Americans and digital knowledge: a snapshot of privacy and security

Moncef Belhadjali
Norfolk State University

Sami M. Abbasi
Norfolk State University

ABSTRACT

For online users, privacy and security constitute the two most troubling concerns. While online, individuals may use a credit card, access email, access social media sites, and store pictures through a cloud storage. The incidents of privacy and security breaches are well-documented throughout the Internet, and through the different news media platforms. Therefore, it in the pressing interest of online users to be well-informed and knowledgeable about the topics of privacy and security as much as possible. How much do Americans know about the topics of privacy and security, and how would they perform on a short quiz on these topics? Do characteristics such gender, age, and party affiliation provide a reliable set of predictors for individuals' knowledge on privacy and security?

This study reports on the analysis of data from a Pew Research Center survey of 3,951 individuals in the USA. The purpose of the analysis is to uncover underlying relationships between individual characteristics and the digital knowledge on privacy and security online. The results of Binary Logistic and Multiple Linear regressions showed that the increase in age is more likely to cause a decrease in the level of knowledge about privacy and security, and that females are more likely to know less about these topics. In addition, the level of knowledge on both topics would increase with Democrats. A series of Chi-square tests showed that that there are statistically significant relationships between individual performance on privacy and security questions and sex, age, and party affiliation.

Keywords: Privacy; Security; Data breaches; digital knowledge; online users

INTRODUCTION

Cyberattacks are well-documented (Aaron, 2021), for example, March 2, 2021 Microsoft Exchange servers were hacked by a group called Hafnium. April 3, 2021, hackers gained access to a database of 533M Facebook users from 106 countries. January 20, 2021, cybercriminals stole data about 2.28M users of the dating site MeetMindful. According to the FBI, there was a rapid increase in the number of cybercrimes since the beginning of the COVID-19 pandemic, as reported to the Internet Crime Complaints Center (IC3) (Miller, 2020). Some of the reported hacking attempts were malicious emails that use the subject coronavirus to trick users into clicking on dangerous links. During the second week of April 2020, Google reported that it blocked an average of 18M of daily malicious coronavirus messages to Gmail users (Sonnemaker, 2020). In addition to blocking these malware and phishing emails, Google blocked about 240M COVID-related daily spam messages.

To find out what consumers think about their privacy and the collection of their data by companies, McKinsey conducted a survey of 1,000 consumers in North America. The results revealed that, consumers are becoming increasingly conscious about what types of data they share—and with whom (Anant et al., 2020). In a study published by the Pew Research Center, the authors reported that only a small proportion (24%) of American citizens surveyed know that private browsing only hides online activities from the users of the same computers, and not from the website and the internet provider (Vogels & Anderson, 2019). "Consumers think they know more than they actually do about personal data security." (Gothmann, 2021). The author reports on a survey of 1,000 consumers from the U.S. and U.K. on their data privacy education, and understanding of how personal data is used. Some of the respondents (28%) indicated that they don't read the terms and conditions. Although, in the context of privacy and security, 82% of the respondents consider themselves proactive, survey results revealed that 44% have a basic understanding of ant-virus technology, 43% have a basic understanding of multi-factor authentication, and 33% have a basic understanding of encryption.

On social network sites, females displayed higher privacy concerns and behavior than did males (Tifferet, 2019). Studies on gender and age for Facebook users found that females and younger users were more concerned about the privacy of sharing their photos (Malik, et al., 2016). Citizens should be concerned about their privacy and security, "Today, however, the biggest risk to our privacy and our security has become the threat of unintended inferences, due to the power of increasingly widespread machine learning techniques." (Burt, 2019). The current study utilizes data from a survey of American citizens to investigate the relationships between individuals' characteristics and their digital knowledge on privacy and security online.

METHODOLOGY

Data

The data used in the study was obtained from the Pew Research Center. The data was collected via an online survey developed by the Pew researchers and administered during the period of June 3, 2019 to June 17, 2019. The respondents are members of the American Trends Panel (ATP), a pool of US citizens of age 18 and older. As a part of this survey, the respondents were asked to take a short quiz that included 3 questions covering privacy topics, and 3 questions covering security topics. The total responses were 4,273, however, this current study used 3,951

cases only. This is due to eliminating records with missing or incomplete data (i.e. respondent refused to answer) for the variables selected. The original data were transformed from multi-values to binary as follows:

- Score & Grade on Privacy questions:
  - Original data: Score: 0; Score: 33%
  - Recoded data: Fail: (0)
  - Original data: Score: 67%; Score: 100%
  - Recoded data: Pass: (1)
- Score & Grade on Security questions:
  - Original data: Score: 0; Score: 33%
  - Recoded data: Fail: (0)
  - Original data: Score: 67%; Score: 100%
  - Recoded data: Pass: (1)
- Demographic Data Coding:
  - SEX: Male: (0); Female: (1)
  - PARTY: Republican: (0); Democrat: (1)
  - AGE:
    - Original data: 18-29; 30-49
    - Recoded data: Under50 (0)
    - Original data: 50-64; 65+
    - Recoded data: 50Plus (1)

Table 1 (Appendix) provides a snapshot of the sample data (N = 3,951) used in this study. Table 2 shows the respondents' performances as continuous variables representing their actual scores on the privacy and security questions, ranging from 0% to 100%. Table 3 shows the respondents' performances as binary variables (Fail/Pass).

Analysis

The first phase of the analysis used two separate Binary Logistic regressions through SPSS. In this case, the dependent variables, representing performances on privacy (PRV) and security (SCR) questions, are binary with possible values Pass (0) and Fail (1). The independent variables sex, age, and party affiliation are categorical. The two models are:

- PRV = f (SEX, AGE, PARTY)
- SCR = f (SEX, AGE, PARTY)

The SPSS outputs for the model PRV = f (SEX, AGE, PARTY) are shown in Tables 4 and 5 (Appendix). The test for overall fit revealed that the model Chi-Square and p values indicate that the model is fitting the data significantly better than the No model. As indicated in Table 5, all variables have a statistically significant predictive power. The positive coefficient for Party (0.321) indicates that moving from Republicans to Democrats is likely to move the performance PRV closer to Pass. This is since Democrat was coded as 1 (Republican: 0). Also, since Female was coded as 1 (Male: 0), the negative coefficient (-0.721) for SEX indicates that the performance PRV would likely move closer to Fail with Females. In addition, the negative coefficient for Age (-0.644) indicates that an increase in age from Under50 to 50Plus is likely to move the performance PRV closer to Fail.

The SPSS outputs for the model SCR = f (SEX, AGE, PARTY) are shown in Tables 6 and 7 (Appendix). The test for overall fit revealed that the model Chi-Square and p values indicate that the model is fitting the data significantly better than the No model. As indicated in

Table 7, all variables have a statistically significant predictive power. The positive coefficient for Party (0.273) indicates that moving from Republicans to Democrats is likely to move the performance SCR closer to Pass. This is since Democrat was coded as 1 (Republican: 0). Also, since Female was coded as 1 (Male: 0), the negative coefficient (-0.686) for SEX indicates that the performance SCR would likely move closer to Fail with Females. In addition, the negative coefficient for Age (-0.572) indicates that an increase in age from Under50 to 50Plus is likely to move the performance SCR closer to Fail. Finally, for both models, PRV and SCR, a series of Chi-square tests showed that that there are statistically significant relationships between individuals' performance on privacy and security questions and sex, age, and party affiliation.

The second phase of the analysis used two separate Multiple Linear regressions through SPSS. In this case, the dependent variables, representing actual scores on privacy (PSR) and security (SSR) questions, are continuous with possible values ranging from 0% to 100%. The independent variables sex, age, and party affiliation are categorical. The two models are:

- PSR = f (SEX, AGE, PARTY)
- SSR = f (SEX, AGE, PARTY)

The SPSS outputs for the model PSR = f (SEX, AGE, PARTY) are shown in Tables 8 and 9 (Appendix). In Table 9, looking at the p-value of the t-test for each predictor, we can say that each of the independent variables contributes to the model. The positive coefficient for Party (0.051) indicates that moving from Republicans to Democrats is likely to increase the score PSR. This is since Democrat was coded as 1 (Republican: 0). Also, since Female was coded as 1 (Male: 0), the negative coefficient (-0.137) for SEX indicates that the score PSR would likely decrease with Females. In addition, the negative coefficient for Age (-0.132) indicates that an increase in age from Under50 to 50Plus is likely to decrease the score PSR.

The SPSS outputs for the model SSR = f (SEX, AGE, PARTY) are shown in Tables 10 and 11 (Appendix). In Table 11, looking at the p-value of the t-test for each predictor, we can say that each of the independent variables contributes to the model. The positive coefficient for Party (0.035) indicates that moving from Republicans to Democrats is likely to increase the score SSR. This is since Democrat was coded as 1 (Republican: 0). Also, since Female was coded as 1 (Male: 0), the negative coefficient (-0.103) for SEX indicates that the score SSR would likely decrease with Females. In addition, the negative coefficient for Age (-0.094) indicates that an increase in age from Under50 to 50Plus is likely to decrease the score SSR.

CONCLUSION

This study reported on the analysis of data from a survey of 3951 individuals in the USA for a better understanding of the relationships between Americans' performances on privacy and security questions and sex, age, and party affiliation. The results revealed that the increase in age is more likely to cause a decrease in the level of knowledge about privacy and security, and that females are more likely to know less about these topics. In addition, the level of knowledge on both topics would increase with Democrats. A series of Chi-square tests showed that that there are statistically significant relationships between individual performance on privacy and security questions and sex, age, and party affiliation. In a future study, the authors plan to include other demographic variables such as ethnicity, education, and income level.

REFERENCES

Aaron, J. (2021). The Worst data breaches of 2021 (Q1 and Q2). July 15, https://auth0.com/blog/the-worst-data-breaches-of-2021-q1-and-q2/

Anant, V., Donchak, L., Kaplan, J., & Soller, H. (2020). The consumer-data opportunity and the privacy imperative. McKinsey & Company, April 27. *https://www.mckinsey.com/business-functions/risk/our-insights/the-consumer-data-opportunity-and-the-privacy-imperative*

Burt, A. (2019). Privacy and cybersecurity are converging. Here is why that matters for people and for companies. Harvard Business Review, January 3. https://hbr.org/2019/01/privacy-and-cybersecurity-are-converging-heres-why-that-matters-for-people-and-for-companies

Gothman, A. (2021). Data privacy day: How much do consumers really know about data privacy? January, 28, https://securityboulevard.com/2021/01/data-privacy-day-how-much-do-consumers-really-know-about-data-privacy/

Malik, A., Hiekkanen, K., Nieminen, M. (2016). Privacy and trust in Facebook photo sharing: Age and gender differences. Program, Vol. 50, No. 4, pp. 462-480.

Miller, M. (2020). FBI sees spike in cyber crime reports during coronavirus pandemic. April 16, https://thehill.com/policy/cybersecurity/493198-fbi-sees-spike-in-cyber-crime-reports-during-coronavirus-pandemic

Sonnemaker, T. (2020). Google blocked an average of 18 million daily malicious coronavirus messages to Gmail users in the past week as hackers try to capitalize on fear and less secure remote-work setups. April 16, https://www.businessinsider.com/google-says-gmail-blocked-18-million-coronavirus-phishing-emails-2020-4

Tifferet, S. (2019). Gender differences in privacy tendencies on social network sites: A meta-analysis. Computers in Human Behaviour, Vol. 93, pp. 1-12.

Vogels, E., & Anderson, M. (2019). Americans and digital knowledge. Pew Research Center, October 9.

APPENDIX

Table 1. Sample data: respondents' demographics, N = 3,951

Sex		Age		Party	
Female:	55.2%	Under50:	47%	Democrat:	55.6%
Male:	44.8%	50Plus:	53%	Republican:	44.4%

Table 2. Sample data: respondents' performance (score) on privacy and security questions, N = 3,951

Performance (Score)	Privacy (PSR)	Security (SSR)
0%	23.7%	19.7%
33%	28.1%	41.7%
67%	32.4%	27.6%
100%	15.8%	11.0%

Table 3. Sample data: respondents' performance on privacy and security questions, N = 3,951

Performance	Privacy (PRV)	Security (SCR)
Fail	51.8%	61.4%
Pass	48.2%	38.6%

Table 4. Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	220.069	3	.000
	Block	220.069	3	.000
	Model	220.069	3	.000

Table 5. Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	SEX	-.721	.067	116.296	1	.000	.486
	AGE	-.644	.066	93.843	1	.000	.525
	PARTY	.321	.067	23.109	1	.000	1.378
	Constant	.486	.073	44.828	1	.000	1.626

a. Variable(s) entered on step 1: SEX, AGE, PARTY.

Table 6. Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	177.580	3	.000
	Block	177.580	3	.000
	Model	177.580	3	.000

Table 7. Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	SEX	-.686	.068	102.036	1	.000	.504
	AGE	-.572	.068	71.292	1	.000	.565
	PARTY	.273	.068	15.936	1	.000	1.313
	Constant	.043	.073	.351	1	.554	1.044

a. Variable(s) entered on step 1: SEX, AGE, PARTY.

Table 8. ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	35.339	3	11.780	111.626	.000[b]
	Residual	416.519	3947	.106		
	Total	451.858	3950			

a. Dependent Variable: PSR

b. Predictors: (Constant), PARTY, SEX, AGE

Table 9. Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	.585	.011		51.134	.000
	AGE	-.132	.010	-.195	-12.649	.000
	SEX	-.137	.010	-.202	-13.103	.000
	PARTY	.051	.011	.076	4.891	.000

a. Dependent Variable: PSR

Table 10. ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	18.708	3	6.236	71.662	.000[b]
	Residual	343.468	3947	.087		
	Total	362.176	3950			

a. Dependent Variable: SSR

b. Predictors: (Constant), PARTY, SEX, AGE

Table 11. Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	.520	.010		50.050	.000
	AGE	-.094	.009	-.155	-9.921	.000
	SEX	-.103	.010	-.169	-10.811	.000
	PARTY	.035	.010	.058	3.694	.000

a. Dependent Variable: SSR