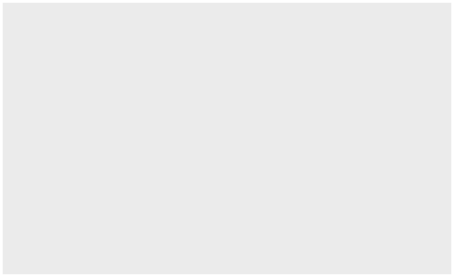
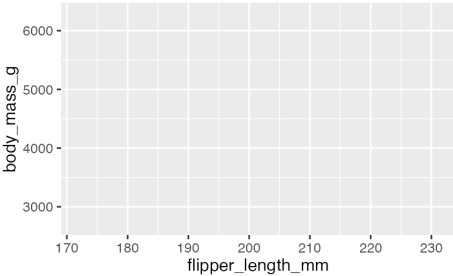
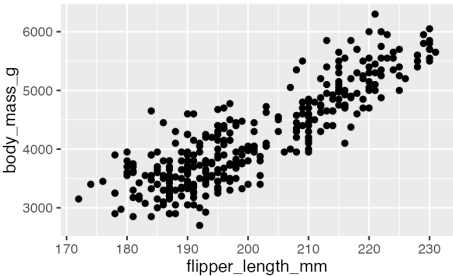


Code Sleuth

Geometries

Inference

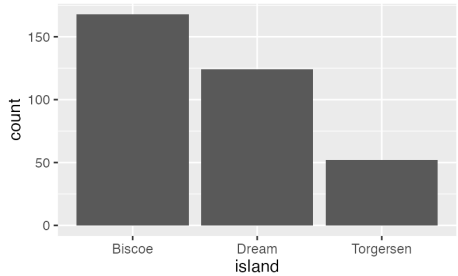
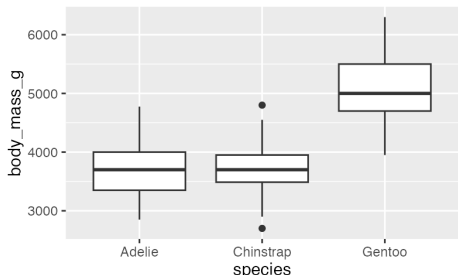
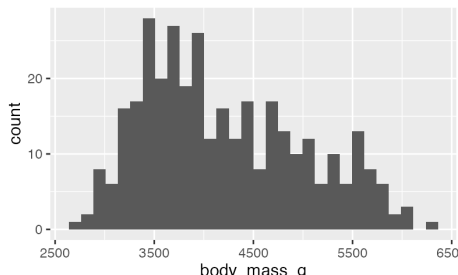
Code example	Output
<pre>ggplot(data = penguins)</pre>	
<pre>ggplot(data = penguins, mapping = aes(x = flipper_length_mm, y = body_mass_g))</pre>	
<pre>ggplot(data = penguins, mapping = aes(x = flipper_length_mm, y = body_mass_g)) + geom_point()</pre>	

How does the “mapping” parameter in ggplot change the figure?

How does adding “geom_point()” change the figure?

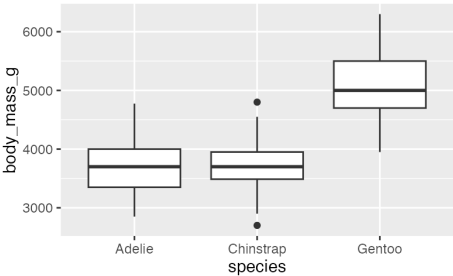
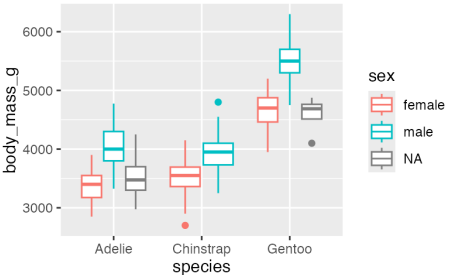
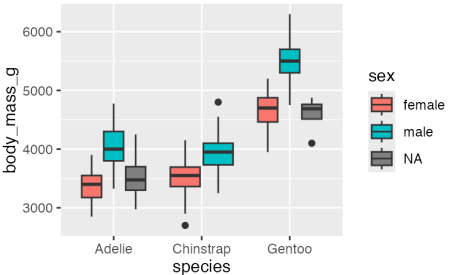
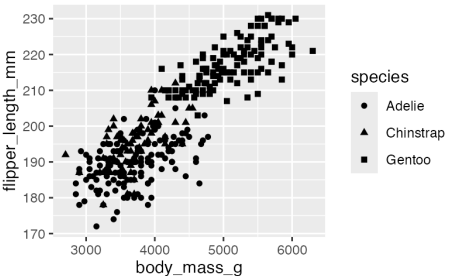
Prediction

Which code chunk do you think is associated with each output?

