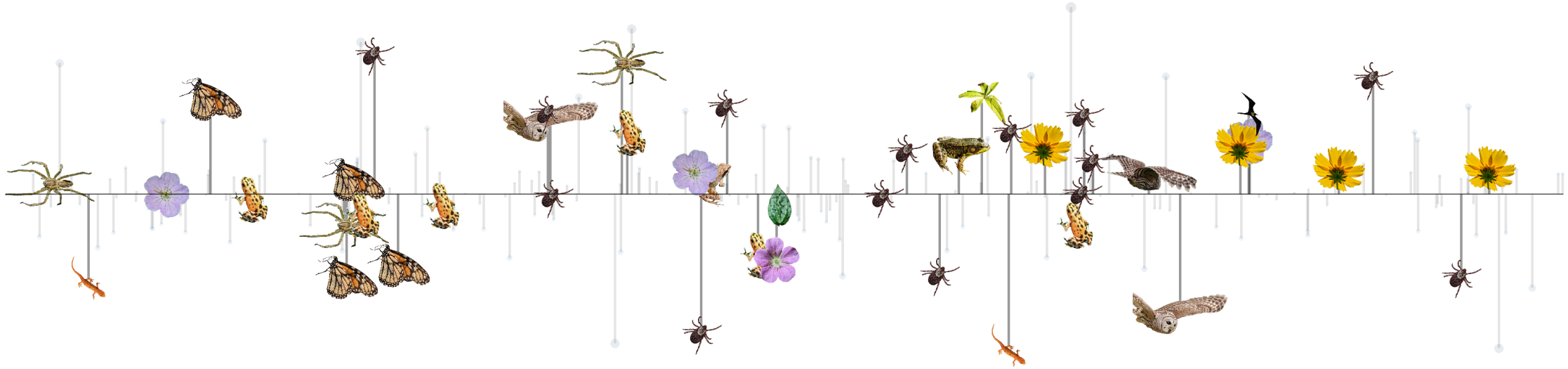
Intro to Quantitative Ecology
UMass Amherst – Michael France Nelson

Deck 5: Graphical Data Exploration



Announcements: Mac Users

- Always download csv data files by right-clicking and choosing ‘save link as’.
- Do not open the file in Numbers.
  - Numbers changes the file format and overwrites the original data without warning you.
  - It makes the file unreadable by R.

R tips of the day: nested function calls

You can use R functions inside other R functions.

- This is called a *nested* function call.

You've already used this syntax:

```
dat = read.csv(here("data", "mander.csv"))
```

In this example `here()` is nested within `read.csv()`

R tips of the day: multi-line commands

You can make long or complicated R calls easier to read by splitting them over multiple lines

```
dat =
  read.csv(
    here("data", "mander.csv"))
```

R tips of the day: indentation

You can also use indentation to make it clear which arguments go with which functions:

```
butterfly =
  as.matrix(
    read.csv(
      here(
        "data",
        "butterfly_table.csv"),
      check.names = FALSE,
      row.names = 1))
```

- `here()` is *nested* in `read.csv()`
- `check.names` and `row.names` are *arguments* to `read.csv()`

Graphical exploration

Two main reasons to use graphs:

1: Inform how to analyze the data

- visualization
- identify patterns
- choose appropriate statistical test

2: Presentation of the data

- summarize results
- communicate results
- publish results

Types of Plots

Exploratory Plots

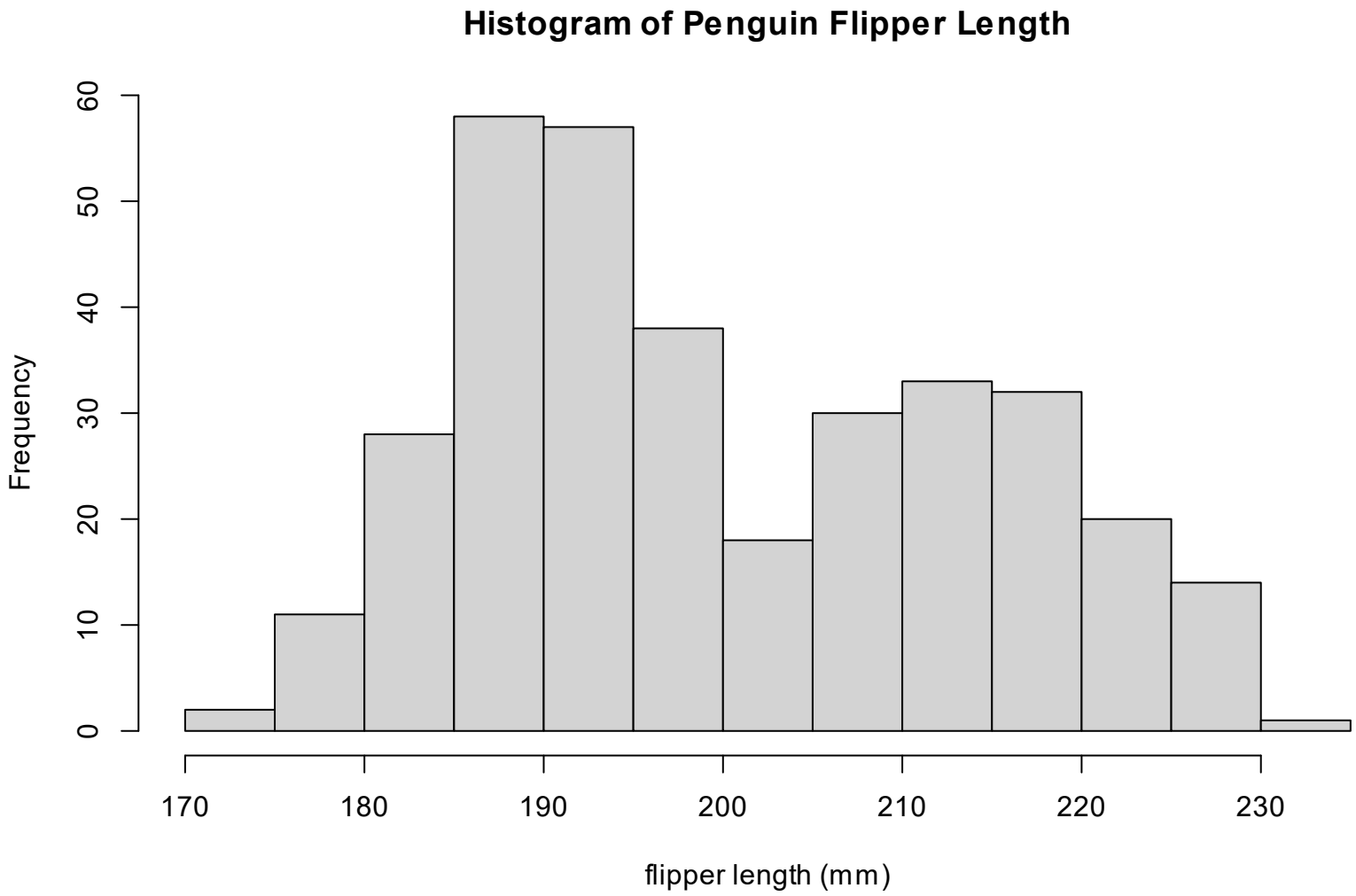
Exploratory graphs help understand the distribution of the data:

