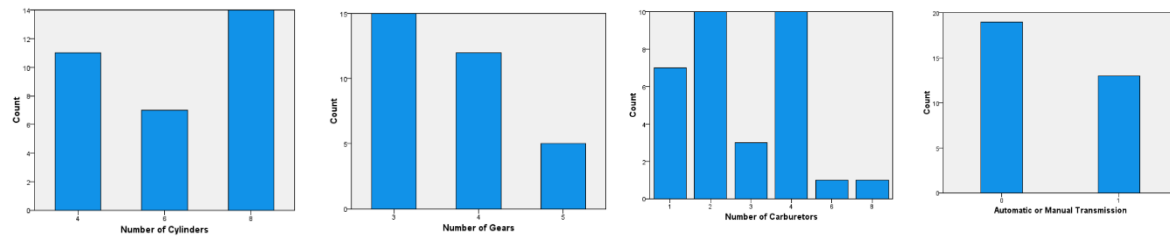


Cars Analysis

Dataset: The data for 32 cars is collected where each car's specifications are shown in the dataset. The parameters mileage per gallon, number of cylinders, number of carburetors, number of gears, weight, automatic/manual, displacement, horse power and quarter mile.

1. Exploratory Data Analysis:

a. Univariate Analysis: All variables in cars data are analyzed:

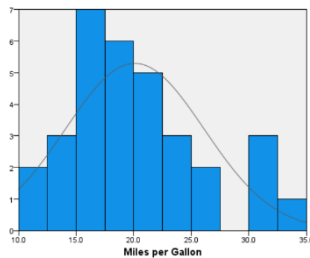


Interpretation: Analyzing the number of cylinders, maximum count of cars i.e. 14 have 8 cylinders, 11 cars have 4 cylinders and Only 7 cars have 6 cylinders. This indicates that large number of cylinders are favored by consumers. More the number of cylinders, higher the power of engine. Therefore, more engine power is preferred by customers.

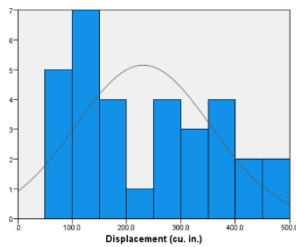
Similarly, among all 32 cars, 15 cars have 3 gears, 12 cars have 4 cylinders and least (5) cars have 5 cylinders. It indicates that less number of gears are preferred which makes the driving comfortable.

When analyzing the number of carburetors, most people (10) prefer either 2 or 4 number of carburetors whereas higher number of carburetors i.e. 6 and 8 are least preferred.

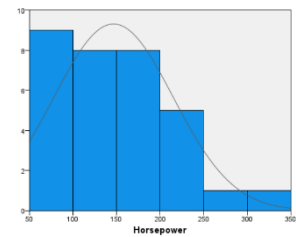
While considering the automatic or manual type of cars, 18 cars have automatic transmission and 13 cars have manual transmission indicating the higher preference for automatic cars.



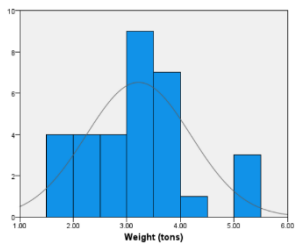
Miles per Gallon	Mean	20.091	1.0654
95% Confidence Interval for Mean	Lower Bound	17.918	
	Upper Bound	22.264	
5% Trimmed Mean		19.893	
Median		19.200	
Variance		36.324	
Std. Deviation		6.0269	
Minimum		10.4	
Maximum		33.9	
Range		23.5	
Interquartile Range		7.5	
Skewness		.672	.414
Kurtosis		-.022	.809



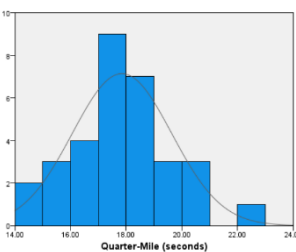
Displacement (cu. in.)	Mean	173.155	54.8082
95% Confidence Interval for Mean	Lower Bound	148.031	
	Upper Bound	200.279	
5% Trimmed Mean		166.300	
Median		163.000	
Variance		12360.800	
Std. Deviation		111.183	
Minimum		71.1	
Maximum		455.0	
Range		383.9	
Interquartile Range		53.2	
Skewness		.450	.614
Kurtosis		-.1088	.808



