



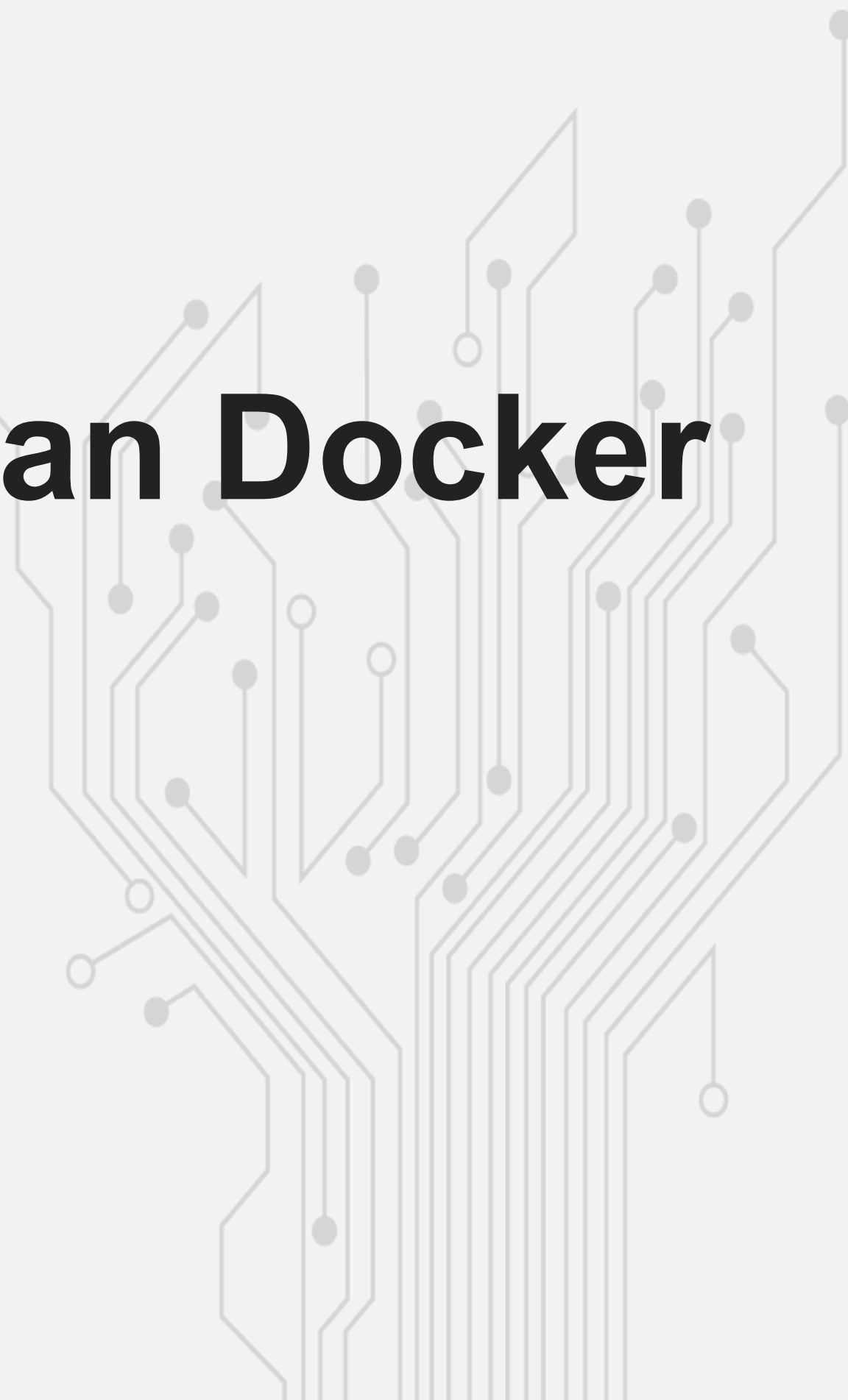
CENTER FOR  
COMPUTATIONAL BIOMEDICINE  
HARVARD MEDICAL SCHOOL

# The CCB NHANES+Stan Docker Container

Andrew Ghazi

Center for Computational Biomedicine

2024-04-26



