

Introduction to reshape

Hadley Wickham

Outline

- Id/measured variable refresher
- Data in packages
- French fries data
- Tasks
- Melting and casting

Id variables or keys

- Identify a record (so must be unique)
- Like indices on a random variable
- Fixed by design of experiment
(known in advance)
- Maybe be single or composite
- As compared to measurements/measured variables

Data in packages

- R packages can contain data as well as functions
- `# install.packages("reshape")`
- `library(reshape)`
- `data(package="reshape")`

Your turn

- Load the reshape package
- Look at the french fries and tips data sets
- Read the help files for those data sets

French fries data

- Sensory data
- Id variables:
 - time: 1-10 weeks
 - treatment: 1-3
 - 12 subjects
 - replicates

French fries data

- Measured variables:
 - potato
 - buttery
 - grassy
 - rancid
 - painty

Tasks

- Create summary tables and graphics
- Explore patterns of missing data
- Investigate inter-rep reliability



First

- First name
- This gets the value of the variable "first" and prints it out
- When you use the `print` function, you have to specify the variables you want to print out (e.g. `print(my_var=5)`)

Your turn

- Melt the french fries data
- Melt the tips data
- Make sure to think about which variables are measured and which are id!
- Look at the molten data - how is it different?

Subject	Age	Height	Weight
John	20	1.95	100
John	21	1.96	NA

Subject	Age	Variable	Value
John	20	Height	1.95
John	20	Weight	100
John	21	Height	1.96

Molten form

- Has a new variable called “variable”
- It's a new id variable, which represents the name of the measured variables
- Just one measured variable - value
- Makes it easy to “cast” into new forms

Casting

- Just like pivot tables (and facetting plots!)
- Row variables, column variables, and a summary function
- `cast(molten, row ~ col, summary)`
- `cast(molten, row ~ ., summary)`
- `cast(molten, . ~ col, summary)`

Missing values

- Sampling vs structural
- Sampling occur when you just don't see something (ie. someone didn't fill out the form)
- Structural missings occur when a measurement just isn't possible
- Structural missings depend on the structure!

Gender	Pregnant	Non-prg
Male	NA	30
Female	21	25

Gender	Pregnant	Value
Male	No	30
Female	No	25
Female	Yes	21

Explore missing values

- `ffm <- melt(french_fries, id=1:4, na.rm=T)`
- `cast(ffm, time ~ subject, length)`
- `cast(ffm, time ~ subject, function(x) 30-length(x))`

Your turn

- Explore the missing values yourself
- Where are they? How do you think they would influence the analysis?

Other summaries

- `cast(ffm, time ~ treatment, mean)`
- `cast(ffm, time ~ treatment, sd)`
- `cast(ffm, time ~ treatment, c(mean, sd))`
- `cast(ffm, treatment + result_variable ~ time, c(mean, sd))`

Your turn

- Use these crude summaries to explore if the treatments work
- (How would you more formally test if they work or not?)
- Create some graphics as well - you might want to use a combination of graphics and reshape

Inter-rep reliability

- Each person tastes fries from the same oil twice – how consistent are they?
- Need to reshape, not aggregate, the data
- `ffrep <- reshape(ffm, ... ~ rep)`
- Now have dataset with two columns, one for each rep
- `qplot(X1, X2, data=ffrep, facets =. ~ variable)`

Your turn

- Try and explore similar a relationships with time?
- Are measurements taken closer together, more or less similar than those taken a long time apart?