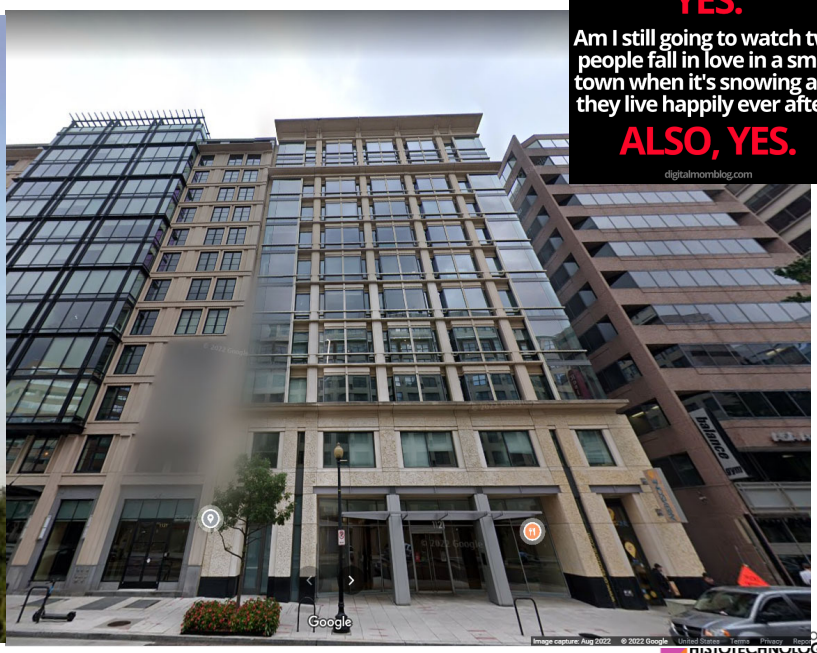
Handout Disclaimer

Disclaimer: the following document was distributed as a handout for the Analyzing Data for Beginner Series and is designed to enhance the sessions you have attended. NSH makes no representations to the factual correctness of any information contained herein. All of the content comprising this handout is the exclusive property of the presenter and the national society for histotechnology. It may not be copied, reproduced, distributed, displayed or transmitted without the consent of the presenter or the National Society for Histotechnology.

Analyzing Data for Beginners: Day 1

Connie Wildeman, MPA
National Society for Histotechnology



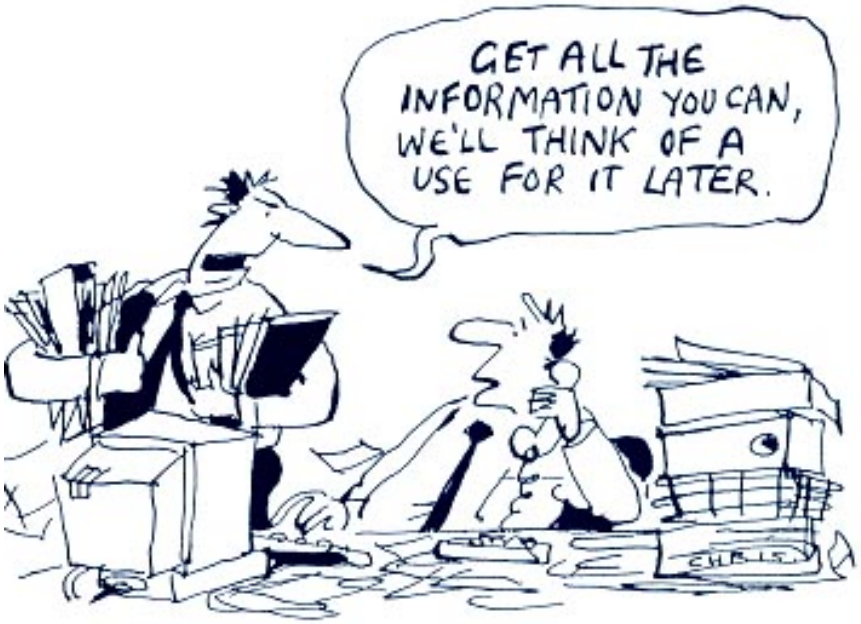
Does every Hallmark Christmas movie have the same plot?

YES.

Am I still going to watch two people fall in love in a small town when it's snowing and they live happily ever after?

ALSO, YES.

digitalmomblog.com



Credit: Gareth Starkey

NATIONAL SOCIETY FOR HISTOTECHNOLOGY
Connecting. Empowering. Innovating.

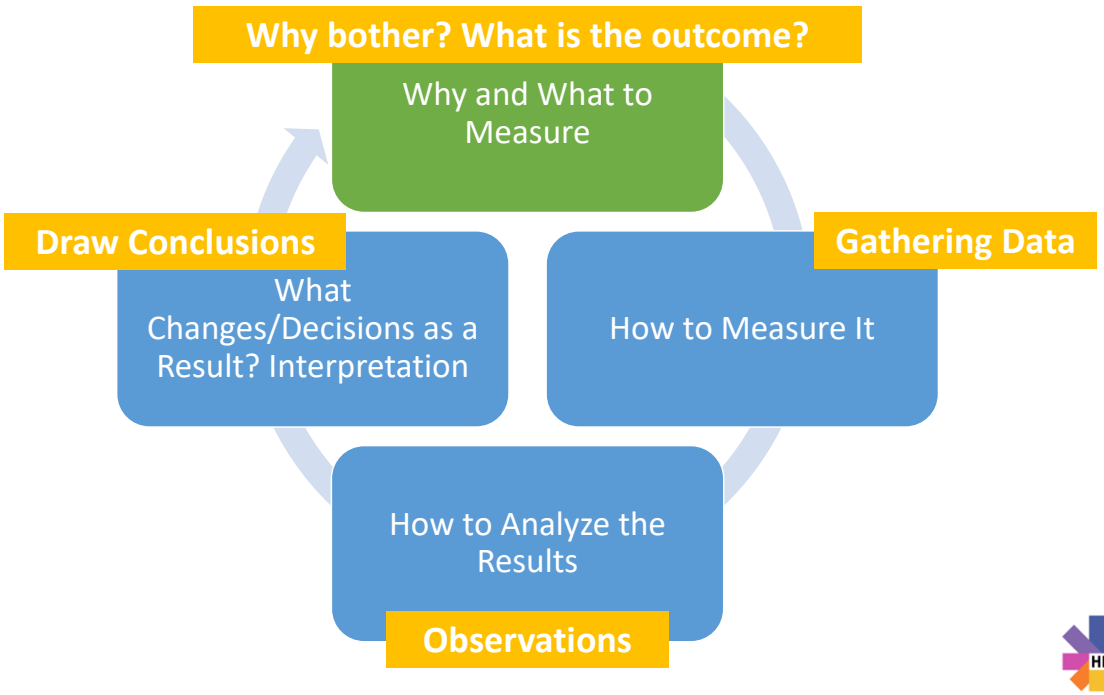
What we will cover today:

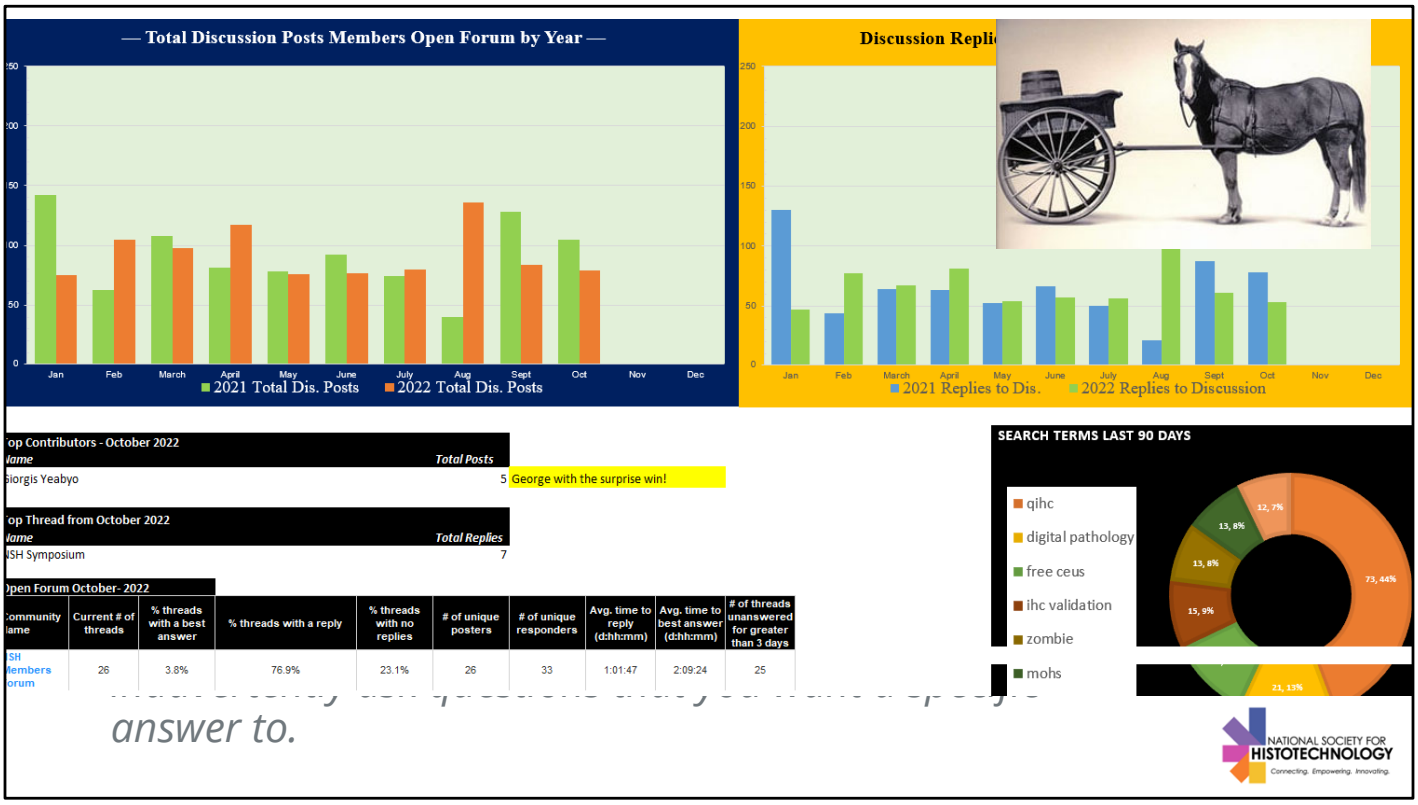
Part 1: What is Data?
Part 2: Preparing Data for Analysis
Part 3: Analyzing Data

Part I: What is Data?

