UK Road Safety Data

Important Information

The UKRoadSafetyData.zip was provided to me to use as test data for analysis practice and demonstration.

Link to dataset -

Tools

- Microsoft SQL Server Management Studio
- Microsoft Visual Studio 2022
- Microsoft Excel (Power Query, CSVs)
- Posit RStudio
- Tableau Public

Languages

- SQL
- R

Dataset First Impressions

- Upon receiving and extracting the 'UKRoadSafetyData.zip' file, I notice there are additional zip files.
- Within each zip file is a single .csv or .xls file.
- Several of these files are missing naming conventions that match other similar files.
  - Renamed files for consistency if needed later.
- The 2015 data is in a single zip, and these are also not following later file naming conventions, and yet more 2015 files exist with apparently duplicate data.
  - Both 2015 Accidents files begin with '201501BS70001' in cell A2 and contain 140057 rows.
  - But without looking further, there is not guarantee these have exactly duplicate data. This will require review.
  - Checked through Power Query, matching both sets of data on for each duplicate (Accidents, Casualties, and Vehicles for 2015).
  - Both sets are a perfect Match. Only need to import one set.
- The Variable Lookup is a set of codes tables. One question is whether I will want to keep these as separate tables, or perform a single codes_table combination.
  - The codes tables can remain separate, but it would require more write scripts.
  - If I combine them into a single table, they could be more manageable on the import.
  - However, I may not need to import these into my database. They may only be needed while visualizing the findings, in which case I can wait until I import them to Tableau. TBD.

Extract, Transform, Load (ETL)

Visual Studio 2022 + SQL

- Created three tables uk_accidents uk_vehicles uk_casualties

uk_accidents

```sql
/*
I had a lot of trouble with the Date column
but finally manged to run through steps in Power Query and SQL
that worked.
*/
--DROP TABLE uk_accidents
--Date VARCHAR(15),
--Date DATE,

CREATE TABLE uk_accidents (
    Accident_Index VARCHAR(60),
    Location_Easting_OSGR INT,
    Location_Northing_OSGR INT,
    Longitude FLOAT,
    Latitude FLOAT,
    Police_Force INT,
    Accident_Severity INT,
    Number_of_Vehicles INT,
    Number_of_Casualties INT,
    Date VARCHAR(15),
    Day_of_Week INT,
    Time TIME,
    Local_Authority_District INT,
    Local_Authority_Highway VARCHAR(20),
    First_Road_Class INT,
    First_Road_Number INT,
    Road_Type INT,
    Speed_limit INT,
    Junction_Detail INT,
    Junction_Control INT,
    Second_Road_Class INT,
    Second_Road_Number INT,
    Pedestrian_Crossing_Human_Control INT,
    Pedestrian_Crossing_Physical_Facilities INT,
    Light_Conditions INT,
    Weather_Conditions INT,
    Road_Surface_Conditions INT,
    Special_Conditions_at_Site INT,
    Carriageway_Hazards INT,
    Urban_or_Rural_Area INT,
    Did_Police_Officer_Attend_Scene_of_Accident INT,
    LSOA_of_Accident_Location VARCHAR(60)
);
```

uk_casualties

```sql
--DROP TABLE uk_vehicles

CREATE TABLE uk_casualties (
    Accident_Index VARCHAR(20),
    Vehicle_Reference INT,
    Casualty_Reference INT,
    Casualty_Class INT,
    Sex_of_Casualty INT,
    Age_of_Casualty INT,
    Age_Band_of_Casualty INT,
    Casualty_Severity INT,
    Pedestrian_Location INT,
    Pedestrian_Movement INT,
    Car_Passenger INT,
    Bus_or_Coach_Passenger INT,
    Pedestrian_Road_Maintenance_Worker INT,
    Casualty_Type INT,
    Casualty_Home_Area_Type INT,
    Casualty_IMD_Decile INT,
);
```

uk_vehicles

```sql
--DROP TABLE uk_casualties

CREATE TABLE uk_vehicles (
    Accident_Index VARCHAR(20),
    Vehicle_Reference INT,
    Vehicle_Type INT,
    Towing_and_Articulation INT,
    Vehicle_Manoeuvre INT,
    Vehicle_Location_Restricted_Lane INT,
    Junction_Location INT,
    Skidding_and_Overturning INT,
    Hit_Object_in_Carriageway INT,
    Vehicle_Leaving_Carriageway INT,
    Hit_Object_off_Carriageway INT,
    First_Point_of_Impact INT,
    Was_Vehicle_Left_Hand_Drive INT,
    Journey_Purpose_of_Driver INT,
    Sex_of_Driver INT,
    Age_of_Driver INT,
    Age_Band_of_Driver INT,
    Engine_Capacity_CC INT,
    Propulsion_Code INT,
    Age_of_Vehicle INT,
    Driver_IMD_Decile INT,
    Driver_Home_Area_Type INT,
    Vehicle_IMD_Decile INT,
);
```

- Importing Accidents 2016 ran into an issue with Speed Limit column, where there were strings of NULL instead of true null values.

- Same issue with the 2017 Accidents in the Lat/Long columns.
- In Power Query, I ran a find/replace NULL for " " on the entire workbook.

- Accidents table: Date
  - Dates were formatted as text in dd/mm/yyyy format. When importing and converting, they were truncated. Niether Excel or SQL were playing nicely with this format.
  - In Power Query, I ran a transformation and re-saved the CSVs.
  - After some difficulty getting dates formatting, and a bit of help from Google and ChatGPT-4, the dates finally worked out.
  - The final CSV format was yyyy-mm-dd, then imported into my database as a VARCHAR
  - Then the final conversion was performed in SQL.

```
/*
I had a lot of trouble with the Date column
but finally manged to run through steps in Power Query and SQL
that worked.
*/
--DROP TABLE uk_accidents
--Date VARCHAR(15),
--Date DATE,

SELECT Date
FROM uk_accidents
WHERE ISDATE(Date) = 0


-- Update the Date column to swap the day and month
UPDATE uk_accidents
SET Date = CONCAT(
    SUBSTRING(Date, 1, 4), '-', -- Year
    SUBSTRING(Date, 9, 2), '-', -- Day
    SUBSTRING(Date, 6, 2)       -- Month
)
WHERE ISDATE(Date) = 0;


ALTER TABLE uk_accidents ADD Date_New DATE;

UPDATE uk_accidents SET Date_New = CAST(Date AS DATE);

--UPDATE uk_accidents SET Date_New = CONVERT(DATE, Date, 103);  -- 103 is for dd/mm/yyyy format

ALTER TABLE uk_accidents DROP COLUMN Date;

EXEC sp_rename 'uk_accidents.Date_New', 'Date', 'COLUMN';
```

- Checking for duplicates in the other tables revealed that while there are duplicate Accident_Index IDs, they are not duplicate observations. Each row is a unique observation.
  - Running the following script is an example of how they are not actually duplicates:

```sql
  WITH CTE_CheckDupsC AS (
    SELECT *
--      c.Accident_Index
--,     c.Casualty_Reference
,     ROW_NUMBER() OVER (PARTITION BY Accident_Index, Casualty_Reference ORDER BY (SELECT NULL)) AS rn
  FROM uk_casualties AS c
  )

  SELECT *
  FROM CTE_CheckDupsC
 WHERE Accident_Index = '201604ED16270'
-- WHERE rn > 1
  ORDER BY CTE_CheckDupsC.Accident_Index
```

Analysis Roadmap

1. Understand the Data: Make notes about each variable, what is it, what signficance does it hold, what potential calculations or analysis can be done on them.
2. Generate Hypotheses: What relationships, correlations, can we expect to find? Are accidents higher in certain regions, road conditions, etc? Are certain vehicles more likely to be in an accident?
3. Exploratory Data Analysis (EDA): Using SQL, R, and Tableau to analyze the datasets, we summarize the data and find patterns. Min, Max, Sum, Count, Mean, Median, etc. will be helpful.
4. Visualization: Once we have some key insights, start creating visualizations (first in R, then in Tableau) to see if anything visually stands out as interesting or if a story begins to emerge.
5. Story Drafting: Draft the narrative around these findings. What story is the data telling? Why is it important? What recommendations or observations can we make?
6. Review and Refine: Review the data story several times, refine it to make sure it's compelling and understandable. Are the visualizations telling the story on their own?
7. Presentation: Put all these findings into a presentation format. Keep it simple but informative. Use your visualizations to support your story. Keep the script out of the visualizations themselves but in a presentable format for those who are not able to attend the presentation.

Questions to Consider:

Conditions

- Roads: Wet/Dry, Urban/Rural, Maintained/Deferred-Maintenance,
- Cities, Regions, Police Depts, Deprivation Index, do any of these show higher accidents or lower?
- Are any types of crossings or conditions more likely to cause accidents?
- Do Seasons, Days, Times of Day, Light, or Weather play a role?

People

- Does Poverty index, age, sex, or any other human factor play a role in accidents?

Analysis

R & RStudio Transitioning to RStudio for Analysis

Connecting to R RStudio Loading Libraries

```
# LOAD LIBRARIES
library(DBI)
```

```
## Warning: package 'DBI' was built under R version 4.3.1
```

```
library(odbc)
```

```
## Warning: package 'odbc' was built under R version 4.3.1
```

```
library(RODBC)
```

```
## Warning: package 'RODBC' was built under R version 4.3.1
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.1
```

```
## Warning: package 'ggplot2' was built under R version 4.3.1
```

```
## Warning: package 'lubridate' was built under R version 4.3.1
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
#library(dplyr)
#library(ggplot2)
#library(scales)
library(sqldf)
```

```
## Warning: package 'sqldf' was built under R version 4.3.1
```

```
## Loading required package: gsubfn
```

```
## Warning: package 'gsubfn' was built under R version 4.3.1

## Loading required package: proto

## Warning: package 'proto' was built under R version 4.3.1

## Loading required package: RSQLite

## Warning: package 'RSQLite' was built under R version 4.3.1
```

```
#library(ggmap)
#library(geosphere)
#library(here)
#library(skimr)
#library(janitor)
#library(Tmisc)
```

Connecting R & RStudio to the database

```
# ESTABLISH CONNECTION TO MY LOCAL DATABASE

connection <- odbcDriverConnect("driver={SQL Server};server=LAPTOP-76LHVPRQ\\SQLEXPRESS;database=UK_Roa
```

Assign Dataframes Assigning the new cleaned database tables as R dataframes:

- UK_AccidentData
- UK_CasualtyData
- UK_VehicleData

```
# ASSIGN THE DATABASE TABLE AS A DATAFRAME (df) VARIABLE FOR EASIER RECALL

UK_AccidentData <- sqlFetch(connection, "dbo.uk_accidents")

UK_CasualtyData <- sqlFetch(connection, "dbo.uk_casualties")

UK_VehicleData <- sqlFetch(connection, "dbo.uk_vehicles")
```

Column Names (Variables)

- What variables do each dataset contain?
- How are they related?
- What might they tell us?

The Accident Index acts as a key. The Accident Table is the key table, there is one Accident Index for each observation (row) in this table. The Casualty and Vehicle tables refer back to this Accident Index key to tie their data into specific accidents. The Number of Vehicles and Number of Casualties are essentially a count of the other related data from the other two tables.

The Vehicle Reference builds on the Accident Index. It is contained in both the Vehicle Table and Casualty Table. For each accident, individual vehicles are numbered (Vehicle 1, Vehicle 2, etc.). This allows us to see how many vehicles were involved in any given accident (from the Vehicle Table) while maintaining a simple Accident Index in the origin table. The Casualty Table can then assign the casualty to which vehicle was involved in that injury.

The Casualty Reference only exists on the Casualty Table, and serves as an index for the individual who was injured.

There are also Key Tables/Codes Tables. During my work in R, I will reference them via CSV/Excel to understand what each one is, as I do not need them for determining averages or means or counts. However, when visualizing this data, I will import these into Tableau in a star schema for more accurate labeling.

```r
colnames(UK_AccidentData)
```

```
##  [1] "Accident_Index"
##  [2] "Location_Easting_OSGR"
##  [3] "Location_Northing_OSGR"
##  [4] "Longitude"
##  [5] "Latitude"
##  [6] "Police_Force"
##  [7] "Accident_Severity"
##  [8] "Number_of_Vehicles"
##  [9] "Number_of_Casualties"
## [10] "Day_of_Week"
## [11] "Time"
## [12] "Local_Authority_District"
## [13] "Local_Authority_Highway"
## [14] "First_Road_Class"
## [15] "First_Road_Number"
## [16] "Road_Type"
## [17] "Speed_limit"
## [18] "Junction_Detail"
## [19] "Junction_Control"
## [20] "Second_Road_Class"
## [21] "Second_Road_Number"
## [22] "Pedestrian_Crossing_Human_Control"
## [23] "Pedestrian_Crossing_Physical_Facilities"
## [24] "Light_Conditions"
## [25] "Weather_Conditions"
## [26] "Road_Surface_Conditions"
## [27] "Special_Conditions_at_Site"
## [28] "Carriageway_Hazards"
## [29] "Urban_or_Rural_Area"
## [30] "Did_Police_Officer_Attend_Scene_of_Accident"
## [31] "LSOA_of_Accident_Location"
## [32] "Date"
## [33] "TimeGroup"
```

```r
colnames(UK_CasualtyData)
```

```
##  [1] "Accident_Index"                     "Vehicle_Reference"
##  [3] "Casualty_Reference"                 "Casualty_Class"
##  [5] "Sex_of_Casualty"                    "Age_of_Casualty"
##  [7] "Age_Band_of_Casualty"               "Casualty_Severity"
##  [9] "Pedestrian_Location"                "Pedestrian_Movement"
## [11] "Car_Passenger"                      "Bus_or_Coach_Passenger"
## [13] "Pedestrian_Road_Maintenance_Worker" "Casualty_Type"
## [15] "Casualty_Home_Area_Type"            "Casualty_IMD_Decile"
```

```r
colnames(UK_VehicleData)
```

```
##  [1] "Accident_Index"                   "Vehicle_Reference"
##  [3] "Vehicle_Type"                     "Towing_and_Articulation"
##  [5] "Vehicle_Manoeuvre"                "Vehicle_Location_Restricted_Lane"
##  [7] "Junction_Location"                "Skidding_and_Overturning"
##  [9] "Hit_Object_in_Carriageway"        "Vehicle_Leaving_Carriageway"
## [11] "Hit_Object_off_Carriageway"       "First_Point_of_Impact"
## [13] "Was_Vehicle_Left_Hand_Drive"      "Journey_Purpose_of_Driver"
## [15] "Sex_of_Driver"                    "Age_of_Driver"
## [17] "Age_Band_of_Driver"               "Engine_Capacity_CC"
## [19] "Propulsion_Code"                  "Age_of_Vehicle"
## [21] "Driver_IMD_Decile"                "Driver_Home_Area_Type"
## [23] "Vehicle_IMD_Decile"
```

UK_Road Safety Data Summary In comparing certain variable's mean and median, we can make some preliminary statements.

