Data visualization Lab

Bibek Sapkota

Data Visualization Lab smooting(qplot)

Task 1: Importing ggplot2 library

```
library (ggplot2)
library(splines)
```

Task 2: Importing diamonds dataset and printing it.

```
data <- diamonds
data
```

```
## # A tibble: 53,940 x 10
##    carat cut       color clarity depth table price     x     y     z
##    <dbl> <ord>     <ord> <ord>   <dbl> <dbl> <int> <dbl> <dbl> <dbl>
##  1  0.23 Ideal     E     SI2      61.5    55   326  3.95  3.98  2.43
##  2  0.21 Premium   E     SI1      59.8    61   326  3.89  3.84  2.31
##  3  0.23 Good      E     VS1      56.9    65   327  4.05  4.07  2.31
##  4  0.29 Premium   I     VS2      62.4    58   334  4.2   4.23  2.63
##  5  0.31 Good      J     SI2      63.3    58   335  4.34  4.35  2.75
##  6  0.24 Very Good J     VVS2     62.8    57   336  3.94  3.96  2.48
##  7  0.24 Very Good I     VVS1     62.3    57   336  3.95  3.98  2.47
##  8  0.26 Very Good H     SI1      61.9    55   337  4.07  4.11  2.53
##  9  0.22 Fair      E     VS2      65.1    61   337  3.87  3.78  2.49
## 10  0.23 Very Good H     VS1      59.4    61   338  4     4.05  2.39
## # i 53,930 more rows
```

Task :Creating a random seed to 1000 and create a subset dsmall containing 100 randomly sampled rows from the diamonds dataset.

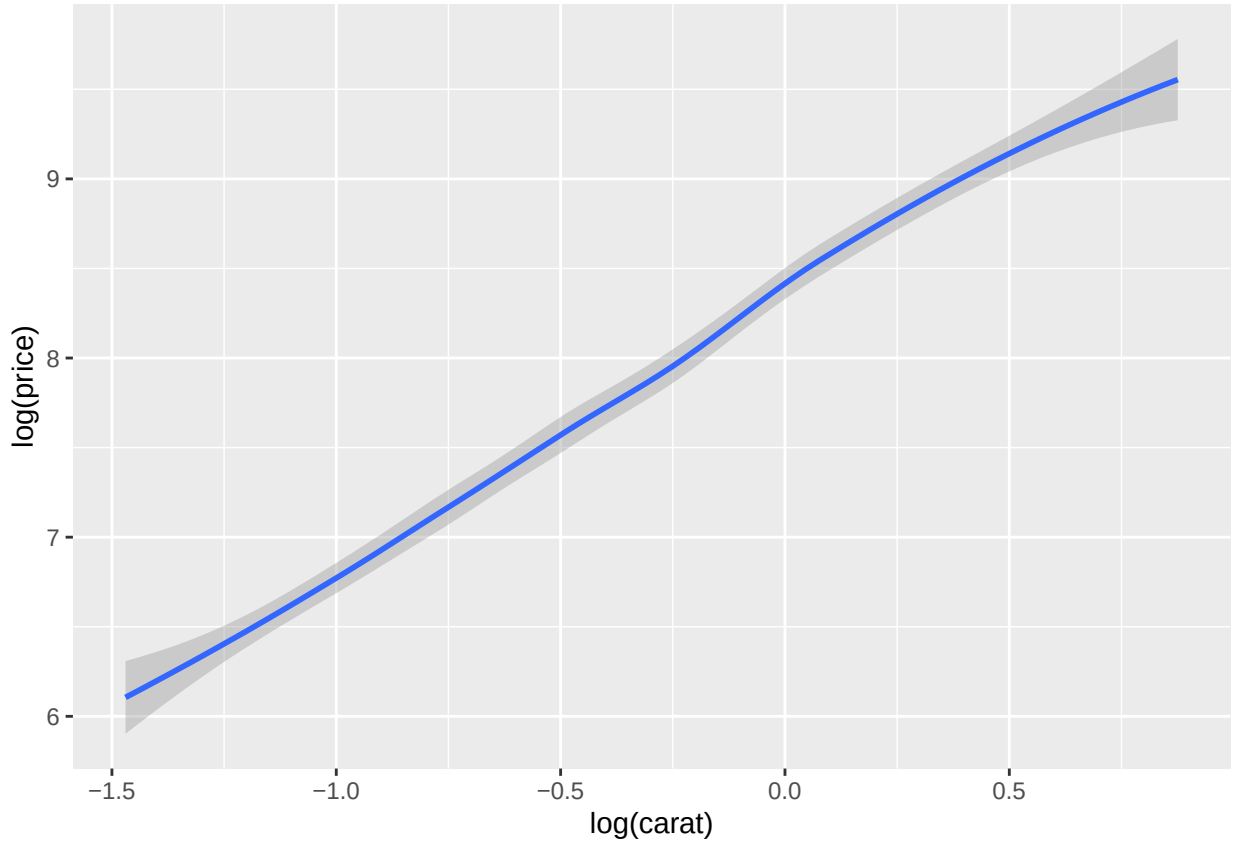
```
set.seed(1000)
dsmall <- diamonds[sample(nrow(diamonds),100),]
```

