINOR&ST3_original$AADT_MINOR<5200
                          &ST3_original$LIGHTING==1&ST3_original
                          $APPROACH_LEFTTURN==1&ST3_original$APPROACH_RIGHTTURN==0),]
Z1111<-ST3_original[which(4200<ST3_original$AADT_MAJOR&ST3_original$AADT_MAJOR<15600
                          &1300<ST3_original$AADT_MINOR&ST3_original$AADT_MINOR<5200
                          &ST3_original$LIGHTING==1&ST3_original
                          $APPROACH_LEFTTURN==1&ST3_original$APPROACH_RIGHTTURN==1),]
```

```
z0000<-Z0000$PDO+Z0000$KA+Z0000$BC
z0001<-Z0001$PDO+Z0001$KA+Z0001$BC
z0010<-Z0010$PDO+Z0010$KA+Z0010$BC
z0011<-Z0011$PDO+Z0011$KA+Z0011$BC
z0100<-Z0100$PDO+Z0100$KA+Z0100$BC
z0101<-Z0101$PDO+Z0101$KA+Z0101$BC
z0110<-Z0110$PDO+Z0110$KA+Z0110$BC
z0111<-Z0111$PDO+Z0111$KA+Z0111$BC
z1000<-Z1000$PDO+Z1000$KA+Z1000$BC
z1001<-Z1001$PDO+Z1001$KA+Z1001$BC
z1010<-Z1010$PDO+Z1010$KA+Z1010$BC
z1011<-Z1011$PDO+Z1011$KA+Z1011$BC
z1100<-Z1100$PDO+Z1100$KA+Z1100$BC
z1101<-Z1101$PDO+Z1101$KA+Z1101$BC
z1110<-Z1110$PDO+Z1110$KA+Z1110$BC
z1111<-Z1111$PDO+Z1111$KA+Z1111$BC
boxplot(z0000,z0001,z0010,z0011,z0100,z0101,z0110,z0111,z1000,z1001,z1010,z1011,z1100,z1101,z1110,z1111,
        names=c("0000","0001","0010","0011","0100","0101","0110","0111","1000","1001","1010","1011","1100","1101","1110","1111"),
        main="Whole crashes side-by-side boxplot of factors")

#########analyze the difference of other two sample
SG4_original<-read.csv("4SG.csv",header = TRUE, sep = ",")
ST4_original<-read.csv("4ST.csv",header = TRUE, sep = ",")

#preparation
boxplot(ST3_original$PDO+ST3_original$KA+ST3_original$BC,SG4_original$PDO+SG4_original$KA+SG4_original$BC,
        ST4_original$PDO+ST4_original$KA+ST4_original$BC,notch = TRUE,
        main="Whole crashes Side-by-side boxplot of three kinds of intersections",
        names=c("3ST","4SG","4ST"))
boxplot(ST3_original$AADT_MAJOR,SG4_original$AADT_MAJOR,ST4_original$AADT_MAJOR,notch = TRUE,
        main="AADT for major roads Side-by-side boxplot of three kinds of intersections",
        names=c("3ST","4SG","4ST"))
```


```
#########analyze the difference of other two sample
SG4_original<-read.csv("4SG.csv",header = TRUE, sep = ",")
ST4_original<-read.csv("4ST.csv",header = TRUE, sep = ",")

#preparation
boxplot(ST3_original$PDO+ST3_original$KA+ST3_original$BC,SG4_original$PDO+SG4_original$KA+SG4_original$BC,
        ST4_original$PDO+ST4_original$KA+ST4_original$BC,notch = TRUE,
        main="Whole crashes Side-by-side boxplot of three kinds of intersections",
        names=c("3ST","4SG","4ST"))
boxplot(ST3_original$AADT_MAJOR,SG4_original$AADT_MAJOR,ST4_original$AADT_MAJOR,notch = TRUE,
        main="AADT for major roads Side-by-side boxplot of three kinds of intersections",
        names=c("3ST","4SG","4ST"))
```


```
#compare the mean of AADT of major road and whole crashes in 3ST and 4ST
#AADT of major road
#Jackknife
X<-ST3_original$AADT_MAJOR;Y<-ST4_original$AADT_MAJOR
estsX<-matrix(0,nrow=(1+length(X)),ncol=2)
n<-length(X)
A<-rep(0,n)
cat(mean(X),log(var(X)),"\n")
estsX[1,]<-c(mean(X),log(var(X)))
for (i in 1:n) {
        cat(mean(X[-i]),log(var(X[-i])),"\n")
        estsX[1+i,]<-c(mean(X[-i]),log(var(X[-i])))
        A[i]<-n*estsX[1,2] - (n-1)*estsX[i+1,2]
}
V1<-var(A)/n


estsY<-matrix(0,nrow=(1+length(Y)),ncol=2)
m <-length(Y)
B<-rep(0,m)
cat(mean(Y),log(var(Y)),"\n")
estsY[1,]<-c(mean(Y),log(var(Y)))
for (i in 1:m) {
        #cat(mean(Y[-i]),log(var(Y[-i])),"\n")
        estsY[1+i,]<-c(mean(Y[-i]),log(var(Y[-i])))
        B[i]<-m*estsY[1,2] - (m-1)*estsY[i+1,2]
}
V2<-var(B)/m

Q<-(mean(B) - mean(A))/(sqrt(V1+V2))
cat("Q=",Q,"\n")
```

```r
#bootstrap
boot<-function(n1,n2,B){
        est<-matrix(0,B,2)
        colnames(est)<-c("mdiff", "vratio")              # column names of matrix est
        out<-vector("list",2)
        names(out)<-c("bootstats","actual")
        actual<-c(mean(X)-mean(Y),var(X)/var(Y))
        names(actual)<-c("mdiff", "vratio")
        for (i in 1:B){                                  # boot loop
                idx1<-sample(1:n1,n1,replace=TRUE)            # select indices for bootstrap sample for Y1
                idx2<-sample(1:n2,n2,replace=TRUE)            # select indices for bootstrap sample for Y2
                Ystar1<-X[idx1]
                Ystar2<-Y[idx2]
                est[i,1]<-mean(Ystar1)-mean(Ystar2)             # difference of two means
                est[i,2]<-var(Ystar1)/var(Ystar2)               # ratio of two variances
        }
        out$bootstats<-est
        out$actual<-actual
        return(out)
}

out<-boot(length(X),length(Y),1000)
mdiff<-sort(out$bootstats[,1])
mdiff[1000*0.025];mdiff[1000*0.975]


#whole crashes
#Jackknife
X<-ST3_original$PDO+ST3_original$KA+ST3_original$BC;Y<-ST4_original$PDO+ST4_original$KA+ST4_original$BC
estsX<-matrix(0,nrow=(1+length(X)),ncol=2)
n<-length(X)
A<-rep(0,n)
cat(mean(X),log(var(X)),"\n")
estsX[1,]<-c(mean(X),log(var(X)))
for (i in 1:n) {
        cat(mean(X[-i]),log(var(X[-i])),"\n")
        estsX[1+i,]<-c(mean(X[-i]),log(var(X[-i])))
        A[i]<-n*estsX[1,2] - (n-1)*estsX[i+1,2]
}
V1<-var(A)/n

estsY<-matrix(0,nrow=(1+length(Y)),ncol=2)
m <-length(Y)
B<-rep(0,m)
cat(mean(Y),log(var(Y)),"\n")
estsY[1,]<-c(mean(Y),log(var(Y)))
for (i in 1:m) {
        #cat(mean(Y[-i]),log(var(Y[-i])),"\n")
        estsY[1+i,]<-c(mean(Y[-i]),log(var(Y[-i])))
        B[i]<-m*estsY[1,2] - (m-1)*estsY[i+1,2]
}
V2<-var(B)/m

Q<-(mean(B) - mean(A))/(sqrt(V1+V2))
cat("Q=",Q,"\n")

#compare the median of AADT of major road and whole crashes in 3ST and 4ST
wilcox.test(ST3_original$AADT_MAJOR,ST4_original$AADT_MAJOR,alternative = "two.sided")
wilcox.test(ST3_original$PDO+ST3_original$KA+ST3_original$BC,
           ST4_original$PDO+ST4_original$KA+ST4_original$BC,alternative="two.sided")
```