FARS Data Set

Analysis and Visualization

Author: Karthik Ranganathan

What is it?

The program collects data for analysis of traffic safety crashes to identify problems, and evaluate countermeasures leading to reducing injuries and property damage resulting from motor vehicle crashes.

Contains Accident Information on
- Date
- Weather
- Person
- Vehicle
- Location

Author: Karthik Ranganathan

Methods and Work

- Cleaning and Wrangling of more than 100 coded elements

- 600 page catalogue to decode

- Understanding the method of data collection

- Finding the key elements for Exploratory Data Analysis

- Visualizing the right information

Author: Karthik Ranganathan

Raw Data

STATE	ST_CASE	VE_FORMS	VEH_NO	PER_NO	STR_VEH	COUNTY	DAY	MONTH	HOUR	MINUTE	RUR_URB	FUNC_SYS	HARM_EV	MAN_COLL	SCH_BUS	MAKE	MAK_MOD	BODY_TYP	MOD_YEAR	TOW_VEH	SPEC_USE
1	10001	1	1	1	0	127	1	1	2	40	1	3	35	0	0	12	12481	31	2003	0	0
1	10002	1	1	1	0	83	1	1	22	13	1	1	34	0	0	49	49040	4	2006	0	0
1	10003	1	1	1	0	11	1	1	1	25	1	3	42	0	0	20	20037	4	2008	0	0
1	10003	1	1	2	0	11	1	1	1	25	1	3	42	0	0	20	20037	4	2008	0	0
1	10004	1	1	1	0	45	4	1	0	57	1	4	53	0	0	12	12481	31	2005	0	0
1	10005	2	1	1	0	45	7	1	7	9	2	3	12	6	0	69	69054	3	2006	0	0
1	10005	2	2	1	0	45	7	1	7	9	2	3	12	6	0	85	85881	63	2015	0	0
1	10006	1	1	1	0	111	8	1	9	59	1	5	1	0	0	12	12471	30	2006	0	0
1	10006	1	1	2	0	111	8	1	9	59	1	5	1	0	0	12	12471	30	2006	0	0
1	10007	1	1	1	0	89	8	1	18	33	2	3	43	0	0	23	23471	30	2002	0	0
1	10007	1	1	2	0	89	8	1	18	33	2	3	43	0	0	23	23471	30	2002	0	0
1	10008	1	1	1	0	73	3	1	21	30	2	5	8	0	0	99	99999	99	9999	0	0
1	10009	1	1	1	0	117	13	1	8	0	2	3	23	0	0	12	12481	31	2000	0	0
1	10010	2	1	1	0	33	5	1	18	45	1	4	12	2	0	12	12017	4	2004	0	0
1	10010	2	2	1	0	33	5	1	18	45	1	4	12	2	0	20	20481	31	2003	0	0
1	10011	2	1	1	0	83	7	1	20	55	2	3	12	7	0	12	12017	4	2005	0	0
1	10011	2	2	1	0	83	7	1	20	55	2	3	12	7	0	20	20401	14	2002	0	0
1	10012	3	1	1	0	95	9	1	6	35	2	3	12	6	0	20	20421	15	1997	0	0
1	10012	3	2	1	0	95	9	1	6	35	2	3	12	6	0	20	20481	31	2003	0	0
1	10012	3	2	2	0	95	9	1	6	35	2	3	12	6	0	20	20481	31	2003	0	0
1	10012	3	2	3	0	95	9	1	6	35	2	3	12	6	0	20	20481	31	2003	0	0
1	10012	3	2	4	0	95	9	1	6	35	2	3	12	6	0	20	20481	31	2003	0	0
1	10012	3	3	1	0	95	9	1	6	35	2	3	12	6	0	20	20037	4	2010	0	0
1	10013	1	1	1	0	87	10	1	10	55	1	5	1	0	0	35	35039	4	2007	0	0
1	10014	1	1	1	0	11	11	1	6	13	1	4	1	0	0	49	49421	15	1999	0	0
1	10015	1	1	1	0	127	11	1	8	30	1	3	1	0	0	23	23481	31	2004	0	0
1	10016	2	1	1	0	81	13	1	17	20	2	4	12	6	0	20	20481	31	1999	0	0
1	10016	2	2	1	0	81	13	1	17	20	2	4	12	6	0	59	59034	4	1998	0	0
1	10017	1	1	1	0	97	14	1	14	40	2	5	42	0	0	49	49040	4	2008	0	0

wing 1 to 23 of 81,622 entries

Author: Karthik Ranganathan

Cleaned Data

State.District	Postal.Code	REST_USE	AIR_BAG	DRINKING	DRUGS	DOA	RACE
Alabama	AL	None Used	Deployed: Front	Unknown-Police Reported	Unknown-Police Reported	Died at Scene	White
Alabama	AL	None Used	Not Deployed	No	No	Died at Scene	White
Alabama	AL	None Used	Deployed: Front	Yes	Unknown-Police Reported	Died at Scene	Black
Alabama	AL	None Used	Deployed: Front	Not Reported	Not Reported	Not Applicable	Not Applicable
Alabama	AL	None Used	Not Deployed	Yes	Unknown-Police Reported	Died at Scene	White
Alabama	AL	Shoulder and Lap Belt	Not Deployed	No	No	Died at Scene	White
Alabama	AL	Shoulder and Lap Belt	Not Applicable	No	No	Not Applicable	Not Applicable
Alabama	AL	None Used	Not Deployed	No	No	Not Applicable	White
Alabama	AL	Shoulder and Lap Belt	Not Deployed	Not Reported	Not Reported	Not Applicable	Not Applicable
Alabama	AL	Shoulder and Lap Belt	Not Deployed	Not Reported	Not Reported	Not Applicable	White
Alabama	AL	Shoulder and Lap Belt	Not Deployed	Not Reported	Not Reported	Not Applicable	Not Applicable
Alabama	AL	Unknown	Deployment Unknown	Unknown-Police Reported	Unknown-Police Reported	Not Applicable	Not Applicable
Alabama	AL	Shoulder and Lap Belt	Deployed: Front	Unknown-Police Reported	Unknown-Police Reported	Died at Scene	White
Alabama	AL	Shoulder and Lap Belt	Deployed: Front	No	No	Died at Scene	White
Alabama	AL	None Used	Deployed: Front	No	No	Not Applicable	Not Applicable
Alabama	AL	Shoulder and Lap Belt	Not Deployed	No	No	Not Applicable	Not Applicable
Alabama	AL	None Used	Not Deployed	No	No	Died at Scene	White
Alabama	AL	Shoulder and Lap Belt	Deployed: Front	Unknown-Police Reported	Unknown-Police Reported	Not Applicable	Not Applicable
Alabama	AL	Shoulder and Lap Belt	Not Deployed	No	No	Not Applicable	Not Applicable
Alabama	AL	Shoulder and Lap Belt	Not Deployed	Not Reported	Not Reported	Not Applicable	Not Applicable
Alabama	AL	None Used	Not Applicable	Not Reported	Not Reported	Not Applicable	Not Applicable
Alabama	AL	None Used	Not Applicable	Not Reported	Not Reported	Not Applicable	Not Applicable
Alabama	AL	None Used	Deployed: Front	Unknown-Police Reported	Unknown-Police Reported	Died at Scene	White
Alabama	AL	None Used	Deployed: Combination	No	No	Died at Scene	White
Alabama	AL	None Used	Not Deployed	No	No	Died at Scene	White
Alabama	AL	None Used	Not Deployed	Unknown-Police Reported	Unknown-Police Reported	Died at Scene	White
Alabama	AL	Shoulder and Lap Belt	Not Deployed	No	No	Not Applicable	Not Applicable
Alabama	AL	Shoulder and Lap Belt	Not Deployed	No	No	Not Applicable	White
Alabama	AL	Shoulder and Lap Belt	Deployed: Front	No	No	Not Applicable	White
Alabama	AL	None Used	Not Applicable	Yes	Yes	Not Applicable	Not Applicable

I to 30 of 81,622 entries
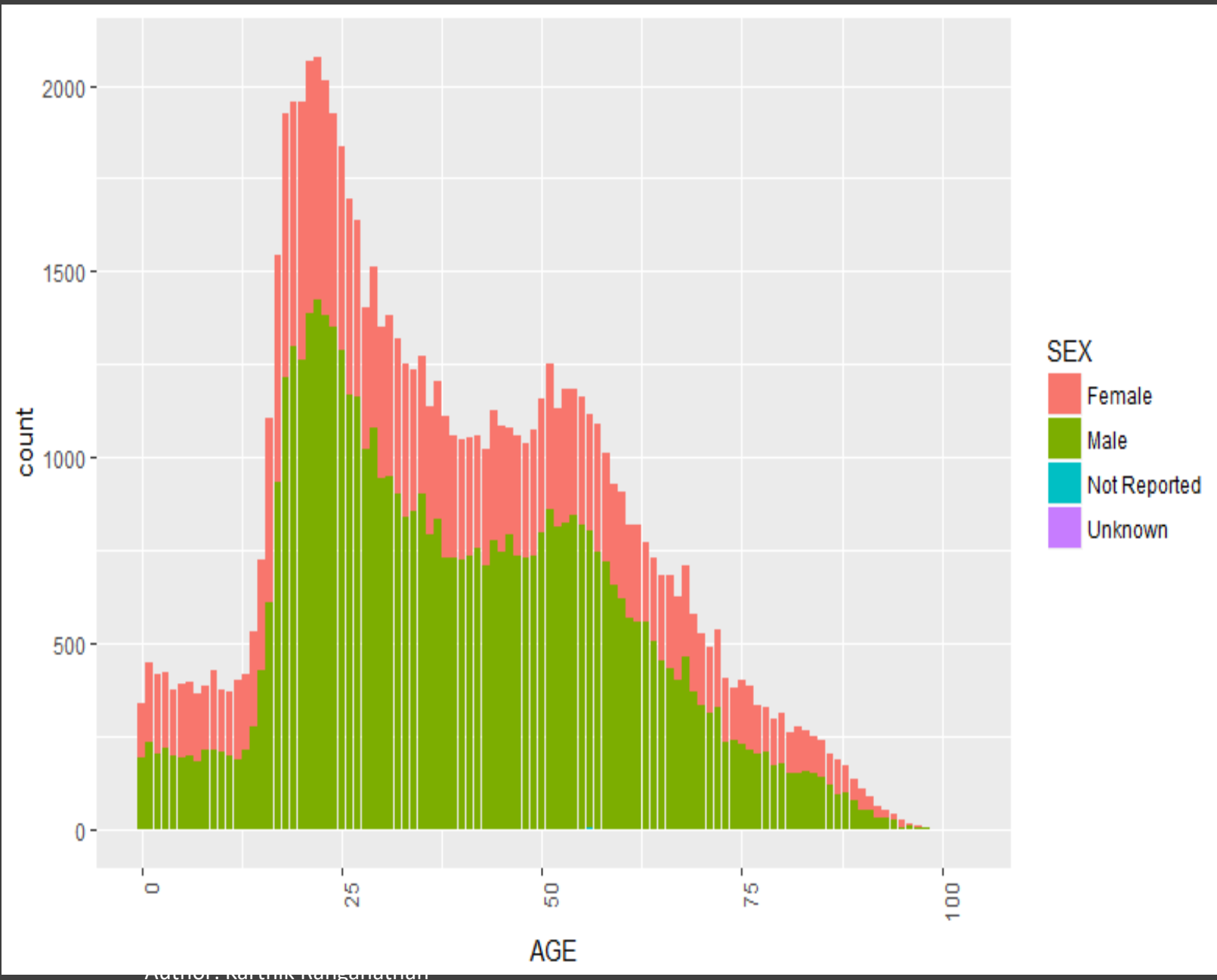
Author: Karthik Ranganathan

Findings

- The data gives a lot of insight into multiple variables playing a role in the crash

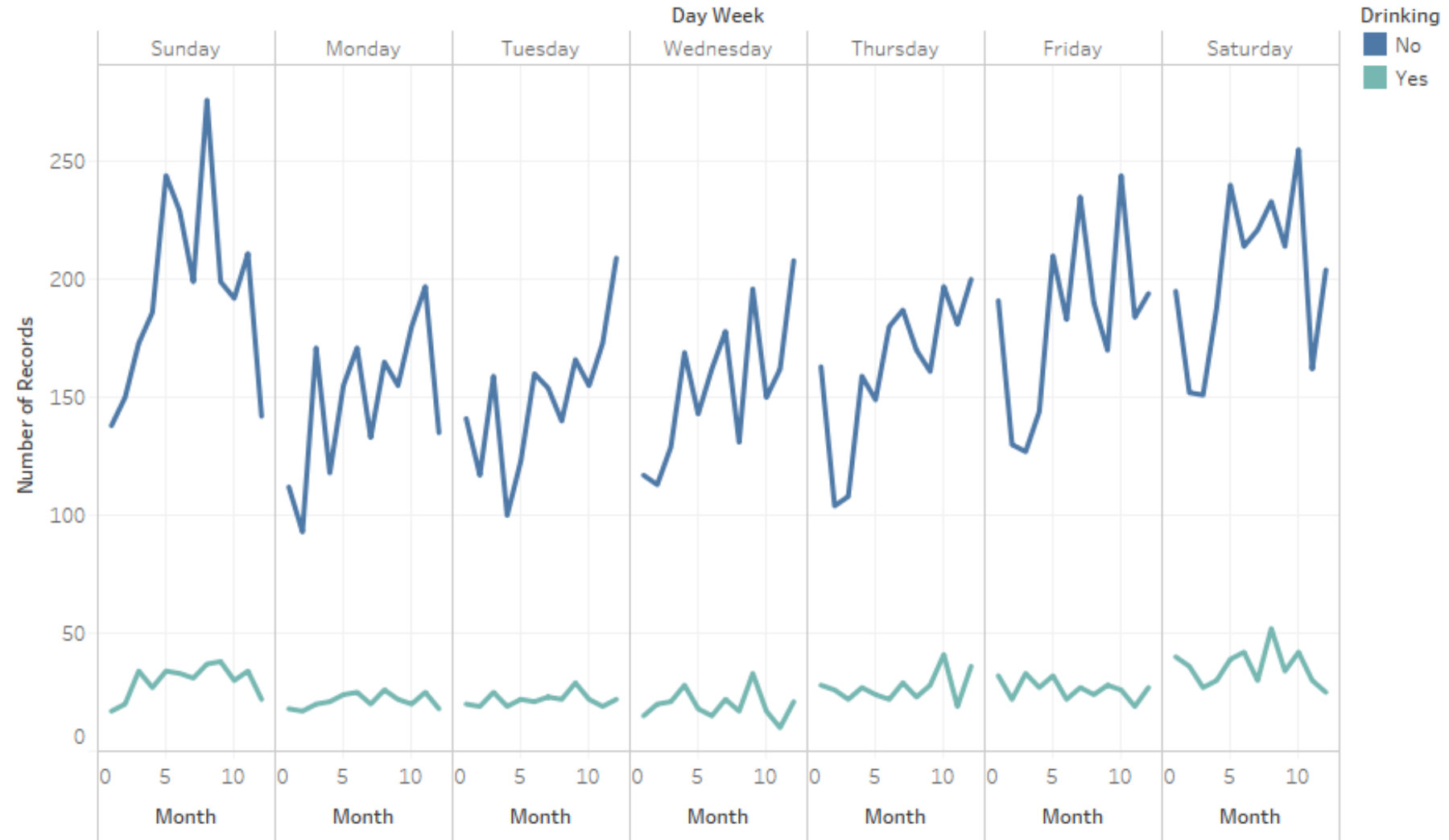- By visualizing more than 50 factors, here are some of the most impactful insights

Author: Karthik Ranganathan

Age

- Overs the years trends have been showing increase in accident rate in age group of 65 years and above

- Aging population?

Note: The high rate of accident on Saturdays and Sundays but surprisingly in the non drinking category

Initial Impact

- An unlikely inference can be made from this data

- We can see that sitting on **the rear right side** of the car will be most safest for initial impacts (given that the seat belt is fastened)

- This is purely a inference and many other data points needs to be collected in order to verify and prove the claim

Author: Karthik Ranganathan
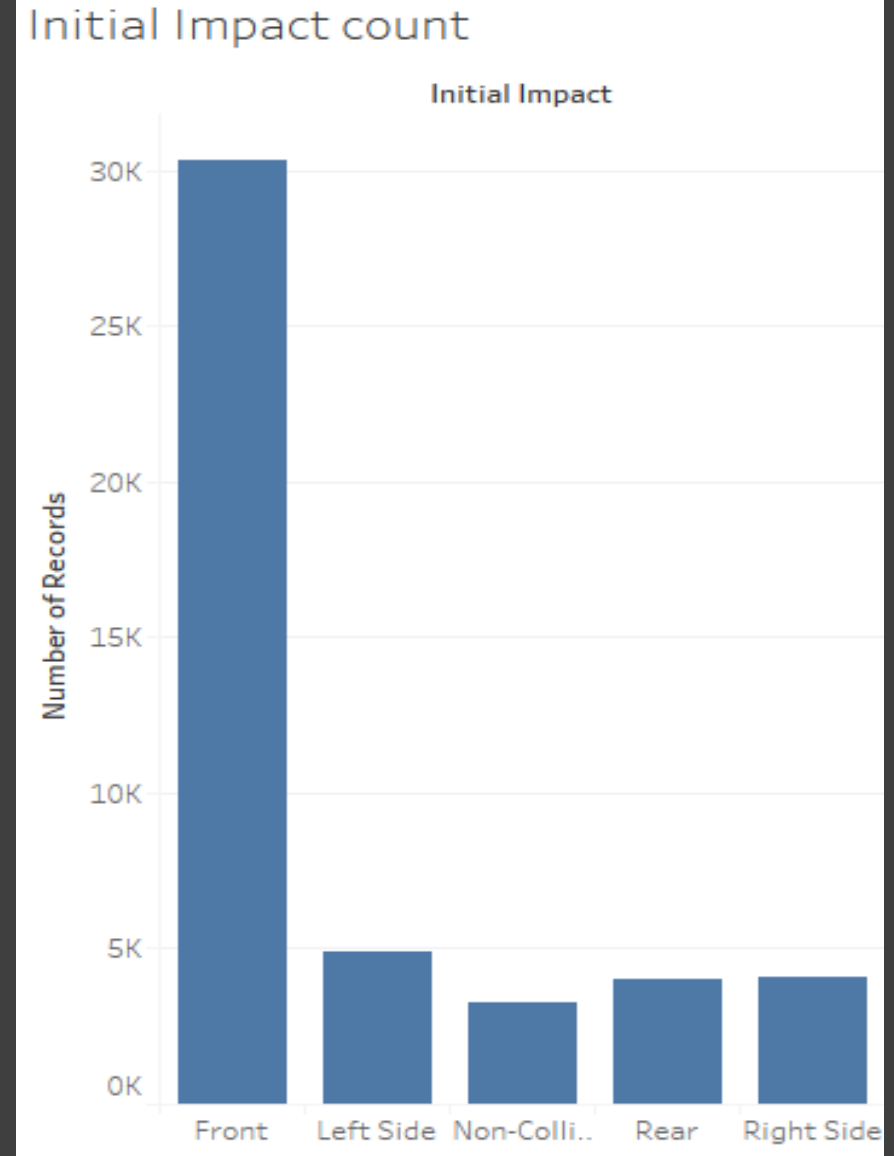
How do the States fair?
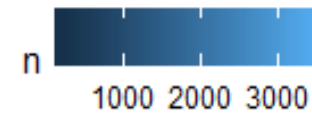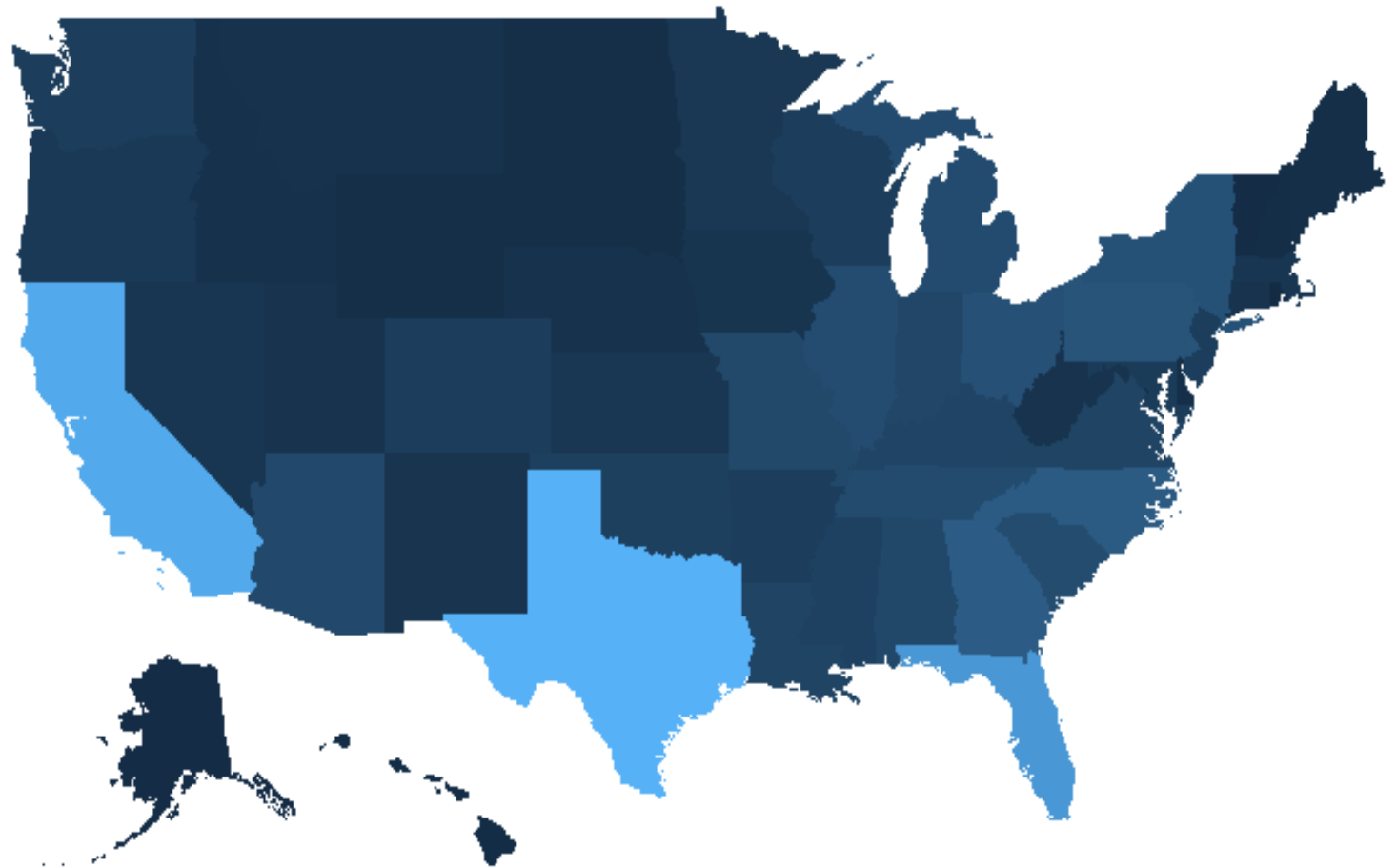
Author: Karthik Ranganathan

Better View
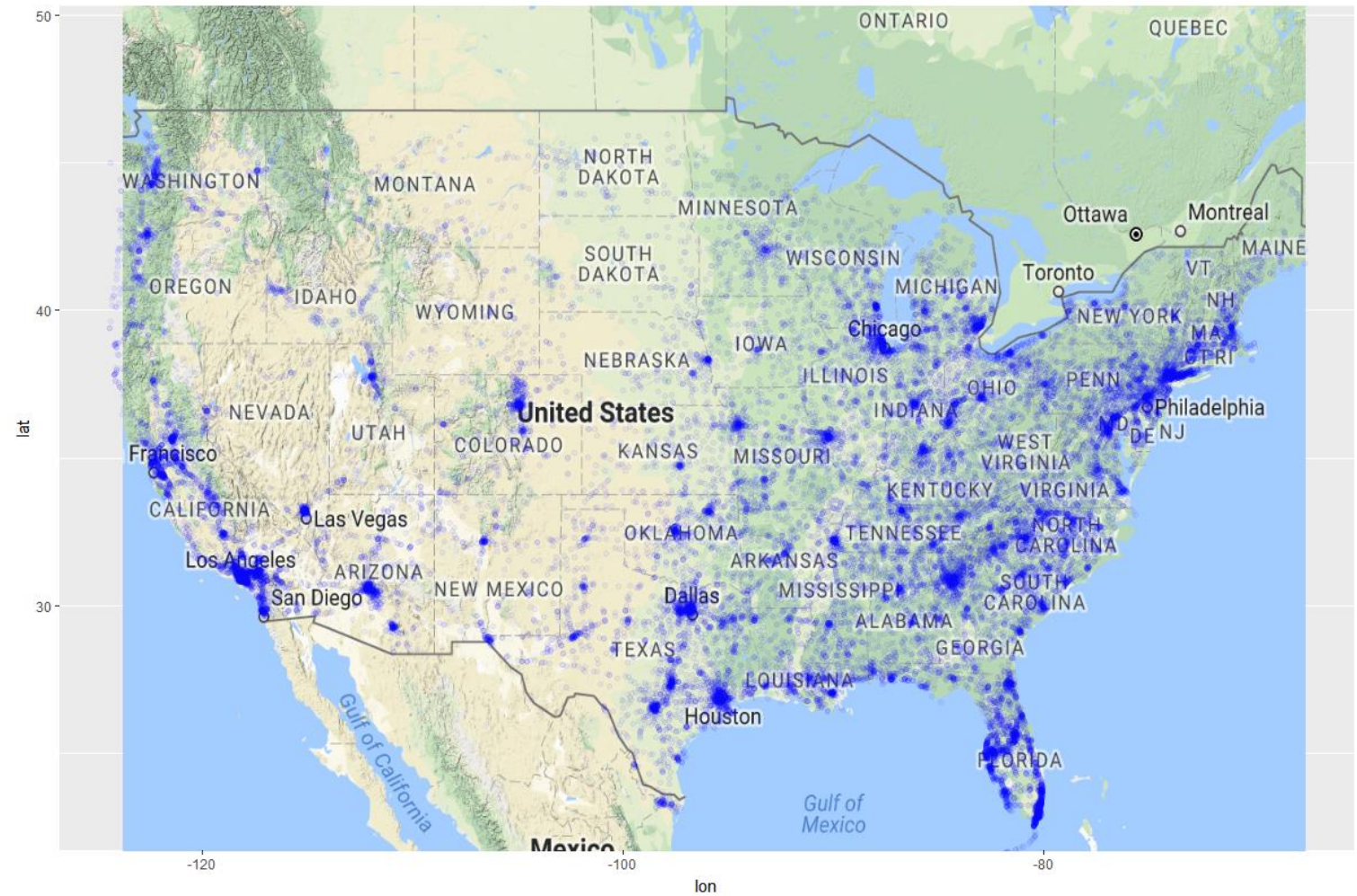
Author: Karthik Ranganathan

California, Florida and Texas

Author: Karthik Ranganathan
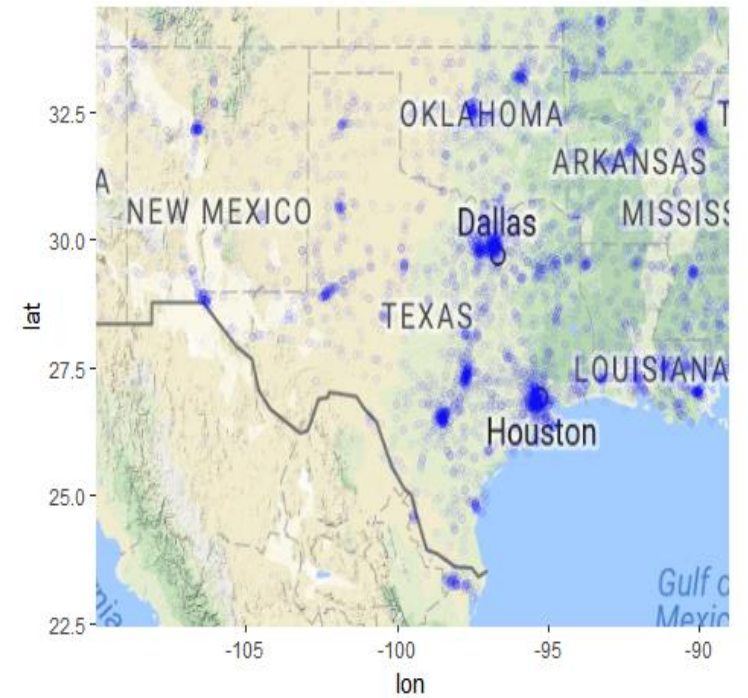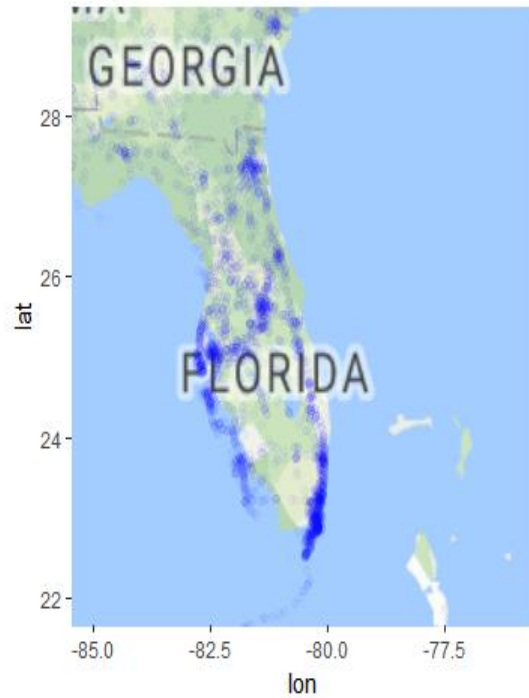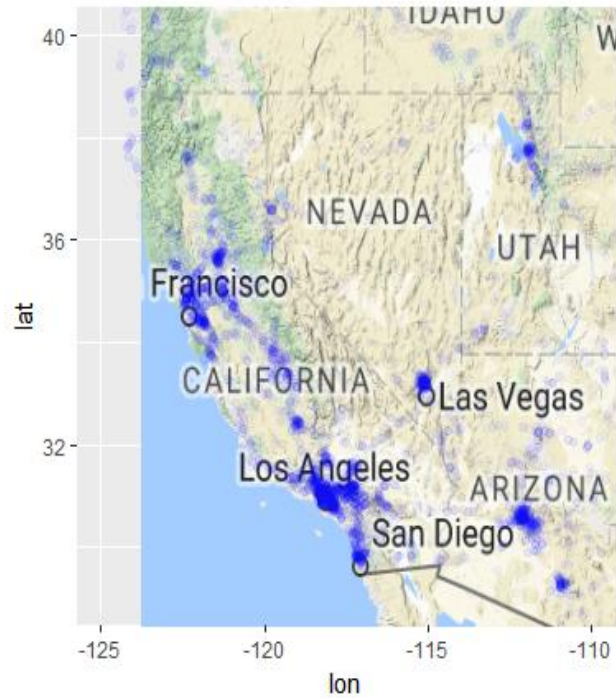
Deaths due to road accidents per state

State	TrafficFatalities
texas	3536
california	3102
florida	2494
north carolina	1284
pennsylvania	1195
georgia	1164
new york	1041
ohio	1006
tennessee	963
illinois	924
michigan	901
south carolina	823
alabama	820
arizona	773

Is it telling us everything? ➡

Total	2014	2015	Percent Change
Traffic Fatalities	32,744	35,485	7.72
Population (x1000)	318,856	321,418	0.80
Licensed Drivers (x1000)	214,093	218,086	1.87
Registered Vehicles (x1000)	260,354	263,608	1.25
Vehicle Miles Traveled (x 100 million)	3,025,655	3,095,373	2.30

Author: Karthik Ranganathan

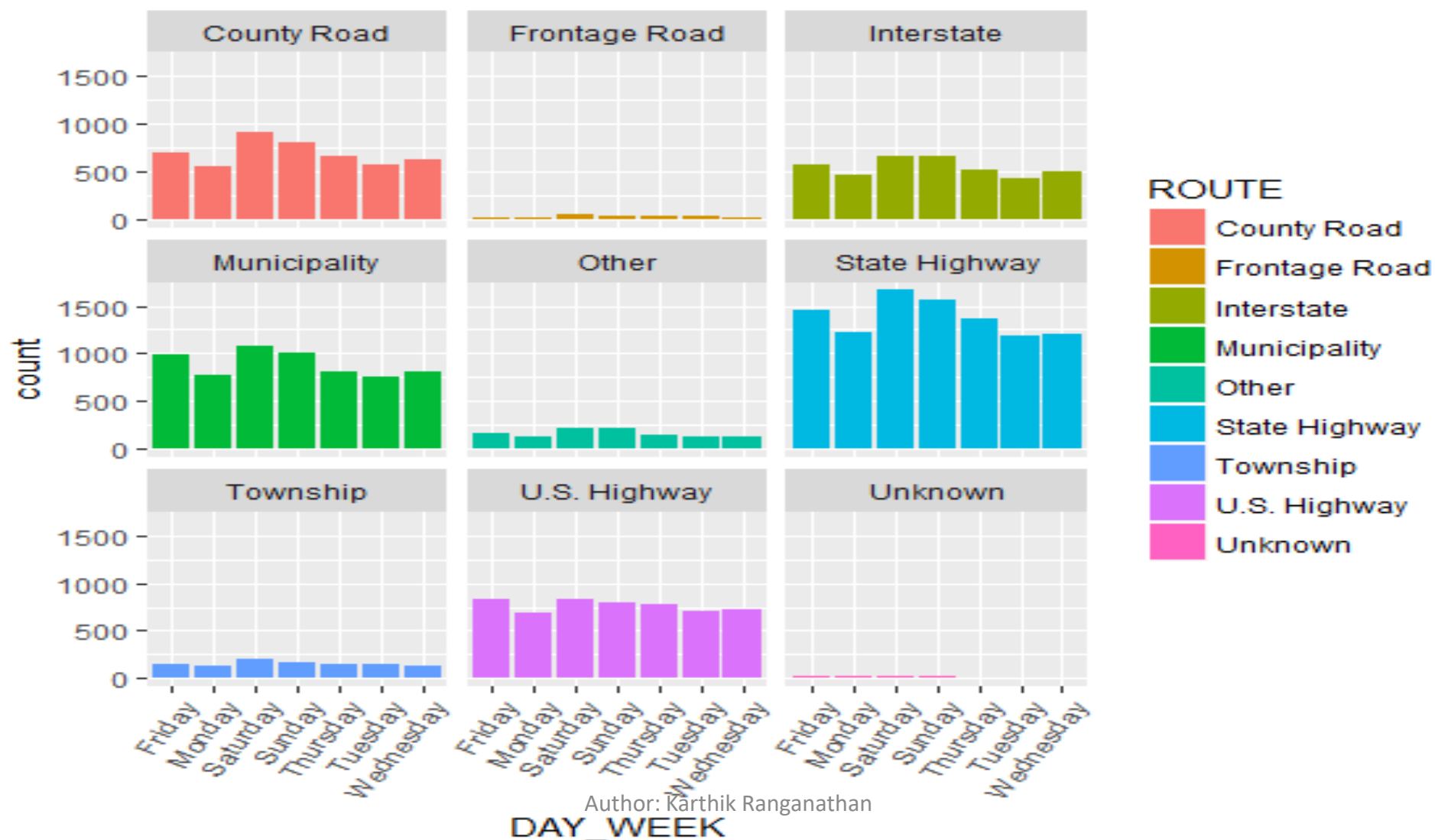Fatalities per 100,000 licensed drivers
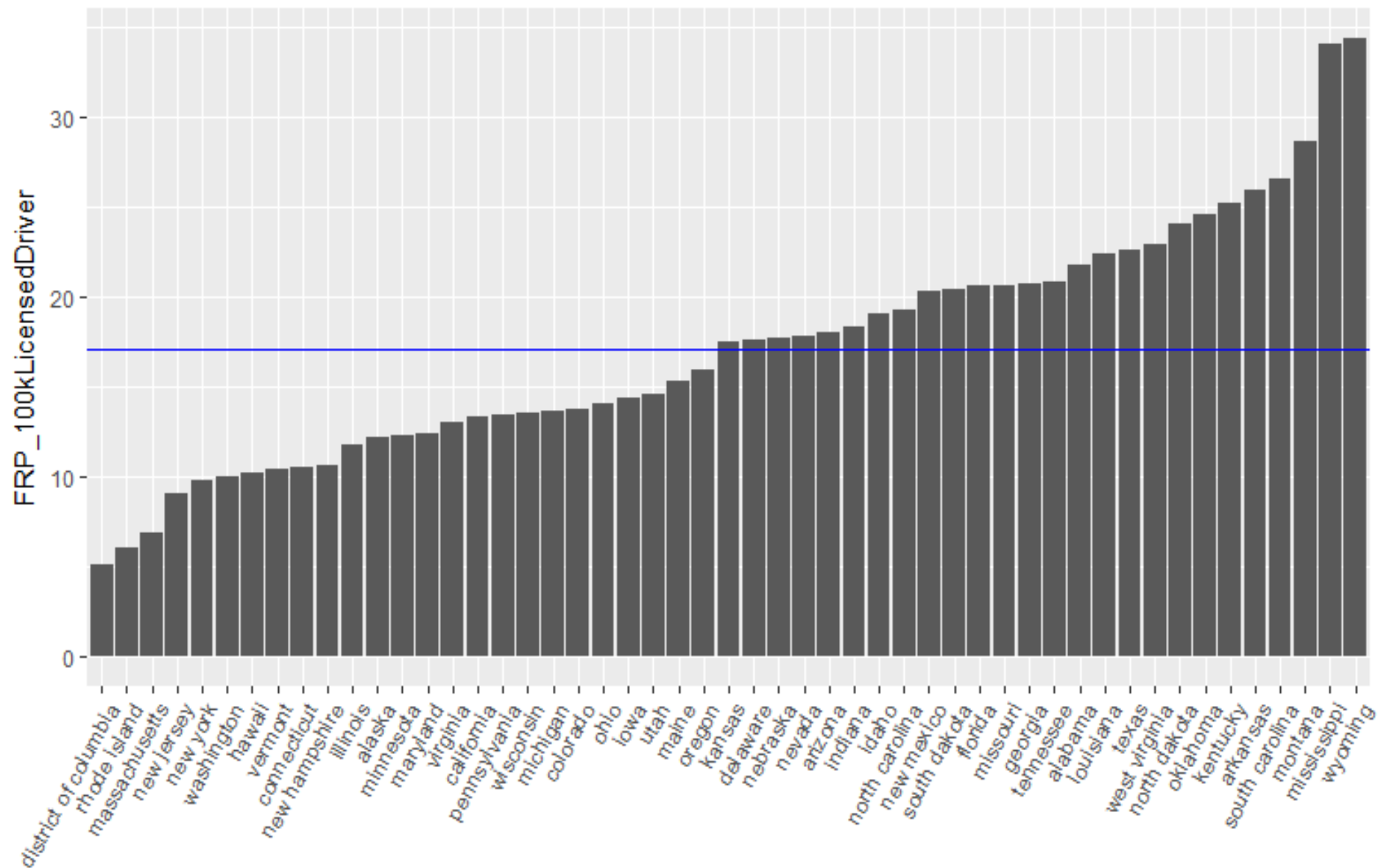
Author: Karthik Ranganathan

Fatalities per 100 million miles travelled
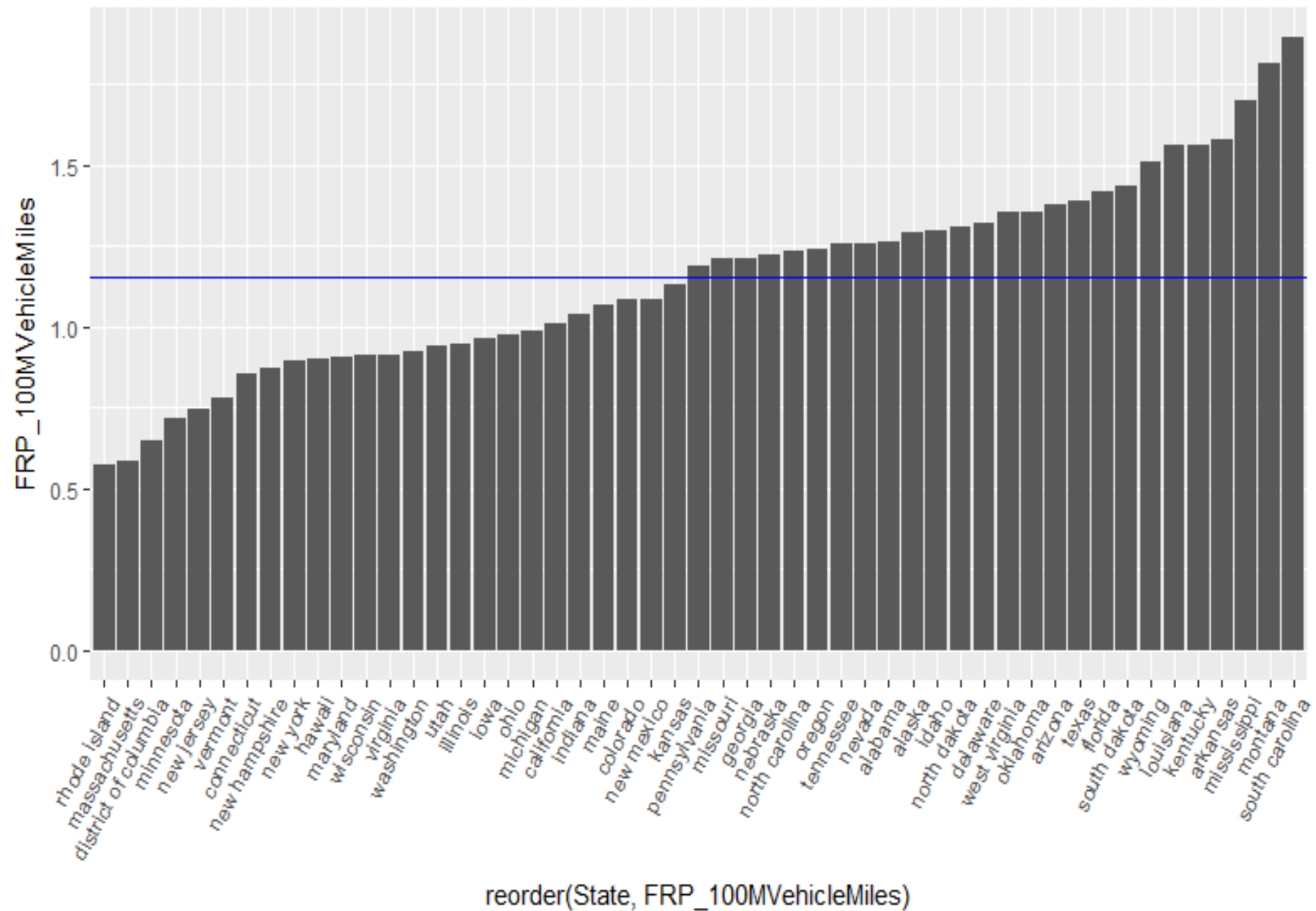
Rhode Island and Massachusetts have the most safest drivers. Really??

