

Participating in the Fragile Families Challenge Activity

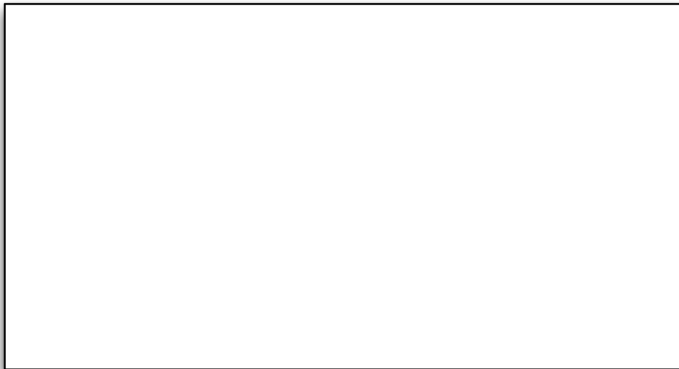
Matthew Salganik, Ian Lundberg, Alex Kindel, Sara McLanahan,
and people from around the world

Funding for FFCWS provided by NICHD (R01HD36916, R01HD39135, R01HD40421) and a consortium of private foundations, including the Robert Wood Johnson Foundation. Funding for FFC provided by the Russell Sage Foundation, NSF, & the Overdeck Fund. FFC Board of Advisors: Jeanne Brooks-Gunn, Kathryn Edin, Barbara Engelhardt, Irwin Garfinkel, Moritz Hardt, Dean Knox, Nicholas Lemann, Karen Levy, Sara McLanahan, Arvind Narayanan, Timothy Nelson, Matthew Salganik, Brandon Stewart & Duncan Watts.



4,242 families

12,942 features
birth to age 9



6 outcomes
age 15

Training

Leaderboard

Holdout

Introducing the outcome variables

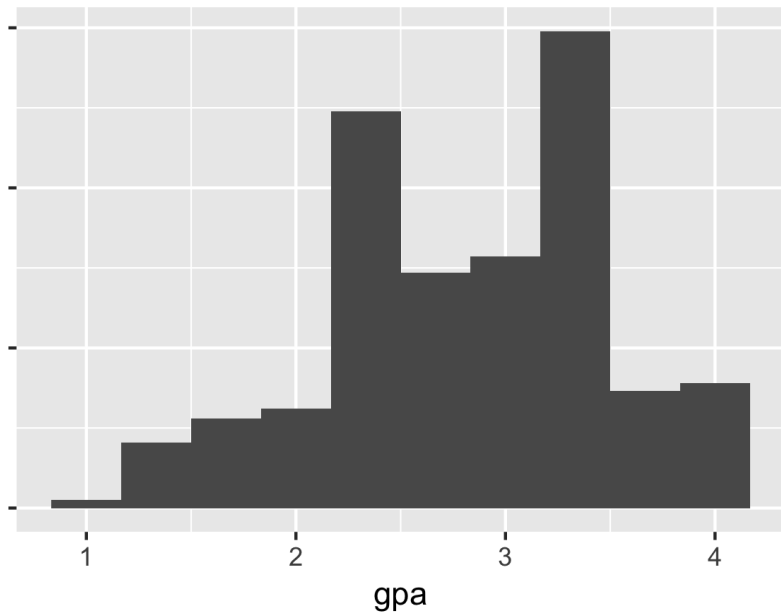
GPA¹

¹Learn more at <http://www.fragilefamilieschallenge.org/gpa/>

B20. At the {most recent grading period/last grading period in the spring} what was your grade in ...

		A	B	C	D OR LOWE R	NO GRADE OR PASS/FAIL	REF	DK	N/A HOMESCHOOLED
B20A	English or language arts? ..	1	2	3	4	5	-1	-2	7 → GO TO B22A
B20B	Math?	1	2	3	4	5	-1	-2	7 → GO TO B22A
B20C	History or social studies? ..	1	2	3	4	5	-1	-2	7 → GO TO B22A
B20D	Science?	1	2	3	4	5	-1	-2	7 → GO TO B22A

²This variable is reverse-coded in the data file so that higher values represent higher GPAs.



Grit³

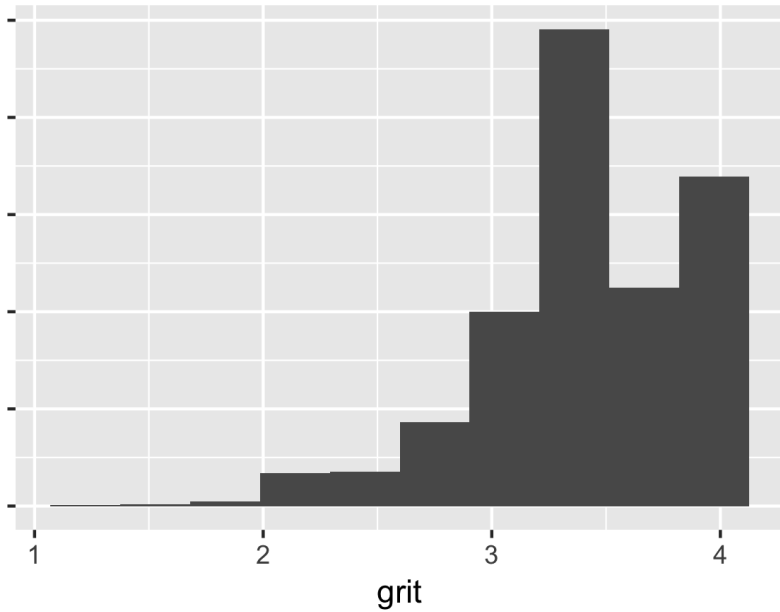
³Learn more at <http://www.fragilefamilieschallenge.org/grit/>

D2. Thinking about how you have behaved or felt during the past four weeks, please tell me whether you strongly agree, somewhat agree, somewhat disagree, or strongly disagree with the following statements.

PROBE: Thinking about the past four weeks, do you strongly agree, somewhat agree, somewhat disagree, or strongly disagree with this statement?

		STRONGLY AGREE	SOMEWHAT AGREE	SOMEWHAT DISAGREE	STRONGLY DISAGREE	REF	DK
D2I.	I keep at my schoolwork until I am done with it.....	1	2	3	4	-1	-2
D2K.	Once I make a plan to get something done, I stick to it.....	1	2	3	4	-1	-2
D2M.	I finish whatever I begin.....	1	2	3	4	-1	-2
D2V.	I am a hard worker	1	2	3	4	-1	-2

⁴This variable is reverse-coded in the data file so that higher values represent more grit.



Material hardship⁵

⁵Learn more at <http://www.fragilefamilieschallenge.org/material-hardship/> » « ≡ » ≡ ↺ 🔍 ↻

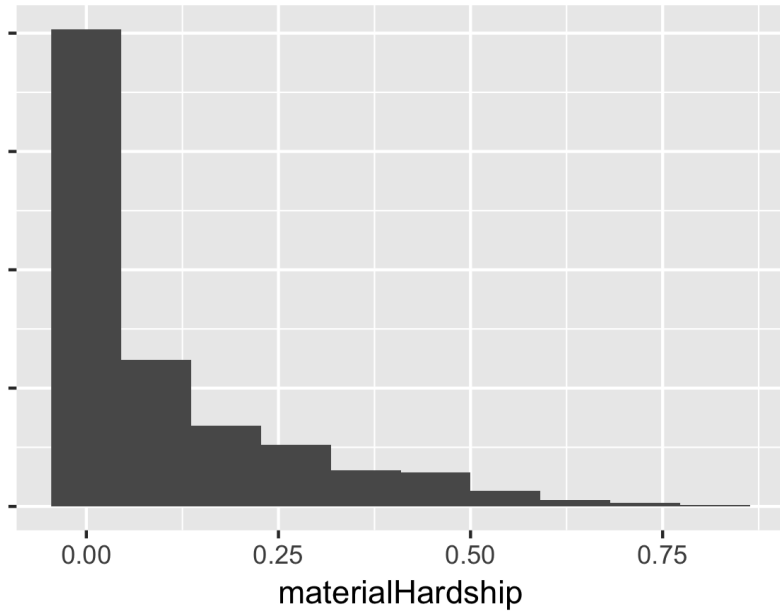
Material hardship

We are also interested in some of the problems that families face making ends meet. In the past twelve months, did you do any of the following because there wasn't enough money?

