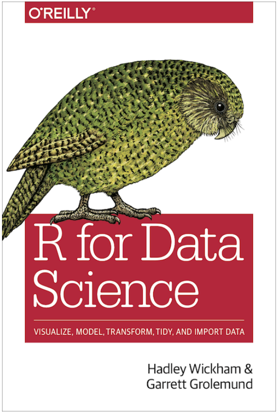
Socio-Informatics 348

Data Visualisation with the Tidyverse
Part 2

Dr Lisa Martin

Department of Information Science
Stellenbosch University

Today's Reading



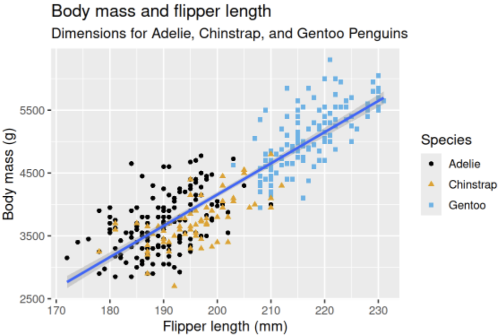
R for Data Science, Wholegame, Visualisation

Side Notes

- **Dark Mode:** Tools ▷ Global Options ▷ Appearance ▷ Editor theme
- **Shortcuts:** Tools ▷ Keyboard Shortcuts Help (Alt+Shift+K)

Where we left off...

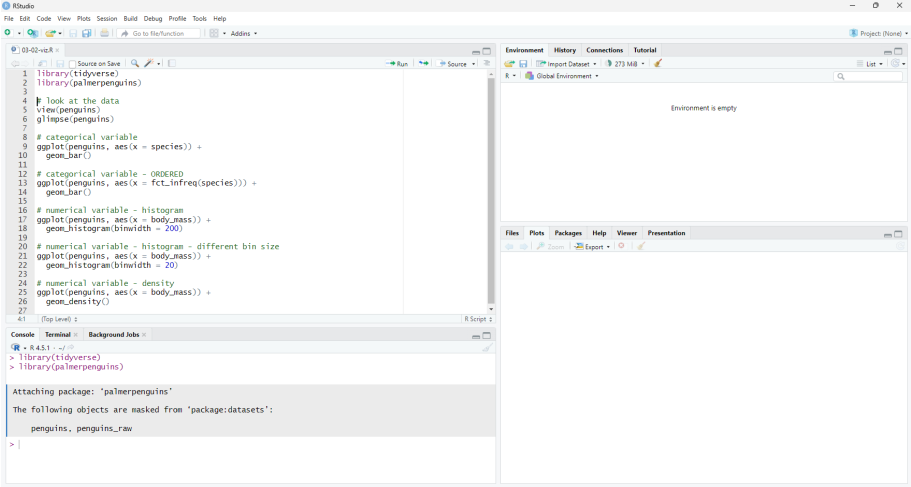
- Scatterplot with palmerpenguins dataset
- ggplot layers
- Data, Aesthetics (aes) and Geometry (geom)



Visualising Distributions

- Understanding the distribution of a variable is important
- Helps to identify patterns, outliers, and the overall shape of the data
- Geom used depends on the type of variable
- Continuous variables: geom_histogram(), geom_density()
- Categorical variables: geom_bar()

Visualising Distributions

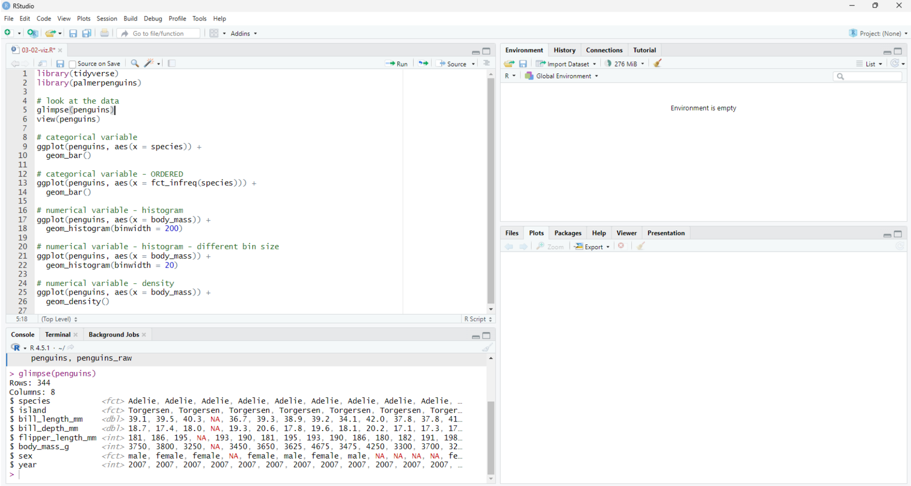


```r
library(tidyverse)
library(palmerpenguins)

# look at the data
view(penguins)
glimpse(penguins)

# categorical variable
ggplot(penguins, aes(x = species)) +
  geom_bar()

# categorical variable - ORDERED
ggplot(penguins, aes(x = fct_infreq(species))) +
  geom_bar()

# numerical variable - histogram
ggplot(penguins, aes(x = body_mass)) +
  geom_histogram(binwidth = 200)

# numerical variable - histogram - different bin size
ggplot(penguins, aes(x = body_mass)) +
  geom_histogram(binwidth = 20)

# numerical variable - density
ggplot(penguins, aes(x = body_mass)) +
  geom_density()

```

```
> library(tidyverse)
> library(palmerpenguins)

Attaching package: ‘palmerpenguins’

The following objects are masked from ‘package:datasets’:

    penguins, penguins_raw

> 
```

Visualising Distributions

- glimpse()



