From the quiz:

Suppose that simple random samples are repeatedly taken from a population, and for each sample a 95% confidence interval for a proportion is calculated. Which of the following statements is FALSE?

A. For each interval, the probability that the true population proportion is between the upper and lower limit of the confidence interval is 95%.

B. The resulting intervals contains the true population proportion approximately 95% of the time.

What is the difference between these two answers?

An applet to illustrate confidence intervals for a proportion:
https://www.stat.tamu.edu/~west/ph/propci.html

An applet to illustrate confidence intervals for a proportion: https://www.stat.tamu.edu/~west/ph/propci.html

Confidence intervals for a proportion

The applet below allows one to visually investigate confidence intervals for a proportion. Specify the sample size **n** and the true proportion **p**. When you click the **Simulate** button, 100 separate samples of size **n** will be selected from a population with a proportion of successes equal to **p**. For each of the 100 samples, a 95% confidence interval and a 99% confidence interval are displayed in the plot to the right. Each of these intervals is computed based using the standard normal approximation. If an interval does not contain the true proportion, it is displayed in red. Note that the 99% confidence interval is always wider than the 95% confidence interval. Additional simulations can be carried out by clicking the **Simulate** button multiple times. The cumulative number of times that each type of interval contains the true proportion is also tabled. Press the **Clear** button to clear existing results and start a new simulation. Things to try with the applet:

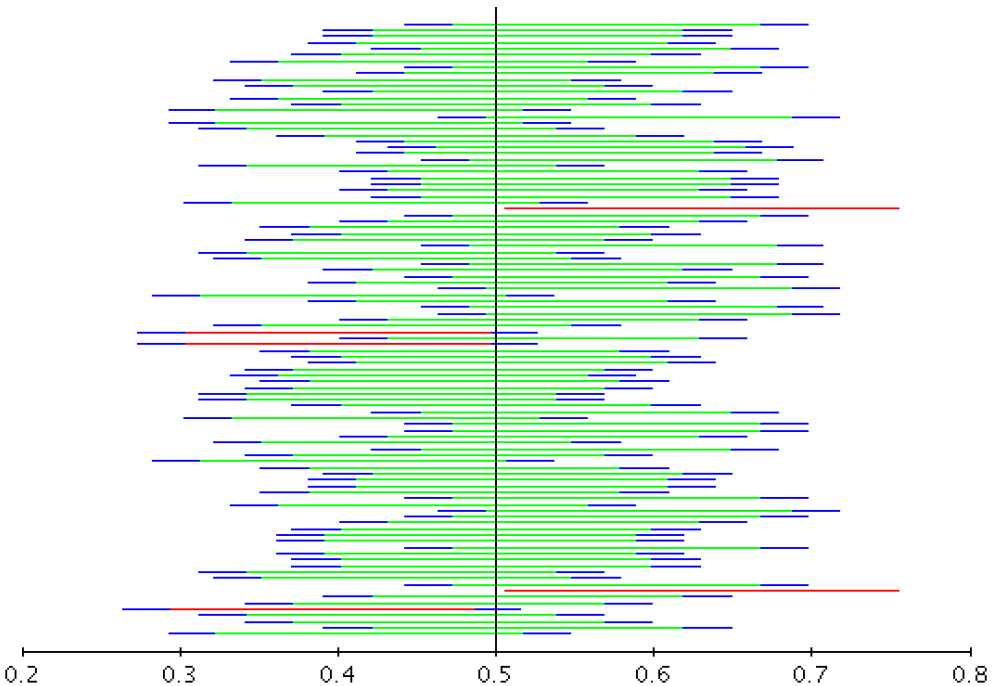
- Simulate at least 1000 intervals with **n** = 100 and **p** = 0.5. What proportion of the 95% confidence intervals contain 0.5? What proportion of the 99% confidence intervals contain 0.5? How does the typical width of these intervals compare to the **n** = 30 and **p** = 0.5 case above?
- Simulate at least 1000 intervals with **n** = 30 and **p** = 0.1. What proportion of the 95% confidence intervals contain 0.1? What proportion of the 99% confidence intervals contain 0.1?
- Simulate at least 1000 intervals with **n** = 30 and **p** = 0.5. What proportion of the 95% confidence intervals contain 0.5? What proportion of the 99% confidence intervals contain 0.5?

n: 100
p: 0.5
Simulate

Cumulative results:

	95% CI	99% CI
Contained p	95	98
Did not contain p	5	2
Prop. contained	0.95	0.98

Clear



The previous slide shows confidence intervals calculated from 100 simulated samples of data.
For each, a random sample of size 100 was taken from a population with p=0.5, the estimate of p was obtained for the sample, and both the 95% and 99% confidence intervals were calculated.

