Basic Course on R:
Basic Plotting Practical Answers

Karl Brand* and Elizabeth Ribble†

18-24 May 2017

Contents

*brandk@gmail.com
†emcclel3@msudenver.edu

1 Basic Plotting

1.1 Use R to do the following exercises on the BOD data.

1.1.1 Display the built-in dataset called BOD by running BOD.

```
BOD
```

```
##   Time demand
## 1    1    8.3
## 2    2   10.3
## 3    3   19.0
## 4    4   16.0
## 5    5   15.6
## 6    7   19.8
```

1.1.2 What is the data structure of BOD? What are the dimensions?

```
str(BOD)
```

```
## 'data.frame': 6 obs. of  2 variables:
##  $ Time  : num  1 2 3 4 5 7
##  $ demand: num  8.3 10.3 19 16 15.6 19.8
##  - attr(*, "reference")= chr "A1.4, p. 270"
```

Using `str` we see that BOD is a data frame with dimensions 6 x 2, each variable (Time and demand) a numeric vector.

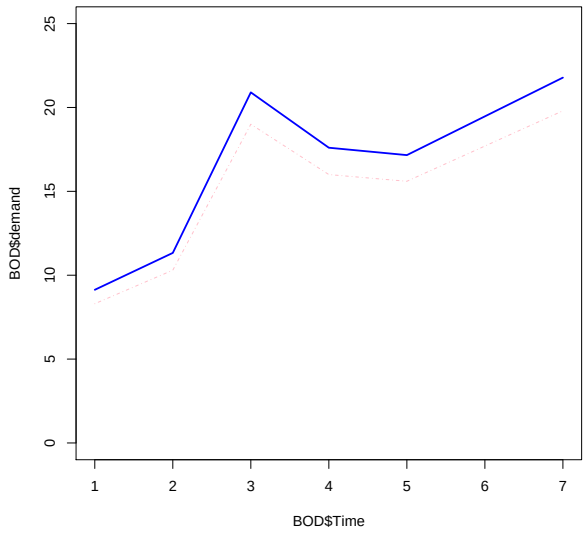
1.1.3 What are the names of BOD? Use a function other than `str`.

```
names(BOD)
```

```
## [1] "Time"   "demand"
```

a line graph of demand versus time, where the line is a pink dot-dashed line [Hint: run `?par` and look for the parameter `lty` to see the line types]. Add a blue dashed line of 1.1 times the demand and give it a thickness of 2 using the line width parameter `lwd`. Make sure both lines are entirely visible by adjusting the range of y using the parameter `ylim` in the original plot.

```
plot(BOD$Time, BOD$demand, type = "l", lty = 4,
     col = "pink", ylim = c(0, 25))
lines(BOD$Time, 1.1 * BOD$demand, lwd = 2, col = "blue")
```



1.2 Use R to do the following exercises on the chickwts data.

1.2.1 Display the built-in chickwts data.

```
chickwts
```

1.2.2 What is the data structure of chickwts? What are the dimensions?

```
str(chickwts)
```

```
## 'data.frame': 71 obs. of  2 variables:
##  $ weight: num  179 160 136 227 217 168 108 124 143 140 ...
##  $ feed  : Factor w/ 6 levels "casein","horsebean",..: 2 2 2 2 2 2 2 2 2 2 ...
```

Using `str` we see that `chickwts` is a data frame with dimensions 71 x 2, with variable `weight` a numeric vector and `feed` a factor with 6 levels.

1.2.3 What are the names of chickwts? Use a function other than str.

```
names(chickwts)

## [1] "weight" "feed"
```

1.2.4 What are the levels of `feed`?

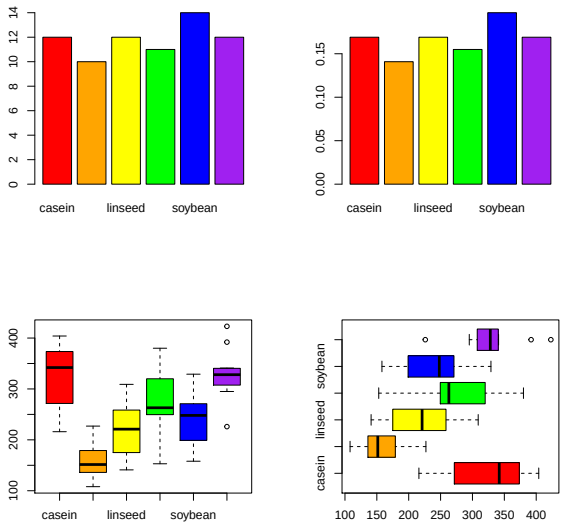
```
levels(chickwts$feed)

## [1] "casein"    "horsebean" "linseed"   "meatmeal"  "soybean"   "sunflower"
```

