SI 618 Homework 8

Getting Data from SQLite Database (10 points)

In the data preparation step, a SQLite database has been created and populated with vehicle data. Now the data in the database is retrieved using R package DBI and RSQLite and stored in a data frame named vehicles. Here are the first 10 rows of the data frame, and the summary.

```
##    year       make                model          VClass cylinders displ
## 1  1985 Alfa Romeo  Spider Veloce 2000     Two Seaters         4   2.0
## 2  1985    Ferrari           Testarossa     Two Seaters        12   4.9
## 3  1985      Dodge              Charger Subcompact Cars         4   2.2
## 4  1985      Dodge B150/B250 Wagon 2WD            Vans         8   5.2
## 5  1993     Subaru    Legacy AWD Turbo    Compact Cars         4   2.2
## 6  1993     Subaru               Loyale    Compact Cars         4   1.8
## 7  1993     Subaru               Loyale    Compact Cars         4   1.8
## 8  1993     Toyota              Corolla    Compact Cars         4   1.6
## 9  1993     Toyota              Corolla    Compact Cars         4   1.6
## 10 1993     Toyota              Corolla    Compact Cars         4   1.8
##               trany city08 highway08 comb08
## 1     Manual 5-spd     19        25     21
## 2     Manual 5-spd      9        14     11
## 3     Manual 5-spd     23        33     27
## 4  Automatic 3-spd     10        12     11
## 5     Manual 5-spd     17        23     19
## 6  Automatic 3-spd     21        24     22
## 7     Manual 5-spd     22        29     25
## 8  Automatic 3-spd     23        26     24
## 9     Manual 5-spd     23        31     26
## 10 Automatic 4-spd     23        30     25
```

Converting to Factors (10 points)

To make downstream analysis easier, we convert the data in columns vehicles$make, vehicles$VClass, vehicles$cylinders, and vehicles$trany into factors. Here is the summary of the data frame after the conversion.

```
##       year             make          model
##  Min.   :1984   Chevrolet: 3635   Length:35719
##  1st Qu.:1990   Ford     : 2958   Class :character
##  Median :1999   Dodge    : 2465   Mode  :character
##  Mean   :1999   GMC      : 2306
##  3rd Qu.:2008   Toyota   : 1821
##  Max.   :2016   BMW      : 1518
##                 (Other)  :21016
##                           VClass        cylinders         displ
##  Compact Cars                : 5160   4      :13596   Min.   :0.600
##  Subcompact Cars             : 4643   6      :12522   1st Qu.:2.200
##  Midsize Cars                : 4035   8      : 7938   Median :3.000
##  Standard Pickup Trucks      : 2354   5      :  759   Mean   :3.328
##  Sport Utility Vehicle - 4WD: 2090   12     :  505   3rd Qu.:4.300
##  Two Seaters                 : 1734   3      :  195   Max.   :8.400
##  (Other)                     :15703   (Other):  204
```

```
##              trany           city08        highway08         comb08
##  Automatic 4-spd:11035   Min.   : 6.00   Min.   : 9.00   Min.   : 7.00
##  Manual 5-spd   : 8252   1st Qu.:15.00   1st Qu.:20.00   1st Qu.:16.00
##  Automatic 3-spd: 3151   Median :17.00   Median :23.00   Median :19.00
##  Manual 6-spd   : 2206   Mean   :17.54   Mean   :23.68   Mean   :19.79
##  Automatic (S6) : 2201   3rd Qu.:20.00   3rd Qu.:27.00   3rd Qu.:22.00
##  Automatic 5-spd: 2179   Max.   :53.00   Max.   :61.00   Max.   :53.00
##  (Other)        : 6695
```

Filter Down Data (30 points)

We will filter down the data such that only 'VClass' with more than 40 vehicles are kept. Here is the summary of the data frame after this subsetting step.

```
library(data.table)
vehicles = as.data.table(vehicles)

library(dplyr)

tt <- table(vehicles$VClass)
vehicles_subset <- subset(vehicles, (VClass %in% names(tt[tt>40])))

summary(vehicles_subset)
```

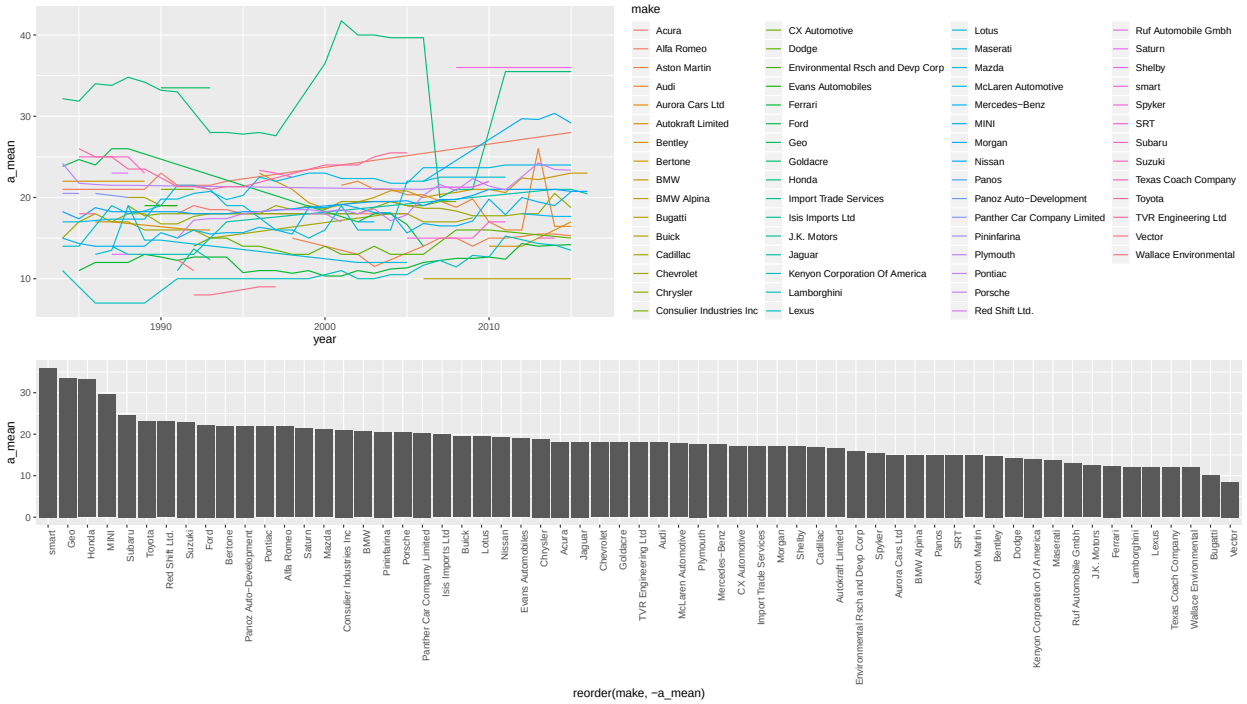
```
##       year             make          model
##  Min.   :1984   Chevrolet: 3633   Length:35708
##  1st Qu.:1990   Ford     : 2958   Class :character
##  Median :1999   Dodge    : 2465   Mode  :character
##  Mean   :1999   GMC      : 2302
##  3rd Qu.:2008   Toyota   : 1821
##  Max.   :2016   BMW      : 1518
##                 (Other)  :21011
##                            VClass        cylinders         displ
##  Compact Cars                 : 5160   4      :13594   Min.   :0.600
##  Subcompact Cars              : 4643   6      :12518   1st Qu.:2.200
##  Midsize Cars                 : 4035   8      : 7933   Median :3.000
##  Standard Pickup Trucks       : 2354   5      :  759   Mean   :3.328
##  Sport Utility Vehicle - 4WD: 2090   12     :  505   3rd Qu.:4.300
##  Two Seaters                  : 1734   3      :  195   Max.   :8.400
##  (Other)                      :15692   (Other):  204
##              trany           city08        highway08         comb08
##  Automatic 4-spd:11026   Min.   : 6.00   Min.   : 9.00   Min.   : 7.00
##  Manual 5-spd   : 8250   1st Qu.:15.00   1st Qu.:20.00   1st Qu.:16.00
##  Automatic 3-spd: 3151   Median :17.00   Median :23.00   Median :19.00
##  Manual 6-spd   : 2206   Mean   :17.54   Mean   :23.68   Mean   :19.79
##  Automatic (S6) : 2201   3rd Qu.:20.00   3rd Qu.:27.00   3rd Qu.:22.00
##  Automatic 5-spd: 2179   Max.   :53.00   Max.   :61.00   Max.   :53.00
##  (Other)        : 6695
```

Fuel Economy of Vehicles of Different Makes (50 points)

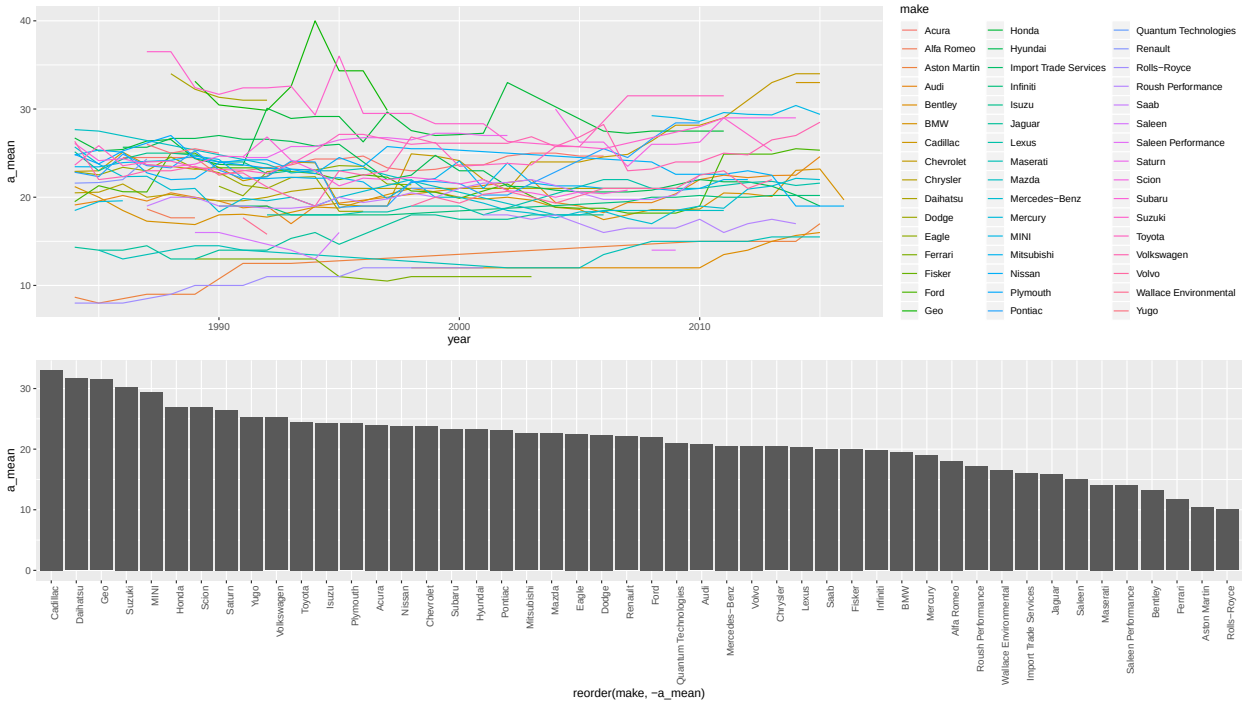
For each vehicle class in filtered down data, we plot the mean combined MPG (average of data in vehicles$comb08) for each vehicle maker every year. And then, we compute the mean combined MPG in all

years for each vehicle maker, and plot it. Both charts are created with ggplot(). Note how the vehicle makers are ranked in the second plot. Use **fig.width=16**. To suppress messages from ggplot regarding groups with only one observation, set **warning=FALSE, message=FALSE** (we recommend setting this option only once your code is complete).

