

Crash Data Introduction:

Crash data are information that derives from the information that comes from a reportable crash. A reportable crash according to Title 75, Pennsylvania Consolidated Statutes, Section 3746(a) is:

An incident that occurs on a highway or traffic way that is open to the public by right or custom and involved at least one motor vehicle in transport. An incident is reportable if it involves:

- Injury to or death of any person, or
- Damage to any vehicle to the extent that it cannot be driven under its own power in its customary manner without further damage or hazard to the vehicle, other traffic elements, or the roadway, and therefore requires towing.
- Crash data does not include non-reportable crashes or near misses
- Crash data may not contain complete information, some elements may be unknown

The Database:

Public downloads of crash data from the PennDOT Open Data Portal are provided as CSV files, one file for each table. You will want to extract your data based on specific criteria by using SQL. You can use Structured Query Language (SQL) to query, update, and manage relational data. Packages such as MSAccess or Crystal Reports (reporting software) also have query builders that aid in the query building process. Before you begin writing your queries it is important to familiarize yourself with the data elements so you can better define your criteria so that the results you desire are accurate.

The tables that have been sent to you are shown here with a general description about the data elements contained within.

Table Name	Description
CRASH	Information about the crash such as: Where: County, Municipality, Work zone When: Date, Time, Day of Week, Hour of Day, Month of Year Item Counts: People, Vehicles, Unbelted, Fatal, etc.
COMMVEH	Information about commercial vehicles, such as carrier information, the cargo body type, Hazmat information, and official agency registration numbers.
CYCLE	Information that pertains to motorcycle/pedal cycles, such as helmet usage and appropriate attire and other accessories such as side bags
FLAG	Series of Yes/No items that help refine lookups for specific factors about the crash such as: Drinking Driver, Use of a Cell Phone, Fatal Crash, Motorcycle involved, and over 60 other crash defining items.
PERSON	Information about all people from all units related to the crash such as: Age, Sex, Drug and alcohol results, Where they sat and in which vehicle, Were they ejected from the vehicle? etc.

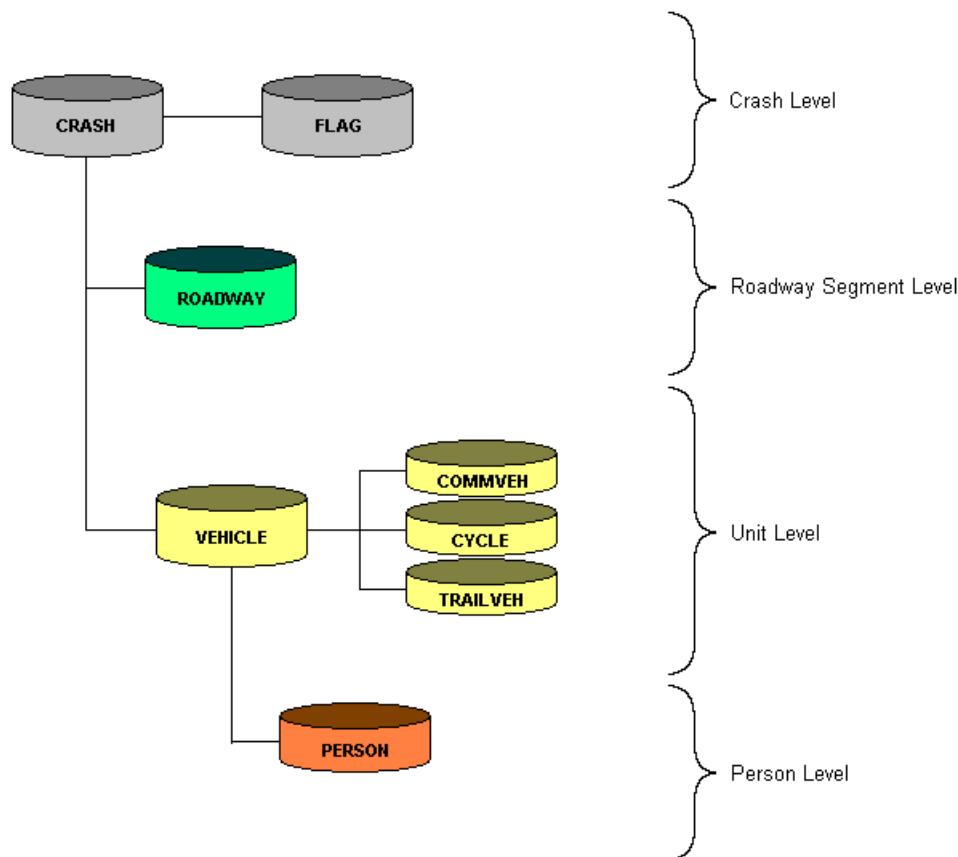
ROADWAY	Information about all the roadways involved in the crash such as: Route number or name, Segment, Offset, Type of Roadway, Rating, and many other Roadway defining elements.
TRAILVEH	Information about the types and kind of trailers that were being towed by vehicles.
VEHICLE	Information about all vehicles involved in the crash such as: Body Type, Commercial Vehicle, Movement, Position, Unit number in the crash and other vehicle related information.

