7. The data file **vehicles.csv** contains the fuel economy data from the EPA, for all of the unique vehicles (vehicles with a unique EPA identifier) that were available from 1985 until 2015. These data are also available from the *fueleconomy* R package, in the **vehicles** dataframe, which can be installed in R using the following R command, *install.packages("fueleconomy")*. The data set contains the twelve variables.

variable	description
id	Unique EPA identifier
make	Manufacturer
model	Model name
year	Model year
class	EPA vehicle size class
trans	Transmission
drive	Drive train
cyl	Number of cylinders
displ	Engine displacement, in litres
fuel	Fuel type
hwy	Highway fuel economy, in mpg
cty	City fuel economy, in mpg

Code to load R packages and save the data to a .csv file.

```r
library(pacman)
p_load(tidyverse, fueleconomy)

write_csv(vehicles, "vehicles.csv")
```

(a) How many unique vehicles (vehicles with *id* values that are unique) were available in the years 1985-2015?

```r
vehicles %>% distinct(id) %>% count()
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1 33442
```

(b) Create a subset of the dataset including *2 wheel drive minivans* for the following 6 manufacturers, *Chrysler*, *Dodge*, *Honda*, *Kia*, *Nissan*, and *Toyota*. How many unique *2 wheel drive minivans* were available in the years 1985-2015?

```r
minivan_2wd <- vehicles %>% filter( make %in% c("Chrysler", "Dodge", "Honda",
                     "Kia", "Nissan", "Toyota") &
                     class == "Minivan - 2WD" )

minivan_2wd %>% count()
```

```
## # A tibble: 1 x 1
```

```
##         n
##     <int>
## 1    172
```

minivan_2wd

```
## # A tibble: 172 x 12
##        id make    model    year class trans  drive    cyl displ fuel     hwy   cty
##     <dbl> <chr>   <chr>   <dbl> <chr> <chr>  <chr>  <dbl> <dbl> <chr>   <dbl> <dbl>
##  1 26110 Chrys~ Town a~  2009 Mini~ Autom~ Front~     6   3.8 Regul~    23    16
##  2 26111 Chrys~ Town a~  2009 Mini~ Autom~ Front~     6   4   Regul~    25    17
##  3 26223 Chrys~ Town a~  2009 Mini~ Autom~ Front~     6   3.3 Gasol~    24    17
##  4 28725 Chrys~ Town a~  2010 Mini~ Autom~ Front~     6   4   Regul~    25    17
##  5 28726 Chrys~ Town a~  2010 Mini~ Autom~ Front~     6   3.3 Gasol~    24    17
##  6 28727 Chrys~ Town a~  2010 Mini~ Autom~ Front~     6   3.8 Regul~    23    16
##  7 31045 Chrys~ Town a~  2011 Mini~ Autom~ Front~     6   3.6 Gasol~    25    17
##  8 31461 Chrys~ Town a~  2012 Mini~ Autom~ Front~     6   3.6 Gasol~    25    17
##  9 32808 Chrys~ Town a~  2013 Mini~ Autom~ Front~     6   3.6 Gasol~    25    17
## 10 33959 Chrys~ Town a~  2014 Mini~ Autom~ Front~     6   3.6 Gasol~    25    17
## # ... with 162 more rows
```

