

Think, Show, Tell



- There are three simple steps to doing Statistics right:

Think first. Know where you're headed and why.

Show is about the mechanics of calculating statistics and making graphical displays, which are important (but are not the most important part of Statistics).

Tell what you've learned. You must explain your results so that someone else can understand your conclusions.

What Are Data?



- Data can be numbers, record names, or other labels.
- Not all data represented by numbers are numerical data (e.g., 1 = male, 2 = female).
- Data are useless without their context...

ex: Zip codes

What do the values 75, 98, 66, 40, 52
represent???

apples ages ###
blueberries grades ###
Temperatures //
floors



Data is useless without context and/or units.

The “W’s”



- To provide context we need the W’s

- Who
- What (and in what units)
- When
- Where
- Why (if possible)
- and How

?????

???

of the data.

- Note: the answers to “who” and “what” are essential.
- Note: Unit III will be devoted to How we collect data.

Data Tables

- The following data table clearly shows the context of the data presented:



Amazon
CD Purchases

Name	Ship to State/Country	Price	Area Code	Previous CD Purchase	Gift?	ASIN	Artist
Katharine H.	Ohio	10.99	440	Nashville	N	B00000I5Y6	Kansas
Samuel P.	Illinois	16.99	312	Orange County	Y	B000002BK9	Boston
Chris G.	Massachusetts	15.98	413	Bad Blood	N	B000068ZVQ	Chicago
Monique D.	Canada	11.99	902	Let Go	N	B000001OAA	Garbage

who

what

- Notice that this data table tells us the **What** (column) and **Who** (row) for these data.



Who

- The *Who* of the data tells us the individual **cases** for which (or whom) we have collected data.
 - Individuals who answer a survey are called **respondents**.
 - People on whom we experiment are called **subjects** or **participants**.
 - Animals, plants, and inanimate subjects are called **experimental units**.

*not necessary
people*

Who (cont.)

- Sometimes people just refer to data values as **observations** and are not clear about the *Who*.
 - But we need to know the *Who* of the data so we can learn what the data say.
- In this course, we will discuss data collected from many different individuals. For example,
 - Students at your school
 - The 50 fastest roller coasters
 - Tomato plants

Who (cont.)



- Population vs. Sample

- Population is the **entire** group of interest
 - Ex: All students at Baker High School
- Sample is **part** of the population of interest
 - Ex: Every tenth person to arrive at school on a day

What and Why

- **Variables** are characteristics recorded about each individual.
- The variables should have a name that identifies *What* has been measured.
- To understand variables, you must *Think* about what you want to know.

What and Why (cont.)



- Some variables have units that tell how each value has been measured and tell the scale of the measurement.

- Examples:

- Distance (meters, feet, etc.)
- Mass (grams, pounds, etc.)
- Time (seconds, years, etc.)
- Temperature
- Speed

grades (points)

PROBLEM: The American Statistical Association sponsors a Web-based project that collects data about primary and secondary school students using surveys. We used the site's "Random Sampler" to choose 40 U.S. high school students who completed the survey in a recent year.⁴ The table displays data for the first 10 students chosen. The rightmost column gives students' answers to the question:

Context
to find W's

Which would you prefer to be? Select one.

<input type="checkbox"/> Rich	<input type="checkbox"/> Happy	<input type="checkbox"/> Famous	<input type="checkbox"/> Healthy
-------------------------------	--------------------------------	---------------------------------	----------------------------------

State	Grade level	Gender	Age	Birth month	Height (cm)	Arm span (cm)	Preferred status
SC	12	Male	17	January	177	161	Famous
UT	9	Female	14	March	162	153	Healthy
NM	12	Female	17	August	164	167	Healthy
CA	12	Female	17	April	153	154	Famous
GA	12	Female	17	June	172	169	Happy
MI	11	Male	17	March	170	173	Famous
IN	12	Female	18	January	168	163	Happy
CO	9	Female	14	June	152	160	Happy
NJ	10	Female	16	November	165	174	Famous
CO	9	Male	15	January	190	177	Rich

Identify the Who's and What's.

Who's: 40 high school students who completed the survey

What's: State, grade, gender, age, birth month, height(cm), arm span(cm), preferred status

On the first day of a statistics course, the instructor gave all 40 students in the class a survey. The table shows data from the first 10 students on the class roster.

Gender	Class	GPA	Pulse rate	Dominant hand	Children in family	Homework last night (min)	Sleep (h)	Have a smart-phone?
F	Fr	3.22	72	R	3	0–14	10	Y
F	Fr	2.3	110	L	3	0–14	8	N
M	Ju	3.8	60	L	6	15–29	7	Y
M	So	3.1	72	R	2	15–29	7.5	Y
F	So	4.0	51	R	1	45–59	7	Y
F	So	3.4	68	R	4	0–14	8.5	Y
F	So	3.0	80	R	3	30–44	7	Y
M	So	3.5	59	R	2	30–44	7	Y
M	Fr	3.9	65	R	2	15–29	6	Y
M	Sr	3.5	104	R	2	0–14	7	N
...								

Identify the Who's and What's.

Who's: 40 students in a statistics class

What's: Gender, class, GPA, pulse rate, dominant hand, children in family, time spent on homework (min), sleep (hours), smart phone (y/n)

Box-office smash According to the Internet Movie Database, *Avatar* is tops based on box-office receipts worldwide. The table displays data on several popular movies.⁵

Who?

Popular movies



Movie	Year	Rating	Time (min)	Genre	Box office (\$)
<i>Avatar</i>	2009	PG-13	162	Action	2,783,918,982
<i>Titanic</i>	1997	PG-13	194	Drama	2,207,615,668
<i>Star Wars: The Force Awakens</i>	2015	PG-13	136	Adventure	2,040,375,795
<i>Jurassic World</i>	2015	PG-13	124	Action	1,669,164,161
<i>Marvel's The Avengers</i>	2012	PG-13	142	Action	1,519,479,547
<i>Furious 7</i>	2015	PG-13	137	Action	1,516,246,709
<i>The Avengers: Age of Ultron</i>	2015	PG-13	141	Action	1,404,705,868
<i>Harry Potter and the Deathly Hallows: Part 2</i>	2011	PG-13	130	Fantasy	1,328,111,219
<i>Frozen</i>	2013	PG	108	Animation	1,254,512,386
<i>Iron Man 3</i>	2013	PG-13	129	Action	1,172,805,920

What?

Name, year, rating, time(min), genre, box office (\$)