HOMEWORK 1   Due: next class   1/27

1. Explain how a sample is related to a population.

    A sample is a sub-group of the population from which data is collected.

2. a.   Explain what is meant by the *distribution* of a categorical variable.

    The distribution of a categorical variable lists all of the values the variable takes and how often it takes each of these values.

    b.   List two ways that this distribution can be displayed visually.

    With a bar chart or a pie chart.

3. For each of the following variables, indicate with Q or C whether it is a quantitative variable or a categorical variable.
    a. the color of a M&M candy   **C**
    b. the weight of an airplane   **Q**
    c. the life expectancy of a nation   **Q**
    d. how many miles a person walks in one day   **Q**
    e. the age of a mother when her first baby born   **Q**
    f. whether or not a student eats breakfast or doesn't   **C**
    g. the length of a snake   **Q**
    h. whether or not a car has automatic transmission or manual transmission/stick shift   **C**
    i. the number of calories in a pint vanilla ice cream   **Q**
    j. whether or not a baby tests HIV-positive   **C**
    k. the running time of a Tom Cruise movie   **Q**
    l. whether or not a state's name consists of one word   **C**
    m. the diameter of a pizza   **Q**
    n. the number of dogs an animal shelter has   **Q**
    o. the height of a sequoia tree   **Q**
    p. the color of a bottle of wine   **C**
    q. the number of books a person owns   **Q**
    r. the race of a person   **C**
   s.

4.   Consider the following pie chart:
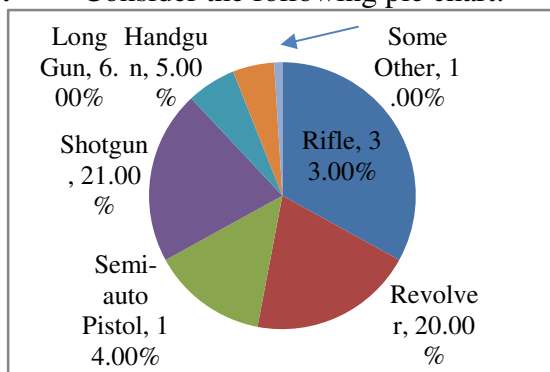
a. What is the variable described in the pie chart?
Type of firearm (gun).
b.   Summarize what the pie chart shows.
Four different gun types are prevalent—rifles, shotguns, revolvers and semi-automatic pistols, with rifles being the most common. Handguns and long guns are much rarer.

5. Consider the following bar graph: Pressure from schoolwork

number of students
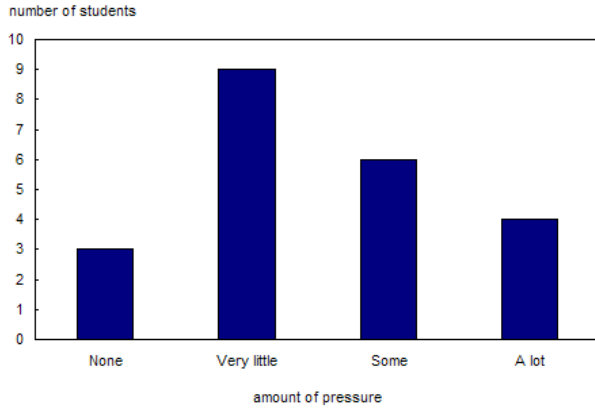
amount of pressure

a. What is the variable described in the bar graph?

Amount of pressure from schoolwork

b. What percent of students feel no pressure from schoolwork? **3/22 = 13.63%**

6. In 2007, 5488 people were killed while working. Here is a breakdown of causes: transportation: 2234; contact with objects or equipment: 916; assaults or violent acts: 839; falls: 835; exposure to harmful substances or harmful environment: 488; fires or explosions:151; others: 25. (The data are from the Bureau of Labor Statistics.) Construct a bar graph.
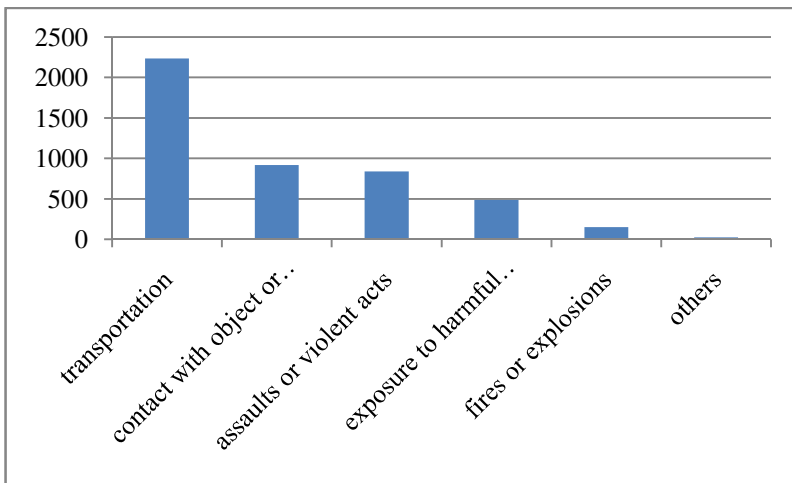
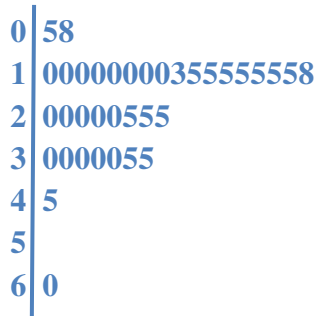**7.** The graph below came from the USA Today Snapshots: Commuting Time.