		YES	NO	REF	DK
J37.	In the past twelve months, did you receive free food or meals?	1	2	-1	-2
J38.	In the past twelve months, were you ever hungry, but didn't eat because you couldn't afford enough food?	1	2	-1	-2
J39.	In the past twelve months, did you ever not pay the full amount of rent or mortgage payments?	1	2	-1	-2
J40.	In the past twelve months, were you evicted from your home or apartment for not paying the rent or mortgage?	1	2	-1	-2
J41.	In the past twelve months, did you not pay the full amount of gas, oil, or electricity bill?	1	2	-1	-2
J42.	In the past twelve months, was your gas or electric services ever turned off, or the heating oil company did not deliver oil, because there wasn't enough money to pay the bills?	1	2	-1	-2
J43.	In the past twelve months, did you borrow money from friends or family to help pay bills?	1	2	-1	-2
J44.	In the past twelve months, did you move in with other people even for a little while because of financial problems?	1	2	-1	-2

Material hardship

J45.	In the past twelve months, did you stay at a shelter, in an abandoned building, an automobile or any other place not meant for regular housing, even for one night?	1	2	-1	-2
J46.	In the past twelve months, was there anyone in your household who needed to see a doctor or go to the hospital but couldn't go because of the cost?	1	2	-1	-2
J47.	In the past twelve months, was your telephone service (mobile or land line) cancelled or disconnected by the telephone company because there wasn't enough money to pay the bill?	1	2	-1	-2



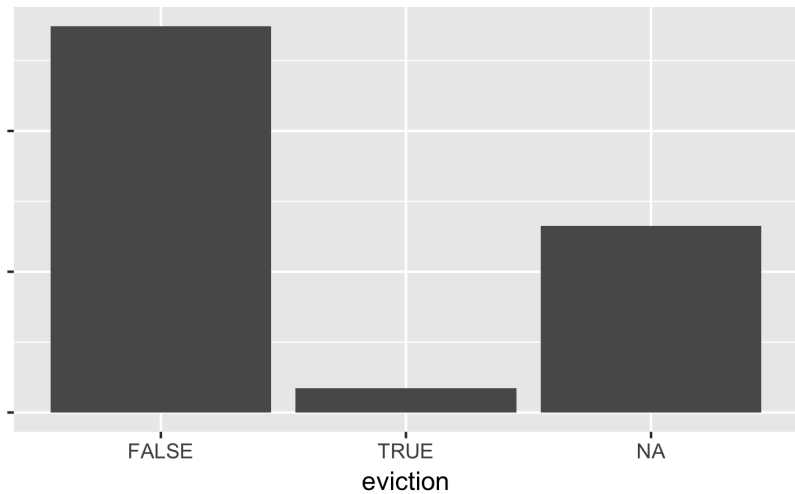
Eviction⁶

⁶Learn more at <http://www.fragilefamilieschallenge.org/eviction/>

Eviction

J51. Since {MONTH AND YEAR COHORT CITY FIELDDED IN YR 9}, were you evicted from your home or apartment for not paying the rent or mortgage?

YES	1
NO	2
REFUSED	-1
DON'T KNOW	-2



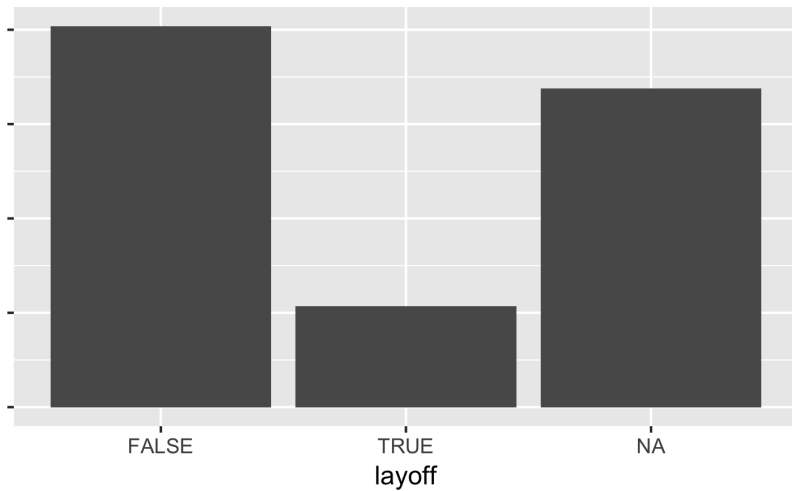
Caregiver layoff⁷

⁷Learn more at <http://www.fragilefamilieschallenge.org/layoff/>

Caregiver layoff

K13. Since {MONTH AND YEAR COHORT CITY FIELDDED IN YR 9}, have you been laid off from your employer for any time?

