Lecture 7

1. Recap:

- The base: plot() functions

2. Today's agenda

- ggplot2()

ggplot2 package

What is ggplot2? 1. Its a package in R called 'ggplot2' 2. An implementation of the grammar of graphics: a description of how a graphics can be broken down into abstract concepts. 3. Written by Hadley Wickham

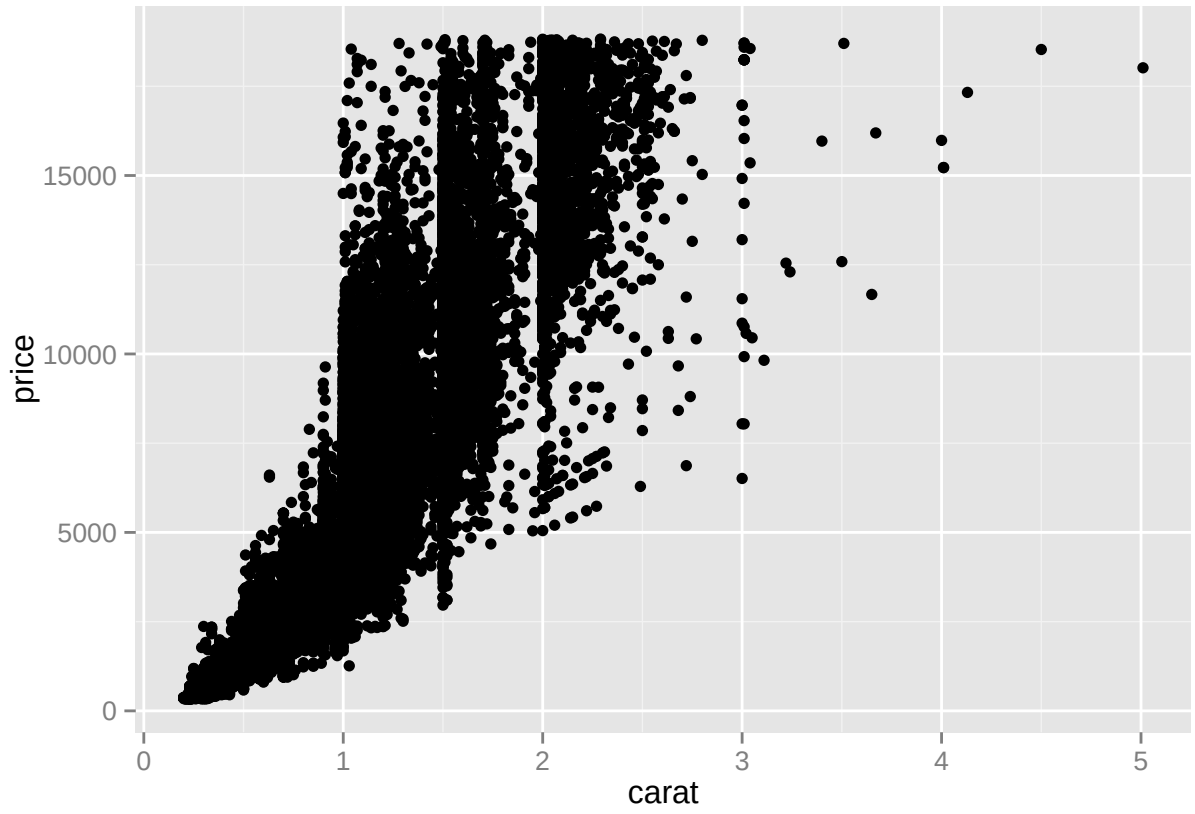
From ggplot2 book: "In brief, the grammar tells us that a statistical graphic is a mapping from data to aesthetic attributes (colour, shape, size) of geometric objects (points, lines, bars). The plot may also contain statistical transformations of the data and is drawn on a specific coordinate system."

Three main modes for plotting in R: - plot functions adding notes and comments - lattice plot - ggplot functions – add layers, and take care of a lot of things automatically.

