PS1_Fan

I. Data Exploration

1. Describe how to access data, where it is stored, who curates it. Make sure to use the original source and curator in addition to the NBER site to which I have linked.

The natality data (http://nber.org/data/vital-statistics-natality-data.html) can be accessed through NBER's website, but is curated by the National Vital Statistics System of the National Center for Health Statistics (NCHS), and many vital registration systems under different juristictions that are responsible for registering those vital events, like births, deaths, marriages, divorces and fetal deaths. It is stored on the NCHS website of Centers for Disease Control and Prevention (https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm).

2. Cite other key papers that have used this data.

Williams & Collins (2001) cites the data for up to 2000 to discuss racial disparities in health.
Gelberg, Andersen, and Leake (2000) cites the data up to 1997 to calibrate for a revised model for access to health by vulnerable populations.
Krieger (1999) reviews that data for up to 1997 in studying how discrimination affects health outcomes.

3. Describe how the data were collected.

The data are a sample of birth certificates in the United States. Prior to 1972, data are based on a 50-percent sample of birth certificates from all States. Beginning in 1972, data are based on a 100-percent sample of birth certificates from some states and on a 50-percent sample from the remaining States.

4. Include a table that gives descriptive statistics for at least 8 key variables (you can do more).

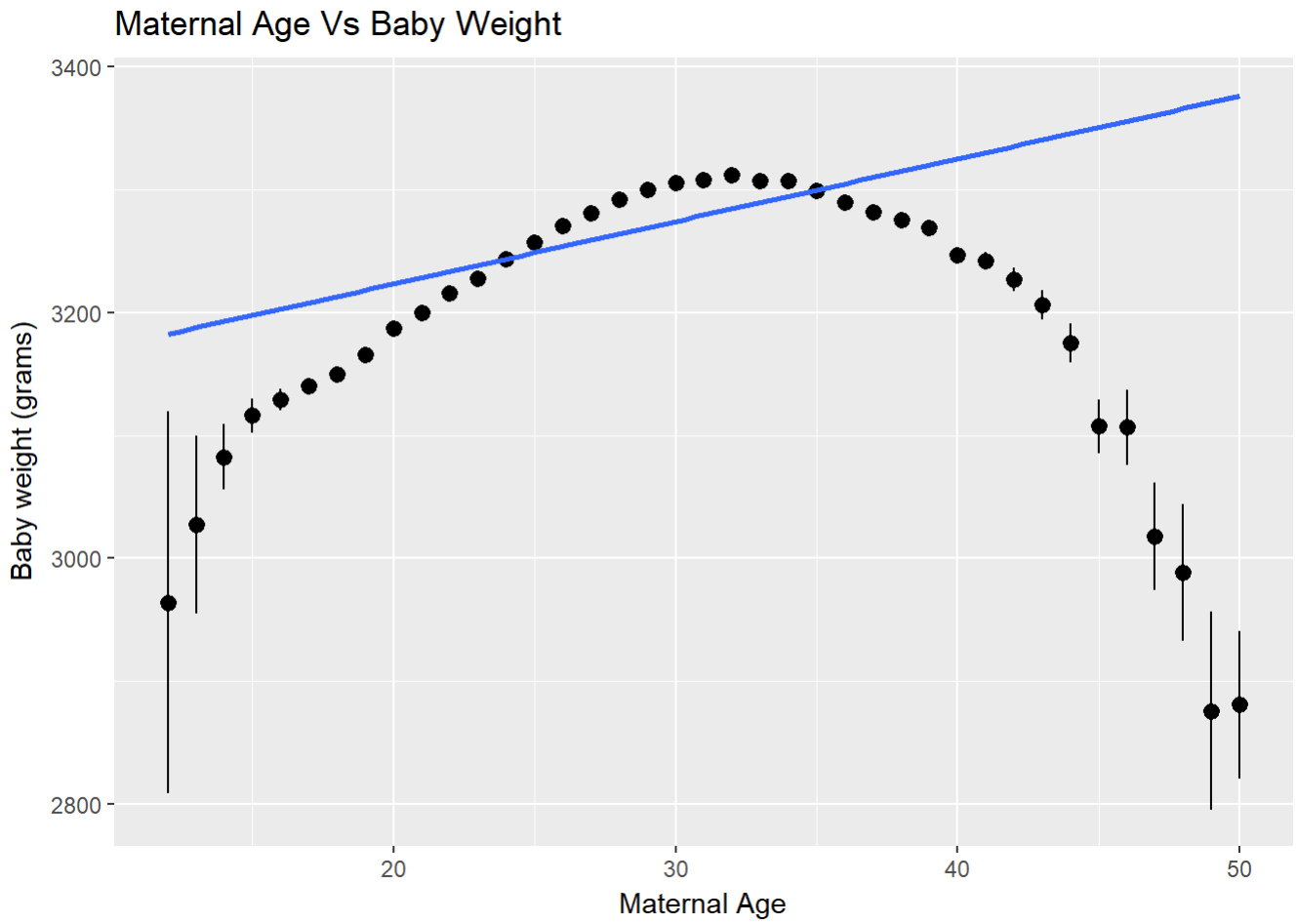
