Bayesian Hierarchical Models

Gavin Shaddick,

Millie Green, Matthew Thomas

University of Bath

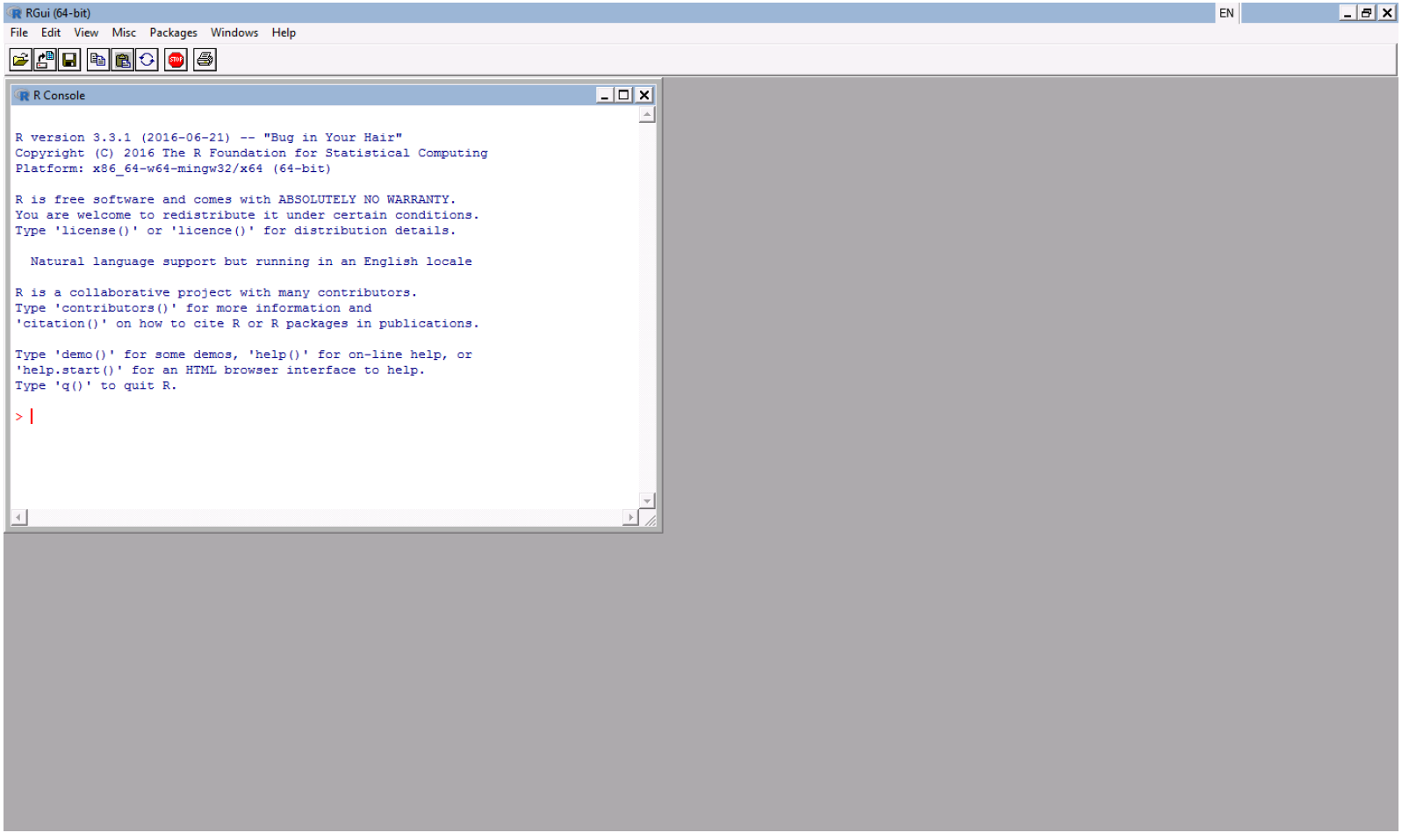
6^{th} - 9^{th} December 2016

IMPLEMENTING BAYESIAN MODELS USING R-INLA

