VISUAL STATISTICS

2.0

DOANE   MATHIESON   TRACY

VISUAL STATISTICS

2.0

David P. Doane
Kieran Mathieson
Ronald L. Tracy

Oakland University

McGraw-Hill Irwin

McGraw-Hill Higher Education

A Division of The McGraw-Hill Companies

VISUAL
STATISTICS
2.0

To

Blythe,
Cullyn and Teagan,
Kathy

Preface

What's New in the Revised Edition

In this new edition of *Visual Statistics* we have made several major enhancements and many minor improvements. Here are a few of the new features.

- Six new modules:
    - Random processes
    - Goodness-of-fit tests
    - Multiple regression analysis
    - Regression assumptions
    - Regression models
    - Binary predictors in regression

- Full integration of the worktext and software:
    - Adobe™ on-line worktext now has color and graphics
    - Jump to any desired part of the worktext chapter using bookmarks
    - Switch easily between worktext and software module
    - Worktext can be magnified, copied, or printed
    - View learning exercises and their solutions simultaneously

- Attractive one-screen main menu that lets you:
    - Select a chapter by clicking it
    - View the worktext for that chapter
    - Run the software module for that chapter
    - Run several modules simultaneously

- Hundreds of new and updated databases and examples
    - Revised examples
    - Updated databases
    - New multivariate databases

- Major changes in appearance and functionality in modules on:
    - Central Limit Theorem
    - Power and type I/II error
    - Trends and seasonality
    - Statistical process control

- Worktext enhancements
    - 2 levels of orientation (basic features, additional features)
    - 3 levels of exercises (basic, intermediate, advanced)

- Module enhancements:
    - Ability to copy/paste graphs as device-independent bitmaps
    - Improved use of color

Background

This textbook began near ten years ago when Tracy walked into Doane's office to see if he would be interested in writing a National Science Foundation (NSF) grant proposal to develop software to teach statistical concepts. The collaboration was natural. Each of us had taught statistics for many years, and had experience developing computer software. As our ideas began to take shape, we agreed that it was time to bring the computer into the *classroom* — not just have students do computer lab projects. We wanted to use computers to teach *concepts* rather than just analyze

data. We wanted to stress visual displays rather than numbers, equations, and calculations. An equation is an explanation to some students, but it is a barrier to others. We felt that a series of pictures or a simple animation could illustrate fairly complex equations. Our proposal asked the National Science Foundation's support to develop software and learning tools that would bring the power of computers into the statistics classroom. After we were awarded the NSF grant, we asked Mathieson if he would serve as a Windows consultant to our project. Soon it was clear that Mathieson's skills were essential, and we became full partners.

Visual Learning

We assume that (1) for many students visualization is a key to learning; (2) most statistical packages are designed for data analysis, not visualizing concepts; (3) individual instructors lack the time and resources to design their own software. Accordingly, we set out to create learning modules to illustrate concepts that cannot easily be shown mathematically or with packages such as Minitab or Excel. We wanted to minimize duplication of software capabilities already available. We wanted to promote learning through active self-discovery (inductive learning) as well as to support the instructor in the traditional lecture/lab setting (deductive learning).

We came up with the idea of a worktext that is closely coupled with software devised to lead a learner through experiments that promote self-discovery. We designed learning projects for teams and individuals that pose less-structured problems to continue this process. Our software was designed to encourage experimentation, so a learner can use simulation to understand the "experimental" side of statistics. These same features can be used effectively by the instructor in the classroom or the computer laboratory to illustrate a concept (e.g., the relationship between a confidence interval and a test statistic) and reinforce it with a simulation experiment.

Overview of Features

Visual Statistics consists of 21 software modules and a worktext. Each chapter of the worktext begins with a list of the key concepts the chapter will cover and a short list of learning objectives. Each concept is reviewed and illustrated. A short (about 15–minute) orientation to each module is provided. *Past users have told us that this orientation is invaluable.* There is also an orientation to additional features that may be skipped if its topics are not required. Learning exercises (basic, intermediate, advanced) are provided for each module. These exercises guide self-discovery learning. There are open-ended learning exercises for individuals and teams. A self-evaluation quiz allows you to test your understanding of the material. Solutions and a guide to answering each question are provided. A glossary at the end of each chapter provides definitions of terms used in the chapter.

