



Week 1: Intro; Data Basics



Professor Kathryn Jacobs



To Do

- Explain class structure; explore canvas
- Stats vocab lesson!
- Experimental design

To Do

- Explain class structure; explore canvas
- Stats vocab lesson!
- Experimental design

A little about me....



Class Structure

Lecture 3x week, Lab 1x

All assignments accessed via Lab canvas site

Lecture

- Textbook based
- Adaptive Assignments
- Exams

Lab

- Data based
- Pre-lab and Group Lab assignments

Canvas

- Syllabus (schedule)
- Textbook
- Lecture slides
- Additional materials (youtube links, worksheets)

To Do

- Explain class structure; explore canvas
- Stats vocab lesson!
- Experimental design

Data!

“Data Matrix” - vocab used in books and Adaptive Assignments

	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Global_Sales
1	1	Wii Sports	Wii	2006	Sports	Nintendo	41.49	29.02	3.77	82.74
2	2	Super Mario Bros.	NES	1985	Platform	Nintendo	29.08	3.58	6.81	40.24
3	3	Mario Kart Wii	Wii	2008	Racing	Nintendo	15.85	12.88	3.79	35.82
4	4	Wii Sports Resort	Wii	2009	Sports	.	15.75	11.01	3.28	33.00
5	5	Pokemon Red/Pokemon Blue	GB	1996	Role-Playing	Nintendo	11.27	8.89	10.22	31.37
6	6	Tetris	GB	1989	Puzzle	Nintendo	23.20	2.26	4.22	30.26
7	7	New Super Mario Bros.	DS	2006	Platform	Nintendo	11.38	9.23	6.50	30.01
8	8	Wii Play	Wii	2006	Misc	Nintendo	14.03	9.20	2.93	29.02
9	9	New Super Mario Bros. Wii	Wii	2009	Platform	Nintendo	14.59	7.06	4.70	28.62
10	10	Duck Hunt	NES	1984	Shooter	Nintendo	26.93	.63	.28	28.31
11	11	Nintendogs	DS	2005	Simulation	Nintendo	9.07	11.00	1.93	24.76
12	12	Mario Kart DS	DS	2005	Racing	Nintendo	9.81	7.57	4.13	23.42
13	13	Pokemon Gold/Pokemon Silver	GB	1999	Role-Playing	Nintendo	9.00	6.18	7.20	23.10
14	14	Wii Fit	Wii	2007	Sports	Nintendo	8.94	8.03	3.60	22.72
15	15	Wii Fit Plus	Wii	2009	Sports	Nintendo	9.09	8.59	2.53	22.00
16	16	Kinect Adventures!	X360	2010	Misc	Microsoft Game Studios	14.97	4.94	.24	21.82
17	17	Grand Theft Auto V	PS3	2013	Action	Take-Two Interactive	7.01	9.27	.97	21.40
18	18	Grand Theft Auto: San Andreas	PS2	2004	Action	Take-Two Interactive	9.43	.40	.41	20.81
19	19	Super Mario World	SNES	1990	Platform	Nintendo	12.78	3.75	3.54	20.61
20	20	Brain Age: Train Your Brain in Minute...	DS	2005	Misc	Nintendo	4.75	9.26	4.16	20.22
21	21	Pokemon Diamond/Pokemon Pearl	DS	2006	Role-Playing	Nintendo	6.42	4.52	6.04	18.36
22	22	Super Mario Land	GB	1989	Platform	Nintendo	10.83	2.71	4.18	18.14
23	23	Super Mario Bros. 3	NES	1988	Platform	Nintendo	9.54	3.44	3.84	17.28

Data: How do we organize it?

Cases: The “things” we are asking questions about. Can be people, sports teams, cities, or colonies of bacteria

Variables: The questions we ask. Height, Win %, Population, and Growth Rate are all possible variables

