Impact of Operational Intensity on Transportation Safety

ECON 370 Final Project

Vikram Iyengar
University of North Carolina at Chapel Hill
2024-12-10

What Are We Studying

Background:

In this study, I define *operational intensity* as factors that indicate the performance and overall attributes of any type of transportation. This can include load factors, vehicular performance in miles traveled, and more.

We want to be able to identify if there is a relationship among improvements or changes in operational intensity, and how that can affect the safety of those who use transportation services.

Research Question:

How do different factors of operational intensity correlate with different metrics of transportation safety?

Motivation

Why do we want to study this?

- Preventing any sort transportation related injuries, incidents and fatalities are important by nature
- Being able to see trends across years of how vehicular performance coincides with these factors can lend insight into potential developments that impact safety
- Moreover, seeing how resource consumption and usage, which often plays a part in vehicular performance, can also lend further insight into safety statistics

Data

Transportation Ridership Data:

Transportation Services Index and Seasonally-Adjusted Transportation Data

- Created by the U.S. Department of Transportation (DOT), Bureau of Transportation Statistics (BTS).
- Measures the movement of freight and passengers.
- Note: Data is seasonally adjusted
  - Allows for measurement of real monthly changes; short and long term patterns of growth or decline; and turning points.

Transportation Safety and Event Data:

Modal Service data and Safety & Security (S&S) public transit time series data

- Created by the U.S. Department of Transportation (DOT), (Federal Transit Administration).
- Measures statistics about transportation in regards to major and non-major safety events

Key Variables

VMT:

Vehicle Miles Traveled (Used as our main measure of operational intensity and our main predictor variable)

Injuries:

Transportation related injuries that occurred. Include:

- Passenger, Operator, People Waiting or Leaving Injuries

Fatalities:

Transportation related fatalities that occurred Include:

- Pedestrian in/not in crosswalk, Employee or Passenger Fatalities

Data Cleaning Process

Loading Data:

```
transport_data = read.csv(
  "/Users/vikram/Documents/ECON 370/Final Project/Data/Transportation_Data.csv")
event_data = read.csv(
  "/Users/vikram/Documents/ECON 370/Final Project/Data/Time_Series_Event_Data.csv")
```

Cleaning Transport_Data:

```
transport_data$OBS_DATE ← as.Date(transport_data$OBS_DATE, format = "%m/%d/%Y")
transport_data$Year ← format(transport_data$OBS_DATE, "%Y")
transport_data$Month = format(transport_data$OBS_DATE, "%m")
filtered_dat = transport_data ▷
  select(OBS_DATE, VMT, Year, Month, PETROLEUM_D11, NATURAL_GAS_D11)

transit_data_monthly = filtered_dat ▷
  filter(Year > 2013 & Year < 2024) ▷
  mutate(Year = as.integer(Year), Month = as.integer(Month)) ▷
  mutate(Month_Year = paste0(Month,"/", Year))
```

Data Cleaning Process Continued...

Cleaning Event_Data:

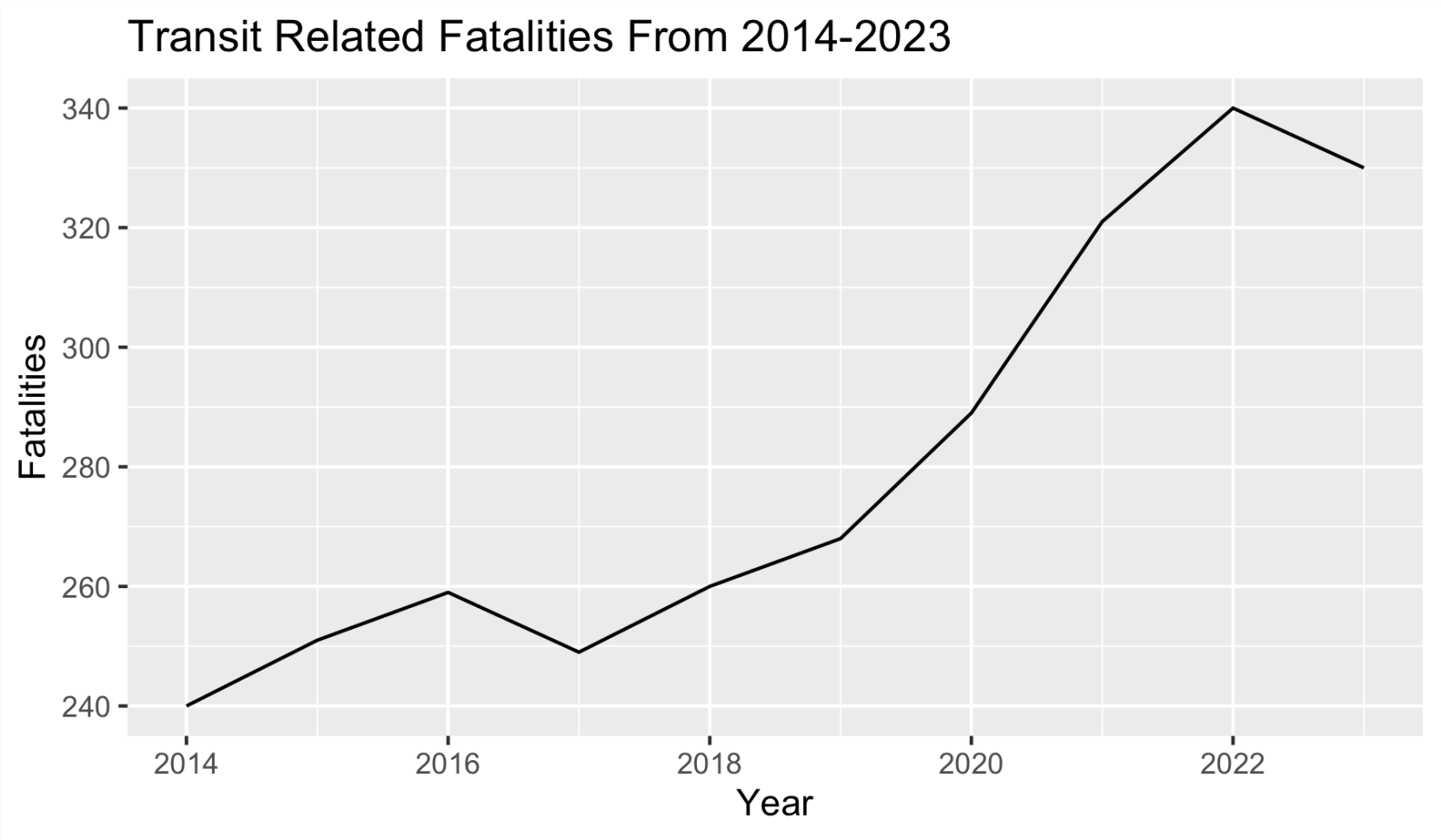
```
event_data_monthly = event_data ▷
  mutate(Month = as.integer(factor(Month, levels = month.name))) ▷
  arrange(Year, Month) ▷
  select(Year, Month, Total.Collisions, Total.Events, Total.Fatalities, Total.Injuries
  filter(Year < 2024) ▷
  group_by(Year, Month) ▷
  summarise(Collisions = sum(Total.Collisions), Events = sum(Total.Events), Fatalities
  mutate(Month_Year = paste0(Month,"/", Year))
```

New Cleaned Event Data

Table: Event Data

Year	Month	Collisions	Events	Fatalities	Injuries	Month_Year
2014	1	420	1670	21	1859	1/2014
2014	2	432	1659	18	1831	2/2014
2014	3	421	1700	23	1971	3/2014
2014	4	388	1588	16	1792	4/2014
2014	5	439	1729	18	2114	5/2014
2014	6	428	1641	20	1948	6/2014

Graphical Representation



New Cleaned Transportation Data

Table: Event Data

OBS_DATE	VMT	Year	Month	PETROLEUM_D11	NATURAL_GAS_D11	Month_Year
2014-01-01	226413	2014	1	207946	2340.5	1/2014
2014-02-01	213949	2014	2	226841	2314.8	2/2014
2014-03-01	253424	2014	3	217994	2283.6	3/2014
2014-04-01	256736	2014	4	218166	2191.2	4/2014
2014-05-01	266237	2014	5	215198	2187.9	5/2014
2014-06-01	263459	2014	6	213778	2112.9	6/2014