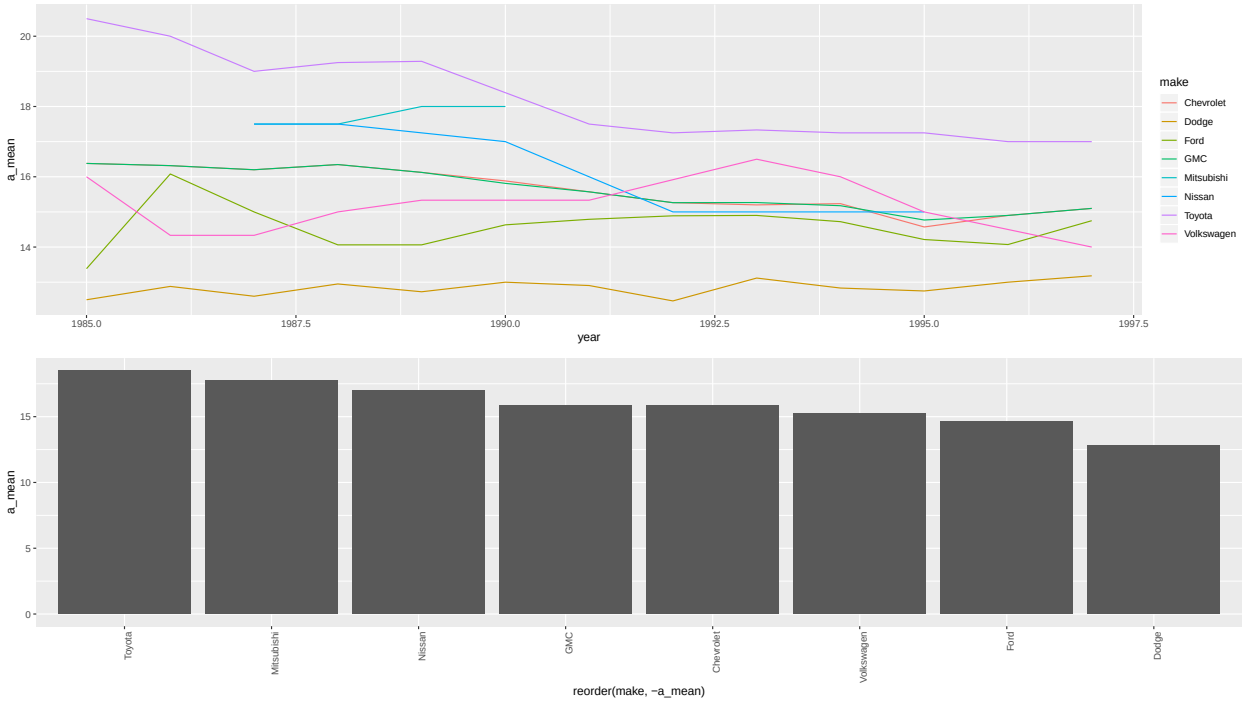
```
x <- vehicles_subset[, "VClass"]
x <- unique(x)
class_list <- as.list(x)

graphing <- function(name1) {
  test <- vehicles_subset[VClass == name1]

  test1 <- test %>% group_by(year, make) %>% summarise(a_mean=(mean(comb08)))
  one <-ggplot(test1,aes(x=year, y=a_mean, color = make))+geom_line()

  test2 <- test %>% group_by(make) %>% summarise(a_mean=(mean(comb08)))


  test2 <- test2[c("make", "a_mean")]
  test2
  two <- ggplot(data=test2, aes(x=reorder(make, -a_mean), y=a_mean)) +
geom_bar(stat="identity") + theme(axis.text.x = element_text(angle = 90, hjust = 1))

  print(one)
  print(two)
}

graphing("Two Seaters")
```



