An introduction to 

Data visualisation and graphics

Noémie Becker & Benedikt Holtmann

Winter Semester 16/17

Course outline

Review – Rearranging and manipulating data

Graphics with base R

- Histograms
- Scatterplots
- Boxplots

Saving plots

Graphics with ggplot2

Review – Rearranging and manipulating data

Reshaping data

	Package	
	tidyr	reshape2
	gather()	melt()
	spread()	dcast()

Combining data sets

Functions to combine data sets in dplyr	
left_join(a, b, by = "x1")	Joins matching rows from b to a
right_join(a, b, by = "x1")	Joins matching rows from a to b
inner_join(a, b, by = "x1")	Returns all rows from a where there are matching values in b
full_join(a, b, by = "x1")	Joins data and returns all rows and columns

Fish survey
Site
Month
Transect
Species

+

Water characteristics
Site
Month
Water temp.
O_2 - content

+

GPS
Site
Transect
Latitude
Longitude

Adding new variables

Three ways adding a new variable (log of FID)

Using **$**

```
Bird_Behaviour$log_FID <- log(Bird_Behaviour$FID)
```

Using **[ ]** - operator

```
Bird_Behaviour[ , "log_FID"] <- log(Bird_Behaviour$FID)
```

Using **mutate()** from **dplyr** package

```
Bird_Behaviour <- mutate(Bird_Behaviour, log_FID = log(FID))
```

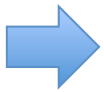
Adding new variables

Split one column into two using **separate()** from **dplyr** package

Combine two columns using **unite()** from **tidyr** package

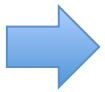
separate()

X1	X2
A	1_1
B	1_2
A	2_1
B	2_2



X1	X2.1	X2.2
A	1	1
B	1	2
A	2	1
B	2	2

unite()



X1	X2
A	1_1
B	1_2
A	2_1
B	2_2

Subsetting data

Subsetting data

- Using [ ] – operator
- Using **subset()**

Operator	Description
>	greater than
>=	greater than or equal to
<	less than
<=	less than or equal to
==	equal to
!=	not equal to
x & y	x and y
x | y	x or y

subset(Bird_Behaviour, FID < 10)
selects all rows with FID smaller than 10m

subset(Bird_Behaviour, FID < 10 & Sex == "male")
selects all rows for males with FID smaller than 10m

subset(Bird_Behaviour, FID > 10 | FID < 15, select = c(Ind, Sex, Year))
selects all rows that have a value of FID greater than 10 or less than 15. We keep only the IND, Sex and Year column

Course outline

Review – Rearranging and manipulating data

Graphics with base R

- Histograms
- Scatterplots
- Boxplots

Saving plots

Graphics with ggplot2

Graphics with base R

Simple graphics using plotting functions in the graphics package

- Base R, installed by default
- Easy and quick to type
- Wide variety of functions

Graphics with base R

Simple graphics using plotting functions in the graphics package

- Base R, installed by default
- Easy and quick to type
- Wide variety of functions

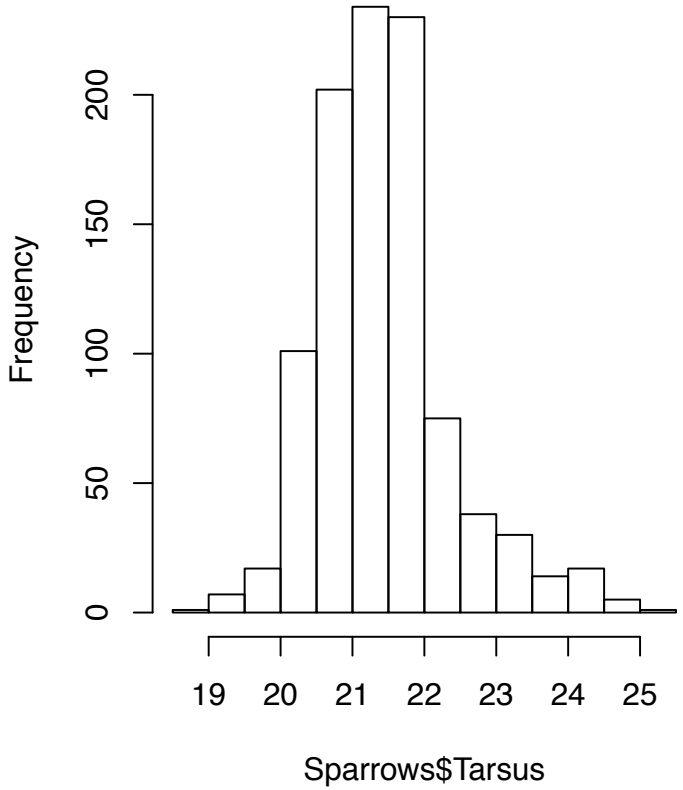
Function	Description
hist()	**Histograms**
plot()	**Scatterplots, etc.**
boxplot()	**Box- and whisker plots**
barplot()	Bar- and column charts
dotchart()	Cleveland dot plots
contour	Contour of a surface (2D)
pie()	Circular pie chart
…	

Graphics with base R

Creating a **histogram** with **hist()**

Example 1:

hist(Sparrows$Tarsus)

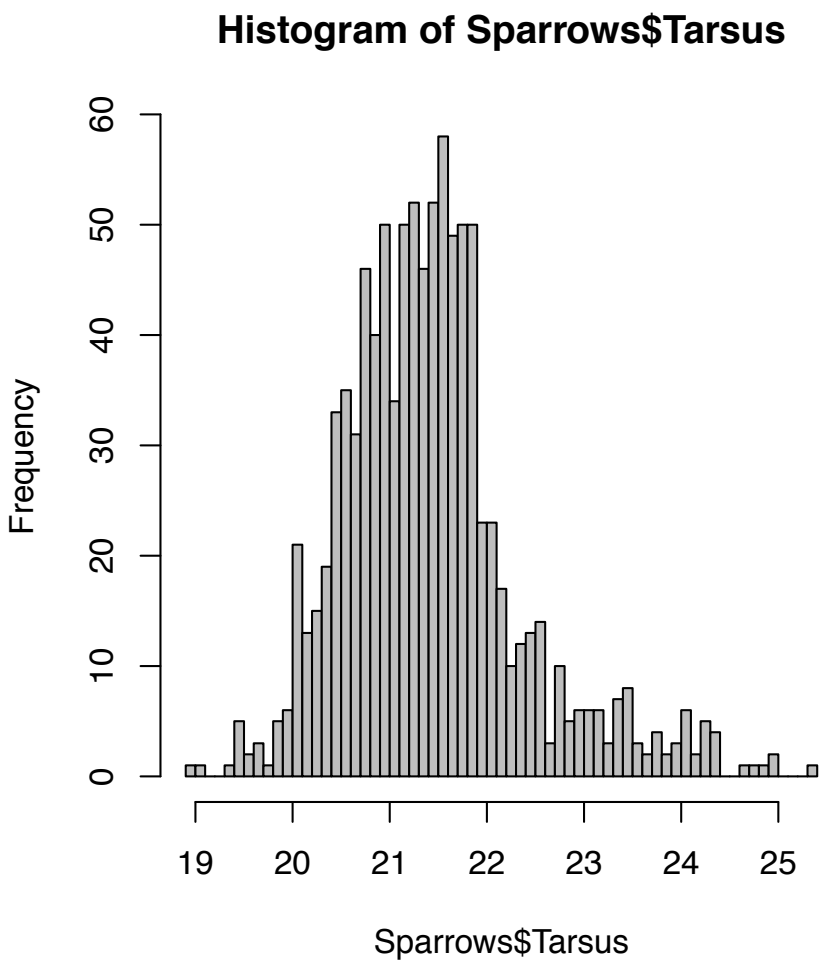


Graphics with base R

Creating a **histogram** with **hist()**

Example 2: Alter **colour** and the **number of bins**

```
hist(Sparrows$Tarsus,
     col = "grey",
     breaks = 50)
```

