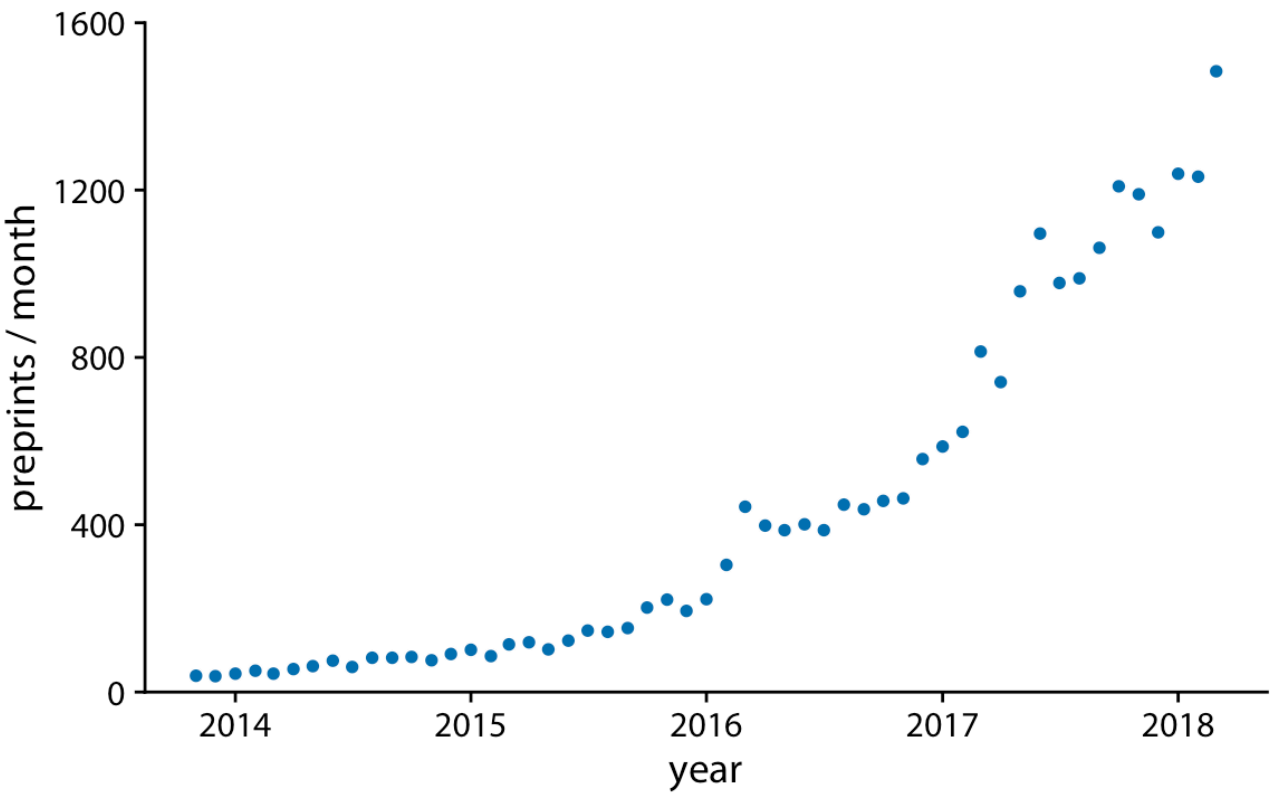
Visualizing time series data -

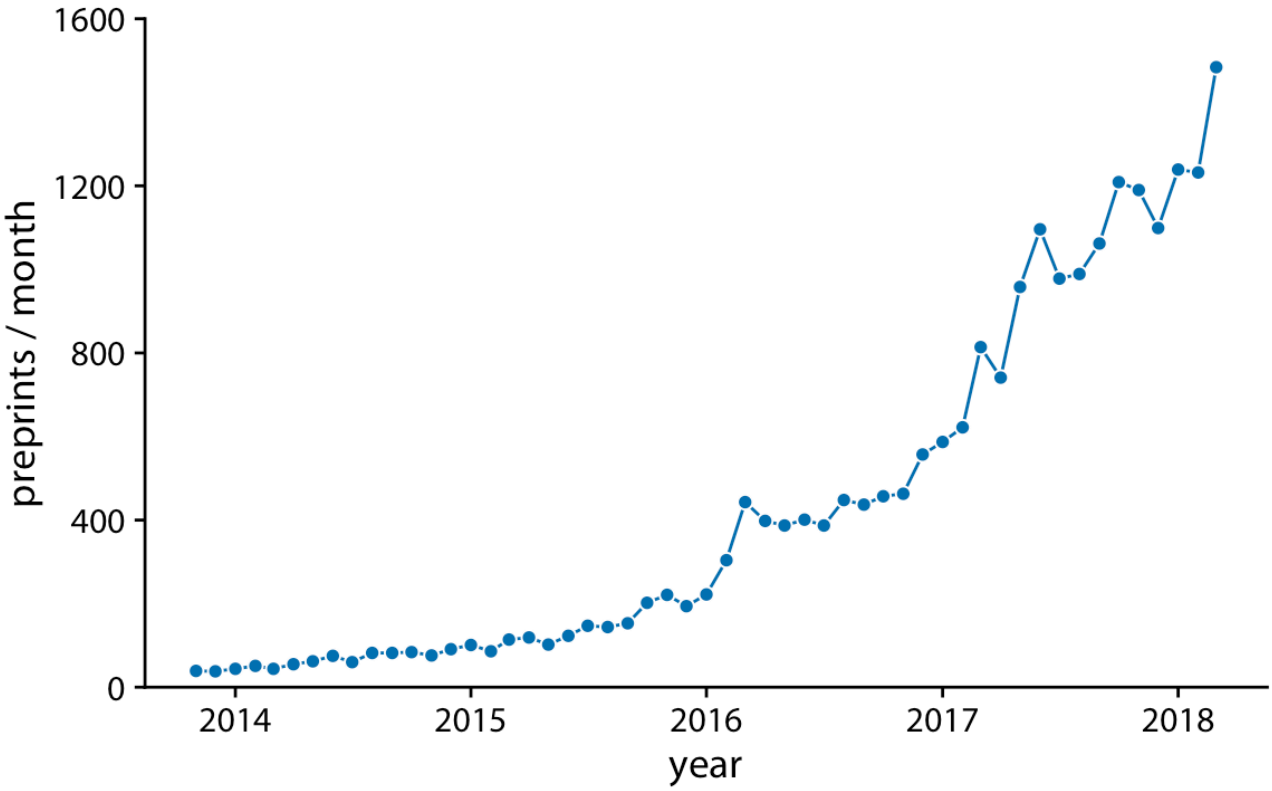
Time adds an additional structure to the data. Line graphs are no limited to show time series, however. They are appropriate whenever one variable imposes an ordering on the data. If we have multiple variables that depend on time, we can either draw separate line plots or we can draw a regular scatter plot and then draw lines to connect the neighbouring points in time.

