

Intelligent Data Analysis

Names:

Daniel

Panagiotis

Mikel

Ignacio

Javier

Introduction

As all chose the Mileage dataset, it contains data about the fuel consumption of different car manufacturers which had a new release every year between 1999 and 2008. The dataset contains 234 records of 11 different variables, which are:

manufacturer: car manufacturer, categorical variable.

Manufacturer	Audi	Chevrolet	Dodge	Ford	Honda	Hyundai	Jeep	Land Rover
n° records	18	19	37	25	9	14	8	4
Manufacturer	Lincoln	Mercury	Nissan	Pontiac	Subaru	Toyota	Volkswagen	
n° records	3	4	13	5	14	34	27	

model: car model name (38 different models), categorical variable.

displ: engine displacement or cylinder admission volume in litres, quantitative variable.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1,6	2,4	3,3	3,472	4,6	7

year: year of manufacturing (1999 or 2008), categorical variable.

n° years	1999	2008
n° entries	117	117

cyl: number of cylinders, categorical variable.

n° gears	4	5	6	8
n° entries	81	4	79	70

trans: type of transmission (automatic or manual), categorical variable.

Transm	Auto	Manual
n° entries	156	77

dvr: drive type (front wheel, rear wheel, 4 wheel), categorical variable.

Drive type	4-wheel	front-wheel	rear-wheel
n° entries	103	105	25

cty: city mileage in milles per gallon, quantitative variable.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
9	14	17	16,83	19	35

hwy: highway mileage in miles per gallon, quantitative variable.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
12	18	24	23,39	27	44

fl: fuel type (petrol, diesel electric, etanol, regular), categorical variable.

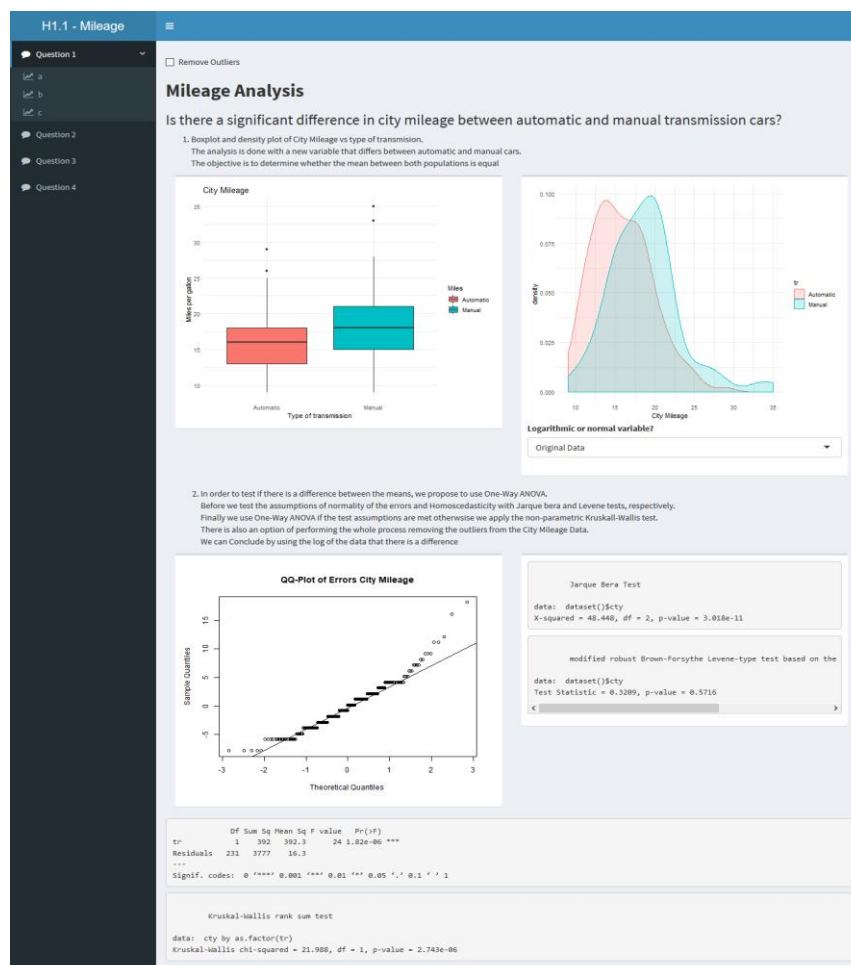
Fuel type	Diesel-elect	Ethanol	Petrol	Regular	"c"
nº entries	5	8	52	168	1

