

Week 6: Graphical Data Exploration

Session 1

Spring 2020

iClicker Question 1

Which of the following should I use to read the file `mydata.csv` into a data frame called `dat` in R?

A `dat = read.csv(mydata.csv)`

B `read.csv(mydata.csv)`

C `dat = read.csv("mydata.csv")`

D `read.csv(mydata.csv, row.names = 1)`

E `read.csv("mydata.csv")`

iClicker Question 2

Which of the following lines of code will make a scatterplot with a custom **title**?

```
##      length      width mass
## 1 0.7209039 2.8777226   37
## 2 0.8757732 1.6052823   29
## 3 0.7609823 0.4713699   19
```

- A `plot(dat$mass, dat$length, col = 3)`
- B `plot(dat$mass, dat$length, type = "l")`
- C `plot(dat$length, dat$mass, type = "p")`
- D `plot(dat$mass, dat$length, main = "data plot")`
- E `plot(dat$length, dat$mass)`

iClicker Question 3

Which is the best location for me to type my R code for in-class and individual activities?

- A A text file in Word or another word procesor
- B The RStudio console pane
- C The RStudio code editor pane
- D The RStudio Environment pane

Announcements

Follow-up questions from the Chapter 5 homework

Error in Tuesday's title slide is now fixed. You can re-download the correct version on Moodle.

Week 5 reading assignment question # 1:

“What is the difference between parametric and a non-parametric distribution?”

- ▶ The book's explanation of *parametric distribution* is slightly misleading. The reading characterizes the Normal distribution as a parametric distribution, however there are many *parametric distributions*. I like to use the term *theoretical distributions* when I'm describing *parametric distributions* to emphasize that they are precisely mathematically defined. The behavior of *parametric distributions* is governed by 1 or more *parameters* in the probability functions.
- ▶ What is a **non-parametric distribution**?

Question 3: “In R, what would be returned if you compute the

For Today

Are there any general questions about data exploration, numeric or graphical?

Short lecture

Group activity: graphical data exploration part 2 - rarefaction data.

We'll start to use *inference* in the coming weeks!

Some graphics pointers

In summary, graphs are a useful data visualization tool

- ▶ summarizing
- ▶ understanding
- ▶ describing
- ▶ presenting/communicating

Some graphics pointers

In summary, graphs are a useful data visualization tool

- ▶ summarizing
- ▶ understanding
- ▶ describing
- ▶ presenting/communicating

BUT we must label the 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!*

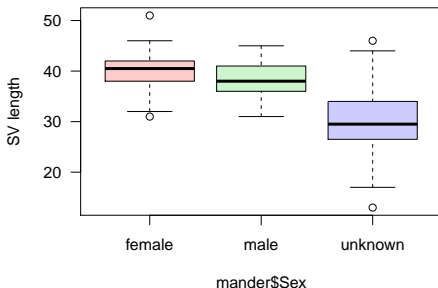
Some graphics pointers

In summary, graphs are a useful data visualization tool

Purpose	Graph Type
Illustrating <i>distribution</i>	Histogram, Density plot Box(-whisker) plot
Illustrating <i>differences</i>	Bar chart, Box plot
Illustrating <i>correlations</i>	Scatter plot
Illustrating <i>associations</i>	Pie chart, Bar chart
Illustrating <i>sample size</i>	Line plot of running avg

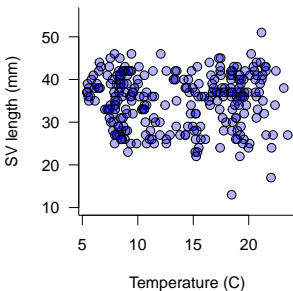
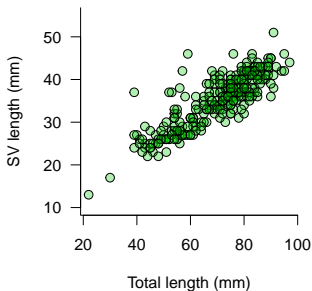
Beyond graphs, Towards statistics

- ▶ Graphs are powerful tools that provide insight and understanding of the patterns and relationships in the data.
- ▶ Don't give us the answer though:
 - ▶ are differences *significant*?
 - ▶ are associations *significant*?



Beyond graphs, Towards statistics

- ▶ Graphs are powerful tools that provide insight and understanding of the patterns and relationships in the data.
- ▶ Don't give us the answer though:
 - ▶ are differences *significant*?
 - ▶ are associations *significant*?



Beyond graphs, Towards statistics

- ▶ Graphs are powerful tools that provide insight and understanding of the patterns and relationships in the data.
- ▶ Don't give us the answer though:
 - ▶ are differences *significant*?
 - ▶ are associations *significant*?
- ▶ Statistics is the tool we use to formally answer these questions!
 - ▶ the differences *are/are not* significant!
 - ▶ are associations *are/are not* significant!

Group graphical activityf