Prediction of pre-university scores

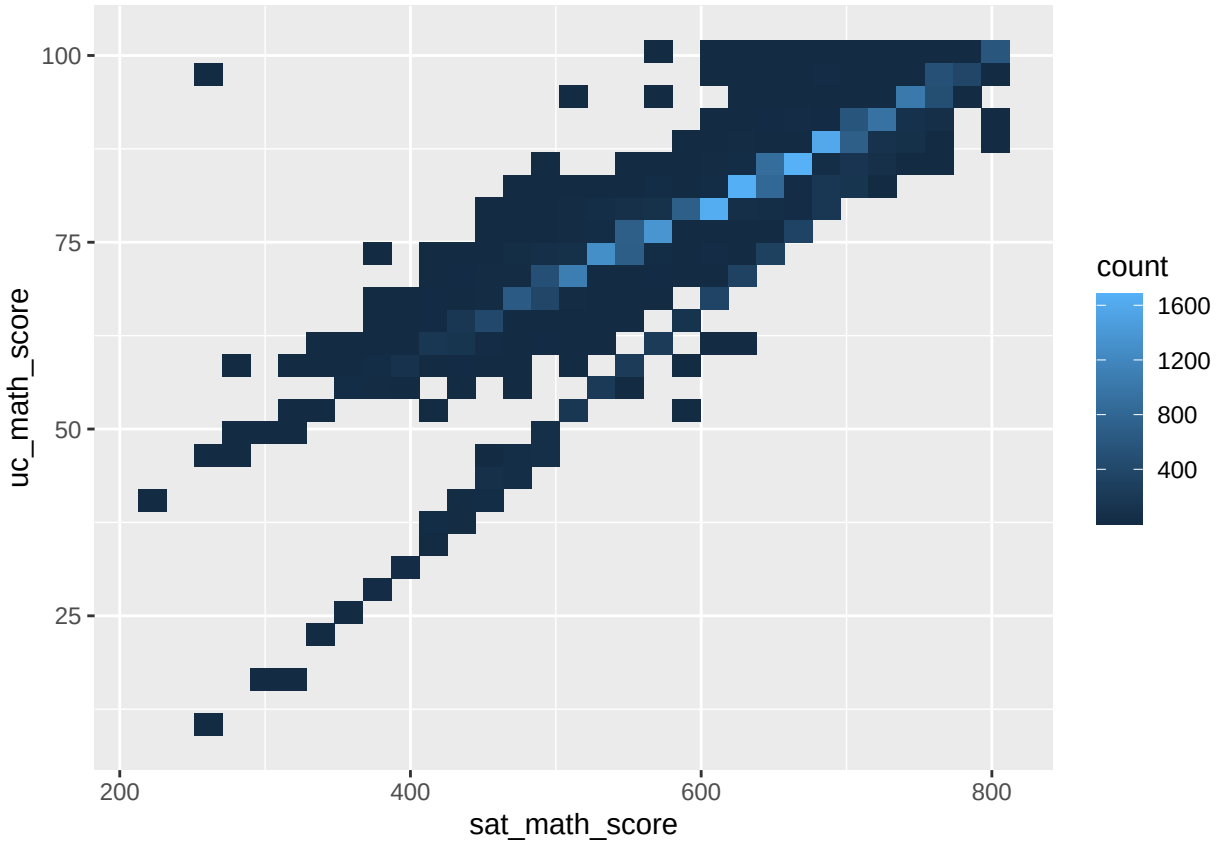
Dominik Glandorf

2022-08-16

```
student_background_data = get_student_background_data()
student_sub = get_student_sub()
bg_sub = student_background_data[student_background_data$mellon_id %in% student_sub$mellon_id,]
```

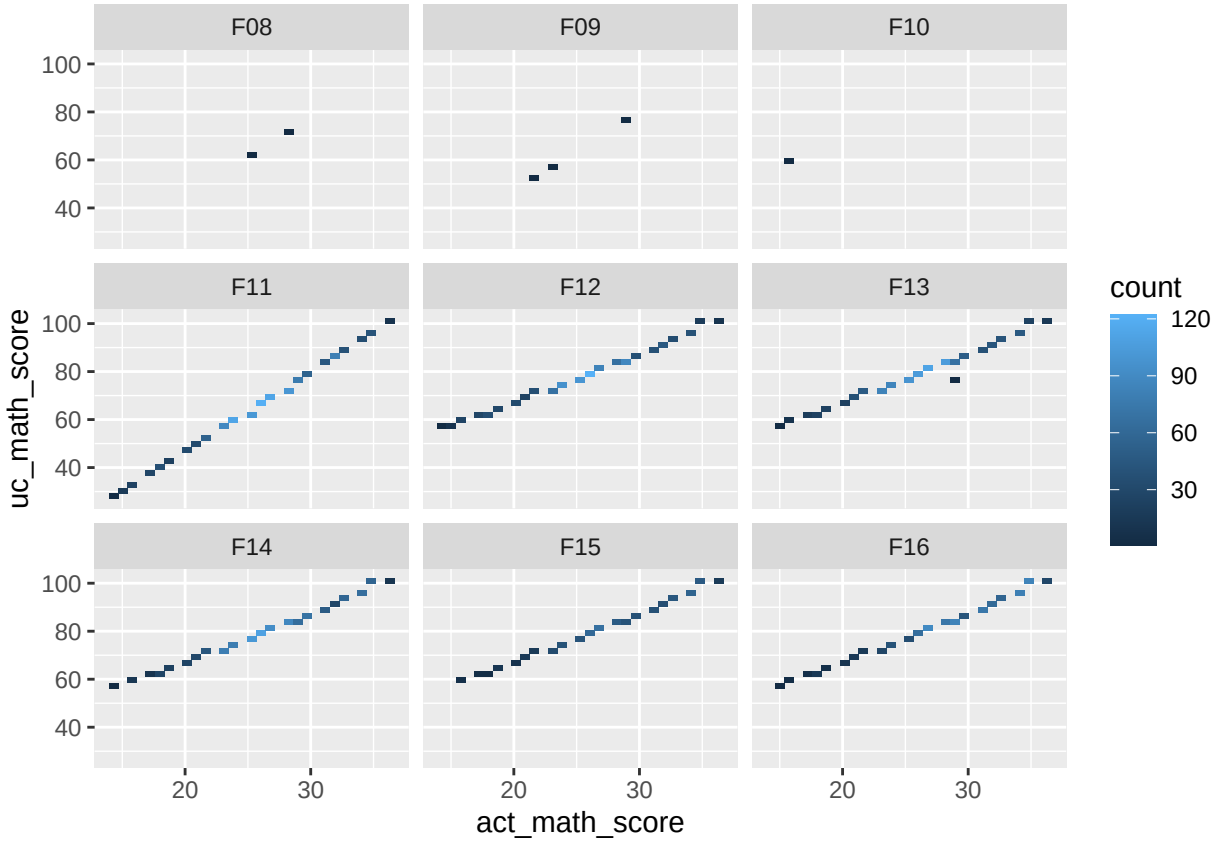
Plot UC math scores for ACT scores

```
ggplot(bg_sub, aes(x=sat_math_score, y=uc_math_score)) +
 geom_bin2d(na.rm=T)
```



