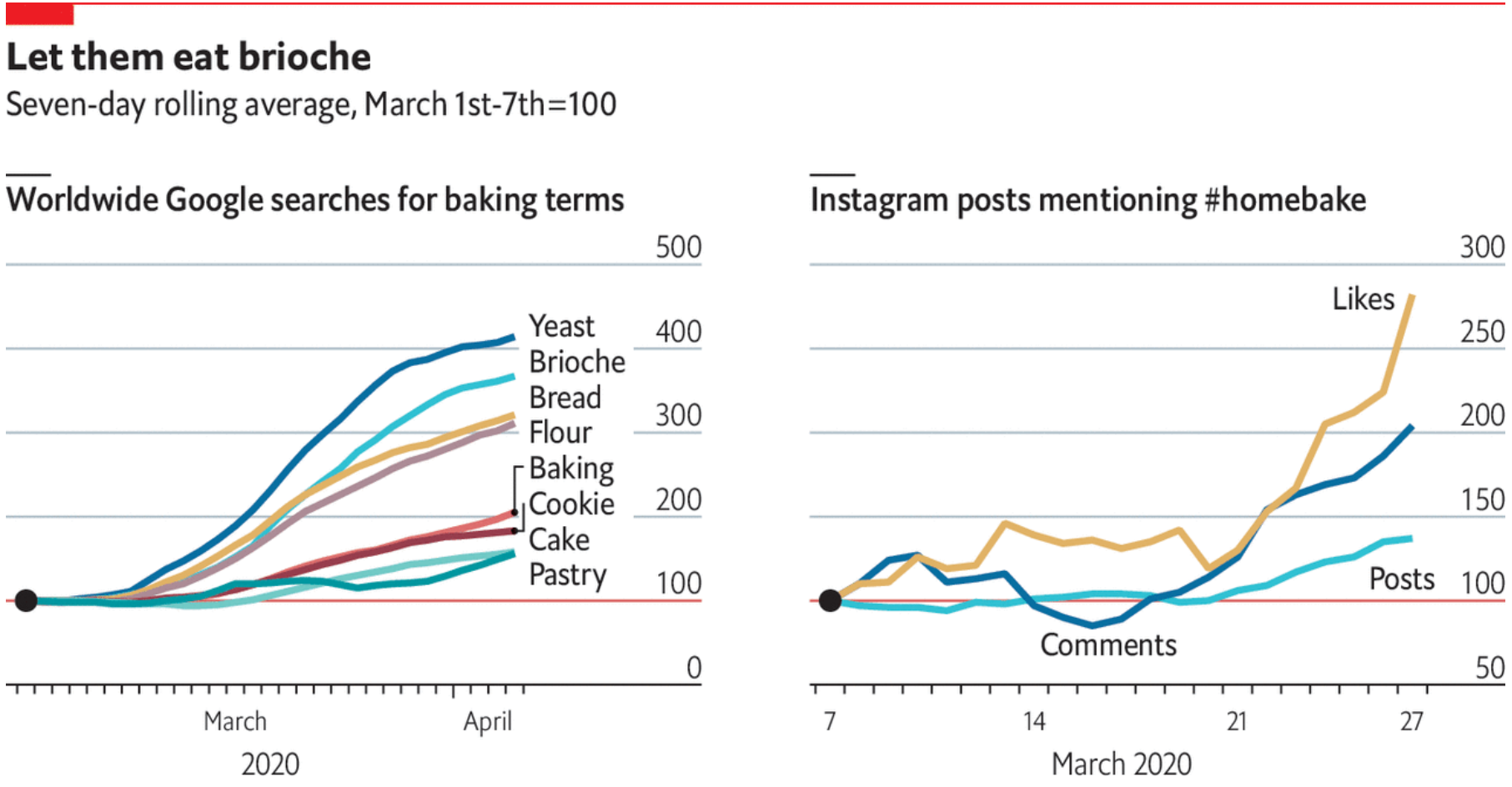
Understanding and presenting your data.

Field Zoology 3
Dr Susan Johnston
Susan.Johnston@ed.ac.uk

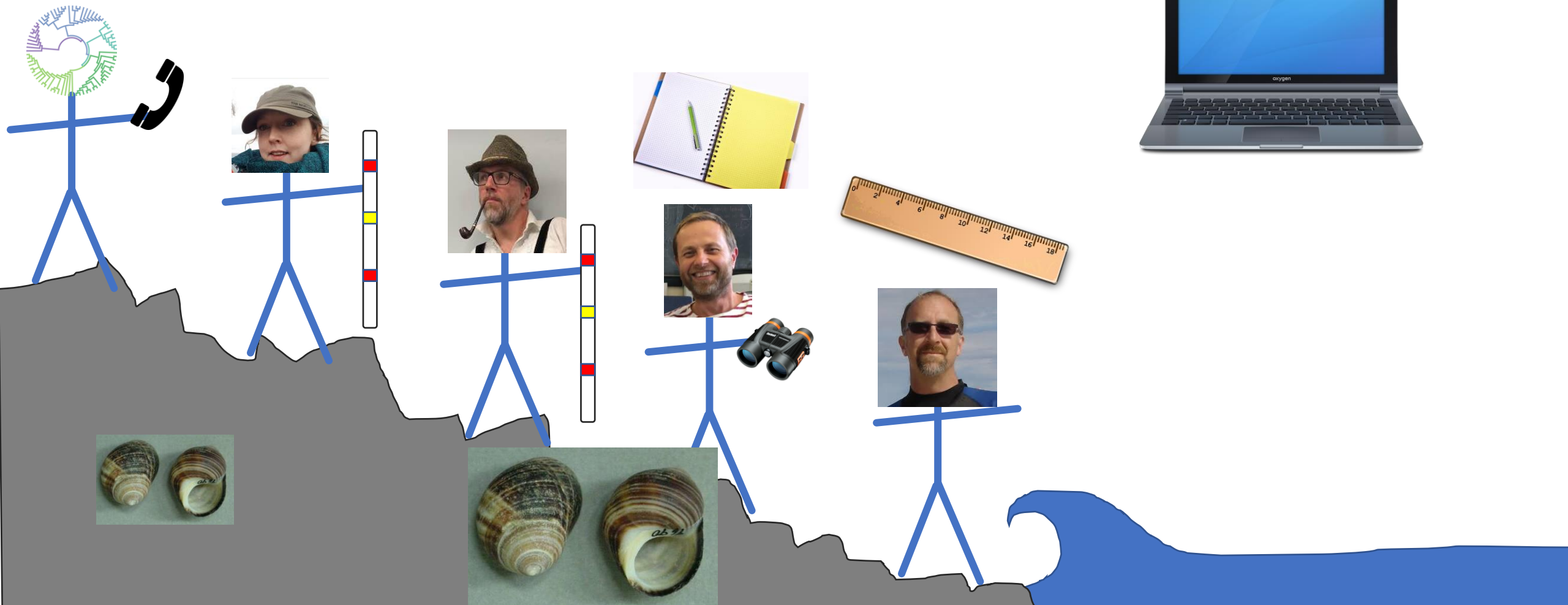
Data visualisation



Overview

- Why visualisation is important for understanding data.
- How visualisation can point you towards the correct statistical tests.
- How to present a graph properly & plot customisations.