(c) Make a table showing the number of unique minivans that were available from each manufacturer. Which manufacturer offered the most unique minivans? Which manufacturer has offered the least number of unique minivans?

```
minivan_2wd %>% group_by(make) %>%
  summarize(n = n()) %>%
  pivot_wider(names_from = make, values_from = n )
```
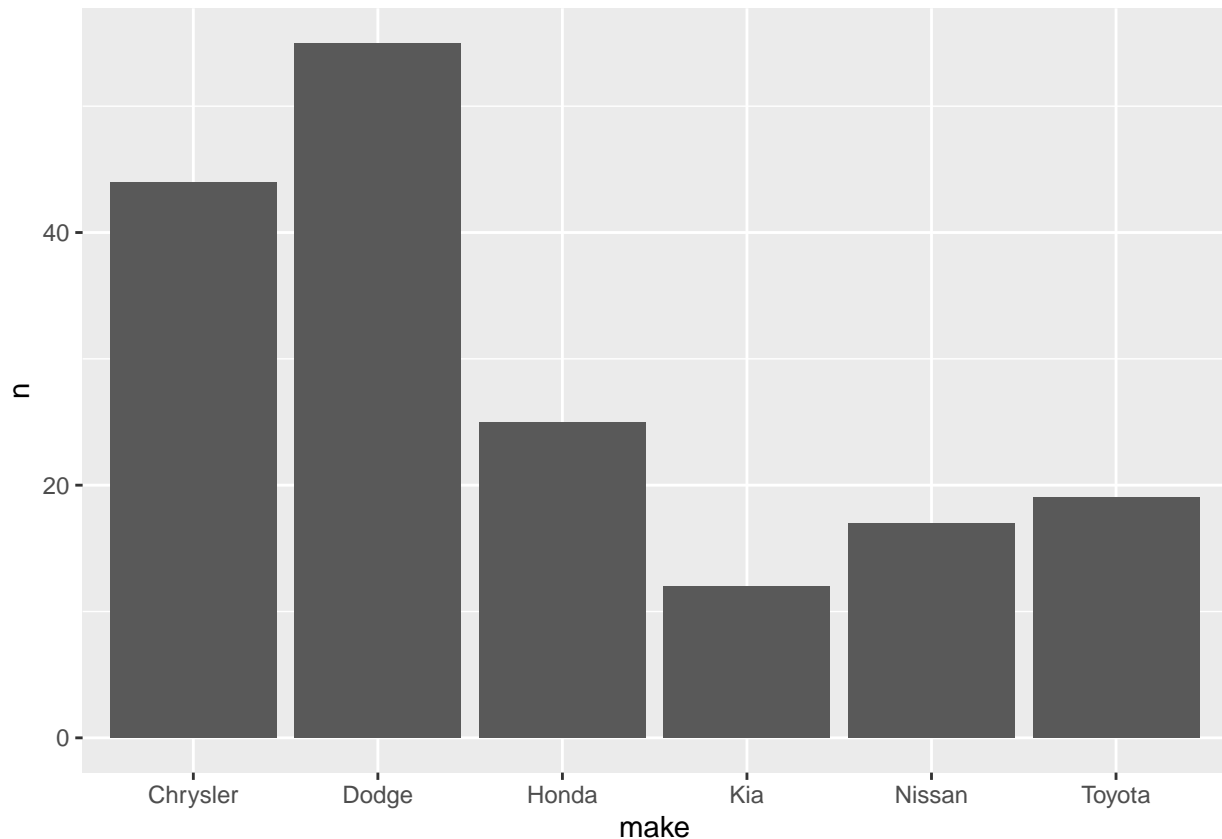
```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 1 x 6
##   Chrysler Dodge Honda   Kia Nissan Toyota
##      <int> <int> <int> <int>  <int>  <int>
## 1       44    55    25    12     17     19
```

(d) Make a bargraph showing the number of unique minivans from each manufacturer computed in the previous part.

```
minivan_2wd %>% group_by(make) %>%
  summarize(n = n()) %>%
  ggplot(aes(x = make, y = n)) +
    geom_col()
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

(e) Make a table showing the average highway miles per gallon (mpg), of the unique minivans that were available, from each manufacturer. Which manufacturer has the best miles per gallon? Which manufacturer has the worst miles per gallon?

```
minivan_2wd %>% group_by(make) %>%
  summarize(hwy_mean = round(mean(hwy))) %>%
  pivot_wider(names_from = make, values_from = hwy_mean )
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 1 x 6
##   Chrysler Dodge Honda   Kia Nissan Toyota
##      <dbl> <dbl> <dbl> <dbl>  <dbl>  <dbl>
## 1       23    24    25    22     23     24
```

(f) Make a bargraph showing the average miles per gallon values of unique minivans for each manufacturer computed in the previous part.

