An Introduction to R Graphics
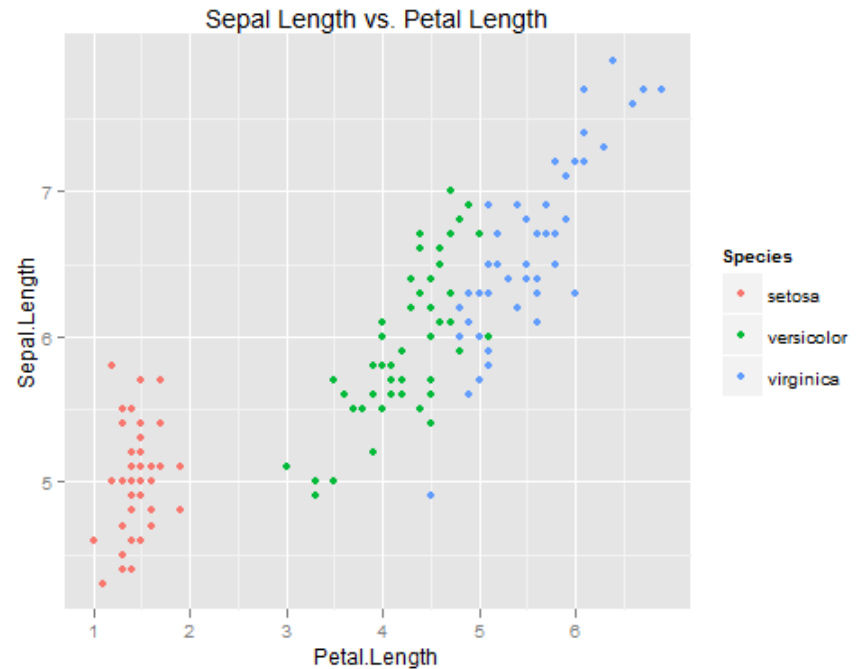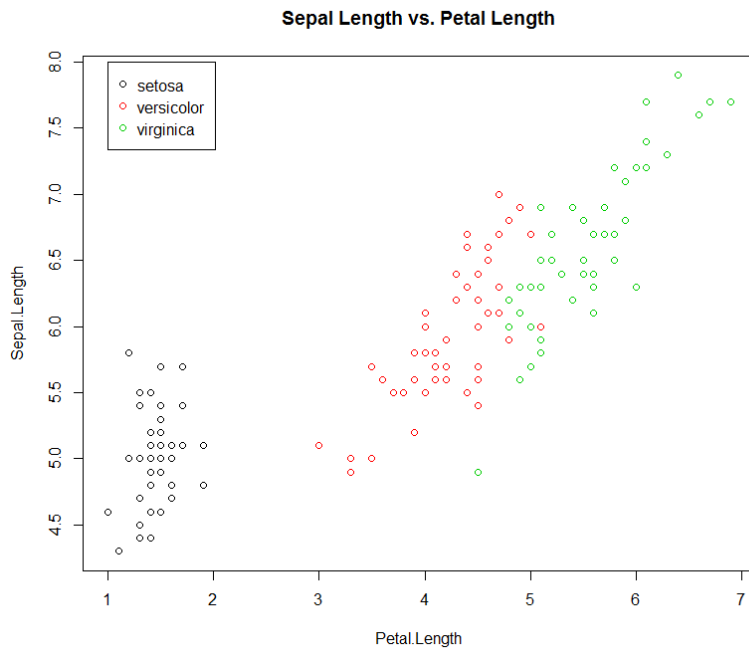
Why use R for graphics?

- Fast data exploration

- Easy automation and reproducibility

- Create publication quality figures

- Customisation of almost every aspect of the plot

- You can feel smug when people still use Excel

A quick recap on vectors and data frames

- Combine for vectors: c()

```
c(1, 6, 4, 7, 9, 4, 2)          # integers
c("A", "B", "C", "D")           # factors
```

- Reading in a data frame from a text file:

```
unicorns<-read.table("unicorns.txt", header=TRUE)
```

- Calling a variable from a data frame using $

```
> head(unicorns)
  birthweight  sex longevity
1    4.478424 Male         1
2    5.753458 Male         0
3    3.277265 Male         0
4    3.929379 Male         0
```

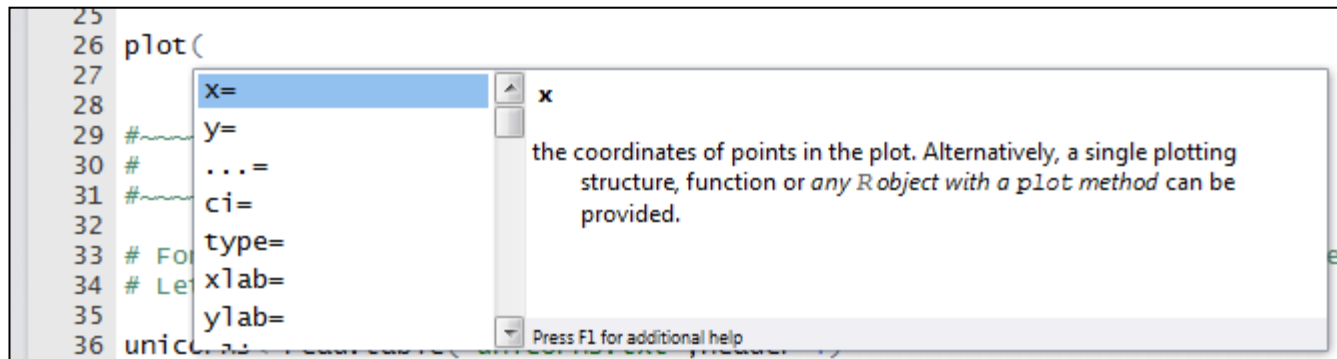To get birth weight alone: unicorns$birthweight

Getting help in R

- For more information on any command in R, use the question mark!

  > ?plot

- Use ?? to search all of the help documents

  > ??boxplot

- Use the tab key in RStudio

Overview

- Part 1: Base graphics

  – Graphical tools already included in R

  – Simple, fast, exploratory graphics

  – Important to know

- Part 2: `ggplot2`

  – More complex, higher quality graphics

  – Fashionable to know

  – Easier to master?

Histogram of Unicorn Birth Weight

Moomin Population Size on Ruissalo 1971 - 2001
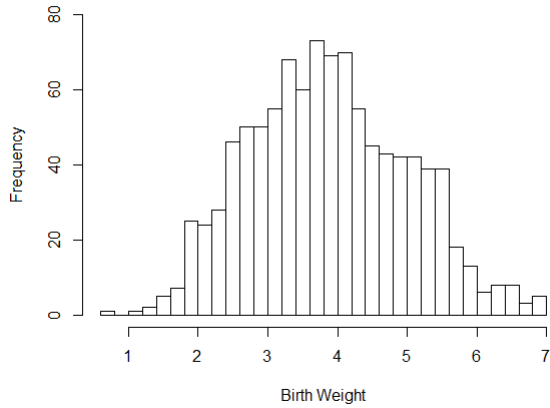
R2 = 0.896
P = 2.615e-15

