Baby_faces_data_analysis

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.4.4     v tibble    3.2.1
v lubridate 1.9.3     v tidyr     1.3.1
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

```
library(dplyr)
```

```
data <-read.csv("~/Documents/2024spring/summer_surf/ICIS_data_analysis/data/BabyFaces2018_
```

```
str(data)
```

```
'data.frame':   2847 obs. of  676 variables:
 $ CHILDID               : int  10004113 10006315 10015113 10016219 10020319 10029114 10030
 $ D1_ID                 : int  338833 581936 311838 906834 370839 969833 909835 331933 5028
 $ CENTERID              : int  962339 538237 722230 186232 -3 507330 355432 751430 503246
 $ CPID                  : int  12142114 77262013 38072112 92636118 38404012 30007110 33899
 $ CLASSID               : int  562937 545718 951045 358842 567916 729945 194653 331876 7266
 $ PI1_DATA              : num  1 1 1 1 1 1 1 1 1 1 ...
 $ PI1_DATE              : chr  "18-Apr" "18-May" "18-Apr" "18-May" ...
 $ PR1_DATA              : int  1 0 1 1 0 1 1 1 1 0 ...
 $ PR1_DATE              : chr  "18-Mar" " " "18-Apr" "18-May" ...
```

```
$ SR1_DATA                : num  1 1 1 1 0 1 1 1 1 1 ...
$ SR1_DATE                : chr  "18-May" "18-May" "18-May" "18-May" ...
$ T_HV_FLAG               : int  1 2 1 1 2 1 1 1 1 1 ...
$ ISPREGNANT              : int  0 0 0 0 0 0 0 0 0 0 ...
$ CHILDMONTHS             : int  19 31 28 31 25 34 29 25 9 22 ...
$ CHILDSEX                : int  1 2 1 2 1 2 2 1 2 1 ...
$ STRAT                   : chr  "M" "M" "M" "M" ...
$ Productive.Vocabulary   : int  15 0 88 100 0 71 73 44 8 0 ...
$ Comprehension.Vocabular: int  66 0 99 100 0 83 95 77 4 0 ...
$ PI_WT                   : num  62.7 64.3 71.8 39 108 ...
$ PR_WT                   : num  50.2 NA 107.7 64.9 NA ...
$ SCR_WT                  : num  50.2 56.8 35.9 49.6 NA ...
$ CHILDNC_WT              : num  50.2 56.8 35.9 39 108 ...
$ PI_CLASS_WT             : num  62.7 NA 71.8 39 NA ...
$ SCRPR_CLASS_WT          : num  50.2 NA 35.9 49.6 NA ...
$ PR_CLASS_WT             : num  50.2 NA 107.7 64.9 NA ...
$ SCR_CLASS_WT            : num  50.2 NA 35.9 49.6 NA ...
$ PI_PR_SCR_WT            : num  62.7 64.3 71.8 49.6 NA ...
$ PI_SCR_WT               : num  62.7 64.3 71.8 49.6 NA ...
$ PI_PR_WT                : num  62.7 NA 107.7 66.1 NA ...
$ PIPRSCRCLASSTEACHER_WT  : num  62.7 NA 71.8 49.6 NA ...
$ PIPRSCR_HV_WT           : num  NA 64.3 NA NA NA ...
$ PI1SERVTYPE             : int  1 2 1 3 2 1 1 1 1 2 ...
$ PI1G04REV               : int  5 -3 5 4 -3 5 5 5 5 -3 ...
$ PI1R_ID                 : int  1 3 1 1 1 1 1 1 3 1 ...
$ PI1P1_ID                : int  3 1 -3 -3 3 -3 6 -3 -3 -3 ...
$ PI1P2_ID                : int  -3 -3 -3 -3 -3 -3 -3 -3 -3 -3 ...
$ PI1R_AGEC               : int  3 6 5 6 6 3 4 4 5 5 ...
$ PI1M_AGEC               : int  3 6 5 6 6 3 4 4 5 5 ...
$ PI1F_AGEC               : int  -9 6 -9 -9 -9 -9 -9 -9 5 -9 ...
$ PI1M_AGE_FSTBC          : int  2 4 3 3 6 1 3 3 4 1 ...
$ PI1R_RACE               : int  2 6 3 1 2 3 1 3 3 3 ...
$ PI1M_RACE               : int  2 6 3 1 2 3 1 3 -9 3 ...
$ PI1F_RACE               : int  2 6 -9 -9 2 -9 -9 -9 3 -9 ...
$ PI1R_YRS_USC            : int  -3 -3 -3 -3 -3 -3 -3 1 3 3 ...
$ PI1M_YRS_USC            : int  -3 3 -3 -3 -3 -3 -3 1 -3 3 ...
$ PI1F_YRS_USC            : int  -3 -3 -3 -3 -3 -3 -3 -3 3 -3 ...
