

# Statistics and Society

Updating the curriculum of an introductory statistical literacy course for the modern student

Ellen Gundlach

Continuing Lecturer and Education Specialist

Department of Statistics

Purdue University

[gundlach@purdue.edu](mailto:gundlach@purdue.edu)

# Class structure for STAT 113

<b>Traditional Supplemental</b>	<b>Fully Online</b>	<b>Flipped</b>
~360 students	80 students	60 students
T/Th lectures in large hall with Ellen Gundlach  Active M recitations with TA	Everything (except exams) done online	Lectures watched online  Th classes with Ellen Gundlach for group work

# How my students have changed 2012-2015

- More have prior statistics experience in other classes, either at Purdue or in high school due to Common Core.
- Many are taking the course to fulfill either Information Literacy or Science core outcomes. (STAT 113 qualifies for both.)
- International and domestic students who need help formulating and judging logical arguments with scientific evidence, having intelligent discussions.
- All students want to understand how this material is relevant to their daily lives and future careers.

# Who the students are

- Liberal arts
- Communications and journalism
- Film studies and theater
- Education
- Nursing
- Financial planning
- Economics and management
- Health and fitness

With sprinklings of:

- Science
- Engineering
- Pharmacy, pre-med, pre-vet

Statistical literacy  $\neq$  introductory statistics “lite”

STAT 113 is a **statistical literacy** course, not a data analysis course.

Where the data comes from, media stories > data analysis and calculations.

Training consumers of statistics and world citizens.

What can we do to add more value to the course with new concepts and activities?

What would set STAT 113 apart from all the other introductory statistics courses in the department?

# Statistical literacy

Statistical literacy: being able to take information and explain it, judge it, evaluate it, and make decisions based on that information.

--Rumsey (2002)

[We need to create] educated citizens [who] understand basic statistical concepts and interpret and critically evaluate statistical messages so that they can detect any misuse of statistics by policymakers, physicians, and others.

--Tishkovskaya and Lancaster (2012)

A judicious man uses statistics, not to get knowledge, but to save himself from having ignorance foisted upon him.

--Thomas Carlyle

# Guidelines for Assessment and Instruction in Statistics Education (GAISE)

Goals for producing statistically literate students include:

- Emphasize statistical literacy and develop statistical thinking.
- Use real data and stories.
- Stress conceptual understanding rather than a cookbook of procedures.
- Foster active learning.

# Statistics and **SOCIETY**: new topics

- Placebo effect and nocebo effect (Is Wind Turbine Syndrome a real thing? Colbert Report video: <http://thecolbertreport.cc.com/videos/wjev3/wind-power-s-health-hazards>)
- Ethics of doing experiments with humans and animals (Facebook experiment)
- Autism and vaccines
- Understanding if an advertiser's claims are valid (Can you trust Dr. Oz?)
- Big data
- Interpreting medical test results (false positives, false negatives)
- Cancer clusters
- Service learning project with Convos concert planning group



# How the course is planned

1<sup>st</sup> third:

Basics of data collection, ethics, causation, government statistics, Big Data.

2<sup>nd</sup> third:

Graphing, summary statistics for one and two variables, Normal distribution.

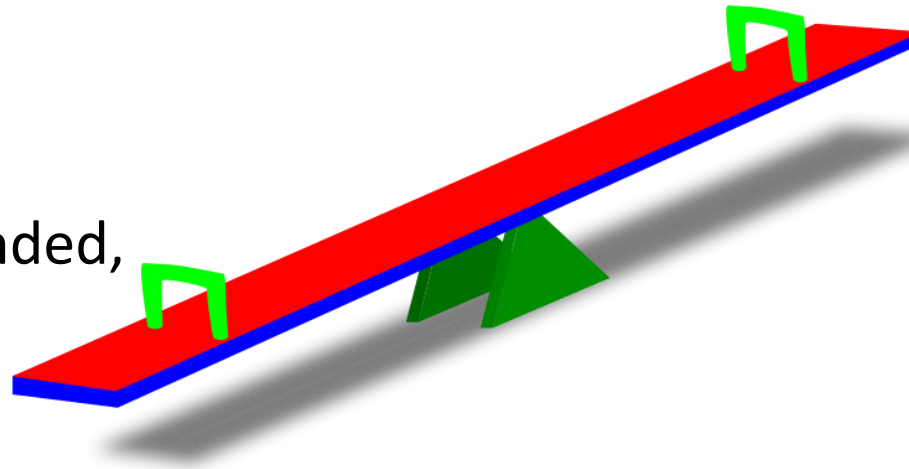
3<sup>rd</sup> third:

Probability, confidence intervals, the logic of hypothesis testing.

# Assessments for these new topics

- Advertisement group project presentation
- Data diary with discussion
- Social media story collection/discussion projects
- Convos service learning project
- Exam questions???

Open-ended,  
creative  
projects



Ease of  
presentation/  
grading for 500  
students

# Advertisement group presentation

- What specific claim is being made by the advertiser?
- Where did the advertiser get this information?
- Find an outside resource that talks about this same issue and summarize the conclusion.
- What do you think? Could the advertiser's claim be true?
- Do you think this is really a case of common response or confounding? What are some possible lurking variables? Explain your answer.

Thanks to Rose Martinez-Dawson for her two eCOTS presentations:

“Using advertisements to teach statistical literacy” <https://www.causeweb.org/ecots/ecots12/breakouts/2/>

“Using creative inquiry to understand the role of statistics in undergraduate research” <https://www.causeweb.org/ecots/ecots14/36/>

# Student feedback on Advertisement project

- "The most interesting concept from this week's lecture to me was watching the different ads and evaluating the claims. I found it interesting because I never thought twice when watching these ads on TV. I just assumed since the information was being presented in a positive manner, the information was correct. Having learned the material from our class, it allows me to evaluate more critically the claims that these companies are making so that I can determine for myself whether or not their product is worth it."
- "I learned how hard it can be to get information about a product, such as how they make and back up their specific claims. Sometimes websites don't want to give this information out to you, which makes you question how they can make claims to the public without giving them the information."

# Big data: where to begin?

- Teaching Big Data analytical techniques to undergraduates has become a hot topic in the Statistics Education field (1 of 3 primary themes for eCOTS 2014).
- How to teach this to a non-analytical statistical literacy class? Make them aware of how their own data is being collected, used, and protected (or not).
- No prepared materials available for this level.
- Read articles and books intended for other audiences, talked with leading statistics educators Rob Gould (UCLA) and Nicholas Horton (Amherst and Smith Colleges). Also talked to the IT people on campus.

Big data is neither good nor bad.  
Big data is neither ethical nor unethical.

The point of this unit is not to scare or outrage you.

The point of this unit is to:

- Introduce you to the idea of big data.
- Show you some of the great things big data can accomplish.
- Make you aware of how much data you produce every day.
- Empower you to ask questions about who collects your data, how they use it, and how they protect your privacy.

*As a society, we are still working out the rules for big data, but we are all affected by it and contribute to it.*

*Data is a currency just like gold. Data is neither good nor bad, neither ethical nor unethical. How it is collected and used makes all the difference.*

# Data brokers

Companies that analyze and sell huge amounts of consumer data for marketing purposes.

One data broker had 3,000 data categories for nearly every individual American consumer.

No way for the consumer to correct errors.

