

MIDWIFE-ATTENDED BIRTHS FROM 2007-2018 IN THE UNITED STATES

Predicting Births, Birth Weights, and Exploring Relationships
between Features



GA Data Science

Cheyanne Baird

May 2020

PROBLEM STATEMENT

- Explore how midwife-attended births result in improved maternal and infant health:
 - Integration of midwives into regional health systems is a key determinant of optimal maternal-newborn outcomes [1].
 - Midwife attendance at births could significantly boost maternal and infant health in the United States [2].
 - US births have decreased since 2007; midwife-attended births show a steady rise. [3]
- More data is needed for this task:
 - Non-midwife-attended births
 - Infant and maternal mortality rates and birth complications
 - Financial and insurance barriers



Sources:

- [1] <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0192523>
- [2] <https://www.propublica.org/article/midwives-study-maternal-neonatal-care>
- [3] <https://www.nbcnews.com/health/health-news/u-s-births-fall-virus-could-drive-them-down-more-n1210956>

METRICS AND ASSUMPTIONS

- Adjusting the original proposal, data was analyzed from the CDC around midwife-attended births in the United States ranging from 2007-2018 to predict midwife-attended births and birth weights, separating train and test sets by year.
- Target variables used for prediction: number of midwife-attended births and birth weights in the United States by state per year.
- Varied feature set: prenatal care, location of birth, maternal age, education of mother, type of birth, and plurality (single or twins).

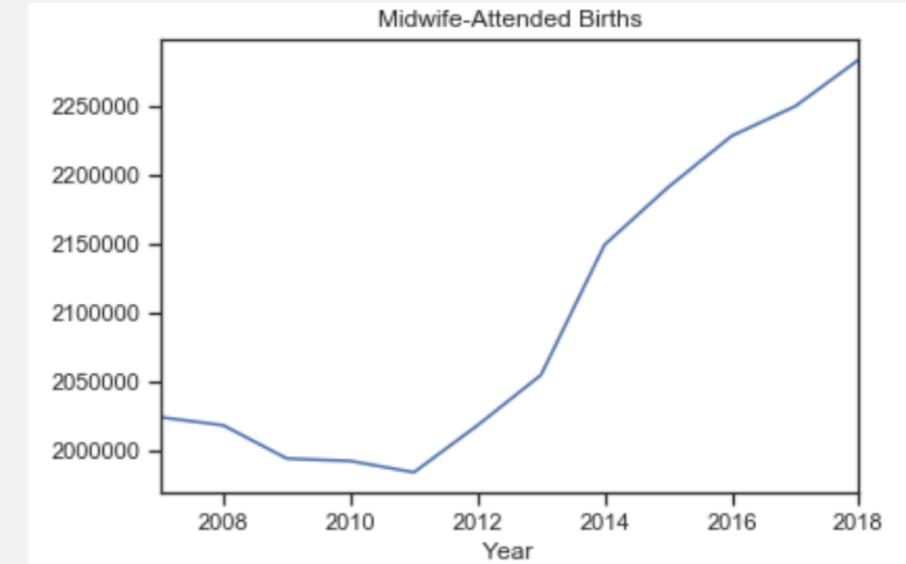
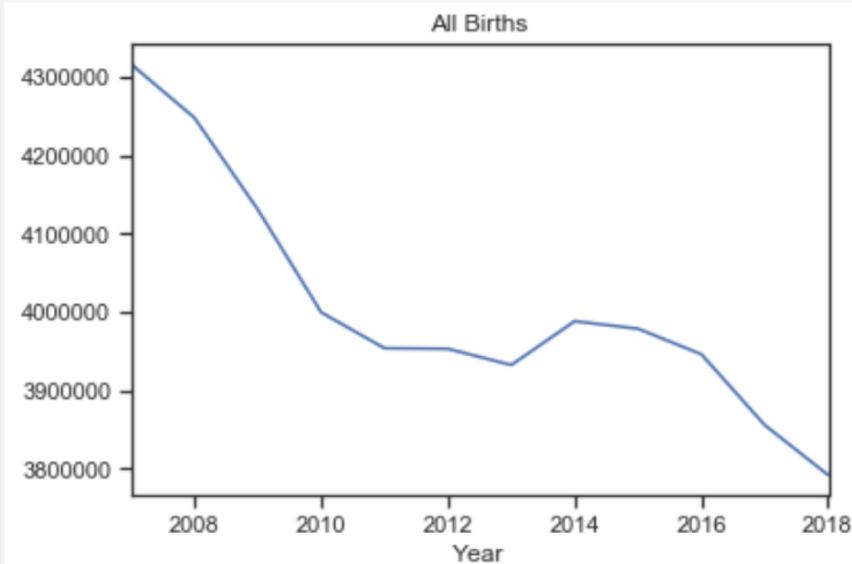


