LOCATIONAL ANALYTICS AND BIG DATA: FOUNDATIONS, EMERGING APPLICATIONS AND RESEARCH

Overview of Spatial Big Data and Analytics

(9:10-9:55am)

Applications and Examples of Spatial Big Data and Analytics

(9:55-10:15am)

Break

Research Opportunities in Location, Analytics, Big Data and GIS for IS Researchers

(10:25-10:45am)

Research Opportunities in Spatial Big Data, Location and Analytics for IS Researchers

Namchul Shin
Pace University
New York, NY, USA
nshin@pace.edu

Pre-ICIS Workshop on Locational Analytics and Big Data: Foundations, Emerging Applications, and Research

Sponsored by SIGGIS
Association for Information Systems

Fort Worth, Texas, December 13, 2015

Definition of Spatial Big Data

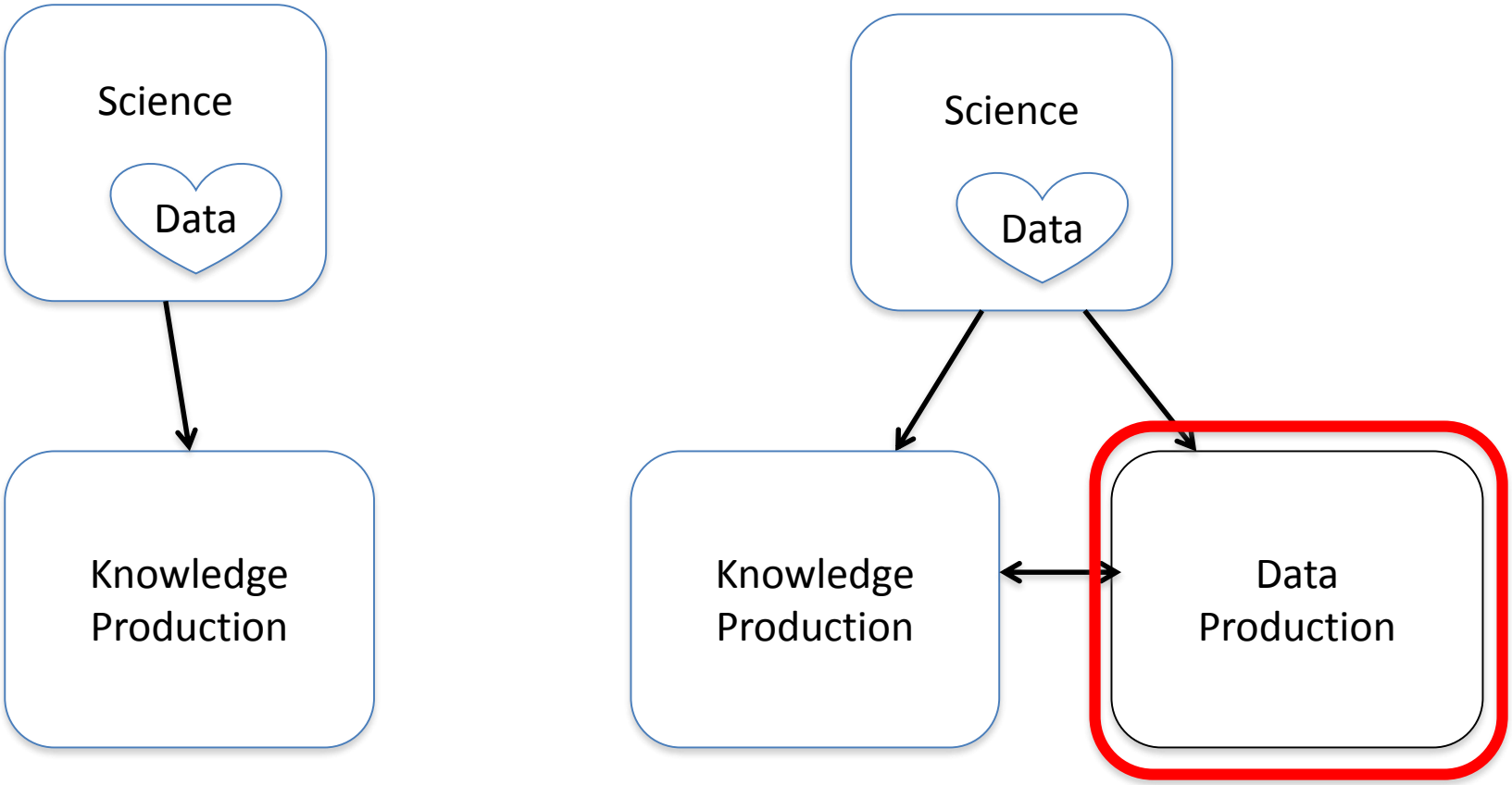
- Big Data are “data sets that are so big they cannot be handled efficiently by common database management systems” (Dasgupta, 2013).
- Spatial Big Data represents Big Data in the form of spatial layers and attributes.
  - There is no standard threshold on minimum size of Big Data or Spatial Big Data, although big data in 2013 was considered one petabyte (1,000 terabytes) or larger (Dasgupta, 2013).

Sources of Spatial Big Data

- *Sources of Spatial Big Data include:*
  - GPS, including GPS-enabled devices
  - Satellite remote sensing
  - Aerial surveying
  - Radar
  - Lidar
  - Sensor networks
  - Digital cameras
  - Location of readings of RFID

(Partially based on Dasgupta, 2013)

Data Production (Data-Driven Science)



Data Production Practices

- Collecting and analyzing
- Processing and managing
- Assembling and organizing
- Preserving and curating
  - Generation of meta data
  - Provenance information

Spatial Big Data Analytics

- Tobler's first law of geography
  - "Everything is related to everything else, but near things are more related than distant things."
- Power of location
  - Location targeting improves the performance of mobile advertising, e.g., Foursquare.
- Grand challenges, such as sustainability and climate change, health, transnationally organized crime, energy, economic development, etc.
  - For example, eco-routing, rather than faster routing

IS Research Areas for Spatial Big Data, Location and Analytics

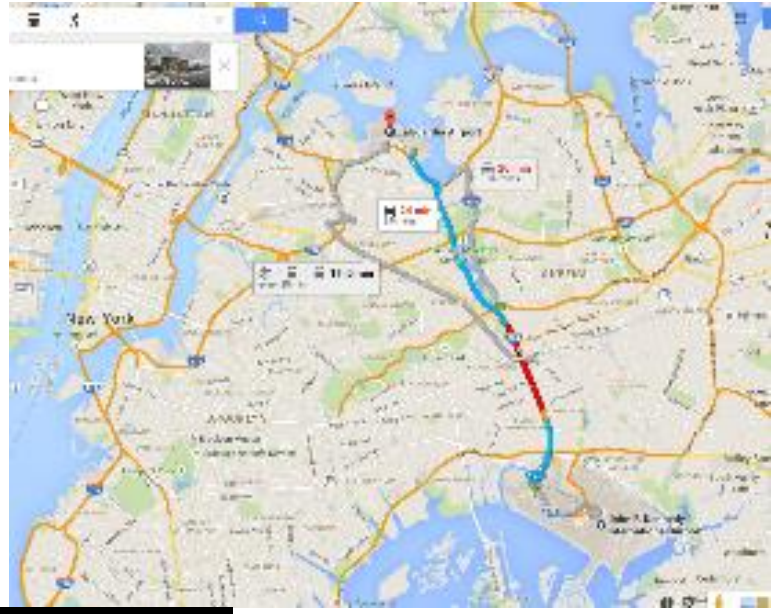
- Politics
- Urban Traffic
- Healthcare
- Spatial business intelligence
  - Spatial data mining and knowledge discovery
- Collaborative spatial decision-support systems
- Location-based services (LBS)
  - Emergency services
  - Tracking/monitoring services
  - Information services
- Marketing/Advertising (a type of LBS)
  - Location-based mobile promotion: personalized and context-aware services
- Supply Chain Management
  - Tracking and monitoring
- Privacy, security, and ethics
- Business value of spatial big data

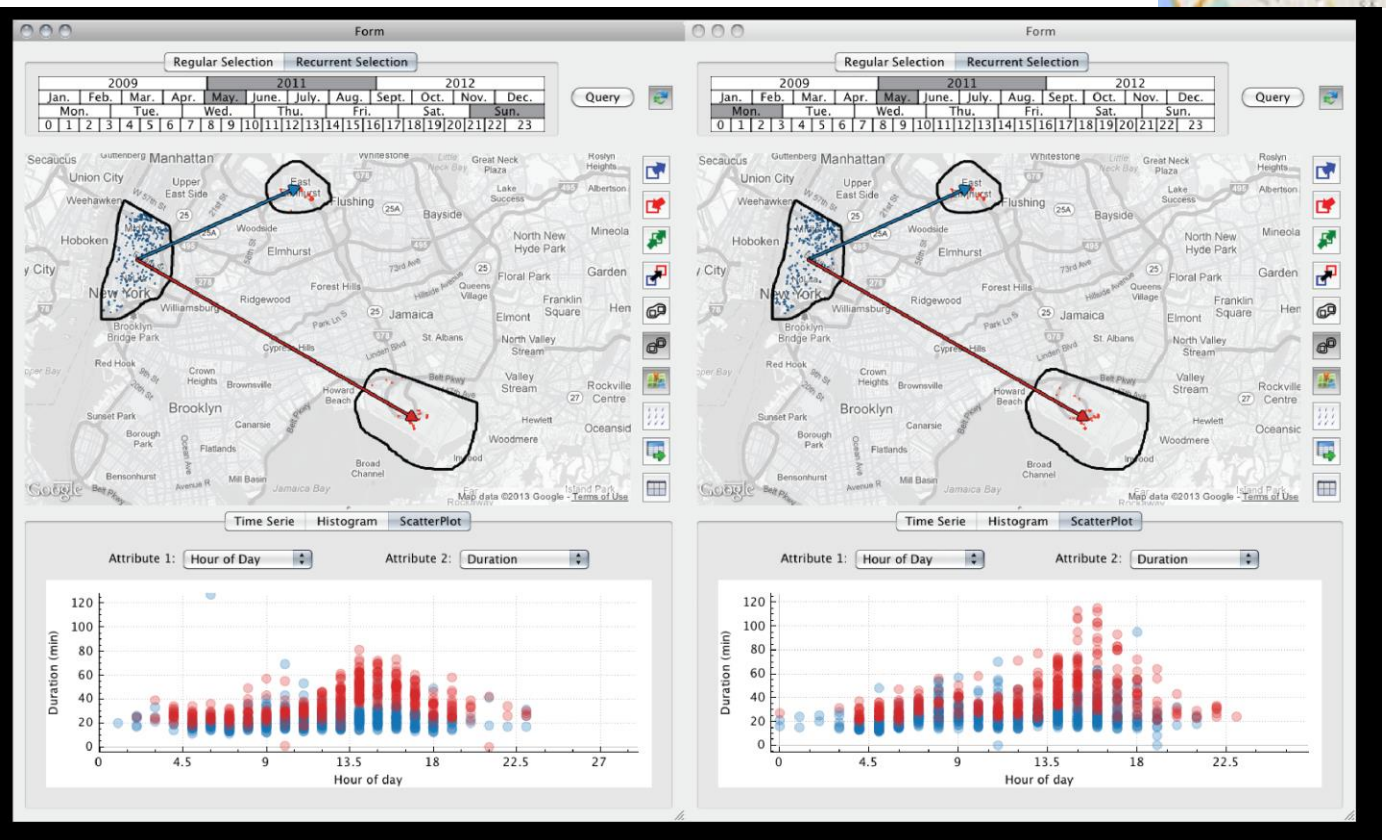
Lack of Research on Spatial Big Data, Location and Analytics

- Little research on spatial big data and analytics published in major MIS journals.
- Studies were published in other journals, but were mostly conceptual.
  - Few exceptions, e.g., Ferreira, Poco, Vo, Freire, and Silva (2013), and Tsou, Yang, Lusher, Han, Spitzberg, Gawron, Gupta, and An (2013)

Spatial Big Data – Example of Locations and Movement of Central New York City Taxicabs, based on space, time, and attributes

TaxiVis: a system supporting visual exploration of big spatial and temporal data (origin-destination)
A user-friendly interface (visual query model) allows users to view and analyze the patterns and movements of 500,000 taxi trips daily in central NYC.
The data from NY Taxi and Limousine Commission gives pickup and drop off locations, time, and attributes.



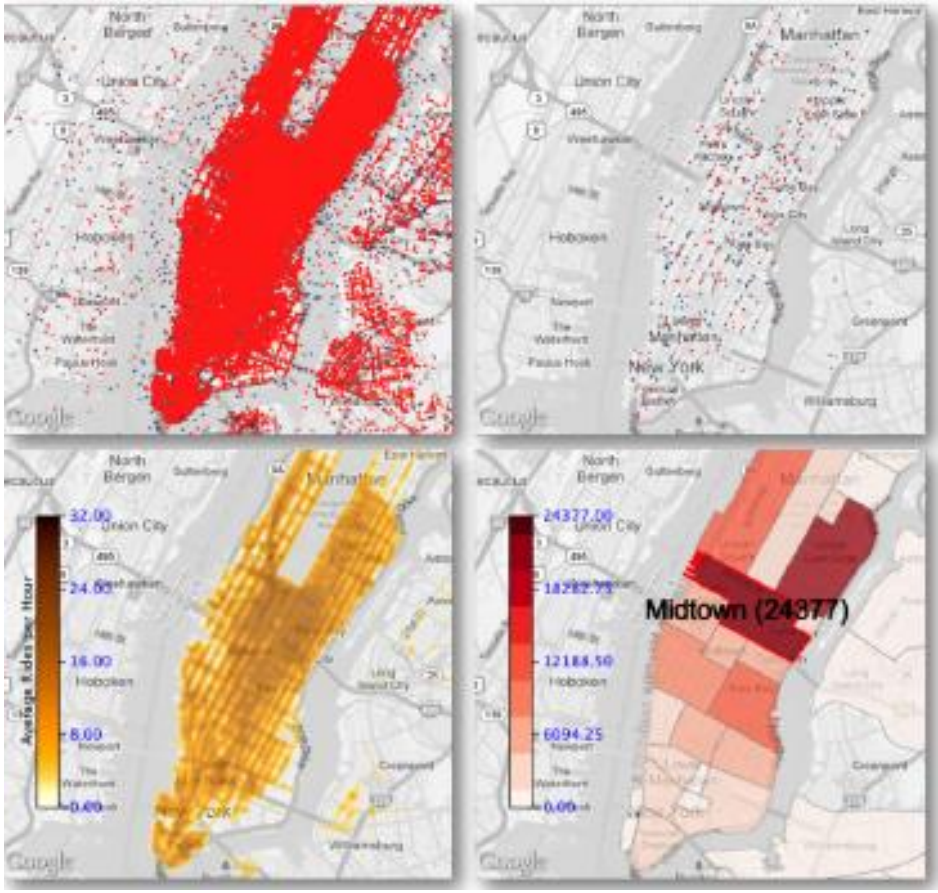


The example shows taxi trips from lower Manhattan area to LaGuardia airport area (upper part of image) and Kennedy airport area (lower part). The volume of trips are given in the lower hourly graphs for Sundays in May 2011 (left) and Monday (right), with blue for LaGuardia and red for Kennedy.

(Source: Ferreira et al., 2013)

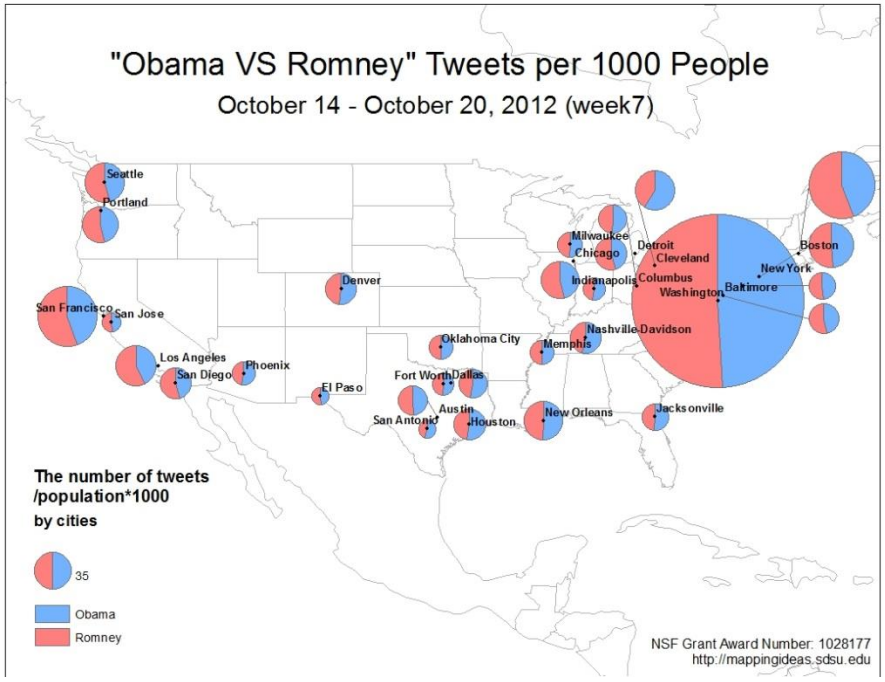
New York City Taxi example – further capabilities

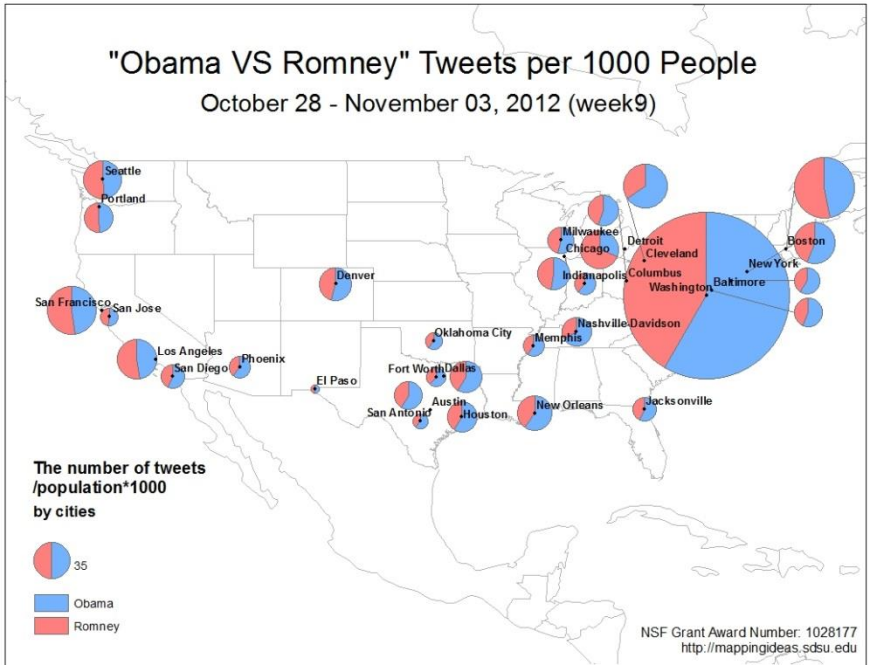
- Side-by-side “sensor” maps over time
  - Visual queries for pick-up AND drop-off (point cloud)
- Level-of-detail reduced the number of points
  - Done by hierarchical sampling of point clouds
- Density heat maps and grid maps
- Constraints of attributes of taxi id, distance traveled, fare, and tip amount
  - Enables economic analysis
- Different visualizations can be derived with filters
- Complex queries
  - Use set-theoretic functions on simple queries



(Source: Ferreira et al., 2013)

Spatial Big Data – Example of Obama vs. Romney Tweets





VISION (Visualizing Information Space in Ontological Network): tracks ideas and social events disseminated in cyberspace from a spatial-temporal perspective.

Case example of Spatial Big Data using social media (a live feed of number of tweets with “Obama” keyword and with “Romney” keyword for largest 30 U.S. cities from Oct. 14-Nov 3, 2012).

