Solution to Series 1

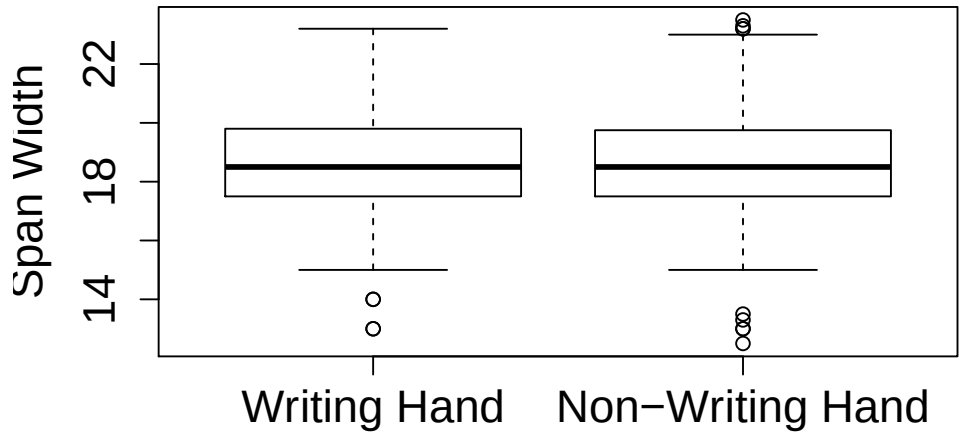
1. **a)** We can get a first overview of the data by looking at the summary statistics:
R-Output: (excerpt)

```
> library(MASS)
> summary(survey)
```

```
     Wr.Hnd           NW.Hnd
 "Min.   :13.00  " "Min.   :12.50  "
 "1st Qu.:17.50  " "1st Qu.:17.50  "
 "Median :18.50  " "Median :18.50  "
 "Mean   :18.67  " "Mean   :18.58  "
 "3rd Qu.:19.80  " "3rd Qu.:19.73  "
 "Max.   :23.20  " "Max.   :23.50  "
 "NA's   :1  "     "NA's   :1  "
```

The summary statistics are very similar and don't give us any evidence for differences between the two hands. This becomes even more clear if we look at the boxplots:

```
> boxplot(survey$Wr.Hnd, survey$NW.Hnd,
          ylab="Span Width", names=c("Writing Hand", "Non-Writing Hand"),
        cex.lab = 1.5, cex.axis = 1.5)
```



We conclude that, just using descriptive statistics, we cannot make out any difference between the span widths of the writing and the non-writing hand.

b) We can e.g. use one of these commands:

```
> survey[rev(order(survey$Age))[1:2],]
```

```
       Sex Wr.Hnd NW.Hnd W.Hnd   Fold Pulse  Clap Exer
171 Female   16.5   17.0 Right L on R    NA Right Some
154   Male   21.5   21.6 Right R on L    69 Right Freq
    Smoke Height      M.I    Age
171 Never 168.00   Metric 73.000
154 Never 172.72 Imperial 70.417
```

```
> survey[rev(order(survey$Age))[1:2],"Smoke"]
```

```
[1] Never Never
Levels: Heavy Never Occas Regul
```

So the two oldest students do not smoke.

c) We can get pairwise plots of all the variables with the command `pairs(survey)` (result not shown for space reasons). From this we see that e.g. the gender of the student, which seems plausible. Surprisingly, the plots don't show a strong dependence between exercise level and pulse. Also, there seem to be differences between left and right handed people. Of course this is not a thorough statistical analysis, but rather a good way to get a first impression of the data.

d) The first line finds all people under 30 and plots their pulse against age. The second line then fits a linear regression model of this restricted data set and adds the regression line to the plot.