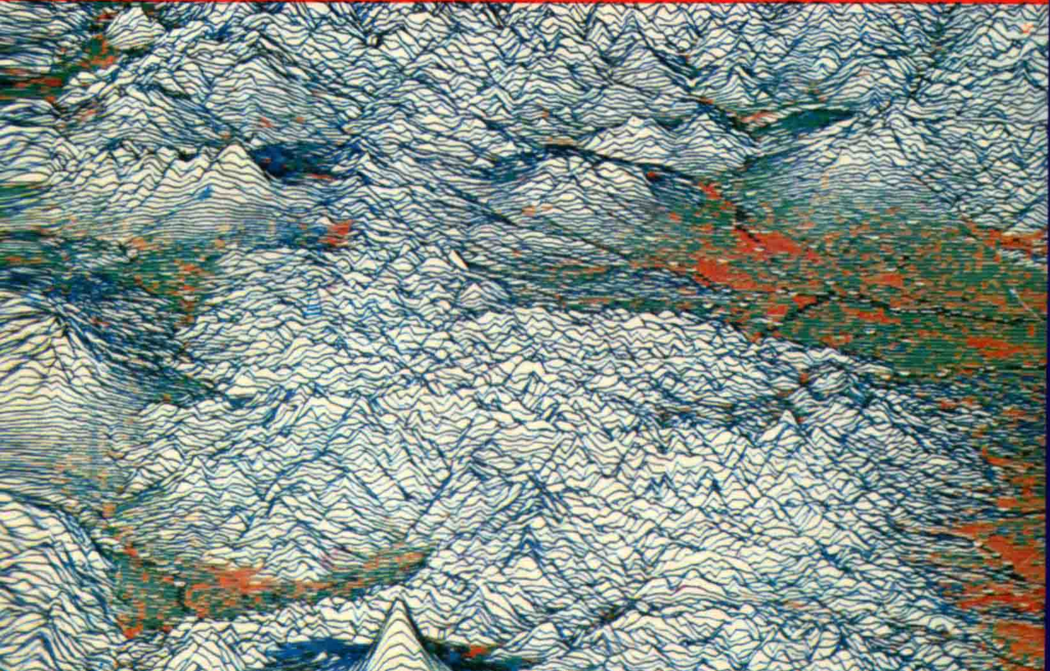
COMPUTER PROGRAMMING FOR GEOGRAPHERS



DAVID J. UNWIN AND JOHN A. DAWSON

Computer programming for geographers

DAVID J. UNWIN AND
JOHN A. DAWSON

LONGMAN
LONDON AND NEW YORK

Longman Group Limited
Longman House, Burnt Mill, Harlow
Essex CM20 2JE, England
Associated companies throughout the world

Published in the United States of America
by Longman Inc., New York

First published 1985

British Library Cataloguing in Publication Data
Unwin, David J.
Computer programming for geographers.
1. Geography – Data processing 2. Electronic digital computers – Programming
I. Title II. Dawson, John A.
001.64'2'02491 G70.2
ISBN 0-582-30095-9

Library of Congress Cataloging in Publication Data
Unwin, David John.
Computer programming for geographers.

Bibliography: p.
Includes index.
1. Geography – Data processing. I. Dawson, John A.
II. Title.
G70.2.U58 1985 001.64'24 84-5690
ISBN 0-582-30095-9

Produced by Longman Group (FE) Limited
Printed in Hong Kong

Acknowledgements

Many of our students and colleagues have helped, knowingly and unknowingly, in the preparation of this text. Undergraduate classes and graduate students at Leicester and Lampeter have freely criticized exercises and helped in their redesign. We hope they have been improved and gratefully acknowledge the help and stimulus provided by these guinea-pig users. Alan Rogers provided helpful comments from the viewpoint of a computer scientist while Graham Sumner and Leigh Sparks commented on various sections of the text and made useful suggestions. The production of the text has been made easier by the artwork of Ruth Rowell who with great skill has turned sketches into finished diagrams. Janice Lewis typed a considerable portion of the text using her sometimes amazing powers of decipherment. We alone are responsible for any errors that remain in the text even after the help of these colleagues and friends. Finally, to Jo and Kathryn to whom this book is dedicated, go our heartfelt thanks for encouragement, patience and control of our excesses during the entire project.

Figures 5.4 and 6.6 are reproduced from *Computing for Geographers* (Dawson, J. A. and Unwin, D. J., 1976, Newton Abbot, David and Charles) and Figures 6.1 and 7.2 are redrawn from the same source. We are grateful to the original publishers for permission to reproduce these works.

