

ANLY 511 Final Project - MANOVA

LINLIN WANG

2022-12-02

MANOVA

MANOVA stands for multivariate analysis of variance. It's basically used to evaluate mean differences on two or more dependent variables simultaneously. That's the main difference compared with ANOVA.

We try to answer below research questions by performing MANOVA. And CO2, CO, and THC are three dependent variables.

Research Questions

1. We want to know if there is statistically significant difference in CO2, CO, and THC between the different types of vehicle.
2. We want to know if there is statistically significant difference in CO2, CO, and THC between vehicle manufacturers.
3. We want to know if there is statistically significant difference in CO2, CO, and THC between the different vehicle transmission types.

Assumptions of MANOVA

There are additional assumptions of MANOVA:

1. Homogeneity of the variances across the range of predictors:

Our data should have equal variance-covariance matrices for each combination formed by each group in the independent variable.

2. Multicollinearity:

Our data should be no multicollinearity among dependent variables.

Loading packages

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.7      v dplyr  1.0.9
## v tidyr   1.2.0      v stringr 1.4.0
```

```
## v readr 2.1.2 v forcats 0.5.1
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()
```

```
library(ggpubr)
library(rstatix)
```

```
##
## Attaching package: 'rstatix'
##
## The following object is masked from 'package:stats':
##
## filter
```

```
library(car)
```

```
## Loading required package: carData
##
## Attaching package: 'car'
##
## The following object is masked from 'package:dplyr':
##
## recode
##
## The following object is masked from 'package:purrr':
##
## some
```

```
library(broom)
library(gplots)
```

```
##
## Attaching package: 'gplots'
##
## The following object is masked from 'package:stats':
##
## lowess
```

```
library(mvnormalTest)
library(heplots)
```

Enter Data

```
# Read cleaned nonelectric data
nonelectric<-read.csv("/Users/linlinw/Desktop/ANLY511-Final-Project-main/data/cardata_nonelectric_clean")
# Remove first column X
nonelectric<-nonelectric[,-1]
# View first couple rows of data
head(nonelectric)
```

##	Model.Year	Vehicle.Manufacturer.Name	Veh.Mfr.Code	Represented.Test.Veh.Make	
## 1	2018	aston martin	ASX	Aston Martin	
## 2	2018	aston martin	ASX	Aston Martin	
## 3	2018	aston martin	ASX	Aston Martin	
## 4	2018	aston martin	ASX	Aston Martin	
## 5	2018	aston martin	ASX	Aston Martin	
## 6	2018	aston martin	ASX	Aston Martin	
##	Represented.Test.Veh.Model	Test.Veh.Displacement..L.	Vehicle.Type		
## 1	DB11	5.2	Car		
## 2	DB11	5.2	Car		
## 3	DB11 V8	4.0	Car		
## 4	DB11 V8	4.0	Car		
## 5	Rapide S	6.0	Car		
## 6	Rapide S	6.0	Car		
##	Rated.Horsepower	X..of.Cylinders.and.Rotors	Tested.Transmission.Type.Code		
## 1	600	12	SA		
## 2	600	12	SA		
## 3	503	8	SA		
## 4	503	8	SA		
## 5	552	12	SA		
## 6	552	12	SA		
##	Tested.Transmission.Type	X..of.Gears	Transmission.Lockup.	Drive.System.Code	
## 1	Semi-Automatic	8	Y	R	
## 2	Semi-Automatic	8	Y	R	
## 3	Semi-Automatic	8	Y	R	
## 4	Semi-Automatic	8	Y	R	
## 5	Semi-Automatic	8	Y	R	
## 6	Semi-Automatic	8	Y	R	
##	Drive.System.Description	Equivalent.Test.Weight..lbs..	Axle.Ratio	N.V.Ratio	
## 1	2-Wheel Drive, Rear	4500	2.70	22.2	
## 2	2-Wheel Drive, Rear	4500	2.70	22.2	
## 3	2-Wheel Drive, Rear	4500	2.70	22.2	
## 4	2-Wheel Drive, Rear	4500	2.70	22.2	
## 5	2-Wheel Drive, Rear	4750	2.73	22.4	
## 6	2-Wheel Drive, Rear	4750	2.73	22.4	
##	Test.Fuel.Type.Description	THC..g.mi.	CO..g.mi.	CO2..g.mi.	RND_ADJ_FE
## 1	Tier 2 Cert Gasoline	0.024700	0.418000	466.87	18.8
## 2	Tier 2 Cert Gasoline	0.001155	0.067334	285.00	30.9
## 3	Tier 2 Cert Gasoline	0.026500	0.070000	386.66	22.7
## 4	Tier 2 Cert Gasoline	0.000500	0.030000	259.74	33.8
## 5	Tier 2 Cert Gasoline	0.026900	0.500000	511.93	17.3
## 6	Tier 2 Cert Gasoline	0.000800	0.060000	296.63	29.9
##	DT.Inertia.Work.Ratio.Rating	DT.Absolute.Speed.Change.Ratg			
## 1	-2.5300000	-1.7300000			
## 2	1.3600000	0.4400000			
## 3	-11.9900000	-9.2600000			
## 4	-3.6400000	-3.2100000			
## 5	0.5655838	0.4420405			
## 6	0.5655838	0.4420405			
##	DT.Energy.Economy.Rating	Target.Coeff.A..lbf.	Target.Coeff.B..lbf.mph.		
## 1	-1.7100000	40.94	0.0169		
## 2	-0.5900000	40.94	0.0169		
## 3	-7.7100000	40.94	0.0169		
## 4	-0.9600000	40.94	0.0169		

