# Data on the Distribution of Cancer Incidence and Death across Age and Sex Groups Visualized using Multilevel Spie Charts

Dror G. Feitelson
School of Computer Science and Engineering
The Hebrew University of Jerusalem, 91904 Jerusalem, Israel

## Abstract

Cancer incidence and death statistics are typically recorded for multiple age and sex brackets, leading to large data tables which are difficult to digest. Effective visualizations of this data would allow practitioners, policy makers, and the general public to comprehend the data more readily and act on it appropriately. We introduce multi-level spie charts to create a combined visualization of cancer incidence and death statistics. Spie charts combine multiple pie charts, where the base pie chart (representing the general population) is used to set the angles of slices, and the superimposed ones use variable radii to portray the cancer data. Spie charts of cancer incidence and death statistics from Israel for 2009—2011 are used as an illustration. These charts clearly show various patterns of how cancer incidence and death distribute across age and sex groups, illustrating (1) absolute numbers and (2) rates per 100 000 population for different age and sex brackets. In addition, drawing separate charts for different cancer types illustrates relative mortality, both (3) across cancer types and (4) mortality relative to incidence. Naturally this graphical depiction can be used for other diseases as well.

**Keywords**: Cancer incidence data, Mortality data, Age/sex distribution, Spie chart, Data visualization

**What is new**?

- A graphical method to present detailed cancer incidence and death statistics in lieu of tables, thereby making the data more accessible and useful.
- Multi-level spie charts which combine both incidence and death data and relate them to the base population.
- An illustration of the methodology showing how patterns are identified.

## 1. Introduction

Humans have always been interested in health, sickness, and death. Health administrators in particular collect and use health and death statistics on a daily basis. Such statistics typically come in either of two forms. Professionals use multiple tables with hundreds of numbers, providing data about multiple factors in excruciating detail. The general public typically sees only artistic infographics, in which very few numbers are illustrated. There have been no effective visualizations of detailed data, despite the fact that graphic visualization of data contributes significantly to making it understandable and actionable, especially when large amounts of data are involved (e.g., [1,2]).



Published in *Journal of Clinical Epidemiology* http://dx.doi.org/10.1016/j.jclinepi.2015.11.001

To appreciate the problem in the specific case of cancer statistics, consider the many different bisections and normalizations that are often used. Overall cancer rates for the entire population may be used for comparison of countries, but are not very meaningful because they change considerably for different ages. There are also differences between the sexes. Data is therefore typically partitioned into multiple age brackets for either sex. These can then be normalized relative to an international standard population, thereby enabling valid international comparisons of real differences in cancer incidence and avoiding artifacts due to differences in population structure. Another choice is to present the data in absolute numbers – thereby placing the focus on the number of individuals affected – or in cases per 100 000 residents – thereby emphasizing hazard rates. In many cases this is also done separately for different ethnic groups, to reflect differences in genetic background. Finally, the whole thing can be repeated for each type of cancer separately.

Spie charts are a statistical graphic based on pie charts which facilitates the comparison of two partitionings [3]. In particular, it has been suggested that this be used to portray hazards for different sex and age groups, such as the risk of being involved in a traffic accident. We extend this to multi-layer spie charts, and use them to portray both the cancer incidence rate and the cancer death rate for different sex and age combinations. In addition, both the relative absolute numbers of cases of different cancer types and the hazard rates per 100 000 residents are shown, using slice areas and radii respectively. It is expected that using such charts it will be much easier to appreciate the patterns in cancer data at a glance.

## 2. Motivating Example

As a motivating example consider cancer incidence and death data for both sexes and multiple age brackets. An example of such data covering the years 2009-2011 from Israel is shown in Table 1. The data in this table is the average number of cases per year for each age and sex combination. All data is from the Israel Central Bureau of Statistics; a span of three years was used to reduce the effect of yearly fluctuations due to small numbers. If in addition we have a similarly-sized table with the total population in each sex and age bracket, we can divide the cancer cases data by the population data to obtain hazard rates, that is the number of cases per 100 000 people in each sex and age group. The result of doing so is shown in Table 2.

| | incidence | | death | |
|---|---|---|---|---|
| ages | males | females | males | females |
| 0-14 | 177.00 | 138.67 | 31.00 | 22.33 |
| 15-24 | 172.33 | 194.67 | 27.33 | 18.00 |
| 25-34 | 314.67 | 841.67 | 41.33 | 45.67 |
| 35-44 | 488.67 | 1387.33 | 100.67 | 158.33 |
| 45-54 | 1171.33 | 2191.00 | 332.00 | 396.33 |
| 55-64 | 3138.67 | 3350.33 | 932.33 | 867.00 |
| 65-74 | 3567.33 | 3013.33 | 1231.67 | 1030.67 |
| 75+ | 3730.33 | 3706.33 | 2429.00 | 2533.33 |

Table 1: Cancer incidence and death data from Israel. Total cases averaged over the years 2009 to 2011.

