Student Name: Xiaohong Chen

Date: 05/10/2020

ATE-252 Final Assignment

Correlation between illumination condition in dark area and car crashes in Pennsylvania

In my previous thought, I think illumination condition in dark area will effect the road safety directly. A no light illumination condition will even cause a fatal accident. However, some people claims that other factors will also cause a fatal accident even if there is a good illumination condition. My purpose is to user R to determine if there is a correlation between illumination condition in dark area and car crashes.

Question:

Will there be a correlation illumination condition in dark area and car crashes? If not, what other factors will affect car crashes?

Reference:

Mr. Mark Egge's tutorial

Thanks for Mr. Mark Egge's and Eric' tutorial, providing me with a bunch of useful library and code demo for R. This project is created by referencing Mr. Mark Egge's tutorial (mark@eateggs.com)

Datasets:

1. Traffic Data(Road segment data) ---- PennDOT's RMS Traffic Open Data ('traffic.shp')

2. Crash Data ---- PennDOT's Public Crash Databases ('CRASH_2018_Statewide.csv ')

Planning:

1. Import traffic data and crash data

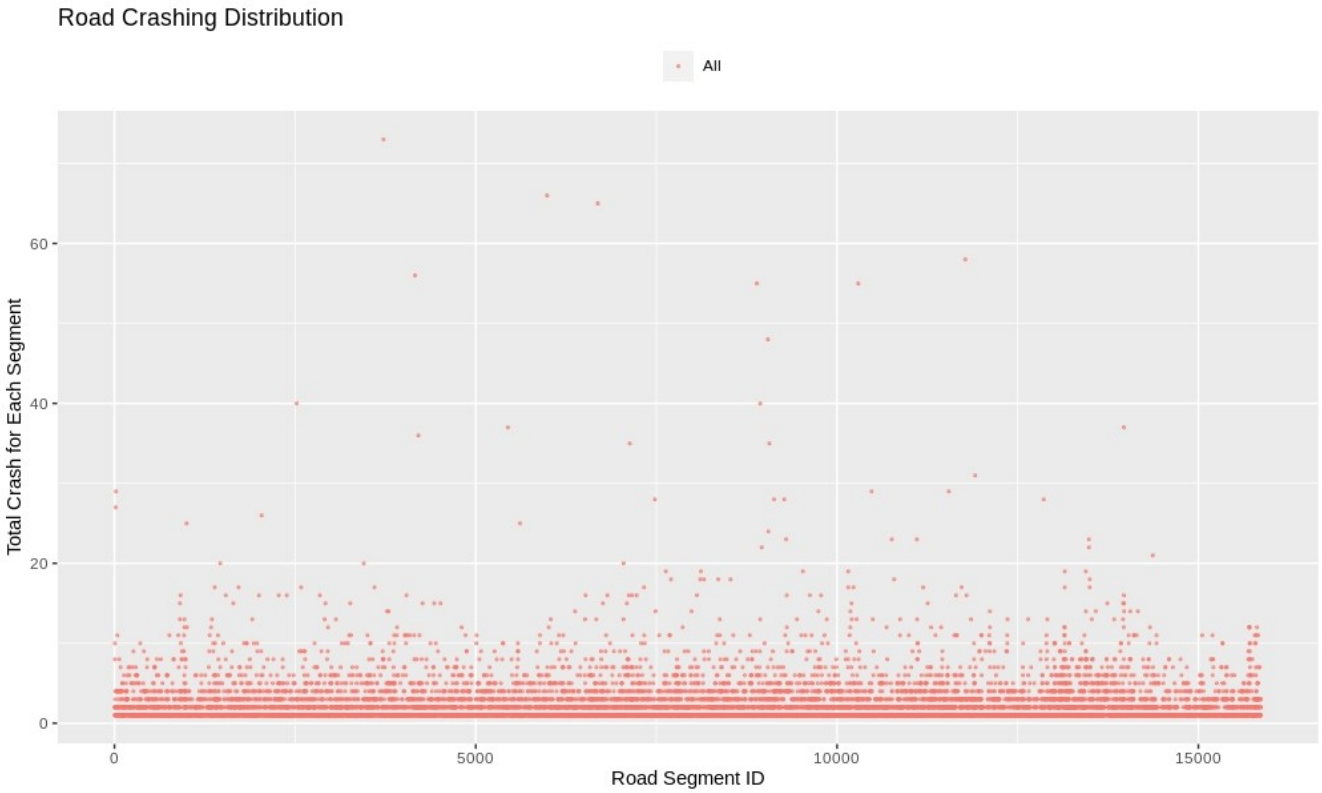
2. Filter illumination condition in dark area from crash data and filter to only crashes rows with "2 - Dark - No Street Lights" or "3 - Dark - Street Lights"

