

非常高兴与大家进行半小时的 R 分享

自我介绍：

徐昊立，毕业于华中科技大学同济医学院，
供职于广东省妇幼保健院，医师，从事出生缺陷筛查与
预防。

一、 R 简介

- R 语言身世
- 为什么要学习 R 语言
- 学习 R 语言基础
- R 的资源
 - 资料与文献
 - 网站
 - 统计分析软件包

R 语言身世

R 由各路无私奉献者成就



Ross Ihaka



Robert Gentleman

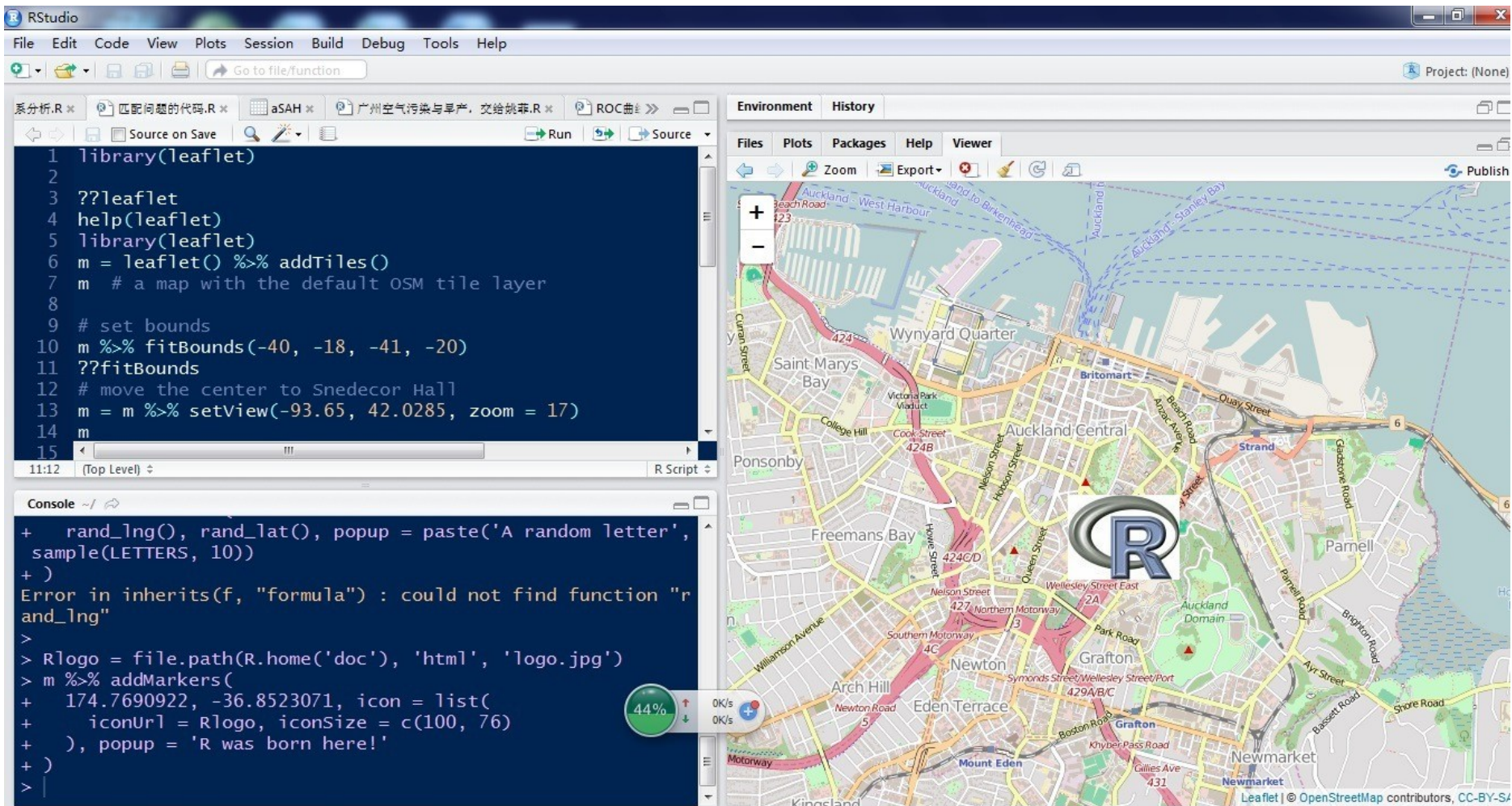


Bill Venables

R 的故乡

```
install.packages("surveillance")
library(leaflet)
??leaflet
help(leaflet)
library(leaflet)
m = leaflet() %>% addTiles()
m # a map with the default OSM tile layer
# set bounds
m %>% fitBounds(-40, -18, -41, -20)
??fitBounds
# move the center to Snedecor Hall
m = m %>% setView(-93.65, 42.0285,
zoom = 17)
m
```

```
# use automatic bounds derived from
lng/lat data
m = m %>% clearBounds()
# marker
m %>% addMarkers(rand_lng(),
rand_lat())
m %>% addMarkers(
  rand_lng(), rand_lat(), popup = paste('A
random letter', sample(LETTERS, 10))
)
Rlogo = file.path(R.home('doc'), 'html',
'logo.jpg')
m %>% addMarkers(
  174.7690922, -36.8523071, icon = list(
    iconUrl = Rlogo, iconSize = c(100, 76)
  ), popup = 'R was born here!'
)
```



为什么要学习 R 语言

- R 是
 - 一个开放 (GPL) 的统计编程环境， 58 啥都有， R 啥都能干
 - 一种软件，是集统计分析与图形直观显示于一体。
- 目前由 R 核心开发小组 (R Development Core Team – 以后用 **R DCT** 表示) 维护，他们完全自愿、努力工作负责，全球写手将各种包打包提供给我们。我们可以通过 R 计划的网站 (<http://www.r-project.org>) 了解有关 R 的最新信息和使用说明，得到最新版本的 R 软件和基于 R 的应用统计软件包。