Do *Littorina* vary in size with shore height?



x = Height on the sea shore (m)
y = Shell size (mm)

Anscombe's Quartet:

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

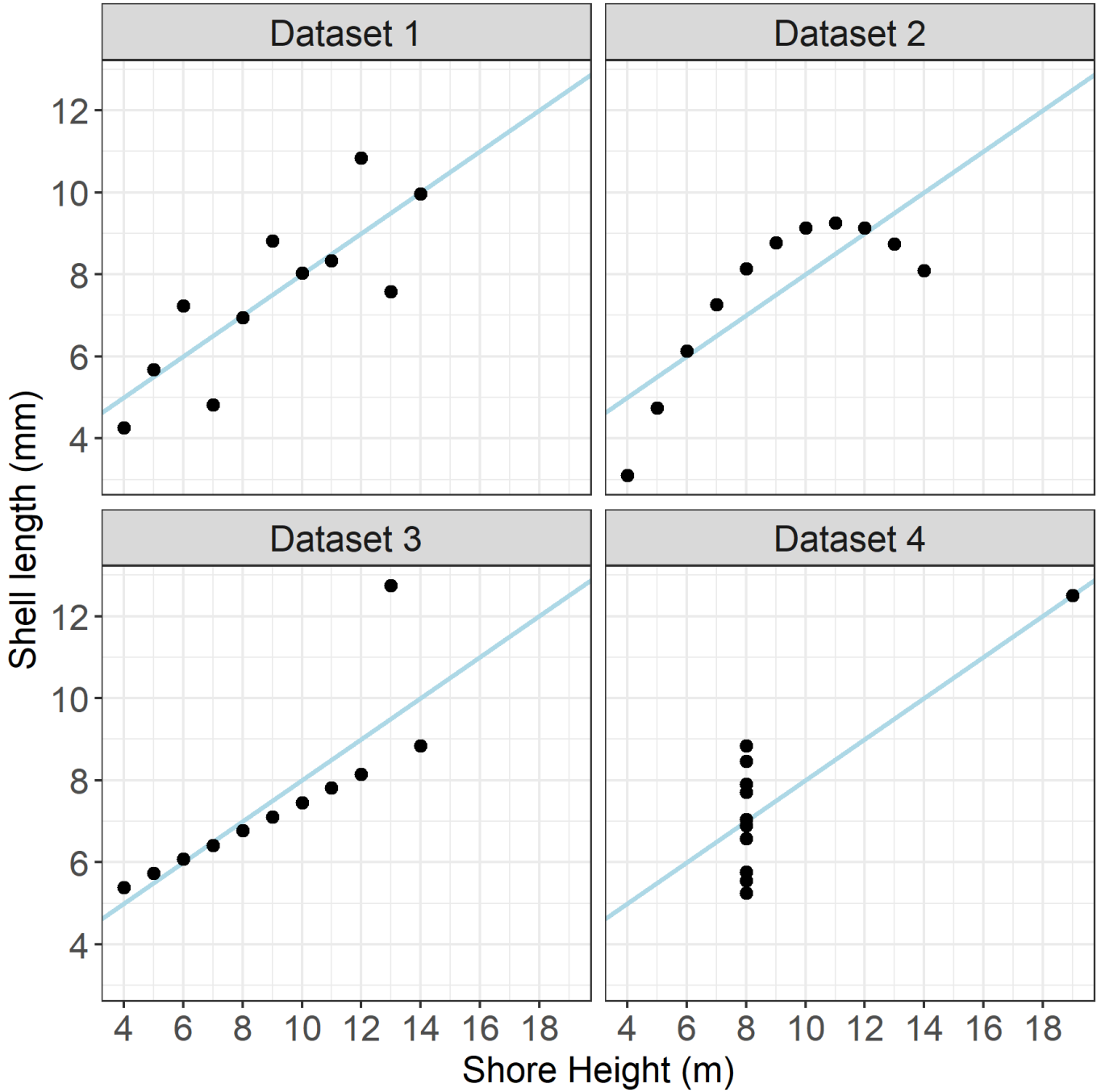
```
> fit1 <- lm(y1 ~ x1, data = anscombe)
> summary(fit1)

Call:
lm(formula = y1 ~ x1, data = anscombe)

Residuals:
     Min       1Q   Median       3Q      Max
-1.92127 -0.45577 -0.04136  0.70941  1.83882

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.0001     1.1247   2.667  0.02573 *
x1            0.5001     0.1179   4.241  0.00217 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.237 on 9 degrees of freedom
Multiple R-squared:  0.6665,    Adjusted R-squared:  0.6295
F-statistic: 17.99 on 1 and 9 DF,  p-value: 0.00217
```


Dataset 1
Dataset 2
Dataset 3
Dataset 4
Shell length (mm)
Shore Height (m)
4
6
8
10
12
14
16
18

Overview

- Why visualisation is important for understanding data.
- How visualisation can point you towards the correct statistical tests.
- How to present a graph properly.

