NYC Transport and Health Status

Topic

Objective: Study the associations of individual health status and transportation in the New York City.

Hypothesis

1. The relationship between EHR patients' hospital visiting frequency/disease and his/her public transportation circumstances.

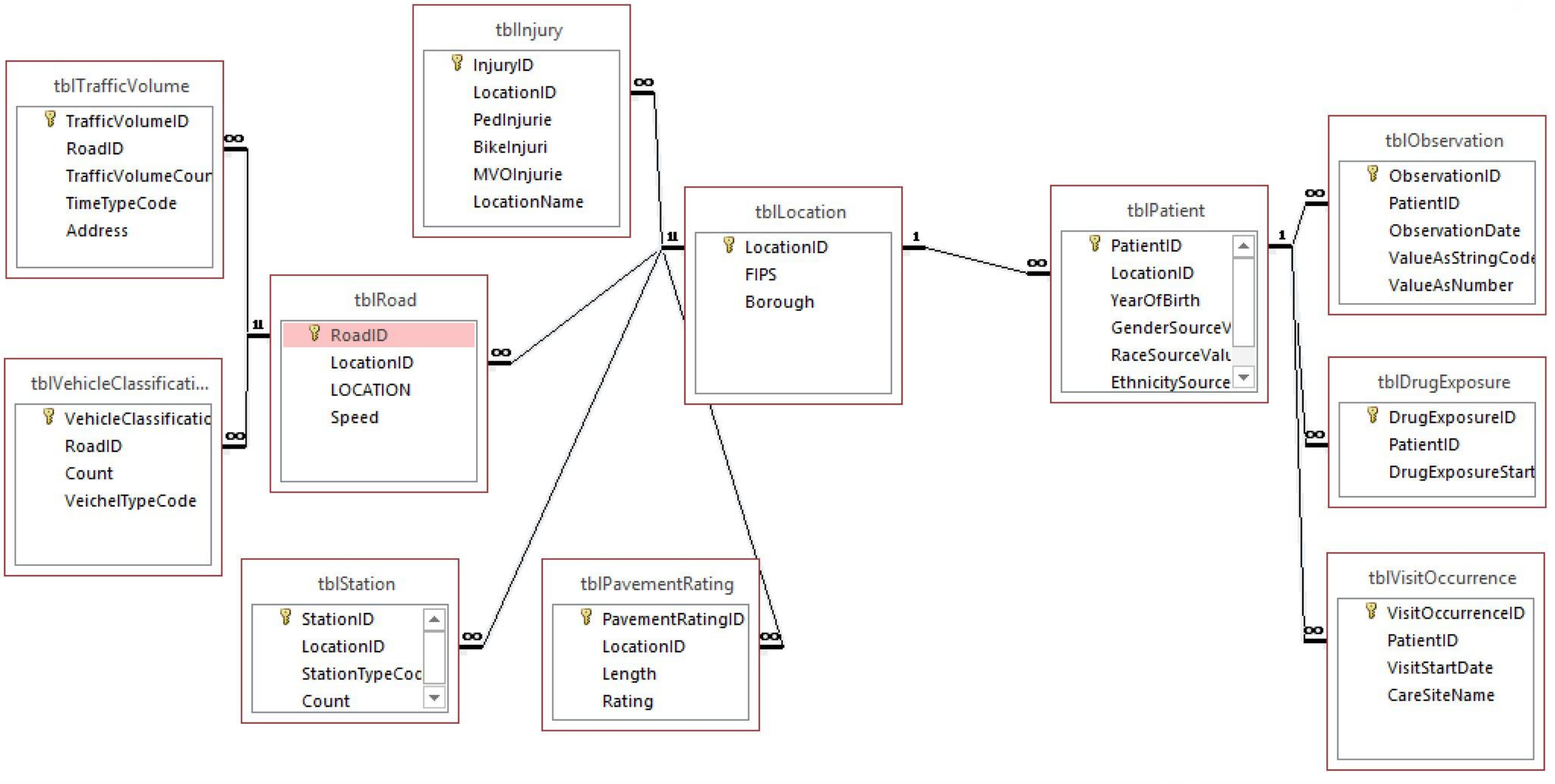
2. The relationship between EHR patients' drug visiting
   frequency and traffic condition/safety condition.