Lots of noise and maybe two different mappings. Let's filter out SAT score entries and plot by admission.

```
ggplot(bg_sub[is.na(bg_sub$sat_math_score),], aes(x=act_math_score, y=uc_math_score)) +
 geom_bin2d(na.rm=T) +
  facet_wrap(vars(admitdate))
```



It looks like there was a change in the mapping starting from admitdate Fall 2012.

```
filter_pre_F12 = as.integer(substr(bg_sub$admitdate, 2,3))<12
act_only_F11 = bg_sub[filter_pre_F12&is.na(bg_sub$sat_math_score),]
table(act_only_F11$act_math_score, act_only_F11$uc_math_score)
```

```
##
##      27  30  33  37  40  43  47  50  53  57  60  63  67  70  73  77  80  83
##   14  3   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   15  0  17   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   16  0   0  24   0   0   0   0   0   0   0   1   0   0   0   0   0   0   0
##   17  0   0   0  27   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   18  0   0   0   0  37   0   0   0   0   0   0   0   0   0   0   0   0   0
##   19  0   0   0   0   0  22   0   0   0   0   0   0   0   0   0   0   0   0
##   20  0   0   0   0   0   0  33   0   0   0   0   0   0   0   0   0   0   0
##   21  0   0   0   0   0   0   0  31   0   0   0   0   0   0   0   0   0   0
##   22  0   0   0   0   0   0   0   0  55   0   0   0   0   0   0   0   0   0
##   23  0   0   0   0   0   0   0   0   0  89   0   0   0   0   0   0   0   0
##   24  0   0   0   0   0   0   0   0   0   0 112   0   0   0   0   0   0   0
##   25  0   0   0   0   0   0   0   0   0   0   0 102   0   0   0   0   0   0
##   26  0   0   0   0   0   0   0   0   0   0   0   0 120   0   0   0   0   0
##   27  0   0   0   0   0   0   0   0   0   0   0   0   0 112   0   0   0   0
```

```
##    28   0   0   0   0   0   0   0   0   0   0   0   0   0   0 101   0   0   0
##    29   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0  78   0   0
##    30   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0  55   0
##    31   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0  45
##    32   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##    33   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##    34   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##    35   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##    36   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##
##       86  87  90  93  97 100
##    14   0   0   0   0   0   0
##    15   0   0   0   0   0   0
##    16   0   0   0   0   0   0
##    17   0   0   0   0   0   0
##    18   0   0   0   0   0   0
##    19   0   0   0   0   0   0
##    20   0   0   0   0   0   0
##    21   0   0   0   0   0   0
##    22   0   0   0   0   0   0
##    23   0   0   0   0   0   0
##    24   0   0   0   0   0   0
##    25   0   0   0   0   0   0
##    26   0   0   0   0   0   0
##    27   0   0   0   0   0   0
##    28   0   0   0   0   0   0
##    29   0   0   0   0   0   0
##    30   0   0   0   0   0   0
##    31   0   0   0   0   0   0
##    32   0  80   0   0   0   0
##    33   0   0  39   0   0   0
##    34   0   0   0  35   0   0
##    35   0   0   0   0  41   0
##    36   0   0   0   0   0   9
```

There is a clear mapping for admitdate=F11.

```
filter_post_F12 = as.integer(substr(bg_sub$admitdate, 2,3))>=12
act_only_postF12 = bg_sub[filter_post_F12 & is.na(bg_sub$sat_math_score),]
table(act_only_postF12$act_math_score, act_only_postF12$uc_math_score)
```

```
##
##       40  56  58  60  62  63  65  67  69  70  71  73  75  77  79  81  83  85
##    14   0   3   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##    15   0   0  12   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##    16   0   0   0  59   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##    17   0   0   0   0  68   0   0   0   0   0   0   0   0   0   0   0   0   0
##    18   0   0   0   0   0 106   0   0   0   0   0   0   0   0   0   0   0   0
##    19   0   0   0   0   0   0  93   0   0   0   0   0   0   0   0   0   0   0
##    20   0   0   0   0   0   0   0 102   0   0   0   0   0   0   0   0   0   0
##    21   0   0   0   0   0   0   0   0 122   0   0   0   0   0   0   0   0   0
##    22   0   0   0   0   0   0   0   0   0   0 161   0   0   0   0   0   0   0
##    23   0   0   0   0   0   0   0   0   0   0   0 291   0   0   0   0   0   0
```

```
##   24   0   0   0   0   0   0   0   0   0   0   0   0 338   0   0   0   0   0
##   25   0   0   0   0   0   0   0   0   0   0   0   0   0 374   0   0   0   0
##   26   0   0   0   0   0   0   0   0   0   0   0   0   0   0 460   0   0   0
##   27   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 442   0   0
##   28   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 379   0
##   29   0   0   0   0   0   0   0   0   0   0   0   0   0   1   0   0   0 342
##   30   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   31   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   32   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   33   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   34   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   35   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   36   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##
##       87  89  92  94  97 100
##   14   0   0   0   0   0   0
##   15   0   0   0   0   0   0
##   16   0   0   0   0   0   0
##   17   0   0   0   0   0   0
##   18   0   0   0   0   0   0
##   19   0   0   0   0   0   0
##   20   0   0   0   0   0   0
##   21   0   0   0   0   0   0
##   22   0   0   0   0   0   0
##   23   0   0   0   0   0   0
##   24   0   0   0   0   0   0
##   25   0   0   0   0   0   0
##   26   0   0   0   0   0   0
##   27   0   0   0   0   0   0
##   28   0   0   0   0   0   0
##   29   0   0   0   0   0   0
##   30 238   0   0   0   0   0
##   31   0 235   0   0   0   0
##   32   0   0 215   0   0   0
##   33   0   0   0 234   0   0
##   34   0   0   0   0 285   0
##   35   0   0   0   0   0 228
##   36   0   0   0   0   0  82
```

Let's create the mapping.

```
uc_scores_11 = unique(act_only_F11$uc_math_score)
uc_scores_11 = uc_scores_11[!is.na(uc_scores_11)][-24] # remove 86
uc_scores_post12 = unique(act_only_postF12$uc_math_score)
uc_scores_post12 = uc_scores_post12[!is.na(uc_scores_post12)]
uc_scores_post12 = c(uc_scores_post12[order(uc_scores_post12)][-c(1, 10)], 100) # remove 40, 70, add 10
uc_act = data.frame(act=14:36, uc_pre_12=uc_scores_11[order(uc_scores_11)], uc_post_12=uc_scores_post12)
uc_act
```

```
##    act uc_pre_12 uc_post_12
## 1   14        27         56
## 2   15        30         58
## 3   16        33         60
```

```
## 4   17        37         62
## 5   18        40         63
## 6   19        43         65
## 7   20        47         67
## 8   21        50         69
## 9   22        53         71
## 10  23        57         73
## 11  24        60         75
## 12  25        63         77
## 13  26        67         79
## 14  27        70         81
## 15  28        73         83
## 16  29        77         85
## 17  30        80         87
## 18  31        83         89
## 19  32        87         92
## 20  33        90         94
## 21  34        93         97
## 22  35        97        100
## 23  36       100        100
```

```
lm(uc_pre_12~act,uc_act)
```

```
##
## Call:
## lm(formula = uc_pre_12 ~ act, data = uc_act)
##
## Coefficients:
## (Intercept)          act
##     -19.953        3.332
```

```
uc_act$uc_pre_12 == round(uc_act$act*10/3-20)
```

```
##  [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [16] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
```

```
lm(uc_post_12~act,uc_act)
```

```
##
## Call:
## lm(formula = uc_post_12 ~ act, data = uc_act)
##
## Coefficients:
## (Intercept)          act
##      26.632        2.036
```

```
uc_act$uc_post_12 == round(uc_act$act*2.036+26.63)
```

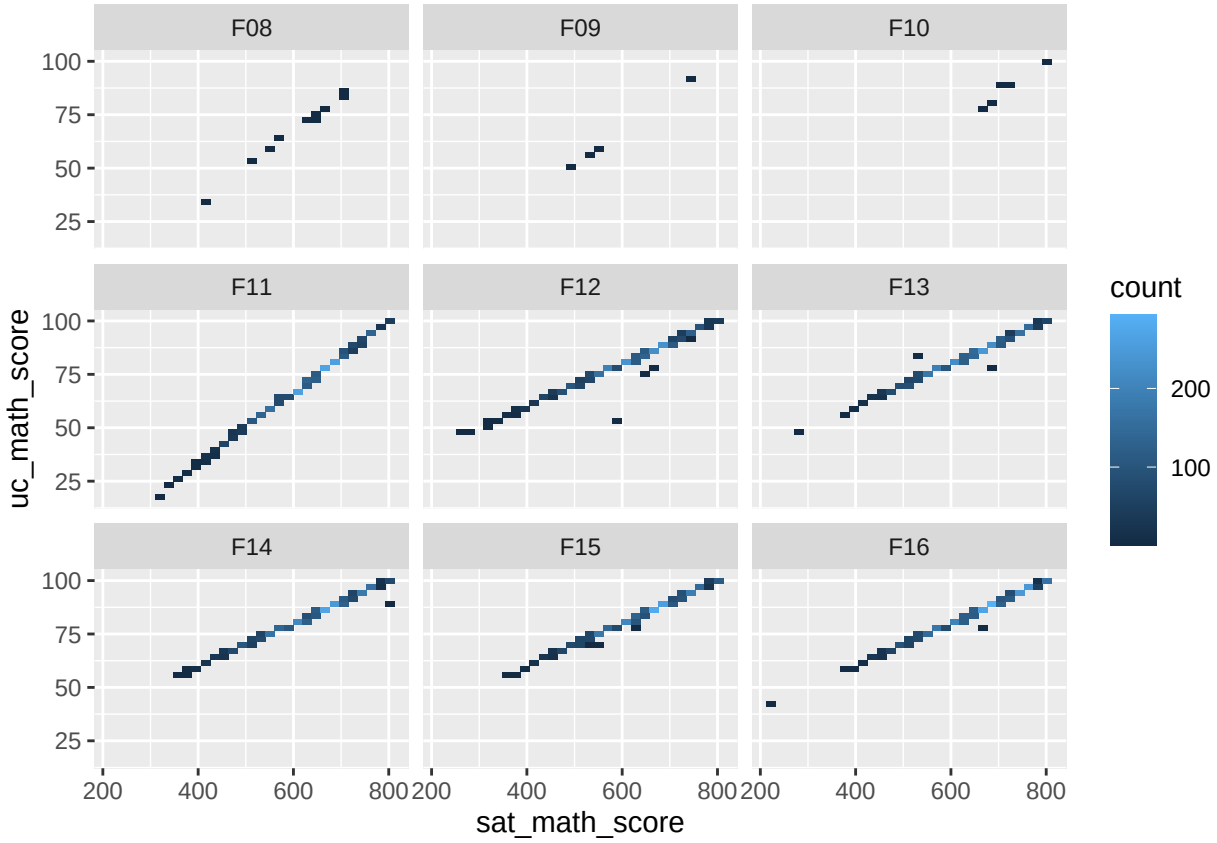
```
##  [1] FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE
## [13] FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE  TRUE
```

