

Collision Analysis for New York City

NYU GA-DS 1007 Term Project

Prof. Gregory Watson

Team: Nan Wu(nw1045), Lingshan Gao (lg2755), Shucheng Yan(sy1253)

Introduction

This program provides a series of user-interactive statistical analysis for the automobile collisions that have happened in New York City during the period of January 2015 to September 2016. The program will provide an analysis on trends for auto vehicle collisions that have happened in the NYC area from January 2015 to October 2016. As you navigate through the program, you will find information related to collisions, such as cause analysis, number of people injured, and heat maps that help you visually examine which area are more prone to car accidents. At certain level, you may be able to compare collision situations across geographic areas in NYC.

Environment Requirement

The program is written using Python 3 syntax. Below is a list of packages and the corresponding versions that are used in the program.

- Numpy 1.11.2
- Pandas 0.19.0
- Matplotlib 1.5.3
- Basemap 1.0.7

An Anaconda platform, although is not required, is highly recommend. Please note that Basemap, which is a Matplotlib toolkit, is not a standard package that is distributed through Anaconda. To install Basemap, please do the following in the Terminal (assuming under a Mac environment).

```
$ conda install basemap
```

Program Structure

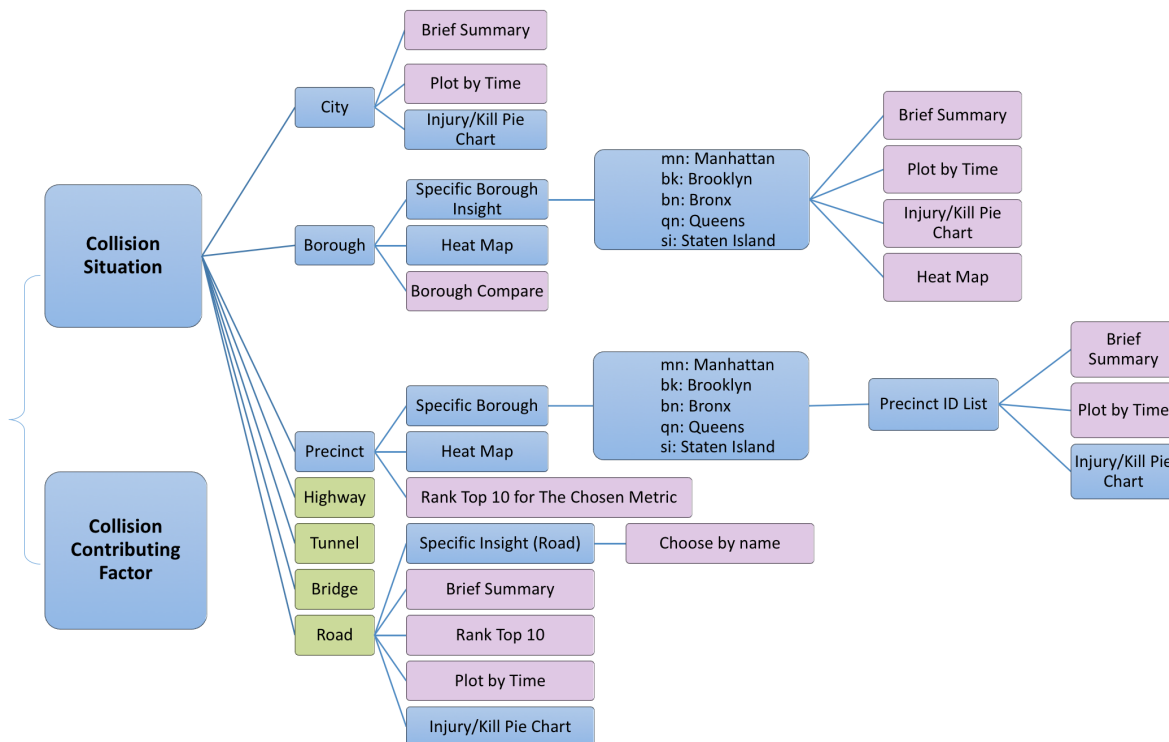
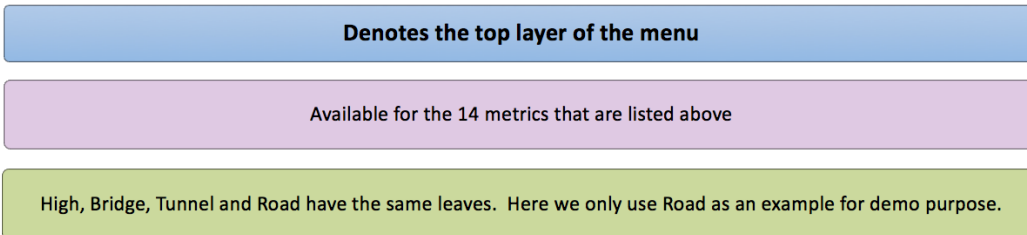
The program has two layers in the top level of the menu. First layer is related to collision situations. You may either view the collision situations, such as number of people injured, during the selection timeframe. Below is the list of 14 metrics that you may view:

1. Number of collisions
2. Collision injury count: number of collisions that are caused injuries
3. Collision killed count: number of collisions that are caused death
4. Person injured: number of people that are injured by collisions
5. Person killed: number of people that are killed by collisions
6. Motorists injured: number of motorist that are injured
7. Motorists killed: number of motorist that are killed
8. Passenger injured: number of passenger that are injured
9. Passenger killed: number of passenger that are killed
10. Cyclists injured: number of cyclists that are injured
11. Cyclists killed: number of cyclists that are killed

12. Pedestrian injured: number of pedestrians that are injured
13. Pedestrian killed: number of pedestrians that are killed
14. Total injuries or death: total number of injuries or death

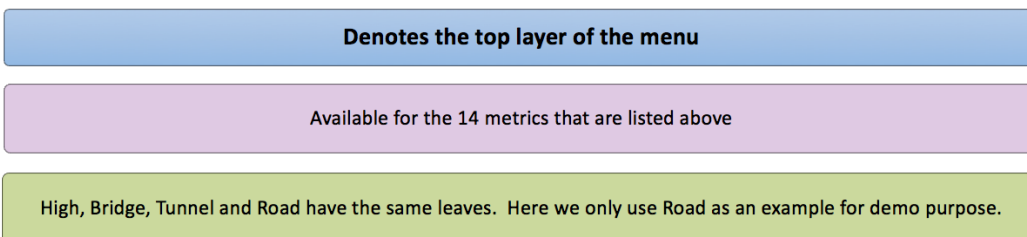
Below is a visualized program structure:

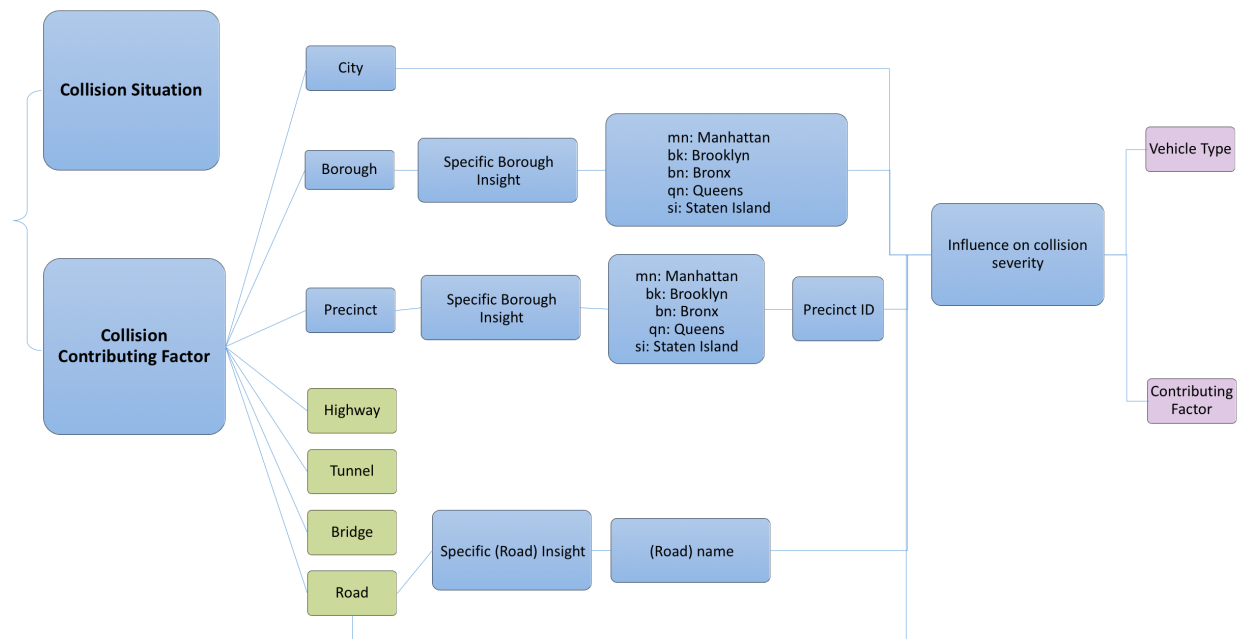
Legends:



The second part of the top level menu is Collision Contributing Factor, where you may view the factors that cause the collision. Below is a visualized structure for this part:

Legends:





How to Run the Program

- Open the terminal, change the directory where the repo <NYC_Collision_Analysis> is saved.
- Change directory to the level of where main program is saved and run the program by doing the command below
 - \$ cd NYC_Collision_Analysis/Final_Project_1007
 - \$ python NYCcollision.py
- At the beginning of the program, input a start and end month to extract the data. The data is only available from January 2015 to October 2016. The input format is <YYYYMM>, for example '201501'. Please keep in mind that the data may take a while to load. We suggest selecting one to two months for testing purpose.
 - Beginning date:
 - \$ 201501
 - Ending date:
 - \$ 201501
- You may then follow the program instruction to look at summary statistics of collision at different levels
- The program includes the following basic commands that you may use at any level of the program:
 - **Exit**: exit the program entirely
 - **Back**: to go back to the previous menu. This works in all levels except for when you are inputting dates
 - **Help**: to access the help menu.

All the charts that are requested by the user, including bar charts, histogram, heat maps, are saved in the “results” folder under “Final_Project_1001.” When the user initiate the program, the program will automatically create a folder in the name of <BeginningDate_EndDate> under “results” directory.

Example Walkthrough:

1. Do the following command, and you will see an introduction to the program.

```
This program will provide an analysis on trends for auto vehicle collisions that have happened in the NYC area from January 2015 to October 2016.
```

```
As you navigate through the program, you will find information related to collisions, such as cause analysis, number of people injured, and heatmaps that help you visually examine which areas are more prone to car accidents.
```

```
At certain level, you may be able to compare collision situations across geographic areas in NYC.
```

```
DATA SOURCE:http://www.nyc.gov/html/nypd/html/traffic\_reports/motor\_vehicle\_collision\_data.shtml
```

```
You may explore the city by area, or by roadway types. There are several sub-menus in each of the two categories:
```

```
Explore by area: City; Borough; Precinct
```

```
Explore by roadway type: Highway; Tunnel; Bridge; Road
```

```
There are available methods under the specific perspective.
```

```
Exit by input : Exit
```

```
Back with: Back
```

```
Now you start exploring. Enjoy!
```

```
>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
```

```
You can set a period for data loading and structure building.
```

```
ALL following analysis will be based on this period.
```

```
Results will be save in a folder named by this period under results folder.
```

```
Example:201501_201601
```

```
Longest Time Interval is 201501-201610.
```

```
Please input a year and a month to begin. The date is inclusive.
```

```
Take note that the data is only available from January 2015.
```

```
Input format: YYYYMM, Example: 201501
```

2. Use January 2015 data as an example, input the start time as 201501 and end time as 201501. Then the program will load the collision data for this month by borough. This may take a while so please be patient. While loading the data, you may see something like “2015 1 si” which stands for “January 2015 Staten Island.” The program will let you know that data has been loaded successfully upon completion.

