

Salty.

in collaboration with



Censorship of Marginalized Communities on Instagram

**A community report by the Salty Algorithmic Bias Collective:
Shakira Smith, Oliver L Haimson, Claire Fitzsimmons, and Nikki Echarte Brown**

**Publish Date:
September 27, 2021**

This work was supported in part by the
National Science Foundation grant #1942125.

TABLE OF CONTENTS

1: Executive Summary and Main Findings	Pg. 3
2: Background	Pg. 5
3: Findings	
• Part I: Overall Experience of Censorship	Pg. 7
• Part II: Narrative Highlights from the Margins	Pg. 9
• Part III: Content Removal and Reasons for Removal by Demographic	Pg. 16
4: Ongoing Efforts and Support	Pg. 24
5: Further Reading on Social Media Censorship and Marginalized Groups	Pg. 25
6: About Salty	Pg. 26
7: Supplementary Materials Available here.	
• Appendix 1: Methods	
○ Survey Summary	
○ Survey Details	
○ Data Collection	
○ Survey Respondents	
○ Preparation and Coding	
○ Analysis	
• Appendix 2: Data Tables from Findings Part III	
• Appendix 3: Open-Source Numerical Data	

EXECUTIVE SUMMARY AND MAIN FINDINGS

The following report was drafted by the Salty Algorithmic Bias Collective.

We found substantial evidence of censorship on Instagram for marginalized groups including people who are transgender and/or nonbinary, LGBTQIA+¹, BIPOC², disabled, sex workers, and/or sex educators. Instagram disproportionately censored people in each of these marginalized groups by removing their content more frequently as compared with people who did not hold these marginalized identities.

Main findings:

- A majority of participants, (61.9%) experienced content removals on Instagram.
- Almost all marginalized groups experienced censorship on Instagram at higher percentages than those from more privileged groups.
- For most marginalized groups the most prominent reason for content removals was Nudity, etc.
- Compared to other marginalized respondents, Sex Workers are more likely to report being censored in general.
- Disabled people are more likely to be report being censored for Self-injury/Self-harm or Violence, etc compared to other marginalized respondents.
- Although nudity was highly reported by all respondents as a reason for censorship, Trans/ Nonbinary respondents experienced the highest rates of removal based on Nudity, etc.
- Compared to other marginalized respondents, BIPOC were most likely to report being censored for Fake news/False information.
- Plus sized respondents report the highest percentage of “Community Guidelines” violations among marginalized respondents.
- Even though most participants who experienced content removals appealed these decisions, over 90% of those who appealed either received no response or their content was not reinstated.
- Narrative quotes from participants highlight the frustration, anger, and disappointment people felt when their content was removed, and the personal and sometimes financial consequences marginalized people face when content removals restrict their ability to express themselves online.
- Our results highlight the problematic and discriminatory ways marginalized people’s bodies and identities are sexualized and policed on social media.
- Taken together, our results show how disproportionate content removals cause marginalized groups to face substantial challenges and consequences when attempting to use online spaces like Instagram.

¹ LGBTQIA+ stands for lesbian, gay, bisexual, queer, intersex, asexual, and additional sexual minority identities. We intentionally separated this category (which relates to sexual orientation) from our trans/nonbinary category (which relates to gender) to highlight the unique disparities each of these marginalized groups face.

² Black, Indigenous, and people of color

EXECUTIVE SUMMARY AND MAIN FINDINGS

Our survey found that almost every marginalized group was censored at higher rates than the privileged comparison group.

Table 1 summarizes the proportion of each marginalized group vs. the relevant privileged group that experienced censorship, as well as the difference in those proportions.

Listed from highest rate of difference to lowest, respondents who identified as sex workers, sex educators, LGBTQIA+, Trans and/or Nonbinary, BIPOC, and disabled experienced censorship on Instagram at greater rates than their privileged counterparts.

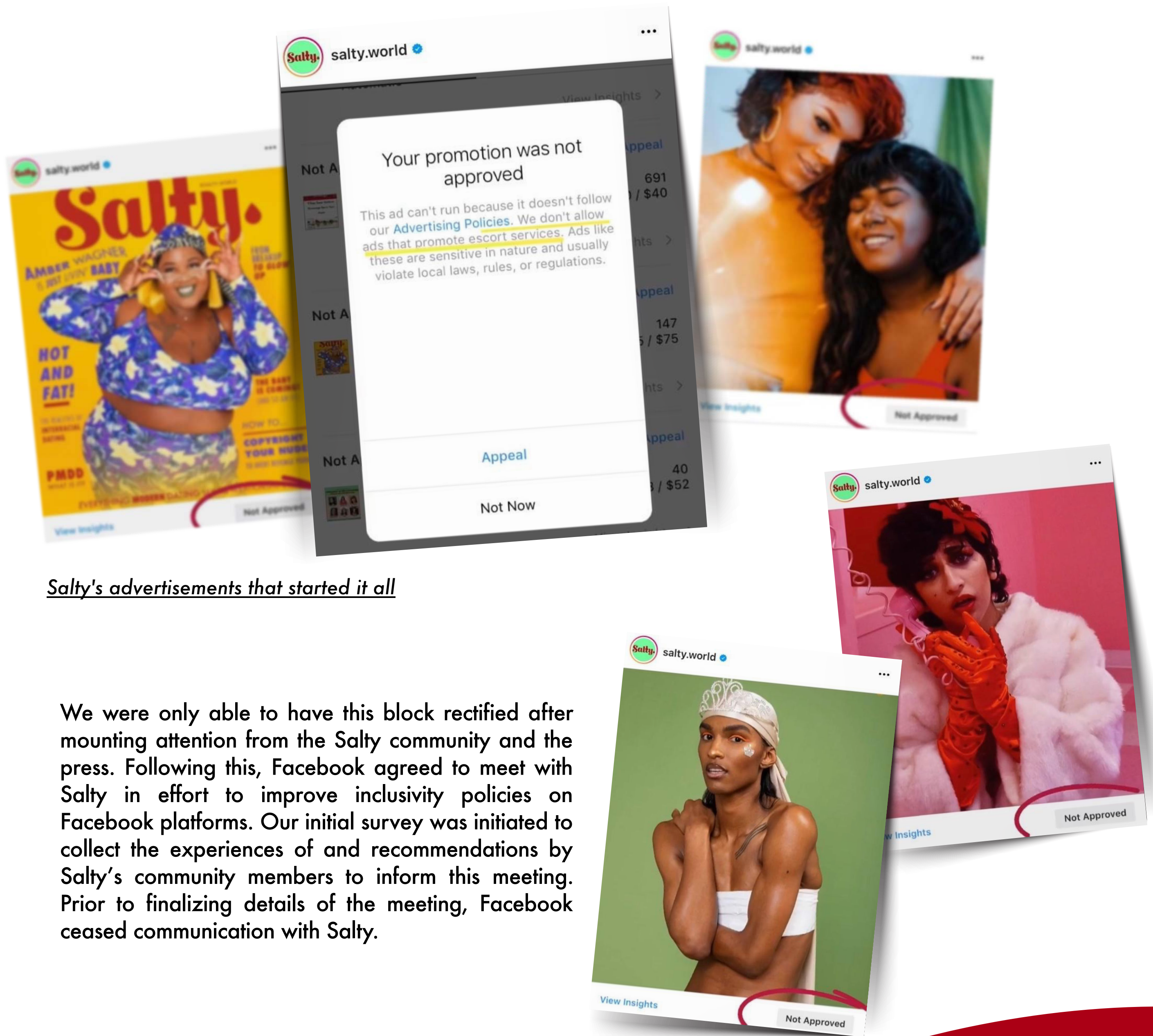
Table 1. Percentage Differences of Content Removal of Marginalized Groups Compared to Privileged Groups

Marginalized Group	Percentage who Experienced Content Removals	Privileged Group	Percentage who Experienced Content Removals	Percentage Difference
Sex Workers	83.9%	Non-Sex Workers	54.2%	29.7%
Sex Educators	77.8%	Non-Sex Educators	58.9%	18.9%
LGBTQIA+	65.7%	Straight	48.8%	16.9%
Trans/Nonbinary	69.7%	Cisgender	58.0%	11.7%
BIPOC	67.1%	White	60.3%	6.8%
Disabled	64.7%	Non-disabled	59.7%	5.0%
Plus-sized	59.3%	Non-plus-sized	63.1%	-3.8%*

**This research is not meant to discount the experience of plus sized people; our finding could be due to a small sample size, and more research needs to be carried out in order to adequately understand how plus sized folks experience censorship online.*

BACKGROUND

Our investigations into social media censorship were instigated by Salty's own experience with censorship on Instagram. Starting around July 2019, Salty's attempts to advertise on Instagram were inappropriately blocked and marked as attempts for sexual solicitation by the platform.



Salty's advertisements that started it all

We were only able to have this block rectified after mounting attention from the Salty community and the press. Following this, Facebook agreed to meet with Salty in effort to improve inclusivity policies on Facebook platforms. Our initial survey was initiated to collect the experiences of and recommendations by Salty's community members to inform this meeting. Prior to finalizing details of the meeting, Facebook ceased communication with Salty.

However, we decided to share the information we had compiled and released it in our first report, *An Investigation into Algorithmic Bias in Content Policing of Marginalized Communities on Instagram and Facebook*, on October 22nd, 2019. Our first report made substantial impact in informing those both within and outside our community about social media censorship, given that minimal public data about content removal on social media platforms exists elsewhere.