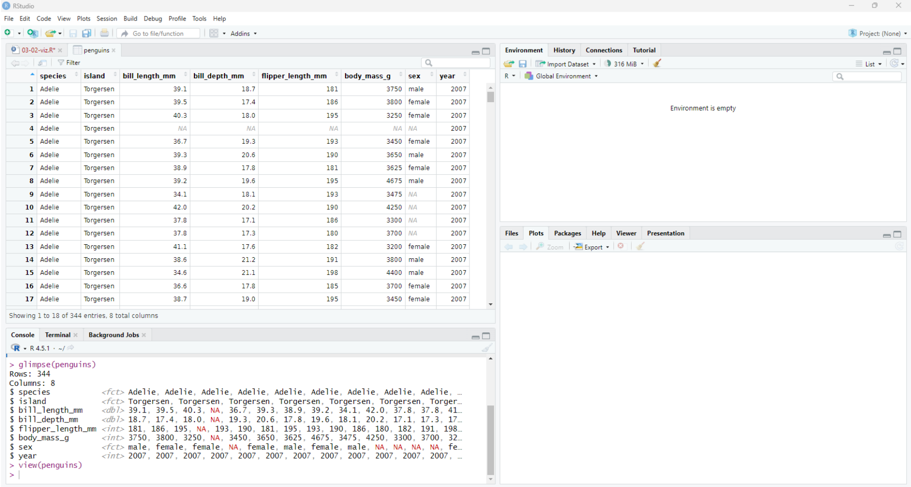
```r
library(tidyverse)
library(palmerpenguins)

# look at the data
glimpse(penguins)
view(penguins)

# categorical variable
ggplot(penguins, aes(x = species)) +
  geom_bar()

# categorical variable - ORDERED
ggplot(penguins, aes(x = fct_infreq(species))) +
  geom_bar()

# numerical variable - histogram
ggplot(penguins, aes(x = body_mass)) +
  geom_histogram(binwidth = 200)

# numerical variable - histogram - different bin size
ggplot(penguins, aes(x = body_mass)) +
  geom_histogram(binwidth = 20)

# numerical variable - density
ggplot(penguins, aes(x = body_mass)) +
  geom_density()

```

```
    penguins, penguins_raw

> glimpse(penguins)
Rows: 344
Columns: 8
$ species           <fct> Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, …
$ island            <fct> Torgersen, Torgersen, Torgersen, Torgersen, Torgersen, Torgersen, Torger…
$ bill_length_mm    <dbl> 39.1, 39.5, 40.3, NA, 36.7, 39.3, 38.9, 39.2, 34.1, 42.0, 37.8, 37.8, 41…
$ bill_depth_mm     <dbl> 18.7, 17.4, 18.0, NA, 19.3, 20.6, 17.8, 19.6, 18.1, 20.2, 17.1, 17.3, 17…
$ flipper_length_mm <int> 181, 186, 195, NA, 193, 190, 181, 195, 193, 190, 186, 180, 182, 191, 198…
$ body_mass_g       <int> 3750, 3800, 3250, NA, 3450, 3650, 3625, 4675, 3475, 4250, 3300, 3700, 32…
$ sex               <fct> male, female, female, NA, female, male, female, male, NA, NA, NA, NA, fe…
$ year              <int> 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, …
> 
```

Environment is empty

Visualising Distributions

- view()



	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	year
1	Adelie	Torgersen	39.1	18.7	181	3750	male	2007
2	Adelie	Torgersen	39.5	17.4	186	3800	female	2007
3	Adelie	Torgersen	40.3	18.0	195	3250	female	2007
4	Adelie	Torgersen	NA	NA	NA	NA	NA	2007
5	Adelie	Torgersen	36.7	19.3	193	3450	female	2007
6	Adelie	Torgersen	39.3	20.6	190	3650	male	2007
7	Adelie	Torgersen	38.9	17.8	181	3625	female	2007
8	Adelie	Torgersen	39.2	19.6	195	4675	male	2007
9	Adelie	Torgersen	34.1	18.1	193	3475	NA	2007
10	Adelie	Torgersen	42.0	20.2	190	4250	NA	2007
11	Adelie	Torgersen	37.8	17.1	186	3300	NA	2007
12	Adelie	Torgersen	37.8	17.3	180	3700	NA	2007
13	Adelie	Torgersen	41.1	17.6	182	3200	female	2007
14	Adelie	Torgersen	38.6	21.2	191	3800	male	2007
15	Adelie	Torgersen	34.6	21.1	198	4400	male	2007
16	Adelie	Torgersen	36.6	17.8	185	3700	female	2007
17	Adelie	Torgersen	38.7	19.0	195	3450	female	2007

Showing 1 to 18 of 344 entries, 8 total columns

Environment is empty

```
> glimpse(penguins)
Rows: 344
Columns: 8
$ species           <fct> Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, …
$ island            <fct> Torgersen, Torgersen, Torgersen, Torgersen, Torgersen, Torgersen, Torger…
$ bill_length_mm    <dbl> 39.1, 39.5, 40.3, NA, 36.7, 39.3, 38.9, 39.2, 34.1, 42.0, 37.8, 37.8, 41…
$ bill_depth_mm     <dbl> 18.7, 17.4, 18.0, NA, 19.3, 20.6, 17.8, 19.6, 18.1, 20.2, 17.1, 17.3, 17…
$ flipper_length_mm <int> 181, 186, 195, NA, 193, 190, 181, 195, 193, 190, 186, 180, 182, 191, 198…
$ body_mass_g       <int> 3750, 3800, 3250, NA, 3450, 3650, 3625, 4675, 3475, 4250, 3300, 3700, 32…
$ sex               <fct> male, female, female, NA, female, male, female, male, NA, NA, NA, NA, fe…
$ year              <int> 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, …
> view(penguins)
> 
```

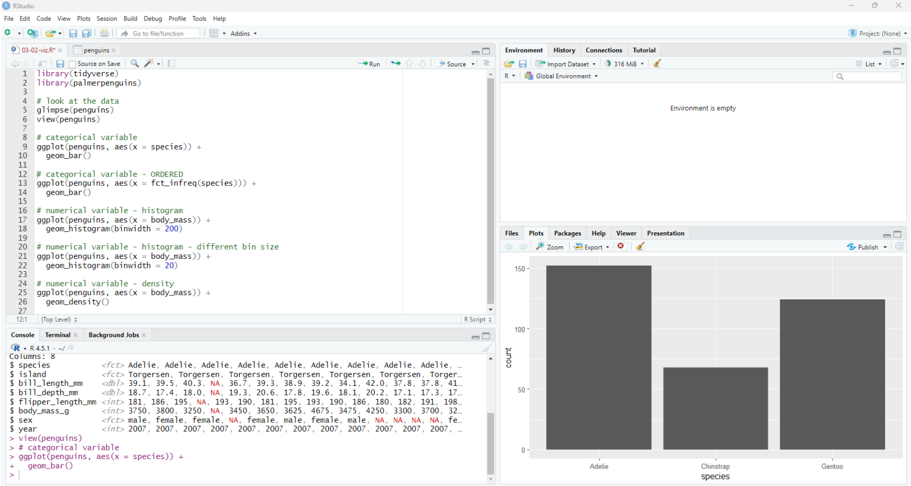
Visualising Distributions

Categorical Variables: geom_bar()

