Internet Use and Online Activities in U.S. States: Geographic Disparities and Socio-economic Influences

James Pick
University of Redlands
james_pick@redlands.edu

Avijit Sarkar
University of Redlands
avijit_sarkar@redlands.edu

Elizabeth Parrish
University of Redlands
elizabeth_parrish@redlands.edu

Abstract

Attention in the digital divide research agenda is shifting gradually from material access of information and communication technologies (ICTs) to actual use of ICTs. As internet usage increases worldwide and online activities concomitantly expand in range and sophistication, it becomes essential to examine patterns and disparities of such usage. This paper examines geographic patterns and disparities along with influences of demographic, economic, and social factors on internet use in U.S. states. Our conceptual model of internet use posits associations of 21 traditional (socio-economic) as well as non-traditional independent variables (social capital, societal openness, innovation, infrastructure, and affordability) with dependent indicators of e-communication, e-commerce, e-education, e-entertainment, e-health, and telework. Age, race/ethnicity, innovation, urban location, managerial and scientific occupations, and social capital are found to predominantly influence internet use spanning a range of online activities. Policy implications of these findings are discussed taking cognizance of geographic disparities in internet use among the fifty states.

1. Introduction

The digital divide is a dynamic and complex phenomenon [29] with social, economic, infrastructural, geographical, and regulatory underpinnings. In the early 2000's, the digital divide was defined as 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 (ICTs) and to their use of the Internet for a wide variety of activities" [17, pp. 5]. A key aspect of this early definition is its emphasis on both access as well as use of ICTs.

Access to ICTs and disparities in access have been the subject of numerous digital divide studies over the years. In the United States, the U.S. Department of Commerce National Telecommunications and Information Administration (NTIA) has been reporting computer and internet use nationwide and in the U.S. states since the late 1990s. The reports are insightful as they expound on demographic, economic, educational attainment, and urban-rural disparities in computer and internet use. Internet usage in some of the early NTIA reports is measured in terms of internet subscriptions or percent of households with internet access, which has been characterized as material access in the digital divide literature [30]. However in more recent NTIA reports [15,16], the purpose of internet use characterized by web-based activities such as emailing, instant messaging, banking or shopping online, and engaging in entertainment activities online has received as much attention as social and economic underpinnings of computer and internet use. This is indicative of a shift in the digital divide discourse from examining aspects of material access to analyzing aspects of purposeful usage of the internet.

This paper analyzes geographic patterns and disparities in purposeful internet use among U.S. states for activities spanning e-communication, e-commerce, e-education, e-entertainment, e-health, and telework. Additionally, this paper examines the influence of traditional demographic, educational, economic factors as well as infrastructural, affordability, social capital, and societal openness related influences on internet use. Our research questions are: (1) Are geographical patterns of internet usage present for the U.S. states as measured by spatial auto-correlation? (2) Are geographic agglomerations of internet usage for a variety of online activities present in U.S. states as estimated by cluster analysis? (3) What are the associations of socio-economic, demographic, affordability, innovation, social capital, and societal openness factors with internet usage in the U.S. states? (4) What are the policy implications of such associations? The remainder of this paper is organized into sections on literature review of internet use, the study's conceptual model of internet use, methodology and data, findings on geographic disparities and socio

HⱢCSS

economic influences on internet usage, implications of findings, and conclusions.

2. Literature Review

Academic interest in the digital divide in the United States has been extensive. Over the years, a number of studies [1,2,4,21,27] have examined various aspects of ICT adoption, diffusion, and access. A number of these studies have focused on examining socio-economic influences on adoption and access for a particular form of ICT, for example, the internet [1,4,28], broadband [8,22], while a few [2,20] have focused on computer adoption in the household along with an array of ICT access.

Some of this academic research has relied on the U.S. Census Bureau's Consumer Population Survey (CPS) data on computer and internet use [1], NTIA and Federal Communications Commission (FCC) datasets [8], survey data from Pew Research Center's Internet and American Life project [28]. In some instances, researchers have employed surveys to collect data on ICT access [4]. However, most of these previous studies have examined the U.S. digital divide through the lens of technology access, or more precisely material access to ICTs. The purpose of ICT use, particularly the range of online activities enabled by the internet has been scarcely examined for the United States. This may be attributed to the fact that until the early 2010's, data on a consistent set of online activities were irregularly collected as part of the CPS, and sporadically reported.

However, with increasing focus on understanding the range of internet use for a variety of activities, recent NTIA reports [15,16] have examined what do Americans do online. For example, NTIA [15] found that researching health plans and finding medical information were relatively common online activities. Online interaction with healthcare professionals was less common, and telemedicine was nascent. The same report stated that using the internet for conducting financial activities, online shopping, accessing entertainment, and using on-the-go location-based services were also popular. Although recent NTIA reports largely describe the usage landscape among Americans, they continue to be illuminating in their analysis of the many influences of demographics, location (urban/rural), affordability, place of access, and other factors on internet access, not use.

Of late, a few studies have examined users' online activities. Some of these studies are situated in very specific contexts. For example, one study [23] surveyed Estonians on their use of the internet for playing games, communicating with peers in chatrooms, within the broader framework of examining

social and cultural aspects of an information society. Another study [6] examined internet use among Aboriginals in a Canadian province. A third study [9] used a comprehensive survey of British nationals to examine different types of internet use (e-entertainment, e-commerce, e-government, civic participation, etc) based upon differences in users' skills, their age, and internet use experience and pedigree.

Recently, using qualitative interview data of a small sample, frequently engaging in email, social networking, searching for work or school-related information, and jobs were found to be popular among low-income U.S. residents [7]. Adoption, diffusion, and use of e-government portals in U.S. states and counties has also been studied [11]. Also in the U.S., demographic differences between all internet users, those who use broadband internet versus dial-up at home for purposes such as conducting online searches, engaging in e-government, e-education, e-communication, e-entertainment, and social networking has been examined [31].

Overall, as focus in the digital divide research shifts towards purposeful internet use and extent and sophistication of online activities, there is a dearth in understanding demographic, socioeconomic, infrastructural, affordability, social capital, and societal openness-related influences on internet use in America. This important gap is addressed in this study. In addition, it investigates geographic patterns of internet use, shedding light on spatial disparities in the usage divide among U.S states.

