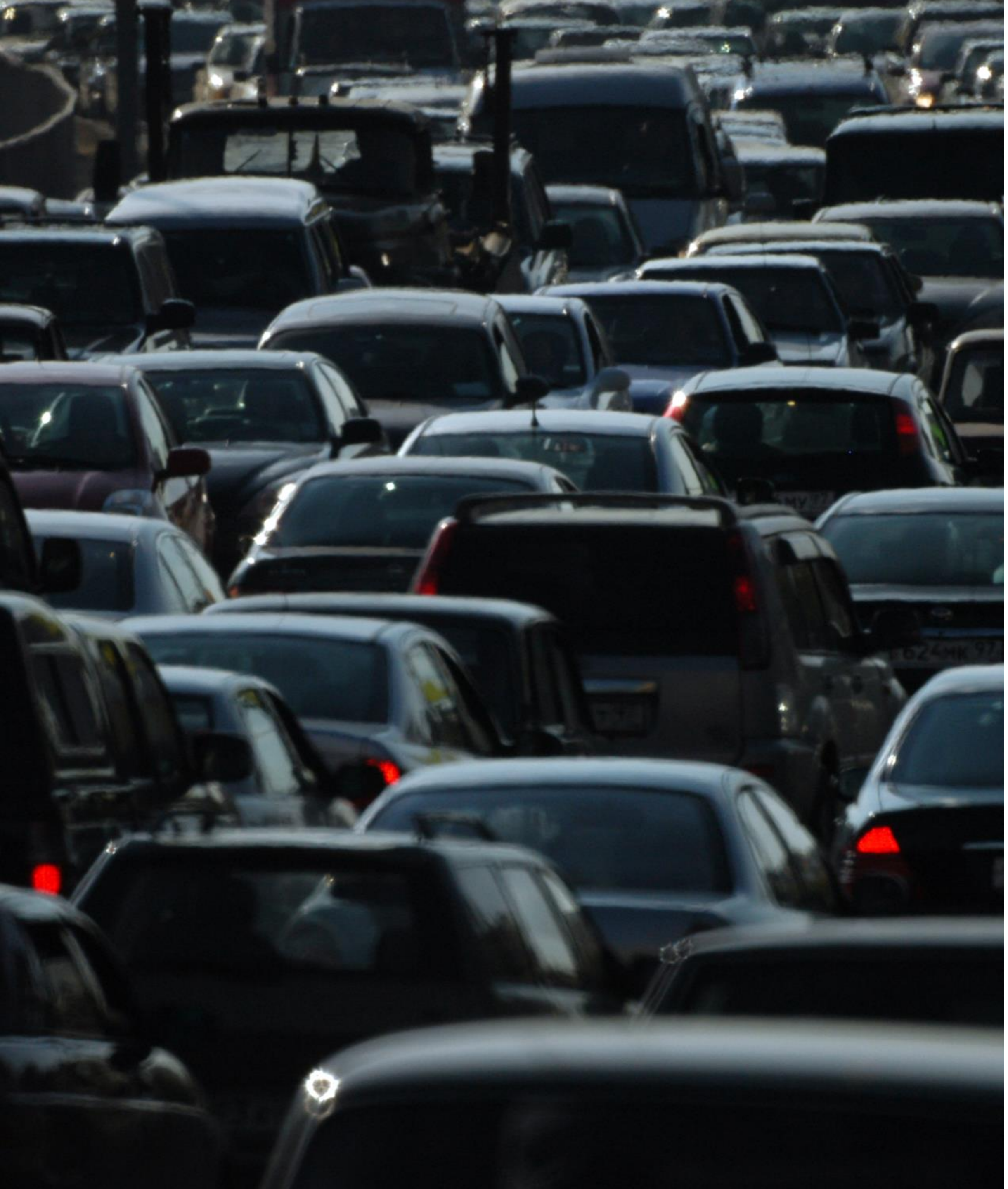


AUTO MPG REGRESSION ANALYSIS

Name: Osama Mufti & Gustavo Orozco

STAT 311- Regression Analysis

Fall 2025

Research Question

How do vehicle characteristics influence fuel efficiency (MPG)?

