Elements of Effective, Fun, Classroom Activities [1]

Deborah Nolan
University of California, Berkeley
Andrew Gelman
Columbia University, New York

1 Introduction

Classroom activities and projects are increasingly popular ways for conveying important statistical concepts to students. Well designed activities make effective teaching tools, but poorly designed ones waste precious class time and may confirm students' preconceptions that statistics is boring and confusing. In this talk, we dissect several demos, classroom exercises, and projects, discussing their strengths and their flaws and making suggestions for improvements. Through these examples we hope to outline some of the key elements of good classroom activities.

2 Goals of classroom activities

DO: Take 5 minutes to think of a classroom activity that you like to use in your teaching or that you enjoyed participating in as a student. Write a description of the activity on a 3x5 card. Find a partner and trade cards and discuss why the activity that you chose was particularly successful. Together with your partner make a list of three reasons why these in-class activities were particularly successful. Write these on the back of your 3x5 cards.

In the next few sections, through examples we present the following ten goals for using classroom activities:

1. Ice-breaker
2. Increase and improve classroom participation
3. Demonstrate a key concept
4. Confront a common misconception
5. Use a punch line to help remember a key concept
6. Calibrate students' understanding of a topic
7. Have students learn from each other
8. Precursor to project work - set the stage and get students to figure out who to work with

[1] Many of the activities presented here have been taken from *Teaching Statistics* by Gelman and Nolan (2002, Oxford University Press)

9. Relate statistics learned in class to current events

10. Make students aware of the use of statistics in addressing important problems

These examples are grouped according to the type of activity: small-group in-class activities, classroom demonstrations, and handouts for students to complete in class. In addition, we discuss some of the pitfalls we have encountered in using activities in class and provide examples of these.

3 Small-Group and Partner Activities

1) Ice-breaker Activity: The goal of an activity need not always be to convey a statistical concept; for example, an activity may be useful as an ice breaker at the beginning of a course and a way for the instructor and students to learn each others names.

At the beginning of a statistics course we traditionally collect data on students—each student is given an index card to write the answers to a series of questions and then these cards are collected and used in illustrations of statistical methods throughout the semester. For example, when we teach sampling, we use the cards as a prop to represent the class population; we mix them up on our desk and have students pick cards at random for a sample. Also, when we teach regression and correlation, we bring to class a scatter plot of student data and, for example, compare a scatter plot of the height and hand span for the students in our class to Pearson's data on university students that was collected over one hundred years ago.

Possible information that we ask students to provide includes questions on student activities outside the classroom, such as the amount of soda drunk yesterday or the time spent watching television the previous night. Physical measurements, such as the span from the thumb to the little finger, can be measured on the spot by handing out a sheet of paper with the photocopy of a ruler on it. Students enjoy completing the handedness form below, which is the basis for a discussion of categorical versus continuous data and for an introduction of distributions and histograms. Try your hand at it – below sketch your guess for the histogram of the scores for the people in this room.

We sometimes reuse the results of an activity to drive a point home, other times we revisit the activity because there are multiple lessons to be gained from it, and further if we collect data from/with the students we analyze the data throughout the semester to illustrate different statistical techniques.

Please indicate which hand you use for each of the following activities by putting a + in the appropriate column, or ++ if you use would never use the other hand for that activity. If in any case you are really indifferent, put + in both columns. Some of the activities require both hands. In these cases the part of the task, or object, for which hand preference is wanted is indicated in parentheses.

Task	Left	Right
Writing		
Drawing		
Throwing		
Scissors		
Toothbrush		
Knife (without fork)		
Spoon		
Broom (upper hand)		
Striking match (hand that holds the match)		
Opening box (hand that holds the lid)		
Total		

Right − Left: Right + Left: $\frac{\text{Right} - \text{Left}}{\text{Right} + \text{Left}}$:

Create a Left and a Right score by counting the total number of + signs in each column. Your handedness score is (Right − Left)/(Right + Left): thus, a pure right-hander will have a score of score $(20 - 0)/(20 + 0) = 1$, and a pure left-hander will score $(0 - 20)/(0 + 20) = -1$.

