Homework 2

Rohit Kamat rgk359

This homework is due on Jan. 31, 2017 at 7:00pm. Please submit as a PDF file on Canvas.

This homework uses data from the `Cars93` data set available in `MASS` package. Each observation in the data frame contains information on passenger cars from 1993. This is a big data frame with 27 columns. We are interested in the information on manufacturer (`Manufacturer`), car model (`Model`), type of car (`Type`), passanger capacity (`Passangers`), midrange price in $1000 (`Price`), maximum horsepower (`Horsepower`), company origin (`Origin`), and fuel tank capacity in gallons (`Fuel.tank.capacity`).

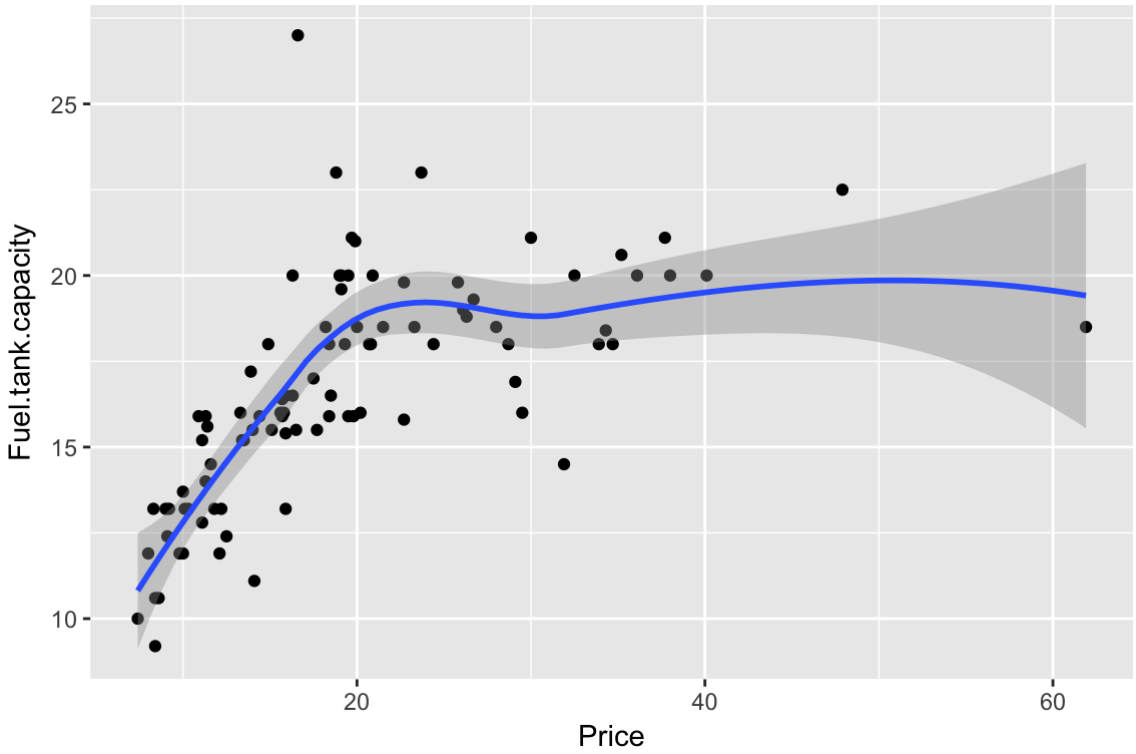
```
library(MASS)
head(Cars93)
```

```
##   Manufacturer   Model    Type Min.Price Price Max.Price MPG.city
## 1        Acura Integra   Small      12.9  15.9      18.8       25
## 2        Acura  Legend Midsize      29.2  33.9      38.7       18
## 3         Audi      90 Compact      25.9  29.1      32.3       20
## 4         Audi     100 Midsize      30.8  37.7      44.6       19
## 5          BMW    535i Midsize      23.7  30.0      36.2       22
## 6        Buick Century Midsize      14.2  15.7      17.3       22
##   MPG.highway            AirBags DriveTrain Cylinders EngineSize
## 1          31               None      Front         4        1.8
## 2          25 Driver & Passenger      Front         6        3.2
## 3          26        Driver only      Front         6        2.8
## 4          26 Driver & Passenger      Front         6        2.8
## 5          30        Driver only       Rear         4        3.5
## 6          31        Driver only      Front         4        2.2
##   Horsepower  RPM Rev.per.mile Man.trans.avail Fuel.tank.capacity
## 1        140 6300         2890             Yes               13.2
## 2        200 5500         2335             Yes               18.0
## 3        172 5500         2280             Yes               16.9
## 4        172 5500         2535             Yes               21.1
## 5        208 5700         2545             Yes               21.1
## 6        110 5200         2565              No               16.4
##   Passengers Length Wheelbase Width Turn.circle Rear.seat.room
## 1          5    177       102    68          37           26.5
## 2          5    195       115    71          38           30.0
## 3          5    180       102    67          37           28.0
## 4          6    193       106    70          37           31.0
## 5          4    186       109    69          39           27.0
## 6          6    189       105    69          41           28.0
##   Luggage.room Weight  Origin          Make
## 1           11   2705 non-USA Acura Integra
## 2           15   3560 non-USA  Acura Legend
## 3           14   3375 non-USA       Audi 90
## 4           17   3405 non-USA      Audi 100
## 5           13   3640 non-USA      BMW 535i
## 6           16   2880     USA Buick Century
```