1.2.5 Make the following plots in one 2 x 2 image:

- A bar chart of the feed types, each bar a different color.
- A bar chart of the proportions of feed types, each bar a different color.
- A boxplot of the weights by feed type, each box a different color.
- A horizontal boxplot of the weights by feed type, each box a different color.

```
par(mfrow = c(2, 2))
barplot(table(chickwts$feed),
        col = c("red", "orange", "yellow",
                "green", "blue", "purple"))
barplot(table(chickwts$feed)/length(chickwts$feed),
        col = c("red", "orange", "yellow",
                "green", "blue", "purple"))
boxplot(chickwts$weight~chickwts$feed,
        col = c("red", "orange", "yellow",
                "green", "blue", "purple"))
boxplot(chickwts$weight~chickwts$feed,
        col = c("red", "orange", "yellow",
                "green", "blue", "purple"),
        horizontal = TRUE)
```



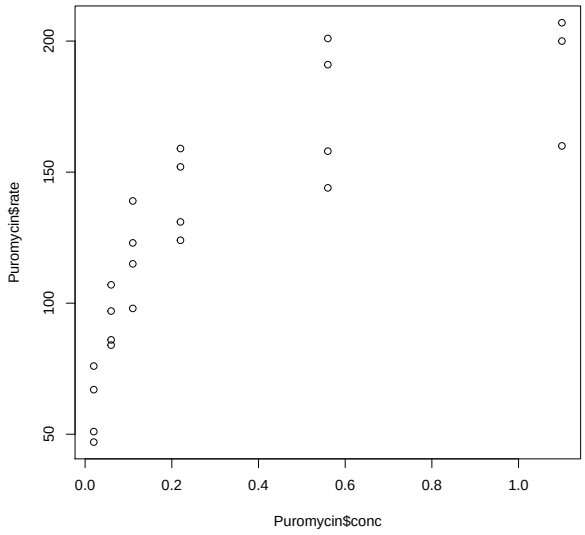
1.3 Use R to do the following exercises on the Puromycin data.

1.3.1 Display the built-in Puromycin data.

```
Puromycin
```

1.3.2 Make a scatterplot of the rate versus the concentration. Describe the relationship.

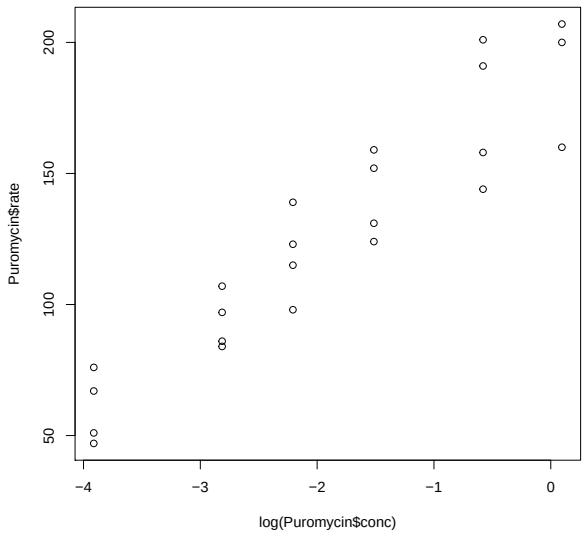
```
plot(Puromycin$conc, Puromycin$rate)
```



The rate increases faster at lower concentrations than at higher concentrations.

