Simpson's Paradox

Donna Dietz

American University

dietz@american.edu

STAT 202 - Spring 2020

First, something that isn't weird...

Before we get into this "Simpson's Paradox" idea, let's talk about something that works the way you would expect it to work.

Let's pretend that we can survey all the students at AU. More than half of the guys say they prefer chocolate to strawberry. More than half of the women agree. Even more than half of the non-binary students agree. What can we conclude?

If we can subdivide a group into non-overlapping subgroups, and we find that over half (or over any ratio) of this group has this trait, that same conclusion will apply to the whole group. So, we can safely conclude that in the entire school, more than half of all students prefer chocolate over strawberry.

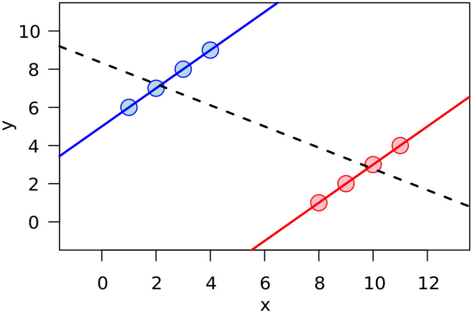
This works as you would expect.

So, what is Simpson's Paradox?

A collection of so-called paradoxes (which some argue aren't true paradoxes)

- Simpson's Paradox
- Lord's Paradox
- Suppression Effects

So, what is Simpson's Paradox?



For the overall group, there is a negative correlation.
However, for each subgroup, there is a positive correlation!

Also

Similar effects happen regardless of whether the data are categorical, ordinal, or numerical.

Example

If I were to plot study time against grades, which I've done, I sometimes find a class where the students who study fewer hours actually have higher grades. But this doesn't mean your grades go up if you don't study. There are students who would bring a B+ to an A by studying but perhaps they don't care. And there are students who have to work very hard just to get a B, and they will do this! So, the data can hide obvious truths if we don't look at it properly.

How long has this been discussed?

You'd think that by now, we'd have it all figured out and this effect would be totally understood. Nope. We seem to never be able to stop talking about it!

- 1899 Karl Pearson
- 1903 Udny Yule
- 1945 Maruice George Kendall
- **1951 Edward H. Simpson**
- 1967 Frederic M Lord
- 2008 Yu-Kang Tu
- 2016 Judea Pearl
- 2019 Carol Nickerson

But this guy gets all the credit...

Edward H. Simpson

Here is a problem

We (educators) brought Simpson's Paradox into the classroom as a prime example of why we needed better quantitative literacy (or whatever buzzword) in the classrooms. We thought people needed to rely more on data instead of relying on intuition to form beliefs. That seems logical right?

But!

But, what we ended up with was an artificial treatment of the subject with most of the examples which are given to the students - as only explaining this phenomenon of **lurking variables**. Lurking variables are quite important, to be sure! However, they only tell part of the story!

Lurking Variables

Lurking Variables

A lurking variable is a part of your model or explanation which is crucial yet missing. It's something that helps to explain what's going on, yet you've somehow overlooked it. Your understanding of the situation is limited, unbeknownst to you.

The entire point

With the push to add Simpson's Paradox to standard classroom fare, most treatments of it overlook its main point!

The data DO NOT TELL you whether or not you need to look at the aggregate data or the grouped data!!!

That sounds like *intuition* to me!

Two examples

Here, I will review the two examples given by Simpson himself in his 1951 paper! The two examples use the selfsame data. Notice that the results are opposite, and there is nothing in the data which could possibly lead to knowing this!

Also...

Also please note: The numbers are all very small. They would not be statistically significant. However, everyone reading Simpson's paper would immediately understand that all the values could simply be multiplied by something like 1000 and suddenly that would not be an issue. So, even though these values aren't significant, that's not important for this discussion. Numbers stay small to keep them easy to calculate and do not distract from the point of the examples.

