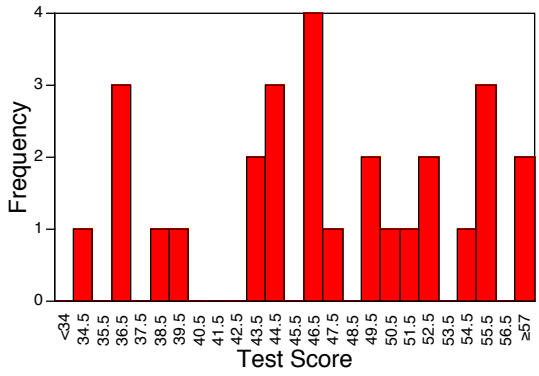
# Chapter 3-Displaying Data

3.1 Katz et al (1990) No Passage Group:




There is too little data to say very much about the shape of this distribution, but it certainly isn't looking normally distributed.

3.3 I would use stems of 3*, 3., 4*, 4. 5*, and 5. for this display. "m = 2" controls the number of steps in the plot.

```
> stem.leaf(score, m = 2)
1 | 2: represents 12
 leaf unit: 1
            n: 25
   1    3* | 4
   6    3. | 66689
  10    4* | 3344
  (6)   4. | 666799
   9    5* | 01224
   4    5. | 5557
```

3.5 Compared to those who read the passages:

a) Almost everyone who read the passages did better than the best person who did not read them. Certainly knowing what you are talking about is a good thing (though not always practiced).

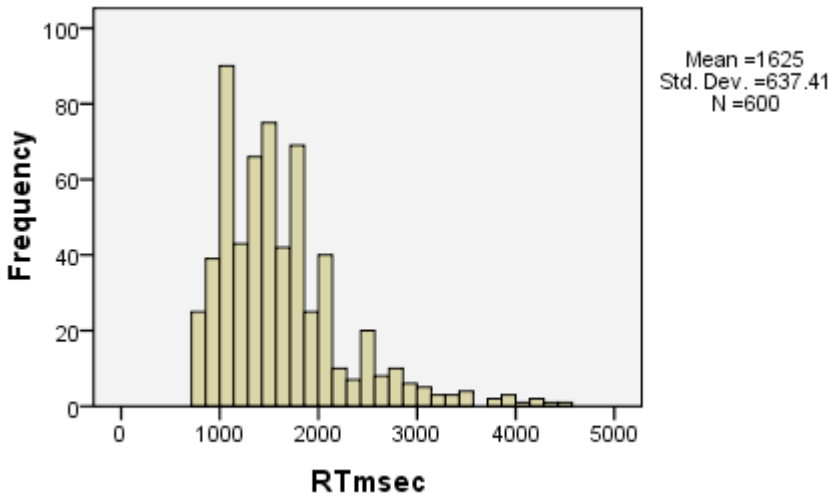
b)

```
> stem.leaf.backback(x = NoPassage, y = Passage)
--------------------------------------
  1 | 2: represents 12, leaf unit: 1
      NoPassage       Passage
--------------------------------------
   1          4|  3* |
   6      98666|  3. |
  10       4433|  4* |
  (6)    997666|  4. |
   9      42210|  5* |
   4       7555|  5. |5669        4
               |  6* |
               |  6. |66          6
               |  7* |11222233   (8)
               |  7. |5           3
               |  8* |
               |  8. |
               |  9* |13          2
               |  9. |
               | 10* |
--------------------------------------
n:            25       17
--------------------------------------
```

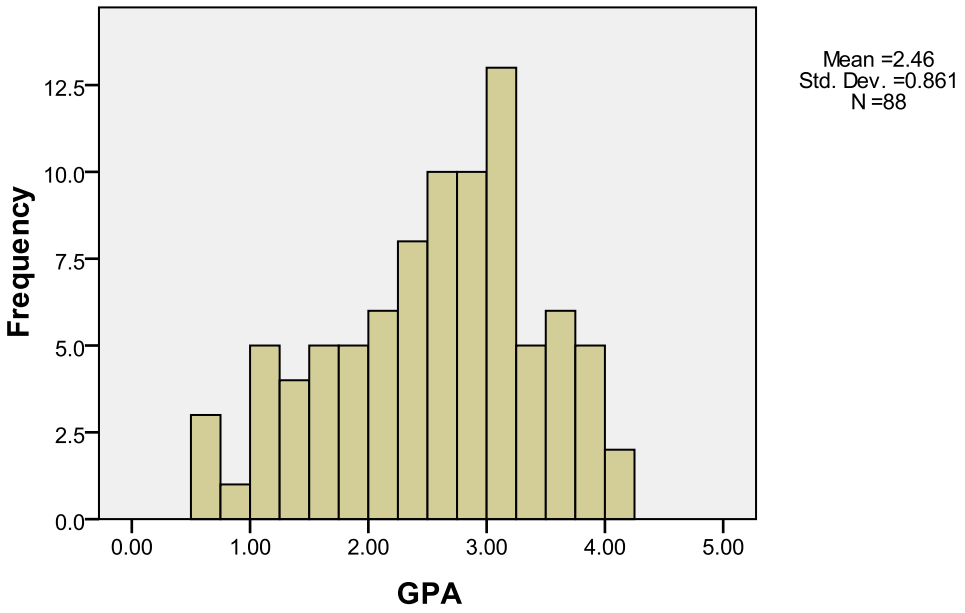
c) It is obvious that the two groups are very different in their performance. We would be worried if they weren't.

d) This is an Internet exercise with no fixed answer. That source is far more advanced than the students would be at this time, but I think that they should be able to read it if they just skip over what they don't understand.

3.7 The following is a plot (as a histogram) of reaction times collapsed across all variables.



3.9 Histogram of GPA scores




3.11 (1) Mexico has very many young people and very few old people, while Spain has a more even distribution. (2) The difference between males and females is more pronounced at most ages in Spain than it is in Mexico. (3) You can see the high infant mortality rate in Mexico.

3.13 The distribution of those whose attendance is poor is far more spread out than the distribution of normal attendees. This would be expected because a few very good students can score well on tests even when they don't attend, but most of the poor attenders are generally poor students who would score badly no matter what. The difference between the average grades of these two groups is obvious.

3.15 As the degree of rotation increases, the distribution of reaction time scores appears to move from left to right—which is also an increase.

I think it is a good idea to get the students to really think through this problem, rather than to just take the answer as given. It is important to make them really see that looking at data can lead to conclusions to scientific questions, even without formal statistical tests. Many students have a hard time seeing the relationship between data and a question they would like to ask. (Probably many older adults do as well.)

3.17 The data points are probably not independent in that data set. As time went on, there would be changes in the subject's performance. At first he might get better with practice, but then fatigue would start to set in. Since the data are given in the order in which they

were collected, at least within each condition, data nearer in time should be more similar than data farther apart in time.

3.19 The amount of shock that a subject delivers to a white participant does not vary as a function of whether or not that subject has been insulted by the experimenter. However, the black participants do suffer more shocks when the subject has been insulted.

3.21 Wikipedia gives an excellent set of data on HIV/AIDS prevalence at http://en.wikipedia.org/wiki/List_of_countries_by_HIV/AIDS_adult_prevalence_rate

3.23 *R* code to reproduce figures above

```
### Households headed by women
percent <- c(.085, .088, .102, .108, .117, .117, .116, .117, .118)
year <- c(1960, 1970, 1975, 1980, 1985, 1987, 1988, 1989, 1990)
famsize <- c(3.33, 3.14, 2.94, 2.76, 2.69, 2.66, 2.64, 2.62, 2.63)
par(mfrow = c(2,1)
plot(percent ~ year, type = "l", ylim = c(.08, .12), ylab = "Percentage", col =
"red", lwd = 3)
plot(famsize ~ year, type = “l”, ylim = c(2.6, 3.4) ylab = “Family Size”, col =
“blue”, lwd = 3)
```

3.25 Down Syndrome There is a tremendous increase in Down syndrome in children botn to older mothers. This increase doesn't really take off until mothers are in their 40s, but with parents delaying having children, this is a potential problem.

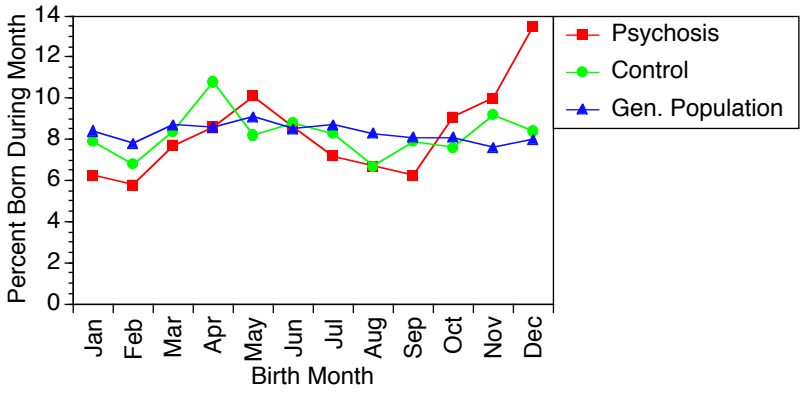
Interestingly, the New York Times, on August 22, 2012, reported that there is evidence that older *fathers* are more likely to have children who suffer from autism or schizophrenia. The cause is apparently random mutations that accumulate over time. The age of the mother appears to be unrelated to differences in either disorder.