Data Sources:

- https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm#Births
- <https://wonder.cdc.gov/controller/datarequest/D66;jsessionid=8385053D65E6CDC2037A9DE69F63F16B>

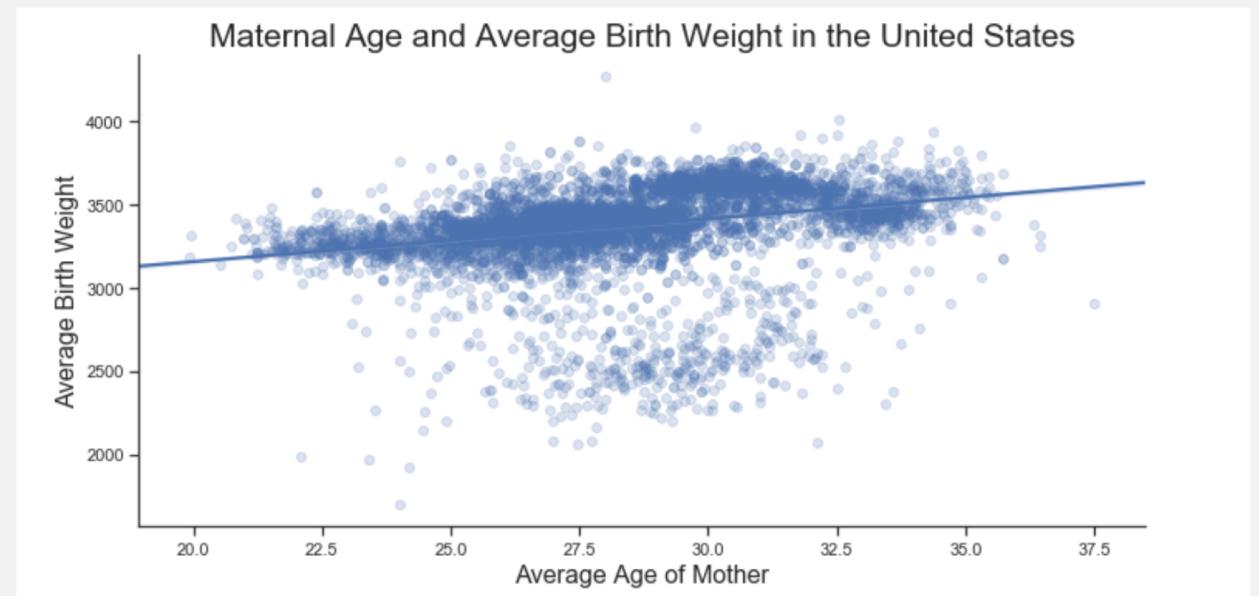
APPROACH AND PROCESS

- Performed exploratory data analysis (EDA), basing models on insights.
- National trends over time:
 - All births decreased after 2007, with a spike in 2014. Steady rise in midwife-attended births with a noticeable dip in 2011.



