Visualising proportions:-

When we break down the an entity or amount into individual pices now each piece represents the proportion of whole.
Eg. Men and women in a group of people, the percentage of people voting for different political parties, or market shares of the company.

When we want see the see changes in proportions over time or across conditions, there is no signal ideal visualisation that always work.

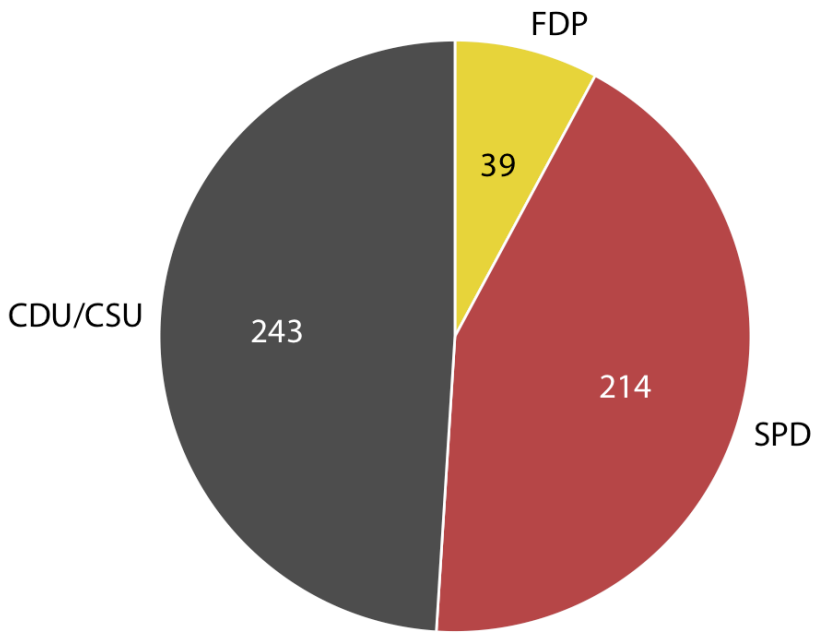
Remember- You always need to pick the visualisation that best fits your specific dataset and that highlight the key data features you want to show.

A Case of pie charts-

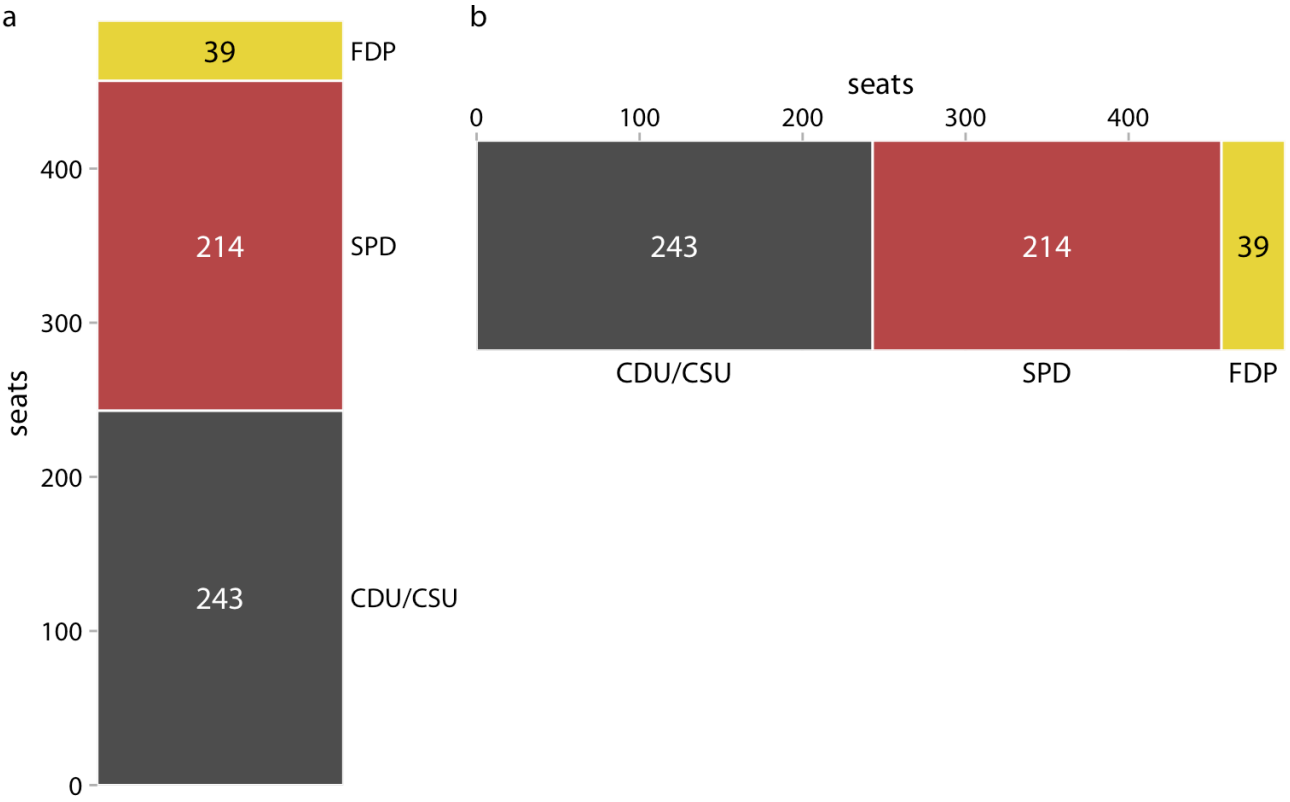
Let's start with an example, From 1961 to 1983, the German parliament was composed of three different parties, CDU/CSU, SPD, and FDP.
In 1976-1980, number of seats different parties have-