Flower Characteristics in Iris

- setosa
- versicolor
- virginica
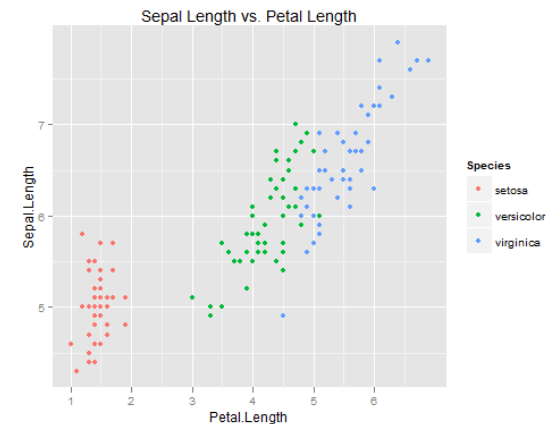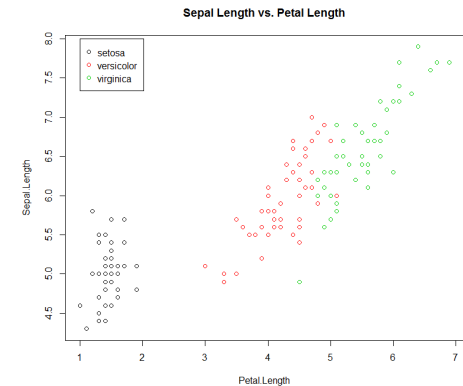
1. Basic Histogram
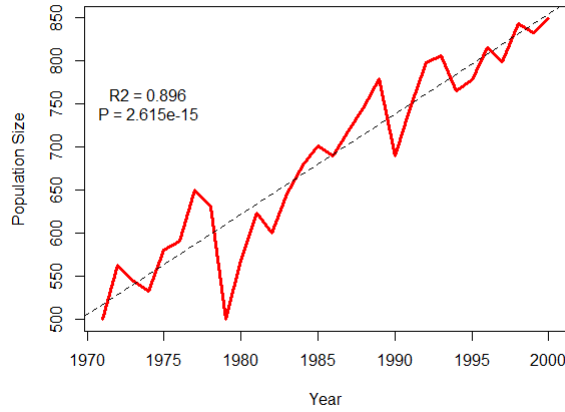
2. Line Graph with Regression
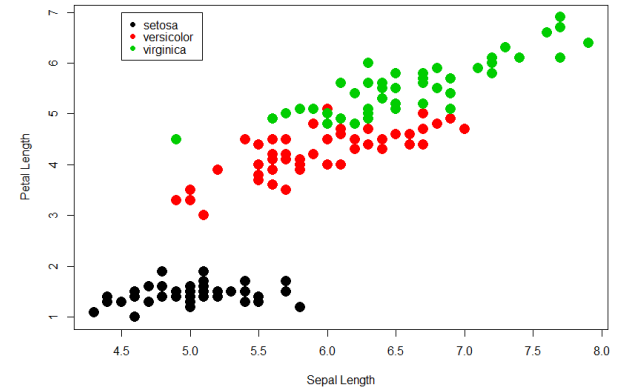
3. Scatterplot with Legend

Sepal Length by Species in Iris

Dragon Talon Length in the UK
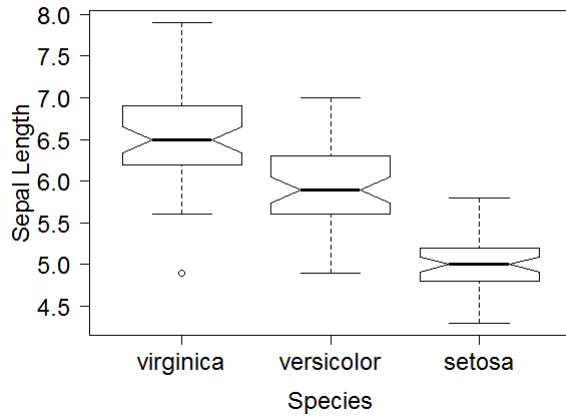
4. Boxplot with reordered/ formatted axes
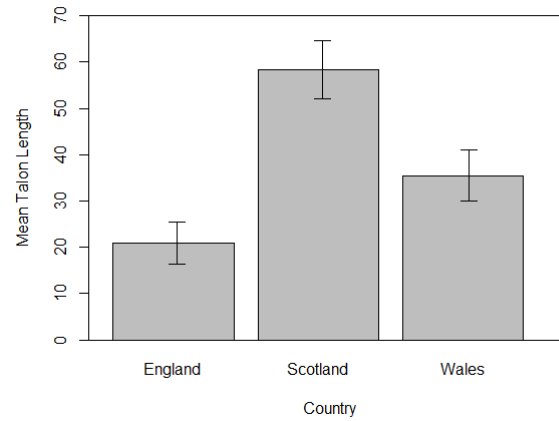
5. Boxplot with Error Bars

Data for Graphs: A Reminder

- See script…

1. *hist* function: Histogram

- Visual representation of data distribution

- Birth weight and longevity in unicorns

1. Basic Histogram
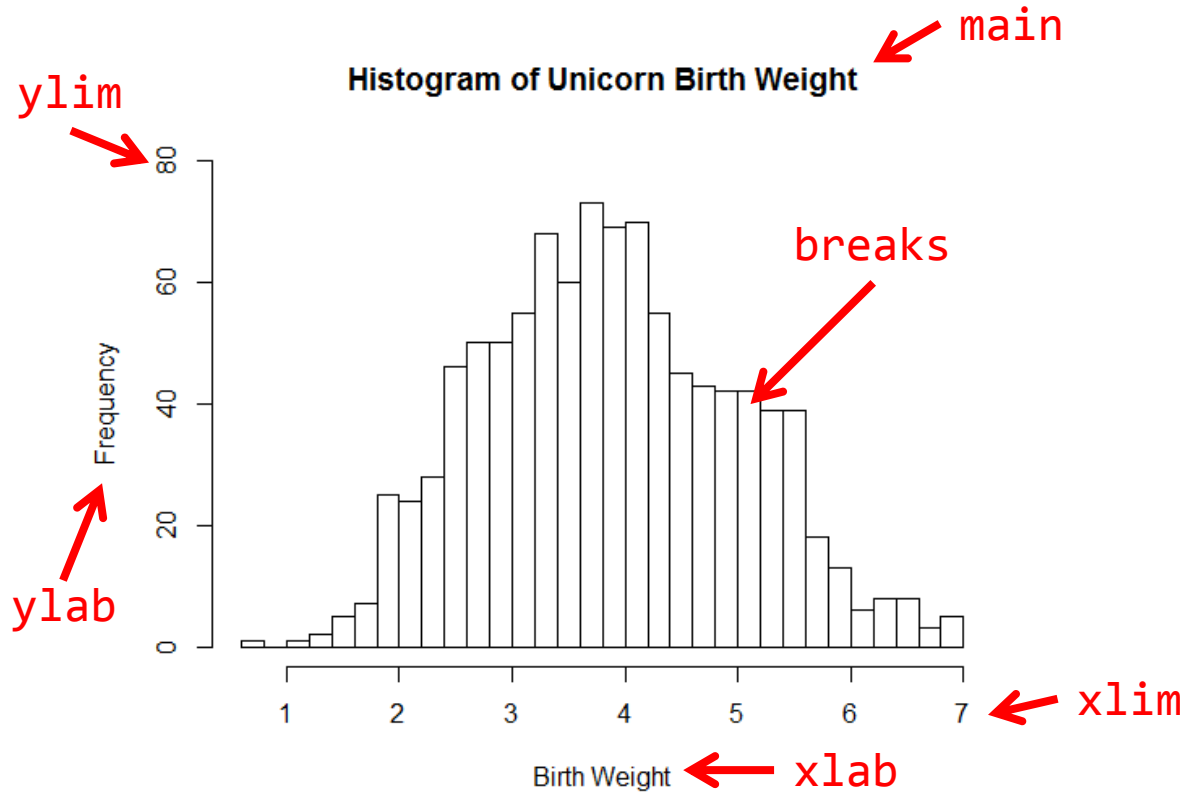

```
 99 #~~ FINAL PLOT:
100
101 hist(unicorns$birthweight,                          # x value
102     breaks = 40,                                     # number of cells
103     xlab = "Birth Weight",                           # x-axis label
104     main = "Histogram of Unicorn Birth Weight",      # plot title
105     ylim = c(0,80))                                  # limits of the y axis (min,max)
106
```

- Moomins are a common pest species in Finland
- Data on population density in Ruissalo from 1971 - 2000

