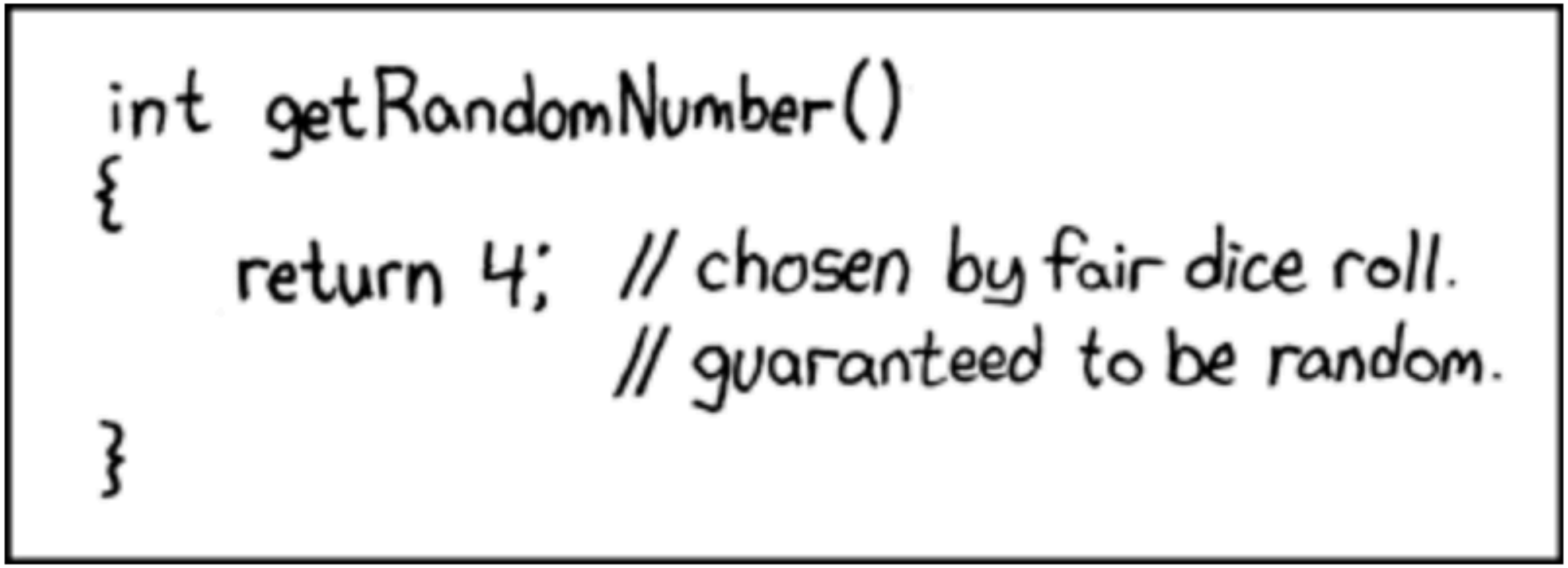
Stat 88: Probability and Statistics in Data Science



https://xkcd.com/221/

Lecture 2: 1/19/2024

Basics, Intersections

1.1, 1.2

Shobhana M. Stoyanov

Agenda

- Quick recap of definitions so far, and continue with terminology
- Idea of probability as a proportion, and assumptions for this
- Social media example from text (not equally likely outcomes)
- Exact calculations of probabilities vs bounds
- If we have time, talk about de Méré’s Paradox

Terminology

- **Experiment**: action that results in exactly one of several possible outcomes or results, and chance or randomness is involved - that is, each time we perform the action, the outcome will be different, and we don't know exactly which outcome will occur.

- An ***event*** is a collection of outcomes.

- A collection of all possible outcomes of an action is called a ***sample space*** or an ***outcome space***. Usually denoted by Ω (sometimes also by S).

- An event is always a subset of Ω. Suppose we call the event A, then we write this as $A \subset \Omega$. We denote the probability of A as $P(A)$.

Warm up: Go to pollev.com/shobhana to answer #1

1. If you have a well-shuffled deck of cards, and deal 1 card from the top, what is the chance of it being the queen of hearts? What is the chance that it is a queen (any suit)?

2. How did you do this? What were your assumptions?

3. Say we roll a die. What is Ω?

4. What is the chance that the die shows a multiple of 3? What were your assumptions?

Chance of a particular outcome

- We usually think of the chance of a particular outcome (roll a 6, coin lands heads etc) as the number of ways to get that outcome divided by the total possible number of outcomes.

$$\frac{\#\ of\ particular\ outcomes\ of\ interest}{total\ \#\ of\ outcomes\ possible}$$

- So if A is an event (subset of Ω), then $P(A) = \dfrac{\#(A)}{\#(\Omega)},\ A \subseteq \Omega$

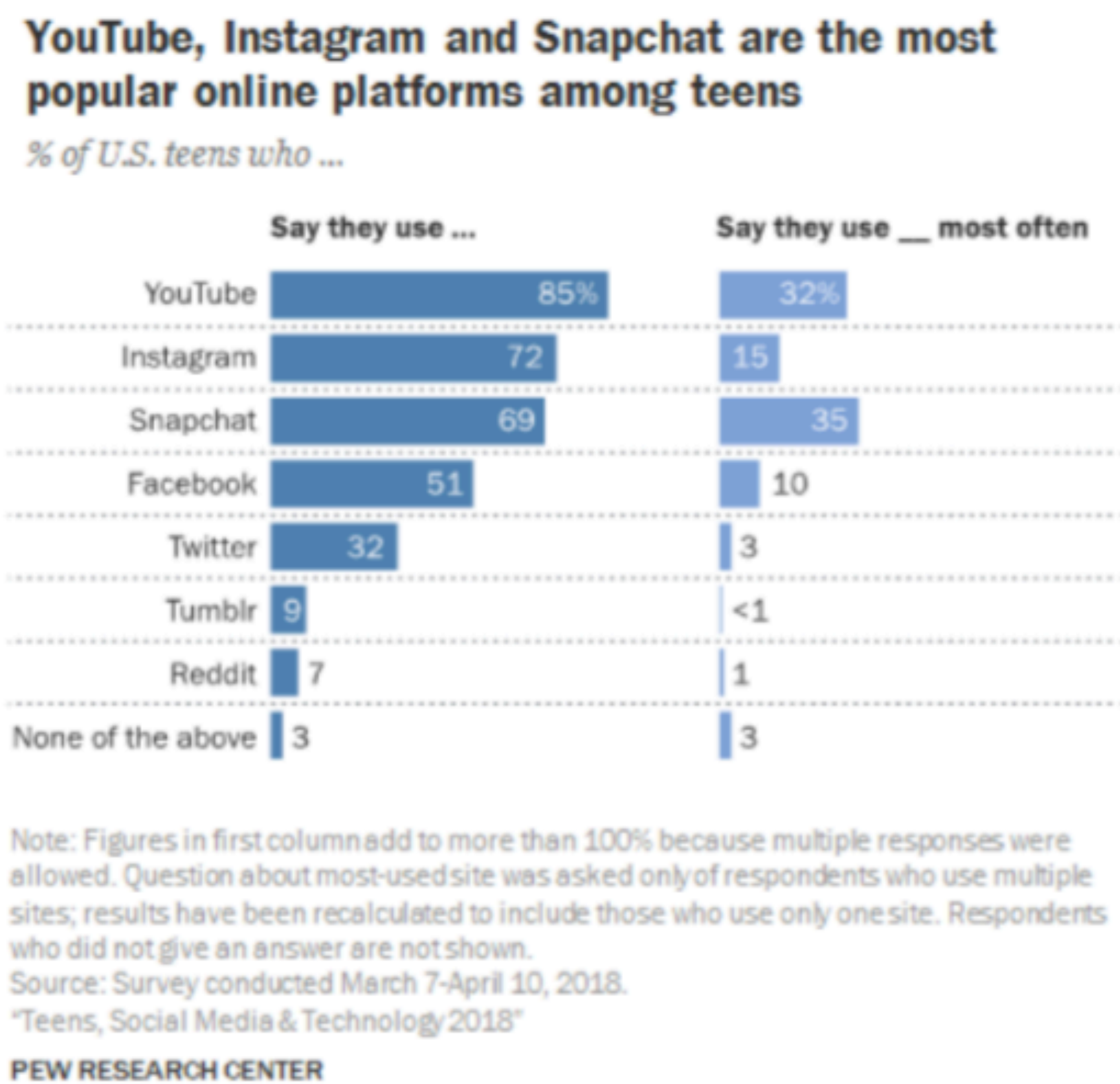
- If an experiment has a **finite** number of possible *equally likely* outcomes, then the probability of an event is the proportion of outcomes that are included in the event.

