Drosophilidae wings — Bolstad et al. 2015

Boldstad et al. (2015) studied the allometric relationships between the length of the wings of Drosophilidae and the length of the L2 vein that runs across the wing. They measured more than 20,000 individuals, belonging to 111 species. In their Figure 1, they show regressions between the log length of the wing size and the log length of the L2 vein. They produce a regression for each species and sex. They then added points showing the average values for each species. The file `data/Bolstad2015_figure1.pdf` contains a simplified version of Figure 1 of the original paper. Reproduce the figure. The data is stored in `ISC/ggplot2/data/Boldstad2015_data.csv`. The logarithm of wing size and L2 length is already taken.

Always start by importing libraries and the data:

```r
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##     filter, lag
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
# read the data
b2015 <- read.csv("../data/Bolstad2015_data.csv")
# explore dimensions of the data
dim(b2015)
```

```
## [1] 20345     6
```

```r
# explore structure of the data
head(b2015)
```

```
##      Species        ID      Date Sex  WingSize  L2Length
## 1 D_acutila ACU1006.TIF 24_Jul_01   F 0.1311220 0.4972620
## 2 D_acutila ACU1009.TIF 24_Jul_01   F 0.1360881 0.4879716
## 3 D_acutila ACU1010.TIF 24_Jul_01   F 0.1953933 0.5401366
## 4 D_acutila ACU1013.TIF 24_Jul_01   F 0.2773285 0.6463595
## 5 D_acutila ACU1018.TIF 24_Jul_01   F 0.1515312 0.4977579
## 6 D_acutila ACU1021.TIF 24_Jul_01   F 0.1751865 0.4919342
```

```r
# number of species
dim(distinct(select(b2015, Species)))
```

```
## [1] 111   1
```

We want to produce a regression per `Species`, and facet by `Sex`. We could use the `aes` colour, but that would produce many different colors. Alternatively, we can define the `group` directly:

```
# draw a regression line for each species and sex
pl <- ggplot(data = b2015, aes(x = WingSize, y = L2Length, group = Species)) +
  geom_smooth(alpha = 0.1, method = "lm", se = FALSE) + facet_grid(.~Sex) + theme_bw()
pl
```
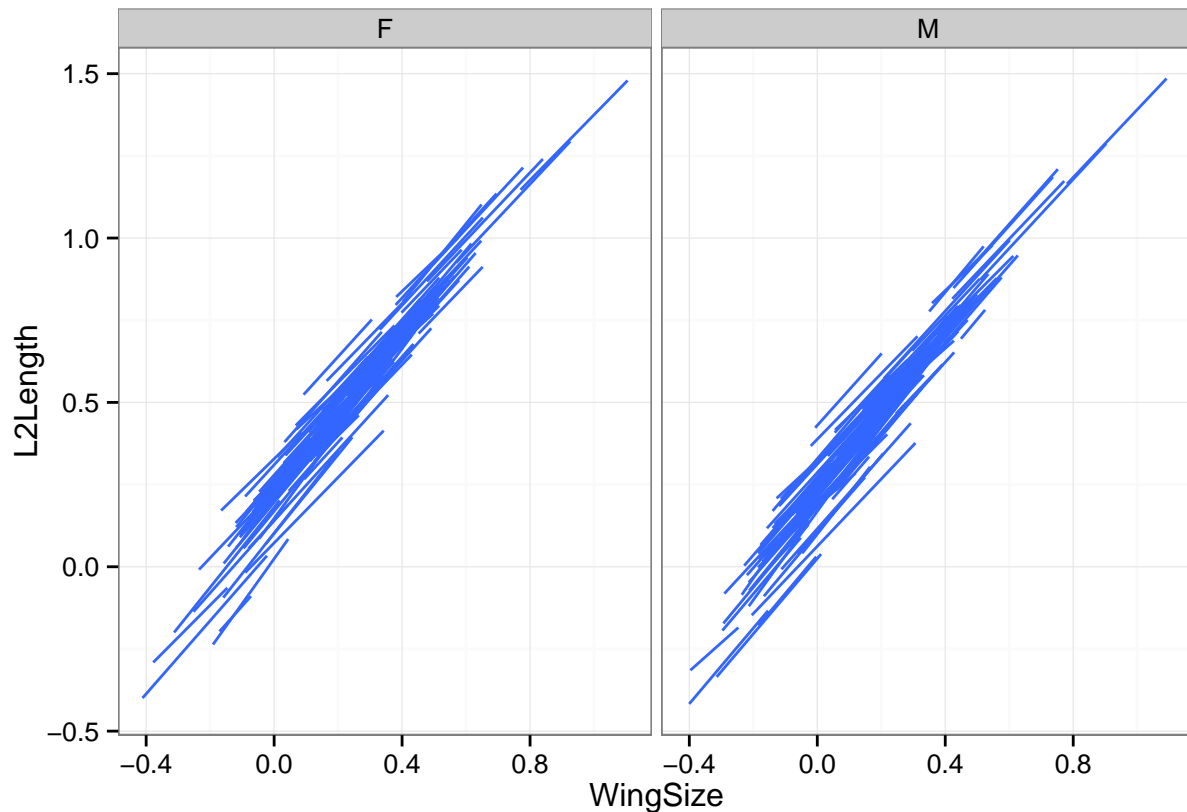

