Unit - IV

Visualizing Time series & Uncertianity

Visualizing time series and other functions of an independent variable.

Time series refers to a sequence of data points that are collected, recorded or observed at regular intervals over a specific period of time.

In Time series, each data point is associated with a specific time period, which allows for a chronological organization of the data.

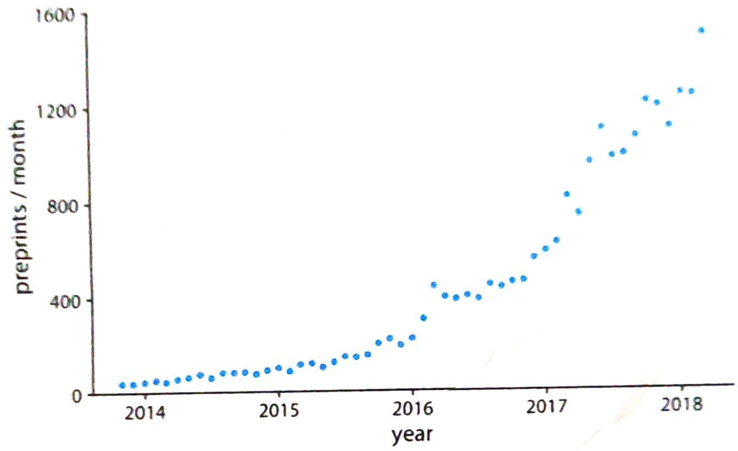
~~Here~~ we visualize three types of time series

1. Individual time series
2. Multiple-time series & dose-response curves
3. Time series of two or more response variables.

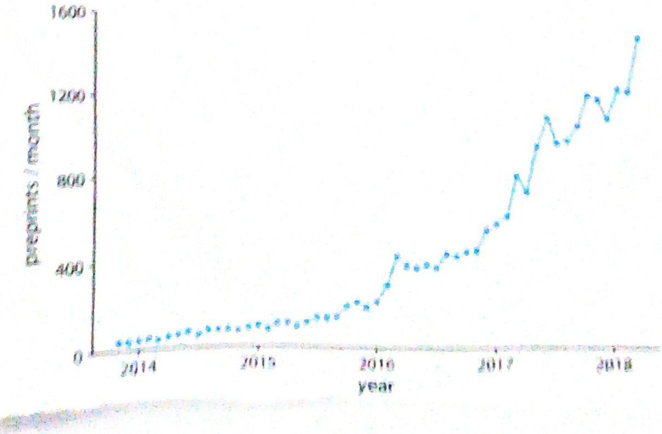
Individual time series

We will consider the pattern of monthly preprint submission in biology. The preprint server bioRxiv, which was found in Nov 2013. specifically for researchers working in the biological sciences, has seen substantial growth in monthly submissions since.

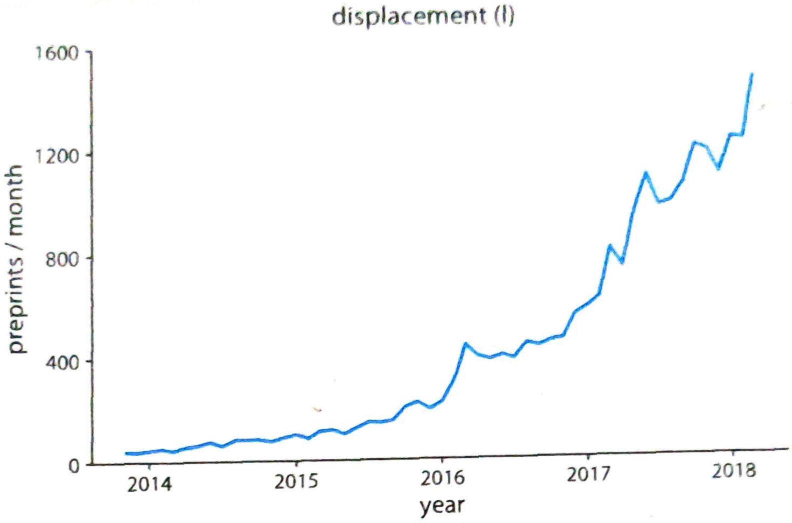
We can visualize this growth by making a form of scatterplot where we draw dots representing the no of submissions in each month. The dots are spaced evenly along the x-axis and there is a defined order among them. Each dot has exactly one left and one right neighbor.