Data visualisation with ggplot2

http://ggplot2.tidyverse.org/reference/

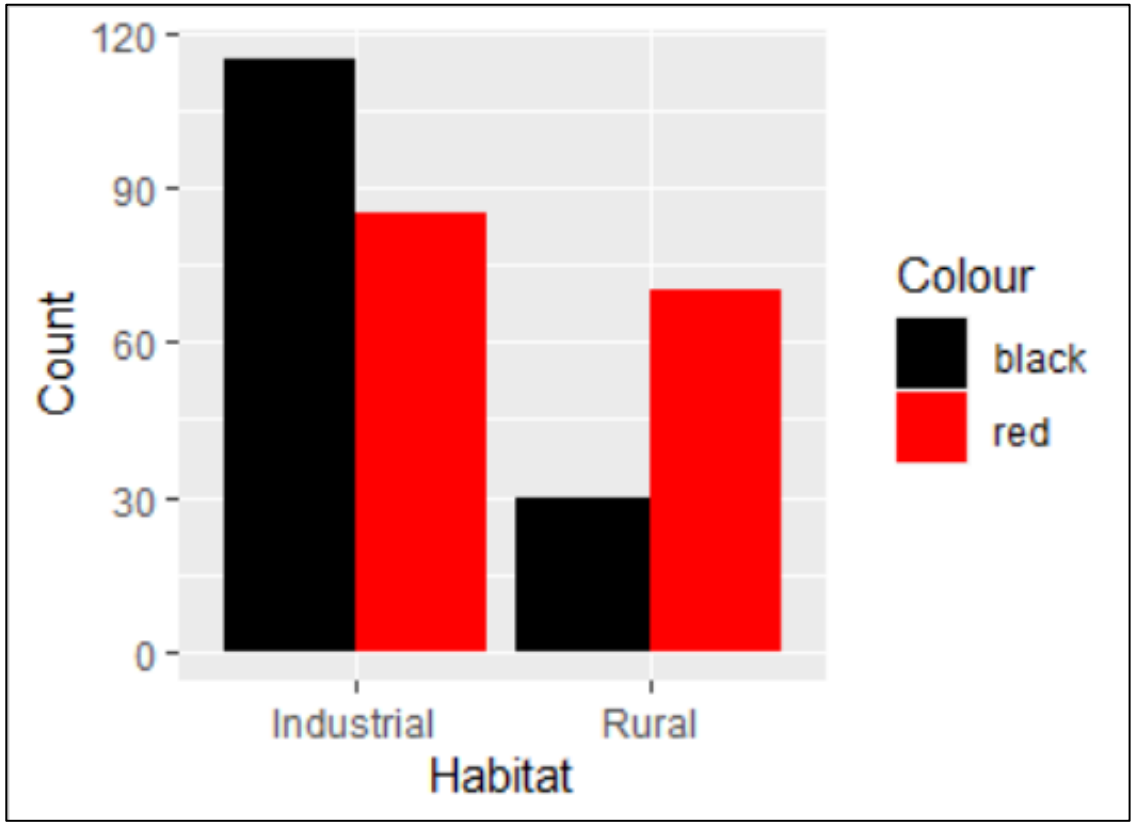
Base graphics…
http://rpubs.com/SusanEJohnston/7953

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

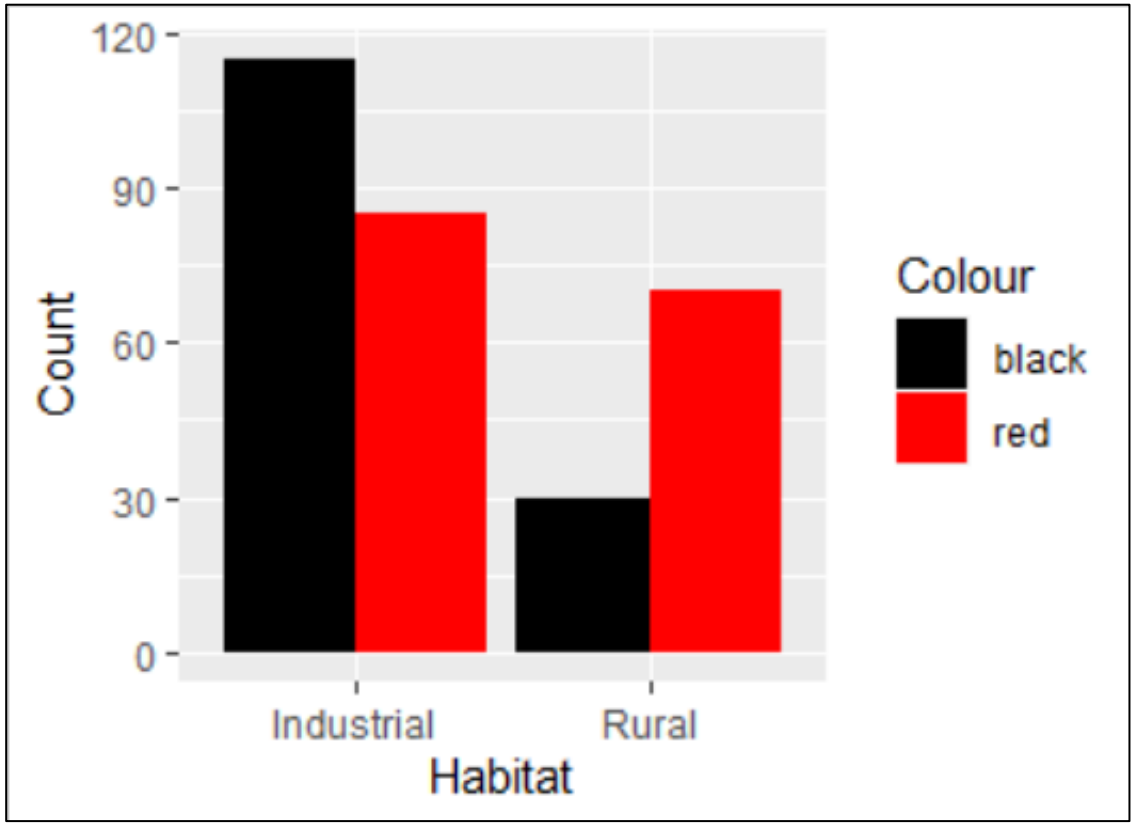
counts of objects/events in categories:
Chi-squared (χ^2)
contingency table

Adalia bipunctata

	Habitat	Colour	Count
1	Industrial	black	115
2	Industrial	red	85
3	Rural	black	30
4	Rural	red	70



	Habitat	Colour	Count
1	Industrial	black	115
2	Industrial	red	85
3	Rural	black	30
4	Rural	red	70

