

# ***How to Lie With Statistics - by Darrell Huff***

**Chapter 1 - *The Sample with the Built-in Bias:*** We find that polls (while informative) are they bias?

**Chapter 2 - *The Well-Chosen Average:*** The word "average" is it mean, median, or mode)

**Chapter 3 - *The Little Figures That Are Not There:*** Independent laboratory says:  
"Science proves that tossed pennies come up heads 80 per cent of the time."

**Chapter 4 - *Much Ado about Practically Nothing:***

**Chapter 5 - *The Gee-Whiz Graph:*** Distorted graphs (choices of scales and origins).

**Chapter 6 - *The One-Dimensional Picture:*** Distorted picture graphs (height vs.. area).

**Chapter 7 - *The Semi attached Figure:*** The reported data may even be irrelevant.  
This mouthwash kills germs (but maybe not cold germs in the mouth).

**Chapter 8 - *Post Hoc Rides Again:*** Cause and effect, or some common cause, or coincidence, or what?  
"There are two clocks which keep perfect time. When "A" points to the hour. "B" strikes. Did "A" cause "B" to strike?

**Chapter 9 - *How to Statistic late:*** Lying with statistics = dishonesty or incompetence?

**Chapter 10 - *How to Talk Back to a Statistic:*** How to detect lying ? Ask yourself does it make sense?

Well Chosen Average

# **Explanation of Neighborhood Example**

## **Scenario 1: The "Upper Class" Neighborhood**

So just how is it possible that one person can claim the average income in the neighborhood is \$150,000 while another can claim it is only \$35,000?

Both are "honest" people who are using legitimate statistics, but they are using different definitions for "average."

