Motivation

One of my good friends and his girlfriend recently brought a new life into this world, and hearing him talk about his child throughout the pregnancy motivated me to explore this dataset about child births in the U.S. in 2016. He would constantly encourage the mother to eat healthy and never miss any checkups with the doctor. Never would I wish anything bad on his newborn, but I started to wonder if their constant prenatal visits and healthy lifestyle choices really made an impact on the health of the baby.

Research Questions:

1. What factors matter most when it comes to Prenatal Care?

2. How does childbirths vary across demographics?

3. How does Tobacco use affect childbirth?

4. How does other risk factors affect childbirth?

Data Source

I obtained my dataset from the Centers for Disease Control and Prevention (CDC) via the National Center for Health Statistics (NCHS) module. The dataset was a 133 MB zip file, but unzipped, the data was a 2.3 GB csv file. There were over 3 million rows and 240 columns in the dataset. ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/DVS/natality/Nat2016us.zip

	mager	mbrace	dmar	meduc	priorlive	priorterm	previs	cig_0	cig_1	cig_2	cig_3	bmi_r	rf_pdiab	rf_phype	ip_gon	ip_syph	ip_chlam	dmeth_rec	dbwt
0	35	1	1	6	3	1	3	0	0	0	0	4	N	N	N	N	N	1	3705
1	32	4	1	7	0	0	10	0	0	0	0	2	N	N	N	N	N	1	3204
2	18	3	2	2	0	1	99	0	0	0	0	2	N	N	N	N	N	1	4499
3	34	1	2	4	0	0	15	99	99	99	99	4	N	N	N	N	N	1	3660
4	31	4	1	6	2	0	9	0	0	0	0	2	N	N	N	N	N	1	2866

The columns that I loaded into a data frame were:

Age	Race	Marriage	Education	Children	Terminations	Prenatal Visits
Cigarettes before Pregnancy	Cigarettes during 1st Trimester	Cigarettes during 2nd Trimester	Cigarettes during 3rd Trimester	BMI	Diabetic	Hypertension
Gonorrhea	Syphilis	Chlamydia	Delivery	Baby Weight		

As the chart above shows, most of the data was mapped to numbers and read in as int64. I decided to remap the data with their corresponding values as well as optimize the data by converting the dtype to either int8 for the numerical data or category for the obkect data.

Research Questions and Methods

I struggled to load the data into a Pandas DataFrame, so instead I used Dask to load the data and filter out the columns that I did not intend to utilize. From here, I could convert the data from a Dask DataFrame to a Pandas DataFrame and perform my computations. Upon looking at the data, I noticed it needed to be cleaned a bit to make it easily readable. I removed all the rows that contained unusable values such as "Unknown" and would not add anything to my analysis. Next, I converted the Baby Weight column from grams to pounds.

1. What factors matter most when it comes to Prenatal Care?

I utilized a Pandas DataFrame with the columns of Age, Race, Marriage_Status, Education, Kids, and Prenatal_Visits to answer this question. I used statistical analysis to calculate the correlation, if any, between the numerical values using Pearson. Secondly, I performed a distributional analysis and created a simple linear regression using JointGrid to visualize the dispersion of the linear data. Bokeh came in handy to plot bar graphs of the average prenatal visits grouped by variables such as race, education status, or marriage status. The final graph that really answered my question was made by calculating the mean difference between the values in the given categories.

2. How does childbirths vary across demographics?

I created another Pandas DataFrame with the variables of Age, Terminations, race, Marriage_Status, and Education. I created a crosstab to visually plot the relationship between variables as a heatmap. I went further to group the data based on a few of the factors given and create a bar chart to display this data. The bar charts are what answered my question and showed the breakdown of demographics of the mothers of children born in 2016.