Data

Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Global_Sales
1	1 Wii Sports	Wii	2006	Sports	Nintendo	41.49	29.02	3.77	82.74
2	2 Super Mario Bros.	NES	1985	Platform	Nintendo	29.08	3.58	6.81	40.24
3	3 Mario Kart Wii	Wii	2008	Racing	Nintendo	15.85	12.88	3.79	35.82
4	4 Wii Sports Resort	Wii	2009	Sports	Nintendo	15.75	11.01	3.28	33.00
5	5 Pokemon Red/Pokemon Blue	GB	1996	Role-Playing	Nintendo	11.27	8.89	10.22	31.37
6	6 Tetris	GB	1989	Puzzle	Nintendo	23.20	2.26	4.22	30.26
7	7 New Super Mario Bros.	DS	2006	Platform	Nintendo	11.38	9.23	6.50	30.01
8	8 Wii Play	Wii	2006	Misc	Nintendo	14.03	9.20	2.93	29.02
9	9 New Super Mario Bros. Wii	Wii	2009	Platform	Nintendo	14.59	7.06	4.70	28.62
10	10 Duck Hunt	NES	1984	Shooter	Nintendo	26.93	.63	.28	28.31
11	11 Nintendogs	DS	2005	Simulation	Nintendo	9.07	11.00	1.93	24.76
12	12 Mario Kart DS	DS	2005	Racing	Nintendo	9.81	7.57	4.13	23.42
13	13 Pokemon Gold/Pokemon Silver	GB	1999	Role-Playing	Nintendo	9.00	6.18	7.20	23.10
14	14 Wii Fit	Wii	2007	Sports	Nintendo	8.94	8.03	3.60	22.72
15	15 Wii Fit Plus	Wii	2009	Sports	Nintendo	9.09	8.59	2.53	22.00
16	16 Kinect Adventures!	X360	2010	Misc	Microsoft Game Studios	14.97	4.94	.24	21.82
17	17 Grand Theft Auto V	PS3	2013	Action	Take-Two Interactive	7.01	9.27	.97	21.40
18	18 Grand Theft Auto: San Andreas	PS2	2004	Action	Take-Two Interactive	9.43	.40	.41	20.81
19	19 Super Mario World	SNES	1990	Platform	Nintendo	12.78	3.75	3.54	20.61
20	20 Brain Age: Train Your Brain in Minute...	DS	2005	Misc	Nintendo	4.75	9.26	4.16	20.22
21	21 Pokemon Diamond/Pokemon Pearl	DS	2006	Role-Playing	Nintendo	6.42	4.52	6.04	18.36
22	22 Super Mario Land	GB	1989	Platform	Nintendo	10.83	2.71	4.18	18.14
23	23 Super Mario Bros. 3	NES	1988	Platform	Nintendo	9.54	3.44	3.84	17.28

What is a case?

A case is every row in your data set. You ask questions about **cases**.

What is a variable?

A **variable** is every column in your data set. Variables are the questions you ask about your cases

Data

Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Global_Sales
1	Wii Sports	Wii	2006	Sports	Nintendo	41.49	29.02	3.77	82.74
2	Super Mario Bros.	NES	1985	Platform	Nintendo	29.08	3.58	6.81	40.24
3	Mario Kart Wii	Wii	2008	Racing	Nintendo	15.85	12.88	3.79	35.82
4	Wii Sports Resort	Wii	2009	Sports	.	15.75	11.01	3.28	33.00
5	Pokemon Red/Pokemon Blue	GB	1996	Role-Playing	Nintendo	11.27	8.89	10.22	31.37
6	Tetris	GB	1989	Puzzle	Nintendo	23.20	2.26	4.22	30.26
7	New Super Mario Bros.	DS	2006	Platform	Nintendo	11.38	9.23	6.50	30.01
8	Wii Play	Wii	2006	Misc	Nintendo	14.03	9.20	2.93	29.02
9	New Super Mario Bros. Wii	Wii	2009	Platform	Nintendo	14.59	7.06	4.70	28.62
10	Duck Hunt	NES	1984	Shooter	Nintendo	26.93	.63	.28	28.31
11	Nintendogs	DS	2005	Simulation	Nintendo	9.07	11.00	1.93	24.76
12	Mario Kart DS	DS	2005	Racing	Nintendo	9.81	7.57	4.13	23.42
13	Pokemon Gold/Pokemon Silver	GB	1999	Role-Playing	Nintendo	9.00	6.18	7.20	23.10
14	Wii Fit	Wii	2007	Sports	Nintendo	8.94	8.03	3.60	22.72
15	Wii Fit Plus	Wii	2009	Sports	Nintendo	9.09	8.59	2.53	22.00
16	Kinect Adventures!	X360	2010	Misc	Microsoft Game Studios	14.97	4.94	.24	21.82
17	Grand Theft Auto V	PS3	2013	Action	Take-Two Interactive	7.01	9.27	.97	21.40
18	Grand Theft Auto: San Andreas	PS2	2004	Action	Take-Two Interactive	9.43	.40	.41	20.81
19	Super Mario World	SNES	1990	Platform	Nintendo	12.78	3.75	3.54	20.61
20	Brain Age: Train Your Brain in Minute.	DS	2005	Misc	Nintendo	4.75	9.26	4.16	20.22
21	Pokemon Diamond/Pokemon Pearl	DS	2006	Role-Playing	Nintendo	6.42	4.52	6.04	18.36
22	Super Mario Land	GB	1989	Platform	Nintendo	10.83	2.71	4.18	18.14
23	Super Mario Bros. 3	NES	1988	Platform	Nintendo	9.54	3.44	3.84	17.28

