Submission

Jake Chanenson and Adriana Knight

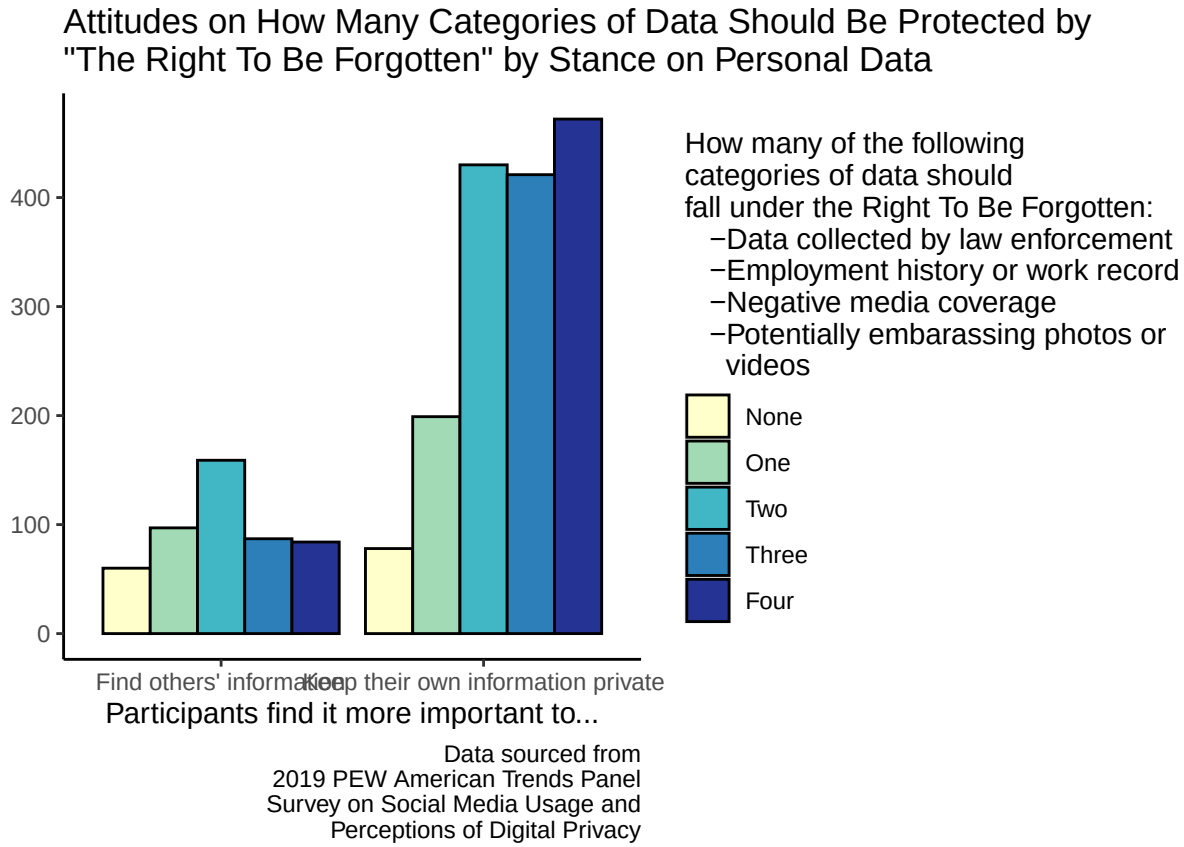
5/7/2021

https://docs.google.com/document/d/1Omlr2cpmMBdgSE8haPFzKQSpPuO8OWO681qwiMvXYfk/edit

Introduction

Our physical and digital lives have been increasingly entwined during the first two decades of the twenty first century. A direct consequence of this entailment is the advent of what social psychologist Shoshana Zuboff calls "Surveillance Capitalism" – an economic model where a user's personal data is commodified for profit generation. The commodification of personal data comes at the cost of user privacy as companies hoover up any and all quantifiable metrics about a given user. As such, it is now more important than ever to understand how the American people feel about issues surrounding the use of their data and their privacy. We choose two aspects of digital privacy to focus on: public sentiment surrounding right to be forgotten (RTBF) – a privilege for people to remove photos, documents, and videos of themselves from public internet search – and the privacy threat of Facial Recognition Technologies.