Cards

- If you deal 2 cards, what is the chance that at least *one* of them is a queen?

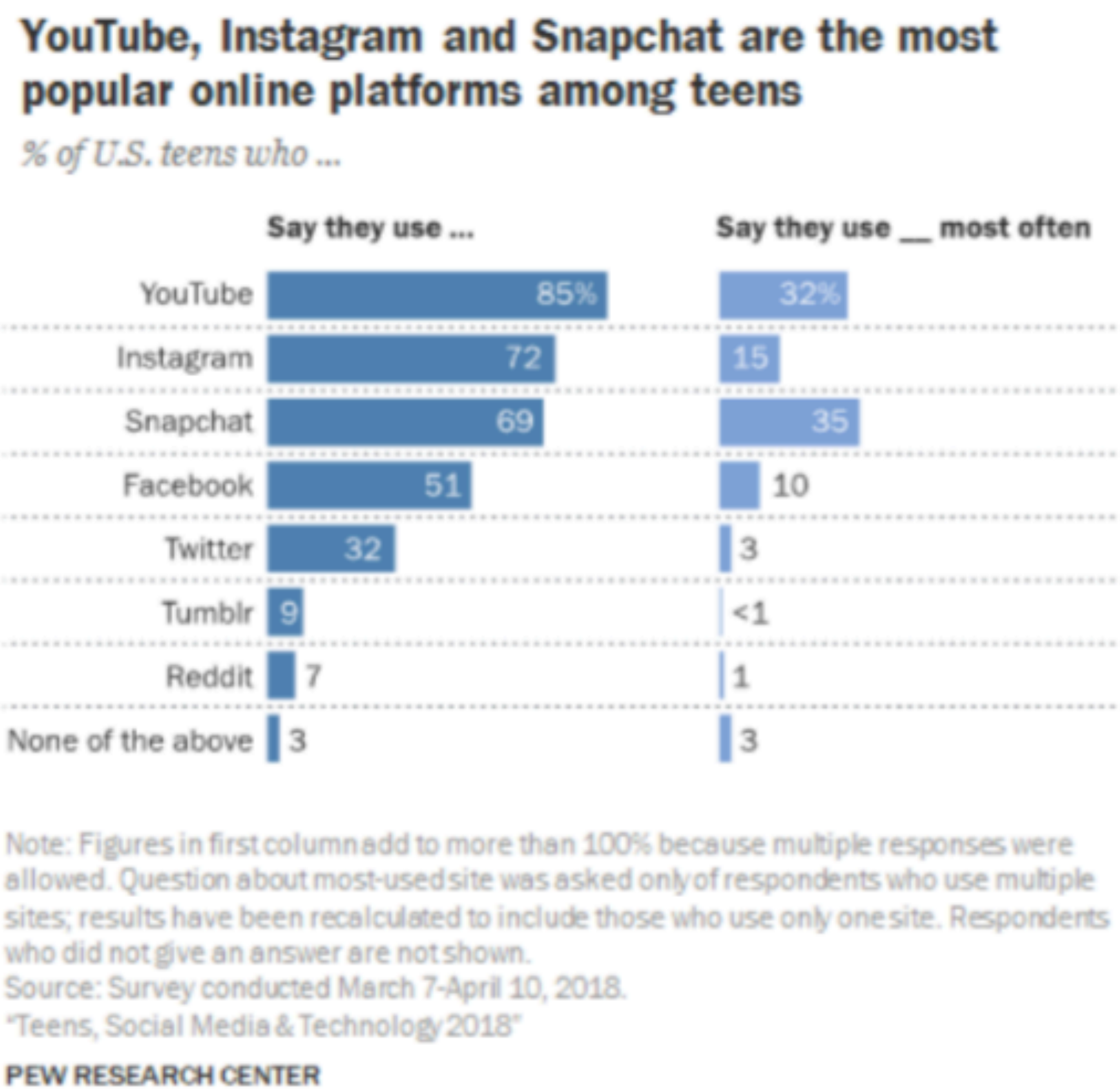
Not equally likely outcomes

- What if our assumptions of equally likely outcomes don't hold (as is often true in life, data is messier than nice examples).
- Here is a graphic from Pew Research displaying the results of a 2018 survey of social media use by US teens.