| | incidence | | death | |
|---|---|---|---|---|
| ages | males | females | males | females |
| 0-14 | 16.19 | 13.34 | 2.84 | 2.15 |
| 15-24 | 28.69 | 33.74 | 4.55 | 3.12 |
| 25-34 | 56.05 | 150.41 | 7.36 | 8.16 |
| 35-44 | 103.44 | 289.39 | 21.31 | 33.03 |
| 45-54 | 306.47 | 543.81 | 86.87 | 98.37 |
| 55-64 | 935.80 | 912.65 | 277.98 | 236.18 |
| 65-74 | 1977.46 | 1414.71 | 682.74 | 483.88 |
| 75+ | 2562.04 | 1736.80 | 1668.27 | 1187.13 |

Table 2: Cancer incidence and death hazard rates (cases per 100 000 individuals). This is obtained by dividing the data in Table 1 by the population of each sex and age bracket.

Given such data, the typical way to visualize it would be some graph where the horizontal axis represents age, and the number of cases or the hazard rate is shown for both sexes as a function of age. However, it is not easy to come up with a graph that shows all the data. One problem is that the scales have different units: numbers of cases are not in the same scale as cases per 100 000 residents. Therefore one has to compromise and show either number of cases or hazard rates, but not both. Wainer suggests that in this case a line plot can be rather effective, for example showing how the hazard rate changes with age for the two sexes, and even comparing the rate at each age bracket to the population-wide average by adding a horizontal line denoting the average [4]. An example of such a graph is shown in Figure 1, where both cancer incidence rates and cancer death rates are plotted for both sexes. Obviously cancer incidence and death grow steeply with age, but deaths tend to occur about 10 years later. The patterns of cancer incidence are different for the two sexes: females are more

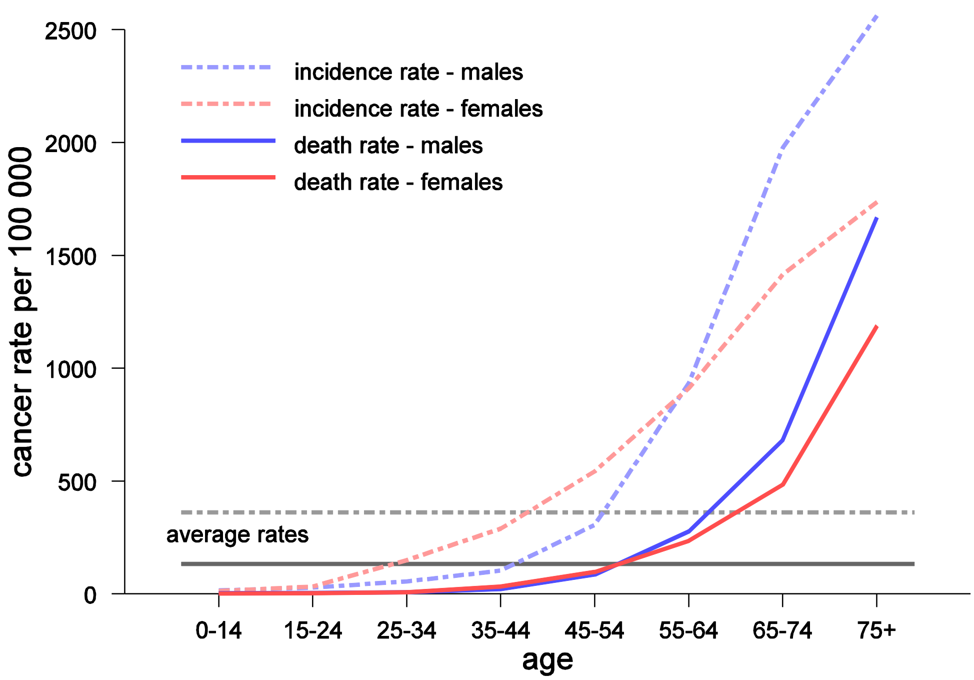


Figure 1: Rendition of cancer incidence and death rates using line plots.

susceptible at younger ages, and males at older ages. In particular, women above age 45 already have a higher than average incidence rate, but for men this happens only at age 55. As expected (and shown below), this is a result of the prevalence of breast cancer in women.

However, this is only half of the data, namely the hazard rates. So while we can see that the incidence and death rates are much higher for older people, we can't tell how many cases are actually involved. To find the number of cases we need to multiply the hazard rate by the population size in each age bracket, and draw a separate line plot where the vertical axis represents cases rather than rates.

So can all the data be shown together? It can by using spie charts. The idea is simple: given that the number of cases is the product of the hazard rate and the population size, we can in principle represent the hazard rate along one dimension, the population along another dimension, and then obtain a representation of the total cases as an area. In spie charts the population dimension is an angular one.

## 3. Basic Spie Charts

Let us start with a general description of spie charts.