1. **Summarize where data fits into the evaluation cycle.**
2. Identify what data is present and available to you.
3. Define data
4. Classify types of data

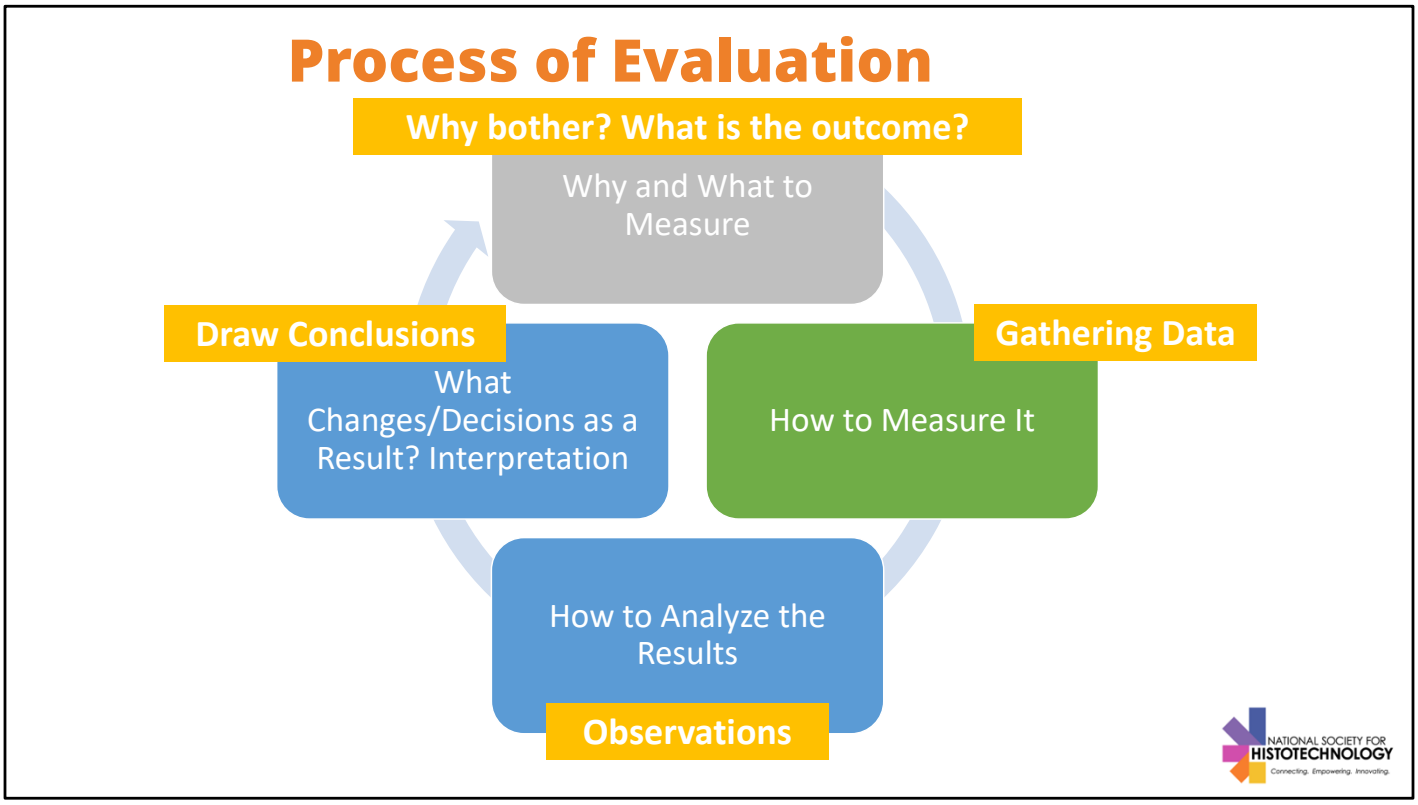
Process of Evaluation




— Total Discussion Posts Members Open Forum by Year —
Jan Feb March April May June July Aug Sept Oct Nov Dec
2021 Total Dis. Posts
2022 Total Dis. Posts
Discussion Repli
2021 Replies to Dis.
2022 Replies to Discussion
op Contributors - October 2022
lame
Total Posts
iorgis Yeabyo
5
George with the surprise win!
op Thread from October 2022
lame
Total Replies
SH Symposium
7
pen Forum October- 2022
ommunity lame
Current # of threads
% threads with a best answer
% threads with a reply
% threads with no replies
# of unique posters
# of unique responders
Avg. time to reply (d:hh:mm)
Avg. time to best answer (d:hh:mm)
# of threads unanswered for greater than 3 days
SH Members orum
26
3.8%
76.9%
23.1%
26
33
1:01:47
2:09:24
25
SEARCH TERMS LAST 90 DAYS
qihc
digital pathology
free ceus
ihc validation
zombie
mohs
12, 7%
13, 8%
13, 8%
73, 44%
15, 9%
21, 13%
answer to.
NATIONAL SOCIETY FOR HISTOTECHNOLOGY
Connecting. Empowering. Innovating.

Examples of Questions

- How does my staff feel about work schedules?
- Should my lab use an automated embedder?
- Should we invest in a new database?



There is a lot of data available to you. In most cases this step is going to be done by someone who is an evaluator, but if you are asked to create the measurement tool (like some kind of ordinal tool, like a liekrt scale)

Part I: What is Data?

1. Summarize where data fits into the evaluation cycle.
2. **Identify what data is present and available to you.**
3. Define data
4. Classify types of data

Lets move onto the last learning objective for this section, and that is really the second part of the cycle. Change slide

- How does my staff feel about work schedules?

What Data Do You Already Have Access To?

How to Measure/Count

- LIS canned reports
- Emails
- Error logs
- Budgets (current and past)
- Page views
- Search terms
- Invoices
- Sign out logs
- Staff meeting agendas
- Leave requests

NATIONAL SOCIETY FOR HISTOTECHNOLOGY
Connecting. Empowering. Innovating.

Data is something that all of you are using the in the lab, whether formally by utilized canned reports from your LIS, invoices, document access reports and logs. Unless you are engaging in a complex research study, this type of data is what you can use.

Okay, so data exists all over the place. But how do I know what I am looking at?

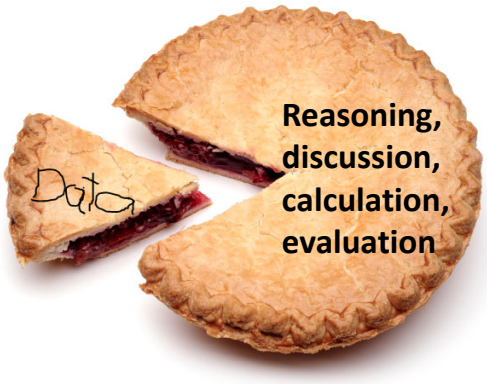
Part I: What is Data?

1. Summarize where data fits into the evaluation cycle.
2. Identify what data is present and available to you.
3. **Define data**
4. Classify types of data

Lets move onto the last learning objective for this section, and that is really the second part of the cycle. Change slide

What is Data?

- **Factual information** (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation (taken from Merriam-Webster)



I think this is great definition of data because it doesn't just mention the calculation. This is because people tend to forget there is a lot of different types of data out there and we use that data to solve problems, makes decisions, and identify patterns every day.

Types of Data

Quantitative: its numerical in nature and can include things like, test scores, temperatures, click rates.

Number of leave requests

Qualitative: its descriptive in nature, like color, types of college degrees, and frequency can be calculated. Also known as categorical.

Type of leave request

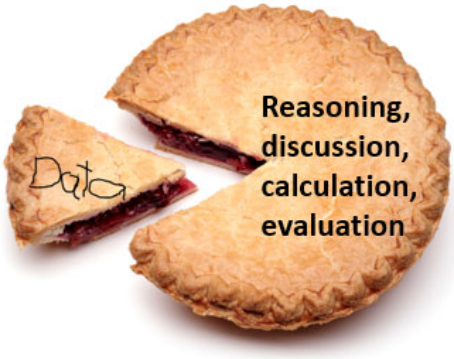
Identify if the following are Qualitative or Quantitative

1. The baby weighs 20 pounds
2. The workshop attendees rated the event highly effective
3. The sky is blue
4. There were 200 IHC requests this month
5. There were 26 sick leave requests in Q2
6. Joe is 6 foot 2

NATIONAL SOCIETY FOR HISTOTECHNOLOGY
Connecting. Empowering. Innovating.

Data is…

- Factual information made up of qualitative and/or quantitative variables that is an important piece in the evaluation and reasoning process.



Data is not a decision

Data is a part of the evaluation or decision making process. It is NOT the decision. It is an AMAZING resource to answer questions though.
You do not have to be processing large data sets to get answers – in fact what we will be working with in the remaining part of this presentation will be sample of data – not full populations.

Part I: What is Data?

1. Summarize where data fits into the evaluation cycle.
2. Identify what data is present and available to you.
3. Define data
4. **Classify types of data**

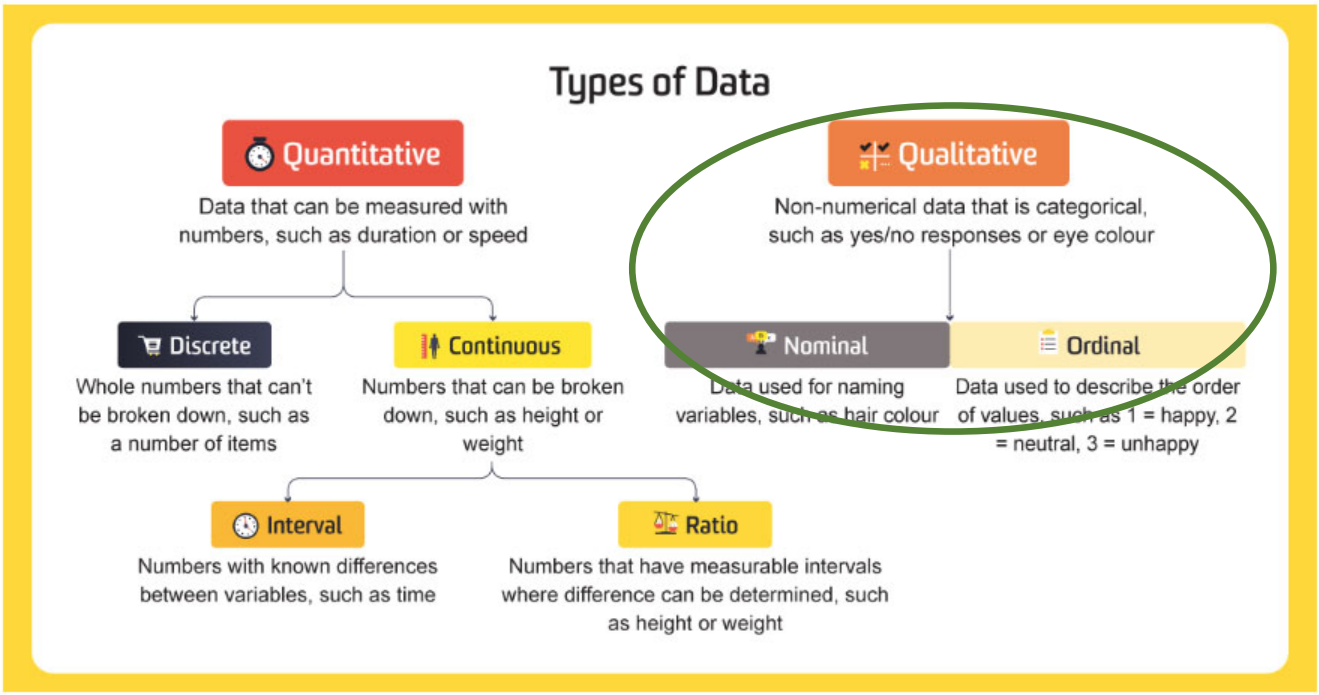
Why does type of data matter?

It determines the kinds of statistical tests and measures that can be used.

How to Measure It

NATIONAL SOCIETY FOR HISTOTECHNOLOGY
Connecting. Empowering. Innovating.

Now that we know that there is qualitative and quantitative data, lets dive even deeper into what kins of qual and quan exists – and how you can, or already are, using it!



https://studyonline.unsw.edu.au/blog/types-of-data

Continuous data is also associated with two types of measurement – internval and ratio.

