



# *Bayesian Data Analysis for Social Scientists*

Mark Andrews & Thom Baguley

Psychology, Nottingham Trent University

✉ mark.andrews@ntu.ac.uk,  @xmjandrews

✉ thomas.baguley@ntu.ac.uk,  @seriousstats

 <https://github.com/lawsofthought/bayes-lmu-2018>

## *Background: ESRC Advanced Training Initiative*

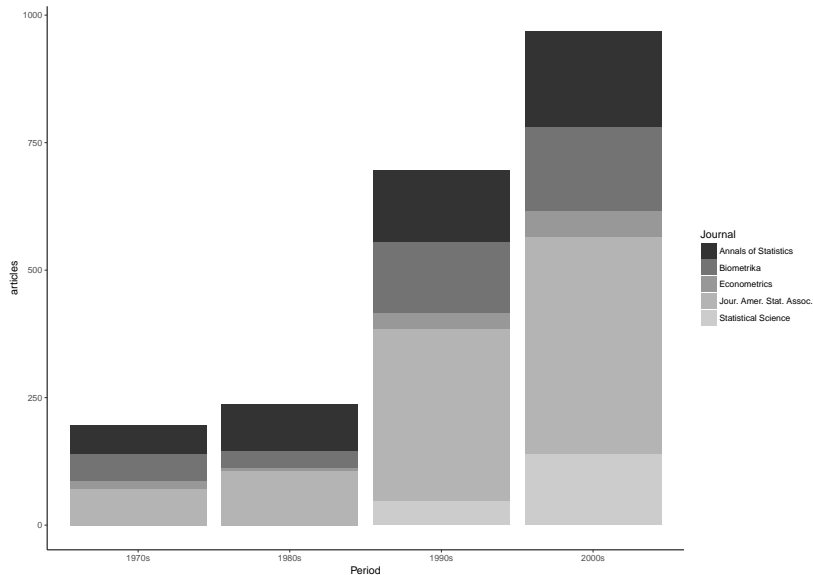
- ▶ The Advanced Training Initiative (ATI) by the Economic and Social Research Council (ESRC) provided grants to support training in advanced social science topics.
- ▶ We were funded to provide a series of workshops on Bayesian data analysis each year for the years 2015, 2016, & 2017:

<http://www.priorexposure.org.uk>

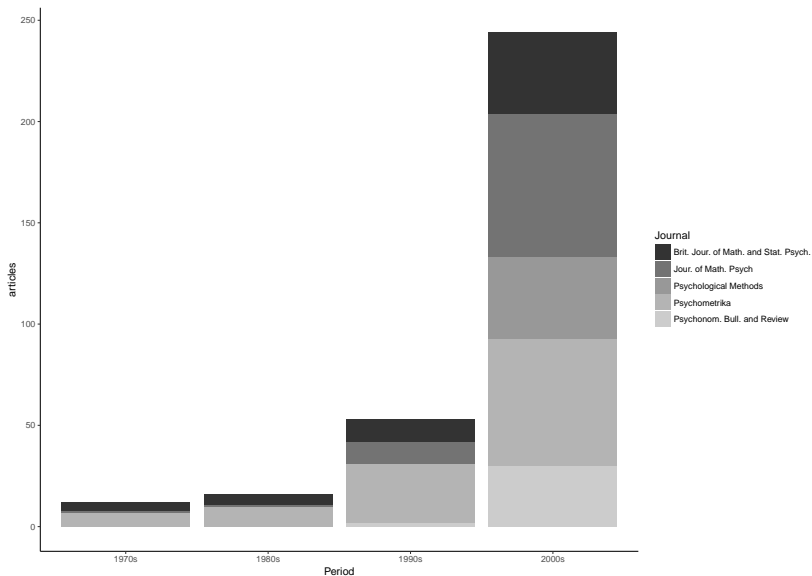
- ▶ Each workshop was limited to around 25 attendees, and could be attended by any UK based social science researchers (post-graduate students and above).
- ▶ We taught 4 workshops in 2015, 9 in 2016, and 9 in 2017

## *Our case for support*

Bayesian methods are growing in popularity, but are not yet part of the social science curriculum.



