Case Study 1

Emily Gentles Phuc Nguyen Joseph Lawson

Jan 21, 2020

Case Discussion

- Data obtained from a subset of women enrolled in the CPP during pregnancy

Goal: Assess how exposure to DDE and PCBs relates to the risk of premature delivery and the relative severity thereof

Exploratory Data Analysis

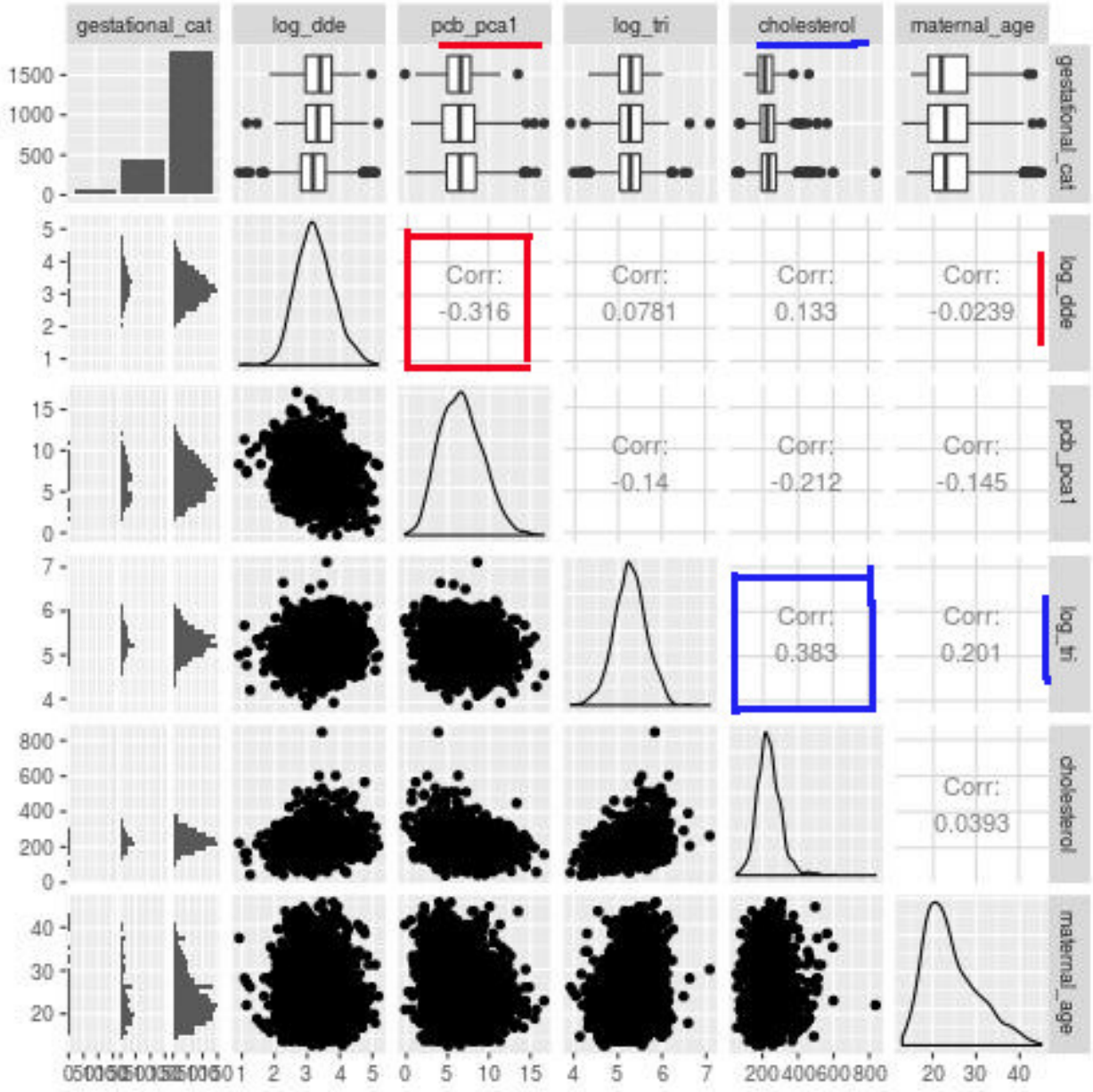
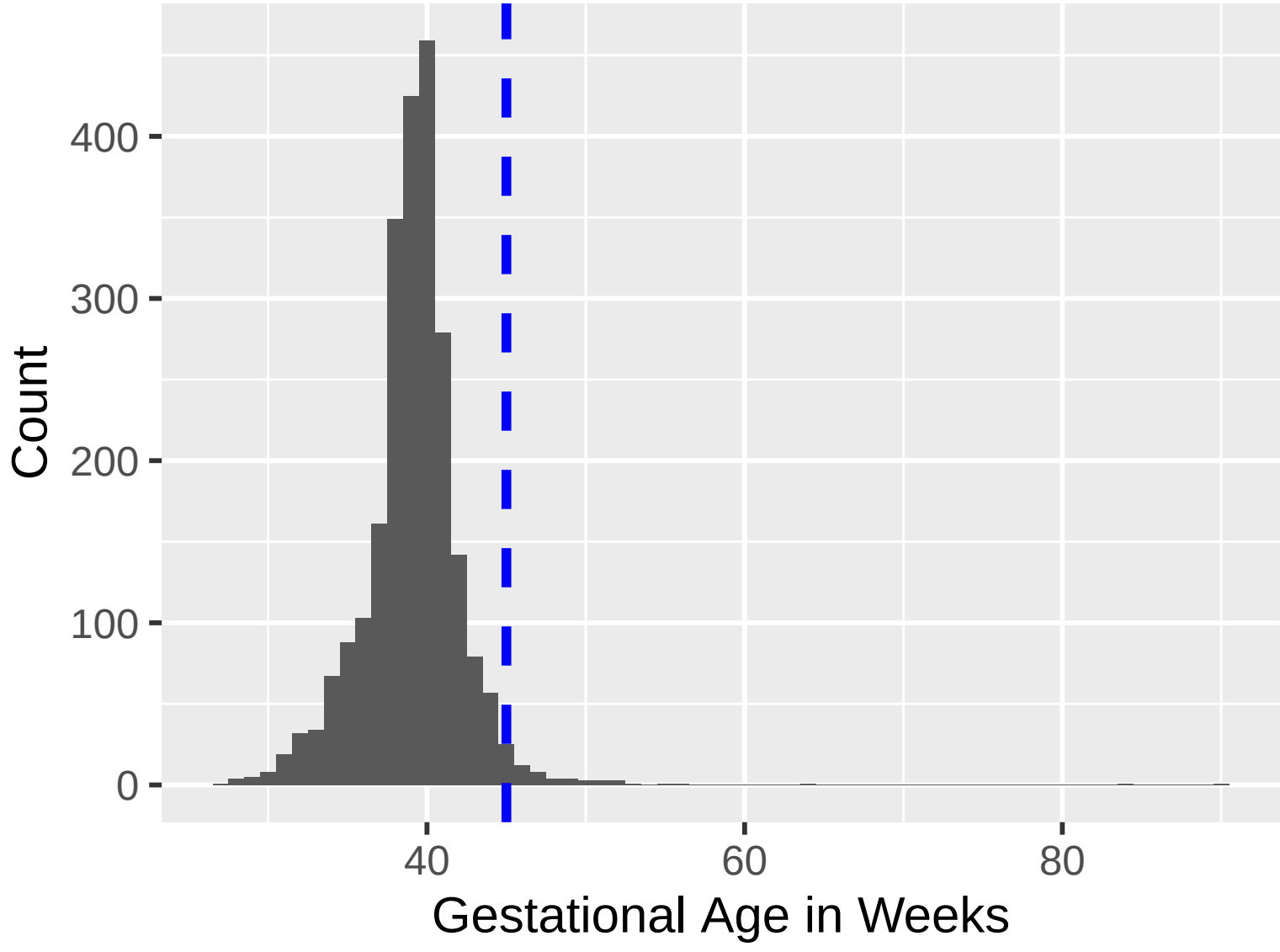
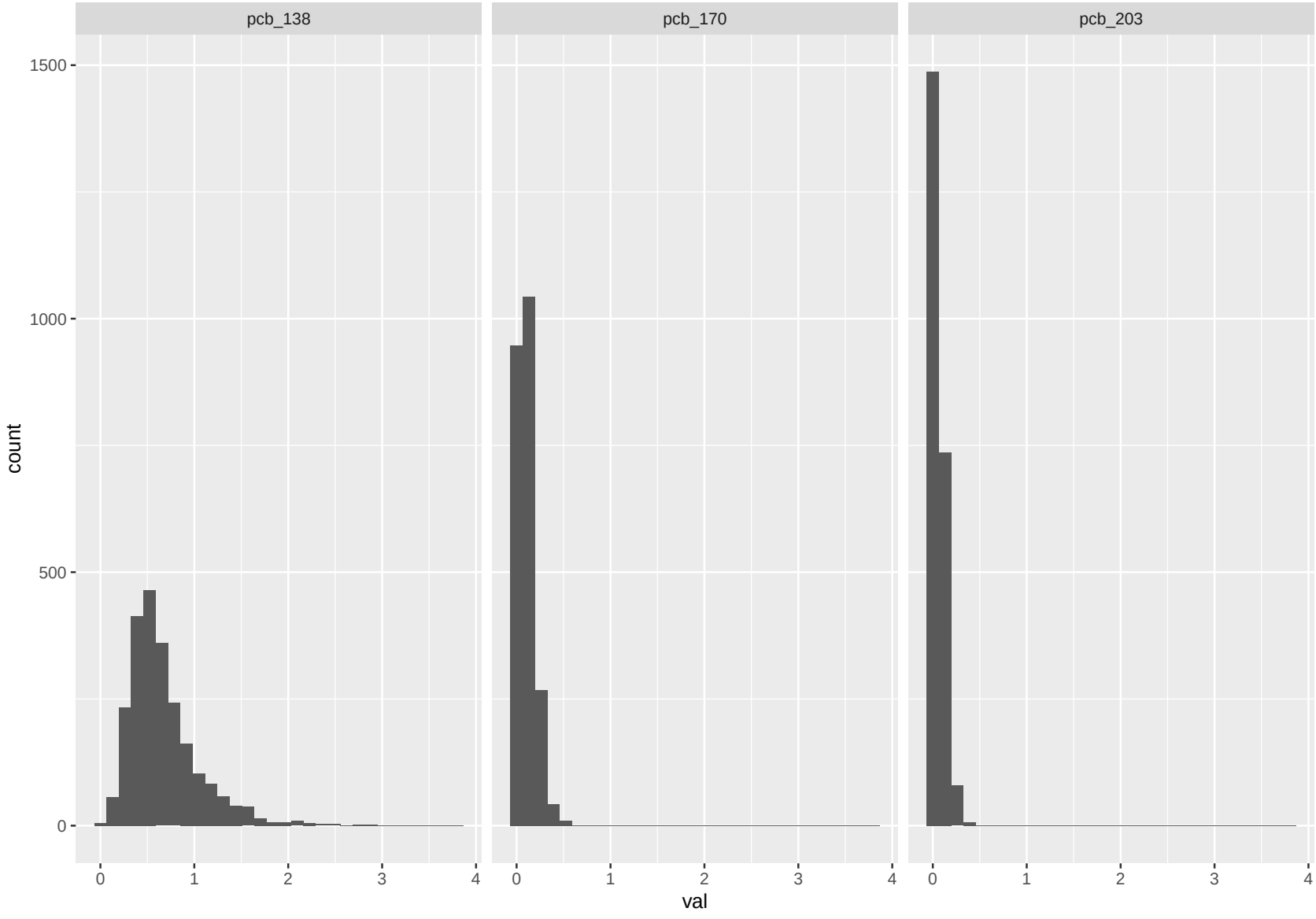
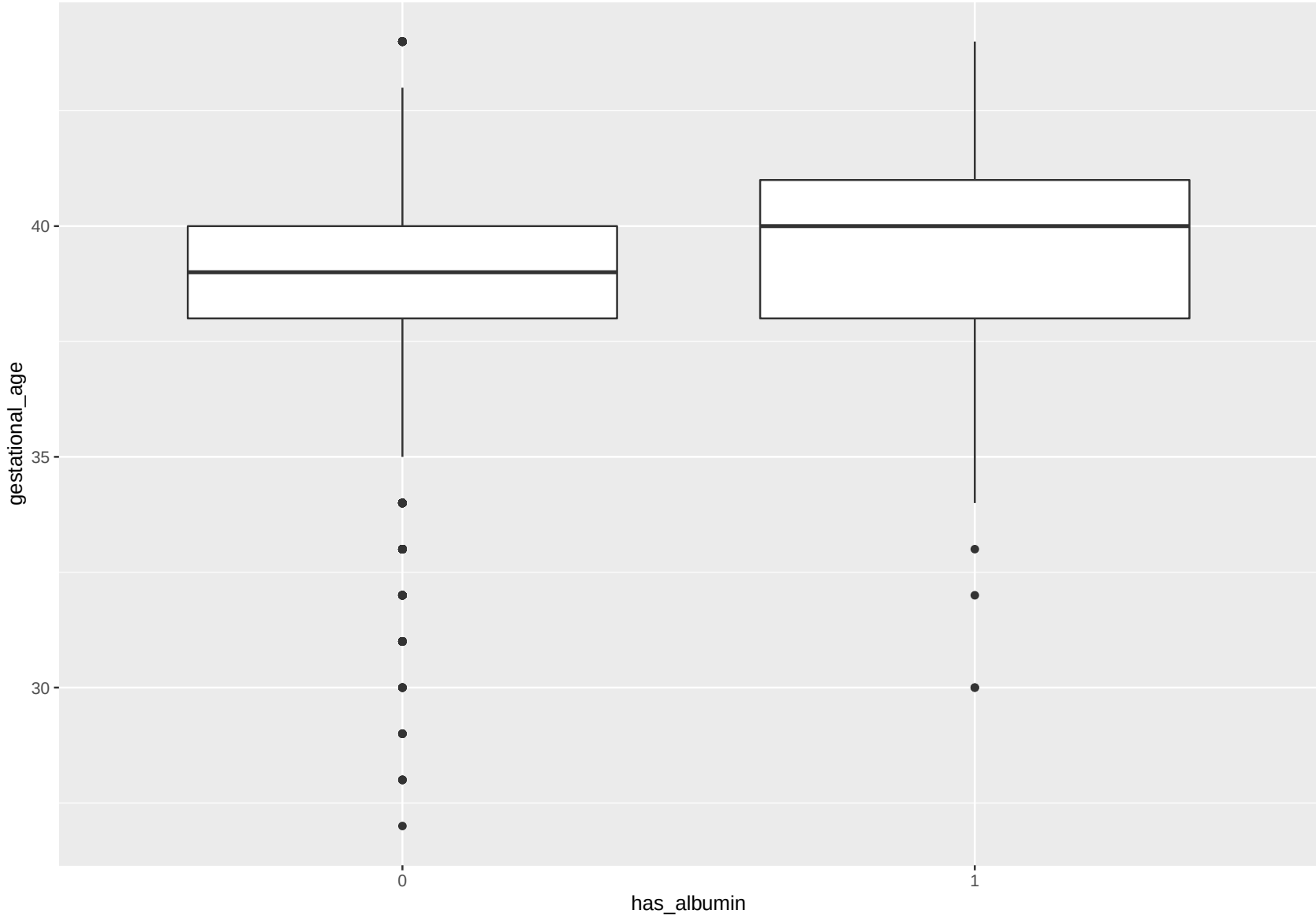


