GEOG 6000
Advanced Geographical Data Analysis
0401: Advanced graphics with **ggplot2**

Simon Brewer

Geography Department
University of Utah
Salt Lake City, Utah 84112

simon.brewer@geog.utah.edu

September 27, 2017

Objectives

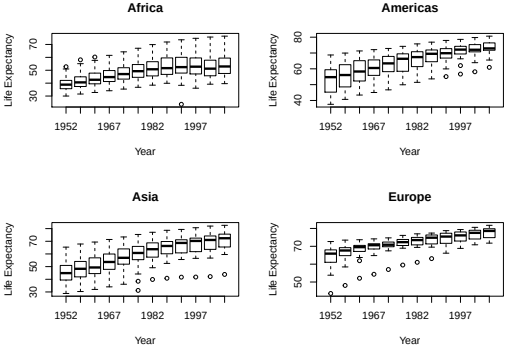
- Introduce **ggplot2**
- Dataframes for **ggplot2**
- Examples

ggplot2

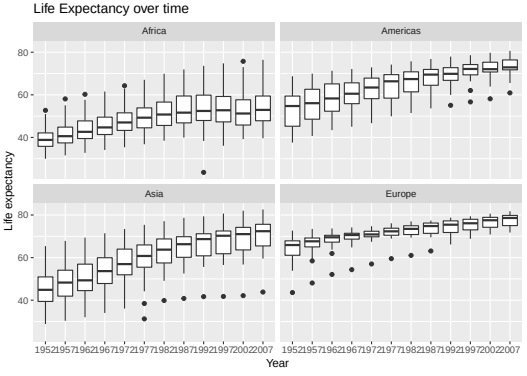
- Based on Leland Wilkinson's Grammar of Graphics
  - All data figures can be represented by the same *grammar*
- Adapted for R by Hadley Wickham
- Provides much easier methods for comparative plots

Base graphics vs. ggplot2

Base graphics: 8 lines

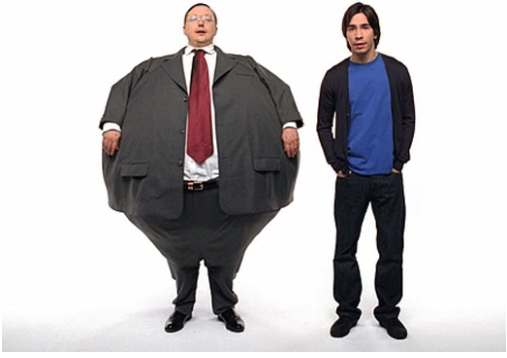


ggplot2: 1 (quite complex) line



Data frames for ggplot2

- Data is often presented as short and fat tables
- Plotting is easier with tall and thin data frames
  - Each variable forms a column
  - Each observation forms a row



Data frames for ggplot2

Short/fat table: good for presenting results

	TreatA	TreatB
Jane Smith	-	2
John Doe	16	11
Mary Jones	3	1

Tall/thin dataframe: preferred for plotting

Name	Treat	Result
Jane Smith	a	-
John Doe	a	16
Mary Jones	a	3
Jane Smith	b	2
John Doe	b	11
Mary Jones	b	1

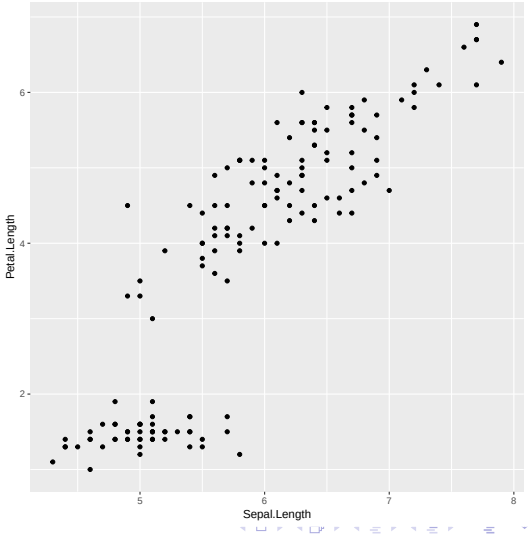
- Support package **reshape2** includes functions to transform between these layouts
- `cast`: thin data frame to table
- `melt`: table to thin data frame

Grammar of Graphics

- Data: as data frame
- Aesthetic: variables used to control position, color, fill, etc
- Geometry: form of the plot, points, lines, bars, etc
- Scale: mapping values into computer values, log scaling, etc
- Statistics: summaries or transformation of data
- Facet: Groups used to split data into multiple graphs

