Report 2024/12/08

Dingyi Nie

1 Categorical variables' case proportions

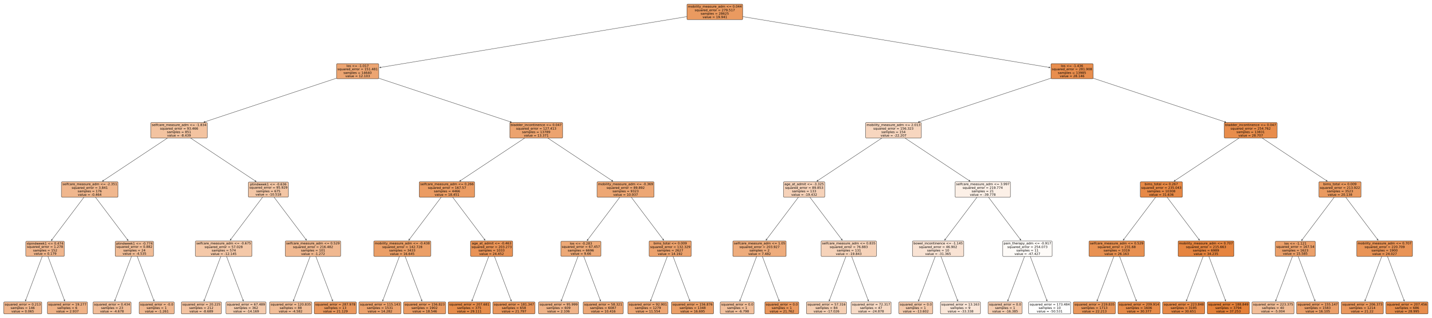
For all 9 categorical variables remaining after manual selection, here are the proportions of cases:

```
Feature: falls_hx
  No: 26714
  Yes: 13847
  Unknown: 2377
Feature: marital_status
  Unmarried: 20703
  Married: 19766
  Unknown: 2469
Feature: depressed_adm
  0: 31768
  1: 6529
  Unknown: 4641
Feature: payor_primary
  Medicare_FFS: 16063
  Not_listed: 14576
  Medicare_Advantage: 12299
Feature: hispanic
  No: 39745
  Yes: 3193
Feature: sex
  male: 22713
  female: 20225
Feature: stroke_type
  Ischemic: 35805
  Hemorrhagic: 6486
  Other: 647
Feature: low_interest_adm
  0: 33620
  1: 4741
  Unknown: 4577
Feature: race
  White: 29020
  Black_or_African_American: 9838
  Unknown: 2413
  Asian: 1321
  American_Indian_or_Alaska_Native: 176
  Multiracial: 89
  Native_Hawaiian_or_Other_Pacific_Islander: 81
```

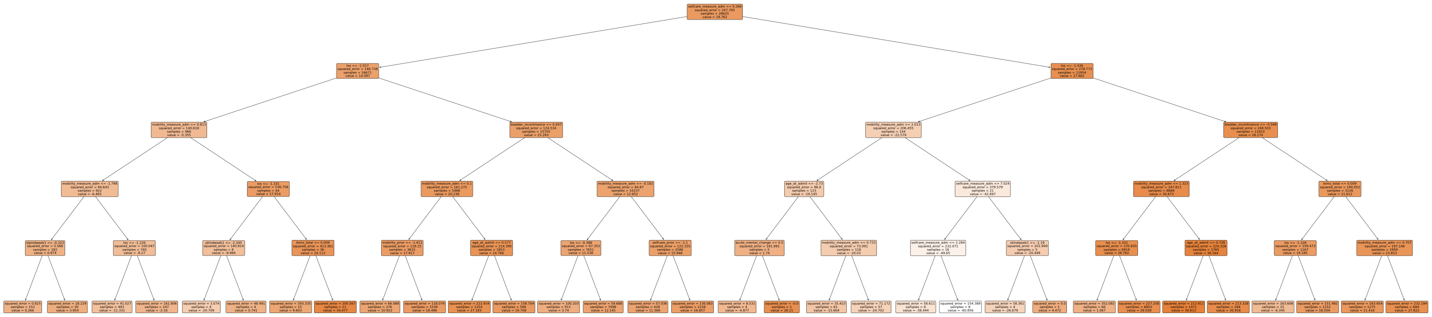
We do not see any drastically low proportion case, except for `race`, where we follow NINDS's grouping. It seems we can not reduce these categories any further, especially considering `Multiracial` is one of the smaller-proportion cases, and grouping it with the other few cases does not make sense logically.