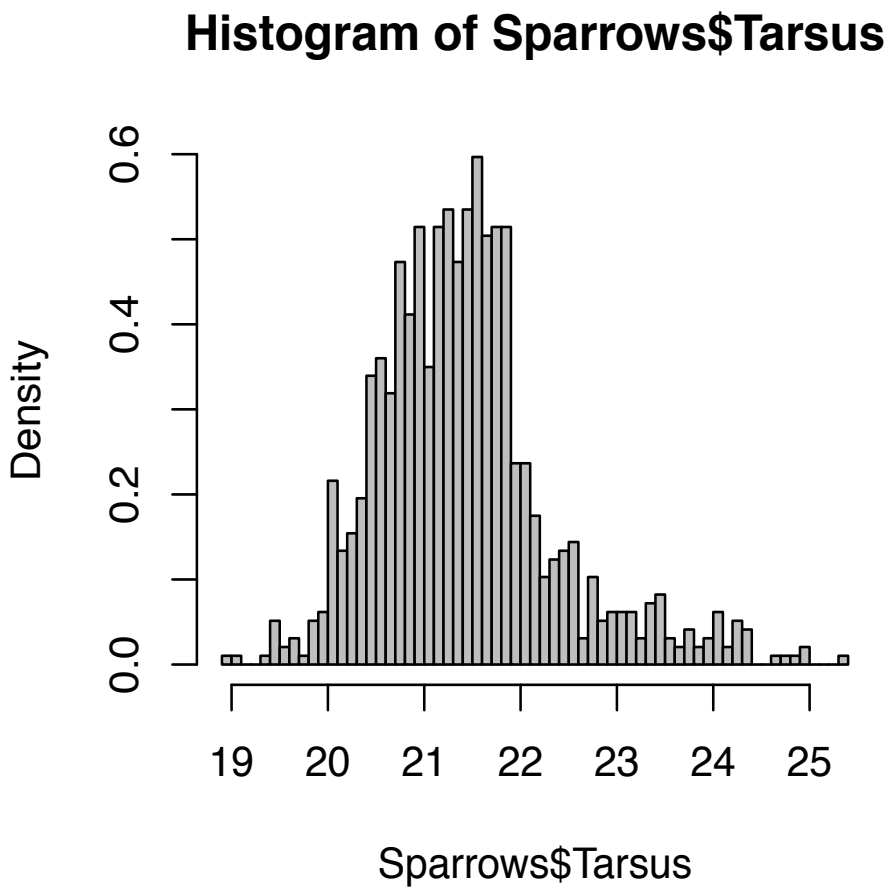


Graphics with base R

Creating a **histogram** with **hist()**

Example 3: Add **density curve**

```
hist(Sparrows$Tarsus,
     col="grey",
     breaks = 50,
     freq = FALSE)
```



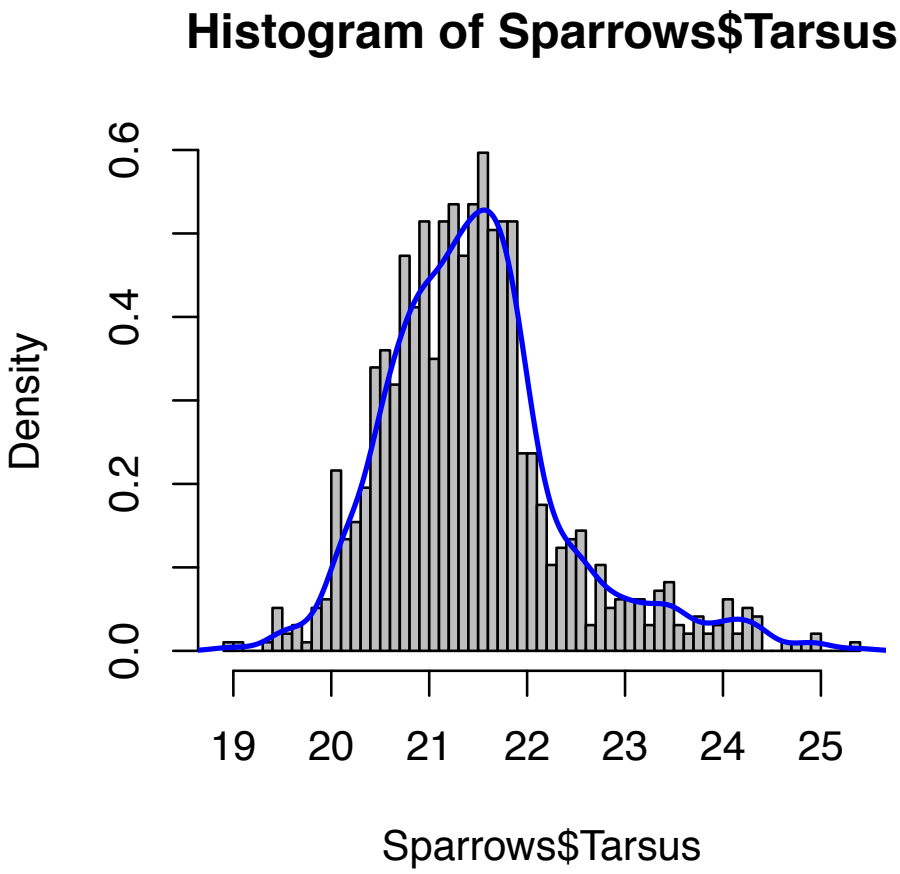
Graphics with base R

Creating a **histogram** with **hist()**

Example 3: Add **density curve**

```
hist(Sparrows$Tarsus,
     col="grey",
     breaks = 50,
     freq = FALSE)

lines(density(Sparrows$Tarsus),
      col = "blue",
      lwd = 2)
```

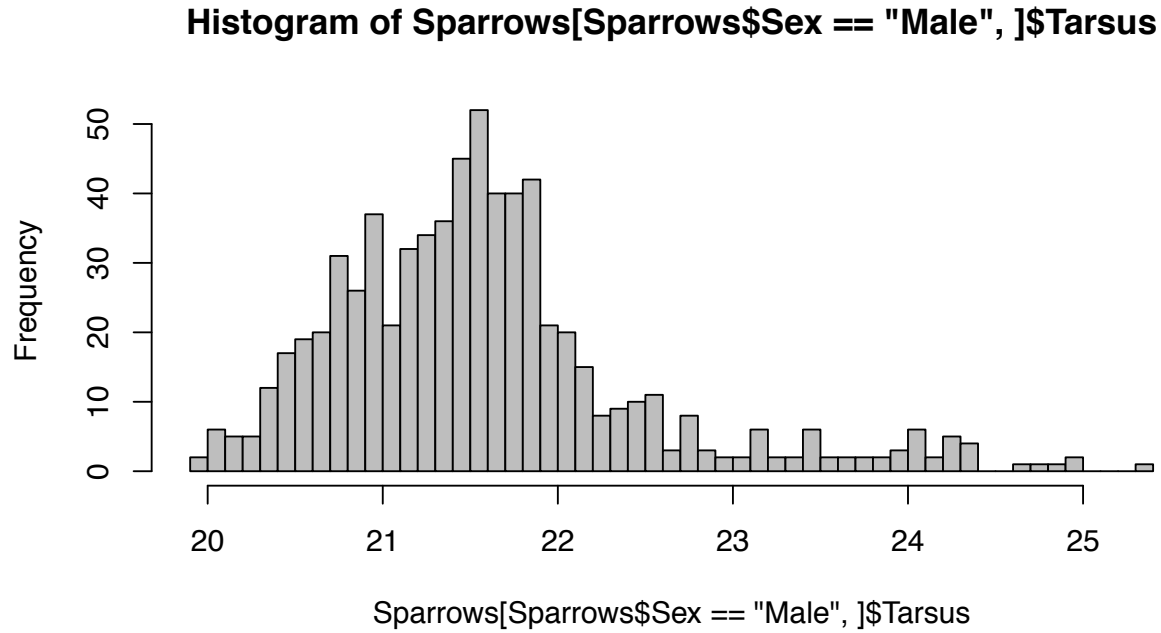


Graphics with base R

Creating a **histogram** with **hist()**

Example 4: Plot only males

hist(Sparrows[Sparrows$Sex == "Male",]$Tarsus, **col** = "grey", **breaks** = 50)