- Can only take one of a small set of values.
- The height of the bars displays how many observations occurred with each x value.
- geom_bar() uses the count of observations by default

Visualising Distributions

Categorical Variables: geom_bar()

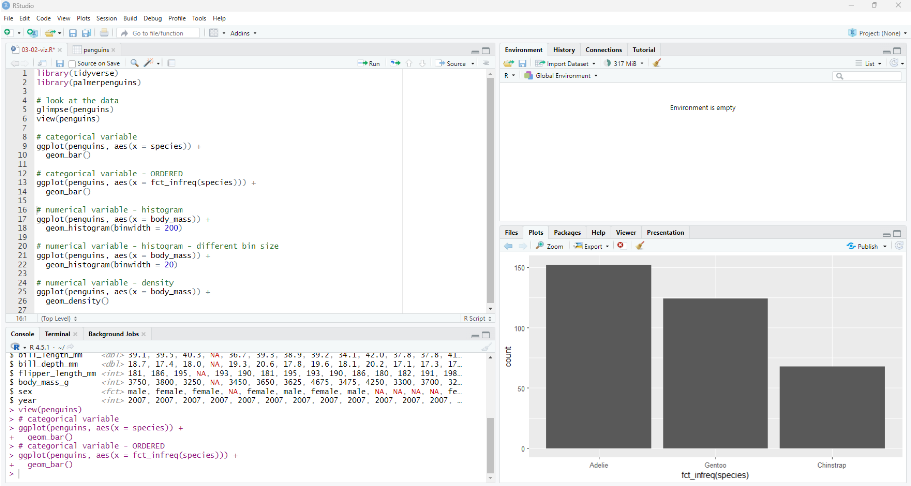


```r
library(tidyverse)
library(palmerpenguins)

# look at the data
glimpse(penguins)
view(penguins)

# categorical variable
ggplot(penguins, aes(x = species)) +
  geom_bar()

# categorical variable - ORDERED
ggplot(penguins, aes(x = fct_infreq(species))) +
  geom_bar()

# numerical variable - histogram
ggplot(penguins, aes(x = body_mass)) +
  geom_histogram(binwidth = 200)

# numerical variable - histogram - different bin size
ggplot(penguins, aes(x = body_mass)) +
  geom_histogram(binwidth = 20)

# numerical variable - density
ggplot(penguins, aes(x = body_mass)) +
  geom_density()

```

```
Columns: 8
$ species           <fct> Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, …
$ island            <fct> Torgersen, Torgersen, Torgersen, Torgersen, Torgersen, Torgersen, Torger…
$ bill_length_mm    <dbl> 39.1, 39.5, 40.3, NA, 36.7, 39.3, 38.9, 39.2, 34.1, 42.0, 37.8, 37.8, 41…
$ bill_depth_mm     <dbl> 18.7, 17.4, 18.0, NA, 19.3, 20.6, 17.8, 19.6, 18.1, 20.2, 17.1, 17.3, 17…
$ flipper_length_mm <int> 181, 186, 195, NA, 193, 190, 181, 195, 193, 190, 186, 180, 182, 191, 198…
$ body_mass_g       <int> 3750, 3800, 3250, NA, 3450, 3650, 3625, 4675, 3475, 4250, 3300, 3700, 32…
$ sex               <fct> male, female, female, NA, female, male, female, male, NA, NA, NA, NA, fe…
$ year              <int> 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, …
> view(penguins)
> # categorical variable
> ggplot(penguins, aes(x = species)) +
+   geom_bar()
> 
```

Visualising Distributions

Categorical Variables: geom_bar() - ordered

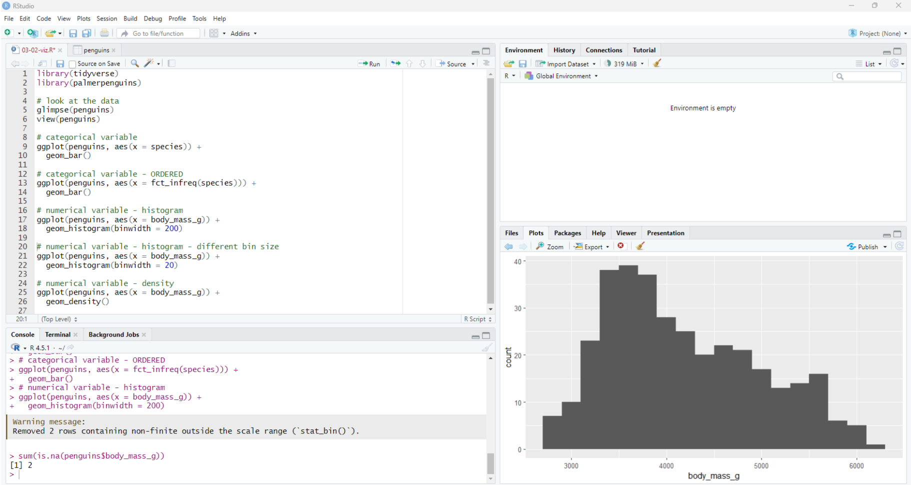


```r
library(tidyverse)
library(palmerpenguins)

# look at the data
glimpse(penguins)
view(penguins)

# categorical variable
ggplot(penguins, aes(x = species)) +
  geom_bar()

# categorical variable - ORDERED
ggplot(penguins, aes(x = fct_infreq(species))) +
  geom_bar()

# numerical variable - histogram
ggplot(penguins, aes(x = body_mass)) +
  geom_histogram(binwidth = 200)

# numerical variable - histogram - different bin size
ggplot(penguins, aes(x = body_mass)) +
  geom_histogram(binwidth = 20)

# numerical variable - density
ggplot(penguins, aes(x = body_mass)) +
  geom_density()