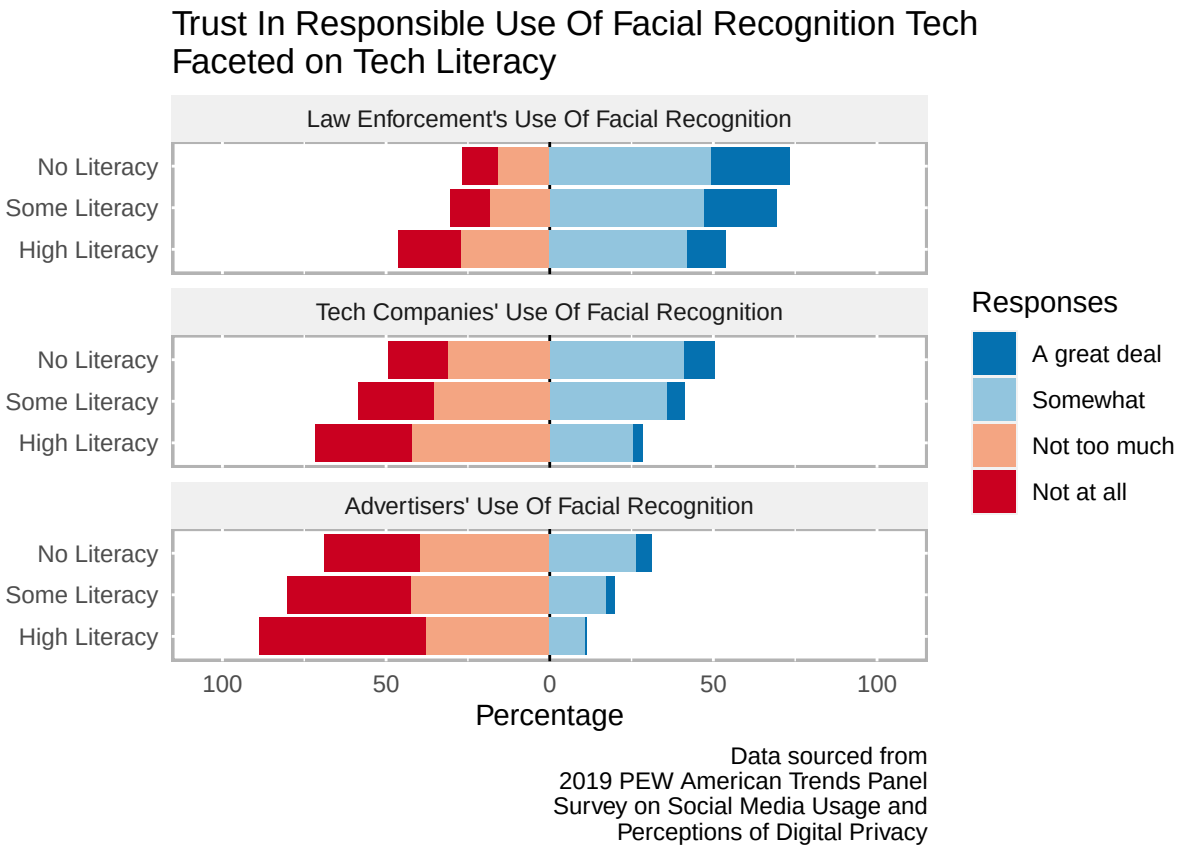
Key Figure 1



Explanation of Key Figure 1

BLAH BLAH BLAH BLAH..... BLAH BLAH BLAH BLAH BLAH

Key Figure 2



Explanation of Key Figure 2

BLAH BLAH BLAH BLAH..... BLAH BLAH BLAH BLAH BLAH

TODO: add a LATEX cmd for pagebreak here (pagebreak) ## Appdendix

Description of Dataset

The dataset we used is the 2019 PEW Survey on American Trends on Technology. The full citation for this data set is as follows:

Pew Research Center. (2019). American Trends Panel (W49). Retrieved from https://www.dropbox.com/sh/adyrtaju2jd7a2d/AAC2fmHoYs2SwVYKqCIkTxOsa?dl=0

The data set is a sample of 4272 respondents of people 18 years or older living in the US. This sample consisted of both English and Spanish-language survey takers. The methodology also details the stratified sampling. I have copied the statement from the methodology section below:

> The ATP subsample was selected by grouping panelists into seven stratum 1. Non-internet panelists. There were 691 total panelists in this stratum and they are sampled at a rate of 100% 2. HS or less panelists. There were 2,027 total panelists in this stratum and they are sampled at a rate of 98.9%. 2,005 panelists were selected for Wave 49. 3. Hispanic, Unregistered or Non-volunteers. There were 5,312 total panelists in this stratum and they are sampled at a rate of 44.8%. 2,380 panelists were selected for Wave 49. 4. Black or 18-34 panelists. There were 1,253 total panelists and they are sampled at a rate of 18.2%. 228 panelists were selected for Wave 49. 5. Other panelists. There were 4,176 total panelists and they are sampled at a rate of 13.5%. 564 panelists were selected for Wave 49.

Variables Used and Their Questions Phrasing Key Figure 1

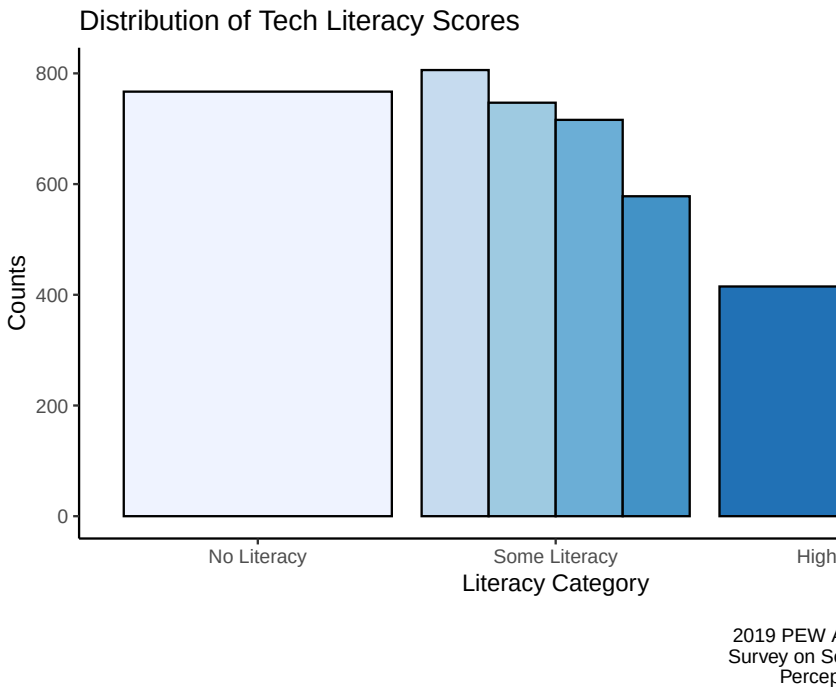
- PUBLICDATA_W49 - "Today a wide range of information about people is searchable online. Do you think it is more important for people to have the ability to. . . "
- RTBF - "Do you think that ALL Americans should have the right to have the following information about themselves removed from public online search results?"
  - RTBFa_W49 - "Data collected by law enforcement, such as criminal records or mugshots"
  - RTBFb_W49 - "Information about their employment history or work record"
  - RTBFc_W49 - "Negative media coverage"
  - RTBFd_W49 - "Potentially embarrassing photos or videos"

Key Figure 2

- FACE1 - "How much have you heard or read about the development of automated facial recognition technology that can identify someone based on a picture or video that includes their face?
- FACE2 - "Based on what you know, how effective do you think facial recognition technology is at the following things?"
  - FACE2a_W49 - "Accurately identifying individual people"
  - FACE2b_W49 - "Accurately assessing someone's gender"
  - FACE2c_W49 - "Accurately assessing someone's race"
- FACE3 - "How much, if at all, do you trust the following groups to use facial recognition technology responsibly?"
  - FACE3a_W49 - "Advertisers"
  - FACE3b_W49 - "Technology companies"
  - FACE3c_W49 - "Law enforcement agencies"
- FACE4 - "In your opinion, is it acceptable or unacceptable to use facial recognition technology in the following situations?"
  - FACE4a_W49 - "Law enforcement agencies assessing potential security threats in public spaces"
  - FACE4b_W49 - "Companies automatically tracking the attendance of their employees"
  - FACE4c_W49 - "Advertisers seeing how people respond to public advertising displays"
  - FACE4d_W49 - "Apartment building landlords tracking who enters or leaves their buildings"
- KNOW1_W49 - "If a website uses cookies, it means that the site. . .
- KNOW3_W49 - "When a website has a privacy policy, it means that the site. . . "
- KNOW4_W49 - "What does it mean when a website has 'https://' at the beginning of its URL, as opposed to 'http://' without the 's'?"
- KNOW7_W49 - "The term 'net neutrality' describes the principle that. . . "
- KNOW8_W49 - "Many web browsers offer a feature known as 'private browsing' or 'incognito mode.' If someone opens a webpage on their computer at work using incognito mode, which of the following groups will NOT be able to see their online activities?"
- KNOW9_W49 - "Some websites and online services use a security process known as two-step or two-factor authentication. Which of the following images is an example of two-factor authentication?"
- CONCERNCO_W49 - "How concerned are you, if at all, about how companies are using the data they collect about you?"

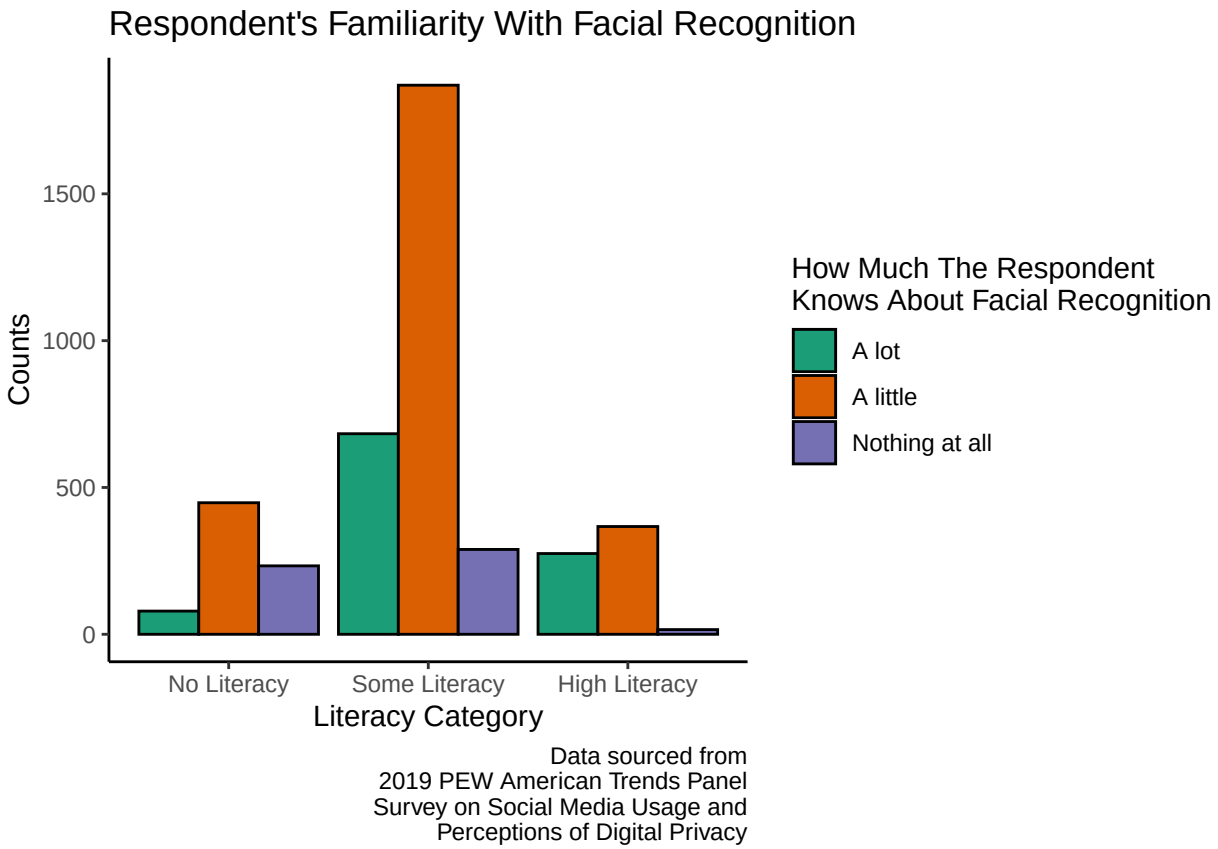
Additional Figures For Key Figure 1

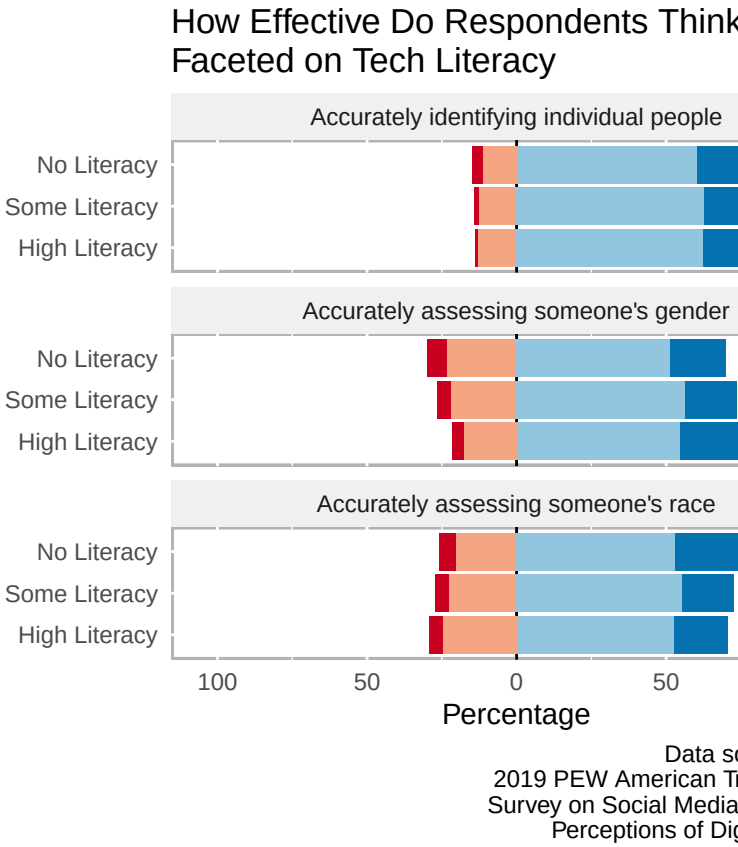
Additional Figures For Key Figure 2



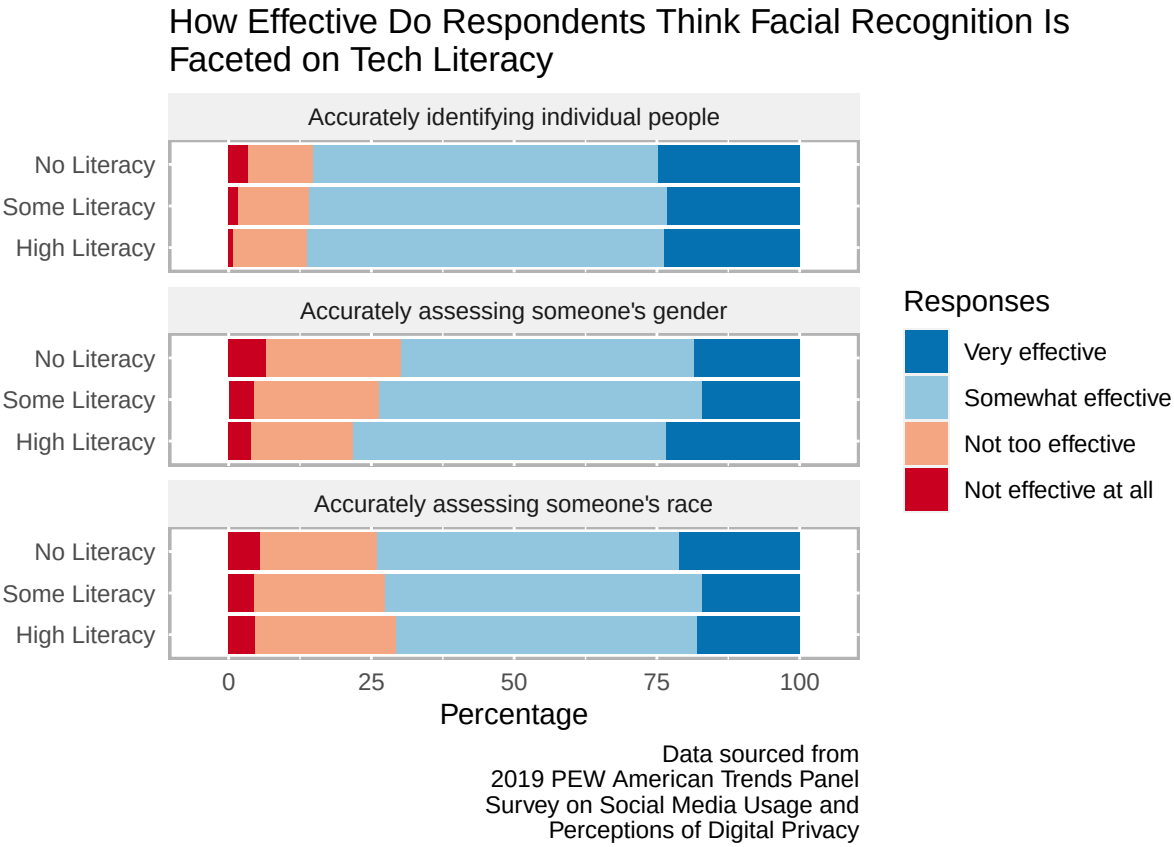
1. Underlying Distribution of Tech Literacy

2. How much have you heard or read about the development of automated facial recognition technology that can identify someone based on a picture or video that includes their face

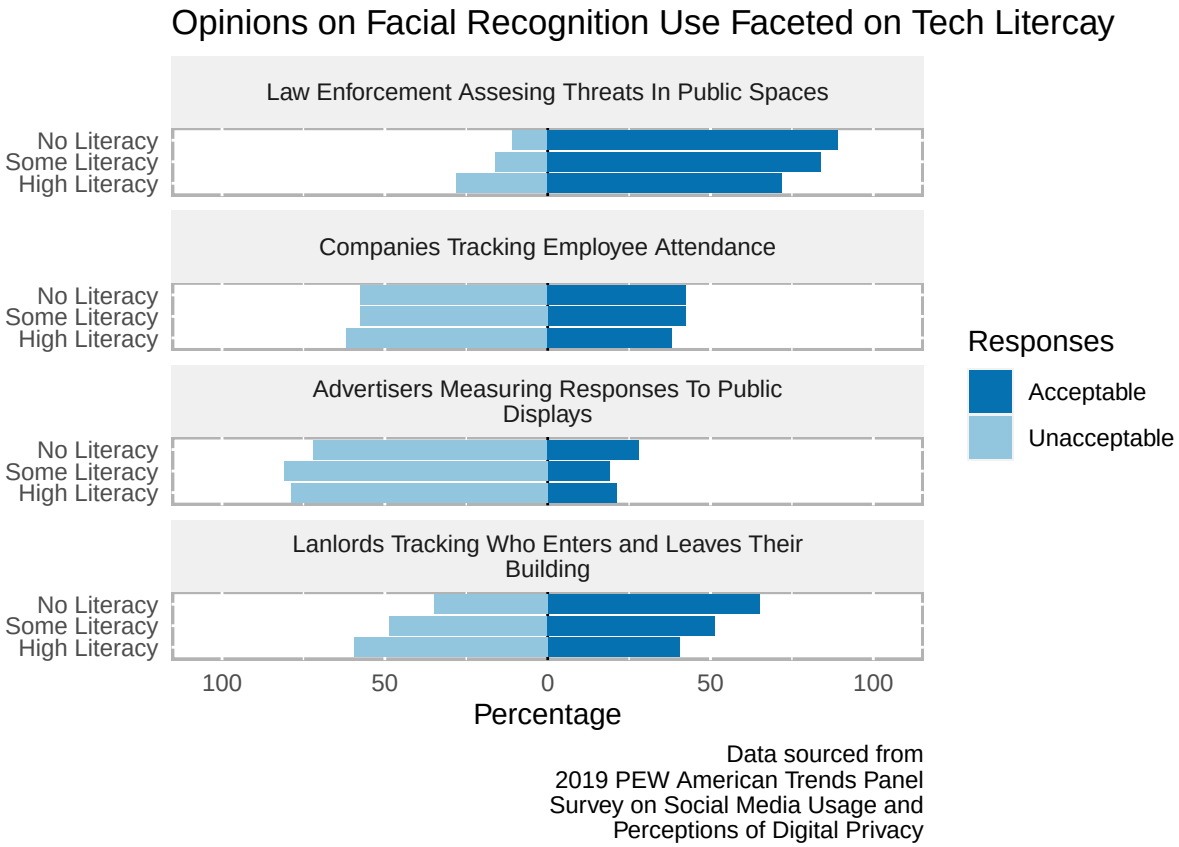

Respondent's Familiarity With Facial Recognition
Counts
1500
1000
500
0
No Literacy
Some Literacy
High Literacy
Literacy Category
How Much The Respondent Knows About Facial Recognition
A lot
A little
Nothing at all
Data sourced from 2019 PEW American Trends Panel Survey on Social Media Usage and Perceptions of Digital Privacy

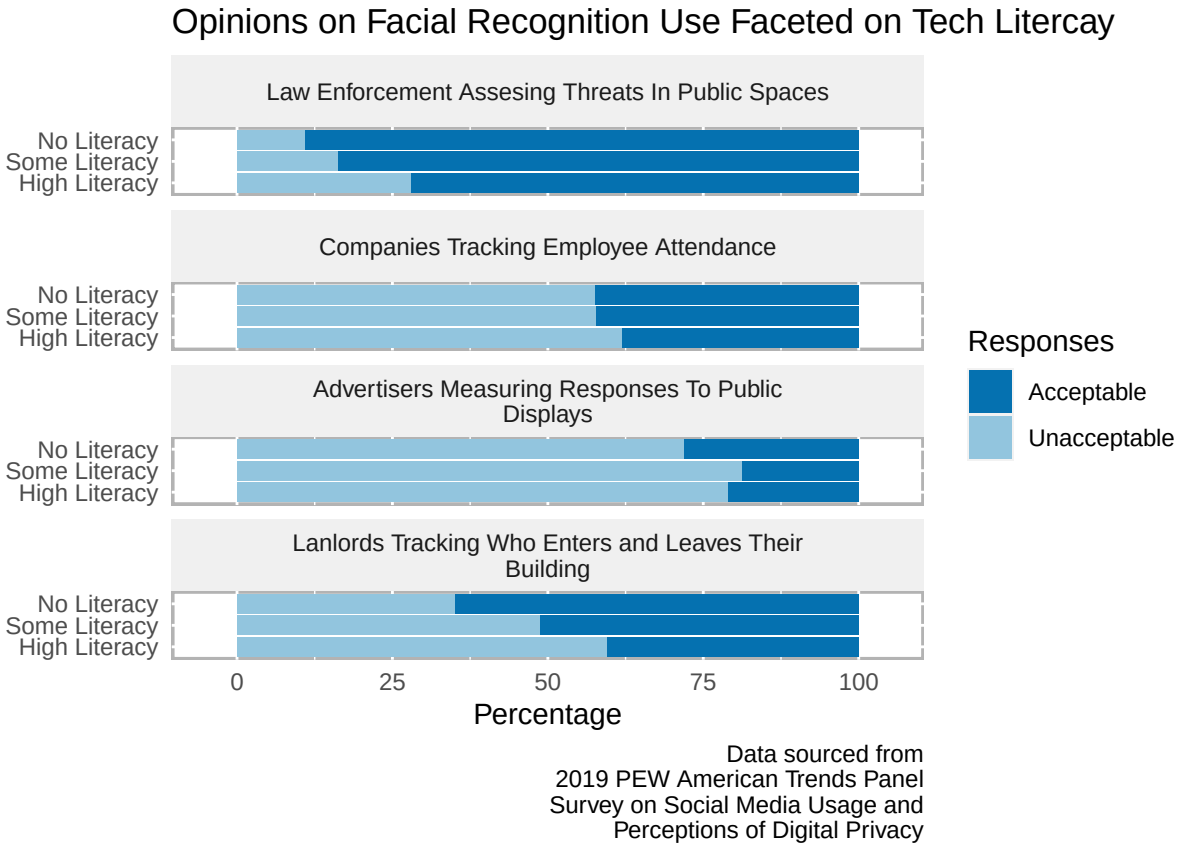


3. People's Understanding of Facial Recognition



4. What situations People Think It Is Acceptable To Use Facial Recognition In the Following Situations?





Additional Tables For Key Figure 1

Table 1: Summary Statistics For Key Figure 1

	Public Data Should Be Private	Public Data Should Be Public
0	89	60
1	213	97
2	442	159
3	430	87
4	479	84

Table 2: Numeric Summary That Involves Uncertainty For Key Figure 1

publicDataNum	mean	sd
0	2.603146	1.183437
1	2.078029	1.248072

Additional Tables For Key Figure 2

Table 3: Summary Statistics For Key Figure 2

Group	Item	low	high	mean
High Literacy	Law Enforcement's Use Of Facial Recognition	46.32238	53.67762	2.461659
High Literacy	Tech Companies' Use Of Facial Recognition	71.51800	28.48200	2.018779
High Literacy	Advertisers' Use Of Facial Recognition	88.73239	11.26761	1.607199
Some Literacy	Law Enforcement's Use Of Facial Recognition	30.48829	69.51171	2.797539
Some Literacy	Tech Companies' Use Of Facial Recognition	58.59468	41.40532	2.237396
Some Literacy	Advertisers' Use Of Facial Recognition	80.26995	19.73005	1.843192
No Literacy	Law Enforcement's Use Of Facial Recognition	26.79612	73.20388	2.862136
No Literacy	Tech Companies' Use Of Facial Recognition	49.51456	50.48544	2.415534
No Literacy	Advertisers' Use Of Facial Recognition	68.73786	31.26214	2.069903

Table 4: Numeric Summary That Involves Uncertainty For Key Figure 2

Group	Item	mean	sd
High Literacy	Law Enforcement's Use Of Facial Recognition	2.461659	0.9325284
High Literacy	Tech Companies' Use Of Facial Recognition	2.018779	0.8149991
High Literacy	Advertisers' Use Of Facial Recognition	1.607199	0.6952671
Some Literacy	Law Enforcement's Use Of Facial Recognition	2.797539	0.9198941
Some Literacy	Tech Companies' Use Of Facial Recognition	2.237396	0.8672941
Some Literacy	Advertisers' Use Of Facial Recognition	1.843192	0.7908986
No Literacy	Law Enforcement's Use Of Facial Recognition	2.862136	0.9030985
No Literacy	Tech Companies' Use Of Facial Recognition	2.415534	0.8921139
No Literacy	Advertisers' Use Of Facial Recognition	2.069903	0.8637574