1.3.3 Make a scatterplot of the rate versus the log of the concentration. Describe the relationship.

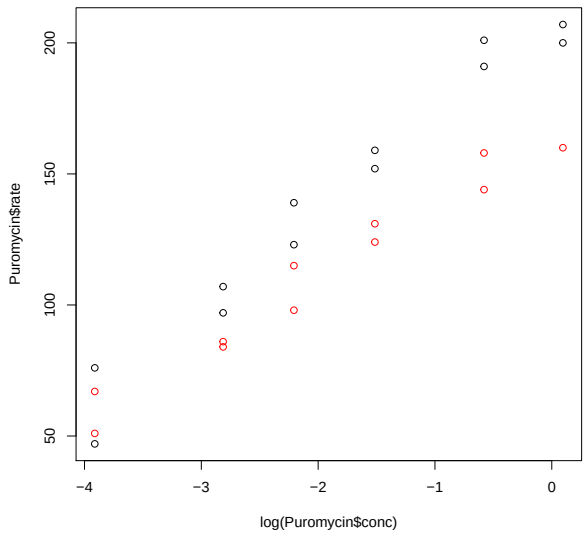
```
plot(log(Puromycin$conc), Puromycin$rate)
```



The two variables have a linear relationship.

1.3.4 Make a scatterplot of the rate versus the log of the concentration and color the points by treatment group (`state`). Describe what you see.

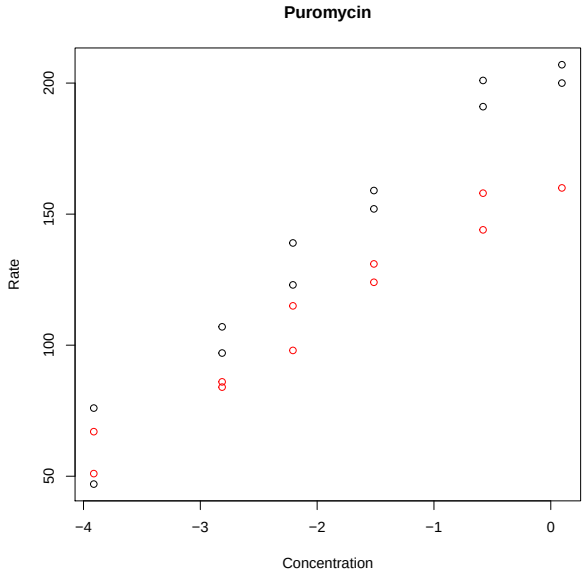
```
plot(log(Puromycin$conc), Puromycin$rate, col = Puromycin$state)
```



It appears that the the treated group has higher rates than the untreated group, on average. (Note that default colors are black for the first level and red for the second level).

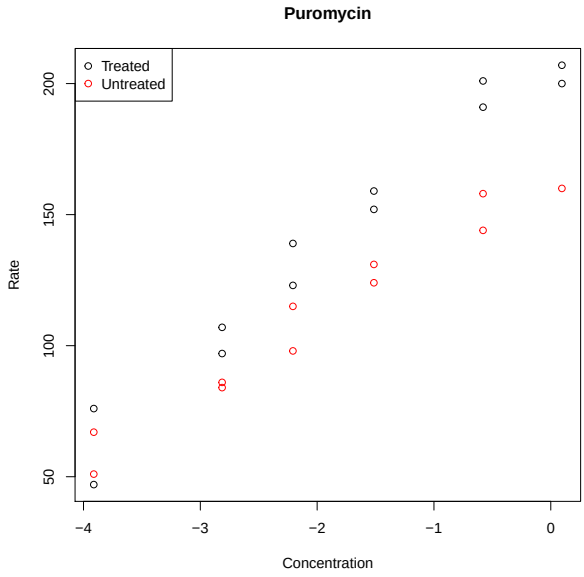
1.3.5 Make a scatterplot of the rate versus the log of the concentration, color the points by treatment group (`state`), label the x-axis "Concentration" and the y-axis "Rate", and label the plot "Puromycin".

```
plot(log(Puromycin$conc), Puromycin$rate, col = Puromycin$state,
     xlab = "Concentration", ylab = "Rate", main = "Puromycin")
```



1.3.6 Add a legend to the above plot indicating what the points represent.

```
plot(log(Puromycin$conc), Puromycin$rate, col = Puromycin$state,
     xlab = "Concentration", ylab = "Rate", main = "Puromycin")
legend("topleft",c("Treated", "Untreated"), col = 1:2, pch = 1)
```