2 Updated RF global summary trees

Selfcare:


mobility_measure_adm <= 0.044 squared_error = 279.517 samples = 28625 value = 19.941
los <= -1.017 squared_error = 151.481 samples = 14640 value = 12.103
los <= -1.436 squared_error = 281.908 samples = 13985 value = 28.146
selfcare_measure_adm <= -1.834 squared_error = 93.466 samples = 851 value = -8.439
bladder_incontinence <= 0.047 squared_error = 127.413 samples = 13789 value = 13.371
mobility_measure_adm <= 2.013 squared_error = 156.323 samples = 154 value = -22.207
bladder_incontinence <= 0.047 squared_error = 254.762 samples = 13831 value = 28.707
selfcare_measure_adm <= -2.351 squared_error = 3.841 samples = 176 value = -0.464
ptindweek1 <= -0.636 squared_error = 95.929 samples = 675 value = -10.518
selfcare_measure_adm <= 0.266 squared_error = 167.57 samples = 4466 value = 18.451
mobility_measure_adm <= -0.369 squared_error = 89.892 samples = 9323 value = 10.937
age_at_admit <= -3.325 squared_error = 89.853 samples = 133 value = -19.432
selfcare_measure_adm <= 3.997 squared_error = 219.774 samples = 21 value = -39.778
bims_total <= 0.267 squared_error = 235.043 samples = 10308 value = 31.636
bims_total <= 0.009 squared_error = 213.922 samples = 3523 value = 20.138
slpindweek1 <= 0.474 squared_error = 1.278 samples = 152 value = 0.179
ptindweek1 <= -0.774 squared_error = 0.882 samples = 24 value = -4.535
selfcare_measure_adm <= -0.675 squared_error = 57.028 samples = 574 value = -12.145
selfcare_measure_adm <= 0.529 squared_error = 216.482 samples = 101 value = -1.272
mobility_measure_adm <= -0.438 squared_error = 142.728 samples = 3433 value = 16.645
age_at_admit <= -0.463 squared_error = 203.273 samples = 1033 value = 24.452
los <= -0.283 squared_error = 67.457 samples = 6696 value = 9.66
bims_total <= 0.009 squared_error = 132.329 samples = 2627 value = 14.192
selfcare_measure_adm <= 1.05 squared_error = 203.927 samples = 2 value = 7.482
selfcare_measure_adm <= 0.835 squared_error = 76.883 samples = 131 value = -19.843
bowel_incontinence <= -1.145 squared_error = 46.902 samples = 10 value = -31.365
pain_therapy_adm <= -0.917 squared_error = 254.073 samples = 11 value = -47.427
selfcare_measure_adm <= 0.529 squared_error = 231.68 samples = 3319 value = 26.163
mobility_measure_adm <= 0.707 squared_error = 215.663 samples = 6989 value = 34.235
los <= -1.121 squared_error = 167.54 samples = 1623 value = 15.585
mobility_measure_adm <= 0.707 squared_error = 220.709 samples = 1900 value = 24.027
squared_error = 0.213 samples = 146 value = 0.065
squared_error = 19.277 samples = 6 value = 2.937
squared_error = 0.434 samples = 23 value = -4.678
squared_error = -0.0 samples = 1 value = -1.261
squared_error = 20.225 samples = 212 value = -8.689
squared_error = 67.489 samples = 362 value = -14.169
squared_error = 120.835 samples = 88 value = -4.582
squared_error = 287.978 samples = 13 value = 21.129
squared_error = 115.143 samples = 1531 value = 14.282
squared_error = 156.823 samples = 1902 value = 18.546
squared_error = 207.681 samples = 375 value = 29.111
squared_error = 181.347 samples = 658 value = 21.797
squared_error = 95.999 samples = 609 value = 2.106
squared_error = 58.321 samples = 6087 value = 10.416
squared_error = 92.901 samples = 1279 value = 11.554
squared_error = 156.876 samples = 1348 value = 16.695
squared_error = 0.0 samples = 1 value = -6.798
squared_error = 0.0 samples = 1 value = 21.762
squared_error = 57.316 samples = 84 value = -17.026
squared_error = 72.317 samples = 47 value = -24.878
squared_error = 0.0 samples = 1 value = -13.602
squared_error = 13.163 samples = 9 value = -33.338
squared_error = 0.0 samples = 1 value = -16.385
squared_error = 173.484 samples = 10 value = -50.531
squared_error = 219.835 samples = 1713 value = 22.213
squared_error = 209.914 samples = 1606 value = 30.377
squared_error = 223.848 samples = 3195 value = 30.651
squared_error = 188.849 samples = 3794 value = 37.253
squared_error = 223.375 samples = 40 value = -5.004
squared_error = 155.147 samples = 1583 value = 16.105
squared_error = 206.373 samples = 1214 value = 21.22
squared_error = 207.456 samples = 686 value = 28.995