Visual Statistics is to be used with a textbook or other course materials. It complements these traditional sources—it does not replace them. However, you'll find topics in these sources easier to understand after you *see* them in *Visual Statistics*.

Every software module has its own Help system, similar to that in other Windows programs. It includes definitions, examples, equations, graphics, and hints. You can use the table of contents and hypertext capabilities to jump to topics you want to learn more about, or do keyword searches using the index. *You will get more out of the learning process if you use the Help system.*

Our software uses familiar Windows visual controls (command buttons, scroll bars, list boxes, option buttons, and so on). Input is mostly from a mouse rather than from the keyboard. This interface will seem quite natural to most people. If you have used other Windows programs, you already know how to use *Visual Statistics* controls.

Every *Visual Statistics* module opens with a virtual notebook (a familiar ring binder with pages you can turn) that organizes the options. Clickable tabs divide the notebook into different sections, including an introduction (general idea of what is in the module), list of concepts (main topics covered), and sources of more information (references to chapters or other textbooks). Other tabs that vary from module to module provide examples (illustrations using real data), scenarios (realistic context for simulations), databases (large number of related variables using real data), visual templates (icons that show a distribution or trend), do-it-yourself controls (control panels to manipulate parameters of an experiment), and a data editor (to enter data or paste data from a spreadsheet). You may return to the notebook while running the program (e.g., to change scenarios, pick a different example, or choose different options).

Level of Complexity
We have tried to write flexible software, so the instructor can omit topics or vary their order. No module requires knowledge of another module, except in a general way (e.g., simple regression precedes multiple regression). Because students learn in different ways and at different rates, each module is aimed at the average learner but with options for those who desire more analytical depth (or who are just adventuresome). For example, in some modules you can set parameters by manipulating do-it-yourself scroll bars (simple level), by typing parameter values (intermediate level), or by choosing a known distribution and its parameters (advanced level).

Acknowledgments
Although we take responsibility for all remaining errors, our project has benefited greatly from field testing, advice, and comments of faculty colleagues and their students at other universities. We thank these individuals for their dedication and interest in our project.
Revised Edition:
    Sung K. Ahn—Washington State University
    Nancy C. Aiello—Northern Virginia Community College
    Priscilla Chaffe-Stengel—California State University- Fresno
    Subha Chakraborti—University of Alabama
    Patti Collings—Brigham Young University
    Bradford R. Crain—Portland State University
    William Duckworth II—Iowa State University
    John C. Dutton—North Carolina State University
    James Flynn—Cleveland State University
    Robert Hannum—University of Denver
    David C. Howell—University of Vermont
    Glenn W. Milligan—Ohio State University
    Pin Ng—University of Illinois
    Roxy Peck—California Polytechnic State University- San Luis Obispo
    Don R. Robinson—Illinois State University
    Mark E. Rush—Washington & Lee University
    Sue B. Schou—Idaho State University
    Robert K. Smidt—California Polytechnic State University- San Luis Obispo
    Michael P. Wegman—Keller Graduate School of Management
Previous Edition:
    Richard Anderson-Sprecher—University of Wyoming
    Mohamed Askalani—Mankato State University

Contents

Introduction

Using *Visual Statistics*

OBJECTIVES

- Understand what *Visual Statistics* is all about

- Learn how to install *Visual Statistics*

- Learn how to use the Notebook, the starting point for every *Visual Statistics* module

- Learn about the worktext

What is *Visual Statistics*?

Statistics is a challenging subject for most people. Most of us have to work hard to understand probability distributions, statistical power, and confidence intervals. However, statistics is becoming more important. It's hard to design a product, create an advertising campaign, or play poker well without understanding statistics.

The main problem is that statistical ideas are abstract. When did you last speak with a probability distribution? Have you ever seen a correlation on the shelf at Sears? Or been almost run down by a wild confidence interval (except on an exam)? Formulas and numbers are fine for some people, but most of us learn best when we can see and interact with the thing we're learning about.