```
## 5          -0.2002973          32.66          0.6085
## 6          -0.2002973          32.66          0.6085
## Target.Coeff.C..lbf.mph..2. Set.Coeff.A..lbf. Set.Coeff.B..lbf.mph.
## 1          0.0271          6.810          0.0807
## 2          0.0271          6.810          0.0807
## 3          0.0271          11.260          0.0919
## 4          0.0271          11.260          0.0919
## 5          0.0198          1.093          2.1980
## 6          0.0198          1.093          2.1980
## Set.Coeff.C..lbf.mph..2. Aftertreatment.Device.Cd Aftertreatment.Device.Desc
## 1          0.0245          TWC          Three-way catalyst
## 2          0.0245          TWC          Three-way catalyst
## 3          0.0251          TWC          Three-way catalyst
## 4          0.0251          TWC          Three-way catalyst
## 5          0.0280          TWC          Three-way catalyst
## 6          0.0280          TWC          Three-way catalyst
## Police...Emergency.Vehicle. Averaging.Method.Cd Averging.Method.Desc
## 1          N          N          No averaging
## 2          N          N          No averaging
## 3          N          N          No averaging
## 4          N          N          No averaging
## 5          N          N          No averaging
## 6          N          N          No averaging
```

Research Question 1: Is there any important difference in CO₂, CO, and THC between the different types of vehicle?

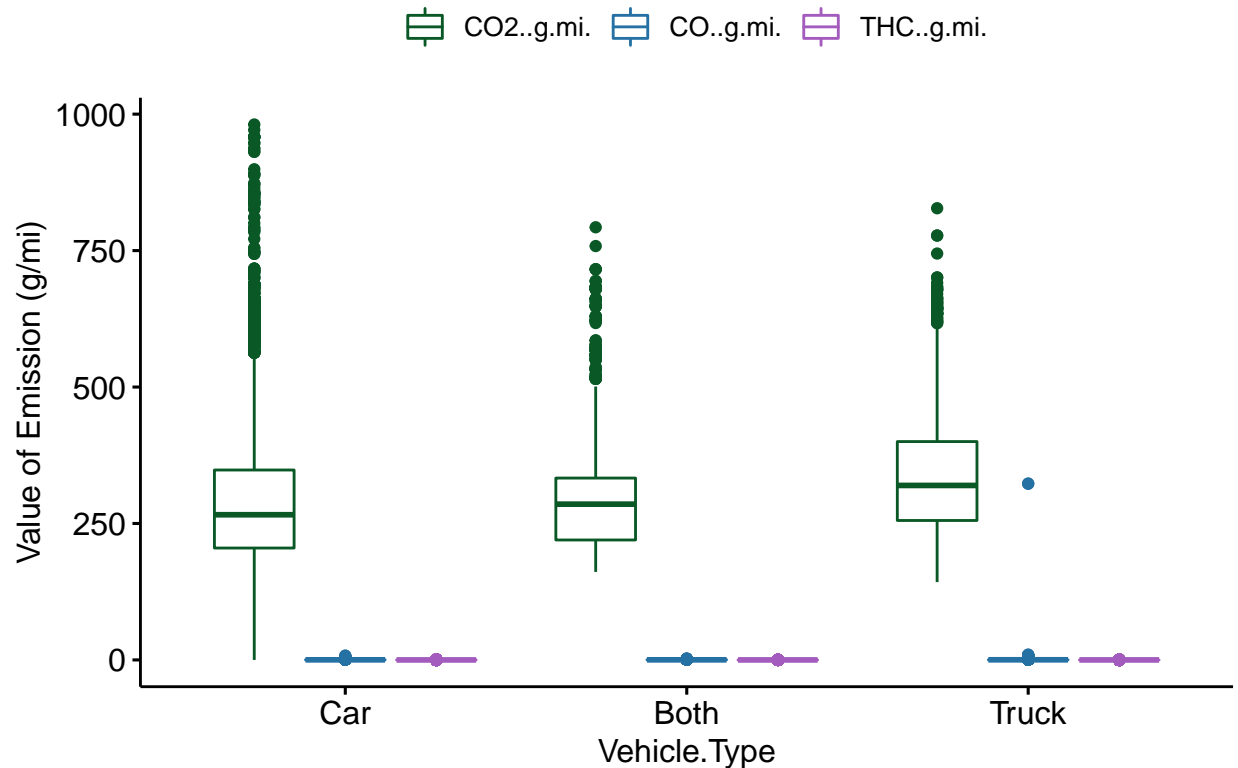
Exploratory Data Analysis:

```
# Create color palettes
Blues <- colorRampPalette(c("#0A146B", "#A9A3DA"))
Purples <- colorRampPalette(c("#3E1370", "#BDA3DA"))
GrBuPuPi <- c("#095826", "#0E7032", "#10913F", "#55A472", "#8CBF9E", "#8CBFB8",
              "#63B7AC", "#2D9A8B", "#137568", "#094E45", "#0B3C5C", "#17547C",
              "#2671A4", "#3C8CC1", "#72B1DB", "#96C3E1", "#B0CDE1", "#B0B3E1",
              "#858ACD", "#4F55AB", "#1923B3", "#0E1468", "#3C1075", "#5821A1",
              "#6B27C4", "#9455E5", "#A278D8", "#A990CA", "#ADA0BF", "#C1A5CB",
              "#B887CA", "#A35CBD", "#762594")

# Visualize dataset
ggboxplot(
  nonelectric, x = "Vehicle.Type", y = c("CO2..g.mi.", "CO..g.mi.", "THC..g.mi."),
  merge = TRUE, palette = c("#095826", "#2671A4", "#A35CBD" ),
  title = "Three Different Emissions for Vehicle.Type",
  ylab = "Value of Emission (g/mi)"
)
```

```
## Warning: 'gather_()' was deprecated in tidyr 1.2.0.
## Please use 'gather()' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was generated.
```

Three Different Emissions for Vehicle.Type



From the above boxplot, we can see there is significant difference in CO2 and CO between vehicle type, and significant difference in CO2 and THC between vehicle type. However, the difference in CO and THC between vehicle type is not obvious.

Test assumption

1. Check for Homogeneity

```
boxM(Y = nonelectric[, c("CO2..g.mi.", "CO..g.mi.", "THC..g.mi.")], group = nonelectric$Vehicle.Type)

##
## Box's M-test for Homogeneity of Covariance Matrices
##
## data: nonelectric[, c("CO2..g.mi.", "CO..g.mi.", "THC..g.mi.")
## Chi-Sq (approx.) = 58734, df = 12, p-value < 2.2e-16
```

Since the p-value is significant for Box's M test, we reject the null hypothesis at 5% significance level and conclude that variance-covariance matrices are not equal for each combination formed by each group in the independent variable. Thus, this assumption is satisfied.

2. Check Multicollinearity

```
cor_co2_co<-cor.test(x = nonelectric$CO2..g.mi., y = nonelectric$CO..g.mi., method = "pearson")$estimate
cor_co2_thc<-cor.test(x = nonelectric$CO2..g.mi., y = nonelectric$THC..g.mi., method = "pearson")$estimate
cor_thc_co<-cor.test(x = nonelectric$THC..g.mi., y = nonelectric$CO..g.mi., method = "pearson")$estimate
result<-cbind(cor_co2_co, cor_co2_thc, cor_thc_co)
result
```

```
##      cor_co2_co cor_co2_thc cor_thc_co
## cor   0.0562401   0.3195182 0.07842236
```

As the correlation coefficient between each dependent variable is smaller than 0.9, there is no multicollinearity. So this assumption is hold.

Perform MANOVA

Hypotheses

H_0 : There is no significant difference in CO₂, CO, and THC between the different types of vehicle.

H_a : There is significant difference in CO₂, CO, and THC between the different types of vehicle.

