

**WELCOME
TO THE 2026
NDACAN
SUMMER
TRAINING
SERIES!**

The session will begin at 12pm EST.

This session is being recorded.

Please submit questions to the Q&A box!

NDACAN SUMMER TRAINING SERIES

National Data Archive on Child Abuse and Neglect

Duke University, Cornell University, UC San Francisco, & Mathematica



Children's Bureau

An Office of the Administration for Children & Families



LAYING THE GROUNDWORK

*FOUNDATIONAL SKILLS FOR USING ADMINISTRATIVE
DATA IN CHILD WELFARE RESEARCH*

NDACAN SUMMER TRAINING SERIES SCHEDULE

- **July 1st, 2026**
 - Overview of NDACAN administrative datasets
- **July 8th, 2026**
 - Data cleaning and management
- **July 15th, 2026**
 - Linking NCANDS and AFCARS
- **July 22rd, 2026**
 - Handling missing data
- **July 29th, 2026**
 - Data presentation and visualization

SESSION AGENDA

- Linking NCANDS and AFCARS
- Demonstration in R
- Q & A

LINKING NCANDS AND AFCARS

WHAT IS RECORD LINKAGE?

- Linkage combines multiple data sources based on one or more shared variables
- Internal record linkage
 - NDACAN administrative data files (NCANDS, AFCARS, NYTD) can be linked to each other at the child level using unique (encrypted) child IDs
- External record linkage
 - Aggregated NDACAN data can be linked at the aggregate level to external sources using common variables:
 - Time: year, month, half-month
 - Place: state, county
 - Demographic groups: sex, race/ethnicity, age

BENEFITS TO LINKING

- Combining sources expands range of measures
 - Individual-level record linkage: within NDACAN administrative data
 - Aggregate-level record linkage: external data sources
- Repeated observations
 - Enhance missing data solutions
 - Help identify and address measurement error
 - Enable longitudinal research designs

PITFALLS OF LINKING

- Myriad errors arise from linking less-than-clean data
- Non-missing values of shared variables may not agree
- Linking may result in (systematic) measurement error
 - NDACAN child IDs are state-specific; interstate moves lead to false negative links
 - NDACAN data reflect variation and changes in record-keeping
- Data linkage can create/amplify missing data problems

DATA LINKAGE AND CHILD WELFARE SYSTEM CONTACT

NCANDS

Child protective history

AFCARS

Foster care experience

NYTD

Transition out of care

MERGING BY ROW: STACKING DATA

- Multiple years of NCANDS or AFCARS data can be combined simply by “stacking” them, or in R, by row-binding year-specific data frames
 - Stacking is usually smooth because NDACAN administrative data have highly consistent structure over time
- Take care to distinguish between the occurrence, recording, and submission of information about events

STACKING AS LINKING

- By stacking multiple years of data within a given source (NCANDS, AFCARS), users can create longitudinal data, i.e. repeated observations
- Record-level linking
 - Users can link children by identifying multiple observations with shared unique child identifiers (ideally StFCID, a combination of child identifier and state)
- Aggregate linking
 - By summarizing data at the state or county level, users can link geographies by identifying multiple observations with shared geographic identifiers (state or county FIPS codes)

LINKING BY STACKING

SubYr	StFCID	RptDt	RptDsp
2019	AL11111	20190208	1
2019	CA22222	20190323	1
2019	KY33333	20190608	2
2019	MA44444	20181023	5
2019	NY55555	20190408	5
2020	AL11111	20191201	2
2020	AZ66666	20200123	1
2020	CT77777	20200501	1
2020	KY33333	20200523	1
2020	NY55555	20191101	5
2020	NY55555	20200623	1

Rearrange by
StFCID and RptDt



SubYr	StFCID	RptDt	RptDsp
2019	AL11111	20190208	1
2020	AL11111	20191201	2
2020	AZ66666	20200123	1
2019	CA22222	20190323	1
2020	CT77777	20200501	1
2019	KY33333	20190608	2
2020	KY33333	20200523	1
2019	MA44444	20181023	5
2019	NY55555	20190408	5
2020	NY55555	20191101	5
2020	NY55555	20200623	1

