Data visualization with ggplot

Mark Andrews

April 5, 2017

Introduction

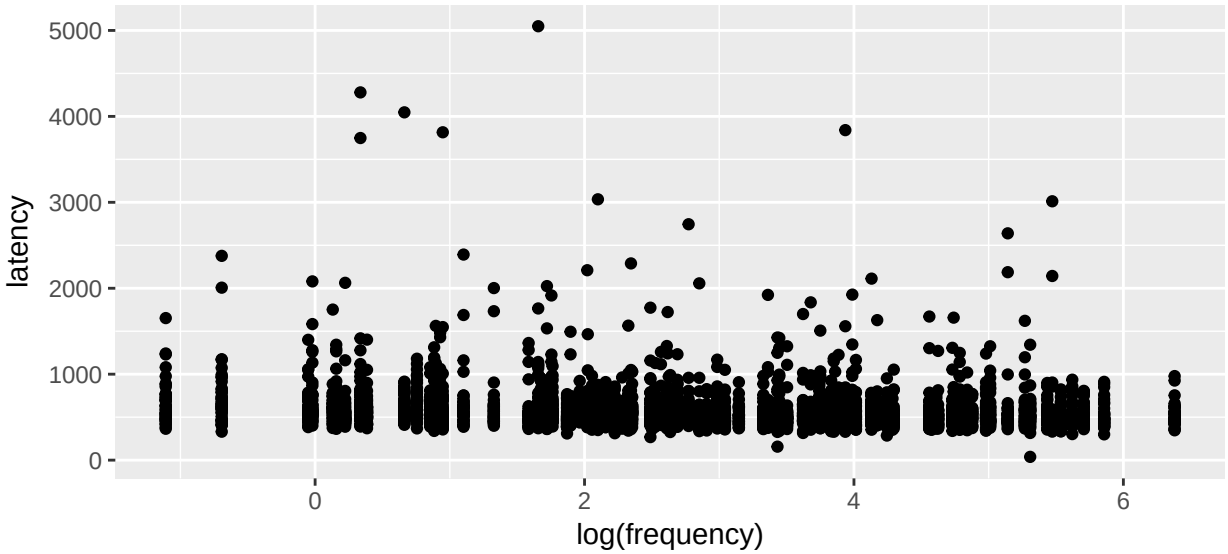
`ggplot` is essentially a mini-language for data visualization. Because of its complexity and versality, it takes some getting used to. It is worth the effort. It allows you to explore your data quickly and efficiently.

```
library(ggplot2)
Df <- read.csv('../data/LexicalDecision.csv', header=T)
```

A simple scatter plot

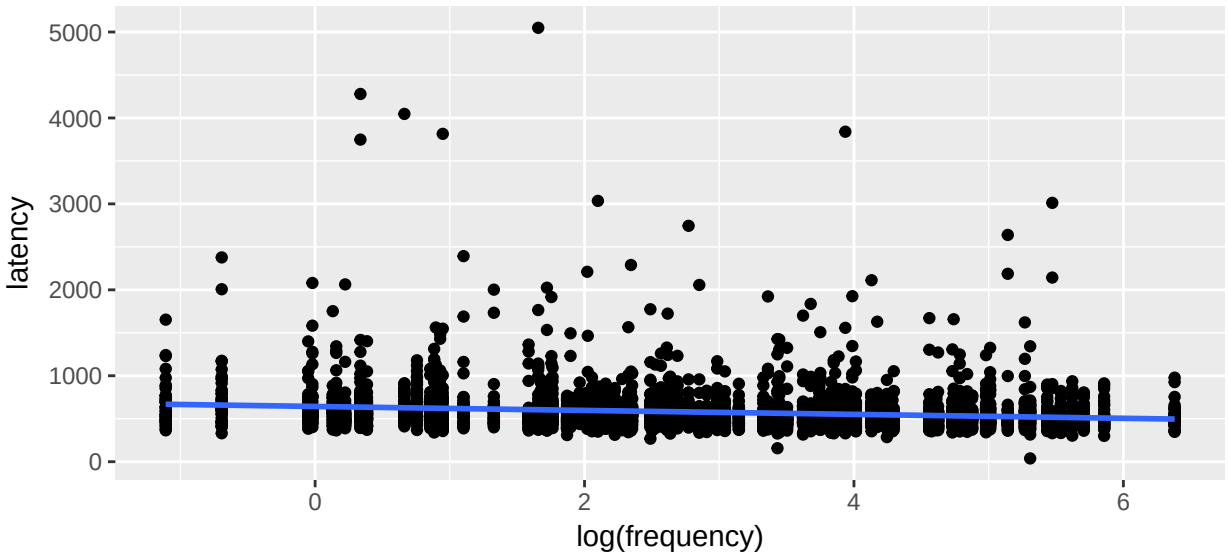
We start by specifying the data frames and which variables should be assigned to the x and y axess. Then we specify that we want points on the values.

```
ggplot(Df, aes(x=log(frequency), y=latency)) + geom_point()
```



