Chi Squared Testing for Project

2022-12-04

Read in Data

```r
# Read in electric car data
e_df <- read.csv('../data/cardata_electric_clean.csv')

# Read in non-electric car data
ne_df <- read.csv('../data/cardata_nonelectric_clean.csv')

(nrow(e_df))
```

```
## [1] 878
```

```r
(nrow(ne_df))
```

```
## [1] 21738
```

```r
head(e_df)
```

```
##   X Model.Year Vehicle.Manufacturer.Name Veh.Mfr.Code Represented.Test.Veh.Make
## 1 1       2018                       BMW          BMX                       BMW
## 2 2       2018                       BMW          BMX                       BMW
## 3 3       2018                       BMW          BMX                       BMW
## 4 4       2018                       BMW          BMX                       BMW
## 5 5       2018                       BMW          BMX                       BMW
## 6 6       2018                       BMW          BMX                       BMW
##   Represented.Test.Veh.Model Test.Veh.Displacement..L. Vehicle.Type
## 1                       330e                         2          Car
## 2                       330e                         2          Car
## 3                       330e                         2          Car
## 4                       330e                         2          Car
## 5                       530e                         2          Car
## 6                       530e                         2          Car
##   Rated.Horsepower Tested.Transmission.Type.Code Tested.Transmission.Type
## 1              180                            SA           Semi-Automatic
## 2              180                            SA           Semi-Automatic
## 3              180                            SA           Semi-Automatic
## 4              180                            SA           Semi-Automatic
## 5              180                            SA           Semi-Automatic
## 6              180                            SA           Semi-Automatic
##   X..of.Gears Transmission.Lockup. Drive.System.Code Drive.System.Description
```

```
## 1          8                    Y                R      2-Wheel Drive, Rear
## 2          8                    Y                R      2-Wheel Drive, Rear
## 3          8                    Y                R      2-Wheel Drive, Rear
## 4          8                    Y                R      2-Wheel Drive, Rear
## 5          8                    Y                R      2-Wheel Drive, Rear
## 6          8                    Y                R      2-Wheel Drive, Rear
##   Equivalent.Test.Weight..lbs.. Axle.Ratio N.V.Ratio Test.Fuel.Type.Description
## 1                          4250       2.93      26.0                Electricity
## 2                          4250       2.93      26.0                Electricity
## 3                          4250       2.93      26.0                Electricity
## 4                          4250       2.93      26.0                Electricity
## 5                          4500       3.23      26.6                Electricity
## 6                          4500       3.23      26.6                Electricity
##   CO2..g.mi. RND_ADJ_FE Target.Coef.A..lbf. Target.Coef.B..lbf.mph.
## 1         NA        0.0                52.9                  -0.113
## 2         NA        0.0                52.9                  -0.113
## 3         NA        0.0                44.9                  -0.063
## 4         NA        0.0                44.9                  -0.063
## 5         NA      122.8                51.1                  -0.114
## 6         NA      122.8                51.1                  -0.114
##   Target.Coef.C..lbf.mph..2. Set.Coef.A..lbf. Set.Coef.B..lbf.mph.
## 1                    0.01826             21.2                0.056
## 2                    0.01826             21.2                0.056
## 3                    0.01831             13.0                0.128
## 4                    0.01831             13.0                0.128
## 5                    0.02015             12.1                0.305
## 6                    0.02015             12.1                0.305
##   Set.Coef.C..lbf.mph..2. Police...Emergency.Vehicle. Averaging.Method.Cd
## 1                 0.01632                           N                   N
## 2                 0.01632                           N                   N
## 3                 0.01611                           N                   N
## 4                 0.01611                           N                   N
## 5                 0.01540                           N                   N
## 6                 0.01540                           N                   N
##   Averging.Method.Desc
## 1         No averaging
## 2         No averaging
## 3         No averaging
## 4         No averaging
## 5         No averaging
## 6         No averaging
```

```
head(ne_df)
```

```
##   X Model.Year Vehicle.Manufacturer.Name Veh.Mfr.Code Represented.Test.Veh.Make
## 1 1       2018              aston martin          ASX              Aston Martin
## 2 2       2018              aston martin          ASX              Aston Martin
## 3 3       2018              aston martin          ASX              Aston Martin
## 4 4       2018              aston martin          ASX              Aston Martin
## 5 5       2018              aston martin          ASX              Aston Martin
## 6 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.Coef.A..lbf. Target.Coef.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.Coef.C..lbf.mph..2. Set.Coef.A..lbf. Set.Coef.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.Coef.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
```

```
# 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")
```

