3_chi_sq

Maiko Hata

chi 1: race_us

— to make the plot using corrplot of national race/exit patterns —

```
# A tibble: 7 x 9
  race              area  exit_total withdrawal_by_parent attempts_to_contact_~1
  <chr>             <chr>      <dbl>                <dbl>                   <dbl>
1 Alaska Native or~ US a~      23783                 3155                    3262
2 Asian             US a~     146252                21737                    6559
3 Black or African~ US a~     415873                48699                   56155
4 Hispanic or Lati~ US a~     905971               109469                   79951
5 Two or More Races US a~     135841                19195                   11676
6 Pacific Islander  US a~      10654                 1467                     977
7 White             US a~    1672185               242489                   99922
# i abbreviated name: 1: attempts_to_contact_unsuccessful
# i 4 more variables: moved_out_of_state <dbl>,
#   part_b_eligible_exiting_part_c <dbl>, complete_or_not_eligible <dbl>,
#   part_b_eligibility_not_determined <dbl>
```

```
    Pearson's Chi-squared test

data:  race_us[, 3:8]
X-squared = 52218, df = 30, p-value < 2.2e-16
```

```
    Pearson's Chi-squared test

data:  race_oregon[, 3:8]
X-squared = 365.56, df = 30, p-value < 2.2e-16
```

step 2: byrace

— US and Oregon entire table. Probably no need as I'll be running the chi-square with residuals anyway. —

```
    Chi-squared test for given probabilities

data:  data_oregon[, 2]
X-squared = 69801, df = 6, p-value < 2.2e-16
```

chi 3: agg_by_area

— withdrawn category —

```
    Pearson's Chi-squared test with Yates' continuity correction

data:  agg_by_area[, 2:3]
X-squared = 120.26, df = 1, p-value < 2.2e-16
```

chi 4: us_data_attempts_BLWH

— DQ for BLACK and WHITE for chi-square and odds ratio

```
    Pearson's Chi-squared test with Yates' continuity correction

data:  us_data_attempts_BLWH[, 2:3]
X-squared = 22556, df = 1, p-value < 2.2e-16
```

All above runs.

chi 5: race_matrix

— I shouldn't need to "call" race_matrix.csv because it's a df but that means I have to recreate it — I HAVE TO READ IT

```
     withdrawal_by_parent attempts_to_contact_unsuccessful moved_out_of_state
[1,]                 3155                             3262                1045
[2,]                21737                             6559                8031
[3,]                48699                            56155               14648
[4,]               109469                            79951               26699
[5,]                19195                            11676                6477
[6,]                 1467                              977                 559
[7,]               242489                            99922               65399
     part_b_eligible_exiting_part_c complete_or_not_eligible
[1,]                           9035                    4395
[2,]                          52840                   26712
[3,]                         139607                   65353
[4,]                         318999                  165654
[5,]                          50699                   28996
[6,]                           3587                    1911
[7,]                         623690                  402568
     part_b_eligibility_not_determined
[1,]                              2480
[2,]                             22848
[3,]                             69687
[4,]                            173416
[5,]                             15813
[6,]                              1532
[7,]                            178843

[1] FALSE

[1] FALSE

[1] FALSE


    Pearson's Chi-squared test

data:  race_matrix
X-squared = 88194, df = 30, p-value < 2.2e-16
```

LINA: Does this number from the next chunk look right? LINA: "It's not adjusted"

```
     withdrawal_by_parent attempts_to_contact_unsuccessful moved_out_of_state
```

```
[1,]           -2.234095                       32.841369          4.903254
[2,]           18.267561                      -47.216930         38.237805
[3,]          -31.865184                      150.672278         -4.860364
[4,]          -46.871821                       41.512674        -45.704335
[5,]            4.759969                        9.210151         19.711489
[6,]            1.784885                        5.970013          8.939672
[7,]           53.637330                     -126.955744         18.662621
     part_b_eligible_exiting_part_c complete_or_not_eligible
[1,]                       3.306337             -11.24173880
[2,]                       3.739375             -23.74837150
[3,]                     -30.371562             -85.23775119
[4,]                     -25.437796             -76.56320235
[5,]                       4.206750              -0.05338682
[6,]                      -3.855291              -6.76085084
[7,]                      39.362366             136.85244266
     part_b_eligibility_not_determined
[1,]                         -17.265174
[2,]                          20.370145
[3,]                          58.876881
[4,]                         163.439764
[5,]                         -28.271052
[6,]                           1.954419
[7,]                        -178.918581
```