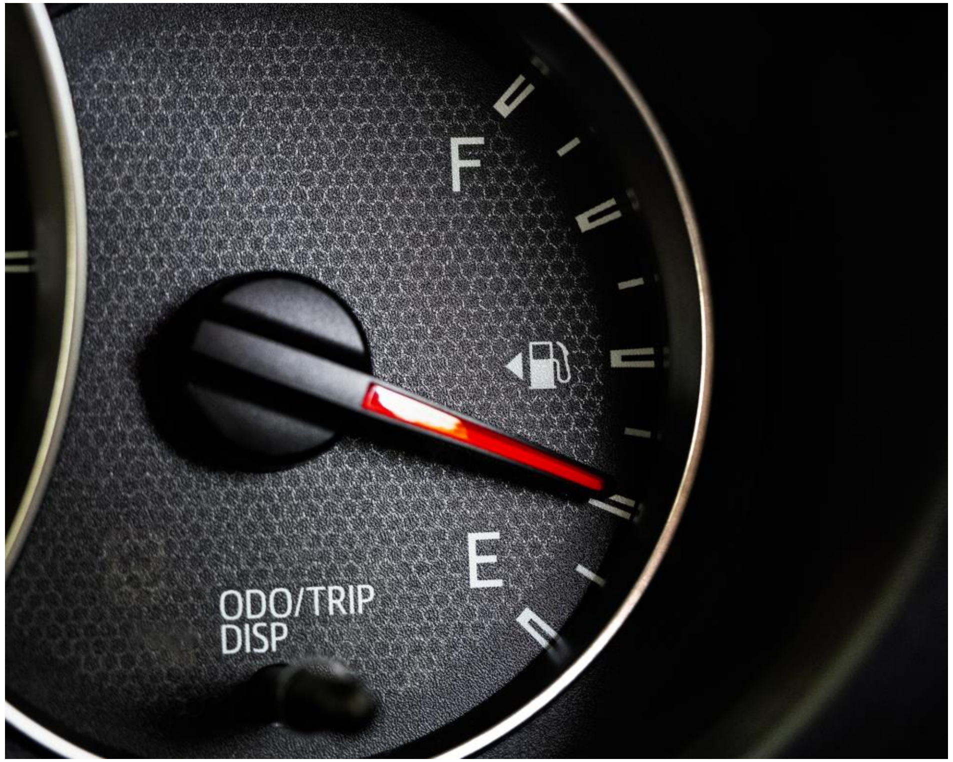
Key Questions:

Does weight affect fuel efficiency?

Did technology improve over time?

Do European and Japanese cars differ from American cars?

Has the relationship between weight and MPG changed over time?



Dataset Overview

Dataset: Auto MPG

Characteristic	Details
Source	UCI Machine Learning Repository/Kaggle.com
Time Period	1970-1982
Original Size	392 observations
After Cleaning	392 observation (6 missing values removed)
Variables	9 total

Key Variables:

- Dependent: MPG (9.0-46.6 mpg)
- Predictors: Weight, Year, Origin, Cylinders, Displacement, Horsepower, Acceleration

Exploratory Data Analysis

EDA: Correlation Matrix

Strong Correlations with MPG:

Multicollinearity Problem:

Cylinders→ Displacement r = 0.95

Displacement ←→ Weight r = 0,93

All engine variables are highly correlated (r>0.84)

Variable	Correlation	Strength
Weight	-0.83	Very Strong
Displacement	-0.81	Very Strong
Cylinder	-0.78	Strong
Horsepower	-0.78	Strong
Year	+0.58	Moderate

MPG Varies by Manufacturing Origin

Key Finding: Japanese cars achieve 50% better fuel efficiency than American Cars!

Origin	Mean MPG	Sample Size
Japan	30.45	79 Cars
Europe	27.60	68 Cars
USA	20.03	245 Cars

EDA-MPG by Origin

Model Building Strategy

Four Models Tested

Model	Predictors	Purpose
Model 1	Weight only	Baseline
Model 2	Weight, Acceleration, Year, Origin	Initial Full model
Model 3	All 8 predictors	Test Multicollinearity
Model 4	Weight, Year, Origin, Weight*Year	Test Interaction