List two things that are wrong with this graph.

(1) The lengths of the dotted lines are not proportional to the percentages. If they were, the line for 32% would be about three times as long as the line for 11% (since 32% ≈ 3 × 11%) and the line for 50% would be between four and five times as long as the 11% line.
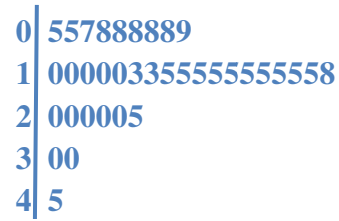
(2) The three percentages do not add up to 100% even though they include all possible commute times.

8. Create a **stemplot** for the time (in minutes) students in your class spend for a typical shower using the class data sheet (question 14 on the questionnaire).
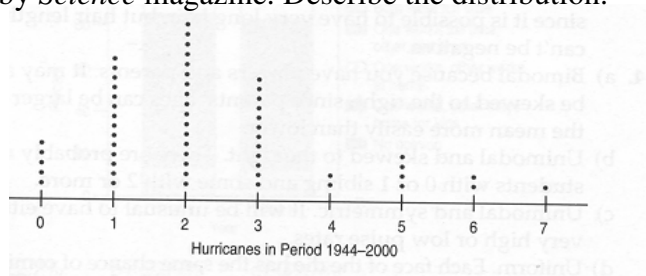
12pm class:

```
0 | 58
1 | 0000000355555558
2 | 00000555
3 | 0000055
4 | 5
5 |
6 | 0
```

1pm class

```
0 | 557888889
1 | 000003355555555558
2 | 000005
3 | 00
4 | 5
```

9. The dotplot below shows the number of hurricanes that happened each year from 1944 through 2000 as reported by *Science* magazine. Describe the distribution.

Hurricanes in Period 1944–2000

The distribution of the number of hurricanes is unimodal and right-skewed. In most of the years between 1944 and 2000 there were 3 or less hurricanes each year. There were only 9 years in this period when 4 or more hurricanes happened in a year. The most frequent number of a year was 2 in this time period. There are no outliers in the distribution.

10. Would you expect the distributions of these variables to be **uniform**, **unimodal**, or **bimodal**? **Symmetric** or **skewed**?

a. Number of times each face of a die shows in 200 tosses
Uniform. Each face of a die has the same chance to occur.

b. Weights of newborn babies born in a particular hospital over the course of a year
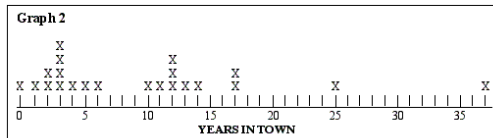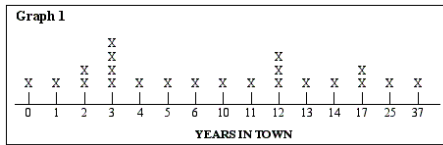**Unimodal, symmetric. I would expect a nice bell-shaped curve.**

c. The scores on a very easy statistics quiz
Unimodal, and skewed to the left. If the quiz was very easy, I expect that most of the students did very well (high scores), and a few did not do well (low scores.)

d. Number of siblings of the students in this class

11. A class of students recorded the number of years their families had each lived in their town. Here are two graphs that the students drew to summarize the data. Which graph gives a more accurate representation of the data? Why?

```
Graph 1
                    X
                    X
            X   X
            X   X                       X
            X   X                       X
    X   X   X   X   X   X   X   X   X   X   X   X   X   X   X
    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
    0   1   2   3   4   5   6   10  11  12  13  14  17  25  37
                        YEARS IN TOWN
```

```
Graph 2
        X
        X                   X
        X X             X
    X X X X X X     X X X X X       X
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
    0       5       10      15      20      25      30      35
                        YEARS IN TOWN
```

The second one. The first graph omits from the axis the values that not occur in the data set, which distorts the shape of the distribution. For example, a student whose family had lived in their town for 37 years has lived there *much* longer than one who has lived there 25 years; this is evident in Graph 2, but not in Graph 1. Graph 2 shows that the distribution is strongly right skewed, with at least one outlier.

12. For the graph below, of heights of singers in a large chorus, please write a complete description of the histogram. Be sure to comment on all the important features.

```
15 ┤
   │
10 ┤
   │
 5 ┤
   │
   └──────────────────────
      60      64      68
        Soprano:Heights
```
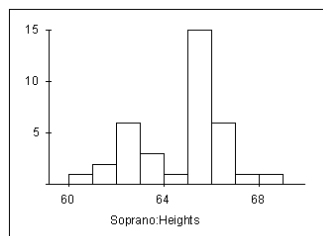
The most striking feature is that the distribution is bimodal, with many singers around 65 to 66 inches tall and also quite a few near 62 to 63 inches. The heights vary between 60 and 69 inches, and there are no outliers.