```

```
$ bill_length_mm    <dbl> 39.1, 39.5, 40.3, NA, 36.7, 39.3, 38.9, 39.2, 34.1, 42.0, 37.8, 37.8, 41…
$ bill_depth_mm     <dbl> 18.7, 17.4, 18.0, NA, 19.3, 20.6, 17.8, 19.6, 18.1, 20.2, 17.1, 17.3, 17…
$ flipper_length_mm <int> 181, 186, 195, NA, 193, 190, 181, 195, 193, 190, 186, 180, 182, 191, 198…
$ body_mass_g       <int> 3750, 3800, 3250, NA, 3450, 3650, 3625, 4675, 3475, 4250, 3300, 3700, 32…
$ sex               <fct> male, female, female, NA, female, male, female, male, NA, NA, NA, NA, fe…
$ year              <int> 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, …
> view(penguins)
> # categorical variable
> ggplot(penguins, aes(x = species)) +
+   geom_bar()
> # categorical variable - ORDERED
> ggplot(penguins, aes(x = fct_infreq(species))) +
+   geom_bar()
> 
```

Visualising Distributions

Numerical Variables: geom_histogram()



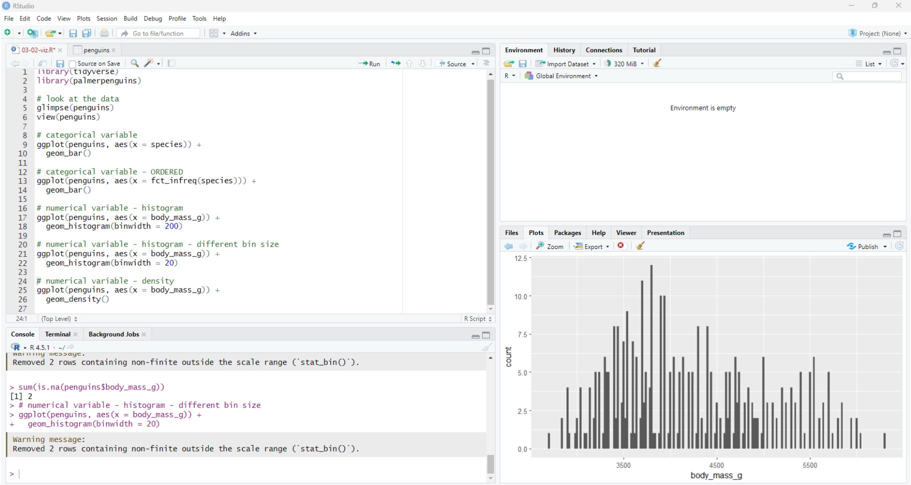
```r
library(tidyverse)
library(palmerpenguins)

# look at the data
glimpse(penguins)
view(penguins)

# categorical variable
ggplot(penguins, aes(x = species)) +
  geom_bar()

# categorical variable - ORDERED
ggplot(penguins, aes(x = fct_infreq(species))) +
  geom_bar()

# numerical variable - histogram
ggplot(penguins, aes(x = body_mass_g)) +
  geom_histogram(binwidth = 200)

# numerical variable - histogram - different bin size
ggplot(penguins, aes(x = body_mass_g)) +
  geom_histogram(binwidth = 20)

# numerical variable - density
ggplot(penguins, aes(x = body_mass_g)) +
  geom_density()

```

```
> # categorical variable - ORDERED
> ggplot(penguins, aes(x = fct_infreq(species))) +
+   geom_bar()
> # numerical variable - histogram
> ggplot(penguins, aes(x = body_mass_g)) +
+   geom_histogram(binwidth = 200)
Warning message:
Removed 2 rows containing non-finite outside the scale range (`stat_bin()`).

> sum(is.na(penguins$body_mass_g))
[1] 2
>
```

Note: Warning about NAs

Visualising Distributions

Numerical Variables: geom_histogram() - bin size



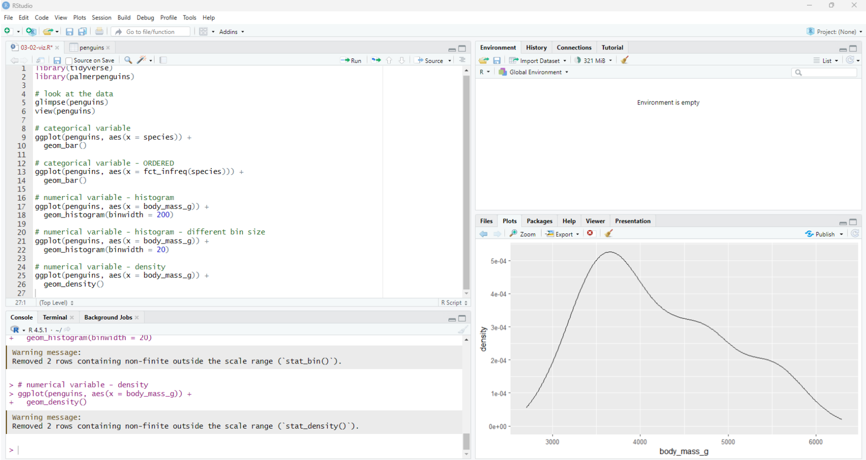
```r
library(tidyverse)
library(palmerpenguins)

# look at the data
glimpse(penguins)
view(penguins)

# categorical variable
ggplot(penguins, aes(x = species)) +
  geom_bar()

# categorical variable - ORDERED
ggplot(penguins, aes(x = fct_infreq(species))) +
  geom_bar()

# numerical variable - histogram
ggplot(penguins, aes(x = body_mass_g)) +
  geom_histogram(binwidth = 200)

# numerical variable - histogram - different bin size
ggplot(penguins, aes(x = body_mass_g)) +
  geom_histogram(binwidth = 20)

# numerical variable - density
ggplot(penguins, aes(x = body_mass_g)) +
  geom_density()

```

```
Warning message:
Removed 2 rows containing non-finite outside the scale range (`stat_bin()`).

> sum(is.na(penguins$body_mass_g))
[1] 2
> # numerical variable - histogram - different bin size
> ggplot(penguins, aes(x = body_mass_g)) +
+   geom_histogram(binwidth = 20)

Warning message:
Removed 2 rows containing non-finite outside the scale range (`stat_bin()`).

> |
```

Visualising Distributions

Numerical Variables: geom_density()

