



China Government Report ANALYSIS

All BY LucasX



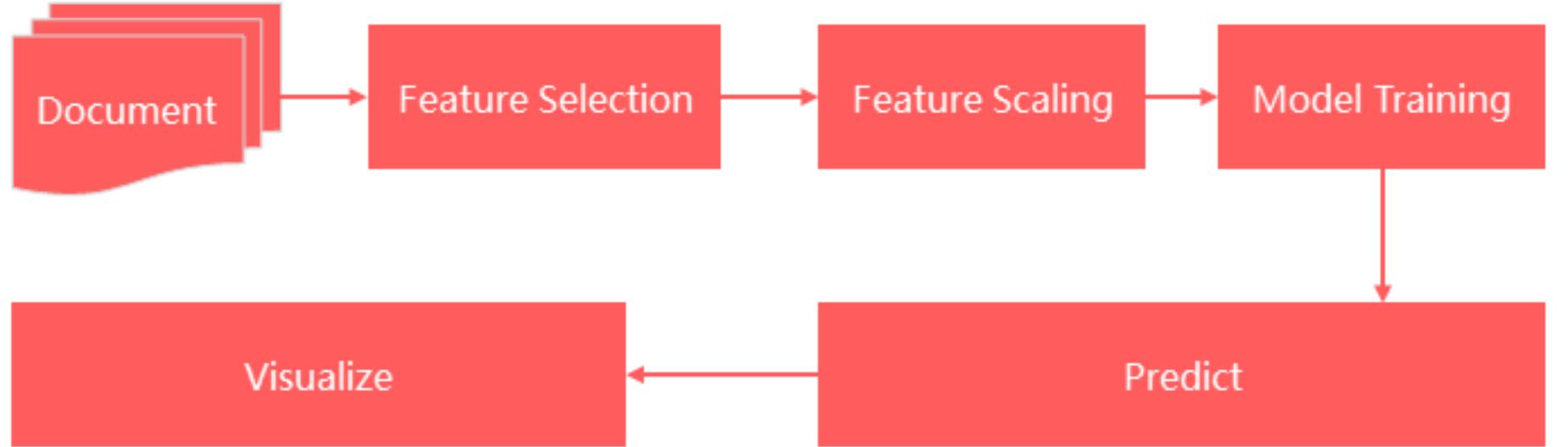
Contents

Part 1 Introduction

Part 2 Theory & Methods

Part 3 Experiment

Part 4 Result



Process of Text Mining



Contents

Part 1 Introduction

Part 2 Theory & Methods

Part 3 Experiment

Part 4 Result



Representation

no curse of dimensionality

suitable for computing

How to represent a word

Contains more information



TF-IDF (term frequency-inverse document frequency) 是一种用于信息检索与数据挖掘的常用加权技术。TF-IDF用以评估一词对于一个文件集或一个语料库中的其中一份文件的重要程度。词的重要性随着它在文件中出现的次数成正比增加，但同时会随着它在语料库中出现的频率成反比下降。

$$TF = \frac{n_{ij}}{\sum_k n_{kj}}$$
$$IDF = \log \frac{|D|}{\{j: t_i \in d_j\}}$$

