When culture and politics cross roads: what is enhancing gender-abuse culture

Anna Maria Fetz

MSc Data Science, Università degli Studi di Trento

Computational Social Science, 2021 - 2022

`annamaria.fetz@studenti.unitn.it`

Github Repository for code and reference datasets

Abstract

*This paper aims to understand which factors have a major impact over sexual harassment (**SH**) tolerance, and whether this can be detected through both statistical indexes and sampled data.*

Keywords: Linear Regression Model (**LM**), Harassment, Rape, Latent Class Analysis (**LCA**), Abuse, Gender.

Research

Previous papers were revised in order to understand the width of the umbrella under which different concepts of sexual harassment were falling, and to further specify the research question. Following literature proved to be of utmost importance while conducting the study *and* doing inference over the obtained results.

1. Kenneth C. Barker[1] discussion over "Sexual Harassment Experience, Psychological Climate, and Sex Effect on Perception of Safety", aiming to study the effects of workplace's sexual harassment experience, and the perception of workplace sexual harassment psychological climate on an employee's perceived safety from sexual harassment, given employee's sex. The outcome showed workplace sexual harassment and perceived workplace sexual harassment psychological climate to be both significant predictors of perceived safety from sexual harassment, *regardless of gender*.

2. Rossalina Latcheva & al.[2] dispute over the multiple facades of sexual harassment, measured by the FRA[3] Survey, ranging from well-known working environment discrimination, to general unwanted sexual attention and/or discrimination. Sexual harassment is not only interpreted through statistical lenses, but socioeconomic ones as well, through which a gap between harassment **conceptualization** and its **actual resemblances** is highlighted, showing the consistent impact culture can have, as well as the eventual inconsistencies one might encounter while doing research over this topic. FRA displays how the observed variations between EU Member States in the prevalence rates of sexual harassment can be explained by:

   - The acknowledgment level of sexual harassment on national legislation and its prioritization;
   - Presence of political debates reflecting women's overall level of awareness;
   - Disclosure level regarding sexual harassment.

3. Due to the variegate interpretative range over SH' concept, Khoshnood & al.[4] have been considered with regards to Sweden, the European country with the highest rate of reported sexual crimes. The scope of this research was to use LCA to try to identify different classes of offenders, according to some observed characteristics. It allowed to highlight two main classes: "high offenders' class" and "low offenders' class". Group A showed both higher education rate and a lower age average, first or second generation immigrants, while group B identified older men, al-

[1] Kenneth C. Barker. "Sexual Harassment Experience, Psychological Climate, and Sex Effect on Perception of Safety." PhD thesis. Minneapolis, Minnesota (USA): Walden University, 2017.

[2] Rossalina Latcheva and Goodey Joanna. "Sexual Harassment in the European Union: A Pervasive but Still Hidden Form of Gender-Based Violence". In: *Journal of Interpersonal Violence* (2017).

[3] *FRA*. URL: `https://fra.europa.eu/en/content/violence-against-women-survey-information`.

[4] Ardavan Khoshnood et al. "Sexual Harassment: A Complex Adaptive System, Viewpoint Swedish rape offenders — a latent class analysis". In: *Forensic Sciences Research* (2021).

most the half having lower-education only. Both groups were convicted for aggravated crimes (e.g; rape).

4. Marina Astakhova & al.[5] have been taken into consideration for assumptions made about individual factors impacting the incidence of Sexual Harassment within society. Namely:

- **Gender**: it is more likely for a female to face sexual harassment and for males to be perpetrators.
- **Age**: the younger (< 35), the more likely it is to face SH for women, while the older (=> 40), the more likely it is for men to act as perpetrator.
- **Job position**: women occupying higher positions tend to be more educated, assertive, and self-defensive, while younger women who occupy lower positions often lack the ability and power to respond properly to SH behaviors. Existing studies on male perpetrators demonstrate that higher job positions are associated with increased proclivity to sexually harass (Uggen & Blackstone, 2004).
- **Sexually permissive attitudes (SPA)** : Males hold SPA that are more permissive than those of females (DuBois, Faley, and Knapp, 2008). The probability of experiencing SH, as either a target or a perpetrator, increases for individuals who hold more SPA.
- **Personal history of SH**: it may be predictive of future SH experience (Chiodo, Wolfe, Crooks, Hughes, & Jaffe, 2009; Lucero et al., 2006), as women who have experienced sexually harassing behaviors are more likely to become a target of additional SH (Fitzgerald, Swan, & Fischer, 1995). This may be due to major SH toleration out of fear or self-blame. Oppositely, research on perpetrators demonstrates that they are likely to repeat harassing behaviors in the future, and even escalate their severity (Lucero et al., 2006).

