The New Digital Divide: a Computational Analysis of Smartphone Dependence

Laurence (w4rner) Warner* Faculty Advisor: Benjamin Soltoff†‡

October 2019

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree in the Master of Arts in Computational Social Science

*MA Candidate in Computational Social Science, University of Chicago. UK Fulbright Scholar 2017-19. w4rner@LaurenceWarner.com, w4rner.tech.

†Lecturer in Computational Social Science, University of Chicago, soltoffbc@uchicago.edu

‡Special thanks also due to the following. Advice: Nora Nickels and Sanja Miklin, Preceptors, Computational Social Science, University of Chicago; Prof. Paolo Parigi, Associate Director for Computational Social Science, IRiSS Stanford. Support: D3.js graphs: Enrique Collin and Sara Zhou. Research Assistance on literature review: Natalie Rose Johnston, co-presenter at the 2018 Fulbright Conference Poster Session 'Life without a smartphone in 2018'; Bhargav Srinivasa Desikan & Abe Pandit, team-mates for FirstWeekFlipPhone.com in 2019 Social New Venture Challenge.

Abstract

Unprecented numbers of Americans have crossed the ”digital divide” by gaining internet access this millenium, firstly through home internet, and now increasingly through mobile devices, leading many scholars of inequality to abandon their previous focus on social groups different access to technology to differential usage. However, in this paper we argue that recent digital divide literature on counter-productive smartphone usage in disadvantaged groups in fact necessitates a re-consideration of device ownership; and in particular of the growing phenomenon of Smartphone Dependence, in which a person is reliant upon their mobile device to access the internet. We bring a previous landmark analysis of Smartphone Dependence up to date, by applying the latest computational tools including complex visualization and Machine Learning algorithms to the latest survey data. Though the great wave of smartphone adoption was led by elites, our best statistical model suggests that by contrast Smartphone Dependence is associated with socioeconomic disadvantage. Smartphone Dependence therefore represents an important new digital divide worthy of further inquiry.

Keywords— machine learning, data visualization, d3.js, digital divide, smartphone dependence, socioeconomic inequality, ownership gap, usage gap

1 Introduction

The internet has caused an overhaul in American society, with the Economist describing it as representing the 'third industrial revolution' (Economist 2012). During the new millennium, Americans gradually began to consume their communication, information and entertainment online (Marler 2018).

Social scientists interested in economic inequality have taken an interest in how technologies spread through society, and the 'digital divide' - the idea that there are those who are reached by technological change and there are others are left behind - has become the dominant heuristic for technological inequality (Norris et al. 2001; Marler 2018).

This heuristic sufficed for empirical analysis of the first decade of the millennium. America's middle classes were able to purchase internet-enabled desktop computers and home broadband to bring the internet into their homes, whilst the internet have-nots, apart from some access promoted by public institutions like schools and libraries, were lagging behind in the internet revolution (Van Dijk 2005; Leigh and Atkinson 2001).

In the second decade, internet access in the US became near-universal, partly due to the widespread availability of cheap smartphones. In response, scholars have started to look beyond mere access (Level 1) and consider differential usage (Level 2) of technology, for example how smartphone devices are particularly susceptible to counter-productive outcomes, such as overuse of leisure apps.

However, in this paper we argue that we must again consider the original question of technology ownership, and in particular whether individuals have the means to access the internet without using a smartphone. In particular, we focus on Smartphone Dependence, which broadly refers to the growing phenomenon of people who rely upon their mobile devices to connect to the internet. We build upon the empirical approach of (Tsetsi and Rains 2017) to investigate their landmark finding that Smartphone Dependence is situated in lower SES groups.

Firstly, we make use of a more recent dataset. Secondly, we make significant computational innovations: including using complex data visualizations (custom-built using D3.js) to explore outliers; and building more sophisticated statistical methodologies. In particular, we train supervised machine learning algorithms to analyze the effect of different explanatory variables on Smartphone Dependence. Our best model brings their finding up to date: Smartphone Dependence is most concentrated in disadvantaged socioeconomic groups in society.

2 Literature Review

2.1 Theory of the digital divide

In a vein of sociology hearkening to Marx's focus on the technological means of production, a digital divide is said to exist when structural factors influence who in society owns and uses certain technologies.

2.1.1 Level 1: Ownership Gap

Early scholarship of the digital divide regards the Ownership Gap. New technology is often expensive: only those in the economic elite can afford them. Early research (Leigh and Atkinson 2001; Van Dijk 2005) focused on the phenomenon of poorer parts of society missing out on adopting broadband internet in their home . As (Veblen 2017) outline, policy makers generally hoped the digital divide problem would be solved when a country's internet access reached near saturation.

As we visualize in Figure 1, internet access in the United States is indeed reaching near saturation. That graph also shows a smartphone ownership quickly following the same trajectory. In 2007 when Apple launched the iPhone, the first user-friendly fully touchscreen design of the internet-enabled smartphone, at a base price of $499, it seemed that many would be still be left out. However, the Open Handset Alliance's development of the Android operating system and spread of cheaper smartphone devices (under $100), has led to wide uptake (Islam and Want 2014). A decade after Steve Jobs's famous 2007 launch, the vast majority of American adults own touchscreen internet-enabled devices (Arora 2019).

2.1.2 Level 2: Usage Gap

Despite policy makers hopes, economic inequality has not improved in the Internet Age (Newman 2013; Stern 2010). As (Marler 2018) outlines, in response to near universal internet access, digital divide scholars have shifted attention to inequalities in usage that persist even in a state of equal access. (Veblen 2017) shows how the earlier theory on device access was renamed 'Level 1', making way for a 'Level 2' digital divide: the Usage Gap.

This concept draws from the sociological theory of the Knowledge Gap: the situation when higher SES groups acquire information from media at a faster rate. The Knowledge Gap is expected to widen over time with a positive feedback mechanism: those who have more knowledge have the literacy to glean more (Lee and Yang 2014). By symmetry, the Usage Gap states that higher SES groups acquire the knowledge and skills to squeeze more benefits out of a given technology by the way they use it (Hamilton et al. 2018).

2.1.3 Usage Gap with smartphones

Conventional wisdom suggests that mobile devices should level the playing field, as they are designed be intuitive and widely usable. Further, due to their novelty, it is understandable that conventional wisdom states that mobile technology is associated with being 'savvy' and productive.

Productivity However, (Napoli and Obar 2016) questioned this conventional wisdom. Mobile technology is limited in two ways: firstly screen size and performance limitations mean that for many tasks - such as research, job applications, tax returns - smartphones are a less efficient device to desktops (Carroll, Heiser, et al. 2010). But it is their second argument which has more future relevance as tablet computers blur the lines between desktop and mobile device: according to the Usage Gap hypothesis, they show that disadvantaged members of society - whom they call "the emerging mobile underclass" - gravitate to less productive activities on the smartphone. This is unsurprising given revelations from Social Media smartphone apps about manipulative design and so-called 'attention hacking' (Papacharissi 2010; Taylor and Bazarova 2018).

In his recent blistering attack on development sector naivety over the benefits of smartphone adoption, Arora (Arora 2019) frames this division as the 'Leisure Divide'. Drawing on an sociology literature dating back to Veblen's 19th Century "Theory of a Leisure Class" (Veblen 2017), Arora identifies a strong tendency among lower SES groups, and in particular when using smartphones, for less productive endeavors. This version of the Usage Gap phenomenon has been observed across multiple societies: from China (Chang, Zhen, and Cao 2016) to Western Europe (Van Deursen and Van Dijk 2014), as well as in America.

Health There is a growing body of research around the concept of internet addiction and problematic internet use (Anderson, Steen, and Stavropoulos 2017; Kwon et al. 2013; Samaha and Hawi 2016; Parasuraman et al. 2017). The smartphone is an exceptionally impactful technological device upon mental health (Wolniewicz et al. 2018; Rozgonjuk and Elhai 2018). Smartphone usage has been shown to be related to poor sleep (Chang, Zhen, and Cao 2016), and smartphone addiction is a growing phenomenon (Parasuraman et al. 2017; Lin et al. 2017), (Liu et al. 2017).

Most strikingly, two recent experiments showed that smartphone usage can do tangible harm. A UT experiment found that the mere presence of a smartphone reduces one's available cognitive capacity (Ward et al. 2017), while the first experimental study of social media detox showed that social media apps were causally linked to depression and loneliness (Hunt et al. 2018).

Smartphone Dependence According to (Tsetsi and Rains 2017), Smartphone Dependence is strictly defined as someone who's only means of accessing the Internet is via a smartphone. More loosely, it is thought of as someone who can only readily access the internet via a smartphone, because they don't own the means necessary for the alternative. This is not a prima facie negative situation (it could just easily be called 'desktop/broadband free'), and in fact many see smartphone ownership as a cultural mark of entrepreneurship and success.

However, given the previously discussed literature on the usage limitations of the device, and the fact that lower SES groups seem to be particularly prone into falling into those patterns, we argue that Smartphone Dependence is indeed a negative scenario for disadvantaged groups. We posit that if it is situated in lower SES groups, policymakers should be seen as a flag for widening inequality.

To investigate whether this is the case empirically, (Tsetsi and Rains 2017) used the 2012 version of the Pew Research national survey, the gold standard for surveys on American digital life. They did not deploy sophisticated machine learning models nor perform complex Exploratory Data Analysis through visualization techniques. Nonetheless, through their more simplistic statistical analysis (chi-squared and ANOVA analysis) they found that lower SES groups were more likely to be Smartphone Dependent. They concluded that Smartphone Dependence was thus a worthy concept for further exploration in the digital divide space. We want to explore whether this general finding holds up with more recent data and more complex analysis.

3 Data

In an effort to promote reproducibility, we have included data and analysis in a public repository. [1]

3.1 2018 data

We use the 2018 version of the survey (Tsetsi and Rains 2017) used. The Internet Project Core Trends Survey can be accessed online [2]. Interviews with a nationally representative sample of 2,002 adults were conducted in January 2018. The target population for the study is non-institutionalized persons age 18 and over living in the US. Most of the interviews were conducted using cellphones (n = 1,502) with the remainder conducted using landlines (n=500); both groups were included in the final sample. We wrangled the raw dataset using the Python library Pandas, leaving us with a cleaned dataset of size N = 1561. For further details on this dataset, see Appendix A.

3.2 Data from previous years

As part of our efforts to leverage data visualization to produce novel insights, we explored the historical context more thoroughly. We cleaned and synthesised the past two decades of Pew

[1] Jupyter Notebook can be found in public repo: https://github.com/w4rner/ma-thesis
[2] http://www.pewinternet.org/datasets/

Internet surveys to form an original dataset [3].

[3] we have made this publicly available on the repo

4 Methods & Results

4.1 Descriptive Statistics

4.1.1 Historical Context

Figure 1: The Explosion of Smartphone Technology

As Figure 1 shows, the gradient of the smartphone time series graph is unprecedentedly steep. A mere half-decade after the iPhone launch, a majority of American adults, now 80%, own such a device. Internet access in total has broadly plateaued, while smartphone ownership is still rising: this gestures toward the trend towards Smartphone Dependence.

4.1.2 Defining our Variables

Outcome Variables For benchmarking purposes, we first apply our analysis to Smartphone Ownership, which can be seen in both Figures 1 and 2 to be at around 80%.

We follow the general approach of (Tsetsi and Rains 2017) in conceptualizing Smartphone Dependence as relying on a Smartphone for internet access. Because we wanted to focus our complexity on the analysis side, we followed Pew Research's own approach in simply defining Smartphone Dependence as smartphone users who lack home broadband facilities (Smith 2018). (Tsetsi and Rains 2017) decided to add more complexity at the variable definition stage, by defining three groups according to device ownership, as they wanted to focus more deeply upon device habits with their two hypotheses which are not relevant to our paper.

The following treemap (mosaic plot more specifically), shows the interaction between Smartphone and Home Broadband ownership:

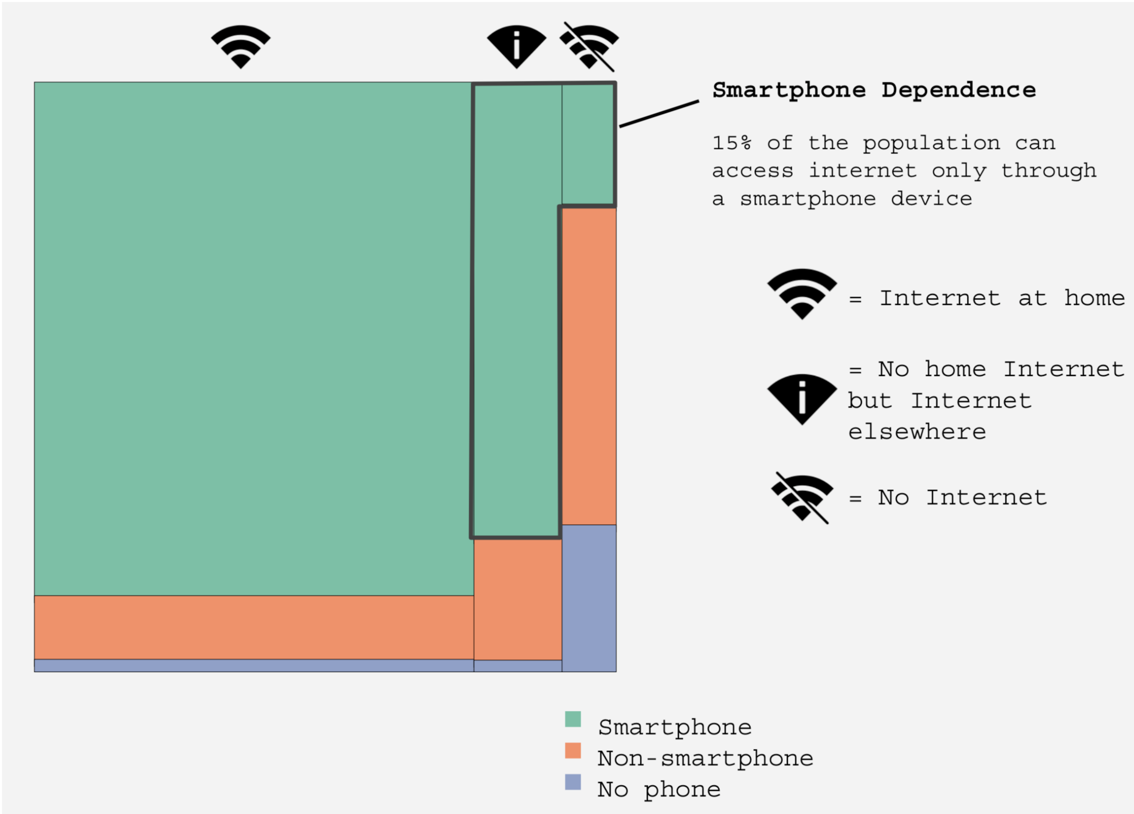

Figure 2: Empirically operationalizing Smartphone Dependence

In Figure 2, we annotate how our new binary variable of Smartphone Dependence is defined as owning a smartphone but not having home broadband we also include the top-right segment, who claim not to access the internet even though they own a smartphone. This constitutes 14% of our population, as opposed to the 3% of the population identified by the (Tsetsi and Rains 2017) definition using the 2012 data.

Categories Education and Income are the SES variables which we focus our attention on. Education is recorded as highest level of school attained, while income is recorded as estimate family income. They are coded as categorical variables.

As an instructive example of the Exploratory Data Analysis we conducted, Figure 3 is a visualization of how the Income variable varies with Smartphone Dependence.

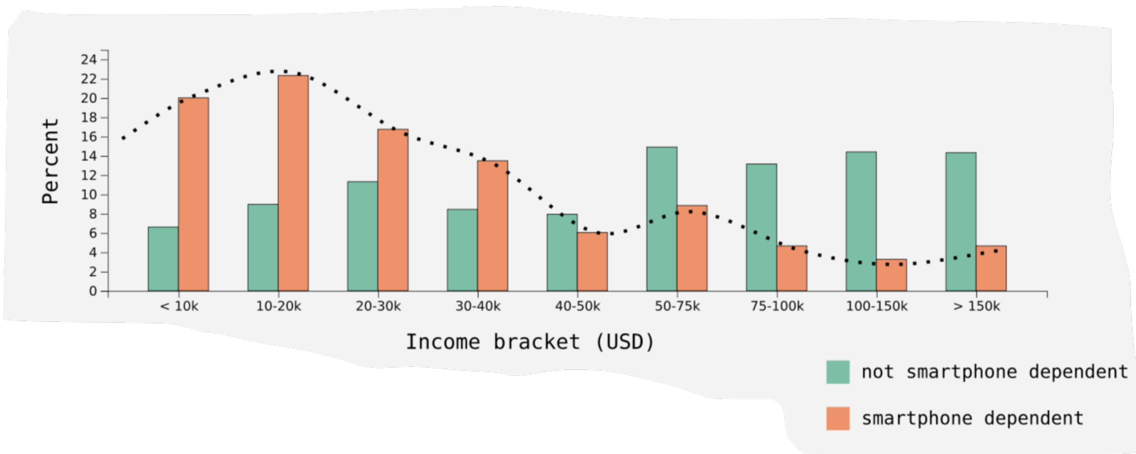

Figure 3: Income Histogram (Divided by Dependent Variable)

4.2 Model Selection

We wanted to use more sophisticated statistical models to analyse the social phenomenon of Smartphone Dependence. We treated this as a Classification Problem on the outcome variable of Smartphone Dependence. We ran a suite of different types of machine learning models. They are outlined below, along with the best accuracy scores for each. See B.3 for an example of the scikit-learn classification reports from which these results were drawn:

model	accuracy
Logistic Regression	87.34
Random Forest	86.29
Support Vector Classifier	85.86
K Nearest Neighbors	85.02
Decision Tree	78.27

Figure 4: Comparing Machine Learning Models

It is interesting to note that apart from Logistic Regression, on accuracy, all of the Machine Learning models are outperformed by a simple predictor that predicts non-Smartphone Dependence for every individual (giving an accuracy of 87%). However, in the Machine Learning literature, this is not uncommon in the case of unbalanced dependent variables like this (García, Sánchez, and Mollineda 2012).

For reassurance on Model Selection, Logistic Regression also ranked highest on average for Precision and Recall scores, so this model was selected for deeper analysis (see Appendix B.3). Incidentally, the Logistic Regression method is more easily human interpretable (see Appendix B.3.1 for explanation) which allows us to interpret our results more clearly.

4.3 Statistical Analysis

4.3.1 Benchmark Model: Smartphone Ownership as Outcome Variable

Before presenting our primary analysis, we present a similar multivariate Logistic Regression model on the outcome variable of Smartphone Ownership as a benchmark to allow for easier interpretation through comparison. Firstly, Figure 5 visualises the key results of the Logistic Regression analysis. It plots each estimated coefficient, with the 95% Confidence Interval as whiskers.

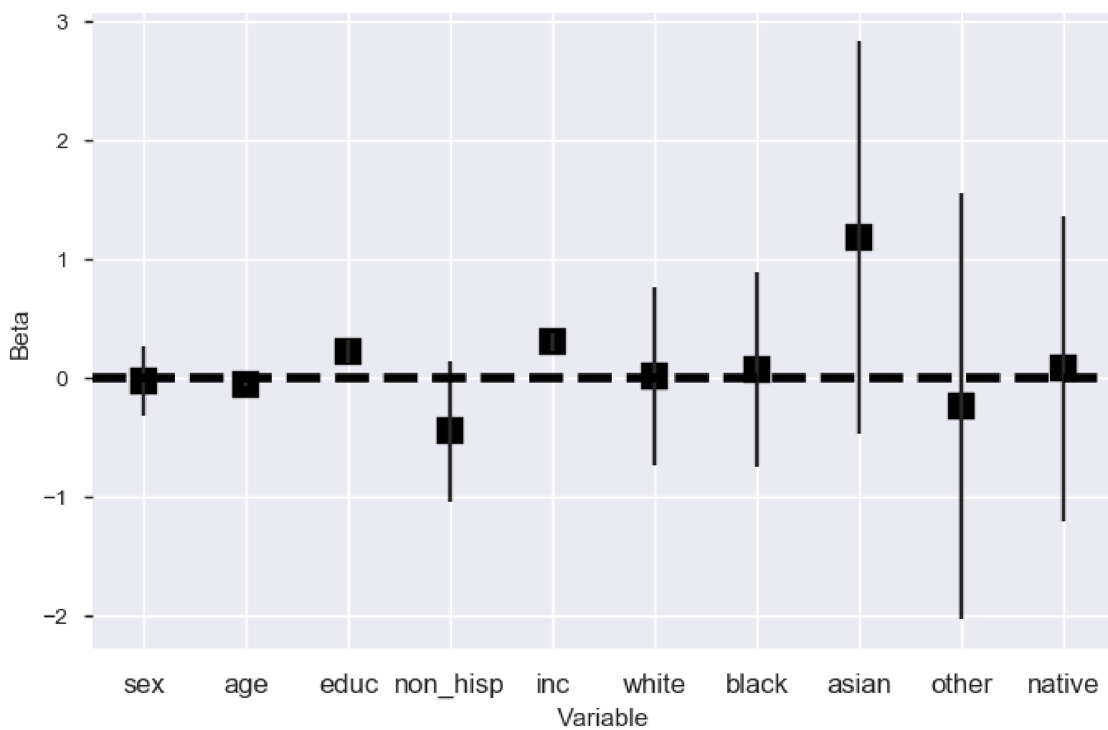

Figure 5: Benchmark Logistic Regression Results for Smartphone Ownership

Education and income are both statistically significant, with positive coefficients. Age is also significant.

4.3.2 Smartphone Dependence as Outcome Variable

Now let us consider the same analysis for our outcome variable of interest: Smartphone Dependence.

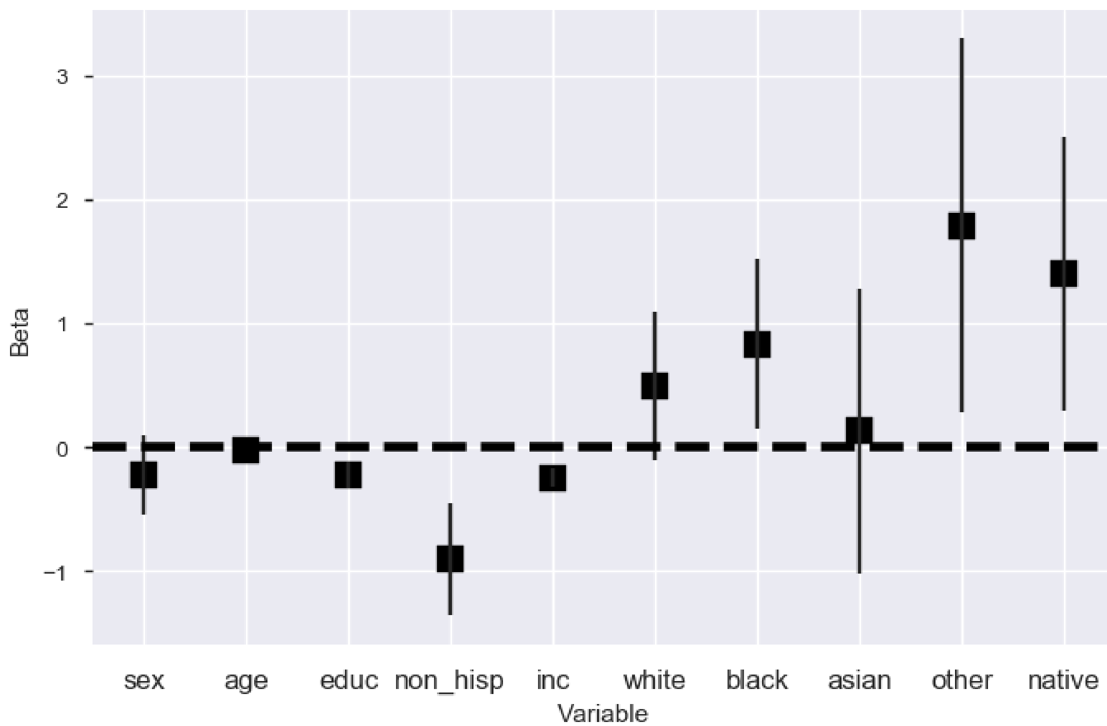

Figure 6: Logistic Regression Results: Smartphone Dependence

Let's observe what has changed. The SES variables Education and Income are again significant.

However, it is important to note that the sign of the coefficients are now negative. In addition, some race variables (non-hisp, black, other, native) are now significant.

4.3.3 Interpreting Results

The log odds ratios are as follows: Education - 0.80; Income - 0.78. For example for education: a one unit increase in Education leads the odds ratio to decrease to 80% of its original level.

Log-odds ratios are not highly intutive, so let's dive deeper into interpreting the effects of the variables, let's consider Marginal Effects analysis.

Firstly, the Marginal Effect for an 'average' individual in the population (mean value for every variable) is -0.021 for Education, and for Income is -0.019. In other words, a one-unit jump up the education hierarchy (such as going from attaining high school to college), leads to a 2.1% decrease in the likelihood of being Smartphone Dependent for the average individual.

Similar conclusions are drawn from analyzing the average Marginal Effects across the population: Education: -0.025, Income: -0.023. In other words, a one-unit jump up the education hierarchy (like from high school to college), leads to an average 2.1% decrease in the likelihood of being Smartphone Dependent for the average individual.

4.4 Complex Data Visualization

To borrow the famous distinction from (Tukey 1980) between Confirmatory and Exploratory data analysis, we have already demonstrated Data Visualization for purely Exploratory purposes, but recent developments in Data Visualization software, in particular the propagation of the D3.js library, allow us to create more complex custom visualizations which can be used for deeper Confirmatory analysis. For example, we can conduct interpretable analysis of variable interaction and outlier detection which may not otherwise be possible.

The following modified scatterplot uses color to encode the dependent variable, and the size of the circle to encode the weight of that category. This allows us to show the interaction between Education and Income on these variables in a static graph in Figure 7.

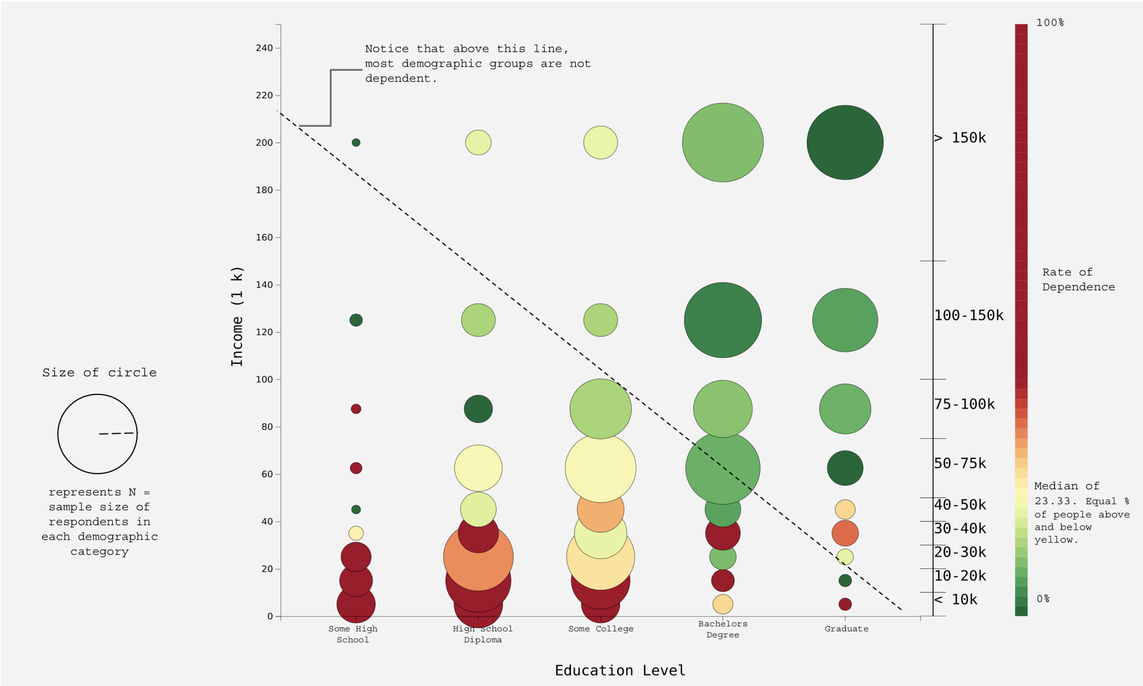

Figure 7: Complex Visualization: Education & Income

The graph echoes the statistical conclusion that there is a negative relationship between each variable and Smartphone Dependence, as Smartphone Dependence is mostly found in the bottomleft of the graph.

It also helps us identify some interesting outliers regarding education. Firsly, consider the green dots in the bottom-right corner. The education effect trumps the income effect for this highly educated segment, who tend to refuse Smartphone Dependence even when very poor. Secondly, consider folks with education of High School or Some College. Even though at lower income levels they are less likely to be Smartphone Dependent, at high income levels they show a median tendency for Smartphone Dependence. In both cases, the small weight of these groups implies that these should be seen as outlier sub-populations.

5 Discussion

5.1 Core Conclusions

With our more sophisticated analysis using the 2018 dataset, we build upon (Tsetsi and Rains 2017)'s findings from the similar 2012 dataset. Even though Smartphone Ownership itself is more prevalent among higher SES groups, Smartphone Dependence is most prevelant among lower SES groups. Though recent scholarship is correct that differential behaviors should be the primary focus for digital divide research, this finding reminds us that when looking at the Usage Gap, we should be mindful of the role that differential ownership in technology, and in particular the phenomenon of Smartphone Dependence, plays in guiding differences in behavior.

In addition, through our visualization methodology, we have identified two new subsets of the population who defy the general trend. Here I hypothesise potential qualitative descriptions of these groups. Firstly, the high-income low-education outlier group who show raised Smartphone Dependence levels, could be made up of young innovators who embody the stereotype of smartphone usage being associated with successful innovation. Secondly, the high-education, low-income outlier could well represent public sector workers, including academics, who have positions which, though lower income, may be stable and have work expectations which focus on desktop computer productivity in long-term projects.

5.2 Implications

Given the literature cited in 2.1.3, showing that smartphones are increasingly damaging people's health and economic prospects, it is concerning to see that groups which rely on these devices are those whose health & economic prospects need the most support in efforts to alleive economic inequality.

Our Smartphone Dependent group is around 14%, though as starkly visualised in Figure 7, for many disadvantaged sub-groups this figure is actually over 50% making it a majority. Smartphone Dependence seems to be on an upward trend. Though our increase from the 3% size in (Tsetsi and Rains 2017), is as much due to our more permissive definition, Pew Research's latest survey, released after this analysis, "Mobile Technology and Home Broadband 2019 survey" now puts this figure at 19% implying that this is indeed a rapidly developing trend [4]. Though the early Level 1 digital divide literature's focus on simply internet access is no longer relevant, digital divide scholars should not throw away the baby - of analyzing device ownership - with the bathwater. The phenomenon (Tsetsi and Rains 2017) first identified, and its increased prevalence, means that future digital divide research must again consider the diet of digital devices different groups have ready access to as a starting point.

[4]https://www.pewinternet.org/2019/06/13/mobile-technology-and-home-broadband-2019/

6 Limitations & Future Directions

6.1 Domain

6.1.1 Mobile devices

Future studies could seek to disambiguate the concept of Smartphone Dependence further, both theoretically in response to technological changes, and empirically.

The core theoretical concept of Smartphone Dependence, is that the person cannot access the internet readily on a more powerful device with a larger screen. Our method, using Home Broadband as a proxy for relying on cellular data to access the internet, will become a decreasingly accurate proxy. Firstly, free public wi-fi is increasingly widespread. Secondly, high-powered tablet devices - which do not fit easily into our binary categorisation of mobile and non-mobile devices - are increasingly gaining Cellular access. They should be given greater attention, as device diets in general should be, in future studies.

Another angle for deepening the study could be to distinguish between Android and iPhone users. University of Chicago scholars found iPhone ownership to be the product to be the single most precise predictor of being in the top income quartile of America (Bertrand and Kamenica 2018). Including such consideration may shed more light on the putative outliers we identified in Figure 7. For example, perhaps the 'innovators' are more likely to be dependent on higher-end Apple smartphones.

6.1.2 Emerging economies

As (Arora 2019) identifies, questions of smartphone usage have even more importance in emerging economies. It would be interesting to deploy the methods here on such data. On the one hand, smartphone device penetration is still catching up, so Smartphone Dependence may not denote lower SES in the same way (Poushter et al. 2016). On the other hand, there have already been some strong sociocultural analyses of device ownership (Ma, Chan, and Chen 2016), and some scholars argue that the Digital Divide has difficult theoretical implications outside the developed world (Rowsell, Morrell, and Alvermann 2017). Pew's 2019 survey of "Mobile Connectivity in Emerging Economies" could be an interesting place to start [5].

6.2 Data Analysis

6.2.1 Variable definitions

To deepen the analysis, more attention could be focussed on the Education and Income variables. A deeper approach to exploring the effect of these SES variables might take into account the following: how family size mediates this family income variable, how cultural capital and social class might mediate purchasing decisions.

In terms of statistical methodologies, it would also be interesting to binarize the variables and conduct analysis in that way, though this may prevent making the general, interpretable conclusions we have been able to make about the direction of the variables.

[5] https://www.pewinternet.org/2019/03/07/mobile-connectivity-in-emerging-economies/

6.2.2 Methods

To improve the accuracy of the Machine Learning models, one methodological approach would be to seek to artificially balance the classes of the outcome variable, perhaps using some techniques outlined in (Chawla 2009). Alas, time alone may also solve this problem, as it appears that Smartphone Dependence is becoming a decreasingly minority technological diet.

References

Anderson, Emma Louise, Eloisa Steen, and Vasileios Stavropoulos (2017). "Internet use and Problematic Internet Use: A systematic review of longitudinal research trends in adolescence and emergent adulthood". In: *International Journal of Adolescence and Youth* 22.4, pp. 430–454.

Andone, Ionut et al. (2016). "How age and gender affect smartphone usage". In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. ACM, pp. 9–12.

Arora, Payal (2019). *The next billion users: Digital life beyond the West*. Harvard University Press.

Bertrand, Marianne and Emir Kamenica (2018). *Coming apart? Cultural distances in the United States over time*. Tech. rep. National Bureau of Economic Research.

Carroll, Aaron, Gernot Heiser, et al. (2010). "An Analysis of Power Consumption in a Smartphone." In: *USENIX annual technical conference*. Vol. 14. Boston, MA, pp. 21–21.

Chang, Enyu, Feng Zhen, and Yang Cao (2016). "Empirical analysis of the digital divide from the perspective of internet usage patterns: A case study of Nanjing". In: *International Review for Spatial Planning and Sustainable Development* 4.1, pp. 49–63.

Chawla, Nitesh V (2009). "Data mining for imbalanced datasets: An overview". In: *Data mining and knowledge discovery handbook*. Springer, pp. 875–886.

Economist, The (2012). *The third industrial revolution.* URL: http://web.archive.org/web/20080207010024/http://www.808multimedia.com/winnt/kernel.htm (visited on 07/28/2018).

García, Vicente, José Salvador Sánchez, and Ramón Alberto Mollineda (2012). "On the effectiveness of preprocessing methods when dealing with different levels of class imbalance". In: *Knowledge-Based Systems* 25.1, pp. 13–21.

Hamilton, Emma C et al. (2018). "The digital divide in adoption and use of mobile health technology among caregivers of pediatric surgery patients". In: *Journal of pediatric surgery* 53.8, pp. 1478–1493.

Hunt, Melissa G et al. (2018). "No more FOMO: Limiting social media decreases loneliness and depression". In: *Journal of Social and Clinical Psychology* 37.10, pp. 751–768.

Islam, Nayeem and Roy Want (2014). "Smartphones: Past, present, and future". In: *IEEE Pervasive Computing* 13.4, pp. 89–92.

Kwon, Min et al. (2013). "Development and validation of a smartphone addiction scale (SAS)". In: *PloS one* 8.2, e56936.

Lee, Hyunwoo and JungAe Yang (2014). "Political knowledge gaps among news consumers with different news media repertoires across multiple platforms". In: *International Journal of Communication* 8, p. 21.

Leigh, Andrew and Robert D Atkinson (2001). "Clear thinking on the digital divide". In: *Progressive Policy Institute*, pp. 1–20.

Lin, Yu-Hsuan et al. (2017). “Incorporation of Mobile Application (App) Measures Into the Diagnosis of Smartphone Addiction.” In: *The Journal of clinical psychiatry* 78.7, pp. 866–872.

Liu, Yueheng et al. (2017). “Problematic smartphone use and its relation with anxiety and depression: a survey in a large sample of undergraduates”. In: *Chinese Journal of Psychiatry* 50.4, pp. 270–275.

Ma, Qi, Alan HS Chan, and Ke Chen (2016). “Personal and other factors affecting acceptance of smartphone technology by older Chinese adults”. In: *Applied ergonomics* 54, pp. 62–71.

Marler, Will (2018). “Mobile phones and inequality: Findings, trends, and future directions”. In: *New Media & Society* 20.9, pp. 3498–3520.

Napoli, Philip M and Jonathan A Obar (2016). “SECOND CLASS NETIZENS”. In: *Race and Gender in Electronic Media: Content, Context, Culture*, p. 79.

Newman, Nathan (2013). “The costs of lost privacy: Consumer harm and rising economic inequality in the age of Google”. In: *Wm. Mitchell L. Rev.* 40, p. 849.

Norris, Pippa et al. (2001). *Digital divide: Civic engagement, information poverty, and the Internet worldwide*. Cambridge University Press.

Papacharissi, Zizi (2010). *A networked self: Identity, community, and culture on social network sites*. Routledge.

Parasuraman, Subramani et al. (2017). “Smartphone usage and increased risk of mobile phone addiction: A concurrent study”. In: *International journal of pharmaceutical investigation* 7.3, p. 125.

Poushter, Jacob et al. (2016). “Smartphone ownership and internet usage continues to climb in emerging economies”. In: *Pew Research Center* 22, pp. 1–44.

Rowsell, Jennifer, Ernest Morrell, and Donna E Alvermann (2017). “Confronting the digital divide: Debunking brave new world discourses”. In: *The Reading Teacher* 71.2, pp. 157–165.

Rozgonjuk, Dmitri and Jon Elhai (2018). “Problematic smartphone usage, emotion regulation, and social and non-social smartphone use”. In: *Proceedings of the Technology, Mind, and Society*. ACM, p. 35.

Samaha, Maya and Nazir S Hawi (2016). “Relationships among smartphone addiction, stress, academic performance, and satisfaction with life”. In: *Computers in Human Behavior* 57, pp. 321–325.

Smith, Aaron (2018). *Declining Majority of Online Adults Say the Internet Has Been Good for Society*. URL: `https://www.pewinternet.org/2018/04/30/declining-majority-of-online-adults-say-the-internet-has-been-good-for-society/` (visited on 10/11/2019).

Stern, Michael J (2010). “Inequality in the Internet age: a twenty-first century dilemma”. In: *Sociological Inquiry* 80.1, pp. 28–33.

Taylor, Samuel Hardman and Natalya N Bazarova (2018). “Social media and subjective well-being: A relational perspective”. In: *A networked self and love*. Routledge, pp. 102–128.

Tsetsi, Eric and Stephen A Rains (2017). “Smartphone Internet access and use: Extending the digital divide and usage gap”. In: *Mobile Media & Communication* 5.3, pp. 239–255.

Tukey, John W (1980). “We need both exploratory and confirmatory”. In: *The American Statistician* 34.1, pp. 23–25.

Van Deursen, Alexander JAM and Jan AGM Van Dijk (2014). “The digital divide shifts to differences in usage”. In: *New media & society* 16.3, pp. 507–526.

Van Dijk, Jan AGM (2005). *The deepening divide: Inequality in the information society.* Sage Publications.

Veblen, Thorstein (2017). *The theory of the leisure class*. Routledge.

Ward, Adrian F et al. (2017). “Brain drain: the mere presence of one?s own smartphone reduces available cognitive capacity”. In: *Journal of the Association for Consumer Research* 2.2, pp. 140–154.

Wolniewicz, Claire A et al. (2018). “Problematic smartphone use and relations with negative affect, fear of missing out, and fear of negative and positive evaluation”. In: *Psychiatry research* 262, pp. 618–623.

A Data

Whilst major ground has been broken regarding smartphone usage habits using the rich data collection that such devices enable, like the Menthal Project or Cambridge University's Device Analyzer, both research projects based upon Android apps, unfortunately such studies exclude a population necessary for our study: non-Smartphone users (Andone et al. 2016). Although there are some issues with surveys such as response accuracy. the best approach is a nationally representative survey, and Pew Research is indeed the best.

According to the Pew Research Center, the landline sample was collected using a proportional sample based on listed telephone households, 3-10 January 2018. The cellphone sample was selected systematically from dedicated wireless numbers. Random digit dialing was used to collect survey responses and the final sample was weighted to represent the American adult population. The sample response rate was 11%, though Pew's statistical corrections remove fear of sampling bias.

In our wrangling, we discarded rows which contained answers including 'Don't Know' or 'Refused to answer'. Our N = 1561 is still easily enough to draw statistical inference in the models we run.

B Methods & Results

B.1 Explanatory variables

We adapted age from a continuous variable into a categorical variable. We left ordered categorical variables - education, income - as ascending integers. Even though this is not a completely accurate scale, it is is a sufficient approximation. We binarized race using one-hot encoding (having to discard the handful of respondents who answered 'Islander' to avoid multicollinearity).

B.2 Dependent Variables

In the process of coding Smartphone Dependence, both of the relevant variables (phone/broadband) were firstly coded into ternary variables of the following form: full ownership, partial ('dumb' internet-disabled phone/use external internet e.g. library) and none.

B.3 Model Selection

Because Logistic Regression performed highest on both Precision & Recall, we can infer that it's Precision-Recall curve would very likely be strictly superior to other models, and therefore deeper AUC-ROC analysis was deemed unnecessary.

Figure 8 shows an example of the Precision and Recall data we considered in the Model Selection process, specifically a Random Forest classifier with 500 estimators.

	precision	recall	f1-score	support
0	0.89	0.95	0.92	412
1	0.44	0.26	0.33	62
avg / total	0.84	0.86	0.84	474

accuracy score: 86.08

Figure 8: Example Classification Report of Machine Learning Models

B.3.1 Logistic Regression

Incidentally, Logistic Regression has the benefit of its results being interpretable.

As an instructive example, we will use the Outcome Variable of smartphone ownership (=1 if smartphone, =0 if none) and a univariate model which includes a single predictor variable: *sex* (=1 if female, =0 if male). The logit model selects beta coefficients in the following model to minimize the sum of the squared errors given the data:

$$logit(p) = log(p/(1-p)) = \beta_0 + \beta_1 * sex \tag{1}$$

Thus the exponent of β_1 can be interpreted as the impact of being female upon the odds ratio: the ratio of the likelihood of having versus not having a smartphone. If the coefficient is negative, being female makes you less likely to have a smartphone, and vice versa. Let's look at the model built from our data:

```
Optimization terminated successfully.
         Current function value: 0.483610
         Iterations 6
                           Logit Regression Results
==============================================================================
Dep. Variable:              smart_bin   No. Observations:                 1567
Model:                          Logit   Df Residuals:                     1565
Method:                           MLE   Df Model:                            1
Date:                Mon, 30 Jul 2018   Pseudo R-squ.:                0.005885
Time:                        03:02:08   Log-Likelihood:                -757.82
converged:                       True   LL-Null:                       -762.30
                                        LLR p-value:                  0.002742
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          2.0228      0.206      9.815      0.000       1.619       2.427
sex           -0.3863      0.129     -2.991      0.003      -0.640      -0.133
==============================================================================
```

Figure 9: Logistic Regression output

The negative coefficient (-0.39) shows that females are indeed less likely to own smartphones in our dataset. This is corroborated by a visual cross-tabulation of the data:

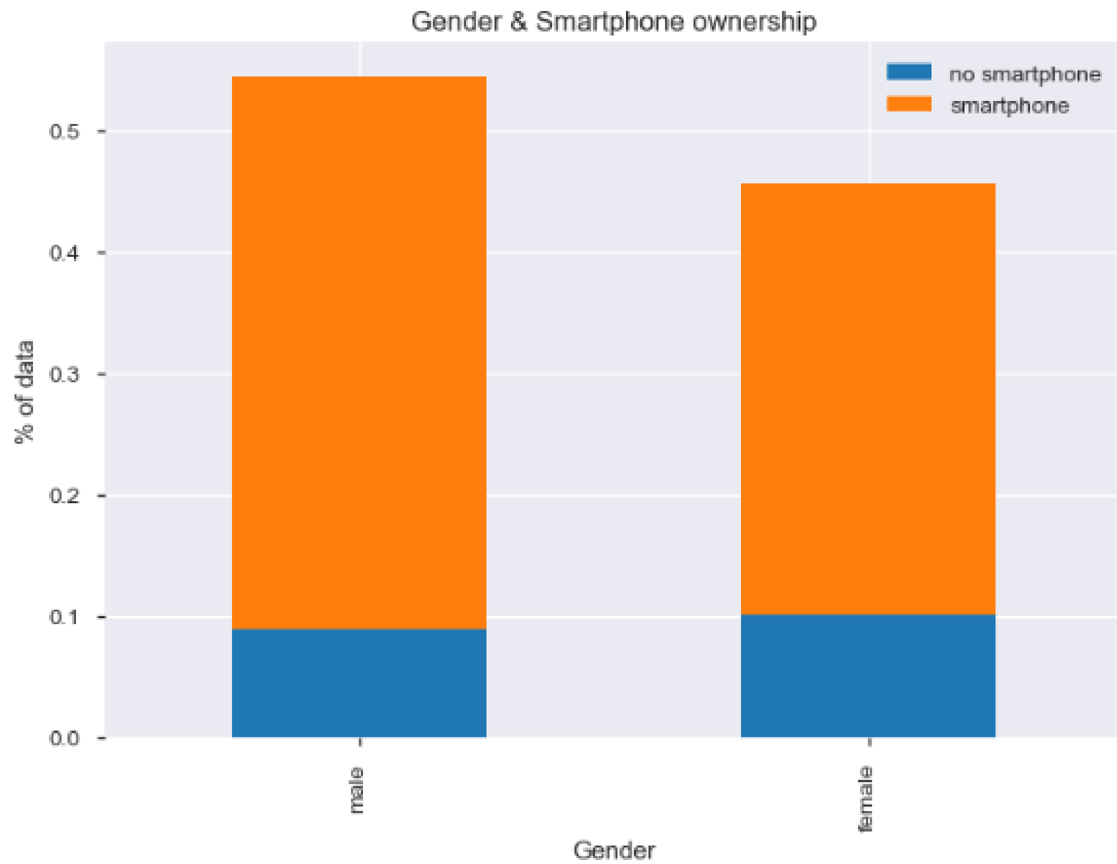

Figure 10: Gender & Smartphone Ownership

Moreover the fact that the p-value was below 5% means that this finding is statistically significant: it is not just a random artefact of a small sample.

We can be even more precise about our inference. By taking the exponent, we can show that the log odds ratio decreases to 0.69 (taking the exponent of the coefficient) from males to females.