- R 是完全免费的！！ 而 SAS 尽管是非常优秀的统计分析软件，可是你需要支付一笔\$。
- R 可以在运行于 UNIX, Windows 和 Macintosh 的操作系统上。
- R 嵌入了一个非常实用的帮助系统。
- R 具有很强的作图能力。
- 很容易学习和掌握 R 语言的语法，傻瓜式的帮助文档。
- 我们可以编制自己的函数来扩展现有的 R 语言 (这就是为什么它在不断升级完善 !!)
- 我的目标是写一个自己的包，替我和我的同行工作。
-

学习R语言基础

统计学基础

- 数据基本知识。（非结构，结构，计量，计数，分类，向量，数据框，列表。。。。）
- 统计方法基本知识，每一种统计方法适用的数据类型，阐明的问题。

编程的基础知识

- 语法
- 函数

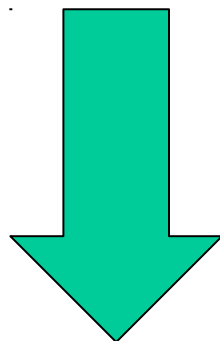
研究领域的专业知识

。。。。血糖值的意义，可能的影响因素，绝对无相关性的因素。。。。

R 的 网 站 资 源

- **R 主页** : <http://www.r-project.org>
- **CRAN (Comprehensive R Archive Network)**,
<http://cran.r-project.org>
- **CRAN 的镜像站点**
<http://cran.r-project.org/mirrors.html>
- **## 各路神仙的 R 使用案例 ##**
<http://www.r-bloggers.com/>
- **## 国内相对专一 R 的论坛 - 统计之都 ##**
<http://cos.name/cn/>

R 最重要的概念



包 (*package*)

以 功 能 为 分 类 基 础

– CRAN 提供了许多统计分析包：

<http://cran.r-project.org/src/contrib/PACKAGES.html>

- stable -- 稳妥 (分布) 广义回归分析
- tseries – 时间序列分析
- VaR – 风险值分析
- matrix – 矩阵运算
- cinterface – C 与 R 的接口
- foreign – 读写由 S, Minitab, SAS, SPSS, Stata 等软件的数据
- normix – 混合正态分布分析
- nortest – 正态分布的 Anderson-Darling 检验
- MCMCpack – 基于 Gibbs 抽样的 MCMC 抽样方法
- fracdiff – 分数差分模型的极大似然估计

还有很多……

入门简介

R_GUI

R console

RStudio

R 运行

- R_GUI

- 启动 R, 我们看到 **R GUI** (graphic user's interface) 的主窗口, 它由三部分组成

- 主菜单 工具条 R console (R 的运行窗口)

- R console

- 你的主要工作是在这里通过发布命令来完成的, 包括数据集的建立, 数据的分析, 作图等.