CDU/CSU	243seats
SPD	214seats
FDP	39seats
	———————-
	total - 496

Such data is most commonly visualised as a pie chart.



A pie chart breaks a circle into slices such that the area of each slice is proportional to the fraction of the total it represents. The same procedure can be performed on a rectangular, and the the result is a stacked bar chart. Depending on whether we slice the bar vertically or horizontally.



If we take the bars and place them side by side rather then stacking, we get a bar chart, this visualisation makes it easier to perform a direct comparison of different groups(3 in this case).

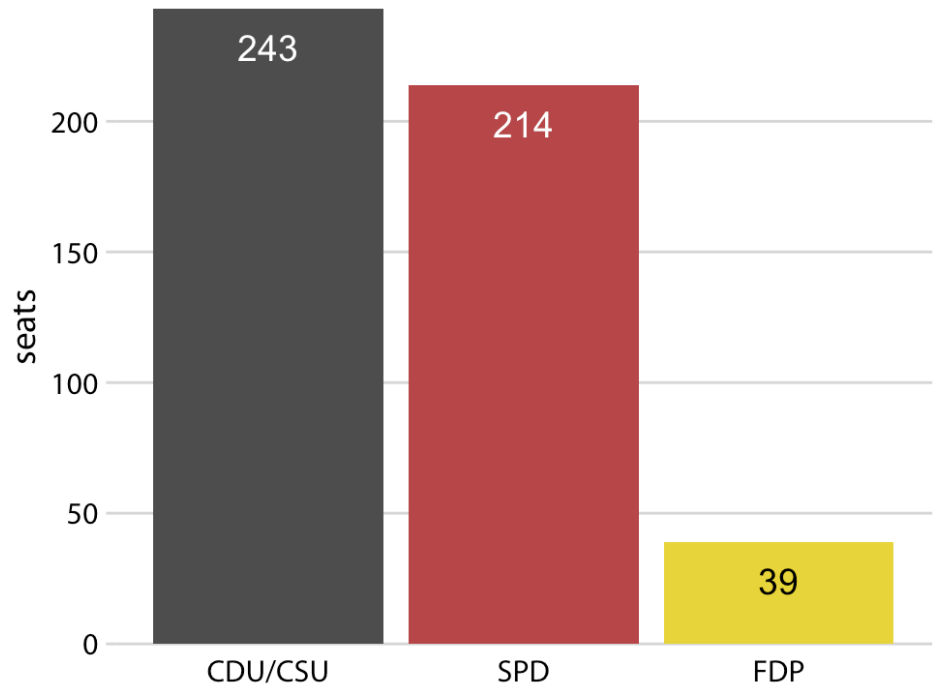
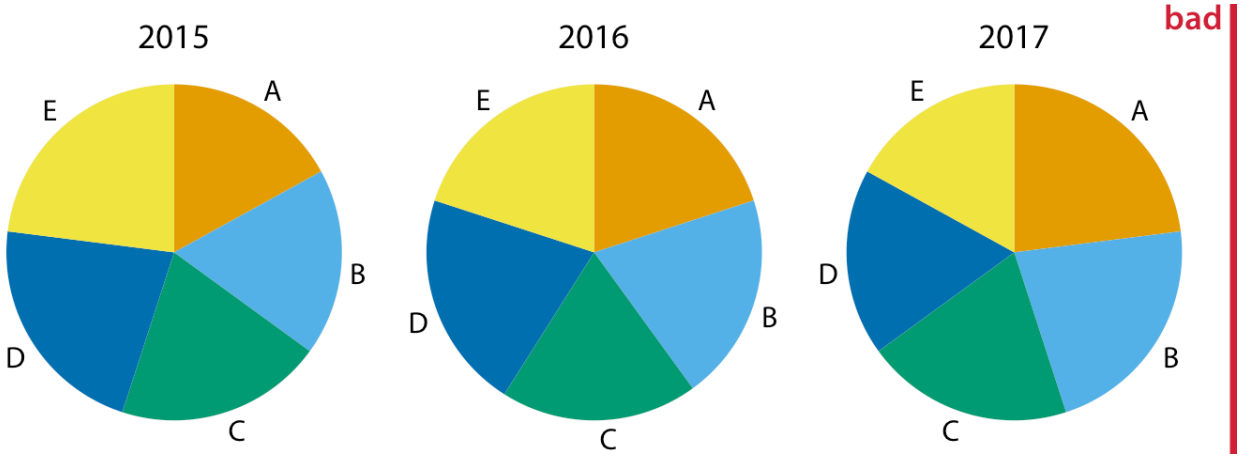