The variables that I am interested in are:
1. mager: Mother's Single Years of Age
2. pwgt_r: Pre-pregnancy Weight Recode
3. cig_0: Cigarettes Before Pregnancy
4. cig_3: Cigarettes 3rd Trimester
5. bmi: body mass index
6. dbwt: Birth Weight - Detail in Grams (Edited)
7. meduc: Mother's Education
8. attend: Attendant at Birth

```
##      mager           pwgt_r          cig_0            cig_3
##  Min.   :12.00   Min.   : 75.0   Min.   : 0.000   Min.   : 0.0000
##  1st Qu.:24.00   1st Qu.:129.0   1st Qu.: 0.000   1st Qu.: 0.0000
##  Median :29.00   Median :150.0   Median : 0.000   Median : 0.0000
##  Mean   :28.71   Mean   :177.3   Mean   : 1.704   Mean   : 0.9917
##  3rd Qu.:33.00   3rd Qu.:180.0   3rd Qu.: 0.000   3rd Qu.: 0.0000
##  Max.   :50.00   Max.   :999.0   Max.   :99.000   Max.   :99.0000
##
##       bmi             dbwt
##  Min.   :13.00   Min.   : 227
##  1st Qu.:22.10   1st Qu.:2970
##  Median :25.50   Median :3310
##  Mean   :28.82   Mean   :3267
##  3rd Qu.:30.80   3rd Qu.:3630
##  Max.   :99.90   Max.   :8165
##                  NA's   :4543
```

```
##
##               8th grade or less                    9th~12th grade
##                          132478                            407598
##                     High school             Some college credit
##                          981410                            809083
##                 Associate degree               Bachelor's degree
##                          322703                            788562
##                 Master's degree  Doctorate or Professional Degree
##                          359830                            102727
```

