ATBC BIOINFORMATICA

STATISTIEK IN R_

Week 4

HAN_UNIVERSITY
OF APPLIED SCIENCES

ONDERWERPEN

- Boxplots
- Toevoegen custom legenda

?BOXPLOT

Description

Produce box-and-whisker plot(s) of the given (grouped) values.

Usage

```
boxplot(x, ...)

## S3 method for class 'formula'
boxplot(formula, data = NULL, ..., subset, na.action = NULL,
        drop = FALSE, sep = ".", lex.order = FALSE)

## Default S3 method:
boxplot(x, ..., range = 1.5, width = NULL, varwidth = FALSE,
        notch = FALSE, outline = TRUE, names, plot = TRUE,
        border = par("fg"), col = NULL, log = "",
        pars = list(boxwex = 0.8, staplewex = 0.5, outwex = 0.5),
        horizontal = FALSE, add = FALSE, at = NULL
```

?BOXPLOT

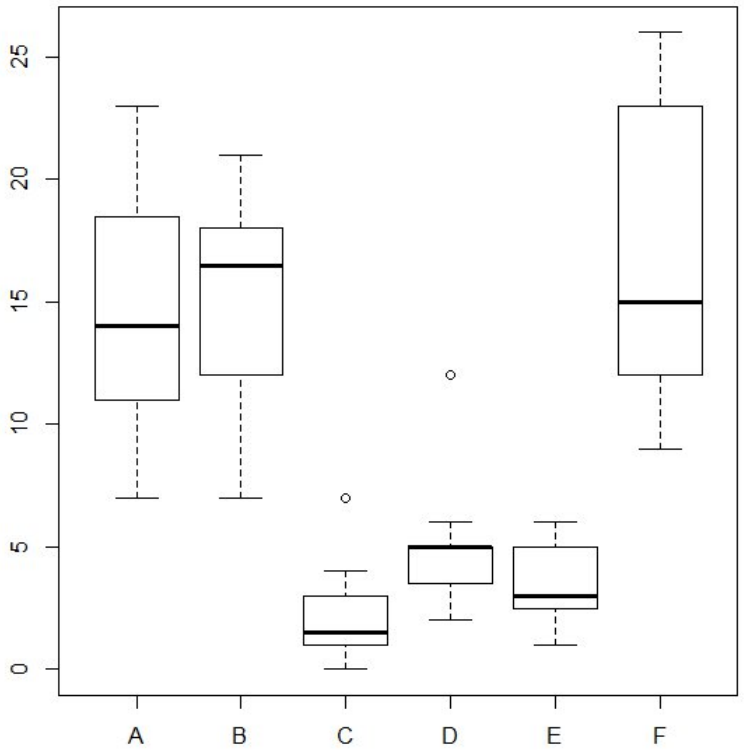
formula	a formula, such as y ~ grp, where y is a numeric vector of data values to be split into groups according to the grouping variable grp(usually a factor).
data	a data.frame (or list) from which the variables in formula should be taken.
subset	an optional vector specifying a subset of observations to be used for plotting.
na.action	a function which indicates what should happen when the data contain NAs. The default is to ignore missing values in either the response or the group.
drop, sep, lex.order	passed to split.default, see there.
x	for specifying data from which the boxplots are to be produced. Either a numeric vector, or a single list containing such vectors. Additional unnamed arguments specify further data as separate vectors (each corresponding to a component boxplot). NAs are allowed in the data.
...	For the formula method, named arguments to be passed to the default method. For the default method, unnamed arguments are additional data vectors (unless x is a list when they are ignored), and named arguments are arguments and graphical parameters to be passed to bxp in addition to the ones given by argument pars (and override those in pars). Note that bxp may or may not make use of graphical parameters it is passed: see its documentation.
range	this determines how far the plot whiskers extend out from the box. If range is positive, the whiskers extend to the most extreme data point which is no more than range times the interquartile range from the box. A value of zero causes the whiskers to extend to the data extremes.
width	a vector giving the relative widths of the boxes making up the plot.
varwidth	if varwidth is TRUE, the boxes are drawn with widths proportional to the square-roots of the number of observations in the groups.
notch	if notch is TRUE, a notch is drawn in each side of the boxes. If the notches of two plots do not overlap this is ‘strong evidence’ that the two medians differ (Chambers *et al*, 1983, p. 62). See boxplot.stats for the calculations used.
outline	if outline is not true, the outliers are not drawn (as points whereas S+ uses lines).

