Does Victim Gender Matter for Justice Delivery? Police and Judicial Responses to Women's Cases in India

Nirvika Jassal

Chuchu Wen & Priscila Stisman

April 3, 2025

Index

1. Introduction
2. Methods
3. Results
4. Extensions
5. Autopsy
6. Suggested Improvements

Introduction

- **Research Question:** Are women disadvantaged when accessing justice in India compared to men? and if women are discriminated against, is it because of their gender or due to the nature of their complaints?
- **Hypothesis:** Women in India face 'multi-stage' discrimination when accessing justice, encountering systemic barriers at sequential stages of the justice process, including police registration, investigation, trial, and verdict
- **Data:** 418,190 individual-level police reports in Haryana from January 2015 to November 2018 merged with court files

Methods

- **OLS Modeling:** to analyze whether women's cases are less likely than men's to be sent to court (+ VAW vs. non-VAW cases)
- **Topic Modeling**: Are there topics in the victims' testimonies that yield low convictions for suspects? Are topics influenced by gender or the classification of the crime as VAW?
- **Topical Inverse Regression Matching**

Structural Topic Models STMs

Why using it?

- Understand the topics in the victims testimonies
- It can help predict whether cases devoted to a topic (e.g. rape) are functions of covariates (e.g. gender)
- Disaggregates what citizens told the police happened to them using statistical associations between words.

Goals

- Give voice to victims by utilizing their own words
- Highlight the severity of claims, especially VAW
- Coarsen high-dimensional data to allow for text matching techniques

STMs with FREX measures

FREX (short for Frequency and Exclusivity) is a measure used in STMs to identify words that are both frequent within a topic and exclusive to that topic (highlight words that make a topic more distinguishable from others)

- High Probability Words: These words appear frequently in a topic but might also appear in other topics
- Exclusive Words: These words are more unique to a single topic but might not be the most frequent ones
- FREX Score: FREX balances frequency and exclusivity to identify words that are both common and distinctive for a topic

STM Steps

- **Step 1:** Data Preprocessing. Cleans and tokenizes the text and retains metadata, which contains additional information about each document (e.g., gender of the complainant, whether the person was convicted, etc.)
- **Step 2:** Preparing Documents for STM. Refine data, filtering out rare words that appear in fewer than 50 documents
- **Step 3:** Running the STM Model ($K = 32$ topics) in different subsets of the corpus
- **Step 4:** Assign Topic Labels

Corpus Subsets

Run STMs in different corpora and analyze topics within those

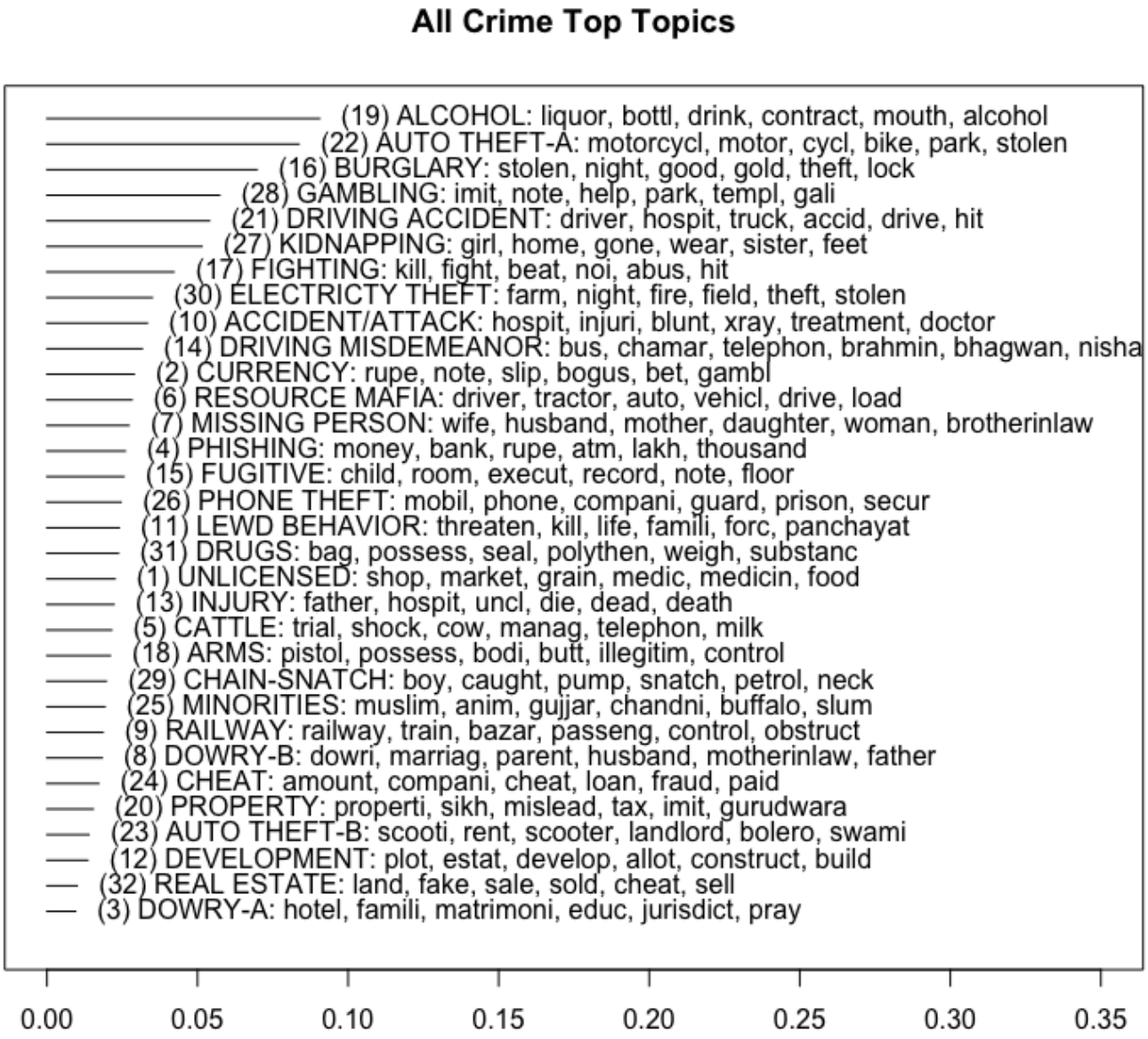
- All Crime
- Female Complainant
- VAW Reports

Top Topics Code Example

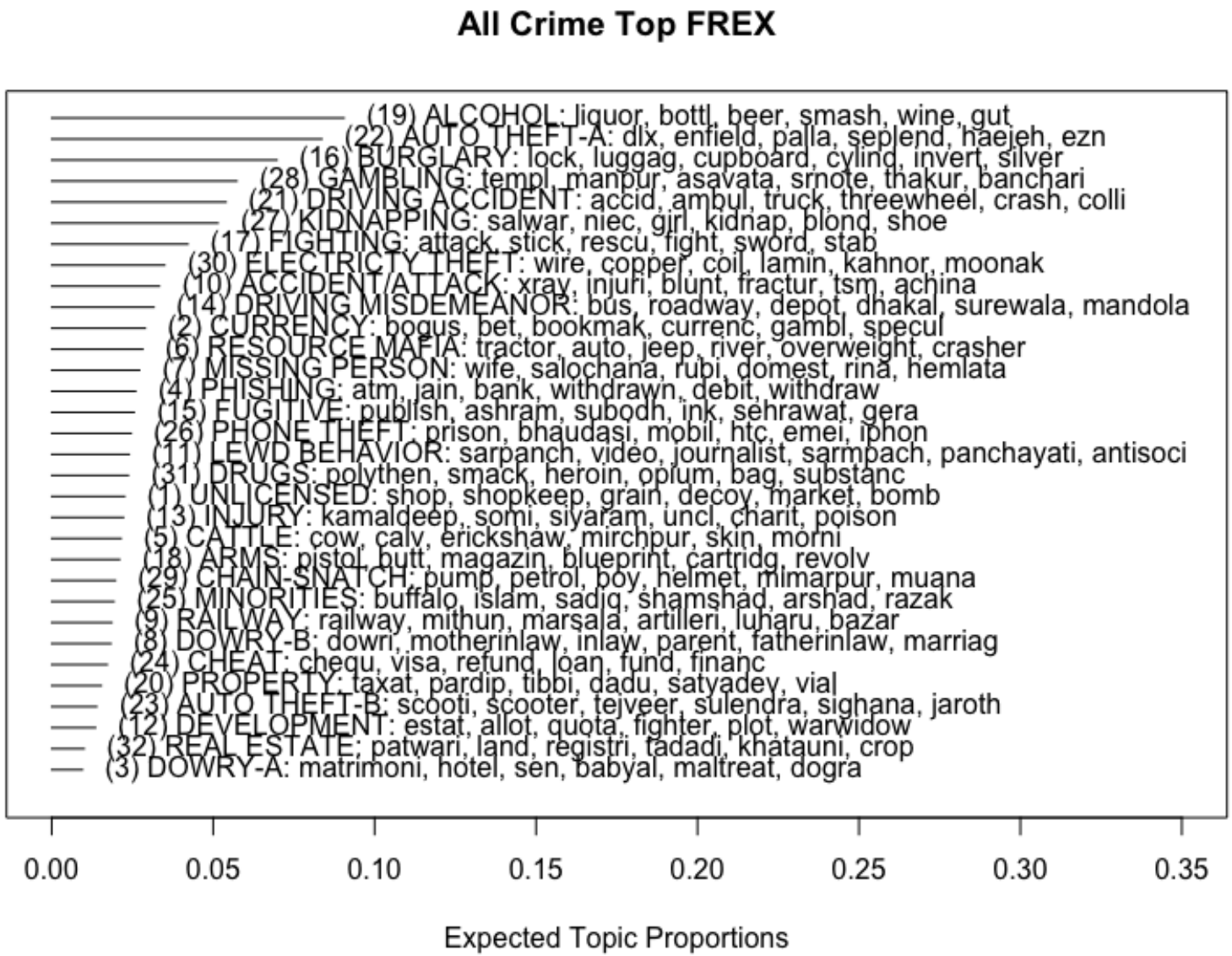
```
set.seed(23456)
processed <- textProcessor(documents = data$text, metadata = data)

out <- prepDocuments(documents = processed$documents,
                     vocab = processed$vocab,
                     meta = processed$meta, lower.thresh = 50)
docs <- out$documents
vocab <- out$vocab
meta <- out$meta
```

```
## Running Model (1)
test2 <- stm(documents = out$documents, vocab = out$vocab,
             K = 32,
             prevalence =~ complainant_gender + gendered + urban + convicted + acquitted + dismissed, # Mult
             max.em.its = 75,
             data = out$meta, init.type = "Spectral",
             verbose = TRUE)
```

