BUSINESS ANALYTICS CLUB

Workshop Series 12.5

Sexy Graphs
w/ ggplot2 in R

Eddie Shim

BAC

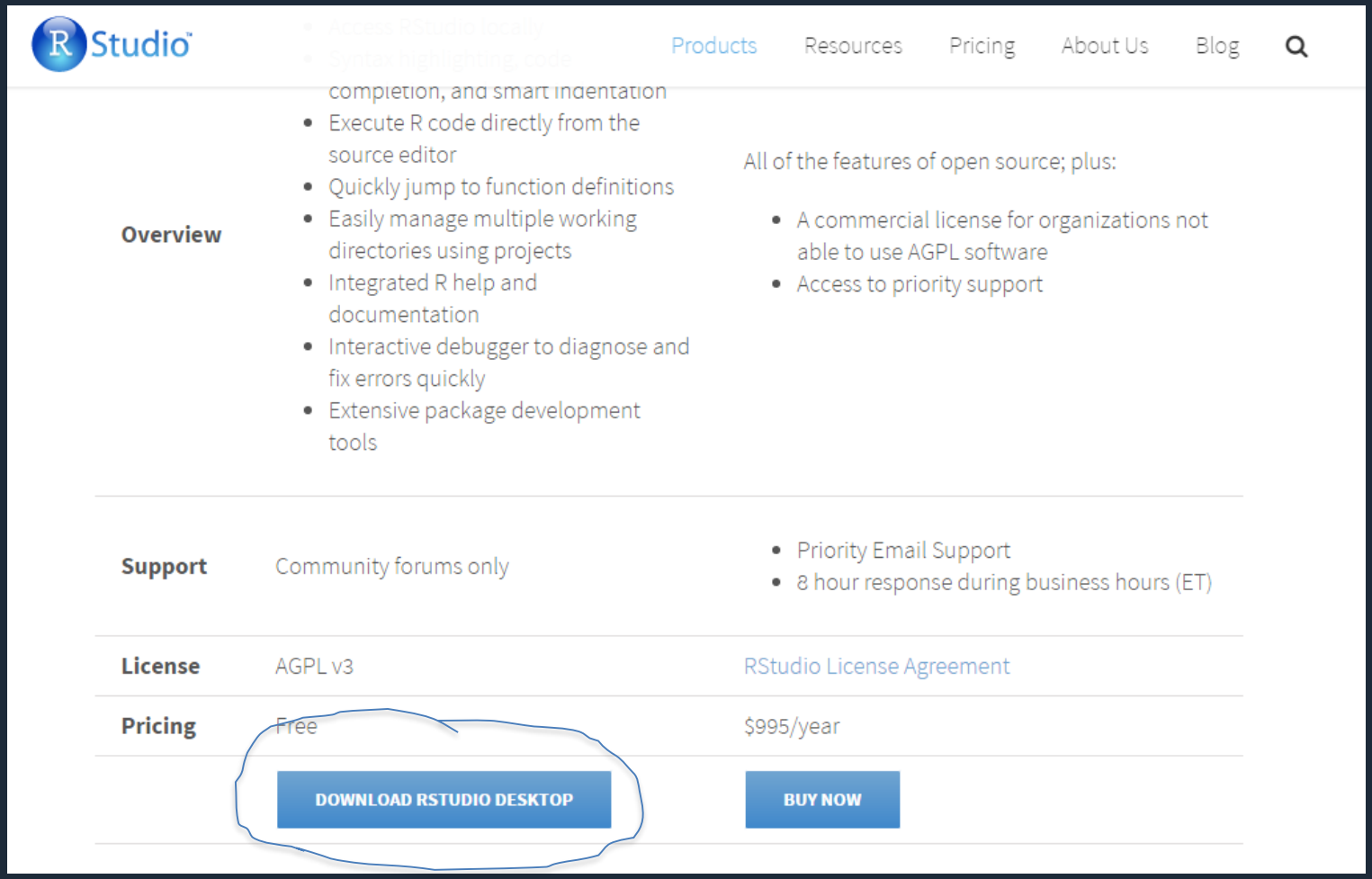
Download R

bit.ly/bacdata

"R and Trading" folder
>Windows R-3.2.2.exe
>Mac R-3.2.2.pkg

Download RStudio

https://www.rstudio.com/products/rstudio/



Setting up ggplot2

```
>install.packages("ggplot2")
>library("ggplot2")
```

Don't forget the quotes! Input parameters must be strings for install.packages() function

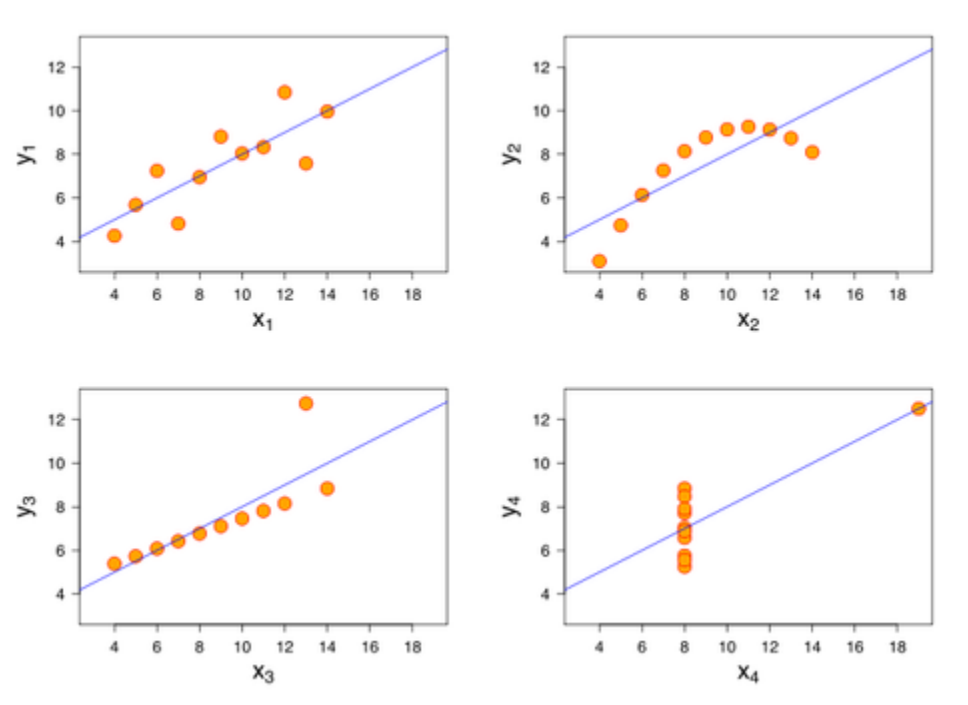
Miscellaneous

- ggplot stands for Grammar of Graphics
- In R, if you press the up arrow, then you can retrieve your last command
- File>New File>R Script
  Allows you to write in script instead of command lines
- Try not to copy paste any of the code. Format errors can occur and you learn more by typing.

Why is data visualization important?