January 1984

David J. Unwin
John A. Dawson

Contents

List of figures

List of tables

1

Computers in geography

1.1 What's it all about?

Nowadays everyone in the developed world encounters computers. In our daily lives we meet them in banks, in government, at airline reservation desks and in numerous other places. Computers are frequently used to control traffic lights, household appliances, arcade video games and many other everyday objects. Whatever we may think or say to the contrary, the chances are that everyone in western society has used a computer at some time or other. In science they have become an indispensable aid to information processing in all disciplines from archaeology to zoology. The reason for this penetration into so many fields is that computers are general-purpose machines capable of being programmed for use in a myriad of situations. It follows that computers may be approached in many ways and from many perspectives. An engineer might see one as a complex set of electronics, a physicist as a means of rapid calculation, a businessman as an automated office, an author as a very good filing and printing system, a student as a personal instructor and so on.

This book describes how computers can be used to assist in the solution of the sorts of problem that a geographer analyses, such as:

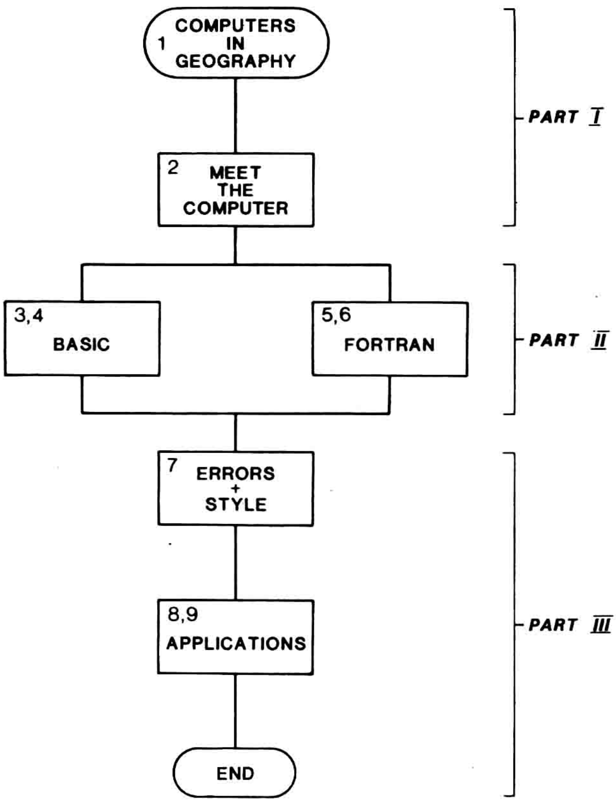
- Collecting and storing data from environmental sensors.
- Processing census or other survey data.
- Drawing maps of spatial data.
- Running a simulation to copy some real-world process.
- Describing and analysing spatial patterns.
- Writing a research report, essay or textbook on computing.

The book is intended for use in two ways – as a practical self-study workbook or as a course text to accompany a short programming or appreciation course of about 20 hours duration. For the first use we have included numerous examples of geographical programs and summary worksheets which suggest possible projects. In addition, embedded within the text in the 'language' chapters are a series of simple exercises that can be used to monitor progress. An instructor taking a course in the classroom might find them useful to relieve the usual one-way monologue. For the second use we have added background material on computer history, advanced applications and other related topics.

We should stress that our worked examples have been written for their effectiveness in illustrating particular programming concepts; in no sense are they intended as polished routines for research use. Similarly, the worksheets are intended to provide examples of what follow-up work seems appropriate and we would anticipate that individual instructors will adapt them to suit local circumstances. For both uses access to a computer, preferably by way of a keyboard and visual display, is absolutely essential; reading a book is no substitute for getting down to work with a machine.

After reading the book it is hoped the student will:

- Have an appreciation of how computers can be used in geography and how problems that arise naturally within the subject can be posed in ways that facilitate computer assistance.
- Be able to program simple tasks in one or other (or both) of two programming languages.
- Have an appreciation of what facilities are available to help solve complex problems beyond the scope of a single individual programmer.



1.1 The organization of this book

The book is organized into three related parts, illustrated by the flow-chart in Fig. 1.1. This and the next chapter present an introduction to computers and problem solving by computer. Chapters 3 to 6 outline the essentials of two programming languages, BASIC and FORTRAN. For many, the chapters on BASIC will suffice, but for those who want to extend the range of what they can do, Chapters 5 and 6 go into FORTRAN. Finally, Chapters 7 to 9 deal with the concepts of programming style and with further, more advanced, geographical applications.