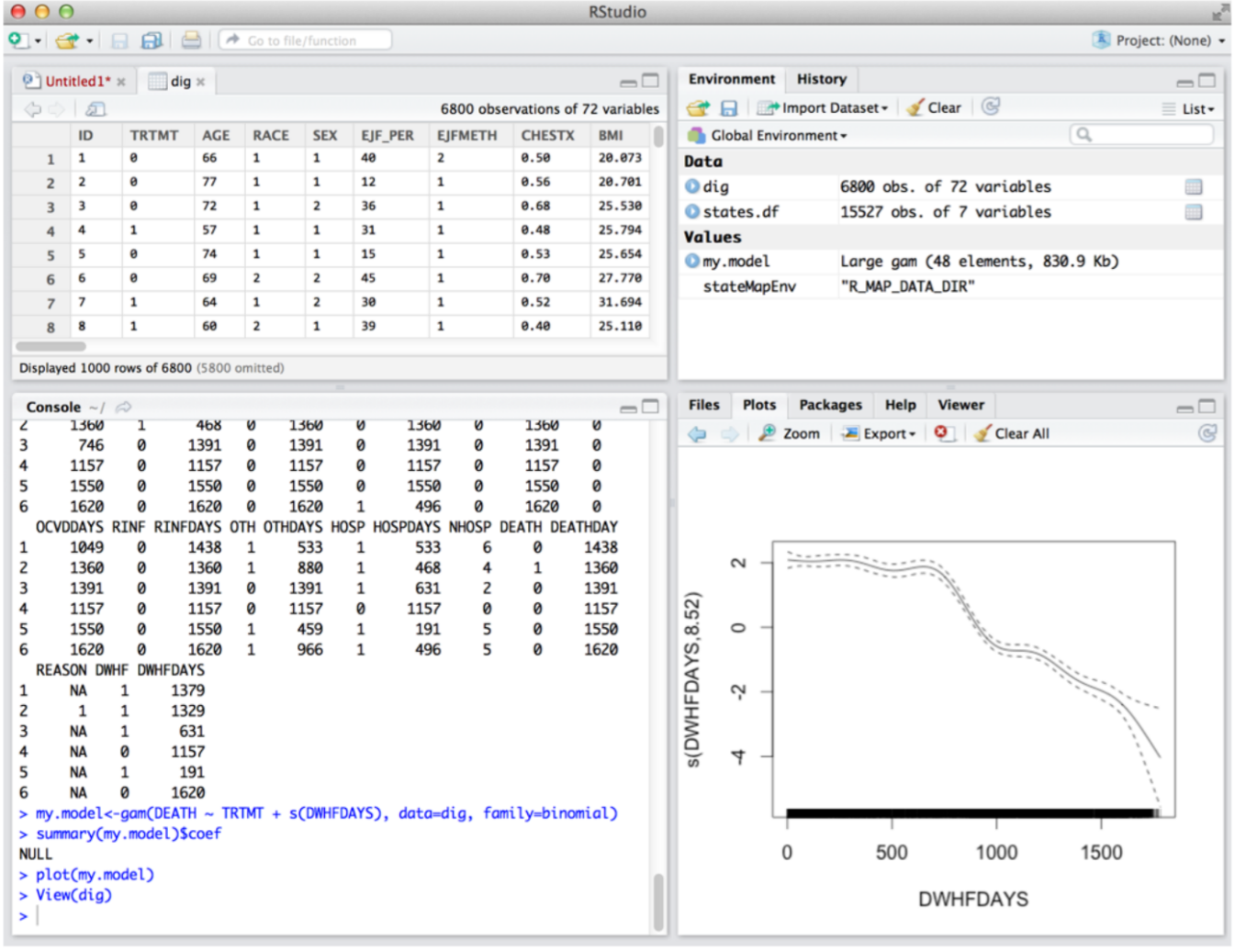
Outline

R and RStudio

R


RGui (64-bit)
EN
File Edit View Misc Packages Windows Help
R Console
R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
>