2. Basic line graph with regression

Moomin Population Size on Ruissalo 1971 - 2001

text() → R2 = 0.896
          P = 2.615e-15

abline(fit1)
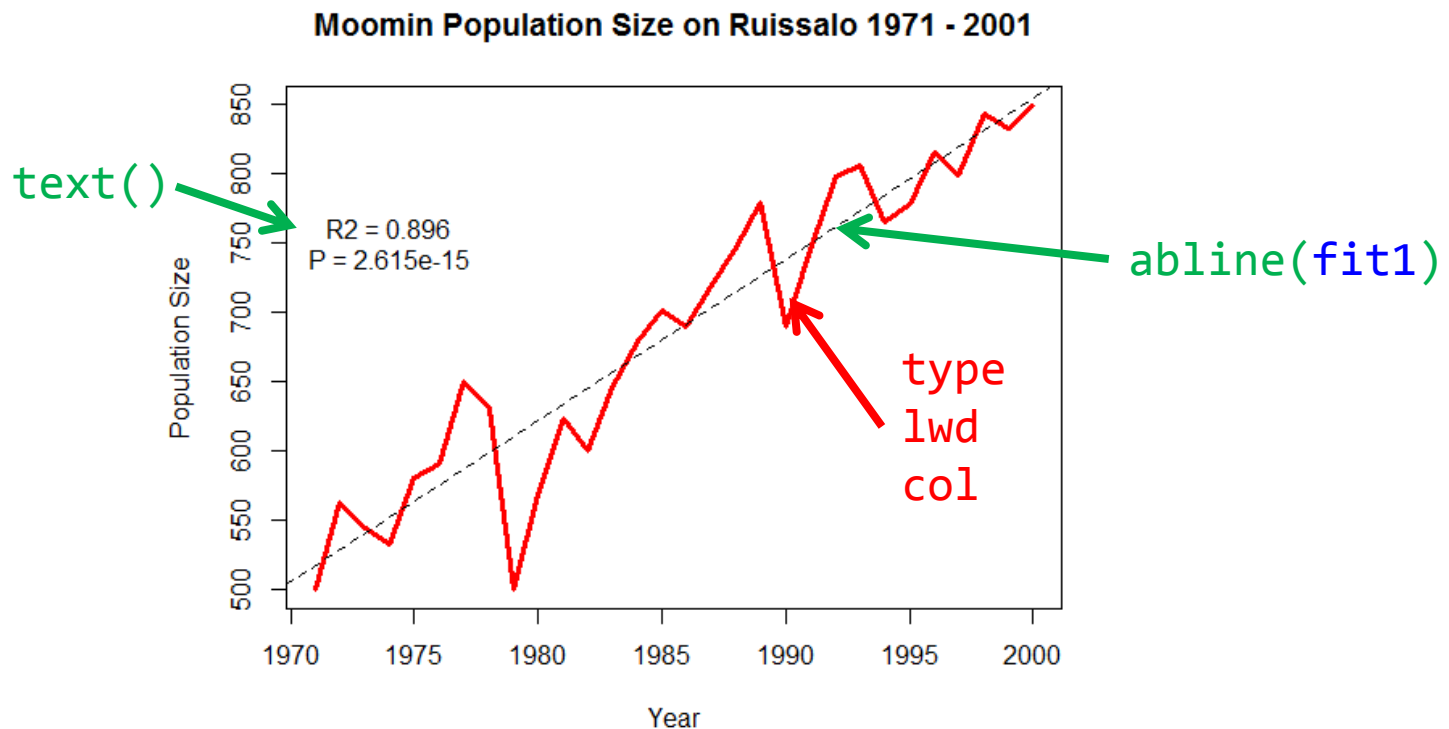
type
lwd
col

```
184 #~~ FINAL PLOT Script
185
186 plot(moomins$Year, moomins$PopSize,                              # x variable, y variable
187     type = "l",                                                  # draw a line graphs
188     col = "red",                                                 # red line colour
189     lwd = 3,                                                     # line width of 3
190     xlab = "Year",                                               # x axis label
191     ylab = "Population Size",                                    # y axis label
192     main = "Moomin Population Size on Ruissalo 1971 - 2001")     # plot title
193 fit1 <- lm (PopSize ~ Year, data = moomins)                      # carry out a linear regression
194 abline(fit1, lty = "dashed")                                     # add the regression line to the plot
195 text(x=1974,y=750,labels="R2 = 0.896\nP = 2.615e-15")           # add a label to the plot at coordinates (x,y)
196
```

3. *plot* function: Scatterplot with legend

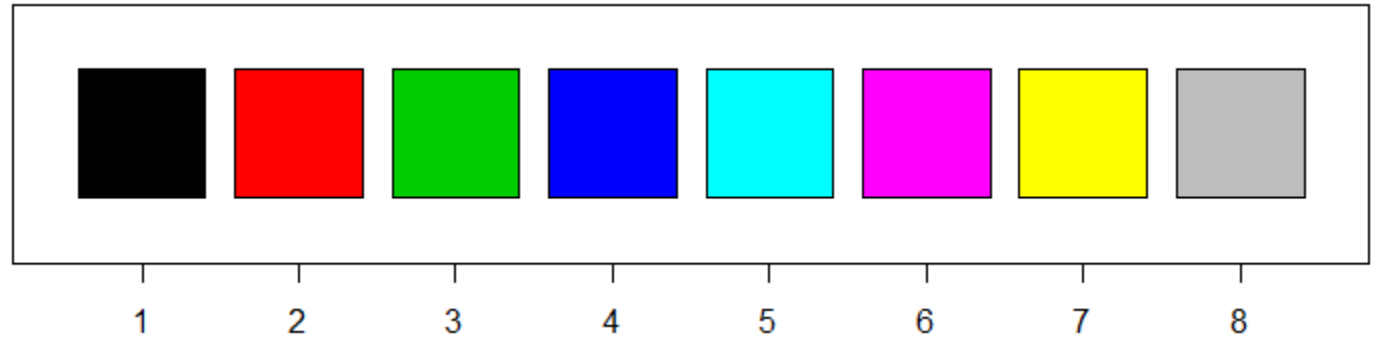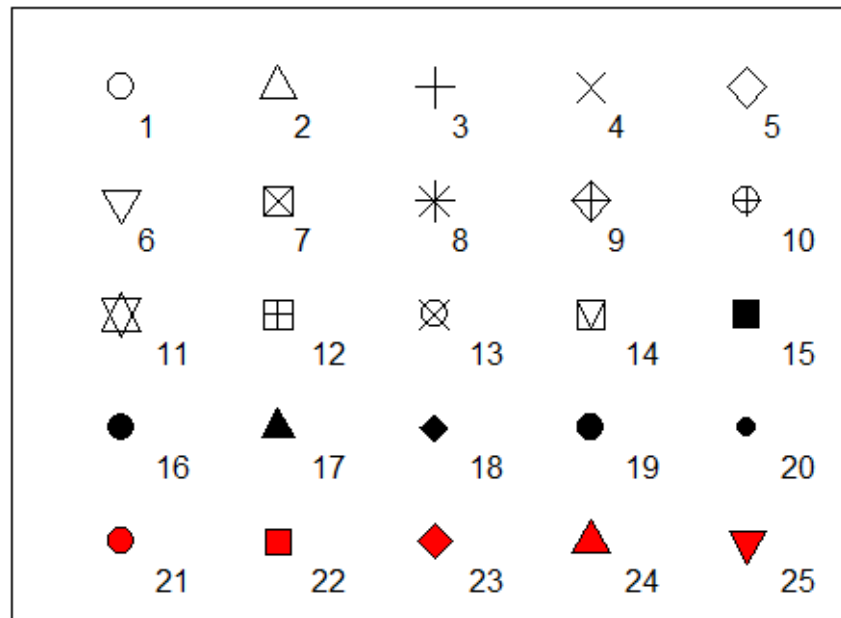*Iris setosa*   *Iris versicolor*   *Iris virginica*

- Measured sepal length, sepal width, petal length and petal width in three species of *Iris* (RA Fisher and ER Anderson).