That's where *Visual Statistics* comes in. Computers have become more powerful and easier to use, so it's possible to explore statistical ideas in new ways. Graphics and animation can bring concepts to life in ways that formulas, tables, and static diagrams cannot. Rather than imagining what data for a particular situation looks like, *Visual Statistics* lets you create the situation and see it for yourself. You can change the circumstances and watch the result. *Visual Statistics* helps you build your statistical intuition, making it easier to understand the ideas behind data analysis.

It's important to understand that while you can analyze data with some *Visual Statistics* modules, that is not their main goal. Instead, *Visual Statistics* will help you *learn statistics*. Even the most powerful analysis tools, like Minitab, SAS, and SPSS, won't help you if you don't know what to do with them. On the other hand, if you are familiar with statistical ideas, you can do wonders with a simple calculator. The tools you use are less important than how you wield them. *Visual Statistics* will help you use statistical techniques more effectively because you will understand what they can be used for and what their limitations are.

Visual Statistics is not a tutorial, however. For example, the regression module won't explain why regression works or how to calculate regression statistics. Instead, it helps you explore regression, so you understand it better than you would without the visual tools. You'll get more value from your statistics text since you'll be able to experiment with the concepts. It'll be easier for you to *understand* statistics — and understanding brings more confidence and less worry.

The *Visual Statistics* suite consists of 21 modules. Each one helps you explore a specific set of statistical ideas. For example, the ANOVA module will help you understand one-way analysis of variance. It is designed for that purpose alone. It's different from the module that helps you understand statistical process control, since the two topics are different. However, every module shares some characteristics. First, they are all easy to use. Run through the 15-minute orientation for each module, and you should be able to use it effectively. If you have trouble, select Help for assistance. Second, they all use the Notebook. The Notebook introduces the module, and lets you choose how you want to explore the relevant statistical ideas. Different modules have different Notebook options. Some let you examine scenarios, while others allow you to analyze real data sets or enter your own data. The Notebook is described in more detail later in this introduction.

The best way to learn what *Visual Statistics* is all about is to try it. So go ahead. Install *Visual Statistics,* start the Univariate Data Analysis module and play around. You'll see how the same data can look very different, depending on how it is displayed.

Installing *Visual Statistics*

Visual Statistics runs under recent versions of Windows (95, 98, 2000, NT). You need at least a Pentium PC with 32 MB of memory. *Visual Statistics* takes about 7 MB of hard disk space for a Compact installation or about 37MB of hard disk space for a Complete installation. This includes the software (21 modules), the worktext (21 chapters plus the introduction), worktext solutions, help files, and databases containing over 1000 variables. The Compact installation runs off of the *Visual Statistics* CD while the Complete installation runs off your computer and only checks to see that you have a *Visual Statistics* CD.

Installing VS
This version of *Visual Statistics* is for a single user. Users needing a network version should contact their McGraw-Hill representative. To install *Visual Statistics*, first close down any other programs you are running. Place the *Visual Statistics* CD in your CD drive. The install program is setup.exe in the Setup folder in the Visual Statistics 2.0 folder. It can be executed in one of three ways:

- Use the Add/Remove Programs dialog box in the Windows Control Panel, *or*
- Double click the setup.exe file in Windows Explorer, *or*
- Press the Start button, select Run, and type d:\visual statistics 2.0\setup\setup.exe (where d: represents the drive letter for your CD reader).

You will be asked to select one of two setup types. *Both require that your Visual Statistics CD be in your computer.*

COMPACT installs the front-end of *Visual Statistics* on your computer and the necessary system files. It takes about 50 Mb LESS room on your hard drive.

COMPLETE installs the entire *Visual Statistics* system on your computer. It takes about 50 Mb MORE room on your hard drive, but runs faster on computers with slower CD readers.

Read the License Agreement during the install. After installing *Visual Statistics* it can be read by selecting the Visual Statistics 2.0 folder in the Programs folder. *Visual Statistics* can be started from this folder or by double clicking the VS icon that will be installed on your desktop. Once the system files have been installed on your computer, *Visual Statistics* can also be run directly from the CD by opening the *VisualStatistics*.exe file on the CD. This can be useful if your VS files have been corrupted and you want to run *Visual Statistics* without reinstalling the software.

Uninstalling VS
Visual Statistics can be uninstalled in two ways:

- Use the Add/Remove Programs dialog box in the Windows Control Panel, *or*
- Press the Start button and select Uninstall Visual Statistics 2.0 from the Visual Statistics 2.0 folder in the Programs folder.