As soon as the dates are given, a folder will be created under “results” folder to save the charts, if any. In this case, the folder will be named “201501 201501”

```
>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
2015 1 qn ...
Loading CSV File about collisions in Intersection...
Loading CSV File about collisions in HighTunBri...
Building Borough Information...
Building Road Information...
Building Highway Information...
Building Bridge Information...
Building Tunnel Information...
>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
2015 1 si ...
Loading CSV File about collisions in Intersection...
Loading CSV File about collisions in HighTunBri...
Building Borough Information...
Building Road Information...
Building Highway Information...
Building Bridge Information...
Building Tunnel Information...
>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
Success...
There are three types of information:
    1-CollisionSituation
    2-CollisionContributingFactors_Vehicle
You can choose one by input the number before the ty
```

3. We are curious about how many auto vehicle collisions has happened January 2015, so I will choose “1-CollisionSituation” first by typing “1”.

```
There are three types of information:
  1-CollisionSituation
  2-CollisionContributingFactors_Vehicle
You can choose one by input the number before the type above:1
Available Perspectives:
  1-City
  2-Borough
  3-Precinct
  4-Highway
  5-Tunnel
  6-Bridge
  7-Road
Please input the number before the perspective you want to explore:
```

4. Let's look at overall city level first by choosing "1-City." We will three available options at City level.

```
Please input the number before the perspective you want to explore:1
There are methods for this level:
1 Brief Summary
2 Plot by Time
3 InjuryKillPie
Please input the number before the method you want to use:
```

5. I will just take a look of the brief summary by typing “1”. We will see a list of 14 metrics

```
Please input the number before the perspective you want to explore:1
There are methods for this level:
1 Brief Summary
2 Plot by Time
3 InjuryKillPie
Please input the number before the method you want to use:1
Please Choose one Indicator from:
1:Number of Collisions
2:CollisionInjuredCount
3:CollisionKilledCount
4:PersonsInjured
5:PersonsKilled
6:MotoristsInjured
7:MotoristsKilled
8:PassengInjured
9:PassengKilled
10:CyclistsInjured
11:CyclistsKilled
12:PedestrInjured
13:PedestrKilled
14:Injury or Fatal
```

6. I choose "Number of Collisions" by giving the program a "1".

[illegible]

7. From the returned results, we can see that there were a total of 15,473 auto collisions happened in January 2015 in NYC. This information is helpful, but most likely we would like some visual aid to help compare across regions. Now we can input anything to proceed. The system will go back to there most recent menu available.

```

Input anything to Continue:
Please Choose one Indicator from:
1: Number of Collisions
2: CollisionInjuredCount
3: CollisionKilledCount
4: PersonsInjured
5: PersonsKilled
6: MotoristsInjured
7: MotoristsKilled
8: PassengInjured
9: PassengKilled
10: CyclistsInjured
11: CyclistsKilled
12: PedestrInjured
13: PedestrKilled
14: Injury_or_Fatal
Your choice: 1

```

8. Now I will type in “Back” twice to go back to choose a specific Borough. Please note that commands are case-sensitive.

```

Please input the number before the method you want to use:Back
Available Perspectives:
1-City
2-Borough
3-Precinct
4-Highway
5-Tunnel
6-Bridge
7-Road
Please input the number before the perspective you want to explore:2
There are methods for this level:
0 Specific Insight
4 Map
6 Borough Compare
Please input the number before the method you want to use:

```

9. I would like to check out the “Map” function, which is a heat map that plots the specific metric for the city by borough. I will type in “4” to choose the Map function, and then input “1” to let the map plot “Number of Collisions” by borough.

