

Variable Exploration

Covariates

1. **wt**

- The birth weight (ounces).
- continuous

2. **gestation**

- The length of the gestation period (days).
- continuous

3. **parity**

- The total number of previous pregnancies (including fetal deaths and still births).
- discrete, not categorical

4. **meth**

- The self-reported ethnicity of the mother:
 - 0-5 = Caucasian,
 - 6 = Mexican,
 - 7 = African-American,
 - 8 = Asian,
 - 9 = Mixed,
 - 10= Other
- categorical

5. **mage**

- The mother's age at termination of pregnancy (years).
- theoretically continuous, in practice probably discrete *double check this*

6. **med**

- The mother's education:
 - 0 = elementary school
 - 1 = middle school
 - 2 = high school
 - 3 = high school + trade school,
 - 4 = high school + some college,
 - 5 = college graduate,
 - 6 = trade school,
 - 7 = highschool unclear.
- categorical

7. **mht**

- The mother's height (inches).
- continuous

8. **mwt**

- The mother's pregnancy weight (pounds).
- continuous

9. **feth**

- The father's ethnicity (same coding as meth).
 - 0-5 = Caucasian,
 - 6 = Mexican,
 - 7 = African-American,
 - 8 = Asian,
 - 9 = Mixed,
 - 10= Other
- categorical

10. **fage**

- The father's age at end of pregnancy (years).
- theoretically continuous, probably discrete in practice *double check*

11. **fed**

- The father's education (same coding as med).
 - 0 = elementary school
 - 1 = middle school
 - 2 = high school
 - 3 = high school + trade school,
 - 4 = high school + some college,
 - 5 = college graduate,
 - 6 = trade school,
 - 7 = highschool unclear.
- categorical

12. **fht**

- The father's height (inches).
- continuous

13. **fwt**

- the father's weight (pounds).
- continuous

14. **marital**

- The mother's marital status:
 - (a) married,
 - (b) legally separated
 - (c) divorced
 - (d) widowed,
 - (e) never married.
- categorical

15. **income**

- The family yearly income in \$2500 (USD) increments:
 - 0 = under 2500
 - 1 = 2500-4999
 - 2 = 5000-7499
 - 3 = 7500-9999
 - 4 = 10000-12499
 - 5 = 12500-14999
 - 6 = 15000-17499
 - 7 = 17500-19999
 - 8 = 20000-22499
 - 9 = over 22500.
- categorical

16. **smoke**

- Does the mother smoke at time of pregnancy?
 - 0 = never
 - 1 = smokes now
 - 2 = until pregnancy
 - 3 = used to, not anymore.
- categorical

17. **time**

- Time since the mother quit smoking before pregnancy:
 - 0 = never smoked,
 - 1 = still smokes,
 - 2 = during pregnancy
 - 3 = less than a year,
 - 4 = 1-2 years,
 - 5 = 2-3 years,
 - 6 = 3-4 years,

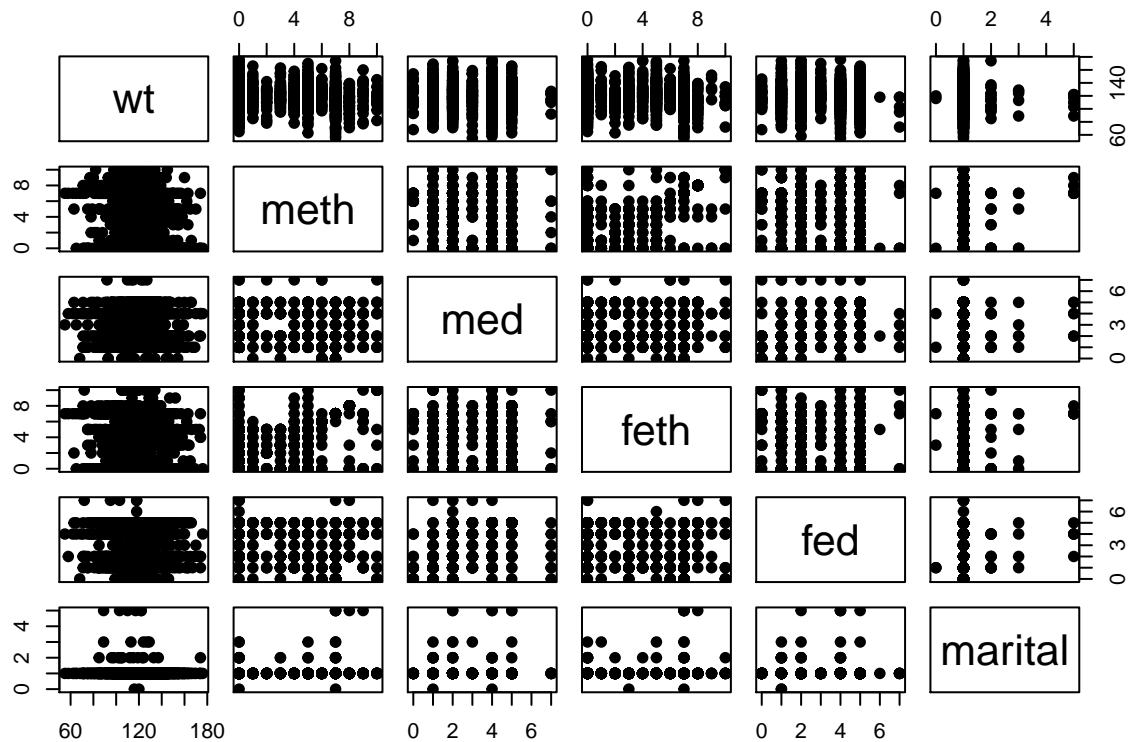
- 7 = 5-9 years,
- 8 = more than 10 years,
- 9 = quit but don't know when.
- categorical