YES.....	1
NO.....	2
REFUSED.....	-1
DON'T KNOW	-2



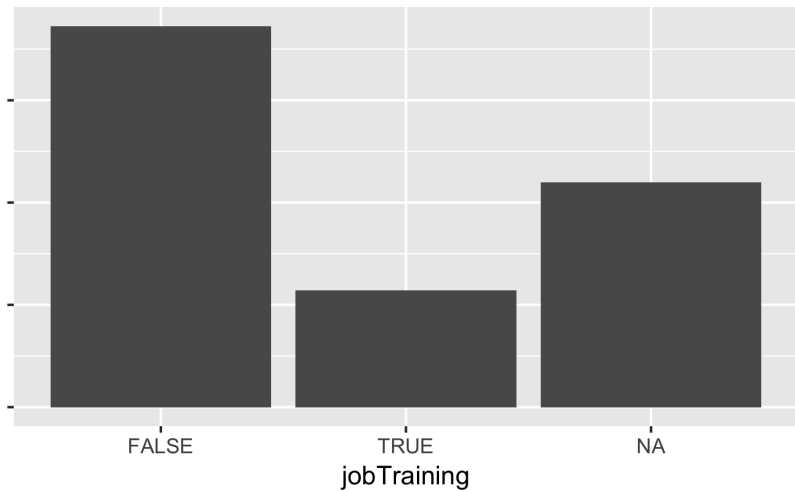
Job training⁸

⁸Learn more at <http://www.fragilefamilieschallenge.org/job-training/>

Caregiver job training

K4. Since {MONTH AND YEAR COHORT CITY FIELDDED IN YR 9}, have you taken any classes to improve your job skills, such as computer training or literacy classes?

YES	1
NO	2
REFUSED	-1
DON'T KNOW	-2



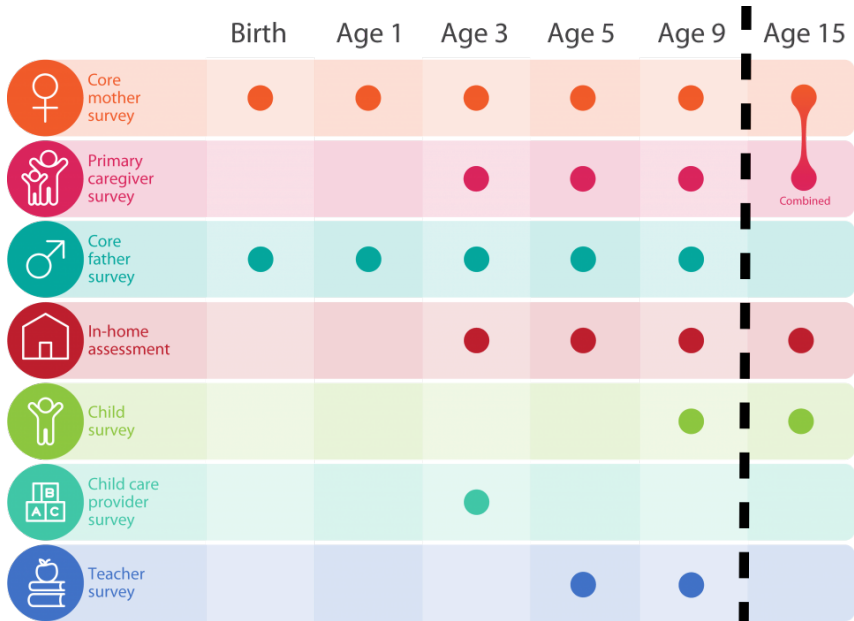
1. Prepare data
2. Create predictive model
3. Upload submission

1. Prepare data
2. Create predictive model
3. Upload submission

The Fragile Families and Child Wellbeing Study is a dataset of real people who have selflessly opened up their lives to us for the last 15 years so that their experiences can contribute to scientific research. By participating in the Fragile Families Challenge activity, you become a collaborator in this project. It is of the utmost importance that you respect the families in the data by using what they have told us responsibly.