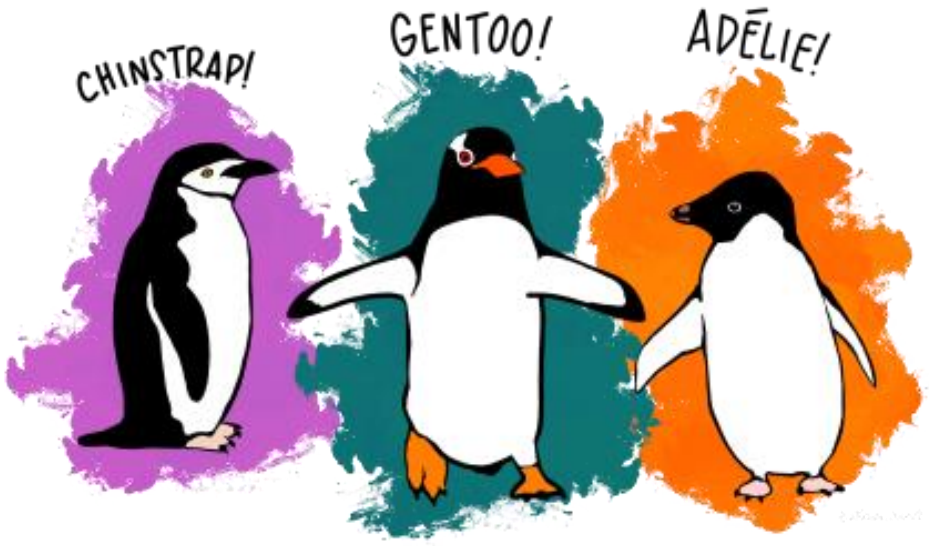


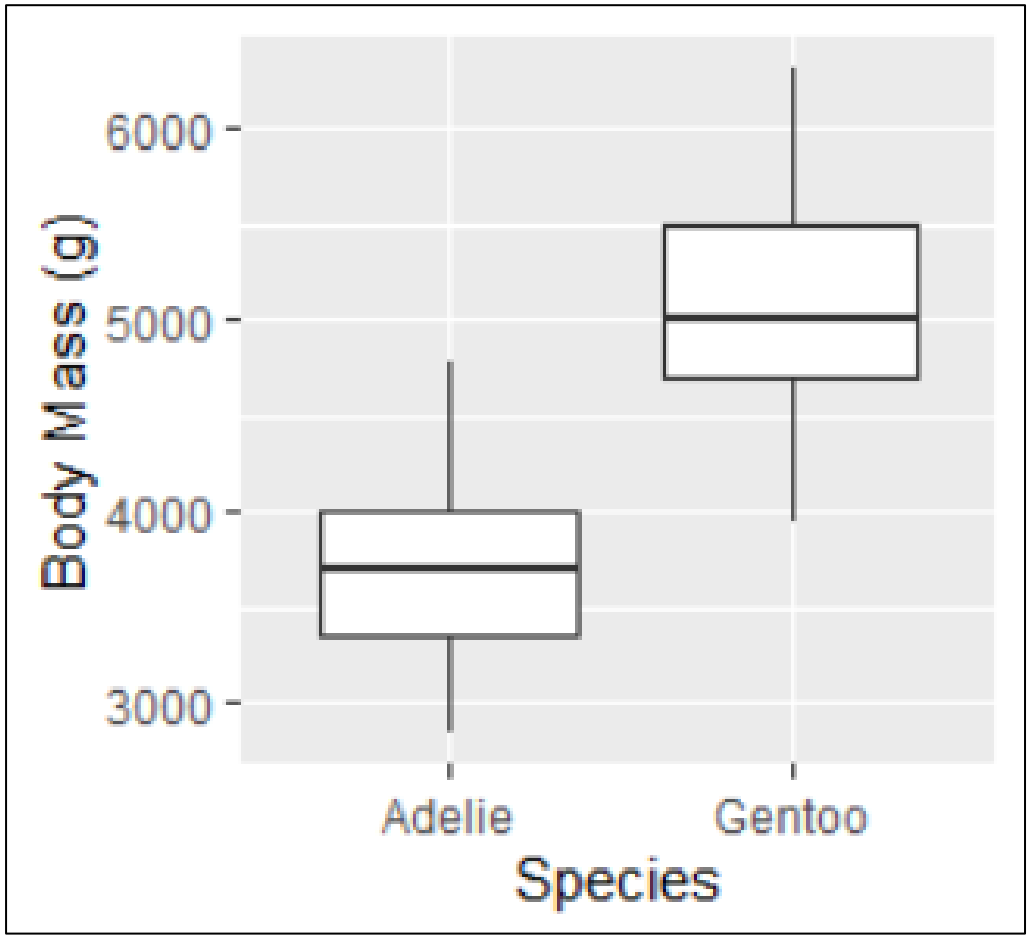

```
ggplot(ladybirds, aes(x = Habitat, y = Count, fill = Colour)) +
  geom_bar(stat = "identity", position = "dodge") +
  scale_fill_identity(guide = "legend")
```

These lines can be copied and pasted to create a barplot in the correct format.

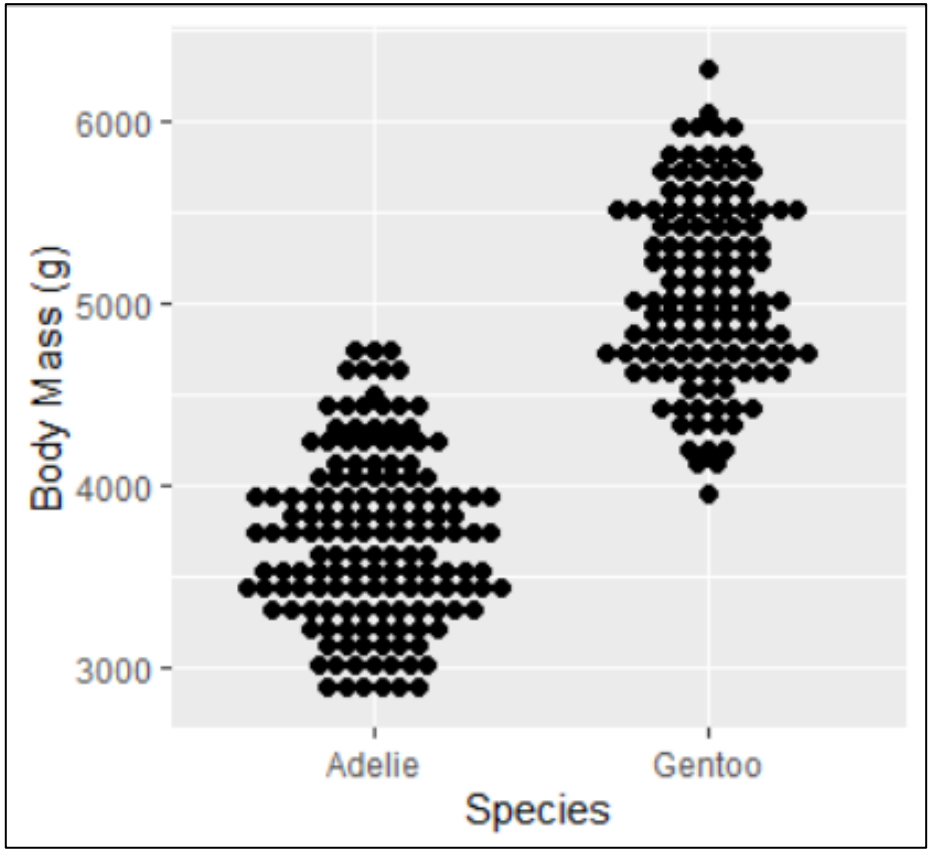
2. Does body weight differ significantly between Adelie and Gentoo penguins?

