Does Breastfeeding Cause Caries?

Brenda Onyango

4/10/2022

Background

Sociological Background on Breastfeeding

The individual decision to breastfeed an infant happens in the context of longstanding social forces including racial stereotypes and corporate advertising for formula with a global reach (Kirksey, 2021; Kaplan, 2008). In the United States, the scope of my analysis, breastfeeding rates are one of many maternal and child health disparities that fall along racial lines. In addition to the sociohistorical motivators/deterrents to breastfeeding, there are also concerns for economic well-being for the breastfeeding parent. Rippeyoung and Noonan document how breastfeeding rather than formula-feeding affects post-birth earnings (2012). Using data from the National Longitudinal Survey of Youth, they find that those who breastfeed for six months or longer experience longer wage loss than those who breastfeed for a shorter period or not at all (Rippeyoung and Noonan, 2012). Indeed, both the ability to breastfeed and the length one breastfeeds is affected by finances and biology. The dynamic between social forces and biological processes are also applicable to caries and other oral health outcomes across the lifecourse.

Breastfeeding and Child Oral Health

A 2017 review summarized the extensive health benefits of breast milk that includes preventing teeth from being misaligned (Peres et al., 2017). The review also cites studies showing breastfeeding up to 12 months has been associated with fewer caries, but there is a positive association with caries when breastfeeding lasts longer (Peres et al., 2017). The authors of the review acknowledge methodological challenges in connecting breastfeeding behavior with oral health outcomes beyond infants aged 12-24 months. This present analysis attempts to understand the causal relationship between breastfeeding and caries. "Caries is a chronic disease process that… is an imbalance between destructive and protective oral forces…cavities are the holes in teeth that result from the underlying chronic caries process" (Centers for Medicare & Medicaid Services, 2015, p. 8-9).

Research Question

The research question is does breastfeeding cause caries? Breastfeeding will be operationalized using definitions from the U.S. Centers for Disease Control and Prevention which collects several questions about breastfeeding via the National Immunization Survey landline sampling frame (U.S. CDC, n.d.). The indicators used in this study are "Infants breastfed at 12 months" and "Infants ever breastfed" (CDC[2]) from 2008; data from before 2009 in this set "represents children born over three years and each birth year can consist of respondent data from up to three years" (CDC[2]). These indicators are the independent variables. The dependent variable of child dental health is operationalized with data from the National Oral Health Surveillance System for third graders in 2016, who would be ~eight years old (CDC[3]). The indicator is "Caries Experience: Percentage of students with caries experience (treated or untreated tooth decay)."

The control variables in this analysis are fluoridation and child poverty rates from CDC[1] and the National Women's Law Center, respectively.

Causal Diagram

Childhood caries formation and breastfeeding behavior have causes not included in the dataset used in this analysis but is captured in the diagram below. These determinants include:

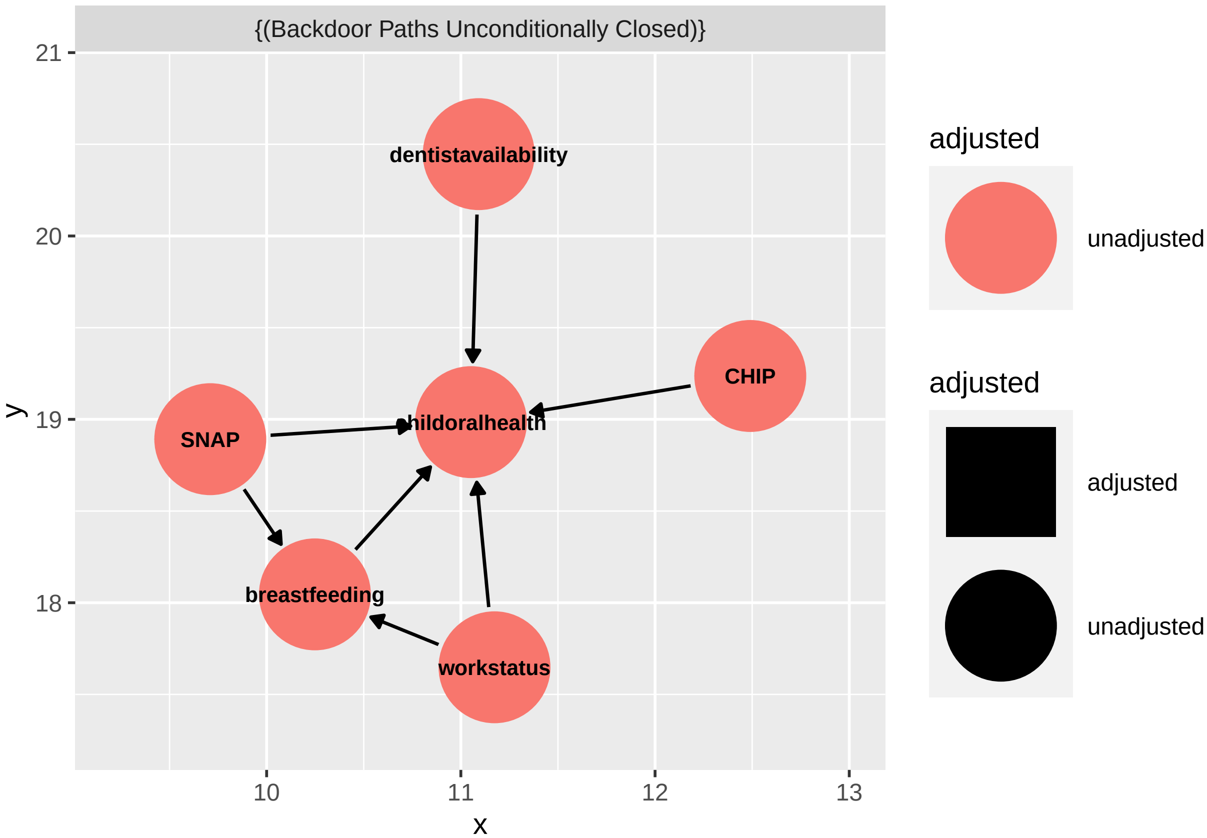
Supplemental Nutrition Assistance Program (SNAP): Colloquially known as food stamps, SNAP likely affects the physical ability to breastfeed and child oral health because people who are lactating require sufficient nutrition to produce milk and children's diets also cause of caries.

Work status: The lactating parent's work status (type of work, wages, and work hours) can be reasonably assumed to affect their ability to breastfeed and the ability to afford dental care for children.

Children's Health Insurance Program (CHIP): CHIP provides more affordable health coverage for children in families that do not qualify for Medicaid. States vary in CHIP structure and eligibility, but all CHIP plans cover dental (HHS). Having dental coverage provides children the possibility of preventive dental services.

Dentist Availability: The availability of dental experts in an area can affect parents' education on how to take care of infant and young children's oral health and the ability to access preventive services.

```
dentaldag <- dagitty::dagitty("dag {

  childoralhealth <- breastfeeding
  childoralhealth <- SNAP -> breastfeeding
  childoralhealth <- workstatus -> breastfeeding
  childoralhealth <- CHIP
  childoralhealth <- dentistavailability


  SNAP [exposure]
  childoralhealth [outcome]
}")

tidy_dagdrive <- tidy_dagitty(dentaldag)  #true exposure or predictor of interest is breastfeeding

ggdag_adjustment_set(dentaldag, node_size = 19, text_size = 2.8, text_col = 'black')
```



Methods

The analysis is linear regression of "caries experience..." on "infants ever breastfed" and "caries experience..." on "infants breastfed at 12 months" controlled for fluoridation and child poverty rates. The most complete data for caries experience came from third graders in 2016 in eight states: Colorado, Connecticut, Florida, Georgia, Idaho, Louisiana, Oregon, and Vermont (CDC[3]). This data can be reasonably assumed to correspond to the 2008 breastfeeding data (when the third graders would have been born or around 1 year old).