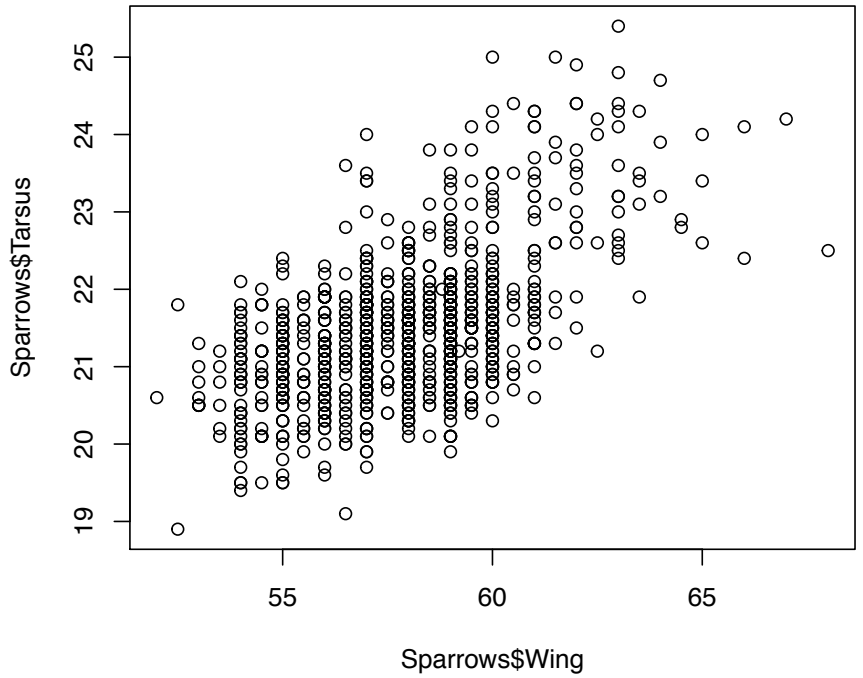
Graphics with base R

Creating a **scatterplot** with **plot()**

Example 1:

plot(Sparrows$Wing, Sparrows$Tarsus)

- **Relationship between two continuous variables**

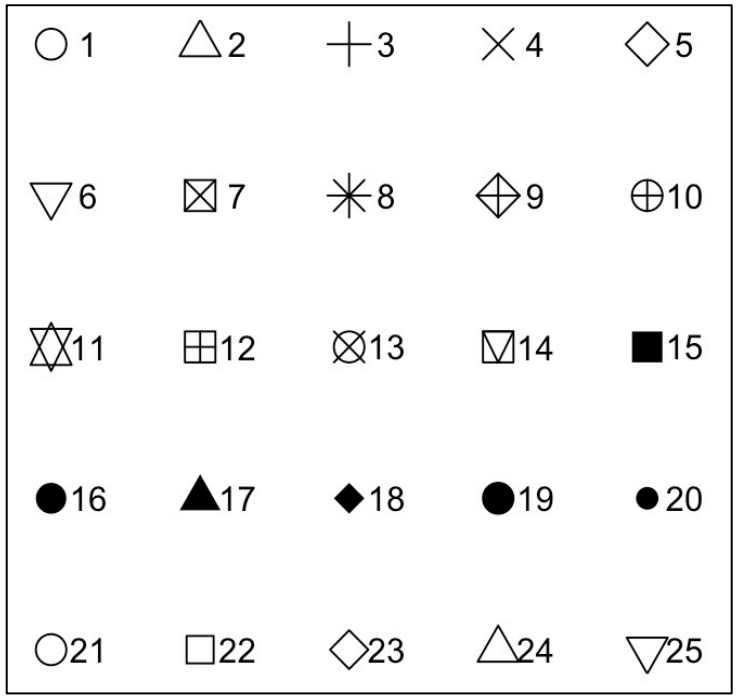


Graphics with base R

Creating a **scatterplot** with **plot()**

Example 2: Alter **axis limits** and **shape** of symbols

```
plot(Sparrows$Tarsus, Sparrows$Wing,
     xlim = c(50, 70),
     pch = 15,
     col = "blue")
```

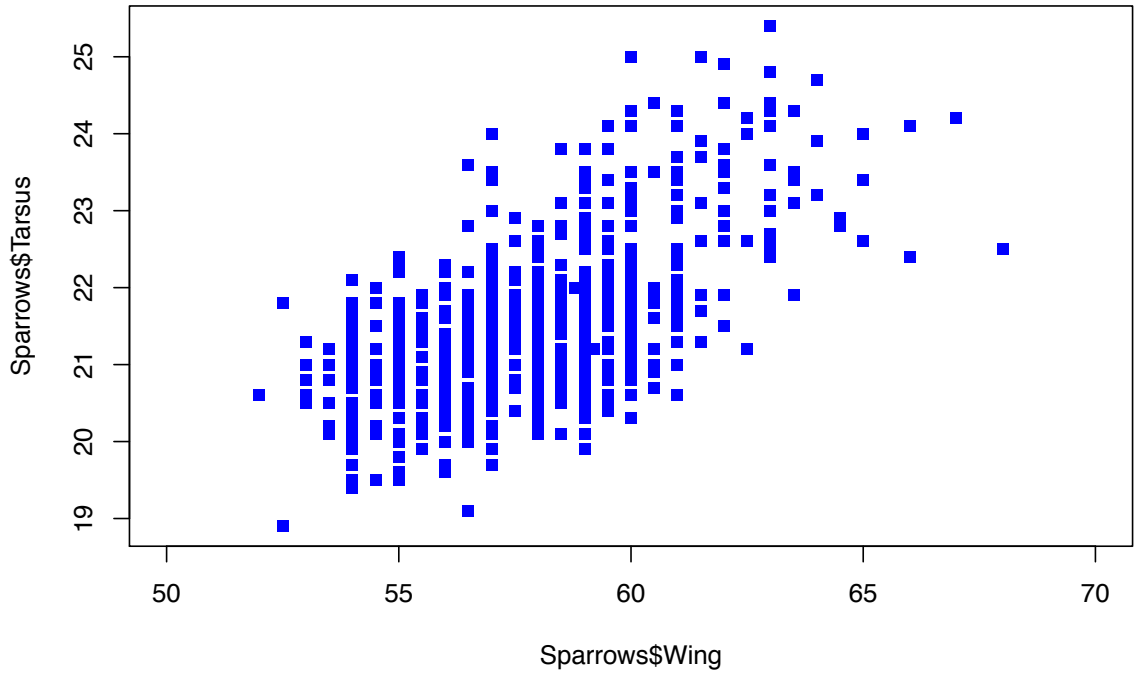


Graphics with base R

Creating a **scatterplot** with **plot()**

Example 2:

```
plot(Sparrows$Tarsus, Sparrows$Wing, xlim = c(50, 70), pch = 15,
     col = "blue")
```

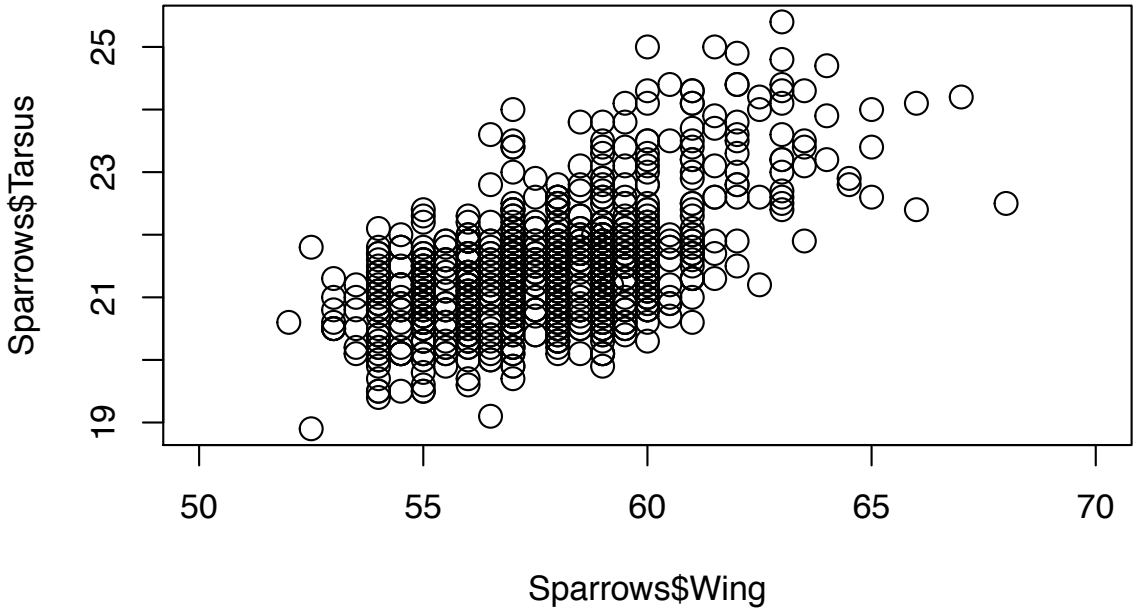


Graphics with base R

Creating a **scatterplot** with **plot()**

Example 3: Alter the **size of plotting symbols**

plot(Sparrows$Wing, Sparrows$Tarsus, xlim = c(50,70), **cex** = 1.5)



