

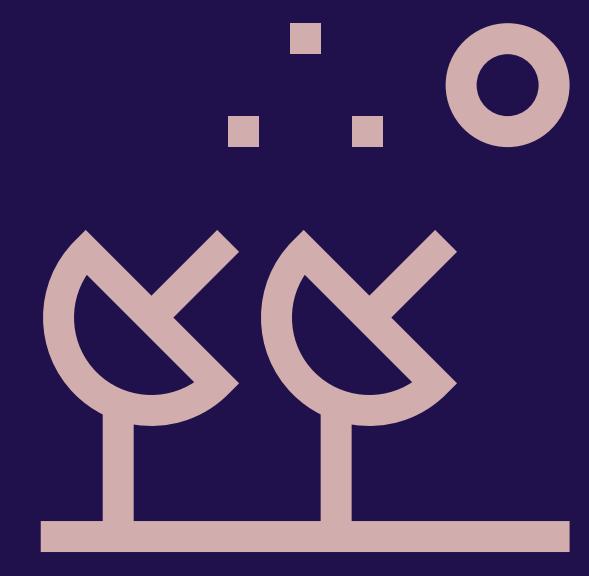
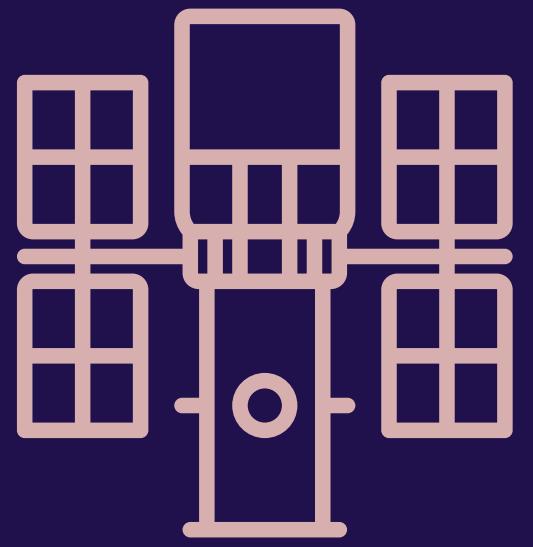
Bayesian Statistics (with M&Ms)

Daniela Huppenkothen

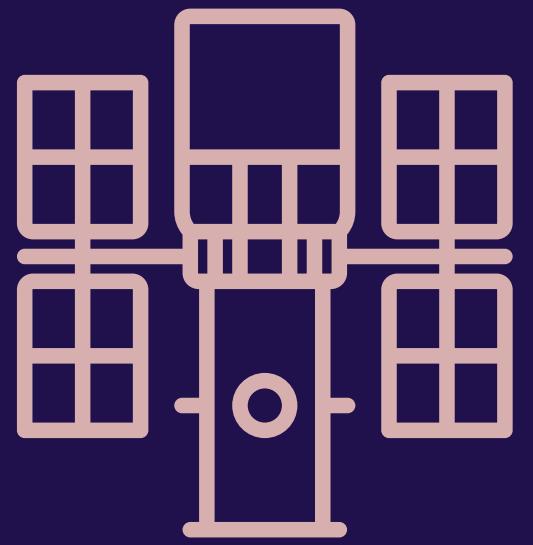
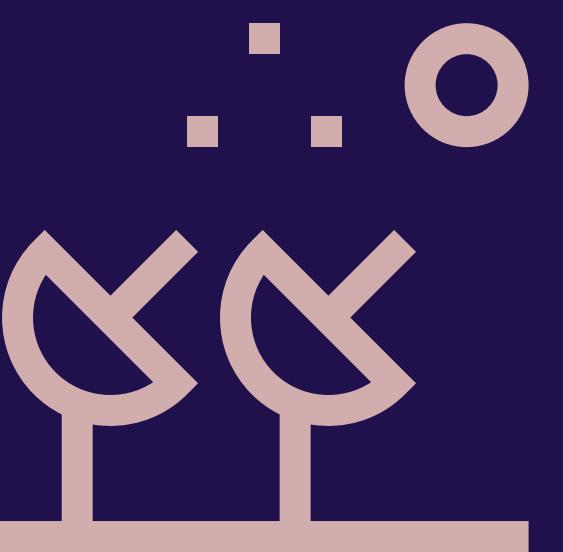
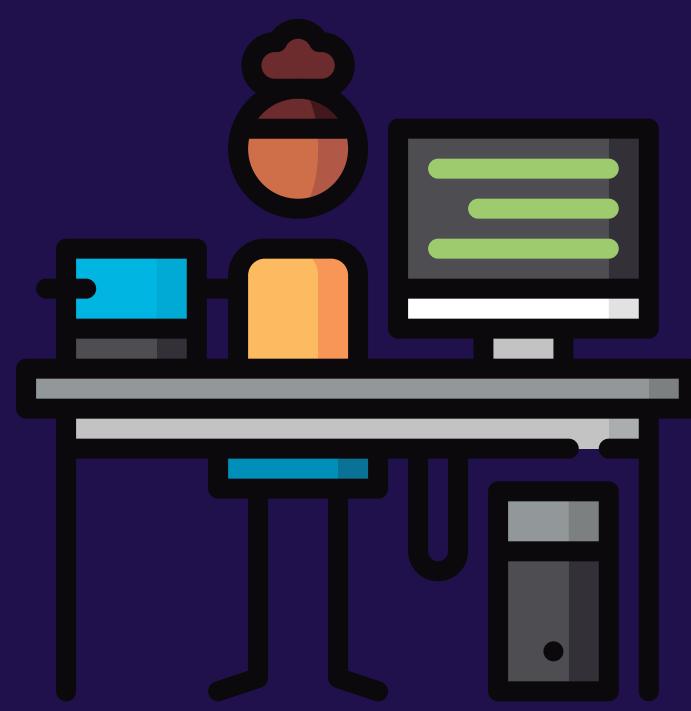
SRON Netherlands Institute for Space Research

All models are **wrong**,
but some are **useful**

– George Box



Statistics!



Part 1: Probabilities

What snacks will we have for coffee
break today?



What snacks will we have for coffee
break today?

p()

p()

Basic rule of probability (1)

$$0 \leq p(\text{cookie}) \leq 1$$

If $p(\text{🍪}) = 0.3$, what is $p(\text{🥐})$?