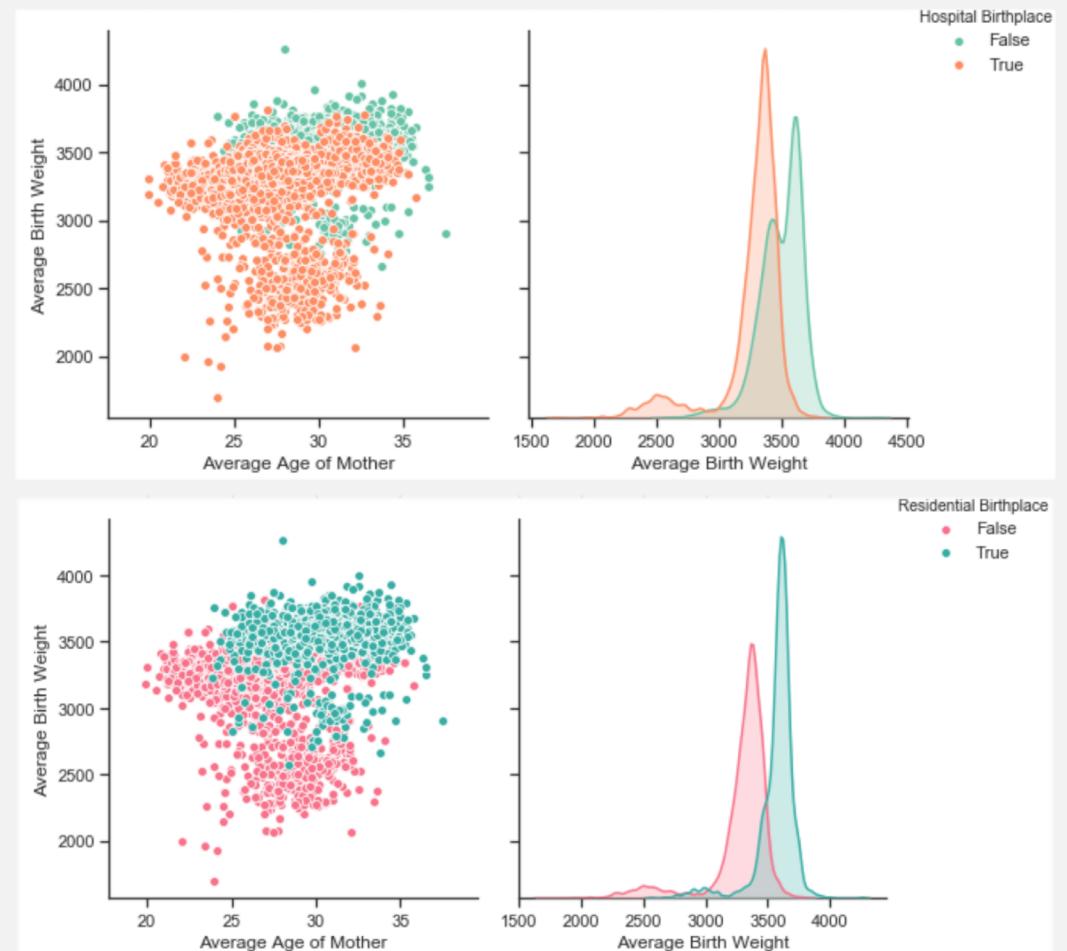
APPROACH AND PROCESS

- Maternal age and birth weight insights:
 - Average maternal age: 28 yrs
 - Average birth weight: 7.4 lbs
 - Average Birth Weight and Average Age of Mother: there are two populations of birth weight and a correlation between the two features.
 - Higher weight babies appear to be born to older mothers.
- Residential vs. hospital birthplace and birth weight:
 - Residential and hospital birthplaces are both correlated with slightly higher birth weights, with residential slightly higher.



APPROACH AND PROCESS

- Residential vs. hospital birthplace and maternal age:
 - Average age of mother is older for residential birthplace.
 - For hospital birthplace, mean is lower than residential for maternal age.
 - For average birth weight, most correlated features are age and residential birthplace.
- Prenatal care vs. lack of prenatal care:
 - Early prenatal care has increased over time.
 - Higher birth weight is correlated with earlier prenatal care as shown in joint plots.
 - Not much of a correlation between birth weight and lack of prenatal care.
- Delivery type:
 - Cesarean deliveries have decreased over time.