Mobility:


selfcare_measure_adm <= 0.266 squared_error = 247.795 samples = 28625 value = 19.762
los <= -1.017 squared_error = 148.738 samples = 16671 value = 14.097
los <= -1.436 squared_error = 278.773 samples = 11954 value = 27.662
mobility_measure_adm <= 0.813 squared_error = 140.618 samples = 966 value = -5.355
bladder_incontinence <= 0.047 squared_error = 124.534 samples = 15705 value = 15.293
mobility_measure_adm <= 2.013 squared_error = 206.455 samples = 144 value = -22.579
bladder_incontinence <= -0.599 squared_error = 248.503 samples = 11810 value = 28.274
mobility_measure_adm <= -1.798 squared_error = 94.643 samples = 922 value = -6.465
los <= -1.331 squared_error = 536.756 samples = 44 value = 17.914
mobility_measure_adm <= 0.1 squared_error = 161.275 samples = 5468 value = 20.238
mobility_measure_adm <= -0.182 squared_error = 84.87 samples = 10237 value = 12.652
age_at_admit <= -2.73 squared_error = 96.0 samples = 123 value = -19.145
selfcare_measure_adm <= 7.024 squared_error = 379.579 samples = 21 value = -42.697
mobility_measure_adm <= 1.323 squared_error = 247.811 samples = 8684 value = 30.673
bims_total <= 0.009 squared_error = 190.052 samples = 3126 value = 21.612
slpindweek1 <= -0.323 squared_error = 5.566 samples = 182 value = 0.874
los <= -1.226 squared_error = 100.047 samples = 740 value = -8.27
ptindweek1 <= -2.345 squared_error = 140.814 samples = 8 value = -9.984
bims_total <= 0.009 squared_error = 413.361 samples = 36 value = 24.113
mobility_prior <= -1.411 squared_error = 118.15 samples = 3615 value = 17.917
age_at_admit <= 0.577 squared_error = 214.395 samples = 1853 value = 24.766
los <= -0.388 squared_error = 67.353 samples = 7651 value = 11.538
selfcare_prior <= -1.1 squared_error = 122.155 samples = 2586 value = 15.948
acute_mental_change <= 0.5 squared_error = 181.991 samples = 5 value = 1.74
mobility_measure_adm <= 0.733 squared_error = 73.091 samples = 118 value = -20.03
selfcare_measure_adm <= 2.284 squared_error = 232.071 samples = 16 value = -49.65
otindweek1 <= -1.19 squared_error = 201.949 samples = 5 value = -20.449
los <= -1.331 squared_error = 235.835 samples = 6919 value = 28.762
age_at_admit <= 0.726 squared_error = 224.316 samples = 1765 value = 38.164
los <= -1.226 squared_error = 159.473 samples = 1167 value = 18.185
mobility_measure_adm <= 0.707 squared_error = 197.106 samples = 1959 value = 23.653
squared_error = 0.825 samples = 152 value = 0.266
squared_error = 18.229 samples = 30 value = 3.954
squared_error = 41.027 samples = 493 value = -11.331
squared_error = 161.806 samples = 247 value = -2.16
squared_error = 3.074 samples = 4 value = -20.709
squared_error = 48.491 samples = 4 value = 0.741
squared_error = 350.335 samples = 15 value = 9.603
squared_error = 200.567 samples = 21 value = 34.477
squared_error = 66.088 samples = 276 value = 10.922
squared_error = 118.074 samples = 3339 value = 18.496
squared_error = 222.934 samples = 1254 value = 27.183
squared_error = 158.704 samples = 599 value = 19.708
squared_error = 100.203 samples = 553 value = 3.74
squared_error = 59.688 samples = 7098 value = 12.145
squared_error = 57.036 samples = 428 value = 11.368
squared_error = 130.083 samples = 2158 value = 16.857
squared_error = 8.531 samples = 4 value = -4.877
squared_error = -0.0 samples = 1 value = 28.21
squared_error = 35.423 samples = 61 value = -15.664
squared_error = 71.172 samples = 57 value = -24.702
squared_error = 58.611 samples = 8 value = -38.444
squared_error = 154.389 samples = 8 value = -60.856
squared_error = 58.361 samples = 4 value = -26.679
squared_error = 0.0 samples = 1 value = 4.472
squared_error = 352.092 samples = 66 value = 1.067
squared_error = 227.258 samples = 6853 value = 29.028
squared_error = 213.911 samples = 1471 value = 39.613
squared_error = 213.328 samples = 294 value = 30.916
squared_error = 163.606 samples = 15 value = -6.345
squared_error = 151.482 samples = 1152 value = 18.504
squared_error = 163.954 samples = 1275 value = 21.416
squared_error = 232.194 samples = 684 value = 27.822