Summary of UK Accidents

- Accident Severity Code 3 is Slight. The Median (3) and Mean (2.81) indicate that the average accident only presents Slight injuries.
- Most accidents involve two vehicles and do involve at least one injury.
- Day of Week: I'd like to see a visual count on this for the "difference" but most accidents occur on a Wednesday (Median 4, Mean 4.104, Code 4 is Wednesday).
- The Mean (37.64) and Median (30) speed limits may indicate the "average" accident occurs at these lower limits. However, a count graphic may indicate whether there was a slight difference or major difference in accidents counts for each limit range.
- Light Conditions. Code 1 is light, Codes 4-7 indicate various lighting conditions in the dark. The Median (1) and Mean (1.993) may indicate the average accident occurs during the day or with good lighting. Graphing this data may help illuminate this key.
- Weather Conditions. Code 1 is Fine. Codes 2-8 are other than fine. 9 is unknown. Median (1) and Mean (1.579) indicate weather is not a major influence "on average", but graphing this will show any trends or difference in the data.

- Police Attendance would be an after-the-fact variable, so while it may indicate the seriousness of the accident, it would not be predictive in any way for preventing accidents. That being said, it does appear the police were involved in the average accident, Median (1), Mean (1.251), with Code 1 being Yes, and Code 2 being No.

Summary of UK Acc*ident Casualty*

Note: By "Casualty", this dataset means "Injury" not necessarily "Death". They are segregated into Fatality, Serious, and Slight.

- Sex of Casualty Code 1, Male, Code 2, Female. Median (1), Mean (1.406); this does not tell me much right now. I'd like to see the data visualized and see if this acts as a component against or with other variables.
- Age of Casualty: Median 33, Mean 36.48. This does seem to indicate something. Although, this is the prime working age, so it's possible this has to do with the number of humans in this age range in the vicinity and not anything to do with their driving. High probability with higher representation. As with the other table, when a casualty is involved, it appears to be serious or slight, but Fatalities are not the average. Although, with the Insurance rates higher for "teens" I would have expected this to trend that direction. I would like to see the full variable visualized.
- Pedestrian Location & Movement reflect Median of 0, and Mean of 0.714/0.5418. Pedestrians are rarely involved.
- Road Maintenance workers are rarely involved. Although, if a road working company wanted to analyze the data for situations in which they could reduce their workers being hurt, they could slice the data for only these situations and see what trends emerge.
- Home Area (1 Urban, 2 Small Town, 3 Rural) - Urban cities seem to be the average place for accidents.
- IMD 4 / 3.714 may indicate a trend toward more casualties in mid-lower income regions? Needs further visualization analysis.

Summary of UK Vehicles in Accidents

- Vehicle Type 9/9.804 is Car. Most people on most trips are in a car, so this holds to the expected pattern. Could be interesting to visualize this data variable to see what the second or third most common types are. Motorcycle, Van, Taxi, etc.
- !!! Manoeuvre (Reversing, Parked, Waiting, Slowing, U-Turn, etc.). Anecdotally, when I was a young driver most of my accidents were in parking lots while reversing. I later learned this is common for ADHD neurotypes. This may be interesting to follow.
- Vehicle Location Median 0 means the central data point is on the main roads. Not surprising.
- Junction Median of 1 (Approaching or waiting at a junction) may indicate this is a common place for accidents. I was rear-ended more than once while waiting at a stoplight.
- Journey or Purpose Unknown (6) may make this a useless variable?
- Several variables show no indication of relevance, all being zero median.
- Age of Driver (35/35.49) matches the Casualty table. Interesting.
- Propulsion Code - Mostly Gasoline/Petrol cars. Not shocking.

- Age of Vehicle - Average car was 5.569 years old. That's may be a reflection of the average age of cars in general, and may only be relevant to the accident data if this was compared to the general average age of cars in the UK. If they were divergent, then it would be relevant. Otherwise it is a reflection and not indicative of anything useful.
- Engine Capacity CC - Median (1390), Mean (1422) actually mean nothing to me because I don't know anything about Engine CCs or how that's relevant. And as with Age of Vehicle, may be a reflection and not an indication.

```r
summary(UK_AccidentData)
```

```
##  Accident_Index     Location_Easting_OSGR Location_Northing_OSGR
##  Length:529294      Min.   : 70860        Min.   :  10235
##  Class :character   1st Qu.:386552        1st Qu.: 176240
##  Mode  :character   Median :455233        Median : 231908
##                     Mean   :449456        Mean   : 286296
##                     3rd Qu.:528110        3rd Qu.: 389580
##                     Max.   :655391        Max.   :1209512
##                     NA's   :108           NA's   :108
##    Longitude          Latitude      Police_Force   Accident_Severity
##  Min.   :-7.4229   Min.   :49.91   Min.   : 1.00   Min.   :1.00
##  1st Qu.:-2.2016   1st Qu.:51.47   1st Qu.: 6.00   1st Qu.:3.00
##  Median :-1.1849   Median :51.97   Median :30.00   Median :3.00
##  Mean   :-1.2934   Mean   :52.46   Mean   :29.19   Mean   :2.81
##  3rd Qu.:-0.1533   3rd Qu.:53.40   3rd Qu.:45.00   3rd Qu.:3.00
##  Max.   : 1.7596   Max.   :60.76   Max.   :98.00   Max.   :3.00
##  NA's   :118       NA's   :118
##  Number_of_Vehicles Number_of_Casualties  Day_of_Week        Time
##  Min.   : 1.000     Min.   : 1.000       Min.   :1.000   Length:529294
##  1st Qu.: 1.000     1st Qu.: 1.000       1st Qu.:2.000   Class :character
##  Median : 2.000     Median : 1.000       Median :4.000   Mode  :character
##  Mean   : 1.843     Mean   : 1.321       Mean   :4.104
##  3rd Qu.: 2.000     3rd Qu.: 1.000       3rd Qu.:6.000
##  Max.   :37.000     Max.   :59.000       Max.   :7.000
##
##  Local_Authority_District Local_Authority_Highway First_Road_Class
##  Min.   :  1              Length:529294           Min.   :1.00
##  1st Qu.: 95              Class :character        1st Qu.:3.00
##  Median :307              Mode  :character        Median :4.00
##  Mean   :336                                      Mean   :4.17
##  3rd Qu.:514                                      3rd Qu.:6.00
##  Max.   :941                                      Max.   :6.00
##
##  First_Road_Number   Road_Type       Speed_limit    Junction_Detail
##  Min.   :   0.0    Min.   :-1.000   Min.   : 0.00   Min.   :-1.000
##  1st Qu.:   0.0    1st Qu.: 6.000   1st Qu.:30.00   1st Qu.: 0.000
##  Median :  60.0    Median : 6.000   Median :30.00   Median : 1.000
##  Mean   : 900.5    Mean   : 5.175   Mean   :37.64   Mean   : 2.252
##  3rd Qu.: 630.0    3rd Qu.: 6.000   3rd Qu.:40.00   3rd Qu.: 3.000
##  Max.   :9918.0    Max.   : 9.000   Max.   :70.00   Max.   : 9.000
##                                     NA's   :37
##  Junction_Control Second_Road_Class Second_Road_Number
##  Min.   :-1.000   Min.   :-1.000    Min.   :  -1.0
##  1st Qu.:-1.000   1st Qu.:-1.000    1st Qu.:   0.0
```

```
##  Median : 2.000   Median : 3.000    Median :   0.0
##  Mean   : 1.657   Mean   : 2.647    Mean   : 320.6
##  3rd Qu.: 4.000   3rd Qu.: 6.000    3rd Qu.:   0.0
##  Max.   : 4.000   Max.   : 6.000    Max.   :9999.0
##
##  Pedestrian_Crossing_Human_Control Pedestrian_Crossing_Physical_Facilities
##  Min.   :-1.000000                 Min.   :-1.0000
##  1st Qu.: 0.000000                 1st Qu.: 0.0000
##  Median : 0.000000                 Median : 0.0000
##  Mean   : 0.004485                 Mean   : 0.8375
##  3rd Qu.: 0.000000                 3rd Qu.: 0.0000
##  Max.   : 2.000000                 Max.   : 8.0000
##
##  Light_Conditions Weather_Conditions Road_Surface_Conditions
##  Min.   :-1.000   Min.   :-1.000     Min.   :-1.000
##  1st Qu.: 1.000   1st Qu.: 1.000     1st Qu.: 1.000
##  Median : 1.000   Median : 1.000     Median : 1.000
##  Mean   : 1.993   Mean   : 1.579     Mean   : 1.286
##  3rd Qu.: 4.000   3rd Qu.: 1.000     3rd Qu.: 2.000
##  Max.   : 7.000   Max.   : 9.000     Max.   : 5.000
##
##  Special_Conditions_at_Site Carriageway_Hazards Urban_or_Rural_Area
##  Min.   :-1.00000           Min.   :-1.00000    Min.   :-1.000
##  1st Qu.: 0.00000           1st Qu.: 0.00000    1st Qu.: 1.000
##  Median : 0.00000           Median : 0.00000    Median : 1.000
##  Mean   : 0.08348           Mean   : 0.05296    Mean   : 1.338
##  3rd Qu.: 0.00000           3rd Qu.: 0.00000    3rd Qu.: 2.000
##  Max.   : 7.00000           Max.   : 7.00000    Max.   : 3.000
##
##  Did_Police_Officer_Attend_Scene_of_Accident LSOA_of_Accident_Location
##  Min.   :-1.000                              Length:529294
##  1st Qu.: 1.000                              Class :character
##  Median : 1.000                              Mode  :character
##  Mean   : 1.251
##  3rd Qu.: 1.000
##  Max.   : 3.000
##
##      Date             TimeGroup
##  Length:529294      Length:529294
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
##
```

```r
summary(UK_CasualtyData)
```

```
##  Accident_Index     Vehicle_Reference Casualty_Reference Casualty_Class
##  Length:699163      Min.   :  1.000   Min.   :  1.000    Min.   :1.00
##  Class :character   1st Qu.:  1.000   1st Qu.:  1.000    1st Qu.:1.00
##  Mode  :character   Median :  1.000   Median :  1.000    Median :1.00
##                     Mean   :  1.488   Mean   :  1.406    Mean   :1.49
##                     3rd Qu.:  2.000   3rd Qu.:  2.000    3rd Qu.:2.00
```

```
##                      Max.   :999.000   Max.   :991.000    Max.   :3.00
##  Sex_of_Casualty  Age_of_Casualty  Age_Band_of_Casualty Casualty_Severity
##  Min.   :-1.000   Min.   : -1.00   Min.   :-1.000       Min.   :1.000
##  1st Qu.: 1.000   1st Qu.: 22.00   1st Qu.: 5.000       1st Qu.:3.000
##  Median : 1.000   Median : 33.00   Median : 6.000       Median :3.000
##  Mean   : 1.406   Mean   : 36.48   Mean   : 6.289       Mean   :2.842
##  3rd Qu.: 2.000   3rd Qu.: 50.00   3rd Qu.: 8.000       3rd Qu.:3.000
##  Max.   : 2.000   Max.   :104.00   Max.   :11.000       Max.   :3.000
##  Pedestrian_Location Pedestrian_Movement Car_Passenger
##  Min.   :-1.000      Min.   :-1.0000     Min.   :-1.0000
##  1st Qu.: 0.000      1st Qu.: 0.0000     1st Qu.: 0.0000
##  Median : 0.000      Median : 0.0000     Median : 0.0000
##  Mean   : 0.714      Mean   : 0.5418     Mean   : 0.2523
##  3rd Qu.: 0.000      3rd Qu.: 0.0000     3rd Qu.: 0.0000
##  Max.   :10.000      Max.   : 9.0000     Max.   : 2.0000
##  Bus_or_Coach_Passenger Pedestrian_Road_Maintenance_Worker Casualty_Type
##  Min.   :-1.00000       Min.   :-1.00000                   Min.   :-1.000
##  1st Qu.: 0.00000       1st Qu.: 0.00000                   1st Qu.: 3.000
##  Median : 0.00000       Median : 0.00000                   Median : 9.000
##  Mean   : 0.07744       Mean   : 0.07626                   Mean   : 7.281
##  3rd Qu.: 0.00000       3rd Qu.: 0.00000                   3rd Qu.: 9.000
##  Max.   : 4.00000       Max.   : 2.00000                   Max.   :98.000
##  Casualty_Home_Area_Type Casualty_IMD_Decile
##  Min.   :-1.000          Min.   :-1.000
##  1st Qu.: 1.000          1st Qu.: 1.000
##  Median : 1.000          Median : 4.000
##  Mean   : 0.986          Mean   : 3.714
##  3rd Qu.: 1.000          3rd Qu.: 7.000
##  Max.   : 3.000          Max.   :10.000
```

```
summary(UK_VehicleData)
```

```
##  Accident_Index     Vehicle_Reference  Vehicle_Type    Towing_and_Articulation
##  Length:975680      Min.   :  1.000   Min.   :-1.000   Min.   :-1.00000
##  Class :character   1st Qu.:  1.000   1st Qu.: 9.000   1st Qu.: 0.00000
##  Mode  :character   Median :  1.000   Median : 9.000   Median : 0.00000
##                     Mean   :  1.562   Mean   : 9.804   Mean   : 0.02147
##                     3rd Qu.:  2.000   3rd Qu.: 9.000   3rd Qu.: 0.00000
##                     Max.   :999.000   Max.   :98.000   Max.   : 5.00000
##  Vehicle_Manoeuvre Vehicle_Location_Restricted_Lane Junction_Location
##  Min.   :-1.00     Min.   :-1.0000                  Min.   :-1.000
##  1st Qu.: 6.00     1st Qu.: 0.0000                  1st Qu.: 0.000
##  Median :17.00     Median : 0.0000                  Median : 1.000
##  Mean   :12.43     Mean   : 0.1404                  Mean   : 2.441
##  3rd Qu.:18.00     3rd Qu.: 0.0000                  3rd Qu.: 5.000
##  Max.   :18.00     Max.   : 9.0000                  Max.   : 8.000
##  Skidding_and_Overturning Hit_Object_in_Carriageway Vehicle_Leaving_Carriageway
##  Min.   :-1.0000          Min.   :-1.0000           Min.   :-1.0000
##  1st Qu.: 0.0000          1st Qu.: 0.0000           1st Qu.: 0.0000
##  Median : 0.0000          Median : 0.0000           Median : 0.0000
##  Mean   : 0.1827          Mean   : 0.2983           Mean   : 0.3094
##  3rd Qu.: 0.0000          3rd Qu.: 0.0000           3rd Qu.: 0.0000
##  Max.   : 5.0000          Max.   :12.0000           Max.   : 8.0000
##  Hit_Object_off_Carriageway First_Point_of_Impact Was_Vehicle_Left_Hand_Drive
```

```
##  Min.   :-1.0000           Min.   :-1.000        Min.   :-1.0000
##  1st Qu.: 0.0000           1st Qu.: 1.000        1st Qu.: 1.0000
##  Median : 0.0000           Median : 1.000        Median : 1.0000
##  Mean   : 0.4375           Mean   : 1.754        Mean   : 0.9695
##  3rd Qu.: 0.0000           3rd Qu.: 3.000        3rd Qu.: 1.0000
##  Max.   :11.0000           Max.   : 4.000        Max.   : 2.0000
##  Journey_Purpose_of_Driver Sex_of_Driver    Age_of_Driver    Age_Band_of_Driver
##  Min.   :-1.000           Min.   :-1.000   Min.   : -1.00   Min.   :-1.000
##  1st Qu.: 2.000           1st Qu.: 1.000   1st Qu.: 23.00   1st Qu.: 5.000
##  Median : 6.000           Median : 1.000   Median : 35.00   Median : 6.000
##  Mean   : 4.714           Mean   : 1.432   Mean   : 35.49   Mean   : 5.959
##  3rd Qu.: 6.000           3rd Qu.: 2.000   3rd Qu.: 50.00   3rd Qu.: 8.000
##  Max.   : 6.000           Max.   : 3.000   Max.   :101.00   Max.   :11.000
##  Engine_Capacity_CC Propulsion_Code  Age_of_Vehicle    Driver_IMD_Decile
##  Min.   :   -1      Min.   :-1.000   Min.   : -1.000   Min.   :-1.000
##  1st Qu.:  113      1st Qu.: 1.000   1st Qu.: -1.000   1st Qu.:-1.000
##  Median : 1390      Median : 1.000   Median :  5.000   Median : 1.000
##  Mean   : 1422      Mean   : 0.936   Mean   :  5.569   Mean   : 2.206
##  3rd Qu.: 1910      3rd Qu.: 2.000   3rd Qu.: 10.000   3rd Qu.: 5.000
##  Max.   :99999      Max.   :12.000   Max.   :105.000   Max.   :10.000
##  Driver_Home_Area_Type Vehicle_IMD_Decile
##  Min.   :-1.0000       Min.   :-1.000
##  1st Qu.: 1.0000       1st Qu.:-1.000
##  Median : 1.0000       Median : 1.000
##  Mean   : 0.8256       Mean   : 2.206
##  3rd Qu.: 1.0000       3rd Qu.: 5.000
##  Max.   : 3.0000       Max.   :10.000
```