CAUTIONARY NOTES ABOUT LINKING BY STACKING

- Child identifiers are state-specific
 - Children experiencing recorded events in multiple states will therefore appear as multiple children
- States change identification procedures
 - Children experiencing recorded events within a state may therefore appear as multiple children
- Prior or future events may be unobserved
 - Children experiencing recorded events outside the sampling frame will appear as having too few events

SELECTED NCANDS CHILD FILE MULTI-YEAR LINKAGE RATES

State	00-01	01-02	02-03	03-04	04-05	05-06	06-07	07-08	08-09	09-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17	17-18	18-19	19-20	20-21
Arizona			0	99	99	99	99	99	100	99	99	99	99	99	99	99	99	99	99	100	
California			99	100	99	100	98	99	99	99	99	99	99	99	99	99	99	99	99	100	99
Florida	0	0	100	98	99	99	0	99	99	0	99	0	0	99	99	98	98	98	98	99	99
Illinois			0	98	99	0	0	0	0	100	100	0	0	100	100	99	99	99	99	99	99
Indiana			94	99	99	99	99	99	100	99	99	0	98	98	99	98	98	98	98	98	98
Massachusetts	0	0	95	100	97	95	96	97	98	96	96	97	97	96	96	95	100	100	98	98	98
Michigan			0	99	99	99	0		96	97	97	97	97	96	98	99	100	100	99	100	100
New York			0	98	96	98	98	98	98	98	98	98	96	96	99	97	98	97	98	98	98
Oregon												0	100	100	100	99	100	100	99	99	99
Pennsylvania	0	0	0	98	0	0	0	0	0	0	0	94	0	95	0	0	0	0	0	0	97
Tennessee				0	95	96	96	97	98	80	0	0	98	97	97	97	97	99	98	97	98
Texas	0	0	0	98	96	97	97	99	98	98	98	98	98	98	98	98	95	95	95	96	96

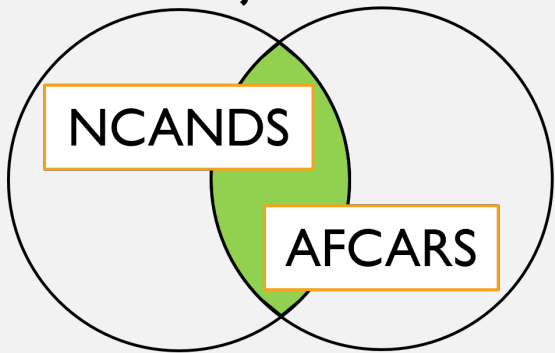
Note: Numbers are percentages of children in NCANDS Child Files, linked by ChID in adjacent years, for whom DOB and sex match (true positive link).

MERGING BY COLUMN: JOINING DATA

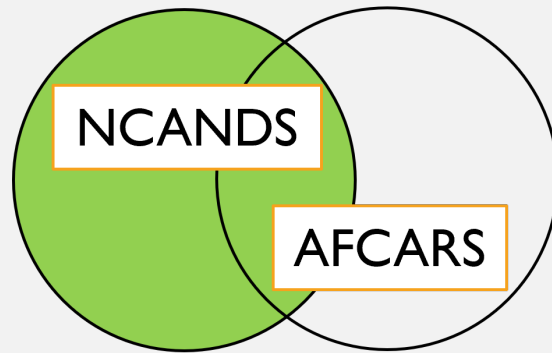
- The NCANDS and AFCARS (as well as NYTD) can be linked to each other by adding columns to each row, or in R, “joining” the data
- Joining is more complex than stacking