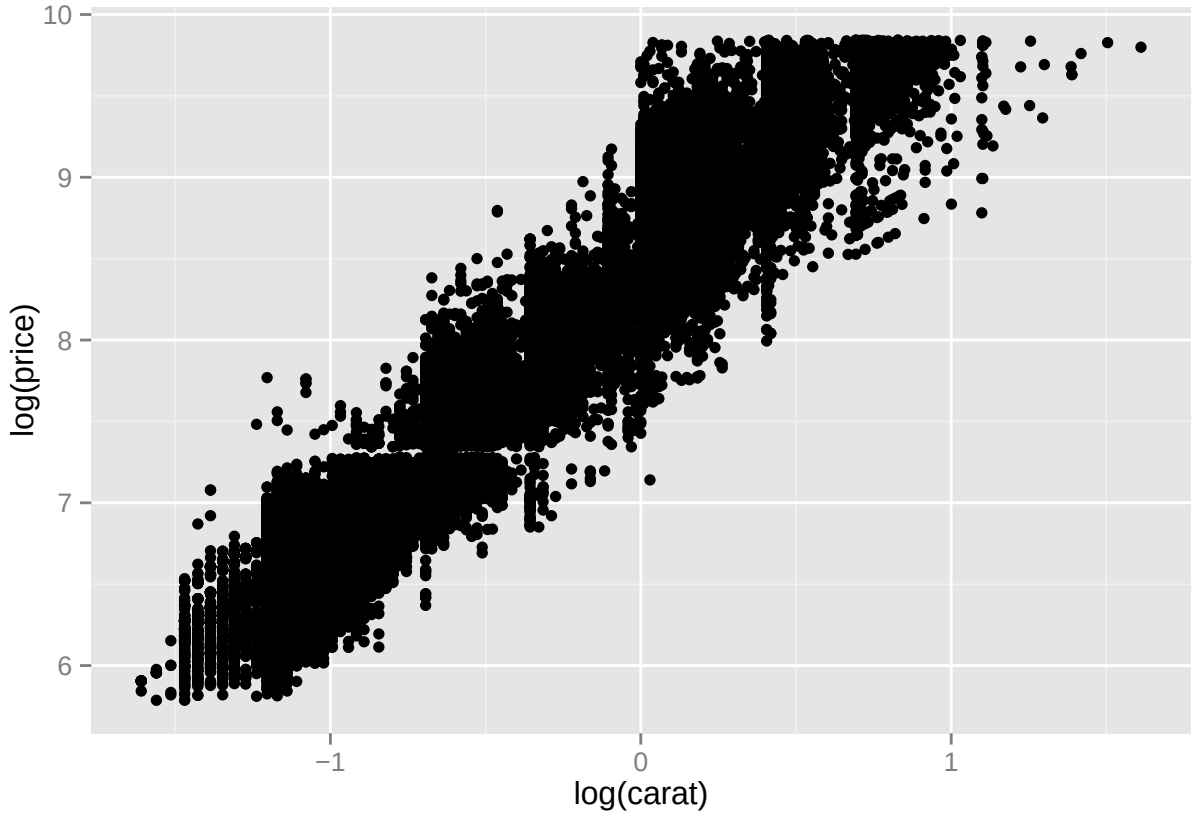
One main issue about plot function: you have to plan in advance, and then you get a whole plot. ### qplot() - Stands for quick plot - Looks very similar to plot function - Needs a data frame. - Plots are made up of aesthetics(size, shape, color) and geoms(points, lines) - Geoms: geometric objects to represent observations - Factors are very important in ggplot - ggplot is the core function and can do things that qplot() can not do. - The diamond example: the data set is available under the package ggplot2.

This dataset containing the prices and other attributes of almost 54,000 diamonds.

