Crime_Dataset_EDA

the_principal_components

2022-07-15

```
mpg_has_luggage <-dplyr::filter(mpg, lv2 != 0 | lv4 != 0)
mpg_has_luggage_and_passenger <-dplyr::filter(mpg, pv2 != 0 | pv4 != 0)
mpg_with_classifier <- mpg_has_luggage_and_passenger %>% mutate( isTwoDoor = (pv2 != 0))
two_door <- dplyr::filter(mpg_with_classifier, isTwoDoor == TRUE) %>% mutate( total_volume = pv2 + lv2)
four_door <- dplyr::filter(mpg_with_classifier, isTwoDoor == FALSE) %>% mutate( total_volume = pv4 + lv4
final_mpg_dataset = rbind(two_door, four_door)
rm(list=c("two_door", "four_door", "mpg_has_luggage", "mpg_has_luggage_and_passenger", "mpg_with_classi
```

```
correlation_table(data=final_mpg_dataset, target="city08")
```

```
## Warning in cor(data, use = "complete.obs"): the standard deviation is zero
```

```
##           Variable city08
## 1           city08   1.00
## 2            UCity   1.00
## 3           comb08   0.99
## 4        highway08   0.92
## 5         UHighway   0.92
## 6     youSaveSpend   0.85
## 7          city08U   0.48
## 8          comb08U   0.45
## 9          feScore   