3.27 Data on birth month and psychiatric diagnosis:

a) You would have to transform the data to percentages /month to put them on the same scale.

b) All three sets of data can be plotted together.

c)

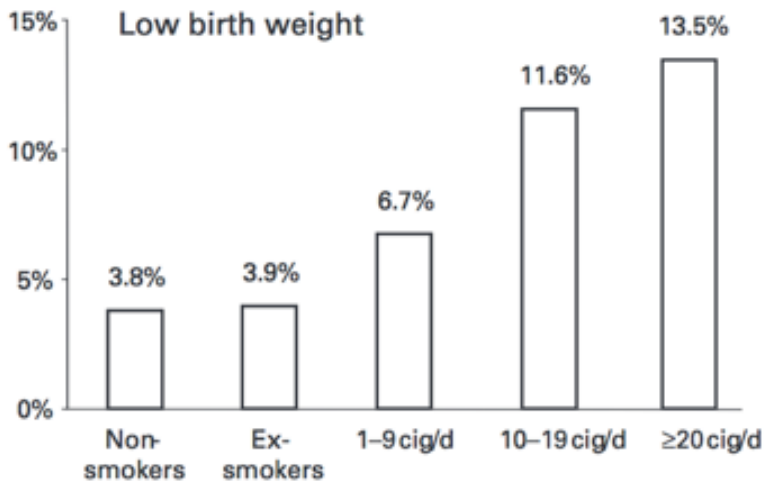



d) There is a lot more variability in the psychosis group, but they have a much smaller sample than either of the other conditions, so that would be expected. The psychosis group is higher than the general population in the last three months of the year, but lower in the first three. I would need a statistical test (to be discussed in Chapter 19) to feel comfortable about saying that there are seasonal variations that are associated with psychosis.

e) The control group allows us to compare two groups of people who have been referred for some sort of treatment. They would allow us to eliminate referral as an explanation. Notice that they stick closer to the general population than do the psychosis group, but the sample size for the control group is quite large.

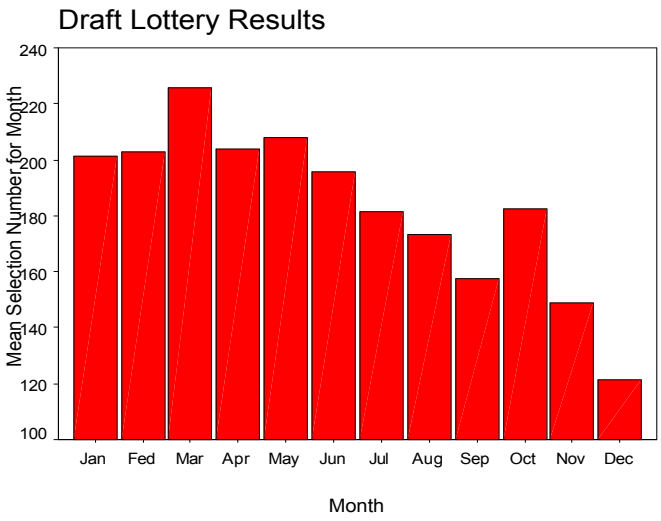
f) The most that I can conclude was already stated in d)

3.29 Birthweight as a function of smoking behavior




3.31 Life expectancy for white and black females.
White females have a longer life expectancy than black females, but the difference has shrunk considerably since 1920, though recent changes have been modest.

3.33 Plot of draft lottery data




There is a clear drop in the mean lottery number across the year. The officials in charge hypothesized that there was probably insufficient mixing of the batches of numbers, which were put into the barrel a month at a time.

3.35 back-to-back stemplots

```
### Creating back to back stem-and-leaf displays.  Not covered in text.
library(aplpack)
```

```
grades <- read.table("http://www.uvm.edu/~dhowell/methods9/DataFiles/Fig2-9.dat", header = TRUE)
attach(grades)
males <- Grade[Sex == 1]
females <- Grade[Sex == 2]
stem.leaf.backback(males,females, m = 2)  # m controls bin size
```