Discussion of Uncertainty & Inference

Key Figure 1

Key Figure 2

Discussion of Analytic Choices

Key Figure 1

Key Figure 2

Discussion Of Other Ways We Could Have Made Our Key Figure

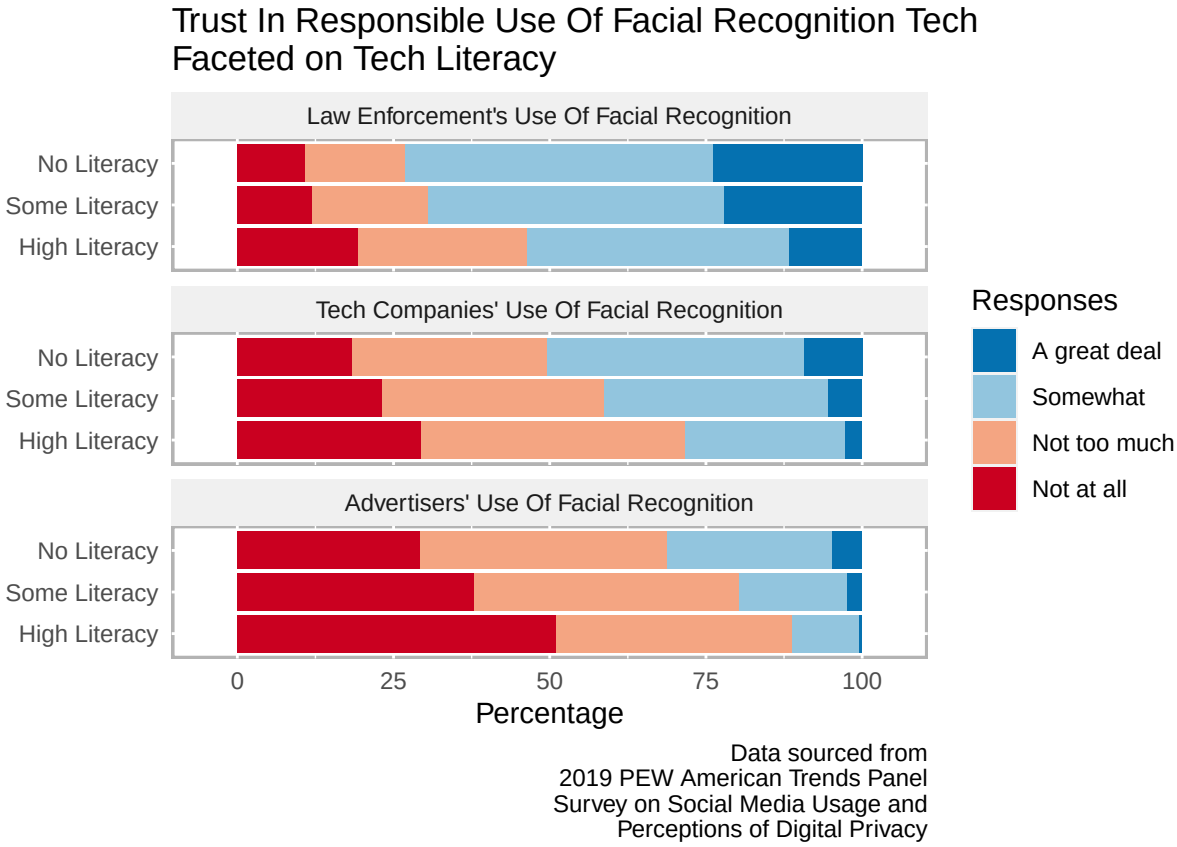
Key Figure 1

BLAH BLAH BLAH

Key Figure 2

When visualizing my key figure, I knew that I wanted to visualize some form of stacked bar chart to communicate the relative differences in comfort of various groups using facial recognition technology. Professor Ella Foster-Molina sent me an excellent article from Datawrapper making the case against diverging stacked bar charts and offering 100% bar charts as the better alternative. The article argues that "the main problem with diverging bars, however, is comparability." On the whole, I agree; to me, the most compelling point of the piece relates to how neutrals are handled in the diverging stacked bar charts. Putting neutrals in the middle means that (1) none of the bars have a common baseline and (2) the chart is implying that part of the

neutrals should be coded as positive and negative respectively. That all being said, none of my charts have neutrals. Thus, I think the diverging bar chart looks just fine because the overall positive-negative trend both within and between facets is more pronounced. Below is what the figure would look like with a 100% bar chart.



NOTE: NEED TO CREATE AND INCLUDE ALT. VERSION OF KEY FIGURE IN THIS SECTION