



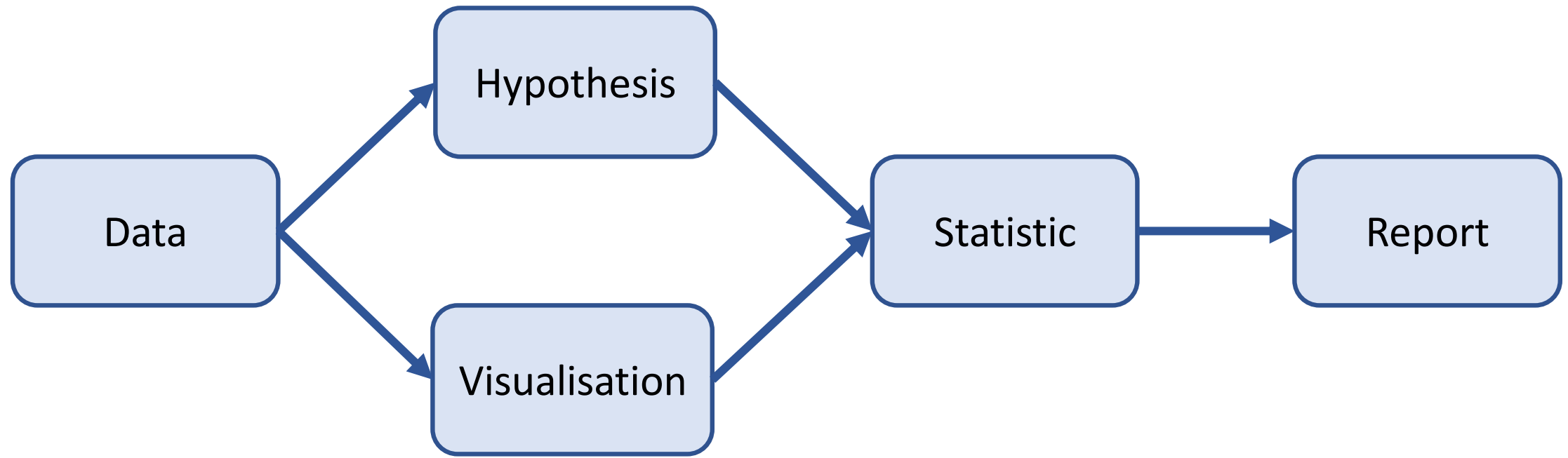
Using R as a Research Tool.

NERC E3 DTP Training – 6th November

Dr Susan Johnston, Institute of Evolutionary Biology

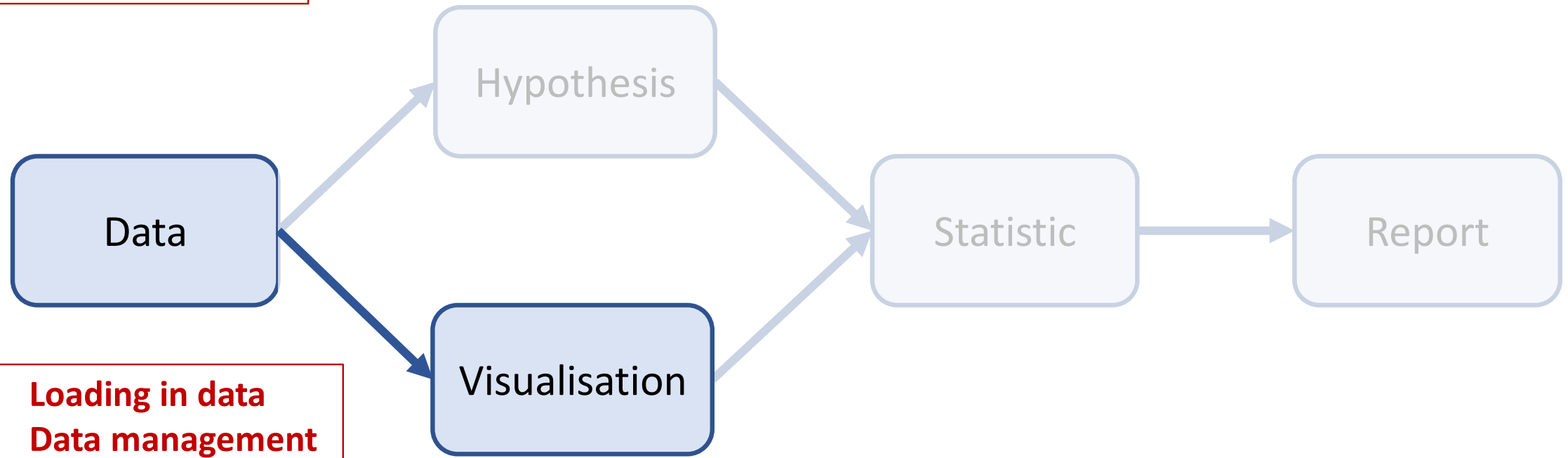
Demonstrators: Gergana Daskalova, John Godlee