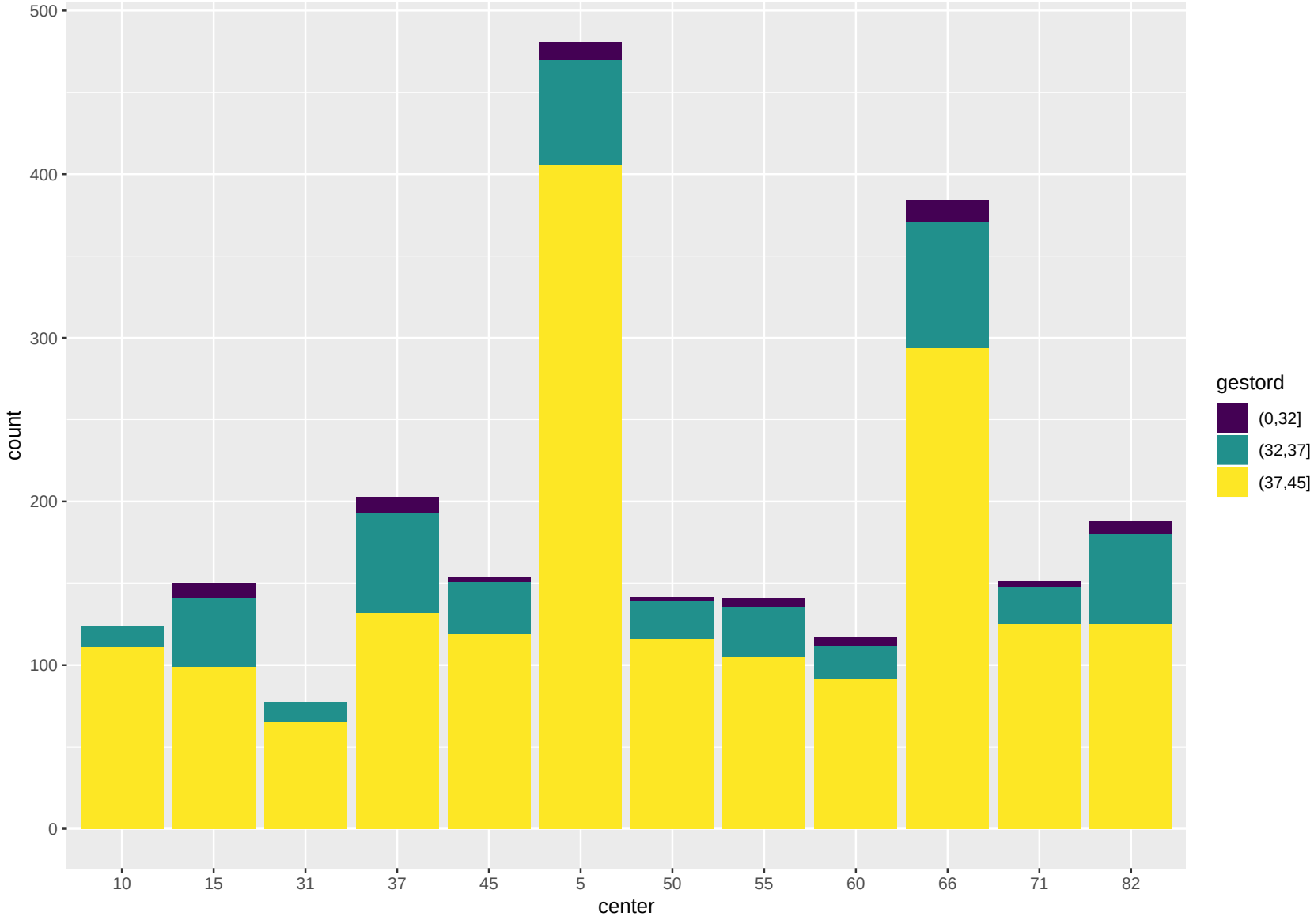
Figure 1: corr-plot

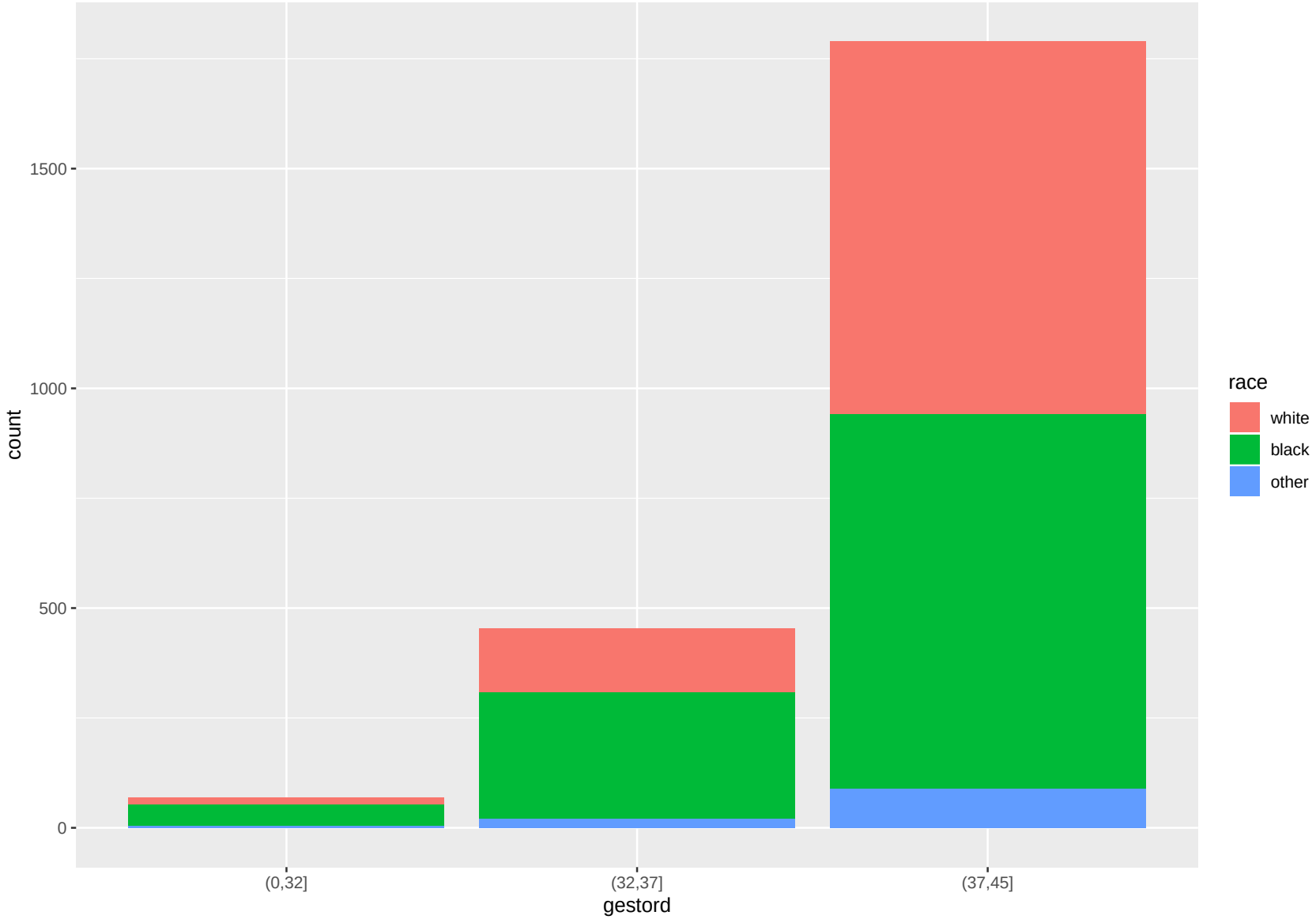
- Data issues: uncertainty, inflation, and missingness

