

R Markdown Report - Cars dataset

Roopak TJ

R Markdown

This is an R Markdown document involving a small analysis on the **mpg** dataset. The dataset is a part of **ggplot** package present in R. Dataset comprises of car details including manufacturer, model, display, year, cylinder, transmission, drive, city, miles per gallon on highway, fl and class

Task-1

The first analysis is to create a dataframe(table) from mpg which satisfies the following conditions:

- Cars with manual transmission
- Cars having 8 cylinders
- Cars with 4 wheel drive
- Table is present in descending miles per gallon

```
# storing mpg dataset from ggplot2 package to cars dataframe
cars <- ggplot2::mpg
# Creating subset which have rows with 8 cylinder, 4 wheel drive and manual transmission
new.car <- subset(cars, cars$cyl == 8 & cars$drv == 4 & grepl("manual", cars$trans))
# Order in descending miles per gallon
new.car <- new.car[order(-new.car$hwy),]
# Display the dataset in table format
knitr::kable(new.car, caption = "Car models with 8 cylinder, 4 wheel drive and manual transmission")
```

Car models with 8 cylinder, 4 wheel drive and manual transmission

manufacturer	model	displ	year	cyl	trans	drv	cty	hwy	fl	class
dodge	dakota pickup 4wd	5.2	1999	8	manual(m5)	4	11	17	r	pickup
dodge	ram 1500 pickup 4wd	4.7	2008	8	manual(m6)	4	12	16	r	pickup
dodge	ram 1500 pickup 4wd	4.7	2008	8	manual(m6)	4	12	16	r	pickup
dodge	ram 1500 pickup 4wd	5.2	1999	8	manual(m5)	4	11	16	r	pickup
ford	f150 pickup 4wd	4.6	1999	8	manual(m5)	4	13	16	r	pickup
dodge	ram 1500 pickup 4wd	4.7	2008	8	manual(m6)	4	9	12	e	pickup

