



新冠求助微博exploration



覃雄派



提纲

- 提纲



新冠求助微博
exploration

新冠求助微博exploration

- 案例：新冠肺炎求助患者数据画像

天边的一朵小马驹_

+关注

2月11日 00:04 来自 肺炎患者求助超话

💡 肺炎患者求助超话 【姓名】 XYZ

【年龄】 56

【所在城市】 武汉

【所在小区、社区】 汉阳区麒麟路麒麟社区 X栋Y单元Z楼

【患病时间】 1月26日

【联系方式】 1397xxx6027

【其他紧急联系电话】 1341xxx1935

【病情详细描述】 病人有高血压，甲状腺癌，发烧十天，高烧39度，双肺感染，咳嗽，无力，2月8日做了试剂盒，结果还没出来，现在病人病危，靠氧气，血氧70。医院不收入院。

【周边家人及隔离情况】 家里还有一个姐姐同住，姐姐有糖尿病重症，也昏睡了十几天，咳嗽，乏力，至今也没隔离，连去医院的劲都没有。家里还有一位父亲抗美援朝的离休老干部90岁，于2月5日在家离世（是不是这个病离世未知）
我已被隔离，妈妈目前情况非常严峻，只能躺床上靠吸氧根本不能动，连CT片都不能拍给我，我们报给社区说只能在家等，根本无力解决，现在情况紧急，生命危在旦夕，求求哪家医院能收治我的家人



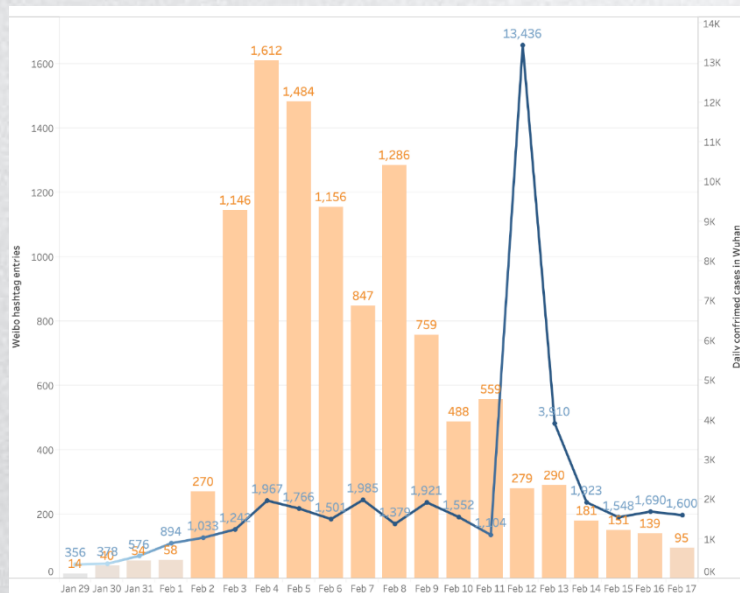
肺炎患者求助超话



数据采集：通过微博API获取

新冠求助微博exploration

- 肺炎患者求助超话的整体分析



- 黄色柱状图
 - 肺炎患者求助超话每天的发帖量
- 蓝色线条图
 - 武汉市卫健委汇报的每日新增新冠确诊数
- 时间区间
 - 1月29日 – 2月17日
- 你能发现什么?



新冠求助微博exploration

- 提出问题 (Insight Questions)
- 我们试图用数据解码数字背后的个体，看清那些宏大叙事之下的具体的人，尤其是那些在社交网络上以个体的身份发出求助的人——
- 他们是谁？
- 他们为什么而发声？
- 他们的处境有何共同点？



新冠求助微博exploration

- 数据准备
- 信息抽取：从文本数据中抽取出结构化信息

天边的一朵小马驹_ +关注

2月11日 00:04 来自 肺炎患者求助超话

肺炎患者求助超话 【姓名】 XYZ

【年龄】 56

【所在城市】 武汉

【所在小区、社区】 汉阳区麒麟路麒麟社区 X栋Y单元Z楼

【患病时间】 1月26日

【联系方式】 1397xxx6027

【其他紧急联系电话】 1341xxx1935

【病情详细描述】 病人有高血压，甲状腺癌，发烧十天，高烧39度，双肺感染，咳嗽，无力，2月8日做了试剂盒，结果还没出来，现在病人病危，靠氧气，血氧70。医院不收入院。

【周边家人及隔离情况】 家里还有一个姐姐同住，姐姐有糖尿病重症，也昏睡了十几天，咳嗽，乏力，至今也没隔离，连去医院的劲都没有。家里还有一位父亲抗美援朝的离休老干部90岁，于2月5日在家离世（是不是这个病离世未知）

我已隔离，妈妈目前情况非常严峻，只能躺床上靠吸氧根本不能动，连CT片都不能拍给我，我们报给社区说只能在家等，根本无力解决，现在情况紧急，生命危在旦夕，求求哪家医院能收治我的家人

id	name	age	city	addr
10815	ZBX	52	湖北武汉	长山社区万科嘉园小区
10842	ZBX	52	湖北武汉	长山社区万科嘉园小区
10473	CTC	75	武汉市洪山区	东湖御院 红星社区
84	DSX	51	武汉	云顶居，清水源社区
248	DSX		武汉	云顶居，清水源社区
565	DSX	51	武汉	云顶居，清水源社区
1205	RC	36	武汉	洪山区
1204	RC	36	武汉	洪山区
725	RC	36	武汉市	洪山区
729	RC	36	武汉市	洪山区
855	RC		湖北省武汉市	洪山区关山大道新竹路保利时代
854	RC	35	湖北省武汉市	保利时代
1826	RC	36	武汉	洪山区

你能观察到真实数据有多“脏”吗？

新冠求助微博exploration

- 数据清洗

name	age	city	addr
ZBX	52	湖北武汉	长山社区万科嘉园小区
ZBX	52	湖北武汉	长山社区万科嘉园小区
CTC	75	武汉市洪山区	东湖御院 红星社区
DSX	51	武汉	云顶居, 清水源社区
DSX		武汉	云顶居, 清水源社区
DSX	51	武汉	云顶居, 清水源社区
RC	36	武汉	洪山区
RC	36	武汉	洪山区
RC	36	武汉市	洪山区
RC	36	武汉市	洪山区
RC		湖北省武汉市	洪山区关山大道新竹路保利时代
RC	35	湖北省武汉市	保利时代
RC	36	武汉	洪山区

- 数据缺失
 - 属性age
- 数据重复
 - 同一个名字的人出现在多条微博中, 是一个人吗?
- 属性值冲突
 - RC的年龄是35还是36?
- 属性值粒度粗细不一
 - 洪山区
 - 保利时代
 - 洪山区关山大道新竹路保利时代

新冠求助微博exploration

- 数据扩充 (Data Enrichment)
 - 能否在地图上定位每位求助者所在的小区
 - 能否找到每位求助者所在小区的平均房价和建造年份
 - 数据扩充
 - 将求助者自己填写的小区名称对齐到链家小区名录中

name	age	city	addr
ZBX	52	湖北武汉	长山社区万科嘉园小区
ZBX	52	湖北武汉	长山社区万科嘉园小区
CTC	75	武汉市洪山区	东湖御院 红星社区
DSX	51	武汉	云顶居, 清水源社区
DSX		武汉	云顶居, 清水源社区
DSX	51	武汉	云顶居, 清水源社区
RC	36	武汉	洪山区
RC	36	武汉	洪山区
RC	36	武汉市	洪山区
RC	36	武汉市	洪山区
RC		湖北省武汉市	洪山区关山大道新竹路保利时代
RC	35	湖北省武汉市	保利时代
RC	36	武汉	洪山区

