Chapter Three Part One: Data Visualization and the Shape of Distributions

Remember, a _____ is a variable that is measured on a numeric scale.

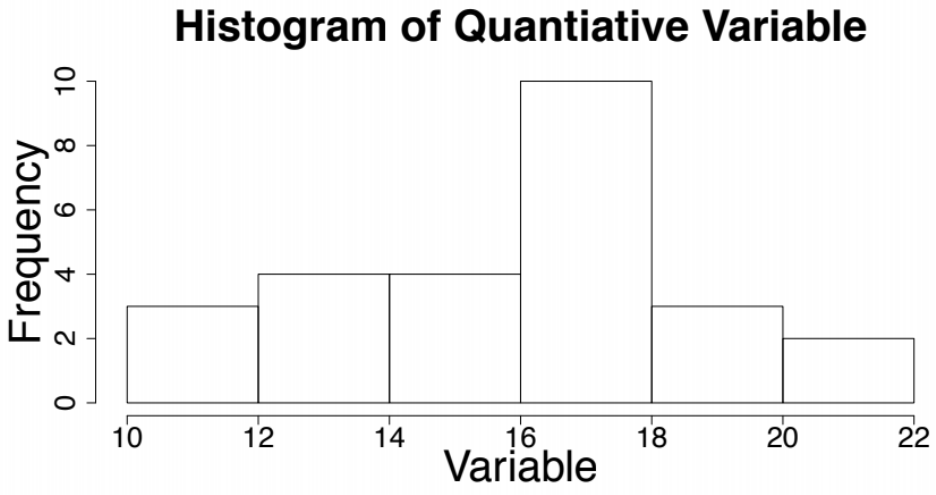
-
-

When trying to glean information from quantitative variables there are two things we look at:

1.
2.

Data Visualization - Histograms

While a histogram looks similar to a bar plot, they mean different things. A histogram is a graphical display of the _____ of a quantitative variable. This is a summary of quantitative information it does not _____ _____.



To build a histogram, we perform the following actions:

1.

2.

3.

4.

5.

Example *The dataset USStates contains information about the 50 states in the US. A porption of the data is shown below:*

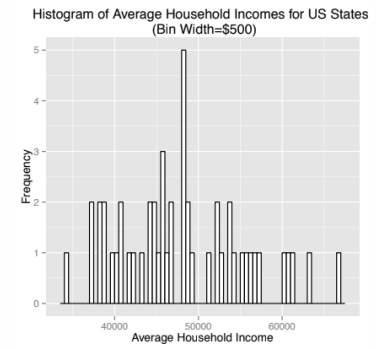
	State	HouseholdIncome	IQ	ObamaMcCain
1	Alabama	38160	95.70	M
2	Alaska	57071	99.00	M
3	Arizona	46693	97.40	M
4	Arkansas	37458	97.50	M
5	California	54385	95.50	O
6	Colorado	53900	101.60	O
7	Connecticut	60551	103.10	O
8	Delaware	52676	100.40	O
9	Florida	45038	98.40	O
10	Georgia	48388	98.00	M

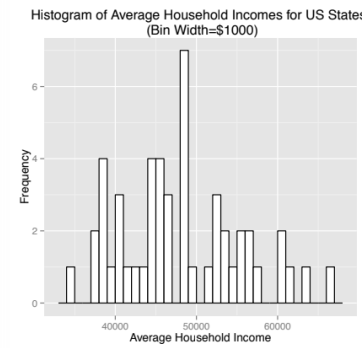
1. *Create bins.*

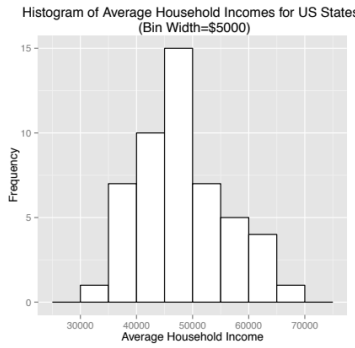
2. *Count the number of cases in each bin.*

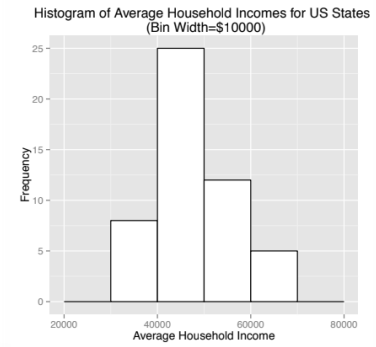
3. Create the Histogram.

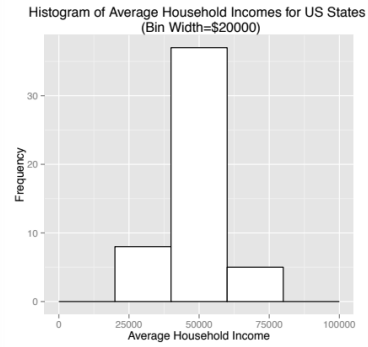
Be careful with how you choose your bins!

