

Accident Data Analysis

"Which driver demographics (age, gender) are most involved in vehicular accidents, and are there correlations with specific vehicle types?"

Analyzing accident statistics data

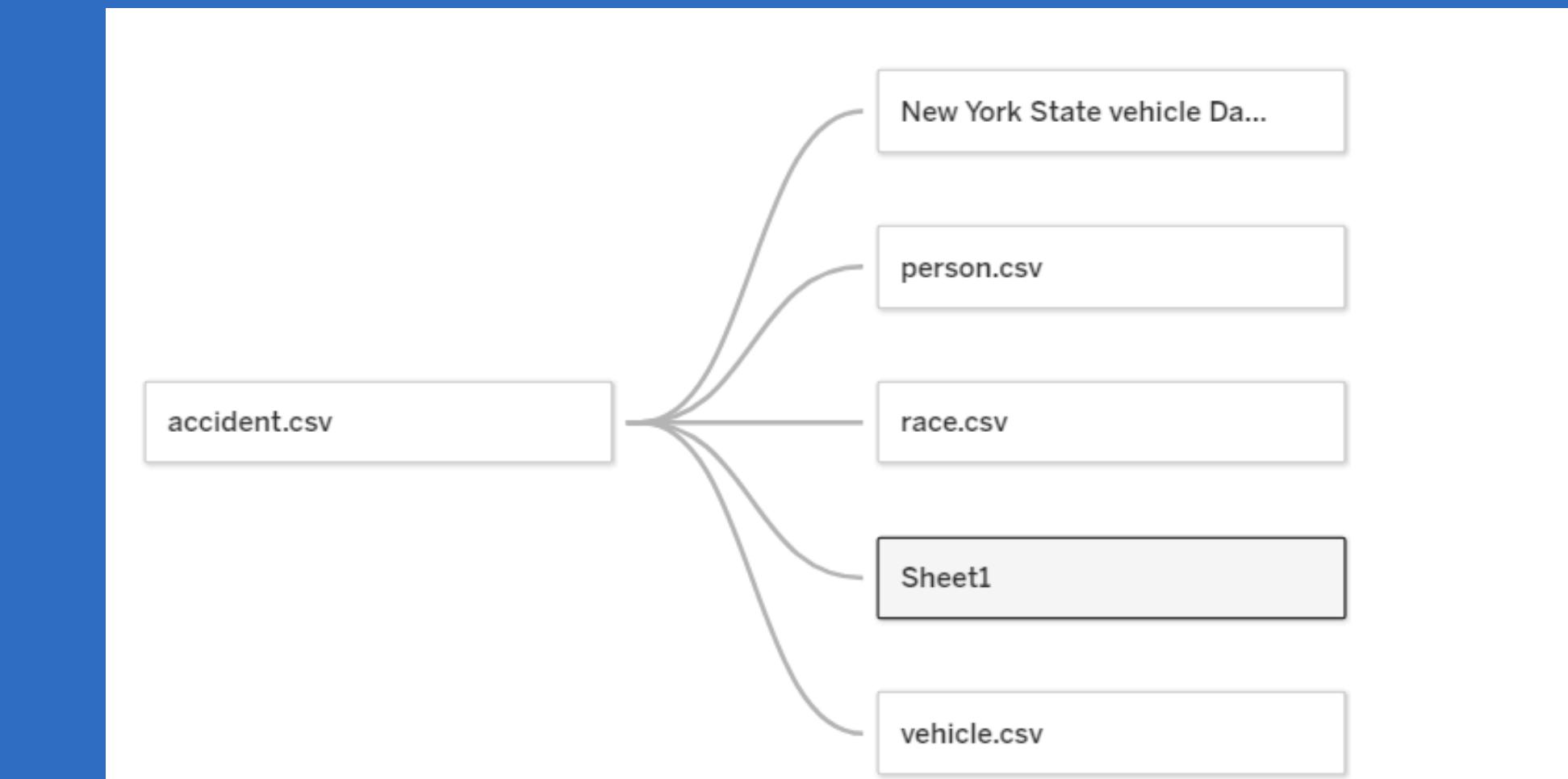




This project proposes a detailed examination of how driver demographics (age, gender) relate to vehicular accidents and investigates if there are discernible patterns connecting specific demographics to certain vehicle types involved in these incidents. Focusing on the 'Person', 'accident' and 'Vehicle' datasets provided by NHTSA, the analysis will illuminate correlations between driver demographics and specific vehicle types involved in accidents.



