

Lecture 4

Descriptive Statistics II

Text: Chapter 3

STAT 8010 Statistical Methods I

August 28, 2019

Review of Last Class

Graphical Summaries
of Quantitative
Variables

Numerical Summaries
of Quantitative
Variables

Whitney Huang
Clemson University

Review of Last Class

Graphical Summaries
of Quantitative
Variables

Numerical Summaries
of Quantitative
Variables

- 1 **Review of Last Class**
- 2 **Graphical Summaries of Quantitative Variables**
- 3 **Numerical Summaries of Quantitative Variables**

- Summarizing **categorical** variables
 - Frequency Table
 - Bar Chart and Pie Chart
- Summarizing **numerical** variables
 - Mean \Rightarrow A measure of **central tendency**
 - Variance / Standard Deviation \Rightarrow A measure of **spread**

Example: Clemson Fact Sheet 2018

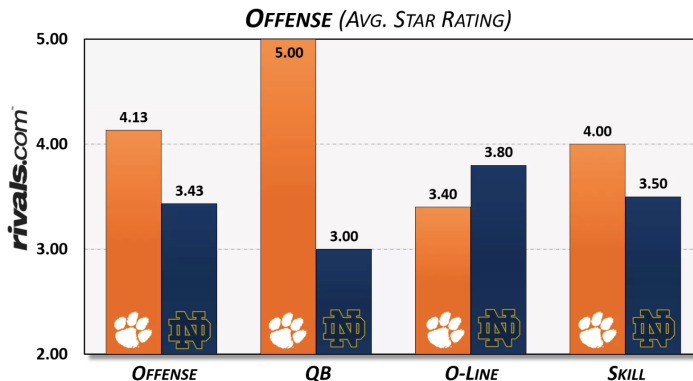
STUDENT DISTRIBUTION

			Full Time	Part Time	Total	Age	Average	
RACE/ETHNICITY/GENDER		Freshman	3,273	10	3,283	16-73	18	
Non-Resident Alien	6%	Sophomore	4,765	61	4,826	17-79	19	
Hispanic	4%	Junior	4,786	122	4,908	17-64	20	
American Indian or Alaskan Native	<1%	Senior	6,121	426	6,547	18-66	22	
Asian	2%	Uncl. U/G	26	79	105	16-73	24	
Black or African American	6%	Master's	1,661	1,629	3,290	19-78	29	
Native Hawaiian or Pacific Islander	<1%	Doctoral	1,291	292	1,583	20-71	30	
White	77%	Certificate		94	94	21-56	35	
Two or More Races (non-Hispanic)	3%	Specialists	1	29	30	26-55	36	
Unknown	1%	Uncl. Grad	16	269	285	23-73	38	
51% Male, 49% Female			Total	21,940	3,011	24,951	16-79	22

Data source: [https:](https://www.clemson.edu/institutional-effectiveness/documents/oir/minis/F18FactSheetUpdated.pdf)

[//www.clemson.edu/institutional-effectiveness/documents/oir/minis/F18FactSheetUpdated.pdf](https://www.clemson.edu/institutional-effectiveness/documents/oir/minis/F18FactSheetUpdated.pdf)

Example: Clemson vs. Notre Dame

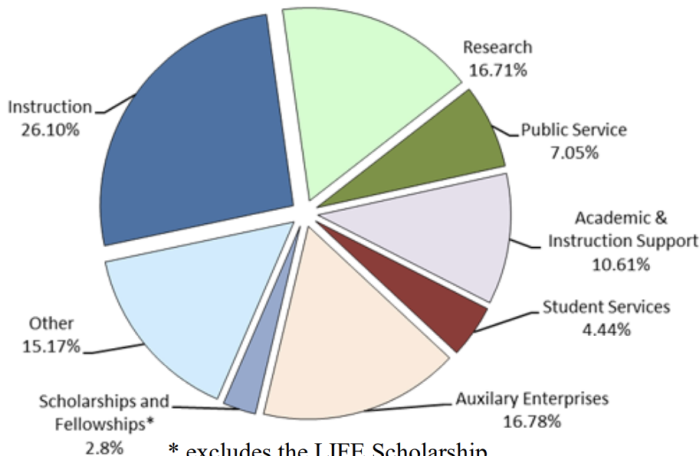


Data source: <https://www.shakinthesouthland.com/2018/12/24/18132204/>

[cotton-bowl-clemson-vs-notre-dame-preview-depth-chart-statistical-analysis](#)

Example: Clemson Expenditures 2017-2018

EXPENDITURES FY 2017-18



* excludes the LIFE Scholarship

+ taken from the Clemson Annual Financial Report

Data source: <https://www.clemson.edu/institutional-effectiveness/documents/oir/minis/F18FactSheetUpdated.pdf>

<https://www.clemson.edu/institutional-effectiveness/documents/oir/minis/F18FactSheetUpdated.pdf>

