Ch2: Exploring Data: Charts

- **HW1 due Thu 10pm**
- **Download and open "SportsShoes.xls"**

13 Sep 2011
BUSI275
Dr. Sean Ho

TRINITY WESTERN UNIVERSITY

Outline for today

- Exploring data with charts:
  - Tallying frequency distributions
  - Gentle intro to Excel, array formulas
  - Qualitative vars: bar, pie
    - Multiple vars: crosstabs, clustered bar
  - Quantitative vars: histogram, line
    - Multiple vars: scatter

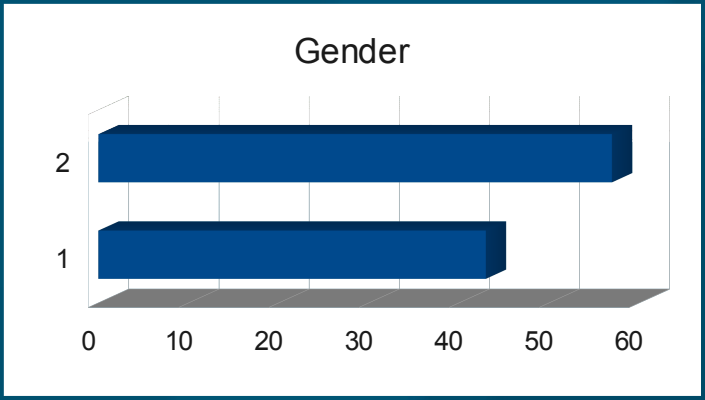
Frequency distributions

- How frequently each value of a variable appears in the dataset (either pop or sample)
- Data usually come as 1 row = 1 participant:

1	Homeroom #	First Name	Last Name	Payment	T-Shirt Color	T-Shirt Size
3	105	Esther	Yaron	7-Oct	Dark Red	Small
4	105	Melissa	White	7-Oct	Heather Grey	Small
5	220-A	Christopher	Peyton-Gomez	Pending	White	Small
6	220-A	Brigid	Ellison	Pending	Dark Red	Small
7	220-B	Windy	Shaw	7-Oct	Heather Grey	Small
8	220-B	Malik	Reynolds	7-Oct	Heather Grey	Small
9	220-B	Michael	Lazar	14-Oct	White	Small
10	105	Christiana	Chen	5-Oct	Dark Red	Medium
11	105	Sidney	Kelly	11-Oct	Dark Red	Medium
12	105	Nathan	Albee	13-Oct	Heather Grey	Medium
13	110	Matt	Benson	11-Oct	White	Medium
14	110	Gabriel	Del Toro	13-Oct	White	Medium
15	135	Chantal	Weller	15-Oct	White	Medium

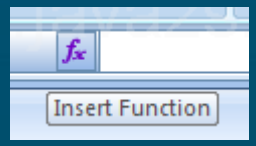
- Compute by tallying up how many occurrences of each value exist in the data:
  - e.g., for "T-Shirt Size" *(level of meas?)*:
    Small: 10; Medium: 20; Large: 15

Excel: freq. dist. & bar chart



- Dataset: SportsShoes.xls
  - Add new sheet: "Charts"
- Frequency distribution:
  - Enter poss. values (Gender coded as 1, 2)
  - FREQUENCY() array formula
  - Relative Frequencies (%): divide by total
    - Use '$' for abs. cell ref.; format as %
- Bar chart:
  - Insert > Bar > 2D > Select Data:
  - Data: freqs; Cat. Axis Labels: values

Excel array formulas

- Regular formulas (functions) take cells or cell ranges as input and produce a single output
  - Array formulas output to a range of cells
- Highlight the range where output will go
- Enter the formula: 

  - =FREQUENCY()
  - Data: highlight Data!M3:M102
  - Bins (values): highlight cells with “1”,”2”
- Don't hit OK yet! Use Ctrl-Shift-Enter instead to indicate it is an array formula