# *Our case for support*



## *Workshops: Overview*

- ▶ Each workshop was planned to be a combination of lecture style teaching and practical exercises.
- ▶ All practical exercises were computer based and used R and Jags<sup>1</sup>.
- ▶ Most lecture teaching involved R and Jags based demonstrations, which could be followed along step by step by attendees.
- ▶ Attendees were required to use their laptops, and details of how to install the required software were provided in advance.
- ▶ Source code and (most) other teaching materials are available at: <https://github.com/lawsofthought/priorexposure>.

---

<sup>1</sup>Why Jags and not its alternatives? See below for discussion.

# *Workshop 1: Bayes for beginners*

- ▶ This workshop aimed to be a general introduction to Bayesian data analysis and how it differs from the more familiar classical approaches to data analysis.
- ▶ Here, we provided a gentle introduction to Bayesian methods. Topics included:
  - ▶ Examples of Bayesian inference and using prior information in simple statistics problems.
  - ▶ Understanding the likelihood function.
  - ▶ Hypothesis testing using Bayes factors.

## *Workshop 2: Doing Bayesian data analysis*

- ▶ This workshop aimed to provide a solid theoretical and practical foundation for real-world Bayesian data analysis in psychology and social sciences.
- ▶ Topics included:
  - ▶ Some detailed examples of analytically tractable Bayesian inference (e.g. inference of Bernoulli random variables, inference of Poisson random variables, inference of means of univariate Normal models, etc.)
  - ▶ Introduction to probabilistic modelling with Jags.
  - ▶ Linear models with Jags.

## *Workshop 3: Introduction to advanced Bayesian data analysis and Bayesian multilevel modelling*

- ▶ This workshop focused on advanced probabilistic modelling in Bayesian data analysis, and in particular, Bayesian data analysis using multilevel regression models.
- ▶ Topics included:
  - ▶ Multilevel linear models.
  - ▶ Multilevel generalized linear models, e.g. logistic regression, Poisson regression.
  - ▶ Examples included models with categorical predictors, interactions, random slope and random intercept models, crossed and nested structures.



## *Workshop 4: Nonlinear and latent variable models*

- ▶ This final workshop focused on Bayesian latent variable modelling, particularly using mixture models, and nonlinear regression.
- ▶ Topics included:
  - ▶ Nonlinear regression modelling using radial basis functions.
  - ▶ Nonlinear regression modelling using Gaussian processes.
  - ▶ Finite mixture modelling.
  - ▶ Nonparametric mixture modelling using Dirichlet processes.

# Participants

- ▶ Attendees were students and researchers from psychology, sociology, criminology, geography, linguistics, neuroscience, economics, epidemiology, education, business studies, etc.
- ▶ A more detailed survey of attendees of this month's (April, 2017) workshops (workshops 1 & 2) showed:
  - ▶ About 50% of attendees are from psychology (usually experimental, cognitive).
  - ▶ About 50% are PhD students.
  - ▶ In terms of general statistical knowledge, attendees rate themselves as around  $\frac{5.5}{10}$  on average.
  - ▶ In terms of statistical computing skill, they rate themselves as around  $\frac{3.5}{10}$  on average.
  - ▶ In terms of knowledge of Bayesian methods, they rate themselves as around  $\frac{2.2}{10}$  on average.
  - ▶ In terms of motivation, about  $\frac{2}{3}$  said they were attending to learn more about hypothesis testing and Bayes factors.

## *Some lessons learned*

- ▶ Delving into mathematical details, e.g. derivations of formulae for posterior distributions, did not prove to be very effective.
- ▶ Learning by building and running Jags models proved much more effective.
- ▶ Being comfortable with R is vital. Pre-workshop R bootcamps were popular and effective.
- ▶ Software installation problems can stymie progress.
- ▶ For many attendees, Bayesian data analysis means Bayesian hypothesis testing (with Bayes factors). While for us, Bayesian data analysis is more about flexible probabilistic modelling.
- ▶ The age of Bugs/Jags has (probably) passed, Stan is now the preferred choice as a probabilistic modelling language.