- Smoothed version of a histogram
- Good for continuous data



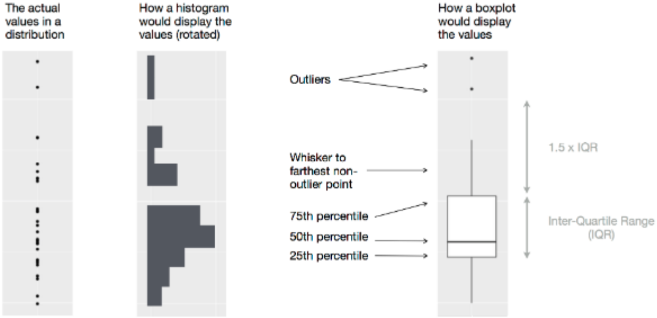
```r
library(tidyverse)
library(palmerpenguins)

# look at the data
glimpse(penguins)
view(penguins)

# categorical variable
ggplot(penguins, aes(x = species)) +
  geom_bar()

# categorical variable - ORDERED
ggplot(penguins, aes(x = fct_infreq(species))) +
  geom_bar()

# numerical variable - histogram
ggplot(penguins, aes(x = body_mass_g)) +
  geom_histogram(binwidth = 200)

# numerical variable - histogram - different bin size
ggplot(penguins, aes(x = body_mass_g)) +
  geom_histogram(binwidth = 20)

# numerical variable - density
ggplot(penguins, aes(x = body_mass_g)) +
  geom_density()
```

```
+   geom_histogram(binwidth = 20)
Warning message:
Removed 2 rows containing non-finite outside the scale range (`stat_bin()`).

> # numerical variable - density
> ggplot(penguins, aes(x = body_mass_g)) +
+   geom_density()
Warning message:
Removed 2 rows containing non-finite outside the scale range (`stat_density()`).
> 
```

Visualising Relationships

Numerical and Categorical: geom_boxplot()

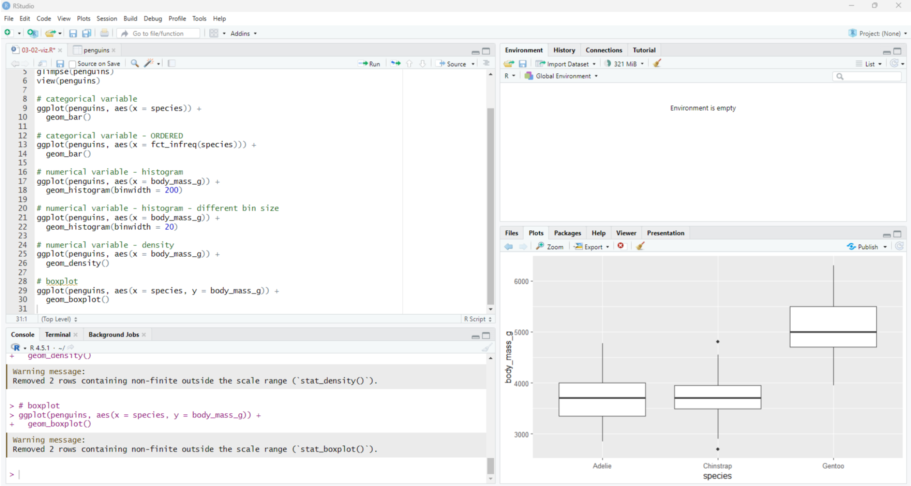
- Visualise the distribution of a continuous variable across categories
- Displays the median, quartiles, and potential outliers
- Spread of the distribution - symmetric or skewed to one side


The actual values in a distribution
How a histogram would display the values (rotated)
How a boxplot would display the values
Outliers
Whisker to farthest non-outlier point
75th percentile
50th percentile
25th percentile
1.5 x IQR
Inter-Quartile Range (IQR)

R4DS, Figure 1.1

Visualising Relationships

Numerical and Categorical: geom_boxplot()

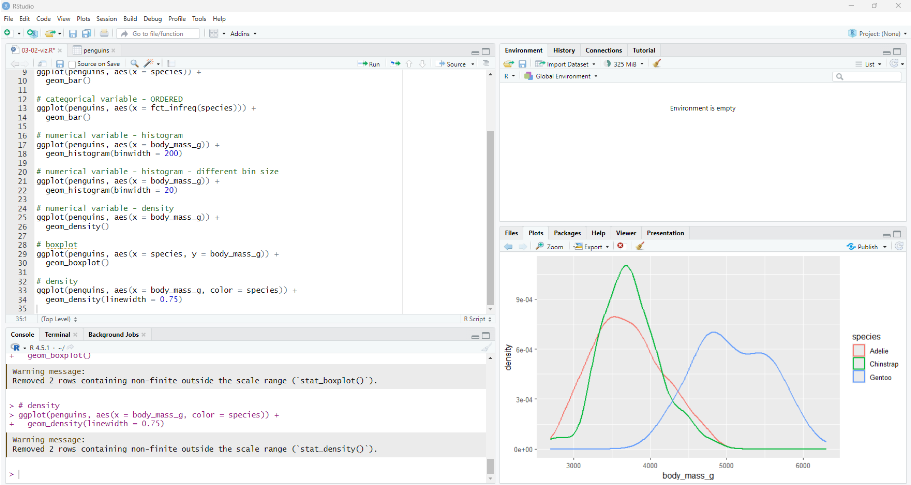


```r
glimpse(penguins)
view(penguins)

# categorical variable
ggplot(penguins, aes(x = species)) +
  geom_bar()

# categorical variable - ORDERED
ggplot(penguins, aes(x = fct_infreq(species))) +
  geom_bar()

# numerical variable - histogram
ggplot(penguins, aes(x = body_mass_g)) +
  geom_histogram(binwidth = 200)

# numerical variable - histogram - different bin size
ggplot(penguins, aes(x = body_mass_g)) +
  geom_histogram(binwidth = 20)

# numerical variable - density
ggplot(penguins, aes(x = body_mass_g)) +
  geom_density()

# boxplot
ggplot(penguins, aes(x = species, y = body_mass_g)) +
  geom_boxplot()
```

```
+   geom_density()
Warning message:
Removed 2 rows containing non-finite outside the scale range (`stat_density()`).

> # boxplot
> ggplot(penguins, aes(x = species, y = body_mass_g)) +
+   geom_boxplot()
Warning message:
Removed 2 rows containing non-finite outside the scale range (`stat_boxplot()`).

> 
```

Visualising Relationships

Numerical and Categorical: geom_density()



