Statistics 120
Multipanel Conditioning Plots

Tonga

Tonga   Trench   Earthquakes
Yellow: 0 − 70 km
Orange: 71 − 300 km
Red: 300 − 800 km.
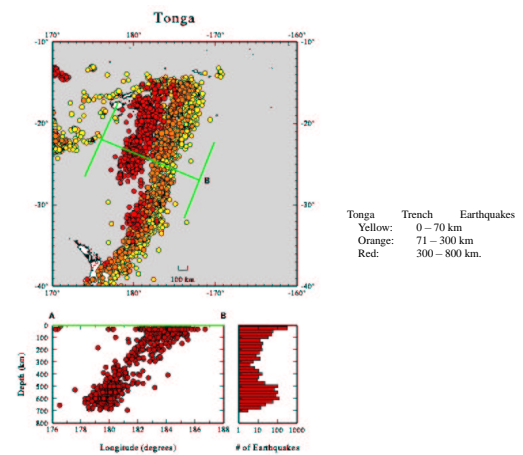
Trellis Graphics

- Trellis Graphics is a family of techniques for viewing complex, multi-variable data sets.

- The ideas have been around for a while, but were formalized by researchers at Bell Laboratories during the 1990s.

- The techniques were given the name *Trellis* because they usually result in a rectangular array of plots, resembling a garden trellis.

- A number of statistical software systems provide multi-panel conditioning plots under the name *Trellis* plots or *Crossplots*.

Problems with this Presentation

- There is a good deal of overplotting and this makes it hard to see all of the structure present in the data.

- The map makes it clear that we are looking down from above on the scene, but deeper quakes appear to be plotted on top of shallower ones.

- The division of depths into three intervals and presentation using colour is relatively crude.


Conditioning

- Trellis plots are based on the idea of *conditioning* on the values taken on by one or more of the variables in a data set.

- In the case of a categorical variable, this means carrying out the same plot for the data subsets corresponding to each of the levels of that variable.

- In the case of a numeric variable, it means carrying out the same plots data subsets corresponding to intervals of that variable.

A Trellis Plot

- We can overcome many of the problems of the previous plot by using a trellis display.

- We create the display by producing a sequence of graphs, each of which presents a different range of depths.

- In this case we will have a slight overlap of the intervals being plotted.

Example: Earthquake Locations

- R contains a data set called quakes which gives the location and magnitude of earthquakes under the Tonga Trench, to the North of New Zealand.

- The spatial distribution of earthquakes in the area is of major interest, because this enables us to "see" the structure of the earthquake faults.

- Here is a plot from the Geology department at Berkeley, which tries to present the the spatial structure.
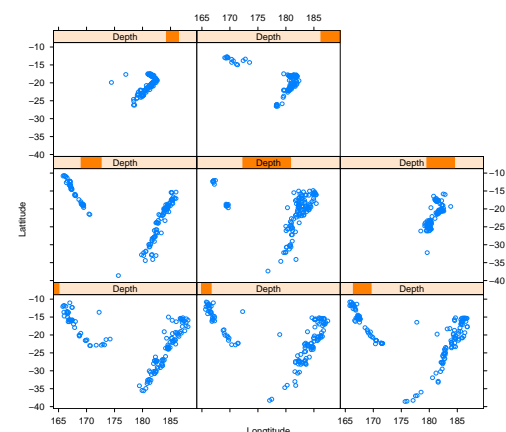
Explanation

- The plot is read left-to-right and bottom-to-top.

- Depth increases progressively through the plot.

- There are eight different depth intervals, each containing approximately the same number of earthquakes.

- Consecutive depth intervals overlap by a small amount.

- The range of depths covered by each interval is indicated in the bar above each plot.
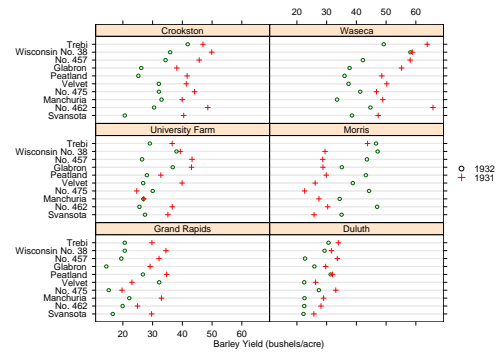
Intepretation

- The shallower earthquakes are concentrated on two inclined fault planes.