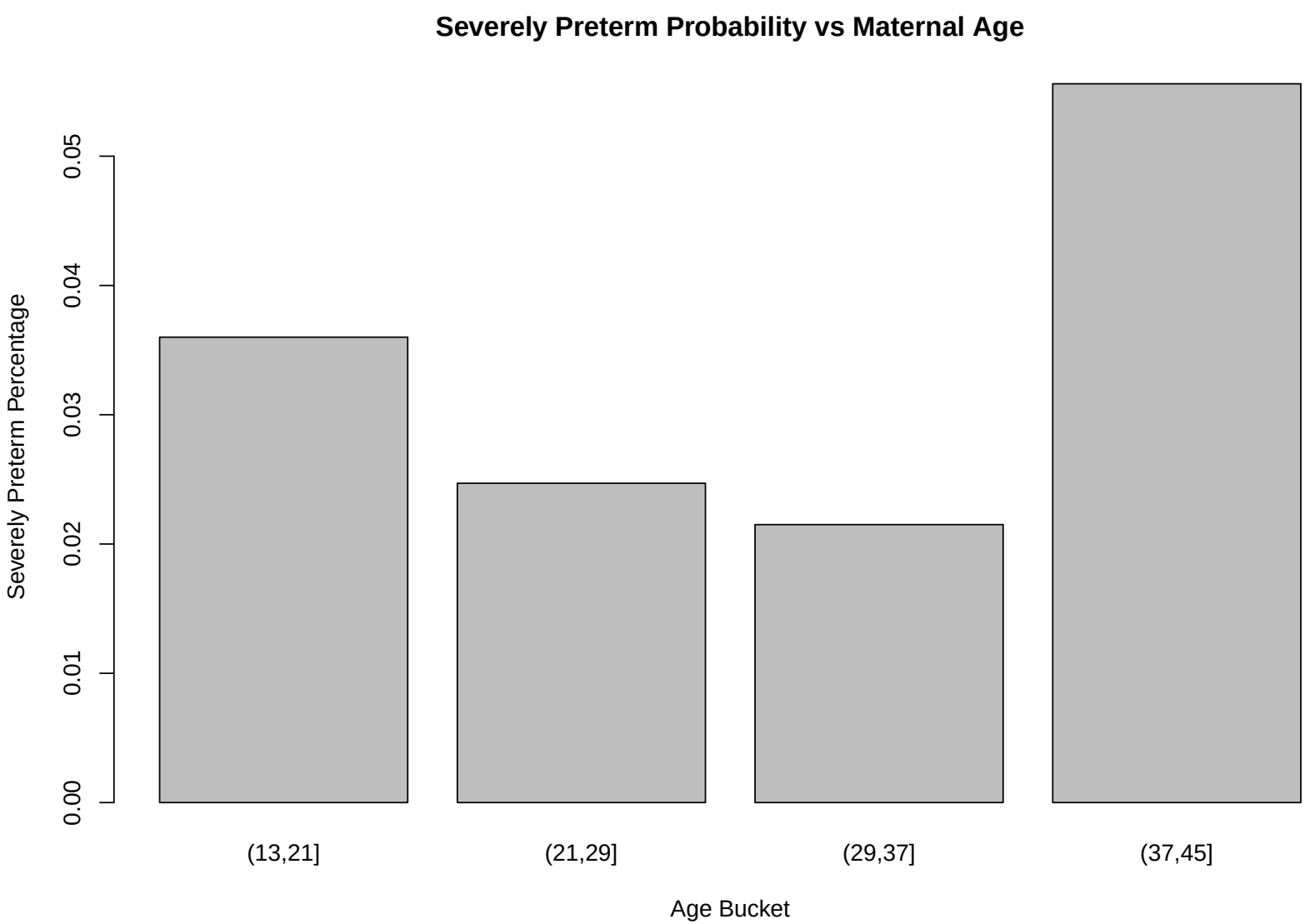



pcb_138
pcb_170
pcb_203
1500
1000
500
0
count
0
1
2
3
4
val


gestational_age
40
35
30
0
1
has_albumin


500
400
300
200
100
0
count
10
15
31
37
45
5
50
55
60
66
71
82
center
gestord
(0,32]
(32,37]
(37,45]


1500
1000
500
0
count
(0,32]
(32,37]
(37,45]
gestord
race
white
black
other


Severely Preterm Probability vs Maternal Age
Severely Preterm Percentage
0.00
0.01
0.02
0.03
0.04
0.05
(13,21]
(21,29]
(29,37]
(37,45]
Age Bucket

Analysis

- Ordinal Logistic Regression with Term, Preterm, and Severely Preterm Gest. Categories
  - Useful interpretation in terms of risk
  - Uses naturally ordinal structure
  - Differentiates between different severities
- Impute score data with MICE to check usefulness
- Remove obs with missing PCB value
- Include blood cholesterol/triglyceride levels, as well as center and SES/Lifestyle metrics
- Use PCA to handle the various PCA measures.

Model Comparison F-Tests

- Model Comparison indicated (p=0.97) that the first principle component of the pcb_* values is sufficient.
- Indication (p=0.74) against including Score Variables (post imputation)