class: vehicle class, categorical variable

Class	2seater	Compact	Midsize	Minivan	Pickup	Subcompact	Suv
nº entries	5	47	41	11	33	35	62

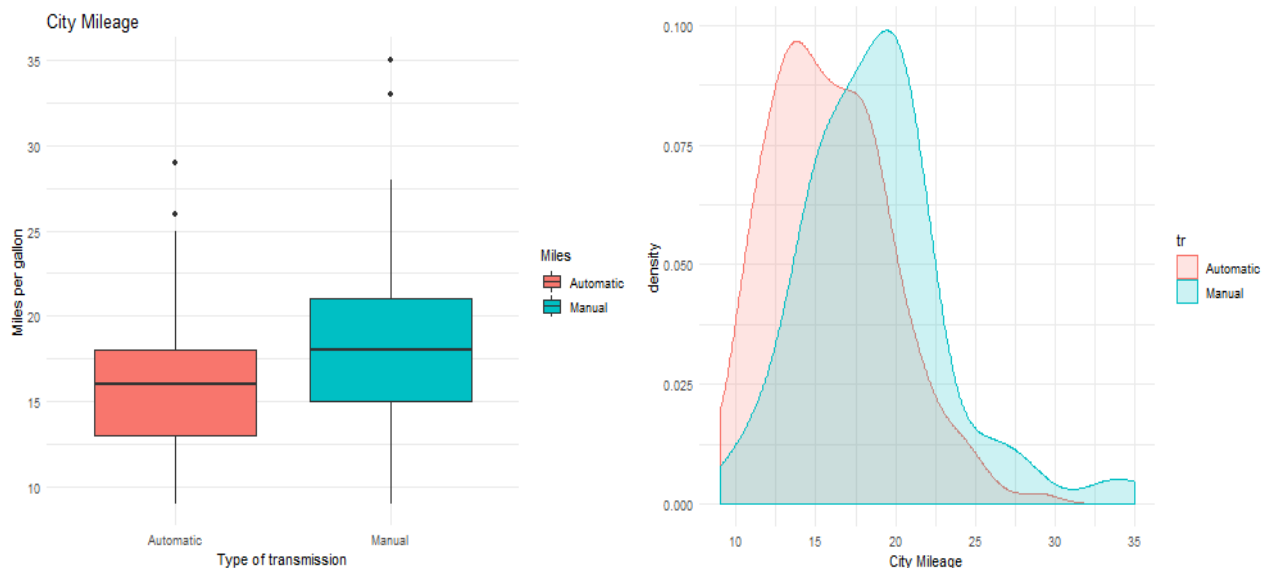
Shiny Dashboard Application

For this exercise we decided to create a shiny dashboard application that allowed us to express our analysis in a better looking and more convenient way. We have included our results of the statistical tests for every question under every plot in our shiny application.



Q1: Is there a significant difference in City / Highway mileage or in Engine Displacement between automatic and manual transmission cars?