- What is the difference b/w 2 charts?
- Why do the % add up to more than 100 in the first graph?
- Second graph gives us a ***distribution*** of teens over the different categories

Not equally likely outcomes



1. What is the chance that a randomly picked teen uses FB most often?

2. What is the chance that a randomly picked teen did *not* use FB most often?

3. What is the chance that FB *or* Twitter was their favorite?

4. What is the chance that the teen used FB, just not most often?

5. Given that the teen used FB, what is the chance that they used it most often?

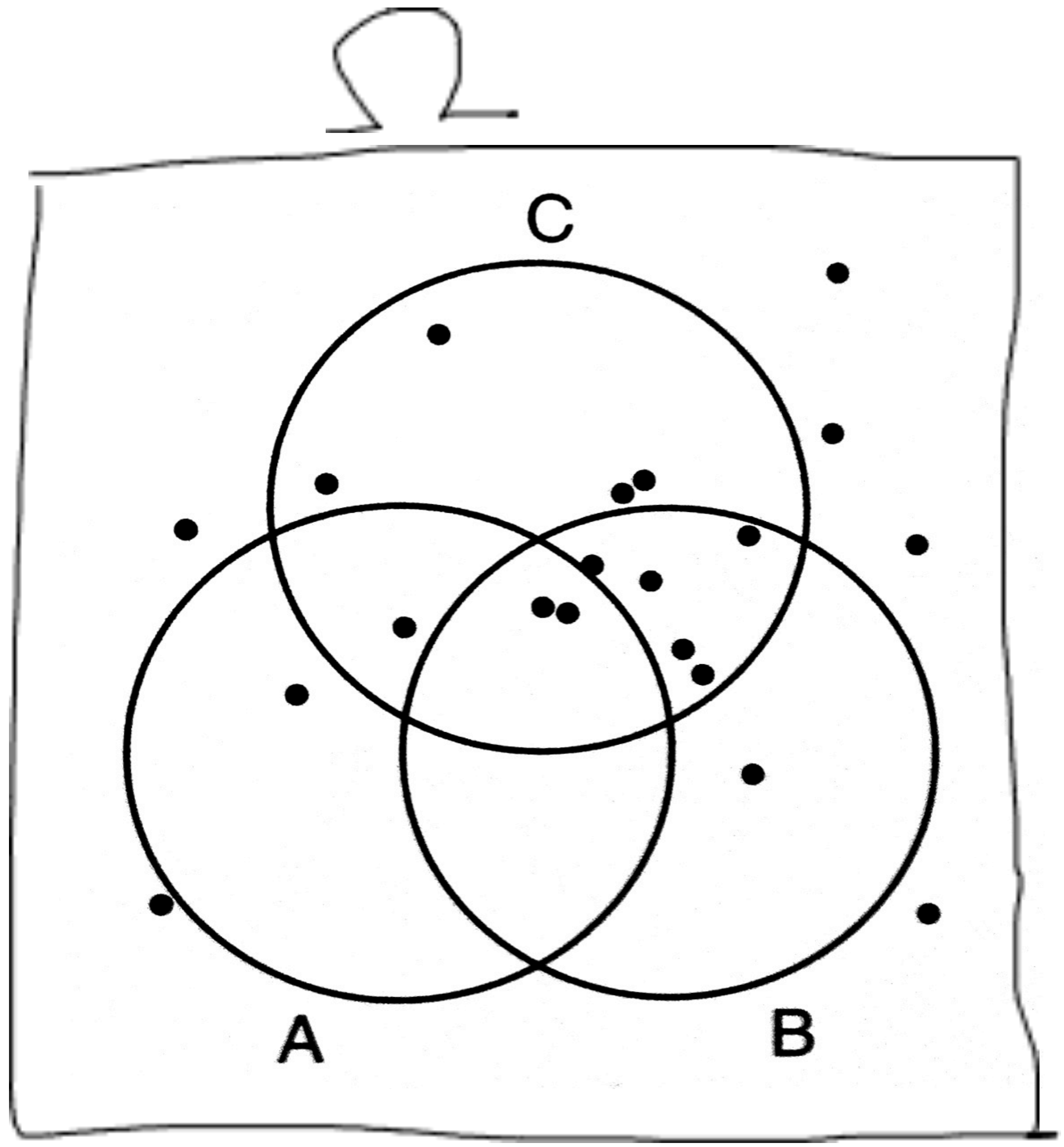
Terminology

- *Experiment*: action that results in exactly one of several possible outcomes or results, and chance or randomness is involved - that is, each time we perform the action, the outcome will be different, and we don't know exactly which outcome will occur.
- An *event* is a collection of outcomes.
- A collection of all possible outcomes of an action is called a ***sample space*** or an ***outcome space***. Usually denoted by Ω (sometimes also by *S*).
- An event is *always* a subset of Ω. Suppose we call the event A, then we write this as $A \subset \Omega$