18. number

- Number of cigarettes smoked per day by mother when she was smoking:
 - 0 = never smoked,
 - 1 = 1-4,
 - 2 = 5-9,
 - 3 = 10-14,
 - 4 = 15-19,
 - 5 = 20-29,
 - 6 = 30-39,
 - 7 = 40-60,
 - 8 = more than 60,
 - 9 = smoked but don't know how much.
- categorical

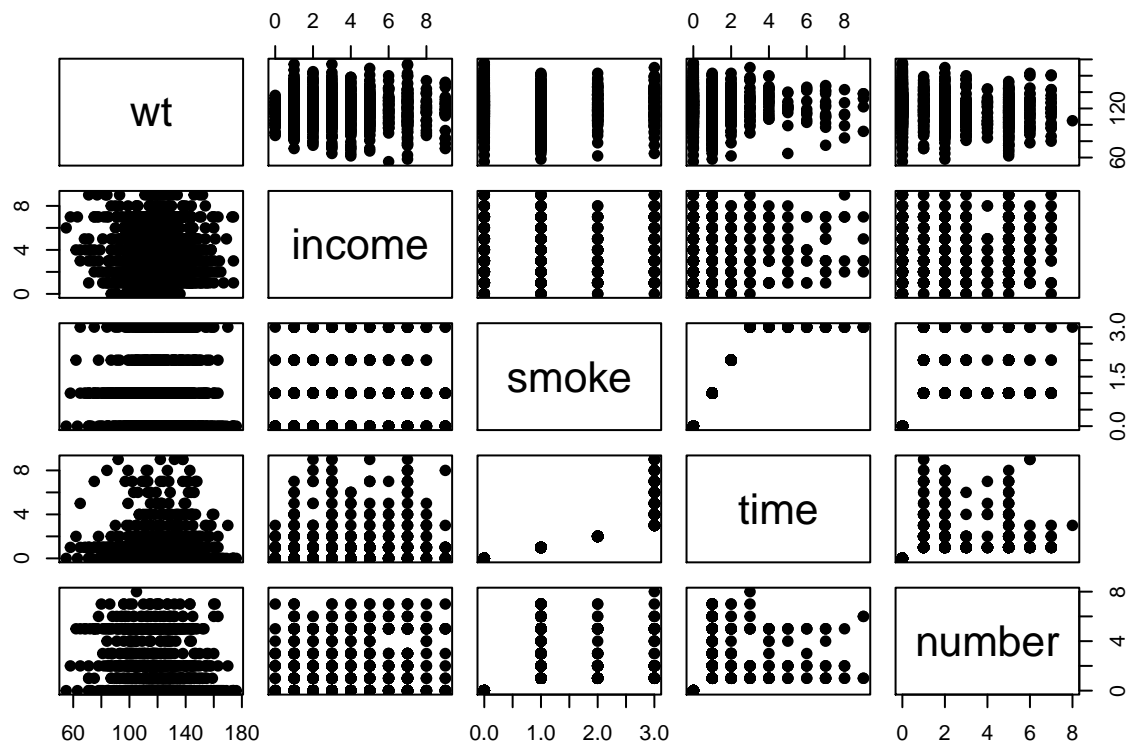
```
cat.inds <- c(4,6,9,11,14,15,16,17,18)
discr.inds <- c(3, 5, 10 )
cts.inds <- c(1, 2,7,8,12,13)
births <- read.csv("chds_births.csv")

pairs(births[, c(1, 4, 6, 9, 11, 14)], pch=19)
```