RSTUDIO


RStudio
Project: (None)
Untitled1*
dig
6800 observations of 72 variables
ID TRTMT AGE RACE SEX EJF_PER EJFMETH CHESTX BMI
1 1 0 66 1 1 40 2 0.50 20.073
2 2 0 77 1 1 12 1 0.56 20.701
3 3 0 72 1 2 36 1 0.68 25.530
4 4 1 57 1 1 31 1 0.48 25.794
5 5 0 74 1 1 15 1 0.53 25.654
6 6 0 69 2 2 45 1 0.70 27.770
7 7 1 64 1 2 30 1 0.52 31.694
8 8 1 60 2 1 39 1 0.40 25.110
Displayed 1000 rows of 6800 (5800 omitted)
Environment History
Import Dataset Clear List
Global Environment
Data
dig 6800 obs. of 72 variables
states.df 15527 obs. of 7 variables
Values
my.model Large gam (48 elements, 830.9 Kb)
stateMapEnv "R_MAP_DATA_DIR"
Console ~/
2 1360 1 468 0 1360 0 1360 0 1360 0
3 746 0 1391 0 1391 0 1391 0 1391 0
4 1157 0 1157 0 1157 0 1157 0 1157 0
5 1550 0 1550 0 1550 0 1550 0 1550 0
6 1620 0 1620 0 1620 1 496 0 1620 0
OCVDDAYS RINF RINFDAYS OTH OTHDAYS HOSP HOSPDAYS NHOSP DEATH DEATHDAY
1 1049 0 1438 1 533 1 533 6 0 1438
2 1360 0 1360 1 880 1 468 4 1 1360
3 1391 0 1391 0 1391 1 631 2 0 1391
4 1157 0 1157 0 1157 0 1157 0 0 1157
5 1550 0 1550 1 459 1 191 5 0 1550
6 1620 0 1620 1 966 1 496 5 0 1620
REASON DWHF DWHFDAYS
1 NA 1 1379
2 1 1 1329
3 NA 1 631
4 NA 0 1157
5 NA 1 191
6 NA 0 1620
> my.model<-gam(DEATH ~ TRTMT + s(DWHFDAYS), data=dig, family=binomial)
> summary(my.model)$coef
NULL
> plot(my.model)
> View(dig)
Files Plots Packages Help Viewer
Zoom Export Clear All
s(DWHFDAYS,8.52)
2
0
-2
-4
0
500
1000
1500
DWHFDAYS

Statistical Analyses

Statistical Analysis in R

- R comes with many statistical tools already installed
  - descriptive statistics
  - visualisation
  - statistical tests
  - model fitting.

Packages

CAN R DO MORE?