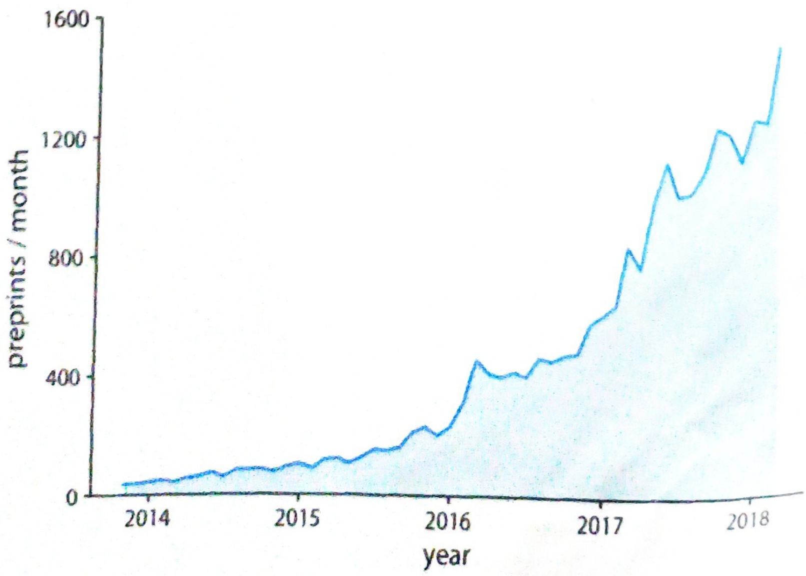
We can visually emphasize this order by connecting neighbouring points with lines. Such a plot is called a line graph.



using lines to represent time series is generally accepted practice however and frequently the dots are omitted altogether without dots, the figure is more emphasis on the overall trend in the data and less on individual observations



We can fill the area under the curve with a solid color. this visualization is only valid if the y axis starts at zero so that the height of the shaded area at each time point represents the data value at that time point.



Multiple Time series and Dose response curves

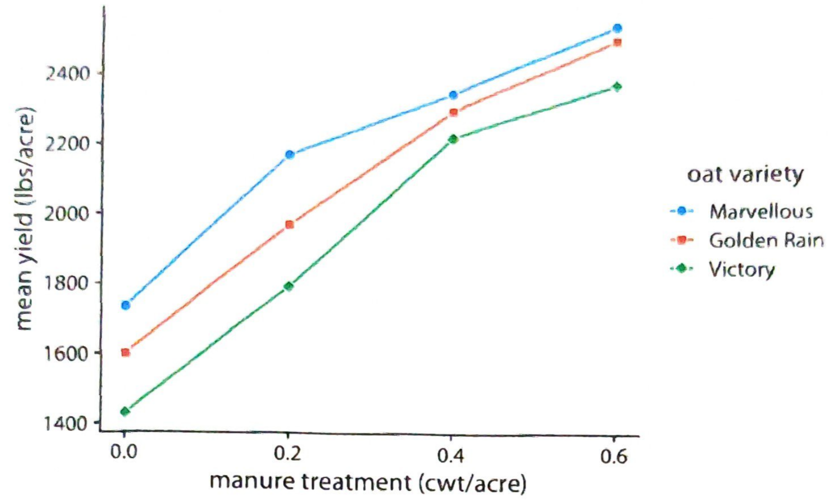
We often have multiple time course that we want to shown at once. In this case we have to be more careful in which how we plot the data. For this type of data it is difficult to draw scatterplot, line graphs.

We use dose-response curves to represent this data. Here we measure how changing some numerical parameter in an experiment (the dose) affects an outcome of interest (the ~~respect~~ response).

Let us consider a experiment of this type, measuring oat yield in response to increasing amount of fertilization. ~~It is a~~ the visualization, highlights how the dose-response curves have a similar shape for the three oat varieties considered but differ in the starting point in the absence of fertilization.

The Dose-response curve showing the mean yield of oat varieties after fertilization with manure. The manure serves as a source of nitrogen and oat yields generally increase as more nitrogen is available, regardless of variety

