Assignment 2

1

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.4.4
```

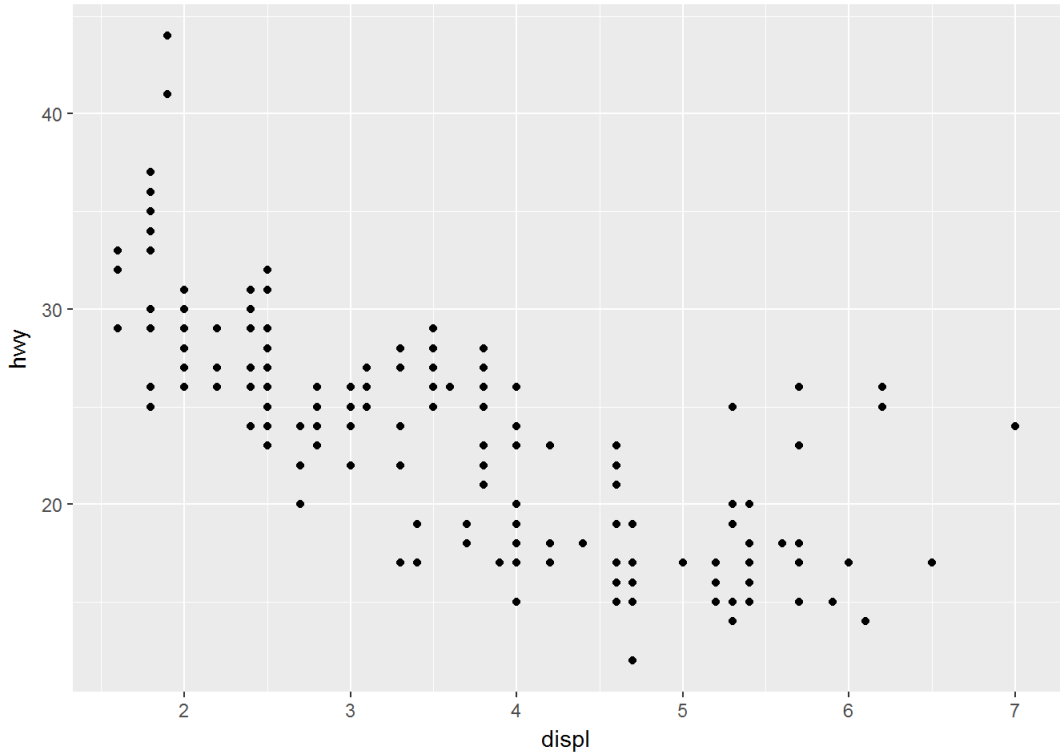
```
## -- Attaching packages ----------------------------------------------- tidyverse 1.2.1 --
```

```
## v ggplot2 2.2.1     v purrr   0.2.4
## v tibble  1.4.2     v dplyr   0.7.4
## v tidyr   0.8.0     v stringr 1.3.0
## v readr   1.1.1     v forcats 0.3.0
```

```
## Warning: package 'forcats' was built under R version 3.4.4
```

```
## -- Conflicts -------------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

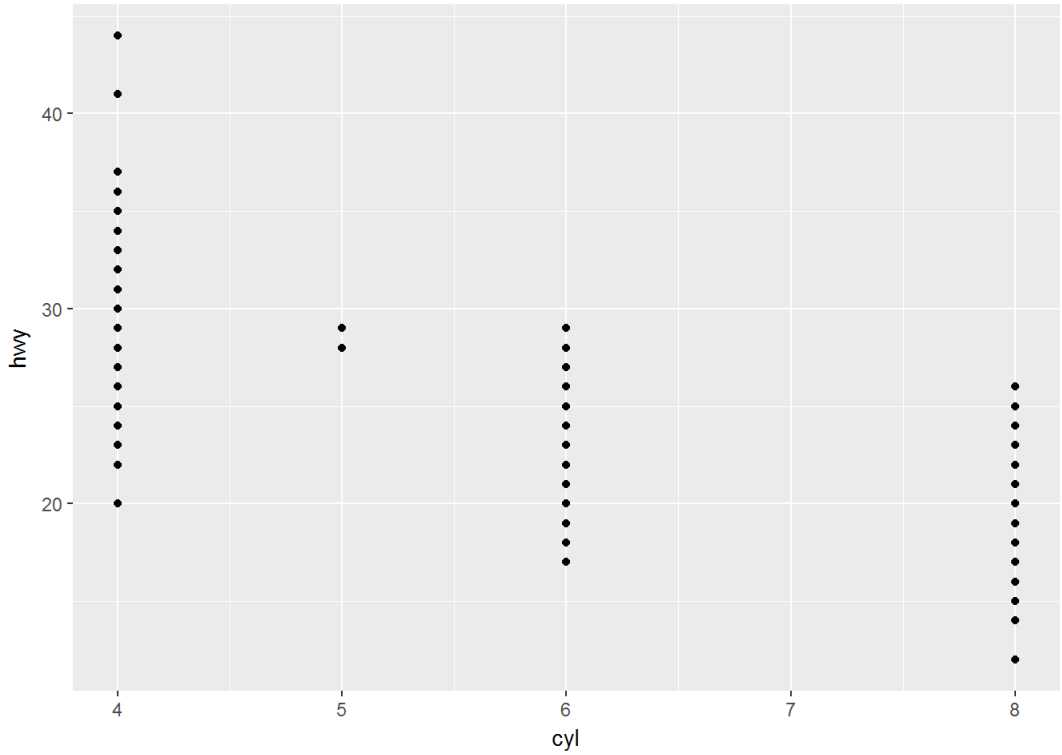
```
library(ggplot2)
data(mpg)
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy))
```



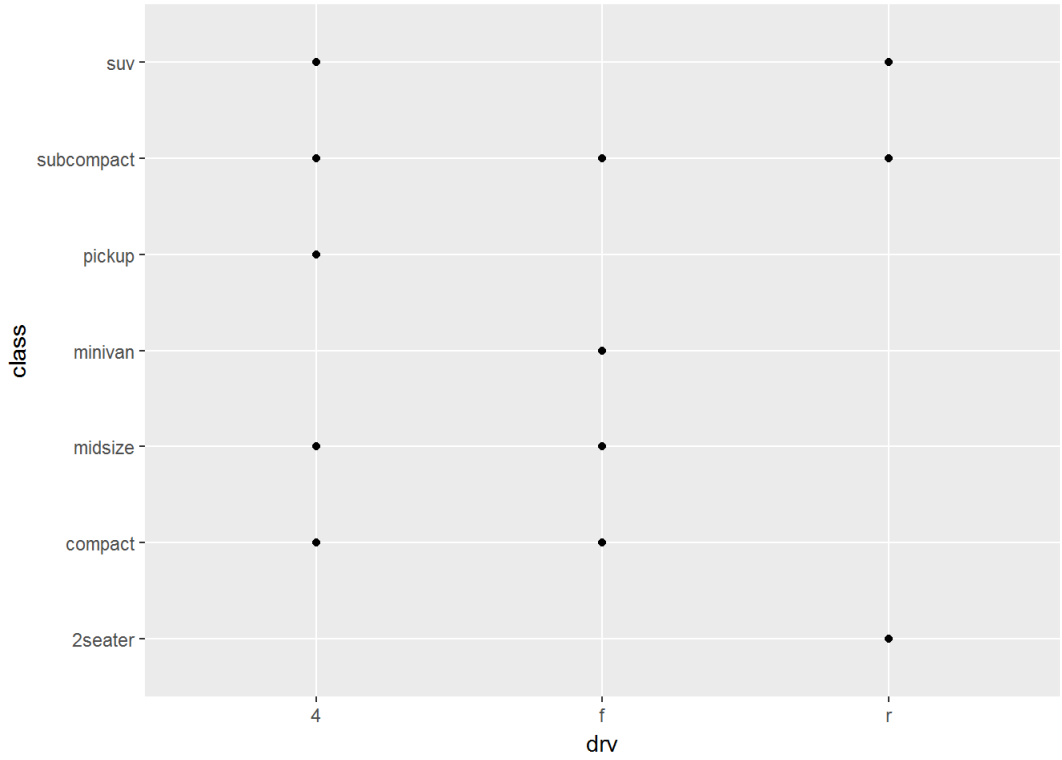

```
# Template for mapping
# ggplot(data = <DATA>) +
#   <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

a. 3.2.4 Exercises #4, #5