Table 10.1: Pros and cons of common approaches to visualizing proportions: pie charts, stacked bars, and side-by-side bars.

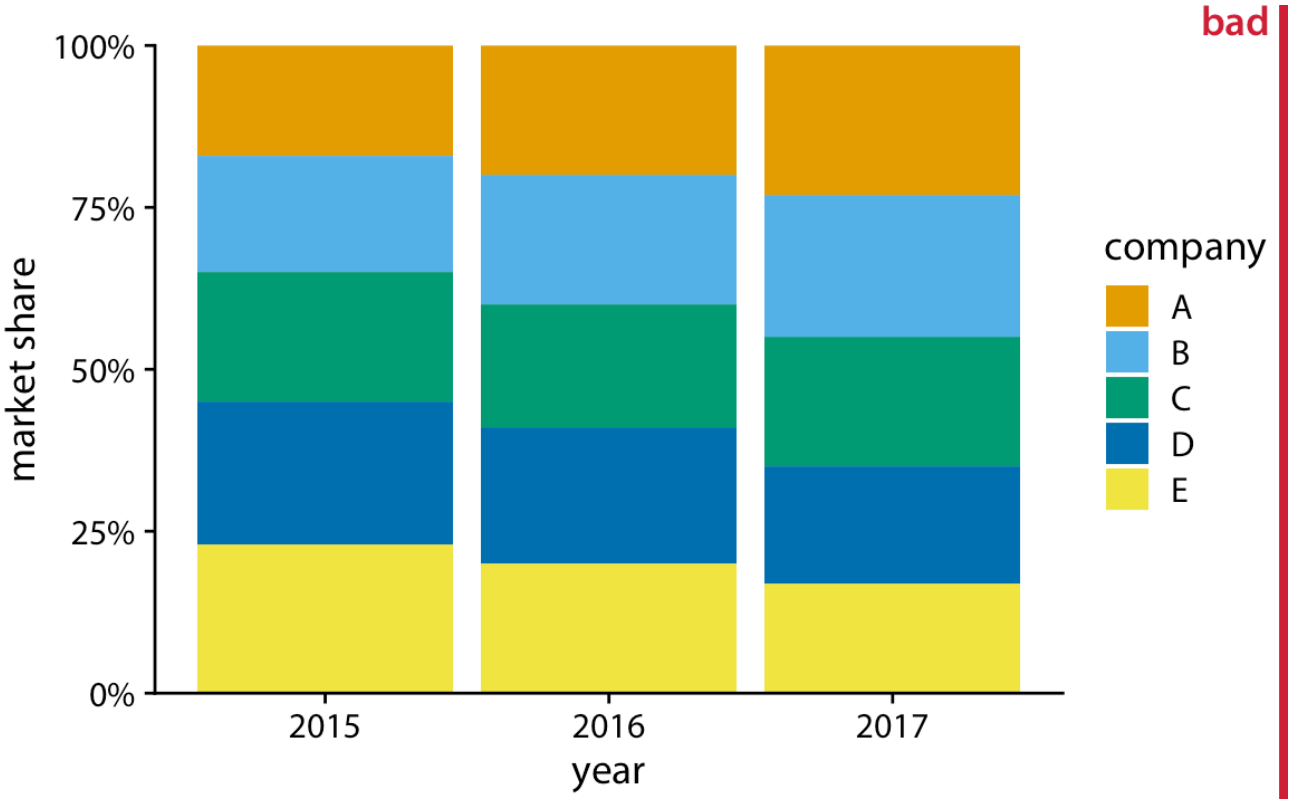
	Pie chart	Stacked bars	Side-by-side bars
Clearly visualizes the data as proportions of a whole	✔	✔	✖
Allows easy visual comparison of the relative proportions	✖	✖	✔
Visually emphasizes simple fractions, such as 1/2, 1/3, 1/4	✔	✖	✖
Looks visually appealing even for very small datasets	✔	✖	✔
Works well when the whole is broken into many pieces	✖	✖	✔
Works well for the visualization of many sets of proportions or time series of proportions	✖	✔	✖