3. How does Tobacco use affect childbirth?

I created another Pandas DataFrame with the variables of Delivery, Weight, amount of cigarettes smoke daily before pregnancy, during their first, second, and third trimesters. I performed analysis on the data frame and calculated the correlation amongst all these variables. A joint plot seemed
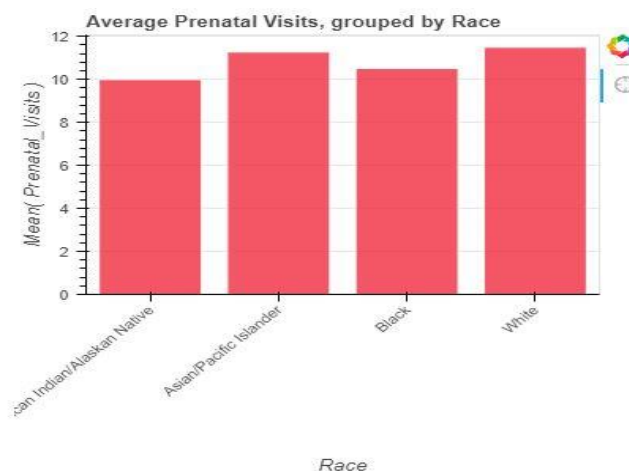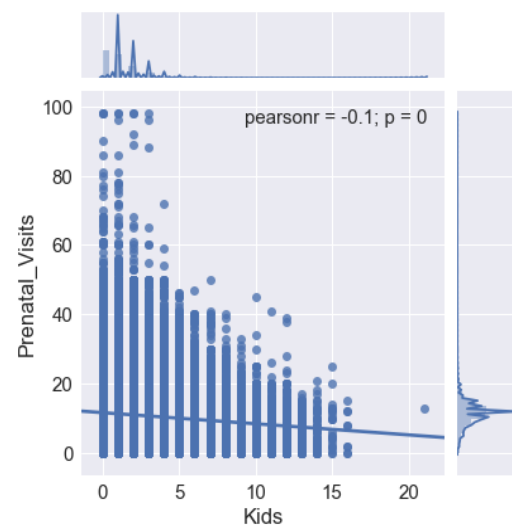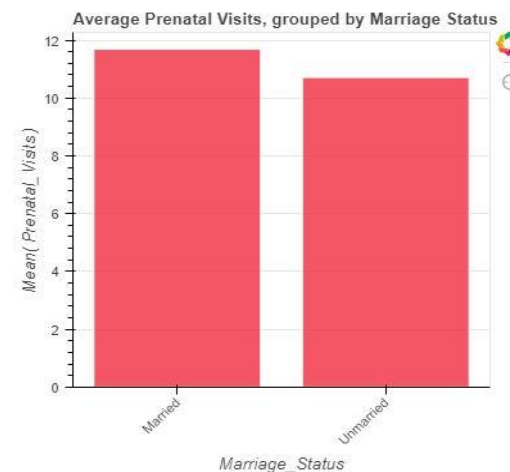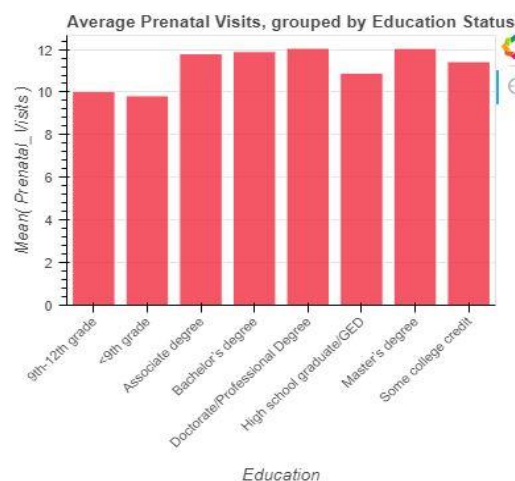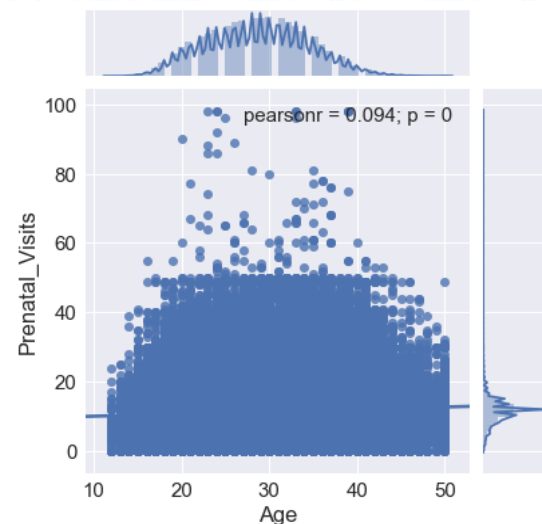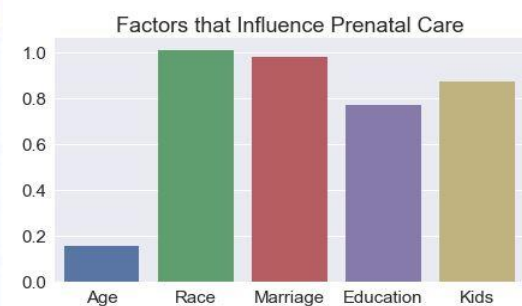
the best visual to show the simple linear relationship between number of cigarettes smoked per day and how it may or may not have affected the baby weight.