Data used for analysis



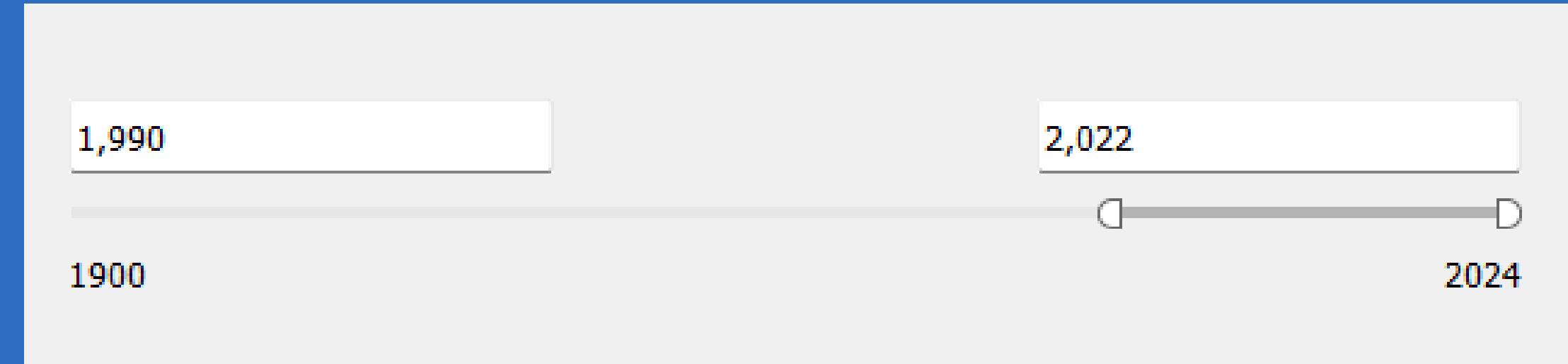
Data Collection and Cleaning

Filter the data based on states, person type, model year

Filter out unknown, missed or incorrect values

Segment the age variable using a 10 year interval and create a new variable

- 10 - 19
- 20 - 29
- 30 - 39
- 40 - 49
- 50 - 59
- 60 - 69
- 70 - 79
- 80 - 89
- 90 - 99



Additional Data used for standardization

Used an additional dataset from New York State Department of Transportation specifying every vehicles in the state of New York with car type

used the data gathered from the additional data to standardize values and created a new dataset containing state, car make, total number of that car make in the state and total number of fatality cases reported with these car makes

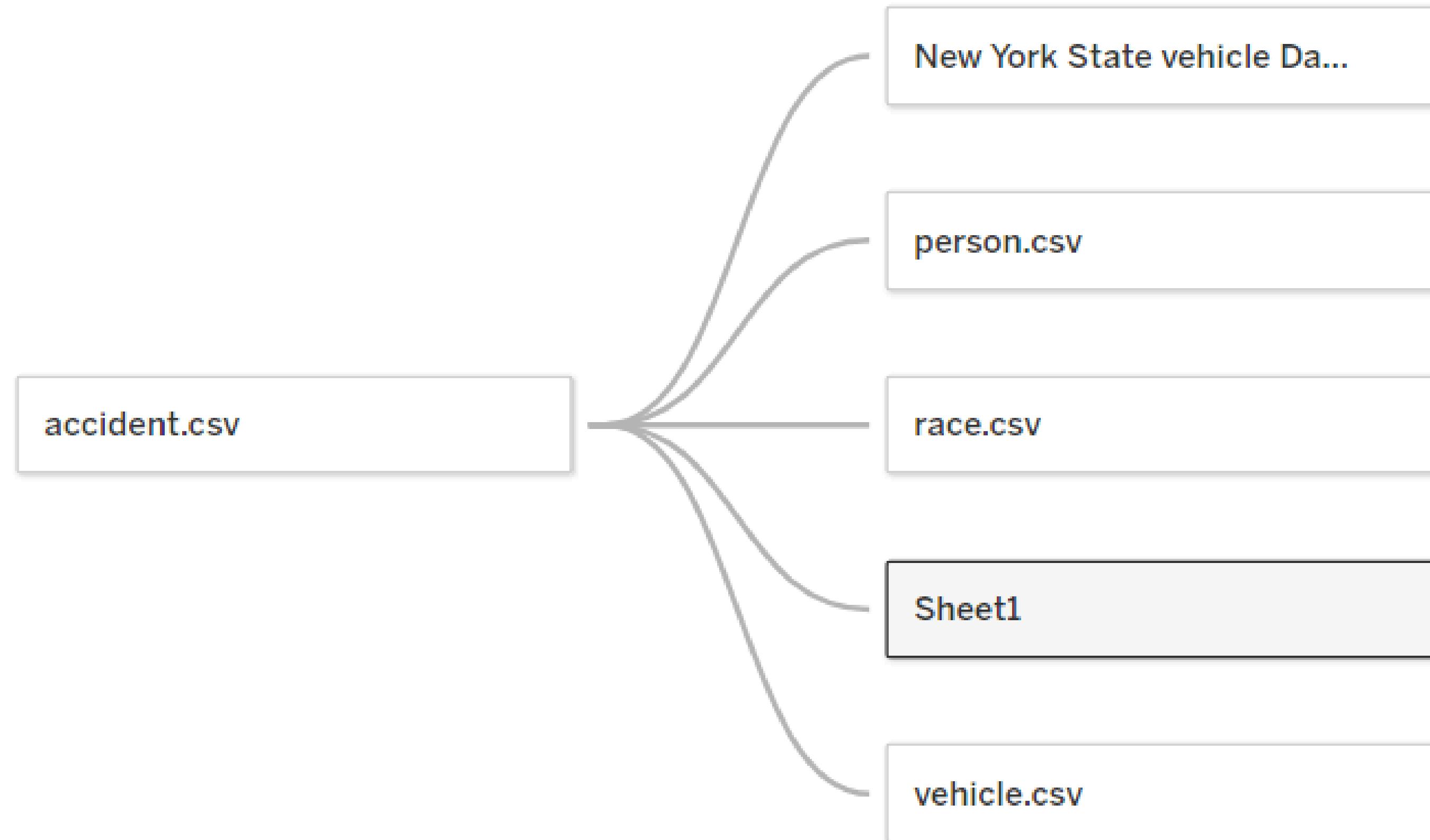
New Dataset that I created

State	Make	num Cars	St case count	num Cars per case
NY	Ford	1,099,871	134	0.01218%
NY	Chevrolet	1,084,657	149	0.01374%
NY	Toyota	1,258,103	116	0.00922%
NY	Nissan	659,137	89	0.01350%
NY	Dodge	293,352	57	0.01943%
NY	Honda	1,203,991	153	0.01271%
NY	Harley Davidson	122,362	58	0.04740%
NY	GMC	261,751	38	0.01452%