Author: Karthik Ranganathan

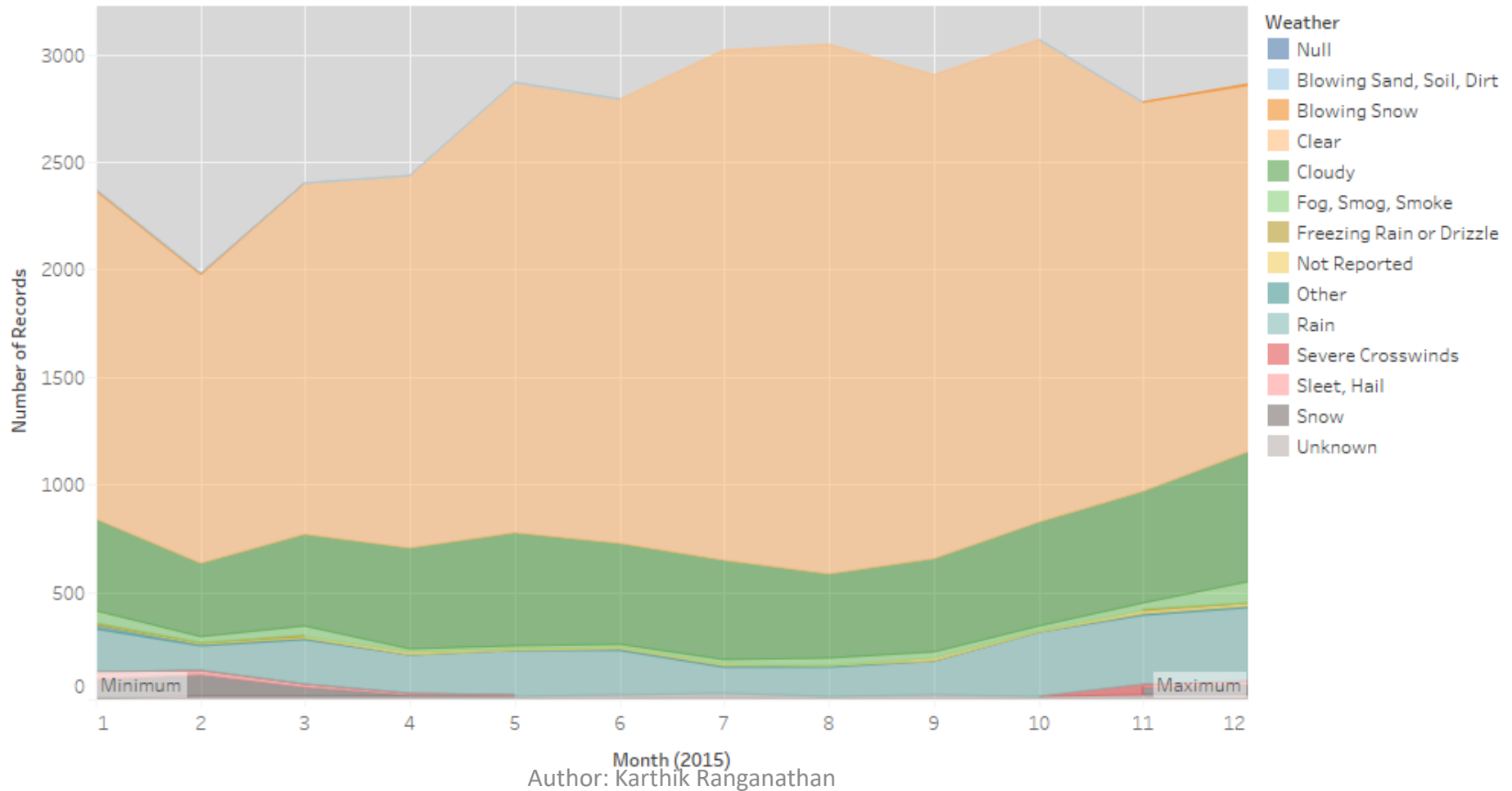Which route to take on which day?