Baby plays with cards

9. An investigator wished to examine whether in a pack of cards the proportion of court cards (King, Queen, Knave) was associated with colour. It happened that the pack which he examined was one with which Baby had been playing, and some of the cards were dirty. He included the classification "dirty" in his scheme in case it was relevant, and obtained the following probabilities:

TABLE 2

	Dirty		*Clean*	
	Court	*Plain*	*Court*	*Plain*
Red . . .	4/52	8/52	2/52	12/52
Black . . .	3/52	5/52	3/52	15/52

It will be observed that Baby preferred red cards to black and court cards to plain, but showed

Baby plays with cards

no second order interaction on Bartlett's definition. The investigator deduced a positive association between redness and plainness both among the dirty cards and among the clean, yet it is the combined table

TABLE 3

	Court	Plain
Red . . .	6/52	20/52
Black . . .	6/52	20/52

which provides what we would call the sensible answer, namely, that there is no such association.

StatCrunch

Contingency table results:
Rows: (Dirty)
Columns: None

Cell format
Count (Expected count)

	Court	Plain	Total
Red	4 (4.2)	8 (7.8)	12
Black	3 (2.8)	5 (5.2)	8
Total	7	13	20

Contingency table results:
Rows: (Clean)
Columns: None

Cell format
Count (Expected count)

	Court	Plain	Total
Red	2 (2.19)	12 (11.81)	14
Black	3 (2.81)	15 (15.19)	18
Total	5	27	32

Same numbers, different conclusion!

10. Suppose we now change the names of the classes in Table 2 thus:

TABLE 4

	Male		*Female*	
	Untreated	*Treated*	*Untreated*	*Treated*
Alive . . .	4/52	8/52	2/52	12/52
Dead . . .	3/52	5/52	3/52	15/52

The probabilities are exactly the same as in Table 2, and there is again the same degree of positive association in each of the 2×2 tables. This time we say that there is a positive association between treatment and survival both among males and among females; but if we combine the tables we again find that there is no association between treatment and survival in the combined population. What is the "sensible" interpretation here? The treatment can hardly be rejected as valueless to the race when it is beneficial when applied to males and to females.

Wait.. how did that happen?

Our two-way table was really one main two-way table which was split up into two sub-categories. You really have a three dimensional data table. In this case we have 8 possible combinations.

When we have three dimensional data, we can fall into this trap of Comparing only one part of the data without seeing it in the whole context properly.

Another example

Options

Contingency table results:
Rows: AllStudents
Columns: None

	NoDetention	Detention	Total
Slytherin	750	250	1000
Griffindor	500	220	720
Total	1250	470	1720

Chi-Square test:

Statistic	DF	Value	P-value
Chi-square	1	6.5059102	0.0108

In this two-way table, we are shown that about 25% of Slytherin students end up getting detention, while about 31% of Griffindor students do. But Griffindor students claim that they tend to be the more well-behaved students!

Split that up!

Options

Contingency table results:
Rows: YoungerStudents
Columns: None

	NoDetention	Detention	Total
Slytherin	150	150	300
Griffindor	400	215	615
Total	550	365	915

Chi-Square test:

Statistic	DF	Value	P-value
Chi-square	1	19.023707	<0.0001

If we split it up by the age of the student, we see that for younger students, nearly 50% of Slytherin go to detention while only about a third of Griffindor do.

And older students as well!

Options

Contingency table results:
Rows: ElderStudents
Columns: None

	NoDetention	Detention	Total
Slytherin	600	100	700
Griffindor	100	5	105
Total	700	105	805

Chi-Square test:

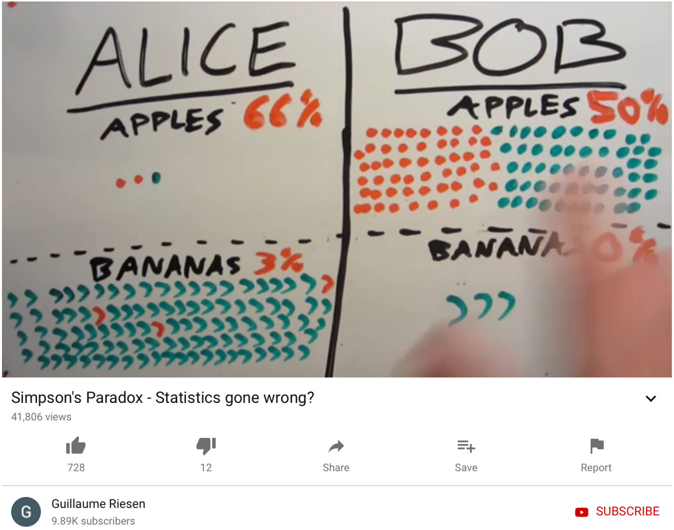
Statistic	DF	Value	P-value
Chi-square	1	7.3015873	0.0069