JOINING DATA

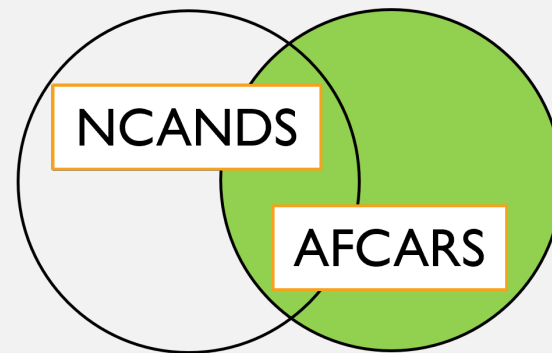
Inner
join



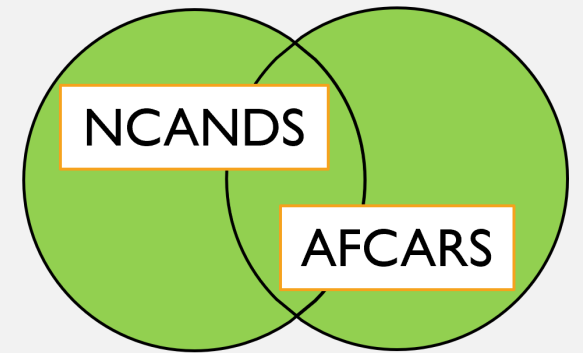
Left join



Right join



Full join



JOINING MULTIPLE OBSERVATIONS

- **One to one**
 - Linking variable is unique in both datasets
- **One to many**
 - Linking variable is unique in one dataset
- **Many to many**
 - Linking variable is not unique in either dataset

LINKING BY JOINING: AGGREGATE DATA

NCANDS

FYs tat	StaTerr	NRpt
2019	AL	4610
2019	CA	68539
2019	DC	2425
2019	FL	29390
2019	GA	12059
2020	AL	5892
2020	CA	77480
2020	DC	3024
2020	FL	27501
2020	GA	13234



AFCARS

FY	StaTerr	NRem
2019	AL	245
2019	CA	3702
2019	DC	165
2019	FL	1492
2019	GA	679
2020	AL	322
2020	CA	3899
2020	DC	154
2020	FL	2090
2020	GA	698



Linked data

FY	StaTerr	NRpt	NRem
2019	AL	4610	245
2019	CA	68539	3702
2019	DC	2425	165
2019	FL	29390	1492
2019	GA	12059	679
2020	AL	5892	322
2020	CA	77480	3899
2020	DC	3024	154
2020	FL	27501	2090
2020	GA	13234	698

LINKING BY JOINING: MICRODATA

- Question:
 - *What are the maltreatment histories of children aged 0-3 (NCANDS) placed into foster care in FY2023 (AFCARS)?*
- Process:
 - Sample using 2023 AFCARS Foster Care File
 - Stack and summarize 2020-2024 NCANDS Child Files
 - Left-join one-to-one using unique child identifier StFCID
 - Analyze maltreatment histories of placed children

DEMONSTRATION IN R

ADDITIONAL RESOURCES

- *R for Data Science* (2e), Hadley Wickham, Mine Çetinkaya-Rundel, and Garrett Grolemund

QUESTIONS?

USER SUPPORT
NDACANSUPPORT@CORNELL.EDU

ALEX ROEHRKASSE
AROHRKASSE@BUTLER.EDU

ALYSSA LINDSEY
ALYSSA.LINDSEY@UCSF.EDU

NEXT WEEK...

