

Title of Resource	Assignment: Using public data to understand descriptive statistics
Author(s)	Sara S. Whipple
Institution	Virginia Military Institute
Brief Description:	This assignment covers several bases. It requires students to use SPSS to interpret and analyze secondary data. Students then report descriptive statistics in an APA style results section.
Keywords:	Statistics; Central Tendency; Variability; Levels of Measurement Communicating Scientific Results; APA Research Reports
Author Contact Information:	whippless@vmi.edu
Additional Information:	Pew Research Center. (2013). <i>April 17-May 19, 2013 online dating</i> [Date file]. Retrieved from http://www.pewinternet.org/datasets/may-2013-online-dating-prelim/

TeachPsychScience.org is made possible with grant support from the *Association for Psychological Science (APS) Fund for Teaching and Public Understanding of Psychological Science* to the site creators Gary Lewandowski, Natalie Ciarocco, and David Strohmets. All materials on this site have been subjected to a peer review process. We welcome additional resources (www.teachpsychscience.org/submissions).

© 2016 Sara S. Whipple. All Rights Reserved. This material may be used for noncommercial educational purposes. All other uses require the written consent of the authors.

Instructors:

I find that students don't fully understand when and why to calculate a median or mode as opposed to mean until they practice in an applied way. This activity is designed with that in mind. It can be used with nearly any dataset, though certain types of variables (ex. a variety of nominal, ordinal, scale) are necessary to demonstrate important learning goals. For best results I suggest downloading the Pew Research Center's publicly available dataset about online dating (the worksheet I provide coincides with this). You can find the link to the dataset here – <http://www.pewinternet.org/datasets/may-2013-online-dating-prelim/>. You must sign in to get the dataset, but it is free for research or teaching purposes. I use this as a take-home assignment, though it could easily be tailored to an in-class activity. The full activity will take a little over an hour, particularly since most students are not familiar with writing results in APA style.

I would recommend familiarizing students with the study itself. Specifically, telephone interviews were conducted by Princeton Survey Research Associates International from April 17 to May 19, 2013. The sample was 2,252 adults, age 18 and older. Interviews were conducted in English and Spanish by landline (1,125) and cell phone (1,127). The Pew website has more information about the study if instructors wish to learn more, or even download the specific questions that were asked. So as not to overwhelm students with excess variables, I delete a large portion of the dataset, save the condensed version, and post the condensed file for students. I keep the following variables: psraid, sex, q1, mar, mar2, mar3, breaka, breakb, date1a, date1bm1, date9a, date9b, date9c, age, educ2, and race. Although I ask students to use SPSS to complete this assignment, the assignment could be completed using any statistical program. Depending on your goals, you may want to request output files from students. Personally, I do not because I want students to practice sharing the results in a common form (APA style) rather than simply give me the results.

Student worksheet

In 2013 the Pew Research Center used a phone survey to collect data on Americans' internet usage and attitudes towards online dating. 2,252 adults, age 18 and older were sampled. Using this data, you will use descriptive statistics to summarize the findings. You will report your results in an APA style Results section. Steps for the full homework are listed below.

Steps:

1. Open the dataset and familiarize yourself with the types of variables. Reminder: you'll want to be in the "Variable View" mode of SPSS.

2. Run the appropriate analyses so that you can explain the following characteristics of the sample. You should report frequencies, measures of central tendency and variability as necessary (in other words, *you must decide* which statistics are most meaningful and appropriate to report for each of the variables below).
 - a) Demographic breakdown of respondents
 - i. include age, sex, and race
 - ii. of those who are married or in a committed relationship, summarize the length of these relationships
 - b) Summarize how many respondents have used an online dating site
 - c) Report the most common dating site(s) people have used
 - d) Based on the data, to what degree do participants agree and disagree with the following statements?
 - i. Online dating is a good way to meet people.
 - ii. Online dating allows people to find a better match for themselves.
 - iii. People who use online dating sites are desperate.
 - e) Summarize how many people have broken up over text or email.
 - f) Summarize the quality of life ratings for those who were surveyed.

3. Write up your results in an APA style results section using the appropriate notation (ex. n , M) for each of the analyses above. Include one table and one figure (in APA style) that appropriately display/summarize some part of the information.

Answer Key:

With the assignment I provide example of APA style results sections, along with Tables and Figures. Here is a simple one you may want to include:

The participants were 300 ninth grade and twelfth grade students from a high school in a suburban, primarily middle-class community in a mid-Atlantic state. Forty percent of these students were ninth graders ($n = 120$), 52% were girls ($n = 156$), 63% were Caucasian ($n = 189$). The average GPA of the students was 2.78 ($SD = .35$).

a) i. Demographics:

Age ($n = 2252$), $M = 52.13$ years, $SD = 20.14$, range is 18-99 years

Sex 1029 males (45.7%), 1223 females (54.3%). In just these first two example, students should begin to differentiate between age as a scale variable and sex as a categorical/nominal variable. It makes sense to report a mean and SD for age but not for sex.

Race: report frequencies/percentages since categorical (see table below).

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	White	1707	75.8	75.8	75.8
	Black or African-American	287	12.7	12.7	88.5
	Asian or Pacific Islander	45	2.0	2.0	90.5
	Mixed race	62	2.8	2.8	93.3
	Native American/American Indian	36	1.6	1.6	94.9
	Other (SPECIFY)	44	2.0	2.0	96.8
	(VOL.) Don't know	9	.4	.4	97.2
	(VOL.) Refused	62	2.8	2.8	100.0
	Total	2252	100.0	100.0	

ii. Of those who are married, or in committed relationship, how many years?

$n = 1428$, $M = 21.81$, $SD = 18.99$

Median = 18, mode = 1

Note that this is an interesting variable. You might question the one participant who has been married for 99 years. Also, although the mean is high, the median, and especially mode are different. Good example to

a) How many have used an online dating site.

See breakdown below: Of those who answered the question, 10.4% have used online dating site ($n = 198$).

DATE1a. Have YOU, personally, ever used an online dating site such as Match.com, eHarmony, or OK Cupid?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	198	8.8	10.4	10.4
	No	1691	75.1	89.2	99.7
	(VOL.) Don't know	2	.1	.1	99.8
	(VOL.) Refused	4	.2	.2	100.0
	Total	1895	84.1	100.0	
Missing	System	357	15.9		
Total		2252	100.0		

b) Most common dating sites

Match.com was people's most frequent first response. Of those who used online dating sites, 81 participants (40.9%) mentioned this site first, followed by eHarmony (17.5%) and OK Cupid (10.6%).

DATE1b. (1ST MENTION) Which online dating sites have you used? [DO NOT READ]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Match.com	81	3.6	40.9	40.9
	eHarmony	35	1.6	17.7	58.6
	OK Cupid	21	.9	10.6	69.2
	Plenty of Fish	18	.8	9.1	78.3
	Christian Mingle	3	.1	1.5	79.8
	Zoosk	2	.1	1.0	80.8
	J Date	2	.1	1.0	81.8
	Adult Friend Finder	1	.0	.5	82.3
	Other (SPECIFY)	19	.8	9.6	91.9
	Don't know	15	.7	7.6	99.5
	Refused	1	.0	.5	100.0
	Total	198	8.8	100.0	
Missing	System	2054	91.2		
Total		2252	100.0		

c) To what degree do participants agree and disagree with the following statements?

- i. Online dating is a **good way to meet people**: 1239 (55%) agreed; 827 (36.7%) disagree
- ii. Online dating allows people to **find better match**: 1097 (48.7%) agree; 937 (41.6%) disagree; 182 (8.1%) don't know
- iii. People who use online dating are **desperate**: 515 (22.9%) agree, 1559 (69.2%) disagree; 138 (6.1%) don't know

e) Broken up over text or email: 106 (16.2%) yes, 545 (83.5%) no

f) Quality of life ratings: 18.1% of participants ($n = 407$) rated the quality of their lives as excellent. 4.9% ($n = 110$) rated their quality of life as poor. The modal response was “good” with 33.8% ($n = 762$) of participants answering this way.

Q1. Overall, how would you rate the quality of life for you and your family today?

Would you say it is excellent, very good, good, fair or poor?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Excellent	407	18.1	18.1	18.1
	Very good	618	27.4	27.4	45.5
	Good	762	33.8	33.8	79.4
	Fair	335	14.9	14.9	94.2
	Poor	110	4.9	4.9	99.1
	(VOL.) Don't know	8	.4	.4	99.5
	(VOL.) Refused	12	.5	.5	100.0
	Total	2252	100.0	100.0	