For older students, about 5% of Griffindor students end up in detention compared to about 14% of Slytherin!

For both age groups, Griffindor has less detention! But the issue here is that most of the Griffindor students are younger and most of the Slytherin students are older. So, it's not fair to compare these kids without considering that the ages are nowhere near proportional.

Another example

The internet abounds ...



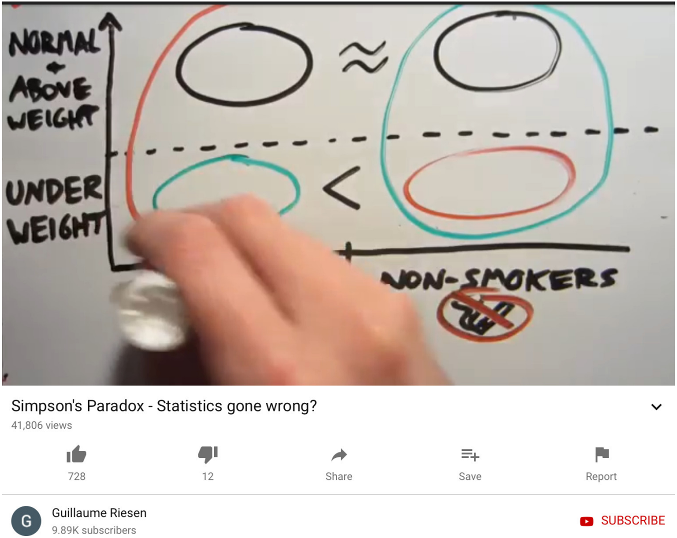
Another example

From Guillaume Riesen:

	Alice	Bob
Apples	2 bad 1 good	50 bad 50 good
Bananas	3 bad 97 good	0 bad 3 good

If you want an apple, you should buy one from Bob.
If you want a banana, you should buy one from Bob.
If you want to buy a random piece of fruit and don't care what it is, buy from Alice!

A classic!


NORMAL
+
ABOVE
WEIGHT
UNDER
WEIGHT
NON-SMOKERS
Simpson's Paradox - Statistics gone wrong?
41,806 views
728
12
Share
Save
Report
G
Guillaume Riesen
9.89K subscribers
SUBSCRIBE

Common Example

Free throws made by player by year

	Williams	Durant
2015	163/187	146/171
2016	38/42	447/498

Williams had a better average in 2015. 87.2% versus 85.4%
Williams had a better average in 2016. 90.5% versus 89.8%
Durant had a better overall average! 88.6% versus 87.8%

Even weirder example

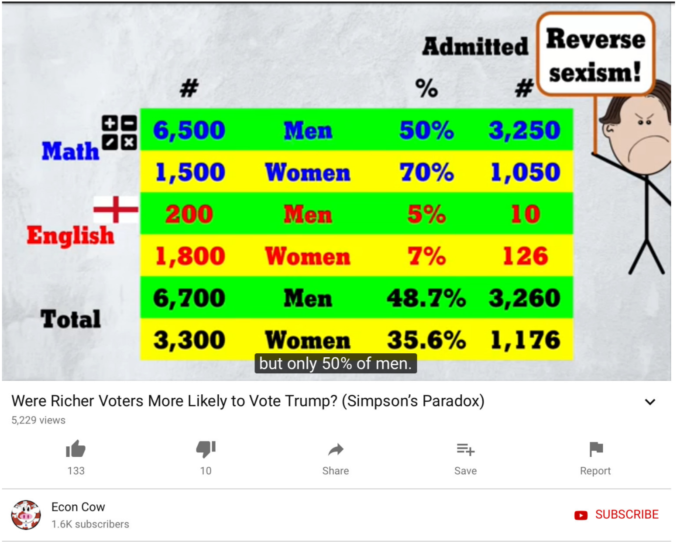
The flip can happen at multiple levels! (Econ Cow on YouTube)
This must be fabricated data, but you can see how it could happen...

Hypothetical Admissions Data

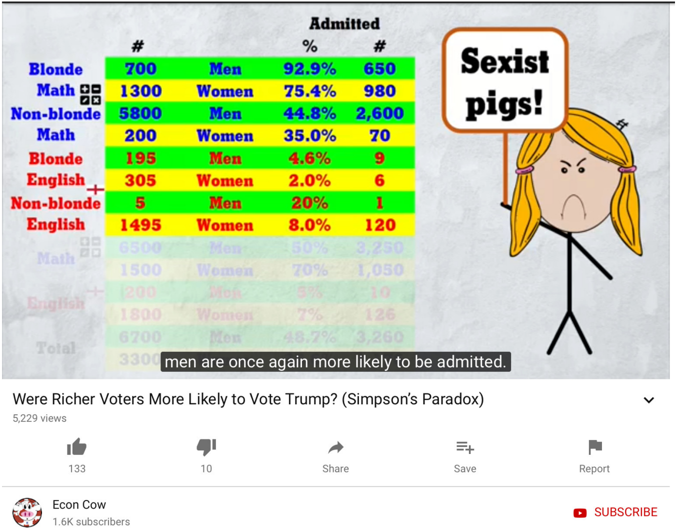
	Men	Women
Math	3250/6500	1050/1500
English	10/200	126/1800

In both Math (70% versus 50%) and English (7% versus 5%), women were admitted more. However, overall (35.6% versus 48.7%), women were admitted less!

Screenshots!



Screenshots!



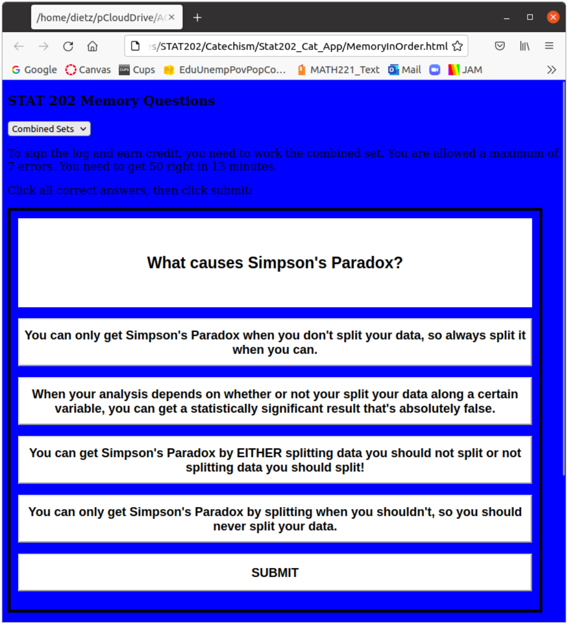
Weirder!

(Math)	Men	Women
Blonde	650/700	980/1300
Other	2600/5800	70/200

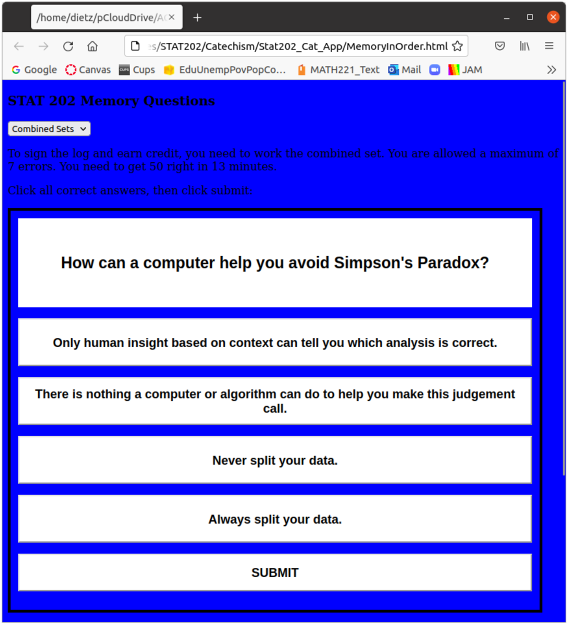
(English)	Men	Women
Blonde	9/195	6/305
Other	1/5	120/1495