Results:

- 5/100 or 5% of the 95% confidence intervals did not include the true proportion (a.k.a. the population proportion, a.k.a. the parameter we are trying to estimate and capture in each confidence interval).

- 2/100 or 2% of the 99% confidence intervals did not include the true population (a.k.a. the population proportion, a.k.a. the parameter we are trying to estimate and capture in each confidence interval).

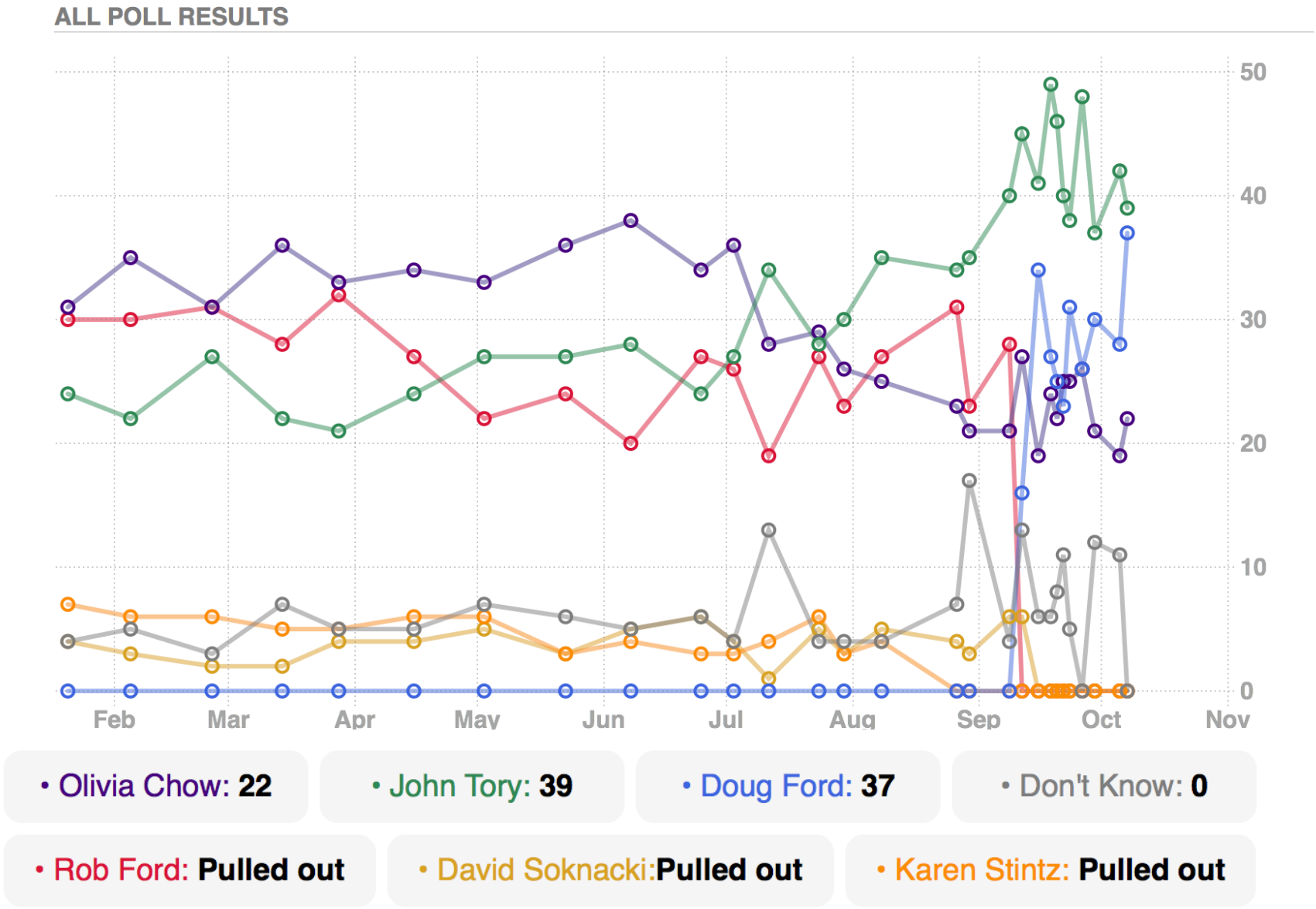
Interpreting Confidence Intervals:

- A confidence interval for a proportion is trying to capture the population proportion or theoretical world probability, *p* (our parameter).
- Although we don't know *p*, it's not random.
- The upper and lower limits of the confidence interval are calculated from the data. They vary with the sample data.
- A 95% CI is designed to capture *p* 95% of the time.
- Why not 100%? Sometimes we get unlucky, and get sample data that are unusual (that is, they have a small probability) and these unusual data result in an estimate of *p* that is unusual, that is, out in the tails of the sampling distribution of $\hat{p}$
- So 5% of the time, we will get a confidence interval that does not include *p* between its upper and lower limits.
- Unfortunately, in real life situations, for any particular confidence interval, we don't know if it's one of the 95% that captured *p*, or one of the 5% that didn't.

Interpretation of a confidence interval:

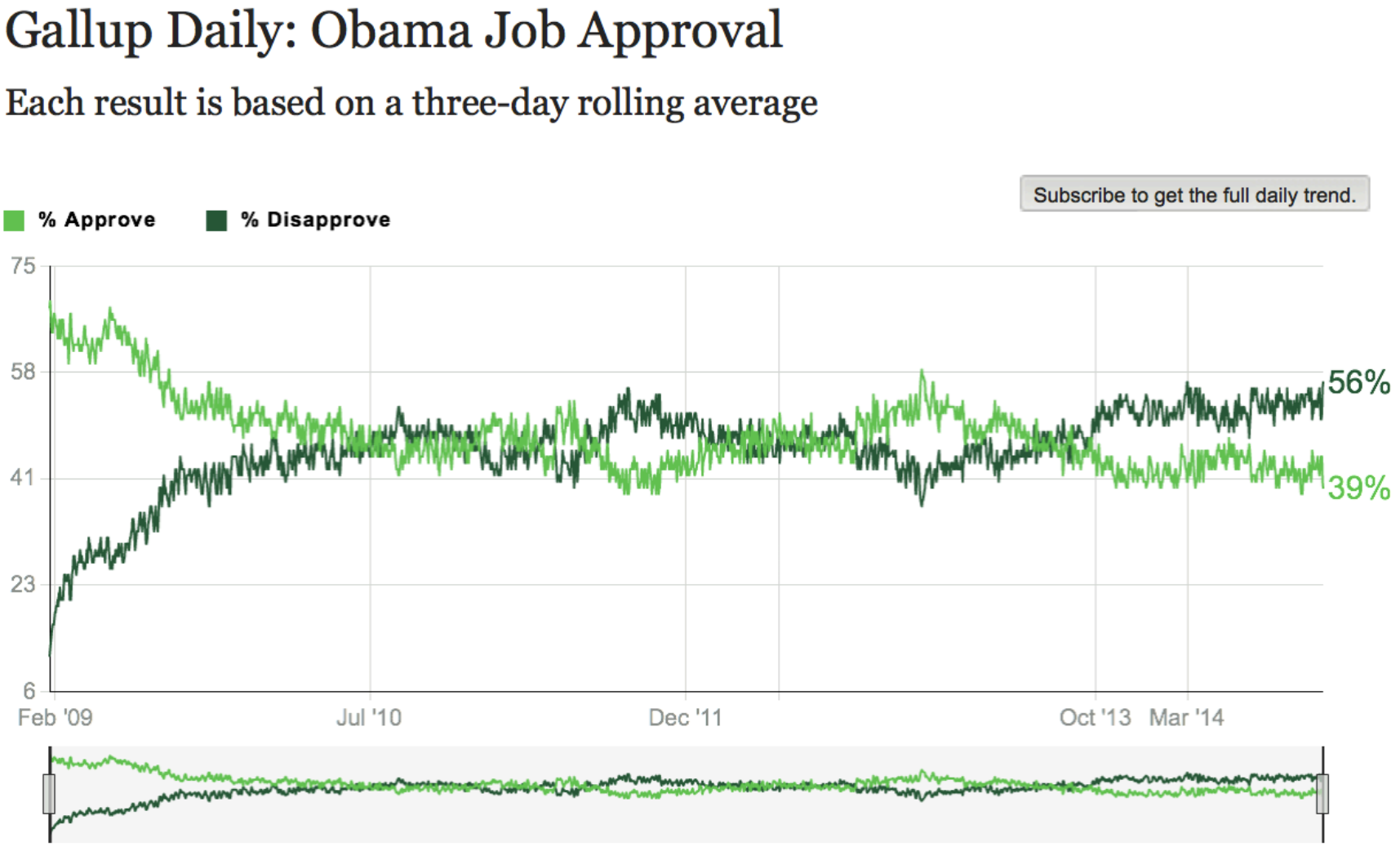
If we perform our data collection procedure (carry out an experiment, or collect a random sample from a population) a large number of times, and each time we use the data we collect to estimate something, and each time we (correctly) calculate a 95% confidence interval for what we're trying to estimate, 95% of the confidence intervals will include the true (or population) value of what is being estimated.