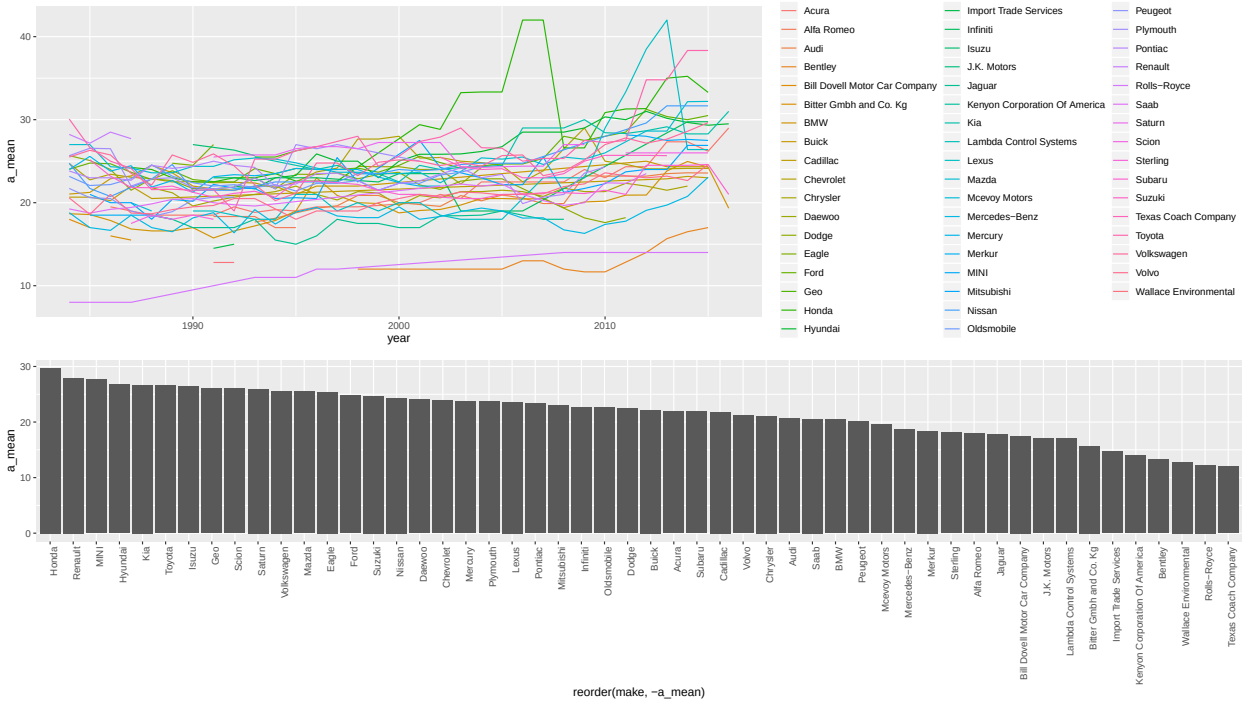
```
graphing("Subcompact Cars")
```



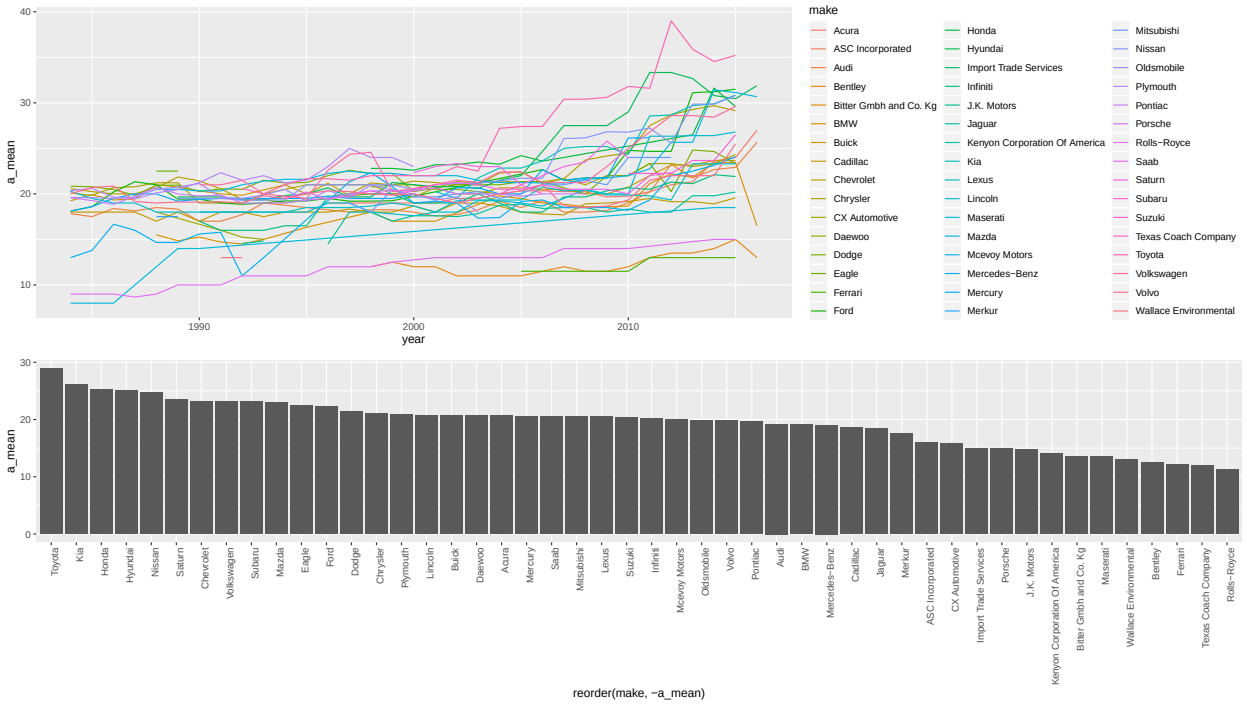

```
graphing("Vans")
```



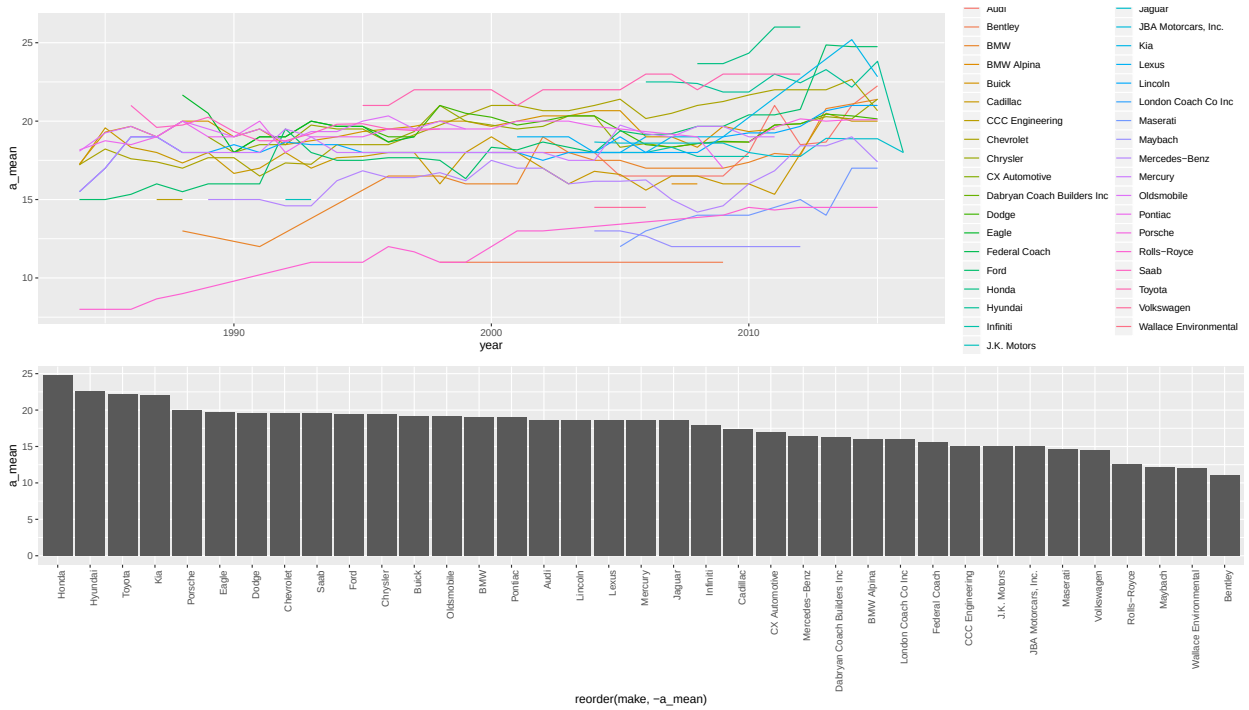

```
graphing("Compact Cars")
```



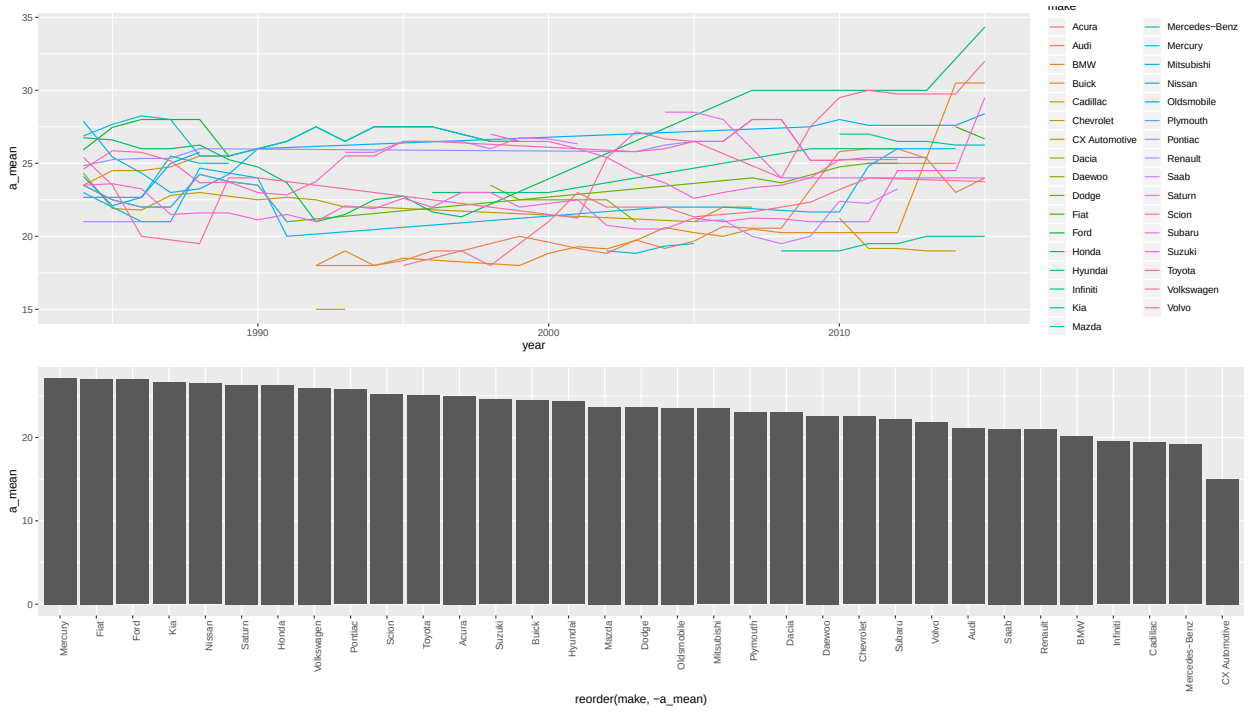

```
graphing("Midsize Cars")
```



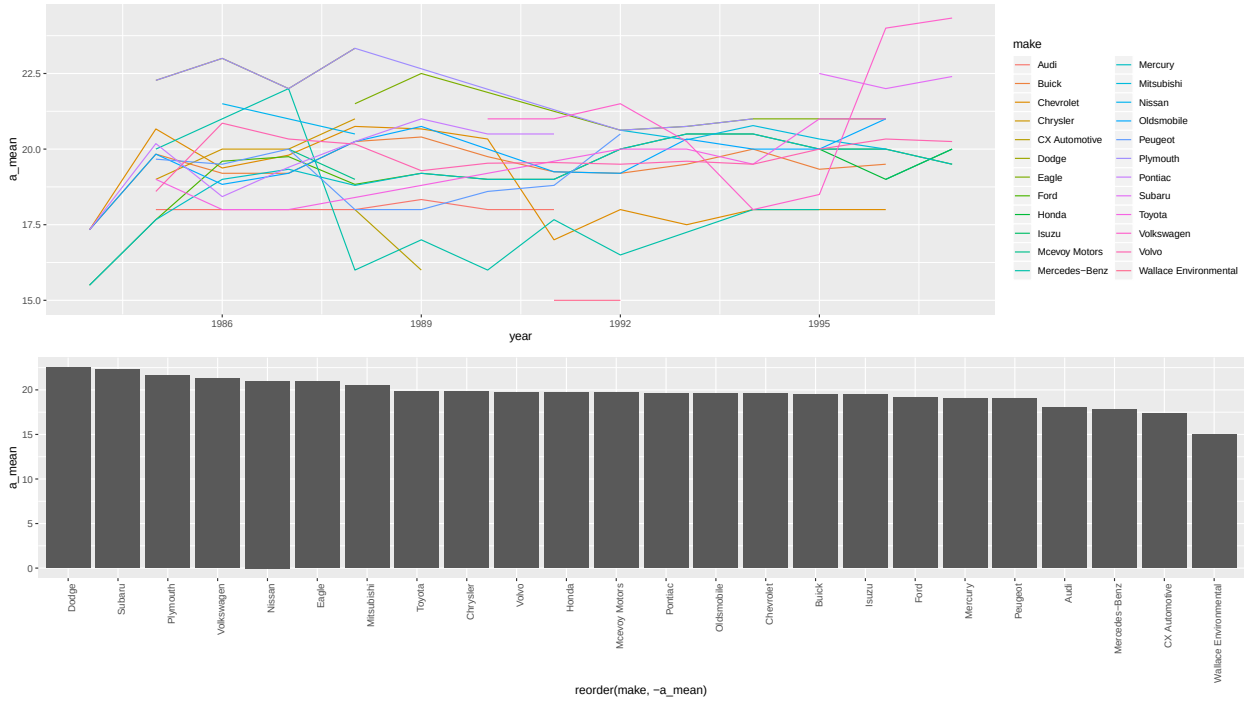

```
graphing("Large Cars")
```



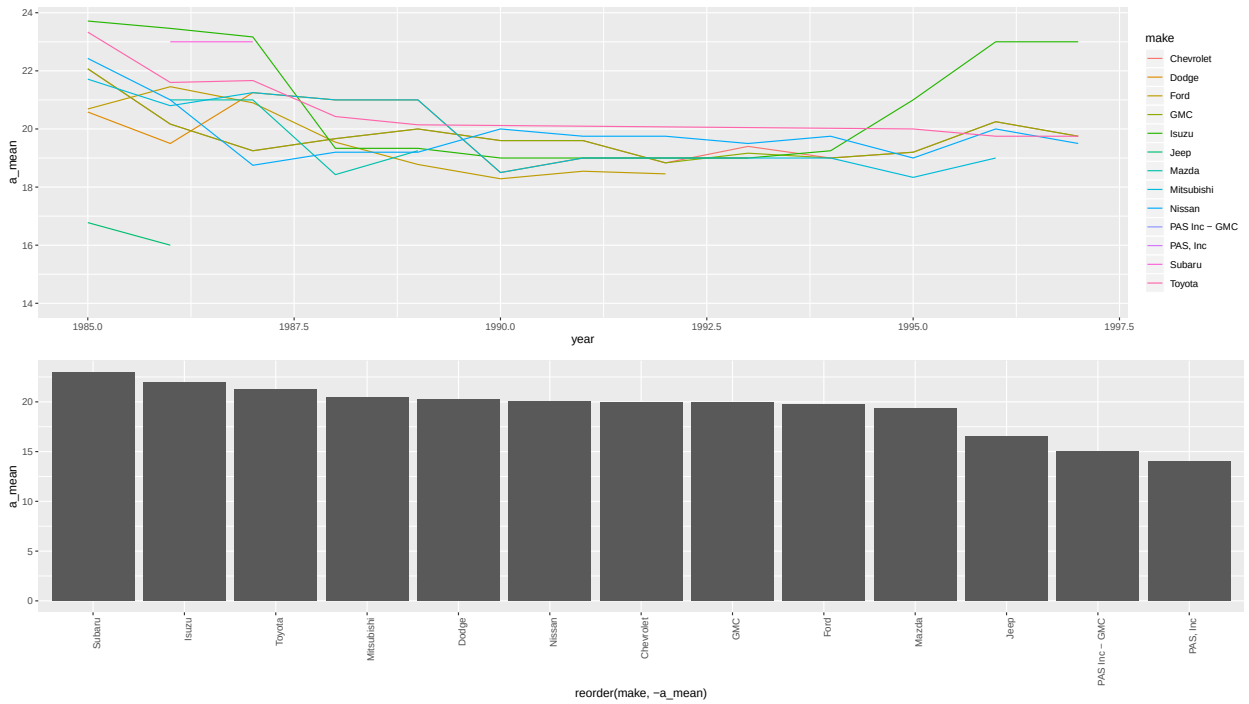

```
graphing("Small Station Wagons")
```



