Data acumen and data numeracy: helping students extract meaning from data

Nicholas J. Horton

Department of Mathematics and Statistics
Amherst College, Amherst, MA, USA

National Numeracy Network Conference, October 12, 2019

nhorton@amherst.edu
http://nhorton.people.amherst.edu/nnn2019-10.pdf

Thanks and acknowledgments

- Project MOSAIC: Danny Kaplan (Macalester College), Randy Pruim (Calvin College), Ben Baumer (Smith College), and Johanna Hardin (Pomona College)
- GAISE College Report (2016) revision committee
- Data Science for Undergraduates (NASEM 2018) members
- Uri Treisman, Roxy Peck, and others at the UT Austin Dana Center
- Hill, Pearl, Scheines, Robins, Hernan, Rubin, Vanderweele, Winship, Morgan, Tchetgen Tchetgen, Cobb, and many others
- Sarah Anoke, Sonia Hernandez-Diaz, Miguel Hernan, Murray Mittelman, Brendan Seto, and Sonja Swanson for many useful suggestions, curricular models, and examples

Overview and motivation (Cobb, 1982)

Let students do what statisticians do: analyze non-trivial datasets by considering a variety of models, **using their imagination and developing their judgment** *in the process.*

I believe it is the use of imagination and judgment that makes our subject appealing. **We owe it to our students not to keep that a secret**.

ASA Undergrad Guidelines, 2014

If statistics is the science of learning from data, then our students need to be able to **"think with data"** *(as Diane Lambert of Google has so elegantly described).*
- Horton and Hardin (TAS, 2015)

Source: causal inference FB

- What do we know about smoking and lung cancer?

Source: causal inference FB

- While clinical trials are wonderful, we live in a world of 'found' data.
- "It is not that i believe an experiment is the only proper setting for discussing causality, but I do feel that it is the simplest such setting" - Holland (1986)

Do we teach multivariate thinking?

As we begin the 21st century, the introductory statistics course appears healthy, with its emphasis on real examples, data production, and graphics for exploration and assumption-checking. Without doubt this emphasis marks a major improvement over introductory courses of the 1960s, an improvement made possible by the vaunted "computer revolution." Nevertheless, I argue that despite broad acceptance and rapid growth in enrollments, the consensus curriculum is still an **unwitting prisoner of history** *(Cobb, TISE 2007).*

Teaching multivariate thinking and confounding

1. landscape of statistics in high school

Changing landscape of K-12 statistics education

Roxy Peck (JSM 2011) noted:

- statistics have been a recommended part of math curriculum for a long time
- recent developments: considerable more emphasis on statistics
- not just AP statistics: expectation for all students

Changing landscape of K-12 statistics education

1. Summarize, represent, and interpret data on a single count or measurement variable
2. Summarize, represent, and interpret data on two categorical and quantitative variables
3. Interpret linear models
4. Make inferences and justify conclusions from sample surveys, experiments, and observational studies

Inferential statistics

- S-IC.3 Recognize the purposes of and differences among sample surveys, experiments, and observational studies; explain how randomization relates to each.
- S-IC.4 Use data from a sample survey to estimate a population mean or proportion; develop a margin of error through the use of simulation models for random sampling.
- S-IC.5 Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant.
- S-IC.6 Evaluate reports based on data.

Changing landscape of K-12 statistics education

- Eight years later, where do things stand?
- Eureka Math Curriculum
- LOCUS (Levels of Conceptual Understanding in Statistics https://locus.statisticseducation.org)
- SAT Math (https://collegereadiness.collegeboard.org/sample-questions/math) (https:greatminds.org)

Eureka Math (Great Minds)

NATIONAL IMPACT

Eureka Math is the most widely used math curriculum in the United States, according to a **study** released by the RAND Corporation. It is also the only curriculum found by **EdReports.org** to align fully with the Common Core State Standards for all grades, K–8. Additionally, over a dozen lessons from *Eureka Math* were rated to be EQuIP exemplars by **Achieve**.

Eureka Math (Great Minds)

- "A nationwide survey by the RAND Corporation found that 57% of elementary school and 47% of secondary teachers use Eureka Math or the version of this curriculum developed for EngageNY.org." (greatminds.org)
- 5 weeks on statistics in grade 9 (descriptive statistics and informal inference)
- 8 weeks on statistics in grade 11 (formal inference using permutation test)

Random selection and random assignment have very different meanings

Random selection: refers to randomly selecting a sample from a population. Random selection allows generalization to a population and is used in well-designed observational studies.

Random assignment: refers to randomly assignment subjects in an experiment to a treatment. Random assignment allows for cause-and-effect conclusions and is used in well-designed experiments

For a given sample, you can find the sample mean:

1. There is variability in the sample mean
2. A graph of the distribution of sample means from many random samples is a simulated sampling distribution
3. Sample means from random samples tend to cluster around the value of the population mean.
4. The variability in the sample mean decreases as the sample size increases.
5. Most sample means are within two standard deviations of the mean of the simulated sampling distributions.

When a single set of values is randomly divided into two groups

1. the two group means will tend to differ just by chance
2. The distribution of random groups' means will be centered at the single set's mean
3. the range of the distribution of the random groups' means will be smaller than the range of the original data
4. the shape of the distribution of the random groups' means will be symmetric

The scatterplot below shows the relationship between height and arm span for a group of students. The least squares line (labeled LS line) and two other lines have been added to the scatterplot.