Nominal:

nom·i·nal ˈnä-mə-nᵊl ˈnäm-nəl

1 : of, relating to, or being a noun or a word or expression taking a noun construction

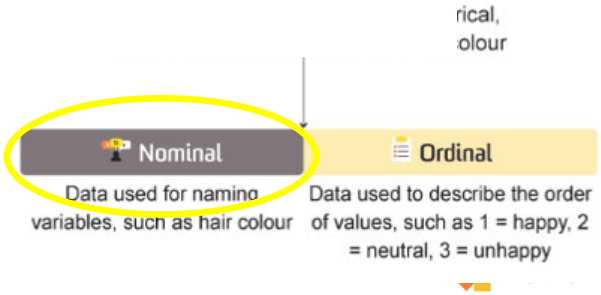
2 a : of, relating to, or constituting a name

b : bearing the name of a person

3 a : existing or being something in name or form only

nominal head of his party

(from Merriam-Webster)

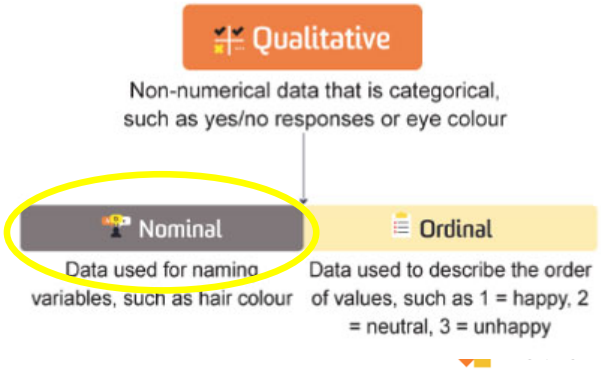


Nominal Data- Categorical

Nominal data are categorized according to labels which are purely descriptive—they don't provide any quantitative or numeric value. **Nominal data cannot be placed into any kind of meaningful order or hierarchy.**

Examples:

- Eye color
- Type of stain
- Gender
- Car color



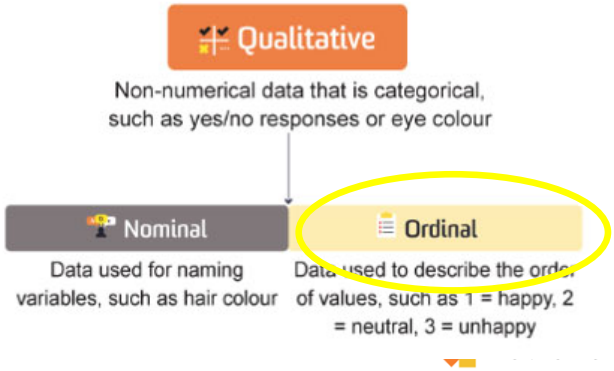
Use mode and frequency to statistically measure nominal.
How responses vary: how do men and women answer the same questions?

Ordinal:

ordinal 2 of 2 *adjective*

1 : of a specified order or rank in a series

2 : of or relating to a taxonomic order

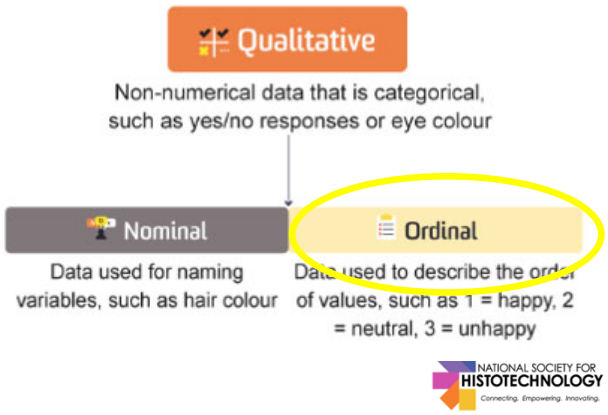


Ordinal Data- Categorical

Ordinal data are categorical (non-numeric) but may use numbers as labels – BUT CAN BE RANKED NATURALLY.

Examples:

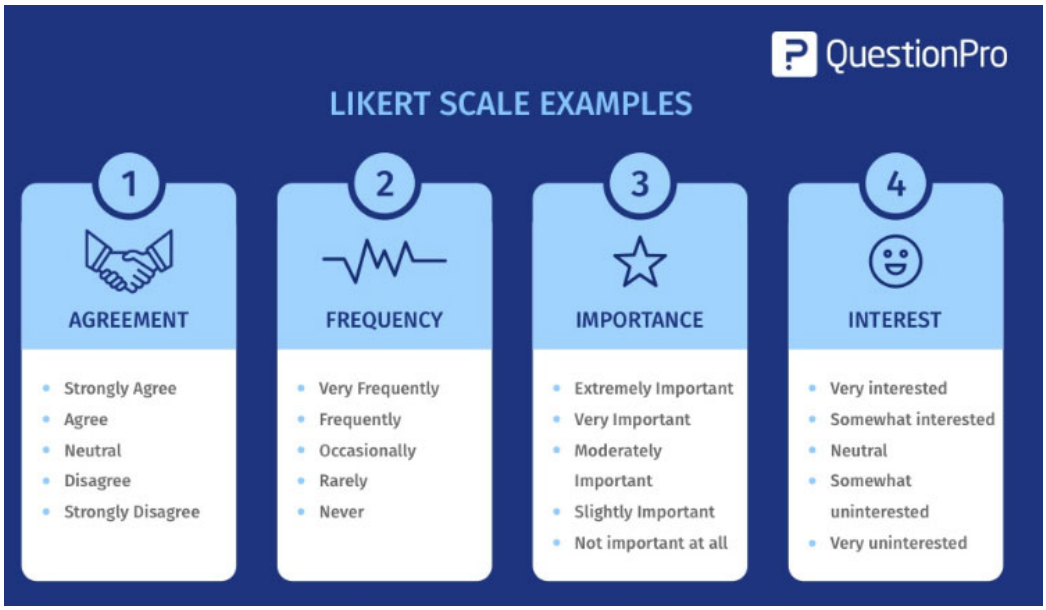
- Grades/Marks (A,B,C,O,ME,etc)
- Education level
- Income level (low, middle, upper middle class, etc.)
- Satisfaction levels



https://careerfoundry.com/en/blog/data-analytics/what-is-ordinal-data/#what-is-ordinal-data-a-definition

Likert Scale – Ordinal (hierarchy, categorical)

https://www.questionpro.com/blog/what-is-likert-scale/

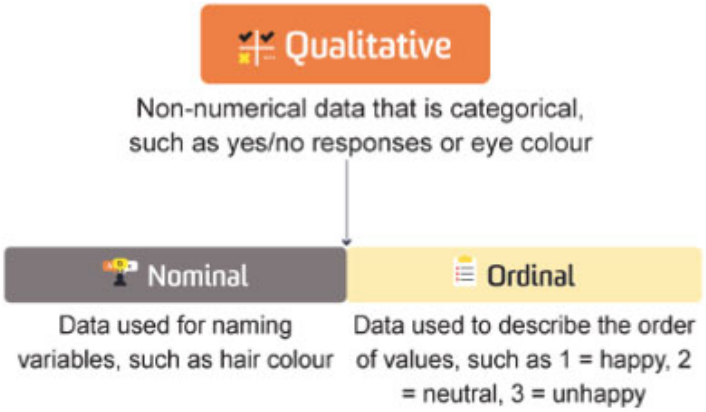


In this example of an LIS – what kind of data is the Task (nominial)
What kind of data is the priority – ordinal

Priority	Block		Container		Task	Lab Responsible Pathologist	
mbedded Priority	PS20-	A1	PS20-	A1-1	H&E		
mbedded Priority	PS20-	A1	PS20-	A1-4	H&E		
mbedded Priority	PS20-	A1	PS20-	A1-7	H&E		
mbedded Priority	PS20-	A1	PS20-	A1-	H&E		
mbedded Priority	PS20-	A1	PS20-	A1-	ERG		
mbedded Priority	PS20-	A1	PS20-	A1-7	H&E		
mbedded Priority	PS20-	A1	PS20-	A1-	H&E FS		
mbedded Priority	PS20-	A1	PS20-	A1-	H&E		
mbedded Priority	PS20-	A1	PS20-	A1-	H&E		
mbedded Priority	SS20-	A1	SS20-	A1-11	H&E		ENT Service
mbedded Priority	SS20-	A1	SS20-	A1-13	H&E		ENT Service
mbedded Priority	SS20-	A1	SS20-	A1-9	H&E		ENT Service
mbedded Priority	SS20-	A1	SS20-	A1-1	ERG		Derm Service
mbedded Priority	SS20-	A2	SS20-	A2-1	ERG		Derm Service
mbedded Priority	SS20-	A3	SS20-	A3-1	H&E		Derm Service
mbedded Priority	SS20-	A4	SS20-	A4-1	H&E		Derm Service
mbedded Priority	SS20-	A5	SS20-	A5-1	H&E		Derm Service
mbedded Priority	SS20-	A6	SS20-	A6-1	ERG		Derm Service
mbedded Routine Surgical	SS20-	A1	SS20-	A1-2	H&E FS Permanent		GYN Surg Service FS/TP Slides
mbedded Routine Surgical	SS20-	A10	SS20-	A10-1	HER2 IHC		GYN Surg Service FS/TP Slides
mbedded Routine Surgical	SS20-	A11	SS20-	A11-1	H&E		GYN Surg Service FS/TP Slides
mbedded Routine Surgical	SS20-	A12	SS20-	A12-1	H&E		GYN Surg Service FS/TP Slides
mbedded Routine Surgical	SS20-	A2	SS20-	A2-1	H&E		GYN Surg Service FS/TP Slides
mbedded Routine Surgical	SS20-	A3	SS20-	A3-1	HER2 IHC		GYN Surg Service FS/TP Slides
mbedded Routine Surgical	SS20-	A4	SS20-	A4-1	H&E		GYN Surg Service FS/TP Slides
mbedded Routine Surgical	SS20-	A5	SS20-	A5-1	H&E		GYN Surg Service FS/TP Slides

In this example of an LIS – what kind of data is the Task (nominial)
What kind of data is the priority – ordinal

Categorical



NATIONAL SOCIETY FOR HISTOTECHNOLOGY
Connecting. Empowering. Innovating.

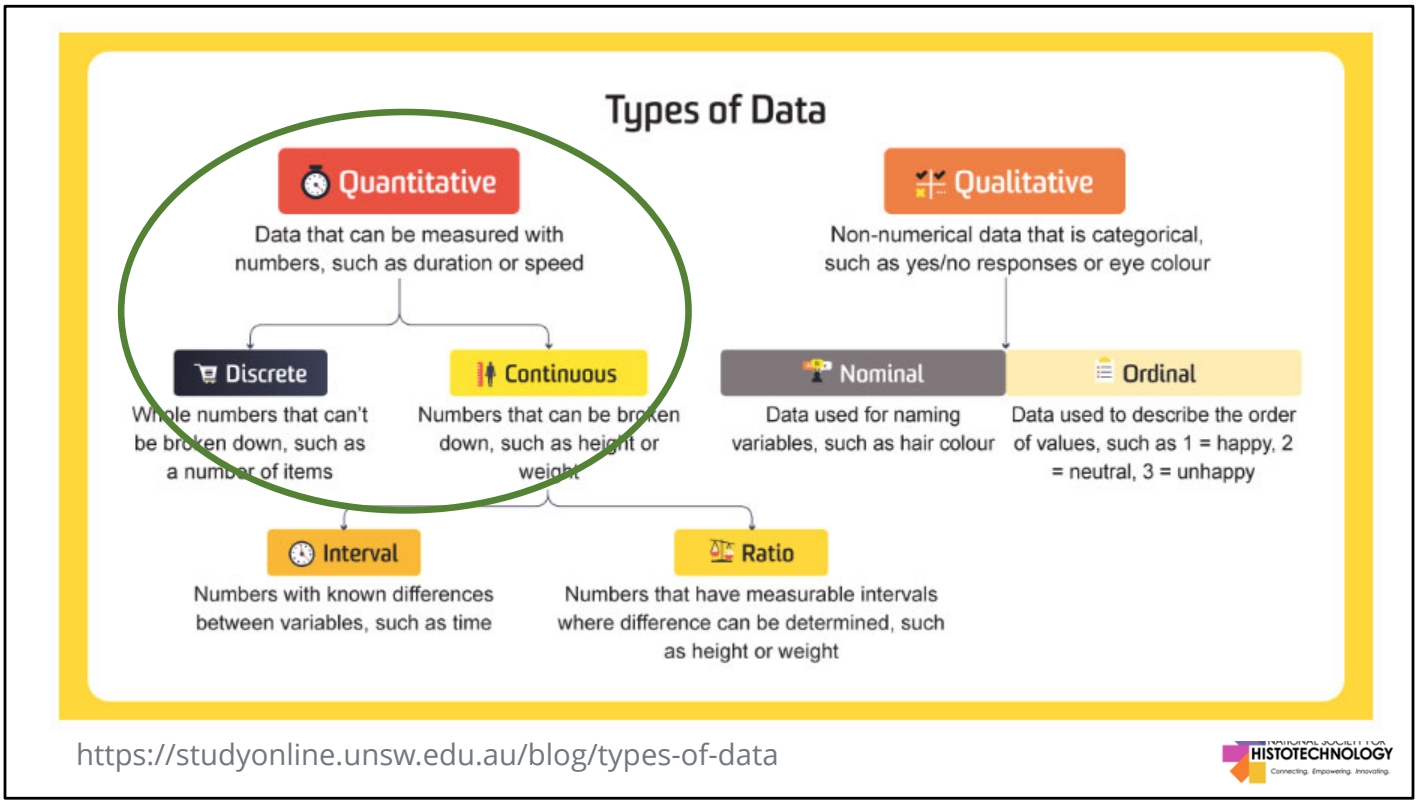
Is the Question Nominal or Ordinal?