3. Conceptual Models of Internet Use in U.S. States

The conceptual model (depicted in Figure 1) includes independent variables that influence each of the 17 dependent variables. The effects of the 21 independent variables on each dependent variable provides a detailed portrait of determinants of the many dimensions of purposeful internet use. This model includes determining whether or not these effects are spatially biased. A model is spatially biased if it is not able to account for geographic influences. In that case the model errors will be significantly agglomerated geographically, distorting the interpretation of the findings. If the error terms are randomly distributed spatially, then the geographical biases are accounted for in the model.

The dependent variables' geographic patterns are examined in an exploratory manner to gauge where they are agglomerated into areas of concentration of high values and low values. This give an overall

picture of purposeful internet uses for the U.S. at the state level. Additionally, each variable is individually assessed for extent of spatial autocorrelation.

This model is termed the Spatially Aware Technology Utilization Model (SATUM) [20,21]. The model's independent factors are divided by categories – demographic, economic, education, infrastructure, affordability, innovation, societal openness, and social capital, all of which have been important in prior research [1-5,19-21]. Each category's inclusion in the conceptual model is based on prior research.

The paper's theory is further informed by the model of the digital divide of J.A.G.M. van Dijk [29,30]. This theory proposes that inequalities of personal position leads to inequalities in individual resources, including material, mental, social and cultural resources [30]. Resource inequalities in turn lead to differential access, which is also influenced by technology characteristics of ICT which influence the behavioral steps in access of motivation, material access, and skills. Access eventually leads to differences in the individual's participation in society, in particular in the economy, social networks, geographic space, culture, politics, and institutions. That social participation in turn loops back and effects individual positional characteristics, so a feedback loop is realized. Although not well known as a mainstream theory of MIS, van Dijk's theory is highly relevant to shed light on mechanisms and relationships in the theory behind this paper's empirical study.

Demographic factors: Location in urban environments has been associated with internet and ICT access and use. In one study, rural location was associated with lower level of fixed broadband usage, as were low income and post-high-school education [22]. In another study of broadband access across the US [8], urban-rural differences were found to be more complex than apparent at first glance, so access groups were categorized by generation as well as by urban-rural context. Nevertheless the most poorly served broadband access was for factory and farm communities, while the best served were prosperous baby boomers located mostly in urban suburbs, enterprising young singles in cities, and high-rise urban dwellers. Hence, urban location is posited to be positively related to internet use.

Many studies have pointed to youth as a factor in internet use in the US [4,10,19,27]. Generally, millennials are known for their advanced access and use of the Internet, while the elderly tend to lag; hence, age is posited to be inversely related to internet use.

Ethnicity has commonly been included in studies of digital divide in the US, but it is less common in international studies, since many nations do not collect widespread and systematic data on their racial/ethnic groups. In the U.S., studies have consistently indicated relative high technology use by Asians [19,22,23] and relatively low technology use by Hispanics [19,22,25]. A Pew study showed that Hispanics are somewhat lower than average in internet use, although the gap widened for four years during and post the recent recession [18].

For African Americans the findings are mixed. Studies largely revealed Blacks to be associated with reduced ICT and internet use [3,4,19,22], although some research on ICT use at the county level has indicated positive associations between the Black population in counties and indicators of entertainment on mobile internet [25]. Thus it is posited that percent Asian will be positively related to internet use, percent Hispanic will be inversely related to internet use, and percent Black will have mixed outcomes (positive and negative) across dependent variables.

In van Dijk's theory, an individual's personal characteristics of age, race/ethnicity, gender, and health lead to opportunities to gain resources, including material, knowledge, social and cultural resources, which favor ICT access. Van Dijk does not indicate the valence of effect. We base the expected valence of personal characteristics by inducing valence from prior literature as explained above.

Economic factors: Economic factors such as income, professional/technical occupations, and service occupations, have often been related to internet and technology uses. In a study of a sample of US counties, employment in professional, scientific, technical services was the leading socio-economic factor associated with revenues and payroll for IS and telecommunications/broadcasting industries, while household income and other services were correlates of telecommunications/broadcasting [2]. In research on a technologically-disadvantaged community in the US, small sample findings indicated that usage was associated with the economic factors of employment status and household income [27]. A national U.S. survey study of 941 respondents of internet access found high family income was among the strongest predictors of internet access, while age and graduate education were even stronger [4]. Creative cities, having occupational thrust towards science, arts, and management, have been identified to have high technology levels as a cornerstone [5]. Hence, median household income, service occupations, and management/science/arts occupations are posited to be related to internet use. In terms of van Dijk's theory [29,30], these factors are initial positional characteristics, which favor development of individual material, mental and social resources. Again we induce the valence of these effects from prior literature.

Education: In a study of e-entertainment use in US counties, educational level was the most influential socio-economic correlate [26], while for mobile entertainment use in US counties college education was one of three dominant correlates for cell phone access to a variety of entertainment apps [25]. College education was a highly significant determinant of home desktop/laptop use, internet access and broadband adoption in a study of US states in 2010 [21]. On the other hand it was inverse for states with high levels of cell-phone-only households and twitter users. A large-scale US survey study of Internet access found graduate education to be second to young age in association with internet access [4]. Van Dijk interprets the mechanism of education's influence as a positional category leading to development of mental, social, and cultural resources [30]. Regarding valence, the digital divide literature has extensive agreement that education is positive in effect. The mechanism of education's influence we posit relates to users' training and experience and to communities/states' greater demand for ICT's productivity and knowledge benefits, leading us to posit that that government funding for higher education is associated with internet use. Educational attainment (specifically, college education) was excluded as a correlate due to high positive correlation with median household income.

Infrastructure: Broadband speed has not had much inclusion in digital divide studies, yet a recent article underscores that broadband speed, although difficult to measure systematically, is a major contributor to broadband levels worldwide. Furthermore, a study that emphasized the extent of population covered by the reach of wireless and fiber networks is fundamental to internet use [8]. Thus we posit that broadband speed, population covered by fiber, and population covered by wireless are associated with internet use. In van Dijk's model, infrastructure provides positive support to behavioral steps in access, in particular it is not enough to have motivation for access without material access i.e. infrastructure to realize the motivation [30].

Affordability: The studies of technology access and use in the US which include cost have found it often to be associated with ICT use. For instance, in research on non-use of the internet at home, cost was confirmed to be the main reason for not using the internet [13]. A Pew research study [10] also found that cost is a primary reason why non-adoption of broadband. Hence, we posit that per capita population who deem high cost to be main factor for internet non-use is related to reduced internet use. Affordability is not included in van Dijk's model. We posit the mechanism for affordability is that a user is more likely to adopt ICT if he/she can pay for its use, within the household budget available to the user. The lower the cost, the more likely the user is to adopt it for purposeful use.

Innovation: It can be argued that states that are innovative are expected to have higher levels of internet use, since such states would have a higher proportion of innovative workers who would tend to make more use of the internet, and, further, that the software and content produced by innovators in the state might have more use at least initially nearer to the center of the innovation. Innovation has been challenging to study within nations, since innovation and R&D data are fairly rare for intra-country units. A study of the provinces of Japan, which provided data on patents as a proxy to R&D, found that innovation was a significant correlate for number of subscribers to broadband, mobile phones, Facebook, and Twitter [14]. However, in a study of US states, with R&D expenditures provided by the National Science Foundation, there was little effect, and the limited findings mainly pointed towards high R&D states having reduced proportion of persons in fixed-phone only households, which alludes to positive association of R&D with ICT levels. In the van Dijk theory, innovation is not explicitly stated, but, over time, it would be a positive influence on the base characteristics of ICT of hardware, software, and content. Based on the findings in [14] and on the van Dijk mechanism as explained, we posit innovation to be associated with internet use.

Societal Openness: Societal openness was an important factor in the empirical findings on a worldwide model of ICT use, which was empirically tested by structural equation modeling [21]. Further in a study of African nations, societal openness, as measured by the extent of ICT-related laws, was shown to be a major determinant of ICT access and use [20]. For these reasons, societal openness is posited to be related to internet use. In terms of van Dijk's model, the mechanism of societal openness is that in an open society, cultural and institutional barriers are lowered to participation i.e. purposeful use of ICT [30].

Social Capital: Social capital is defined as social linkages that citizens in a community develop with each other based on their resource and inter-personal communication needs. The concept, originally developed in sociological research [24], has been occasionally included in US digital divide studies and proven to be a significant factor associated with internet access [1,4,21]. Therefore, social capital is posited to be associated with enhanced internet use. Van Dijk's theory includes social networks, in the older sense of social capital, as a way to enhance participation i.e. purposeful ICT use in society.

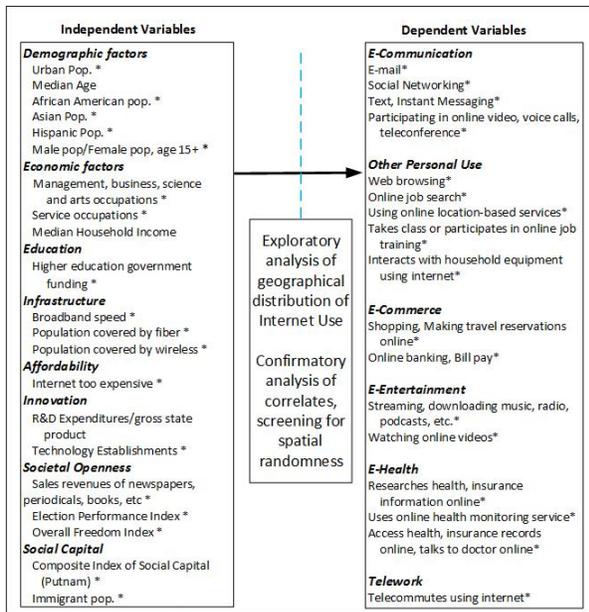The conceptual model is depicted in Figure 1.

Figure 1. Conceptual model of internet use, U.S. states

4. Methodology and Data

The research methodology is comprised of the following steps: (1) Descriptive statistics (Table 1) for all dependent indicators of internet use and their independent correlates are computed, and Pearson correlations are estimated for the independent variables for the diagnosis of multicollinearity. (2) Dependent variables were then mapped. Descriptive mapping provides important visual cues about spatial patterns and geographic disparities in internet use for various activities. (3) Subsequently, we diagnose statistically significant agglomeration (or randomness) of dependent variables using the Moran's I test statistic. The null hypothesis is values of a dependent variable are randomly distributed spatially (Moran's I close to 0). Moran's I values closer to +1 or -1 are indicative of spatial autocorrelation, in other words, spatial bias is present in the dependent variable's geographic distribution. It is essential for our OLS regression-based associations to account for spatial bias to ensure validity of regression results. (4) To obtain a more holistic understanding of spatial patterns of internet use in U.S. states, we subsequently conduct K-means cluster analysis (with K=6). K-means cluster analysis is exploratory, not confirmatory, yet it provides a high-level overview of agglomerations of states that lead the U.S. in internet use for a variety of online activities, versus those states that are laggards. K-means clusters are mapped and characterized in terms of their demographic, socio-economic, innovation, and other

attributes. This provides important clues about disparities between leaders and laggards. (5) Finally, OLS regression analysis is employed, in stepwise fashion, allowing in only those independent variables with significance levels equal or less than 0.05, to test posited associations of independent correlates with dependent variables. A cutoff value of 5.0 for the Variance Inflation Factor (VIF) is employed to screen multicollinearity problems. Regression assumptions are tested using three diagnostic tests (Joint Wald, Koenker, and Jarque-Bera) and regression residuals are mapped and tested to diagnose if spatial bias continues to influence regression results.

Table 1. Variable definitions, descriptive statistics

		Definition	Source	MIN*	MAX*	MEAN*	STD. DEV*	VAR
	Dependent Vars.							
E-Commerce	ECOMM_15	Shops, Makes Travel Reservations, or Uses Other Consumer Services Online		0.56	0.80	0.69	0.05	0.003
	FIN_15	Uses Online Financial Services Like Banking, Investing, Paying Bills		0.53	0.77	0.64	0.05	0.003
E-Communication	EMAIL_15	Uses Email		0.84	0.96	0.91	0.02	0.001
	SOCNET_15	Uses Online Social Networks		0.66	0.79	0.71	0.03	0.001
	TXTIM_15	Uses Text Messaging or Instant Messaging		0.78	0.94	0.86	0.03	0.001
	VOIP_15	Participates in Online Video or Voice Calls or Conferencing		0.26	0.60	0.36	0.06	0.004
E-Entertainment	AUDIO_15	Streams or Downloads Music, Radio, Podcasts, etc.		0.42	0.77	0.54	0.05	0.003
	VIDEO_15	Watches Videos Online		0.53	0.84	0.67	0.05	0.003
E-Health	HLTHINFO_15	Researches Health Information Online	NTIA	0.36	0.57	0.49	0.05	0.003
	HLTHMTR_15	Uses Health Monitoring Service that Connects to the Internet		0.02	0.10	0.06	0.02	0.000
	HLTHREC_15	Access Electronic Health Records or Insurance Records, or Talks w Doctor Online		0.13	0.39	0.25	0.06	0.004
E-Other	EDU_15	Takes Class or Participates in Job Training Online		0.15	0.30	0.21	0.03	0.001
	JOBSRCH_15	Searches for a Job Online		0.19	0.44	0.27	0.04	0.002
	LOC_15	Uses Online Location-Based (On-the-Go) Services		0.62	0.83	0.70	0.04	0.002
	WEB_15	Browses the Web		0.84	0.94	0.91	0.02	0.001
	HHIOT_15	Interacts with Household Equipment Using the Internet		0.04	0.11	0.06	0.02	0.000
Telework	TELEWK_15	Telecommutes Using the Internet		0.14	0.38	0.21	0.04	0.002
	Independent Vars.							
Affordability	NOLCOST_13	Main Reason for Household Not Online at Home: Too Expensive	NTIA	0.18	0.40	0.27	0.01	0.045
Innovation	RDEXP_13	R&D expenditures / gross state product 2013	NSF	0.00001091	0.00040939	0.00001905	0.00001204	0.000
	TECHBUS_13	Tech establishments as a percentage of total establishments	ACS2013	2.00	7.50	4.00	0.19	1.335
Education	EDFUND_13	Higher education government funding per capita	SHEEO	150.72	672.71	246.77	15.61	111.506
Infrastructure	BBHIGH_14	Broadband speeds ≥ 6 mbps, and Upload speeds ≥ 1.5 mbps	NTIA NBM	8.00	116.00	42.73	3.48	24.850
	FIBRCOVR_14	Percent Population covered by Fiber availability	NTIA NBM	1.70	97.90	25.63	3.17	22.613
	WRLSCOVR_14	Percent Population covered by Wireless availability	NTIA NBM	93.90	100.00	99.29	0.16	1.145
Societal Openness	NEWSPPR_12	Sales revenues of newspapers, periodicals, books, etc.($)/Total Sales Revenues($)	ECONCEN2012	0.00	0.06	0.01	0.01	0.000
	ELCTN_14	Election performance index 2014	PEW	0.49	0.84	0.69	0.01	0.083
	FRDM_14	Overall Freedom Ranking 2014	CATO	0.0039	1.31	0.97	0.03	0.247
Demographic	AFRAM_13	African American, percent of total population	ACS2013	0.0040	0.49	0.11	0.02	0.109
	ASIAN_13	Asian, percent of total population	ACS2013	0.01	0.38	0.04	0.01	0.055
	HSPN_13	Hispanic/Latino, percent of total population	ACS2013	0.01	0.46	0.10	0.01	0.098
	MEDAGE_13	Median Age	ACS2013	29.60	43.20	37.64	0.33	2.352
	POPM2F_13	Male Pop 15+/ Female Pop 15+ *100	ACS2015	88.04	111.17	95.99	3.90	15.240
	URBAN_10	Urban population as percent of total population	CENSUS2010	0.39	1.00	0.74	0.02	0.149
Economic	EMPBUS_13	Total employment in management, business, science and arts occupations, as percent of population 16+ in labor force	ACS2013	27.80	60.20	36.08	0.67	4.779
	EMPSVC_13	Total employment in service occupations, as percent of population 16+ in labor force	ACS2013	15.70	27.70	18.00	0.27	1.909
	MEDHHINC_10	Median Household Income ($)	ACS2013	39,031.00	73,538.00	53,530.27	1,229.51	8,780.43
Social Capital	IMMIGRT_13	Percentage Population Immigrant (not US citizen at time of birth)	ACS2013	1.40	27.00	8.86	0.85	6.039
	SOCCAP_15	Social Capital, 2015	Esri BA	0.25	0.32	0.28	0.0023	0.016

* Sample Size, n = 51 (50 states & District of Columbia)

NTIA	National Telecommunications & Information Administration, https://www.ntia.doc.gov/page/download-digital-nation-datasets		
NTIA NBM	National Telecommunications & Information Administration, National Broadband Map, https://www.broadbandmap.gov/		
NSF	National Science Foundation	ACS, 2013/2015	American Community Survey Datasets, 2013/2015
CENSUS2010	U.S. Census Bureau, 2010 Data	ECONCEN2012	U.S. Census Bureau, Economic Census, 2012
PEW	PEW Research Center	CATO	CATO Institute
SHEEO	State Higher Education Executive Officers Association	Esri BA	Esri Business Analyst Datasets, 2015

Data for all dependent indicators of internet use is obtained from the NTIA's Digital Nation Data Explorer. This data is collected as part of the Census Bureau's Current Population Survey (CPS) and is published with major findings as part of the NTIA's well-known "Digital Nation" report series. Data on demographic and economic independent variables are also sourced from the Census Bureau's American Community Survey. NTIA datasets were used to procure data for affordability and infrastructural independent variables. Societal openness appears in the conceptual model as three independent variables, one of which, the Election Performance Index is from the Pew Research Center, and another, the Freedom Index is collected from the "Freedom in the Fifty States" project, originally at George Mason University, and now part of a public policy think tank. Finally, a composite construct of social capital was developed as

an index combining participation in public activity, serving in a local committee, volunteering in a charitable organization, and participation in voting [24]. Information for these four elements of social capital were sourced from datasets of Esri's Business Analyst software. All variables were normalized by population or converted to a ratio. Variable definitions, data sources, year of data collection, and descriptive statistics are in Table 1.

5. Spatial Patterns of Internet Use

K-means (K=6) cluster analysis and 17 dependent indicators of internet use yields six agglomerations of states as shown in Figure 2.

Cluster 6 has only one "outlier" member, Washington D.C. – the state with the highest overall use of the internet among the 50 states with cluster center values that are 1.06 to 2.31 times higher than those of the lowest cluster 1, comprised of 9 states (Alabama, Arkansas, Louisiana, Mississippi, Nebraska, New Mexico, Oklahoma, South Carolina, West Virginia). In terms of internet use, Washington D.C. exceeds all other states in terms of online activities spanning all 17 dependent variables. As the nation's capital and a hotspot of political and media activity, internet usage for e-communication is expected to be high. This is also expected to spur other forms of online activities. Compared to member states in cluster 1, Washington D.C.'s exclusively urban population is younger, wealthier, more gender-balanced, and markedly more employed in management, science, and arts occupations. Cluster 1 states are agglomerated in the South and the Appalachian region with the exceptions of New Mexico and Nebraska. This region has been identified as a laggard in prior studies [21,32]. Cluster 2, with the exception of Kansas, is comprised of states that are sandwiched between lowest use cluster 1 states in the south and moderate-low use states in the Great Lakes region of cluster 3.

Cluster 3, a major agglomeration of 20 states spanning Arizona in the West, prairie states such as North Dakota and Wyoming, Wisconsin, Iowa, and Missouri in the Midwest, rustbelt states of Michigan and Ohio, the north-northeastern states from New York to Maine, and Georgia and Florida in the southeast, have low to moderate levels of online activities. These states are 4 to 6 years older compared to cluster 5 and 6 states, with close to a third or more of the population in rural areas. Median household income is also lower in these 20 states by as much as $15,000 in a few instances compared to clusters 5 and 6 states underlining socio-economic disparities between moderate-low use and high internet use states.
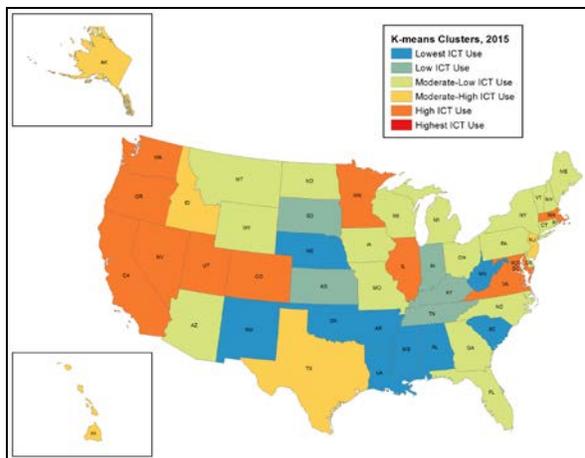

Figure 2. K-means clusters of internet use in U.S. states, 2015

Finally cluster 5, the highest internet use cluster states (excluding Washington D.C.) are situated along the Pacific coast in the West, the Boston-Washington D.C. megalopolitan area in the east, the Rockies (Nevada, Utah, Colorado) and Illinois and Minnesota in the Midwestern Great Lakes region. Ethnically, this cluster is as diverse as states in clusters 2, 3, and 4, with slight differences in median age. Median household income is however significantly higher, so is the extent of urban population. Employment in management, science, and arts occupations is also somewhat higher in these states compared to states in clusters 1 – 4.

Overall, online activities in U.S. states largely resemble patterns of material access to the internet [26]. The similarities reinforce that patterns of online activities in the US states are remarkably consistent in intensity, even with a half decade difference and a more traditional set of ICT access variables in 2010 versus a more contemporary set of e-communication, e-commerce, e-entertainment, e-health, and other online uses in 2015.

6. Regression Results

The regression findings (in Tables 2, 3) are highly significant overall, with adjusted r square values that vary between .307 and .692, all highly significant. The regression diagnostics are nearly all supportive that the OLS regression assumptions have been met. The exceptions are three significant Jarque-Bera indicators, for the dependent variables of online shopping/travel, online banking/bill paying, and browsing the web. In these instances, the skewness and kurtosis of the regression residuals do not correspond to a normal distribution; hence these regression findings will be interpreted cautiously. Moran's I tests on the

regression residuals all indicate lack of spatial bias demonstrating the model's overall robustness.

For the demographic dimensions, the strong inverse association with age is apparent for about a third of the regressions. The dependent factors that are age-associated tend to involve internet uses that are characteristic of younger users, including listening to music, radio, and podcasts, watching videos online, using social networks, and texting/instant messaging. This corresponds to prior research on advantages of youth for internet use [4,10,19,27]. Urban population is also a significant correlate for about a third of regressions, and some of these effects are characteristic of urban and city environments, for example using online location-based services and connecting to household devices. The findings for gender indicate positive associations, i.e. ones reflecting male influence, on online banking/bill paying, participation in online video calls, and online education/training. Further research is needed to confirm these effects and to gain understanding of the reasons. We argue, without proof, that the enhanced male online education/training might reflect the national declining sex ratio of students enrolled in universities, so males might have more motivation to take up online for lower traditional educational participation.

The findings for ethnic/racial groups show that for Asians there is mostly no association with internet use. There are three positive associations, for watching videos online, monitoring health online, and texting/instant messaging, as well as one surprising inverse association for online banking/bill paying, which is unexplained. For Hispanics, again associations are lacking, with only two inverse effects, for online health monitoring and social networking. The predominant lack of effect across the dependent variables signals that Hispanics are converging with the population as a whole, a gap narrowing that is also reflected over time in a Pew study, in particular for 2009-2015 [18]. For African Americans there are mixed findings. Higher African American proportion in states is linked to reduced online video watching, looking up health information online, and online health monitoring, findings in concert with some prior studies [3,4,19,22]. On the other hand, proportion African American is related to texting/instant messaging, similar to the positive relationships reported for recent data at the county level [25]. The latter result points a technology area where Black usage has been converging with other ethnic/racial groups, which is in concert with a long-term, gradual convergence on internet use seen in a Pew survey for 2000-2015 [18]. Overall, the ethnic/racial findings are slight across the 17 use variables, implying that ethnicity/race is overall of limited importance for internet use at the state level.

Results for economic factors are moderate and selective across the many internet/e-communication uses. Income, a traditionally strong and consistent correlate of technology use, has limited impact, with associations only for online shopping/travel reservations, e-mail use, and browsing the web. The former effect indicates that higher-income states have higher prevalence on online shopping/travel reservations, which can be ascribed to greater relative purchasing capability for consumption of online goods and services. In higher-income states, e-mail communications may be comparatively more affordable for the populace, whereas in low income states, less expensive alternative of texting/instant messaging may have greater relative usage. Possibly this argument would apply to the affordability of browsing the web, but further research is called for. Findings indicate that more creative occupational categories are associated with listening to music/radio/podcasts, watching videos, participating in online video calls, online education, and online job searching. Most of these uses are ones more creative occupations would be attracted to; the explanation for increased online job searching is unexplained.