This question is divided in three parts depending on the three variables we are analysing: city milage,highway milage and engine displacement.On this page only the plots for City mileage are presented:



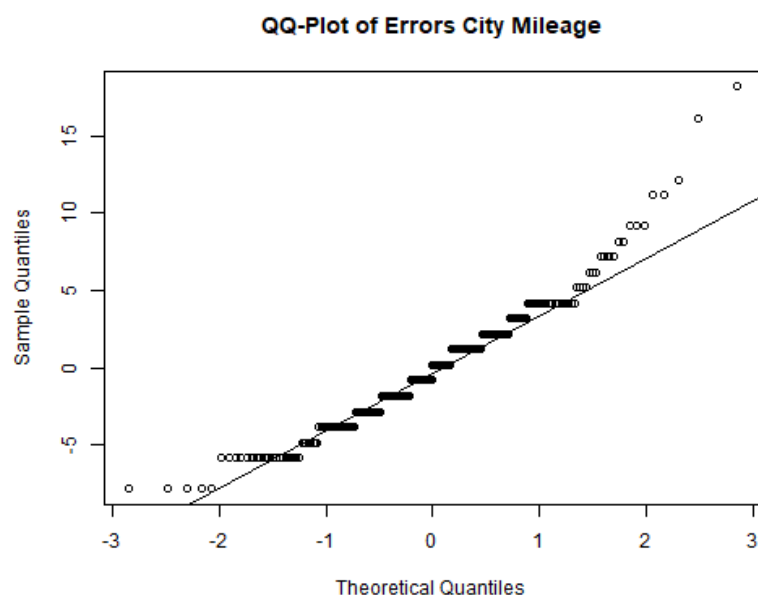
Boxplot and density plot of City Mileage vs type of transmission.

The analysis is done with a new variable that differs between automatic and manual cars.

The objective is to determine whether the mean between both populations is equal.

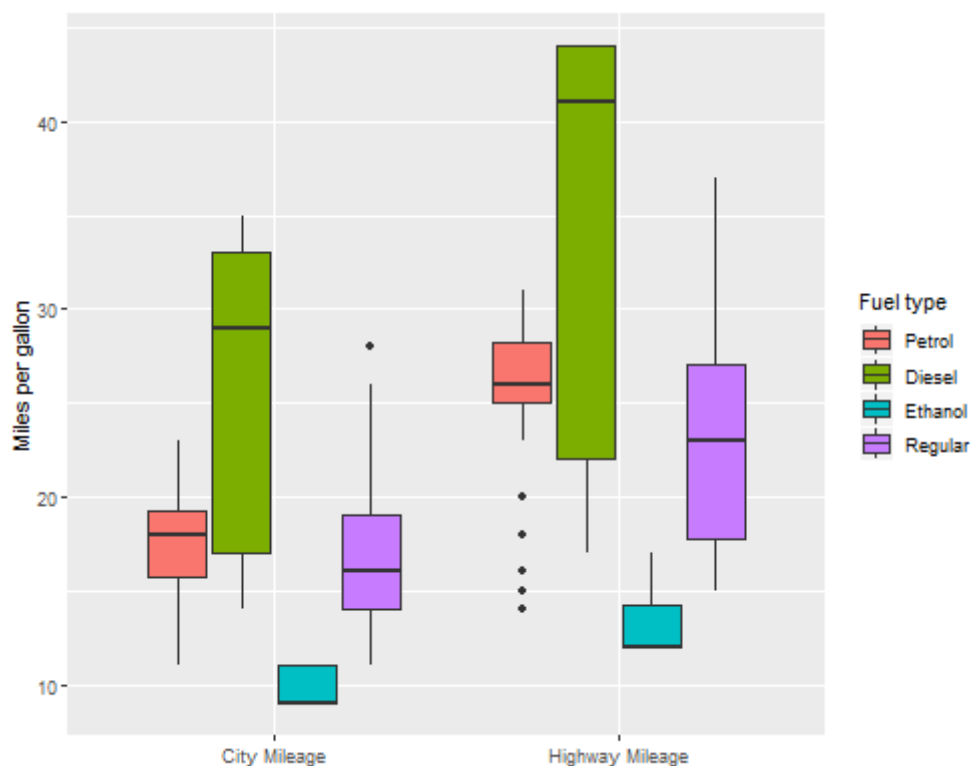
In order to test if there is a difference between the means, we propose to use One-Way ANOVA.