New Questions

Based on the summary review, here are some questions to pursue.

- Accidents Table
  - How many accidents involve 0, 1, 2, or more vehicles?
  - How many accidents involve a casualty/injury of any kind?
  - How many accidents occur at various speed limit ranges?
  - How many accidents occur on which days of the week?
  - How many accidents occur at various light conditions?
  - How many accidents occur at various weather conditions?
  - How many accidents for each IMD?
- Casualty Table
  - How many casualties for each sex/gender?
  - How many casualties for each age band?
  - How many casualties for each Home Areas?
  - How many casualties for each IMD?
- Vehicles Table
  - How many of each vehicle type? We know cars are most common, but which are 2nd or 3rd?
  - Which Manoeuvres are most common?
  - Besides main roads, which other Locations are common?

– Explore variables, see what else sticks out.

UK_AccidentData Variable Analysis

The following represents a look at the counts for occurrences of variables in the UK_AccidentData dataset. Comments above each represent findings.

```
# As expected, most accidents involved two vehicles, then one, then three.
# Numbers drop off significantly after this.
UK_AccidentData %>%
  count(Number_of_Vehicles)
```

```
##    Number_of_Vehicles      n
## 1                   1 155544
## 2                   2 319053
## 3                   3  42056
## 4                   4   9372
## 5                   5   2182
## 6                   6    637
## 7                   7    256
## 8                   8     87
## 9                   9     51
## 10                 10     27
## 11                 11      9
## 12                 12      4
## 13                 13      3
## 14                 14      3
## 15                 15      1
## 16                 16      5
## 17                 18      1
## 18                 23      1
## 19                 24      1
## 20                 37      1
```

```
# At first glance, this doesn't appear to say much. Weekends are fewer.
UK_AccidentData %>%
  count(Day_of_Week)
```

```
##   Day_of_Week     n
## 1           1 59060
## 2           2 74984
## 3           3 78967
## 4           4 79955
## 5           5 80881
## 6           6 86219
## 7           7 69228
```

```
# There are a LOT more accidents at 30 MPH than at any other speed limit.
UK_AccidentData %>%
  count(Speed_limit)
```

```
##   Speed_limit      n
## 1           0      1
## 2          10      2
## 3          20  31333
## 4          30 327671
## 5          40  44098
## 6          50  21652
## 7          60  69574
## 8          70  34926
## 9          NA     37
```

```
# The vast majority of accidents occur during the day (when most are driving)
# Next is darkness in well lit aready (again, where most people are driving)
UK_AccidentData %>%
  count(Light_Conditions)
```

```
##   Light_Conditions      n
## 1               -1     14
## 2                1 382690
## 3                4 106713
## 4                5   3362
## 5                6  27075
## 6                7   9440
```

```
# The lion share of accidents occur when weather is fine (when most people are driving)
# Of the remaining, the most occur in rain.
UK_AccidentData %>%
  count(Weather_Conditions)
```

```
##    Weather_Conditions      n
## 1                  -1     33
## 2                   1 430121
## 3                   2  57140
## 4                   3   2602
## 5                   4   5990
## 6                   5   6130
## 7                   6    755
## 8                   7   2480
## 9                   8   9507
## 10                  9  14536
```

```
# Not useful, needs better grouping.
#UK_AccidentData %>%
#  count(Time)

# This will be very useful for Visualization in Tableau,
# added this new column to table in SQL

UK_AccidentData %>%
  mutate(TimeGroup = case_when(
    Time >= "06:00:00.0000000" & Time < "9:00:00.0000000" ~ "Morning_Commute",
    Time >= "09:00:00.0000000" & Time < "11:00:00.0000000" ~ "Morning_Late",
```

```
    Time >= "11:00:00.0000000" & Time < "13:00:00.0000000" ~ "Lunch_Hours",
    Time >= "13:00:00.0000000" & Time < "17:00:00.0000000" ~ "Afternoon",
    Time >= "17:00:00.0000000" & Time < "20:00:00.0000000" ~ "Evening_Commute",
    Time >= "20:00:00.0000000" & Time < "21:00:00.0000000" ~ "Evening_Late",
    TRUE ~ "Late_Night"
  )) %>%
  count(TimeGroup)
```

```
##         TimeGroup      n
## 1      Late_Night  28984
## 2 Morning_Commute 500310
```

```
# Most accidents occur in the Afternoon (1p-5p)
# Second most occur during the evening commute (5p-8p)

UK_AccidentData %>%
  count(TimeGroup)
```

```
##         TimeGroup      n
## 1       Afternoon 147299
## 2 Evening_Commute 112939
## 3    Evening_Late  20148
## 4      Late_Night  69267
## 5     Lunch_Hours  57138
## 6 Morning_Commute  73060
## 7    Morning_Late  49443
```

```
# The following three seem evenly dispersed.
# Maybe let's get a dashboard with these?
# See if anything pops out visually?

# Seems fairly evenly dispersed?
UK_AccidentData %>%
  count(Police_Force) %>%
  arrange(-n) %>%
  head(10)
```

```
##    Police_Force      n
## 1             1 101812
## 2            20  23173
## 3            46  19116
## 4            13  18757
## 5            43  18022
## 6            44  16770
## 7            47  16216
## 8            50  14458
## 9            45  14215
## 10            4  14120
```

```
# Seems fairly evenly dispersed?
```

```r
UK_AccidentData %>%
  count(Local_Authority_District) %>%
  arrange(-n) %>%
  head(10)
```

```
##    Local_Authority_District     n
## 1                       300 11095
## 2                       204  7164
## 3                         1  6319
## 4                         9  5114
## 5                       596  4759
## 6                       926  4456
## 7                         5  4405
## 8                       200  4375
## 9                       215  4375
## 10                        8  4160
```

```r
# Seems fairly evenly dispersed?
UK_AccidentData %>%
  count(Local_Authority_Highway) %>%
  arrange(-n) %>%
  head(10)
```

```
##    Local_Authority_Highway     n
## 1                E10000016 16625
## 2                E10000030 14215
## 3                E10000012 11190
## 4                E10000014 11130
## 5                E08000025 11095
## 6                E10000017 11089
## 7                E10000015  9447
## 8                E10000019  7912
## 9                E10000032  7862
## 10               E08000035  7164
```

```r
# No Intersection & T Intersections are most common accident types.
# That is surprising, given the MPH 30 thing? Let's look at MPH, Junctions, and Time.

UK_AccidentData %>%
  count(Junction_Detail)
```

```
##    Junction_Detail      n
## 1               -1   1488
## 2                0 218682
## 3                1  44888
## 4                2   6698
## 5                3 162272
## 6                5   7828
## 7                6  50729
## 8                7   5077
## 9                8  15585
## 10               9  16047
```

```r
# Mostly missing data (-1)
UK_AccidentData %>%
  count(Junction_Control)
```

```
##   Junction_Control      n
## 1               -1 223831
## 2                0    388
## 3                1   1375
## 4                2  56285
## 5                3   2945
## 6                4 244470
```

```r
# Mostly missing data (-1), filter these out.
# With (-1, missing data) filtered out, these show a strong preference for 4.
# Uncontroled intersections lead the way.
UK_AccidentData %>%
  filter(Junction_Control != -1) %>%
  count(Junction_Control)
```

```
##   Junction_Control      n
## 1                0    388
## 2                1   1375
## 3                2  56285
## 4                3   2945
## 5                4 244470
```

```r
# One lane roads dominate
UK_AccidentData %>%
  count(Road_Type)
```

```
##   Road_Type      n
## 1        -1      1
## 2         1  34444
## 3         2  12742
## 4         3  80196
## 5         6 388887
## 6         7   5875
## 7         9   7149
```

```r
# Most accidents (383,735) occurred on dry roads.
# 2nd Most (131,236) on wet roads.
# As these are the two most common road conditions, that is only so helpful to know.
UK_AccidentData %>%
  count(Road_Surface_Conditions)
```

```
##   Road_Surface_Conditions      n
## 1                      -1   4213
## 2                       1 383735
## 3                       2 131236
## 4                       3   2325
## 5                       4   7205
## 6                       5    580
```

```r
# Roughly 3/4 of accidents had a Police Officer
# Roughly 1/4 did not
# A fraction had no police officer but did self-report via a form.
UK_AccidentData %>%
  count(Did_Police_Officer_Attend_Scene_of_Accident)
```

```
##   Did_Police_Officer_Attend_Scene_of_Accident      n
## 1                                           -1     10
## 2                                            1 399445
## 3                                            2 126704
## 4                                            3   3135
```