EDA

To formulate our hypotheses, we first perform EDA on the dataset

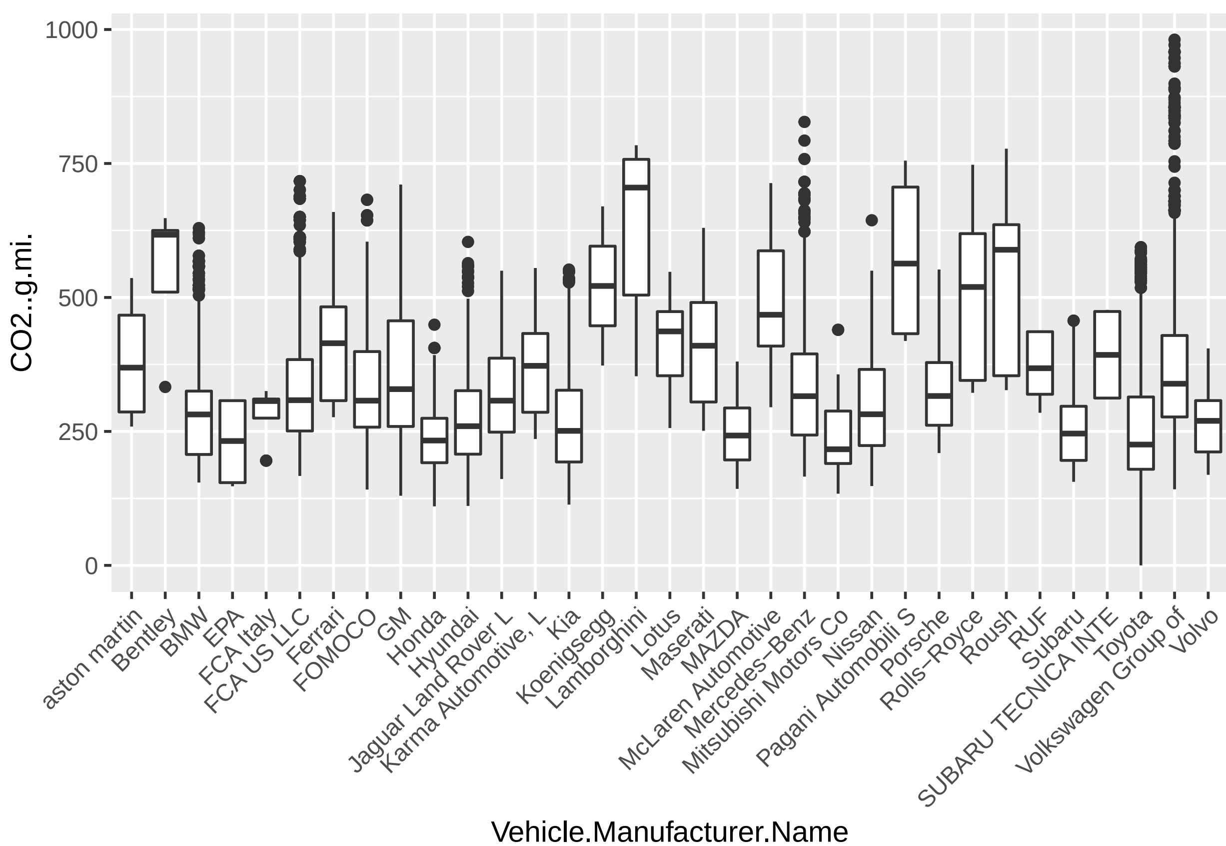
```
#Vehicle.Manufacturer.Name
#CO2..g.mi.
library(ggplot2)
library(tidyr)

# Drop NAs for the emissions column
ne_df <- ne_df %>% drop_na(CO2..g.mi.)

ne_df$emission_cat[ne_df$CO2..g.mi. < 250] <- "low"
ne_df$emission_cat[ne_df$CO2..g.mi. >= 250 & ne_df$CO2..g.mi. < 500] <- "medium"
ne_df$emission_cat[ne_df$CO2..g.mi. >= 500] <- "high"

head(ne_df)
```

```
##   X Model.Year Vehicle.Manufacturer.Name Veh.Mfr.Code Represented.Test.Veh.Make
## 1 1       2018              aston martin          ASX              Aston Martin
## 2 2       2018              aston martin          ASX              Aston Martin
## 3 3       2018              aston martin          ASX              Aston Martin
## 4 4       2018              aston martin          ASX              Aston Martin
## 5 5       2018              aston martin          ASX              Aston Martin
## 6 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
```