- are the data normally distributed?
- Normality is an important assumption in statistics
- Normality determines how data are analyzed
- what is the central tendency?
- what is the spread?
- general summaries of the data

Exploratory: *Histogram*

- Width of bars are defined data bins or intervals
- Height of bars represent bin-specific frequencies

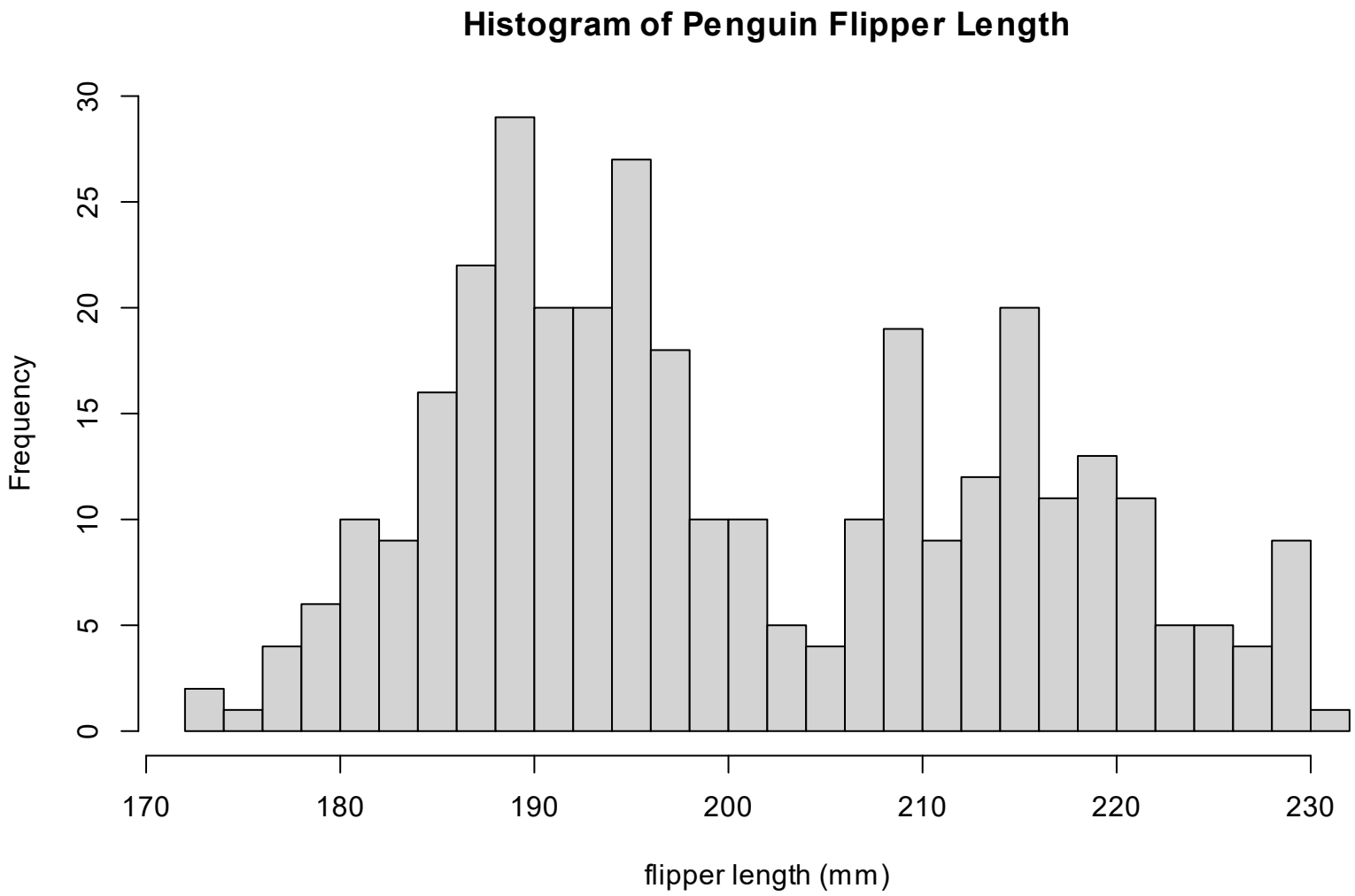
```
hist(
  penguins$flipper_length_mm,
  main = "Penguin Flipper Length",
  xlab = "flipper length (mm)")
```



Exploratory: *Histogram*

You can change the number and widths of the bins with the breaks argument in R:

```
hist(
  penguins$flipper_length_mm,
  main = "Penguin Flipper Length",
  xlab = "flipper length (mm)",
  breaks = 30)
```

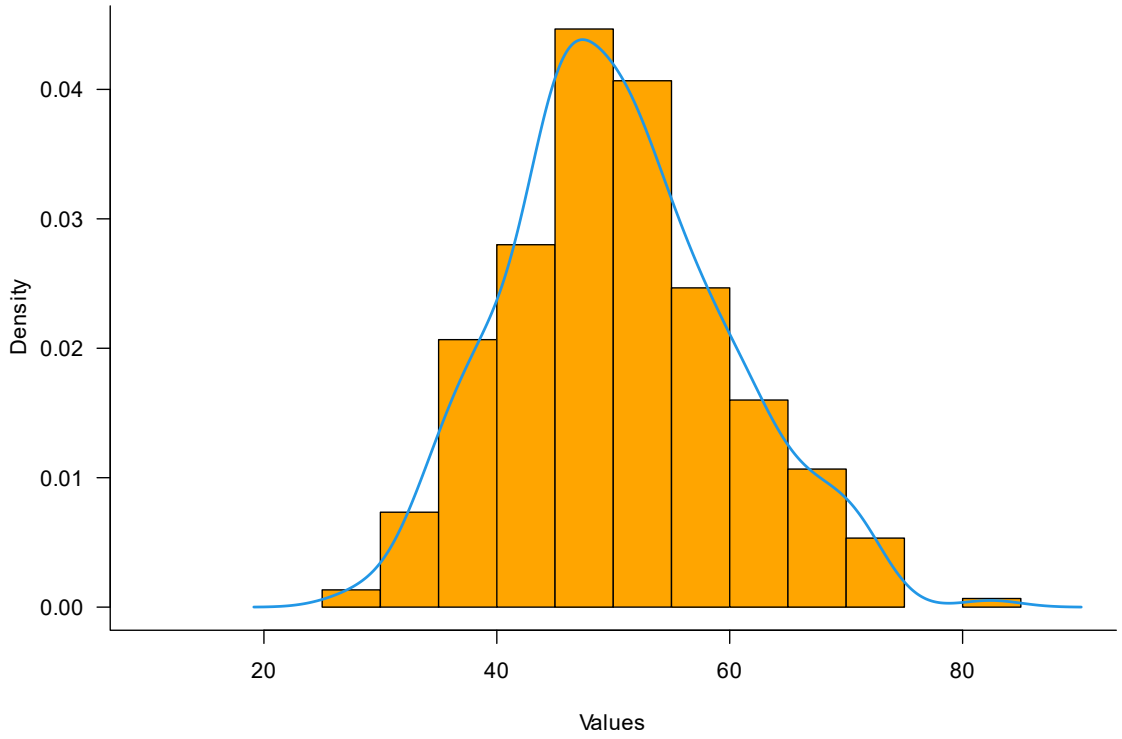


Exploratory: *Histogram + Density Plot*

A *density plot*: smoothed version of histogram

- To overlay on a histogram, tell `hist()` to plot the *probability* version of the histogram:

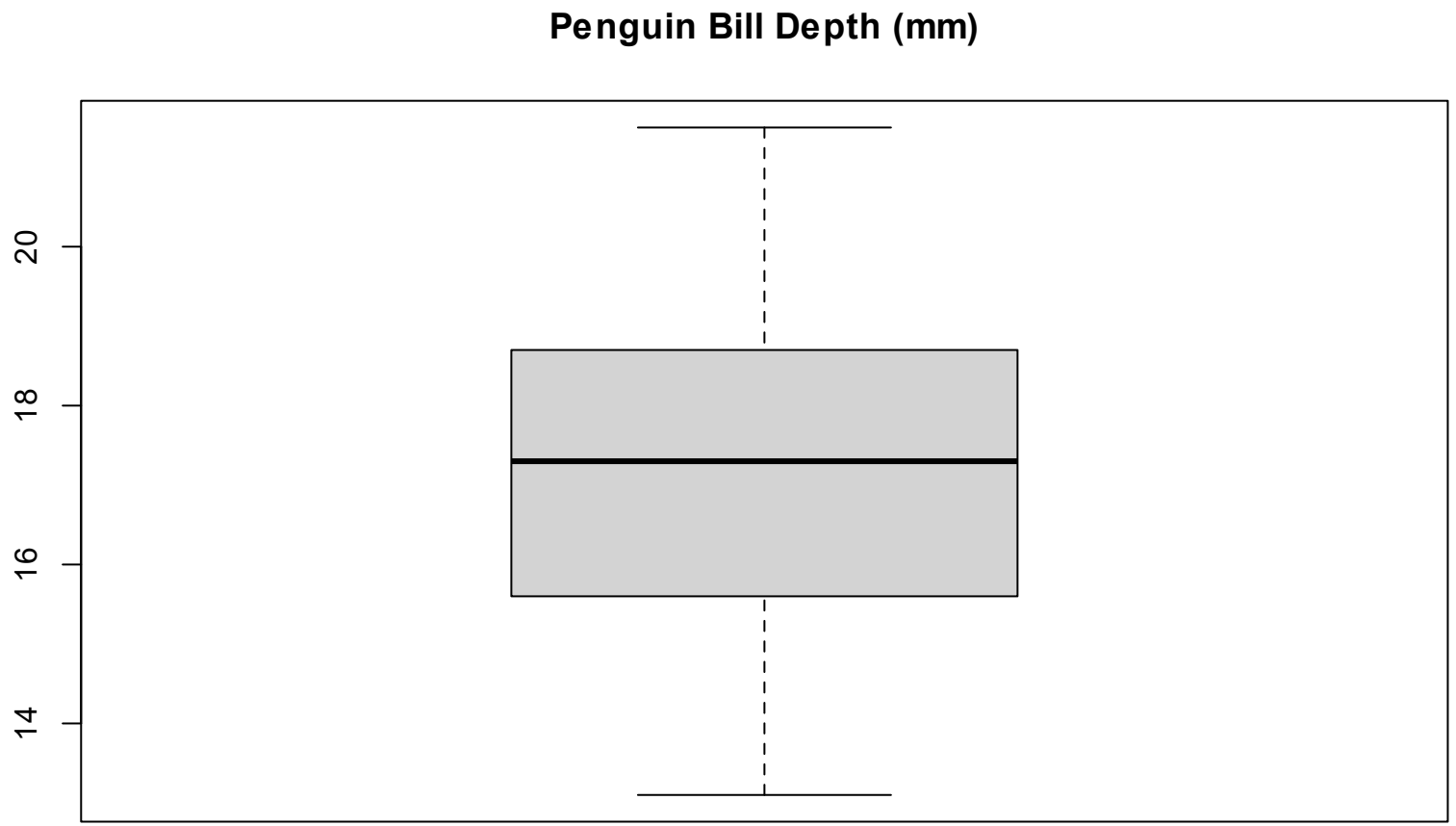
```
hist(values, probability = TRUE)
lines(density(values))
```



Exploratory: *Box-whisker/Box plot*

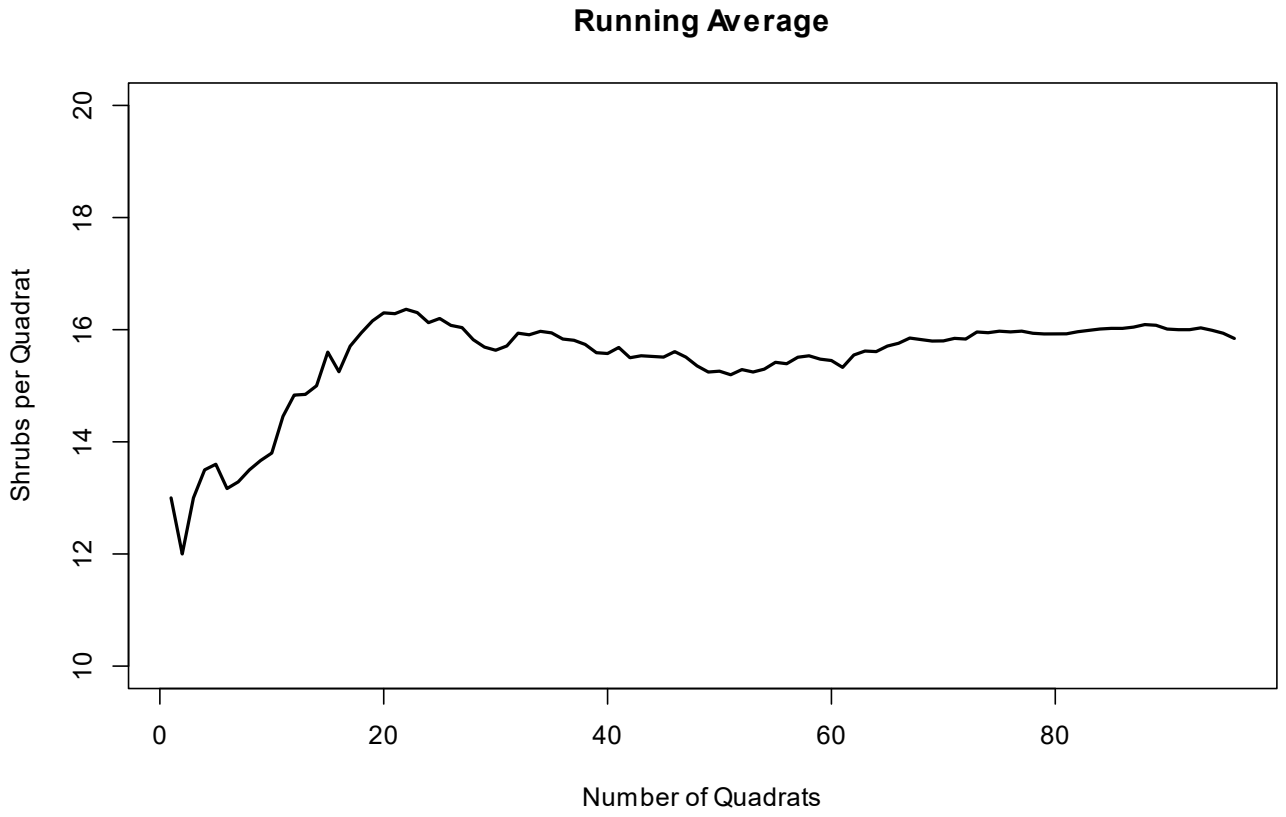
Boxplots summarize the spread (inter-quartile range), symmetry, and center (median).

```
boxplot(
  penguins$bill_depth_mm,
  main = "Penguin Bill
Depth (mm)")
```