UK_CasualtyData Variable Analysis

The following represents a look at the counts for occurrences of variables in the UK_CasualtyData dataset. Comments above each represent findings.

```r
# UK_CasualtyData

colnames(UK_CasualtyData)
```

```
##  [1] "Accident_Index"                     "Vehicle_Reference"
##  [3] "Casualty_Reference"                 "Casualty_Class"
##  [5] "Sex_of_Casualty"                    "Age_of_Casualty"
##  [7] "Age_Band_of_Casualty"               "Casualty_Severity"
##  [9] "Pedestrian_Location"                "Pedestrian_Movement"
## [11] "Car_Passenger"                      "Bus_or_Coach_Passenger"
## [13] "Pedestrian_Road_Maintenance_Worker" "Casualty_Type"
## [15] "Casualty_Home_Area_Type"            "Casualty_IMD_Decile"
```

```r
# Roughly double M vs F
UK_CasualtyData %>%
  count(Sex_of_Casualty)
```

```
##   Sex_of_Casualty      n
## 1              -1    180
## 2               1 414486
## 3               2 284497
```

```r
# The data set is too long, using age bands is better.
# Unlike the Time in the Accident table, this one is already grouped.
#UK_CasualtyData %>%
 # count(Age_of_Casualty)

# Probably a bell curve?
# Highest at 6 (26-35) and 7 (36-45) with tapering on both sides.
# Probably a reflection of the age average among the general population.
UK_CasualtyData %>%
  count(Age_Band_of_Casualty)
```

```
##    Age_Band_of_Casualty      n
## 1                    -1  10942
## 2                     1  13262
## 3                     2  19412
## 4                     3  29392
## 5                     4  74991
## 6                     5  86653
## 7                     6 142824
## 8                     7 105425
## 9                     8  97143
## 10                    9  57019
## 11                   10  34093
## 12                   11  28007
```

```r
# Casualties roughly divided between Vehicle 1 & 2
# Some casualties in multi-car crashes, with fewer and fewer the more cars involved.
# As expected.
UK_CasualtyData %>%
  count(Vehicle_Reference)
```

```
##    Vehicle_Reference      n
## 1                  1 392140
## 2                  2 281738
## 3                  3  20406
## 4                  4   3584
## 5                  5    770
## 6                  6    277
## 7                  7    107
## 8                  8     52
## 9                  9     22
## 10                10     17
## 11                11      4
## 12                12      5
## 13                13      5
## 14                14      4
## 15                15      3
## 16                16      5
## 17                18      2
## 18                19      1
## 19                20      2
## 20                21      3
## 21                22      4
## 22                23      1
## 23                24      2
## 24                25      1
## 25                26      1
## 26                27      3
## 27                28      1
## 28                32      1
## 29               101      1
## 30               999      1
```

```r
# Most casualties "Not Car Passenger"?
# Does that mean they weren't in THIS car, or weren't in any car?
# Does that mean most were pedestrians? But that doesn't align with other data?
UK_CasualtyData %>%
  count(Car_Passenger)
```

```
##   Car_Passenger      n
## 1            -1   1828
## 2             0 567397
## 3             1  81621
## 4             2  48317
```

```r
# Most (vast majority) not a Pedestrian, as stated above.
UK_CasualtyData %>%
  count(Pedestrian_Location)
```

```
##    Pedestrian_Location      n
## 1                   -1     11
## 2                    0 605313
## 3                    1  14376
## 4                    2    439
## 5                    3    214
## 6                    4   7227
## 7                    5  40928
## 8                    6   9725
## 9                    7    490
## 10                   8   3370
## 11                   9   9642
## 12                  10   7428
```

```r
# Most (vast majority) not a Pedestrian, as stated above.
UK_CasualtyData %>%
  count(Pedestrian_Movement)
```

```
##    Pedestrian_Movement      n
## 1                   -1     16
## 2                    0 605311
## 3                    1  32064
## 4                    2   6550
## 5                    3  19377
## 6                    4   4356
## 7                    5   4608
## 8                    6    727
## 9                    7   1273
## 10                   8   2134
## 11                   9  22747
```

```r
# Vast majority not a bus or coach passenger, as expected.
UK_CasualtyData %>%
  count(Bus_or_Coach_Passenger)
```

```
##   Bus_or_Coach_Passenger      n
## 1                     -1    163
## 2                      0 683507
## 3                      1    619
## 4                      2    692
## 5                      3   4426
## 6                      4   9756
```

```r
# Urban 485,508
# Small Town 52,300
# Rural 65,163
# Large number no data
UK_CasualtyData %>%
  count(Casualty_Home_Area_Type)
```

```
##   Casualty_Home_Area_Type      n
## 1                      -1  96192
## 2                       1 485508
## 3                       2  52300
## 4                       3  65163
```

```r
# For those with data, seems evenly spread Poverty to Affluent
UK_CasualtyData %>%
  count(Casualty_IMD_Decile)
```

```
##    Casualty_IMD_Decile      n
## 1                   -1 151742
## 2                    1  65220
## 3                    2  67435
## 4                    3  64260
## 5                    4  60894
## 6                    5  56618
## 7                    6  53529
## 8                    7  49743
## 9                    8  46876
## 10                   9  44246
## 11                  10  38600
```

```r
# Driver 450,659
# Passenger 154,656
# Pedestrian 93,848
UK_CasualtyData %>%
  count(Casualty_Class)
```

```
##   Casualty_Class      n
## 1              1 450659
## 2              2 154656
## 3              3  93848
```

```r
# Fatality 7,099
# Seriously Injured 96,587
# Slight Injury 595,477
```

```r
# This rings true to logic, but I actually thought fatalities would have been higher.
UK_CasualtyData %>%
  count(Casualty_Severity)
```

```
##   Casualty_Severity      n
## 1                 1   7099
## 2                 2  96587
## 3                 3 595477
```

```r
# At least 347 road workers died 2015-2018
# While small, that's not nothing given the population size.
UK_CasualtyData %>%
  count(Pedestrian_Road_Maintenance_Worker)
```

```
##   Pedestrian_Road_Maintenance_Worker      n
## 1                                 -1    434
## 2                                  0 671681
## 3                                  1    347
## 4                                  2  26701
```

UK_VehicleData Variable Analysis

The following represents a look at the counts for occurrences of variables in the UK_VehicleData dataset. Comments above each represent findings.

```r
# UK_VehicleData

# Repeated from other analysis.
# Most vehicles in accidents are in a one or two care accident.
# As expected, some accidents involve more than two cars, but far fewer.
UK_VehicleData %>%
  count(Vehicle_Reference) %>%
  arrange(-n) %>%
  head(20)
```

```
##    Vehicle_Reference      n
## 1                  1 529172
## 2                  2 373763
## 3                  3  54768
## 4                  4  12657
## 5                  5   3274
## 6                  6   1092
## 7                  7    452
## 8                  8    195
## 9                  9    107
## 10                10     56
## 11                11     30
## 12                12     20
## 13                13     16
```

```
## 14                 14      13
## 15                 15      10
## 16                 16       9
## 17                 22       6
## 18                 17       4
## 19                 18       4
## 20                 21       4
```

```
# Cars most, then vans and motorcycles.
UK_VehicleData %>%
  count(Vehicle_Type)
```

```
##    Vehicle_Type      n
## 1            -1    968
## 2             1  75566
## 3             2   7638
## 4             3  34624
## 5             4   9035
## 6             5  25696
## 7             8  21210
## 8             9 690344
## 9            10   1818
## 10           11  19890
## 11           16    386
## 12           17   2032
## 13           18     81
## 14           19  51542
## 15           20   5457
## 16           21  16721
## 17           22    967
## 18           23    174
## 19           90   5985
## 20           97   1346
## 21           98   4200
```

```
# The vast majority of vehicles were simply driving straight ahead
# I've heard about "Road Hypnosis", I wonder if this is at play?
# I think a visualization will tell more of the story for the others.
UK_VehicleData %>%
  count(Vehicle_Manoeuvre)
```

```
##    Vehicle_Manoeuvre      n
## 1                 -1  15550
## 2                  1  13833
## 3                  2  39615
## 4                  3  58108
## 5                  4  71680
## 6                  5  44667
## 7                  6   7735
## 8                  7  32377
## 9                  8   4656
## 10                 9  93441
## 11                10  14243
```

```
## 12                 11   8394
## 13                 12   9030
## 14                 13  18850
## 15                 14  11203
## 16                 15   6503
## 17                 16  28891
## 18                 17  33275
## 19                 18 463629
```

```r
# Most accidents did not involve getting towed.
UK_VehicleData %>%
  count(Towing_and_Articulation)
```

```
##   Towing_and_Articulation      n
## 1                      -1   9759
## 2                       0 951029
## 3                       1   9694
## 4                       2    197
## 5                       3    586
## 6                       4   3216
## 7                       5   1199
```

```r
# Vast majority were on a Main road not in a restricted lane.
# Again, statistically, wouldn't this be expected?
UK_VehicleData %>%
  count(Vehicle_Location_Restricted_Lane)
```

```
##    Vehicle_Location_Restricted_Lane      n
## 1                                -1  15237
## 2                                 0 935869
## 3                                 1    427
## 4                                 2   4443
## 5                                 3    350
## 6                                 4   3458
## 7                                 5   1242
## 8                                 6   2636
## 9                                 7    535
## 10                                8   1088
## 11                                9  10395
```

```r
# A slight majority were not at an intersection
# The rest were doing some road change.
UK_VehicleData %>%
  count(Junction_Location)
```

```
##    Junction_Location      n
## 1                 -1  11466
## 2                  0 394357
## 3                  1 216258
## 4                  2  52839
## 5                  3  14488
## 6                  4  26125
```

```
## 7                     5  24089
## 8                     6  41257
## 9                     7   3331
## 10                    8 191470
```

```r
# Most did not skid
UK_VehicleData %>%
  count(Skidding_and_Overturning)
```

```
##   Skidding_and_Overturning      n
## 1                       -1  16530
## 2                        0 858351
## 3                        1  64699
## 4                        2  16556
## 5                        3    291
## 6                        4    172
## 7                        5  19081
```

```r
# Most did not strike an object other than an in motion vehicle
# Of those that did strike an on-road object:
# 4 Parked Vehicle
# 10 Curb/Kerb
UK_VehicleData %>%
  count(Hit_Object_in_Carriageway)
```

```
##    Hit_Object_in_Carriageway      n
## 1                         -1  16643
## 2                          0 916666
## 3                          1    427
## 4                          2    545
## 5                          4  15349
## 6                          5    104
## 7                          6    558
## 8                          7   4515
## 9                          8   1688
## 10                         9   1005
## 11                        10  14290
## 12                        11   2812
## 13                        12   1078
```

```r
# Most did not strike an object other than an in motion vehicle
# Of those that4 did strike an off-road object:
# 4 Tree
# 10 "Other" Permanent Object
# 11 Wall or Fence
UK_VehicleData %>%
  count(Hit_Object_off_Carriageway)
```

```
##    Hit_Object_off_Carriageway      n
## 1                          -1  13838
## 2                           0 897151
## 3                           1   6309
```

```
## 4                              2   5236
## 5                              3   1997
## 6                              4  10490
## 7                              5    442
## 8                              6   4791
## 9                              7   4849
## 10                             8     67
## 11                             9   5877
## 12                            10  13326
## 13                            11  11307
```

```r
# The majority were not leaving the road,
# This aligns with the fact most were driving straigh ahead
# Oh those that did, the majority were
# Nearside (1)
# or Offside (7)
UK_VehicleData %>%
  count(Vehicle_Leaving_Carriageway)
```

```
##    Vehicle_Leaving_Carriageway      n
## 1                           -1  15860
## 2                            0 860391
## 3                            1  52897
## 4                            2   6387
## 5                            3   3744
## 6                            4   3817
## 7                            5   2480
## 8                            6    802
## 9                            7  26094
## 10                           8   3208
```

```r
#Defining Unfamiliar Terms:
# - **Nearside**: This is the side of the vehicle closest to the curb or side of the road.

# In the UK, where driving is on the left-hand side of the road,
#   the nearside would be the left side of the vehicle.

# - **Offside**: This is the side of the vehicle that is closest to the middle of the road.
#     In the UK, this would be the right side of the vehicle.

# These terms are used to describe where a vehicle left the carriageway
#   (i.e., the main part of the road). For example:
# - **"Nearside"**: The vehicle left the road and ended up on the side closest to the curb.
# - **"Offside"**: The vehicle left the road and ended up on the side closest to the middle of the road
#   or possibly even crossed to the opposite side.

# Related to side, left-handed or not?
# Most were, in the UK where almost all cars were, no shockers here.
UK_VehicleData %>%
  count(Was_Vehicle_Left_Hand_Drive)
```

```
##   Was_Vehicle_Left_Hand_Drive      n
## 1                          -1  17808
```

```
## 2                           1 952022
## 3                           2   5850
```

```r
# Most firs point of impact was front.
# This makes sense as there were a lot more vehicle one (529,172)
# than there were vehicle two (373,763) accidents.
# This means a significant amount were single car accidents striking something else.
UK_VehicleData %>%
  count(First_Point_of_Impact)
```

```
##   First_Point_of_Impact      n
## 1                    -1  10685
## 2                     0  57071
## 3                     1 478211
## 4                     2 168582
## 5                     3 137494
## 6                     4 123637
```

```r
# Reminder for sql in R:
#view(sqldf("sql here lower case"))
#variable_name <- sqldf("sql here lower case")
#view (variable_name)
# (Vehicle_Reference, First_Point_of_Impact, Hit_Object_in_Carriageway, Hit_Object_off_Carriageway) %>%


conn <- dbConnect(odbc(),
      Driver = "SQL Server",
       Server = "LAPTOP-76LHVPRQ\\SQLEXPRESS",
        Database = "UK_RoadSafety",
         Trusted_Connection = "True")

VehicleNum_vs_Impact <- sqldf(
  "select
  Vehicle_Reference,
  First_Point_of_Impact,
  Hit_Object_in_Carriageway,
  Hit_Object_off_Carriageway,
  Vehicle_Manoeuvre,
  count(Accident_Index) as count_of_incidents
  from UK_VehicleData
  where Vehicle_Reference in ('1','2')
  and First_Point_of_Impact <> '-1'
  and Hit_Object_in_Carriageway  <> '-1'
  and Hit_Object_off_Carriageway  <> '-1'

  group by
  Vehicle_Reference,
  First_Point_of_Impact,
  Hit_Object_in_Carriageway,
  Hit_Object_off_Carriageway,
  Vehicle_Manoeuvre

  --having count(Accident_Index) = '1'
```

```
  order by Vehicle_Reference, count_of_incidents DESC;")

# View SQL aggregates
# View(VehicleNum_vs_Impact)
summary(VehicleNum_vs_Impact)
```

```
##  Vehicle_Reference First_Point_of_Impact Hit_Object_in_Carriageway
##  Min.   :1.000     Min.   :0.000         Min.   : 0.000
##  1st Qu.:1.000     1st Qu.:1.000         1st Qu.: 0.000
##  Median :1.000     Median :2.000         Median : 6.000
##  Mean   :1.396     Mean   :2.133         Mean   : 5.099
##  3rd Qu.:2.000     3rd Qu.:3.000         3rd Qu.:10.000
##  Max.   :2.000     Max.   :4.000         Max.   :12.000
##  Hit_Object_off_Carriageway Vehicle_Manoeuvre count_of_incidents
##  Min.   : 0.000             Min.   :-1.00     Min.   :     1.0
##  1st Qu.: 0.000             1st Qu.: 5.00     1st Qu.:     1.0
##  Median : 4.000             Median :12.00     Median :     3.0
##  Mean   : 4.443             Mean   :10.93     Mean   :   220.3
##  3rd Qu.: 9.000             3rd Qu.:16.00     3rd Qu.:    11.0
##  Max.   :11.000             Max.   :18.00     Max.   :134574.0
```

```
head(VehicleNum_vs_Impact)
```

```
##   Vehicle_Reference First_Point_of_Impact Hit_Object_in_Carriageway
## 1                 1                     1                         0
## 2                 1                     1                         0
## 3                 1                     4                         0
## 4                 1                     3                         0
## 5                 1                     1                         0
## 6                 1                     1                         0
##   Hit_Object_off_Carriageway Vehicle_Manoeuvre count_of_incidents
## 1                          0                18             134574
## 2                          0                 9              32382
## 3                          0                18              27445
## 4                          0                18              23147
## 5                          0                 5              16417
## 6                          0                 4              15256
```