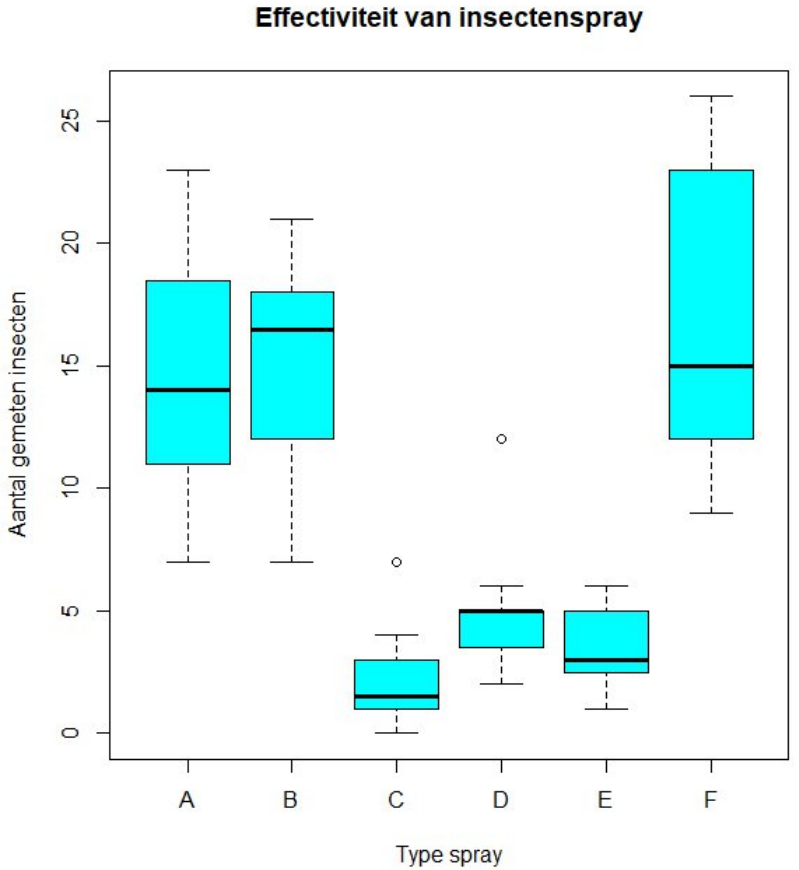
?BOXPLOT

names	group labels which will be printed under each boxplot. Can be a character vector or an expression (see plotmath).
boxwex	a scale factor to be applied to all boxes. When there are only a few groups, the appearance of the plot can be improved by making the boxes narrower.
staplewex	staple line width expansion, proportional to box width.
outwex	outlier line width expansion, proportional to box width.
plot	if TRUE (the default) then a boxplot is produced. If not, the summaries which the boxplots are based on are returned.
border	an optional vector of colors for the outlines of the boxplots. The values in border are recycled if the length of border is less than the number of plots.
col	if col is non-null it is assumed to contain colors to be used to colour the bodies of the box plots. By default they are in the background colour.
log	character indicating if x or y or both coordinates should be plotted in log scale.
pars	a list of (potentially many) more graphical parameters, e.g., boxwex or outpch; these are passed to bxp (if plot is true); for details, see there.
horizontal	logical indicating if the boxplots should be horizontal; default FALSE means vertical boxes.
add	logical, if true *add* boxplot to current plot.
at	numeric vector giving the locations where the boxplots should be drawn, particularly when add = TRUE; defaults to 1:n where n is the number of boxes.

