**Topic 2**

**Data and Distributions**

Recall that data is the set of values that represent a variable or variables for a certain set of observational units (or OUs).

The set of data values for a characteristic of a population has a *distribution.*

**Distribution -** an arrangement of values of a variable showing their observed or theoretical frequency of occurrence[[1]](#endnote-1)

The distribution of a variable refers to it pattern of variation. Recall that in topic 1 we discussed that variables vary. They often follow a pattern of variation. This pattern is called a variable’s distribution.

At this point, we will only discuss a *sample distribution*, or a distribution of values from data.

How the distribution is handled depends on the variable type.

Categorical Variables

* + Display with a bar graph.[[2]](#footnote-1)
	+ The order of the categories is not important
	+ The distribution does not have a distinct shape[[3]](#footnote-2)
	+ Determine distribution with **proportional values**[[4]](#footnote-3)

Ordinal Variables

* + Display with
		- Histogram
		- Dotplot if n (the number OUs) is small
	+ Values are in determined order
	+ Distributional shape is important
	+ Distribution may also be displayed using proportional values

Quantitative Variable

* + Display with
		- Histogram
		- Dotplot if n (the number OUs) is small
	+ Values are in numerical order
	+ Distributional shape is very important
1. http://www.thefreedictionary.com/statistical+distribution [↑](#endnote-ref-1)
2. When discussing the distribution of values for a categorical variable, then bar graph generated from the data is not called a ‘distribution’ as would a graph (or histogram) generated by ordinal or quantitative data. [↑](#footnote-ref-1)
3. This is because changing the order of the categories changes the locations of the different bars, and thus the overall shape of the graph. [↑](#footnote-ref-2)
4. See Proportionsv2.docx in blackboard content section. (may be v3 or v4 or another number) [↑](#footnote-ref-3)