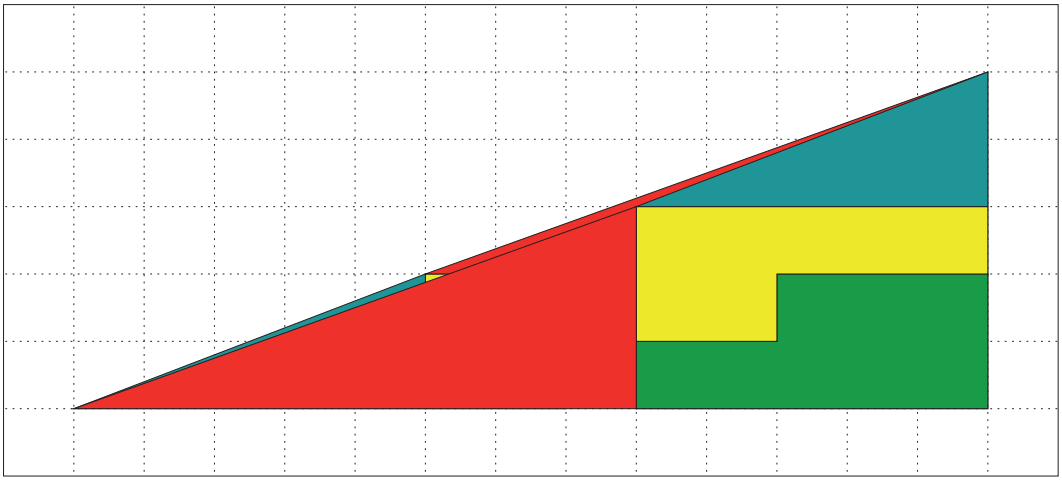
Principles of Data Visualization: Answers Exercises part 2

Ronald Geskus
Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam

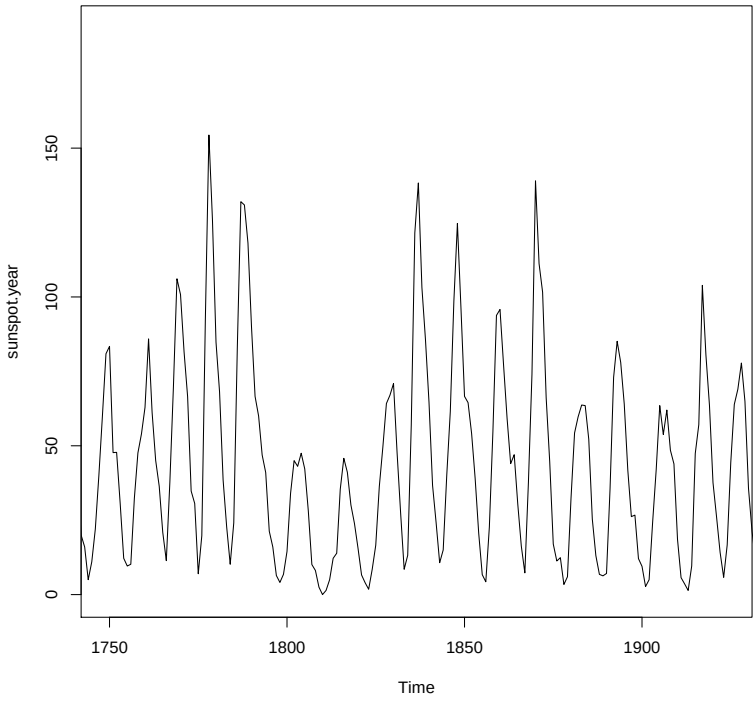
March 12, 2025

1. When you look at the top figure in the question, it looks like the pieces fit together to form a triangle. This isn't the case. The slope of the hypotenuse (the longest side) of the dark grey triangle is $2/5 = .4$, while the slope of the hypotenuse of the gray triangle is $3/8 = .375$. If the top and bottom figures are superimposed you can see there is a thin sliver of difference between the areas. The area of this sliver is precisely the area of the extra square.

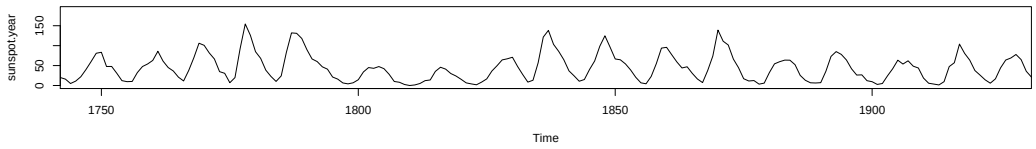
   Humans are very bad at making slope judgements. You may be fooled here because you cannot detect the slope difference in the triangles.