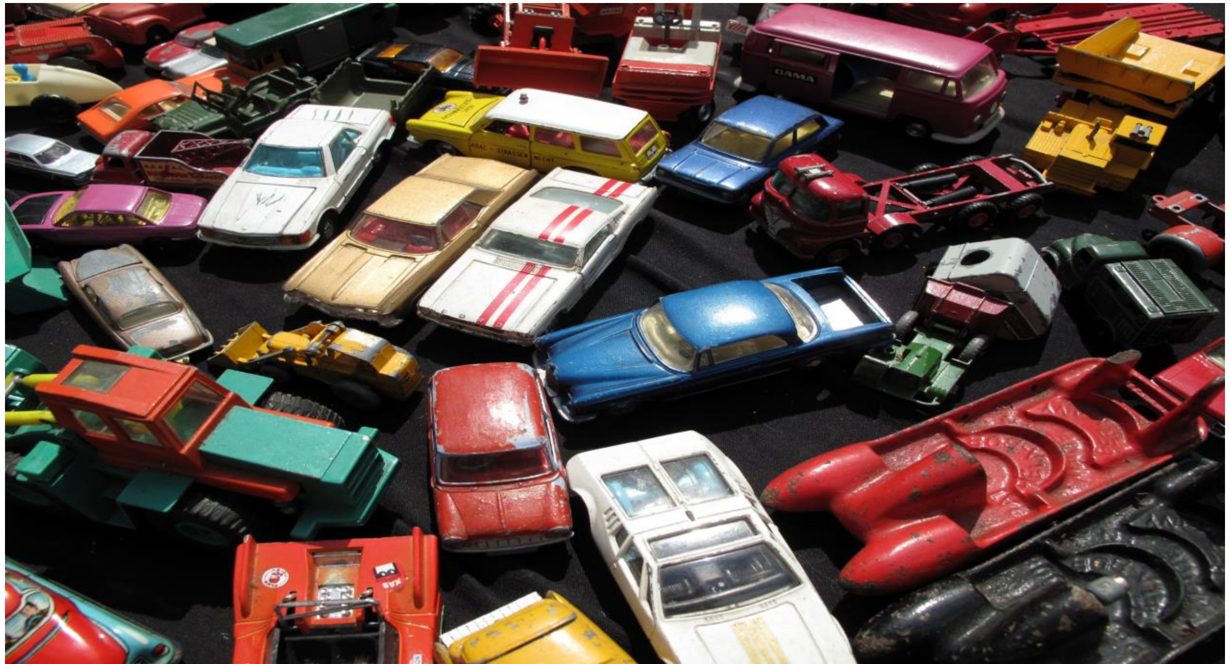
Why These Models?

Model 1: Establish the baseline with the strongest predictor

Model 2 Test Main effects

Model 3: Demonstrate the multicollinearity problem

Model 4: Test if weight effect changes over time?



Model Comparison Model Performance Comparison

Model	Rsquare	Adj Rsquare	RMSE	All sig	Issue
Model 1	69.2%	69.1%	4.33	Sig	Limited
Model 2	81.9%	81.7%	3.34	No Sig	Acceleration NS
Model 3	82.4%	82.1%	3.31	No Sig	Multicollinearity
Model 4	84.3%	84..1%	3.11	Sig	None

Winner: Model 4

Highest R^2 (84.3%)

Lowest RMSE (3.11 MPG)

All predictor Significant

No Multicollinearity

Key Hypothesis Test: Interaction

Testing the Interaction Effect

H_0: Weight effect does NOT depend on year (β_5 =0)

H_1:Weight effect DEPENDS on year ($\beta_5 \neq 0$)

Result:

Coefficient: β_5 = -0.0000451

t-ratio: 7.72

P-value: <.0001

Decision: Reject $\boldsymbol{H_0}$

Conclusion: The interaction is highly significant!
The weight penalty decreased over time

Model 4- Hypothesis Tests

All Predictors Significant:

Predictor	P-value	Significant?
Weight	<0.0001	Significant
Year	<.0001	Significant
Origin_Europe	<.0001	Significant
Origin_Japan	0.0006	Significant
Weight * Year	<0.0001	Significant

Speaking engagement metrics

Impact factor	Measurement	Target	Achieved
Audience interaction	Percentage (%)	85	88
Knowledge retention	Percentage (%)	75	80
Post-presentation surveys	Average rating	4.2	4.5
Referral rate	Percentage (%)	10	12
Collaboration opportunities	# of opportunities	8	10

Model 4 – The Interaction Effect

Year	Weight Effect (per 1,000 lbs)	Change
1970	-7.36 mpg	Baseline (Heavy penalty
1976	-7.09 mpg	Improved by 0.27 mpg
1982	-6.81 mpg	Improved by 0.55 mpg

What Does the Interaction Mean?

Weight Effect Changes Over Time:

Total Improvement: 7.5% reduction in weight penalty

Why Did This Happen?