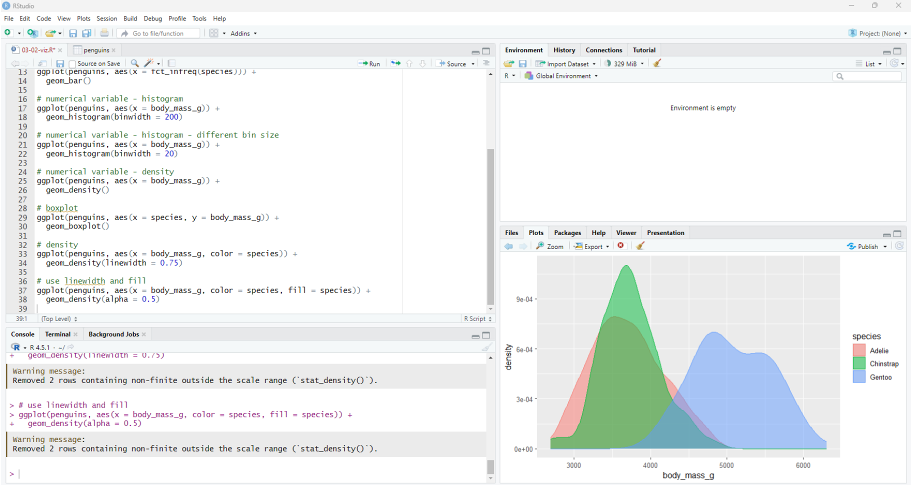
```
ggplot(penguins, aes(x = species)) +
  geom_bar()

# categorical variable - ORDERED
ggplot(penguins, aes(x = fct_infreq(species))) +
  geom_bar()

# numerical variable - histogram
ggplot(penguins, aes(x = body_mass_g)) +
  geom_histogram(binwidth = 200)

# numerical variable - histogram - different bin size
ggplot(penguins, aes(x = body_mass_g)) +
  geom_histogram(binwidth = 20)

# numerical variable - density
ggplot(penguins, aes(x = body_mass_g)) +
  geom_density()

# boxplot
ggplot(penguins, aes(x = species, y = body_mass_g)) +
  geom_boxplot()

# density
ggplot(penguins, aes(x = body_mass_g, color = species)) +
  geom_density(linewidth = 0.75)

```

```
+   geom_boxplot()
Warning message:
Removed 2 rows containing non-finite outside the scale range (`stat_boxplot()`).

> # density
> ggplot(penguins, aes(x = body_mass_g, color = species)) +
+   geom_density(linewidth = 0.75)
Warning message:
Removed 2 rows containing non-finite outside the scale range (`stat_density()`).

> 
```

Note: linewidth to set the thickness of the line

Visualising Relationships

Numerical and Categorical: geom_density()



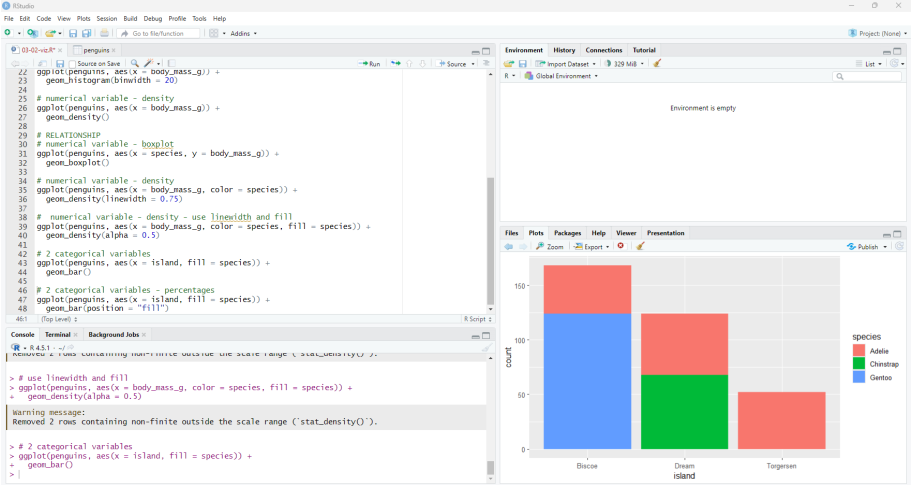
```r
ggplot(penguins, aes(x = fct_infreq(species))) +
  geom_bar()

# numerical variable - histogram
ggplot(penguins, aes(x = body_mass_g)) +
  geom_histogram(binwidth = 200)

# numerical variable - histogram - different bin size
ggplot(penguins, aes(x = body_mass_g)) +
  geom_histogram(binwidth = 20)

# numerical variable - density
ggplot(penguins, aes(x = body_mass_g)) +
  geom_density()

# boxplot
ggplot(penguins, aes(x = species, y = body_mass_g)) +
  geom_boxplot()

# density
ggplot(penguins, aes(x = body_mass_g, color = species)) +
  geom_density(linewidth = 0.75)

# use linewidth and fill
ggplot(penguins, aes(x = body_mass_g, color = species, fill = species)) +
  geom_density(alpha = 0.5)
```

```
+   geom_density(linewidth = 0.75)
Warning message:
Removed 2 rows containing non-finite outside the scale range (`stat_density()`).

> # use linewidth and fill
> ggplot(penguins, aes(x = body_mass_g, color = species, fill = species)) +
+   geom_density(alpha = 0.5)
Warning message:
Removed 2 rows containing non-finite outside the scale range (`stat_density()`).

> 
```

Note: linewidth and fill to set the thickness and colour of the area under the curve alpha to set the transparency of the fill

Visualising Relationships

Two Categorical: geom_bar()