Individual time series -

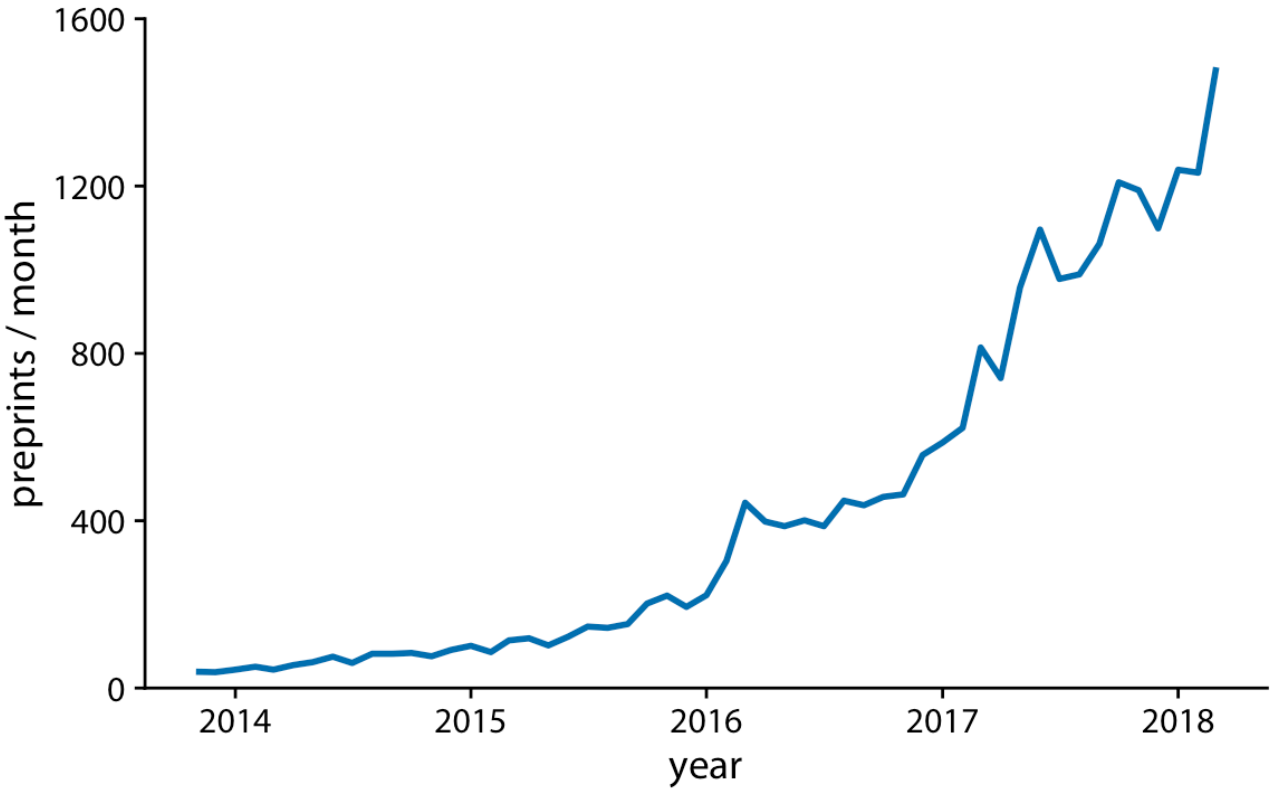
Starting with example, we are considering the the pattern of monthly preprint submission in biology. Preprints are scientific articles that researchers post online before formal peer review and public in a scientific journal. In this scatter plot we are representing the number of submissions in each month.