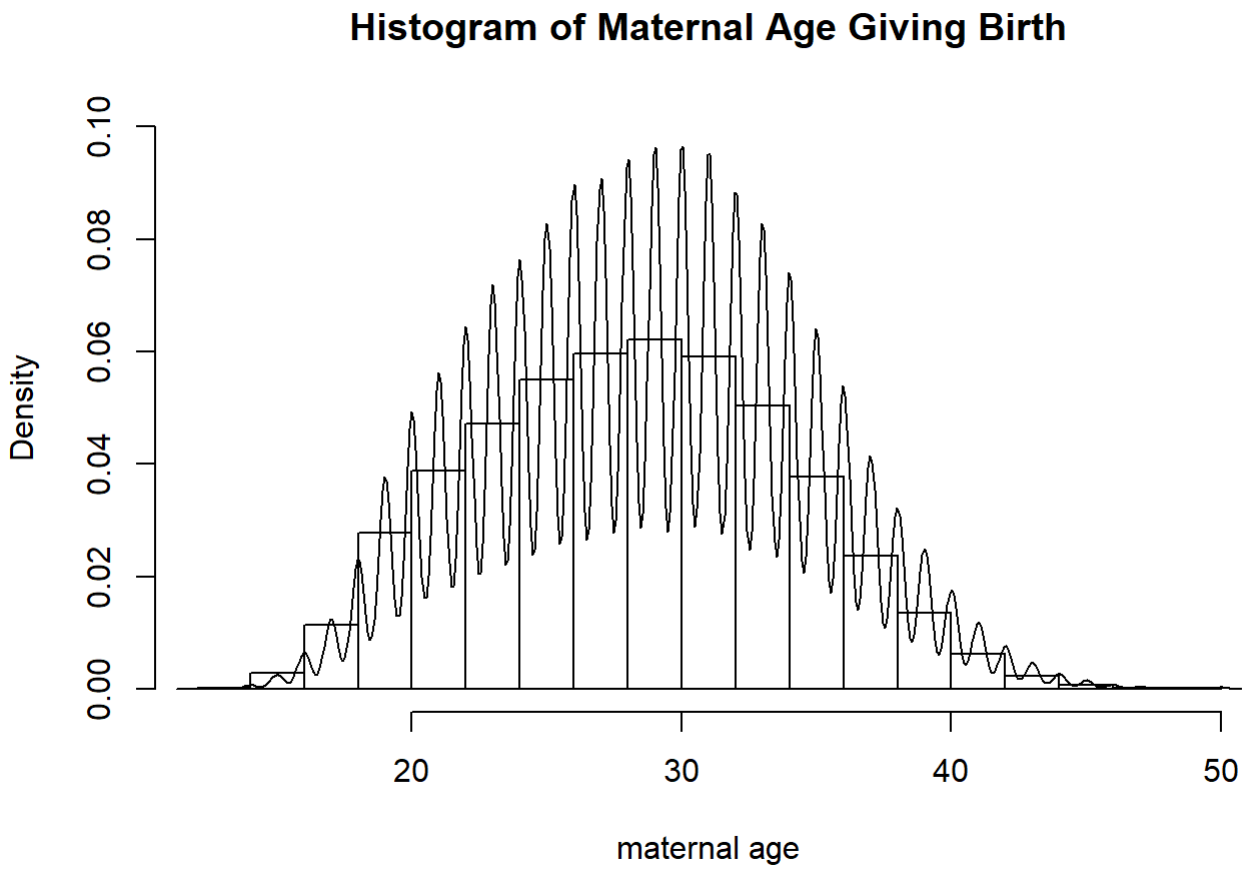
```
##
##       Doctor of Medicine (MD)    Doctor of Osteopathy (DO)
##                       3248843                       291083
## Certified Nurse Midwife (CNM)                Other Midwife
##                        348745                         31653
##                         Other
##                         33375
```