Date: July 22nd

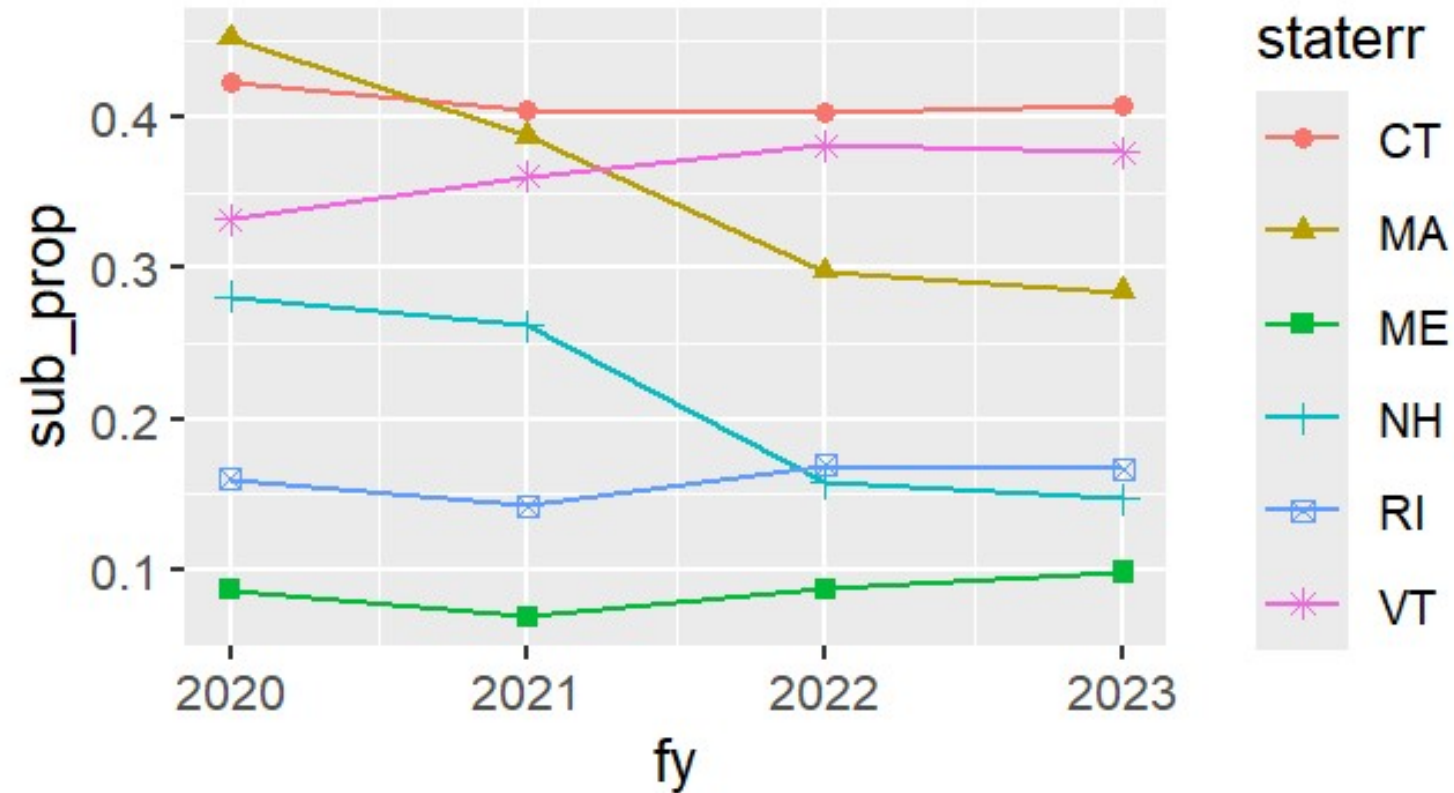
Topic: Handling missing data

Instructor: Alex Roehrkasse

COMMENTS IN THE ZOOM CHAT

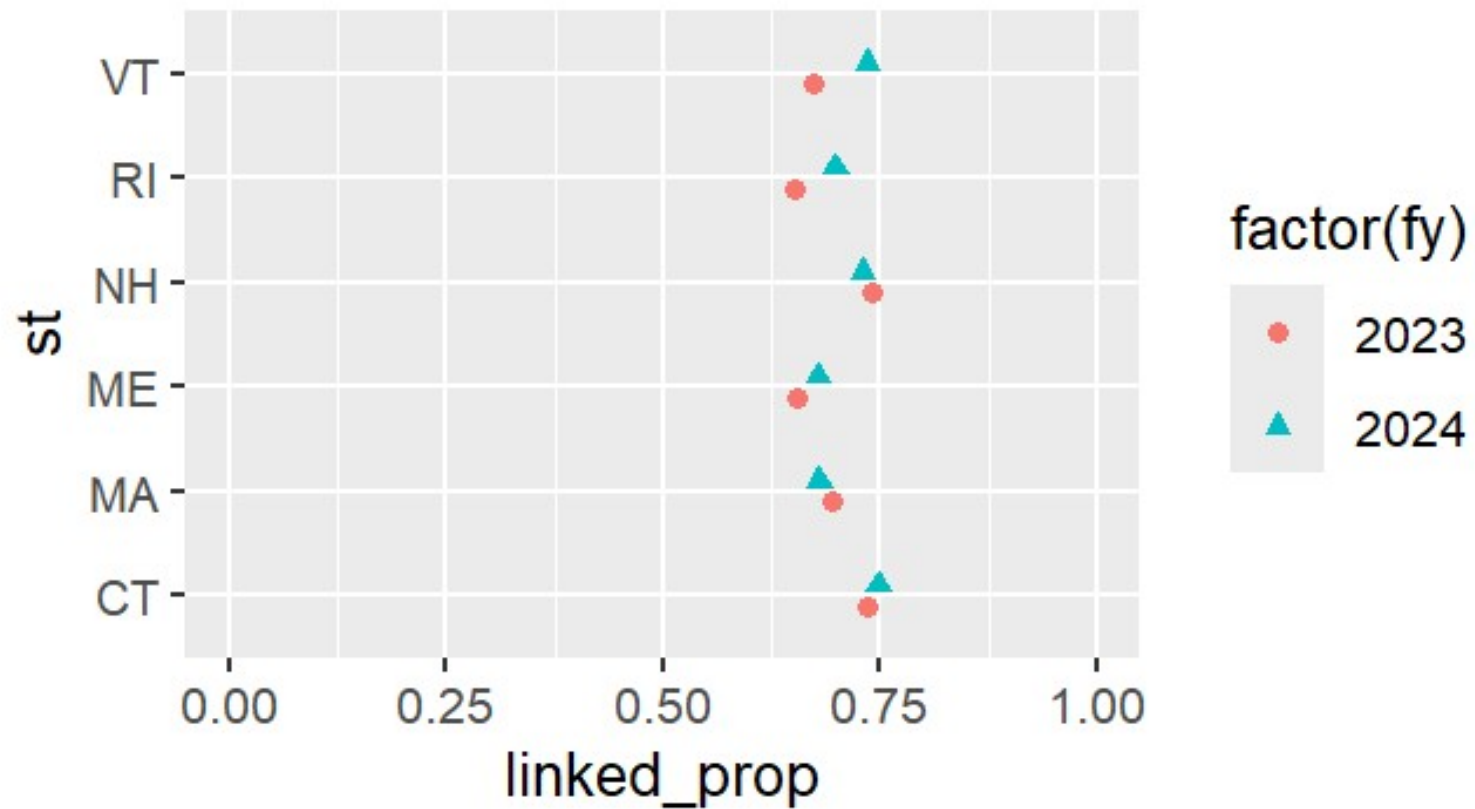
- Email the NDACAN archive at NDACAN@cornell.edu with general questions, and email NDACANSUPPORT@cornell.edu with technical support questions.
- The NDACAN archive website <https://www.ndacan.acf.hhs.gov> contains the following:
 - The R code and output of this presentation.
 - Information and instructions for joining the Child-Maltreatment-Research-L (CMRL) list serve.
 - The FAQ Frequently Asked Questions about using the NDACAN archive data including AFCARS, NCANDS, and NYTD.
 - Instructions for ordering restricted datasets