A Case of side-by-side chart-

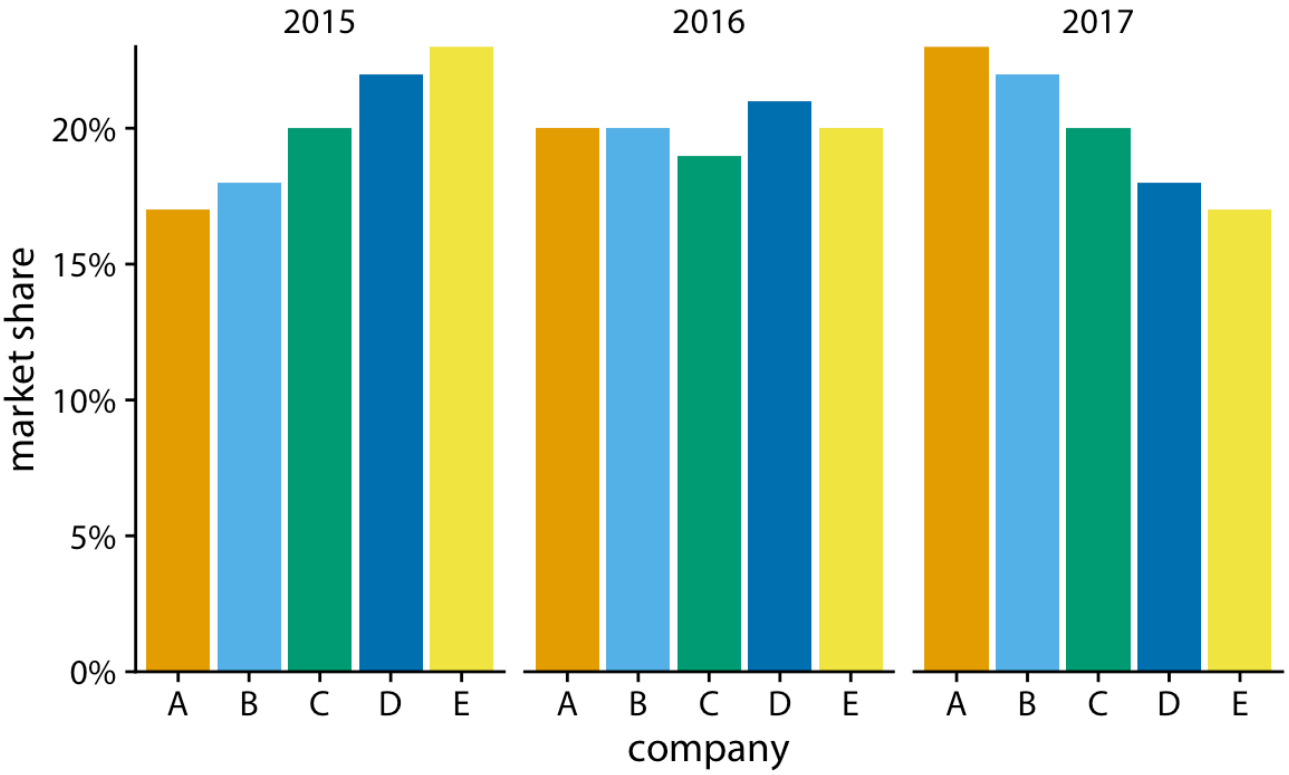
I will now demonstrate a case where pie chart fail. Consider a scenario of five companies A, B, C, D and E, who all have roughly comparable market share of approximately 20%. This hypothetical dataset list the market share of each company for three consecutive years. When we visulaize this dataset with pie chart, it is difficult to sense the time comparison in companies how companies growing or shrinking.