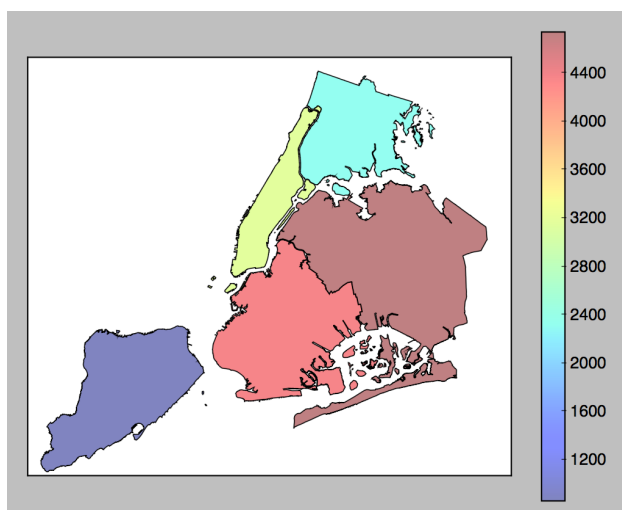
This is what we will see in the Terminal:

```

There are methods for this level:
0 Specific Insight
4 Map
6 Borough Compare
Please input the number before the method you want to use:4
Please Choose one Indicator from:
1: Number of Collisions
2: CollisionInjuredCount
3: CollisionKilledCount
4: PersonsInjured
5: PersonsKilled
6: MotoristsInjured
7: MotoristsKilled
8: PassengInjured
9: PassengKilled
10: CyclistsInjured
11: CyclistsKilled
12: PedestrInjured
13: PedestrKilled
14: Injury_or_Fatal
Your choice: 1
Input anything to Close the Figure and Continue

```

This heat map will pop up:

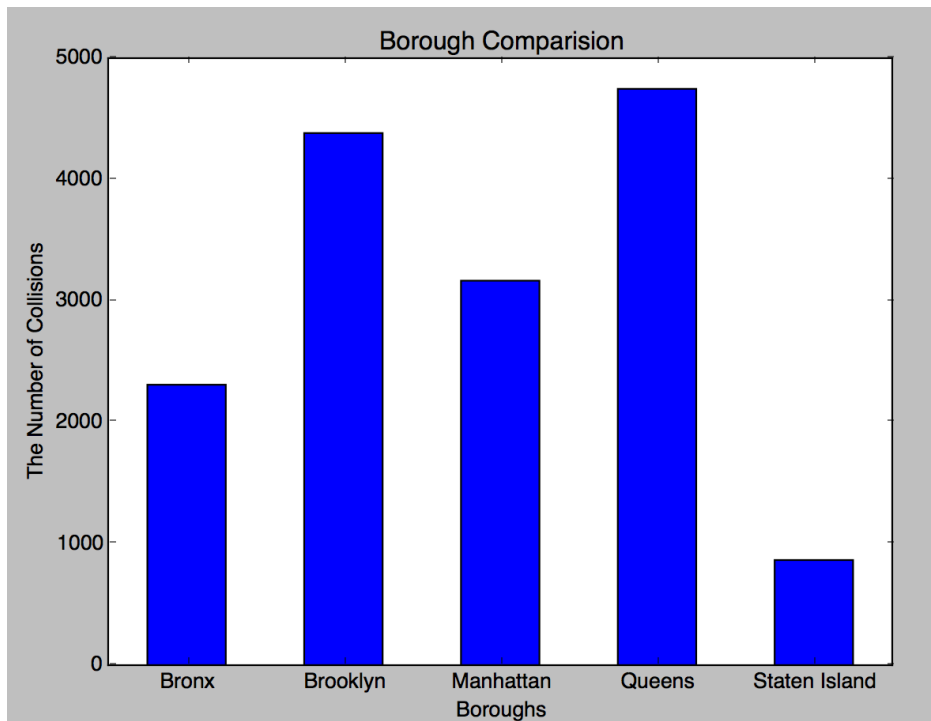


10. As we can tell from the map above, the 15,000 of collisions mostly happened in Queens, and Staten Island has the least number of collisions. Now I can input anything to continue, and the plot will be saved in the corresponding folder.
11. Now instead of a heat map to get a general understanding, I also want to compare number of collisions across boroughs by creating a bar chart. So let's go back to the previous level of menu and select "6 Borough Compare" by typing in 6, and still select the metric "1 Number of Collisions"

This is what we will in the Terminal:

```
Please input the number before the perspective you want to explore:2
There are methods for this level:
0 Specific Insight
4 Map
6 Borough Compare
Please input the number before the method you want to use:6
Please Choose one Indicator from:
1: Number of Collisions
2: CollisionInjuredCount
3: CollisionKilledCount
4: PersonsInjured
5: PersonsKilled
6: MotoristsInjured
7: MotoristsKilled
8: PassengInjured
9: PassengKilled
10: CyclistsInjured
11: CyclistsKilled
12: PedestrInjured
13: PedestrKilled
14: Injury_or_Fatal
Your choice: 1
Input anything to Close the Figure and Continue
```