Task 3:visualize the relationship between the logarithmically transformed log(carat) and log(price) of diamonds in the dsmall dataset using a scatter plot with a smoothed trend line

```
qplot(log(carat),log(price),data=dsmall,geom = "smooth")
```

```
## Warning: `qplot()` was deprecated in ggplot2 3.4.0.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

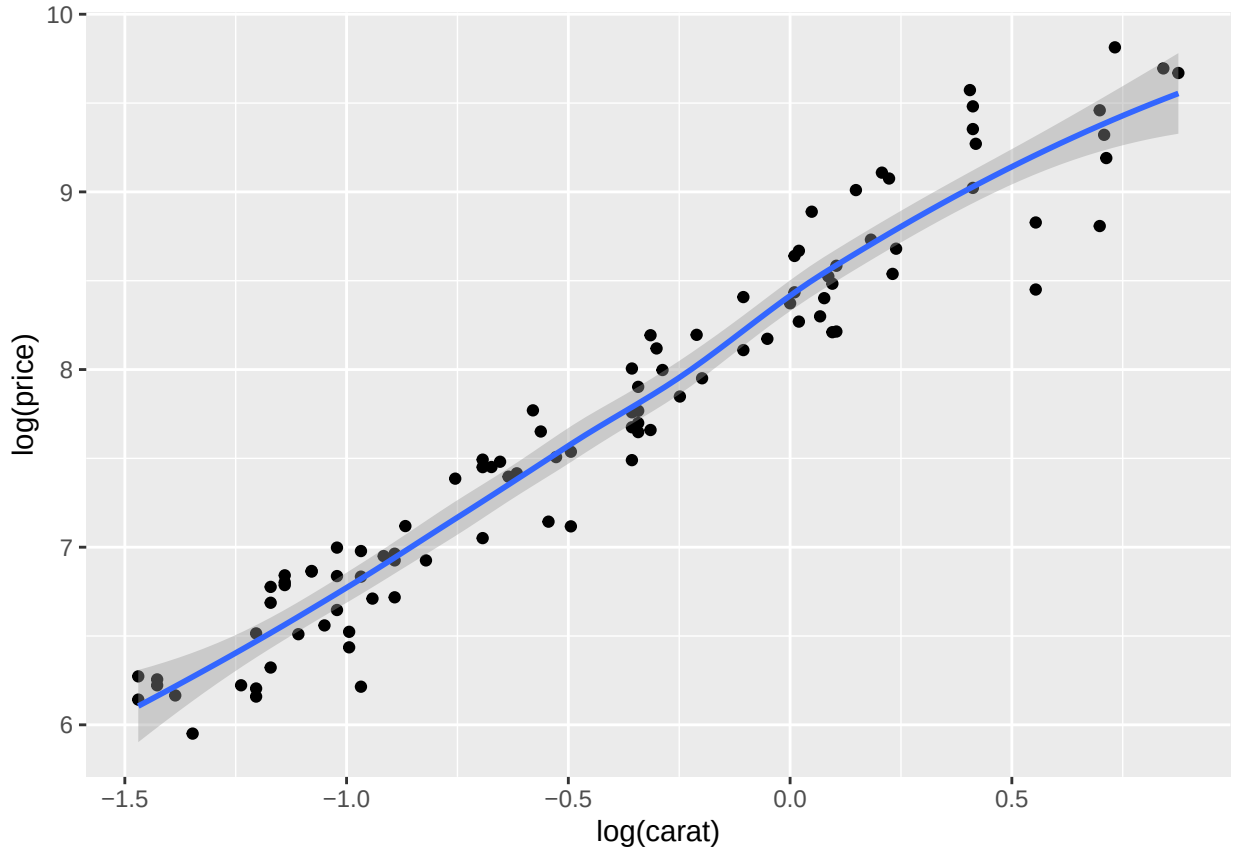
```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