- a) 0.3
- b) 0.5
- c) 0.01
- d) 0.7

Join at

slido.com
#3643 573

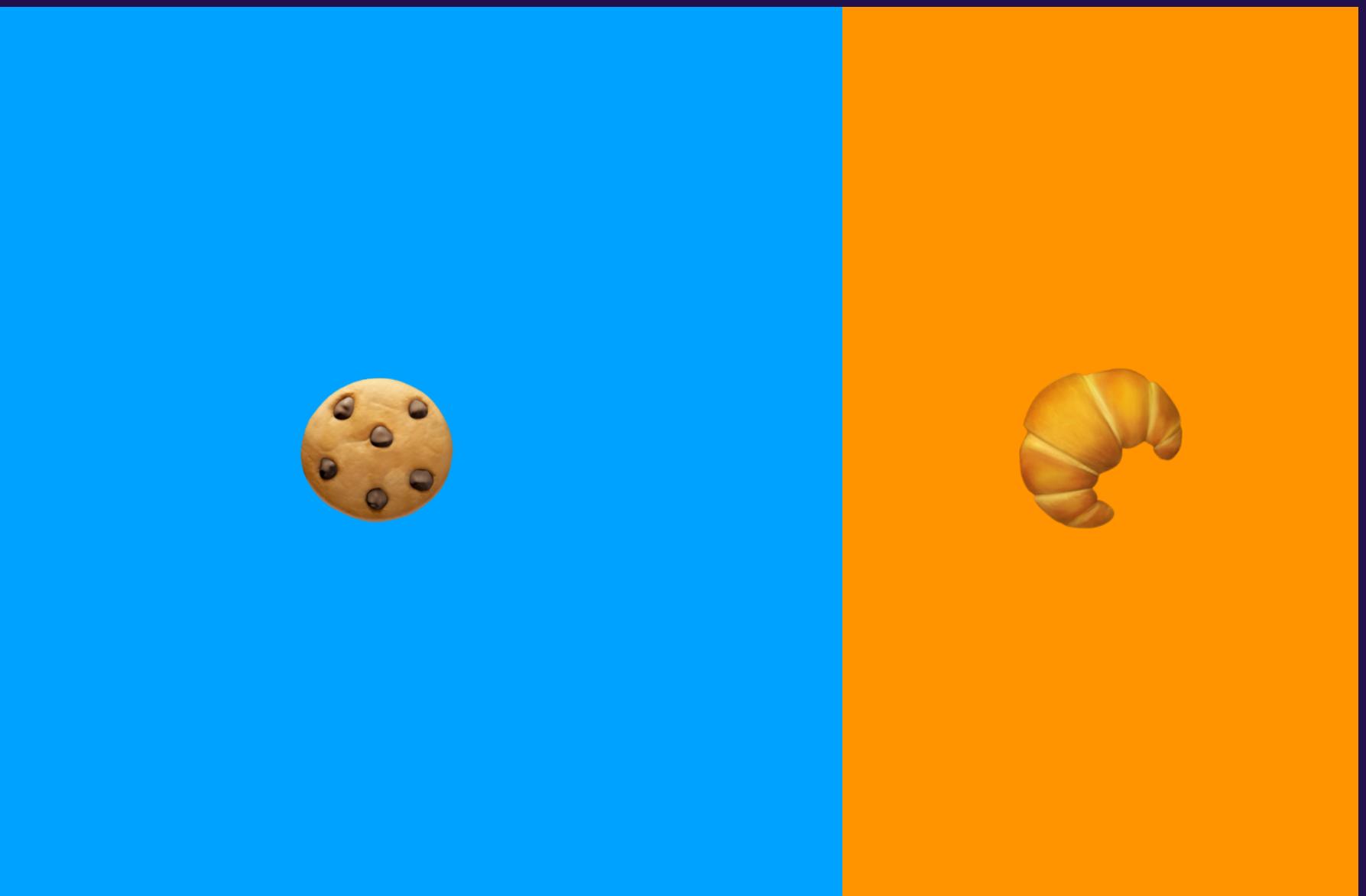


$$p(\cookie) + p(\croissant) = 1$$

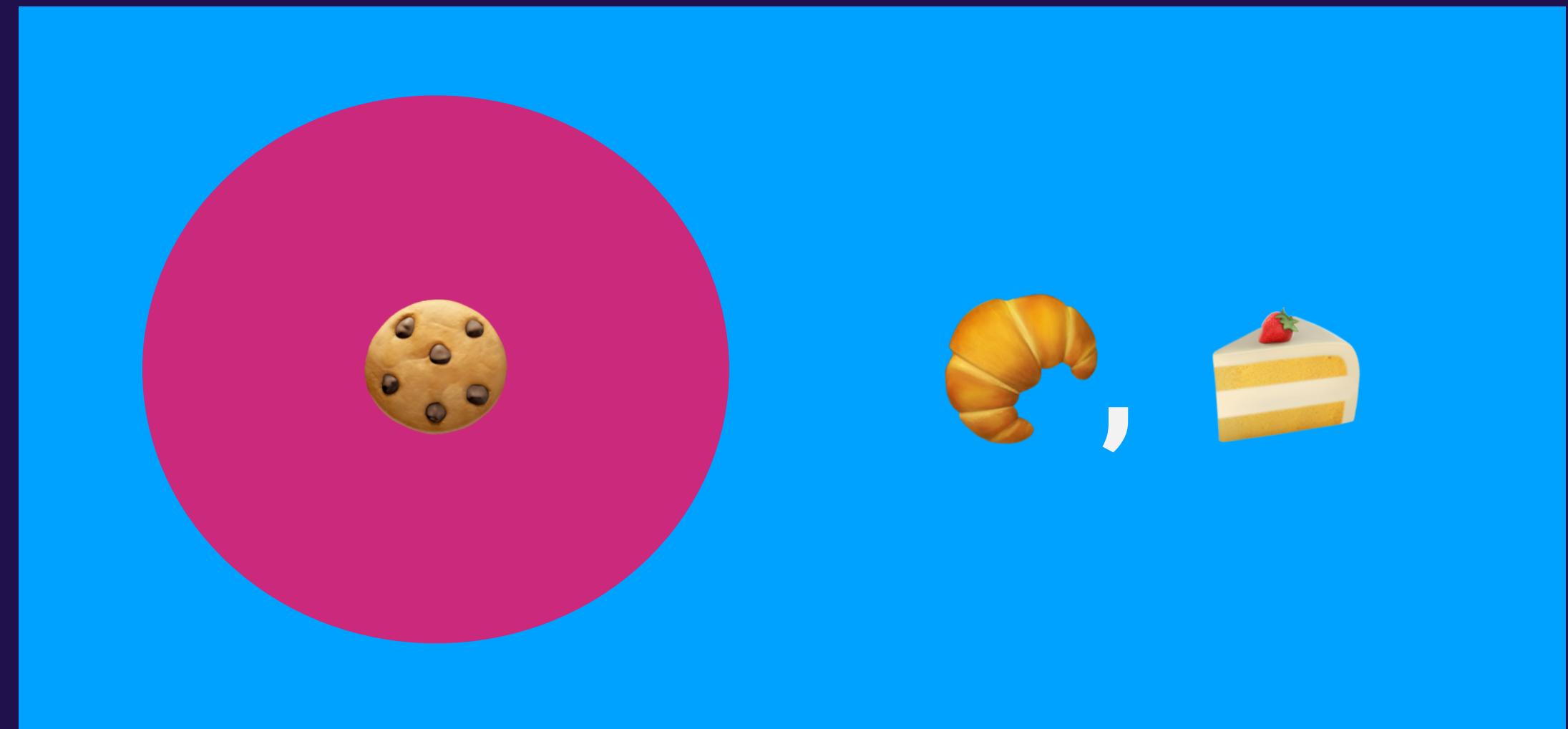


$$p(\cookie) + p(\croissant) = 1$$

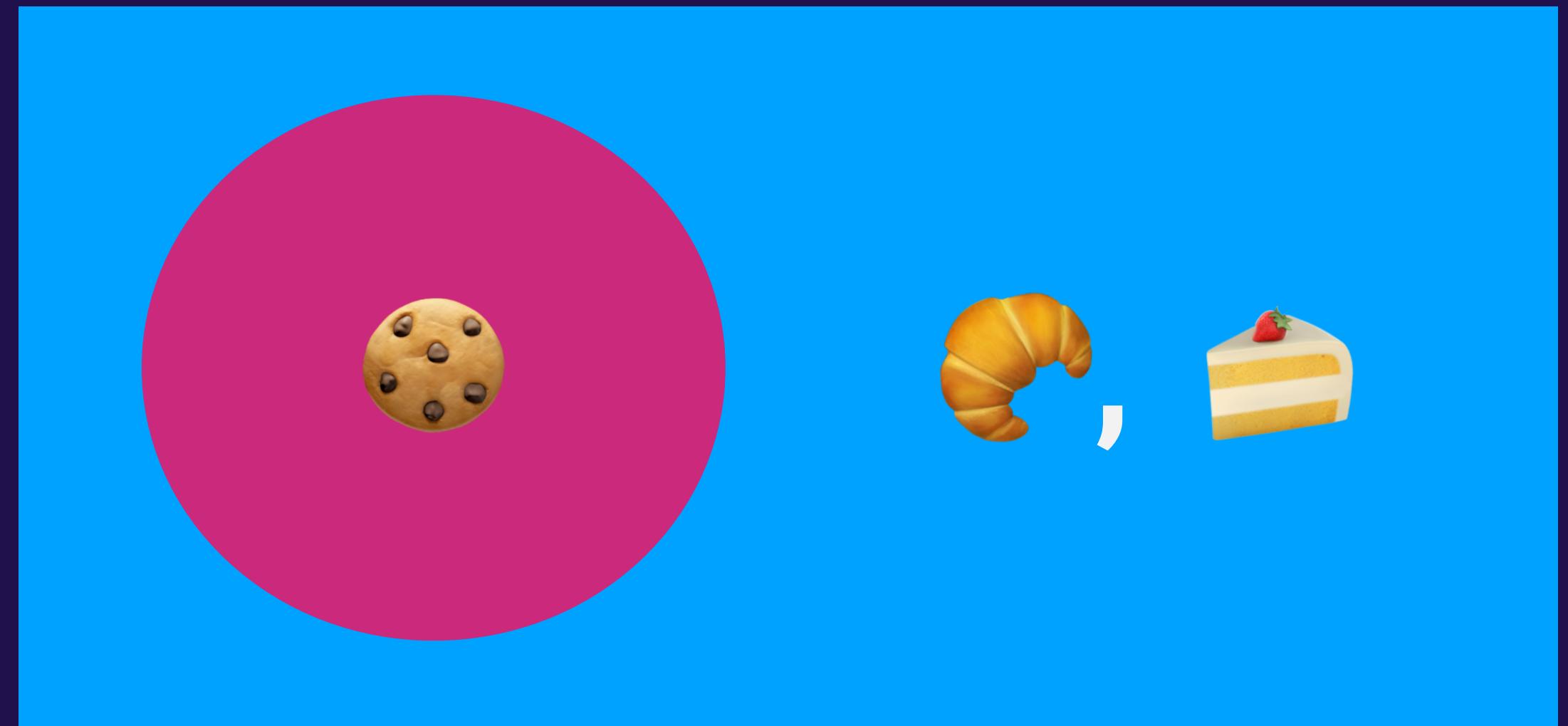
$$\sum_i P_i(x) = 1$$



$$p(\text{🍪}) + p(\text{🥐}) + p(\text{🍰}) = 1$$

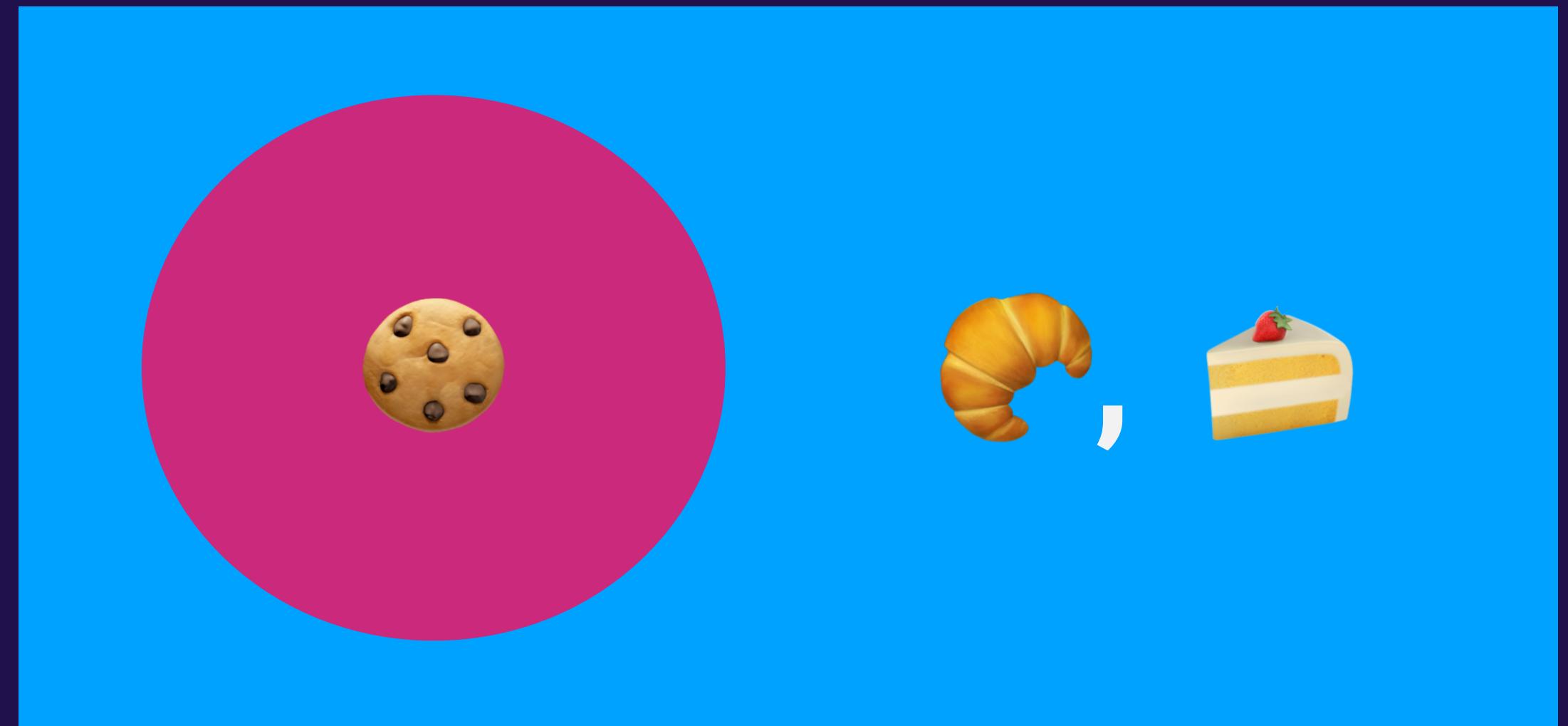


$$p(\text{🍪}) + p(\text{🥐}) + p(\text{🍰}) = 1$$



$$p(\text{🥐}) + p(\text{🍰}) = p(\text{not } \text{🍪}) \quad \text{"complement"}$$

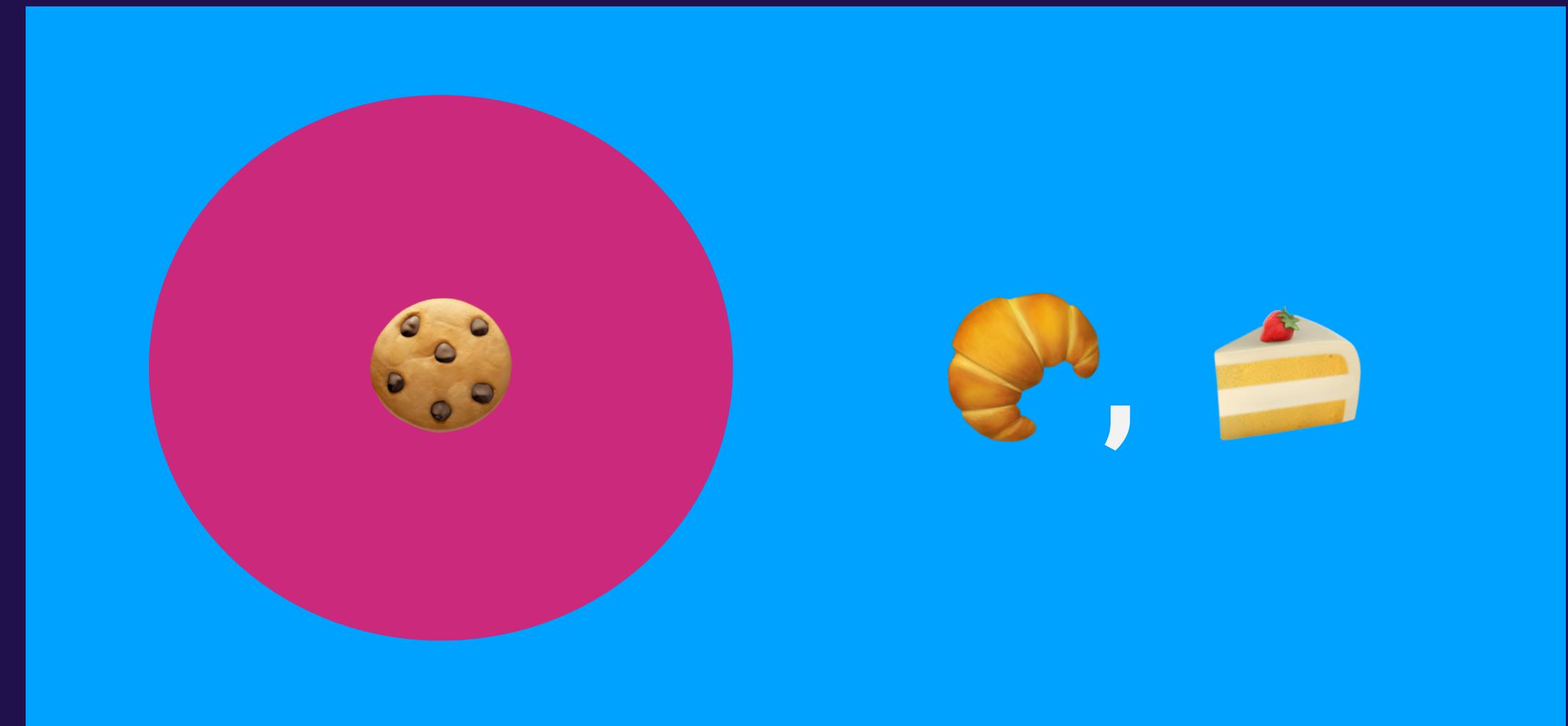
$$p(\text{Cookie}) + p(\text{Croissant}) + p(\text{Slice}) = 1$$



$$p(\text{Croissant}) + p(\text{Slice}) = p(\text{not Cookie}) \quad \text{"complement"}$$

$$p(\text{Cookie}) =$$

$$p(\text{🍪}) + p(\text{🥐}) + p(\text{🍰}) = 1$$



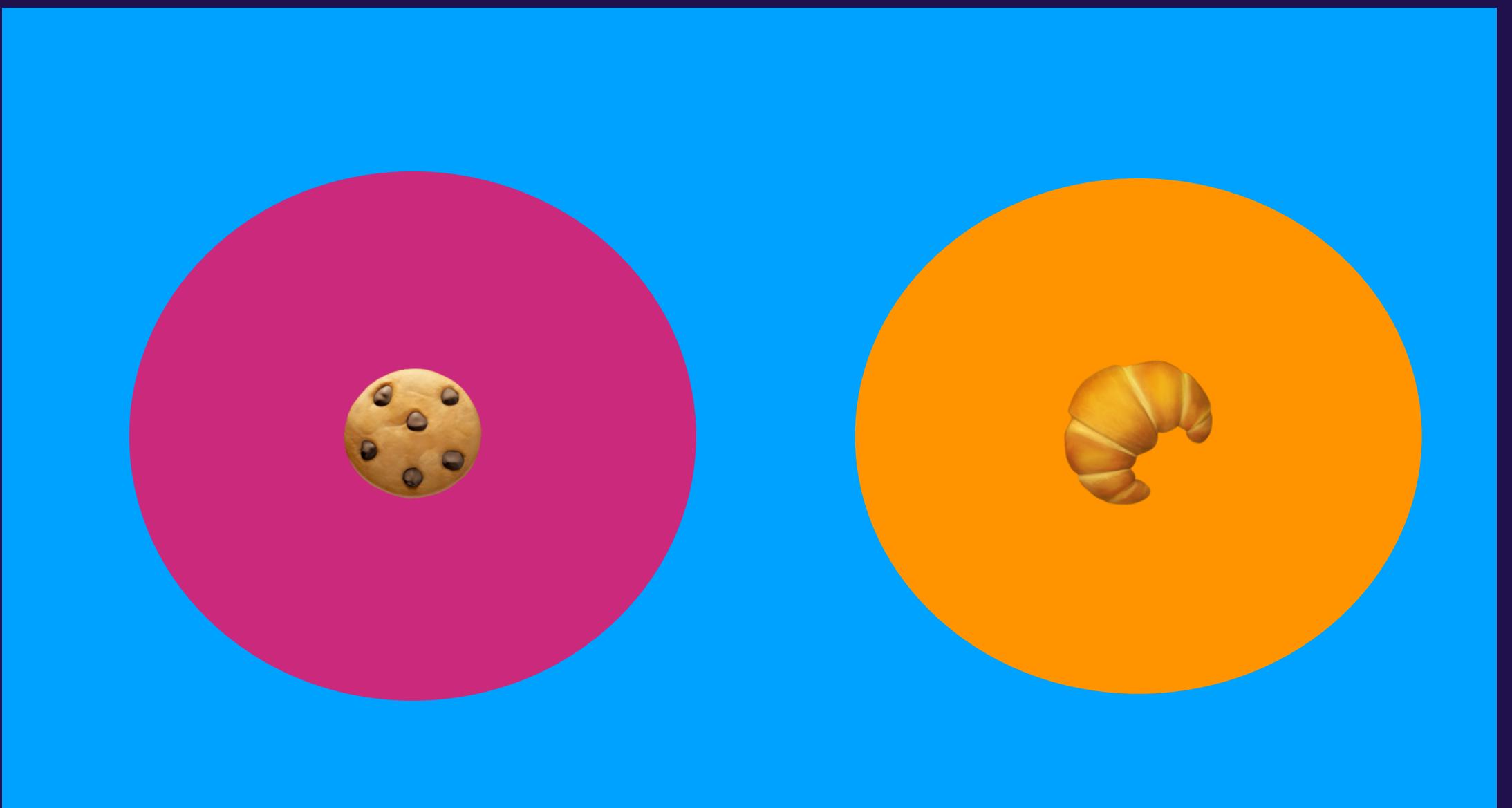
$$p(\text{🥐}) + p(\text{🍰}) = p(\text{not } \text{🍪})$$

“complement”

$$p(\text{🍪}) = 1 - p(\text{not } \text{🍪})$$

$p(\text{cookie} \cap \text{croissant}) = "p(\text{cookie and croissant})"$

$p(\text{cookie} \cup \text{croissant}) = "p(\text{cookie or croissant})"$



$p(\text{cookie} \cap \text{croissant}) = "p(\text{cookie and croissant})"$

$p(\text{cookie} \cup \text{croissant}) = "p(\text{cookie or croissant})"$

cookie , croissant are independent:



$p(\text{cookie} \cap \text{croissant}) = "p(\text{cookie and croissant})"$

$p(\text{cookie} \cup \text{croissant}) = "p(\text{cookie or croissant})"$

cookie , croissant are independent:

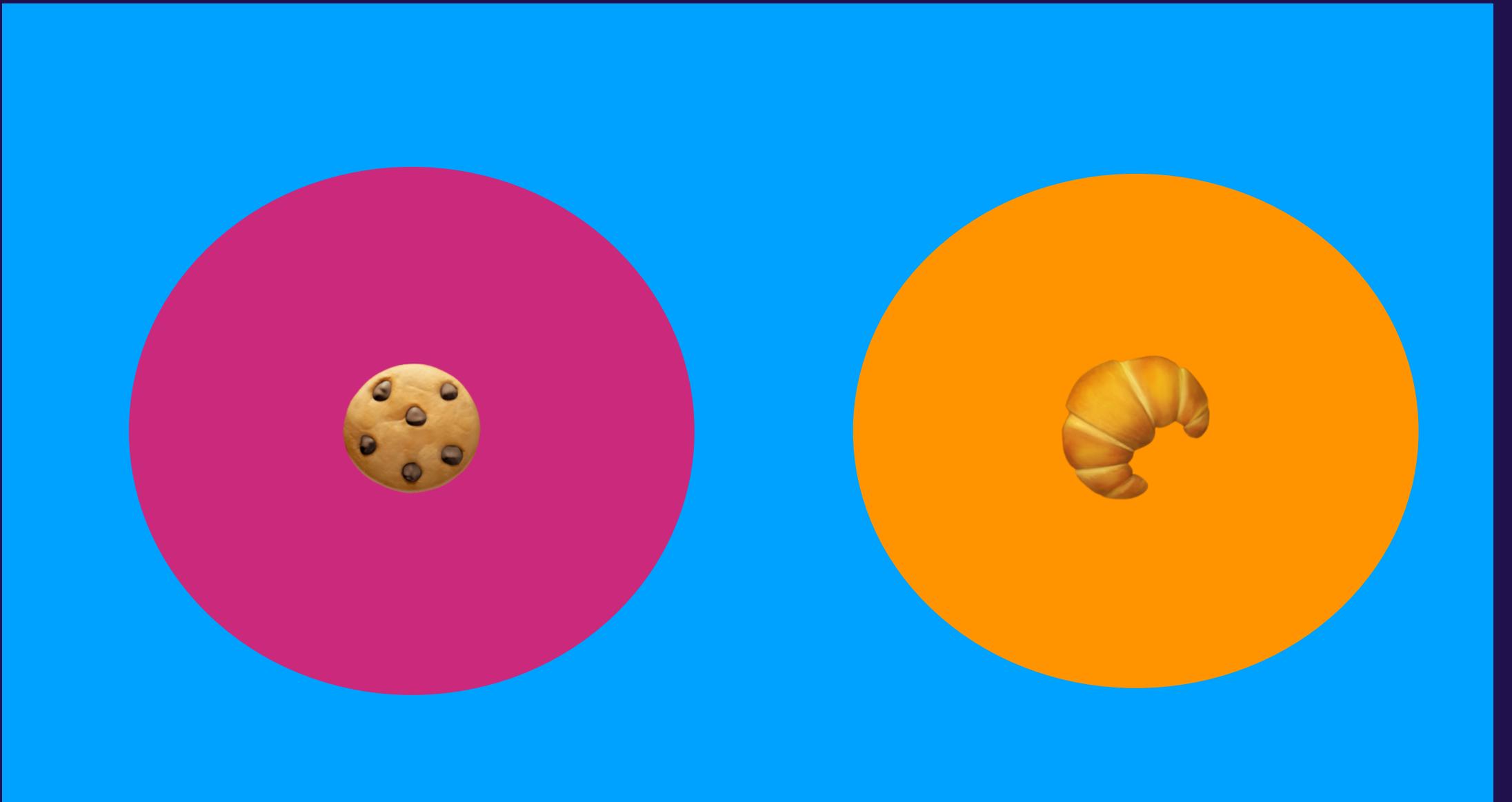
$p(\text{cookie} \cap \text{croissant}) = 0$