Comparing a continuous variable between groups: t-test, ANOVA

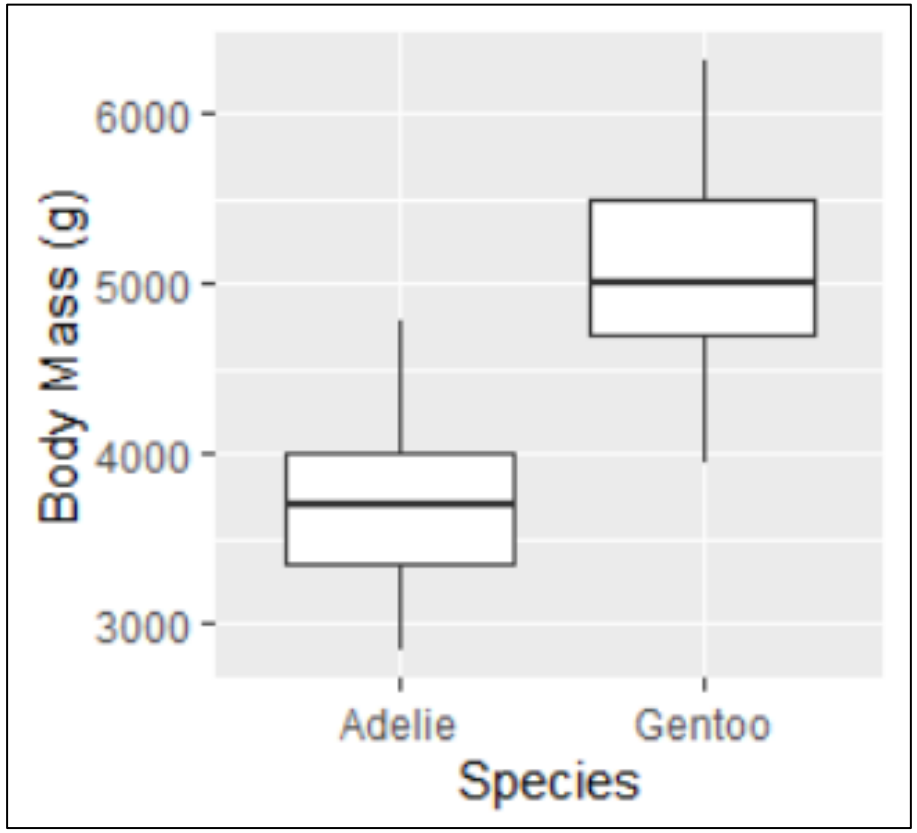




Explore your data….



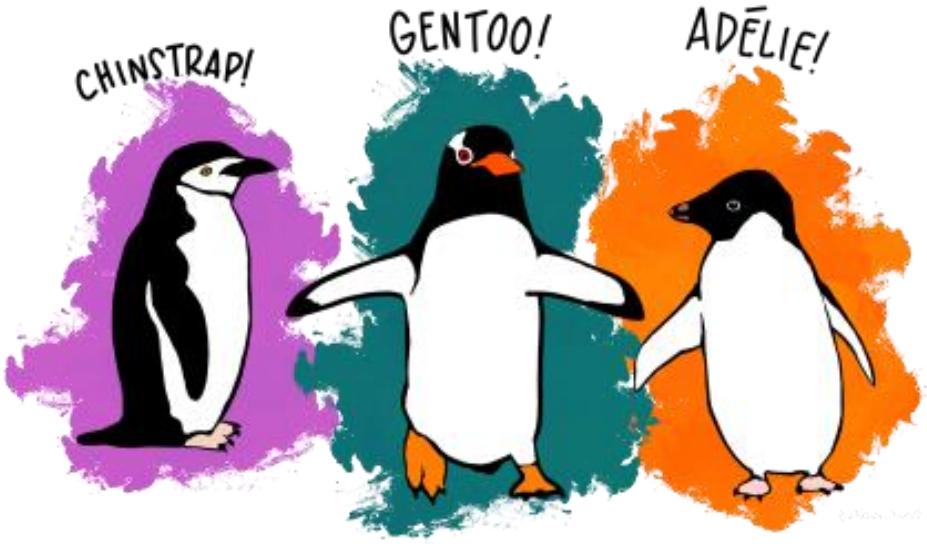

```
ggplot(penguins, aes(x = species, y = body_mass_g)) +
  geom_dotplot(binaxis = "y", stackdir = "center")
```