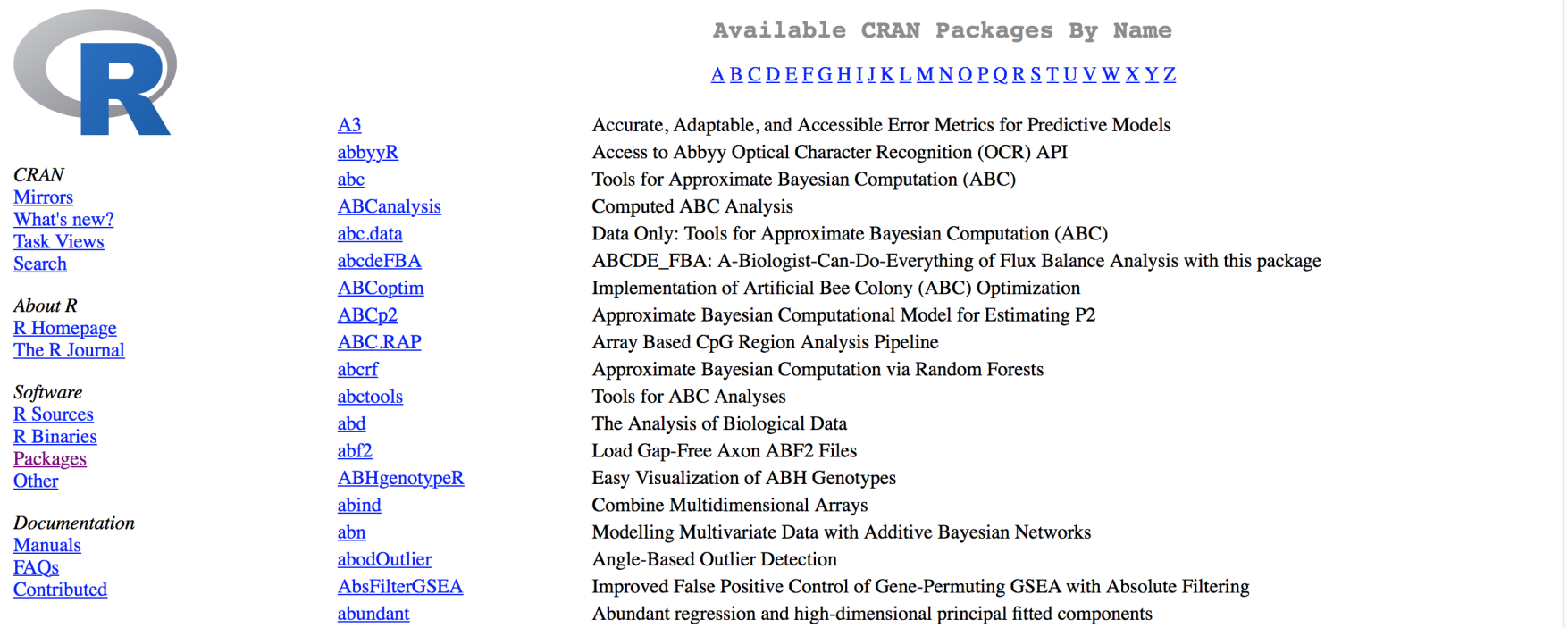
- The default installation of R has a comprehensive set of tools for statistical analyses.
- To meet the specific needs of data scientists, many other statistical tools are readily available in the form of packages.
- Packages are collections of functions and data.
- "During the last decade, the momentum coming from both academia and industry has lifted R to become the single most important tool for computational statistics, visualisation and data science."

R Packages: Examples used in this course

- `ggplot2`
- `raster`
- `Rmisc`
- `mgcv`
- `maptools`
- ... many many more!!

Other R Packages

- A list of R Packages can be seen and downloaded from `https://cran.r-project.org`



CRAN
Mirrors
What's new?
Task Views
Search

About R
R Homepage
The R Journal

Software
R Sources
R Binaries
Packages
Other

Documentation
Manuals
FAQs
Contributed

Available CRAN Packages By Name

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

A3	Accurate, Adaptable, and Accessible Error Metrics for Predictive Models
abbyyR	Access to Abbyy Optical Character Recognition (OCR) API
abc	Tools for Approximate Bayesian Computation (ABC)
ABCanalysis	Computed ABC Analysis
abc.data	Data Only: Tools for Approximate Bayesian Computation (ABC)
abcdeFBA	ABCDE_FBA: A-Biologist-Can-Do-Everything of Flux Balance Analysis with this package
ABCoptim	Implementation of Artificial Bee Colony (ABC) Optimization
ABCp2	Approximate Bayesian Computational Model for Estimating P2
ABC.RAP	Array Based CpG Region Analysis Pipeline
abcrf	Approximate Bayesian Computation via Random Forests
abctools	Tools for ABC Analyses
abd	The Analysis of Biological Data
abf2	Load Gap-Free Axon ABF2 Files
ABHgenotypeR	Easy Visualization of ABH Genotypes
abind	Combine Multidimensional Arrays
abn	Modelling Multivariate Data with Additive Bayesian Networks
abodOutlier	Angle-Based Outlier Detection
AbsFilterGSEA	Improved False Positive Control of Gene-Permuting GSEA with Absolute Filtering
abundant	Abundant regression and high-dimensional principal fitted components