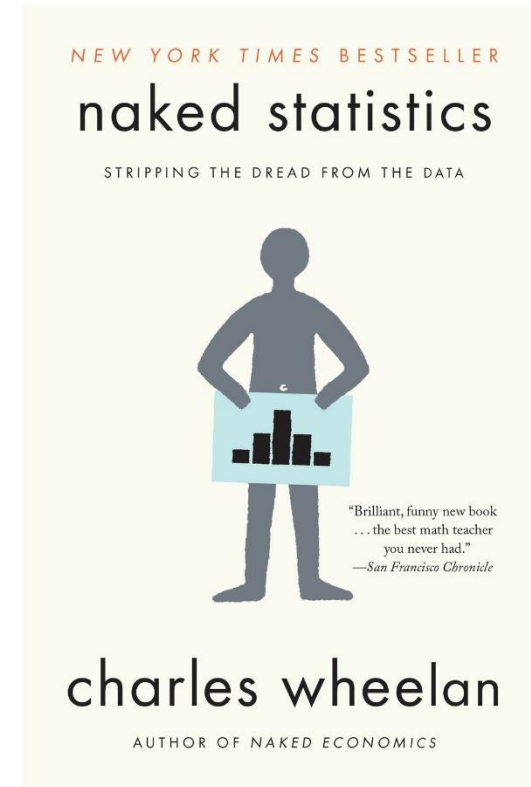
Federal Trade Commission has very little control over them, partly because the consumer is not the customer but the product.

# How much does Target know about you?

**WATCH:** <http://www.c-span.org/video/?c4503500/naked-statistics-target-data> (5 1/2 minutes, good story!)

## Naked Statistics Target Data

Naked Statistics book reading, story about Target's data collection and the pregnant teenager.



From (whole thing is an hour and really good, including the Q&A at the end):  
<http://www.c-span.org/video/?310325-1/book-discussion-naked-statistics>



# Trust the Education Department With a Student Database?

May 12, 2014 *Chronicles of Higher Education*

A proposal for a detailed database of all college students has surfaced. The proposed "unit record" database would require every college student to submit extensive personal information to the government as a condition of receipt of federal student aid and college enrollment. Information collected may include: family socioeconomic background, elementary- and secondary-school records, and health records.

**Good for research.**

**Bad because:**

- Can you trust the federal government with sensitive data security?
- Informed consent?

# Ginger.io: health care with your cell phone



<http://vimeo.com/73484315>

# The ethics of educational uses of big data

- Are you able to give informed consent when using educational tools like Blackboard?
- Does the college administration let students know their academic behaviors are being tracked?
- Yet another set of concerns arises because a lot of the new educational data collection is proprietary. Companies like Pearson, Blackboard and Coursera each have information on millions of learners.
- Who owns this data? The student, the institution, the company or some combination? Who gets to decide what is done in whose best interest?

<http://www.npr.org/blogs/ed/2014/07/04/327745863/big-data-comes-to-college>

<http://www.educause.edu/ero/article/ethics-big-data-and-analytics-model-application>

# Big data and data privacy in the news daily

White House's draft of a consumer privacy bill is out — and even the FTC is worried The White House's consumer privacy bill of rights draft is already facing growing criticism.

<http://wapo.st/1BJT8fl>

Washington Post, 2/27/15

"In this digital age, particularly as big data innovations drive advances across our economy, more and more data about Americans is being collected and stored," a fact sheet announcing the draft legislation said. "And, even though responsible companies provide us with tools to control privacy settings and decide how our personal information is used, too many Americans still feel they have lost control over their data."

# Nicholas Felton: A Quantified Life

[http://www.nytimes.com/video/technology/100000003050767/nicholas-felton-a-quantified-life.html?emc=edit\\_th\\_20140820&nl=todaysheadlines&nid=61564825](http://www.nytimes.com/video/technology/100000003050767/nicholas-felton-a-quantified-life.html?emc=edit_th_20140820&nl=todaysheadlines&nid=61564825)

Short video about a man who records EVERYTHING about his life and wants you to know what is being recorded from yours.