Spie charts allow for an easy comparison of two partitions by combining two corresponding pie charts. 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 that its area reflects the desired relative size. Slices that now extend beyond the circle of the original pie chart indicate that their relative size has grown in the second partitioning, 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. The total area of the superimposed slices is the same as the total area of the base pie chart.

More precisely, assume a base partition $B = (b_1, b_2, \cdots, b_n)$ and a second partition $C = (c_1, c_2, \cdots, c_n)$, both having $n$ corresponding parts. The sums $b_1 + b_2 + \cdots + b_n$ and $c_1 + c_2 + \cdots + c_n$ need not be the same, and in fact in most cases (including ours) they are not, as we are only interested in the relative sizes of the parts. The base pie chart is just a normal pie chart of the base partition. In such a pie chart the angles of the slices reflect the parts in the partition. Denoting the angle of the $i$th slice by $\alpha_i$, as measured in radians, we get

$$\alpha_i = \frac{b_i}{\sum_{j=1}^{n} b_j} 2\pi$$

for $i = 1, \cdots, n$. In the base pie chart, the radius is a constant; we shall arbitrarily decide that it is 1, so that the total area of the pie is $\pi$.

The superimposed pie chart has the same total area $\pi$, but we set the radius of each slice $r_i$ so that the area of the slice reflects its part in the second partition while using the same angle as in the base pie. Thus the equation for the area of the $i$th slice is

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

which leads to

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

Jacobs shows how to do all this in R to generate spie charts from data [5].

Figure 3 shows the result of applying this derivation to the cancer incidence data introduced above (for the moment we deal only with incidence; the death data will be added later). The base pie chart is shown in Figure 2. This portrays the general population of Israel and how it was partitioned into sex and age groups in 2010. This is the data that was used to normalize the values in Table 1 and derive Table 2. Looking at this pie chart we can observe that the population of Israel is heavily skewed towards young ages: the topmost slices – representing the youngest ages – are the largest, and they grow progressively narrower as we move downwards toward older ages. (But note that the first slice is especially wide partly due to its representing a span of 15 years, as opposed to 10 years for all other slices except the last one.)

The superimposed pie chart (in Figure 3) uses the same partition for the population of newly detected cancer cases in the years 2009 to 2011, namely the incidence data shown in Table 1. Thus what we are doing is to compare the partitioning of the cancer incidence cases across age and sex groups to the partitioning of the general population across these same age and sex groups. As a result we can easily see which age and sex groups are under-represented among cancer victims, meaning that cancer incidence is relatively low in these age and sex groups: they are the groups where the slices are shorter than in the base pie. And

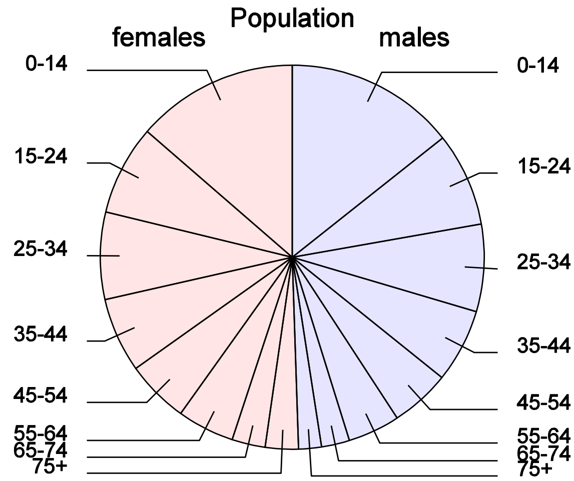


Figure 2: Base pie chart depicting the general population.

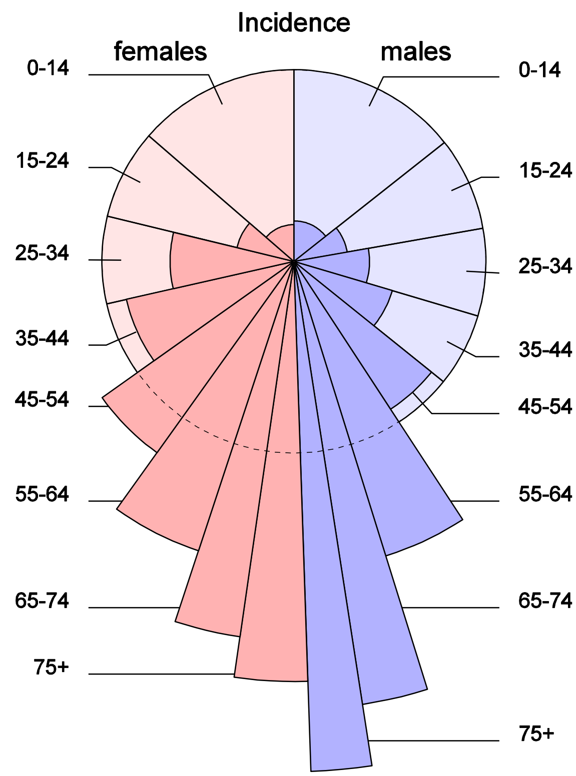


Figure 3: Example of a spie chart, showing cancer incidence rates for sex and age groups relative to the general population.

contrariwise, groups with slices that extend beyond the original pie are over-represented among cancer victims, meaning that people in these groups suffer from a higher cancer rate than in the general population. Thus the circumference of the original pie serves the same role as the line showing the population-wide average in Figure 1. Basing the rendition on a pie chart emphasizes the fact that we are looking at parts of a whole – all new cancer cases relative to the whole population of the country.