- ▶ When you applied to get access to the data, you completed a data use agreement

- ▶ When you applied to get access to the data, you completed a data use agreement
- ▶ After this activity you should delete the data from your computer unless you plan to keep using it for research



RStudio Source Editor																						
f5f x																						
Filter																						
	f5f1	f5f1a	f5f2	f5f3	f5f3a	f5f3b	f5f3b1	f5f4	f5f4a	f5f4b	f5f5	f5f6	f5f7a	f5f7b	f5f7c	f5f8a1	f5f8b1	f5f8c1	f5f8a2	f5f8b2	f5f8c2	f5f8a:
1	-9	-9	-9	-9.00	-9.00	-9.000000	-3	-9.000000	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	
2	-9	-9	-9	-9.00	-9.00	-9.000000	-3	-9.000000	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	
3	-9	-9	-9	-9.00	-9.00	-9.000000	-3	-9.000000	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	
4	-6	-6	-6	-6.00	-6.00	-6.000000	-3	-6.000000	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	
5	-6	-6	-6	-6.00	-6.00	-6.000000	-3	-6.000000	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	
6	2	-6	-6	-6.00	-6.00	-6.000000	-3	-6.000000	-6	-6	-6	-6	2	2	2	2	-6	-6	1	12	600	
7	-6	-6	-6	-6.00	-6.00	-6.000000	-3	-6.000000	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	
8	1	3	2	-6.00	-6.00	-6.000000	-3	196.26903	1	2	2	2	2	2	2	2	-6	-6	2	-6	-6	
9	-6	-6	-6	-6.00	-6.00	-6.000000	-3	-6.000000	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	
10	1	1	1	-6.00	-6.00	-6.000000	-3	1310.50458	-10	2	2	2	2	2	2	2	-6	-6	2	-6	-6	
11	1	5	4	149734....	11087.07	1.271596	-3	-6.000000	-6	-6	-6	-6	1	2	2	2	-6	-6	2	-6	-6	
12	-9	-9	-9	-9.00	-9.00	-9.000000	-3	-9.000000	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	
13	1	2	1	-6.00	-6.00	-6.000000	-3	1024.43600	3	2	2	2	2	2	2	2	-6	-6	2	-6	-6	
14	-6	-6	-6	-6.00	-6.00	-6.000000	-3	-6.000000	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	
15	1	5	3	-6.00	-6.00	-6.000000	-3	-6.000000	-6	-6	2	-6	2	2	2	2	-6	-6	2	-6	-6	
16	2	-6	-6	-6.00	-6.00	-6.000000	-3	-6.000000	-6	-6	-6	-6	2	2	2	2	-6	-6	2	-6	-6	
17	-6	-6	-6	-6.00	-6.00	-6.000000	-3	-6.000000	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	
18	1	3	1	-6.00	-6.00	-6.000000	-3	877.82490	1	2	2	2	1	1	2	1	3	176	2	-6	-6	
19	-9	-9	-9	-9.00	-9.00	-9.000000	-3	-9.000000	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	
20	-6	-6	-6	-6.00	-6.00	-6.000000	-3	-6.000000	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	
21	-9	-9	-9	-9.00	-9.00	-9.000000	-3	-9.000000	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	
22	1	1	1	-6.00	-6.00	-6.000000	-3	720.26305	1	2	2	2	1	2	2	2	-6	-6	2	-6	-6	

Showing 1 to 24 of 4,242 entries

How do I know what these variables are?

FF Fragile Families & Child Wellbeing Study

PRINCETON | COLUMBIA



Search...



Home

About

People

Publications

Data and Documentation

Contact

Spotlight on

Spotlight on FFCWS and Incarceration Research

Recent publications using the Fragile Families & Child Wellbeing data provide a family-wide view of the effects

[General Documentation](#)[Baseline](#)[Year 1](#)

**Links to
documentation
for each wave**

[Year 3](#)[Year 5](#)[Year 9](#)[Year 15](#)[FAQ](#)

Data and Documentation

Data

Data are free to download from Princeton University's [Office of Population Research \(OPR\) data archive](#).

Currently, there are five waves of publicly available data including baseline and Year 1, Year 3, Year 5, and Year 9 follow-ups. In order to protect the confidentiality of survey respondents, geographic identifiers, medical records data, contextual data (i.e., census tract characteristics), macroeconomic indicators, and genetic biomarkers are not available in the public use data files. Researchers may apply for these data via a [restricted use contract](#).