Another example showing that humans are not good at judging slopes is from the book by Cleveland. Compare



and



Only in the latter figure it can be seen that the decline is slower than the increase.

2. First figure:

   - A polar coordinate system is used
   - Geometric object: barplot (bars next to each other); each type of antibiotics is a bar
   - Aesthetic attributes:
     - MIC is mapped to the radius, on a reversed scale; a logarithmic transformation is used. Grid lines are used to read off the values
     - type of antibiotics is mapped to the colours red, blue and black. A legend is placed in the middle.
     - type of bacteria is mapped to the angle axis. I would not call them facets, because they are all plotted together. (Think of an equivalent bar chart in a cartesian coordinate system as in slide 34.)
     - Gram negative and Gram positive bacteria are shown separately. Orange and light blue background distinguish between them. A legend for type of bacteria is placed below the figure.

   For the second figure

   - A cartesian coordinate system is used
   - Geometric object: barplot
   - Aesthetic attributes:
     - MIC is mapped to the y-axis; a logarithmic transformation is used. No grid lines are used
     - type of antibiotics is mapped to the colours red, blue and black. A legend is placed on top of the figure
     - type of bacteria is mapped to the x-axis. I would not call them subpanels (facets), because they are part of the same barplot. A special character is used to mark Gram-positive bacteria
     - genus name is given by coloured text, and a legend is added
   - Facets by row; separate subpanels by resistance pattern are used along the vertical direction.

   The following things can be observed

   - The bacteria *Diplococcus pneumoniae* has a resistance pattern that is very similar to to bacteria from the *Streptococcus* family. In fact, in 1974 this bacteria was recognized as a *Streptococcus*.
   - *Streptococcus fecalis* seems to have a resistance behaviour that is different from the other *Streptococcus* bacteria. Indeed, in 1984 its genus was changed and it was named *Enterococcus fecalis*.

   A nice lecture on the topic of visualisation is on YouTube: https://www.youtube.com/watch?v=vc1bq0qIKoA. From minute seven onwards this example is mentioned.

3. (a)

   - A polar coordinate system is used.
   - Geometric object: bar
   - Aesthetic attributes:
     - calendar time is scaled to position along the angle axis
     - area, measured from the centre, represents percentage of deaths. Use of area makes it impossible to read off the actual numbers. It would have been easier if radius (distance from the center) had been used.
     - cause of death is mapped to colours that fill the wedges (bars). Because the bars overlap, a black and red line are used to distinguish "deaths from wounds" and "other deaths"
   - No (log) transformation of values. Another statistical transformation may have been used: numbers are added.
   - No position adjustment is made. All bars start from the center and overlap, it's not a stacked barchart, nor a dodged barchart (as the first figure in the previous exercise).
   - Separate subplots (facets) are made for the time periods before and after the intervention

   (b) The statement is not correct. Values are mapped to area size, not to radius. Although the graph is a truthful reflection of the numbers, it's impossible to learn the actual numbers from the graph. Also, total number of deaths were shown in the plot, not the numbers standardized by number of soldiers. Another issue is that the wedges overlap. And one may wonder whether a polar coordinate system is the best one. Why not use a cartesian coordinate system?

   (c) The R code gives some alternative ways of plotting, using a cartesian coordinate system. You can also find some at http://mbostock.github.io/protovis/ex/crimea-rose.html

   There has been a discussion as to whether the Nightingale Rose can be seen as a pie chart. I think it cannot. In a pie chart, the numbers are mapped to angle (as the height in a bar chart), while the radius is the same for each category. In the Nightingale Rose, the numbers are mapped to area (in the graph of the previous exercise they were mapped to radius), while the angle represents month and each month covers the same angle size.

4.
   - A polar coordinate system is used.
   - Geometric object: bar
   - Aesthetic attributes:
     - Percentage cost is mapped to position along the angle axis, using a linear scale. Labels are provided that give the percentages.
     - Type of cost is mapped to colour, and a legend is provided. Type of cost is in alphabetical order.
   - No transformation and only one layer

- A position adjustment is made (the percentage values do not start from the same location), but it's difficult to say whether the bars are stacked or dodged. In a cartesian coordinate system a stacked barplot does position adjustment in the y-direction, while a dodged barchart does position adjustment in the x-direction. In a pie chart (polar bar chart) there is only one axis, so we cannot tell.
- Separate subplots (facets) are made for the combination of ventilation status and disease type.