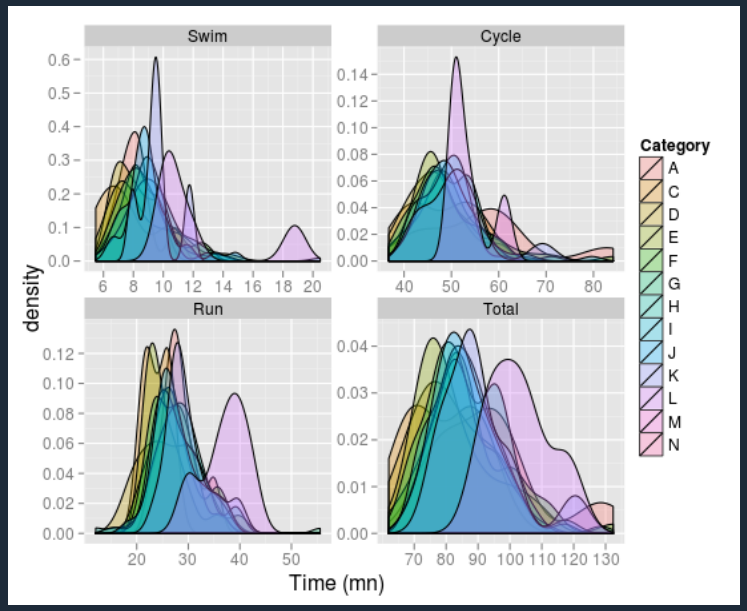
- For yourself: data visualization should be the first step in your analysis process after data collection

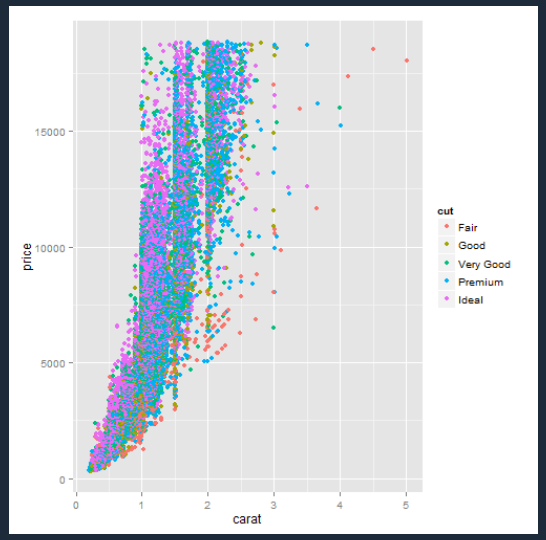
Example: Anscombe's Quartet Data Set

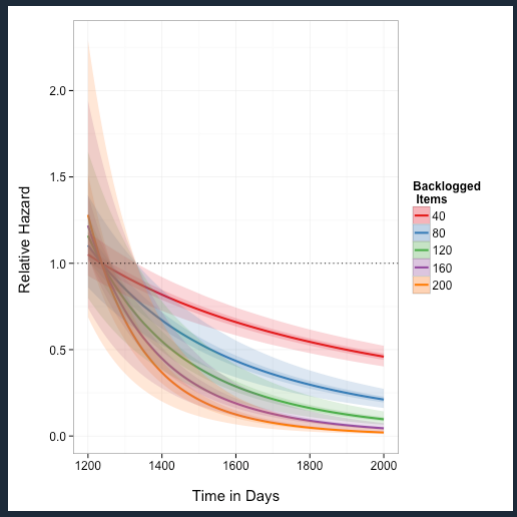


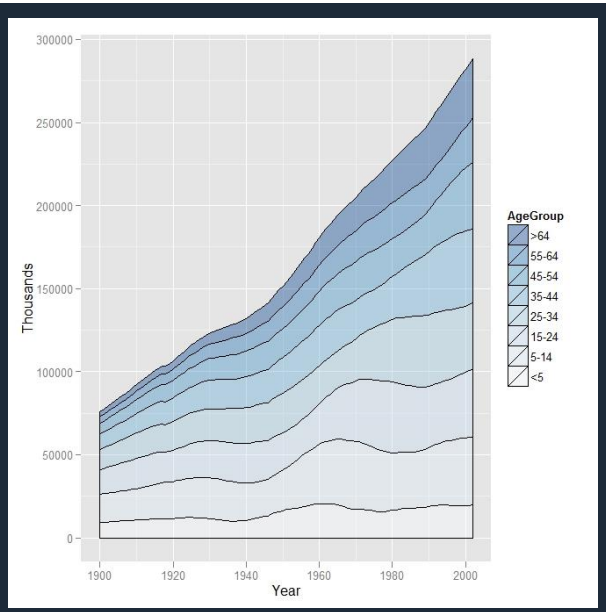
- For others: analysis without clear communication is worthless

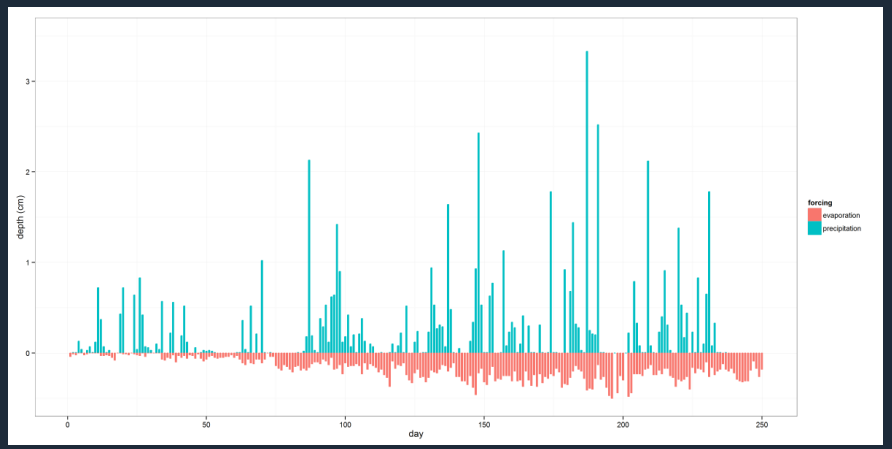
Clear data is beautiful











The qplot() function

```
>?qplot
    # This accesses the R documentation for any function
```

qplot {ggplot2} R Documentat

Quick plot

Description

`qplot` is the basic plotting function in the ggplot2 package, designed to be familiar if you're used to `plot` from the base package. It is a convenient wrapper for creating a number of different types of plots using a consistent calling scheme. See http://had.co.nz/ggplot2/book/qplot.pdf for the chapter in the `ggplot2` book which describes the usage of `qplot` in detail.