### **Neighborhood Incomes**

Garnett \$1,000,000

Jackson \$225,000

Nelson \$80,000

Olson \$60,000

Smith \$40,000

Jones \$30,000

Howard \$20,000

Johnson \$20,000

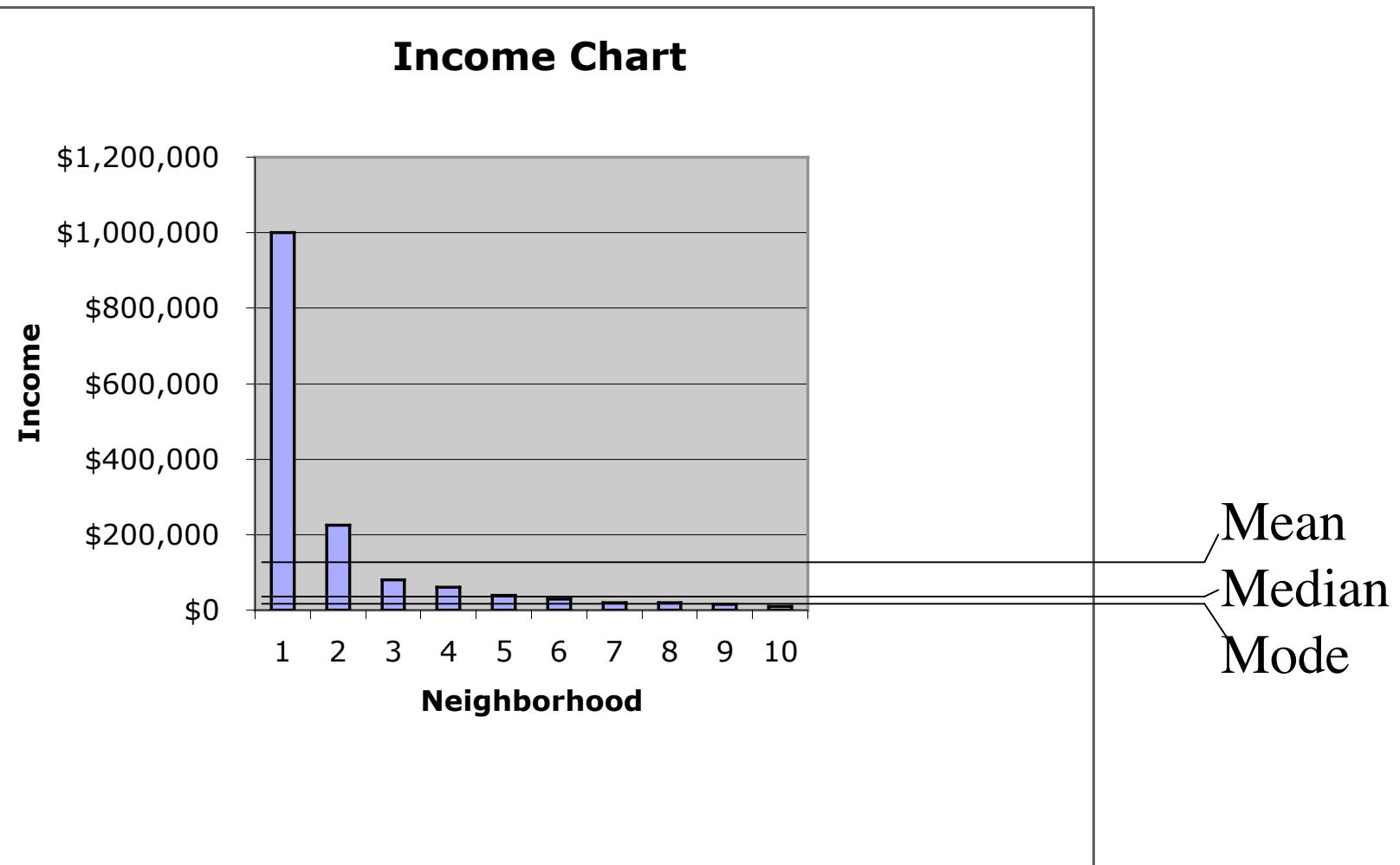
Hamlin \$15,000

D'Amario \$10,000

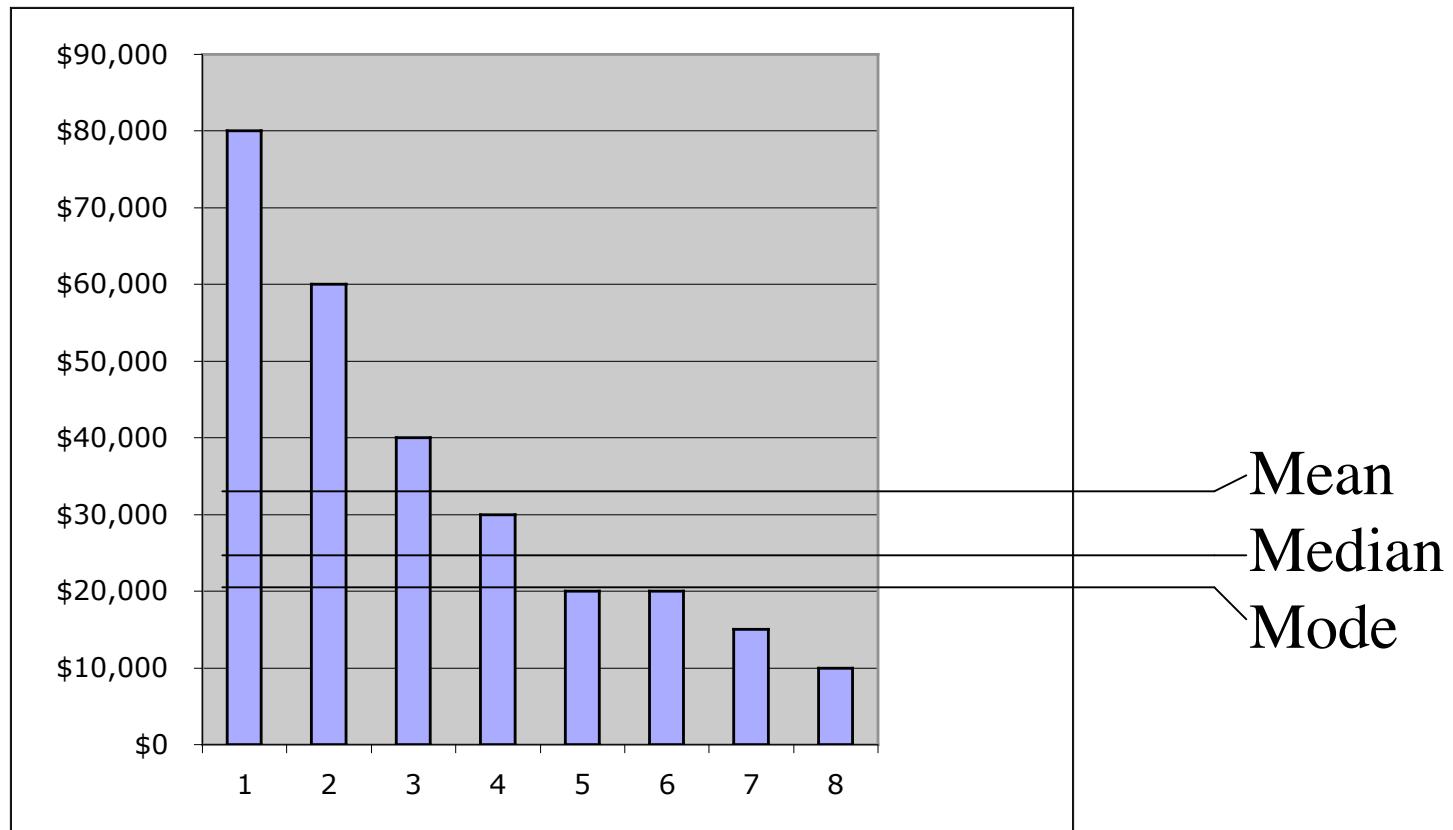