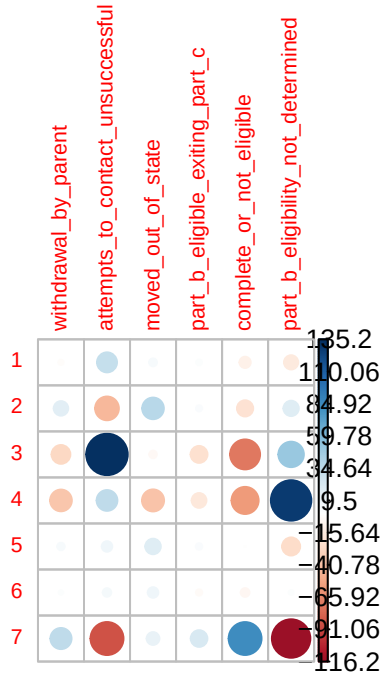
Open res_race_matrix as a df so I can show Lina the numbers.

Trying to get ADJUSTED standardized residuals

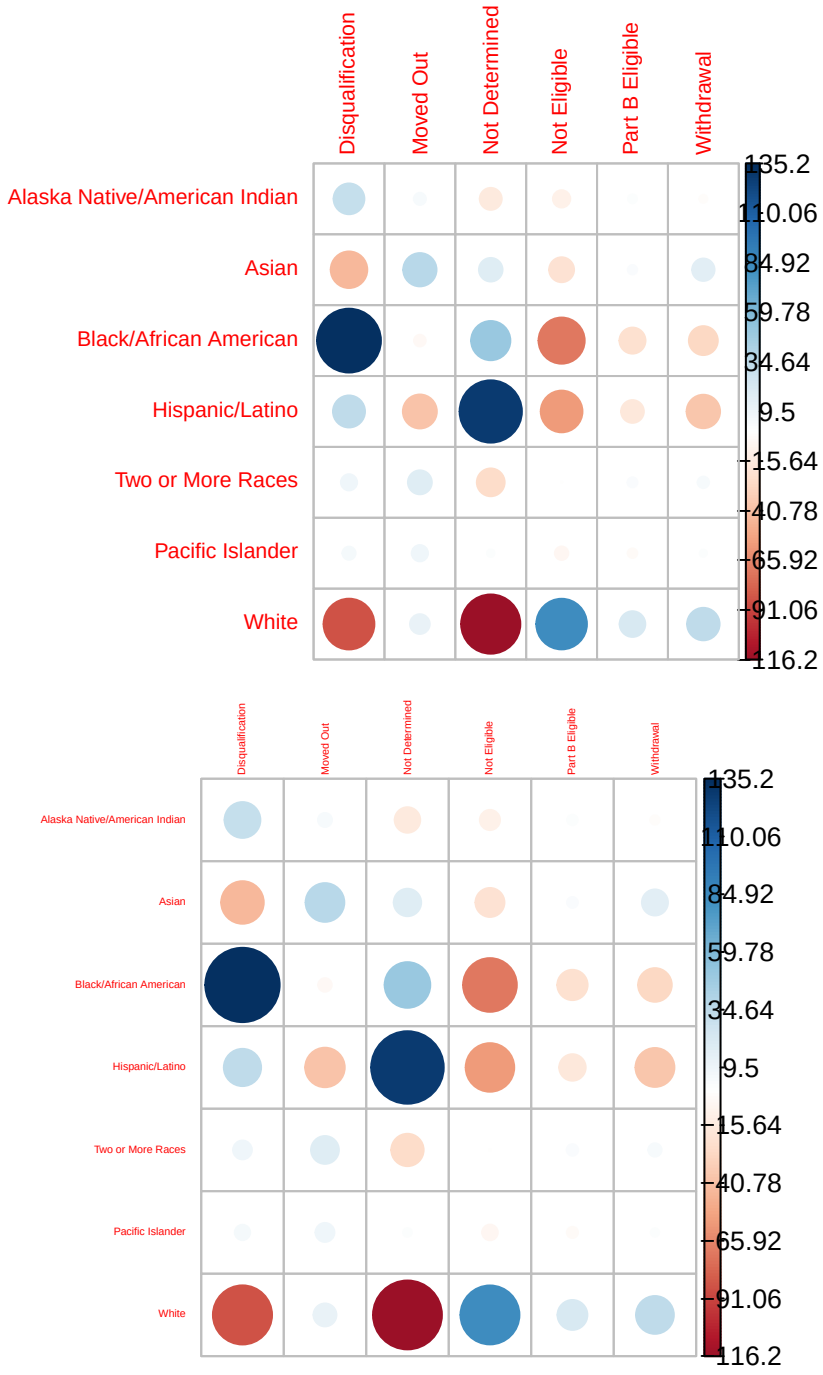
```
    Pearson's Chi-squared test

data:  race_matrix
X-squared = 88194, df = 30, p-value < 2.2e-16
```

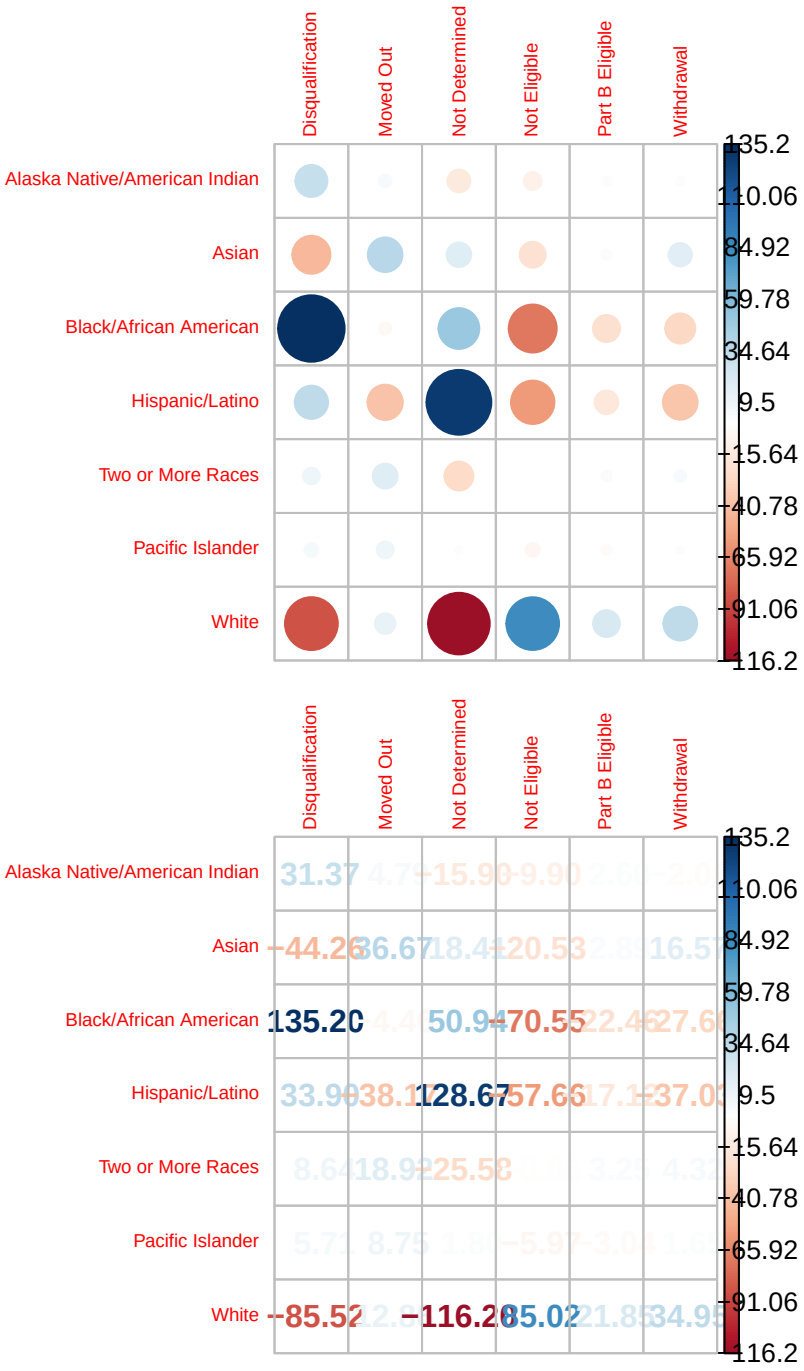
```
null device
          1
```

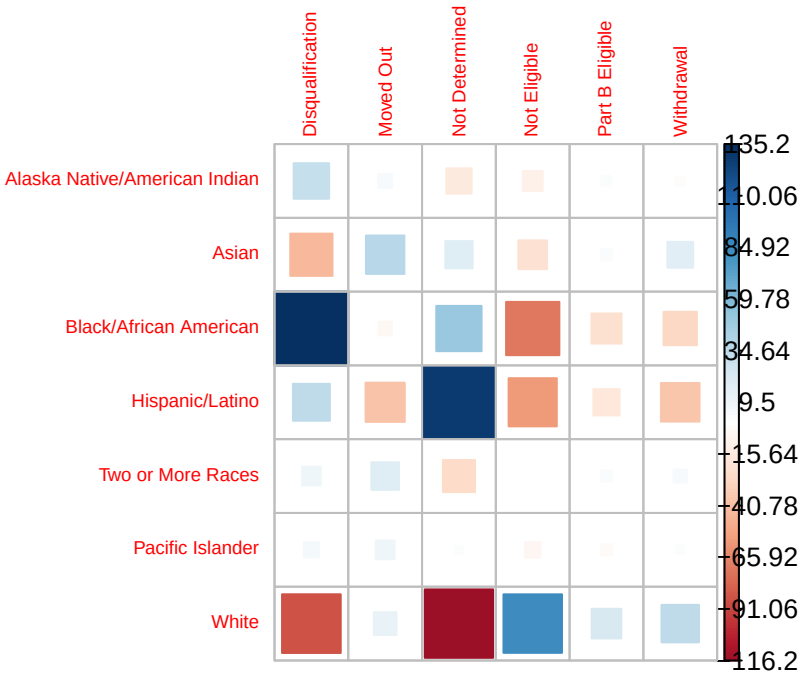
corrplot race/exit categories for race: renaming columns/rows for chi_results

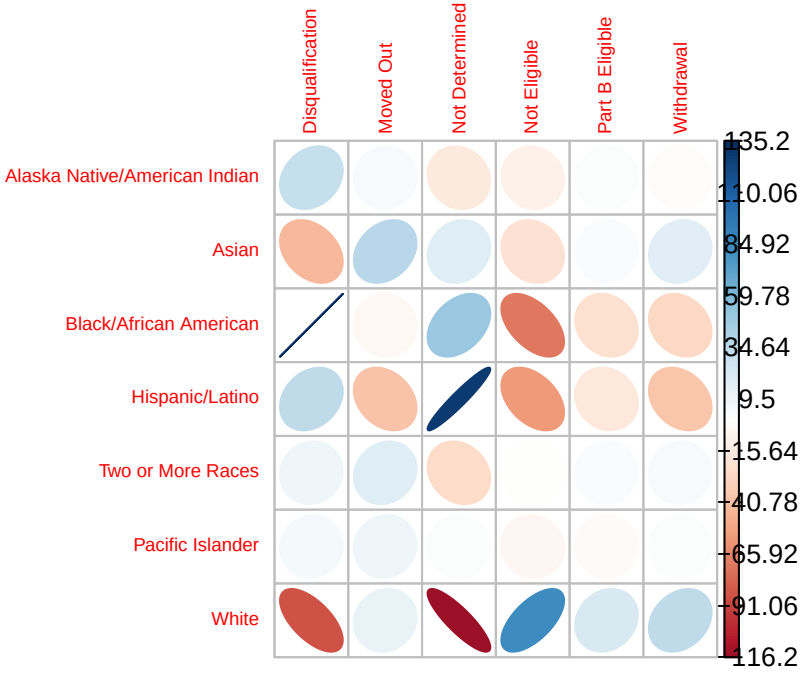


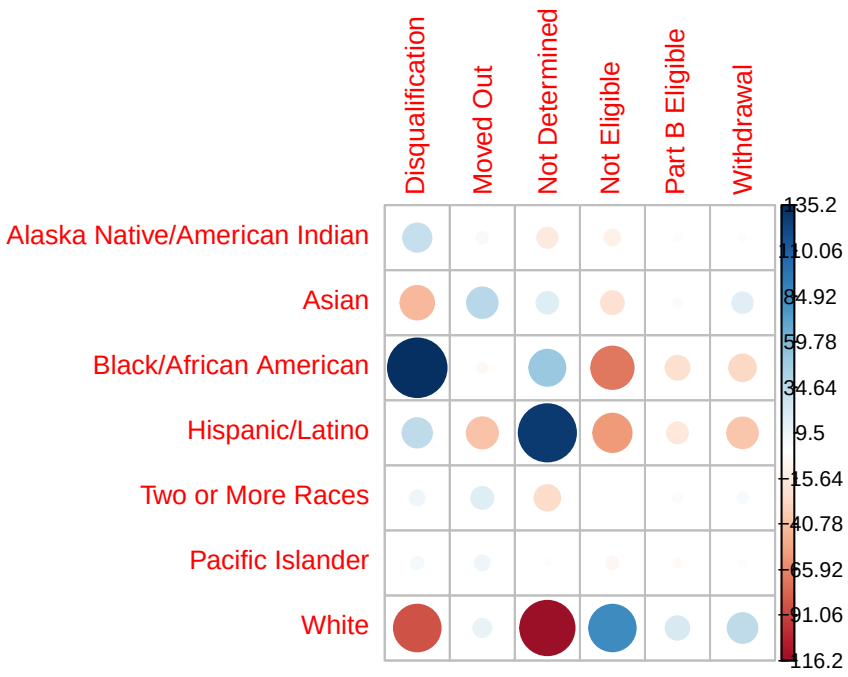

```
[1] "withdrawal_by_parent"                "attempts_to_contact_unsuccessful"
[3] "moved_out_of_state"                  "part_b_eligible_exiting_part_c"
[5] "complete_or_not_eligible"            "part_b_eligibility_not_determined"
```