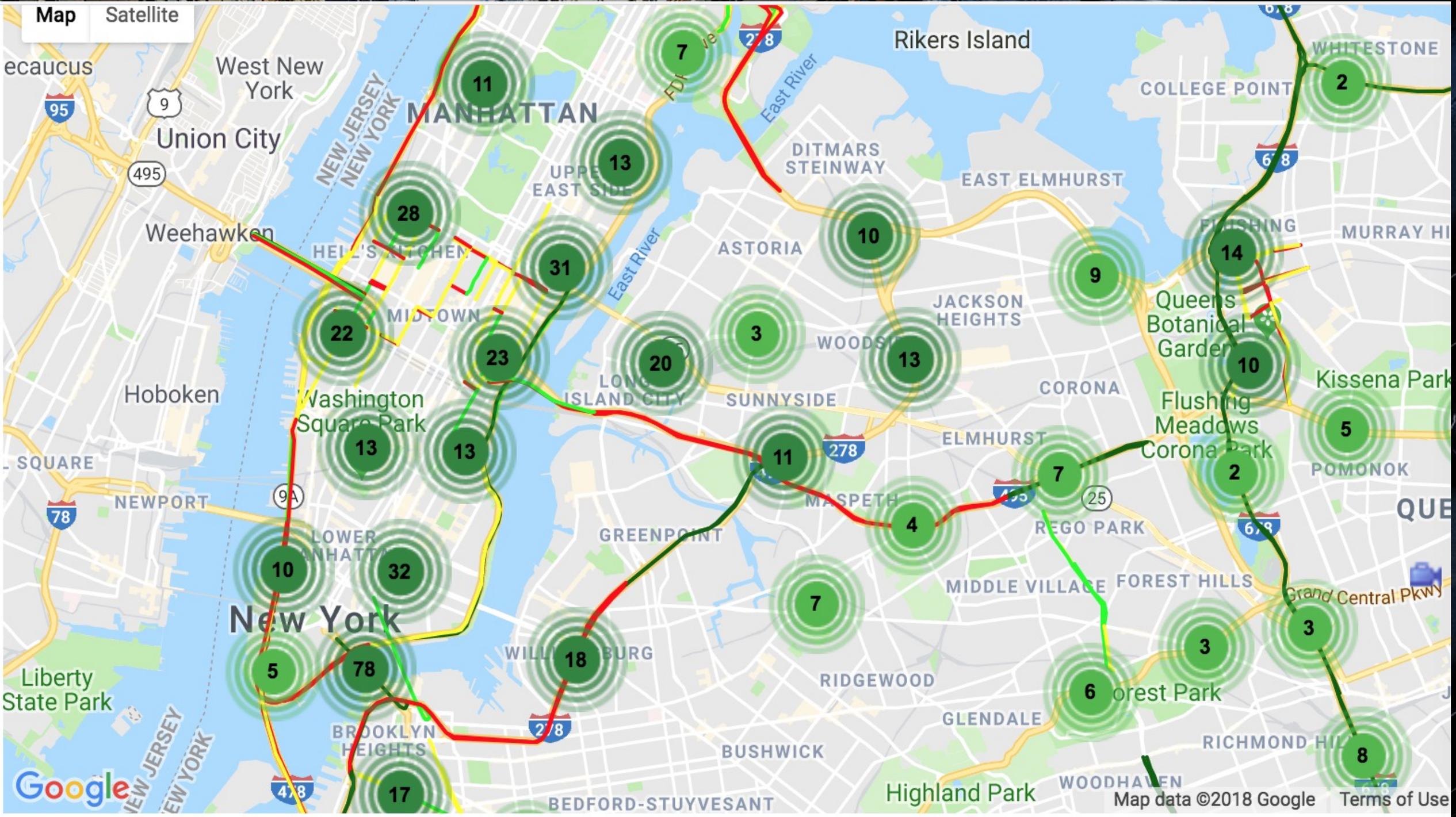
EHR Data Selection

- Patient: basic information - age, gender, race and type of disease
- Drug Exposure: the date patients took the drug
- Visit Occurrence: the date and name of hospital
- Observation: health conditions - BMI, weight and blood pressure

tblVisitOccurrence			
VisitOccurrenceID	PatientID	VisitStartDate	CareSiteName
1	1	6/21/2008	NYP
2	2	7/10/2007	WeillCornell
3	3	5/21/2006	MountSaina
4	4	3/30/2009	HospSpecialSurgery
5	5	9/23/2005	ColumbiaMed
6	6	8/30/2010	NYULangone
7	7	12/12/2007	NorthwellHealth