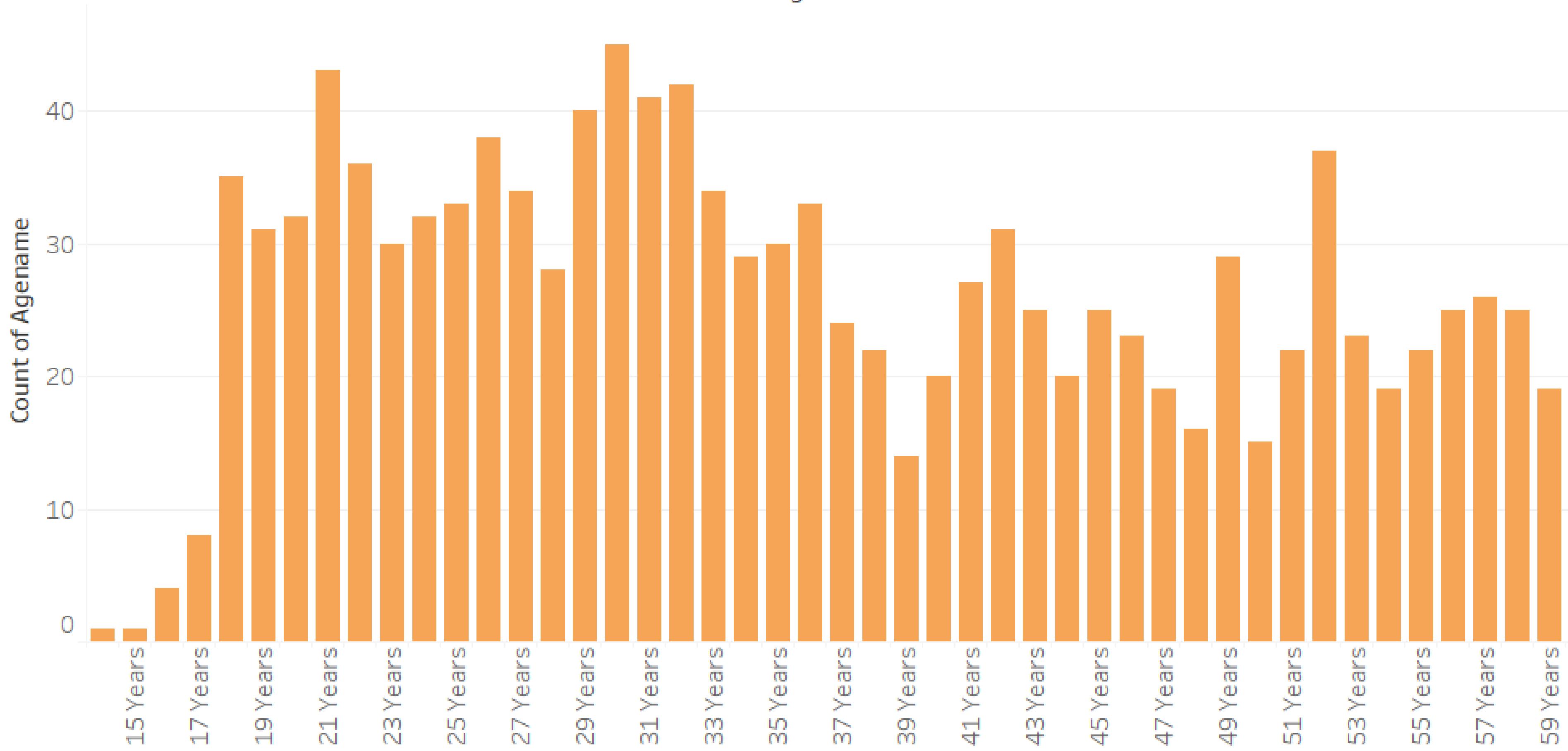
"Accidents hurt - safety first
always, prevention saves
lives and pain."