MODEL BUILDING

- After performing EDA, a list of questions to explore was compiled along with which models would be best to use in exploring these questions, deciding whether the problem was a regression or classification problem.
- A variety of models were built:
 - Linear Regression, including r-squared test and train accuracy scores with RSME and Ridge Regression.
 - Logistic Regression, including r-squared test accuracy scores with F1 score and null model score analysis.
 - Random Forest Regression, including r-squared test and train accuracy scores, GridSearchCV, and feature importances.
 - Random Forest Classification, including r-squared test and train accuracy scores with null model score analysis.



PERFORMANCE EVALUATION

- Models are not perfect, and some additional model tuning is needed. Overfitting is a issue with several models.

Explore correlation between maternal age and weight of baby.

Best Model	Train	Test	Features	Target	Region	Notes
Random Forest Regression Model 1	99%	92%	All	Birth Weight	CA	Overfitting



PERFORMANCE EVALUATION

Predict number of births for 2017 and 2018 based on past years for specific states (California or New York). GridSearchCV on Random Forest showed that birth weight, maternal age, and vaginal delivery were the most important features, with an optimal max_depth of 3.

LinReg Models	Train	Test	Features	Target	Region	Notes
Linear Regression Model 3	76%	75%	All	Births	CA	RMSE on test low in comparison to max births; ridge reg shows errors in my model = 74% smaller than the null model
Linear Regression Model 5	75%	75%	All (- age)	Births	NY	RMSE on test low in comparison to max births; ridge reg shows errors in my model = 75% smaller than the null model

Best Model	Train	Test	Features	Target	Region	Notes
Random Forest Regression Model 4	99%	90%	All	Births	NY	Overfitting

PERFORMANCE EVALUATION

Residential and Hospital birthplaces correlated with slightly higher birth weights, with residential being slightly higher. Hospital birthplace mean lower than residential for maternal age.

Best Models	Train	Test	F1	Features	Target	Region	Notes
Random Forest Classification Model 1	100%	94%		Birth Weight	Residential Birthplace	NY	Overfitting; score > null model
Logistic Regression Model 9		90%	87%	Birth Weight	Residential Birthplace	CA	Overfitting; score > null model
Logistic Regression Model 5		87%	94%	Birth Weight	Hospital Birthplace	CA	Overfitting; score > null model

Relationship between delivery type, prenatal care, and maternal age.

Best Models	Train	Test	F1	Features	Target	Region	Notes
Logistic Regression Model 8		81%	50%	1stMoPrenatal, Age	Vaginal Delivery	NY	F1 score poor; score > null model
Random Forest Classification Model 4		81%		1stMoPrenatal, Age	Vaginal Delivery	NY	score > null model
Logistic Regression Model 3		80%	44%	1stMoPrenatal, Age	Vaginal Delivery	CA	F1 score poor; score < null model
Random Forest Classification Model 5	81%	80%		1stMoPrenatal, Age	Vaginal Delivery	CA	score = null model

IMPACT

- In concert with additional data and analysis, these findings could be used to explore the need to increase midwife-attended births and early prenatal care in the United States and how this improves infant and maternal health.
- This could be extended to analyze the impact of doulas and lactation consultants.



FUTURE WORK

- Model tuning, especially to prevent overfitting.
 - Random forest: increase amount of n-estimators (tried, did not work), reduce feature set, feature subsetting, OOB error, adjust max_depth.
 - Logistic regression: regularization, additional data, and additional cross-validation.
 - Linear regression: additional data and features.
- Steady decrease in overall births since 2007 with a spike in 2014, steady rise in midwife-attended births with noticeable dip in 2011. Decreases due to recession?
- Analyze prenatal care in relation to birth weight and relationship between maternal age and birthplace. Analysis showed two distinct populations and a correlation between birth weight and maternal age.
- Obtain data from non-midwife-attended births, infant/maternal mortality rates, birth complications, financial/insurance barriers, 1st/2nd births per mother, newly released 2019 CDC data, 2020 CDC data to analyze how Coronavirus and restricted midwife/doula support in hospital impacted births. Increase in home births? Decrease in overall births?
- Look at average age and weight by state or region.