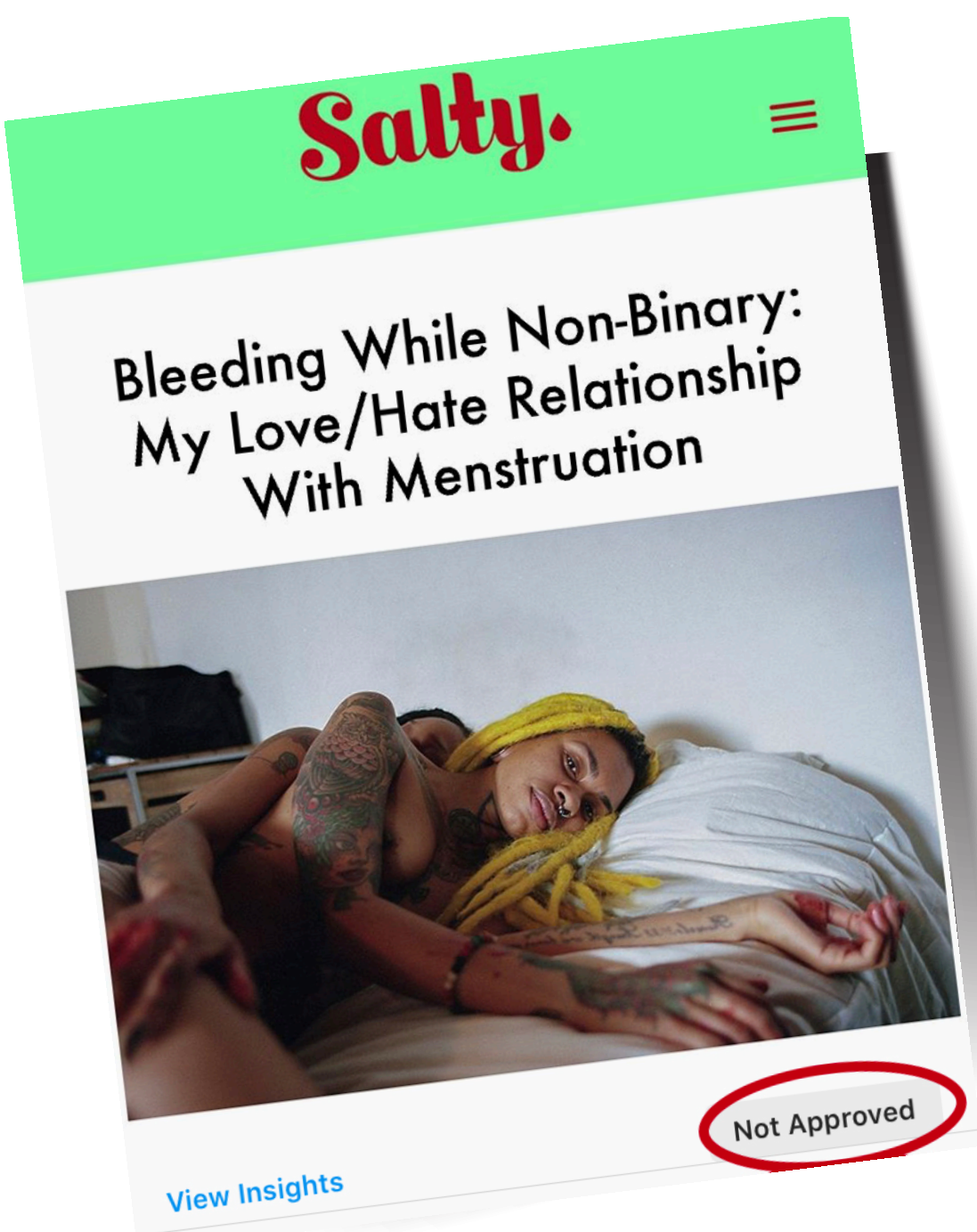


Press coverage from the Daily Beast and Rolling Stone on Salty's ad censorship

The present publication, *Censorship of Marginalized Communities on Instagram*, represents new survey data collected from October 2019 to February 2021. Our findings are detailed within.

We have additionally included an open-source Appendix that details the methods and the majority of the numerical data from our survey (in aggregate to protect participants' privacy). We encourage other researchers to use, analyze, and cite this data in their own investigations.

We have exciting future research planned as well! Salty's Algorithmic Bias Collective has released a new survey for its forthcoming third report. Please consider adding your voice to this novel and important documentation of experiences of censorship! If you'd like to participate, **click here**. Please see Section 4: Ongoing Efforts and Support for details on the Collective's ongoing projects.



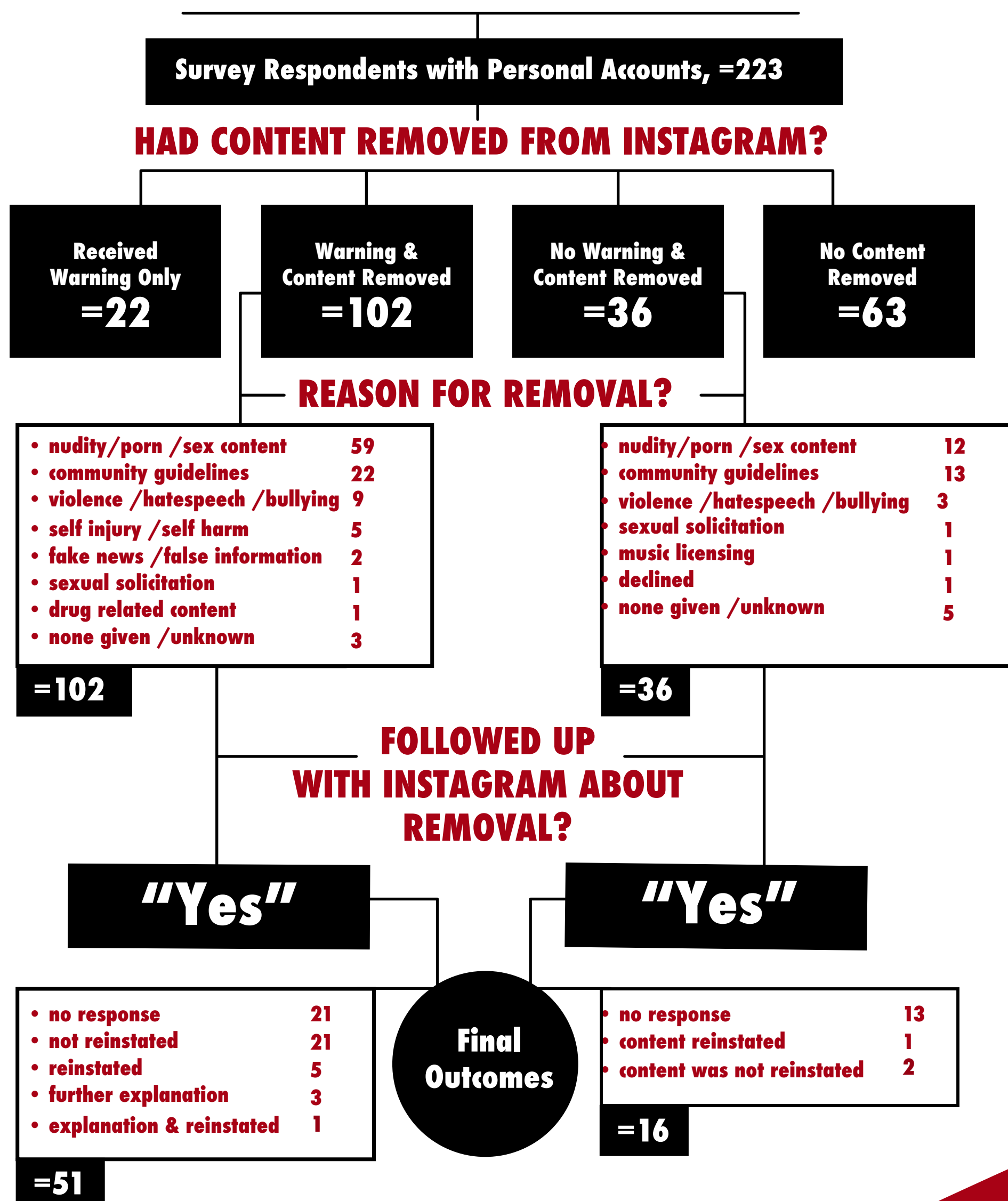
Another example of Salty content which is rejected for promotion on Instagram.

FINDINGS

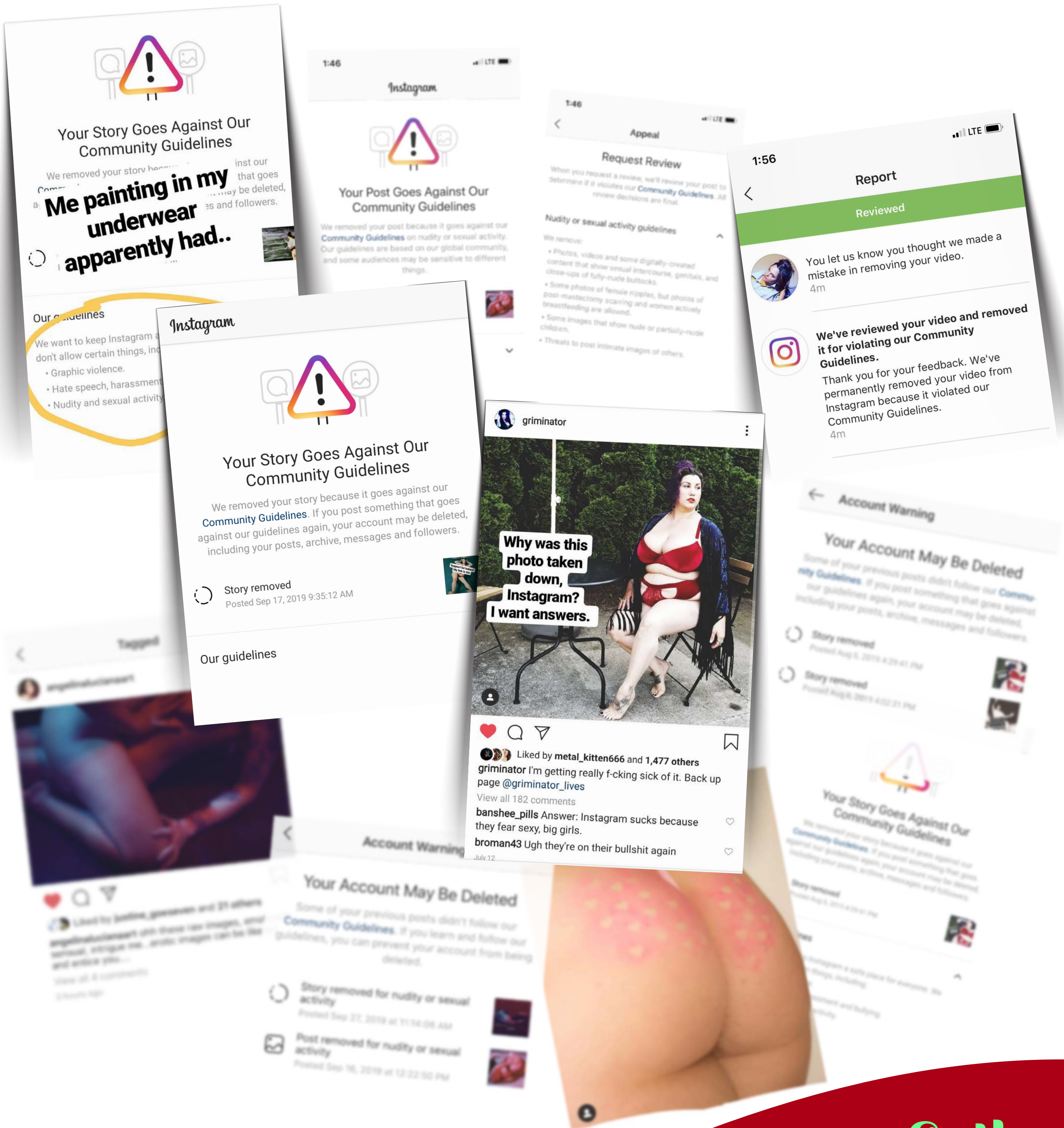
Part I: Overall Experience of Censorship