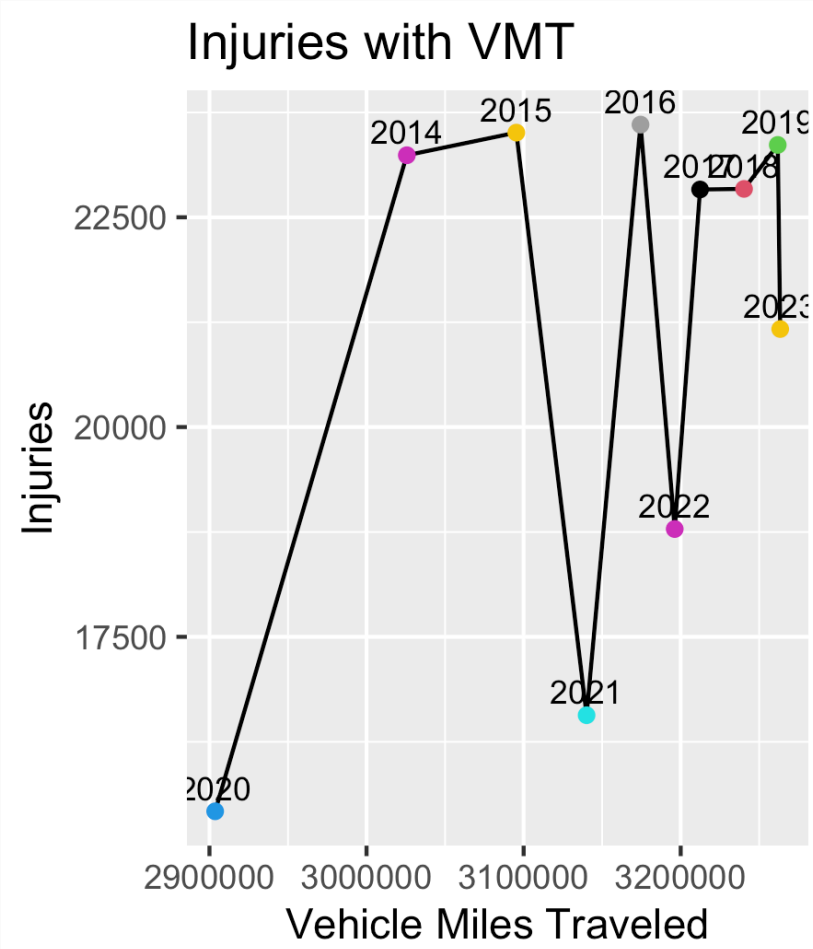
Merging Data

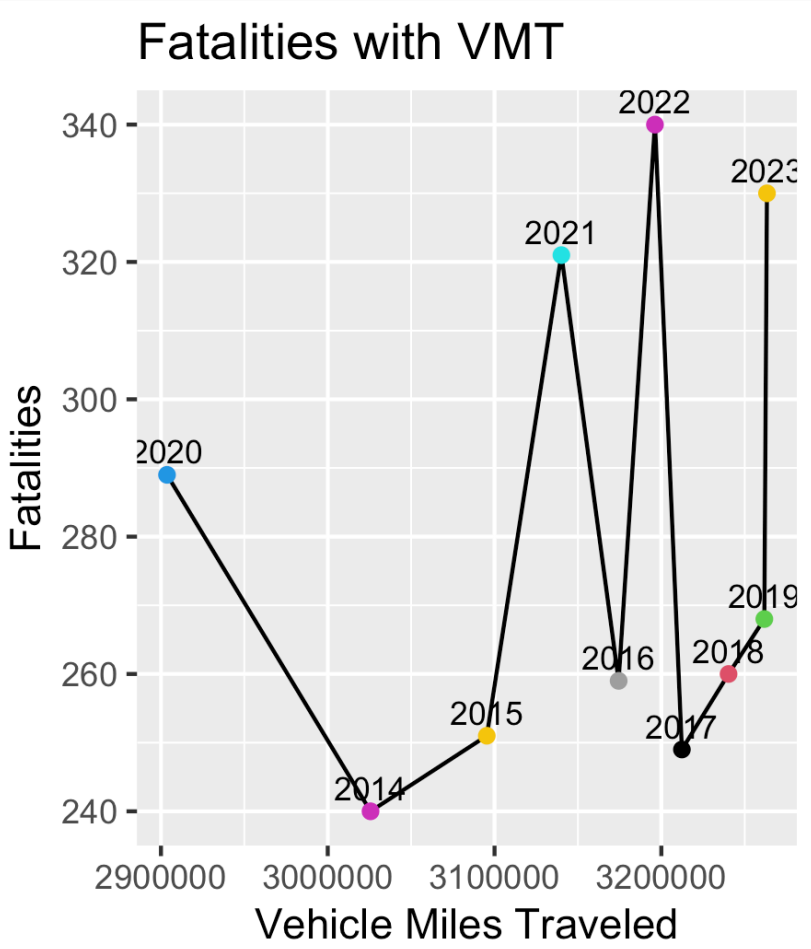
Below contains some of the columns from the fully merged dataset, comprising of elements from the event and transport datasets.

