Final Project DSC530

Milestone 1: Choose and look at datasets.

I’ve chosen a dataset released by the NHTSA for US traffic accidents in 2015. After a nearly 50 year decline, fatality rates increased by 7.2%. We would like to

Milestone 2:

Are younger drivers more likely to be involved in fatal accidents? Codeset: <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/AUXF_A>

Interesting variables: A\_D15\_19 et all – accidents involving younger drivers

A\_DOW (Day of week) A\_Dist (distracted driver) A\_DROWSY (Drowsy driver) A\_SPCRASH A\_RU (rural or urban) A\_TOD(Time of day)

I’d be interested to see if DOW, DIST, DROWSY, SPCRASH, RU, TOD were better insights than the age of the driver. I can probably compare the two.

Variables used: dayofweek, holiday, age, numfatal, sex. We’re looking at day of the week, if it’s a holiday, the number of fatalities, the age of the driver, and the sex of the driver.

Our question: are younger drivers more likely to be involved in accidents that are deadlier?

Variables: