# Consulting for a Dose Response Modeling Package in R

MANGOSOLUTIONS
data analysis that delivers

Tom Taverner, ttaverner@mango-solutions.com

MANGOSOLUTIONS
data analysis that delivers

# Structure of this talk

- About Mango and my background
- The “R consulting” experience
- Delivering the product using software engineering and R
- Outcome

MANGOSOLUTIONS
data analysis that delivers

# Overview of Mango Solutions

- Private Company formed in 2002
- Global Team of ~70
- Cross-Sector Software and Services
- ISO 9001 Accredited

MANGOSOLUTIONS
data analysis that delivers

# Located here ...

Bath, UK

London, UK

Shanghai, CN

MANGOSOLUTIONS
data analysis that delivers

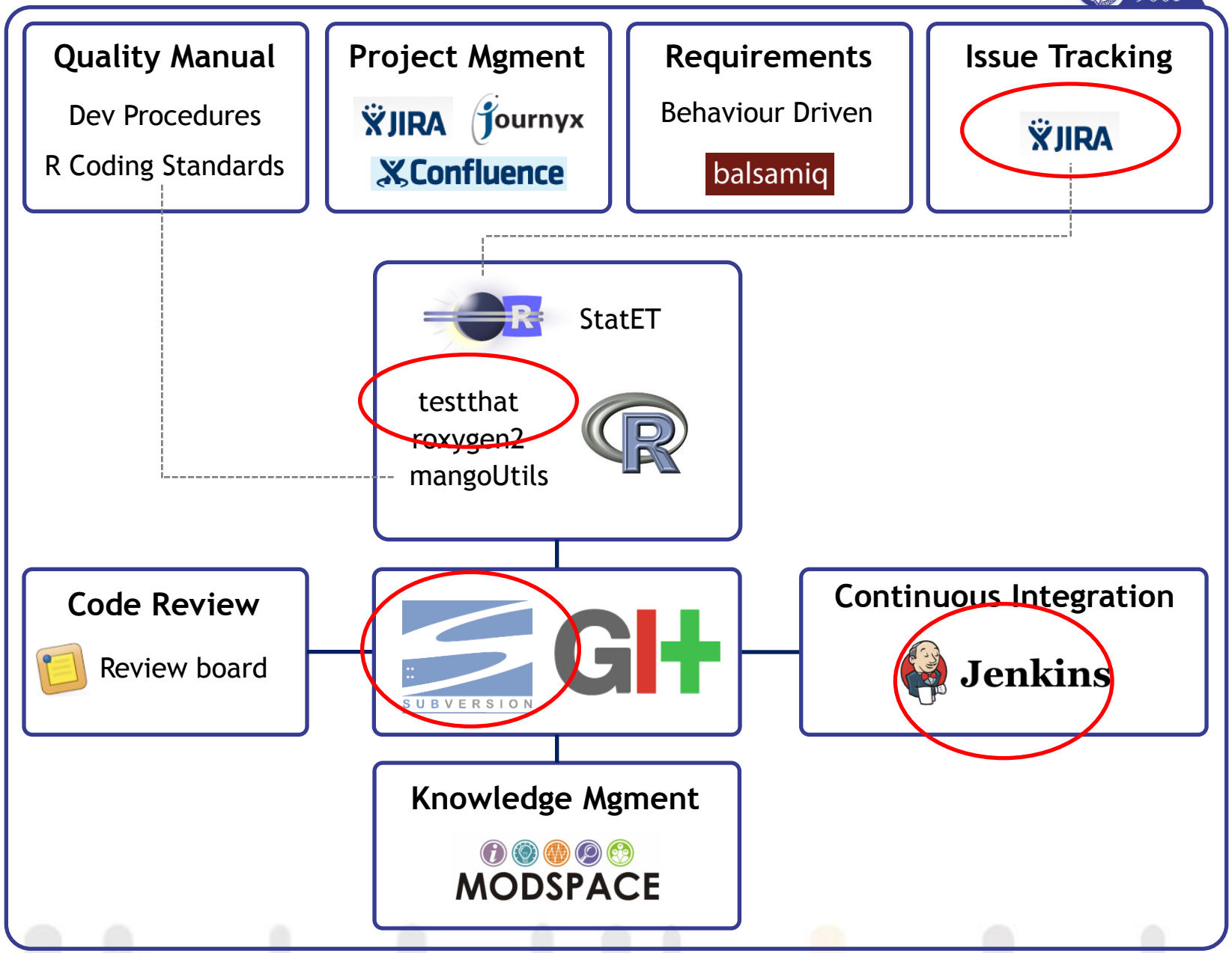
# The consultant experience

- My background – biochemistry/bioinformatics
- PhD Cambridge 2006
- Did a couple of post-docs in the USA and got interested in statistics, wrote a lot of R
- MSc in statistics, Warwick, 2012 and went to work for a consulting firm, Mango
- Mango's main product – Navigator – R based front end

MANGOSOLUTIONS
data analysis that delivers

# Formal R Development

- Creating sophisticated analytic applications requires a formal development approach
- This mostly means taking standard development practices and applying it to analytics
- Mango's formal R development procedures and structure has been evolving since its inception ~2004


Quality Manual
Dev Procedures
R Coding Standards
Project Mgment
JIRA
Journyx
Confluence
Requirements
Behaviour Driven
balsamiq
Issue Tracking
JIRA
StatET
testthat
roxygen2
mangoUtils
Code Review
Review board
SUBVERSION
GIT
Continuous Integration
Jenkins
Knowledge Mgment
MODSPACE

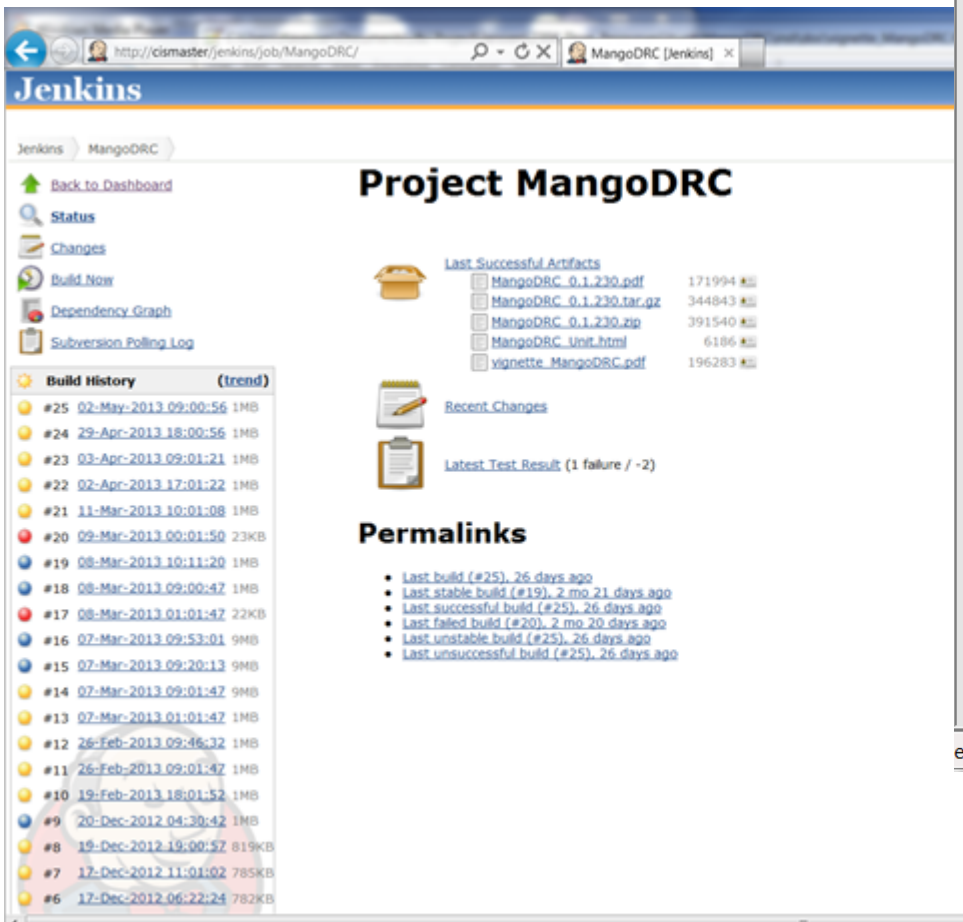
MANGOSOLUTIONS
data analysis that delivers

# R Quality Standards

- Coding standards have grown over years
- Quality and coverage of comments
- Coding style, structure, nesting
- Roxygen2 headers
- Unit testing (Runit to testthat)
- Naming conventions

MANGOSOLUTIONS
data analysis that delivers

# Continuous integration with Jenkins




```
1  <project name = "MangoDRC trunk" default = "output
2
3      <property environment = "env" />
4      <property name = "package.name" value = "Mango
5      <property name = "major.version" value = "0.1"
6
7      <condition property = "hostname" value="${env.
8          <os family = "windows" />
9      </condition>
10
11     <property file = "${hostname}.properties" />
12     <property name = "path.old" value = "${env.Pat
13     <property name = "path.new" value = "${miktex.
14     <property name = "build.script" value = "build
15     <property name = "build.folder" value = "${bas
16     <property name = "output.folder" value = "${ba
17     <property name = "exe.Rscript" value = "${r.ho
18     <property name = "exe.R" value = "${r.home}/bi
19     <property name = "package.version" value = "${
20
21     <target name="message">
22         <echo message="R HOME is ${env.R_HOME}"/>
```

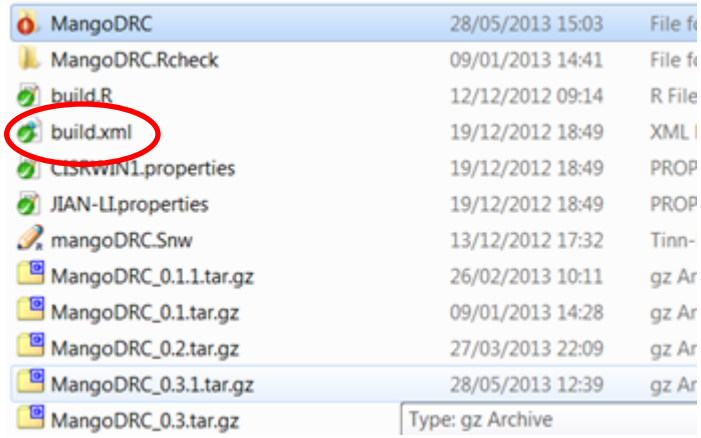
Ant build script

eXtensible Markup Language file

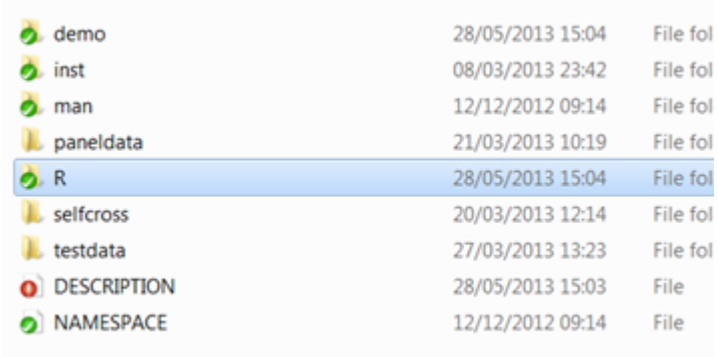
MANGOSOLUTIONS
data analysis that delivers

# Exemplar package structure

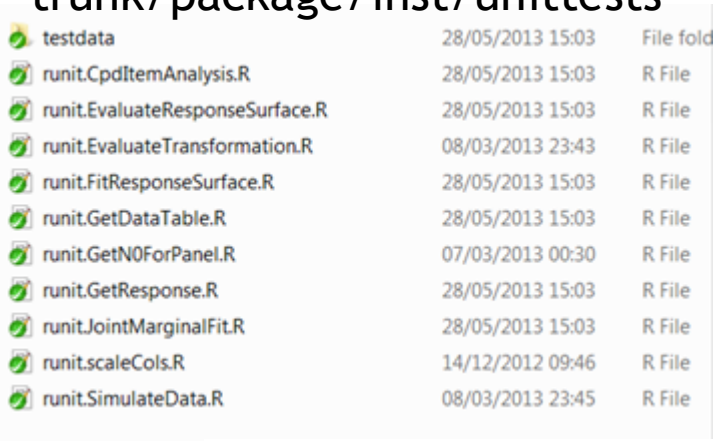
trunk/



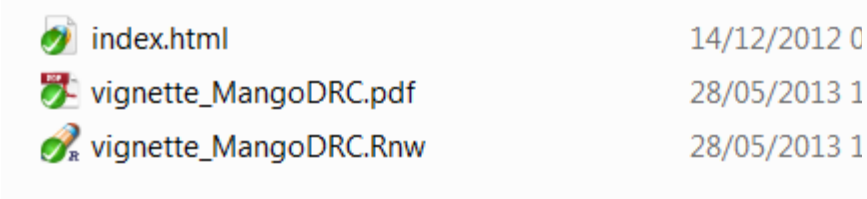

trunk/package/




trunk/package/inst/unittests




trunk/package/inst/doc



MANGOSOLUTIONS
data analysis that delivers

# Case Studies

- Ranges from information we can fully disclose to only being able to say vague things about the customer
- Only so much info we can give today – please see us after or contact us and we can step through things in more detail

Richard Pugh = rpugh@mango-solutions.com

Andy Vodden = avodden@mango-solutions.com

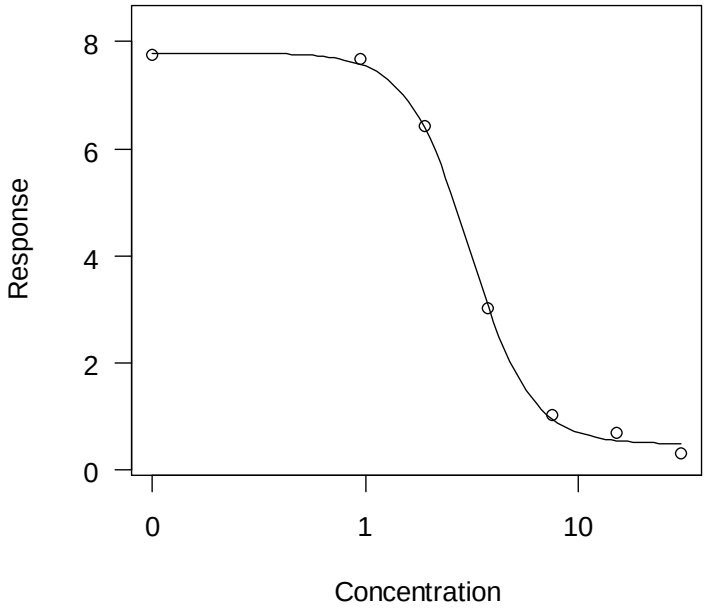
Tom Taverner = ttaverner@mango-solutions.com

MANGOSOLUTIONS
data analysis that delivers

# CASE STUDY
# DOSE RESPONSE MODELING

# Overview

- Client wants to model dose-response curves in R
  - Current package is rather slow
    - Doesn't model interactions
  - Proposed a new mathematical model for drug interactions
    - Implementation and memory requirements



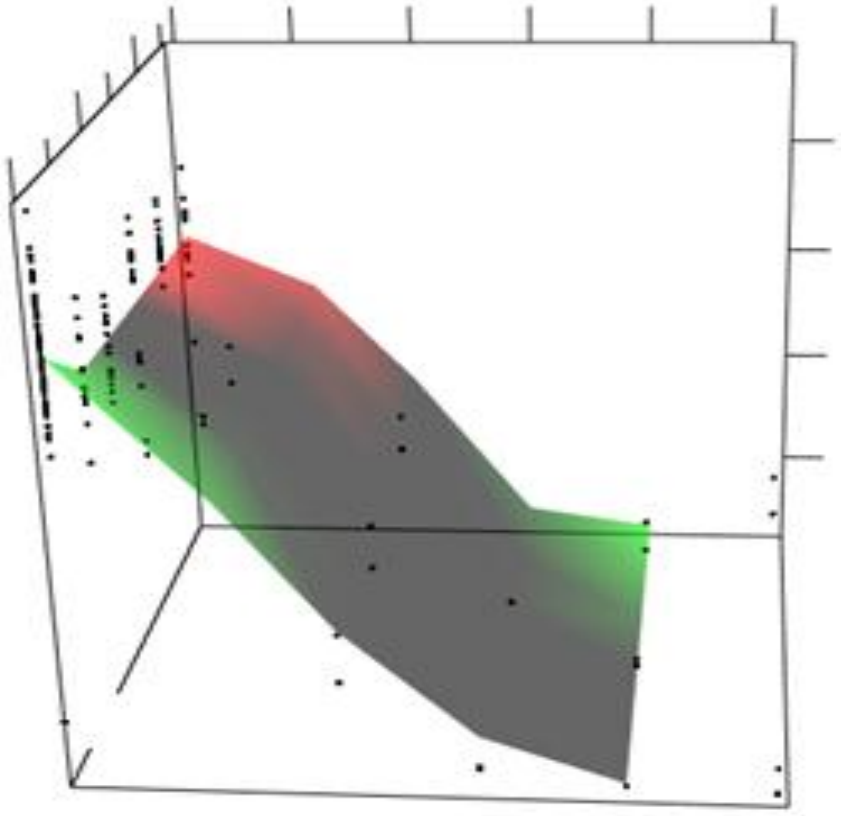

$$f(x) = c + \frac{d - c}{1 + \exp(b(\log(x) - \log(e)))}$$

# R findings

- Existing package did not vectorize code
- Use of optim() needed to be checked more carefully
- Signs of parameters were confusing
- Package had too many layers of abstraction
- It was very helpful to have the existing package to check results against

# Enhancing the delivery using R's packages

- rgl
- mgcv
- testthat



MANGOSOLUTIONS
data analysis that delivers

# Working with the client

- We had weekly meetings and constant email contact
- We shared package builds through Dropbox for regular testing
- A number of iterations were needed and some issues arose
  - Equation formatting didn't survive Word documents!
  - Turning ideas into code sometimes exposed hidden assumptions
  - Needed data quality checking and backfitting
  - Estimating times for project completion

MANGOSOLUTIONS
data analysis that delivers

# Consultancy role

- Requires completely different skills from academic expertise
- Much shorter time frames
- Results oriented
- Confidentiality and security
- Interesting problems
- Great opportunity to learn about new area (pharma)

MANGOSOLUTIONS
data analysis that delivers

# Case Study: Dose Response Modelling

## Outcome

- The package in general was a success
- We increased efficiency substantially and created a platform for more sophisticated data analysis and modeling
- We  improved graphical outputs, making them attractive, interactive and 3 dimensional
- Continuous integration and unit tests ensure builds work each time
- Commercial benefit in that it allows screening in shorter time

MANGOSOLUTIONS
data analysis that delivers

# Summary

- Thank you
- We could only cover certain amount of detail in time, so ask us for more if interested!