Data

To conduct this analysis I combined data on community water fluoridation (CDC[1]), breastfeeding (CDC[2]), child dental health (CDC[3]), and child poverty (National Women's Law Center).

First I installed and loaded packages needed to import data.

Then I imported the datasets for the independent and dependent variables .

```
bfsixmonths <- read.csv("/Users/brendaonyango/Downloads//ExportCSV (6).csv") #breastfedthroughsixmonths

bfever <- read.csv("/Users/brendaonyango/Downloads//ExportCSV (5).csv") #ever breastfed

bfandformulasixmonths <-  read.csv("/Users/brendaonyango/Downloads//ExportCSV (7).csv") #breastfed and

bftwelvemonths <- read.csv("/Users/brendaonyango/Downloads//ExportCSV (8).csv") #breastfed 12 months
```

```
childoralhealth <- read.csv("/Users/brendaonyango/Downloads//NOHSS_Child_Indicators (1).csv") #dependen
```

Data Wrangling Indepedent and Dependent Variables

Dependent Variable After importing the datasets, I need to clean and combine them.

I started with wrangling the childoralhealth set.

```
childoralhealth <- childoralhealth |> arrange(LocationAbbr) |> group_by(LocationAbbr, SchoolYearStart, (
```

```
childoralhealth2 <- childoralhealth |> filter(SchoolYearStart %in% c("2016")) |> group_by(LocationAbbr,
```

```
childoralhealthcleaner <- childoralhealth2  |> filter(Indicator == "Caries Experience: Percentage of stu
```

```
childoralhealthcleaner <- childoralhealthcleaner |> select(LocationAbbr, LocationDesc, SchoolYearStart,
```

```
names(childoralhealthcleaner)[names(childoralhealthcleaner) == 'Data_Value'] <- 'Percentevercav' #switch
```

```
childoralhealthcleaner <- childoralhealthcleaner[-(6),] #removing IHS from data
```

I had the most extensive data for children who began third grade in 2016; these students were approximately 8 years old meaning they were born around 2008. Therefore, I used breastfeeding data from 2008.

Independent Variable Next I cleaned the independent variable. After some consideration I decided to focus on using breastfeed at twelve months because this data point better temporally connects breastfeeding to the cavity data. I will also use infantseverbreastfed to capture periods of breastfeeding that were less than 12 months.

```
bftwelvemonthsclean <- bftwelvemonths |> filter(LocationAbbr == "FL"| LocationAbbr == "CO" | LocationAbb
```

```
names(bftwelvemonthsclean)[names(bftwelvemonthsclean) == 'Data_Value'] <- 'PercentBFdozenmonths' #switch
```

```
bfeverclean <- bfeverclean |> select(YearStart, LocationAbbr, LocationDesc, Data_Value, Low_Confidence_I
```

```
names(bfeverclean)[names(bfeverclean) == 'Data_Value'] <- 'percenteverbreastfed' #switching name of data
```

Next I joined the breastfed at 12 months and ever breastfed sets.

```
bfdata <- right_join(bftwelvemonthsclean, bfeverclean, by = "LocationAbbr")
```

Next, I combined the datasets with independent and dependent variables.

```
combinedset <- left_join(bfdata, childoralhealthcleaner, by = "LocationAbbr")
```

```
combinedset <- combinedset |> select(-Indicator)
```

Control Variables Because it was readily available, I included state fluoridation and childhood poverty levels for 2016 to use as control variables to better isolate the effects of breastfeeding on cavity rates.

```
library(readxl)
state_fluoride <- read_excel("state fluoride.xlsx",
    col_types = c("text", "numeric", "numeric"))
```

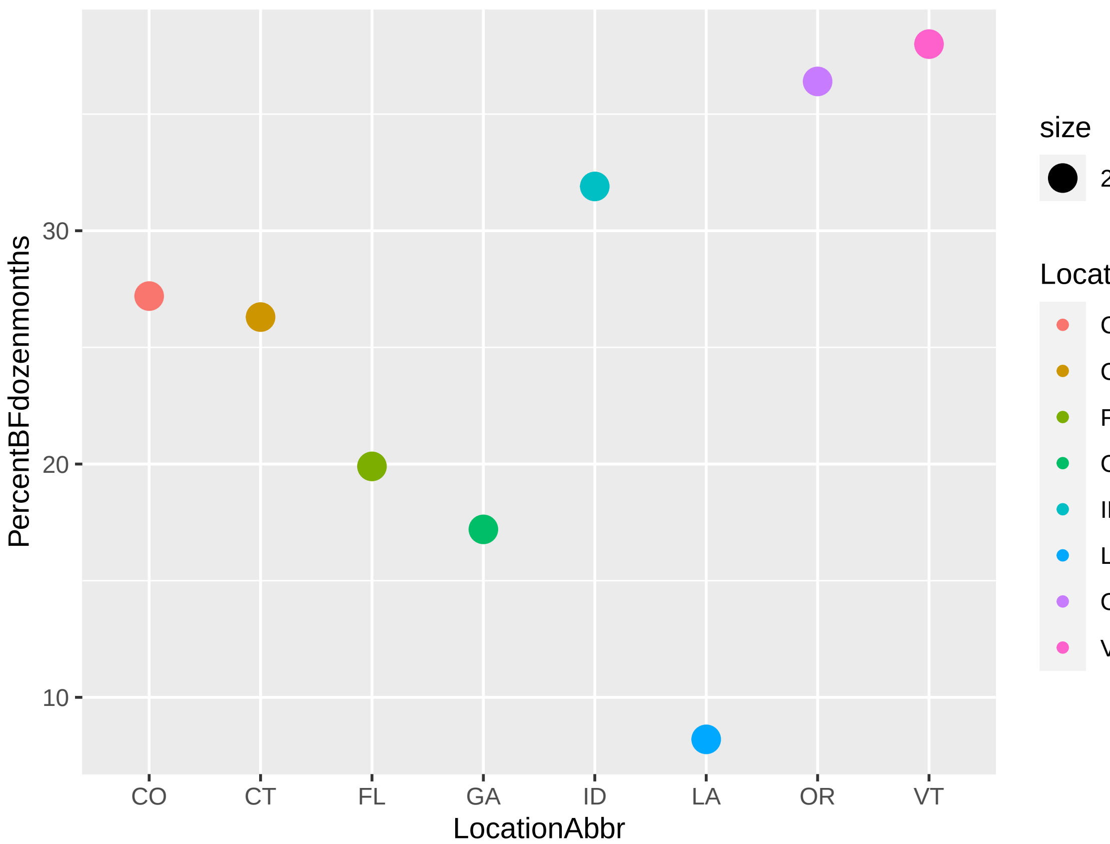
```
combinedset <- left_join(combinedset, state_fluoride, by = "LocationAbbr")
```

Before I performed regressions and visualizations, I confirmed that my combinedsetfinal object is a dataframe and not a list.

```
combinedsetfinal <- as.data.frame(combinedset) #ensures object I created in wrangling is a dataframe
```

