

# 基于R软件的传染病数据可视化及 统计分析平台

张兵

邮箱: zhangbing4502431@outlook.com

主页: <http://spatial-r.github.io/>

浙江省疾病预防控制中心免疫所

# Outlines

## 平台背景

传染病简介

传染病监测

免疫规划

## 平台简介

## 平台构建

## 资料推荐

传染病统计分析平台

张兵

### 2 平台背景

传染病简介

传染病监测

免疫规划

平台简介

平台构建

传染病数据获取和整理

传染病可视化平台

传染病统计分析平台

时间序列分析

空间统计分析

时空统计分析

传染病动力学模型

疫苗接种评估

干预效果评估

资料推荐

# 传染病简介

- ▶ 传染病(Communicable Diseases)是由病原微生物(如：细菌、病毒、寄生虫或真菌)所引起的能够在人与人之间直接或间接传播的疾病，可以认为是感染症中的一种。
- ▶ 法定传染病是指《中华人民共和国传染病防治法》中所规定的传染病，包括甲类(鼠疫和霍乱)；乙类(SARS、艾滋病、病毒性肝炎、脊髓灰质炎、人感染高致病性禽流感、麻疹、流行性出血热、狂犬病、流行性乙型脑炎、登革热、炭疽、细菌性痢疾、肺结核、伤寒副伤寒、流行性脑脊髓膜炎、百日咳、白喉、猩红热、布鲁氏菌病、淋病、梅毒、新生儿破伤风、钩端螺旋体病、血吸虫病、疟疾和甲型H1N1流行性感冒)；丙类(流行性感冒、流行性腮腺炎、风疹、急性出血性结膜炎、麻风、流行性或地方性斑疹伤寒、黑热病、包虫病、丝虫病、感染性腹泻和手足口病)。

传染病统计分析平台

张兵

平台背景

3 传染病简介

传染病监测

免疫规划

平台简介

平台构建

传染病数据获取和整理

传染病可视化平台

传染病统计分析平台

时间序列分析

空间统计分析

时空统计分析

传染病动力学模型

疫苗接种评估

干预效果评估

资料推荐

# 传染病历史

人类健康的发展史就是一部与传染病的斗争史：

- ▶ 公元前5世纪的雅典瘟疫，“人们像羊群一样地死亡着。病人裸着身体在街上游荡，寻找水喝直到倒地而死。存活下来的人不是没了指头、脚趾、眼睛，就是丧失了记忆。”
- ▶ 14世纪，黑死病席转了整个欧洲，造成欧洲1/3人口死亡，包括西班牙国王阿尔方索十一世。
- ▶ 16-18世纪，天花开始流行于世界，被称为世界上最可怕的疾病，仅在20世纪就造成了多达5亿人死亡。
- ▶ 1918年世界性流感暴发，疫情持续了近一年，造成了约2100万人口死亡。
- ▶ 1992年，由O139霍乱弧菌引起的新型霍乱席卷了印度和孟加拉国等地区，疫情至今未完全平息。

传染病统计分析平台

张兵

平台背景

4 传染病简介

传染病监测

免疫规划

平台简介

平台构建

传染病数据获取和整理

传染病可视化平台

传染病统计分析平台

时间序列分析

空间统计分析

时空统计分析

传染病动力学模型

疫苗接种评估

干预效果评估

资料推荐

# 传染病监测

传染病监测(Infectious disease surveillance)是指对特定环境、人群进行流行病学、血清学、病原学、临床症状以及其它有关影响因素的调查研究，来预测传染病发生、发展和流行规律，提出检疫措施并评价预防效果。

- ▶ 1950年我国建立了传染病报告系统，要求报告的病种有15种。
- ▶ 1987年全国疫情统计数据进行点对点联网，实现疫情报告的计算机系统化管理，各区县收集逐级汇总至省级，再通过全国疾病计算机网络系统上传至中国预防医学科学院。
- ▶ 1994年建立“传染病疫情统计报告信息系统”，基于国家公用数据分组交换网实行统计数据网络报告，结束手工报表的历史。
- ▶ 2004年建立“传染病与突发公共卫生事件报告系统”，全面实现传染病“在线、个案、直报”的监测信息系统。

传染病统计分析平台

张兵

平台背景

传染病简介

5

传染病监测

免疫规划

平台简介

平台构建

传染病数据获取和整理

传染病可视化平台

传染病统计分析平台

时间序列分析

空间统计分析

时空统计分析

传染病动力学模型

疫苗接种评估

干预效果评估

资料推荐

## 传染病监测数据特点:

- ▶ 时间属性: 出生日期、发病日期、诊断日期。
- ▶ 空间属性: 住址信息, 点或者汇总成区域信息。
- ▶ 其他信息: 如年龄、性别、疫苗接种情况、人群分类等

数据非独立, 存在: 时间自相关、空间自相关和空间异质性。

父	15669125	#####	20岁	34162319930916	浙江省宁波市江北区	33020307	本市其它区	散居儿童	2013/12/10	2013/12/13	
男	13586801	#####	8月		浙江省宁波市	33022562	本市其它区	散居儿童	2013/12/11	2013/12/20	
女	15867510	#####	7月		浙江省宁波市	33021201	本市其它区	散居儿童	2013/12/7	2013/12/20	
女	18368275	#####	8月		浙江省宁波市	33020503	本县	散居儿童	2013/12/9	2013/12/23	
女	13819166	2012/7/6	17月	无	浙江省杭州	33012731	本县	散居儿童	2013/12/8	2013/12/12	
男	13567905	#####	35岁		浙江省宁波市	33021103	本市其它区	家政、家政	2013/12/5	2013/12/23	
女	65003879	#####	40岁		浙江省宁波市	33022502	本县	农民	2013/12/4	2013/12/17	
女	13968213	#####	41岁	' 33022219720119	浙江省宁波市	33028239	本市其它区	家政、家政	2013/11/27	2013/11/30	
女	13968395	#####	39岁		浙江省宁波市	33022501	本县	农民	2013/11/23	2013/11/28	
男	15967317	#####	9月		浙江省嘉兴	33041135	本市其它区	散居儿童	2013/11/22	2013/12/5	
女	13566505	1969/2/5	44岁		象山县	浙江省宁波市	33022501	本县	餐饮食品	2013/11/13	2013/11/20
女	15958223	#####	7月		浙江省宁波市	33028204	本市其它区	散居儿童	2013/11/8	2013/11/13	
女	13566633	#####	30岁	' 33100319	英菲尼	浙江省宁波市	33020307	本市其它区	商业服务	2013/11/3	2013/11/15
男	13586845	#####	21月	无	浙江省宁波市	33022532	本县	散居儿童	2013/11/7	2013/11/13	
男	13967515	#####	3岁		浙江省绍兴	33068303	本市其它区	散居儿童	2013/11/3	2013/11/14	
男		#####	43岁	无	浙江省宁波	33022565	本县	农民	2013/10/30	2013/11/7	
男	13857470	#####	41岁	' 33021115	镇海发	浙江省宁波	33021101	本县	干部职员	2013/10/28	2013/11/7
男	18857871	#####	8月		浙江省杭州	33011102	本县	散居儿童	2013/10/21	2013/10/26	
女	13735395	#####	59岁		浙江省绍兴	33068332	本县	农民	2013/10/18	2013/11/5	
男	15157332	#####	8月		浙江省嘉兴	33048302	本市其它区	散居儿童	2013/10/17	2013/10/24	

免疫规划(Immunization program)是指按照国家或者省、自治区、直辖市确定的疫苗品种、免疫程序或者接种方案，在人群中有计划地进行预防接种，以预防和控制特定传染病的发生和流行。

- ▶ 疫苗接种是预防和控制疫苗针对传染病最有效的方法。
- ▶ 按照科学的免疫程序接种，才能充分发挥疫苗的免疫效果。
- ▶ 免疫程序中的疫苗接种**起始时间**、**接种剂量**、**接种针次**、**接种间隔**、**接种途径**、**不同疫苗同时接种**、**加强免疫**等都与免疫效果有一定关系。
- ▶ 制定免疫程序时需考虑：**传染病的流行特征**、**疫苗的生物**  
**学特性和免疫效果**、**实施条件**。

# Outlines

平台背景

平台简介

平台构建

资料推荐

传染病统计分析平台

张兵

平台背景

传染病简介

传染病监测

免疫规划

8 平台简介

平台构建

传染病数据获取和整理

传染病可视化平台

传染病统计分析平台

时间序列分析

空间统计分析

时空统计分析

传染病动力学模型

疫苗接种评估

干预效果评估

资料推荐



# 必要性和可行性

## 必要性:

- ▶ 传染病人類已经成功消灭了天花，控制了脊髓灰质炎，但传染病防治形势依然严峻。
- ▶ 深层次地发掘传染病数据内在信息，更有利于采取相应措施降低传染病危害。

## 可行性:

- ▶ 传染病监测数据有其**固定的格式**，因而从理论上而言，搭建平台来处理和分析传染病数据是可行的。
- ▶ R软件中包含了众多可用于传染病数据处理和统计分析的程序包，可为平台的核心分析模块提高支持。
- ▶ 基于R软件中的shiny程序包，可快速打造数据产品。

传染病统计分析平台

张兵

平台背景

传染病简介

传染病监测

免疫规划

9 平台简介

平台构建

传染病数据获取和整理

传染病可视化平台

传染病统计分析平台

时间序列分析

空间统计分析

时空统计分析

传染病动力学模型

疫苗接种评估

干预效果评估

资料推荐

平台的搭建，是为使得数据处理和分析更**简易**且有**条理**。基于R软件中已有程序包来构建传染病数据可视化及统计分析平台，主要实现如下功能：

- ▶ 公共平台上传染病共享数据的获取
- ▶ 传染病数据的清理和整理
- ▶ 传染病数据静态和交互式可视化
- ▶ 描述传染病数据在时间、空间和人群的分布规律
- ▶ 探索传染病蔓延的危险因素，建立暴发预警机制
- ▶ 揭示传染病可能的高危区域和高危人群
- ▶ 评估不同干预措施对传染病的影响，筛选最优防治策略

# Outlines

平台背景

平台简介

平台构建

传染病数据获取和整理

传染病数据可视化平台

传染病统计分析平台

资料推荐

传染病统计分析平台

张兵

平台背景

传染病简介

传染病监测

免疫规划

平台简介

11 平台构建

传染病数据获取和整理

传染病可视化平台

传染病统计分析平台

时间序列分析

空间统计分析

时空统计分析

传染病动力学模型

疫苗接种评估

干预效果评估

资料推荐

# 传染病数据获取和整理

传染病统计分析平台

张兵

平台背景

传染病简介

传染病监测

免疫规划

平台简介

平台构建

12 传染病数据获取和整理

传染病可视化平台

传染病统计分析平台

时间序列分析

空间统计分析

时空统计分析

传染病动力学模型

疫苗接种评估

干预效果评估

资料推荐

## 传染病数据获取:

- ▶ 中国传染病监测信息系统(2004至今): 直接导出
- ▶ 美国流感周监测数据(2014至今): Package `cdcfluview`
- ▶ <http://www.healthdata.gov/>: Package `rHealthDataGov`
- ▶ 美国法定传染病周报告数据( [Project Tycho](#)): 直接导出

## 传染病数据整理:

- ▶ 已清理过的数据利用Package `dplyr`进行再度整合。
- ▶ 未清理过的数据(中国传染病监测信息系统)需利用Package `stringi`、`dplyr`和`reshape2`等进行清理整合。

# 传染病数据可视化平台: 静态图

静态图(如散点图、直方图、箱式图、密度函数图、时序图、地图)的实现基于Packag **ggplot2**, 通过定义如下要素构建图形:

- ▶ 数据(Data)和映射(Mapping)
- ▶ 几何对象(Geometric)
- ▶ 坐标系统(Coordinante)
- ▶ 统计变换(Statistics)
- ▶ 标尺(Scale)
- ▶ 图层(Layer)
- ▶ 分面(Facet)
- ▶ 主题(Theme)

其他扩展程序包如:

- ▶ **cowplot**: 提供了部分学术主题
- ▶ **directlabels**: 标签设置
- ▶ **ggdendro**: 树状图
- ▶ **ggExtra**: 底图的边缘增加直方图、密度函数图和箱式图
- ▶ **ggsn**: 提供指南针和比例尺

传染病统计分析平台

张兵

平台背景

传染病简介

传染病监测

免疫规划

平台简介

平台构建

传染病数据获取和整理

13

传染病可视化平台

传染病统计分析平台

时间序列分析

空间统计分析

时空统计分析

传染病动力学模型

疫苗接种评估

干预效果评估

资料推荐

# 传染病数据可视化平台：交互式图形

交互式传染病可视化平台主要侧重传染病数据在时间、空间和人群三个维度上的可视化。

## ► 时间维度：

**dygraphs**和**rCharts**：前者基于dygraphs JavaScript，后者基于Highcharts，NVD3，Polychart等JavaScript。

## ► 空间维度：

**rleafmap**和**leafletR**：基于leaflet JavaScript。

## ► 时空维度：

**surveillance**和 **animation**：点时空数据和区域时空数据可视化。

## ► 人群维度：

**rCharts**：不同性别、年龄和干预措施下传染病的分布。

传染病统计分析平台

张兵

平台背景

传染病简介

传染病监测

免疫规划

平台简介

平台构建

传染病数据获取和整理

14 传染病可视化平台

传染病统计分析平台

时间序列分析

空间统计分析

时空统计分析

传染病动力学模型

疫苗接种评估

干预效果评估

资料推荐

# 时间序列分析

传染病数据的时间序列分析主要集中于疫情预测(Prospective)和暴发探测(Retrospective)，但都需考虑传染病数据携带的长期趋势、周期性、季节性和短期波动等现象。

## 疫情预测

- ▶ **forecast**: 提供了包括指数平滑处理在内的多种预测方法。
- ▶ **bsts**: 贝叶斯状态空间模型。
- ▶ **tscount**: 基于广义线性模型进行拟合及干预效果评估。

## 暴发探测

- ▶ **surveillance**: 单个异常点和异常时段的暴发探测，既适合单  
时序数据也包含多分类时序数据暴发探测方法。
- ▶ **changepoint**和**ecp**: 常规时序数据中异常值的探测。

传染病统计分析平台

张兵

平台背景

传染病简介

传染病监测

免疫规划

平台简介

平台构建

传染病数据获取和整理

传染病可视化平台

传染病统计分析平台

15 时间序列分析

空间统计分析

时空统计分析

传染病动力学模型

疫苗接种评估

干预效果评估

资料推荐

# 空间序列分析: 点格局

传染病数据的空间点格局分析主要探讨传染病发病个案在空间层面的分布特征, 发现热点区域, 可分为全局和局部探测。

- ▶ **spatstat**: 自由定义研究区域, 分类点过程和空间协变量
- ▶ **spatialkernel**: 多元点过程分析中不同核函数和方法。
- ▶ **ads**: 一阶和二阶多尺度空间点格局分析(Ripley K)。
- ▶ **DCluster**: 传染病疾病的热点探测(Moran's I, Geary's C和Beasg-newell)
- ▶ **SpatialEpi**: 在DCluster程序包的基础上加上了贝叶斯热点探测和分层探测的统计方法。
- ▶ **SPODT**: 空间斜分类回归树来定义空间分布格局。

传染病统计分析平台

张兵

平台背景

传染病简介

传染病监测

免疫规划

平台简介

平台构建

传染病数据获取和整理

传染病可视化平台

传染病统计分析平台

时间序列分析

16 空间统计分析

时空统计分析

传染病动力学模型

疫苗接种评估

干预效果评估

资料推荐



# 空间序列分析: 回归模型

空间回归模型是在考虑空间效应的基础上来研究自变量与因变量之间的关系，或同时考虑事件生存时间的影响。

- ▶ **spdep**: 空间滞后模型、空间误差模型、空间杜宾模型等。
- ▶ **gwr**和 **GWmodel**: 地理加权回归(主成分分析、判别分析)。
- ▶ **McSpatial**: 非参空间模型如空间局部加权回归、空间线性logit和probit模型及空间分位数回归等。
- ▶ **CARBayes**: 贝叶斯分层空间模型(二项、高斯或泊松分布)
- ▶ **spatcounts**: 条件自回归模型来处理区域计数数据。
- ▶ **spatsurv**和**spBayesSurv**: 空间生存分析。
- ▶ **splm**: 最大似然比和广义矩阵法拟合空间面板模型。

传染病统计分析平台

张兵

平台背景

传染病简介

传染病监测

免疫规划

平台简介

平台构建

传染病数据获取和整理

传染病可视化平台

传染病统计分析平台

时间序列分析

17 空间统计分析

时空统计分析

传染病动力学模型

疫苗接种评估

干预效果评估

资料推荐

# 时空统计分析

从**时间轴**看，在揭示传染病发病随时间变化的同时，也揭示其在每个时点上发病的空间差异。

从**空间轴**看，在揭示传染病发病随区域变化的同时，也揭示其在每个区域上发病的时间差异。

- ▶ **lgcp**: log-Gaussian Cox process拟合时空数据。
- ▶ **INLA**: 基于Integrated nested Laplace approximation对隐高斯随机场的统计推断。
- ▶ **surveillance**: 时空多层分模型(双成分应用于点数据，三成分模型应用于区域数据)。
- ▶ **stppResid**: 时空点数据模型的残差分析。
- ▶ **mwa**: 基于滑动窗口和统计匹配的方法进行时空点数据的归因分析。

传染病统计分析平台

张兵

平台背景

传染病简介

传染病监测

免疫规划

平台简介

平台构建

传染病数据获取和整理

传染病可视化平台

传染病统计分析平台

时间序列分析

空间统计分析

18 时空统计分析

传染病动力学模型

疫苗接种评估

干预效果评估

资料推荐

# 传染病动力学模型

传染病动力学模型通过传播机制将目标人群分为若干状态并定义不同状态间的转变特性，可从理论角度定量模拟传染病的演变规律。

- ▶ **EpiModel**: 确定性和随机性传染病动力学网络模型。
- ▶ **EpiDynamics**: 《Modeling Infectious Diseases in Humans and Animals》一书中介绍的传染病动力学模型。
- ▶ **pomp**: 传染病动力学模型和模拟仿真技术相结合。
- ▶ **deSolve**: 常微分、偏微分和时滞微分方程求解。
- ▶ **R0**: 基本再生数和实时再生数(基于罹患率、指数增长、最大似然、时间依赖和贝叶斯序贯等方法)。
- ▶ **EpiEstim**: 采用参数、非参和不确定方法定义序列间隔(series interval)来计算不同时间点再生数。

传染病统计分析平台

张兵

平台背景

传染病简介

传染病监测

免疫规划

平台简介

平台构建

传染病数据获取和整理

传染病可视化平台

传染病统计分析平台

时间序列分析

空间统计分析

时空统计分析

19 传染病动力学模型

疫苗接种评估

干预效果评估

资料推荐

# 疫苗接种评估

传染病统计分析平台

张兵

平台背景

传染病简介

传染病监测

免疫规划

平台简介

平台构建

传染病数据获取和整理

传染病可视化平台

传染病统计分析平台

时间序列分析

空间统计分析

时空统计分析

传染病动力学模型

20 疫苗接种评估

干预效果评估

资料推荐

R软件中与疫苗相关的已有程序包如下：

- ▶ **hbim**: 基于Hill/Bliss Independence模型计算联合疫苗的相对危险度和保护率。
- ▶ **sglr**: 序贯广义似然比检验疫苗临床试验不良反应发生率。
- ▶ **vacem**: 接种率调查数据及疫苗接种常规数据来估算疫苗接种率。

总体而言，R软件中与疫苗相关的程序包较少，开发相关程序包用以客观评估疫苗接种对疫苗针对传染病的防控效果，可以较好地指导卫生部门调整防控策略。

# 其他与传染病相关的程序包

- ▶ BCEA: 贝叶斯成本-效益分析。
- ▶ BCEs0: 当存在结构零成本时的贝叶斯成本-效益模型。
- ▶ DALY: 计算随机性伤残调整寿命年。
- ▶ seroincidence: 基于横断面抗体水平计算传染病感染率。
- ▶ EpiContactTrace: 传染病接触历程追踪。
- ▶ OutbreakTools: 传染病暴发分析。
- ▶ outbreaker: 根据流行病数据和基因数据重构传染病暴发。

传染病统计分析平台

张兵

平台背景

传染病简介

传染病监测

免疫规划

平台简介

平台构建

传染病数据获取和整理

传染病可视化平台

传染病统计分析平台

时间序列分析

空间统计分析

时空统计分析

传染病动力学模型

21

疫苗接种评估

干预效果评估

资料推荐

26

浙江省疾病预防控制中心

# 干预效果评估

一般而言，针对传染病的干预措施主要集中在三个方面：**传染源**、**传播途径**和**易感人群**，具体而言体现在疫苗接种（常规接种和应急接种）、虫媒密度处理、特殊人群干预等。

## 评价方法:

1. 基于seroincidence获取传染病的传播力系数。
2. 基于pomp获取传染病动力学模型中未知参数的后验分布。
3. 基于pomp重新构建传染病动力学模型预测未采取干预措施后传染病发病情况。
4. 与传染病实际情况对比，得到干预措施后的个案数改变量、伤残调整寿命年(DALY)、成本效益(BCEA)、基本可再生数( $R_0$ )。

传染病统计分析平台

张兵

平台背景

传染病简介

传染病监测

免疫规划

平台简介

平台构建

传染病数据获取和整理

传染病可视化平台

传染病统计分析平台

时间序列分析

空间统计分析

时空统计分析

传染病动力学模型

疫苗接种评估

22 干预效果评估

资料推荐

26 浙江省疾病预防控制中心

# Outlines

平台背景

平台简介

平台构建

资料推荐

传染病统计分析平台

张兵

平台背景

传染病简介

传染病监测

免疫规划

平台简介

平台构建

传染病数据获取和整理

传染病可视化平台

传染病统计分析平台

时间序列分析

空间统计分析

时空统计分析

传染病动力学模型

疫苗接种评估

干预效果评估

23 资料推荐

26 浙江省疾病预防控制中心

# 书籍推荐：传染病介绍

- ▶ 《大流感》：作者依据大量的历史资料重绘了1918年-1919年横扫世界的流感疫情发生、发展及其肆虐全球的过程以及当时医学工作者在巨大压力下表现出的勇气、信仰和研究态度。
- ▶ 《目击小汤山》：小汤山医院，作为世界第一所治疗SARS的专科医院，本文再现了小汤山医院30余个不平凡日夜发生的一段段感人故事。
- ▶ 《鼠疫》：描述北非一个奥兰的城市突发鼠疫后，以主人公里厄医生为代表的一群人面对瘟疫奋力抗争的故事。该书作者写的《霍乱时期的爱情》曾获诺贝尔文学奖。
- ▶ 《国土无双伍连德》：伍连德是我国卫生防疫事业的奠基人，成功扑灭了1911年东三省得大鼠疫流行。

传染病统计分析平台

张兵

平台背景

传染病简介

传染病监测

免疫规划

平台简介

平台构建

传染病数据获取和整理

传染病可视化平台

传染病统计分析平台

时间序列分析

空间统计分析

时空统计分析

传染病动力学模型

疫苗接种评估

干预效果评估

24 资料推荐

26 浙江省疾病预防控制中心



# 书籍推荐：传染病分析

## 空间流行病学分析：

- ▶ Applied Spatial Data Analysis with R
- ▶ Applied Spatial Statistics for Public Health Data
- ▶ Spatial Epidemiology Notes
- ▶ Case Studies in Spatial Point Process Modeling
- ▶ 传染病时空聚集性探测与预测预警方法

## 传染病动力学模型：

- ▶ Modeling Infectious Disease Parameters Based on Serological and Social Contact Data
- ▶ Modeling Infectious Diseases in Humans and Animals

传染病统计分析平台

张兵

平台背景

传染病简介

传染病监测

免疫规划

平台简介

平台构建

传染病数据获取和整理

传染病可视化平台

传染病统计分析平台

时间序列分析

空间统计分析

时空统计分析

传染病动力学模型

疫苗接种评估

干预效果评估

25 资料推荐

26 浙江省疾病预防控制中心

# Thanks

希望与志同道合者一起合作，促进R软件  
在公共卫生领域的应用