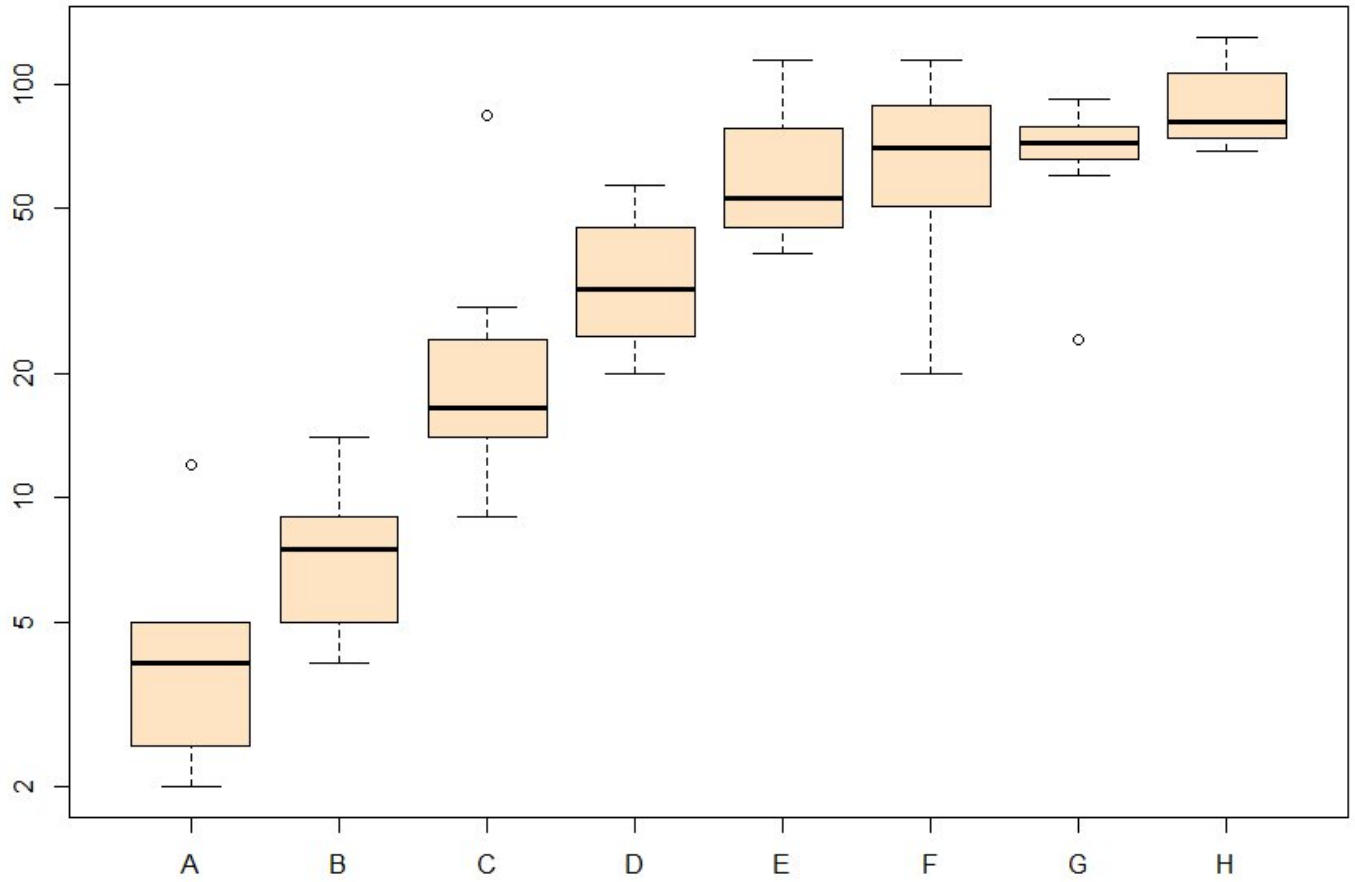
BOXPLOT(COUNT ~ SPRAY, DATA = INSECTSPRAYS)

Maak er deze grafiek van!





boxplot(decrease ~ treatment, data = orchardsprays, log = "y", col = "bisque")



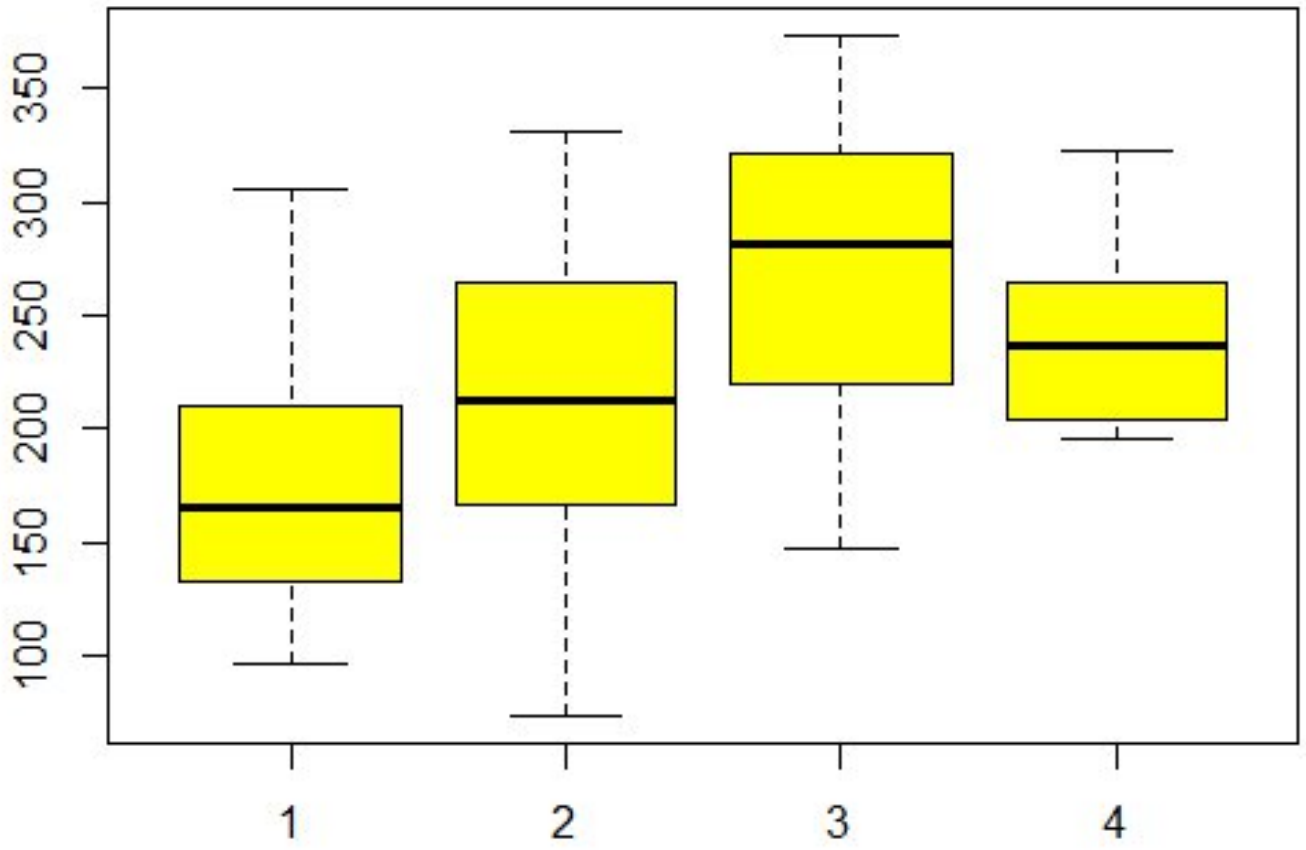
Logaritmische schaal gemakkelijk aangezet