The picture becomes clearer if we use stacked bar chart instead of pie chart. Now we can see the trend of a growing market share for company A and a shrinking market share for company E. However, It is still difficult to see the relative market share of companies B, C and D across years because bars are shifted relative to each other across year. This is general problem with stacked bar plots.



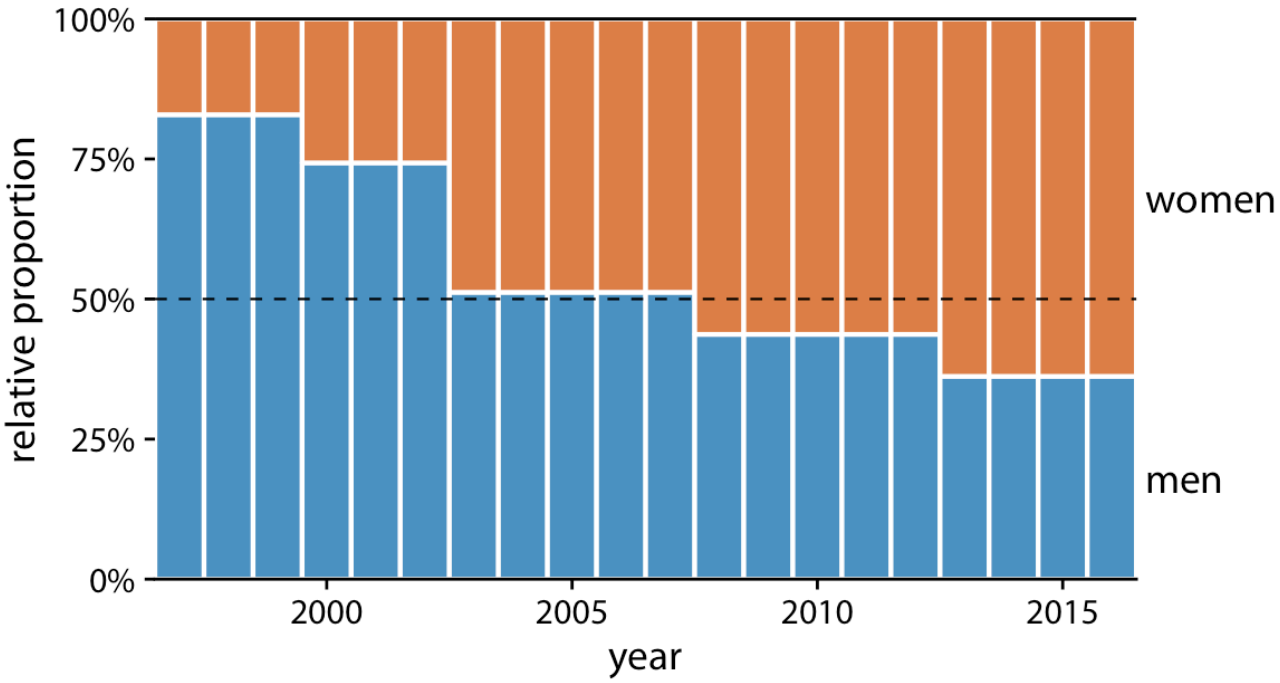
For this hypothetical dataset, side-by-side bars ate the best choice. This visualisation highlights that both companies A and B have increased their market share 2015 to 2017 while both companies D and E have reduces there. It also shows that market shares increase sequentially from company A to E and similarly decrease in 2017.



