

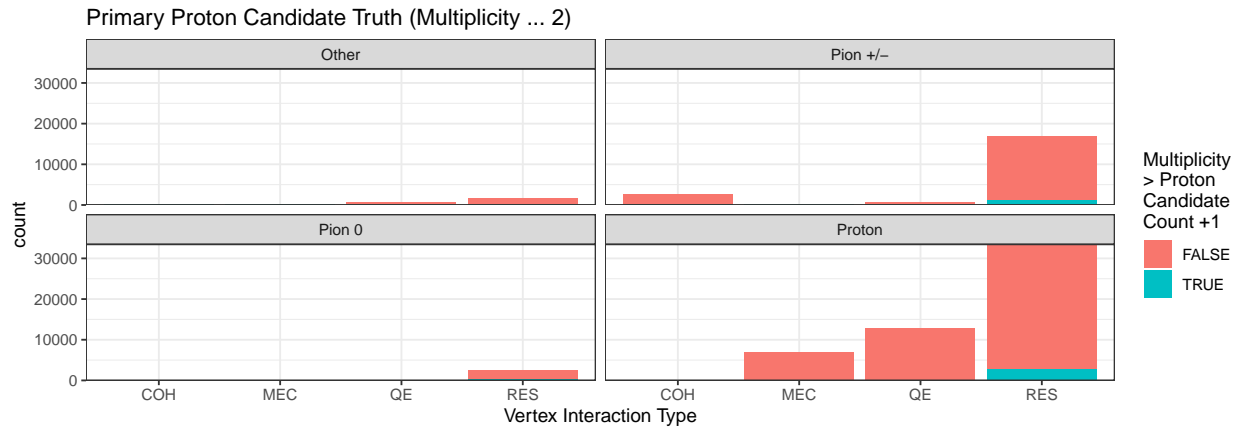
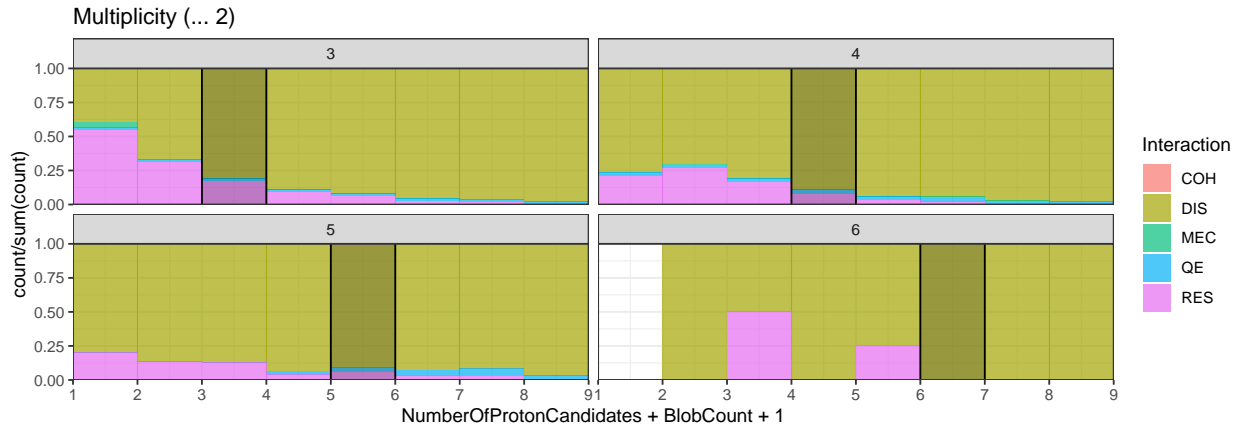
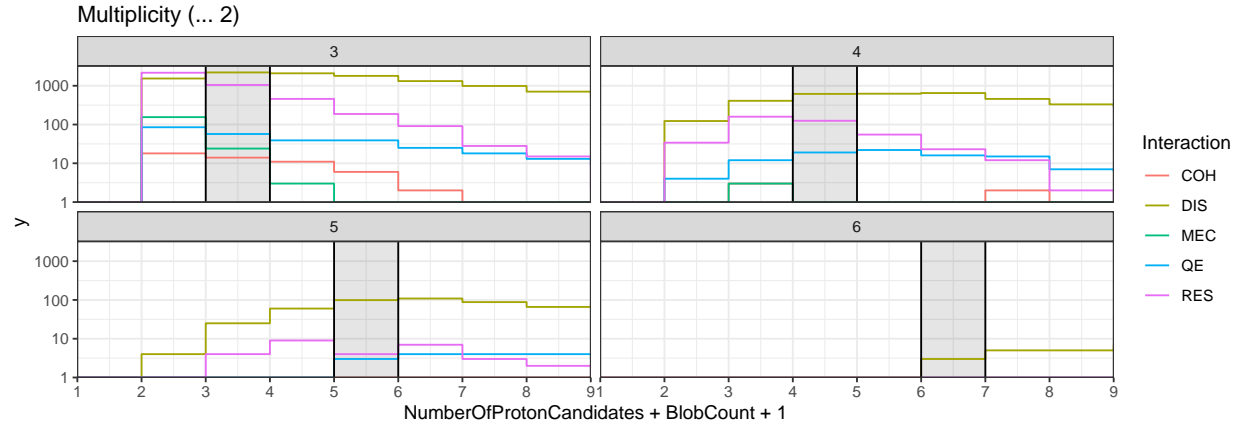
# Proton Classification

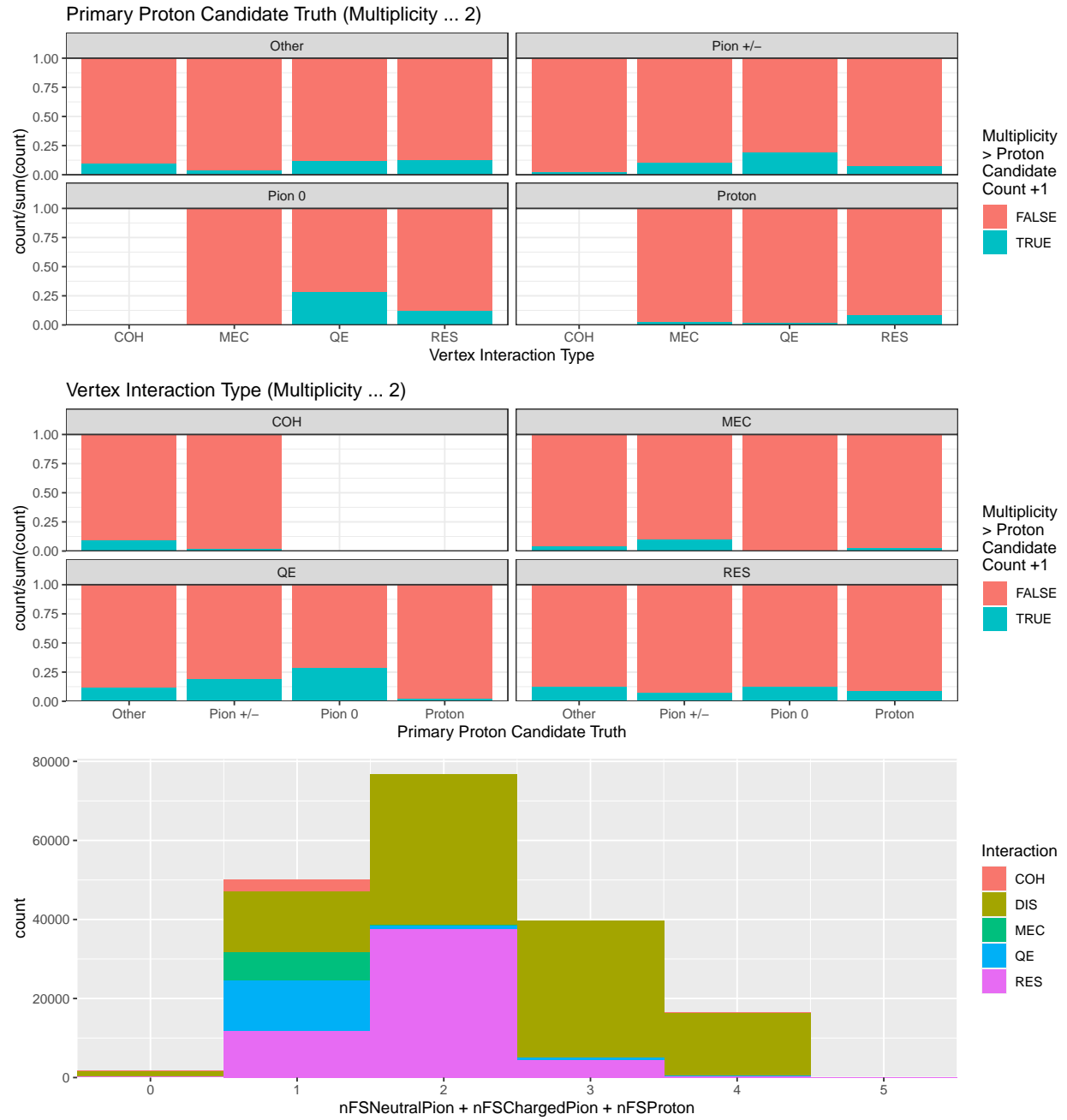
Sean Gilligan

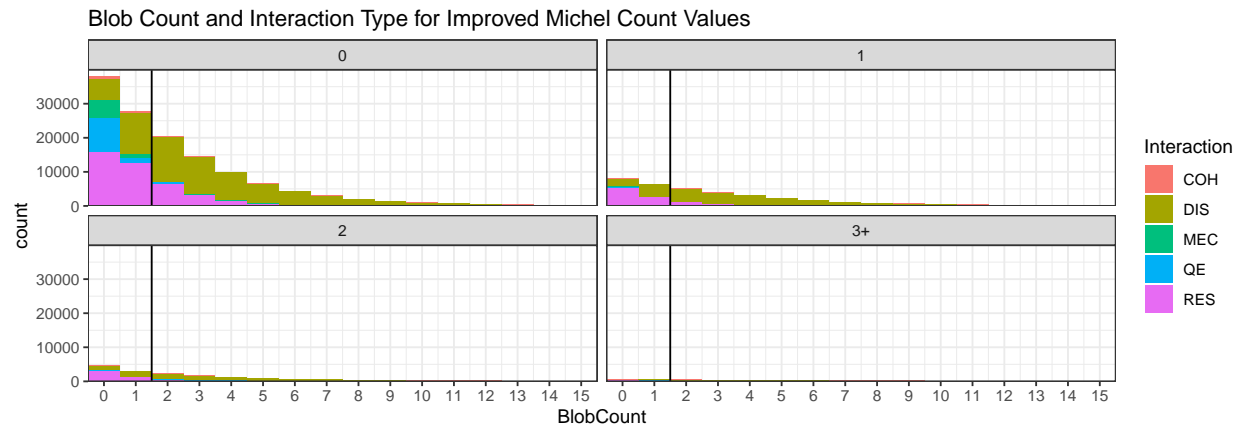
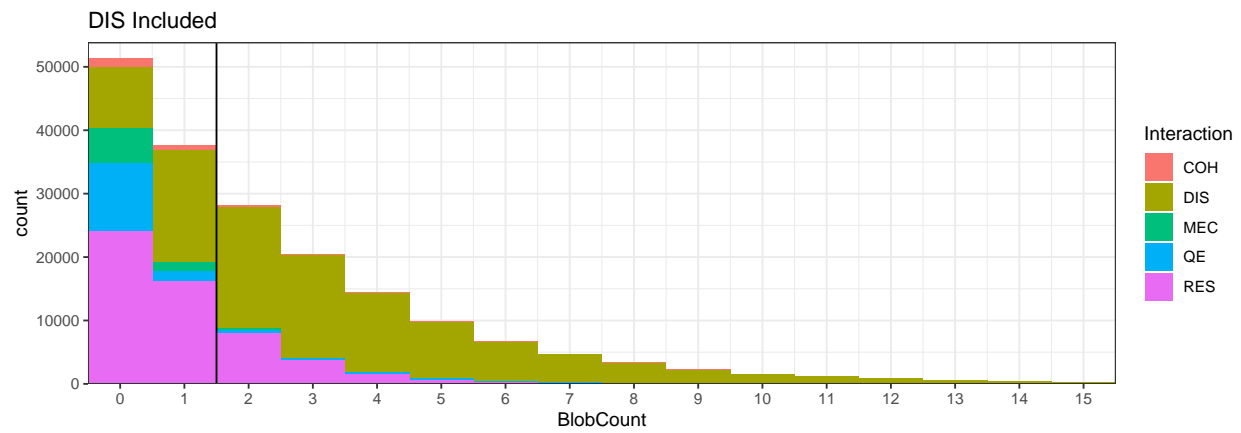
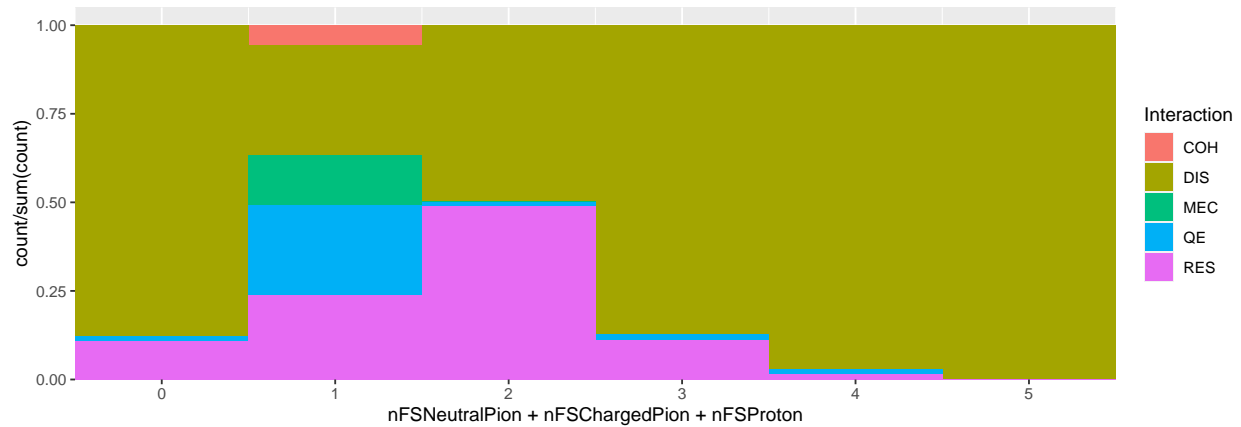
2023-01-12

## Variables in CSV File

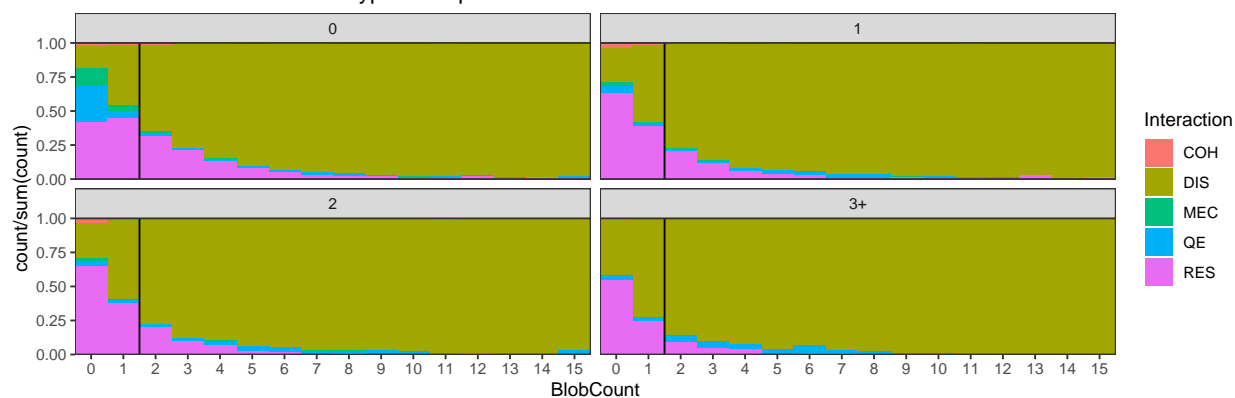
|   |                                   |
|---|-----------------------------------|
| ## [1] "Entry"                              | "BlobCount"                       |
| ## [3] "EnuCCQE"                            | "EventID"                         |
| ## [5] "ImprovedMichelCount"                | "Multiplicity"                    |
| ## [7] "NumClustsPrimaryProtonEnd"          | "NumClustsSecProtonEnd_1"         |
| ## [9] "NumClustsSecProtonEnd_2"            | "NumberOfProtonCandidates"        |
| ## [11] "PrimaryProtonAngle"                | "PrimaryProtonCandidatePDG"       |
| ## [13] "PrimaryProtonFractionEnergyInCone" | "PrimaryProtonScore"              |
| ## [15] "PrimaryProtonScore1"               | "PrimaryProtonScore2"             |
| ## [17] "PrimaryProtonTfromdEdx"            | "PrimaryProtonTrackLength"        |
| ## [19] "PrimaryProtonTrackVtxGap"          | "PrimaryProtonTrueKE"             |
| ## [21] "Q2QE"                              | "SecProtonAngle_1"                |
| ## [23] "SecProtonAngle_2"                  | "SecProtonCandidatePDG_1"         |
| ## [25] "SecProtonCandidatePDG_2"           | "SecProtonFractionEnergyInCone_1" |
| ## [27] "SecProtonFractionEnergyInCone_2"   | "SecondaryProtonScore1_1"         |
| ## [29] "SecondaryProtonScore1_2"           | "SecProtonTfromdEdx_1"            |
| ## [31] "SecProtonTfromdEdx_2"              | "SecProtonTrackVtxGap_1"          |
| ## [33] "SecProtonTrackVtxGap_2"            | "SecProtonTrueKE_1"               |
| ## [35] "SecProtonTrueKE_2"                 | "TotalPrimaryProtonEnergy"        |
| ## [37] "TotalSecProtonEnergy_1"            | "TotalSecProtonEnergy_2"          |
| ## [39] "ptmu"                              | "pzmu"                            |
| ## [41] "recoil"                            | "zvtx"                            |
| ## [43] "Interaction"                       | "nFSPart"                         |
| ## [45] "nFSChargedPion"                    | "nFSNeutralPion"                  |
| ## [47] "nFSProton"                         | "nFSNeutron"                      |
| ## [49] "nFSNegMuon"                        | "nFSGamma"                        |
| ## [51] "Arachne"                           | "Particle_0"                      |
| ## [53] "Particle_1"                        | "Particle_2"                      |
| ## [55] "Particle"                          |                                   |



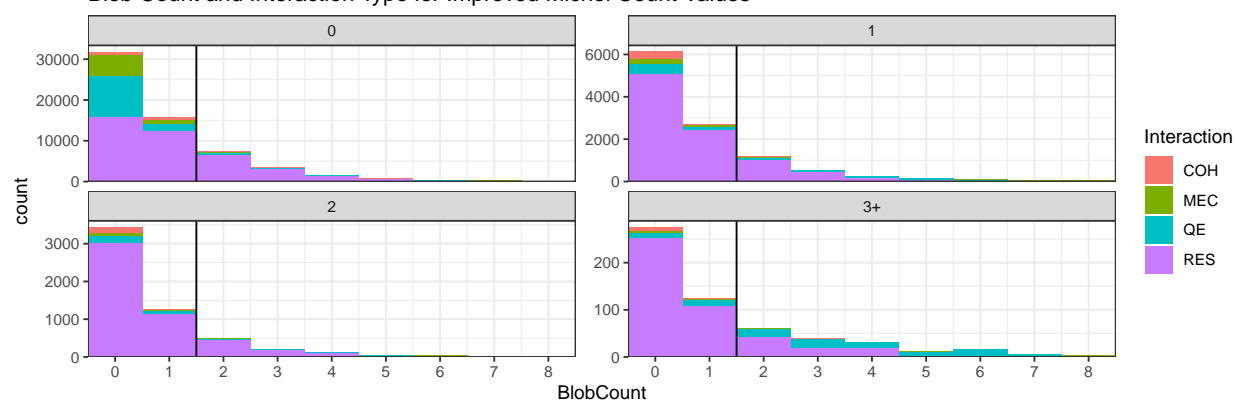




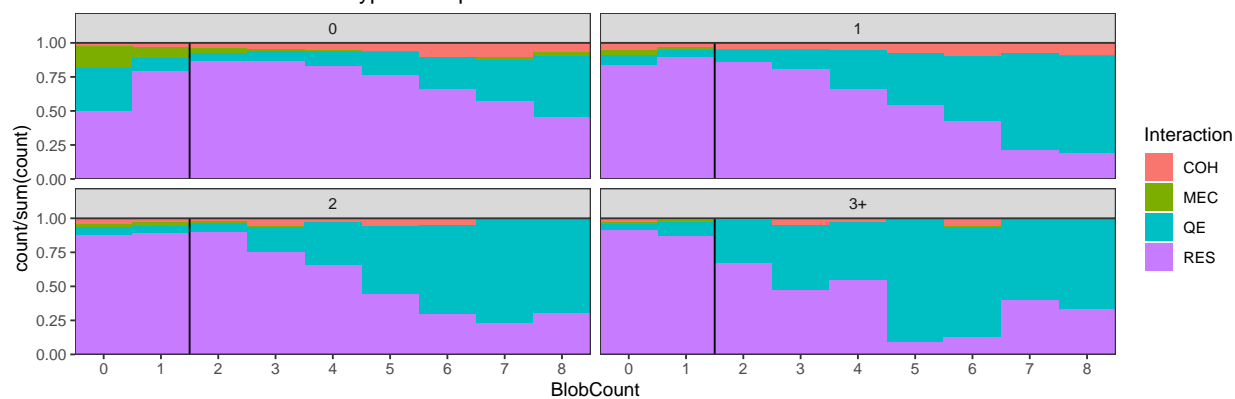
Blob Count and Interaction Type for Improved Michel Count Values

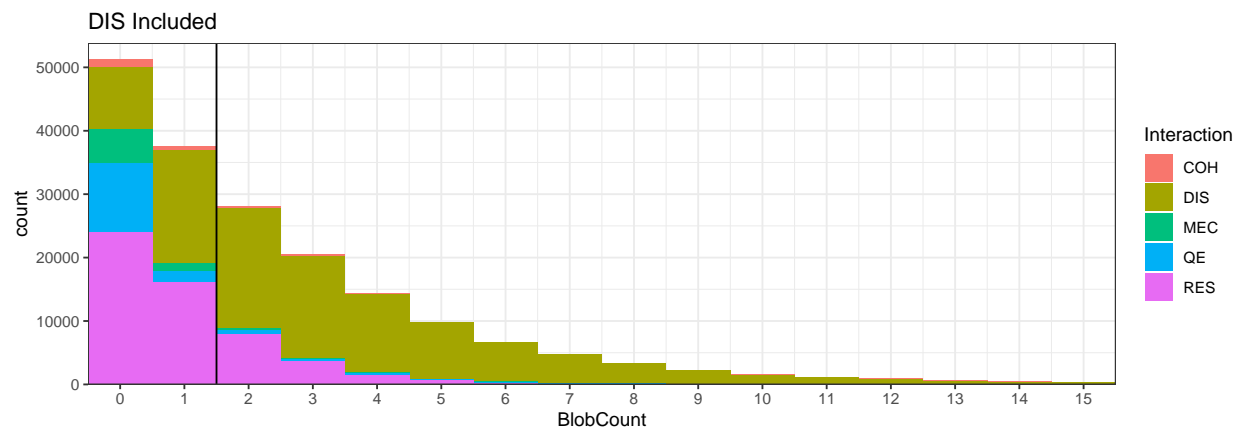
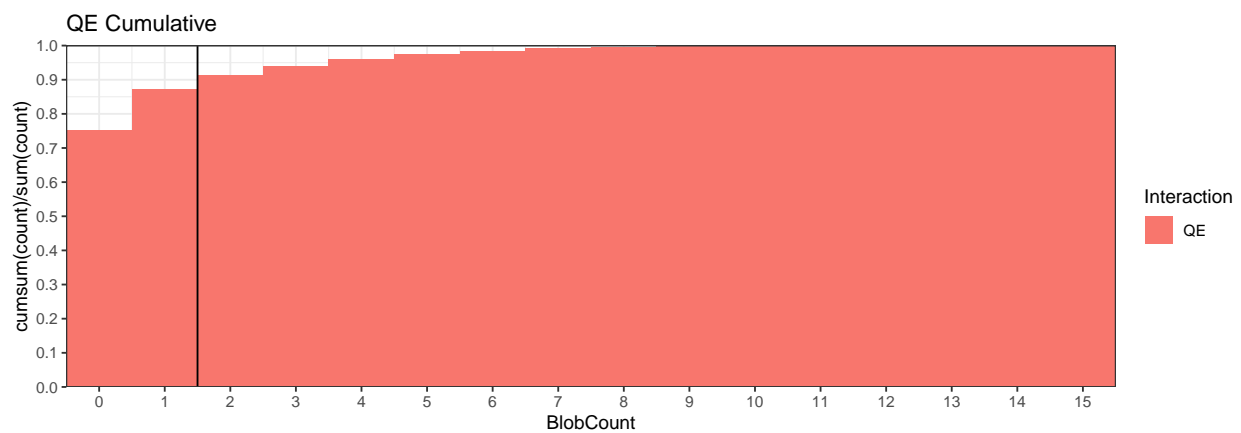
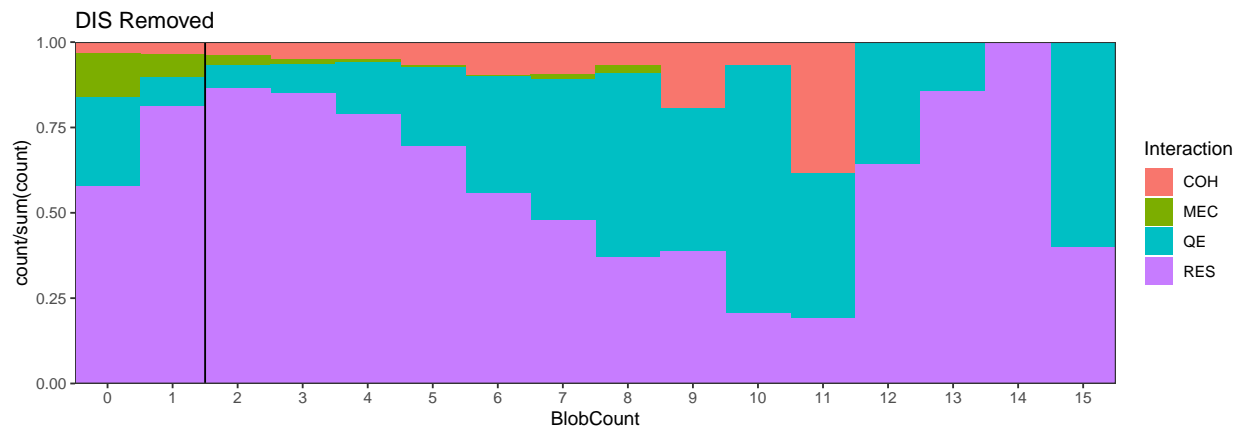


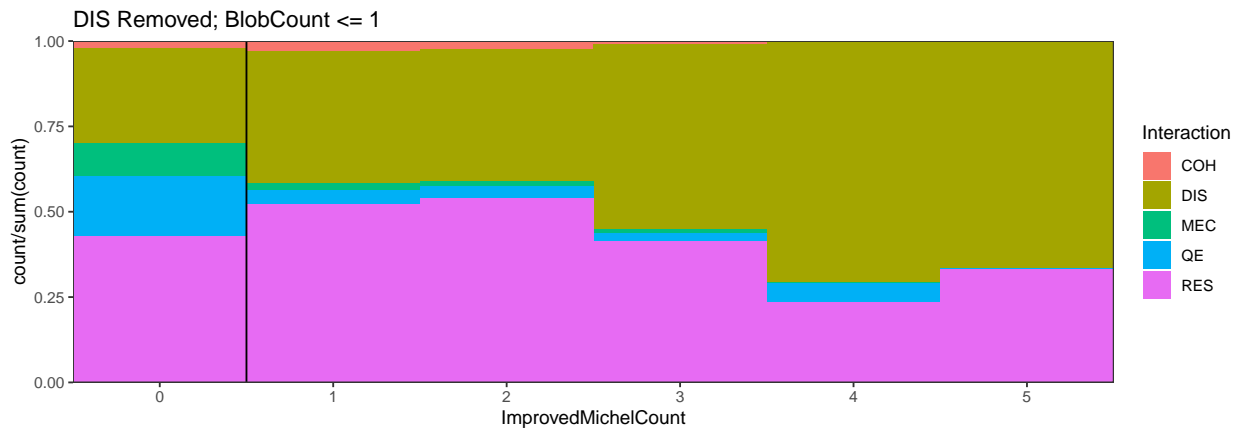
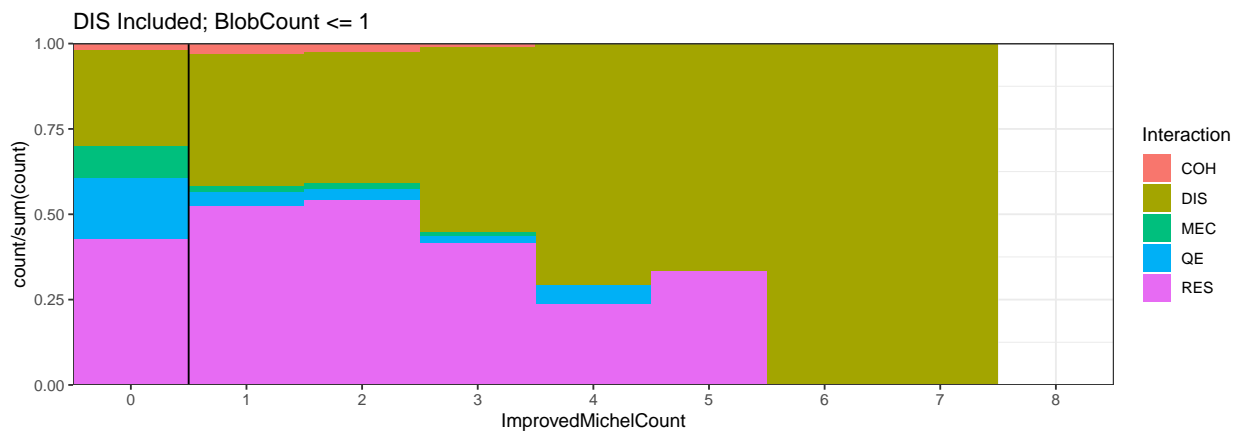
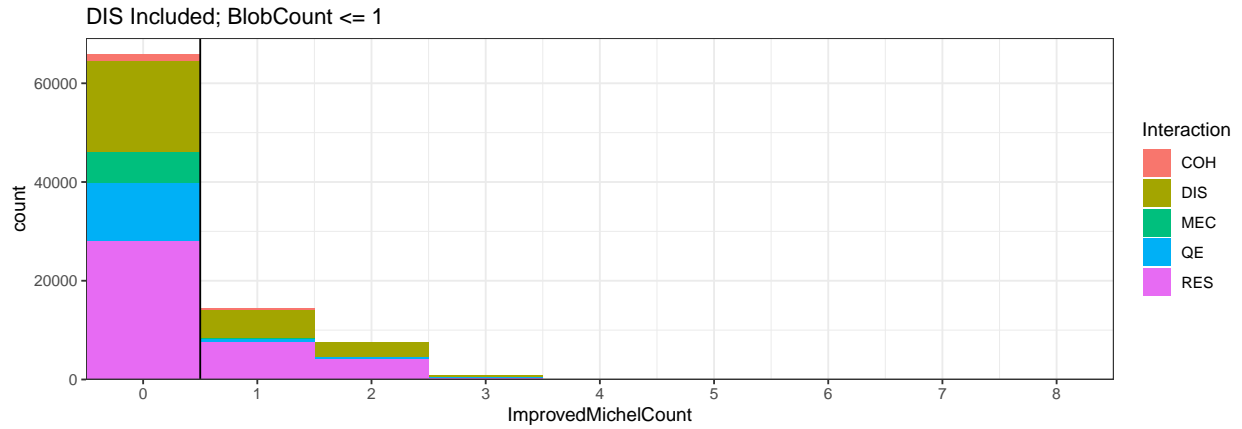
Blob Count and Interaction Type for Improved Michel Count Values

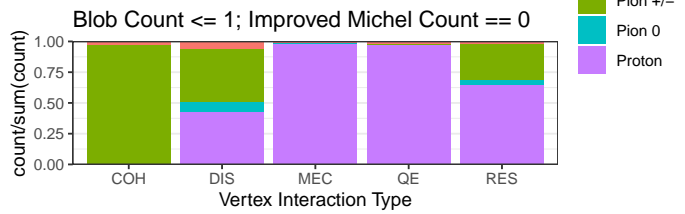
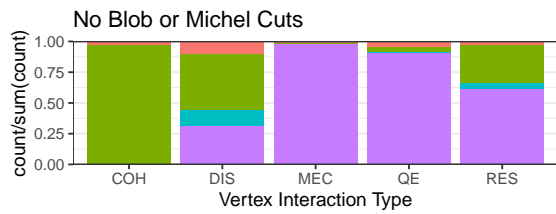
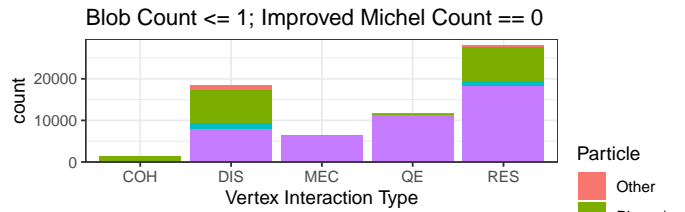
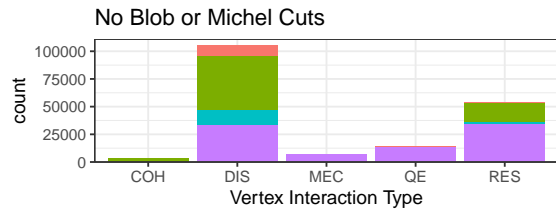
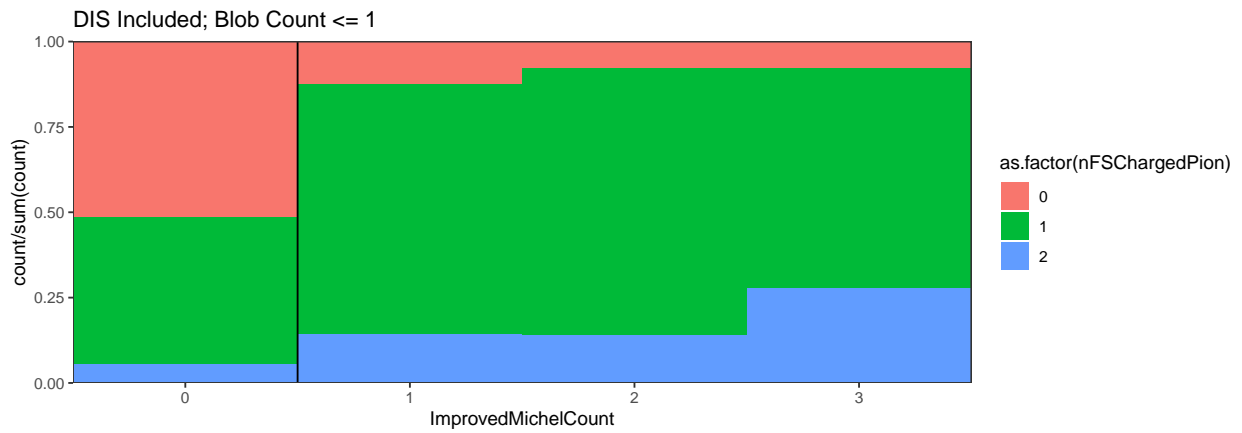
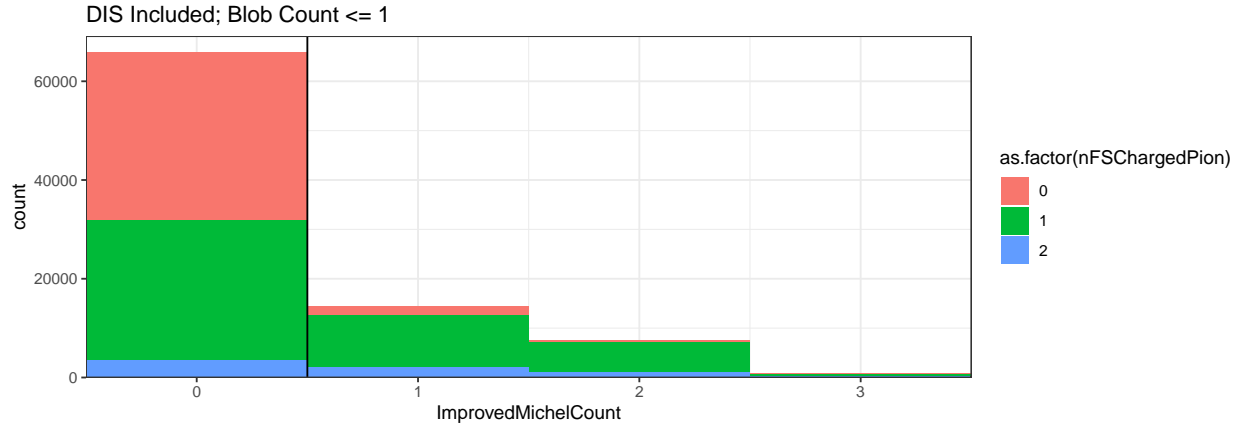


Blob Count and Interaction Type for Improved Michel Count Values

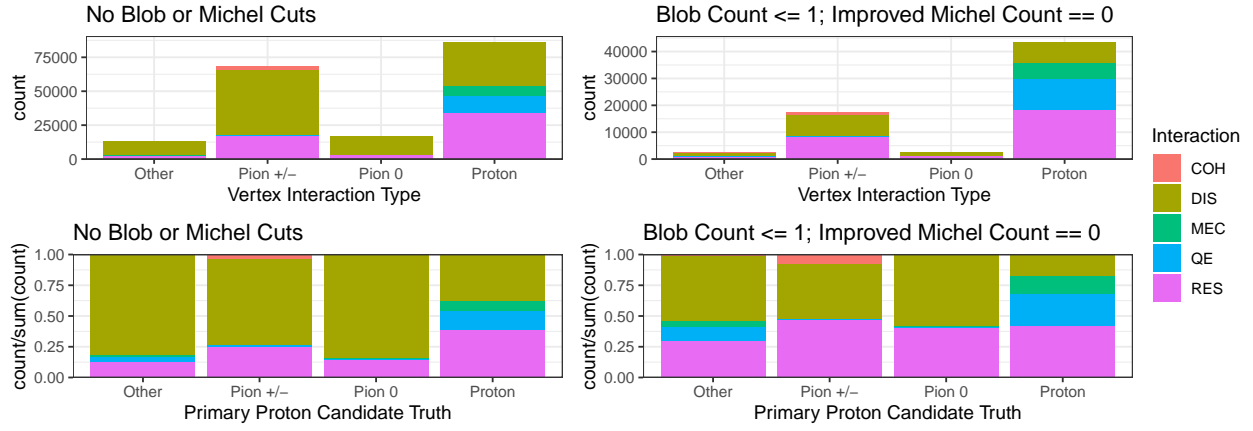






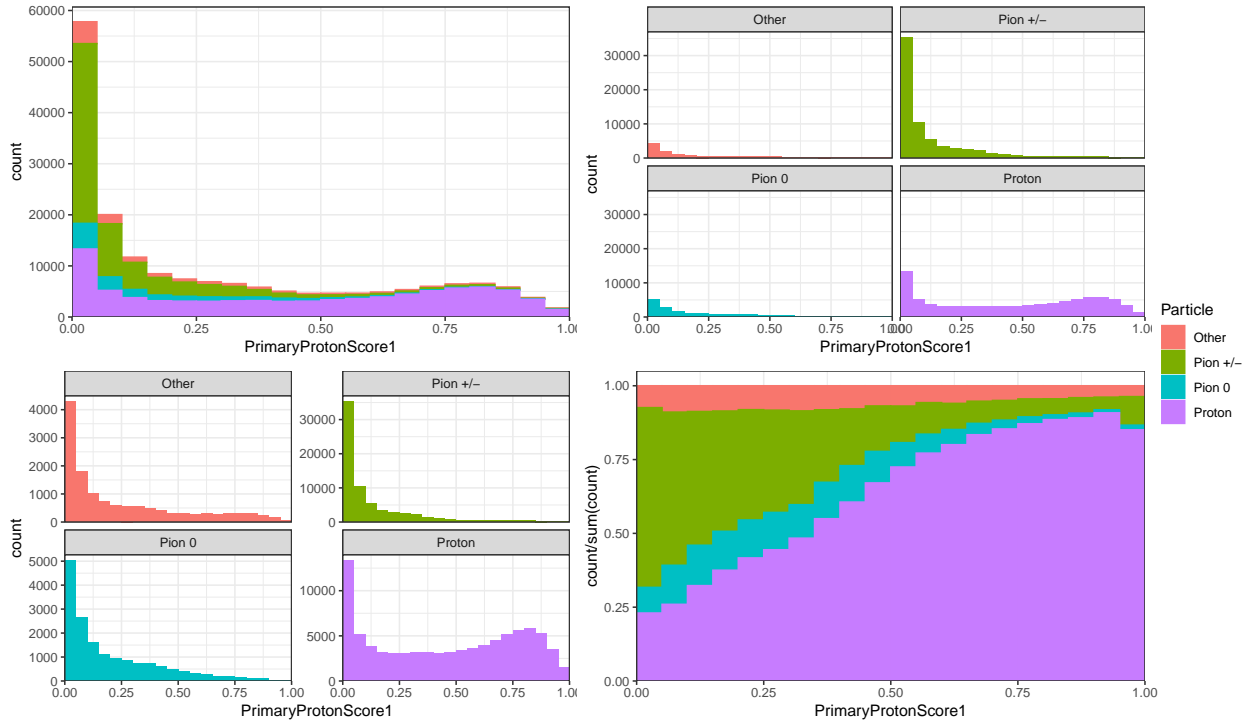






# Data Transformation

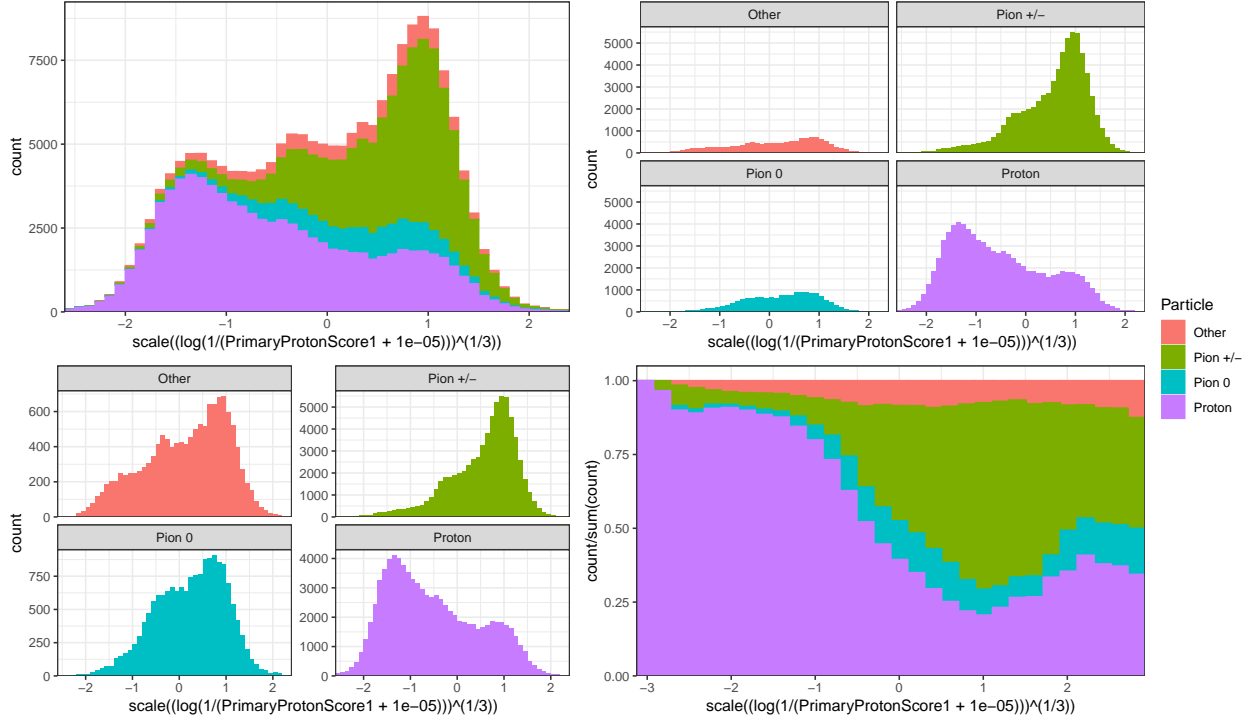
## PrimaryProtonScore1



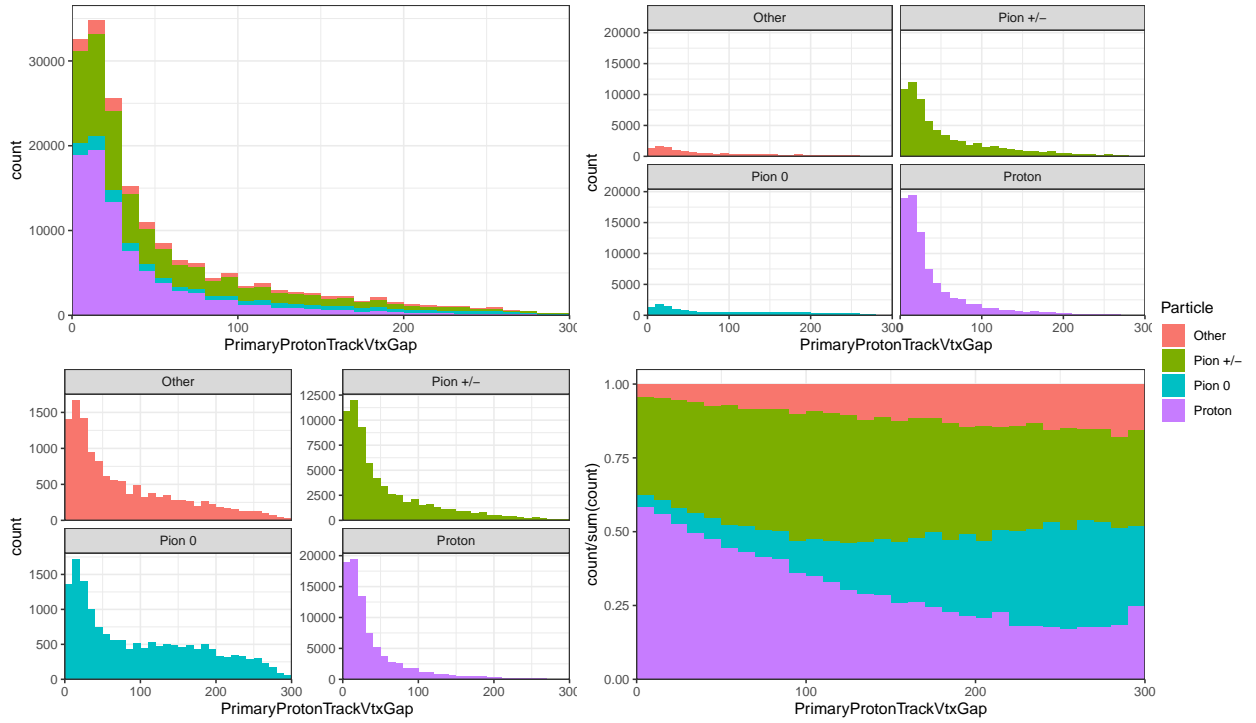
Modeled as a Beta distribution, transformed as below to “standardize”.

$$X' = \left[ \log \left( \frac{1}{X + 10^{-5}} \right) \right]^{1/3}$$

$$X'' = \frac{X' - \hat{\mu}_{X'}}{\hat{\sigma}_{X'}}$$

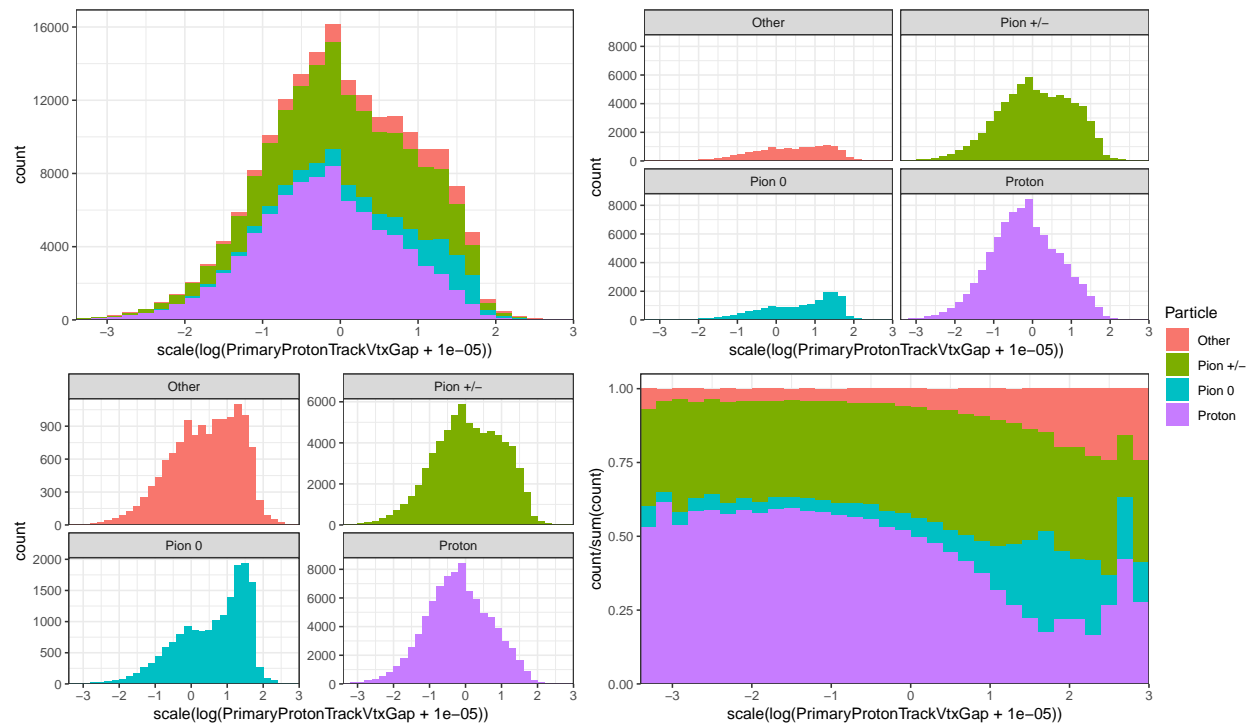


## PrimaryProtonTrackVtxGap

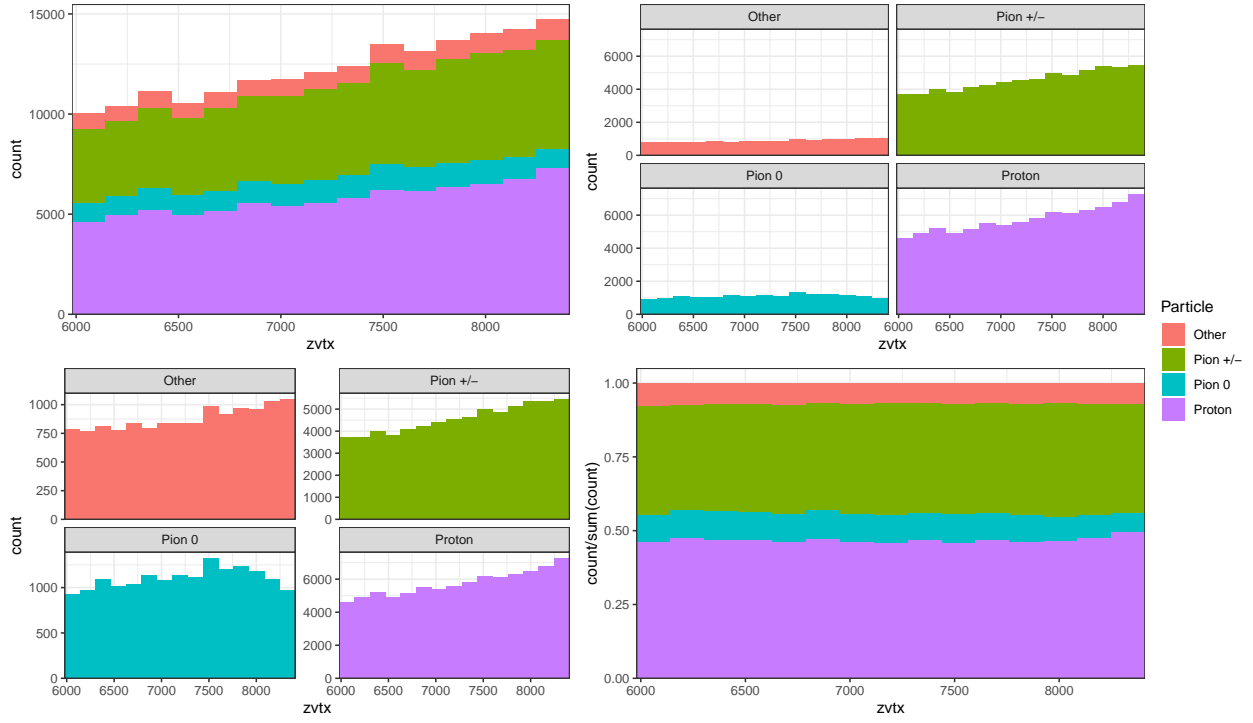


Modeled as an Exponential distribution, standardized as below (minus  $X'' = f(X')$  part)

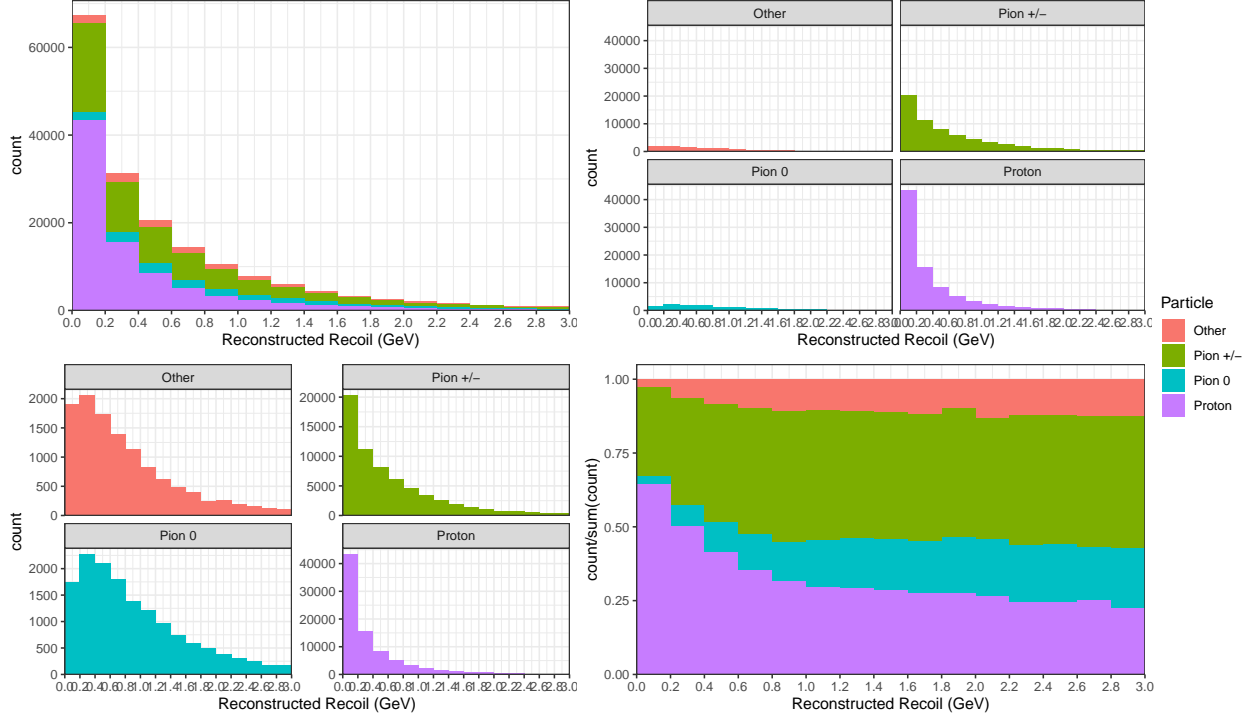
$$X' = \log(X + 10^{-5})$$



zvtx

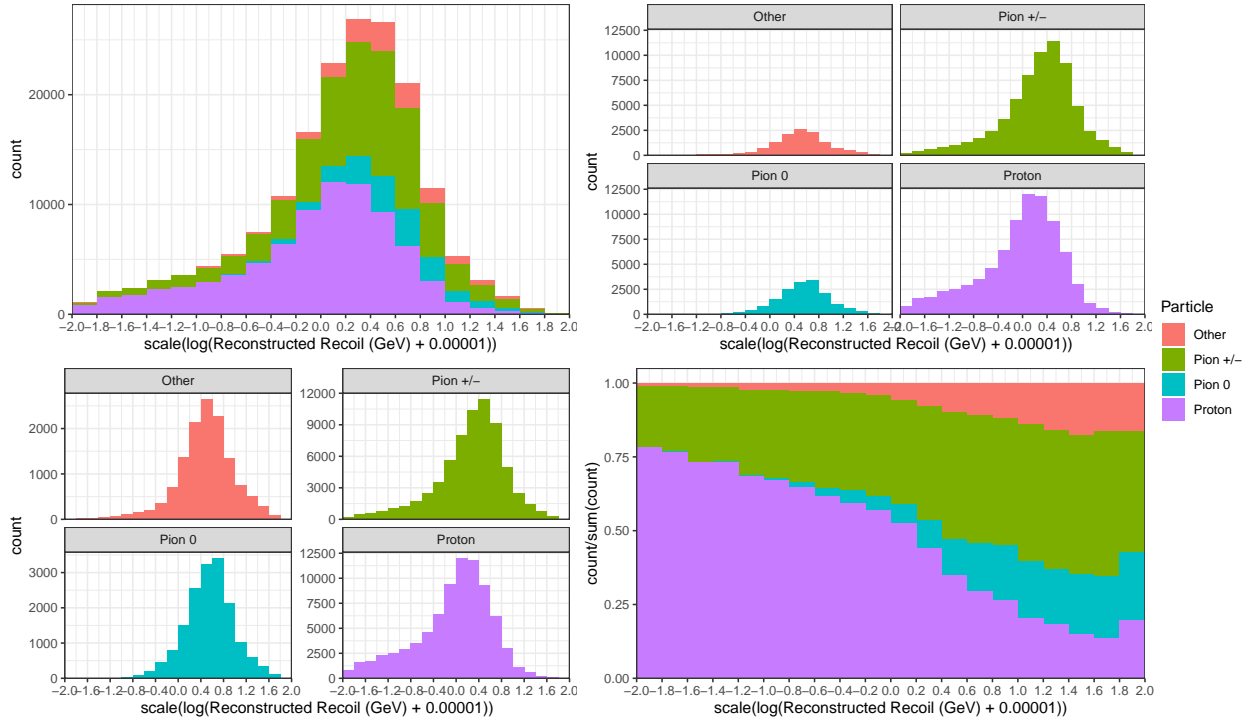


# recoil

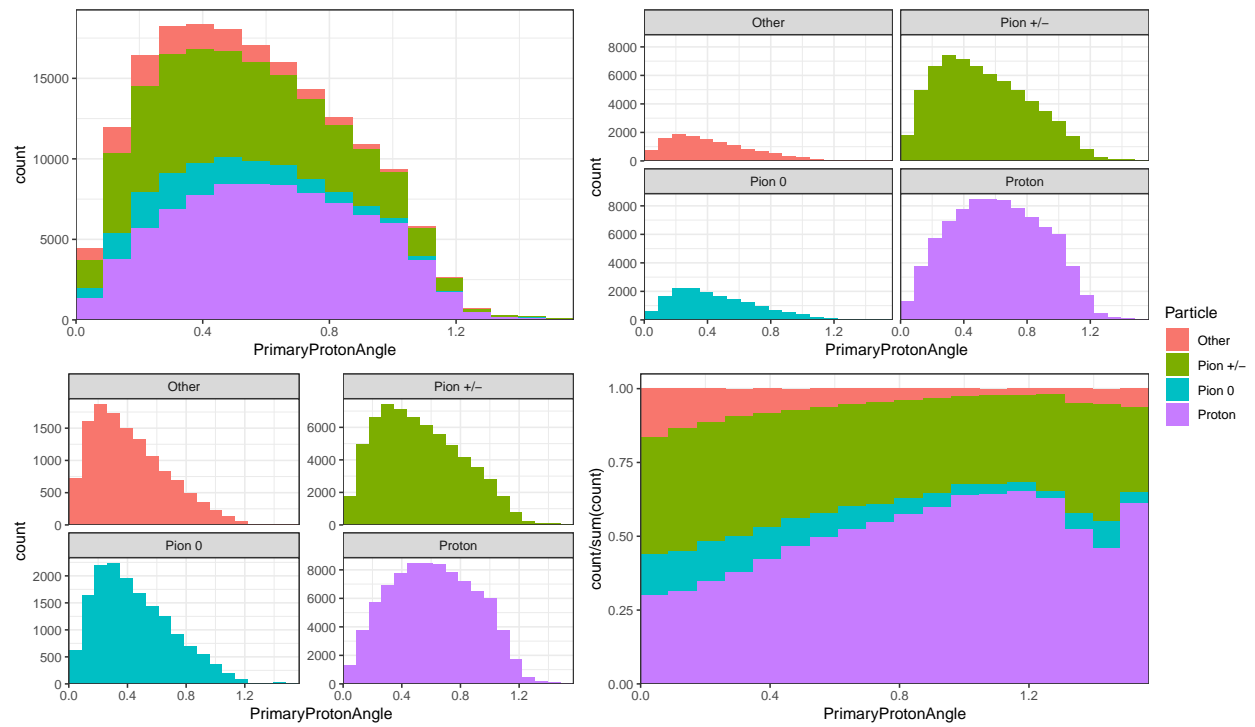


Modeled as an Exponential distribution, standardized as below (minus  $X'' = f(X')$  part)

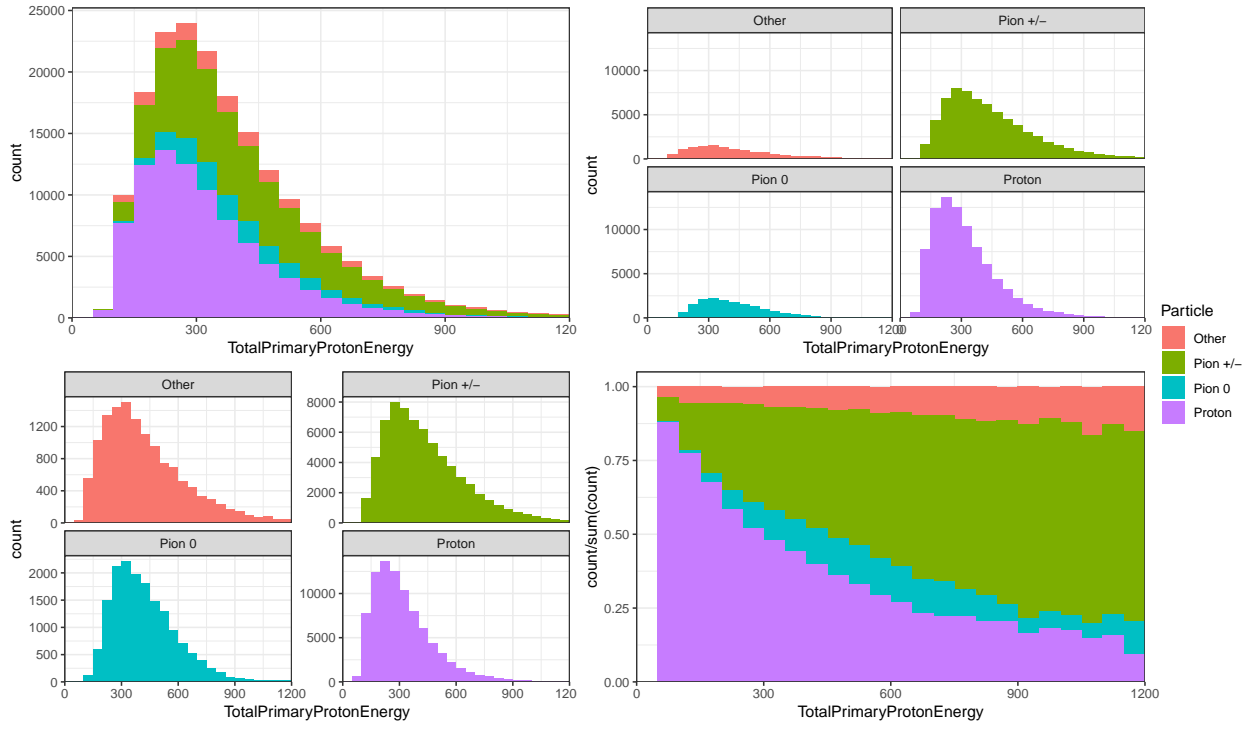
$$X' = \log(X + 10^{-5})$$



# PrimaryProtonAngle

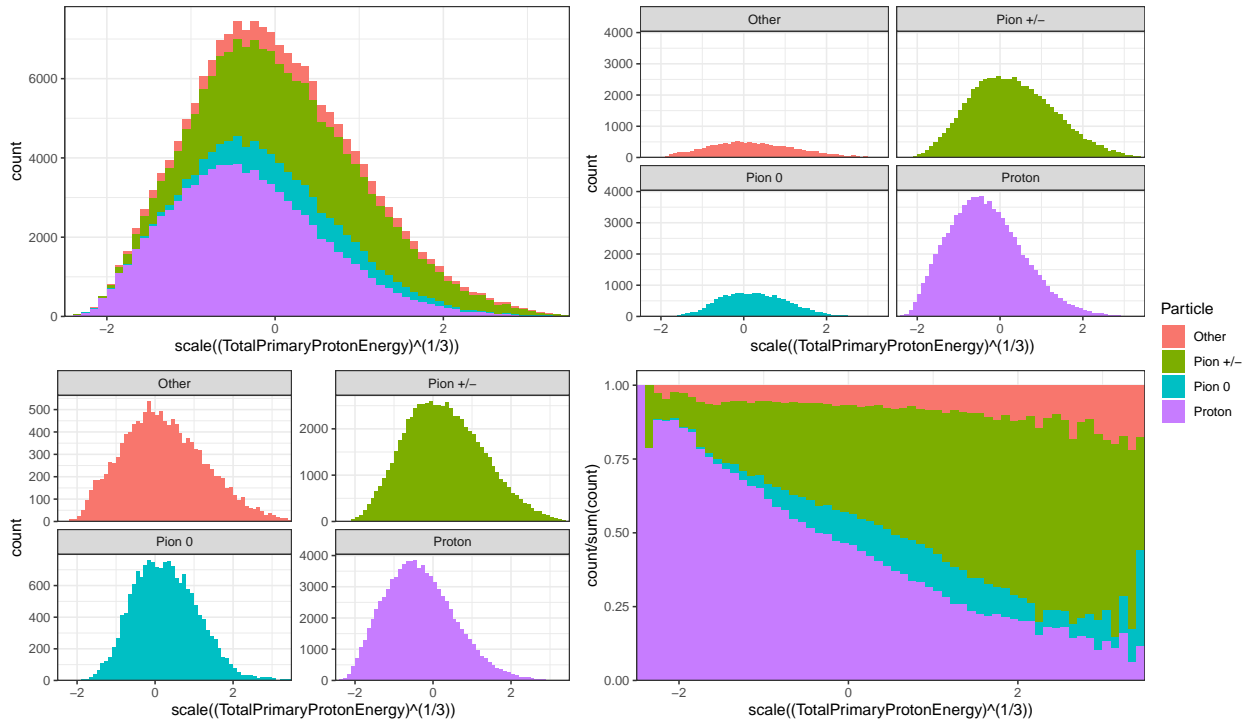


# TotalPrimaryProtonEnergydEdxAndClusters



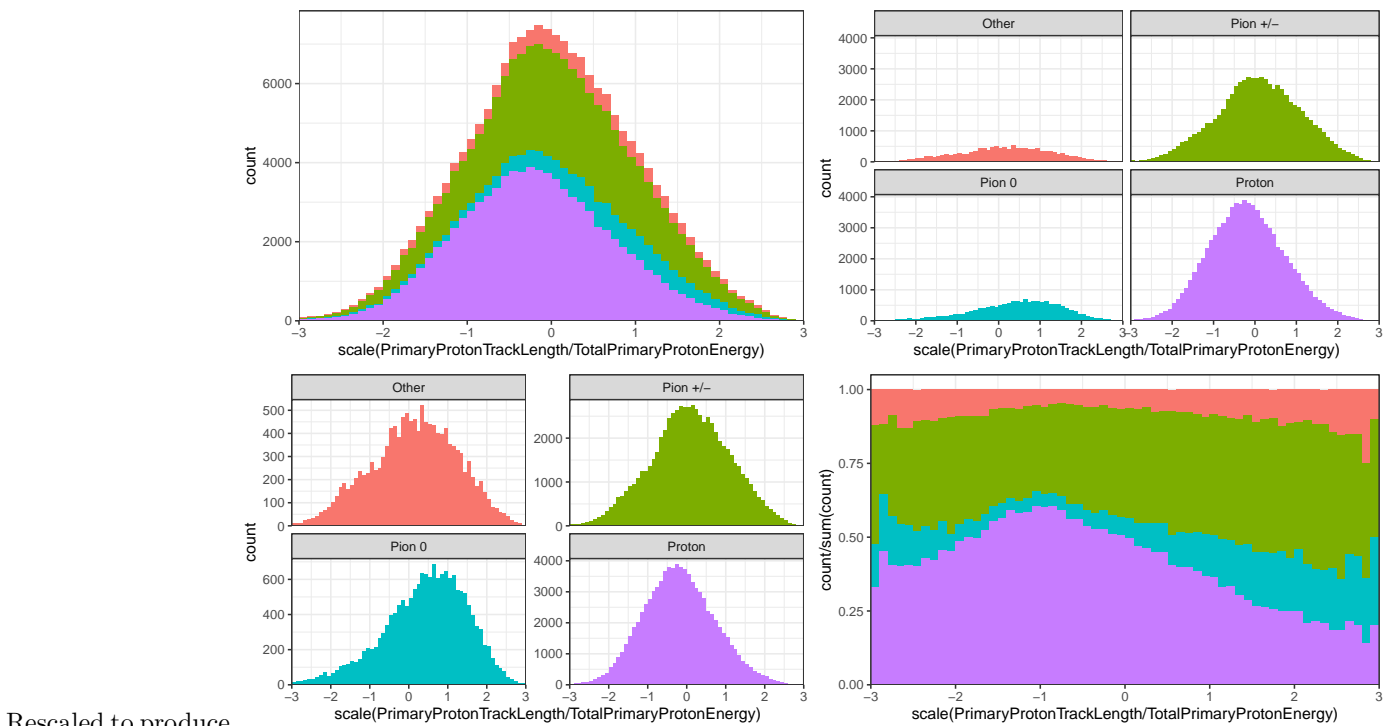
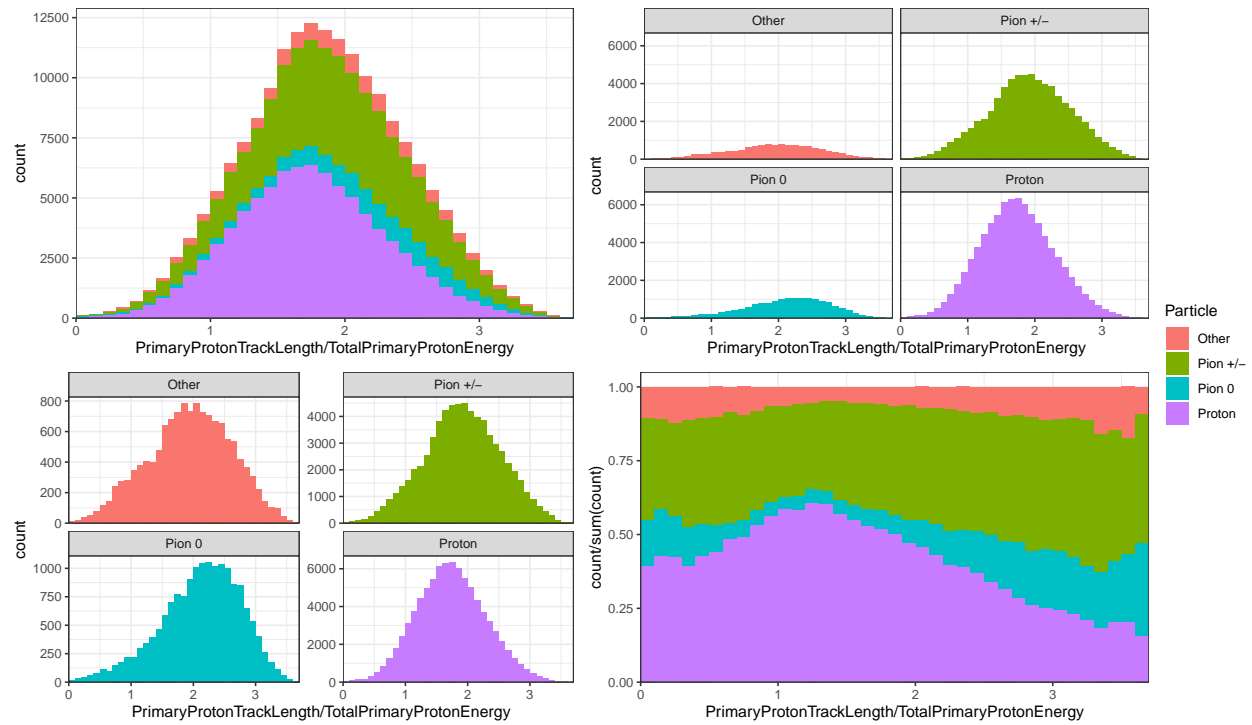
Standardized as below (minus  $X'' = f(X')$  part)

$$X' = X^{1/3}$$



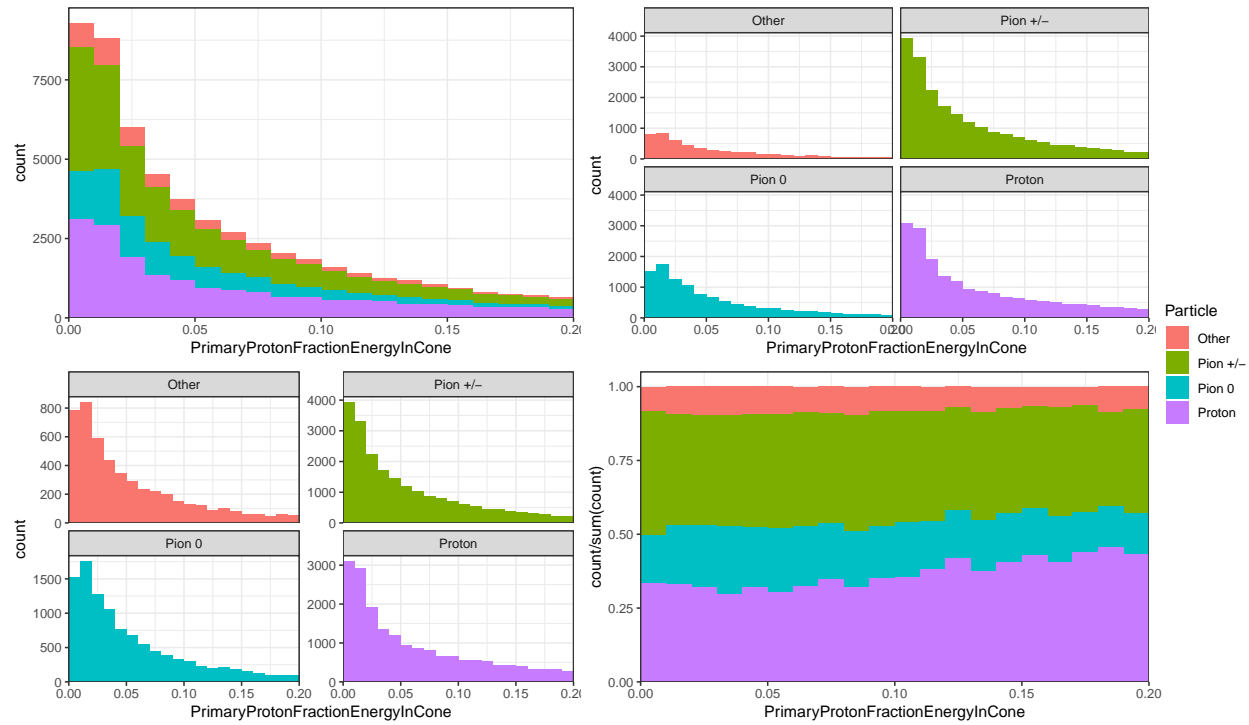


# PrimaryProtonTrackLength/TotalPrimaryProtonEnergydEdxAndClusters;

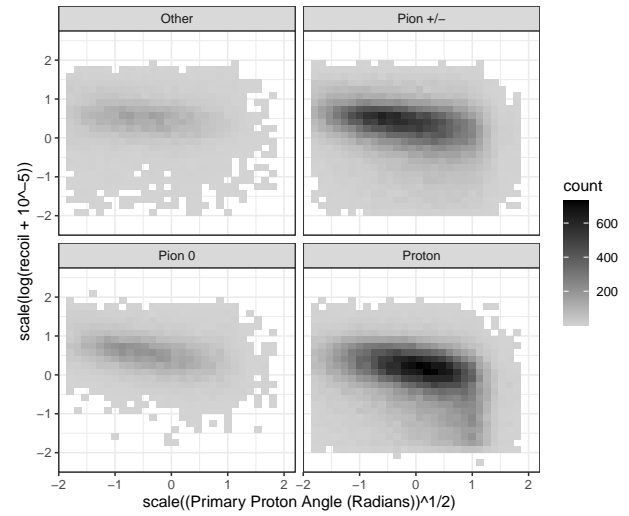
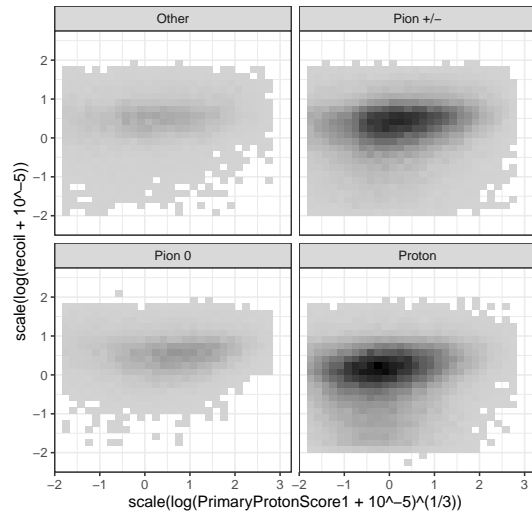
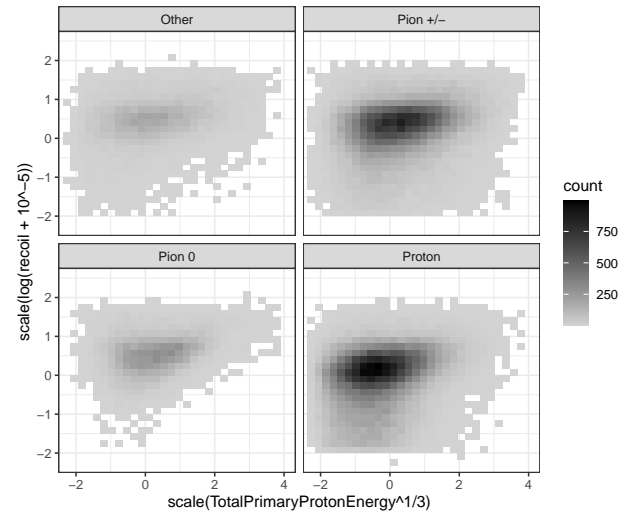
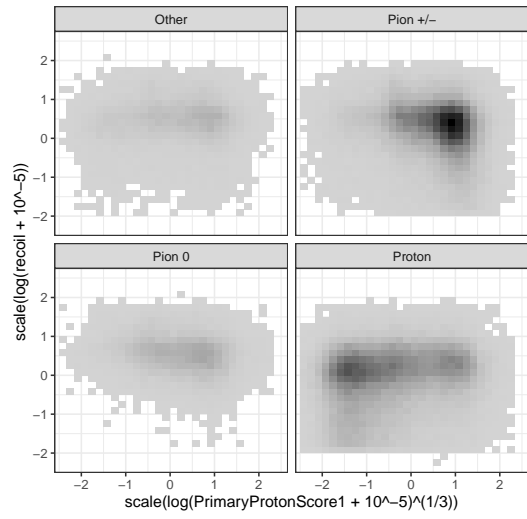
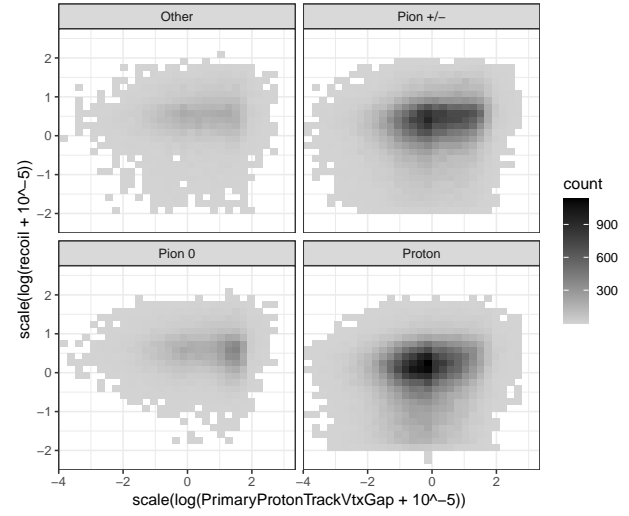
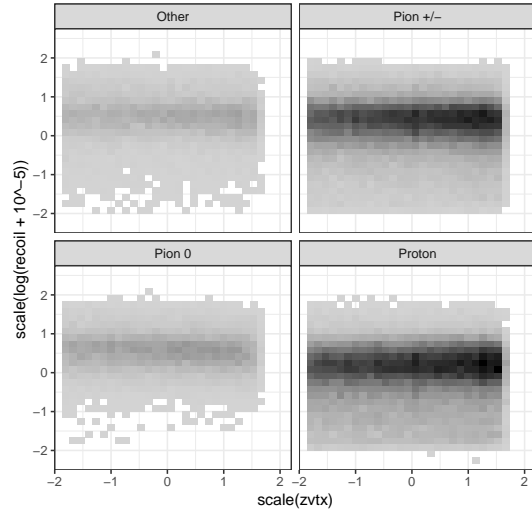


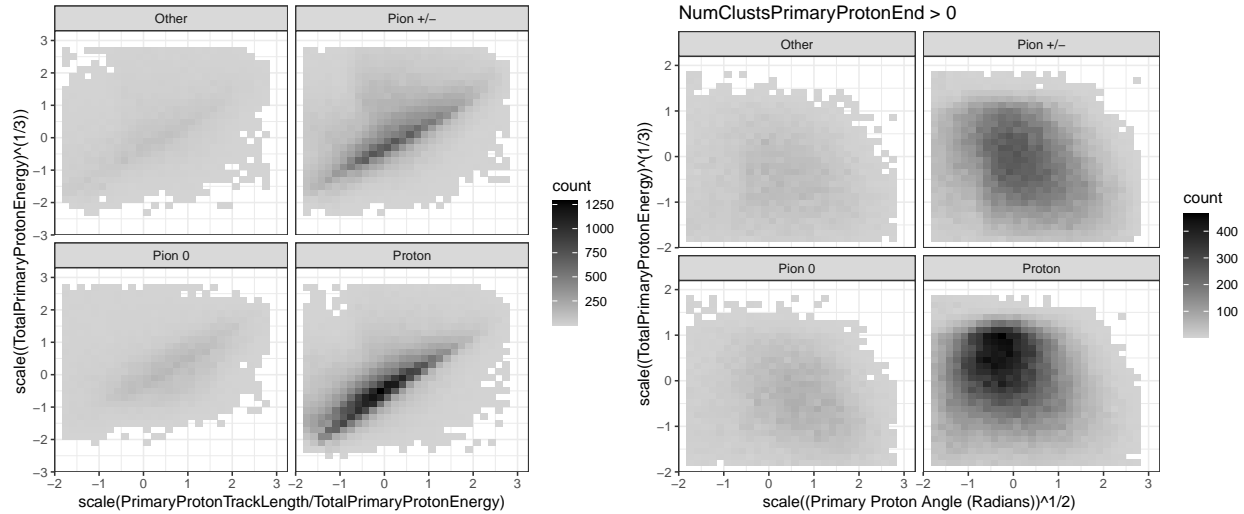
Rescaled to produce

# PrimaryProtonFractionEnergyInCone



Rescaled to produce





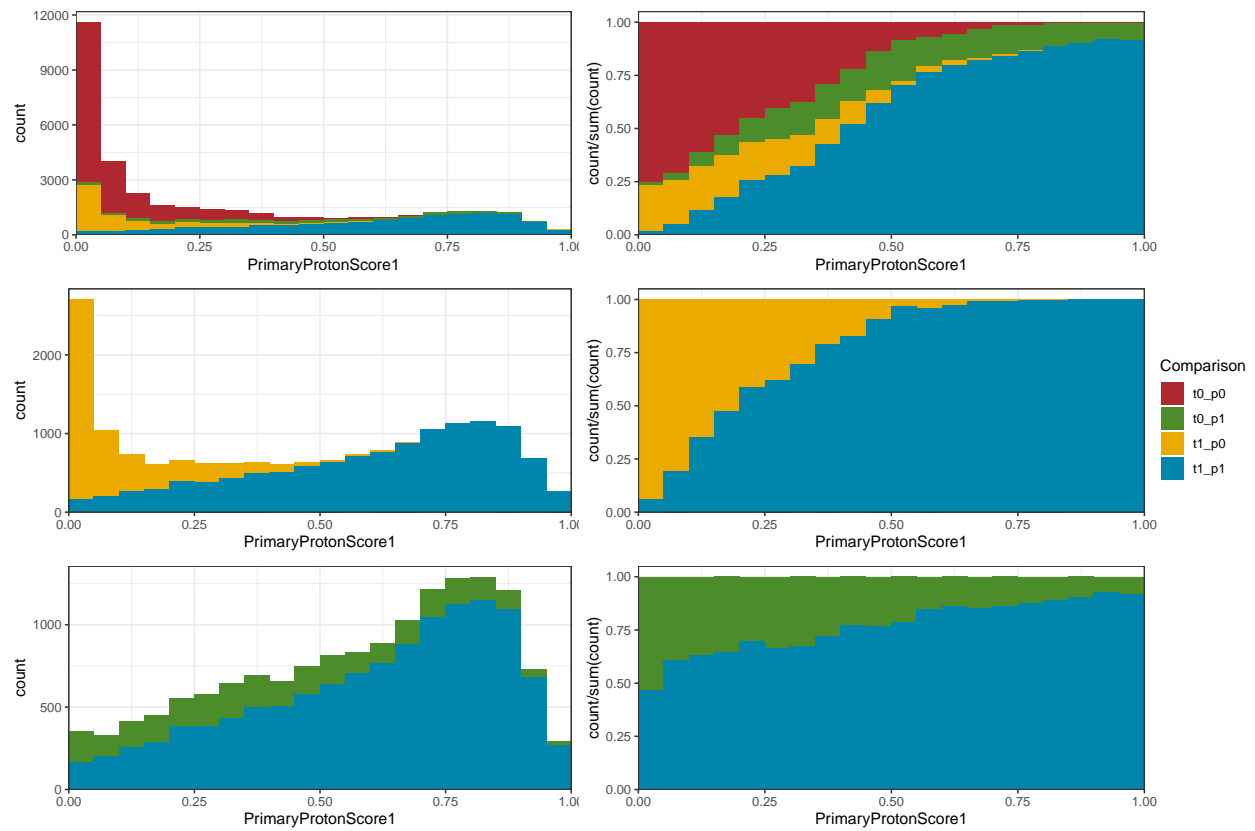
```
if(DEBUG){
  ggplot(stdvar, aes(TotalE, PVGap, color = Particle)) +
    geom_point(pch = 21) +
    scale_x_continuous(expand = expansion(mult = c(0, .05)), limits = c(-2.5,4)) +
    scale_y_continuous(expand = expansion(mult = c(0, .05)), limits = c(-4,3))
  ggplot(stdvar, aes(zvtx, PVGap, color = Particle)) +
    geom_point(pch = 21) +
    scale_x_continuous(expand = expansion(mult = c(0, .05)), limits = c(-2,2)) +
    scale_y_continuous(expand = expansion(mult = c(0, .05)), limits = c(-4,3))
  ggplot(stdvar, aes(ProtonScore1, PVGap, color = Particle)) +
    geom_point(pch = 21) +
    scale_x_continuous(expand = expansion(mult = c(0, .05)), limits = c(-2.6,2.4)) +
    scale_y_continuous(expand = expansion(mult = c(0, .05)), limits = c(-4,3))
  ggplot(stdvar, aes(TrackLen, PVGap, color = Particle)) +
    geom_point(pch = 21) +
    scale_x_continuous(expand = expansion(mult = c(0, .05)), limits = c(-2.6,3.5)) +
    scale_y_continuous(expand = expansion(mult = c(0, .05)), limits = c(-4,3))

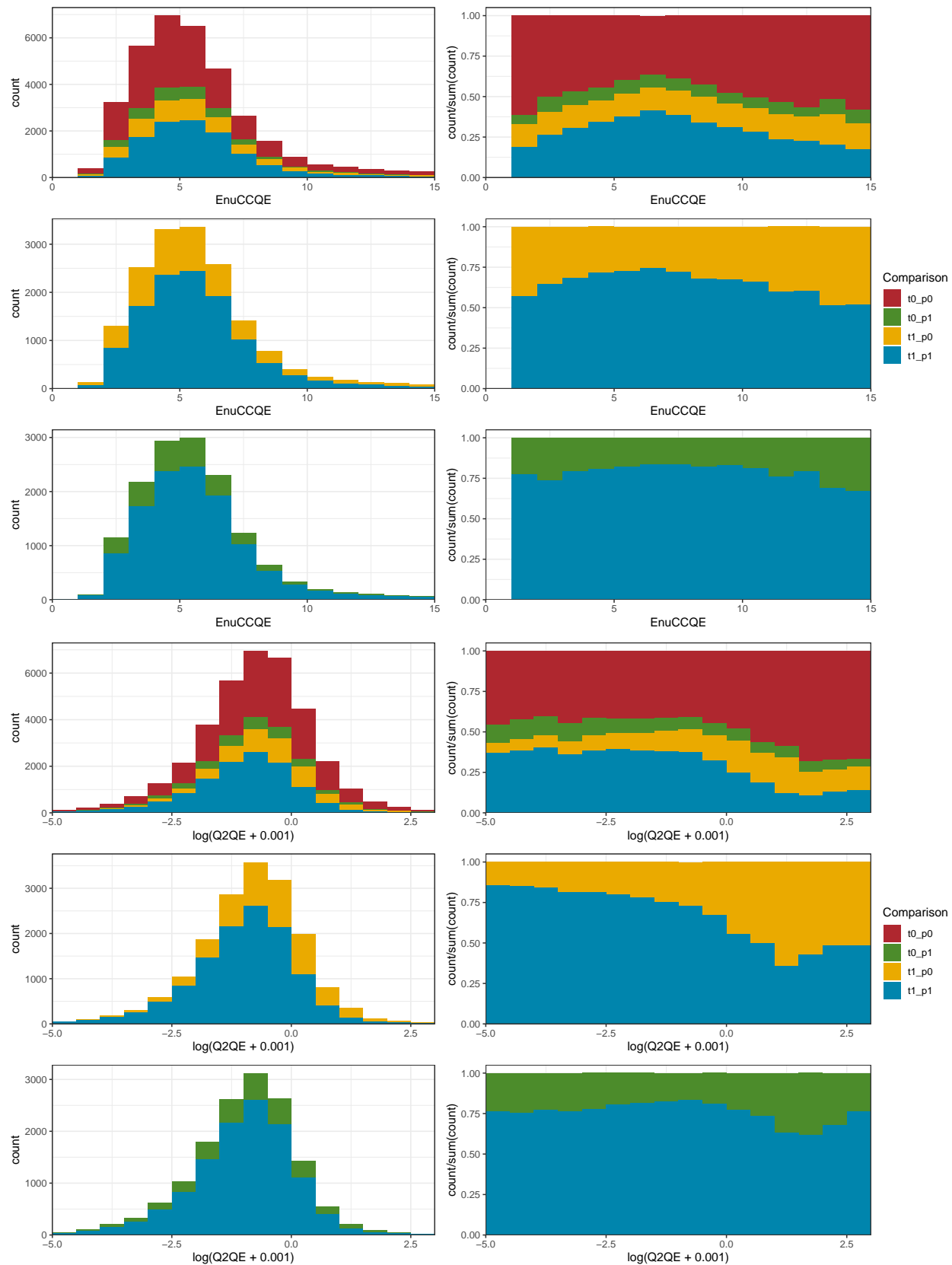
  ggplot(stdvar, aes(TotalE, PVGap)) +
    geom_bin2d() +
    theme_bw() +
    scale_fill_gradient(low = "lightgray", high = "black") +
    scale_x_continuous(expand = expansion(mult = c(0, .05)), limits = c(-2.5,4)) +
    scale_y_continuous(expand = expansion(mult = c(0, .05)), limits = c(-4,3)) +
    facet_wrap(vars(Particle))
  ggplot(stdvar, aes(zvtx, PVGap)) +
    geom_bin2d() +
    theme_bw() +
    scale_fill_gradient(low = "lightgray", high = "black") +
    scale_x_continuous(expand = expansion(mult = c(0, .05)), limits = c(-2,2)) +
    scale_y_continuous(expand = expansion(mult = c(0, .05)), limits = c(-4,3)) +
    facet_wrap(vars(Particle))
  ggplot(stdvar, aes(ProtonScore1, PVGap)) +
    geom_bin2d() +
    theme_bw() +
    scale_fill_gradient(low = "lightgray", high = "black") +
    scale_x_continuous(expand = expansion(mult = c(0, .05)), limits = c(-2.6,2.4)) +
```