R CODE IMAGE I



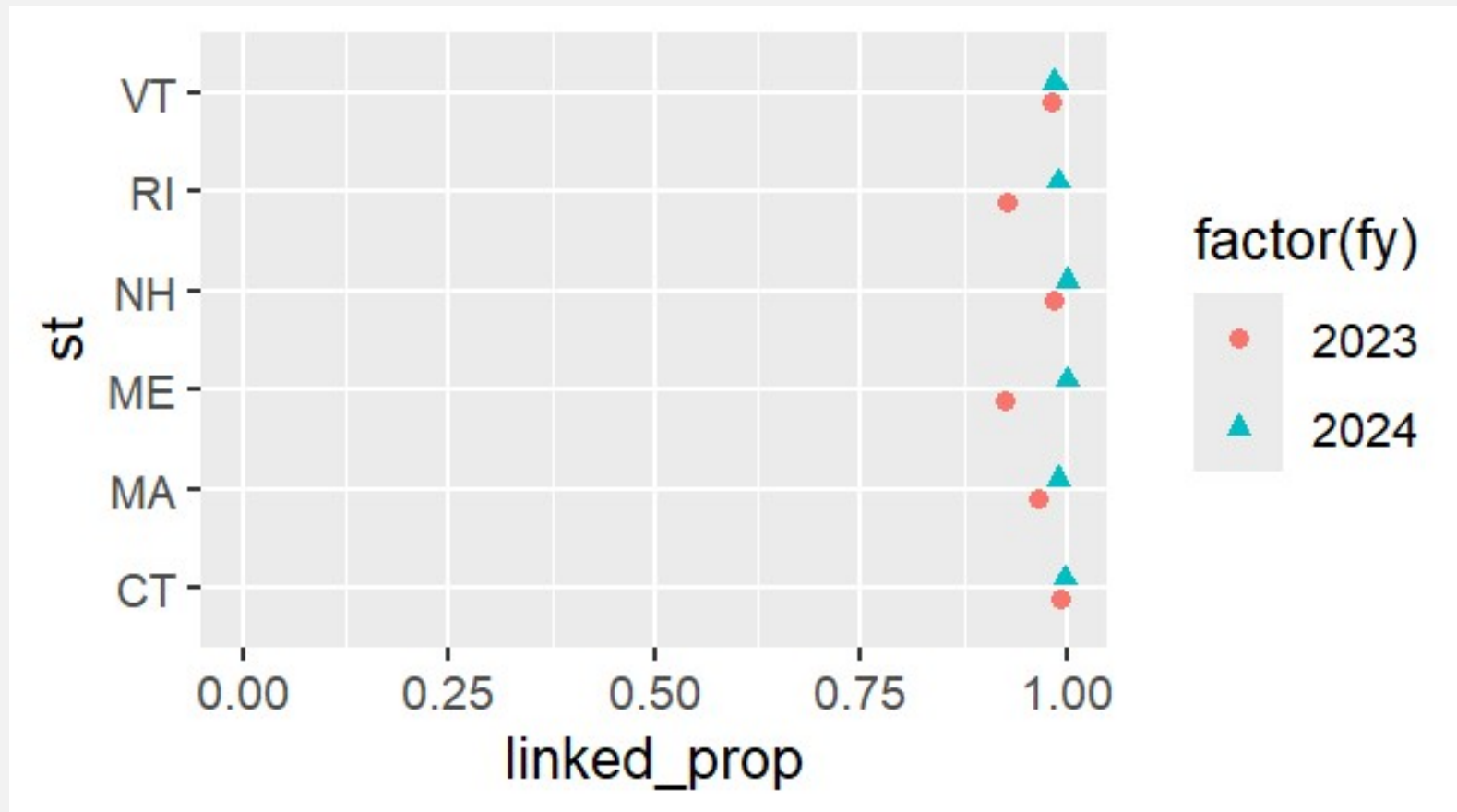
A panel of state-years with the linking variables (a combination of state and year) and the outcome of interest over time (substantiation proportion).