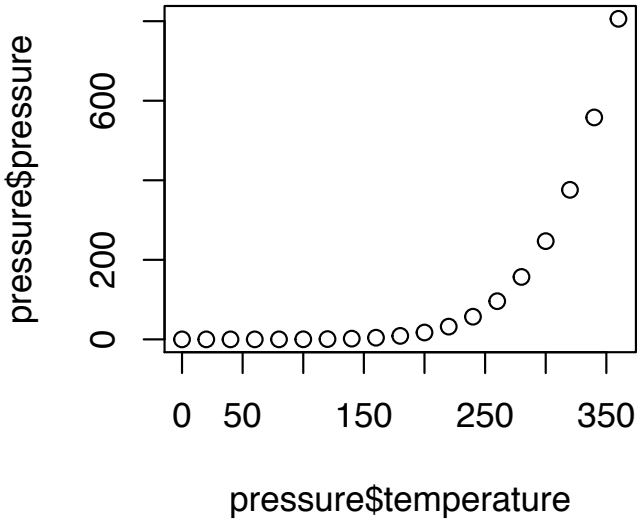
Graphics with base R

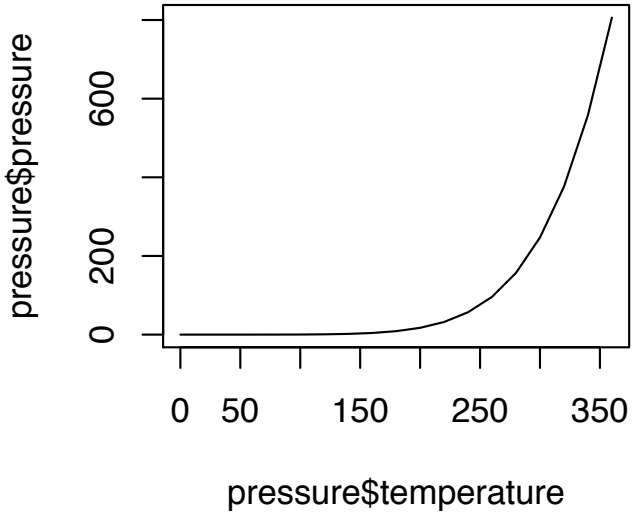
Creating a **line graphs** with **plot()**

Example 1:

plot(pressure$temperature, pressure$pressure)

plot(pressure$temperature, pressure$pressure, **type = "l"**)





Graphics with base R

Use the **type** argument to specify the type of plot

Possible types	
"p"	points
"l"	lines
"b"	points connected by lines
"o"	points overlaid by lines
"h"	vertical lines from points to the zero axis
"s"	steps
"n"	nothing, only the axes

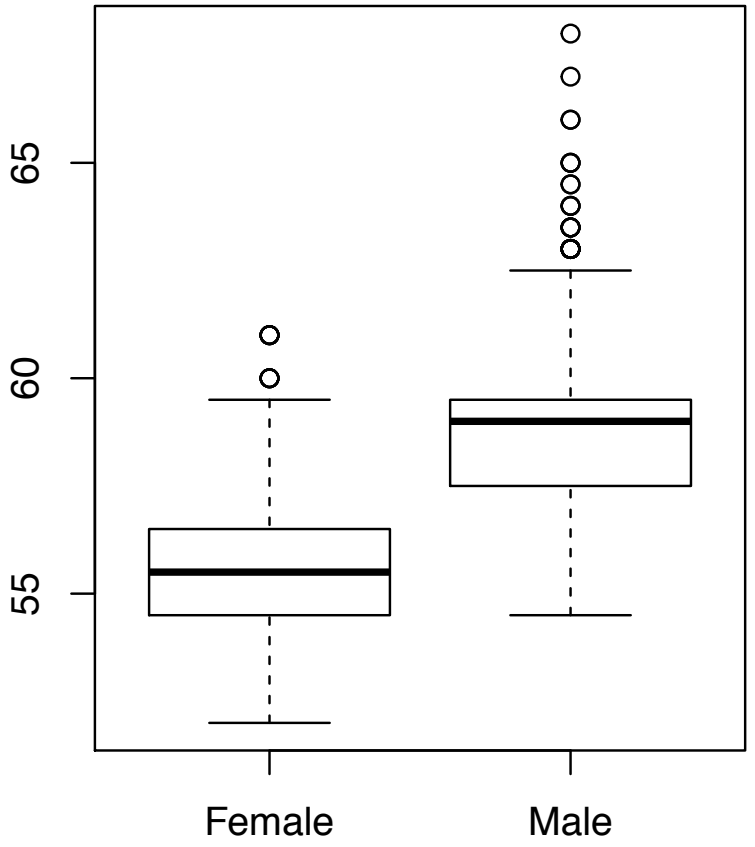
Graphics with base R

Creating a **boxplot** with **boxplot()**

Example 1:

boxplot(Wing ~ Sex, data = Sparrows)

- **Relationship between continuous and categorical variables**



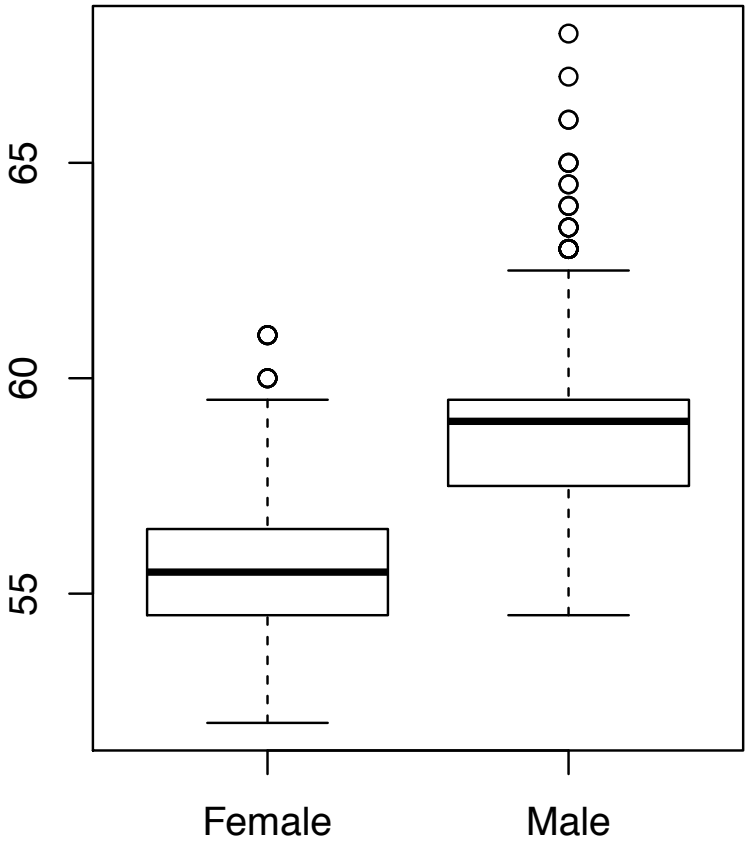
Graphics with base R

Creating a **boxplot** with **boxplot()**

Example 1:

boxplot(Wing ~ Sex, data = Sparrows)