1. Are you left handed or right handed?
2. How satisfied are you with your pizza delivery service? 1 – not satisfied, 2 – satisfied, 3 – very satisfied
3. What kind of house do you live in?
4. What kind of pet do you have?
5. Are you willing to work extra shifts? I am willing to work any extra shifts, I am willing to work extra shifts if given notice, I am willing to work extra shifts occasionally, I am not willing to work extra shifts.

Remember, nominal has no hierarchy…

NATIONAL SOCIETY FOR HISTOTECHNOLOGY
Connecting. Empowering. Innovating.

Write your response down, or type in the chat
Nominial
Ordinial
Nominal
Nominial
Ordinial



https://studyonline.unsw.edu.au/blog/types-of-data

Continuous data is also associated with two types of measurement – internval and ratio.

Discrete:

discrete adjective

dis·crete di-ˈskrēt ˈdis-

1 : constituting a separate entity : individually distinct

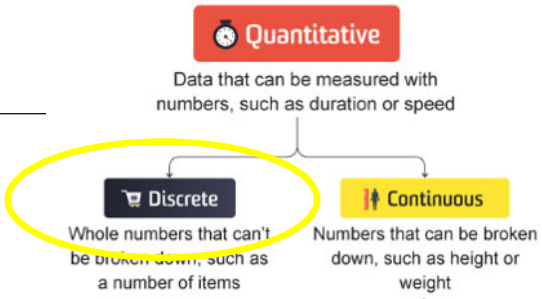
> several *discrete* sections

2 a : consisting of distinct or unconnected elements : NONCONTINUOUS

b : taking on or having a finite or countably infinite number of values

> *discrete* probabilities

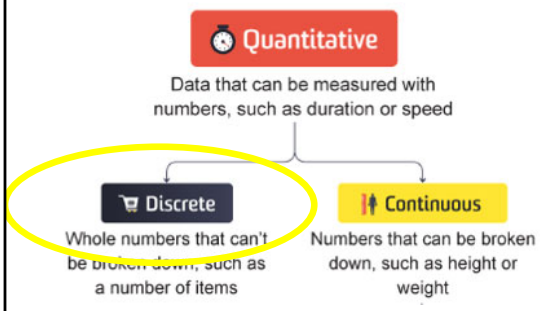
> a *discrete* random variable



NATIONAL SOCIETY FOR HISTOTECHNOLOGY
Connecting. Empowering. Innovating.

Discrete Data

Discrete data is numeric. They do not have to be whole numbers. **You count discrete data.**



Examples:

- Number of employees
- Number of IHC tests run
- Favorite ice cream flavor among a group (counted)
- Number of responses to a survey

NATIONAL SOCIETY FOR HISTOTECHNOLOGY
Connecting. Empowering. Innovating.

Its how we make categorical data count!

Discrete Data Example

Priority	Block		Container		Task	Lab Responsible Pathologist	Case Flags
Embedded Priority	PS20-	A1	PS20-	A1-1	H&E		RENAL
Embedded Priority	PS20-	A1	PS20-	A1-4	H&E		RENAL
Embedded Priority	PS20-	A1	PS20-	A1-7	H&E		RENAL
Embedded Priority	PS20-	A1	PS20-	A1-1	H&E		
Embedded Priority	PS20-	A1	PS20-	A1-4	ERG		
Embedded Priority	PS20-	A1	PS20-	A1-7	H&E		
Embedded Priority	PS20-	A1	PS20-	A1-1	H&E FS		
Embedded Priority	PS20-	A1	PS20-	A1-4	H&E		
Embedded Priority	PS20-	A1	PS20-	A1-7	H&E		
Embedded Priority	SS20-	A1	SS20-	A1-11	H&E		
Embedded Priority	SS20-	A1	SS20-	A1-13	H&E		
Embedded Priority	SS20-	A1	SS20-	A1-9	H&E		
Embedded Priority	SS20-	A1	SS20-	A1-1	ERG		
Embedded Priority	SS20-	A2	SS20-	A2-1	ERG		
Embedded Priority	SS20-	A3	SS20-	A3-1	H&E		Derm Service
Embedded Priority	SS20-	A4	SS20-	A4-1	H&E		Derm Service
Embedded Priority	SS20-	A5	SS20-	A5-1	H&E		Derm Service
Embedded Priority	SS20-	A6	SS20-	A6-1	ERG		Derm Service
Embedded Routine Surgical	SS20-	A1	SS20-	A1-2	H&E FS Permanent		GYN Surg Service FS/TP Slides
Embedded Routine Surgical	SS20-	A10	SS20-	A10-1	HER2 IHC		GYN Surg Service FS/TP Slides
Embedded Routine Surgical	SS20-	A11	SS20-	A11-1	H&E		GYN Surg Service FS/TP Slides
Embedded Routine Surgical	SS20-	A12	SS20-	A12-1	H&E		GYN Surg Service FS/TP Slides
Embedded Routine Surgical	SS20-	A2	SS20-	A2-1	H&E		GYN Surg Service FS/TP Slides
Embedded Routine Surgical	SS20-	A3	SS20-	A3-1	HER2 IHC		GYN Surg Service FS/TP Slides
Embedded Routine Surgical	SS20-	A4	SS20-	A4-1	H&E		GYN Surg Service FS/TP Slides

Goes from ordinal data to discrete!

Going back to this LIS data, we can calculate the frequency of the priority or the task. Then it goes from ordinal data to discrete!

Continuous:

continuous *adjective*

con·tin·u·ous kən-ˈtin-yü-əs

1 : marked by uninterrupted extension in space, time, or sequence

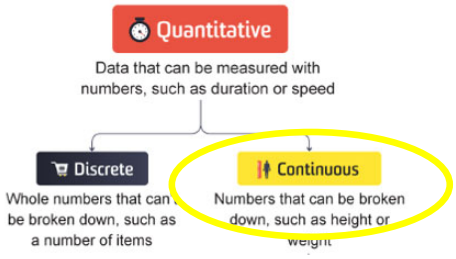
> The batteries provide enough power for up to five hours of *continuous* use.

2 **of a function** : having the property that the absolute value of the numerical difference between the value at a given point and the value at any point in a neighborhood of the given point can be made as close to zero as desired by choosing the neighborhood small enough

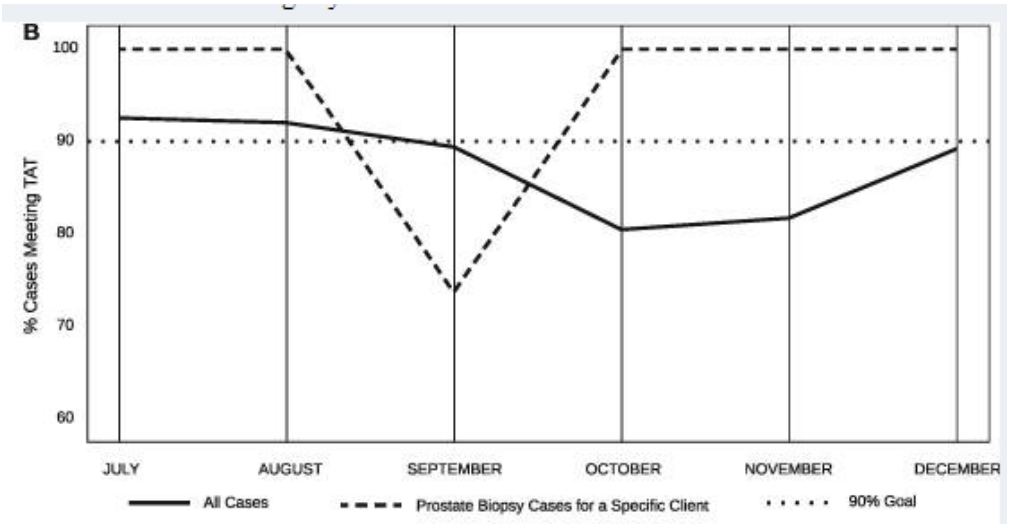
NATIONAL SOCIETY FOR HISTOTECHNOLOGY
Connecting. Empowering. Innovating.

Continuous Data

Continuous data is numeric and you measure it (height, weight, temperature). Its real defining factor is that it can be counted and change over time. It's the most complex, but also can be analyzed in the most number of ways!



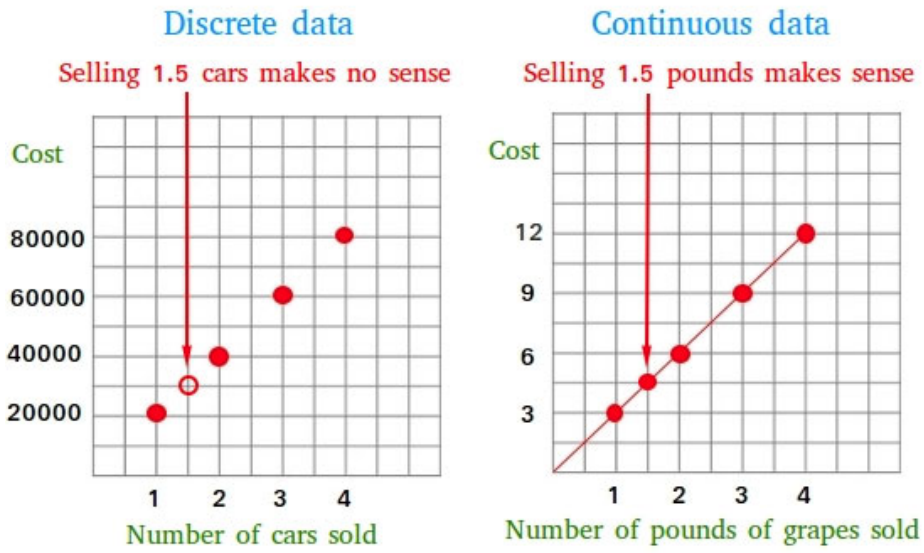
Continuous Data



Notice that we cannot connect the points since the numbers between 1 and 2, 2 and 3, 3 and 4 do not exist.

https://www.basic-mathematics.com/discrete-and-continuous-data.html

Discrete vs Continuous



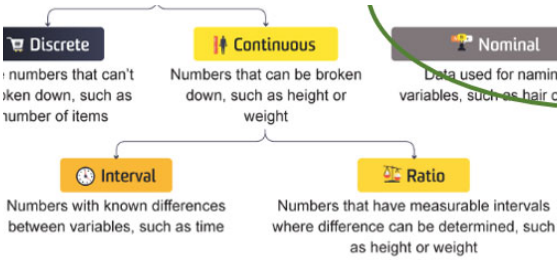
Identify if the following are Discrete or Continuous

1. The baby weighs 7 lbs 6 ounces pounds
2. There were 200 IHC requests this month
3. The lab was 65 degree Fahrenheit
4. The lab has 6 employees
5. It took the staff 2 days and 3 hours to complete the safety training.

