

Box Plots and Scatter Plots

1 Load the Hot Dog Data

Fill in the blanks below to read the file `hotdog.csv` into R and reorder the levels of the `Day` factor. You should have this file from the previous workshop. Make sure the file is in your working directory.

```
myFile <- "hotdog.csv"
hd <- read.csv(file = _____,
               skip = _____,
               nrows = _____,
               header = _____)

hd$Day <- factor(hd$Day,
                levels = c(____, ____, ____, ____, ____))
```

Check that the file was read in correctly and that the levels of the `Day` factor are reordered. The structure of your data should look like this:

```
str(hd)

## 'data.frame': 54 obs. of 5 variables:
## $ Day      : Factor w/ 5 levels "Mon","Tue","Wed",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Brand    : Factor w/ 2 levels "Nathans","OscarMayer": 1 1 1 1 1 2 2 2 2 2 ...
## $ Type     : Factor w/ 3 levels "Beef","Meat",...: 2 3 3 3 3 1 1 1 2 3 ...
## $ Calories: int  175 129 102 135 142 190 153 132 173 152 ...
## $ Sodium  : int  507 430 542 426 513 587 401 253 458 588 ...
```

If you can't get this structure, ask me for help.

1.1 Questions

1. What do the arguments `'skip='`, `'nrows='` and `'header='` do in the `read.csv()` function?
2. Why do we reorder the levels of the `Day` column?
3. **True or False:** We reorder the levels of the `Day` column by overwriting the original column?

2 Box Plots

Fill in the blank arguments to get the box plot on the next page. Run your code to see if you get the same plot.

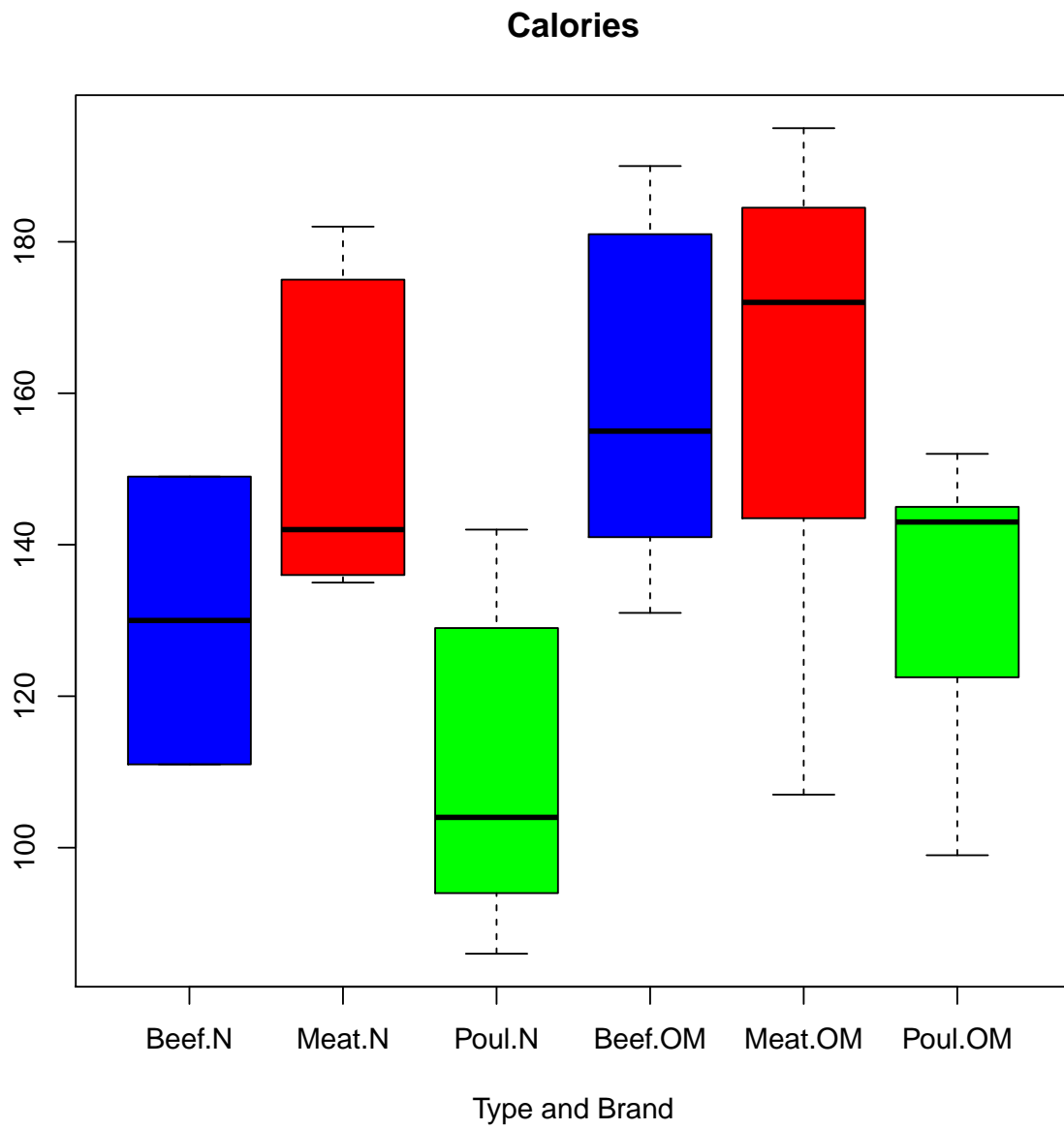
```
boxplot(Calories ~ _____ * _____,
        data = _____,
        main = _____,
        xlab = _____,
        col = c(_____, _____, _____, _____, _____, _____),
        xaxt = "n")

x.ticks <- c(_____, _____, _____, _____, _____, _____)

axis(side = 1,
     at = 1:__,
     labels = _____)
```