$p(\text{cookie} \cap \text{croissant}) = "p(\text{cookie and croissant})"$

$p(\text{cookie} \cup \text{croissant}) = "p(\text{cookie or croissant})"$

cookie , croissant are independent:

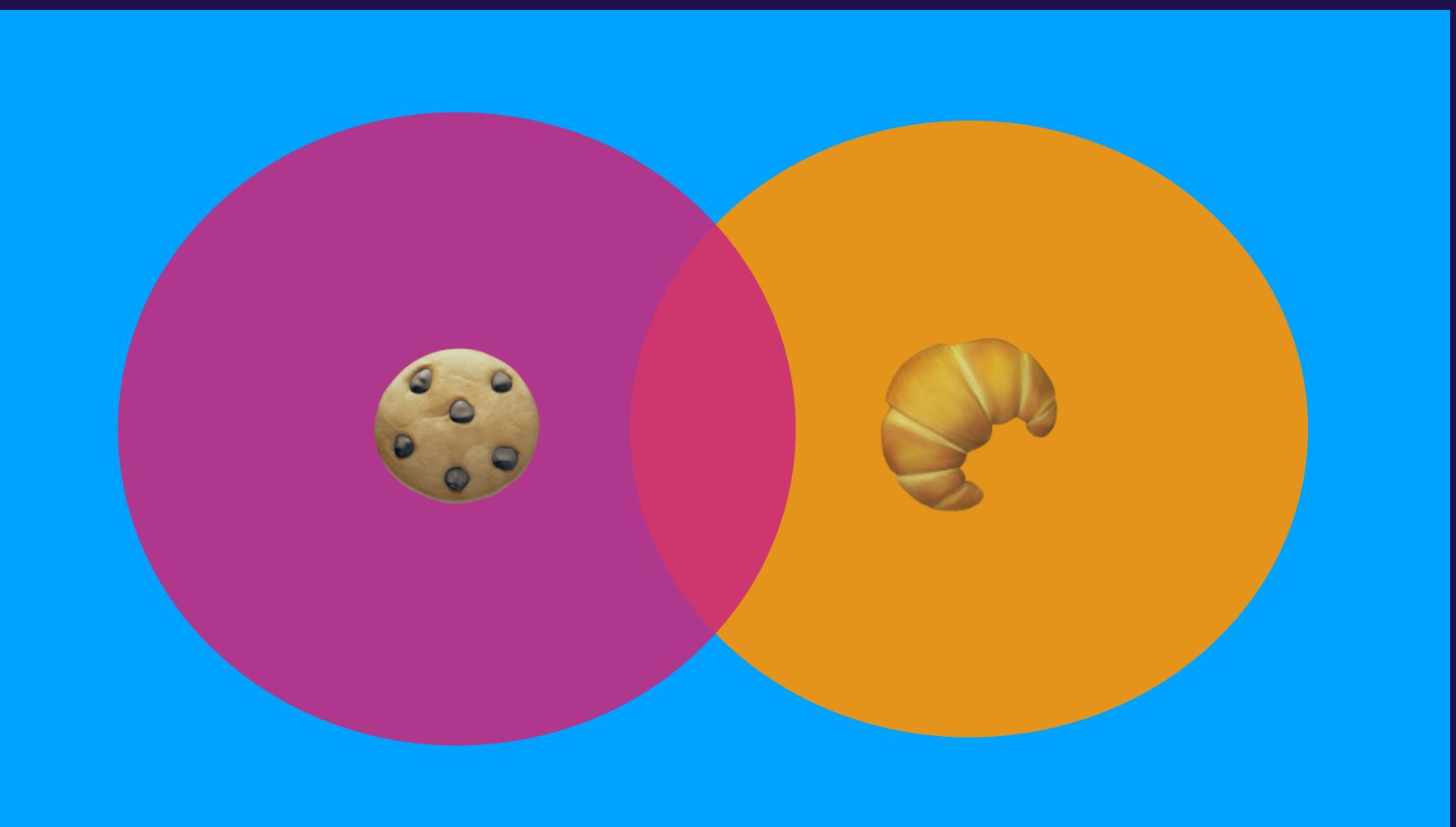


$p(\text{cookie} \cap \text{croissant}) = 0$

$p(\text{cookie} \cup \text{croissant}) = p(\text{cookie}) + p(\text{croissant})$

$p(\text{Cookie} \cap \text{Croissant}) = "p(\text{Cookie and Croissant})"$

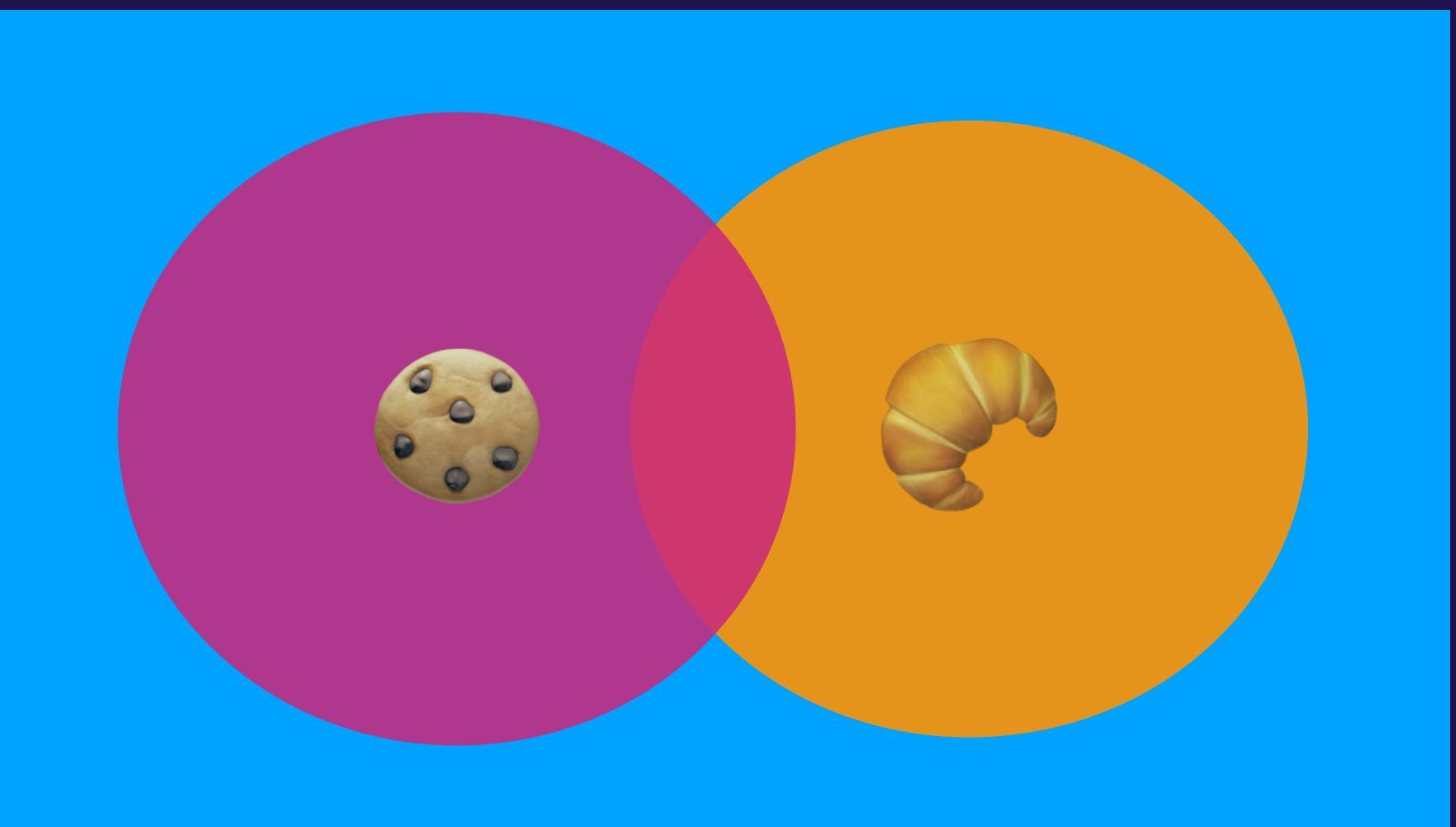
$p(\text{Cookie} \cup \text{Croissant}) = "p(\text{Cookie or Croissant})"$



$p(\text{Cookie} \cap \text{Croissant}) = "p(\text{Cookie and Croissant})"$

$p(\text{Cookie} \cup \text{Croissant}) = "p(\text{Cookie or Croissant})"$

Cookie , Croissant are **not** independent:

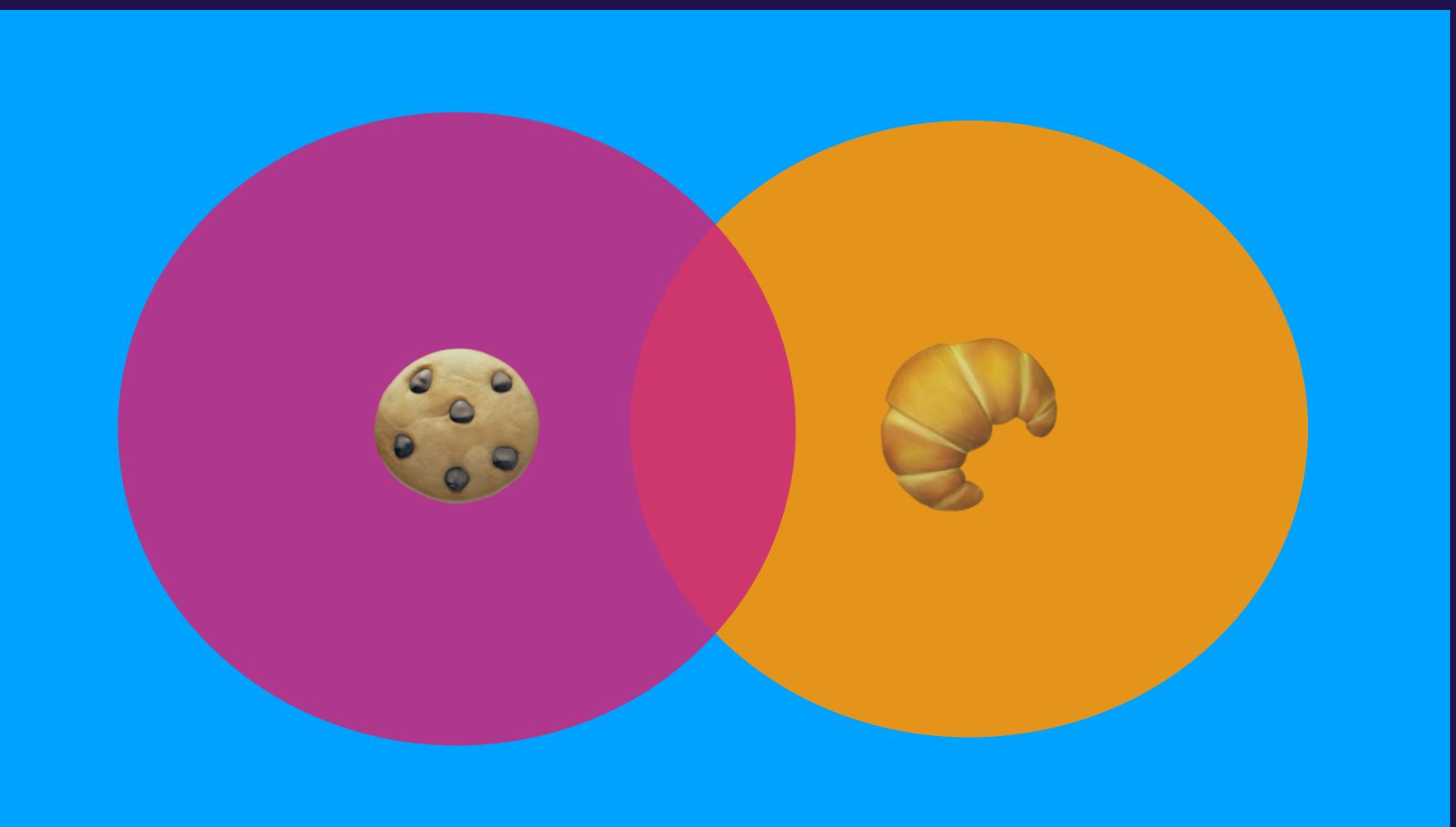


$p(\text{cookie} \cap \text{croissant}) = "p(\text{cookie and croissant})"$

$p(\text{cookie} \cup \text{croissant}) = "p(\text{cookie or croissant})"$

cookie , croissant are **not** independent:

$p(\text{cookie} \cap \text{croissant}) = x$



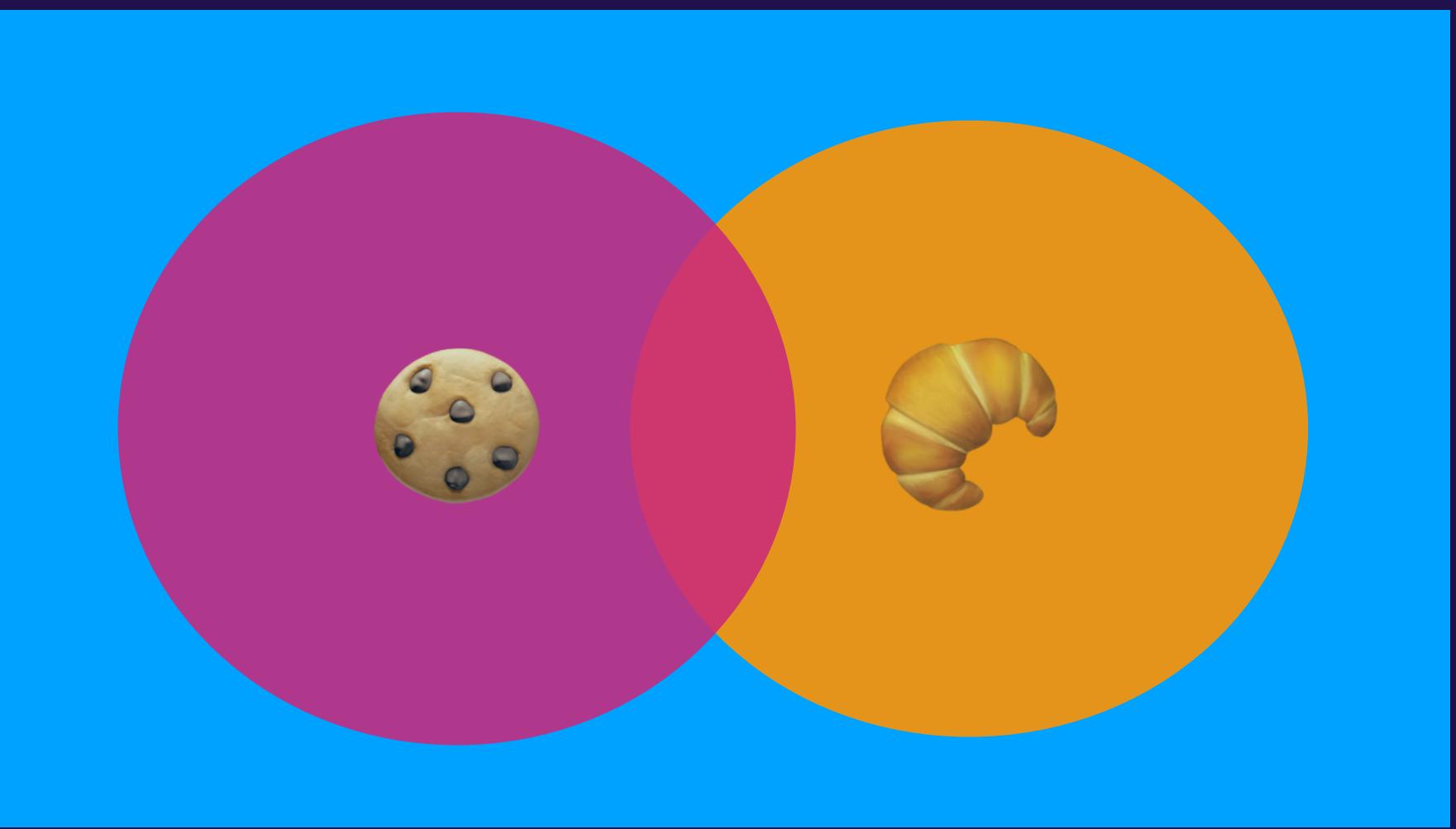
$p(\text{cookie} \cap \text{croissant}) = "p(\text{cookie and croissant})"$