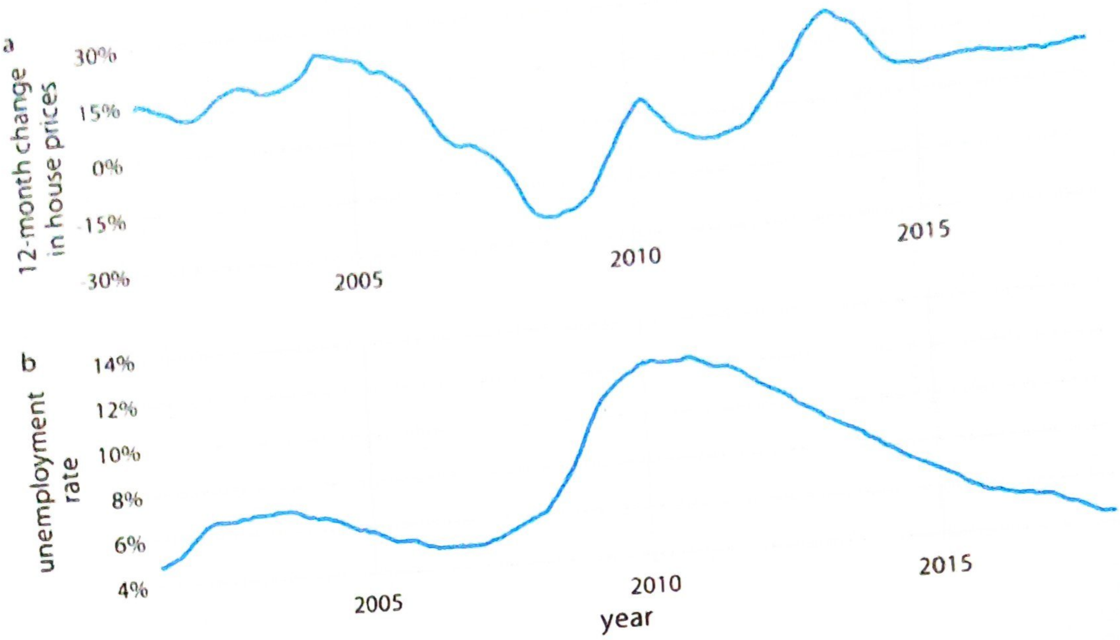
Here, manure application is measured in cwt (hundredweight) per acre.



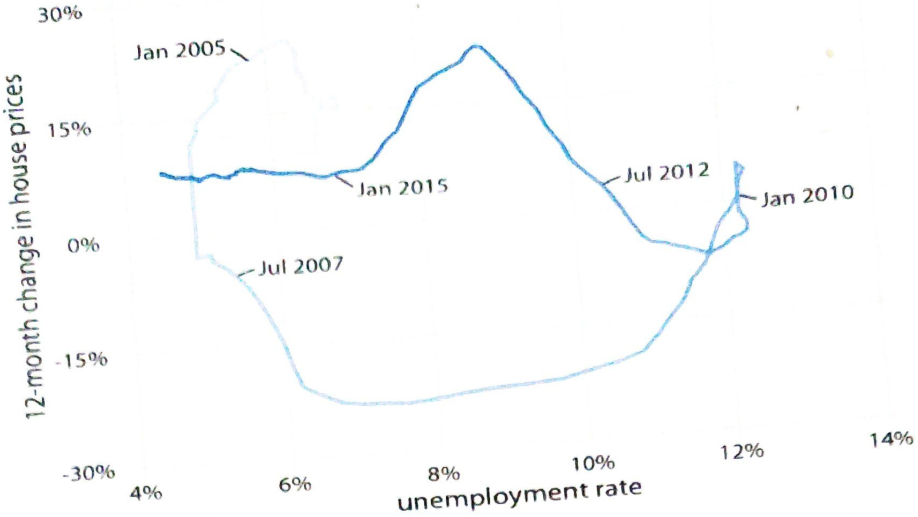
Time series of Two or More Response Variables:

This situations arises commonly in macroeconomics. For example we may be interested in the change in house price from the previous 12 months as it relates to the unemployment rate. we may expect that house prices rise when the unemployment rate is low and viceversa.

We can visualize such data as two separate line graphs stacked on the top of each other.



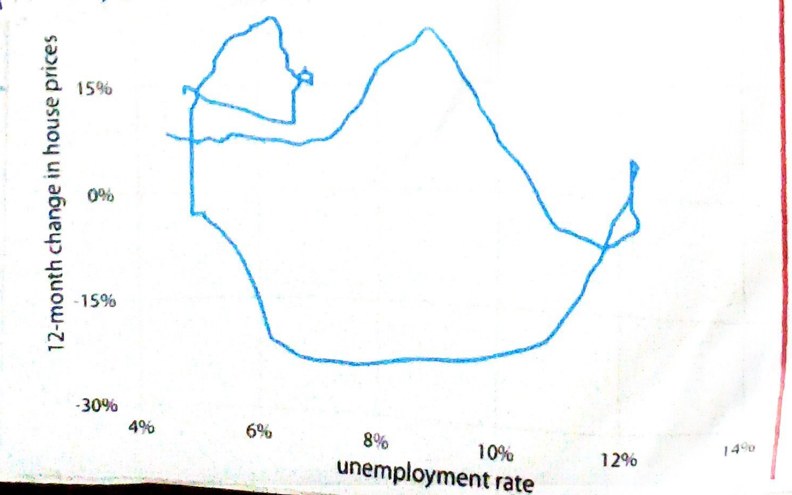
We can plot the two variables against each other, tracing a path that leads from the earliest time point to the latest. Such a visualization is called a connected scatterplot. We are technically making a scatterplot of the two variables against each other and then are connecting neighbouring points



In a connected scatterplot, lines going in the direction from the lower left to the upper right represent correlated movement between the two variables ~~lasa~~ ie as one variable grows, so does the other.

If the lines going in the perpendicular direction, from the upper left to the lower right represent anticorrelated movement ie as one variable grows the other shrinks.

If two variables have a somewhat cyclic relationship, we will see circles or spirals in the connected scatterplot.



Visualizing uncertainty
Framing probabilities of frequencies

The concept of uncertainty mostly in the context of future events. We deal with uncertainty by the concept of probability.

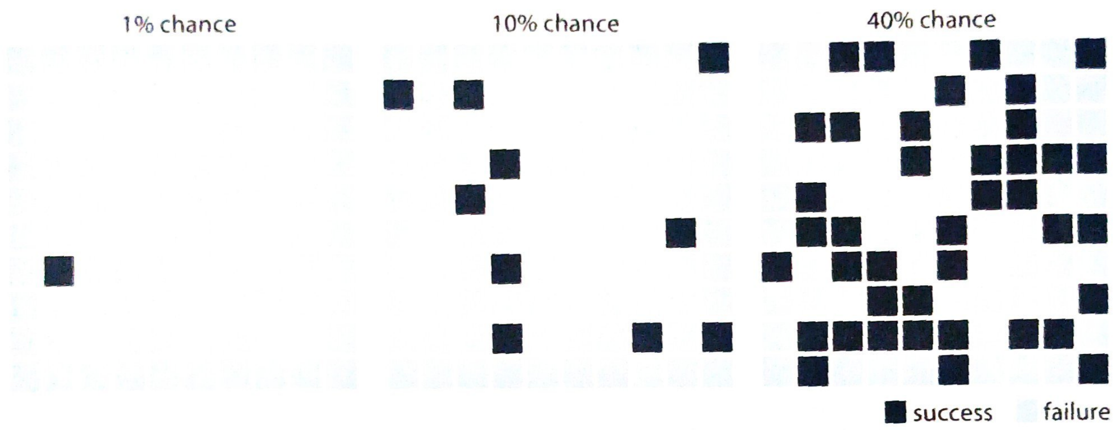
Assume you perform some sort of random trial such as a coin flip or rolling a die and look for a particular outcome. you can call this outcome success and any other outcome as failure.

The concept of probability tangible by creating a graph that emphasizes both the frequency aspect and the unpredictability of a random trial

Let us consider an example of visualizing probability as frequency. There are 100 squares in a grid and each square represents either success of failure in some random trial.

A 1% chance of success corresponds to one dark and 99 light squares, a 10% chance of success corresponds to ten dark and 90 light squares and 40% chance of success corresponds to 40 dark and 60 light squares. By randomly placing the dark squares among the light squares.

we can create a visual impression of randomness that emphasizes the uncertainty of the outcome of a single trial. when we want to visualize two discrete outcomes, success or failure then this visualization is fine.

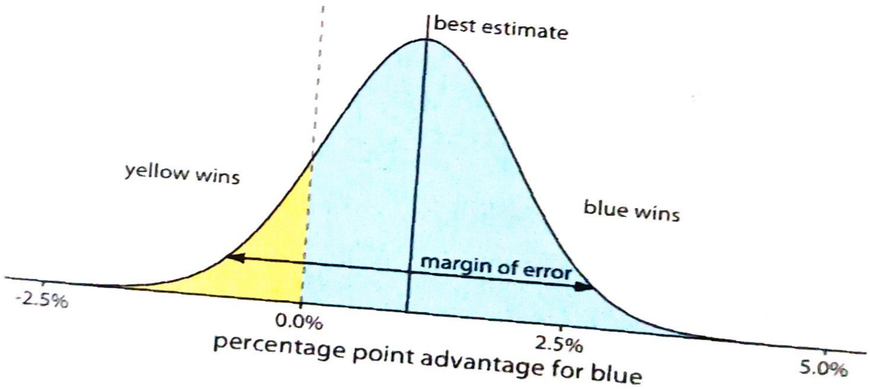


If we are dealing with more complex scenarios where the outcome of a random trial is a numeric value. we use the probability distribution.

