Distracted Driving

A Look at
US Motor Vehicle Accident
Data

Angela Gunn

The Problem

OK I GET IT NOW...
NO **TEXTING** WHILE DRIVING!
WWW.CARTOONADAY.COM

Distracted Driving - 1

How often (and severe) are injuries to the distracted driver compared with other persons involved?

Distracted Driving - 2

What events are most common?

- 🚗 first/most harmful event
- 🚗 events prior to accident
- 🚗 manner of collision
- 🚗 location of accident

Distracted Driving - 3

Which distractions are related to the most accidents?

- 🚗 phone
- 🚗 eating
- 🚗 smoking
- 🚗 adjusting climate controls

Data

Source: National Highway Traffic Safety Administration (NHTSA)

Database: National Accident Sampling System (NASS) General Estimates System (GES)

- Survey of police-reported traffic crashes in USA

2013 Summary Statistics	Crashes	Vehicles	People
Unweighted (Weighted)	34,516 (5,584,299)	59,350 (9,880,971)	85,779 (13,533,802)

Challenges **(beyond my learning curve)**

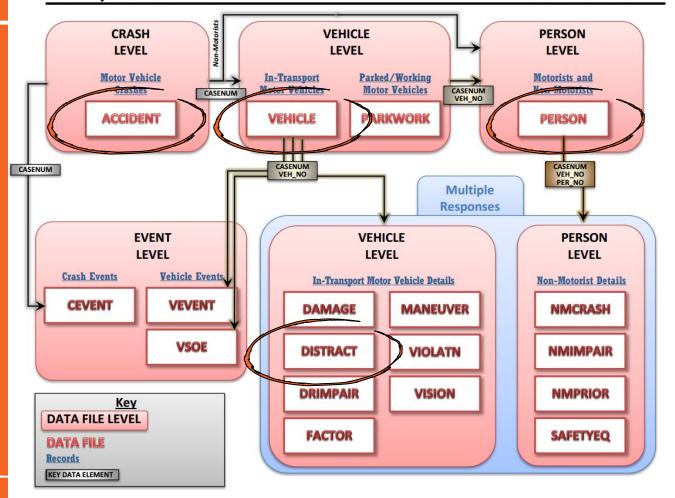| Challenge 1 | Challenge 2 | Challenge 3 |

Files! Files! Files

- 🚗 multiple files per year (current = 18)
- 🚗 not all files available every year
  - 🚨 new
  - 🚨 discontinued
  - 🚨 missing

Codes! Codes! Codes!

- 🚗 Everything is a code
- 🚗 Codes may (and do) change every year
  - 🚨 8 in one year has meaning, the next means "unknown"

Model & Deliver

- 🚗 Create a model with consistent codes and structure
- 🚗 Create ways of accessing the data

4

5

─

Transition

- 🚗 Integers -> String Values
- 🚗 Roll-up Values and Split Elements
- 🚗 Every Year Different

Table	GES Elements	Project Elements
Accident	53	21
Vehicle	72	18
Person	32	15
Distract	8	5

Accident
- Case_Number
- Year
- Month
- Region
- Day_of_Week
- Hour
- Num_Vehicles
- Num_ParkWork
- Persons_Ped
- Persons_NotTransit
- Persons_InTransit
- First_Harm_Event
- Manner_Collision
- Within_Interchange
- Relation_to_Junction
- Intersection_Type
- Work_Zone
- Weather1
- Weather2
- Weather3
- Max_Injury
- Num_Injury
- Alcohol_Involved

Vehicle
- Case_Number
- Vehicle_Number
- Year
- Num_Occupants
- Body_Type
- Model_Year
- Extent_Damage
- Most_Harm_Event
- Num_Injury
- Max_Injury
- Driver_Drinking
- Speeding
- Travel_Speed
- Pre_Event_Movement
- Critical_Event_PreCrash

Person
- Case_Number
- Vehicle_Number
- Person_Number
- Year
- Age
- Sex
- Person_Type
- Injury_Severity
- Seating_Position
- Restraint_Use
- Drinking
- Drugs
- Striking_Vehicle_Number

Distract
- Case_Number
- Vehicle_Number
- Year
- Factor

Storage & Retrieval

Storage	Retrieval
Hive Database	**PySpark Scripts**
🚓 Tables partitioned by year	🚓 Create CSV
	🚨 full table retrieval
🚓 Stored as text files in HDFS	🚨 2 tables joined
	🚓 Create Histogram
🚓 Views useful in querying distracted driving	🚨 Any table.field
	🚨 Specify Year

For Each Year of Data:

Extract data and move to HDFS	→	Create external tables based on files extracted	→	Define tables for target as per ERD	→	Script to transform data to common format	→	Script to populate tables, partitioned by year

GOAL: One Command to Run it All

Limitations

- Static data source

  - Each year demands special attention.

  - 2014 data just added; changes to 2 of fields of interest

- Summarized data

  - Rolled up data loses information

- Choice of database

  - Postgres or other relational database would make some things easier

  - Hard-coded strings in code; one code file per year

Dear Owen,
As a user, I would like to be able to...