```
#4
ggplot(data = mpg) +
  geom_point(mapping = aes(x=cyl,y=hwy))
```



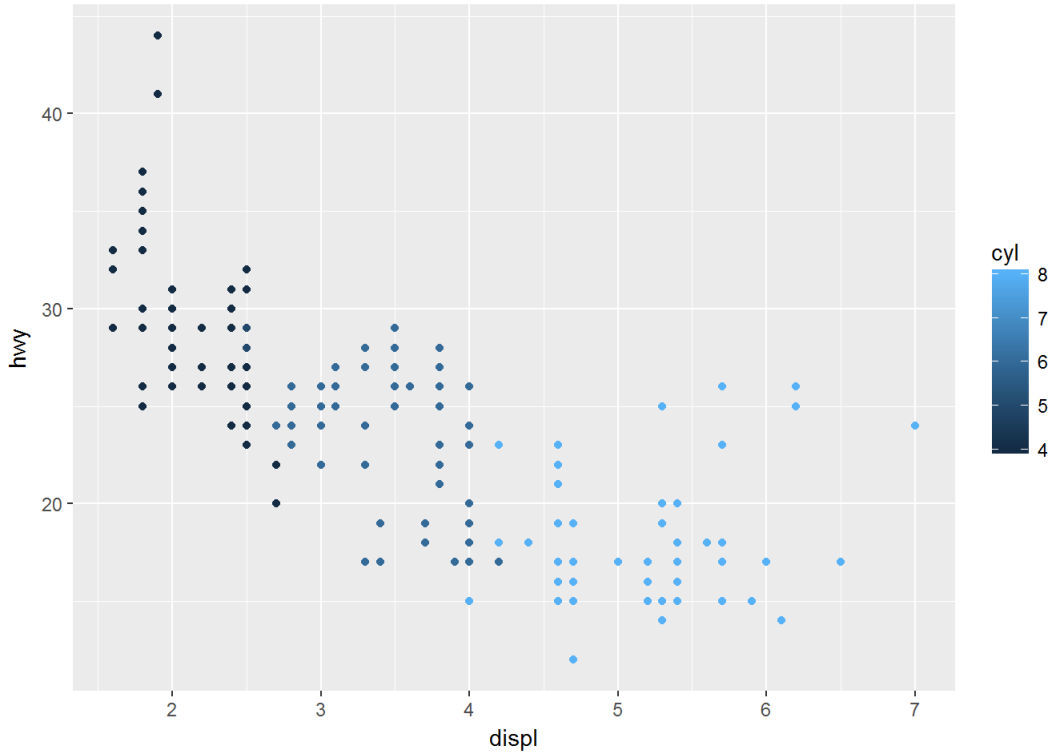

```
#5
ggplot(data = mpg) +
  geom_point(mapping = aes(x=drv,y=class))
```



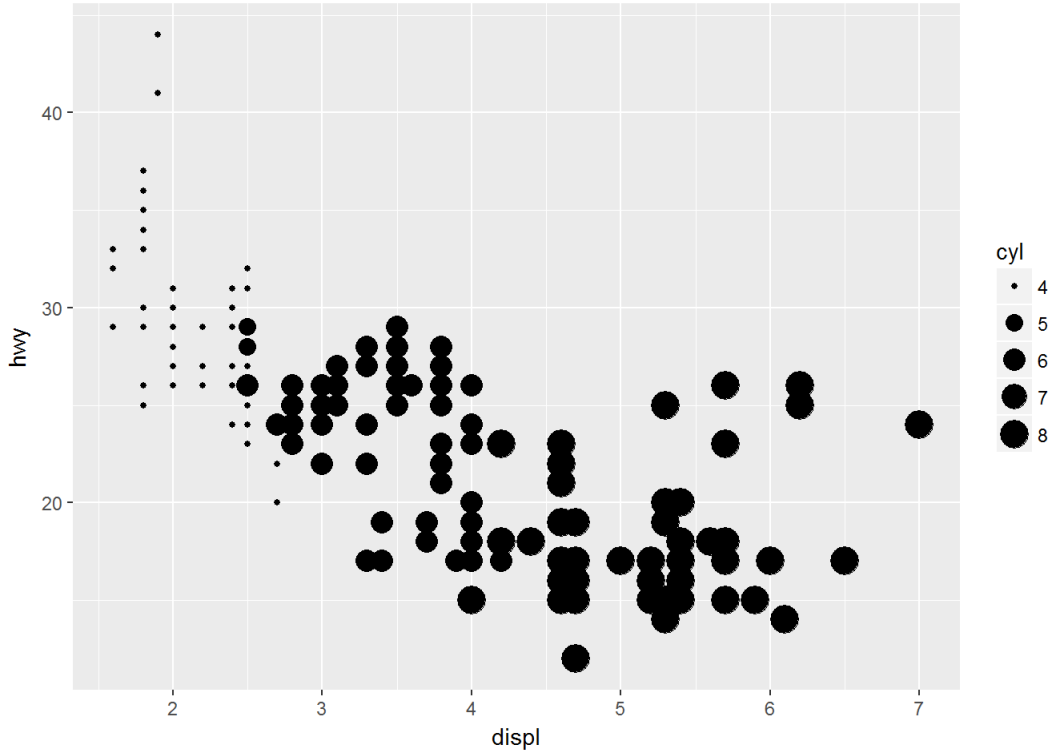
The plot for #5 is of not much use as it doesnt any correlation between the two varibles [“type of car” and “type of drive (front wheel, rear wheel or all wheels)”]

a. 3.3.1 Exercises #3, #4, #6

```
#3
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, color = cyl))
```



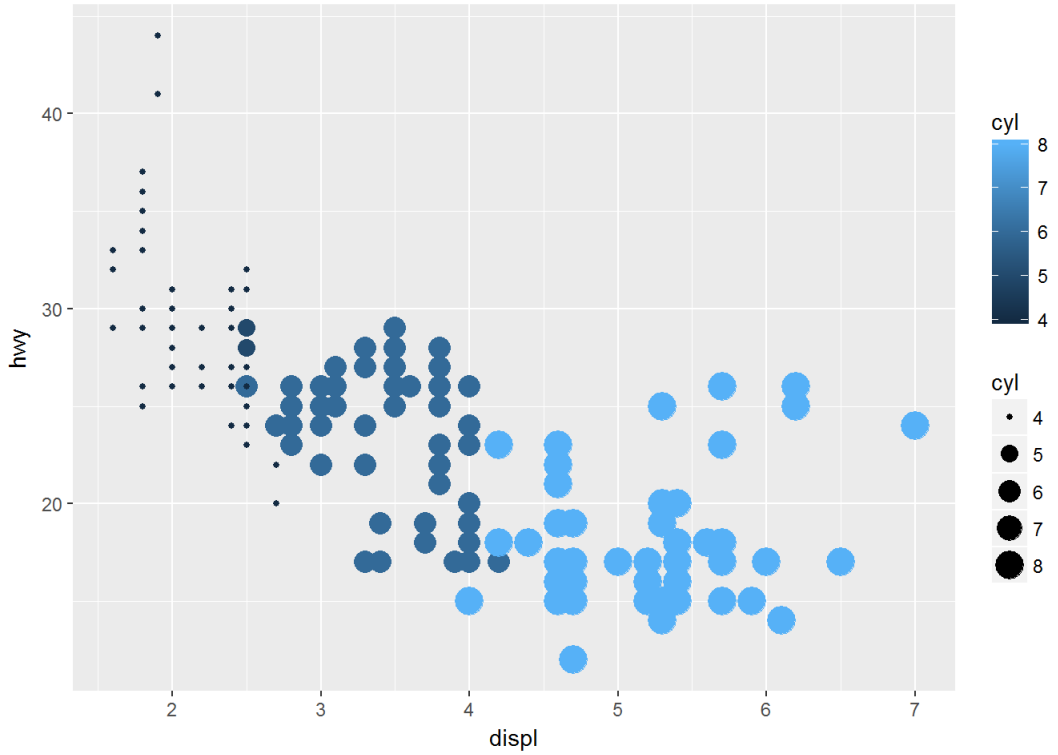

```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, size = cyl))
```



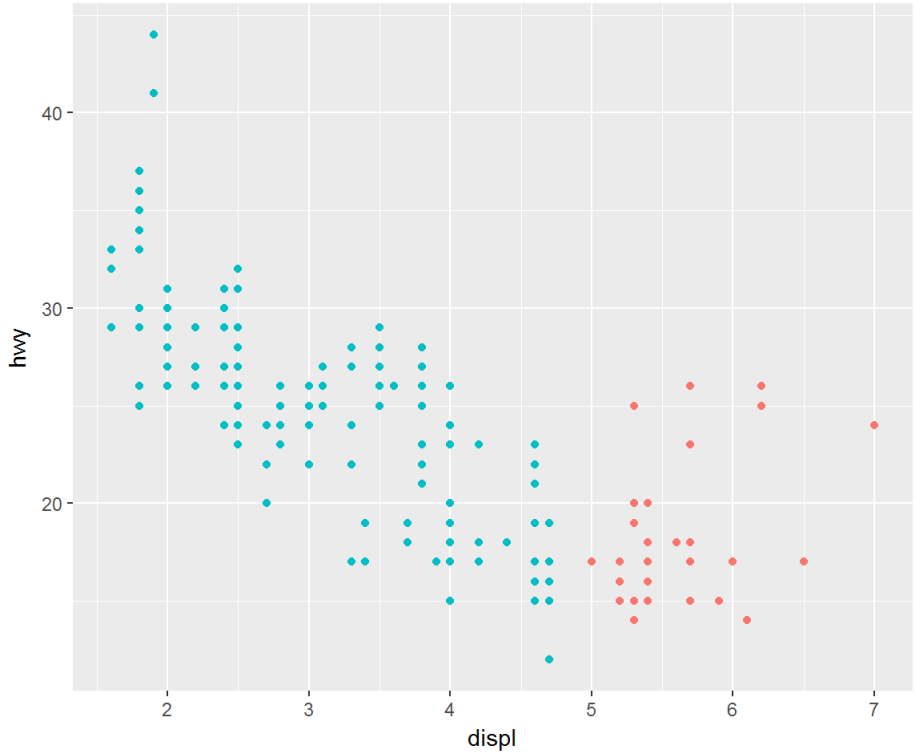

```
#ggplot(data = mpg) +
 # geom_point(mapping = aes(x = displ, y = hwy, shape = cyl))
#Error: A continuous variable can not be mapped to shape
```

While continuous variables do not get mapped to shape, categorical variables do. Also the categorical variables when represented through size do not give logical information as opposed to continuous variables.

```
#4
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, color = cyl, size=cyl))
```



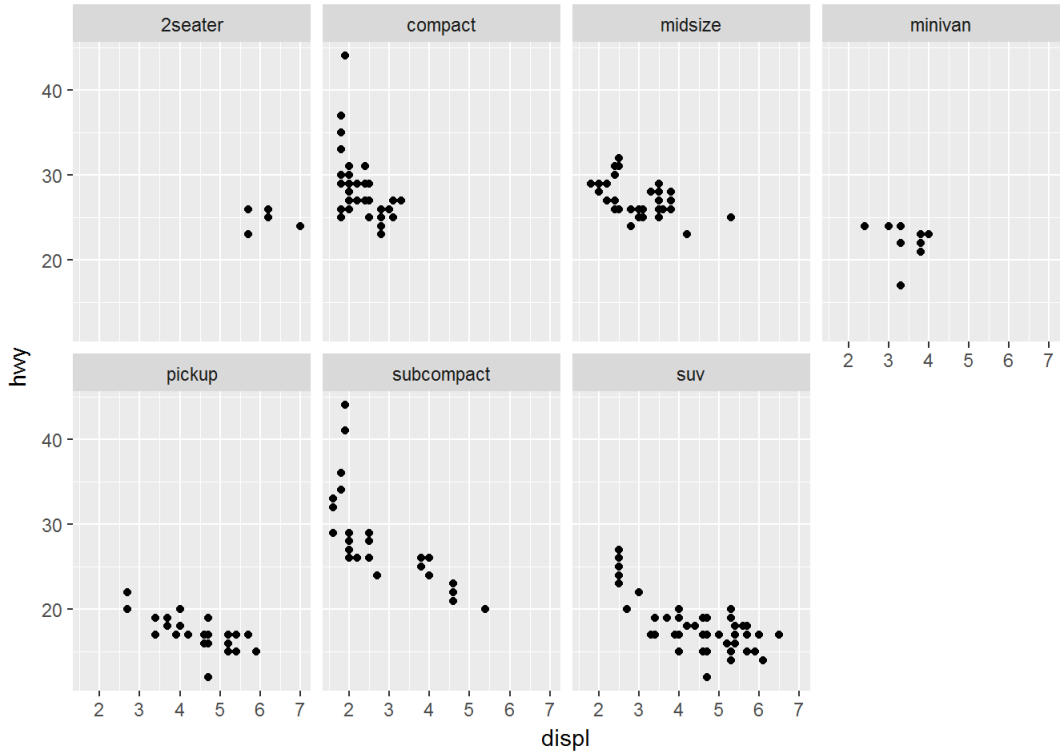

```
#The variable is represented through the multiple asethetics

#6
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, colour = displ < 5))
```



It differentiates the cars that have displ more than 5 from the ones that have it less than 5 using color coding.

a. 3.5.1 Exercises #4

```
#4
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy)) +
  facet_wrap(~ class, nrow = 2)
```

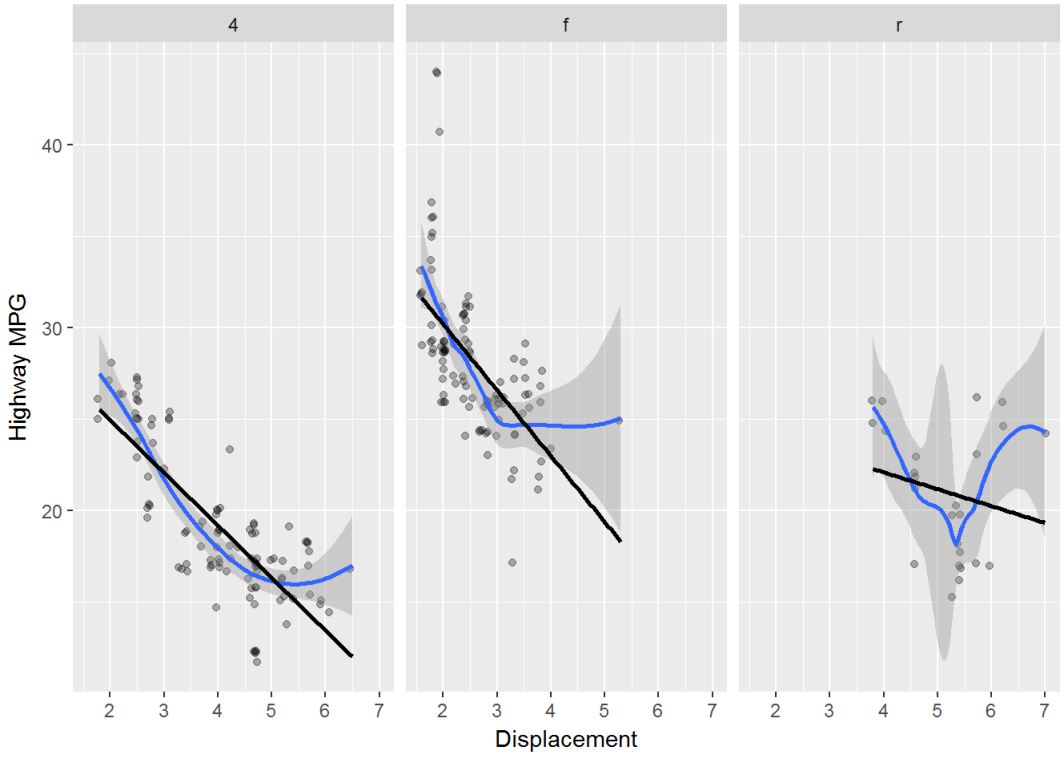


Faceting makes the difference amongst the varibles stand out, especially when the number of varibles to be represented is higher. Colouring makes it difficult to infer specific differentiation amongst the variables. However faceting doesnt provide the bigger picture as opposed to coloring, wherein you observe the entire dataset as a whole. Also faceting takes time to read while colours are easy on the eye when in shorter numbers.

b.

```
base1 <- ggplot(data = mpg, mapping = aes(x = displ, y = hwy))
base1 + geom_point(alpha=0.3,position = "jitter") + geom_smooth()+ geom_smooth(method=lm,colour="black",se=FALSE)+
  facet_grid(.~drv)+xlab("Displacement")+ylab("Highway MPG")
```

```
## `geom_smooth()` using method = 'loess'
```



2 Generating data and advanced density plots

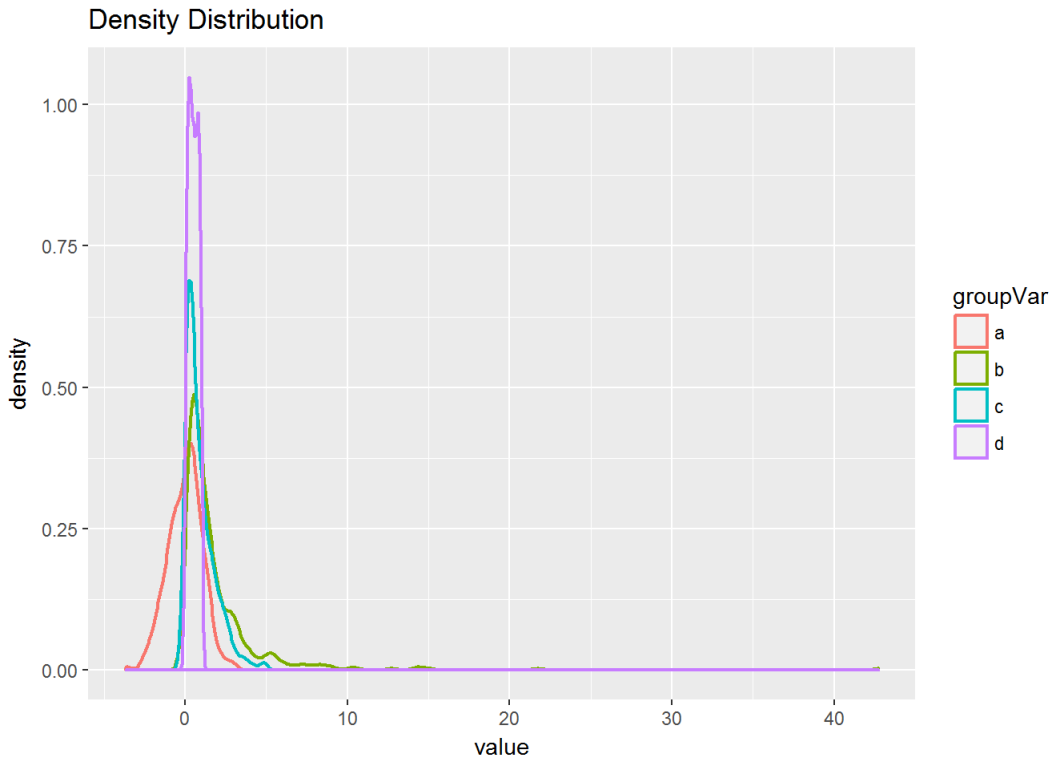
A.