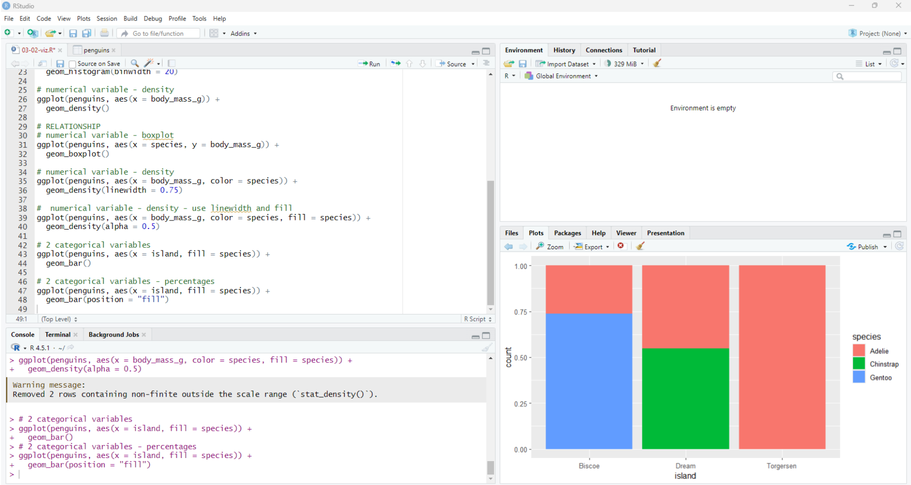
```
ggplot(penguins, aes(x = body_mass_g)) +
  geom_histogram(binwidth = 20)

# numerical variable - density
ggplot(penguins, aes(x = body_mass_g)) +
  geom_density()

# RELATIONSHIP
# numerical variable - boxplot
ggplot(penguins, aes(x = species, y = body_mass_g)) +
  geom_boxplot()

# numerical variable - density
ggplot(penguins, aes(x = body_mass_g, color = species)) +
  geom_density(linewidth = 0.75)

#  numerical variable - density - use linewidth and fill
ggplot(penguins, aes(x = body_mass_g, color = species, fill = species)) +
  geom_density(alpha = 0.5)

# 2 categorical variables
ggplot(penguins, aes(x = island, fill = species)) +
  geom_bar()

# 2 categorical variables - percentages
ggplot(penguins, aes(x = island, fill = species)) +
  geom_bar(position = "fill")
```

```
Removed 2 rows containing non-finite outside the scale range (`stat_density()`).

> # use linewidth and fill
> ggplot(penguins, aes(x = body_mass_g, color = species, fill = species)) +
+   geom_density(alpha = 0.5)
Warning message:
Removed 2 rows containing non-finite outside the scale range (`stat_density()`).

> # 2 categorical variables
> ggplot(penguins, aes(x = island, fill = species)) +
+   geom_bar()
> |
```

Environment is empty

Visualising Relationships

Two Categorical: geom_bar()



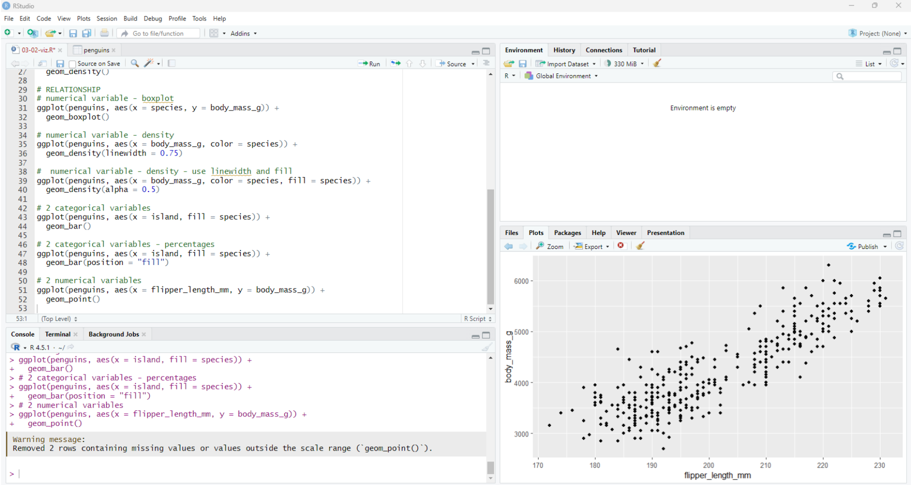
```
  geom_histogram(binwidth = 20)

# numerical variable - density
ggplot(penguins, aes(x = body_mass_g)) +
  geom_density()

# RELATIONSHIP
# numerical variable - boxplot
ggplot(penguins, aes(x = species, y = body_mass_g)) +
  geom_boxplot()

# numerical variable - density
ggplot(penguins, aes(x = body_mass_g, color = species)) +
  geom_density(linewidth = 0.75)

#  numerical variable - density - use linewidth and fill
ggplot(penguins, aes(x = body_mass_g, color = species, fill = species)) +
  geom_density(alpha = 0.5)

# 2 categorical variables
ggplot(penguins, aes(x = island, fill = species)) +
  geom_bar()

# 2 categorical variables - percentages
ggplot(penguins, aes(x = island, fill = species)) +
  geom_bar(position = "fill")
```

```
> ggplot(penguins, aes(x = body_mass_g, color = species, fill = species)) +
+   geom_density(alpha = 0.5)
Warning message:
Removed 2 rows containing non-finite outside the scale range (`stat_density()`).

> # 2 categorical variables
> ggplot(penguins, aes(x = island, fill = species)) +
+   geom_bar()
> # 2 categorical variables - percentages
> ggplot(penguins, aes(x = island, fill = species)) +
+   geom_bar(position = "fill")
> 
```

Note: Use of position = "fill" to show proportions instead of counts

Visualising Relationships

Two Numerical: geom_point()

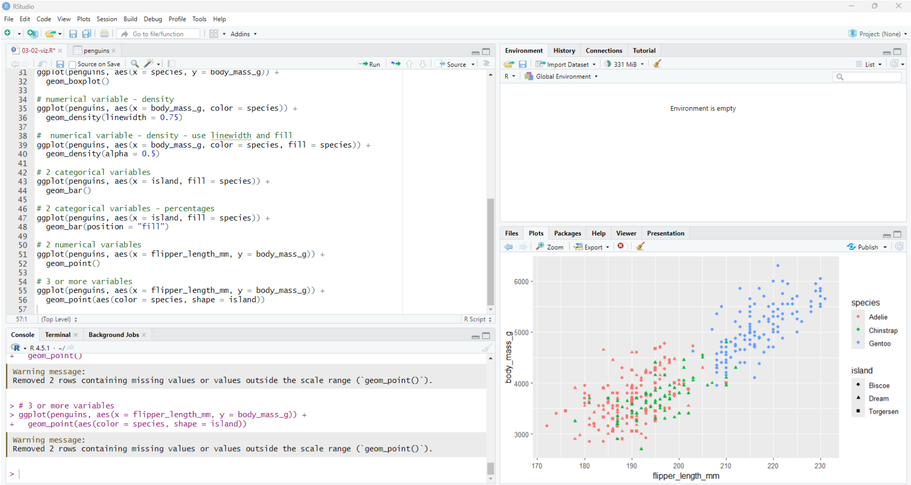


```
  geom_density()

# RELATIONSHIP
# numerical variable - boxplot
ggplot(penguins, aes(x = species, y = body_mass_g)) +
  geom_boxplot()

# numerical variable - density
ggplot(penguins, aes(x = body_mass_g, color = species)) +
  geom_density(linewidth = 0.75)

#  numerical variable - density - use linewidth and fill
ggplot(penguins, aes(x = body_mass_g, color = species, fill = species)) +
  geom_density(alpha = 0.5)

# 2 categorical variables
ggplot(penguins, aes(x = island, fill = species)) +
  geom_bar()

# 2 categorical variables - percentages
ggplot(penguins, aes(x = island, fill = species)) +
  geom_bar(position = "fill")

# 2 numerical variables
ggplot(penguins, aes(x = flipper_length_mm, y = body_mass_g)) +
  geom_point()