- Quantify the morphological variation in the three related species

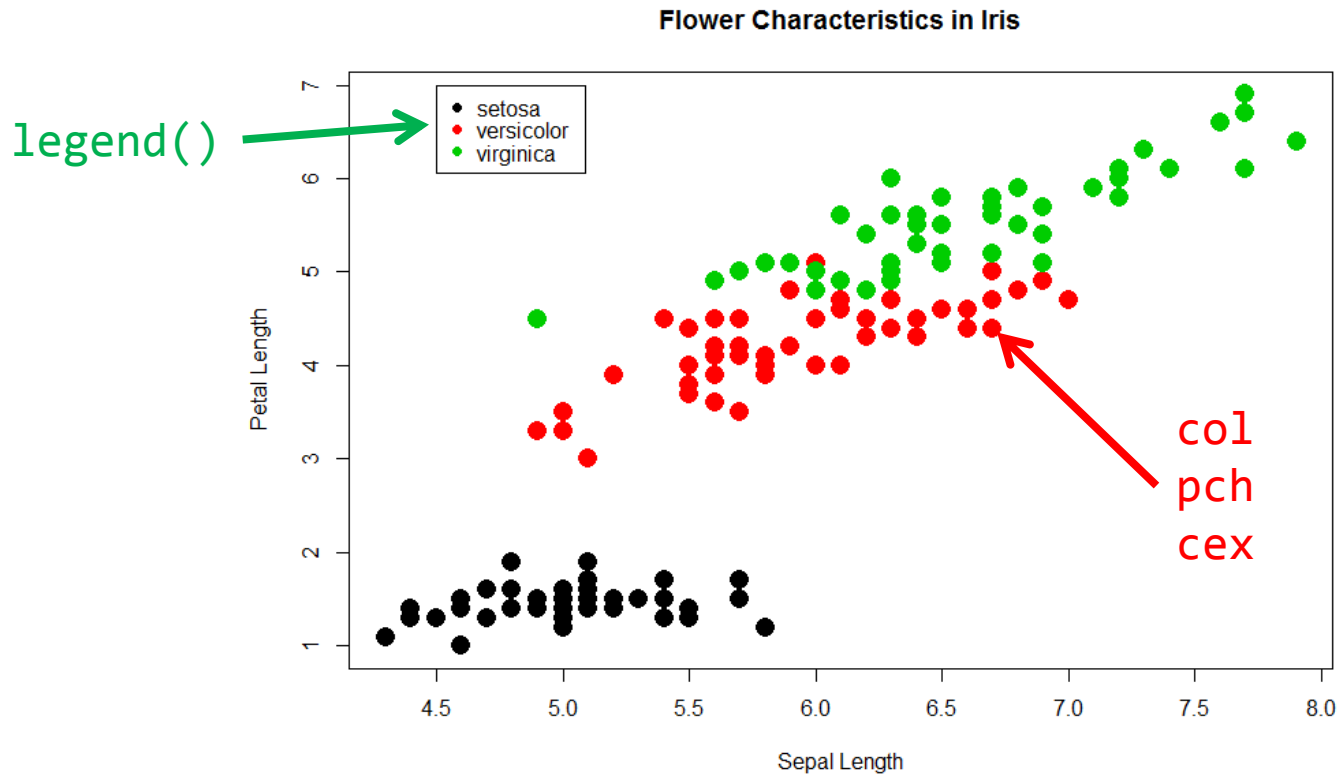- Dataset is included in R

col

pch

3. Scatterplot with legend

Flower Characteristics in Iris

legend()

col
pch
cex

```
261  # FINAL PLOT
262
263  plot(iris$Sepal.Length, iris$Petal.Length,        # x variable, y variable
264       col = iris$Species,                          # colour by species
265       pch = 16,                                     # type of point to use
266       cex = 2,                                      # size of point to use
267       xlab = "Sepal Length",                       # x axis label
268       ylab = "Petal Length",                       # y axis label
269       main = "Flower Characteristics in Iris")     # plot title
270
271  legend (x = 4.5, y = 7, legend = levels(iris$Species), col = c(1:3), pch = 16)
272  # legend with titles of iris$Species and colours 1 to 3, point type pch at coords (x,y)
273
```
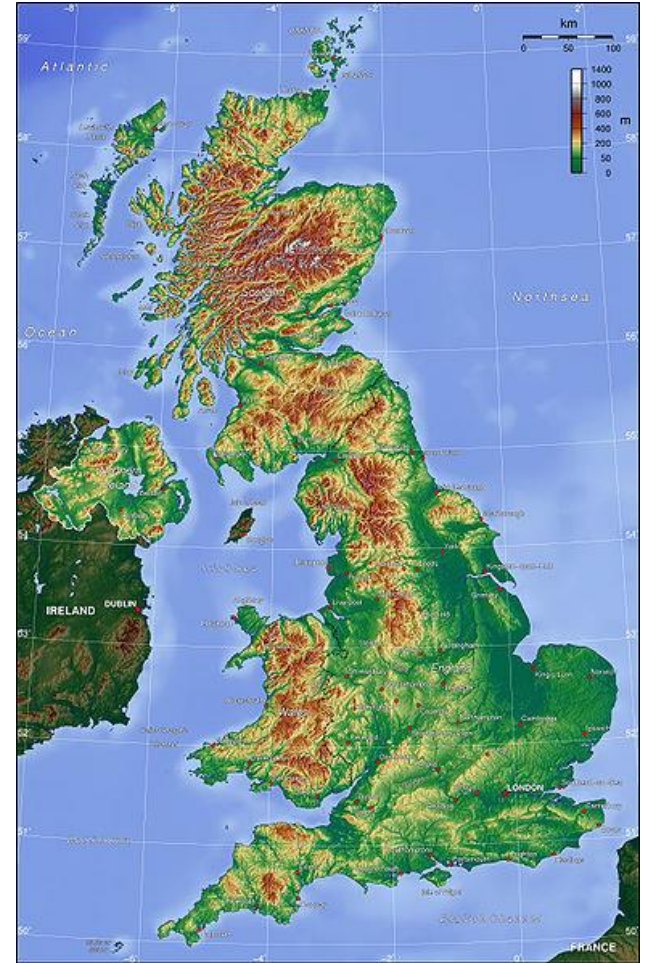
4. *boxplot* function

- Continue with the same dataset…

4. *boxplot* function

Sepal Length by Species in Iris

las, cex.main, cex.lab, notch, cex.axis labels point to a boxplot showing Sepal Length by Species (virginica, versicolor, setosa).

```
354  #~~ FINAL PLOT
355
356  iris$Species<-factor(iris$Species, levels = c("virginica","versicolor","setosa"))
357
358  boxplot(iris$Sepal.Length ~ iris$Species,          # x variable, y variable
359          notch = T,                                  # Draw notch
360          las = 1,                                    # Orientate the axis tick labels
361          xlab = "Species",                           # X-axis label
362          ylab = "Sepal Length",                      # Y-axis label
363          main = "Sepal Length by Species in Iris",   # Plot title
364          cex.lab = 1.5,                              # Size of axis labels
365          cex.axis = 1.5,                             # Size of the tick mark labels
366          cex.main = 2)                               # Size of the plot title
367
```
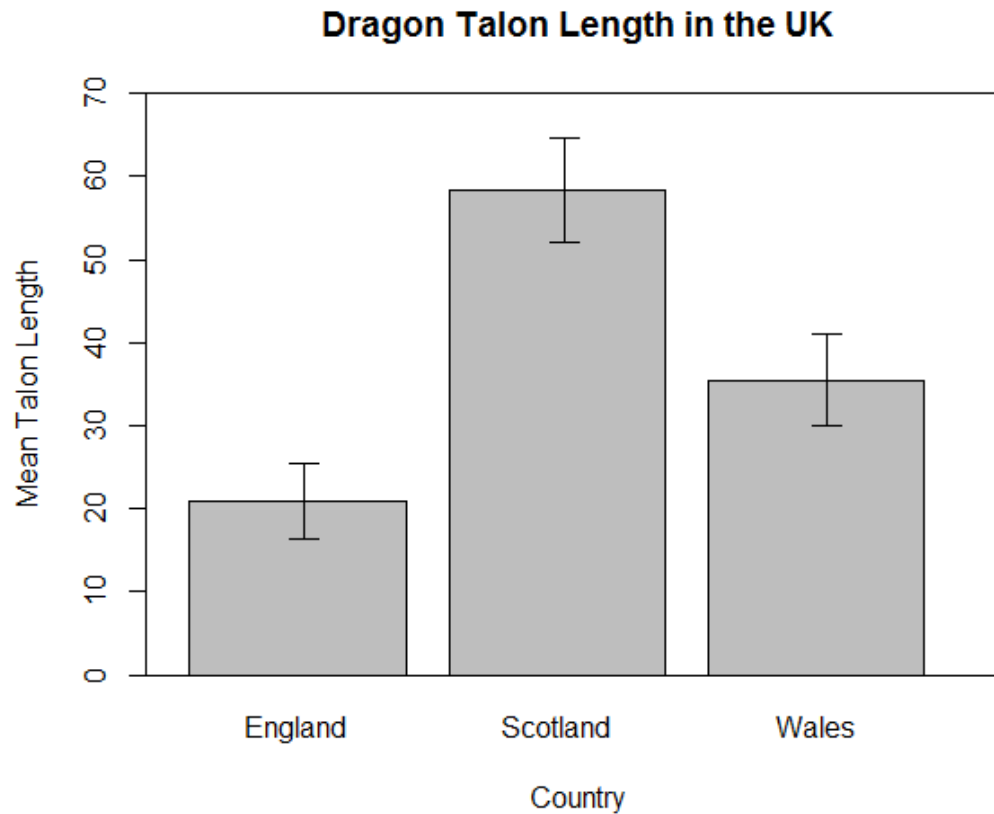
5. *barplot* function

