Q3

Dataset URL (Click Below):

Kitchener Car Collision(https://data.waterloo.ca/datasets/KitchenerGIS::traffic-collisions/data?geometry=-80.703%2C43.374%2C-80.175%2C43.461)

Load and Display Data

```
traffic <- read.csv('Traffic_Collisions.csv',header=T)
head(traffic)
```

```
##                              LOCATION              LIGHT OBJECTID ACCIDENTNUM
## 1 CEDAR ST N @ KING ST E / CEDAR ST S         01-Daylight        1   13-118637
## 2          GLASGOW ST @ BELMONT AVE W         01-Daylight        2   13-019811
## 3 CEDAR ST N @ KING ST E / CEDAR ST S         01-Daylight        3   12-089689
## 4           BELMONT AVE W @ ARGYLE ST         01-Daylight        4   12-127500
## 5             STRASBURG RD @ HURON RD 08-Dark, artificial        5   13-012564
## 6          DRIFTWOOD DR @ HIGHVIEW DR                            6   12-044704
##   ACCIDENTDATE        ACCIDENT_LOCATION                IMPACT_LOCATION
## 1     20130525  02-Intersection related                   02-Thru lane
## 2     20130124 04-At/near private drive                   02-Thru lane
## 3     20120418      01-Non intersection              09-Right shoulder
## 4     20120601       03-At intersection         01-Within intersection
## 5     20130115       03-At intersection         01-Within intersection
## 6     20120225       03-At intersection 11-Not on roadway - right side
##   ENVIRONMENT_CONDITION COLLISION_TYPE TRAFFICCOLLISIONSID
## 1              01-Clear   INTERSECTION                  NA
## 2              01-Clear   INTERSECTION                  NA
## 3              01-Clear   INTERSECTION                  NA
## 4               02-Rain   INTERSECTION                  NA
## 5              01-Clear   INTERSECTION                  NA
## 6               03-Snow   INTERSECTION                  NA
```

Description and Population Justification

This is a dataset taken from City of Waterloo Open Data website. The dataset records all intersection and mid-block traffic collisions happened in the city of Kitchener from 2013 to present, which means that new collisions will continue to be recorded into the table in the future. The whole dataset represents a population of $\mathcal{N} = 12201$ units.

It is a population because:

1. There are limited numbers of collisions in the dataset. We are able to count how many rows in the table. Therefore, the data is finite.

2. All collisions are recorded in the given time interval. They are not selected based on an unique characteristic or feature. Therefore, the data is complete.

3. The table contains comprehensive info of every single collision, not limited to just a few variates.

Therefore, the dataset represents a population.

Unit and Variates

The unit in the dataset is: a collision happened in the city of Kitchener

Two Variates:

1. Location: The variate describes the specific location in the city where the collision happened.

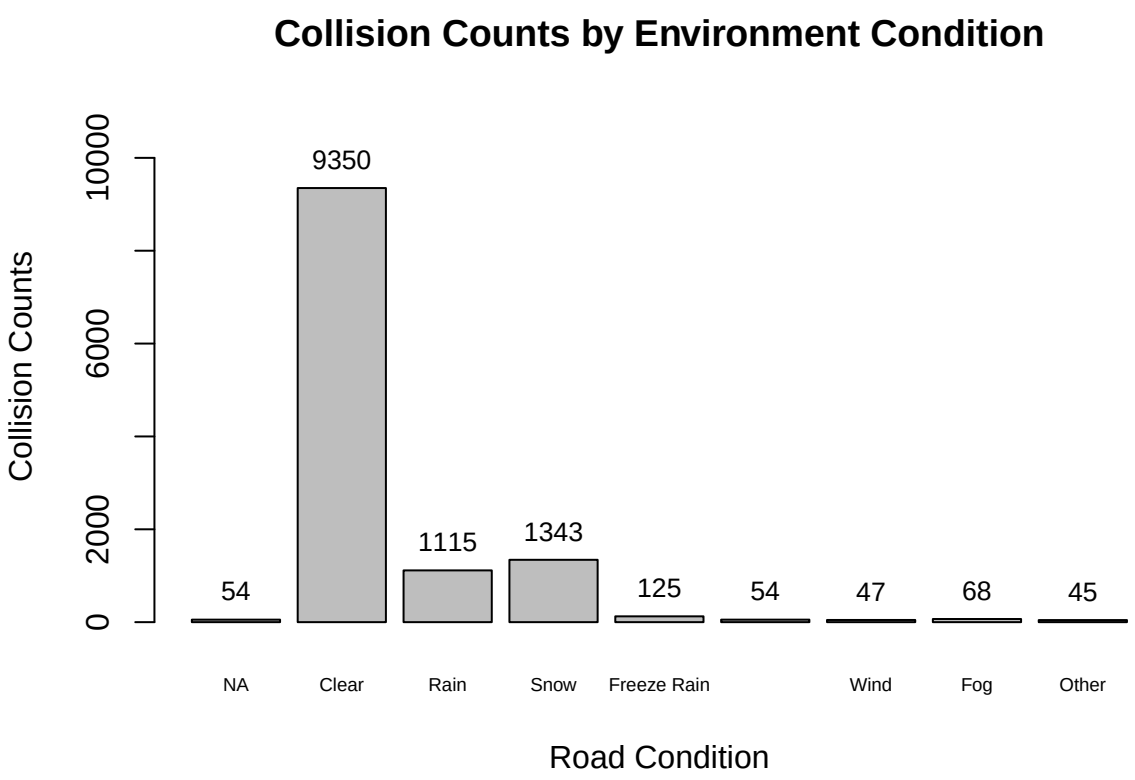
2. Environment Condition: The variate describes the weather condition on the road when the collision happened.

A single Graph

```
levels(as.factor(traffic$ENVIRONMENT_CONDITION))
```

```
## [1] ""                          "01-Clear"
## [3] "02-Rain"                   "03-Snow"
## [5] "04-Freezing Rain"          "05-Drifting Snow"
## [7] "06-Strong wind"            "07-Fog, mist, smoke, dust"
## [9] "99-Other"
```

```
counts <- table(traffic$ENVIRONMENT_CONDITION)
freqs <- as.numeric(counts)
graph <- barplot(counts, main = "Collision Counts by Environment Condition",
                 xlab = "Road Condition",
                 ylab = "Collision Counts",
                 cex.names = 0.6,
                 ylim = c(0,11000),
                 names.arg = c("NA", "Clear", "Rain", "Snow", "Freeze Rain",
                               "Drift Snow", "Wind", "Fog", "Other"))
text(x = graph, y = freqs, label = freqs, pos = 3, cex = 0.8, col = "black")
```



Interesting Feature

From the graph, we can see an interesting feature that the maximum count of collision by environment condition is the clear day. Based on the common sense, we normally think that terrible weathers will cause a higher risk of collision. In those days, bad weathers can do a damage to the car. The car will easily lose control. However, from the data, most of the collision happened in clear days. The reason behind this situation is that drivers tend to lose focus when they can see the road clearly. When they drive in bad weathers, they are more aware of the potential danger and drive more cautiously.