

ONE QUANTITATIVE VARIABLE

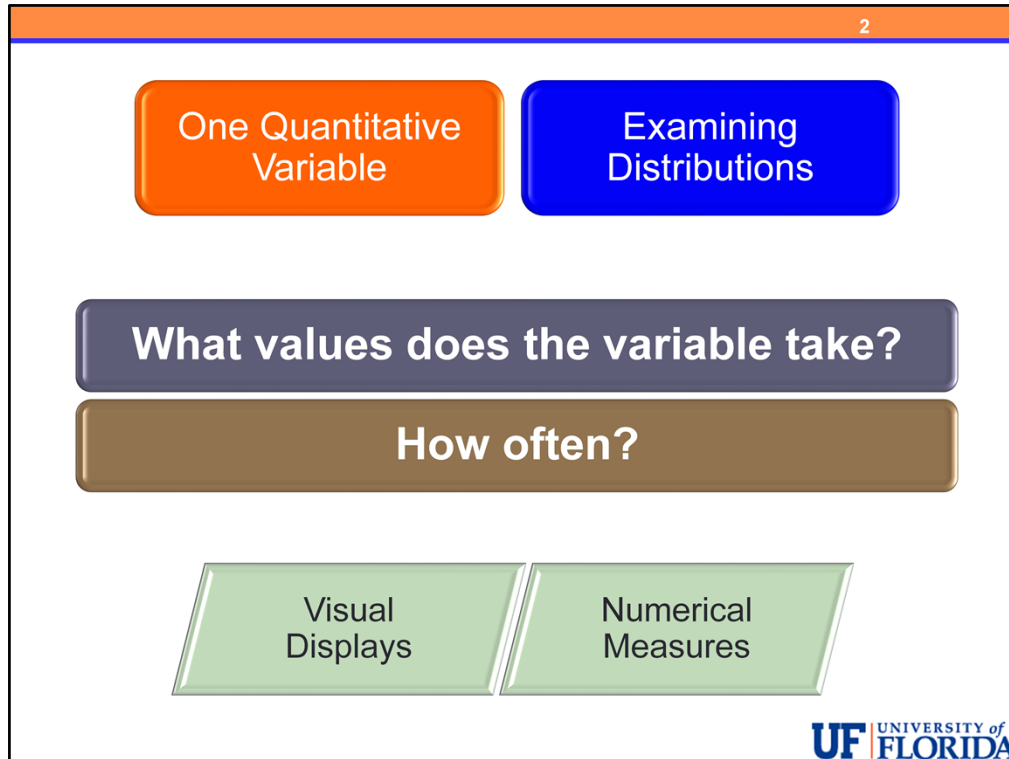
Describing Distributions
Unit 1: Exploratory Data Analysis



UF UNIVERSITY of
FLORIDA

How do we summarize one quantitative variable?

What visual displays and numerical measures are appropriate?



Recall that, when we say Distribution we mean

What values the variable can take

And

How often the variable takes those values.

Exploratory Data Analysis for one quantitative variable is more involved than that for one categorical variable because we have additional considerations when we study the distributions of one quantitative variable such as the shape, center, and spread of the distribution as well as identification of outliers and finding numerical measures to summarize some of these aspects of the distribution.

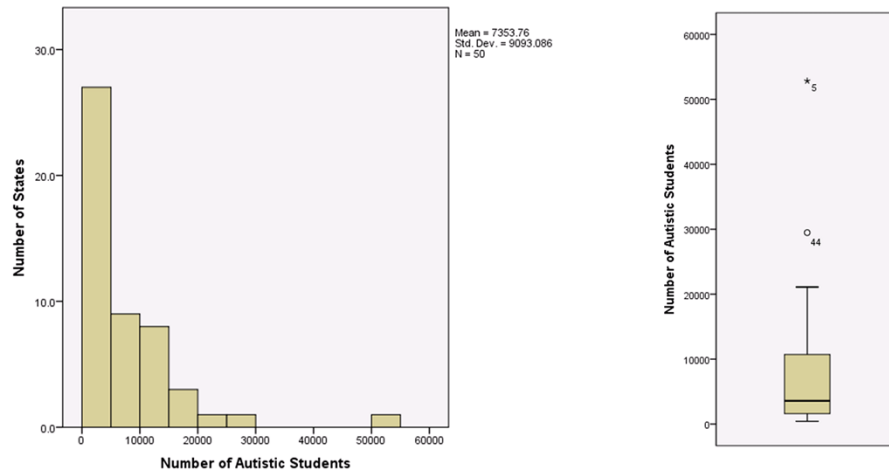
Again we will discuss both Visual Displays and Numerical Measures.

For visual displays we typically use a histogram for a single quantitative variable, however, we will also discuss boxplots in detail as these are very useful for comparing groups and will be used often. By-hand displays such as the stem plot and dot plot will be briefly explained.