```
minivan_2wd %>% group_by(make) %>%
  summarize(hwy_mean = round(mean(hwy))) %>%
  ggplot(aes(x = make, y = hwy_mean)) +
    geom_col()
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```
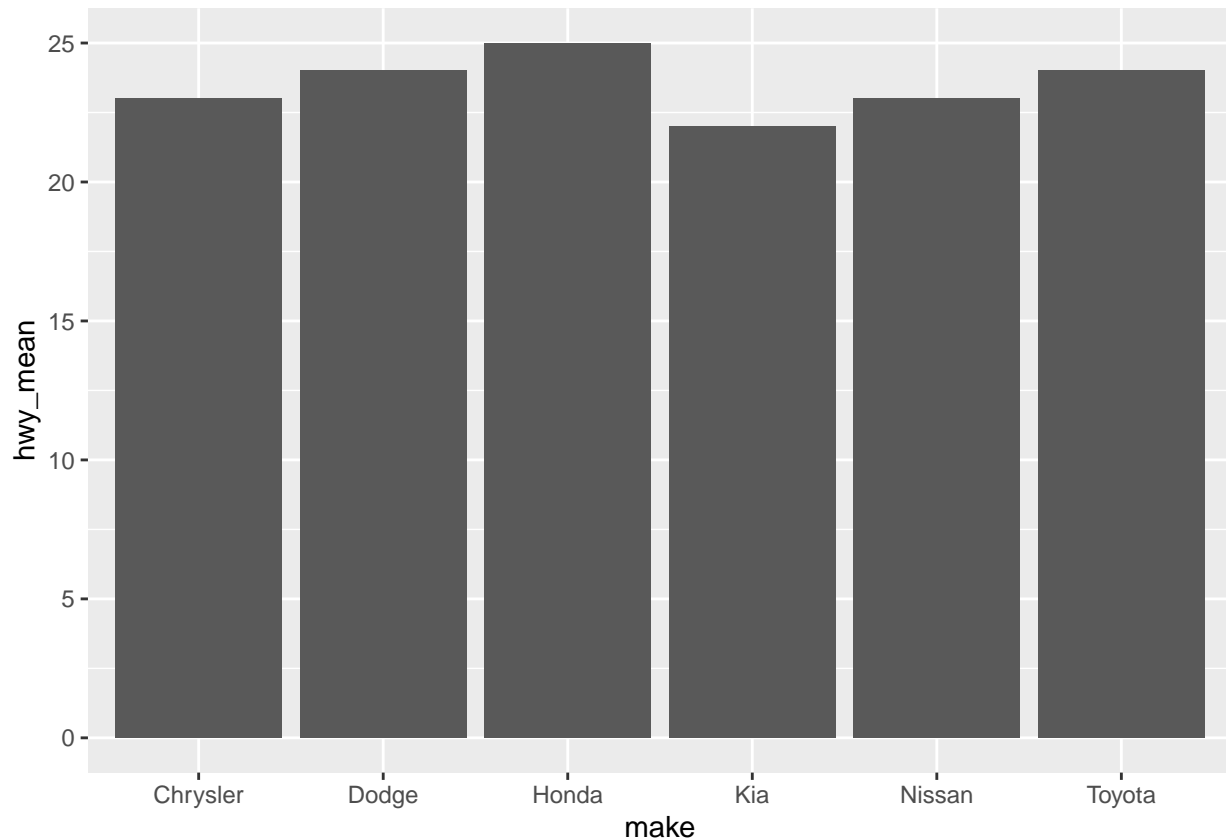
(g) Plot the highway miles per gallon versus model year of unique minivans faceted for each manufacturer, using the same scales. Include a linear regression smoother on each plot. Make the plots again adding color for the fuel type. Identify the outlier in the dataset.

