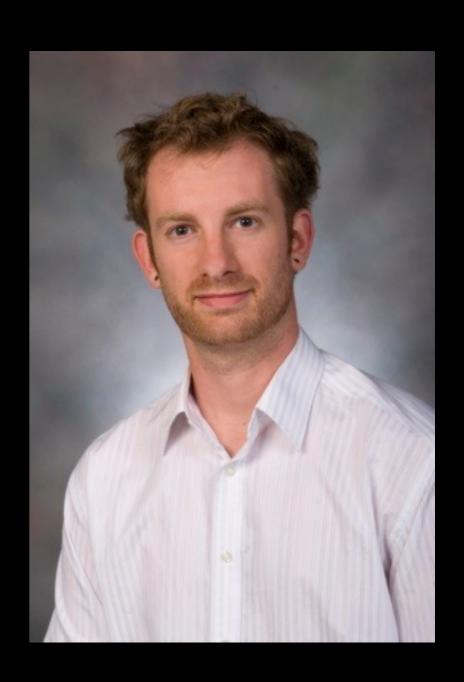# plyr
## Introduction

## Hadley Wickham

1. About me

2. Active learning

3. Resources

4. Outline of the course

# About me

New assistant professor at Rice University.

Interested in developing tools that make analysis easier - focussing more on data cleaning, organisation and exploration than traditional statistics.

Have written 17 R packages at last count.

# Active learning

Hard to absorb much information from pure lecture format (especially after lunch!), so we'll be breaking things up with some short interactive activities.

Please take 30 seconds now to introduce yourself to your neighbours. You'll be working with them in a bit.

# Resources

All code, data & slides available at
http://had.co.nz/plyr/09-user

Tutorial covers most of the material in
http://had.co.nz/plyr/plyr-intro.pdf, but
with more examples, and at a higher level.

Problems or questions?  Use the
manipulatr mailing list:
http://groups.google.com/group/manipulatr

# Outline

Basic strategy: split-apply-combine.

Exploring US baby name trends

Modelling Texas house sales

Wrap up: how it all fits together

Can't go into much detail in 3 hours, but hopefully this tutorial will get you off to a good start.

# Code

`bnames-explore.r`—what we'll work through next

`bnames-cluster.r`—expanded example showing how to find cluster of similar names

`tx-explore-houston.r`—introduction to housing data

`tx-explore-all.r`—process of model building for large data

`nnet.r`—fitting a neural network with many random starts and varying parameters

# Split-apply-combine

**Split** up a big dataset

**Apply** a function to each piece

**Combine** all the pieces back together

Keep this theme in mind as we work through the examples.