**Mean \$150,000**

**Median \$35,000**

**Mode \$20,000**



## Income Chart - revisited



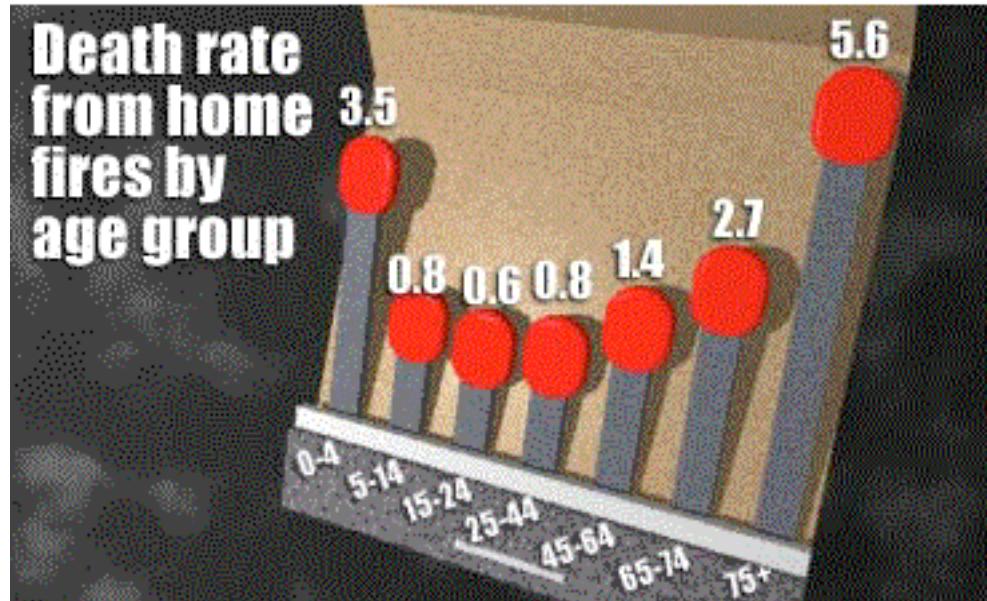
The Garnett & Jackson households are quite wealthy (outlier)

If they did not live here; the mean, median, and mode would be \$34,375, \$25,000, & \$20,000 respectively).