实体链接



小区名	区县	街道	年份	二手房价
洪山区 027社区	洪山	老南湖	2008	18869
黄陂区 08经典	黄陂	盘龙城	2010	12231
硚口区 2008城市花	硚口	长丰常码头	2007	13682
武昌区 2008新长江	武昌	积玉桥	2005	22585
青山区 20街坊	青山	青山	1987	17574
青山区 22街坊	青山	青山	1985	16752

链家小区名录

新冠求助微博exploration

- 数据扩充 (Data Enrichment)
 - 能否在地图上定位每位求助者所在的小区
 - 能否找到每位求助者所在小区的平均房价和建造年份
 - 数据扩充
 - 通过百度地图获取求助者对应到链家小区的经纬度坐标

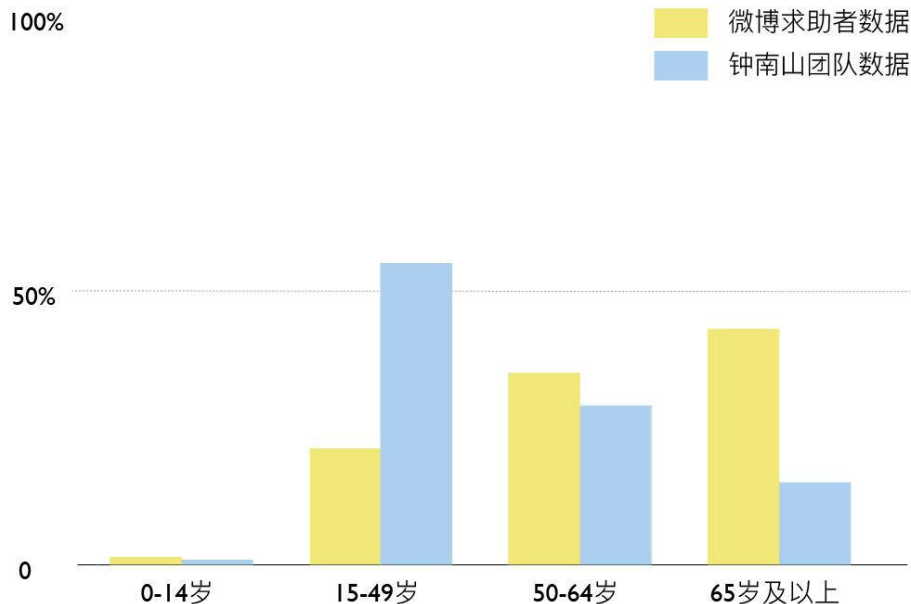
小区名	区县	街道	年份	二手房价
洪山区 027社区	洪山	老南湖	2008	18869
黄陂区 08经典	黄陂	盘龙城	2010	12231
硚口区 2008城市花	硚口	长丰常码头	2007	13682
武昌区 2008新长江	武昌	积玉桥	2005	22585
青山区 20街坊	青山	青山	1987	17574
青山区 22街坊	青山	青山	1985	16752