- Dragons are commonly studied in the United Kingdom

- Extensive data on talon length in three countries.

5. *barplot* function

Dragon Talon Length in the UK


```
440  #~~ FINAL PLOT
441
442  # Do it in ggplot2
443
```

What are the limitations of base graphics?

- Just the tip of the iceberg…
  - You could feasibly do anything you require in base graphics, but…

- Some common actions are not straightforward
  - Legends
  - Dodged plots
  - Faceting (lattice)
  - Error Bars (gplots)
  - Formatting axes and plot area

- Complex graphs are time-consuming.

- My advice
  - Base graphics best for quick and dirty exploratory graphics
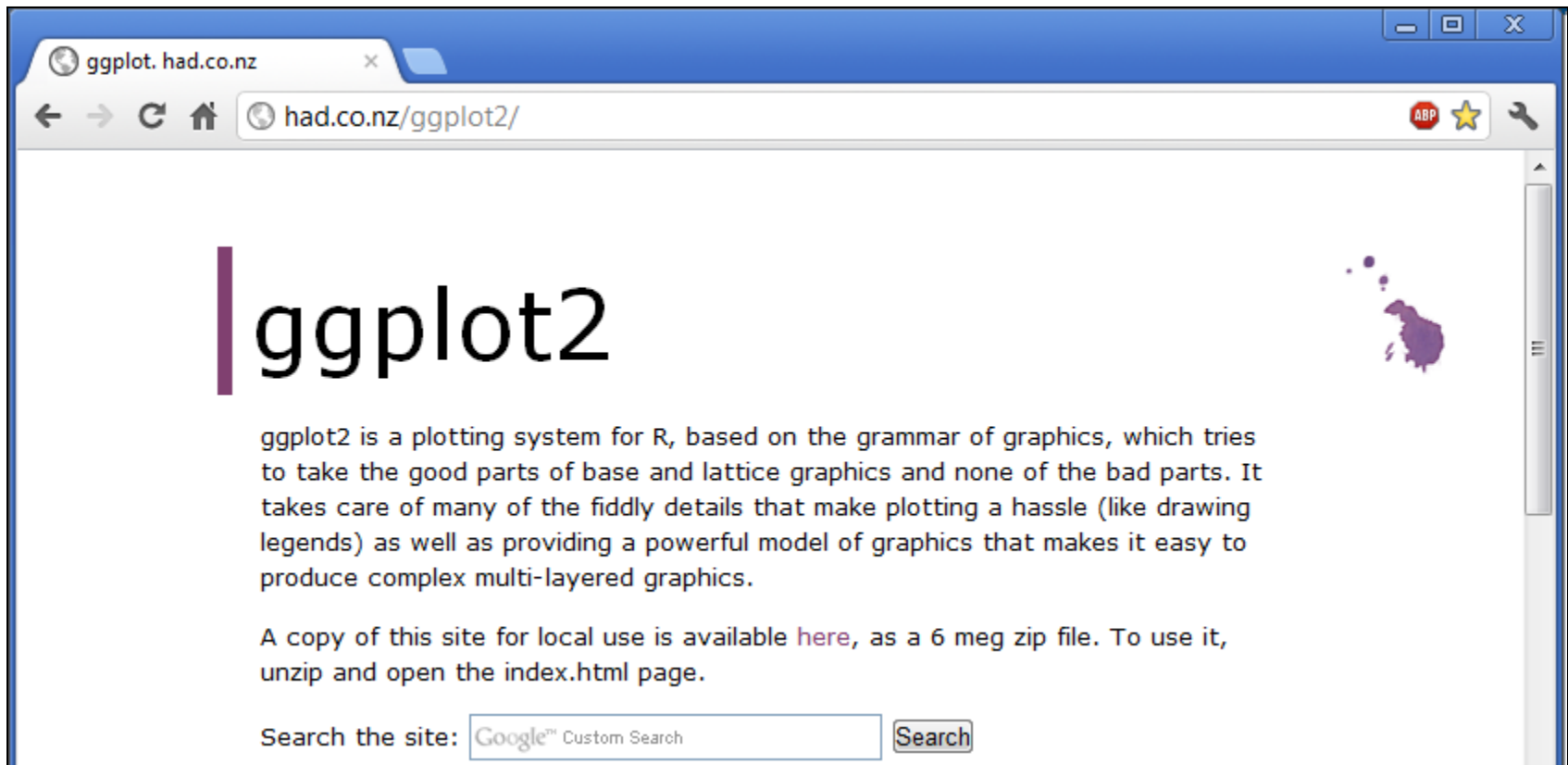 - ggplot2 is best for everything else

Overview

- **Part 1: Base graphics**

  – Simple, fast, exploratory graphics

  – Important to know

- **Part 2: `ggplot2`**

  – More complex, higher quality graphics

  – Fashionable to know

  – Easier to master
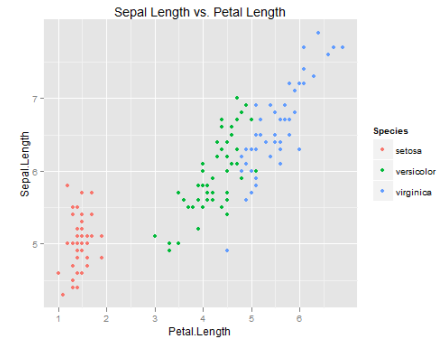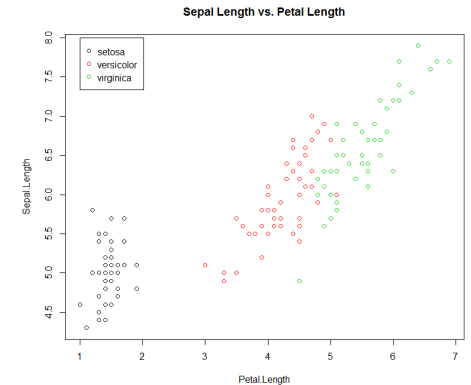
Library ggplot2

- Created by Hadley Wickham: had.co.nz/ggplot2/

ggplot2 uses three components to construct a graph.