Cases

= video games

Variables

**=rank, platform,
year, genre,
publisher, Sales**

**If categories are
unique to each case,
it is NOT a variable**

Variables: Types

Quantitative (numerical)

Made of numbers

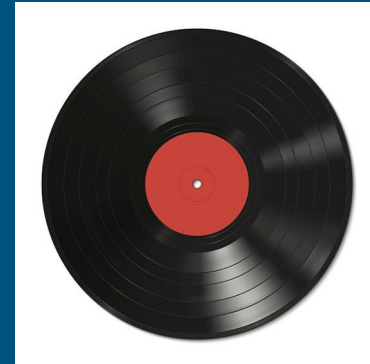
Can add, subtract, order



Categorical/Qualitative

Made of categories

Adding/subtracting makes no sense



Variables: Types

Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Global_Sales
1	Wii Sports	Wii	2006	Sports	Nintendo	41.49	29.02	3.77	82.74
2	Super Mario Bros.	NES	1985	Platform	Nintendo	29.08	3.58	6.81	40.24
3	Mario Kart Wii	Wii	2008	Racing	Nintendo	15.85	12.88	3.79	35.82
4	Wii Sports Resort	Wii	2009	Sports	Nintendo	15.75	11.01	3.28	33.00
5	Pokemon Red/Pokemon Blue	GB	1996	Role-Playing	Nintendo	11.27	8.89	10.22	31.37
6	Tetris	GB	1989	Puzzle	Nintendo	23.20	2.26	4.22	30.26
7	New Super Mario Bros.	DS	2006	Platform	Nintendo	11.38	9.23	6.50	30.01
8	Wii Play	Wii	2006	Misc	Nintendo	14.03	9.20	2.93	29.02
9	New Super Mario Bros. Wii	Wii	2009	Platform	Nintendo	14.59	7.06	4.70	28.62
10	Duck Hunt	NES	1984	Shooter	Nintendo	26.93	.63	.28	28.31
11	Nintendogs	DS	2005	Simulation	Nintendo	9.07	11.00	1.93	24.76
12	Mario Kart DS	DS	2005	Racing	Nintendo	9.81	7.57	4.13	23.42
13	Pokemon Gold/Pokemon Silver	GB	1999	Role-Playing	Nintendo	9.00	6.18	7.20	23.10
14	Wii Fit	Wii	2007	Sports	Nintendo	8.94	8.03	3.60	22.72
15	Wii Fit Plus	Wii	2009	Sports	Nintendo	9.09	8.59	2.53	22.00
16	Kinect Adventures!	X360	2010	Misc	Microsoft Game Studios	14.97	4.94	.24	21.82
17	Grand Theft Auto V	PS3	2013	Action	Take-Two Interactive	7.01	9.27	.97	21.40
18	Grand Theft Auto: San Andreas	PS2	2004	Action	Take-Two Interactive	9.43	.40	.41	20.81
19	Super Mario World	SNES	1990	Platform	Nintendo	12.78	3.75	3.54	20.61
20	Brain Age: Train Your Brain in Minutes!	DS	2005	Misc	Nintendo	4.75	9.26	4.16	20.22
21	Pokemon Diamond/Pokemon Pearl	DS	2006	Role-Playing	Nintendo	6.42	4.52	6.04	18.36
22	Super Mario Land	GB	1989	Platform	Nintendo	10.83	2.71	4.18	18.14
23	Super Mario Bros. 3	NES	1988	Platform	Nintendo	9.54	3.44	3.84	17.28

Two Major Types

1. Quantitative (Numerical)

Math!

2. Categorical

Can't math!