- A: Compared to the other lines, Line 1 has the smallest sum of squared residuals.
- B: The sum of squared residuals for Line 1 is greater than the sum of squared residuals for Line 2.
- C: Compared to the other lines, the least squares line has the smallest sum of squared residuals.
- D: The sum of squared residuals for the least squares line is greater than the sum of squared residuals for Line 2.

The city of Gainesville hosted two races last year on New Year's Day. Individual runners chose to run either a 5K (3.1 miles) or a half-marathon (13.1 miles). One hundred thirty four people ran in the 5K, and 224 people ran the half-marathon. The mile time, which is the average amount of time it takes a runner to run a mile, was calculated for each runner by dividing the time it took the runner to finish the race by the length of the race. The histograms below show the distributions of mile times (in minutes per mile) for the runners in the two races.

SAT Math test (example question)

A research assistant randomly selected 75 undergraduate students from the list of all students enrolled in the psychology-degree program at a large university. She asked each of the 75 students, "How many minutes per day do you typically spend reading?" The mean reading time in the sample was 89 minutes, and the margin of error for this estimate was 4.28 minutes. Another research assistant intends to replicate the survey and will attempt to get a smaller margin of error.

Which of the following samples will most likely result in a smaller margin of error for the estimated mean time students in the psychology-degree program read per day?

SAT Math test

A researcher wanted to know if there is an association between exercise and sleep for the population of 16-year-olds in the US. She obtained survey responses from a random sample of 2000 US 16-year-olds and found convincing evidence of a positive association between exercise and sleep. Which of the following conclusions is well supported by the data?

- A: There is a positive association between exercise and sleep for 16-year-olds in the US
- B: There is a positive association between exercise and sleep for 16-year-olds in the world.
- C: Using exercise and sleep as defined by the study, an increase in sleep is caused by an increase of exercise for 16-year-olds in the US.
- D: Using exercise and sleep as defined by the study, an increase in sleep is caused by an increase of exercise.

Changing landscape of K-12 statistics education

Executive Summary: Roxy Peck was right!

- variability, sampling distributions, linear regression, and hypothesis tests now part of the curriculum
- these topics are being assessed, therefore they are being taught
- intro stats or QL/QR course will **not** be the first exposure for many topics
- (this is helpful, since GAISE K-12 report talks about the need for repeated exposure)
- some material may be reviewed rather than taught from scratch
- allows us to explore ways to introduce multivariate thinking and develop "data acumen" in our students (more later)

Teaching multivariate thinking and confounding

1. landscape of statistics in high school
2. what are we teaching?

Course descriptions

MA 113 (Boston University) Basic concepts of estimation and tests of hypotheses, ideas from probability; one-, two-, and multiple-sample problems. Applications in social sciences. Primarily for students in the social sciences who require a one-semester introduction to statistics.

Course descriptions

MATH-1342 Elementary Statistics (Austin Community College) A first course in statistics for students in business; nursing; allied health; or the social, physical, or behavioral sciences; or for any student requiring knowledge of the fundamental procedures for data organization and analysis. Topics include frequency distributions, graphing, measures of location and variation, the binomial and normal distributions, z-scores, t-test, chi-square test, F-test, hypothesis testing, analysis of variance, regression, and correlation.

AP Stats: consensus curriculum from 1996

- much of the material now appears in the curriculum for many students
- still focused almost exclusively on univariate and bivariate relationships
- how to make sense of observational data? What do students conclude from the class when asked to interpret found data?

AP Stats: consensus curriculum from 1996

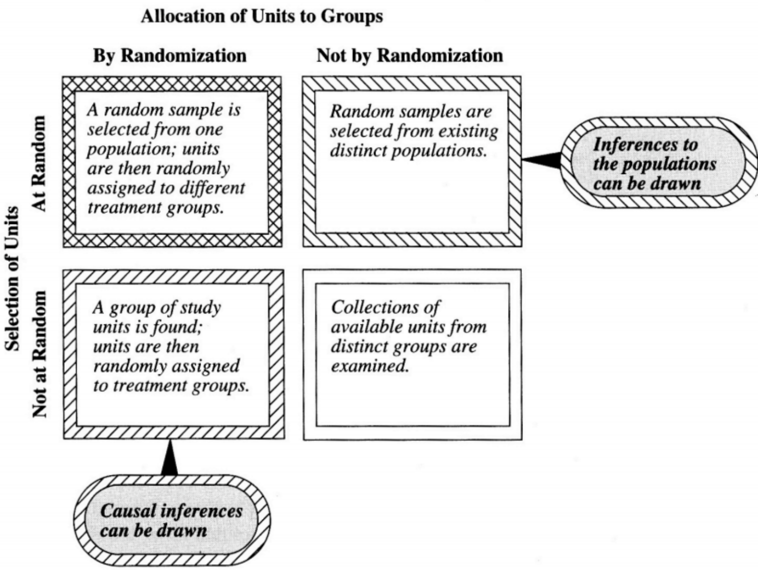
- much of the material now appears in the curriculum for many students
- still focused almost exclusively on univariate and bivariate relationships
- how to make sense of observational data? What do students conclude from the class when asked to interpret found data?
- **My worry: students leave the intro stats course feeling paralyzed!**