# Additional reading on data privacy

- Big data's impact in the world

<http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html>

- Mobile Users Privacy Bill of Rights

<https://www.eff.org/deeplinks/2012/03/best-practices-respect-mobile-user-bill-rights>

- Ginger.io privacy policy

<http://ginger.io/privacy-policy/>

- World Economic Forum big data paper

<http://www.weforum.org/reports/personal-data-emergence-new-asset-class>

- White House paper on data privacy

<http://www.whitehouse.gov/sites/default/files/privacy-final.pdf>

# Why smart statistics are the key to fighting crime

Anne Milgam's TED Talk (12 minutes 41 seconds, but very interesting)

<https://www.youtube.com/watch?v=ZJNESMhIxQ0>



# Data diary project

- Fill out the 24-hour data diary.
  - Time of day
  - Activity/website
  - What type of data was collected from you?
- What is the most concerning data that is being collected from you? Why does it concern you?
- Go read the full privacy policy for a site you use. Does it specifically list which pieces of information the company collects from you? If so, what are they?
- Search online to learn how big data is improving something that interests you. (E.g., sports, health, farming, shopping, entertainment)
- Write a big data privacy “bill of rights”
  - What is ok for these companies to know about you?
  - How should they be able to ask about this information
  - What should they be allowed to do with your data?
  - How should they have to protect your data?



# Student feedback on Data Diary project

- "The Data Diary project is awesome!! I never realized how much information websites can get from you without you knowing. Even when they ask you something directly they can get a lot more out of the question than you think they can. It was interesting to actually read a Privacy Policy, even though I will admit I was dreading that part of the assignment. It made me think about the information that I give out and that people can probably easily find out a lot of things about me."
- "I have learned that there are many, many times I get on my computer to do something that is leaving a data trail. And every time I pick up my phone I could be clicking on an app that is doing the same. Until you are assigned a project to pay close attention to these things, it's like we never realize everything we do."
- "I think Big Data has been the most interesting topic so far. I think that it is so cool that companies can see statistical data that represents certain aspects of society based on usage of the internet, phones, and many more devices, and go ahead and use those statistics to make products that would be more effective for the consumers. Although it may seem a bit questionable in regards to privacy, I think it is really cool in how technology can provide information to better serve general aspects of living such as movies, sports, shopping, and so on."

# Convos service learning project



<https://www.purdue.edu/convocations/portfolio/pentatonix-10-1-14/>

# Convos service learning project

Video for our introductory “consulting” meeting: [http://youtu.be/tJu\\_Xq-fw2g](http://youtu.be/tJu_Xq-fw2g)

Convos needs help figuring out which artists to bring to campus for the student concerts.

Surveys have low response rate and don’t represent study body.

Students are unreliable about buying tickets for shows the surveys showed were popular.

Why?

**CONVOS**  
PURDUE CONVOCATIONS

**SCC**  
STUDENT CONCERT  
COMMITTEE

# Convos service learning project: Phase 1

Our students work in groups to collect lists of reasonably priced artists suggested by the student body.

Convos gave us a budget of \$100,000 for the artist's fees (spending another \$100,000 on tech, catering, security, etc.) and a website to look up artist's fees: <http://eventresourcespresents.com/natartist.cfm>

Students provide discussion of:

- Their sampling methods
- Potential problems with their methods (bias, nonresponse, undercoverage, etc.)
- Whether their sample truly represents the student body
- How Convos should contact students and reduce nonresponse for the official survey later
- How Convos can improve consistency between students' survey answers and actual purchasing behavior
- If you redid this project, what would you do differently?
- What did you learn by doing this project that you didn't learn through lecture or homework?

# Convos service learning project: Part 2

- Convos uses our consulting reports to create an official online survey for the student body.
- Convos uses our suggestions for disseminating the survey.
- Our students will receive class participation points for completing the survey. Hopefully they will encourage their friends to do so.
- In the fall 2015 semester, an artist selected by the survey will come to Purdue to perform.

# Student feedback on Convos project

- “Even though conducting an experiment or a SRS is the best way to find data that is unbiased, sometimes you have to use a convenience sample.”
- “I learned how difficult it is to design a practical, useful survey and how it is quite a challenge to get people to participate.”
- “You have to be very specific when conducting a survey. You may think the question is clear because you created it, but your population doesn’t know what you’re looking for completely.”

# Mixable media story collection/discussion projects

#1: Find a story about an **observational study or an experiment**. Discuss features of how the data was collected such as sample size, how the sample was selected, if we know specific details about the questions asked to the individuals, potential lurking variables, whether the principles of good experimental design were used, whether the experiment was conducted ethically, etc.