0.44
## 10        ghgScore   0.44
## 11      highway08U   0.41
## 12             hpv   0.21
## 13              id   0.21
## 14        cityA08U   0.19
## 15             hlv   0.19
## 16            year   0.19
## 17        phevCity   0.19
## 18        phevComb   0.19
## 19         cityA08   0.18
## 20        combA08U   0.18
## 21          UCityA   0.18
## 22         phevHwy   0.18
## 23          cityUF   0.17
## 24         combA08   0.17
## 25      combinedUF   0.17
## 26     highwayA08U   0.17
## 27      highwayA08   0.16
## 28       highwayUF   0.16
## 29             pv4   0.15
## 30       charge240   0.14
```

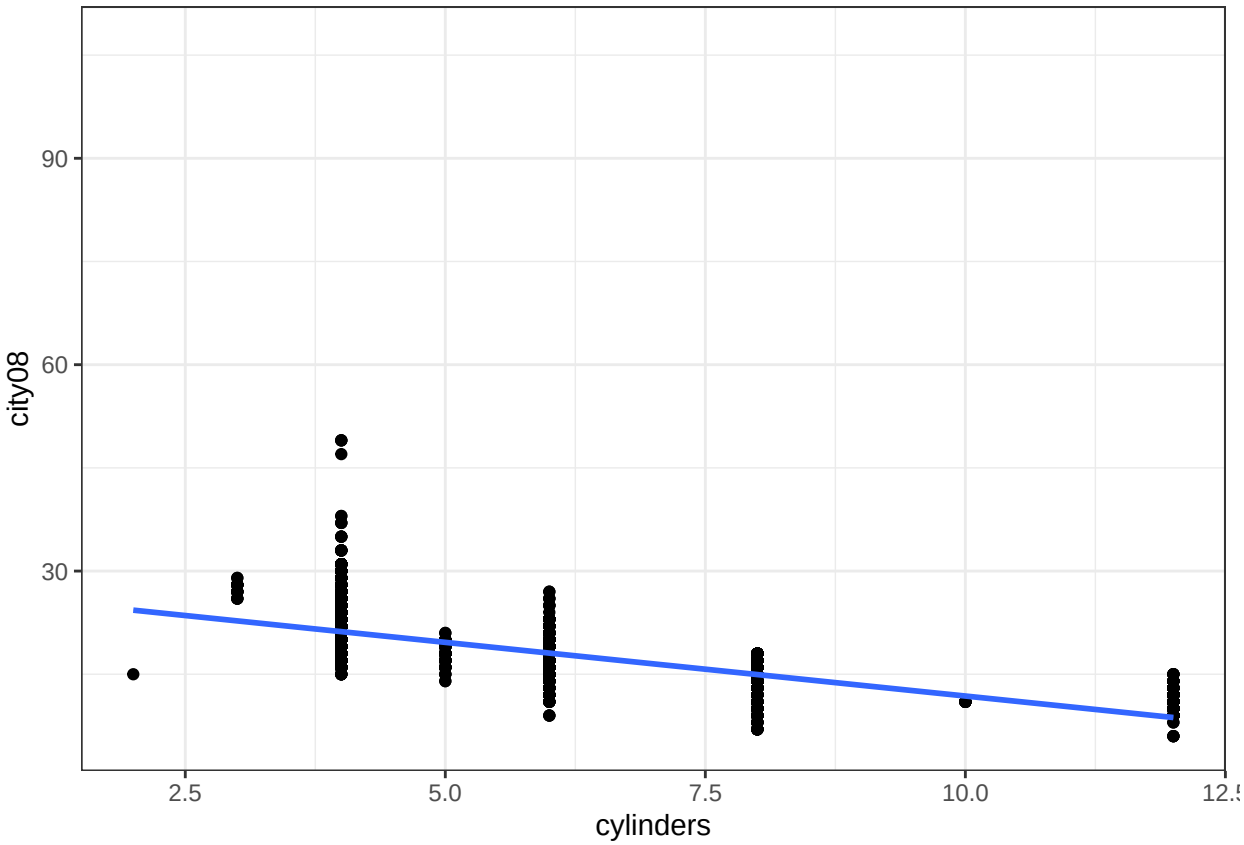
```
## 31      rangeCityA   0.14
## 32       rangeHwyA   0.14
## 33        highwayE   0.13
## 34           cityE   0.12
## 35           combE   0.12
## 36             co2   0.09
## 37             lv4   0.07
## 38      combinedCD   0.04
## 39       highwayCD   0.04
## 40      barrelsA08   0.03
## 41          cityCD   0.03
## 42       UHighwayA   0.03
## 43           engId   0.00
## 44       ghgScoreA   0.00
## 45            