

Story Telling with Data:

You will be doing research based on the first wave of the **Adolescent Health Study** (ADDHEALTH) which includes a representative U.S. sample of adolescents in grades 7 through 12.

Following each step, you should add new text to your poster.

STEP 1: Reading the code book

Read through the **Adolescent Health Study (Wave 1)** code book.

STEP 2: Choosing a research question

Based on your reading of the codebook, choose a specific **topic of interest** and then a **second topic** you believe is related to the first topic. Prepare a short codebook of your own.

For example, I am interested in **gender (topic 1)**

1. Respondent's biological gender			BIO_SEX	num 1
3147	1	male		
3356	2	female		
1	6	refused		

and how it is related to being **older relative to one's grade level (topic 2).**

4. Current age of respondent			AGE	num 2
4837	11.68 to 20.93	(calculated from date of birth)		
1667	.	missing		

20. What grade {ARE/WERE} you in? [If school doesn't have grade levels of this kind, enter "99."]			H1GI20	num 2
979	7	seventh grade		
992	8	eighth grade		
1107	9	ninth grade		
1144	10	tenth grade		
1122	11	eleventh grade		
993	12	twelfth grade		
1	96	refused		
128	97	legitimate skip [not in school]		
3	98	don't know		
35	99	school doesn't have grade levels of this kind or not applicable		

STEP 3: Examining Data: Writing your first program

Write a basic program that tells SAS where to get the data set (**map icon**), calls in the data set so that you can use it (**arrow icon**), sorts the data (**car sort icon**), requests results (**i information icon**) and ends the program (**run icon**). Add **variable names** that you have chosen to the **request for results** statements (**yellow highlight**)

Example program



```
libname mydata "/courses/d1406ae5ba27fe300" access=readonly;
```



```
data new; set mydata.addhealth_teen;
```



```
proc sort; by aid;
```



```
proc freq; tables BIO_SEX AGE H1GI20;
```



```
run;
```

STEP 4: Managing Data

Add **data management statements** to your program (**puzzle icon**). This will include **setting aside missing data** for each of our chosen variables and one or more of the following a) creating a new variable by collapsing response categories; b) creating a new variable by aggregating across more than one variable; and c) labeling your variables.

```
libname mydata "/courses/d1406ae5ba27fe300" access=readonly;
```

```
data new; set mydata.addhealth_teen;
```



```
/*Setting aside missing data*/
if BIO_SEX=6 then BIO_SEX=.;
if H1GI20 GE 96 then H1GI20=.;

/*creating a new variable by collapsing response categories*/
if AGE EQ . then AGEGROUP=.;
else if AGE LE 16.51 then AGEGROUP=1;
else if AGE GT 16.51 then AGEGROUP=2;

/*Creating a new variable by aggregating more than one variable*/
if H1GI20 EQ . or AGE eq . then OLDER4GRADE=.;
else if (H1GI20=7 and AGE GE 14) or
(H1GI20=8 and AGE GE 15) or (H1GI20=9 and AGE GE 16) or (H1GI20=10 and AGE GE 17)
or (H1GI20=11 and AGE GE 18) or (H1GI20=12 and AGE GE 19) then OLDER4GRADE=1;
else OLDER4GRADE=0;

/*labeling variables*/
label AGEGROUP='age group'
      BIO_SEX='gender'
      H1GI20='grade level'
      OLDER4GRADE='older for grade';
```

```
proc sort; by AID;
```

```
proc freq; tables AGE BIO_SEX H1GI20 AGEGROUP OLDER4GRADE;
```

```
proc means; var AGE H1GI20;
```

```
run;
```

Note: use **PROC MEANS** only to examine **quantitative variables** that have meaningful number values including decimals (e.g. age, height, weight) rather than discrete number values that include only whole number (e.g. grade level = 7, 8, 9, 10, 11, 12).

STEP 5: Univariate Graphing

Display and describe univariate (i.e. **one variable**) graphs for each of your newly data managed variables.

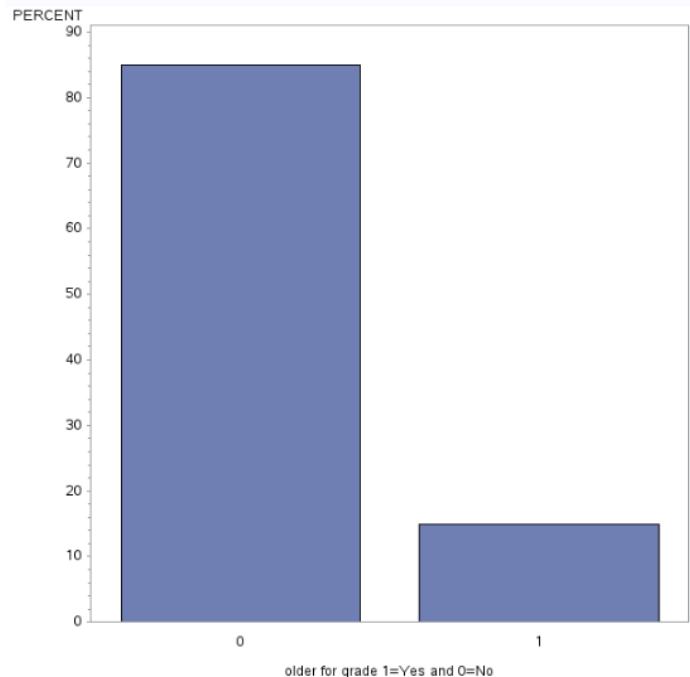
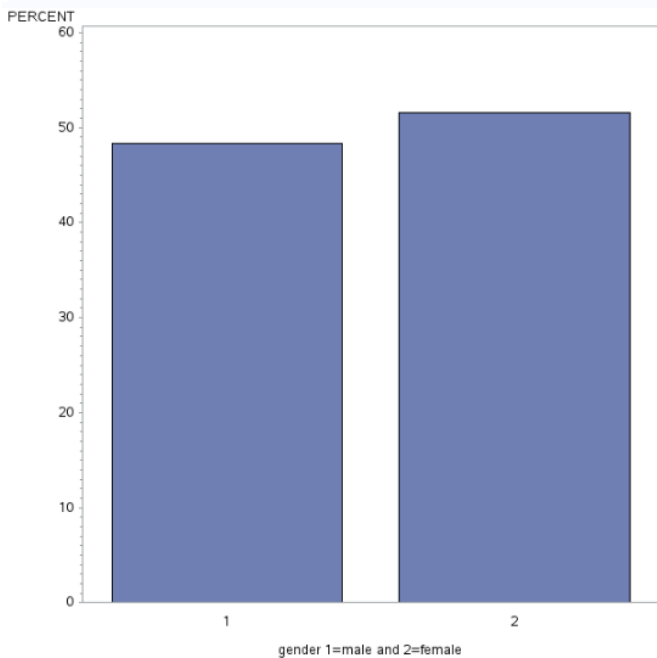
For **quantitative variables**, you will be examining shape, center and spread of the graph. For **categorical variables**, you will be examining the number or percent of observations in each category.

Categorical variable

```
Proc freq; tables BIO_SEX OLDER4GRADE;
Proc gchart; vbar BIO_SEX OLDER4GRADE / Discrete type=PCT
width=30;
```

gender 1=male and 2=female				
BIO_SEX	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	3147	48.39	3147	48.39
2	3356	51.61	6503	100.00
Frequency Missing = 1				

older for grade 1=Yes and 0=No				
OLDER4GRADE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	4025	84.99	4025	84.99
1	711	15.01	4736	100.00
Frequency Missing = 1768				

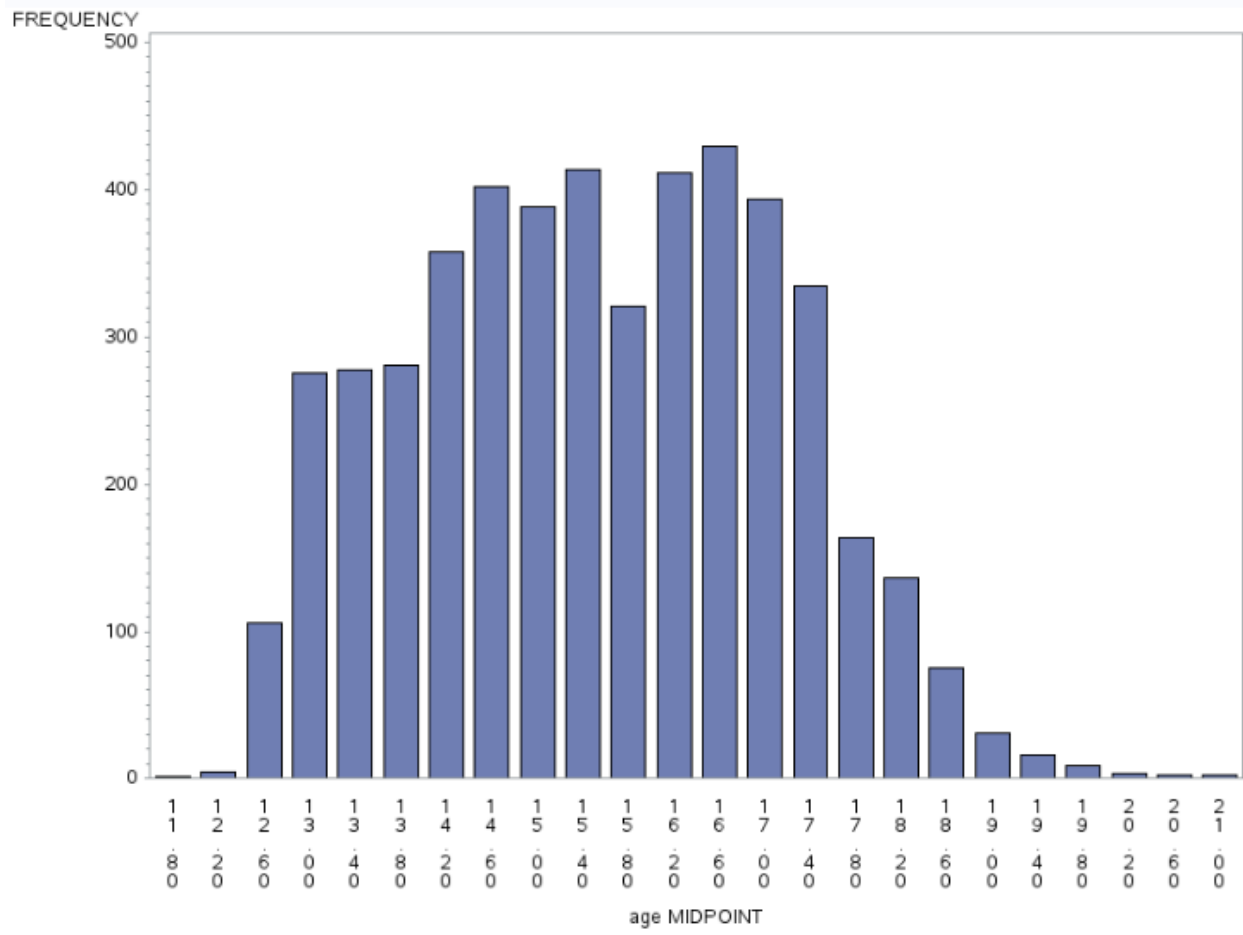


Quantitative variable

```
proc means; var AGE;
proc gchart; vbar AGE;
```

The MEANS Procedure

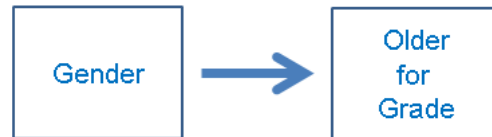
Analysis Variable : age				
N	Mean	Std Dev	Minimum	Maximum
4837	15.5415482	1.5819929	11.6767123	20.9342466



STEP 6: Examining one variable by another variable (an association)

Display and describe the association between your two topics (i.e. two data managed variables)

```
proc sort; by BIO_SEX;
proc freq; tables OLDER4GRADE;
  where BIO_SEX ne .;
  by BIO_SEX;
```



This code requests a table for OLDER4AGE and splits the sample by BIO_SEX (i.e. gender) creating two tables, one for males and one for females.

gender=1 Male

older for grade				
OLDER4GRADE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	1848	81.27	1848	81.27
YES 1	426	18.73	2274	100.00
Frequency Missing = 873				

The FREQ Procedure

gender=2 Female

older for grade				
OLDER4GRADE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	2177	88.42	2177	88.42
YES 1	285	11.58	2462	100.00
Frequency Missing = 894				

Males are more likely to be older for their grade than females. A total of 18.7% of males are older for their grade compared to only 11.6% of females (Note: we round to one decimal place).

STEP 7: Choose a third topic

Based on your reading of the codebook, choose a **third topic** that may be related to the first two topics. Add the variable to your short codebook..

Though I am interested in whether males or females (**topic 1**) are more likely to be older for their grade (**topic 2**), I also wonder whether being older for your grade is related to **feeling socially accepted (topic 3)** and whether the answer to that question is different for males and females.

35. You feel socially accepted.		H1PF35	num 1
1857	1	strongly agree	
3674	2	agree	
667	3	neither agree nor disagree	
241	4	disagree	
40	5	strongly disagree	
10	6	refused	
15	8	don't know	

STEP 8: Manage data for your third topic

Add **data management** to your program. This will include **setting aside missing data** for each of our chosen variables and one or more of the following a) creating a new variable by collapsing response categories; b) creating a new variable by aggregating across more than one variable; and c) labeling your variables. After the PROC SORT statement, **ask for a frequency table for your new variable**.

New code is highlighted in yellow.

```

/*setting aside missing data*/
if BIO_SEX=6 then BIO_SEX=.;
if H1GI20 GE 96 then H1GI20=.;
if H1PF35 GE 6 then H1PF35=.;

/*creating a new variable by collapsing response categories*/
if AGE EQ . then AGEGROUP=.;
else if AGE LE 15.51 then AGEGROUP=1;
else if AGE GT 15.51 then AGEGROUP=2;

/*Creating a new variable 'OLDER4GRADE' by aggregating across more than
one variable*/
if H1GI20 EQ . or AGE EQ . then OLDER4GRADE=.;
else if (H1GI20=7 and AGE GE 14) or (H1GI20=8 and AGE GE 15) or (H1GI20=9
and AGE GE 16) or
(H1GI20=10 and AGE GE 17) or (H1GI20=11 and AGE GE 18) or (H1GI20=12 and
AGE GE 19) then OLDER4GRADE=1;
else OLDER4GRADE=0;

/*third topic - feeling socially accepted*/
if H1PF35 =. then FEELSOCACCEPT=.;
else if h1PF35 le 2 then FEELSOCACCEPT=1;
else if h1PF35 le 5 then FEELSOCACCEPT=0;

/*labeling variables*/
Label  AGEGROUP='age group'
       BIO_SEX='gender'
       H1GI20='grade level'
       OLDER4GRADE='older for grade'
       FEELSOCACCEPT='feel socially accepted';
proc sort; by AID;

proc freq; tables FEELSOCACCEPT;
run;

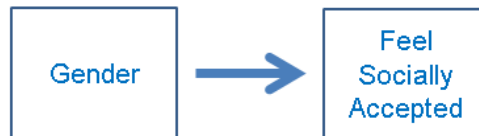
```

STEP 9: Examine the third variable by each of your first two topics

Display and describe the associations among your three topics.

1. Are males or females more likely to feel socially accepted?

```
proc sort; by BIO_SEX;
proc freq; tables FEELSOCACCEPT;
where BIO_SEX ne .;
by BIO_SEX;
```



The FREQ Procedure

gender=1 Male

feel socially accepted

FEELSOCACCEPT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	396	12.64	396	12.64
YES 1	2737	87.36	3133	100.00
Frequency Missing = 14				

The FREQ Procedure

gender=2 Female

feel socially accepted

FEELSOCACCEPT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	552	16.50	552	16.50
YES 1	2794	83.50	3346	100.00
Frequency Missing = 10				

Males are only slightly more likely to feel socially accepted (87.4%) than females (83%).

2. Are those who are older for their grade compared to those who are not older for their grade more likely to feel socially accepted?

```
proc sort; by OLDER4GRADE;
proc freq; tables FEELSOCACCEPT;
where OLDER4GRADE ne .;
by OLDER4GRADE;
```



older for grade=0 **Not older for grade**

feel socially accepted				
FEELSOCACCEPT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	559	13.91	559	13.91
YES 1	3459	86.09	4018	100.00
Frequency Missing = 7				

The FREQ Procedure

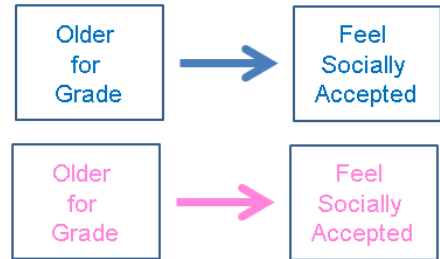
older for grade=1 **Older for grade**

feel socially accepted				
FEELSOCACCEPT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	125	17.76	125	17.76
YES 1	579	82.24	704	100.00
Frequency Missing = 7				

Adolescents who are older for their grade are slightly less likely to feel socially accepted (82.2%) than adolescents who are not older for their grade (86.1%).

Is the relationship between being older for your grade and feeling socially accepted different for males and females?

```
proc sort; by BIO_SEX OLDER4GRADE;
proc freq; tables FEELSOCACCEPT;
  where BIO_SEX ne . and OLDER4GRADE ne .;
by BIO_SEX OLDER4GRADE;
```



gender=1 older for grade=0 **Male not older for grade**

feel socially accepted				
FEELSOCACCEPT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	221	11.98	221	11.98
YES 1	1624	88.02	1845	100.00
Frequency Missing = 3				

The FREQ Procedure

gender=1 older for grade=1 **Male older for grade**

feel socially accepted				
FEELSOCACCEPT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	58	13.74	58	13.74
YES 1	364	86.26	422	100.00
Frequency Missing = 4				

gender=2 older for grade=0 **Female not older for grade**

feel socially accepted				
FEELSOCACCEPT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	338	15.55	338	15.55
YES 1	1835	84.45	2173	100.00
Frequency Missing = 4				

The FREQ Procedure

gender=2 older for grade=1 **Female older for grade**

feel socially accepted				
FEELSOCACCEPT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	67	23.76	67	23.76
YES 1	215	76.24	282	100.00
Frequency Missing = 3				

Males who are or are not older for their grade are similarly likely to feel socially accepted (86.3% vs. 88.0%), while females who are older for their grade are less likely to feel socially accepted (76.2%) than females who are not older for their grade (84.5%).