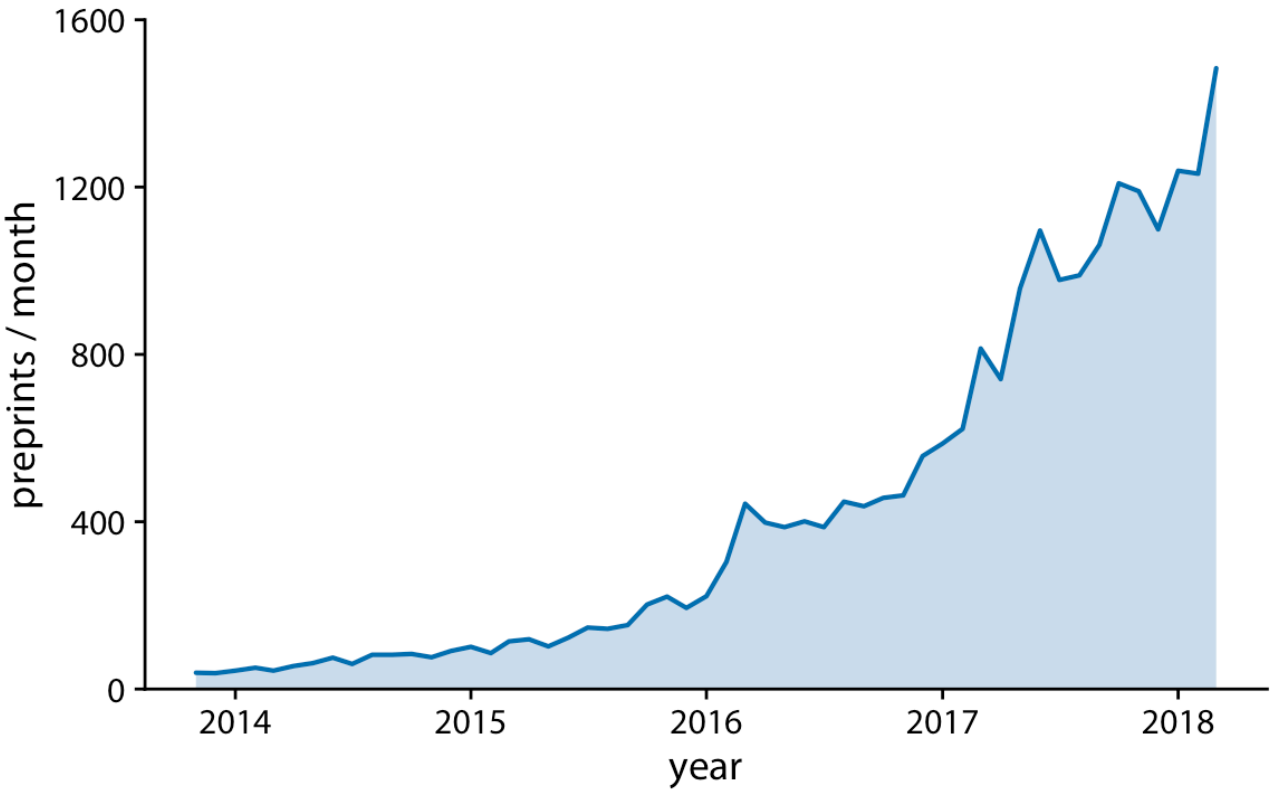
The dots are evenly spaced along the x axis, and there is a defined order among them. We can visually emphasise this order by connecting neighbouring points with lines. Such a plot is called line graph.



Without dots, the figure places more emphasis on the overall trend in the data and less on individual observation. A figure without dots is also visually less busy. In general, the denser the time series, the less important it is show individual observation.i think omitting the dots is fine.

