

# Data Analysis of the Fatality Analysis Reporting System (FARS) from the National Highway Traffic Safety Administration (1975-1981)



U.S. Department  
of Transportation  
**National Highway  
Traffic Safety  
Administration**



DOT HS 812 602

October 2018

## **Fatality Analysis Reporting System (FARS) Analytical User's Manual 1975-2017**

Name: Audrey Ward

Grade 9

4-H Club: Sunshiners

Project: Computer

Division: Advanced

# Introduction

Data science is becoming more and more popular with the internet and interconnected technologies producing lots of data. Computers and data scientists need to know how to deal with all the massive amounts of data. It can provide exciting new careers for people who have learned about all the different technologies for interpreting Big Data.

My dad is the director of The Data Mine at Purdue, and he encouraged me to work through the materials they used in their introductory seminar to gain a familiarity with these cutting-edge tools. The curriculum for The Data Mine introductory seminar is publicly accessible, so I set out to work through some of it this year.

## Method

I learned how to use the Scholar computing cluster at Purdue. I learned with the R programming language and many of its commands. I also learned how to read and analyze large datasets withing the Jupyter Notebook environment.

I worked through many of the tutorials for The Data Mine seminar, but I also used my new tools to explore the data in

original ways. My project contains answers to some assigned questions as well as individual explorations.

While working through The Data Mine materials I got to practice working with lots of big datasets, i.e., from Disney parks, weather station data, Goodreads....

I chose to work with the FARS dataset because I liked all the data provided. It allowed me to explore a lot of different situations and new information about accidents and drunk driving which are particularly relevant for high schoolers as they learn to drive.

This dataset contains 43,445 accident entries with 45 variables. This provides 1,955,025 data points for analysis.

## References:

The companion book for the course The Data Mine 190:

<https://thedatamine.github.io/the-examples-book/>

Specially I worked through many of the projects for the fall 2020 semester found here:

<https://thedatamine.github.io/the-examples-book/fall2020-projects.html>

This manual was provided for the FARS dataset:

<https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812602>

I worked through many of the tutorials suggested by The Data Mine course. I also spent time solving the questions posed at the ends of the projects as well as extending my understanding with further explorations of the graphs and data.