

# Comparing Partitions With Spie Charts

Dror G. Feitelson

School of Computer Science and Engineering

The Hebrew University of Jerusalem

91904 Jerusalem, Israel

## Abstract

Statistical graphics are important both in exploratory data analysis and in presenting results to non-professionals. Pie charts are often used to show a partition, but are hard to compare with each other. A spie chart is a combination of two pie charts: one sets the angles of the slices, and the other sets their areas, by manipulating the radius of each slice individually. This enables an easy comparison of the partitions represented by the two charts. It is useful for presenting change (e.g. a new division of a budget compared to a previous division) or risk (e.g. effected groups relative to their fraction in the population).

Keywords: Statistical graphics, Pie chart, Relative size

## 1. Introduction

The first pie chart is thought to have been drawn by William Playfair in 1801. Since then pie charts have become one of the most widely used graphics in the popular press and executive briefings. They are used to show partitions of a whole, e.g. how a budget is divided among ministries or departments, how seats in the parliament are divided among political parties, etc. This is done by dividing a circle into slices, such that the size of each slice represents one of the parts.

Despite their widespread use, pie charts have been unpopular with some professionals involved in statistical graphics (e.g. [Tufté, 1983, Cleveland and McGill, 1984]). This was based on the fact that it is harder to estimate area than length; it would therefore appear that estimating the relative height of bars along a common scale would be easier than estimating the relative sizes of slices of a pie [Cleveland and McGill, 1984]. However, in the special case of pie charts it is actually the angle that defines the area. In fact, pie charts have been shown to be an effective technique for conveying proportions, that is quickly perceived and leads to low absolute errors [Spence and Lewandowsky, 1991, Hollands and Spence, 1998].

Nevertheless, the use of pie charts suffers from two main drawbacks. One is that it is hard to estimate the relative size of similar slices; This is typically solved by labeling each slice with the percentage value that it represents. The other is that it is hard to compare two different partitions. In this case writing the percentages next to the slices is no better than simply writing them in a table to begin with.

## 2. Spie Charts

Spie charts allow for an easy comparison of two partitions by combining two pie charts into one. The first pie chart serves as a base, and its partition sets the angle of each slice. The second pie chart is superimposed on the first, using the same angles. Its partition is then expressed by changing the radius of each slice so as to reflect its relative size. Slices that now extend beyond the circle of the original pie chart indicate that their relative size has grown, while slices that are smaller than the original circle indicate that their relative size has shrunk. This provides an immediate and visually striking display of the change from the first partition to the second one, at the price of losing the easy comparison of slice sizes for the second partition.

More precisely, assume a base partition  $B = (b_1, b_2, \dots, b_n)$  and a second partition  $C = (c_1, c_2, \dots, c_n)$ , both having  $n$  parts. We want angles to reflect parts in the base partition. If angles are measured in degrees, we get

$$\alpha_i = 360 \frac{b_i}{\sum_{j=1}^n b_j}$$

In the base pie chart, the radius is a constant; we shall arbitrarily decide that it is 1. In the superimposed pie chart we set the radius of each slice to reflect its part in the second partition. Thus we have

$$\frac{\alpha_i}{360} r_i^2 = \frac{c_i}{\sum_{j=1}^n c_j}$$

which leads to

$$r_i = \sqrt{\frac{c_i / \sum_{j=1}^n c_j}{b_i / \sum_{j=1}^n b_j}}$$

It should be noted that using pie charts in which slices have different radii is not new. Florence Nightingale used such charts to show army casualties for successive months of the year in 1858 [Small, 1998], and they are also used in four-fold displays of  $2 \times 2$  tables [Friendly, 1994]. However, in these previous designs the angles are all equal. Different radii are also used in Kiviat graphs [Jain, 1991] and star graphs [Ward and Lipchak, 2000], but there the data refers to different categories rather than to a partition. Spie charts are unique in combining two pie charts into one, using both angles and radii.

