

## 可视化中国的方言

董元烨、樊玥、王书睿、王新昊

汉语方言的数据可视化工作对于语言 学研究和文化传承具有重要意义。本项目 以中国方言可视化为主题, 收集各地区方! 言音频文本数据,使用VUE、Echarts等框 架插件开发方言可视化系统。

• 乌鲁木齐

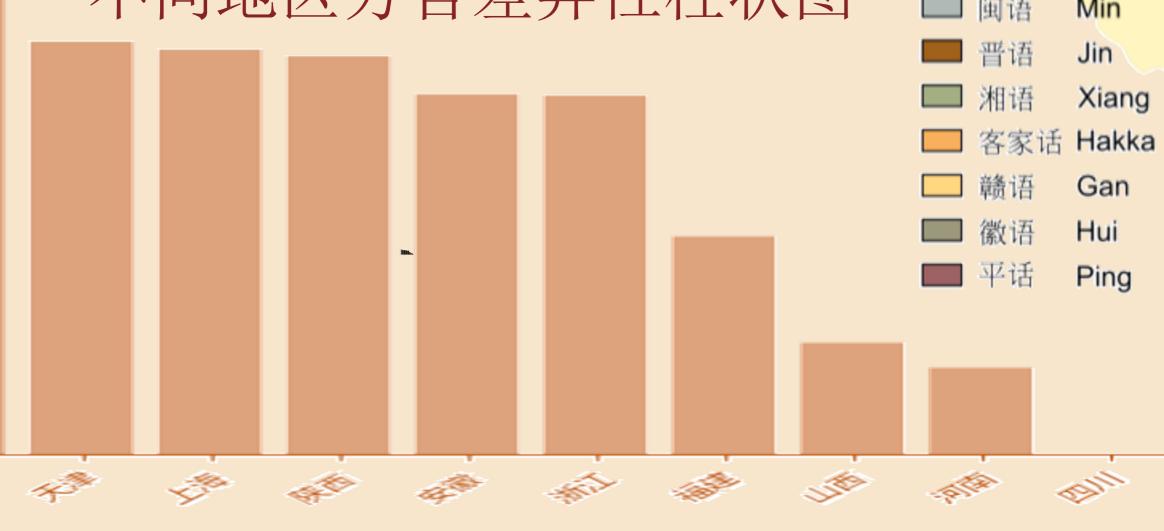
分布、方言相似度比较、 方言收集等。旨在促进方 言的研究和保护工作。

• 哈尔滨

沈阳

吴语

不同地区方言差异性柱状图



## 数据描述

L.O

0.8

0.6

0.4 -



文本



统计



中国汉语言分布

Xiang

Gan

Mandarin 836 million (全球)

Cantonese 71 million (全球)

77 million

36 million

31 million

60 million (包括台湾)

34 million (全球)

45 million (常算为官话的方言)

3.2 million (常算为吴语的方言)

2 million (常算为粤语的方言)

音频

文本数据: 收集不同语气典型话语8句 我们是在车站买的车票。

你算算看,这点钱够不够花。.....

统计数据: 十大汉语言分布

西南官话 19.12%	中原官话 16.30%	昊 8.4	语 3%	粤语 6.81%
<u> </u>		—————————————————————————————————————	吾 <b>多%</b>	晋语 4.81%
8.09%	5.95% 2.		语 /7%	其它 2.24%
东北官话 7.20%	胶辽官话   官	注银 言话 湘 56% 2.8	语 9%	客家话 2.18%

音频数据:来自34个地区8句话的音频



不同省份方言相似度mds图

• 兰州

●成都

●昆明

晋语

• 西安

湘语。长沙 赣语

客家话

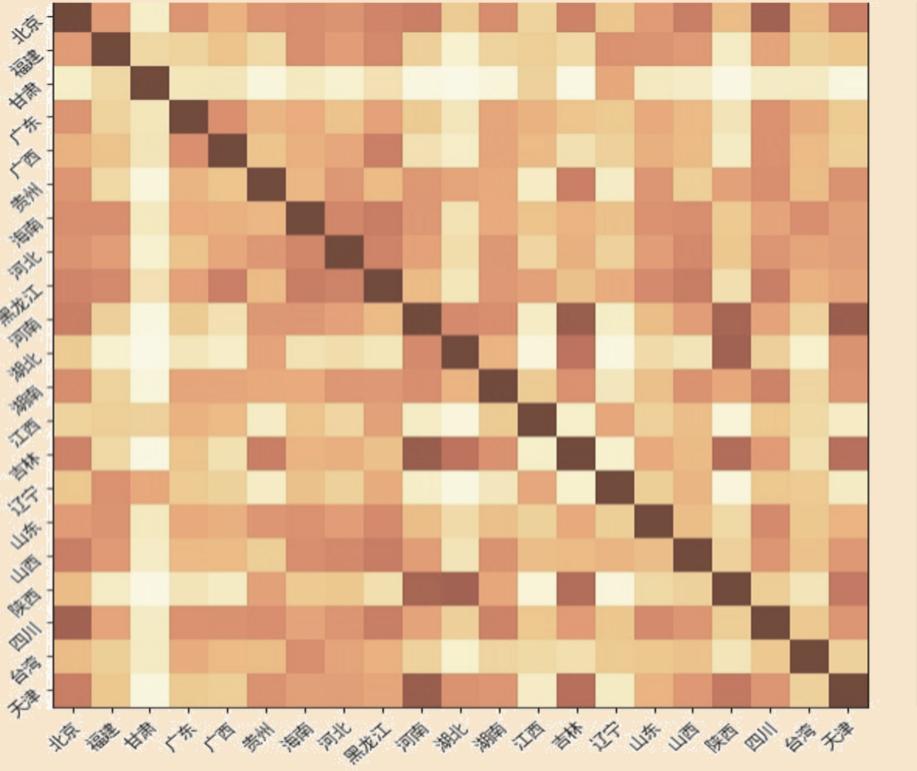
●重庆

地图分布 方言数目统计 mds图、热力图相似度比较 模块

语音收集 用户方言 录音收集

不同语句 音频播放 查看波形图

不同省份方言相似度热力图



开发框架: VUE+JavaScript! 热力图等: python 动态图表: **Echarts**