Confidence intervals for proportions often occur in the media as the results of polls with an estimate or a proportion given with its margin of error. Sometimes the media tracks an issues with regular polls.



How should you interpret daily tracking polls?

(from http://www.gallup.com/poll/113980/gallup-daily-obama-job-approval.aspx)



Gallup tracks daily the percentage of Americans who approve or disapprove of the job Barack Obama is doing as president. Daily results are based on telephone interviews with approximately 1,500 national adults; Margin of error is ±3 percentage points.

The Gallup daily tracking poll of the proportion of Americans who approve of the job Obama is doing:

- 1500 people, margin of error of 3%
- If the margin of error is for a 95% confidence interval:
  Margin of error = 1.96 * sqrt(.5*.5/1500) = 0.025 = 2.5%
- So let's assume Gallup rounded to 3% and, for each daily poll for have a confidence interval with margin of error = 3%.

The Gallup daily tracking poll of the proportion of Americans who approve of the job Obama is doing:

- 1500 people, margin of error of 3%
- If the margin of error is for a 95% confidence interval:
  Margin of error = 1.96 * sqrt(.5*.5/1500) = 0.025 = 2.5%
- So let's assume Gallup rounded to 3% and, for each daily poll for have a confidence interval with margin of error = 3%.

From January 25, 2009 to October 8, 2014, there were 2082 daily polls. The latest poll (October 8) had an approval rating of 44%, and a corresponding confidence interval of (41%, 47%).

The Gallup daily tracking poll of the proportion of Americans who approve of the job Obama is doing:

- 1500 people, margin of error of 3%
- If the margin of error is for a 95% confidence interval:
  Margin of error = 1.96 * sqrt(.5*.5/1500) = 0.025 = 2.5%
- So let's assume Gallup rounded to 3% and, for each daily poll for have a confidence interval with margin of error = 3%.

From January 25, 2009 to October 8, 2014, there were 2082 daily polls. The latest poll (October 8) had an approval rating of 44%, and a corresponding confidence interval of (41%, 47%).

- Do you think this CI captured the percentage of all Americans who, on the day the poll was taken, approved of the job Obama is doing?

- Do you think that all 2082 CIs captured the percentage of all Americans who, on the day each of the polls was taken, approved of the job Obama is doing?

A 99% confidence interval for a proportion is calculated from a single sample of data and is found to be (.4, .6). It can be interpreted as:

A. In 99% of samples, the estimated proportion for the sample will be between .4 and .6.

B. There is a 1% chance that the true (population) proportion is less than .4 or greater than .6.

C. 99% of all population values are within the interval from .4 to .6.

D. Both A. and B.

E. None of the above.

A 99% CI for a proportion is calculated as (.4, .6). It can be interpreted as:

To join this session, send a message using the keyword **ALISONGIBBS**

Text your message to **37607**

Submit responses at **PollEv.com/alisongibbs**

In 99% of samples, the estimated proportion will be between .4 and .6.	**A**
There is a 1% chance that the population proportion is less than .4 or greater than .6.	**B**
99% of all population values are within the interval from .4 to .6.	**C**
Both A and B	**D**
None of the above	**E**

Total Results: 0

From the quiz:

Suppose that simple random samples are repeatedly taken from a population, and for each sample a 95% confidence interval for a proportion is calculated. Which of the following statements is FALSE?

A. For each interval, the probability that the true population proportion is between the upper and lower limit of the confidence interval is 95%.

B. The resulting intervals contains the true population proportion approximately 95% of the time.

What is the difference between these two answers?

Another quiz question:

The probability that a 99% confidence for a proportion captures *p* is:

A. 0

B. 1

C. 0.99

D. 0.01

E. Either 0 or 1