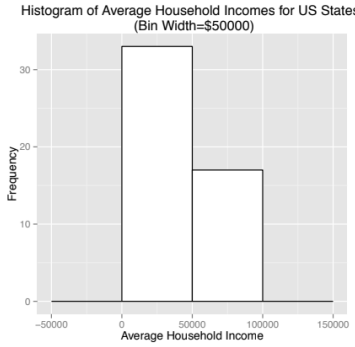




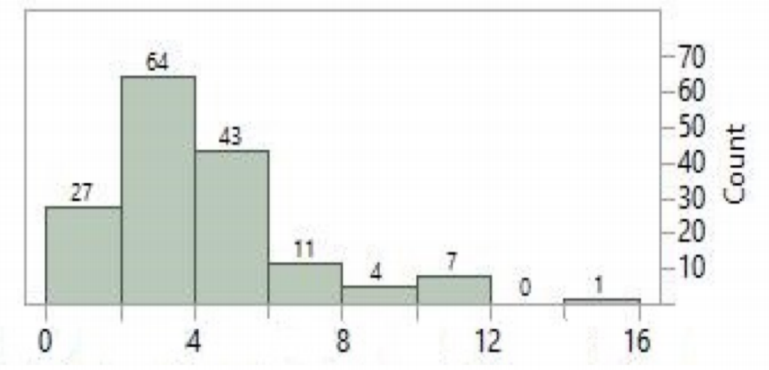








Example *Below is a histogram of the tips(in dollars) received by a server at a restaurant in one week.*



The fourth largest bin has a count of 11. This indicates:

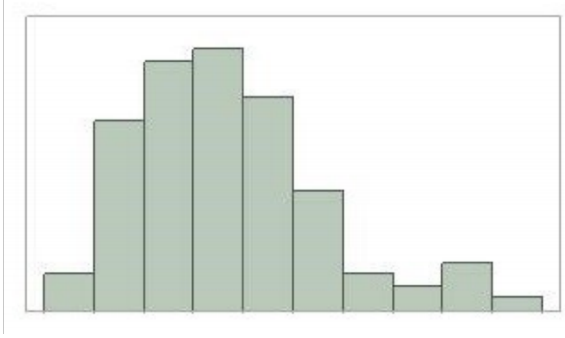
(a) There were 11 seven dollar tips.

(b) There were 11 tips between six to eight dollars.

(c) There were 11 eight dollar tips.

(d) There were 11 six dollar tips.

Example *A histogram for the number of states in the U.S. visited by a random sample of 100 STAT 101 students is given below:*



Choose the best description of the vertical axis in this image.

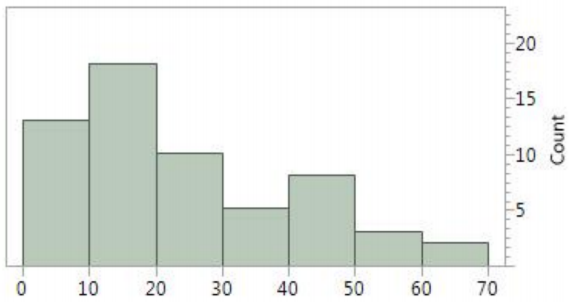
(a) The number of states

(b) The mean number of states in each bin

(c) The number of students

(d) The mean number of students in each bin

Example *Below is a histogram for arrival delays of flights for an airline.*



How many flights were delayed between 10 and 20 minutes?

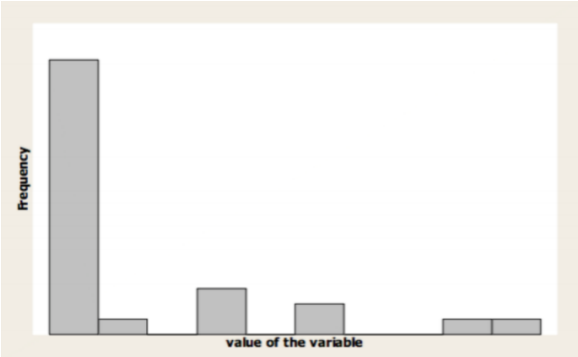
(a) 12

(b) 5

(c) 18

(d) 10

Example *Students at another university completed an in class survey. Consider the following variables and select the one you think is displayed in the histogram below:*



(a) Hours of sleep on a typical weeknight

(b) Monetary amount in dollars of carried coins

(c) Randomly selected integer between 0 and 9

(d) Height in inches

Stem-and-Leaf Plot

A stem-and-leaf plot is another way of displaying the distribution of a _____.

-
-
-
  -
  -

To build a stem-and-leaf plot we:

1.
2.
3.
4.

Note: JMP orders stems in increasing order from bottom to top (backwards from usual)

Example *Suppose we have data of the weight filled soda cans in grams.*

348, 348, 349, 351, 352, 353, 353, 354, 355, 355, 355, 356

357, 362, 365, 366, 367, 368, 368, 369, 369, 370, 370, 373, 378

Create a stem-and-leaf plot from the data given.