$p(\text{cookie} \cup \text{croissant}) = "p(\text{cookie or croissant})"$

cookie , croissant are **not** independent:

$p(\text{cookie} \cap \text{croissant}) = x$

$p(\text{cookie} \cup \text{croissant}) = p(\text{cookie}) + p(\text{croissant})$



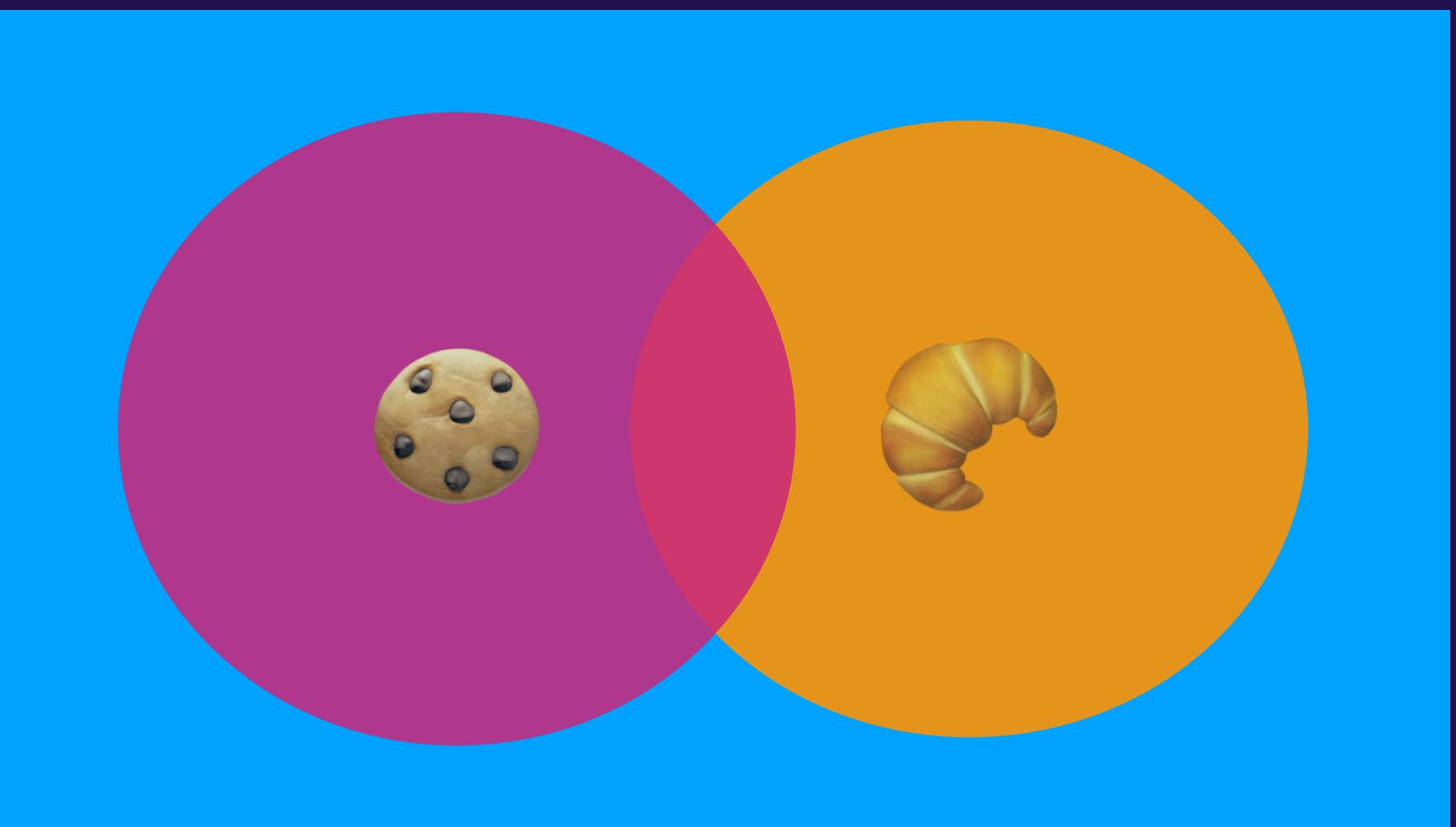
$p(\text{cookie} \cap \text{croissant}) = "p(\text{cookie and croissant})"$

$p(\text{cookie} \cup \text{croissant}) = "p(\text{cookie or croissant})"$

cookie , croissant are **not** independent:

$p(\text{cookie} \cap \text{croissant}) = x$

$p(\text{cookie} \cup \text{croissant}) = p(\text{cookie}) + p(\text{croissant}) - p(\text{cookie} \cap \text{croissant})$



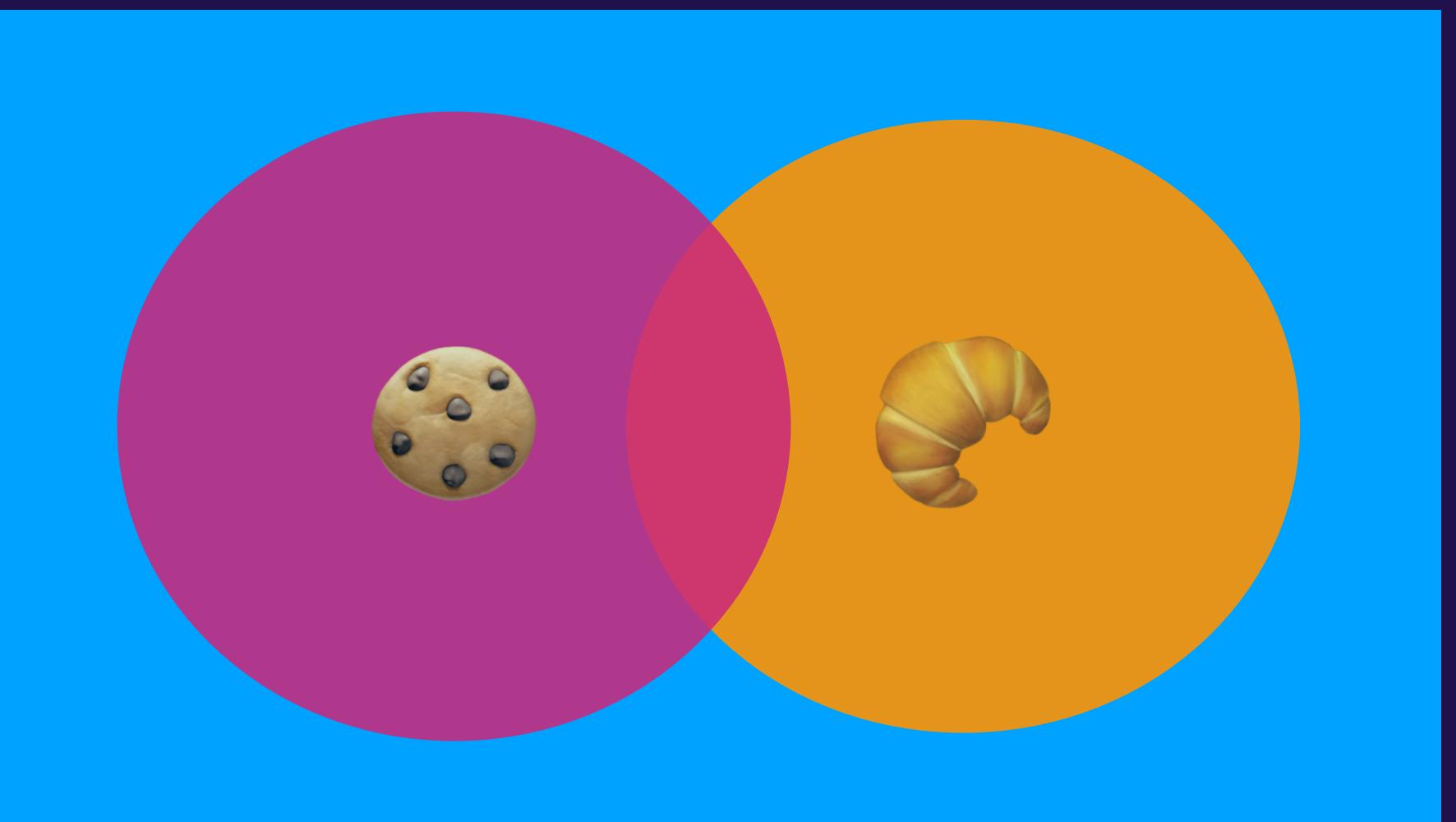
$p(\text{Cookie} \cap \text{Croissant}) = "p(\text{Cookie and Croissant})"$

$p(\text{Cookie} \cup \text{Croissant}) = "p(\text{Cookie or Croissant})"$

Cookie , Croissant are **not** independent:

$p(\text{Cookie} \cap \text{Croissant}) = x$

$p(\text{Cookie} \cup \text{Croissant}) = p(\text{Cookie}) + p(\text{Croissant}) - p(\text{Cookie} \cap \text{Croissant})$



Let's add another category

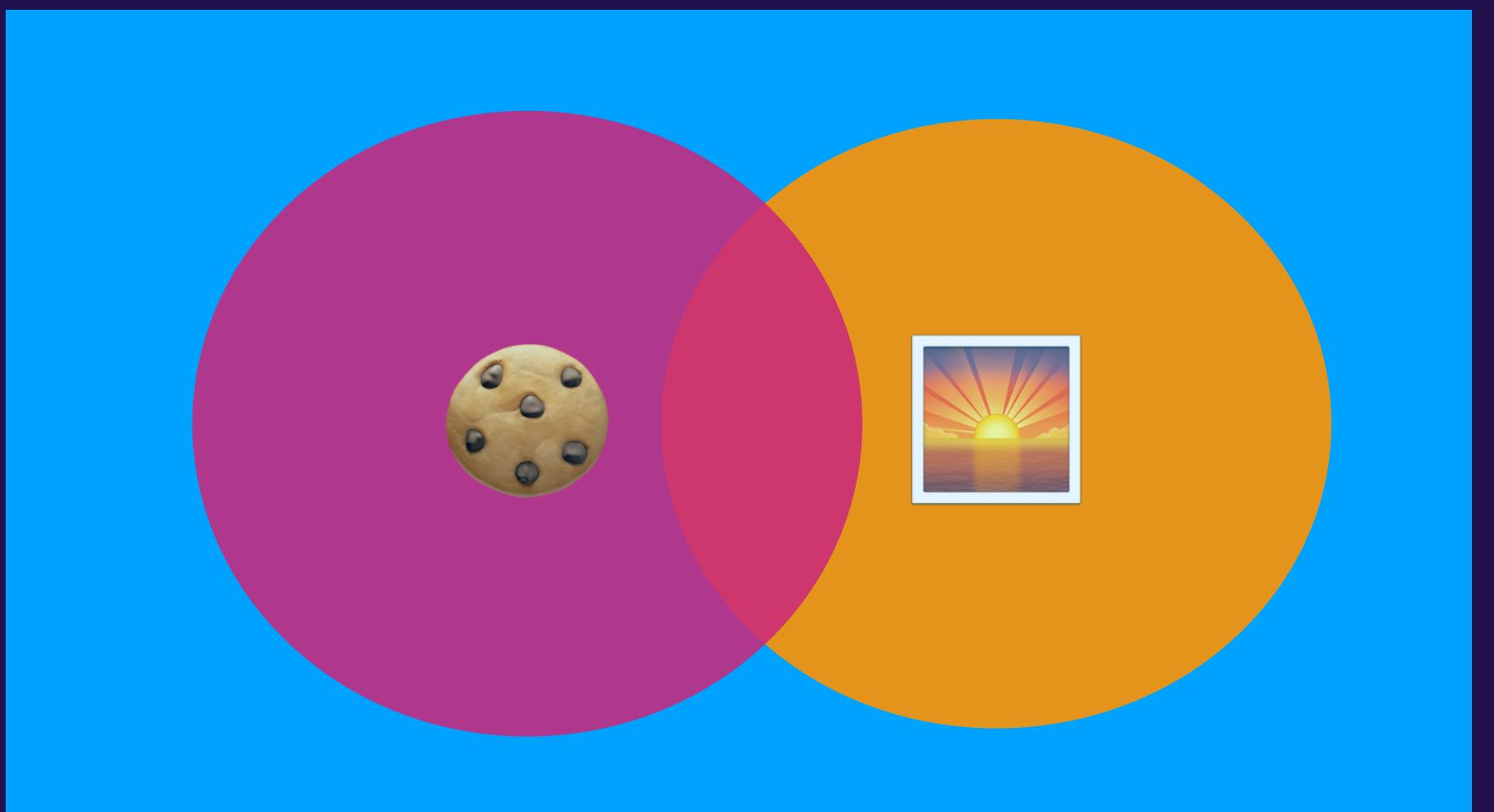
Is there a break with snacks in the morning or the afternoon?

$p(\square)$ = there are snacks in the morning

$p(\text{😴})$ = there are snacks in the afternoon

Do we get cookies more often in the morning or in the afternoon?