3. Join crashes data with illumination condition into road segment data

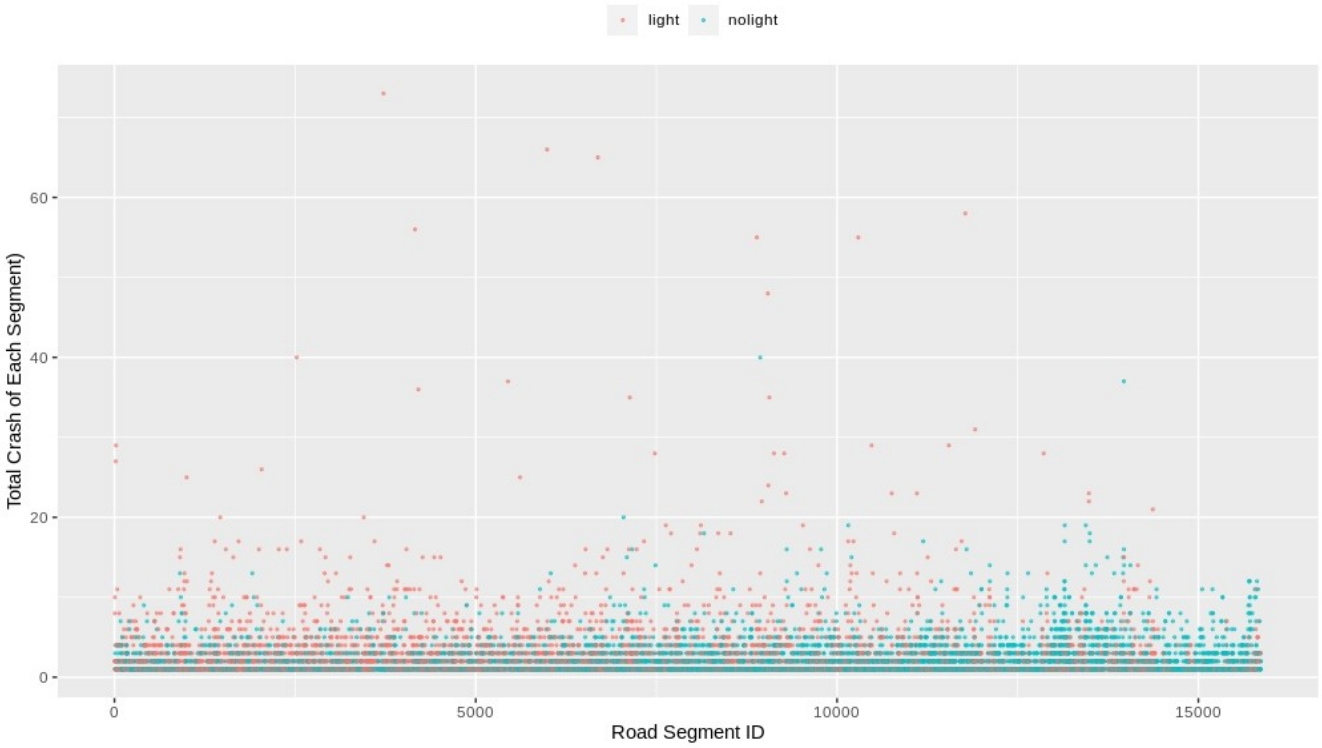
4. Using ggplot() function to get chart comparing "No Street Lights" data with " Street Lights" data

5. Find other factors that will affect car crashes

Crashing Distribution Chart for All Road Segments:



Crashing distribution with no street lights V.S. Crashing distribution with street lights in Dark area.