### Lessons to learn:

1. The word "average" is meaningless unless you know whether it refers to the "mean", "median", or "mode."
2. A few outliers can greatly skew the summary statistics. The mean is usually more sensitive to outliers than the median or mode.

Gee Whiz Graph



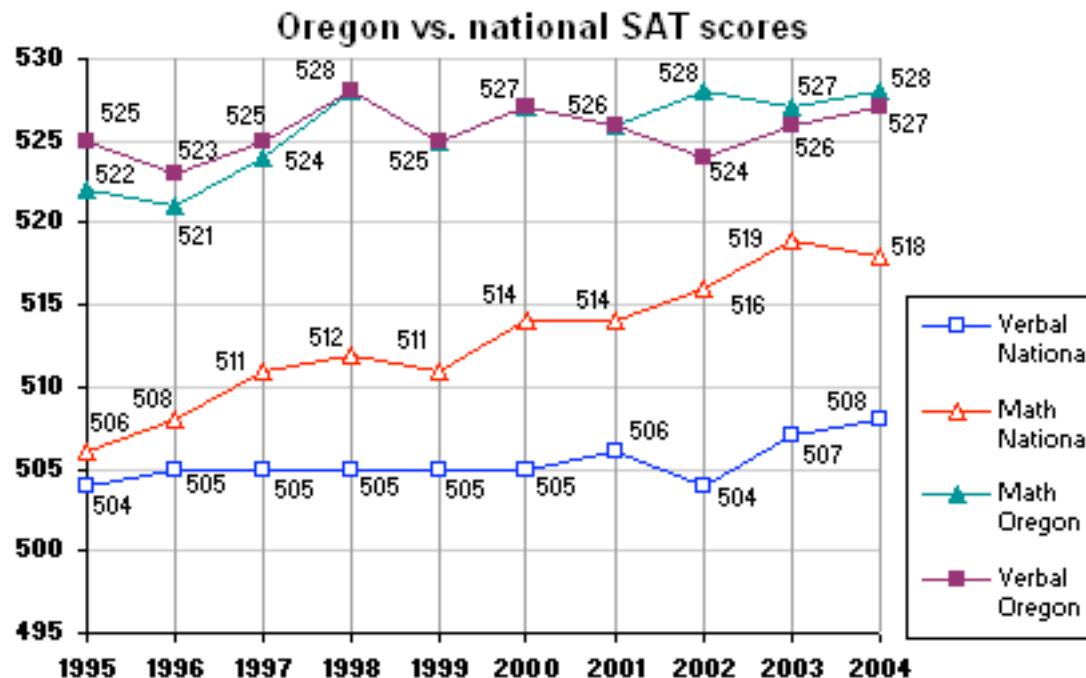
Source: National Safety Council

What does this information tell us?

Notice that the age ranges are not constant.

What relevance do the size of the matches have?

## Oregon SAT scores again second in nation



Source: Oregon School Board Association 2204

Look at the reports from  
the College Boards

Can you create an accurate graph?

"Oregon's SAT class of 2004 again scored second in the nation on the SAT among the 23 states that had at least 50 percent of their high school graduates tested."

### Why 23 states not 50 states

"Washington, a Northwest rival in SAT test performance, outperformed Oregon by three points on the math section and one point on the verbal section of the test."

### Is three points significant ?

### Is it possible?

Oregon improved by one point over last year's score on the math section of the test for a total of 528.

They gained another point over last year's score on the verbal test, posting a 527. The 2004 national average SAT scores were 518 for math and 508 for verbal

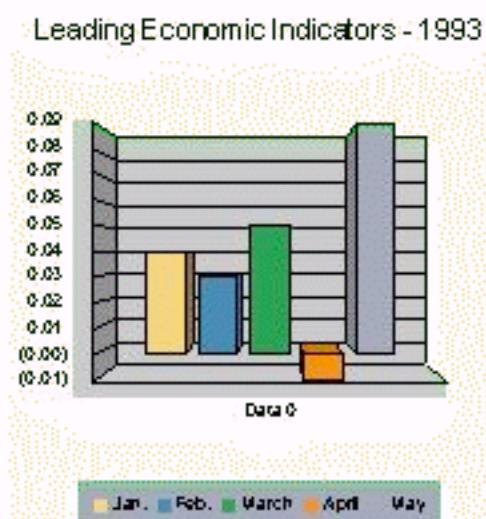
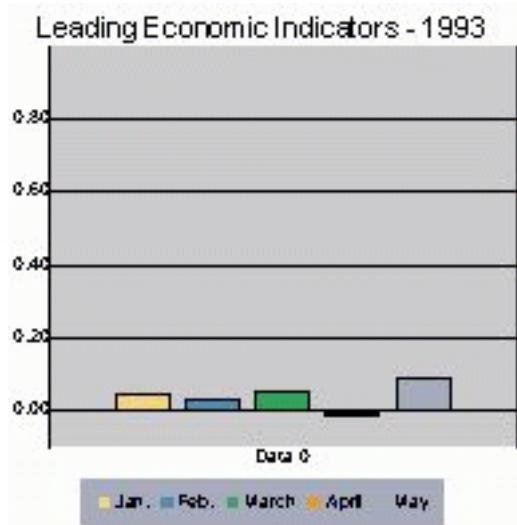
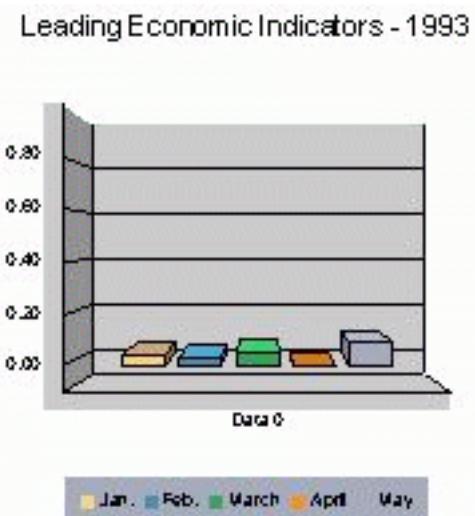
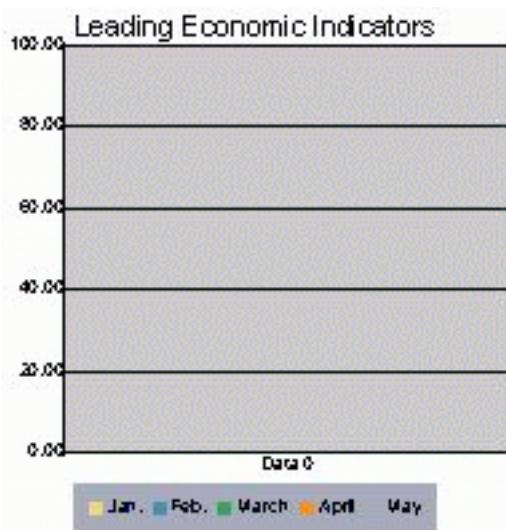
### What is the mean?

### What is the standard deviation?

Why is the scale 495-530? The score are from 200-800.