- Indication (p=0.21) against including Center interactions
- Indication (p=0) for including Center as variable
  - Indicates heterogeneity in preterm birth risk accross medical centers

- Indication (p=0.46) against PCB-DDE interaction effect
- Indication (p=0.15) (weakly) against Triglyceride interaction with PCE/DDE
  - Hypothesized that there might be because of fat-solubility of the contaminants of interest
- Indication (p=0.02) for inclusion of quadratic term in maternal age
  - Makes sense given above chart/intuition of increased risk at extremes of age.
- Strong indication that the indicator of testing for Albumin is associated with longer gestational period on the margin

- Control Variables of Significance (marginally) in Final Model:
  - Center
  - Triglycerides (g/dL)
  - Cholesterol (g/dL)
  - Maternal Age (years)
  - Albumin (Testing Indicator)
  - Race

Final Model

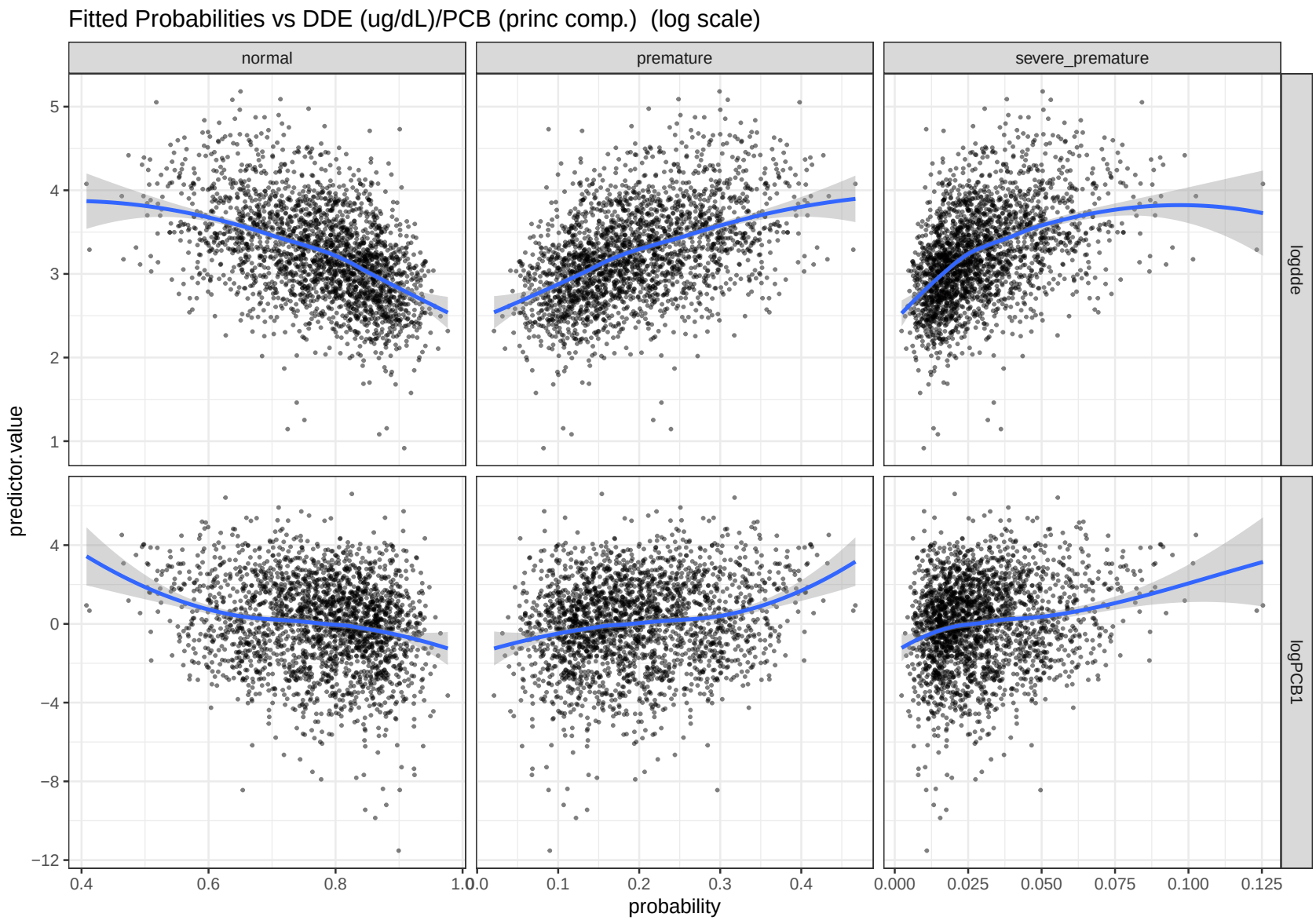
$$
\begin{aligned}
GestCategory \sim\ & Center + log(dde) + logPCB1 && (1)\\
& + log(trigl.) + Poly(MaternalAge, 2) && (2)\\
& + Smoking + log(Cholest.) + AlbuminTested && (3)\\
& + race && (4)
\end{aligned}
$$