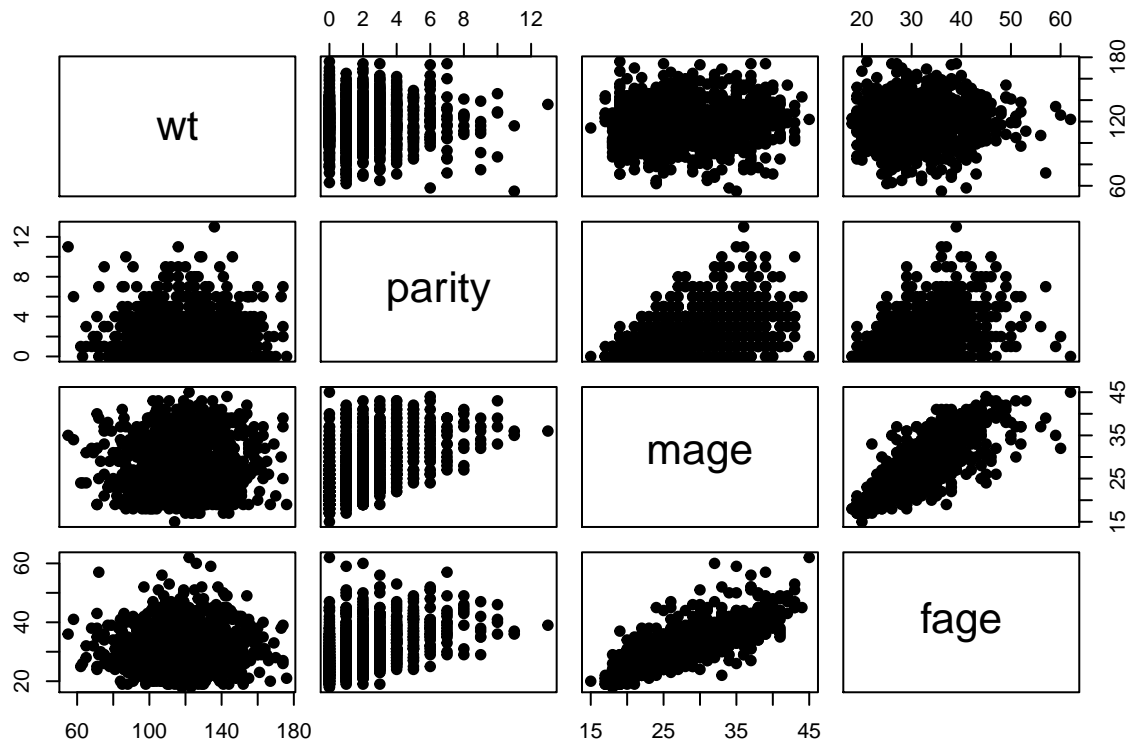
Variables to include: marital, fed, med, maybe meth, maybe feth

```
pairs(births[, c(1, 15, 16, 17, 18)], pch=19)
```