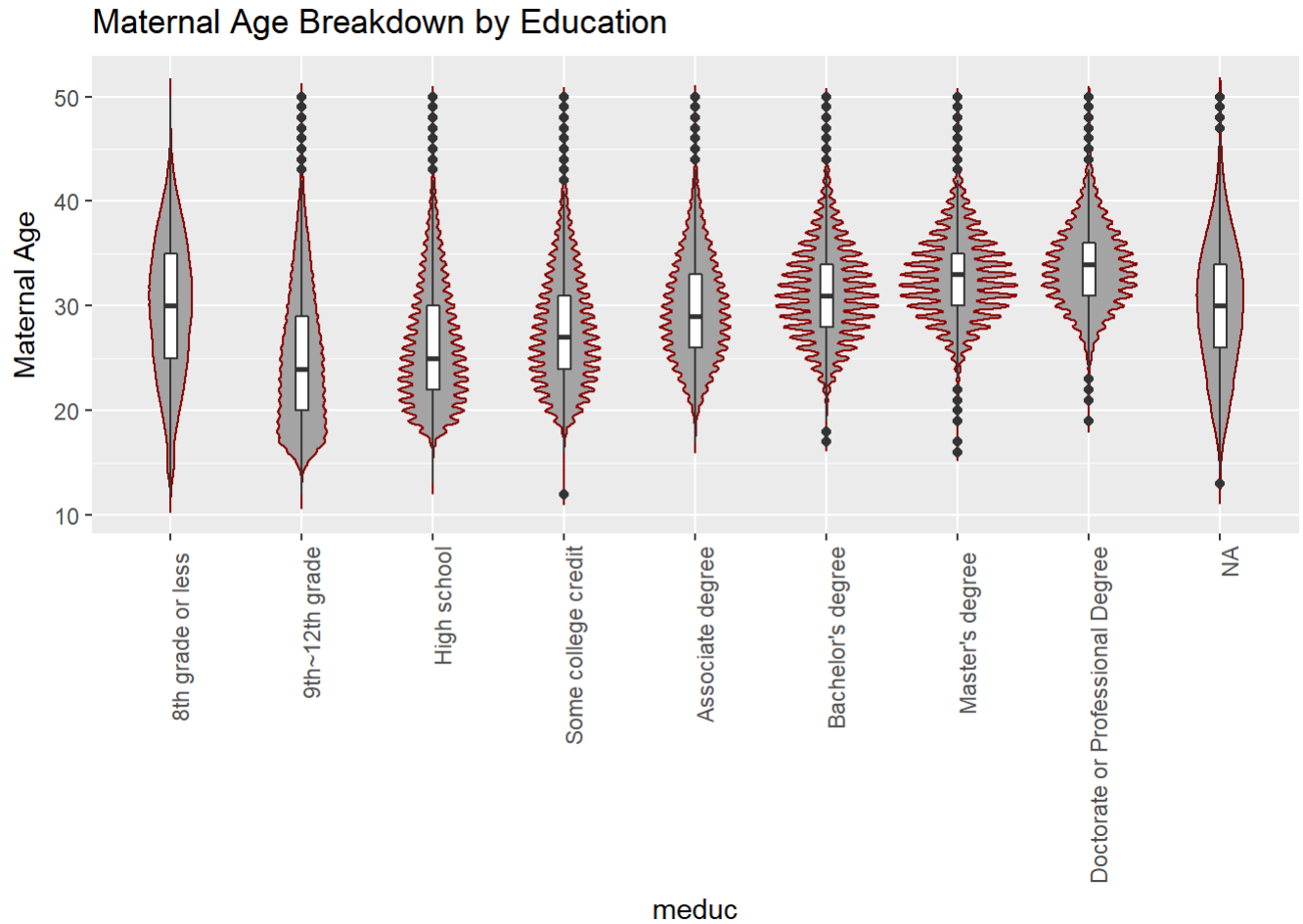
5. Include at least one key visualization of the data that exhibits an interesting characteristic.



Baby's weight has long been established to be a good indicator for baby's healt: the heavier the healthier. As shown in the graph, Baby's weight peaks at the age of 30~32 for the mother, which coincides with the average maternal age of the country. The convex shape of the curve could be explained twofolds: younger mothers tend to have less experience with neonatal care/ parenting/ financing for neonatal care, while the older mothers face genetic deficiencies of the babies due to a postponed age of giving birth.

6. Show at least one conditional (slice) description of the data (e.g., all variable descriptive statistics by nationality of survey respondent). This can be a table or visualization.


Histogram of Maternal Age Giving Birth
Density
0.00
0.02
0.04
0.06
0.08
0.10
20
30
40
50
maternal age


Maternal Age Breakdown by Education
Maternal Age
50
40
30
20
10
8th grade or less
9th~12th grade
High school
Some college credit
Associate degree
Bachelor's degree
Master's degree
Doctorate or Professional Degree
NA
meduc

The national average of giving birth to a child is around 28~29. One would suspect the more education a woman receives, the later she would postpone her age of perception. Interestingly, this suspection is generally true except for one case: those who receive education of less than 8th grade give birth at around the same time as national average, while those who receive education up until 9~12th grade are the youngest mothers.

II. Research Paper

1. State the research question of your assigned paper.

Roach Anleu and Mack (2015) investigated the performance of authority of judges by employing an obervational dataset on how the judges engage the defendent in court. They specifically looked at two measures of the magitrates' behaviors in court: 1) whether the magistrate spoke or looked at the defendent, and 2) the order of results VS reasons in her ruling. They also looked at how the results would differ with or without the presence of representation. Their hypothesis was that the more solemn trials (i.e. sentencing decisions) require more personal engagemant from the magistrate to the defendent to sustain legitimacy.

