

The Effects of Using Big Data Sets in an Introductory Statistics Class

John J Weber III, Georgia Perimeter College

Stacy Thrall, Georgia Gwinnett College

20 November 2015

GAISE Recommendations

1. Emphasize statistical literacy and develop statistical thinking.
2. Use real data.
3. Stress conceptual understanding, rather than mere knowledge of procedures.
4. Foster active learning in the classroom.
5. Use technology for developing concepts and analyzing data.

Data

1. <http://www.Data.gov>.
2. <https://www.cia.gov/library/publications/the-world-factbook/>.
3. <http://www.census.gov/data.html>.
4. <http://open-data.europa.eu/en/data/>.
5. <http://data.gov.uk/>.
6. <https://www.healthdata.gov/>.
7. <http://www.hscic.gov.uk/home>.
8. <http://aws.amazon.com/datasets>.
9. <http://www.inside-r.org/howto/finding-data-internet>.

Internet Research Data

1. Omnibus Pew Internet Research dataset

[Omnibus Internet Research Data](#)

2. Internet Research Data Description

[Omnibus Internet Research Data Description](#)

Importing Database

.xls or .csv are the easiest to import into most software packages.

Omnibus Internet Research Data															
<div>AppletsEditDataStatGraphHelp</div>															
Row	psraid	sample	state	cregion	usr	eminuse	intmob	pialql1	smart1	SmartPhone	pial1	pial2	pial3	pial4	pial5
1	100009	1	37	3	U	1	1	1	8	2	2	2	2	2	2
2	100019	1	37	3	U	1	1	1	1	1	2	2	2	2	2
3	100022	1	51	3	U	1	1	1	2	2	2	2	1	2	2
4	100025	1	42	1	S	1	1	1	1	1	2	2	2	2	2
5	100026	1	18	2	U	2	2	1	2	2					2
6	100027	1	25	1	S	1	2	1	1	1	2	2	2	2	2
7	100032	1	12	3	U	1	1	1	1	1	1	1	2	1	2
8	100035	1	36	1	U	1	2	1	2	2	2	2	2	2	2
9	100038	1	18	2	R	2	2	1	2	2					2
10	100044	1	47	3	U	1	1	1	1	1	1	1	2	1	2
11	100049	1	37	3	S	1	1	1	2	2	1	1	1	1	2
12	100050	1	25	1	S	1	1	1	2	2	2	2	2	2	2
13	100060	1	1	3	U	1	2	1	2	2	1	2	2	2	2
14	100070	1	17	2	S	1	2	1	2	2	2	2	2	2	2
15	100081	1	17	2	S	1	1	1	1	1	2	2	2	2	2
16	100083	1	29	2	S	2	2	2		3					
17	100093	1	5	3	R	2	2	2		3					
18	100095	1	18	2	R	1	1	1	1	1	2	2	2	2	2
19	100105	1	18	2	R	1	1	1	1	1	1	1	1	1	2
20	100109	1	18	2	S	2	2	1	1	1					2
21	100111	1	12	3	R	1	1	1	1	1	1	1	2	2	2
22	100118	1	29	2	S	1	1	1	1	1	2	2	2	2	2
23	100136	1	33	1	R	1	1	2		3	1	1	2	1	
24	100139	1	29	2	S	2	2	2		3					
25	100143	1	12	3	S	1	1	1	2	2	2	2	2	2	2
26	100145	1	37	3	R	1	2	1	2	2	2	2	2	2	2
27	100148	1	12	3	U	1	1	1	1	1	1	1	1	1	2
28	100152	1	9	1	S	2	2	1	2	2					2
29	100156	1	55	2	S	2	2	1	2	2					2
30	100157	1	51	3	S	1	2	1	1	1	2	2	2	2	2
31	100159	1	27	2	R	2	2	1	2	2					2
32	100163	1	29	2	U	1	1	1	1	1	1	1	2	2	2
33	100165	1	13	3	S	1	2	2		3	1	1	2	2	

Code Book for Data

Code books are usually linked on the same page as the data set.

PEW INTERNET PROJECT

PSRAI Omnibus October Week 1 (October 3-6, 2013)

Revised: October 1, 2013

ASK ALL:

EMINUSE Do you use the internet or email, at least occasionally? {PIAL trend}

- 1 Yes
- 2 No
- 8 **(DO NOT READ)** Don't know
- 9 **(DO NOT READ)** Refused

ASK ALL:

INTMOB Do you access the internet on a cell phone, tablet or other mobile handheld device, least occasionally? {PIAL trend}

- 1 Yes
- 2 No
- 8 **(DO NOT READ)** Don't know
- 9 **(DO NOT READ)** Refused

ASK LANDLINE SAMPLE (SAMPLE=1):

PIALQ1 Do you have a cell phone? {PIAL trend}

- 1 Yes
- 2 No
- 8 **(DO NOT READ)** Don't know
- 9 **(DO NOT READ)** Refused

ASK IF HAVE CELL PHONE (PIALQL1=1 or SAMPLE=2):

PIAL5 Do you ever use the mobile app called "SNAP-CHAT" that is used to take and share photos and videos, or do you not use that app?