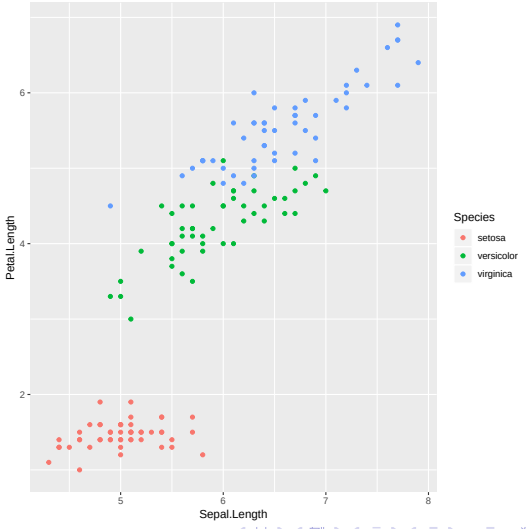
Simple scatterplot

```
myplot = ggplot(data=iris,
                aes(x=Sepal.Length, y=Petal.Length))
myplot + geom_point()
```



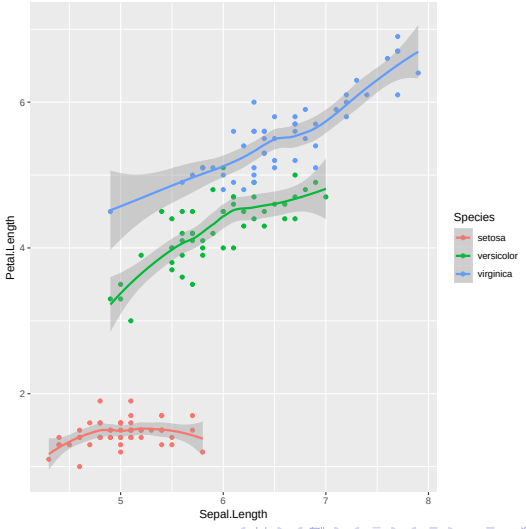
Simple scatterplot

```
myplot = ggplot(data=iris,
                aes(x=Sepal.Length, y=Petal.Length, col=Species))
myplot + geom_point()
```



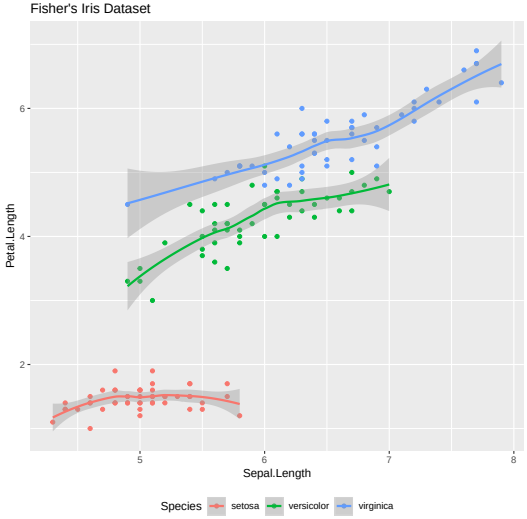
Simple scatterplot

```
myplot = ggplot(data=iris,
                aes(x=Sepal.Length, y=Petal.Length, col=Species))
myplot = myplot + geom_point() + geom_smooth()
print(myplot)
```



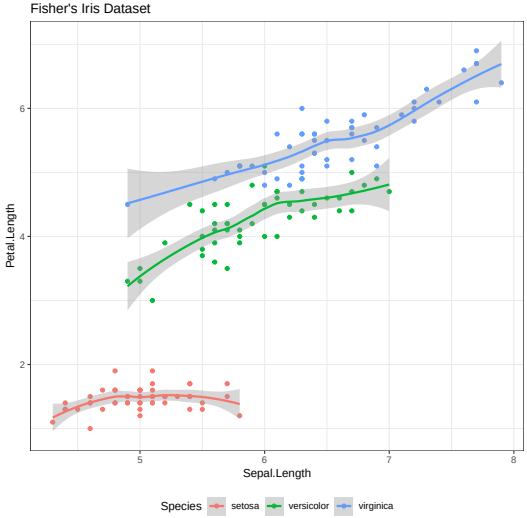
Themes and title

```
myplot = ggplot(data=iris,
                aes(x=Sepal.Length, y=Petal.Length, col=Species))
myplot = myplot + geom_point() + geom_smooth()
myplot = myplot + ggtitle("Fisher's Iris Dataset") +
  theme(legend.position="bottom")
print(myplot)
```