- median
- 25th and 75th percentiles
- maximum and minimum values
- outliers



Graphics with base R

Creating a **boxplot** with **boxplot()**

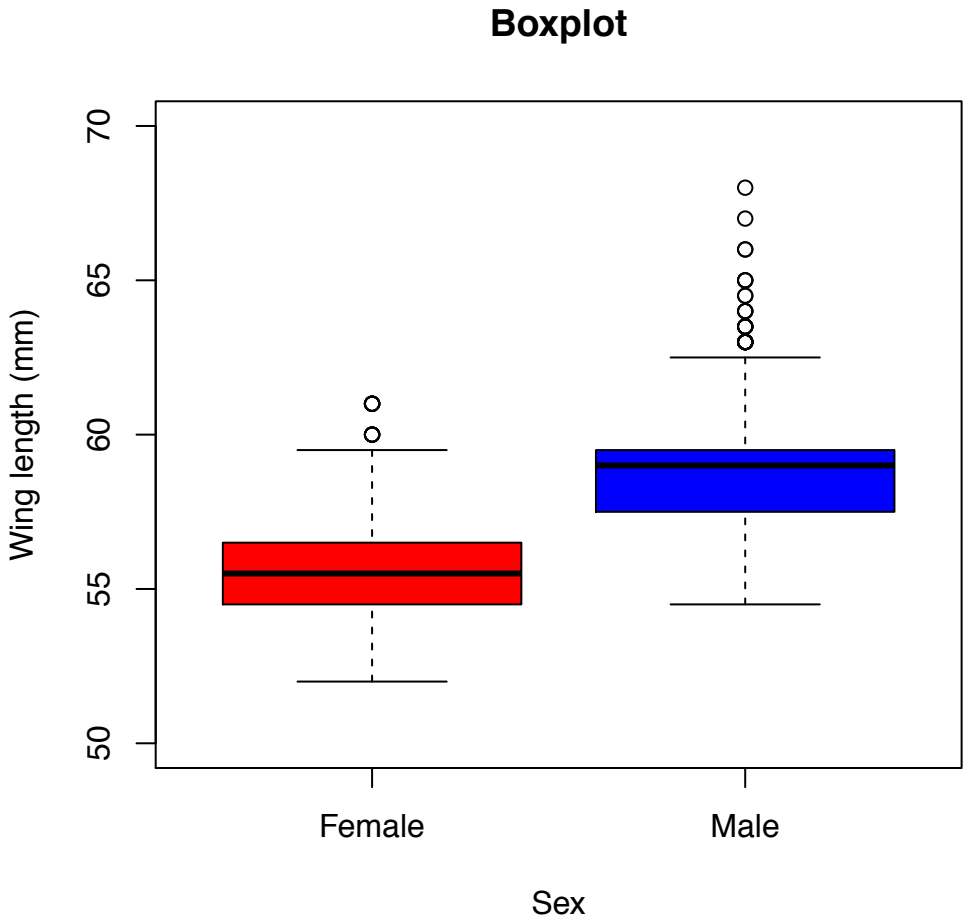
Example 2:

```
boxplot(Wing ~ Sex, data = Sparrows,
      xlab = 'Sex',                          # Adds label to x-axis
      ylab = 'Wing length (mm)',             # Adds label to y-axis
      col=c("red", "blue"),                  # Adds colour
      ylim = c(50,70),                       # Changes axis limits
      main = "Boxplot"))                     # Adds title
```

Graphics with base R

Creating a **boxplot** with **boxplot()**

Example 2:



Graphics with base R

Creating a **boxplot** with **boxplot()**

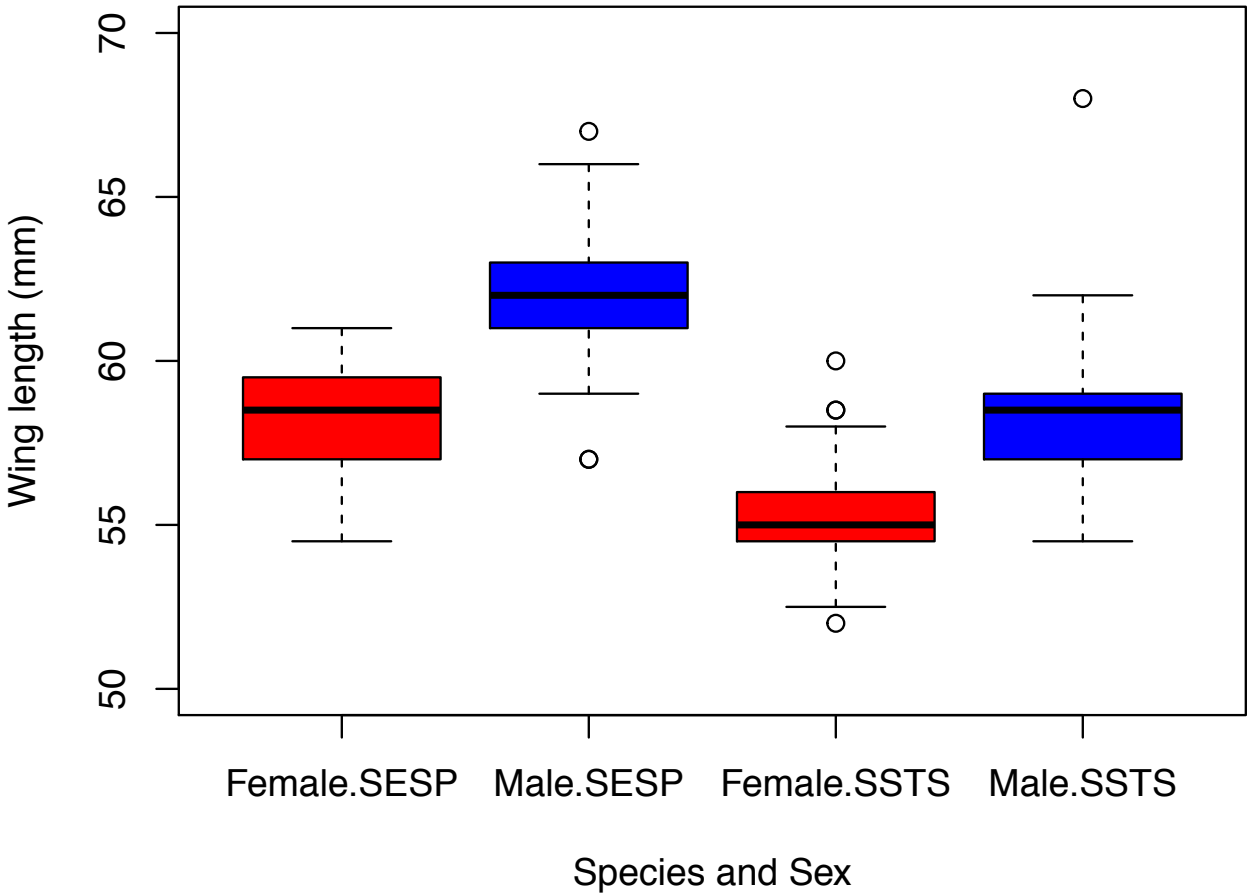
Example 3: Multiple grouping variables

```
boxplot(Wing ~ Sex + Species, data = Sparrows,
      xlab = ’Species and Sex',
      ylab = 'Wing length (mm)',
      col=c("red", "blue"),
      ylim = c(50,70),
      main = "Boxplot”))
```

Graphics with base R

Creating a **boxplot** with **boxplot()**

Example 3:



Quickly exploring data

Common parameters in graphics	
main	title of the plot
xlab	label of x-axis
ylab	label of y-axis
xlim	range/limits of x-axis
ylim	range/limits of y-axis
col	colour of the points, bars, etc. can be character string or hexadecimal colour (e.g. #RRGGBB)
breaks	number of bins
pch	shape of symbol
cex	size of symbols
lty	line type
lwd	line width

Multiple plots on one page

The **par()** function

- comes with an extensive list of graphical parameters you can change (see ?par)

