

Day 5: Mass collaboration and machine learning

Matthew J. Salganik
Department of Sociology
Princeton University



Happy Juneteenth

<https://www.pbs.org/wnet/african-americans-many-rivers-to-cross/history/what-is-juneteenth/>

Today is our final organized lunch tables (you are free to self-organize)

Today is our final organized lunch tables (you are free to self-organize)

► Juneteenth

Today is our final organized lunch tables (you are free to self-organize)

- ▶ Juneteenth
- ▶ Hosting a partner location for SICSS 2021

Today is our final organized lunch tables (you are free to self-organize)

- ▶ Juneteenth
- ▶ Hosting a partner location for SICSS 2021
- ▶ Other topics

If you want to learn about hosting a partner location but can't attend, talk to Chris or Matt during office hours in week 2

Feedback on the feedback

Feedback on the feedback

- ▶ People liked groups of 3

Feedback on the feedback

- ▶ People liked groups of 3
- ▶ Group composition seems to matter

Feedback on the feedback

- ▶ People liked groups of 3
- ▶ Group composition seems to matter
- ▶ Still a few struggles with remote collaboration

Feedback on the feedback

- ▶ People liked groups of 3
- ▶ Group composition seems to matter
- ▶ Still a few struggles with remote collaboration
- ▶ We heard good things about Zoom remote control

Feedback on the feedback

- ▶ People liked groups of 3
- ▶ Group composition seems to matter
- ▶ Still a few struggles with remote collaboration
- ▶ We heard good things about Zoom remote control
- ▶ Tension between structured and open-ended activities

Feedback on the feedback

- ▶ People liked groups of 3
- ▶ Group composition seems to matter
- ▶ Still a few struggles with remote collaboration
- ▶ We heard good things about Zoom remote control
- ▶ Tension between structured and open-ended activities
- ▶ Keep the feedback coming

Other logistics questions/comments

Schedule for today:

[https:](https://compsocialscience.github.io/summer-institute/2020/duke/schedule)

[//compsocialscience.github.io/summer-institute/2020/duke/schedule](https://compsocialscience.github.io/summer-institute/2020/duke/schedule)

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.



Learning objectives:

Learning objectives:

- ▶ Practice converting raw survey data to a suitable format

Learning objectives:

- ▶ Practice converting raw survey data to a suitable format
- ▶ Practice building predictive models

Learning objectives:

- ▶ Practice converting raw survey data to a suitable format
- ▶ Practice building predictive models
- ▶ Experience the common task method

4,242 families

Birth to age 9
12,942 variables

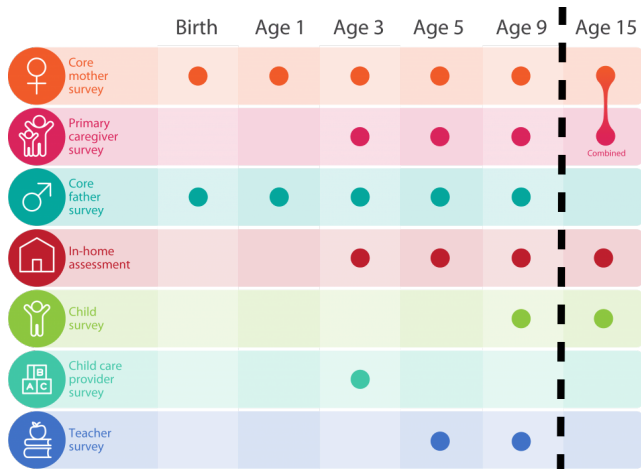
Background

Age 15
6 outcomes

Training

Leaderboard

Holdout



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. Honor that agreement.

- ▶ When you applied to get access to the data, you completed a data use agreement. Honor that agreement.
- ▶ After this activity you should delete the data from your computer. If you plan to do more research, you can download it again.

Many submissions to the Challenge had three main steps:

- ▶ data cleaning

Many submissions to the Challenge had three main steps:

- ▶ data cleaning
- ▶ feature selection

Many submissions to the Challenge had three main steps:

- ▶ data cleaning
- ▶ feature selection
- ▶ statistical learning

Many submissions to the Challenge had three main steps:

- ▶ data cleaning
- ▶ feature selection
- ▶ statistical learning

A reasonable goal for today is to upload a submission to the leaderboard

Questions about the activity?

Good luck and enjoy