Task 4: ploting the logarithm of diamond carat vs. price, displaying both points and a smoothed trend line.

```
qplot(log(carat),log(price),data = dsmall,geom = c("point","smooth"))
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

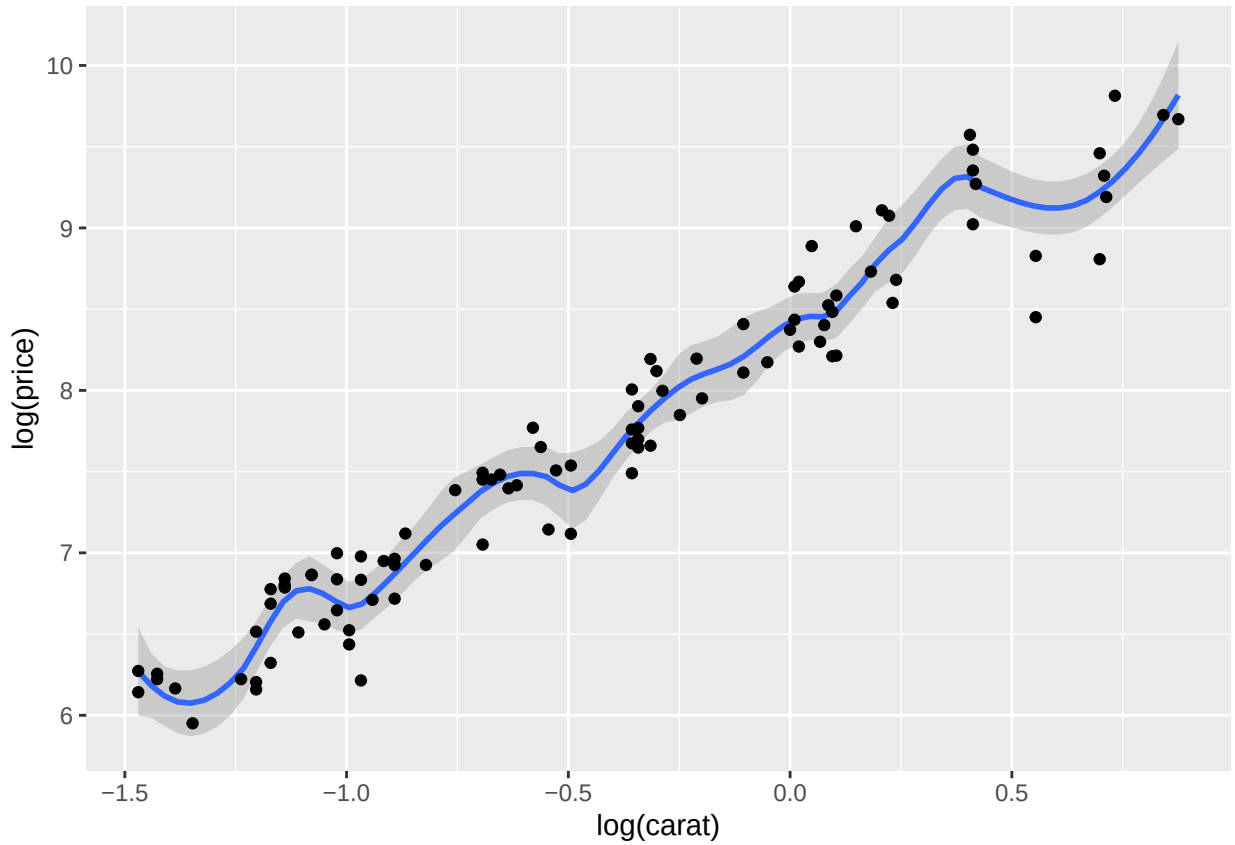


Task 5: Creating a scatter plot of the logarithm of diamond carat versus price from the dsmall dataset, displaying both points and a smoothed trend line with a span of 0.2 for the smoothing parameter.

```
qplot(log(carat),log(price),data = dsmall,geom = c("smooth","point"),span=0.2)
```

```
## Warning in geom_point(span = 0.2): Ignoring unknown parameters: ‘span‘
```