Before we test the assumptions of normality of the errors and Homoscedasticity with Jarque bera and Levene tests, respectively.Finally we use One-Way ANOVA if the test assumptions are met otherwise we apply the non-parametric Kruskal-Wallis test. There is also an option of performing the whole process removing the outliers from the City Mileage Data. We can Conclude by using the log of the data that there is a difference.

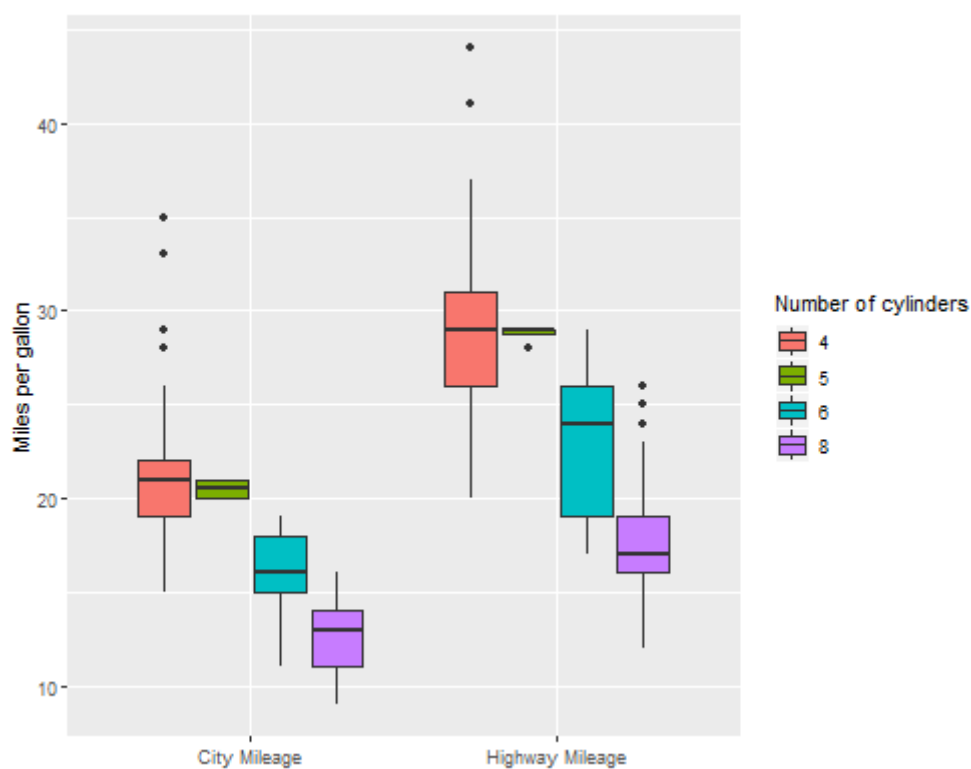


Q2: Has fuel type (or cylinders) any influence on highway mpg or city mpg?

For this question we have created Boxplots comparing the City Mileage to the Highway Mileage with respect to : A)The fuel type of each automobile.