AP Stats: aside on technology

> *Currently, the graphing calculator is the only computational aid that is available to students for use as a tool for data analysis on the AP Exam (AP Stats course description).*

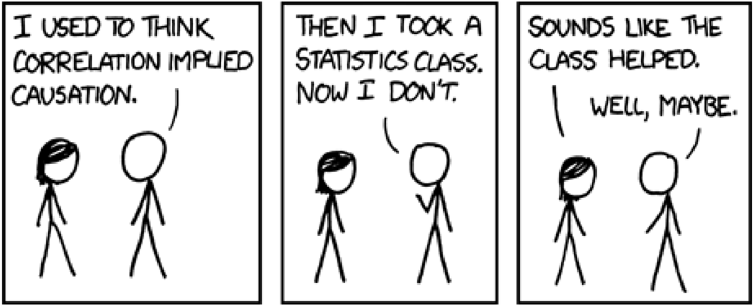
Unfortunately, statistics in the modern world does not use calculators. Can't develop much "data acumen" using them.

Other factors may be responsible for observed associations



(Ramsey and Schafer, **Statistical Sleuth**)

Do we teach in a way that encourages paralysis? thanks to xkcd



Example from SDM4 (De Veaux, Velleman, and Bock) p. 575

Exercise 20.41: It's widely believed that regular mammogram screening may detect breast cancer early, resulting in fewer deaths from that disease. One study that investigated this issue over a period of 18 years was published during the 1970's. Among 30,565 who had never had mammograms, 196 died of breast cancer (0.64%) while only 153 of 30,131 who had undergone screening died of breast cancer (0.50%).

Do these results suggest that mammograms may be an effective screening tool to reduce breast cancer deaths?

Solution to Exercise 20.41 SDM4 (De Veaux, Velleman, and Bock) p. 575

$H_0 : p_1 - p_2 = 0$ vs. $H_A : p_1 - p_2 > 0$ (one-sided test? That's a different sermon.)

Solution to Exercise 20.41 SDM4 (De Veaux, Velleman, and Bock) p. 575

$H_0 : p_1 - p_2 = 0$ vs. $H_A : p_1 - p_2 > 0$ (one-sided test? That's a different sermon.) where p_1 is the proportion of women who never had mammograms who died of breast cancer and p_2 is the proportion of women who had undergone screening who died of breast cancer (z=2.17, p=0.015).

With a p-value this low, we reject H_0. The data suggest that mammograms may reduce breast cancer deaths.

Solution to Exercise 20.41 SDM4 (De Veaux, Velleman, and Bock) p. 575

$H_0 : p_1 - p_2 = 0$ vs. $H_A : p_1 - p_2 > 0$ (one-sided test? That's a different sermon.) where p_1 is the proportion of women who never had mammograms who died of breast cancer and p_2 is the proportion of women who had undergone screening who died of breast cancer (z=2.17, p=0.015).

With a p-value this low, we reject H_0. The data suggest that mammograms may reduce breast cancer deaths.

(But what about possible confounders?)

Open Intro Statistics

6.34 Prenatal vitamins and Autism. Researchers studying the link between prenatal vitamin use and autism surveyed the mothers of a random sample of children aged 24 - 60 months with autism and conducted another separate random sample for children with typical development. The table below shows the number of mothers in each group who did and did not use prenatal vitamins during the three months before pregnancy (periconceptional period).[57]

		Autism		
		Autism	Typical development	Total
Periconceptional	No vitamin	111	70	181
prenatal vitamin	Vitamin	143	159	302
	Total	254	229	483

(a) State appropriate hypotheses to test for independence of use of prenatal vitamins during the three months before pregnancy and autism.

(b) Complete the hypothesis test and state an appropriate conclusion. (Reminder: Verify any necessary conditions for the test.)

(c) A New York Times article reporting on this study was titled "Prenatal Vitamins May Ward Off Autism". Do you find the title of this article to be appropriate? Explain your answer. Additionally, propose an alternative title.[58]

1. Since the p-value $< \alpha$, we reject H_0. There is strong evidence of a difference in the rates of autism of children of mothers who did and did not use prenatal vitamins during the first three months before pregnancy.
2. The p-value is small and we reject H0. The data provide convincing evidence to suggest that autism and vitamin use in women are associated.

1. Since the p-value $< \alpha$, we reject H_0. There is strong evidence of a difference in the rates of autism of children of mothers who did and did not use prenatal vitamins during the first three months before pregnancy.
2. The p-value is small and we reject H0. The data provide convincing evidence to suggest that autism and vitamin use in women are associated.
3. Yes, this is an observational study. Based on this study we can't deduce that taking vitamins leads to less autism. **There may be other factors, lurking variables, that may cause autism and be associated with vitamin use.**

Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report 2016

revised GAISE College Report (2016)

1. Teach statistical thinking.
2. Focus on conceptual understanding.
3. Integrate real data with a context and purpose.
4. Foster active learning.
5. Use technology to explore concepts and analyze data.
6. Use assessments to improve and evaluate student learning.

revised GAISE College Report (2016)

1. Teach statistical thinking.
   - Teach statistics as an investigative process of problem-solving and decision-making.
   - Give students experience with **multivariable thinking**.
2. Focus on conceptual understanding.
3. Integrate real data with a context and purpose.
4. Foster active learning.
5. Use technology to explore concepts and analyze data.
6. Use assessments to improve and evaluate student learning.

