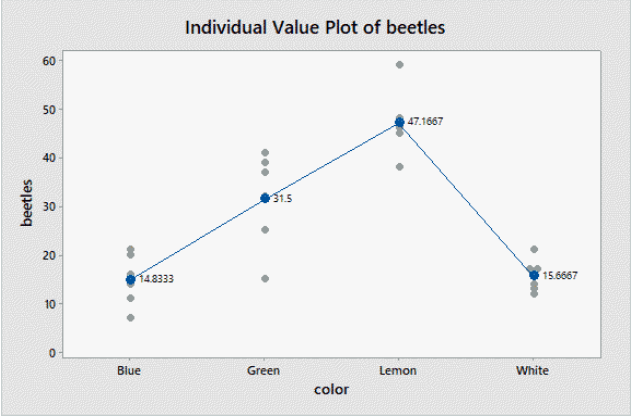
## Supplementary exercise 2.7 of IPS7e

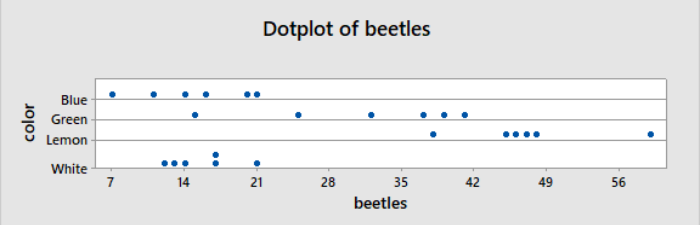
A study on trapping of cereal leaf beetles in an oat field on boards of different colours. Four board colours (lemon yellow, white, green, blue) were used, and there were six boards per colour.

(a) We use Minitab to display the counts in an Individual Value Plot, and to compute descriptive statistics. In Minitab 18, the mean values can be added directly to the Individual Value Plot by selecting Means as the Label in the `Labels-Data Labels` submenu. It is also possible to add symbols and connecting lines for the means in the `Data View` submenu; the resulting plot is shown below. However, it is not obvious why one would want to connect the means across the different categories (colours); such connecting lines could be viewed as misleading because the order of the categories has no real meaning.

Another good and simple option to display the data points is a dotplot, separated by groups. This plot will not show the means, but still gives a good impression of the differences between groups. A boxplot will instead display the 5-number summary, but with only 6 observations per group these statistics are not estimated very well, so there is a risk of over-interpreting the visual display. It is however possible to add observed means to the plot (using the same steps as described above).

```
Indplot ( 'beetles' ) * 'color';
  MeaLabel;
    YValue;
  Individual;
  Mean;
  CMean.
Dotplot ( 'beetles' ) * 'color'.
Describe 'beetles';
  By 'color'.
```




**Descriptive Statistics: beetles**

**Statistics**

| Variable | color | N | N* | Mean | SE Mean | StDev | Minimum | Q1 | Median | Q3 | Maximum |
|---|---|---|---|---|---|---|---|---|---|---|---|
| beetles | Blue | 6 | 0 | 14.83 | 2.18 | 5.34 | 7.00 | 10.00 | 15.00 | 20.25 | 21.00 |
| | Green | 6 | 0 | 31.50 | 4.05 | 9.91 | 15.00 | 22.50 | 34.50 | 39.50 | 41.00 |
| | Lemon | 6 | 0 | 47.17 | 2.77 | 6.79 | 38.00 | 43.25 | 46.50 | 50.75 | 59.00 |
| | White | 6 | 0 | 15.67 | 1.36 | 3.33 | 12.00 | 12.75 | 15.50 | 18.00 | 21.00 |

(b) Although the data set is small, it seems that the lemon yellow boards attract most beetles, followed by the green boards. There is a little overlap in the distributions for these two colours. The white and blue boards seem to attract substantially less beetles than the lemon yellow board because the beetle counts for those two colours are all less than those for the lemon yellow board; in other words, the distributions appear completely separated. The largest spread in the beetle counts is observed for the green boards, with the lowest count as low as for the blue and white

boards and with the highest counts as high as for the lemon yellow board. Without a proper statistical analysis it is impossible to say whether the differences observed could have happened by chance alone (that is, whether they are statistically significant), but intuitively one would expect at least the differences between the lemon yellow and the two boards with lowest counts (blue and white) to be significant.

(c) No. The board colour is a categorical variable, and we can only speak about positive or negative associations between quantitative variables.