2.1 Questions and Challenges

1. What does the argument 'axt = "n"' do in the `boxplot()` function?
2. What do the arguments 'side = 1' and 'at = 1:__' do in the `axis()` function?
3. Create a box plot of Sodium by Type. Include the following:
 - A main title
 - Custom x and y axes titles
 - Make the box plots horizontal
 - Different colors for each hot dog type



3 Scatter Plots

Fill in the blank arguments to get the scatter plot on the next page. Run your code to see if you get the same plot.

```
plot(_____ ~ _____,
     data = _____,
     pch = c(15, 16)[_____],
     col = c(_____, _____, _____)[hd$Type],
     main = _____)

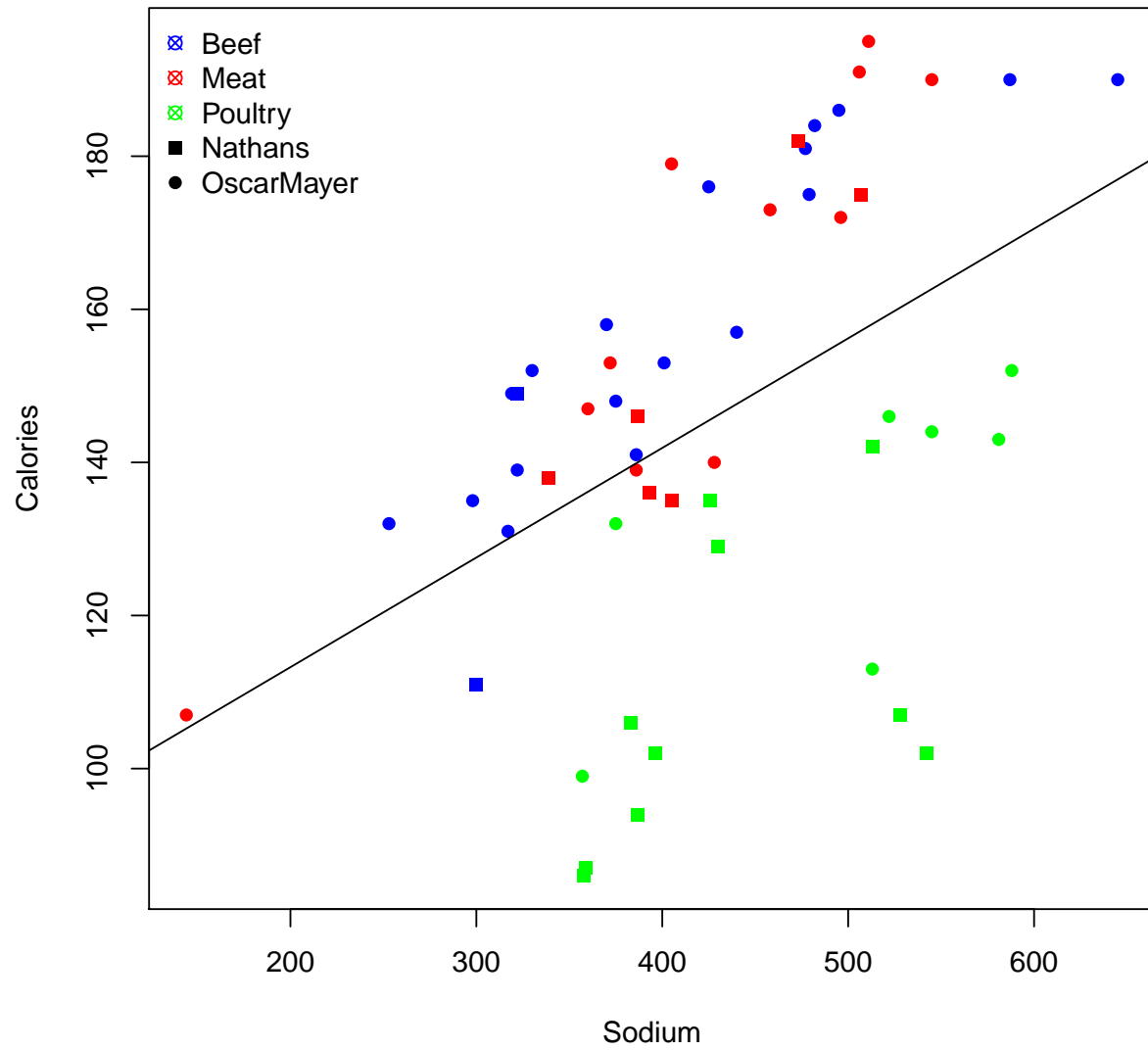
abline(lm(Calories ~ Sodium, data = hd))

legend(x = "top____",
       legend = c(levels(hd$_____), levels(hd$_____)),
       pch = c(rep(13, 3), 15, 16),
       col = c(_____, _____, _____, _____, _____),
       bty = _____)
```

3.1 Questions and Challenges

1. What does the argument 'pch=' do in the `plot()` function?
2. What does the command 'c(15, 16)[_____]' do?
3. Instead of using 'x = "top____"' in the `legend()` function what can you use?
4. What does the argument 'bty=' do in the `legend()` function?
5. Create a scatter plot of **Sodium** and **Type**. Include the following:
 - **Sodium** on the y-axis, **Calories** on the x-axis
 - A main title
 - Different colors for each brand
 - Different shapes for each day (hint: `pch`)
 - **Bonus:** Add a line of best fit and a legend

Calories ~ Sodium



R Session Information

R version 3.0.2 (2013-09-25)

Platform: x86_64-apple-darwin10.8.0 (64-bit)

locale:

[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:

[1] stats graphics grDevices utils datasets methods base

other attached packages:

[1] knitr_1.5

loaded via a namespace (and not attached):

[1] digest_0.6.4 evaluate_0.5.1 formatR_0.10 highr_0.3

[5] stringr_0.6.2 tools_3.0.2