Data Science for Undergraduates (NASEM 2018)

> *Recommendation 2.3 To prepare their graduates for this new data-driven era, academic institutions should encourage the development of a basic understanding of data science in all undergraduates.*

`https://nas.edu/envisioningds`

Data Science for Undergraduates (NASEM 2018)

Defined "data acumen" as a key capacity for all undergraduates

> *A critical task in the education of future data scientists is to instill data acumen. This requires exposure to key concepts in data science, real-world data and problems that can reinforce the limitations of tools, and ethical considerations that permeate many applications.*

https://nas.edu/envisioningds

Data Science for Undergraduates (NASEM 2018)

Components of data acumen

- Mathematical foundations
- Computational foundations
- Statistical foundations
- Data management and curation
- Data description and visualization
- Data modeling and assessment
- Workflow and reproducibility
- Communication and teamwork
- Domain-specific considerations
- Ethical problem solving

Data Science for Undergraduates (NASEM 2018)

Components of data acumen

- Mathematical foundations
- Computational foundations
- **Statistical foundations**
- Data management and curation
- Data description and visualization
- Data modeling and assessment
- Workflow and reproducibility
- Communication and teamwork
- Domain-specific considerations
- Ethical problem solving

Data Science for Undergraduates (NASEM 2018)

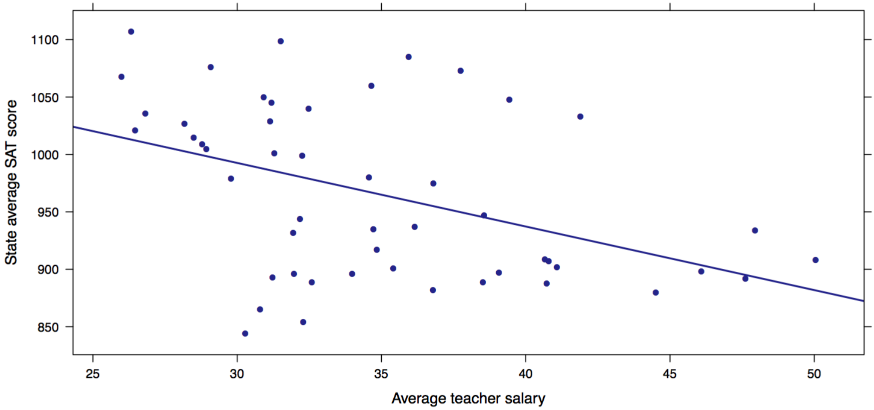
Important statistical foundations include:

- Variability, uncertainty, sampling error, and inference
- **Multivariate thinking**
- Nonsampling error, **design**, experiments (e.g., A/B testing), **biases, confounding, and causal inference**
- Exploratory data analysis
- **Statistical modeling** and model assessment; and
- Simulations and experiments

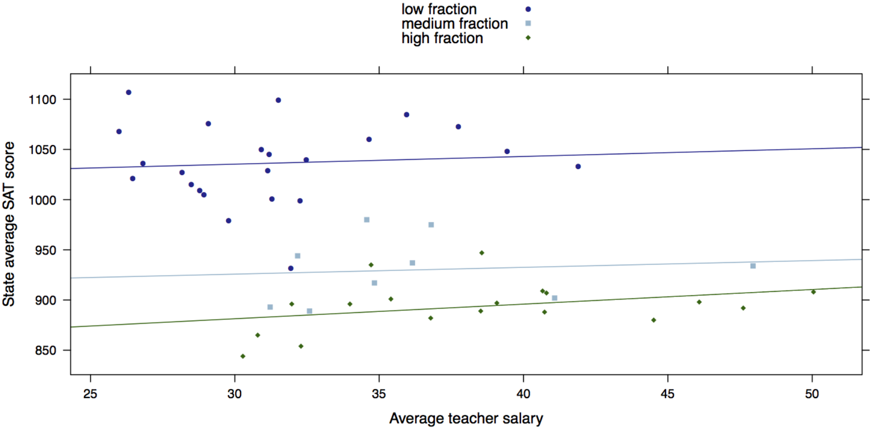
Teaching multivariate thinking and confounding

1. landscape of statistics in high school
2. what are we teaching?
3. multivariate thinking and confounding 101 and 201

SAT scores and teacher salaries (state data from 2010)

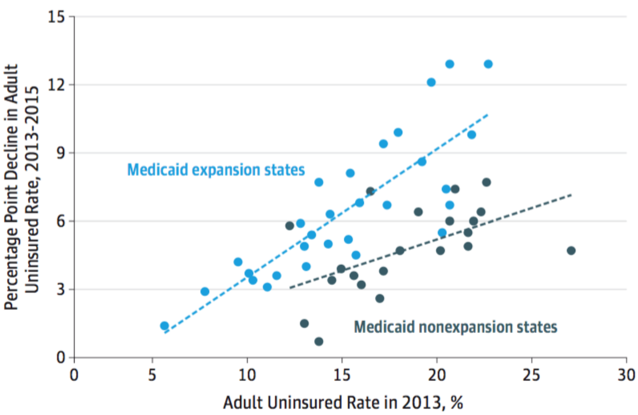


SAT scores and teacher salaries (state data from 2010)



stratification and/or multiple regression: Obama's 2016 single author JAMA paper

Figure 2. Decline in Adult Uninsured Rate From 2013 to 2015 vs 2013 Uninsured Rate by State