Relationships





Show Speeds

- No Data
- Heavy
- Moderate to Heavy
- Moderate
- Light

External Data Selection

NYC OpenData: https://data.cityofnewyork.us

New York City DOT: http://www.nyc.gov/

- Transportation methods

Bus stop shelter, bike parking shelter, subway station locations

- Traffic conditions

Traffic speed, traffic volume counts, bridge rating, street pavement rating,

vehicle classification counts

- Injuries
- Code Tables: VehicleTypeCode

Preprocessing - API

Convert longitude and latitude into FIPS: block API

```
> for( i in 1:21){
+   x<- mapply(latlong2fips,latitude=a$lat[i], longitude=a$long[i])
+   return(x)
+ }
> for( i in 1:21){
+   x<- mapply(latlong2fips,latitude=a$lat[i], longitude=a$long[i])
+   print(x)
+ }
[1] "360050281003001"
[1] "360050319001000"
```

Convert location to longi

```
> gGeoCode <- function(address,verbose=FALSE) {
+   if(verbose) cat(address,"\n")
+   u <- construct.geocode.url(address)
+   doc <- getURL(u)
+   x <- fromJSON(doc,simplify = FALSE)
+   if(x$status=="OK") {
+     lat <- x$results[[1]]$geometry$location$lat
+     lng <- x$results[[1]]$geometry$location$lng
+     return(c(lat, lng))
+   } else {
+     return(c(NA,NA))
+   }
+ }
> construct.geocode.url("W 96 ST")
[1] "http://maps.google.com/maps/api/geocode/json?address=W%2096%20ST&sensor=false"
>
> x <- gGeoCode("W 96 ST")
> x
[1]  40.79406 -73.97036
```