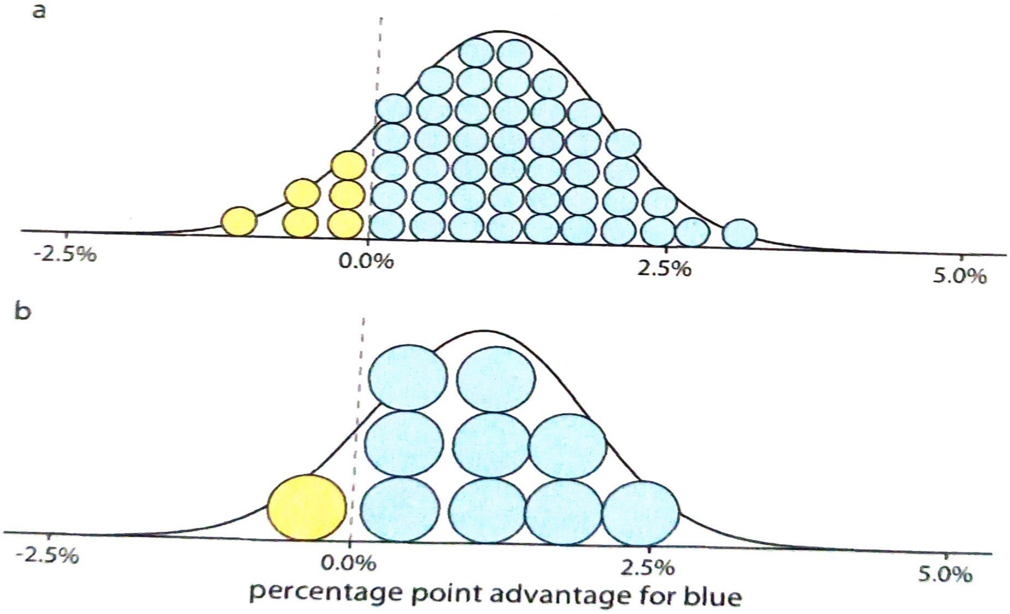
Let us consider the example hypothetical prediction of an election outcome. The blue party is predicated to win over the yellow party by approxi-mately one percentage point. but the prediction has a margin of error. The area shaded in blue, corresponding to

87.1%. of the total, represents all outcomes under which blue would win. In this example blue has an 87%. chance of winning the election.

Here the range of possible outcomes with their associated likelihoods is called a probability distribution and we can draw it as a smooth curve that rises and then falls over the range of possible outcomes as shown in the figure below.



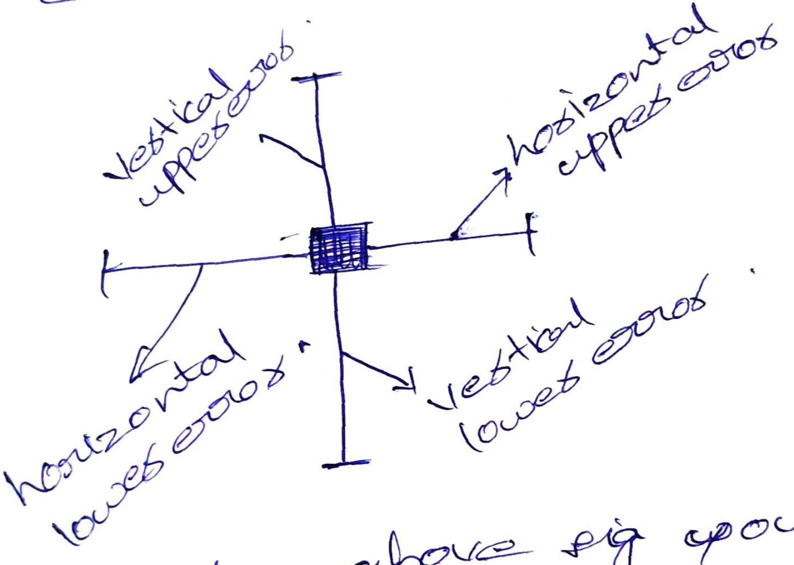
The other representation is the quantile dotplot. In the quantile dotplot, we subdivide the total area under the curve into evenly sized units and draw each unit as a circle. we then stack the circles such that their arrangement approximately represents the original distribution curve as shown in the figure below.