Joined Tables

By joining the tables on the Primary Key correlations reference keys, we might be able to see some interesting overlaps or correlations.

```
# ESTABLISH CONNECTION TO MY LOCAL DATABASE

connection <- odbcDriverConnect("driver={SQL Server};server=LAPTOP-76LHVPRQ\\SQLEXPRESS;database=UK_Roa

# ASSIGN THE DATABASE TABLE AS A DATAFRAME (df) VARIABLE FOR EASIER RECALL
```

```r
UK_AccidentData <- sqlFetch(connection, "dbo.uk_accidents")
#write.csv(UK_AccidentData, "C:/Users/darre/My Drive (dwolfe.data@gmail.com)/!Datasets/Amplify_Intervie


UK_CasualtyData <- sqlFetch(connection, "dbo.uk_casualties")
#write.csv(UK_CasualtyData, "C:/Users/darre/My Drive (dwolfe.data@gmail.com)/!Datasets/Amplify_Intervie

UK_VehicleData <- sqlFetch(connection, "dbo.uk_vehicles")
#write.csv(UK_VehicleData, "C:/Users/darre/My Drive (dwolfe.data@gmail.com)/!Datasets/Amplify_Interview

All_Accident_Data <- sqldf(
                    "select *
                    from UK_AccidentData as a
                    join UK_VehicleData as v
                      on a.Accident_Index=v.Accident_Index
                    join UK_CasualtyData as c
                      on a.Accident_Index=c.Accident_Index
                      and v.Vehicle_Reference=c.Vehicle_Reference;")

# View(All_Accident_Data)
head(All_Accident_Data)
```

```
##   Accident_Index Location_Easting_OSGR Location_Northing_OSGR Longitude
## 1  201501BS70055                527920                 179050 -0.158650
## 2  201501BS70056                523110                 181540 -0.227035
## 3  201501BS70057                525540                 179380 -0.192799
## 4  201501BS70058                526940                 177450 -0.173334
## 5  201501BS70060                525230                 180630 -0.196819
## 6  201501BS70061                525520                 179460 -0.193059
##   Latitude Police_Force Accident_Severity Number_of_Vehicles
## 1 51.49593            1                 3                  2
## 2 51.51937            1                 3                  2
## 3 51.49943            1                 3                  2
## 4 51.48177            1                 3                  1
## 5 51.51073            1                 3                  2
## 6 51.50015            1                 3                  2
##   Number_of_Casualties Day_of_Week             Time Local_Authority_District
## 1                    1           1 23:05:00.0000000                       12
## 2                    1           5 08:35:00.0000000                       11
## 3                    1           5 16:30:00.0000000                       12
## 4                    1           6 08:52:00.0000000                       12
## 5                    1           6 13:27:00.0000000                       12
## 6                    1           2 21:16:00.0000000                       12
##   Local_Authority_Highway First_Road_Class First_Road_Number Road_Type
## 1               E09000020                3              3216         6
## 2               E09000013                4               412         6
## 3               E09000020                5                 0         6
## 4               E09000020                3              3220         6
## 5               E09000020                6                 0         6
## 6               E09000020                5                 0         6
##   Speed_limit Junction_Detail Junction_Control Second_Road_Class
## 1          30               0               -1                -1
## 2          30               0               -1                -1
```

```
## 3         30              3                4                 6
## 4         30              0               -1                -1
## 5         30              0               -1                -1
## 6         30              3                4                 6
##   Second_Road_Number Pedestrian_Crossing_Human_Control
## 1                  0                                 0
## 2                  0                                 0
## 3                  0                                 0
## 4                  0                                 0
## 5                  0                                 0
## 6                  0                                 0
##   Pedestrian_Crossing_Physical_Facilities Light_Conditions Weather_Conditions
## 1                                       0                4                  1
## 2                                       0                1                  2
## 3                                       0                1                  1
## 4                                       5                1                  1
## 5                                       0                1                  1
## 6                                       0                4                  1
##   Road_Surface_Conditions Special_Conditions_at_Site Carriageway_Hazards
## 1                       1                          0                   0
## 2                       2                          0                   0
## 3                       1                          0                   0
## 4                       1                          0                   0
## 5                       1                          0                   0
## 6                       1                          0                   0
##   Urban_or_Rural_Area Did_Police_Officer_Attend_Scene_of_Accident
## 1                   1                                           1
## 2                   1                                           1
## 3                   1                                           1
## 4                   1                                           1
## 5                   1                                           1
## 6                   1                                           1
##   LSOA_of_Accident_Location       Date       TimeGroup Accident_Index
## 1                 E01002863 2015-08-02      Late_Night  201501BS70055
## 2                 E01001874 2015-05-02 Morning_Commute  201501BS70056
## 3                 E01002816 2015-05-02       Afternoon  201501BS70057
## 4                 E01002840 2015-06-02 Morning_Commute  201501BS70058
## 5                 E01002884 2015-06-02       Afternoon  201501BS70060
## 6                 E01002816 2015-09-02      Late_Night  201501BS70061
##   Vehicle_Reference Vehicle_Type Towing_and_Articulation Vehicle_Manoeuvre
## 1                 1            9                       0                18
## 2                 2            1                       0                18
## 3                 2            3                       0                 7
## 4                 1           11                       0                 4
## 5                 2            1                       0                18
## 6                 1            1                       0                18
##   Vehicle_Location_Restricted_Lane Junction_Location Skidding_and_Overturning
## 1                                0                 0                        0
## 2                                0                 0                        0
## 3                                0                 8                        0
## 4                                0                 0                        0
## 5                                0                 0                        0
## 6                                0                 2                        0
##   Hit_Object_in_Carriageway Vehicle_Leaving_Carriageway
```

```
## 1                          0                          0
## 2                          0                          0
## 3                          0                          0
## 4                          0                          0
## 5                          8                          0
## 6                          0                          0
##   Hit_Object_off_Carriageway First_Point_of_Impact Was_Vehicle_Left_Hand_Drive
## 1                          0                     1                           1
## 2                          0                     3                           1
## 3                          0                     0                           1
## 4                          0                     0                           1
## 5                          0                     1                           1
## 6                          0                     1                           1
##   Journey_Purpose_of_Driver Sex_of_Driver Age_of_Driver Age_Band_of_Driver
## 1                         6             2            20                  4
## 2                         6             1            42                  7
## 3                         6             1            29                  6
## 4                         1             1            63                  9
## 5                         6             1            25                  5
## 6                         6             1            56                  9
##   Engine_Capacity_CC Propulsion_Code Age_of_Vehicle Driver_IMD_Decile
## 1               3498               1              1                -1
## 2                 -1              -1             -1                -1
## 3                 -1              -1             -1                -1
## 4               4500               2              1                -1
## 5                 -1              -1             -1                -1
## 6                 -1              -1             -1                -1
##   Driver_Home_Area_Type Vehicle_IMD_Decile Accident_Index Vehicle_Reference
## 1                     1                 -1  201501BS70055                 1
## 2                     1                 -1  201501BS70056                 2
## 3                     1                 -1  201501BS70057                 2
## 4                     1                 -1  201501BS70058                 1
## 5                     1                 -1  201501BS70060                 2
## 6                     1                 -1  201501BS70061                 1
##   Casualty_Reference Casualty_Class Sex_of_Casualty Age_of_Casualty
## 1                  1              2               2              20
## 2                  1              1               1              42
## 3                  1              1               1              29
## 4                  1              2               2              26
## 5                  1              1               1              25
## 6                  1              1               1              56
##   Age_Band_of_Casualty Casualty_Severity Pedestrian_Location
## 1                    4                 3                   0
## 2                    7                 3                   0
## 3                    6                 3                   0
## 4                    6                 3                   0
## 5                    5                 3                   0
## 6                    9                 3                   0
##   Pedestrian_Movement Car_Passenger Bus_or_Coach_Passenger
## 1                   0             1                      0
## 2                   0             0                      0
## 3                   0             0                      0
## 4                   0             0                      3
## 5                   0             0                      0
```

```
## 6                   0             0                     0
##   Pedestrian_Road_Maintenance_Worker Casualty_Type Casualty_Home_Area_Type
## 1                                  0             9                      -1
## 2                                  0             1                       1
## 3                                  0             3                       1
## 4                                  0            11                       1
## 5                                  0             1                       1
## 6                                  0             1                       1
##   Casualty_IMD_Decile
## 1                  -1
## 2                   3
## 3                   6
## 4                   8
## 5                   3
## 6                   5
```

```r
summary(All_Accident_Data)
```

```
##  Accident_Index     Location_Easting_OSGR Location_Northing_OSGR
##  Length:699163      Min.   : 70860        Min.   :  10235
##  Class :character   1st Qu.:384909        1st Qu.: 176620
##  Mode  :character   Median :451891        Median : 241761
##                     Mean   :447498        Mean   : 288741
##                     3rd Qu.:527257        3rd Qu.: 391268
##                     Max.   :655391        Max.   :1209512
##                     NA's   :149           NA's   :149
##    Longitude          Latitude      Police_Force   Accident_Severity
##  Min.   :-7.4229   Min.   :49.91   Min.   : 1.00   Min.   :1.000
##  1st Qu.:-2.2264   1st Qu.:51.47   1st Qu.: 6.00   1st Qu.:3.000
##  Median :-1.2317   Median :52.06   Median :30.00   Median :3.000
##  Mean   :-1.3214   Mean   :52.49   Mean   :29.61   Mean   :2.793
##  3rd Qu.:-0.1654   3rd Qu.:53.42   3rd Qu.:45.00   3rd Qu.:3.000
##  Max.   : 1.7596   Max.   :60.76   Max.   :98.00   Max.   :3.000
##  NA's   :163       NA's   :163
##  Number_of_Vehicles Number_of_Casualties  Day_of_Week        Time
##  Min.   : 1.000     Min.   : 1.000       Min.   :1.000   Length:699163
##  1st Qu.: 1.000     1st Qu.: 1.000       1st Qu.:2.000   Class :character
##  Median : 2.000     Median : 1.000       Median :4.000   Mode  :character
##  Mean   : 1.945     Mean   : 1.781       Mean   :4.102
##  3rd Qu.: 2.000     3rd Qu.: 2.000       3rd Qu.:6.000
##  Max.   :37.000     Max.   :59.000       Max.   :7.000
##
##  Local_Authority_District Local_Authority_Highway First_Road_Class
##  Min.   :  1.0            Length:699163           Min.   :1.000
##  1st Qu.:104.0            Class :character        1st Qu.:3.000
##  Median :321.0            Mode  :character        Median :3.000
##  Mean   :341.3                                    Mean   :4.094
##  3rd Qu.:516.0                                    3rd Qu.:6.000
##  Max.   :941.0                                    Max.   :6.000
##
##  First_Road_Number   Road_Type       Speed_limit    Junction_Detail
##  Min.   :    0.0    Min.   :-1.000   Min.   : 0.00   Min.   :-1.000
##  1st Qu.:    0.0    1st Qu.: 6.000   1st Qu.:30.00   1st Qu.: 0.000
##  Median :   62.0    Median : 6.000   Median :30.00   Median : 1.000
```

```
##  Mean   : 906.4    Mean   : 5.156   Mean   :38.97   Mean   : 2.238
##  3rd Qu.: 633.0    3rd Qu.: 6.000   3rd Qu.:50.00   3rd Qu.: 3.000
##  Max.   :9918.0    Max.   : 9.000   Max.   :70.00   Max.   : 9.000
##                                     NA's   :47
##  Junction_Control Second_Road_Class Second_Road_Number
##  Min.   :-1.000   Min.   :-1.000    Min.   :  -1.0
##  1st Qu.:-1.000   1st Qu.:-1.000    1st Qu.:   0.0
##  Median : 2.000   Median : 3.000    Median :   0.0
##  Mean   : 1.608   Mean   : 2.576    Mean   : 319.8
##  3rd Qu.: 4.000   3rd Qu.: 6.000    3rd Qu.:   0.0
##  Max.   : 4.000   Max.   : 6.000    Max.   :9999.0
##
##  Pedestrian_Crossing_Human_Control Pedestrian_Crossing_Physical_Facilities
##  Min.   :-1.000000                 Min.   :-1.0000
##  1st Qu.: 0.000000                 1st Qu.: 0.0000
##  Median : 0.000000                 Median : 0.0000
##  Mean   : 0.005251                 Mean   : 0.8015
##  3rd Qu.: 0.000000                 3rd Qu.: 0.0000
##  Max.   : 2.000000                 Max.   : 8.0000
##
##  Light_Conditions Weather_Conditions Road_Surface_Conditions
##  Min.   :-1.000   Min.   :-1.00      Min.   :-1.000
##  1st Qu.: 1.000   1st Qu.: 1.00      1st Qu.: 1.000
##  Median : 1.000   Median : 1.00      Median : 1.000
##  Mean   : 2.019   Mean   : 1.56      Mean   : 1.295
##  3rd Qu.: 4.000   3rd Qu.: 1.00      3rd Qu.: 2.000
##  Max.   : 7.000   Max.   : 9.00      Max.   : 5.000
##
##  Special_Conditions_at_Site Carriageway_Hazards Urban_or_Rural_Area
##  Min.   :-1.00000           Min.   :-1.00000    Min.   :-1.000
##  1st Qu.: 0.00000           1st Qu.: 0.00000    1st Qu.: 1.000
##  Median : 0.00000           Median : 0.00000    Median : 1.000
##  Mean   : 0.08974           Mean   : 0.05498    Mean   : 1.374
##  3rd Qu.: 0.00000           3rd Qu.: 0.00000    3rd Qu.: 2.000
##  Max.   : 7.00000           Max.   : 7.00000    Max.   : 3.000
##
##  Did_Police_Officer_Attend_Scene_of_Accident LSOA_of_Accident_Location
##  Min.   :-1.000                              Length:699163
##  1st Qu.: 1.000                              Class :character
##  Median : 1.000                              Mode  :character
##  Mean   : 1.225
##  3rd Qu.: 1.000
##  Max.   : 3.000
##
##      Date             TimeGroup          Accident_Index     Vehicle_Reference
##  Length:699163      Length:699163      Length:699163      Min.   :  1.000
##  Class :character   Class :character   Class :character   1st Qu.:  1.000
##  Mode  :character   Mode  :character   Mode  :character   Median :  1.000
##                                                           Mean   :  1.488
##                                                           3rd Qu.:  2.000
##                                                           Max.   :999.000
##
##   Vehicle_Type    Towing_and_Articulation Vehicle_Manoeuvre
##  Min.   :-1.000   Min.   :-1.000000       Min.   :-1.00
```

```