E-R Diagram

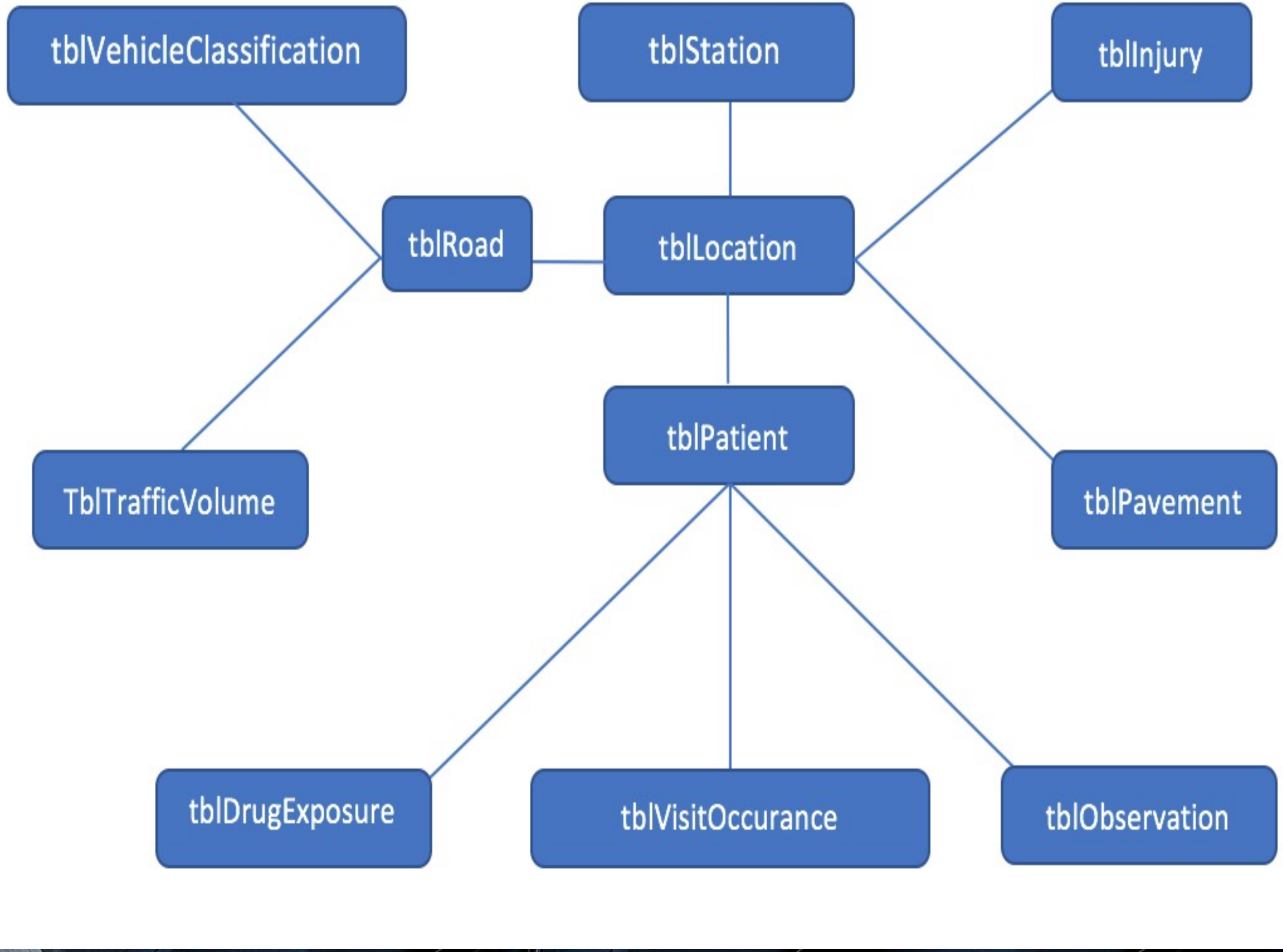

tblVehicleClassification
tblStation
tblInjury
tblRoad
tblLocation
tblPatient
TblTrafficVolume
tblPavement
tblDrugExposure
tblVisitOccurance
tblObservation

Table Schema

tlbLocation (PatientID, FIPS, Borough)

tblVisitOccurrence (VisitOccurrenceID, PatientID, VisitStartDate, CareSiteName)

tblPatient (PatientID, YearOfBirth, GenderSourceValue, RaceSourceValue, EthnicitySourceValue, LocationSourceValue, ConceptName)

tblObservation (ObservationID, PatientID, ObservationDate, ValueAsString, ValueAsNumber)

tblRoad (LocationID, Location , Speed)

tblDrugExposure (DrugExposureID, PatientID, DrugExposureStartDate)

tblInjury (LocationID, PedInjurie, BikeInjuri, MVOInjurie, LocationName)

tblPavementRating (PavementID, LocationID, Length, Rating, FIPS, Borough)

ect.

SQL Queries

1. Provide a list of the frequency (count) of patient visiting the NYP, Weill Cornell, MountSaina and MSK during 2005 and 2008 and the total amount of bike shelters, bus stops and subway stations. This query can help to analyze the relationship between patients' hospital visiting frequency and his/her public transportation circumstances.

2. Provide a list of whether patients have mental disease and the average traffic speed. This query is to analyze the relationship between patients' mental health and traffic condition.