1. **Layers**: data with aesthetic properties
   - Data: iris
   - Aesthetic properties of the data
     `x = Sepal.Length, y = Petal.Length, colour = Species`

2. **Geoms**: the type of plot you make.
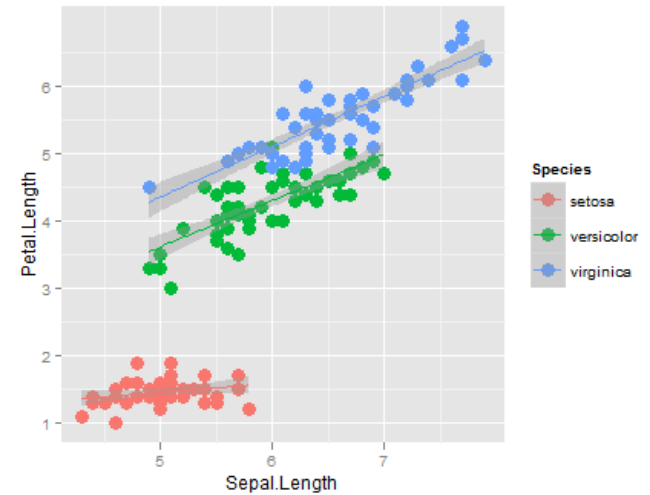   - a line graph, a scatterplot, a boxplot

3. **Stats:** statistical transformations
   - e.g. assigning data to bins, smoothing lines, etc.
   - Every geom has a default statistic, so this is not always specified.
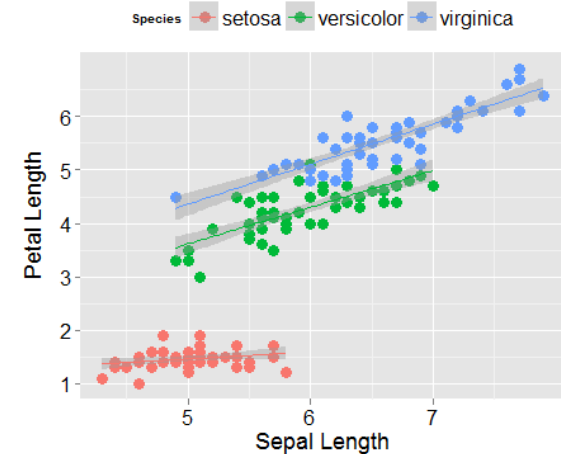
- This should become clearer as we go along.
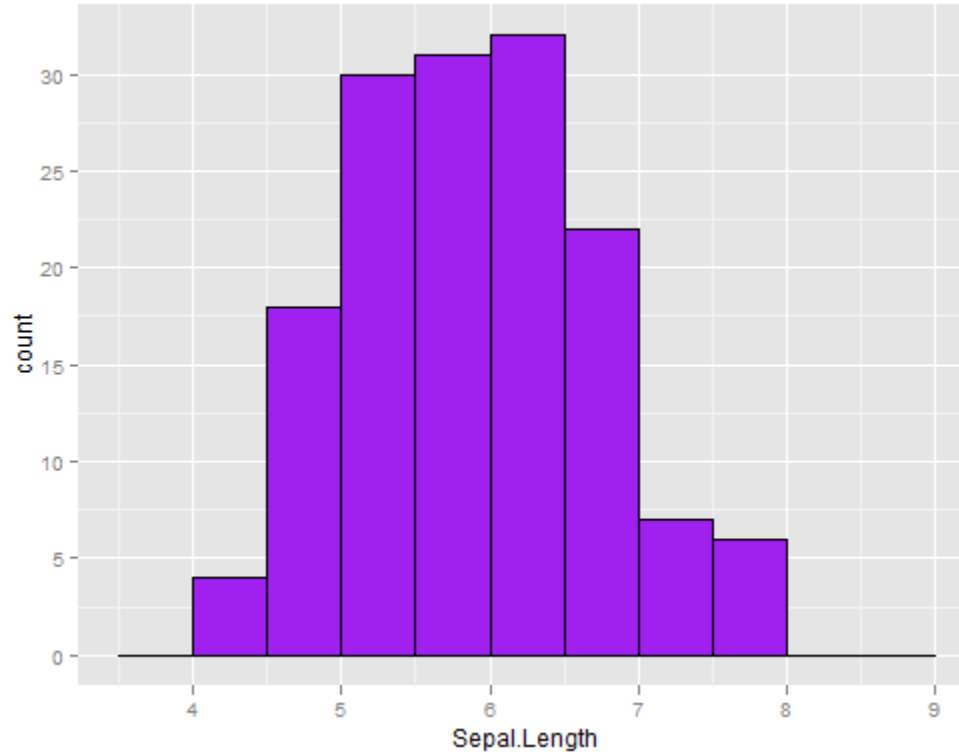
Overview

1. Learn how to build plots

- Data + aesthetic values

- Geoms

- Stats

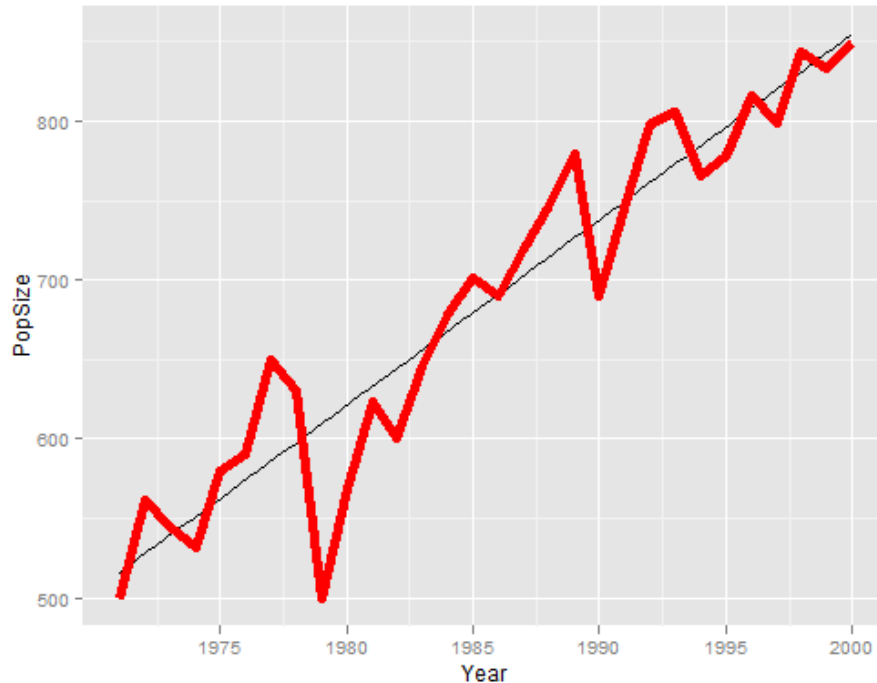- Faceting and Dodging

2. Customise the layout

1. Histogram


```
ggplot(iris, aes(x = Sepal.Length)) +        ⟵ DATA

  geom_histogram(binwidth = 0.5,  col = "black", fill = "purple")   ⟵ GEOM
```
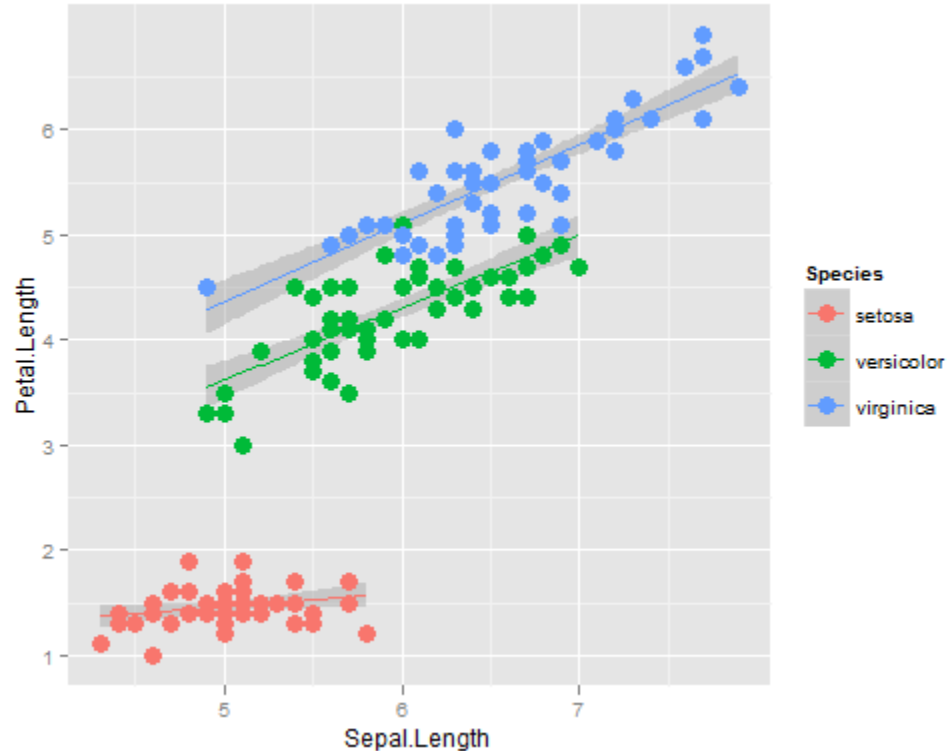
2. Line Graph


```
ggplot(moomins, aes(x = Year, y = PopSize)) +      ← DATA
    stat_smooth(method = "lm", col = "black", se = F) +   ← STAT
    geom_line(col = "red", size = 2)                ← GEOM
```