Remember, counting vs measuring…

Continuous
Discrete
Continuous
Discrete
Continuous

Two Types of Continuous Data



Interval:

Year

Credit score

SAT test

Temperature (C or F)

Ratio:

Weight

Error rates

Crime rate

Length of time

Provides:	Nominal	Ordinal	Interval	Ratio
The "order" of values is known		✔	✔	✔
"Counts," aka "Frequency of Distribution"	✔	✔	✔	✔
Mode	✔	✔	✔	✔
Median		✔	✔	✔
Mean			✔	✔
Can quantify the difference between each value			✔	✔
Can add or subtract values			✔	✔
Can multiple and divide values				✔
Has "true zero"				✔

https://www.mymarketresearchmethods.com/types-of-data-nominal-ordinal-interval-ratio/

Part I: What is Data?

- ✓Summarize where data fits into the evaluation cycle.
- ✓Define data.
- ✓Identify what data is present and available to you.
- ✓Classify types of data.

Part 1 Recap!

Launch the quiz!!!

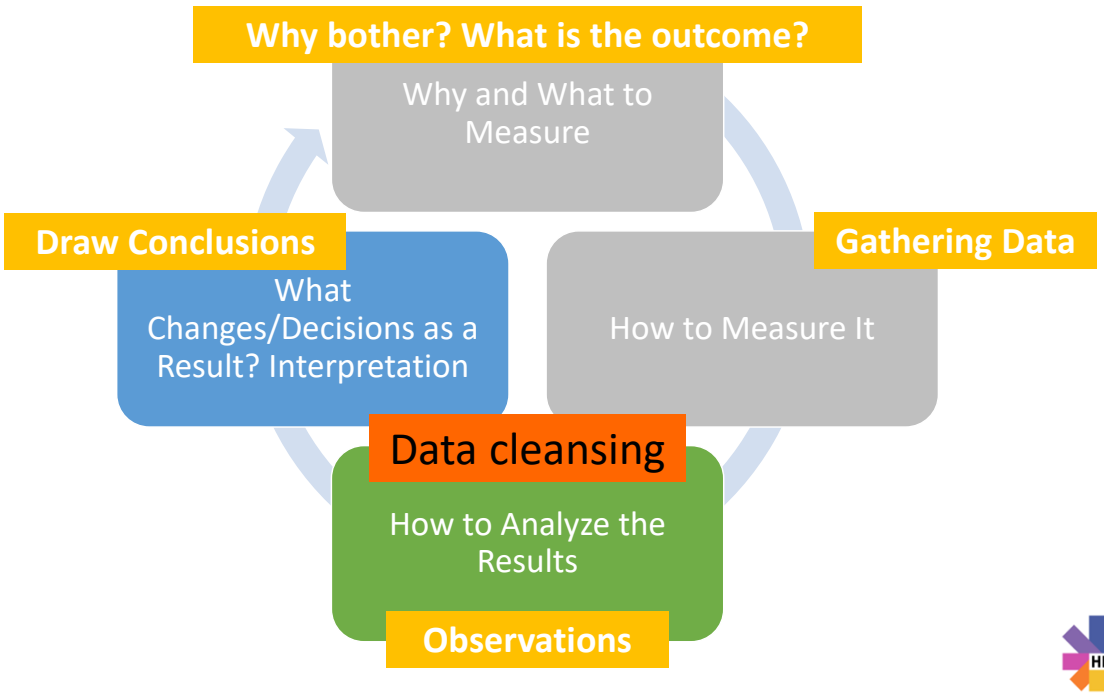
1. Determine purpose, Gather data, Analyze results, Interpret data
2. Qualitative
3. Discrete
4. Ordinal
5. True
6. Continuous

Part II: Preparing Data for Analysis

1. **Explain the impact of dirty data**
2. Dissect a data cleanse

Cleansing check list, activity with excel short cuts, use of find and replace, etc. Do a observation breakout with the data. 25 minutes – 10 lecture, 5 observation, 5 review. Take it back to quantifying the qualitative.

Evaluation Life Cycle



Preparing the Data: Cleansing

60-80% of a data scientists time is spent cleaning data.

- Data Scrubbing: "The procedure of modifying or removing incomplete, incorrect, inaccurately formatted, or repeated data in a database." (Technopedia).
- AKA: Cleaning, scrubbing, dirty data, unclean data
- Extractions of data or databases

These are the types of issues you may encounter…move to next slide

Example

Im trying to make a budget or staffing decision based on the testing numbers.

Test Cost

SU = $25.00

TC = $20.00

TN = $15.00

SU = 8, total cost $200.00

TC = 15, total cost $300.00

TN = 51, $765.00

So, I budget $1265

	Company	Testing Site	Unique ID	Test Date	State	Test
	2337	655715	SU25791	6/3/2022	State=FL	SU
	2337	662695	SU25791	6/3/2022	State=FL	SU
	2337	662699	SU47021	5/26/2022	State=FL	SU
	2337	687630	TC48255	6/12/2022	StateFL	TC
	2337	759333	TC48255	6/12/2022	State=FL	TC
	2337	770665	TC48255	6/12/2022	State=FL	TC
	2337	770677	TC51146	4/2/2022	State=FL	TC
	2337	810308	TC51146	6/14/2022	State=FL	TC
0	2337	810310	TN47756	6/10/2022	State=FL	TN
1	2337	813164	TN47756	6/11/2022	State=FL	TN
2	2337	813168	TN47756	4/1/2022	State=FL	TN
3	2337	813170	TN47756	4/24/2022	State=FL	SU
4	2337	813172	TN47756	4/1/2022	State=FL	SU
5	2337	813174	TN47756	6/17/2022	State=FL	TN
6	2337	813176	TN47756	6/18/2022	State=FL	TN
7	2337	813246	TN51377	5/27/2022	State=FL	TN
8	2337	813250	TN51377	5/25/2022	State=FL	TN
9	2337	813252	TN51377	5/24/2022	State=FL	TN
0	2337	813256	tn51377	5/26/2022	State=FL	TN
1	2337	813264	TN51377	5/24/2022	State=FL	TN
2	2337	813268	TN51377	5/25/2022	State=FL	TN
3	2337	813270	TN51377	5/25/2022	State=FL	TN
4	2337	813274	TN51377	5/27/2022	State=FL	TN
5	2337	813276	TN51377	5/26/2022	State=FL	TN
6	2337	842802	TN51377	5/27/2022	StateFL	TN
7	2337	842806	TN51377	5/26/2022	State=FL	TN
8	2337	842810	TN51377	5/26/2022	State=FL	TC
9	2337	842812	TN47756	6/10/2022	State=FL	TC
0	2337	870834	TN51377	5/14/2022	State=FL	TC

So I count them, or the frequency. And here is what I come up with.

But…

Duplicates!

SU = 5, total cost $125.00
TC = 10, total cost $200.00
TN = 35, $525.00

SU = $25.00
TC = $20.00
TN = $15.00

So, I budgeted $1265 BUT it really only cost $850.00. That could impact so many other parts of my budget. Not to mention staffing and time allotment.

After removing them the total counts go down! And maybe, that SU test requires some higher level testing that only 1-2 staff members can perform and based on the old total of 8 I bring in someone extra…or ask them to work when they don't really need to. This is all hypothetical, but you get the idea.

Part II: Preparing Data for Analysis

1. Explain the impact of dirty data
2. **Dissect a data cleanse**

Cleansing check list, activity with excel short cuts, use of find and replace, etc. Do a observation breakout with the data. 25 minutes – 10 lecture, 5 observation, 5 review. Take it back to quantifying the qualitative.

Connie's Customized Checklist

1. Remove duplicates (if you are not counting frequency)

As a reminder, this can take a long time! A keen attention to detail is needed. When you are having conversation about "what the data is telling you" - be sure you are calculating any time for data cleaning.

Duplicate Values

- When you have one qualified response entered more than one time.
- CRMs, CMXs, etc typically have a duplicate value finder or preventer – but they are far from perfect.

Tip: Have protocols/SOPs in place to remove duplicates. Staff turnover and people unfamiliar with your database may not understand what to do when they encounter a duplicate and that can cause dirty data!

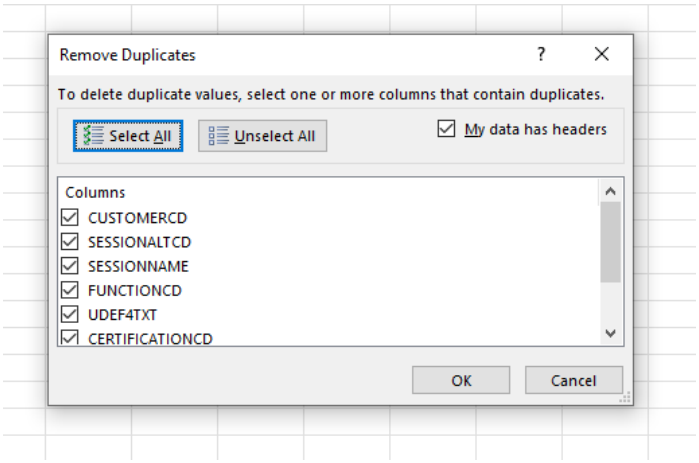
When dealing with CRM data, put a protocol in place to remove duplcaited

000083539I	SC2021	47th NSH	WS10	Histogonc	NSH	09/14/21	1	A	Allison Eck; Brad Flowers
000083539I	SC2021	47th NSH	WS11	Adhesive	NSH	09/14/21	1	A	David Prine
000083539I	SC2021	47th NSH	WS21	Diagnosti	NSH	15-Sep	1	A	Richard Ormesher
000083539I	SC2021	47th NSH	WS32	Discovery	NSH	09/16/21	1	A	Kim Pickard; Michele Levitt
000083539I	SC2021	47th NSH	WS40	Using The	NSH	09/16/21	1	A	Gerelyn Henry
000083539I	SC2021	47th NSH	WS11	Adhesive	NSH	09/14/21	1	A	David Prine
000083539I	SC2021		YOGADAY	Morning '	NSH	09/15/21	1	A	Gerelyn Henry
000083539I	SC2021	47th NSH	YOGADAY	Morning '	NSH	09/15/21	1	A	Gerelyn Henry
000083539I	SC2021	47th NSH	YOGADAY	Morning '	NSH	09/16/21	1	A	Gerelyn Henry
000082078I	SC2021	47th NSH	ET05	Stains Be	NSH	09/15/21	0.5	A	Jean Mitchell; Surena Becraft
	SC2021	47th NSH	ET01	Modificat	NSH	09/21/21	0.5	A	Elizabeth Chlipala
	SC2021	47th NSH	ET01		NSH	09/25/21	0.5	A	Elizabeth Chlipala
	SC2021	47th NSH	ET01	Modificat	NSH	09/27/21	0.5	A	Elizabeth Chlipala
	SC2021	47th NSH	ET01	Modificat	NSH	09/30/21	0.5	A	Elizabeth Chlipala
	SC2021	47th NSH	ET02	Chemical	NSH	09/15/21	0.5	A	Steven Goodman
	SC2021	47th NSH	ET02	Chemical	NSH	09/15/21	0.5	A	S Goodman
	SC2021	47th NSH	ET02	Chemical	NSH	09/20/21	0.5	A	
	SC2021	47th NSH	ET02	Chemical	NSH	09/24/21	0.5	A	Steven Goodman
	SC2021	47th NSH	ET02	Chemical	NSH	09/25/21	0.5	A	Steven Goodman
	SC2021	47th NSH	ET02	Chemical	NSH	09/27/21	0.5	A	Steven Goodman
	SC2021	47th NSH	ET02	Chemical	NSH	09/27/21	0.5	A	Steven Goodman

NATIONAL SOCIETY FOR HISTOTECHNOLOGY
Connecting. Empowering. Innovating.