- 在这里你可以得到在线帮助

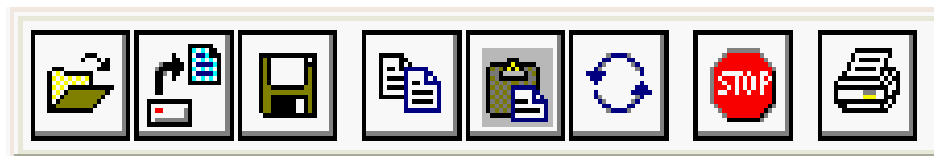
- help()** 得到相应函数的帮助, 例如 **help(plot)**

- help.start()** HTML 格式的关于 R 的帮助文件

- demo()** 得到 R 提供的几个示例; **demo(fm)**

- q()** 退出 R

- 工具条介绍



Source R code

(打开程序脚本)

Load image

(载入映像)

Save image

(保存映像)

Copy and Paste

(打印)
Stop current computation
(中断目前的计算)

- Source R code

执行 R 文件 (*.R 或 *.r)

- Save image
*.RData

保存工作空间, 文件名为

- Load image

打开已有的工作空间

- Stop current computation

»主菜单介绍（相应有中文对照）

The image shows the RGui application window with its menu system. The title bar is 'RGui'. The main menu bar includes 'File', 'Edit', 'Misc', 'Packages', 'Windows', and 'Help'. The 'File' menu is open, showing options like 'Source R code...', 'Display file(s)...', 'Load Workspace...', 'Save Workspace...', 'Load History...', 'Save History...', 'Change dir...', 'Print...', 'Save to File...', and 'Exit'. The 'Edit' menu is also open, showing 'Copy', 'Paste', 'Copy and Paste', 'Select all', 'Clear console', 'Data editor...', and 'GUI preferences...'. The 'Misc' menu is open, showing 'Stop current computation', 'Buffered output', 'List objects', 'Remove all objects', and 'List search path'. The 'Packages' menu is open, showing 'Load package...', 'Install package(s) from CRAN...', 'Install package(s) from local zip files...', 'Update packages from CRAN', 'Install package(s) from Bioconductor...', and 'Update packages from Bioconductor'. The 'Help' menu is open, showing 'Console', 'FAQ on R', 'FAQ on R for Windows', 'Manuals', 'R functions (text)...', 'Html help', 'Search help...', 'Apropos...', and 'About'. A separate box on the right lists 'An Introduction to R', 'R Reference Manual', 'R Data Import/Export', 'R Language Manual', and 'Writing R Extensions'. Pink arrows point from the title bar to the 'Edit' menu, from the 'Edit' menu to the 'Packages' menu, and from the 'Manuals' menu item to the separate box on the right.

RGui

File **Edit** **Misc** **Packages** **Windows** **Help**

Source R code...
Display file(s)...

Load Workspace...
Save Workspace...

Load History...
Save History...

Change dir...

Print...
Save to File...

Exit

Edit **Misc** **Packages** **Windows** **Help**

Copy Ctrl+C
Paste Ctrl+V
Copy and Paste Ctrl+X
Select all
Clear console Ctrl+L
Data editor...
GUI preferences...

Misc **Packages** **Windows** **Help**

Stop current computation ESC
✓ Buffered output Ctrl+W
List objects
Remove all objects
List search path

Packages **Windows** **Help**

Load package...
Install package(s) from CRAN...
Install package(s) from local zip files...
Update packages from CRAN
Install package(s) from Bioconductor...
Update packages from Bioconductor

Help

Console
FAQ on R
FAQ on R for Windows
Manuals
R functions (text)...
Html help
Search help...
Apropos...
About

An Introduction to R
R Reference Manual
R Data Import/Export
R Language Manual
Writing R Extensions

RStudio

File Edit Code View Plots Session Build Debug Tools Help

Go to file/function

Project: (N)