A case for stacked bars-

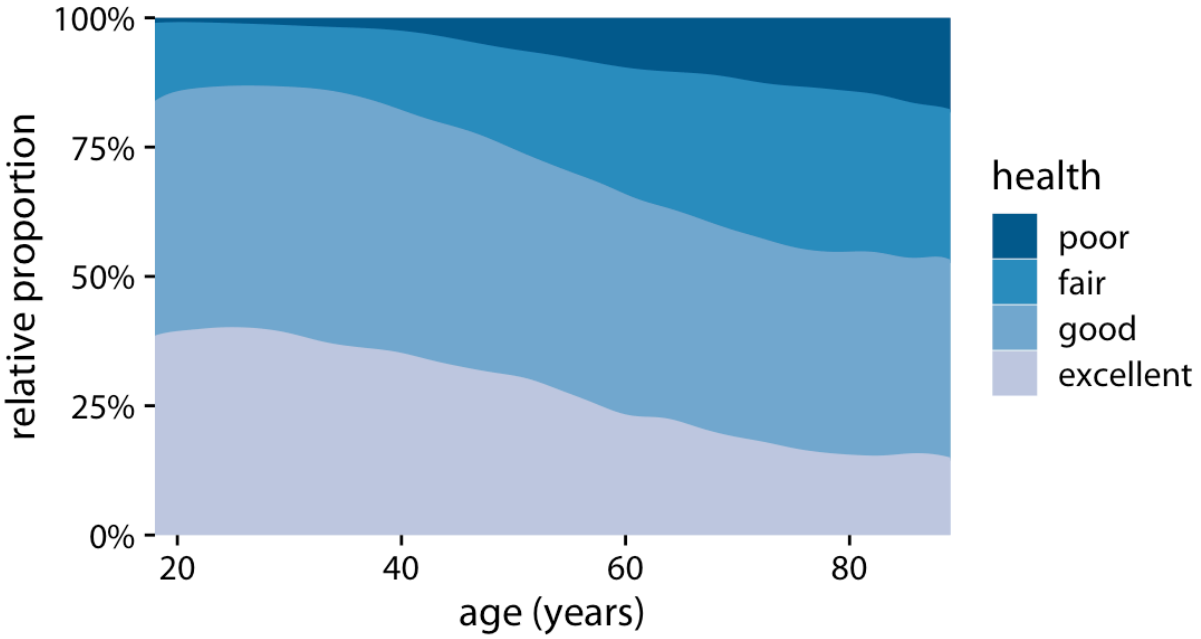
As an example, Consider the proportion of women a country's national parliament. In country Rawanda, which as of 2016 tops the list of countries with the highest female proportion of female parliament members. To visualize how the proportion of women in the Rawanda parliament has changed over time, we can draw a sequence of stacked bars. This provides an immediate visual representation of the changing proportion over Time, we can draw a sequence of stacked bar graphs. This provides an immediate visual representation of the changing proportion over time. To see the when the majority turned female here I used sashed horizontal line at 50%.

Without this line it is difficult to determine whether from 2003 to 2007 the majority was male or female.



Staked Densities graph-

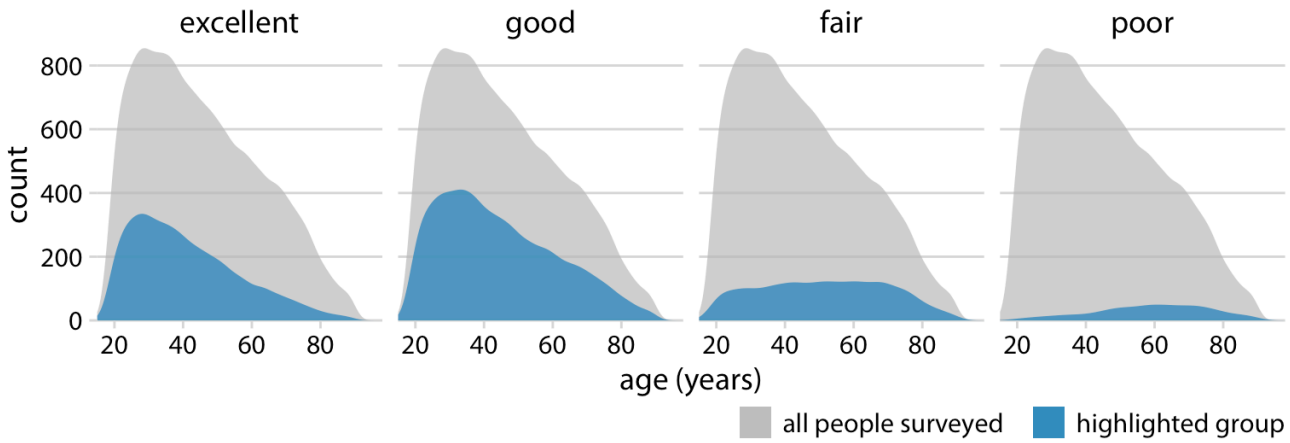
If we want to visualize how proportions change in response to a continuous variable, we can use staked densities. It can be tough as a limiting case, where infinitely many small stacked bars arranged side-by-side. The densities in staked-density are typically obtained from kernel density estimation(KDE).
Eg. Consider the health of people as a function of age. Age can be considered as a continuous variable. Here four categories are described by colour gradient. We can see clearly that overall health declines as people age, and we can also see that despite this trend, over half of the population remain in good or excellent health until very old age.