Show them how you do it in excel.

Excel Trick: Remove Duplicates



Connie's Customized Checklist

1. Remove duplicates (if you are not counting frequency)
2. Identify formatting issues:
   - Formatting for quantitative entries like dates, decimals
   - Formatting for categorical information like states, names, etc.

As a reminder, this can take a long time! A keen attention to detail is needed. When you are having conversation about "what the data is telling you" - be sure you are calculating any time for data cleaning.

Formatting Issues

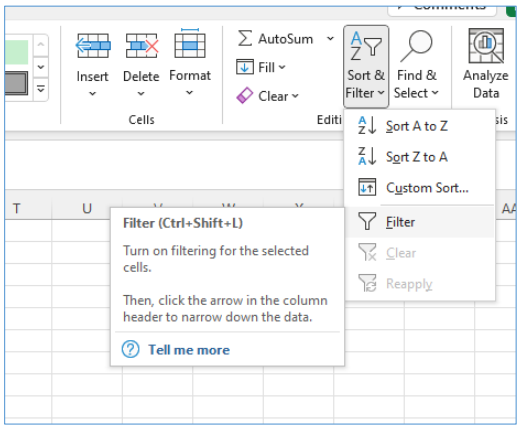
- Inconsistencies with data fields.
- Open ended fields will always need attention.

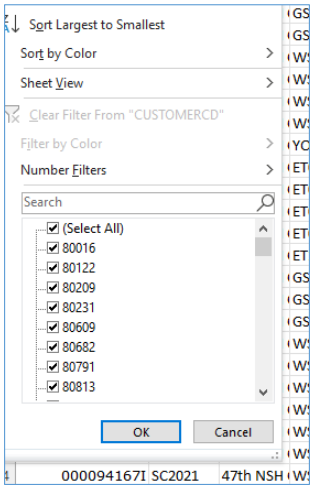
Tip: Consider drop downs, and prefilled items when you can. Test with people to see if there are issues that are preventable (i.e. are dropdowns missing fields, etc.)

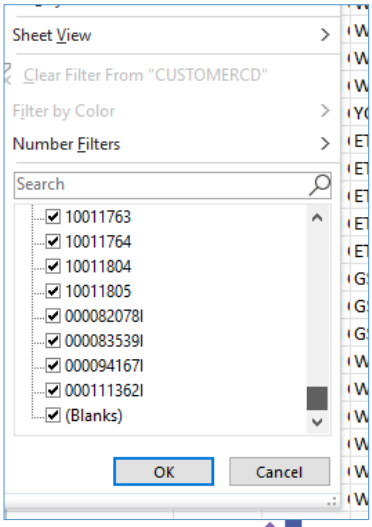
000083539I	SC2021	47th NSH	WS10	Histogonc	NSH	09/14/21	1	A	Allison Eck; Brad Flowers
000083539I	SC2021	47th NSH	WS11	Adhesive	NSH	09/14/21	1	A	David Prine
000083539I	SC2021	47th NSH	WS21	Diagnosti	NSH	15-Sep	1	A	Richard Ormesher
000083539I	SC2021	47th NSH	WS32	Discovery	NSH	09/16/21	1	A	Kim Pickard; Michele Levitt
000083539I	SC2021	47th NSH	WS40	Using The	NSH	09/16/21	1	A	Gerelyn Henry
000083539I	SC2021	47th NSH	WS11	Adhesive	NSH	09/14/21	1	A	David Prine
000083539I	SC2021		YOGADAY	Morning \	NSH	09/15/21	1	A	Gerelyn Henry
000083539I	SC2021	47th NSH	YOGADAY	Morning \	NSH	09/15/21	1	A	Gerelyn Henry
000083539I	SC2021	47th NSH	YOGADAY	Morning \	NSH	09/16/21	1	A	Gerelyn Henry
000082078I	SC2021	47th NSH	ET05	Stains Be	NSH	09/15/21	0.5	A	Jean Mitchell; Surena Becraft
	SC2021	47th NSH	ET01	Modificat	NSH	09/21/21	0.5	A	Elizabeth Chlipala
	SC2021	47th NSH	ET01		NSH	09/25/21	0.5	A	Elizabeth Chlipala
	SC2021	47th NSH	ET01	Modificat	NSH	09/27/21	0.5	A	Elizabeth Chlipala
	SC2021	47th NSH	ET01	Modificat	NSH	09/30/21	0.5	A	Elizabeth Chlipala
	SC2021	47th NSH	ET02	Chemical	NSH	09/15/21	0.5	A	Steven Goodman
	SC2021	47th NSH	ET02	Chemical	NSH	09/15/21	0.5	A	S Goodman
	SC2021	47th NSH	ET02	Chemical	NSH	09/20/21	0.5	A	
	SC2021	47th NSH	ET02	Chemical	NSH	09/24/21	0.5	A	Steven Goodman
	SC2021	47th NSH	ET02	Chemical	NSH	09/25/21	0.5	A	Steven Goodman
	SC2021	47th NSH	ET02	Chemical	NSH	09/27/21	0.5	A	Steven Goodman
	SC2021	47th NSH	ET02	Chemical	NSH	09/27/21	0.5	A	Steven Goodman

NATIONAL SOCIETY FOR HISTOTECHNOLOGY
Connecting. Empowering. Innovating.

Excel Trick: Filter or Sort





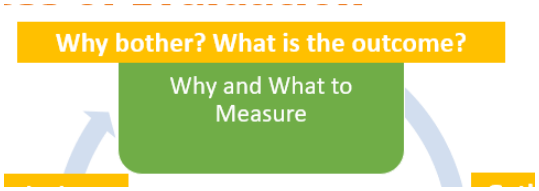


Connie's Customized Checklist

1. Remove duplicates (if you are not counting frequency)
2. Identify formatting issues:
   - Formatting for quantitative entries like dates, decimals
   - Formatting for categorical information like states, names, etc.
3. Find missing data and fill in or remove.

As a reminder, this can take a long time! A keen attention to detail is needed. When you are having conversation about "what the data is telling you" - be sure you are calculating any time for data cleaning.

Missing Data



- Data fields that are not complete.
- "Missing data can reduce the statistical power of a study and can produce biased estimates, leading to invalid conclusions."

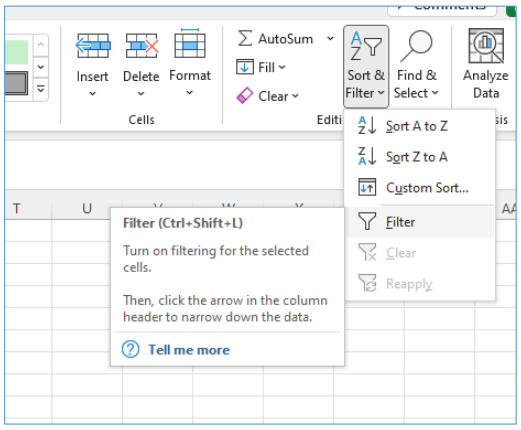
Tip: Collection desires vs realities (think about the amount of time it would take a person to enter fully the information, or accurately)

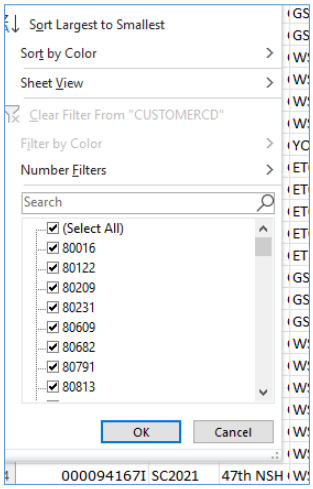
This leads us to the concept that we need to be realistic about our questions. If we are finding that we are missing data, and it's a lot, maybe we made a mistake in asking the question.

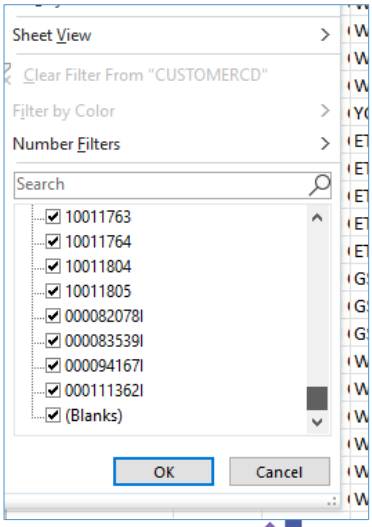
000083539I	SC2021	47th NSH	WS10	Histogonc	NSH	09/14/21	1	A	Allison Eck; Brad Flowers
000083539I	SC2021	47th NSH	WS11	Adhesive	NSH	09/14/21	1	A	David Prine
000083539I	SC2021	47th NSH	WS21	Diagnosti	NSH	15-Sep	1	A	Richard Ormesher
000083539I	SC2021	47th NSH	WS32	Discovery	NSH	09/16/21	1	A	Kim Pickard; Michele Levitt
000083539I	SC2021	47th NSH	WS40	Using The	NSH	09/16/21	1	A	Gerelyn Henry
000083539I	SC2021	47th NSH	WS11	Adhesive	NSH	09/14/21	1	A	David Prine
000083539I	SC2021		YOGADAY	Morning \	NSH	09/15/21	1	A	Gerelyn Henry
000083539I	SC2021	47th NSH	YOGADAY	Morning \	NSH	09/15/21	1	A	Gerelyn Henry
000083539I	SC2021	47th NSH	YOGADAY	Morning \	NSH	09/16/21	1	A	Gerelyn Henry
000082078I	SC2021	47th NSH	ET05	Stains Be	NSH	09/15/21	0.5	A	Jean Mitchell; Surena Becraft
	SC2021	47th NSH	ET01	Modificat	NSH	09/21/21	0.5	A	Elizabeth Chlipala
	SC2021	47th NSH	ET01		NSH	09/25/21	0.5	A	Elizabeth Chlipala
	SC2021	47th NSH	ET01	Modificat	NSH	09/27/21	0.5	A	Elizabeth Chlipala
	SC2021	47th NSH	ET01	Modificat	NSH	09/30/21	0.5	A	Elizabeth Chlipala
	SC2021	47th NSH	ET02	Chemical	NSH	09/15/21	0.5	A	Steven Goodman
	SC2021	47th NSH	ET02	Chemical	NSH	09/15/21	0.5	A	S Goodman
	SC2021	47th NSH	ET02	Chemical	NSH	09/20/21	0.5	A	
	SC2021	47th NSH	ET02	Chemical	NSH	09/24/21	0.5	A	Steven Goodman
	SC2021	47th NSH	ET02	Chemical	NSH	09/25/21	0.5	A	Steven Goodman
	SC2021	47th NSH	ET02	Chemical	NSH	09/27/21	0.5	A	Steven Goodman
	SC2021	47th NSH	ET02	Chemical	NSH	09/27/21	0.5	A	Steven Goodman

NATIONAL SOCIETY FOR HISTOTECHNOLOGY
Connecting. Empowering. Innovating.

Excel Trick: Filter or Sort







Connie's Customized Checklist

1. Remove duplicates (if you are not counting frequency)
2. Identify formatting issues:
   - Formatting for quantitative entries like dates, decimals
   - Formatting for categorical information like states, names, etc.
3. Find missing data and fill in or remove.
4. Remove duplicates again!

NAL SOCIETY FOR HISTOTECHNOLOGY
Connecting. Empowering. Innovating.

As a reminder, this can take a long time! A keen attention to detail is needed. When you are having conversation about "what the data is telling you" - be sure you are calculating any time for data cleaning.