Environment History

Import Dataset

Global Environment

Data

Arthritis	84 obs. of 5 variables
aSAH	113 obs. of 7 variables
DATA	3196 obs. of 33 variables
data	num [1:50, 1:8] 3615 365 2212 2110 ...
DATA5.9	1377 obs. of 37 variables
detg	24 obs. of 5 variables
df	1 obs. of 2 variables
examdata	32 obs. of 11 variables
examdatacars	32 obs. of 11 variables
hot410mon	2354 obs. of 34 variables
labortest1	100 obs. of 31 variables
...	...

```
1 setwd("E:/Rway")
2 install.packages("surveillance")
3 library(leaflet)
4
5 ??leaflet
6 help(leaflet)
7 library(leaflet)
8 m = leaflet() %>% addTiles()
9 m # a map with the default OSM tile layer
10
11 # set bounds
12 m %>% fitBounds(-40, -18, -41, -20)
13 ??fitBounds
14 # move the center to Snedecor Hall
15
```

1:1 (Top Level) R Script

Console ~/

Type 'license()' or 'licence()' for distribution details.

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.

Files Plots Packages Help Viewer

R: Add a raster image as a layer Find in Topic

addRasterImage {leaflet}

R Documentation

Add a raster image as a layer

例 1 在 $[-1, 2]$ 上画 $y = e^{2x} + \sin(3x^2)$ 的图形

数学函数

abs , sqrt : 绝对值, 平方根 log, log10, log2 , exp
: 对数与指数函数 sin , cos , tan , asin , acos
, atan , atan2 : 三角函数
sinh , cosh , tanh , asinh , acosh , atanh : 双
曲函数

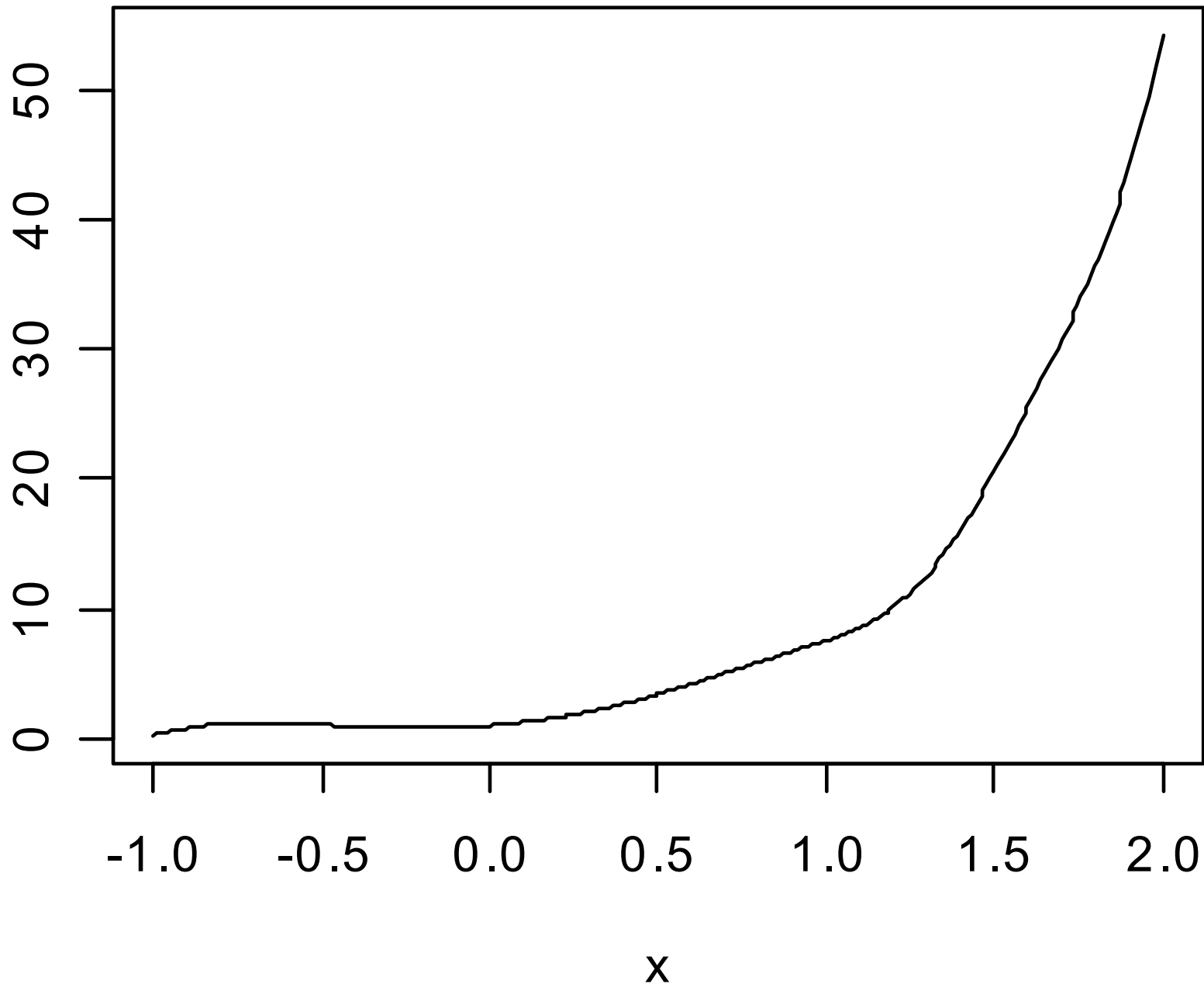
```
> x1 <- 100:200
```

```
> x <- x1/100
```

```
> y <- exp(2*x)+sin(3*x*x)
```

```
> plot(x,y, type='l')
```