This bar plot will pop up:



12. As expected, the number of collisions that happened in Queens in January 2015 is five times as many as in Staten Island. This helps has visualize metric ratios across boroughs.
13. Now I am curious what are the contributing factors to these accidents. So I went back to the menu where I can select contributing factors.

```
Please input the number before the perspective you want to explore:Back
There are three types of information:
1-CollisionSituation
2-CollisionContributingFactors_Vehicle
You can choose one by input the number before the type above:
```

14. Select "CollisionContributingFactors_Vehicle" by typing "2", and select "Borough" level by typing "2", eventually, we would like to view the contributing metric to the accidents in Queens which has the most number of collisions in January 2015.

```
2-CollisionContributingFactors_Vehicle
You can choose one by input the number before the type above:2
Available Perspectives:
1-City
2-Borough
3-Precinct
4-Highway
5-Tunnel
6-Bridge
7-Road
Please input the number before the perspective you want to explore
There are methods for this level:
0 Specific Insight
Please input the number before the method you want to use:0
You can choose from:
bx : Bronx
bk : Brooklyn
mn : Manhattan
qn : Queens
si : Staten_Island
Please input the abbreviation (two letters) for the borough:
```

15. To choose Queens, input its abbreviation "qn":

```
Please input the abbreviation (two letters) for the borough:qn
There are methods for this level:
1 Influence On Collision Severity
Please input the number before the method you want to use:
```

16. There is only one available option "Influence on Collision Severity," so input its number "1".

```
Please Choose one Influencer from:
1:VehicleType
2:ContributingFactor
Your choice:
```

17. Now we can either view the vehicle type of these vehicles that had collisions, or we can further look at the contributing factors. Here I am interested in the vehicle type, and choose the metric of number of collisions:

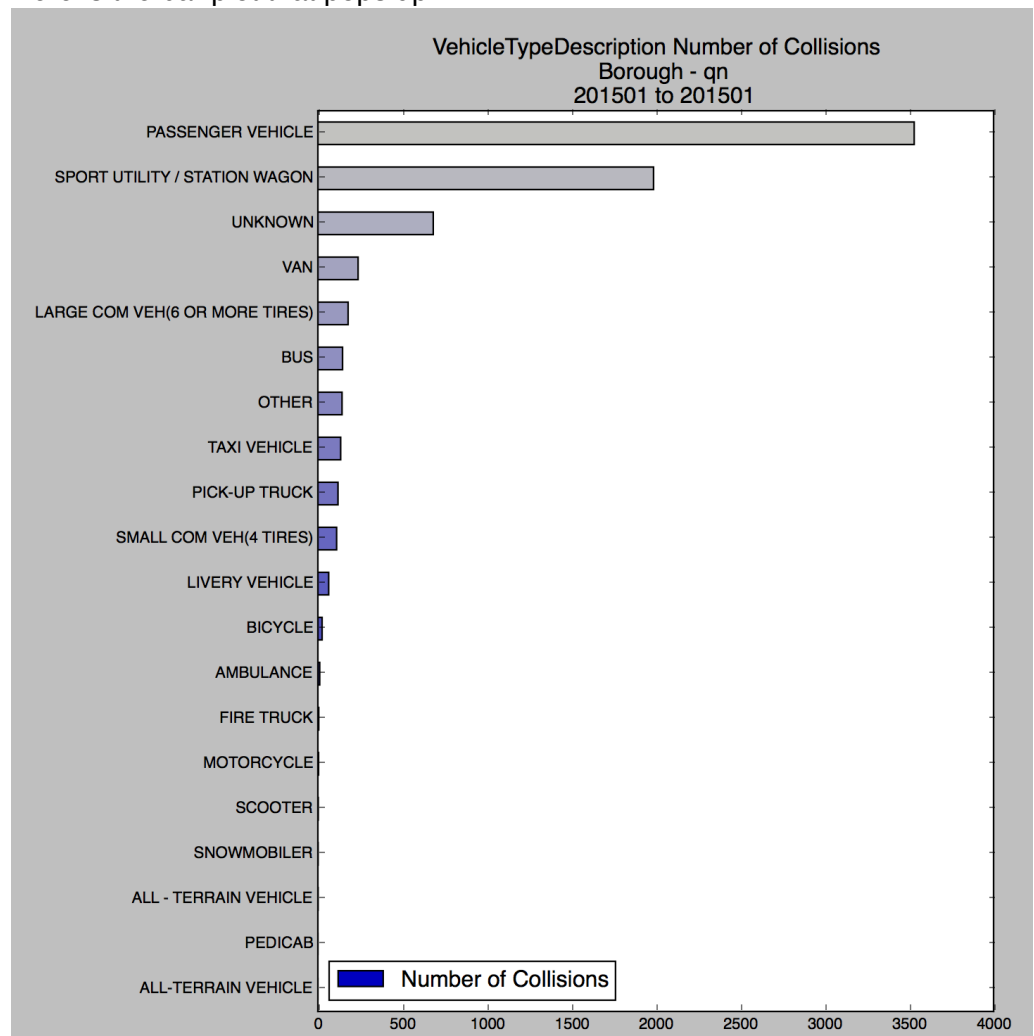
Here is the display in the Terminal:

```

Please Choose one Influencer from:
1:VehicleType
2:ContributingFactor
Your choice: 1
Please Choose one metric from:
1:Number of Collisions
2:CollisionInjuredCount
3:CollisionKilledCount
4:PersonsInjured
5:PersonsKilled
6:MotoristsInjured
7:MotoristsKilled
8:PassengInjured
9:PassengKilled
10:CyclistsInjured
11:CyclistsKilled
12:PedestrInjured
13:PedestrKilled
14:Injury_or_Fatal
Your choice: 1
Input anything to Close the Figure and Continue