See also "Statistical methods in the NEJM" (2007)

Kidney stones (Wikipedia Simpson's Paradox)

	Treatment A	Treatment B
Small stones	*Group 1* **93% (81/87)**	*Group 2* 87% (234/270)
Large stones	*Group 3* **73% (192/263)**	*Group 4* 69% (55/80)
Both	78% (273/350)	**83% (289/350)**

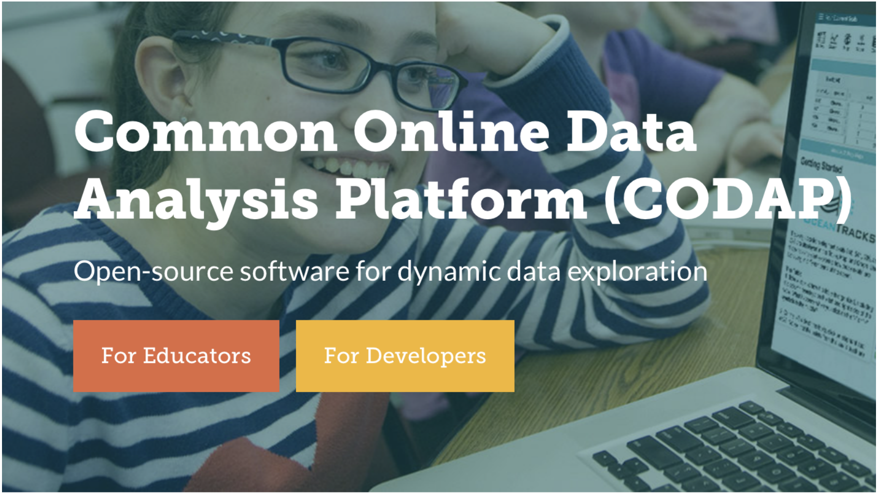
How to handle more than two variables (in R)?

- "Data Viz on Day One" (TISE, http://escholarship.org/uc/item/84v3774z)

How to handle more than two variables (in R)?

- "Data Viz on Day One" (TISE, `http://escholarship.org/uc/item/84v3774z`)
- stratification
- multiple regression (early and often)
- straightforward to use mosaic package "Less Volume, More Creativity" approach to modeling (*R Journal*, `https://journal.r-project.org/archive/2017/RJ-2017-024` and related Little Books)

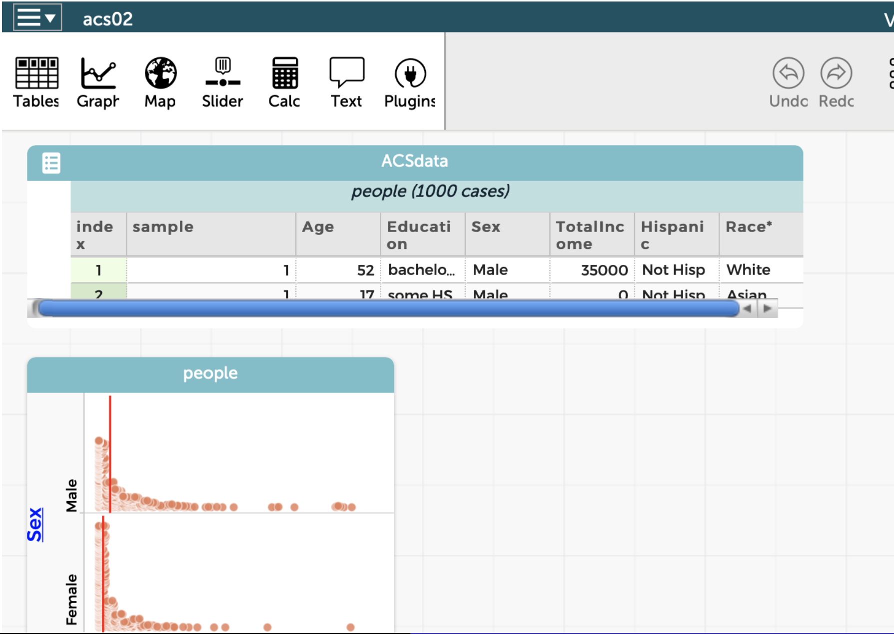
How to handle more than two variables (in CODAP)?



More on CODAP

- CODAP is free educational software for data analysis.
- This web-based data science tool is designed as a platform for developers and as an application for students in grades 6-14.
- See interactive examples from this summer's Teach Data Science blog: `https://teachdatascience.com/codap`

How to handle more than two variables?