```
# Fit the MANOVA model
fit1 = manova(cbind(CO2..g.mi., CO..g.mi., THC..g.mi.) ~ Vehicle.Type, data = nonelectric)
summary(fit1, intercept = TRUE)
```

```
##              Df  Pillai approx F num Df den Df      Pr(>F)
## (Intercept)    1 0.88848    57714     3 21733 < 2.2e-16 ***
## Vehicle.Type    2 0.04618     171     6 43468 < 2.2e-16 ***
## Residuals      21735
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p-value for Vehicle.Type variable is smaller than the significance level 0.05, we can reject the null hypotheses at 5% significance level, and conclude that there is statistically significant difference in CO₂, CO, and THC between the different types of vehicle.

However, we are unclear about which emissions are affected by vehicle type. We perform univariate ANOVAs to figure it out.

```
summary.aov(fit1)
```

```
## Response CO2..g.mi. :
##              Df    Sum Sq Mean Sq F value    Pr(>F)
## Vehicle.Type    2 10263186 5131593  421.74 < 2.2e-16 ***
## Residuals      21735 264465545   12168
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Response CO..g.mi. :
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Vehicle.Type    2   126   62.795   6.4026 0.00166 **
## Residuals      21735 213170    9.808
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Response THC..g.mi. :
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Vehicle.Type    2  0.147  0.073417  35.578 3.751e-16 ***
## Residuals      21735 44.852  0.002064
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can see from the output that the p-value for all univariate ANOVAs are smaller than significance level 0.05, which indicates that vehicle type has a statistically significant effect on CO₂, CO, and THC emissions.

Visualizing Group Means

Visualizing the Group means for each level of our independent variable vehicle type is helpful to get a better understanding of our results.

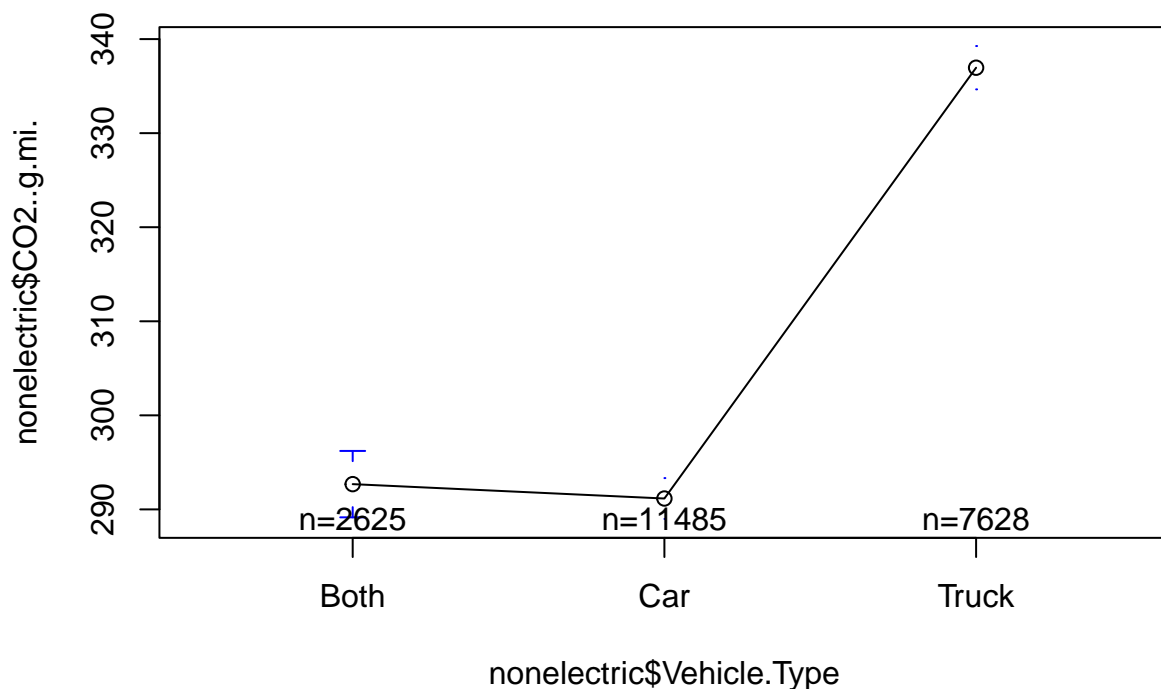
```
#visualize mean CO2 by vehicle type
plotmeans(nonelectric$CO2..g.mi. ~ nonelectric$Vehicle.Type)

## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```



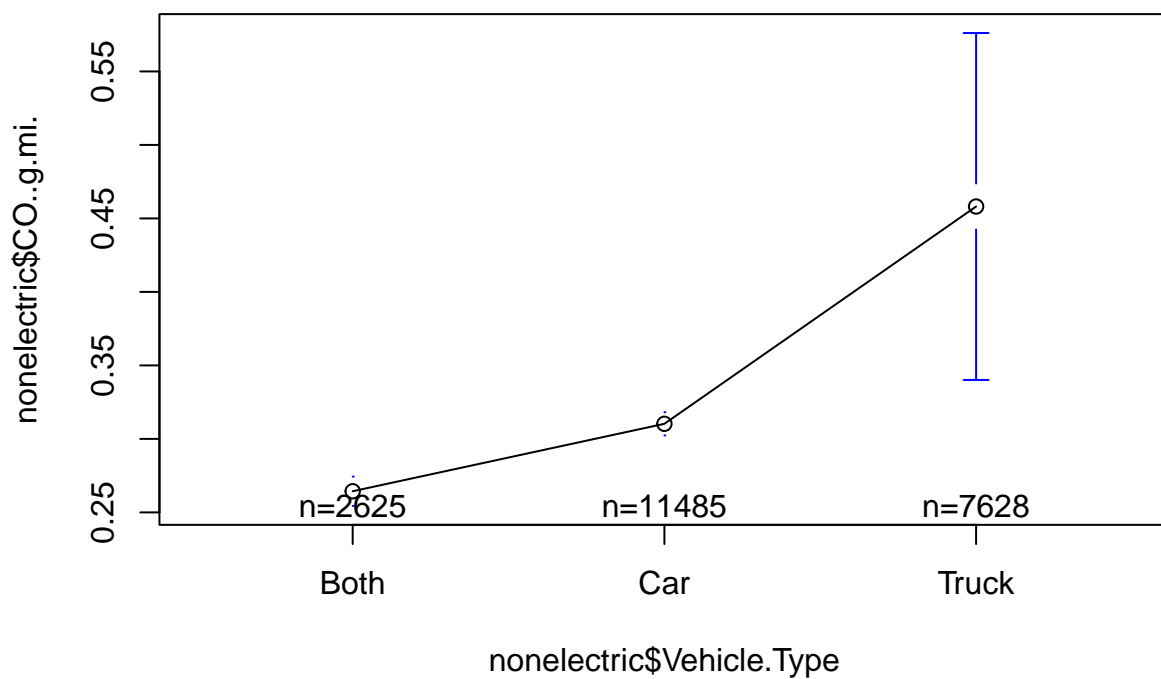
```
#visualize mean CO by vehicle type
plotmeans(nonelectric$CO..g.mi. ~ nonelectric$Vehicle.Type)
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

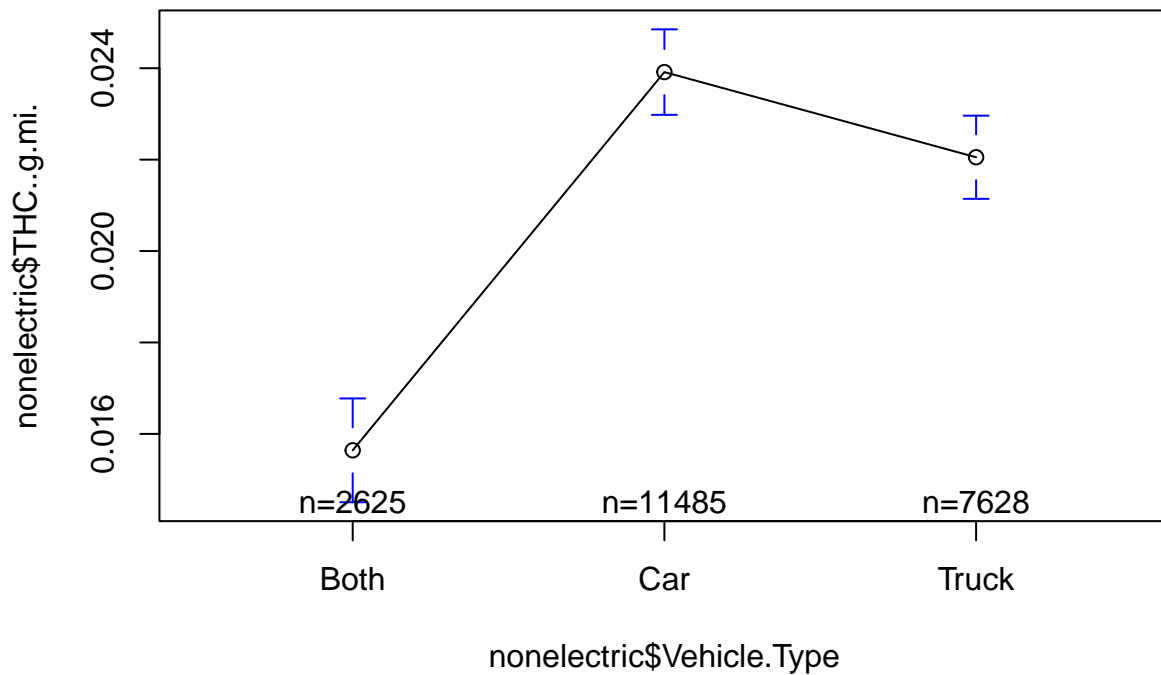
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```



```
#visualize mean THC by vehicle type
plotmeans(nonelectric$THC..g.mi. ~ nonelectric$Vehicle.Type)
```

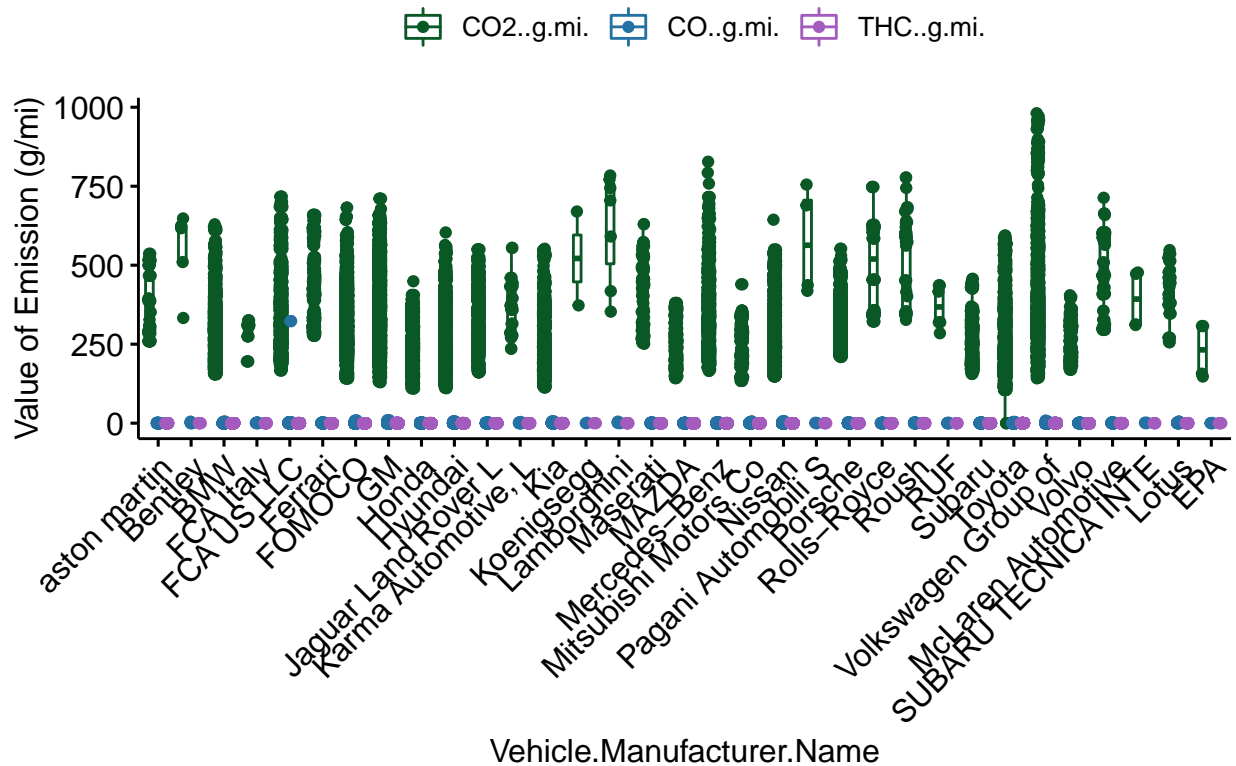



From the above three plots, we can see that the mean CO₂, CO, and THC varies quite a bit by vehicle types. This matches the results from our MANOVA, which indicates that there is statistically significant difference in three emissions based on vehicle types.

Research Question 2: Is there any important difference in CO₂, CO, and THC between the different vehicle manufacturers?

