1_clean_data

Maiko Hata

csv1: 10 exit categories

(for Kable table)

csv2: byrace

(imported OSEP excel, 2013-22, US/OR, deceased/continuing in Part C removed, mutated complete_or_not_eligible to combine 3 similar categories)

csv3: agg_by_race_and_state

based on "byrace", dropped the combined 3 categories, so just total + 6 exit categories

csv 4: race_oregon

based on agg_by_race_and_state but Oregon only

csv 5: race_us

csv 6: race_us_chart

US Exit reasons by race (deleted "area")

```
[1] "race"                              "withdrawal_by_parent"
[3] "attempts_to_contact_unsuccessful"  "moved_out_of_state"
[5] "part_b_eligible_exiting_part_c"    "complete_or_not_eligible"
[7] "part_b_eligibility_not_determined" "area"
[9] "exit_total"
```

csv 7: race_oregon_chart

Oregon Exit reasons by race chart

csv 8: area_by_race & other_by_race

— df w. % of each racial group represented in the EI exit total data for OREGON and NATIONAL

csv 9: other_by_race

— same as csv 8, but for NATIONAL data, filtered by US and Outlying Areas

csv 10: oregon_by_race

csv 11a & 11b: us_data_long & us_data_wide

wide

csv 12: agg_by_area

— TOTAL EXITS vs WITHDRAWAL in OREGON and US but NO RACE so not too meaningful, so I didn't save as CSV

csv 13: data_national (exit vs withdrawn)

— national data comparing TOTAL EXITS vs WITHDRAWAL by RACE

csv 14: us_data_attempts_BLWH

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

csv 15: oregon_data_long & oregon_data_long

wide

csv 16: oregon_data_wide_DQ

csv 17: us_data_DQ

— Very similar to us_data_attempts_BLWH which has more data, us_data_DQ only has BLACK/WHITE DQs, saved for chi-square —

csv 18: us_data_DQ_proportion

— df for making the plot 2 —-

CLEAND UP TO HERE

The chi-square indicated that there was a statistically significant association between children being Black/African American or White and them leaving EI due to being disqualified nationally. The chi-square test indicated, X-squared (222556.00, N = 2,088,058), p < 2.2e-16 or 0.0000000000000002 (p < .001).

Cohen's h was calculated to evaluate the effect size of the analysis**.** The result indicated a small to medium effect size, h = 0.259. However, even though effect size shows the magnitude of the difference, it is not necessarily considered to be a direct indication of the importance of the findings (Morgan et al., 2020).

We used R version 4.4.1 [@base] and the following R packages: corrplot v. 0.95 [@corrplot2024], distill v. 1.6 [@distill], DT v. 0.33 [@DT], epitools v. 0.5.10.1 [@epitools], gt v. 0.11.1 [@gt], gtsummary v. 2.0.3 [@gtsummary], here v. 1.0.1 [@here], janitor v. 2.2.0 [@janitor], kableExtra v. 1.4.0 [@kableExtra], knitr v. 1.48 [@knitr2014; @knitr2015; @knitr2024], lme4 v. 1.1.35.5 [@lme4], patchwork v. 1.3.0 [@patchwork], pwr v. 1.3.0 [@pwr], quarto v. 1.4.4 [@quarto], rcartocolor v. 2.1.1 [@rcartocolor], rcompanion v. 2.4.36 [@rcompanion], reactable v. 0.4.4 [@reactable], rio v. 1.2.3 [@rio], rmarkdown v. 2.28 [@rmarkdown2018; @rmarkdown2020; @rmarkdown2024], scales v. 1.3.0 [@scales], sjPlot v. 2.8.16 [@sjPlot], tidylog v. 1.1.0 [@tidylog], tidyverse v. 2.0.0 [@tidyverse], tinytex v. 0.53 [@tinytex2019; @tinytex2024].