$$TF - IDF = TF * IDF$$



Contents

Part 1 Introduction

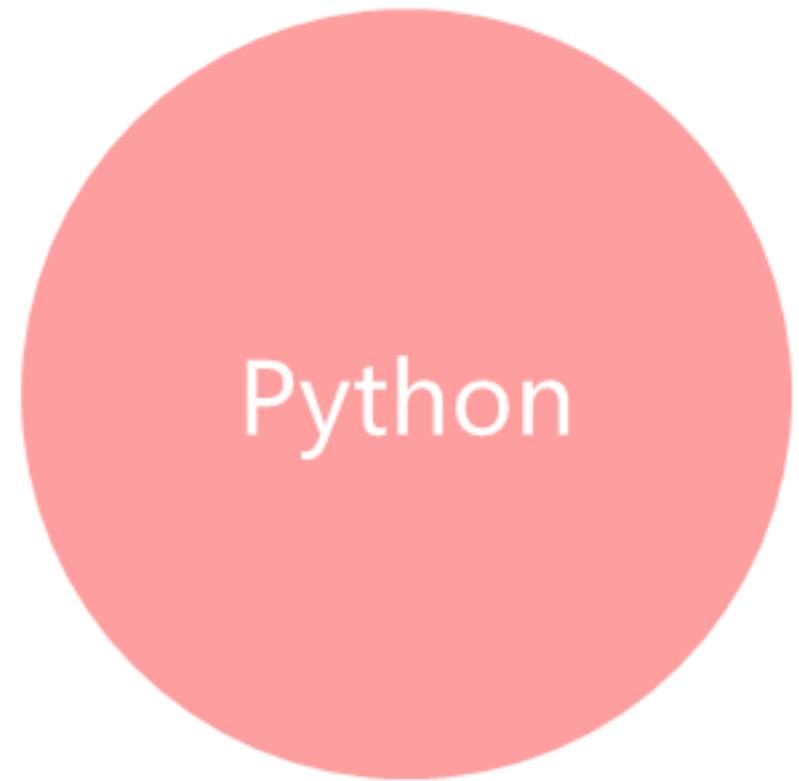
Part 2 Theory & Methods

Part 3 Experiment

Part 4 Result



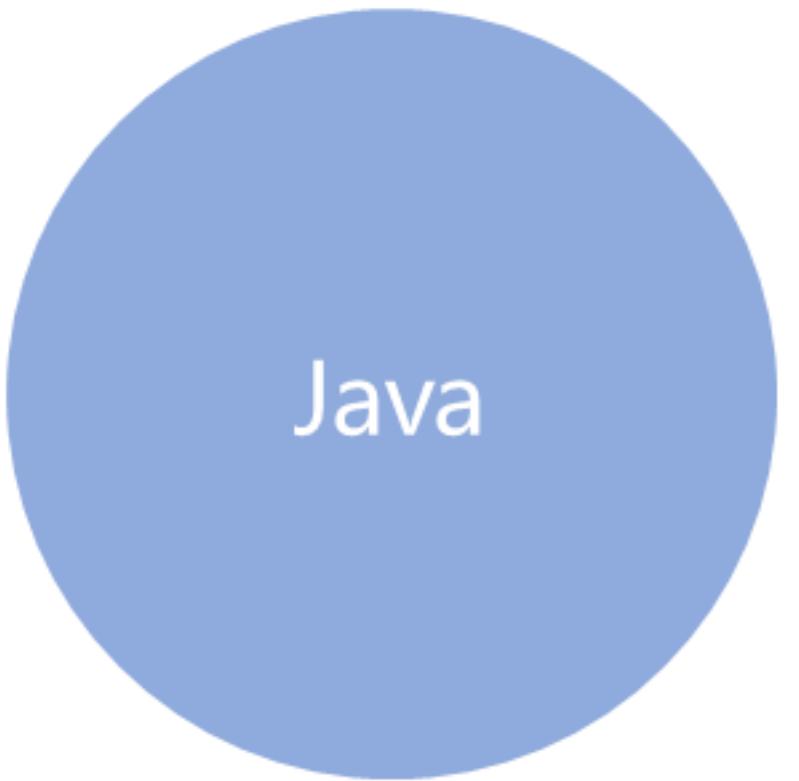
Experiment



Python



Web Spider, NLP toolbox
in Python3.5



Java



Data handling
in Java8



HTML5



Data visualization
in HTML5



Contents

Part 1 Introduction

Part 2 Theory & Methods

Part 3 Experiment

Part 4 Result



执政热度词 TF-IDF

根据不同时代领导人执政时期
的政府工作报告，计算热
度词，生成可视化字符云





毛爷爷时期执政热度词



邓小平时期执政热度词



投资
重视
积极
科技
增加
投入
国民
经济
体系
五年
完善
基础
工作
农村
管理
五人
加强
加大
邓小平
九五
继续
建立
生产
基础设施
人民
财政
中央
社会
推进
整顿
社会主义
群众
提高
合作
大力
鼓励
稳定
支持
经济
文化
实现
制度
地区
干部
粮食
国有
各级同志
改善
关系
解决
逐步
搞好
结构
粮食
国有
国家
资源
产业
世贸组织
国务院
农民
市场
方针
深化
坚持
发挥
开发
社会保障
经济体制
世界
教育
发展
农业
经济体制
特别
方面
建设
基本
生活
切实
问题
促进
企业
国际
规范
现代化
生产
继续
建立
邓小平
九五
精神
文明
我国
事业
地方
增长
增加
投入
国民
经济
体系
五年
完善
基础
工作
农村
管理
五人
加强
加大
邓小平
九五
继续
建立
生产
基础设施
人民
财政
中央
社会
推进
整顿
社会主义
群众
提高
合作
大力
鼓励
稳定
支持
经济
文化
实现
制度
地区
干部
粮食
国有
各级同志
改善
关系
解决
逐步
搞好
结构