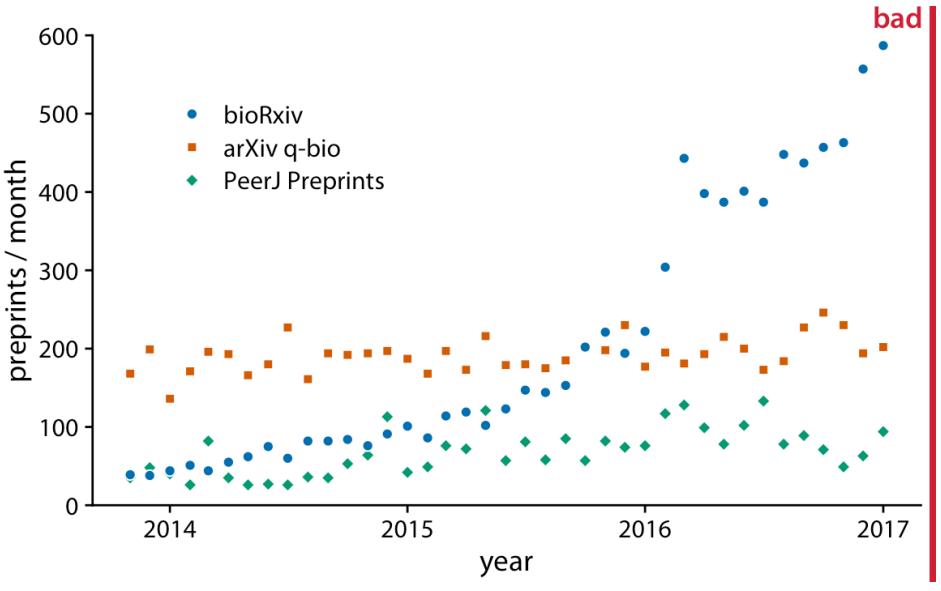


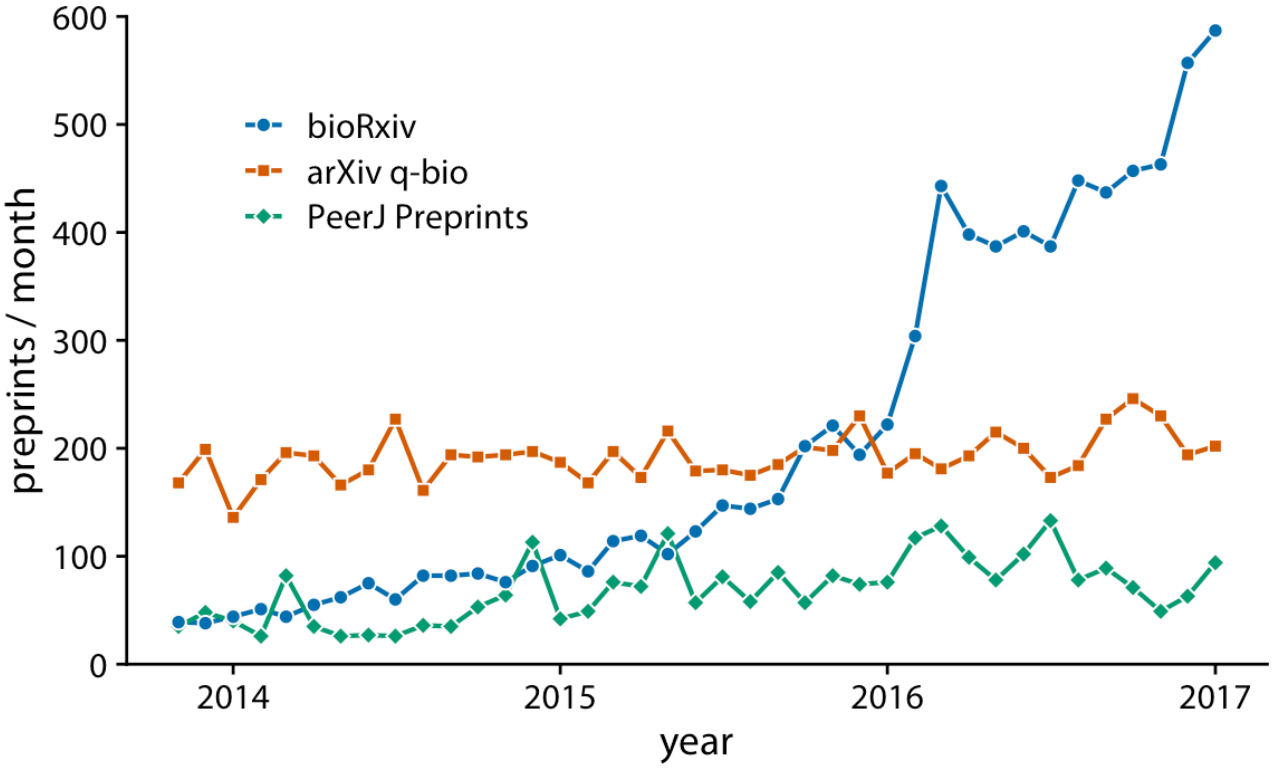
We can also fill the area under the curve with a solid colour. This choice further emphasises trend in data. This visualization is only valid if the y axis start at zero, so that the height of the shaded area at each time point.