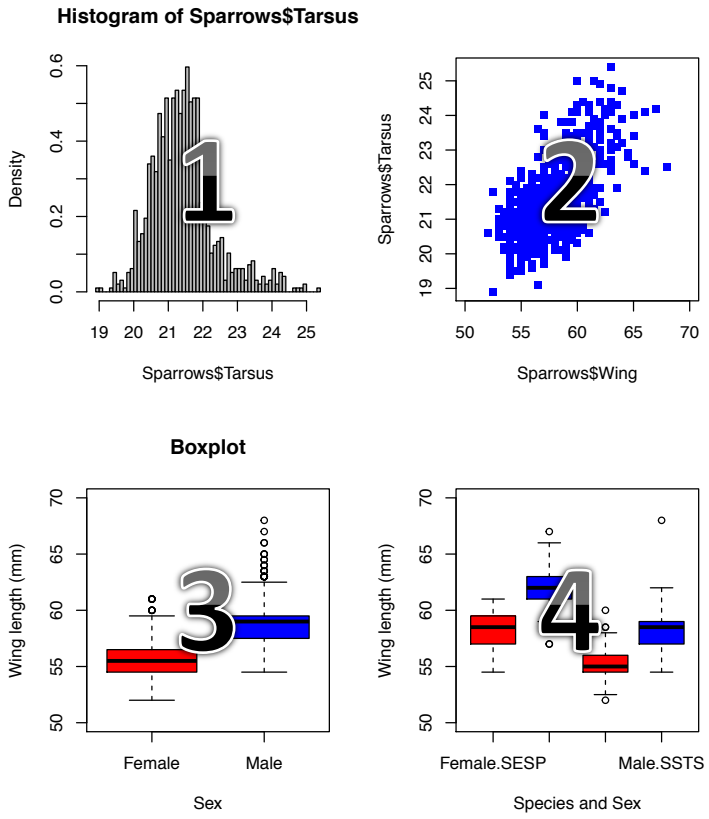
- Some options are helpful; others you may never use

To plot multiple charts within the same window, you can use the **mfcol** or **mfrow** parameter

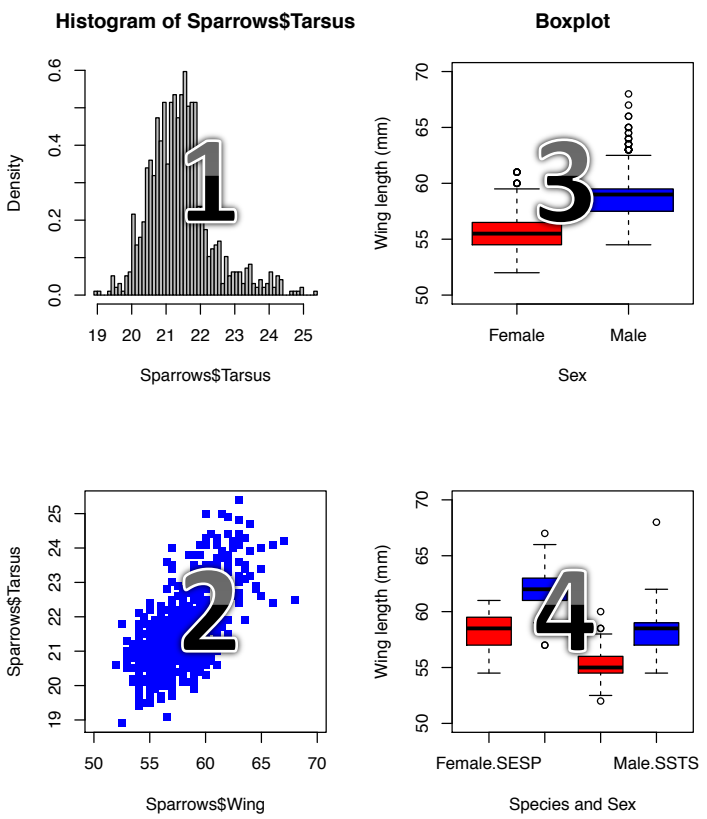
For example, **par(mfrow=c(2, 2)** divides the graphic window into four panels (two rows and two columns)

Multiple plots on one page

par(**mfrow** = c(2,2))



par(**mfcol** = c(2,2))



Saving plots

There are several possibilities saving a plot

1. dev.print()

Example:

```
plot(x, y, ….)          # Make a plot

# After you are finished with the plot use:

dev.print(device = pdf, file = "filename.pdf")
```

Important:

When you are done, you have to close the printing device!

```
dev.off()   # shuts down current device
```

Saving plots

There are several possibilities saving a plot

2. savePlot()

Example:

```
plot(x, y, ….)          # Make a plot

savePlot(filename = "Figure1.pdf", type = "pdf")
```

Important:

It is possible that it does not work for your system!

(uses X11 device, most Unix systems)

Saving plots

There are several possibilities saving a plot

3. Plot directly into a file

Example:

```
pdf("Figure2.pdf", width= 4, height = 4)     # width and height are in inches

hist(x)                  # You can execute multiple graphing commands
plot(x, y, ….)           # The result of each will go into the pdf file
```

dev.off()

But file is not printed on screen!

Saving plots

There are several possibilities saving a plot

3. Plot directly into a file

Functions to save plots	
pdf()	Opens a pdf-file as device
postscript()	Opens a postscript-file as device
png()	Opens a png-file as device
jpeg()	Opens a jpeg-file as device
tiff()	Opens a tiff-file as device
bmp()	Opens a bmp-file as device

Graphics with ggplot2

Why use **ggplot2**?

- Many users, a lot of support
- Check out the ggplot2 documentation at http://docs.ggplot2.org/
- Very flexible and powerful
- Sophisticated plots for publication

Graphics with ggplot2

To create a plot you use the **ggplot()** function

Basic structure:

```
ggplot(data,                          # data frame with variables to plot
    aes(x variable, y variable)) +    # specifies which variables to plot
    geom_object()                     #  specifies the geometric objects
```

Commonly used geometric objects:

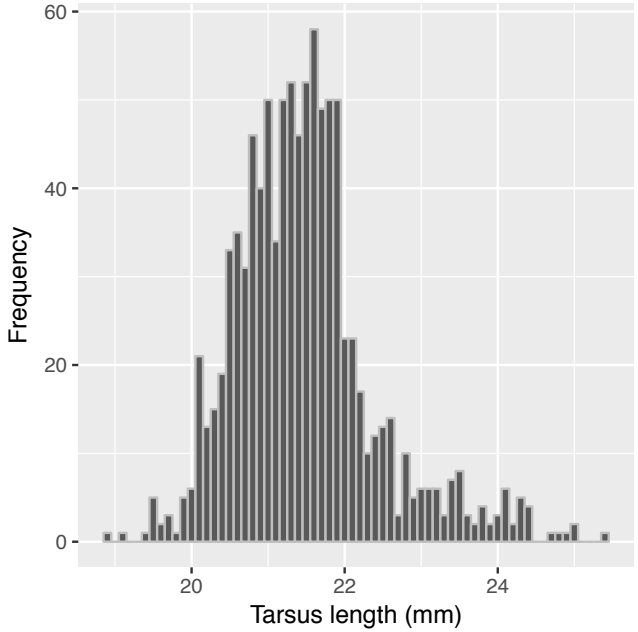
Histogram: + geom_histogram()
Scatterplot: + geom_point()
Boxplot: + geom_boxplot()

Graphics with ggplot2

Creating a **histogram** with **ggplot()**

Example 1:

```
ggplot(Sparrows, aes(Tarsus)) +
        geom_histogram(col = "grey", binwidth = 0.1) +
        xlab("Tarsus length (mm)") +
        ylab("Frequency")
```