Chart 1 shows our conceptualization of participants' content removal experiences and the reasons Instagram provided for removing content. In total, 138 participants (61.9%) experienced content removals. The majority of participants who experienced content removals appealed Instagram's content removal decisions (52.9%), but most of them either received no response or their content was not reinstated. Only six participants (8.2% of those who appealed their content removal) had their content reinstated by Instagram.

CHART 1: CONTENT REMOVAL FROM INSTAGRAM



These screenshots were submitted by survey participants documenting their interactions with Instagram and the content that was removed from their pages.



FINDINGS

Part II: Narrative Highlights from the Margins

In this section, we include some of participants' responses to our open-ended question:

How do you feel about your content being taken down?

These quotes highlight the frustration, anger, and disappointment people felt when their content was removed, and the personal and sometimes financial consequences marginalized people face when content removals restrict their ability to express themselves online.

Many participants had multiple marginalized identities (e.g., trans people of color); our categories are not meant to simplify these identities, but instead to provide a loose grouping to contextualize participants' quotes.

TRANS AND/OR NONBINARY.

“ I’m angry. I see brands posting the same type of content as I am that get approved. But because I’m Hispanic and transgender, my content gets blocked? Can the algorithm really not contain all of this beauty? I was simply posting a photo of me, showing no more skin than any white thin cis straight model does on Instagram. But I’m not one of those. Yet. Trans femmes of color are coming for you. We are here, we are beautiful, and we are not going anywhere. ”

“ It was extremely frustrating to feel silenced. Even now I’m concerned about posting too much about human rights or social issues on my story because I don’t know what else would’ve triggered the ban. ”

“ I feel as if I’m being policed for being a femme afab³ trans person. I’ve had pictures of my face taken down. I follow their guidelines. They just want to erase me. ”

³ assigned female at birth

LGBQIA+.

“

It's incredibly disheartening. it was a post for my partner's f*cking birthday of all things. They replaced the post but ever since, I have lost followers, and my stories have gone from over 200 views to maybe 60. I'm trying to fundraise for my college tuition and I use Instagram as a platform to promote my gofundme. Now my engagement with my community is incredibly restricted which has real consequences for me.

”

“

Disappointed that the algorithm favours a certain category of people.

”

“

It makes me feel like my body's not okay to be seen

”

“

Very angry as my posts are not vulgar, if anything I promote body positivity and self love and care. When I see posts of people literally wearing the tiniest bit of clothing, and then for mine when the post that have been removed aren't even anything to be bothered about, it really does irritate me.

”

“

Like I'm being bullied

”

BIPOC.

“ Instagram targeted me constantly for challenging actual hate speech, other forms of oppression, and white supremacy specifically. ”

“ Very discouraged and blue. On Thanksgiving I reposted a series of infographics and stats... about indigenous health and welfare, highlighting in impropriety of celebrating genocide... the next morning, all my stories were gone, no explanation no post. Being censored for being a female is tough... being silenced for sharing real talk on First Nations erasure is heartbreaking. ”

“ When I post something pro black or to do with data surveillance education, my reach with my followers is much less than any other stories... It is anti black. I research all my information & cross reference my sources but feel like I'm being gaslit to think it's false. ”

“ I was incredibly angry. To be clear, my content wasn't taken down, my entire account was deleted. No warning, no explanation. ”

“ Politically surveilled ”

“ Disappointed and upset! It was a picture of a painting from a museum of a topless black woman along with other paintings of black subjects (fully clothed). I captioned the post talking about how amazing it was to see museums... take accountability for their lack of diversity in art and history by showcasing untold stories of black folks. Oh the irony! ”

DISABLED.

“

I felt extremely angry and upset as the picture was just of me in a swim suit, there was no recent self harm visible and I should not be forced to hide my scars.

”

“

I feel like it silenced those of us struggling with mental illness and takes away a community where people come to find like minded individuals who suffer as well.

”

“

I was really concerned. I initially thought someone was posting content on my account without my knowledge until I realized Instagram flagged my account for simply having a disability.

”

PLUS-SIZED.

“

Unfair. It's obviously fatphobia. It effects my income.

”

“

My content has been removed simply for having large, naked tummies visible. No boobs; no genitals. Just fat. If a thin tummy was exposed, it would never be removed. The double standard here is obvious and harmful.

”

“

Really shitty. These policies are supposed to be protecting marginalized groups, not policing them for punching up at white supremacy and the patriarchy.

”

“

It feels as though I'm not meant to be seen. Instagram has a clear bias against fat, disabled, sexual and LGBTQ bodies.

”

“

Frustrating to see neo nazis and white skinny sexually suggestive posts not be targeted as much as mine.

”

SEX WORKER.

“

I feel like my creativity is being stifled, and that my body is being policed. I can't express my creativity on this platform due to unreasonable regulations. I'm being targeted for doing what I do (stripper) and I don't post nearly as much as I would otherwise, if the platform didn't target sex workers. I don't see it as a safe platform to educate about sexuality and sex work, but I do it anyway because people need to learn.

”

“

It makes me feel angry and upset. I know the reason it happens is because I don't have loads of followers and I'm not rich, as I see celebs and verified accounts posting similar content and not getting taken down. At one point my Instagram was growing like 100 followers a month (which was good by my standards), but since the shadow ban I haven't had new followers or interactions and my account feels stagnant - which is annoying because I want to use Instagram to get clients for sex work, meet photographers, and promote the art that I do. Meaning the shadow ban could have lost me money and opportunities.

”

“

Male photographers are artists - but female models are 'whores.' My same photo will appear on MANY other accounts, usually male-run accounts, but removed on mine. My image is used on multiple accounts, but I am not allowed to use my own image.

I became shadowbanned after several impersonation accounts surfaced- pretending to be me, and soliciting other users via DM. When I contacted Instagram about said impersonation accounts, at first they helped take them down, but later said they could not confirm that I was me - despite me sending them the requested photo of my face, next to a paper written with my full (real) name and temporary code I gave.

Then, I was punished and shadowbanned. Then deleted. Yet my images still remain on countless male-run accounts. I do not own my own body, apparently. Men do. Instagram does.

”

SEX EDUCATOR.

“ I censor all the content I post on Instagram, I think it’s wrong for them to delete my posts (and my account, twice!) Without any option for dispute. I know my account is being censored despite following guidelines and it frustrates me to see the obvious inconsistency and bias against creators like me. ”

“ It made me spiral into a really dark place. My confidence was so low after my account was deleted. I felt shamed... I had my account since at least 2009, they said the reason was for nudity and my account got instantly disabled because I reached my three warnings.... over 10 years... ”

“ I adhere to the guidelines but they’re applied in ways that are unpredictable, so sometimes I have no idea my post will be removed. It’s frustrating and annoying. ”



FINDINGS

Part III: Content Removal and Reasons for Removal by Demographic

We asked two questions of our data:

How does one's experience of censorship on Instagram compare for marginalized and privileged groups?

For what reasons are marginalized groups' content censored on Instagram?

In all but one category, we found that those with marginalized identities experienced censorship on Instagram at higher percentages than those from more privileged groups.

Censorship disparities were observed for sex workers, sex educators, LGBTQIA+ people, trans and nonbinary people, BIPOC, and disabled people.

It was also observed that different marginalized identities were more likely to report being censored for different reasons. For example, disabled people are more likely to be report being censored for Self Harm/Self Injury and Violence, etc compared to other marginalized respondents. Compared to other marginalized respondents, BIPOC were most likely to report being censored for Fake news/False information, and although nudity was highly reported by all respondents as a reason for censorship, trans/nonbinary respondents reported the highest rates of removal based on Nudity, etc.

When reviewed in context of the narrative highlights in Part II, it becomes clear that these disproportionate content removals cause marginalized groups substantial challenge to attempting to exist online.

Again, note that many participants had multiple marginalized identities (e.g., trans people of color). Whereas in Part II participants' responses are only included in one grouping, in Part III, participants are represented in each category where they self-identified.

Please note that numerical data is provided in the [Supplemental Materials](#) for our report; we encourage our allied investigators to utilize this numerical data to answer remaining questions they may have of our data.

TRANS AND/OR NONBINARY.

CHART 2: PERCENTAGE OF CONTENT REMOVAL BY TRANS/ NONBINARY STATUS

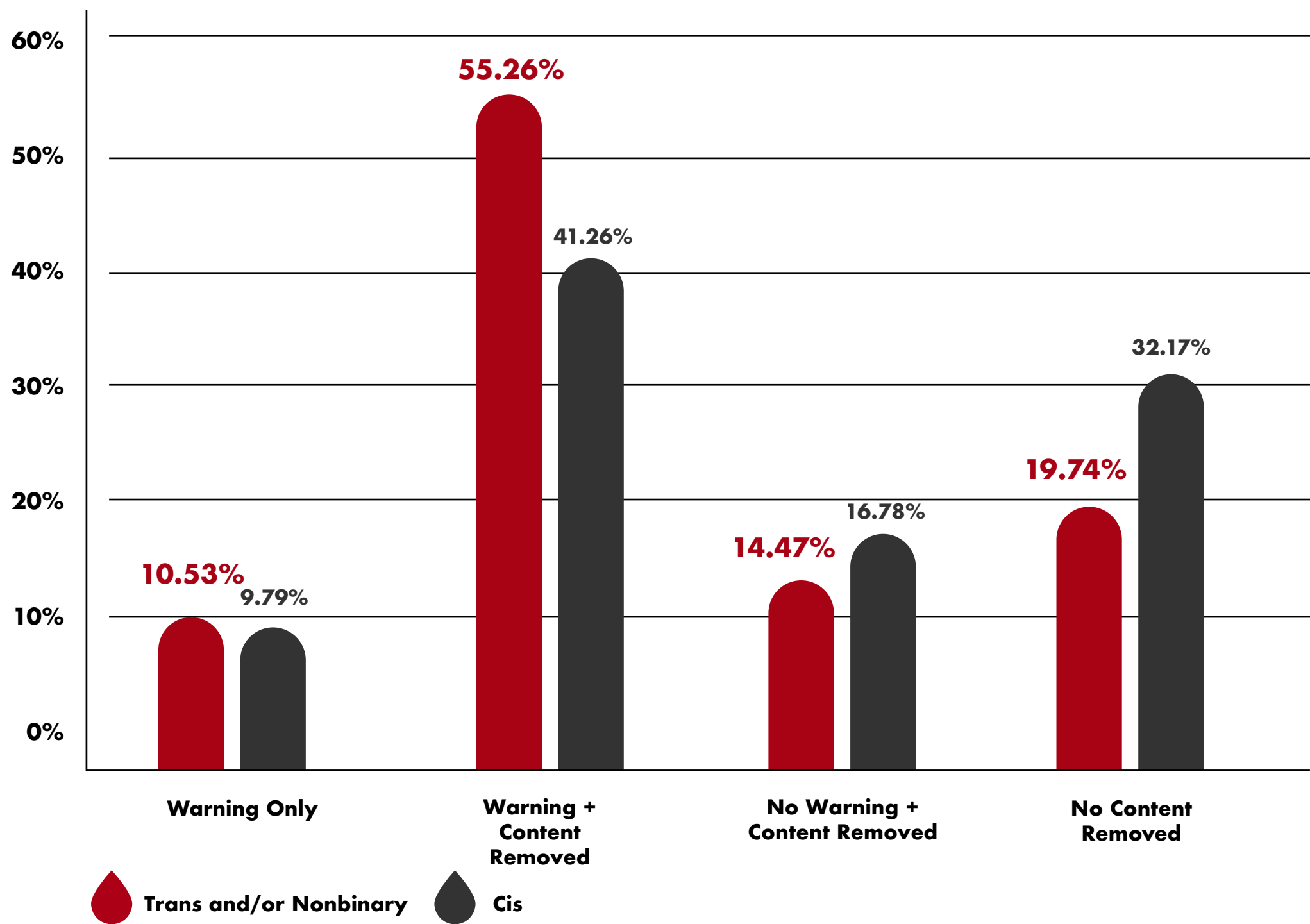


CHART 3: PERCENTAGE OF REASONS FOR REMOVAL: TRANS/ NONBINARY



LGBQIA+.

CHART 4: PERCENTAGE OF CONTENT REMOVAL BY LGBQIA+ STATUS

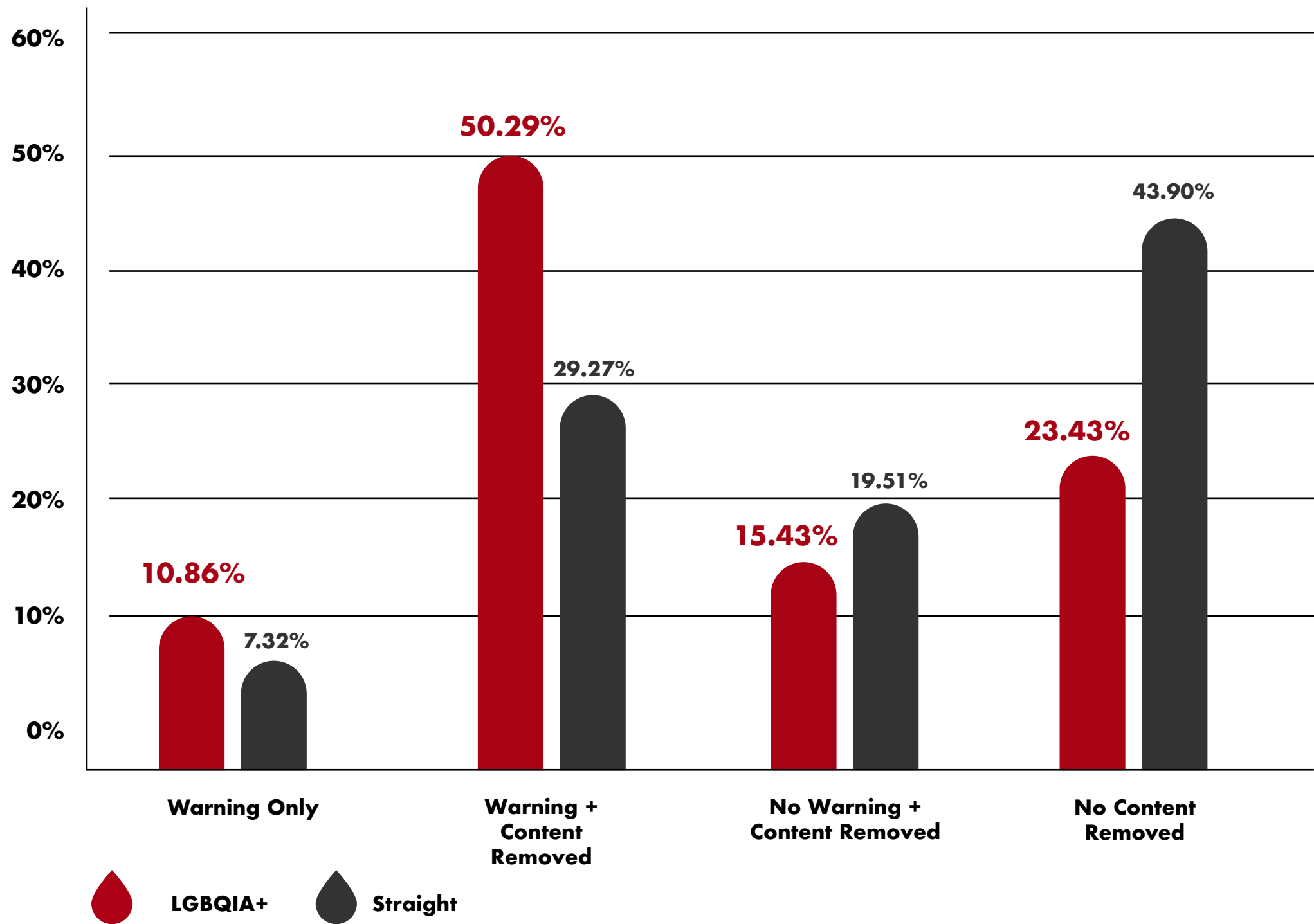
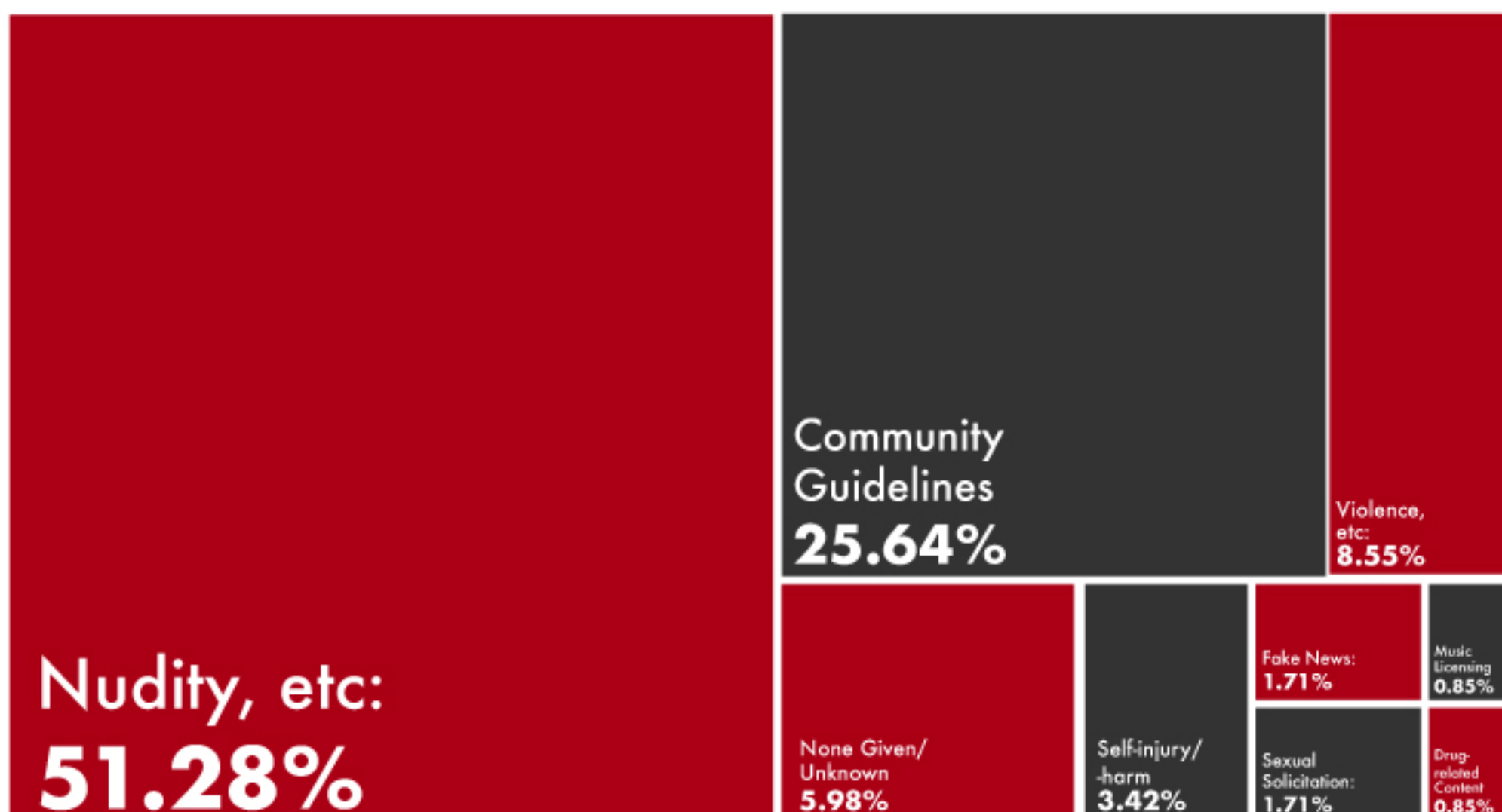


CHART 5: PERCENTAGE OF REASONS FOR REMOVAL: LGBQIA+



BIPOC.

CHART 6: PERCENTAGE OF CONTENT REMOVAL BY BIPOC STATUS

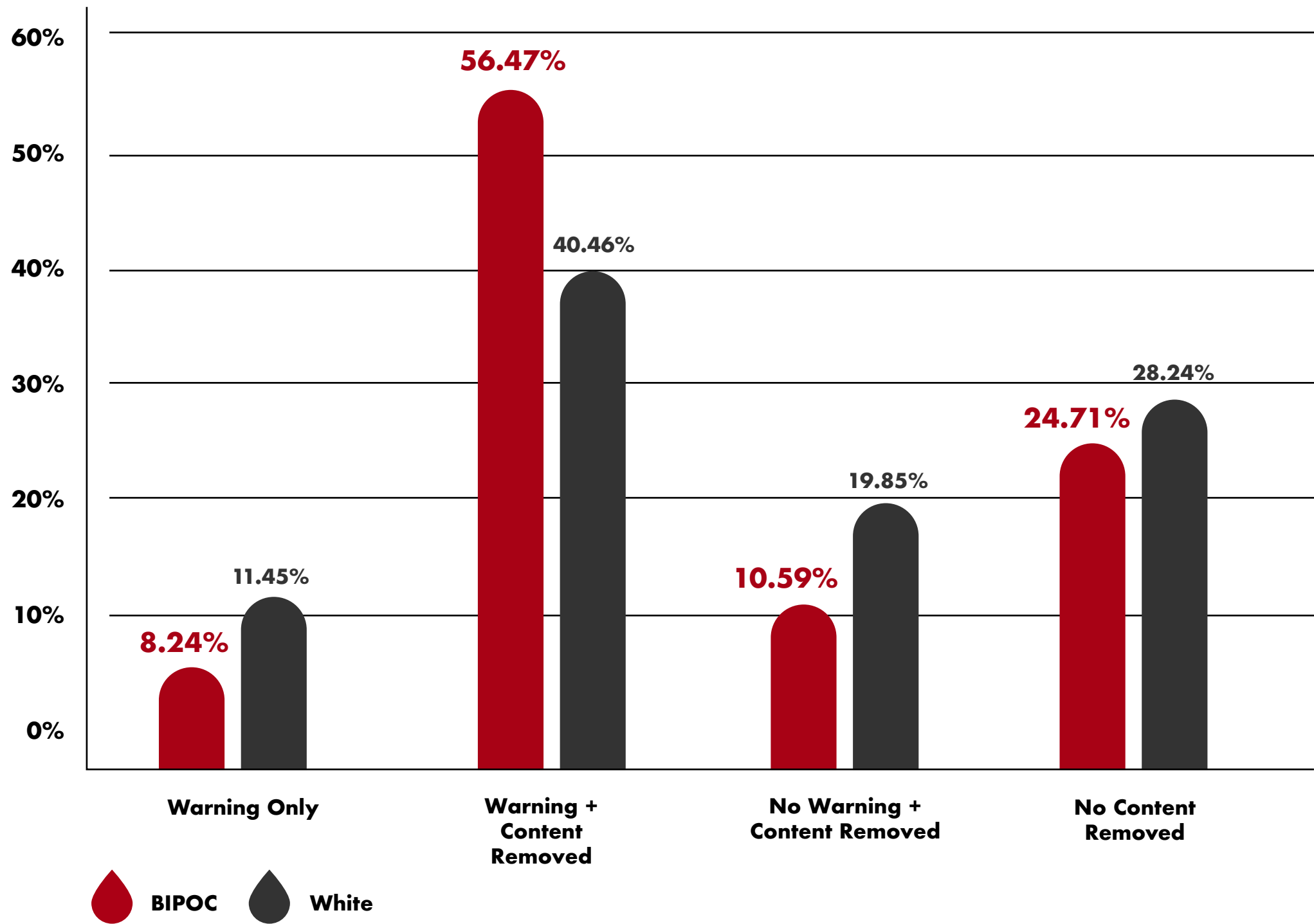


CHART 7: PERCENTAGE OF REASONS FOR REMOVAL: BIPOC



DISABLED.

CHART 8: PERCENTAGE OF CONTENT REMOVAL BY DISABLED STATUS

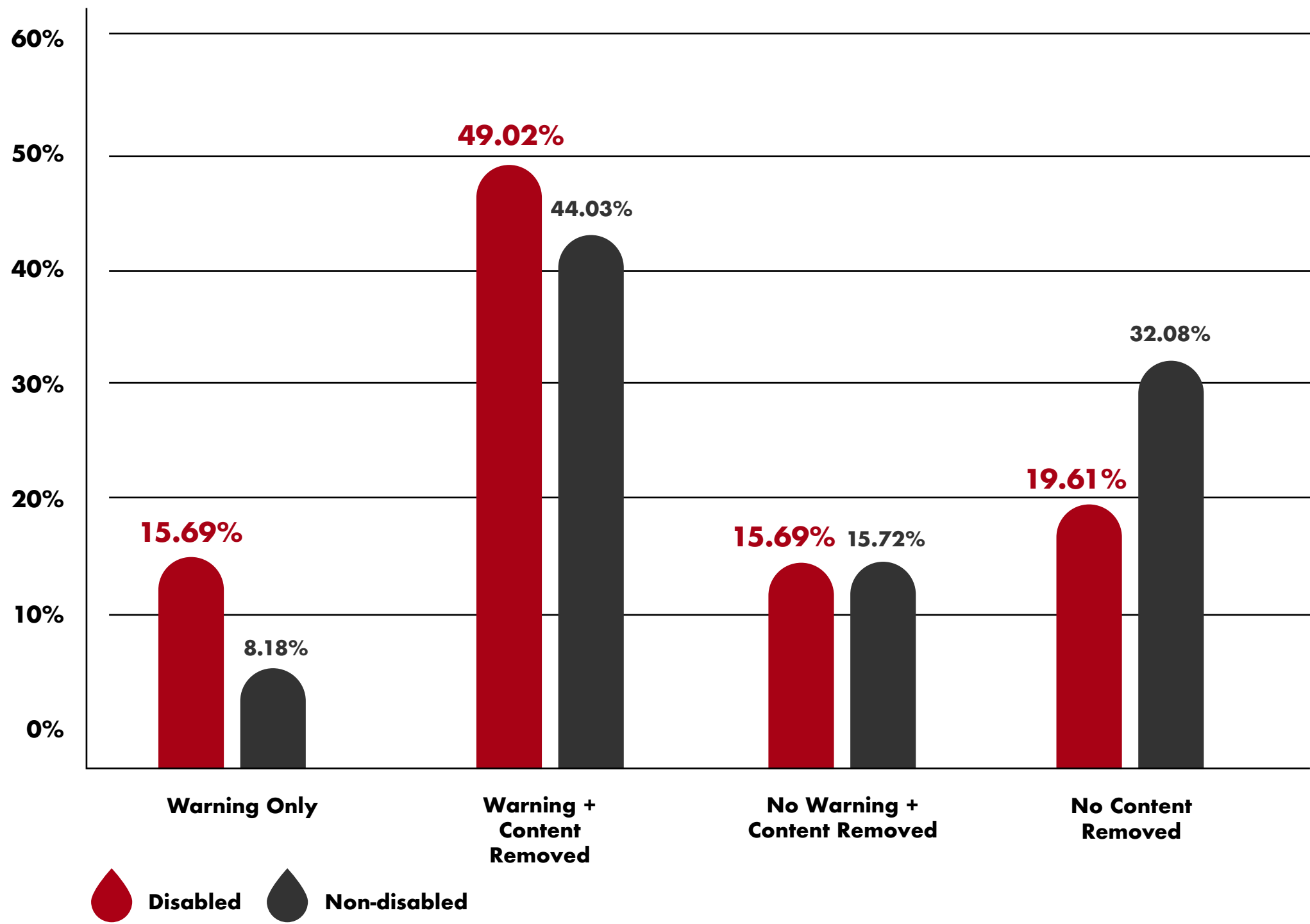


CHART 9: PERCENTAGE OF REASONS FOR REMOVAL: DISABLED



PLUS-SIZED.

CHART 10: PERCENTAGE OF CONTENT REMOVAL BY PLUS-SIZED STATUS

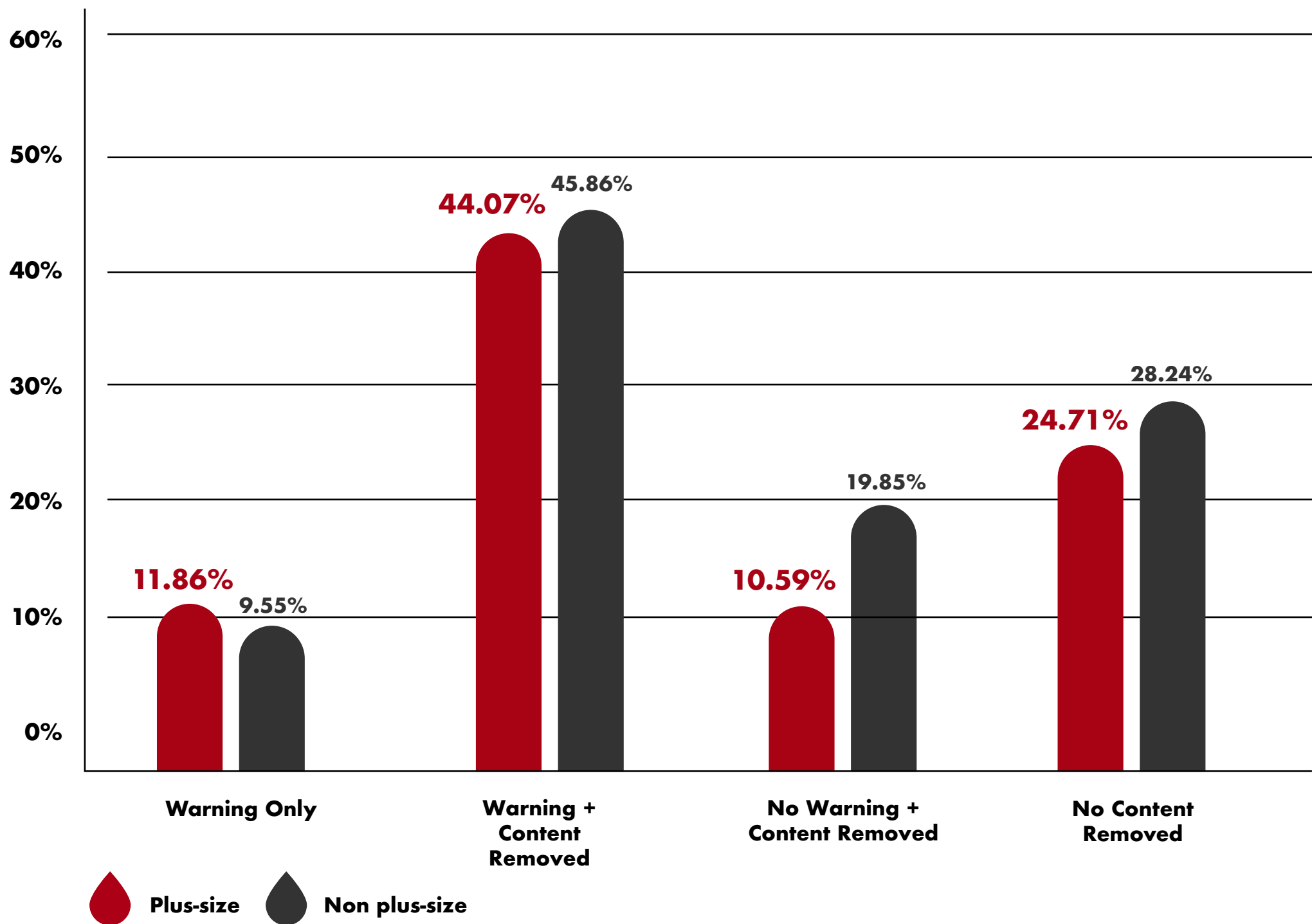
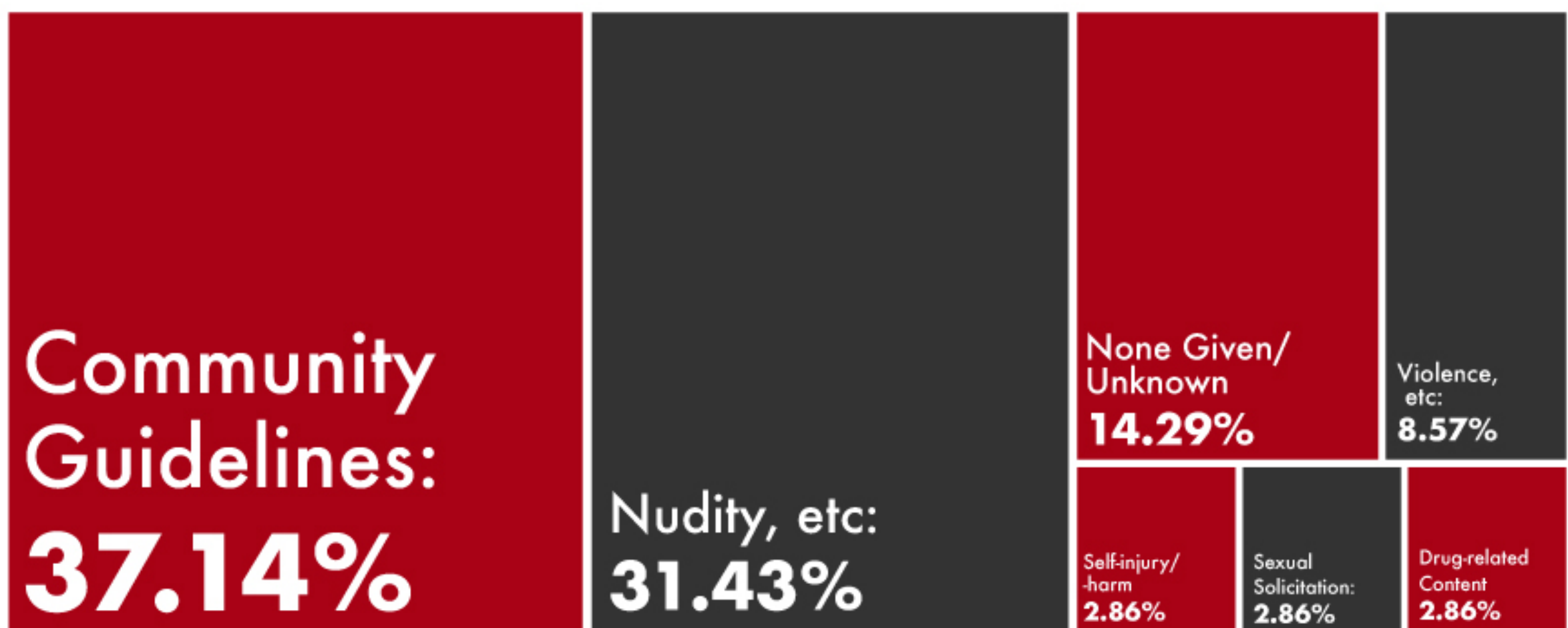


CHART 11: PERCENTAGE OF REASONS FOR REMOVAL: PLUS-SIZED



SEX WORKER.

CHART 12: PERCENTAGE OF CONTENT REMOVAL BY SEX WORKER STATUS

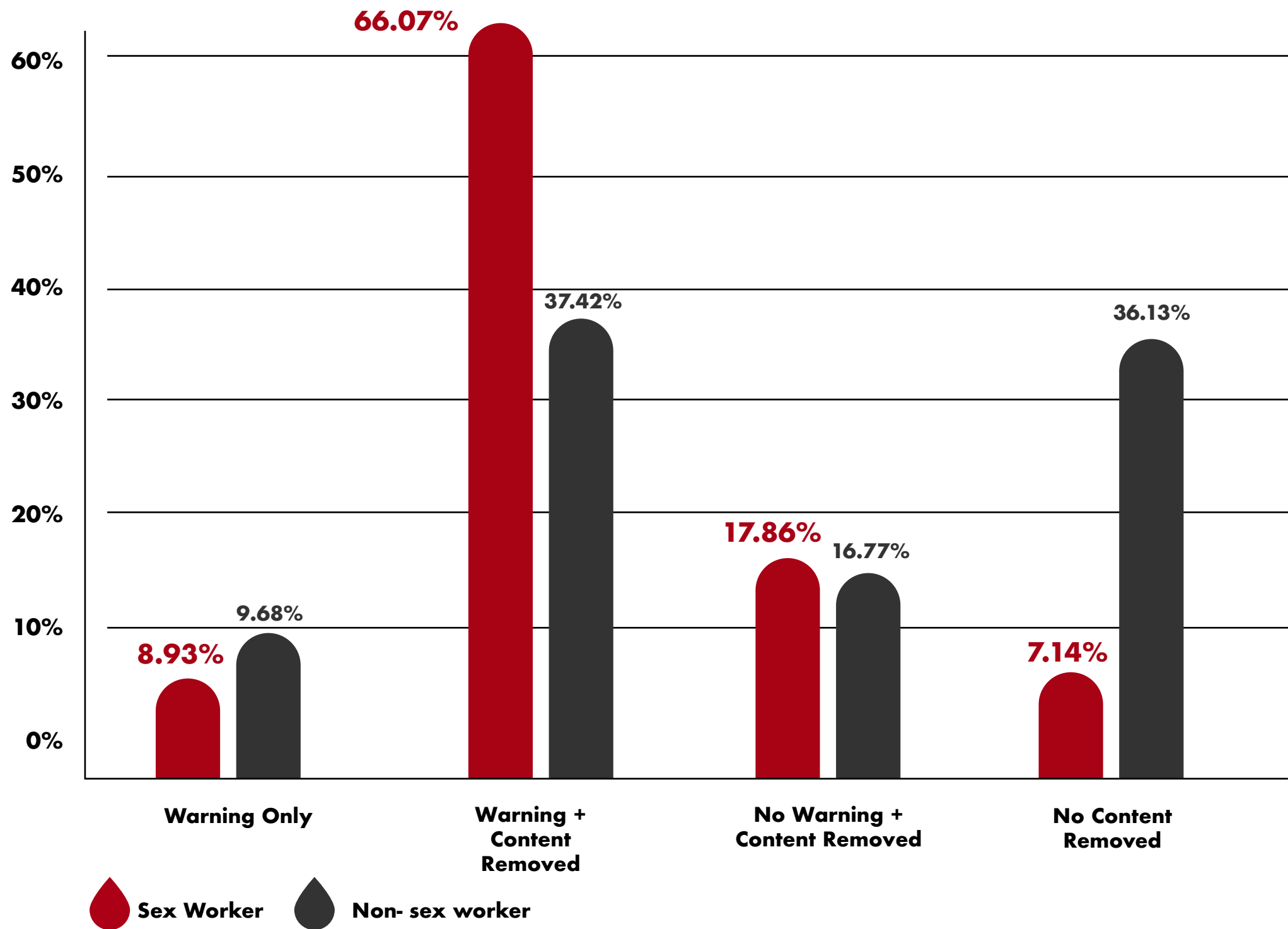
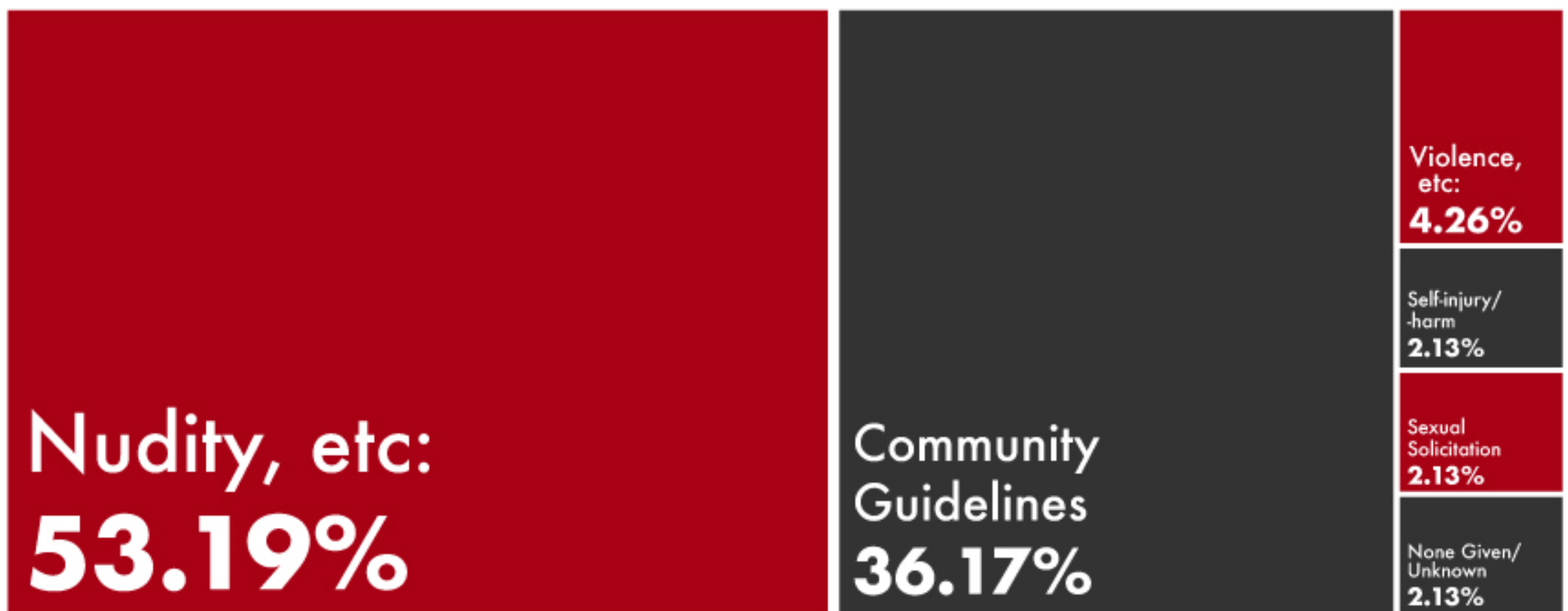


CHART 13: PERCENTAGE OF REASONS FOR REMOVAL: SEX WORKER



SEX EDUCATOR.

CHART 14: PERCENTAGE OF CONTENT REMOVAL BY SEX EDUCATOR STATUS

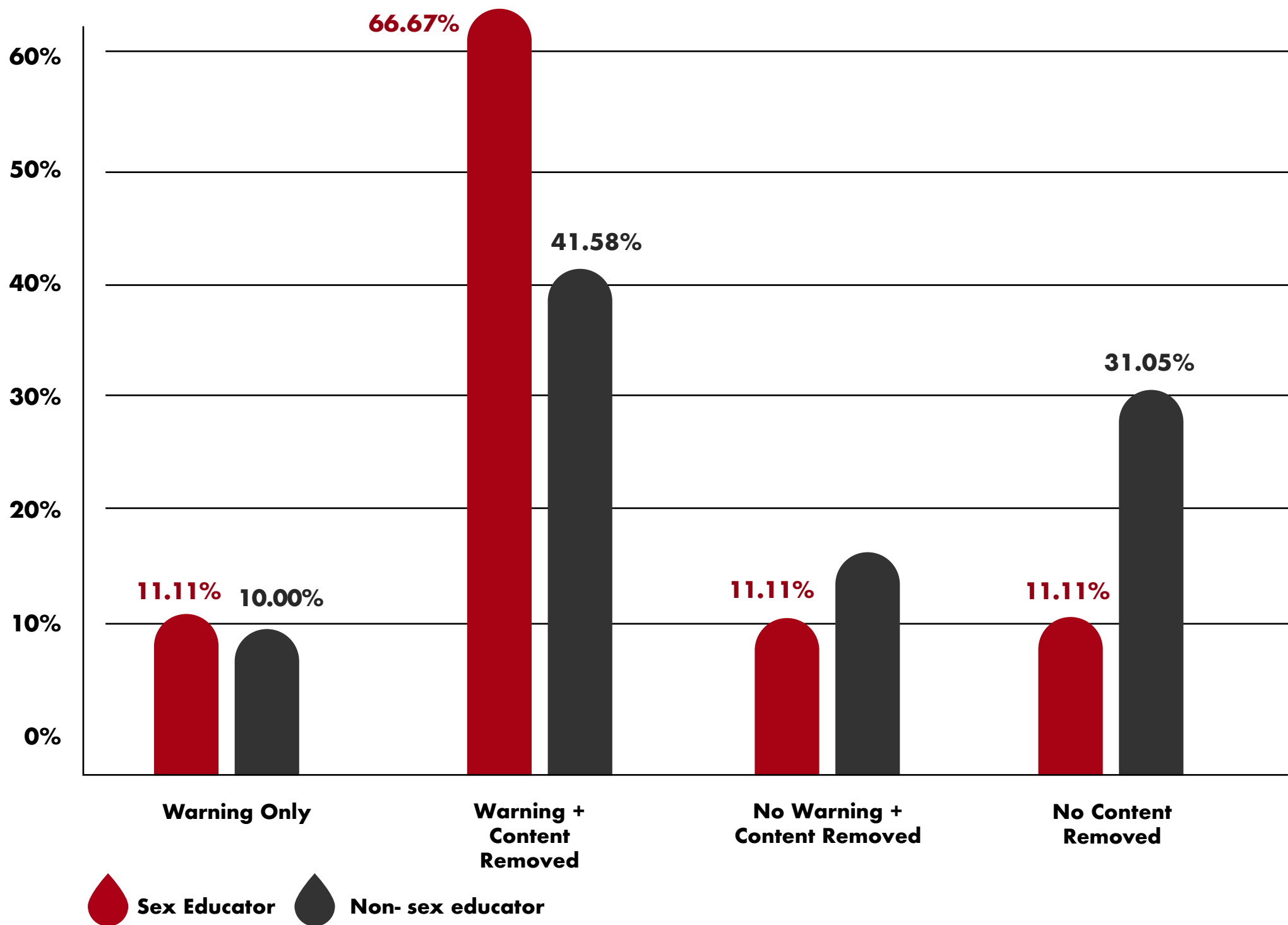
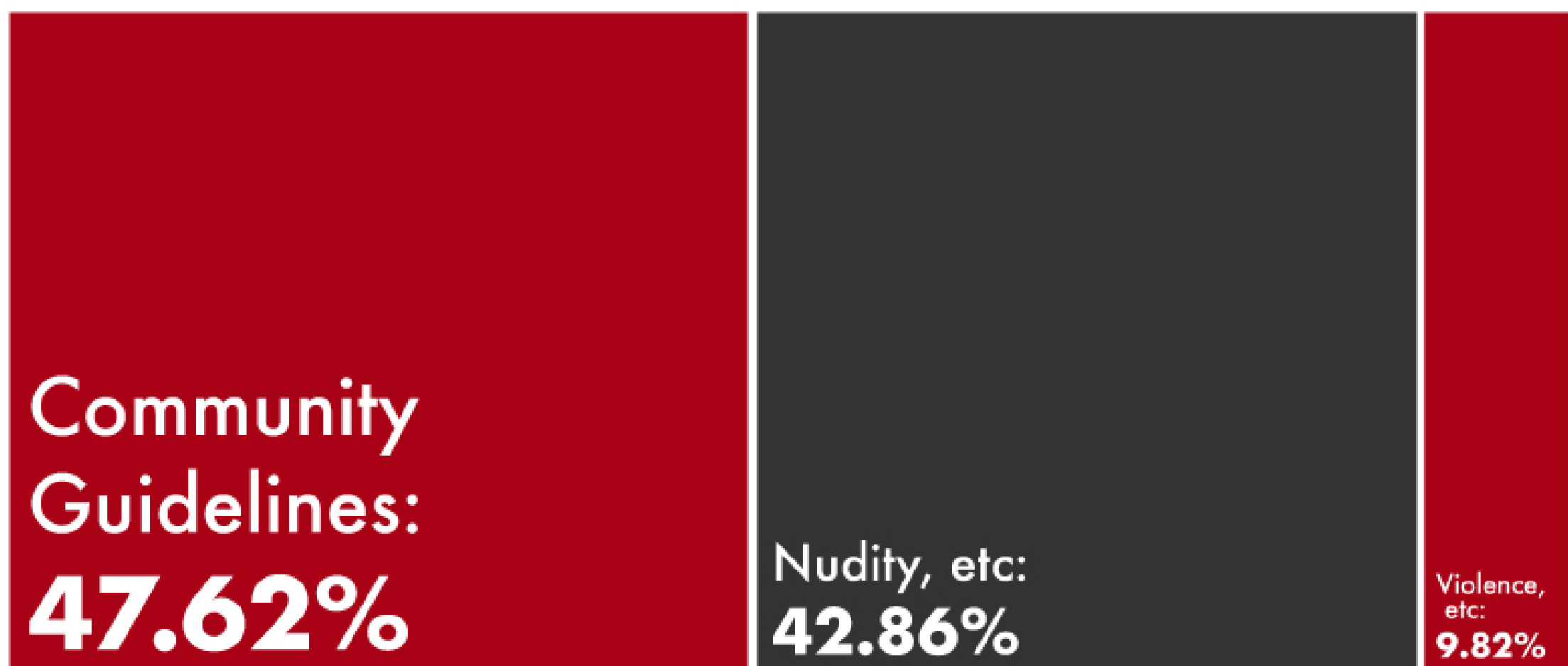