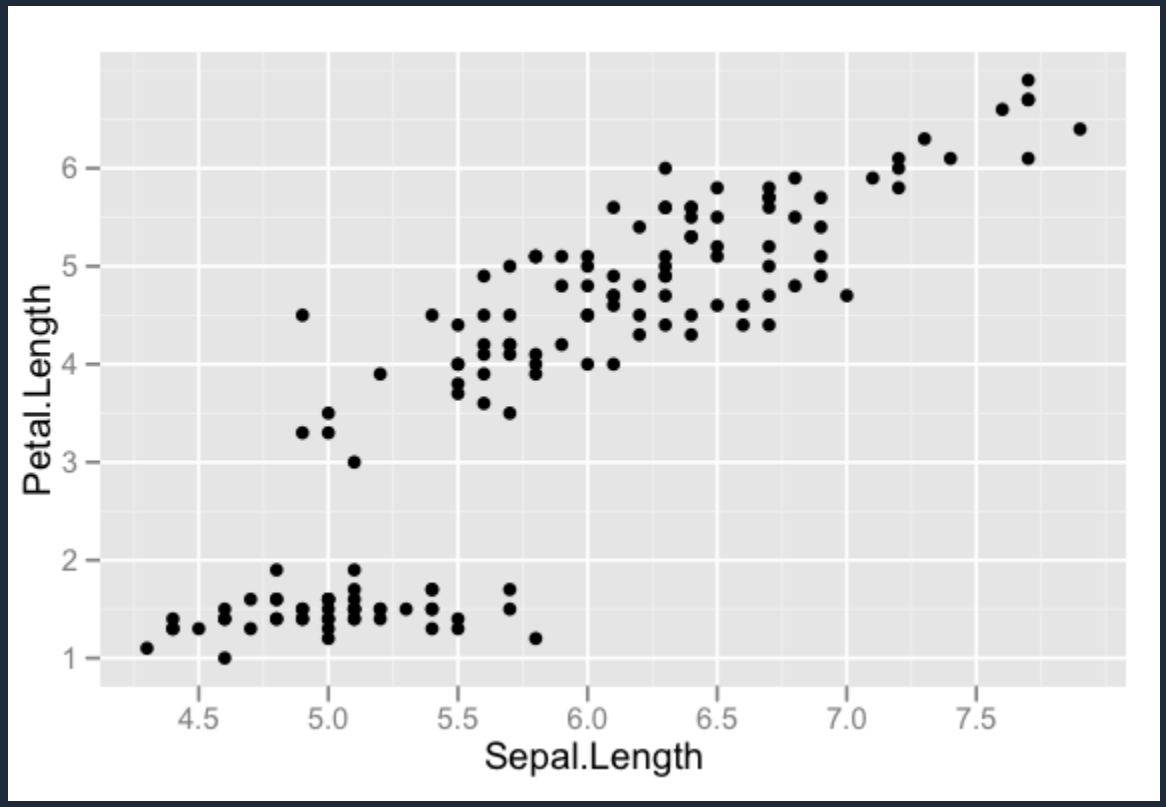
Usage

```
qplot(x, y = NULL, ..., data, facets = NULL, margins = FALSE,
  geom = "auto", stat = list(NULL), position = list(NULL), xlim = c(NA,
  NA), ylim = c(NA, NA), log = "", main = NULL,
  xlab = deparse(substitute(x)), ylab = deparse(substitute(y)), asp = NA)
```

qplot

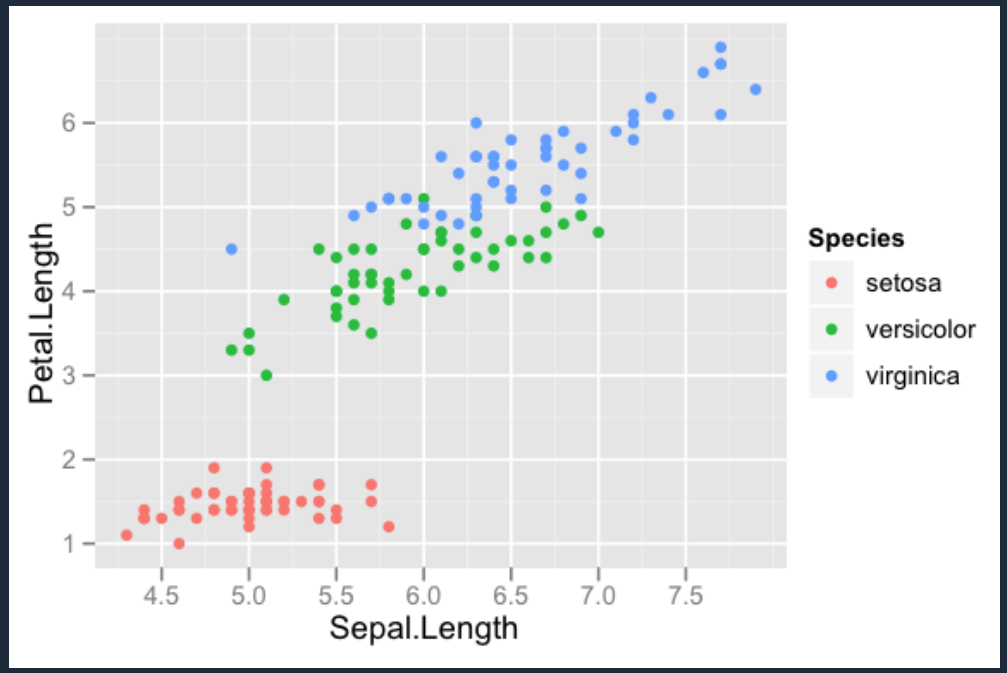
Use the head() function to peek at data

```
>head(iris)
>qplot(Sepal.Length, Petal.Length,
data = iris)
```



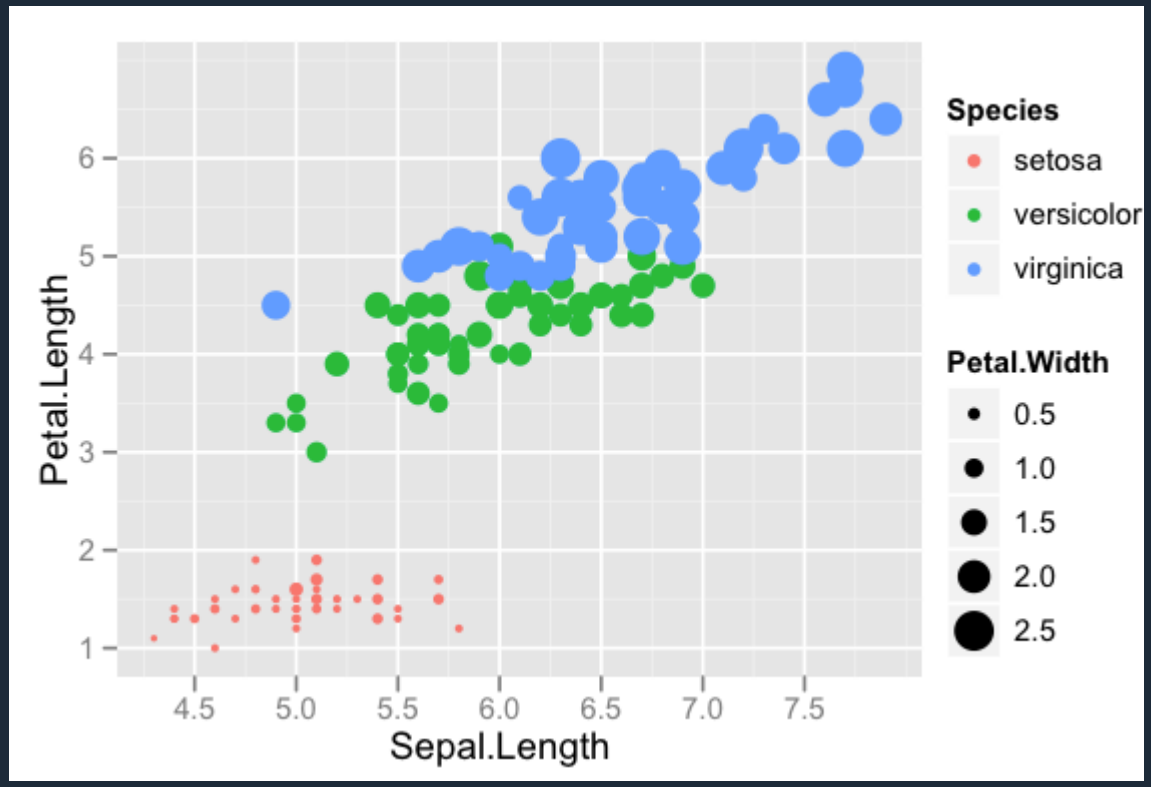
qplot: color

```
>qplot(Sepal.Length, Petal.Length,
data = iris, color = Species)
```



qplot: size

```
>qplot(Sepal.Length, Petal.Length,
data = iris, color = Species, size =
Petal.Width)
```