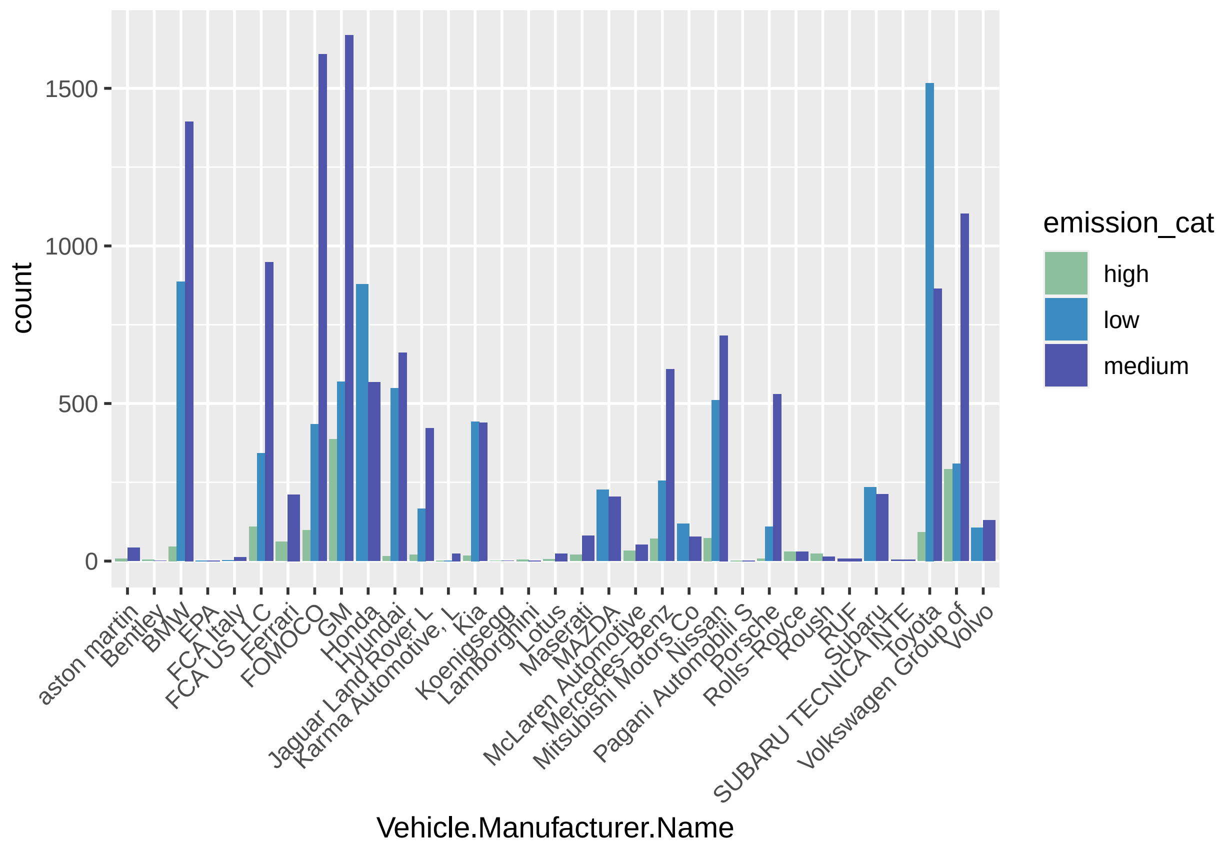
```
## 3                  DB11 V8                        4.0         Car
## 4                  DB11 V8                        4.0         Car
## 5                Rapide S                        6.0         Car
## 6                Rapide S                        6.0         Car
##   Rated.Horsepower X..of.Cylinders.and.Rotors 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.Coef.A..lbf. Target.Coef.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.Coef.C..lbf.mph..2. Set.Coef.A..lbf. Set.Coef.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.Coef.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
##   emission_cat
## 1       medium
## 2       medium
## 3       medium
## 4       medium
## 5         high
## 6       medium
```

```
# Make bar plot of transmission type and CO2 emissions
ggplot(data=ne_df, aes(x=Vehicle.Manufacturer.Name,y=CO2..g.mi.))+geom_boxplot()+ scale_x_discrete(guide
```