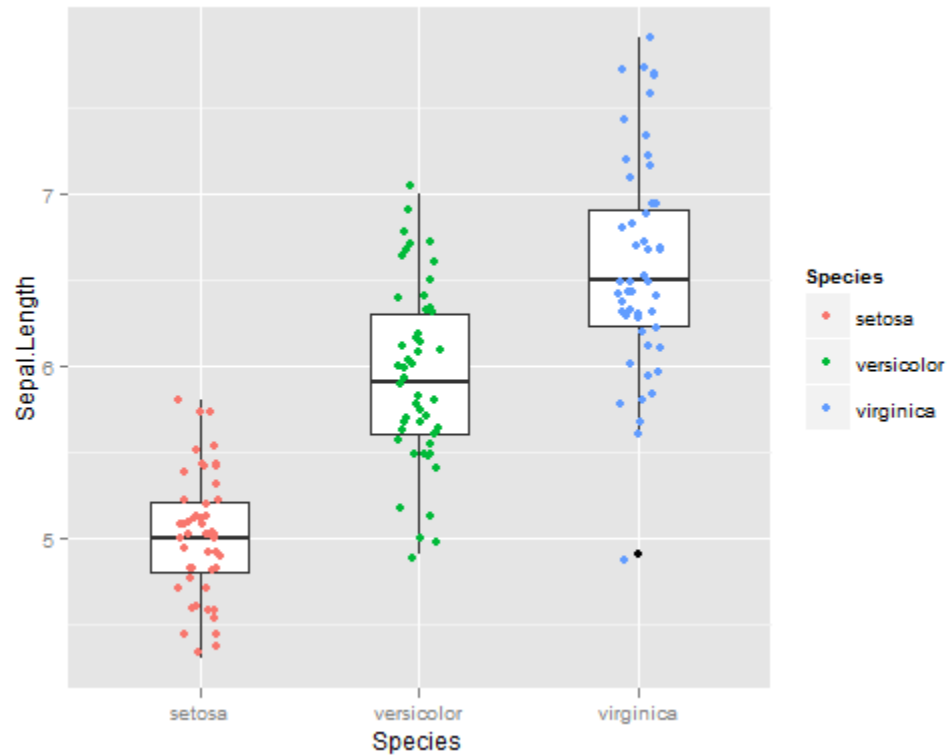# 3. Scatterplot


```
ggplot(iris, aes(x=Sepal.Length, y=Petal.Length, col=Species)) +    ← DATA

   stat_smooth(method = "lm") +                                      ← STAT

   geom_point(size = 4)                                              ← GEOM
```

4. Boxplot

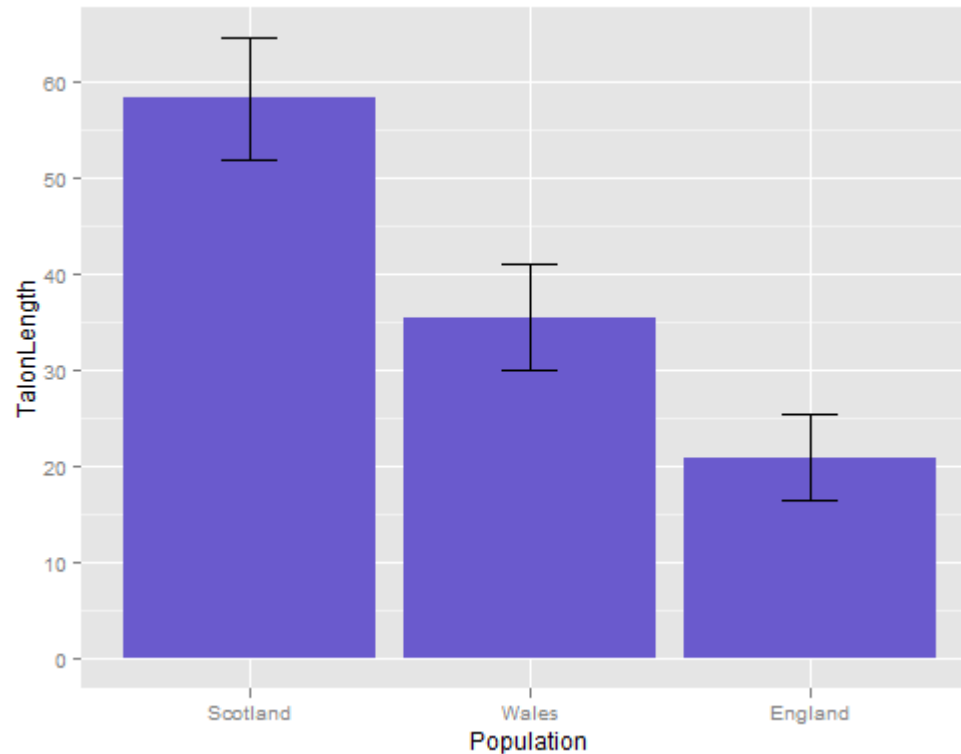```
ggplot(iris, aes(x = Species, y = Sepal.Length)) +   DATA

