Database Phase 1 -- updated ver.

1.Members: Tianyi Lin, Yang Cao
2.Target domain: Our database describes the locations of crashes involving bikes and cars in the Chapel Hill Region of North Carolina.We download the data online, and the original source is “from police-reported bicycle-motor vehicle and pedestrian-motor vehicle collisions that occurred on the public roadway network, public vehicular areas and private properties (if reported)”, as stated on the website.

3. List of queries (Note: for completion of phase 2, we also write the input & output and stored procedures name for each question)

1) What kind of crash type (pedestrian AND bike) is the most frequent (top 5) around certain area?
    a) input: city (drop down)
    b) output:
        i) tables (1 and 2): most frequent time of the crash, severity of ped/biker
        ii) table 3 (legend): kind of crash type
    c) procedures: CrashTypeRate_Bike(city VARCHAR(18)), CrashTypeRate_Ped(city VARCHAR(18)), ShowCrashTypes_Comb()

2) What areas have higher accident rate (e.g. bicycle crash)?What areas have higher accident rate at night and what areas have higher accident rate at day (based on time range)?
    a) input: time range
    b) output: accident rate, city, crash type, severity of injury
    c) procedures: AccidentRate_Bike(t_from NUMERIC(4,1), t_to NUMERIC(4,1)), AccidentRate_Ped(t_from NUMERIC(4,1), t_to NUMERIC(4,1))

3) Does light condition correlate with bicycle/pedestrian crash rate? (light_bike, light_ped)
    a) input:condition
    b) output: both table with columns of: type of light condition && corresponding # of crashes, % of light condition = && crash most frequent severity

4) Does road surface correlated with bicycle/pedestrian crash rate? (road_charact_bike, road_charact_ped)
    a) input:condition
    b) output: road surface type, count of crash, weather, severity

5) Does weather correlated with bicycle/pedestrian crash rate? (weather_bike,weather_ped)
    a) input:condition
    b) output: type of weather & # of crashes, several frequent month…

6) Does exceeding the speed limit (driver) correlate with crash rate? (output: coordinate with highest rate, link: other factors, navigate to corresponding pages)?
    a) input:condition
    b) output: percentage of exceedLim/all, percentage of belowLim/all
    c) procedures: ExceedSp_Bike(), ExceedSp_Ped()

7) For bicycle crashes, does Alcohol Detected for driver correlate with crash rate?
    a) input:
    b) output: percentage of alco/all for biker, percentage of alco/all for driver, time of the day, location

8) Do bike/pedestrian crashes have higher severity of injury?
    a) input:
    b) output:
        i) all types of injury
        ii) 2 tables, percentage of severe injury (level B/C), time, weather, alcohol, county
    c) procedures: ShowInjuryTypes(), Injury_Bike(), Injury_Ped()

9) Correlation between ambulance response and severity of injury
    a) input: time
    b) output:
        i) table 1 (ped): for ambulance=yes/no, the most frequent severity level
        ii) table 2 (bike): for ambulance=yes/no, the most frequent severity level
    c) procedures: AmbulanceSevri_Bike(t_from NUMERIC(4,1), t_to NUMERIC(4,1)), AmbulanceSevri_Ped(t_from NUMERIC(4,1), t_to NUMERIC(4,1))

10) For a selected age group, show all crashes data for the victim.
    a) variable: time
    b) input: age group (drop down?)
    c) output:
        i) table 1 (bike) count, Biker.bikeage_gr, bike_injur, bike_race, bike_dir, bike_sex, bike_pos, bike_alc_d
        ii) table 2 (ped) count, pedInjure.pedage_grp, ped_pos, ped_race, ped_injury, ped_sex #need to check

d) procedures: AgeGpAccidentRate_Bike(age VARCHAR(7)), AgeGpAccidentRate_Ped(age VARCHAR(7))

11) correlation between time(input) and hit and run rate
    a) input: time
    b) output: hit and run rate, weather
    c) procedures: HitRun_Bike(t_from NUMERIC(4,1), t_to NUMERIC(4,1)), HitRun_Ped(t_from NUMERIC(4,1), t_to NUMERIC(4,1))

12) which type of location (rural/urban) has more frequent crashes?
    a) input: rural/urban (drop down)
    b) output: hit and run rate, time, alcohol detected, weather, severity
    c) procedures: LocAccidentRate_Bike(loc VARCHAR(5)), LocAccidentRate_Ped(loc VARCHAR(5)) # need to check

13) driver information
    a) input: pedestrian/bike (drop down)
    b) output: summary of driver info: sex, race, age, crash type, driver severity
    c) procedures: DriverInfo(type VARCHAR(10))

14) do crashes happen more often at intersection/non-intersection
    a) input:
    b) output: 1 table (union ped and bike crashes), crash rate (intersection/all, or non-intersection/all), severity, weather, light cond, num_lanes
    c) procedures: IntersectAccidentRate()

15) do crashes happen more often when there's no traffic control?
    a) input:
    b) output: 2 tables, traffic control rate, severity, weather, light cond, num_lanes
    c) procedures: Traffic_Bike(), Traffic_Ped()

4.Relational data model:

- 1.PedCrashRdCond
    - rd_defects,rural_urba,city,locality,rd_feature,light_cond,rd_charact. ,rd_surface, developmen, traff_cntr,rd_conditi,region, rd_class, weather, num_lanes, rd_config

PedCrashID	rd_defects	rural_urba	city	locality	rd_feature	light_cond	rd_charact

1	“None”	“Urban”	“Chapel Hill”	“Urban (>70% Developed)”	“No Special Feature”	“Dark - Roadway Not Lighted”	“Straight - Level”

rd_surface	developmen	traff_cntr	rd_conditi	region	rd_class	weather	num_lanes	rd_config
“Smooth Asphalt”	“Commercial”	“No Control Present”	“Dry”	“Piedmont”	“Public Vehicular Area”	“Clear”	“Unknown”	“Unknown”

- 2.PedInjure
    - for injured people profile
    - ped_pos,ped_race,pedage_grp,ped_age, ped_injury,ped_sex

pedInjure

PedCrashID	ped_pos	ped_race	pedage_grp	ped_age	ped_injury	ped_sex

- 3. DiverBiker_PedCrash
    - driver/biker profile
    - speed_limi, drvr_vehty, drvr_injur, drvr_sex, drvrage_gr, drvr_race, drvr_age, drvr_estsp, drvrage_gr, drvr_alc_d

PedCrashID	speed_limi	drvr_vehty	drvr_injur	drvr_sex	drvrage_gr	drvr_race	drvr_age	drvrage_gr	drvr_estsp	drvr_alc_d

- 4. PedCrashDetail
    - crash detail
    - crsh_sevri,“ambulancer”, crash_time, crash_year, county, longitude,latitude, crash_mont,crash_type,city,locality,ped_pos,drvr_injur,crashday,crash_loc,

crash_hour,geo_shape, crash_date, crash_grp,hit_run

PedCrashID	crsh_sevri	ambulancer	crash_time	crash_year	county	longitude	latitude	crash_mont	crash_type

city	locality	ped_pos	drvr_injur	crash day	crash_loc	crash_hour	geo_shape	crash_date	crash_grp	hit_run

- 5. ReasonPed
    - Bike/Pedestrian Alcohol Detected, Driver Estimated Speed, Speed Limit,Driver Alcohol Detected

PedCrashID	crashalcoh	excsspdind	ped_pos	drvr_injur	hit_run	drvr_estsp	exceed Speed	speed_limi

- 6.BikeCrashTime

BikeCrashID	crash_time	crash_hour	crashday	crash_mont	crash_year
1	“10:12”	10.0	“Saturday”	“July”	2011

- 7.BikeCrashLoc

BikeCrashID	lat	lon	county	city	rural_urba	crash_loc	developmen
1	35.9100670923	-79.0745027481	“Orange”	“Carrboro”	“Urban”	“Non-Intersection”	“Commercial

- 8.BikeCrashRdCond

BikeCrashID	rd_defects	rd_feature	rd_charact	rd_surface	rd_conditi	speed_limi	traff_cntr	weather	rd_config	num_lanes	development	light_cond
1	“Non”e	“No Special Feature”	“Straight - Level”	“Smooth Asphalt”	“Dry”	“20 - 25 MPH”	“No Control Present”	“Clear”	“Two-Way, Not Divided”	“Unknown”	“Commercial”	“Daylight”

- 9.BikeCrashResult

BikeCrashID	ambulancer	crash_type	crsh_sevri	hit_run	bike_injur	drvr_injur
1	“Yes”	“Motorist Overtaking - Bicyclist Swerved”	“K: Killed”	“No”	“K: Killed”	“O: No Injury”

- 10.Biker

BikeCrashID	bike_injur	bike_race	bike_dir	bikeage_gr	bike_age	bike_sex	bike_pos	bike_alc_d
1	“K: Killed”	“White”	“With Traffic”	“70+”	“70+”	“Male”	“Travel Lane”	“No”

- 11.Driver_BikeCrash

BikeCrashID	drvr_vehty	drvr_injur	drvr_sex	drvr_race	drvr_age	drvrage_gr	drvr_estsp	drvr_alc_d
1	“Passenger Car”	“O: No Injury”	“Male”	“White”	“70+”	“70+”	“11 - 15 mph”	“No”

- 12.BikeCrashReason

BikeCrashID	crashalcoh	excsspdind	drvr_alc_d	bike_alc_d	bike_pos	bike_dir	drvr_estsp	on_rd

5. Sql statements are in Procedures.sql
6. how to load database
    - We first put the variables and their corresponding value in a java object, we populate the data into the database.

7.
    - We expect the output to be table of data, and our work includes combine multiple rows of data and present it to the user in a user friendly view. Possibly some design on the table format to make the information easier to read.

8.
    - Since we are in section 315, we choose to minorly focus on complex data extraction. So far, our raw data consists of both csv and json files, and we also plan to do some statistical analysis/calculation on these files before we push all of the datasets to the database.