链家小区名录

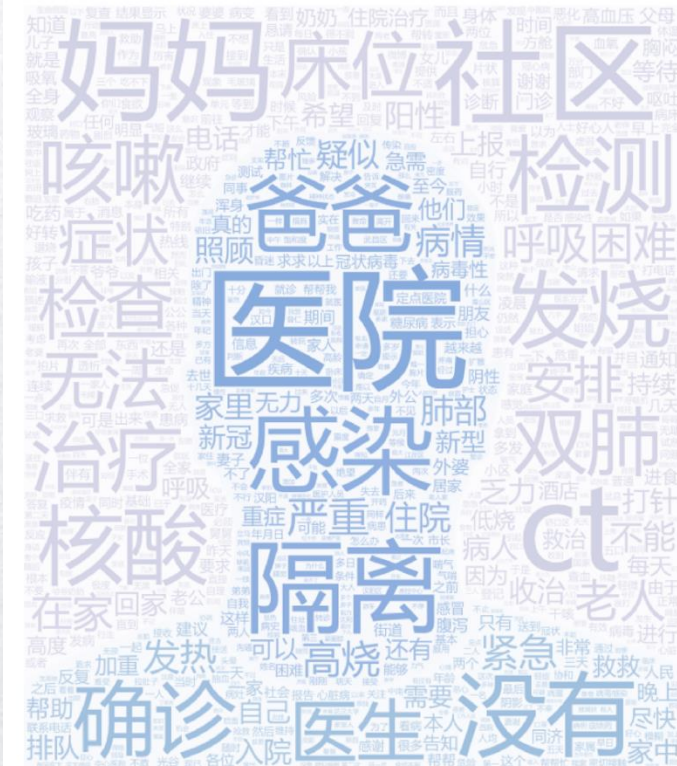
调用百度API



- 数据分析
 - 围城之中，越老越无助

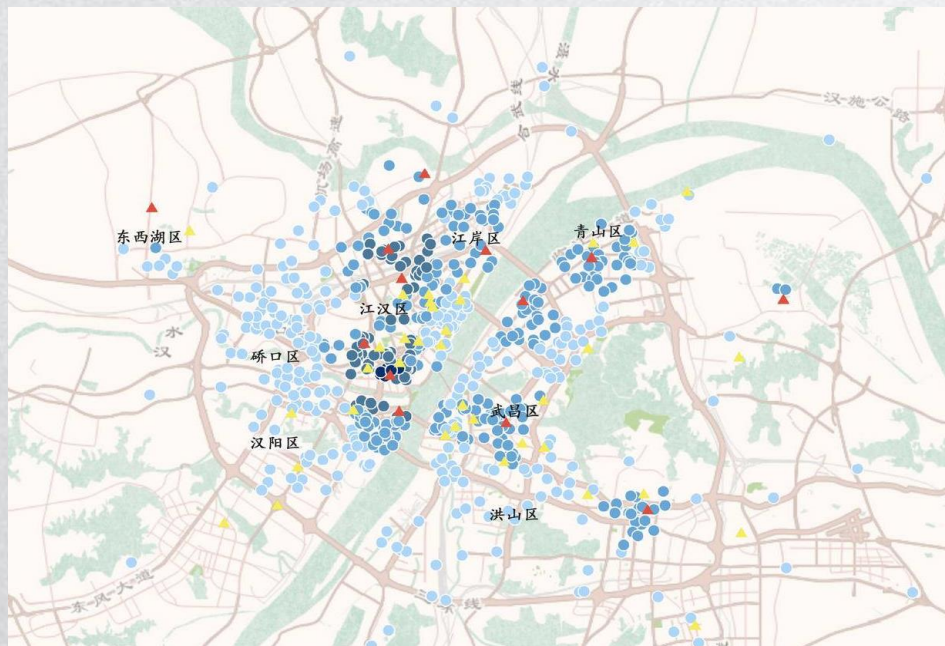


1183条求助微博的词频统计

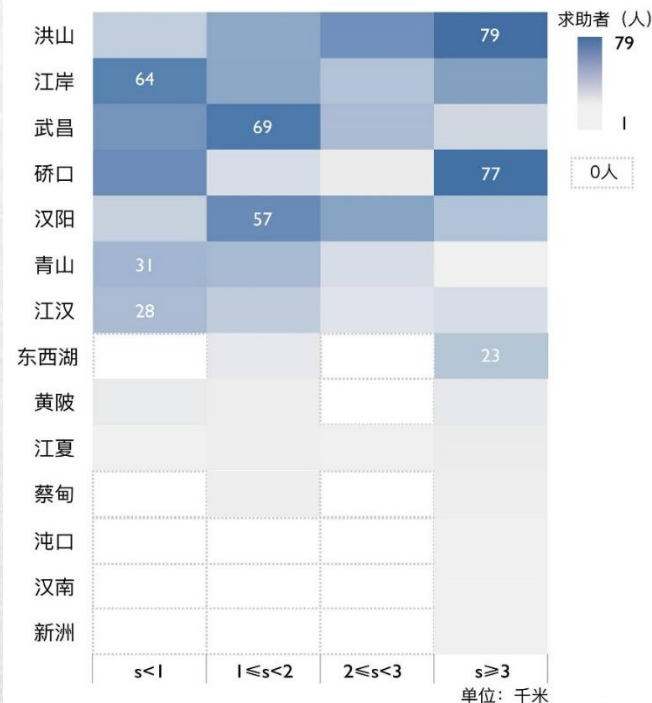


新冠求助微博exploration

- 数据分析
 - 求助者离医院有多远?



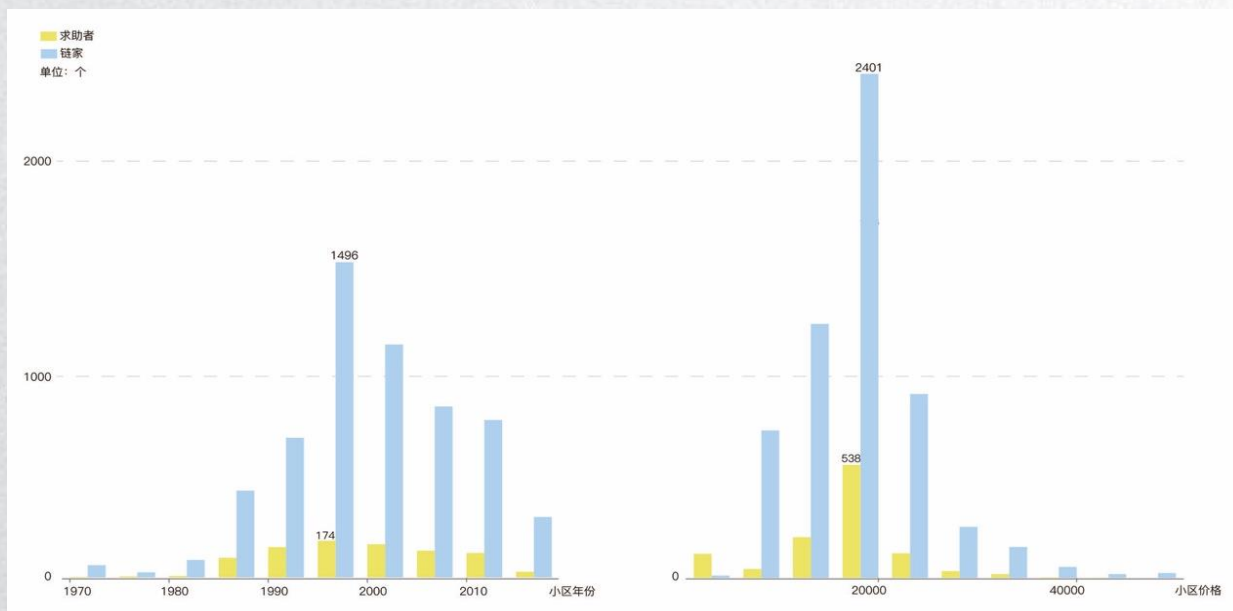
各区县求助者距最近医院的距离分布



新冠求助微博exploration

• 数据分析

- 求助者所居住的房屋，无论是价格还是年份，都与武汉住宅的整体分布相当





新冠求助微博exploration

- 数据分析结论

- 时代的雨滴打在每一个人头上，数字背后甚至被排除在数字之外的，是一个个鲜活的人，用文字、图片、视频、转发、点赞.....拼凑聚合成历史本身，而不甘愿沦为其注脚。
- 正如作家李静睿所言，“灾难面前，我们不是旁观，只是幸存。”这些陷入窘境的人、苦苦挣扎的人、奋力求生的人，不是弱者也不是他者，正是另一种处境下的你我。
- 照见他们，也是看到我们自己。