# National Health and Nutrition Examination Survey (NHANES)



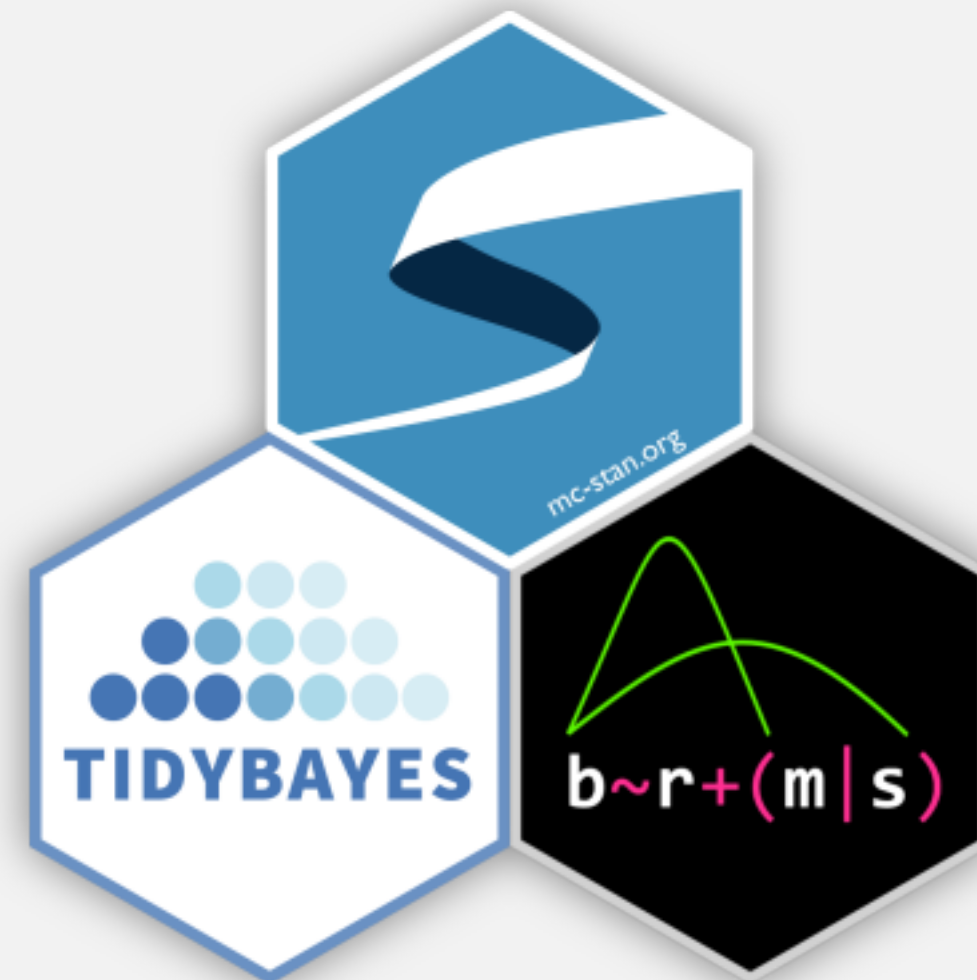
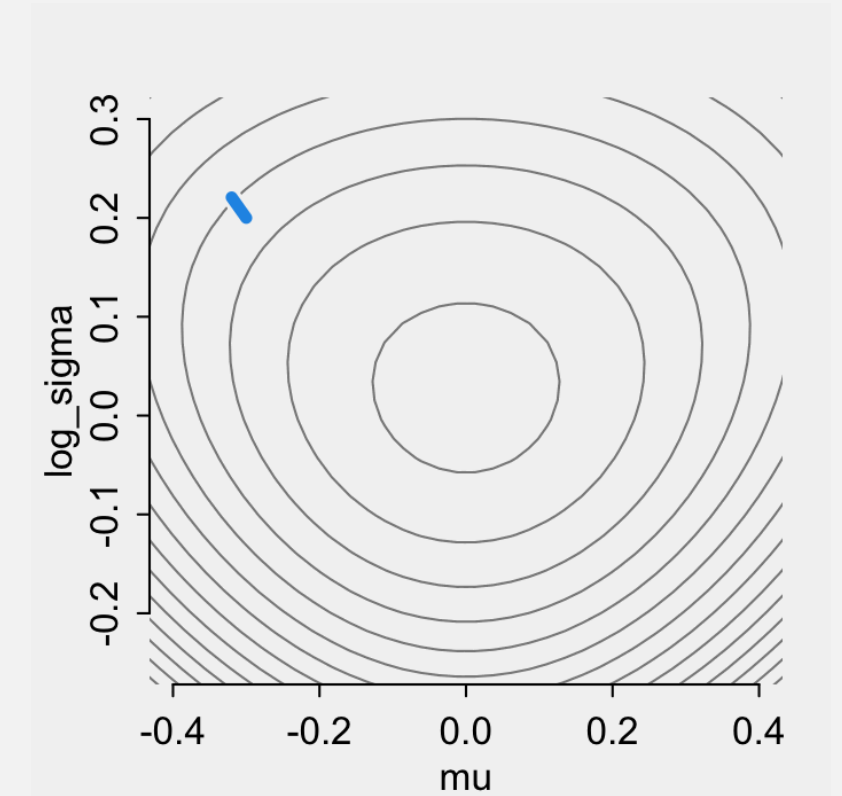
- 100s of variables: Demographic, dietary, medical exams, labs, questionnaires
- Across the US in cohorts of 5k
- Biannually since 1999, intermittently since 1959
- **Separate SAS transport files for each cohort and variable set**
- **Variable descriptions must be scraped from HTML**