- 1 Yes
- 2 No
- 8 **(DO NOT READ)** Not sure/Don't know
- 9 **(DO NOT READ)** Refused

ASK IF HAVE CELL PHONE (PIALQL1=1 or SAMPLE=2):

PIAL7 Do you ever use INSTAGRAM on your cell phone, or do you not use that app?

- 1 Yes
- 2 No
- 8 **(DO NOT READ)** Not sure/Don't know
- 9 **(DO NOT READ)** Refused

ASK ALL INSTAGRAM USERS (PIAL7=1):

PIAL8 How often do you use Instagram? Would you say... [READ]

- 1 Several times a day
- 2 About once a day
- 3 3 to 5 days a week
- 4 1 to 2 days a week
- 5 Every few weeks (OR)
- 6 Less often
- 8 **(DO NOT READ)** Not sure/Don't know
- 9 **(DO NOT READ)** Refused

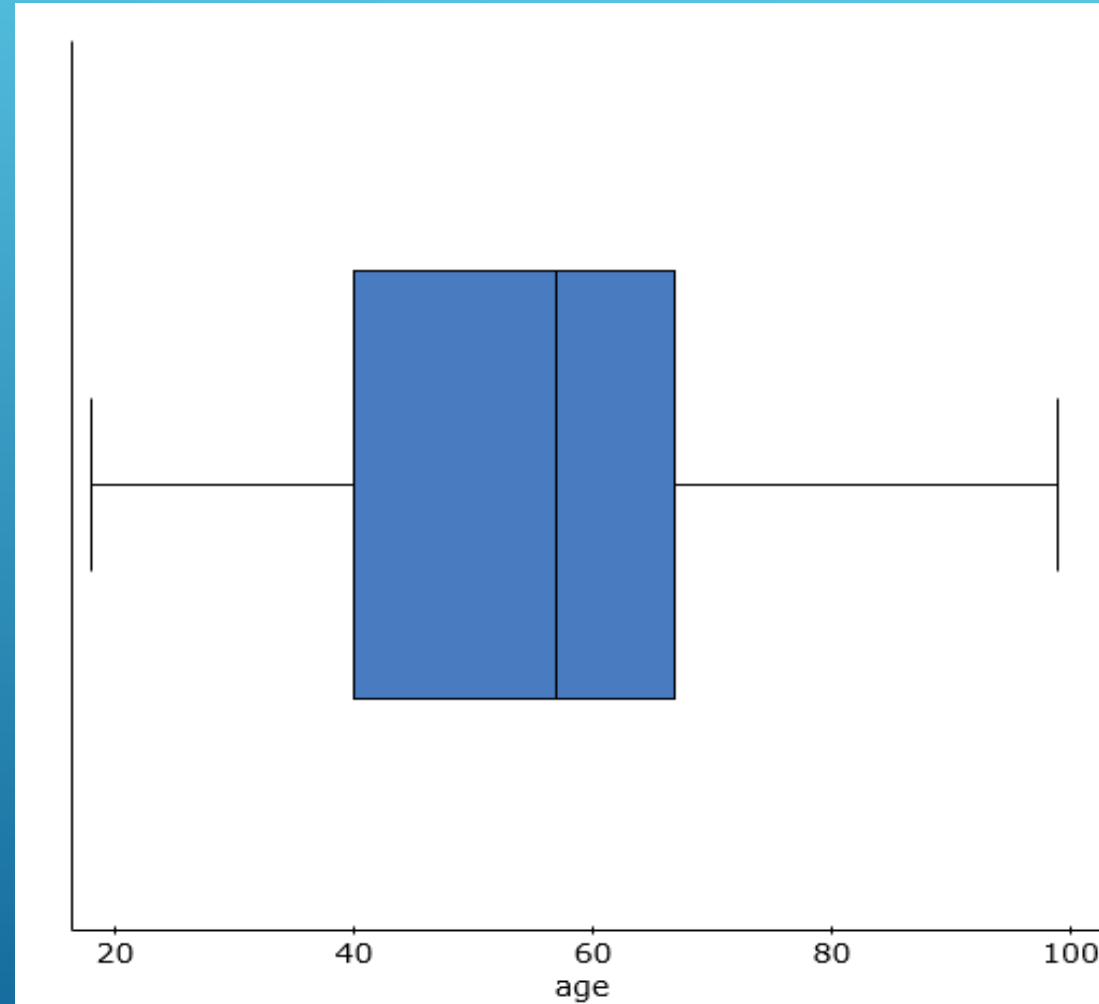
Internet Research Data Set

Variables for Potential Explorations by Students:

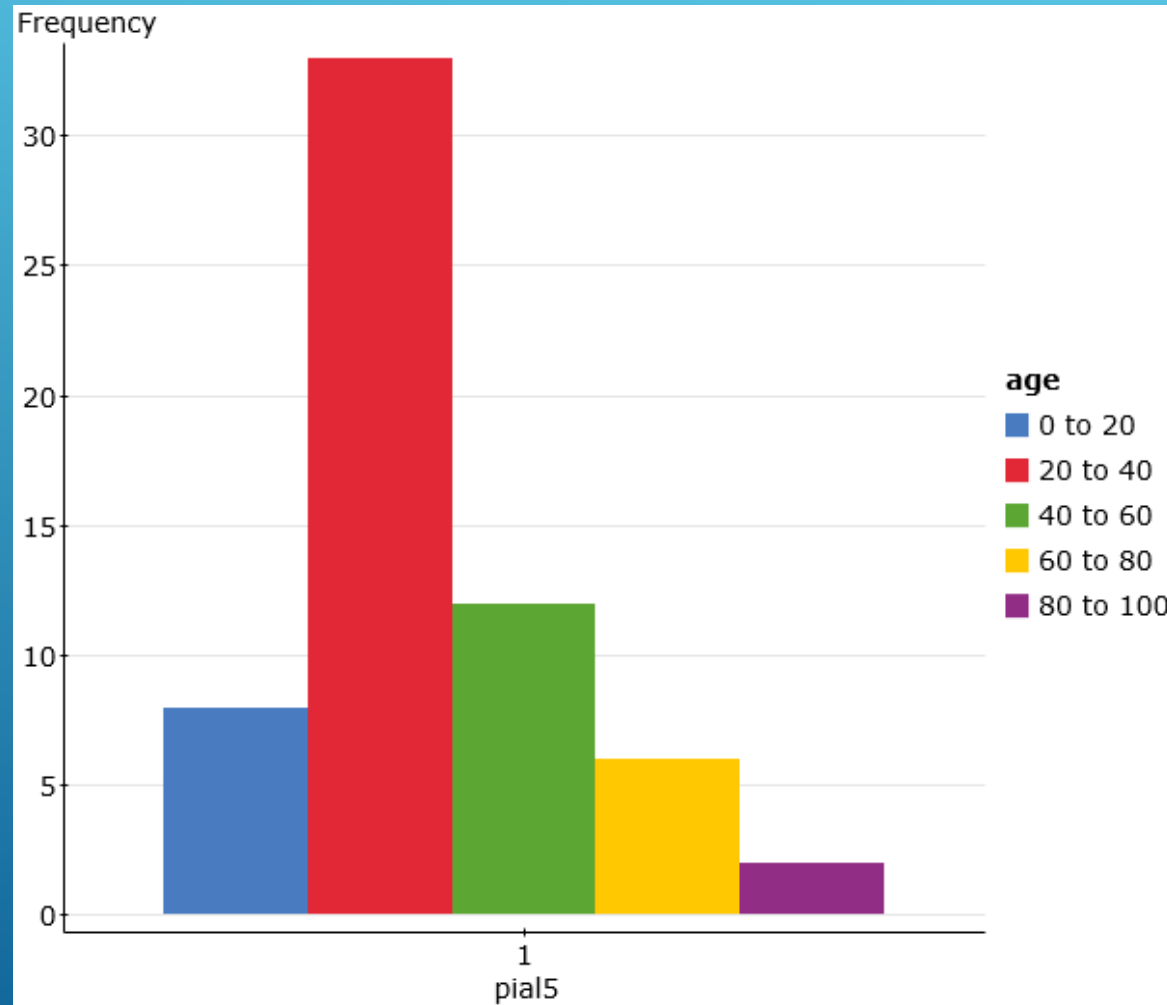
1. intmob - Use of cell phones/tablets for internet purposes
2. pial1 – posting of photos
3. pial5 - Use of snapchat
4. pial7 - Use of Instagram

While still having the variables of age, race, income, etc. to group by.

Box and Whisker Plot for the age variable in the data set

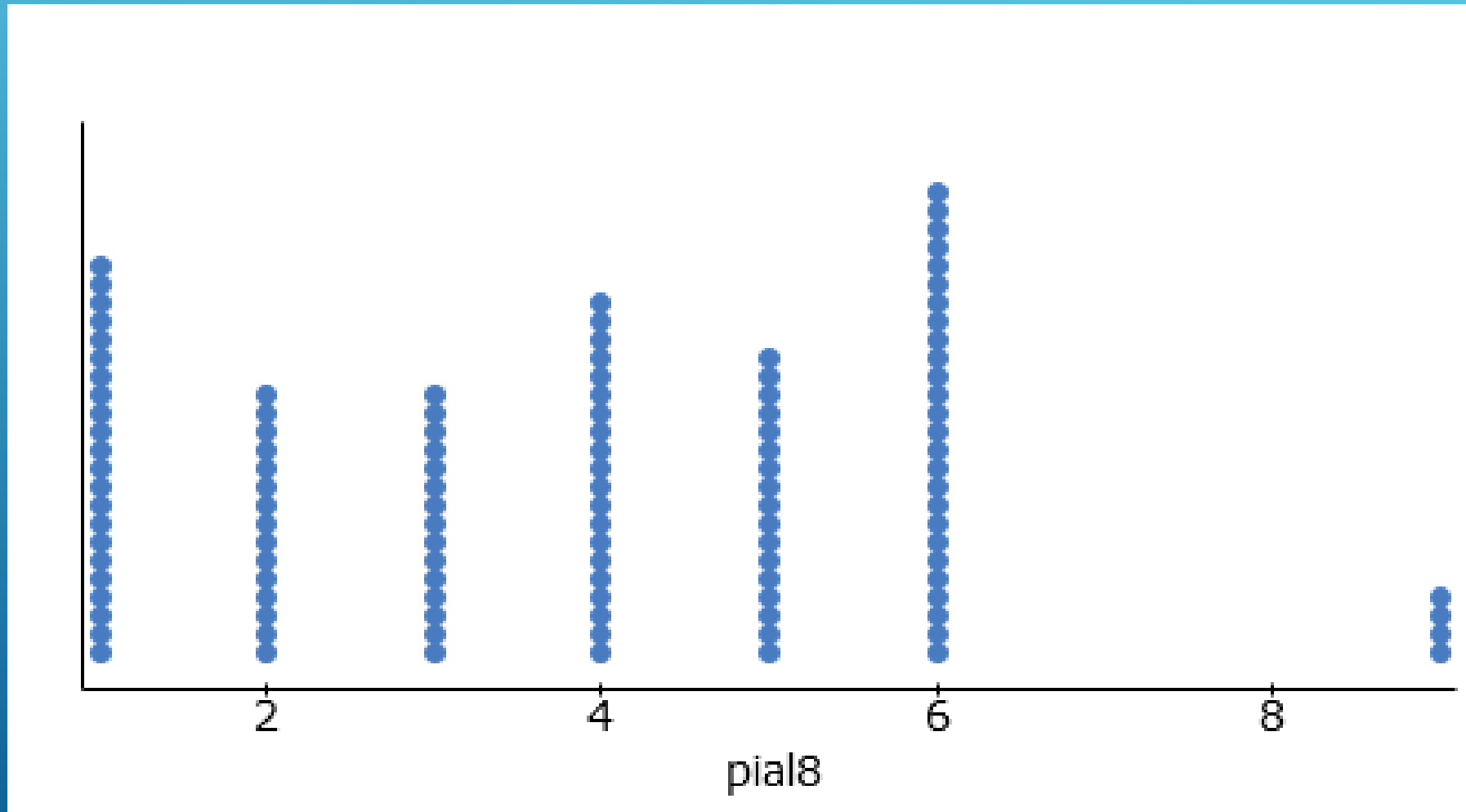


Do you ever use the mobile app called "SNAP-CHAT" that is used to take and share photos and videos? (pia15).
Those who responded 1 for "Yes" by age group



DOT PLOT OF HOW OFTEN INSTAGRAM USERS POST PHOTOS TO THE SITE (PIAL8)

frequency



Potential Value for Students

1. Allows the instructor to find an interesting data set the students will enjoy exploring.
2. Great exposure to the data sets they will see in the real worlds of business, industry, healthcare, etc.
3. Instead of just memorizing the processes, they are applying the concepts to real world data and situations.
4. Students enjoyed the computer lab interactive set up during class, increased participation.

Thank You

Any Questions?

John J. Weber

Georgia Perimeter College

John.Weber@gpc.edu

Stacy Thrall

Georgia Gwinnett College

sthrall@ggc.edu