3 Updated Lasso regression

Increased Lasso CV search of Alpha's grainularity. Now alpha is searched in

```
alpha_range = [0.001, 0.05, 0.03, 0.01, 0.5, 0.1, 0.5, 1, 5, 10, 50, 100]
```

For selfcare task, the best alpha found by CV is now 0.03.

Following 1-SE rule, for all alpha values that satisfy `mean_error <= min_error + min_error_se`, the largest is 0.1. Fit a new Lasso model on selfcare task with alpha = 0.1 yields the following results:

```
Best alpha (min error): 0.03
Alpha selected by 1SE rule: 0.1

---------- Train Set Results ----------
MSE: 346.2007
R²: 0.2403
```

```
---------- Eval Set Results ----------
MSE: 338.3352
R²: 0.2534

---------- Test Set Results ----------
MSE: 340.5518
R²: 0.2417

---------- Model Summary: Non-zero Coefficients ----------
                           Feature  Coefficient
0             mobility_measure_adm     6.830357
1                       ptindweek1     2.158104
2                       otindweek1     1.740079
3                       bims_total     1.499827
4                   selfcare_prior     1.122635
5            understand_verbal_adm     1.058634
6         marital_status_Unmarried     0.807883
7                       race_White     0.654775
8                      slpindweek1     0.519224
9                   mobility_prior     0.362948
10             pain_activities_adm     0.337819
11      payor_primary_Medicare_FFS     0.315076
12                             los     0.203444
13        payor_primary_Not_listed     0.183014
14                      vision_adm    -0.272099
15              bowel_incontinence    -0.283727
16             acute_mental_change    -0.392702
17  race_Black_or_African_American    -0.421941
18                      ptindweek2    -0.486597
19              pressure_ulcer_adm    -0.666926
20                  health_lit_adm    -0.765210
21                    falls_hx_Yes    -0.806571
22                        sex_male    -0.837267
23                    age_at_admit    -1.625570
24            selfcare_measure_adm    -2.542324
25            bladder_incontinence    -3.203255
```

Now number of variables reduce from 49 to 26.