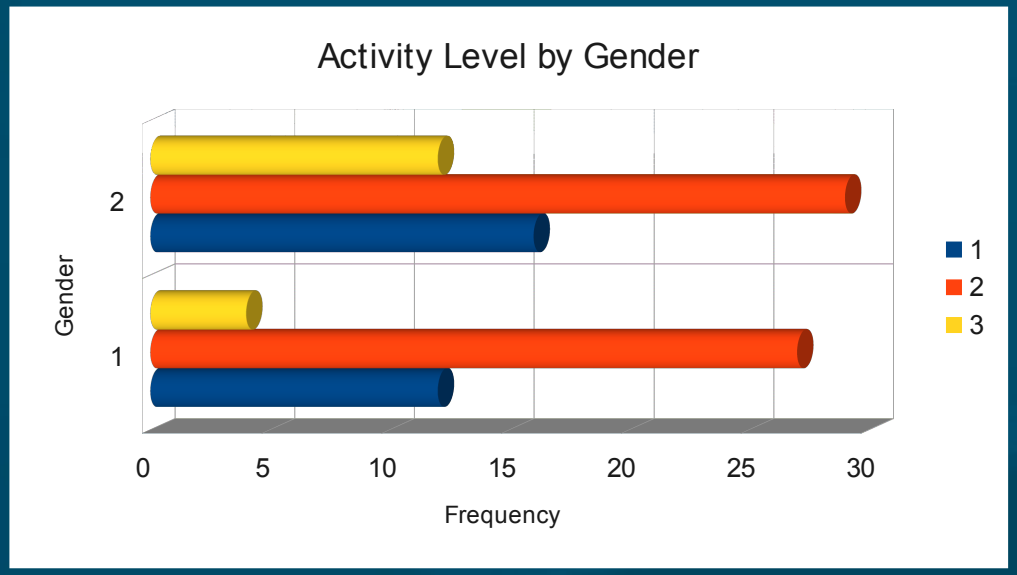
Multiple vars: crosstabs

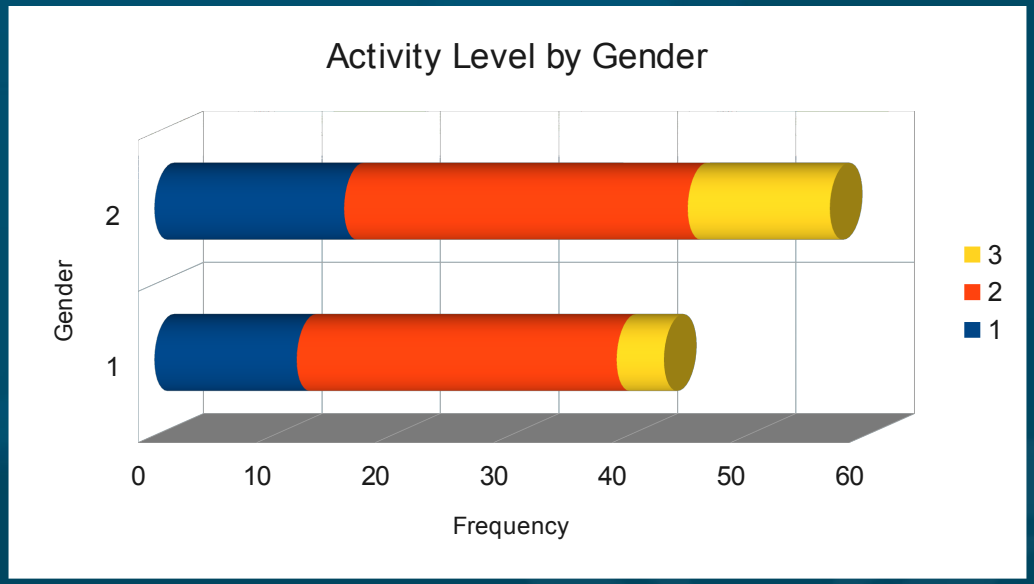
- Consider all combinations of values:
  - e.g., Gender: 1 or 2; Activity: 1, 2, 3
    so there are 6 combos of (Gender, Act)
- Cross-tabulations (Pivot Tables, Joint freq. dist):
  - Insert > Pivot Table
  - Select Range: L2:M102
  - Row Labels: Gender
  - Col Labels: Activity
  - Values: either
  - Summarize By: Count

Count - Activity	Activity			
Gender	1	2	3	**Total Result**
1	12	27	4	**43**
2	16	29	12	**57**
Total Result	**28**	**56**	**16**	**100**