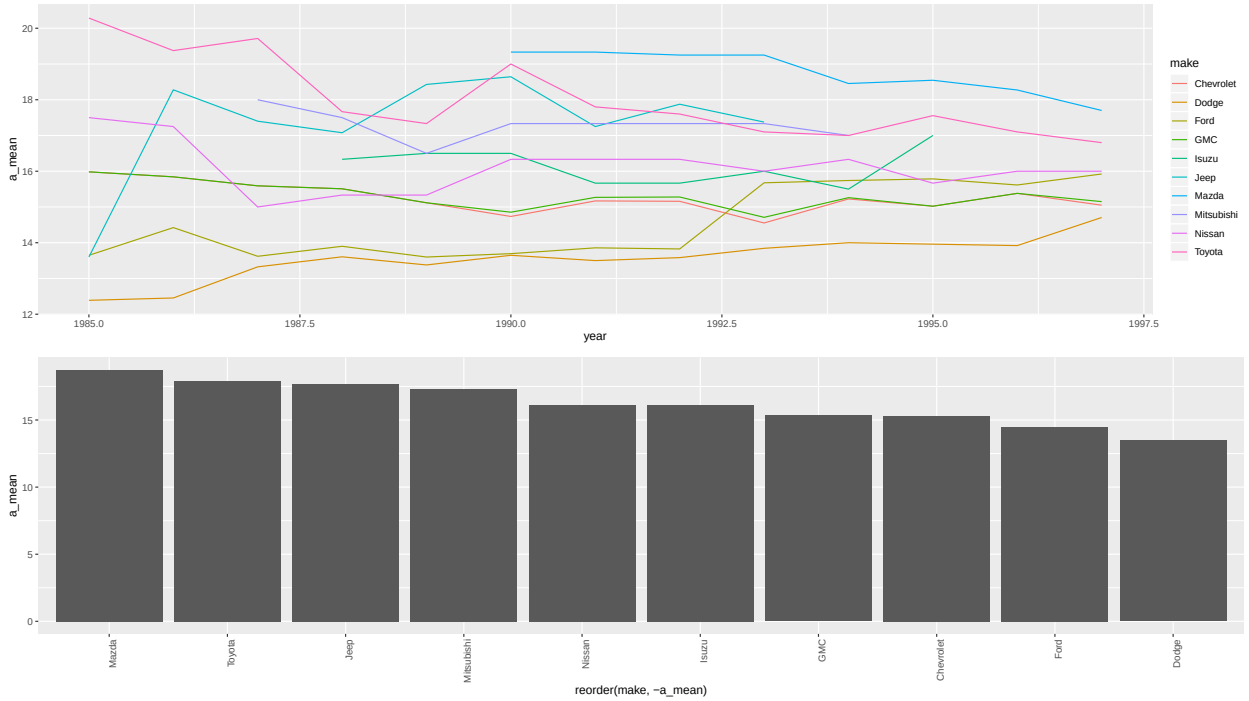

```
graphing("Midsize-Large Station Wagons")
```



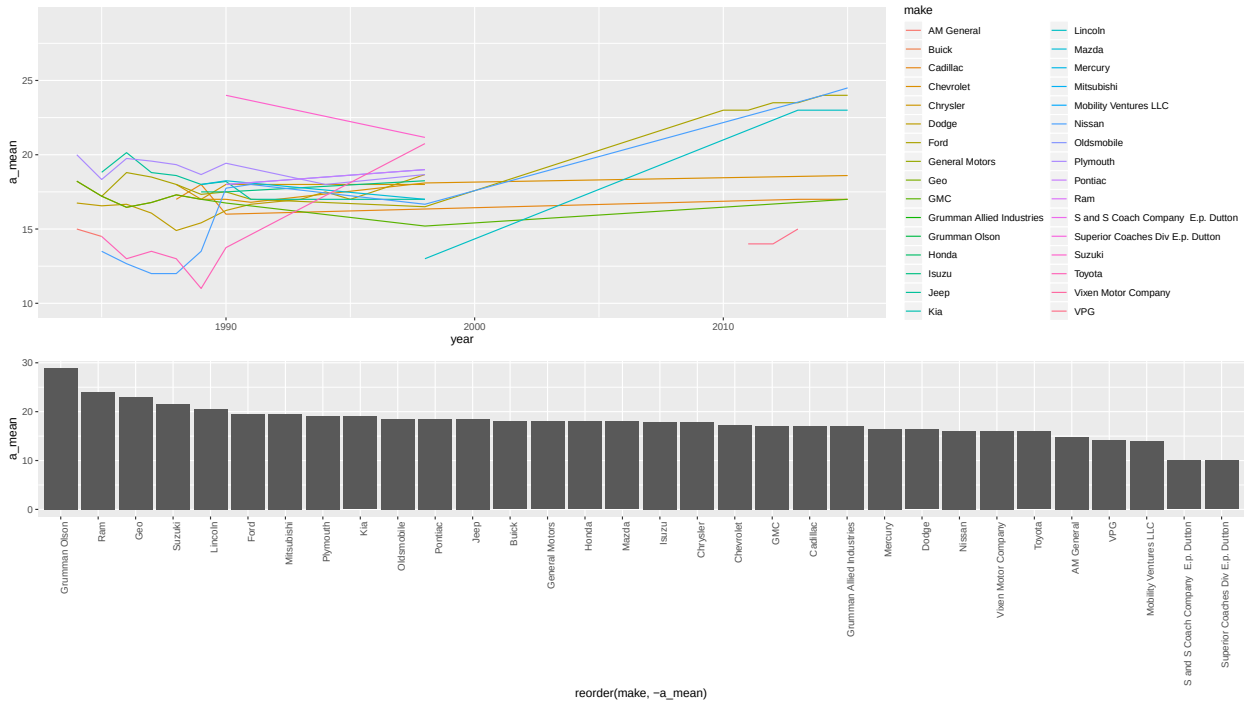

```
graphing("Small Pickup Trucks")
```



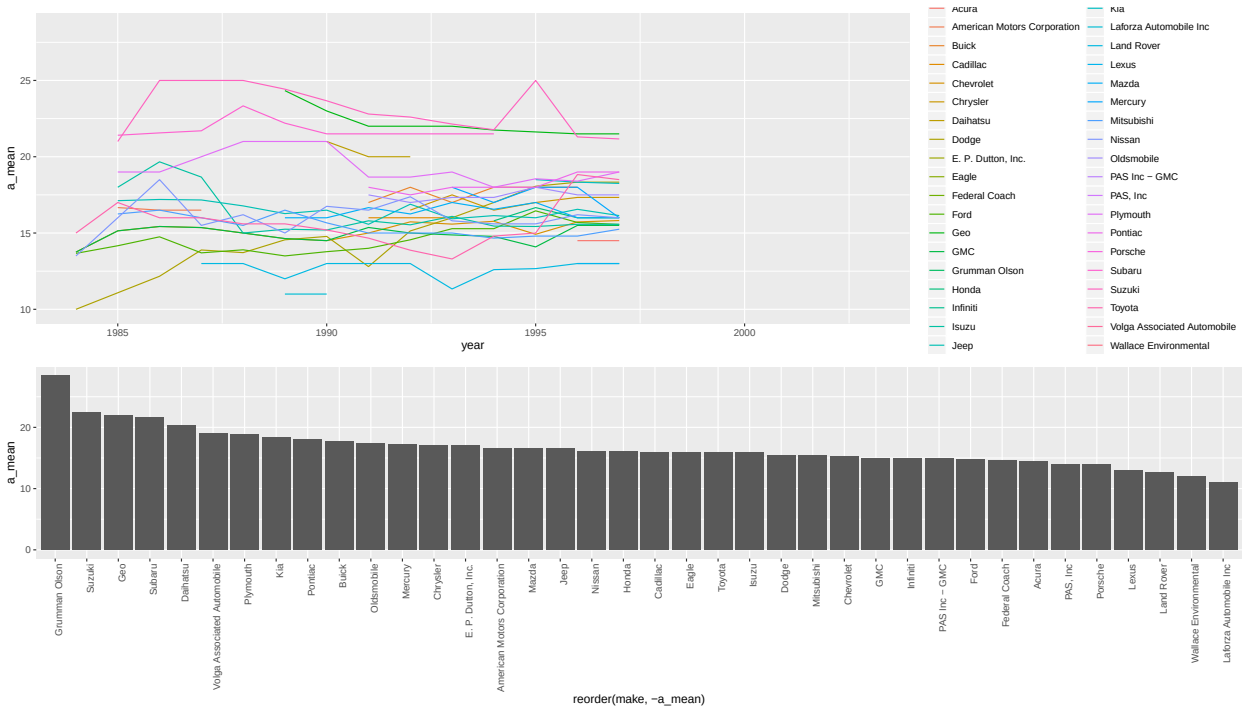

```
graphing("Standard Pickup Trucks")
```



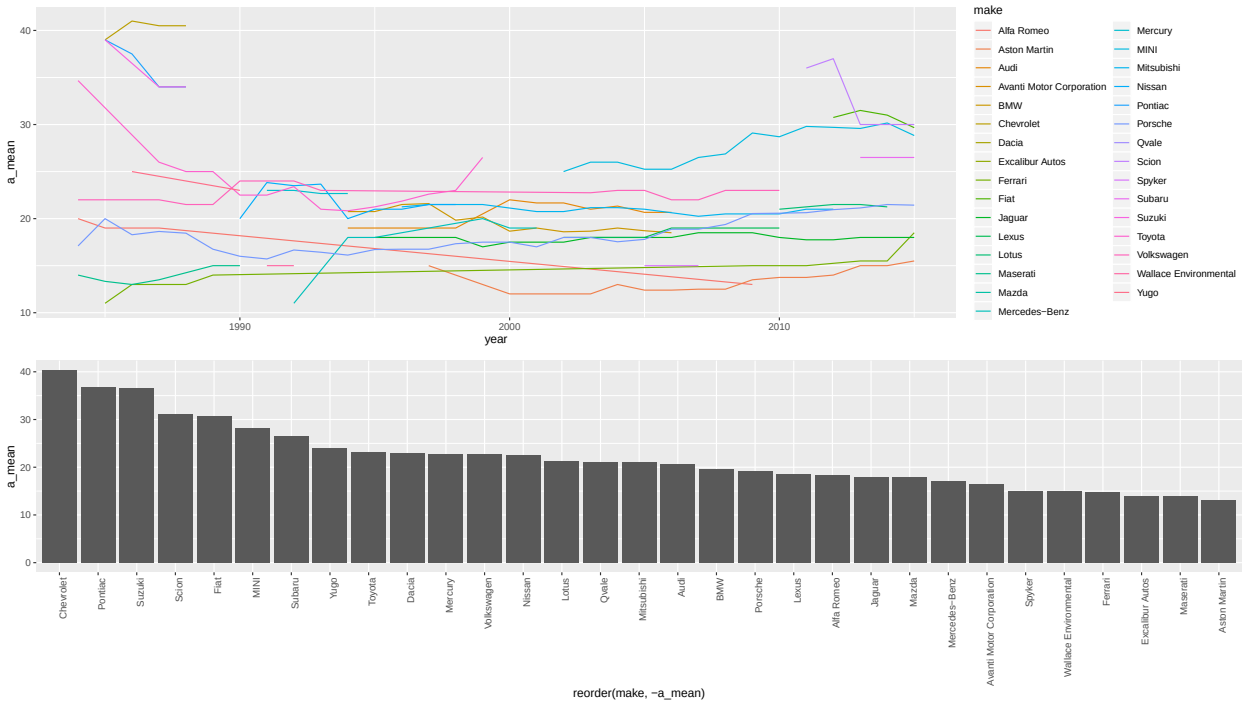

```
graphing("Special Purpose Vehicle 2WD")
```