Scale the Solution

🚗 Add data from more files

    🚨 Used 4/18 available files (based on 2013 year)

🚗 Add more years

    🚨 Data collection started in 1988; distraction data started 2002
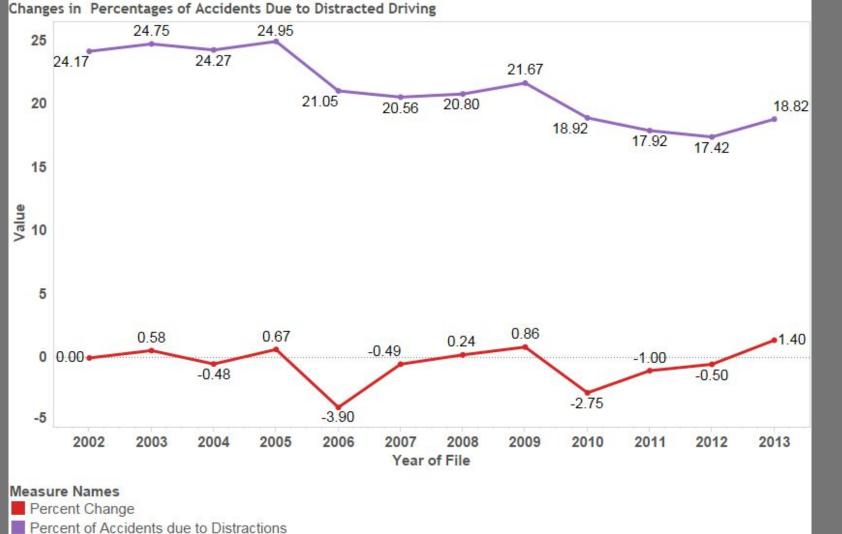
🚗 Determine better model for generic querying

🚗 Add data from other countries

    🚨 Does anyone know French?

Questions & Results

Selection of Images
From Tableau on
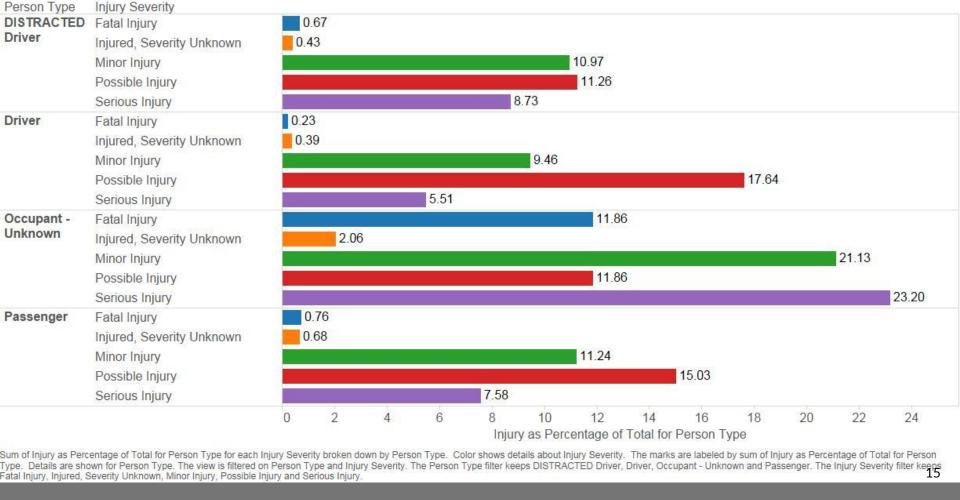Data Exported from
Accident Project

Changes in Percentages of Accidents Due to Distracted Driving

Pre-Crash Events by Driver Attention

	Driver Attention	
Pre-Crash Event	Not Distracted	Distracted
Animal in Road	2.37	0.45
Object in Road	0.75	0.32
Other Event	3.70	5.03
Other Motor Vehicle Encrouching Into Lane	28.46	4.52
Other Motor Vehicle in Lane	30.88	34.99
Pedestrian, Pedacyclist or Other Non-Motorist in Road	2.60	2.82
Unknown	0.73	0.61
Vehicle Crossing Intersection	5.04	9.52
Vehicle Decelerating	2.24	0.18
Vehicle Loss of Control	7.16	6.31
Vehicle Travelling on Road Edge	8.54	23.13
Vehicle Turning at Junction	7.54	12.10

Sum of Percent Drivers in Driver Attention Category broken down by Driver Attention vs. Pre-Crash Event. Color shows sum of Percent Drivers in Driver Attention Category. The marks are labeled by sum of Percent Drivers in Driver Attention Category. The view is filtered on Driver Attention, which keeps Distracted and Not Distracted.

Severity of Damage to Vehicles by Driver Attention

Injuries to People Involved in Distracted Accidents

Person Type	Injury Severity	
DISTRACTED Driver	Fatal Injury	0.67
	Injured, Severity Unknown	0.43
	Minor Injury	10.97
	Possible Injury	11.26
	Serious Injury	8.73
Driver	Fatal Injury	0.23
	Injured, Severity Unknown	0.39
	Minor Injury	9.46
	Possible Injury	17.64
	Serious Injury	5.51
Occupant - Unknown	Fatal Injury	11.86
	Injured, Severity Unknown	2.06
	Minor Injury	21.13
	Possible Injury	11.86
	Serious Injury	23.20
Passenger	Fatal Injury	0.76
	Injured, Severity Unknown	0.68
	Minor Injury	11.24
	Possible Injury	15.03
	Serious Injury	7.58

Injury as Percentage of Total for Person Type

Sum of Injury as Percentage of Total for Person Type for each Injury Severity broken down by Person Type. Color shows details about Injury Severity. The marks are labeled by sum of Injury as Percentage of Total for Person Type. Details are shown for Person Type. The view is filtered on Person Type and Injury Severity. The Person Type filter keeps DISTRACTED Driver, Driver, Occupant - Unknown and Passenger. The Injury Severity filter keeps Fatal Injury, Injured, Severity Unknown, Minor Injury, Possible Injury and Serious Injury.

Sample of Distraction Factors over Time As Percent of all Distractions