

Research Project Using Mobility Data

Data Source

- ▶ Mandy Hu (AP Marketing @ CUHK)
- ▶ SHT: 2nd largest wireless carrier in Shanghai
 - ▶ 6.5 millions cellphone users in Shanghai (20% market share)
 - ▶ 7 million landline users (88% market share)
- ▶ Big Data Center: data available since 2014
- ▶ User info, calls, internet usage (URL), mobility (BTS)

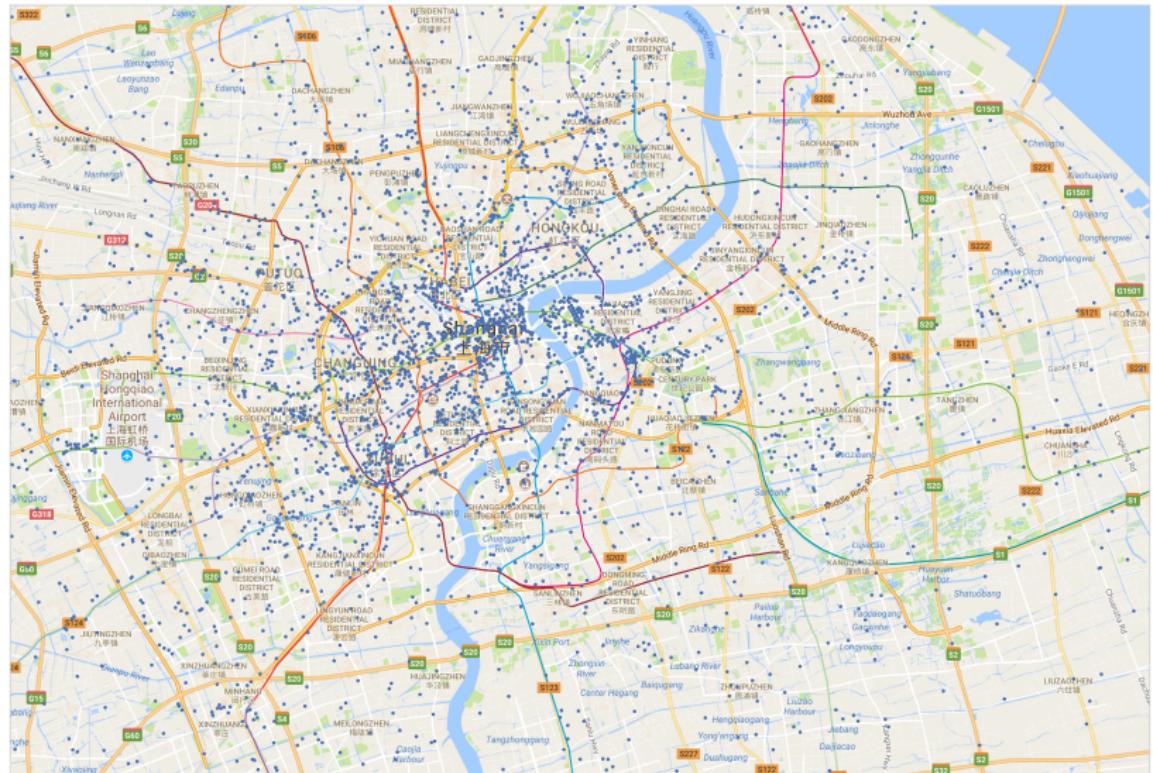


Figure: Locations of Base Transceiver Station (BTS)

A	B	C
PROD_INST_NUM	START_TIME	BASE_STATION
手机号码	开始时间	基站编号
XXXXXXXXXXXXXXXXXXXX	12/17/15 6:51	C0079
XXXXXXXXXXXXXXXXXXXX	12/17/15 7:02	C0078
XXXXXXXXXXXXXXXXXXXX	12/17/15 7:12	C0753
XXXXXXXXXXXXXXXXXXXX	12/17/15 15:27	C8398
XXXXXXXXXXXXXXXXXXXX	12/17/15 15:58	C0753
XXXXXXXXXXXXXXXXXXXX	12/17/15 20:18	C8398
XXXXXXXXXXXXXXXXXXXX	12/17/15 20:49	C0753
XXXXXXXXXXXXXXXXXXXX	12/17/15 0:27	C8396
XXXXXXXXXXXXXXXXXXXX	12/17/15 2:13	C1643
XXXXXXXXXXXXXXXXXXXX	12/17/15 2:23	C8396

Figure: Raw Mobility Records

	A	B	C	D	E	F	G	H	I
1	Monthly Aggregate Data (100k)				Daily Aggregate Data (10k)				
2	Id	YM	BTS	Duration		Id	YMD	BTS	Duration
3	1	16/01	1	5h		1	16/01/01	1	50m
4	1	16/01	2	3h		1	16/01/01	2	40m
5	1	16/01	3	2h		1	16/01/01	3	20m
6	1	16/01	4	0.5h		1	16/01/02	1	60m
7	1	16/01	5	0.2h		1	16/01/02	2	20m
8						1	16/01/02	3	15m
9						1	16/01/02	4	10m
10						1	16/01/02	5	5m
..									

Figure: Aggregate Data Collected

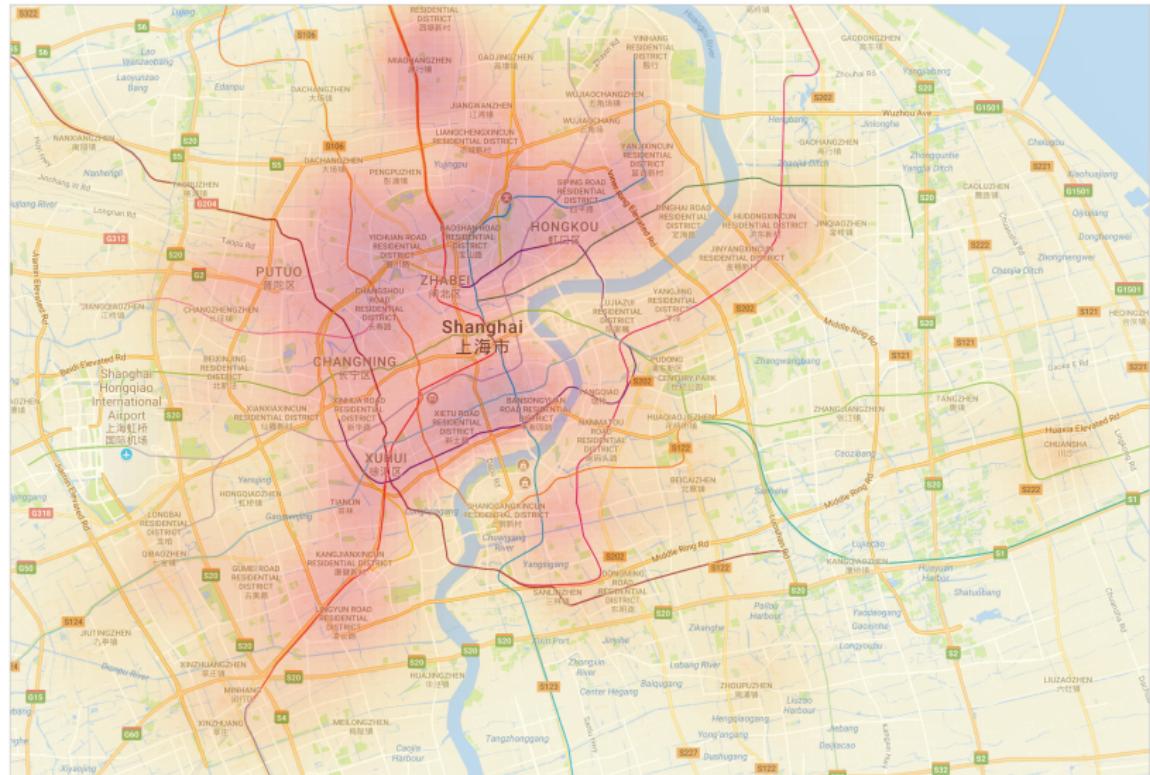


Figure: Distribution of Home Locations

	uid	Gender	Age	IDCity	Pmt1702	IntTraffic	IntDur	IntEvent	Phone	Car	UserCluster	uniNum	uniNumIn
0	1	1	67	310222	13	555	184	379	iPhone 6s Plus	0	中高端语音_后付费	78	66
1	2	2	52	310104	56	109	188	562	iPhone 6 Plus	0	中端_后付费	38	33
2	3	1	47	310227	153	448	322	763	iPhone 6s	0	中高端语音_后付费	-1	-1
3	4	1	52	310225	66	273	373	664	iPhone 6s Plus	0	中高端语音_后付费	136	92
4	5	1	-1	310229	155	830	283	675	FRD-AL00	0	中高端语音_后付费	78	52

Figure: User Characteristics

Reserach Projects

- ▶ Optimal Design of Commercial Districts (Mandy Hu, Lei Xu)
 - ▶ CD characteristics: including price, positioning, store composition
- ▶ Effect of Weather/Pollution on Location Choices (Mandy Hu, Matthew Osborne (Toronto), Lei Xu)
 - ▶ Focus on consumer dynamic choices in response to weather & pollution: e.g. intertemporal substitution

Optimal Design of Commercial Districts

- ▶ Gov policy: high-level CD planning
- ▶ Decisions: number, location, positioning
- ▶ Local CD Manager: store composition: restaurants, movie theaters, etc
- ▶ Externality on other CDs

Other Data Collected

- ▶ Dianping.com: Chinese TripAdvisor
- ▶ detailed store information: location, CD, type, price, reviews, isClosed, time when store is listed, etc

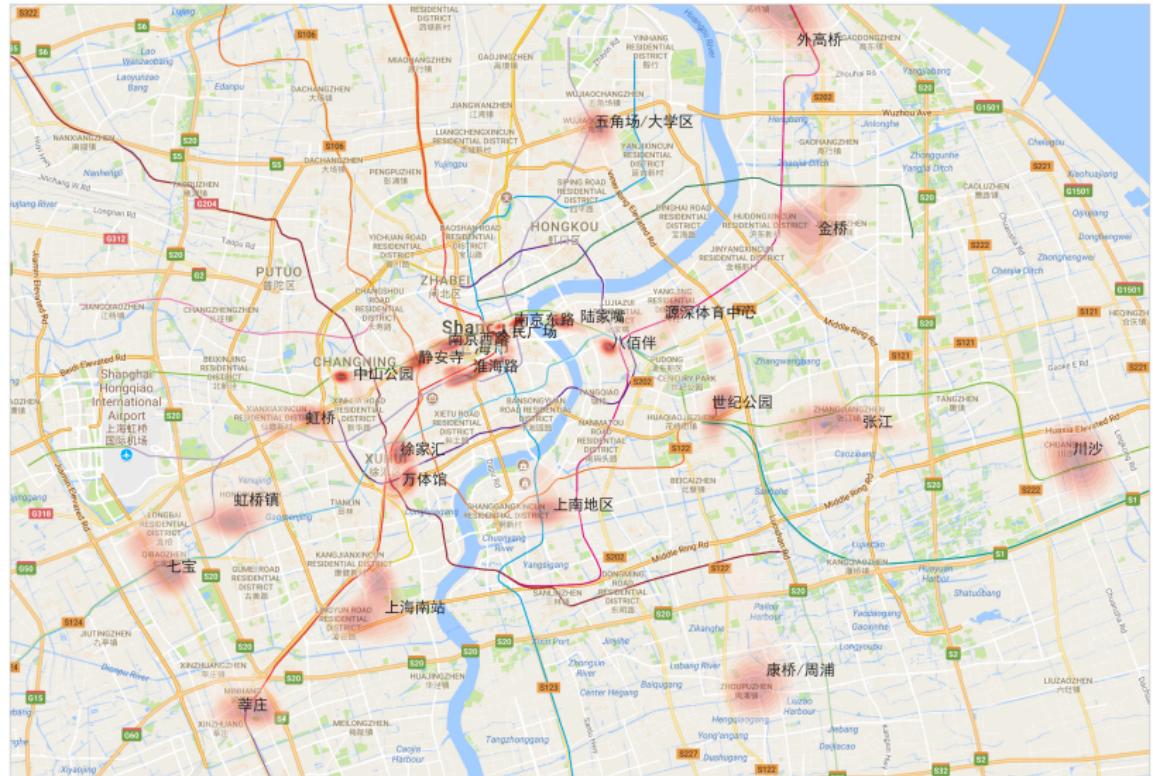


Figure: Business Clustering by Commerical Districts (CD)

Distance Measure

- ▶ Number of BTS: 5046
- ▶ BTS Pairs: 13 million

	OD	DDist	DDur	DTraffic	TDist	TDur	Direct
1	58506175	6359	604	749	4535	2861	3582
2	58505910	1046	191	201	852	648	640
3	58506033	8136	861	1096	7427	3904	6062
4	58506128	7448	830	1043	5117	2958	4155
5	58506079	7705	973	1089	8689	3429	5929
6	58506101	8425	885	1123	9231	3911	6508
7	58506049	9252	1067	1338	9587	3101	7079
8	58506316	3265	626	695	2802	1924	2182
9	58508398	1370	268	278	899	690	750
10	58506317	8000	795	1022	7334	3839	5803

Basic Logit Model

$$s_{idt} = \frac{\exp(U_{idt})}{1 + \sum_{d' \in D} \exp(U_{id't})}$$

$$\begin{aligned} U_{idt} = & \alpha_i + \beta_1 Gender_i + \beta_2 Age_i + \beta_3 \bar{p}_{dt} + \beta_4 Shop_{dt} \\ & + \gamma_1 Dist_{id} + \gamma_2 Dist_{id}^2 + \epsilon_{idt} \end{aligned}$$

where

- ▶ consumer i \in 1, 2, 3, ..., 100k
- ▶ time t: 2014/08 - 2017/02
- ▶ commercial district: d \in D={1,2,3, }

Problems

- ▶ Endogeneity: Shop entry/exit is endogenous
 - ▶ Though some government level planning, each shop still needs to decide whether to enter or not
- ▶ Modeling: Logit is for discrete choice, not continuous choice
 - ▶ Comparison to Nested Logit: only CD choices, no shop choice data
 - ▶ Comparison to Dubois et al. (2014): treat each CD as an item with various characteristics such as types of shops, etc