# Stan

... is a state-of-the-art platform for statistical modeling and high-performance statistical computation.

- Adaptive Hamiltonian Monte Carlo sampler
- Ecosystem of modeling packages in R



# CCB Epiconductor Docker images

- CCB NHANES base image provides:

- R
- Rstudio
- tidyverse
- SQL Server
- NHANES Snapshot



National Health and Nutrition Examination Survey

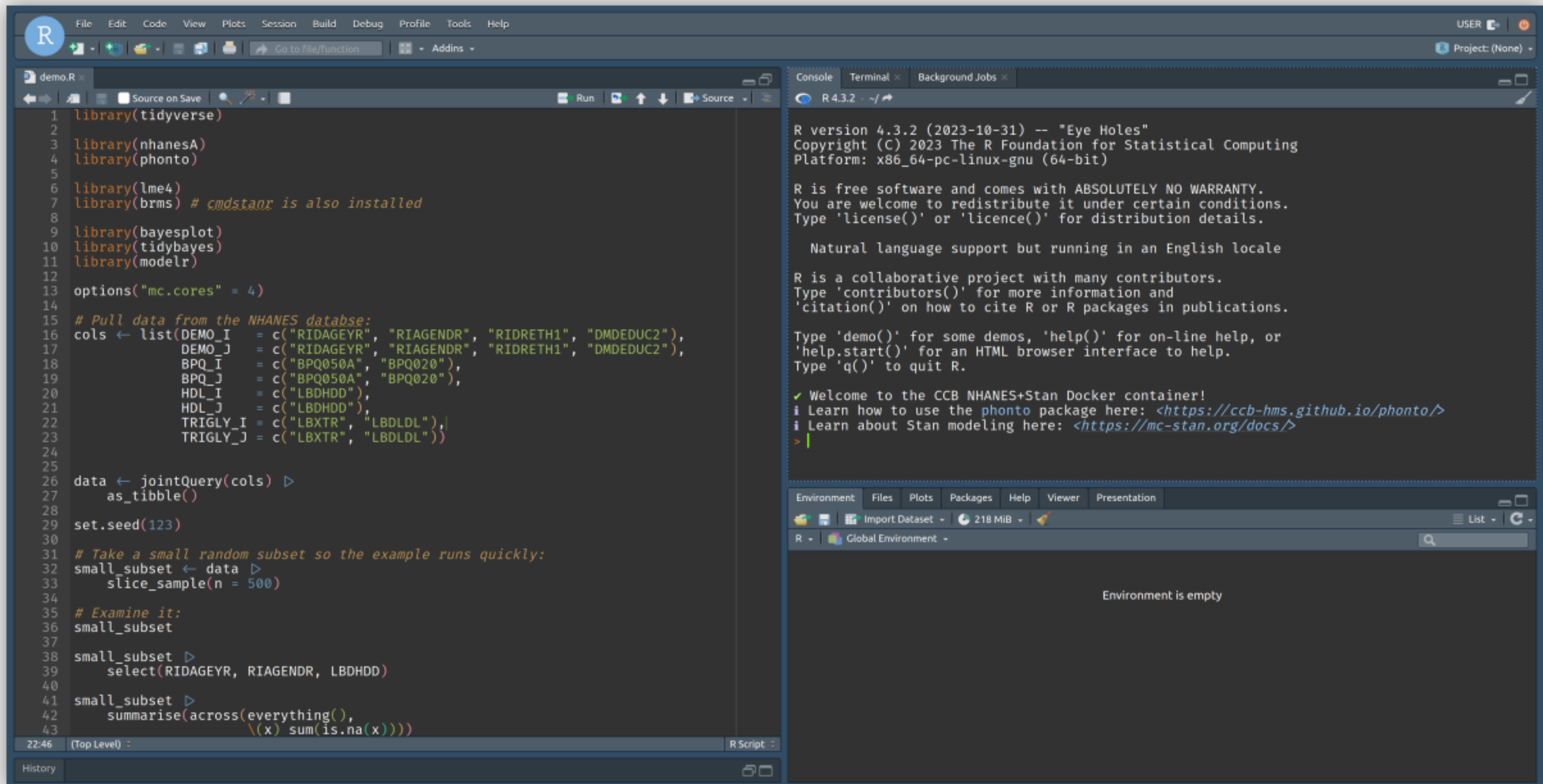
- NHANES+Stan image:

- Stan
- brms, cmdstanr, modelr, tidybayes, bayesplot, loo, ...
- beautifying bells & whistles

- Build everything and run with two commands.

+







# CCB NHANES+Stan Docker container

```
1 library(tidyverse)
2 library(nhanesA)
3 library(brms)
4 library(loo)
5
6 nhanesSearch("HDL")
7
8 cols = list(DEMO_I = c("RIDAGEYR", "RIAGENDR", "RIDRETH1"),
9             DEMO_J = c("RIDAGEYR", "RIAGENDR", "RIDRETH1"),
10             HDL_I = c("LBDHDD"),
11             HDL_J = c("LBDHDD"),
12             TRIGLY_I = c("LBXTR", "LBDLDL"),
13             TRIGLY_J = c("LBXTR", "LBDLDL"))
14
15 dat = jointQuery(cols) |>
16       filter(!is.na(LBXTR) & !is.na(LBDHDD))
17
18 m1 = brm(LBDHDD ~ RIDAGEYR*RIAGENDR, data = dat)
19
20 m2 = brm(LBDHDD ~ RIDAGEYR*RIAGENDR + log10(LBXTR), data = dat)
21
22 loo_compare(loo(m1), loo(m2))
```

# Thank you!

CCB is hiring! Careers page:

