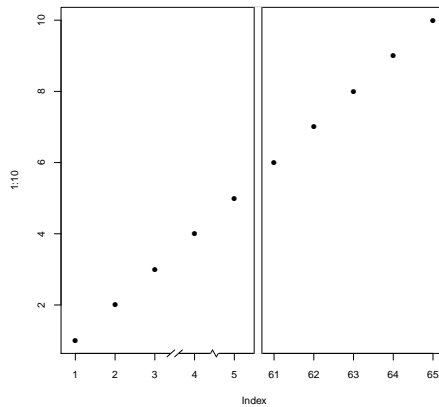


Exercises for the course
“An introduction to R”
 Sheet 07

Exercise 35: *How to produce gaps on the axis*

`axis.break()` and `gap.plot()` are commands in the library `plotrix`. If you cannot load this library, then install it first.

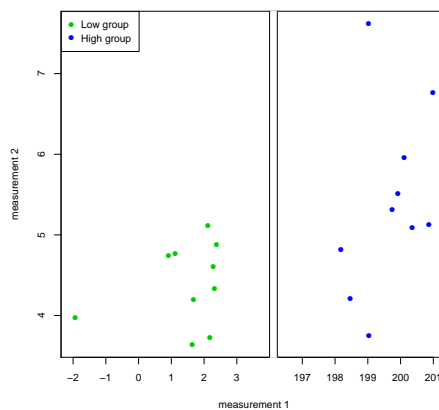
- Use `axis.break` to produce a plot similar to the following plot. (The data vector is 1:10).



- Execute the commands

```
set.seed(1111)
twogrp.x <- c( rnorm(10)+1, rnorm(10)+200 )
twogrp.y <- c( rnorm(10, sd=0.5)+4, rnorm(10)+5 )
```

Plot `twogrp.y` against `twogrp.x` to see how it looks like without gap. Then use `gap.plot()` to produce a plot similar to the following plot.



Use the color 'green' for the points to the left of the axis break and 'blue' for the points to the right of the axis break.

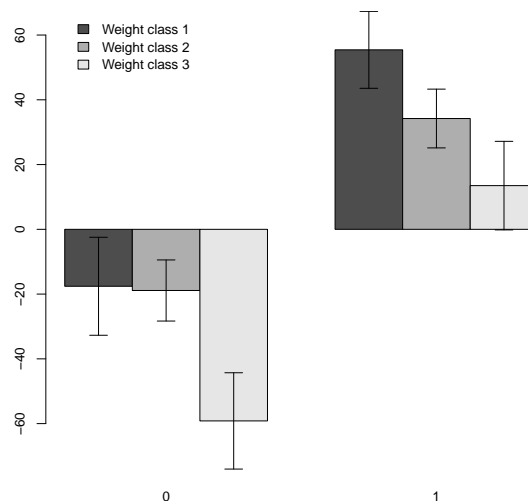
Exercise 36: *Supporting a hypothesis*

Apply a suitable test:

- The built-in data set `sleep` shows the effect of two soporific drugs (increase in hours of sleep compared to the control group) on 10 patients. Are the increases in sleeping hours of the two groups significantly different?
- Recall `heartbeats` from Exercise 12. Is the increase in weight of the heartbeat group compared to the control group significant? Then answer this question for every weight class separately.
- The common red spider (Bohnen-spinnmilbe) is a vermin of agricultural plants to which plants react with the production of toxic substances. Do plants “remember” a former attack? Two groups of each 20 cotton plants are being infected with the mite. One group has never encountered this mite before. The second group has already survived a former attack. After a certain time, the mites on the plants are counted. You find the result in the file `mite.txt` on the web page. Is there a significant difference between the two groups? Compare the two groups visually (produce an appropriate plot) and apply a suitable test. Formulate an answer!

Exercise 37: *How to produce barplots with error bars*

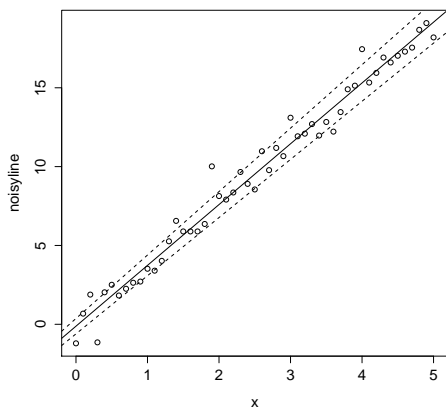
Install and load the library `sciplot`. Look at the help page `?bargraph.CI` and the example with `ToothGrowth` at the end of the help page. Then load the data `heartbeats.txt` from the web page. Use `bargraph.CI` from the library `sciplot` to barplot `weightincr` for every `treatment` and grouped according to `wghtcls`. Your result should resemble the following figure.



Then make a complete figure out this (what information is missing so far?). Hints: Change the location of the legend with `x.legend=1` and change the labels of the legend.

Exercise 38: Set the seed to 1234 to get the same picture. Define two vectors `x <- seq(from=0,to=5,by=0.1)` and `noisyline <- 4*x + rnorm(length(x))`. Explain the variable `noisyline` with a linear model of the independent variable `x`. Use the command `lm()` for this. Denote the returned object as `regr`. What is the Anova table of this regression? Read off the

intercept and the slope of the fitted line with `coef()`. Extract the p-values of the intercept and of the slope from `summary(regr)`. What is the fraction of the total variation in `noisyline` that is explained by the regression? This fraction is called 'r squared' and is printed by `summary(regr)` under 'Multiple R-squared'. Check that this value is equal to $(\text{cor}(x, \text{noisyline}))^2$. In order to visualize the linear model, plot `noisyline` against `x`. Add the regression line with `abline(regr)`. Next calculate confidence intervals of the fitted parameters with `confint()` applied to `regr`. Denote the object returned by this command as `cf`. Enter `cf` to view it. The two columns of `cf` define two lines. Add these two lines to the plot with line type "dashed". Your result should resemble the following figure.



Exercise 39: Recall the data set `cars` from Exercise 21. Plot `speed` as a function of `dist`. The plot suggests that `speed` depends linearly on `dist`. Find this linear dependence with a linear regression in which `dist` is the response variable. Store the returned regression object into the variable `regr`. What is the Anova table of this regression? Read off the p-values of the intercept and of the slope from `summary(regr)`. What is the fraction of the total variation in `dist` that is explained by the regression, that is, what is r-squared? Then add the regression line to the plot. Extract intercept and slope of the regression line from `regr` and round both values to 3 significant digits. These values are used for the legend of the following figure. In addition add the two lines which you get from the confidence intervals for the intercept and for the slope of the regression line.