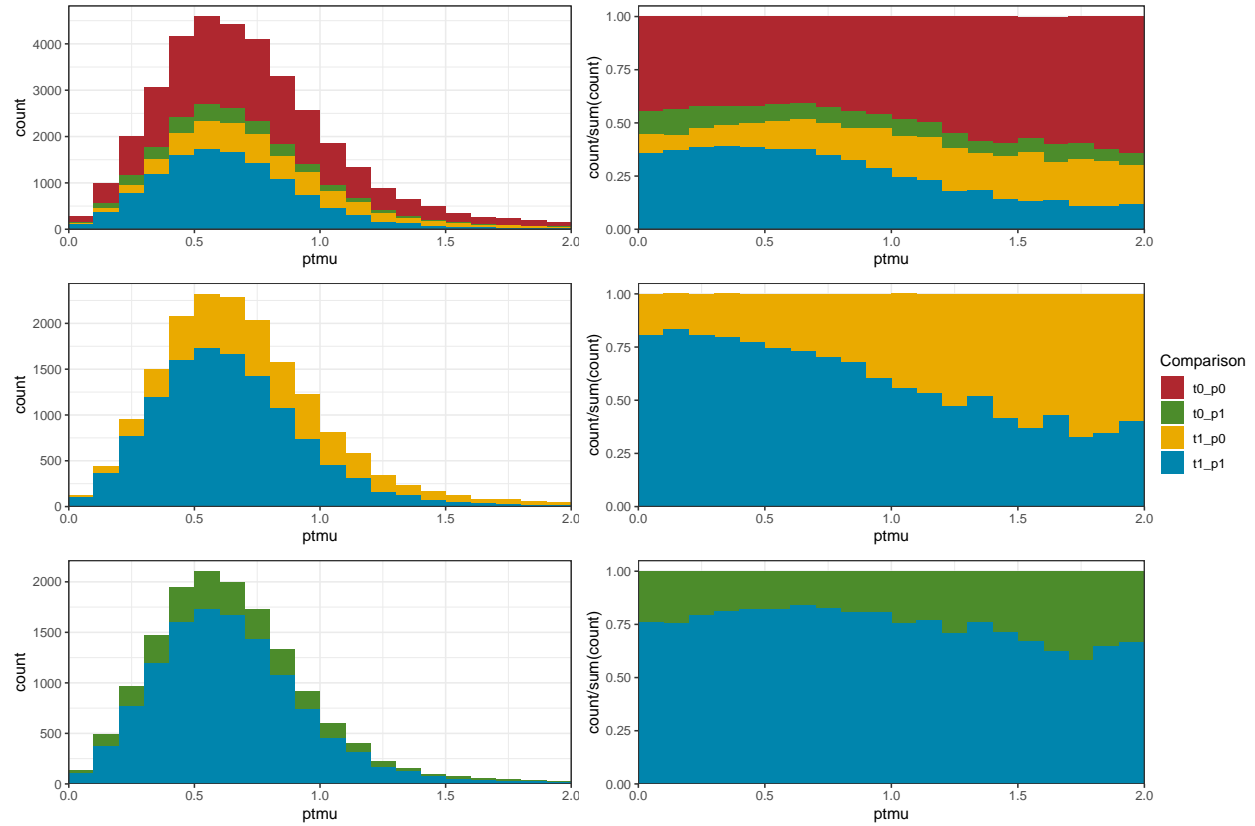
```

scale_y_continuous(expand = expansion(mult = c(0, .05)), limits = c(-4, 3)) +
facet_wrap(vars(Particle))
}

```







## ## Confusion Matrix and Statistics

##

## Reference

## Prediction Other Pion +/- Pion 0 Proton

|             |       |       |       |       |
|-------------|-------|-------|-------|-------|
| ## Other    | 10418 | 1     | 0     | 2     |
| ## Pion +/- | 59    | 54619 | 61    | 141   |
| ## Pion 0   | 1     | 2     | 13103 | 3     |
| ## Proton   | 80    | 125   | 32    | 68771 |

##

## Overall Statistics

##

## Accuracy : 0.9966

## 95% CI : (0.9962, 0.9969)

## No Information Rate : 0.4675

## P-Value [Acc > NIR] : < 2.2e-16

##

## Kappa : 0.9945

##

## McNemar's Test P-Value : < 2.2e-16

##

## Statistics by Class:

##

## Class: Other Class: Pion +/- Class: Pion 0 Class: Proton

## Sensitivity 0.98674 0.9977 0.99295 0.9979

## Specificity 0.99998 0.9972 0.99996 0.9970

## Pos Pred Value 0.99971 0.9952 0.99954 0.9966

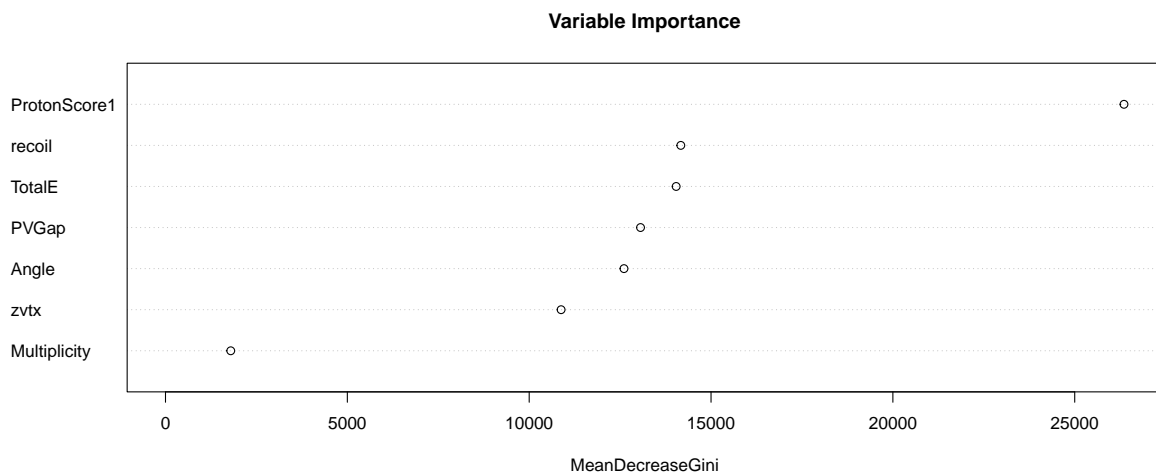
## Neg Pred Value 0.99898 0.9986 0.99931 0.9981

```

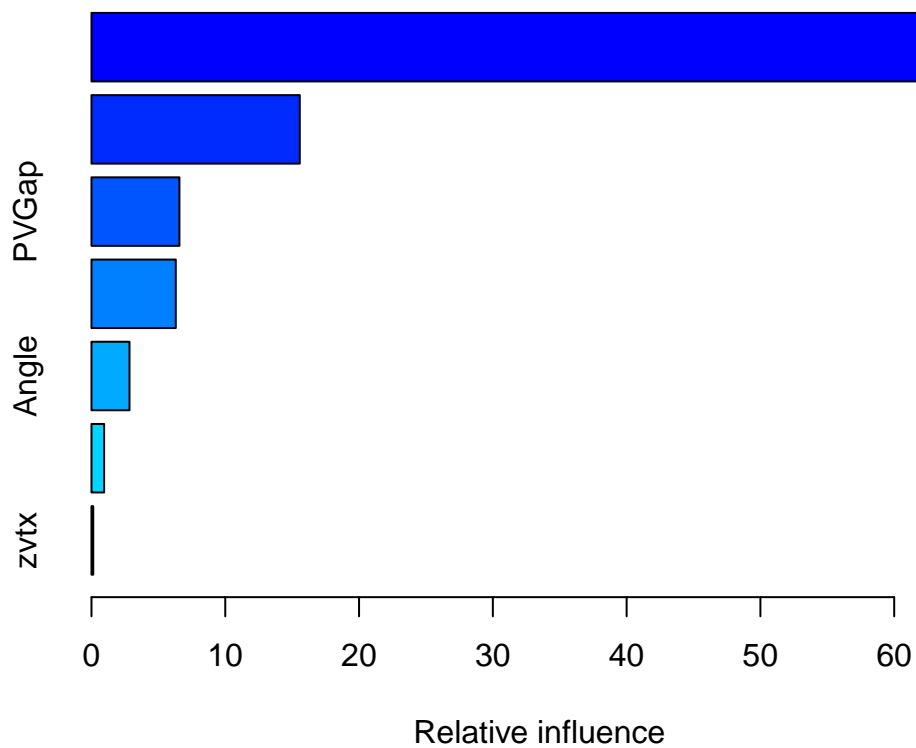
## Prevalence          0.07162          0.3714          0.08951          0.4675
## Detection Rate      0.07067          0.3705          0.08888          0.4665
## Detection Prevalence 0.07069          0.3723          0.08892          0.4681
## Balanced Accuracy    0.99336          0.9974          0.99645          0.9974

## Confusion Matrix and Statistics
##
##              Reference
## Prediction Other Pion +/- Pion 0 Proton
##   Other      114      148      81      127
##   Pion +/-   1410     10078    1894    4117
##   Pion 0      227      531      655      310
##   Proton      880      2809      693    12780
##
## Overall Statistics
##
##              Accuracy : 0.6411
##              95% CI : (0.6362, 0.646)
##      No Information Rate : 0.4703
##      P-Value [Acc > NIR] : < 2.2e-16
##
##              Kappa : 0.4029
##
##      McNemar's Test P-Value : < 2.2e-16
##
## Statistics by Class:
##
##              Class: Other Class: Pion +/- Class: Pion 0 Class: Proton
## Sensitivity          0.043330          0.7429          0.19711          0.7373
## Specificity          0.989598          0.6813          0.96815          0.7755
## Pos Pred Value       0.242553          0.5759          0.38015          0.7447
## Neg Pred Value       0.930821          0.8198          0.92406          0.7687
## Prevalence           0.071390          0.3681          0.09017          0.4703
## Detection Rate       0.003093          0.2735          0.01777          0.3468
## Detection Prevalence 0.012753          0.4748          0.04675          0.4657
## Balanced Accuracy     0.516464          0.7121          0.58263          0.7564

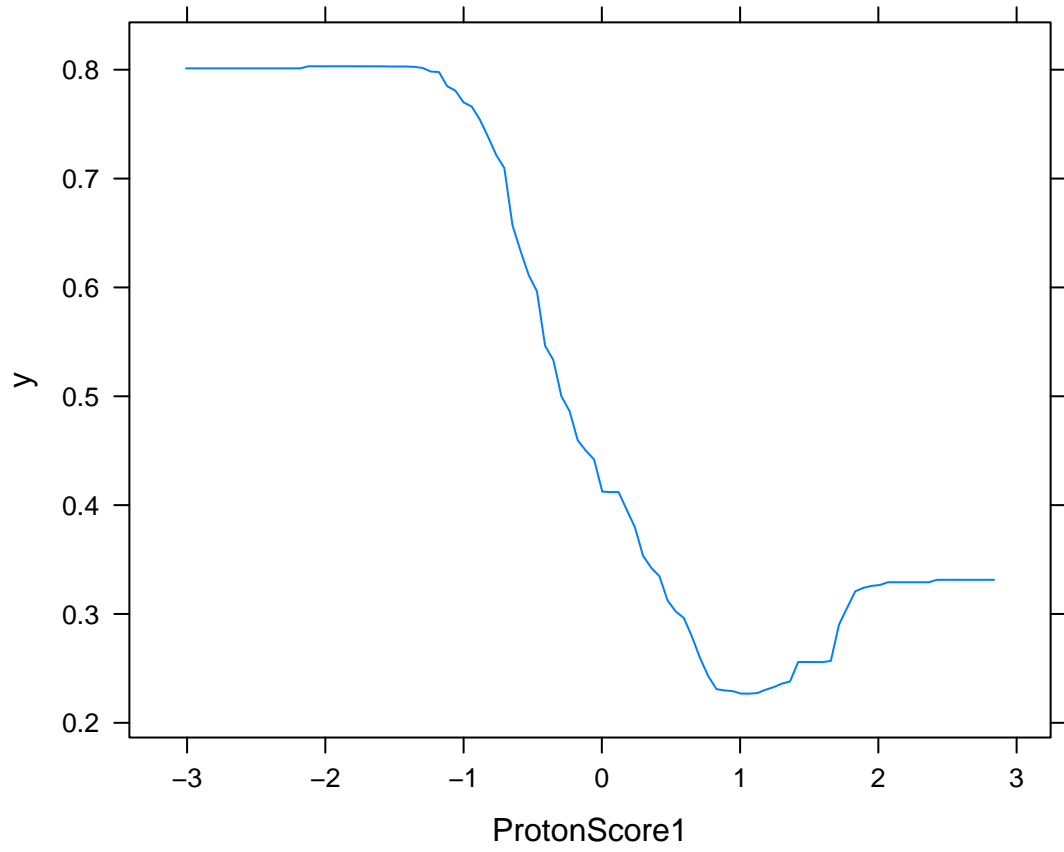
```

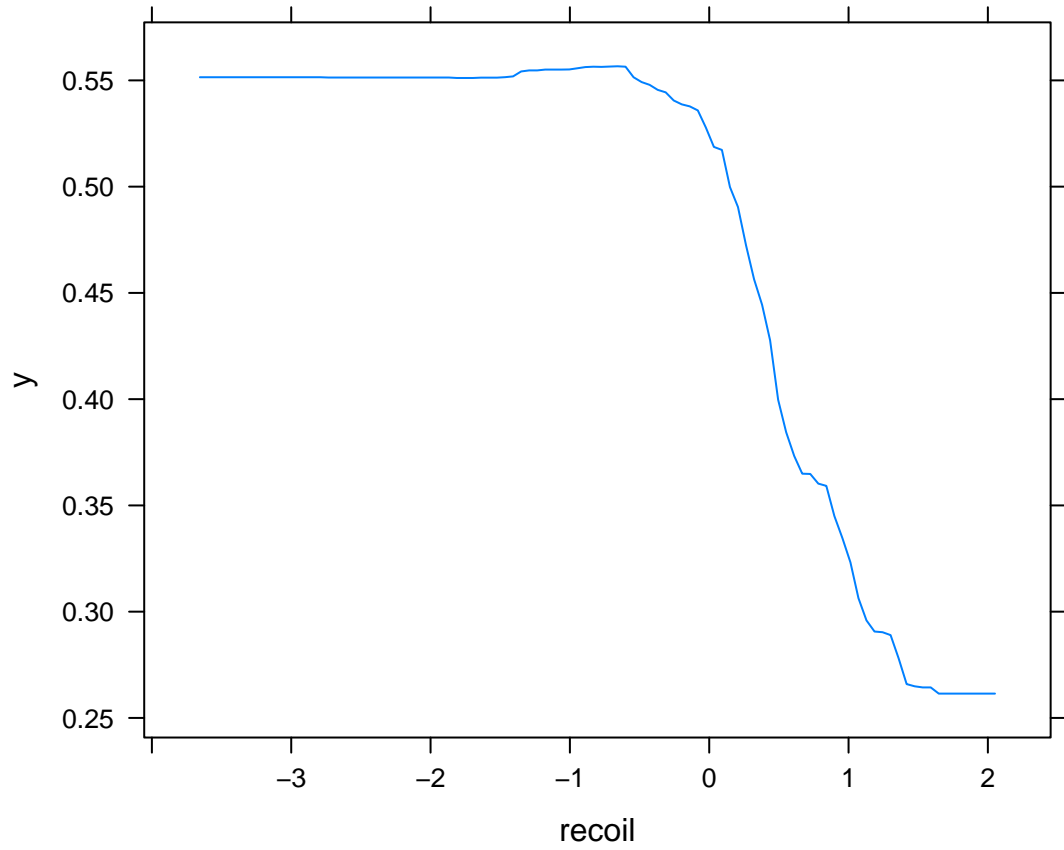


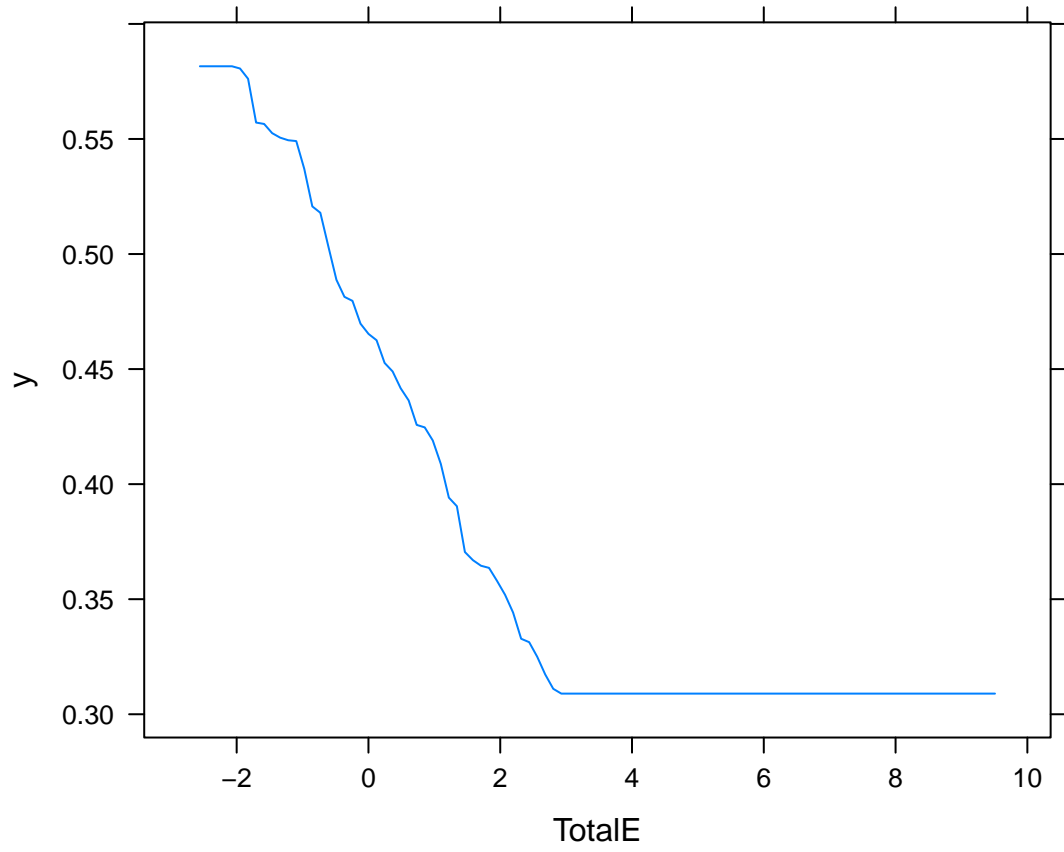


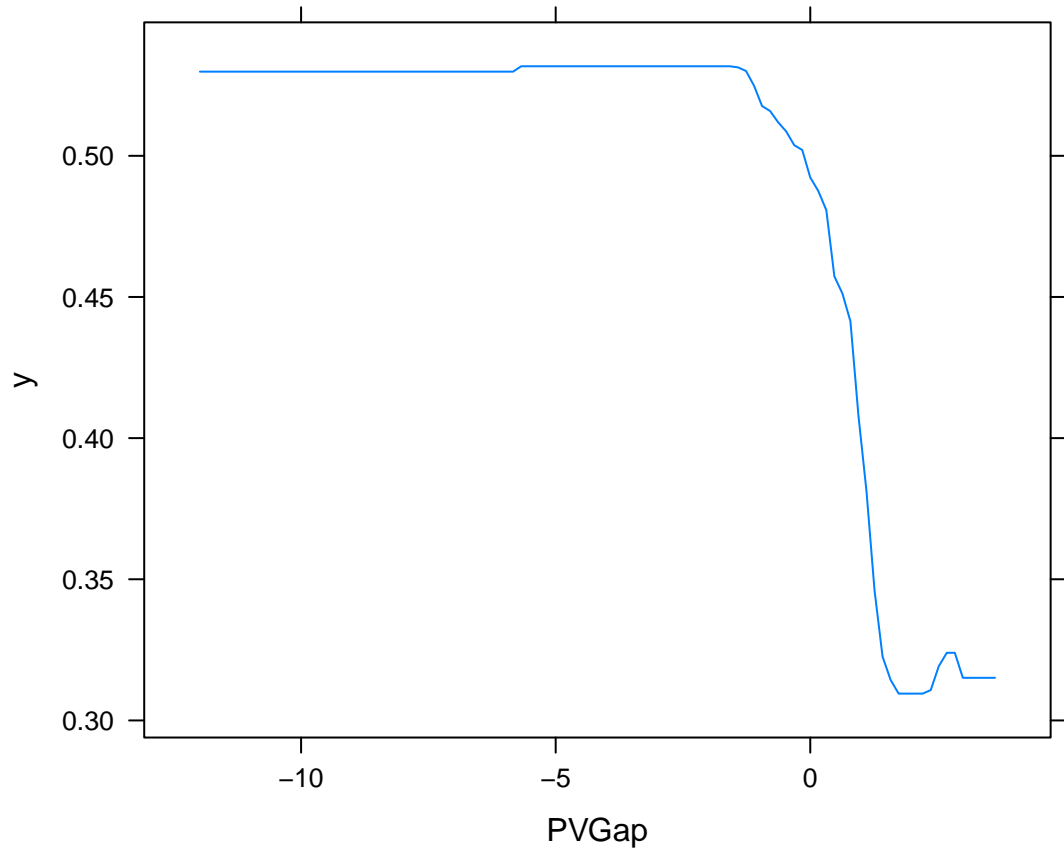


```
##          var    rel.inf
## ProtonScore1 ProtonScore1 67.6371605
## recoil      recoil 15.5698028
## PVGap       PVGap 6.5747611
## TotalE      TotalE 6.3006264
## Angle       Angle 2.8414155
## Multiplicity Multiplicity 0.9452803
## zvtx        zvtx 0.1309534
```

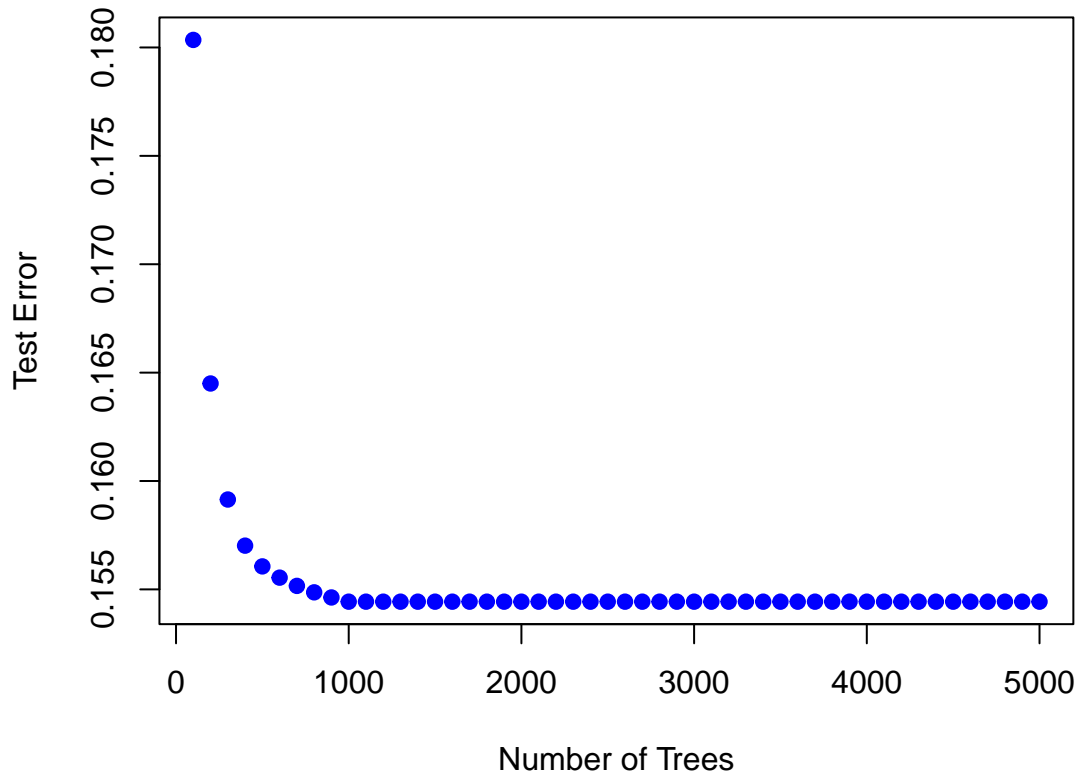








## Perfomance of Boosting on Test Set



```
##      [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##     [37] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##     [73] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##    [109] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##    [145] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##    [181] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##    [217] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##    [253] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##    [289] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##    [325] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##    [361] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##    [397] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##    [433] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##    [469] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##    [505] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##    [541] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##    [577] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##    [613] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##    [649] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##    [685] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##    [721] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##    [757] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##    [793] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

[illegible]

[illegible]



[illegible]



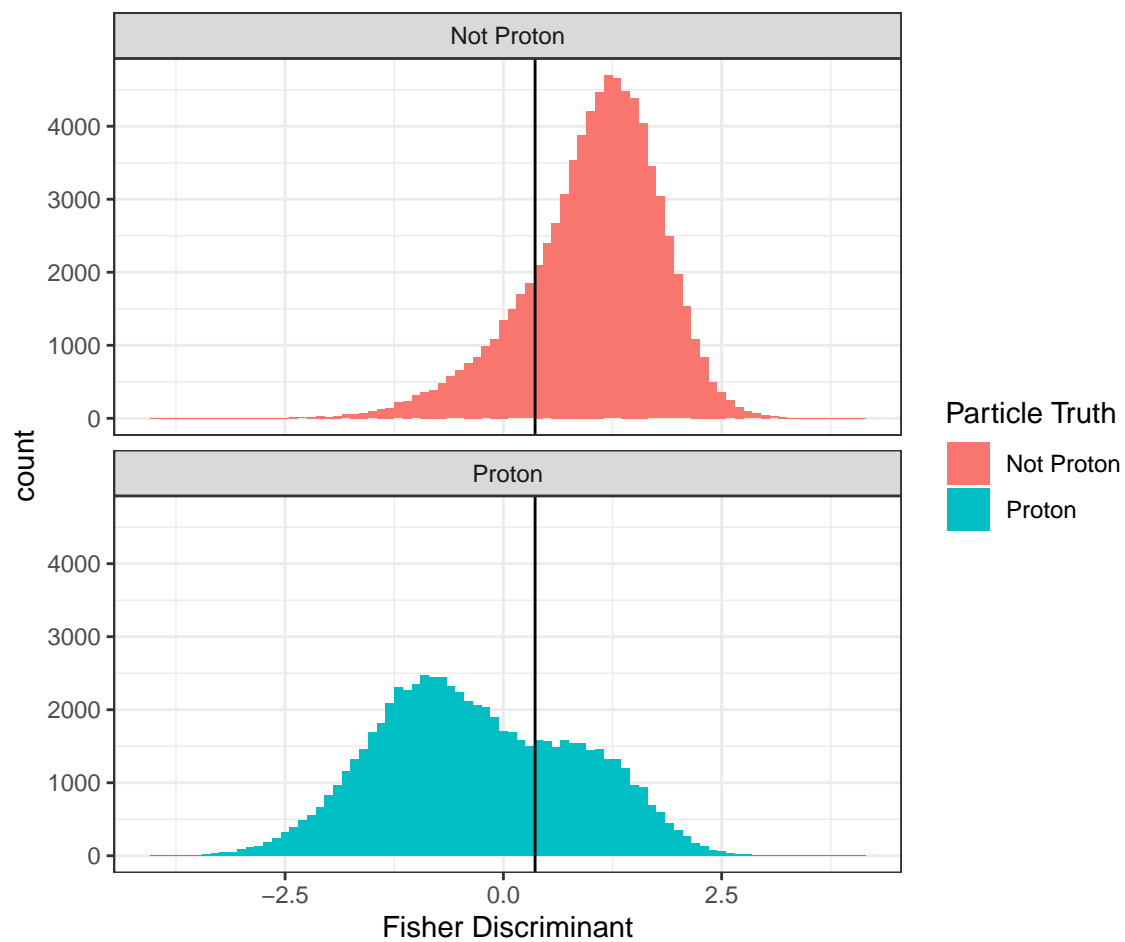


[illegible]

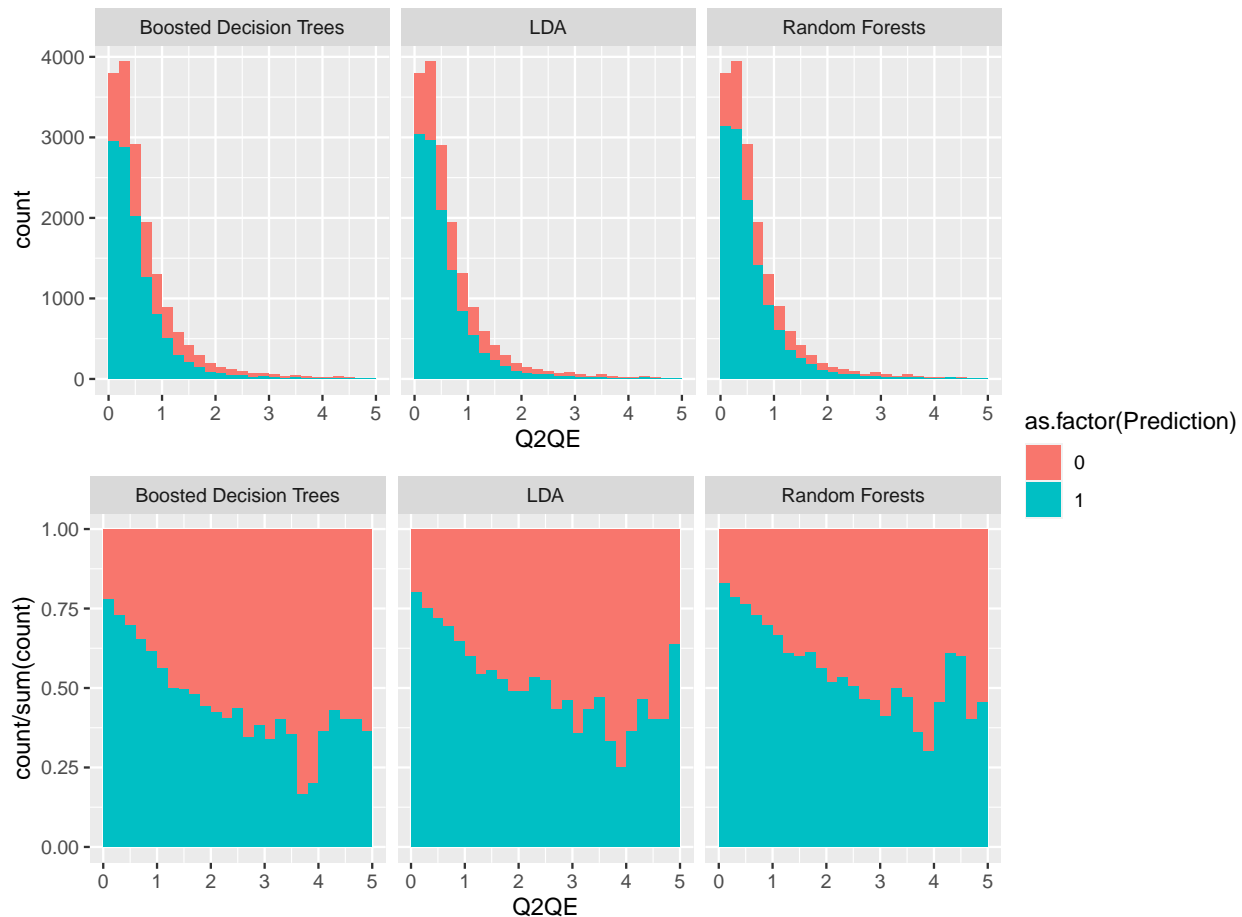
[illegible]

[illegible]









```
## # A tibble: 6 x 4
## # Groups:   Predictor, Prediction [6]
##   Predictor      Prediction    n `n/N`
##   <chr>          <dbl> <int> <dbl>
## 1 Boosted Decision Trees      0  5782 0.334
## 2 Boosted Decision Trees      1 11552 0.666
## 3 LDA                        0  5221 0.301
## 4 LDA                        1 12113 0.699
## 5 Random Forests              0  4534 0.262
## 6 Random Forests              1 12800 0.738

ggplot(events.R %>% filter(Multiplicity == 2,
                           lda_pred == 1,
                           Truth == 0),
        aes(x=ProtonScore1, y=TrackVtxGap)) +
  geom_bin2d() +
  theme_bw() +
  facet_wrap(vars(q2qe_case)) +
  scale_fill_gradient(low = "lightgray", high = "black") +
  ggtitle("Truth: Not Proton, Prediction: Proton")

ggplot(events.R %>% filter(Multiplicity == 2,
                           lda_pred == 0,
                           Truth == 1),
```

```

    aes(x=ProtonScore1, y=TrackVtxGap)) +
  geom_bin2d() +
  theme_bw() +
  facet_wrap(vars(q2qe_case)) +
  scale_fill_gradient(low = "lightgray", high = "black") +
  ggtitle("Truth: Proton, Prediction: Not Proton")

events.R %>% filter(lda_pred == 0, Truth == 1, TrackVtxGap <= 150)
events.R %>% filter(lda_pred == 1, Truth == 0, TrackVtxGap <= 150)

```