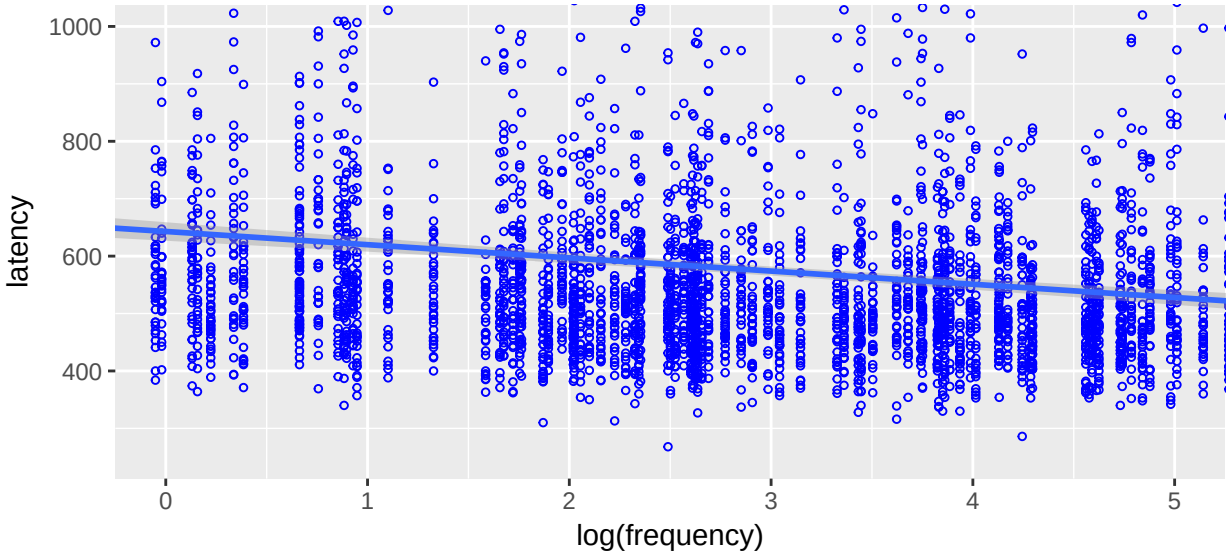
We can add another layer that is a line of best fit, plus standard error:

```
ggplot(Df, aes(x=log(frequency), y=latency)) +
  geom_point() +
  stat_smooth(method='lm')
```



We can play with the properties of everything:

```
ggplot(Df, aes(x=log(frequency), y=latency)) +
  geom_point(size=1, shape=1, colour='blue') +
  stat_smooth(method='lm') +
  coord_cartesian(xlim = c(0, 5), ylim=c(250, 1000))
```