Visualizing the uncertainty of point estimates

When we visualizing the uncertainty of point estimates we uses two methods Error bars and Bayesian ~~credible~~ credible intervals.

Error bars are graphical representations of the variability of data and used on graphs to indicate the error or uncertainty in a reported measurement. The error bars can be represented as

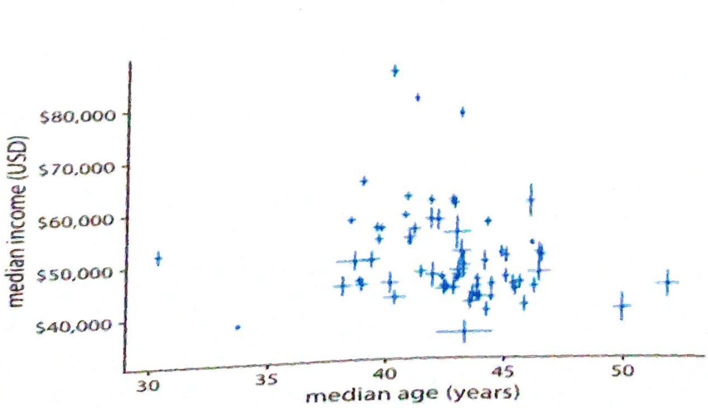


From the above fig you can choose to show only one of the error bars or any combination of them.

The length of an error bar indicates the uncertainty of the value. For example for an average value, a long error bar means that the concentration of the values the average was calculated on is low and thus that the average value is uncertain. Conversely a short error bar means that the concentration of the value is high.

and thus that the average value is more certain.

The error bars have one important advantage over more complex display of uncertainty. They can be combined with many other types of plots. For any visualization we may have we can add some indication of uncertainty by adding error bars. For example we can show amounts with uncertainty by drawing a bar plots with error bars.



Visualizing the uncertainty of concepts

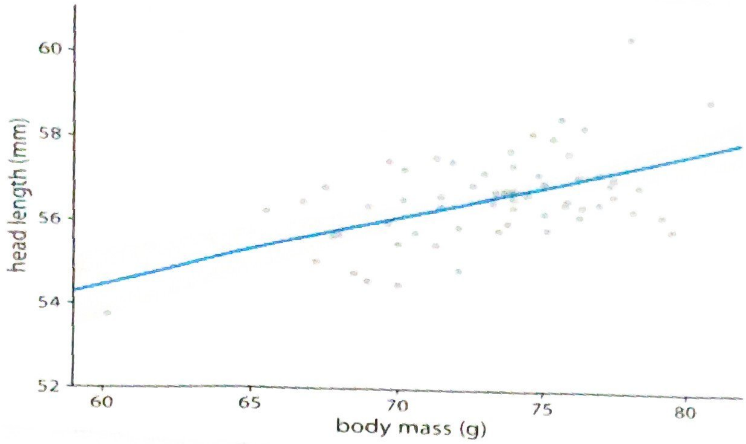
Here we represent how to show a trend in a dataset by fitting a straight line or curve to the data.

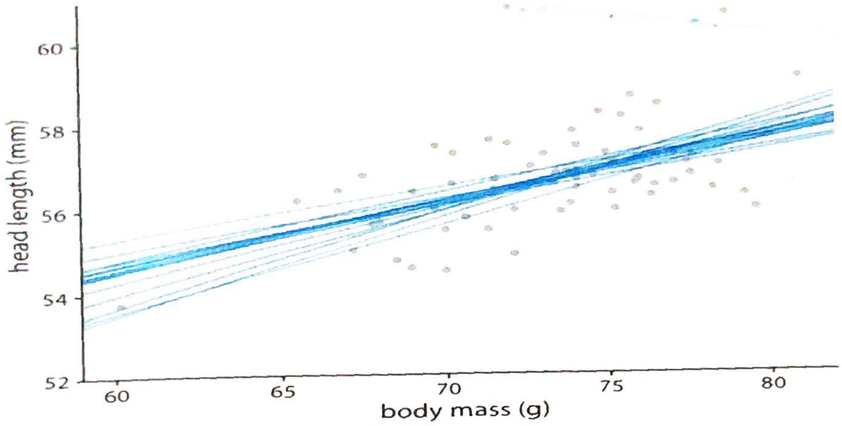
The confidence band provides us with a range of different fit lines that would be compatiable with the data. The reason for the curvature is that straight line fit produces a confidence band can in two distinct directions.