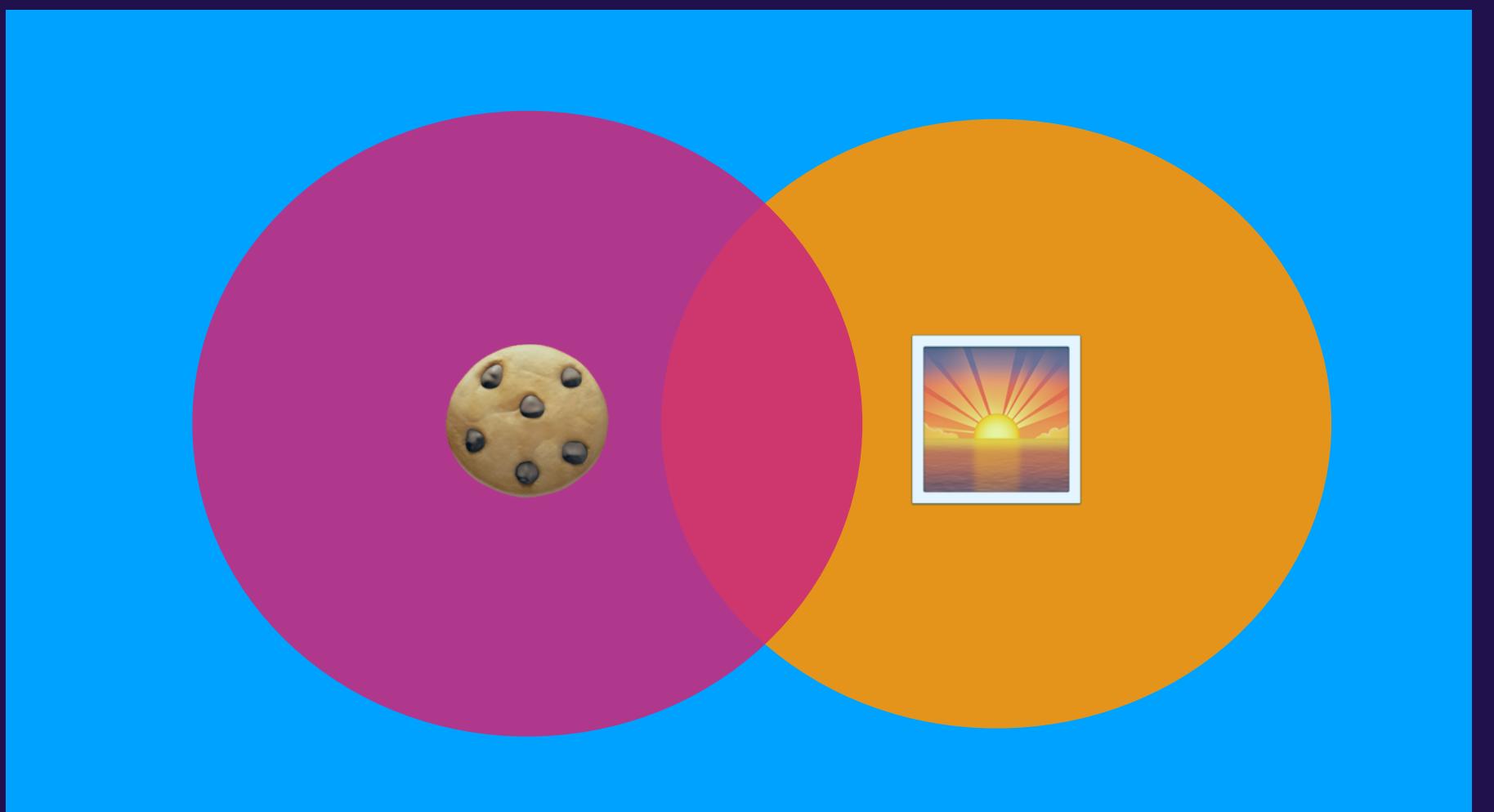
$p(\text{cookie} \mid \text{sunrise})$ = “probability of cookie given sunrise”



Do we get cookies more often in the morning or in the afternoon?

$p(\text{cookie} | \text{sun})$ = “probability of cookie given sun”

$p(\text{cookie} | \text{sun}) = p(\text{cookie} \cap \text{sun}) / p(\text{sun})$

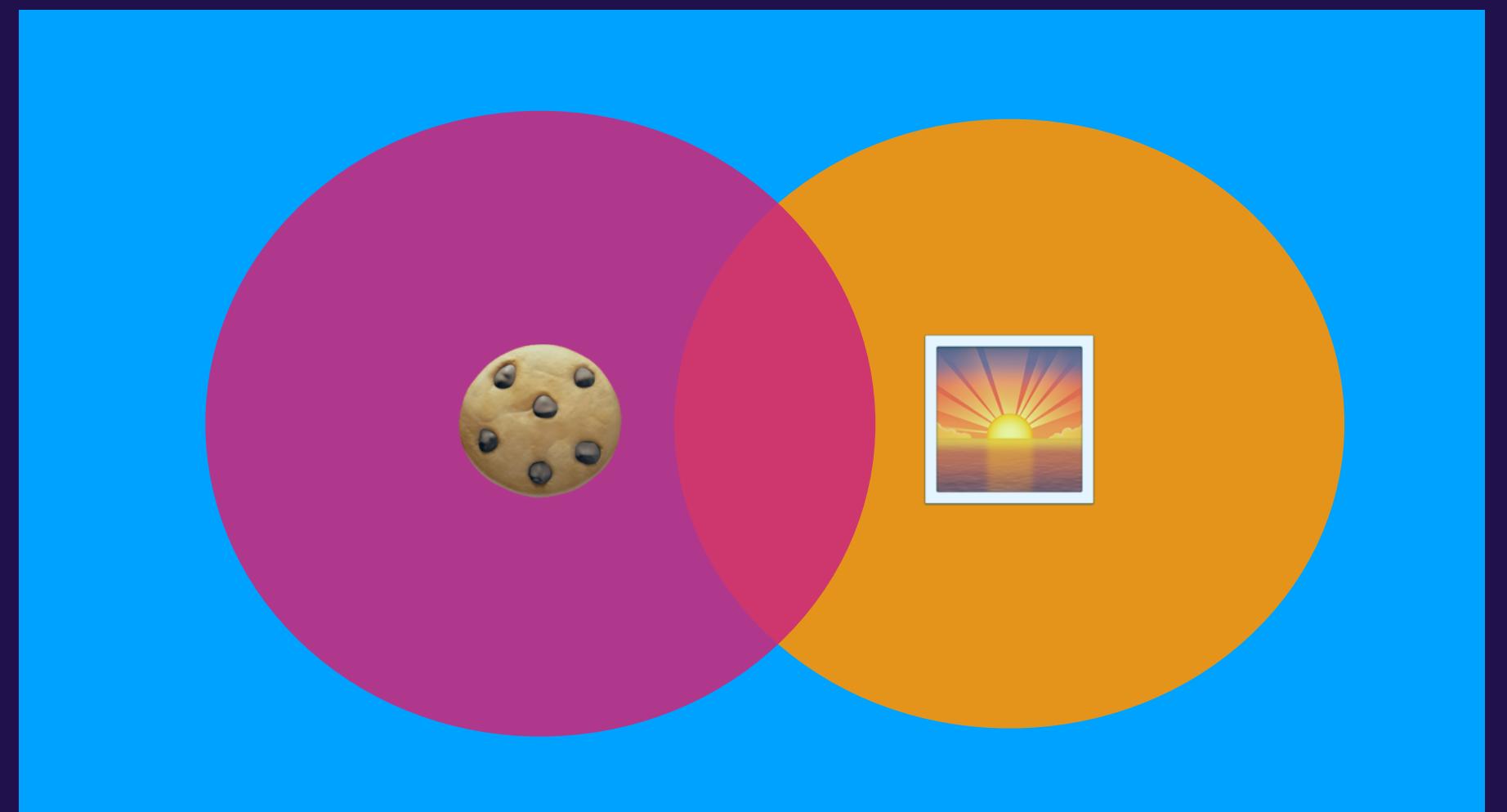


Do we get cookies more often in the morning or in the afternoon?

$p(\text{cookie} | \text{sun})$ = “probability of cookie given sun”

$p(\text{cookie} | \text{sun}) = p(\text{cookie} \cap \text{sun}) / p(\text{sun})$

$p(\text{cookie} | \text{sun}) = p(\text{cookie}, \text{sun}) / p(\text{sun})$



Which snacks do we get at which time of the day?

	$p(\text{cookie})$	$p(\text{croissant})$	$p(\text{cake})$
$p(\text{sunset})$	0,25	0,06	0,29
$p(\text{sleepy})$	0,25	0,14	0,01

Join at
slido.com
#3643 573



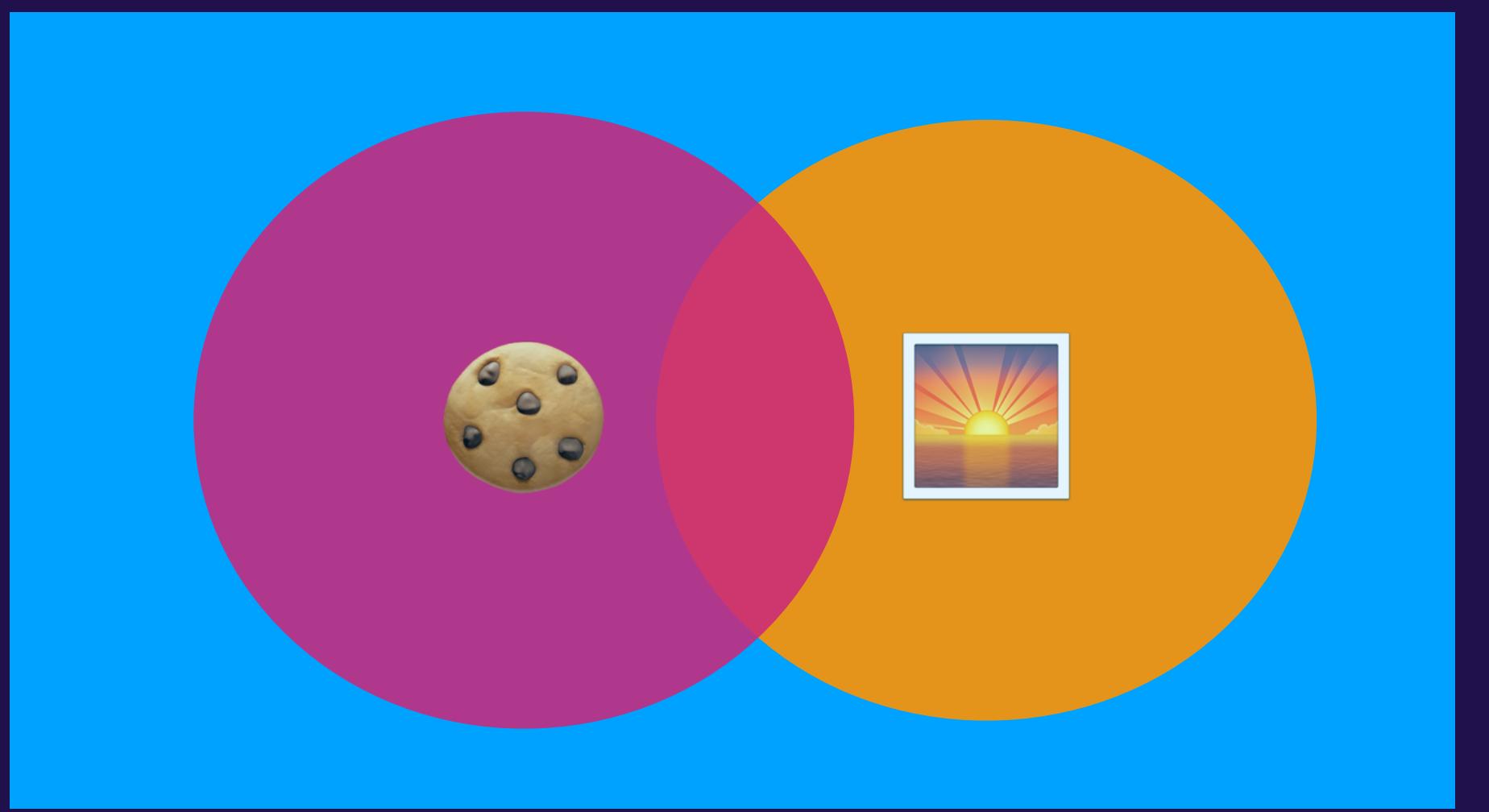
Which snacks do we get at which time of the day?

	$p(\text{cookie})$	$p(\text{croissant})$	$p(\text{cake})$
$p(\text{sunset})$	0,25	0,06	0,29
$p(\text{sleepy})$	0,25	0,14	0,01

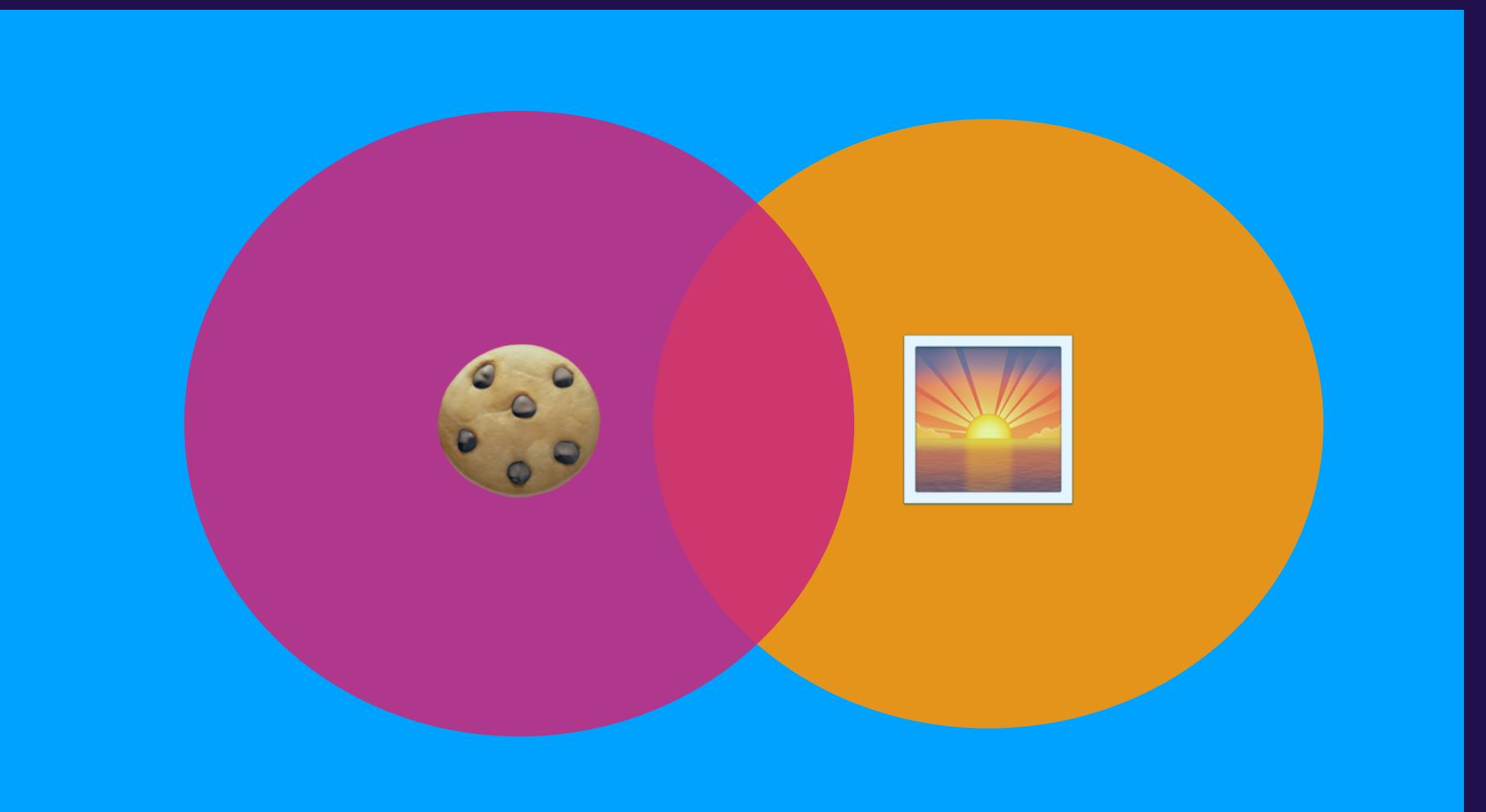
- $p(\text{cake}) = ?$
- $p(\text{sunset} \mid \text{croissant}) = ?$
- $p(\text{cake} \mid \text{sunset}) = ?$

Join at
slido.com
#3643 573



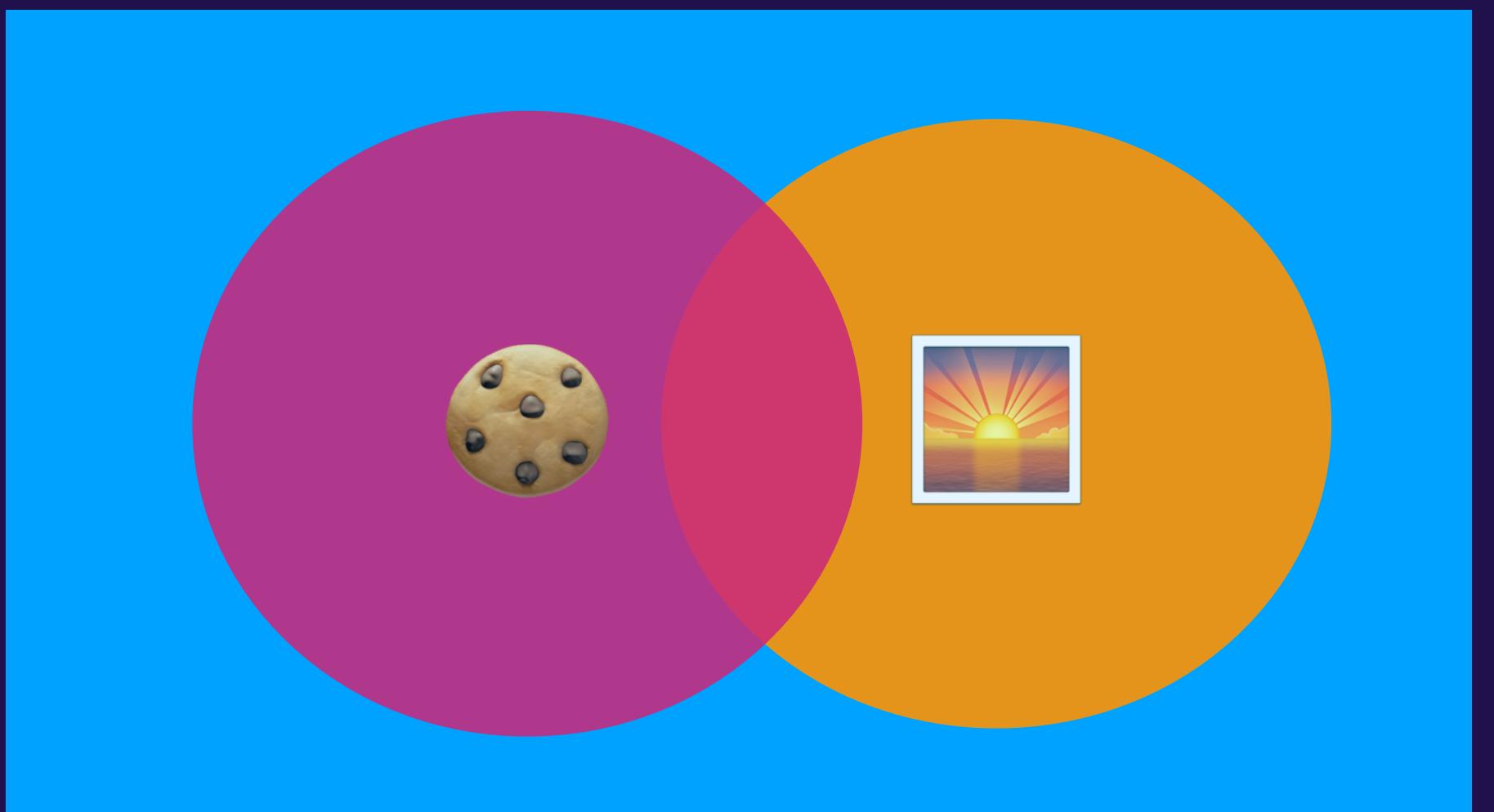


$$p(\text{🍪} | \boxed{\text{☀}}) = p(\text{🍪}, \boxed{\text{☀}}) / p(\boxed{\text{☀}})$$



$$p(\text{🍪} | \text{🌅}) = p(\text{🍪}, \text{🌅}) / p(\text{🌅})$$

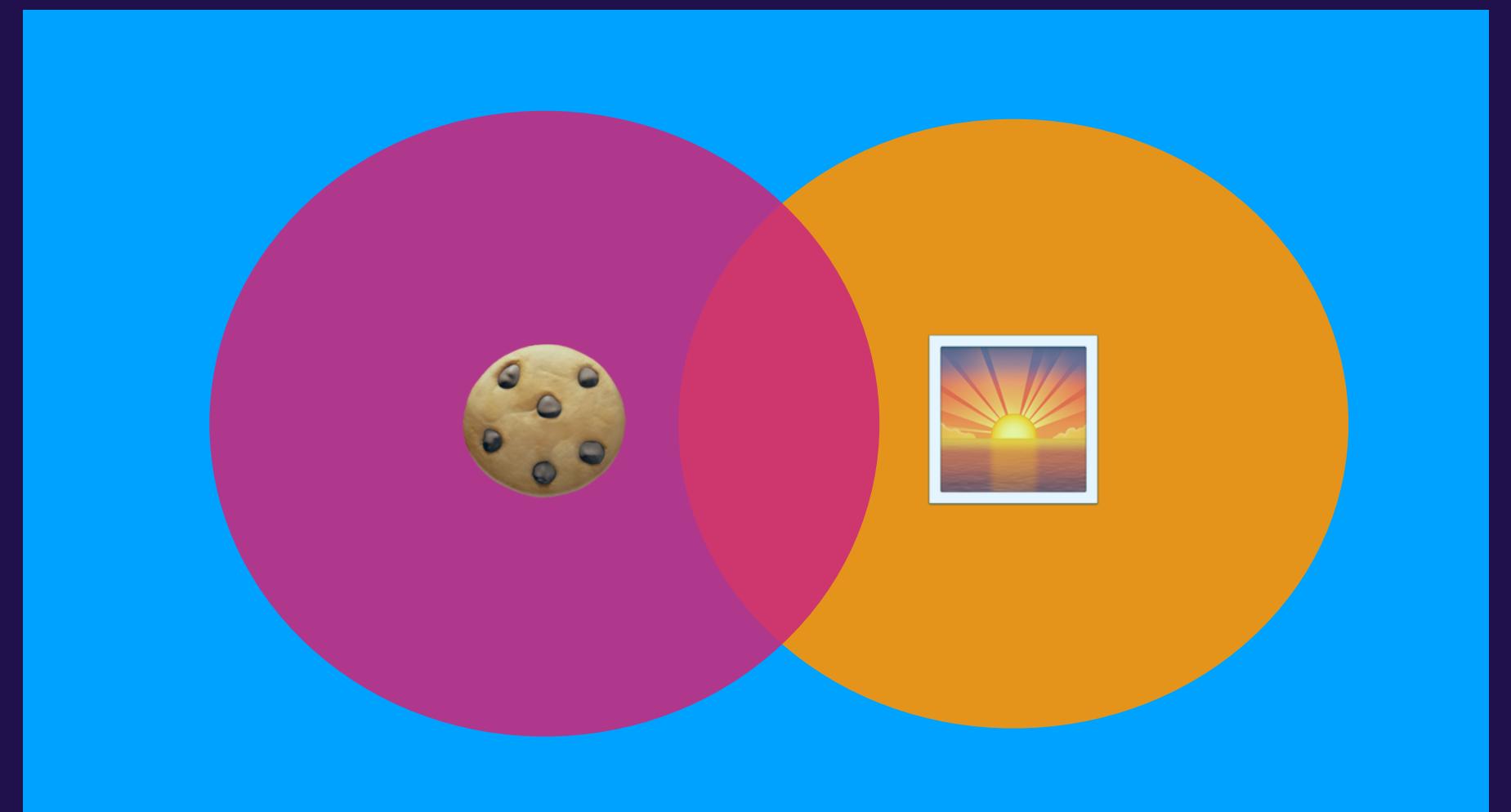
What is $p(\text{🌅} | \text{🍪})$?

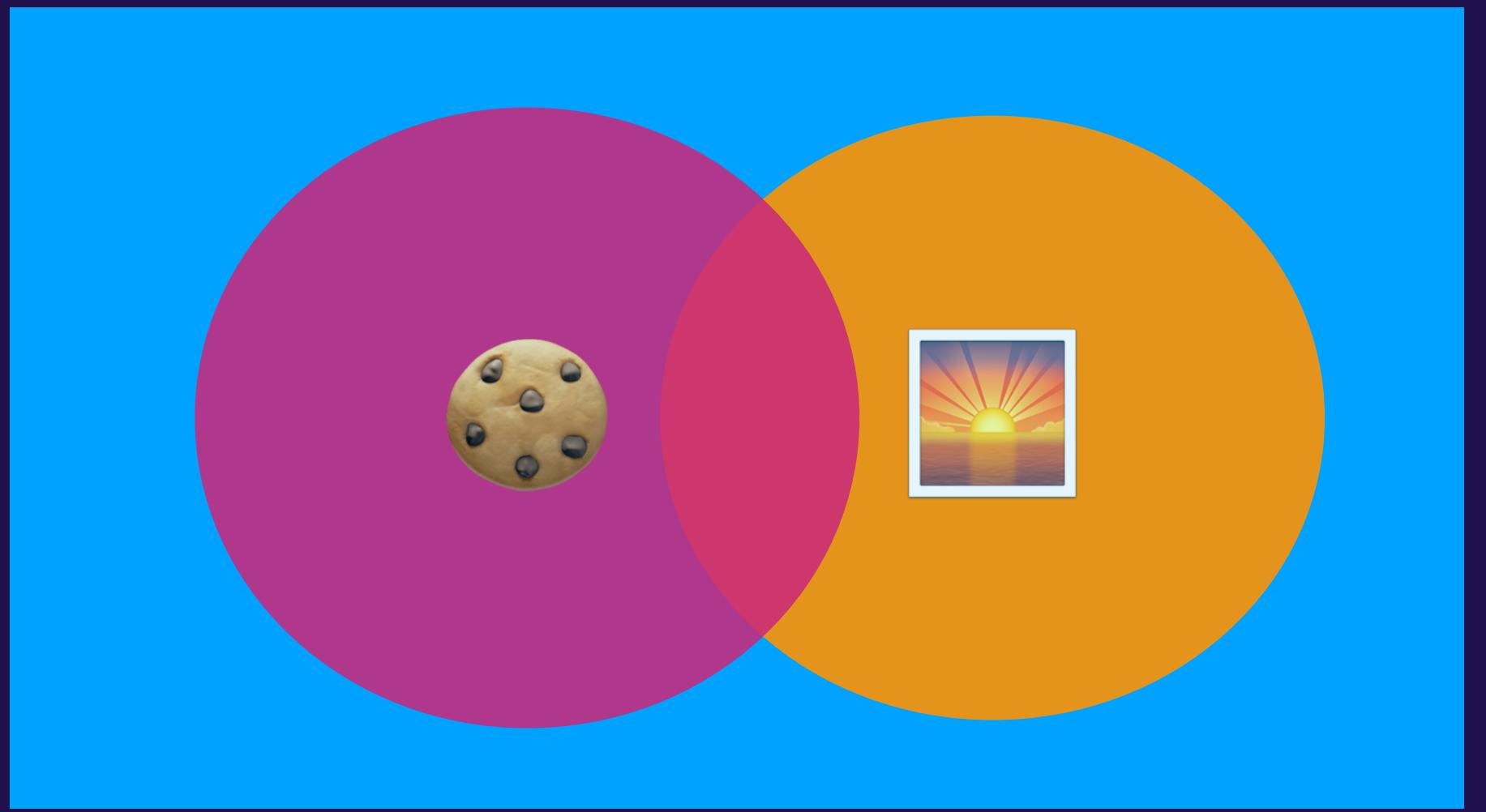


$$p(\text{🍪} \mid \text{🌅}) = p(\text{🍪}, \text{🌅}) / p(\text{🌅})$$

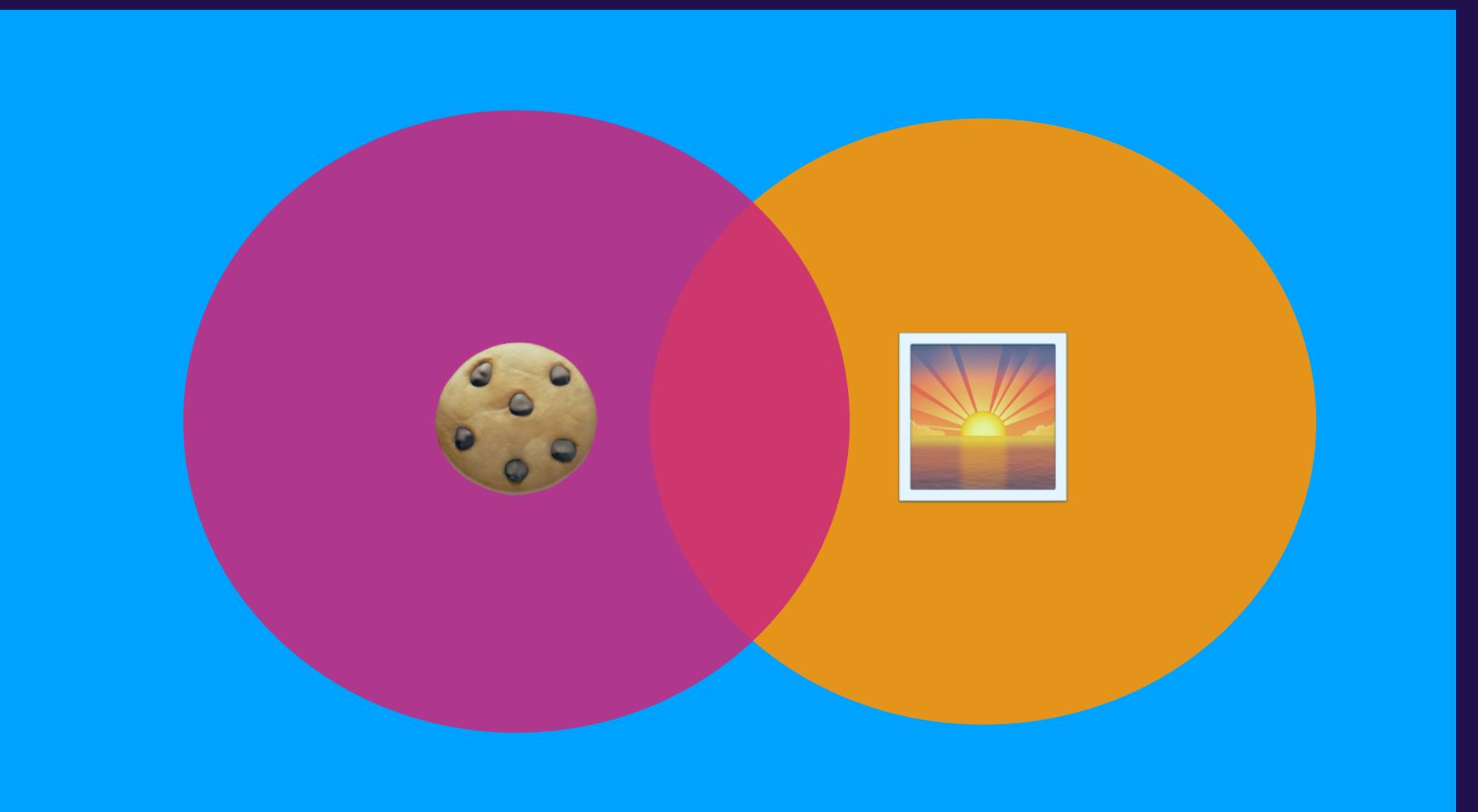
What is $p(\text{🌅} \mid \text{🍪})$?

$$p(\text{🌅} \mid \text{🍪}) = p(\text{🍪}, \text{🌅}) / p(\text{🍪})$$



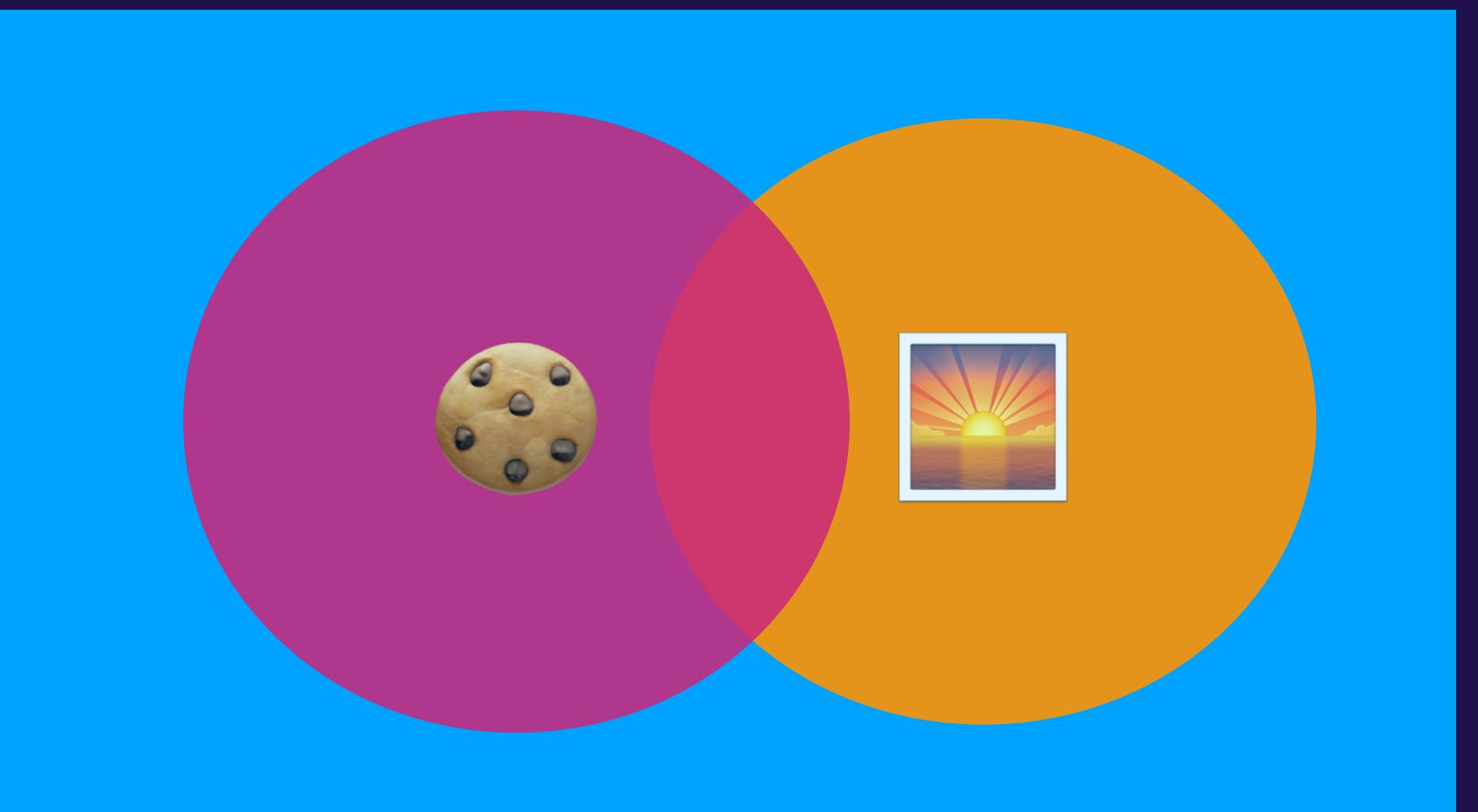


$$p(\text{🍪} \mid \text{🌅}) = p(\text{🍪}, \text{🌅}) / p(\text{🌅})$$



$$p(\text{🍪} | \text{🌅}) = p(\text{🍪}, \text{🌅}) / p(\text{🌅})$$

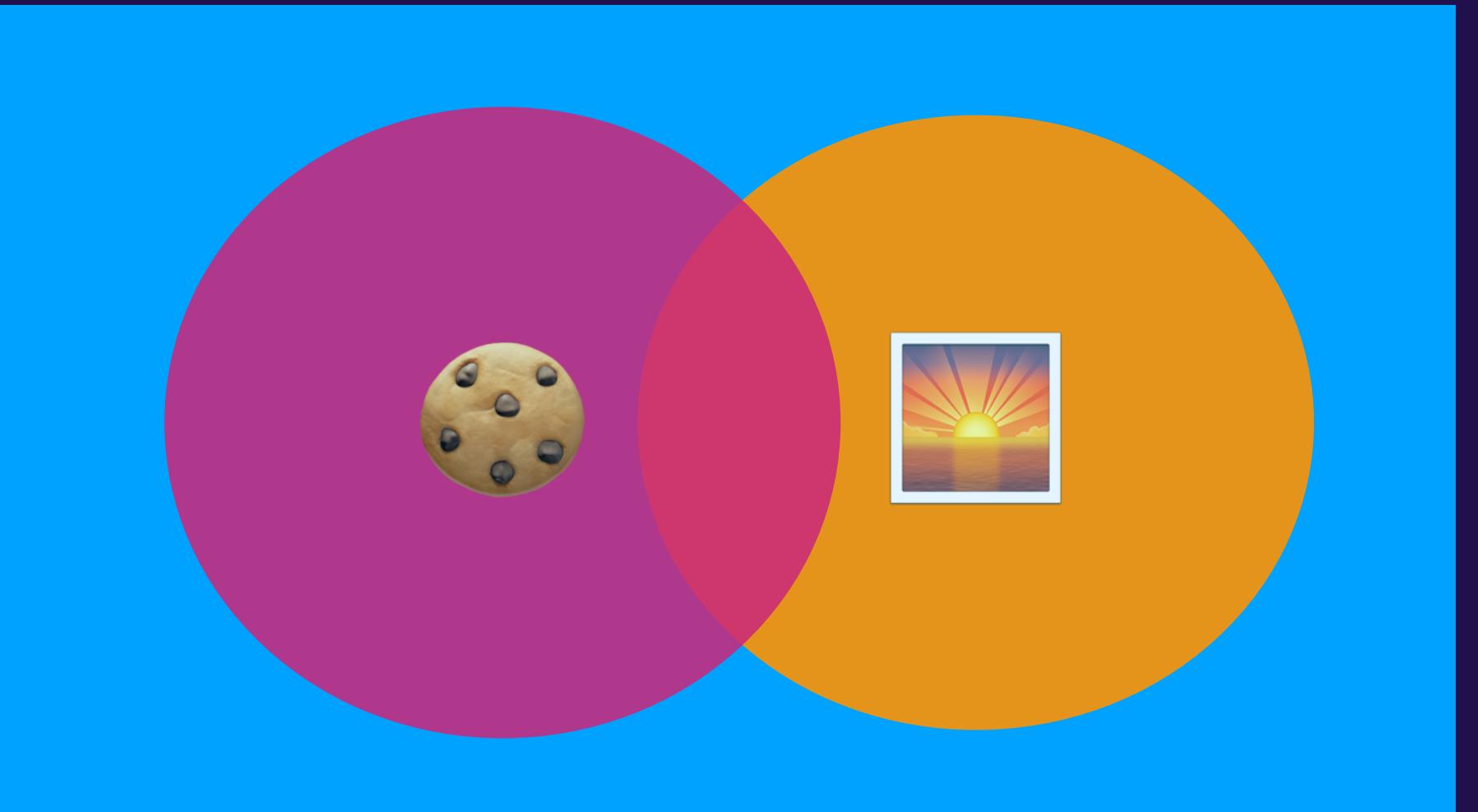
$$p(\text{🌅} | \text{🍪}) = p(\text{🍪}, \text{🌅}) / p(\text{🍪})$$



$$p(\text{Cookie} | \text{Sunset}) = p(\text{Cookie}, \text{Sunset}) / p(\text{Sunset})$$

$$p(\text{Sunset} | \text{Cookie}) = p(\text{Cookie}, \text{Sunset}) / p(\text{Cookie})$$

$$p(\text{Sunset} | \text{Cookie}) p(\text{Cookie}) = p(\text{Cookie} | \text{Sunset}) p(\text{Sunset})$$

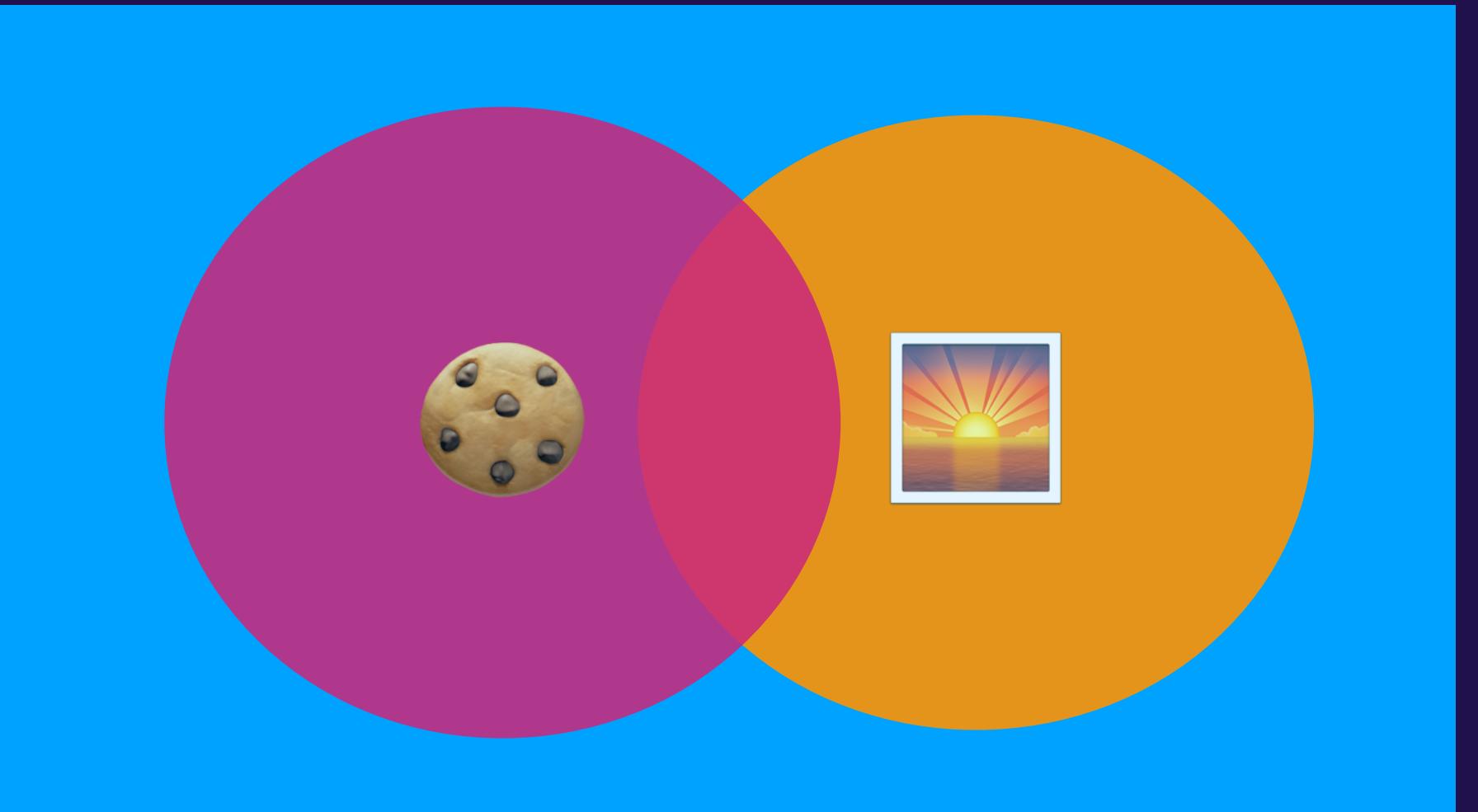


$$p(\text{Cookie} | \text{Sunset}) = p(\text{Cookie}, \text{Sunset}) / p(\text{Sunset})$$

$$p(\text{Sunset} | \text{Cookie}) = p(\text{Cookie}, \text{Sunset}) / p(\text{Cookie})$$

$$p(\text{Sunset} | \text{Cookie}) p(\text{Cookie}) = p(\text{Cookie} | \text{Sunset}) p(\text{Sunset})$$

Bayes rule

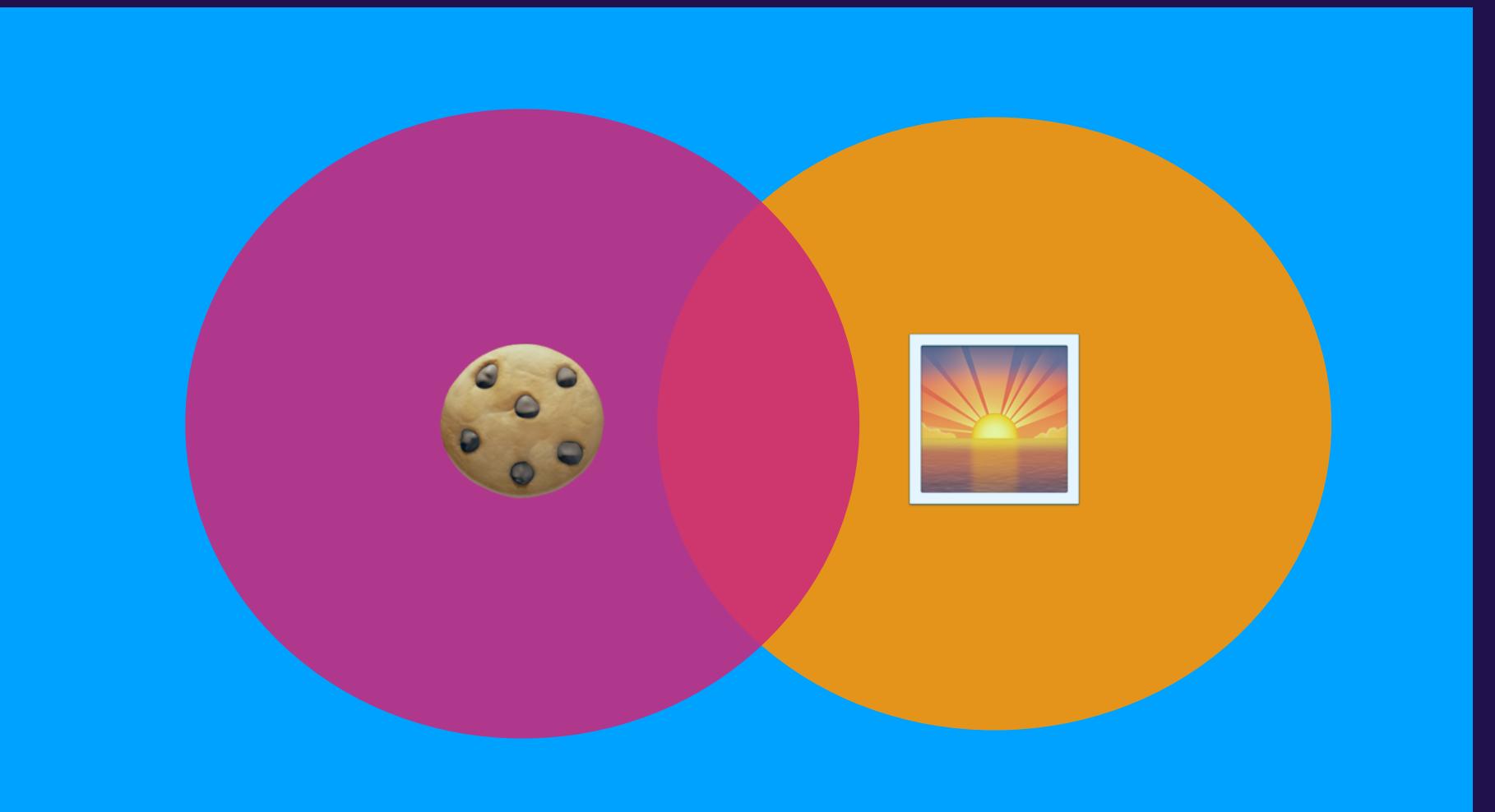
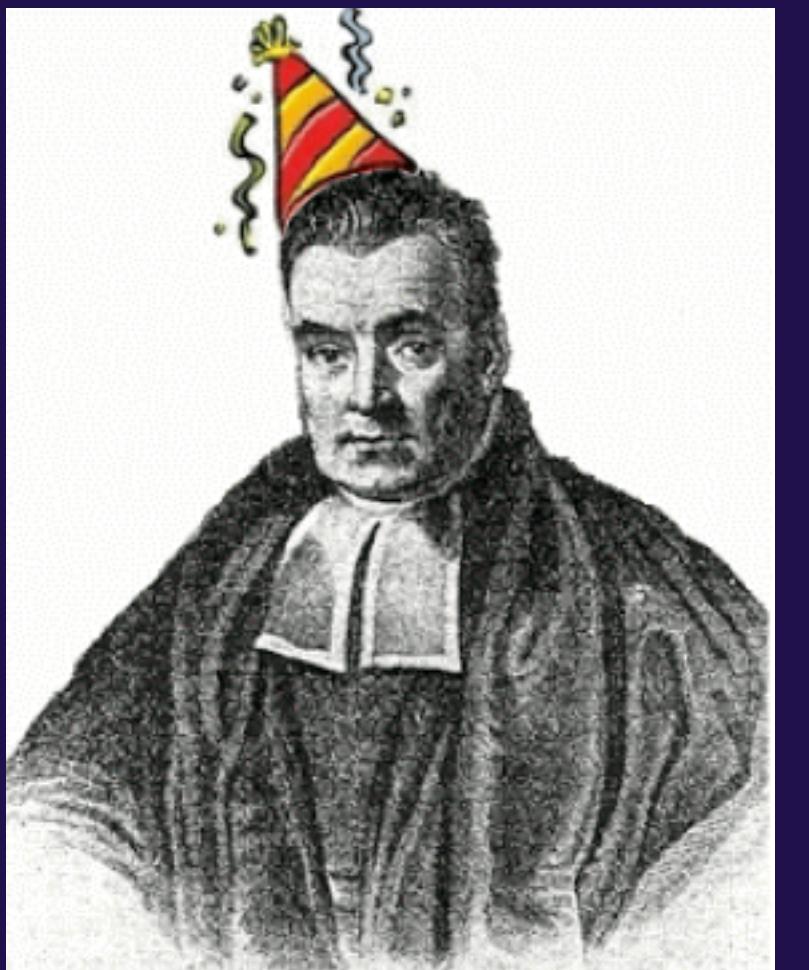


$$p(\text{Cookie} | \text{Sunset}) = p(\text{Cookie}, \text{Sunset}) / p(\text{Sunset})$$

$$p(\text{Sunset} | \text{Cookie}) = p(\text{Cookie}, \text{Sunset}) / p(\text{Cookie})$$

$$p(\text{Sunset} | \text{Cookie}) p(\text{Cookie}) = p(\text{Cookie} | \text{Sunset}) p(\text{Sunset})$$