Review of Last Class

Graphical Summaries
of Quantitative
Variables

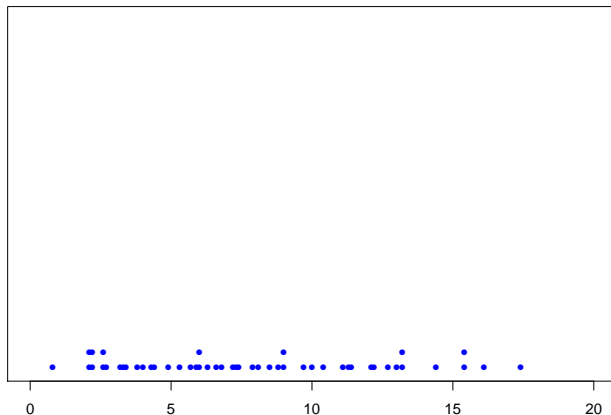
Numerical Summaries
of Quantitative
Variables

Example: Murder arrests (per 100,000) in US States in 1973

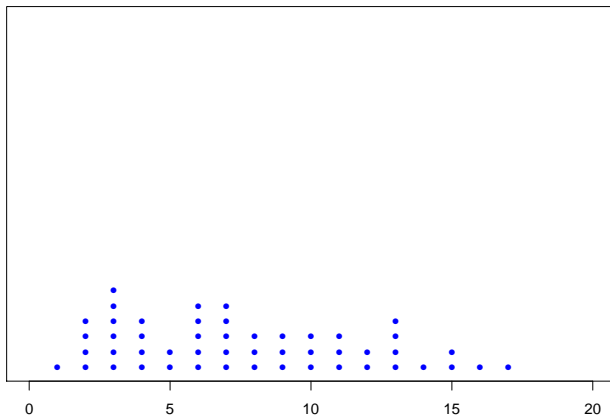
Data: 13.2, 10.0, 8.1, 8.8, 9.0, 7.9, 3.3, 5.9,
15.4, 17.4, 5.3, 2.6, 10.4, 7.2, 2.2, 6.0,
9.7, 15.4, 2.1, 11.3, 4.4, 12.1, 2.7, 16.1,
9.0, 6.0, 4.3, 12.2, 2.1, 7.4, 11.4, 11.1,
13.0, 0.8, 7.3, 6.6, 4.9, 6.3, 3.4, 14.4, 3.8,
13.2, 12.7, 3.2, 2.2, 8.5, 4.0, 5.7, 2.6, 6.8.

Question: How to graphically summarize this data set?

Dotplot of Murder Arrest Rate (per 100,000)



Rounded Murder Arrest Rate (per 100,000)

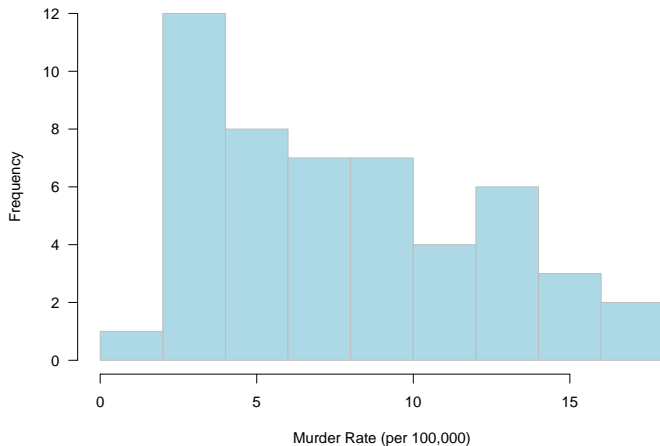


Stem-and-Leaf Plot

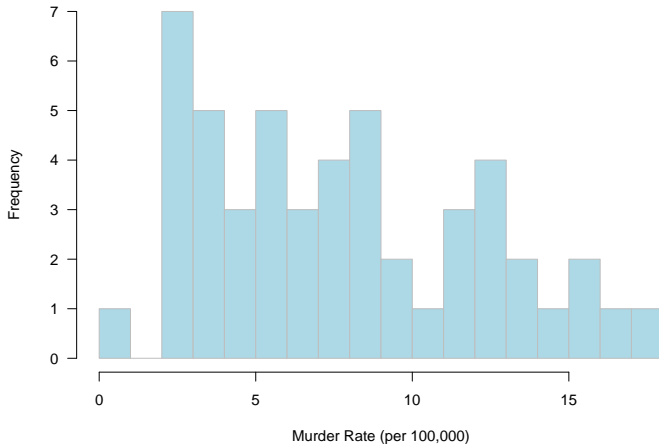
The decimal point is at the |

```
0 | 8
1 |
2 | 1122667
3 | 2348
4 | 0349
5 | 379
6 | 00368
7 | 2349
8 | 158
9 | 007
10 | 04
11 | 134
12 | 127
13 | 022
14 | 4
15 | 44
16 | 1
17 | 4
```

