

# Exploratory Data Analysis

## Section 2-7

---

---

---

---

---

---

---

---

# Exploratory Data Analysis

**the process of using statistical tools (such as graphs, measures of center, and measures of variation) to investigate the data sets in order to understand their important characteristics**

---

---

---

---

---

---

---

---

# Outliers

- ❖ **a value located very far away from almost all of the other values**
- ❖ **an extreme value**
- ❖ **can have a dramatic effect on the mean, standard deviation, and the histogram so that the true nature of distribution is totally obscured**

---

---

---

---

---

---

---

---

# Boxplot

(Box-and-Whisker Diagram)

Reveals the:

- ❖ center of the data
- ❖ spread of the data
- ❖ distribution of the data
- ❖ presence of outliers

Excellent for comparing two or more data sets

---

---

---

---

---

---

---

---

# Boxplot

5 - number summary

- ❖ Minimum
- ❖ first quartile  $Q_1$
- ❖ Median ( $Q_2$ )
- ❖ third quartile  $Q_3$
- ❖ Maximum

---

---

---

---

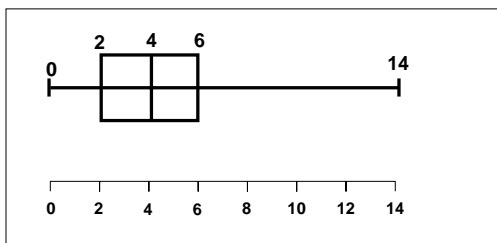
---

---

---

---

# Boxplot



Boxplot of Qwerty Word Ratings

---

---

---

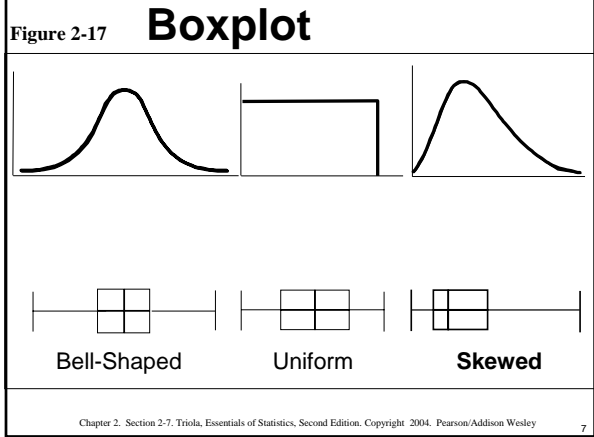
---

---

---

---

---




---

---

---

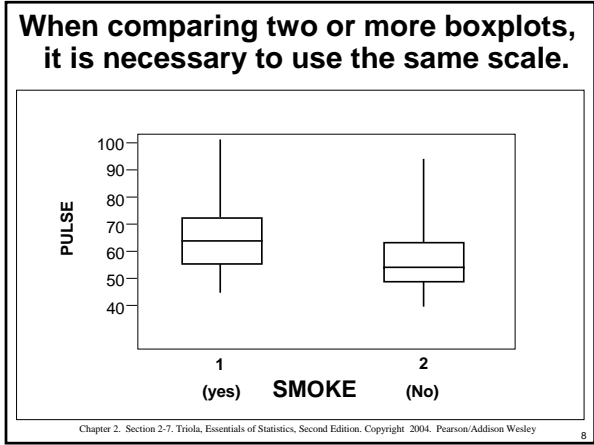
---

---

---

---

---




---

---

---

---

---

---

---

---

**Exploring**

- ❖ **Measure of center:** mean, median, mode, midrange
- ❖ **Measure of variation:** Standard deviation, range, variance
- ❖ **Measures of relative standing:** minimum, median, maximum, and quartiles, z-scores
- ❖ **Unusual values:** outliers
- ❖ **Distribution:** histogram, stem-leaf plot, and boxplot

Chapter 2, Section 2-7, Triola, Essentials of Statistics, Second Edition, Copyright 2004, Pearson/Addison Wesley 9

---

---

---

---

---

---

---

---