CHART 15: PERCENTAGE OF REASONS FOR REMOVAL: SEX EDUCATOR



ONGOING EFFORTS AND SUPPORT

In this survey and report, we identified several important future directions for new research on social media censorship. For instance, many participants noted negative financial consequences resulting from their content removals. Additionally, many participants discussed instances of shadowbanning in their open-ended responses, rather than outright content removals. Thus, in our ongoing research efforts, we will expand our focus to include social media censorship's financial impacts and on shadowbanning experiences (extending Hacking/Hustling's important work in these areas [Blunt, Coombes, Mullin, & Wolf, 2000; Blunt & Wolf, 2000]).

Since the present survey was distributed to the Salty community, it primarily included people who are marginalized in one or more ways. Thus, the comparisons to privilege groups detailed in this report are likely much different than they would be if we compared participants' experiences to the general population. It is likely that the difference in proportion of content removals observed in the present report would be even more disparate in a wider, more generalized sample (e.g., as found in Haimson et al. 2021). For this reason, our next survey and report will attempt to survey a wider audience.

Finally, one major limitation in our data preparation and analysis relates to coding participants' demographic information. We chose to prompt participants to enter their gender status, LGBTQIA+ status, and BIPOC status demographic information via fill-in-the-blank to allow the most autonomy to self-describe. However, to quantify this information as needed for this report's analyses, participants' responses had to be aggregated, which certainly led to folks being categorized in ways they may not have chosen for themselves. This issue has been a priority adaptation in our next survey.

To participate in our newest survey, and to contribute your voice and experiences with social media censorship (on any platform!), please follow this [link](#).

Salty relies on contributions and volunteers to survive. If you believe the work we are doing with this kind of research has value – please [click here](#) to become a Member or [here](#) to make a one off contribution.

For follow up questions regarding this report, please email shakira@saltyworld.net

FURTHER READING ON SOCIAL MEDIA CENSORSHIP AND MARGINALIZED GROUPS

Carolina Are. 2020. **How Instagram's algorithm is censoring women and vulnerable users but helping online abusers.** *Feminist Media Studies* 20, 5: 741–744. <https://doi.org/10.1080/14680777.2020.1783805>

Danielle Blunt and Ariel Wolf. 2020. **Erased: The Impact of FOSTA-SESTA & the Removal of Backpage.** *Hacking//Hustling*. <https://hackinghustling.org/wp-content/uploads/2020/01/HackingHustling-Erased.pdf>

Ysabel Gerrard. 2020. **Social media content moderation: six opportunities for feminist intervention.** *Feminist Media Studies* 20, 5: 748–751. <https://doi.org/10.1080/14680777.2020.1783807>

Oliver L Haimson, Dan Delmonaco, Peipei Nie, and Andrea Wegner. 2021. **Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas.** *Proc. ACM Hum.-Comput. Interact.* <http://oliverhaimson.com/PDFs/HaimsonDisproportionateRemovals.pdf>

Dottie Lux and Lil Miss Hot Mess. 2017. **Facebook's Hate Speech Policies Censor Marginalized Users.** *Wired*. <https://www.wired.com/story/facebook-hate-speech-policies-censor-marginalized-users/>

Sarah T. Roberts. 2019. **Behind the Screen: Content Moderation in the Shadows of Social Media.** Yale University Press, New Haven, CT.

Kristen Vaccaro, Christian Sandvig, and Karrie Karahalios. 2020. **"At the End of the Day Facebook Does What It Wants": How Users Experience Contesting Algorithmic Content Moderation.** *Proc. ACM Hum.-Comput. Interact* 4, CSCW2: 22. <https://dl.acm.org/doi/abs/10.1145/3415238>

Carolina Are. 2021. **The Shadowban Cycle: an autoethnography of pole dancing, nudity and censorship on Instagram.** *Feminist Media Studies*: 1–18. <https://doi.org/10.1080/14680777.2021.1928259>

Christina Dinar. 2021. **The state of content moderation for the LGBTQIA+ community and the role of the EU Digital Services Act.** Heinrich-Böll-Stiftung. <https://us.boell.org/en/2021/06/21/state-content-moderation-lgbtqia-community-and-role-eu-digital-services-act-1>

Ysabel Gerrard and Helen Thornham. 2020. **Content moderation: Social media's sexist assemblages.** *New Media & Society* 22, 7: 1266–1286. <https://doi.org/10.1177/1461444820912540>

Shagun Jhaver, Darren Scott Appling, Eric Gilbert, and Amy Bruckman. 2019. **"Did You Suspect the Post Would be Removed?": Understanding User Reactions to Content Removals on Reddit.** *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW: 192:1-192:33. <https://doi.org/10.1145/3359294>

Brandeis Marshall. 2021. **Algorithmic misogyny in content moderation practice.** Heinrich-Böll-Stiftung. <https://us.boell.org/en/2021/06/21/algorithmic-misogyny-content-moderation-practice-1>

Salty. 2019. **Exclusive: An Investigation into Algorithmic Bias in Content Policing on Instagram.** Salty. <https://www.saltyworld.net/algorithmicbiasreport-2/>

Danielle Blunt, Emily Coombes, Shanelle Mullin, and Ariel Wolf. 2020. **Posting into the Void.** *Hacking//Hustling*. <https://hackinghustling.org/wp-content/uploads/2020/09/Posting-Into-the-Void.pdf>

Electronic Frontier Foundation. 2019. **EFF Project Shows How People Are Unfairly "TOSsed Out" By Platforms' Absurd Enforcement of Content Rules.** Electronic Frontier Foundation. <https://www.eff.org/press/releases/eff-project-shows-how-people-are-unfairly-tossed-out-platforms-absurd-enforcement>

Tarleton Gillespie. 2018. **Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media.** Yale University Press, New Haven.

Knight Foundation and Gallup, Inc. 2020. **Free Expression, Harmful Speech and Censorship in a Digital World.** https://knightfoundation.org/wp-content/uploads/2020/06/KnightFoundation_Panel6-Techlash2_rprt_061220-v2_es-1.pdf

Sarah Myers West. 2018. **Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms.** *New Media & Society* 20, 11: 4366–4383. <https://doi.org/10.1177/1461444818773059>

Sanna Spišák, Elina Pirjatanniemi, Tommi Paalanen, Susanna Paasonen, and Maria Vihlman. 2021. **Social Networking Sites' Gag Order: Commercial Content Moderation's Adverse Implications for Fundamental Sexual Rights and Wellbeing.** *Social Media + Society* 7, 2. <https://doi.org/10.1177/20563051211024962>

Salty.

ABOUT SALTY

Salty is a 100% independent, membership supported newsletter, bouyed by the voices of women, trans and nonbinary contributors from all over the world.

Salty launched on International Women's Day in 2018. (But was kicked off Mailchimp a few hours later, lol.)

Salty's mission is to pass the mic to Salty babes across the world and amplify their unique stories on our platforms – in our newsletter, on our website and via our social media channels.

We fight for digital visibility for women, trans and non binary people and are working everyday to make sure our stories are not erased from the Internet. Salty's Algorithmic Bias Research Collective explores and exposes algorithmic bias against our community.

The Salty website has been recognized by the United States Library of Congress to be a digital artifact of historical importance and is kept in the national digital archive.

We need your support to survive. If you think that the work Salty does has value, please become a paying member or make a one-off contribution. Our choose-what-you-pay memberships start at \$4.99 per month.

BECOME A MEMBER : SELECT YOUR PAYMENT LEVEL

\$4.99 / month

\$9.99 / month

\$19.99 / month

\$55 / year

\$100 / year

\$200 / year

Make a one-off contribution

Salty.

SALTYWORLD.NET

M UNIVERSITY OF
MICHIGAN

**To contribute your voice and experiences
with social media censorship**

**Please participate
in our newest survey.**

tinyurl.com/saltysurvey2