江泽民时期执政热度词



胡锦涛时期执政热度词



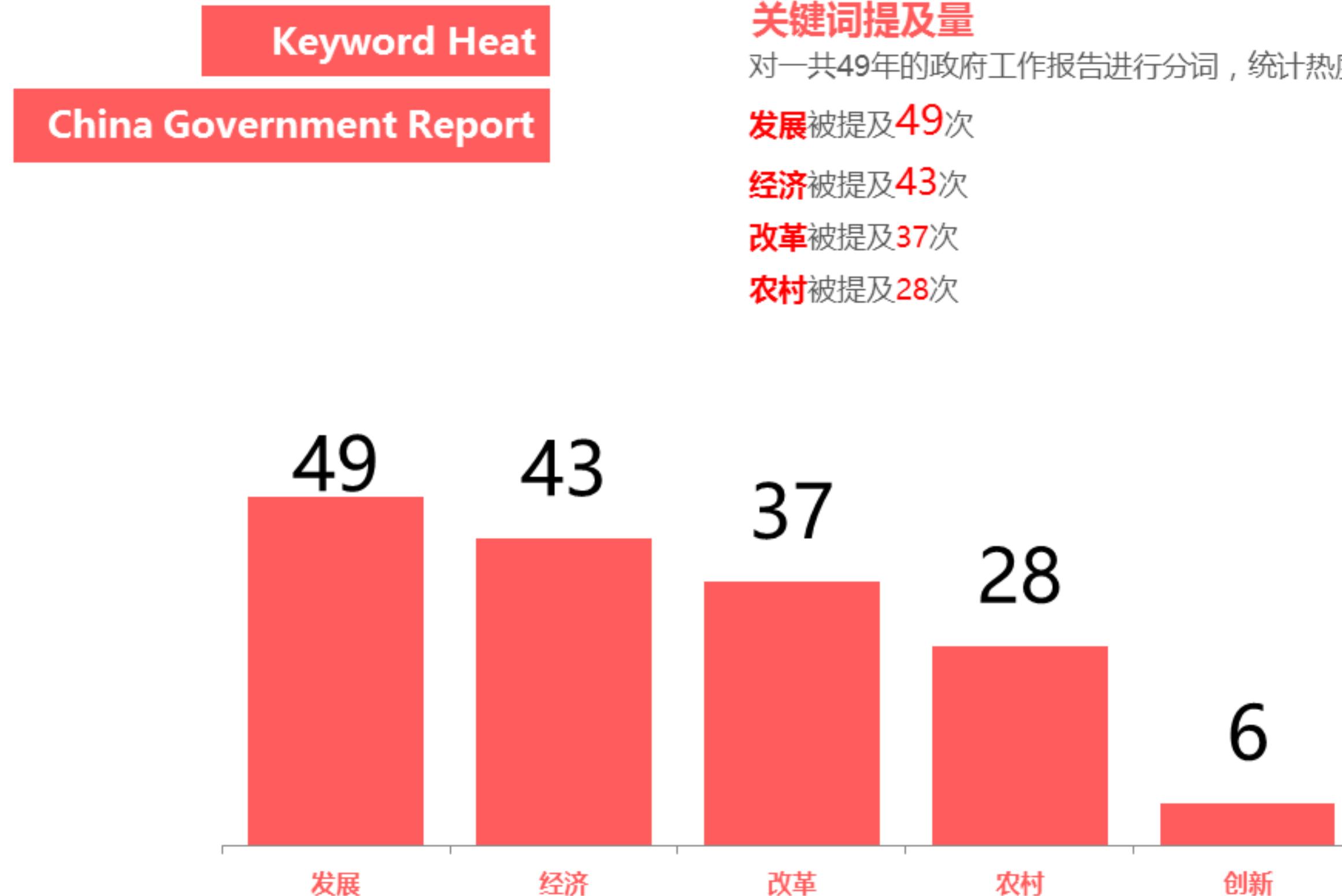
发展

城乡稳定 政策机制 深入国家 取得创新
健全 改善完善 支持制度重大 消费创业 体系生态
我国 改革开放 扩大安全 服务健全 依法政府建设
改革 落实深化 提高经济 推动群众 今年民生 建设统筹
增加 实现全面 实施战略 国际产能 工作结构
制度 全面推进 继续投资 动能农业 发展
坚决 加强合作 现代化 治理服务业 重点地区 试点五年
实施 保持事业结构性 农村经济社会 维护