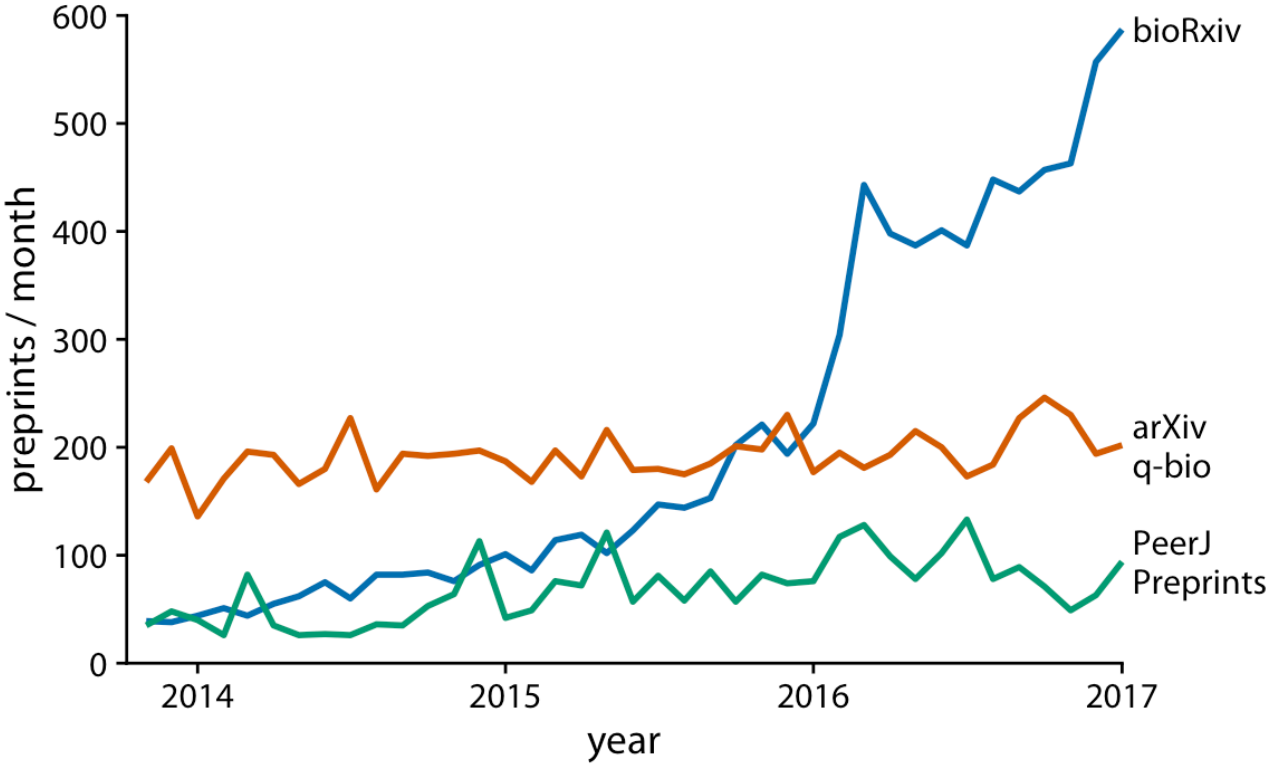
Multiple time series and dose-response curves -

We stern have to show multiple time course that we want to show at once. In this case scatter plot is not a good idea, because the individual time courses run into each other.