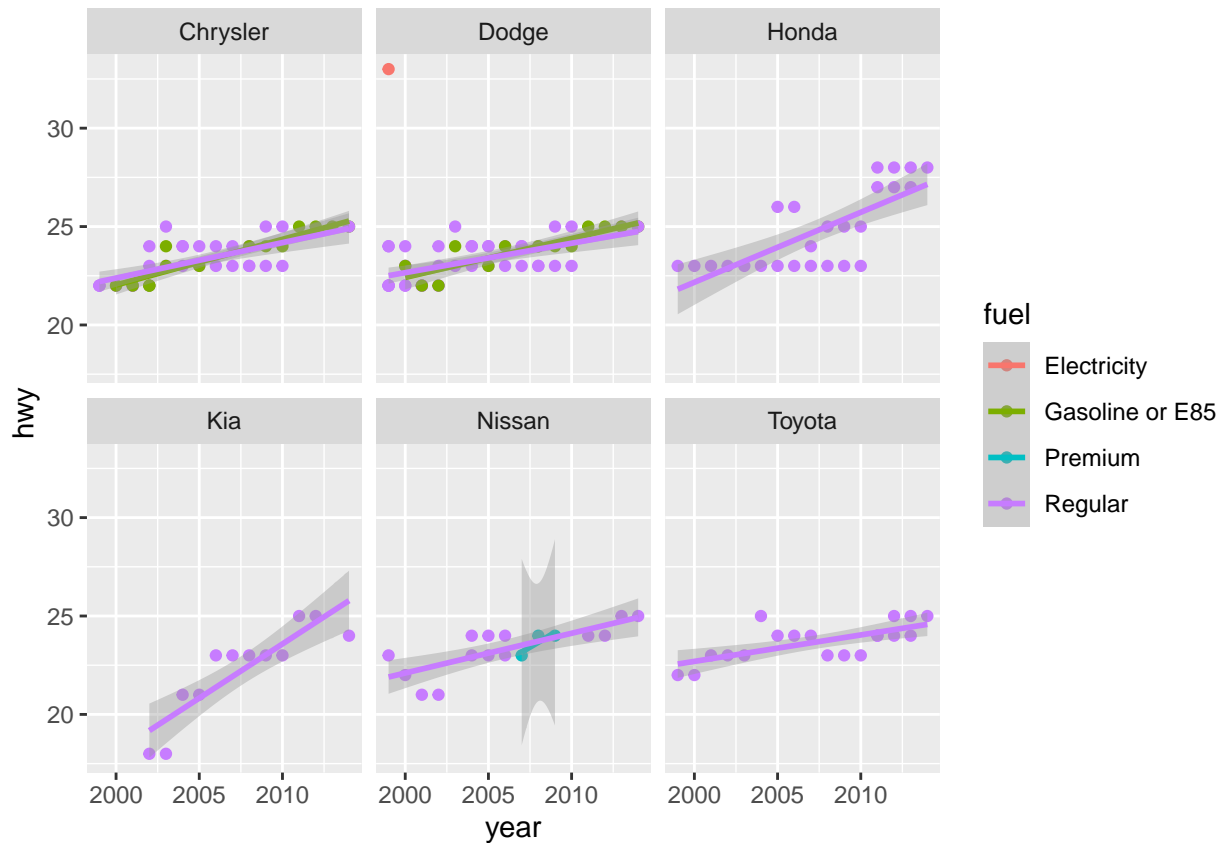**Answer:** Outlier 1999 Dodge Caravan/Grand Caravan 2WD, Electric

```
minivan_2wd %>%
  ggplot(aes(x = year, y = hwy)) +
  geom_point() +
  geom_smooth(method = "lm") +
  facet_wrap(~ make)
```