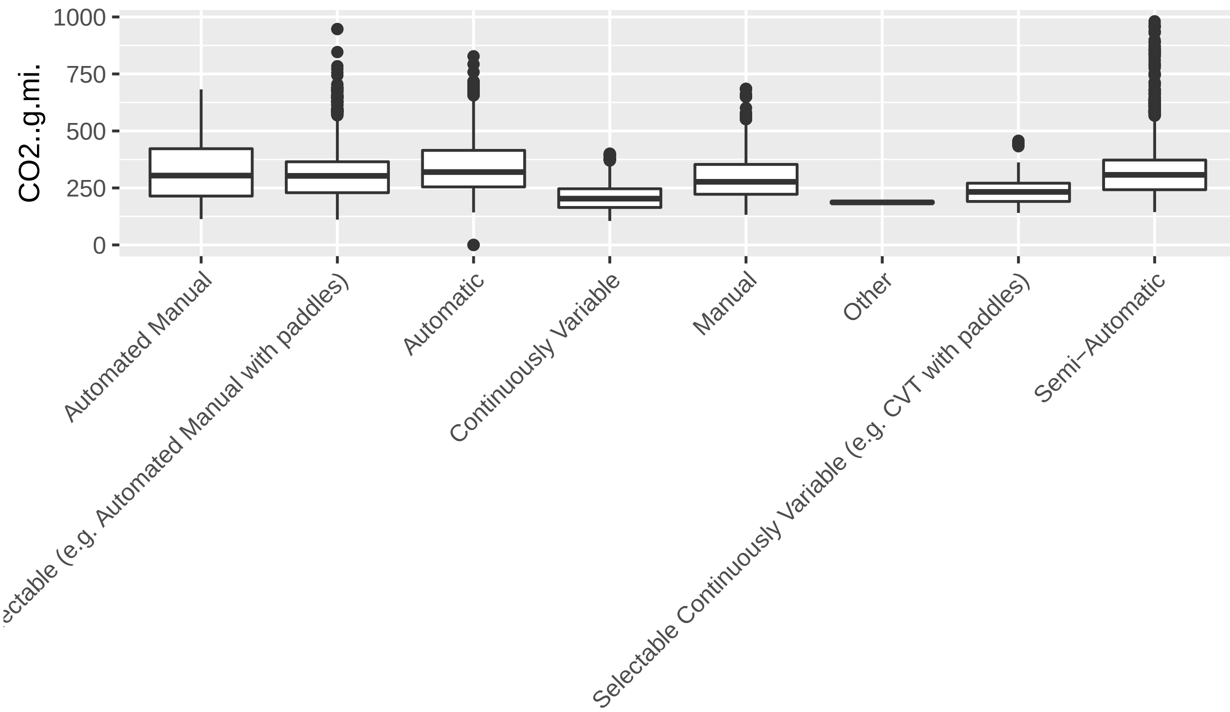


```r
# make barplot of emissions categories

ggplot(ne_df, aes(x=Vehicle.Manufacturer.Name, fill=emission_cat)) + geom_bar(position="dodge")+ scale_
```



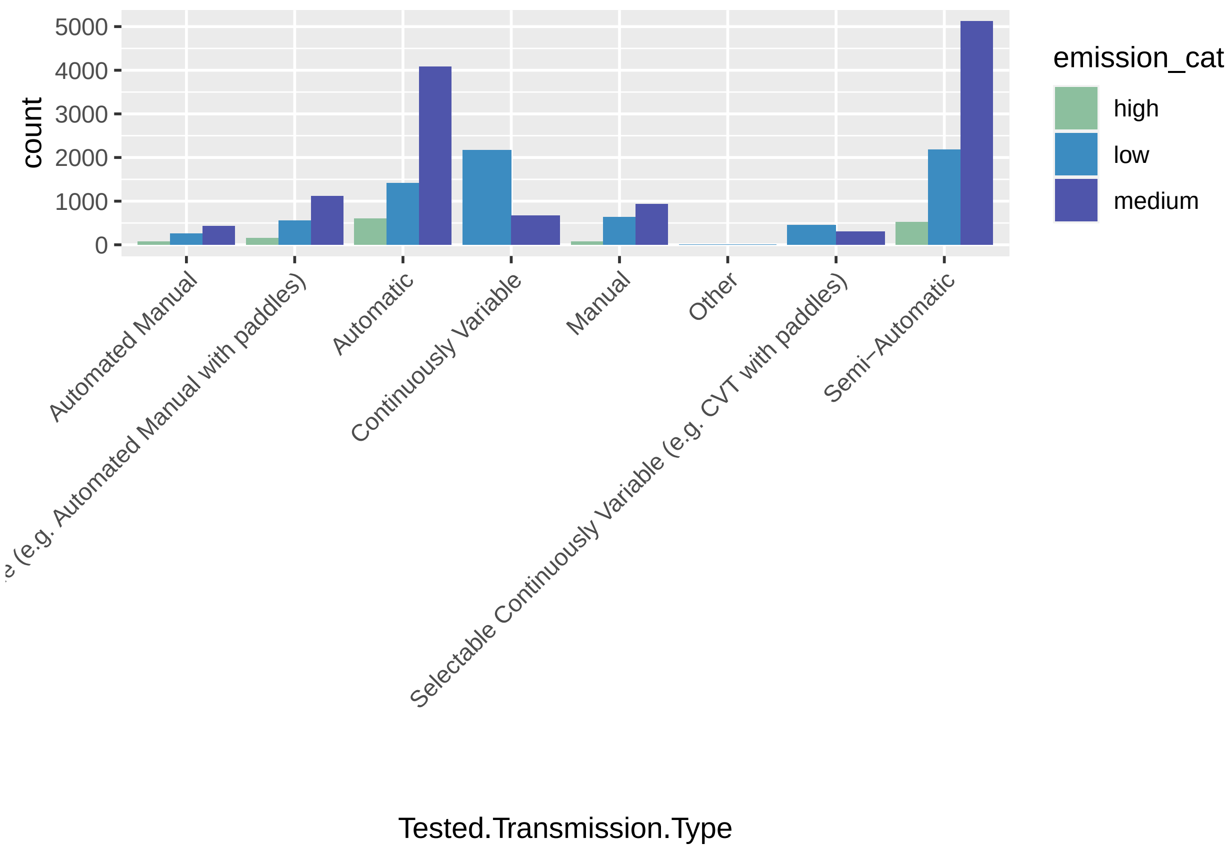
We definitely see some car manufacturers have higher average fuel emissions. For instance, Lamborghini, Bentley, and Rolls-Royce have higher emissions likely due to them being luxury brands. Honda and Mitsubishi, on the other hand, are more affordable brands and have lower average emissions.

```r
# Make bar plot of transmission type and CO2 emissions
ggplot(data=ne_df, aes(x=Tested.Transmission.Type,y=CO2..g.mi., fill=CO2..g.mi.))+geom_boxplot()+ scale_
```



