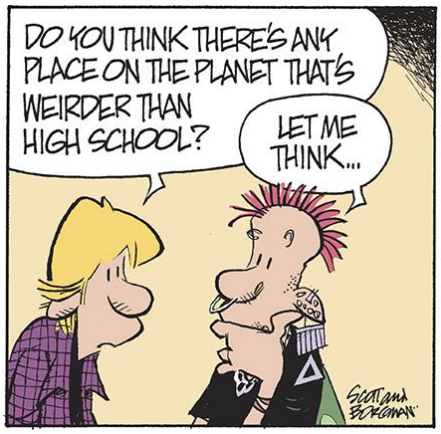
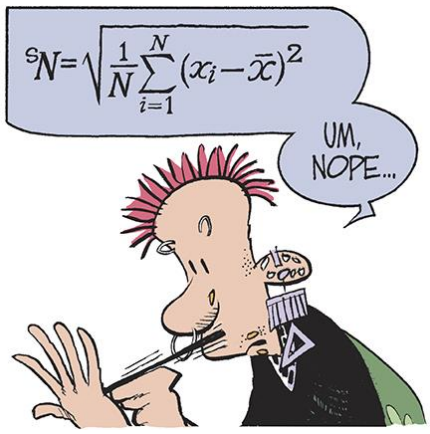
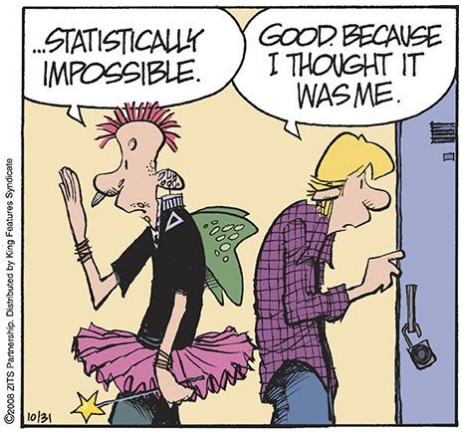
Chapter 1: Exploring Data

Key Vocabulary:

- individual
- variable
- frequency table
- relative frequency table
- distribution
- pie chart
- bar graph
- two-way table
- marginal distributions
- conditional distributions
- side-by-side bar graph
- association
- dotplot
- stemplot
- histogram
- SOCS
- outlier
- symmetric
- Σ
- $\bar{x}$
- spread
- variability
- median
- quartiles
- Q_1, Q_3
- IQR
- five-number summary
- minimum
- maximum
- boxplot
- resistant
- standard deviation
- variance







Data Analysis: Making Sense of Data (pp.2-6)

1. *Individuals* are…

2. A *variable* is …

3. When you first meet a new data set, ask yourself:

    - Who…
    - What…
    - Why, When, Where and How…

4. Explain the difference between a *categorical* variable and a *quantitative* variable. Give an example of each.

5. Give an example of a categorical variable that has number values.

6. Define *distribution*:

7. What are the four steps to *exploring data*?

    - Begin by….
    - Study relationships…
    - Start with a …
    - Then add…

8. Answer the two questions for the *Check Your Understanding* on page 5*:*

9. Define *inference.*

1.1 Analyzing Categorical Data (pp.8-22)

1. A *frequency* table displays…

2. A *relative frequency table* displays…

3. What type of data are *pie charts* and *bar graphs* used for?

4. *Categories* in a bar graph are represented by _____ and the *bar heights* give the category _____.

5. What is a *two-way table*?

6. Define *marginal distribution*.

7. What are the two steps in examining a marginal distribution?

8. Answer the two questions for the *Check Your Understanding* on page 14.

9. What is a *conditional distribution*? Give an example demonstrating how to calculate one set of conditional distributions in a two-way table.

10. What is the purpose of using a *segmented bar graph*?

11. Answer question one for the *Check Your Understanding* on page 17.

12. Describe the four steps to organizing a statistical problem:

    - State…
    - Plan…
    - Do…
    - Conclude…

13. Explain what it meant by an *association* between two variables.

1.2 Analyzing Categorical Data (pp.27-42)

1. What is a *dotplot*? Draw an example.

2. When examining a distribution, you can describe the overall pattern by its

S_____ O_____ C_____ S_____

3. If a distribution is *symmetric*, what does it look like?

4. If a distribution is *skewed to the right*, what does it look like?

5. If a distribution is *skewed to the left*, what does it look like?

6. Describe and illustrate the following distributions:

    a. Unimodal

    b. Bimodal

    c. Multimodal

7. Answer questions 1-4 for the *Check Your Understanding* on page 31.

8. How are a *stemplot* and a *histogram* similar?

9. When is it beneficial to *split the stems* on a stemplot?

10. When is it best to use a *back-to-back stemplot*?

11. List the three steps involved in making a histogram.

12. Why is it advantageous to use a relative frequency histogram instead of a frequency histogram?

13. Answer questions 2-4 for the *Check Your Understanding* on page 35.

1.3 Analyzing Categorical Data (pp.50-67)

1. What is the most common *measure of center*?

2. Explain how to calculate the *mean,* $\bar{x}$.

3. What is the meaning of $\sum$?

4. Explain the difference between $\bar{x}$ and μ.

5. Define *resistant measure*.

6. Explain why the mean is not a resistant measure of center.

7. What is the *median* of a distribution? Explain how to find it.

8. Explain why the median is a resistant measure of center?

9. How does the shape of the distribution affect the mean and median?

10. What is the *range*?

11. Is the range a resistant measure of spread? Explain.

12. How do you find *first quartile* Q_1 and *third quartile* Q_3?

13. What is the *Interquartile Range* (IQR)?

14. Is the IQR and the quartiles a resistant measure of spread? Explain.

15. How is the IQR used to identify *outliers*?

16. What is the *five-number summary* of a distribution?

17. Explain how to use the five-number summary to make a *boxplot*.

18. What does the *standard deviation* measure? How do we calculate it?

19. What is the relationship between *variance* and *standard deviation*?

20. What are the *properties* of the standard deviation as explained on page 64?

21. How should one go about choosing measures of center and spread?