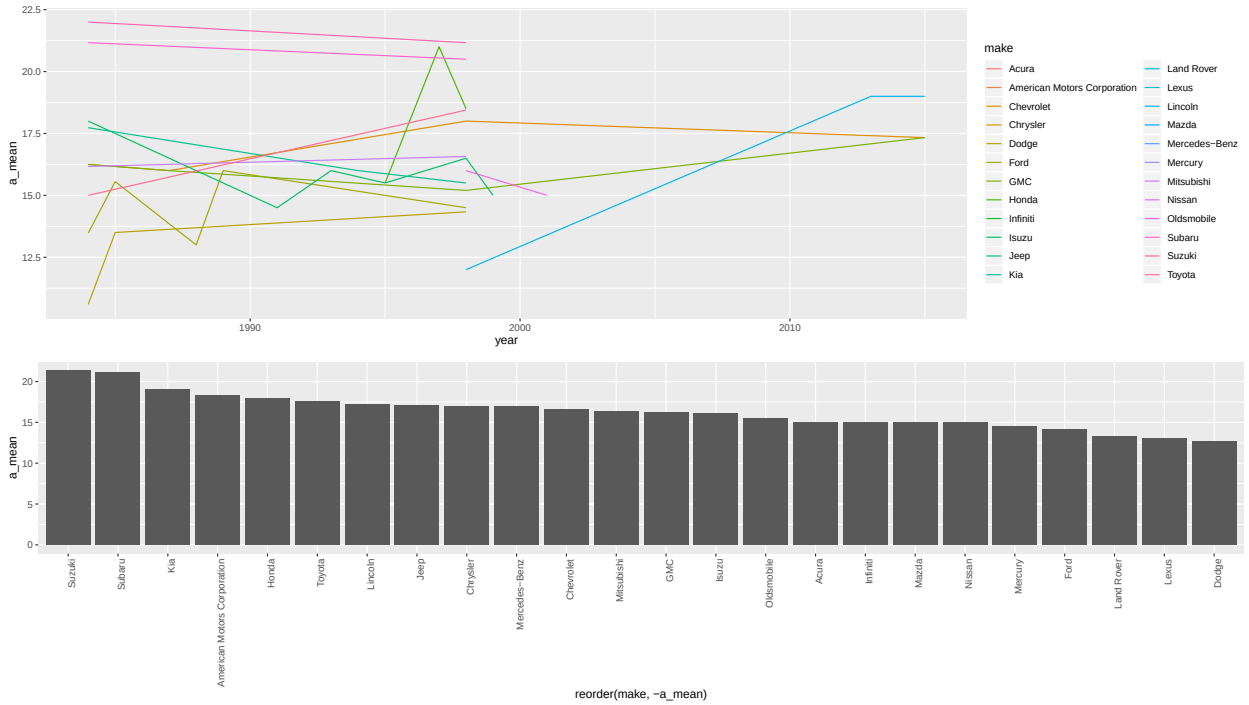

```
graphing("Special Purpose Vehicles")
```



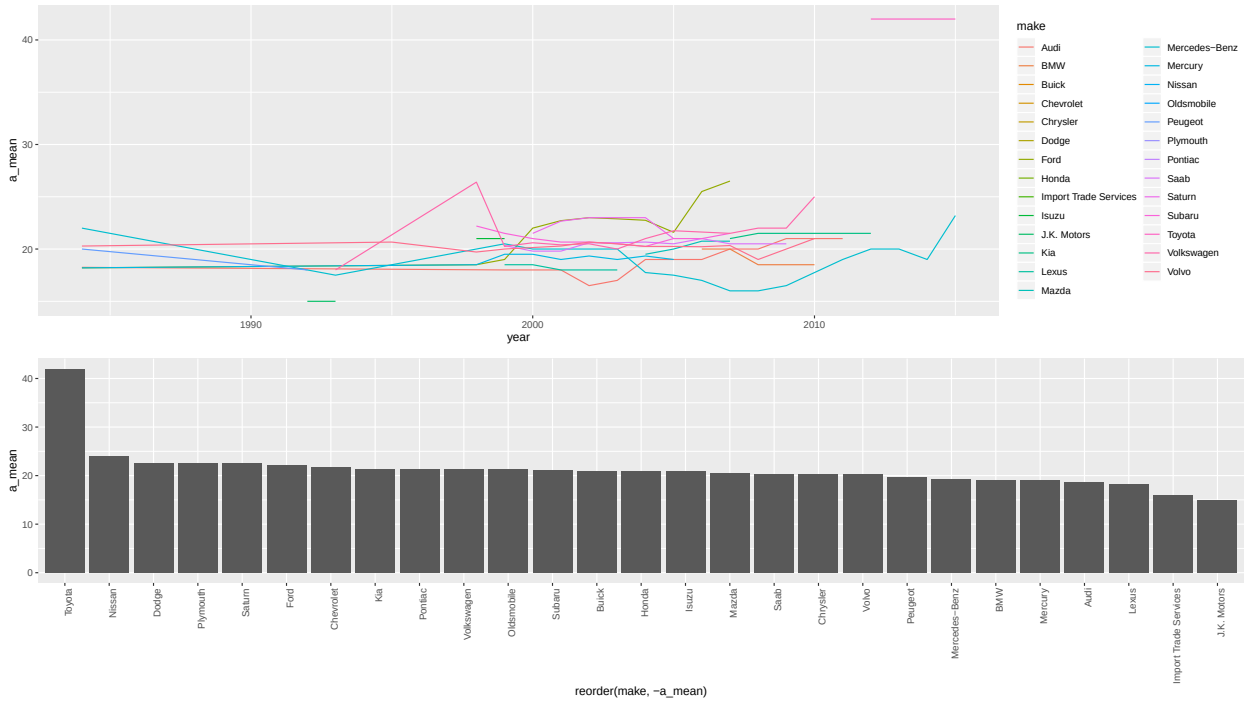

```
graphing("Minicompact Cars")
```



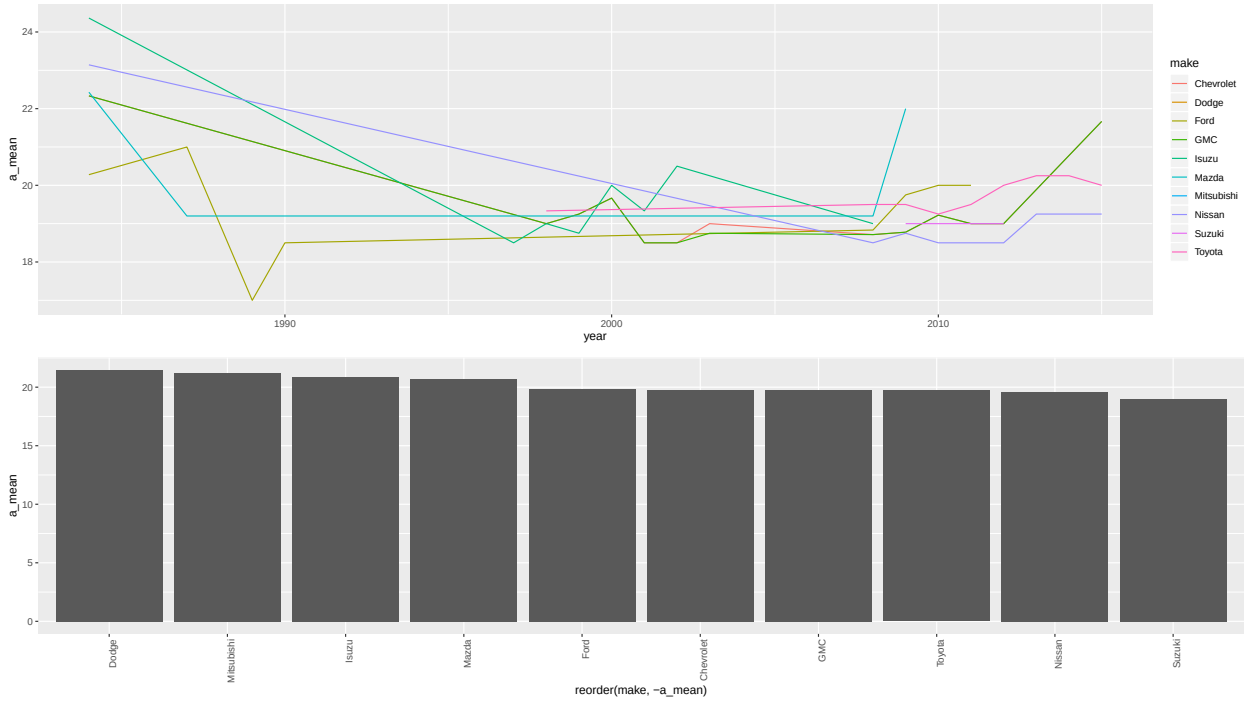

```
graphing("Special Purpose Vehicle 4WD")
```



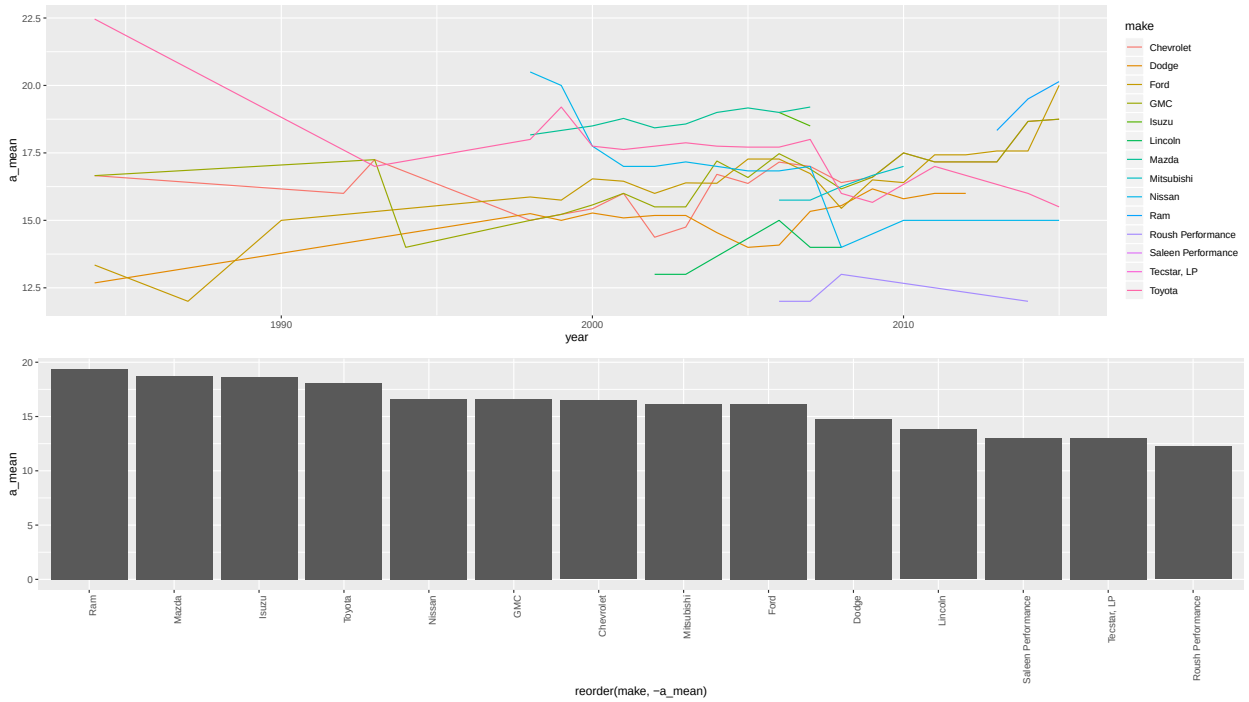

```
graphing("Midsize Station Wagons")
```