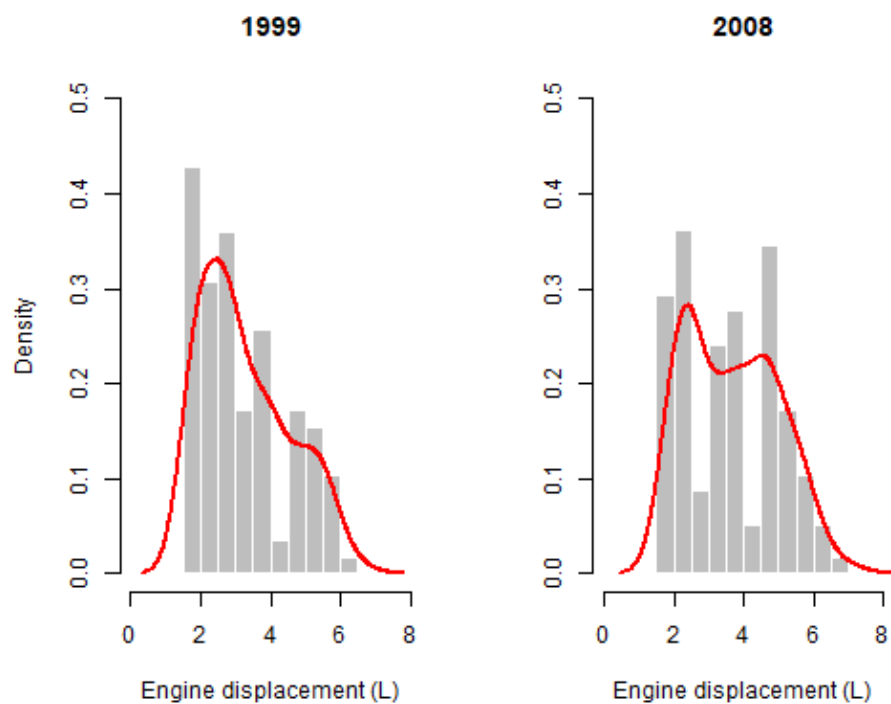
B)The number of cylinders each automobile has.



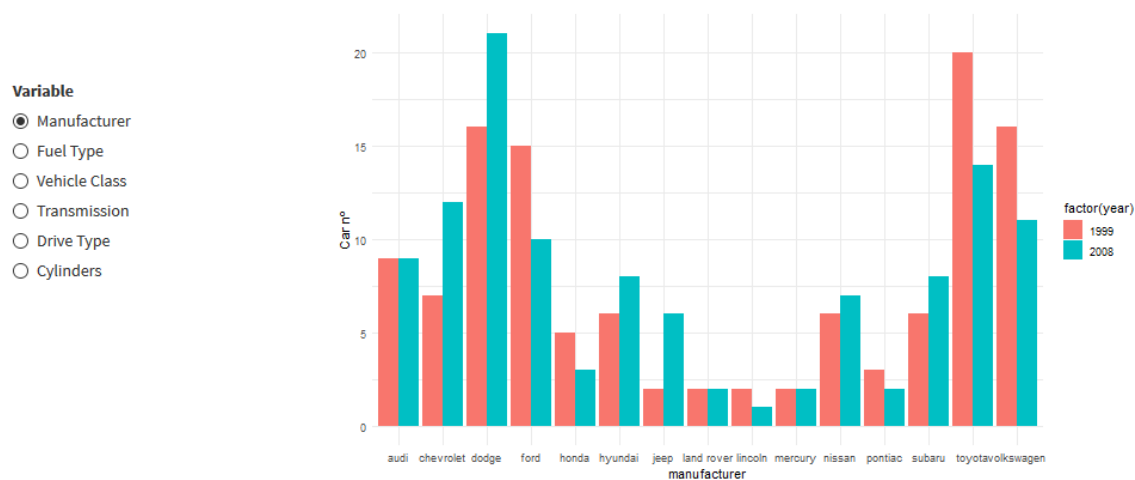
Q3: Is there any difference in car requests between the years 1999 and 2008?

For this question we compared the distribution of some of the variables in years 1999 and 2008 in order to determine whether the requests on automobiles have been changed.