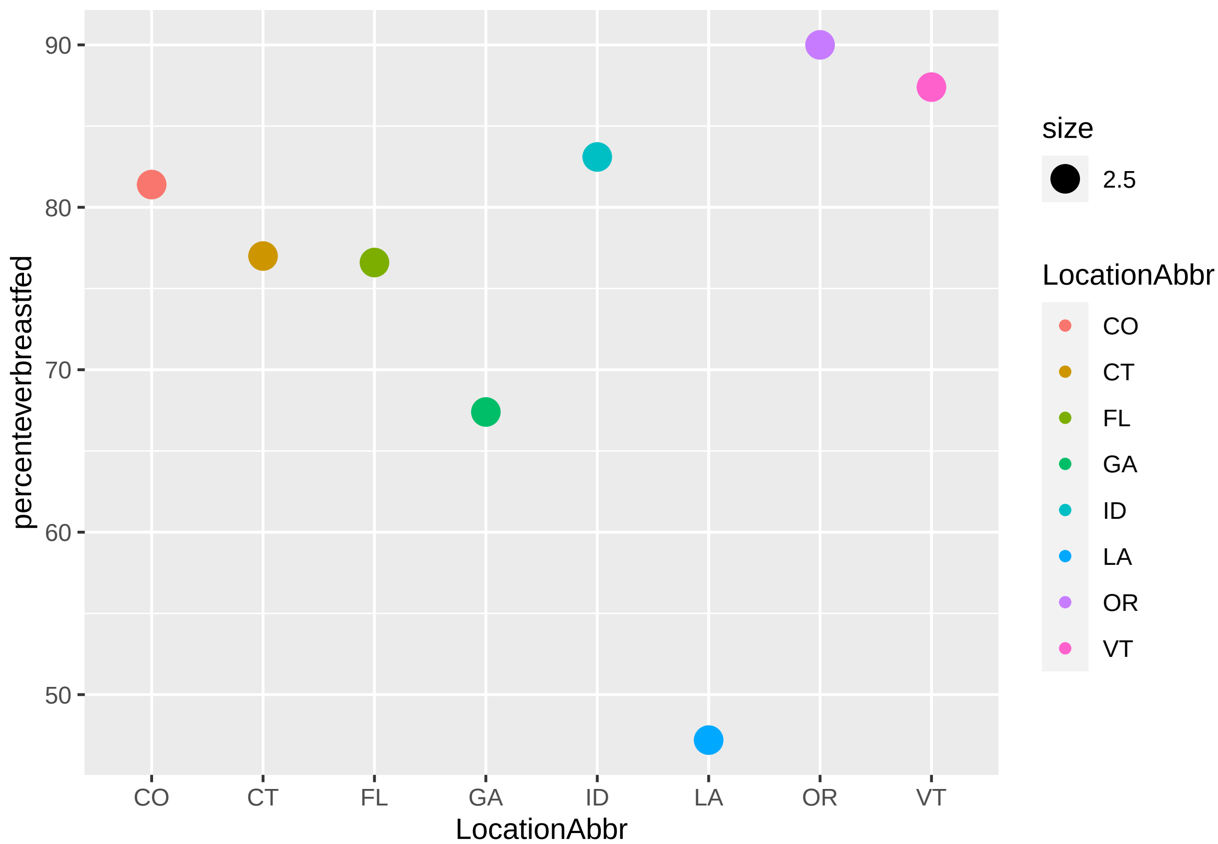
Data Visualization

```
ggplot(combinedsetfinal) + geom_point(aes(x= LocationAbbr, y=PercentBFdozenmonths, color = LocationAbbr
```



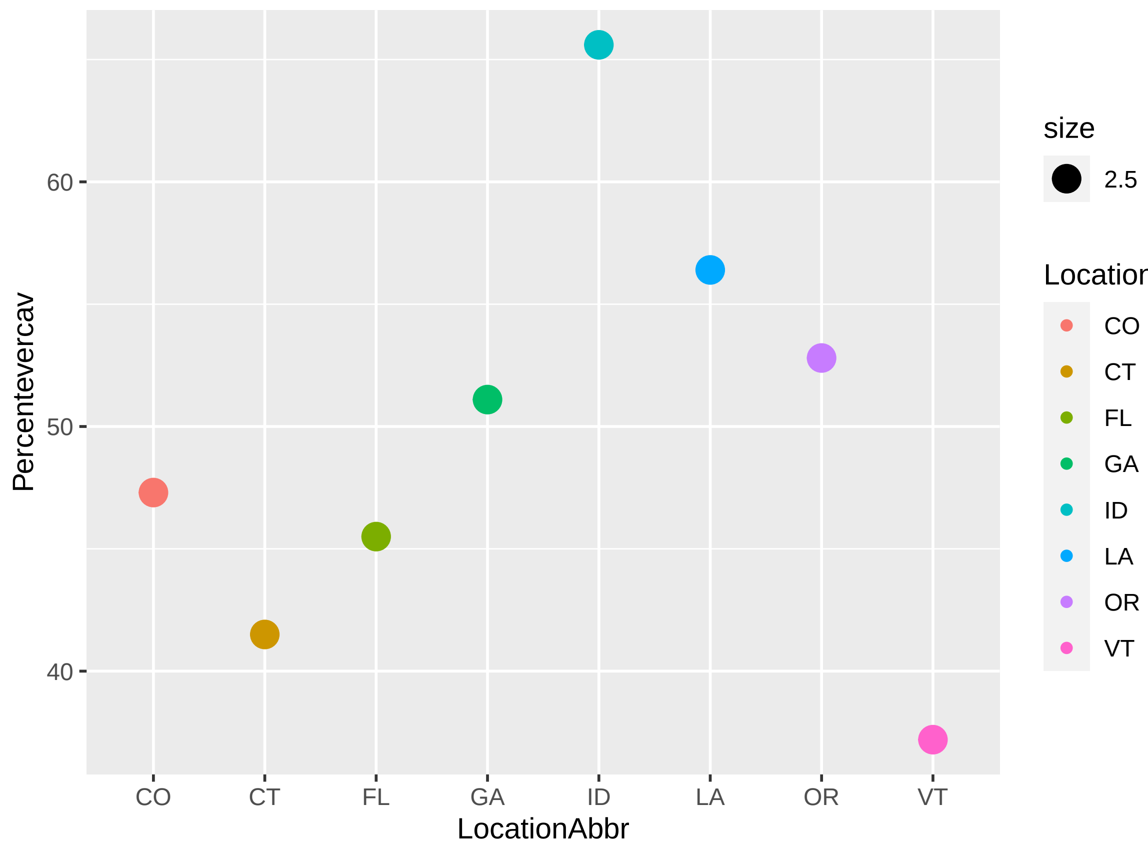
Independent Variables

```
ggplot(combinedsetfinal) + geom_point(aes(x= LocationAbbr, y=percenteverbreastfed, color =LocationAbbr,
```



From the above plots we see variation in percent of children being breastfed at 12 months and ever being breastfed in 2008, respectively. Oregon and Vermont are among the highest among both independent variables and Louisiana is the lowest.

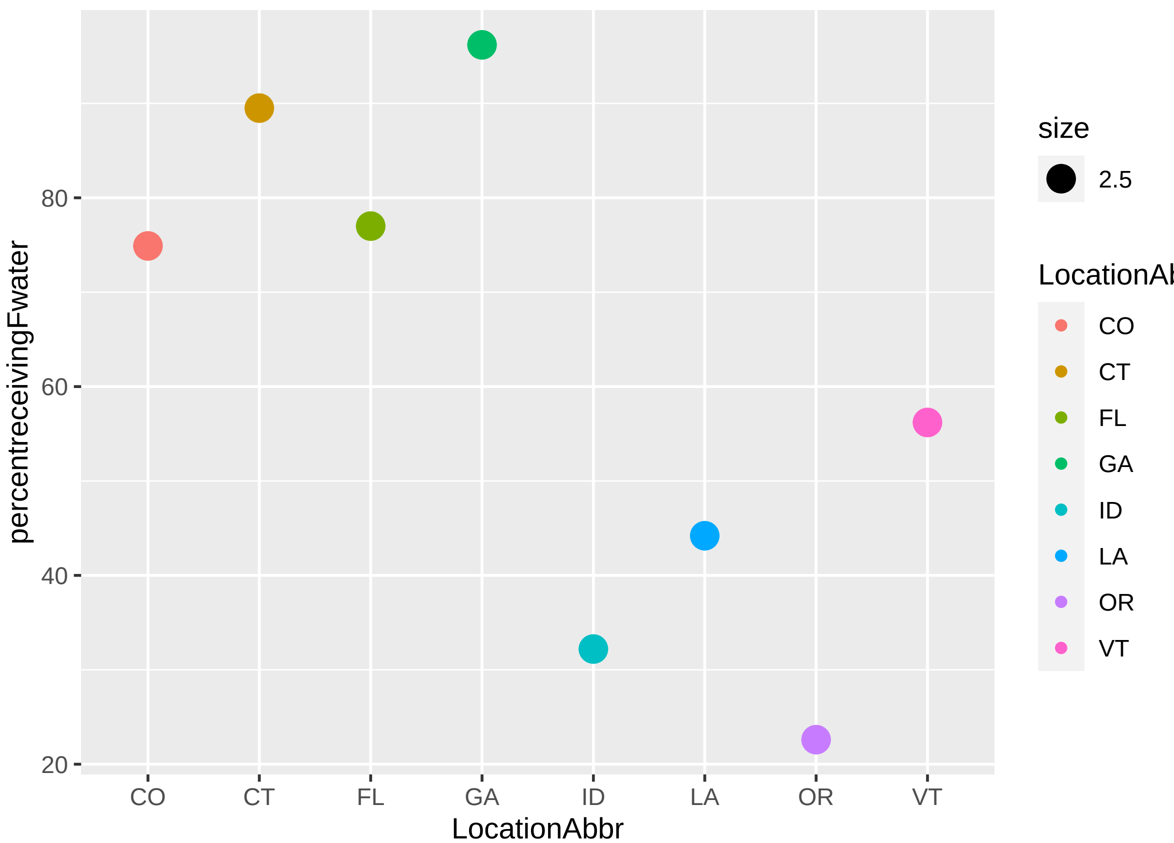
```
ggplot(combinedsetfinal) + geom_point(aes(x= LocationAbbr, y=Percentevercav, color =LocationAbbr, size =
```



Dependent Variable

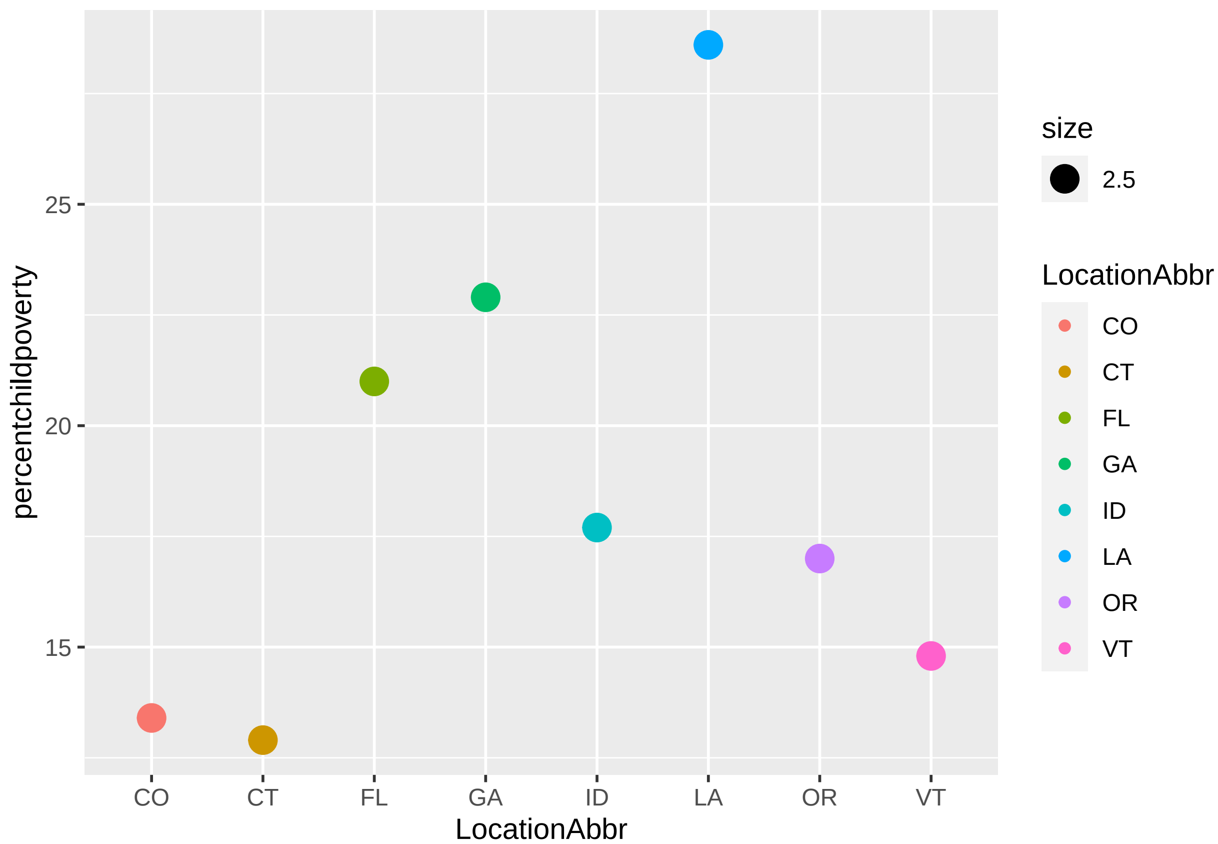
From this visualization we see variation in percent ever having caries (treated or untreated) among third graders in 2016. Idaho has the highest and Vermont the lowest.

```
ggplot(combinedsetfinal) + geom_point(aes(x= LocationAbbr, y=percentreceivingFwater, color =LocationAbb
```



Control Variables

```
ggplot(combinedsetfinal) + geom_point(aes(x= LocationAbbr, y=percentchildpoverty, color =LocationAbbr, s
```



Louisiana had the highest child (individuals under age 18) poverty in 2016 (28.6%). Oregon had the lowest level of the percent of the population being served by a community water system receiving fluoridated water.

These data visualizations show variation across the independent, dependent, and control variables.

Linear Regression

```
library(modelsummary)
```

```
#performing one predictor regression on both independent variables and plotting regression

mbf12 <- lm(Percentevercav ~ PercentBFdozenmonths, data = combinedsetfinal)
meverfed <- lm(Percentevercav ~ percenteverbreastfed, data = combinedsetfinal)

msummary(list(mbf12, meverfed),
         stars = TRUE)
```

```
## Warning: In version 0.8.0 of the `modelsummary` package, the default significance markers produced by
## This warning is displayed once per session.
```

```
library(flextable) #installed after error knitting plots
```

```
##
## Attaching package: 'flextable'
```

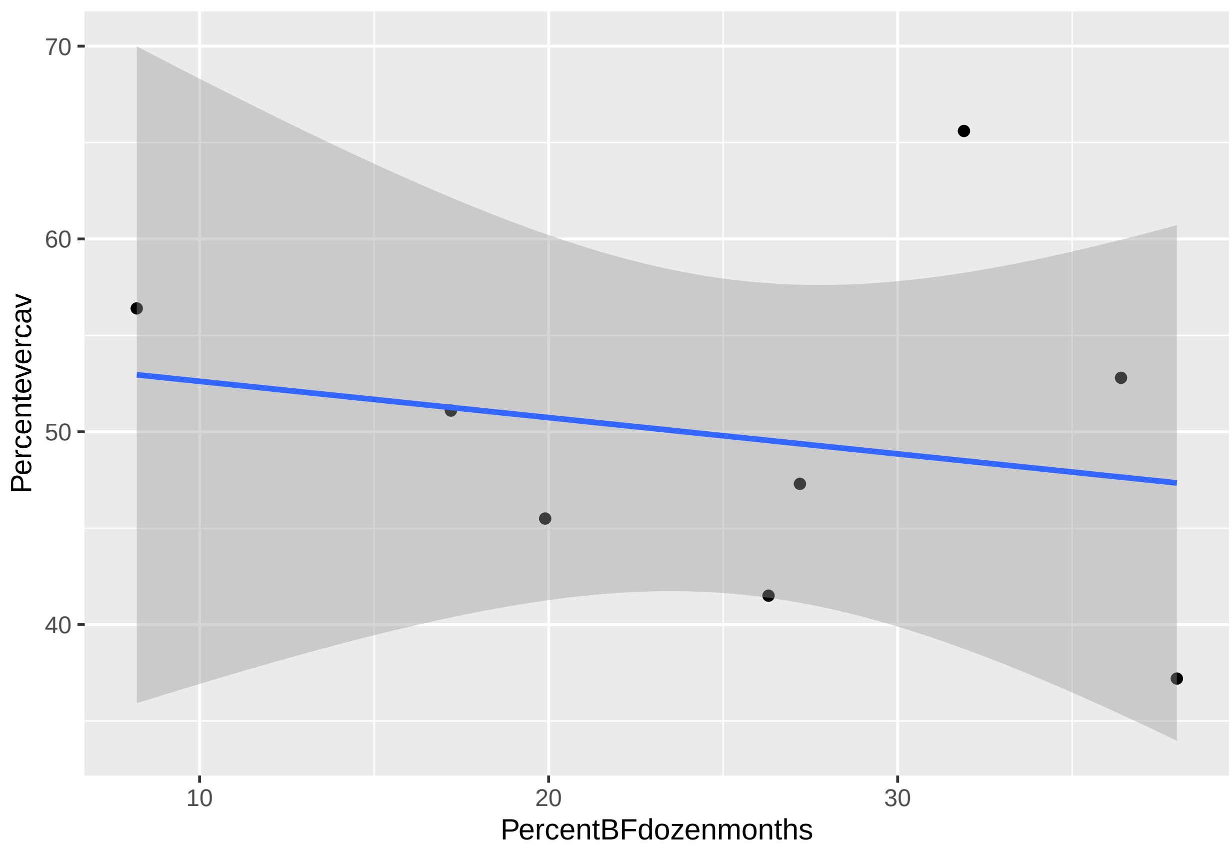
	Model 1	Model 2
(Intercept)	54.501**	62.184*
	(9.587)	(19.909)
PercentBFdozenmonths	−0.188	
	(0.351)	
percenteverbreastfed		−0.164
		(0.257)
Num.Obs.	8	8
R2	0.046	0.063
R2 Adj.	−0.113	−0.093
AIC	62.3	62.1
BIC	62.5	62.4
Log.Lik.	−28.130	−28.056
F	0.288	0.406

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

```
## The following object is masked from 'package:purrr':
##
##     compose
```

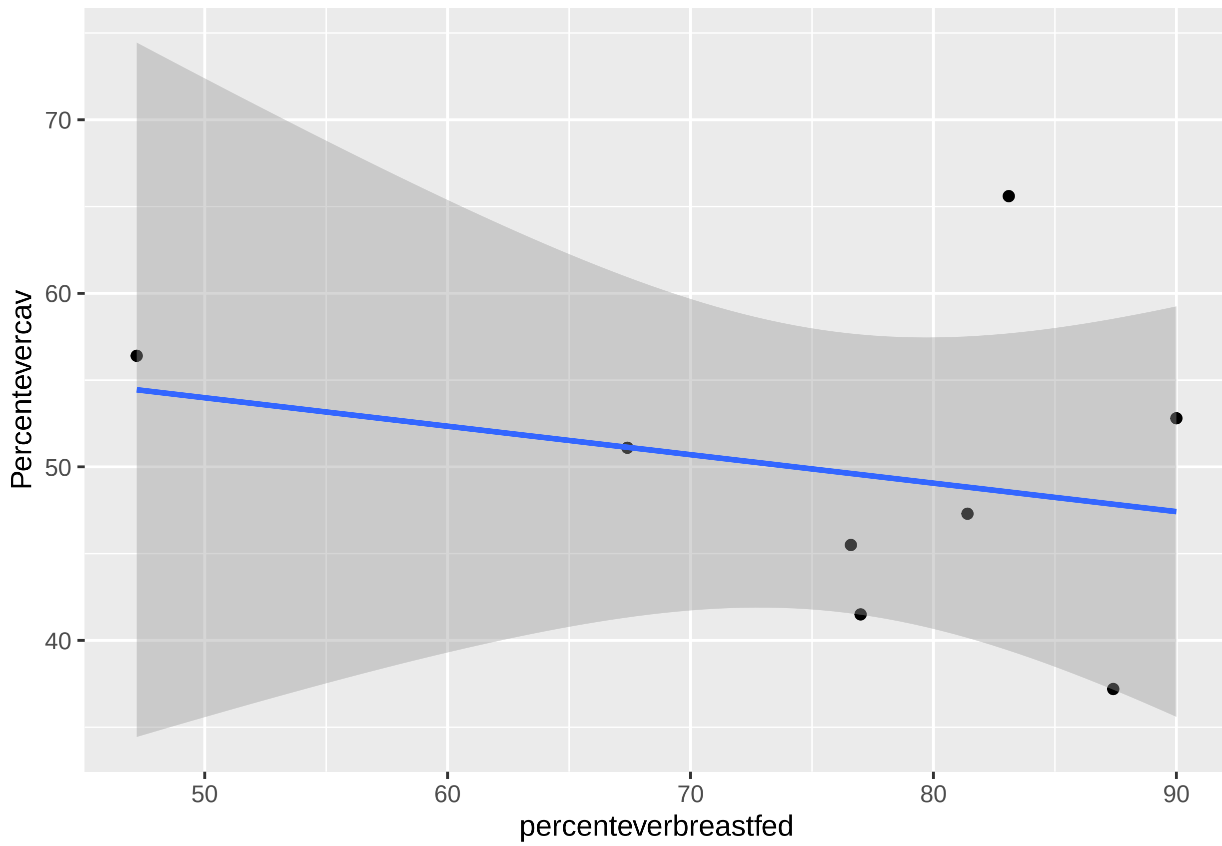
```
ggplot(combinedsetfinal, aes(PercentBFdozenmonths, Percentevercav)) + geom_point() +
                                                                                geom_smooth(meth
```

```
## `geom_smooth()` using formula 'y ~ x'
```



```
ggplot(combinedsetfinal, aes(percenteverbreastfed, Percentevercav)) + geom_point() +
                                                                                   geom_smooth(meth
```

```
## ‘geom_smooth()‘ using formula ’y ~ x’
```




```
#performing one predictor regression on both independent variables and one control variable: fluoridati

mbf12F <- lm(Percentevercav ~ PercentBFdozenmonths + percentreceivingFwater, data = combinedsetfinal)
meverfedF <- lm(Percentevercav ~ percenteverbreastfed + percentreceivingFwater, data = combinedsetfinal

msummary(list(mbf12F, meverfedF),
         stars = TRUE)
```

```
#performing one predictor regression on both independent variables and one control variable: childpover

mbf12p <- lm(Percentevercav ~ PercentBFdozenmonths + percentchildpoverty, data = combinedsetfinal)
meverfedp <- lm(Percentevercav ~ percenteverbreastfed + percentchildpoverty, data = combinedsetfinal)

msummary(list(mbf12p, meverfedp),
         stars = TRUE)
```

	Model 1	Model 2
(Intercept)	75.139**	80.637**
	(12.449)	(19.954)
PercentBFdozenmonths	−0.422	
	(0.301)	
percentreceivingFwater	−0.238+	−0.203
	(0.113)	(0.114)
percenteverbreastfed		−0.242
		(0.225)
Num.Obs.	8	8
R2	0.493	0.427
R2 Adj.	0.291	0.198
AIC	59.2	60.2
BIC	59.5	60.5
Log.Lik.	−25.598	−26.089
F	2.435	1.864

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

	Model 1	Model 2
(Intercept)	16.119	4.974
	(32.928)	(54.500)
PercentBFdozenmonths	0.367	
	(0.568)	
percentchildpoverty	1.303	1.314
	(1.073)	(1.169)
percenteverbreastfed		0.267
		(0.459)
Num.Obs.	8	8
R2	0.263	0.252
R2 Adj.	−0.032	−0.047
AIC	62.2	62.3
BIC	62.5	62.6
Log.Lik.	−27.097	−27.155
F	0.893	0.844

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

	Model 1	Model 2
(Intercept)	80.044	54.398
	(56.236)	(66.413)
PercentBFdozenmonths	−0.490	
	(0.823)	
percentreceivingFwater	−0.250	−0.173
	(0.185)	(0.144)
percentchildpoverty	−0.131	0.540
	(1.453)	(1.291)
percenteverbreastfed		−0.054
		(0.514)
Num.Obs.	8	8
R2	0.494	0.451
R2 Adj.	0.115	0.040
AIC	61.2	61.8
BIC	61.6	62.2
Log.Lik.	−25.590	−25.918
F	1.304	1.096

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

```
#performing regression on both independent variables and 2 control variable: fluoridation and child pov

mbf12b <- lm(Percentevercav ~ PercentBFdozenmonths + percentreceivingFwater + percentchildpoverty, data
meverfedb <- lm(Percentevercav ~ percenteverbreastfed + percentreceivingFwater + percentchildpoverty, da

msummary(list(mbf12b, meverfedb),
         stars = TRUE)
```

Log Transformation and Regression

The challenge with the above results is that they are based on 8 (small sample size even though it represents thousands of kids) states - some of which have outliers in child poverty, fluoridation, and caries experience - and each variable is expressed as a percentage. Linear regression is often used for continuous variables and these percentages are proportions. Recognizing this challenge, there are several pathways to provide additional information to answer "does breastfeeding cause cavities?" One option is to log transform all the variables to give them statistical properties that are more compatible with regression.

```
mbf12log <- lm(log(Percentevercav) ~ log(PercentBFdozenmonths) + log(percentreceivingFwater) + log(perc
meverfedlog <- lm(log(Percentevercav) ~ log(percenteverbreastfed) + log(percentreceivingFwater) + log(pe

msummary(list(mbf12log, meverfedlog),
         stars = TRUE)
```

Second Linear Regression

Another option is to do a similar linear regression with a different dataset with more observations. I created a second dataset using the number of births in 2008 with data from Infoplease; breastfeeding data from CDC[2]; and percent of kids (ages 1-17) with oral health problems data from 2020 from KFF. Using the percentage

	Model 1	Model 2
(Intercept)	4.362	4.277
	(2.472)	(3.535)
log(PercentBFdozenmonths)	−0.070	
	(0.259)	
log(percentreceivingFwater)	−0.193	−0.179
	(0.148)	(0.134)
log(percentchildpoverty)	0.182	0.240
	(0.462)	(0.393)
log(percenteverbreastfed)		−0.083
		(0.529)
Num.Obs.	8	8
R2	0.505	0.499
R2 Adj.	0.134	0.124
AIC	−1.6	−1.5
BIC	−1.2	−1.1
Log.Lik.	5.793	5.746
F	1.362	1.330

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

data I estimated the number of kids born in 2008 that were ever breastfed, breastfed at six months, breastfed at 12 months, and developed dental problems. This dataset has many assumptions.

```
 #importing second constructed dataset

OneStateData <- read_excel("OneStateData.xlsx",
    col_types = c("text", "numeric", "numeric",
        "numeric", "numeric", "numeric",
        "numeric", "numeric", "numeric",
        "numeric", "numeric", "numeric"))
```

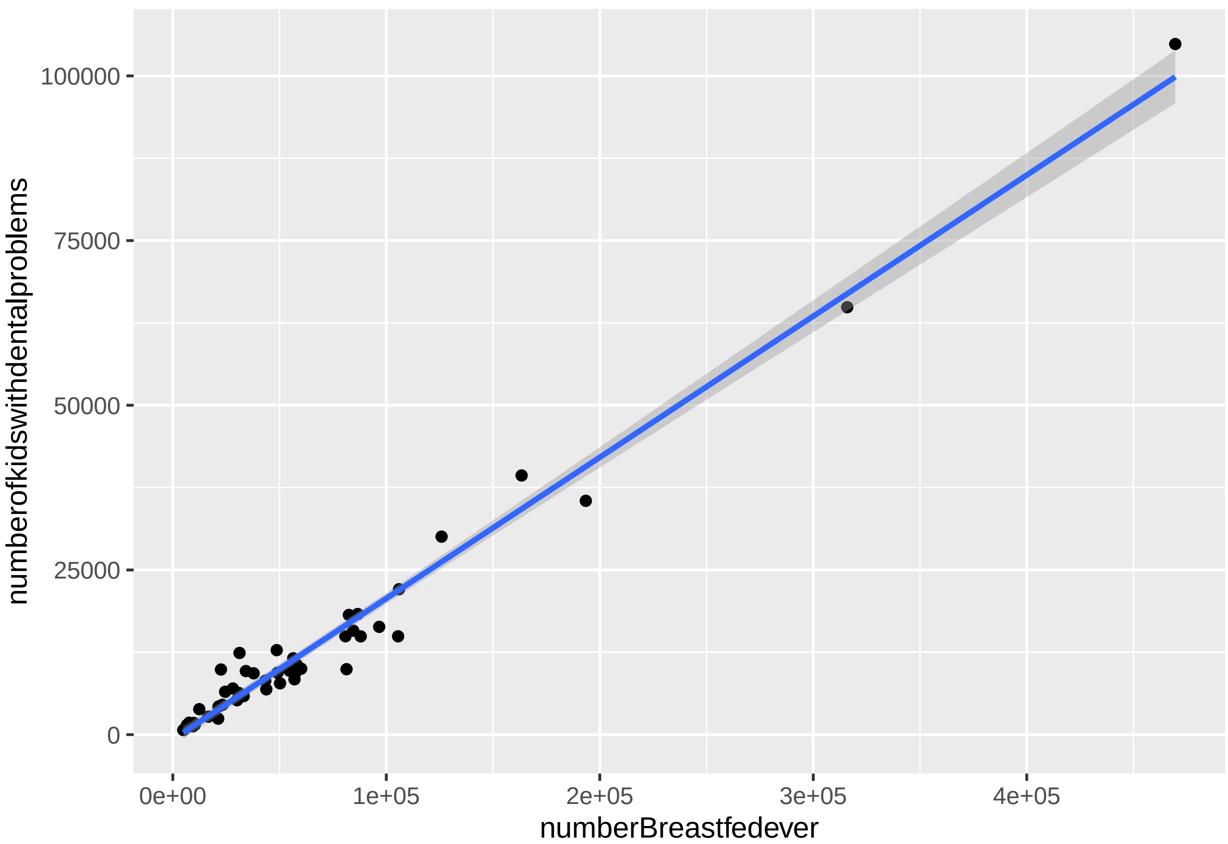
```
bfever2 <- lm(numberofkidswithdentalproblems ~ numberBreastfedever + numberkidsinpoverty, data = OneSta

bf6months2 <- lm(numberofkidswithdentalproblems ~ numberbf6months + numberkidsinpoverty, data = OneState

bf12months2 <- lm(numberofkidswithdentalproblems ~ numberbf12months + numberkidsinpoverty, data = OneSta

msummary(list(bfever2, bf6months2, bf12months2),
         stars = TRUE)
```

```
ggplot(OneStateData, aes(numberBreastfedever, numberofkidswithdentalproblems)) + geom_point() +
                                                                                   geom_smooth(meth
```

```
## `geom_smooth()` using formula 'y ~ x'
```

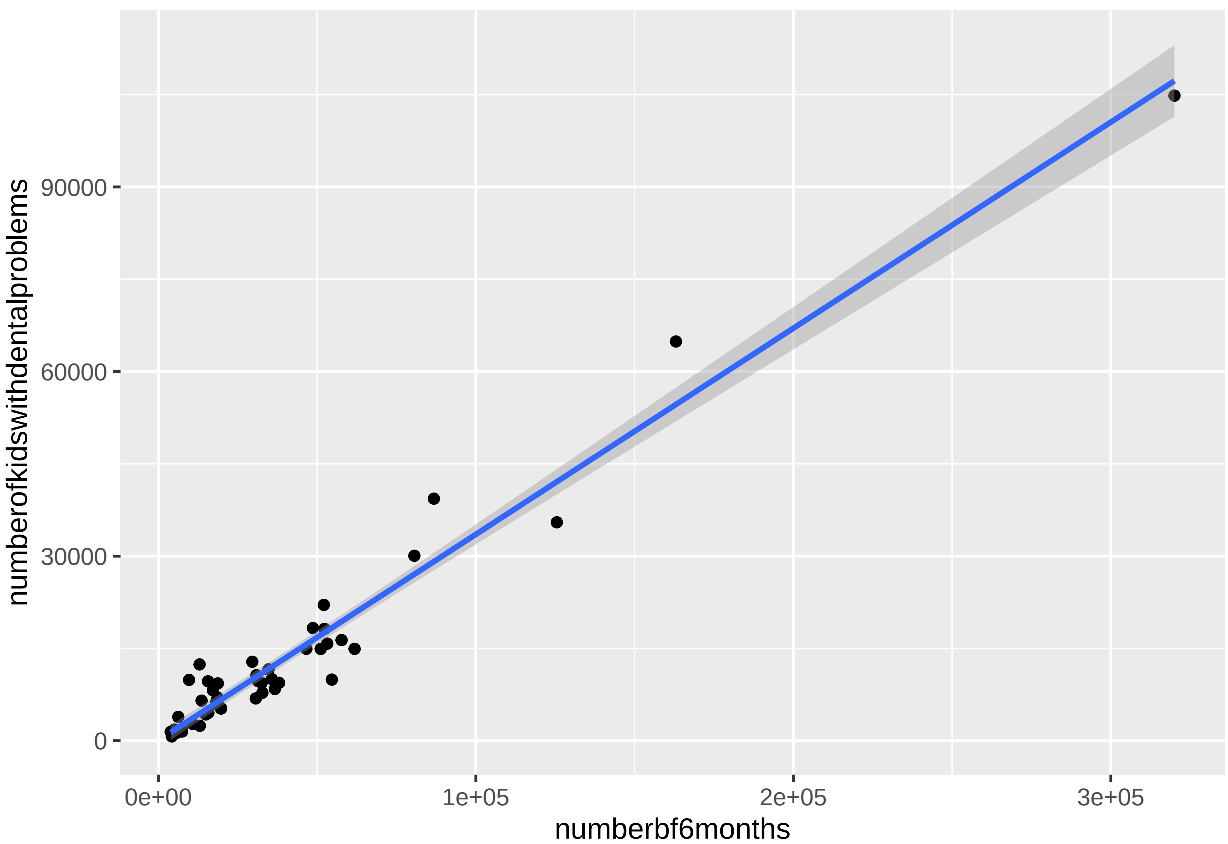
	Model 1	Model 2	Model 3
(Intercept)	−899.469+	−735.875+	−281.314
	(454.240)	(417.237)	(410.031)
numberBreastfedever	0.158***		
	(0.020)		
numberkidsinpoverty	0.224**	0.404***	0.455***
	(0.078)	(0.048)	(0.041)
numberbf6months		0.179***	
		(0.020)	
numberbf12months			0.273***
			(0.029)
Num.Obs.	51	51	51
R2	0.980	0.983	0.984
R2 Adj.	0.979	0.982	0.983
AIC	949.6	940.9	938.1
BIC	957.3	948.6	945.9
Log.Lik.	−470.795	−466.458	−465.072
F	1147.466	1364.630	1442.207