```
df <- data.frame(matrix(ncol = 4, nrow = 500))
colnames(df) <- c("a","b","c","d")
df$a=rnorm(500)
df$b=rlnorm(500)
df$c=rexp(500)
df$d=runif(500)
library(tidyr)
df2=gather(df, key = "groupVar", value = "value",a:d)
head(df2)
```

```
##   groupVar      value
## 1        a  0.1933813
## 2        a  0.2610091
## 3        a  1.3188167
## 4        a -0.6898556
## 5        a -1.2657890
## 6        a  0.4891208
```

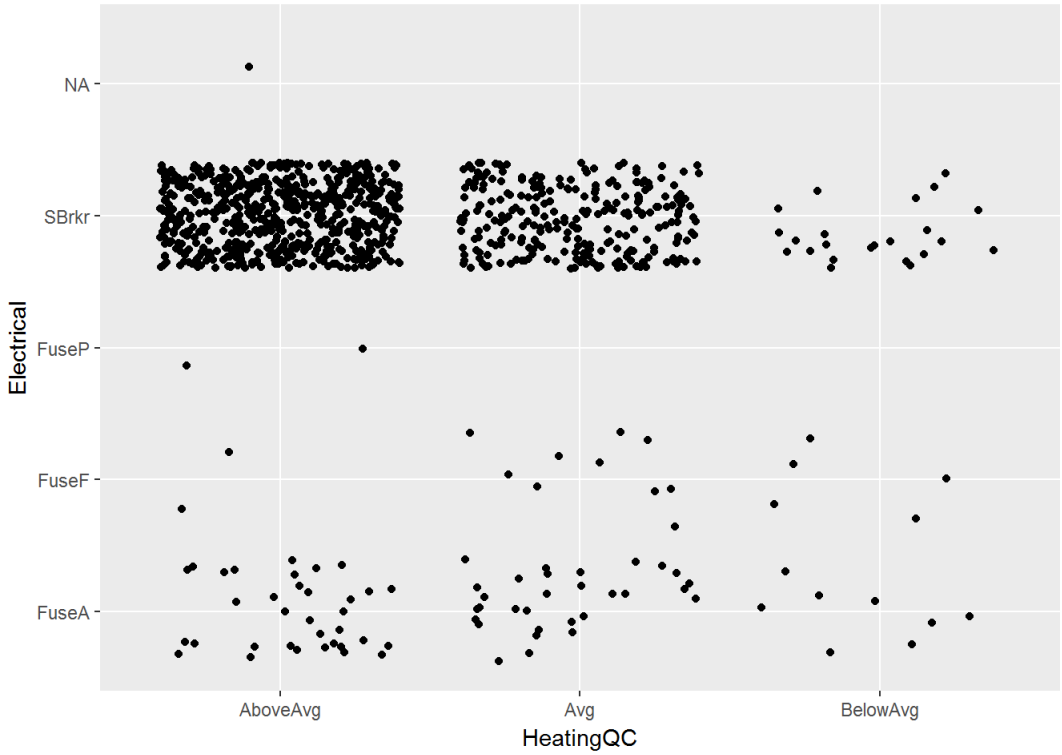
B.

```
ggplot(df2,aes(value))+geom_density(alpha=0.3,aes(colour=groupVar),size=0.8)+
  labs(title="Density Distribution")
```

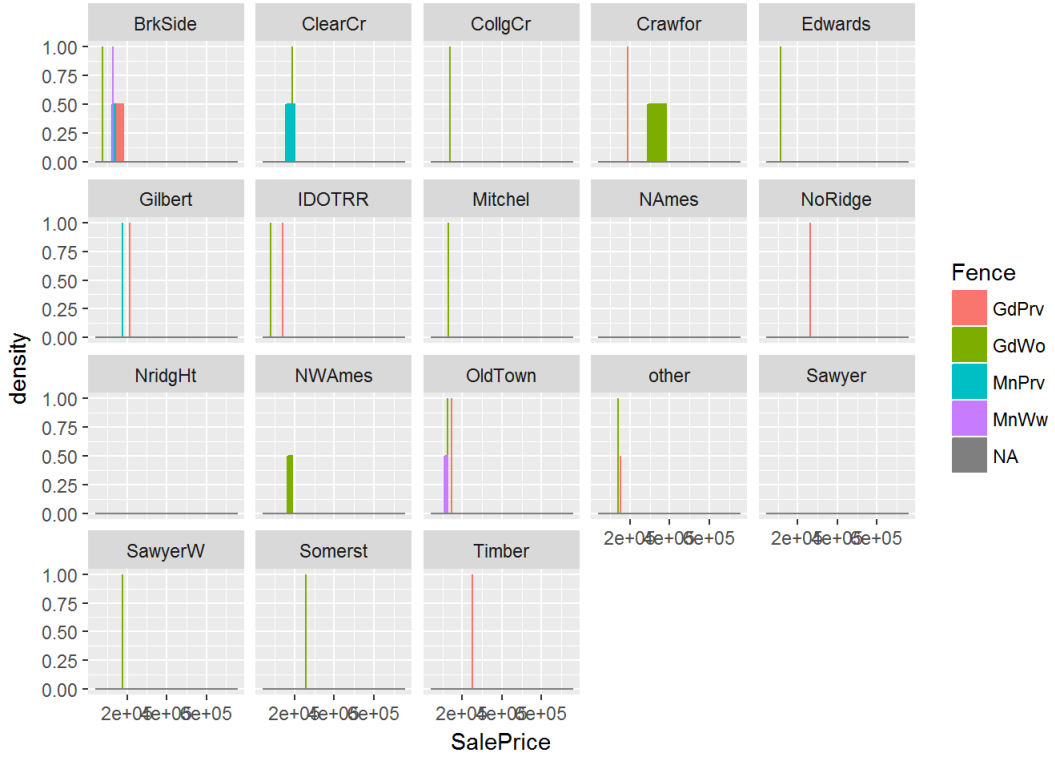


3 House prices data

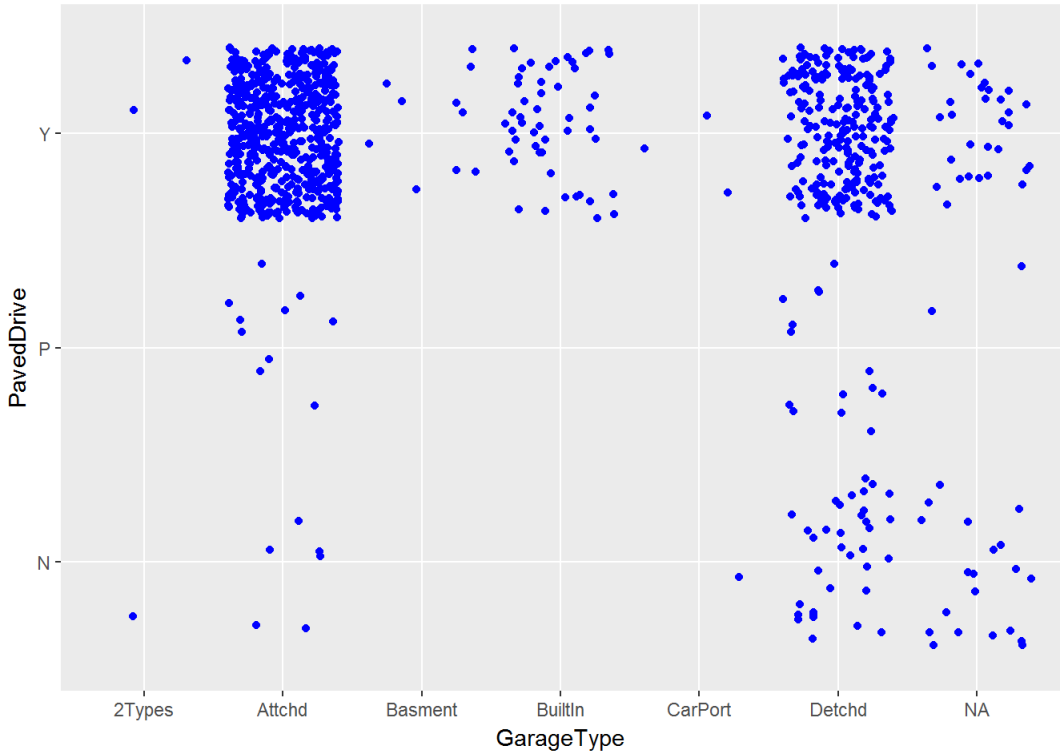
```
house=read.csv("housingData.csv")
ggplot(house)+geom_jitter(aes(x=HeatingQC,y=Electrical))
```