The next visual on the following page will illustrate at what level of the crash each table pertains to and a general understanding of how the tables relate to one another.

For example if you are writing SQL against any of the gray cylinders pictured below you will be counting the number of crashes where your specific criteria occurred.

As you move down through the table structure you will very often be counting other attributes involved in the crash.

For example, if you were looking at drivers from the PERSON Table one crash event may be counted more than once as more than one driver involved in the crash may meet your query criteria.



Please familiarize yourself with the data dictionary that contains all the fields for these tables along with their coding/constraints criteria.

The fields that have a “* See Column Code” under the Constraints column only have a certain amount of valid code values that may be entered into that field. Since you will be working directly with the raw data you will want to be careful and only apply valid criteria as your filter to a field. Follow the instructions at the beginning of the Data Dictionary.doc file to learn how to navigate back and forth on these lookup values.

Defining Criteria:

When writing your query you will want to decide upon a few things before you begin:

- What exactly do I want to count or lookup?
- What tables contain the data I need?
- How do these tables relate to one another?
- Do I want any calculated fields?
- Do I want to limit my results with any specific criteria?
- How do I want to sort my records?

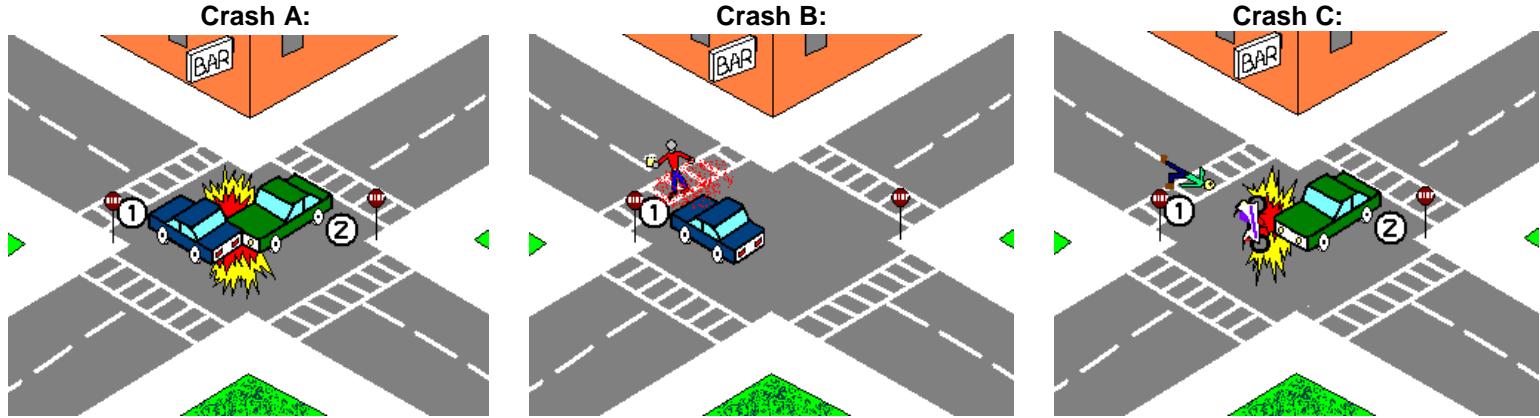
These questions may seem relatively basic, but it is very important to know exactly what you want and are able to extract it correctly as expensive decisions may be made on the basis of the results. The data in these tables was compiled mainly to make informed decisions to reduce the number and severity of crashes. It cannot analyze and interpret the data for you. Therefore you must analyze the data to determine its value in helping you make decisions with your own set of criteria. Before accepting the results of your queries try and verify its accuracy before making any decisions based upon the output. Things you may ask yourself:

- Does this data make sense?
- Are these results what I expected?
- Do the crash types match the roadway's geometry?

Additionally, you should:

- Compare data to older reports for the same location
- Compare data to other reports for similar locations
- Use maps and other GIS tools to confirm locations
- Start with small pieces of a report and combine them together to build a complex report.
(For example, if you are looking for causations along an entire route look at one causation first, and then another. Alternatively, you can look at road segments rather than the entire road.)