(i) It can move up and down ie it have different intercepts.

(ii) It can rotate ie it have different slopes.

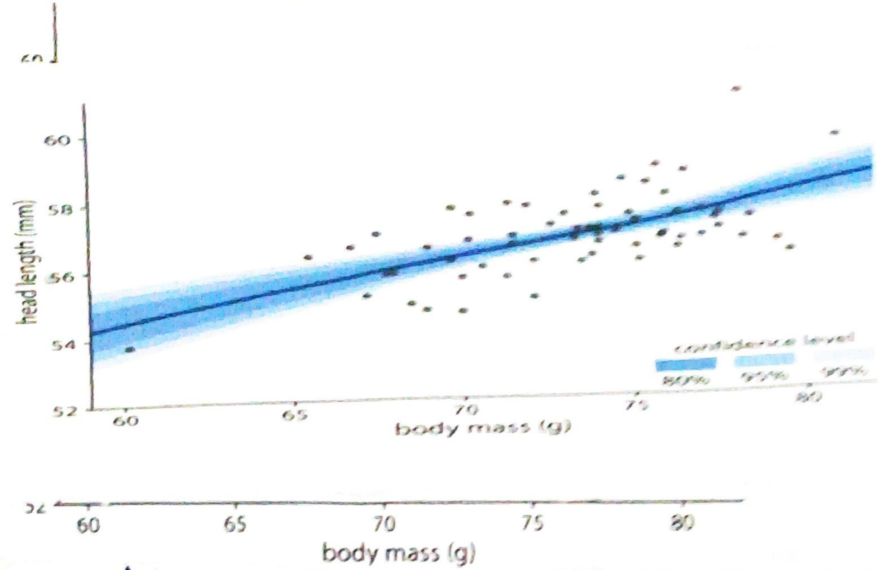
Let us consider an example the head length versus body mass for male blue jays, the straight line represents the best linear fit to the data and the gray band around the line shows the uncertainty in the linear fit. The gray band represents a 95% confidence level.



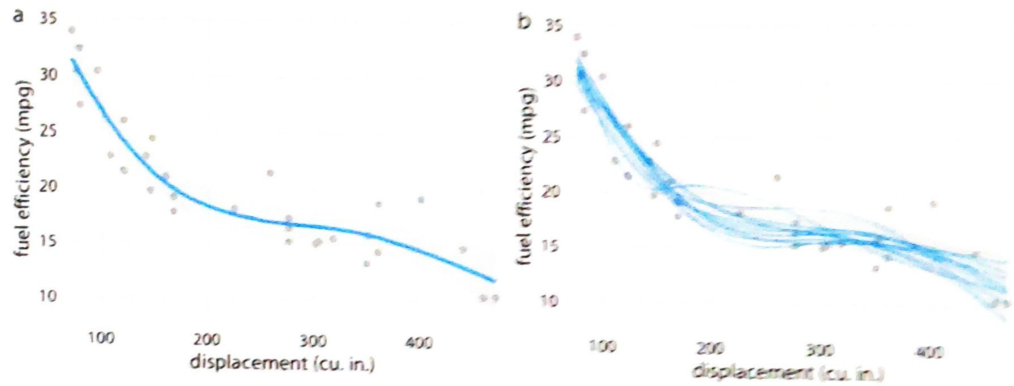


The straight blue lines now represent equally likely alternative fits randomly drawn from the posterior distribution.

To draw a confidence band, we need to specify a confidence level and just as we saw for error bars and posterior probabilities it can be useful to highlight different levels of confidence. This leads us to the graded confidence band, which shows several confidence levels at once.



We can also draw confidence band for non-linear curve fits. Such confidence band look nice but can be difficult to interpret.