2) Increase participation – Students can contribute to a classroom discussion in many ways other than the traditional approach of raising a hand to answer aloud a question raised by the instructor. For example, students can be asked to write responses on paper, which are read aloud by the instructor in class; they can write an answer to a question on the board, where the instructor leads the class through the answer; or, students working in groups can prepare a presentation for the rest of the class, where each student has a part in the presentation. These approaches provide effective alternatives to the typical classroom discussion and they help set the expectation that students are to be actively engaged while in the classroom.

As an example, a news clipping that reports on a sample survey with the potential for several types of bias works well to get students discussing issues central to survey sampling. The newspapers are full of reports on the latest survey results, and clippings can quickly generate discussion among students, especially at the start of a course. For one of our first activities, we hand out a news clipping and after reading the article to the class, students work in pairs and write down one criticism and one positive comment about the study. We use these critiques to begin defining various types of bias in survey sampling—nonresponse, measurement, selection, question-wording, and so forth. Students hand in their comments with no names on them, and without sorting or looking through the papers, we take the top sheet from the stack, read it aloud, and discuss with the class. This approach gives students a chance to think and discuss a topic in a small group before it is discussed with the entire class, which helps prepare them for classroom discussion. We continue reading papers, taking them one by one from the stack until we have heard and discussed a variety of ideas.

After class, we read through the remainder of the papers and select those with ideas that we did not cover and think would be good to address. We also make a pile of papers that contain incorrect statements or that indicate basic misconceptions. We bring these two sets of papers to the next class and continue our discussion. We have found it helpful to identify and address basic misunderstandings as quickly as possible. Also, we have found that maintaining anonymity of the authors of the papers is crucial for guaranteeing complete student participation. And, as always, having students work in pairs reduces the pressure of performance and helps make this a fun activity.

4 Demos

We have found student-participation demonstrations to be effective in dramatizing concepts that students often find difficult (for example, numeracy, conditional probability, the difference between an experiment and a survey, statistical and practical significance, the sampling distribution of confidence intervals). Demonstrations are also useful in clearing up commonly held misconceptions. When students put their misconception to use where it fails them, it makes a good lesson because it surprises the students and challenges their strongly held beliefs.

3) Key Concept Demo: In our experience, a classroom demonstration that includes student helpers or subjects tends to keep the interest of all of the students, not unlike the magicians request for a volunteer from the audience. We use the famous tea-tasting experiment of R.A. Fisher's as a model for one classroom demo, where we have one student attempt to distinguish between two kinds of soda. This taste test illustrates the principles of randomized experiments and hypothesis testing.

We bring to class two types of chilled soda, e.g. one diet and one regular cola; a bottle of water, small paper cups, paper towels for spills, and eight marbles in a sack, four each of two colors. We also have ready the hypergeometric probabilities for $n = 4$, $M = 4$, $N = 8$. The demo begins by us asking if any students can tell the difference between the two types of soda. We ask one of the "experts" to volunteer to demonstrate his or her abilities.

Before setting up the demonstration, we discuss the following question: What should be done about variations in the temperature, color, and so on? We tell them about some of the pitfalls we encountered in previous times we conducted this experiment. For example, one time our student expert noted that he could distinguish the two kinds of soda by the amount of bubbles he saw in the cups. He in fact was able to separate the cups perfectly, but then guessed which set of four was which and got it wrong! We ask students how they could guard against that happening again. While we can control some ways in which the two sodas differ from each other, other ways must always be dealt with by randomization. We wind up the discussion by going over the protocol used by Fisher, which we use in the demo: four pairs of cups, where each pair includes both types of soda. Sometimes we extract other possible designs from the class, and discuss their pros and cons.

We then have the student-expert leave the room with another student whose job is to keep watch over the expert while we set up the experiment. Two other student volunteers divide the eight cups into two sets of four. They mark each of the four cups in one set in an inconspicuous place, and fill the cups with the same kind of soda. The unmarked cups are filled with the other kind of soda. We draw marbles from the sack to determine the order in which the cups are put into a line - this is the randomization. When the student-expert returns, she sips the soda, separates the cups into two groups of four, and tells us which group is which. The water is to help her clear her palate between sips of soda. If the student makes no mistakes, we declare her an expert, and with one or more mistakes we say that she did as well as one might expect someone who was just guessing.