Installing Adobe Reader 4.0
When you run the *Visual Statistics* install program, it will detect if you have already installed an Adobe Reader. If you do *not* have an Adobe Reader, VS will install the Adobe 4.0 Reader for you. If you *do* have an Adobe Reader, VS will *not* install the Adobe 4.0 Reader. In this case, you *should* determine which version of the Adobe Reader is on your computer. Although *Visual Statistics* will

work with versions earlier than 4.0, the presentation of graphics improved considerably with version 4.0. If you have an earlier version, it is strongly recommended that you uninstall it and install the 4.0 version by executing the ar405en.exe file in the Reader folder on the *Visual Statistics* CD or by getting the latest version from Adobe at www.Adobe.com. Adobe warns that earlier versions should be uninstalled *before* installing version 4.0. If the earlier version is uninstalled *after* the 4.0 version is installed, problems will develop with the 4.0 Reader according to Adobe Corporation.

Starting *Visual Statistics*

Main Menu

To start *Visual Statistics*, click its icon on your desktop, or select *Visual Statistics* 2.0 in the Programs folder in the Start menu. After opening the cover, you will see the a menu like the one shown in Figure 1. From this menu, you can do three things: (1) view a chapter in the worktext (the Show Worktext button); (2) run a software module (the Run Module button); (3) exit *Visual Statistics* (the Exit button). Subject to your system's memory, you may simultaneously use as many chapters or modules as you wish (they will appear on the taskbar at the bottom of your screen). When a chapter is selected, its title will turn white and its icon will turn yellow. In Figure 1, for example, Chapter 1 (Visualizing Univariate Data Analysis) has been selected.
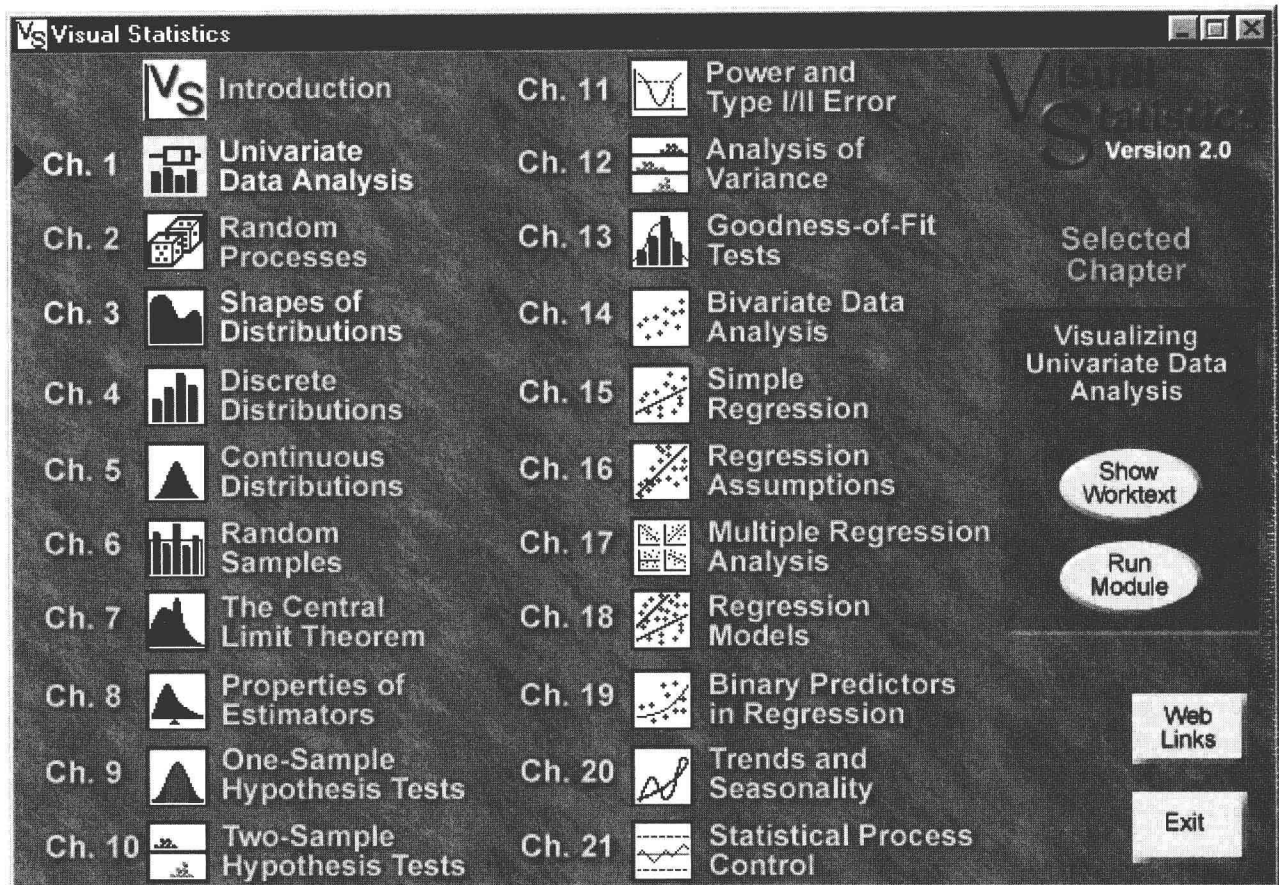

