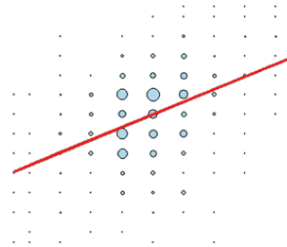
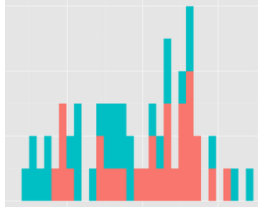


# R语言入门



# 提纲

- 什么是R？
- 为什么选择R？
- 掌握了R有什么用？

# 什么是R？

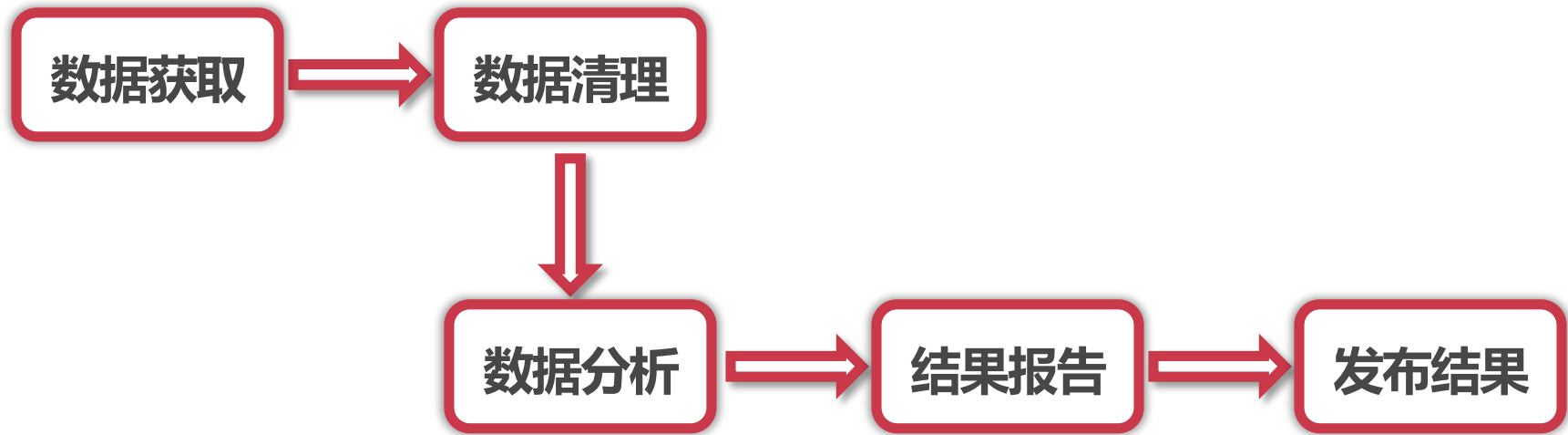
- 最受欢迎的**数据分析和可视化**平台之一
- 首次出现 1993年; 2011年随大数据的爆发而流行起来

# 为什么选择R？

- 免费、支持Windows/Mac OS/Linux
- 开源
  - 有强大的工具包
  - 可以贡献自己的工具包

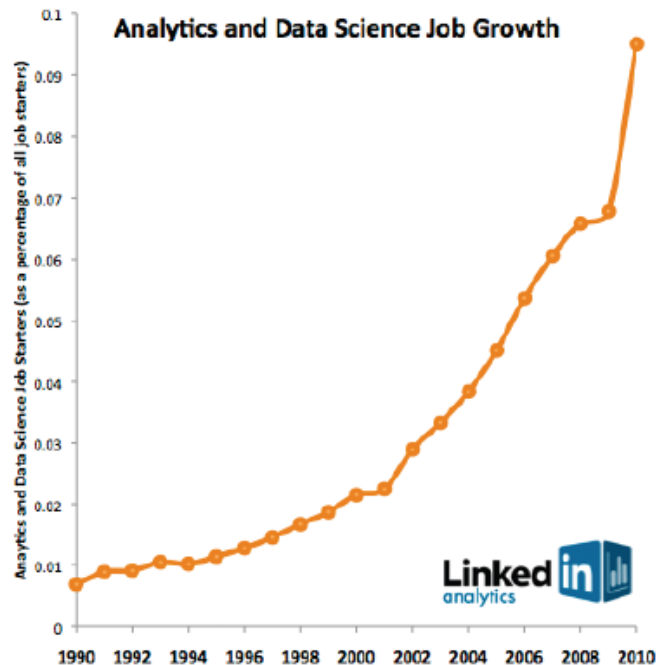
# 为什么选择R？

- 可以完成**数据分析**涉及的几乎所有**步骤**



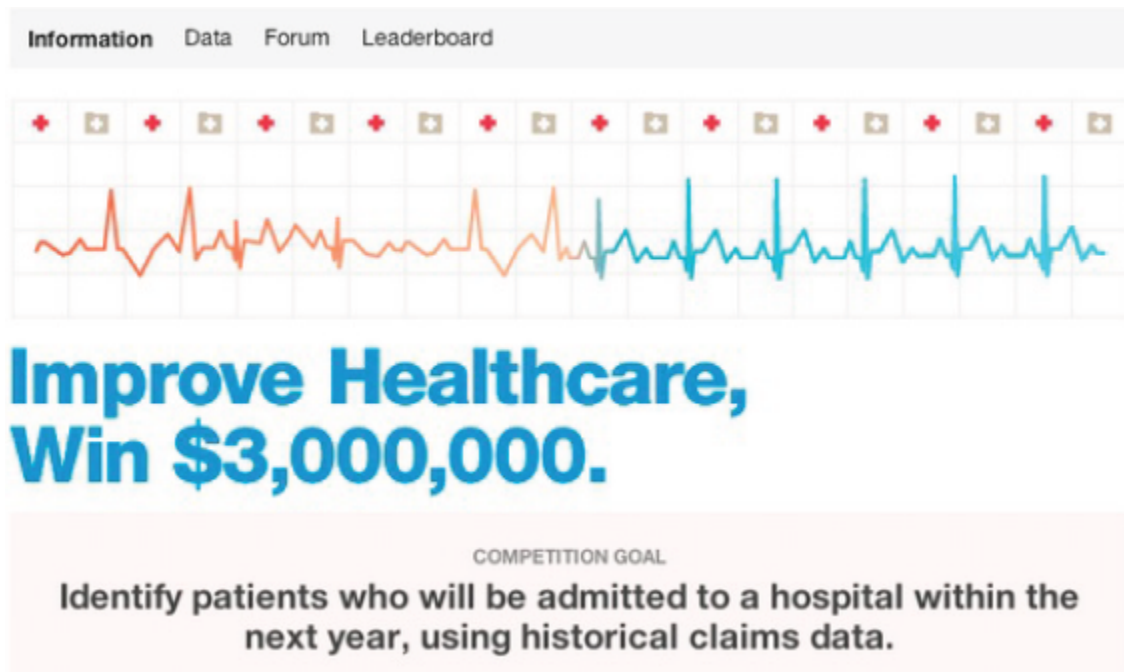
# 掌握了R, 我可以...

- 有钱赚 !!!
  - 3/5
  - 17.5万美元; 20%-30%
- 缺口大 !!!



# 掌握了R, 我可以...

- 有趣味！！！！有价值！！！！





# 提纲

## 数据分析

- 探索性数据分析
- 统计推断
- 回归分析
- 机器学习- 分类问题

## 结果报告

## 发布结果

- 开发数据产品

# 探索性数据分析

- 数据分析中的必要步骤
- 了解数据
- 作图

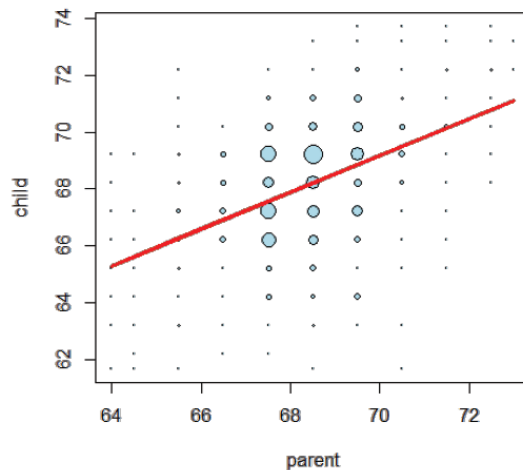
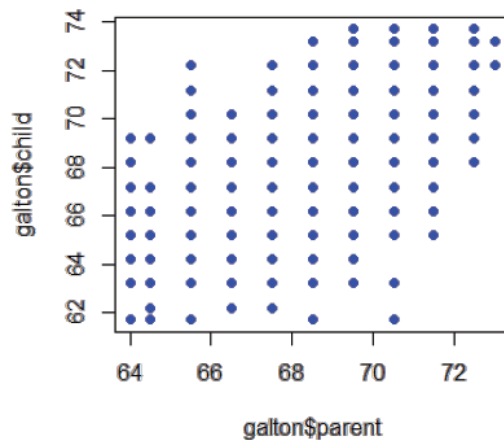
# 统计推断

- **基于数据得出正式结论的过程**
  - **不确定性**
- **选举：候选人A vs. 候选人B 谁会胜出？**
  - **结论 + 结论是错误的概率**
- **药效：药A vs. 药B 谁更有效？**
  - **结论 + 结论是错误的概率**

# 回归分析

- 线性模型拟合数据
  - 预测变量
  - 结果变量

- 预测



# 机器学习- 分类问题

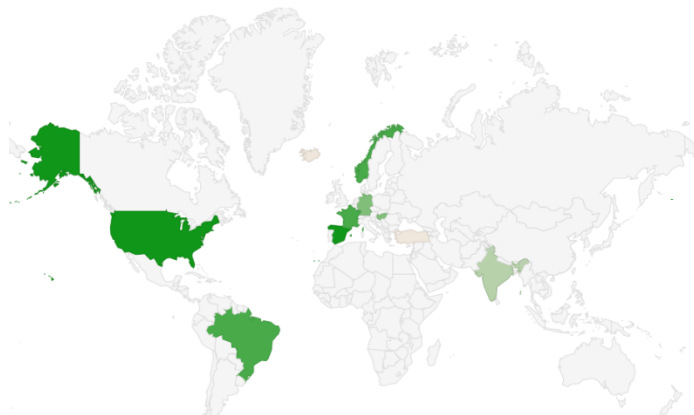
- 训练模型 + 预测
- 分类问题



汽车？

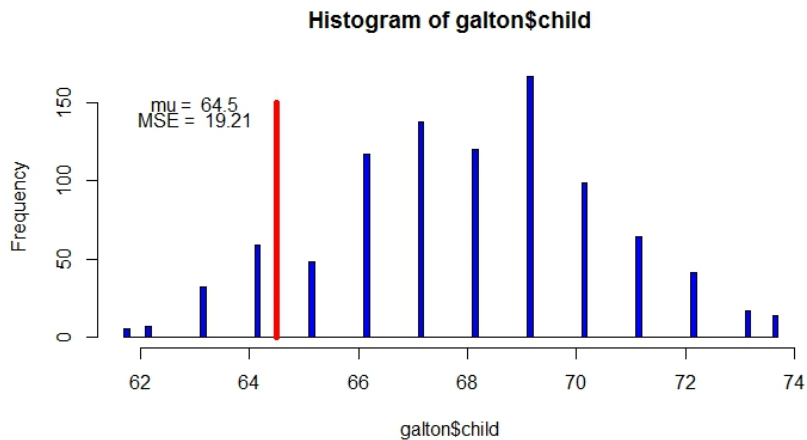
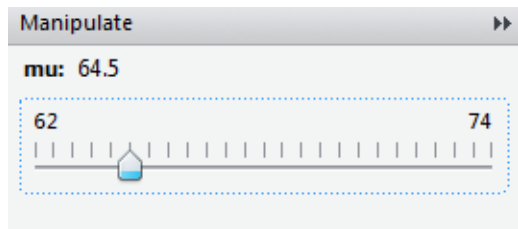
# 开发数据产品