Research project – typical workflow.



Research project – typical workflow.

Create an R Project

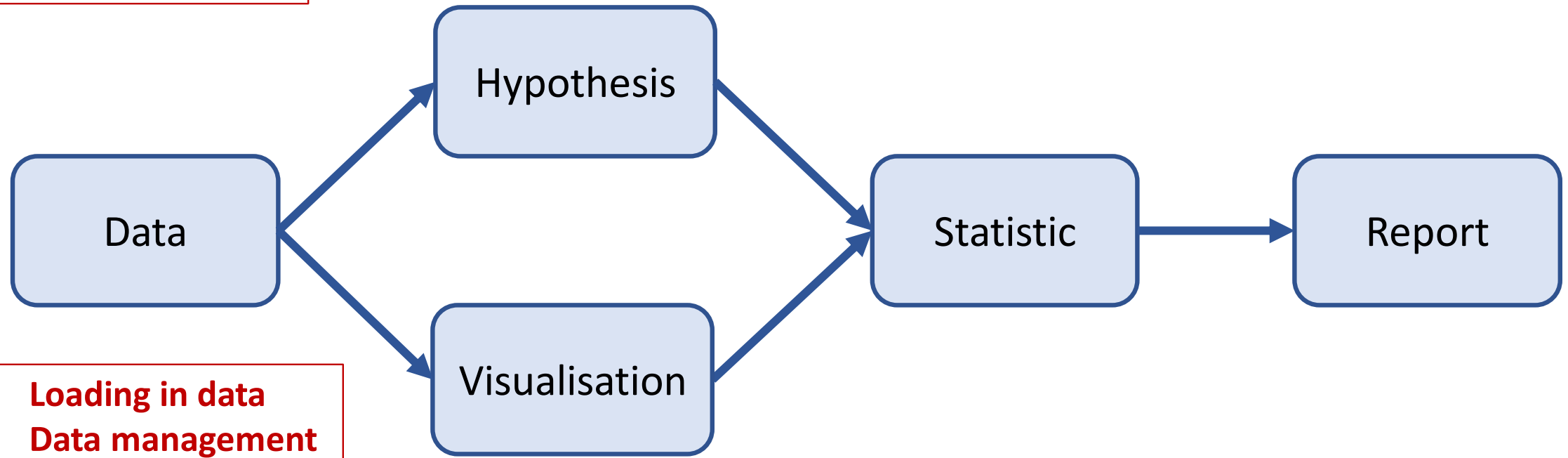


- **Loading in data**
- **Data management with dplyr**

- **Data visualisation with ggplot2**

Research project – typical workflow.

Create an R Project



- **Loading in data**
- **Data management with dplyr**

- **Data visualisation with ggplot2**

Today's session:

- Carry out basic statistics and visualisations, including:
 - Chi-squared test with `chisq.test()`
 - 2-sample t-test with `t.test()`
 - Linear regression with `lm()`
- Write, embed and render code and results into an HTML document.