Table: Transit Event Data

Month_Year	Fatalities	Injuries	VMT	PETROLEUM_D11	NATURAL_GAS_D11
1/2014	21	1859	226413	207946	2340.5
2/2014	18	1831	213949	226841	2314.8
3/2014	23	1971	253424	217994	2283.6
4/2014	16	1792	256736	218166	2191.2
5/2014	18	2114	266237	215198	2187.9
6/2014	20	1948	263459	213778	2112.9

How Does VMT Compare to Safety Events





Observe multiple spikes within the graphs. This motivates the regression that we want to run, which will consider other factors like Natural Gas and Petroleum Consumption in vehicles.

Initial Regression

We want to observe if the outcome of both transportation fatalities and injuries can be explained by changes in three predictors:

- Vehicle Miles Traveled (VMT)
- Consumption of Natural Gas (NatGas)
- Consumption of Petroleum (Pet)

Note: These variables are taken across multiple different vehicles in different locations across the United States, from 2014-2023.

Fitting the Model

Methodology: We will create two multiple linear regression models. They will take on the following format:

- $$\hat{Fatalities} = \beta_0 + \beta_1 VMT + \beta_2 NatGas + \beta_3 Pet$$
- $$\hat{Injuries} = \beta_0 + \beta_1 VMT + \beta_2 NatGas + \beta_3 Pet$$

Regression Statistics

Fatality Statistics

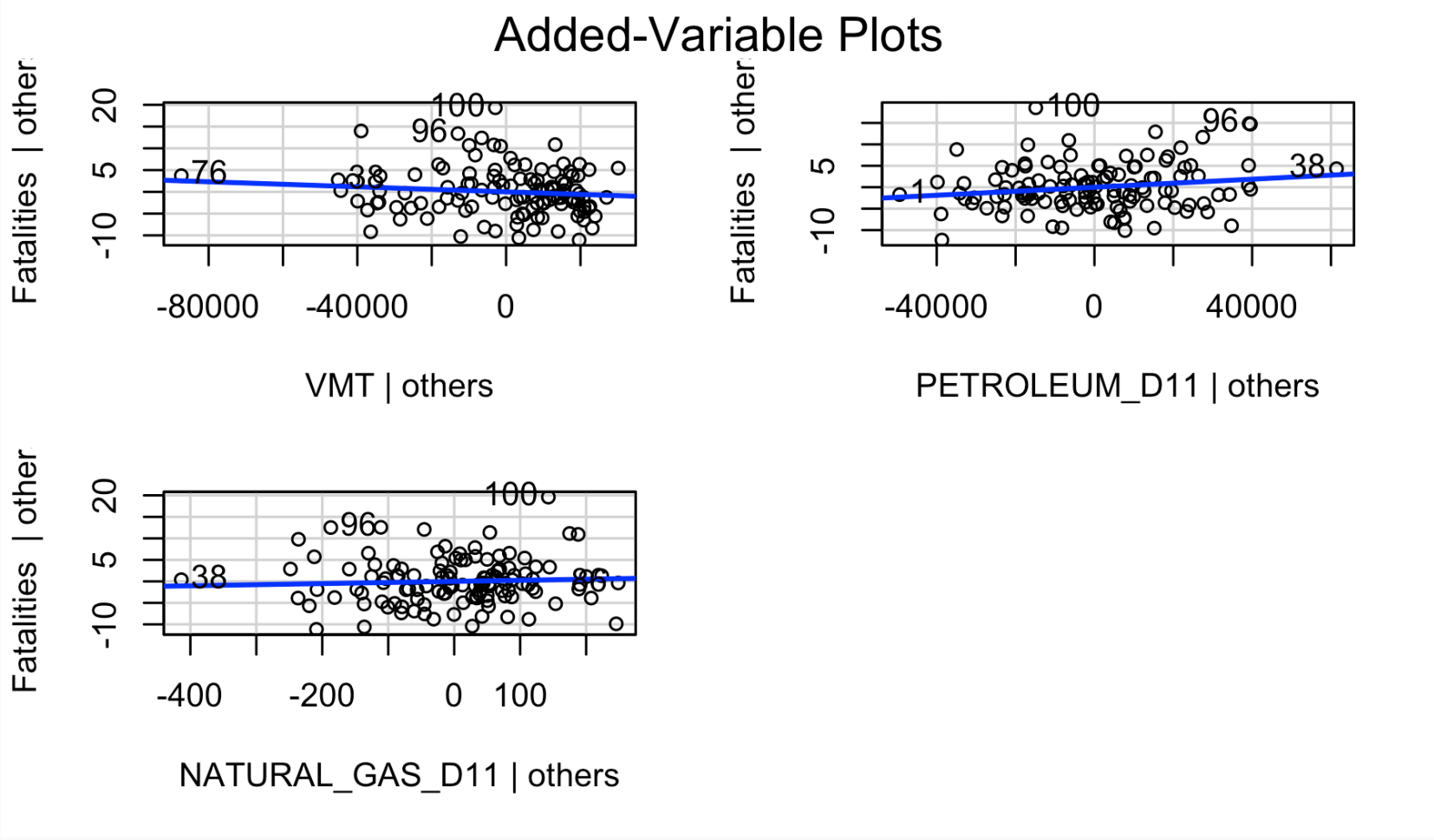
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	11.2308524	9.0518620	1.2407229	0.2172116
VMT	-0.0000291	0.0000253	-1.1502417	0.2524101
PETROLEUM_D11	0.0000470	0.0000246	1.9111200	0.0584569
NATURAL_GAS_D11	0.0025633	0.0043922	0.5836007	0.5606221