```
# Visualize dataset
p2<-ggboxplot(
  nonelectric, x = "Vehicle.Manufacturer.Name", y = c("CO2..g.mi.", "CO..g.mi.", "THC..g.mi."),
  merge = TRUE, palette = c("#095826", "#2671A4", "#A35CBD" ),
  title = "Three Different Emissions for Vehicle.Manufacturer",
  ylab = "Value of Emission (g/mi)",
  add = "jitter"
)
p2 + rotate_x_text(45)
```

Three Different Emissions for Vehicle.Manufacturer



From this boxplot, we can get same conclusion with first plot. There is no important difference in CO and THC between vehicle manufacturer.

Hypotheses2

H_0 : There is no significant difference in CO₂, CO, and THC between vehicle manufacturer.

H_a : There is significant difference in CO₂, CO, and THC between vehicle manufacturer.

Fit the MANOVA model

```
fit2 = manova(cbind(CO2..g.mi., CO..g.mi., THC..g.mi.) ~ Vehicle.Manufacturer.Name, data = nonelectric)
summary(fit2, intercept = TRUE)
```

```
##               Df  Pillai approx F num Df den Df    Pr(>F)
## (Intercept)      1  0.90421    68286      3 21703 < 2.2e-16 ***
## Vehicle.Manufacturer.Name 32  0.22328      55      96 65115 < 2.2e-16 ***
## Residuals      21705
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p-value for vehicle manufacturer variable is smaller than the significance level 0.05, we can reject the null hypotheses at 5% significance level, and conclude that there is statistically significant difference in CO₂, CO, and THC between the different types of vehicle.

Then we perform univariate ANOVAs to figure it out which emissions are affected by vehicle manufacturer

```
summary.aov(fit2)
```

```
## Response CO2..g.mi. :
##               Df      Sum Sq Mean Sq F value    Pr(>F)
## Vehicle.Manufacturer.Name    32  51944440 1623264  158.15 < 2.2e-16 ***
## Residuals                21705 222784291   10264
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Response CO..g.mi. :
##               Df Sum Sq Mean Sq F value    Pr(>F)
## Vehicle.Manufacturer.Name    32    701 21.8911   2.235 7.695e-05 ***
## Residuals                21705 212595   9.7947
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Response THC..g.mi. :
##               Df Sum Sq Mean Sq F value    Pr(>F)
## Vehicle.Manufacturer.Name    32   1.726 0.053943  27.057 < 2.2e-16 ***
## Residuals                21705 43.273 0.001994
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can see from the output that the p-value for all univariate ANOVAs are smaller than significance level 0.05, which indicates that Vehicle Manufacturer has a statistically significant effect on CO₂, CO, and THC emissions.

Visualizing Group Means

Visualizing the Group means for each level of our independent variable Vehicle.Manufacturer.Name is helpful to get a better understanding of our results.

```
#visualize mean CO2 by vehicle type
plotmeans(nonelectric$CO2..g.mi. ~ nonelectric$Vehicle.Manufacturer.Name) +rotate_x_text(45)
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```

[illegible]

```

## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

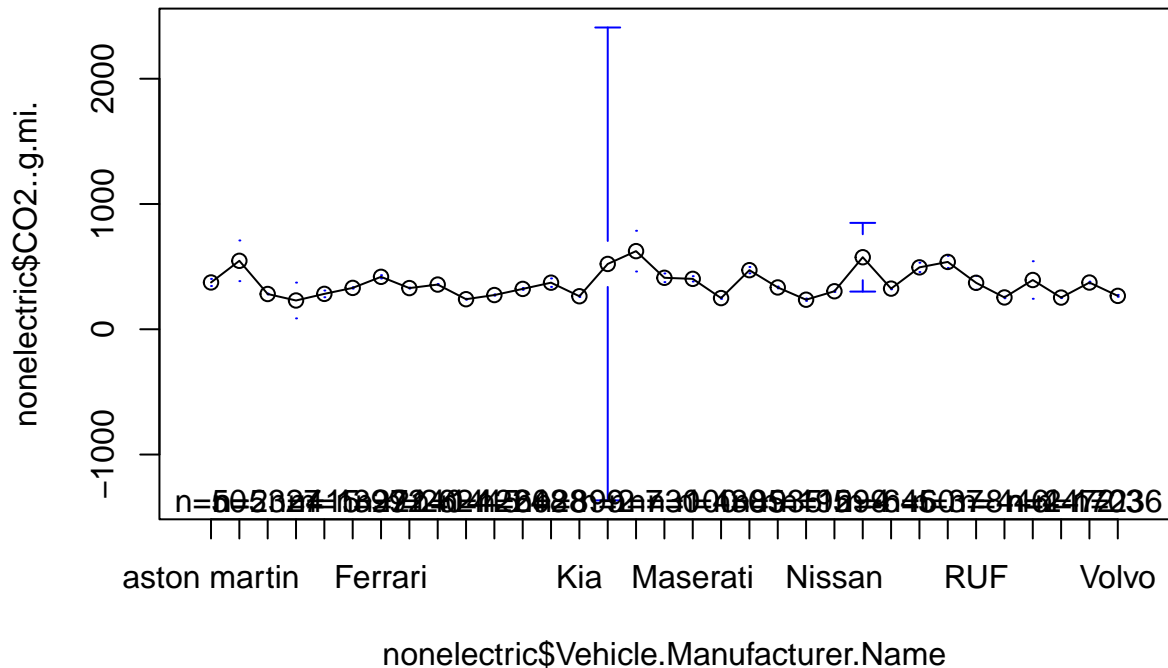
## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

```



```
## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```



```
## NULL
```

```
#visualize mean CO by vehicle type
```

```
plotmeans(nonelectric$CO..g.mil. ~ nonelectric$Vehicle.Manufacturer.Name)
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```