Documentation

Year 9

The Year 9 follow-up wave of data collection took place from 2007 to 2010, which makes the data useful for researchers interested in the effects of the Great Recession on children and families. It is different from previous waves because the home visit was integral to the wave procedures. In previous waves, we conducted core interviews before proceeding to the in-home components. At year 9, our initial interview was with the child's primary caregiver (usually the mother) and we scheduled a home visit at the time of that initial interview. As part of the home visit, we interviewed focal children for the first time. We attempted teacher surveys through the mail. Similar to previous waves, we have core interviews with mothers and fathers. Restricted Data at this wave include **census tract characteristics** of mother and father residences, **macroeconomic indicators**, administrative data on children's **school characteristics**, and **genetic** data from saliva samples from the mother and focal child.

PRIMARY CAREGIVER

[Primary Caregiver Survey](#)[Primary Caregiver Self-Administered](#)

SCALES

[Scales documentation](#)

MOTHER

[Questionnaire](#)[Codebook](#)

FATHER

[Questionnaire](#)[Codebook](#)

CHILD

[Child Survey](#)[Home Visit Workbook](#)[Interviewer Observations](#)[Codebook](#)

TEACHER

Each survey has
a questionnaire
and
a codebook

Questionnaire:

BOX A3A2

**IF PCG=BIOFATHER IN THE PCG IDENTIFIER IN THE SCREENER,
GO TO A3C.**

**ELSE IF PCG= NON-PARENT AND RELATIONSHIP = MATERNAL
GRANDPARENT(S), PATERNAL GRANDPARENT(S), OTHER
RELATIVES OR FRIEND IN THE PCG IDENTIFIER IN THE
SCREENER, GO TO A3B1A.**

**ELSE IF PCG=NON-PARENT AND RELATIONSHIP=FOSTER CARE
IN THE PCG IDENTIFIER IN THE SCREENER, GO TO A3B.**

**ELSE IF PCG=NON-PARENT AND RELATIONSHIP = OTHER,
SPECIFY IN THE PCG IDENTIFIER IN THE SCREENER, GO TO
A3B1A.**

**ELSE IF PCG= "NOT MOTHER" IN THE PCG IDENTIFIER GO TO
A3D.**

A3B. Are {CHILD}'s foster parents related to you?

YES 1
NO 2
REFUSED -1



<https://www.talk2solicitors.co.uk/blog/limits-frustration-contract/>

Improving Metadata Infrastructure for Complex Surveys: Insights from the Fragile Families Challenge

**Alexander T. Kindel¹, Vineet Bansal¹, Kristin D. Catena¹,
Thomas H. Hartshorne¹, Kate Jaeger¹, Dawn Koffman¹,
Sara McLanahan¹, Maya Phillips¹, Shiva Rouhani¹, Ryan Vinh¹,
and Matthew J. Salganik¹**

Kindel et al. (2019)

Abstract

Researchers rely on metadata systems to prepare data for analysis. As the complexity of data sets increases and the breadth of data analysis practices grow, existing metadata systems can limit the efficiency and quality of data preparation. This article describes the redesign of a metadata system supporting the Fragile Families and Child Wellbeing Study on the basis of the experiences of participants in the Fragile Families Challenge. The authors demonstrate how treating metadata as data (i.e., releasing comprehensive information about variables in a format amenable to both automated and manual processing) can make the task of data preparation less arduous and less error prone for all types of data analysis. The authors hope that their work will facilitate new applications of machine-learning methods to longitudinal surveys and inspire research on data preparation in the social sciences. The authors have open-sourced the tools they created so that others can use and improve them.