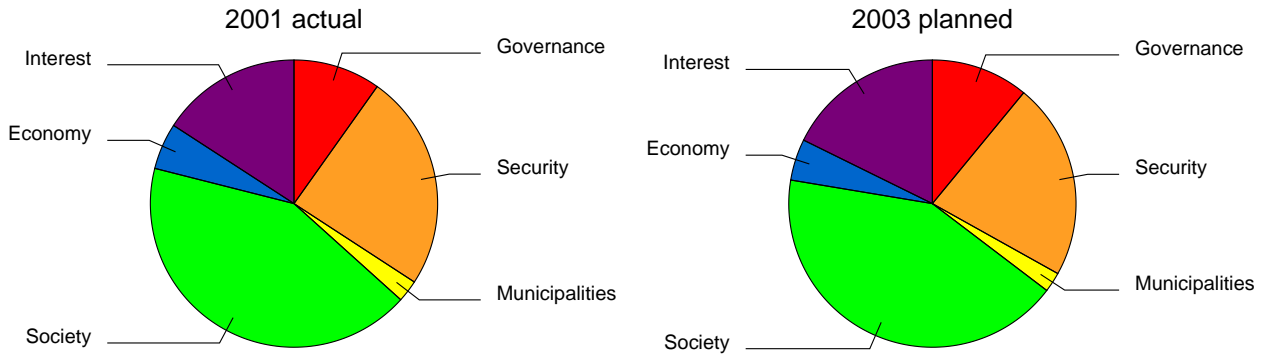


Figure 1: Comparison of the actual 2001 Israeli government budget and the 2003 budget plan, using conventional pie charts.

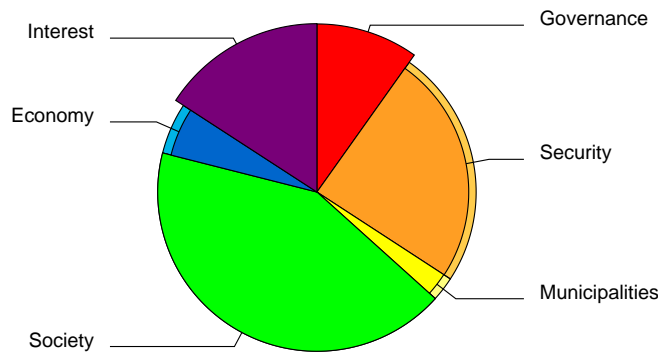


Figure 2: Comparison of the actual 2001 Israeli government budget and the 2003 budget plan, using a spie chart.

### 3. Examples

As examples of using spie charts, consider the following case studies. The first compares the top-level allocations in Israel's governmental budget plan for 2003 with the actual allocations for 2001. As the differences in allocations are relatively small, drawing two pie charts for the two allocations leads to graphs that are indistinguishable with the naked eye (Fig. 1). But using a spie chart the differences are brought out. In particular, this shows that a conservative government thought to be supportive of small government and security, and against social programs, actually increased the budget for governance, reduced the budget for security, and left the social programs as they were (Fig. 2).

The second case study compares the election results to the Israeli Knesset (parliament) in 1999

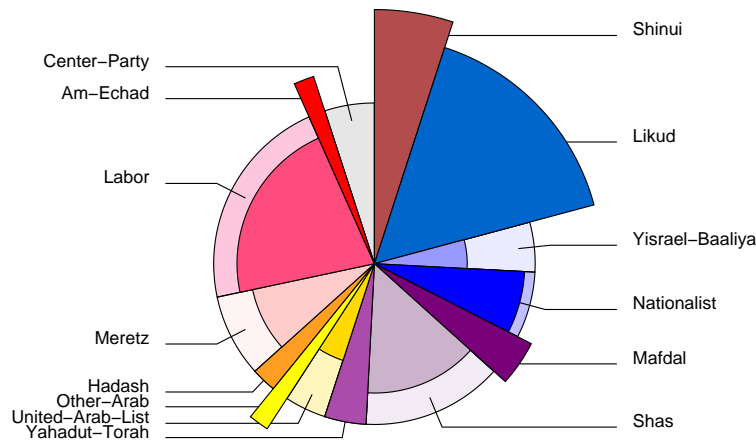


Figure 3: Comparison of the 2003 election results to the Israeli Knesset with the 1999 results, showing the collapse of the left wing and Shas and the strengthening of the right wing.