```
#shows that the Standard Circuit Breakers & Romex are the best circuit systems
#to be used as the show the Above Average Heating conditions

ggplot(house, aes(x=SalePrice, color=Fence)) +
  geom_density(aes(fill=Fence)) +
  facet_wrap(~Neighborhood)
```

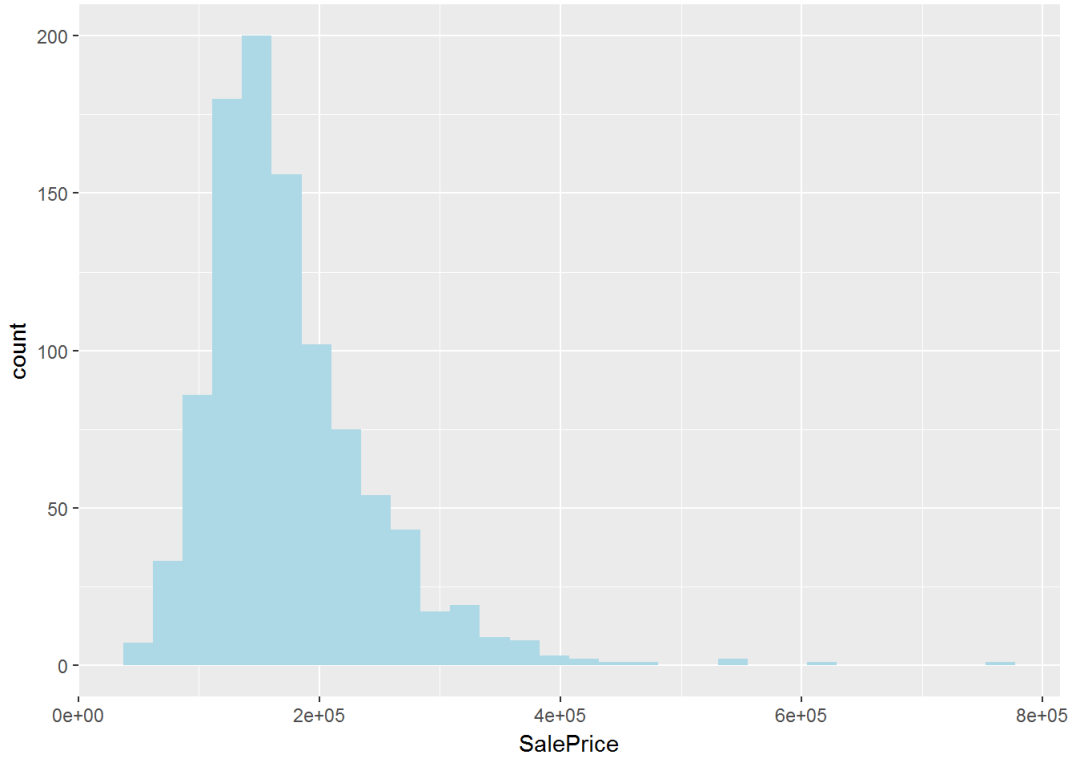



```
#We find that Brookside neighborhood has lower priced property giving good privacy
#through its fencing. Similarly we find Crawford giving good privacy too.

ggplot(house,aes())+geom_jitter(aes(y=PavedDrive,x=GarageType),colour="Blue")
```

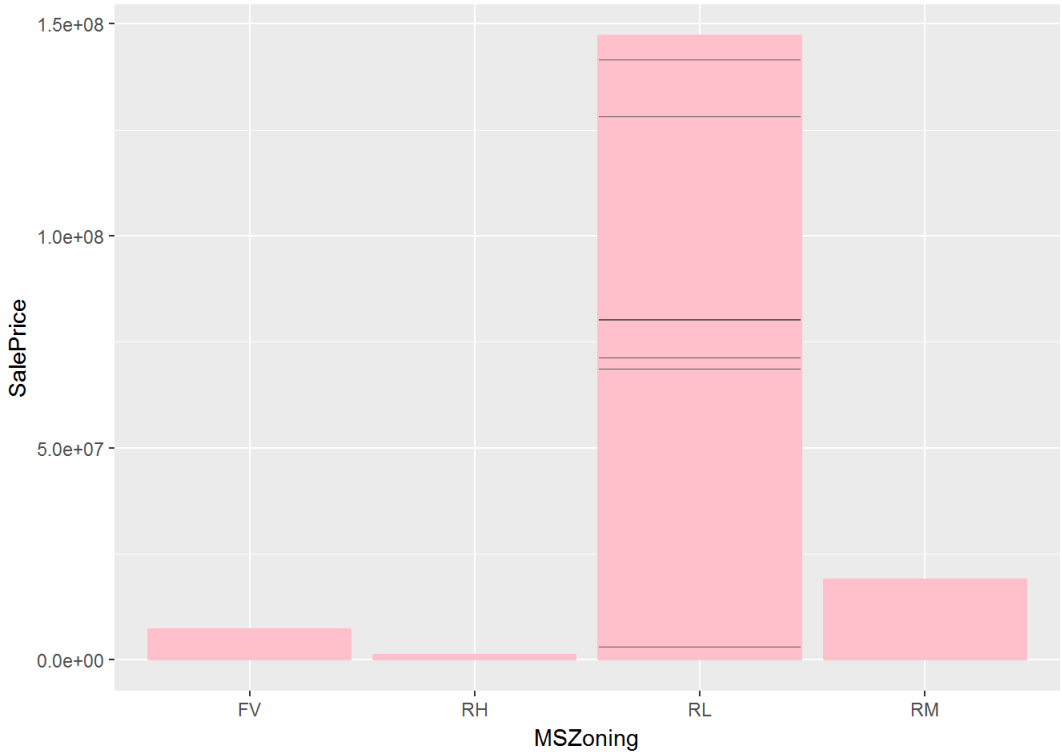



```
#Most of the Garages that are attached to the home have a Paved Drive.

ggplot(house, aes(x=SalePrice)) +
  geom_histogram(stat="bin",fill="Light Blue")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
#Most of the sales happened for the range of 100000 to 200000 dollars

ggplot(house, aes(x=MSZoning,y=SalePrice)) +
  geom_bar(stat="identity",colour="pink")
```




```
#We see the residential Low density zoned houses to be costlier than the highly
#populated places.
```

4 Missing Data

```
#a
library(Amelia)
```

```
## Warning: package 'Amelia' was built under R version 3.4.4
```

```
## Loading required package: Rcpp
```

```
## Warning: package 'Rcpp' was built under R version 3.4.4
```

```
## ##
## ## Amelia II: Multiple Imputation
## ## (Version 1.7.5, built: 2018-05-07)
## ## Copyright (C) 2005-2018 James Honaker, Gary King and Matthew Blackwell
## ## Refer to http://gking.harvard.edu/amelia/ for more information
## ##
```

```
data(freetrade)
library(VIM)
```

```
## Warning: package 'VIM' was built under R version 3.4.4
```

```
## Loading required package: colorspace
```

```
## Loading required package: grid
```

```
## Loading required package: data.table
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
## The following object is masked from 'package:purrr':
##
##     transpose
```

```
## VIM is ready to use.
##  Since version 4.0.0 the GUI is in its own package VIMGUI.
##
##           Please use the package to use the new (and old) GUI.
```

```
## Suggestions and bug-reports can be submitted at: https://github.com/alexkowa/VIM/issues
```

```
##
## Attaching package: 'VIM'
```

```
## The following object is masked from 'package:datasets':
##
##     sleep
```

```
library(mice)
```

```
## Warning: package 'mice' was built under R version 3.4.4
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'mice'
```

```
## The following object is masked from 'package:tidyr':
##
##     complete
```

```
## The following objects are masked from 'package:base':
##
##     cbind, rbind
```

```
freetrade[complete.cases(freetrade),] #Give the results for observed data
```

```
##     year     country tariff polity       pop     gdp.pc  intresmi signed
## 3   1983    SriLanka   41.3      5  15417000   489.2266 1.6639363      1
## 5   1985    SriLanka   31.0      5  15837000   525.5609 2.2591159      0
## 7   1987    SriLanka   27.3      5  16361000   540.0475 1.4229829      0
## 8   1988    SriLanka   27.3      5  16587000   545.8610 1.0596240      1
## 10  1990    SriLanka   28.3      5  16993000   579.9548 1.6636324      0
## 11  1991    SriLanka   26.9      5  17247000   597.6987 2.2852130      1
## 13  1993    SriLanka   24.2      5  17628420   652.6205 4.2803612      0
## 14  1994    SriLanka   26.0      5  17865000   680.0408 4.3899121      0
## 15  1995    SriLanka   20.0      5  18112000   707.6591 3.9959190      0
## 17  1997    SriLanka   20.0      5  18552000   763.0638 3.5143008      0
## 20  1981       India   74.3      8 702821248   236.0894 5.5428457      1
## 25  1986       India  100.0      8 781892992   271.3088 6.0622978      0
## 26  1987       India   98.8      8 798680000   278.2737 5.7676215      0
## 29  1990       India   81.8      8 849515008   323.8349 2.0127988      0
## 30  1991       India   79.2      8 866530432   318.8200 3.2970126      1
## 31  1992       India   53.0      8 882300032   330.0941 3.6984935      0
## 32  1993       India   47.8      8 898200000   340.3172 5.1200824      0
## 33  1994       India   47.8      8 913600000   359.1398 6.3580613      0
## 34  1995       India   41.0      9 929358016   380.0712 4.8719177      0
## 35  1996       India   38.7      9 945611776   399.5068 4.9265795      0
## 36  1997       India   35.0      9 962377664   410.4688 5.3202863      0
## 42  1984   Indonesia   37.0     -7 160075456   593.2498 2.8442452      0
```