With numerical measures, we wish to quantify the center and spread of the distribution as well as measure the location of a particular value in the distribution.

Let's return to the examples from the disability dataset on Autistic students and briefly introduce some of these ideas.

Number of Autistic Students

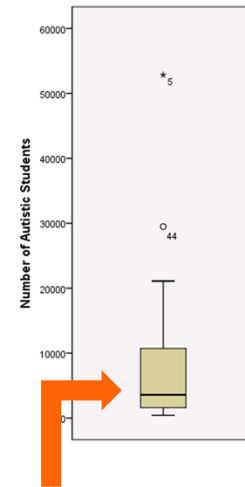
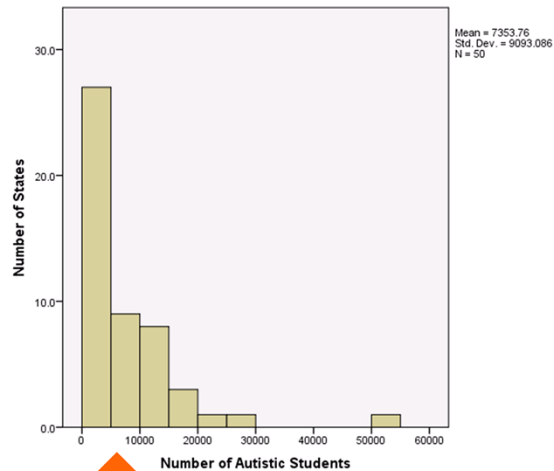


For one quantitative variable, the most common visual display is a histogram.

In this histogram, we can see that most states have low numbers of autistic students with fewer states having higher numbers of autistic students enrolled. The shape of this distribution for this variable (Number of Autistic Students Enrolled) is called skewed right.

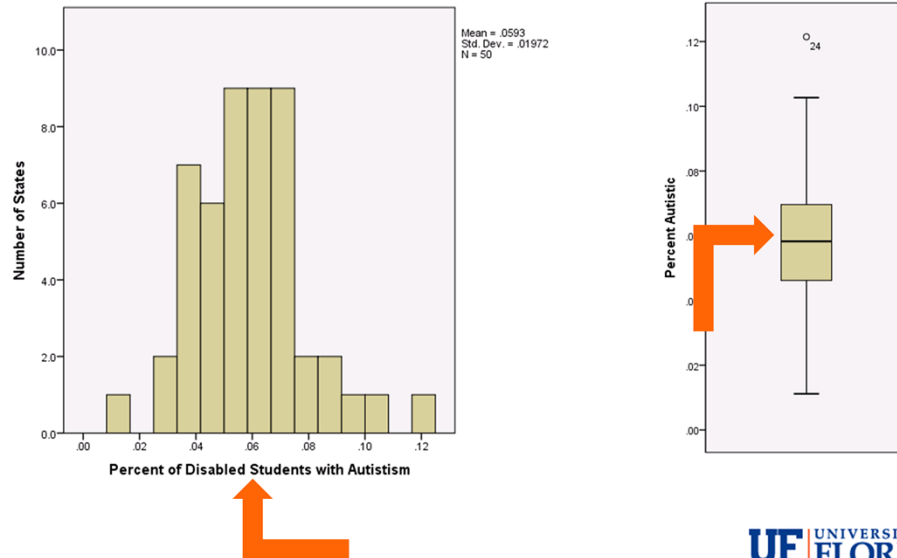
There are also a few outliers representing two states with unusually large values for this variable.