While the student is tasting and separating the cups, we continue the class discussion and focus on the hypothesis test. We ask, If we used ten pairs of cups, and our expert classified nine correctly, should we call her an expert or not? Many students at first think that the expert must properly classify all cups correctly, whether she has to taste 4 or forty, in order to be an expert. We also take this time to discuss what would the results of the experiment look like if the person wasn't an expert at all. We consider the probability that someone with no powers of discrimination could wind up with a particular outcome, where an outcome here is a certain number of correctly classified cups. That is, if someone can't distinguish between the two, then by just guessing, she should have a small chance of correctly determining which cups are which.

After the student presents her results, the volunteers determine if she is correct or not, and we compute the p-value and come to a conclusion, declaring the student as an expert or a fraud. We wind up the demo by discussing the protocol again, we interview the student asking her to reveal how she did it, and we ask the class what changes would they make to the experimental process if they had the opportunity to do it over again? Once in our post-experiment discussion, our expert confessed to having a stuffy nose and being unable to distinguish anything because of it!

4) Precursor to Project After our soda-tasting demonstration, we have students work in groups of two or three to conduct their own taste test in a directed project outside of class. Typically they compare two brands of food, such as peanut butter, chocolate, or coffee. We encourage them to be creative and conduct an experiment that they care about; not to simply settle for the milk test (2% versus fat-free) because it is easily available in the dining hall.

The classroom discussion that we carry out with the demo above gives them ideas for different experimental designs and it shows them the importance of identifying and controlling for factors that may influence the expert's ability to discriminate. Rather than simply copying the classroom experiment with a change in soda brands, students use the classroom demonstration as a springboard for creating their own experiments. One year, a group of students performed a rum-and-coke taste test. Since then, we make sure we discuss the ethics of experimentation on human subjects before letting them loose on the project.

5) Common Misconception Demo: Through a demo, students can be surprised to discover that they have a basic misunderstanding of a statistical concept. For example, the following coin-flipping demonstration highlights the differences between the leyman's understanding of "random" and the probabilistic concept of randomness, which is central to much of statistics.

The demonstration proceeds as follows. We pick two students to be "judges" and one to be the "recorder" and divide the others in the class into two groups. One group is instructed to flip a coin 100 times, or flip 10 coins 10 times each, or follow some similarly defined protocol, and then to write the results, in order, on a sheet of paper, writing heads as "1" and tails as "0". The second group is instructed to create a sequence of 100 "0"s and "1"s that are intended to *look like* the result of coin flips—but they are to do this without flipping any coins or using any randomization device (or consulting with the other group of students)—and to write this sequence on a sheet of paper. The recorder is instructed to copy these sequences onto two blackboards. We announce that the instructor and the judges will leave the room for five minutes while the students create their sequences, and then we will return and try to guess which sequence is from actual coin flips and which was made up.

Below are two binary sequences produced by students. Can you figure out which is the actual sequence of 100 coin flips and which is the fake?

00111000110010000100	01000101001100010100
00100010001000000001	11101001100011110100
00110010101100001111	01110100011000110111
11001100010101100100	10001001011011011100
10001000000011111001	01100100010010000100

We return to the room, examine the two sequences written on the blackboard, and ask the judges to guess which sequence is real. We then identify real and fake sequences; almost always the identification is correct, and the students are impressed. How did we do it? Well, even as the sequences are being written on the blackboard, the students notice a difference: the sequence of fake coin flips looks "random" in an orderly sort of way, with frequent switches between 0's and 1's, whereas the sequence of real coin flips has a "streaky" look to it, with one or more long runs of successive 0's or 1's. People generally believe that a sequence of coin flips should have a haphazard pattern, including frequent (but not regular) alternations between heads and tails. In fact, it is quite common for long runs of heads and tails to appear in sequences of random coin flips.