Task-2

Plot to show the difference in mpg on highway each year for Audi, Nissan and Honda and show 95% CI on the mean

```
# Creating a subset of Cars considering only Audi, Honda and Nissan as manufacturers
new_table<-subset(cars,grepl("audi", cars$manufacturer) | grepl("nissan", cars$manufac
turer) | grepl("honda", cars$manufacturer))

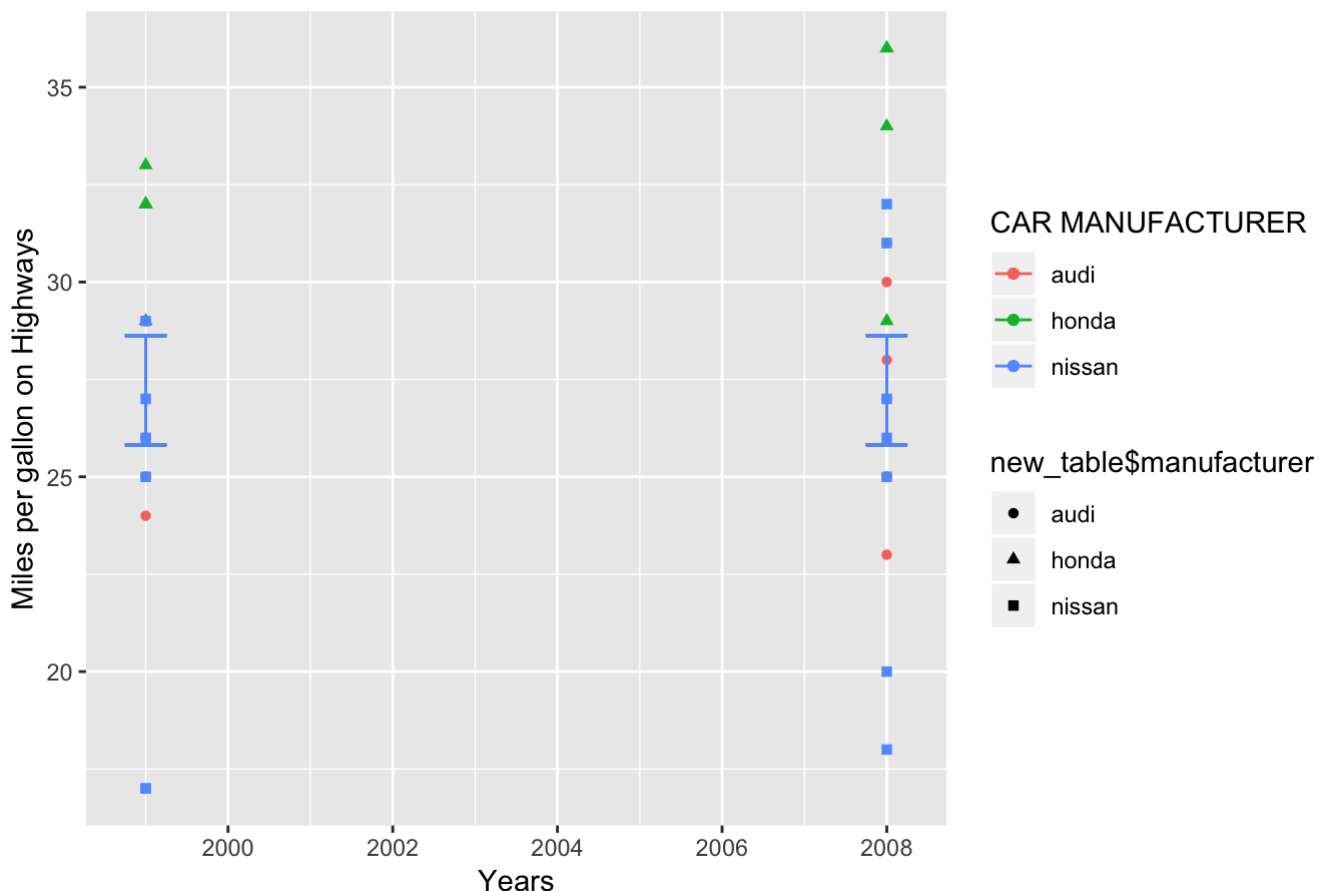
# Calculating 95% confidence Interval for the mean value
m <- mean(new_table$hwy)
v <- var(new_table$hwy)
sd <- sqrt(v)
n <- length(new_table$hwy)
lower.range <- m - 1.96*(sd/sqrt(n))
higher.range <- m + 1.96*(sd/sqrt(n))

require(ggplot2)
```

```
## Loading required package: ggplot2
```