co2A  -0.02
## 46     fuelCostA08  -0.03
## 47 co2TailpipeAGpm  -0.04
## 48    total_volume  -0.04
## 49             lv2  -0.12
## 50             pv2  -0.14
## 51       cylinders  -0.66
## 52           displ  -0.68
## 53      fuelCost08  -0.84
## 54       barrels08  -0.86
## 55  co2TailpipeGpm  -0.87
## 56           range     NA
## 57       rangeCity     NA
## 58        rangeHwy     NA
## 59      charge240b     NA
```

```
ggplot(dplyr::filter(final_mpg_dataset, isTwoDoor == TRUE), aes(x=cylinders, y=city08)) +
  geom_point()+
  geom_smooth(method=lm)
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: Removed 1 rows containing non-finite values (stat_smooth).
```

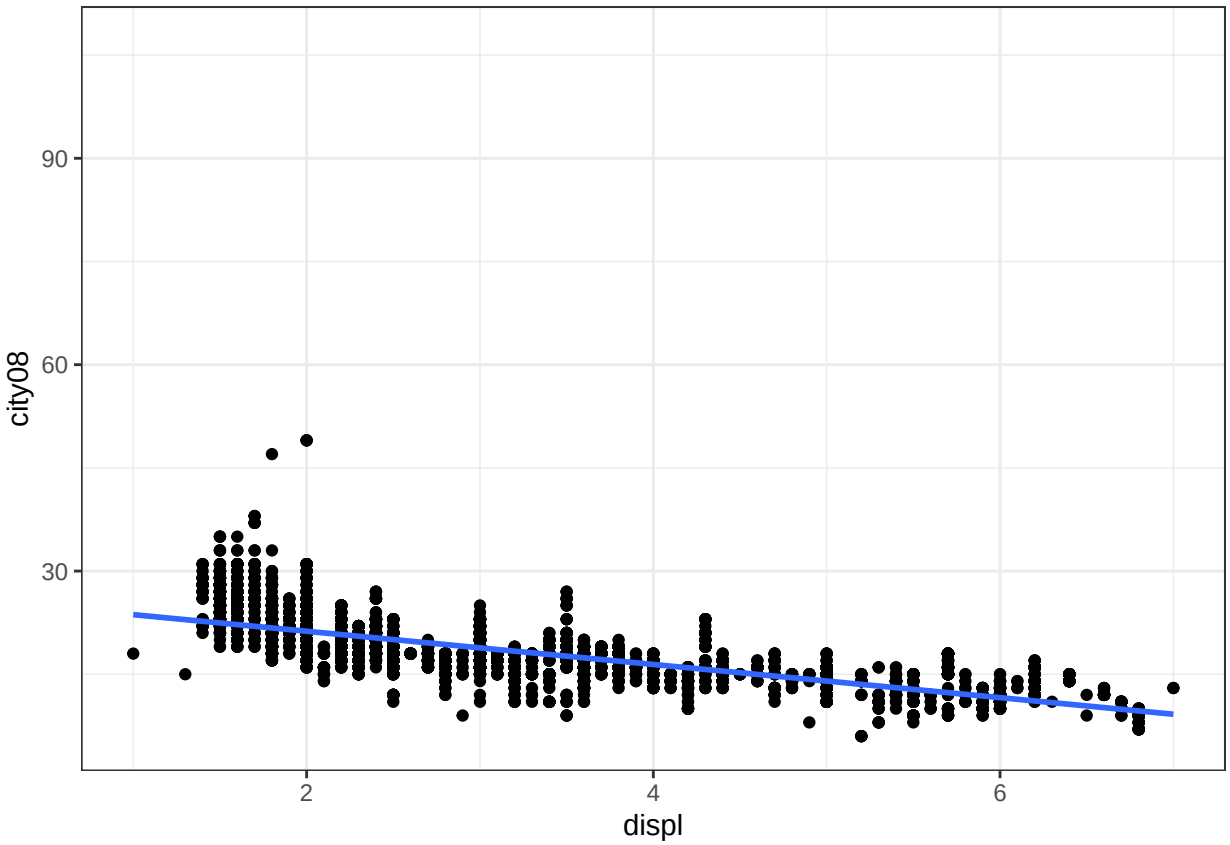
```
## Warning: Removed 1 rows containing missing values (geom_point).
```




```
ggplot(dplyr::filter(final_mpg_dataset, isTwoDoor == TRUE), aes(x=displ, y=city08)) +
  geom_point()+
  geom_smooth(method=lm)
```

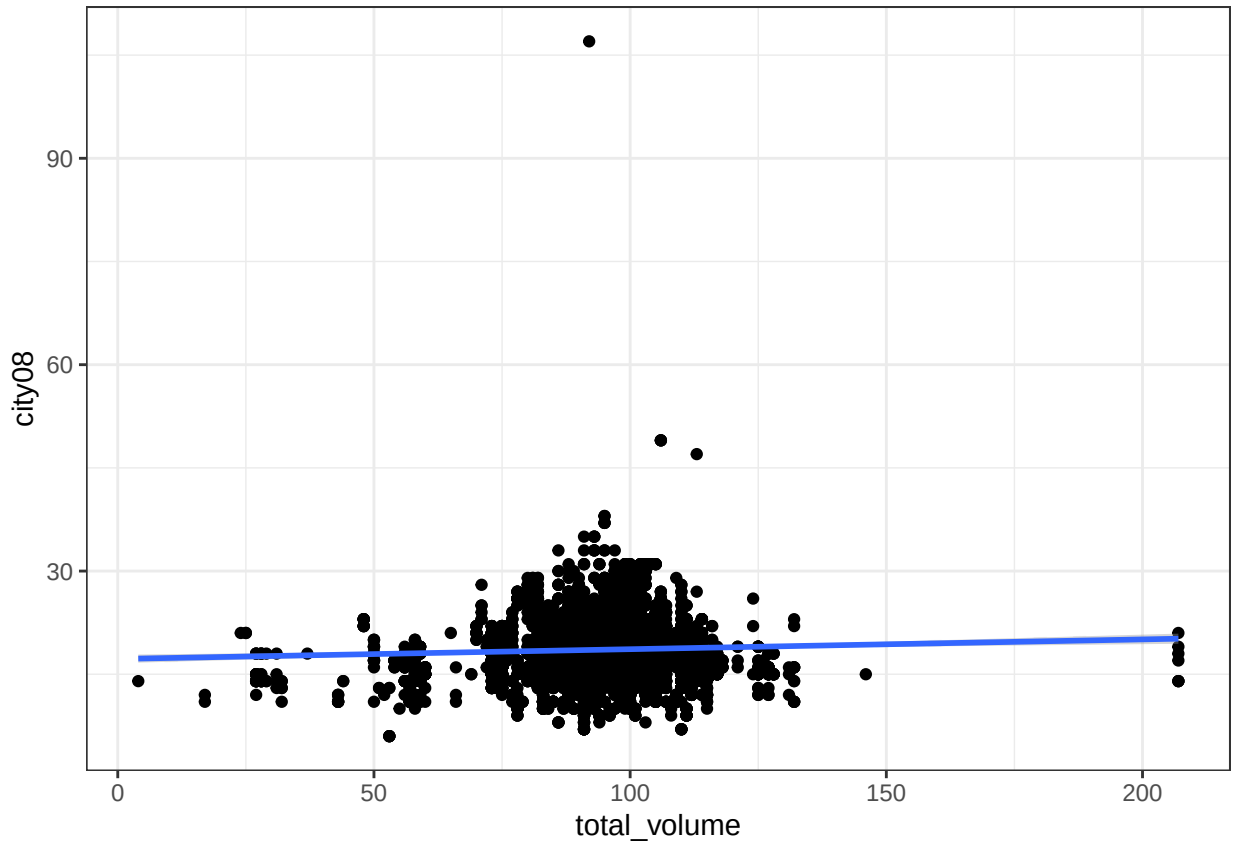
```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: Removed 1 rows containing non-finite values (stat_smooth).
