Data Management and Visualization

Week 3 assignment

Data management

Introduction:

Data management I did:

1. collapsing the responses for grade, creating a new variable gradegroup to represent the low grade (7-8), middle grade (9-10) and high grade (11-12);

2. setting aside all the missing values of every variable;

3. creating a new variable named H1GH59 (labeled as height in inch) to represent the height of each respondent;

4. subsetting all the variables I am going to use in future research;

5. showing the frequency distribution of each variable.

According to the frequency distribution, the gender and grade group are evenly distributed.

More than half respondents thought their health was excellent or very good. No more than 5% of them felt headache or weak every day. No more than half of them always felt happy. More respondents felt close to mother than father.

My code:

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

/*data management for grade: select the 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;
/*data management for grade: collapse*/
if H1GI20 le 8 then gradegroup=1;
else if H1GI20 le 10 then gradegroup=2;
else gradegroup=3;

/*data management for other variables: set missing value*/
if H1GH1=6 then H1GH1=.; if H1GH1=8 then H1GH1=.;
if H1GH2=6 then H1GH2=.; if H1GH2=8 then H1GH2=.;
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 H1GH59A=.; if H1GH59B=98 then H1GH59A=.; if H1GH59B=99 then H1GH59A=.;
if H1GH60=996 then H1GH60=.; if H1GH60=998 then H1GH60=.; if H1GH60=999 then H1GH60=.;
if H1FS11=6 then H1FS11=.; if H1FS11=8 then H1FS11=.;
if H1WP9=6 then H1WP9=.; if H1WP9=7 then H1WP9=.;
if H1WP9=8 then H1WP9=.; if H1WP9=9 then H1WP9=.;
if H1WP13=6 then H1WP13=.; if H1WP13=7 then H1WP13=.;
if H1WP13=8 then H1WP13=.; if H1WP13=9 then H1WP13=.;

/*data management for height: create a new variable to show the height*/
H1GH59=H1GH59A * 12 + H1GH59B;

/*subset the interesting variables*/
keep AID BIO_SEX gradegroup H1GH1 H1GH2 H1GH6
     H1GH59 H1GH60 H1FS11 H1WP9 H1WP13;

label AID="respondent ID"
      BIO_SEX="gender"
      gradegroup="grade group"
      H1GH1="general health"
      H1GH2="frequency of headache"
      H1GH6="frequnecy of feeling weak"
      H1GH60="weigt"
      H1FS11="feeling happy"
      H1WP9="how close with mother"
      H1WP13="how close with father"
      H1GH59="height in inch";
proc sort; by AID;/*sorted by AID*/
proc freq; tables BIO_SEX gradegroup H1GH1 H1GH2 H1GH6 H1FS11 H1WP9 H1WP13;
run;
```

Output results:

gender				
BIO_SEX	Frequency	Percent	Cumulative frequency	Cumulative percent
1	3074	48.51	3074	48.51
2	3263	51.49	6337	100.00

grade group				
gradegroup	Frequency	Percent	Cumulative frequency	Cumulative percent
1	1971	31.10	1971	31.10
2	2251	35.52	4222	66.62
3	2115	33.38	6337	100.00

general health				
H1GH1	Frequency	Percent	Cumulative frequency	Cumulative percent
1	1806	28.51	1806	28.51
2	2549	40.24	4355	68.76
3	1563	24.68	5918	93.43
4	391	6.17	6309	99.61
5	25	0.39	6334	100.00
Frequency missing = 3				

frequency of headache				
H1GH2	Frequency	Percent	Cumulative frequency	Cumulative percent
0	555	8.76	555	8.76
1	3924	61.94	4479	70.70
2	1408	22.23	5887	92.93
3	378	5.97	6265	98.90
4	70	1.10	6335	100.00
Frequency missing = 2				

frequnecy of feeling weak				
H1GH6	Frequency	Percent	Cumulative frequency	Cumulative percent
0	3474	54.86	3474	54.86
1	2314	36.54	5788	91.41
2	411	6.49	6199	97.90
3	100	1.58	6299	99.48
4	33	0.52	6332	100.00
Frequency missing = 5				

feeling happy				
H1FS11	Frequency	Percent	Cumulative frequency	Cumulative percent
0	164	2.59	164	2.59
1	1188	18.77	1352	21.36
2	2637	41.67	3989	63.03
3	2340	36.97	6329	100.00
Frequency missing = 8				

how close with mother				
H1WP9	Frequency	Percent	Cumulative frequency	Cumulative percent
1	25	0.42	25	0.42
2	152	2.54	177	2.95
3	473	7.89	650	10.85
4	1208	20.16	1858	31.00
5	4135	69.00	5993	100.00
Frequency missing = 344				

how close with father				
H1WP13	Frequency	Percent	Cumulative frequency	Cumulative percent
1	70	1.57	70	1.57
2	178	3.99	248	5.56
3	595	13.34	843	18.90
4	1188	26.64	2031	45.54
5	2429	54.46	4460	100.00
Frequency missing = 1877				