CHICKWEIGHT

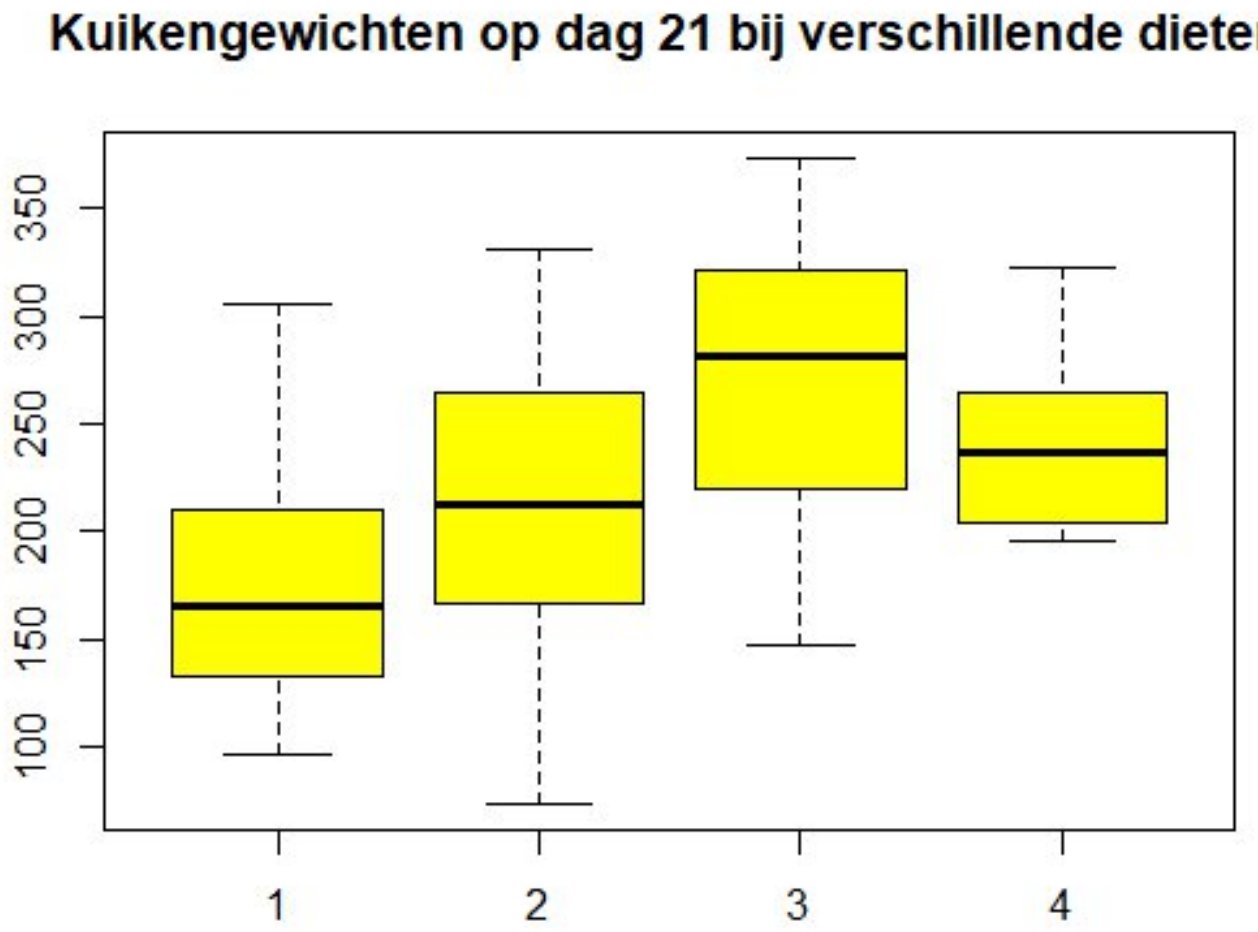
Haal de dataset ChickWeight op uit het package datasets.