# One Dimensional Picture

# Look what a few changes can do?

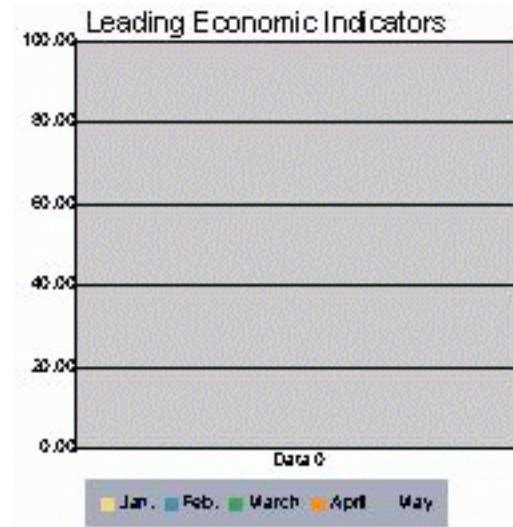


## Leading Economic Indicators for the U.S -1993

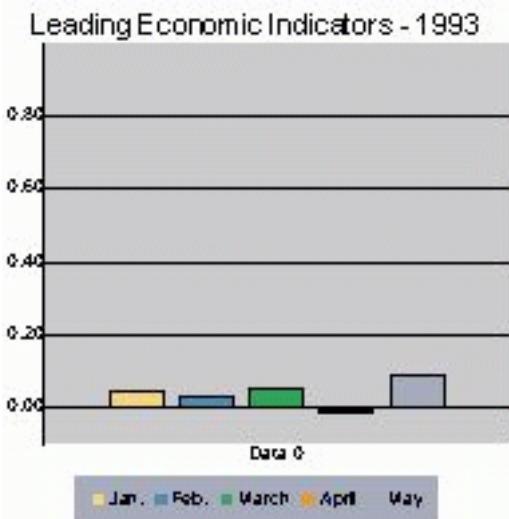
The data for the first five months of the year are as follows:

- January - +0.04
- February - +0.03
- March - +0.01
- April - -0.01
- May - +0.09

• Let's graph it



Well, that doesn't show much, does it?



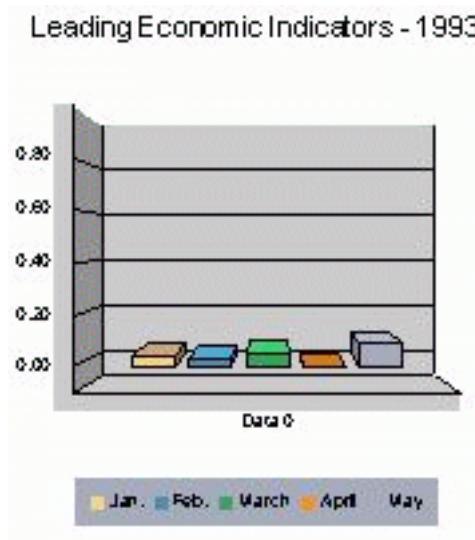
Instead of using a 0-100% scale, why don't we reduce the top end of the scale to 1% and extend the bottom of the scale to -.1%.

Let's add a little visual interest

Now, we are cooking

But what conclusions can I reach from this data ?

Do we have a story ?



Looking at the graph -- the change from April to May was not that great.

# *Post Hoc Rides:*

Cause and effect, or some common cause, or coincidence, or what?

"There are two clocks which keep perfect time. When "A" points to the hour. "B" strikes. Did "A" cause "B" to strike?"

# *The Semi attached Figure:*

The reported data may even be irrelevant.

This mouthwash kills germs (but maybe not cold germs in the mouth).

# *How to Statistic late:*

Lying with statistics = dishonesty or incompetence?

## The USA Today web site:

*Computers seem to be taking over the world these days and no where do they seem as important as in our schools. But the availability of computers in the schools is spotty at best. Some of the lowest computer to student ratios are in the West.*

*Some of the highest are in the East.*

The headline seems to indicate that the percentages represent the number of students per computer in the respective school districts.

Wyoming will have 7.0 students per computer (an average we assume).



*Source: General Accounting Office*

A problem: What is 0.6 of a student?