There is almost no association of government funding of higher education with internet/e-communication uses. This is surprising given extensive research corroborating education's importance for ICT and internet use (for instance, [4,19,22]). We argue that this might relate to the variable which emphasizes educational funding, whereas much of the prior literature includes educational attainment instead. More investigation is needed to determine why the funding dimension of education does not predict internet use in states. Innovation has strong associations, in particular presence of tech establishments is associated with online banking/bill paying; online health information gathering, monitoring and health insurance look up; online education/training; and online location services. These uses are more complex and often involve greater knowledge and interaction from the user, implying that states with higher innovation have users who are somewhat more sophisticated in online uses. Two innovation results are unexplained, the association of R&D expenditure and e-mail use and the inverse relationship of tech establishment presence with online job searching. Full explanation of the latter could be investigated through a future survey study on how tech industry workers procure job information.

Infrastructure has only slight relationship with internet usage and e-communication. The three variables register only three positive influences: fiber coverage on telework over the internet and broadband speed with online video calls and online job search.

Because telecommuters may be located outside metropolitan or urban areas, it makes sense that greater fiber coverage in a state would relate to more telecommuters using the internet. Likewise higher broadband speeds statewide would be encouraging for online video calling, which is bandwidth intensive. The relationship of broadband speeds with online job search might be explained by the appeal of quicker and more powerful online job searching made possible in more intensive broadband settings.

Affordability is associated only with e-mail use and text/instant messaging (Table 4). This association is more interpretable for lower income and elderly users, for whom affording these entry-level e-communication uses would be an important first step into advanced e-communication and broader internet use. Further study is needed to corroborate this finding and analyze it further relative to entry-level users.

Measures of societal openness have slight impacts. Election performance is associated only with looking up online health/insurance records and with browsing the web, effects that are unexplained; while overall freedom index has no associations. For sales of newspapers, periodicals/books, there are associations with looking up health/insurance records, with social networking and browsing the web. The former is unexplained, while latter two can be interpreted as citizens in content-rich states tending to browse for content more and exchange information more through social networking.

For social capital, although proportion of immigrant population in a state has no association with internet use and e-communication, the social capital index has strong associations for online shopping / travel reservations, listening to music/radio/podcasts, looking up health information online, and e-mail. These findings reinforce that social capital can be highly important for technology use [1,4,21]. The mechanisms underlying these specific associations involve processes of interpersonal networking and communications stimulating certain online uses/e-collaborations, which bear further study.

7. Discussion of Results

This research has provided a profile of the geographic arrangement of internet/e-communication uses for the states of the US, as well as the associations of demographic, social, economic, infrastructural, and social capital variables in eight prominent categories of independent variables. This broad study is based on the conceptual model, a variation of the SATUM model, described earlier. The model is tested through recent availability of internet use data from the NTIA.

The geographic findings demonstrate in several ways that internet/e-collaboration variables on purposeful uses tend to agglomerate together in statewide groupings of high intensity and low intensity of use. For each of the 17 dependent variables, testing of Moran's I indicates that 11 of them have significant positive spatial autocorrelation, which implies the individual variable in its geography has agglomeration between groups of neighboring states, and the other six variables have random arrangements spatially.

Regarding independent-variable associations with internet use and e-communications, nearly all the independent variables were influential on some outcome variables, with the exception of overall freedom index and immigrant population. The presence of predominantly influential predictors, which largely have no spatial bias, validates the conceptual model. The absence of association of immigrant population might be due to immigrant groups' social capital being more focused on sharing common experiences, networking physically with others, forming support groups, rather than influencing online activities. Overall freedom index may be less important a differentiating factor for states in a highly democratic nation, which is distinctive from major significant associations with ICT identified for a freedom factor for nations in the African continent [21]. Immigrant population and Freedom Index are therefore recommended to be dropped from the conceptual model as applied to US states.

The most important independent variables overall in this study are age, race/ethnicity, innovation with the proxy of tech establishments, urban location, managerial/science/arts occupations, and social capital. It can be argued that these factors typify the citizenry of states that are highly successful in contemporary online uses. Perhaps middle-class citizens in Washington DC would be an example or residents of Silicon Valley and Silicon Beach in California. At the same time, a finer-grained inspection across 17 different outcome variables indicates that the "young, contemporary" variable set is only partially expressed for each individual regression. In other words, portions of this overall set of factors influences one variable, e.g. age, ethnicity, tech establishments, and social capital influence looking up health info online, whereas tech establishments and urban population influence online location services.

Most of the mechanisms of use implied by our empirical findings can be explained referring to van Dijk's theory. Age and race/ethnicity are positional characteristics that encourage development of social, cultural, material, and behavioral resources [30]. Innovation leads to improved hardware, software, and extent of content, which in turn encourage the

individual to progress from access motivation to actual usage. Urban location is associated with several of van Dijk's favorable positional characteristics, including education and labor force position i.e. occupation. The latter is also supported by the present finding of importance of managerial/science/arts occupations. Favorable position leads to the individual obtaining material, social, and cultural resources. Findings for social capital are interpreted as follows: greater ICT usage favoring greater social participation through physical "social networks" [30]. The mechanism for presence of tech establishments, according to van Dijk's theory, is through its encouragement of the technological underpinnings of software, hardware, and content, which in turn influence the individual's progression from motivation to access ICT to material access, skills access and finally to usage access [30].

Finally, the geographic findings of this study are supported by van Dijk's including "space/geography" as a part of how ICT access encourages participation in society i.e. purposeful use of ICT [30]. In the present research, geography favors purposeful use in certain US states, which over time feeds back on enhancing favoring positional categories particularly education and labor force.

8. Policy Implications and Limitations

Policies considered in this section are based on the findings in this research study and are focused on government policies at the US state and federal levels. Some policies might also be considered by nonprofit organizations such as universities, public TV and radio, and foundations providing educational support. Most policies being recommended are long-term for achievement of positive policy outcomes, so government steadfastness and patience are necessary.

State and federal policies can be focused on providing training, education, and technology access to technology-deprived or trailing age categories, which are late middle age or elderly persons. For instance, programs might be set up to start or enrich e-communication and purposeful internet use in elder communities. For late middle-aged displaced workers, services and equipment could be supported for public centers of retraining that emphasize technology.