Horsepower	Mean	146.69	12.120
95% Confidence Interval for Mean	Lower Bound	121.97	
	Upper Bound	171.41	
5% Trimmed Mean		142.76	
Median		123.00	
Variance		4700.967	
Std. Deviation		68.563	
Minimum		52	
Maximum		335	
Range		283	
Interquartile Range		85	
Skewness		.799	.414
Kurtosis		.275	.809



Weight (tons)	Mean	3.3173	17297
95% Confidence Interval for Mean	Lower Bound	2.8445	
	Upper Bound	3.5700	
5% Trimmed Mean		3.1889	
Median		3.3250	
Variance		.657	
Std. Deviation		.8146	
Minimum		1.51	
Maximum		5.42	
Range		3.91	
Interquartile Range		1.19	
Skewness		.466	.414
Kurtosis		.417	.809

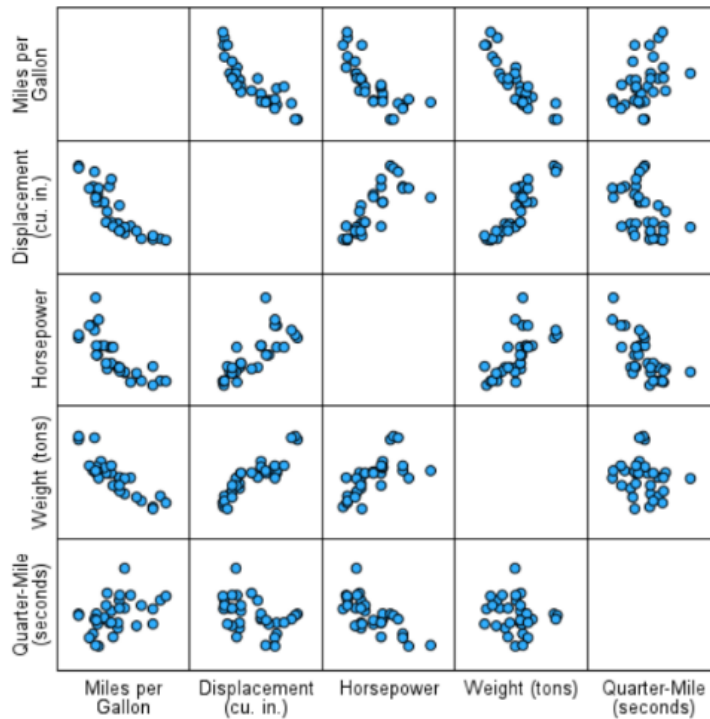


Quarter-Mile (seconds)	Mean	17.8488	.31589
95% Confidence Interval for Mean	Lower Bound	17.2045	
	Upper Bound	18.4930	
5% Trimmed Mean		17.8079	
Median		17.7100	
Variance		3.193	
Std. Deviation		1.78694	
Minimum		14.50	
Maximum		22.90	
Range		8.40	
Interquartile Range		2.02	
Skewness		.406	.414
Kurtosis		.865	.809

Interpretation: Mileage per gallon follows normal distribution ranging from 10.4 to 33.9. The mean value is 20.1, most cars are clustered around mean having 20.1 mpg. Horse power is a right skewed distribution as mean > median and is varying between 52 and 335. It indicates that most values are clustered towards the mean but due to presence of extreme values the distribution is right skewed i.e. most cars have horse power on

146.7. The displacement variable has less difference in mean and median indicating the normal distribution which means that most cars have 230.7 cu in of displacement. The weight and quarter mile graph follows a normal distribution which shows that most cars are clustered around mean i.e. have weight of 3.22 tons and quarter mile of 17.8 seconds.

b. Bivariate Analysis:



Interpretation: The scatterplot of all variables shows that there exists a strong negative relationship between miles per gallon vs displacement, horsepower and weight. It indicates that with the increase in value of one variable (displacement, horse-power and weight), the value of other variable decreases. The relationship between variables displacement-horse power; weight-horse power and displacement-horse power is positive indicating that with the increase in one variable, the value of other variable also increases.