Problem 1: (2 pts) Use ggplot2 to create a scatter plot of the fuel tank capacity versus the car prices. In the same plot, fit a curve to these data using `geom_smooth()`. In one sentence, what broad trend do you observe in fuel tank capacity for different car prices?

```
ggplot(Cars93,aes(x=Price, y=Fuel.tank.capacity)) + geom_point() + geom_smooth()
```

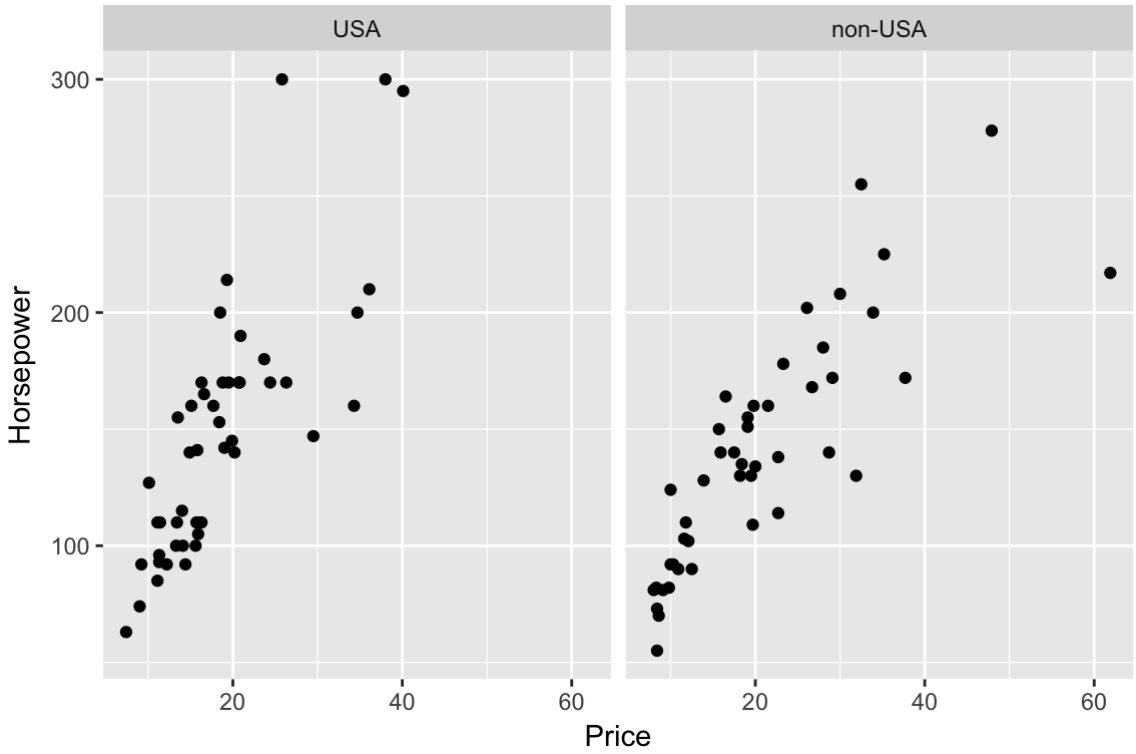
```
## `geom_smooth()` using method = 'loess'
```



As the price increases the fuel tank capacity of the car increases.

Problem 2: (4 pts) Next, create a scatter plot of horsepower against car price, facetted by origin. Make two observations about the data from this plot. State each in 1 sentence.