For a better representation let's examine some activity at an intersection and determine the crash statistics that pertain to both the individual crashes and to the location itself. The example on the following page will contain three crashes for an intersection, some information about each crash, and some questions.



<p>In the crash above car one driven by a 76 year old man accompanied by his wife 74 ran a stop sign and was struck by car two, luckily they were wearing their seatbelts and even though his wife suffered major injuries both will survive. Unfortunately, the driver of car two, age 17, wasn't wearing her seatbelt and was talking on her cellular phone at the time of the crash and was killed in the collision while her friend of the same age, also not wearing a seatbelt, was also killed.</p>	<p>In the crash above a pedestrian who had too many drinks on Memorial Day wandered into the path of a moving vehicle, shown as car one. The pedestrian later recovered from minor injuries. The driver of the vehicle who wasn't wearing his seatbelt at the time was unharmed in the incident.</p>	<p>In the crash above vehicle one, a 17 year old male speeding on his motorcycle while not wearing a helmet, ran head-on into oncoming traffic and was struck by vehicle two. He was thrown from his bike and was killed immediately. Upon questioning the driver of vehicle two who was wearing a seatbelt and unharmed, the police became suspicious of his behavior and suspected alcohol. Test results later revealed the driver to be above the legal limit.</p>
<p>Items to note on Crash A:</p> <p>This was a very severe crash. How many injuries were there and what were they? There was only one major injury in this crash. Fatalities are not considered injuries, and elevate this crash to the fatal crash category. Could this crash be attributed to aggressive driving? Well we find that the driver action for car two was that she was talking on her cell phone, had she not been on her cell phone she may have been able to avoid the crash, but that is speculation and also not an aggressive action as defined by PennDOT. However, car one ran a stop sign and that is considered aggressive driving. So the answer is yes. How many fatal crashes would show up in a report based on this crash? Although there were two fatalities, this is only one fatal crash incident. Would this crash show up on a report for not wearing restraining devices? As long as anyone in any unit wasn't wearing a seatbelt then the answer would be Yes. Lastly, PennDOT also flags specific driver age categories. This specific instance would show up on a report that would ask for either drivers over 75 or drivers age 17.</p>	<p>Items to note on Crash B:</p> <p>We notice that this crash only involved one vehicle. The other item involved happened to be a pedestrian. Would this crash even be recorded by PennDOT as a reportable crash? If a vehicle had to be towed or there was an injury then Yes this is a reportable crash. The crash would also be classified as an unbelted crash, and the drinking pedestrian would also classify this crash as alcohol related, but not a Drinking Driver related crash.</p>	<p>Items to note on Crash C:</p> <p>We have another severe crash here and this crash will also be classified as a fatal crash with one fatality. We have a number of items that will be recorded in the database. The fact that The motorcycle driver wasn't wearing a helmet would be recorded; the crash impact would be recorded as head-on. The two questions here are; Would this go as an unbelted crash? and would it also qualify as a crash where a driver was ejected? The answer to both of these is no, You can't be ejected from a motorcycle and motorcycles do not qualify as having a restraining device. This crash will however be DUI and alcohol related as the driver later turned out to be over the legal limit.</p>

Items to note on the crash location: We will call this site the intersection of X street and Y street. The three crashes above will be recorded for X street they will also be recorded for Y street. So be very careful when looking at all crashes on both roadways as you may get the same crashes twice, and come up with 6 crashes instead of 3. How many major injury crashes were there at this intersection? The answer is none, Fatal crashes take priority and they would be classified as fatal crashes. Other items to note: there would be 2 unbelted crashes, 2 alcohol related crashes, and 2 aggressive driver related crashes. The next section will help describe crash location criteria.

Crash Location Criteria:

Intersection versus Mid-block locations:

- *Mid-block Locations*: PennDOT will populate the database with one and only one roadway. Normally it will be numbered as roadway **number 3**.
- *Intersections*: PennDOT will populate the database with a minimum of two roadways. Generally the roads will be numbered as 3 and 4; however it could go up to roadway 9 depending on the configuration of the intersection.

Categories of Roadways:

State Routes are PennDOT maintained roadways which include Interstates, US (Federal) Routes, and State Roads. These will contain:

- Two digit County
- Four digit State Route Number
- Four digit Segment (or 9999 for unknown)
- Four digit Offset (or 9999 for unknown)
- They May or may not have a Street Name and Street Ending

Local Roads include roads maintained by the County, Local Municipality, or privately maintained roadway open to the public. These will contain:

- Two digit County
- Route number (usually blank but might contain a 3 digit route number preceded by a letter, most commonly T for Township Routes or C for County Routes)
- Street Name
- Street Ending

Defining location on a State Route:

Each State Route is separated into distinct but not necessarily contiguous segments. The segments commonly begin with segment 0010 and increment by 10 until the end of each route within each county. The segments and offset within the segments increase as you head either east or north. Upon reaching a county line, the segment will usually start again at 0010 (except for interstates).