Report writing

Using R as a Research Tool.

Dr Susan Johnston: Susan.Johnston@ed.ac.uk

Demonstrators: Gergana Dalaskova, John Godlee.
Hat-Tips to Kyle Dexter, The Coding Club and R4all.

November 6, 2017

1 Introduction

1.1 What is R?

R began its life in New Zealand in 1993 as a language and environment for statistical computing and graphics. It is an interpreted programming language, meaning that rather than pointing and clicking, the user types in commands. It is **free** and works across all platforms.

1.2 Why use R?

R Base Graphics: An Idiot's Guide

Comments (-)

Share

Hide Toolbars

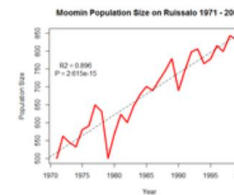
One of the most powerful functions of R is its ability to produce a wide range of graphics to quickly and easily visualise data. Plots can be replicated, modified and even publishable with just a handful of commands.

Making the leap from chiefly graphical programmes, such as Excel and Sigmaplot, may seem tricky. However, with a basic knowledge of R, just investing a few hours could completely revolutionise your data visualisation and workflow. Trust me - it's worth it.

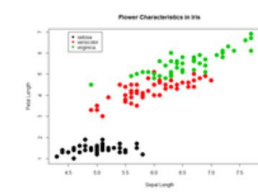
Last year, I presented an informal course on the basics of R Graphics University of Turku. In this blog post, I am providing some of the slides and the full code from that practical, which shows how to build different plot types using the basic (i.e. pre-installed) graphics in R, including:



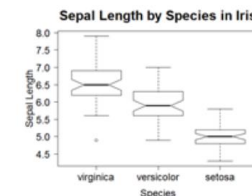
1. Basic Histogram



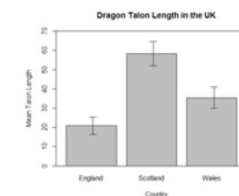
2. Line Graph with Regression



3. Scatterplot with Legend



4. Boxplot with reordered/
formatted axes



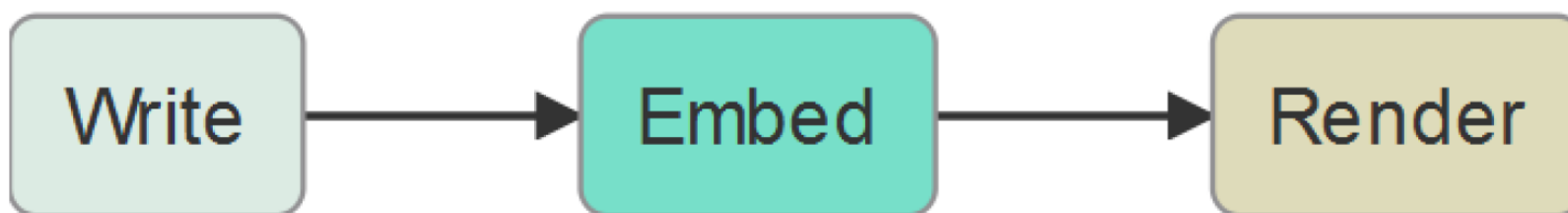
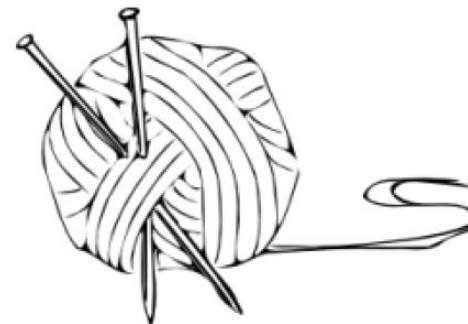
5. Boxplot with Error Bars

LaTeX and R Sweave

knitr to HTML

knitr

Elegant, flexible and fast
dynamic report generation with R



The knitr package allows R code and document templates to be compiled into a single report containing text, results and figures.