SQL Queries

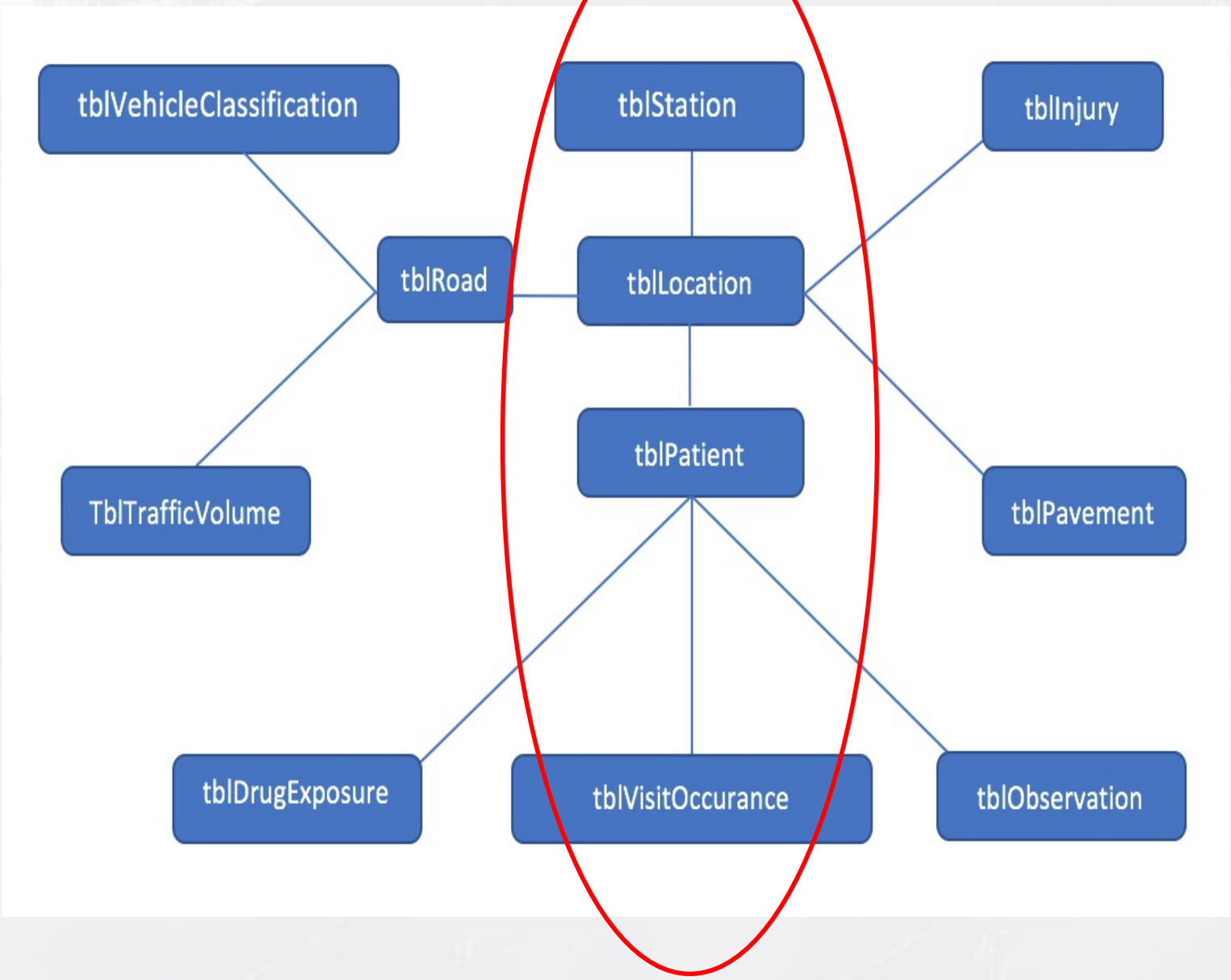
3. Provide a list of Hispanic male patients' **weights** and **pavement rating**. This query can help to analyze the relationship between Hispanic male patients' weights and his/her traffic condition.

4. Provide a list of subway and patients in **different age group**. This query is to analyze whether patients at different ages have a preference for taking public transportation.

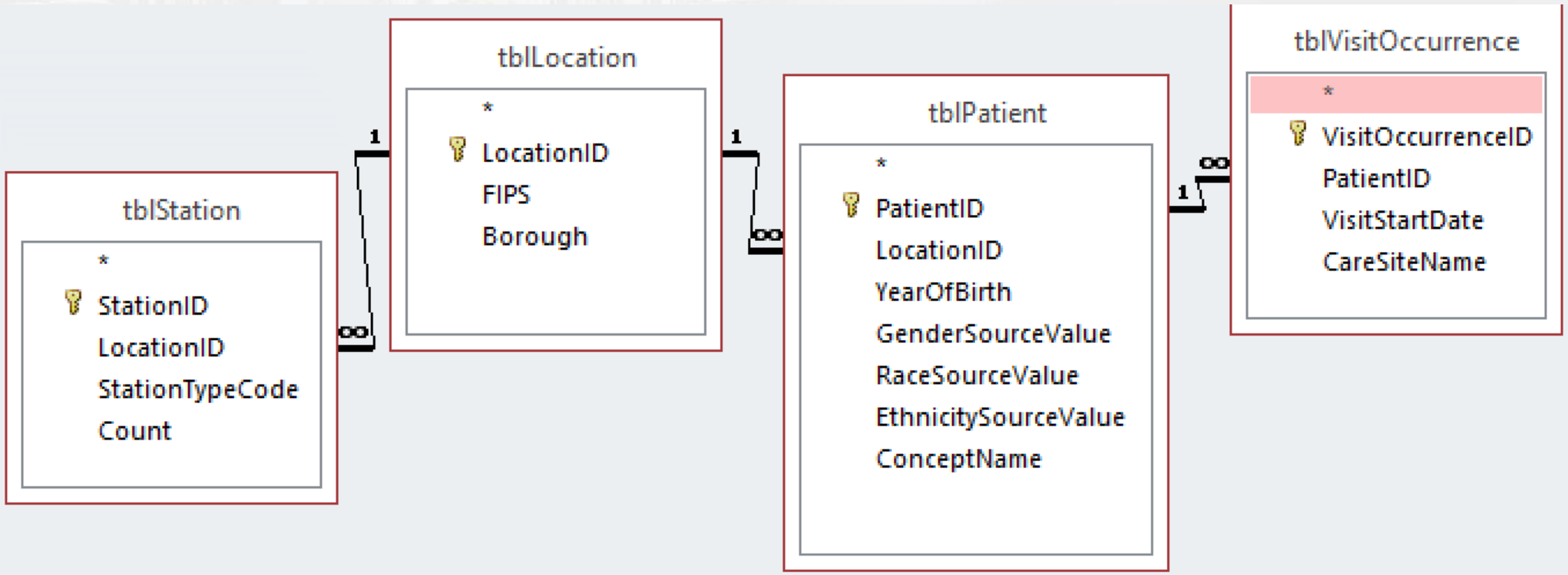
5. Provide a list of the frequency (count) of patient **visiting** the drug service providers and Motor Vehicle injuries. This query is to analyze the relationship between patients' drug visiting frequency and traffic safety condition.