Hypothesis Test:

Null Hypothesis : The factor of not having street lights condition in dark area and the factor of having street lights condition have the same influence on car crashes

Alternative Hypothesis : They have the difference on car crashes

T-test:

data: light_segment_crashes and nolight_segment_crashes

t = 9.0321, df = 7394, **p-value < 2.2e-16**

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

0.4388736 0.6821818

sample estimates:

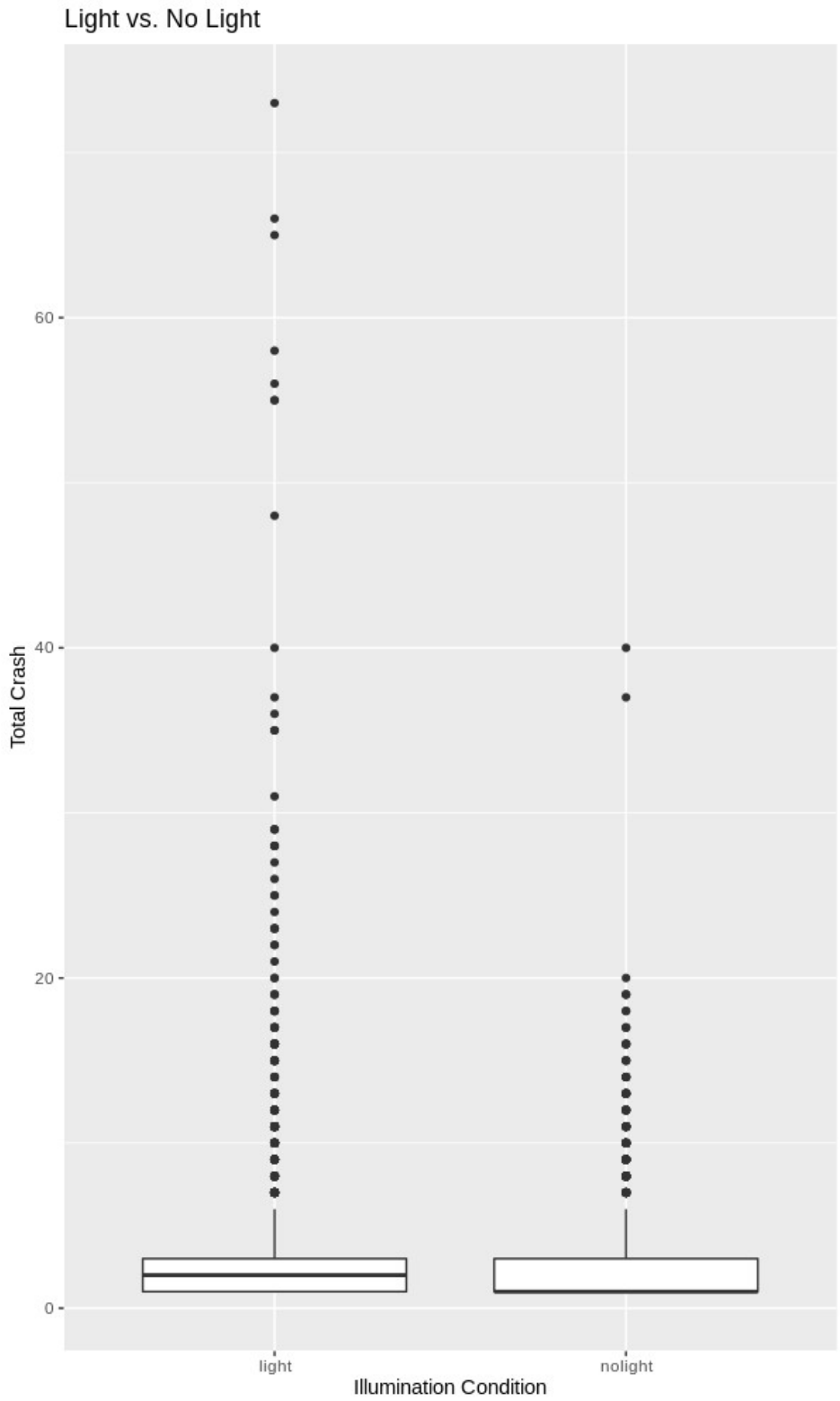
mean of x mean of y

2.781345 2.220817

Interpretation:

By doing a t. test, we can find the p-value less than 0.05. A significant difference can be detected, which means we have a very low chance to incorrectly reject the Null Hypothesis. So we can conclude The factor of not having street lights condition in dark area and the factor of having street lights condition have difference influence on car crashes