Disqualification
Moved Out
Not Determined
Not Eligible
Part B Eligible
Withdrawal
Alaska Native/American Indian
Asian
Black/African American
Hispanic/Latino
Two or More Races
Pacific Islander
White
135.2
110.06
84.92
59.78
34.64
9.5
-15.64
-40.78
-65.92
-91.06
-116.2

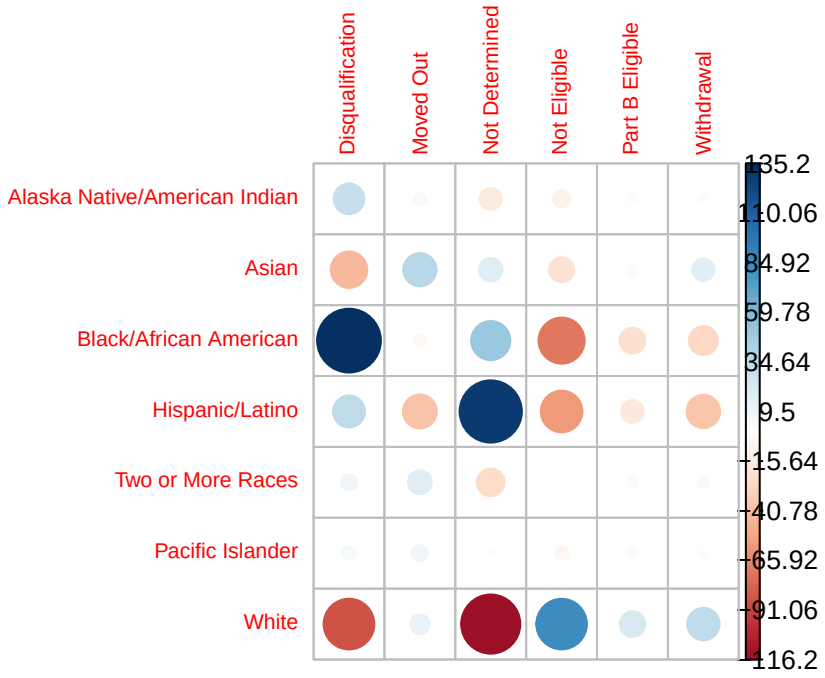

Disqualification
Moved Out
Not Determined
Not Eligible
Part B Eligible
Withdrawal
Alaska Native/American Indian
Asian
Black/African American
Hispanic/Latino
Two or More Races
Pacific Islander
White
135.2
110.06
84.92
59.78
34.64
9.5
−15.64
−40.78
−65.92
−91.06
−116.2
31.37
−15.90
−9.90
−44.26
−20.53
16.57
135.20
50.94
−70.55
33.90
38.11
128.67
57.66
−37.03
−85.52
−116.28
85.02
21.85
34.95


Disqualification
Moved Out
Not Determined
Not Eligible
Part B Eligible
Withdrawal
Alaska Native/American Indian
Asian
Black/African American
Hispanic/Latino
Two or More Races
Pacific Islander
White
135.2
110.06
84.92
59.78
34.64
9.5
-15.64
-40.78
-65.92
-91.06
-116.2


Disqualification
Moved Out
Not Determined
Not Eligible
Part B Eligible
Withdrawal
Alaska Native/American Indian
Asian
Black/African American
Hispanic/Latino
Two or More Races
Pacific Islander
White
135.2
110.06
84.92
59.78
34.64
9.5
-15.64
-40.78
-65.92
-91.06
-116.2




```
    Pearson's Chi-squared test

data:  race_matrix
X-squared = 88194, df = 30, p-value < 2.2e-16
```



IGNORE THESE CHUNKS — I DON'T THINK I NEED THIS There used to be a Var1 and Var2 now there's only Var1. WHERE DID I MAKE residuals_matrix

it's NULL so I need to assign names first it says

I'm HERE RIGHT NOW (2/19 1:17)

IGNORE: chi 5: residuals_df

Idoesn't render but it almost does - it's rendering pdf and stops. What's wrong with the following chunk?