STAT 545A
Class meeting #4
Monday, September 16, 2013

Dr. Jennifer (Jenny) Bryan

Department of Statistics and Michael Smith Laboratories

any questions from tutorial re: R objects and indexing?

Have a look around the bullet lists for HW 1 and HW 2 and take a hard look at your links and filenames. Do yours look funny? Don't wait for Song or me to alert you. Fix it.

No class this Wed.

HW #3 "Data Aggregation". Due before class next Monday. Details coming soon to web. Submission process likely to change

Other stuff to do before next class: learn how to get help and participate in the discussion about datasets

A few words from Matt G re: anti-aliasing and PNGs and Windows; see his post on the Google Group and/or his demo on Rpubs.

How to ask a question to maximize chance of getting an answer.

The R Inferno

Patrick Burns[1]

30th April 2011

Read about the 9th circle of R hell in The R Inferno.

Circle 9

Homework reading

Unhelpfully Seeking Help

Here live the thieves, guarded by the centaur Cacus. The inhabitants are bitten by lizards and snakes.

There's a special place for those who—not being content with one of the 8 Circles we've already visited—feel compelled to drag the rest of us into hell.

The road to writing a mail message should include at least the following stops:

9.1 Read the appropriate documentation.

"RTFM" in the jargon. There is a large amount of documentation about R, both official and contributed, and in various formats. A large amount of documentation means that it is often nontrivial to find what you are looking for—especially when frustration is setting in and blood pressure is rising.

Breathe.

There are various searches that you can do. R functions for searching include `help.search`, `RSiteSearch` and `apropos`.

If you are looking for particular functionality, then check the Task Views (found on the left-side menu of CRAN).

If you have an error, then look in rather than out—debug the problem. One way of debugging is to set the `error` option, and then use the `debugger` function:

```
options(error=dump.frames)
# command that causes the error
debugger()
```

The `debugger` function then provides a menu of the stack of functions that have been called at the point of the error. You can inspect the state of the objects inside these functions, and hopefully understand what the problem is.

The R Inferno

Patrick Burns[1]

30th April 2011

Circle 9

Unhelpfully Seeking Help

Here live the thieves, guarded by the centaur Cacus. The inhabitants are bitten by lizards and snakes.

There's a special place for those who—not being content with one of the 8 Circles we've already visited—feel compelled to drag the rest of us into hell.

"If someone has the wit and knowledge to answer your question, they probably have other things they would like to do. Making your message clear, concise and user-friendly gives you the best hope of at least one of those strangers diverting their attention away from their life towards your problem."

How to ask a question to maximize chance of getting an answer.