Exploratory: *Line graph*

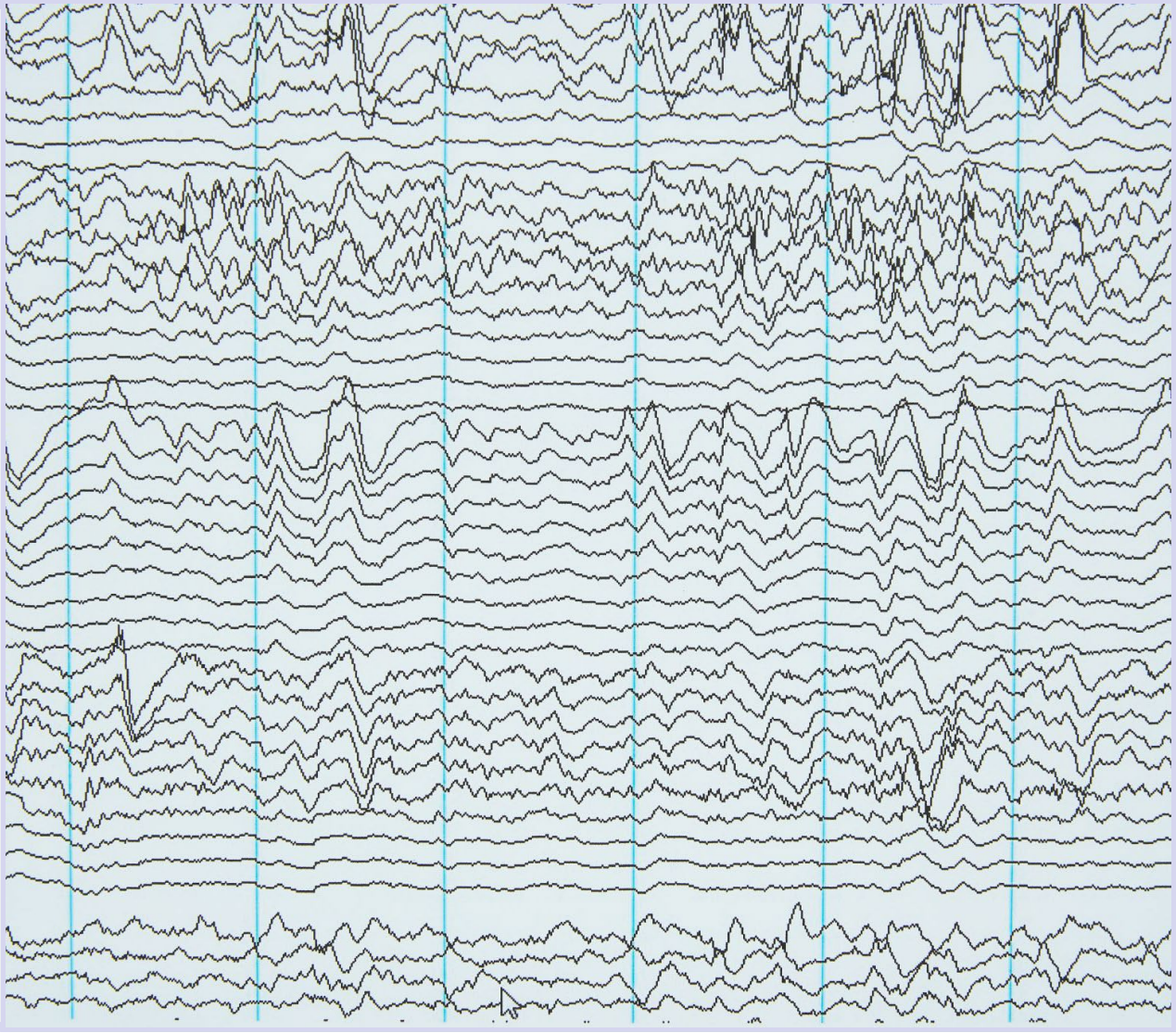
Line graph is a useful plot for running average or time series data



Differences

To visualize differences between groups

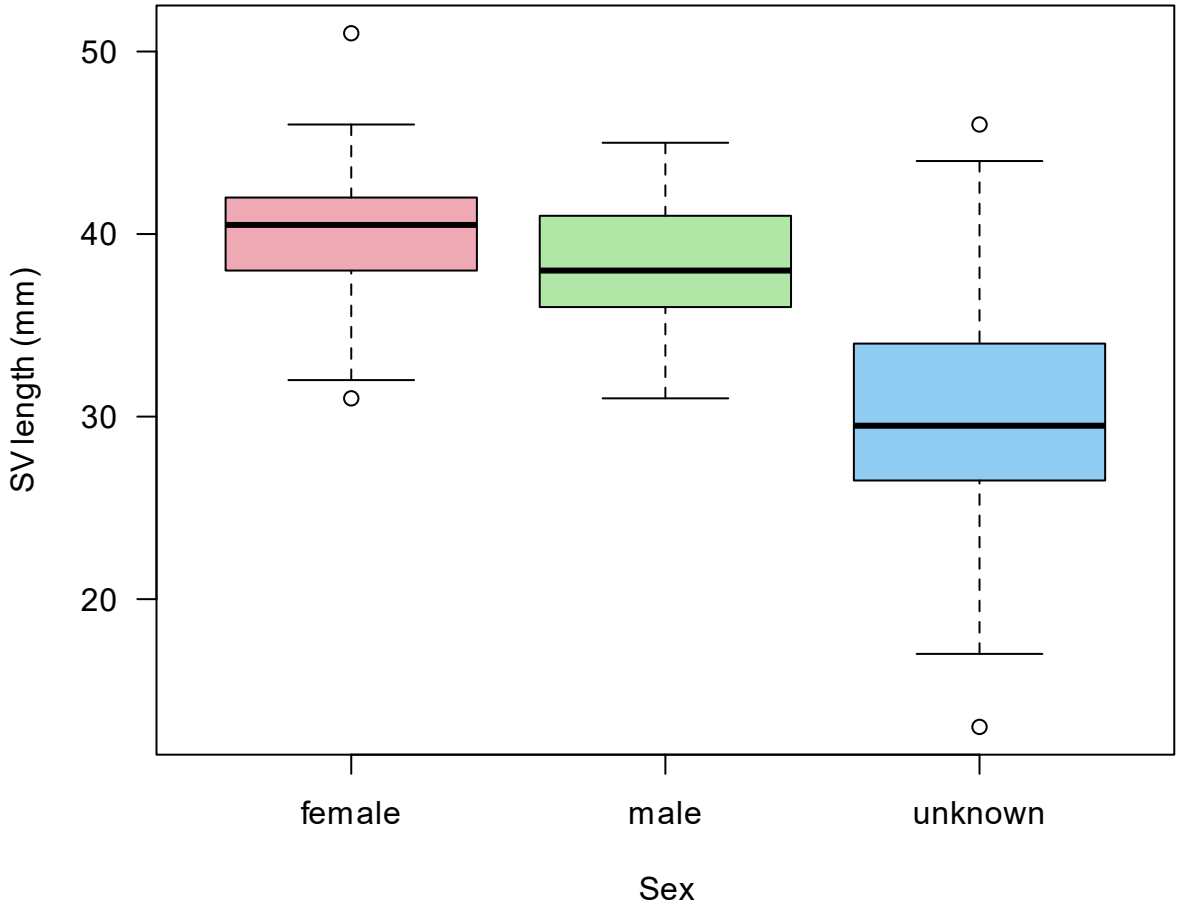
- box-whisker plots
  - compares averages
  - compares distribution
- bar charts
  - compares averages