- The most easterly of these fault planes is the one which bisects New Zealand.

- The Westerly fault plane has mainly shallow earthquakes, while the Easterly fault plane has both shallow and deep earthquakes.

- The deep earthquakes show distinct small angular fishhook structure which is not visible in the earlier plot.

Interpretation

- When the data are present in this way it is easy to see what is odd about them.

- The yields for 1931 were higher than those for 1932 at all sites except Morris, where the pattern is reversed.

- The clear implication is that the yield values at Morris were switched at some point.

Example: Barley Yields

- This example is concerned with the yields obtained from field trials of barley seed.

- The data comes from the 1930s so there is no direct genetic modification going here.

- The trials were conducted in 1931 and 1932, using:
  - 10 different strains of barley
  - 6 different growing sites

- There are $2 \times 10 \times 6 = 120$ observations.

- It was suspected for a long time that there was something odd about this data set.

The Trellis Technology

- There are a variety of displays which can be produced by Trellis, including:
  - Bar Charts
  - Dot Charts
  - Box and Whisker Plots
  - Histograms
  - Density Traces
  - QQ Plots
  - Scatter Plots

- A common framework is used to produce all these plots.

The Trellis Plot

- The plot we will look at shows that barley yields for each of the 10 strains at the 6 sites and for each year.

- The results for each site are plotted on a separate graph – i.e. we are working conditional on the site.

- The yields from the two years are superimposed on each of the plots.

Some Terminology

- Every Trellis display consists of a series of rectangular *panels*, laid out in a regular row-by-column array.

- The indexing of the array is left-to-right, bottom-to-top.

- The x axes of all the panels are identical. This is also true for the y axes.

- Each panel of the a display corresponds to conditioning, either on the levels of a factor, or on sub-intervals of the range of a numeric variable.

Shingles

- The conditioning carried out in the earthquake plot is described by a *shingle*.

- A shingle consists of a number of overlapping intervals (like the shingles on a roof of a house).

- Assuming that the earthquake depths are contained in the variable depth, the shingle is created as follows.

```
> Depth = equal.count(depth,
                       number=8,
                       overlap=.1)
```

- The shingle assigned to Depth has 8 intervals with adjacent intervals having 10% of their values in common.

The Plot Formula

- The first argument to xyplot is a symbolic formula describing the plot.

- In this case the formula is:

```
lat ~ long | Depth
```

which is an instruction to plot lat on the *y* axis against long on the *x* axis with conditioning intervals as described in Depth.

- The second argument to xyplot specifies which data frame the data for the plot should be obtained from.

- Additional arguments control other aspects of the plot.

Shingles

- A shingle contains the numerical values it was created from and can be treated like a copy of that variable. For example:

```
> range(Depth)
[1]   40 680

> range(depth)
[1]   40 680
```
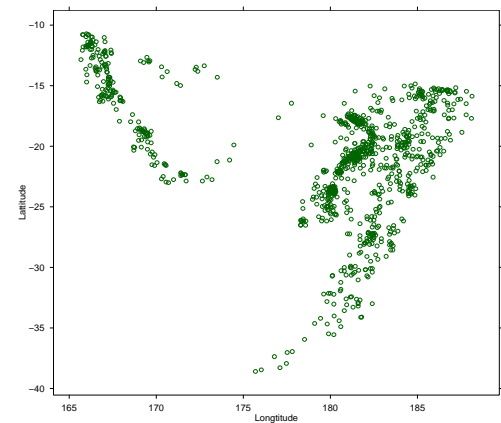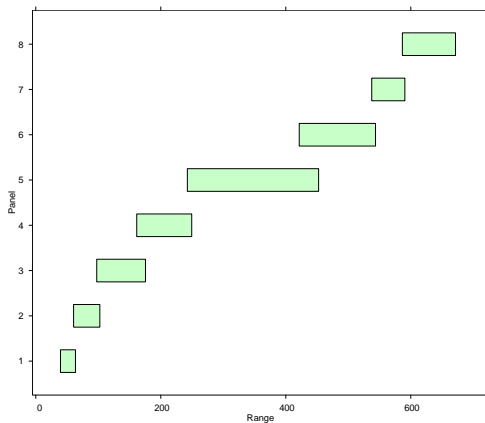
- A shingle also has the information attached to it. This can be displayed by printing or plotting the shingle.

```
> plot(Depth)
```

Unconditional Plots

- The xyplot function can be used to produced an unconditional plot by omitting the conditioning specification from the plot formula.

```
> xyplot(lat ~ long, data = quakes,
         xlab = "Longtitude",
         ylab = "Lattitude")
```


Producing the Plot

- The display of the earthquakes is produced by the function xyplot, which is the Trellis variant of a scatter plot function.