Project Presentation

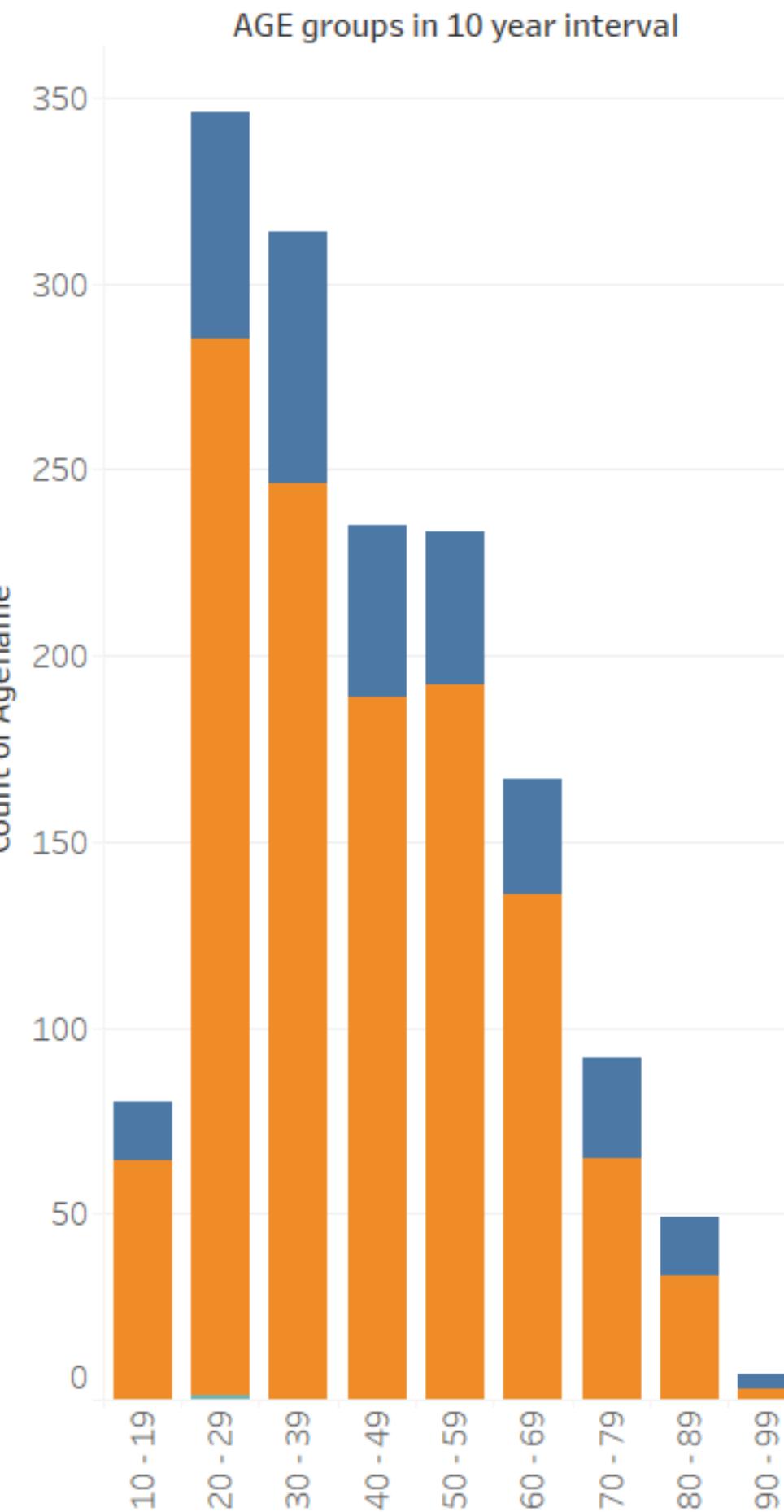


Count of people in each age

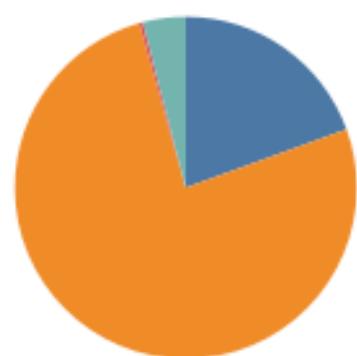
Agename



Number of people in age groups



Sex



Sexname

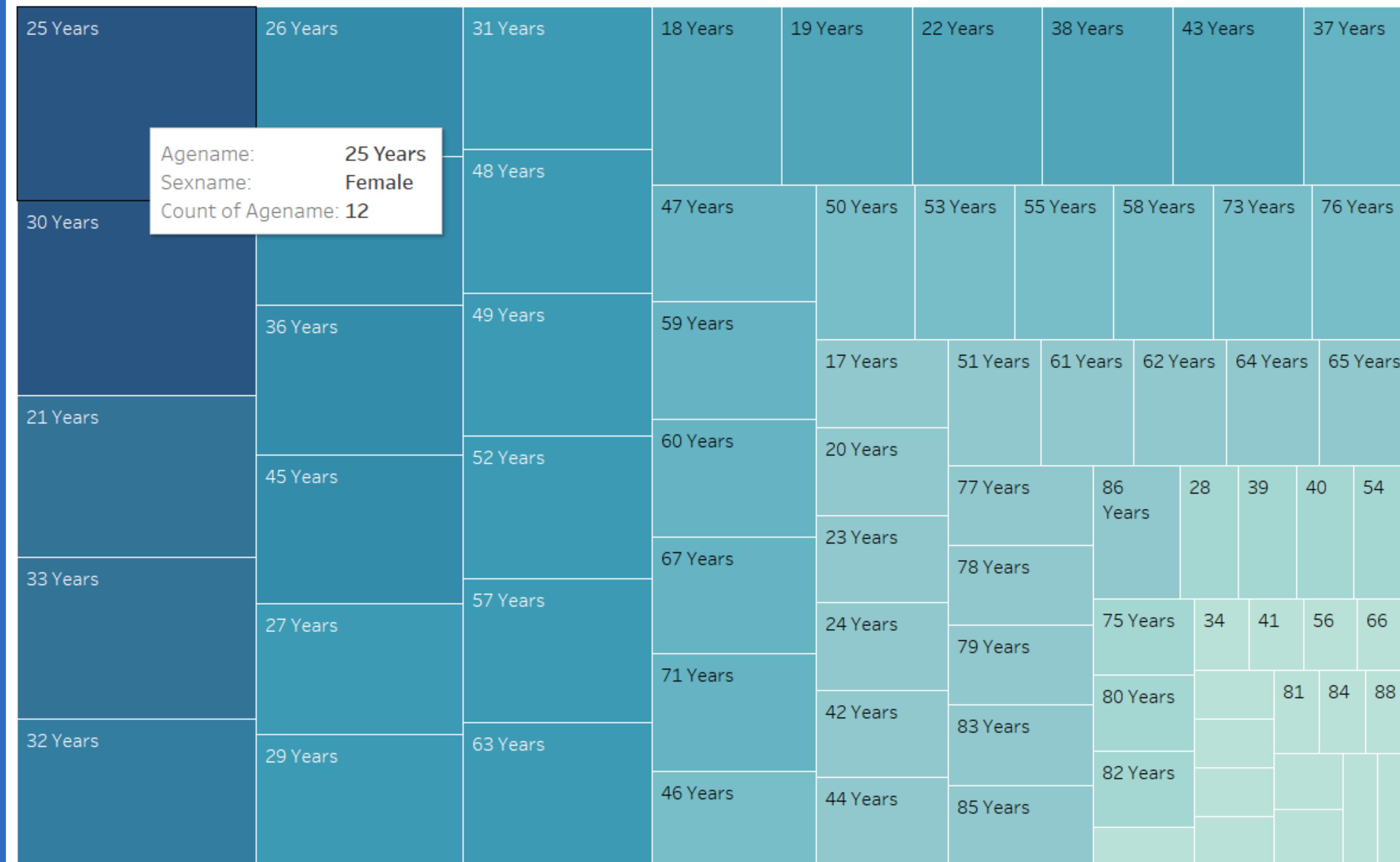