This graph has a major limitation: by visualising the proportion of the four health condition as percent of the total, the figure obscures that there are many more young people than old people in dataset. Thus, even though the percentage of people reporting to be in good health remains approximately unchanged. The absolute number of people in good health declined unchanged as the total number of people at a given age declines.

Visualizing proportions separately as part of the total-

We are plotting above dataset here in efficient way. Side-by-side bars have the problem that they don't clearly visualize the size of individual parts relative to the whole and staked have the problem that the different cannot be compared easily because they have different baselines. We can resolve these two issues by making separate plot for each part and in each plot showing the respective part relative to the whole.
The overall age distribution for is shown as the shaded gray and the age distribution foreach health status are shown in blue.

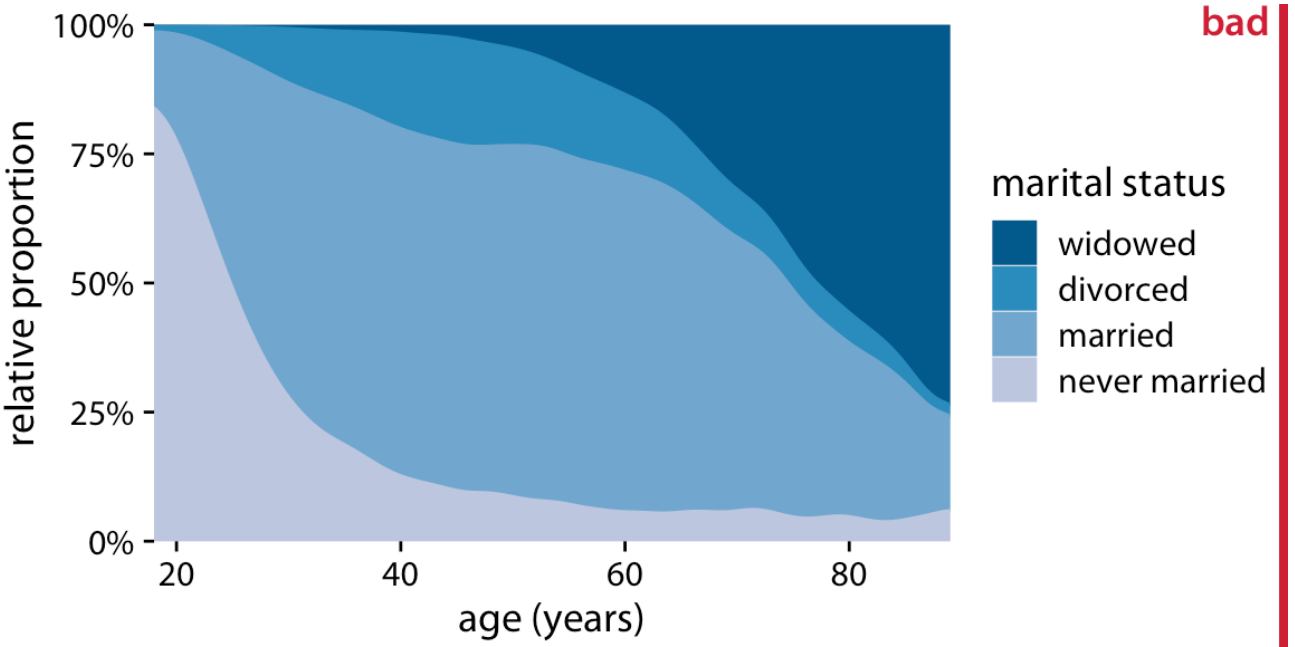


Now in this graph we can clearly see that the number of people with excellent or good health declines past ages 30-40, while the number of people with fair health remains approximately constant across all age.