Hispanic and African Americans are seen from this and other studies to be two ethnic/racial groups that could benefit from special training in e-communication and internet uses. Although this has been attempted in certain states and cities, it has sometimes failed due to weak, one-shot approaches that ignore the continuing need for follow-up and opening up opportunities to use the training [12]. The policies recommended include a longer-term involvement based on metrics of ongoing

use by trainees. Specific uses identified by the present study as currently not being associated with Hispanic or Black citizenry can be emphasized more, but also a generic long-term gap is present for these two groups [18], which comprehensive training could aspire to narrow or eliminate. In a higher education setting, technology-enhancement opportunities for these minorities could be supported, and the study points to particularly large benefits for students entering managerial, science, and arts occupations.

State or federal policies could establish incentives and funding to stimulate development of technological industry, especially in states with smaller tech sectors, for example Mississippi and West Virginia. Also, policies to increase income levels in low-income states would have benefits for particular types of internet or e-communication uses.

This research indicates that social capital has particularly strong impacts on personal uses such as online shopping/travel, listening to music/radio/podcasts, and online look-up of health information. Long-term government policies could set a goal of encouraging growth in social capital in communities across a state, for example by providing support to nonprofit organizations with purposes benefitting society or funding universities to provide outreach to community organizations.

Lastly, geographically informed federal policies can be developed which provide greater technology-related assistance to states weaker in online and e-communication uses. Such policies could be further targeted to rural areas within those states, based on the enhanced technology levels in urban settings for some of the study's dependent variables.

The study has limitations. Although it is able to identify associations between independent and dependent variables, it does not study mechanisms for many of the identified associations. Methodologically it is limited by a relatively small sample size (n=51), which precludes structural equation modelling, path analysis, and other multi-path designs. The research provides findings only at the state level, so its conclusions would not apply for smaller geographies such as counties and cities, for which testing, based on models attuned to those units of analysis, would be necessary. This study also is limited to a cross-sectional view, whereas a continuation of NTIA's annual survey in future years would allow statistical analysis of temporal changes.

9. Conclusions

This study is the first systematic attempt to examine purposeful internet use among U.S. users in contrast to prior studies which have investigated material access to

the internet. Spatial patterns and socioeconomic, innovation, social capital, affordability, and infrastructural dimensions of internet use in the United States are analyzed based on a contemporary set of online activities. As internet usage becomes untethered from PCs and migrates to mobile devices, wearables, and sensors, understanding extent and patterns of online activities for traditional usage such as instant messaging, to more sophisticated use such as connecting to the internet-of-things is imperative. This has implications for the usage dimension and related disparities within the broader context of the U.S. digital divide. This also represents a shift in the digital divide conversation, from technology access to actual usage, and heralds further investigation of longitudinal changes in internet use and patterns of online activities and subsequent focus on outcomes and impacts of internet use in ensuing research.

10. References

[1] R. Agarwal, A. Animesh, and K. Prasad, "Research note—social interactions and the 'digital divide': Explaining variations in internet use," Information Systems Research, 20(2), 2009, pp. 277-294.

[2] R. Azari, & J.B. Pick, "Technology and society: Socioeconomic influences on technological sectors for United States counties," International Journal of Information Management, 25(1), 2005, pp. 21-37.

[3] J. Chakraborty, & M.M. Bosman, "Measuring the digital divide in the United States: Race, income, and personal computer ownership", The Professional Geographer, 57(3), 2005, pp. 395-410.

[4] W. Chen, "The implications of social capital for the digital divides in America", The Information Society, 29, 2013, pp. 13-25.

[5] R. Florida, The Rise of the Creative Class – Revisited. New York: Basic Books, 2012.

[6] S. Gauvin, K. Granger, & M. Lorthiois. "Analog citizens". Electronic Commerce Research, 15(3), 2015, pp. 365-386.

[7] A. Gonzales. "The contemporary US digital divide: from initial access to technology maintenance", Information, Communication & Society, 19(2), 2016, pp. 234-248

[8] T. H. Grubesic, "A spatial taxonomy of broadband regions in the United States," Information, Economics and Policy, 18, 2004, pp. 335-358.

[9] E.J. Helsper, & R. Eynon. "Digital natives: where is the evidence?" British Educational Research Journal, 36(3), 2010, pp. 503-520.

[10] J.B. Horrigan and M. Duggan. "Home Broadband 2015," Pew Research Center, Washington, DC, December, 2015.

[11] Z. Huang. "A comprehensive analysis of U.S. counties' e-Government portals: Development status and functionalities", European Journal of Information Systems 16, 2007, pp. 149-164.

[12] L. Kvasny, and M. Keil, "The challenges of redressing the digital divide: a tale of two cities," Information Systems Journal, 16, 2006, pp. 23-53.

[13] G. McHenry, E. Carlson, M. Lewis, R.M. Goldberg, J. Goss, and C. Chen, "The Digital divide is closing, even as new fissures surface." The 44th Research Conference on Communication, Information and Internet Policy 2016.

[14] T. Nishida, J.B. Pick, & A. Sarkar, "Japan′s prefectural digital divide: A multivariate and spatial analysis," Telecommunications Policy, 38(11), 2014, pp. 992-1010.

[15] NTIA. Exploring the Digital Nation: Embracing the Mobile Internet, National Telecommunications and Information Administration, Washington D.C., 2014.

[16] NTIA. Exploring the Digital Nation: America's Emerging Online Experience. National Telecommunications and Information Administration, Washington D.C., 2013.

[17] OECD. Understanding the digital divide. Organization for Economic Cooperation and Development, Paris, France, 2001.

[18] A. Perrin, and M. Duggan, "Americans' internet access: 2000-2015", Pew Research Center, Washington, DC, June, 2015.

[19] J. Pick, & A. Sarkar, "Theories of the Digital Divide: Critical Comparison," In Proceedings of 49th Hawaii International Conference on System Sciences (HICSS), pp. 3888-3897, IEEE, 2016.

[20] J.B. Pick, & A. Sarkar, The global digital divides: Explaining change, Springer, 2015.

[21] J.B. Pick, A. Sarkar, & J. Johnson, "United States digital divide: State level analysis of spatial clustering and multivariate determinants of ICT utilization," Socio-Economic Planning Sciences, 49, 2015 pp. 16-32.

[22] J. E. Prieger, "The broadband digital divide and the economic benefits of mobile broadband for rural areas", Telecommunications Policy, 37(6), 2013, pp. 483-502.