This does not look linear enough. Let's then print the mapping for the config.

```
## [1] "data.frame(act=c(14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36),uc_pre_1
```

Plot SAT to UC math score by admission

```
ggplot(bg_sub[is.na(bg_sub$act_math_score),], aes(x=sat_math_score, y=uc_math_score)) +
 geom_bin2d(na.rm=T) +
  facet_wrap(vars(admitdate))
```



Just a little bit of noise but also a change between 11 and 12. So, let's also separate for the admit data to find the mapping.

```
options(max.print=3000)
sat_only_pre_F12 = bg_sub[filter_pre_F12&is.na(bg_sub$act_math_score),]
mappings = table(sat_only_pre_F12$sat_math_score, sat_only_pre_F12$uc_math_score)
mappings
```

```
##
##       18  23  25  27  28  30  32  33  35  37  38  40  42  43  45  47  48  50
##   300  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   310  2   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   330  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   340  0   2   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   350  0   0   4   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   360  0   0   0   5   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   370  0   0   0   0   5   0   0   0   0   0   0   0   0   0   0   0   0   0
##   380  0   0   0   0   0   4   0   0   0   0   0   0   0   0   0   0   0   0
##   390  0   0   0   0   0   0   5   0   0   0   0   0   0   0   0   0   0   0
```

```
##   400   0   0   0   0   0   0   0   6   0   0   0   0   0   0   0   0   0   0
##   410   0   0   0   0   0   0   0   0  12   0   0   0   0   0   0   0   0   0
##   420   0   0   0   0   0   0   0   0   0  15   0   0   0   0   0   0   0   0
##   430   0   0   0   0   0   0   0   0   0   0  11   0   0   0   0   0   0   0
##   440   0   0   0   0   0   0   0   0   0   0   0  16   0   0   0   0   0   0
##   450   0   0   0   0   0   0   0   0   0   0   0   0  19   0   0   0   0   0
##   460   0   0   0   0   0   0   0   0   0   0   0   0   0  42   0   0   0   0
##   470   0   0   0   0   0   0   0   0   0   0   0   0   0   0  34   0   0   0
##   480   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0  26   0   0
##   490   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0  39   0
##   500   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0  39
##   510   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   520   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   530   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   540   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   550   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   560   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   570   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   580   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   590   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   600   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   610   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   620   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   630   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   640   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   650   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   660   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   670   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   680   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   690   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   700   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   710   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   720   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   730   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   740   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   750   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   760   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   770   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   780   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   790   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   800   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##
##        52  53  55  57  58  60  62  63  65  67  68  70  72  73  75  77  78  80
##   300   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   310   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   330   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   340   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   350   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   360   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   370   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   380   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   390   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   400   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   410   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
```

```
##   420   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   430   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   440   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   450   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   460   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   470   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   480   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   490   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   500   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   510  50   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   520   0  58   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   530   0   0  48   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   540   0   0   0  88   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   550   0   0   0   0  64   0   0   0   0   0   0   0   0   0   0   0   0   0