1973 Oil Crisis: Urgent need for efficiency

1975 CAFÉ Standard Regulatory Pressure

Technology: Electronic fuel Injection, better aerodynamics, Lighter material terial
Heavy cars benefited MORE from these improved aerodynamics

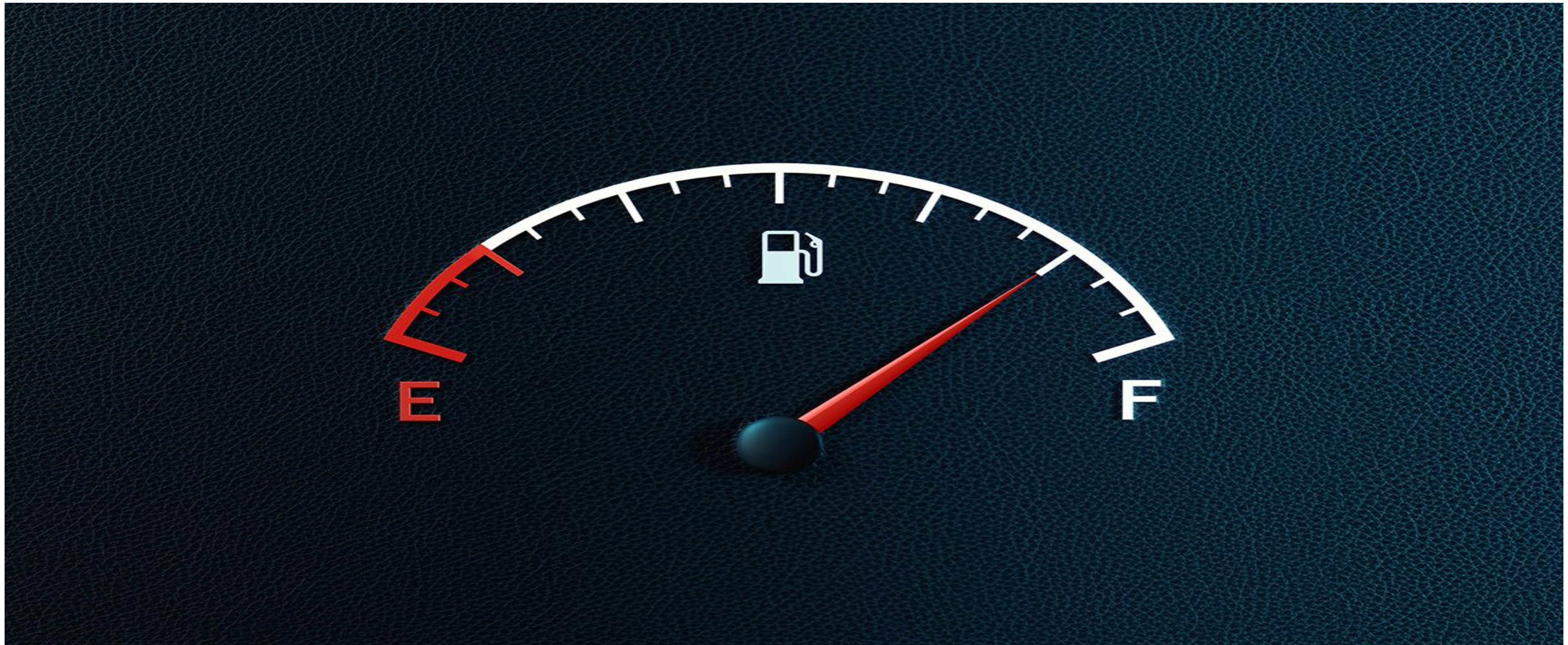
Model 4 Results-Coefficients

Model Performance:

Rsquare = 84.3%

RMSE = 3.11 mpg

Predictor	Coefficient	P-value	Interpretation
Weight	+0.885	<.0001	Modified by interaction
Year	+2.041	<.0001	-0.76 mpg improvement per year
Origin_Europe	+2.145	<.0001	+2.15 mpg vs USA
Origin_Japan	+1.682	0.0006	+1.68 mpg vs USA
Weight * Year	-0.000451	<.0001	Reduces weight penalty over time



Key Findings Summary

Weight is the Dominant Factor
Strongest predictor (r =-0.83)
Effect: -6.8 to –7.4 mpg per 1,000 Lbs

2. Significant Interaction Discovered
   Weight penalty decreased 7.5% from 1970 to 1982
   Technology reduced the cost of weight

3. Technology Improved Efficiency
   Average improvement: 0.76 mpg per year
   Total 1970-1982: 9-mpg improvement

4. Origin Matters
   Japan: +1.68 mpg vs USA
   Europe:+2.15 mpg vs USA

5. Model 4 is Best
   Rsquare = 84.3% all predictors significant

Conclusions

Model 4 successfully answers the research question

Main Conclusions:

1.Weight dominates fuel efficiency (Strongest effect)

2. Interaction is real-technology reduced weight penalty over time

3.Origin matters-Japanese/European cars more efficient

4.Model 4 is best-84.3% Rsquare, all significant, no violation