Divided Highways are separated by a barrier, land, or a minimum of 4 feet of painted lines and are heading in opposing directions will have two separate segment designations, one for each direction. These are known as even/odd segment pairs. The odd segment corresponding to an even segment will always be one digit higher than the even segment. For example, US Route 15 in Adams County, starting at the Maryland state line, has 2 northbound lanes and 2 southbound lanes separated by a median. The northbound lanes are considered segment 0010 and the southbound lanes are considered segment 0011.

Null Segments:

There are numerous examples of State Routes that are disjoint. The point from which the first section of a state route ends to the point where the next section of the same state route picks up again within any given county is referred to as a null segment. This can occur where either a roadway is shared by multiple state routes (example 22/322 north of Harrisburg) or where a section of state route has been turned back to a local municipality (example Route 72 in the City of Lebanon)



In Tionesta Borough, Forest County, US Route 62 heads north into Tionesta Station, crosses the Allegheny River then heads north towards Warren. Route 36 heads north into Tionesta, crosses the river and then heads north towards Titusville.

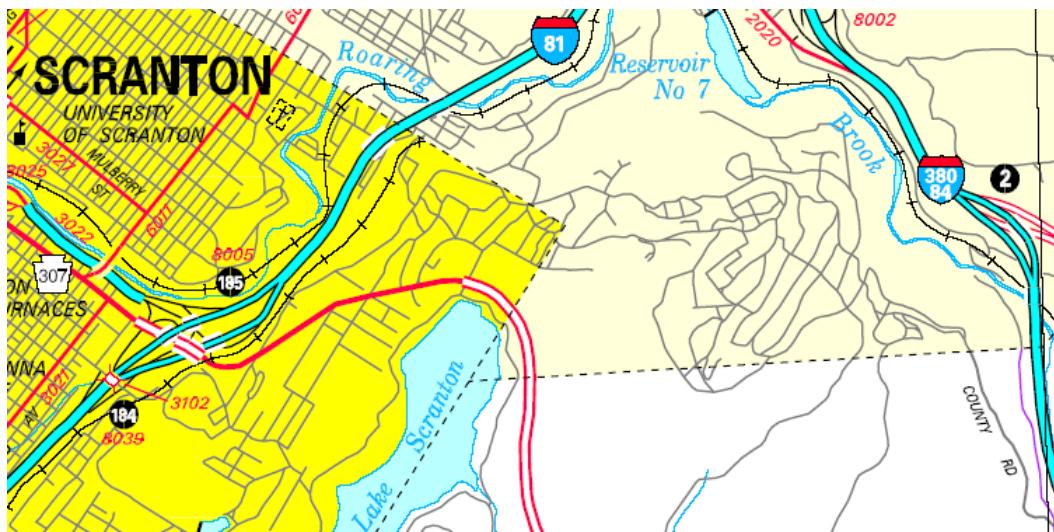
- Any crashes that would occur on the bridge but not at an intersection on either side would be coded as County 27 Route 0062.
- If the crash were to happen at an intersection at either end of the bridge, then both roadways would be included.
- Between the two intersections, a **null segment** would exist for route 36, and no crashes would be coded against that null segment.

When a roadway has multiple route designations, the way to determine which will be the null segment is based on 2 factors. The first factor is route designation. Interstates always take priority, followed by US Routes, and finally State Highways. If there are multiple routes that are in the same category then the lowest number prevails.

For Example:



For US Route 62/US Route 322/Route 8 in the city of Franklin. SR0062 and SR0322 are US highways and SR0008 is a state highway. That section of roadway is considered SR0062 and would be a **null segment** for both SR0008 and SR0322.

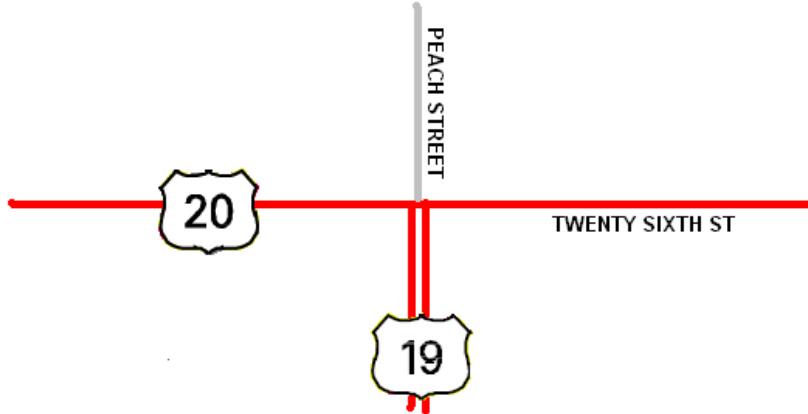


Interstate 380/84 in Lackawanna County would be coded as SR0084 and would be a **null segment** for Interstate 380 (SR0380).

