Today: Organizing tabular data, forming indices, more about simulation.

Recall from Monday …

Birth and Death Rates, per 1000 population, for 69 Countries
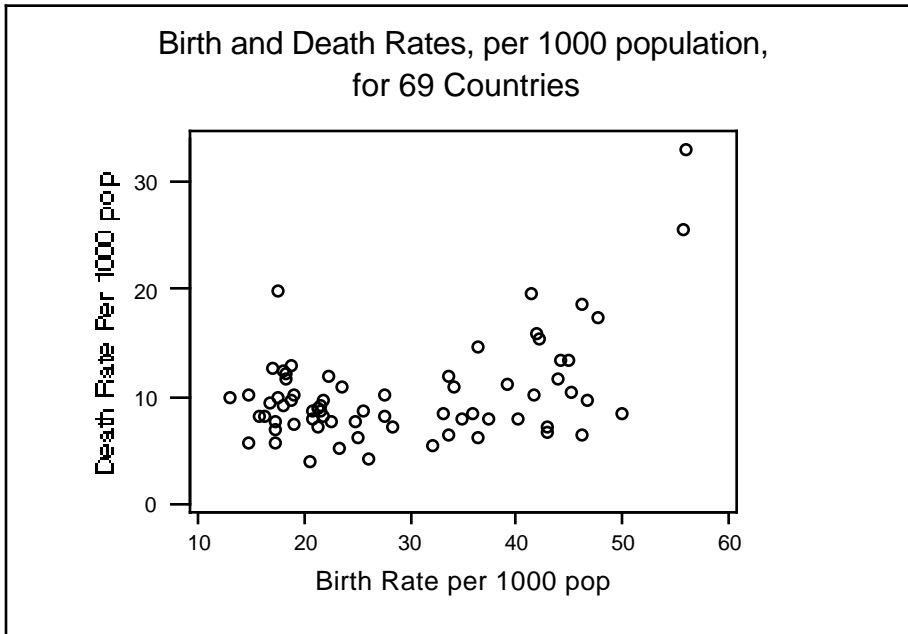
The missing information about country name is important! Shown here is just a few lines of the complete data set.

Row	birth-rt	death-rt	country	Contin.
1	36.4	14.6	alg	AFRICA
2	37.3	8.0	con	AFRICA
13	32.1	5.5	hkg	ASIA
14	20.9	8.8	ind	ASIA
27	24.8	7.8	can	AMERICA
28	49.9	8.5	cra	AMERICA
45	18.8	12.8	aus	EUROPE
46	17.1	12.7	bel	EUROPE
68	21.6	8.7	ast	OCEANIA
69	25.5	8.8	nzl	OCEANIA

But is this a good way to arrange the table? Note that sorting the rows often helps.
Sort by birth rate, or by death rate, or by ratio birth/death rate, for example.
The general idea is that if a feature of a display is arbitrary, it may sometimes be
re-organized to advantage.

Last time it was suggested that time series of birth and death rates would be of interest. A
simple web search produced this for the last 25 years in 29 countries.
http://www.infoplease.com/ipa/A0004395.html
Amongst other things, this shows that the data displayed last day is 25 years old!
So now we will use the data that brings us up to date.

What might we do to help use view this data: tabular? graphical?

Sort by?

Consider "pre-processing" the data. Construction of indices which measure some aspect
of interest.
For example:
Take ratio (birth rate)/(death rate), and compute the change - then sort by change over the
27 year period? Think about how you might summarize this data.

Homework: (not to hand in, yet)
1. Propose a method for numerical summary of this data. By eye-balling the table,
anticipate what your summary would show, and express this in words.
2. Read Tanur pp 227-238 (Economic Indicators) and pp 161-169(Food Study).

More about Simulation:

Sequence of coin tosses:

Model for:

Sequence of sport matches
Sequence of goals scored for or against a given team during the season.
Selecting a committee from a poopulation of 50% male, 50% female.
Stock Market Daily Directions

```
MTB > random 10 c1-c5;
SUBC> bern .5.
MTB > prin c1-c5
```

Data Display

Row	C1	C2	C3	C4	C5
1	0	1	0	1	1
2	0	0	1	1	1
3	0	1	0	0	0
4	1	1	1	0	1
5	0	0	0	0	0
6	0	0	0	0	1
7	0	1	0	1	1
8	0	0	1	0	0
9	0	0	0	1	0
10	1	1	0	1	1

But what about
Days on which a fatal accident occurs in Greater Vancouver?
Does not work since the rate is not .5 per day.