R CODE IMAGE 2



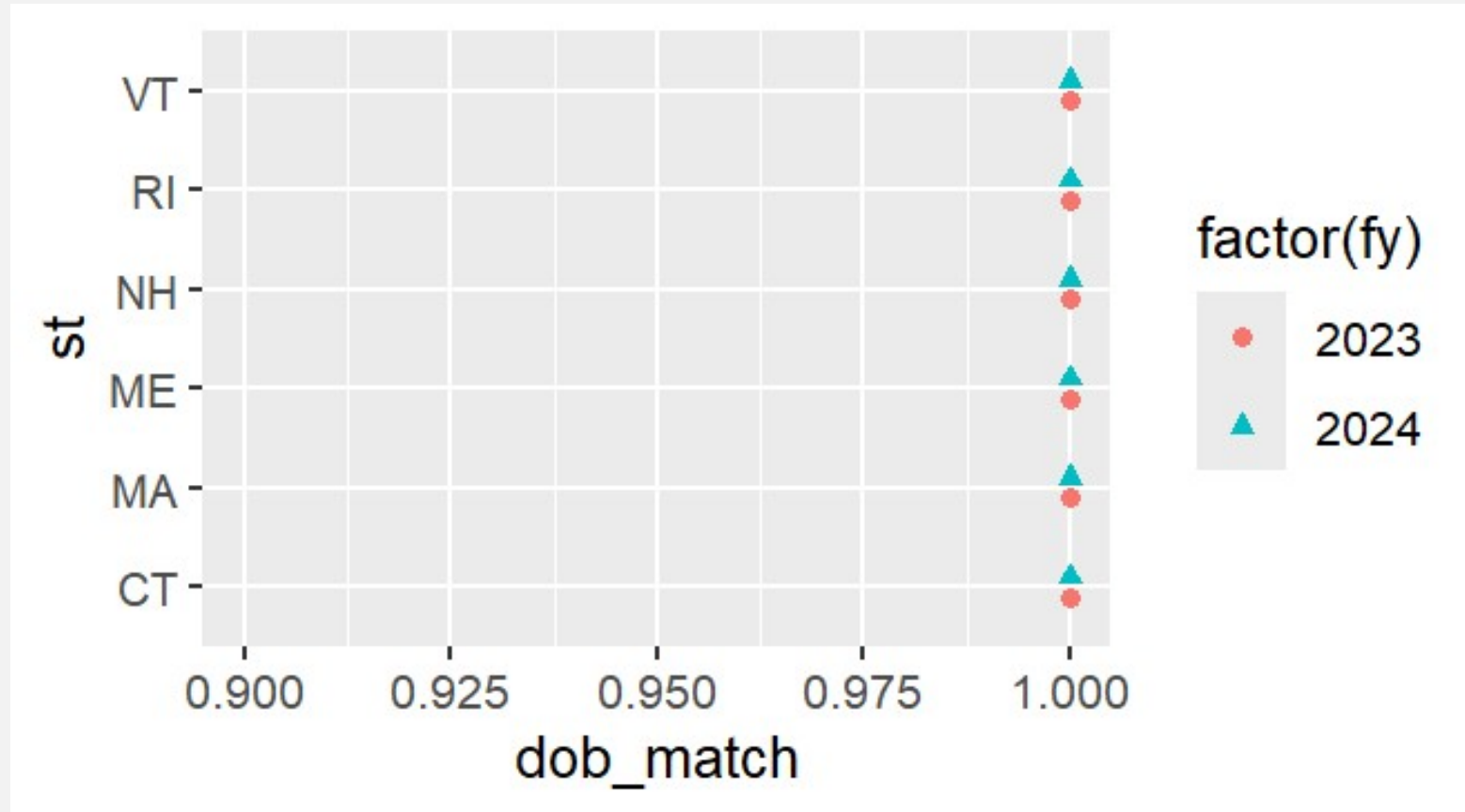
Graph of proportion of children with records in 2023 and in 2024.