Differences: Boxplots

Compare salamander snout-vent lengths by three sexes:

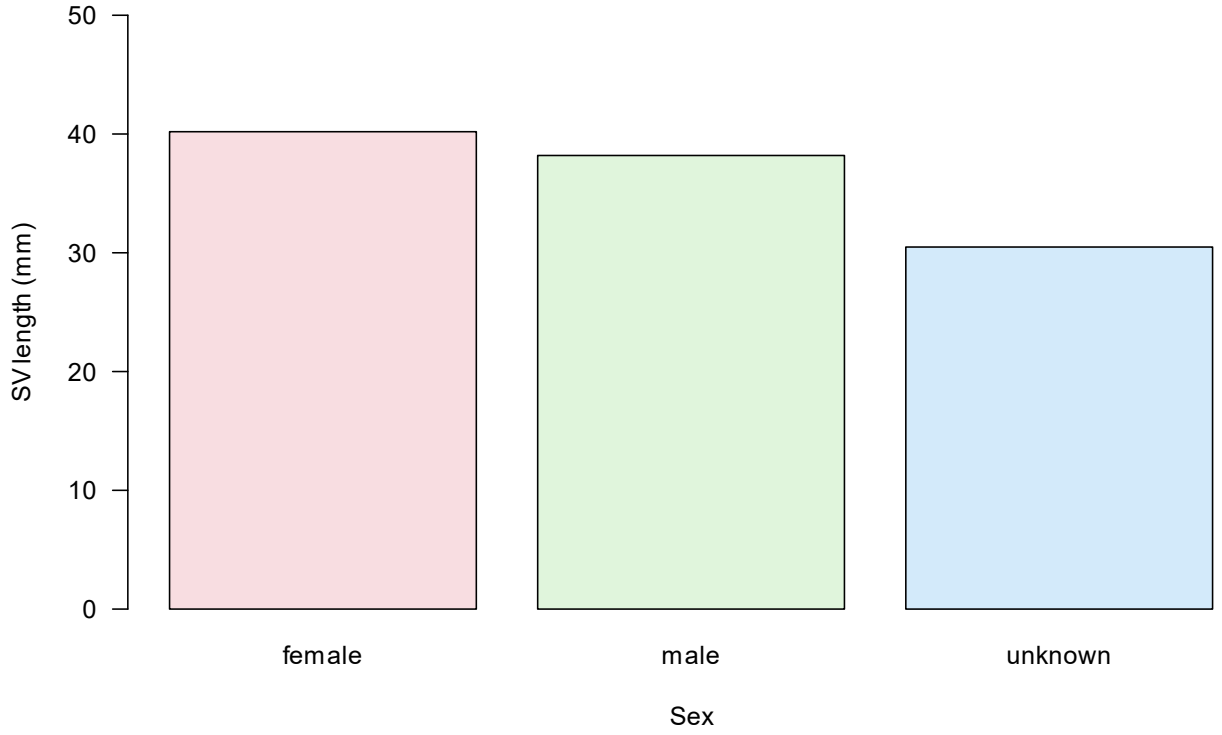
```
boxplot(
      mander$SVL ~ mander$Sex
#formula notation
```



Differences: *Bar chart*

Compare salamander snout-vent lengths by three sexes:

```
bars = tapply(mander$SVL,mander$Sex,mean)
#create matrix
barplot(bars)
# plot it
```



Links

Two main approaches for graphing relationships between data:

1: Correlations

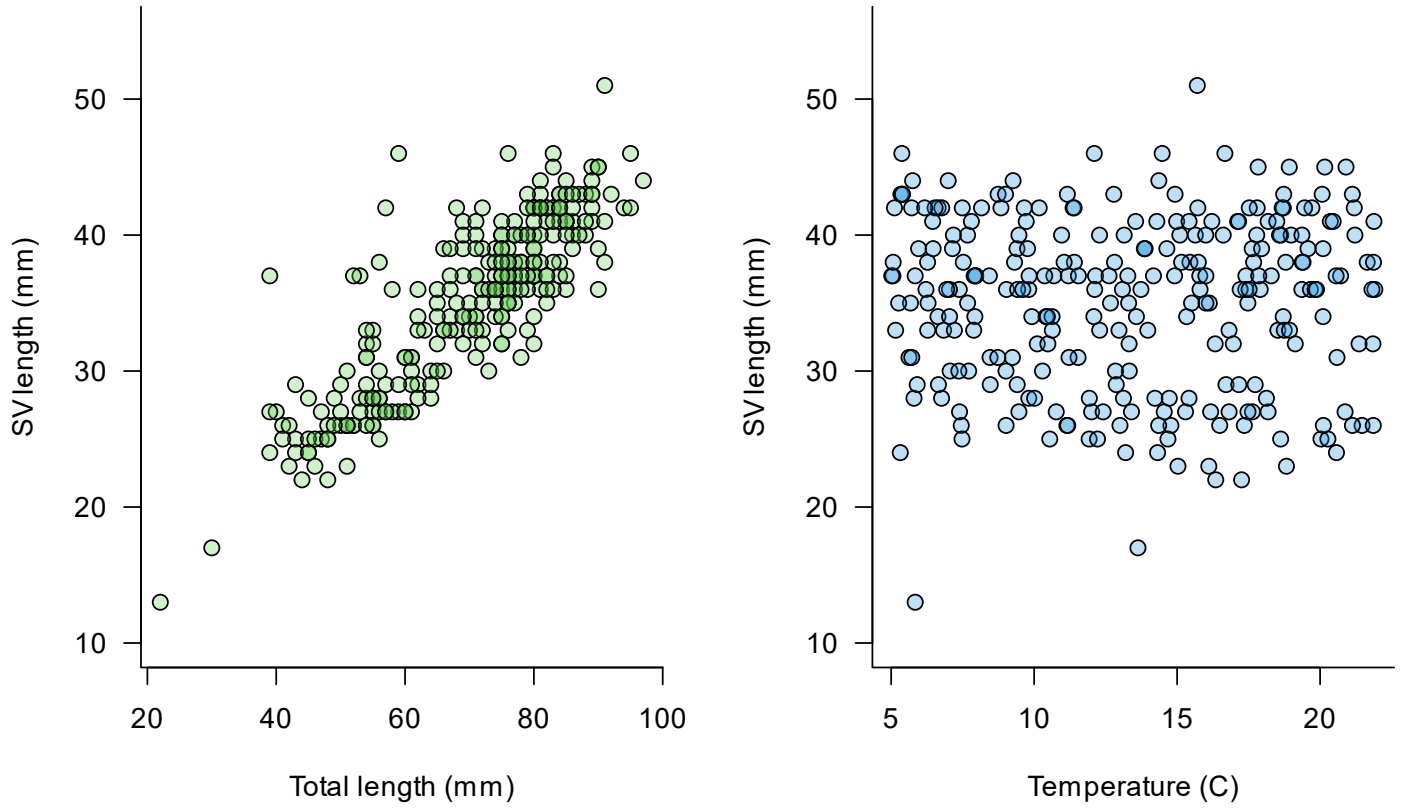
- two numeric variables
- dependent variable (of primary interest: y-axis)
- independent variable (explanatory variable: x-axis)