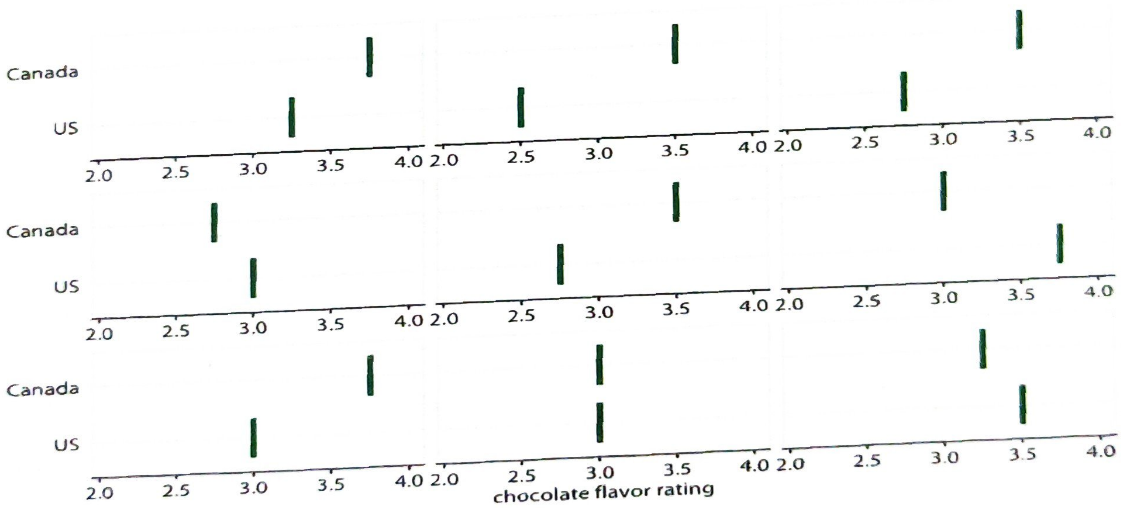
~~Visualizing~~ uncertainty

Hypothetical outcome plots :- (Hop)

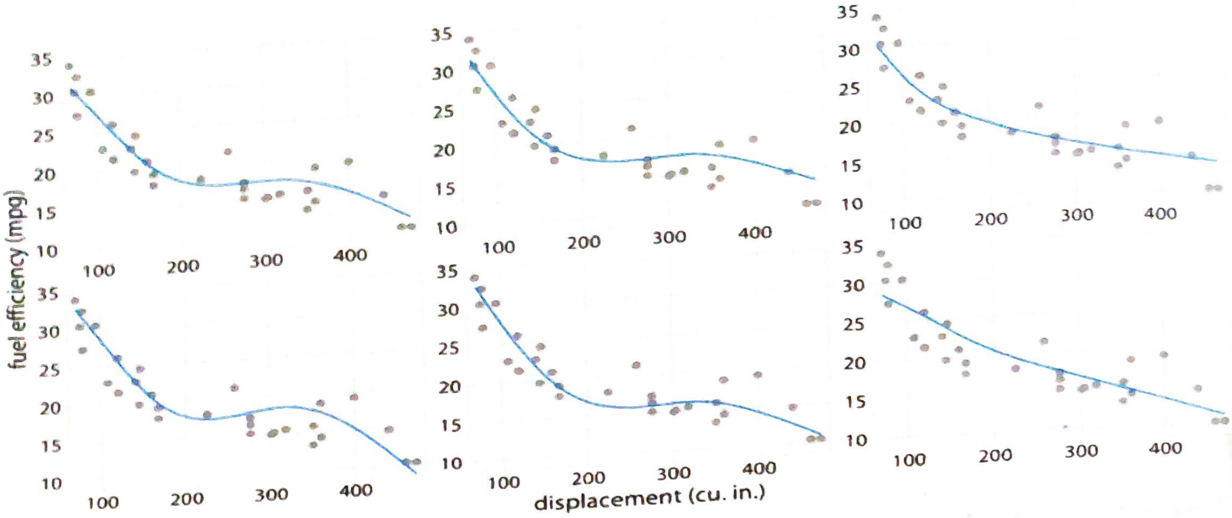
Hop is a approach to visualizing uncertain data. Rather than showing a continuous probability distribution, Hops visualize a set of drawn from a distri-bution where each draw is shown as a new plot in either a small multiples or animated form : ~~H~~

Hop's are not possible in print medium they can be very effective in online settings where animated visualizations can be provided in the form of GIFs or MP4 videos.

To illustrate the concept of Hop. Let us consider the following figure of point edition. schematic of a Hop for chocolate bar ratings of candian and U.S manufactured bars. Each vertical green bar represents the rating for one bar and each panel shows a comparison of two randomly chosen bars, one each from a canadian manufacturer and a U.S manu-facturer.



Let us consider another example schematic of a Hop for fuel efficiency Vs displacement. Each dot represents one car and the smooth lines were obtained by fitting a cubic regression spline with 5 knots. Each line in each panel represents one alternative fit outcome, drawn from the posterior distribution of the fit parameters.



Assignment

1) Explain the framing probabilities & frequencies

2) Define uncertainty, Explain visualizing the uncertainty using point estimates and Hop's

3) Define Time series?, Explain the time series of Dose-Response curves.

4) Explain the time series of two or more response variables.