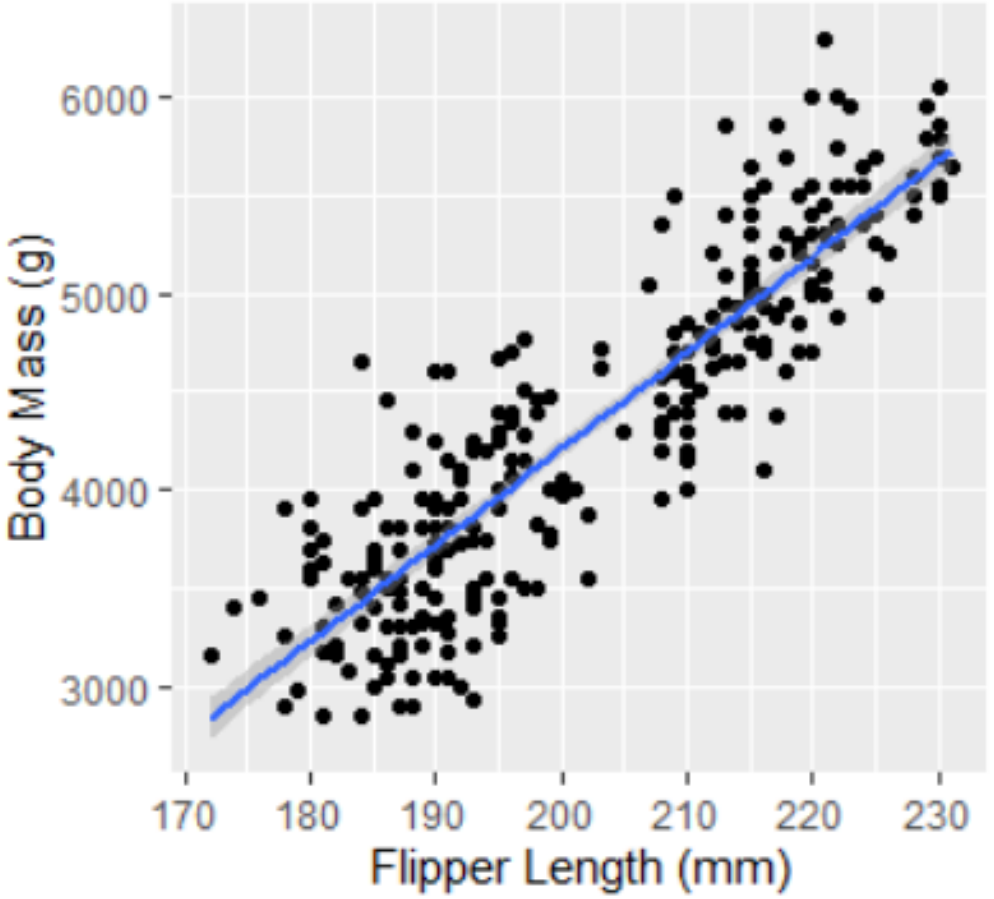


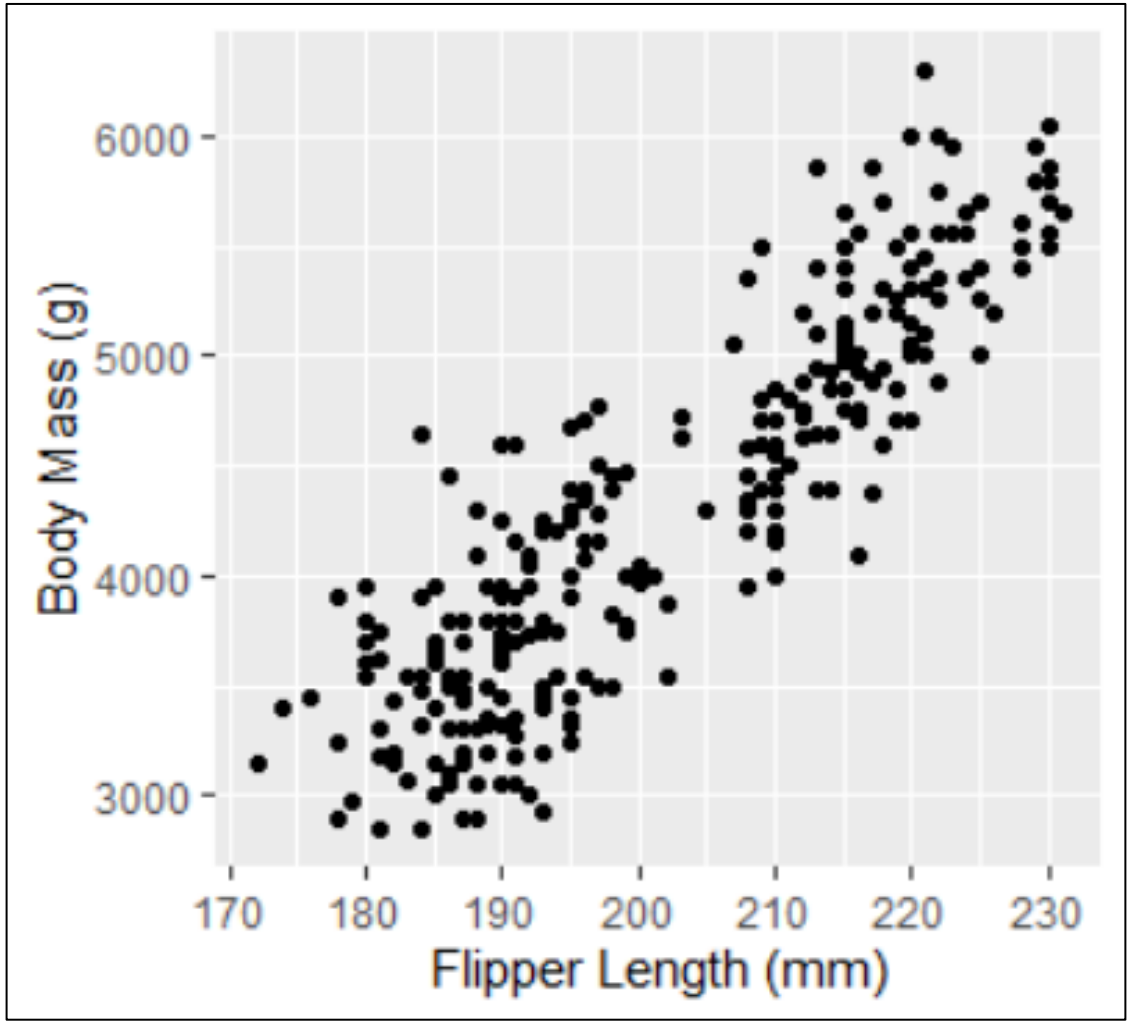

```
ggplot(penguins, aes(x = species, y = body_mass_g)) +
  geom_boxplot() +
  labs(x = "Species", y = "Body Mass (g)")
```