- Female
- Male
- Not Reported
- Reported as Unknown

Count of Sexname

1,594

age with high cases in Male sex

Age with high cases in female sex



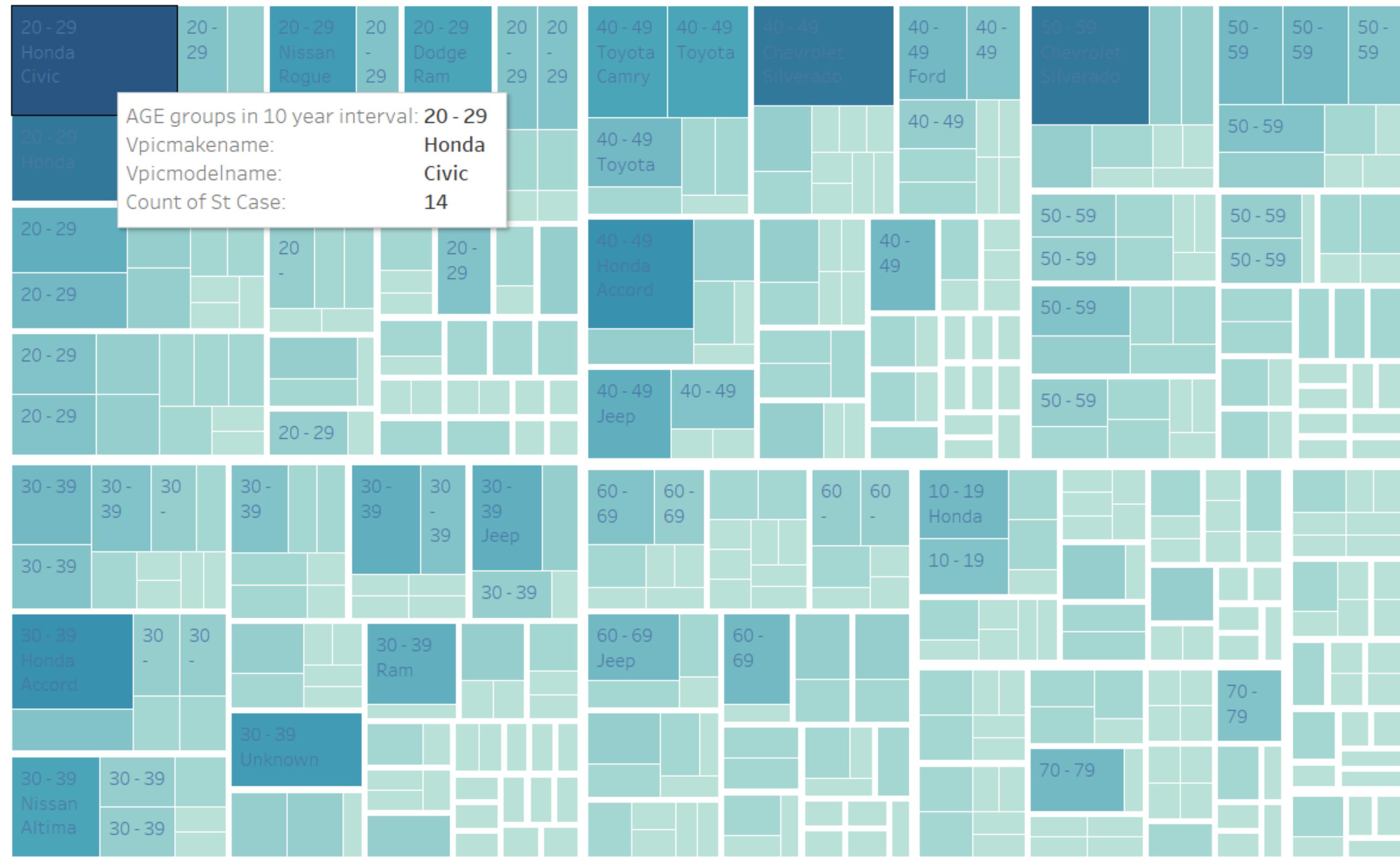
Distribution of car makes over age groups