Figure 1: Main Menu

Selecting a Chapter

To select a chapter, click on its chapter number, its icon, or its title. If a chapter has not been installed, its title will be "grayed out" to indicate that it is inactive. When you click the chapter, a comet will streak across the screen to the Selected Chapter panel on the right, and the Show Worktext and Run Module buttons will appear in the panel. If either the module or worktext was not installed, the button will be "grayed out" and will be inactive. Each software module corresponds to a chapter in the worktext. The chapter's learning exercises will require that you run the corresponding software module. You can run several software modules at once, subject to your system's memory.

The *Visual Statistics* **Worktext**

The worktext has one chapter for each module. Each chapter in the worktext contains:

- A list of concepts and learning objectives
- An overview of concepts and illustrations of concepts
- An orientation to software features
- Structured learning exercises (basic, intermediate, and advanced)
- Suggested independent projects (team, individual)
- A bank of self-test questions
- A glossary of terms
- Answers to the self-test questions

Each chapter starts with a list of relevant concepts and learning objectives. Next, it offers an overview and illustration of each concept. Then there is a short section entitled Orientation to Basic Features. *Make sure you read this section.* It will quickly show you how to use the most important elements of the module. It only takes a few minutes and is well worth it. The Orientation to Additional Features illustrates other important capabilities of the software, which will be necessary to complete some of the exercises. These features are not necessarily more complex, but they intended to suggest priorities that may assist a busy learner or that might be recommended by an instructor if time permits.

The Learning Exercises ask questions you can answer using the module. They are to be completed in front of a PC. You will have a better understanding of the statistical ideas covered in the chapter once you run through the exercises. The Basic, Intermediate, and Advanced exercises are arranged in gradually increasing level of challenge. Your instructor might ask you to turn in some of the exercises. However, you can grade your own work using the Solutions.

The worktext also includes Learning Projects for individuals and teams. Where the Learning Exercises guide you through the module, the Learning Projects leave you to figure out how to use the module and do the analyses yourself or with your team. You may be asked to do a written report or make an oral presentation based on the Learning Projects. It is a good idea to do the Learning Exercises before you try the Learning Projects.

The Self-Evaluation Quiz lets you test your understanding of the ideas covered in the chapter. Try it after you have completed the Learning Exercises. The answers are at the back of each chapter, after the chapter's Glossary. You should do the entire quiz first, before you look at the answers. If you answer a question incorrectly, there are references to relevant sections of *Visual Statistics* that will help you understand the right answer. The exercises are arranged somewhat in increasing order of challenge. Some of the exercises can be answered by looking at the Overview, Illustration, and Glossary, while others may require using the software or working certain exercises.

The Glossary defines the terms used in the chapter. Much of the Glossary information (plus a lot more, such as formulas) will also be included in the module's Help file. Click Help on the menu bar to look at the Help file. It contains definitions, formulas, examples and explanations as well as usage instructions. The Help files are an oft-overlooked resource.