R CODE IMAGE 3



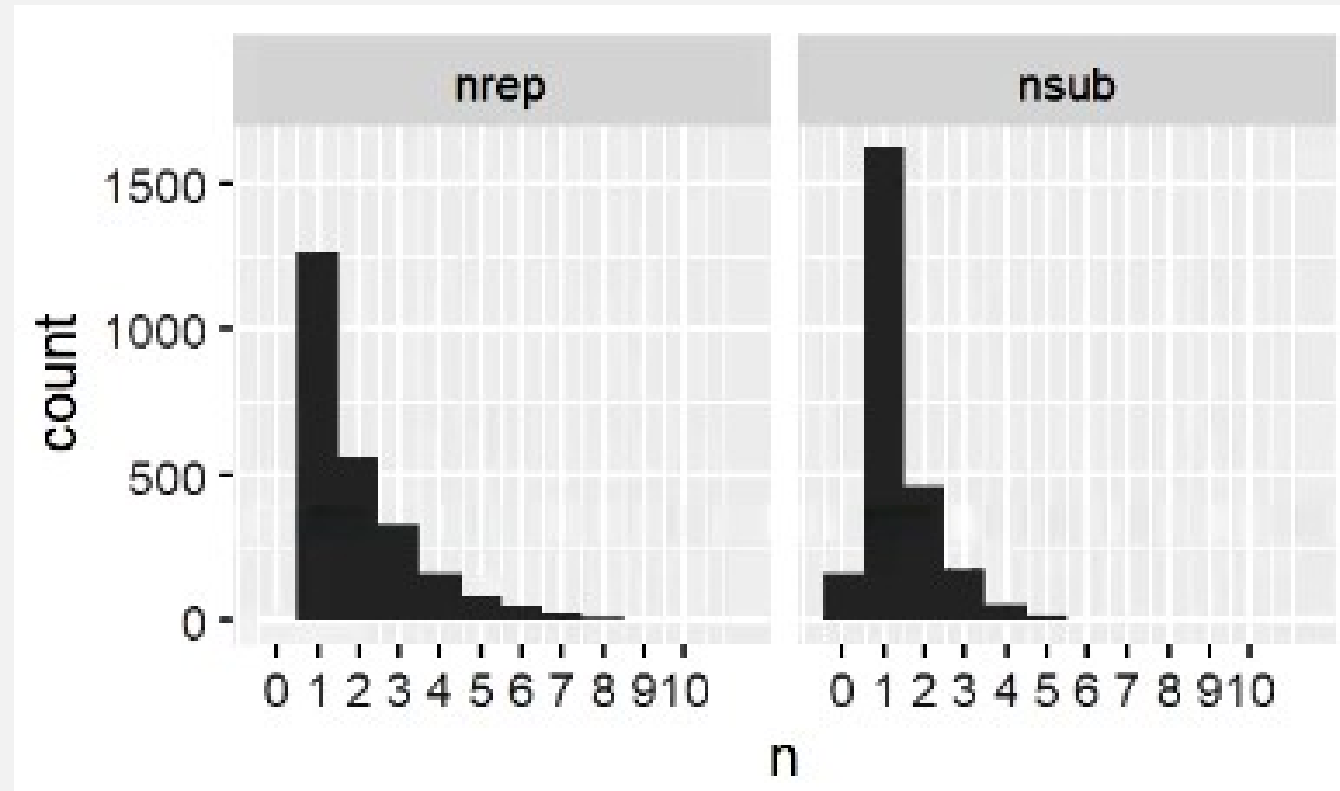
Graph of proportion of children who were in foster care at the very end of 2023 or in foster care at the very start of 2024.

R CODE IMAGE 4



Graph of proportion of child records we have a linked and have a matching date of birth.

R CODE IMAGE 5



Histogram showing distribution of maltreatment reports and substantiated maltreatment reports experienced by children aged 0-3 placed into foster care in 2023.