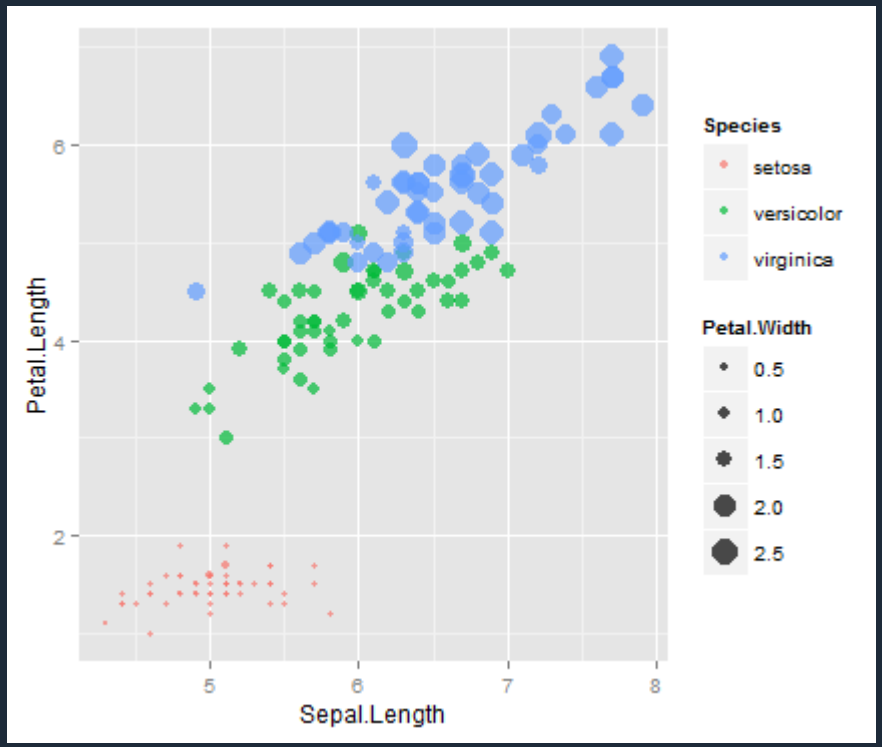


qplot: alpha

```
>qplot(Sepal.Length, Petal.Length,
data = iris, color = Species, size =
Petal.Width, alpha = I(.7))
```

- Between 0(transparent) and 1(opaque)

alpha = I(.1))

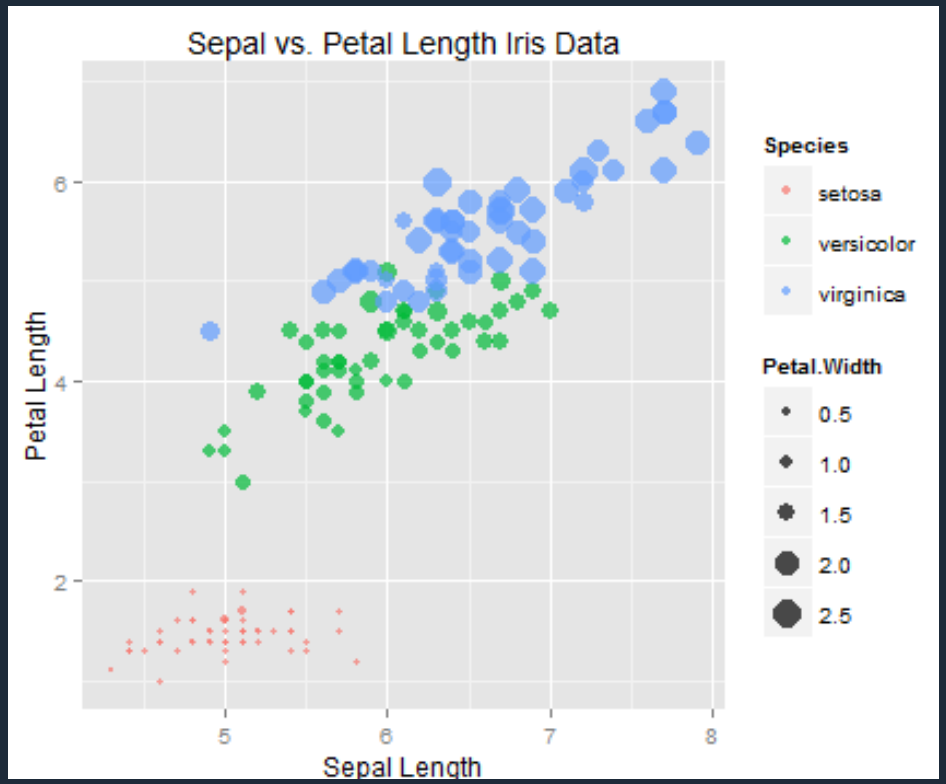


alpha = I(.7))

BAC

qplot: labeling

```
>qplot(Sepal.Length, Petal.Length, data = iris,
color = Species, size = Petal.Width, alpha =
I(.7), xlab = "Sepal Length", ylab = "Petal
Length", main = "Sepal vs. Petal Length Iris
Data")
```



qplot: geom

- point (default)
- bar
- line
- boxplot
- histogram
- path
- dotplot
- smooth
- density

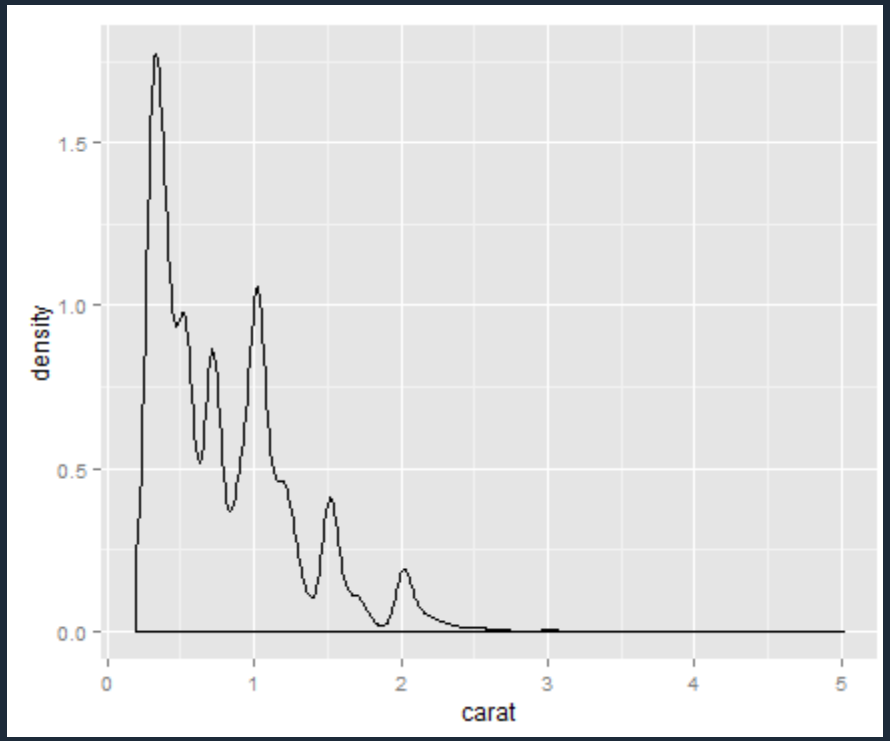
Example:

```
>qplot(Sepal.Length, Petal.Length,
data = iris, geom = "point")
```

qplot: geom density

Example:

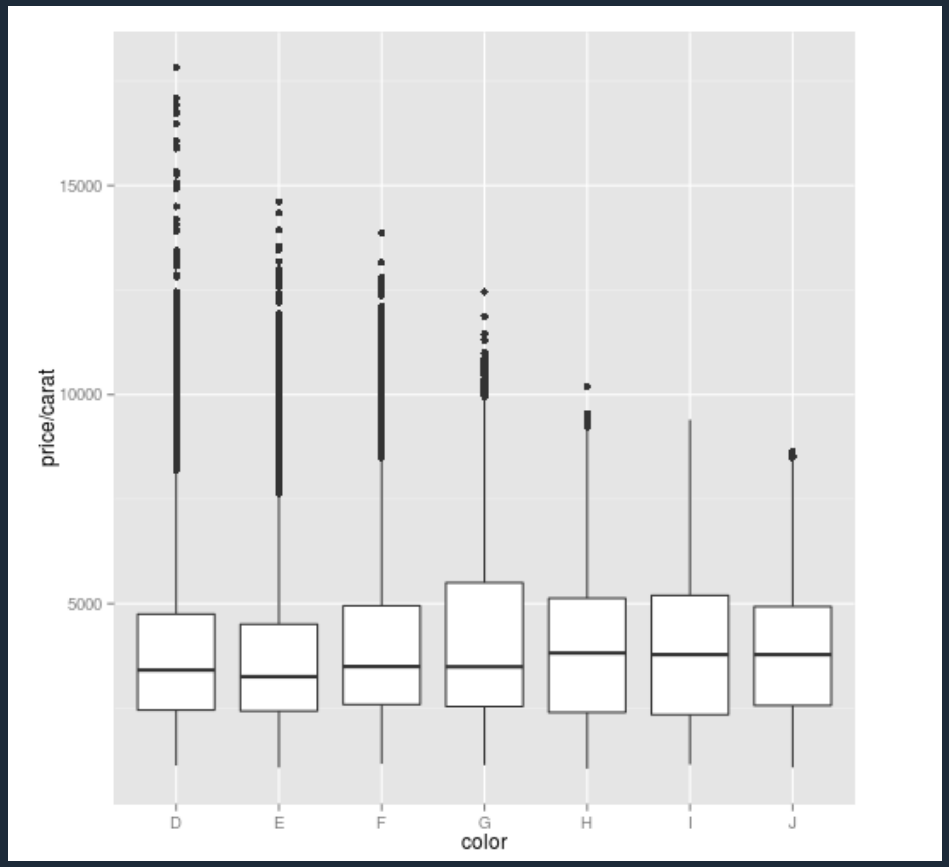
```
>qplot(carat, data = diamonds, geom =
"density")
```



qplot: geom boxplot

Example:

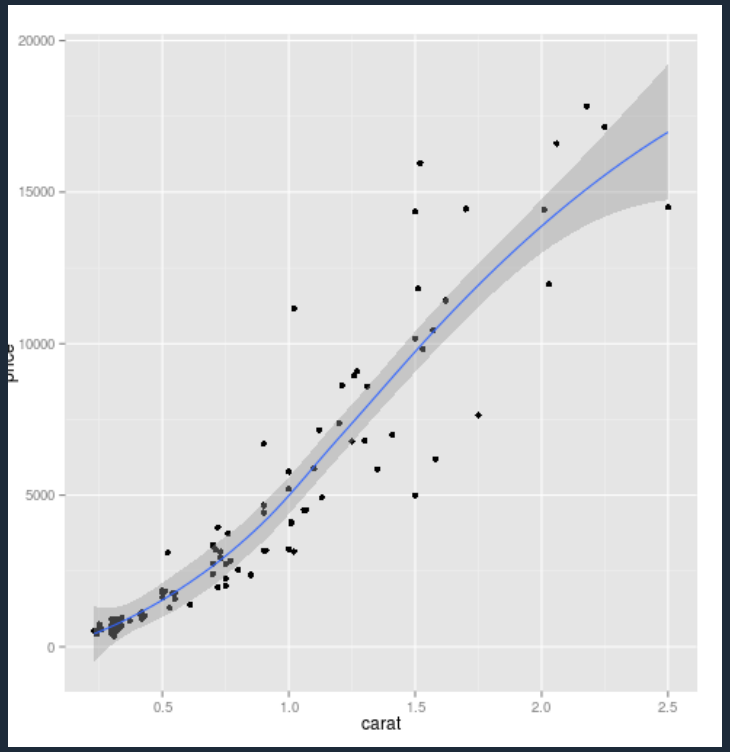
```
>qplot(color, price / carat, data = diamonds, geom = "boxplot")
```



qplot: geom smooth

Example:

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

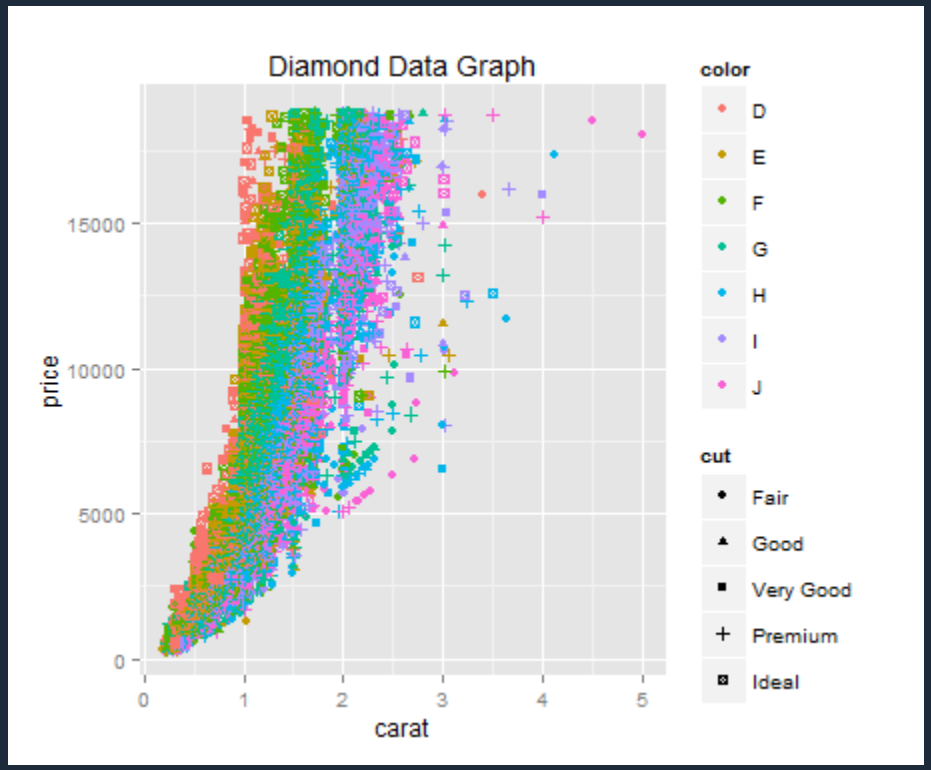


Make this graph:

Data set is diamonds

Start by observing data:

```
>head(diamonds)
```



Hints:

Start with:

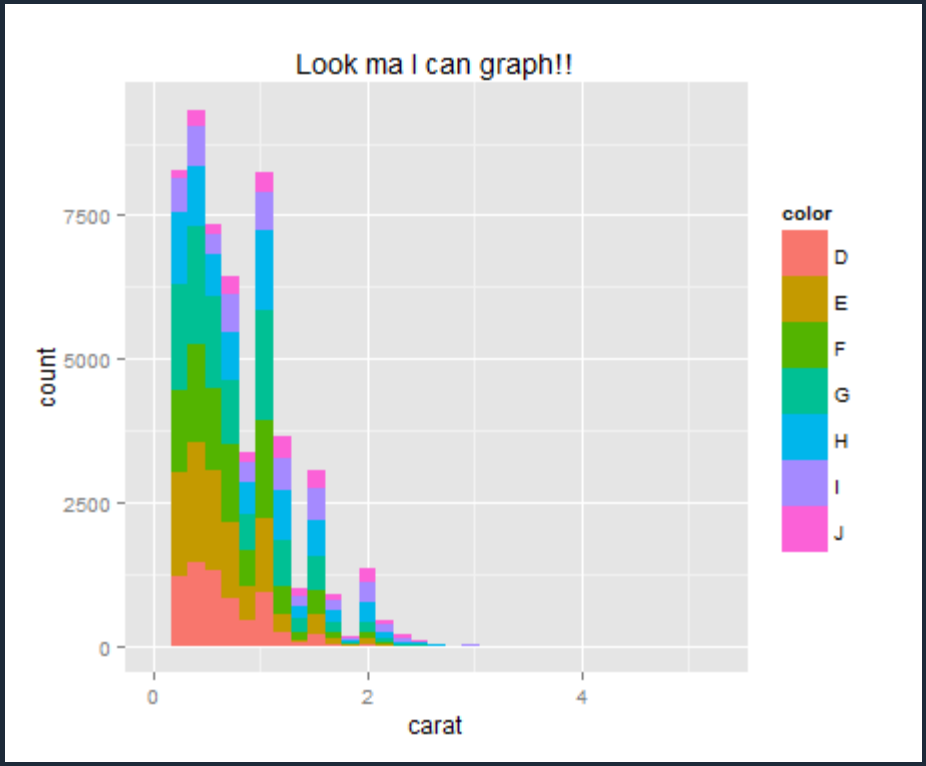
```
>qplot(carat,price,data=diamonds)
```

Check off list:

- Title []
- Color []
- Shape []

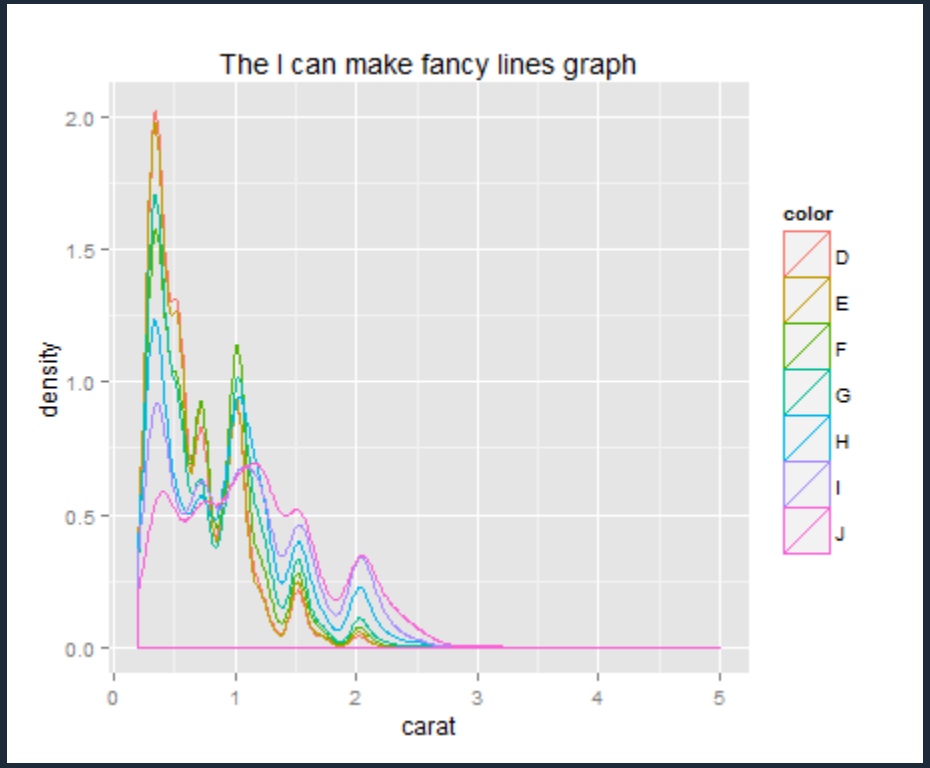
Make this graph:

Hint: geom is histogram



Make this graph:

No hints!!



The ggplot() function

- ggplot() is another function in the ggplot2 package
- It is more powerful and flexible than qplot
- If you're more familiar with R:
  - If using data.frame, easier to use ggplot()
  - If using separate variables, easier to use qplot()

- In general:
- qplot – for quick plots
- ggplot – for fine, granular control of everything

Dataset setup

Download dataset:

Go to bit.ly/bacdata > ggplot2 workshop > download Chicago-nmmaps.csv

Copy paste & run this code:

```
nmmaps<-read.csv("chicago-nmmaps.csv", as.is=T)

nmmaps$date<-as.Date(nmmaps$date)

nmmaps<-nmmaps[nmmaps$date>as.Date("1996-12-
31"),]

nmmaps$year<-substring(nmmaps$date,1,4)

head(nmmaps)
```

BAC

ggplot()

How it works:

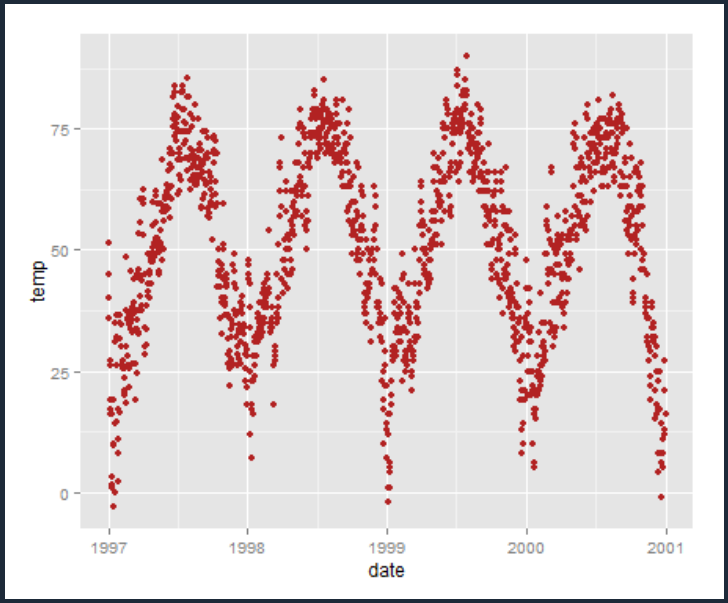
You save your code as a variable object.
You can add layers to the variable by the operand "+"

ggplot(): color

Run this code:

```
g<-ggplot(nmmaps, aes(date,temp)) +
geom_point(color="firebrick")

g
```



Check out the colors available in ggplot2!
http://www.stat.columbia.edu/~tzheng/files/Rcolor.pdf

BAC

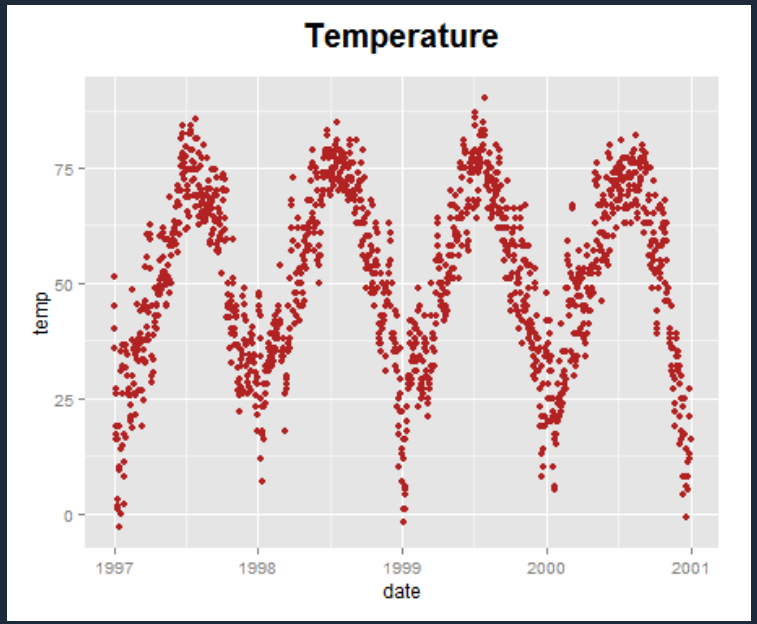
ggplot(): title

Title:

```
g<-g+ggtitle("Temperature")
```

Title adjustments:

```
g+theme(plot.title = element_text(size=20,
face="bold", vjust=2))
```



ggplot(): axis

X and Y Axis:

```
g<-g+labs(x="Date", y="Fahrenheit")
```

Axis Customization:

```
g+theme(
    axis.title.x = element_text(color="darkorchid", vjust=-0.35)
    axis.title.y = element_text(color = "darksalmon")
)
```

BAC

ggplot(): legends

Putting in seasonality:

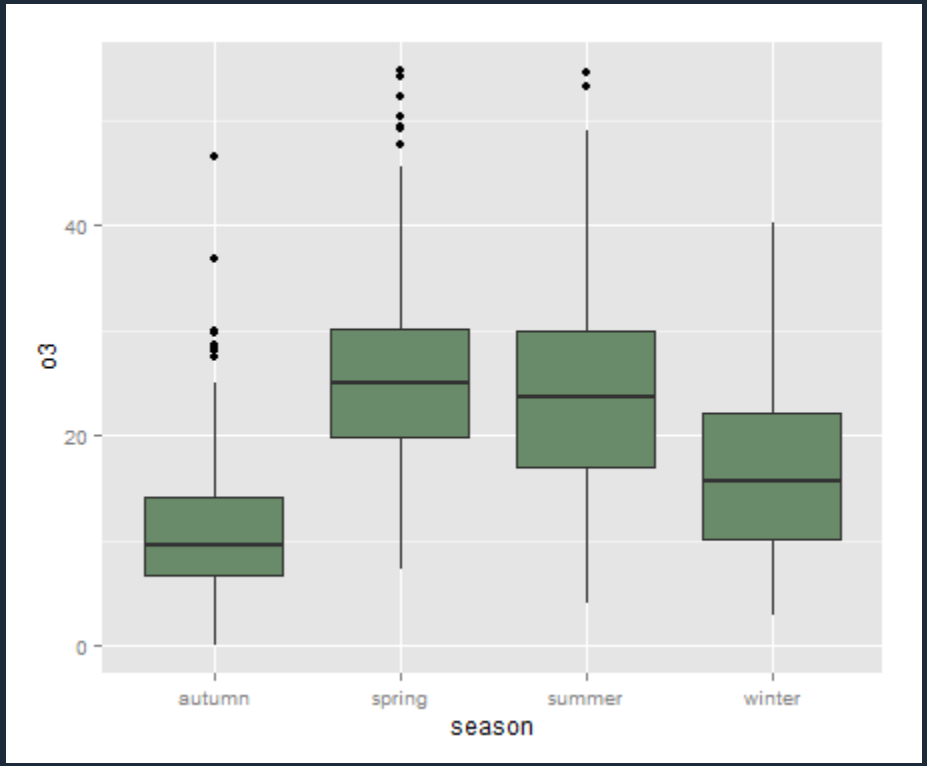
```
g<-ggplot(nmaps, aes(date, temp, color =
factor(season))) + geom_point()
```

Legend title:

```
g+theme(legend.title= element_text(colour =
“chocolate”, size=16, face = “bold”))
```