```
## ‘geom_smooth()‘ using method = ’loess’ and formula = ’y ~ x’
```

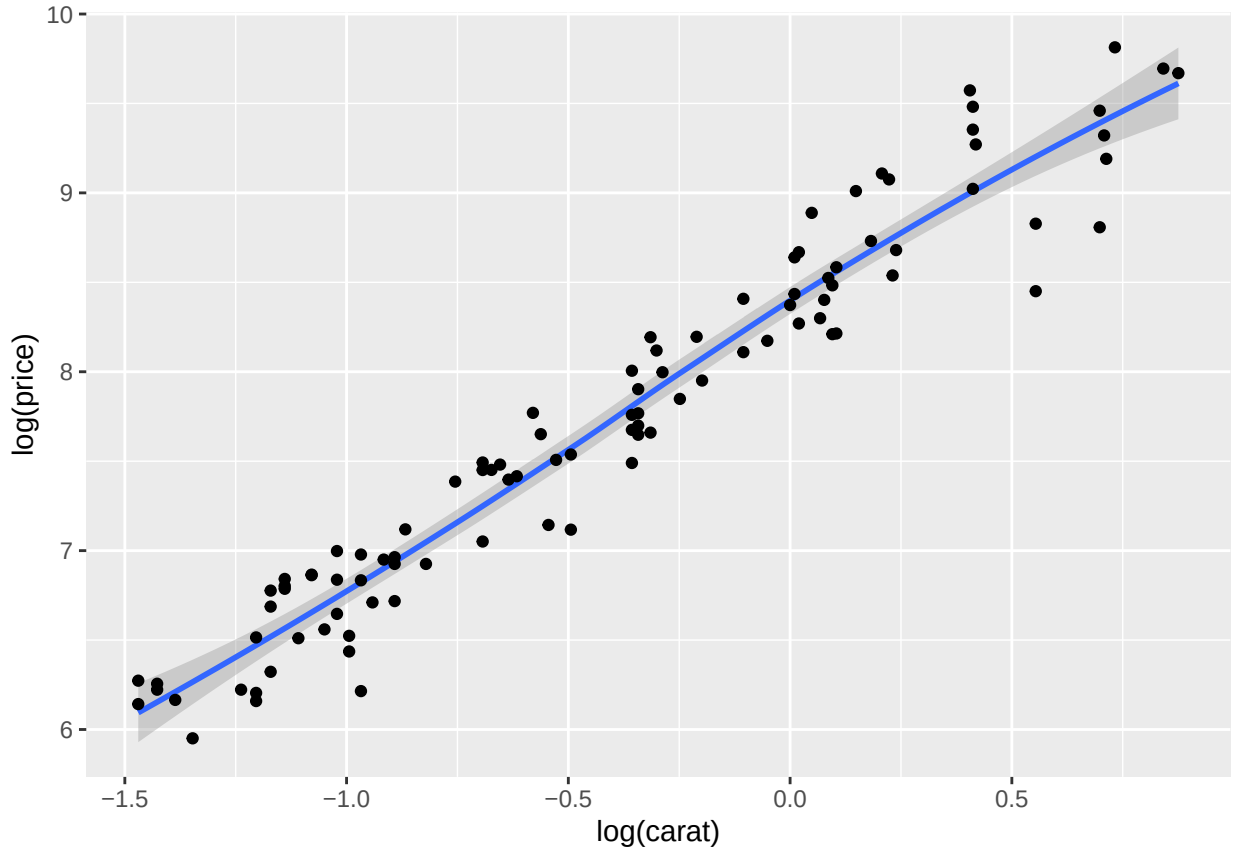


Task 6: Creating a scatter plot of the logarithm of diamond carat vs price from the dsmall dataset, displaying both points and a smoothed trend line with a smoothing span parameter set to 1.

```
qplot(log(carat),log(price),data = dsmall,geom = c("smooth","point"),span=1)
```

```
## Warning in geom_point(span = 1): Ignoring unknown parameters: ‘span‘
```

```
## ‘geom_smooth()‘ using method = ’loess’ and formula = ’y ~ x’
```



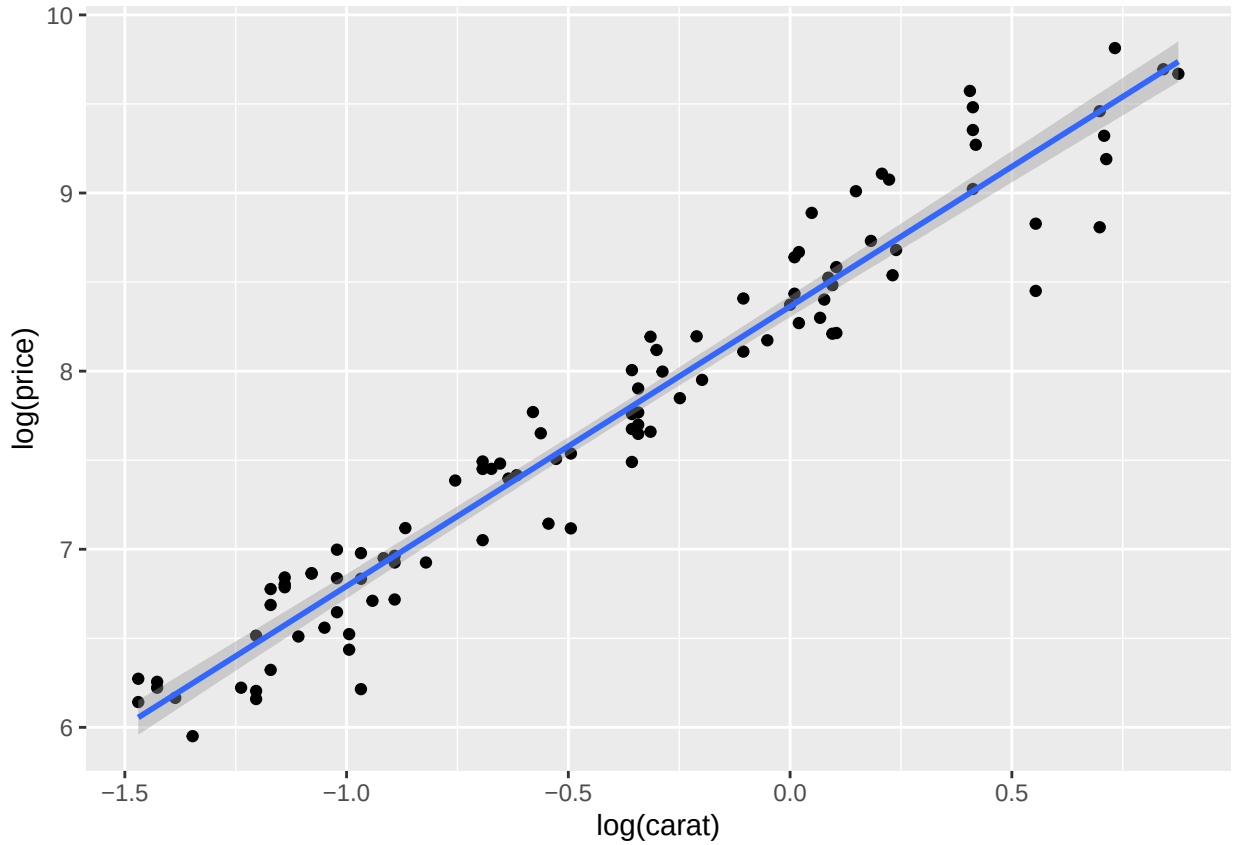
Fit a linear model instead by using method = "lm" or robust linear model "rlm" which is less sensitive to outlier

Task 1: Visualizing the relationship between the (log(carat)) and (log(price)) of diamonds in the dsmall dataset using a scatter plot with individual points and a linear regression smooth line

```
qplot(log(carat), log(price), data = dsmall, geom = c("point", "smooth"), method = "lm")
```

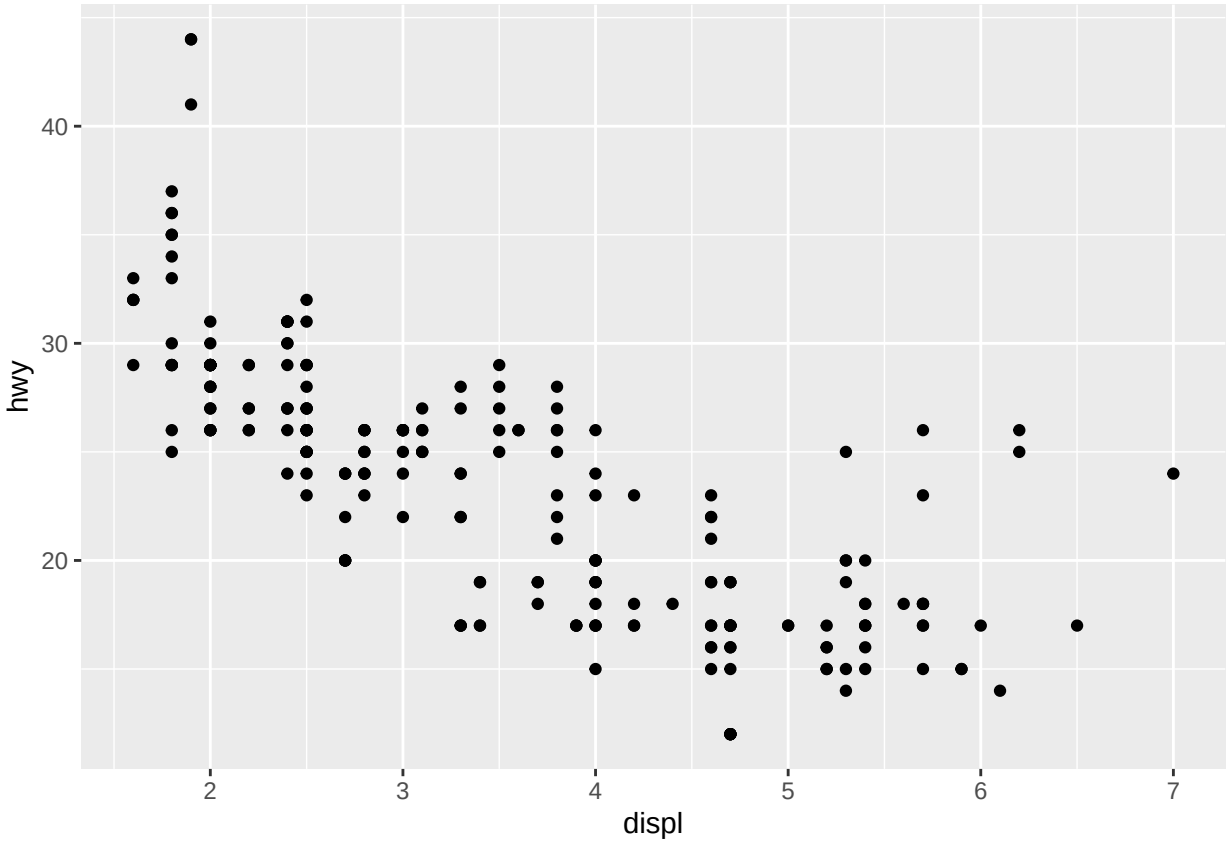
```
## Warning in geom_point(method = "lm"): Ignoring unknown parameters: `method`
```

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



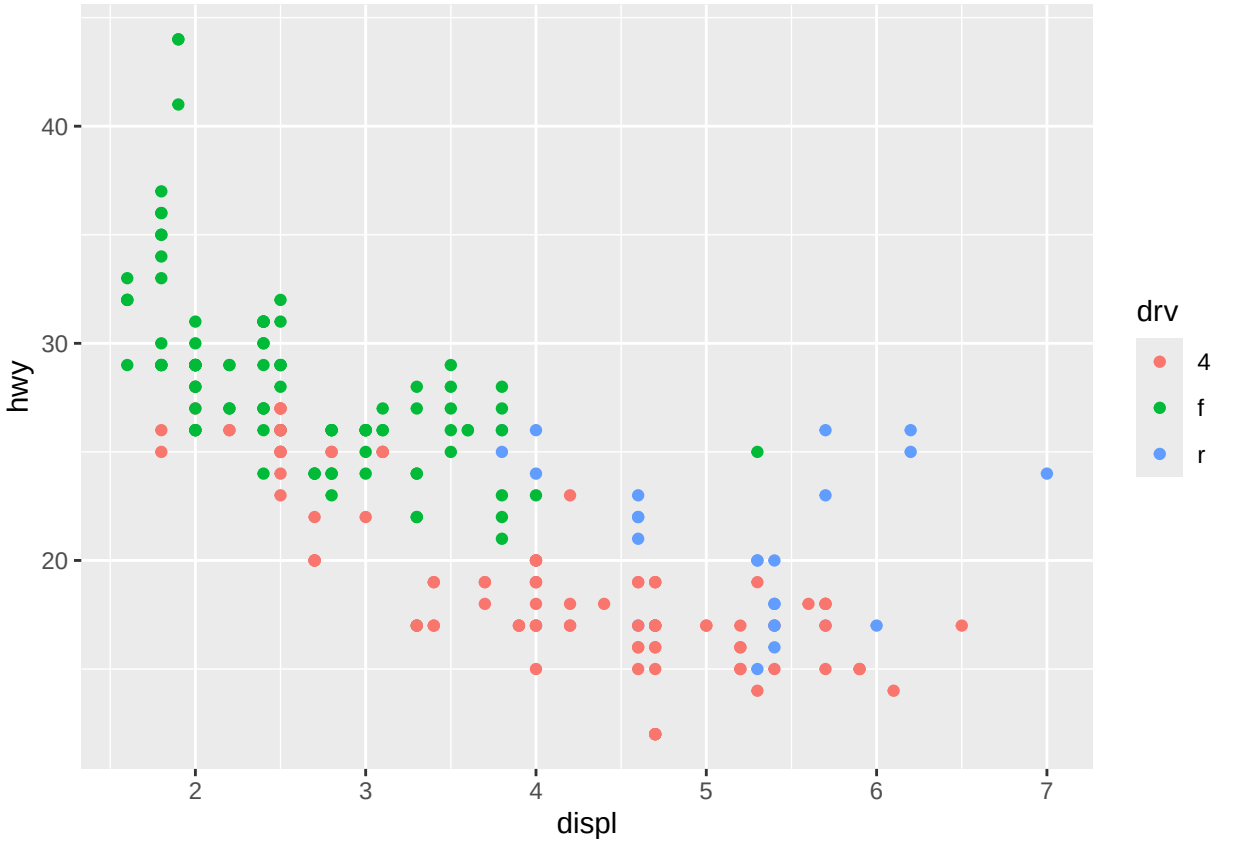
Task 2: Creating a scatter plot to visualize the relationship between engine displacement (displ) and highway miles per gallon (hwy) for cars in the mpg dataset.