- **GoogleVis API**
  - R制作html, 调用Google charts
  - 交互式html图表



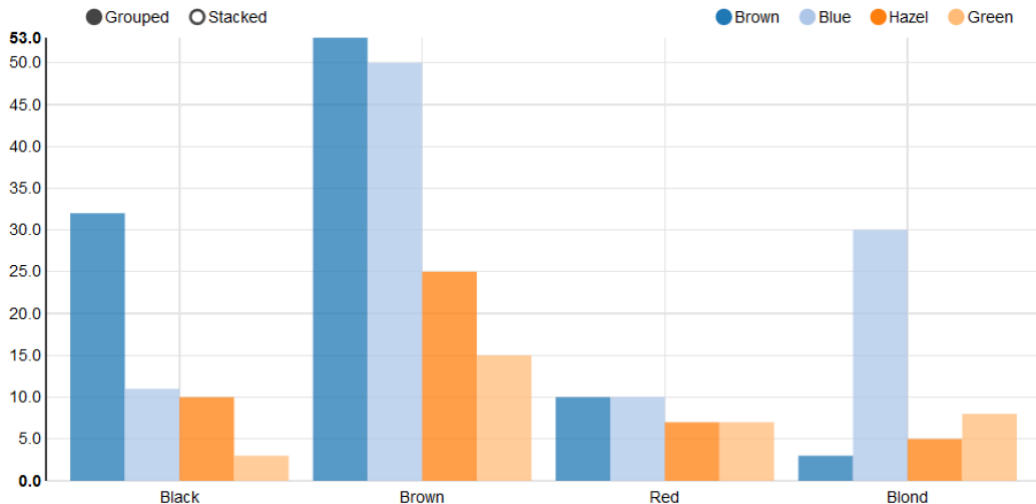
# 开发数据产品

- Manipulate



# 开发数据产品

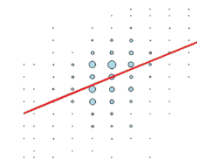
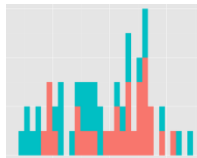
- rCharts
  - 使用R制作交互式javascript可视化产品





# 开发数据产品

- **Shiny**
  - 制作嵌入网页的交互式R程序的平台
- **Slidify**
  - 制作和发布基于R的报告（ppt）



**R**Pubs

# 提纲

**R与Rstudio的获取与安装**

**包**

**如何获取帮助**

# R的获取与安装

- [cran.r-project.org](https://cran.r-project.org)

# RStudio的获取与安装

- 集成开发环境(IDE), R的用户接口
- [rstudio.com](https://rstudio.com)

# 包 (package)

- 扩展R基本功能的机制 / 集成了众多函数
- CRAN / Bioconductor / GitHub 等
  - `install.packages()`
  - `install_github()`

# 获取帮助

- **?函数名 ( R的帮助文档 )**
- **Google / Stackoverflow**
- **如何问问题**
  - **操作系统、版本、哪一步产生的错误、预期是什么、得到的结果是什么、其他有用的信息**
  - **例如 : Win7 R 3.2.0 lm() “seg fault on large data frame”**