4. How does other risk factors affect childbirth?

The final question, I choose to look at the variables of BMI, Diabetes, Hypertension, Gonorrhea, Syphilis, Chlamydia, Delivery, and Baby Weight. I grouped the data based on each variable and compared it to the method of delivery and weight of the baby. The best illustrator for risk factor seemed to be BMI.
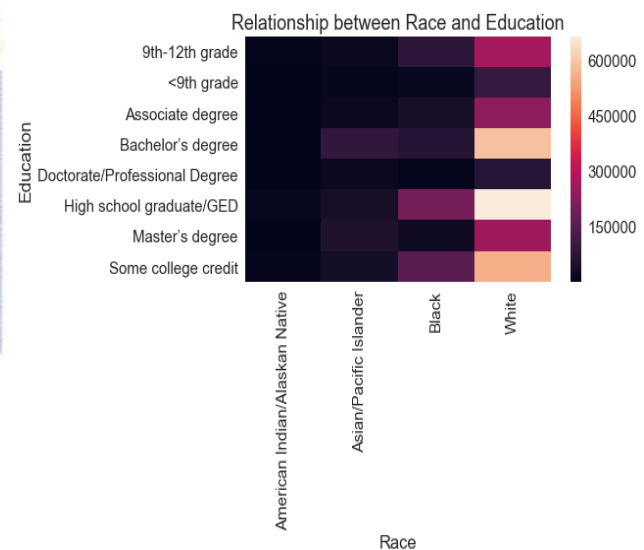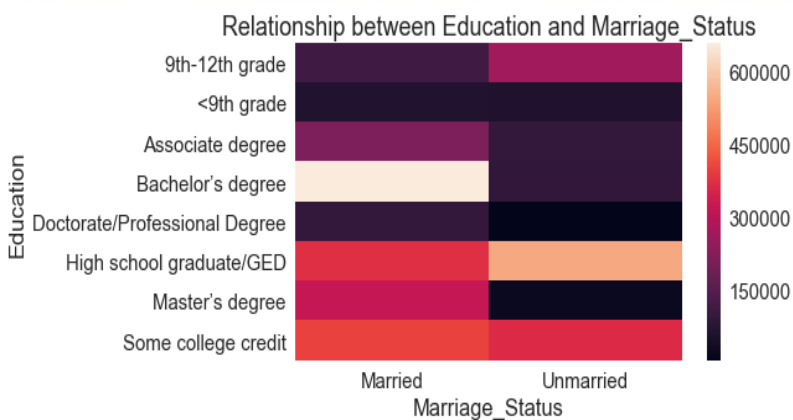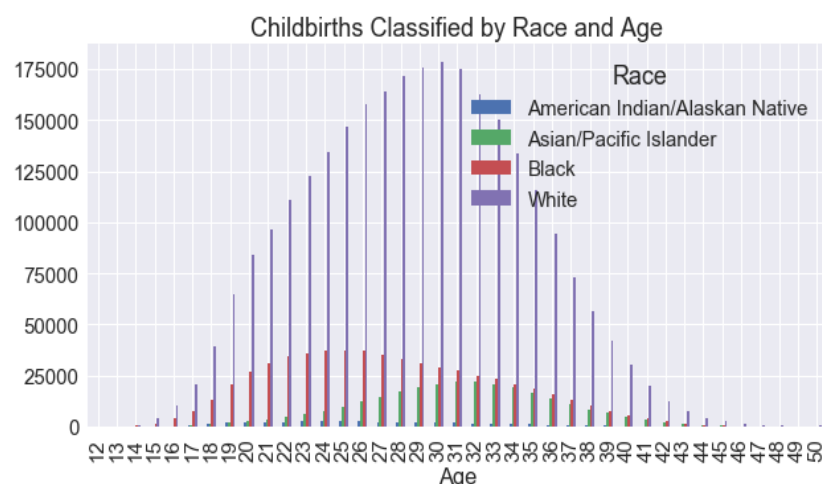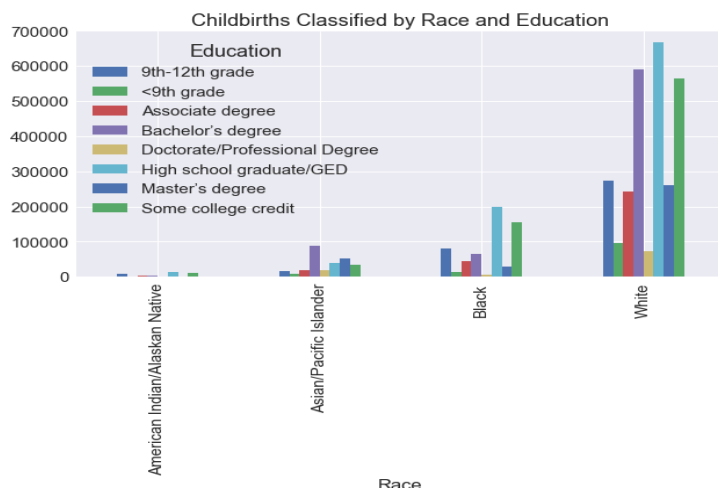
<u>Results and Analysis</u>

	Age	Race	Marriage_Status	Education	Kids	Prenatal_Visits
0	35	White	Married	Bachelor's degree	3	3
1	32	Asian/Pacific Islander	Married	Master's degree	0	10
2	31	Asian/Pacific Islander	Married	Bachelor's degree	2	9
3	33	White	Married	Some college credit	2	9
4	28	Asian/Pacific Islander	Married	High school graduate/GED	8	3

Factors that Influence Prenatal Care

Average Prenatal Visits, grouped by Education Status

Average Prenatal Visits, grouped by Marriage Status

Average Prenatal Visits, grouped by Race

These charts and graphs shows a variety of things about my data. The regression plots both show that there is not a significant correlation between kids and prenatal visits nor is there correlation between age and prenatal visits, however there is a linear relationship. The relationship between kids and prenatal visits suggests that as people have more kids, they tend to go to fewer prenatal visits. This makes sense because they may have gotten the hang of prenatal care, therefore, they might not want to or feel a need to visit a doctor as often. The bar charts show the average prenatal visits grouped by race as well as education and lastly marriage status. We can notice a large difference between marriage status as well as differences across the scales for race and education. My final graph shows the difference in means based on each group category. For each column, I calculated the difference for each row of values and then calculated the mean of those differences to get a representation of the spread of the data. It appears that the factors that have the biggest influence on prenatal visits is race and marriage status.