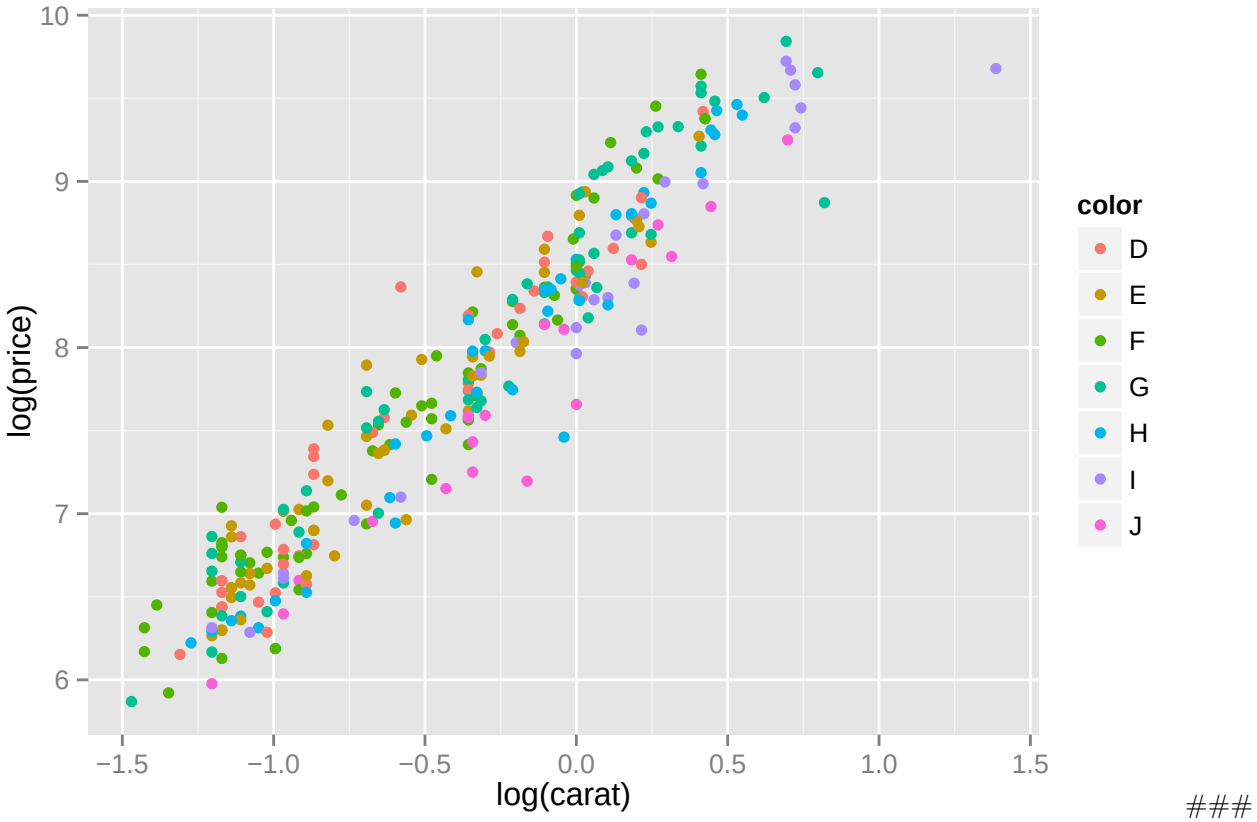
```
library(ggplot2)
qplot(carat, price, data = diamonds)
```




```
qplot(log(carat), log(price), data = diamonds)
```



```
dsmall <- diamonds[sample(nrow(diamonds), 300), ]
qplot(log(carat),log(price), colour = color, data = dsmall)
```



###

ggplot() When we used qplot(), it did a lot of things for us: it created a plot object, added layers, and displayed the result, using many default values along the way.

Some basics about ggplot() function:

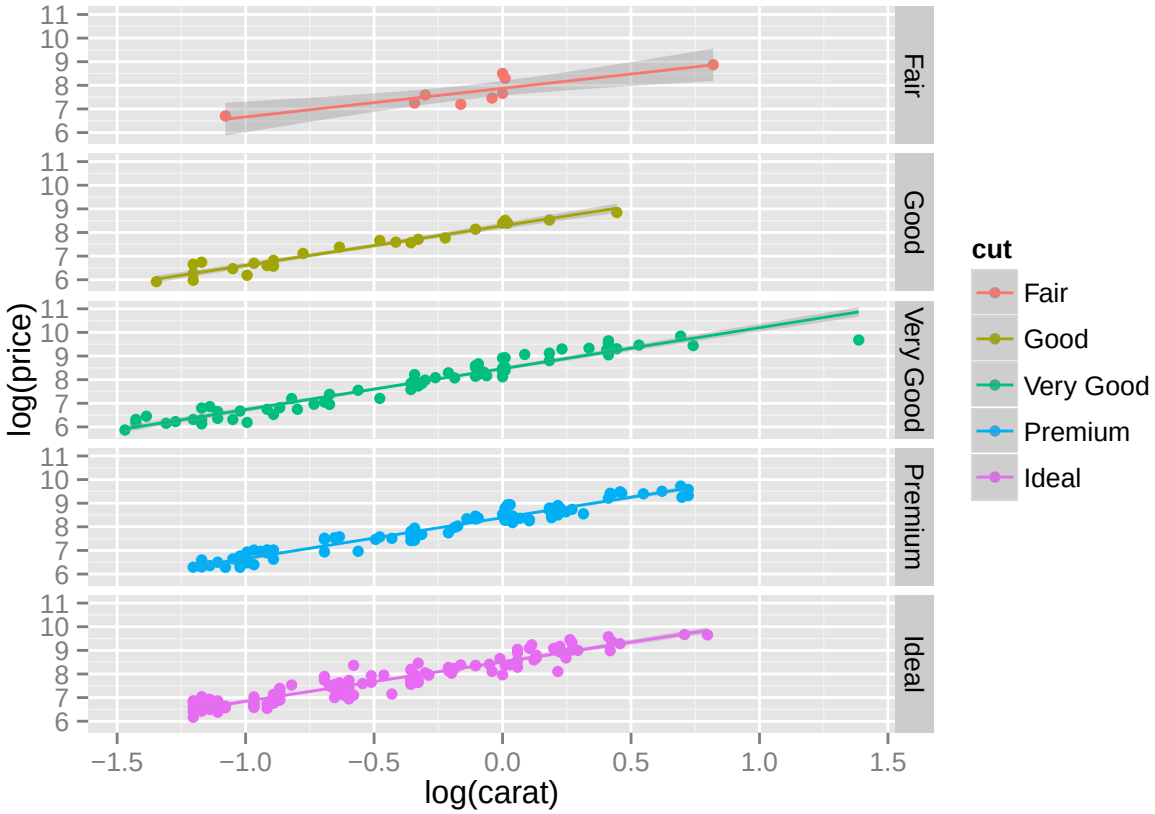
1. Needs a data frame
2. aesthetic mapping
3. geoms: geometric objects
4. facets: multiple panels
5. stats: statistical transformations
6. scales
7. coordinate system

