

# R Graphics with ggplot2

Dr. Aijun Zhang  
STAT3622 Data Visualization

19 September 2016

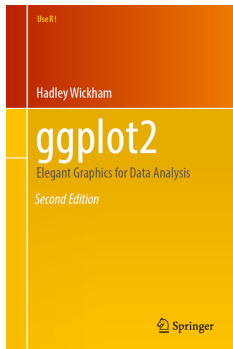


Department of 統計及精算學系  
**Statistics & Actuarial Science**

# Outline

- 1 What is ggplot2?
- 2 Quick Plots with qplot()
- 3 Layer-by-Layer ggplot()
- 4 Options and Themes

# What is ggplot2?



Wickham (2016; Springer 2ed)

- One of most commonly downloaded R packages, written by Hadley Wickham
- Based on the **Grammar of Graphics** by Wilkinson (2005; Springer 2ed)
- **base** → **lattice** → **ggplot2**
- *“ggplot2, started in 2005, is an attempt to take the good things about base and lattice graphics and improve on them with a strong underlying model”*
- ggplot2 version 2.x.x since Dec.2015

# Grammar of Graphics (GG)

*"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. Facetting can be used to generate the same plot for different subsets of the dataset. It is the combination of these independent components that make up a graphic."*

*Quote from the ggplot2 book  
based on Wilkinson (2005)*

# What is ggplot2?

- Installation: `install.packages("ggplot2")` on R  $\geq 3.1$

## ggplot2: An Implementation of the Grammar of Graphics

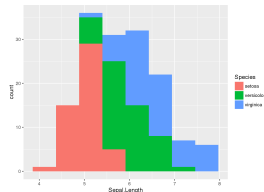
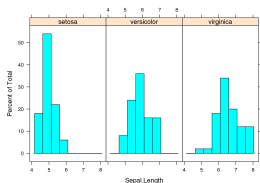
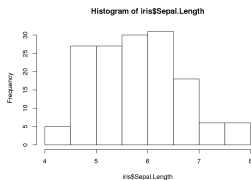
An implementation of the grammar of graphics in R. It combines the advantages of both base and lattice graphics: conditioning and shared axes are handled automatically, and you can still build up a plot step by step from multiple data sources. It also implements a sophisticated multidimensional conditioning system and a consistent interface to map data to aesthetic attributes. See <http://ggplot2.org> for more information, documentation and examples.

Version: 2.1.0  
Depends: R ( $\geq 3.1$ )

Source: <https://cran.r-project.org/>

- The package provides two ways/levels to build graphs:
  - **qplot()** — **quick plot**, supplies many defaults
  - **ggplot()** — **grammar of graphics plot**, allows more control
- Online documentation at <http://docs.ggplot2.org/>
- Remember to download the RStudio Cheatsheet from [here](#)

# First Impression of Base, Lattice and ggplot2 Graphs



```
hist(iris$Sepal.Length) # Base
```

```
histogram(~Sepal.Length|Species, iris) # Lattice
```

```
ggplot(iris, aes(x=Sepal.Length, fill=Species))  
  + geom_histogram(bins=8) # ggplot2
```

# Outline

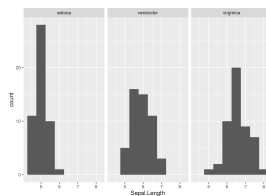
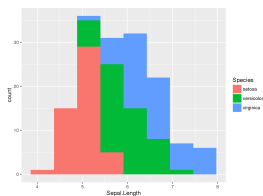
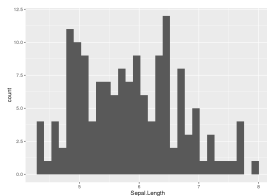
- 1 What is ggplot2?
- 2 Quick Plots with qplot()
- 3 Layer-by-Layer ggplot()
- 4 Options and Themes

# Quick Plots with qplot()

- `qplot()` may create a quick plot with minimum typing.
- Automatic use of default settings to make life easier.
- It defines a plot in a single call with the basic syntax:  
`qplot(variables, [geom], dataset, options)`
- A sensible geom will be picked by default if it is not supplied.
- `qplot()` is analog to base `plot()`. Learning it is very quick!

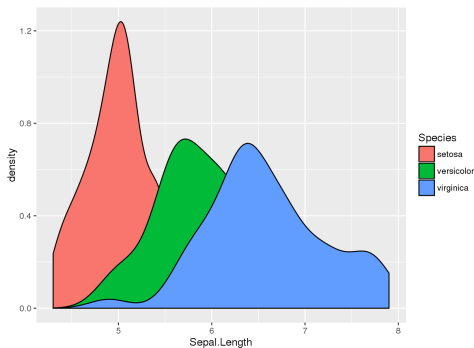


# Histogram



```
qplot(Sepal.Length, geom="histogram", data = iris) #
  histogram, default bins=30
qplot(Sepal.Length, data = iris, fill = Species, bins
      =8) # options: fill, bins/binwidth
qplot(Sepal.Length, data = iris, facets = .~Species,
      binwidth = 0.5) # options: facets
```