Types of graphs:

- *Scatterplots* can show how one variable is related to another

Links: *Scatter plot*

```
plot(x,y) # x and y are numeric vectors
```



Links

Two main approaches for graphing relationships between data:

2: Associations

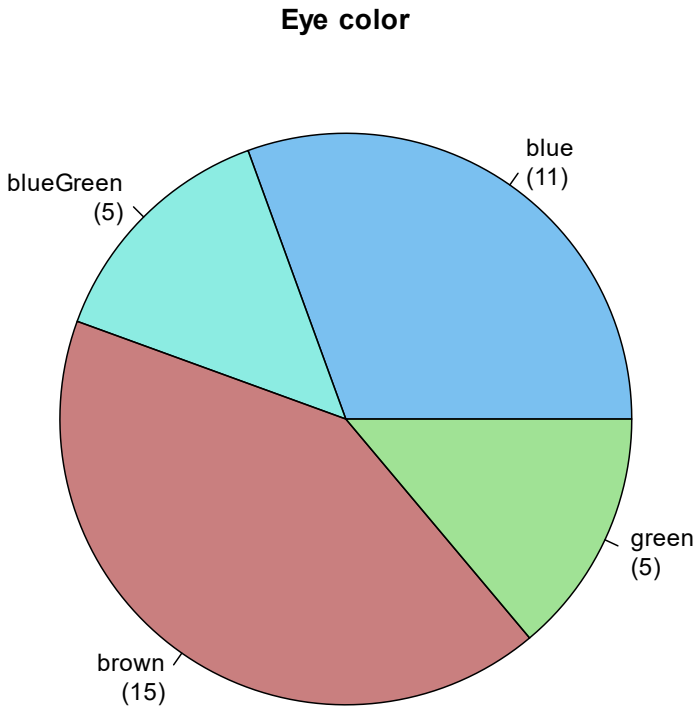
- categorical data
- summarize categories
- counts
- proportions
- by rows and/or columns of a table

Types of graphs:

1. *pie charts* for single categories
2. *bar graphs* for several categories

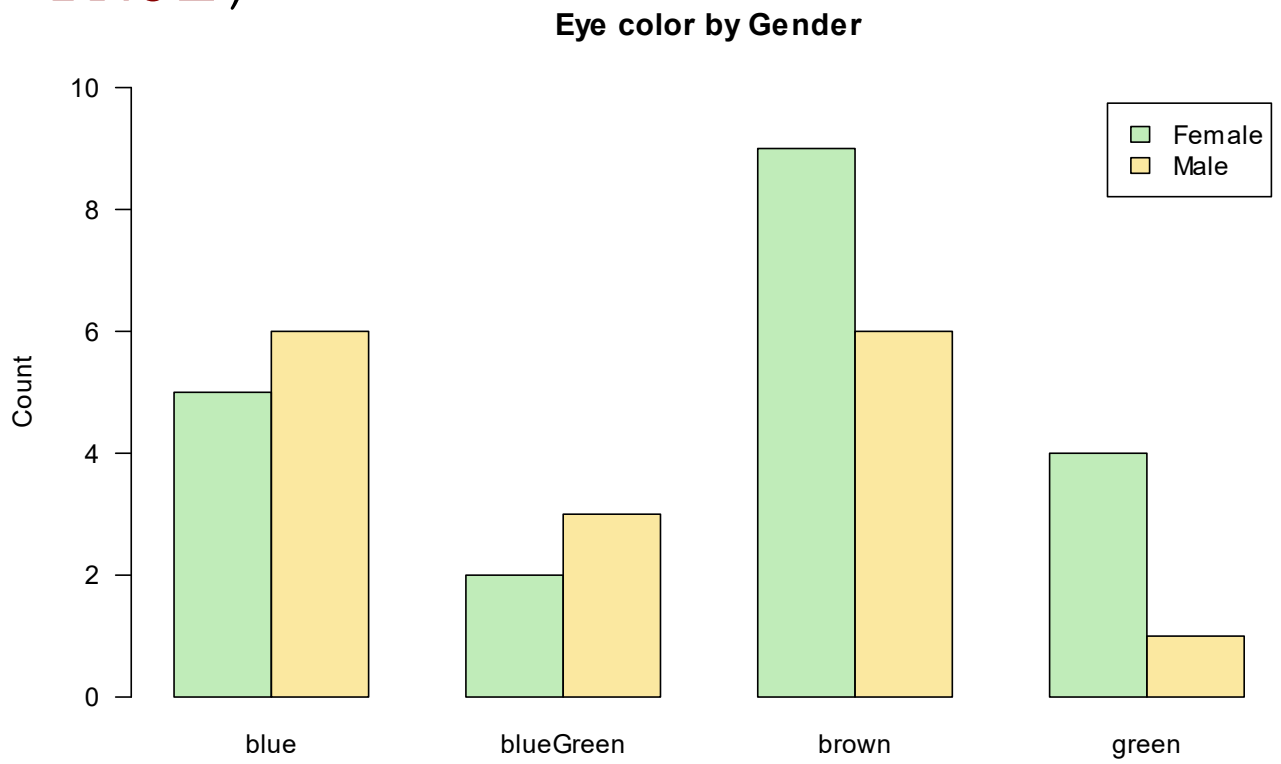
Links: *Pie chart*

```
pietab <- table(classData$Eyes)
pie(pietab) # number of people with each eye color
```



Links: *Bar chart*

```
bartab <- table(classData$Gender, classData$Eyes)
#(number of each gender with each eye color)
barplot(pietab, beside=TRUE)
```



Some graphics pointers

In summary, graphs are a useful data visualization tool

Some graphics pointers

In summary, graphs are a useful data visualization tool

- summarizing
- understanding
- describing
- presenting/communicating

Some graphics pointers: label your plots!