3. Does flipper length vary relative to body weight in penguins?

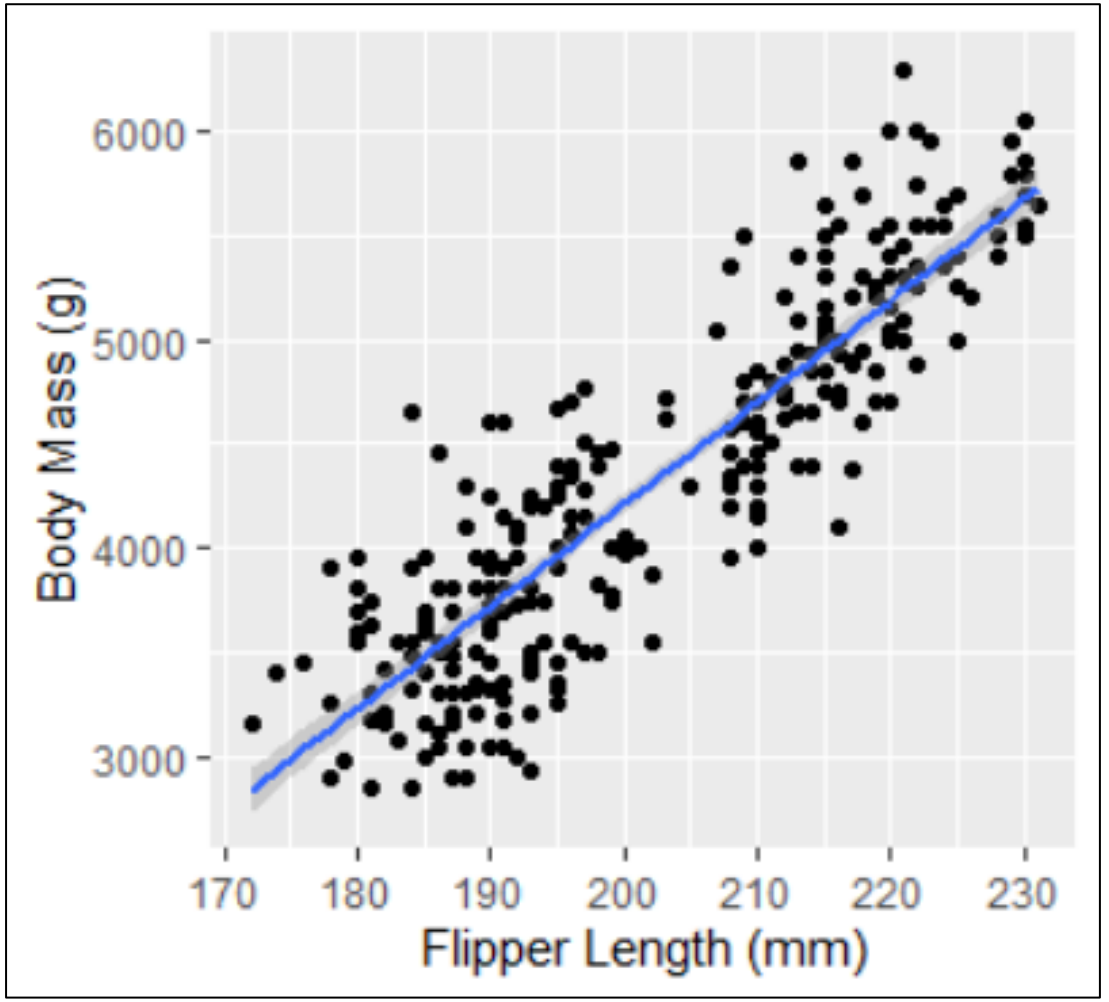
Relationship between continuous variables: correlation & regression.



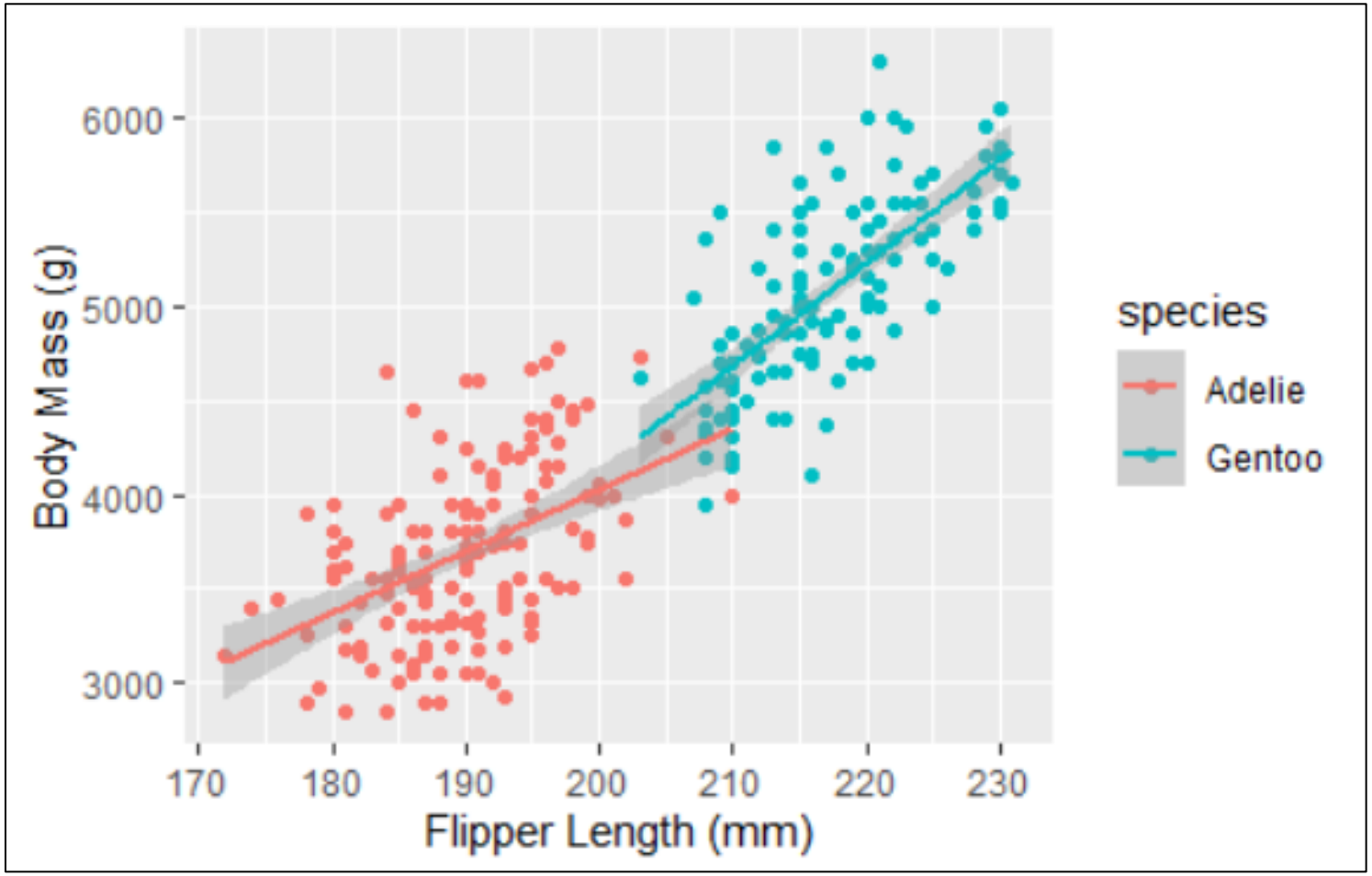





```
ggplot(penguins, aes(x = flipper_length_mm, y = body_mass_g)) +
  geom_point() +
  labs(x = "Flipper Length (mm)", y = "Body Mass (g)")
```




```
ggplot(penguins, aes(x = flipper_length_mm, y = body_mass_g)) +
  geom_point() +
  labs(x = "Flipper Length (mm)", y = "Body Mass (g)") +
  stat_smooth(method = "lm")
```

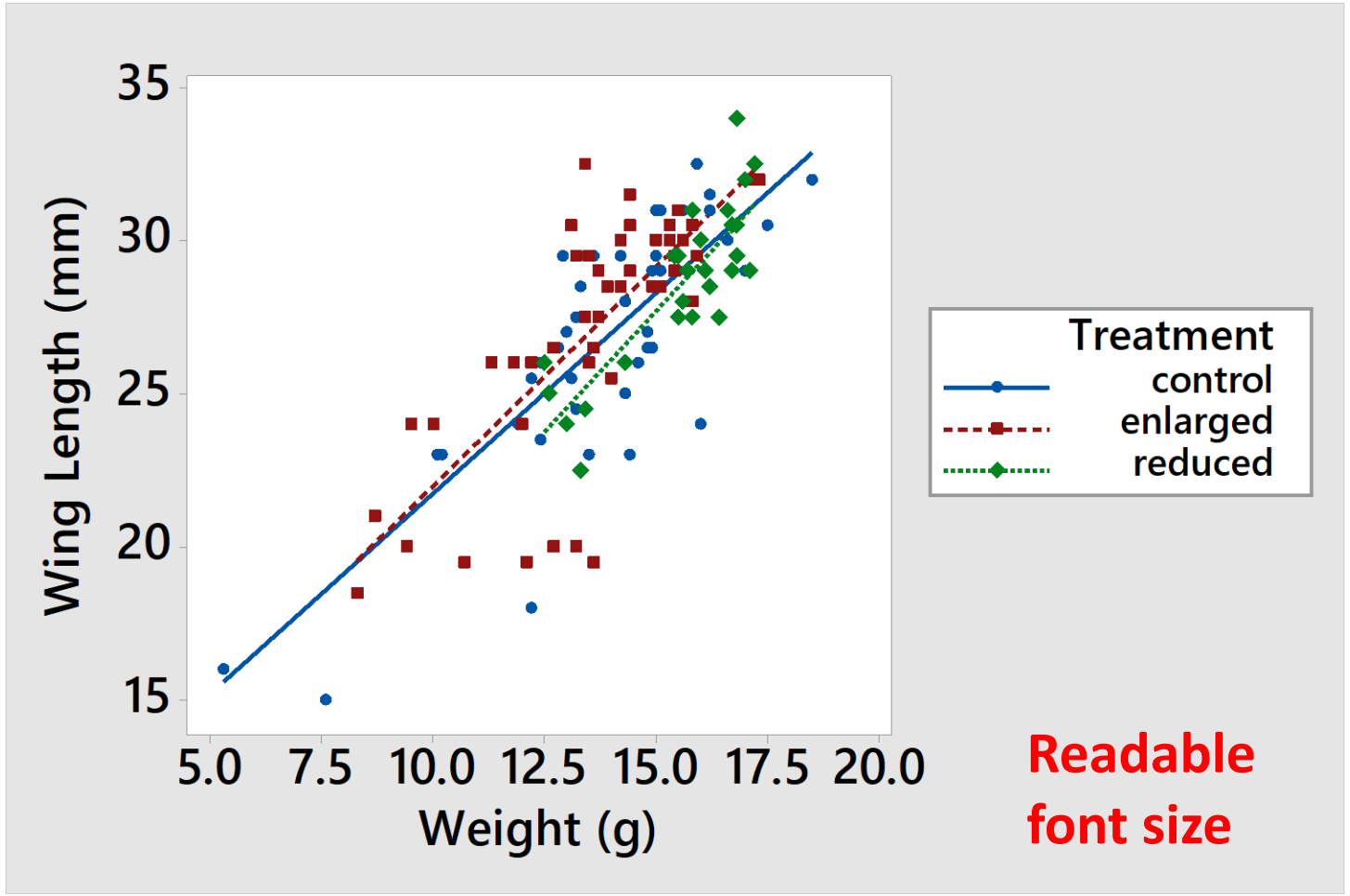



```
ggplot(penguins, aes(x = flipper_length_mm, y = body_mass_g, colour = species)) +
  geom_point() +
  labs(x = "Flipper Length (mm)", y = "Body Mass (g)") +
  stat_smooth(method = "lm")
```

Overview

- Why visualisation is important for understanding data.
- How visualisation can point you towards the correct statistical tests.
- How to present a graph properly.

Label the axes, including units



Use legends if needed, don't obscure the data.

Readable font size

Figure 1. Correlation between wing length (mm) and weight (g) in savannah sparrows. Symbols indicate the nest brood-size treatment: blue circles are the control treatment, red squares the enlarged nests and green diamonds are the reduced nests. There was a significant association between wing length and weight in all three treatment groups (linear regression, $P < 0.001$)

Title and description underneath

Make it easy for the reader/marker!

Customisation…

- See FZ3_Data_Visualisation_in_R_Examples.R

Questions?