If categories are unique to each case, it is NOT a variable

Variables: Types

Nominal

Ordinal

Interval

Ratio



Variables: Types



Nominal

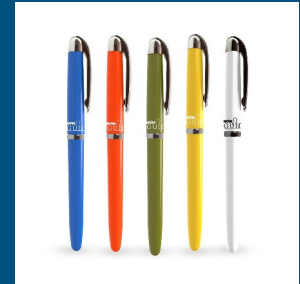
Ordinal

Interval

Ratio

- Categorical
- Order doesn't matter

Ex: Football teams, color of pens



Variables: Types



Nominal

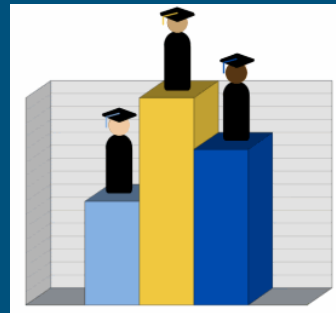
Ordinal

- Categorical
- Order DOES matter

Interval

Ex: place in a race, school rank

Ratio



Variables: Types



Nominal

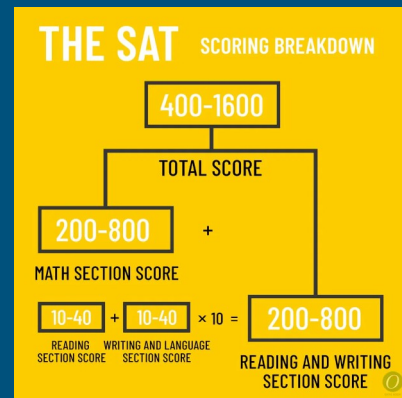
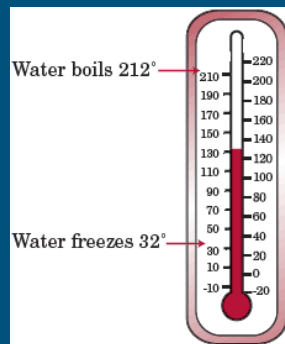
Ordinal

Interval

Ratio

- Quantitative (numerical)
- Order matters, but no set 0

Ex: Fahrenheit, SAT score



Variables: Types



Nominal

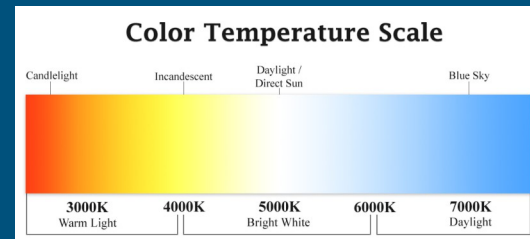
Ordinal

Interval

Ratio

- Quantitative
- 0 (zero) means 0, a total lack of stuff

Ex: Weight, temp in Kelvins



THINK ~ 1 min



Identify the type of variable for each of the following:

- The square footage of houses in St Paul
- [Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree]
- Home state, with 1 = Alabama; 50 = Wyoming
- Number of pairs of shoes someone owns

Nominal - categories, order doesn't matter

Ordinal - categories, order DOES matter

Interval - numbers, no set 0 point

Ratio - numbers, 0 means 0

PAIR ~ 2 mins



Identify the type of variable for each of the following:

- The square footage of houses in St Paul
- [Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree]
- Home state, with 1 = Alabama; 50 = Wyoming
- Number of pairs of shoes someone owns

Nominal - categories, order doesn't matter

Ordinal - categories, order DOES matter

Interval - numbers, no set 0 point

Ratio - numbers, 0 means 0

SHARE ~ 5 mins



Identify the type of variable for each of the following:

- The square footage of houses in St Paul
- [Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree]
- Home state, with 1 = Alabama; 50 = Wyoming
- Number of pairs of shoes someone owns

Nominal - categories, order doesn't matter

Ordinal - categories, order DOES matter

Interval - numbers, no set 0 point

Ratio - numbers, 0 means 0

SHARE ~ 5 mins



Identify the type of variable for each of the following:

- The square footage of houses in St Paul
- [Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree]
- Home state, with 1 = Alabama; 50 = Wyoming
- Number of pairs of shoes someone owns

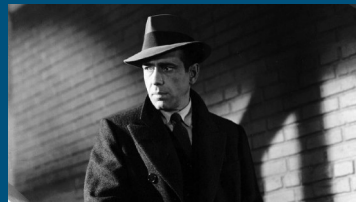
Nominal - categories, order doesn't matter