Using the *Visual Statistics* Worktext Reader

To view a worktext chapter, click the Show Worktext button (if the chapter has not been installed, the button will be "grayed out" to indicate that it is inactive) to bring up the Adobe reader, illustrated in Figure 2. Each chapter opens to its title page. You may jump to any section of the chapter by clicking the appropriate title in the Bookmarks window of the Adobe reader (on the left side of the screen) or you may use the scroll bar on the right. The Run Module button takes you to the chapter's software module. The Show Solutions button will display the answers to the learning exercises in a separate window at the bottom of the screen. The Return to VS button displays the *Visual Statistics* main menu.
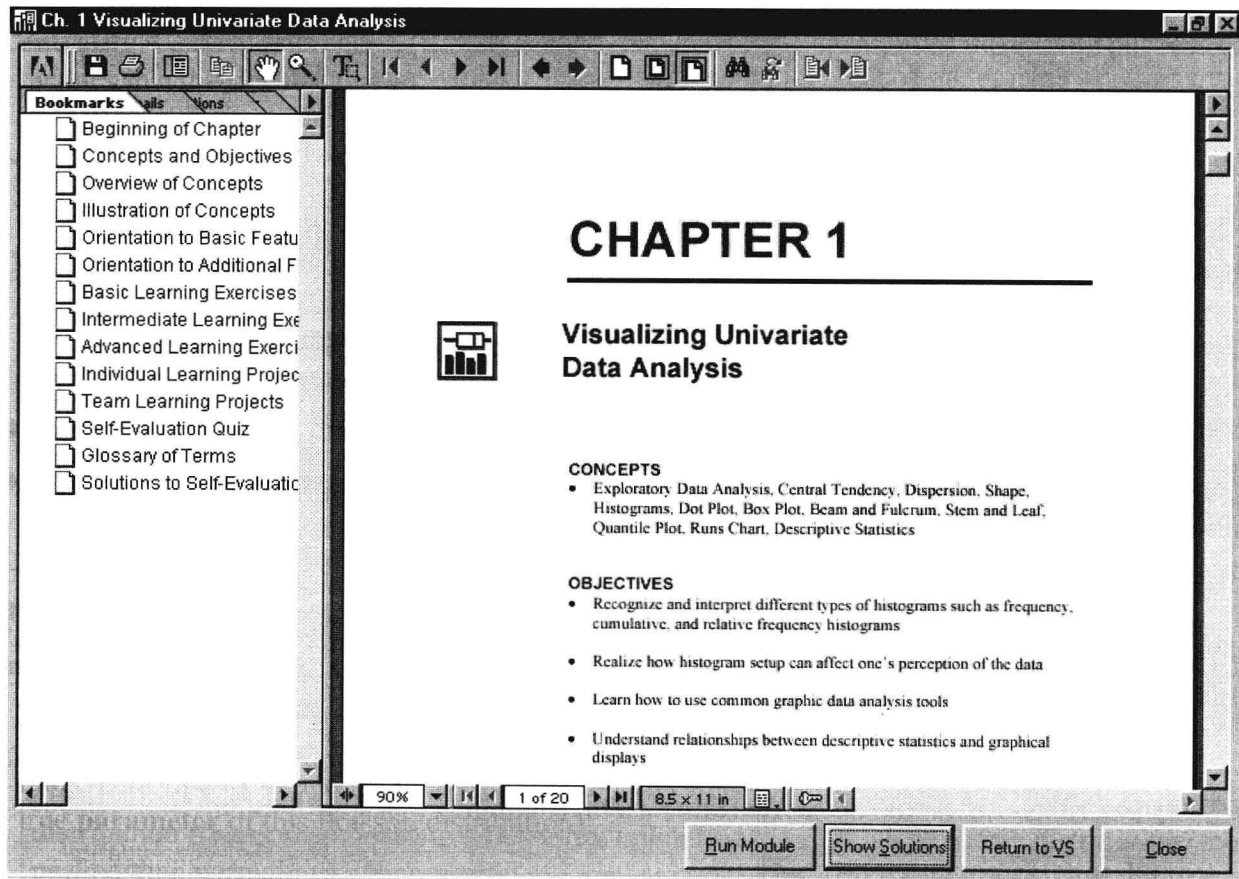

Figure 2: Adobe Reader