Roadway Data at Intersections:

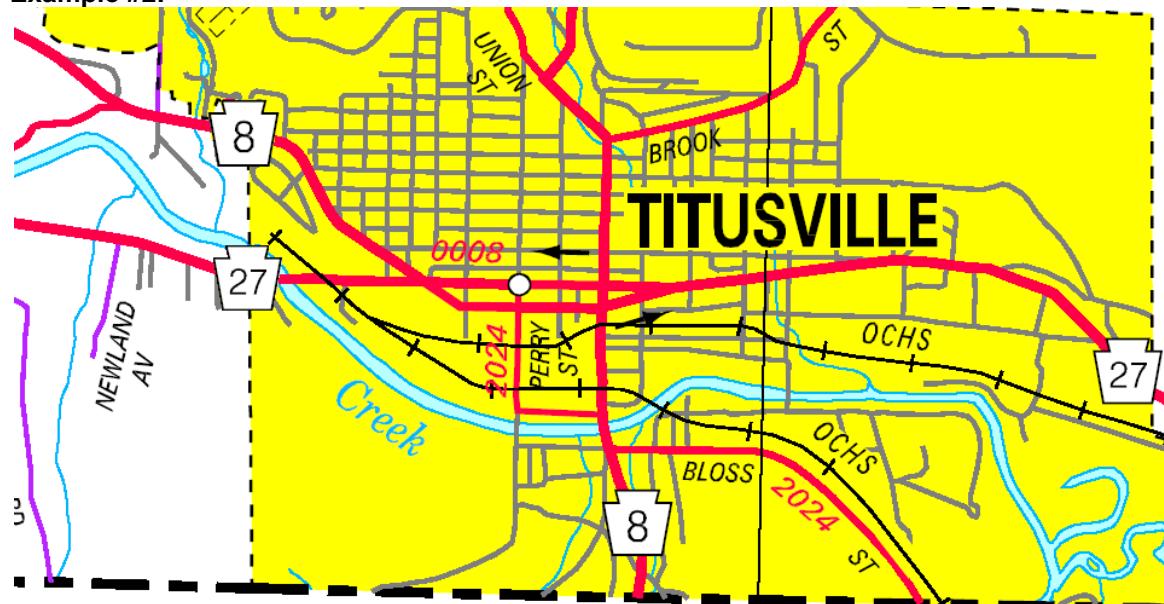
When a crash is located at an intersection with a divided highway, both the even and odd segment will be included with the roadway data. The exception would be when the divided highway is made up of two separate named roadways, usually within a city or borough.

Example #1



In the City of Erie, US Route 19 ends at US Route 20 at the corner of 26th and Peach. US Route 19 is a divided highway and US Route 20 is a non-divided highway. Peach Street becomes a local one-way street north of US20. The 4 roadways that will be included with this location are the even and odd segments of the divided highway (SR0019, segments 0410 and 0411), the non-divided highway (SR0020, segment 0580) and the local road (Peach Street).

Example #2:



In the City of Titusville, Route 8 splits and becomes two separate one way streets, one carrying north bound traffic (Central Avenue) and the other carrying south bound traffic (Spring Street).

- The intersection of Perry Street (SR2024) and Spring Street (SR0008) would have only 2 roadways listed. SR2024 segment 0010 offset 0231 and SR0008 segment 0031 offset 0896.
- The intersection of Perry Street (SR2024) and Center Avenue would have 3 roadways listed. SR2024 segment 0010 offset 0000, SR0008 segment 0030 offset 0896, and the local portion of Perry Street.

Non-typical Roadway data:

Local Roads:

Local roads do not follow the same conventions as State Routes in that they do not get stored with a numeric route, segment and offset. Compiling data on local roads cannot be done with as much accuracy in regard to location as with state roads. Finding an intersection between 2 local roads can be achieved by finding all crashes that happen on each road separately and combining the subset. Be sure to narrow down your search to only a specific municipality (see municipal code table). Note: Crashes involving local roads prior to 2008 will often be missing Latitude and Longintude due to limited GIS capabilities.