Gebruik R om in een overzicht boxplots te maken van het gewicht van kuikens op dag 21 bij de verschillende diëten.

boxplot(weight~diet, data=ChickWeight, subset=time==21, col="yellow")

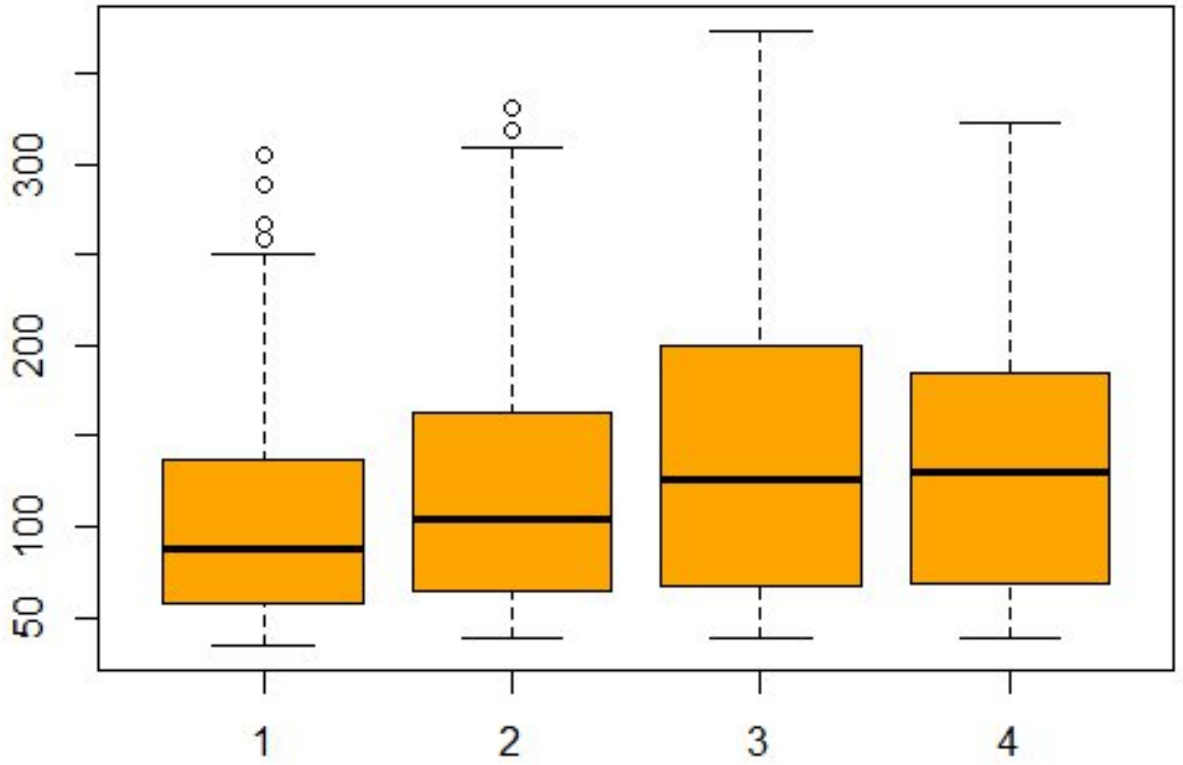


title("kuikengewichten op dag 21 bij verschillende dieten")



(ZONDER) SUBSET

- boxplot(weight~Diet, data=ChickWeight, col="orange")