$ PI1R_EDUC               : int  2 5 5 2 2 6 2 2 5 2 ...
$ PI1M_EDUC               : int  2 4 5 2 2 6 2 2 5 2 ...
$ PI1F_EDUC               : int  2 5 3 -8 7 1 1 -8 5 1 ...
$ PI1R_EMPLY              : int  1 1 2 1 1 5 5 1 1 1 ...
$ PI1M_EMPLY              : int  1 5 2 1 1 5 5 1 -9 1 ...
$ PI1F_EMPLY              : int  5 1 -9 -9 1 -9 -9 -9 1 -9 ...
```

```
$ PI1C_BIRTHWEIGHTC      : int  1 2 1 1 2 1 2 1 1 1 ...
$ PI1C_RACE              : int  2 6 3 1 2 3 1 -9 3 3 ...
$ PI1C_RACE4             : int  2 4 3 1 2 3 1 -9 3 3 ...
$ PI1HH_BIO              : int  1 1 2 2 1 2 2 2 3 2 ...
$ PI1C_LNG_SPKTO         : int  1 1 3 1 1 3 1 2 5 3 ...
$ PI1C_LNG_SPKBY         : int  1 1 1 1 1 3 1 2 -3 -3 ...
$ PI1C_LNG_SPKTO_SP      : int  -3 -3 2 -3 -3 2 -3 1 -3 2 ...
$ PI1C_LNG_SPKTO_OTH     : int  -3 -3 -3 -3 -3 -3 -3 -3 2 -3 ...
$ PI1P_IMGRNT            : int  0 1 0 0 0 0 0 1 1 1 ...
$ PI1M_PREG              : int  0 0 0 0 0 0 0 0 0 0 ...
$ PI1R_MALE              : int  0 1 0 0 0 0 0 0 1 0 ...
$ PI1P1_MALE             : int  1 0 -3 -3 1 -3 1 -3 -3 -3 ...
$ PI1P2_MALE             : int  -3 -3 -3 -3 -3 -3 -3 -3 -3 -3 ...
$ PI1R_AGE               : int  20 36 33 38 41 21 26 26 31 33 ...
$ PI1M_AGE               : int  20 38 33 38 41 21 26 26 32 33 ...
$ PI1F_AGE               : int  -9 36 -9 -9 -9 -9 -9 -9 31 -9 ...
$ PI1M_AGE_FCB           : int  19 35 31 36 39 18 23 24 31 31 ...
$ PI1F_AGE_FCB           : int  -9 33 -9 -9 -9 -9 -9 -9 31 -9 ...
$ PI1M_AGE_FSTB          : int  18 25 22 22 36 17 23 23 25 17 ...
$ PI1R_BORNINUS          : int  1 1 1 1 1 1 1 0 0 0 ...
$ PI1M_BORNINUS          : int  1 0 1 1 1 1 1 0 -9 0 ...
$ PI1F_BORNINUS          : int  1 1 -9 -9 1 -9 -9 -9 0 -9 ...
$ PI1M_YRS_US            : int  -3 17 -3 -3 -3 -3 -3 5 -3 32 ...
$ PI1F_YRS_US            : int  -3 -3 -3 -3 -3 -3 -3 -3 22 -3 ...
$ PI1M_CLASS             : int  0 0 1 0 1 1 1 1 -9 0 ...
$ PI1F_CLASS             : int  -9 0 -9 -9 -9 -9 -9 -9 0 -9 ...
$ PI1M_EMPLY12           : int  -3 0 -3 -3 -3 1 1 -3 -9 -3 ...
$ PI1F_EMPLY12           : int  0 -3 -9 -9 -3 -9 -9 -9 -3 -9 ...
$ PI1M_JOB_TR            : int  0 0 1 0 0 0 0 0 -9 0 ...
$ PI1F_JOB_TR            : int  -9 0 -9 -9 -9 -9 -9 -9 0 -9 ...
$ PI1C_MALE              : int  1 0 1 0 1 0 0 1 0 1 ...
$ PI1CAGE_MTH            : num  19.9 31.5 29.2 31.6 26 34.5 29.8 26.1 9.8 25.2 ...
$ PI1CAGE_MTHC           : int  2 3 3 3 3 3 3 3 1 3 ...
$ PI1C_WEEKSEARLY        : int  2 2 -3 -3 4 -3 5 -3 -3 -3 ...
$ PI1CBPREM              : int  0 0 0 0 1 0 1 0 0 0 ...
$ PI1C_BIRTHWEIGHT       : num  7.81 4.31 8.81 -9 5.31 ...
$ PI1C_USBORN            : int  1 1 1 1 1 1 1 1 1 1 ...
$ PI1HH_NEMPLY           : int  1 1 1 1 2 0 1 1 1 1 ...
$ PI1HH_COMP             : int  3 4 4 4 4 3 3 5 4 6 ...
$ PI1HH_KIDS             : int  1 2 3 3 2 1 2 2 3 5 ...
$ PI1HH_ADULTS           : int  2 2 1 1 2 2 1 3 1 1 ...
$ PI1P1_NUMPAR           : int  2 2 1 1 2 1 2 1 1 1 ...
$ PI1HH_PRNT             : int  2 2 1 1 2 1 1 1 1 1 ...
```

```
 $ PI1MF_MAR              : int  0 1 1 0 0 0 0 0 0 0 ...
 $ PI1F_LV_IN             : int  1 1 0 0 1 0 0 0 1 0 ...
 $ PI1HH_IG               : int  0 0 0 0 0 0 0 0 0 0 ...
 $ PI1HHSTRUC             : int  2 1 4 4 2 4 3 4 5 4 ...
  [list output truncated]
```

```
data_cleaned<-data|>
  filter(PI1CESDRT >= 0 )|>
  filter(PR1BITSPR >=0)|>
  filter(PR1BITSCR>= 0)
```