y



简单统计量

sum, mean, var(方差), sd (标准差) , min, max, range,

Median (中位数) , IQR (四分位间距) 等为统计量,

sort , order , rank 与排序有关,

其它还有 ave (平均值) , fivenum , mad , quantile (分位数) ,

stem 等。

个人应用例子

广州市炎热指数对早产滞后效应的时间序列分析

徐昊立, 武丽, 吴云涛, 李玉萍, 缪华章, 马远珠, 梁志江, 李兵, 赵庆国

广东省妇幼保健院保健部, 广东 广州 511442

摘要:目的 了解广州市炎热指数对早产的滞后效应。方法 以广州市白云区、越秀区 2004 年 1 月 1 日至 2011 年 12 月 31 日间分娩的产妇为研究对象, 符合纳入标准的共 293 849 人, 其中发生早产者 22702 例。收集该时期产妇分娩数据、气象资料和空气污染物数据。采用分布滞后非线性模型, 以广州市年均炎热指数为参照, 建立炎热指数在滞后 0~20 d 内对逐日早产例数的回归模型, 分析产妇暴露于不同炎热指数后 0~20 d 内发生早产的相对危险度。结果 研究期间广州市白云区、越秀区的早产发生率为 7.70%; 炎热指数波动区间为 45~136 °F, 年均炎热指数为 (101±20) °F, 以炎热指数 101 °F 为参照, 炎热指数升高 20 °F 时, 滞后第 5~7 日的早产发生危险性升高, 其中第 6 日 RR 值最高 (1.014, 95%CI: 1.005~1.023); 第 10~12 日、19 日的早产发生危险性降低, 其中第 19 日 RR 值最低 (0.985, 95%CI: 0.974~0.997)。炎热指数降低 20 °F 时, 滞后第 5~7 日的早产发生危险性降低, 其中第 6 日 RR 值最低 (0.986, 95%CI: 0.978~0.995); 第 10~12 日、19 日的早产发生危险性升高, 其中第 19 日 RR 值最高 (1.015, 95%CI: 1.003~1.027)。结论 广州市炎热指数对早产的发生呈非线性滞后影响。炎热指数升高时, 早产相对危险度先增加后降低; 炎热指数降低时, 早产相对危险度先降低后增加。