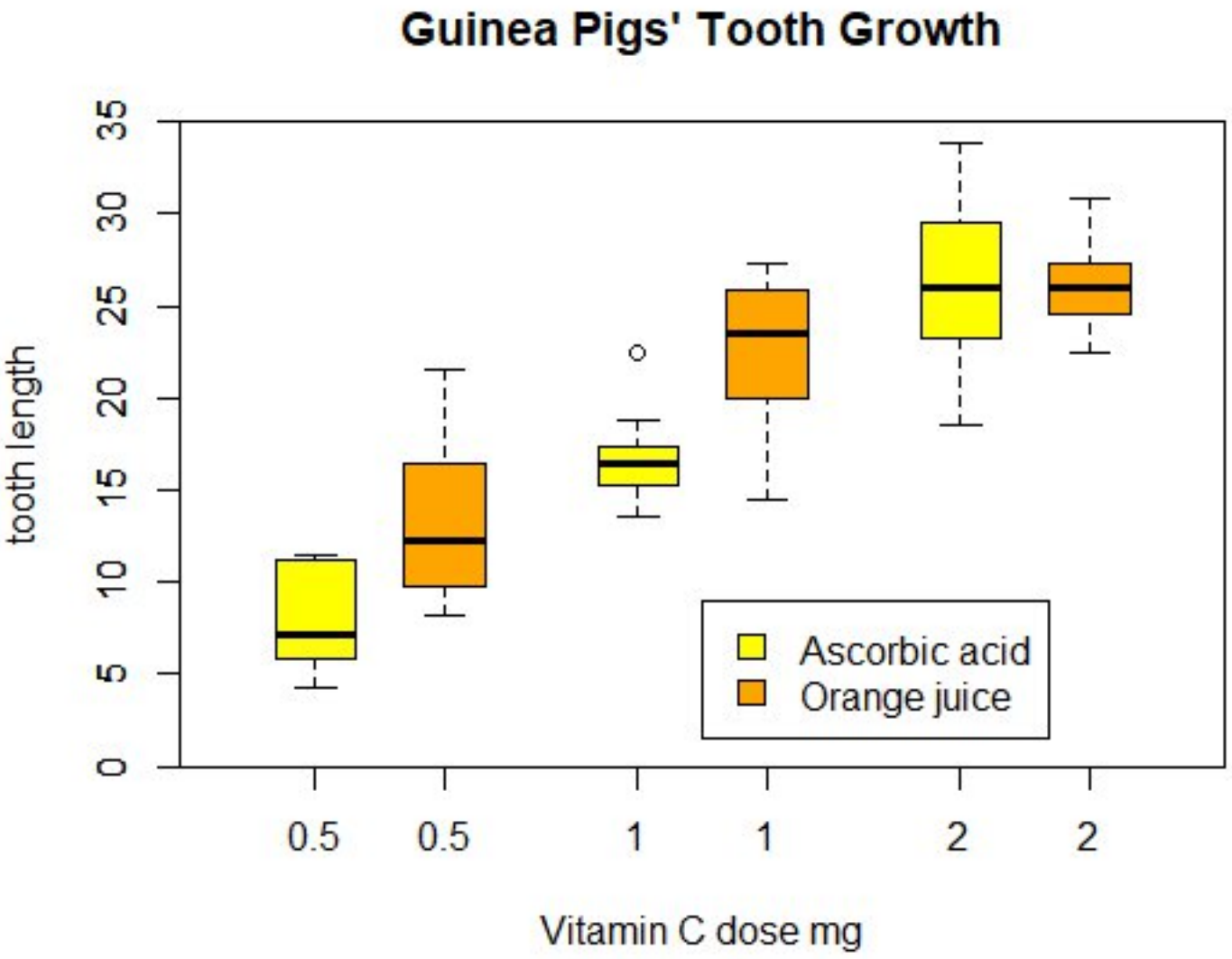


Zonder subset krijg je alle gewichten van alle dagen op 1 hoop gegooid

TOOTHGROWTH (CODE UIT ?BOXPLOT)

```
boxplot(len ~ dose, data = ToothGrowth,
        boxwex = 0.25, at = 1:3 - 0.2,
        subset = supp == "VC", col = "yellow",
        main = "Guinea Pigs' Tooth Growth",
        xlab = "Vitamin C dose mg",
        ylab = "tooth length",
        xlim = c(0.5, 3.5), ylim = c(0, 35), yaxs = "i")
boxplot(len ~ dose, data = ToothGrowth, add = TRUE,
        boxwex = 0.25, at = 1:3 + 0.2,
        subset = supp == "OJ", col = "orange")
legend(2, 9, c("Ascorbic acid", "Orange juice"),
       fill = c("yellow", "orange"))
```