1.2 The computer and the geographer

A notable feature of many analyses in geography is that they involve large numbers of often simple tasks which, if attempted by hand, are inherently very tedious and error prone. This is particularly the case with statistical computations but it also includes the routine reading and recording of instrumental data, recording grid co-ordinates, measuring areas and distances, drawing maps and writing reports.

The easiest of these tasks to automate are those involving calculation and the idea of making machines to do this is very old. The Chinese abacus is one of the oldest forms of 'calculator', with improvements on it being relatively recent, dating from the seventeenth century when devices such as the logarithm, slide rule and Pascal's adding machine were invented. All these certainly help speed calculation procedures but in a similar way to modern electronic calculators their effectiveness is severely limited by the need for the frequent intervention of a human operator to provide data, to control the sequence and nature of the operations and to write down the results. In the modern jargon such machines lack 'intelligence'. Clearly, a major advance would be to develop machines that could not only calculate but could also execute a series of stored instructions without human intervention. During the nineteenth century several such devices were invented. The first were the 'difference' and 'analytical' engines proposed by a visionary mathematician, Charles P. Babbage. Although mechanical, his machines embodied many of the features of a modern computer. The second was a primitive form of program control by punched cards developed by and named after a Lyons silk weaver, J. M. Jacquard.

By 1900 the idea of programmable, intelligent machines was well established but further development had to await improvements in electronic engineering. If we regard a computer as an electronic machine for processing information in a predetermined way by means of stored instructions, the first such devices were developed and used for military intelligence work during the Second World War. From then on development has been extremely rapid. The first peacetime machines were produced in British and American universities from 1946 onwards with the first commercial machines appearing around 1950. By today's standards this first generation of machines was very primitive. Relying on vacuum tubes and other standard electronic components, they were physically large, consumed a lot of power and even though intelligent had very little memory. Even more significant was that they could be programmed only in machine codes, which were difficult to understand by the average scientist. Consequently computer programming began as an obscure and esoteric craft tailored more to the needs of the electronics engineer than to the eventual user of the results. We can find no examples of use by geographers of these early machines but in a related field a very early application was in numerical weather prediction.

Towards the end of the 1950s a second generation of machines based on semiconductor devices was introduced and gave a large increase in speed and memory. Equally important was making the machines easier to program by the provision of 'low' and 'high-level' programming languages. A set of instructions, the program, should specify a series of operations to be performed, the quantities or symbols on which these operations should take place, and where any results should be stored. In the computer all this must be specified by a meaningful sequence of binary digits (0 and 1, as in 00101110) because ultimately it can only recognize the two states 'on' or 'off'. These sequences of binary digits are termed 'machine code' and first-generation machines required them to be provided by the user but in the second generation it was realized that the machine itself could be made capable of translating from a simple form of language into machine code. To do this the computer was given a general program to carry out the translation and so allow the use of more meaningful symbols than 0 or 1. Low-level assembly languages allowed the programmer to use simple mnemonics such as ADD or MULT for single-step operations and labels like X and Y for the quantities. Even so, assembly language programming remained for most people a difficult task far removed from the problems that they wanted to solve. Assembly languages are still available on modern machines and are frequently used where execution speed is a major concern but they remain rather difficult to understand by the average user. Fortunately, we can go a step further and allow one program statement to generate a whole series of machine instructions by relying on a second special program called either an interpreter or a compiler resident in the machine to do the necessary conversion. Often this provision of a compiler results in a loss of absolute computing speed but it enables 'high-level' languages to be defined with a structure and syntax which resembles normal English or mathematics. The first of these was FORTRAN, an acronym for FORmula TRANslation, developed by IBM for scientific use around 1954 but it was followed by COBOL (COmmon Business Oriented Language) for commercial use, BASIC (Beginners' All-purpose Symbolic Instruction Code) for educational use and many others. To a very large extent these languages do not depend upon the features of any particular computer so that, in theory at least, programs written in them ought to be able to be run on any machine with the appropriate interpreter or compiler.

Second-generation machines were much easier to use than those which had gone before and it is from the mid-1950s that the first geographical applications date (see Barry 1960 ; Kao 1963). Historians of geographical thought, such as Burton (1963), have spoken of a 'quantitative revolution' in the subject that took place from the mid-1940s onward until around the mid-1960s, changing geography from an essentially qualitative and descriptive discipline into one which increasingly concerned itself with the quantitative testing of general statements about spatial distributions. Given these concerns it is hardly surprising that researchers turned to the newly available machines for assistance. But well into the 1960s, computer power was a scarce and valuable resource and these initial geographical applications were limited to processing numerical data in the more obviously quantitative parts of the discipline. The impact on the mainstream of the subject was really quite limited.