Need to consider using biased coin. Or computer.

```
MTB > random 10 c1-c5;
SUBC> bern .1.
MTB > prin c1-c5
```
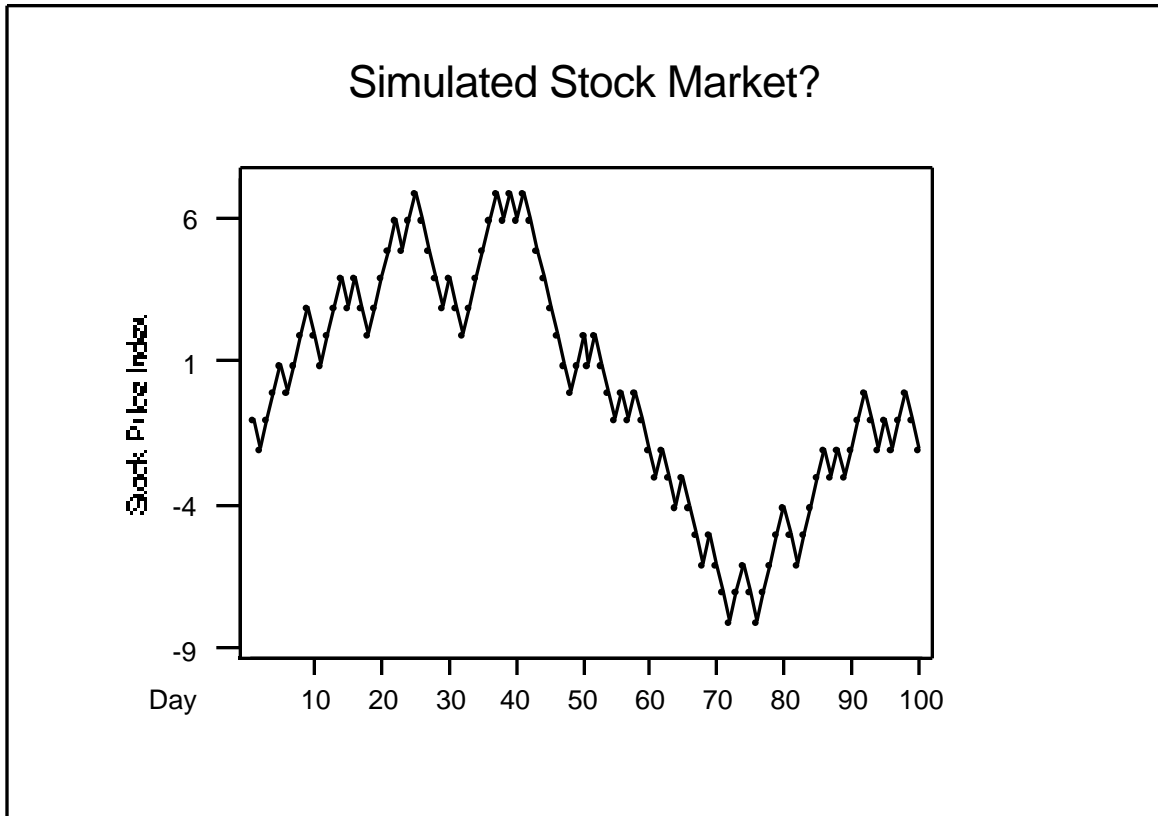
Data Display

Row	C1	C2	C3	C4	C5
1	1	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	1	0	0	0
5	0	0	0	0	0
6	0	0	0	0	0
7	0	0	1	0	0
8	0	0	0	0	0
9	0	0	0	0	0
10	0	0	0	0	0

Why would this simulation be useful to the insurance company?

Now consider a simulated stock market daily price …

What graphical displays would be useful for viewing a sequence of coin-tosses?

Simulated Stock Market?

Fairly Realistic, BUT moves each day can only be +1 or –1.
Need to allow more flexibility in size of move.   Suggestions?