descriptive_statistics

```r
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 becom
```

```r
library(dplyr)
```

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

The sample size for the full dataset after cleaning the observations which have missing value
is 1130

```r
data_cleaned <- data %>%
  select("PI1CESDRT", "Comprehension.Vocabular", "Productive.Vocabulary", "PR1BITSPR",
         "PR1BITSCR", "PR1HFPIPC", "PR1CPRCL", "PI1FCONF", "PI1CHAOS", "PR1HFPISS",
         "PR1PSITS", "PR1PSPCDI", "PI1CESDRM") %>%
  filter(if_all(everything(), ~ . >= 0))
```

descriptive statistics

full data

```
  summary(data_cleaned)
```

```
   PI1CESDRT        Comprehension.Vocabular Productive.Vocabulary    PR1BITSPR
 Min.   : 0.000   Min.   :  0.00          Min.   :  0.00         Min.   : 0.00
 1st Qu.: 0.000   1st Qu.: 37.00          1st Qu.: 12.00         1st Qu.: 6.00
 Median : 2.000   Median : 71.00          Median : 35.00         Median :10.00
 Mean   : 4.404   Mean   : 61.05          Mean   : 41.15         Mean   :10.82
 3rd Qu.: 6.000   3rd Qu.: 92.00          3rd Qu.: 67.75         3rd Qu.:14.00
 Max.   :52.000   Max.   :100.00          Max.   :100.00         Max.   :51.00
   PR1BITSCR        PR1HFPIPC        PR1CPRCL         PI1FCONF         PI1CHAOS
 Min.   : 4.00    Min.   :18.00    Min.   : 7.00    Min.   :1.00    Min.   : 0.00
 1st Qu.:16.00    1st Qu.:43.00    1st Qu.:29.00    1st Qu.:1.00    1st Qu.: 6.00
 Median :18.00    Median :46.00    Median :32.00    Median :1.40    Median :10.00
 Mean   :17.72    Mean   :45.59    Mean   :31.12    Mean   :1.43    Mean   :10.36
 3rd Qu.:20.00    3rd Qu.:49.00    3rd Qu.:34.00    3rd Qu.:1.60    3rd Qu.:14.00
 Max.   :22.00    Max.   :50.00    Max.   :35.00    Max.   :3.60    Max.   :41.00
   PR1HFPISS        PR1PSITS         PR1PSPCDI        PI1CESDRM
 Min.   : 5.00    Min.   : 36.00   Min.   :12.00    Min.   :0.00000
 1st Qu.:19.00    1st Qu.: 42.00   1st Qu.:13.00    1st Qu.:0.00000
 Median :23.00    Median : 56.00   Median :16.00    Median :0.00000
 Mean   :21.29    Mean   : 58.47   Mean   :17.97    Mean   :0.01062
 3rd Qu.:25.00    3rd Qu.: 72.00   3rd Qu.:22.00    3rd Qu.:0.00000
 Max.   :25.00    Max.   :145.00   Max.   :43.00    Max.   :1.00000
```

standard deviation

```
  lapply(data_cleaned[, 1:13], sd)
```

```
$PI1CESDRT
[1] 7.199452

$Comprehension.Vocabular
[1] 35.28845

$Productive.Vocabulary
```

```
[1] 32.22861
```

```
$PR1BITSPR
[1] 6.64009
```

```
$PR1BITSCR
[1] 3.178269
```

```
$PR1HFPIPC
[1] 4.554516
```

```
$PR1CPRCL
[1] 3.933821
```

```
$PI1FCONF
[1] 0.3905519
```

```
$PI1CHAOS
[1] 5.936798
```

```
$PR1HFPISS
[1] 4.518988
```

```
$PR1PSITS
[1] 17.76198
```

```
$PR1PSPCDI
[1] 5.683715
```

```
$PI1CESDRM
[1] 0.1025476
```