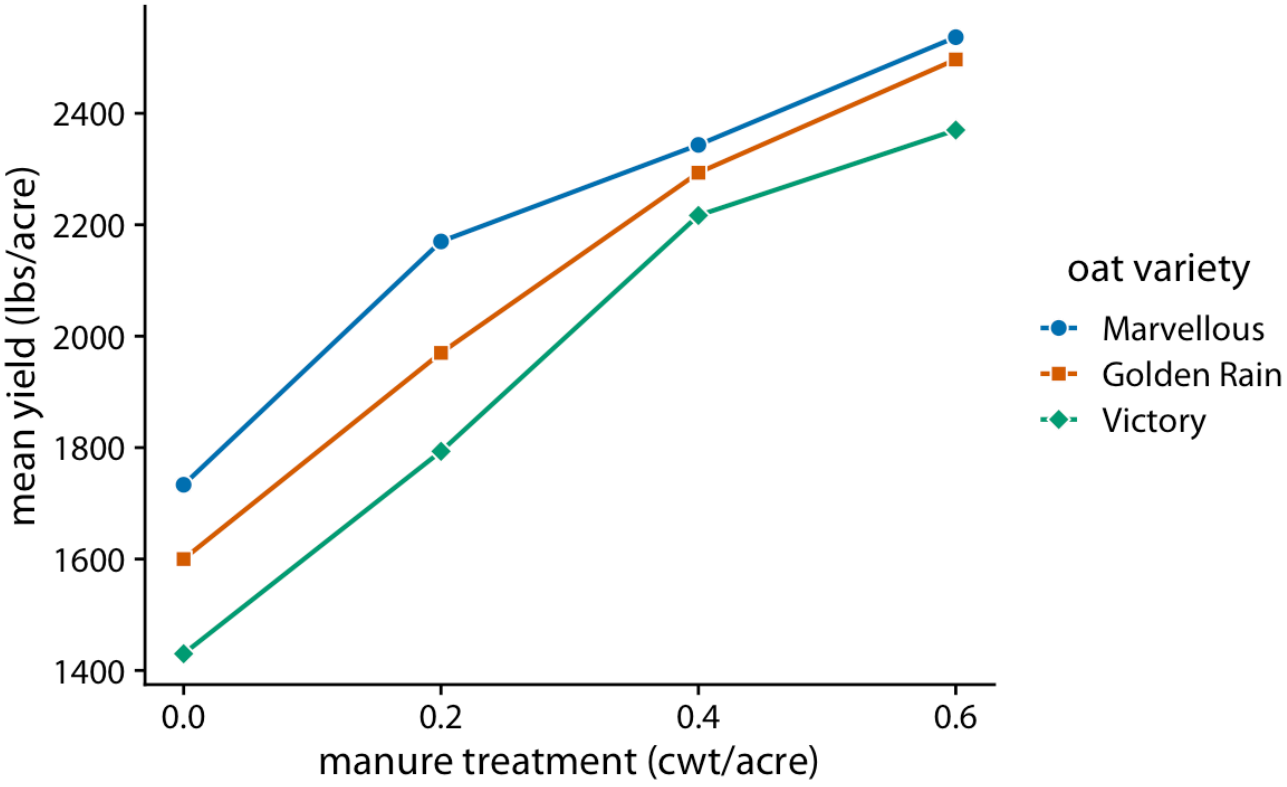




The above plot is acceptable in this case. But we can further improve it by removing dots.



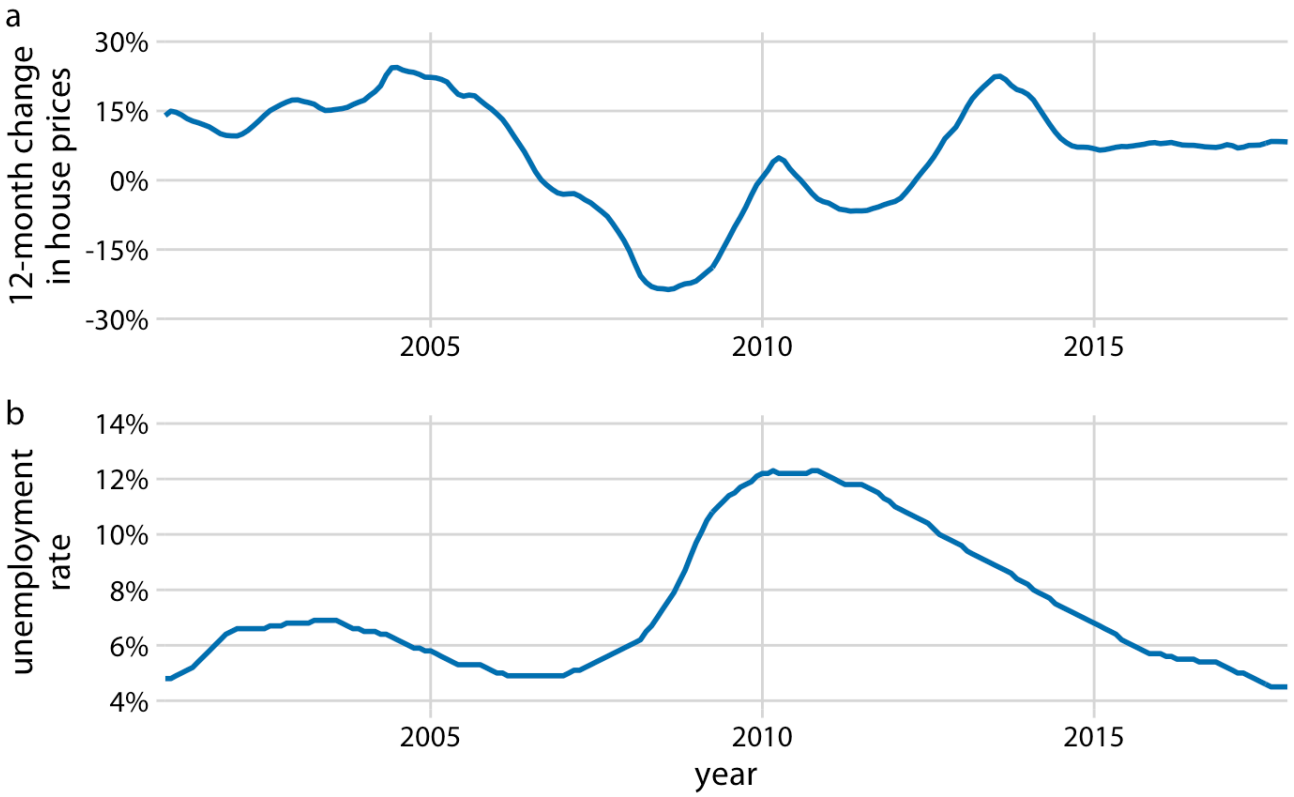
They are appropriate whenever data points have natural order. For example, in dose-response curves, where we measure how changing some numerical parameter in an experiment affects an outcome of interest. In plot given below the dose-response curve has similar shape but difference is at the starting point where in the absence of fertiliser some varieties of oats yields higher than others naturally.



