

# Advanced graphics: practical 2

## Jumping Rivers

This practical aims to guide you through some of the key ideas in data manipulation. I've tried to construct this practical in such a way that you get to experiment with the various tools. Feel free to experiment!

### 1 Factors

When using ggplot2, the easiest way of rearranging the graph or to alter labels is to manipulate the data set. Consider the mpg data set:

```
data(mpg, package = "ggplot2")
```

Suppose we generate a scatter plot of engine displacement against highway mpg.

```
g = ggplot(data=mpg, aes(x=displ, y=hwy)) +  
  geom_point()
```

Next, we add a loess line, conditional on the drive type:

```
g + stat_smooth(aes(colour=drv))
```

While this graph is suitable for exploring the data; for publication, we would like to rename the axis and legend labels. To change the axis labels, we can rename the data frame columns or use xlab and ylab. To change the order of the legend, we need to manipulate the data. Since drv is a character, we could use:

```
mpg[mpg$drv == "4",]$drv = "4wd"  
mpg[mpg$drv == "f",]$drv = "Front"  
mpg[mpg$drv == "r",]$drv = "Rear"
```

However the legend will still be ordered alphabetically. Instead, we can use a factor:

```
##Reload the data just to make sure  
data(mpg, package="ggplot2")  
mpg$drv = factor(mpg$drv, labels = c("4wd", "Front", "Rear"))
```

To change the order of the, we need to use the factor function:

```
mpg$drv = factor(mpg$drv,  
  levels = c("Front", "Rear", "4wd"))
```

The legend now displays the labels in the order: Front, Rear and 4wd.

### 2 Aphids

This data set consists of seven observations on cotton aphid counts on twenty randomly chosen leaves in each plot, for twenty-seven

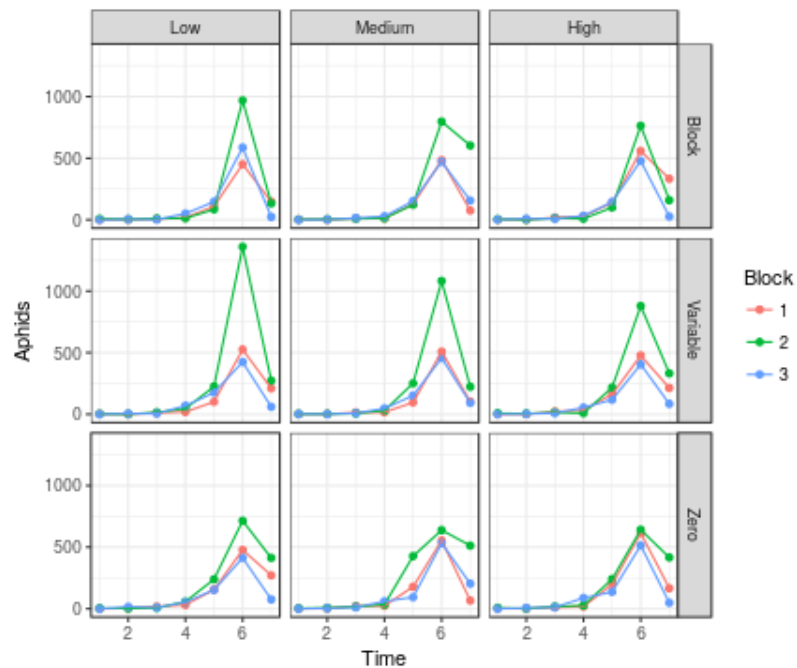


Figure 1: Final figure from section 2.

treatment-block combinations. The data were recorded in July 2004 in Lamesa, Texas. The treatments consisted of three nitrogen levels (blanket, variable and none), three irrigation levels (low, medium and high) and three blocks, each being a distinct area. Irrigation treatments were randomly assigned within each block as whole plots. Nitrogen treatments were randomly assigned within each whole block as split plots.

```
data(aphids, package="jrSouth")
```

The sampling times are once per week.

REPRODUCE figure 1. Here are some hints to get you started. The key idea is to think of the plot in terms of layers. So

- Leave the ordering of factors to the end
- The plot contains a combination of `geom_line` and `geom_point`.
- You can change the x-axis label using

```
+ xlab("Time")
```

- Change the theme using `theme_bw()`

## *Solutions*

Solutions are contained within this package:

```
library(jrSouth)  
vignette("ggplot2_solutions2", package="jrSouth")
```