##  1st Qu.: 9.000   1st Qu.: 0.000000       1st Qu.: 9.00
##  Median : 9.000   Median : 0.000000       Median :18.00
##  Mean   : 8.714   Mean   : 0.006878       Mean   :13.37
##  3rd Qu.: 9.000   3rd Qu.: 0.000000       3rd Qu.:18.00
##  Max.   :98.000   Max.   : 5.000000       Max.   :18.00
##
##  Vehicle_Location_Restricted_Lane Junction_Location Skidding_and_Overturning
##  Min.   :-1.0000                  Min.   :-1.000    Min.   :-1.000
##  1st Qu.: 0.0000                  1st Qu.: 0.000    1st Qu.: 0.000
##  Median : 0.0000                  Median : 1.000    Median : 0.000
##  Mean   : 0.1509                  Mean   : 2.364    Mean   : 0.294
##  3rd Qu.: 0.0000                  3rd Qu.: 5.000    3rd Qu.: 0.000
##  Max.   : 9.0000                  Max.   : 8.000    Max.   : 5.000
##
##  Hit_Object_in_Carriageway Vehicle_Leaving_Carriageway
##  Min.   :-1.0000           Min.   :-1.0000
##  1st Qu.: 0.0000           1st Qu.: 0.0000
##  Median : 0.0000           Median : 0.0000
##  Mean   : 0.4397           Mean   : 0.5006
##  3rd Qu.: 0.0000           3rd Qu.: 0.0000
##  Max.   :12.0000           Max.   : 8.0000
##
##  Hit_Object_off_Carriageway First_Point_of_Impact Was_Vehicle_Left_Hand_Drive
##  Min.   :-1.0000            Min.   :-1.000        Min.   :-1.0000
##  1st Qu.: 0.0000            1st Qu.: 1.000        1st Qu.: 1.0000
##  Median : 0.0000            Median : 1.000        Median : 1.0000
##  Mean   : 0.7287            Mean   : 1.776        Mean   : 0.9775
##  3rd Qu.: 0.0000            3rd Qu.: 3.000        3rd Qu.: 1.0000
##  Max.   :11.0000            Max.   : 4.000        Max.   : 2.0000
##
##  Journey_Purpose_of_Driver Sex_of_Driver    Age_of_Driver    Age_Band_of_Driver
##  Min.   :-1.000            Min.   :-1.000   Min.   : -1.00   Min.   :-1.000
##  1st Qu.: 3.000            1st Qu.: 1.000   1st Qu.: 25.00   1st Qu.: 5.000
##  Median : 6.000            Median : 1.000   Median : 36.00   Median : 7.000
##  Mean   : 4.738            Mean   : 1.362   Mean   : 37.88   Mean   : 6.468
##  3rd Qu.: 6.000            3rd Qu.: 2.000   3rd Qu.: 50.00   3rd Qu.: 8.000
##  Max.   : 6.000            Max.   : 3.000   Max.   :101.00   Max.   :11.000
##
##  Engine_Capacity_CC Propulsion_Code   Age_of_Vehicle    Driver_IMD_Decile
##  Min.   :   -1      Min.   :-1.0000   Min.   : -1.000   Min.   :-1.000
##  1st Qu.:  124      1st Qu.: 1.0000   1st Qu.: -1.000   1st Qu.:-1.000
##  Median : 1360      Median : 1.0000   Median :  5.000   Median : 2.000
##  Mean   : 1323      Mean   : 0.9135   Mean   :  5.755   Mean   : 2.434
##  3rd Qu.: 1797      3rd Qu.: 2.0000   3rd Qu.: 10.000   3rd Qu.: 5.000
##  Max.   :91000      Max.   :12.0000   Max.   :105.000   Max.   :10.000
##
##  Driver_Home_Area_Type Vehicle_IMD_Decile Accident_Index     Vehicle_Reference
##  Min.   :-1.0000       Min.   :-1.000     Length:699163      Min.   :  1.000
##  1st Qu.: 1.0000       1st Qu.:-1.000     Class :character   1st Qu.:  1.000
##  Median : 1.0000       Median : 2.000     Mode  :character   Median :  1.000
##  Mean   : 0.9852       Mean   : 2.434                        Mean   :  1.488
##  3rd Qu.: 1.0000       3rd Qu.: 5.000                        3rd Qu.:  2.000
##  Max.   : 3.0000       Max.   :10.000                        Max.   :999.000
##
```

```
##  Casualty_Reference Casualty_Class Sex_of_Casualty  Age_of_Casualty
##  Min.   :  1.000    Min.   :1.00   Min.   :-1.000   Min.   : -1.00
##  1st Qu.:  1.000    1st Qu.:1.00   1st Qu.: 1.000   1st Qu.: 22.00
##  Median :  1.000    Median :1.00   Median : 1.000   Median : 33.00
##  Mean   :  1.406    Mean   :1.49   Mean   : 1.406   Mean   : 36.48
##  3rd Qu.:  2.000    3rd Qu.:2.00   3rd Qu.: 2.000   3rd Qu.: 50.00
##  Max.   :991.000    Max.   :3.00   Max.   : 2.000   Max.   :104.00
##
##  Age_Band_of_Casualty Casualty_Severity Pedestrian_Location Pedestrian_Movement
##  Min.   :-1.000       Min.   :1.000     Min.   :-1.000      Min.   :-1.0000
##  1st Qu.: 5.000       1st Qu.:3.000     1st Qu.: 0.000      1st Qu.: 0.0000
##  Median : 6.000       Median :3.000     Median : 0.000      Median : 0.0000
##  Mean   : 6.289       Mean   :2.842     Mean   : 0.714      Mean   : 0.5418
##  3rd Qu.: 8.000       3rd Qu.:3.000     3rd Qu.: 0.000      3rd Qu.: 0.0000
##  Max.   :11.000       Max.   :3.000     Max.   :10.000      Max.   : 9.0000
##
##  Car_Passenger     Bus_or_Coach_Passenger Pedestrian_Road_Maintenance_Worker
##  Min.   :-1.0000   Min.   :-1.00000       Min.   :-1.00000
##  1st Qu.: 0.0000   1st Qu.: 0.00000       1st Qu.: 0.00000
##  Median : 0.0000   Median : 0.00000       Median : 0.00000
##  Mean   : 0.2523   Mean   : 0.07744       Mean   : 0.07626
##  3rd Qu.: 0.0000   3rd Qu.: 0.00000       3rd Qu.: 0.00000
##  Max.   : 2.0000   Max.   : 4.00000       Max.   : 2.00000
##
##  Casualty_Type    Casualty_Home_Area_Type Casualty_IMD_Decile
##  Min.   :-1.000   Min.   :-1.000          Min.   :-1.000
##  1st Qu.: 3.000   1st Qu.: 1.000          1st Qu.: 1.000
##  Median : 9.000   Median : 1.000          Median : 4.000
##  Mean   : 7.281   Mean   : 0.986          Mean   : 3.714
##  3rd Qu.: 9.000   3rd Qu.: 1.000          3rd Qu.: 7.000
##  Max.   :98.000   Max.   : 3.000          Max.   :10.000
##
```

```
#write.csv(All_Accident_Data, "C:/Users/darre/My Drive (dwolfe.data@gmail.com)/!Datasets/Amplify_Interv


Accidents_Vehicles_Casualties <- sqldf(
  "-- How many accidents had how many vehicles and casualties?
  select
  count(a.Accident_Index) AS Count_Incidents,
  v.Vehicle_Reference,
  c.Casualty_Reference

  from UK_AccidentData as a
  join UK_VehicleData as v
    on a.Accident_Index=v.Accident_Index
  join UK_CasualtyData as c
    on a.Accident_Index=c.Accident_Index
    and v.Vehicle_Reference=c.Vehicle_Reference

  group by
  v.Vehicle_Reference,
  c.Casualty_Reference
```

```
  order by
  v.Vehicle_Reference,
  c.Casualty_Reference")

# View(All_Accident_Data)
head(Accidents_Vehicles_Casualties)
```

```
##   Count_Incidents Vehicle_Reference Casualty_Reference
## 1          329081                 1                  1
## 2           41015                 1                  2
## 3           14104                 1                  3
## 4            4901                 1                  4
## 5            1663                 1                  5
## 6             568                 1                  6
```

```
summary(Accidents_Vehicles_Casualties)
```

```
##  Count_Incidents    Vehicle_Reference Casualty_Reference
##  Min.   :     1.0   Min.   :  1.000   Min.   :  1.00
##  1st Qu.:     1.0   1st Qu.:  2.000   1st Qu.:  5.00
##  Median :     2.5   Median :  3.000   Median : 13.00
##  Mean   :  2515.0   Mean   :  9.932   Mean   : 22.28
##  3rd Qu.:    13.8   3rd Qu.:  8.000   3rd Qu.: 29.00
##  Max.   :329081.0   Max.   :999.000   Max.   :991.00
```

```
#write.csv(Accidents_Vehicles_Casualties, "C:/Users/darre/My Drive (dwolfe.data@gmail.com)/!Datasets/Am




Accidents_Light_Weather <- sqldf(
  "-- How many accidents had how many vehicles and casualties?
select
  count(a.Accident_Index) AS Count_Incidents,
  a.Light_Conditions,
  a.Weather_Conditions,
  a.Road_Surface_Conditions,
  v.Vehicle_Reference,
  c.Casualty_Reference

  from UK_AccidentData as a
  join UK_VehicleData as v
    on a.Accident_Index=v.Accident_Index
  join UK_CasualtyData as c
    on a.Accident_Index=c.Accident_Index
    and v.Vehicle_Reference=c.Vehicle_Reference

  where  a.Light_Conditions <> '-1'
  and a.Weather_Conditions <> '-1'
  and a.Road_Surface_Conditions <> '-1'

  group by
```

```
  a.Light_Conditions,
  a.Weather_Conditions,
  a.Road_Surface_Conditions,
  v.Vehicle_Reference,
  c.Casualty_Reference

  order by
  v.Vehicle_Reference,
  c.Casualty_Reference")

# View(All_Accident_Data)
head(Accidents_Light_Weather)
```

```
##   Count_Incidents Light_Conditions Weather_Conditions Road_Surface_Conditions
## 1          170261                1                  1                       1
## 2           20631                1                  1                       2
## 3             131                1                  1                       3
## 4            2040                1                  1                       4
## 5              34                1                  1                       5
## 6             325                1                  2                       1
##   Vehicle_Reference Casualty_Reference
## 1                 1                  1
## 2                 1                  1
## 3                 1                  1
## 4                 1                  1
## 5                 1                  1
## 6                 1                  1
```

```
summary(Accidents_Light_Weather)
```

```
##  Count_Incidents    Light_Conditions Weather_Conditions Road_Surface_Conditions
##  Min.   :      1.0   Min.   :1.000    Min.   :1.000      Min.   :1.000
##  1st Qu.:      1.0   1st Qu.:1.000    1st Qu.:1.000      1st Qu.:1.000
##  Median :      3.0   Median :4.000    Median :3.000      Median :2.000
##  Mean   :    262.6   Mean   :3.605    Mean   :3.935      Mean   :2.222
##  3rd Qu.:     12.0   3rd Qu.:6.000    3rd Qu.:7.000      3rd Qu.:3.000
##  Max.   :170261.0   Max.   :7.000    Max.   :9.000      Max.   :5.000
##  Vehicle_Reference Casualty_Reference
##  Min.   :  1.000   Min.   :  1.00
##  1st Qu.:  1.000   1st Qu.:  2.00
##  Median :  2.000   Median :  3.00
##  Mean   :  3.295   Mean   :  5.94
##  3rd Qu.:  3.000   3rd Qu.:  5.00
##  Max.   :999.000   Max.   :991.00
```

```
#write.csv(Accidents_Light_Weather, "C:/Users/darre/My Drive (dwolfe.data@gmail.com)/!Datasets/Amplify_


All_Accident_Data <- sqldf(
                    "select *
                    from UK_AccidentData as a
                    join UK_VehicleData as v
```

```r
                      on a.Accident_Index=v.Accident_Index
                    join UK_CasualtyData as c
                      on a.Accident_Index=c.Accident_Index
                      and v.Vehicle_Reference=c.Vehicle_Reference;")

# View(All_Accident_Data)
head(All_Accident_Data)
```

```
##   Accident_Index Location_Easting_OSGR Location_Northing_OSGR Longitude
## 1  201501BS70055                527920                 179050 -0.158650
## 2  201501BS70056                523110                 181540 -0.227035
## 3  201501BS70057                525540                 179380 -0.192799
## 4  201501BS70058                526940                 177450 -0.173334
## 5  201501BS70060                525230                 180630 -0.196819
## 6  201501BS70061                525520                 179460 -0.193059
##   Latitude Police_Force Accident_Severity Number_of_Vehicles
## 1 51.49593            1                 3                  2
## 2 51.51937            1                 3                  2
## 3 51.49943            1                 3                  2
## 4 51.48177            1                 3                  1
## 5 51.51073            1                 3                  2
## 6 51.50015            1                 3                  2
##   Number_of_Casualties Day_of_Week             Time Local_Authority_District
## 1                    1           1 23:05:00.0000000                       12
## 2                    1           5 08:35:00.0000000                       11
## 3                    1           5 16:30:00.0000000                       12
## 4                    1           6 08:52:00.0000000                       12
## 5                    1           6 13:27:00.0000000                       12
## 6                    1           2 21:16:00.0000000                       12
##   Local_Authority_Highway First_Road_Class First_Road_Number Road_Type
## 1               E09000020                3              3216         6
## 2               E09000013                4               412         6
## 3               E09000020                5                 0         6
## 4               E09000020                3              3220         6
## 5               E09000020                6                 0         6
## 6               E09000020                5                 0         6
##   Speed_limit Junction_Detail Junction_Control Second_Road_Class
## 1          30               0               -1                -1
## 2          30               0               -1                -1
## 3          30               3                4                 6
## 4          30               0               -1                -1
## 5          30               0               -1                -1
## 6          30               3                4                 6
##   Second_Road_Number Pedestrian_Crossing_Human_Control
## 1                  0                                 0
## 2                  0                                 0
## 3                  0                                 0
## 4                  0                                 0
## 5                  0                                 0
## 6                  0                                 0
##   Pedestrian_Crossing_Physical_Facilities Light_Conditions Weather_Conditions
## 1                                       0                4                  1
## 2                                       0                1                  2
```

```
## 3                                  0              1                1
## 4                                  5              1                1
## 5                                  0              1                1
## 6                                  0              4                1
##   Road_Surface_Conditions Special_Conditions_at_Site Carriageway_Hazards
## 1                       1                          0                   0
## 2                       2                          0                   0
## 3                       1                          0                   0
## 4                       1                          0                   0
## 5                       1                          0                   0
## 6                       1                          0                   0
##   Urban_or_Rural_Area Did_Police_Officer_Attend_Scene_of_Accident
## 1                   1                                           1
## 2                   1                                           1
## 3                   1                                           1
## 4                   1                                           1
## 5                   1                                           1
## 6                   1                                           1
##   LSOA_of_Accident_Location       Date       TimeGroup Accident_Index
## 1                 E01002863 2015-08-02      Late_Night  201501BS70055
## 2                 E01001874 2015-05-02 Morning_Commute  201501BS70056
## 3                 E01002816 2015-05-02       Afternoon  201501BS70057
## 4                 E01002840 2015-06-02 Morning_Commute  201501BS70058
## 5                 E01002884 2015-06-02       Afternoon  201501BS70060
## 6                 E01002816 2015-09-02      Late_Night  201501BS70061
##   Vehicle_Reference Vehicle_Type Towing_and_Articulation Vehicle_Manoeuvre
## 1                 1            9                       0                18
## 2                 2            1                       0                18
## 3                 2            3                       0                 7
## 4                 1           11                       0                 4
## 5                 2            1                       0                18
## 6                 1            1                       0                18
##   Vehicle_Location_Restricted_Lane Junction_Location Skidding_and_Overturning
## 1                                0                 0                        0
## 2                                0                 0                        0
## 3                                0                 8                        0
## 4                                0                 0                        0
## 5                                0                 0                        0
## 6                                0                 2                        0
##   Hit_Object_in_Carriageway Vehicle_Leaving_Carriageway
## 1                         0                           0
## 2                         0                           0
## 3                         0                           0
## 4                         0                           0
## 5                         8                           0
## 6                         0                           0
##   Hit_Object_off_Carriageway First_Point_of_Impact Was_Vehicle_Left_Hand_Drive
## 1                          0                     1                           1
## 2                          0                     3                           1
## 3                          0                     0                           1
## 4                          0                     0                           1
## 5                          0                     1                           1
## 6                          0                     1                           1
##   Journey_Purpose_of_Driver Sex_of_Driver Age_of_Driver Age_Band_of_Driver
```

```
## 1                    6            2           20                4
## 2                    6            1           42                7
## 3                    6            1           29                6
## 4                    1            1           63                9
## 5                    6            1           25                5
## 6                    6            1           56                9
##   Engine_Capacity_CC Propulsion_Code Age_of_Vehicle Driver_IMD_Decile
## 1               3498               1              1                -1
## 2                 -1              -1             -1                -1
## 3                 -1              -1             -1                -1
## 4               4500               2              1                -1
## 5                 -1              -1             -1                -1
## 6                 -1              -1             -1                -1
##   Driver_Home_Area_Type Vehicle_IMD_Decile Accident_Index Vehicle_Reference
## 1                     1                 -1  201501BS70055                 1
## 2                     1                 -1  201501BS70056                 2
## 3                     1                 -1  201501BS70057                 2
## 4                     1                 -1  201501BS70058                 1
## 5                     1                 -1  201501BS70060                 2
## 6                     1                 -1  201501BS70061                 1
##   Casualty_Reference Casualty_Class Sex_of_Casualty Age_of_Casualty
## 1                  1              2               2              20
## 2                  1              1               1              42
## 3                  1              1               1              29
## 4                  1              2               2              26
## 5                  1              1               1              25
## 6                  1              1               1              56
##   Age_Band_of_Casualty Casualty_Severity Pedestrian_Location
## 1                    4                 3                   0
## 2                    7                 3                   0
## 3                    6                 3                   0
## 4                    6                 3                   0
## 5                    5                 3                   0
## 6                    9                 3                   0
##   Pedestrian_Movement Car_Passenger Bus_or_Coach_Passenger
## 1                   0             1                      0
## 2                   0             0                      0
## 3                   0             0                      0
## 4                   0             0                      3
## 5                   0             0                      0
## 6                   0             0                      0
##   Pedestrian_Road_Maintenance_Worker Casualty_Type Casualty_Home_Area_Type
## 1                                  0             9                      -1
## 2                                  0             1                       1
## 3                                  0             3                       1
## 4                                  0            11                       1
## 5                                  0             1                       1
## 6                                  0             1                       1
##   Casualty_IMD_Decile
## 1                  -1
## 2                   3
## 3                   6
## 4                   8
## 5                   3
```

```
## 6                   5
```

```r
summary(All_Accident_Data)
```

```
##  Accident_Index     Location_Easting_OSGR Location_Northing_OSGR
##  Length:699163      Min.   : 70860        Min.   :  10235
##  Class :character   1st Qu.:384909        1st Qu.: 176620
##  Mode  :character   Median :451891        Median : 241761
##                     Mean   :447498        Mean   : 288741
##                     3rd Qu.:527257        3rd Qu.: 391268
##                     Max.   :655391        Max.   :1209512
##                     NA's   :149           NA's   :149
##    Longitude          Latitude      Police_Force   Accident_Severity
##  Min.   :-7.4229   Min.   :49.91   Min.   : 1.00   Min.   :1.000
##  1st Qu.:-2.2264   1st Qu.:51.47   1st Qu.: 6.00   1st Qu.:3.000
##  Median :-1.2317   Median :52.06   Median :30.00   Median :3.000
##  Mean   :-1.3214   Mean   :52.49   Mean   :29.61   Mean   :2.793
##  3rd Qu.:-0.1654   3rd Qu.:53.42   3rd Qu.:45.00   3rd Qu.:3.000
##  Max.   : 1.7596   Max.   :60.76   Max.   :98.00   Max.   :3.000
##  NA's   :163       NA's   :163
##  Number_of_Vehicles Number_of_Casualties  Day_of_Week        Time
##  Min.   : 1.000     Min.   : 1.000       Min.   :1.000   Length:699163
##  1st Qu.: 1.000     1st Qu.: 1.000       1st Qu.:2.000   Class :character
##  Median : 2.000     Median : 1.000       Median :4.000   Mode  :character
##  Mean   : 1.945     Mean   : 1.781       Mean   :4.102
##  3rd Qu.: 2.000     3rd Qu.: 2.000       3rd Qu.:6.000
##  Max.   :37.000     Max.   :59.000       Max.   :7.000
##
##  Local_Authority_District Local_Authority_Highway First_Road_Class
##  Min.   :  1.0            Length:699163           Min.   :1.000
##  1st Qu.:104.0            Class :character        1st Qu.:3.000
##  Median :321.0            Mode  :character        Median :3.000
##  Mean   :341.3                                    Mean   :4.094
##  3rd Qu.:516.0                                    3rd Qu.:6.000
##  Max.   :941.0                                    Max.   :6.000
##
##  First_Road_Number   Road_Type       Speed_limit    Junction_Detail
##  Min.   :   0.0    Min.   :-1.000   Min.   : 0.00   Min.   :-1.000
##  1st Qu.:   0.0    1st Qu.: 6.000   1st Qu.:30.00   1st Qu.: 0.000
##  Median :  62.0    Median : 6.000   Median :30.00   Median : 1.000
##  Mean   : 906.4    Mean   : 5.156   Mean   :38.97   Mean   : 2.238
##  3rd Qu.: 633.0    3rd Qu.: 6.000   3rd Qu.:50.00   3rd Qu.: 3.000
##  Max.   :9918.0    Max.   : 9.000   Max.   :70.00   Max.   : 9.000
##                                     NA's   :47
##  Junction_Control Second_Road_Class Second_Road_Number
##  Min.   :-1.000   Min.   :-1.000    Min.   :  -1.0
##  1st Qu.:-1.000   1st Qu.:-1.000    1st Qu.:   0.0
##  Median : 2.000   Median : 3.000    Median :   0.0
##  Mean   : 1.608   Mean   : 2.576    Mean   : 319.8
##  3rd Qu.: 4.000   3rd Qu.: 6.000    3rd Qu.:   0.0
##  Max.   : 4.000   Max.   : 6.000    Max.   :9999.0
##
##  Pedestrian_Crossing_Human_Control Pedestrian_Crossing_Physical_Facilities
##  Min.   :-1.000000                 Min.   :-1.0000
```

```
##  1st Qu.: 0.000000                 1st Qu.: 0.0000
##  Median : 0.000000                 Median : 0.0000
##  Mean   : 0.005251                 Mean   : 0.8015
##  3rd Qu.: 0.000000                 3rd Qu.: 0.0000
##  Max.   : 2.000000                 Max.   : 8.0000
##
##  Light_Conditions Weather_Conditions Road_Surface_Conditions
##  Min.   :-1.000   Min.   :-1.00      Min.   :-1.000
##  1st Qu.: 1.000   1st Qu.: 1.00      1st Qu.: 1.000
##  Median : 1.000   Median : 1.00      Median : 1.000
##  Mean   : 2.019   Mean   : 1.56      Mean   : 1.295
##  3rd Qu.: 4.000   3rd Qu.: 1.00      3rd Qu.: 2.000
##  Max.   : 7.000   Max.   : 9.00      Max.   : 5.000
##
##  Special_Conditions_at_Site Carriageway_Hazards Urban_or_Rural_Area
##  Min.   :-1.00000           Min.   :-1.00000    Min.   :-1.000
##  1st Qu.: 0.00000           1st Qu.: 0.00000    1st Qu.: 1.000
##  Median : 0.00000           Median : 0.00000    Median : 1.000
##  Mean   : 0.08974           Mean   : 0.05498    Mean   : 1.374
##  3rd Qu.: 0.00000           3rd Qu.: 0.00000    3rd Qu.: 2.000
##  Max.   : 7.00000           Max.   : 7.00000    Max.   : 3.000
##
##  Did_Police_Officer_Attend_Scene_of_Accident LSOA_of_Accident_Location
##  Min.   :-1.000                              Length:699163
##  1st Qu.: 1.000                              Class :character
##  Median : 1.000                              Mode  :character
##  Mean   : 1.225
##  3rd Qu.: 1.000
##  Max.   : 3.000
##
##      Date            TimeGroup         Accident_Index     Vehicle_Reference
##  Length:699163      Length:699163      Length:699163      Min.   :  1.000
##  Class :character   Class :character   Class :character   1st Qu.:  1.000
##  Mode  :character   Mode  :character   Mode  :character   Median :  1.000
##                                                           Mean   :  1.488
##                                                           3rd Qu.:  2.000
##                                                           Max.   :999.000
##
##   Vehicle_Type    Towing_and_Articulation Vehicle_Manoeuvre
##  Min.   :-1.000   Min.   :-1.000000       Min.   :-1.00
##  1st Qu.: 9.000   1st Qu.: 0.000000       1st Qu.: 9.00
##  Median : 9.000   Median : 0.000000       Median :18.00
##  Mean   : 8.714   Mean   : 0.006878       Mean   :13.37
##  3rd Qu.: 9.000   3rd Qu.: 0.000000       3rd Qu.:18.00
##  Max.   :98.000   Max.   : 5.000000       Max.   :18.00
##
##  Vehicle_Location_Restricted_Lane Junction_Location Skidding_and_Overturning
##  Min.   :-1.0000                  Min.   :-1.000    Min.   :-1.000
##  1st Qu.: 0.0000                  1st Qu.: 0.000    1st Qu.: 0.000
##  Median : 0.0000                  Median : 1.000    Median : 0.000
##  Mean   : 0.1509                  Mean   : 2.364    Mean   : 0.294
##  3rd Qu.: 0.0000                  3rd Qu.: 5.000    3rd Qu.: 0.000
##  Max.   : 9.0000                  Max.   : 8.000    Max.   : 5.000
##
```

```
##  Hit_Object_in_Carriageway Vehicle_Leaving_Carriageway
##  Min.   :-1.0000           Min.   :-1.0000
##  1st Qu.: 0.0000           1st Qu.: 0.0000
##  Median : 0.0000           Median : 0.0000
##  Mean   : 0.4397           Mean   : 0.5006
##  3rd Qu.: 0.0000           3rd Qu.: 0.0000
##  Max.   :12.0000           Max.   : 8.0000
##
##  Hit_Object_off_Carriageway First_Point_of_Impact Was_Vehicle_Left_Hand_Drive
##  Min.   :-1.0000            Min.   :-1.000        Min.   :-1.0000
##  1st Qu.: 0.0000            1st Qu.: 1.000        1st Qu.: 1.0000
##  Median : 0.0000            Median : 1.000        Median : 1.0000
##  Mean   : 0.7287            Mean   : 1.776        Mean   : 0.9775
##  3rd Qu.: 0.0000            3rd Qu.: 3.000        3rd Qu.: 1.0000
##  Max.   :11.0000            Max.   : 4.000        Max.   : 2.0000
##
##  Journey_Purpose_of_Driver Sex_of_Driver    Age_of_Driver    Age_Band_of_Driver
##  Min.   :-1.000            Min.   :-1.000   Min.   : -1.00   Min.   :-1.000
##  1st Qu.: 3.000            1st Qu.: 1.000   1st Qu.: 25.00   1st Qu.: 5.000
##  Median : 6.000            Median : 1.000   Median : 36.00   Median : 7.000
##  Mean   : 4.738            Mean   : 1.362   Mean   : 37.88   Mean   : 6.468
##  3rd Qu.: 6.000            3rd Qu.: 2.000   3rd Qu.: 50.00   3rd Qu.: 8.000
##  Max.   : 6.000            Max.   : 3.000   Max.   :101.00   Max.   :11.000
##
##  Engine_Capacity_CC Propulsion_Code   Age_of_Vehicle    Driver_IMD_Decile
##  Min.   :   -1      Min.   :-1.0000   Min.   : -1.000   Min.   :-1.000
##  1st Qu.:  124      1st Qu.: 1.0000   1st Qu.: -1.000   1st Qu.:-1.000
##  Median : 1360      Median : 1.0000   Median :  5.000   Median : 2.000
##  Mean   : 1323      Mean   : 0.9135   Mean   :  5.755   Mean   : 2.434
##  3rd Qu.: 1797      3rd Qu.: 2.0000   3rd Qu.: 10.000   3rd Qu.: 5.000
##  Max.   :91000      Max.   :12.0000   Max.   :105.000   Max.   :10.000
##
##  Driver_Home_Area_Type Vehicle_IMD_Decile Accident_Index     Vehicle_Reference
##  Min.   :-1.0000       Min.   :-1.000     Length:699163      Min.   :  1.000
##  1st Qu.: 1.0000       1st Qu.:-1.000     Class :character   1st Qu.:  1.000
##  Median : 1.0000       Median : 2.000     Mode  :character   Median :  1.000
##  Mean   : 0.9852       Mean   : 2.434                        Mean   :  1.488
##  3rd Qu.: 1.0000       3rd Qu.: 5.000                        3rd Qu.:  2.000
##  Max.   : 3.0000       Max.   :10.000                        Max.   :999.000
##
##  Casualty_Reference Casualty_Class Sex_of_Casualty  Age_of_Casualty
##  Min.   :  1.000    Min.   :1.00   Min.   :-1.000   Min.   : -1.00
##  1st Qu.:  1.000    1st Qu.:1.00   1st Qu.: 1.000   1st Qu.: 22.00
##  Median :  1.000    Median :1.00   Median : 1.000   Median : 33.00
##  Mean   :  1.406    Mean   :1.49   Mean   : 1.406   Mean   : 36.48
##  3rd Qu.:  2.000    3rd Qu.:2.00   3rd Qu.: 2.000   3rd Qu.: 50.00
##  Max.   :991.000    Max.   :3.00   Max.   : 2.000   Max.   :104.00
##
##  Age_Band_of_Casualty Casualty_Severity Pedestrian_Location Pedestrian_Movement
##  Min.   :-1.000       Min.   :1.000     Min.   :-1.000      Min.   :-1.0000
##  1st Qu.: 5.000       1st Qu.:3.000     1st Qu.: 0.000      1st Qu.: 0.0000
##  Median : 6.000       Median :3.000     Median : 0.000      Median : 0.0000
##  Mean   : 6.289       Mean   :2.842     Mean   : 0.714      Mean   : 0.5418
##  3rd Qu.: 8.000       3rd Qu.:3.000     3rd Qu.: 0.000      3rd Qu.: 0.0000
```

```
##  Max.   :11.000       Max.   :3.000     Max.   :10.000      Max.   : 9.0000
##
##  Car_Passenger     Bus_or_Coach_Passenger Pedestrian_Road_Maintenance_Worker
##  Min.   :-1.0000   Min.   :-1.00000       Min.   :-1.00000
##  1st Qu.: 0.0000   1st Qu.: 0.00000       1st Qu.: 0.00000
##  Median : 0.0000   Median : 0.00000       Median : 0.00000
##  Mean   : 0.2523   Mean   : 0.07744       Mean   : 0.07626
##  3rd Qu.: 0.0000   3rd Qu.: 0.00000       3rd Qu.: 0.00000
##  Max.   : 2.0000   Max.   : 4.00000       Max.   : 2.00000
##
##  Casualty_Type    Casualty_Home_Area_Type Casualty_IMD_Decile
##  Min.   :-1.000   Min.   :-1.000          Min.   :-1.000
##  1st Qu.: 3.000   1st Qu.: 1.000          1st Qu.: 1.000
##  Median : 9.000   Median : 1.000          Median : 4.000
##  Mean   : 7.281   Mean   : 0.986          Mean   : 3.714
##  3rd Qu.: 9.000   3rd Qu.: 1.000          3rd Qu.: 7.000
##  Max.   :98.000   Max.   : 3.000          Max.   :10.000
##
```

```r
# write.csv(All_Accident_Data, "C:/Users/darre/My Drive (dwolfe.data@gmail.com)/!Datasets/Amplify_Inter



Accidents_Vehicles_Casualties <- sqldf(
  "-- How many accidents had how many vehicles and casualties?
  select
  count(a.Accident_Index) AS Count_Incidents,
  v.Vehicle_Reference,
  c.Casualty_Reference

  from UK_AccidentData as a
  join UK_VehicleData as v
    on a.Accident_Index=v.Accident_Index
  join UK_CasualtyData as c
    on a.Accident_Index=c.Accident_Index
    and v.Vehicle_Reference=c.Vehicle_Reference

  group by
  v.Vehicle_Reference,
  c.Casualty_Reference

  order by
  v.Vehicle_Reference,
  c.Casualty_Reference")

# View(All_Accident_Data)
head(Accidents_Vehicles_Casualties)
```

```
##   Count_Incidents Vehicle_Reference Casualty_Reference
## 1          329081                 1                  1
## 2           41015                 1                  2
## 3           14104                 1                  3
## 4            4901                 1                  4
## 5            1663                 1                  5
```

```
## 6             568                 1                 6
```

```r
summary(Accidents_Vehicles_Casualties)
```

```
##  Count_Incidents    Vehicle_Reference Casualty_Reference
##  Min.   :     1.0   Min.   :  1.000   Min.   :  1.00
##  1st Qu.:     1.0   1st Qu.:  2.000   1st Qu.:  5.00
##  Median :     2.5   Median :  3.000   Median : 13.00
##  Mean   :  2515.0   Mean   :  9.932   Mean   : 22.28
##  3rd Qu.:    13.8   3rd Qu.:  8.000   3rd Qu.: 29.00
##  Max.   :329081.0   Max.   :999.000   Max.   :991.00
```

```r
# write.csv(Accidents_Vehicles_Casualties, "C:/Users/darre/My Drive (dwolfe.data@gmail.com)/!Datasets/A



Accidents_Light_Weather <- sqldf(
  "-- How many accidents had how many vehicles and casualties?
select
  count(a.Accident_Index) AS Count_Incidents,
  a.Light_Conditions,
  a.Weather_Conditions,
  a.Road_Surface_Conditions,
  v.Vehicle_Reference,
  c.Casualty_Reference

  from UK_AccidentData as a
  join UK_VehicleData as v
    on a.Accident_Index=v.Accident_Index
  join UK_CasualtyData as c
    on a.Accident_Index=c.Accident_Index
    and v.Vehicle_Reference=c.Vehicle_Reference

  where  a.Light_Conditions <> '-1'
  and a.Weather_Conditions <> '-1'
  and a.Road_Surface_Conditions <> '-1'

  group by
  a.Light_Conditions,
  a.Weather_Conditions,
  a.Road_Surface_Conditions,
  v.Vehicle_Reference,
  c.Casualty_Reference

  order by
  v.Vehicle_Reference,
  c.Casualty_Reference")

# View(All_Accident_Data)
head(Accidents_Light_Weather)
```

```
##   Count_Incidents Light_Conditions Weather_Conditions Road_Surface_Conditions
```

```
## 1         170261                1                  1                       1
## 2          20631                1                  1                       2
## 3            131                1                  1                       3
## 4           2040                1                  1                       4
## 5             34                1                  1                       5
## 6            325                1                  2                       1
##   Vehicle_Reference Casualty_Reference
## 1                 1                  1
## 2                 1                  1
## 3                 1                  1
## 4                 1                  1
## 5                 1                  1
## 6                 1                  1
```

```
summary(Accidents_Light_Weather)
```

```
##  Count_Incidents    Light_Conditions Weather_Conditions Road_Surface_Conditions
##  Min.   :     1.0   Min.   :1.000    Min.   :1.000      Min.   :1.000
##  1st Qu.:     1.0   1st Qu.:1.000    1st Qu.:1.000      1st Qu.:1.000
##  Median :     3.0   Median :4.000    Median :3.000      Median :2.000
##  Mean   :   262.6   Mean   :3.605    Mean   :3.935      Mean   :2.222
##  3rd Qu.:    12.0   3rd Qu.:6.000    3rd Qu.:7.000      3rd Qu.:3.000
##  Max.   :170261.0   Max.   :7.000    Max.   :9.000      Max.   :5.000
##  Vehicle_Reference Casualty_Reference
##  Min.   :  1.000   Min.   :  1.00
##  1st Qu.:  1.000   1st Qu.:  2.00
##  Median :  2.000   Median :  3.00
##  Mean   :  3.295   Mean   :  5.94
##  3rd Qu.:  3.000   3rd Qu.:  5.00
##  Max.   :999.000   Max.   :991.00
```

```
# write.csv(Accidents_Light_Weather, "C:/Users/darre/My Drive (dwolfe.data@gmail.com)/!Datasets/Amplify
```