关键词: 炎热指数; 早产; 滞后效应; 时间序列研究

中图分类号: R181.3 文献标志码: A 文章编号: 1001-5914(2015)10-0000-00

he lag effect of heat index on preterm birth in Guangzhou; a time-series study

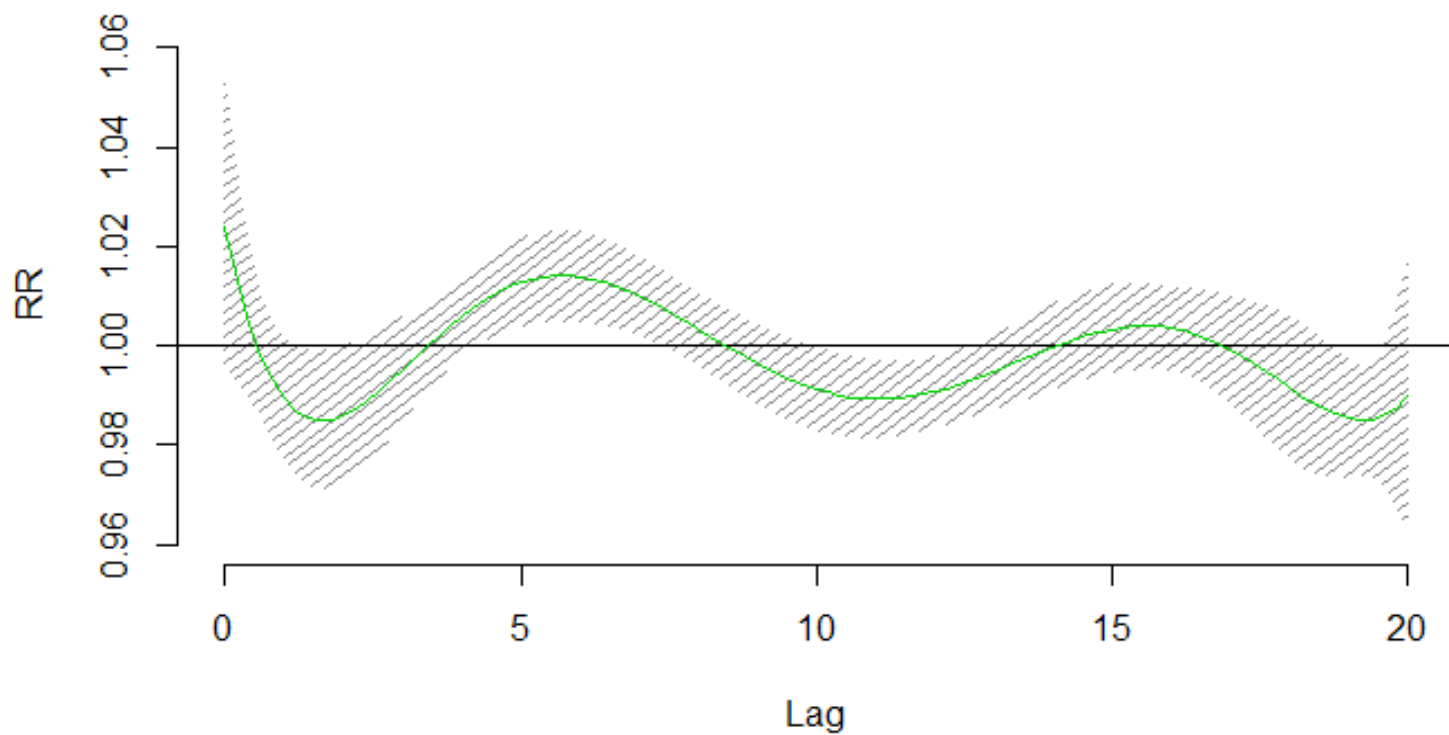
XU Hao-li, WU Li, WU Yun-tao, LI Yu-ping, MIAO Hua-zhang, MA Yuan-zhu, LIANG Zhi-jiang, LI Bing, ZHAO Qing-guo
Healthcare Department of Guangdong Women and Children Hospital, Guangzhou, Guangdong 511442, China

Corresponding author: ZHAO Qing-guo, E-mail: zqgfrost@126.com

Abstract: Objective To estimate the effects of Heat Index on preterm birth in Guangzhou. Methods Daily preterm births, meteorological and air pollution data during 2004–2011 used to explore the association between Heat Index and preterm birth. Distributed lag non-linear time series model was fitted for access the lag effects of each heat index value on preterm birth, based on the poisson regression. Results The exposure-response functions appeared to be lag-non-linear. During 2004–1–1 to 2011–12–31, preterm birth rate was 7.70%, mean Heat Index was (101±20) °F. For the exposure-response effects from lag0 to lag20 with Heat Index reference at 101 (°F), with a 20 unit increase in heat index, the preterm birth risks are different with statistical significant in periods lag 5 day to lag7 day lag10 day to lag12 day and lag19 day the RR value of lag6 day and

个人应用例子

Association with a 20-unit increase in heat index



讨论阶段

1. 为什么要学 R?

关键词：优化、开放性、精确、生物统计，
数据积累到应用、找工作。

2. 分享 3 个身边的故事（暂保密）