For the mobility task, 1-SE rule also yields alpha = 0.1, and the results are as follows:

```
Best alpha (min error): 0.01
Alpha selected by 1SE rule: 0.1

---------- Train Set Results ----------
MSE: 344.2645
R²: 0.1851
```

```
---------- Eval Set Results ----------
MSE: 336.2865
R²: 0.2023

---------- Test Set Results ----------
MSE: 334.4508
R²: 0.1999

---------- Model Summary: Non-zero Coefficients ----------
                           Feature  Coefficient
0                       ptindweek1     2.238204
1             selfcare_measure_adm     2.041735
2             mobility_measure_adm     1.960027
3                       otindweek1     1.356835
4                      slpindweek1     1.273856
5                   selfcare_prior     0.906201
6         payor_primary_Not_listed     0.850155
7                       bims_total     0.767928
8                   mobility_prior     0.682834
9            understand_verbal_adm     0.626199
10                      race_White     0.605145
11        marital_status_Unmarried     0.351502
12                             los     0.221773
13             pain_activities_adm     0.184571
14                     slpindweek2     0.000979
15                    pvd_comorbid    -0.121013
16  race_Black_or_African_American    -0.304452
17                      otindweek2    -0.304661
18              bowel_incontinence    -0.423265
19                      vision_adm    -0.448847
20                      ptindweek2    -0.530227
21                  health_lit_adm    -0.667423
22              pressure_ulcer_adm    -0.821829
23                    age_at_admit    -1.428901
24                    falls_hx_Yes    -1.470242
25            bladder_incontinence    -2.615774
```

4 Updated modeling with interaction terms

Now, before creating interaction terms, OT/PT/SLP are combined for week1 and week2 respectively, hence only two combined variables (`sumindweek1` and `sumindweek2`) will be interacting with the rest of the variables.

After adding interaction terms to the data, rerun the forementioned Lasso experiments with 1-SE rule, the results are as follows:

Selfcare:

```
Best alpha (min error): 0.03
Alpha selected by 1SE rule: 0.1

---------- Train Set Results ----------
MSE: 327.2569
R²: 0.2818

---------- Eval Set Results ----------
MSE: 322.7168
R²: 0.2879

---------- Test Set Results ----------
MSE: 321.1996
R²: 0.2848

---------- Model Summary: Non-zero Coefficients ----------
                                 Feature  Coefficient
0                   mobility_measure_adm     6.303209
1                             bims_total     1.681771
2                  understand_verbal_adm     1.157852
3                         selfcare_prior     1.115792
4               marital_status_Unmarried     0.726547
5                            sumindweek1     0.578239
6     sumindweek1_x_mobility_measure_adm     0.525772
7     sumindweek2_x_bladder_incontinence     0.462465
8                                    los     0.430909
9                         mobility_prior     0.414889
10                            race_White     0.394299
11     sumindweek2_x_acute_mental_change     0.321403
12    sumindweek1_x_selfcare_measure_adm     0.287278
13              sumindweek1_x_bims_total     0.226504
14                   pain_activities_adm     0.213853
15   sumindweek1_x_understand_verbal_adm     0.203225
16      sumindweek2_x_pressure_ulcer_adm     0.162224
17                sumindweek1_x_sex_male     0.150836
18            sumindweek2_x_falls_hx_Yes     0.133327
19          sumindweek2_x_health_lit_adm     0.126407
20  sumindweek1_x_marital_status_Unknown     0.116839
21                        expression_adm     0.058529
22          sumindweek1_x_expression_adm     0.052702
23            payor_primary_Medicare_FFS     0.052049
24          sumindweek1_x_mobility_prior     0.045845
25     sumindweek2_x_pain_activities_adm     0.036738
26    sumindweek2_x_selfcare_measure_adm     0.032978
27                sumindweek2_x_sex_male     0.022398
28          sumindweek1_x_selfcare_prior     0.017972
29          sumindweek2_x_selfcare_prior     0.010290
30             sumindweek2_x_sumindweek1    -0.002467
31              sumindweek1_x_vision_adm    -0.004501
```

```
32        sumindweek2_x_socisolation_adm    -0.015323
33        sumindweek1_x_socisolation_adm    -0.041551
34          sumindweek1_x_health_lit_adm    -0.053187
35            sumindweek2_x_age_at_admit    -0.118749
36            sumindweek1_x_age_at_admit    -0.125633
37              sumindweek2_x_race_White    -0.168922
38                           sumindweek2    -0.187202
39    sumindweek1_x_bladder_incontinence    -0.187727
40                            vision_adm    -0.301072
41    sumindweek2_x_stroke_type_Ischemic    -0.304803
42              sumindweek2_x_bims_total    -0.346274
43             sumindweek1_x_sumindweek2    -0.353788
44   sumindweek2_x_understand_verbal_adm    -0.382435
45                    bowel_incontinence    -0.558118
46    sumindweek2_x_mobility_measure_adm    -0.603572
47                          falls_hx_Yes    -0.650124
48        race_Black_or_African_American    -0.719531
49                              sex_male    -0.736162
50                   acute_mental_change    -0.753878
51                        health_lit_adm    -0.808181
52                    pressure_ulcer_adm    -0.843482
53                     sumindweek2_x_los    -0.996046
54                          age_at_admit    -1.687008
55                  selfcare_measure_adm    -2.586528
56                  bladder_incontinence    -3.211919
```