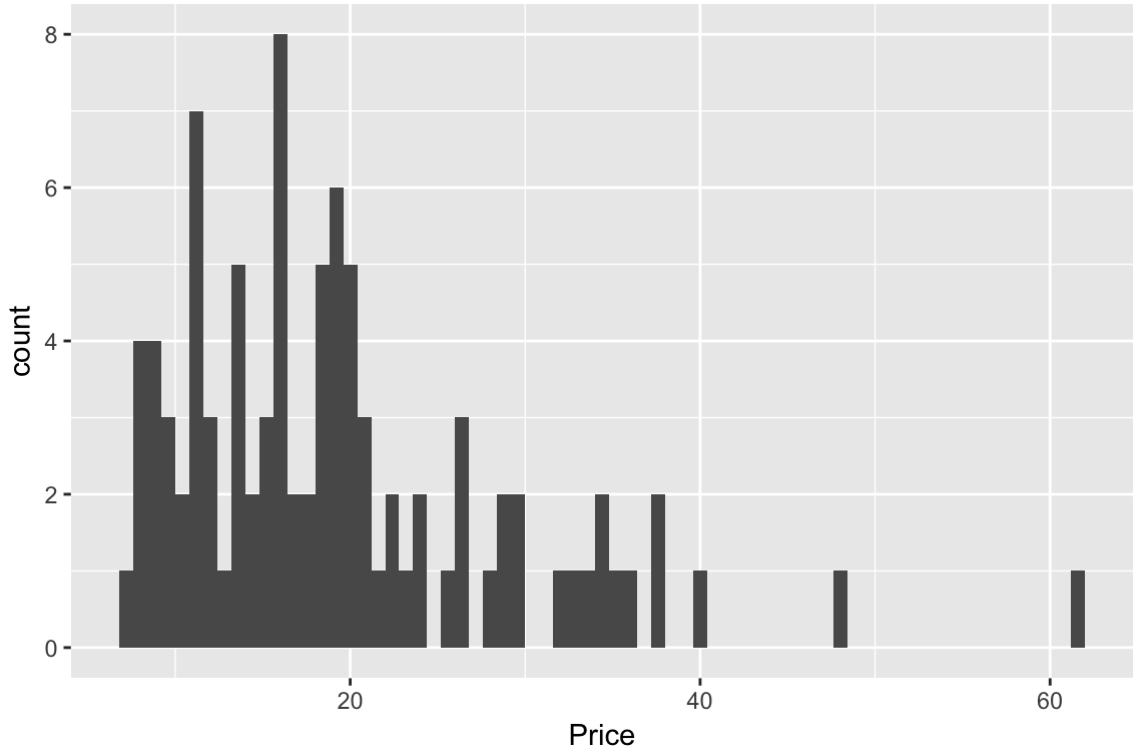
```
ggplot(Cars93,aes(x=Price,y=Horsepower)) + facet_wrap(~Origin) + geom_point()
```



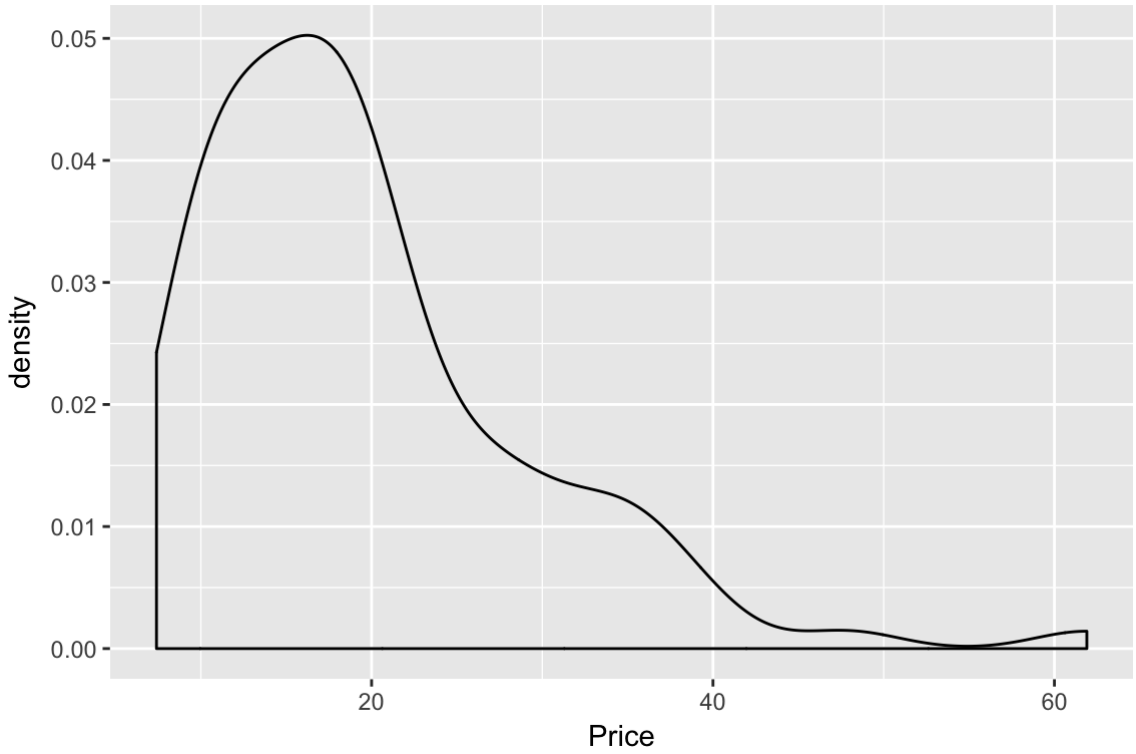
Both USA and Non-USA have a similiar positive trend in regards to the relationship of Horsepower vs. Price. Most of the data for both USA and non-USA fall within 0 to 300 horsepower.

Problem 3: (2 pts) Plot the distribution of price, once using `geom_histogram()` and once using `geom_density()`.

```
ggplot(Cars93,aes(x=Price))+ geom_histogram(binwidth = .8)
```




```
ggplot(Cars93,aes(x=Price))+ geom_density()
```



Problem 4: (2 pts) What does the y-axis in your histogram represent? In your density plot, what is the *total* area under the curve? Please give a single number as your answer. **HINT**: You do not need to do any additional calculations to determine the area under the curve.

The y-axis in the histogram tells the amount of cars that are under a certain price. The total area under the curve for the density plot is 1.