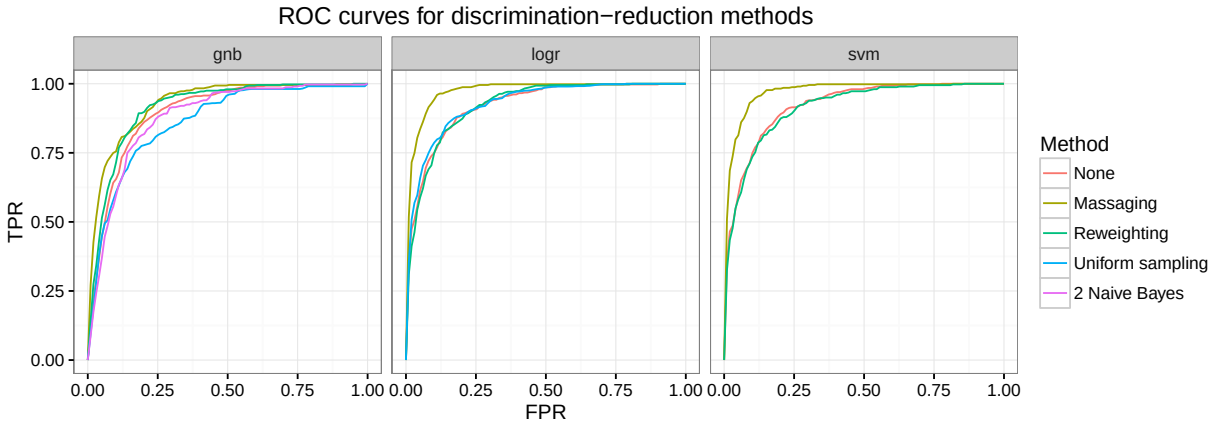
# analysis

*Selin Jessa*

*April 16, 2017*

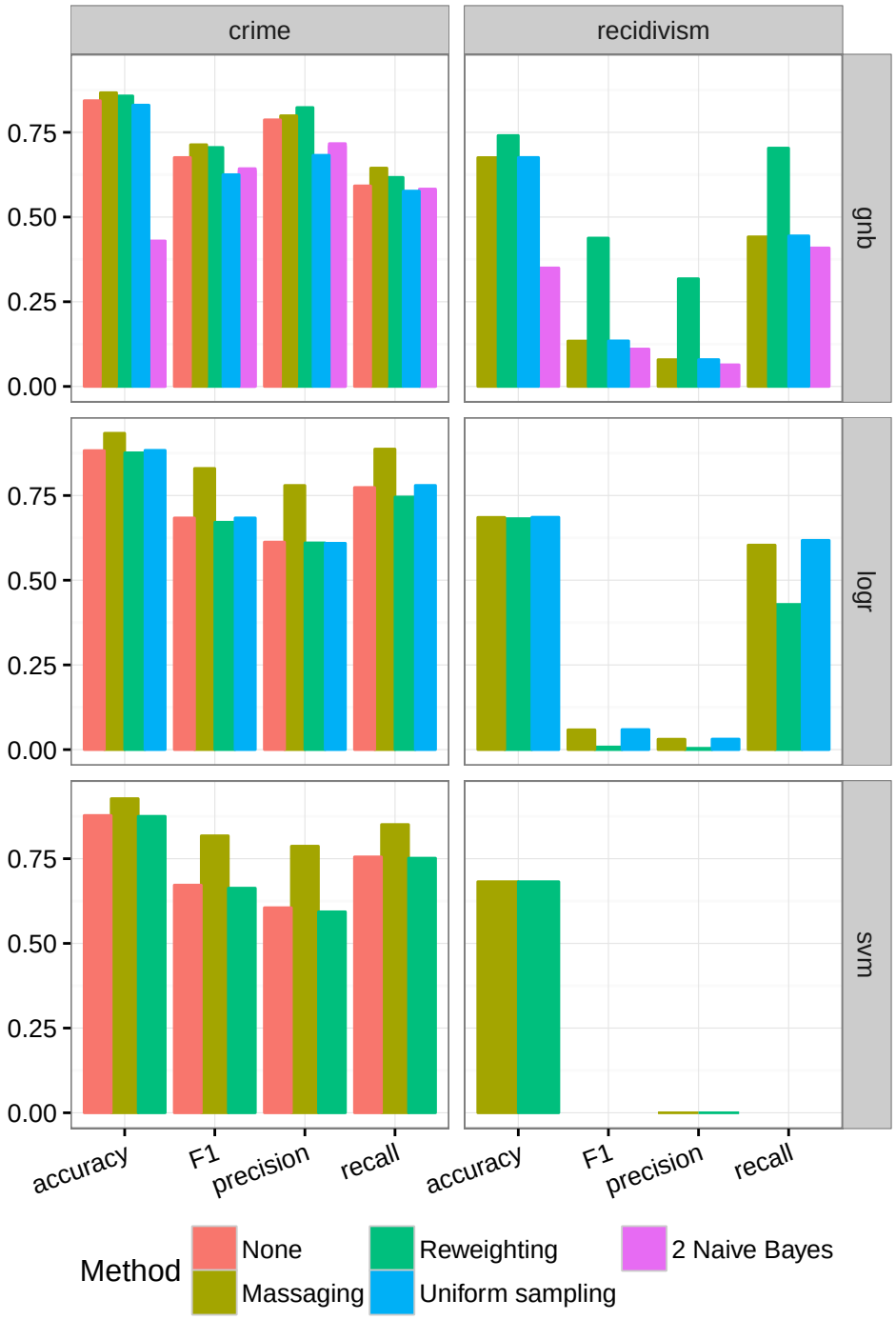
## ROC curves

```
roc %>% ggplot(aes(x = fpr, y = tpr, colour = method)) +
  geom_line() +
  facet_wrap(~ model) +
  xlab("FPR") + ylab("TPR") +
  ggtitle("ROC curves for discrimination-reduction methods") +
  guides(colour = guide_legend(title = "Method")) +
  theme_bw()
```




## Accuracy

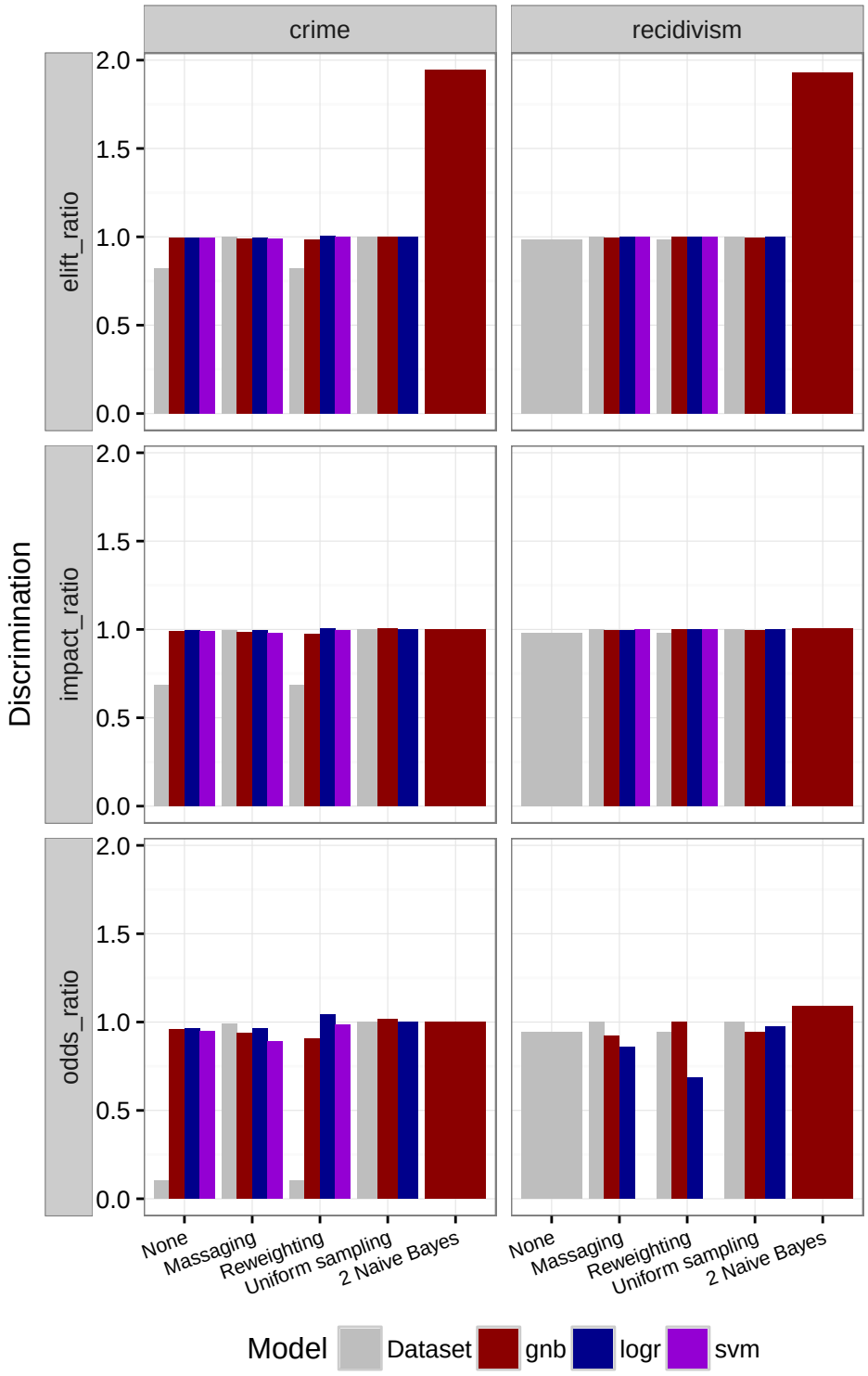
```
acc %>% gather(statistic, value, accuracy, precision, recall, F1) %>%
  ggplot(aes(x = statistic, y = value)) +
  geom_bar(stat = "identity", aes(fill = method, colour = method), position = "dodge") +
  xlab(NULL) + ylab(NULL) +
  guides(fill = guide_legend(title = "Method", nrow = 2), colour = guide_legend(title = "Method")) +
  facet_grid(model ~ dataset) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 20, hjust = 1),
        legend.position = "bottom")
```




## Discrimination

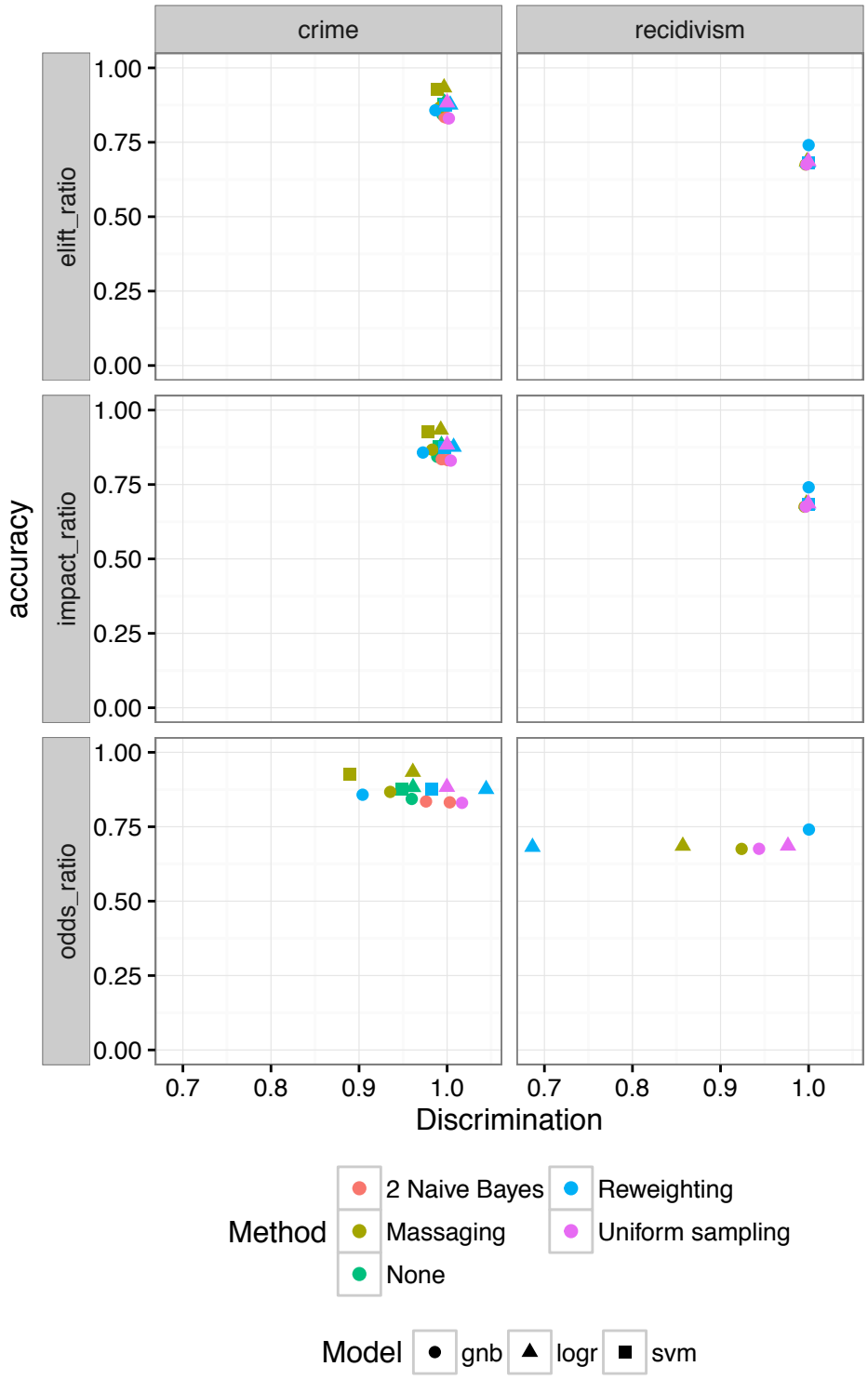
```
comp_tidy %>% gather(measure, value, impact_ratio, odds_ratio, elift_ratio) %>%
  ggplot(aes(x = method, y = value)) +
  geom_bar(stat = "identity", aes(fill = model), position = "dodge") +
```

```
  guides(fill = guide_legend(title = "Model")) +
  scale_fill_manual(values = c("gray", "darkred", "darkblue", "darkviolet")) +
  xlab(NULL) + ylab("Discrimination") +
  facet_grid(measure ~ dataset, switch = "y") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 20, hjust = 1, size = 8),
        legend.position = "bottom")
```


crime
recidivism
elift_ratio
impact_ratio
odds_ratio
Discrimination
2.0
1.5
1.0
0.5
0.0
None
Massaging
Reweighting
Uniform sampling
2 Naive Bayes
Model
Dataset
gnb
logr
svm

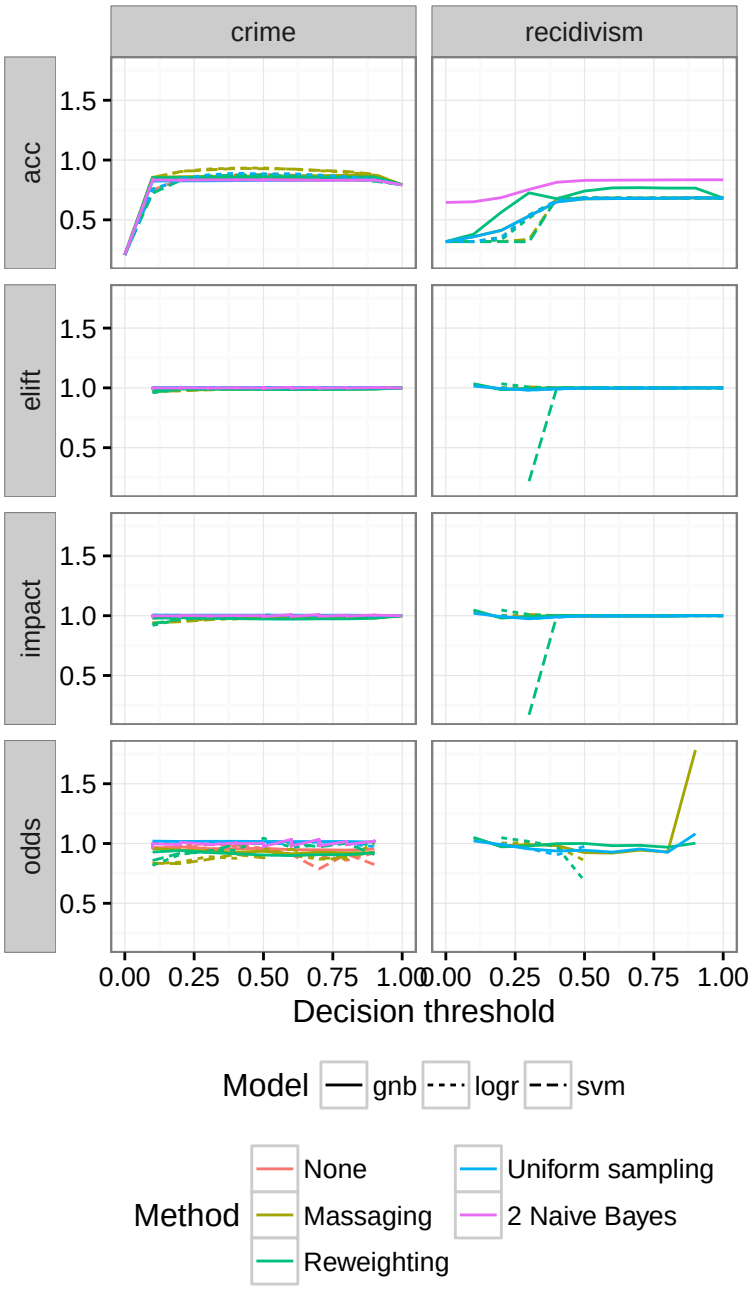
## Accuracy vs. discrimination

```
all_tidy %>%
  ggplot(aes(x = value, y = accuracy)) +
  geom_point(aes(colour = method, shape = model), size = 2) +
  guides(colour = guide_legend(title = "Method", nrow = 3)) +
  guides(shape = guide_legend(title = "Model")) +
  ylim(0, 1) +
  xlab("Discrimination") +
  #geom_text_repel(aes(x = value, y = accuracy, label = model)) +
  facet_grid(measure ~ dataset, switch = "y") +
  theme_bw() +
  theme(legend.position = "bottom")
```


crime
recidivism
elift_ratio
impact_ratio
odds_ratio
accuracy
1.00
0.75
0.50
0.25
0.00
0.7
0.8
0.9
1.0
Discrimination
Method
2 Naive Bayes
Massaging
None
Reweighting
Uniform sampling
Model
gnb
logr
svm

## Impact of threshold

```
thresh %>% gather(statistic, value, elift, impact, odds, acc) %>%
  ggplot(aes(x = thresholds, y = value)) +
  xlab("Decision threshold") + ylab(NULL) +
  geom_line(aes(colour = method, linetype = model)) +
  guides(colour = guide_legend(title = "Method", nrow = 3)) +
  guides(linetype = guide_legend(title = "Model")) +
  facet_grid(statistic ~ dataset, switch = "y") +
  theme_bw() +
  theme(legend.position = "bottom")
```


crime
recidivism
acc
elift
impact
odds
1.5
1.0
0.5
0.00
0.25
0.50
0.75
1.00
Decision threshold
Model
gnb
logr
svm
Method
None
Massaging
Reweighting
Uniform sampling
2 Naive Bayes