So what can we see in the spie chart of Figure 3 relative to the line plots of Figure 1? The main observation, that cancer incidence grows steeply with age, is obvious. It is also easy to see that women up to age 44 and men up to age 54 are underrepresented among cancer victims, whereas above these ages they are overrepresented, and generally that females are more susceptible at younger ages and males at older ages. But the comparison of men to women is slightly different: in the spie chart the differences manifest themselves as asymmetry in the shape of the overlay, instead of the direct comparison that is possible with line plots.

The main difference, however, is that the spie chart shows the actual number of cases in each age and sex bracket in addition to the hazard rate. The incidence rate per 100 000 residents or hazard is encoded in the radii of the different slices. To see this, note that $r_i = f(c_i/b_i)$, where $c_i$ is the number of cases in the $i$th sex and age group, and $b_i$ is the size of this group. However, the function $f$ includes a square root, so the scale is not linear: doubling the radius represents a four-fold increase in hazard. (Concentric grid lines may be used to enable observers to read out the factor by which each slice differs from the average hazard rate for the entire population.) But in exchange for this non-linear scale, we add the information about the actual number of cases in each age and sex group. This information is encoded by the areas of the superimposed slices, and was missing from the line plots of Figure 1. For example, in Figure 1 we can see that the incidence of cancer in males aged 75 and above is about 2.5 times higher than in males ages 55-64, but we have no way of knowing how many people are actually affected. In the spie chart of Figure 3 we can estimate that the numbers are similar, because the slice for the 55-64 age bracket, while shorter, is also much wider (namely, there are many more men aged 55-64 than men aged 75 and above). As a result the areas of the two slices appear similar. And checking back to the original data in Table 1, we find that the number of cases for men aged 75 and above is indeed only 19% higher than for men aged 55-64, despite the much higher hazard rate.

## 4. Combining Incidence and Death Data

Cancer incidence rates do not tell the whole story. Death rates are different for different types of cancer, and naturally also depend on age and sex. We can add this to the spie chart by adding another overlay on top of the previous one, using darker shading, creating a multi-level spie chart.

In technical terms, we now have a third partition $D = (d_1, d_2, \cdots, d_n)$ in addition to the base partition $B$ and the original overlay partition $C$. Assume that the partitions $C$ and $D$ are related, as they are in our case, where death due to cancer necessarily follows the incidence of cancer. We can then facilitate the comparison of the two superimposed partitions by using the sum total of the first one in the normalization of both. Thus the radii of the slices in the second superimposed partition will be calculated as

$$r'_i = \sqrt{\frac{d_i / \sum_{j=1}^{n} c_j}{b_i / \sum_{j=1}^{n} b_j}}$$

As a result of this normalization the total area of all the slices of the second superimposed partition will not be the same as that of the first one and the base partition. Instead, its area will reflect the size of this partition (in our case, total deaths) relative to the previous one (in our case, total incidence). In other words, the total area of the second overlay will be $\frac{d_1+d_1+\cdots+d_n}{c_1+c_2+\cdots+c_n}$ the area of the first. To emphasize this, a solid circle is added to show the area of the second superimposed partition more clearly. As death rates are naturally lower than incidence rates, this circle is smaller than the base pie. It also denotes the average rate for the second overlay, just like the circumference of the base pie does for the first overlay.

The resulting multi-level spie chart is shown in Figure 4. This shows, for example, that the distributions of death ages for the two sexes are quite similar, much more so than the distributions of incidence ages, as manifest by the fact that the darker overlay is much more symmetrical than the lighter one (of course, this could also be seen in Figure 1). However, men still have a higher death rate at the very highest ages.

It should be noted that this specific graphic shows the number of new cases and the number of deaths in the same three years. Naturally most of the deaths that occurred in this period do not correspond to new cases in this same period. In fact, many of the deaths at high ages probably correspond to diagnoses that had occurred at earlier ages many years before. Therefore, in principle, slices of the second overlay (deaths) may extend beyond and occlude slices of the first overlay (incidence). However, this does not actually happen in our data. It is also possible to consider alternative normalizations, for example showing a prediction of how many of the new cases will eventually die rather than how many actually did die in each age bracket, thereby guaranteeing that $d_i \leq c_i$ for all $i$.

