Problem session prep

Natalie Nelson, PhD, Biological and Agricultural Engineering

BAE 495/590: Applied Statistics for Env. & Ag. Data Analysis

January 19, 2018

NC STATE

Recap

Deterministic vs. stochastic processes

What is “**uncertainty**” and how does it relate to deterministic and stochastic processes?

Recap

What are important considerations to make when comparing **samples** collected from different **populations**?

Recap

What are two ways in which you can call a column in a dataframe?

Recap

What does the str( ) function do?

I. Exploratory Data Analysis (EDA)

II. Quick R vocab

Exploratory Data Analysis

Exploratory data analysis is the process of exploring your data, and it typically includes examining the structure and components of your dataset, the distributions of individual variables, and the relationships between two or more variables.

Data visualization is arguably the most important tool for exploratory data analysis because the information conveyed by graphical display can be very quickly absorbed and because it is generally easy to recognize patterns in a graphical display.

Peng and Matsui (2017), *The Art of Data Science*

EDA Checklist

1. Formulate your question	Pre-EDA
2. Read in your data 3. Check the packaging 4. Look at the top and bottom of your data 5. ABC: Always Be Checking your "n"s 6. Validate with at least one external data source 7. Make a plot	Perform iteratively throughout EDA
8. Try the easy solution first 9. Follow up	Post-EDA

Peng and Matsui (2017), *The Art of Data Science*

Some R vocab for today's analysis

- Function(arguments). For example:

```
mean(x)
```

- # is used to comment out lines of code (meaning, R won't read lines of code that begin with "#"). Use comments frequently in order to organize your code, as well as make notes to yourself. For example:

```
# Mean of variable x
mean(x)
```

- To assign the output of a function to an object, use "<-". Think of this is an arrow pointing the function to the object name. For example:

```
y <- mean(x) # assign the mean of x to an object named y
```