Finding crashes for a portion of a local road can be problematic since there is no systematic method used for determining the distinct crash location. If latitude and longitude data are available for crashes on a specific road, crashes can be filtered by coordinate range.

Local roads that are delineated with a direction can be very inconsistent, especially for older data. Example MAIN ST might be N MAIN ST, or NORTH MAIN ST.

Another problem is that street names are not always consistent. Though the majority of streets are kept consistent there are cases where streets are referred to by multiple names, commonly misspelled, have a space or hyphen in the name that aren't used consistently, or are referred to by both a local name and township/county route number. You may also see the street ending twice such as MAIN STREET ST

Interchanges and Ramps

Ramps at interchanges are a separate category of state roads and are normally designated with a state route designation between 8001 and 8999. These route numbers are distinct for each interchange within each county (virtually every county has a state route 8001). Each of the ramps within one interchange has the same SR number, but each ramp will have a different segment number. The segments are generally assigned in a clock wise manner.

Crashes that occur at interchanges that involve a ramp will either be a midblock crash on the ramp or an intersection between the ramp and another roadway.

Determining location ranges:

When PennDOT is asked to produce a crash history for a specific location we generally follow certain guidelines.

-  For intersections, we include all crashes that occur within 100 feet of the intersection.
-  For interchanges, we include all crashes on all ramps and on the state route within 100 feet of those ramps.
-  Point A to point B requests are done for each route separately when multiple routes are involved (common where null segments exist, or when a named local street has multiple state route designations).
-  For those that involve interchanges (such as Interstate X from Exit 10 to Exit 15) the report is typically run to include 100 feet beyond all entrance and exit ramps. For multiple sections (i.e. Interstate X from Exit 10 to Exit 11 and Exit 11 to Exit 12), the delineation between the two requests is where the main roads cross each other.
-  When a request spans a county line, each county is run separately.

Examples:

Question: How many fatal crashes were attributed to aggressive driving last summer?

Let's break down this question with the checklist we mentioned earlier.

What exactly do I want to count or lookup?

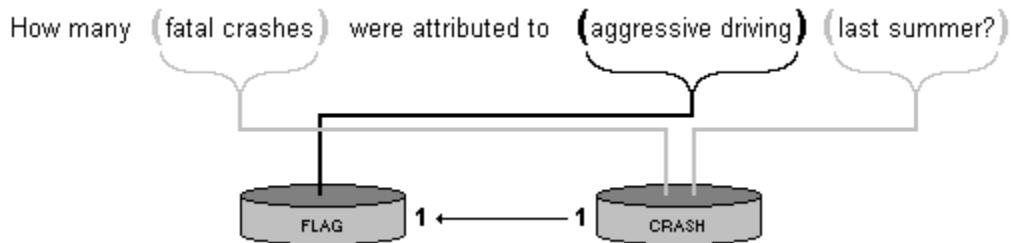
We want to count fatal crashes not fatalities. We don't need to know how many people died, just crashes where at least one person was killed.

What tables contain the data I need?

AGGRESSIVE_DRIVING is located in the FLAG table.
CRASH_DATE and FATAL_COUNT are located in table CRASH.

How do these tables relate to one another?

Every Crash record could have one related Flag record. The Flag record contains a series of Yes/No classifications that the crash falls under. According to the diagram shown here we have established our elements and the tables where they come from and how those tables are related.



Do I want to limit my results with any specific criteria?

For fatal crashes we need field FATAL_COUNT from CRASH table where FATAL_COUNT is greater than zero.

Last summer is very interesting as it is not specific enough. Technically it's June 21st thru September 21st, but to some people it may mean Memorial Day thru Labor Day, to others it may be June, July and August. In this case contact the requestor and verify the dates required. In our example let's say the data requested turned out to be June, July, and August, and our criteria would be:

Field CRASH_DATE from CRASH table greater than 20050531 and less than 20050901. Note: *crash date is stored as a text value in yyyyymmdd format.*

For Aggressive Driving we need field AGGRESSIVE_DRIVING from FLAG table where AGGRESSIVE_DRIVING = 1.

How do I want to sort my records?

We may want to order our results by date, and we could put and ascending sort on our CRASH_DATE field as well.

When compiling the data for the new system the developers decided to create a flag table to make it easier to query. It was appropriately named FLAG, and it basically sets a yes/no value to designate when

certain criteria was met for a specific item. The yes/no's are stored in table FLAG as one's and zero's. A one denotes a "Yes" value and a zero denotes a "No" value.

Note: The flag table is a great place to begin looking for the items that you want to include or exclude in your record set.