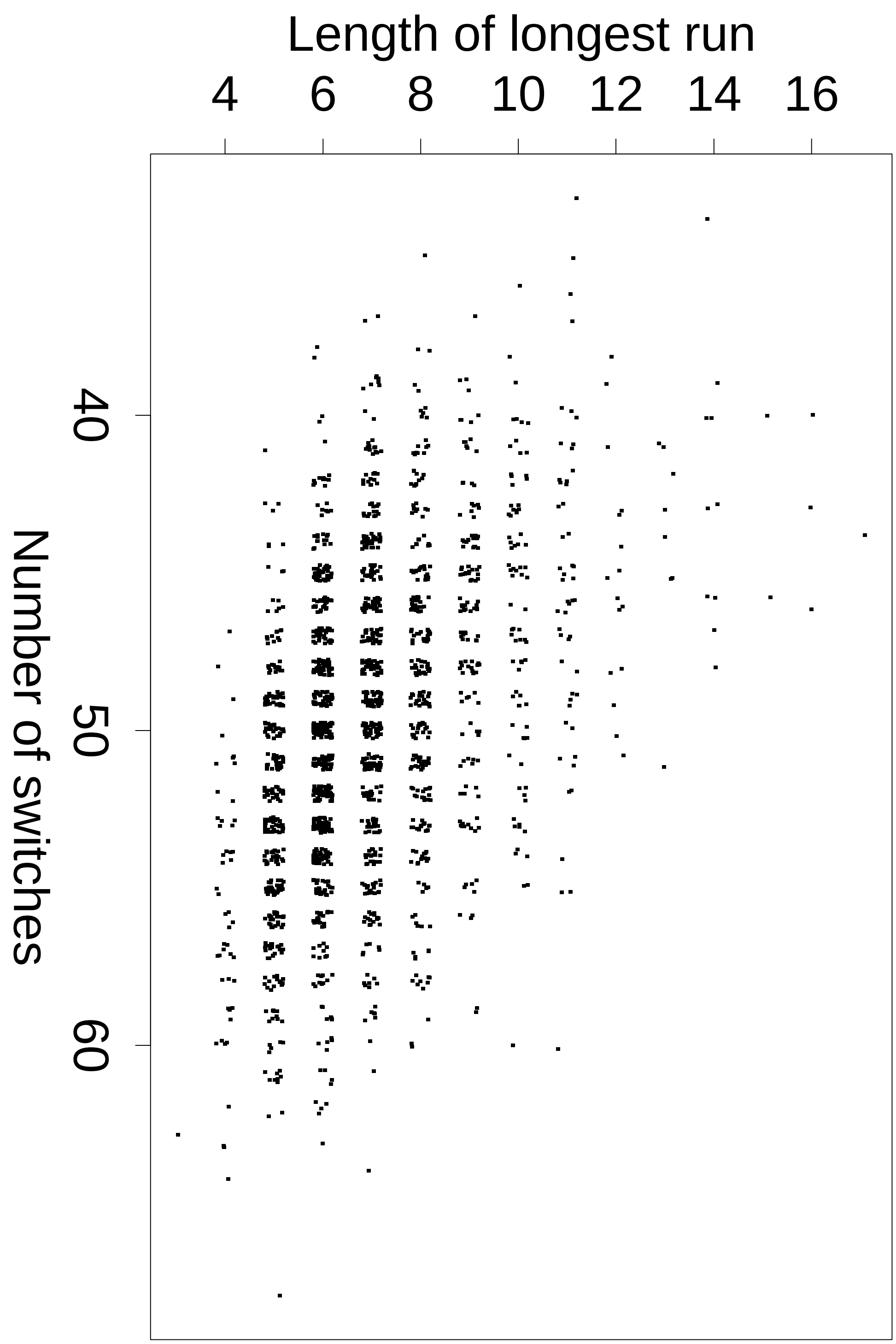
One could distinguish between real and fake sequences using a formal rule based on the longest run length, but we find that we can make the distinction more effectively based on a visual inspection of the sequences, which implicitly takes into account much more information.

We picked out the real sequence using our experience and knowledge of coin flips. How can this reasoning be formalized? For each of the two sequences on the blackboards, we count the number of runs (sequences of 0's and 1's) and the length of the longest run. We then hand out copies of the following figure which shows the *probability distribution* of these two statistics, as simulated from 2000 independent computer simulations of 100 coin flips. The students are instructed to circle on the scatter plot the locations of the values for the sequences on the blackboard. Most of the times we have used this example in class, the sequence of real coin flips is near the center of the scatter plot, and the sequence of fake coin flips has too many runs and too short a longest run, compared to this distribution.

In addition to its "magic trick" aspect, this demonstration appeals to students because it dramatically illustrates an important point for the interpretation of data: seemingly surprising patterns (long sequences of heads or tails) can occur entirely at random, with no external cause. Long runs in real coin-flip data surprise students because they expect that any part of a random sequence will itself look "random"—that is, typical of the whole. This can be the starting point of a discussion of the general phenomenon that small samples can be unrepresentative of a population. Familiar examples include biological data (for example, a family can have several boys or girls in a row) and sports (a basketball player can have "hot" and "cold" streaks that are consistent with random fluctuation).