# Improved Graphics



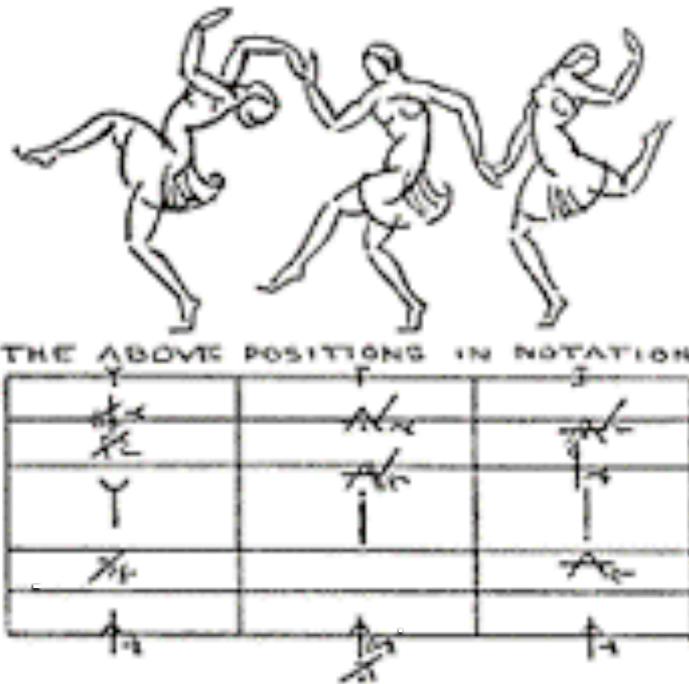
TEE-SHIRTS

Source Envisioning Information by Edward Tufte



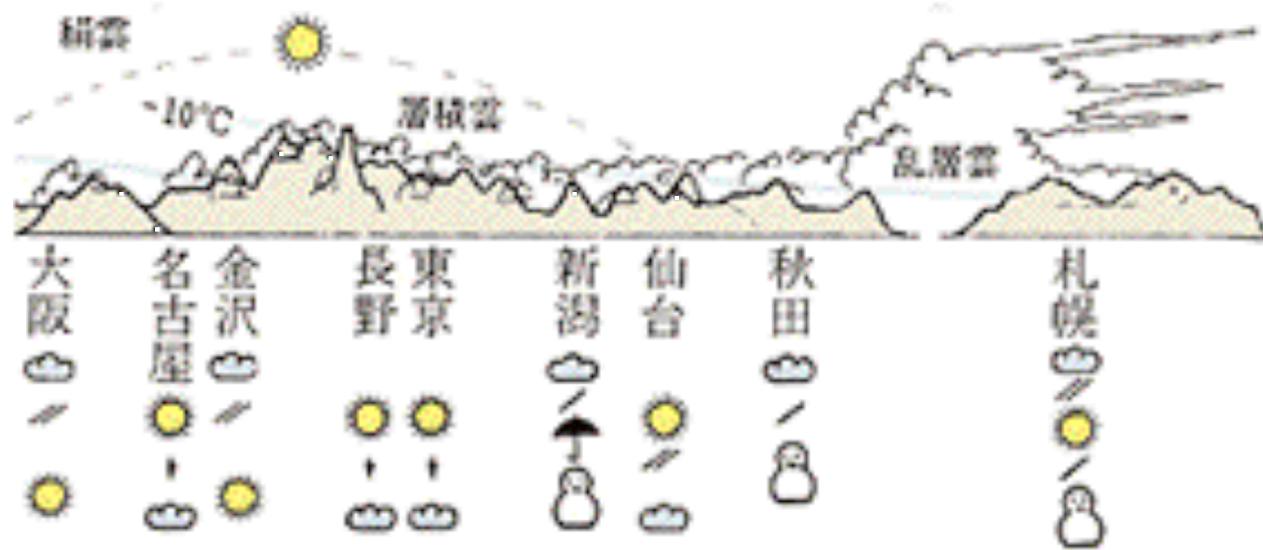
MUSIC SCORE WITH DANCE NOTATION

Source [Envisioning Information](#) by Edward Tufte



THREE DANCERS WITH DANCE NOTATION

Source Envisioning Information by Edward Tufte



JAPANESE WEATHER MAP

Source [Envisioning Information](#) by Edward Tufte