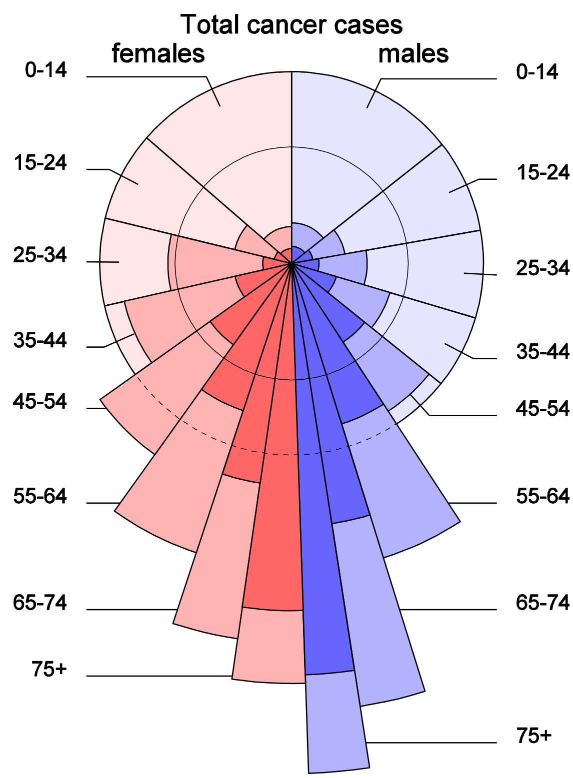


Figure 4: Double spie chart with both cancer incidence rates and cancer death rates superimposed on sex and age groups relative to the general population.

## 5. Comparing Multiple Locations

The above charts were concerned with total incidence of cancer and deaths from cancer. It is also possible to draw them for a specific type of cancer. Here we have two options. The first is useful when we want to focus on a certain type of cancer in isolation. In this case the top slices will reflect the partitions of this type's incidence and deaths, normalized by the total incidence of this type only. The second is when we want to consider several types of cancer together. In this case we draw multiple charts, each with overlaid slices reflecting the incidence and deaths from one type of cancer, but normalizing by the total incidence from all types of cancer.

Using a common normalization leads to slice sizes that are comparable across spie charts depicting different cancer types. Moreover, we can add a dashed circle to reflect the total incidence of each type, and a solid circle to reflect the total deaths of each type. The area of the dashed circle relative to the base pie then reflects the fraction of all cancer incidence cases that are of this specific type. The area of the solid circle reflects the deaths from this type of cancer.

A nice feature of these two circles is that they allow us to visualize two distinct interpretations of the term “death rate”. The size of the solid circle (relative to the base pie) reflects the death rate from this type of cancer, meaning the fraction of all cancer patients who died from it. The similarity of the two circles to each other reflects the death rate for those who have contracted this specific cancer: the closer the circles are to each other, the higher the fraction of people who contracted it who die from it.

An example using ten multi-level spie charts to portray ten of the more common types of cancer is shown in Figure 5. This is easily seen to illustrate various well-known properties of different types of cancer, such as the following:

1. Breast cancer has by far the highest incidence rate, as indicated by the dashed circle which is larger than for any other type of cancer. Colon and rectum cancer and prostate cancer also have relatively high incidence rates.
2. Despite its high incidence rate, the total number of deaths due to breast cancer is lower than those of colon and rectum or lung cancers, as indicated by the solid circle which is smaller.
3. The high incidence of breast cancer at relatively low ages is the reason women are much more susceptible to cancer than men in the 35—54 age range.
4. The most deaths occur as a result of lung and colon and rectal cancers, as indicated by these cancers having the largest solid circles.
5. Cancers of the lung and ovary are the most deadly if you contract them, with Stomach and leukemia following them. This observation is based on the fact that the solid circle is nearly as large as the dashed one. Prostate cancer appears to be the least deadly as the difference between the circles is the largest.
6. Malignant lymphoma and Leukemia are the only cancers with significant incidence in children. In other cancer types, there are no overlaid slices at young ages.
7. Excluding cancers of sexual organs, cancers of the bladder and lung are the most skewed, with much higher incidence and death rates in men, especially above the age of 55.
8. In some cancers, notably those of the prostate and bladder, deaths are concentrated at the oldest