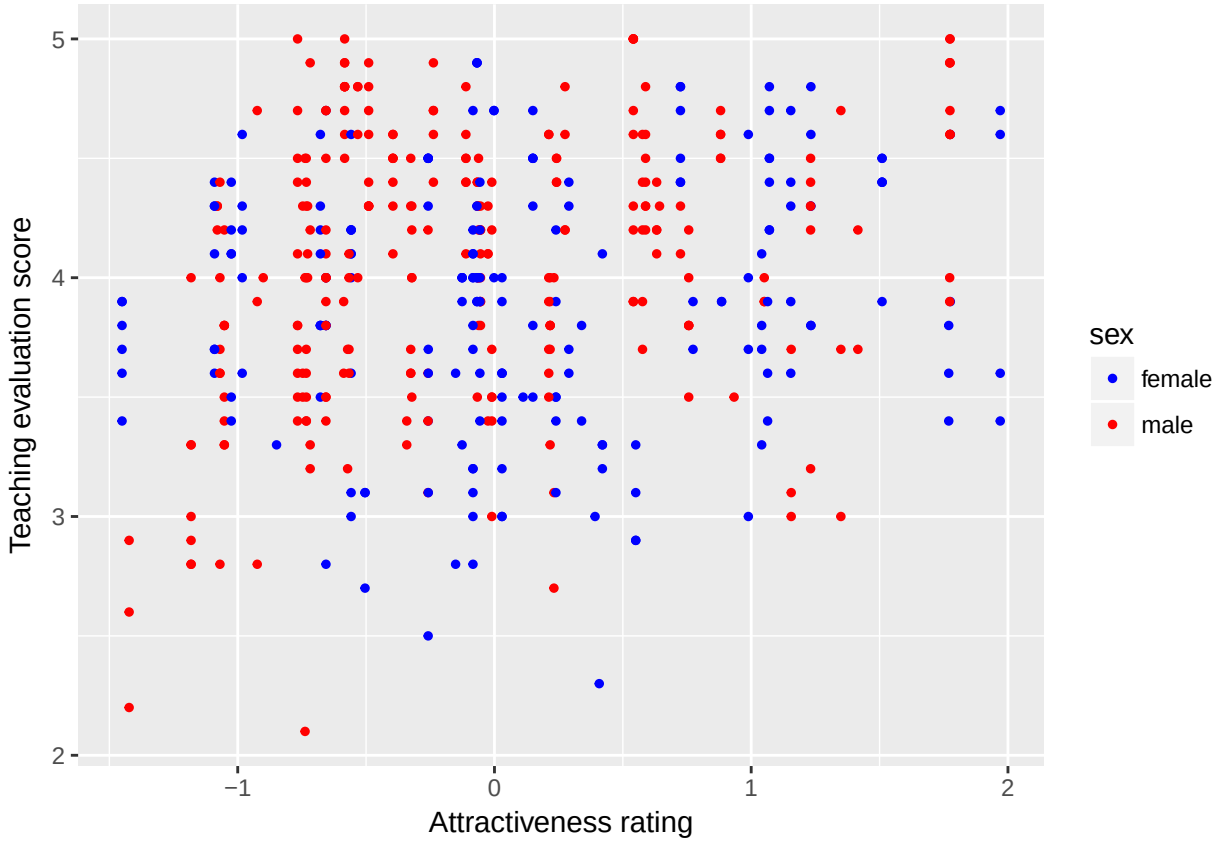


Colour and shape codes

One of the great things about ggplot is how it allows us to color code our data. For example, here we make a scatter plot and color code the points that belong to males and females. This is possible in base R plots, but it is just much easier here.

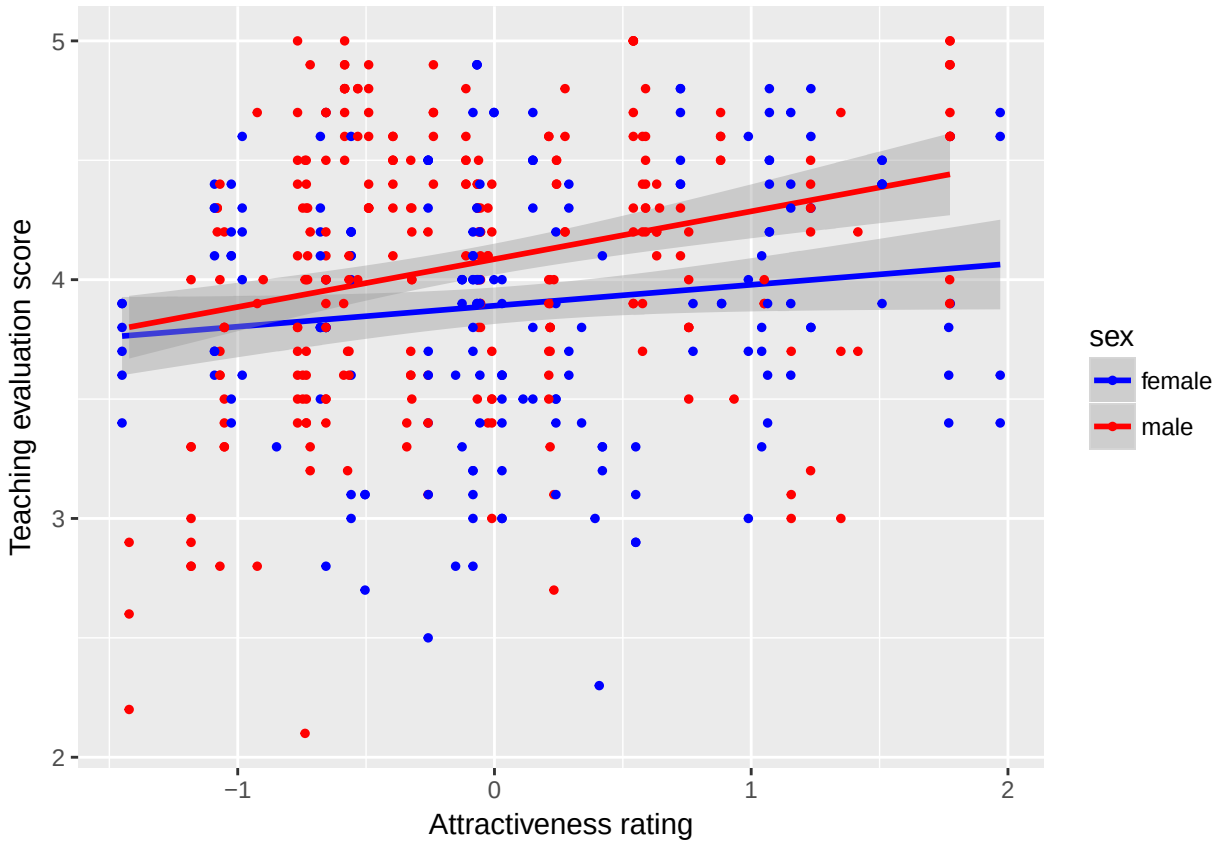
```
# Load a new data set
load('../data/beautyeval.Rda')

# Make scatterplot, point colour by male/female
```

```
ggplot(beautydata, aes(x=beauty, y=eval, colour=sex)) +
  geom_point(size=1) +
  scale_color_manual(values=c('blue', 'red')) +
  scale_y_continuous(name='Teaching evaluation score') +
  scale_x_continuous((name='Attractiveness rating'))
```



As before, we can superimpose lines of best fit, etc.

