



武汉大学  
WUHAN UNIVERSITY



STRINGdata  
——字符串——



科学软件网  
SCIENCE SOFTWARE NETWORK

# Stata与中文地图

**Link Stata to Chinese Map and Beyond**

李春涛

武汉字符串数据科技有限公司

中南财经政法大学金融学院

# 中文地址中城市名称的提取

- 读入银行机构设立数据

```
clear all
cap mkdir E:\正则表达式\
cap mkdir E:\正则表达式\地址\
cd E:\正则表达式\地址\
local ad1 https://stata-club-1257787903.
copy "`ad1'cos.ap-chengdu.myqcloud.com/regex/4.xls" ///
    近期机构设立情况.xls, replace
import excel using 近期机构设立情况.xls, first
```

# 中文地址中城市名称的提取

- 
- ↵ **机构地址** ↵
- 
1. ↵ 合肥市庐阳区濉溪路 269 号 ↵
  2. ↵ 马鞍山市博望新区南环路精华模具厂综合大楼 ↵
  3. ↵ 合肥市安徽大市场 D1-363. 348 一层和安徽大市场 D1-364. 347 一层 ↵
  4. ↵ 安徽省安庆市潜山县梅城镇梅苑路 185 号 ↵
  5. | ↵ 东至县大渡口镇南大街安徽长江大市场综合楼 1018 室 ↵
  6. ↵ 安庆市光彩大市场四期元山路 1 号楼 9、11、13 号门面 ↵
  7. ↵ 芜湖县湾沚镇保沙路与荆山河路交口处城南农贸市场门面 ↵
  8. ↵ 安徽省蚌埠市固镇县连城镇中学路 96 号 ↵
  9. ↵ 繁昌县繁阳镇迎春西路台客隆超市门面 ↵
  10. ↵ 巢湖市巢庐路军分区教导队西侧碧桂园滨湖商业广场 102 号 ↵
-

# 中文地址中城市名称的提取

```
gen city1 = ustrregexs(1) if ustrregexm(机构地址, "安徽省(.+?市)")  
list
```

	机构地址	city1
1.	合肥市庐阳区濉溪路269号	
2.	马鞍山市博望新区南环路精华模具厂综合大楼	
3.	合肥市安徽大市场D1-363.348一层和安徽大市场D1-364.347一层	
4.	安徽省安庆市潜山县梅城镇梅苑路185号	安庆市
5.	东至县大渡口镇南大街安徽长江大市场综合楼1018室	
6.	安庆市光彩大市场四期元山路1号楼9、11、13号门面	
7.	芜湖县湾沚镇保沙路与荆山河路交口处城南农贸市场门面	
8.	安徽省蚌埠市固镇县连城镇中学路96号	蚌埠市
9.	繁昌县繁阳镇迎春西路台客隆超市门面	
10.	巢湖市巢庐路军分区教导队西侧碧桂园滨湖商业广场102号	

# 中文地址中城市名称的提取

```
gen city1 = ustrregexs(1) if ustrregexm(机构地址, "(安徽省)?(.+?市)")
```

	机构地址	city1
1.	合肥市庐阳区濉溪路269号	合肥市
2.	马鞍山市博望新区南环路精华模具厂综合大楼	马鞍山市
3.	合肥市安徽大市场D1-363.348一层和安徽大市场D1-364.347一层	合肥市
4.	安徽省安庆市潜山县梅城镇梅苑路185号	安庆市
5.	东至县大渡口镇南大街安徽长江大市场综合楼1018室	<u>东至县大渡口镇南大街安徽长江大市</u>
6.	安庆市光彩大市场四期元山路1号楼9、11、13号门面	安庆市
7.	芜湖县湾沚镇保沙路与荆山河路交口处城南农贸市场门面	芜湖县湾沚镇保沙路与荆山河路交口处城南农贸市
8.	安徽省蚌埠市固镇县连城镇中学路96号	蚌埠市
9.	繁昌县繁阳镇迎春西路台客隆超市门面	繁昌县繁阳镇迎春西路台客隆超市
10.	巢湖市巢庐路军分区教导队西侧碧桂园滨湖商业广场102号	<u>巢湖市</u>

# 中文地址中城市名称的提取

```
gen city2 = ustrregexs(2) if ustrregexm(机构地址, "(安徽省)?(."+?(?!超)市(?!场))")  
list city2
```

	机构地址	city2
1.	合肥市庐阳区濉溪路269号	合肥市
2.	马鞍山市博望新区南环路精华模具厂综合大楼	马鞍山市
3.	合肥市安徽大市场D1-363.348一层和安徽大市场D1-364.347一层	合肥市
4.	安徽省安庆市潜山县梅城镇梅苑路185号	安庆市
5.	东至县大渡口镇南大街安徽长江大市场综合楼1018室	
6.	安庆市光彩大市场四期元山路1号楼9、11、13号门面	安庆市
7.	芜湖县湾沚镇保沙路与荆山河路交口处城南农贸市场门面	
8.	安徽省蚌埠市固镇县连城镇中学路96号	蚌埠市
9.	繁昌县繁阳镇迎春西路台客隆超市门面	
10.	巢湖市巢庐路军分区教导队西侧碧桂园滨湖商业广场102号	巢湖市

# 中文地址中城市名称的提取

```
replace city = "池州市" if ustrregexm(机构地址, "(安徽省)?东至县?")
replace city = "芜湖市" if ustrregexm(机构地址, "(安徽省)?繁昌县?")
replace city = "合肥市" if ustrregexm(机构地址, "(安徽省)?巢湖市?")
```

# 中文地址中城市名称的提取

- 行政区划表

```
clear all
cap mkdir E:\正则表达式\
cap mkdir E:\正则表达式\地址\
cd E:\正则表达式\地址\
local ad1 https://stata-club-1257787903.cos.
copy "`ad1'ap-chengdu.myqcloud.com/regex/5.dta" 安徽行政区划.dta, replace
use 安徽行政区划.dta, clear
list in 1/10
```



# 行政区划表

	省份	城市	区县
1.	安徽省	亳州市	谯城区
2.	安徽省	亳州市	涡阳县
3.	安徽省	亳州市	蒙城县
4.	安徽省	亳州市	利辛县
5.	安徽省	六安市	金安区
6.	安徽省	六安市	裕安区
7.	安徽省	六安市	叶集区
8.	安徽省	六安市	霍邱县
9.	安徽省	六安市	舒城县
10.	安徽省	六安市	金寨县

# 行政区划表

	城市	区县			cmd
1.	亳州市	谯城区	replace city = "亳州市"	if ustrregexm(机构地址, "^(安徽省)?谯城区?")	
2.	亳州市	涡阳县	replace city = "亳州市"	if ustrregexm(机构地址, "^(安徽省)?涡阳县?")	
3.	亳州市	蒙城县	replace city = "亳州市"	if ustrregexm(机构地址, "^(安徽省)?蒙城县?")	
4.	亳州市	利辛县	replace city = "亳州市"	if ustrregexm(机构地址, "^(安徽省)?利辛县?")	
5.	六安市	金安区	replace city = "六安市"	if ustrregexm(机构地址, "^(安徽省)?金安区?")	
6.	六安市	裕安区	replace city = "六安市"	if ustrregexm(机构地址, "^(安徽省)?裕安区?")	
7.	六安市	叶集区	replace city = "六安市"	if ustrregexm(机构地址, "^(安徽省)?叶集区?")	
8.	六安市	霍邱县	replace city = "六安市"	if ustrregexm(机构地址, "^(安徽省)?霍邱县?")	
9.	六安市	舒城县	replace city = "六安市"	if ustrregexm(机构地址, "^(安徽省)?舒城县?")	
10.	六安市	金寨县	replace city = "六安市"	if ustrregexm(机构地址, "^(安徽省)?金寨县?")	

# 行政区划表

```
clear all
cap mkdir E:\正则表达式\
cap mkdir E:\正则表达式\地址\
cd E:\正则表达式\地址\
copy "https://stata-club-1257787903.cos.ap-chengdu.myqcloud.com/regex/5.dta" 安
use 安徽行政区划.dta, clear
keep 城市 区县
gen cmd = `replace city = "" + 城市 + `"' if ustrregexm(机构地址, "^(安徽省)?"
list in 1/10
```

# 行政区划表

```
clear all
cd E:\正则表达式\地址\
frame rename default region
copy "https://stata-club-1257787903.cos.ap-chengdu.myqcloud.com/regex/5.dta" 安
use 安徽行政区划.dta, clear
keep 城市 区县
local region_N =_N
mkf bank
cwf bank
copy "https://stata-club-1257787903.cos.ap-chengdu.myqcloud.com/regex/4.xls" 近
import excel using 近期机构设立情况.xls, first
gen str15 city=""
forvalue i = 1/`region_N' {
    local county = _frval(region,区县,`i')
    local city = _frval(region,城市,`i')
    replace city = "`city'" if ustrregexm(机构地址, "^(安徽省?)?`cou
}
```

# 行政区划表

```
clear all
cap mkdir E:\正则表达式\
cap mkdir E:\正则表达式\地址\
cd E:\正则表达式\地址\
local ad1 https://stata-club-1257787903.
copy "`ad1'cos.ap-chengdu.myqcloud.com/regex/4.xls" ///
    近期机构设置情况.xls, replace
import excel using 近期机构设置情况.xls, first
keep in 1/10

local bdk RkwfPwjwfrn3P5XZoNKz7BScyor0nZvW
cngcode, baidukey(`bdk') fulladdress(机构地址) ///
    lat(bank_lat) long(bank_lng)
cnaddress, baidukey(`bdk') long(bank_lng) lat(bank_lat) ///
    country(country) province(prov) city(city) ///
    district(county) address(address)
```

# 行政区划

```
. list 机构地址 prov city county
```

	机构地址	prov	city	county
1.	合肥市庐阳区濉溪路269号	安徽省	合肥市	庐阳区
2.	马鞍山市博望新区南环路精华模具厂综合大楼	安徽省	马鞍山市	博望区
3.	合肥市安徽大市场D1-363.348一层和安徽大市场D1-364.347一层	安徽省	合肥市	瑶海区
4.	安徽省安庆市潜山县梅城镇梅苑路185号	安徽省	安庆市	潜山市
5.	东至县大渡口镇南大街安徽长江大市场综合楼1018室	安徽省	池州市	东至县
6.	安庆市光彩大市场四期元山路1号楼9、11、13号门面	安徽省	安庆市	宜秀区
7.	芜湖县湾沚镇保沙路与荆山河路交口处城南农贸市场门面	安徽省	芜湖市	湾沚区
8.	安徽省蚌埠市固镇县连城镇中学路96号	安徽省	蚌埠市	固镇县
9.	繁昌县繁阳镇迎春西路台客隆超市门面	安徽省	芜湖市	繁昌區
10.	巢湖市巢庐路军分区教导队西侧碧桂园滨湖商业广场102号	安徽省	合肥市	巢湖市

# 我们赋予**Stata**如下的功能

- 把中文地址转换为经纬度（cngcode）
- 把经纬度转换为中文地址（cnaddress）
- 搜索某一半径范围内的地铁站、医院、商场等 (cnmapsearch)
- 计算两个位置之间的交通距离和通勤时间 (cntraveltime)

# 命令开发者

Chuntao LI  
China Stata Club(爬虫俱乐部)  
Wuhan, China  
cht1@zuel.edu.cn

Yuan Xue  
China Stata Club(爬虫俱乐部\华中科技大学博士生)  
Wuhan, China  
xueyuan@hust.edu.cn

Xueren Zhang  
China Stata Club(爬虫俱乐部\武汉大学博士生)  
Wuhan, China  
zhijunzhang\_hi@163.com



# 安装命令

```
local url https://stata-club-1257787903.cos.ap-chengdu.  
net install cnmapsearch,from("`url'myqcloud.com/cngcode/cngcode")  
net install cnmapsearch,from("`url'myqcloud.com/cnmapsearch/")  
net install cnmapsearch,from("`url'myqcloud.com/cntravelttime/")
```

# 中文地址到经纬度

- cngcode
- extract longitude and latitude from a given Chinese address
- from Baidu Map API(<http://api.map.baidu.com>)

# 例1

```
clear all
local bdk RkwfPwjwfrn3P5XZoNKz7BScyor0nZvW
input str15 prov str15 city str20 county str100 address
"湖北省" "武汉市" "洪山区" "华中科技大学南一楼"
"湖北省" "武汉市" "武昌区" "武汉大学经济管理学院"
"北京市" "" "海淀区" "中央财经大学中财大厦"
"广东省" "深圳市" "南山区" "深圳大学"
end

gen fulladdress = prov+city+county+address
cngcode, baidukey(`bdk') ///
    province(prov) city(city) ///
    district(county) address(address)
```

# 例1

	county	address	fulladdress	longitude	latitude
1	洪山区	华中科技大学南一楼	湖北省武汉市洪山区华中科技大学南一楼	114.42016	30.515493
2	武昌区	武汉大学经济管理学院	湖北省武汉市武昌区武汉大学经济管理学院	114.37953	30.545136
3	海淀区	中央财经大学中财大厦	北京市海淀区中央财经大学中财大厦	116.34753	39.963927
4	南山区	深圳大学	广东省深圳市南山区深圳大学	113.94265	22.53945

# 例1

```
cngcode, baidukey(`bdk`) ///  
    fulladdress(fulladdress) ///  
    lat(lat2) long(long2)
```

运行结果如下：

	fulladdress	longitude	latitude	long2	lat2
1	湖北省武汉市洪山区华中科技大学南一楼	114.42016	30.515493	114.42016	30.515493
2	湖北省武汉市武昌区武汉大学经济管理学院	114.37953	30.545136	114.37953	30.545136
3	北京市海淀区中央财经大学中财大厦	116.34753	39.963927	116.34753	39.963927
4	广东省深圳市南山区深圳大学	113.94265	22.53945	113.94265	22.53945

# 例1

```
cngcode, baidukey(`bdk`) ///  
    district(county) ///  
    lat(lat3) long(long3)  
  
cngcode, baidukey(`bdk`) ///  
    province(prov) district(county) ///  
    lat(lat4) long(long4)  
  
cngcode, baidukey(`bdk`) ///  
    province(prov) city(city) district(county) ///  
    lat(lat5) long(long5)
```

## 经纬度偏移

- 百度地图默认bd09ll坐标系，即百度BD-90经纬度坐标系，GCJ-02基础上加密的结果
- 高德地图和谷歌地图在中国内地区域使用的坐标系是GCJ-02坐标系下的经纬度
- 另外在 `cngcode` 和 `cnaddress` 增加了 `coordtype()` 选项，可以选择提交和获取的经纬度的坐标系类型
  - `coordtype(gcj02ll)` 获得和提交的就是GCJ-02坐标系下的经纬度
  - `coordtype(bd09ll)` 获得和提交的就是BD-90坐标系下的经纬度

# 经纬度偏移

```
cngcode, baidukey(`bdk`) ///  
    province(prov) city(city) district(county) ///  
    lat(lat6) long(long6) coordtype(gcj0211)  
cngcode, baidukey(`bdk`) ///  
    province(prov) city(city) district(county) ///  
    lat(lat7) long(long7) coordtype(bd0911)
```



# 把经纬度转换为中文地址

(标志性地点名称)

```
clear all
local bdk RkwfPwjwfrn3P5XZoNKz7BScyor0nZvW
input str15 prov str15 city str20 county str100 address

"河南省" "开封市" "顺河回族区" "河南大学老校区"
"河南省" "开封市" "顺河回族区" "明伦街85号"
"湖北省" "武汉市" "武昌区" "武汉大学经济管理学院"
"湖北省" "武汉市" "武昌区" "八一路299号"
end

gen fulladdress = prov+city+county+address
```

## 例2

- 利用 `cngcode` 获取经纬度，然后用 `cnaddress` 将获取到的经纬度转换为中文地址

```
local bdk RkwfPwjwfrn3P5XZoNKz7BScyor0nZvW
cngcode, baidukey(`bdk`) province(prov) city(city) ///
    district(county) address(address)
cnaddress, baidukey(`bdk`) long(longitude) ///
    lat(latitude) country(country) ///
    province(new_prov) city(new_city) ///
    district(new_county) address(new_address)
```

## 例2

运行结果如下：

	city	county	address	fulladdress	longitude	latitude	country	new_prov	new_city	new_county	street	new_address
1	开封市	顺河回族区	河南大学老校区	河南省开封市顺河回族区河南大学老校区	114.37568	34.817677	中国	河南省	开封市	顺河回族区	琢玉路	河南省开封市顺河回族区琢玉路
2	开封市	顺河回族区	明伦街85号	河南省开封市顺河回族区明伦街85号	114.37606	34.814213	中国	河南省	开封市	顺河回族区	明伦街	河南省开封市顺河回族区明伦街85号
3	武汉市	武昌区	武汉大学经济管理学院	湖北省武汉市武昌区武汉大学经济管理	114.37953	30.545136	中国	湖北省	武汉市	武昌区	东湖南路	湖北省武汉市武昌区东湖南路
4	武汉市	武昌区	八一路299号	湖北省武汉市武昌区八一路299号	114.32255	30.559567	中国	湖北省	武汉市	武昌区	明惠路	湖北省武汉市武昌区明惠路1-1号

# 计算通勤距离和通勤时间

```
cntraveltime, baidukey(`bdk`) ///  
  startlat(taxb_lat) startlng(taxb_lng) ///  
  endlat(mall_lat) endlng(mall_lng) ///  
  mode("car") tactic(4)
```

# (1) 交通模式--bus

- 交通模式
  - bus
    - bus 0: default, recommendation
    - bus 1: Less transfer
    - bus 2: less walk
    - bus 3: no subway
    - bus 4: as quickly as possible
    - bus 5: subway
  - car

## (2) 交通模式--car

- 交通模式
  - bus
  - car
    - car 0: default
    - car 3: avoid high speed
    - car 4: high speed priority
    - car 5: avoid congested sections
    - car 6: avoiding toll stations
    - car 7: both 4 and 5

## (3) 交通模式--bike

- 交通模式
  - bus
  - car
  - bike
    - bike 0: default, common
    - bike 1: electric bicycle

## 例3：寻找地铁站

```
clear all
local bdk RkwfPwjwfrn3P5XZoNKz7BScyor0nZvW
input str15 prov str15 city str20 county str100 address
"河南省" "郑州市" "郑东新区" "郑州市郑东新区河南大学龙子湖校区"
"湖北省" "武汉市" "武昌区" "武汉大学经济管理学院"
end

gen fulladdress = prov+city+county+address

cngcode, baidukey(RkwfPwjwfrn3P5XZoNKz7BScyor0nZvW) ///
        fulladdress(fulladdress) lat(univ_lat) long(univ_lng)
```



## 例3：寻找地铁站

- 运行结果如下：

	prov	city	county	address	fulladdress	univ_lng	univ_lat	
1	河南省	郑州市	郑东新区	郑州市郑东新区河南大学龙子湖校区	河南省郑州市郑东新区郑州市郑东新区河南大学龙子湖校区	113.69722	34.818233	
2	湖北省	武汉市	武昌区	武汉大学经济管理学院	湖北省武汉市武昌区武汉大学经济管理学院	114.37953	30.545136	

# 用 `cnmapsearch` 寻找附近的地铁站

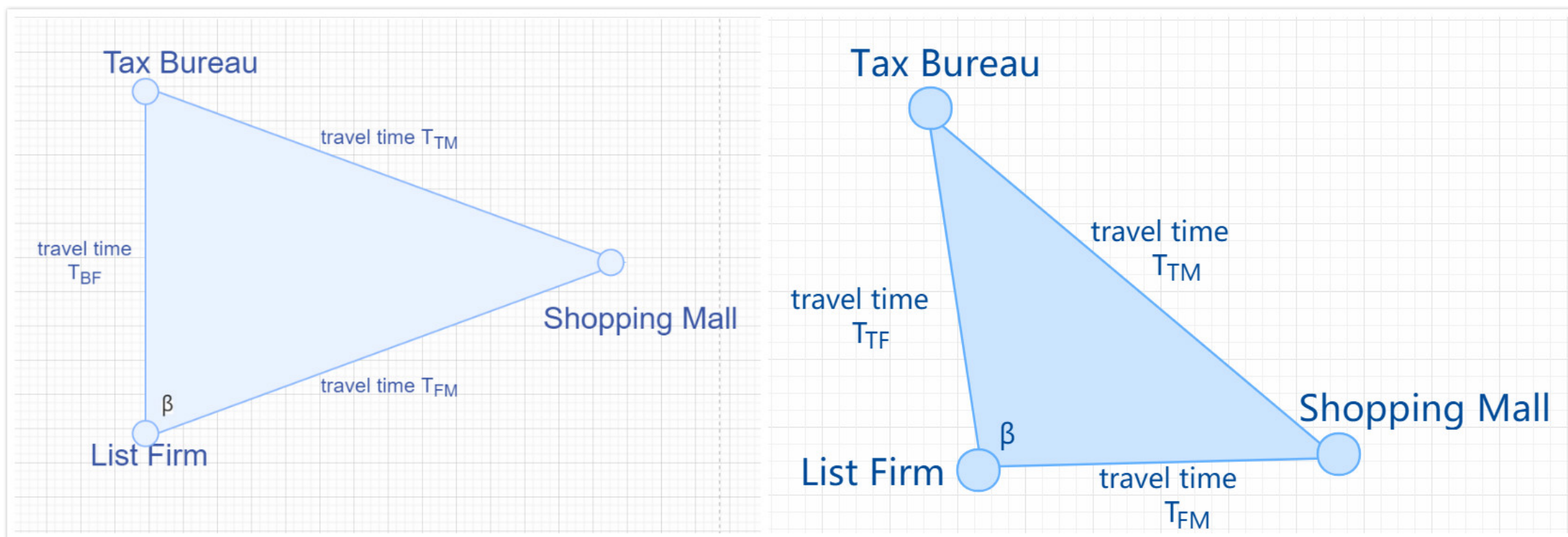
```
cnmapsearch, baidukey(`bdk`) ///  
             latitude(univ_lat) longitude(univ_lng) ///  
             keyword("地铁") radius(10000)  
  
keep if index(tag, "地铁站")  
sort centerid distance  
by centerid: keep if _n==1 //保留最近的地铁站
```

## 例3：寻找地铁站

运行结果如下：

centerid	locid	prov	city	county	fulladdress	univ_lng	univ_lat	name	address	telephone	tag	distance	loc_lat	loc_lng	
1	1	4	河南省	郑州市	郑东新区	河南省郑州市郑东新区郑州市郑东新区河南大学龙子湖校区	113.69722	34.818233	沙门	地铁2号线		地铁站	1137	34.825294	113.68821
2	2	22	湖北省	武汉市	武昌区	湖北省武汉市武昌区武汉大学经济管理学院	114.37953	30.545136	广埠屯	地铁2号线		地铁站	1999	30.528971	114.37038

# 例4: 从税务局去Shopping Mall 经过上市公司吗?



# (1) 爬取上市公司中文地址

- 分析源网页：以平安银行（000001）为例，在源网页中找到目标信息--“办公地址”
- [http://vip.stock.finance.sina.com.cn/corp/go.php/vCI\\_CorpInfo/stoc\\_kid/000001.phtml](http://vip.stock.finance.sina.com.cn/corp/go.php/vCI_CorpInfo/stoc_kid/000001.phtml)



# 爬取一家公司的地址

```
copy "http://vip.stock.finance.sina.com.cn/corp/go.php/vCI_CorpInfo/stockid/00
infix strL v 1-100000 using temp.txt, clear
replace v = ustrfrom(v, "gb18030", 1)
keep if index(v[_n-1], "<td class="ct">办公地址: </td>")
replace v = ustrregexra(v, "<.*?>", "")
```

运行结果如下：

	v
1	广东省深圳市罗湖区深南东路5047号,中国广东省深圳市福田区益田路5023号平安金融中心B座

- 爬取多家公司的地址
  - `cnstock` 获取公司代码，随机保留10家上市公司

```
clear all  
cnstock all  
sample 10, count
```



- 爬取多家公司的地址

```
mkf address stkcd strL address
cwf address
levelsof stkcd, local(stkcd)
foreach stk in `stkcd' {
    local stk: disp %06.0f `stk'
    cap copy "http://vip.stock.finance.sina.com.cn/corp/go.php/vCI_CorpInfo/stoc
    while _rc != 0 {
        sleep 5000
        cap copy "http://vip.stock.finance.sina.com.cn/corp/go.php/vCI_CorpInfo/sto
    }
    infix strL v 1-100000 using temp.txt, clear
    replace v = ustrfrom(v, "gb18030", 1)
    keep if index(v[_n-1], "<td class="ct">办公地址: </td>"')
    replace v = ustrregexra(v, "<.*?>", "")
    frame post address (`stk') (v[1])
}
```

运行结果如下：

	stkcd	address		
1	000892	重庆市江北区北城天街15号富力海洋广场6幢2201室		
2	000927	天津市西青区京福公路578号		
3	002035	广东省中山市小榄镇工业大道南华园路1号		
4	002292	广东省广州市天河区珠江新城金穗路62号侨鑫国际金融中心37楼		
5	300110	山东省青岛市高科技工业园株洲路187号		
6	300196	江苏省常州市武进区遥观镇塘桥村		
7	300294	江西抚州高新技术产业开发区惠泉路333号		
8	300856	江苏省南京市江宁经济技术开发区苏源大道19号		
9	600783	山东省济南市经十路9999号黄金时代广场C座4层		
10	900952	辽宁省锦州经济技术开发区锦港大街1段1号		

## (2) `cngcode` 将中文地址转换为经纬度

```
cwf address  
local bdk RkwfPwjwfrn3P5XZoNKz7BScyor0nZvW  
cngcode, baidukey(`bdk`) fulladdress(address) ///  
    lat(firm_lat) long(firm_lng)
```

运行结果如下图：

	stkcd	address	firm_lng	firm_lat
1	000892	重庆市江北区北城天街15号富力海洋广场6幢2201室	106.54459	29.588128
2	000927	天津市西青区京福公路578号	117.06318	39.087002
3	002035	广东省中山市小榄镇工业大道南华园路1号	113.29489	22.587701
4	002292	广东省广州市天河区珠江新城金穗路62号侨鑫国际金融中心37楼	113.32872	23.128621
5	300110	山东省青岛市高科技工业园株洲路187号	120.44623	36.138827
6	300196	江苏省常州市武进区遥观镇塘桥村	120.06342	31.692293
7	300294	江西抚州高新技术产业开发区惠泉路333号	116.38503	27.947908
8	300856	江苏省南京市江宁经济技术开发区苏源大道19号	118.81642	31.916551
9	600783	山东省济南市经十路9999号黄金时代广场C座4层	117.11991	36.667025
10	900952	辽宁省锦州经济技术开发区锦港大街1段1号	121.0674	40.843103

# 获取目标地点半径**10**公里内的税务机关及其经纬度

cnmapsearch

```
cnmapsearch, baidukey(`bdk`) ///  
             latitude(firm_lat) longitude(firm_lng) ///  
             keyword("税务") radius(10000)  
keep if index(tag, "政府机构")
```

# 按距离排序，为每家公司保留距其最近的一家税务机关

```
sort stkcd distance
by stkcd: keep if _n==1

drop locid centerid
rename (loc_lat loc_lng name address) ///
      (taxb_lat taxb_lng taxb_name taxb_address)
drop telephone tag distance
```

运行结果如下图所示：

	firm_lng	firm_lat	taxb_name	taxb_address	taxb_lat	taxb_lng
1	106.54459	29.588128	国家税务总局重庆市江北区税务局观音桥税务所	重庆市江北区兴隆路26号	29.585648	106.54656
2	117.06318	39.087002	国家税务总局天津市西青区税务局张家窝所	天津市西青区嘉和路与祥和大道交叉路口向北约50米	39.069964	117.0428
3	113.29489	22.587701	国家税务总局中山市税务局东升税务分局	广东省中山市东升镇同乐大街100号	22.613559	113.30277
4	113.32872	23.128621	国家税务总局广州市天河区税务局猎德税务所	华利路59号保利大厦西域1、2、14、15楼	23.123879	113.32634
5	120.44623	36.138827	青岛市李沧区税务局第二税务所	大崂路1003号	36.168943	120.43851
6	120.06342	31.692293	中国税务(青盛东)	常州市武进区中吴大道与新乐路交叉口南150米	31.700749	120.10452
7	116.38503	27.947908	国家税务总局抚州高新技术产业开发区税务局	江西省抚州市临川区金梌大道678号	27.945373	116.3842
8	118.81642	31.916551	淳化街道个体税收征收服务点	江苏省南京市江宁区天元东路38号附近	31.93838	118.85478
9	117.11991	36.667025	国家税务总局济南市税务局	济南市历城区舜泰北路1001号	36.671062	117.14633
10	121.0674	40.843103	中国税务	昆仑山路8号附近	40.849352	121.06931

# 目标地点半径10000米内的商场

- 税务机关经纬度作为中心
- 获取中心点半径10公里范围内的购物中心

```
local bdk RkwfPwjwfrn3P5XZoNKz7BScyor0nZvW
cnmapsearch, baidukey(`bdk`) ///
    latitude(taxb_lat) longitude(taxb_lng) ///
    keyword("商场") radius(10000)

keep if index(tag, "购物中心") | index(tag, "百货商场")
```



# 找到距离税务机关最近的一家商场

```
sort stkcd distance
by stkcd: keep if _n==1

rename (loc_lat loc_lng name address) ///
      (mall_lat mall_lng mall_name mall_address)
drop telephone tag distance
```

# 找到距离税务机关最近的一家商场

运行结果如下图所示：

	taxb_name	taxb_address	taxb_lat	taxb_lng	mall_name	mall_address	mall_lat	mall_lng
1	国家税务总局重庆市江北区税务局观音桥税务所	重庆市江北区兴隆路26号	29.585648	106.54656	万汇乐奇世界	重庆市江北区北城天街38号	29.585842	106.54742
2	国家税务总局天津市西青区税务局张家窝所	天津市西青区嘉和路与祥和大道交叉路口向北约50-	39.069964	117.0428	朝向城	天津市西青区玉锦路8号	39.070008	117.04955
3	国家税务总局中山市税务局东升税务分局	广东省中山市东升镇同乐大街100号	22.613559	113.30277	东升商业广场	葵兴大道106号	22.623925	113.3029
4	国家税务总局广州市天河区税务局猎德税务所	华利路59号保利大厦西域1、2、14、15楼	23.123879	113.32634	广州国金天地购物中心	广东省广州市天河区珠江西路5号	23.124481	113.32912
5	青岛市李沧区税务局第二税务所	大崂路1003号	36.168943	120.43851	万达广场(青岛李沧店)	山东省青岛市李沧区巨峰路178号	36.172992	120.44101
6	中国税务(青盛东)	常州市武进区中吴大道与新乐路交叉口南150米	31.700749	120.10452	常州爱琴海购物公园奥特莱斯	常州市武进区东方东路165号	31.760154	120.05821
7	国家税务总局抚州高新技术产业开发区税务局	江西省抚州市临川区金棍大道678号	27.945373	116.3842	硕果时代购物中心	江西省抚州市临川区竹山路与金巢大道交叉-	27.960385	116.37096
8	淳化街道个体税收征收服务点	江苏省南京市江宁区天元东路38号附近	31.93838	118.85478	中国女人街	江宁区东山镇金箔路468号	31.95329	118.85876
9	国家税务总局济南市税务局	济南市历城区舜泰北路1001号	36.671062	117.14633	雨滴广场	济南市历下区新泺大街2008号	36.678558	117.13696
10	中国税务	昆仑山路8号附近	40.849352	121.06931	天街·橙购物中心	辽宁省锦州市太和区珠江街52-86号	40.889907	121.07909

# 计算两点间驾车距离和行车时间

## cntraveltime

计算税务机关-->商场的距离和行车时间

```
local bdk RkwfPwjwfrn3P5XZoNKz7BScyor0nZvW
cntraveltime, baidukey(`bdk`) ///
    startlat(taxb_lat) startlng(taxb_lng) ///
    endlat(mall_lat) endlng(mall_lng) ///
    mode("car") tactic(4)
rename (distance duration) (C_D C_T)
```

- 计算税务机关-->上市公司的距离和行车时间

```
cntraveltime, baidukey(`bdk`) ///  
    startlat(taxb_lat) startlng(taxb_lng) ///  
    endlat(firm_lat) endlng(firm_lng) ///  
    mode("car") tactic(4)  
rename (distance duration) (A_D A_T)
```

- 计算上市公司-->商场的距离和行车时间

```
cntraveltime, baidukey(`bdk`) ///  
  startlat(firm_lat) startlng(firm_lng) ///  
  endlat(mall_lat) endlng(mall_lng) ///  
  mode("car") tactic(4)  
rename (distance duration) (B_D B_T)
```

运行结果如下图所示：

	taxb_lat	taxb_lng	mall_name	mall_address	mall_lat	mall_lng	C_D	C_T	A_D	A_T	B_D	B_T
1	29.585648	106.54656	万汇乐奇世界	重庆市江北区北城天街38号	29.585842	106.54742	72	8	1717	461	899	190
2	39.069964	117.0428	朝向城	天津市西青区玉锦路8号	39.070008	117.04955	950	151	3518	400	2530	410
3	22.613559	113.30277	东升商业广场	葵兴大道106号	22.623925	113.3029	2259	347	4326	751	6087	966
4	23.123879	113.32634	广州国金天地购物中心	广东省广州市天河区珠江西路5号	23.124481	113.32912	836	408	1940	500	1231	308
5	36.168943	120.43851	万达广场(青岛李沧店)	山东省青岛市李沧区巨峰路178号	36.172992	120.44101	578	123	4212	683	4862	518
6	31.700749	120.10452	常州爱琴海购物公园奥特莱斯	常州市武进区东方东路165号	31.760154	120.05821	10977	1232	6274	815	12426	1361
7	27.945373	116.3842	硕果时代购物中心	江西省抚州市临川区竹山路与金巢大道交叉-	27.960385	116.37096	2938	357	527	62	2743	343
8	31.93838	118.85478	中国女人街	江宁区东山镇金箔路468号	31.95329	118.85876	2525	469	6035	903	8156	1245
9	36.671062	117.14633	雨滴广场	济南市历下区新泺大街2008号	36.678558	117.13696	1559	309	3474	621	3382	606
10	40.849352	121.06931	天街·橙购物中心	辽宁省锦州市太和区珠江街52-86号	40.889907	121.07909	5315	581	1229	201	6331	659

# 余弦定理

计算角度 $\beta$ 的大小：

假设角 $\beta$ 对应的边TM为 $b$ ,其余两边BF、FM为边 $a$ 、 $c$ 。

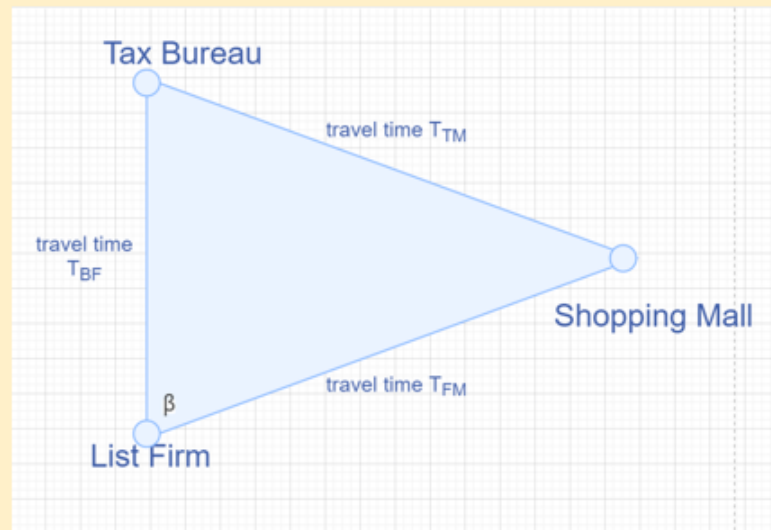
根据余弦定理： $b^2 = a^2 + c^2 - 2accos\beta$

$$\text{得 } \cos\beta = \frac{(a^2 + c^2) - b^2}{2ac}$$

所以  $\beta = \arccos \beta$

根据角度 $\beta$ 的大小即可判断税务局工作人员前往上市公司是顺路还是绕路。

接下来，将具体讲述如何在stata里实现上述模型。



# 将行车时间或距离作为边长计算公司的 $\beta$ 角

- 将距离作为边长计算上市公司的 $\cos\beta$ 与 $\beta$

```
gen cos_beta_D =(A_D^2+ B_D^2-C_D^2)/(2*A_D*B_D)
```

```
gen beta_D = acos(cos_beta_D)
```



- 将行车时长作为边长计算上市公司的 $\cos\beta$ 与 $\beta$

```
gen cos_beta_T =(A_T^2+ B_T^2-C_T^2)/(2*A_T*B_T)
```

