Clustering and Gradient Boosting Domestic Violence Data to Tailor Outreach Strategies
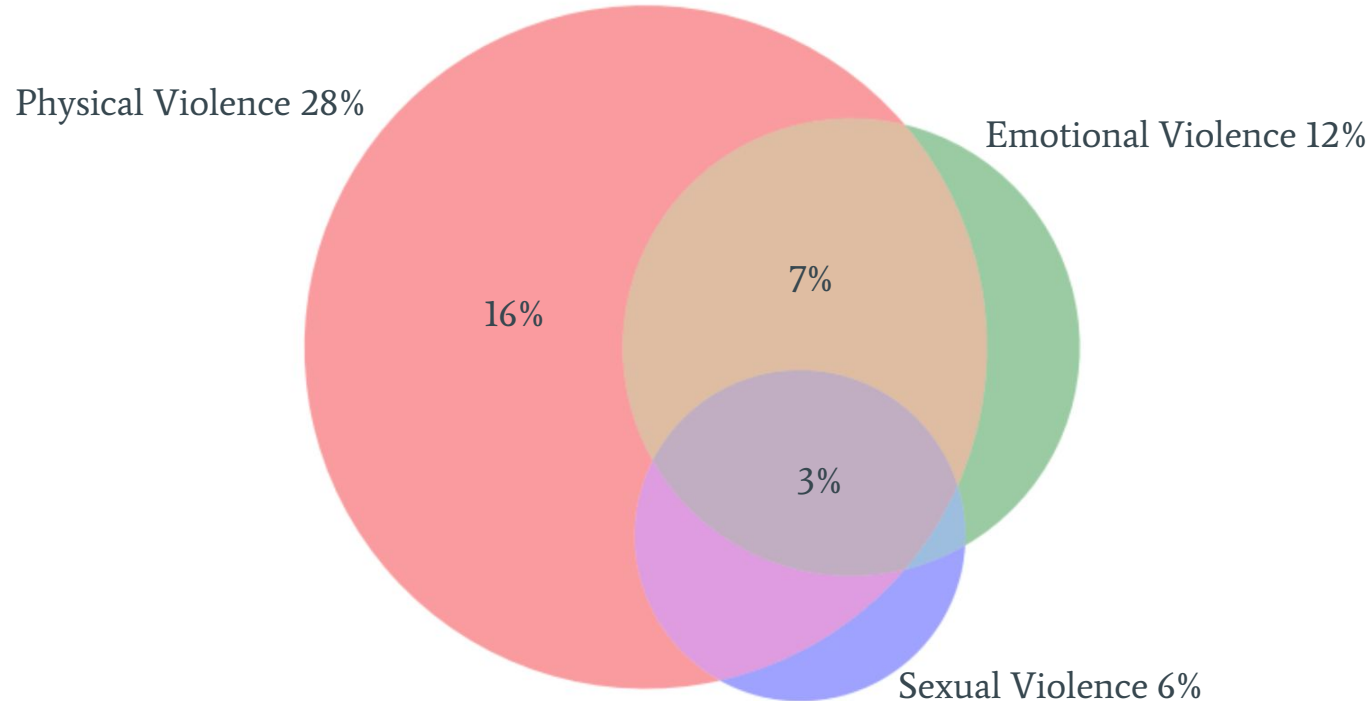
● ● ●

India DHS Surveys 2015-2016
Mary Scott Sanders 10/2018
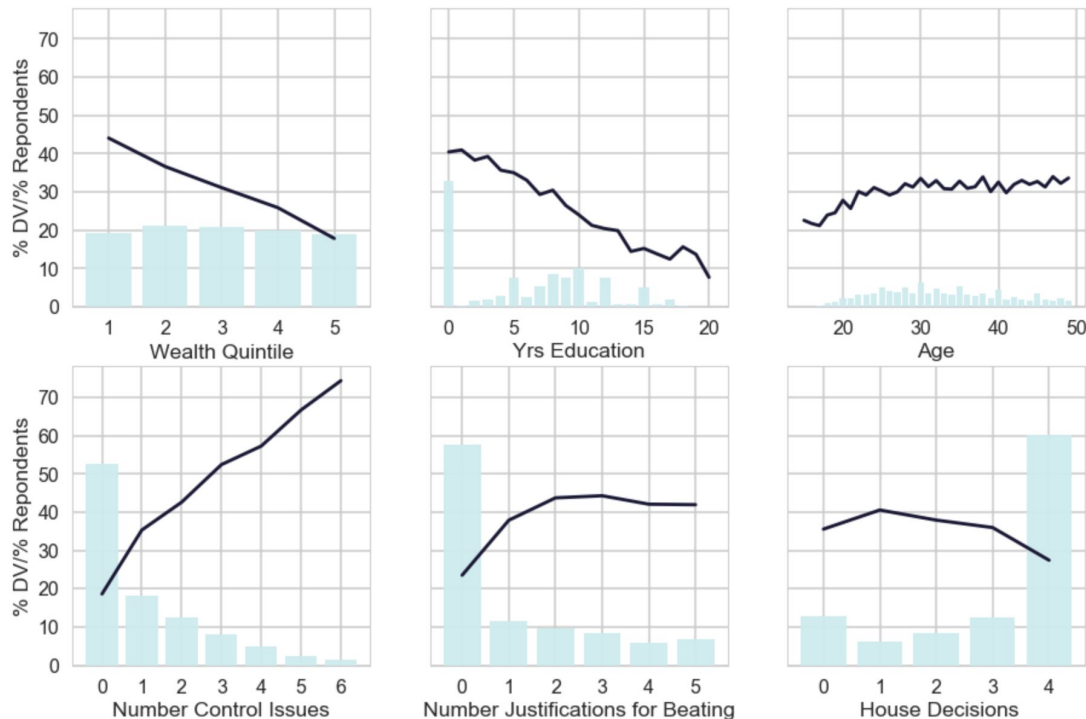
Domestic Violence is Pervasive in India

31% of Women in Unions have Experienced DV

Physical Violence 28%

Emotional Violence 12%

7%

16%

3%

Sexual Violence 6%

Domestic Violence is a Complex Problem

Percent of Women in Unions that Have Experienced Domestic Violence (% DV)

Domestic Violence Requires a Complex Set of Solutions

- Proper Legislation

- Police and Prosecutor Training

- Medical Professional Training and Additional Crisis Centers

- **Social Campaigns ⬅ portion this analysis aims to address**

Results from Clustering by Experience with DV

Feature & Description	Abused	In Danger	Less Harmed
cluster_perc: Percent of sample in cluster	24%	55%	21%
dv_phys_less: Less severe physical abuse reported	98%	3%	11%
dv_sex: Sexual abuse reported	21%	2%	2%
dv_emo: Emotional abuse reported	39%	3%	4%
control_issues: Count of controlling behaviors partner displays	2.0	0.8	0.6
afraid_most: Afraid of partner most of the time	29%	12%	0%
afraid_sometimes: Afraid of partner sometimes	67%	88%	0%
justs: Respondent agrees with 1+ justifications for wife beating	61%	37%	36%

- **Method:** K-Means clustering set to find 3 groups (k=3)
- **Abused Group:** Experience many forms of dv, often afraid.
- **In Danger Group:** Mostly do not report dv. However, partners display controlling behaviors, often afraid
- **Less Harmed Group:** Mostly do not report dv, not afraid of partners

Results from Clustering by Experience with DV

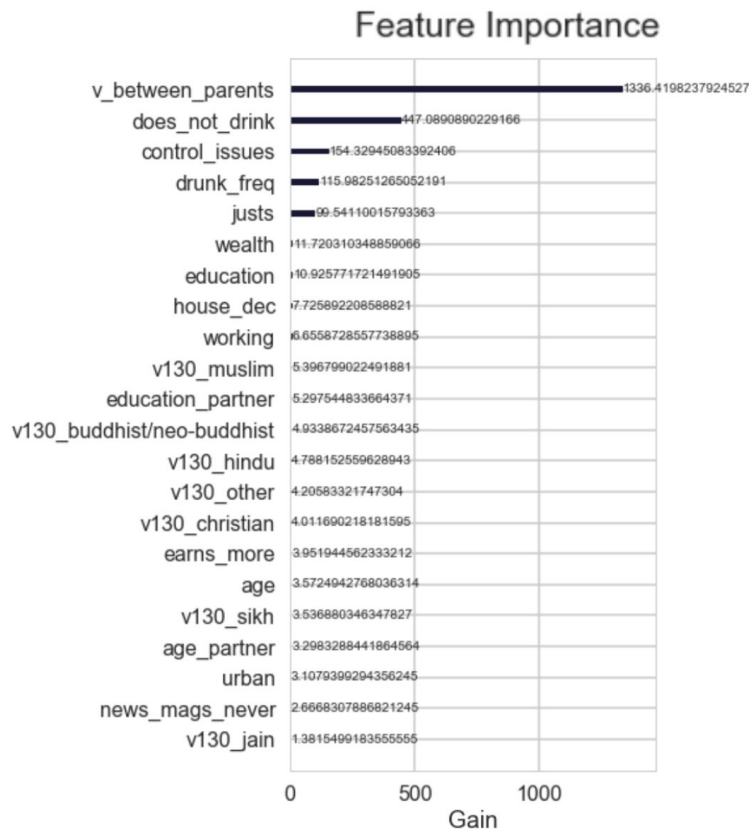| Feature & Description	Abused	In Danger	Less Harmed
tv_never: Respondent never watches tv	33%	25%	18%
news_mags_never: Respondent never reads newspapers/magazines	77%	65%	58%
radio_never: Respondent never listens to the radio	86%	85%	84%

- Abused group has least interactions with media, is least accessible
- **Out of channels, Abused interacts with tv the most (average of 3.5 times a week)**
- TV which is the most equitable medium overall

Results from Clustering by Experience with DV

Feature & Description	Abused	In Danger	Less Harmed
wealth: Wealth quintile	2.5	3.1	3.3
age: Respondent's age	33.2	32.6	33.3
education: Respondent's years of education	4.3	6.3	7.3
education_partner: Partner's years of education	6.1	7.9	8.6

- Wealth and education are negatively correlated with abuse.

- The Abused group has the least education, therefore in order for education to have an impact, curriculum regarding dv must begin early.

- Less Harmed group is best educated, wealthier, and the most access to technology
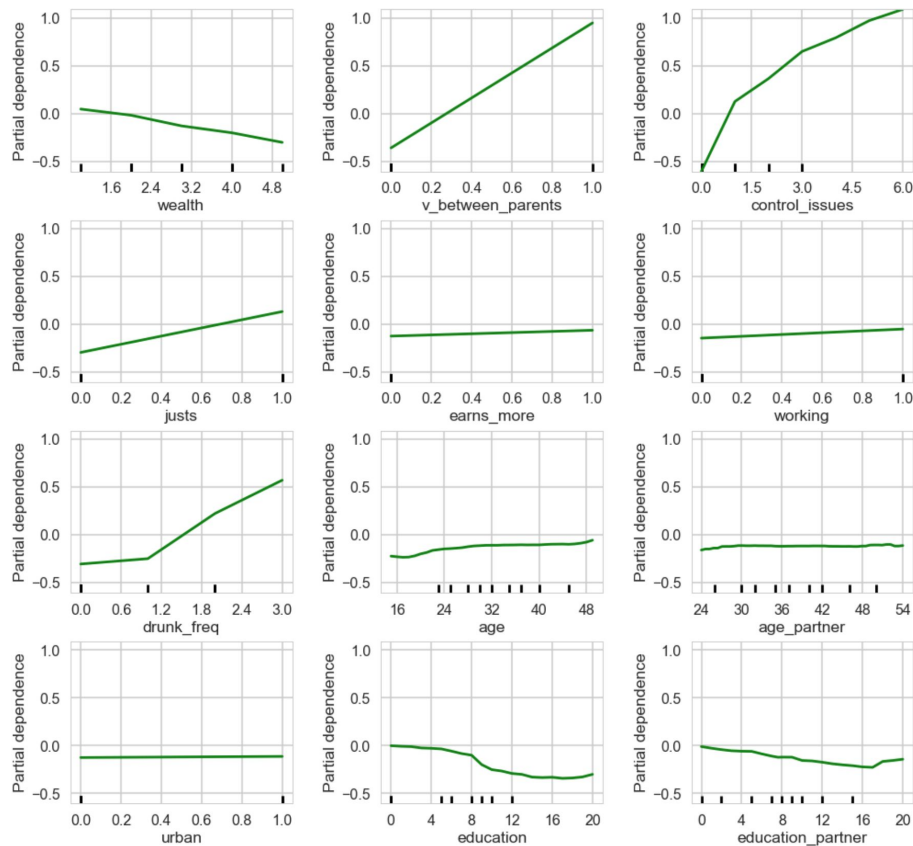
Results from Gradient Boosting Classification

Feature Importance

- Used XGBClassification to predict if a woman would report any violence

- Results: AUC of 67%, 75% accuracy, 45% recall, 68% precision.

- Best performance with 30% of columns used by each tree and max depth of 5.

- Valuable in understanding relative importance of features (rather than in predicting dv)

Partial Dependence of Women Experiencing Any Violence

Partial Dependence of Women Experiencing Any Type of Violence (any_v) on Features

- Behavioral factors dominate (parental violence, drinking, and control issues)

- Women's age and education seem to matter more than men's

- **While socioeconomic features are useful in targeting high risk audiences, behavioral features more directly drive dv.**

Recommendations

Mechanisms

- **TV**: "Abused" group hardest to reach, but most accessible via tv
- **Schools**: Should be leveraged, but need to start early, "Abused" group's partners average 6 yrs education

Messages

- Target less educated, lower income, more rural audiences
- Address associated behavioral issues in addition to domestic violence
- Convert 20% of partner's from "not abusive or antagonistic" to "advocates."
- Convert 50% of partners from "antagonistic but not abusive" to "not antagonistic."
- Convert 30% of partners from "abusive" to "not abusive."
- Convert 42% women from agreeing with justifications for beating. Affirm worth.
- Educate 20% of women who are abused but not reporting on how to access resources