Ordinal - categories, order DOES matter

Interval - numbers, no set 0 point

Ratio - numbers, 0 means 0

SHARE ~ 5 mins



Identify the type of variable for each of the following:

- The square footage of houses in St Paul
- [Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree]
- Home state, with 1 = Alabama; 50 = Wyoming
- Number of pairs of shoes someone owns

Nominal - categories, order doesn't matter

Ordinal - categories, order DOES matter

Interval - numbers, no set 0 point

Ratio - numbers, 0 means 0

SHARE ~ 5 mins



Identify the type of variable for each of the following:

- The square footage of houses in St Paul
- [Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree]
- Home state, with 1 = Alabama; 50 = Wyoming
- Number of pairs of shoes someone owns

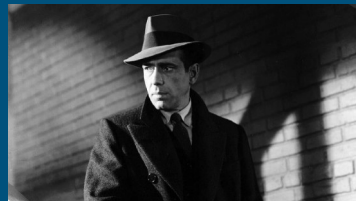
Nominal - categories, order doesn't matter

Ordinal - categories, order DOES matter

Interval - numbers, no set 0 point

Ratio - numbers, 0 means 0

SHARE ~ 5 mins



Identify the type of variable for each of the following:

- The square footage of houses in St Paul
- [Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree]
- Home state, with 1 = Alabama; 50 = Wyoming
- Number of pairs of shoes someone owns

Nominal - categories, order doesn't matter

Ordinal - categories, order DOES matter

Interval - numbers, no set 0 point

Ratio - numbers, 0 means 0

Variables: Types

For quantitative variables, we have one additional distinction!

Continuous

Vs.

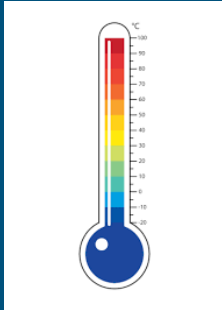
Discrete

Variables: Types; Quantitative

Continuous

All decimal places possible

Ex: Temperatures; Weight



Discrete

Only whole numbers

Ex: Goals in a hockey game



Variables: Types; Quantitative

Time it takes someone to run a 5k

Latitude of a city

Continuous

Age of patients at a clinic

Number of pairs of shoes someone owns

Discrete

Vocab reminder!

Proportion vs. Percent- when do we use each?

Out of 96 students in this course, 35 are sophomores.

What proportion of students are sophomores?

What percent of students are sophomores?

Percent Vs. Proportion

Proportion = how much out of 1

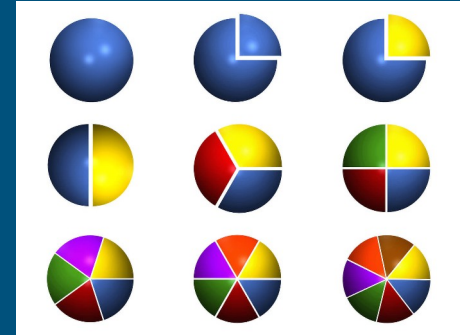
Percent = how many out of 100

Proportion of 0.52 = 52%

To convert proportion to percent: multiply by 100

To convert percent to proportion: divide by 100

What proportion is a percent of 0.034%?



Vocab reminder!

Proportion vs. Percent- how to calculate?

Out of 96 students in this course, 35 are sophomores.

What proportion of students are sophomores?

$$35/96 = 0.365$$

What percent of students are sophomores?