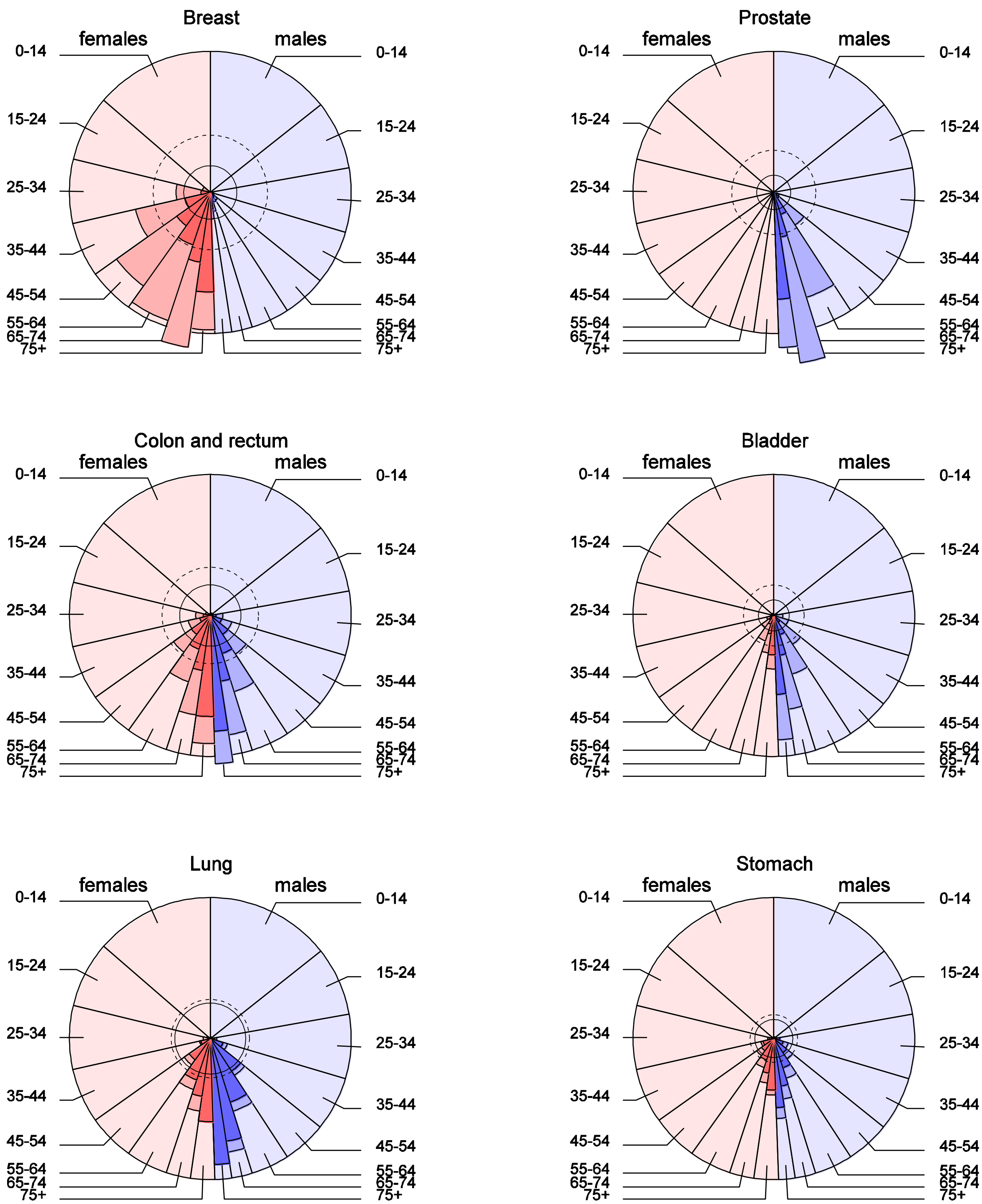


Figure 5: Spie charts comparing distributions of incidence and death rates for different sexes and age brackets for different cancer locations.