Histogram of US Murder Rate in 1973



Histogram of US Murder Rate in 1973



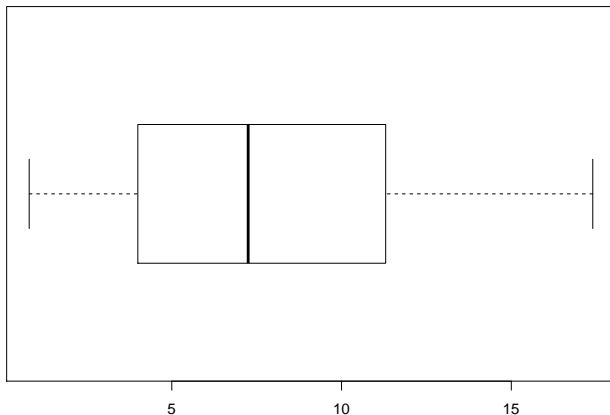
Box-and-Whisker Plot

Review of Last Class

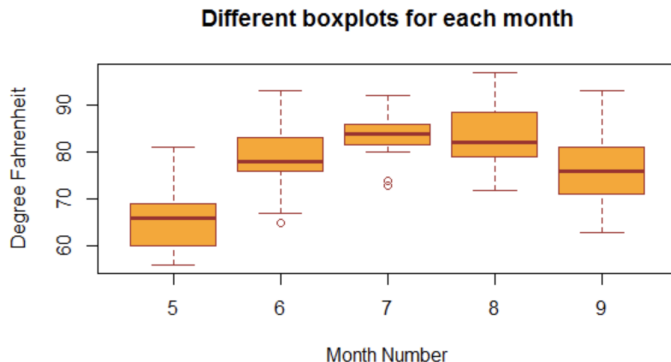
Graphical Summaries
of Quantitative
Variables

Numerical Summaries
of Quantitative
Variables

Murder Rate (per 100,000)

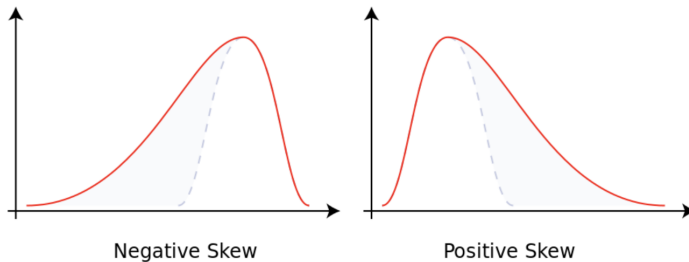


Side-by-Side Boxplots



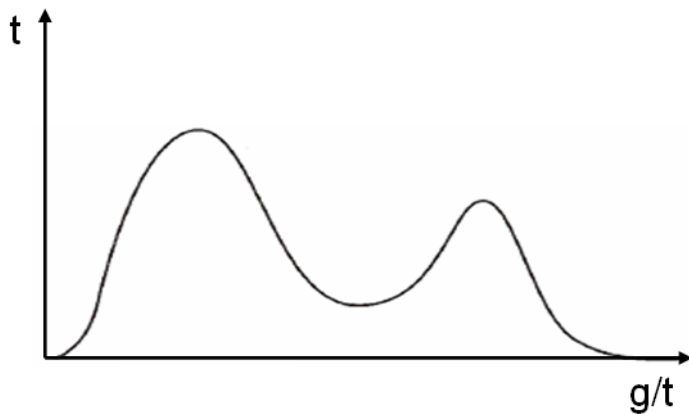
Source: <https://www.datamentor.io/r-programming/box-plot/>

Shape of Distributions



Source: [Skewness - Wikipedia](#)

Shape of Distributions cont'd



Source: [Multimodal distribution - Wikipedia](#)

- A **measure of center** attempts to report a “typical” value for the variable
- When a measure of center is calculated with **sample data** it is a **statistic**
- When a measure of center is calculated with popular (e.g., census data) it is a **parameter**
- **Measures:** Mean, Median, Mode, ...

- The **population mean**, denoted by μ_X , is the sum of all the population values ($\{X_i, \dots, X_N\}$) divided by the total number (N) of population values. That is,

$$\mu_X = \frac{\sum_{i=1}^N X_i}{N}$$

- The **sample mean**, denoted by \bar{X} is the sum of all the sample values ($\{X_1, \dots, X_n\}$) divided by the total number of sample values (n). That is

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

Example

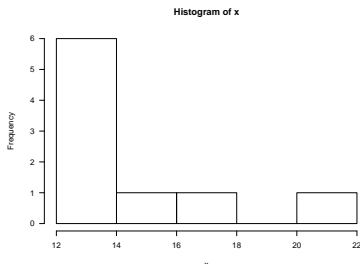
Suppose we have the following list of values: 13, 18, 13, 14, 13, 16, 14, 21, 13

- Plot this “data set” and describe the shape of the distribution

Review of Last Class

Graphical Summaries
of Quantitative
Variables

Numerical Summaries
of Quantitative
Variables



- Find the mean (both sample and population means)

$$\bar{X} = \mu_X = \sum_{i=1}^9 \frac{13 + 18 + 13 + 14 + 13 + 16 + 14 + 21 + 13}{9} = 15$$