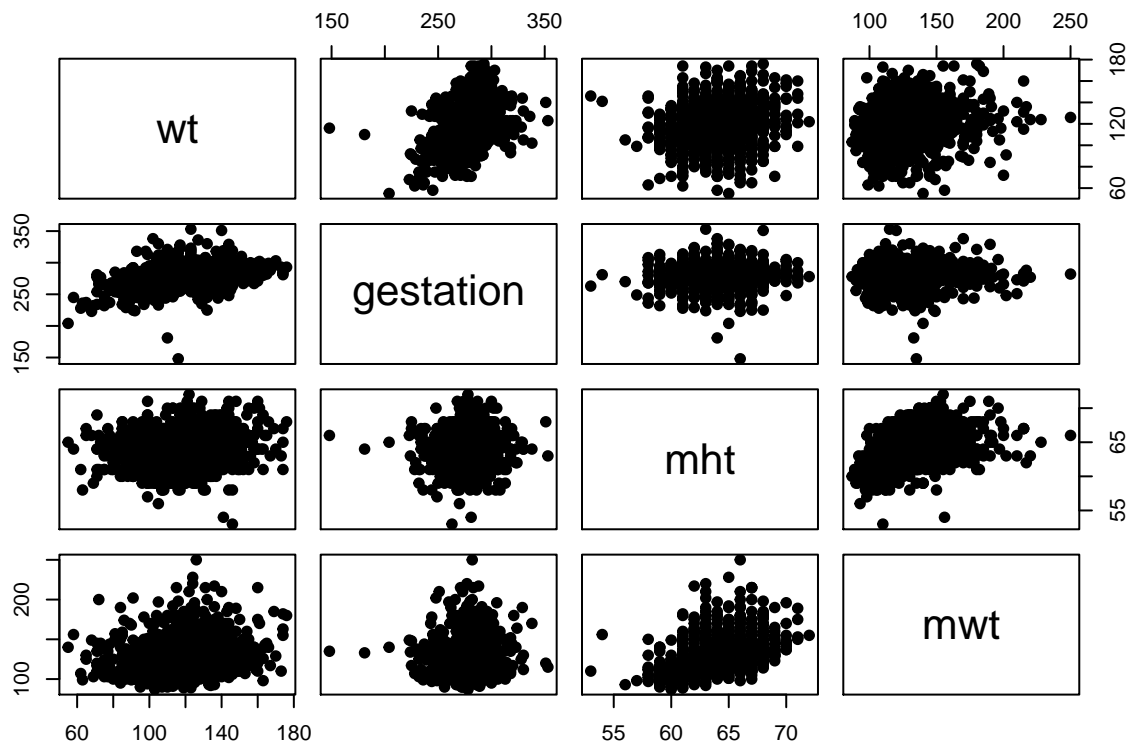
time, number, income, maybe smoke (should try grouping 0 and 3 together; 1 and 2 together)

```
pairs(births[, c(1, 3, 5, 10)], pch=19)
```



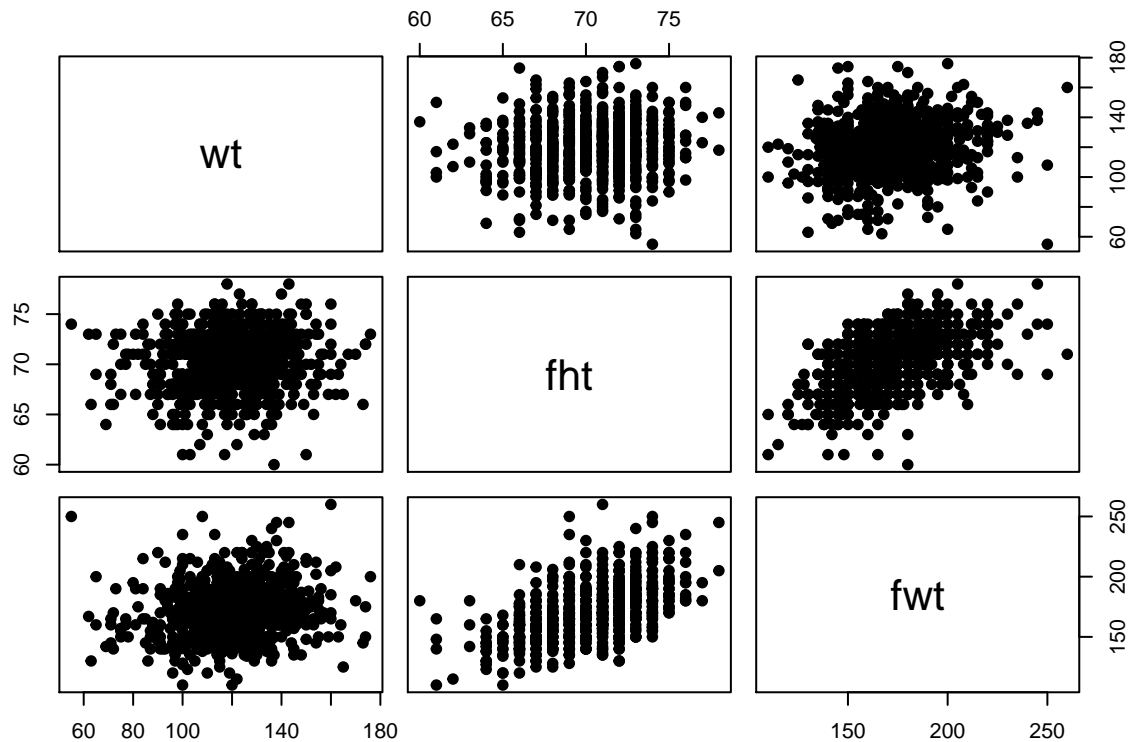
Variables to include: parity, maybe mage or fage (note that mage and fage are correlated)

```
pairs(births[, c(1, 2, 7, 8)], pch=19)
```