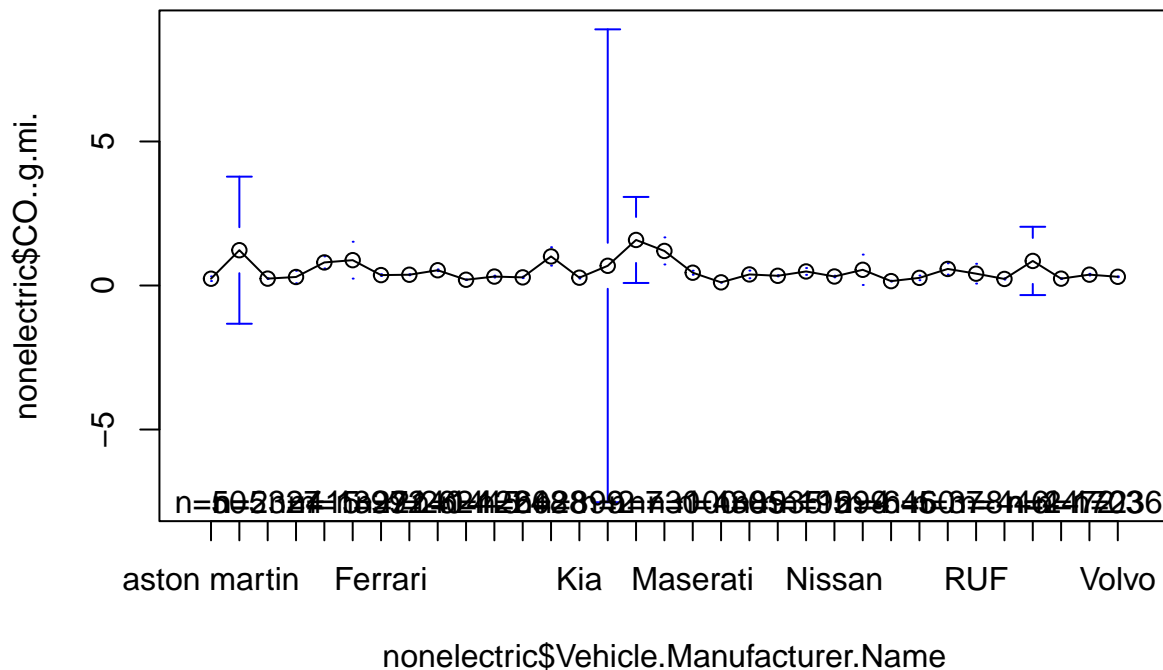
```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
```

[illegible]

[illegible]

[illegible]



```
#visualize mean THC by vehicle type
```

```
plotmeans(nonelectric$THC..g.mi. ~ nonelectric$Vehicle.Manufacturer.Name)
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-  
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-  
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-  
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-  
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-  
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-  
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-  
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
```

[illegible]


```

## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

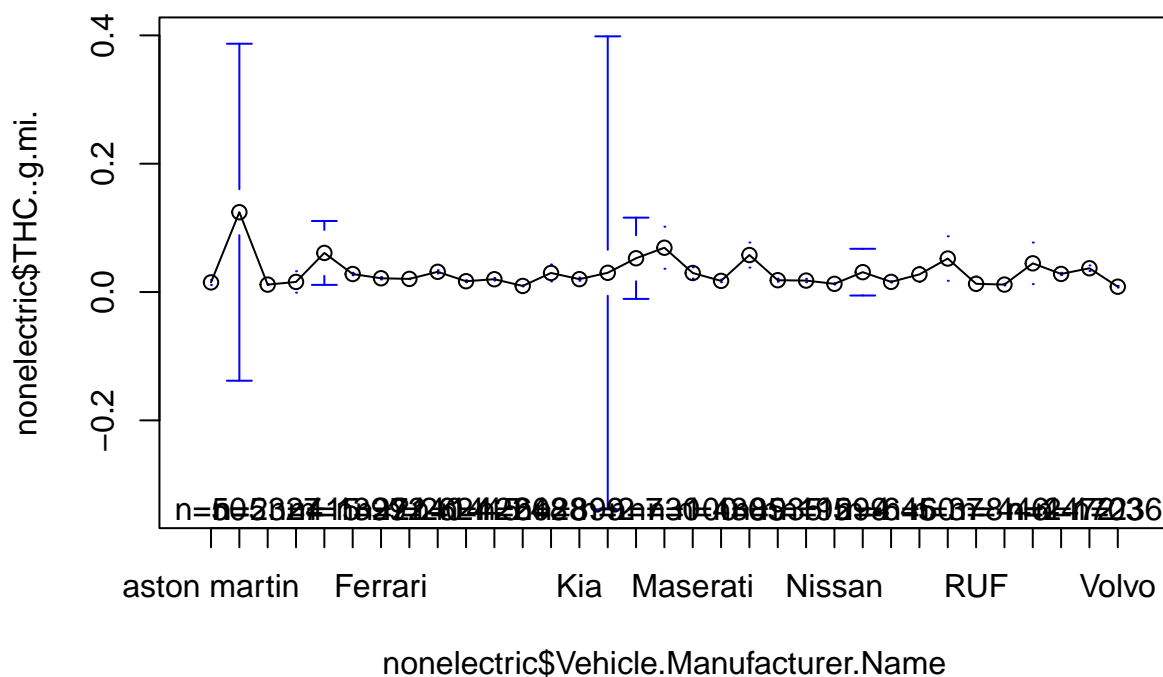
## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

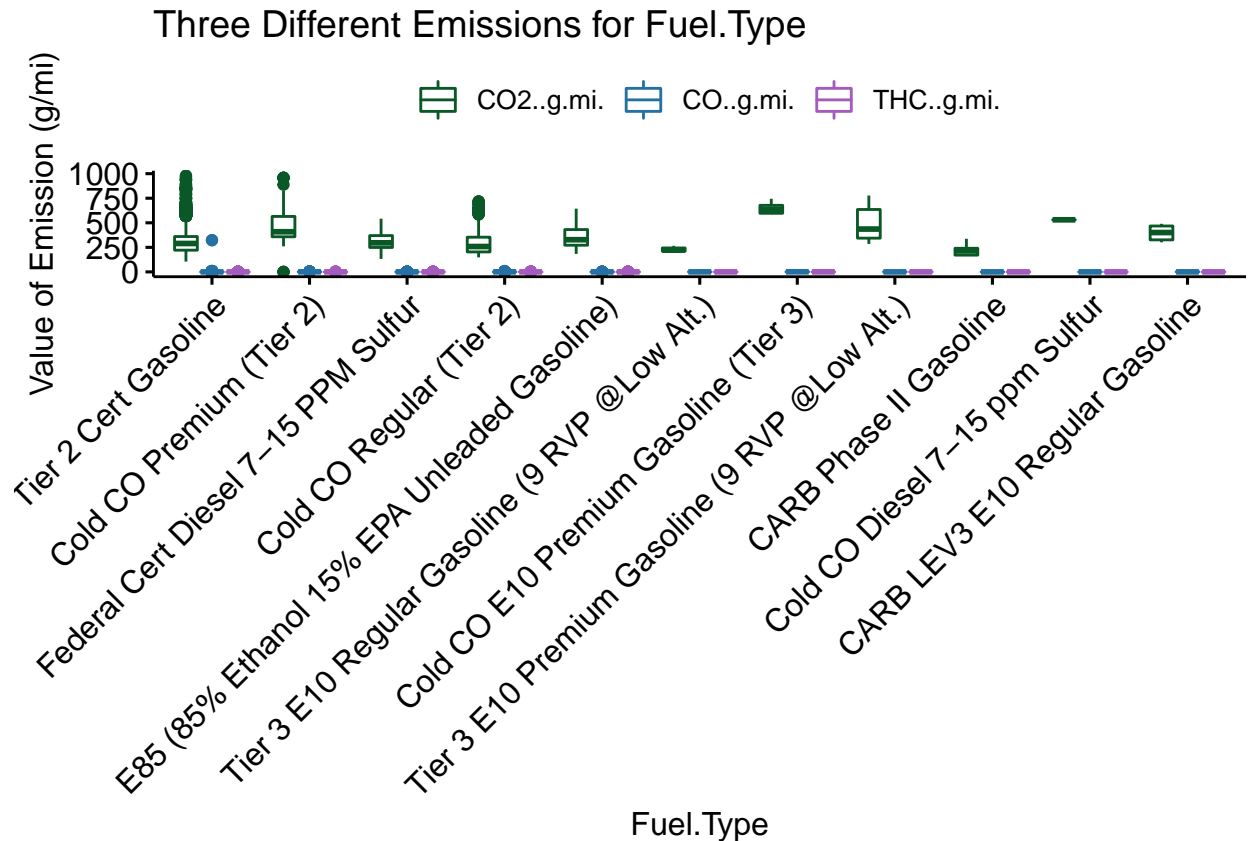
```