CUSTOMERCD	CONTACT	CITY	STATECD	ZIP	ISMEMBERFLG	CUSTOMERTYPE	STATUSSTT
10013333		Orland Park	IL	60462	N	NONMEMBER	ACTIVE
10012128		fergana		15011	N	NONMEMBER	ACTIVE
222905		Glendale	CA	91201	N	NONMEMBER	Active
201568		Needham	MA	2494	N	NONMEMBER	Active
10012018		bronx	NY	10453	N	NONMEMBER	ACTIVE
10001892		Reading	MA	1867	N	NONMEMBER	ACTIVE
10000679		Omdrman			N	NONMEMBER	ACTIVE
145814		San Jose		11501	N	NONMEMBER	Active
10012780		Enugu		400001	N	NONMEMBER	ACTIVE
10011882		Spotsylvania	VIC	22551	N	NONMEMBER	ACTIVE
112211		Lafayette	LA	70503-2573	Y	ENHANCED	Active
10013780		Seattle	WA	98119	N	NONMEMBER	ACTIVE
180270		SOUTH SAN FRANCISCO	CA	94080	N	NONMEMBER	Active
223196		Washington	DC	20007-2126	N	CORE	FORMER
10011949		Beaumont	CA	92223	N	NONMEMBER	ACTIVE
10003077		Springfield	MO	65804	N	NONMEMBER	ACTIVE
190551		San Diego	CA	92122-4586	N	CORE	FORMER
10012022		Gaffney	SC	29341	N	NONMEMBER	ACTIVE
109680		Downey	CA	90242-2041	Y	ENHANCED	Active
10001940		Richmond	TX	77069	N	NONMEMBER	ACTIVE
10012147		Khartoum	BC	12345	N	NONMEMBER	ACTIVE
109013		karachi		75190	N	NONMEMBER	Active
10011862		Risharon	TX	77583	N	NONMEMBER	ACTIVE
101728		Madison	WI	53527	N	CORE	FORMER
10013961		Calabar		54201	N	NONMEMBER	ACTIVE
10001661		Sydney, Fairfield	ACT	2165	N	NONMEMBER	ACTIVE
10011232		Auckland	OC		N	NONMEMBER	ACTIVE
10014787		Spokane Valley	WA	99037	Y	ENHANCED	Active
10012275		Columbia	TN	38401	Y	CORE	Active

What's Wrong With This Picture?

NATIONAL SOCIETY FOR HISTOTECHNOLOGY
Connecting. Empowering. Innovating.

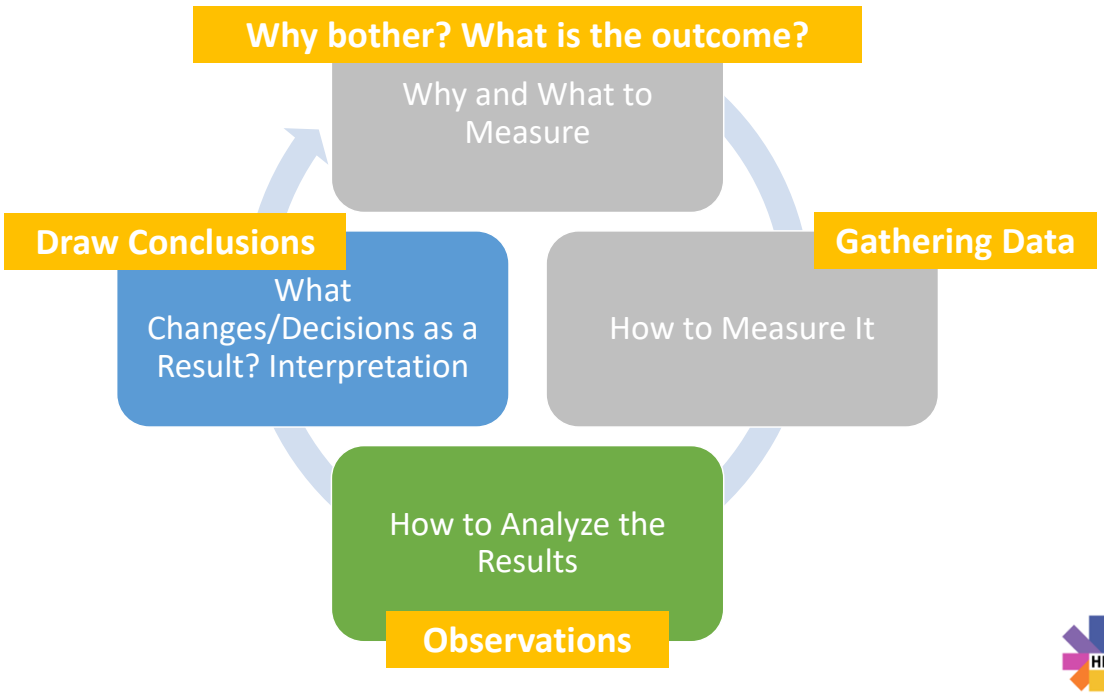
Missing state
Formatting of zip
Font sizes
Formatting of certain words
Suppose I don't want international records how can I fix that?

Part III: Analyzing Data

1. **Interpret measures of central tendency.**
2. Analyze data

Cleansing check list, activity with excel short cuts, use of find and replace, etc. Do a observation breakout with the data. 25 minutes – 10 lecture, 5 observation, 5 review. Take it back to quantifying the qualitative.

Evaluation Life Cycle



Two Types of Quantitative Analysis

- Descriptive: focus on describing a sample of a population.
- Inferential: makes predictions based on the sample about the ENTIRE population.

"In other words, we use one group of statistical methods – descriptive statistics – to investigate the slice of cake, and another group of methods – inferential statistics – to draw conclusions about the entire cake."

-Karryn Warren, PhD

NATIONAL SOCIETY FOR HISTOTECHNOLOGY
Connecting. Empowering. Innovating.

When a research question and/or a design created the type of analysis matters. There are two types. Desriptive and Inferential. For our purpose we will be focusing on descriptive statistics. If you would like to read more, I would suggest an evaluation book, like insert name Research Design. If you think about. If you are going cake level, you are likely going to be working with an evaluator or statistician and they will stil provide you with descriptive statsistic, but then the insight for the infnerential.

For example, car color and accidents

Central Tendency

- A measure of central tendency is a single value that attempts to describe a set of data by identifying the central position within that set of data.
- A single measure may not give you the full picture because extreme values can impact each measurement.

Frequency

The rate at which something occurs or is repeated (over time or in a sample). For example, Embed Priority appears 18 times.

Priority	Block		Container		Task	Lab Responsible Pathologist	Case Flags
Embedded Priority	PS20-	A1	PS20-	A1-1	H&E		RENAL
Embedded Priority	PS20-	A1	PS20-	A1-4	H&E		RENAL
Embedded Priority	PS20-	A1	PS20-	A1-7	H&E		RENAL
Embedded Priority	PS20-	A1	PS20-	A1-1	H&E		RENAL
Embedded Priority	PS20-	A1	PS20-	A1-4	ERG		RENAL
Embedded Priority	PS20-	A1	PS20-	A1-7	H&E		RENAL
Embedded Priority	PS20-	A1	PS20-	A1-1	H&E FS		RENAL
Embedded Priority	PS20-	A1	PS20-	A1-4	H&E		RENAL
Embedded Priority	PS20-	A1	PS20-	A1-7	H&E		RENAL
Embedded Priority	SS20-	A1	SS20-	A1-11	H&E		ENT Service
Embedded Priority	SS20-	A1	SS20-	A1-13	H&E		ENT Service
Embedded Priority	SS20-	A1	SS20-	A1-9	H&E		ENT Service
Embedded Priority	SS20-	A1	SS20-	A1-1	ERG		Derm Service
Embedded Priority	SS20-	A2	SS20-	A2-1	ERG		Derm Service
Embedded Priority	SS20-	A3	SS20-	A3-1	H&E		Derm Service
Embedded Priority	SS20-	A4	SS20-	A4-1	H&E		Derm Service
Embedded Priority	SS20-	A5	SS20-	A5-1	H&E		Derm Service
Embedded Priority	SS20-	A6	SS20-	A6-1	ERG		Derm Service
Embedded Routine Surgical	SS20-	A1	SS20-	A1-2	H&E FS Permanent		GYN Surg Service FS/TP Slides
Embedded Routine Surgical	SS20-	A10	SS20-	A10-1	HER2 IHC		GYN Surg Service FS/TP Slides
Embedded Routine Surgical	SS20-	A11	SS20-	A11-1	H&E		GYN Surg Service FS/TP Slides
Embedded Routine Surgical	SS20-	A12	SS20-	A12-1	H&E		GYN Surg Service FS/TP Slides
Embedded Routine Surgical	SS20-	A2	SS20-	A2-1	H&E		GYN Surg Service FS/TP Slides

NATIONAL SOCIETY FOR HISTOTECHNOLOGY
Connecting. Empowering. Innovating.

Mean

- Average
- Total all numbers in the data set and divide by the number of responses.
- **Very sensitive to outliers**

Example:

20, 16, 15, 12, 22, 13, 16, 19, 31, 32, 12, 16, 20, 20, 16

Excel Formula
=AVERAGE(A2:A16)

Balanced Distribution: Use Mean

NATIONAL SOCIETY FOR HISTOTECHNOLOGY
Connecting. Empowering. Innovating.

Median

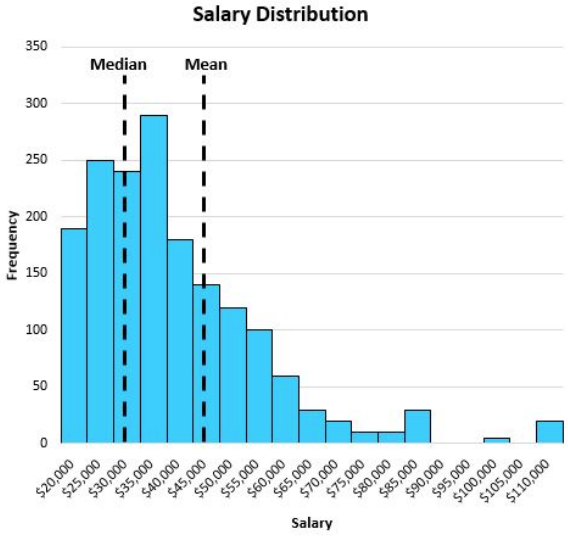
- The value that appears in the center of the data set.
- If the number of data points is odd, the median is the middle data point in the list.
- If the number of data points is even, the median is the average of the two middle data points in the list (add the two middle values and divide by 2)
- **Very useful if there are extreme outliers**

Example:

12, 12, 13, 15, 16, 16, 16, 16, 18, 19, 20, 20, 20, 20, 21, 31, 32

16 + 18 = **17**

Not Balanced Distribution: Use Median



NATIONAL SOCIETY FOR HISTOTECHNOLOGY
Connecting. Empowering. Innovating.