Time series of two or more response variable -

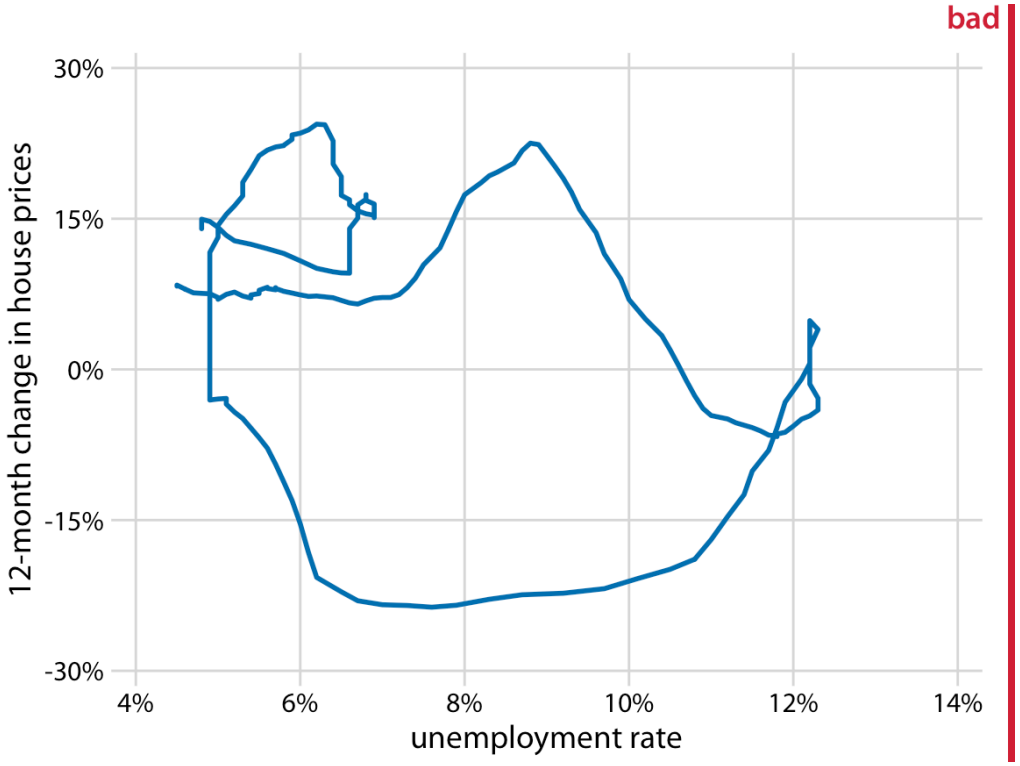
Such situations arises commonly in macroeconomics where we have more than one response variable. For example, we may be interested in the house price from previous 12 months as it relates to the unemployment rate. We may expect that house prices rise when the unemployment rate is low and vice versa.

This plot shows the two variables of interest(change in house price and unemployment rate). If we want to identify temporal/time-related regions when both variables move in the same or in opposite directions, we need to compare the relative slope of the two curves.