Tested.Transmission.Type

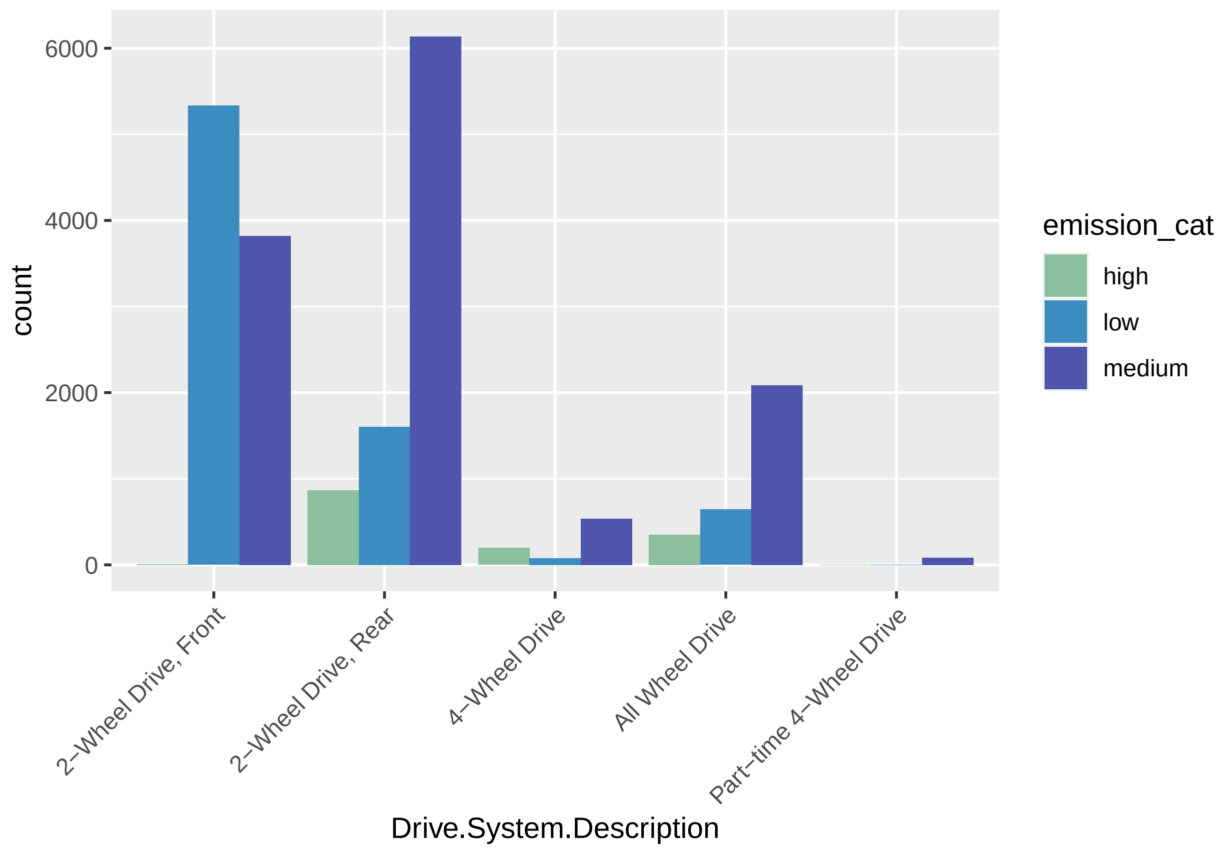
```
ggplot(ne_df, aes(x=Tested.Transmission.Type, fill=emission_cat)) + geom_bar(position="dodge")+ scale_x_
```



Interestingly, it appears like the high emissions cars are mostly within the automatic and semi-automatic categories. Manual tends to have mostly low and medium with very few high emissions cars.