$$0.365 * 100 = 36.5\%$$

Variables: Relationships

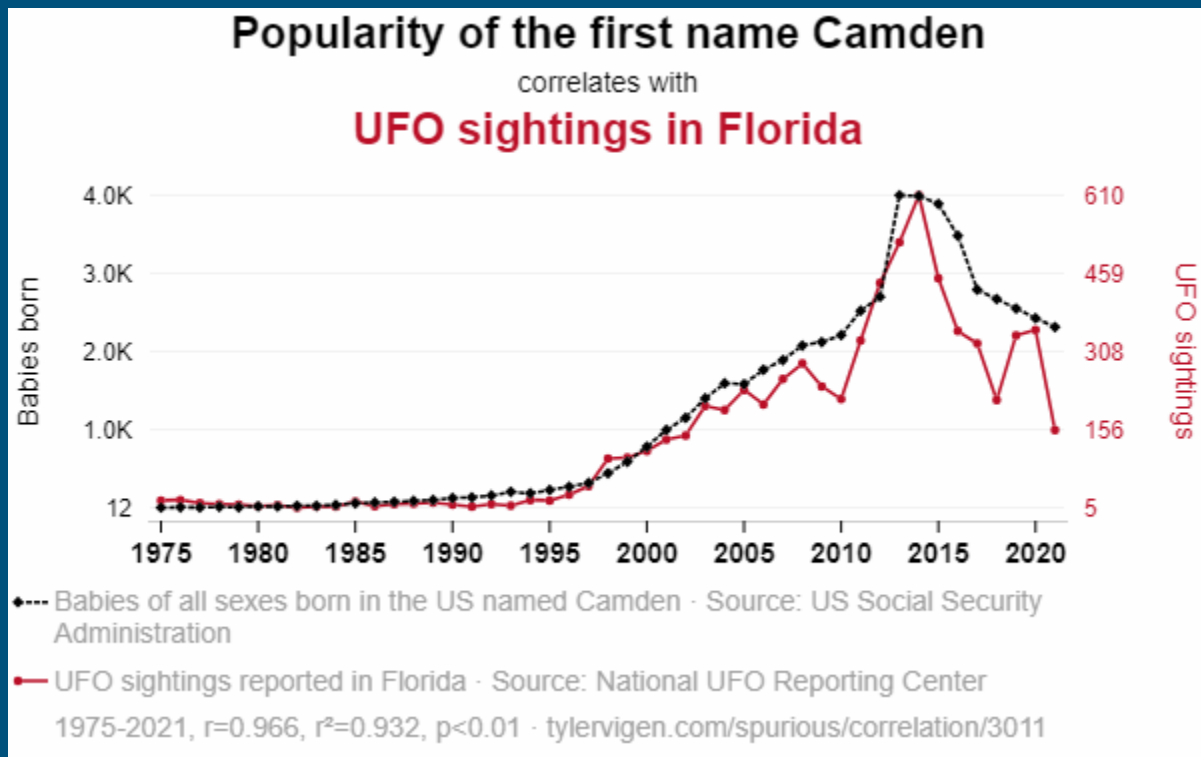
Either

Independent

OR

Associated

Associated Variables: Correlation \neq Causation



Associated Variables: Explanatory vs. Response variables

Does **explanatory variable** explain differences in **response variable**?

Ex:

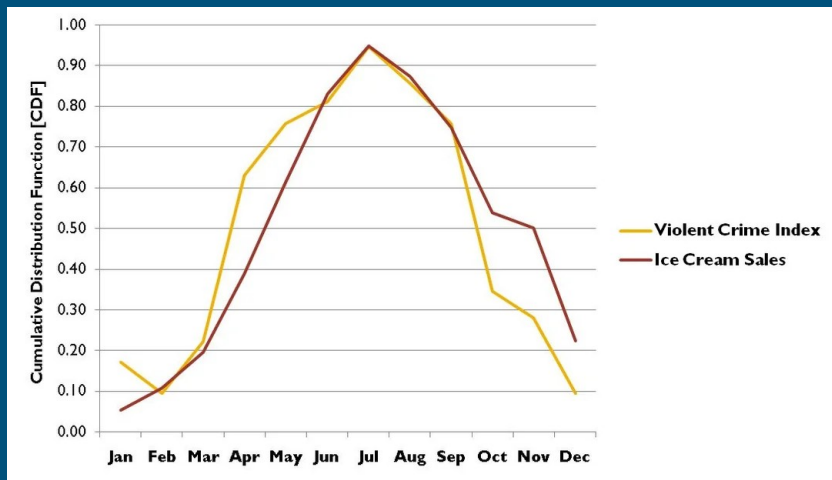
Are **crime rates** affected by **temperature**?

Does **sugar consumption** cause **hyperactivity in children**?

If you change the **explanatory variable**, you expect the **response variable** to also change

Associated Variables: Confounding variables

Confounding variables: when two variables appear to be directly related, but each are actually related to a third (hidden) variable



Should we ban **ice cream** to reduce **violent crime**?