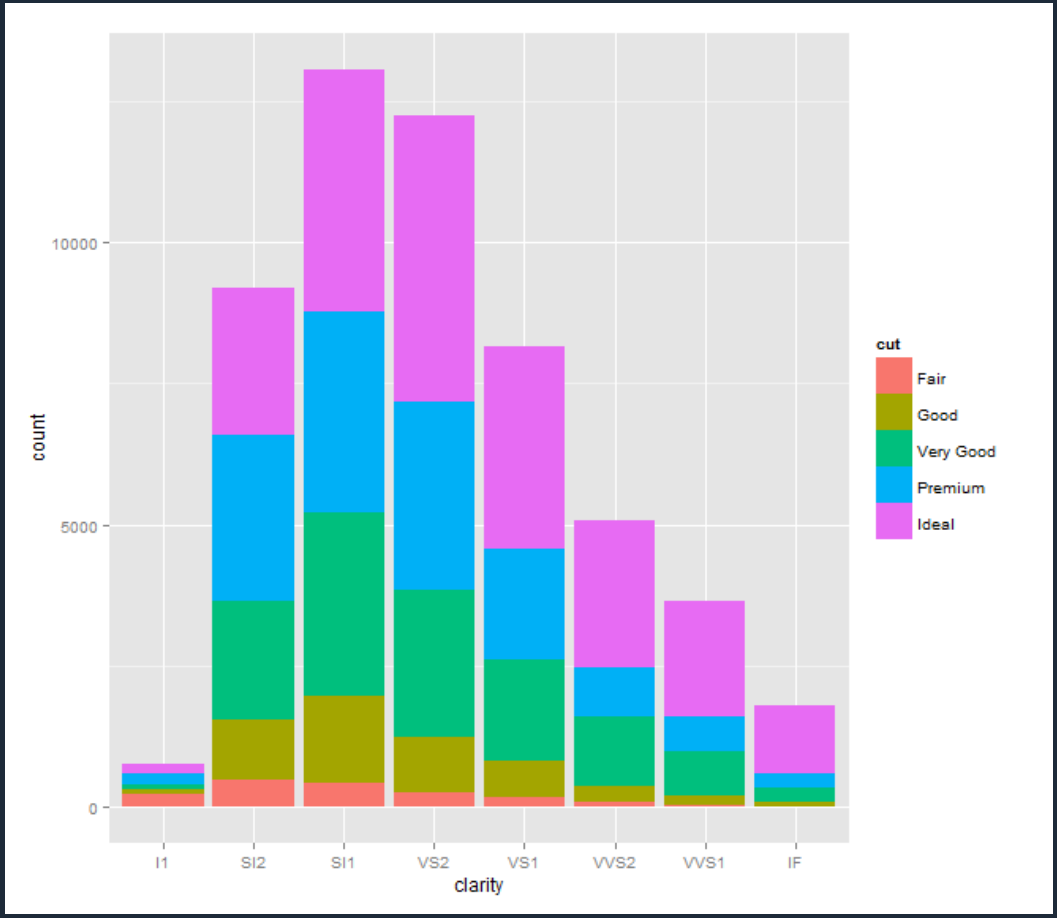
geom:boxplot

```
g<-ggplot(nmmaps, aes(x=season, y=o3))
g+geom_boxplot(fill="darkseagreen4")
```



geom: bar()

```
g <- ggplot(diamonds,aes(clarity,fill=cut))
g+geom_bar()
```



Contest

You have the next 20 minutes to:

1. Choose any dataset from R's library. List of datasets are at: https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/00Index.html
   No need to download. These are all already in R.
2. Make us a sexy graph!
3. Post your graph on https://www.facebook.com/SternBAC/

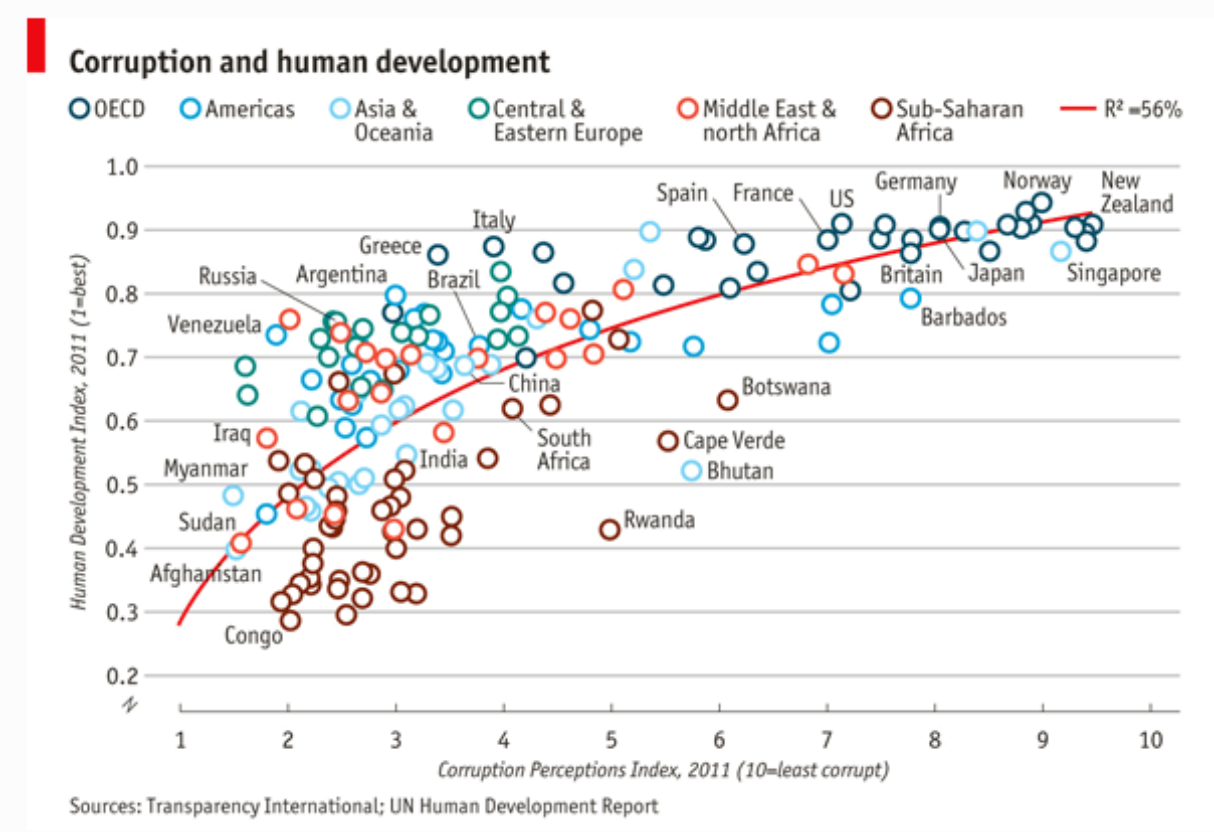
Select winners will be chosen & featured on our Facebook page!

BAC

Advanced ggplot2

Want to keep learning? Let us know in our workshop feedback:
Bit.ly/thurfdbk

Medium/Advanced ggplot2 workshop (Faceting, Statistical Transformations, Themes)
http://tutorials.iq.harvard.edu/R/Rgraphics/Rgraphics.html#orgheadline19



Instructions Outline

- [ ] add a trend line
- [ ] change the point shape to open circle
- [ ] change the order and labels of Region
- [ ] label select points
- [ ] fix up the tick marks and labels
- [ ] move color legend to the top
- [ ] title, label axes, remove legend title
- [ ] theme the graph with no vertical guides
- [ ] add model R^2 (hard)
- [ ] add sources note (hard)
- [ ] final touches to make it perfect (use image editor for this)

Acknowledgement & Sources

Edwin Chen
http://blog.echen.me/2012/01/17/quick-introduction-to-ggplot2/

ggplot2 Documentation
http://docs.ggplot2.org/dev/vignettes/qplot.html

ZevRoss
http://zevross.com/blog/2014/08/04/beautiful-plotting-in-r-a-ggplot2-cheatsheet-3/#quicksetup-the-dataset

The Institute for Quantitative Social Science & Harvard University
http://tutorials.iq.harvard.edu/R/Rgraphics/Rgraphics.html#orgheadline19

Hadley Wickham, Rice University
http://www.ceb-institute.org/bbs/wp-content/uploads/2011/09/handout_ggplot2.pdf

BAC

Answers

```
>qplot(carat, price, data=diamonds, colour = color, shape = cut,
main = “Diamond Data Graph”)
>qplot(carat, data=diamonds, geom=“density”, colour = color)
>qplot(carat, data=diamonds, geom=“histogram”, fill = color)
```