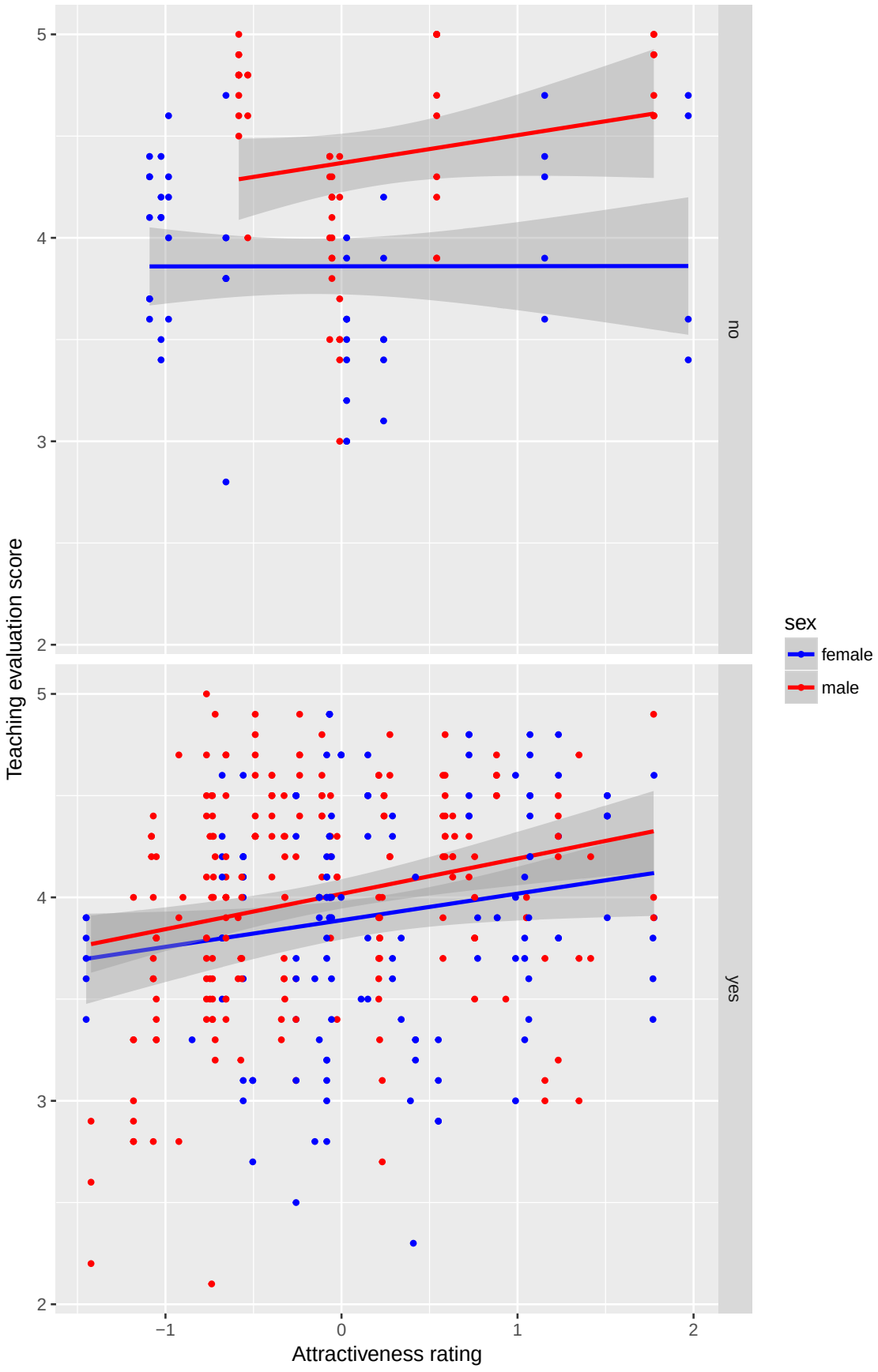
```
# Scatterplot, with lines of best fit and errors
ggplot(beautydata, aes(x=beauty, y=eval, colour=sex)) +
  stat_smooth(method='lm') +
  geom_point(size=1) +
  scale_color_manual(values=c('blue', 'red')) +
  scale_y_continuous(name='Teaching evaluation score') +
  scale_x_continuous((name='Attractiveness rating'))
```



Facets

Facets allow us to create multiple plots in the same way, with each one showing some subset of the data.

```
# Scatterplot, with lines of best fit and errors
# one per tenure group
ggplot(beautydata, aes(x=beauty, y=eval, colour=sex)) +
  stat_smooth(method='lm') +
  geom_point(size=1) +
  scale_color_manual(values=c('blue', 'red')) +
  scale_y_continuous(name='Teaching evaluation score') +
  scale_x_continuous((name='Attractiveness rating')) +
  facet_grid(tenure ~ .)
```


Teaching evaluation score
Attractiveness rating
5
4
3
2
−1
0
1
2
no
yes
sex
female
male