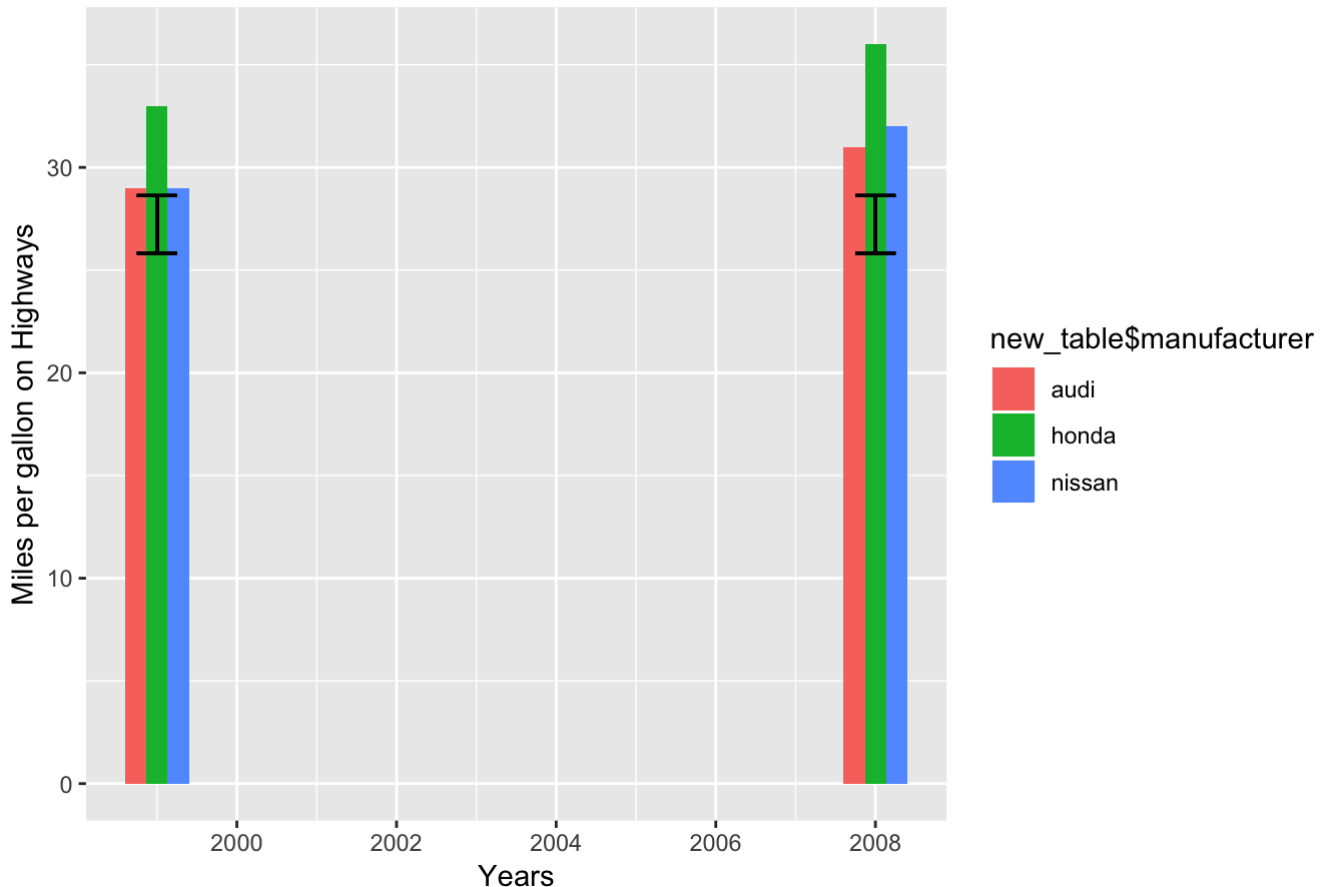
```
# Plotting a Point chart displaying the mpg on highway for Audi, Honda and Nissan. Al
so showing 95% mean CI
g1 <- ggplot(new_table, aes(new_table$year, new_table$hwy, color = new_table$manufac
turer)) +
  geom_point(aes(shape = new_table$manufacturer)) +
  geom_errorbar(aes(ymin=lower.range, ymax=higher.range, width = .5)) +
  xlab("Years") +
  ylab("Miles per gallon on Highways") +
  ggtitle("Point Chart with Mean 95% CI") +
  labs(colour='CAR MANUFACTURER')
g1
```

Point Chart with Mean 95% CI



```
# Plotting a Bar chart displaying the mpg on highway for Audi, Honda and Nissan. Also showing 95% mean CI
g <- ggplot(new_table, aes(new_table$year, new_table$hwy, fill = new_table$manufacturer)) +
  geom_bar(stat="identity", width = .8, position = "dodge") +
  geom_errorbar(aes(ymin=lower.range, ymax=higher.range, width = .5)) +
  xlab("Years") +
  ylab("Miles per gallon on Highways") +
  ggtitle("Bar Chart with Mean 95% CI") +
  labs(colour='CAR MANUFACTURER')
g
```

Bar Chart with Mean 95% CI



The graphs shows that we have data for just 2 years - 1999 and 2008. 95% confidence interval for the whole data lies between 25.82 and 28.62