## Removed 1 rows containing missing values (geom_point).
```



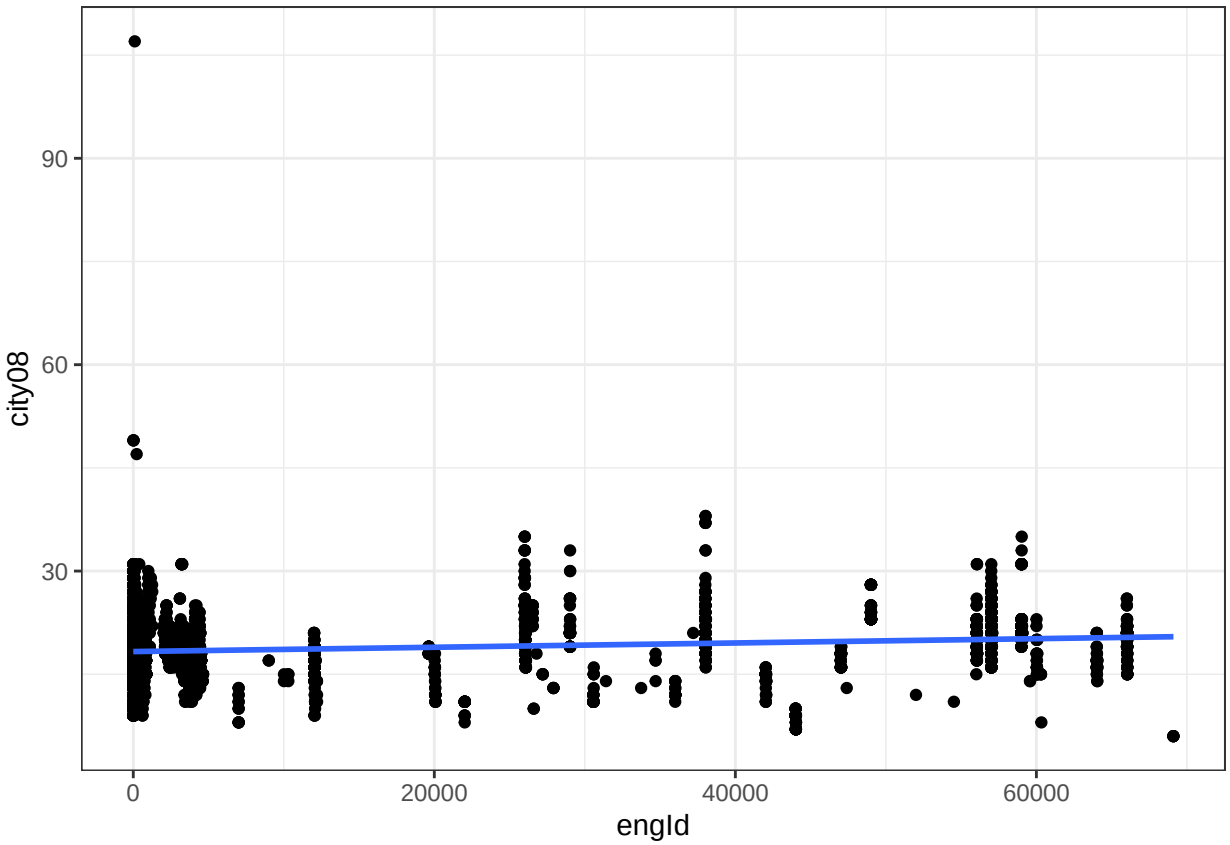

```
ggplot(dplyr::filter(final_mpg_dataset, isTwoDoor == TRUE), aes(x=total_volume , y=city08)) +
  geom_point()+
  geom_smooth(method=lm)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



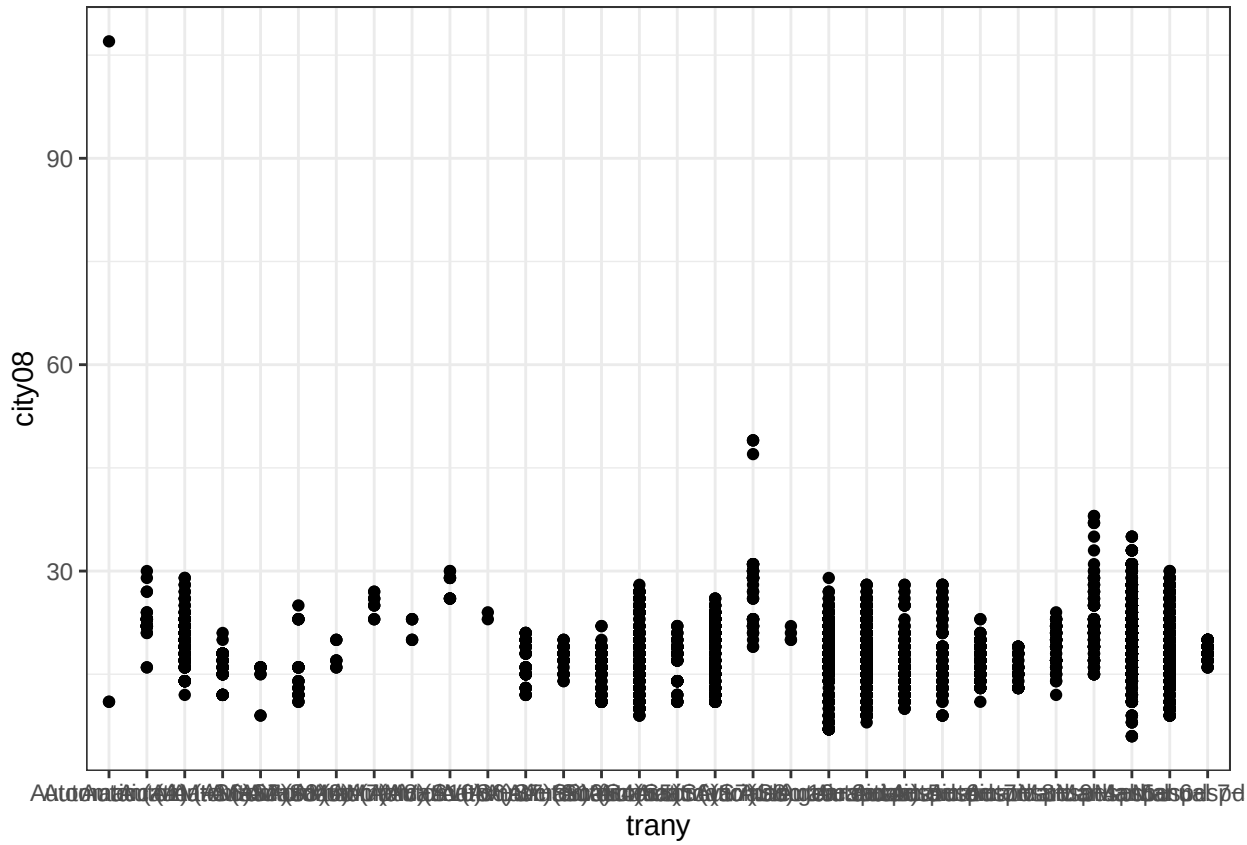

```
ggplot(dplyr::filter(final_mpg_dataset, isTwoDoor == TRUE), aes(x=engId, y=city08)) +
  geom_point()+
  geom_smooth(method=lm)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



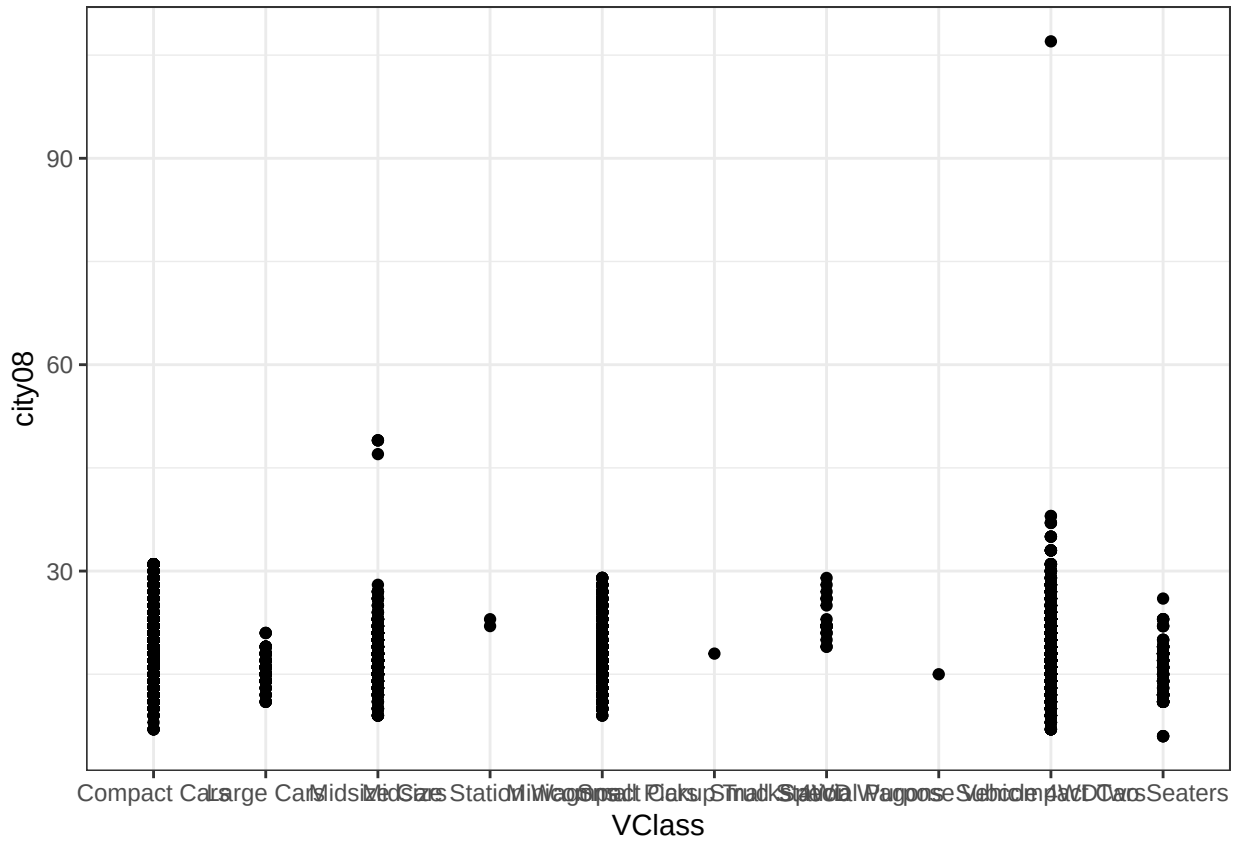

```
ggplot(dplyr::filter(final_mpg_dataset, isTwoDoor == TRUE), aes(x=trany, y=city08)) +
  geom_point()+
  geom_smooth(method=lm)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



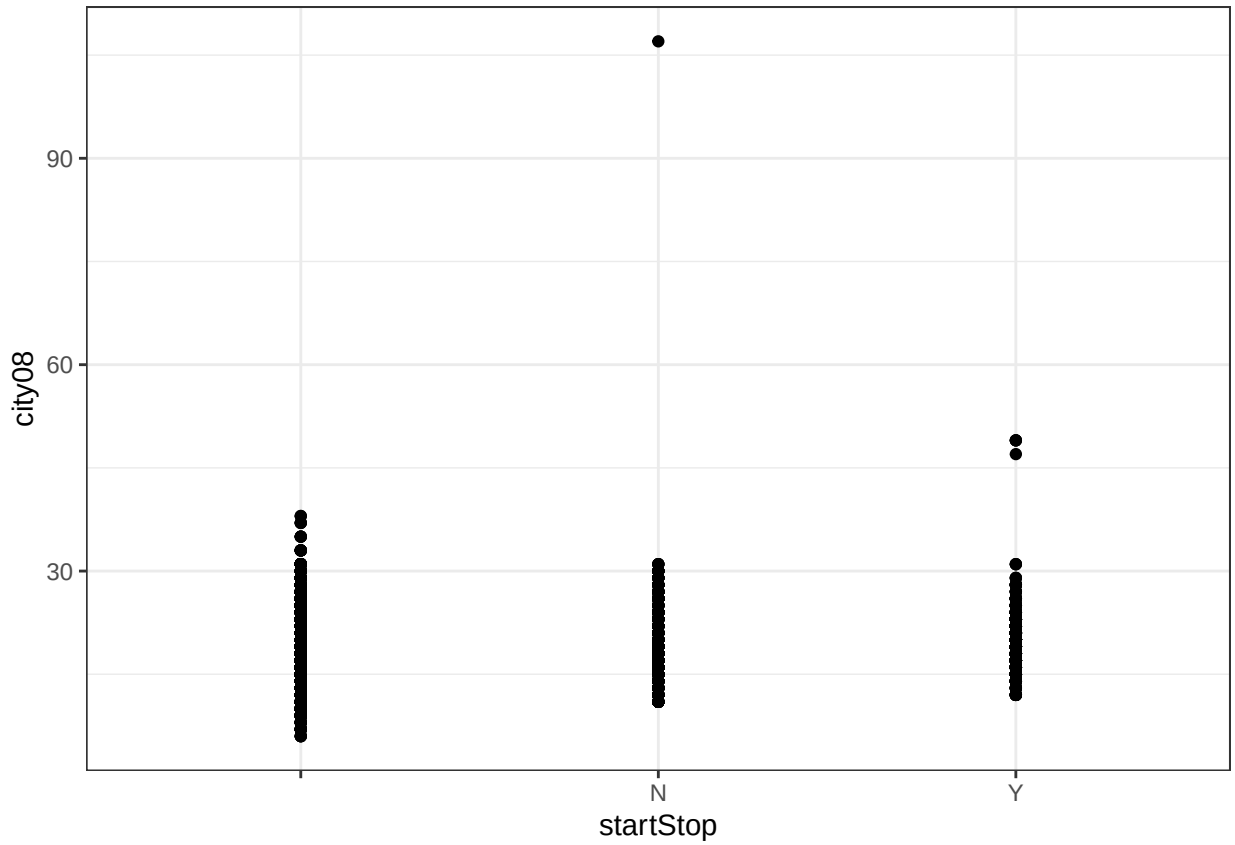

```
ggplot(dplyr::filter(final_mpg_dataset, isTwoDoor == TRUE), aes(x=VClass, y=city08)) +
  geom_point()+
  geom_smooth(method=lm)
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
ggplot(dplyr::filter(final_mpg_dataset, isTwoDoor == TRUE), aes(x=startStop , y=city08)) +
  geom_point()+
  geom_smooth(method=lm)
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
mpg_fit_01 <- final_mpg_dataset %>% lm(city08 ~ displ + total_volume + VClass, data = .)
mpg_fit_01_se <- mpg_fit_01 %>% vcovHC(type = "HC1") %>% diag() %>% sqrt()
stargazer(mpg_fit_01, type="text",se = list(mpg_fit_01_se))
```

```