Vpicmakename	AGE groups in 10 year interval						
	10 - 19	20 - 29	30 - 39	40 - 49	50 - 59	60 - 69	70 - 79
Ford		28	21	28	17	18	12
Chevrolet		30	26	25	33	14	10
Toyota		17	15	29	27	14	
Honda	14	45	30	21	19	11	
Dodge		16	14				
Harley Davidson		13		11	10	11	
Yamaha			14				
Subaru			12				
Kawasaki		10	11				

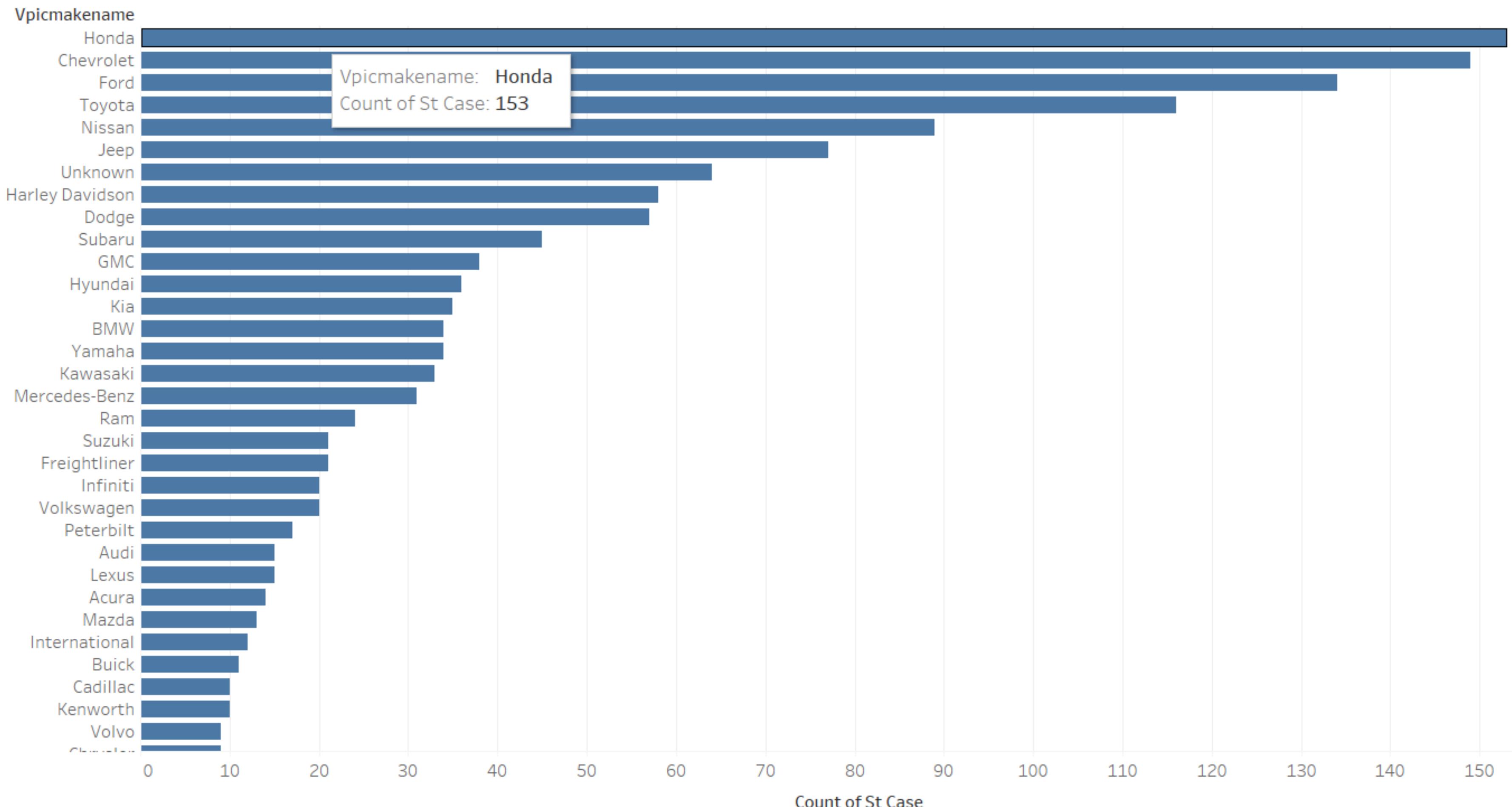
Distribution of car makes over sex

Vpicmakename	Sexname	
	Female	Male
Honda	40	109
Chevrolet	33	116
Ford	25	109
Toyota	32	86
Nissan	24	63
Jeep	27	51
Harley Davidson	2	56
Dodge	9	43
Subaru	17	27
GMC	9	29
Yamaha	1	33
Kia	11	23
Kawasaki		33
BMW	5	28
Mercedes-Benz	5	23
Ram	3	21
Suzuki	1	20