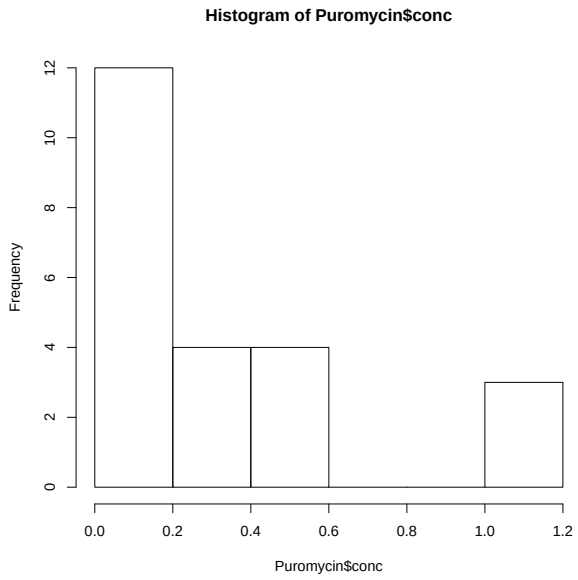
1.3.7 Make a boxplot of the treated versus untreated rates. Using the function `pdf`, save the image to a file with a width and height of 7 inches.

```
pdf("puromycin.pdf",width = 7, height = 7)
boxplot(Puromycin$rate~Puromycin$state)
dev.off()

## pdf
##   2
```

1.3.8 Make a histogram of the frequency of concentrations. What is the width of the bins?

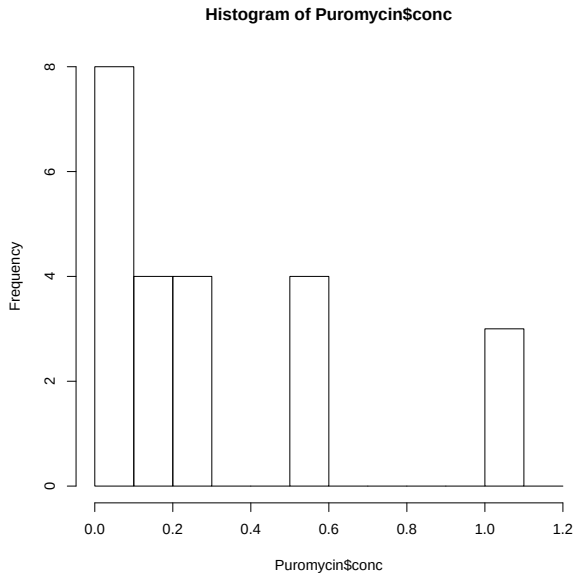
```
hist(Puromycin$conc)
```



The bin width is 0.20.

1.3.9 Make a histogram of the frequency of concentrations with a bin width of 0.10. How is this different from the histogram above?

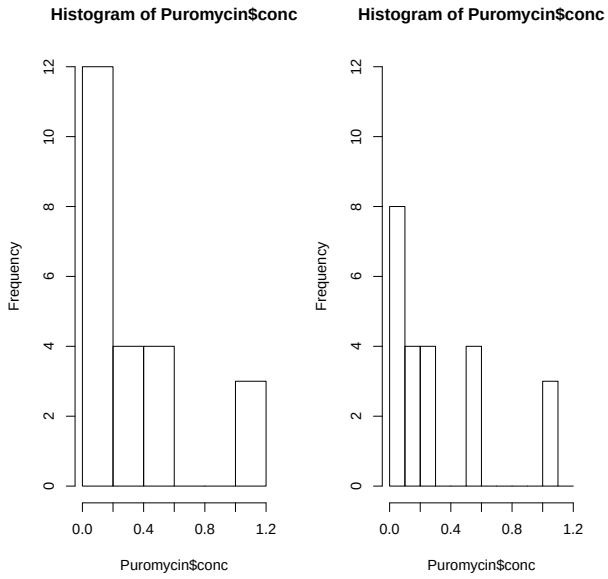
```
hist(Puromycin$conc, breaks = seq(0, 1.2, .10))
```



The bins are narrower, so we see in finer detail the distribution of the concentrations.

1.3.10 Plot the two histograms (bin widths 0.20 and 0.10) side by side in the same graphic window and make sure they have the same range on the y-axis. Does this make it easier to answer the question of how the two histograms differ?

```
par(mfrow = c(1, 2))
hist(Puromycin$conc, ylim = c(0, 12))
hist(Puromycin$conc, breaks = seq(0, 1.2, .10), ylim = c(0, 12))
```



In some situations it may be of use to view plots simultaneously. In this case, on the right we see clearly that more values are between 0 and 0.10 than 0.10 and 0.20 whereas the plot on the left does not display this information. In the histogram on the right we see that no concentrations fall between 0.30 and 0.50, whereas this is not apparent in the histogram on the left.

If you want to save your work: save your `R` session and/or source code!