De Veaux, Velleman, and Bock (Intro Stats, 5th edition)

- introduces multivariate thinking in most chapters in chapters 1-8
- introduces multiple regression early in the book (chapter 9) purely descriptively
- returns to inference for multiple regression later
- also discusses stratification and similar approaches
- see `https://nhorton.people.amherst.edu/is5/` for illustrated examples in R using mosaic "Less Volume"

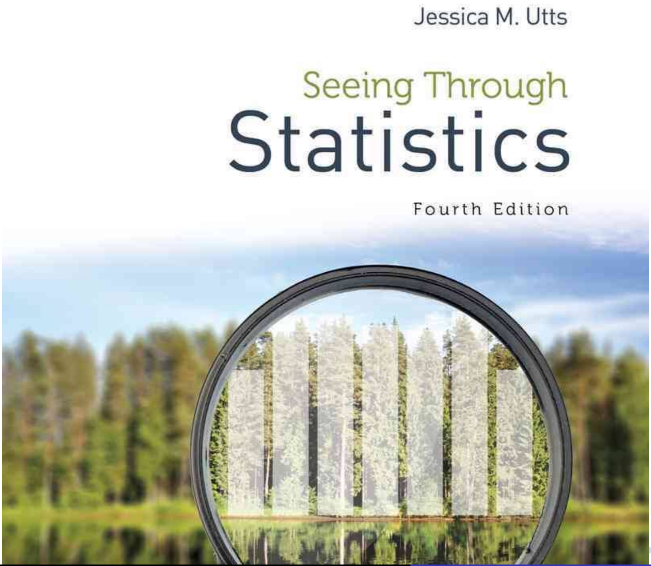
Multivariate thinking and confounding

AP Statistics Vocabulary

Both Sides

confounding

when the levels of one factor are associated with the levels of another factor so their effects cannot be separated

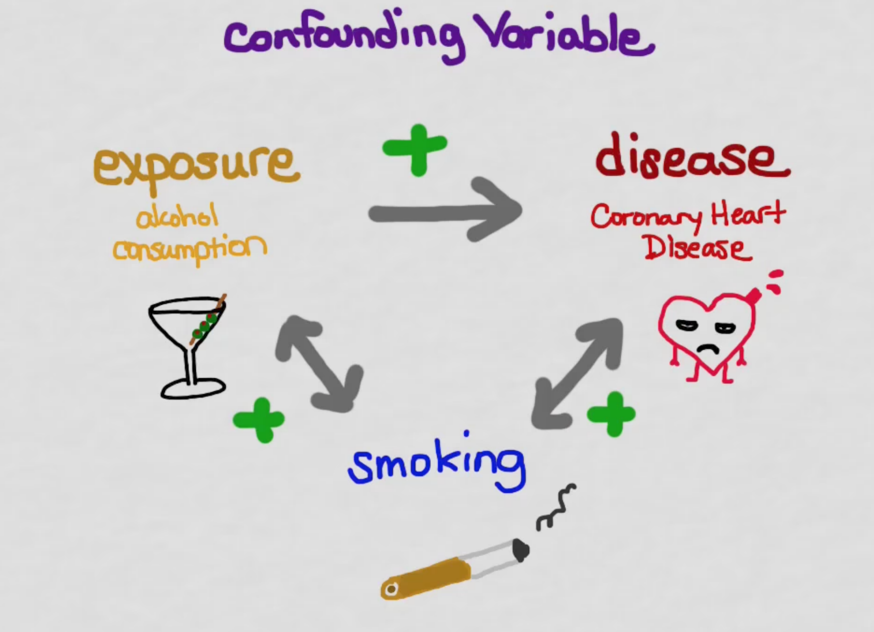

Jessica M. Utts
Seeing Through
Statistics
Fourth Edition

Utts: Seeing through Statistics (4th edition)

A confounding variable is one that has two properties.

1. A confounding variable is related to the explanatory variable in the sense that individuals who differ for the explanatory variable are also likely to differ for the confounding variables.
2. A confounding variable affects the response variable. Because of these two properties, the effect of a confounding variable on the response variable cannot be separated from the effect of the explanatory variable on the response variable.

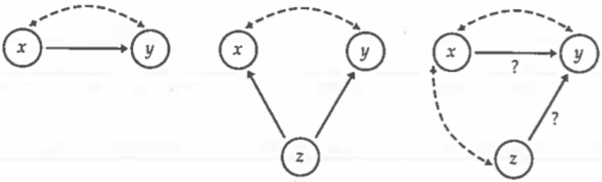
causal graphs (thanks to Elizabeth Lynch)



Explaining association: causation

Figure 2.29 shows in outline form how a variety of underlying links between variables can explain association. The dashed double-arrow line represents an observed association between the variables x and y. Some associations are explained by a direct cause-and-effect link between these variables. The first diagram in Figure 2.28 shows "x causes y" by a solid arrow running from x to y.

Items 1 and 2 in Example 2.42 are examples of direct causation. *Even when direct causation is present, very often it is not a complete explanation of an association between two variables.* The best evidence for causation comes from experiments that actually change x while holding all other factors fixed. If y changes, we have good reason to think that x caused the change in y.



Causation
(a)

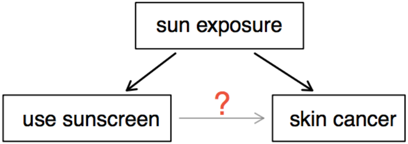
Common response
(b)

Confounding
(c)

Example

Suppose an observational study tracked sunscreen use and skin cancer, and it was found that the more sunscreen someone used, the more likely the person was to have skin cancer. Does this mean sunscreen causes skin cancer?