	Age	Terminations	Race	Marriage_Status	Education
0	35	1	White	Married	Bachelor's degree
1	32	0	Asian/Pacific Islander	Married	Master's degree
2	31	0	Asian/Pacific Islander	Married	Bachelor's degree
3	33	0	White	Married	Some college credit
4	28	0	Asian/Pacific Islander	Married	High school graduate/GED

Relationship between Race and Education

Relationship between Education and Marriage_Status

Childbirths Classified by Race and Education

Childbirths Classified by Race and Age

These sets of visuals can explain and show how the childbirths in the U.S. was broken down demographically. The heatmaps shows the breakdown of education levels based on race as well as education levels based on marriage status. The bar graph shows the childbirths group by race and education. I found it interesting how the childbirths group by education was highest among blacks and whites for a high school diploma or GED, but it was highest among Asians with a bachelor's degree. Secondly, it was highest among whites with bachelor's degree followed by some college credit, but with black it was highest secondly with some college credit followed by a 9th-12th grade education. And once again, with Asians, the second highest is among Master's degree followed by high school graduate. The last plot is a nice visual showing a normal distributing of childbirths based on race and ages with the average age being mid 20s – early 30s.

	Cig_Before	Tri_1	Tri_2	Tri_3	Delivery	Weight
0	0	0	0	0	Vaginal	8.171875
1	0	0	0	0	Vaginal	7.062500
2	0	0	0	0	Vaginal	6.320312
3	0	0	0	0	Vaginal	9.671875
4	0	0	0	0	Vaginal	1.697266
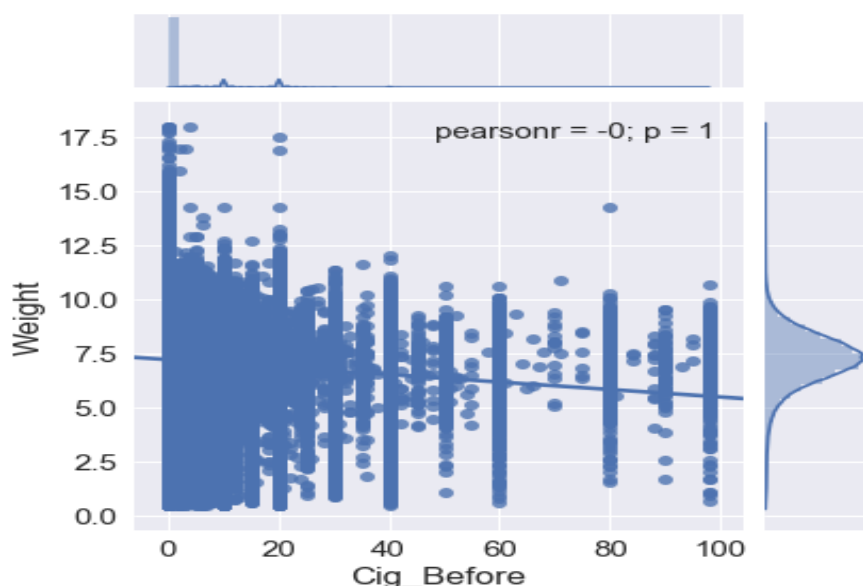
For this data, it was hard to find any solid correlation between smoking and delivery method or weight. However, I still generated a regression plot for the number of daily cigarettes smoked before pregnancy and how it affected baby weight. As we can see, the data is heavily skewed for weight, leading me to see that there were many outliers in the data. As a result of the analysis, I can't find any relation between cigarettes affecting child birth.

