A Mosaic of Television Ratings

J.A. HARTIGAN and BEAT KLEINER*

A mosaic is a graphical display of cross-classified data in which each count is represented by a rectangle of area proportional to the count. The positions and sides of the rectangles are set to encourage comparisons between counts in the figures. Mosaics are useful for discovering unusually high or small counts and for discovering dependencies between variables. In principle, mosaics may be used for any number of cross-classifying variables, but six seems to be a practical maximum. A mosaic is given for a four-way classification of Nielsen ratings.

KEY WORDS: Graphics; High-dimensional data; Contingency tables; Cross-tabulated data; Cross-classified data.

1. INTRODUCTION

Although the statistical analysis of many-way cross classifications has recently been much developed (Bishop, Fienberg, and Holland 1975), few graphical techniques are available to help suggest hypotheses, check assumptions, and interpret results. We propose the use of mosaics (Hartigan and Kleiner 1981), in which counts are represented by rectangles of area proportional to the count as in the familiar histogram for one-way classifications.

Two-way contingency tables have frequently been represented as series of one-dimensional histograms. The novelty of a mosaic is in the careful choice of positions and sides of the rectangles for counts, which makes it possible to compare visually many interesting sets of counts. Bertin (1967) proposes different choices of position and side for two-way tables.

A mosaic is constructed for Nielsen television ratings in 1977 and 1978, classified by network, time of day, day of week, and week of month.

2. THE DATA

Commercial television is supported by advertising priced by the number of homes reached, so the income of a station or network is directly determined by the popularity of its shows. The number of homes reached is estimated in the Nielsen report, which is compiled at three-month intervals for various marketing areas. Our data are based on the Nielsen report for the Hartford-New Haven metropolitan areas: In May 1978, 667,000 households had television sets in this area, and 991 of these households were keeping a Nielsen diary (see Head 1972 for more details on survey methodology). About 250 of these households report their viewing during each week of the month; thus, if the percentage watching *Happy Days* is 20%, this corresponds to 50 households; thus the standard deviation is about 3%.

We consider only the three principal networks, CBS, NBC, and ABC, ignoring the comparatively small number of viewers watching public television and independent stations. We examine the six half-hour prime-time periods from 8:00 p.m. to 11:00 p.m., for each day of the week, for each of four weeks in the three "ratings" or "sweeps" months (November 1977, February 1978, and May 1978). Note that some shows last more than a half hour, so high correlations should be expected in the proportion of viewers watching in those contiguous half hours; and while some time slots of a particular day of the week are occupied by the same show for all 12 weeks, it frequently happens that "special" shows or new series replace old ones.

3. CONSTRUCTING MOSAICS

A statistical analysis of these ratings, as in Goodhardt, Ehrenberg, and Collins (1975), handles continuity and competition using regression methods. Our purpose here is to illustrate a graphical method for examining cross-tabulated data that we call a *mosaic* (Hartigan and Kleiner 1981). In a mosaic, each count is represented by a rectangle or *tile* of area proportional to the count; the dimensions and position of the rectangle are chosen to facilitate comparisons between similar counts.

In Figure 1, a mosaic is given for the two variables—day and week. There is a tile for each of seven days and each of twelve weeks; the area of the tile is proportional to the total number of people recorded as watching television for that day and week, summed over the three networks and the six time slots. For example, Thursday in the fourth week of November, Thanksgiving Day, has fewer viewers than any other day. The height of each tile is proportional to the total number of viewers for that day—the heights are the same for each week within a day. Fewer viewers watch on Thursday, Friday, and Saturday. The width of each tile is proportional to the number of viewers for that week, compared with other weeks within the same day. There are more viewers in February, the middle four weeks, and fewer in May, the last four weeks.

The complete mosaic is obtained by breaking down each of the tiles in Figure 1 according to network and time of day. The first block in Figure 1 was broken into

*J.A. Hartigan is Professor of Statistics, Yale University, New Haven, CT 06520. Beat Kleiner is a supervisor at Bell Laboratories, Murray Hill, NJ 07974. This research was supported in part by National Science Foundation Grant MCS-8102280.