[23] P. Pruulmann-Vengerfeldt, "Exploring social theory as a framework for social and cultural measurements of the information society", The Information Society, 22(5), 2006, pp. 303-310.

[24] R. Putnam, "Social capital: Measurement and Consequences," Canadian Journal of Policy Research, 2, 2015, pp. 41-51.

[25] A. Sarkar, J. Pick, & G. Moss, "Geographic Patterns and Socio-Economic Influences on Mobile Internet Access and Use in United States Counties," In Proceedings of the 50th Hawaii International Conference on System Sciences (HICSS), pp. 4148-4158, IEEE, 2017.

[26] A. Sarkar, J. Pick, & J. Rosales, "ICT-enabled E-Entertainment Services in United States Counties: Socio-economic Determinants and Geographic Patterns," in Choudrie J., Kurnia, S. and Tsatsou, P. (Eds.), Innovative ICT enabled Services and Social Inclusion: Routledge, 2017.

[27] J.C. Sipior, B.T. Ward, and R. Connolly. "The digital divide and t-government in the United States: using the technology acceptance model to understand usage". European Journal of Information Systems, 20, 2011, pp. 308-328.

[28] D. E. Sylvester, & A. J. McGlynn, "The digital divide, political participation, and place", Social Science Computer Review, 28(1), 2010, pp. 64-74.

[29] J. van Dijk, & K. Hacker, "The digital divide as a complex and dynamic phenomenon", The Information Society, 19(4), 2003, pp. 315-326.

[30] J.A.G.M. van Dijk, The Deepening Divide: Inequality in the Information Society, SAGE Publications, Thousand Oaks, CA, 2005.

[31] M. Warschauer, and T. Matuchniak, "New technology and digital worlds: Analyzing evidence of equity in access, use, and outcomes", Review of Research in Education, 34, 2010, pp.179-225.

[32] White House, Mapping the digital divide. Council of Economic Advisers Issue Brief, 2015.

Table 2. Regression Findings: E-Commerce, E-entertainment, E-health, Telework

Category	Variable	Online Shopping, Trv. Resrv.	Online Banking, Bill Pay	Listen to Music, Radio, Podcast	Watches Videos Online	Looks up Health Info Online	Online Health Monitoring	Looks up Health, Insurance Record Online	Telecommutes using internet
		E-Commerce		E-Entertainment		E-Health			Telework
Demographic	African American				-.237*	-.347**	-.321**		
	Asian		-.252**		.272*		.264*		
	Hispanic						-.270*		
	Male Pop. 15+/Female Pop.15+ *100		.553***						
	Median Age	-.309*		-.687***	-.306**	-.271*			
	Urban Pop.		.360**						.293*
Economic	Median Household Income	.348*							
	Mgmt, Bus, Science, Arts Occup.			.460*	.583***				.663***
	Service Occupation			.297**					
Education	Higher Ed. Govt. Funding		-.257**						
Innovation	Tech Establishments		.539***			.284*	.596***	.459***	
	R&D Expenditure						.281*		
Infrastructure	Broadband Speed				0.187				
	Pop. Covered by Fiber			-.209*					.153*
	Pop. Covered by Wireless								
Affordability	Internet too expensive								
Societal Openness	Election Performance Index							.300***	
	Overall Freedom Index	0.173							
	Sales revenues of newspapers, periodicals, books, etc.($)/Total Sales Revenues($)	.200*						.396***	
Social Capital	Immigrant Pop.								0.215
	Social Capital Index	.547**		.458*		.619***			
	Adjusted R-squared	.662***	.739***	.643***	.504***	.507***	.425***	.661***	.801***
	Moran's I of Dependent Variables	.181***	.089**	0.013	-.013	.149***	-.005	.103**	.241***
	Joint Wald	1258.797*	1327.443*	694.764*	272.572*	455.721*	404.846*	589.562*	842.204*
	Koenker	17.376	12.811	17.120	16.592	11.716	23.468	16.800	17.003
	Jarque-Bera	7.662*	9.858*	4.108	0.823	1.810	1.744	1.644	0.361
	Moran's I of Regression Residuals	-.010	-.056	-.033	-.033	-.017	-.108	-.028	-.033

* p<.05, **p<.01, ***p<.001

Table 3. Regression Findings: E-Communication and other online activities

Category	Variable	Email	Social Network	Text, Instant Messaging	Participate in online video call	Education, Training Online	Job Search Online	Online Location Servcs.	Browses Web	Connects to Households Device
		E-Communication				Misc. Personal Use				
Demographic	African American			.304***						
	Asian			.268***						
	Hispanic		-.360**							
	Male Pop. 15+/Female Pop.15+ *100				.325***	.617***				
	Median Age		-.514***	-.599***						-.221
	Urban Pop.				.252*		.507***	.311*		.464***
Economic	Median Household Income	.454***							.362**	
	Mgmt, Bus, Science, Arts Occup.				.775***	.324*	.789***			
	Service Occupation				.313**					
Education	Higher Ed. Govt. Funding									
Innovation	Tech Establishments					.437***	-.506***	.484***		
	R&D Expenditure	.251**								
Infrastructure	Broadband Speed	0.192			.191*		.281**			
	Pop. Covered by Fiber									
	Pop. Covered by Wireless									
Affordability	Internet too expensive	.295**		.211*						
Societal Openness	Election Performance Index								.246*	
	Overall Freedom Index									
	Sales revenues of newspapers, periodicals, books, etc.($)/Total Sales Revenues($)		.246*						.258*	
Social Capital	Immigrant Pop.									
	Social Capital Index	.372**								
	Adjusted R-squared	.547***	.290***	.692***	.690***	.575***	.598***	.513***	.331***	.307***
	Moran's I of Dependent Variables	.127***	-.064	0.051	.090**	.079*	.162***	.159***	0.06	.068*
	Joint Wald	773.959*	237.124870*	723.238*	938.576*	903.784*	255.139*	225.563*	205.853*	140.765*
	Koenker	15.301	21.772	31.915	22.273	16.528	14.097	26.046	11.978	13.706
	Jarque-Bera	1.145	0.160	0.068	2.084	0.618	1.210	1.037	7.198*	0.783
	Moran's I of Regression Residuals	-.045	-.033	-.066	-.077	-.033	-.033	-.033	-.011	-.0118

* p<.05, **p<.01, ***p<.001