The UNIT_TYPE field found in table VEHICLE classifies the type of unit involved in the incident. In our drinking driver filter we find the values of 01,03,05,06 and this is filtering for only motor vehicles in transit and would exclude pedestrians and bicyclists.. *Note: When dealing with traffic studies you will always want to use these four codes as a filter for your data.*

Code	Description
01	Motor vehicle in transport
02	Legally parked
03	Illegally parked
05	Hit and run vehicle
06	Disabled from a previous crash
11	Non-motorized
21	Train
31	Pedestrian
32	Pedestrian on skates/ in wheelchair/etc.
51	Phantom vehicle

For our next example:

List all DUI crashes on New Years Eve and New Years Day for 04/05 by police agency?

What exactly do I want to count or lookup?

Crashes that involved a Drinking Driver on 12/31/04 and 01/01/05.

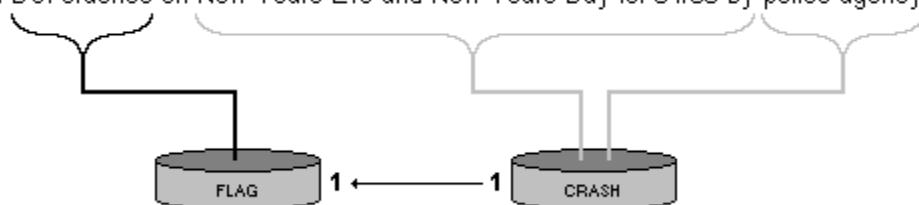
What tables contain the data I need?

DRINKING_DRIVER is located in the FLAG table.

CRASH_DATE and POLICE_AGCY are located in table CRASH.

How do these tables relate to one another?

List all DUI crashes on New Years Eve and New Years Day for 04/05 by police agency?



Do I want to limit my results with any specific criteria?

Field CRASH_DATE from CRASH table = 20041231 or CRASH_DATE from CRASH table = 20050101.

For DUI crashes we need field DRINKING_DRIVER from FLAG table where DRINKING_DRIVER = 1.

How do I want to sort my records?

We want to order our results by POLICE_AGCY

Sample Queries:

Copies of these queries are bundled with the data tables you will be receiving in MSAccess and the names of the queries will be listed with each example shown here. Feel free to edit and modify them to fit your needs.

The following query will generate a detailed output for crashes that occurred in a work zone and displays the counts of fatal injuries and total injuries sorted by municipality.

```
SELECT CRASH.CRASH_YEAR, FLAG.WORK_ZONE, CRASH.CRN, CRASH.MUNICIPALITY, CRASH.WORK_ZONE_LOC,  
CRASH.FATAL_COUNT, CRASH.TOT_INJ_COUNT  
FROM CRASH INNER JOIN FLAG ON CRASH.CRN = FLAG.CRN  
WHERE ((FLAG.WORK_ZONE)=1))  
ORDER BY CRASH.MUNICIPALITY;
```

The following query will generate a detailed output for 16-18 year old drivers. Note: PCR Unit type is filtering for only motor vehicles in transit and would exclude pedestrians and bicyclists.

```
SELECT PERSON.AGE, CRASH.CRN, CRASH.COUNTY, PERSON.INJ_SEVERITY, VEHICLE.VEH_TYPE, PERSON.UNIT_NUM  
FROM (CRASH INNER JOIN VEHICLE ON CRASH.CRN = VEHICLE.CRN) INNER JOIN PERSON ON (VEHICLE.UNIT_NUM =  
PERSON.UNIT_NUM) AND (VEHICLE.CRN = PERSON.CRN)  
WHERE (((PERSON.AGE)>=16 And (PERSON.AGE)<=18) AND ((VEHICLE.UNIT_TYPE)='01' Or (VEHICLE.UNIT_TYPE)='03' Or  
(VEHICLE.UNIT_TYPE)='05' Or (VEHICLE.UNIT_TYPE)='06') AND ((PERSON.PERSON_TYPE)='1'));
```

The following query will generate a detailed output for crashes on Downyflake Lane in the city of Allentown. The municipality was used to filter for Allentown, and Like 'DOWN' was used to filter the street name. This was done as some officers filing the report may have misspelled Downyflake Lane so this brings in all occurrences of DOWN with any other characters following it.

```
SELECT CRASH.CRN, CRASH.CRASH_YEAR, CRASH.ROAD_CONDITION, CRASH.ILLUMINATION, CRASH.WEATHER,  
CRASH.DAY_OF_WEEK, CRASH.HOUR_OF_DAY, CRASH.COLLISION_TYPE, CRASH.MAX_SEVERITY_LEVEL,  
CRASH.CRASH_MONTH, VEHICLE.VEH_TYPE, CRASH.FATAL_COUNT, CRASH.MAJ_INJ_COUNT, CRASH.MOD_INJ_COUNT,  
CRASH.MIN_INJ_COUNT, CRASH.UNK_INJ_DEG_COUNT, CRASH.UNK_INJ_PER_COUNT, CRASH.TOT_INJ_COUNT,  
DRIVER_ACTIONS.DVR_ACTION, ENV_FACTORS.ENV_RD_FACTOR, VEHICLE.UNIT_NUM, ROADWAY.STREET_NAME,  
CRASH.MUNICIPALITY  
FROM ((CRASH INNER JOIN DRIVER_ACTIONS ON CRASH.CRN = DRIVER_ACTIONS.CRN) INNER JOIN ENV_FACTORS ON  
CRASH.CRN = ENV_FACTORS.CRN) INNER JOIN ROADWAY ON CRASH.CRN = ROADWAY.CRN) INNER JOIN VEHICLE ON  
CRASH.CRN = VEHICLE.CRN  
WHERE (((ROADWAY.STREET_NAME) Like 'DOWN*') AND ((CRASH.MUNICIPALITY)=39301));
```

The following query will generate a detailed output for all aggressive driving related crashes in Lehigh County. The area of interest will be on state route 0145(P) between segment 40 offset 0 and segment 0082 offset 2185 or state route 0145(S) between segment 41 offset 0 and segment 0083 offset 2185.

Another item of interest is that Segment was concatenated with Offset, which was formatted to show 4 digits and to supplant zeros in place of blanks. That is shown in Italics and underlined.

```
SELECT CRASH.CRASH_DATE, CRASH.DISTRICT, CRASH.COUNTY, ROADWAY.ROUTE, ROADWAY.ADJ_RDWY_SEQ,  
ROADWAY.SEGMENT, ROADWAY.OFFSET, CRASH.COLLISION_TYPE, CRASH.INTERSECT_TYPE, CRASH.URBAN_RURAL,  
CRASH.FATAL_COUNT, CRASH.MAX_SEVERITY_LEVEL, CRASH.PERSON_COUNT, CRASH.VEHICLE_COUNT,  
CRASH.TOT_INJ_COUNT, CRASH.ILLUMINATION, CRASH.WEATHER, CRASH.ROAD_CONDITION, CRASH.MHE,  
CRASH.CRASH_YEAR, CRASH.CRN  
FROM (CRASH INNER JOIN FLAG ON CRASH.CRN = FLAG.CRN) INNER JOIN ROADWAY ON CRASH.CRN = ROADWAY.CRN  
WHERE (((CRASH.COUNTY)=39) AND ((ROADWAY.ROUTE)="0145") AND ((ROADWAY.ADJ_RDWY_SEQ)=3) AND  
((FLAG.STATE_ROAD)=1) AND ((FLAG.AGGRESSIVE_DRIVING)=1) AND (([SEGMENT] & Format([OFFSET],"0000"))>="00400000"  
And ([SEGMENT] & Format([OFFSET],"0000"))<="00839999"))  
ORDER BY CRASH.COUNTY, ROADWAY.ROUTE, ROADWAY.SEGMENT, ROADWAY.OFFSET, CRASH.CRASH_DATE;
```

The last example generates its output similar to a spreadsheet. This type of query is commonly referred to as a cross tab query. It follows the same structure as a SELECT statement and its clauses but you will notice it has 2 additional clauses (TRANSFORM and PIVOT) at the beginning and the end respectively. Some other items to note with cross tab queries is you must specify at least one field as your row header, one as your column header and at least one field to count occurrences of. The example shown here counts all persons that were killed while riding in an automobile, small truck, van or SUV, and were not wearing a restraining device categorized by month and illumination type.

```
TRANSFORM Count(PERSON.CRN) AS CountOfCRN
SELECT CRASH.ILLUMINATION
FROM (CRASH INNER JOIN VEHICLE ON CRASH.CRN = VEHICLE.CRN) INNER JOIN PERSON ON (VEHICLE.UNIT_NUM = PERSON.UNIT_NUM) AND (VEHICLE.CRN = PERSON.CRN)
WHERE (((VEHICLE.VEH_TYPE)="01" Or (VEHICLE.VEH_TYPE)="04" Or (VEHICLE.VEH_TYPE)="06" Or
(VEHICLE.VEH_TYPE)="07") AND ((PERSON.INJ_SEVERITY)="1") AND ((PERSON.RESTRAINT_HELMET)="00"))
GROUP BY CRASH.ILLUMINATION
PIVOT CRASH.CRASH_MONTH;
```