Sample SQL Query1

1. Provide a list of the frequency (count) of patient visiting the NYP, Weill Cornell, MountSaina and MSK during 2005 and 2008 and the total amount of bike shelters, bus stops and subway stations. This query can help to analyze the relationship between patients' hospital visiting frequency and his/her public transportation circumstances.



Tables


tblStation
*
StationID
LocationID
StationTypeCode
Count
tblLocation
*
LocationID
FIPS
Borough
tblPatient
*
PatientID
LocationID
YearOfBirth
GenderSourceValue
RaceSourceValue
EthnicitySourceValue
ConceptName
tblVisitOccurrence
*
VisitOccurrenceID
PatientID
VisitStartDate
CareSiteName
1
∞
1
∞
1
∞

Sample SQL Query1

```
select*
from
(select tblPatient.PatientID, tblPatient.LocationID, Bus, Bike, Subway
from tblPatient
    left join
    (select tblLocation.LocationID, Bus, Bike, Subway
    from tblLocation
        left join
        (SELECT qryStationView.LocationID
        , Max(IIf(StationTyPeCode='bus',Count,Null)) AS Bus
        , Max(IIf(StationTyPeCode='bike',Count,Null)) AS Bike
        , Max(IIf(StationTyPeCode='subway',Count,Null)) AS Subway
        FROM qryStationView
        GROUP BY qryStationView.LocationID
        ORDER BY qryStationView.LocationID
        ) as tmp
        On tblLocation.LocationID = tmp.LocationID
    ) as tmp2
    On tblPatient.LocationID = tmp2.tblLocation.LocationID
) as tmp4
RIGHT join
(Select PatientID, count(VisitOccurrenceID) as n_visit
From tblVisitOccurrence
WHERE (VisitStartDate between #01/01/2005# and #12/31/2008#) AND ((CareSiteName = "WeillCornell")
  OR (CareSiteName = "NYP") OR (CareSiteName = "MountSaina") OR (CareSiteName = "MSK"))
Group by PatientID
) as tmp3
On tmp4.tblPatient.PatientID = tmp3.PatientID
```

Result1 table output

qryPublicTransportationVisiting					
LocationID	PatientID	Bus	Bike	Subway	n_visit
1	1	1	1	2	1
2	2	1	1	1	1
3	3	1	1	2	2
4	4	1	1	2	1
6	6	1	1	1	1
9	9	1	1	4	1
10	10	1	1	1	1
11	11	1	1	2	1
12	12	1	1	2	1
13	13	1	1	1	1
14	14	1	1	1	1
15	15	1	1	1	1
16	16	2	1	1	1
19	19	1	1	1	2
20	21	1	1	1	1