From the above three plots, we can see that the mean CO₂, CO, and THC do not vary a lot by vehicle manufacturer. This doesn't matches the results from our MANOVA.

Research Question 3: Is there any important difference in CO₂, CO, and THC between the different vehicle transmission types?

```
# Visualize dataset
p3<-ggboxplot(
  nonelectric, x = "Test.Fuel.Type.Description", y = c("CO2..g.mi.", "CO..g.mi.", "THC..g.mi."),
  merge = TRUE, palette = c("#095826", "#2671A4", "#A35CBD" ),
  title = "Three Different Emissions for Fuel.Type",
  ylab = "Value of Emission (g/mi)",
  xlab = "Fuel.Type"
)
p3 + rotate_x_text(45)
```



Hypotheses 3

H_0 : There is no significant difference in CO2, CO, and THC between the different types of fuel

H_a : There is significant difference in CO2, CO, and THC between the different types of fuel

Fit the MANOVA model

```
fit3 = manova(cbind(CO2..g.mi., CO..g.mi., THC..g.mi.) ~ Test.Fuel.Type.Description, data = nonelectric)
summary(fit3, intercept = TRUE)
```

```
##               Df  Pillai approx F num Df den Df    Pr(>F)
## (Intercept)      1  0.88787    57342      3 21725 < 2.2e-16 ***
## Test.Fuel.Type.Description  10  0.52439     460     30 65181 < 2.2e-16 ***
## Residuals                21727
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p-value for Fuel.Type variable is smaller than the significance level 0.05, we can reject the null hypotheses at 5% significance level, and conclude that there is statistically significant difference in CO2, CO, and THC between the different types of fuel.

However, we are unclear about which emissions are affected by fuel type. We perform univariate ANOVAs to figure it out.

```
summary.aov(fit3)
```

```
## Response CO2..g.mi. :
```



```
##              Df      Sum Sq Mean Sq F value    Pr(>F)
## Test.Fuel.Type.Description    10  11752782 1175278  97.101 < 2.2e-16 ***
## Residuals                21727 262975949   12104
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Response CO2.g.mi. :
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Test.Fuel.Type.Description    10    439  43.911  4.4822 2.379e-06 ***
## Residuals                21727 212856    9.797
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Response THC.g.mi. :
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Test.Fuel.Type.Description    10 21.274  2.12738  1948.2 < 2.2e-16 ***
## Residuals                21727 23.725  0.00109
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can see from the output that the p-value for all univariate ANOVAs are smaller than significance level 0.05, which indicates that fuel type has a statistically significant effect on CO₂, CO, and THC emissions.

Visualizing Group Means

Visualizing the Group means for each level of our independent variable vehicle type is helpful to get a better understanding of our results.

```
#visualize mean CO2 by vehicle type
plotmeans(nonelectric$CO2.g.mi. ~ nonelectric$Test.Fuel.Type.Description)
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```

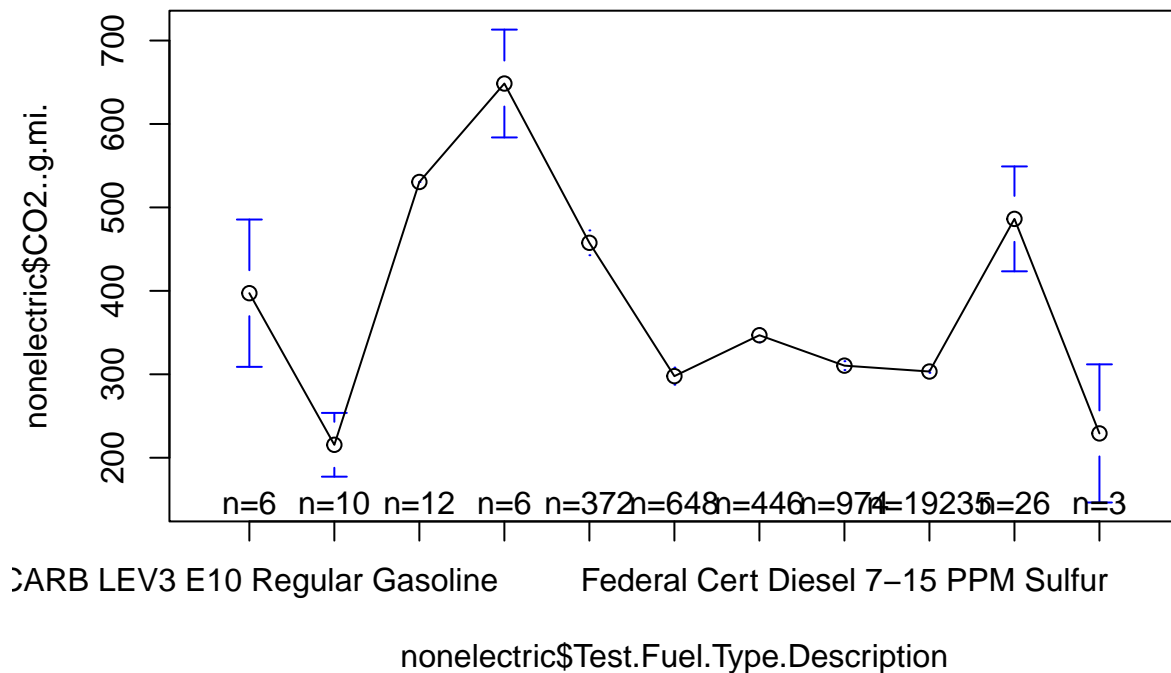
```
## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```



```
#visualize mean CO by vehicle type
plotmeans(nonelectric$CO..g.mi. ~ nonelectric$Test.Fuel.Type.Description)
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```