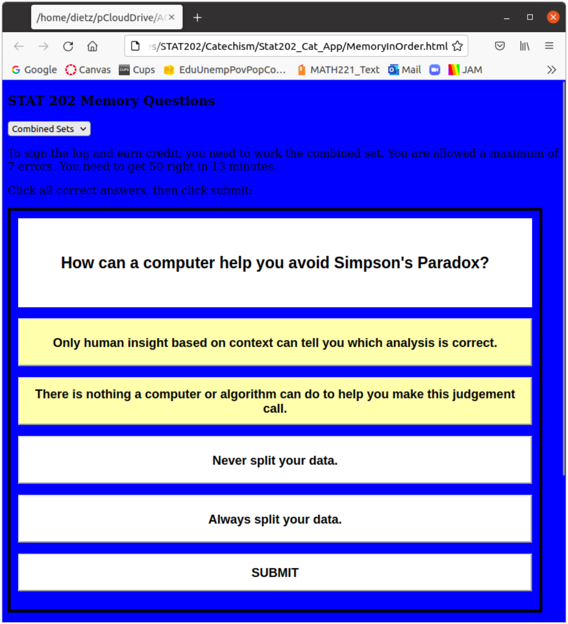
Worksheet time!

MEMORY QUESTIONs


/home/dietz/pCloudDrive/A
s/STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html
Google
Canvas
Cups
EduUnempPovPopCo...
MATH221_Text
Mail
JAM
STAT 202 Memory Questions
Combined Sets
To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.
Click all correct answers, then click submit:
What causes Simpson's Paradox?
You can only get Simpson's Paradox when you don't split your data, so always split it when you can.
When your analysis depends on whether or not your split your data along a certain variable, you can get a statistically significant result that's absolutely false.
You can get Simpson's Paradox by EITHER splitting data you should not split or not splitting data you should split!
You can only get Simpson's Paradox by splitting when you shouldn't, so you should never split your data.
SUBMIT


/home/dietz/pCloudDrive/A
s/STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html
Google
Canvas
Cups
EduUnempPovPopCo...
MATH221_Text
Mail
JAM
STAT 202 Memory Questions
Combined Sets
To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.
Click all correct answers, then click submit:
What causes Simpson's Paradox?
You can only get Simpson's Paradox when you don't split your data, so always split it when you can.
When your analysis depends on whether or not your split your data along a certain variable, you can get a statistically significant result that's absolutely false.
You can get Simpson's Paradox by EITHER splitting data you should not split or not splitting data you should split!
You can only get Simpson's Paradox by splitting when you shouldn't, so you should never split your data.
SUBMIT


/home/dietz/pCloudDrive/A
s/STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html
Google
Canvas
Cups
EduUnempPovPopCo...
MATH221_Text
Mail
JAM
STAT 202 Memory Questions
Combined Sets
To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.
Click all correct answers, then click submit:
How can a computer help you avoid Simpson's Paradox?
Only human insight based on context can tell you which analysis is correct.
There is nothing a computer or algorithm can do to help you make this judgement call.
Never split your data.
Always split your data.
SUBMIT


/home/dietz/pCloudDrive/A
s/STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html
Google
Canvas
Cups
EduUnempPovPopCo...
MATH221_Text
Mail
JAM
STAT 202 Memory Questions
Combined Sets
To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.
Click all correct answers, then click submit:
How can a computer help you avoid Simpson's Paradox?
Only human insight based on context can tell you which analysis is correct.
There is nothing a computer or algorithm can do to help you make this judgement call.
Never split your data.
Always split your data.
SUBMIT