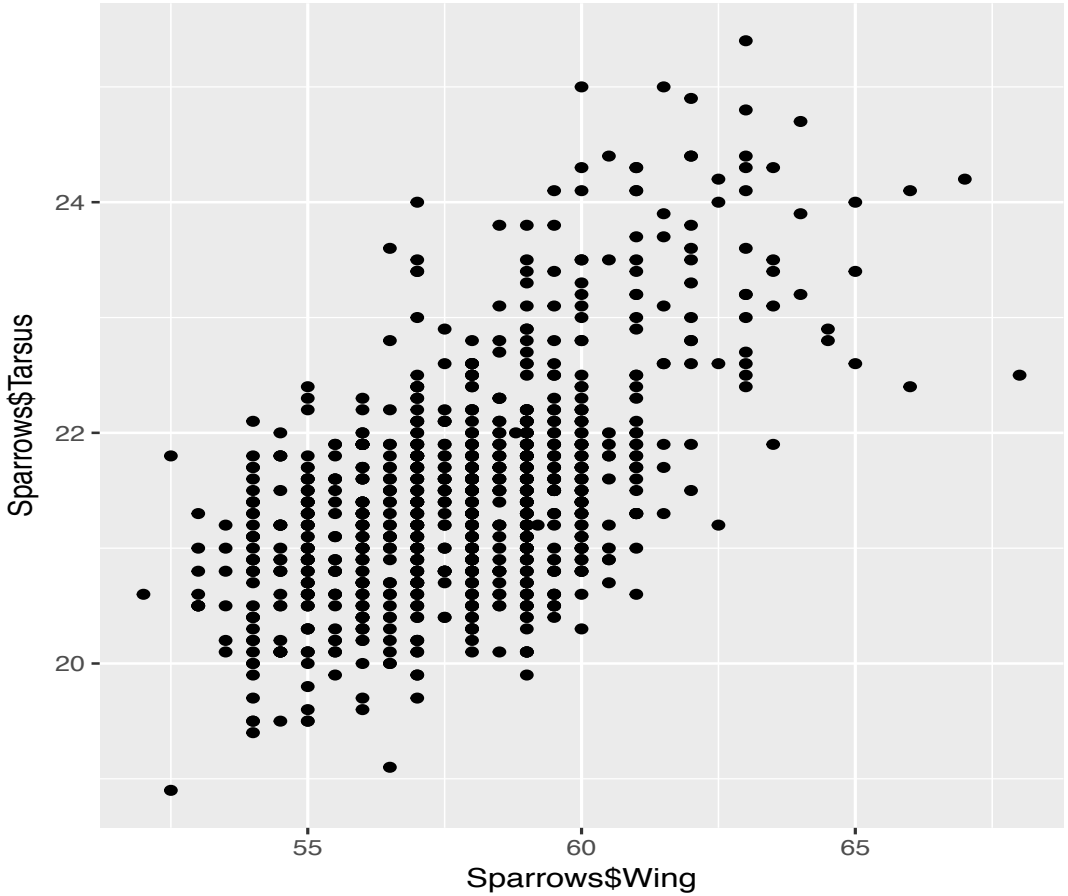


Graphics with ggplot2

Creating a **scatterplot** with **ggplot()**

Example 1:

```
ggplot(Sparrows,
       aes(x = Wing,
           y = Tarsus)) +
       geom_point()
```



Graphics with ggplot2

Creating a **scatterplot** with **ggplot()**

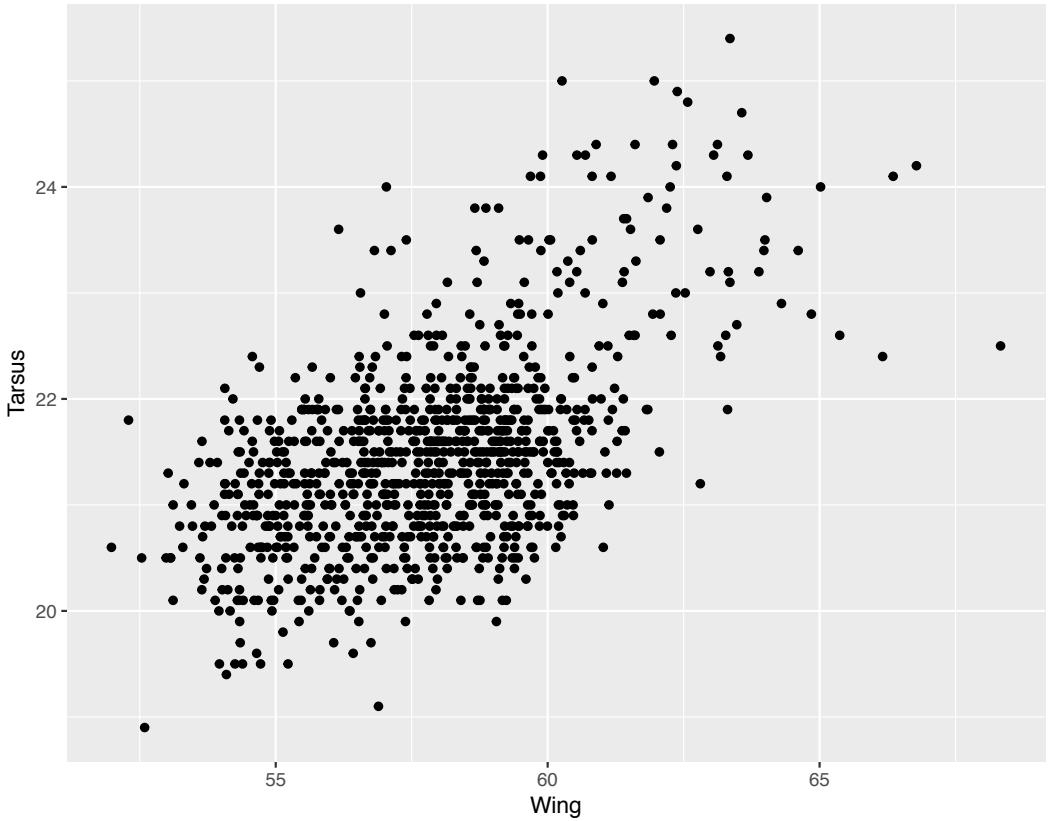
Example 2: Avoid overplotting of symbols

```
ggplot(Sparrows,
       aes(x = Wing, y = Tarsus))+
       geom_point(position=position_jitter(width=0.5, height=0))
```

Graphics with ggplot2

Creating a **scatterplot** with **ggplot()**

Example 2: Avoid overplotting of symbols



Graphics with ggplot2

Creating a **scatterplot** with **ggplot()**

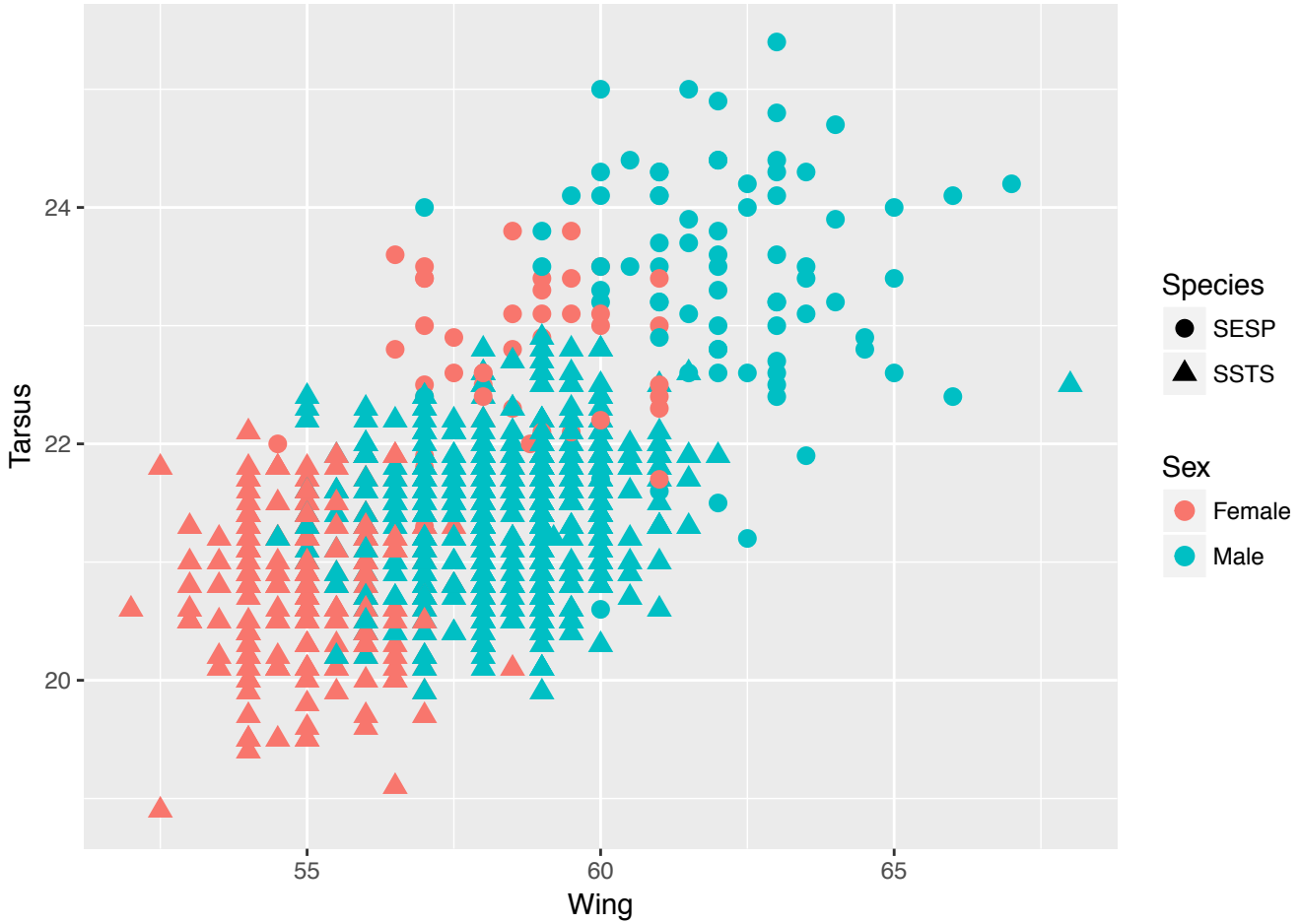
Example 3: Alter **colour, shape,** and **size** of symbols

```
ggplot(Sparrows,
       aes(x = Wing, y = Tarsus, colour = Sex, shape = Species)) +
       geom_point(size = 2)
```