2. What data did the paper use?

The paper draws its primary data from prior observational studies of courtroom behaviour and practice. The data includes courts in different geographical locations, different demographics of magitrates, general criminal list (proceedings related to offences), magistrates' daily work, and etc. The two coauthors then use standardized coding to record information of the defendant, the offences, aspects of the magistrate's interaction, and decisions and outcome.

3. What theory did the paper reference in order to interpret the data? (Note: it is possible that the paper has no reference to theory.)

The theory that the paper heavily relies on is Weber's theory of authority (Weber 1978), which makes the connection of legitimacy, especially the magistrate's 'belief in legitimacy', and authority, reflected by how authority is exercised. In their experiment, legitimacy can be fostered/sustained by interactions with the defendent.

4. Was your assigned paper a descriptive study, an identification exercise, a numerical solution to system of equations study, or some combination of the three? (These are the three classifications we discussed in class.)

The paper is a desciptive one looking no further than performing simple statistical tests, like t-tests and χ^2 tests, on population percentages. All can be simply achieved by two-way and three-way tables and some statistical packages. It does not look to establish a causal relationship (or predictive relationship) between two variables.

5. What computational methods did this paper use to answer the research question? What was their result or answer to the question?

As mentioned above, the paper uses simple descriptives like percentages of the population, t-test, χ^2 tests to answer the research question. And agreeing with their hypothesis, in sentencing trials, there are significantly more looking and speaking to the defendents as compared to other trials. The presence of attorney does not affect the amount of such interactions in sentencing trials, but significant decreases such interactions in non-sentencing trials. Following the same logic, compliant with the Weberian theory, in sentencing trials the percentage of summaries+decision, compared to decision+reasons or only decisions significantly triumphs over that in non-centencing trials. The phenomenon preserves itself with or without the presence of representation, but the absence of representation decreases the percentage a bit more in sentencing trials.

6. Think of yourself as an academic referee. Give two suggestions to the author(s) of your assigned paper of things the authors might do to improve their results or strengthen their evidence for the answer to the question.

- 1). The paper really throws away the richness of the data (i.e. the demographics of magistrates/offender, features of the court) by simply running descriptives. These data can be used to better describe what factors can predict, or causally influence a magistrate's interactions with the defendent. Regressions or predictive machine learning methods can both serve this purpose well.
- 2). The paper fails to give a quantifiable outcome measure in how well authority/legitimacy is carried out. Some easily attainable measures can include second appeal rate, duration of trial, or even surveyed defendent perception of the fairness of the trials. With the solid outcome measure, the authors can further establish causal or non-causal relationships to see whether magistrates' interaction with the defendent increases legitamacy and authority, persuant to the Weberian theory.

Reference:

Gelberg, Lillian, Ronald M Andersen, and Barbara D Leake. 2000. "The Behavioral Model for Vulnerable Populations: Application to Medical Care Use and Outcomes for Homeless People." *Health Services Research* 34 (6). Health Research & Educational Trust:1273.

Krieger, Nancy. 1999. "Embodying Inequality: A Review of Concepts, Measures, and Methods for Studying Health Consequences of Discrimination." *International Journal of Health Services* 29 (2). SAGE Publications Sage CA: Los Angeles, CA:295–352.

Roach Anleu, Sharyn, and Kathy Mack. 2015. "Performing Authority: Communicating Judicial Decisions in Lower Criminal Courts." *Journal of Sociology* 51 (4). SAGE Publications Sage UK: London, England:1052–69.

Weber, Max. 1978. *Economy and Society: An Outline of Interpretive Sociology*. Vol. 1. Univ of California Press.

Williams, David R, and Chiquita Collins. 2001. "Racial Residential Segregation: A Fundamental Cause of Racial Disparities in Health." *Public Health Reports* 116 (5). SAGE Publications:404.