Whisker Plot

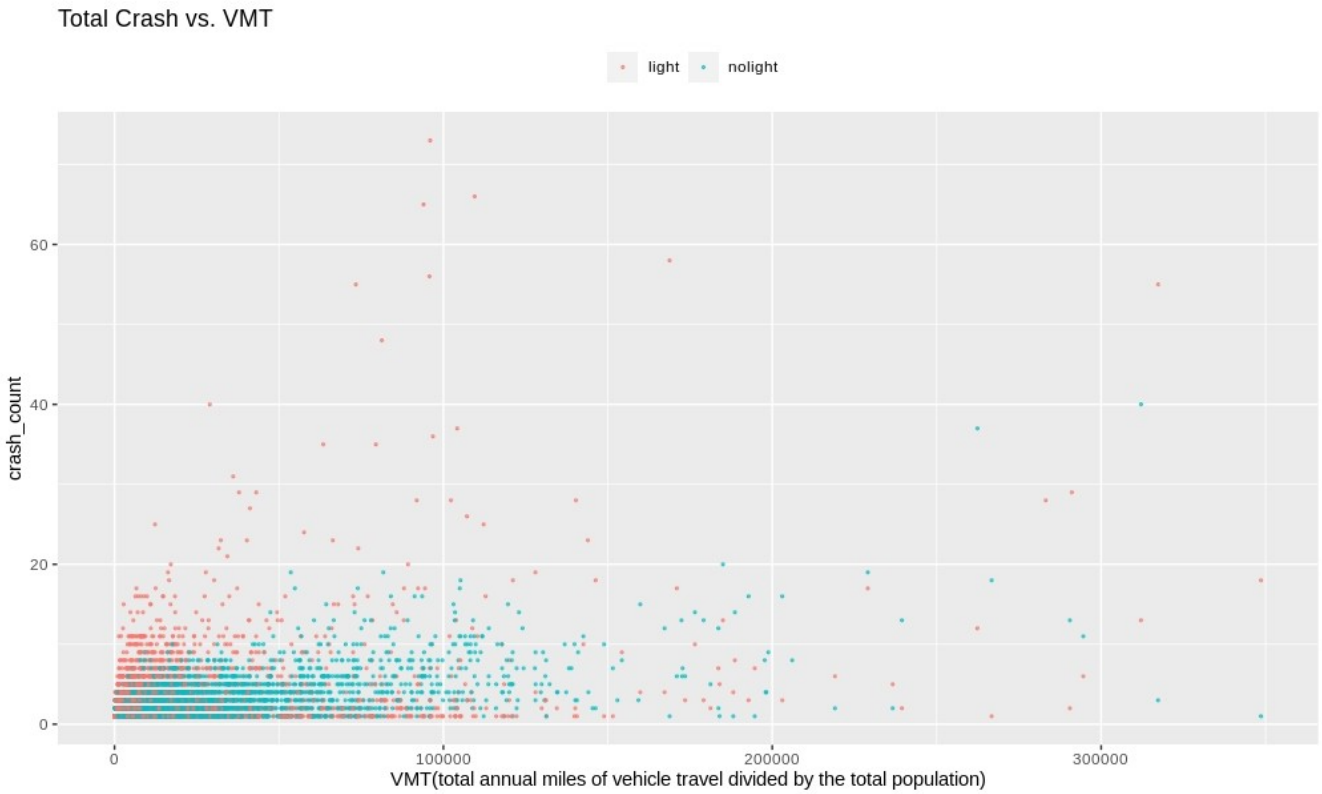


Conclusion-1:

By doing a whisker plot, we get the distributions of two samples and conclude that there will be more likely to have a car crash in a dark area with a good illumination condition. So there might be some factors that affect car crashing.

Correlation between car crashes and VMT

(VMT: total annual miles of vehicle travel divided by the total population)



Conclusion-2:

By comparing the distribution of car accident data and VMT data, we found that VMT also has an impact on car accidents. The range from 0 to 100,000 VMT, with the increase of vmt, road accidents increase especially under a good lighting condition. In the range from 100,000 to 300,000 VMT, the increase in car accidents is not very obvious. I guess the road with VMT of 100,000 to 300,000 may be highways. Therefore, as VMT increases, car accidents decrease.