Peng and Matsui (2017), *The Art of Data Science*

Boxplots

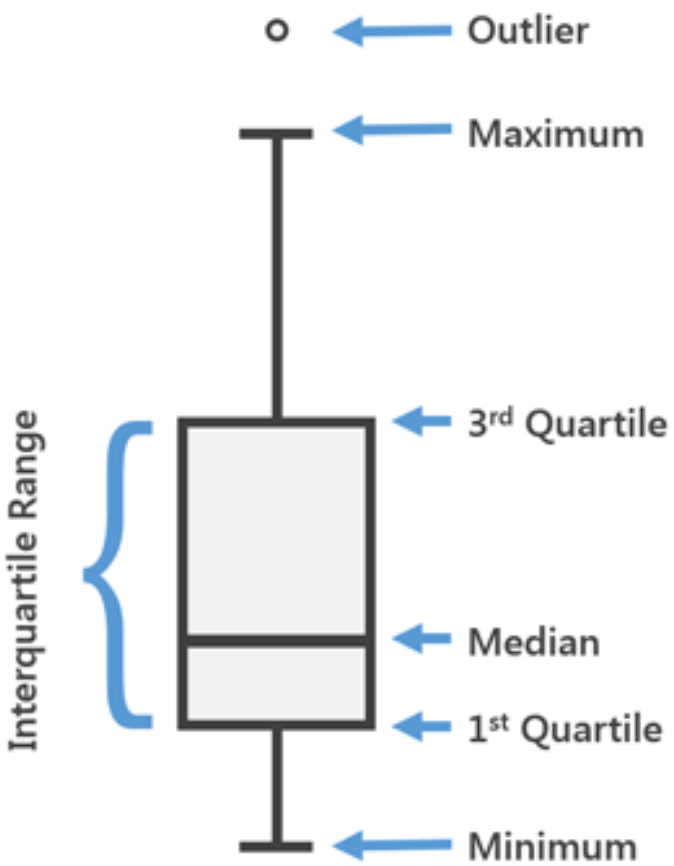

Image credit: ArcGIS Pro

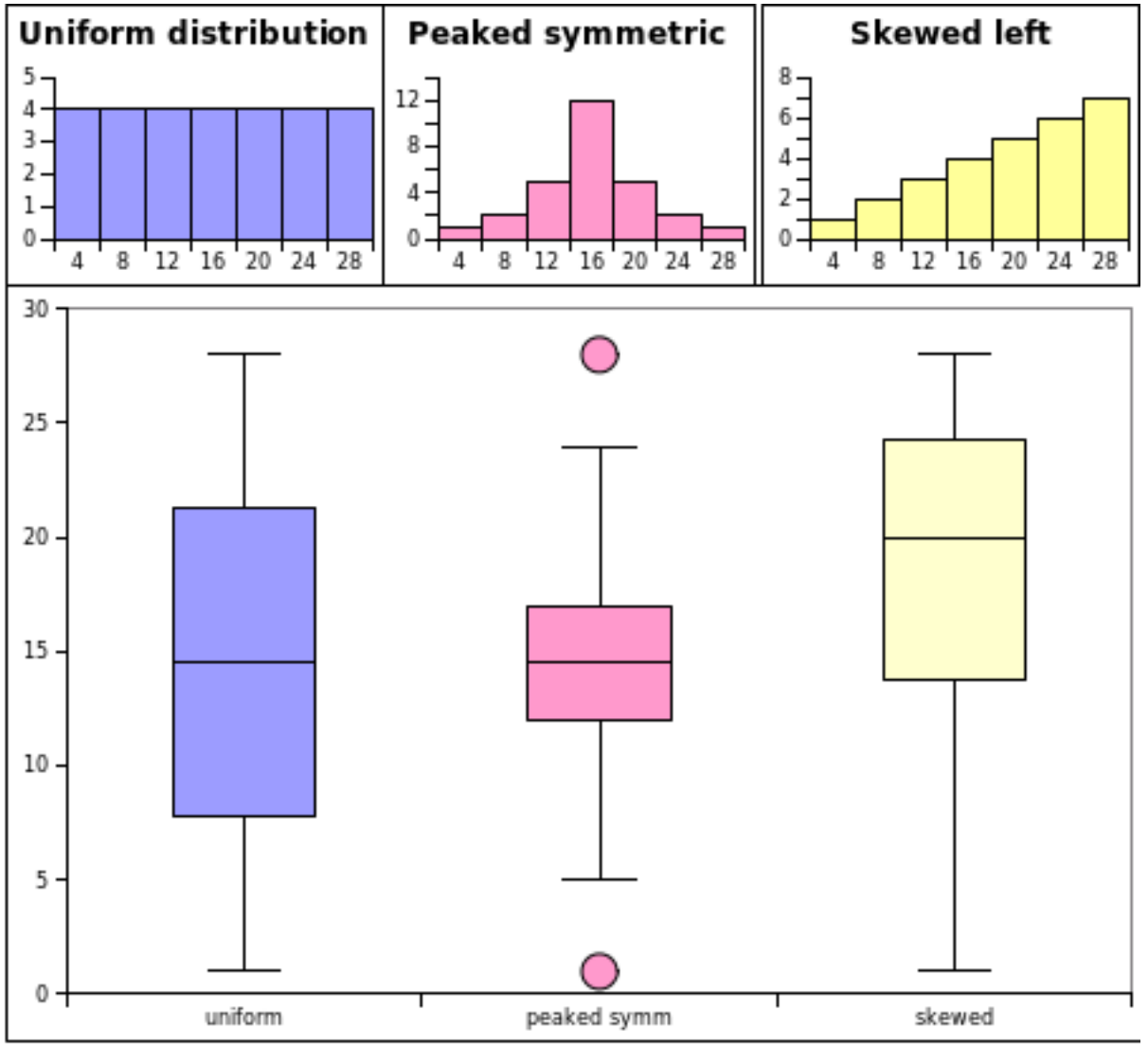

Image credit: Helical IT Solutions