Variables to include: gestation, maybe mht/mwt (note that these are slightly correlated)

```
pairs(births[, c(1, 12, 13)], pch=19)
```



fht and fwt do not seem to be good variables to include (and are definitely correlated)

Missing Data

Note that 5-10% of observations = 61.8-123.6

```
summary(births)
```

```
##           wt           gestation           parity           meth
##  Min.      : 55.0    Min.      :148.0    Min.      : 0.000    Min.      : 0.000
##  1st Qu.:108.8    1st Qu.:272.0    1st Qu.: 0.000    1st Qu.: 0.000
##  Median :120.0    Median :280.0    Median : 1.000    Median : 3.000
##  Mean   :119.6    Mean   :279.3    Mean   : 1.932    Mean   : 3.129
##  3rd Qu.:131.0    3rd Qu.:288.0    3rd Qu.: 3.000    3rd Qu.: 7.000
##  Max.   :176.0    Max.   :353.0    Max.   :13.000    Max.   :10.000
##                NA's      :13                NA's      :1
##           mage           med           mht           mwt
##  Min.      :15.00    Min.      :0.000    Min.      :53.00    Min.      : 87.0
##  1st Qu.:23.00    1st Qu.:2.000    1st Qu.:62.00    1st Qu.:114.8
##  Median :26.00    Median :2.000    Median :64.00    Median :125.0
##  Mean   :27.26    Mean   :2.917    Mean   :64.05    Mean   :128.6
##  3rd Qu.:31.00    3rd Qu.:4.000    3rd Qu.:66.00    3rd Qu.:139.0
##  Max.   :45.00    Max.   :7.000    Max.   :72.00    Max.   :250.0
##  NA's      :2      NA's      :1      NA's      :22    NA's      :36
##           feth           fage           fed           fht
##  Min.      : 0.000    Min.      :18.00    Min.      :0.000    Min.      :60.0
##  1st Qu.: 0.000    1st Qu.:25.00    1st Qu.:2.000    1st Qu.:68.0
##  Median : 3.000    Median :29.00    Median :4.000    Median :71.0
##  Mean   : 3.154    Mean   :30.35    Mean   :3.127    Mean   :70.2
##  3rd Qu.: 7.000    3rd Qu.:34.00    3rd Qu.:5.000    3rd Qu.:72.0
```

##	Max.	:10.000	Max.	:62.00	Max.	:7.000	Max.	:78.0
##	NA's	:31	NA's	:7	NA's	:13	NA's	:492
##	fwt		marital		income		smoke	
##	Min.	:110.0	Min.	:0.000	Min.	:0.000	Min.	:0.0000
##	1st Qu.	:155.0	1st Qu.	:1.000	1st Qu.	:2.000	1st Qu.	:0.0000
##	Median	:170.0	Median	:1.000	Median	:3.000	Median	:1.0000
##	Mean	:171.2	Mean	:1.038	Mean	:3.701	Mean	:0.8018
##	3rd Qu.	:185.0	3rd Qu.	:1.000	3rd Qu.	:5.000	3rd Qu.	:1.0000
##	Max.	:260.0	Max.	:5.000	Max.	:9.000	Max.	:3.0000
##	NA's	:499			NA's	:124	NA's	:10
##	time		number					
##	Min.	:0.0000	Min.	:0.00				
##	1st Qu.	:0.0000	1st Qu.	:0.00				
##	Median	:1.0000	Median	:1.00				
##	Mean	:0.9625	Mean	:1.76				
##	3rd Qu.	:1.0000	3rd Qu.	:3.00				
##	Max.	:9.0000	Max.	:8.00				
##	NA's	:10	NA's	:21				

Variables to remove due to missing data:

- fht (492)
- fwt (499)
- income (124)