group by depression

0-no depression 1-depression

the sample size for depression is 12 and the sample size for parents without depression is 1118

```
by(data_cleaned, data_cleaned$PI1CESDRM, summary)
```

```
data_cleaned$PI1CESDRM: 0
```

```
     PI1CESDRT        Comprehension.Vocabular Productive.Vocabulary   PR1BITSPR
 Min.   : 0.000    Min.   :  0.00      Min.   :  0.00       Min.   : 0.00
 1st Qu.: 0.000    1st Qu.: 37.00      1st Qu.: 12.00       1st Qu.: 6.00
 Median : 1.000    Median : 71.00      Median : 35.00       Median :10.00
 Mean   : 4.025    Mean   : 61.09      Mean   : 41.16       Mean   :10.77
 3rd Qu.: 5.000    3rd Qu.: 92.00      3rd Qu.: 68.00       3rd Qu.:14.00
 Max.   :44.000    Max.   :100.00      Max.   :100.00       Max.   :51.00
    PR1BITSCR          PR1HFPIPC          PR1CPRCL          PI1FCONF
 Min.   : 4.00     Min.   :18.00     Min.   : 7.00     Min.   :1.000
 1st Qu.:16.00     1st Qu.:43.00     1st Qu.:29.00     1st Qu.:1.000
 Median :18.00     Median :46.00     Median :32.00     Median :1.400
 Mean   :17.75     Mean   :45.63     Mean   :31.13     Mean   :1.428
 3rd Qu.:20.00     3rd Qu.:49.00     3rd Qu.:34.00     3rd Qu.:1.600
 Max.   :22.00     Max.   :50.00     Max.   :35.00     Max.   :3.600
     PI1CHAOS          PR1HFPISS          PR1PSITS          PR1PSPCDI         PI1CESDRM
 Min.   : 0.00     Min.   : 5.00     Min.   : 36.00    Min.   :12.00     Min.   :0
 1st Qu.: 6.00     1st Qu.:19.00     1st Qu.: 42.00    1st Qu.:13.00     1st Qu.:0
 Median :10.00     Median :23.00     Median : 56.00    Median :16.00     Median :0
 Mean   :10.29     Mean   :21.36     Mean   : 58.11    Mean   :17.87     Mean   :0
 3rd Qu.:14.00     3rd Qu.:25.00     3rd Qu.: 71.00    3rd Qu.:22.00     3rd Qu.:0
 Max.   :34.00     Max.   :25.00     Max.   :125.00    Max.   :41.00     Max.   :0
------------------------------------------------------------
data_cleaned$PI1CESDRM: 1
     PI1CESDRT        Comprehension.Vocabular Productive.Vocabulary   PR1BITSPR
 Min.   :24.00     Min.   :  0.0       Min.   :  6.00       Min.   : 8.00
 1st Qu.:32.50     1st Qu.:42.0        1st Qu.:15.25        1st Qu.: 9.75
 Median :39.50     Median :60.5        Median :37.00        Median :13.73
 Mean   :39.67     Mean   :57.5        Mean   :39.92        Mean   :15.59
 3rd Qu.:46.75     3rd Qu.:81.5        3rd Qu.:59.75        3rd Qu.:21.20
 Max.   :52.00     Max.   :98.0        Max.   :98.00        Max.   :28.00
    PR1BITSCR          PR1HFPIPC          PR1CPRCL          PI1FCONF
 Min.   : 8.80     Min.   :29.00     Min.   :19.00     Min.   :1.000
 1st Qu.:11.75     1st Qu.:39.25     1st Qu.:25.75     1st Qu.:1.400
 Median :15.50     Median :44.00     Median :31.00     Median :1.600
 Mean   :15.32     Mean   :42.15     Mean   :29.33     Mean   :1.617
 3rd Qu.:18.50     3rd Qu.:47.84     3rd Qu.:34.25     3rd Qu.:1.850
 Max.   :21.00     Max.   :50.00     Max.   :35.00     Max.   :2.200
     PI1CHAOS          PR1HFPISS          PR1PSITS          PR1PSPCDI         PI1CESDRM
 Min.   : 5.00     Min.   : 6.00     Min.   : 52.00    Min.   :14.00     Min.   :1
 1st Qu.: 9.00     1st Qu.: 9.75     1st Qu.: 76.25    1st Qu.:17.50     1st Qu.:1
 Median :17.50     Median :12.00     Median : 92.00    Median :25.00     Median :1
 Mean   :17.42     Mean   :14.25     Mean   : 92.58    Mean   :26.83     Mean   :1
 3rd Qu.:18.50     3rd Qu.:19.25     3rd Qu.:104.00    3rd Qu.:35.00     3rd Qu.:1
```

```
Max.   :41.00    Max.    :25.00    Max.    :145.00    Max.    :43.00    Max.    :1
```

parents with depression-sd

```r
data_cleaned_depression <- data_cleaned|>
  filter(PI1CESDRM == 1)
lapply(data_cleaned_depression[, 1:13], sd)
```

```
$PI1CESDRT
[1] 8.947659

$Comprehension.Vocabular
[1] 32.64269

$Productive.Vocabulary
[1] 30.81752

$PR1BITSPR
[1] 7.209265

$PR1BITSCR
[1] 4.036613

$PR1HFPIPC
[1] 6.961873

$PR1CPRCL
[1] 5.836147

$PI1FCONF
[1] 0.376185

$PI1CHAOS
[1] 10.49206

$PR1HFPISS
[1] 6.22495

$PR1PSITS
[1] 28.44279
```

$PR1PSPCDI
[1] 10.76892

$PI1CESDRM
[1] 0

parents without depression-sd

```r
data_cleaned_nodepression <- data_cleaned|>
  filter(PI1CESDRM == 0)
lapply(data_cleaned_nodepression[, 1:13], sd)
```

$PI1CESDRT
[1] 6.172389

$Comprehension.Vocabular
[1] 35.32736

$Productive.Vocabulary
[1] 32.25636

$PR1BITSPR
[1] 6.618589

$PR1BITSCR
[1] 3.160144

$PR1HFPIPC
[1] 4.512276

$PR1CPRCL
[1] 3.90785

$PI1FCONF
[1] 0.3903837

$PI1CHAOS
[1] 5.830968

$PR1HFPISS
[1] 4.440887

```
$PR1PSITS
[1] 17.27069

$PR1PSPCDI
[1] 5.536824

$PI1CESDRM
[1] 0
```