problem behaviors

```
model1 <- lm(PR1BITSPR~PI1CESDRT, data = data_cleaned)
summary(model1)
```

```
Call:
lm(formula = PR1BITSPR ~ PI1CESDRT, data = data_cleaned)

Residuals:
    Min      1Q  Median      3Q     Max
-14.287  -4.857  -1.027   3.322  45.096

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.02746    0.18175  55.172  < 2e-16 ***
PI1CESDRT    0.16919    0.02106   8.034 1.67e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.59 on 1834 degrees of freedom
Multiple R-squared:  0.034, Adjusted R-squared:  0.03347
F-statistic: 64.54 on 1 and 1834 DF,  p-value: 1.674e-15
```

The model's slope coefficient is 0.16919, and the p-value is 1.67e-15. Since this p-value is smaller than the significance level of 0.05, it indicates that there is significant relationship between parental depression and the problem behaviors in children.

social competence

```
model3 <- lm(PR1BITSCR~PI1CESDRT, data = data_cleaned)
summary(model3)
```

```
Call:
lm(formula = PR1BITSCR ~ PI1CESDRT, data = data_cleaned)

Residuals:
     Min       1Q   Median       3Q      Max
-13.8396  -1.8051   0.4711   2.2390   4.9199

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 17.83963    0.08639 206.507  < 2e-16 ***
PI1CESDRT   -0.03453    0.01001  -3.449 0.000575 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.132 on 1834 degrees of freedom
Multiple R-squared:  0.006445,  Adjusted R-squared:  0.005903
F-statistic:  11.9 on 1 and 1834 DF,  p-value: 0.0005751
```

The model's slope coefficient is -0.035, and the p-value is 0.000575. Since this p-value is smaller than the significance level of 0.05, it indicates that there is significant relationship between parental depression and the social competence in children.

language skills

comprehensive volcabulary

```
model4 <- lm(Comprehension.Vocabular~PI1CESDRT, data = data_cleaned)
summary(model4)
```

```
Call:
lm(formula = Comprehension.Vocabular ~ PI1CESDRT, data = data_cleaned)
```

```
Residuals:
   Min     1Q Median     3Q    Max
-61.68 -23.68  10.38  32.32  40.78

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 61.67879    0.98690   62.50   <2e-16 ***
PI1CESDRT   -0.06066    0.11435   -0.53    0.596
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 35.78 on 1834 degrees of freedom
Multiple R-squared:  0.0001534, Adjusted R-squared:  -0.0003918
F-statistic: 0.2814 on 1 and 1834 DF,  p-value: 0.5959
```

The model's slope coefficient is -0.2151, and the p-value is 0.0522. Since this p-value is greater than the significance level of 0.05, it indicates that there is no significant relationship between parental depression and the comprehensive vocabulary development in children.

productive volcabulary

```r
model5 <- lm(Productive.Vocabulary~PI1CESDRT, data = data_cleaned)
summary(model5)
```

```
Call:
lm(formula = Productive.Vocabulary ~ PI1CESDRT, data = data_cleaned)

Residuals:
    Min      1Q  Median      3Q     Max
-43.015 -30.152  -6.368  28.271  64.160

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  43.0154     0.9068  47.435   <2e-16 ***
PI1CESDRT    -0.1434     0.1051  -1.365    0.172
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 32.88 on 1834 degrees of freedom
Multiple R-squared:  0.001015,  Adjusted R-squared:  0.0004702
F-statistic: 1.863 on 1 and 1834 DF,  p-value: 0.1724
```

The model's slope coefficient is -0.2084, and the p-value is 0.0267. Since this p-value is smaller than the significance level of 0.05, it indicates that there is significant relationship between parental depression and the productive vocabulary development in children.