Number of Autistic Students



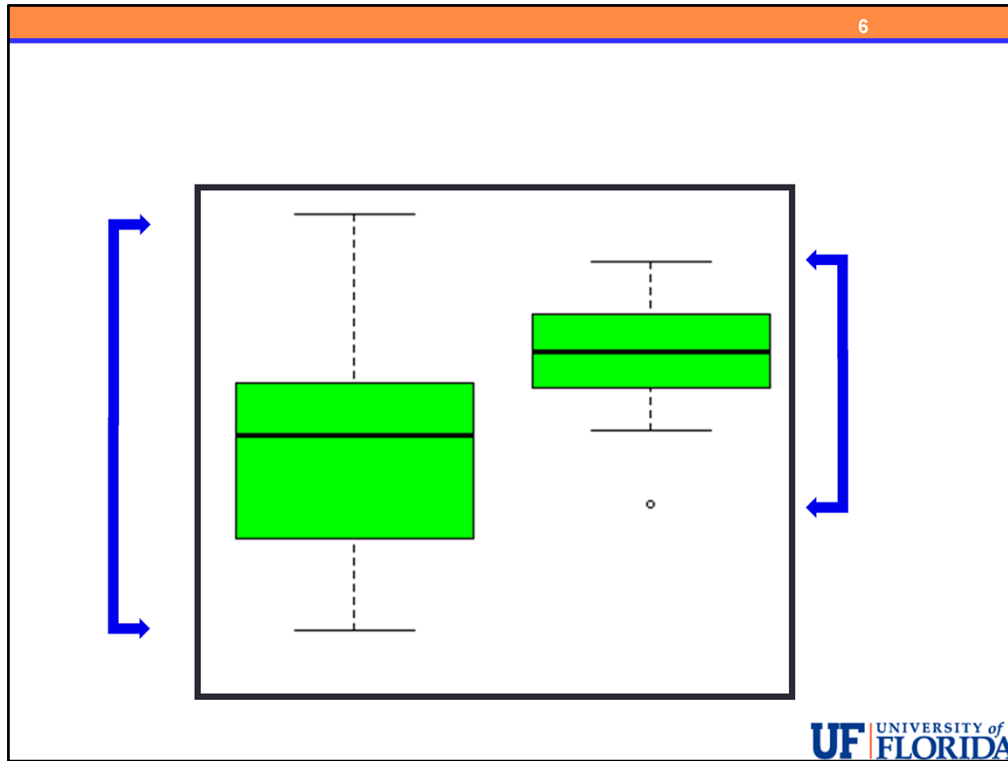
The “center” of this distribution is difficult to define, however, we tend to desire a value that locates the “bulk” of the data.

Relative to Total Number Disabled



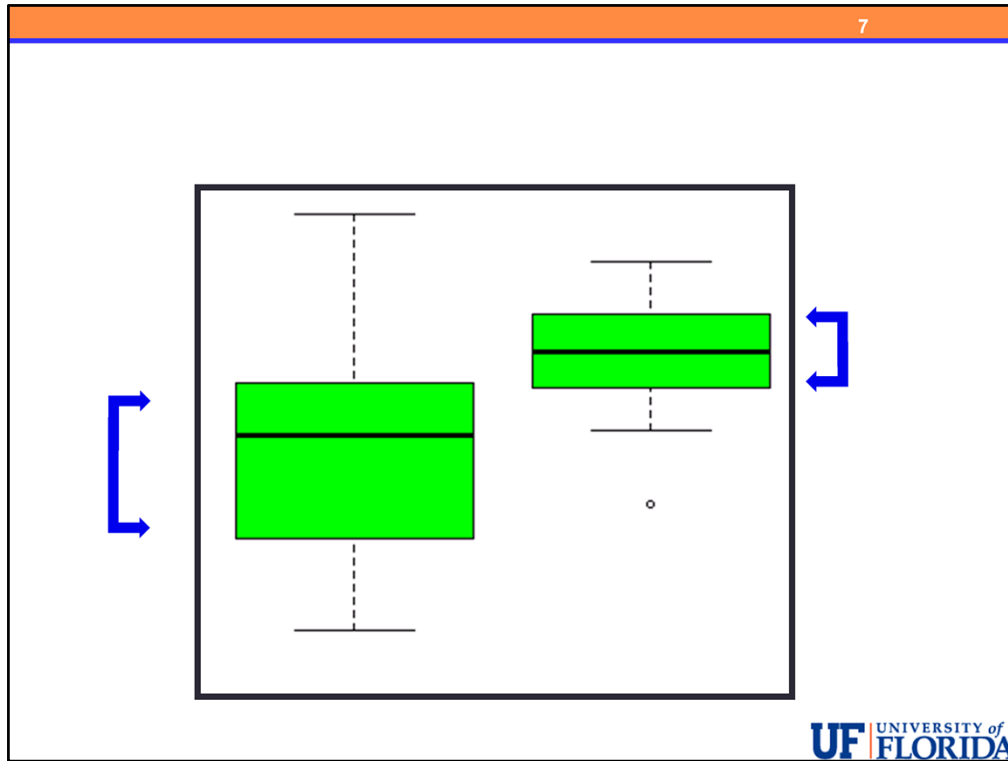
The distribution for the variable “Percent of Disabled Students with Autism” is reasonably symmetric with one unusually large outlier.

For this distribution the center is much easier to define and locate.



We will define three ways to measure the spread of the distribution two of which we will see are easily visible in the boxplot.

The range is the distance from the minimum to the maximum.



The IQR or Inner Quartile Range is the length of the box.

For these two distributions, the one on the left has more variation, is more variable, than the one on the right.

We can also say that the center of the distribution for the distribution on the left is less than that of the distribution on the right.

Interactive Applet

- We can Analyze One Quantitative Variable with this

[One-Variable Statistical Calculator](#)

- From online content for [Introduction to the Practice of Statistics](#), Seventh Edition (Moore and McCabe)



This applet will quickly show the results for a number of datasets. You can also enter your own data and view the results.

