Creating Cleveland Dotplots in R

Ralph Mansson

Introduction

Dotplots

To illustrate the Cleveland dot plot we consider data on area equipped for irrigation in four regions of the World. The R code to create the data object is shown here:

\[
\]

Base Graphics

In the base graphics system we can build up the dotplot step by step. The first function call creates the graph region based on the data set but we do not plot any data by setting the type = "n" option. The axis labels for the horizontal and vertical scales are set along with the title:

\[
\text{plot(irrigation.df$Area, irrigation.df$Region, xlab = "Area", ylab = "Region", main = "Irrigation Area by Region", type = "n")}
\]

To add the points with separate colours for each of the four years we use the points function and subset to the particular year by testing a condition on the year.
The `ggplot2` version of the graph is shown:

```r
ggplot(irrigation.df, aes(x = Area, y = Region, colour = Year)) + geom_point() +
  opts(title = "Irrigation Area by Region")
```

The lattice graphics function is used to create the dot plot:

```r
dotplot(Region ~ Area, data = irrigation.df, groups = Year, main = "Irrigation Area by Region")
```

The code is rather long winded compared to the other two graphics packages. We can add a legend to the graph so that the years can be identified:

```r
```

The graph is basic but we can consider the changes over time for the four regions. One downside is that the regions have been labelled with numbers rather than text strings.

Lattice Graphics

The `lattice` graphics package has a function `dotplot` that is used to create dot plots. The first argument to the function is a formula describing the variables to use for the horizontal and vertical axes. We also specify the data frame to use for the graph and which column to determine different symbols and/or colours to highlight groupings within the plot.

```r
dotplot(Region ~ Area, data = irrigation.df, groups = Year, main = "Irrigation Area by Region")
```

The lattice variant of the graph is shown here:

The ggplot2 version of the graph is shown:

```r
ggplot(irrigation.df, aes(x = Area, y = Region, colour = Year)) + geom_point() +
  opts(title = "Irrigation Area by Region")
```

The `ggplot2` function is used to create the dot plot where we first specify the data frame with the information to be displayed and then use the `aes` argument to list the variables to plot on the horizontal and vertical axes. The `colour` argument determines the variable to use for assigning colours to (usually) a categorical variable.

```r
ggplot(irrigation.df, aes(x = Area, y = Region, colour = Year)) + geom_point() +
  opts(title = "Irrigation Area by Region")
```

The graph is simple and very similar to the one produced using the base graphics with the advantage that the code is not as complicated.

```r
dotplot(Region ~ Area, data = irrigation.df, groups = Year, main = "Irrigation Area by Region")
```

The lattice version of the graph is shown here: