Four Principles for Statistical Computing

Before the 40 Years

- 1940s: Birth of the computer
- 1950s: "Two Cultures" (C. P. Snow)
- 1957: Sputnik & Science's Glory Days
- 1959-60: Meeting the computer and Fortran

The Computer: An Epiphany

Encountering the computer and Fortran ca. 1960 was a mind-altering experience for a science undergraduate.

1. Digital computing produced "true" answers -- well defined and testable! (in contrast to the slide rule)

2. Fortran provided a scientific language for programming. (Fortran: *For*mula *tran*slator)

Striving for "true" computations and providing an appropriate "language" remain challenges today. Indeed, here are the first two principles:

Four Principles for Statistical Computing

1. Report the truth about data.

2. Provide the language of statistics for computing.

1. Report the Truth about Data

- Accuracy of numerical results and simulations
- Correctness of programming (defensible coding)
- Truthful visualization and reports

Statistical Software at Bell Labs, 1965 - 2005

A few Markers

- 1965 BLISS (prototype only): PL-1 Wrapper
- 1967-1976 Fortran libraries
- 1976-1981 S1, S2: Language, Interfaces & Macros
- 1988 S3: Objects and Functions
- 1991 Models: Formulas and Data Frames
- 1998 S4: Classes, Methods, Connections

Statistical Software at Bell Labs, 1965 - 2005

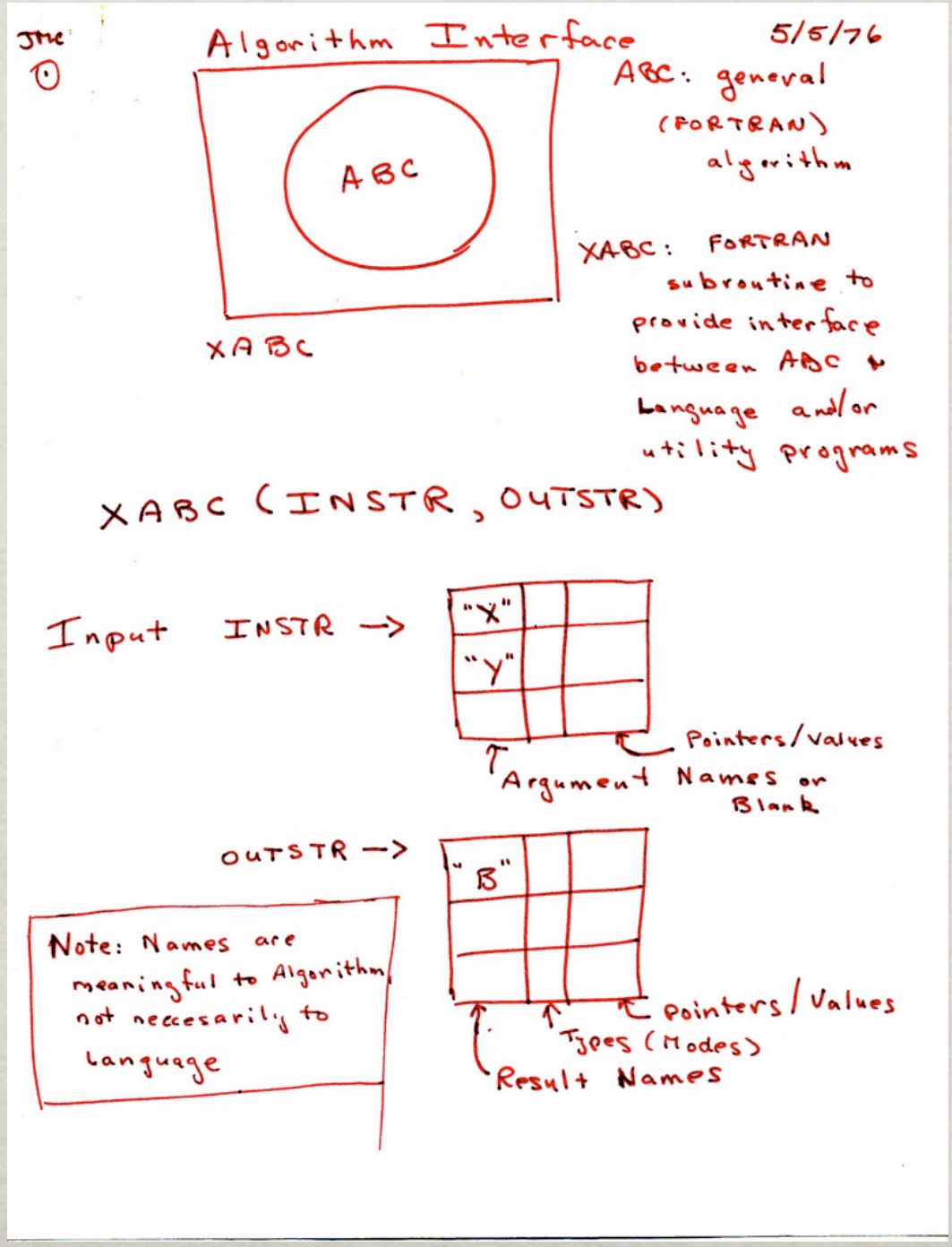
A few Markers

- 1965 BLISS (prototype only): PL-1 Wrapper
- 1967-1976 Fortran libraries
- 1976-1981 S1, S2: Language, Interfaces & Macros
- 1988 S3: Objects and Functions
- 1991 Models: Formulas and Data Frames
- 1998 S4: Classes, Methods, Connections

1976 - Interfaces: Enabling Access to Algorithms

May 1976: first meetings leading to S.

This sketch proposed an interface between a Fortran routine and the interactive user language.



Enable Ideas

The interface to an algorithm gives access to the idea, enabling it to be used flexibly in research or applications.

Early S, and later versions too, can be seen to be **enabling ideas**, rather than just providing specific statistical analyses.

Enabling important ideas effectively can in turn influence the practice of statistics (e.g., how we deal with models).

The third principle:

Four Principles for Statistical Computing

1. Report the truth about data.

2. Provide the language of statistics for computing.

3. Enable ideas as software.

S and the Community

- The S software.
- The S language.
- The community of users and programmers.

- The S language.

  Described in books from Bell Labs.

  Versions (S2, S3, S4) over 20-year period.

  Then along came 

- The community of users and programmers.

  Contributed software became important soon after distribution began (remember Statlib?)

  R now supplies the largest, most essential flow of ideas to (and **by**) the statistical community.

  This is critical. Why?

  Statistics can only thrive if new ideas are truthfully implemented as software. For this the community ***must*** be involved.

  And this is the fourth principle:

Four Principles for Statistical Computing

1. Report the truth about data.

2. Provide the language of statistics for computing.

3. Enable ideas as software.

4. Involve and reward the community.

and beyond?