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$



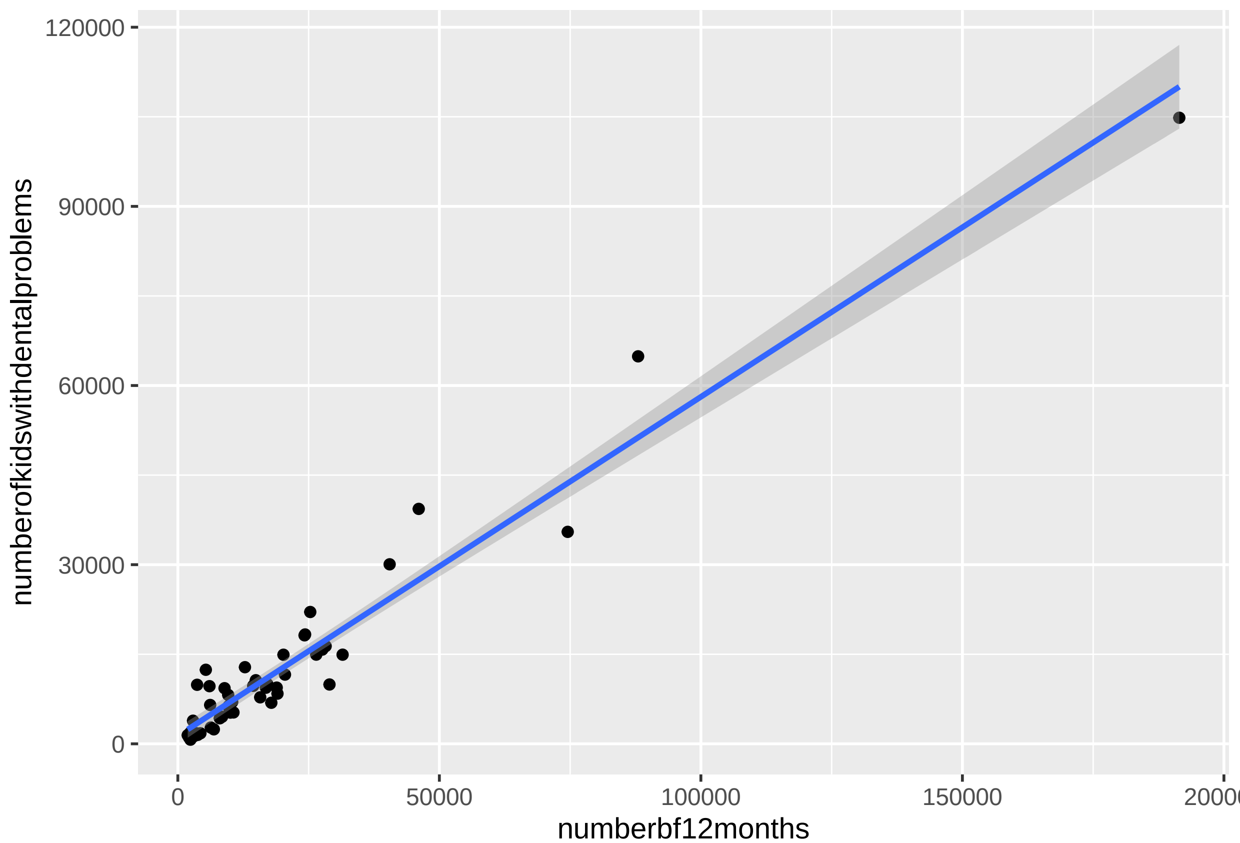

```
ggplot(OneStateData, aes(numberbf6months, numberofkidswithdentalproblems)) + geom_point() +
                                                                           geom_smooth(meth
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
ggplot(OneStateData, aes(numberbf12months, numberofkidswithdentalproblems)) + geom_point() +
                                                                         geom_smooth(meth
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Results

Without log transforming variables, regressions of percent ever with caries on percent breastfed at 12 months and on percent ever breastfed showed that a one unit increase in breastfeeding at 12 months and ever is associated with a decrease in percent ever having a caries experience by 0.49 and 0.05 percentage points, respectively when controlling for fluoridation and child poverty. These results were not statistically significant.

Based on the regressions with log-transformed variables, controlling for fluoridation and child poverty, a one percent increase in breastfed at 12 months is associated with a 0.07% decrease in ever having caries in the third grade at the state level. These results are not statistically significant. A one percent increase in ever breastfed results in a 0.08% decrease in ever having caries in the third grade, controlling for fluoridation and child poverty, at the state level. These result are also not statistically significant.

The regression done with estimated figures for number of children ever breastfed, breastfed at six months, breastfed at 12 months, ever having any dental problems, and child poverty had statistically significant results showing that breastfeeding is associated with dental problems (defined by KFF as having toothpain in the past 12 months, bleeding gums, or decayed teeth/cavities) with the highest coefficient being at 12 months; a one unit increase in number of kids being breastfed at 12 months resulted in a 0.27 unit increase in kids with dental problems, when controlling for child poverty, and is significant at 99% level.

Discussion

Limitations

The first set of linear regressions based on data from eight states has a small sample size, but its strengths are that the data is from reputable organizations and relies on fewer assumptions then the second regression. The 8 states are also varied and could be thought to be representative of 50 states. These regressions also account for fluoridation, an important causal factor for the development of caries.

The second regression is based on data I constructed by assuming I could multiply percentage data on poverty, dental problems, and breastfeeding from various sources and a few different years to the number of births for 50 states and the District of Columbia in 2008 to estimate how many third graders (8 year olds) had caries. The fault with this assumption is that number of births does not correspond to number of children still alive; some metrics for breastfeeding in 2008 mean by definition the child was born in 2007; the dental health data used in this assumption was for all kids age 1-17; and the poverty data was from 2016 (and not year of birth or year of being 8). All that said, these sets of regressions, which controlled for child poverty, had statistically significant results.

Both sets of regressions did not account for demographics of kids (age and race/racism experience of parents), dental workforce availability, or children's diet.

These regressions also approach breastfeeding as a behavior. Perhaps breastfeeding behavior is associated with other parental behaviors that affect dental health like diet and age at first dentist visit. The regressions do not incorporate the biological pathways between breast milk (its acidity, sugar content, etc.), formula, and early baby food on caries development. A regression that accounts for these other factors would provide stronger evidence of the relationship between breastfeeding as behavior and caries development since a randomized trial is likely unethical. More research is needed to distinguish the behavioral and biological relationship between breastfeeding and caries.

References

Centers for Disease Control and Prevention[1]. Community Water Fluoridation: 2016 Flouridation Statistics [online]. [accessed April 07, 2022]. URL: https://www.cdc.gov/fluoridation/statistics/2016stats.htm

Centers for Disease Control and Prevention[2]. National Center for Chronic Disease Prevention and Health Promotion, Division of Nutrition, Physical Activity, and Obesity. Data, Trend and Maps Category: Breastfeeding [online]. [accessed Apr 03, 2022]. URL: https://www.cdc.gov/nccdphp/dnpao/data-trends-maps/index.html.

Centers for Diseease Control and Prevention National Oral Health Surveillance System (NOHSS)[3]. NOHSS Child Indicators Oral Health. [access Apr 03, 2022]. URL: https://chronicdata.cdc.gov/Oral-Health/NOHSS-Child-Indicators/qcai-zfj9

Centers for Medicare & Medicaid Services.(2015). Reducing Early Childhood Tooth Decay: Approaches in Medicaid. [online PowerPoint]. [accessed April 08, 2022]. URL: https://www.medicaid.gov/sites/default/files/2019-12/learninglabslides12.pdf

Infoplease. Births, Birth Rates, and Fertility Rates 2008. [accessed April 09, 2022]. URL: https://www.infoplease.com/us/population/births-birth-rates-and-fertility-rates-2008

Kaplan, D. L., & Graff, K. M. (2008). Marketing breastfeeding—reversing corporate influence on infant feeding practices. Journal of Urban Health, 85(4), 486-504.

KFF. (2020). Percent of Children (ages 1-17) with Oral Health Problems. [accessed April 09, 2022]. URL: https://www.kff.org/other/state-indicator/children-with-oral-health-problems/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D

Kirksey, K. (2021). A social history of racial disparities in breastfeeding in the United States. Social Science & Medicine, 289, 114365.

National Women's Law Center. Poverty Rates State by State, 2016. [accessed April 08, 2022]. URL: https://nwlc.org/wp-content/uploads/2017/09/Poverty-Rates-State-by-State-2016.pdf

Rippeyoung, P. L., & Noonan, M. C. (2012). Is breastfeeding truly cost free? Income consequences of breastfeeding for women. American Sociological Review, 77(2), 244-267.

Peres, K. G., Chaffee, B. W., Feldens, C. A., Flores-Mir, C., Moynihan, P., & Rugg-Gunn, A. (2018). Breastfeeding and Oral Health: Evidence and Methodological Challenges. Journal of dental research, 97(3), 251–258. https://doi.org/10.1177/0022034517738925

U.S. Department of Health and Human Services (HHS). Medicaid & CHIP: The Children's Health Insurance Program (CHIP). [accessed April 08, 2022]. URL: https://www.healthcare.gov/medicaid-chip/childrens-health-insurance-program/