#2: Find a story which includes one of the **graphs** we have studied. Discuss whether the graph was used appropriately, what type and how the data was collected, if another graph might be more appropriate, whether the correct conclusions were presented, whether the graph was clearly displayed, etc.

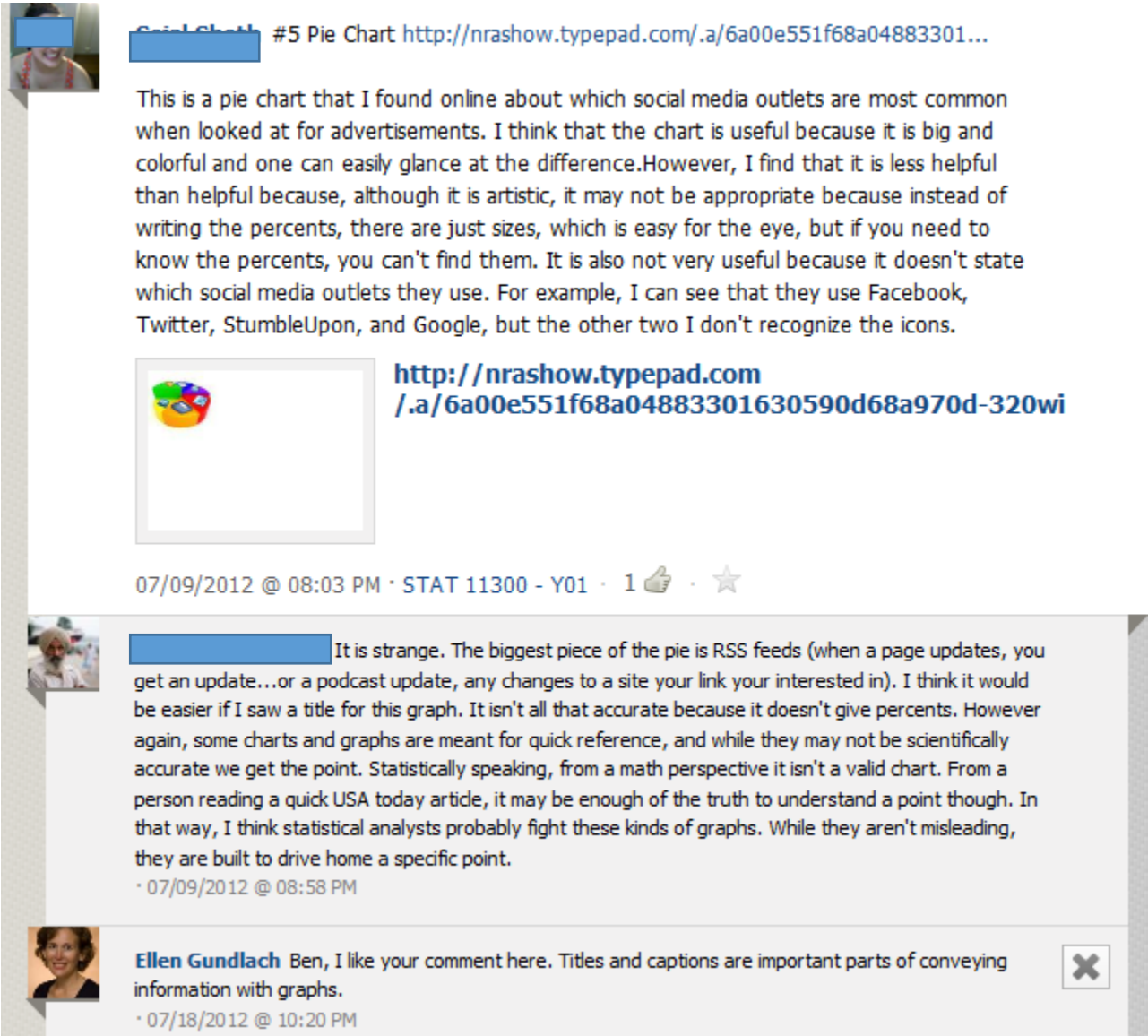
#3: Investigate one of your own **good or bad habits** with a research-based article or video. Discuss the quality of the evidence and important features of the research design.

