Multivariate and Geospatial Analysis of ICT Utilization in US States

James B. Pick
Avijit Sarkar
Jeremy Johnson

2014 Esri User Conference
San Diego CA
July 2014

Goals of research

- Examine the digital divide in the United States.
  - Enhance understanding of factors associated with availability and utilization of **information and communication technologies (ICTs)** at the **state level** in the US.
- Digital divide: "*the gap between individuals, households, businesses and geographic areas at different socio-economic levels with regard both to their opportunities to access information and communication technologies and to their use of the Internet for a wide variety of activities*'' (OECD, 2011).
- Overall research question
  - What factors determine ICT adoption and utilization for US states and how do ICT adoption and utilization vary geographically?

Organization

- Introduction
- Literature/prior research
- Conceptual model
- Research questions
- Methodology
- Findings
- Discussion and Practical Implications
- Conclusion

Introduction: ICT Landscape in USA

ICT	Users (mil)	% Penetration	World Ranking	Behind
Landlines	**146 (2011)**	47	**2**	China
Mobile cellular	**290.3 (2011)**	94.25	**3**	China, India
Internet users	**245 (2009)**	78.3	**2**	China
# of Internet hosts	**505 (2012)**		**1**	------
Facebook (subscribers)	**166 (June 2012)**	53	**1**	------
Internet Penetration		78.3% (end of 2011)	**27**	Nordic nations, South Korea, Japan, Germany, UK, Netherlands, etc.
Broadband penetration	NA	22	**7**	Netherlands, South Korea, Sweden, Canada, UK, France
ICT Development Index (IDI)	IDI = 7.53 Highest IDI = 8.57 (South Korea)		**17**	S. Korea, Nordic nations, Japan, UK, Netherlands, Australia, Switzerland, Singapore, New Zealand, etc.

Sources:
ITU,
Internet World Stats,
CIA World Factbook

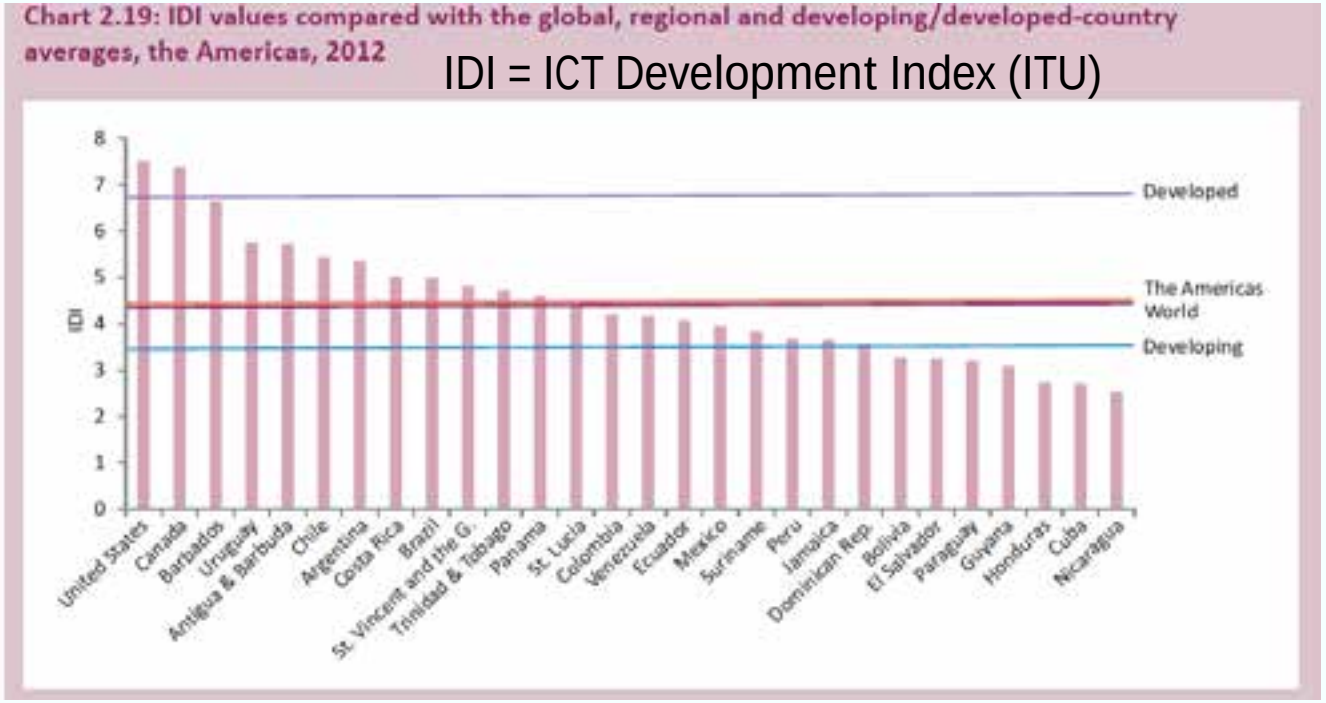
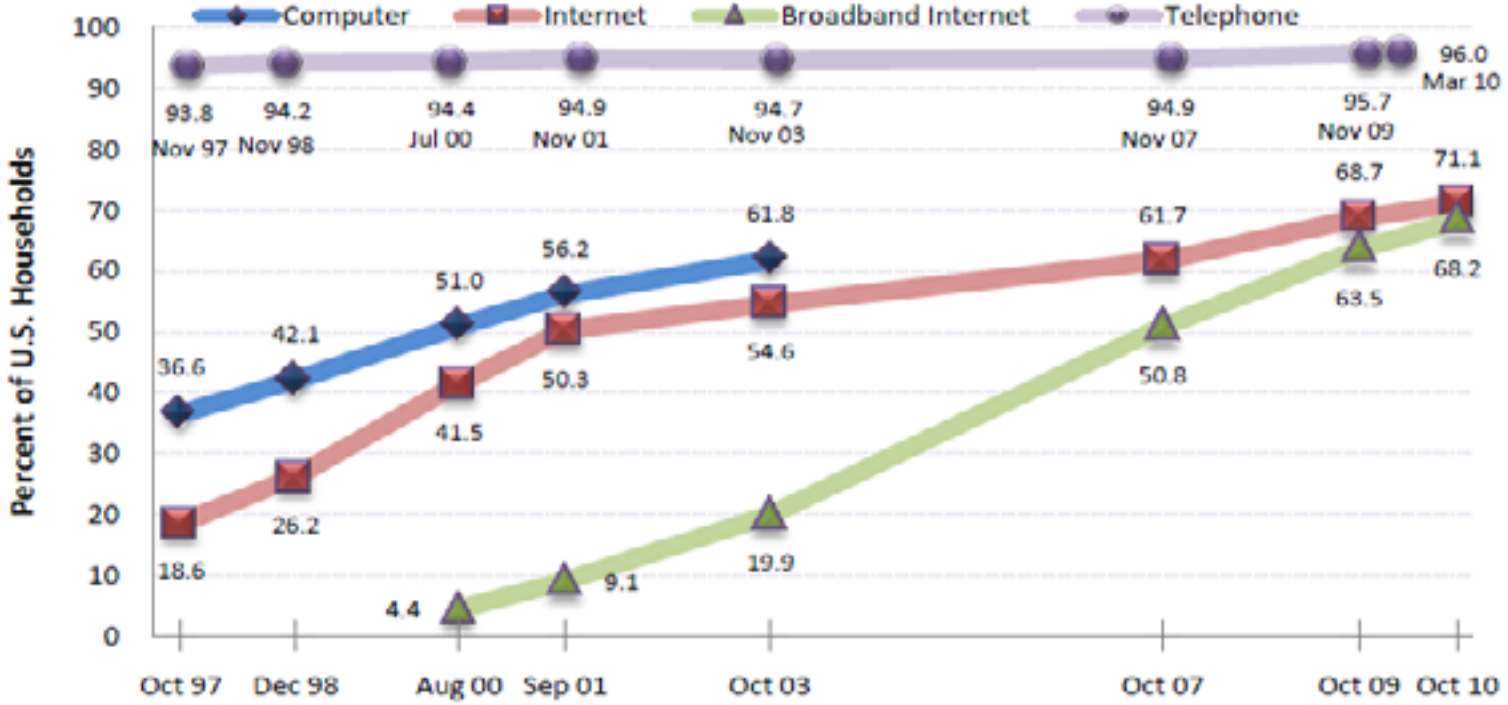
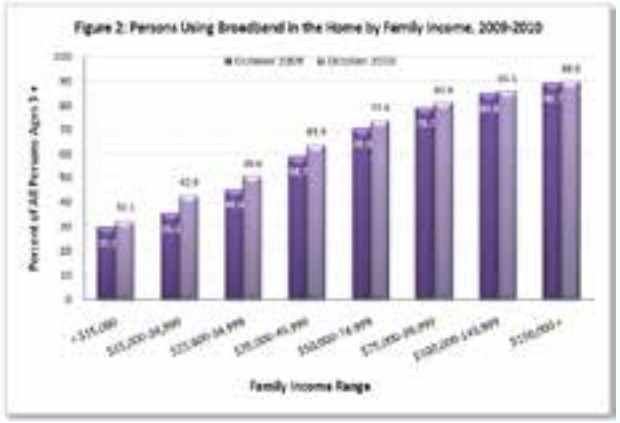
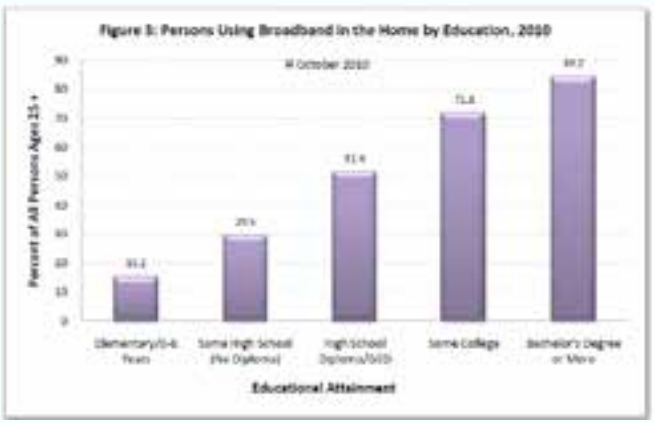


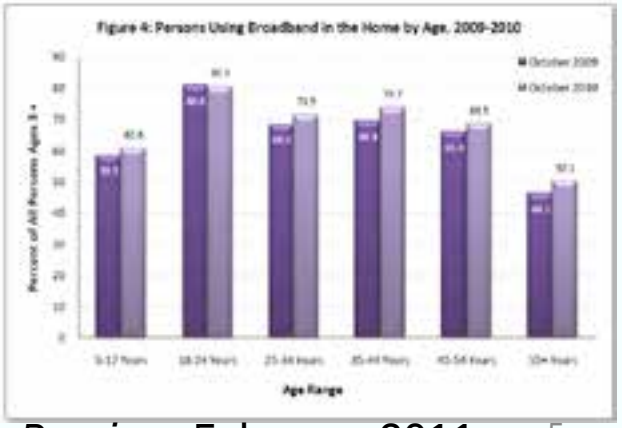
Figure 1: Households with Computers, Telephone Subscriptions, and Internet Access, Selected Years, 1997-2010*



* Note: 2001-2010 use 2000 Census-based weights and earlier years use 1990 Census-based weights.







Source: Digital Nation: Expanding Internet Usage, *NTIA Research Preview,* February 2011

Literature/prior research

- Digital divide differences have been studied extensively from a variety of perspectives.
- Significant empirical literature explains influences on access, adoption, and use of ICTs.
- Important determinants of ICT adoption and utilization from prior studies are:

q	Education	q	Income	q	Age	q	Ethnicity
q	Urban location	q	Higher education govt funding	q	ICT expenditure	q	Infrastructure
q	Innovation & R&D	q	Newspaper, magazine, & book publishers	q	Labor force	q	Societal openness

- For United States
  - U.S. Studies conducted for a variety of geographies
    - Individuals & households (Chen, 2013; Sipior, Ward, & Connolly, 2011).
    - Counties in U.S. (Azari & Pick, 2004, 2005)
    - Specific cities (Kvasny & Keil, 2006
    - National small areas (Grubesic, 2006)
  - Digital have-nots are often on the wrong side of social inequalities (Chen, 2013).
  - Digital divide will persist in the absence of remediation of social, economic, and locational inequities (Kvasny & Keil, 2006).
- Digital divide studies at the state/provincial/prefectural level exist for India, China, Japan, Europe.

Research Questions

1. Are geographical patterns of ICT access and use present for the U.S. states as measured by spatial auto-correlation?
2. Is there significant geographic clustering of states based on the access and use of ICTs?
3. What are the associations of socio- economic, political, demographic, innovation, and societal openness factors with access and use of ICTs?
4. Can a regression model of these associations on access and use of ICTs account for the influence of the geographical proximity of states?

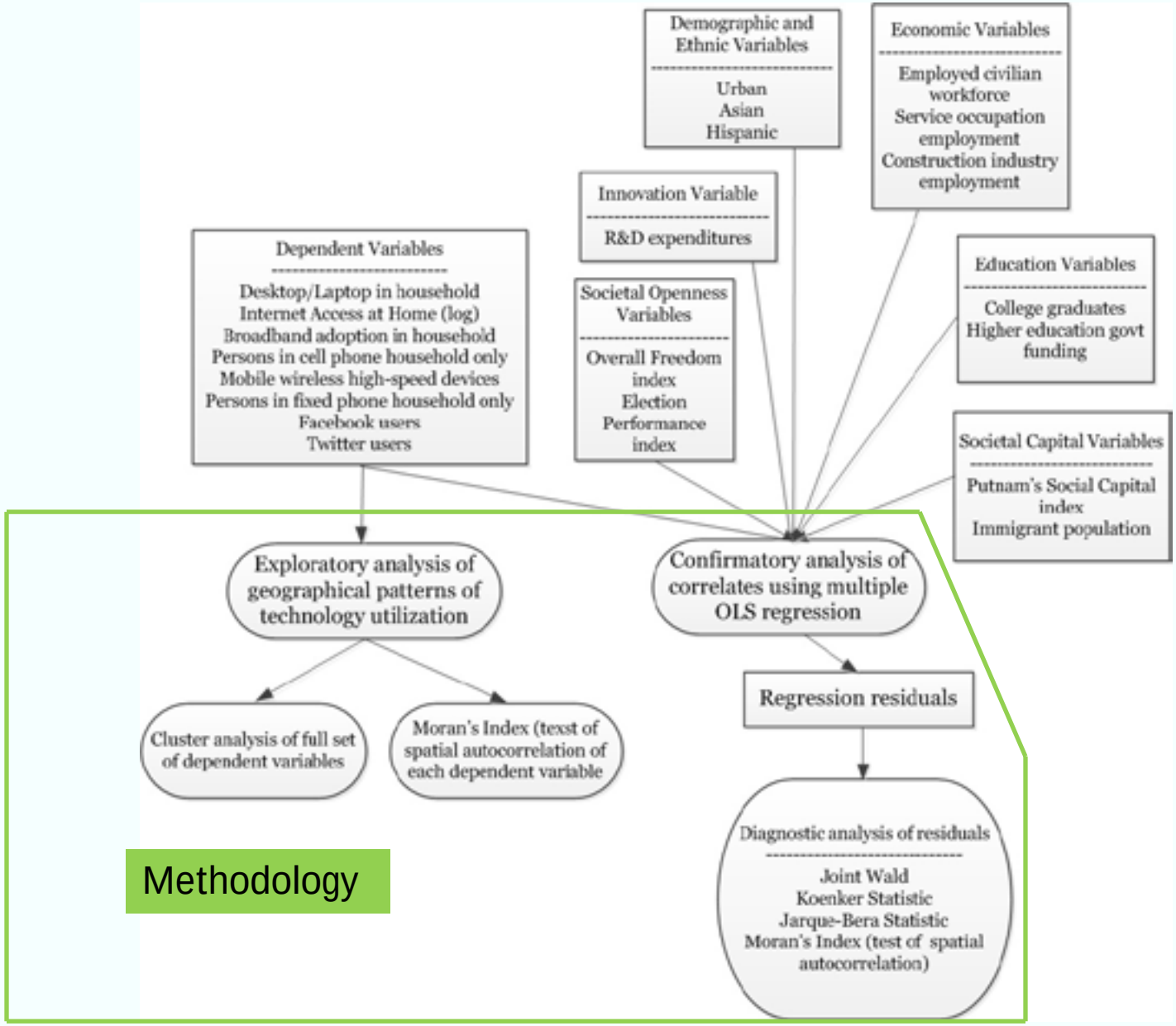
Unique features of this work

1. Examination of factors that influence the digital divide at the state level in the US does not exist.
   - Empirical analysis of ICT utilization **at state level** has not been conducted for US.
2. Inclusion of social media technologies of Facebook & Twitter among dependent variables.
3. Spatial analysis (clustering and detection of spatial autocorrelation that may bias regression results) are new to MIS research.

Conceptual model

- Our conceptual model is broadly based on:
  - Prior literature on institutional factors in technology utilization and innovation (King et al., 1994),
  - A theoretical model that builds on these factors (Simon, 2004), and
  - On literature on government support for IT, legal framework, and social openness (Baliamoune-Lutz, 2003, Pick and Azari, 2011, Pick and Nishida, 2013).
- Induction from prior studies is utilized to establish each construct in the conceptual model.
- The model consists of independent variables that influence 8 dependent ICT utilization variables, with exploratory screening using spatial analysis methods.

Conceptual model of ICT Utilization



Dependent & Independent Variables

Category	Variable	Source*	Year	Definition
DEPENDENT VARIABLES				
Technology Use	Desktop/Laptop in Household	CENCPS10	2010	Persons three years or older who live in household with desktop or laptop
Technology Use	Internet Access at Home	CENCPS10	2010	Log of persons three years or odler who access the Internet at home
Technology Use	Broadband Adoption in Household	NTIA	2010	Percent of totalhouseholds with broadband adoption
Technology Use	Persons in Cellphone-only Households	NCHS	2010	Percent of persons 18 years or older living in households with wire-less only phones
Technology Use	Mobile Wireless High-Speed Devices	FCC	2008	Subscribers to mobile wireless high-speed devices per capita
Technology Use	Persons in Fixed-phone-only Household	NCHS	2011	Percent of persons 18 years or older living in households with fixed-phones only
Technology Use	Facebook Users	DCI	2010	Facebook users per capita
Technology Use	Twitter Users	DCI	2010	Twitter users per capita
INDEPENDENT VARIABLES				
Demographic	Urban	CENPOP10	2010	Urban percent of total population
Demographic	Asian	CENPOP10	2009	Asian percent of total population
Demographic	Hispanic	CENPOP10	2009	Hispanic/Latino percent of total population
Demographic	Black	CENPOP10	2009	Black percent of total population
Social Capital	Putnam's Social Capital Index	Putnam (2000)	2000	Putnam's Social Capital Index by state
Economic	Employed Civilian Workforce	CENPOP10	2010	Total employment in civilian occupations per population 16+ in labor force
Economic	Service Occupation Employment	CENPOP10	2009	Total employment in service occupations per population 16+ in labor force
Economic	Construction Industry Employment	CENPOP10	2009	Total employment in construction occupations per population 16+ in labor force
Education	College Graduates	CENPOP10	2009	Population with any college degreee (associate, bachelor, graduate) per population
Innovation	R&D Expenditures	NSF	2008	R&D expenditures/gross state product
Innovation	Publisher Annual Sales (Newspaper, Periodical, Book, Directory)	CENECON07	2007	Sales revenues of newspaper, periodical, book, and directory publishers/annual sales revenues
Societal Openness	Overall Freedom Index	MERCATUS	2009	Overall freedom index
Societal Openness	Election Performance Index	PEW	2008	Pew index measuring state elections performance

* DCI = Digital America Report, DCI Group

FCC = Federal Communications Commission

MERCATUS = Marcatus Center, George Mason University

NCHS = National Center for Health Statistics

NTIA = National Telecommunications and Information Administration

NSF = National Science Foundation

PEW = Pew Research Center

Spatial Autocorrelation Patterns Measured by Moran's I

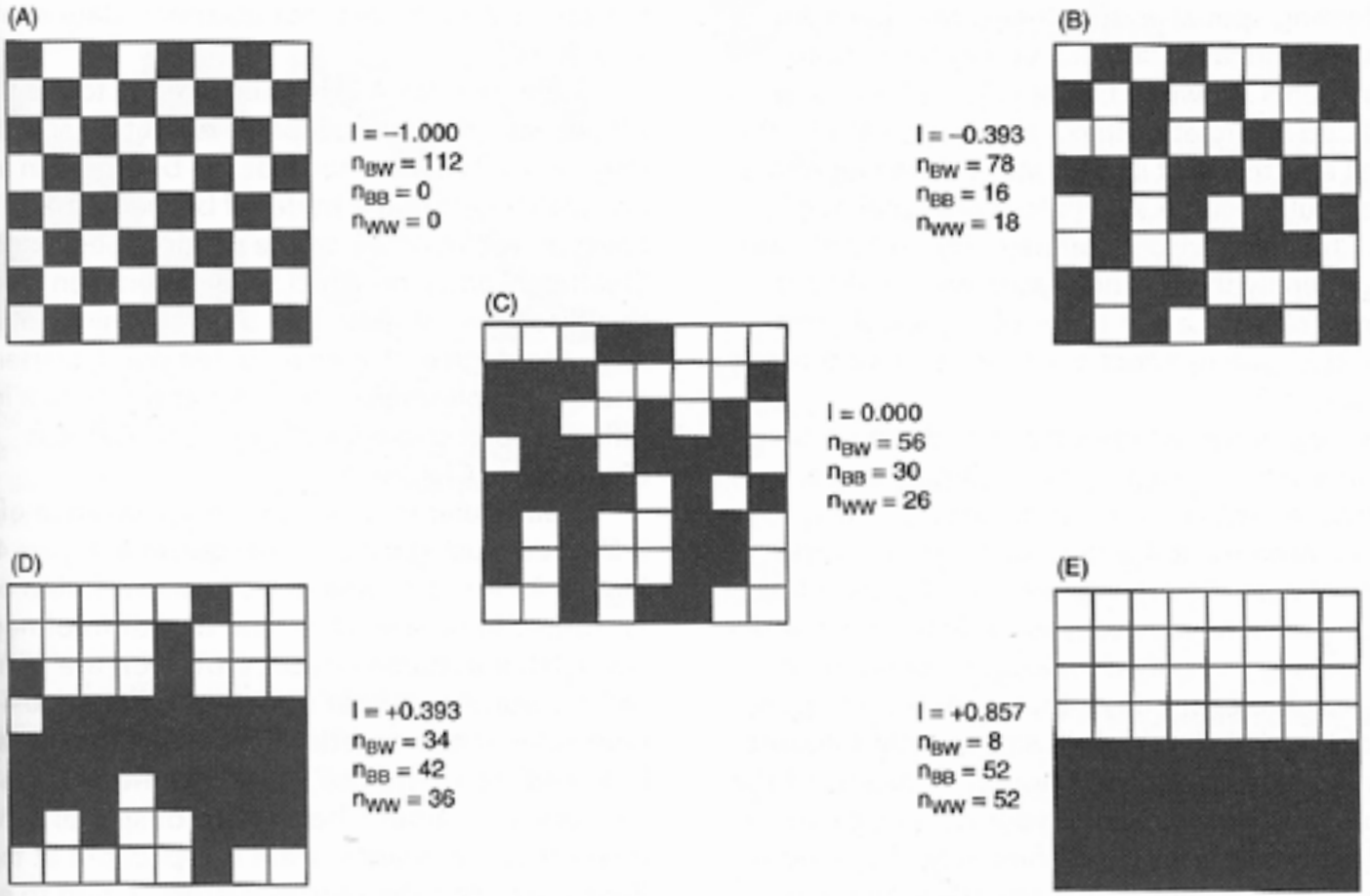
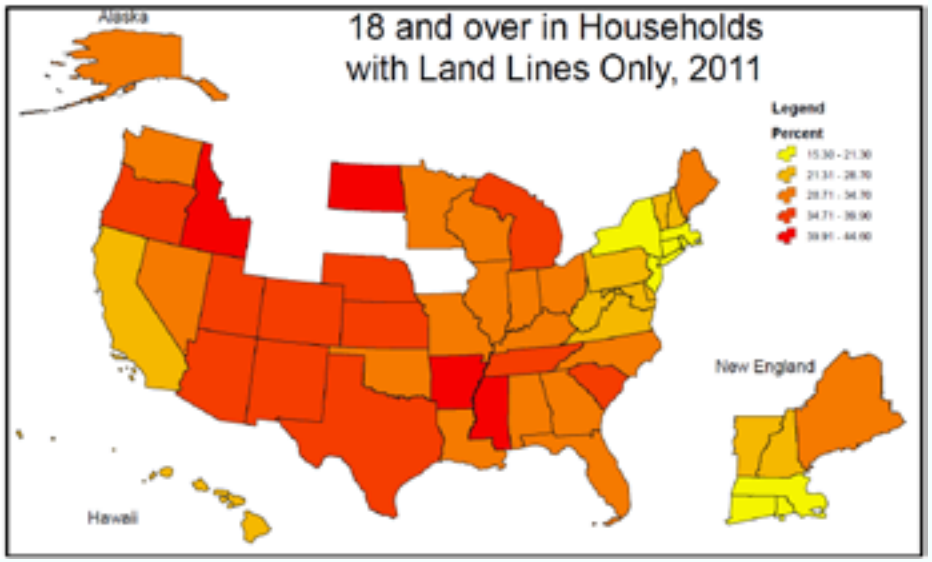
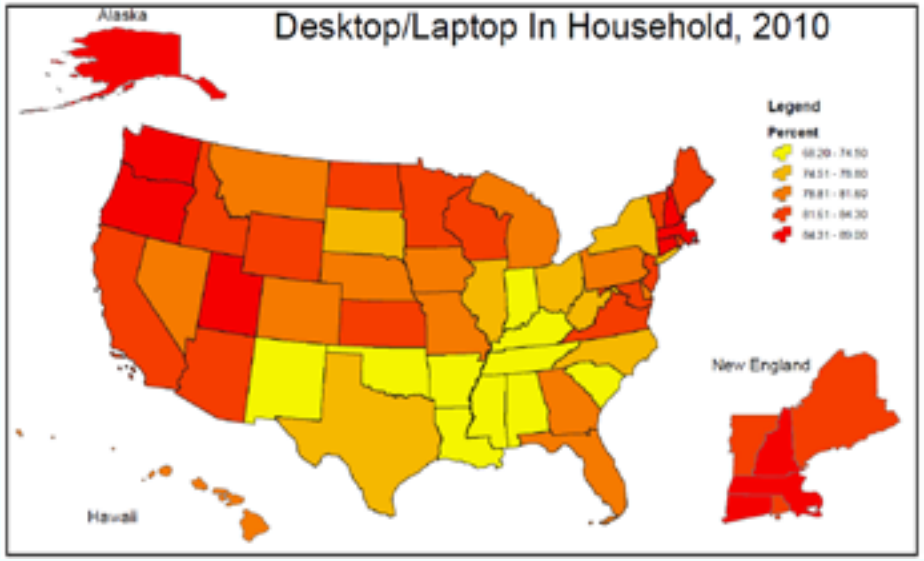


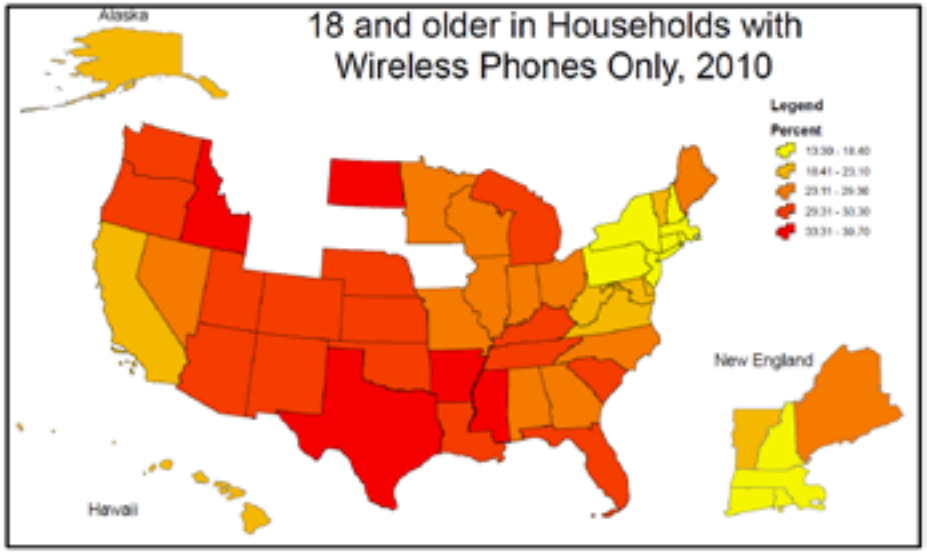
Figure 4.1 Field arrangements of blue and white cells exhibiting: (A) extreme negative spatial autocorrelation; (B) a dispersed arrangement; (C) spatial independence; (D) spatial clustering; and (E) extreme positive spatial autocorrelation. The values of the I statistic are calculated using the equation in Section 4.6. (Source: Goodchild 1986 CATMOG, GeoBooks, Norwich)

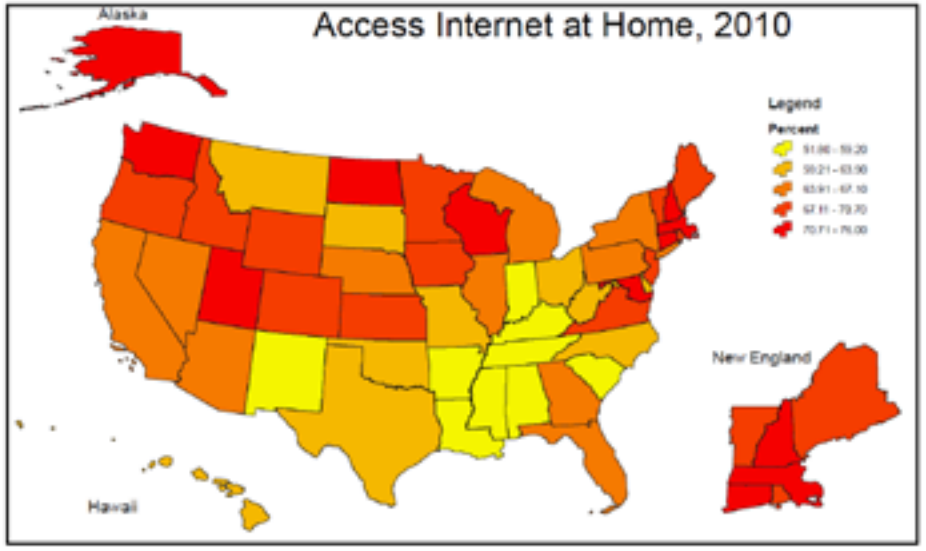
Source: Longley, P. et al. (2011). *Geographic Information Systems & Science*, Wiley, p. 103.

Maps of Dependent Variables I

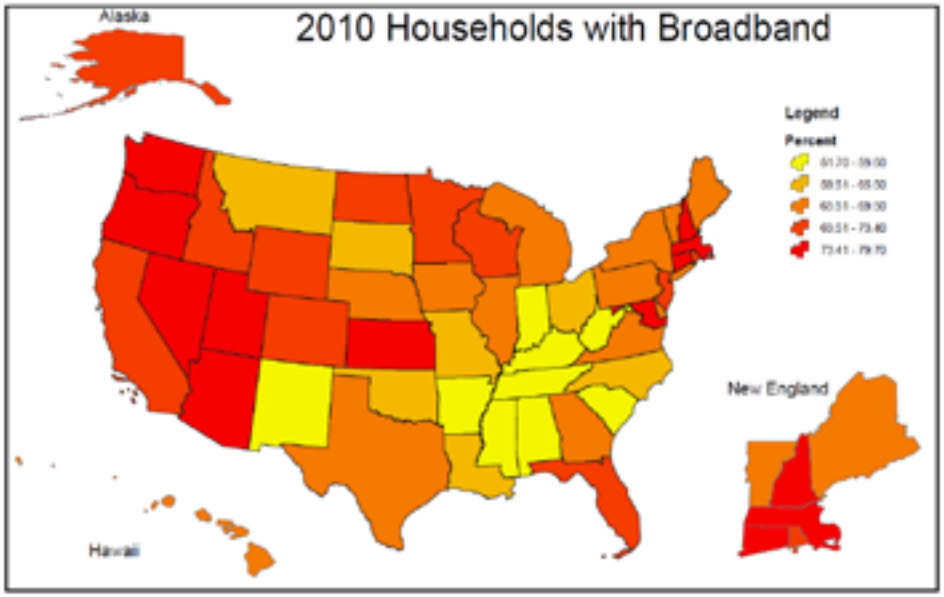


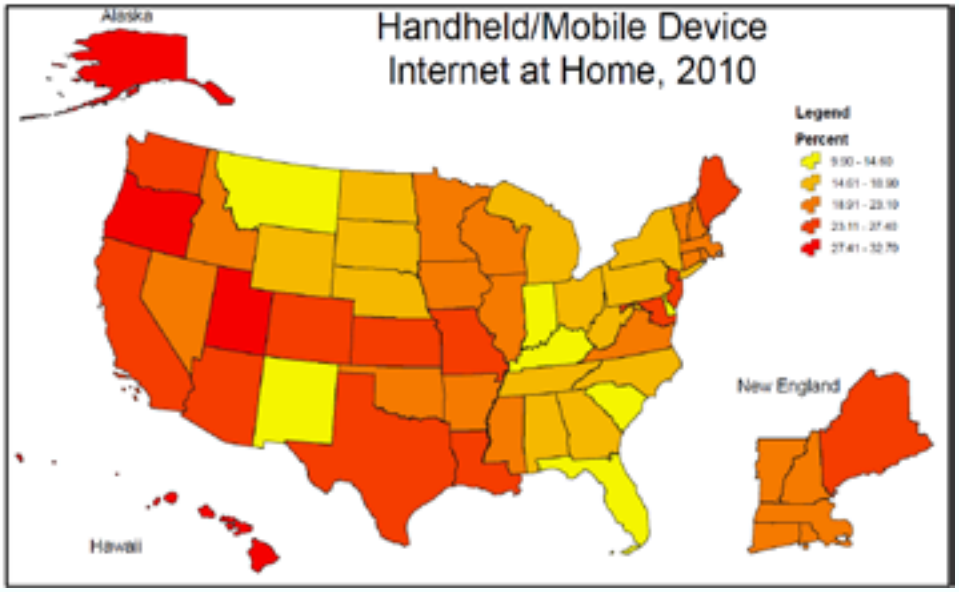


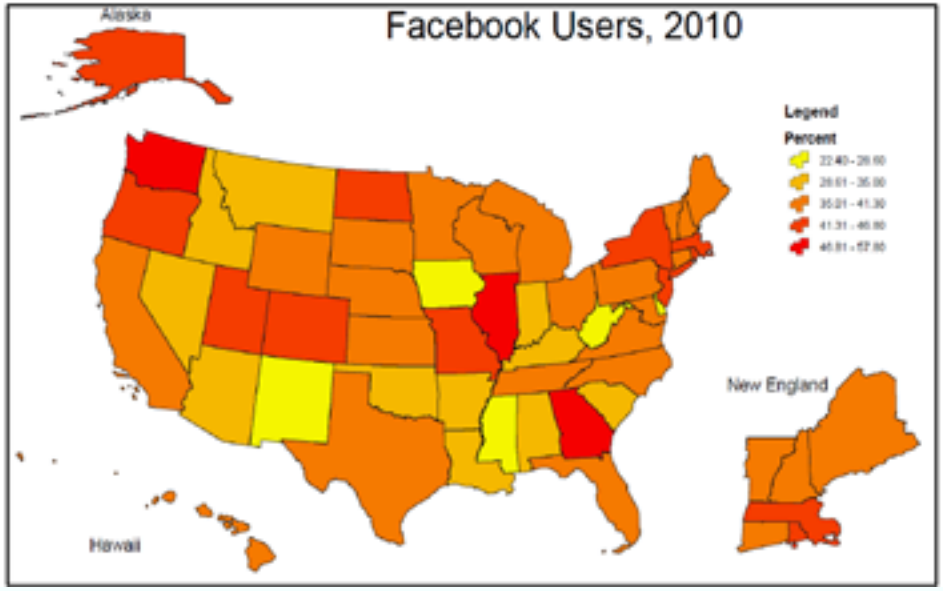


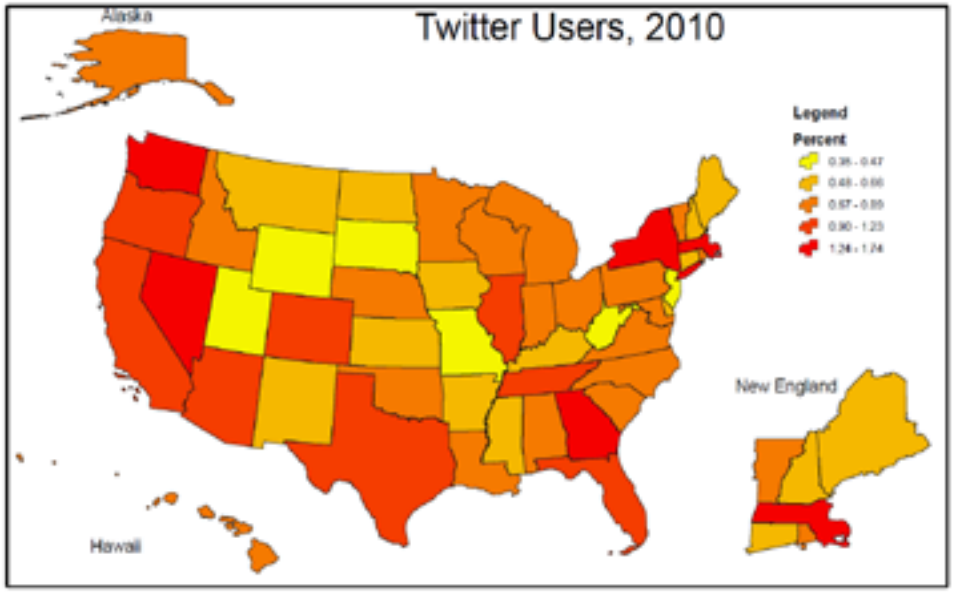


Maps of Dependent Variables II









Spatial Autocorrelation Findings

- All dependent variables except Facebook & Twitter have highly significant spatial autocorrelation.
  - Possibly due to appeal of social media to a more youthful base of users/consumers that is not limited or bound in its ICT use to ICT-intensive geographies.

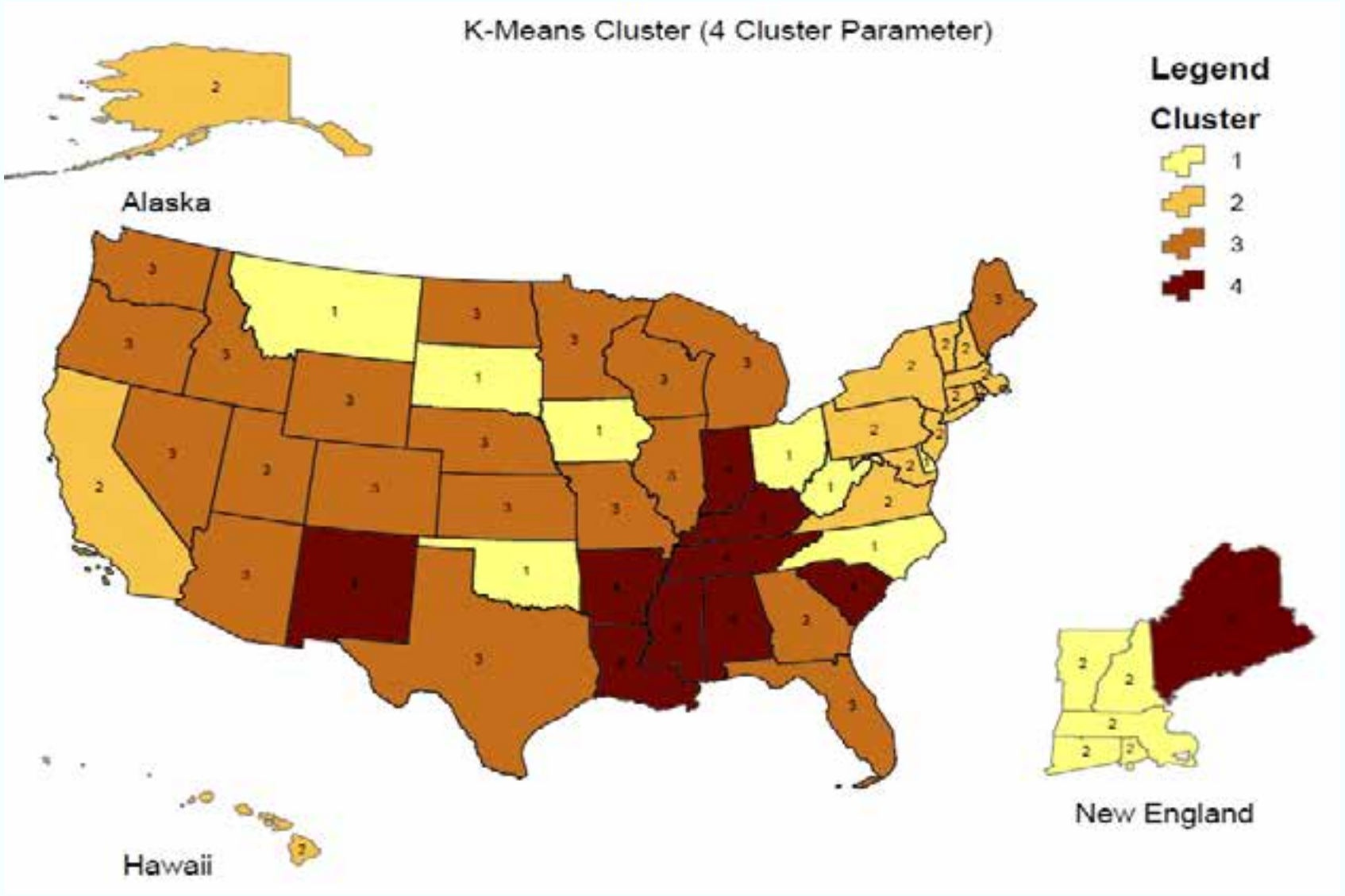
Moran's I values for United States, N=50, 2010								
Desktop/ Laptop in Household	Log of Internet Access at Home	Broadband Adoption in Household	Persons in Cellphone-Only Household	Mobile Wireless High-Speed Devices	Persons in Fixed-phone-only Household	Facebook Users	Twitter Users	AVERAGE
0.547***	0.471***	0.457***	0.621***	0.230***	0.648***	-0.004	0.069	0.381
Moran's I values for China, N=31, 2009								
PCs per 100 Urban Families	PCs per 100 Rural Families	Internet Users per 100 pop.	Broadband Subscribers per 100 pop.	Mobile Telephone Subscribers per 100 pop.		Urban Fixed Phone Subscribers per Capita	Number of Domain Names per 100 pop.	Number of Web Pages per Capita	AVERAGE
0.346***	0.206*	0.264**	0.272**	0.205*		0.252**	0.615***	0.086*	0.285
Moran's I values for China, N=29, Excluding Beijing and Shanghai, 2009									
0.343***	0.239**	0.137	0.258**	0.061		0.143	0.236**	0.177*	0.199

(Source for China, author, 2013a)

Spatial Autocorrelation of Dependent Variables, US 2010 and China 2009, as measured by Moran's I Statistic

K-means Clusters for the US 2010

(1 = Intermediate ICT cluster; 2 = High ICT cluster;
3 = Highest ICT Cluster; 4 = Lowest ICT Cluster)



Cluster Characteristics

Dependent Variable	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Ratio of High to Clusters
Desktop/Laptop in Household	78.30	83.30	82.20	71.90	1.16
Internet Access at Home	63.30	69.00	68.10	56.70	1.22
Broadband Adoption in Household	64.16	72.12	71.20	57.06	1.26
Persons in Cellphone-Only Household	26.18	18.76	30.90	31.89	1.70
Mobile Wireless High-Speed Devices	17.20	23.60	22.50	17.00	1.39
Persons in Fixed-phone-only Household	31.22	23.43	36.15	36.54	1.56
Facebook Users	32.40	40.80	41.90	31.10	1.35
Twitter Users	0.64	0.82	0.89	0.72	1.39
	Delaware	Alaska	Arizona	Alabama	
	Maryland	California	Colorado	Arkansas	
	North Carolina	Connecticut	Florida	Indiana	
	Ohio	Hawaii	Georgia	Kentucky	
	Oklahoma	Maryland	Idaho	Louisiana	
	South Dakota	Massachusetts	Illinois	Mississippi	
	West Virginia	New Hampshire	Iowa	New Mexico	
	Wyoming	New Jersey	Kansas	South Carolina	
		New York	Maine	Tennessee	
		Pennsylvania	Michigan		
		Rhode Island	Minnesota		
		Vermont	Missouri		
		Virginia	Montana		
			Nebraska		
			Nevada		
			North Dakota		
			Oregon		
			Texas		
			Utah		
			Washington		
			Wisconsin		

1 = Intermediate ICT cluster
2 = High ICT cluster
3 = Highest ICT Cluster
4 = Lowest ICT Cluster

K-means Cluster Analysis Findings

- *Cluster 1:* "Selected non-Metropolitan." Technology access/use levels are intermediate. Most of the states are more rural, and are low to medium in their proportion metropolitan.
- *Cluster 2*: "Northeast, California, Hawaii, Alaska." Technology levels are high and resemble Cluster 3. Most states are in the Boston-Washington megalopolis and California, which are regarded politically as "blue states."
- *Cluster 3*: "Western, Sunbelt Cluster."
  - This highest cluster overall in ICT access/use is similar to Cluster 2, but higher in cell-phone-only, fixed-phone-only households, and Twitter use.
  - It comprises twenty states mostly in the Rocky Mountain region, and some in the upper and western Midwest, as well as Georgia, Maine, and the Sunbelt states of Arizona, Texas, and Florida, while only three in are in the East.
  - The states tend to be large in land area.
- *Cluster 4*: "South, non-Sunbelt states, Indiana, New Mexico." The cluster has the lowest technology access/use levels, with broadband adoption in 57% households, and computers in 72% households. These states have lowered educational and income levels, and few large metropolitan areas.
- **Conclusions**
  - Higher technological clusters (2 and 3) tend to be in the Northeast, prosperous Sunbelt states, Pacific Northwest, and Rocky Mountain areas.
  - Lower technology clusters (1 and 4) are in more rural, mostly interior parts of the country.
- Overall, the four clusters identified are mostly agglomerated as geographical regions. This is consistent Tobler's Law (Longley et al., 2011).

Standardized Regression Results: Full Sample

with Putnam's Social Capital Index

		Desktop/Laptop in Household	Log of Internet Access at home	Broadband Adoption in Household	Persons in Cell-Phone only Household	Mobile Wireless High-Speed Devices	Persons in Fixed Phones-only Household	Facebook Users	Twitter Users
Demographic	Urban	0.415**	.455***	0.524***					
Ethnicity	Asian					0.488***			
Ethnicity	Hispanic	-0.254	-0.329**				.245*	-435*	
Economic	Employed civilian workforce								
Economic	Service occupation employment					-0.261			
Economic	Construction industry employment								
Education	College graduates	0.295*	0.311**	0.279**	-0.354**		-.202	.300	-.265*
Education	Higher ed government funding				0.247*		.189		-.207
Innovation	R&D expenditures						-.301*		
Societal Openness	Publisher Annual Sales (Newspaper, Periodical, Book, Directory)								
Societal Openness	Overall Freedom Index				0.440***		0.485***		.224
Societal Openness	Election performance Index								
Social Capital	Putnam's Social Capital Index	.493***	0.486***	0.446***				.321*	
Social Capital	Immigrant population							.487*	0.764***
	Regression adjusted R squared and significance level	0.515***	0.562***	0.608***	0.422***	0.163**	0.494***	0.294***	0.372***
	Sample size	50	50	50	50	50	50	50	50
	OLS Regression Tests								
	Joint Wald Statistic	74.32585***	88.058247***	139.435129***	66.680747***	15.443465***	89.609108***	39.005191***	10.834381*
	Koenker (BP) Statistic	7.636497	6.287538	2.711848	2.439081	0.995471	4.163691	2.265298	5.498212
	Jarque-Bera Statistic	1.643593	1.620257	1.041716	0.596326	1.03644	0.077525	5.125463	3.400511
	SPATIAL AUTOCORRELATION OF RESIDUALS								
	Moran's Index	0.409564***	0.287848***	0.236272**	0.042322	0.289023***	-0.016117	-0.009846	0.06977
	* signif. at 0.05		** signif. at 0.01			*** signif at 0.001			

Regression Findings: Full Sample

1. Social Capital (Putnam's SCI) significantly associated with *computer, internet, broadband in household*.
2. Societal openness factor (freedom index) significant influence on *mobile only & landline only households*.
3. Ethnicity is a strong predictor of *mobile high-speed wireless devices (*Asian) *& internet access at home* (negative association with Hispanic ethnicity).
4. Urban is significant for three of eight ICT factors: *computer, internet, broadband in household*.
5. Other significant factors
   - Education (for cellphone-only households),
   - R&D (inversely associated with fixed-phone-only households).

- Diagnostics
  - Findings meet the OLS diagnostic tests entirely for all regressions.

Standardized Regression Results: Reduced Sample (n=42)

8 states with urban population >=90% excluded

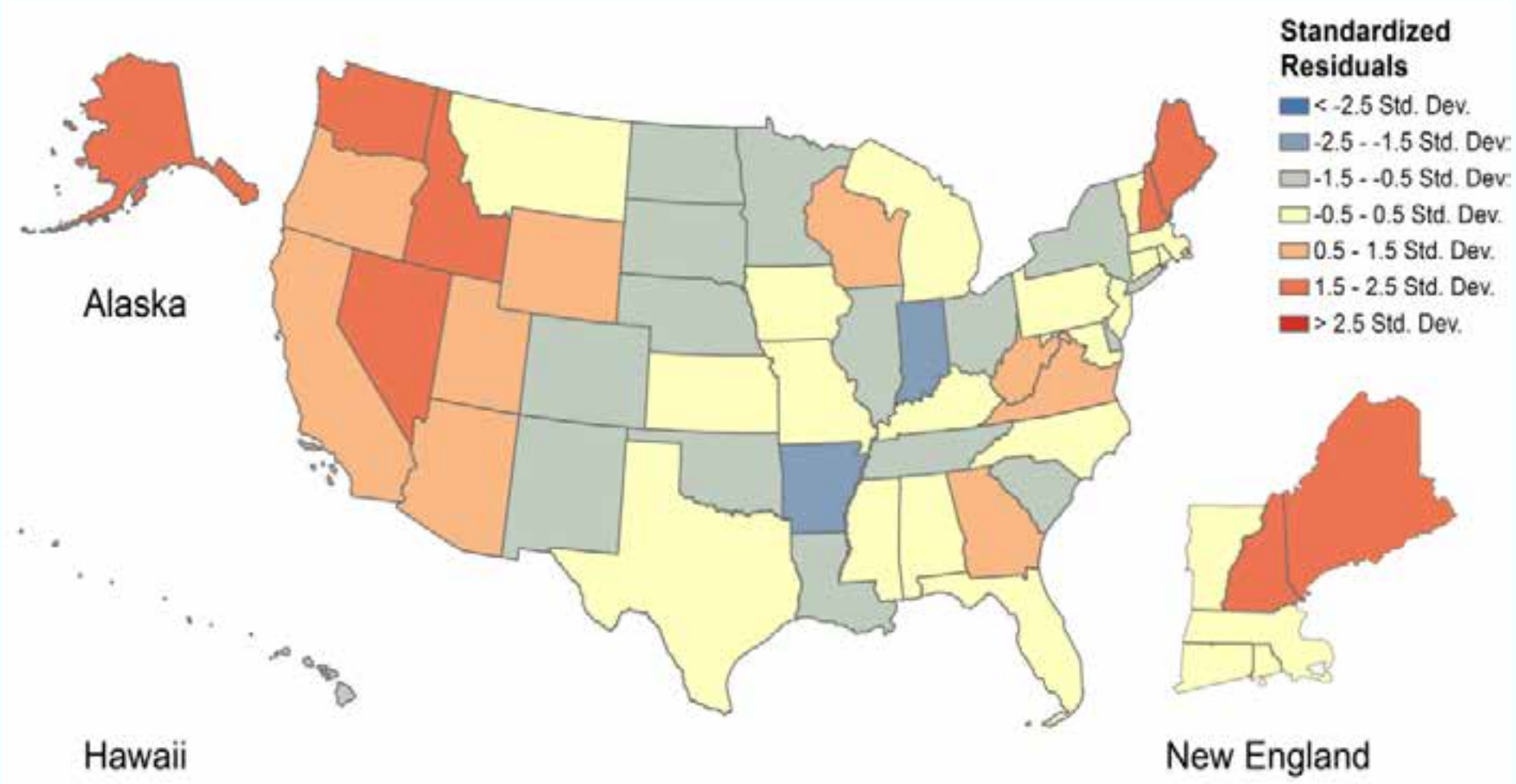
		Desktop/Laptop in Household	Log of Internet Access at home	Broadband Adoption in Household	Persons in Cell-Phone only Household	Mobile Wireless High-Speed Devices	Persons in Fixed Phones-only Household	Facebook Users	Twitter Users
Demographic	Urban		.257						
Ethnicity	Asian	.417***	0.288*	.235		.513***		0.635***	
Ethnicity	Hispanic	-0.268*	-0.416***		0.353**		0.377**		
Economic	Employed civilian workforce								
Economic	Service occupation employment								
Economic	Construction industry employment								
Education	College graduates	0.314**	0.303**	0.296**	-0.464***		-0.405**		-0.373*
Education	Higher ed government funding								
Innovation	R&D expenditures				-0.340*		-0.394**	-0.306	
Societal Openness	Publisher Annual Sales (Newspaper, Periodical, Book, Directory)								
Societal Openness	Overall Freedom Index				.225		0.248*		
Openness	Election performance Index							.341*	
Social Capital	Putnam's Social Capital Index	0.500***	0.513***	0.491***					
Social Capital	Immigrant population								0.606***
	Regression adjusted R squared and significance level	0.614***	0.644***	0.647***	0.509***	0.244***	0.512***	0.315***	.254***
	Sample size	42	`	42	42	42	42	42	42
	OLS Regression Tests								
	Joint Wald Statistic	81.617779***	117.745156***	66.744692***	58.003309***	14.652591***	47.184128***	25.286995***	14.827865***
	Koenker (BP) Statistic	8.485696	9.099237	4.636538	2.889868	0.001553	2.152501	6.050845	4.423228
	Jarque-Bera Statistic	1.084778	1.378454	0.554106	1.008805	0.62894	1.077467	1.17947	1.062528
	SPATIAL AUTOCORRELATION OF RESIDUALS								
	Moran's Index	0.181663	0.259442	-0.083256	-0.033149	0.147392	-0.011108	0.021444	-0.00219
	* signif. at 0.05		** signif. at 0.01			*** signif at 0.001			

Regression Findings: Reduced Sample (n=42)

8 states with urban population >=90% excluded

- Social Capital (Putnam's SCI) continues to be significantly associated with *computer, internet, broadband in household*.
- But, societal openness loses significance relative to full sample.
- A striking difference is that ethnicity exceeds urban in the demographic category.
  - Directionally, ethnicity is positive for Asian and mixed for Hispanic.
- Urban is not significant for any of the ICT dependent variables.
- Findings for college education are consistent with full sample.

OLS Regression Residuals for Desktop & Laptop in Household, US 2010



Discussion: Clusters of ICT

- Cluster size
  - US: Somewhat even (approx. 8 – 20 states in each cluster)
  - China & India: Unique 1-state clusters (Beijing & Shanghai in China, Delhi in India)
- Ratio of high-to-low ICT utilization
  - US: 1.2 – 1.7
  - China & India: 3 – 20
- Unique metropolitan high-tech provinces and states in China and India skew the high-low ratios more than for the US states.
- Greater ICT use evenness among US states might be due to fewer economic barriers in provision of technological capacities.
  - In China the distribution of ICT capacities is centrally planned & controlled, yielding very high emphasis on its two most important metropolises.
  - In India, Delhi as the national capital has received higher per capita technology investment and infrastructure, and it has highly educated consumers.

Discussion: Social Capital

- Computer, internet, broadband in household: Positive association with social capital but negatively associated with Black ethnicity.
- Ethnic disparity result largely consistent with findings of NTIA (2011).
- Strong influence of peer effects on internet use documented in Agarwal et al (2009).
- Chen (2013): "*information want-nots" tend to be social capital have-nots……closing the digital divides calls for a three-pronged approach that enriches individual, community, and network resources."*
- **Our study offers first validation of the association of social capital with 3 important forms of ICT at the state level for the US.**

Discussion: Education

- Major predictors associated with higher ICT levels are **education** and social capital.
- ICT & Digital Divide literature has consistently supported the relation of education to levels of ICTs for national, state/provincial, and individual units of analysis.
  - Inverse relationship of education to fixed-phone-only households in “less urban” states reflects that households in such states trail in educational levels.
  - Inverse relationship of education with cellphone-only households (N=42) is surprising.
    - We speculate it reflects that those households are more prevalent per capita in more rural and remote states in the U.S., where fixed phone access is less available, and fixed phone technology is aging, and less reliable.

Discussion: Social Media Technologies

- Facebook: Positive ethnic association with Asian (N=42) but negative association with Hispanic (N=50).
- Twitter: Highly significant association with Immigrant Population (N=42).
- Relationship of Asian ethnicity to mobile wireless high-speed devices – supported by a Pew Foundation finding that mobile wireless connectivity of Asian Americans is 77% (20% higher than the total U.S. population)
- Findings also corrobrate studies of the U.S. by noted urban studies theorist Richard Florida,
  - Creative and technological activities are more common in ethnically diverse metropolitan areas (Florida, 2012).

Conclusions

- Conceptual model of predictors of ICT utilization in US states largely robust.
  - Construction labor force could be dropped and replaced by labor force.
    - This reflects that the 21st century U.S. economy is much more a service and knowledge economy.
  - Likewise, newspaper, magazine and book publishers (used as proxies for innovation in prior studies) can be dropped and replaced by an indicator of electronic content.
- 6 ICT variables (desktops/laptops, internet, broadband, cell phones, fixed phones, and mobile wireless devices) are highly agglomerated by state into regional clusters.
- Social media variables of Facebook and Twitter are randomly distributed.
- Exploratory K-means cluster analysis shows 4 distinctive agglomerations
  - Most US states are in two multi-state, high-technology clusters.
  - Some states in mid-America cluster are at intermediate technological level, and
  - A Southern cluster plus New Mexico are at a relatively low level.
- Spatial influences are completely accounted for by the model.
- The model's socioeconomic categories especially associated with the dependent variables are social capital and education.
  - Both are backed by conceptual ideas and empirical findings in the literature, **particularly for developed nations**.

Implications

- For States:
  1. invest and support higher education,
  2. strive to emphasize inclusiveness in society,
  3. emphasize R&D for newer forms of technology, and
  4. favor inclusion in technology initiatives of ethnically diverse segments of the population.
- Examples
  - Georgia Enterprise Technology Services (GETS).
  - Virginia Information Technologies Agency (VITA).
- For Systems Sciences researchers and professionals:
  - Spatial analysis is a useful tool for state governments & the federal government to understand
    - how states compare with each other,
    - which regional agglomerations are present,
    - what their distinctive characteristics are, and
    - how technology is advancing throughout the U.S.

Limitations

- Lack of a longitudinal sample, so trends over time cannot be analyzed. However, the present set of variables is not available in its entirety for a prior time points of 3-5 years ago or earlier.
- Regression methods are unidirectional, rather than considering a system of simultaneous interacting relationships. Nevertheless, a majority of digital divide studies have been unidirectional.
- Sample size. Although quite suited for regression, the prefectural sample is not robust enough for broader models, such as structural equation modeling.

DISCUSSION AND QUESTIONS

Contact: James B. Pick
james_pick@redlands.edu