Plots are built up in layers:

1. Plot the data
2. Overlay a summary: a statistics summary, regression line
3. Metadata and annotation

We just learned that:

```
qplot(log(carat), log(price), data = dsmall, facets = cut~., geom = c("point", "smooth"), method = 'lm'
```

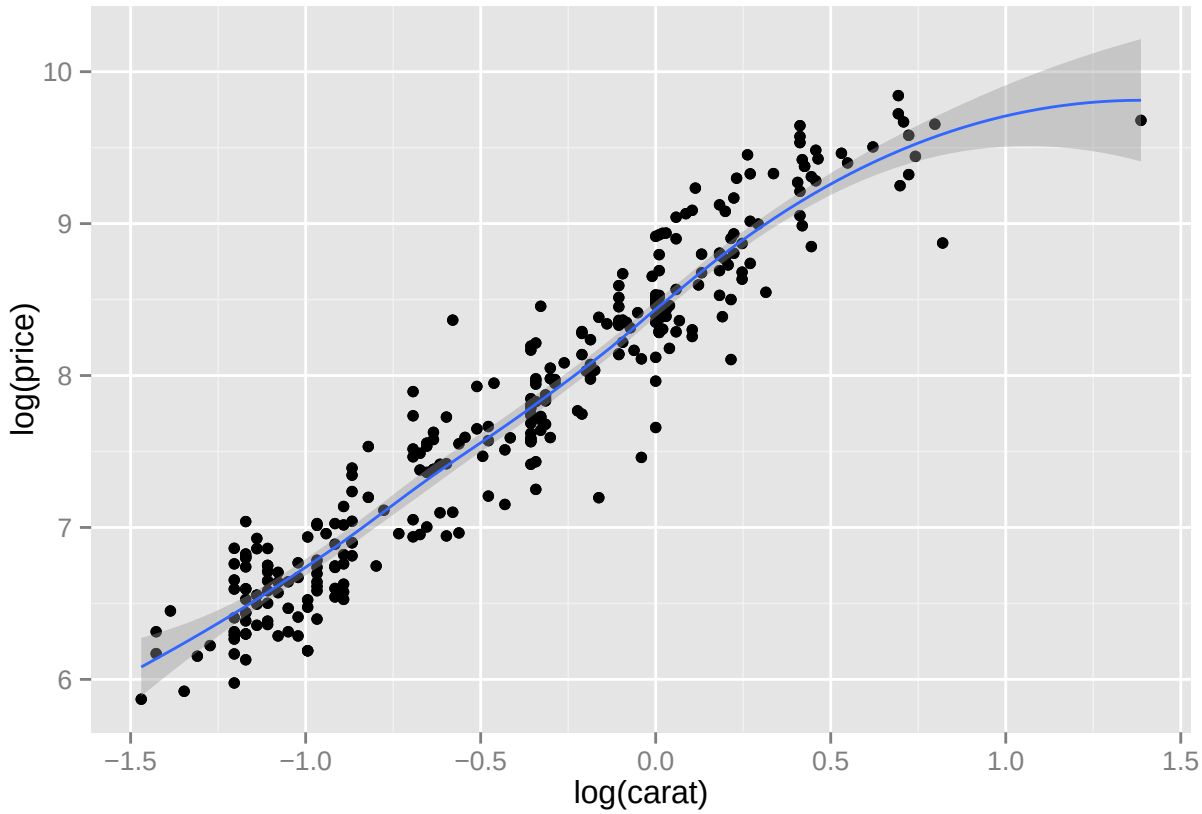


Now, we learn how to build with ggplot() function: Building ggplot():

```
g = ggplot(dsmall, aes(log(carat), log(price))) +  geom_point()
```

