Week 3 – Graphics in R

Checkpoint 1: Do you have questions about loading data? Looking at the data frame, what is

a statistical individual in this case? Stated differently, what is the CDC’s unit of sampling?

Checkpoint 2: Use the subset() function (or any other suitable function to answer the following

questions.

Checkpoint 3: How would the plot change if we wrote par(mar = c(15,5,4,2) instead?

Checkpoint 4: Make a barplot illustrating the number of outbreaks across states in 2023, the

most recent year of the data. Make sure your script can reproduce this figure. You will receive

credit for the appropriate figure, so make sure it has a title, axis labels where appropriate, and

is interpretable by a scientist.

Checkpoint 5: Which plot (or combination of plots) would you use to show that most epidemics

are very small, but there are a few very, very big ones? Provide some justification for your

answer.

Checkpoint 6: Write code to create a figure of the distribution of outbreak sizes in the state of

New York.

Checkpoint 7: What does this graph tell you about the relationship between outbreak size and

outbreak severity?

Checkpoint 8: Make a plot of the median outbreak size in North Carolina over time.

Checkpoint 9: Construct a good time series of the median enteric disease outbreak size in New

York over time. Save the figure as a .pdf.

Lab 4 – Probability

Checkpoint 1: From the str() function, answer the following questions.

• What are the statistical individuals in this data set?

**Plant Species**

• How many characteristics are measured about each statistical individual?

**16 (maybe 17 if includes species names)**

• How many characteristics are categorical and how many are numerical.

**2 numerical. 14 categorical (maybe 15)**

Checkpoint 2: Search where this species is found. Where is it found and what is its common

name?

## **dragon withe; Puerto rico**

Checkpoint 3: Select a random tree and show how it is pollinated and what its sexual system

is.

Checkpoint 4: Is sampling a tree and sample a needle leaf plan mutually exclusive? Explain

why.

**No, because a tree can be a needleleaf plan plant.**

Checkpoint 5: Based on this figure, which sexual system has the highest probability of being

selected? Which sexual system by shade tolerance combination has the highest probability of

being selected?

**Hermaphrodites are most common sexual system.**

**Hermaphrodites with intermediate shade tolerance look most likely**

Checkpoint 6: Make a mosaic plot to visualize whether the presence of fleshy fruits is independent

of whether or not the plant species has a seed bank. Based on this figure, present an

argument for or against independence of these two characters.

**They look independent to me. There are only moderate differences in the prevalence of fleshy fruits of plant species with and without seedbanks. Could be sampling variation.**

#### Checkpoint 7: Create a Bernoulli random variable that describes whether an individual has a mutant allele. Let the probability of the individual having the allele be 0.01. Sample an individual from this random variable and tell me what it is. Make sure to set the seed in your code so that it is repeatable on my computer.

#### Checkpoint 8: What is the probability of catching 3 trout in this sampling design?

**dbinom(3, 10, 0.1) = 0.0574**

#### Checkpoint 9: Write code to show the sampling distribution of the number of trout in 20 fish, rather than 10, under the assumption that $p\_{trout} = 0.1$.

Checkpoint 10: Plot the probability distribution and cumulative distribution for the following sample design. You are sampling 50 individuals for infection status where the probability that a single individual is infected is 0.13 and you want to know how many individuals in your sample are infected. Also, find the 30th and 80th quantiles for this distribution.

**5 and 8**

Checkpoint 11: Write code that produces a two panel figure. In one, show the probability distribution for a normal with mean 5 and standard deviation of 20. In the other, show that distribution's cumulative density.

Checkpoint 12: Write code to find the 63rd and 12th quantiles of a normal distribution with mean 5 and standard deviation of 20.

*X0.63* = 11.637 *X0.12* = -18.500