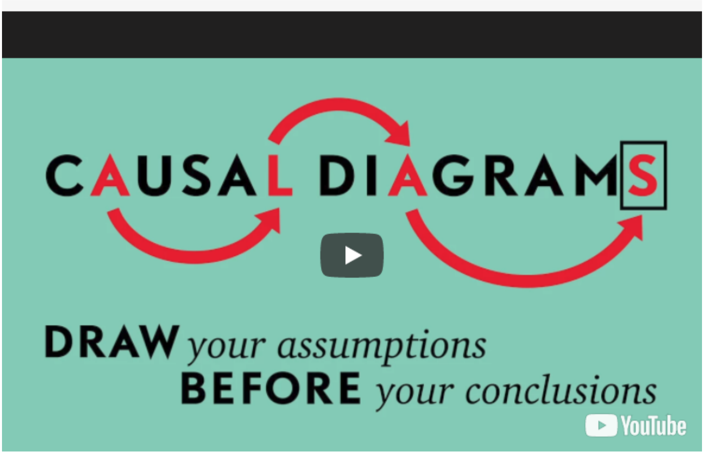
Causal graphs version 3.0 (Open Intro)



Sun exposure is what is called a **confounding variable**,[13] which is a variable that is correlated with both the explanatory and response variables. While one method to justify making causal conclusions from observational studies is to exhaust the search for confounding variables, there is no guarantee that all confounding variables can be examined or measured.

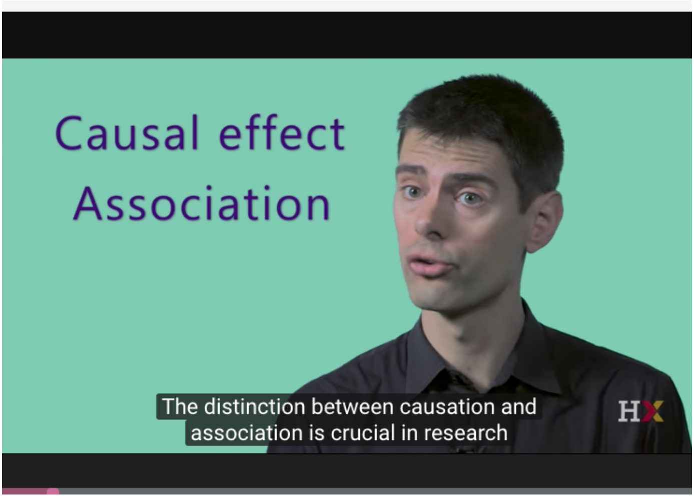
version 3.0: EdX Causal Diagrams (Draw Your Assumptions Before Your Conclusions course)

2. What is a DAG?



version 3.0: EdX Causal Diagrams (Draw Your Assumptions Before Your Conclusions course)

Association vs. Causation



version 3.0: EdX Causal Diagrams (Draw Your Assumptions Before Your Conclusions course)

1. Causal DAGs
2. Confounding
3. Selection Bias
4. Measurement Bias and putting it all together

Learning outcomes (EPI524 Hernan and Swanson)

At the completion of the course, you will be able to:

1. Explain why models are necessary for confounding control
2. Control for confounding using various modeling approaches
3. Identify the relative advantages and disadvantages of each modeling approach
4. Recognize and formulate well defined questions concerning causal effects

My proposal (a reinvigorated intro stat or QR course)

- Include multivariate thinking as part of descriptive statistics
- Include multiple regression as part of that introduction
- Address confounding by stratification and multiple regression control
- Introduce idea of a causal graph
- (Avoid paralysis and paranoia re: "other factors")

My proposal (second course in statistics)

- Extend ideas of causal graph and formalize causal inference
- Incorporate Hernan EdX course to save on class time
- Focus on what causal graphs tell us to do in terms of multiple regression, exposure to one or more other methods to address confounding

My proposal (Capstone and other courses)

- Reinforce and extend these ideas in later courses, projects, workshops, co-curricular experiences

Making this happen

- design and confounding are arguably the most important statistical foundation topics (after the concept of variability)
- computing environments are free and much easier to use
- new curricular models and materials have been created (more needed)
- need to rethink how we integrate this material into our courses (and promulgate the approach)

Teaching multivariate thinking and confounding

1. landscape of statistics in high school
2. what are we teaching?
3. multivariate thinking and confounding 101 and 201
4. closing thoughts: developing data acumen

Closing thought: need to avoid paralysis

- ensure that students don't get stuck (conclude they can't make any headway if data don't arise from a randomized trial)
- teach (modern) design early and often
- reinforce key aspects (observational data vs. randomized trials) when we teach inference
- teach techniques to move beyond two-sample t-test (stratification and multiple regression)
- make room for more components of *data acumen* by simplifying (what if all datasets were $n > 100$? what if p-values were de-emphasized? ATOM: Accept uncertainty. Be Thoughtful, Open, and Modest (Wasserstein et al., 2019))