```
qplot(displ,hwy,data = mpg)
```



Task 3: Creating a scatter plot of engine displacement (displ) vs highway miles per gallon (hwy) from the mpg dataset, with points colored by the drive type (drv).

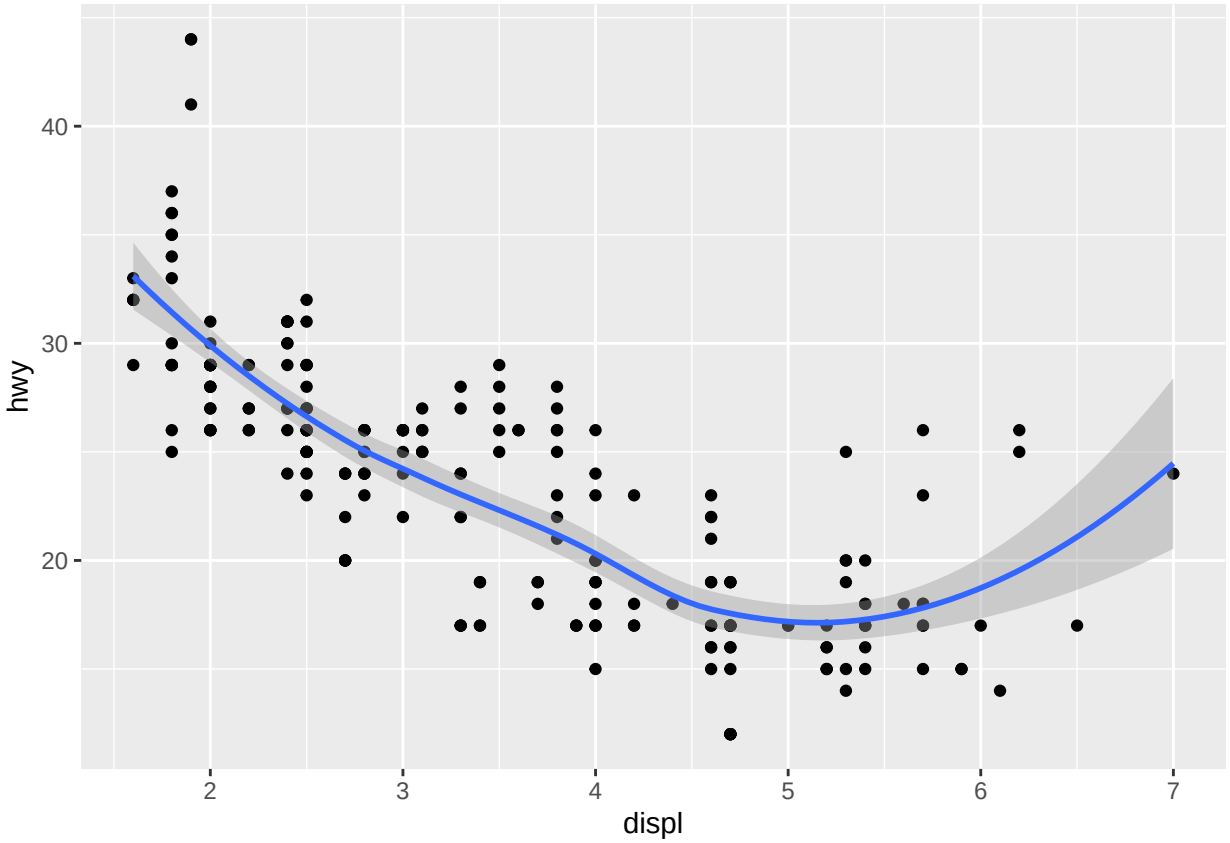
```
qplot(displ, hwy, data = mpg, colour = drv)
```



Task 4: Creating a scatter plot of engine displacement (displ) vs highway miles per gallon (hwy) from the mpg dataset, displaying both points and a smoothed trend line.