Freightliner	2	19
Volkswagen	7	13
Infiniti	1	17
Lexus	5	10
Audi	2	13
Mazda	3	10
Acura	4	9
Buick	5	6
Cadillac	3	7
Chrysler	2	7
Mitsubishi	2	5
Tesla		4

Car make and model distribution over age groups

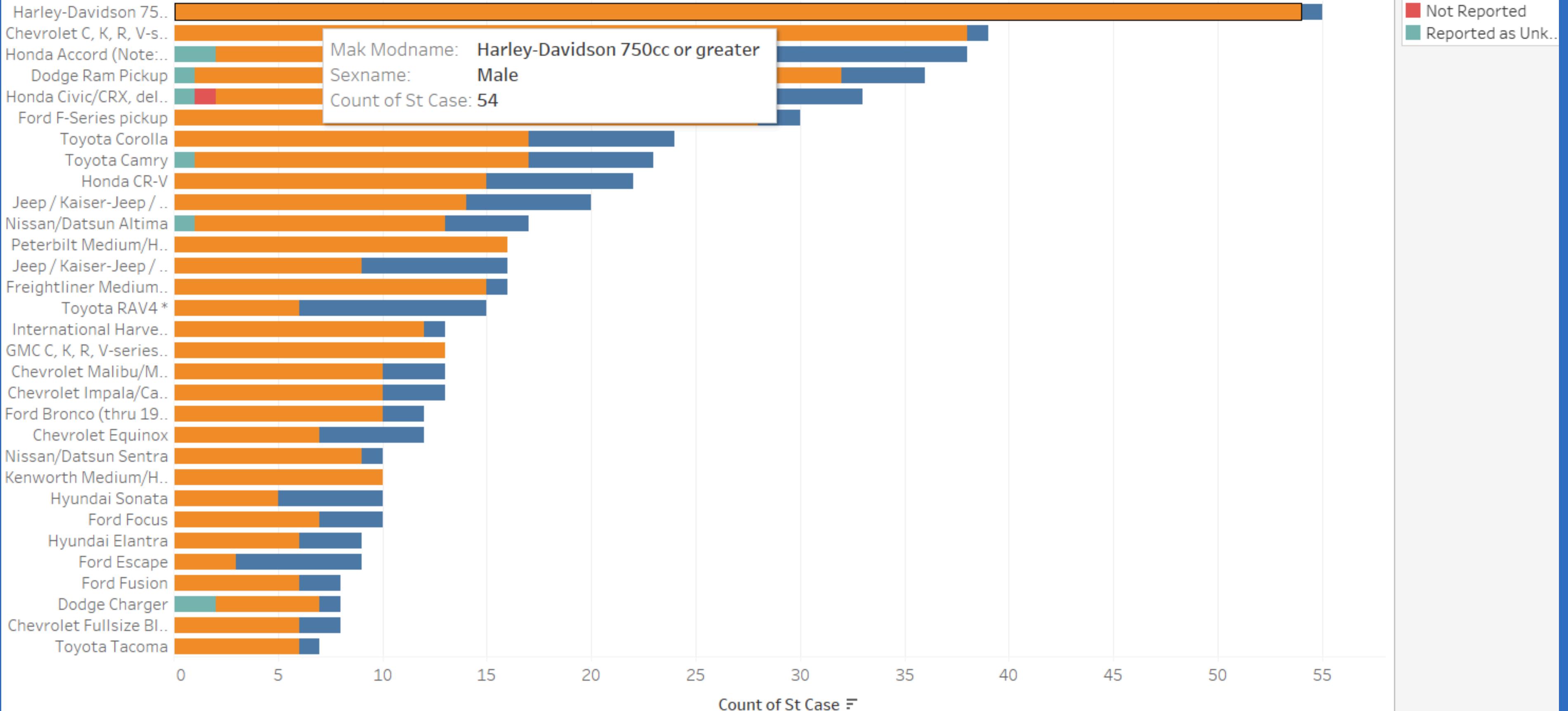


car makes with high cases



Car models with high cases

Mak Modname F:



Number of car makes in the state of New York

Make1
TOYOT 1,258,103
HONDA 1,203,991
FORD 1,099,871
CHEVR 1,084,657
NISSA 659,137
JEEP 591,521
SUBAR 499,515
HYUND 393,946
DODGE 293,352
BMW 274,895
GMC 261,751
ME/BE 256,200
KIA 235,987
VOLKS 207,512
RAM 201,948
LEXUS 196,961
MAZDA 185,266
ACURA 158,274
AUDI 134,262
HA/DA 122,362