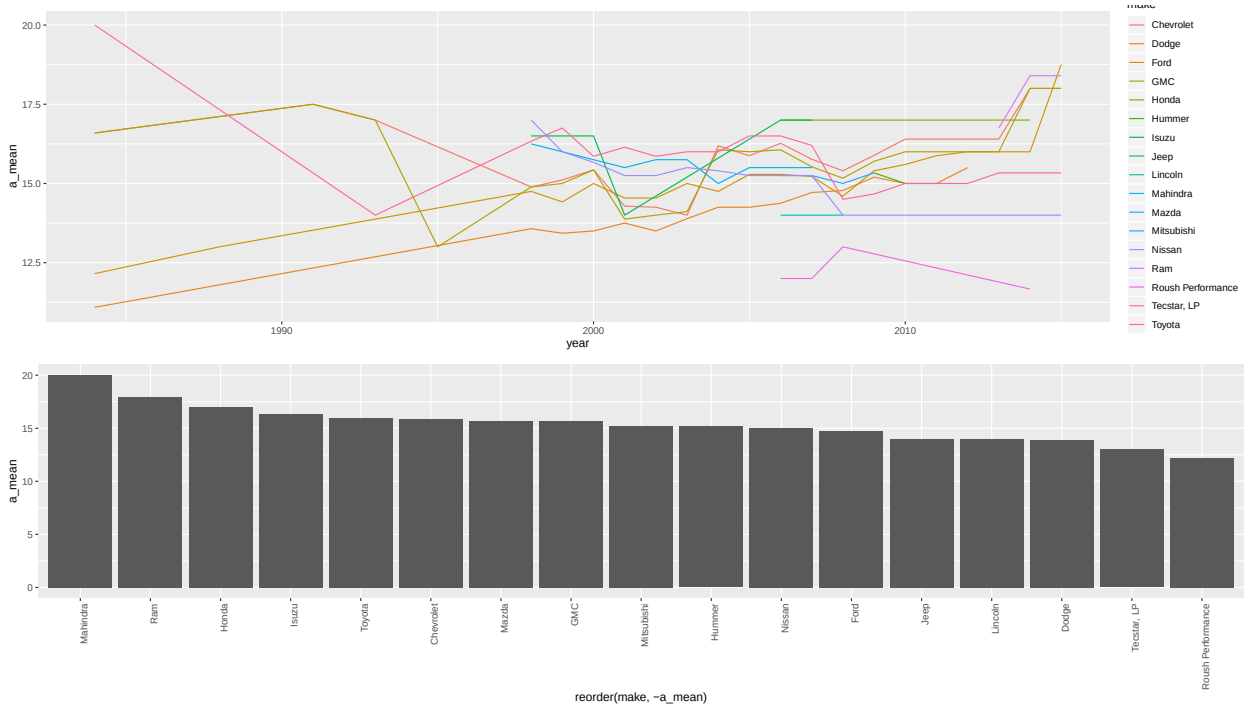

```
graphing("Small Pickup Trucks 2WD")
```



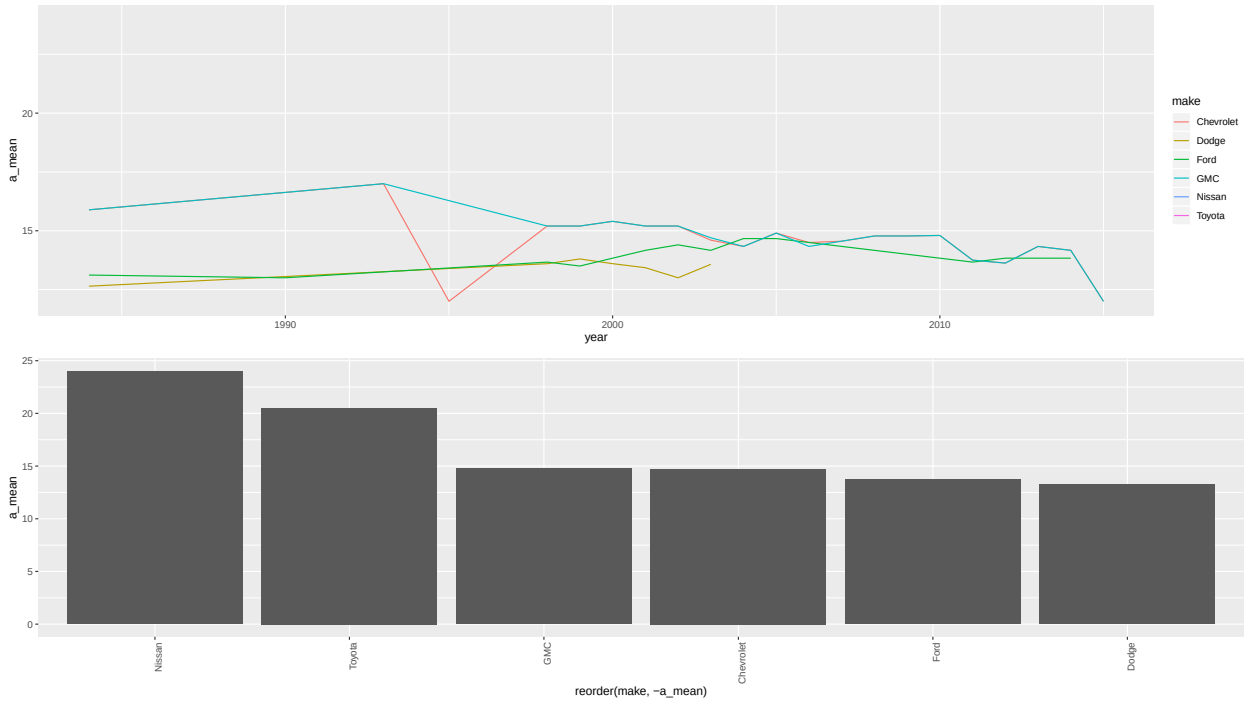

```
graphing("Standard Pickup Trucks 2WD")
```



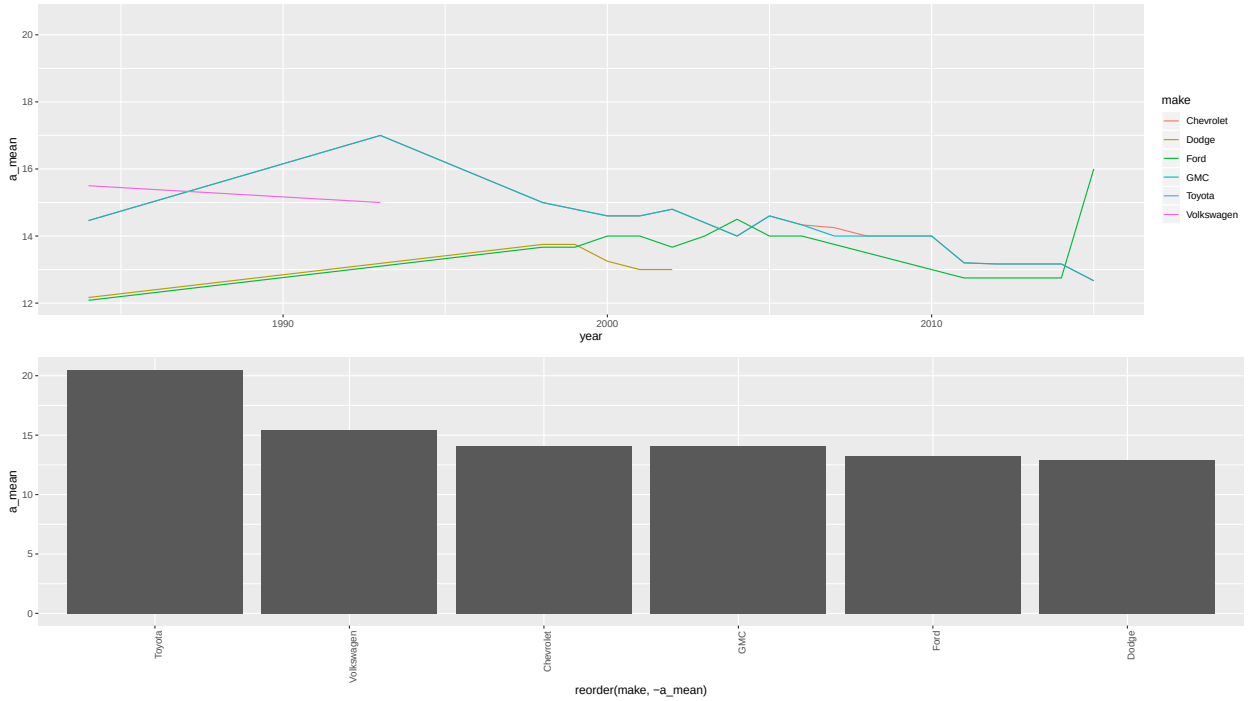

```
graphing("Standard Pickup Trucks 4WD")
```