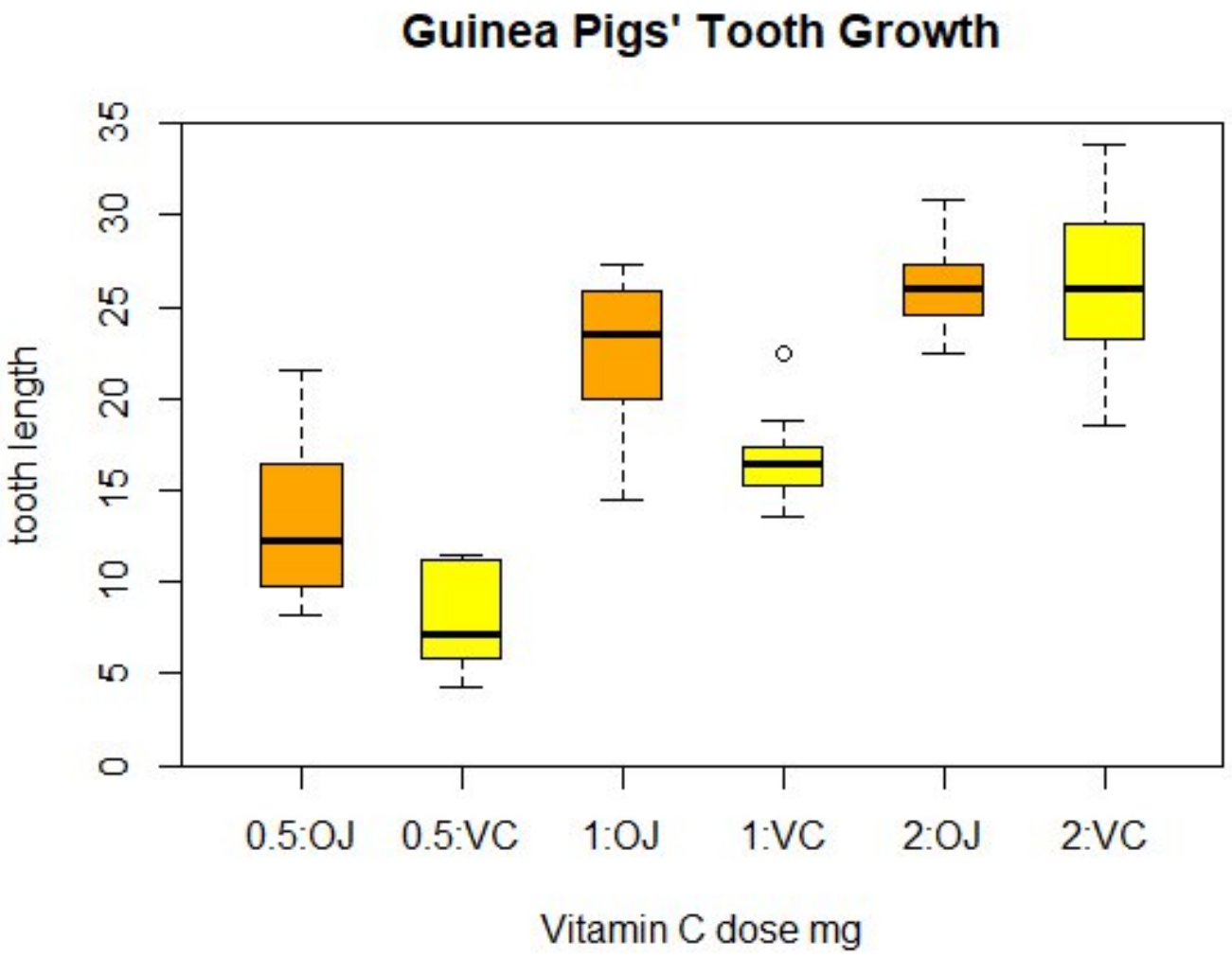
TOOTHGROWTH



TOOTHGROWTH EASY

```
boxplot(len ~ dose:supp, data = ToothGrowth,
        boxwex = 0.5, col = c("orange", "yellow"),
        main = "Guinea Pigs' Tooth Growth",
        xlab = "Vitamin C dose mg", ylab = "tooth length",
        sep = ":", lex.order = TRUE, ylim = c(0, 35), yaxs = "i")
```

TOOTHGROWTH EASY



TOEVOEGEN CUSTOM LEGENDA

- Afgelopen week : legend.text=TRUE
- Maar wat als je meer opties wilt?

- legend()