Now we want to add another layer, with a point for each `Species`/`Sex` combination. We create a new data frame using the function `summarise` of `dplyr`:

```
# compile averages by species and sex
by_species_sex <- summarise(group_by(b2015, Species, Sex),
                            WingSize = mean(WingSize),
                            L2Length = mean(L2Length))
# see the result
by_species_sex
```

```
## Source: local data frame [220 x 4]
## Groups: Species [?]
##
##         Species    Sex   WingSize   L2Length
##          (fctr) (fctr)      (dbl)      (dbl)
## 1   C_procnemis      M 0.07531273 0.1904114
## 2     D_acutila      F 0.19390128 0.5319778
## 3     D_acutila      M 0.07009950 0.4083030
## 4     D_affinis      F 0.19271682 0.4567225
```

```
## 5      D_affinis     M 0.07317315 0.3128763
## 6      D_albostr     F 0.30348696 0.6383540
## 7      D_albostr     M 0.18746097 0.5073995
## 8      D_algonqu     F 0.34727439 0.6329803
## 9      D_algonqu     M 0.21745274 0.4923027
## 10     D_america     F 0.47858859 0.7935839
## ..         ...      ...       ...       ...
```

```r
# add points to current plot
pl <- pl + geom_point(data = by_species_sex)
# add labels for x- and y-axis
pl <- pl + xlab("Wing size (ln mm)") + ylab("Vein L2 length (ln mm)")
# plot the result
pl
```
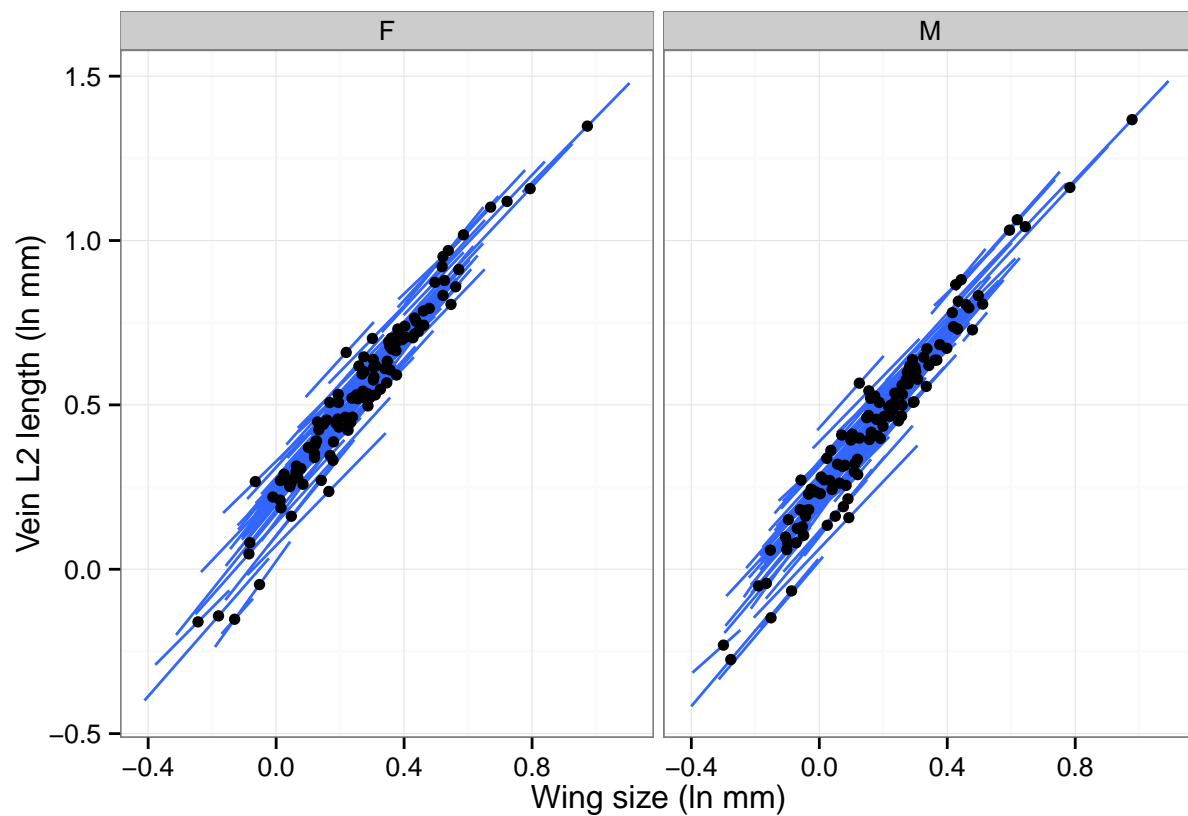


```r
# save the graph
ggsave(pl, filename = "../data/Bolstad2015_figure1.pdf", width = 9, height = 4)
```