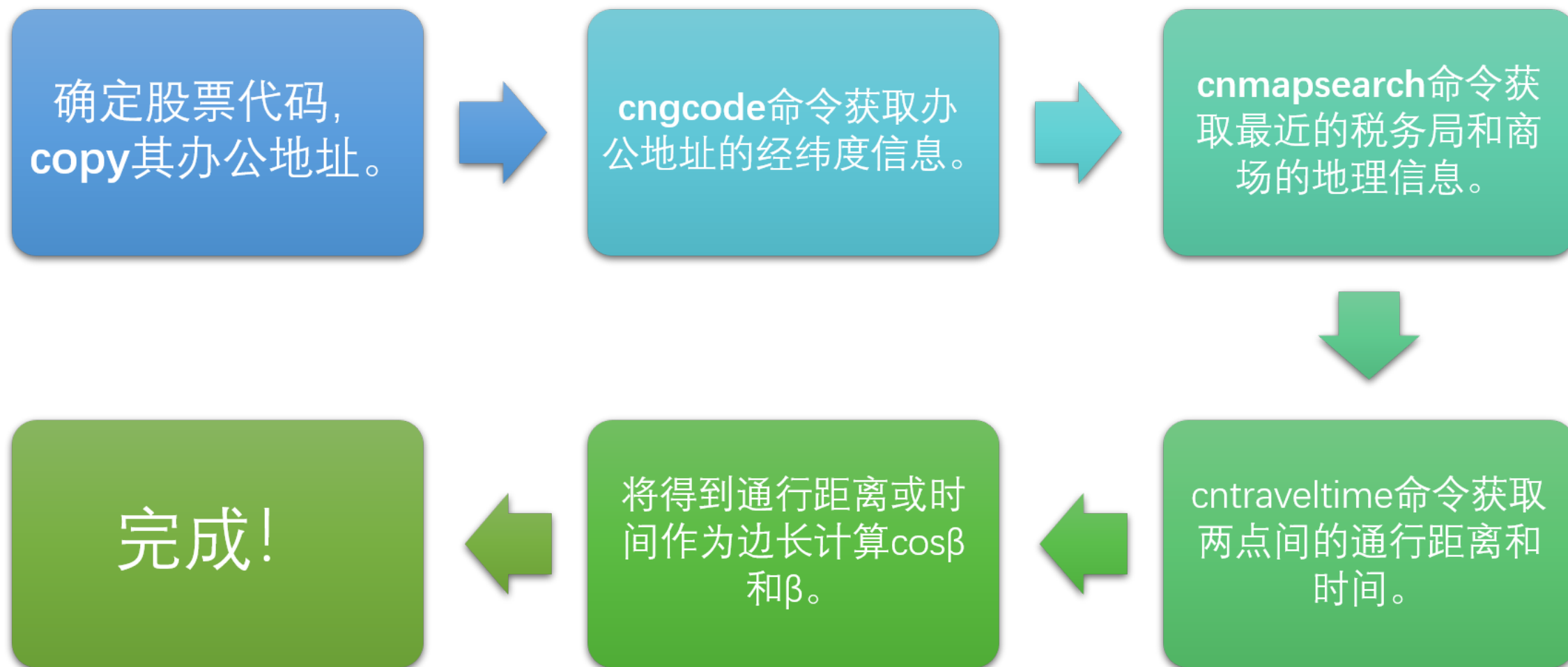
```
gen beta_T = acos(cos_beta_T)
```

运行结果如下图所示：

cos_beta_distance	beta_distance	cos_beta_time	beta_time
1.215065	.	1.418866	.
1.004137	.	.9307896	.3742293
.9619867	.2766101	.9488715	.3211549
.9589189	.2876301	.5792208	.9530239
1.002159	.	1.017095	.
.4699419	1.081571	.4501932	1.103815
-.2871086	1.862003	-.1400357	1.711294
.9809332	.1955896	.9541924	.3038475
.8969273	.4580256	.8734395	.5085745
.8574169	.5405673	.5175979	1.026755

# 流程图

最后，我们总结一下这几个地图命令在案例中的使用流程：



# 欢迎关注爬虫俱乐部

- Email: [Statatraining@163.com](mailto:Statatraining@163.com)
- Web: [www.wuhanstring.com](http://www.wuhanstring.com)
- 公众号:



# 我们开发的命令

- 地图系列 cngcode cnaddress cnmapsearch cntraveltime
- 股票系列 cntrade cnar cnintraday cnstock cntop10
- 结果输出 reg2docx sum2docx t2docx corr2docx ttable2
- 文本系列 wordconvert subinfile addbefore
- 其它 psemial ttable2 eventstudy addbefore

# 我们输送海内外的优秀研究助理

- 香港大学 薛原 司海涛 赵一 徐鑫
- 香港理工大学 张计宝 李子健 孙雪丽
- 香港岭南大学 高金凤 薛原 刘贝贝 潘铖 宋凯
- 武汉大学 张学人
- 中欧国际商学院 杨慧琳

# 我们的毕业生读博去向

- 北京大学光华管理学院（1人）
- 清华大学五道口金融学院（1人）
- 上海交通大学（1人）
- 西安交通大学（2人）
- 武汉大学（3人）
- 华中科技大学（1人）
- 上海财经大学（2人）
- 财政部财政科学研究所（1人）
- 中南财经政法大学金融学院（5人）

# 我们的课程曾经在如下学校讲授

浙江大学	武汉大学	华中科技大学	厦门大学
中央财经大学	西南财经大学	新疆财经大学	中南财经政法大学
郑州大学	云南财经大学	兰州财经大学	山东财经大学
湖北大学	福州大学	广西大学	南京信息工程大学
河南大学	华侨大学	华中农业大学	广东外语外贸大学
江汉大学		湖北经济学院	河南牧业经济学院
温州商学院	黄淮学院	长沙理工大学	石河子大学
华东政法大学	西南政法大学		





**Thank you!**