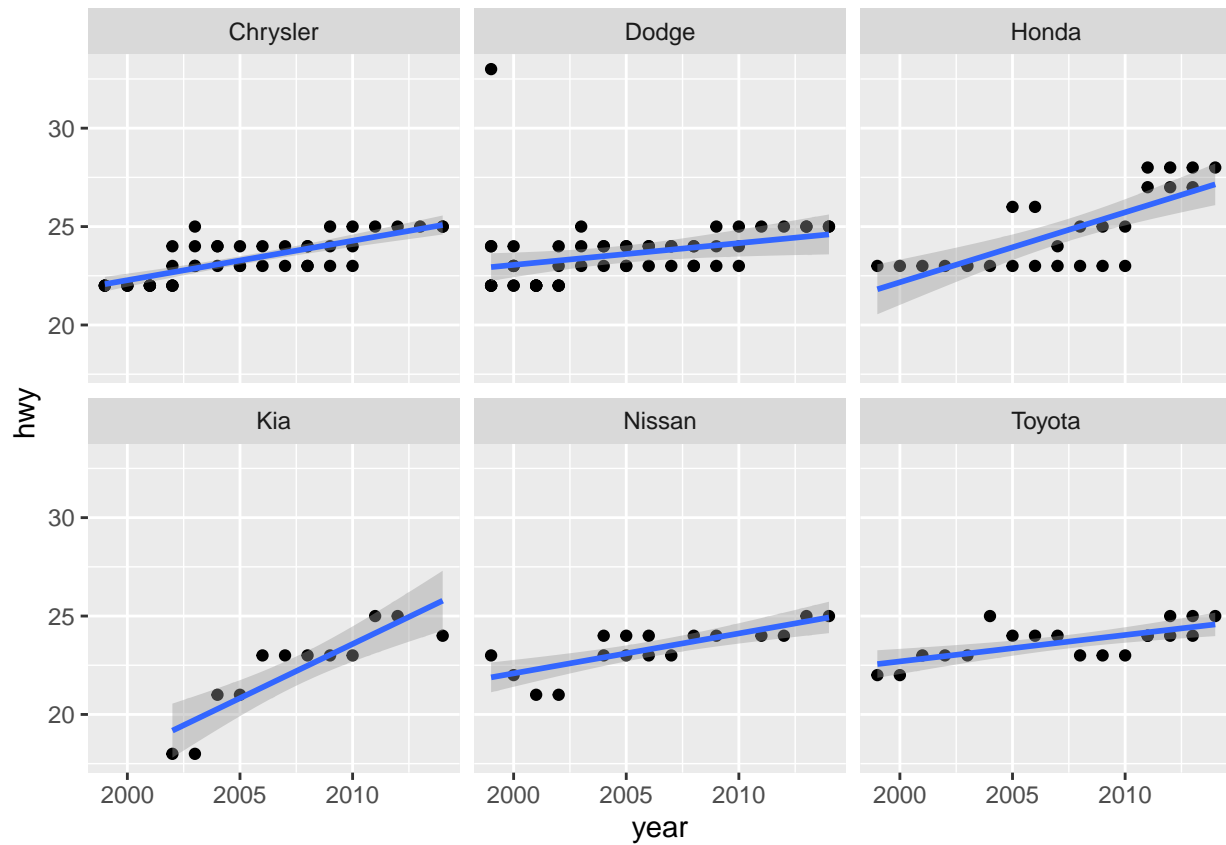
```
## `geom_smooth()` using formula 'y ~ x'
```

```
minivan_2wd %>%
  ggplot(aes(x = year, y = hwy, color = fuel)) +
  geom_point() +
  geom_smooth(method = "lm") +
  facet_wrap(~ make)
```