Students often have difficulty thinking about summary statistics as random variables with probability distributions. This demonstration, which also alerts students to misconceptions about randomness, motivates the concept of the sampling distribution. We can also use this demo when we teach hypothesis testing. We task of identifying the real and fake sequences of coin flips was certainly a hypothesis test, and the question arises of how to quantify it. We do not go further with that example but rather use it as a general motivation for introducing the theory of hypothesis testing, which we introduce in the context of data examples we have already discussed.

6) Surprise Twist Demo: A punch line helps students remember key concepts: the double take of a surprise ending gets students thinking and talking about why things didn't turn out as expected. It is well known among statis-


Length of longest run
4
6
8
10
12
14
16
Number of switches
40
50
60

ticians that when you take a "haphazard" sample without using any formal probability sampling, you are likely to oversample the more accessible units. As an example, we consider the issue of confidence intervals and biased sampling, and we have found students to respond well to the following demonstration based on estimating the weight of a collection of objects.

We pass around the room a small digital kitchen scale along with a bag which, we (truthfully) tell the class, we filled ahead of time with 100 wrapped candies of different shapes and sizes (for example, 20 full-sized candy bars and 80 assorted small candies). We divide the class into pairs and tell each pair of students to estimate the total weight of the candies in the bag by first selecting a "representative or random sample" of five candies out of the bag, then weighing the sample and multiplying by 20 to estimate the total weight. We ask each pair of students to write down their measurement and estimate silently (so as not to influence the other students), then put their sample back in the bag, shake up the bag, and pass to the next pair.

This demonstration takes about two minutes to explain, and then it proceeds while the lecture takes place, thus giving all the students the opportunity to participate without taking away lecture time. While this is happening, we set up the blackboard for data collection, and we lecture on sampling distributions, its expectation and variance. As usual, we have the students work in pairs so they focus more consciously on the task.

Once all the pairs of students have weighed their candies and estimated the total weight; we go around the classroom and have the students shout out their estimates and their names. We write all of these weights on the blackboard and then display them as a histogram. While we draw the histogram, we pass the bag of candy around the classroom for consumption.

This histogram illustrates the sampling distribution of the estimated weights. We then get another student to weigh the entire bag and state the true total weight. It is invariably lower than most or even all of the sample-based estimates, and this shocks the students. Why did this happen? The students realize that the larger candies are more accessible (and also are more likely to remain on the top of the bag after it has been shaken). Even though they tried to get a representative or random sample, they could not help oversampling the large candies. Sometimes to increase the impact, we come to class with a sealed envelope, which we hand to the student who weighs the whole bag of candy. We ask him to open and read aloud the message inside the envelope after having weighed the bag of candies. The note says, "Your estimates are too high!"

5 Handouts

We find that handouts where students work in pairs or small groups in class to complete the questions on the handout can be very useful in calibrating and building students confidence and understanding of a topic. We provide examples of three types of handouts: problem sets, worksheets, and drills.

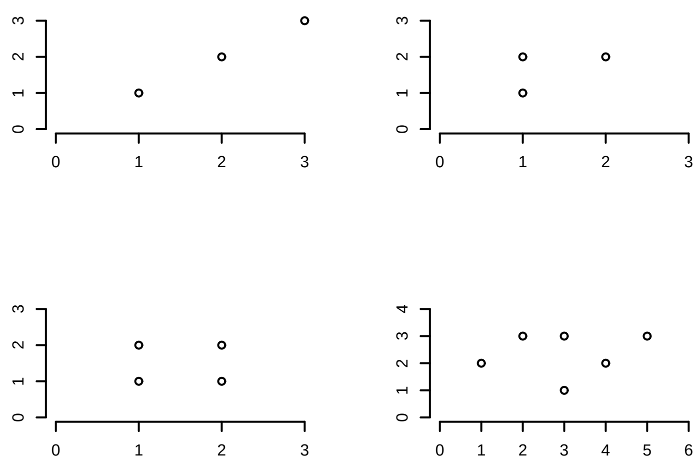
7) Group Problem Sets: When we cover probability, we typically give students a graduated set of probability problems to solve in groups and present in class. We begin by introducing a new concept via an example that the class works on with the instructor. After completing the example, the students split into groups of two to four to work on problems that further develop the material just introduced. Typically we set a time limit for the groups to work on the problems, say 10 to 15 minutes, and we circulate among the students to keep them on track and offer advice. When many groups encounter the same point of confusion, we write hints or points of clarification on the board. In the last part of the class period, we ask each group to write up a solution to one problem on the board. When there are more groups than problems, we sometimes have more than one group write a solution.