Keywords

metadata, survey research, data sharing, quantitative methodology, computational social science

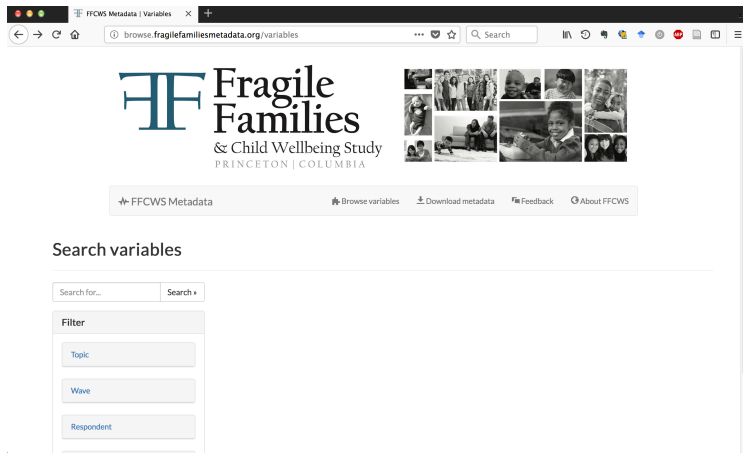
Kindel et al. (2019)

- ▶ Data preparation is a barrier to progress in predictive modeling in the social sciences.

- ▶ Data preparation is a barrier to progress in predictive modeling in the social sciences.
- ▶ New approach to metadata: Human-readable → machine-actionable.

Kindel et al. (2019)

How do I know what these variables are?



`http://metadata.fragilefamilies.princeton.edu/variables`

Introducing cm1relf

<http://metadata.fragilefamilies.princeton.edu/variables/cm1relf>

You can filter variables by:

- ▶ Topic
- ▶ Wave
- ▶ Respondent
- ▶ Variable Type (e.g., continuous, ordered categorical, unordered categorical, etc).

Direct access to the metadata:

- ▶ API: `api.metadata.fragilefamilies.princeton.edu`
- ▶ Python package: `github.com/fragilefamilieschallenge/ffmetadata-py`
- ▶ R package: `github.com/fragilefamilieschallenge/ffmetadata`
- ▶ Raw metadata: `api.fragilefamiliesmetadata.org/get_metadata`

Why `cm1re1f`?

Why cm1relf?

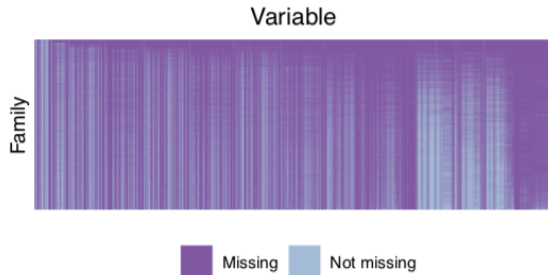
Response type	Respondent	Wave	Leaf
(blank): questionnaire c: constructed	m - mother f - father p - primary caregiver k - child t - teacher h - home o - observations n - non parental caregiver d - child care center r - family care center u - post center observations q - couple	1 - baseline 2 - year 1 3 - year 3 4 - year 5 5 - year 9 6 - year 15	(letter): survey section + (number): question number OR (string): Constructed variable ID OR (<u>string</u>): national or city weight

Advice for data preparation

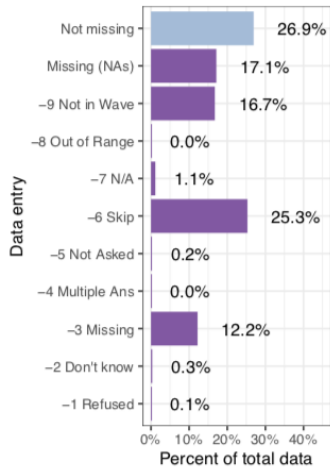
- ▶ do something about missing data
- ▶ do something about unordered categorical variables

Advice for data preparation

Deal with missing data



(a)



(b)

Advice for data preparation

Deal with unordered categorical variables

F3B1. Whose name is on the mortgage for this house?

CODE ONE

FATHER'S NAME ONLY 1➡ **GO TO BOX F3C**

MOTHER'S OR CURRENT PARTNER'S NAME
ONLY 2➡ **GO TO BOX F3C**

BOTH FATHER'S NAME AND MOTHER'S
OR CURRENT PARTNER'S NAMES 3➡ **GO TO BOX F3C**

FAMILY MEMBER(S) ON THE
MOTHER'S OR CURRENT PARTNER'S SIDE 4➡ **GO TO BOX F3C**

FAMILY MEMBER(S) ON THE FATHER'S
SIDE 5➡ **GO TO BOX F3C**

OTHER (SPECIFY)..... 91➡ **GO TO BOX F3C**

REFUSED..... -1➡ **GO TO BOX F3C**

DON'T KNOW..... -2➡ **GO TO BOX F3C**

BOX F3C

1. Prepare data
2. Create predictive model
3. Upload submission

1. Prepare data
2. Create predictive model
3. Upload submission

Here are the approaches that participants used in the Challenge (Key ideas are flexibility and over-fitting):