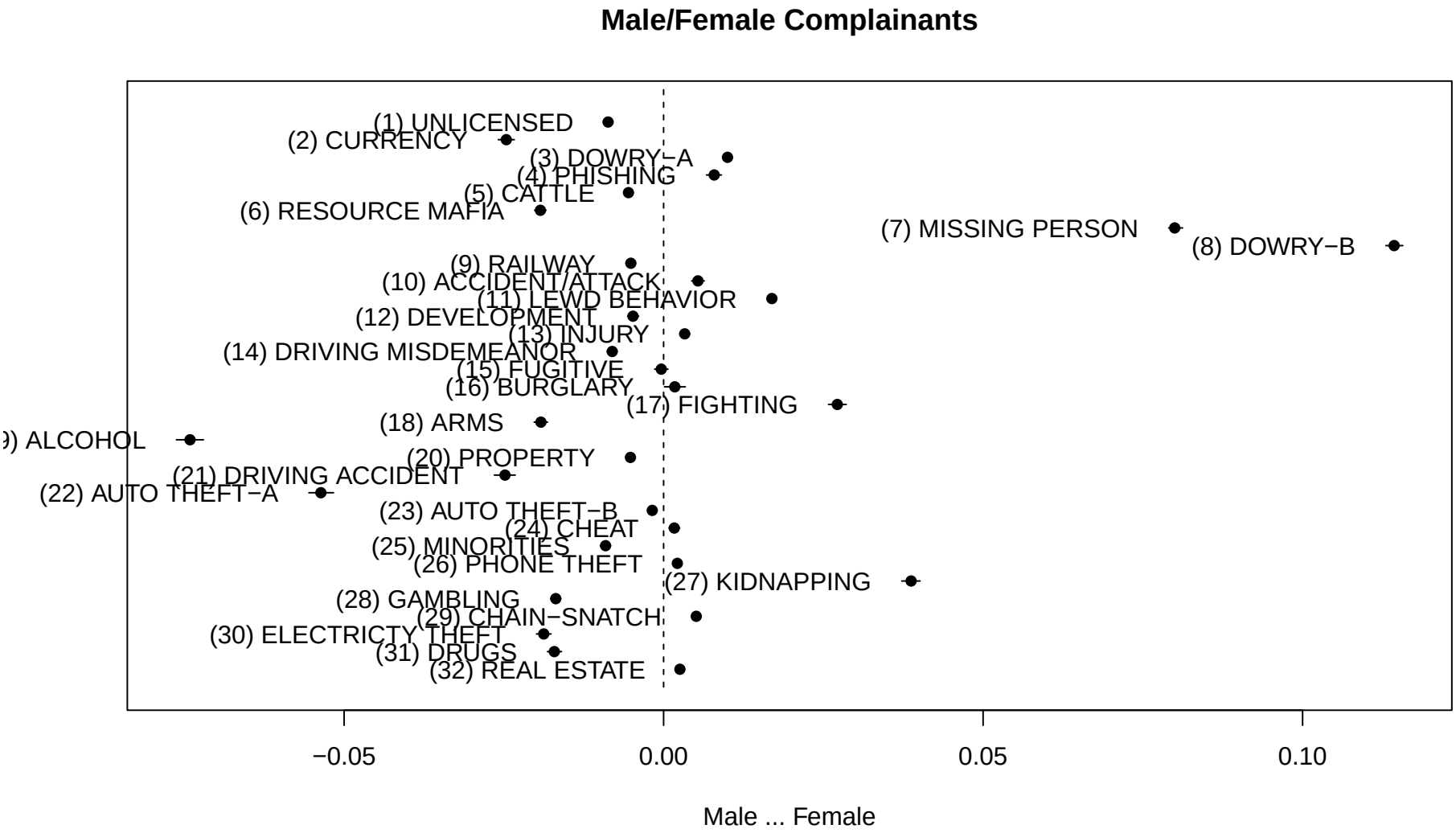
Results


All Crime Top Topics
(19) ALCOHOL: liquor, bottl, drink, contract, mouth, alcohol
(22) AUTO THEFT-A: motorcycl, motor, cycl, bike, park, stolen
(16) BURGLARY: stolen, night, good, gold, theft, lock
(28) GAMBLING: imit, note, help, park, templ, gali
(21) DRIVING ACCIDENT: driver, hospit, truck, accid, drive, hit
(27) KIDNAPPING: girl, home, gone, wear, sister, feet
(17) FIGHTING: kill, fight, beat, noi, abus, hit
(30) ELECTRICTY THEFT: farm, night, fire, field, theft, stolen
(10) ACCIDENT/ATTACK: hospit, injuri, blunt, xray, treatment, doctor
(14) DRIVING MISDEMEANOR: bus, chamar, telephon, brahmin, bhagwan, nisha
(2) CURRENCY: rupe, note, slip, bogus, bet, gambl
(6) RESOURCE MAFIA: driver, tractor, auto, vehicl, drive, load
(7) MISSING PERSON: wife, husband, mother, daughter, woman, brotherinlaw
(4) PHISHING: money, bank, rupe, atm, lakh, thousand
(15) FUGITIVE: child, room, execut, record, note, floor
(26) PHONE THEFT: mobil, phone, compani, guard, prison, secur
(11) LEWD BEHAVIOR: threaten, kill, life, famili, forc, panchayat
(31) DRUGS: bag, possess, seal, polythen, weigh, substanc
(1) UNLICENSED: shop, market, grain, medic, medicin, food
(13) INJURY: father, hospit, uncl, die, dead, death
(5) CATTLE: trial, shock, cow, manag, telephon, milk
(18) ARMS: pistol, possess, bodi, butt, illegitim, control
(29) CHAIN-SNATCH: boy, caught, pump, snatch, petrol, neck
(25) MINORITIES: muslim, anim, gujjar, chandni, buffalo, slum
(9) RAILWAY: railway, train, bazar, passeng, control, obstruct
(8) DOWRY-B: dowri, marriag, parent, husband, motherinlaw, father
(24) CHEAT: amount, compani, cheat, loan, fraud, paid
(20) PROPERTY: properti, sikh, mislead, tax, imit, gurudwara
(23) AUTO THEFT-B: scooti, rent, scooter, landlord, bolero, swami
(12) DEVELOPMENT: plot, estat, develop, allot, construct, build
(32) REAL ESTATE: land, fake, sale, sold, cheat, sell
(3) DOWRY-A: hotel, famili, matrimoni, educ, jurisdict, pray
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35

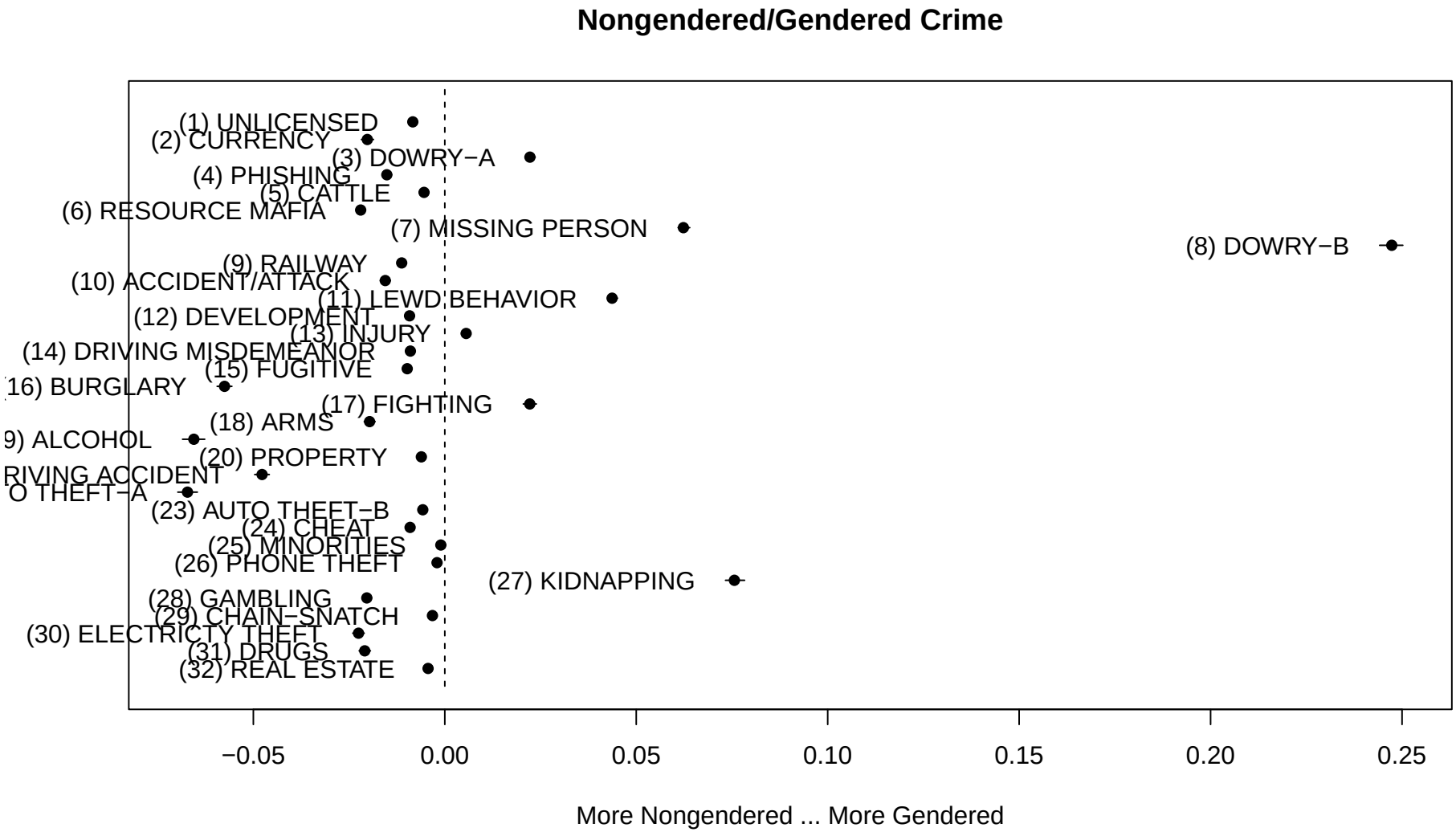
Results


All Crime Top FREX
(19) ALCOHOL: liquor, bottl, beer, smash, wine, gut
(22) AUTO THEFT-A: dlx, enfield, palla, seplend, haejeh, ezn
(16) BURGLARY: lock, luggag, cupboard, cylind, invert, silver
(28) GAMBLING: templ, manpur, asavata, srnote, thakur, banchari
(21) DRIVING ACCIDENT: accid, ambul, truck, threewheel, crash, colli
(27) KIDNAPPING: salwar, niec, girl, kidnap, blond, shoe
(17) FIGHTING: attack, stick, rescu, fight, sword, stab
(30) ELECTRICTY THEFT: wire, copper, coil, lamin, kahnor, moonak
(10) ACCIDENT/ATTACK: xray, injuri, blunt, fractur, tsm, achina
(14) DRIVING MISDEMEANOR: bus, roadway, depot, dhakal, surewala, mandola
(2) CURRENCY: bogus, bet, bookmak, currenc, gambl, specul
(6) RESOURCE MAFIA: tractor, auto, jeep, river, overweight, crasher
(7) MISSING PERSON: wife, salochana, rubi, domest, rina, hemlata
(4) PHISHING: atm, jain, bank, withdrawn, debit, withdraw
(15) FUGITIVE: publish, ashram, subodh, ink, sehrawat, gera
(26) PHONE THEFT: prison, bhaudasi, mobil, htc, emei, iphon
(11) LEWD BEHAVIOR: sarpanch, video, journalist, sarmpach, panchayati, antisoci
(31) DRUGS: polythen, smack, heroin, opium, bag, substanc
(1) UNLICENSED: shop, shopkeep, grain, decoy, market, bomb
(13) INJURY: kamaldeep, somi, siyaram, uncl, charit, poison
(5) CATTLE: cow, calv, erickshaw, mirchpur, skin, morni
(18) ARMS: pistol, butt, magazin, blueprint, cartridg, revolv
(29) CHAIN-SNATCH: pump, petrol, boy, helmet, mimarpur, muana
(25) MINORITIES: buffalo, islam, sadiq, shamshad, arshad, razak
(9) RAILWAY: railway, mithun, marsala, artilleri, luharu, bazar
(8) DOWRY-B: dowri, motherinlaw, inlaw, parent, fatherinlaw, marriag
(24) CHEAT: chequ, visa, refund, loan, fund, financ
(20) PROPERTY: taxat, pardip, tibbi, dadu, satyadev, vial
(23) AUTO THEFT-B: scooti, scooter, tejveer, sulendra, sighana, jaroth
(12) DEVELOPMENT: estat, allot, quota, fighter, plot, warwidow
(32) REAL ESTATE: patwari, land, registri, tadadi, khatauni, crop
(3) DOWRY-A: matrimoni, hotel, sen, babyal, maltreat, dogra
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
Expected Topic Proportions

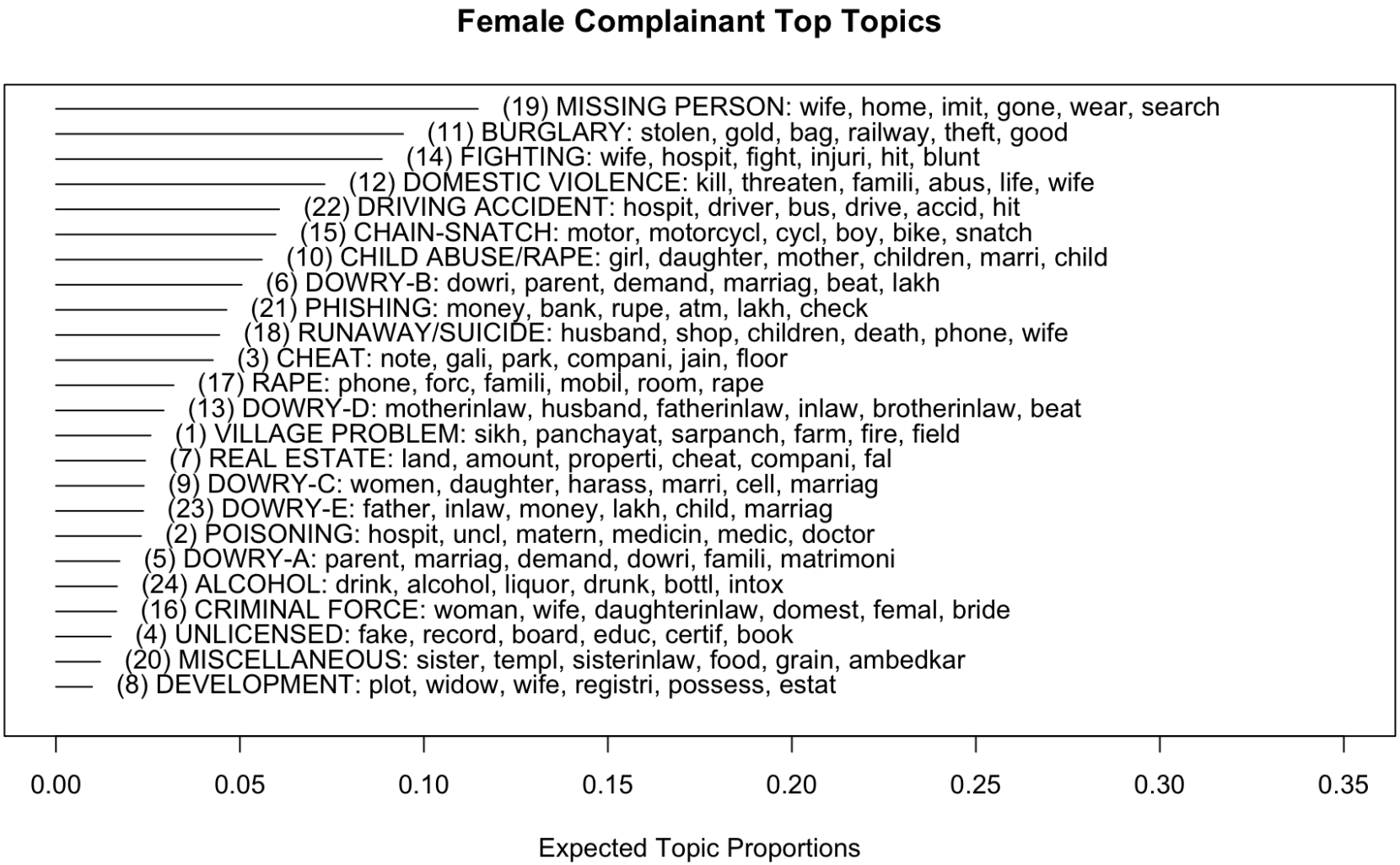
Results


Male/Female Complainants
(1) UNLICENSED
(2) CURRENCY
(3) DOWRY−A
(4) PHISHING
(5) CATTLE
(6) RESOURCE MAFIA
(7) MISSING PERSON
(8) DOWRY−B
(9) RAILWAY
(10) ACCIDENT/ATTACK
(11) LEWD BEHAVIOR
(12) DEVELOPMENT
(13) INJURY
(14) DRIVING MISDEMEANOR
(15) FUGITIVE
(16) BURGLARY
(17) FIGHTING
(18) ARMS
) ALCOHOL
(20) PROPERTY
(21) DRIVING ACCIDENT
(22) AUTO THEFT−A
(23) AUTO THEFT−B
(24) CHEAT
(25) MINORITIES
(26) PHONE THEFT
(27) KIDNAPPING
(28) GAMBLING
(29) CHAIN−SNATCH
(30) ELECTRICTY THEFT
(31) DRUGS
(32) REAL ESTATE
−0.05
0.00
0.05
0.10
Male ... Female

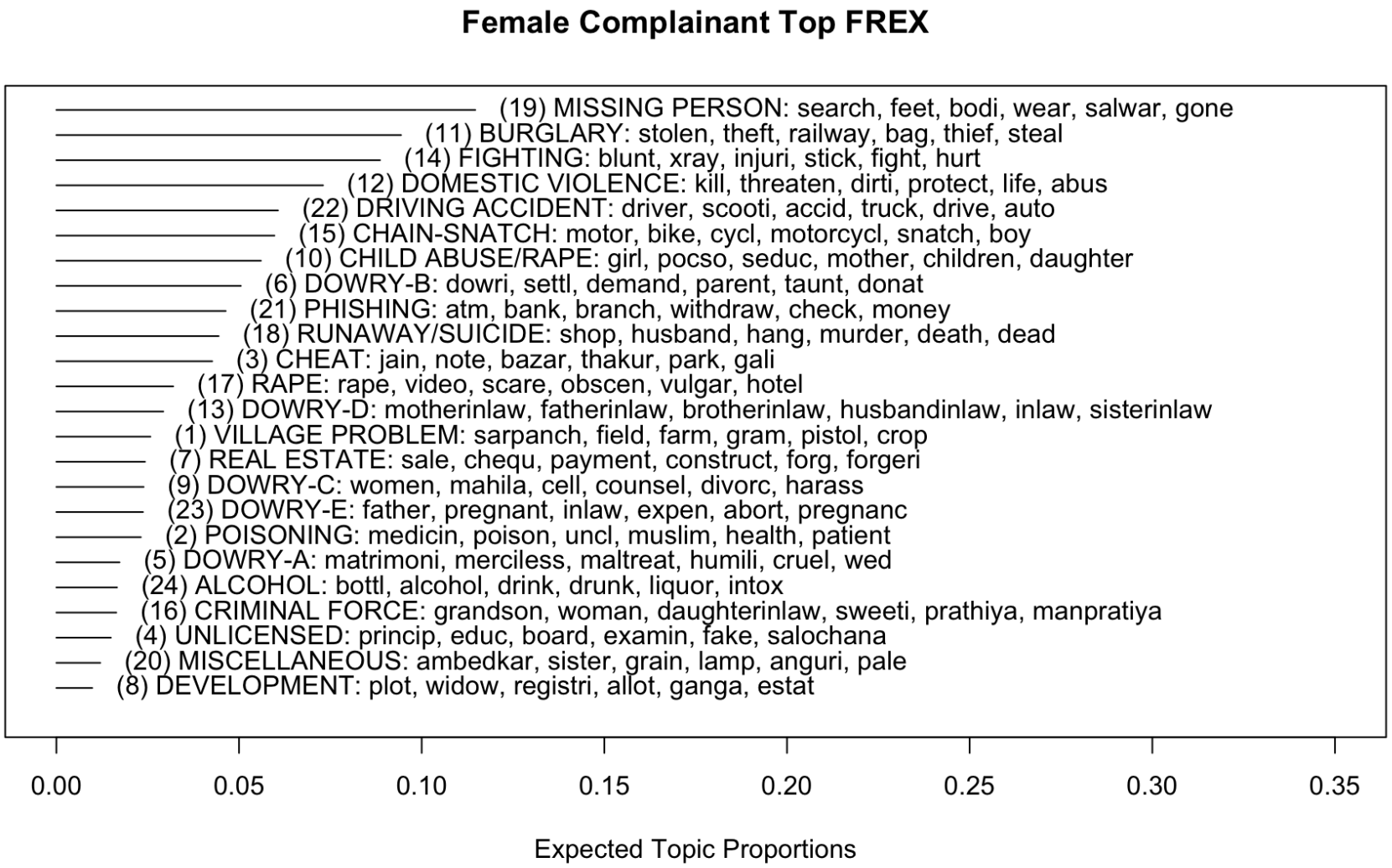
Results


Nongendered/Gendered Crime
(1) UNLICENSED
(2) CURRENCY
(3) DOWRY−A
(4) PHISHING
(5) CATTLE
(6) RESOURCE MAFIA
(7) MISSING PERSON
(8) DOWRY−B
(9) RAILWAY
(10) ACCIDENT/ATTACK
(11) LEWD BEHAVIOR
(12) DEVELOPMENT
(13) INJURY
(14) DRIVING MISDEMEANOR
(15) FUGITIVE
16) BURGLARY
(17) FIGHTING
(18) ARMS
9) ALCOHOL
(20) PROPERTY
RIVING ACCIDENT
O THEFT−A
(23) AUTO THEFT−B
(24) CHEAT
(25) MINORITIES
(26) PHONE THEFT
(27) KIDNAPPING
(28) GAMBLING
(29) CHAIN−SNATCH
(30) ELECTRICTY THEFT
(31) DRUGS
(32) REAL ESTATE
−0.05
0.00
0.05
0.10
0.15
0.20
0.25
More Nongendered ... More Gendered

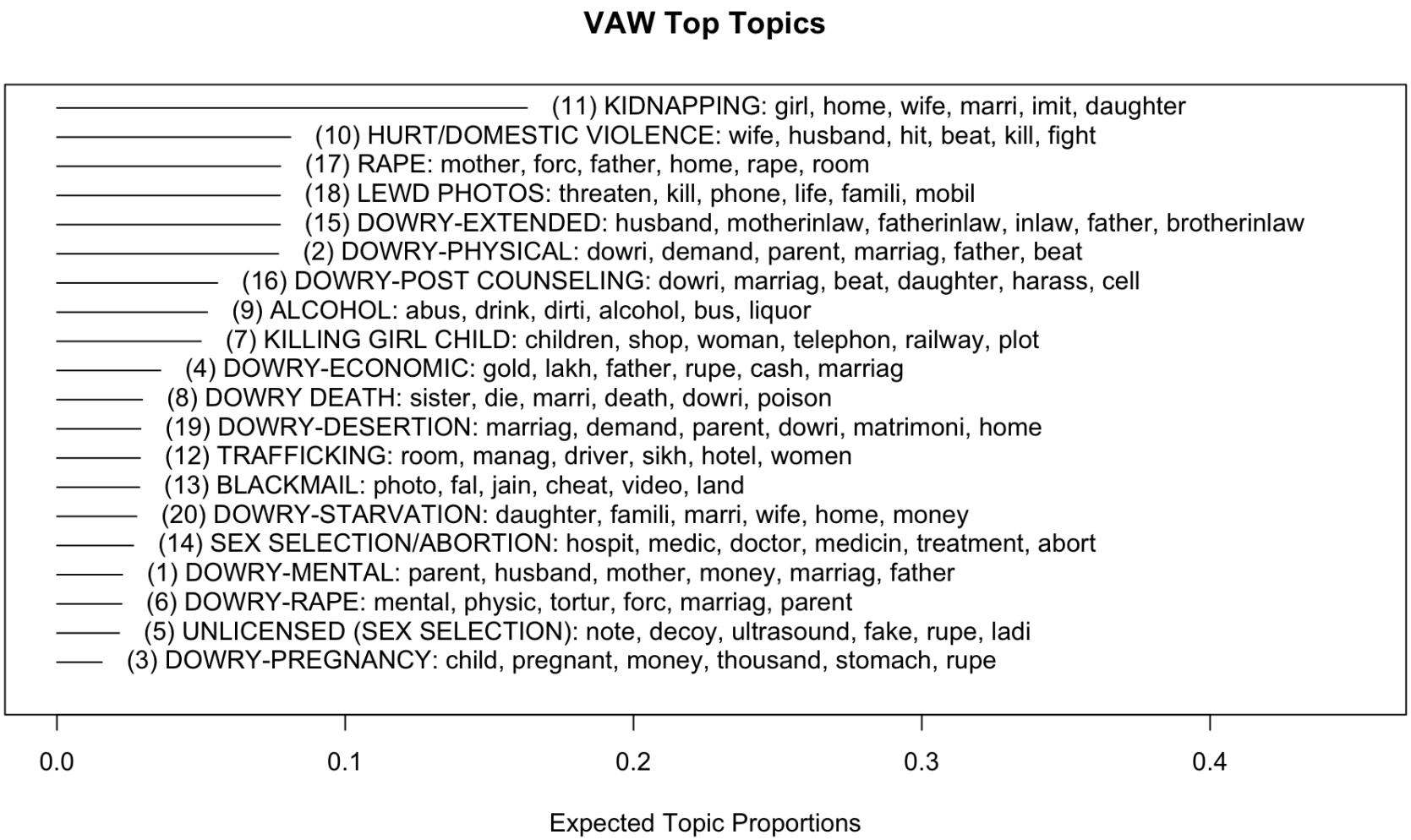
Results


Female Complainant Top Topics
(19) MISSING PERSON: wife, home, imit, gone, wear, search
(11) BURGLARY: stolen, gold, bag, railway, theft, good
(14) FIGHTING: wife, hospit, fight, injuri, hit, blunt
(12) DOMESTIC VIOLENCE: kill, threaten, famili, abus, life, wife
(22) DRIVING ACCIDENT: hospit, driver, bus, drive, accid, hit
(15) CHAIN-SNATCH: motor, motorcycl, cycl, boy, bike, snatch
(10) CHILD ABUSE/RAPE: girl, daughter, mother, children, marri, child
(6) DOWRY-B: dowri, parent, demand, marriag, beat, lakh
(21) PHISHING: money, bank, rupe, atm, lakh, check
(18) RUNAWAY/SUICIDE: husband, shop, children, death, phone, wife
(3) CHEAT: note, gali, park, compani, jain, floor
(17) RAPE: phone, forc, famili, mobil, room, rape
(13) DOWRY-D: motherinlaw, husband, fatherinlaw, inlaw, brotherinlaw, beat
(1) VILLAGE PROBLEM: sikh, panchayat, sarpanch, farm, fire, field
(7) REAL ESTATE: land, amount, properti, cheat, compani, fal
(9) DOWRY-C: women, daughter, harass, marri, cell, marriag
(23) DOWRY-E: father, inlaw, money, lakh, child, marriag
(2) POISONING: hospit, uncl, matern, medicin, medic, doctor
(5) DOWRY-A: parent, marriag, demand, dowri, famili, matrimoni
(24) ALCOHOL: drink, alcohol, liquor, drunk, bottl, intox
(16) CRIMINAL FORCE: woman, wife, daughterinlaw, domest, femal, bride
(4) UNLICENSED: fake, record, board, educ, certif, book
(20) MISCELLANEOUS: sister, templ, sisterinlaw, food, grain, ambedkar
(8) DEVELOPMENT: plot, widow, wife, registri, possess, estat
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
Expected Topic Proportions

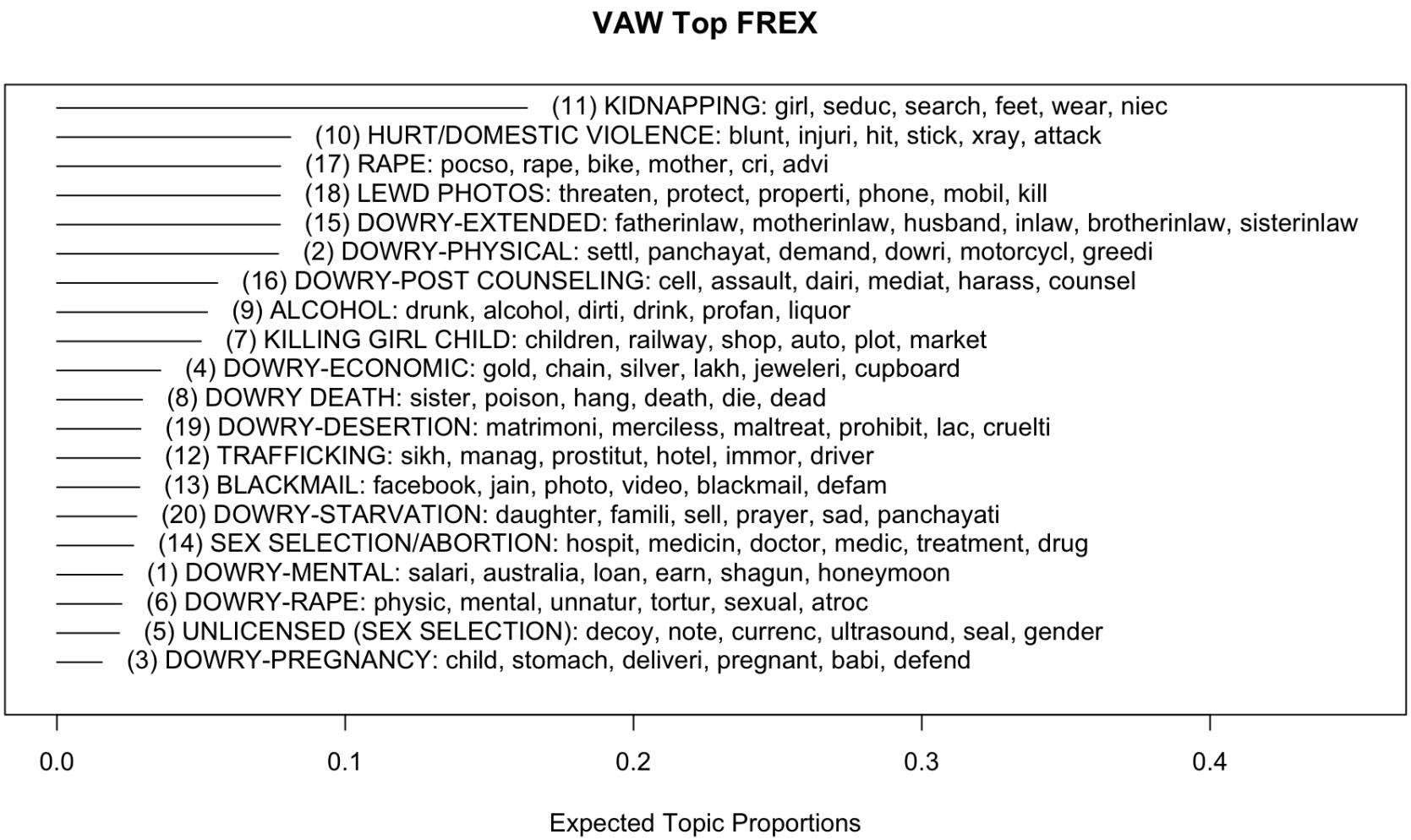
Results


Female Complainant Top FREX
(19) MISSING PERSON: search, feet, bodi, wear, salwar, gone
(11) BURGLARY: stolen, theft, railway, bag, thief, steal
(14) FIGHTING: blunt, xray, injuri, stick, fight, hurt
(12) DOMESTIC VIOLENCE: kill, threaten, dirti, protect, life, abus
(22) DRIVING ACCIDENT: driver, scooti, accid, truck, drive, auto
(15) CHAIN-SNATCH: motor, bike, cycl, motorcycl, snatch, boy
(10) CHILD ABUSE/RAPE: girl, pocso, seduc, mother, children, daughter
(6) DOWRY-B: dowri, settl, demand, parent, taunt, donat
(21) PHISHING: atm, bank, branch, withdraw, check, money
(18) RUNAWAY/SUICIDE: shop, husband, hang, murder, death, dead
(3) CHEAT: jain, note, bazar, thakur, park, gali
(17) RAPE: rape, video, scare, obscen, vulgar, hotel
(13) DOWRY-D: motherinlaw, fatherinlaw, brotherinlaw, husbandinlaw, inlaw, sisterinlaw
(1) VILLAGE PROBLEM: sarpanch, field, farm, gram, pistol, crop
(7) REAL ESTATE: sale, chequ, payment, construct, forg, forgeri
(9) DOWRY-C: women, mahila, cell, counsel, divorc, harass
(23) DOWRY-E: father, pregnant, inlaw, expen, abort, pregnanc
(2) POISONING: medicin, poison, uncl, muslim, health, patient
(5) DOWRY-A: matrimoni, merciless, maltreat, humili, cruel, wed
(24) ALCOHOL: bottl, alcohol, drink, drunk, liquor, intox
(16) CRIMINAL FORCE: grandson, woman, daughterinlaw, sweeti, prathiya, manpratiya
(4) UNLICENSED: princip, educ, board, examin, fake, salochana
(20) MISCELLANEOUS: ambedkar, sister, grain, lamp, anguri, pale
(8) DEVELOPMENT: plot, widow, registri, allot, ganga, estat
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
Expected Topic Proportions

Results


VAW Top Topics
(11) KIDNAPPING: girl, home, wife, marri, imit, daughter
(10) HURT/DOMESTIC VIOLENCE: wife, husband, hit, beat, kill, fight
(17) RAPE: mother, forc, father, home, rape, room
(18) LEWD PHOTOS: threaten, kill, phone, life, famili, mobil
(15) DOWRY-EXTENDED: husband, motherinlaw, fatherinlaw, inlaw, father, brotherinlaw
(2) DOWRY-PHYSICAL: dowri, demand, parent, marriag, father, beat
(16) DOWRY-POST COUNSELING: dowri, marriag, beat, daughter, harass, cell
(9) ALCOHOL: abus, drink, dirti, alcohol, bus, liquor
(7) KILLING GIRL CHILD: children, shop, woman, telephon, railway, plot
(4) DOWRY-ECONOMIC: gold, lakh, father, rupe, cash, marriag
(8) DOWRY DEATH: sister, die, marri, death, dowri, poison
(19) DOWRY-DESERTION: marriag, demand, parent, dowri, matrimoni, home
(12) TRAFFICKING: room, manag, driver, sikh, hotel, women
(13) BLACKMAIL: photo, fal, jain, cheat, video, land
(20) DOWRY-STARVATION: daughter, famili, marri, wife, home, money
(14) SEX SELECTION/ABORTION: hospit, medic, doctor, medicin, treatment, abort
(1) DOWRY-MENTAL: parent, husband, mother, money, marriag, father
(6) DOWRY-RAPE: mental, physic, tortur, forc, marriag, parent
(5) UNLICENSED (SEX SELECTION): note, decoy, ultrasound, fake, rupe, ladi
(3) DOWRY-PREGNANCY: child, pregnant, money, thousand, stomach, rupe
0.0
0.1
0.2
0.3
0.4
Expected Topic Proportions

Results

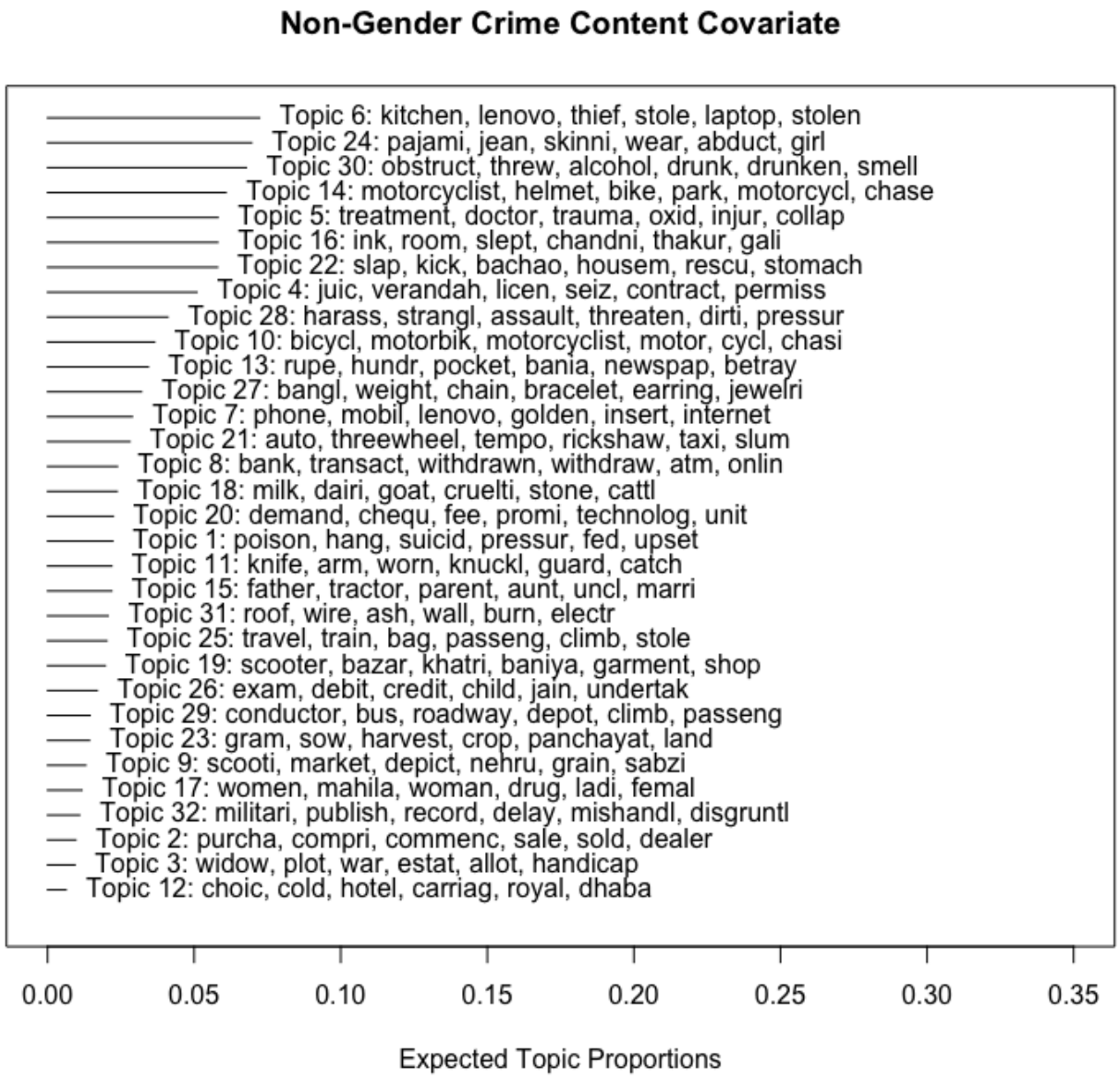

VAW Top FREX
(11) KIDNAPPING: girl, seduc, search, feet, wear, niec
(10) HURT/DOMESTIC VIOLENCE: blunt, injuri, hit, stick, xray, attack
(17) RAPE: pocso, rape, bike, mother, cri, advi
(18) LEWD PHOTOS: threaten, protect, properti, phone, mobil, kill
(15) DOWRY-EXTENDED: fatherinlaw, motherinlaw, husband, inlaw, brotherinlaw, sisterinlaw
(2) DOWRY-PHYSICAL: settl, panchayat, demand, dowri, motorcycl, greedi
(16) DOWRY-POST COUNSELING: cell, assault, dairi, mediat, harass, counsel
(9) ALCOHOL: drunk, alcohol, dirti, drink, profan, liquor
(7) KILLING GIRL CHILD: children, railway, shop, auto, plot, market
(4) DOWRY-ECONOMIC: gold, chain, silver, lakh, jeweleri, cupboard
(8) DOWRY DEATH: sister, poison, hang, death, die, dead
(19) DOWRY-DESERTION: matrimoni, merciless, maltreat, prohibit, lac, cruelti
(12) TRAFFICKING: sikh, manag, prostitut, hotel, immor, driver
(13) BLACKMAIL: facebook, jain, photo, video, blackmail, defam
(20) DOWRY-STARVATION: daughter, famili, sell, prayer, sad, panchayati
(14) SEX SELECTION/ABORTION: hospit, medicin, doctor, medic, treatment, drug
(1) DOWRY-MENTAL: salari, australia, loan, earn, shagun, honeymoon
(6) DOWRY-RAPE: physic, mental, unnatur, tortur, sexual, atroc
(5) UNLICENSED (SEX SELECTION): decoy, note, currenc, ultrasound, seal, gender
(3) DOWRY-PREGNANCY: child, stomach, deliveri, pregnant, babi, defend
0.0
0.1
0.2
0.3
0.4
Expected Topic Proportions

Covariates Code

```r
set.seed(23456)
processed <- textProcessor(documents = data$text, metadata = data)

out <- prepDocuments(documents = processed$documents,
                     vocab = processed$vocab,
                     meta = processed$meta, lower.thresh = 50)
docs <- out$documents
vocab <- out$vocab
meta <- out$meta
```

```r
stm.out.c <- stm(
  out$documents,      # The processed documents
  out$vocab,          # The vocabulary of the documents
  K=K,                # Number of topics (32 in this case)
  prevalence=~female_complainant,    # Covariate affecting topic prevalence
  content=~female_complainant,       # Covariate affecting topic content
  data=out$meta,      # Metadata associated with the documents
  max.em.its=25,      # Maximum number of EM (Expectation-Maximization) iterations
  seed=1033311        # Random seed for reproducibility
)
```

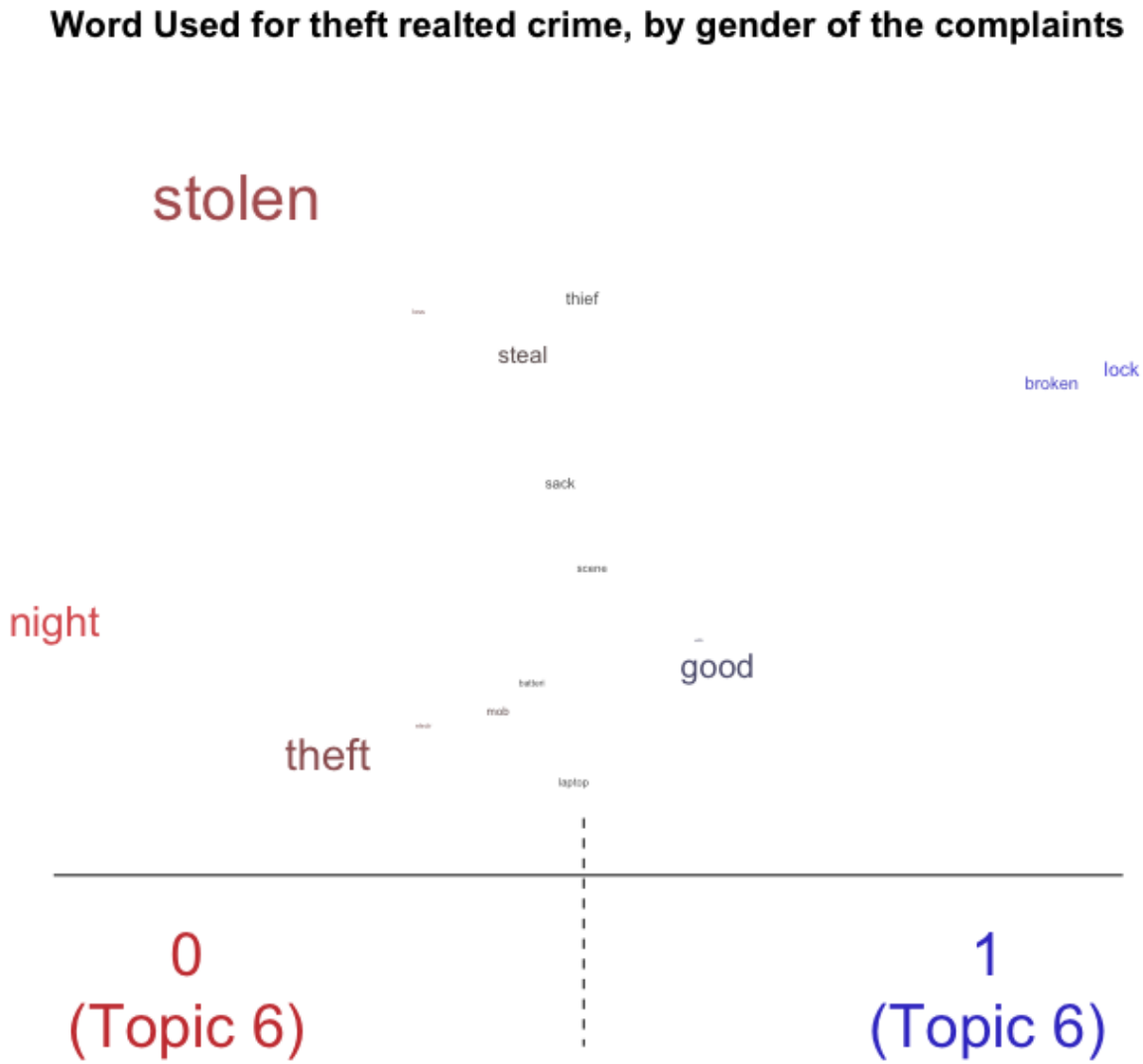
Results


Non-Gender Crime Content Covariate
Topic 6: kitchen, lenovo, thief, stole, laptop, stolen
Topic 24: pajami, jean, skinni, wear, abduct, girl
Topic 30: obstruct, threw, alcohol, drunk, drunken, smell
Topic 14: motorcyclist, helmet, bike, park, motorcycl, chase
Topic 5: treatment, doctor, trauma, oxid, injur, collap
Topic 16: ink, room, slept, chandni, thakur, gali
Topic 22: slap, kick, bachao, housem, rescu, stomach
Topic 4: juic, verandah, licen, seiz, contract, permiss
Topic 28: harass, strangl, assault, threaten, dirti, pressur
Topic 10: bicycl, motorbik, motorcyclist, motor, cycl, chasi
Topic 13: rupe, hundr, pocket, bania, newspap, betray
Topic 27: bangl, weight, chain, bracelet, earring, jewelri
Topic 7: phone, mobil, lenovo, golden, insert, internet
Topic 21: auto, threewheel, tempo, rickshaw, taxi, slum
Topic 8: bank, transact, withdrawn, withdraw, atm, onlin
Topic 18: milk, dairi, goat, cruelti, stone, cattl
Topic 20: demand, chequ, fee, promi, technolog, unit
Topic 1: poison, hang, suicid, pressur, fed, upset
Topic 11: knife, arm, worn, knuckl, guard, catch
Topic 15: father, tractor, parent, aunt, uncl, marri
Topic 31: roof, wire, ash, wall, burn, electr
Topic 25: travel, train, bag, passeng, climb, stole
Topic 19: scooter, bazar, khatri, baniya, garment, shop
Topic 26: exam, debit, credit, child, jain, undertak
Topic 29: conductor, bus, roadway, depot, climb, passeng
Topic 23: gram, sow, harvest, crop, panchayat, land
Topic 9: scooti, market, depict, nehru, grain, sabzi
Topic 17: women, mahila, woman, drug, ladi, femal
Topic 32: militari, publish, record, delay, mishandl, disgruntl
Topic 2: purcha, compri, commenc, sale, sold, dealer
Topic 3: widow, plot, war, estat, allot, handicap
Topic 12: choic, cold, hotel, carriag, royal, dhaba
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
Expected Topic Proportions

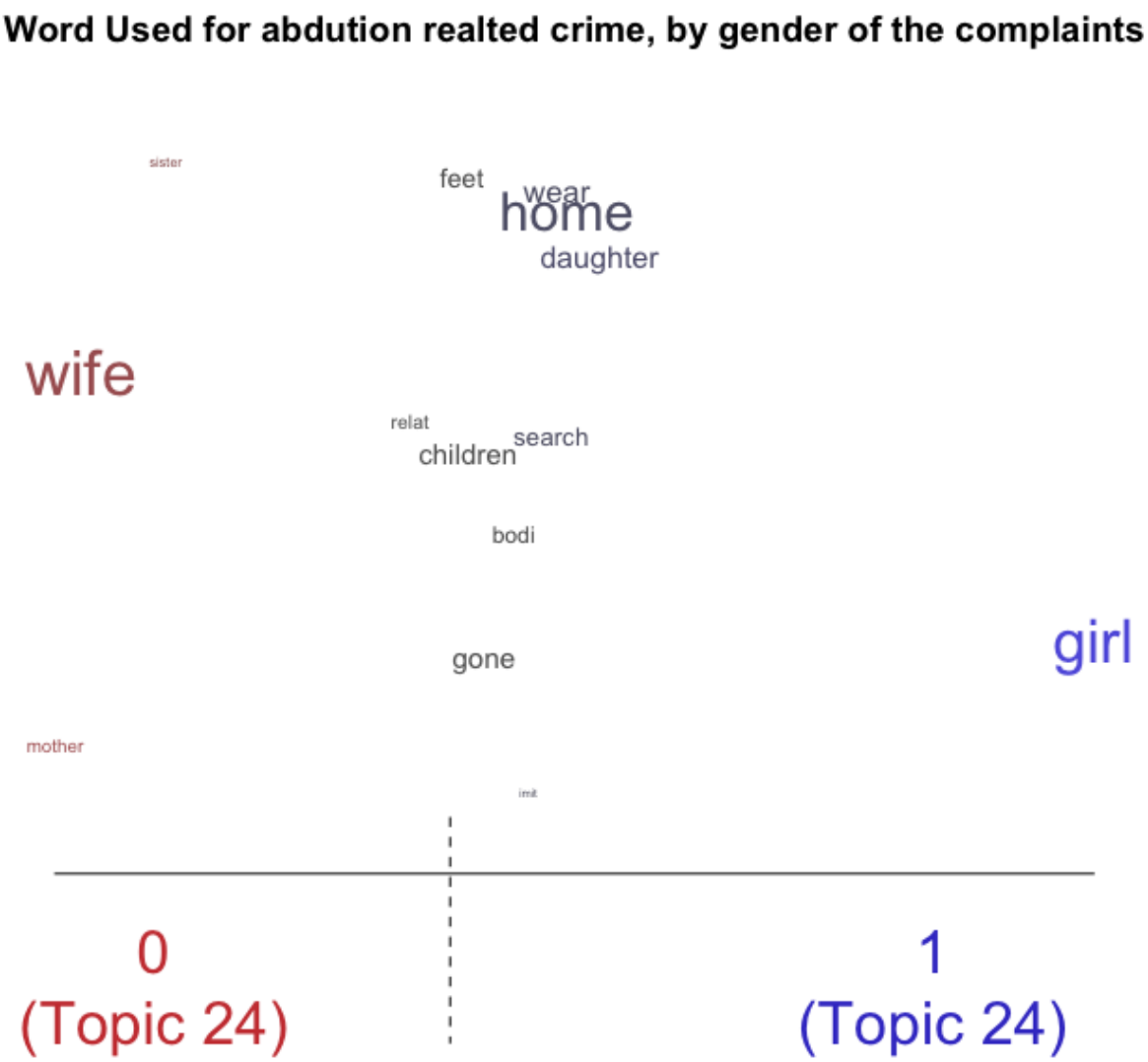
Extensions

- We analyze the most frequently used words by men and women within the same topic in police reports.
- Rationale: Do men and women use certain words differently or with varying frequency when reporting a criminal case?
- For example, in alcohol-related incidents, what words are predominantly used by men, and which ones are more common among women?
- We perform this exercise in three topics in non-gendered police reports

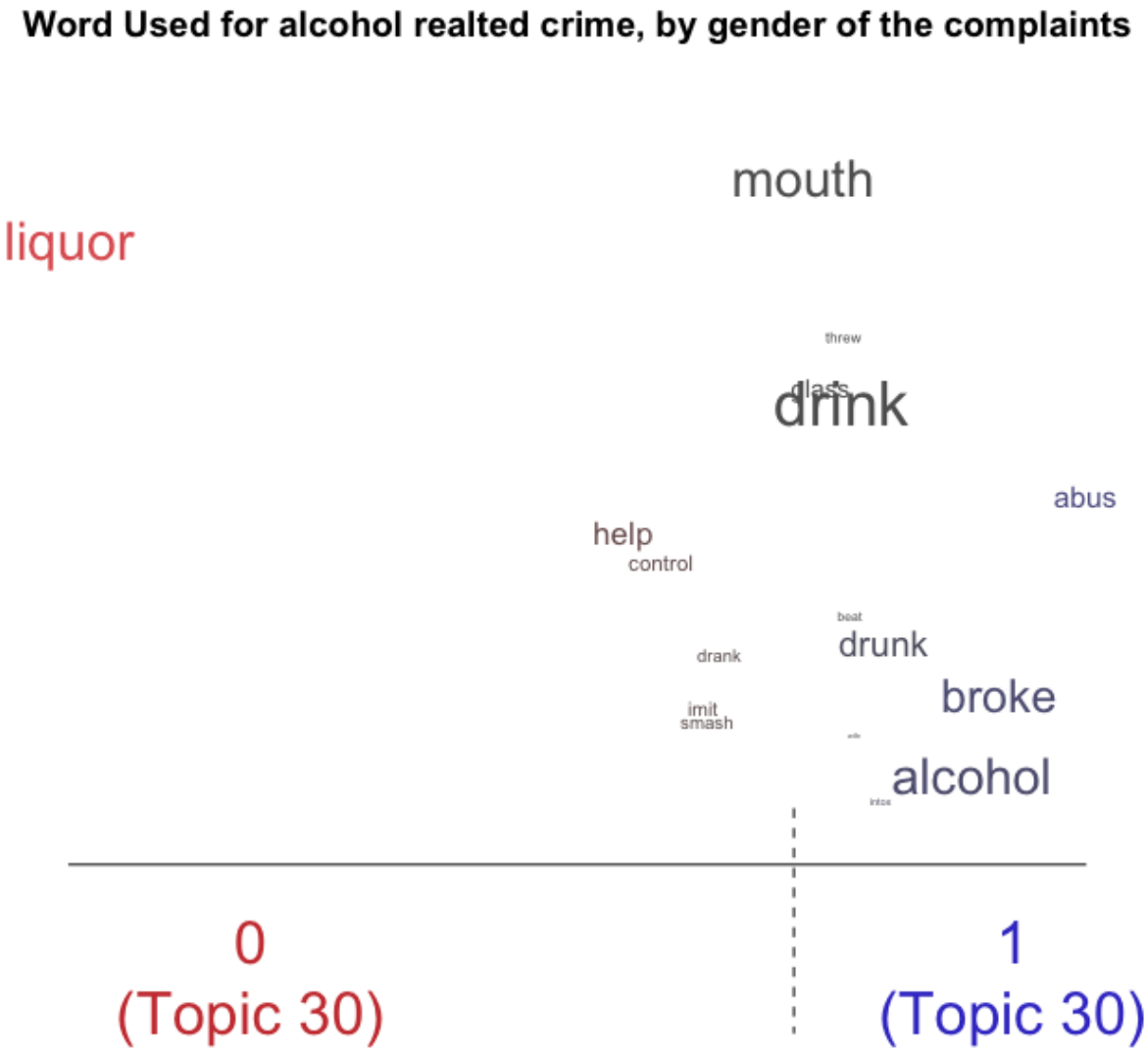
Extensions Results



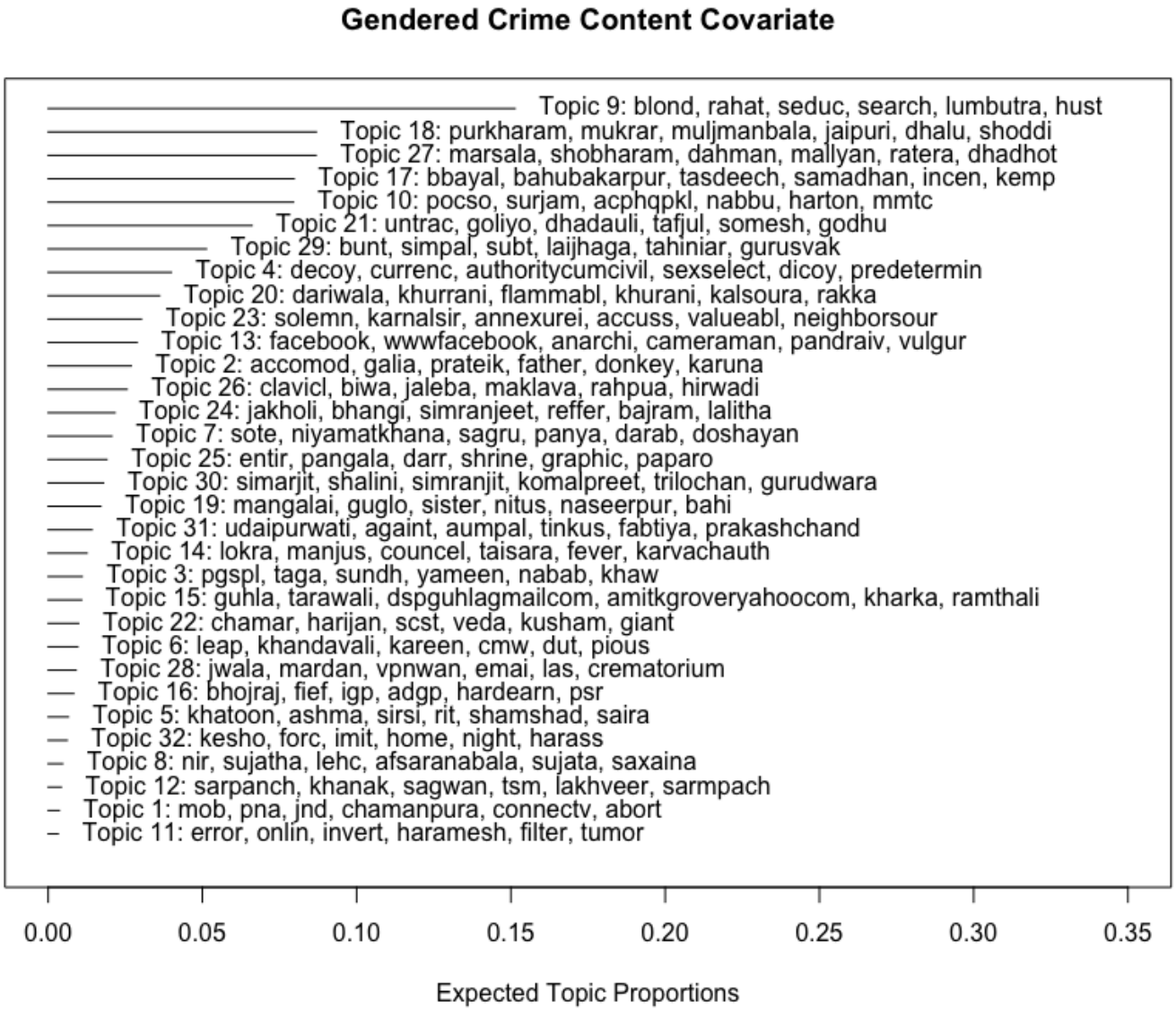
Extensions Results



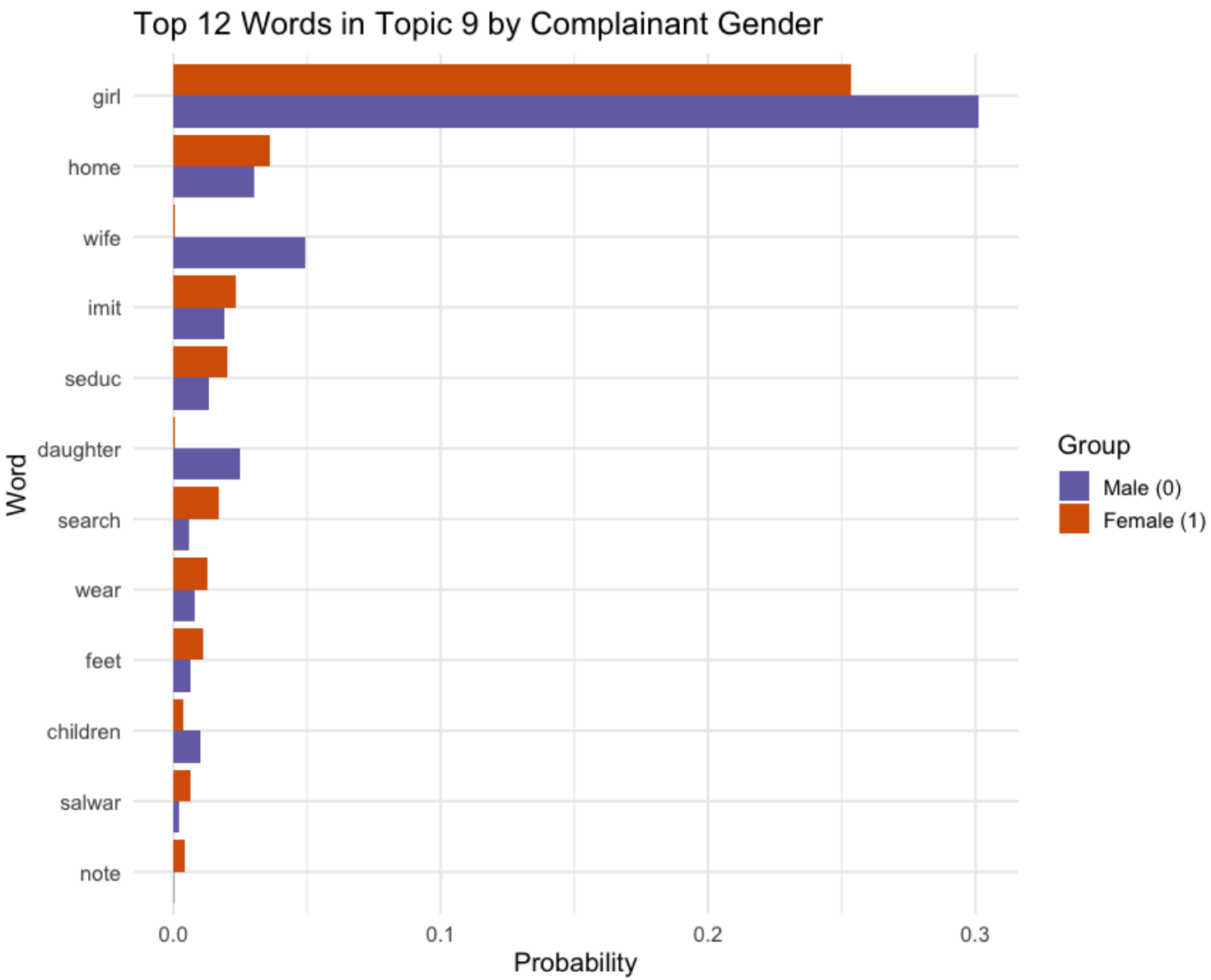
Extensions Results



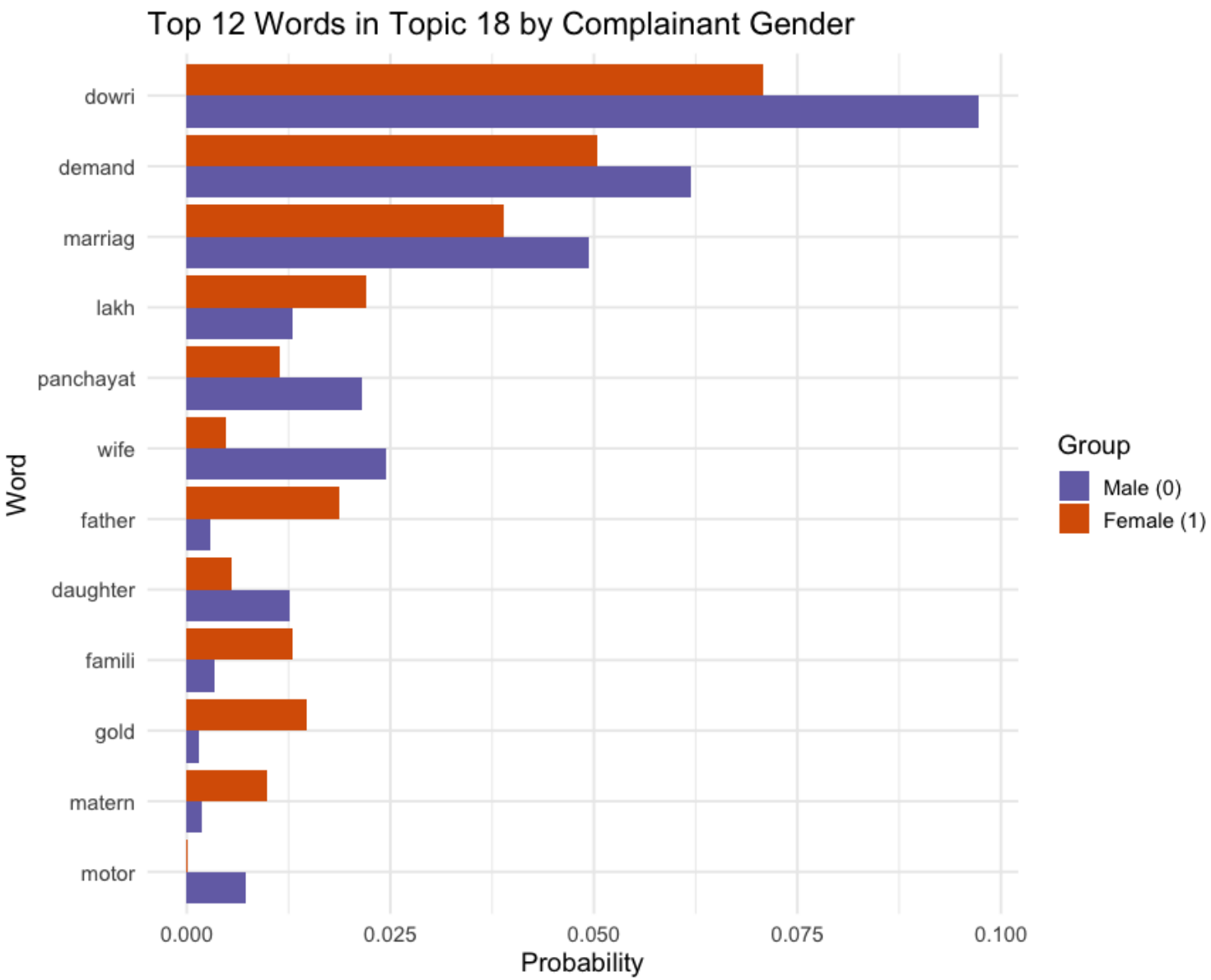
Extensions Results



Extensions Results



Extensions Results



Autopsy

- Highly satisfied with our STM results
- All STM results were successfully replicated, producing identical outcomes since we used the same seed
- We included the STM for all types of crimes, an analysis presented in the author's code but not in the paper
- Our extensions were straightforward to implement and provided valuable insights
- Great learning experience with STMs, FREX and covariate analysis (metadata, prevalence, content, etc)

Suggested Improvements

- It is unclear how and where the translation from Hindi to English occurs in the author's code
- Part of the corpus is not translated, which makes challenging to easily interpret results
- Lots of typos in the corpus