```
## 42  1984   Indonesia   37.0     -7 160075456   595.2496 2.6442452      0
## 43  1985   Indonesia   27.0     -7 163036000   602.7329 3.2447402      0
## 44  1986   Indonesia   31.5     -7 166015056   627.2222 3.1365039      0
## 47  1989   Indonesia   25.2     -7 175063344   726.6517 2.9932704      0
## 48  1990   Indonesia   20.6     -7 178232000   777.9812 3.1374667      0
## 49  1991   Indonesia   20.3     -7 181397024   832.6502 3.2799203      0
## 50  1992   Indonesia   20.0     -7 184556192   877.4985 3.3316619      0
## 51  1993   Indonesia   19.4     -7 187707264   925.3455 3.3837914      0
## 54  1996   Indonesia   13.2     -7 197156480  1105.4620 3.4949601      0
## 59  1982       Korea   23.7     -5  39326000  4171.4204 1.1417191      0
## 60  1983       Korea   23.7     -5  39910000  4583.0039 0.9219497      1
## 61  1984       Korea   21.9     -5  40406000  4919.5259 0.9729462      0
## 65  1988       Korea   18.9      6  41975000  6984.6323 2.5869560      0
## 66  1989       Korea   14.9      6  42380000  7359.5601 2.6701207      0
## 67  1990       Korea   13.3      6  42869000  7967.3906 2.2559617      0
## 68  1991       Korea   11.4      6  43268000  8622.3135 1.7905118      0
## 69  1992       Korea   10.1      6  43663000  9008.9160 2.1899016      0
## 70  1993       Korea    8.9      6  44056000  9418.8623 2.4988637      0
## 73  1996       Korea   13.4      6  45545000 11467.4053 2.2771194      0
## 74  1997       Korea   13.3      6  45991000 11925.2627 1.3825226      1
## 77  1981    Malaysia   10.6      4  14105080  2397.0015 3.6918523      0
## 82  1986    Malaysia   15.8      4  16143010  2540.9355 5.0967517      0
## 83  1987    Malaysia   13.6      4  16633580  2598.8804 5.5652966      0
## 84  1988    Malaysia   13.0      4  17144390  2772.0222 3.9440835      0
## 85  1989    Malaysia   17.0      4  17669600  2933.2661 3.6590879      0
## 87  1991    Malaysia   16.9      4  18656950  3317.3928 3.2112672      0
## 88  1992    Malaysia   12.8      4  19127100  3523.3594 4.4357429      0
## 89  1993    Malaysia   14.3      4  19609110  3776.8167 5.8374891      0
## 90  1994    Malaysia   13.0      4  20103260  4023.3499 4.3134141      0
## 92  1996    Malaysia    8.7      3  21129230  4624.7827 3.4246469      0
## 93  1997    Malaysia    9.1      3  21667000  4840.2520 2.6002250      0
## 98  1983       Nepal   22.1     -2  15663490   149.5029 4.7332282      0
## 102 1987       Nepal   21.0     -2  17382740   166.7793 4.9377427      1
## 103 1988       Nepal   22.6     -2  17836420   175.0473 4.4469566      0
## 108 1993       Nepal   16.1      5  20244250   193.6161 7.9346037      0
## 110 1995       Nepal   11.0      5  21272000   206.3166 5.0656819      0
## 116 1982    Pakistan   77.6     -7  87436144   345.6995 3.0479228      0
## 117 1983    Pakistan   77.0     -7  89831888   359.2879 4.5125327      0
## 118 1984    Pakistan   77.0     -7  92284304   367.4550 2.5139329      0
## 119 1985    Pakistan   77.0     -4  94794432   384.8838 2.2084200      0
## 120 1986    Pakistan   66.0     -4  97353880   395.3834 2.2022588      0
## 121 1987    Pakistan   68.9     -4  99953232   409.9492 2.2189448      0
## 122 1988    Pakistan   69.0      8 102621984   429.7351 1.5439408      1
## 123 1989    Pakistan   65.0      8 105269632   439.7046 1.6029303      0
## 124 1990    Pakistan   64.8      8 107975056   447.8007 1.2045232      0
## 125 1991    Pakistan   66.0      8 110750016   460.3921 1.2399414      0
## 126 1992    Pakistan   61.1      8 113561960   484.1700 1.4281969      0
## 127 1993    Pakistan   56.0      8 116444160   481.2280 1.6599897      1
## 128 1994    Pakistan   51.0      8 119401848   487.5894 3.4947820      1
## 129 1995    Pakistan   50.7      8 122374952   500.1006 2.0452147      1
## 130 1996    Pakistan   41.7      8 125409848   506.8169 0.9035686      0
## 134 1981 Philippines   34.6     -8  49740480  1171.4153 2.9328992      0
## 135 1982 Philippines   31.4     -7  51039800  1182.9126 1.7864840      0
## 136 1983 Philippines   29.5     -6  52262480  1176.8947 0.9042833      1
## 137 1984 Philippines   28.8     -6  53456040  1066.3496 1.0449501      0
## 138 1985 Philippines   27.6     -6  54668000   966.5225 1.5818520      0
## 140 1987 Philippines   27.9      8  57020400   999.6287 2.7225583      0
## 141 1988 Philippines   27.9      8  58176680  1045.9196 2.1998901      0
## 142 1989 Philippines   27.6      8  59383040  1088.2557 1.9586543      1
## 143 1990 Philippines   27.8      8  60687000  1097.2126 1.4865717      0
## 144 1991 Philippines   26.0      8  62148312  1065.2173 3.2610879      1
## 145 1992 Philippines   24.3      8  63644808  1043.6821 3.3457856      0
## 146 1993 Philippines   22.6      8  65177340  1040.7100 3.1525953      0
## 147 1994 Philippines   21.7      8  66746768  1060.8284 3.0627861      0
## 148 1995 Philippines   20.0      8  68354000  1084.3506 2.6057911      1
## 149 1996 Philippines   14.3      8  69913752  1122.1346 3.1971264      0
## 150 1997 Philippines   13.4      8  71421040  1155.4116 1.9547111      0
## 153 1981    Thailand   32.3      2  47688900  1161.3615 2.7221076      1
## 157 1985    Thailand   41.2      2  51146000  1333.0128 3.0223274      1
## 161 1989    Thailand   40.8      3  54618620  1833.6932 4.2566657      0
## 162 1990    Thailand   39.8      3  55595000  2002.6646 4.4117045      0
## 163 1991    Thailand   37.8     -1  56454008  2140.9771 4.8411651      0
## 165 1993    Thailand   45.6      9  57796980  2449.7668 5.3830957      0
```

```
## 166 1994    Thailand   23.3      9  58271528  2647.3428 5.3520565      0
## 167 1995    Thailand   23.1      9  58610000  2866.3428 5.0294995      0
##     fiveop     usheg
## 3     12.3 0.2655022
## 5     12.3 0.2952431
## 7     12.5 0.2734092
## 8     12.6 0.2756469
## 10    12.7 0.2608332
## 11    12.8 0.2589872
## 13    13.2 0.2812928
## 14    13.2 0.2783585
## 15    13.2 0.2627195
## 17    13.2 0.2844036
## 20    12.4 0.2593112
## 25    12.5 0.2886563
## 26    12.5 0.2734092
## 29    12.7 0.2608332
## 30    12.8 0.2589872
## 31    13.1 0.2623017
## 32    13.2 0.2812928
## 33    13.2 0.2783585
## 34    13.2 0.2627195
## 35    13.2 0.2681700
## 36    13.2 0.2844036
## 42    12.3 0.2988009
## 43    12.3 0.2952431
## 44    12.5 0.2886563
## 47    12.6 0.2785387
## 48    12.7 0.2608332
## 49    12.8 0.2589872
## 50    13.1 0.2623017
## 51    13.2 0.2812928
## 54    13.2 0.2681700
## 59    12.5 0.2558008
## 60    12.3 0.2655022
## 61    12.3 0.2988009
## 65    12.6 0.2756469
## 66    12.6 0.2785387
## 67    12.7 0.2608332
## 68    12.8 0.2589872
## 69    13.1 0.2623017
## 70    13.2 0.2812928
## 73    13.2 0.2681700
## 74    13.2 0.2844036
## 77    12.4 0.2593112
## 82    12.5 0.2886563
## 83    12.5 0.2734092
## 84    12.6 0.2756469
## 85    12.6 0.2785387
## 87    12.8 0.2589872
## 88    13.1 0.2623017
## 89    13.2 0.2812928
## 90    13.2 0.2783585
## 92    13.2 0.2681700
## 93    13.2 0.2844036
## 98    12.3 0.2655022
## 102   12.5 0.2734092
## 103   12.6 0.2756469
## 108   13.2 0.2812928
## 110   13.2 0.2627195
## 116   12.5 0.2558008
## 117   12.3 0.2655022
## 118   12.3 0.2988009
## 119   12.3 0.2952431
## 120   12.5 0.2886563
## 121   12.5 0.2734092
## 122   12.6 0.2756469
## 123   12.6 0.2785387
## 124   12.7 0.2608332
## 125   12.8 0.2589872
## 126   13.1 0.2623017
## 127   13.2 0.2812928
## 128   13.2 0.2783585
```