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

```
minivan_2wd %>% filter(fuel == "Electricity")
```
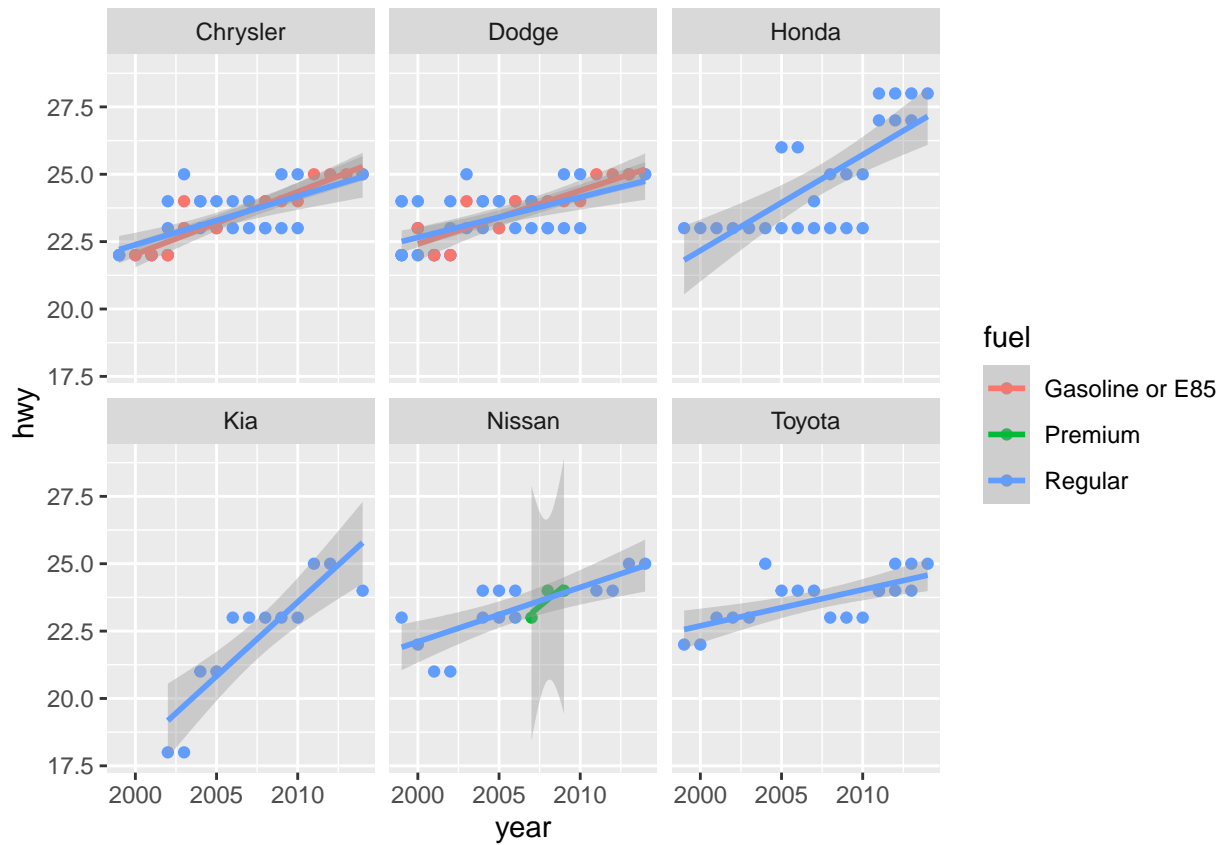
```
## # A tibble: 1 x 12
##       id make  model       year class   trans  drive     cyl displ fuel    hwy   cty
##    <dbl> <chr> <chr>      <dbl> <chr>   <chr>  <chr>   <dbl> <dbl> <chr> <dbl> <dbl>
## 1 30973 Dodge Caravan~   1999 Miniva~ Autom~ 2-Whe~     NA    NA Elec~    33    37
```

(h) Remove the outlier and remake the visualization from the previous part.

```
minivan_2wd %>%
  filter(!(fuel == "Electricity")) %>%
  ggplot(aes(x = year, y = hwy, color = fuel)) +
  geom_point() +
  geom_smooth(method = "lm") +
  facet_wrap(~ make)
```

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

(i) Does it appear that the fuel economy for the minivans from these manufacturers was improving over the years 1985 to 2015? Which manufacturer(s) produced minivans that used gasoline or E85 for fuel over the years 1985 to 2015?

Answer: Yes. Chrysler and Dodge.