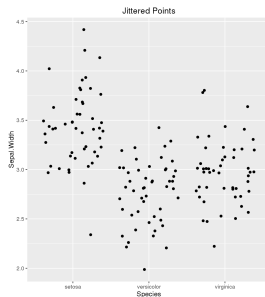
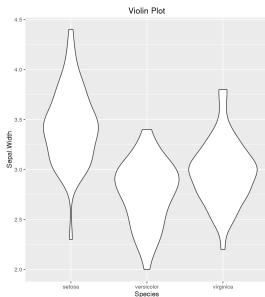
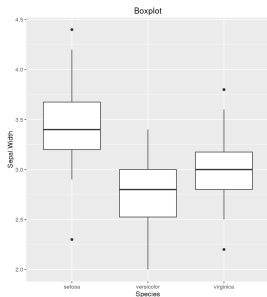
# Density Plot



```
qplot(Sepal.Length, geom = "density", data = iris,  
      fill = Species) # density
```

Compare it with the grouped histogram (in counts) in previous page.

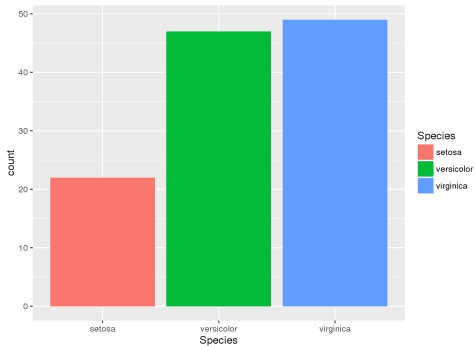
# Boxplot



```
qplot(Species, Sepal.Width, geom="boxplot", data =  
      iris, main="Boxplot")  
qplot(Species, Sepal.Width, geom="violin", data = iris  
      , main="Violin Plot")  
qplot(Species, Sepal.Width, geom="jitter", data = iris  
      , main="Jittered Points")
```

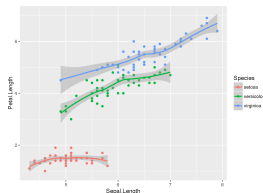
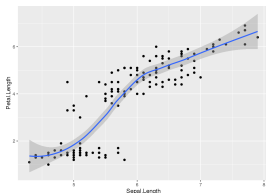
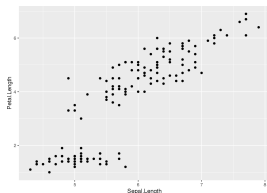
# Bar Chart

Bar chart for a categorical/discrete variable



```
qplot(Species, geom="bar", data = subset(iris, Sepal.Length>5), fill=Species)
```

# Scatterplot



```
qplot(Sepal.Length, Petal.Length, geom = "point", data = iris) # Scatterplot
qplot(Sepal.Length, Petal.Length, geom = c("point", "smooth"), data = iris) # With smooth
qplot(Sepal.Length, Petal.Length, geom = c("point", "smooth"), data = iris, color = Species) # grouping
```

# Outline

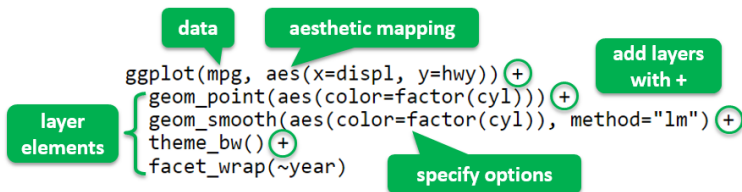
- 1 What is ggplot2?
- 2 Quick Plots with qplot()
- 3 Layer-by-Layer ggplot()**
- 4 Options and Themes

# Basic Components of ggplot2

- **Data:** a data.frame to visualize
- **Aesthetics:** mapping variables of the data to aesthetic attributes (position, size, shape, color, fill, transparency, ...)
- **Scales:** mapping values of the data to visual values for each aesthetic (e.g. position, color, fill and shape scales)
- **Geometric objects:** point, line, polygon, histogram, quantile, bar, ...
- **Statistical transformations:** bin, boxplot, density, contour, function, ...
- **Coordinate system:** Cartesian, polar, map projection, ...
- **Facet:** display split data in multi-panels (aka conditioning)
- **Theme:** control non-data visual elements (title, axes, tick, ...)

# Layer-by-Layer ggplot()

- `ggplot()` builds a plot layer by layer, with the syntax:

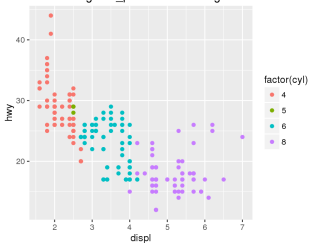


- `ggplot()` provides more control than `qplot()`;
- `ggsave()` the `last_plot()` with formats `.png`, `.jpg`, `.pdf`, ...