The groups often work at different speeds; some will not have completed all the problems in the allotted time, and others will finish quickly. We bring extra problems for the rapid learners, and those still working on the first set of problems are asked to write up a problem that they have finished. At the end of the "lecture" period, we bring everyone together to review the solutions, make corrections if needed, tie up loose ends, and draw connections to the big picture. This kind of group problem-solving also works well when each group works on a different problem.

With group work we can cover a lot of basic material quickly. A typical handout includes problems that illustrate the main concepts and just-for-fun, difficult problems about sleazy gambling joints, winning the lottery twice, and random cuts of spaghetti. For the well known birthday problem, we go around the class looking for same-birthday pairs. We write students' birthdays on the board and count the matches. Once when we did this with a class of 60 students, we found we had twins in the class (we thought they were just brothers until then), and another student shared the twins' birthday! All together we had five matches. After we talked about the assumptions we made in computing the probability of a same-birthday pair, we computed the expected number of pairs, and found it to be $60 \cdot 59/(2 \cdot 365) = 4.8$, another surprise for the students.

8) Concept Worksheet: We have found that students often have difficulty with the simple regression line. To get them started, we use simple scatter plots to introduce straight-line fitting. We bring to class a handout of four scatter plots, divide the students into pairs, and ask the each pair to draw the best-fitting line for predicting y from x for the data on each plot. We don't go into the exact definition of a best fit yet—we just tell them that we want to predict y from x and let them loose on the problem.

Once they have drawn their lines, we have them compute their predictions for each y_i and the sum of squared errors. Then we pass out red pencils and ask each pair to mark a red $\times$ at the average y-value for each observed x. Then they find the best-fitting line to the crosses, in some cases these lines are quite different, which provides a good lead-in to the method of least squares. We also have them compute the sum of squared errors for predicting y_i from the red


0 1 2 3
0 1 2 3
0 1 2 3
0 1 2 3
0 1 2 3
0 1 2 3
0 1 2 3 4
0 1 2 3 4 5 6

line.

For a couple of the scatter plots, the placement of the points suggests the wrong line to the naive line-fitter. The students are surprised to find the red line outperforms their line in these cases, and at this point we begin the discussion of minimizing square error in the y-direction.

9) Current events Worksheet: An important theme in an introductory statistics course is the connection between statistics and the outside world. We describe here an assignment that we have found useful in getting students to learn how to gather and process information presented in the newspaper articles and scientific reports they read. Students work through prepared instructional packets, where each packet contains a newspaper article that reports on a scientific study or statistical analysis, excerpts from the original report on which the article was based, and a worksheet with guidelines for summarizing the reported study and a series of questions.

Through this type of assignment, students connect the statistics that they learn in the classroom to current events, and they learn how to think critically about the information found in the newspaper. An important component of this assignment, which sets it apart from other projects that use current newspaper articles, is the inclusion of the original source in the analysis. We have found that students are better able to evaluate the merits of the study and the quality of the news reporting when given excerpts from the original reports, even when these reports are quite technical.

An standard worksheet for a news story appears on the next page. Students work in pairs or groups of 3-4 to answer the questions. We sometimes use worksheets that are more tailored to the particular news article, and we sometimes, ask them to write answers to these questions on the board for class discussion.

News Story Summary Worksheet

What is the source of newspaper article? wire-service, in-depth article, feature story, other
Kind of report: medical/technical journal, press release, book, other

Provide a one-sentence summary of the newspaper article.

According to the original report, what was the objective of the study?

What type of study was conducted?
(a) *randomized experiment*: treatments under the control of the experimenter and assigned randomly
(b) *nonrandomized experiment*: treatments under the control of the experimenter and assigned non-randomly
(c) *observational study*: treatments not under the control of the experimenter (for example, the patients choose whether to smoke)
(d) *sample survey*
(e) *model-based analysis* (for example, an estimate of the economic effects of immigration)
(f) *meta-analysis*: an examination of several earlier studies.