- The plot was produced as follows:

```
> data(quakes)
> Depth = equal.count(quakes$depth,
                      number=8,
                      overlap=.1)
> xyplot(lat ~ long | Depth, data = quakes,
         xlab = "Longtitude", ylab = "Lattitude")
```

- There are three steps here (i) loading the data, (ii) creating the shingle and (iii) producing the display.
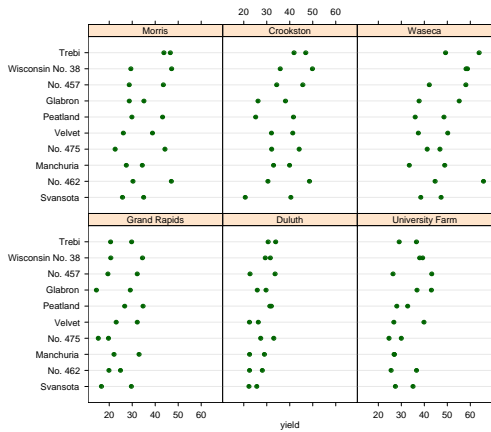
The Barley Yield Plot

- The barley yield plot is produced by the function dotchart which can be used to numeric values against a categorical variable.

- In this case, the numeric variable is the barley yield and the categorical variable is the seed strain.

- We also condition on the value of another variable, the growing site.

A First Attempt

- The following code is a simple attempt at creating a dot chart using similar code to that for the earthquakes.
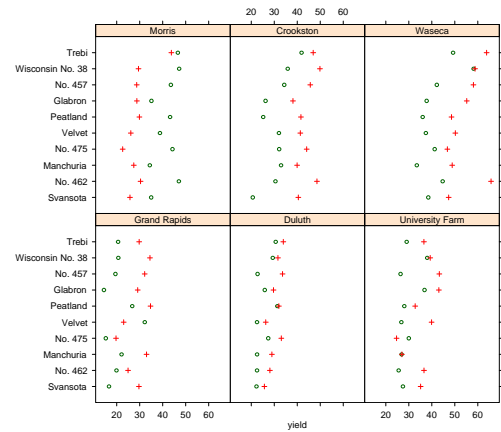
```
> dotplot(variety ~ yield | site,
        data=barley)
```


A Second Attempt

- We could also try conditioning on both site and year.

```
> dotplot(variety ~ yield | site * year,
        data=barley)
```


A Third Attempt

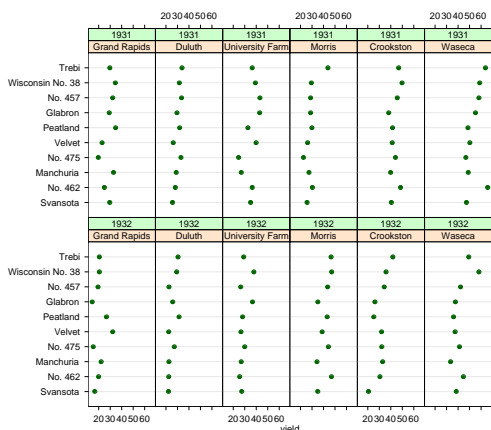- What we need is to superimpose the two years for each site on a single panel.

```
> dotplot(variety ~ yield | site, data=barley,
        panel = panel.superpose,
        group = year,
        pch = c(1,3))
```


A Fourth Attempt

- The last plot is quite close.

- We need to add a legend which indicates which year is which.

```
> dotplot(variety ~ yield | site, data=barley,
        panel = panel.superpose,
        group = year, pch = c(1,3),
        key = list(space="right",
                transparent = TRUE,
                points=list(pch=c(1,3),
                col=1:2),
                text=list(c("1932",
                        "1931"))))
```