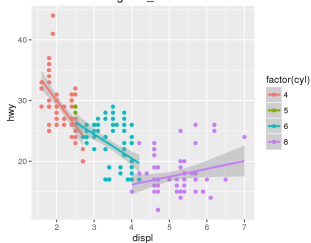


# Layer-by-Layer Scatterplot

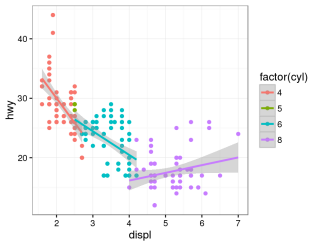
Add geom\_point with coloring



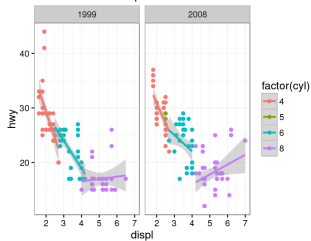
Add geom\_smooth



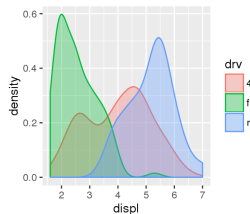
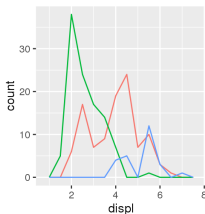
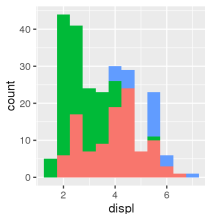
Add a theme



Add multipanel facets

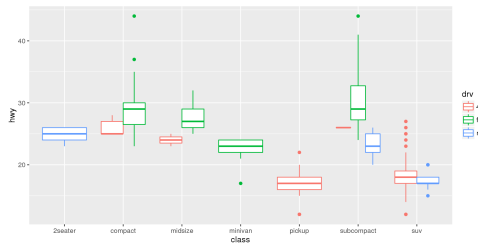
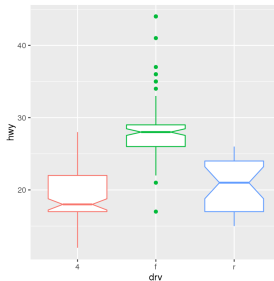
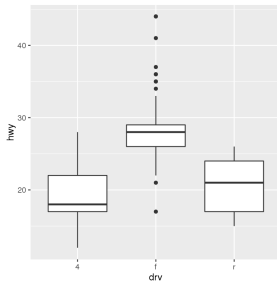


# Histogram, Freqpoly and Density Plots

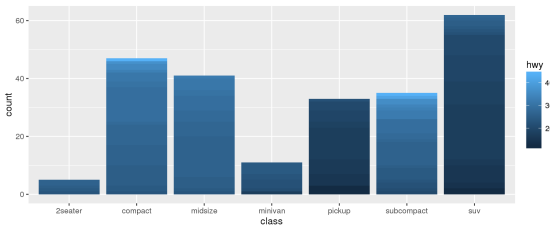
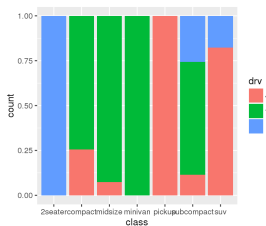
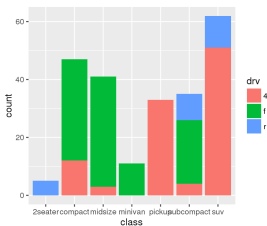


```
p1 = ggplot(mpg, aes(displ, fill=drv)) +
  geom_histogram(binwidth = 0.5)
p2 = ggplot(mpg, aes(displ, colour=drv)) +
  geom_freqpoly(binwidth = 0.5)
p3 = ggplot(mpg, aes(displ, fill=drv, colour=drv)) +
  geom_density(alpha=1/3)
grid.arrange(p1, p2, p3, ncol=3) # require(gridExtra)
```

# Boxplots



# Bar Charts

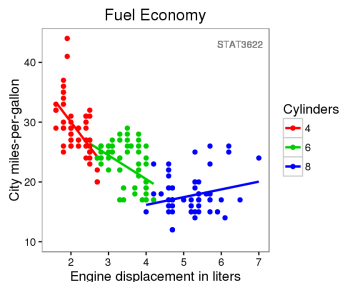
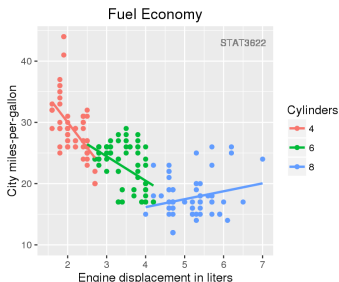


# Outline

- 1 What is ggplot2?
- 2 Quick Plots with qplot()
- 3 Layer-by-Layer ggplot()
- 4 Options and Themes**

# Options

Many options are available for polishing the graphs and making them sophisticated. To give only an example:



(Refer to the attached R markdown for the source codes.)

# Themes

- ggplot2 provides `theme_bw()`, `theme_minimal()`, ...
- More cool stuff are provided by the package `ggthemes`:  
`theme_economist()`, `theme_wsj()`, `theme_solarized()`, ...