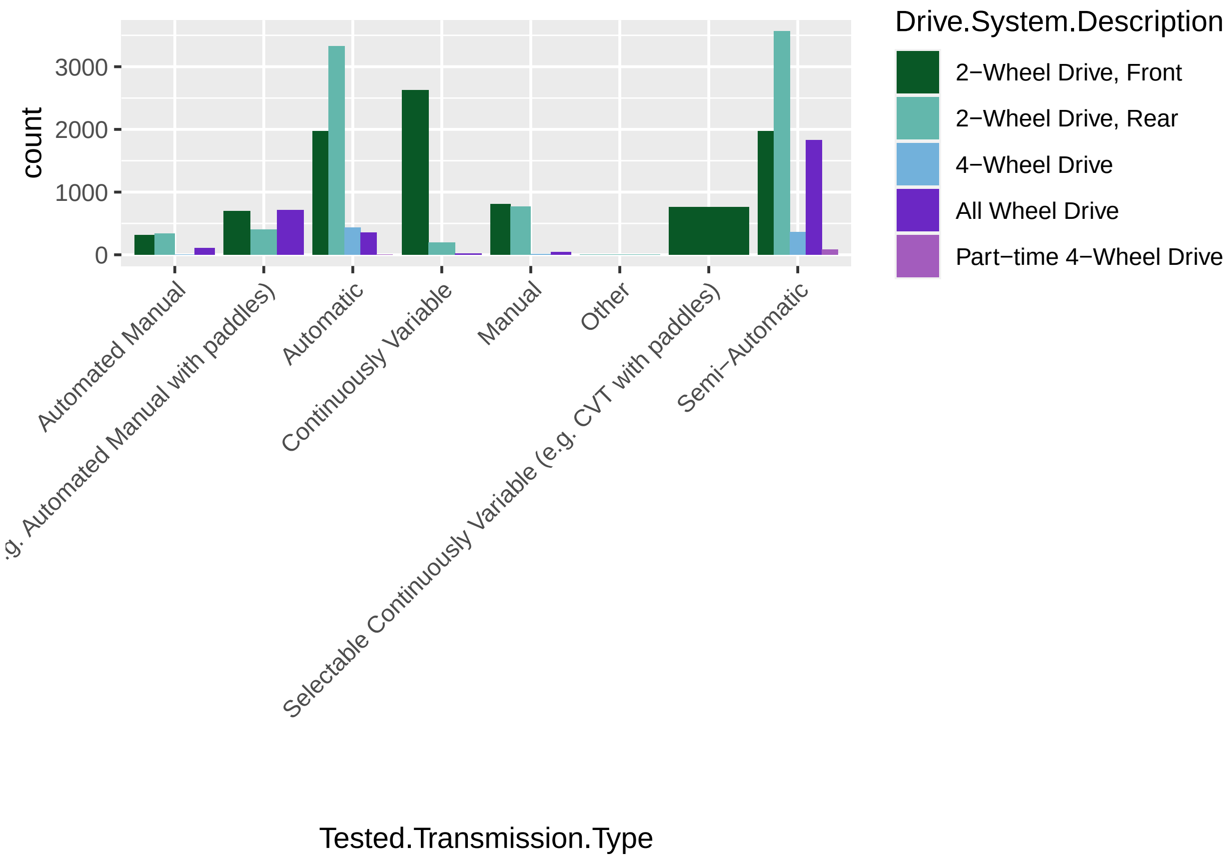
Drive system vs Emission category. Looks like 2 wheel drive front does not have high emissions compared to the 4 wheel drive and all wheel drive. Perhaps fuel emissions increase the more the wheel drive increases.

```
ggplot(ne_df, aes(x=Drive.System.Description, fill=emission_cat)) + geom_bar(position="dodge")+ scale_x
```



Next, check observe the transmission type vs the drive system type

```
ggplot(ne_df, aes(fill=Drive.System.Description, x=Tested.Transmission.Type)) + geom_bar(position="dodg
```



Define Hypotheses

Question 1: Is there a relationship between car brand and fuel emission level?

Null hypothesis: Car Brand and fuel emissions level are independent. The fuel emissions level does not depend on the car brand

Alternative Hypothesis: Car Brand and fuel emissions level are dependent. The fuel emissions level does depend on the car brand

Question 2: Is there a relationship between transmission type and fuel emission level?

Null hypothesis: Transmission type and fuel emissions level are independent. The fuel emissions level does not depend on the transmission type

Alternative Hypothesis: Transmission type and fuel emissions level are dependent. The fuel emissions level does depend on the Transmission type

Question 3: Is there a relationship between transmission type and the drive system type?

Null hypothesis: Transmission type and drive system type are independent. The transmission type does not depend on the drive system type

Alternative Hypothesis: Transmission type and drive system type are dependent. The transmission type does depend on the drive system type

Question 4: Is there a relationship between drive system and the fuel emission level type?