```
qplot(displ, hwy, data = mpg, geom = c("point", "smooth"))
```

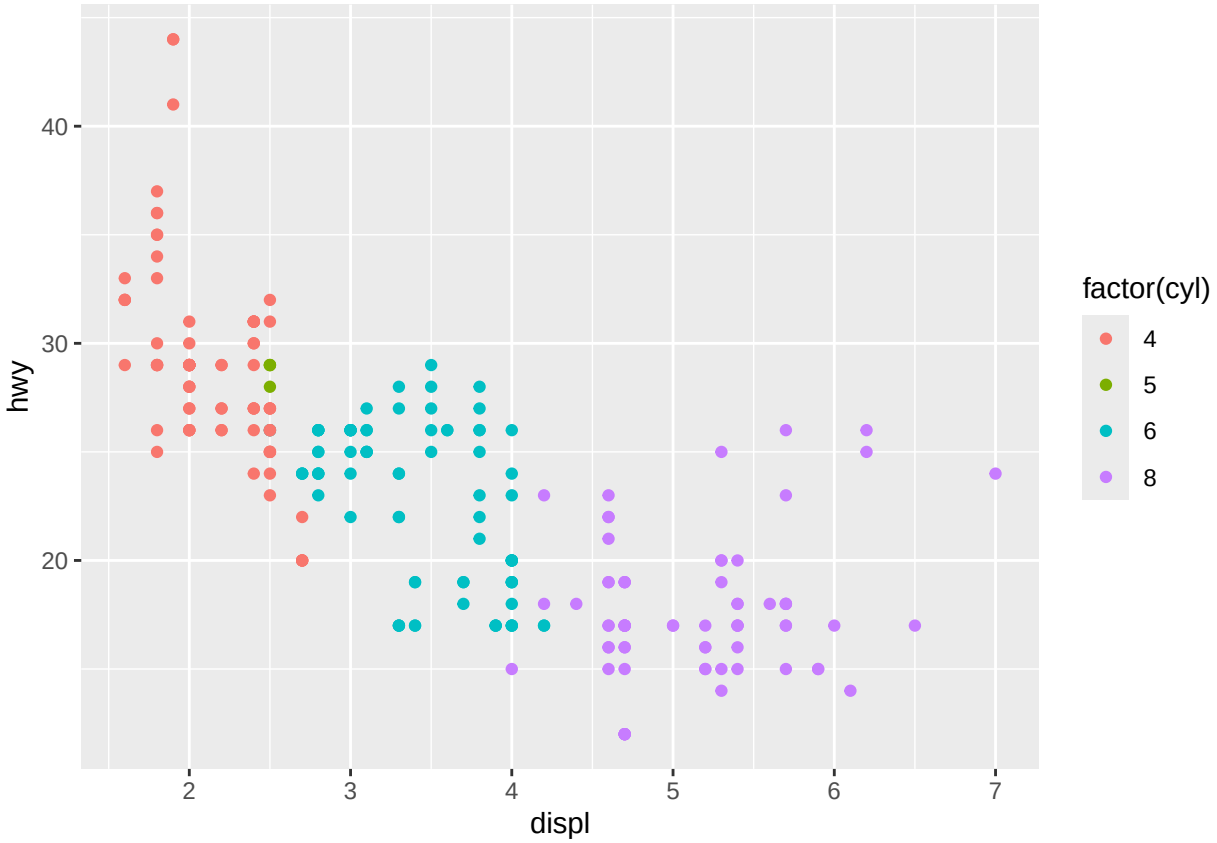
```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