```
## 129   13.2 0.2627195
## 130   13.2 0.2681700
## 134   12.4 0.2593112
## 135   12.5 0.2558008
## 136   12.3 0.2655022
## 137   12.3 0.2988009
## 138   12.3 0.2952431
## 140   12.5 0.2734092
## 141   12.6 0.2756469
## 142   12.6 0.2785387
## 143   12.7 0.2608332
## 144   12.8 0.2589872
## 145   13.1 0.2623017
## 146   13.2 0.2812928
## 147   13.2 0.2783585
## 148   13.2 0.2627195
## 149   13.2 0.2681700
## 150   13.2 0.2844036
## 153   12.4 0.2593112
## 157   12.3 0.2952431
## 161   12.6 0.2785387
## 162   12.7 0.2608332
## 163   12.8 0.2589872
## 165   13.2 0.2812928
## 166   13.2 0.2783585
## 167   13.2 0.2627195
```

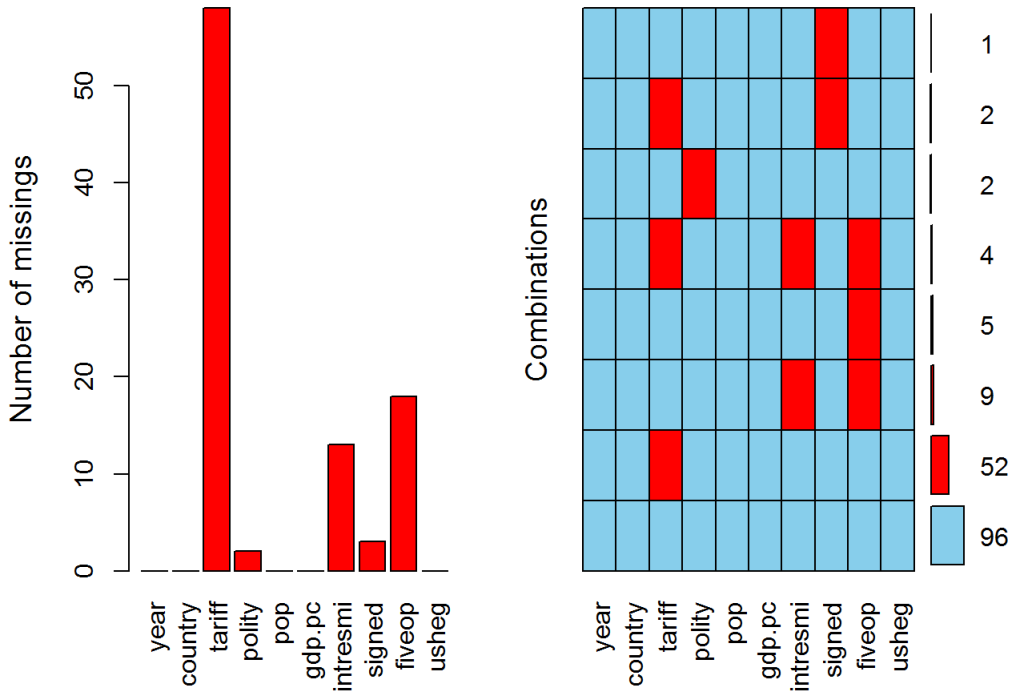
```
freetrade[!complete.cases(freetrade),] #Give the results for unobserved data
```

