

Programming and Analyzing in R: Data, Models, and Plots

Wouter van Atteveldt <http://vanatteveldt.com>

May 2016

Combining data

- Adding rows: `rbind`
- Adding columns: `cbind`

Merging data

- Merge on common key column(s)
 - `merge(a, b)`
 - `merge(a, b, by="year")`
 - `merge(a, b, by="year", all.x=T)`
 - `merge(a, b, by.x="year", by.y="jaar")`

Matching data

- Match returns the position of elements in another list
- `match(a$year, b$year)`
- `a$fr = b$fr[match(a$year, b$year)]`

Reshape and aggregate

- Package reshape2:
- `melt(data, id.vars=)`
 - Variables to cases
- `dcast(data, rows~cols, value.var=)`
 - Cases to variables / tabulation
 - `fun.aggregate=mean`
- `aggregate(data, by=)`
 - Aggregating

Simple two variable tests

- `t.test(x, y, paired=.)`
- `cor.test(x, y)`
- `chisq.test(x,y)`

Linear models (and anova/ancova)

- `lm(y ~ x1 + x2)`
- `lm(Y ~ x1*x2)`
- `lm(Y ~ x1 + x2 + x1:x2)`
- `lm(Y ~ x1 + cat)`
- `lm(Y ~ x1 + cat - 1)`
- `aov(y ~ X)`

Modeling results

- `m = lm(...)`
- `summary(m)`
- `anova(m)`
- `coef(m)`
- `fitted(m)`
- `residuals(m)`
- `plot(m)`
- `anova(m1, m2)`
- <http://www.statmethods.net/stats/rdiagnostics.html>

Hands-on 2b

- 4_transforming.html
- 5_modeling.html

for loops, functions, lags

- `for (var in values) {commands}`
 - `for (i in 1:10) {print(i)}`
- `f = function(x) {commands}`
 - `f = function(x) {length(unique(x))}`
- `d$lagx = c(NA, d$x[-ncol(d)])`
- `d$leadx = c(d$x[-1], NA)`

Plotting

- `plot(x, y, type=., ...)`
- `lines(x, y, ...)`
- `points(x, y, ...)`
- `abline(lm(...))`
- `text(.)`
- `axis`
- `legend`

Hands-on 3a

- `http://vanatteveldt.com/wp-content/uploads/rcourse/handson3a.html`

Simple two variable tests

- `t.test(x, y, paired=.)`
- `cor.test(x, y)`
- `chisq.test(x,y)`

Linear models (and anova/ancova)

- `lm(y ~ x1 + x2)`
- `lm(Y ~ x1*x2)`
- `lm(Y ~ x1 + x2 + x1:x2)`
- `lm(Y ~ x1 + cat)`
- `lm(Y ~ x1 + cat - 1)`

Modeling results

- `m = lm(...)`
- `summary(m)`
- `anova(m)`
- `coef(m)`
- `fitted(m)`
- `residuals(m)`
- `plot(m)`
 - (do `par(mfrow=c(2,2))` first)
- `anova(m1, m2)`
- `http:`
`//www.statmethods.net/stats/rdiagnostics.html`

Dimensionality reduction

- `cmdscale`
 - `fit$points`
- `princomp(d, cor=TRUE)`
 - `plot(fit), biplot(fit)`
- `factanal(d, rotate=)`
- `hclust(d, method=)`
 - `plot(fit), rect.hclust(fit, k=)`

Hands-on 3b

- <http://vanatteveldt.com/wp-content/uploads/rcourse/handson3b.html>
- <http://vanatteveldt.com/wp-content/uploads/rcourse/handson3c.html>
- <http://www.openintro.org/stat/labs.php>, labs 7-8