##   560   0   0   0   0   0  97   0   0   0   0   0   0   0   0   0   0   0   0
##   570   0   0   0   0   0   0  76   0   0   0   0   0   0   0   0   0   0   0
##   580   0   0   0   0   0   0   0  79   0   0   0   0   0   0   0   0   0   0
##   590   0   0   0   0   0   0   0   0  80   0   0   0   0   0   0   0   0   0
##   600   0   0   0   0   0   0   0   0   0 103   0   0   0   0   0   0   0   0
##   610   0   0   0   0   0   0   0   0   0   0 157   0   0   0   0   0   0   0
##   620   0   0   0   0   0   0   0   0   0   0   0 112   0   0   0   0   0   0
##   630   0   0   0   0   0   0   0   0   0   0   0   0 131   0   0   0   0   0
##   640   0   0   0   0   0   0   0   0   0   0   0   0   0 123   0   0   0   0
##   650   0   0   0   0   0   0   0   0   0   0   0   0   0   0 113   0   0   0
##   660   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 137   0   0
##   670   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 130   0
##   680   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 127
##   690   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   700   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   710   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   720   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   730   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   740   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   750   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   760   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   770   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   780   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   790   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   800   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##
##        82  83  85  86  87  88  90  92  93  95  97  98 100
##   300   0   0   0   0   0   0   0   0   0   0   0   0   0
##   310   0   0   0   0   0   0   0   0   0   0   0   0   0
##   330   0   0   0   0   0   0   0   0   0   0   0   0   0
##   340   0   0   0   0   0   0   0   0   0   0   0   0   0
##   350   0   0   0   0   0   0   0   0   0   0   0   0   0
##   360   0   0   0   0   0   0   0   0   0   0   0   0   0
##   370   0   0   0   0   0   0   0   0   0   0   0   0   0
##   380   0   0   0   0   0   0   0   0   0   0   0   0   0
##   390   0   0   0   0   0   0   0   0   0   0   0   0   0
##   400   0   0   0   0   0   0   0   0   0   0   0   0   0
##   410   0   0   0   0   0   0   0   0   0   0   0   0   0
##   420   0   0   0   0   0   0   0   0   0   0   0   0   0
##   430   0   0   0   0   0   0   0   0   0   0   0   0   0
```

```
##   440   0   0   0   0   0   0   0   0   0   0   0   0   0
##   450   0   0   0   0   0   0   0   0   0   0   0   0   0
##   460   0   0   0   0   0   0   0   0   0   0   0   0   0
##   470   0   0   0   0   0   0   0   0   0   0   0   0   0
##   480   0   0   0   0   0   0   0   0   0   0   0   0   0
##   490   0   0   0   0   0   0   0   0   0   0   0   0   0
##   500   0   0   0   0   0   0   0   0   0   0   0   0   0
##   510   0   0   0   0   0   0   0   0   0   0   0   0   0
##   520   0   0   0   0   0   0   0   0   0   0   0   0   0
##   530   0   0   0   0   0   0   0   0   0   0   0   0   0
##   540   0   0   0   0   0   0   0   0   0   0   0   0   0
##   550   0   0   0   0   0   0   0   0   0   0   0   0   0
##   560   0   0   0   0   0   0   0   0   0   0   0   0   0
##   570   0   0   0   0   0   0   0   0   0   0   0   0   0
##   580   0   0   0   0   0   0   0   0   0   0   0   0   0
##   590   0   0   0   0   0   0   0   0   0   0   0   0   0
##   600   0   0   0   0   0   0   0   0   0   0   0   0   0
##   610   0   0   0   0   0   0   0   0   0   0   0   0   0
##   620   0   0   0   0   0   0   0   0   0   0   0   0   0
##   630   0   0   0   0   0   0   0   0   0   0   0   0   0
##   640   0   0   0   0   0   0   0   0   0   0   0   0   0
##   650   0   0   0   0   0   0   0   0   0   0   0   0   0
##   660   0   0   0   0   0   0   0   0   0   0   0   0   0
##   670   0   0   0   0   0   0   0   0   0   0   0   0   0
##   680   0   0   0   0   0   0   0   0   0   0   0   0   0
##   690 131   0   0   0   0   0   0   0   0   0   0   0   0
##   700   0 112   0   0   0   0   1   0   0   0   0   0   0
##   710   0   0  97   0   0   0   0   0   0   0   0   0   0
##   720   0   0   0   0  45   0   0   0   0   0   0   0   0
##   730   0   0   0   0   0  71   0   0   0   0   0   0   0
##   740   0   0   0   0   0   0  56   0   0   0   0   0   0
##   750   0   0   0   0   0   0   0  68   0   0   0   0   0
##   760   0   0   0   0   0   0   0   0  47   0   0   0   0
##   770   0   0   0   0   0   0   0   0   0  87   0   0   0
##   780   0   0   0   0   0   0   0   0   0   0  25   0   0
##   790   0   0   0   0   0   0   0   0   0   0   0   4   0
##   800   0   0   0   0   0   0   0   0   0   0   0   0  46
```

```
mappings['700','90']=0 # fix the only outlier
uc_sat_pre_12 = data.frame(sat = as.integer(rownames(mappings)[(row(mappings)[which(mappings != 0)])]),
                           uc = as.integer(colnames(mappings)[(col(mappings)[which(mappings != 0)])]))
uc_sat_pre_12
```

```
##    sat  uc
## 1  310  18
## 2  340  23
## 3  350  25
## 4  360  27
## 5  370  28
## 6  380  30
## 7  390  32
## 8  400  33
## 9  410  35
## 10 420  37
```

```
## 11 430  38
## 12 440  40
## 13 450  42
## 14 460  43
## 15 470  45
## 16 480  47
## 17 490  48
## 18 500  50
## 19 510  52
## 20 520  53
## 21 530  55
## 22 540  57
## 23 550  58
## 24 560  60
## 25 570  62
## 26 580  63
## 27 590  65
## 28 600  67
## 29 610  68
## 30 620  70
## 31 630  72
## 32 640  73
## 33 650  75
## 34 660  77
## 35 670  78
## 36 680  80
## 37 690  82
## 38 700  83
## 39 710  85
## 40 720  87
## 41 730  88
## 42 740  90
## 43 750  92
## 44 760  93
## 45 770  95
## 46 780  97
## 47 790  98
## 48 800 100
```

Try to find a formula.

```
lm(uc ~ sat, uc_sat_pre_12)
```

```
##
## Call:
## lm(formula = uc ~ sat, data = uc_sat_pre_12)
##
## Coefficients:
## (Intercept)          sat
##    -33.4138       0.1668
```

```
uc_sat_pre_12$uc == round((uc_sat_pre_12$sat - 200) / 6)
```

```
##  [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [16] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [31] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [46] TRUE TRUE TRUE
```

```
sat_only_post_F12 = bg_sub[filter_post_F12&is.na(bg_sub$act_math_score),]
mappings = table(sat_only_post_F12$sat_math_score, sat_only_post_F12$uc_math_score)
mappings
```

```
##
##       40  42  47  48  49  51  52  53  54  55  56  57  58  59  60  61  62  63
##   220  0   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   270  0   0   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   280  0   0   0   3   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   290  0   0   0   0   2   0   0   0   0   0   0   0   0   0   0   0   0   0
##   310  0   0   0   0   0   3   0   0   0   0   0   0   0   0   0   0   0   0
##   320  0   0   0   0   0   0   1   0   0   0   0   0   0   0   0   0   0   0
##   330  0   0   0   0   0   0   0   2   0   0   0   0   0   0   0   0   0   0
##   340  0   0   0   0   0   0   0   0   4   0   0   0   0   0   0   0   0   0
##   350  0   0   0   0   0   0   0   0   0   5   0   0   0   0   0   0   0   0
##   360  0   0   0   0   0   0   0   0   0   0  10   0   0   0   0   0   0   0
##   370  0   0   0   0   0   0   0   0   0   0   0  10   0   0   0   0   0   0
##   380  0   0   0   0   0   0   0   0   0   0   0   0  17   0   0   0   0   0
##   390  0   0   0   0   0   0   0   0   0   0   0   0   0  18   0   0   0   0
##   400  0   0   0   0   0   0   0   0   0   0   0   0   0   0  43   0   0   0
##   410  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0  40   0   0
##   420  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0  51   0
##   430  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0  72
##   440  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   450  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   460  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   470  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   480  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   490  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   500  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   510  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   520  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   530  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   540  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   550  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   560  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   570  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   580  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   590  0   0   0   0   0   0   0   1   0   0   0   0   0   0   0   0   0   0
##   600  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   610  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   620  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   630  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   640  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   650  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
```

```
##   660   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   670   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   680   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   690   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   700   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   710   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   720   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   730   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   740   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   750   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   760   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   770   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   780   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   790   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   800   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##
##        64  65  66  67  68  69  70  71  72  73  74  75  76  77  78  79  80  81
##   220   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   270   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   280   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   290   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   310   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   320   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   330   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   340   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   350   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   360   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   370   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   380   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   390   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   400   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   410   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   420   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   430   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   440  92   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   450   0  99   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   460   0   0 144   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   470   0   0   0 188   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   480   0   0   0   0 171   0   0   0   0   0   0   0   0   0   0   0   0   0
##   490   0   0   0   0   0 199   0   0   0   0   0   0   0   0   0   0   0   0
##   500   0   0   0   0   0   0 258   0   0   0   0   0   0   0   0   0   0   0
##   510   0   0   0   0   0   0   0 324   0   0   0   0   0   0   0   0   0   0
##   520   0   0   0   0   0   0   0   0 300   0   0   0   0   0   0   0   0   0
##   530   0   0   0   0   0   0   0   0   0 355   0   0   0   0   0   0   0   0
##   540   0   0   0   0   0   0   0   1   0   0 348   0   0   0   0   0   0   0
##   550   0   0   0   0   0   1   0   0   0   0   0 393   0   0   0   0   0   0
##   560   0   0   0   0   0   0   0   0   0   0   0   0 426   0   0   0   0   0
##   570   0   0   0   0   0   0   0   0   0   0   0   0   0 405   0   0   0   0
##   580   0   0   0   0   0   0   0   0   0   0   0   0   0   0 430   0   0   0
##   590   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 441   0   0
##   600   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 511   0
##   610   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 554
##   620   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   630   0   0   0   0   0   0   0   0   0   0   0   0   0   0   1   0   0   0
```

```
##     640   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     650   0   0   0   0   0   0   0   0   0   0   0   1   0   0   0   0   0   0
##     660   0   0   0   0   0   0   0   0   0   0   0   0   0   1   0   1   0   0
##     670   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     680   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     690   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   1   0   0
##     700   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     710   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     720   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     730   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     740   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     750   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     760   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     770   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     780   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     790   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     800   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##
##         82  83  84  85  86  87  88  89  90  91  92  93  94  95  96  97  98  99
##     220   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     270   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     280   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     290   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     310   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     320   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     330   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     340   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     350   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     360   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     370   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     380   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     390   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     400   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     410   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     420   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     430   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     440   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     450   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     460   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     470   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     480   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     490   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     500   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     510   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     520   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     530   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     540   0   0   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     550   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     560   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     570   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     580   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     590   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     600   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##     610   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
```

```
##   620 577   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   630   0 566   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   640   0   0 588   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   650   0   0   0 599   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   660   0   0   0   0 634   0   0   0   0   0   0   0   0   0   0   0   0   0
##   670   0   0   0   0   0 639   0   0   0   0   0   0   0   0   0   0   0   0
##   680   0   0   0   0   0   0 655   0   0   0   0   0   0   0   0   0   0   0
##   690   0   0   0   0   0   0   0 590   0   0   0   0   0   0   0   0   0   0
##   700   0   0   0   0   0   0   0   0 582   0   0   0   0   0   0   0   0   0
##   710   0   0   0   0   0   0   0   0   0 476   0   0   0   0   0   0   0   0
##   720   0   0   0   0   0   0   0   0   0   0 452   0   0   0   0   0   0   0
##   730   0   0   0   0   0   0   0   0   0   0   0 324   0   0   0   0   0   0
##   740   0   0   0   0   0   0   0   0   0   0   0   0 526   0   0   0   0   0
##   750   0   0   0   0   0   0   0   0   0   0   1   0   0 369   0   0   0   0
##   760   0   0   0   0   0   0   0   0   0   0   0   0   0   0 349   0   0   0
##   770   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 446   0   0
##   780   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 169   0
##   790   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 143
##   800   0   0   0   0   0   0   1   0   0   0   0   0   0   0   0   0   0   0
##
##       100
##   220   0
##   270   0
##   280   0
##   290   0
##   310   0
##   320   0
##   330   0
##   340   0
##   350   0
##   360   0
##   370   0
##   380   0
##   390   0
##   400   0
##   410   0
##   420   0
##   430   0
##   440   0
##   450   0
##   460   0
##   470   0
##   480   0
##   490   0
##   500   0
##   510   0
##   520   0
##   530   0
##   540   0
##   550   0
##   560   0
##   570   0
##   580   0
##   590   0
```

```
##   600   0
##   610   0
##   620   0
##   630   0
##   640   0
##   650   0
##   660   0
##   670   0
##   680   0
##   690   0
##   700   0
##   710   0
##   720   0
##   730   0
##   740   0
##   750   0
##   760   0
##   770   0
##   780   0
##   790   0
##   800 459
```

```
 # fix outliers
mappings['590','53']=0
mappings['550','69']=0
mappings['540','71']=0
mappings['540','84']=0
mappings['630','78']=0
mappings['650','75']=0
mappings['660','77']=0
mappings['660','79']=0
mappings['690','79']=0
mappings['750','92']=0
mappings['800','88']=0
#mappings
```

```
uc_sat_post_12 = data.frame(sat = as.integer(rownames(mappings)[(row(mappings)[which(mappings != 0)])])
                            uc = as.integer(colnames(mappings)[(col(mappings)[which(mappings != 0)])]))
uc_sat_post_12$uc == (uc_sat_post_12$sat + 200) / 10
```

```
##  [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [16] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [31] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [46] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
```