# Mixable

Works just like Facebook, but FERPA compliant.

Post and discuss a relevant article/video that uses specific statistics concepts.

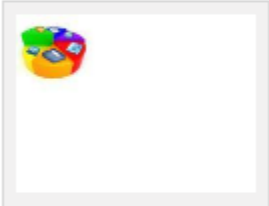
Make 4 “statistically intelligent” comments on other students’ posts.





The screenshot shows a social media interface with three posts. The top post is by a user with a blue profile picture, titled "#5 Pie Chart" with a link to a Typepad page. The post text discusses a pie chart about social media outlets. Below the text is a placeholder for a pie chart image and a full URL. The post has a timestamp of 07/09/2012 @ 08:03 PM, 11300 views, and 1 like. The middle comment is by a user with a blue profile picture, discussing the accuracy of the pie chart. The bottom comment is by Ellen Gundlach, praising the comment and mentioning titles and captions.

**#5 Pie Chart** <http://nrashow.typepad.com/.a/6a00e551f68a04883301...>


This is a pie chart that I found online about which social media outlets are most common when looked at for advertisements. I think that the chart is useful because it is big and colorful and one can easily glance at the difference. However, I find that it is less helpful than helpful because, although it is artistic, it may not be appropriate because instead of writing the percents, there are just sizes, which is easy for the eye, but if you need to know the percents, you can't find them. It is also not very useful because it doesn't state which social media outlets they use. For example, I can see that they use Facebook, Twitter, StumbleUpon, and Google, but the other two I don't recognize the icons.

 <http://nrashow.typepad.com/.a/6a00e551f68a04883301630590d68a970d-320wi>

07/09/2012 @ 08:03 PM · STAT 11300 - Y01 · 1  · 

**It is strange.** The biggest piece of the pie is RSS feeds (when a page updates, you get an update...or a podcast update, any changes to a site your link your interested in). I think it would be easier if I saw a title for this graph. It isn't all that accurate because it doesn't give percents. However again, some charts and graphs are meant for quick reference, and while they may not be scientifically accurate we get the point. Statistically speaking, from a math perspective it isn't a valid chart. From a person reading a quick USA today article, it may be enough of the truth to understand a point though. In that way, I think statistical analysts probably fight these kinds of graphs. While they aren't misleading, they are built to drive home a specific point.

· 07/09/2012 @ 08:58 PM

**Ellen Gundlach** Ben, I like your comment here. Titles and captions are important parts of conveying information with graphs. 

· 07/18/2012 @ 10:20 PM



## What I've learned

- Autism-vaccines has been tricky to talk about—important to present the science while also expressing sympathy for smart but scared parents who were really trying to do the best for their kids. Give historical perspective of ice cream/polio story. Perfect timing with Disneyland measles outbreak this year.
- Big data can be a scary topic if not presented correctly. However, my students' generation is not nearly as shocked by the invasion of privacy as my generation is.
- The students love that the Convos project is “real” instead of just a class assignment. The opportunity to see how difficult it is to collect data well is important. They needed to be reassured it was actually going to be shared with Convos instead of being just an elaborate homework assignment.

# Challenges for the future

- Am I spending too much time pointing out flaws in the media instead of showing what is done well? Will they trust anybody after taking this class? *(And should they?)*
- Figuring out a better definition for what big data actually is.
- How to make group work more effective?
- Students said having 2 weeks to collect data for the Convos project would have been better than 1 week during an exam-heavy time of the semester. Also need a better way to organize all the data the students collect.
- Purdue has over 20% international students, and sometimes we run into cultural differences we weren't expecting. I am coordinating with international student specialists this semester to learn new strategies for explanations, active learning, and group work.