Class Notes

this time: histograms, stem-plots
next time: measures of center and spread
-reading: DD ch 3
-PPP ch 4
-permission codes are currently available
-reader will be ready Monday (in library this weekend)

relationships: some are causal
ex: smoking causes cancer and heart disease but others are not
ex: drinking soda pop causes polio

<table>
<thead>
<tr>
<th>Year</th>
<th>Time of Year</th>
<th>Amount of Soda Consumed</th>
<th># of Polio Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summer</td>
<td>high</td>
<td>high</td>
<td>high</td>
</tr>
<tr>
<td>Fall</td>
<td>medium</td>
<td>medium</td>
<td>medium</td>
</tr>
<tr>
<td>Winter</td>
<td>low</td>
<td>low</td>
<td>low</td>
</tr>
<tr>
<td>Spring</td>
<td>medium</td>
<td>medium</td>
<td>Medium</td>
</tr>
</tbody>
</table>

*Here there was an "association," but it was non-causal.

- decision making is about "predicting the future" under different sets of possibilities and choosing the desired one.
  - ex: [Iraq example in first handout]
- Statistics helps deal with uncertainty associated with real decision making, but data must be gathered
- data gathering: 1) experimental design
  2) sample surveys
- Descriptive measures (ex: histograms):
  1) Graphical (hist)
  2) Numerical (mean, median, mode, std. dev.)
* Factual data
- Counter-factual data: "What if...?"
  - Compliments the factual data by examining the other possibilities

[Look at page 11 of handout]
"Tracking the Markets" Stocks:

![Graph of stock index over time]

< "time trend plot" or "time series plot" tracks something over time

[Page 12]
"Associated" but not necessarily causal

[Page 16] - Case Study 1
"The Chicago Engineer Exam"
- High scores (top 15) will get jobs

| Variable | Subjects | Variables: thing that may vary per subject
|----------|---------|-----------------------------------------|
|          | Subjects: people or things with measured variables

The sample size, \( n \), is 223. The data is sorted from smallest to largest, but not in order recorded.

<table>
<thead>
<tr>
<th>Value</th>
<th>Raw freq.</th>
<th>( \text{rel. freq.} = \left(\frac{\text{raw freq.}}{# \text{ of subjects}}\right) \times 100% )</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>1</td>
<td>( \left(\frac{1}{223}\right) \times 100% = 0.45% )</td>
</tr>
<tr>
<td>27</td>
<td>4</td>
<td>( \left(\frac{4}{223}\right) \times 100% = \ldots )</td>
</tr>
<tr>
<td>95</td>
<td>2</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>223</td>
<td></td>
<td>( \ldots )</td>
</tr>
</tbody>
</table>

There was a gap in the right tail of the distribution. The gap and the fact that there were exactly 15 people in the right tail proved cheating.

[page 25 of reader]