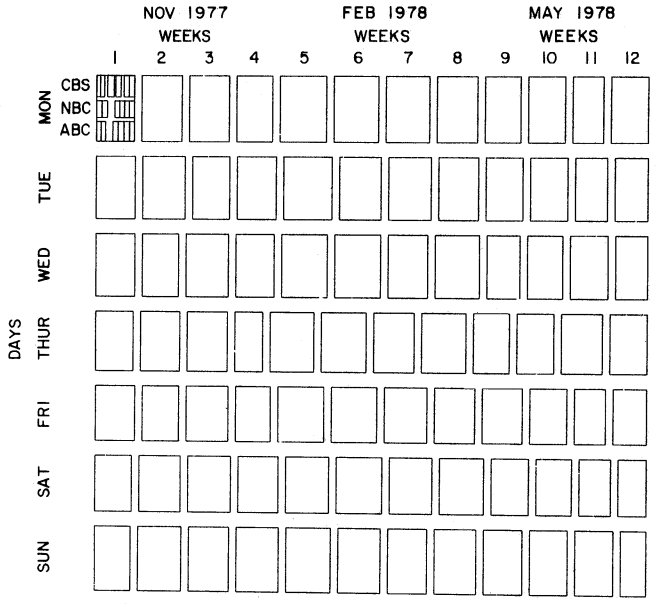

Figure 1. Mosaic of Nielsen ratings for the Hartford-New Haven area in November 1977, February 1978, and May 1978. The area of each rectangle is proportional to the number of Nielsen households watching television on a particular day, summed over the three networks and over the six prime-time half hours.

smaller tiles to illustrate the decomposition. The tile at the top left corresponds to the 15% of Nielsen households watching *Logan's Run* on CBS Monday night at 8:00 p.m. in the first week of November 1977. The tile immediately below it corresponds to the 17% watching NBC's *Little House on the Prairie,* and the tile below that corresponds to the 17% watching *Lindsay Wagner* on ABC. The second tile in the top row corresponds to the 17% watching *Logan's Run* at 8:30 p.m.; the 27% watching *Betty White* on CBS at 9:00 p.m.; the 22% watching *Maude* at 9:30 p.m.; and the 21% and 22% watching an hour of *Rafferty*.

The lengths of the vertical edges of the tiles are proportional to the average percentages watching CBS, NBC, and ABC on Monday night: 21%, 16%, and 19%. The horizontal edges are determined by requiring the tile area to be proportional to the percentage in the corresponding cell of the cross classification. We see that CBS, after a slow start, expanded with *Betty White* and continued with *Rafferty*; NBC is weak throughout; and ABC sustained itself with *Monday Night Football* in the last four time slots.

The whole mosaic in Figure 2 is constructed as follows:

1. The data are broken down by the day of the week. The seven large blocks (looking like horizontal stripes) are proportional in area to the percentages watching each day of the week. Since these blocks are equally long, their heights are proportional to the percentages watching each day. Spacings within a block occupy an area proportional to the block area, so both the area of the block in outline and the sum of areas of tiles within a block are proportional to the daily percentages.

2. For each day of the week, the block is broken down into 12 smaller blocks representing the 12 weeks in the three monthly periods; the width of each of these 12 blocks is proportional to the percentage watching that week, although the constant of proportionality varies between days. Likewise, the areas of the 12 blocks, including spacing within blocks, and the areas of the tiles within the 12 blocks, are proportional to the percentage watching that week. At this point, the 7×12 table represents the percentages watching in the 12 weeks on various days—it gives the relative percentages on days and the relative percentages over weeks given days. Areas of blocks correspond to percentages watching on a given day and week—for example, the fewest viewers watched on Thursday in the 4th week of November.

3. For each day and week, the data are divided into three blocks of height proportional to the CBS, NBC, and ABC percentages—here the conditional probability of network given the day and week is represented.

4. Finally, for each day, week, and network, the blocks are divided into six tiles of width proportional to the percentage watching at a particular time given the day, week, and network.