```

Here is the bar plot that pops up:



18. From the graph above, seems like all these auto collisions happened in Queens, most of them involved passenger vehicle. As always, the plot will be saved in the corresponding folder under results.
19. Assuming the user is a graduate student that lives near NYU, and he is more interested in the injuries that are caused by the collisions near West 4th than somewhere in Queens. So let's go back to the previous menu and select a region and a metric that serves the interest for this user.
20. Let's go back to the top level of the menu and follow the command below.
 - select "2-CollisionSituation" by inputting 2
 - select "Specific Insight" by inputting 0
 - Here is what we will see:

```
Please input the number before the perspective you want to explore:Back
There are three types of information:
  1-CollisionSituation
  2-CollisionContributingFactors_Vehicle
You can choose one by input the number before the type above:1
Available Perspectives:
  1-City
  2-Borough
  3-Precinct
  4-Highway
  5-Tunnel
  6-Bridge
  7-Road
Please input the number before the perspective you want to explore:7
There are methods for this level:
0 Specific Insight
5 Rank_TOP10
1 Brief Summary
2 Plot by Time
3 InjuryKillPie
Please input the number before the method you want to use:0
Please specify the first character of the road you want to explore.
You can choose from ABCDEFGHIGKLMNOPQRSTUVWXYZ or - for others
Input a CAPITAL letter or - :
```

21. Here input the first letter or number for the street/road. The user is interested in West 4th Street. So let's input a capital letter "W". Now you will see a rather long list of road and street

WEST 136 STREET
WATER STREET
WEST TREMONT AVENUE
WEST 184 STREET
WEST 178 STREET
WALNUT AVENUE
WEST 54 STREET
WEST 162 STREET
WIMAN AVENUE
WESTCOTT BOULEVARD
WEST 34 STREET
WHITESTONE EXPRESSWAY E SVC RD
WEST 259 STREET
WEST 118 STREET
WEST 37 STREET
WOODWARD AVENUE
WATTS STREET
WEST 113 STREET
WESTMINSTER ROAD
WHITLOCK AVENUE
WEST 35 STREET
WATERS PLACE
WATERLOO PLACE
Please input the name:

```
Please input the name:WEST 4 STREET
There are methods for this level:
1 Brief Summary
2 Plot by Time
3 InjuryKillPie
Please input the number before the method you want to use:
```

[illegible]

25. The demo will end here, but there are a lot more to explore in the program. Have fun!