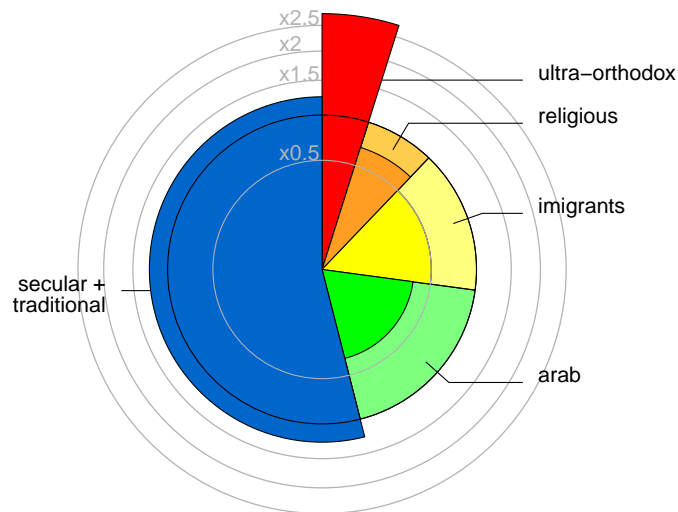


Figure 4: Relative representation of specific sectors in the Israeli Knesset.

and 2003. The Knesset typically includes representatives from about 15 different parties. In the 1999 elections it was dominated by the left-wing Labor party (26 seats out of 120), the right-wing Likud (19 seats), and the religious Shas party (17 seats). The base pie chart depicts right-wing parties to the right in shades of blue, left wing parties to the left in shades of red, and religious (purple) and Arab (yellow) parties at the bottom. The 2003 elections saw a big shift in power, with Labor dropping to 19 seats, Likud doubling to 38, and Shas dropping to 11. But the biggest story was the Shinui party which grew from a mere 6 seats to being third largest with 15 (Fig. 3).

Another view of the 2003 election results is shown in Fig. 4. Quite a few of the Israeli political parties position themselves as catering for specific sectors of the society that may have special

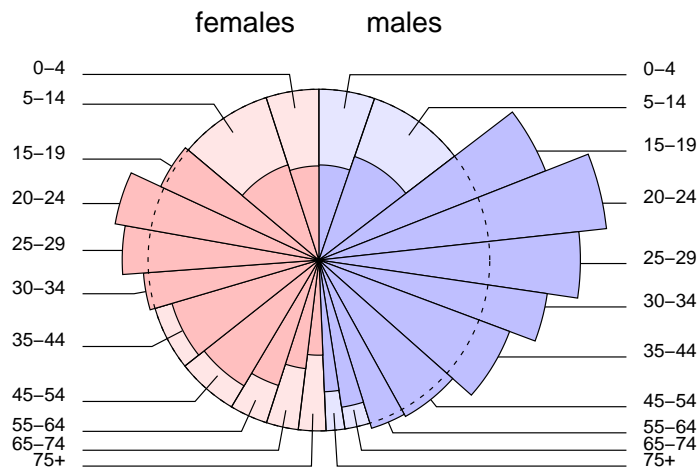


Figure 5: *Distribution of road accident casualties by age and gender, relative to the size of the population. Data from the Israel Bureau of Statistics, 2002.*

needs. But as the figure shows, their representation in the Knesset does not match the relative sizes of these sectors in the population. The ultra-orthodox parties have a much larger share than would be expected, partly because of the strict voting regiment imposed by their rabbis, and partly due to their appeal to traditional Jews who are themselves not orthodox. Adding gridlines (actually, circles) that represent different areas, we easily see that they surpass their population share by a factor of more than 2.5. The representation of Arabs, on the other hand, is much lower than their share in the population, largely due to their reluctance to participate in the Israeli elections; in fact, they achieve less than half the representation that they should. The share of immigrant-oriented parties is also rather low, which might be seen as an indication of successful assimilation, where immigrant prefer parties that present a general agenda rather than an immigrant-specific agenda.

The third case study concerns road casualties. In this case the base pie chart is the partition of the general population of Israel into age and gender groups. The superimposed pie chart uses the same partition for the population of road casualties (Fig. 5). Obviously the main age group hit is 20–24, and males are much more prone to being involved in accidents than females. In fact, males are over-represented in the casualties population for all ages from 15 to 64, whereas for females this is true only for the ages of 15 to 34.