R-INLA

R-INLA

- The R–INLA package provides a practical implementation of Integrated Nested Laplace Approximations (INLA).
- The class of models that can be expressed in this form and thus can be used with R–INLA is very large and includes, amongst others, the following:
  - Dynamic linear models.
  - Stochastic volatility models.
  - Generalised linear (mixed) models.
  - Generalised additive (mixed) models.
  - Spline smoothing.
  - Semi–parametric regression.
  - Disease mapping.
  - Log–Gaussian Cox–processes.
  - Model–based geostatistics.
  - Spatio–temporal models.
  - Survival analysis.

The syntax of R–INLA

- There are three main parts to fitting a model using R–INLA:
  1. The data.
  2. Defining the model formula.
  3. The call to the INLA program.
- The basic syntax of running models in R–INLA is very similar in appearance to that of `glm` in R and takes the general form `formula, data, family` but with the addition of the specification of the nature of the random effects, `f()`.
- For the latter component, common examples include
  - `f(i, model="iid")` (independent)
  - `f(i, model="rw")` (random walk of order one)
  - `f(i, model="ar")` (autoregressive of order p).

Fitting a Poisson regression model in R–INLA

- An extension of the standard Poisson model to include log–normal random effects in the linear predictor

$$\log \mu_l = \beta_0 + \beta_{0i} + \beta_1 X_l + \beta_d X_l + \epsilon_l \tag{1}$$

where β_l represents the effect of exposure, β_d is the effect of an area-level covariate and β_{0i} denotes the random effect for area i.

- The syntax of the R–INLA code to fit this model is very similar to that of a standard `glm` in R.

Fitting a Poisson regression model in R–INLA

```
> formula = Y ~ X1+X2 + f(i, model="iid")
> model   = inla(formula, family="poisson", data=data)

Call:
"inla(formula = formula, family = "poisson", data = data)"

Time used:
 Pre-processing    Running inla Post-processing          Total
       0.278389        0.286911        0.125699       0.690999

Integration Strategy: Central Composite Design

Model contains 1 hyperparameters
The model contains 3 fixed effect (including a possible
    intercept)

Likelihood model: poisson

The model has 1 random effects:
1.'i' is a IID model
```

Fitting a Poisson regression model in R–INLA

```
> summary(model)

Call:
"inla(formula = formula, family = "poisson", data = data)"

Time used:
 Pre-processing     Running inla Post-processing         Total
         0.2784           0.2869           0.1257        0.6910

Fixed effects:
                mean    sd 0.025quant 0.5quant 0.975quant
```

Fitting a Poisson regression model in R–INLA

```
(Intercept) 2.4960 0.0713    2.3553    2.4962      2.6355
X1          0.1187 0.0310    0.0578    0.1186      0.1796
X2          0.0578 0.0074    0.0433    0.0578      0.0722

Random effects:
Name       Model
 i    IID model

Model hyperparameters:
                   mean      sd 0.025quant 0.5quant 0.975quant
Precision for i 3.784 0.3548       3.131    3.769       4.525

Expected number of effective parameters(std dev):
    321.42(3.926)
Number of equivalent replicates : 1.223

Marginal Likelihood:  -1513.92
```

Fitting models in R–INLA

- Future details can be found on the R–INLA webpage: `http://www.R-INLA.org`.