- ▶ general linear model (linear and logistic regression)
- ▶ Tree-based methods (e.g., random forrest, gradient boosted trees)
- ▶ Regularization approaches (e.g., LASSO, Ridge, Elastic Net)

1. Prepare data
2. Create predictive model
3. Upload submission

1. Prepare data
2. Create predictive model
3. Upload submission

Building a submission

Submissions include:

1. Predictions
2. Code
3. Narrative explanation

- ▶ Submission preparation instructions:
www.fragilefamilieschallenge.org/upload-your-contribution/

Building a submission

Submissions include:

1. Predictions
2. Code
3. Narrative explanation

- ▶ Submission preparation instructions:
www.fragilefamilieschallenge.org/upload-your-contribution/
- ▶ We've created a template submission for you template

Building a submission


Submissions include:

1. Predictions
2. Code
3. Narrative explanation

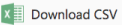
- ▶ Submission preparation instructions:
www.fragilefamilieschallenge.org/upload-your-contribution/
- ▶ We've created a template submission for you template
- ▶ You can upload to the SICSS 2020 leaderboard.

Get on the leaderboard

← → ↻ codalab.fragilefamilieschallenge.org/#results 🔍 ☆ 📄 ⋮

 **Fragile Families**
CHALLENGE

Help Sign Up Sign In

 Download CSV

Results

#	User	GPA ▲	Grit ▲	Material hardship ▲	Eviction ▲	Layoff ▲	Job training ▲
1	wjlei1990	0.36854 (1)	0.21896 (3)	0.02436 (1)	0.05341 (7)	0.17435 (5)	0.20224 (3)
2	OldDriver.ffc	0.37099 (2)	0.22979 (18)	0.02471 (2)	0.05341 (7)	0.17435 (5)	0.20224 (3)
3	yjpeng	0.37120 (3)	0.21759 (2)	0.02493 (3)	0.05223 (2)	0.17048 (1)	0.20169 (2)
4	hamidrezaomidvar	0.37136 (4)	0.22191 (13)	0.02523 (5)	0.05227 (3)	0.18784 (7)	0.21409 (7)
5	t.f.schaffner	0.37143 (5)	0.21755 (1)	0.02499 (4)	0.05660 (8)	0.22453 (9)	0.27736 (9)
6	andrewor	0.37143 (5)	0.21755 (1)	0.02499 (4)	0.06038 (10)	0.26792 (13)	0.30755 (13)
7	pc12	0.37583 (6)	6.18762 (29)	0.03536 (24)	0.94340 (18)	0.77547 (19)	0.72264 (18)
8	mannyg	0.37789 (7)	0.21997 (7)	0.02880 (17)	0.05341 (7)	0.17435 (5)	0.20224 (3)
9	ppz	0.37810 (8)	0.23896 (19)	0.02859 (14)	0.12830 (16)	0.30755 (15)	0.36981 (16)
10	lazs	0.38407 (9)	0.22054 (9)	0.02877 (16)	0.05660 (8)	0.22453 (9)	0.27736 (9)
11	mloyola	0.38644 (10)	0.21969 (4)	0.02880 (17)	0.05341 (7)	0.17435 (5)	0.20224 (3)
12	agalle	0.38846 (11)	0.22137 (11)	0.02740 (8)	0.05341 (7)	0.17435 (5)	0.20224 (3)
13	weggert	0.38868 (12)	0.24682 (21)	0.02546 (6)	0.05660 (8)	0.24528 (12)	0.29245 (11)
14	jeremvfreese	0.39077 (13)	0.22060 (10)	0.02803 (12)	0.05295 (5)	0.17379 (4)	0.20132 (1)

Powered by Codalab v0.1.1

1. Prepare data
2. Create predictive model
3. Upload submission

Remember this is not easy, especially in a short amount of time. Many valid submissions to the Challenge did not beat the 4 variable regression model.

At the end of the activity we will open the holdout data.
Anyone that beats the best submission to the Challenge
will receive a prize.

Good luck and enjoy