Mode

- The number that appears the most in the data set. You can have more than one mode.
- **Mode is best with large datasets because smaller data sets may not even have a mode!**

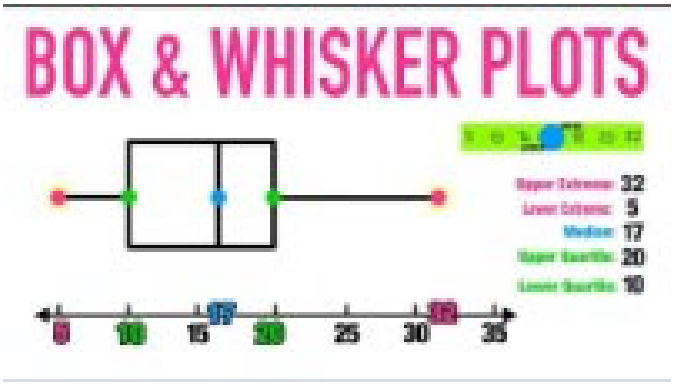
Example:

202, 116, 115, 112, 212, 113, 116, 119, 319, 320, 120, 126, 202, 201, 132

Why Central Tendency Matters

13: Other (please describe):Which type of POL/Specialty Lab?	14: Non-Registered HT/HTL:How many full time personnel work in your lab with each of the following job functions? Leave blank if no staff perform a particular function.	14: Registered HT/HTL:How many full time personnel work in your lab with each of the following job functions? Leave blank if no staff perform a particular function.	14: Laboratory Assistant:How many full time personnel work in your lab with each of the following job functions? Leave blank if no staff perform a particular function.
MEDIAN	2	4	2
MODE	1	1	1
MEAN	4.158054711	6.45	4.506527415

Box and Whisker Plots!



https://www.youtube.com/watch?v=fJZv9YeQ-qQ

Range

- Describes the difference between the smallest and largest value and describes how well the central tendency represents the data. If the range is large, the central tendency is not as representative of the data as it would be if the range was small.
- Sensitive to outliers

Example:

210, 116, 115, 112, 212, 113, 116, 119, 319, 320, 120, 126, 202, 301, 136

Quiz Time!

1. Range
2. Arrange in ascending order
3. True
4. Median
5. 5
6. True
7. Descriptive

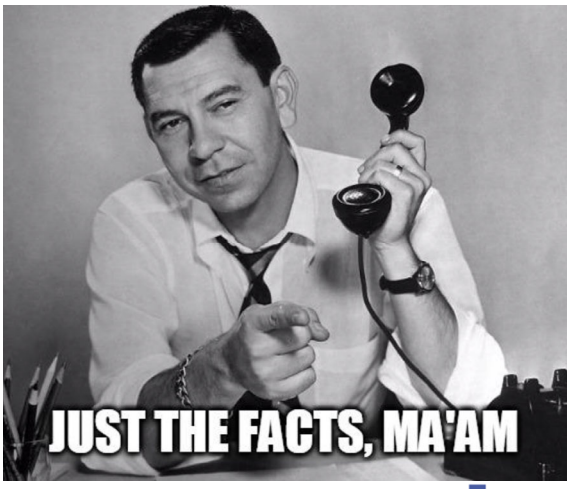
Part III: Analyzing Data

1. Interpret measures of central tendency.
2. **Analyze data**

Cleansing check list, activity with excel short cuts, use of find and replace, etc. Do a observation breakout with the data. 25 minutes – 10 lecture, 5 observation, 5 review. Take it back to quantifying the qualitative.

OBSERVATION - Analysis

- Observation = FACTS
- Analysis is just what the numbers or information say.



This is the step that people tend to skip. They go right from the assessment to making assumptions and rationalizing. And that's not good because you lose information. We focus in on one data point and move on…but the story is usually much bigger.

ASCP BOC QUALIFICATION STATISTICS: 2006 - 2021

QUALIFICATION TYPE	YEAR	MEAN	RANGE OF SCORES		TOTAL # TAKING QUALIFICATION	TOTAL PASS		TOTAL FAIL		TOTAL QUALIFIED
IMMUNOHISTO CHEMISTRY (QIHC)	2006	465	240	640	60	42	70%	18	30%	1822 (First Year Qualified 1994)
	2007	485	260	680	52	42	81%	10	19%	
	2008	488	260	820	70	58	83%	12	17%	
	2009	471	300	680	76	62	82%	14	18%	
	2010	499	270	680	88	80	91%	8	9%	
	2011	463	210	680	84	69	82%	15	18%	
	2012	475	220	680	104	84	81%	20	19%	
	2013	433	220	820	123	74	60%	49	40%	
	2014	430	100	680	133	72	54%	61	46%	
	2015	455	217	906	113	57	50%	56	50%	
	2016	459	250	814	124	66	53%	58	47%	
	2017	389	212	759	131	46	35%	85	65%	
	2018	410	189	759	135	57	42%	78	58%	
	2019	428	236	759	135	67	50%	68	50%	
	2020	418	212	759	131	64	49%	67	51%	
	2021	408	217	749	141	50	35%	91	65%	

- More test takers passed the QIHC by percent in 2010 than any other year measured. (continuous data example)
- The total number of test takers in 2021 was 141. (discrete data example)

Remember, counting vs measuring…

NATIONAL SOCIETY FOR HISTOTECHNOLOGY
Connecting. Empowering. Innovating.

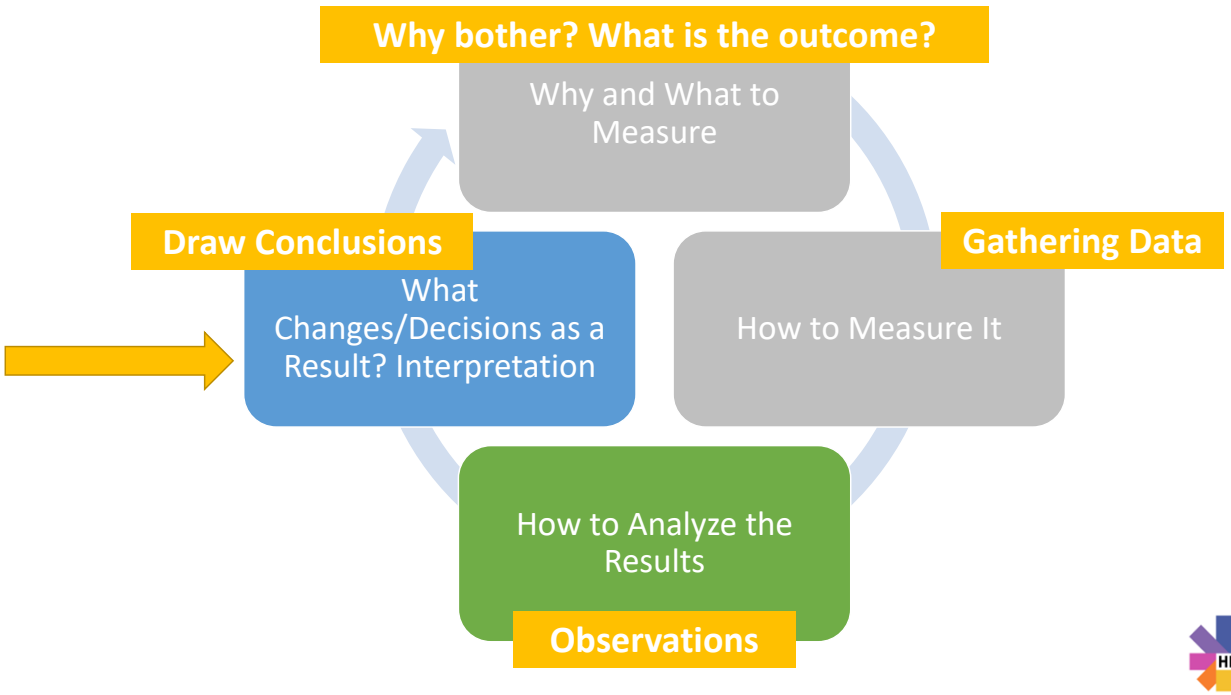
Is this qualitative or quantitive data?
Is this

Why can't I make conclusions like...

The QIHC exam is too hard.

- Because the data doesn't show us that. Maybe the test takers didn't have enough study tools, maybe the technology is outdated for taking the exam.
- More importantly, we aren't at that step yet!

Evaluation Life Cycle



Instructions

- Looking at the data set, type in an observation about the data.

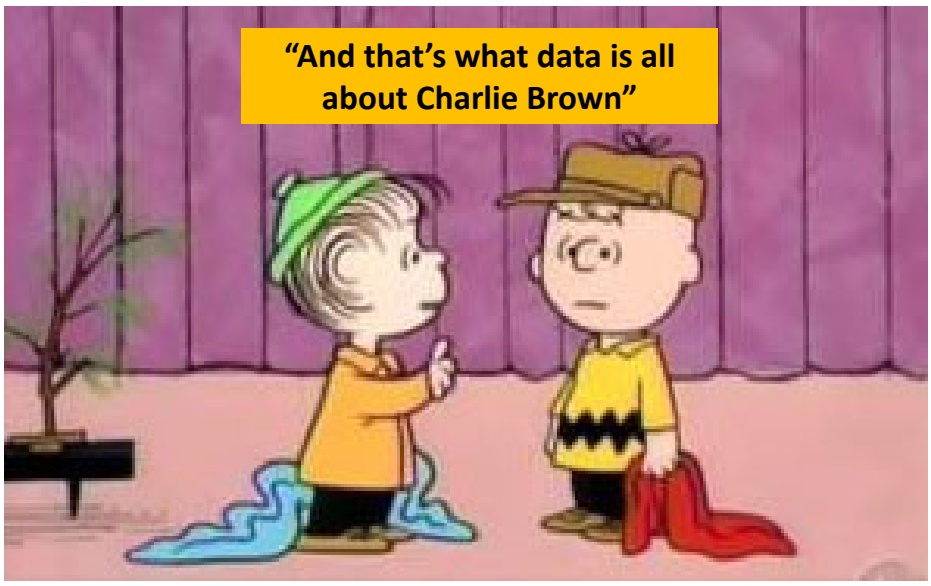
197
Sessionname (C), and status (I), function code (D)
Statis (I), cert date g

	Total	Hospital	Private Independent
Sample Size	1,088	809	279
Grossing Questions			
Who performs grossing in your lab?			
Non-Registered HT/HTL	9%	7%	15%
Registered HT/HTL	31%	26%	44%
Laboratory Assistant	3%	2%	5%
Laboratory Technician	8%	7%	11%
Grossing Assistant	22%	20%	27%
Supervisor	9%	9%	11%
Manager	3%	3%	3%
Pathologist	31%	35%	21%
Pathologist's Assistant	53%	62%	25%
Resident	12%	16%	2%
Other (please specify)	2%	2%	2%
What are the average number of cassettes grossed per hour?			
Mean	46.9	46.1	48.7
Median	30.0	30.0	35.0
What are the average number of containers grossed per hour?			
Mean	27.6	23.3	36.1

NATIONAL SOCIETY FOR ISTOTECHNOLOGY
Connecting. Empowering. Innovating.

Diversity of Thought Makes Data Better!

- We all look at things differently, even numbers
- Using many sets of eyes and brains can analyze data better



NATIONAL SOCIETY FOR HISTOTECHNOLOGY
Connecting. Empowering. Innovating.

For Your Reading Pleasure

- https://uxdesign.cc/data-for-design-f33fd7419cc8
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7886543/
- https://towardsdatascience.com/9-common-mistakes-beginner-data-scientists-make-91255ddd1311
- Excel Formula Tutorial: https://www.youtube.com/watch?v=KYALCR6Bdxk

References

- Evaluation, A Systematic Approach: Seventh Edition, Rossi, Lipsey, and Freeman
- DataFloq, 10 Really Cool Data Cartoons You Have to See! https://datafloq.com/read/10-really-cool-data-cartoons-you-have-to-see/
- Merriam-Webster Dictionary, 2022
- ASCP QIHC Pass Rates – published Jan 18, 2022
- 2022 NSH Workload Study
- Kang H. The prevention and handling of the missing data. Korean J Anesthesiol. 2013 May;64(5):402-6. doi: 10.4097/kjae.2013.64.5.402. Epub 2013 May 24. PMID: 23741561; PMCID: PMC3668100.
- Seifert RP, Casler V, Al Qaysi N, Gothi SR, Williams L, Christensen PR, Flax S, Chamala S. Informatics driven quality improvement in the modern histology lab. JAMIA Open. 2020 Nov 30;3(4):530-535. doi: 10.1093/jamiaopen/ooaa066. PMID: 33623889; PMCID: PMC7886543.