Prediction of math score from SAT, ACT and Combination

- SAT scores to UC are two affine transformations:

```
print(get_uc_from_sat)
```

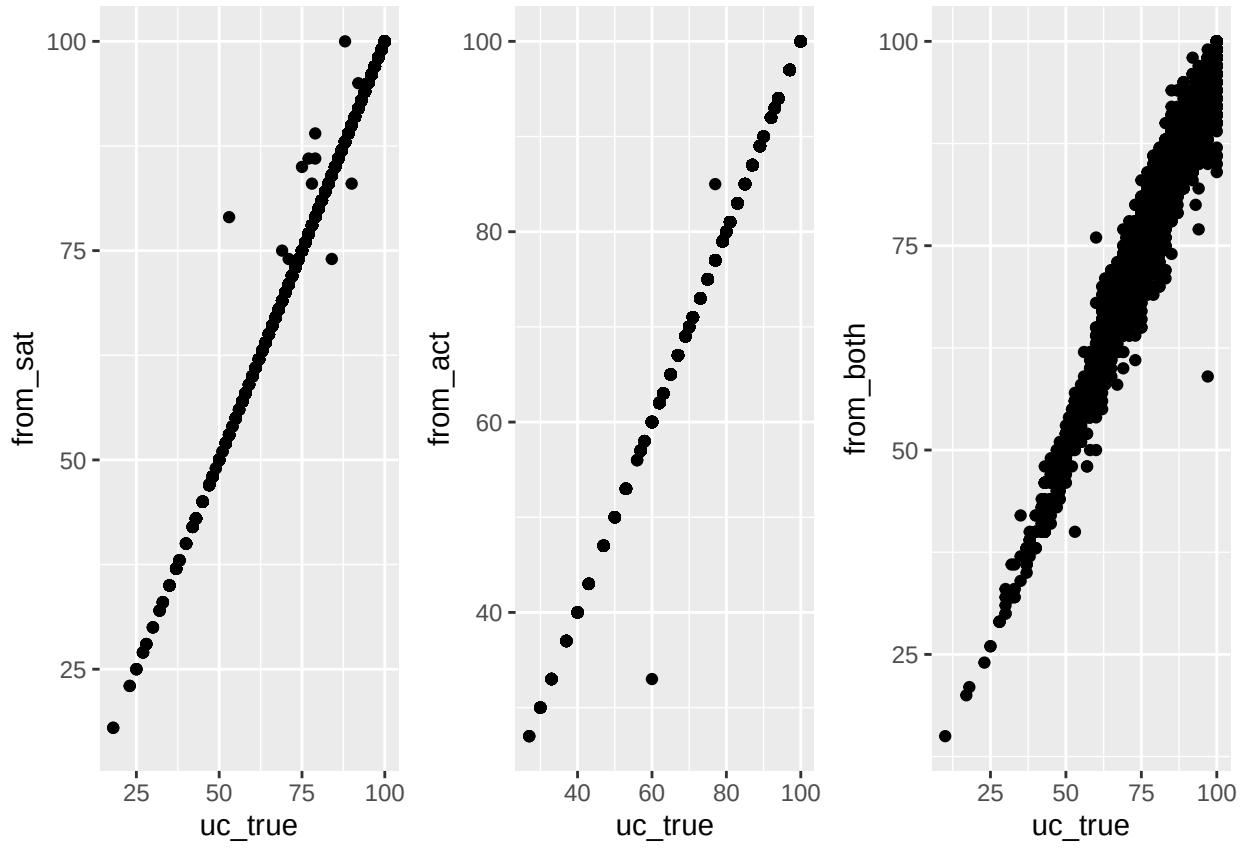
```
## function (sat, pre_12 = F)
## {
##     if (pre_12)
##         return(round((sat - 200)/6))
##     return(round((sat + 200)/10))
## }
```

- ACT follows an almost-linear mapping

```
uc_math_from_sat = mapply(get_uc_from_sat, bg_sub$sat_math_score, filter_pre_F12)
uc_math_from_act = mapply(get_uc_from_act, bg_sub$act_math_score, filter_pre_F12)
uc_math_from_sat_and_act = round(0.75 * uc_math_from_sat + 0.25 * uc_math_from_act)
```