```

```
> ggplot(penguins, aes(x = island, fill = species)) +
+   geom_bar()
> # 2 categorical variables - percentages
> ggplot(penguins, aes(x = island, fill = species)) +
+   geom_bar(position = "fill")
> # 2 numerical variables
> ggplot(penguins, aes(x = flipper_length_mm, y = body_mass_g)) +
+   geom_point()

Warning message:
Removed 2 rows containing missing values or values outside the scale range (`geom_point()`).

> 
```

Visualising Relationships

Two or more variables: Additional aesthetics

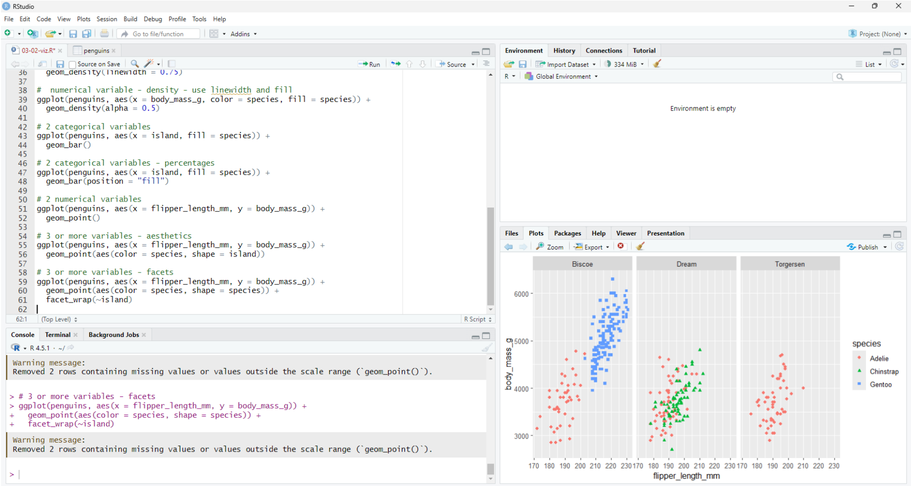


```r
ggplot(penguins, aes(x = species, y = body_mass_g)) +
  geom_boxplot()

# numerical variable - density
ggplot(penguins, aes(x = body_mass_g, color = species)) +
  geom_density(linewidth = 0.75)

#  numerical variable - density - use linewidth and fill
ggplot(penguins, aes(x = body_mass_g, color = species, fill = species)) +
  geom_density(alpha = 0.5)

# 2 categorical variables
ggplot(penguins, aes(x = island, fill = species)) +
  geom_bar()

# 2 categorical variables - percentages
ggplot(penguins, aes(x = island, fill = species)) +
  geom_bar(position = "fill")

# 2 numerical variables
ggplot(penguins, aes(x = flipper_length_mm, y = body_mass_g)) +
  geom_point()

# 3 or more variables
ggplot(penguins, aes(x = flipper_length_mm, y = body_mass_g)) +
  geom_point(aes(color = species, shape = island))
```

```
+   geom_point()
Warning message:
Removed 2 rows containing missing values or values outside the scale range (`geom_point()`).

> # 3 or more variables
> ggplot(penguins, aes(x = flipper_length_mm, y = body_mass_g)) +
+   geom_point(aes(color = species, shape = island))
Warning message:
Removed 2 rows containing missing values or values outside the scale range (`geom_point()`).
> 
```

Visualising Relationships

Two or more variables: Facets

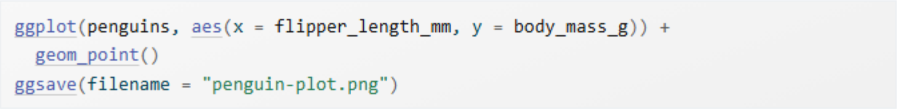


```
    geom_density(linewidth = 0.75)

# numerical variable - density - use linewidth and fill
ggplot(penguins, aes(x = body_mass_g, color = species, fill = species)) +
    geom_density(alpha = 0.5)

# 2 categorical variables
ggplot(penguins, aes(x = island, fill = species)) +
    geom_bar()

# 2 categorical variables - percentages
ggplot(penguins, aes(x = island, fill = species)) +
    geom_bar(position = "fill")

# 2 numerical variables
ggplot(penguins, aes(x = flipper_length_mm, y = body_mass_g)) +
    geom_point()

# 3 or more variables - aesthetics
ggplot(penguins, aes(x = flipper_length_mm, y = body_mass_g)) +
    geom_point(aes(color = species, shape = island))

# 3 or more variables - facets
ggplot(penguins, aes(x = flipper_length_mm, y = body_mass_g)) +
    geom_point(aes(color = species, shape = species)) +
    facet_wrap(~island)
```

```
Warning message:
Removed 2 rows containing missing values or values outside the scale range (`geom_point()`).

> # 3 or more variables - facets
> ggplot(penguins, aes(x = flipper_length_mm, y = body_mass_g)) +
+   geom_point(aes(color = species, shape = species)) +
+   facet_wrap(~island)

Warning message:
Removed 2 rows containing missing values or values outside the scale range (`geom_point()`).

> |
```

Save your plots

ggsave()



```r
ggplot(penguins, aes(x = flipper_length_mm, y = body_mass_g)) +
  geom_point()
ggsave(filename = "penguin-plot.png")
```