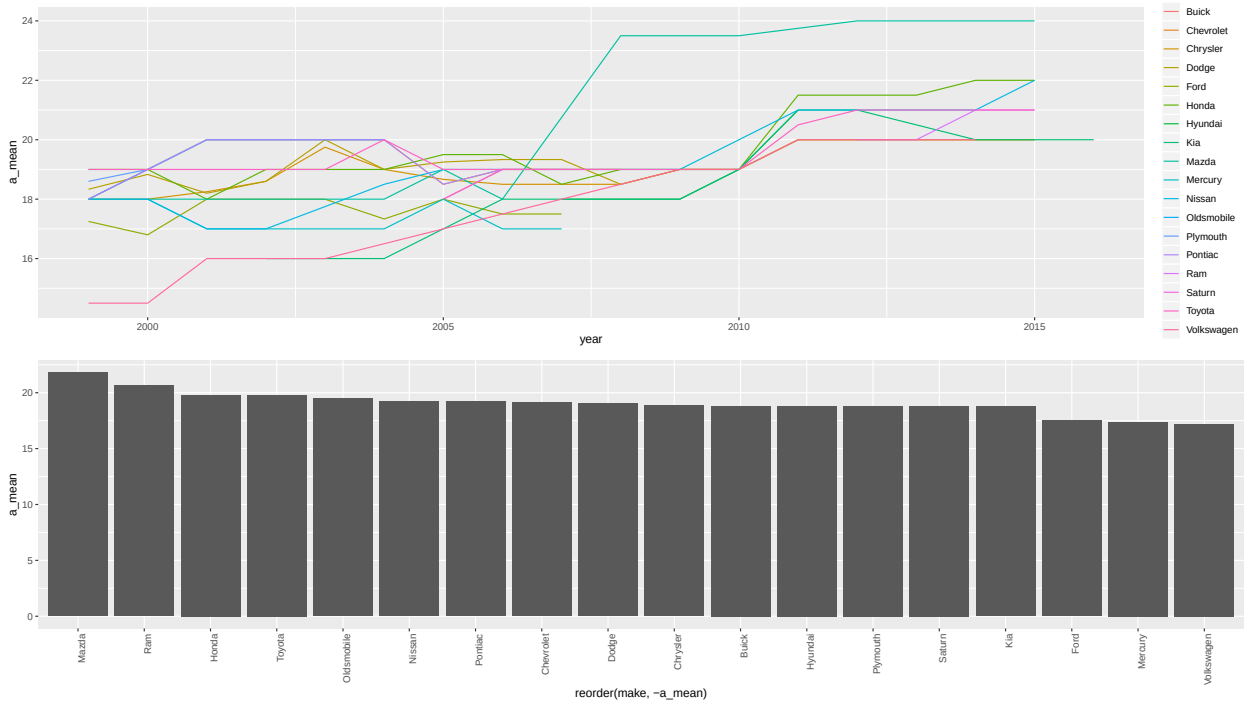

```
graphing("Vans, Cargo Type")
```



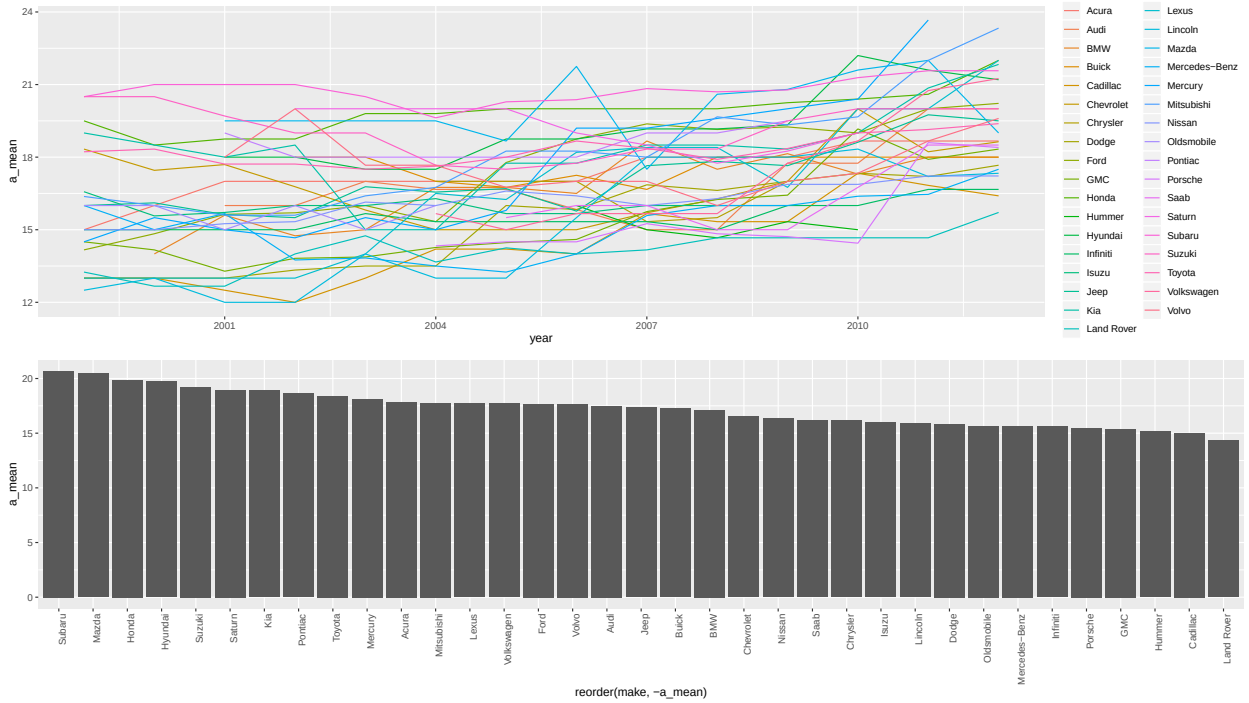

```
graphing("Vans, Passenger Type")
```



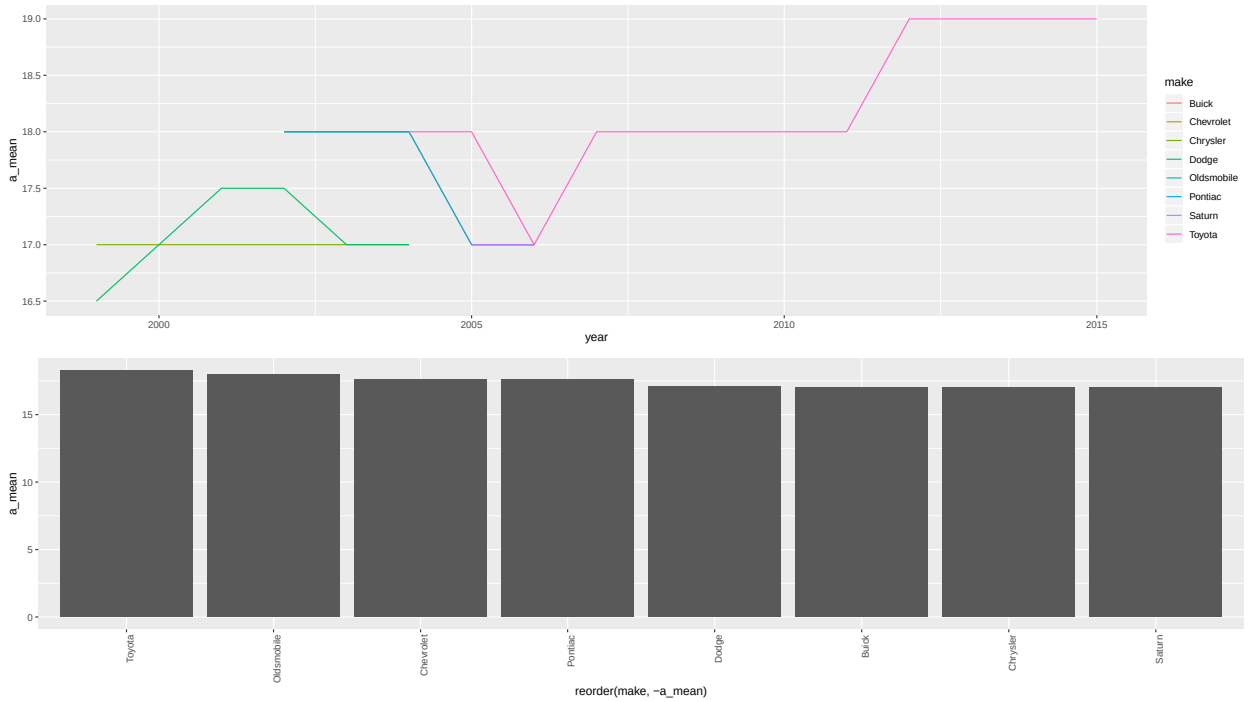

```
graphing("Minivan - 2WD")
```



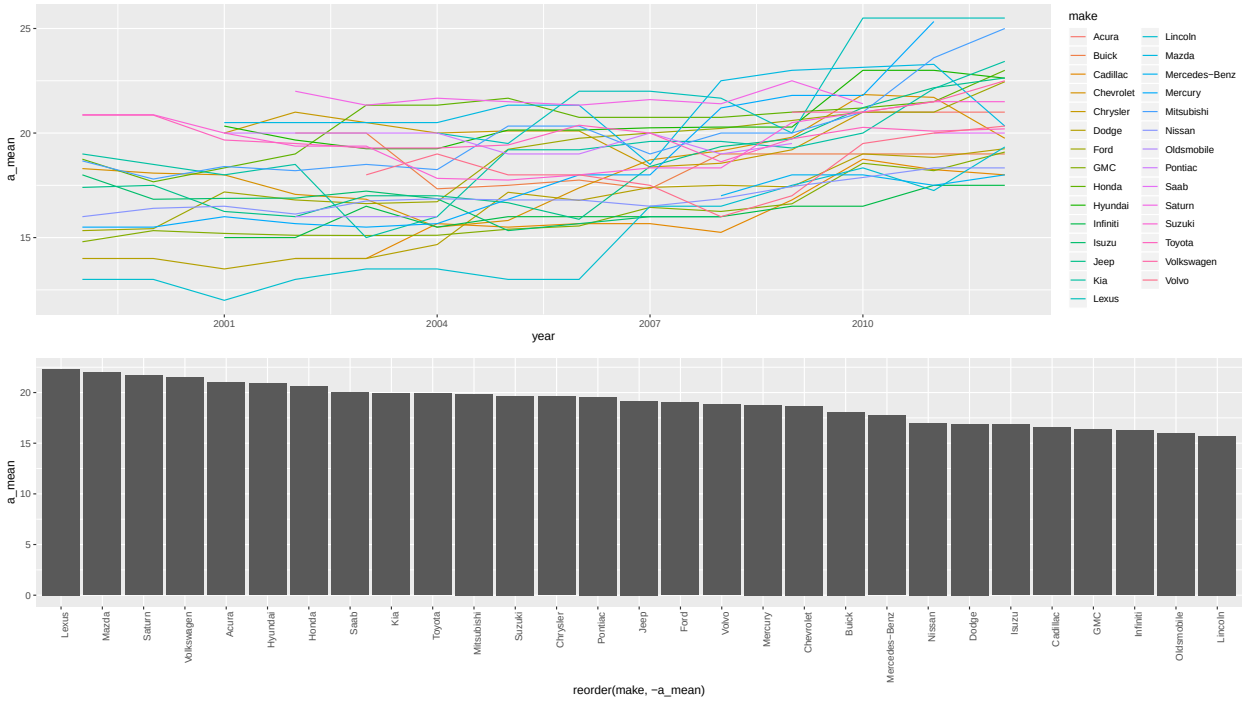

```
graphing("Sport Utility Vehicle - 4WD")
```