Who are the subjects/participants?

What population do they represent?

What is the main outcome measurement?

Name some of the control variables.

Describe the nonresponse.

What were the stated conclusions?

Name three difficulties of generalizing to the real world or other problems of the study.

How do the results relate to the rest of the scientific literature?

How accurately did the newspaper article summarize the report?

10) Drill sheets: A drill sheet is a structured series of short problems that students should be able to solve almost automatically once they have mastered the relevant course material. In combination with lecture we often do one problem first with the entire class. Or alternatively, students are first given the basic problem and work in pairs on it. Then the class regroups to go over a solution that the instructor adapts from one of the students papers. The drill sheet is then handed out and students quickly complete the problems, we ask students to write solutions on the board as they progress through the set of problems. The class reconvenes to discuss the open ended questions.

6 Drawbacks

DO: With your same partner, now list three potential problems with classroom activities, including any cautionary notes as to what may go wrong with the activity that you have just described.

Many of the reasons given above for using classroom activities have counterpart arguments for why an activity can be troublesome. For example, rather than clearing up a common misconception, the activity may be so confusing that students transfer their frustration with the activity to the concept. Or, a punch line that is interrupted by the bell, may lose all of its impact in the retelling at the the next class meeting. Below are a few potential pitfalls with classroom activities and ideas for how to avoid them:

11) Time: Activities take up more class time than the concept warrants.

We work hard to streamline our activities to take up no more than 5 to 10 minutes, and depending on the activity, we often continue teaching while activity is being carried out. With new activities, we test them out first before using them in class with an eye to reducing the activity to the essential bits.

12) Complex: Some activities are too confusing or have too many steps to carry out to get to punch line.

We picked out an activity from a teaching journal that involved simulating a capture-recapture sampling method. It involved a game board and a dozen or more game pieces, all made from paper, that we discovered tended to fly off the board in class or stick to students' fingers. After playing the game for 10 minutes, the students still had not reached the point where they could see how the sampling method worked. This activity took too much time, was too complex, and the point was lost on the students because of it. In this case, I jettisoned the activity rather than attempt to fix it and replaced it with a simple worksheet involving cluster sampling.

13) Muddled: The point of the activity is not clear, and therefore the take-away message is lost.

See the above example.

14) Poor Practice: Due to the limitations of carrying out an activity in the classroom, it may be difficult to avoid activities that demonstrate poor practice of statistics.

Examples of such transgressions include: using the students in the classroom as a simple random sample, when they are in fact a sample of convenience; using data from a population as a sample; or acknowledging that a statistical method does not quite apply to this case, but using it anyway to demonstrate the technique. We try to stick by the old adage, "practice what you preach" and not use the activity if it does not represent sound statistical practice.

15) Pointless or Unreliable: An activity seems unrelated to real statistics, or a demo is a bust.

We have an example of the latter that occurred with a variant of the coin-flipping demo. In this alternative version we divide the students into four groups, having two groups create sequences of real coin flips and two groups make fake sequences. The goal was to involve more students, but the result was to create a much harder problem for the instructor to discern the real from the fake coin flips—it's a matter of picking the one correct answer out of six possibilities (4 choose 2), rather than one out of two—and we did occasionally picked it wrong. For maximum dramatic impact, we recommend just two groups so that it is easy to make the correct identification quickly.

TIME Including activities and demos in your lectures takes time: time to develop, prepare, maintain, streamline, and carry out. We end with a few additional thoughts on time-saving techniques.

- **Preparation** Activities and demos take time to prepare, but adequate preparation pays off in terms of saving precious class time. We often anticipate the outcomes of an activity and come to class prepared with for example pre-computed standard errors and p-values to avoid wasting time on these calculations.

- **Maintenance** Think of the demos as living things that we continually update, and add twists to. To help accomplish this, consider keeping a diary where you jot down after class a re-cap of what worked and what didn't.

- **Class-time management** As instructors, we can summarize, explain, and cover solutions more succinctly than our students. We take advantage of this by having students write solutions on the board while others are still working on problems, having the instructor present solutions with assistance from the students, and continuing instruction in parallel with a classroom activity.