?LEGEND

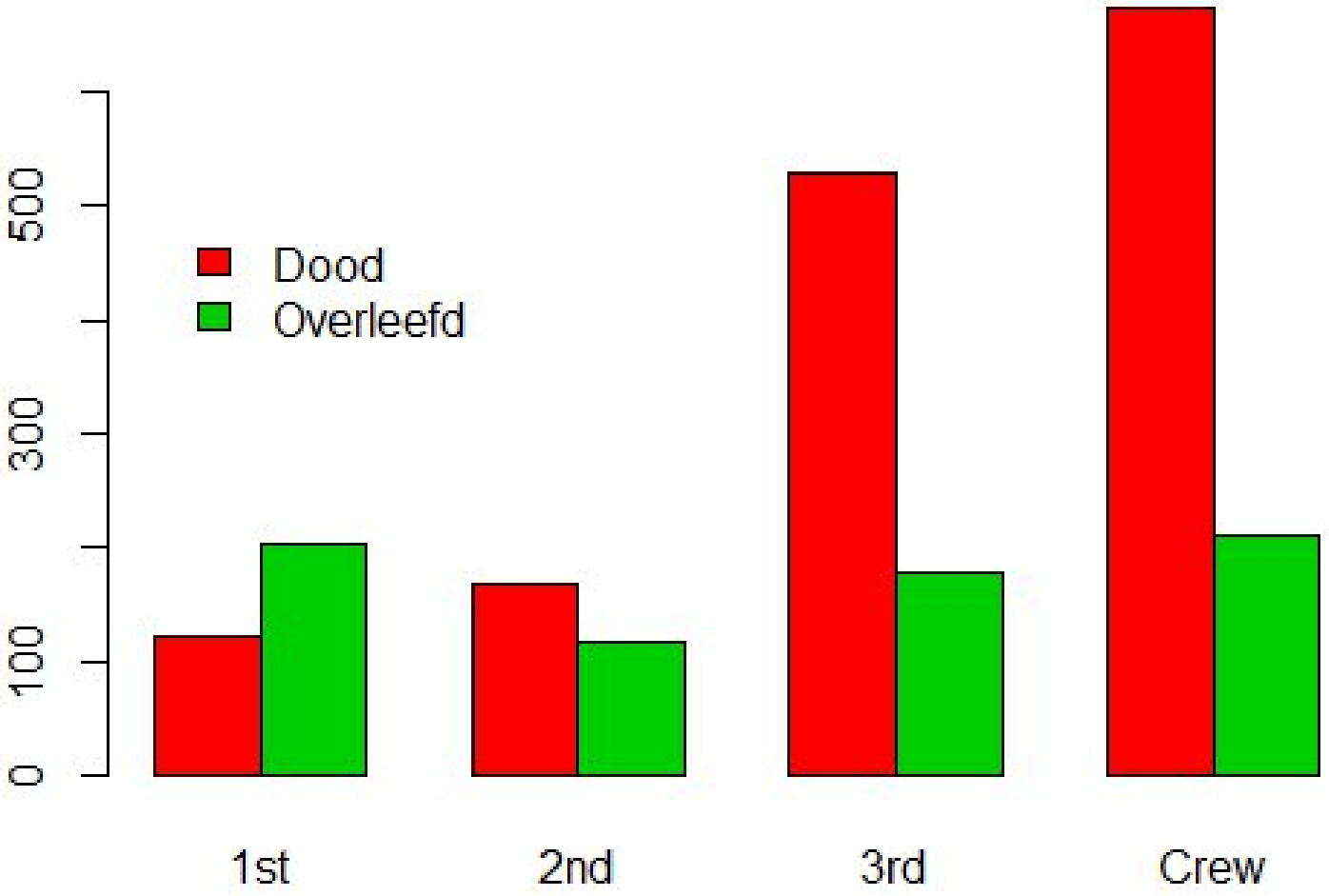
```
legend(x, y = NULL, legend, fill = NULL, col = par("col"),
       border = "black", lty, lwd, pch,
       angle = 45, density = NULL, bty = "o", bg = par("bg"),
       box.lwd = par("lwd"), box.lty = par("lty"), box.col = par("fg"),
       pt.bg = NA, cex = 1, pt.cex = cex, pt.lwd = lwd,
       xjust = 0, yjust = 1, x.intersp = 1, y.intersp = 1,
       adj = c(0, 0.5), text.width = NULL, text.col = par("col"),
       text.font = NULL, merge = do.lines && has.pch, trace = FALSE,
       plot = TRUE, ncol = 1, horiz = FALSE, title = NULL,
       inset = 0, xpd, title.col = text.col, title.adj = 0.5,
       seg.len = 2)
```

TITANIC LEGENDA

- barplot(margin.table(Titanic,c(4,1)),legend.text = TRUE, beside = TRUE, col = c(2,3))

- barplot(margin.table(Titanic,c(4,1)), beside = TRUE, col = c(2,3))
- legend(1,500, fill=c(2,3), legend=c("Dood", "Overleefd"), bty="n")

CUSTOM LEGENDA



OPDRACHT

- Maak verschillende boxplots
- Gebruik de dataset airquality
  - Ozon
  - Temperatuur vs maand
  - Ozon vs maand
  - Temperatuur en Ozon vs maand in 1 boxplot met 2 verschillende kleuren, voeg een legenda toe
- Dataset mtcars
  - Gewicht vs aantal cilinders, in een volledige grafiek
  - Brandstofverbruik vs aantal cilinders

OPDRACHT

- Dataset chickwts
- Boxplot met de gewichten per voedingssupplement.
- Voeg een lijn toe ter hoogte van de mediaan van het gewicht van alle voedingssupplementen.

VOLGENDE WEEK

- Statistische toetsen
- Neem je statistiek boek mee!

VERANTWOORDING

- In deze uitgave is géén auteursrechtelijk beschermd werk opgenomen
- Alle teksten © Gonny Henkes tenzij expliciet externe bronnen zijn aangegeven
- Screenshots op basis van vernoemde websites
- Eventuele images zijn opgenomen met vermelding van bron