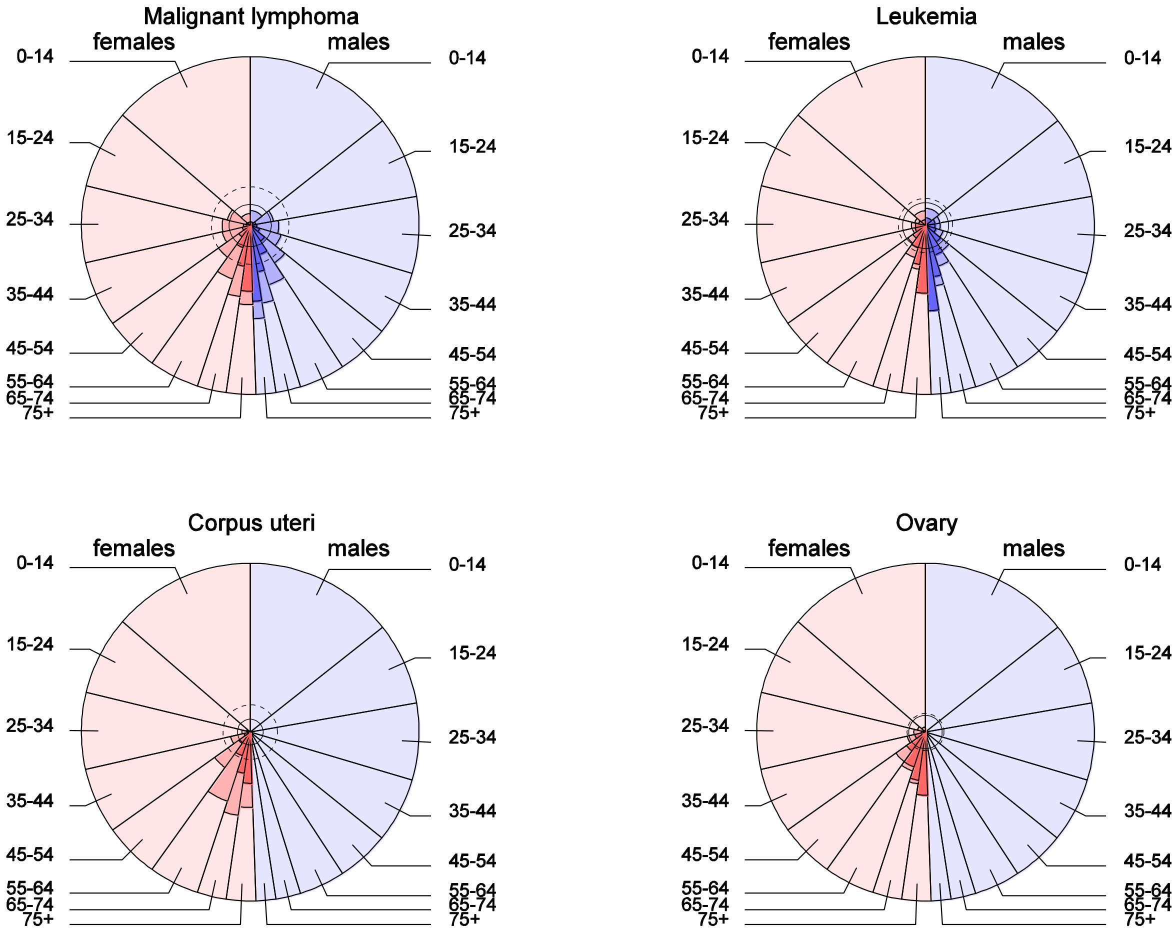


Figure 5: (continued)

ages, despite incidence occurring many years earlier, implying that patients live for many years with these cancers. In lung cancer, on the other hand, deaths are spread over the same ages groups as incidence.

9. In breast and prostate cancer incidence does not grow monotonically with age: the incidence rate drops slightly for the most elderly.

These observations will probably not surprise practitioners in the field, but the visualization using spie charts makes all of them much more self-evident than they appear when studying tables of data.

## 6. Discussion

Spie charts are a form of polar diagrams, and in particular are related to polar area charts. These are sometimes also called polar histograms, and consist of a pie with a certain number of slices that all have the same angles but different radii. Perhaps the most famous use of such charts was made by Florence Nightingale to show army casualties for successive months of the year in the Crimean War in 1854-1856

[6]. However, they are more appropriate when the data can in fact be expected to repeat itself in a cyclic manner, as in the case of infant mortality as a function of months of the year [7] or crime events as a function of hours of the day [8]. Alternatively, they may be appropriate to depict angular data, such as the prevalence of winds from different directions, or for four-fold displays of $2 \times 2$ tables [9]. Other types of polar diagrams use lines instead of slices; these include Kiviat graphs [10] and star (or "radar") plots [11,12], which are typically used to illustrate multivariate categorical data for different elements in a set, e.g. the features of different brands of cars. The lines then connect points along multiple radial axes, but again, the angles between these axes are all the same.

Spie charts are unique in using both angles and radii (and as a result, also area) to convey data, and were designed specifically for comparing one partition to another. Thus the use of a polar format is not just to create a recognizable pattern, but also to emphasize that we are dealing with parts of a whole. For example, in our use of spie charts to display cancer data, the rendition is less "cases as a function of age" than "distribution of cases across age and sex groups". Spie charts have been shown to be superior for displaying multivariate healthcare data [13], and previous use includes showing the distribution of healthcare expenses by age, which is important in an aging society [5]. They have also been used, for example, to illustrate shifting results in successive multi-party elections, the execution of a budget relative to the original budgetary plan, and to compare the distribution of publications across fields of study in a certain institution relative to the whole country.

A possible criticism of spie charts is that, as in pie charts, they suffer from the fact that estimating areas is much less precise than estimating lengths [14], and in particular, that people tend to underestimate large areas [15]. And in spie charts it may be especially hard to estimate the relative sizes of slices that have different angles, radii, and orientations. This problem has prompted some influential scholars to recommend that even simple pie charts be avoided [1,14]. Fenton parrots these arguments and has a whole category in his blog devoted to pie chart hatred [16], including a post that calls spie charts "stupidly complex". And indeed, many pie charts are very bad for various reasons as Fenton's examples demonstrate. But the argument that they are bad because they do not support accurate reading of values is debatable.