From about 1963 a third generation of computers became available and brought with it a massive increase in computer use throughout science, business and industry. These machines, many of which remain in service today, used medium- and large-scale integrated circuits, were configured into complex systems, had what at the time seemed to be very large memories and were able to operate at very, very high speed. Most geographical computer use dates from their introduction into the universities of Britain and America. Geographers made use of the high speeds of calculation to perform a variety of statistical techniques such as factor analysis, cluster analysis and so on (see King 1969 and for a review Haggett 1969). Usually computers in higher education were operated in a manner which encouraged this kind of computational work with very little interaction between them and their users. Typically a program and its data were input on pre-prepared punched cards or punched paper tape, one job at a time, and the output was printed on a high-speed lineprinter. Inevitably such working involved a delay caused by the queue of jobs awaiting processing that could be of several hours or even days duration between submitting a fairly simple job and getting any results. Program development was slow and data preparation and checking were tedious in the extreme. Nevertheless, by 1970 almost a quarter of all geographical research reported in two major national journals used computer-derived results (Dawson and Unwin 1976: 20) and most departments of geography had several graduate students and one or two staff who were reasonably familiar with computer use.

During the 1970s and 1980s developments in computer technology have led to quite fundamental changes in the relationship between computers and their users, including geographers. What was once a costly resource of very limited availability has become both cheap and plentiful and in response the number of geographers using computers has increased many times. At the time of writing two developments seem most significant. First, a fourth generation of machines has become available making use of very large-scale integrated circuits, extremely compact main memory and high-capacity magnetic disk 'backing store'. The speed of operation of these machines is such that they can be configured into systems that can process a very large number of jobs more or less simultaneously by 'time sharing' and making it possible for each individual user to interact with his program as it runs (see Ch. 2). Instead of punched cards and lineprinter output a great deal of the interaction with such a machine can be through the keyboard and screen of a visual display unit terminal (VDU) and a modern mainframe machine will be able to handle a large number of these simultaneously. From a user's point of view this is a very easy environment in which to learn computing and to get useful results. Because programs can be entered, stored and run from a keyboard program development is rapid, easy and without the often long delays associated with punched cards and line-printer. Large files of information entered at the keyboard and stored on magnetic disk can be read and processed at electronic speed and programs can be written that allow the user to direct the exact flow of work while they run. It goes without saying that such a complex computer system with one or more processors servicing a large number of terminals needs an extremely sophisticated collection of programs in its own operating system to control it, but most users do not need to know very much about the detail of this.

Despite their technological complexity and sophistication, fourth-generation machines are very much easier to use than their predecessors. Often a large number of programming languages are available to suit almost any application and there will be a library of 'package' programs for common tasks as well as many utility programs to help users edit and store their information. Nowadays these machines frequently are joined in networks. Users in one place are able to access facilities on a computer that might be physically very distant from them but they are largely unaware of this distance.

The same improvements in integrated circuitry that led to the fourth generation of large mainframe computers have also led to a parallel development that will prove to be even more significant. It has become possible to put all the necessary logic circuits of a computer processor onto a single silicon chip called a 'microprocessor'. By themselves microprocessors can be used in many devices but if they have attached to them further chips for memory, a keyboard and some form of display such as a domestic television, they become a microcomputer. Microcomputers first became available in the late 1970s and are very much smaller, lighter, more portable and cheaper than mainframes. By fourth-generation mainframe standards microcomputers have limited memory and tend to operate very slowly. They can do nothing that could not have been done many years ago by a second or third-generation machine. The difference is that their extremely low cost and portability have led to a major qualitative and quantitative leap forward in computing. Whereas in the past only a small section of society had access to computers, now it is possible for virtually anyone with sufficient interest to purchase his own machine for work, home or hobby use at a price that is less than the average weekly industrial wage. From a university computing environment it is all too easy to think that the typical computer user of the 1980s will have access by way of some sort of terminal to large, shared mainframe machines operating in an extensive network. In fact the reported sales of the most popular microcomputers make it quite clear that the typical computer user of the 1980s will sit at the keyboard of his own personal machine. The chances are that it will have been bought 'off the shelf' in a store and that its user will never have touched a punched card. His