Author: Karthik Ranganathan
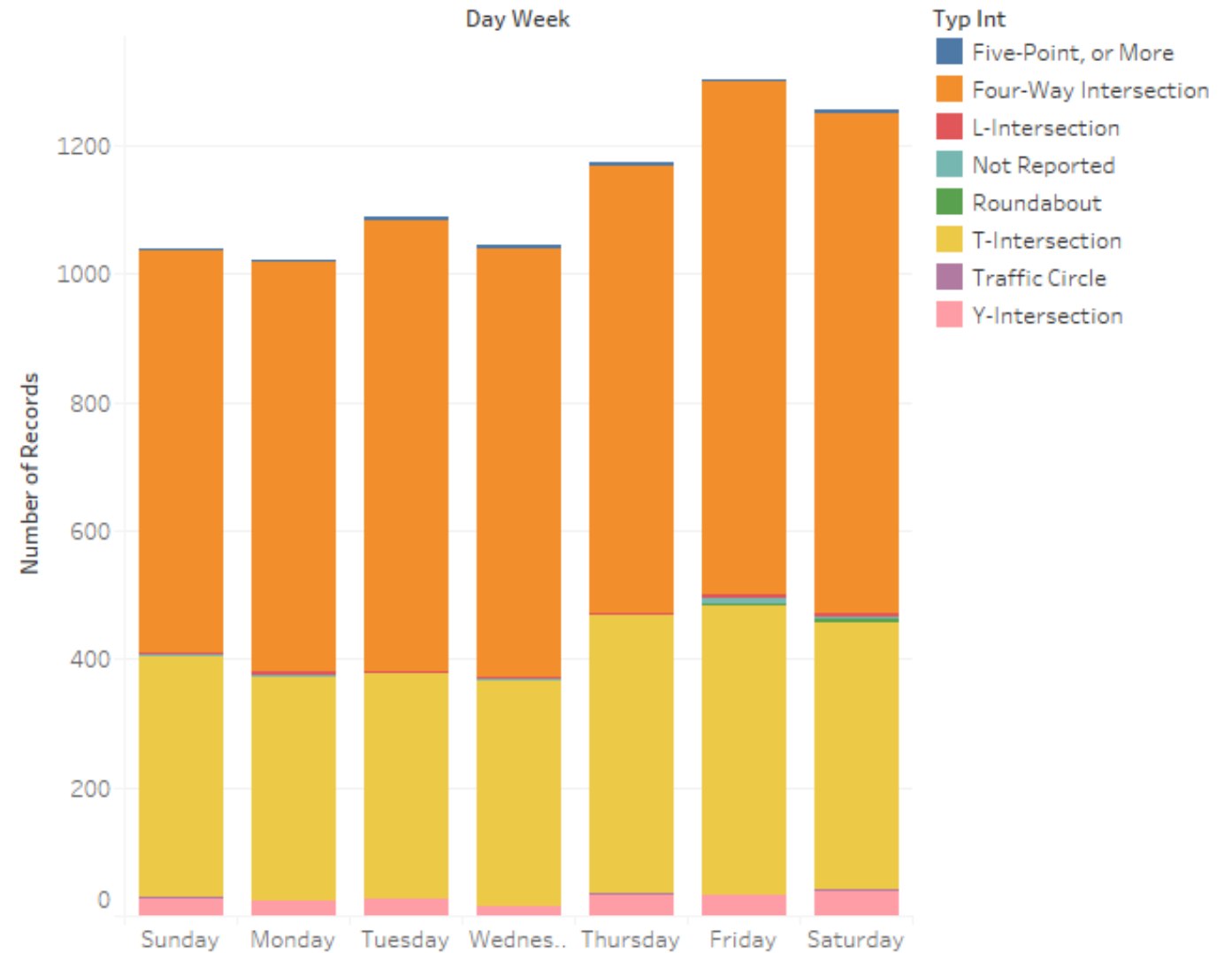
Reasons

Author: Karthik Ranganathan

Weather Condition?
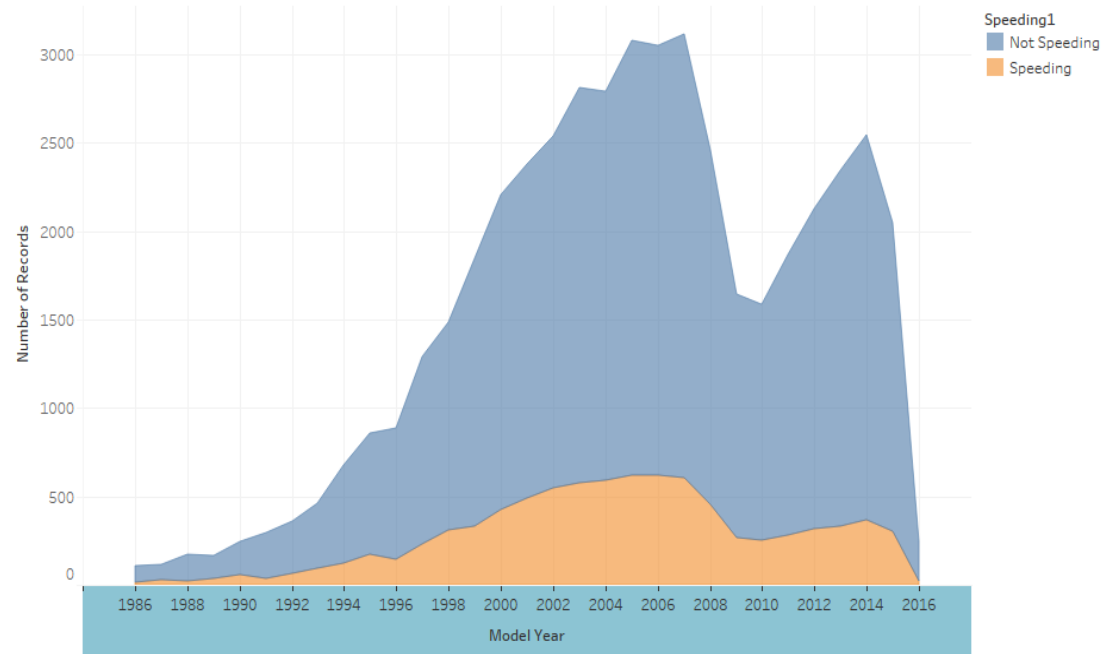

Fatalities - Weather Condition and Month

Author: Karthik Ranganathan

Car Model Year and Speeding (2015)

Car Model Year and Speeding (2014)

May be Car Model Year?

Author: Karthik Ranganathan
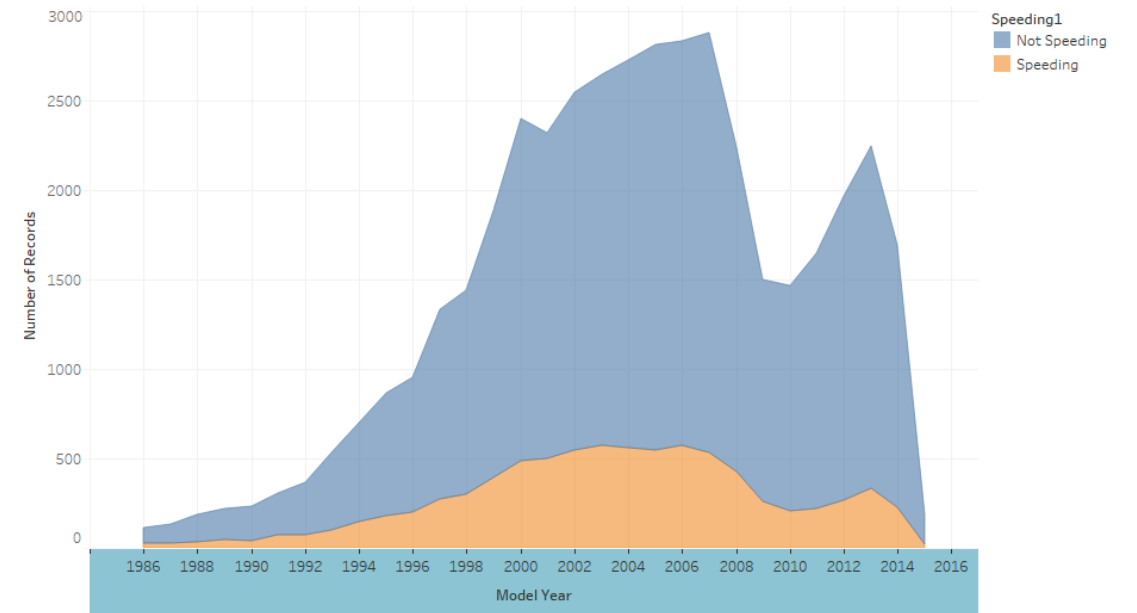
Conclusions

- Need to increase surveillance of accident prone areas by thorough analysis of the heat map

- Increase in safety checks of cars which are in the age bracket of 8-17 years

- Higher police presence and installation of speed cameras in Four-Way Intersection and T-Intersection

- Run driver education campaigns in Mississippi, Montana, South Carolina and, Wyoming to tackle the high rate of incidence

- Also major policy changes need to be implemented in California, Florida and, Texas with respect to Speed restriction and heavier fines

Author: Karthik Ranganathan

Take Away

- The devil is in the details

- Could have looked into more data which included injuries and other incidence reports

- Every situation in unique and the categorization of data has to be more complex

- Generalized lot of data to simplify the visualization

- By toning the data down to specific roads which have higher accident rates we can suggest detailed plans to reduce the rate

- Additional Years could give us information on various trends and accurate insights

Author: Karthik Ranganathan

Thank You

Author: Karthik Ranganathan