Plot predictions

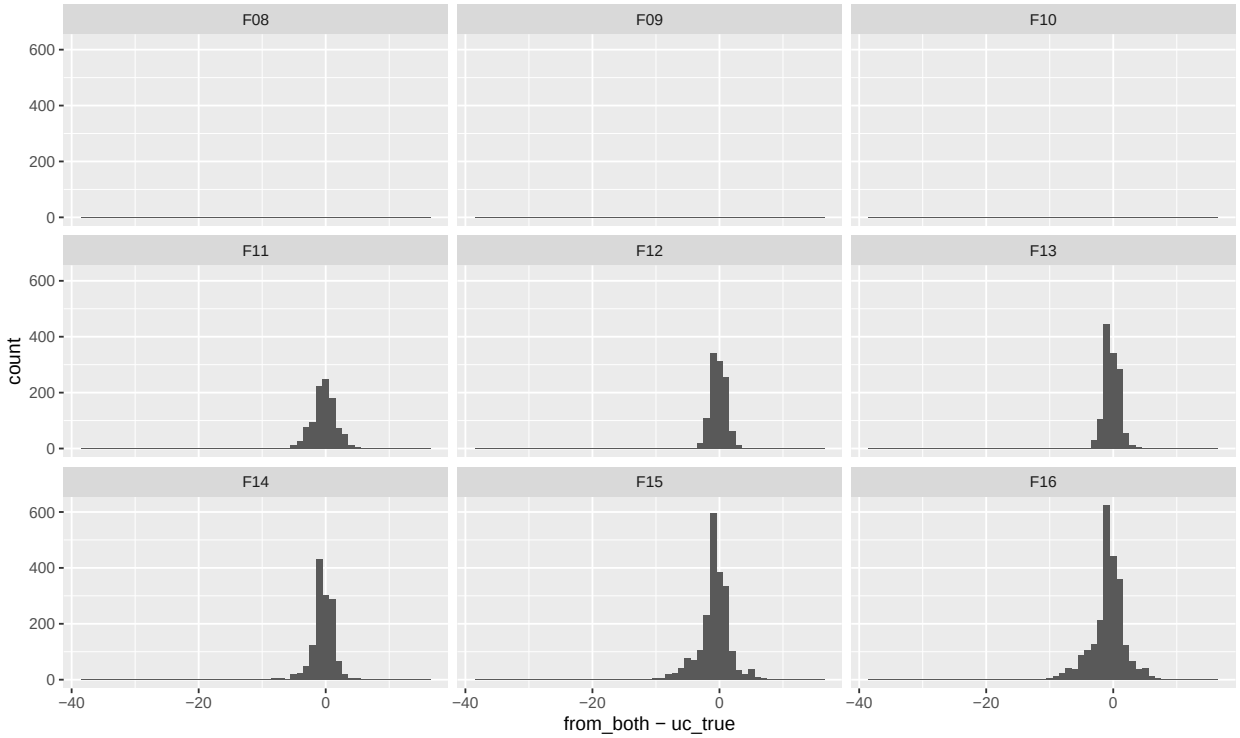
```
uc_math_scores = data.frame(uc_true=bg_sub$uc_math_score,
                            from_sat=uc_math_from_sat,
                            from_act=uc_math_from_act,
                            from_both=uc_math_from_sat_and_act,
                            admitdate=bg_sub$admitdate)
p1 = ggplot(uc_math_scores[is.na(bg_sub$act_math_score),], aes(x=uc_true, y=from_sat)) +
  geom_point(na.rm=T)
p2 = ggplot(uc_math_scores[is.na(bg_sub$sat_math_score),], aes(x=uc_true, y=from_act)) +
  geom_point(na.rm=T)
p3 = ggplot(uc_math_scores, aes(x=uc_true, y=from_both)) +
  geom_point(na.rm=T)
grid.arrange(p1, p2, p3, ncol=3)
```




```
print('Break grid arrange')
```

```
## [1] "Break grid arrange"
```

Plot prediction errors by admission



- combination has some problems, that do not depend on admission year.

```
model = lm(uc_true ~ from_act + from_sat, uc_math_scores)
summary(model)
```

```
##
## Call:
## lm(formula = uc_true ~ from_act + from_sat, data = uc_math_scores)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -16.544  -1.099  -0.101   0.901  37.735
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.142748   0.167153   0.854    0.393
## from_act    0.252396   0.004392  57.473   <2e-16 ***
## from_sat    0.753046   0.004557 165.249   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.166 on 9273 degrees of freedom
##   (37132 Beobachtungen als fehlend gelöscht)
## Multiple R-squared:  0.9593, Adjusted R-squared:  0.9592
## F-statistic: 1.092e+05 on 2 and 9273 DF,  p-value: < 2.2e-16
```

Let's check predictions

```
table(uc_math_scores$uc - round(predict(model, uc_math_scores)))
```

```
##
##  -17  -10   -9   -8   -7   -6   -5   -4   -3   -2   -1    0    1    2    3    4
##    1    1    4    5   37   42   58  185  389  941 2211 2418 1520  706  248  159
##    5    6    7    8    9   10   11   12   13   14   16   17   38
##  172   53   43   46   15    6    6    2    4    1    1    1    1
```

Although majority is predicted very well, it looks like the formula does not quite get it right. Maybe let's see what information we have in the important three years.

```
table(!is.na(bg_sub$act_math_score[filter_pre_F12]))
```

```
##
## FALSE  TRUE
## 15878  2295
```

```
table(!is.na(bg_sub$sat_math_score[filter_pre_F12]))
```

```
##
## FALSE  TRUE
##  4220 13953
```

```
table(is.na(bg_sub$sat_math_score[filter_pre_F12]) & !is.na(bg_sub$act_math_score[filter_pre_F12]))
```

```
##
## FALSE  TRUE
## 16905  1268
```

We mostly have SAT scores and only 1268 cases where we exclusively know the ACT math score, which means that we might ignore the remaining residuals.

Next steps: - use the prediction for math scores - investigate on meaning of UC score: https://blog.prepscholar.com/historical-act-percentiles-2015-2014-2013-2012-2011 - check predictiveness of UC scores for dropout