Markdown Language

- Lightweight mark-up language
- Basic text formatting, adding images, creating lists, etc.
- Can embed code and results, making it useful for reproducible research.
- Looks very much like regular text with a few extra characters (#, *, etc)

R Markdown Cheat Sheet

learn more at rmarkdown.rstudio.com

rmarkdown 0.2.50 Updated: 8/14



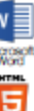
1. Workflow R Markdown is a format for writing reproducible, dynamic reports with R. Use it to embed R code and results into slideshows, pdfs, html documents, Word files and more. To make a report:

i. **Open** - Open a file that uses the .Rmd extension.

ii. **Write** - Write content with the easy to use R Markdown syntax

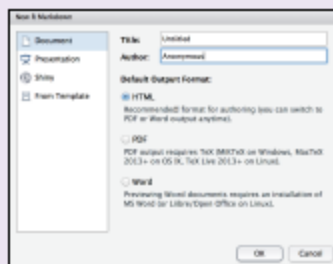
iii. **Embed** - Embed R code that creates output to include in the report

iv. **Render** - Replace R code with its output and transform the report into a slideshow, pdf, html or ms Word file.



2. Open File Start by saving a text file with the extension .Rmd, or open an RStudio Rmd template

- In the menu bar, click **File ► New File ► R Markdown...**
- A window will open. Select the class of output you would like to make with your .Rmd file
- Select the specific type of output to make with the radio buttons (you can change this later)
- Click OK



3. Markdown Next, write your report in plain text. Use markdown syntax to describe how to format text in the final report.

syntax

Plain text

End a line with two spaces to start a new paragraph.

`*italics*` and `_italics_`

`**bold**` and `__bold__`

`superscript^2*`

`--strikethrough--`

`[link](www.rstudio.com)`

`# Header 1`

`## Header 2`

`### Header 3`

`#### Header 4`

`##### Header 5`

`##### Header 6`

`endash: --`

`emdash: ---`

`ellipsis: ...`

`inline equation: $A = \pi * r^2$`

`image: `

`horizontal rule (or slide break):`

`***`

`> block quote`

`* unordered list`

`* item 2`

`+ sub-item 1`

`+ sub-item 2`

`1. ordered list`

`2. item 2`

`+ sub-item 1`

`+ sub-item 2`

`Table Header | Second Header`

`Table Cell | Cell 2`

`Cell 3 | Cell 4`

becomes

Plain text

End a line with two spaces to start a new paragraph.

italics and italics

bold and bold

^{superscript²}

~~strikethrough~~

[link](http://www.rstudio.com)

Header 1

Header 2

Header 3

Header 4

Header 5

Header 6

endash: --

emdash: ---

ellipsis: ...

inline equation: $A = \pi * r^2$

image: 

horizontal rule (or slide break):

> block quote

* unordered list

* item 2

+ sub-item 1

+ sub-item 2

1. ordered list

2. item 2

+ sub-item 1

+ sub-item 2

Table Header | Second Header

Table Cell | Cell 2

Cell 3 | Cell 4

4. Choose Output Write a YAML header that explains what type of document to build from your R Markdown file.

YAML

A YAML header is a set of key: value pairs at the start of your file. Begin and end the header with a line of three dashes (---)

```
---
title: "Untitled"
author: "Anonymous"
output: html_document
---
```

This is the start of my report. The above is metadata saved in a YAML header.

The RStudio template writes the YAML header for you

The output value determines which type of file R will build from your .Rmd file (in Step 6)

`output: html_document` html file (web page)

`output: pdf_document` pdf document

`output: word_document` Microsoft Word .docx

`output: beamer_presentation` beamer slideshow (pdf)

`output: ioslides_presentation` ioslides slideshow (html)



Worked example...

Today's session:

- Carry out basic statistics and visualisations, including:
 - Chi-squared test with `chisq.test()`
 - 2-sample t-test with `t.test()`
 - Linear regression with `lm()`
- Write, embed and render code and results into an HTML document.