Shows the period before Hurricane Sandy hit East Coast (it hit on Oct. 29 and ended on Nov. 5)--There is a major shift towards Obama during this two week interval, which is more prominent in the northeast.

Most tweets originate with mobile devices. Errors include re-tweeting, robot tweets, city definitions, and positive or negative emotion of the tweet.

(Source: M-H. Tsou, et al., 2013)

Spatial Big Data for 2012 Presidential Election

- Data source: millions of tweets were examined and analyzed for the same keywords.
    - Techniques used were “commercial web search engines (Yahoo and Bing APIs), Twitter search engine API, IP geo-location methods, and GIS software functions of kernel density and raster-based map algebra methods” (Tsou et al., 2013).
- Privacy is opt-in
    - Locational referencing for Twitter is an opt-in service, so when a user decides to use Twitter, he/she is legally accepting the locational referencing option.
- Valence of “Obama” and “Romney” tweets was unknown
    - A limitation is that whether the candidate was being referred to favorably or unfavorable was unknown. Results were interpreted as positive valence, but there is a data quality issue present.
    - Although the emotion is not captured, more sophisticated natural language processing could possibly capture it.
    - Need more comprehensive data analysis methods and data cleaning algorithms.

Articles on Spatial Big Data, Location and Analytics Published in Selected Leading MIS Journals

	2011-2015		1998-2012
Journals	Spatial Big Data	Big Data	GIS
MIS	0	0	2
ISR	1	1	1
EJIS	1	1	0
JMIS	1	0	1
DSS	6	0	13
CACM	1	11	4
CAIS	0	3	2
Tele Policy	1	0	5
Tech Forecast	4	2	5
Interfaces	1	1	2

Articles on Spatial Big Data, Location and Analytics Published in Selected Leading MIS Journals (2011-2015)

MISQ (none)

ISR (1 article)

Fang, Zheng, Bin Gu, Xueming Luo, and Yunjie Xu. 2015. Contemporaneous and delayed sales impact of location-based mobile promotions. *ISR* 26(3): 552–564

EJIS (1 article)

Constantiou, Ioanna D., Christiane Lehrer and Thomas Hess. 2014. Changing information retrieval behaviours: an empirical investigation of users' cognitive processes in the choice of location-based services. *EJIS* 23(5): 513-528.

JMIS (1 article)

Shi, Zhan and Andrew B. Whinston. 2013. Network Structure and Observational Learning: Evidence from a Location-Based Social Network. *JMIS* 30(2): 185-212.

Articles on Spatial Big Data, Location and Analytics Published in Selected Leading MIS Journals (2011-2015)

Decision Support Systems (6 articles)

A personalization process for spatial data warehouse development

The personalization privacy paradox: An exploratory study of decision making process for location-aware marketing

Leveraging location-based services for couponing and infomediation

On the brink: Predicting business failure with mobile location-based checkins

Building a targeted mobile advertising system for location-based services

The effects of location personalization on individuals' intention to use mobile services

Communications of the ACM (1 article)

Challenges and Business Models for Mobile Location-based Services and Advertising

Supplementary Literature Review

- Papers published in second-tier journals in both MIS and other related fields (2011-2015)
  - Communications of the AIS (0)
  - Telecommunications Policy (1)
  - Technological Forecasting and Social Change (4)
  - Interfaces (1)
- Not many papers published as well
- Focused on spatial analysis and locations

Supplementary Literature Review

- Papers (on Location and Analytics) published in GIS journals (2011-2015)
  - Transactions in GIS (8, 0)
  - International Journal of GIS (17, 5)
  - Applied Geography (16, 3)
  - The Professional Geographer (4, 0)
- Lots of papers published were on geospatial (or spatial) topics.
- There were no papers published on big data

Examples of Research from Leading MIS Journals

Fang, Zheng, Bin Gu, Xueming Luo, and Yunjie Xu, ISR 2015

- Approach: randomized field experiment for sales impact (contemporaneous and delayed purchases of a movie ticket) of mobile location-based promotion (LMP): LMP in the geo-fenced area (treatment) and LMP not in the geo-fenced area/no LMP in the geo-fenced area (two control groups).
- Research question: sales impact of location-based mobile promotion
- Theory: real-time marketing and planned behavior
- Method: logistic regression
- Findings:
  - LMP makes a significant impact on both contemporaneous and delayed sales purchases (importance of location and temporal proximities, but also motivation of need recognition for planned behaviors)
  - Importance: research in an actual field setting (actual customer sales)/research not only on same-day purchases, but also subsequent-days sales, which was not studied before.
- Future research: other location data for customers, e.g., where they live?/research on other product categories, such as restaurants or retailing?