- A *distribution* of the outcomes over some categories represents the proportion of outcomes in each category (each outcome appears in one and only one category)

- The *complement* of an event A is an event consisting of all the outcomes that are not in A. It is denoted by A^C and we have that $P(A^C) = 1 - P(A)$ (*Complement Rule)*

Venn Diagrams



- You worked on this picture in section today. This kind of drawing is called a ***Venn Diagram.*** The events are usually represented as roughly circular, though they don't have to be, and the outcome space as a rectangle containing it. Often drawing Venn diagrams and other pictures will make your problem more clear.

So far:

- If all the possible outcomes are *equally likely*, then each outcome has probability $1/n$, where $n = \#(\Omega)$
- Let $\mathrm{A} \subseteq \Omega$, $\quad P(A) = \dfrac{\#(A)}{\#(\Omega)}$
- Probabilities as proportions
- Sum of the probabilities of all the distinct outcomes should add to 1
- $0 \leq P(A) \leq 1$, $\mathrm{A} \subseteq \Omega$
- A *distribution* of the outcomes over different categories is when each outcome appears in one and only one category.
- What would happen if we get some information about the outcome or event whose probability we want to figure out?
- Our outcome space reduces (number of possible outcomes), incorporating that information, so we recompute the probability.

Conditional probability

- In the last question, we used the information that the teen used FB. We were told the teen used FB, and *then* asked to compute the chance that FB was their favorite.
- This is called the *conditional probability that the teen used Facebook most often, given that they used Facebook* and denoted by:

Conditional probability

- This probability we computed is called a ***conditional probability***. It puts a condition on the teen, and *changes* (restricts) the universe (the sample space) of the next outcome, a teen who likes FB best.
- To compute a conditional probability:
  - First restrict the set of all outcomes as well as the event to *only* the outcomes that *satisfy* the given **condition**
  - Then calculate proportions accordingly
- How do the probabilities in #1 and #5 compare?