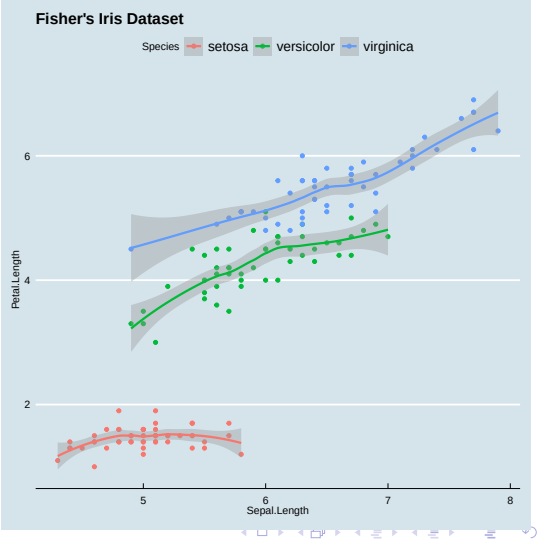
Themes and title

```
myplot = ggplot(data=iris,
                aes(x=Sepal.Length, y=Petal.Length, col=Species))
myplot = myplot + geom_point() + geom_smooth()
myplot = myplot + ggtitle("Fisher's Iris Dataset") +
  theme_bw() + theme(legend.position="bottom")
print(myplot)
```



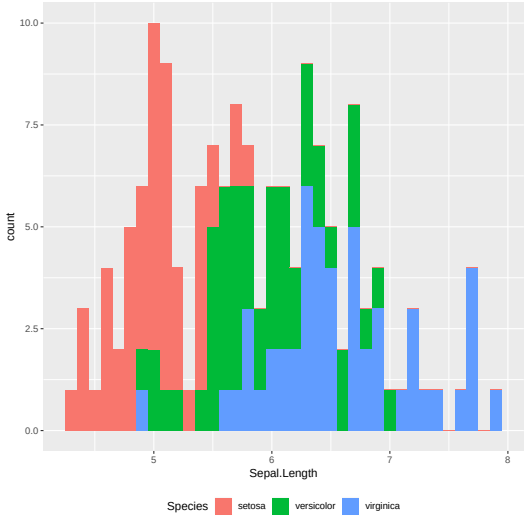
Themes and titles

```
require(ggthemes)
myplot = ggplot(data=iris,
                aes(x=Sepal.Length, y=Petal.Length, col=Species))
myplot = myplot + geom_point() + geom_smooth()
myplot = myplot + ggtitle("Fisher's Iris Dataset") +
  theme_economist()
print(myplot)
```



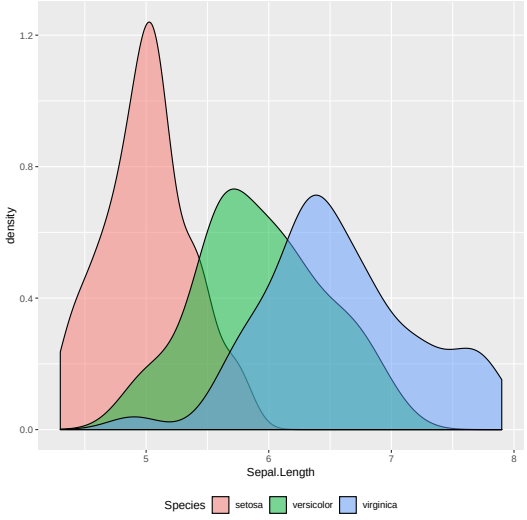
Histograms

```
myplot = ggplot(data=iris, aes(x=Sepal.Length, fill=Species))
myplot = myplot + geom_histogram(binwidth = 0.1) +
  theme(legend.position="bottom")
print(myplot)
```



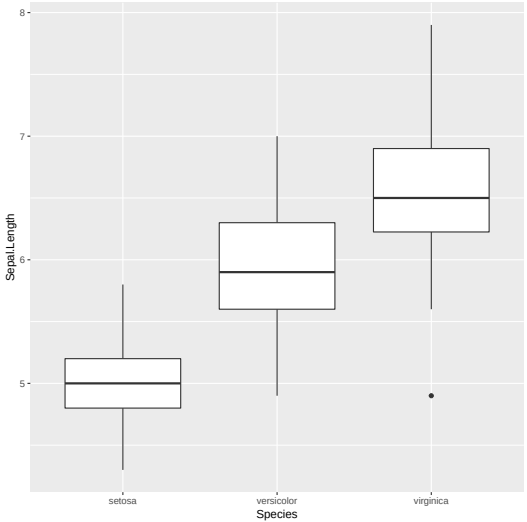
Kernel densities

```
myplot = ggplot(data=iris, aes(x=Sepal.Length, fill=Species))
myplot = myplot + geom_density(alpha=0.5) +
  theme(legend.position="bottom")
print(myplot)
```