	BMI	Diabetes	Hypertension	Gonorrhea	Syphilis	Chlamydia	Delivery	Weight
0	Obesity I	No	No	No	No	No	Vaginal	8.171875
1	Normal	No	No	No	No	No	Vaginal	7.062500
2	Normal	No	No	No	No	No	Vaginal	6.320312
3	Overweight	No	No	No	No	No	Vaginal	9.671875
4	Overweight	No	No	No	No	No	Vaginal	1.697266
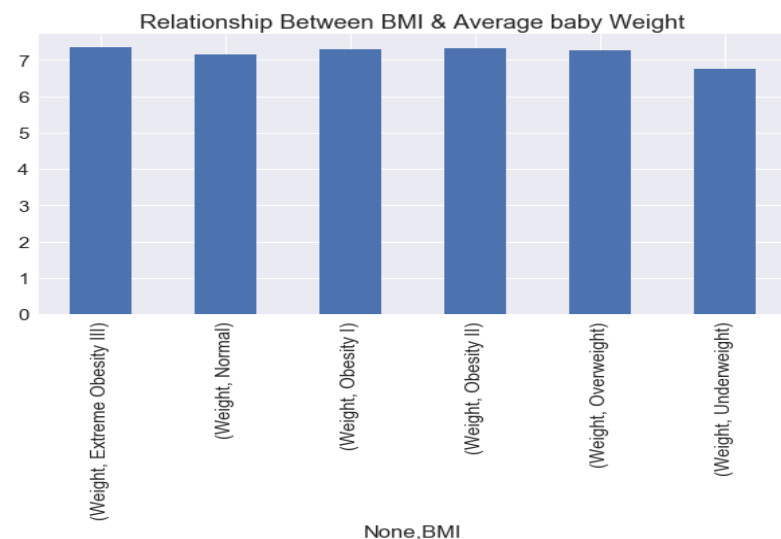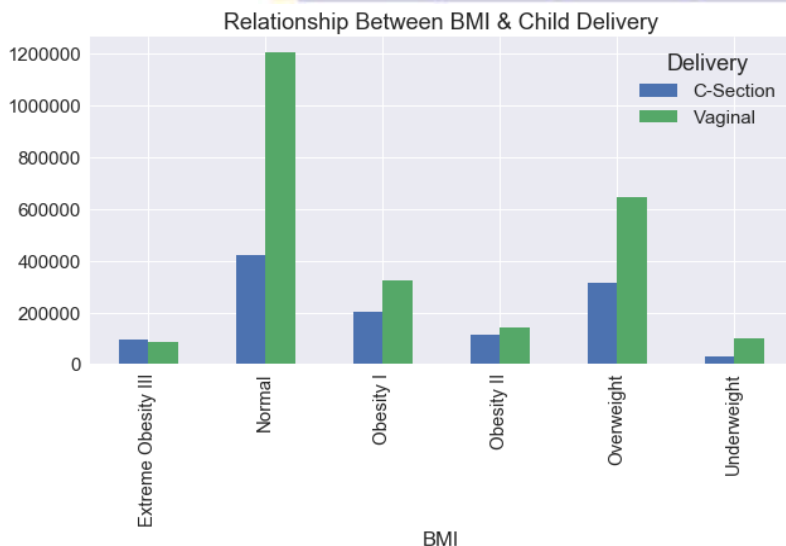
The final virilizations were to show any other risk factors that may affect childbirth and in what ways. The first chart comparing BMI to method of delivery shows a direct relationship to the data. For normal BMI, there was many vaginal delivery, however as BMI increased, the method of delivery started to converge to almost equal between vaginal and C-Section. Interestingly, as BMI reached Extreme Obesity, C-Sections were more than vaginal deliveries. The second graph shows the relationship between BMI and baby weight. There is also a relationship here, showing that as BMI increased, so did the baby weight.

Future Steps

Continuing my exploration of this data in the future, I would like to gather data from multiple years and compare trends in the data based on demographics. I would also like to include other lifestyle factors such as eating habits and other changes or factors that a child and mother may be exposed to during pregnancy.