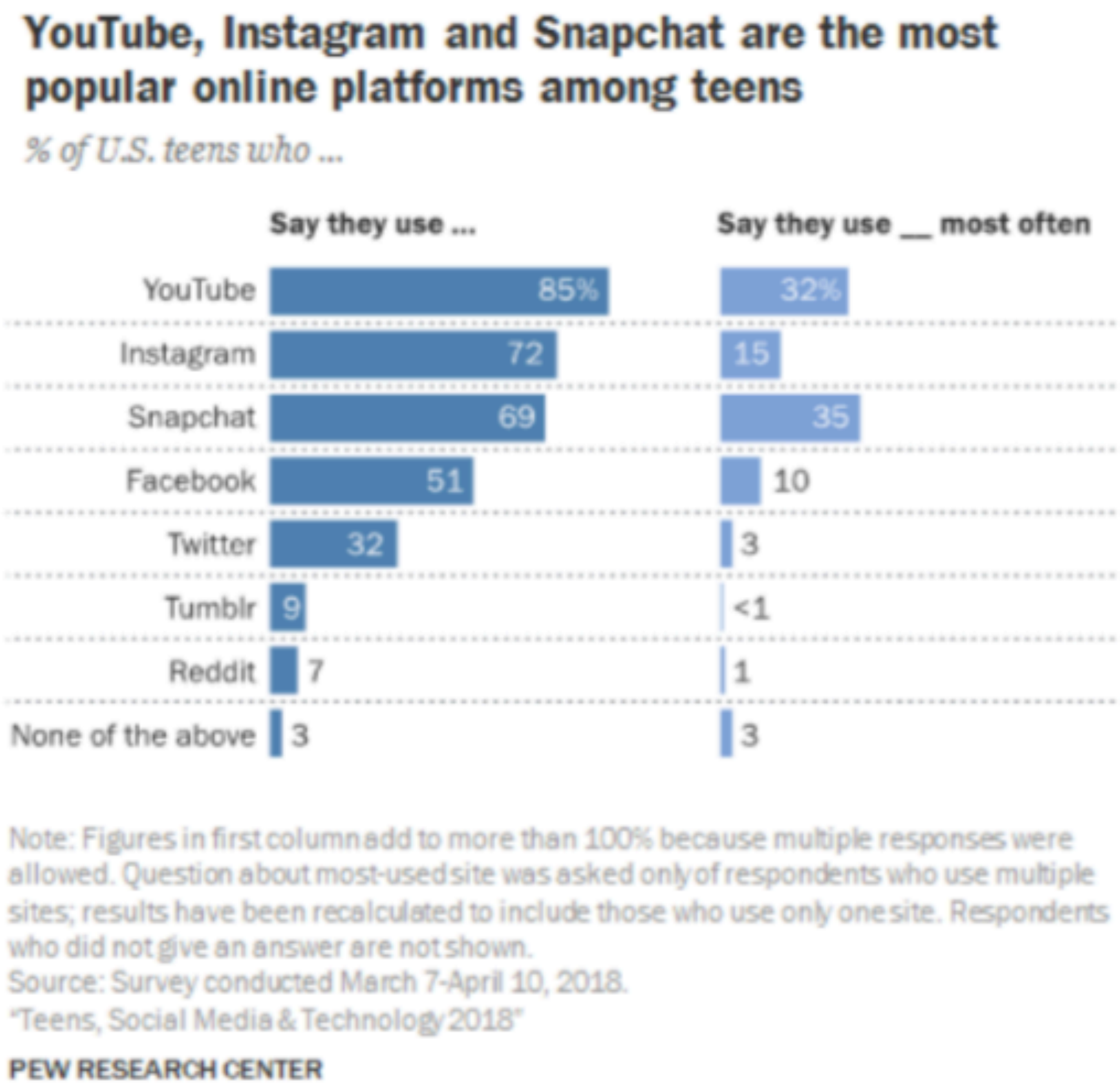
Example

- A six-sided fair die is rolled twice:

  - If the first roll lands on 1, what is the chance that the second roll lands on a number bigger than 2?

Exercise: Find the probability that the second number is greater than the twice the first number.

Section 1.2: Exact Calculations, or Bound?



Recall #3 about FB or Twitter. What was the answer? What can you say about the chance that a randomly selected teen used FB or Twitter (not necessarily most often)?

Example with bounds

- Let A be the event that you catch the bus to class instead of walking, P(A) = 70%
- Let B be the event that it rains, P(B) = 50%
- What is the chance of **at least** one of these two events happening?

- What is the chance of **both** of them happening?

Exercise: what about if we have 3 events?

- Let A be the event that you catch the bus to class instead of walking, P(A) = 70%
- Let B be the event that it rains, P(B) = 50%
- Let C be the event that you are on time to class, P(C) = 10%
- What is the chance of **at least** one of these three events happening?

- What is the chance of **all three** of them happening?

Rules that we used:

- If all the possible outcomes are *equally likely*, then each outcome has probability ***1/n***, where ***n*** = number of possible outcomes.
- If an event A contains k possible outcomes,
  then *P(A) = k/n*.

- Probabilities are between 0 and 1

- If two events A and B don't overlap, then the probability of A or B = P(A) + P(B) (since we can just add the number of outcomes in one and the other, and divide by the number of outcomes in Ω)

Rules of probability

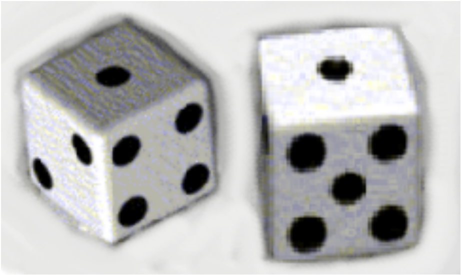
- Let's think about what rules we can lay down, based on what we have seen so far.

De Méré’s Paradox

- We can think about probability as a numerical measure of uncertainty, and we will define some basic principles for computing these numbers.
- These basic computational principles have been known for a long time, and in fact, gamblers thought about these ideas a lot. Then mathematicians investigated the principles.

- Famous problem: will the probability of **at least one six** in **four** throws of a die be equal to prob of **at least a double six** in 24 throws of a pair of dice.

- Note: single = die, plural = dice:



Origins of probability: de Méré's paradox

Questions that arose from gambling with dice.

Antoine Gombaud,
Chevalier de Méré

Blaise Pascal

Pierre de Fermat

The dice players
Georges de La Tour
(17th century)