In summary, graphs are a useful data visualization tool

- summarizing
- understanding
- describing
- presenting/communicating

BUT we must label them well or they are useless!

- label both axes
- provide a main title for your graph
- avoid clutter
- make it readable
- *I expect graphs to be properly labeled from now on*!

Salamander Data Exploration

Let’s work on the first part in class!

R Markdown Practice!

Import the salamander data

Salamander Group Assignment

Your group will:

- Perform a numerical and graphical data exploration.
- Write a report on your methods.
- Let's work together to import the data!
  - You'll need the `mander.csv` data file.

Right now, before you form your groups, let's write code to import the data individually.

- You'll need 'mander.csv'
- Create a new R Markdown document (not an R notebook).

In-Class Histograms

Note to Mike: Enable assignment in Moodle!!!

Announcements

- In-class histograms: Moodle group settings were incorrect. You'll need to re-self-select and re-upload your document. Sorry, I know it's a pain!
- Potential point of confusion: there are two group assignments in a row:
  - Salamander Data Exploration (Due March 20th)
  - Salamander Tests for Differences (Due March 27th)

Signs of Spring [Break]

- Next week is spring break!
- Please review your gradebook on Moodle and contact us with any questions.
  - We’re human, sometimes we make grading mistakes!
  - The earlier you let us know, the earlier it is to make corrections!

Plot type summary

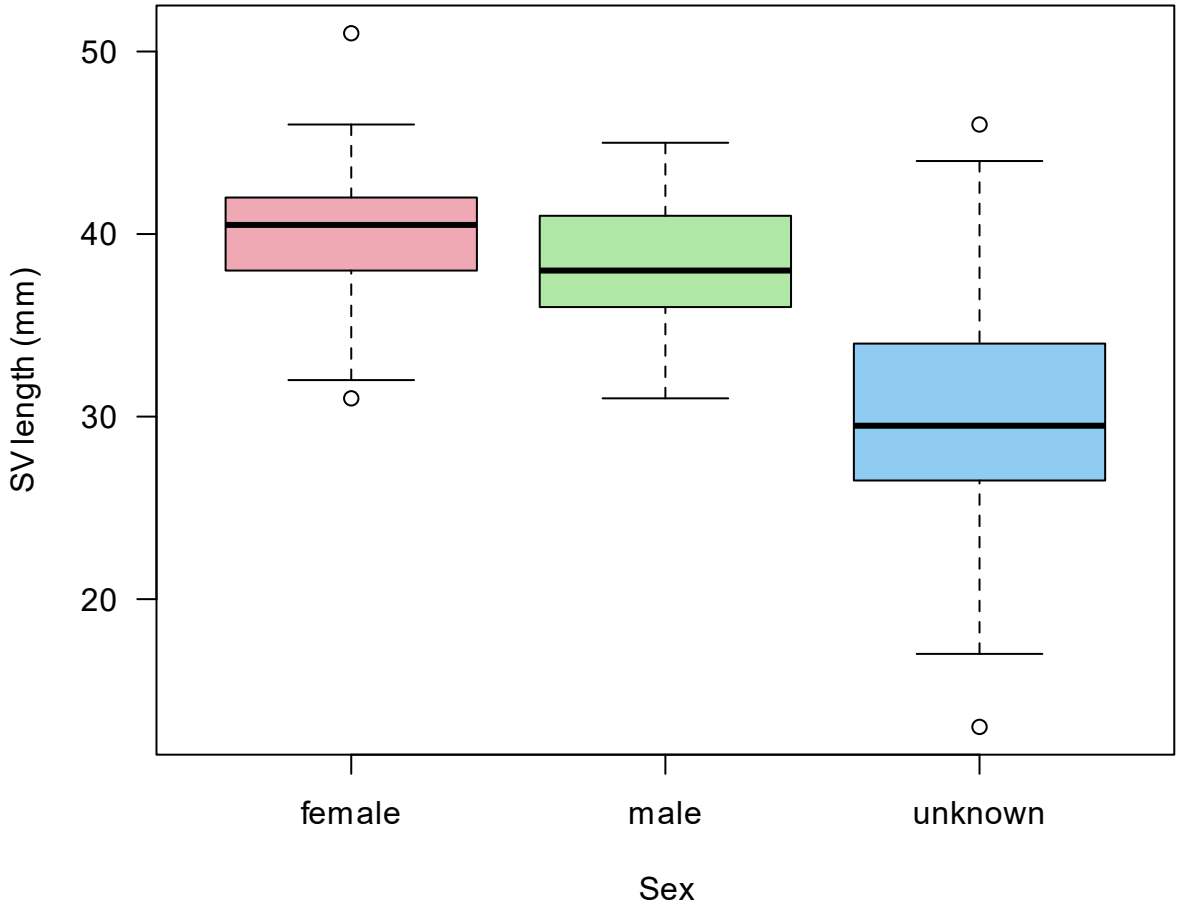
Purpose	Graph Type
Illustrating *distribution*	Histogram, Density plot, Boxplots(+whisker) plots
Illustrating *differences*	Bar chart, Box plot
Illustrating *correlations*	Scatter plot
Illustrating *associations*	Pie chart, Bar chart
Illustrating *sample size*	Line plot of running avg

Beyond graphs, Towards statistics

- Graphs are powerful tools that provide insight and understanding of the patterns and relationships in the data.
- Graphs alone don’t give us the complete answer. We need to **quantify** the relationships we see in our plots.