Chi-squared (χ^2) contingency table

- Analysis of count data



Adalia bipunctata

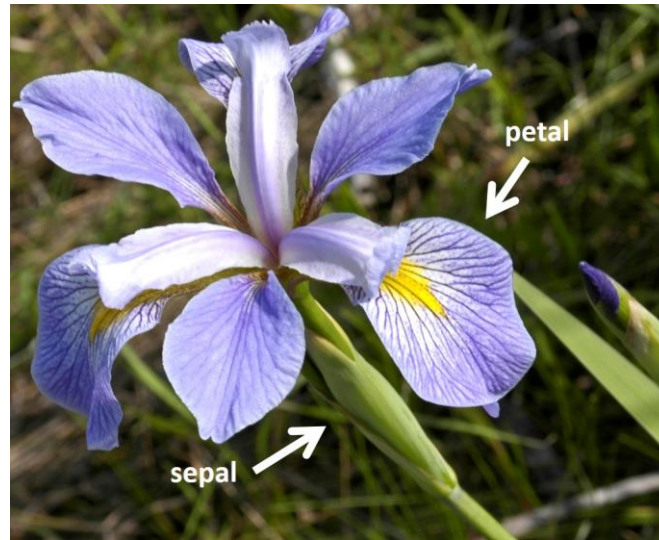
Are lark ladybirds more likely to live in industrial (dark) backgrounds?

Two sample t-test

- Compares the mean of two groups.



Iris versicolor

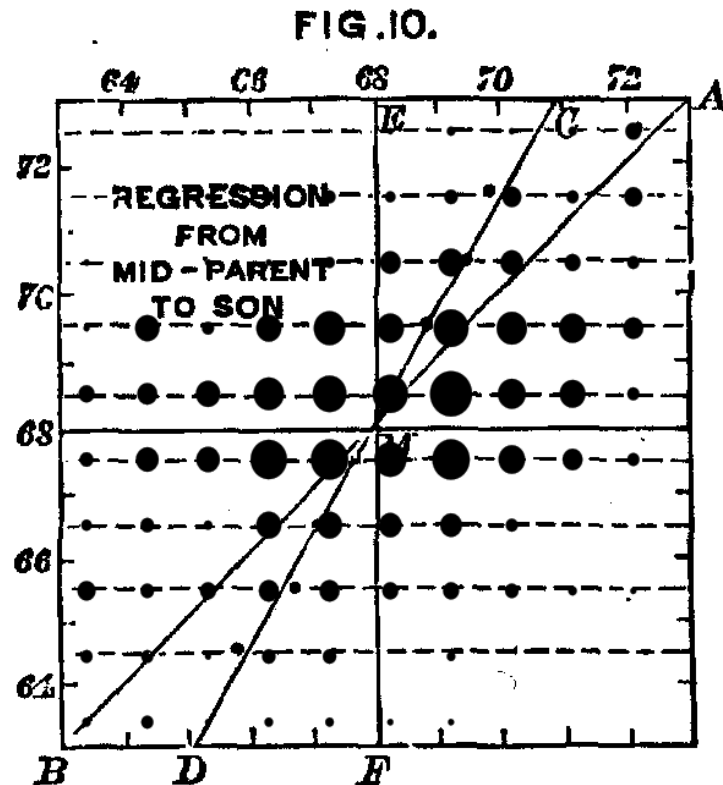


Iris virginica

Are sepal lengths significantly different between these two species of *Iris*?

Simple linear regression

- Summarise and study relationships between two continuous variables.



Francis Galton's Human
Height Data Set 1886

Does daughter height vary
relative to the mean
height of her parents?

Need help?

- **ASK US!**
- Use `?`, `str()`, `glimpse()` to explore functions and objects.
- Coding Club tutorials: <https://ourcodingclub.github.io/tutorials/>
- Google & StackOverflow
- Feel free to write as a normal R script!