Closing thought: back to data acumen

> *The ability to take data – to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it's going to be a hugely important skill in the next decades, not only at the professional level but even at the educational level for elementary school kids, for high school kids, for college kids. Because now we really do have essentially free and ubiquitous data. So the complimentary scarce factor is the ability to understand that data and extract value from it.*

`https://flowingdata.com/2009/02/25/googles-chief-economist-hal-varian-on-statistics-and-data`

Closing thought: other aspects of data acumen

Key **communication and teamwork** concepts/skills that would be important for all students in their data science programs and critical for their success in the workforce are the following:

- Ability to understand client needs
- Clear and comprehensive reporting
- Conflict resolution skills
- Well-structured technical writing without jargon, and
- Effective presentation skills

Closing thought: other aspects of data acumen

Key **workflow and reproducibility** concepts/skills that would be important for all students in their data science programs and critical for their success in the workforce are the following:

- Workflows and workflow systems
- Reproducible analysis
- Documentation and code standards
- Source code (version) control systems, and
- Collaboration

Closing thought: other aspects of data acumen

Key aspects of **ethics** needed for all data scientists (and for that matter, all educated citizens) include the following:

- Ethical precepts for data science and codes of conduct
- Privacy and confidentiality
- Responsible conduct of research
- Ability to identify "junk" science, and
- Ability to detect algorithmic bias

Final closing thought: AP Computer Science Principles

Big Idea 3: Data and Information

Data and information facilitate the creation of knowledge. Computing enables and empowers new methods of information processing, driving monumental change across many disciplines — from art to business to science. Managing and interpreting an overwhelming amount of raw data is part of the foundation of our information society and economy. People use computers and computation to translate, process, and visualize raw data and to create information. Computation and computer science facilitate and enable new understanding of data and information that contributes knowledge to the world. Students in this course work with data using a variety of computational tools and techniques to better understand the many ways in which data is transformed into information and knowledge.

AP Computer Science Principles: taught for first time in 2017

Enduring Understandings (Students will understand that ...)	**Learning Objectives** (Students will be able to ...)
EU 3.1 People use computer programs to process information to gain insight and knowledge.	**LO 3.1.1** Find patterns and test hypotheses about digitally processed information to gain insight and knowledge. [P4]

AP Computer Science Principles: project based learning

LO 3.1.3 Explain the insight and knowledge gained from digitally processed data by using appropriate visualizations, notations, and precise language. [P5]

EK 3.1.3A Visualization tools and software can communicate information about data.

EK 3.1.3B Tables, diagrams, and textual displays can be used in communicating insight and knowledge gained from data.

EK 3.1.3C Summaries of data analyzed computationally can be effective in communicating insight and knowledge gained from digitally represented information.

EK 3.1.3D Transforming information can be effective in communicating knowledge gained from data.

EK 3.1.3E Interactivity with data is an aspect of communicating.

AP Computer Science Principles: 200 page course description

EU 3.2 Computing facilitates exploration and the discovery of connections in information.

LO 3.2.1 Extract information from data to discover and explain connections or trends. [P1]

AP Computer Science Principles: example question

8 of 9

Biologists often attach tracking collars to wild animals. For each animal, the following geolocation data is collected at frequent intervals.

- The time
- The date
- The location of the animal

Which of the following questions about a particular animal could NOT be answered using only the data collected from the tracking collars?

AP Computer Science Principles: example question

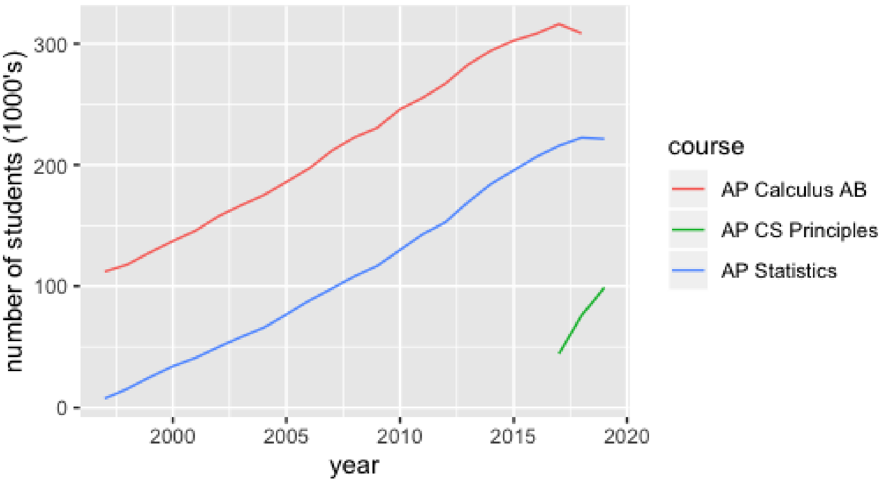
Approximately how many miles did the animal travel in one week?

Does the animal travel in groups with other tracked animals?

Do the movement patterns of the animal vary according to the weather?

In what geographic locations does the animal typically travel?

Advanced Placement Statistics



Data acumen and data numeracy: helping students extract meaning from data

Nicholas J. Horton

Department of Mathematics and Statistics
Amherst College, Amherst, MA, USA

National Numeracy Network Conference, October 12, 2019

nhorton@amherst.edu
http://nhorton.people.amherst.edu/nnn2019-10.pdf