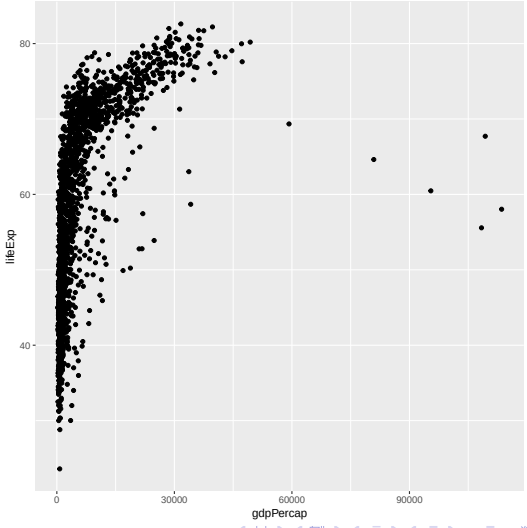
Boxplots

```
myplot = ggplot(data=iris, aes(x=Species, y=Sepal.Length))
myplot = myplot + geom_boxplot()
print(myplot)
```



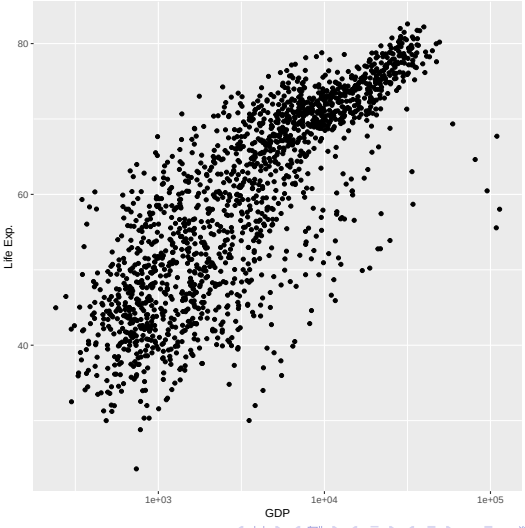
Scales

```
myplot = ggplot(data=gapdata, aes(x=gdpPercap, y=lifeExp))
myplot = myplot + geom_point()
print(myplot)
```



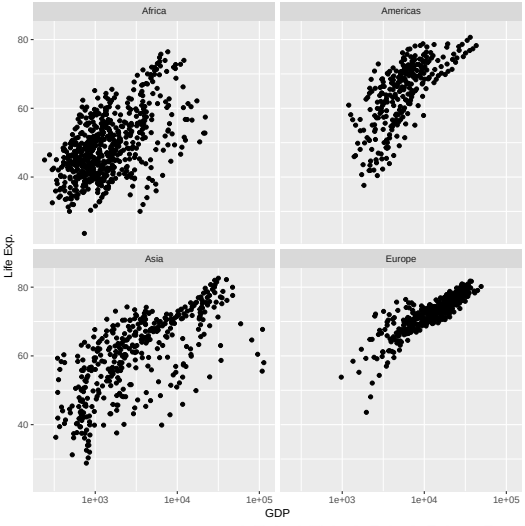
Scales

```
myplot = ggplot(data=gapdata, aes(x=gdpPercap, y=lifeExp))
myplot = myplot + geom_point() +
  scale_x_log10("GDP") + scale_y_continuous("Life Exp.")
print(myplot)
```



Facets

```
myplot = ggplot(data=gapdata, aes(x=gdpPercap, y=lifeExp))
myplot = myplot + geom_point() +
  scale_x_log10("GDP") + scale_y_continuous("Life Exp.")
myplot = myplot + facet_wrap(~continent)
print(myplot)
```



Facets

```
myplot = ggplot(data=gapdata, aes(x=gdpPercap, y=lifeExp))
myplot = myplot + geom_point() +
  scale_x_log10("GDP") + scale_y_continuous("Life Exp.")
myplot = myplot + facet_grid(year~continent)
myplot = myplot + geom_smooth(method="lm")
print(myplot)
```