Number of Cylinders

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	4	11	34.4	34.4	34.4
	6	7	21.9	21.9	56.3
	8	14	43.8	43.8	100.0
	Total	32	100.0	100.0	

Interpretation: 43% of cars in data has 8 cylinders, 34.4% of cars have 4 cylinders and only 22% of cars has 6 cylinders indicating that mostly cars have either highest (8) no. of cylinders or least (4) no. of cylinders.

Number of Gears

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	3	15	46.9	46.9	46.9
	4	12	37.5	37.5	84.4
	5	5	15.6	15.6	100.0
	Total	32	100.0	100.0	

Interpretation: 47% of cars in data has only 3 gears, 37.5% of cars have 4 gears and only 15.6% of cars has 5 gears indicating that lesser number of gears are preferred.

Automatic or Manual Transmission

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	19	59.4	59.4	59.4
	1	13	40.6	40.6	100.0
	Total	32	100.0	100.0	

Interpretation: 60% of cars in data has Automatic transmission and 40.6% of cars have Manual transmission indicating that automatic cars are preferred over manual cars.

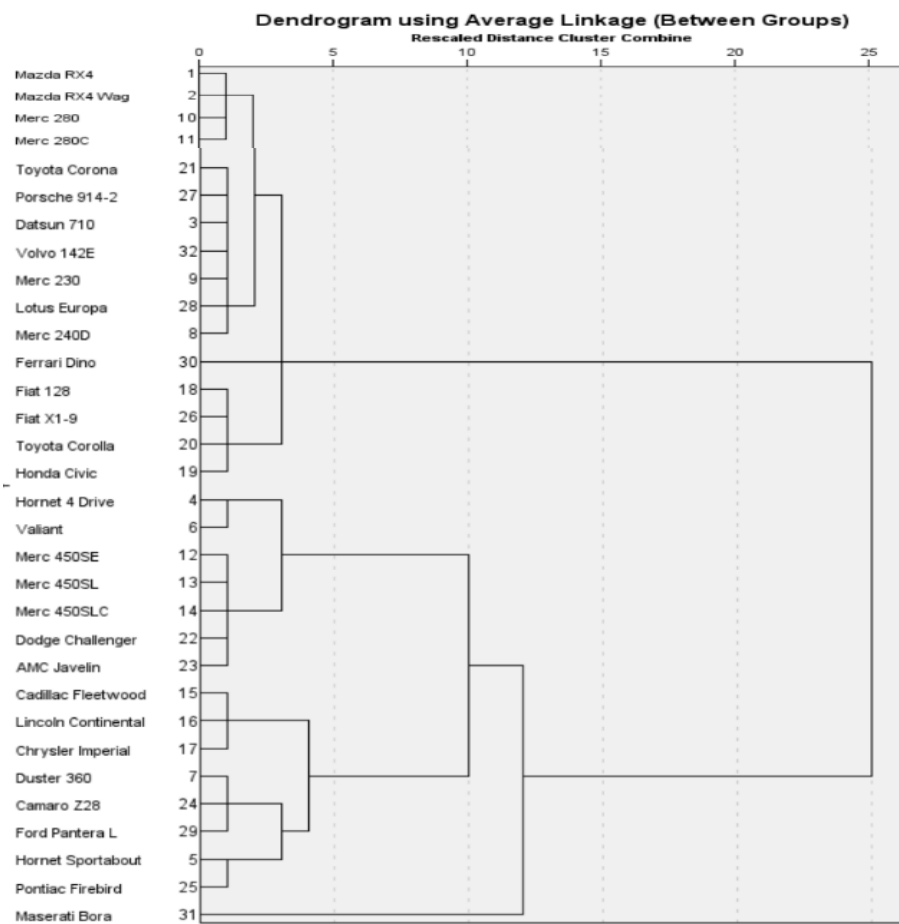
Number of Carburetors

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	7	21.9	21.9	21.9
	2	10	31.3	31.3	53.1
	3	3	9.4	9.4	62.5
	4	10	31.3	31.3	93.8
	6	1	3.1	3.1	96.9
	8	1	3.1	3.1	100.0
	Total	32	100.0	100.0	