Mobility:

```
Best alpha (min error): 0.03
Alpha selected by 1SE rule: 0.1

---------- Train Set Results ----------
MSE: 325.2945
R²: 0.2300

---------- Eval Set Results ----------
MSE: 321.9947
R²: 0.2362

---------- Test Set Results ----------
MSE: 316.1936
R²: 0.2436

---------- Model Summary: Non-zero Coefficients ----------
                                     Feature  Coefficient
0                       selfcare_measure_adm     1.882533
1                       mobility_measure_adm     1.461744
```

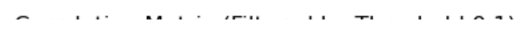
```
2                             selfcare_prior     0.933889
3                                 bims_total     0.868202
4         sumindweek1_x_mobility_measure_adm     0.770175
5                             mobility_prior     0.768207
6                   payor_primary_Not_listed     0.676640
7                      understand_verbal_adm     0.664408
8                                sumindweek1     0.632398
9         sumindweek2_x_bladder_incontinence     0.393663
10         sumindweek2_x_acute_mental_change     0.358847
11                  marital_status_Unmarried     0.299967
12                sumindweek2_x_falls_hx_Yes     0.248308
13                                race_White     0.217153
14                                       los     0.175723
15              sumindweek2_x_health_lit_adm     0.170146
16        sumindweek1_x_selfcare_measure_adm     0.159483
17      sumindweek2_x_marital_status_Unknown     0.145298
18       sumindweek1_x_understand_verbal_adm     0.136549
19          sumindweek2_x_pressure_ulcer_adm     0.129064
20                    sumindweek1_x_sex_male     0.110202
21                  sumindweek1_x_bims_total     0.108379
22              sumindweek1_x_mobility_prior     0.094699
23                sumindweek2_x_age_at_admit     0.068384
24                       pain_activities_adm     0.053893
25                sumindweek1_x_hispanic_Yes     0.044573
26        sumindweek2_x_selfcare_measure_adm     0.034050
27              sumindweek1_x_expression_adm     0.031052
28                  sumindweek2_x_vision_adm     0.016671
29              sumindweek2_x_pain_sleep_adm     0.013657
30              sumindweek2_x_expression_adm     0.004204
31          sumindweek1_x_bowel_incontinence    -0.000132
32                 sumindweek2_x_sumindweek1    -0.003778
33                 sumindweek1_x_hearing_adm    -0.024736
34              sumindweek2_x_mobility_prior    -0.033383
35                  sumindweek1_x_vision_adm    -0.045027
36                         sumindweek1_x_los    -0.049604
37                  sumindweek2_x_race_White    -0.052906
38            sumindweek2_x_socisolation_adm    -0.059858
39              sumindweek1_x_health_lit_adm    -0.071914
40                       acute_mental_change    -0.078821
41  sumindweek2_x_payor_primary_Medicare_FFS    -0.102238
42        sumindweek1_x_bladder_incontinence    -0.135312
43                sumindweek1_x_falls_hx_Yes    -0.160446
44                sumindweek1_x_age_at_admit    -0.220517
45                  sumindweek2_x_bims_total    -0.276977
46                 sumindweek1_x_sumindweek2    -0.321876
47       sumindweek2_x_understand_verbal_adm    -0.380237
48        sumindweek2_x_stroke_type_Ischemic    -0.420057
49                                vision_adm    -0.487702
50            race_Black_or_African_American    -0.625619
51                            health_lit_adm    -0.701054
```