```
##      year     country tariff polity       pop     gdp.pc  intresmi signed
## 1    1981    SriLanka     NA      6  14988000   461.0236 1.9373473      0
## 2    1982    SriLanka     NA      5  15189000   473.7634 1.9644299      0
## 4    1984    SriLanka     NA      5  15599000   508.1739 2.7974622      0
## 6    1986    SriLanka     NA      5  16117000   538.9237 1.8325486      0
## 9    1989    SriLanka     NA      5  16806000   551.1353 1.1375573      0
## 12   1992    SriLanka   25.0      5  17405000   618.3329 2.8778768     NA
## 16   1996    SriLanka     NA      5  18300000   727.0039 3.6767628      0
## 18   1998    SriLanka     NA      5  18778000   789.3003        NA      0
## 19   1999    SriLanka     NA      6  18985000   814.2686        NA      0
## 21   1982       India     NA      8 718425600   239.6947 5.9615273      0
## 22   1983       India     NA      8 734071936   252.0074 5.6787739      0
## 23   1984       India     NA      8 749676928   255.8381 5.6151156      0
## 24   1985       India     NA      8 765147008   264.3729 5.6777949      0
## 27   1988       India     NA      8 815590016   299.5461 3.8956485      0
## 28   1989       India     NA      8 832534976   312.7361 3.2554514      0
## 37   1998       India   30.0      9 979672896   430.4632 5.6912227      0
## 38   1999       India   32.2      9 997515200   450.2433        NA      0
## 39   1981   Indonesia     NA     -7 151304976   534.1083 2.9180412      0
## 40   1982   Indonesia     NA     -7 154244608   529.7137 2.0481887      0
## 41   1983   Indonesia     NA     -7 157156544   563.8297 2.2353837      0
## 45   1987   Indonesia     NA     -7 168989536   648.8398 4.0184855      0
## 46   1988   Indonesia     NA     -7 171994064   678.0231 3.2952881     NA
## 52   1994   Indonesia     NA     -7 190847968   978.7410 3.2307374      0
## 53   1995   Indonesia     NA     -7 193976000  1042.0466 2.9021389      0
## 55   1997   Indonesia     NA     -7 200390288  1138.7395 2.9548082      1
## 56   1998   Indonesia    9.5     -5 203678368   974.6287 5.2577205      0
## 57   1999   Indonesia   10.9      7 207021616   961.8376        NA      0
## 58   1981       Korea     NA     -5  38723000  3937.6125 1.0570642      1
## 62   1985       Korea     NA     -5  40806000  5190.2485 1.0485235      1
## 63   1986       Korea     NA     -5  41184000  5736.8276 1.0933198      0
## 64   1987       Korea   22.9     NA  41575000  6337.4648 0.9516343      0
## 71   1994       Korea     NA      6  44453000 10104.8682 2.5817108      0
## 72   1995       Korea     NA      6  44995000 10873.6016 2.4654078      0
## 75   1998       Korea   11.1      8  46430000 11022.3311 5.0923381      0
## 76   1999       Korea    8.7      8  46858000 12086.2324        NA      0
## 78   1982    Malaysia     NA      4  14465810  2476.0813 3.2455084      0
## 79   1983    Malaysia     NA      4  14847000  2563.2966 2.8454931      0
## 80   1984    Malaysia     NA      4  15250250  2689.2180 2.5748692      0
## 81   1985    Malaysia     NA      4  15677000  2586.6553 3.7058082      0
## 86   1990    Malaysia     NA      4  18201900  3104.0334 3.6044145      0
## 91   1995    Malaysia     NA      3  20609860  4310.1919 3.1658885      0
```

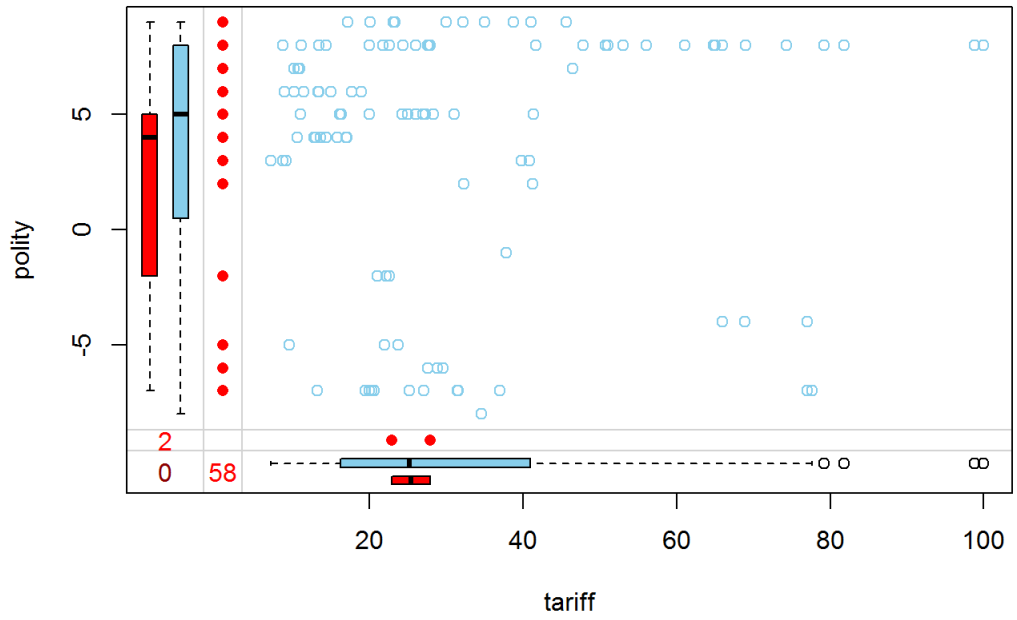
```
## 94  1998    Malaysia    7.1      3  22180000  4379.9424        NA      0
## 95  1999    Malaysia     NA      4  22710000  4525.9541        NA      0
## 96  1981       Nepal     NA     -2  14874560   156.3544 7.7400808      0
## 97  1982       Nepal     NA     -2  15262920   158.1348 7.7678432      0
## 99  1984       Nepal     NA     -2  16076710   159.7617 3.6014018      0
## 100 1985       Nepal     NA     -2  16503000   165.1985 2.4010794      1
## 101 1986       Nepal     NA     -2  16938310   168.3015 3.0573549      0
## 104 1989       Nepal     NA     -2  18299470   178.0051 4.2687879      0
## 105 1990       Nepal     NA      5  18772000   181.5673 5.4928718      0
## 106 1991       Nepal     NA      5  19253590   188.2990 6.1504669      0
## 107 1992       Nepal     NA      5  19744320   191.1591 6.6667037      1
## 109 1994       Nepal     NA      5  20753450   204.3828 7.0388422      0
## 111 1996       Nepal     NA      5  21794680   212.0982 4.5800462      0
## 112 1997       Nepal     NA      5  22321160   217.5510 4.2696733      0
## 113 1998       Nepal   16.3      5  22851110   218.7851 5.7470169      0
## 114 1999       Nepal   17.7      6  23384180   222.2162        NA      0
## 115 1981    Pakistan     NA     -7  85096000   333.4097 2.5595610      0
## 131 1997    Pakistan     NA      7 128457312   499.8134 1.2861589      0
## 132 1998    Pakistan   46.5      7 131582000   500.3802        NA      0
## 133 1999    Pakistan     NA     -6 134790000   507.8308        NA      0
## 139 1986 Philippines   27.9     NA  55866680   978.1002 3.8589563      1
## 151 1998 Philippines   10.7      7  72871008  1123.7935 3.0419078      1
## 152 1999 Philippines   10.1      7  74258872  1137.7983        NA      0
## 154 1982    Thailand     NA      2  48633368  1199.7605 3.0265975      1
## 155 1983    Thailand     NA      2  49535192  1243.6954 2.4778438      0
## 156 1984    Thailand     NA      2  50378140  1293.2310 2.5394394      0
## 158 1986    Thailand     NA      2  51952480  1384.9415 3.7410393      0
## 159 1987    Thailand     NA      2  52799040  1492.4540 3.8199978      0
## 160 1988    Thailand     NA      3  53687208  1662.8018 3.6140218      0
## 164 1992    Thailand     NA      9  57189720  2284.2720 5.0973606      0
## 168 1996    Thailand     NA      9  58976152  3017.3818 5.1052136      0
## 169 1997    Thailand     NA      9  59370472  2946.9143 4.0518551     NA
## 170 1998    Thailand   20.1      9  59793500  2628.5005        NA      0
## 171 1999    Thailand   17.1      9  60245800  2717.2185        NA      0
##     fiveop     usheg
## 1     12.4 0.2593112
## 2     12.5 0.2558008
## 4     12.3 0.2988009
## 6     12.5 0.2886563
## 9     12.6 0.2785387
## 12    13.1 0.2623017
## 16    13.2 0.2681700
## 18      NA 0.2956678
## 19      NA 0.3083147
## 21    12.5 0.2558008
## 22    12.3 0.2655022
## 23    12.3 0.2988009
## 24    12.3 0.2952431
## 27    12.6 0.2756469
## 28    12.6 0.2785387
## 37      NA 0.2956678
## 38      NA 0.3083147
## 39    12.4 0.2593112
## 40    12.5 0.2558008
## 41    12.3 0.2655022
## 45    12.5 0.2734092
## 46    12.6 0.2756469
## 52    13.2 0.2783585
## 53    13.2 0.2627195
## 55    13.2 0.2844036
## 56      NA 0.2956678
## 57      NA 0.3083147
## 58    12.4 0.2593112
## 62    12.3 0.2952431
## 63    12.5 0.2886563
## 64    12.5 0.2734092
## 71    13.2 0.2783585
## 72    13.2 0.2627195
## 75      NA 0.2956678
## 76      NA 0.3083147
## 78    12.5 0.2558008
## 79    12.3 0.2655022
## 80    12.3 0.2988009
```

```
## 81    12.3 0.2952431
## 86    12.7 0.2608332
## 91    13.2 0.2627195
## 94      NA 0.2956678
## 95      NA 0.3083147
## 96    12.4 0.2593112
## 97    12.5 0.2558008
## 99    12.3 0.2988009
## 100   12.3 0.2952431
## 101   12.5 0.2886563
## 104   12.6 0.2785387
## 105   12.7 0.2608332
## 106   12.8 0.2589872
## 107   13.1 0.2623017
## 109   13.2 0.2783585
## 111   13.2 0.2681700
## 112   13.2 0.2844036
## 113     NA 0.2956678
## 114     NA 0.3083147
## 115   12.4 0.2593112
## 131   13.2 0.2844036
## 132     NA 0.2956678
## 133     NA 0.3083147
## 139   12.5 0.2886563
## 151     NA 0.2956678
## 152     NA 0.3083147
## 154   12.5 0.2558008
## 155   12.3 0.2655022
## 156   12.3 0.2988009
## 158   12.5 0.2886563
## 159   12.5 0.2734092
## 160   12.6 0.2756469
## 164   13.1 0.2623017
## 168   13.2 0.2681700
## 169   13.2 0.2844036
## 170     NA 0.2956678
## 171     NA 0.3083147
```

```
#to visualize the missing data we use VIM
aggr(freetrade, prop = F, numbers = T)
```




```
#missingness beteen variables
marginplot(freetrade[,c("tariff","polity")]) #just using two variables to show the functioning
```




```
#b
#Using anova to check for relation between the missingness of tariff and country
aov(freetrade$tariff~freetrade$country,freetrade)
```

```
## Call:
##    aov(formula = freetrade$tariff ~ freetrade$country, data = freetrade)
##
## Terms:
##                 freetrade$country Residuals
## Sum of Squares           37348.55  13097.83
## Deg. of Freedom                 8       104
##
## Residual standard error: 11.22233
## Estimated effects may be unbalanced
## 58 observations deleted due to missingness
```

```
summary(aov(freetrade$tariff~freetrade$country,freetrade))
```

```
##                   Df Sum Sq Mean Sq F value Pr(>F)
## freetrade$country   8  37349    4669   37.07 <2e-16 ***
## Residuals         104  13098     126
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 58 observations deleted due to missingness
```

```
#removing of Nepal from the data
t1=freetrade[freetrade$country!="Nepal",]
aov(t1$tariff~t1$country,t1) # retrying the ANOVA test
```

```
## Call:
##    aov(formula = t1$tariff ~ t1$country, data = t1)
##
## Terms:
##                 t1$country Residuals
## Sum of Squares    35980.54  12995.36
## Deg. of Freedom          7        98
##
## Residual standard error: 11.51546
## Estimated effects may be unbalanced
## 46 observations deleted due to missingness
```

```
summary(aov(t1$tariff~t1$country,t1))
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## t1$country   7  35981    5140   38.76 <2e-16 ***
## Residuals   98  12995     133
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 46 observations deleted due to missingness
```

```
#removing of Philipines
t2=freetrade[freetrade$country!="Philipines",]
aov(t2$tariff~t2$country,t2) # retrying the ANOVA test
```

```
## Call:
##    aov(formula = t2$tariff ~ t2$country, data = t2)
##
## Terms:
##                 t2$country Residuals
## Sum of Squares    37348.55  13097.83
## Deg. of Freedom          8       104
##
## Residual standard error: 11.22233
## Estimated effects may be unbalanced
## 58 observations deleted due to missingness
```

```
summary(aov(t2$tariff~t2$country,t2))
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## t2$country    8  37349    4669   37.07 <2e-16 ***
## Residuals   104  13098     126
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 58 observations deleted due to missingness
```

Upon using the ANOVA test we see that the F value is high. Hence indicating no relation in the missingness of the two variables tariff and country. Upong removal of the Country Philipines the F value doesnt changes, while Nepals exclusion creates a change in the F value, stating that the variables can be related for their missingness if the countries are ommitted.