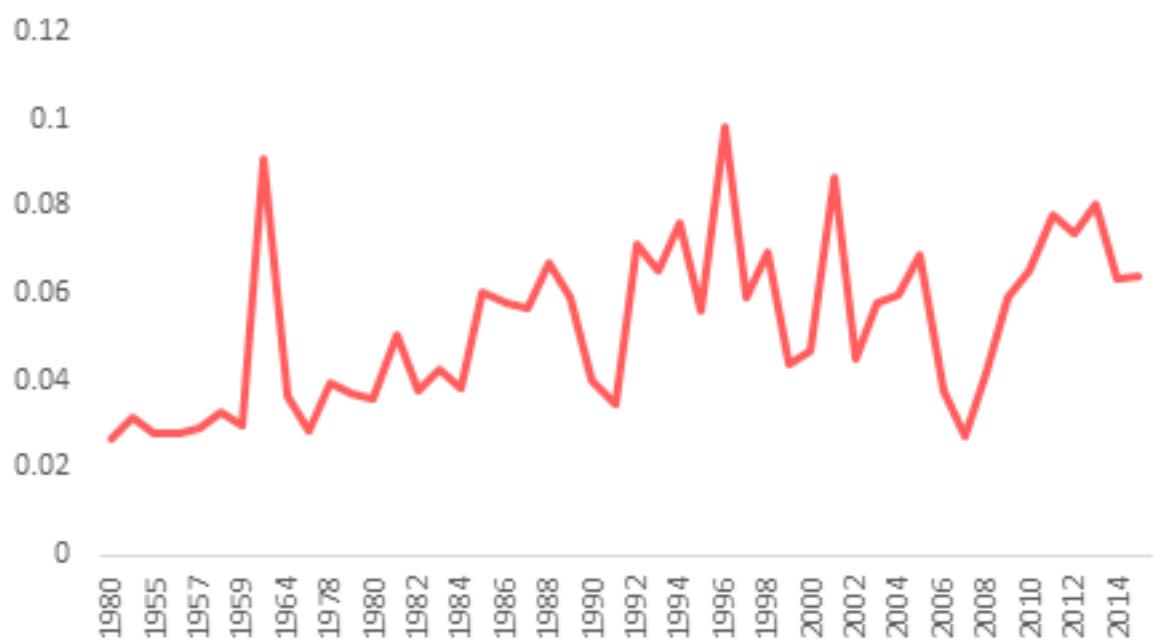
习大大时期执政热度词



Result

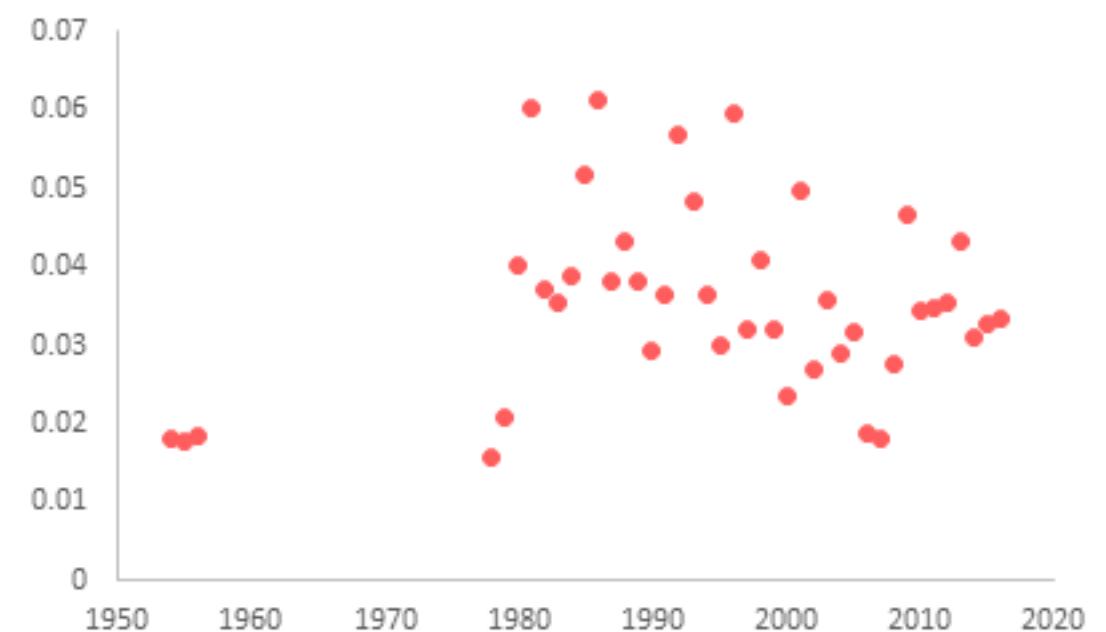


“改革” 热度指数



可以明显看到，在每次世界经济危机前后，“改革”热度指数明显激增。毕竟“改革是发展的动力”也是我朝政府的治国之道。

“经济” 热度散点图



可以明显看到，我朝自打1978年改革开放实施，“经济”在政府报告中的曝光度陡增。毕竟“以经济建设为中心”是四项基本原则之一。



未完待续

We are trying to add more features in this repository...



THANK
YOU

All BY LucasX