Result1 linear regression

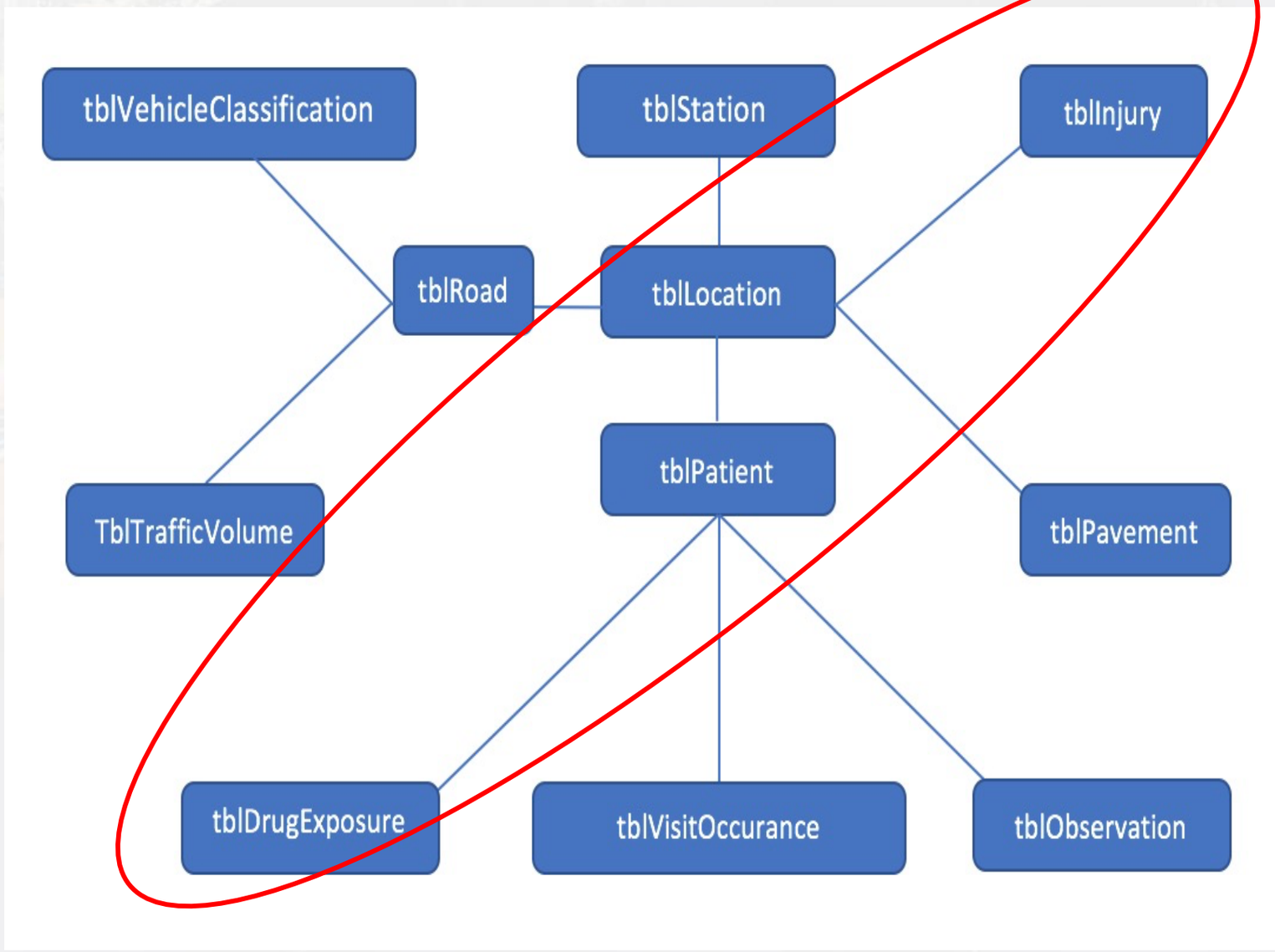
```
Call:
lm(formula = n_visit ~ Bus + Subway, data = qdat)

Residuals:
     Min       1Q   Median       3Q      Max
-0.17308 -0.09615 -0.05769 -0.05769  0.90385

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.07692    0.42106   2.558   0.0285 *
Bus         -0.05769    0.32024  -0.180   0.8606
Subway       0.03846    0.10212   0.377   0.7143
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3006 on 10 degrees of freedom
Multiple R-squared:  0.02083,   Adjusted R-squared:  -0.175
F-statistic: 0.1064 on 2 and 10 DF,  p-value: 0.9001
```