  geom_boxplot(width=0.6) +   GEOM

  geom_jitter(position=position_jitter(width=0.1), aes(col = Species)) GEOM
```

# 5. Barplot
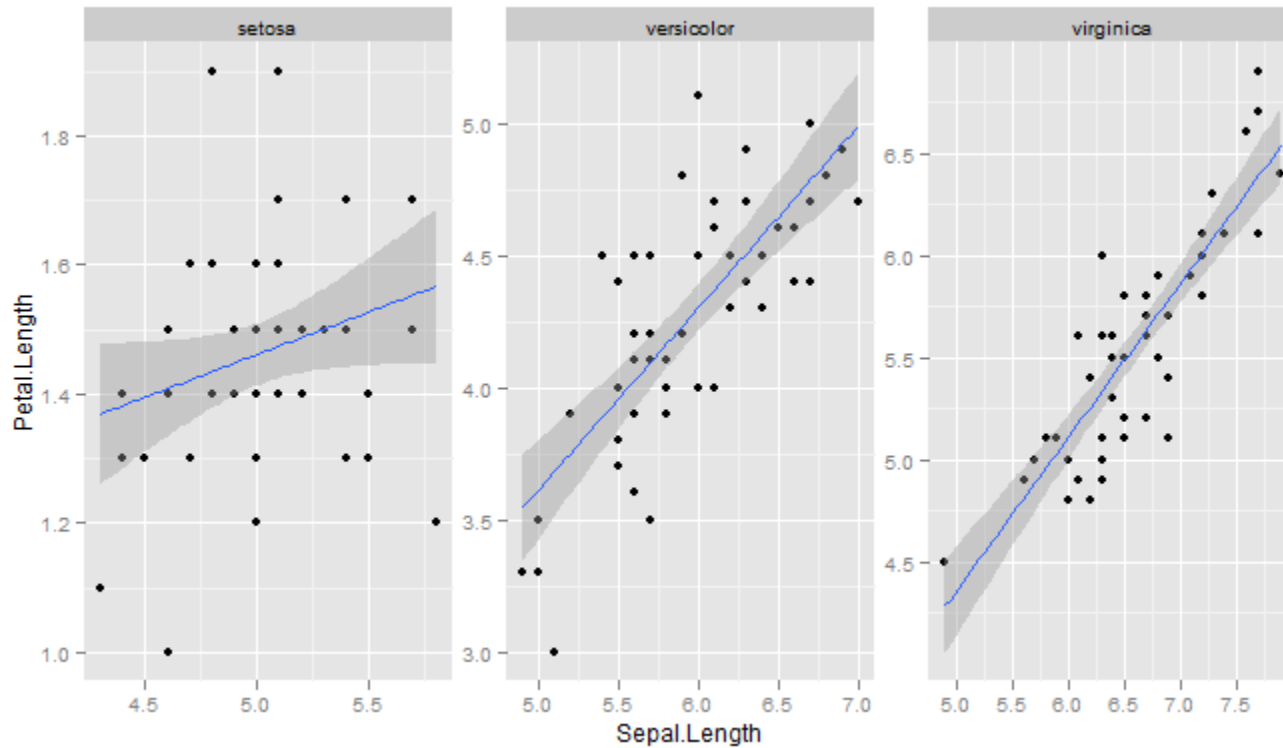


```
ggplot(dragons, aes(x = Population, y = TalonLength)) +   DATA
    geom_bar(fill = "slateblue") +   GEOM
    geom_errorbar(aes(ymax = TalonLength + SE,
                      ymin = TalonLength - SE), width = 0.2) GEOM
```
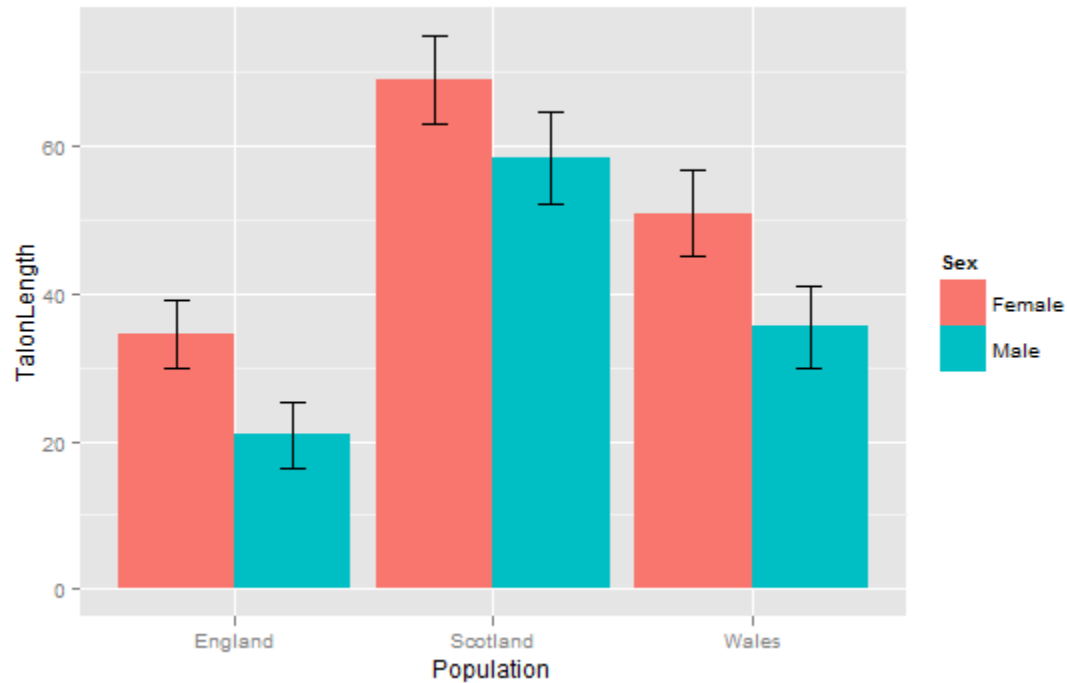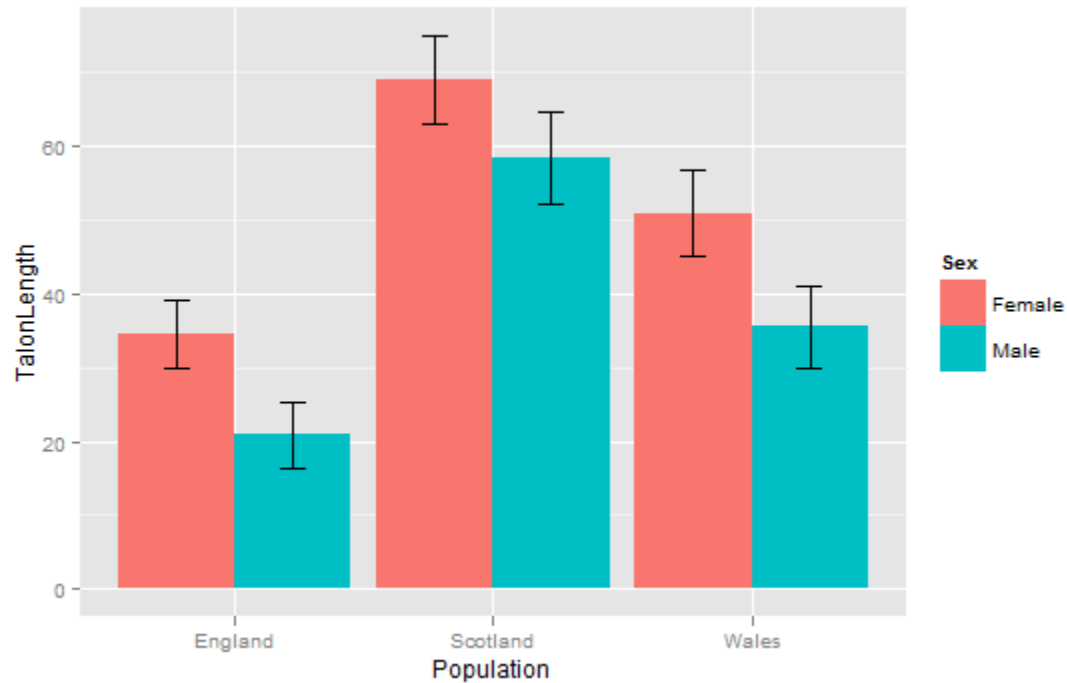
# 6. Faceting



```
ggplot(iris, aes(x = Sepal.Length, y = Petal.Length)) +
  geom_point() +
  stat_smooth(method="lm") +
  facet_wrap(~ Species, scales = "free")
```

7. Dodging


```
ggplot(dragons2, aes(x = Population, y = TalonLength, fill = Sex)) +
  geom_bar(position = position_dodge(width = 0.9)) +
  geom_errorbar(position = position_dodge(width = 0.9),
      aes(ymax = TalonLength + SE, ymin = TalonLength - SE),width = 0.2)
```

Customisation


```
ggplot(dragons2, aes(x = Population, y = TalonLength, fill = Sex)) +
  geom_bar(position = position_dodge(width = 0.9)) +
  geom_errorbar(position = position_dodge(width = 0.9),
        aes(ymax = TalonLength + SE, ymin = TalonLength - SE),width = 0.2)
```

Where can I get help for R Graphics?

- R-Bloggers: r-bloggers.com/

- Stack Overflow: stackoverflow.com/

- R mailing lists archive: tolstoy.newcastle.edu.au/R/

- ggplot2 documentation: had.co.nz/ggplot2/

- R colour chart: research.stowers-institute.org/efg/R/Color/Chart/ColorChart.pdf