5. A different mosaic is obtained for each ordering of the variables.

6. Spacings between the blocks make computation more difficult and interpretation easier. The same relative spacing must be used at each level of the display so that areas with spacings included are proportional to areas without spacings. Consider, for example, the spacings between networks within each day by week block: The total space must be proportional to the height of the block, changing across day but not across week within day. Larger spacings are used for the initial variables in the dissection. We have used the same spacings between different categories of the same variable, except for the final variable, time of day. Zero spacings are used between time slots occupied by the same show. (This causes a slight embarrassment when one show occupies all six time slots, since the outline of the day-week block is no longer rectangular.) It is thus possible to vary spacings between categories of the same variable while retaining proportionality of block areas to corresponding percentages.

4. RESULTS

What we see in Figure 2 are the following:

1. From the heights of the blocks corresponding to days, Thursday, Friday, and Saturday have fewer viewers.

2. From the weeks blocks within each day, there are more viewers in February, fewer in May, and apparently a decline throughout the weeks of May. Thursday, in the 4th week of November, seems unusually low: That day is Thanksgiving Day.

3. From the networks blocks, within weeks and days, there is no really uniform pattern over the 12 weeks for

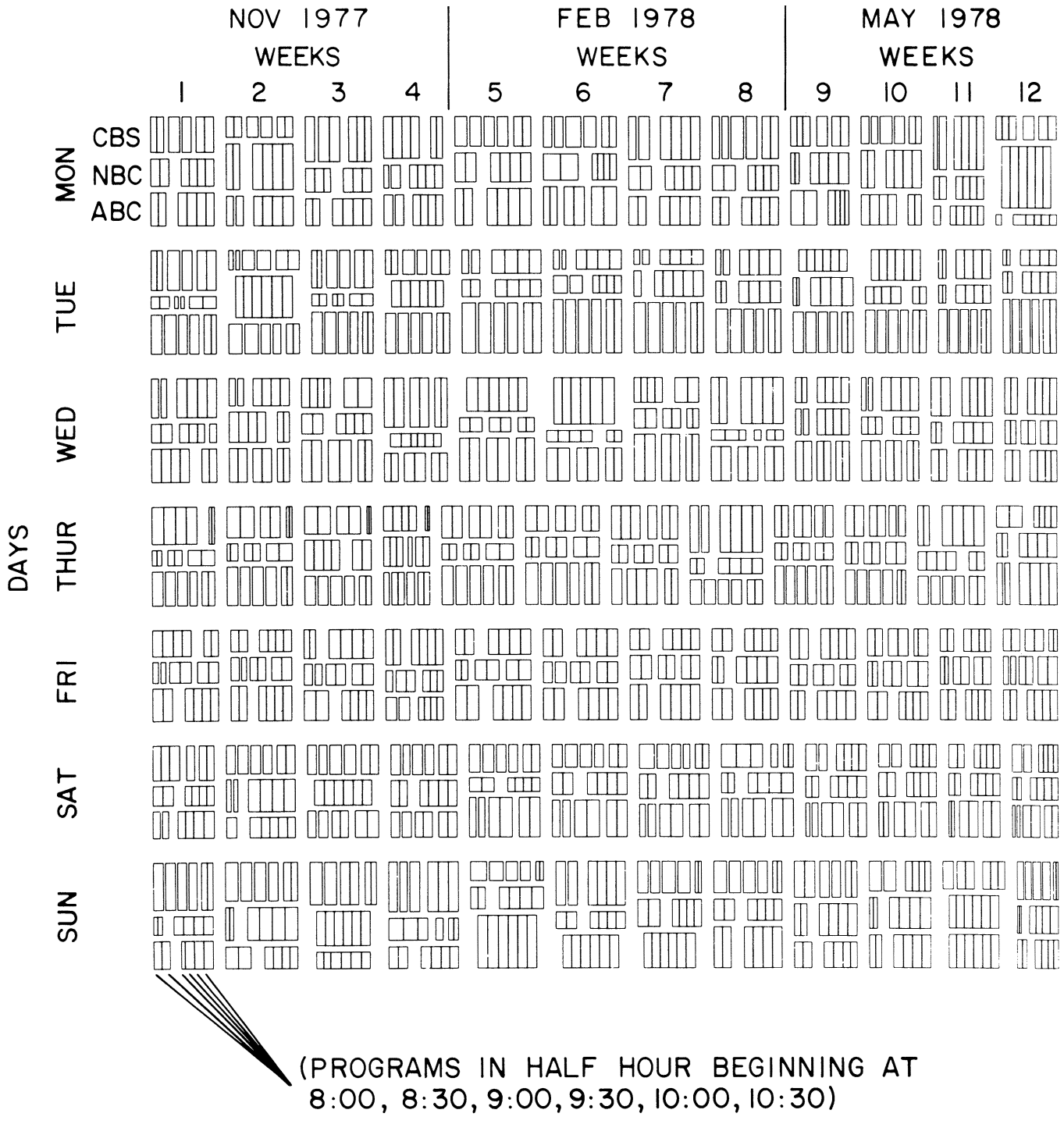

Figure 2. Mosaic of Nielson ratings broken down by days, weeks, networks, and time slots. The area of each rectangle is proportional to the number of Nielson households watching television on a particular day, viewing a particular network during a particular prime-time half hour.

the share of the audience among the three networks: NBC is nearly always third, with CBS and ABC exchanging first and second place; ABC dominates Tuesday nights. Results are fairly consistent within each of the three months—for example, CBS did well on Sunday nights in November but less well in later months. Perhaps this consistency could be represented by a further vertical dissection according to week of month. There are some surprising NBC blocks, such as Monday and Tuesday, during the second week of November. These blocks represent the *Monday Night Movie* and *The Big Event,* both specials taking two and three hours, respectively. Similarly, on Monday in the last week of May, NBC had a three-hour *Bob Hope Special.* The only other time NBC was first was for a *Saturday Night Movie* in the second week of November. Any analysis should treat these special programs differently from regularly scheduled programs. In contrast ABC has consistently large blocks on Tuesdays in February, with *Happy Days, Laverne and Shirley, Three's Com-*