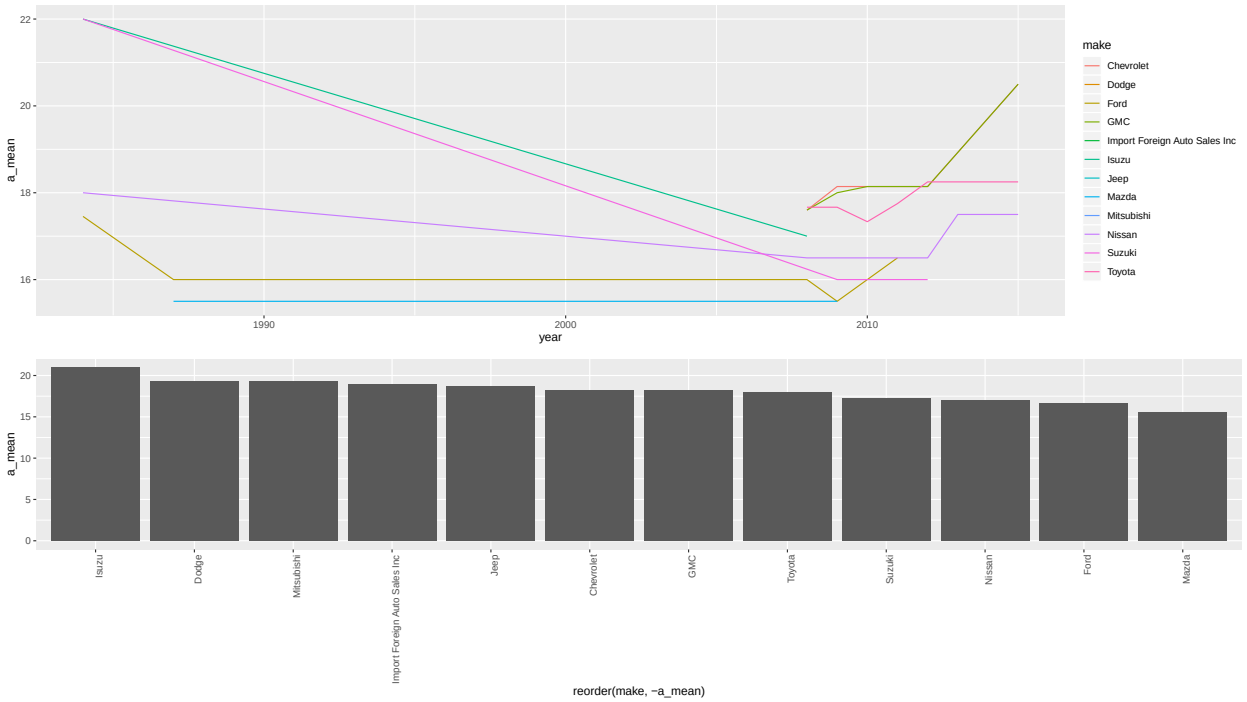

```
graphing("Minivan - 4WD")
```



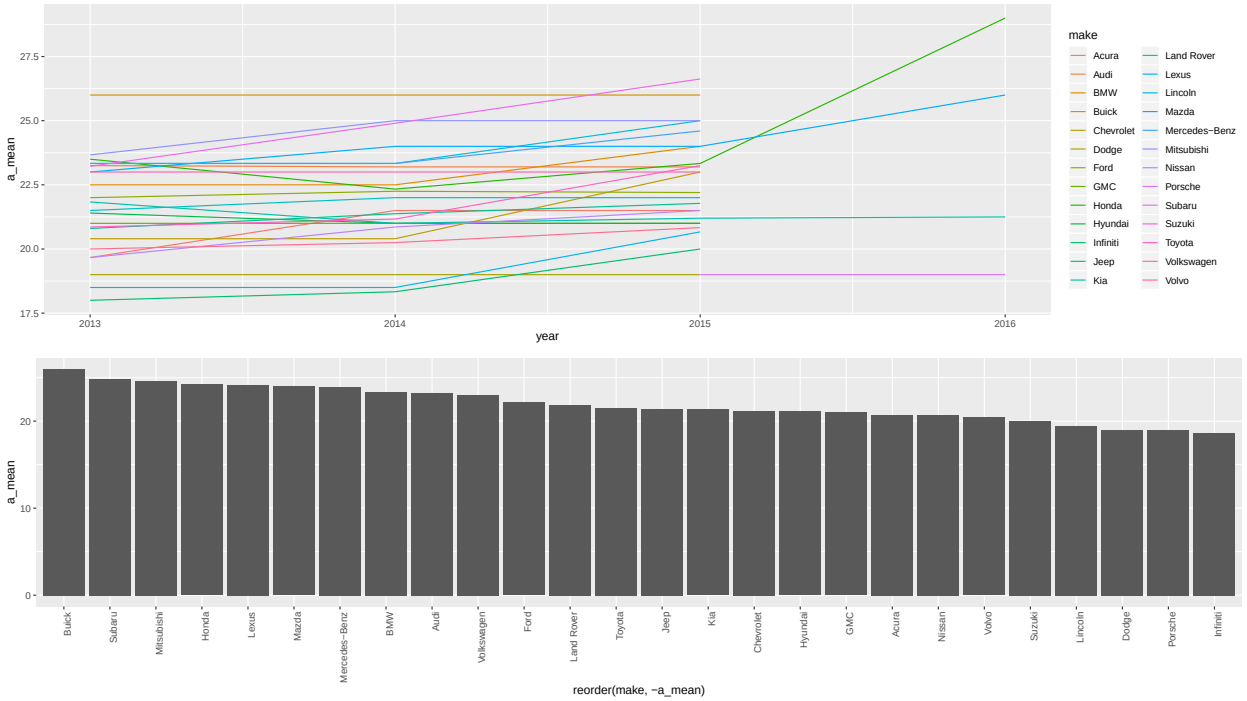

```
graphing("Sport Utility Vehicle - 2WD")
```



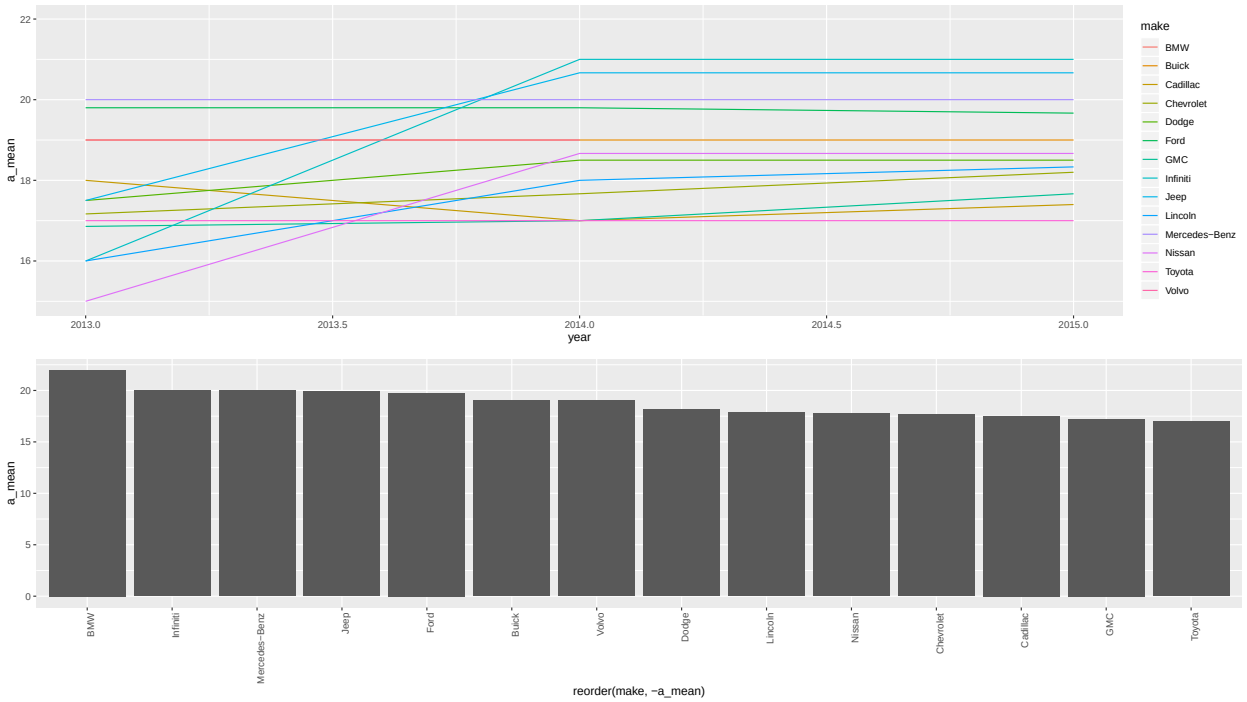

```
graphing("Small Pickup Trucks 4WD")
```



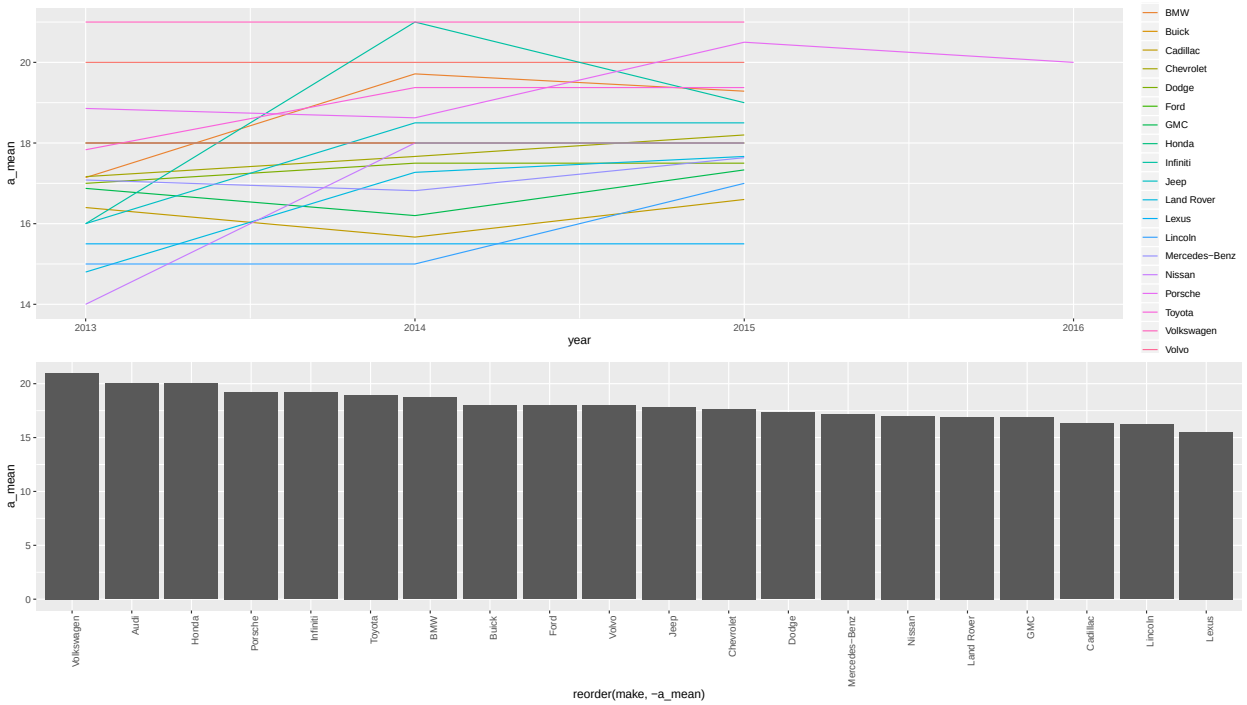

```
graphing("Small Sport Utility Vehicle 4WD")
```




```
graphing("Standard Sport Utility Vehicle 2WD")
```




```
graphing("Standard Sport Utility Vehicle 4WD")
```




```
graphing("Small Sport Utility Vehicle 2WD")
```