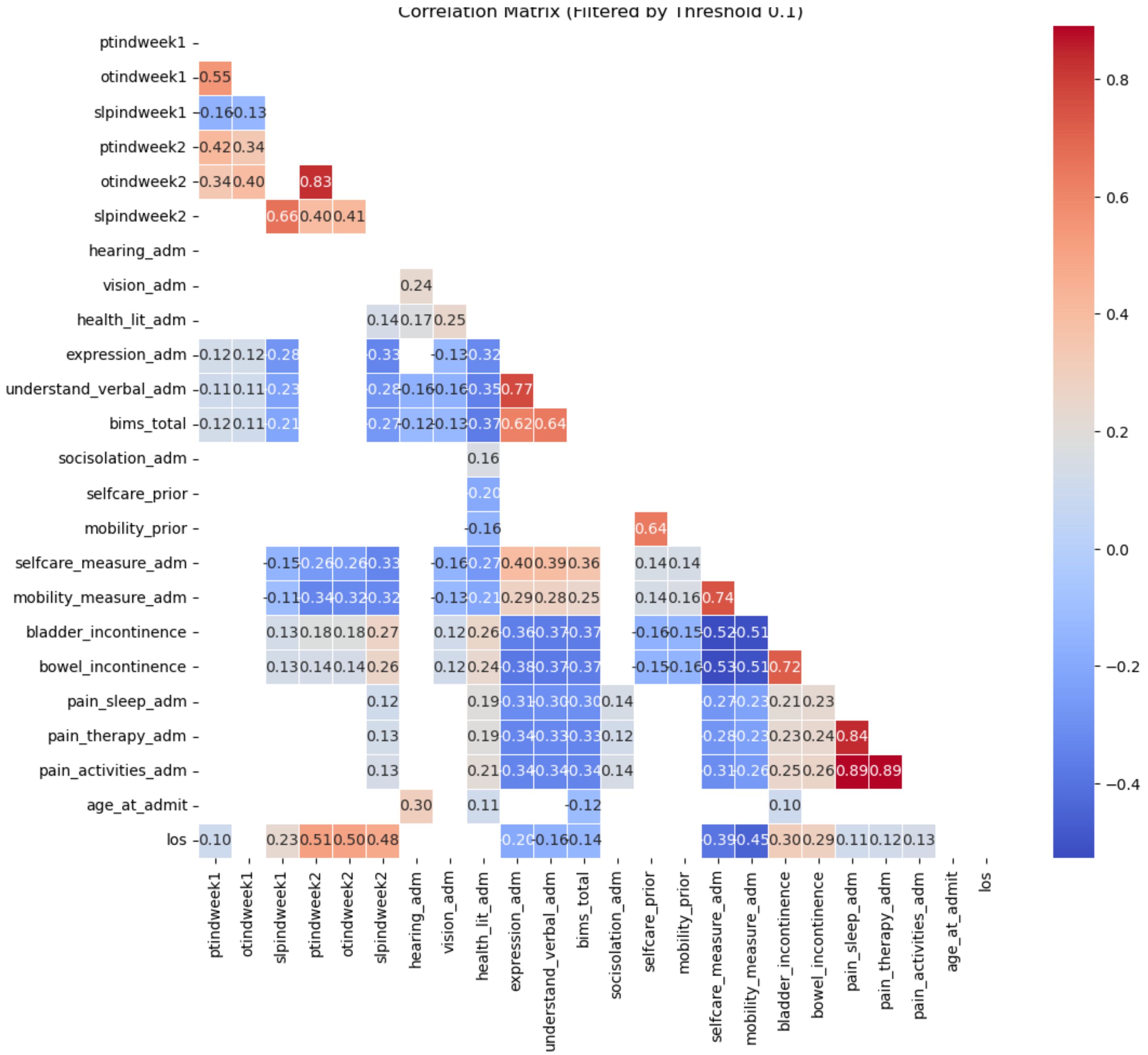
```
51                           health_lit_adm    -0.701054
52                        bowel_incontinence    -0.707825
53                         sumindweek2_x_los    -0.761137
54         sumindweek2_x_mobility_measure_adm    -0.820422
55                        pressure_ulcer_adm    -0.915766
56                              falls_hx_Yes    -1.365534
57                              age_at_admit    -1.512892
58                      bladder_incontinence    -2.650849
```

We've managed to reduce the number of variables (including interaction terms) down to under 60, while keeping similar predictive power to our previous models.

5 Correlation matrix

Binary variables are removed, and only correlation absolute value >= 0.1 is highlighted:




Correlation Matrix (Filtered by Threshold 0.1)
ptindweek1
otindweek1 0.55
slpindweek1 -0.16 0.13
ptindweek2 0.42 0.34
otindweek2 0.34 0.40 0.83
slpindweek2 0.66 0.40 0.41
hearing_adm
vision_adm 0.24
health_lit_adm 0.14 0.17 0.25
expression_adm 0.12 0.12 0.28 0.33 -0.13 0.32
understand_verbal_adm 0.11 0.11 0.23 0.28 -0.16 -0.16 0.35 0.77
bims_total 0.12 0.11 0.21 0.27 -0.12 -0.13 0.37 0.62 0.64
socisolation_adm 0.16
selfcare_prior 0.20
mobility_prior -0.16 0.64
selfcare_measure_adm -0.15 0.26 0.26 0.33 -0.16 0.27 0.40 0.39 0.36 0.14 0.14
mobility_measure_adm -0.11 0.34 0.32 0.32 -0.13 0.21 0.29 0.28 0.25 0.14 0.16 0.74
bladder_incontinence 0.13 0.18 0.18 0.27 0.12 0.26 0.36 0.37 0.37 -0.16 0.15 0.52 0.51
bowel_incontinence 0.13 0.14 0.14 0.26 0.12 0.24 0.38 0.37 0.37 -0.15 0.16 0.53 0.51 0.72
pain_sleep_adm 0.12 0.19 0.31 0.30 0.30 0.14 0.27 0.23 0.21 0.23
pain_therapy_adm 0.13 0.19 0.34 0.33 0.33 0.12 0.28 0.23 0.23 0.24 0.84
pain_activities_adm 0.13 0.21 0.34 0.34 0.34 0.14 0.31 0.26 0.25 0.26 0.89 0.89
age_at_admit 0.30 0.11 -0.12 0.10
los 0.10 0.23 0.51 0.50 0.48 0.20 -0.16 0.14 0.39 0.45 0.30 0.29 0.11 0.12 0.13
ptindweek1
otindweek1
slpindweek1
ptindweek2
otindweek2
slpindweek2
hearing_adm
vision_adm
health_lit_adm
expression_adm
understand_verbal_adm
bims_total
socisolation_adm
selfcare_prior
mobility_prior
selfcare_measure_adm
mobility_measure_adm
bladder_incontinence
bowel_incontinence
pain_sleep_adm
pain_therapy_adm
pain_activities_adm
age_at_admit
los
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4