The main goal of this research is to assess the evolution of political and social factors on SH *conception* and *perception* across European countries, given the current legislation. With the aim of answering the above-mentioned question, two main approaches will be used: on the one hand, official information[6] regarding rape and abuse reported cases will be examined through standard linear regression, to understand the most consistent predictors determining rape rate in a country. On the other hand, as Latcheva[7] states, due to the latency of collected outcomes throughout years and countries, the major impact of individual perception on SH, and cultural and psychological aspects, further assessments are set through LCA, based on survey data (2021). Due to the issue of misrepresentation, for the sake of the project many different definitions were included within *sexual harassment*. Namely: sexual harassment, rape, sexism, and gender abuse.

Lastly, the lack of an uniform definition for *sexual harassment* and *rape* across Europe, is worth noticing. Namely, only 11 countries have laws defining rape as sex without consent (Belgium, Croatia, Cyprus, Denmark, Germany, Greece, Iceland, Ireland, Luxembourg, Malta, Sweden), and only the *Equal Treatment Directive/2002/73/EC* and its recast *Directive2006/54/EC* handle the concept of SH. Therefore, statistically relevant outcomes are analyzed while generalizing to a broader SH term.

Data and Methodology

Both the online-survey and official data required most of the socioeconomic factors aligned with the ones analysed in the FRA.[8] The most consistent resource for *official* data has been OECD,[9] from which four main indicators have been considered:

- **Violence against women**, split into: attitude towards violence index, prevalence of violence in the lifetime percentage, and presence of laws on domestic violence percentage.
- **Discriminatory family code**, from which early marriage percentage has been extracted.
- **Women in politics**, from which the percentage of women in parliaments has been extracted.
- **Social institutions and gender (SIGI)**, split into: restricted civil liberties index, and discrimination in the family index.

EIGE[10]'s has been used as main reference for:

- **Sexual Violence**, split into sexual assault, rape, sexual harassment.

[5]Marina Astakhova and Cathy L.Z. Dubois. "Sexual Harassment: A Complex Adaptive System Viewpoint". In: *Gender Technology and Development* (2015).

[6]*OECD Statistics*. URL: https : / / www . oecd - ilibrary.org/statistics.

[7]Latcheva and Joanna, "Sexual Harassment in the European Union: A Pervasive but Still Hidden Form of Gender-Based Violence".

[8]*FRA*.

[9]*OECD Statistics*.

[10]*eige EU*. URL: https://eige.europa.eu/gender-statistics/dgs/browse/genvio.

- **Attitudes and perception**, split into: perception of incidence of gender-based violence, and opinion on gender-based violence.

Lastly, data regarding gender pay-gap, education level and rape rate per country have been collected from Eurostat,[11] as well as the share of people agreeing on SH being a relevant issue in their country from Statista.[12] Thus, collected data comprehend both qualitative and quantitative demographic measures, aligned with the ones used in the FRA.

Several techniques were involved to understand the extent of political and psychological impact on SH perception: firstly, a data frame was created with the following columns: country, percentage of women agreeing to violence, women representation in parliament, population percentage considering sexual harassment as an issue, index of gender discrimination within family, index of restricted civil liberties according to gender, laws against domestic violence, percentage of multiple violence occurred during lifetime, population share with a maximum level of secondary education, gender pay gap, rape rate and early marriage. Secondly, preliminary analyses were applied to understand relationships across variables. Then, a hybrid model subset selection strategy was adopted to understand the ideal amount of explanatory predictors, given the large number of explanatory variables. Two LMs were then applied on the data frame: one considering Sweden, and the other discarding it. Main reason of this being that Sweden has got the highest rate of reported rapes and SH cases in Europe, mainly due to different and more detailed legislation regarding SH definition than other European countries.[13] Therefore, due to the irregular amount of reported cases, it is considered as an outlier. Simple linear regression was used due to its bias-variance trade off management, and fairly consistent predicted outcomes despite the little information. Finally, LCA was performed on the outcomes of the online survey, in order to retrieve some unmeasured, community-level variable, such as different shapes in culture and beliefs. The survey presents a three-level decoupling: firstly, age, provenance and gender were required. Then, questions regarding **factual** happenings have been asked, such as whether some kind of SH has ever happened and whether this has been reported. Lastly, it was asked to rate the self-perception of SH being problematic in one's country and public perception. These three levels allow to respond to three different requirements:

1. Social placement
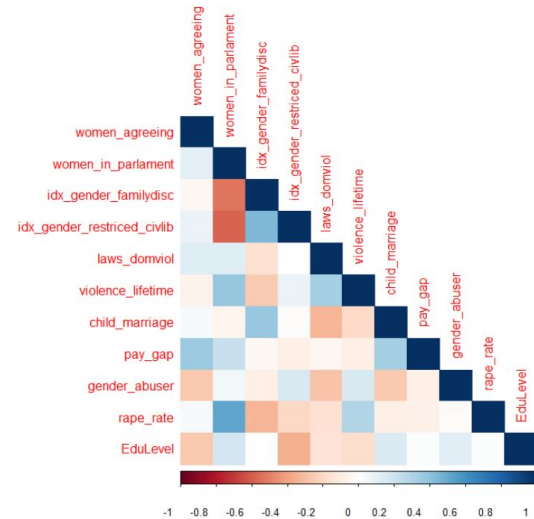2. Factual Experience
3. Personal Perception

Analysis

Figure 1: Correlation Matrix

As figure 1 shows, one may notice that women parliamentarian representation is negatively correlated with restricted civil liberties and discrimination within family (e.g; whether women and men have the same right to be the legal guardian of a child during marriage), both being positively correlated between each other. Worth of notice is rape rate being positively correlated with the amount of women in the parliament as well. The main conclusion to be drawn perfectly reflects what literature[14] states, namely, higher-educated women tend to recognize discrimination more often and are more aware of laws protecting them, thus the major number of reported cases. Additionally, the higher the education level, the higher the job position, thus being able to reflect a social issue better. The slight negative correlation between women agreeing to violence, and the educational level, tend to enhance this statement. All variables, besides provenance country, were considered in the LM. However, the difficulty of being able to measure what is considered as SH, and personal SH perception, immediately arises when applying the model over rape rate by gender, as shown in figure 2. As it can be noticed, only two vari-

---

[11]*Eurostat.* URL: `https://ec.europa.eu/eurostat/statistics-explained`.

[12]*Statista.* URL: `https://www.statista.com`.

[13]Khoshnood et al., "Sexual Harassment: A Complex Adaptive System, Viewpoint Swedish rape offenders — a latent class analysis".

[14]Latcheva and Joanna, "Sexual Harassment in the European Union: A Pervasive but Still Hidden Form of Gender-Based Violence".

```
Call:
lm(formula = df$rape_rate ~ ., data = df[, -c(1, 4)])

Residuals:
     Min      1Q  Median      3Q     Max
-13.6692 -5.3400 -0.9038  2.7830 28.2152

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               25.39088   28.22349   0.900   0.3817
women_agreeing            -0.01465    0.46791  -0.031   0.9754
women_in_parlament         0.85534    0.37120   2.304   0.0350 *
idx_gender_familydisc     -0.17238    0.44465  -0.388   0.7034
idx_gender_restricted_civlib 0.42798  0.44516   0.961   0.3506
laws_domviol             -24.34240   12.01736  -2.026   0.0598 .
violence_lifetime          0.32598    0.50570   0.645   0.5283
child_marriage            -1.26525    1.96122  -0.645   0.5280
pay_gap                   -0.42195    0.48283  -0.874   0.3951
gender_abuser             -0.47667    0.34003  -1.402   0.1801
EduLevel                   0.02424    0.28276   0.086   0.9327
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.21 on 16 degrees of freedom
Multiple R-squared:  0.4891,    Adjusted R-squared:  0.1697
F-statistic: 1.531 on 10 and 16 DF,  p-value: 0.2157
```

Figure 2: Linear regression over Rape Rate

```
Call:
lm(formula = df$rape_rate ~ women_in_parlament + laws_domviol +
    pay_gap + idx_gender_restriced_civlib + gender_abuser, data = df)

Residuals:
     Min      1Q  Median      3Q     Max
-15.075  -5.415  -3.086   2.915  31.622

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                10.8116    20.2866   0.533  0.59967
women_in_parlament          1.0137     0.2706   3.746  0.00119 **
laws_domviol              -17.7087     9.2627  -1.912  0.06964 .
pay_gap                    -0.5958     0.3801  -1.567  0.13197
idx_gender_restriced_civlib 0.4537     0.3365   1.348  0.19187
gender_abuser              -0.3295     0.2616  -1.260  0.22164
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.18 on 21 degrees of freedom
Multiple R-squared:  0.4462,    Adjusted R-squared:  0.3144
F-statistic: 3.384 on 5 and 21 DF,  p-value: 0.02131
```

Figure 3: Linear regression over Rape Rate after hybrid subset selection

```
Call:
lm(formula = df_new$rape_rate ~ women_in_parlament + laws_domviol +
    pay_gap + idx_gender_restriced_civlib + gender_abuser, data = df_new)

Residuals:
    Min     1Q Median     3Q    Max
-7.242 -3.110 -1.606  1.952 15.738

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               11.54752   11.08239   1.042    0.310
women_in_parlament         0.31717    0.17744   1.787    0.089 .
laws_domviol              -3.40368    5.44655  -0.625    0.539
pay_gap                   -0.26201    0.21291  -1.231    0.233
idx_gender_restriced_civlib -0.04912  0.19699  -0.249    0.806
gender_abuser             -0.10983    0.14621  -0.751    0.461
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.564 on 20 degrees of freedom
Multiple R-squared:  0.2779,    Adjusted R-squared:  0.09739
F-statistic: 1.54 on 5 and 20 DF,  p-value: 0.2226
```

Figure 4: Linear regression over Rape Rate without Sweden

ables seem to have high explanatory power: the presence of laws on domestic violence, and political representation of women in countries. However, the adjusted R^2 shows that the model has only 17% of explanatory power over rape rate in each country. Thus, there may be some unexplored/unmeasured variables which should be taken into account. This can be further noticed when analysing the outcome of a hybrid-subset approach, stating that 5 variables would be ideal for a good RSS- Explainability trade-off. Thus, re-sampled variables were: the amount of women in parliament, the presence of laws on domestic violence, gender pay gap, abuser gender and restricted civil liberties. This led to the confirmation of the previous assessment, namely, laws on domestic violence and women in parliament being the main proxies of lower rape rate (figure 3). However, explanatory power of the model increased to 31%, with a significantly lower p-value. Thus, although variables do not seem ideal in explaining rape rate on a qualitative level, they seem to have enough quantitative explanatory power. It is worth noticing women in parliament and laws on domestic violence's sign: the former is positive, while the latter is strongly negative, suggesting women in powerful positions to recognize abuse more often, and laws against SH to have strong impact. This can be also proved when checking the LM without Sweden, the country with most Laws against domestic violence: predictor's relevance drops significantly. Thus, due to Swedish detailed legislation against SH, the outcomes are fully biased. Therefore, after re-applying the same procedure on a new data frame not displaying this "outlier", the goodness of the outcomes proves to be worse, as figure 4 shows. Namely, only political representation seems to be a relevant driver of rape rate, together with gender pay gap, in a country, as followed-up in figure 5.

These poor outcomes can be justified in multiple ways, first of which being the latency of rape-reporting across countries and people themselves. As literature shows, most of women tend to report only when the assault is not socially accepted, while underestimating remaining SH. Thus, the main proxy driver introducing such biased results may be the high rate of under-reporting across countries. Moreover, culture tends to consistently influence SH conception and definition even across households.[15] Therefore, strong and non-measurable variability is introduced in the model itself, not allowing reliable outcomes. Due to the lack of consistent official data, an online-survey has been previously built, answers of which have been used to build a LCA in order to distinguish some groups within respondents. The questionnaire has been shared across multiple European and non-European countries, ranging from Age <18 to 50 > in order to grasp a wider population range. 208 respondents were analysed. LCA proved to be successful in defining two main latent

---

[15]Barker, "Sexual Harassment Experience, Psychological Climate, and Sex Effect on Perception of Safety."; Latcheva and Joanna, "Sexual Harassment in the European Union: A Pervasive but Still Hidden Form of Gender-Based Violence"; Astakhova and Dubois, "Sexual Harassment: A Complex Adaptive System Viewpoint".

```
Call:
lm(formula = df_new$rape_rate ~ women_in_parlament + pay_gap,
    data = df_new)

Residuals:
    Min      1Q  Median      3Q     Max
-7.0749 -3.2780 -0.9734  1.8972 16.0311

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)          1.2943     3.7781   0.343   0.7350
women_in_parlament   0.3144     0.1193   2.635   0.0148 *
pay_gap             -0.2332     0.1976  -1.180   0.2500
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.326 on 23 degrees of freedom
Multiple R-squared:  0.2389,    Adjusted R-squared:  0.1727
F-statistic:  3.61 on 2 and 23 DF,  p-value: 0.04331
```

Figure 5: Linear regression over Rape Rate without Sweden, only two proxies

groups of people having answered the survey, no values being held fixed. LCA (figure 6) has been com-
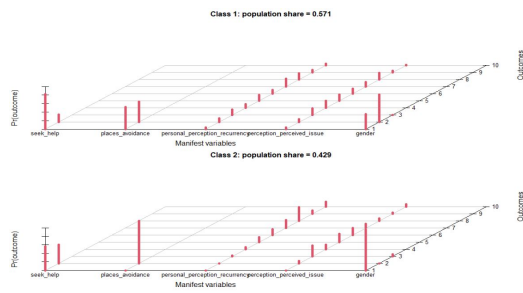

Figure 6: 1st Latent Class Analysis

puted for 2-4 latent classes, with respect to: places avoidance, gender, seek for help, perception of SH as an issue, and perception of how society perceives SH as an issue. However, only 2 and 3 groups proved to be statistically significant, the former being the most explanatory and having the lowest AIC and BIC values. The following may be assessed: respectively to each class, 57% and 43% share of the population are displayed. Although population shares seem to be equally split among classes, given that gender levels are displayed in R as 1 for females, 2 for males, and 3 for other, remarkable is on the one hand, the gender gap in class 1. Indeed, the latter displays 65% males against 35% of women, and no further gender distinctions are present. On the other hand, class 2 shows an even higher disproportion among gender: 95% of respondents are women, 34% are men, and the remaining chose to specify other. When considering class 1, this group of respondents seems to be prone to seek help (while keeping in mind that levels for 'Yes' and 'No' are respectively 2 and 1), indifferent to whether avoid or not certain places (meaning that some people would and some other wouldn't), and with an overall neutral position towards both personal perception of SH as an

issue, and social perception of it. Namely, the distribution is more or less even, not showing very strong peaks anywhere. Oppositely, class 2 is less prone to seek help, and directly avoids certain places considered to be dangerous. Interestingly, both perception scales seem to follow opposite patterns, both with respect to class 1, and with respect to each other: on the one hand, some peaks can be eventually measured on higher values of the scale, meaning that these people personally perceive the issue as relevant. On the other hand, the perception of the perceived issue displays higher peaks on lower values of the scale, meaning that they don't recognize society to be perceiving SH as an issue. Lastly, to further discriminate groups, LCA has been applied by considering whether SH has been experienced during one's lifetime or not as well. As figure 7 shows, this further specification creates an even stronger class-distinction, although both AIC and BIC result in higher values than the previously-adopted model (AIC: 2581 vs AIC: 2789). Groups are split
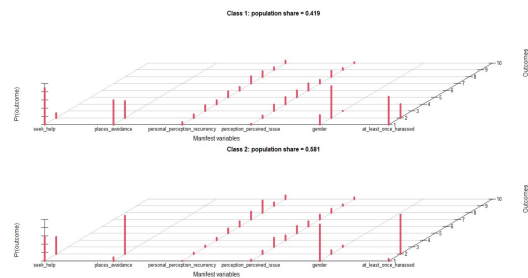

Figure 7: 2nd Latent Class Analysis

into 42% and 58%, and both gender and SH happenings present a stronger disproportion than the previous model. Indeed, class 1 introduces 73% of men, 63% of the population has never been harassed, and is less prone to avoid places and seek for help against SH. On the opposite, class 2 shows 88% of women, 8.8% of men and almost 2% of other gender, SH experience accounts for 95% of times, places avoidance 92% of times, and less than the half of respondents would seek for help (40%). For both classes the same discussion as above can be held for both perception distributions.

Research strengths and vulnerabilities

Despite the interesting outcomes, this research would require some strong enhancements:

- Firstly, although generalizing different SH occasions within SH umbrella may help getting good outline results, research on its whole does not account for differences in definitions. Therefore,

outcomes may be severely biased by this strong generalization.

- Secondly, data collected for linear regression account for Europe only, while the 207 valid questionnaire respondents accounted for non-EU countries as well. Further bias might be introduced by political and social cultures strongly different from European ones. Thus, comparisons between LM outputs and LCA's ones, may be accomplished carefully, as LM notably has been chosen for the lower bias of the model, this may be introduced both by LCA's and by the above-mentioned under-reporting rate and SH generalization.

- Next, some downfalls of an almost-random survey need to be accounted for: although dropouts were discarded, and very good population coverage has been obtained, sampling, inaccuracy errors and coverage may still be considered. Firstly, little inclusion has been accomplished for non-binary and further genders. Secondly, despite having collected answers from a fairly heterogeneous audience, a 38% share belonged to age range 20-25 and 25% to 50 > age. Considering age range 20-25 a target area for SH,[16] some strong bias could be introduced by the fact that respondents could have been experiencing SH episodes during or close to their answer to the questionnaire. Additionally, a self-rating scale for perception may be deceiving.

- Although LM was considered as a good modeling choice given the available data and the overall independence assumption across variables, a Multi-level Linear Model (MLM) may have been more suited for research purposes, as it would have allowed to take into account different levels of each discriminatory index that has been gathered.

- Lastly, distinctions of the **type** of SH asked in the questionnaire, have not been considered, but could have been considered for a further MLM.

Results interpretation and conclusions

LCA's outcomes fully summarize the hypothesized gap between cultural SH triggering factors, and political intervention. Moreover, they introduce a strong causality dilemma between political and cultural interactions.

Namely, although SH is generally perceived within society, having experienced SH at least once in a lifetime severely enhances the perception of it as a recurrent issue[17] and yields to common behaviors, such as being more careful in new places or with new people, and being less talkative about the problem. This introduces the cultural and the psychological impact on an individual, biasing one's actions that could be officially captured, such as not reporting the offender, as it can be noticed by the poor results obtained with linear regression. In light of this, it is also worth noticing how highly politics and female representation in the field impact over SH reported cases. Therefore, on the one hand, SH-accepting cultures, and SH episodes, tend to shape the way individuals act and perceive the issue. On the other, political action seems to have almost 30% of explanatory power in the decrease of SH rate, which is especially true when Swedish regulations and high rape-reporting are taken into account.

References

Astakhova, Marina and Cathy L.Z. Dubois. "Sexual Harassment: A Complex Adaptive System Viewpoint". In: *Gender Technology and Development* (2015).

Barker, Kenneth C. "Sexual Harassment Experience, Psychological Climate, and Sex Effect on Perception of Safety." PhD thesis. Minneapolis, Minnesota (USA): Walden University, 2017.

eige EU. URL: https://eige.europa.eu/gender-statistics/dgs/browse/genvio.

Eurostat. URL: https://ec.europa.eu/eurostat/statistics-explained.

FRA. URL: https://fra.europa.eu/en/content/violence-against-women-survey-information.

Khoshnood, Ardavan et al. "Sexual Harassment: A Complex Adaptive System, Viewpoint Swedish rape offenders — a latent class analysis". In: *Forensic Sciences Research* (2021).

Latcheva, Rossalina and Goodey Joanna. "Sexual Harassment in the European Union: A Pervasive but Still Hidden Form of Gender-Based Violence". In: *Journal of Interpersonal Violence* (2017).

OECD Statistics. URL: https://www.oecd-ilibrary.org/statistics.

Statista. URL: https://www.statista.com.

---

[16]Barker, "Sexual Harassment Experience, Psychological Climate, and Sex Effect on Perception of Safety."; Latcheva and Joanna, "Sexual Harassment in the European Union: A Pervasive but Still Hidden Form of Gender-Based Violence"; Astakhova and Dubois, "Sexual Harassment: A Complex Adaptive System Viewpoint".

---

[17]Barker, "Sexual Harassment Experience, Psychological Climate, and Sex Effect on Perception of Safety."