

## **Thursday/Friday, August 6/7: Intro; Graphing Categorical Variables**

The Statistical Process:

1. Formulating Questions
2. Collecting Data
3. Analyzing Data
4. Interpreting Results

In the 2008 playoffs, NBA basketball star LeBron James of the Cleveland Cavaliers was criticized in the media for not performing up to his usual standards.

**1. Question:** Was this a fair criticism or was the media being too harsh?

**2. Collecting Data:** For this investigation, we will look at three-point shooting percentage. According to [www.basketball-reference.com](http://www.basketball-reference.com), here are LeBron's performances:

Regular Season: Made 113 of 359 = 31.5%

Playoffs: Made 18 of 70 = 25.7%

**3. Analyzing Data:** Data analysis should always start with graphs.

This type of data is called categorical because the outcomes of each shot fall into categories (made, missed). Other types of variables such as the distance a golf ball is hit or the time it takes to swim 50m are called numerical variables since the outcomes are numerical (e.g. 250 yards, 35.2 seconds).

There are 3 different ways to graph categorical data:

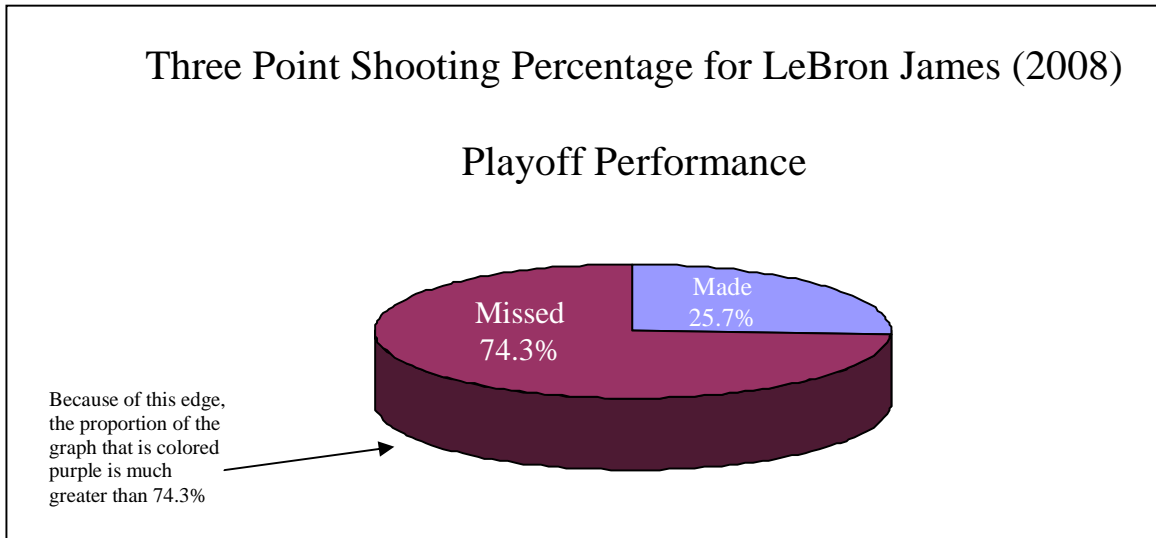
1. pie charts:

## 2. segmented bar charts

## 3. comparative bar charts

Note: When making comparative bar charts, it is almost always better to use relative frequencies (percents) rather than frequencies (counts). This is especially true when the number of observations is very different like in the regular season vs. the playoffs.

Note: All of these graphs must follow the “Area Principle” which says that the area representing each category should be proportional to the number of observations in each category. A common violation of this principle occurs with 3-D graphs such as the following:



Note: At the end of the packet there are instructions for how to make each of these graphs using Excel.

**HW #1: Unit 1, Problems 1, 2**

## **Monday, August 10: Modeling Athletic Performance**

What is the difference between *Performance* and *Ability*?

Example: Shooting Free Throws

Model:  $Performance = Ability + Random\ Chance$

- *Performance* is what an athlete actually does in an athletic contest or series of contests. This is what is typically referred to as an athlete's statistics.
- *Ability* is a concept that imagines what an athlete would do if given an unlimited number of opportunities at a specific point in the athlete's career.
- *Random Chance* is an acknowledgement that there is variability in athletic *Performance*. In other words, even in identical conditions, a particular athlete's *Performance* will not always be the same.

Example: Flipping a coin

What should happen if we flip the coin more times?

Back to LeBron: If we assume his *Ability* to make 3-pointers is 31.5%, what kinds of *Performances* should we expect due to *Random Chance*?

**HW #2: Unit 1, problems 3, 5**

**Tuesday/Wednesday, August 11/12: Simulations with Technology**

**HW #3: Unit 1, Problems 4, 6**

**Thursday/Friday, August 13/14: Another Example: Eric Byrnes, 2007**

In 2007, Eric Byrnes of the Arizona Diamondbacks finished 11<sup>th</sup> in the MVP balloting as he helped lead the D-backs to the playoffs. During the season, his batting average was 0.286 (179/626). However, in the playoffs he only had 6 hits in 29 at-bats (0.207). Did he choke?

Data from: [www.baseball-reference.com](http://www.baseball-reference.com).

Discuss Projects/Rubric/etc.

**HW #4: Unit 1, Problems 7, 8**

**Monday, August 17: Computer Lab!**

Go to the computer lab to research data for Project

**HW #5: Unit 1, Problem 9; Work on Projects**

**Tuesday/Wednesday, August 18/19: Summary of Unit 1**

Connections: What's coming up

Stats 101: How these ideas are taught in a traditional course

**HW: Review for Test; Work on Projects**

**Thursday/Friday, August 20/21: Projects Due / Test on Unit 1**