Advanced graphics: practical 1

Dr Colin S. Gillespie

First install the `ggplot2` package

```
install.packages("ggplot2")
```

then install the course package

```
##Windows & Linux users
install.packages("nclRggplot2",
                 repos="http://R-Forge.R-project.org")
##Apple users
install.packages("nclRggplot2",
                 repos="http://R-Forge.R-project.org",
                 type="source")
```

To load the package, use

```
library(nclRggplot2)
```

1 Practical 1

To get you familiar with the underlying `ggplot2` concepts, we'll recreate some standard graphics. Some these of plots aren't particularly useful, we are just using them for illustration purposes.

To BEGIN WITH, install the package

```
install.packages("ggplot2")
```

then load the package:

```
library(ggplot2)
```

Next we load the beauty data set:[1]

[1] Details of the beauty data set can be found at the end of this practical.

```
library(nclRggplot2)
data(Beauty)
```

When loading in data, it's always a good idea to carry out a sanity check. I tend to use the commands

```
head(Beauty)
colnames(Beauty)
dim(Beauty)
```

2 *Scatter plots*

Scatter plots are created using the point geom. Let's start with a basic scatter plot

```
ggplot(data=Beauty) + geom_point(aes(x=age, y=beauty))
```

To save typing, we can also store the plot as a variable:

In this practical, we are creating the plots in a slightly verbose way.

```
g = ggplot(data=Beauty)
g1 = g + geom_point(aes(x=age, y=beauty))
```

To view this plot, type `g1`.

The arguments `x` and `y` are called aesthetics. For `geom_point`, these parameters are required. This particular geom has other aesthetics: `shape`, `colour`, `size` and `alpha`.[2] Here are some things to try out.

[2] These aesthetics are usually available for most geoms.

- Experiment with other aesthetics. For example,

  ```
  g + geom_point(aes(x=age, y=beauty, colour=gender))
  ```

  or

  ```
  g + geom_point(aes(x=age, y=beauty,
                     alpha=evaluation, colour=gender))
  ```

  Some aesthetics, like `shape` must be discrete. So use `shape = factor(tenured)`.

- Are there any differences between numeric values like `tenured` and characters like `gender` for some aesthetics? What happens if you convert `tenured` to a factor in the `colour` aesthetic. For example, `colour = factor(tenured)`.

- What happens if you set `colour` (or some other aesthetic) outside of the `aes` function? For example, compare

  ```
  g + geom_point(aes(x=age, y=beauty, colour="blue"))
  ```

  to

  ```
  g + geom_point(aes(x=age, y=beauty), colour="blue")
  ```

- What happens when you set an aesthetic to a constant value. For example, `colour=2`. What happens if you put this argument outside of the `aes` function?

3 *Box plots*

The box plot geom has the following aesthetics: `x`, `y`, `colour`, `fill`, `linetype`, `weight`, `size` and `alpha`. We can create a basic boxplot using the following commands:

Adjustment	Description
`dodge`	Adjust position by overlapping to the side
`fill`	Stack overlapping elements; standardise stack height
`identity`	Do nothing
`jitter`	Jitter points
`stack`	Stack overlapping elements

Table 1: Position adjustments - table 4.5 in the ggplot2 book.

```
g + geom_boxplot(aes(x=gender, y=beauty))
```

Similar to the point geom, we can add in aesthetics:

```
g + geom_boxplot(aes(x=gender, y=beauty,
                             colour=factor(tenured)))
```

Why do you think we have to convert `tenured` to a discrete factor?

As before, experiment with the different aesthetics. For some of the aesthetics, you will need to convert the continuous variables to discrete variables. For example, this will give an error:

```
g + geom_boxplot(aes(x=gender, y=beauty, colour=tenured))
```

while this is OK

```
g + geom_boxplot(aes(x=gender, y=beauty, colour=factor(tenured)))
```

Make sure you play about with the different aesthetics.

4 *Combining plots*

The key idea with `ggplot2` is to think in terms of layers not in terms of plot "types". For example,

In the lectures we will discuss what this means.

```
g + geom_boxplot(aes(x=gender, y=beauty,
                           colour=factor(tenured))) +
  geom_point(aes(x=gender, y=beauty))
```

- What happens to the plot if you swap the order of the `geom_boxplot` and `geom_point` function calls?
- In this case, `geom_point` isn't that great. Try using `geom_jitter`:

  We have a bit too much data for `geom_jitter`, but you get the point.

  ```
  g + geom_boxplot(aes(x=gender, y=beauty,
                             colour=factor(tenured))) +
    geom_jitter(aes(x=gender, y=beauty))
  ```

tenured	minority	age	evaluation	gender	students	beauty
0	1	36	4.3	Female	43	0.202
1	0	59	4.5	Male	20	-0.826
1	0	51	3.7	Male	55	-0.660
1	0	40	4.3	Female	46	-0.766
0	0	31	4.4	Female	48	1.421

Table 2: The first five rows of the beauty data set. There are a total of 463 course evaluations.

5 *Bar plots*

The bar geom has the following aesthetics: `x`, `colour`, `fill`, `size`, `linetype`, `weight` and `alpha`. Here is a command to get started:

```
g + geom_bar(aes(x=factor(tenured)))
```

- As before, try different aesthetic combinations. Convert parameters to discrete versions as needed using `factor(...)`.
- Let's get a bit more fancy. First, we round ages to the nearest decade:

  ```
  Beauty$dec = signif(Beauty$age, 1)
  ```

  then plot:

  ```
  g = ggplot(data=Beauty)
  g + geom_bar(aes(x=gender, fill=factor(dec)))
  ```

  We can adjust the layout of this bar plot using ggplot's position adjustments. The five possible adjustments are listed in table 1. The default adjustment is `stack`

  ```
  g + geom_bar(aes(x=gender, fill=factor(dec)),
               position="stack")
  ```

- Try the other adjustments.

6 *The beauty data set*

This data set is from a study where researchers were interested in whether a lecturers' attractiveness affected their course evaluation.[3] This is a cleaned version of the data set and contains the following variables:

[3]

- `evaluation` - the questionnaire result.
- `tenured` - does the lecturer have tenure; 1 == Yes. In R, this value is continuous.

- `minority` - does the lecturer come from an ethnic minority (in the USA).
- `age`.
- `gender`.
- `students` - number of students in the class.
- `beauty` - each of the lecturers' pictures was rated by six undergraduate students: three women and three men. The raters were told to use a 10 (highest) to 1 rating scale, to concentrate on the physiognomy of the professor in the picture, to make their ratings independent of age, and to keep 5 in mind as an average. The scores were then normalised.

Table 2 gives the first few rows of the data set.