Code example	Output
<pre>ggplot(data = penguins, mapping = aes(x = body_mass_g)) + geom_histogram()</pre>	 A histogram showing the distribution of body mass for penguins across three islands: Biscoe, Dream, and Torgersen. The x-axis is labeled 'island' and the y-axis is labeled 'count'. The bars are dark gray. Biscoe has the highest count (around 160), followed by Dream (around 120), and Torgersen (around 50).
<pre>ggplot(data = penguins, mapping = aes(x = island)) + geom_bar()</pre>	 A boxplot showing the distribution of body mass for penguins across three species: Adelle, Chinstrap, and Gentoo. The x-axis is labeled 'species' and the y-axis is labeled 'body_mass_g'. The boxes are white with black outlines. Adelle and Chinstrap have similar distributions, while Gentoo has a higher median and more spread.
<pre>ggplot(data = penguins, mapping = aes(x = species, y = body_mass_g)) + geom_boxplot()</pre>	 A histogram showing the distribution of body mass for penguins across three species: Adelle, Chinstrap, and Gentoo. The x-axis is labeled 'body_mass_g' and the y-axis is labeled 'count'. The bars are dark gray. The distribution is skewed to the right, with a peak around 3500g.

Aesthetics

Inference

Code example	Output
<pre>ggplot(data = penguins, mapping = aes(x = species, y = body_mass_g)) + geom_boxplot()</pre>	 A boxplot showing the distribution of body mass (in grams) for three penguin species: Adelie, Chinstrap, and Gentoo. The y-axis is labeled 'body_mass_g' and ranges from 3000 to 6000. The x-axis is labeled 'species'. The plot shows that Gentoo penguins have the highest median body mass, followed by Adelie, and then Chinstrap.
<pre>ggplot(data = penguins, mapping = aes(x = species, y = body_mass_g, color = sex)) + geom_boxplot()</pre>	 A boxplot showing the distribution of body mass (in grams) for three penguin species, faceted by sex. The y-axis is labeled 'body_mass_g' and ranges from 3000 to 6000. The x-axis is labeled 'species'. The legend indicates that red represents female, teal represents male, and grey represents NA. The plot shows that for each species, males generally have higher body mass than females.
<pre>ggplot(data = penguins, mapping = aes(x = species, y = body_mass_g, fill = sex)) + geom_boxplot()</pre>	 A boxplot showing the distribution of body mass (in grams) for three penguin species, faceted by sex. The y-axis is labeled 'body_mass_g' and ranges from 3000 to 6000. The x-axis is labeled 'species'. The legend indicates that red represents female, teal represents male, and grey represents NA. The plot shows that for each species, males generally have higher body mass than females.
<pre>ggplot(data = penguins, mapping = aes(x = body_mass_g, y = flipper_length_mm, shape = species)) + geom_point()</pre>	 A scatter plot showing the relationship between body mass (in grams) on the x-axis and flipper length (in millimeters) on the y-axis. The x-axis ranges from 3000 to 6000, and the y-axis ranges from 170 to 230. The legend indicates that circles represent Adelle, triangles represent Chinstrap, and squares represent Gentoo. The plot shows a strong positive correlation between body mass and flipper length across all species.

Topic: *Grammar of graphics*

The *aesthetics* of the figure are defined by the “mapping” parameter, using the “`aes()`” function. What do you think each of the following aesthetics does in the figure?