Results

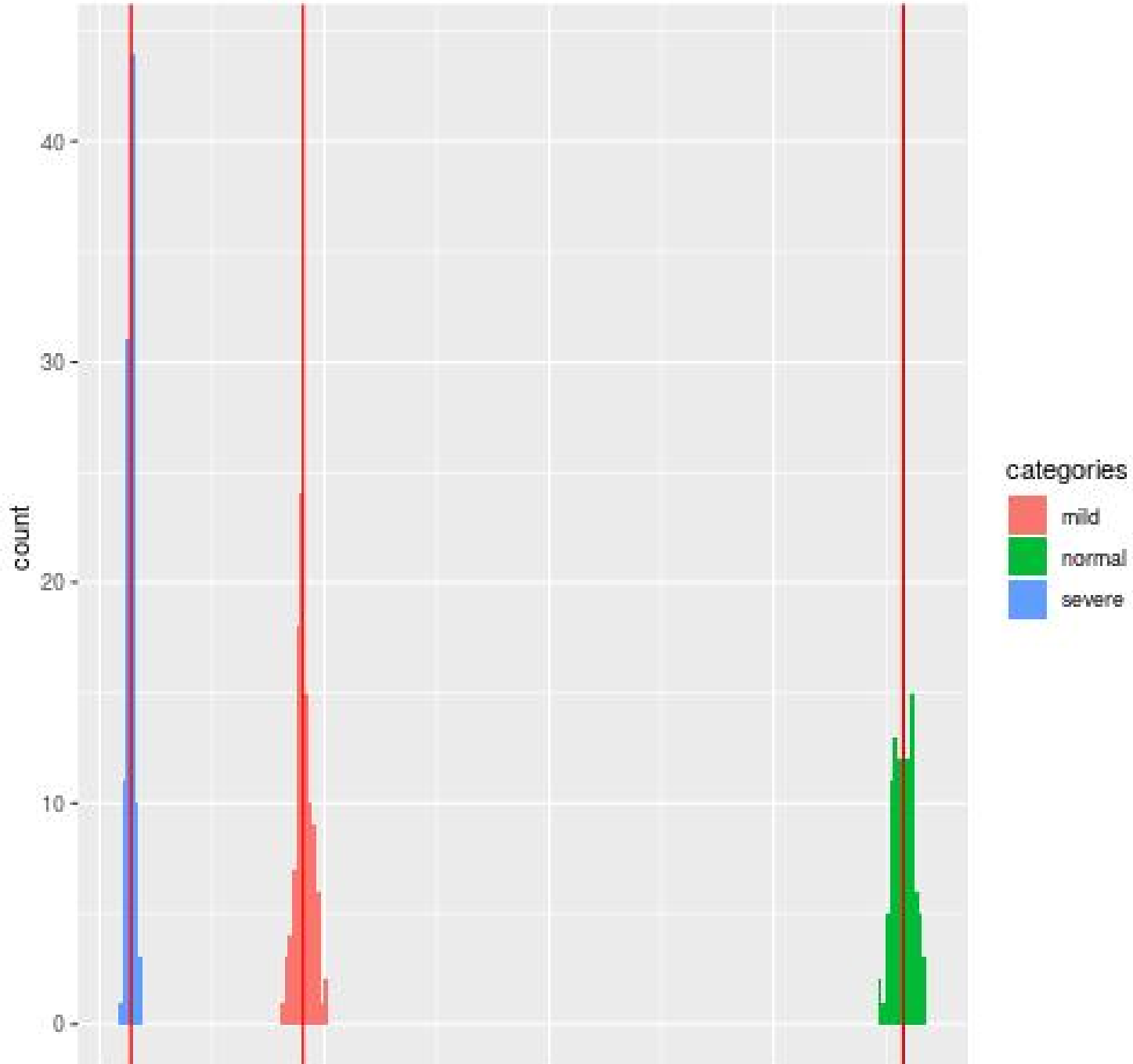
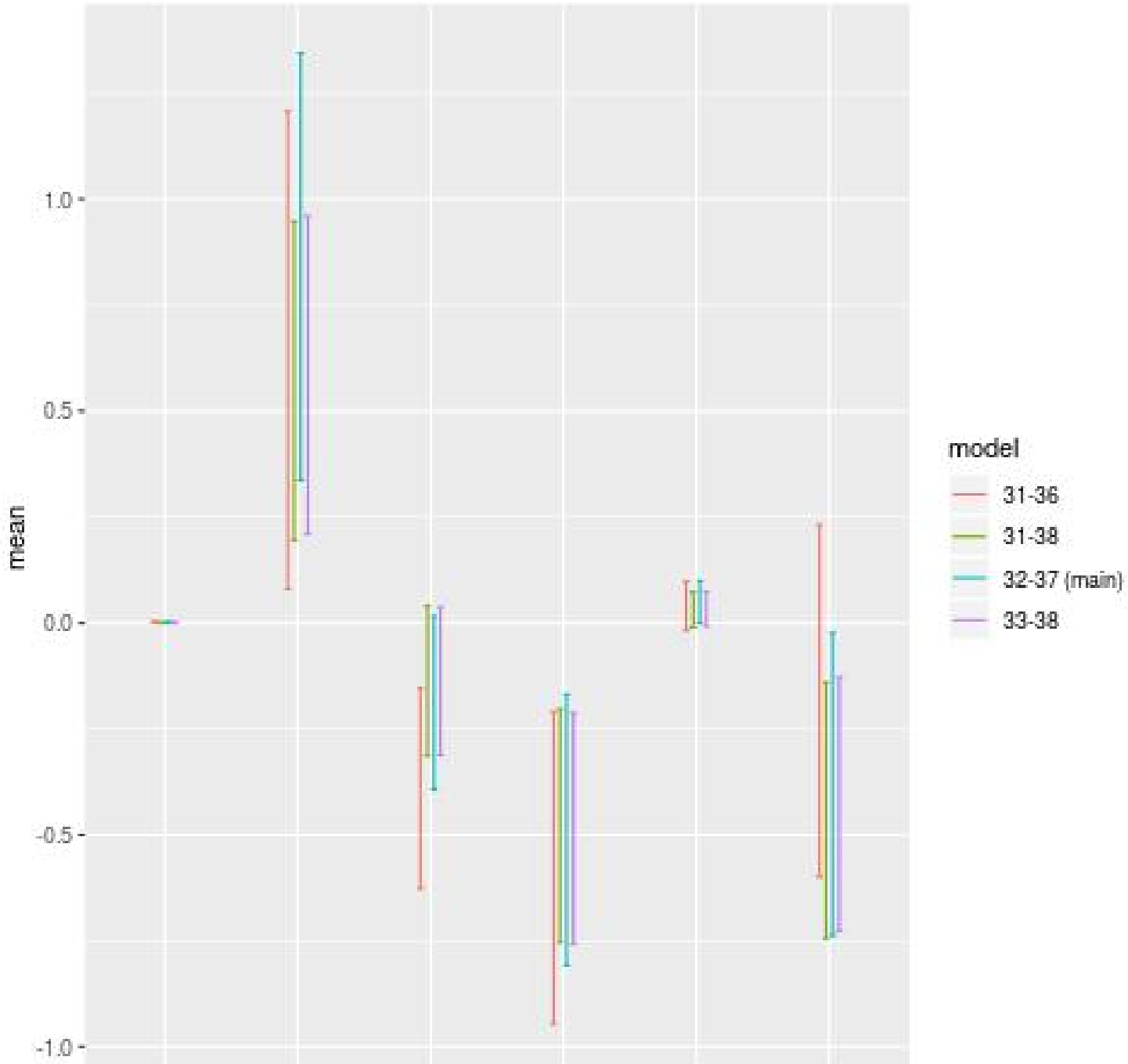
- (log) DDE and PCB are (collectively) significantly associated with preterm delivery likelihood even when adjusting for other factors
- Interpretation that 1% decrease in DDE is approximately associated with a -0.18% change, and a unit increase in logPCB1 with a -0.06 change, in the log odds of of a gestational age occurrence above a given ordinal threshold.
  - Note for interpretability that the PCB coefficient is relative to the (standardized/centered) first principle component of the pcb_* measurements
- Testing inclusion of both DDE and PCB vs control indicate (p=0.005) that they provide increased explanatory power vs control, but marginally in the full model each is just at the edge of significance, as seen in the following table:

	Coef Est	2.5 %	97.5 %
log(dde)	-0.183	-0.388	0.022
logPCB1	-0.056	-0.113	0.000

- Note the overlap of the confidence intervals with 0.


Fitted Probabilities vs DDE (ug/dL)/PCB (princ comp.) (log scale)
normal
premature
severe_premature
logdde
logPCB1
predictor.value
probability
5
4
3
2
1
4
0
−4
−8
−12
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
0.000
0.025
0.050
0.075
0.100
0.125

Sensitivity Analysis



Sensitivity Analysis