##
## ===============================================================
##                                         Dependent variable:
##                                     ---------------------------
##                                               city08
## ---------------------------------------------------------------
## displ                                        -2.920***
##                                               (0.025)
##
## total_volume                                 0.021***
##                                               (0.004)
##
## VClassLarge Cars                             0.399***
##                                               (0.118)
##
## VClassMidsize-Large Station Wagons           -2.194***
##                                               (0.186)
##
## VClassMidsize Cars                           0.580***
##                                               (0.091)
##
## VClassMidsize Station Wagons                 -1.652***
```

```
##                                                    (0.197)
##
## VClassMinicompact Cars                              0.127
##                                                    (0.128)
##
## VClassSmall Pickup Trucks 4WD                     -5.646***
##                                                    (0.239)
##
## VClassSmall Sport Utility Vehicle 2WD              0.652***
##                                                    (0.171)
##
## VClassSmall Sport Utility Vehicle 4WD               0.188
##                                                    (0.250)
##
## VClassSmall Station Wagons                          0.070
##                                                    (0.129)
##
## VClassSpecial Purpose Vehicle 2WD                 -2.798***
##                                                    (0.080)
##
## VClassSpecial Purpose Vehicle 4WD                 -3.241***
##                                                    (0.563)
##
## VClassSport Utility Vehicle - 2WD                 -2.822***
##                                                    (0.171)
##
## VClassSport Utility Vehicle - 4WD                 -2.330***
##                                                    (0.186)
##
## VClassStandard Sport Utility Vehicle 2WD          -1.308***
##                                                    (0.361)
##
## VClassStandard Sport Utility Vehicle 4WD          -2.092***
##                                                    (0.709)
##
## VClassSubcompact Cars                               0.093
##                                                    (0.077)
##
## VClassTwo Seaters                                   0.213
##                                                    (0.258)
##
## Constant                                          25.807***
##                                                    (0.367)
##
## ------------------------------------------------------------------------
## Observations                                        20,681
## R2                                                   0.480
## Adjusted R2                                          0.480
## Residual Std. Error                          3.637 (df = 20661)
## F Statistic                                1,005.624*** (df = 19; 20661)
## ========================================================================
## Note:                                        *p<0.1; **p<0.05; ***p<0.01
```