Injury Statistics

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1679.3423788	378.1851639	4.440530	0.0000206
VMT	0.0063924	0.0010575	6.044559	0.0000000
PETROLEUM_D11	-0.0001632	0.0010266	-0.159013	0.8739351
NATURAL_GAS_D11	-0.6287087	0.1835039	-3.426132	0.0008476

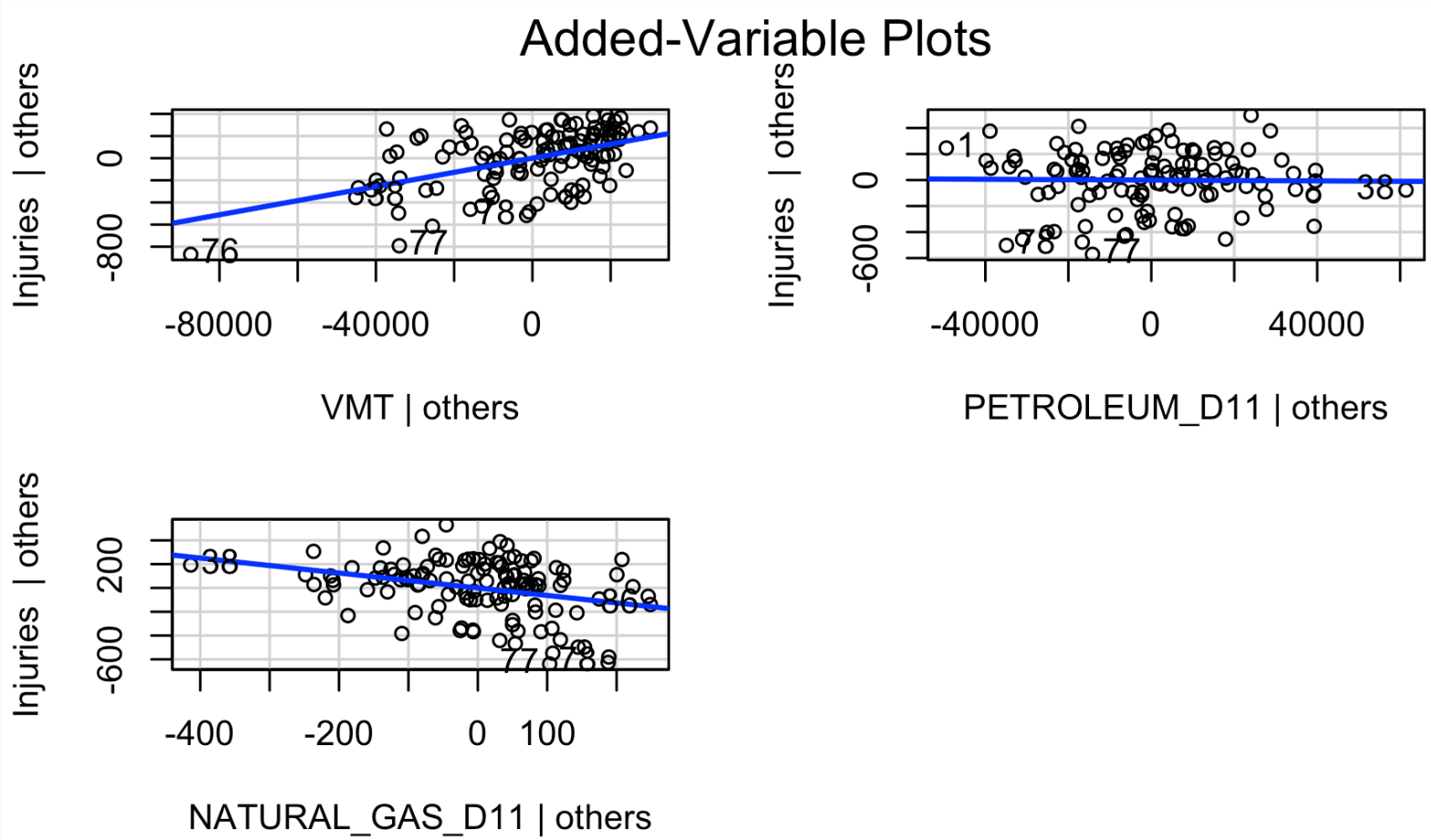
Graphical Representation

Below are the added variable plots for each predictor variable in my fatality model



Graphical Representation (Injuries)

Below are the added variable plots for each predictor variable in my injury model



Analysis of Fatality Regression

Coefficients vs. P-values

Even though our fatality regression has coefficients on each predictor variable, the p-values explain the actual significance of them.

Moreover, it is hard to reject the hypothesis that all the predictor variables have no effect on the response variable (Fatalities) as each p-value is not statistically significant at the 95% confidence level.

Analysis of Injury Regression

Coefficients vs. P-values

Our injury regression tells a different story though in comparison to our fatality regression. The coefficients on the predictor variables **VMT** and **NatGas** both have p-values much less than 0.05.

This means that for those two predictor variables, there is a chance they could have some indication/correlation with how injury statistics are predicted.

Overall Validity From AvPlots

The AvPlots do a good job of showing how each predictor variable individually contributes to the outcome of our response variable.

Fatalities

- Each predictor variable individually shows a weak to almost 0 relationship with the Fatality variable.
- This is further explained through our of our analysis of each p-value of our predictor variables not indicating statistical significance.

Injuries

- The predictors *VMT* and *NatGas* both indicated a positive relationship/correlation, especially VMT.