{web address of One-Variable Statistical Calculator applet:

http://content.bfwpub.com/webroot_pubcontent/Content/BCS_4/IPS7e/Student/Statistical%20Applets/onevar.html

}

{web address of textbook: <http://bcs.whfreeman.com/ips7e> }

Interactive Applet: Pick Data

Data Sets | Data | Histogram | Statistics | Stem Plot

- ☐ IQ scores of fifth graders (Table 1.1 page 13)
- ☐ Scores on a first exam in an introductory statistics course (Exercise 1.10, page 14)
- ☐ Service times at a call center (Table 1.2, page 15)
- ☐ Carbon dioxide emissions by country (Table 1.3, page 26)
- ☒ Acidity of rainwater (Exercise 1.38, page 26)
- ☐ Diameters of longleaf pine trees (Exercise 1.68, page 46)
- ☐ Blood proteins of children from Papua New Guinea (Exercise 1.69, page 47)

We will use the acidity of rainwater. Once you select the dataset you wish to use, you can view the data and results in the other four tabs:

Interactive Applet: Pick Data

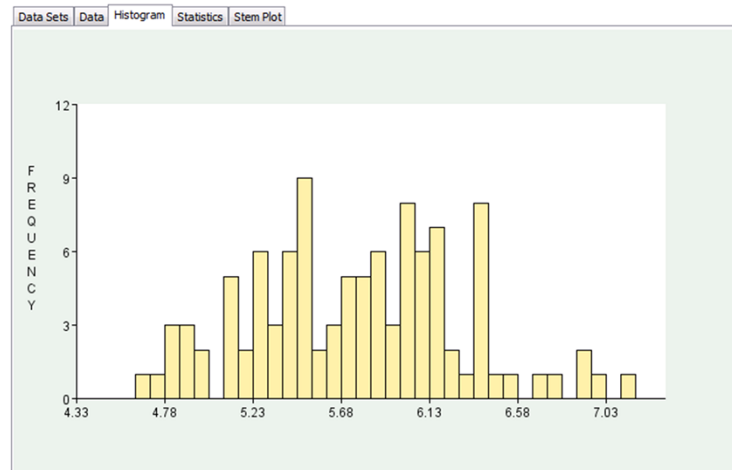
Acidity of rainwater (Exercise 1.38, page 26)

	pH	
1	4.33	▼
2	4.38	
3	4.48	
4	4.48	
5	4.50	
6	4.55	
7	4.59	
8	4.59	
9	4.61	
10	4.61	🗑️

Data: Here we can view all of the data for this dataset. In this case, there are 105 observations on the pH of rainwater.

On this tab you can click on the trashcan to remove this data and enter your own data.

Interactive Applet: Pick Data



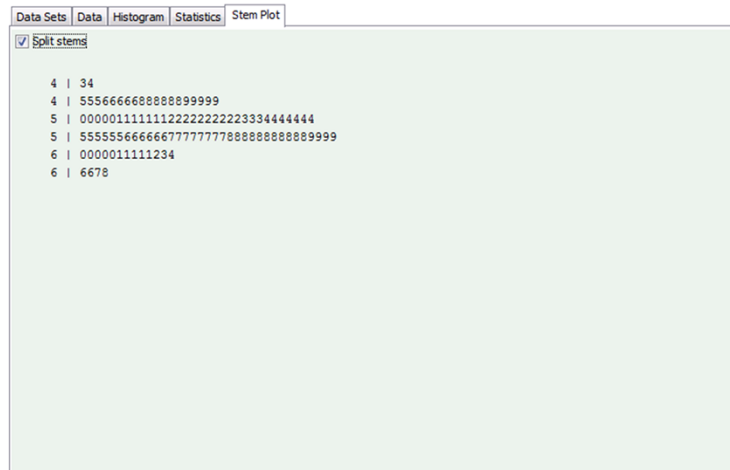
Histogram: Here we see a histogram of the data. We will discuss histograms further but likely you have seen graphs similar to this before.

Interactive Applet: Pick Data



Statistics: This tab gives some numerical measures including the number of observations (labeled capital N), the mean (or average) which is around 5.4, the standard deviation, and the five-number summary which includes the minimum, maximum, first and third quartiles, and the median.

Interactive Applet: Pick Data



Stem plot: These displays are useful for creating a quick display by-hand and sometimes are provided by software packages so it is a good idea to at least be exposed to this way of organizing data. In this tab you can show the stem plot with or without split stems by checking the box in the top left corner.

This applet can be a useful tool for extra practice and understanding as you review the material in this section on examining the distribution of one quantitative variable.



ONE QUANTITATIVE VARIABLE

Describing Distributions

Unit 1: Exploratory Data Analysis



There is a lot of information to cover regarding examining the distribution of one quantitative variable.

Try to keep our broad goal in mind – examining the distribution of the variable – as we discuss the specifics of particular visual displays and numeric measures.