First, studies have shown that pie charts are actually quite effective for conveying proportions, and lead to quick perception with low absolute errors [15,17]. Second, pie charts should indeed not be used if your goal is to find precise numbers. Numbers are obviously best read from tables. But when the data becomes complex with multiple connections and relationships, a graphical rendition is preferred [18]. The reason is that graphs allow observers to appreciate patterns at a glance, which are impossible to see when the data is presented in other (especially tabular) forms. Spie charts showing hazard rates for age and sex groups have already been found to exhibit some easily recognized patterns: a "bowtie" where the greatest danger of being hurt in traffic accidents occurs for ages 15-34 for both sexes [3], a "horse-show award ribbon" where the greatest risk of cancer occurs at old ages (this paper), a "leaf" where healthcare costs are biggest for old ages but also high for babies under 1 year of age [5], and a "wing" where the vast majority of casualties in armed conflict are young men. At a finer granularity, one can observe the symmetry (or asymmetry) between males and females, degrees of concentration in a narrow age range as opposed to a spread across a wider range, and more. Wainer, in his inspection of pie and spie charts, writes that “I am forced to concede that, for some purposes, this variation on the pie theme works” [4].

It is suggested that multi-level spie charts are likewise useful. In particular, multi-level spie charts, as introduced in this paper, enable the comparison of two partitions to a base partition. And using a set of double spie charts as in Figure 5 shows all of the following:

- The distribution of the incidence of new cancer cases of each type with respect to age and sex groups.
- The distribution of deaths due to each type of cancer with respect to age and sex groups.
- The susceptibility of different age and sex groups to various types of cancer, as reflected by the relative sizes of corresponding slices in the different charts.
- The relative incidence rate of each cancer type as a fraction of all cancers, reflected by the area of the dashed circle (or of the colored slices) relative to the base pie.
- The relative incidence rates of different cancer types, reflected by the areas of the dashed circles (or the colored slices) relative to each other.
- The relative death rate of each cancer type, reflected by the area of the solid circle relative to the dashed circle (or the darker colored slices relative to the lighter colored slices).
- The relative death rates of different cancer types, reflected by the areas of the solid circles (or the darker colored slices) relative to each other.

No other graphical device allows all this data to be shown together so concisely allowing patterns to be observed.

## Acknowledgements

This study was supported in part by an IBM Faculty Award.

## References

1. Tufte ER. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press; 1983.
2. Ware C. *Information Visualization*. San Fransisco, CA: Morgan Kaufmann; 2nd ed., 2004.
3. Feitelson DG. *Comparing partitions with spie charts*. Technical Report 2003-87, Hebrew University, Dec 2003.
4. Wainer H. Pies, spies, roses, lines, and symmetries (Visual Revelations column). *Chance* 2010; 23(4):58-61.
5. Jacobs A. Visualizing Health Expenditure using Spie Charts (and R). URL http://unconj.ca/blog/visualizing-health-expenditure-using-spie-charts-and-r.html/ Oct 2014.
6. Small H. *Florence Nightingale's statistical diagrams*. URL http://www.florence-nightingale-avenging-angel.co.uk/GraphicsPaper/Graphics.htm, Mar 1998.
7. Levasseur E. La statistique graphique. *J. Stat. Soc. London*, Jubilee volume, pp. 218-250, Jun 1885. (graph on page 232)
8. Brinton WC. *Graphic Presentation*. NY: Brinton Assoc., 1939. (graph on page 85)
9. Friendly M. *A fourfold display for 2 by 2 by K tables*. Technical Report 217, York University, Psychology Dept., Jan 1995.

10. Jain R. *The Art of Computer Systems Performance Analysis*. New York, NY: Wiley; 1991.
11. Ward MO and Lipchak BN. A visualization tool for exploratory analysis of cyclic multivariate data. *Metrika* 2000; 51:27-37.
12. Saary MJ. Radar plots: a useful way for presenting multivariate health care data. *J Clin Epidemiol*. 2008; 61(4):311-7
13. Stafoggia M, Lallo A, Fusco D, Barone AP, D'Ovidio M, Sorge C, and Perucci CA. Spie charts, target plots, and radar plots for displaying comparative outcomes of health care. *J Clin Epidemiol*. 2011; 64(7):770-778.
14. Cleveland WS and McGill R. Graphical perception: theory, experimentation, and application to the development of graphical methods. *J. Am. Stat. Assoc.* 1984; 79(387):531–554.
15. Spence I and Lewandowsky S. Displaying proportions and percentages. *App. Cognitive Psychol.* 1991; 5:61–77.
16. Fenton S. Blog category: pie charts. URL https://www.stevefenton.co.uk/category/pie-charts/
17. Hollands JG and Spence I. Judging proportions with graphs: the summation model. *App. Cognitive Psychol.* 1998; 12:173–190.
18. Friendly M and Kwan E. Comment (on Gelman / Why tables are really much better than graphs). *J. Comput. Graphical Stat.* 2011; 20(1):18-27.