Data Management and Visualization

Week 4 assignment

Creating graphs

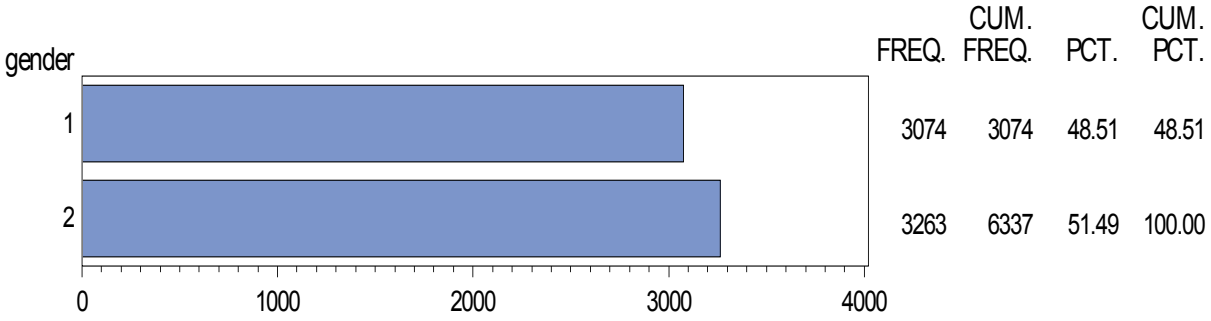
I am going to show the relationships between the following variables respectively;

1. gender (BIO_SEX) and how the respondents are close to their mother (H1WP9);

2. general health (H1GH1) and body mass index (BMI), which is categorical to quantitative;

3. height (H1GH59) and weight (H1GH60), which is quantitative to quantitative.

First of all, gender and closeness with mother was analyzed.



	FREQ.	CUM. FREQ.	PCT.	CUM. PCT.
1	3074	3074	48.51	48.51
2	3263	6337	51.49	100.00

Figure 1 bar plot of gender

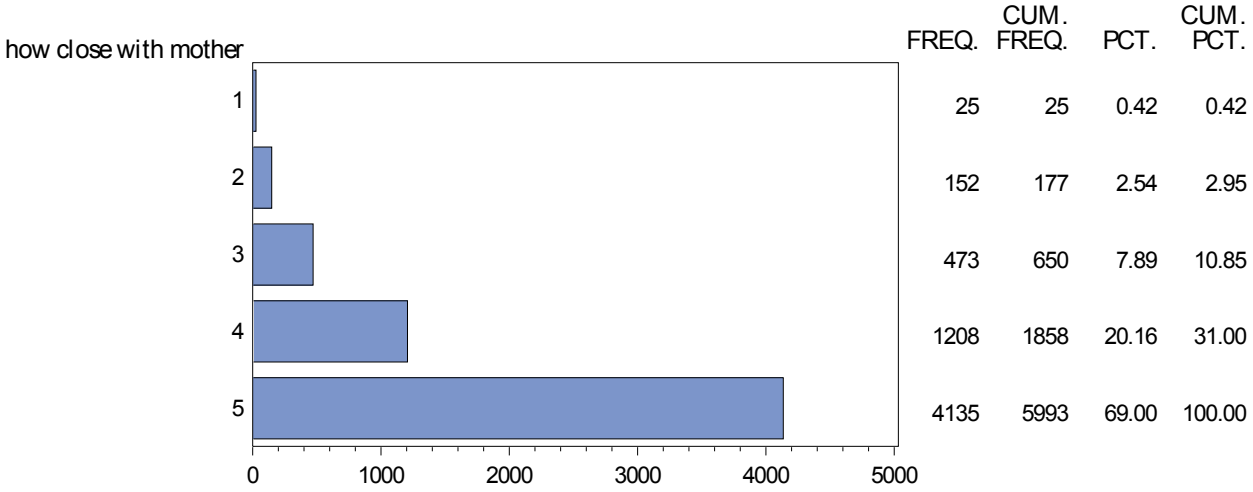


Figure 2 bar plot of closeness with mother

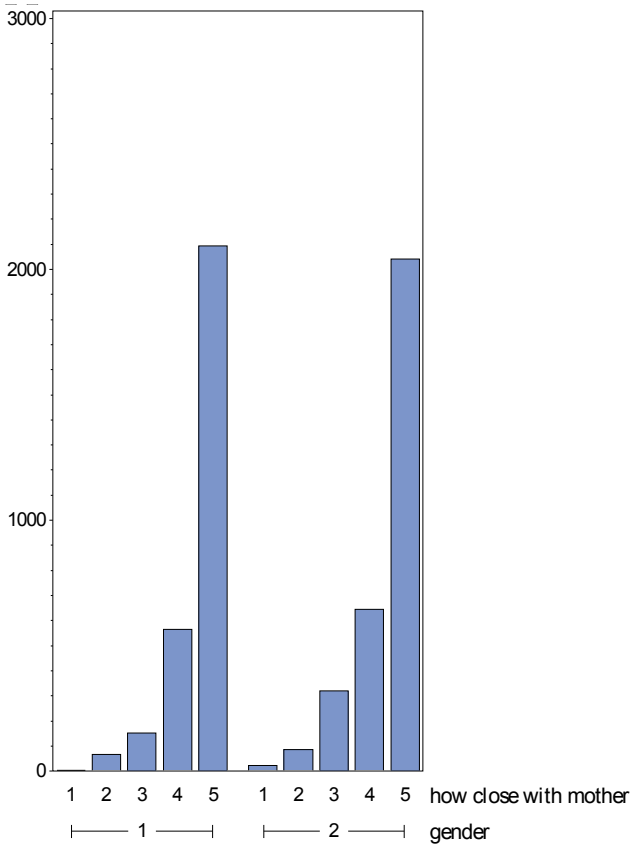
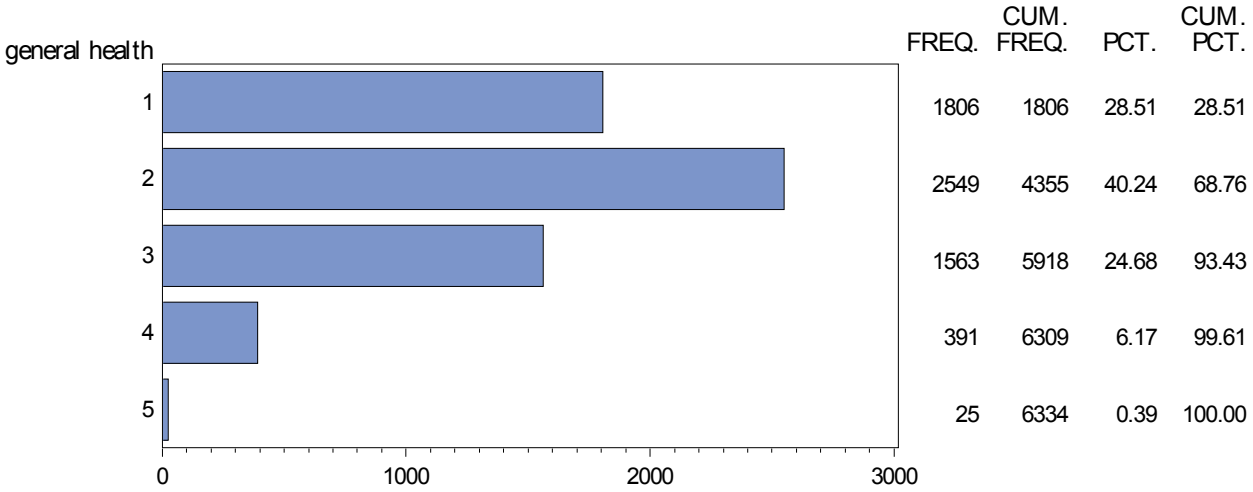


Figure 3 relationship between gender and closeness with mother

Figure 3 shows that the gender has no significant effect on the closeness with mother.

Because both gender and closeness with mother are categorical variables, bar plot is used to show their relationship.

Second, general health and BMI were analyzed.



	FREQ.	CUM. FREQ.	PCT.	CUM. PCT.
1	1806	1806	28.51	28.51
2	2549	4355	40.24	68.76
3	1563	5918	24.68	93.43
4	391	6309	6.17	99.61
5	25	6334	0.39	100.00

Figure 4 bar plot of general health

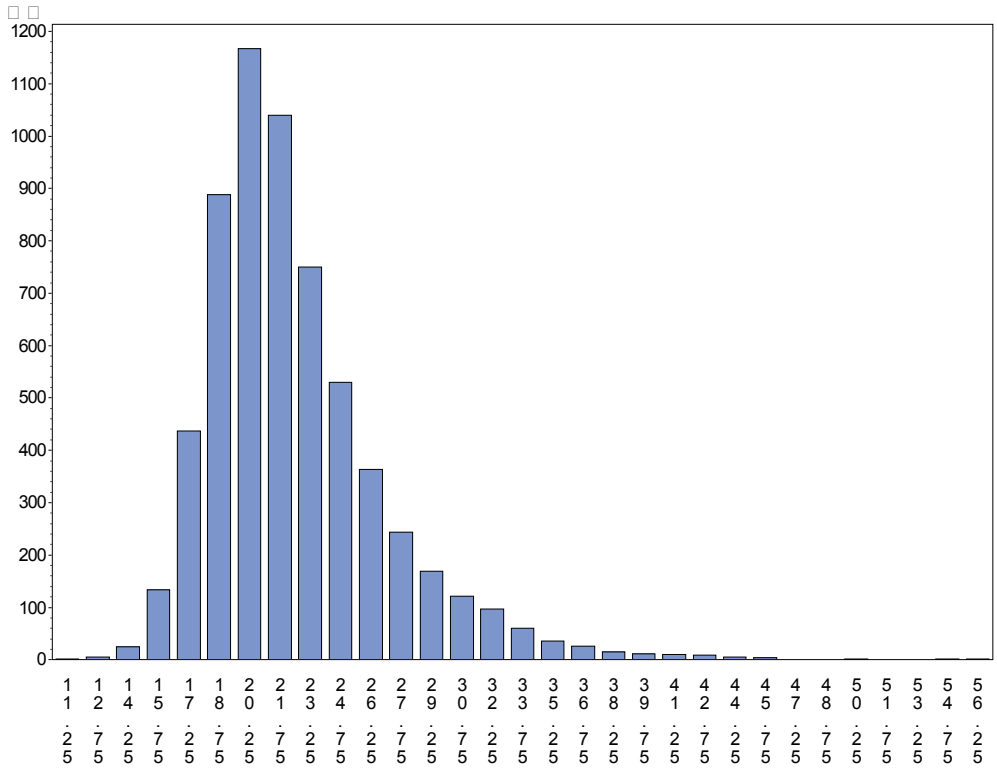


Figure 5 bar plot of BMI

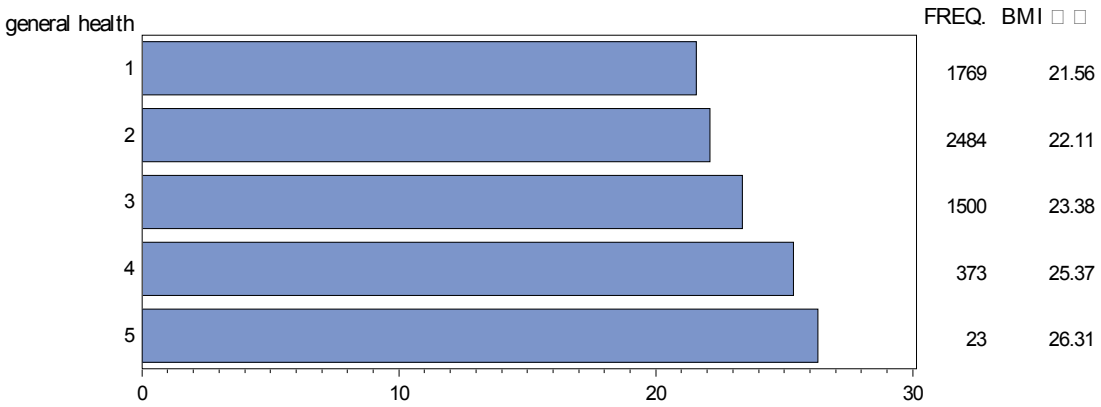


Figure 6 relationship between general health and BMI

Figure 6 shows that people are healthier with smaller BMI than people with larger BMI.

Because general health is categorical variable and BMI is quantitative variable, bar plot is used to show their relationship.

Finally, weight and height were analyzed.

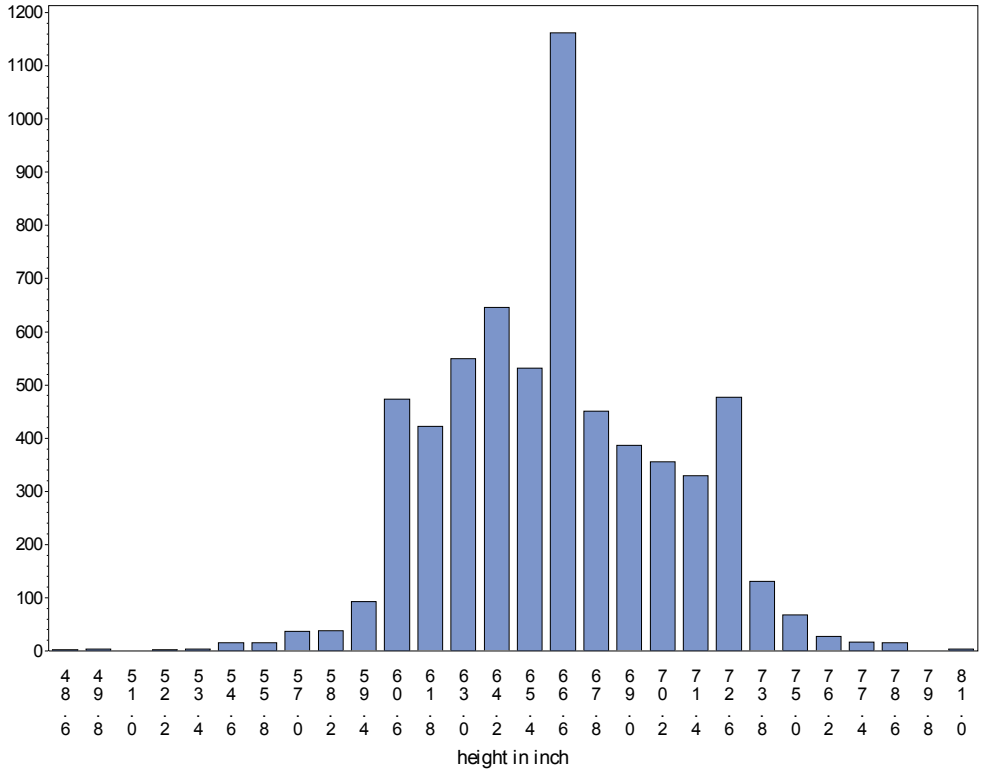


Figure 7 bar plot of height

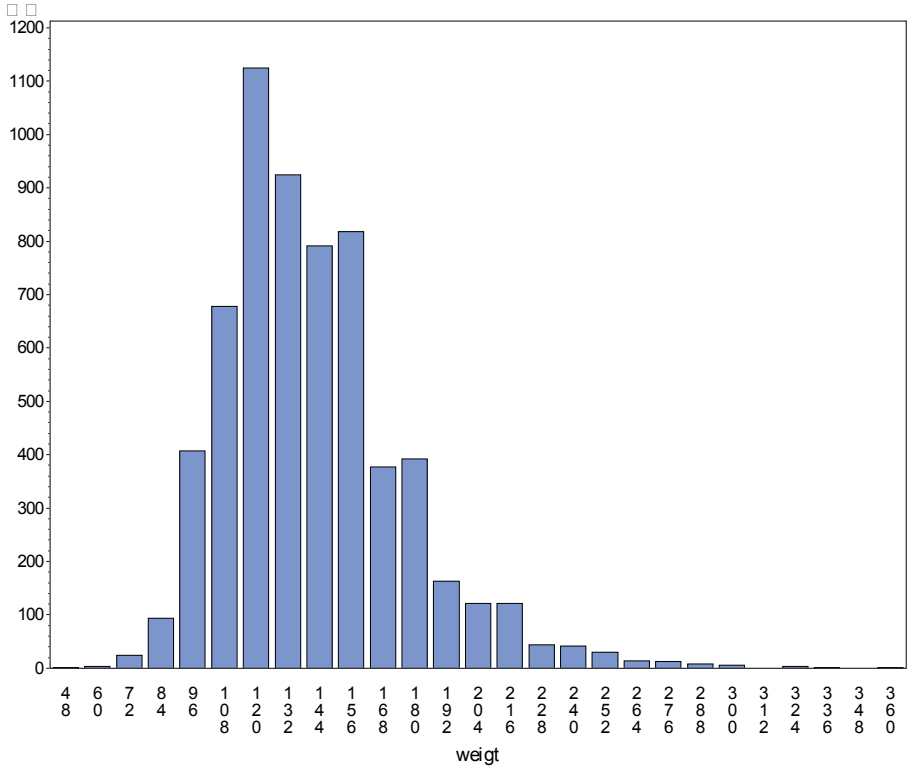


Figure 8 bar plot of weight

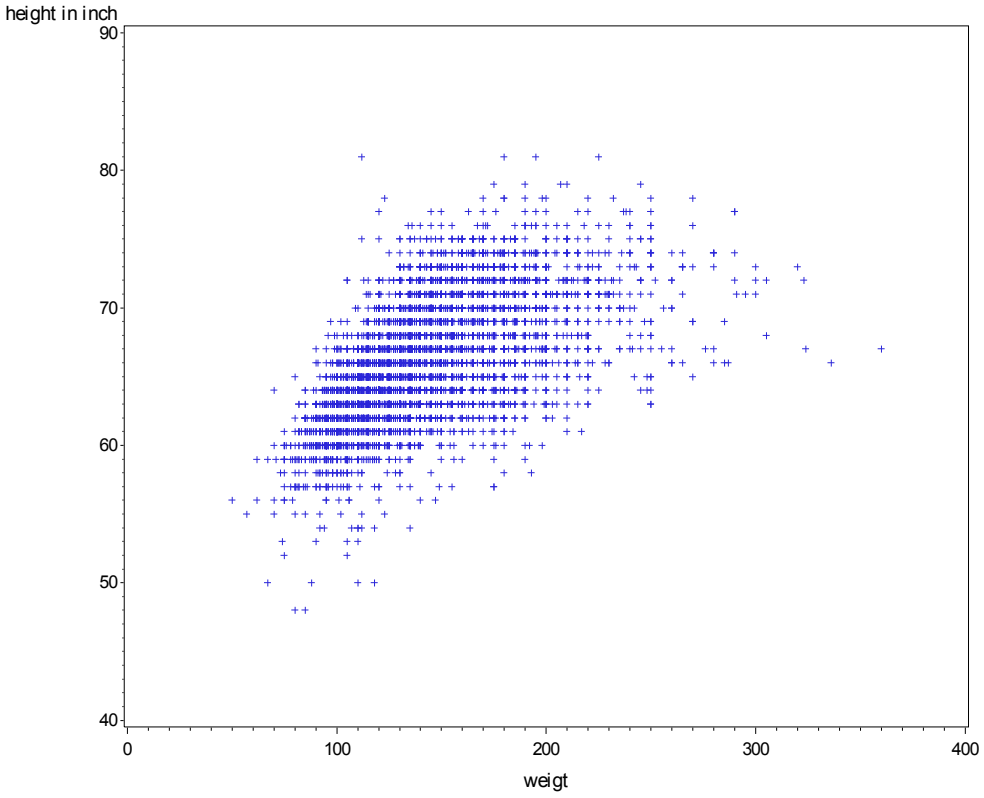


Figure 9 relationship between weight and height

Figure 9 shows that as the weight gets larger, the height also gets larger.

Because both of height and weight are quantitative variables, scatter plot is used to show their relationship.

My code:

```
LIBNAME mydata "/courses/d1406ae5ba27fe300" access=readonly;
data new; set mydata.addhealth_pds;

/*choose the respondent in grade from 7 to 12*/
if H1GI20=97 then delete; if H1GI20=99 then delete; if H1GI20=96 then delete;
if H1GI20=98 then delete;

/*set aside mising value*/
if H1GH1=6 then H1GH1=.; if H1GH1=8 then H1GH1=.;
if H1GH6=6 then H1GH6=.; if H1GH6=8 then H1GH6=.;
if H1GH59A=96 then H1GH59A=.; if H1GH59A=98 then H1GH59A=.; if H1GH59A=99 then H1GH59A=.;
if H1GH59B=96 then H1GH59B=.; if H1GH59B=98 then H1GH59B=.; if H1GH59B=99 then H1GH59B=.;
if H1GH60=996 then H1GH60=.; if H1GH60=998 then H1GH60=.; if H1GH60=999 then H1GH60=.;
if H1WP9=6 then H1WP9=.; if H1WP9=7 then H1WP9=.;
if H1WP9=8 then H1WP9=.; if H1WP9=9 then H1WP9=.;

/*calculate the height*/
H1GH59=H1GH59A * 12 + H1GH59B;

/*calculate the body mass index and collapse the index*/
BMI=H1GH60 * 0.454/(H1GH59 * 0.0254)**2;
if BMI le 15 then bmigroup=1;
else if BMI le 35 then bmigroup=2;
else if BMI le 55 then bmigroup=3;
else bmigroup=4;

/*add label*/
label AID="respondent ID"
      BIO_SEX="gender"
      H1GH1="general health"
      H1GH60="weigt"
      H1WP9="how close with mother"
      H1GH59="height in inch";

/*collapse the height*/
if H1GH59 le 60 then heightgroup=1;
else if H1GH59 le 70 then heightgroup=2;
else if H1GH59 le 80 then heightgroup=3;
else heightgroup=4;

/*collapse the weight*/
if H1GH60 le 100 then weightgroup=1;
else if H1GH60 le 200 then weightgroup=2;
else if H1GH60 le 300 then weightgroup=3;
else weightgroup=4;

/*select variables of interest*/
```

```
keep AID BIO_SEX H1GH1 H1GH60 H1WP9 H1GH59 BMI heightgroup weightgroup bmigroup;
proc sort; by AID;

/*show frequency table of each variable*/
proc freq; tables BIO_SEX H1GH1 bmigroup weightgroup heightgroup H1WP9;
run;

/*show the relationship between gender and how close the respondents are to their mother*/
proc gchart; hbar BIO_SEX H1WP9/discrete;
proc gchart; vbar H1WP9/discrete group=BIO_SEX;

/*show the relationship between general health and body mass index*/
proc gchart; hbar H1GH1/discrete;
proc gchart; vbar BMI;
proc gchart; hbar H1GH1/discrete type=mean sumvar=BMI;

/*show the relationship between height and weight*/
proc gchart; vbar H1GH59 H1GH60;
proc gplot; plot H1GH59*H1GH60;
```