Graphics with ggplot2

Creating a **scatterplot** with **ggplot()**

Example 3:

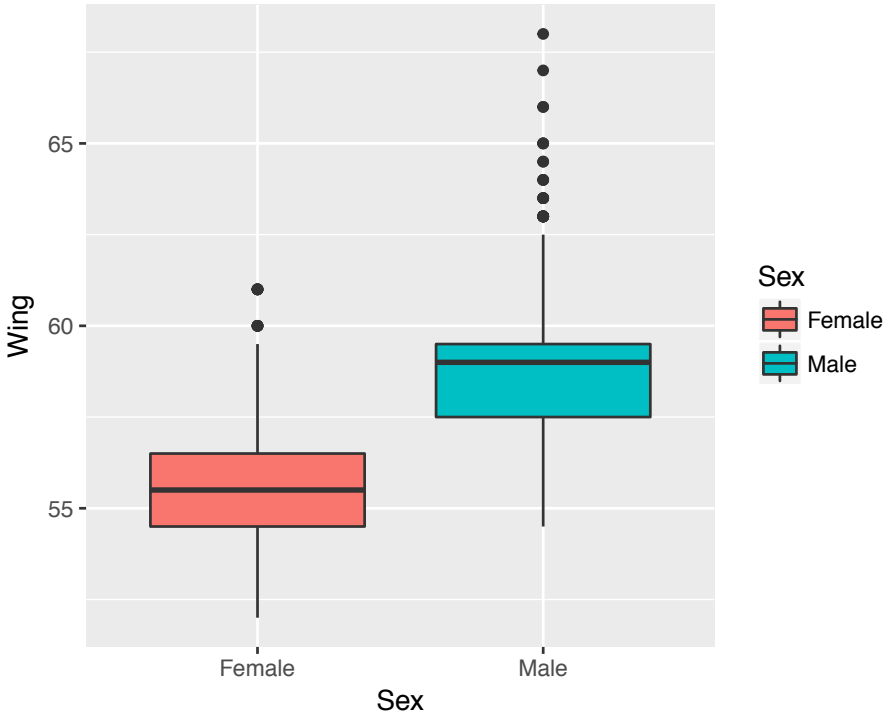


Graphics with ggplot2

Creating a **boxplot** with **ggplot()**

Example 1:

```
ggplot(Sparrows, aes(Sex, Wing, fill=Sex)) +
        geom_boxplot()
```



Graphics with ggplot2

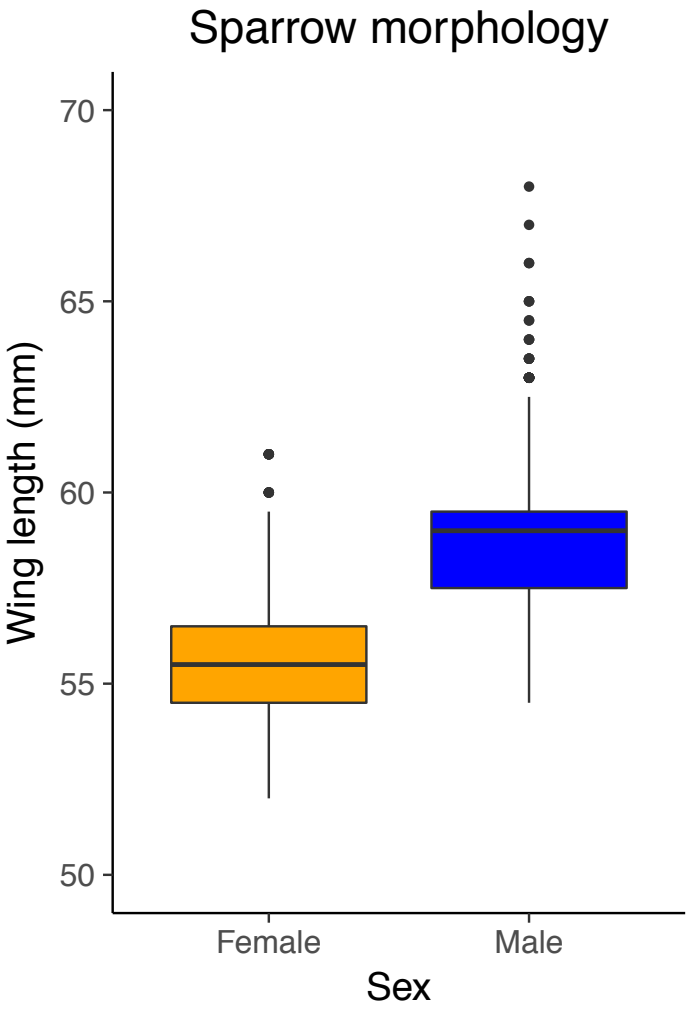
Preparing plots for publication

- Title and axis labels
- Range of axes
- Colours
- Overall appearance (themes)
- Text size
- Legend

Graphics with ggplot2

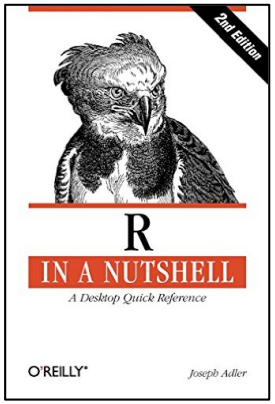
Preparing plots for publication

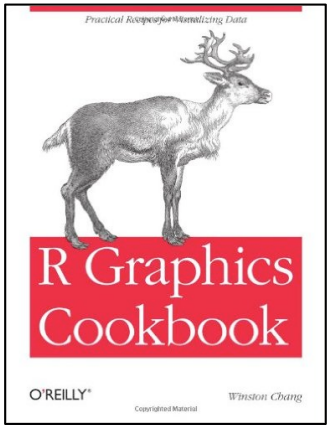
- Title and axis labels
- Range of axes
- Colours
- Overall appearance (themes)
- Text size
- Legend

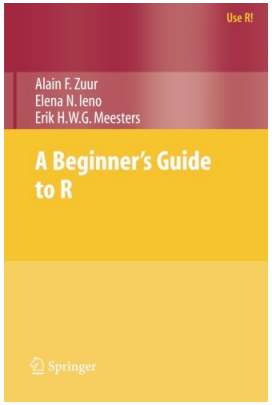


Further reading

Books







Internet

http://docs.ggplot2.org/

http://www.cookbook-r.com/