Homework reading

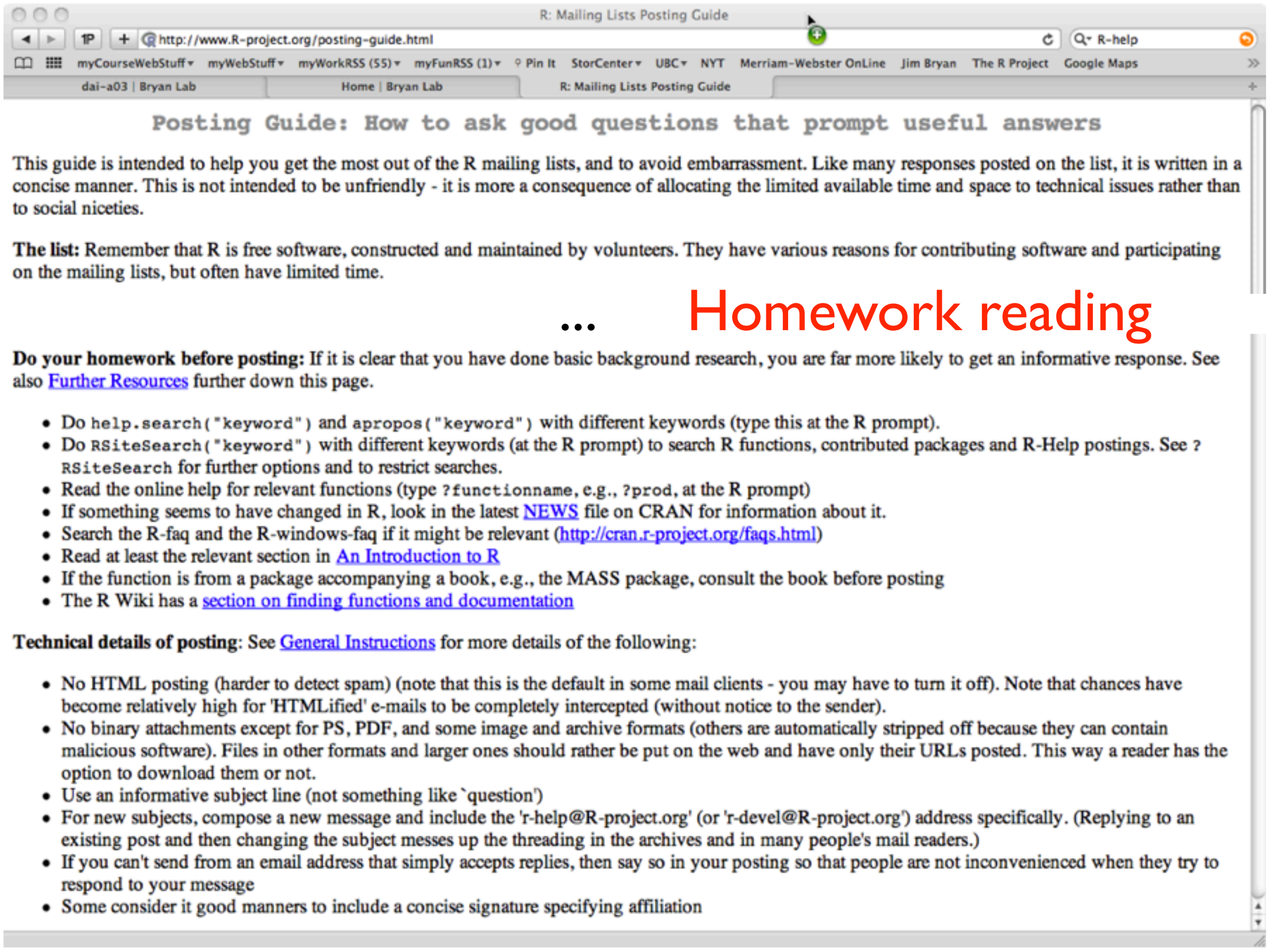
"How To Ask Questions The Smart Way"
by Eric Raymond and Rick Moen

It's OK to be ignorant; it's not OK to play stupid.

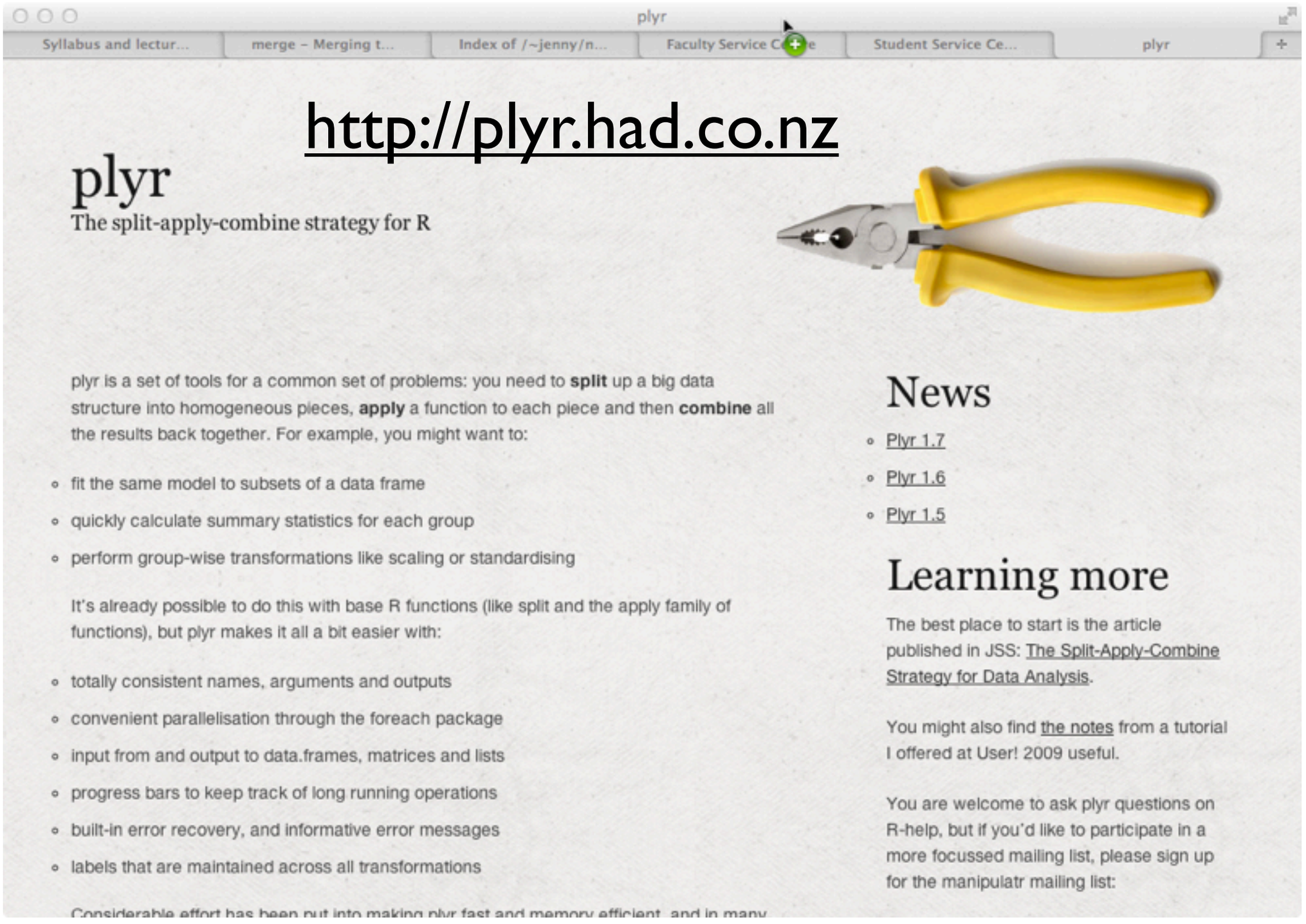
So, while it isn't necessary to already be technically competent to get attention from us, it is necessary to **demonstrate the kind of attitude that leads to competence: alert, thoughtful, observant, willing to be an active partner in developing a solution**.

The best way to get a rapid and responsive answer is to ask it like a person with smarts, confidence, and clues who just happens to need help on one particular problem.

R-help (you probably shouldn't be posting here!) has a good posting guide



The plyr package is what I advise long-term for data aggregation.



JB found it hard to get started with plyr by reading documentation for individual functions. You need to get the big picture and then it will all come into focus. Read this paper!

Hadley Wickham.
The split-apply-combine strategy for data analysis.
Journal of Statistical Software, vol. 40, no. 1, pp. 1–29, 2011.
http://www.jstatsoft.org/v40/i01/paper

Journal of Statistical Software
April 2011, Volume 40, Issue 1. *http://www.jstatsoft.org/*

The Split-Apply-Combine Strategy for Data Analysis

Hadley Wickham
Rice University

Abstract

Many data analysis problems involve the application of a split-apply-combine strategy, where you break up a big problem into manageable pieces, operate on each piece independently and then put all the pieces back together. This insight gives rise to a new R package that allows you to smoothly apply this strategy, without having to worry about the type of structure in which your data is stored.

The paper includes two case studies showing how these insights make it easier to work with batting records for veteran baseball players and a large 3d array of spatio-temporal ozone measurements.

Keywords: R, apply, split, data analysis.

split apply combine

Input \ *Output*	Array	Data frame	List	Discarded
Array	`aaply`	`adply`	`alply`	`a_ply`
Data frame	`daply`	`ddply`	`dlply`	`d_ply`
List	`laply`	`ldply`	`llply`	`l_ply`

- `a*ply(.data, .margins, .fun, ..., .progress = "none")`
- `d*ply(.data, .variables, .fun, ..., .progress = "none")`
- `l*ply(.data, .fun, ..., .progress = "none")`

we spent the rest of class time going through this tutorial:

http://www.stat.ubc.ca/~jenny/STAT545A/block04_dataAggregation.html