Null hypothesis: Drive system and fuel emissions level are independent. The fuel emissions level does not depend on the drive system

Alternative Hypothesis: Drive System and fuel emissions level are dependent. The fuel emissions level does depend on the drive system

Hypothesis Testing

```
head(ne_df)
```

```
##   X Model.Year Vehicle.Manufacturer.Name Veh.Mfr.Code Represented.Test.Veh.Make
## 1 1       2018              aston martin          ASX              Aston Martin
## 2 2       2018              aston martin          ASX              Aston Martin
## 3 3       2018              aston martin          ASX              Aston Martin
## 4 4       2018              aston martin          ASX              Aston Martin
## 5 5       2018              aston martin          ASX              Aston Martin
## 6 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.Coef.A..lbf. Target.Coef.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.Coef.C..lbf.mph..2. Set.Coef.A..lbf. Set.Coef.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.Coef.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
##   emission_cat
## 1       medium
## 2       medium
## 3       medium
## 4       medium
## 5         high
## 6       medium
```

Question 1

Question 1: Is there a relationship between car brand and fuel emission level?

Null hypothesis: Car Brand and fuel emissions level are independent. The fuel emissions level does not depend on the car brand

Alternative Hypothesis: Car Brand and fuel emissions level are dependent. The fuel emissions level does depend on the car brand

```
# Make contingency table
cont <- table(ne_df$Vehicle.Manufacturer.Name, ne_df$emission_cat)
cont
```

```
##
##                         high  low medium
##   aston martin             8    0     42
##   Bentley                  4    0      1
##   BMW                     46  886   1395
##   EPA                      0    2      2
##   FCA Italy                0    3     12
##   FCA US LLC             109  342    948
##   Ferrari                 61    0    211
##   FOMOCO                  98  434   1608
##   GM                     387  569   1668
##   Honda                    0  878    567
##   Hyundai                 15  548    661
##   Jaguar Land Rover L     20  167    421
##   Karma Automotive, L      2    2     24
##   Kia                     17  443    439
##   Koenigsegg               1    0      1
##   Lamborghini              5    0      2
##   Lotus                    6    0     24
##   Maserati                20    0     80
##   MAZDA                    0  226    204
##   McLaren Automotive      33    0     52
##   Mercedes-Benz           71  255    609
##   Mitsubishi Motors Co     0  118     77
##   Nissan                  73  510    716
##   Pagani Automobili S      2    0      2
##   Porsche                  7  109    529
##   Rolls-Royce             30    0     30
##   Roush                   23    0     14
##   RUF                      0    0      8
##   Subaru                   0  234    212
##   SUBARU TECNICA INTE      0    0      4
##   Toyota                  91 1517    864
##   Volkswagen Group of    292  309   1102
##   Volvo                    0  106    130
```

```r
chisq.test(cont)
```

```
## Warning in chisq.test(cont): Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  cont
## X-squared = 4126, df = 64, p-value < 2.2e-16
```

Repeat test with the Yates correction

```r
chisq.test(cont, correct = TRUE)
```

```
## Warning in chisq.test(cont, correct = TRUE): Chi-squared approximation may be
## incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  cont
## X-squared = 4126, df = 64, p-value < 2.2e-16
```

Since the Yates correction was not enough, we can switch over to Fisher's Exact Test

```r
fisher.test(cont, simulate.p.value=TRUE)
```

```
##
##  Fisher's Exact Test for Count Data with simulated p-value (based on
##  2000 replicates)
##
## data:  cont
## p-value = 0.0004998
## alternative hypothesis: two.sided
```

Question 2

Question 2: Is there a relationship between transmission type and fuel emission level?

Null hypothesis: Transmission type and fuel emissions level are independent. The fuel emissions level does not depend on the transmission type

Alternative Hypothesis: Transmission type and fuel emissions level are dependent. The fuel emissions level does depend on the Transmission type

```r
# Make contingency table
cont <- table(ne_df$Tested.Transmission.Type, ne_df$emission_cat)
cont
```

```
##
##                                                                  high  low
##    Automated Manual                                                79  255
##    Automated Manual- Selectable (e.g. Automated Manual with paddles)  151  553
##    Automatic                                                      603 1411
##    Continuously Variable                                            0 2168
##    Manual                                                          72  634
##    Other                                                            0    4
##    Selectable Continuously Variable (e.g. CVT with paddles)        0  454
##    Semi-Automatic                                                 516 2179
##
##                                                                  medium
##    Automated Manual                                                 430
##    Automated Manual- Selectable (e.g. Automated Manual with paddles)   1115
##    Automatic                                                       4081
##    Continuously Variable                                            671
##    Manual                                                           932
##    Other                                                              0
##    Selectable Continuously Variable (e.g. CVT with paddles)         308
##    Semi-Automatic                                                  5122
```

```
chisq.test(cont)
```

```
## Warning in chisq.test(cont): Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  cont
## X-squared = 3047.7, df = 14, p-value < 2.2e-16
```

Repeat test with the Yates correction

```
chisq.test(cont, correct = TRUE)
```

```
## Warning in chisq.test(cont, correct = TRUE): Chi-squared approximation may be
## incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  cont
## X-squared = 3047.7, df = 14, p-value < 2.2e-16
```

Since the Yates correction was not enough, we can switch over to Fisher's Exact Test

```
fisher.test(cont, simulate.p.value=TRUE)
```

```
##
##  Fisher's Exact Test for Count Data with simulated p-value (based on
##  2000 replicates)
##
## data:  cont
## p-value = 0.0004998
## alternative hypothesis: two.sided
```

Question 3

Question 3: Is there a relationship between transmission type and the drive system type?