- `x`
- `y`
- `color`
- `fill`
- `shape`

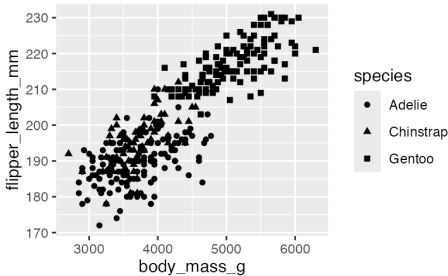
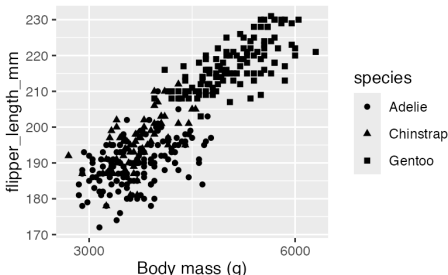
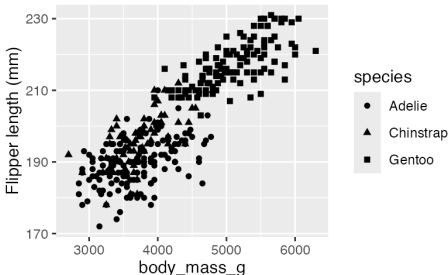
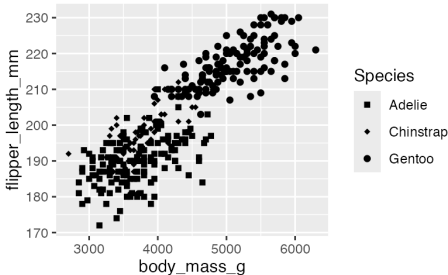
Prediction

What is the difference between the figures made by the following code examples?

Code example	Difference
<pre>ggplot(data = penguins, mapping = aes(x = body_mass_g, y = flipper_length_mm, shape = species)) + geom_point()</pre>	
<pre>ggplot(data = penguins, mapping = aes(x = body_mass_g, y = flipper_length_mm, color = species)) + geom_point()</pre>	

Scales

Inference

Code example	Output
<pre>ggplot(data = penguins, mapping = aes(x = body_mass_g, y = flipper_length_mm, shape = species)) + geom_point()</pre>	 A scatter plot showing the relationship between body mass (g) on the x-axis and flipper length (mm) on the y-axis for three penguin species. The x-axis ranges from 3000 to 6000, and the y-axis ranges from 170 to 230. The legend indicates that points are colored by species: Adelle (blue), Chinstrap (orange), and Gentoo (green). The plot shows a positive correlation between body mass and flipper length, with distinct clusters for each species.
<pre>ggplot(data = penguins, mapping = aes(x = body_mass_g, y = flipper_length_mm, shape = species)) + geom_point() + scale_x_continuous(name = "Body mass (g)", n.breaks = 2)</pre>	 A scatter plot similar to the first one, but with a single major break on the x-axis. The x-axis is labeled "Body mass (g)" and has two major tick marks at 3000 and 6000. The y-axis is labeled "flipper_length_mm" and ranges from 170 to 230. The legend indicates that points are colored by species: Adelle (blue), Chinstrap (orange), and Gentoo (green). The plot shows a positive correlation between body mass and flipper length, with distinct clusters for each species.
<pre>ggplot(data = penguins, mapping = aes(x = body_mass_g, y = flipper_length_mm, shape = species)) + geom_point() + scale_y_continuous(name = "Flipper length (mm)", n.breaks = 4)</pre>	 A scatter plot similar to the first one, but with 4 major breaks on the y-axis. The x-axis is labeled "body_mass_g" and ranges from 3000 to 6000. The y-axis is labeled "Flipper length (mm)" and ranges from 170 to 230. The legend indicates that points are colored by species: Adelle (blue), Chinstrap (orange), and Gentoo (green). The plot shows a positive correlation between body mass and flipper length, with distinct clusters for each species.
<pre>ggplot(data = penguins, mapping = aes(x = body_mass_g, y = flipper_length_mm, shape = species)) + geom_point() + scale_shape_manual(name = "Species", values = c(Adelle = "square", Chinstrap = "diamond", Gentoo = "circle"))</pre>	 A scatter plot similar to the first one, but with shapes mapped to species. The x-axis is labeled "body_mass_g" and ranges from 3000 to 6000. The y-axis is labeled "flipper_length_mm" and ranges from 170 to 230. The legend indicates that shapes are mapped to species: Adelle (square), Chinstrap (diamond), and Gentoo (circle). The plot shows a positive correlation between body mass and flipper length, with distinct clusters for each species.

Topic: *Grammar of graphics*

What do each of the “scale_*” functions change about the figure?

- `scale_x_continuous()`
- `scale_y_continuous()`
- `scale_shape_manual()`

What do each of the following parameters control about the scale?

- `name`
- `n.breaks`
- `values`