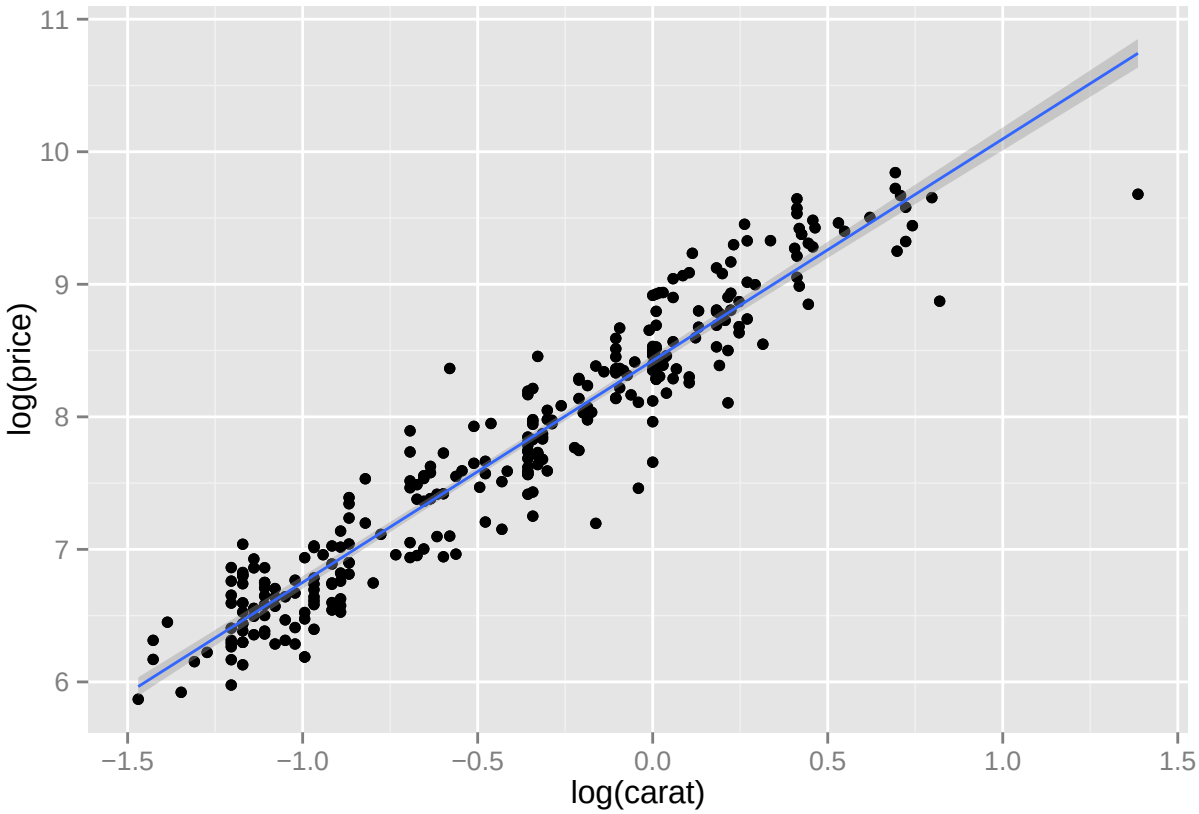
Adding a smoother:

```
g + geom_point() + geom_smooth()
```



Adding another smoother:

```
g + geom_point() + geom_smooth(method = 'lm')
```



Adding Facets:

```
g + geom_point() +facet_grid(cut~.)+geom_smooth(method = 'lm')
```

Adding Colors:

```
g + geom_point(aes(colour = cut)) +facet_grid(cut~.) + geom_smooth(method = 'lm')
```

Remarks:

1. Orders do not matter
2. Make sure that your data has meaningful factor levels.

Some other changes could be done to your plot:

1. Modifying Aesthetics: `g + geom_point(aes(colour = cut), size = 4, alpha = 1/4)`
2. Annotation: You can change the labels of x, y axis by adding xlab(), ylab(), ggtitle()
3. The background of your ggplot can be changed using `theme_gray()` default setting `theme_bw()` ---change to blacks
4. Change the font by `{r} base_family = "Times"`
5. You can also delete your legend by issuing `theme(lengend.position = "none"")`