Null hypothesis: Transmission type and drive system type are independent. The transmission type does not depend on the drive system type

Alternative Hypothesis: Transmission type and drive system type are dependent. The transmission type does depend on the drive system type

```
# Make contingency table
cont <- table(ne_df$Tested.Transmission.Type, ne_df$Drive.System.Description)
cont
```

```
##
##                                                                 2-Wheel Drive, Front
##   Automated Manual                                                               315
##   Automated Manual- Selectable (e.g. Automated Manual with paddles)              700
##   Automatic                                                                     1974
##   Continuously Variable                                                         2623
##   Manual                                                                         809
##   Other                                                                            0
##   Selectable Continuously Variable (e.g. CVT with paddles)                       762
##   Semi-Automatic                                                                1976
##
##                                                                 2-Wheel Drive, Rear
##   Automated Manual                                                              336
##   Automated Manual- Selectable (e.g. Automated Manual with paddles)             404
##   Automatic                                                                    3330
##   Continuously Variable                                                         196
##   Manual                                                                        770
##   Other                                                                           4
##   Selectable Continuously Variable (e.g. CVT with paddles)                        0
##   Semi-Automatic                                                               3565
##
##                                                                 4-Wheel Drive
##   Automated Manual                                                          2
##   Automated Manual- Selectable (e.g. Automated Manual with paddles)         0
##   Automatic                                                               434
##   Continuously Variable                                                     0
##   Manual                                                                   12
##   Other                                                                     0
##   Selectable Continuously Variable (e.g. CVT with paddles)                  0
##   Semi-Automatic                                                          363
##
##                                                                 All Wheel Drive
##   Automated Manual                                                            111
##   Automated Manual- Selectable (e.g. Automated Manual with paddles)           715
##   Automatic                                                                   355
##   Continuously Variable                                                        20
##   Manual                                                                       47
##   Other                                                                         0
##   Selectable Continuously Variable (e.g. CVT with paddles)                      0
##   Semi-Automatic                                                             1830
##
##                                                                 Part-time 4-Wheel Drive
##   Automated Manual                                                                    0
##   Automated Manual- Selectable (e.g. Automated Manual with paddles)                   0
##   Automatic                                                                           2
##   Continuously Variable                                                               0
##   Manual                                                                              0
##   Other                                                                               0
##   Selectable Continuously Variable (e.g. CVT with paddles)                            0
##   Semi-Automatic                                                                     83
```

```r
chisq.test(cont)
```

```
## Warning in chisq.test(cont): Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  cont
## X-squared = 7472.1, df = 28, p-value < 2.2e-16
```

Repeat test with the Yates correction

```
chisq.test(cont, correct = TRUE)
```

```
## Warning in chisq.test(cont, correct = TRUE): Chi-squared approximation may be
## incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  cont
## X-squared = 7472.1, df = 28, p-value < 2.2e-16
```

Since the Yates correction was not enough, we can switch over to Fisher's Exact Test

```
fisher.test(cont, simulate.p.value=TRUE)
```

```
##
##  Fisher's Exact Test for Count Data with simulated p-value (based on
##  2000 replicates)
##
## data:  cont
## p-value = 0.0004998
## alternative hypothesis: two.sided
```

Question 4: Is there a relationship between drive system and the fuel emission level type?

Null hypothesis: Drive system and fuel emissions level are independent. The fuel emissions level does not depend on the drive system

Alternative Hypothesis: Drive System and fuel emissions level are dependent. The fuel emissions level does depend on the drive system

```
# Make contingency table
cont <- table(ne_df$Drive.System.Description, ne_df$emission_cat)
cont
```

```
##
##                              high  low medium
##   2-Wheel Drive, Front          7 5332   3820
##   2-Wheel Drive, Rear         866 1603   6136
##   4-Wheel Drive               197   78    536
##   All Wheel Drive             350  643   2085
##   Part-time 4-Wheel Drive       1    2     82
```

Repeat test with the Yates correction

```r
chisq.test(cont, correct = TRUE)
```

```
##
##  Pearson's Chi-squared test
##
## data:  cont
## X-squared = 4416.8, df = 8, p-value < 2.2e-16
```

Since the Yates correction was not enough, we can switch over to Fisher's Exact Test

```r
fisher.test(cont, simulate.p.value=TRUE)
```

```
##
##  Fisher's Exact Test for Count Data with simulated p-value (based on
##  2000 replicates)
##
## data:  cont
## p-value = 0.0004998
## alternative hypothesis: two.sided
```