```

## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

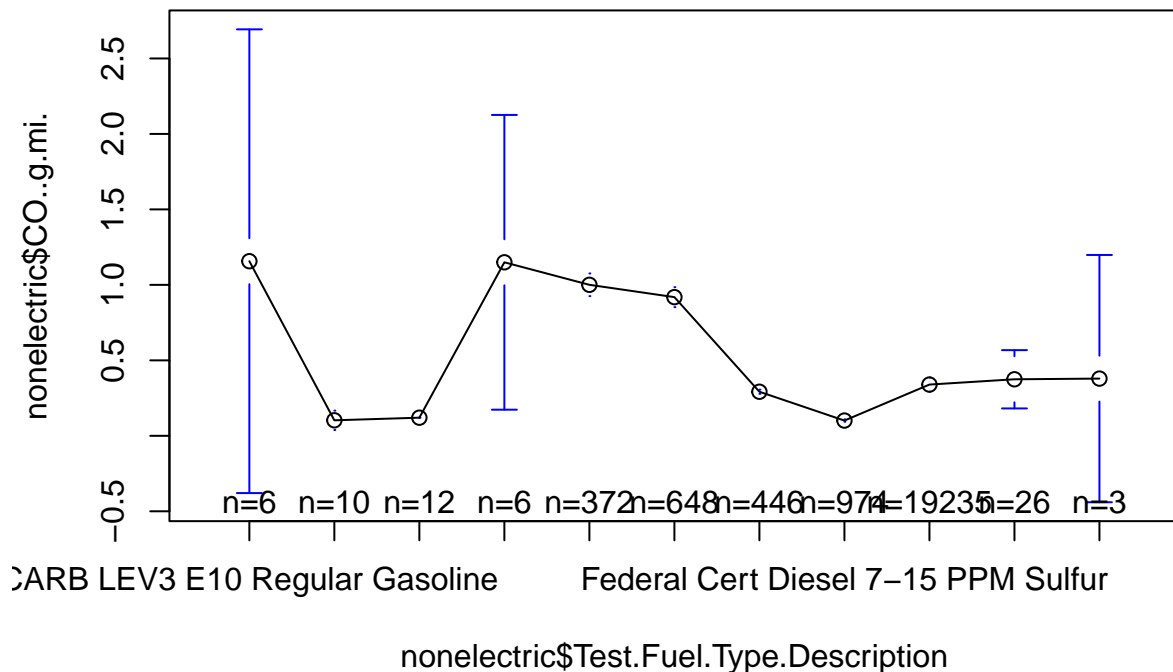
## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

```



```
#visualize mean THC by vehicle type
```

```
plotmeans(nonelectric$THC..g.mi. ~ nonelectric$Test.Fuel.Type.Description)
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-  
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-  
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-  
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-  
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-  
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-  
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-  
## length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
```

```

## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

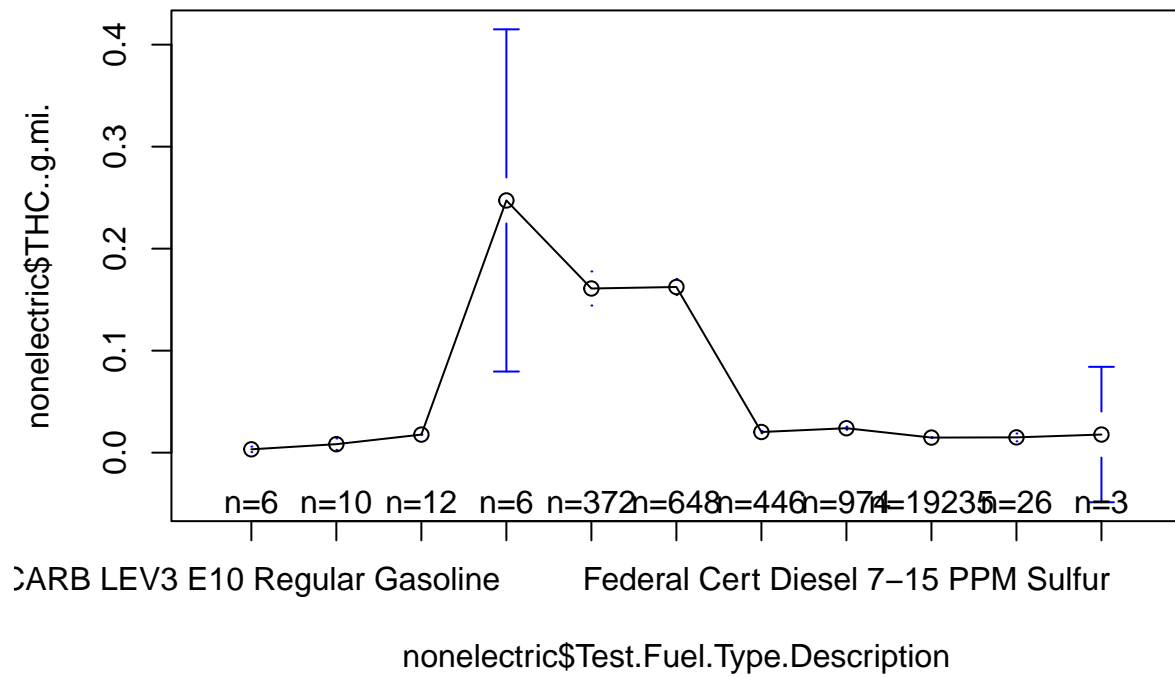
## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

```



From the above three plots, we can see that the mean CO₂, CO, and THC varies quite a bit by fuel types. This matches the results from our MANOVA, which indicates that there is statistically significant difference in three emissions based on fuel types.