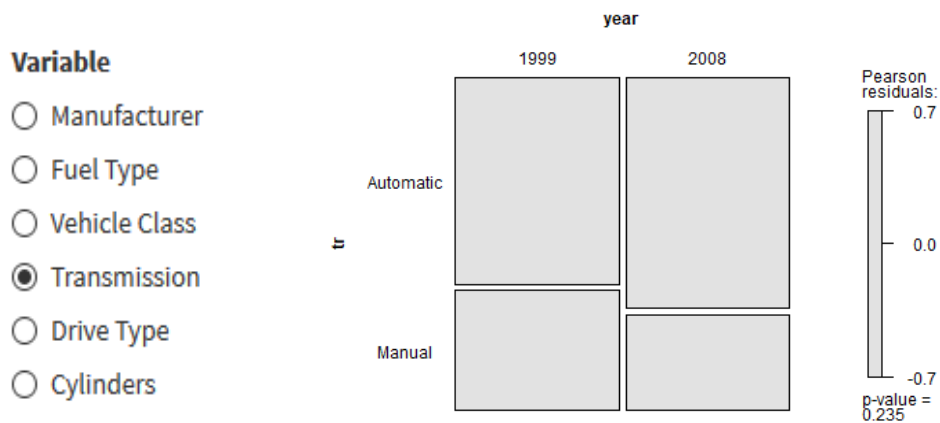
On the histogram plot we compare the distribution of the engine displacement for each automotive through the years.



We have used a reactive bar chart to visualize how the different levels of the categorical variables are distributed between 1999 and 2008.



Following with a reactive Mosaic plot.

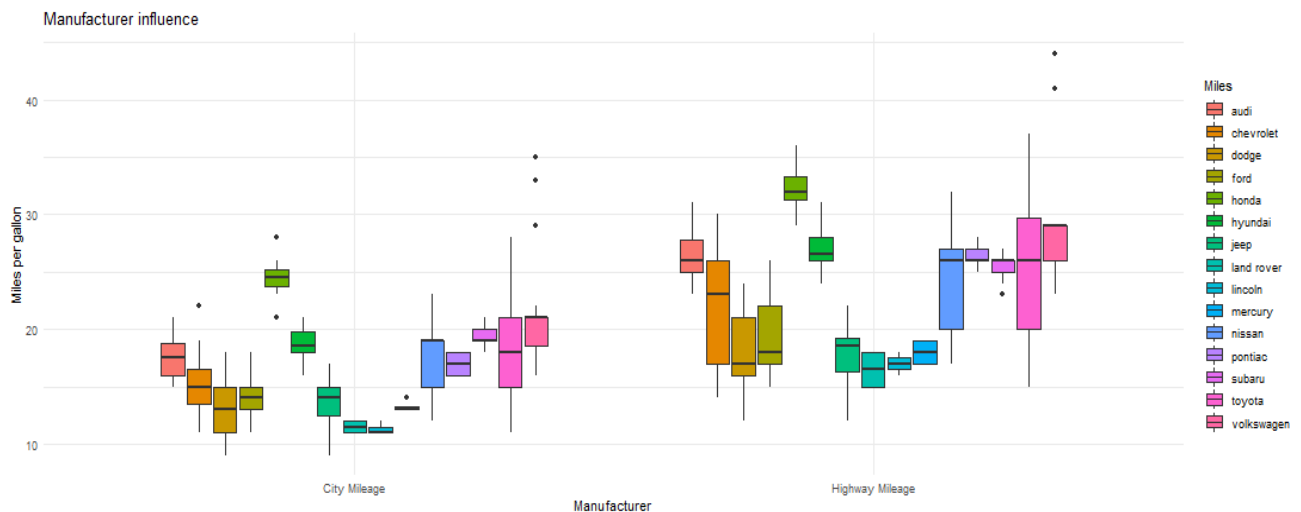


We have also included a reactive chi-test table:

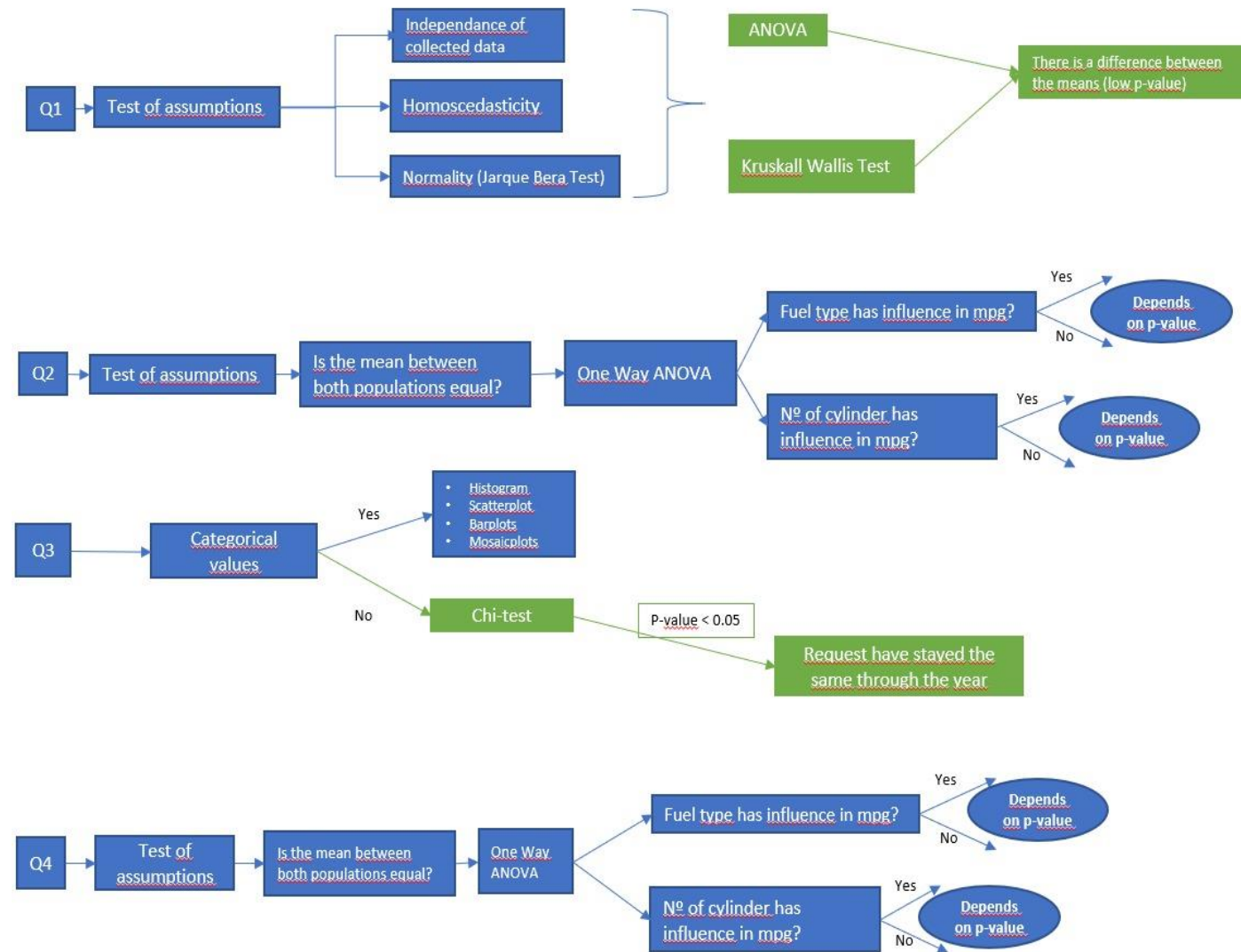
Variable	Variable	Chi_Test_p_value
<input type="radio"/> Manufacturer	d	0.179712494879
<input checked="" type="radio"/> Fuel Type	p	0.0960923294556731
<input type="radio"/> Vehicle Class	r	0.164914822553301
<input type="radio"/> Transmission	e	NA
<input type="radio"/> Drive Type		
<input type="radio"/> Cylinders		

Q4: Is there any difference in city and highway mpg between manufacturers?

For this question we have used a double box plot to show how the different manufacturers engines behave in terms of consumption for City driving and Highway driving.



Data Analysis Plan



Q1: Is there a significant difference in city mileage between automatic and manual transmission cars?
 Q2: Has fuel type (or cylinders) any influence on highway mpg or city mpg?
 Q3: Has the requests on automobiles changed from 1999 to 2008?
 Q4: Is there any difference in city and highway mpg between manufacturers?