Fig. 6 shows this data again, this time dividing each slice into three stacked segments. The segments represent pedestrians, car riders, and bicycle or motorcycle riders. The first division is roughly circular, indicating that pedestrian casualties are roughly proportional to the population size. The main exception is old people who are slightly overrepresented, probably indicating that they are more prone to be hit; it is also noticeable that babies are slightly under-represented, and

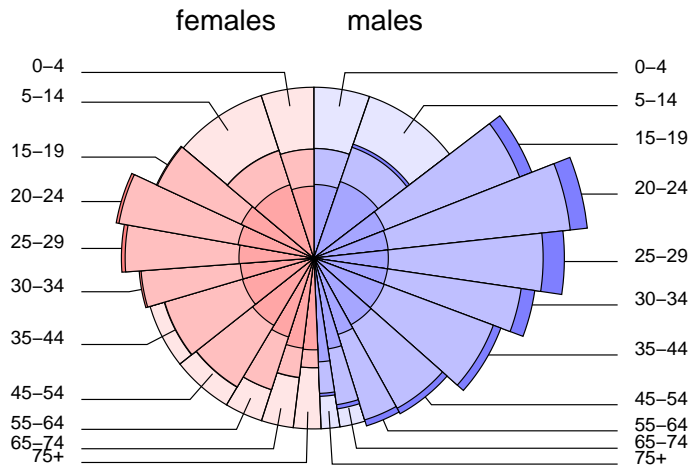


Figure 6: *Distribution of road accident casualties by age and gender, with division into pedestrians, car riders, and bicycle or motorcycle riders from the center outwards.*

boys 5–14 are slightly overrepresented. The third division shows that only a small fraction of accident casualties are bicycle or motorcycle riders, and that they are overwhelmingly male.

#### 4. Discussion

The above examples show the effectiveness of spie charts in comparing partitions, or rather, in showing the change from a base partition to a derived partition. A quantification of the degree of change can be achieved by using grid lines as in Fig. 4. But perception of the derived partition suffers. Thus, if the derived partition itself is of prime interest, a separate pie chart should be used. But if the difference is the focus of the story, this is brought out more dramatically by a spie chart.

A limitation of spie charts is that both partitions must be into the same groups. This need not always be so, and indeed it was not so in the case of the Knesset elections: the set of represented parties in 2003 was not the same as in 1999. Parties that appear only in the base partition can be rendered, but then it is not clear whether they do not appear in the second partition or maybe they appear with exactly the same relative size. This is the case for the Center Party in Fig. 3. Worse, parties that were absent from the base partition cannot be drawn as they would have an angle of zero. This can be side-stepped by grouping parties into an “other” category. For example, the “Other Arab” category in the figure includes the National Democratic Alliance from 1999 and the Balad movement from 2003.

A related problem is that partitions that shrink considerably may still be perceived to be rather important, as their angle is based on the original partition. This happens for the United Arab List and Yisrael Baaliya, which both got only 2 seats in 2003, but look much more important

than Hadash or Am-Echad who got 3. A related problem is that humans are much better able to gauge lengths than areas [Cleveland and McGill, 1984]. While this is immaterial for conventional pie charts, in which all slices have the same radius, it becomes important for spie charts. For example, a spie chart of election results does not allow for easy assessment of how many seats each party actually got. As with the original pie charts, this can be rectified by including the actual numbers in the slice labels. As it is known that humans tend to underestimate larger areas [Spence and Lewandowsky, 1991], it might also be appropriate to adjust the calculated radius so that the perceived area (rather than the real area) match the desired percentage.

## References

- [Cleveland and McGill, 1984] W. S. Cleveland and R. McGill, “*Graphical perception: theory, experimentation, and application to the development of graphical methods*”. *J. Am. Stat. Assoc.* **79(387)**, pp. 531–554.
- [Friendly, 1994] M. Friendly, *A Fourfold Display for 2 by 2 by K Tables*. Technical Report 217, York University, Psychology Dept. URL <http://www.math.yorku.ca/SCS/Papers/4fold/>.
- [Hollands and Spence, 1998] J. G. Hollands and I. Spence, “*Judging proportions with graphs: the summation model*”. *Applied Cognitive Psychology* **12**, pp. 173–190.
- [Jain, 1991] R. Jain, *The Art of Computer Systems Performance Analysis*. John Wiley & Sons.
- [Small, 1998] H. Small, “*Florence Nightingale’s statistical diagrams*”. URL <http://www.florence-nightingale.co.uk/small.htm>.
- [Spence and Lewandowsky, 1991] I. Spence and S. Lewandowsky, “*Displaying proportions and percentages*”. *Applied Cognitive Psychology* **5**, pp. 61–77.
- [Tufté, 1983] E. R. Tufté, *The Visual Display of Quantitative Information*. Graphics Press.
- [Ward and Lipchak, 2000] M. O. Ward and B. N. Lipchak, “*A visualization tool for exploratory analysis of cyclic multivariate data*”. *Metrika* **51(1)**, pp. 27–37.