Question:How are engine size and fuel economy related?

Task 1: creating a scatter plot of engine displacement (displ) versus highway miles per gallon (hwy) from the mpg dataset, with points colored according to the number of cylinders (cyl).

```
qplot(displ,hwy,data = mpg,colour=factor(cyl))
```

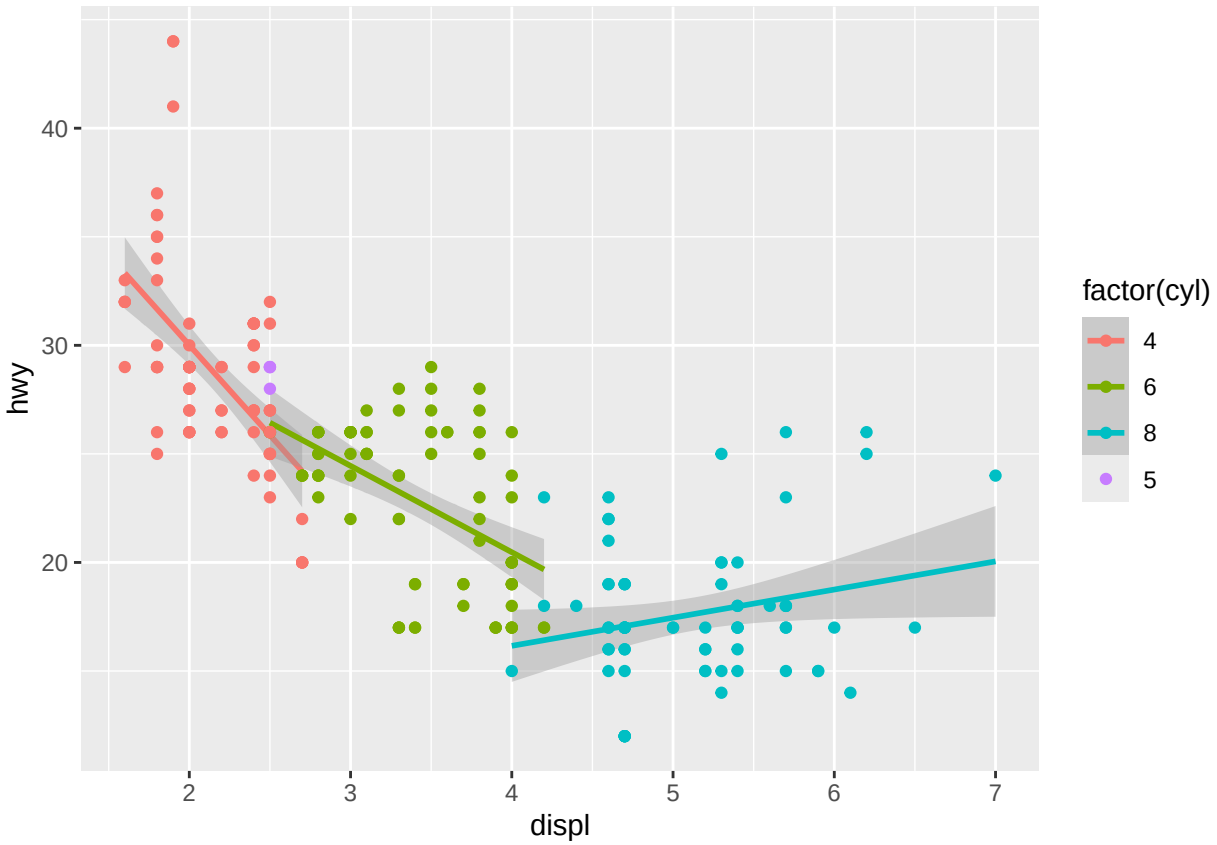


task 2: creating a scatter plot of engine displacement (displ) versus highway miles per gallon (hwy) from the mpg dataset, with points colored by the number of cylinders (cyl) and including a linear regression smooth line.

```
qplot(displ,hwy,data = mpg,colour=factor(cyl),geom = c("smooth","point"),method="lm")
```

```
## Warning in geom_point(method = "lm"): Ignoring unknown parameters: `method`
```

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


40
30
20
hwy
2
3
4
5
6
7
displ
factor(cyl)
4
6
8
5