Interpretation: 31.3% of cars in data has 2 or 4 carburetors and 22% of cars have 1 carburetor and only 3.1% of cars has 6 or 8 carburetors indicating that mostly cars have

either 2 or 4 carburetors which improves the engine power but engine power higher than it, is least preferred by the customers.

Hierarchical Clustering



Interpretation: A dendrogram is a tree-like structure that explains the relationship between all the data points in the system. The dendrogram shows clusters of vehicles with similar characteristics. Eg. Mazda RX4, Mazda RX4 Wag, Merc 280, Merc 280c has similar characteristics, forming a single cluster. Majorly, there are two clusters where one group is from Mazda RX4 to Honda Civic and another cluster is from Hornet 4 Drive to Maserati Bora.

Similarly, Fiat 128, FiatX1-9, Toyota Corolla and Honda Civic forms another cluster with similar features. All of them have 4 number of cylinders, lower engine power in the range of 52 to 66. All cars in the cluster are manual with 4 gears and less (1 or 2) number of carburetors.

Another cluster of Duster 360, Camor Z28 and Ford Pantera L have lower mileage per gallon, having 8 cylinders, displacement in high range indicating high power. The cars in this cluster has higher engine capacity of 245 or 264 hp.

Regression Analysis

Dependent variable: Miles per gallon

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.931 ^a	.867	.821	2.5479

a. Predictors: (Constant), Number of Carburetors, Automatic or Manual Transmission, Quarter-Mile (seconds), Displacement (cu. in.), Number of Gears, Horsepower, Number of Cylinders, Weight (tons)

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	976.734	8	122.092	18.807	<.001 ^b
	Residual	149.314	23	6.492		
	Total	1126.047	31			

a. Dependent Variable: Miles per Gallon

b. Predictors: (Constant), Number of Carburetors, Automatic or Manual Transmission, Quarter-Mile (seconds), Displacement (cu. in.), Number of Gears, Horsepower, Number of Cylinders, Weight (tons)

In regression output, the p-value for the F-statistic, which tests the overall significance of the regression model, is shown as 0.001 in the ANOVA table. The p-value is less than 0.05 indicating that the model as a whole explains a significant amount of variance in the dependent variable, meaning the independent variables collectively have a meaningful effect.

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	15.865	16.244		.977
	Number of Cylinders	-.307	.907	-.091	.738
	Displacement (cu. in.)	.014	.017	.289	.416
	Horsepower	-.021	.020	-.244	.297
	Weight (tons)	-3.897	1.788	-.633	.040
	Quarter-Mile (seconds)	.852	.654	.253	.205
	Automatic or Manual Transmission	2.605	1.910	.216	.186
	Number of Gears	.721	1.430	.088	.619
	Number of Carburetors	-.129	.777	-.035	.870

a. Dependent Variable: Miles per Gallon

Interpretation: As our model has many independent variables, therefore adjusted R-square is considered to assess the model accuracy. Adjusted R-squared adjusts the standard R-squared by penalizing the inclusion of additional predictors that don't improve the model's fit. The value of adjusted R^2 of 0.82 indicates that 82% of variance in dependent variable (mileage per gallon) is explained by the independent variables. This states that regression model fits the data well.

In the coefficients table, the significance column consists of the p-value for each independent variable in the model. It is concluded that only weight variable has p-value of 0.04 which is < 0.05 , indicating that the weight has significant relationship with dependent variable (miles per gallon)

The value of beta coefficient of independent variables shows the strength of relationship between dependent variable (mpg) and independent variable. Here, weight has highest strength of -0.63 indicating the strong relationship between weight and mpg. However -ve sign indicates that as weight decreases, the mpg increases.