Visualizations in Tableau

Initial Analysis and Views complete for now with a handful of key insights. I really want to see the data in visualizations. I saved the three cleaned tables as CSVs, as well as a few key views that might be interesting.

I will port these plus the codes tables to Tableau and see what happens when we begin visualizing these insights and looking for stuff I cannot do in R (because Google maps requires a paid API to use the plot map in R).

The Data Stories

The following data stories came to light throughout this analysis:

In 2015-2018, there were an aggregate total of 151.269 accidents and 593.391 casualties.

Accidents decreased dramatically during these years, possibly indicating a combination of safer vehicles, driver awareness, new laws, and other factors.

There were no indications that any particular month or season was more dangerous.

Weekdays are the most dangerous, especially Fridays.

While Fatal and Serious accidents do not show a preference for the day of the week, Slight injuries are reported at higher frequency on weekdays, with the highest on Fridays. Although the number of accidents have steadily dropped from 2015-2018 (the timeframe of the study).

Afternoon and Evening Commutes are most dangerous. The hours of 1p to 5p are the most likely times for an accident, followed closely by the hours of 5p to 8p. Most accidents occurred during daylight, following by well lit nights, which coincides with these findings.

Note: the "Evening Late/Late Night" categories could have been spread across times better. This was an oversight, and would be corrected upon an updated version.

Types of Roads and Lanes did not show any particularly interesting patterns, but two maps were included as reference.

While there were more male drivers than female, and more male casualties than female, the distribution of injuries Fatal, Serious, and Slight were correlated to those numbers and did not reveal anything interesting.

However, while males were the driver more often, females were injured as passangers more often. And rear-seat injuries were far less common than front seat injuries.

One vehicle Manoeuvre shows to be involved in more accidents than any other, "Going ahead", which indicates a state known as Road Hypnosis.(Shi et al. 2023)

Ages: While at first ages seem to show more accidents for those 26-35 with a bell curve on either side, this also correlates with the working population. Therefore, this may only be a reflection and not an indication. When comparing this to the population data (from WikiMedia Commons), this guess held true.

Speed Limit: One speed limit (30 MPH (or is that KMH?) has a higher representation of accidents than all others. This seems highly correlated with the fact that most accidents occur in urban city areas and during afternoon or commute hours.

Vehicles 26-35 years old played an outsized role in accidents, however, the age bands showed a bell curve, which is likely an indication of the population of vehicles available for accident incidents.

The Index of Multiple Deprivation (IMD) did show that accidents in deprived regions were higher, this may reflect multiple factors including access to safer vehicles and local budgets for road maintenance.

Accidents involving one vehicle were most common, followed by those involving two, with a representation for third vehicles. Very few involved more than three.

Surprisingly, most accidents occurred with one vehicle in daylight hours on dry single-lane roads, while driving straight ahead (not while navigating a turn or intersection).

This reminded me of a term I had heard called Road Hypnosis; which is especially important in North Idaho, USA. This occurs when the driver is taking a well known route and driving straight ahead. The human brain has evolved to only store important information, and a well-known straight route does not fall into this category. The brain goes into a form of autopilot, and drivers sometimes tell stories of having arrived at home after work without remembering the drive. In this auto-pilot state, one can miss clues that something has changed or even drift off the road entirely.

Conclusions

Some conditions can make it more likely for an accident to occur.

Driving in the afternoons or on the way home after work.

Driving on long straight routes in good weather (leading to carelessness).

Driving an older vehicle, in urban city regions, or in areas where the IMD index indicates the possibility of dangerous factors.

Driving in regions where the speed limit is lower may lead to accidents (for reasons not investigatged in this analysis).

Note: If a passenger, chosing the rear seat may prevent or lessen injury in case of an accident.

Presentation

- Technical Review: Provide the Word Doc and PDF version of the Rmd file for those wanting to look into the technical review.
  - UK Road Safety, Technical Analysis, Darrell Wolfe, (Temporary Demo Page - Topos Creative, LLC)
  - Google Drive: Word Doc & PDF Doc
- Tableau Public: Provide link to the Tableau Public presentation for the users to click and explore the data.
  - Link Here
- GitHub: Provide the GitHub link to this project:
  - GitHub_UK Road Safety Repo
- Google Drive: Provide link to files used for this exercise, including the Power Query transformation and re-organized CSVs.
  - UK Road Safety, Darrell Wolfe Analysis, Google Drive Files
- Some final data story format… Slides? Rmd to PDF? Undecided.

The End of this document… but not the data stories.

Shi, Huili, Longfei Chen, Xiaoyuan Wang, Bin Wang, Gang Wang, and Fusheng Zhong. 2023. "Research on Recognition of Road Hypnosis in the Typical Monotonous Scene." *Sensors (Basel, Switzerland)* 23 (3): 1701. https://doi.org/10.3390/s23031701.