```r
coeftest(mpg_fit_01, vconv = vcovHC(type = "HC1"))
```

```
##
## t test of coefficients:
##
##                                            Estimate Std. Error   t value
## (Intercept)                               25.8073808  0.3600947   71.6683
## displ                                     -2.9198636  0.0232379 -125.6509
## total_volume                               0.0205067  0.0034396    5.9619
## VClassLarge Cars                           0.3992574  0.1182282    3.3770
## VClassMidsize-Large Station Wagons        -2.1944501  0.2190810  -10.0166
## VClassMidsize Cars                         0.5796790  0.0805332    7.1980
## VClassMidsize Station Wagons              -1.6522677  0.2054690   -8.0414
## VClassMinicompact Cars                     0.1269058  0.1481553    0.8566
## VClassSmall Pickup Trucks 4WD             -5.6462664  3.6441462   -1.5494
## VClassSmall Sport Utility Vehicle 2WD      0.6517156  0.3685762    1.7682
## VClassSmall Sport Utility Vehicle 4WD      0.1880404  0.4371963    0.4301
## VClassSmall Station Wagons                 0.0696522  0.1179829    0.5904
## VClassSpecial Purpose Vehicle 2WD         -2.7976933  1.4869462   -1.8815
## VClassSpecial Purpose Vehicle 4WD         -3.2409893  2.1027673   -1.5413
## VClassSport Utility Vehicle - 2WD         -2.8219607  0.5412104   -5.2142
## VClassSport Utility Vehicle - 4WD         -2.3304875  0.3473094   -6.7101
## VClassStandard Sport Utility Vehicle 2WD  -1.3078760  1.2949842   -1.0100
## VClassStandard Sport Utility Vehicle 4WD  -2.0922679  0.8459209   -2.4734
## VClassSubcompact Cars                      0.0931763  0.0839990    1.1093
## VClassTwo Seaters                          0.2126235  0.3462830    0.6140
##                                            Pr(>|t|)
## (Intercept)                               < 2.2e-16 ***
## displ                                     < 2.2e-16 ***
## total_volume                              2.534e-09 ***
## VClassLarge Cars                          0.0007341 ***
## VClassMidsize-Large Station Wagons        < 2.2e-16 ***
## VClassMidsize Cars                        6.319e-13 ***
## VClassMidsize Station Wagons              9.353e-16 ***
## VClassMinicompact Cars                    0.3916912
## VClassSmall Pickup Trucks 4WD             0.1212992
## VClassSmall Sport Utility Vehicle 2WD     0.0770426 .
## VClassSmall Sport Utility Vehicle 4WD     0.6671236
## VClassSmall Station Wagons                0.5549570
## VClassSpecial Purpose Vehicle 2WD         0.0599176 .
## VClassSpecial Purpose Vehicle 4WD         0.1232598
## VClassSport Utility Vehicle - 2WD         1.864e-07 ***
## VClassSport Utility Vehicle - 4WD         1.995e-11 ***
## VClassStandard Sport Utility Vehicle 2WD  0.3125286
## VClassStandard Sport Utility Vehicle 4WD  0.0133929 *
## VClassSubcompact Cars                     0.2673331
## VClassTwo Seaters                         0.5392112
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
mpg_fit_o1_vif = ols_vif_tol(mpg_fit_01)
```