pany, Soap, and *Family,* and they have nearly the same lineup in November and May. The other noteworthy ABC block is its *Silver Anniversary Special* on the first Sunday in February. They also do well with a two-hour movie in the fourth Sunday in February, and with a two-hour *Bing Crosby Special* on the fourth Thursday of May.

4. From the time of evening, within networks, weeks, and days, we see a consistent decline in number of viewers as the evening wears on, and we see considerable loyalty to the same show. For example, on Thursdays in November CBS's *The Waltons* (8:00–9:00) and *Hawaii Five-O* (9:00–10:00) do well consistently followed by a dying local show, *This Week.* The pattern is changed in February when *This Week* is replaced by *Barnaby Jones;* the abrupt change in the last week in May is caused by a two-hour showing of the weak NBA championship. Four strong Sunday shows on CBS in November are *Rhoda, On Our Own, All in the Family,* and *Alice,* followed by the less popular *Kojak.* On the fourth Sunday of November the pattern is changed when the last three shows are replaced by a movie. In May the same pattern appears, although *Kojak* has been replaced by an even less popular local news show called *This Week.* In the second week of May a movie was screened during the last two hours.

5. SUMMARY AND CONCLUSIONS

Mosaics are practicable by hand for moderate-sized tables, especially if no spacing is inserted. The only computations necessary are the frequencies of the first variable, the conditional frequencies of the second given the first, down to the conditional frequencies of the last given all previous variables. Intelligent choice of spacings can dramatize selected features of the table while still preserving faithful representation of counts proportional to areas of tiles.

[*Received October 1982. Revised July 1983.*]

REFERENCES

BERTIN, J. (1967), *Semiologie Graphique,* Paris: Gauthiers-Villars.

BISHOP, Y.M., FIENBERG, S.E., and HOLLAND, P.W. (1975), *Discrete Multivariate Analysis: Theory and Practice,* Cambridge, Mass.: M.I.T. Press.

GOODHARDT, G.J., EHRENBERG, A.S.C., and COLLINS, M.A. (1975), *The Television Audience: Patterns of Viewing,* Westmead, England: Saxon House.

HARTIGAN, J.A., and KLEINER, B. (1981), "Mosaics for Contingency Tables," *Proceedings of the 13th Symposium on the Interface Between Computer Science and Statistics,* ed. W.F. Eddy, New York: Springer-Verlag.

HEAD, S.W. (1972), *Broadcasting in America,* New York: Houghton Mifflin Co.