Result2 table output

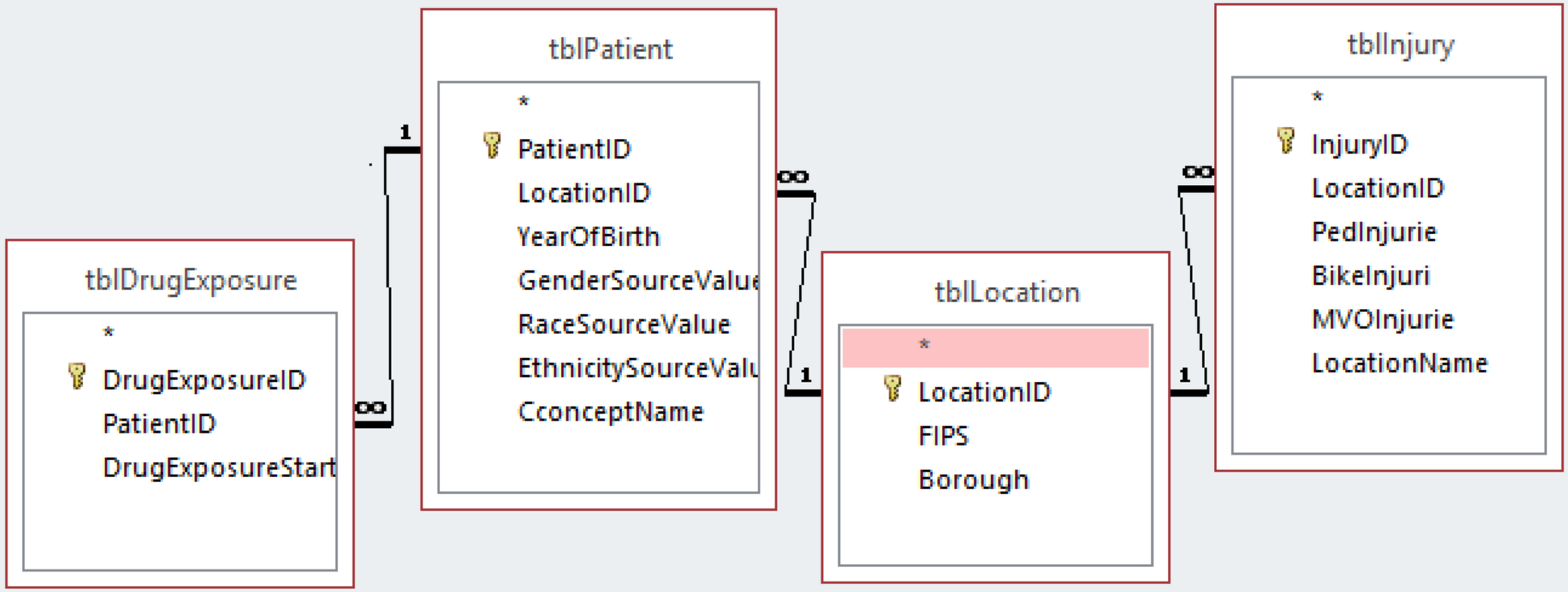
qryDrugInjuryRelationship			
PatientID	LocationID	n_injuries	n_drug
1	1	1	2
2	2	3	2
3	3	1	3
4	4	5	3
5	5	3	3
6	6	5	3
7	7	1	3
8	8	3	3
9	9	1	3
10	10	3	3
11	11	3	2
12	12	2	2
13	13	5	2
14	14	6	2
15	15	4	2
16	16	2	2
17	17	5	2
18	18	6	2
19	19	5	2
20	20	8	2
21	20	8	2

Sample SQL Query2

4. Provide a list of the frequency (count) of patient visiting the drug service providers and MOVinjuries. This query is to analyze the relationship between patients' drug visiting frequency and traffic safety condition.



Tables


tblPatient
*
PatientID
LocationID
YearOfBirth
GenderSourceValue
RaceSourceValue
EthnicitySourceValu
CconceptName
tblDrugExposure
*
DrugExposureID
PatientID
DrugExposureStart
tblLocation
*
LocationID
FIPS
Borough
tblInjury
*
InjuryID
LocationID
PedInjurie
BikeInjuri
MVOInjurie
LocationName
1
∞
1
∞
1
∞

Sample SQL Query2

```
select tmp4.tblPatient.PatientID, tblPatient.LocationID, n_injuries, n_drug
from
(select*
from tblPatient
    left join
    (select *
    from tblLocation
        left join
        (SELECT tblInjury.LocationID, Count(tblInjury.MVOInjurie) as n_injuries
        FROM tblInjury
        GROUP BY tblInjury.LocationID
        ORDER BY tblInjury.LocationID
        ) as tmp
        On tblLocation.LocationID = tmp.LocationID
    ) as tmp2
    On tblPatient.LocationID = tmp2.tblLocation.LocationID
) as tmp4
RIGHT join
(Select tblDrugExposure.PatientID, count(tblDrugExposure.DrugExposureID) as n_drug
From tblDrugExposure
Group by tblDrugExposure.PatientID
) as tmp3
On tmp4.tblPatient.PatientID = tmp3.PatientID
```

Conclusion

For this project, we built a database that linked traffic burden, traffic safety and transportation tools to the EHR data. And according to this database, we can extract data to analyze the relationship between NYC transportation and residents' health status.