- *VMT* is shown to have the most positive slope with transportation related injuries, indicating that increases in VMT are correlated with increased in transportation related injuries.

Conclusion

What Did We Find?

- Most worthwhile to continue exploring how increases in Vehicle Miles Traveled contributes to overall transportation related injuries
- A weak negative correlation could exist among Natural Gas Consumption for Vehicles and transportation related injuries

What Needs to Be Done?

- Further analysis of transportation related injuries across different regions
- A further breakdown into the types of injuries could be worthwhile to explore
- Possible transformation of the data in order to better normalize the predictor variables to align better with general linear model assumptions

Limitations of My Study

Generalizing the Data

I cleaned the data and made it very generalized for the purpose of applying the analysis more broadly. However, this fails to take into account a lot of the specifics of our different predictors.

Omitted Variable Bias

Similar to the generalization of my study, I did not account for a lot of other variables and factors which could explain the variance and results of my response variables. The decrease in transportation during the Pandemic is a big one, as transportation was very limited around that time.

Relevance of Predictor Variables

The predictor variables I chose, while worthwhile to study, may not have been the best when it came down to predicting different outcomes of safety incidents. Predictors such as Transportation Regulations or state-to-state policies may have been more effective.

Thank You

Note about AI Usage: AI was used to help with understanding AVPlots and helping organize date formats when cleaning the data. StackOverflow was used for general knowledge questions regarding specific libraries such as knitr and the kable command, as well as the cowplot library.