Here is an example of a stem-and-leaf plot in JMP.

Stem	Leaf	Count
7	01	2
6	8899999	7
6	666677	6
6	444444455555	12
6	2223333	7
6	000000011111111	15
5	88899	5
5	666666677	9
5	4444555	7
5	2233	4
5	0	1

5|0 represents 50

What we notice from the JMP plot:

- JMP puts things in a non-intuitive order
- Key at the bottom of the plot
- Manageable amount of cases, but generally, _____

Describing the Shape of a Distribution

There are three main attributes that we use to discuss the shape of a distribution, especially when we are describing _____.

-
-
-

The acronym I use to remember these three attributes is **MiSO** soup!

Modes

Modes refer to the "humps" in the histogram, or rather where _____.
We refer to the specific number of "humps" in our distribution using the following names:

- **Uniform:**
- **Unimodal:**
- **Bimodal:**
- **Multimodal:**

Let's draw a uniform, unimodal, and bimodal distribution.

Symmetry

The symmetry of a distribution refers to whether both sides of a distribution are roughly equal or symmetric. For example, if the plot were to be folded in half, the distribution would cover itself up.

Example *Example of symmetric and non-symmetric distributions*

Another aspect of symmetry is whether or not the distribution is _____.

- **Skewed Right:**
- **Skewed Left:**

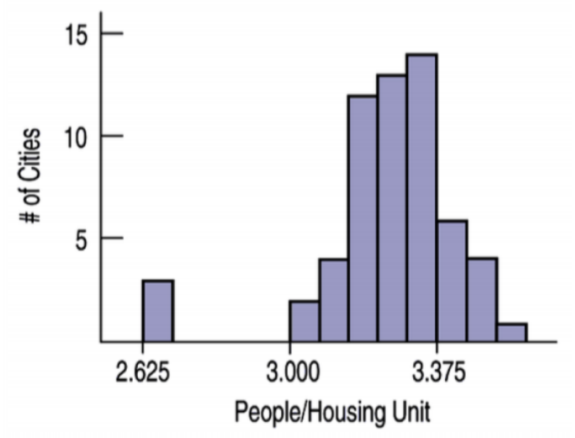
Example *Example of right and left skewed distributions*

Outliers

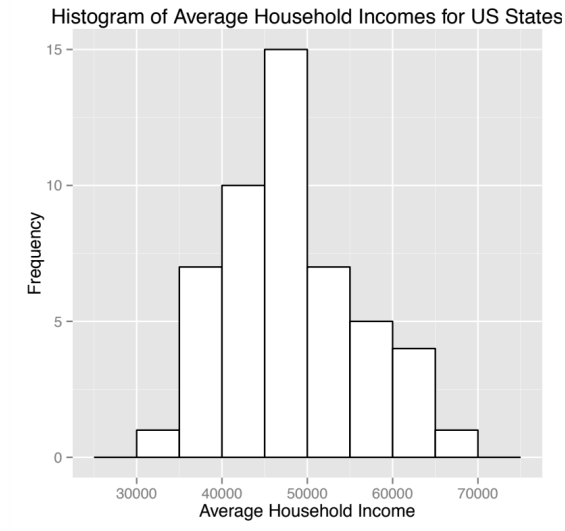
Outliers are extreme values present in our distribution that differentiate themselves from the rest of the distribution. Typically found in the _____ of the distribution. Understanding and making note of outliers is important for three key reasons:

- Can occasionally be informative or indicative of some sort of error
- Impact the statistical methods we employ when analyzing data
- Can dramatically impact results

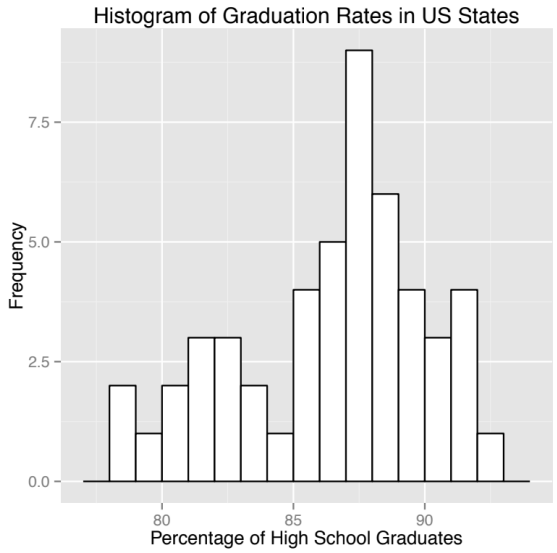
To look for outliers on a histogram we look for _____.



Example *Describe the shape of the distribution:*



Example *Describe the shape of the distribution:*



Example *Describe the shape of the distribution:*