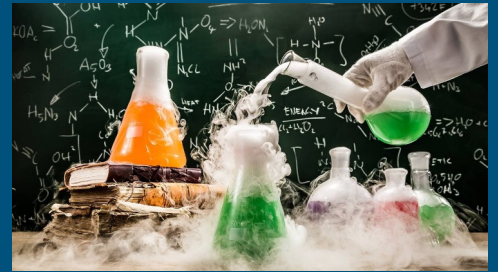
What might our confounding variable here be?

To Do

- Explain class structure; explore canvas
- Stats vocab lesson!
- Experimental design



Experimental Design



First!

An observational study is NOT an experiment

Not experiment language:

“survey”

“case study”

“observed”

Experiment language:

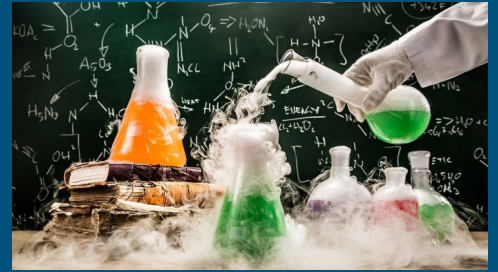
“control group”

“tested”

“significant”

You NEED experiments in order to conclude cause and effect

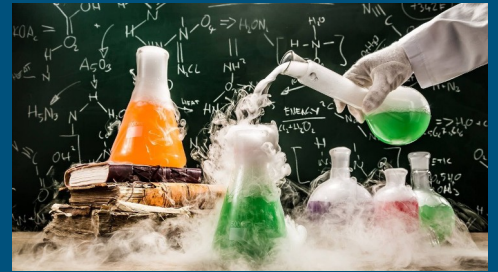
Experimental Design



In order to have an experiment, you need a few things:

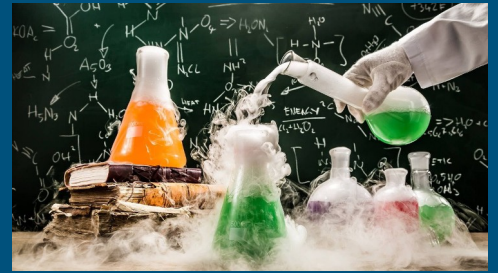
1. At least 2 groups
2. Experimenter control
3. Random selection/assignment

Experimental Design: Blinds



- When assigning people to groups, you generally don't want them to know which group they're in - this is called a Blind
- If the participants AND the experimenters do not know which group in which, this is called a Double Blind experiment
- Double blind experiments help reduce experimenter bias

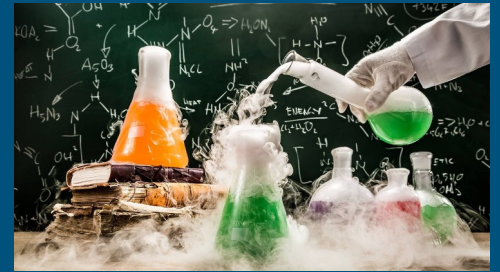
Think ~ 1 min



An experimenter wants to study the relationship between lung cancer and smoking. She splits a group of people aged 50-75 into two groups: smokers and non-smokers. She finds that the smokers are 50% more likely to develop lung cancer within 10 years.

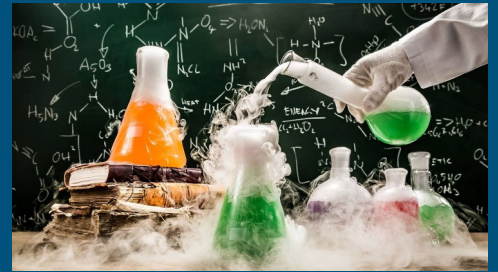
- a) What are the explanatory and response variables here?
- b) Is this an experiment or an observational study?
- c) Can we determine that smoking causes lung cancer? Why or why not?

Pair ~ 2 min



An experimenter wants to study the relationship between lung cancer and smoking. She splits a group of people aged 50-75 into two groups: smokers and non-smokers. She finds that the smokers are 50% more likely to develop lung cancer within 10 years.

- a) What are the explanatory and response variables here?
- b) Is this an experiment or an observational study?
- c) Can we determine that smoking causes lung cancer? Why or why not?



An experimenter wants to study the relationship between lung cancer and smoking. She splits a group of people aged 50-75 into two groups: smokers and non-smokers. She finds that the smokers are 50% more likely to develop lung cancer within 10 years.

- What are the explanatory and response variables here?
- Is this an experiment or an observational study?
- Can we determine that smoking causes lung cancer? Why or why not?

Fin