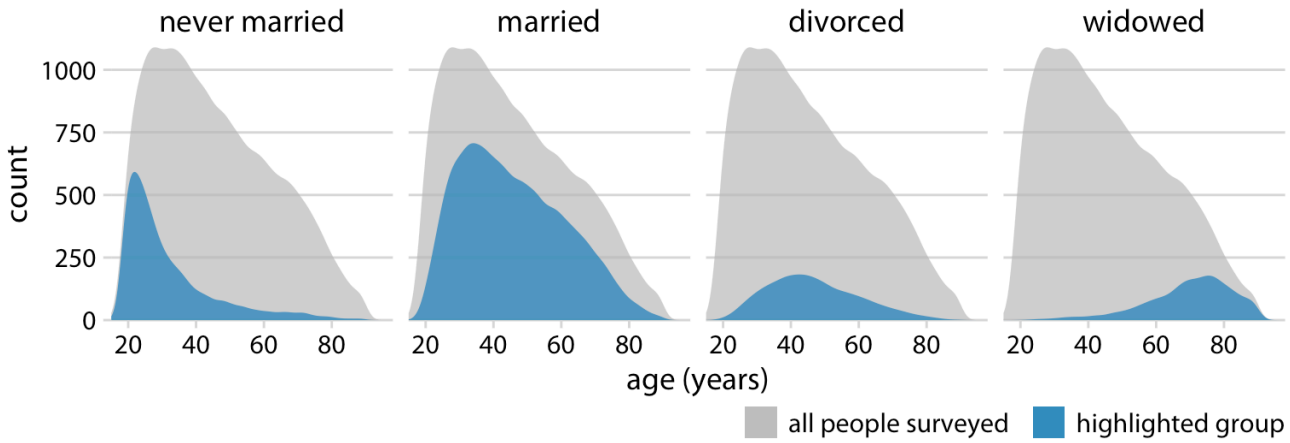
Let's take another example and plating the same-
Consider a case of marital status. Marital status changes much more drastically with age than does the health status and staked densities is not illuminating.

I have labels this "bad" because the frequency of people who never been married or are widowed changes so drastically with the age distribution of married and divorced propel are highly distorted and difficult to interpret.

The same dataset visualised as partial denotes is much clearer that the proportion of married people peaks around the late 30's , the proportion of

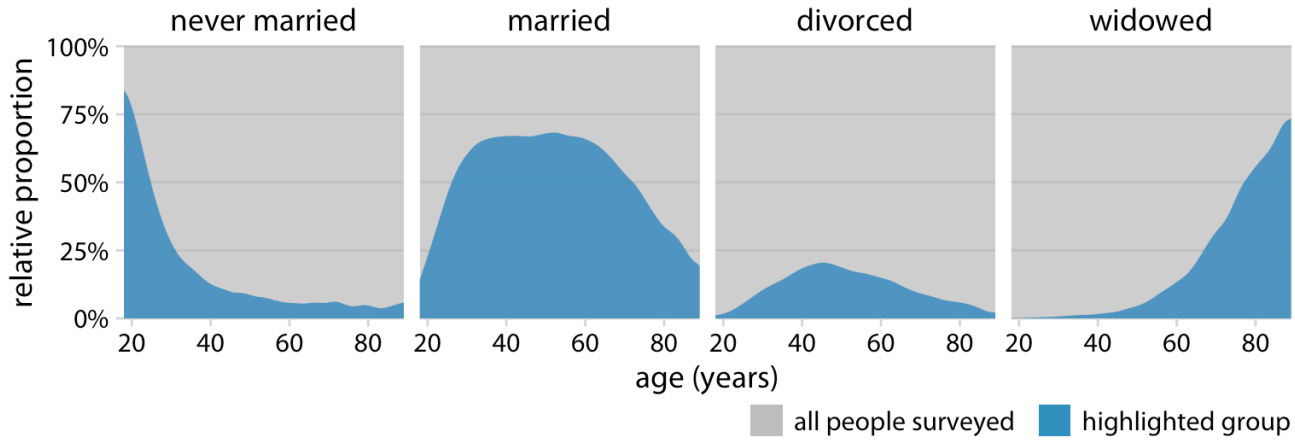


divorced people peaks around the early 40's, and the proportion of wowed people peaks around the mid 70s



The coloured areas show the density estimates of the ages of people with the respective marital status, and the gray areas show the overall age distribution.

However, one downside of this, is that this representation doesn't make it easy to determine relative proportions at any given point in time for example If we wanted to know at what age more than 50% of all people served are Married. To answer this question, we can instead use the same of type of plot but show relative proportion instead of absolute counts along the y axis.



Now we see that married people are in the majority starting in their late 20's, and widowed people are in the majority starting in their mid 70's.