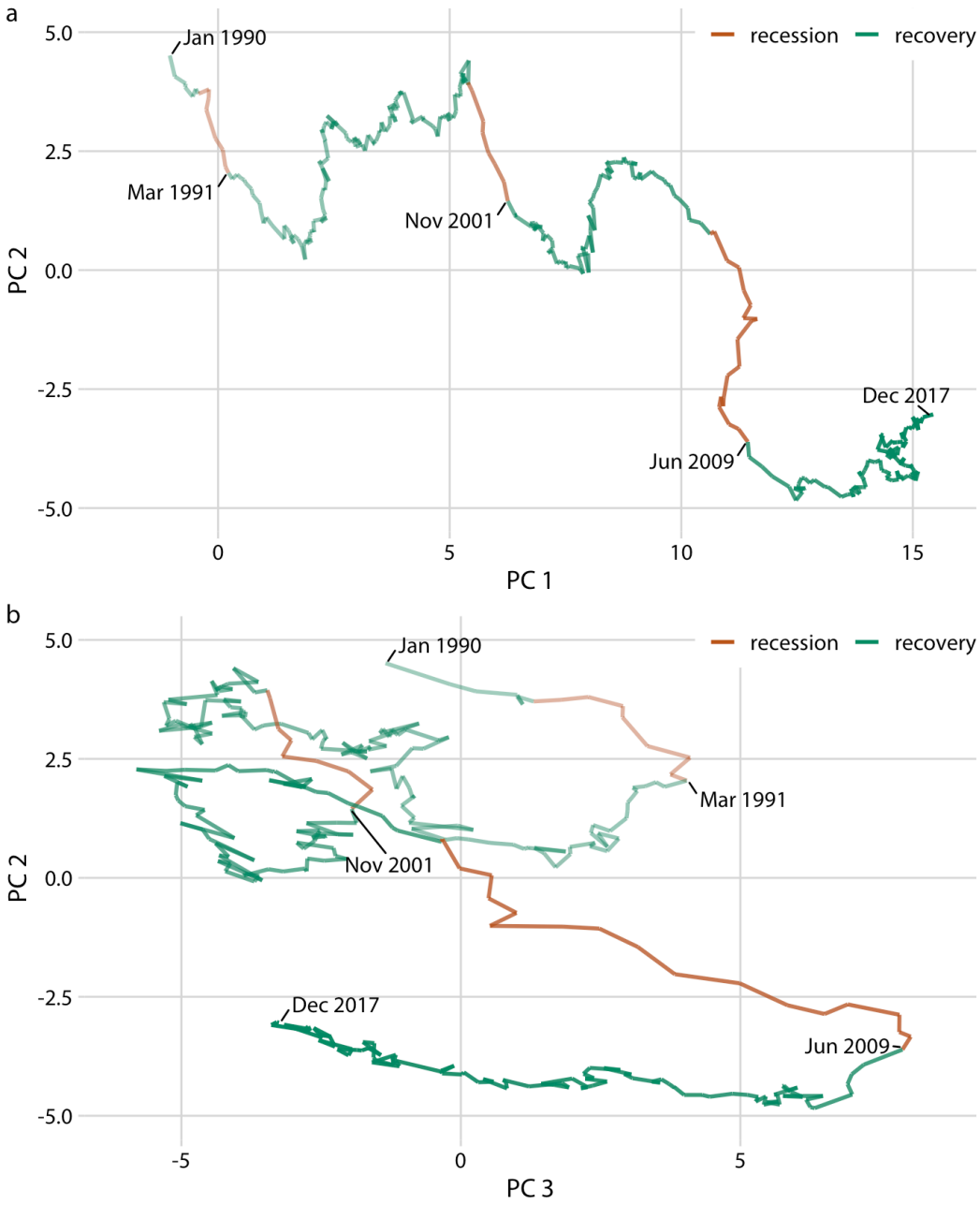
As an alternative, instead of showing two separate line graphs, we can plot the two variables against each other(unemployment rate on x axis and change in house price on y axis) we can choose any axis for variables. Such a visualization is called a *connected scatter plot* because we are making a scatter lot of two variables against each other then connecting neighbouring points. This is also called a *phase portrait,* because it is commonly used to represent movement in phase space. In this plot we are plotted the daily temperature normals in Houston, Texas, versus in San Diego.

When drawing a connected scatter plot, it is important that we indicate both the direction and the temporal/time-related scale of the data on the curve.otherwise It is a meaningless scribble.



Separate line graphs are tend to be easier to read, but once people are used to connected scatter plot than they may be also able to extract information/ patterns from it that is hard to see in line graph.

Even though connected scatter plots can show only two variables at a time, we can also use them to visualize higher-dimensional dataset.. as an example, we will visualize a dataset of monthly observations over 100 macroeconomic indicators. We perform a principal component analysis of all indicators and then draw a connected scatter plot of PC2 versus PC1 and versus PC3.



Here, PC1 approximately measures the overall size of the economy, which rarely decreases over time.

By colouring the connected scatter plot by times of recession/slowdown and recovery, we can see that slowdown with a drop in PC2. The recoveries seems to be correspond to a drop in PC3. Moreover, in the PC2 versus PC3 plot, we see that the line follows the shape of a clockwise spiral. This pattern emphasizes the cyclical nature of the economy, with recessions following recovers and vice versa.