Examples of Research from Leading MIS Journals

Constantiou, Ioanna D., Christiane Lehrer and Thomas Hess, EJIS 2014

- Approach: interviews and dairy survey of young adults (non-random sample; higher probability of location-based services (LBS) use in everyday life) in the German mobile telecommunications market (one of the most advanced markets in Europe)
- Research question: what are the cognitive processes involved in the choice of LBS and how do these processes influence information retrieval behaviors?
- Theory: Theory of behavioral decision-making
- Method: descriptive statistics of qualitative data—semi-structured interview data (40 participants) and diary survey data (103 instances of 16 participants) of daily LBS use
- Findings:
  - LBS use is the outcome of intuitive processes—1) availability heuristics are likely to be used by experienced users. 2) representativeness heuristics are likely to be used by novice users.
  - Comparative processes (perceived value in contexts of use) are also used for LBS use, for example, when time-sensitive information is needed (or while on the move), people highly appreciate the positioning functionality and information accuracy of LBS.
  - In a dynamic context of use, mobile users are willing to pay for an LBS when it addresses unmet needs and enables new information retrieval routines.
- Future research: A large-scale quantitative study or longitudinal study for the choice of LBS

Examples of Research from Leading MIS Journals

Shi, Zhan and Andrew B. Whinston, 2013

- Approach: Analysis of primary data collected from a major location-based social networking Web site in China
- Research question: How does friends' activity sharing in social media help users make their own economic decisions on experienced goods (visiting venues, such as restaurants, shopping centers, nightlife sites, and tourist attractions)?
- Theory: observational learning
- Method: log-log regressions (maximum likelihood method)
- Findings:
  - The proportion of checked-ins made by friends is not positively associated with the likelihood of a new visit.
  - However, the network proximity-weighted proportion of check-ins made by friends (close friends) is a better indicator of the likelihood of a new visit; repeated check-ins by friends have a pronounced effect.
  - Internet search domain incorporates individual identity and social relationships into the methodology that determines search results; users may see a personalized list based on his or her network (e.g., Facebook) friends' "likes" in addition to a list based on popularity by anonymous others.
- Future research:
  - A measure for similarity of tastes of friends (other measures of social network proximity?)
  - A longitudinal study
  - Difference in learning effect for different types of venues (e.g., restaurants versus shopping centers)

Current Gaps and Limitations

- Data quality (citizen science)
  - Big data has very low density in value in itself
  - Biased (user-generated content/volunteer geographic information)
- Small data versus big data
  - Marginalization of small data studies
  - What data are captured is shaped by the technology used, the context in which data are generated and the data ontology employed (Kitchin, 2013).
- Evolving analytics for spatial big data
  - When to analyze whole unstructured big data-set versus analyzing selective structured slices.
  - New and evolving analytic techniques for spatial and non-spatial dimensions of big data.
- Corporate secrecy and proprietary limitations.
- Corporate case studies
- Need research about spatial big data as well as studies using spatial big data.

Summary: Research on Spatial Big Data, Location and Analytics

- Since the preponderance of data is, or can be, geo-referenced, the size of spatial big data is vast.
- The limited documented examples illustrate the power of location and discovery aspects.
- Computer Science and GIScience are taking the lead.
- Analytics are needed since the extent of map visualization is overwhelming.
  - A more comprehensive data analysis
  - Analysis with the combination of location and other attributes
- There are lots of questions and much future work to be done
  - Business models for location-based mobile promotion or services for different stakeholders
  - Location-based social network promotion
  - Post-adoption (or continued use) of location-based services
- MIS has an important role to play…….