Multiple vars: clustered bars

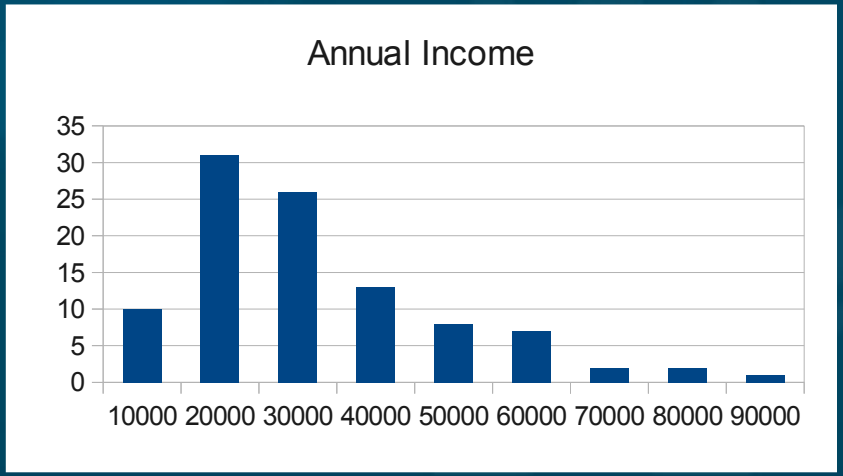
- If one of the nominal variables only has a few possible values (categories), then
- We can use clustered or stacked bar charts:





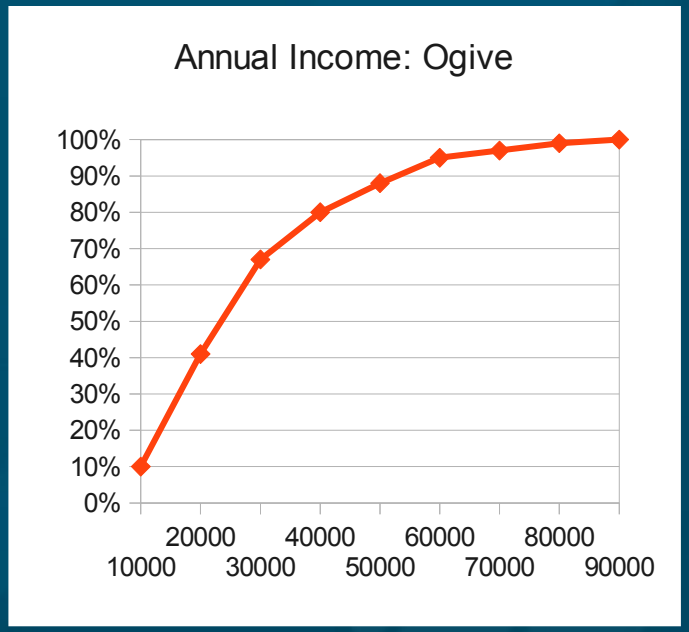
Quantitative vars: histograms

- For quantitative vars (scale, ratio), must group data into classes
  - e.g., length: 0-10cm, 10-20cm, 20-30cm... (class width is 10cm)
  - Specify class boundaries: 10, 20, 30, ...
- How many classes? for sample size of n, use k classes, where $2^k \geq n$
- Can use FREQUENCY() w/ column chart, or
- Data > Data Analysis > Histogram



Cumulative distrib.: ogive

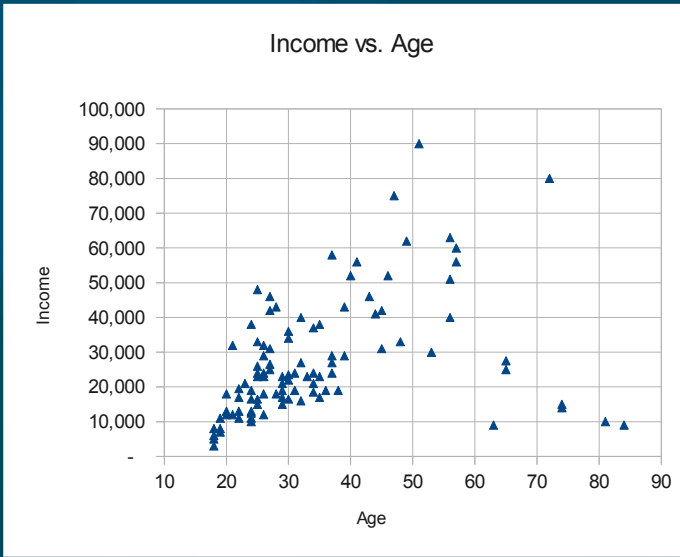
- The ogive is a curve showing the cumulative distribution on a variable:
  - Frequency of values equal to or less than a given value
- Compute cumul. freqs.
- Insert > Line w/Markers



- Pareto chart is an ogive on a nominal var, with bins sorted by decreasing frequency
  - Sort > Sort by: freq > Order: Large to small

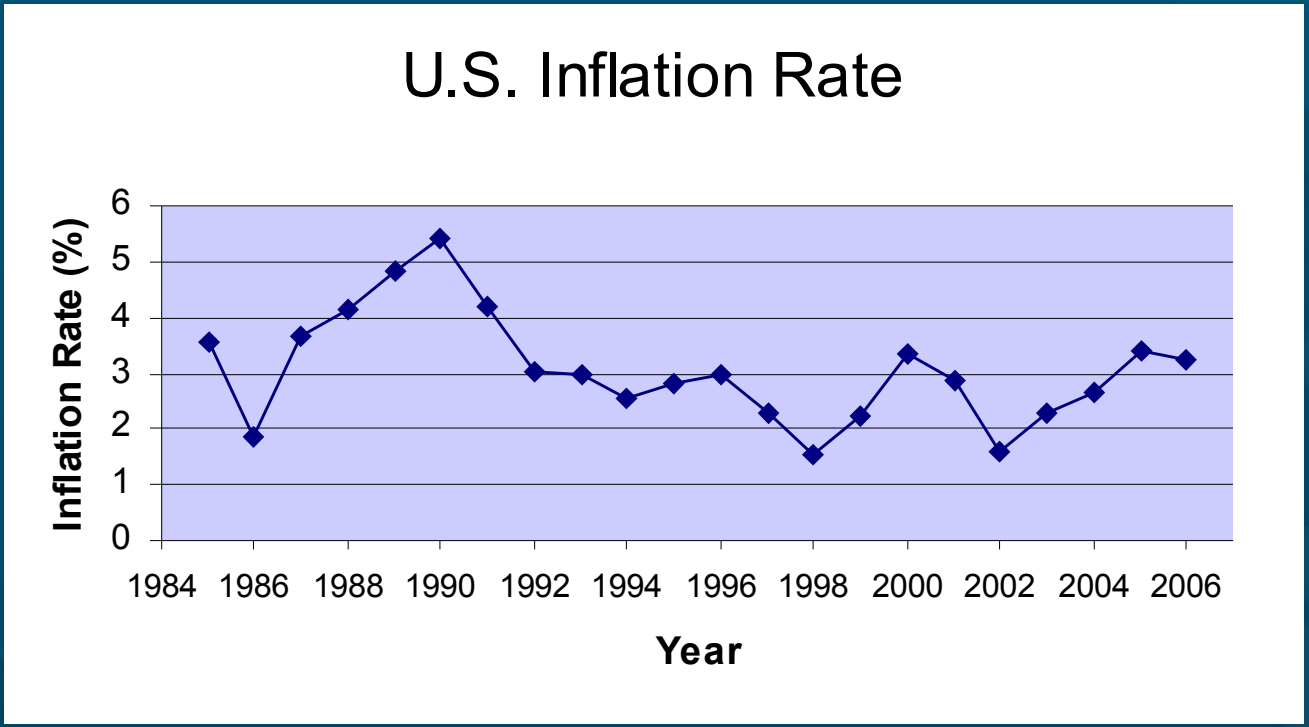
2 quant. vars: scatterplot

- Each participant in the dataset is plotted as a point on a 2D graph
  - (x,y) coordinates are that participant's observed values on the two variables
- Insert > XY Scatter
- If more than 2 vars, then either
  - 3D scatter (hard to see), or
  - Match up all pairs: matrix scatter



Time series: line graph

- Think of time as another variable
  - Horizontal axis is time
- Insert > Line > Line



TODO

- HW1 (ch1-2): due this Thu 15Sep at 10pm
  - Format as a clear, neat document
  - Also upload your Excel spreadsheet
  - HWs are to be individual work
- Get to know your classmates and form teams
  - Email me when you know your team
  - You can come up with a good name, too
- Discuss topics/variables you are interested in
  - Find existing data, or gather your own?