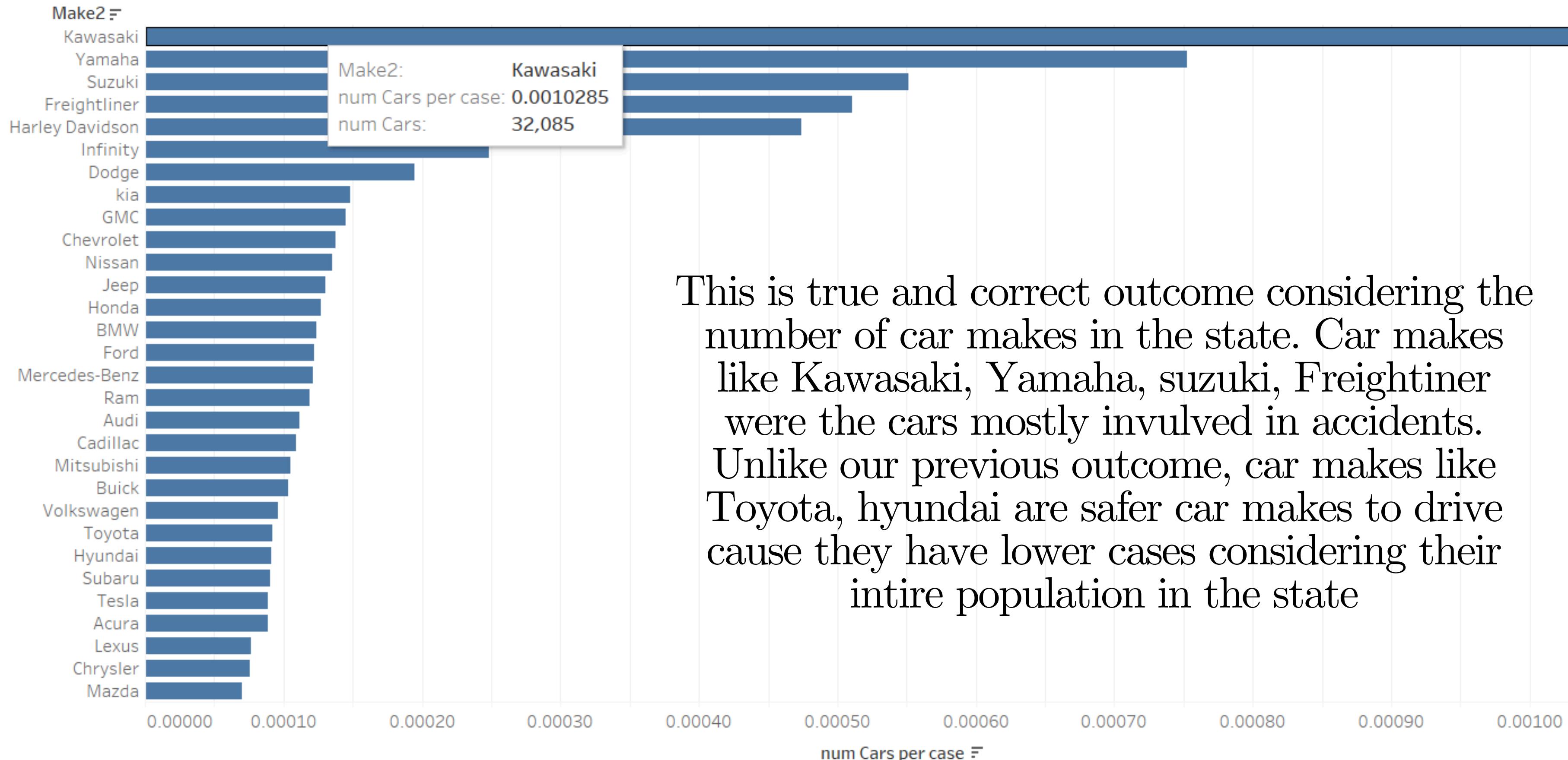
Used this data to standardize the outcome given before. Based on the previous outcome, Honda, chevrolet, ford and toyota were the car makes with high cases without considering the population of the car makes in the state of New York

State	Make	num Cars	St case count	num Cars per case
NY	Ford	1,099,871	134	0.01218%
NY	Chevrolet	1,084,657	149	0.01374%
NY	Toyota	1,258,103	116	0.00922%
NY	Nissan	659,137	89	0.01350%
NY	Dodge	293,352	57	0.01943%
NY	Honda	1,203,991	153	0.01271%
NY	Harley Davidson	122,362	58	0.04740%
NY	GMC	261,751	38	0.01452%

Used the percentage of how many cars were reported in accidents over the entire number of car make

$$\text{num Cars per case} = \frac{\text{St case count}}{\text{num Cars}} * 100$$

Standardized (which cars are most likely to be involved in accidents)



Conclusion

The age group 20 - 29 have highest rate of being involved in an accident and the car make involved was Honda

31 years old male has the highest cases compared to other ages

25 years old female has the highest cases compared to other ages

Most car makes have even distribution among the age groups. however, motercycles like Kawasaki and yamaha are driven by young people with the exception of harley davidson which has mostly older drivers

Toyota, Honda and ford have high population the the state of New york

Based on this findings, people driving car makes like Kawasaki, yamaha, suzuki, Freightliner and harley davidson should be very careful in the sate of New York