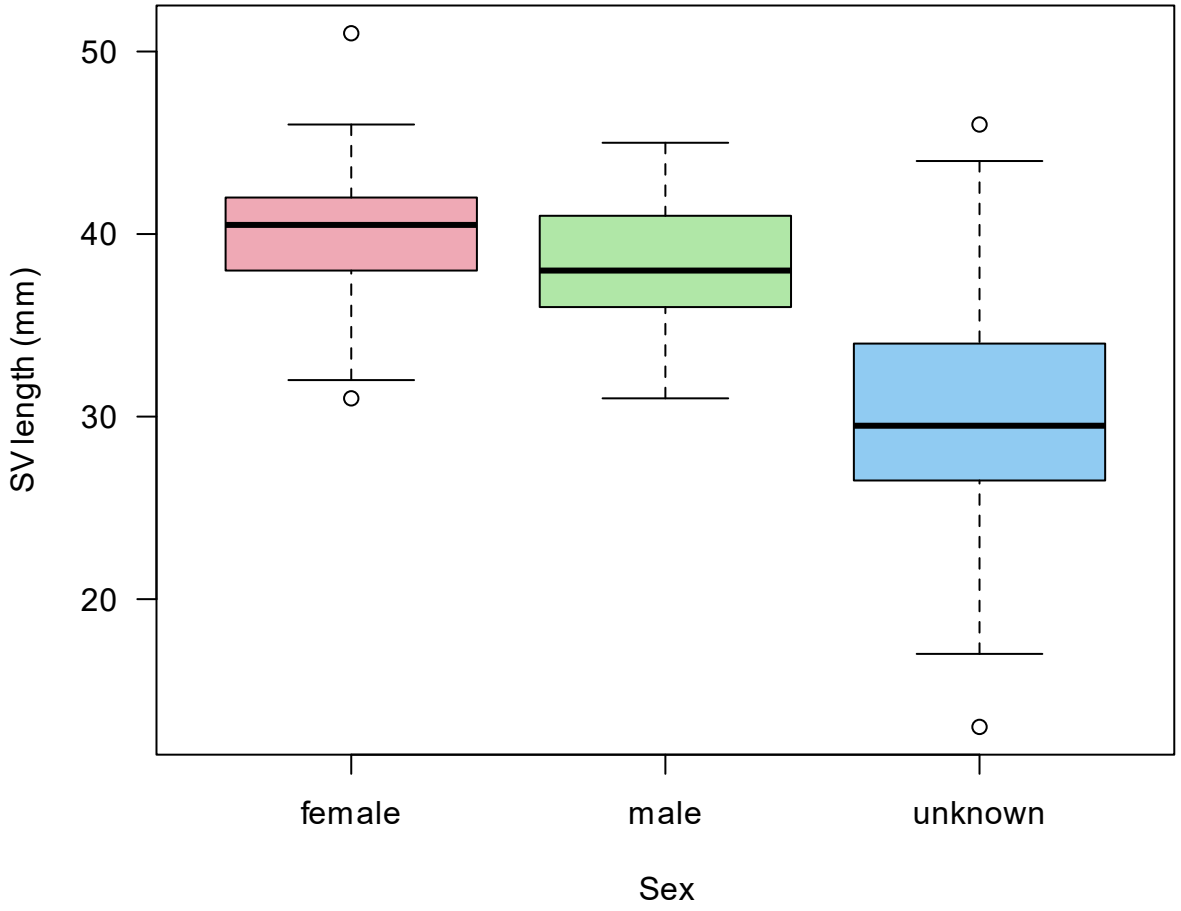
Beyond graphs, Towards statistics

- How can we **quantify** our evidence for relationships?
- Are differences between groups *significant*?



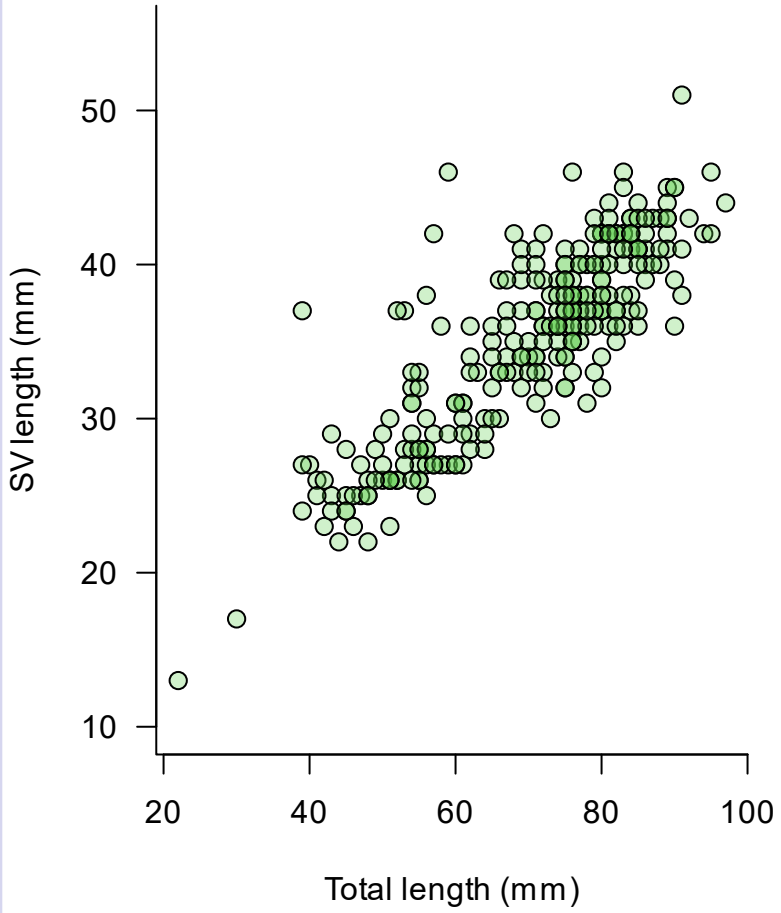
Beyond graphs, Towards statistics

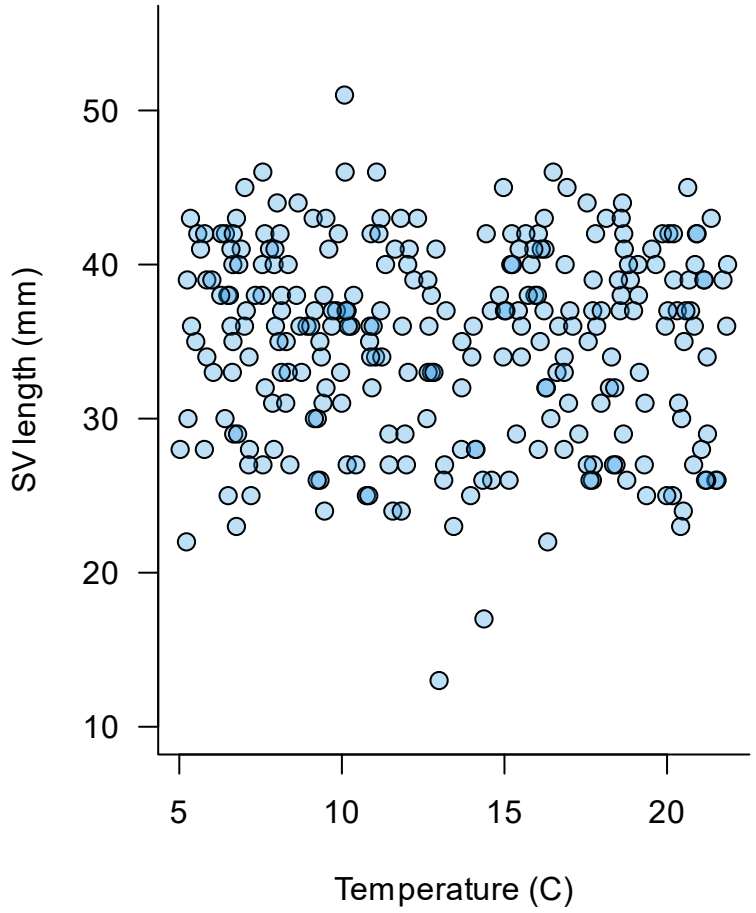
- How can we **quantify** our evidence for relationships?
- Are differences between groups *meaningful*?



Beyond graphs, Towards statistics

- How can we **quantify** our evidence for relationships?
- Are correlations *significant*?
- Are correlations *meaningful*?





Beyond graphs, Towards statistics

- Graphs are powerful tools that provide insight and understanding of the patterns and relationships in the data.
- Graphs alone don't give us the complete answer. We need to **quantify** the relationships we see in our plots.
- Statistics is the tool we use to formally answer these questions:
- Are differences, correlations, or associations *significant*?
  - What do we even mean by significant?