Bayes rule



$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$

Bayes rule

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$

Bayes rule

Important: $P(A | B) \neq P(B | A)$

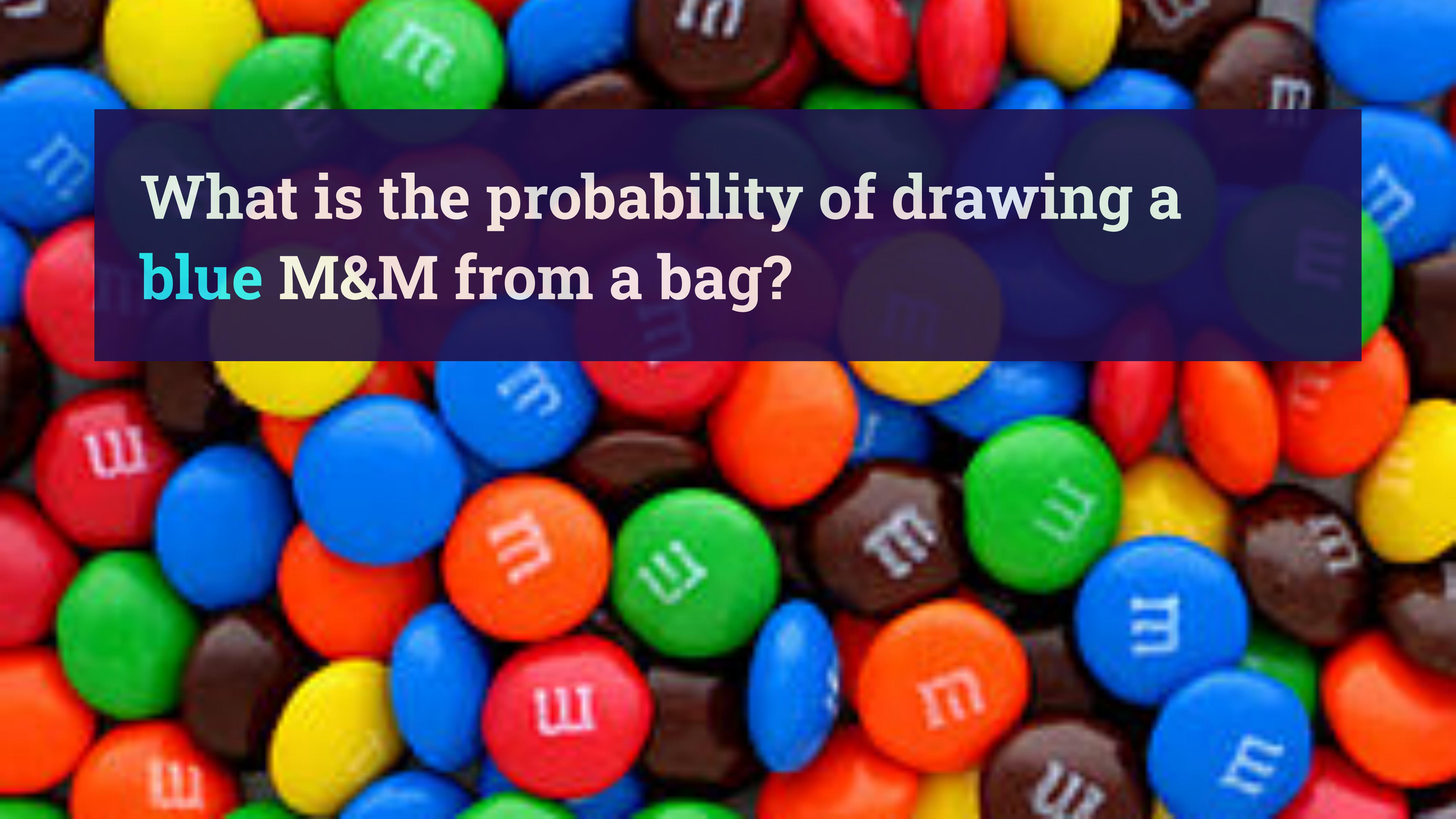
$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$

Bayes rule

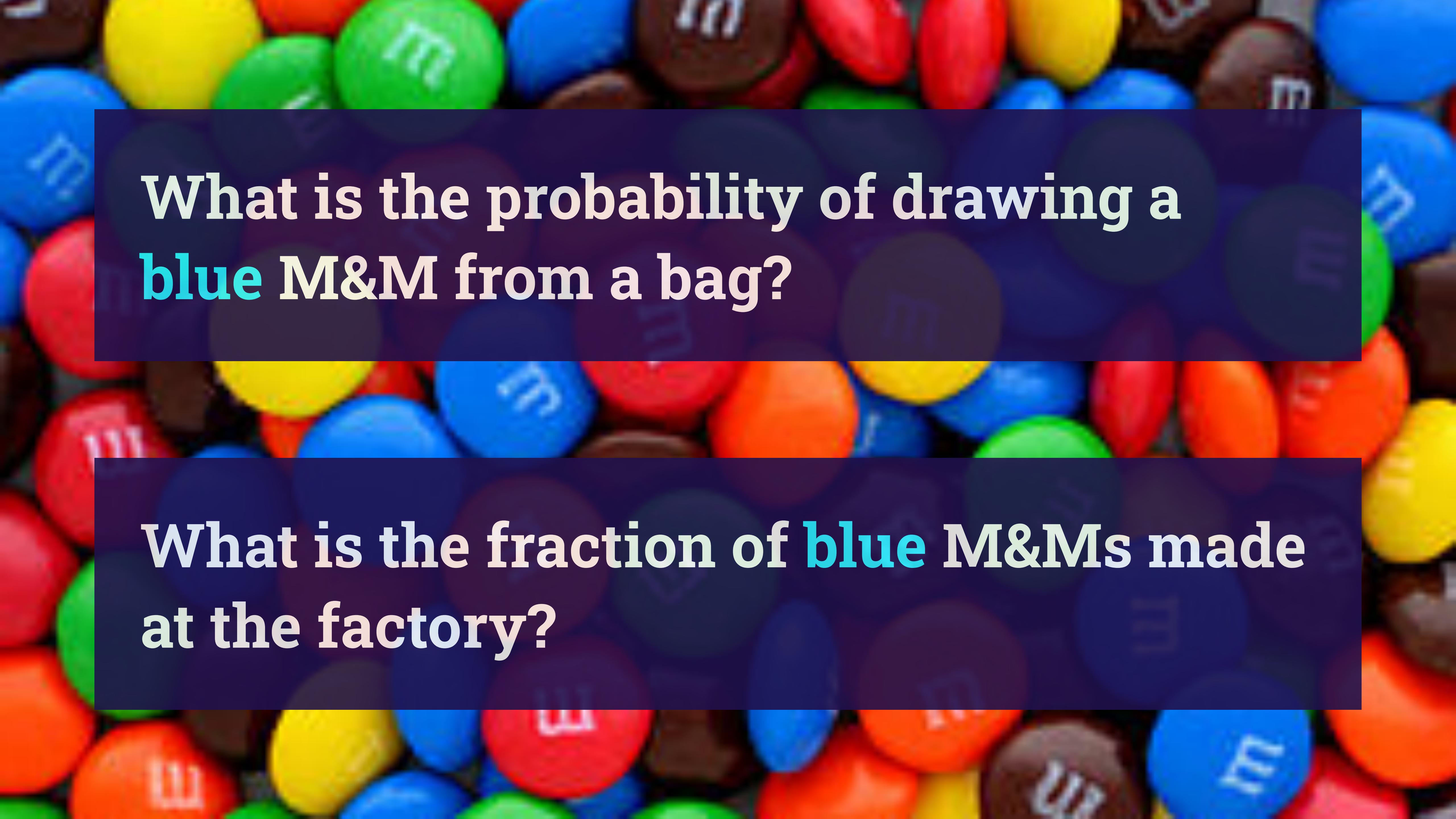
Important: $P(A | B) \neq P(B | A)$

$$p(\text{rainy} | \text{cloudy}) \neq p(\text{cloudy} | \text{rainy})$$





What is the probability of drawing a
blue M&M from a bag?



What is the probability of drawing a
blue M&M from a bag?

What is the fraction of **blue M&Ms** made
at the factory?



1. Blue M&Ms are the best.



1. Blue M&Ms are the best.

2. Why are there sometimes more blue M&Ms in a package and sometimes fewer?



1. Blue M&Ms are the best.

2. Why are there sometimes more blue M&Ms in a package and sometimes fewer?

3. What's the distribution of colours in a package of M&Ms? Are all colours equally distributed?



Why are there sometimes more blue smarties
in a package and sometimes fewer?



Why are there sometimes more blue smarties
in a package and sometimes fewer?



Why are there sometimes more blue smarties
in a package and sometimes fewer?



posterior

likelihood

prior

$$P(\theta | y) \propto P(y | \theta)P(\theta)$$

Find the posterior distribution for the percentage of **blue** m&ms made at the factory

likelihood

$$P(q \mid N, k) \propto P(k \mid N, q) P(q)$$

likelihood

$$P(q \mid N, k) \propto P(k \mid N, q) P(q)$$

$k =$

likelihood

$$P(q \mid N, k) \propto P(k \mid N, q) P(q)$$

$k =$ number of successes (blue m&ms)

likelihood

$$P(q \mid N, k) \propto P(k \mid N, q) P(q)$$

$k =$ number of successes (blue m&ms)

$N - k =$

likelihood

$$P(q \mid N, k) \propto P(k \mid N, q) P(q)$$

k = number of successes (blue m&ms)

$N - k$ = number of failures (not-blue m&ms)

likelihood

$$P(q \mid N, k) \propto P(k \mid N, q) P(q)$$

$k =$ number of successes (blue m&ms)

$N - k =$ number of failures (not-blue m&ms)

$q =$

likelihood

$$P(q \mid N, k) \propto P(k \mid N, q) P(q)$$

k = number of successes (blue m&ms)

$N - k$ = number of failures (not-blue m&ms)

q = probability of drawing a blue m&m

I have drawn the following sequence of blue (b) and not-blue (t) m&ms:

I have drawn the following sequence of blue (b) and not-blue (t) m&ms:

$S = b \ b \ t \ b \ t \ t \ t \ b \ t \ b$

I have drawn the following sequence of blue (b) and not-blue (t) m&ms:

$S = b \ b \ t \ b \ t \ t \ t \ b \ t \ b$

What's the probability of drawing exactly that sequence?

I have drawn the following sequence of blue (b) and not-blue (t) m&ms:

$$S = b \ b \ t \ b \ t \ t \ t \ b \ t \ b$$

What's the probability of drawing exactly that sequence?

$$p(S) = qq(1 - q)q(1 - q)(1 - q)(1 - q)q(1 - q)q$$

I have drawn the following sequence of blue (b) and not-blue (t) m&ms:

$$S = b \ b \ t \ b \ t \ t \ t \ b \ t \ b$$

What's the probability of drawing exactly that sequence?

$$\begin{aligned} p(S) &= qq(1 - q)q(1 - q)(1 - q)(1 - q)q(1 - q)q \\ &= q^5(1 - q)^5 \end{aligned}$$

The Binomial Distribution

Observed fraction of blue M&Ms in a single bag True fraction of blue M&Ms produced at the factory

$$p(\text{ } \text{ } | \text{ } \text{ }) = p(k | N, q) \sim \text{Binom}(N, k, q)$$



“The probability of observing k blue M&Ms out of N given that the true fraction of M&Ms is q ”

The Binomial Distribution

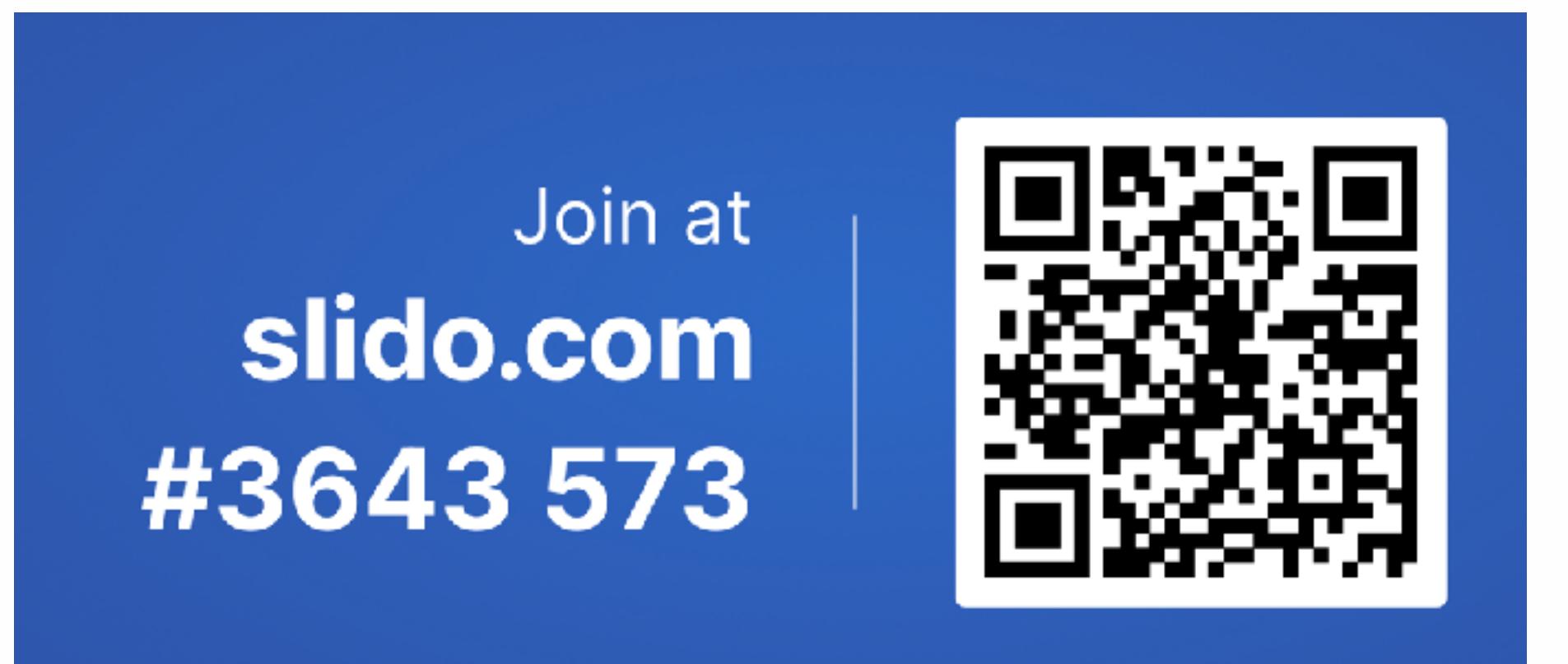
Observed fraction of blue M&Ms in a single bag True fraction of blue M&Ms produced at the factory

$$p(\text{ } \text{ } | \text{ } \text{ }) = p(k | N, q) \sim \text{Binom}(N, k, q)$$

$$= \binom{N}{k} q^k (1 - q)^{(N-k)}$$

“The probability of observing k blue M&Ms out of N given that the true fraction of M&Ms is q ”





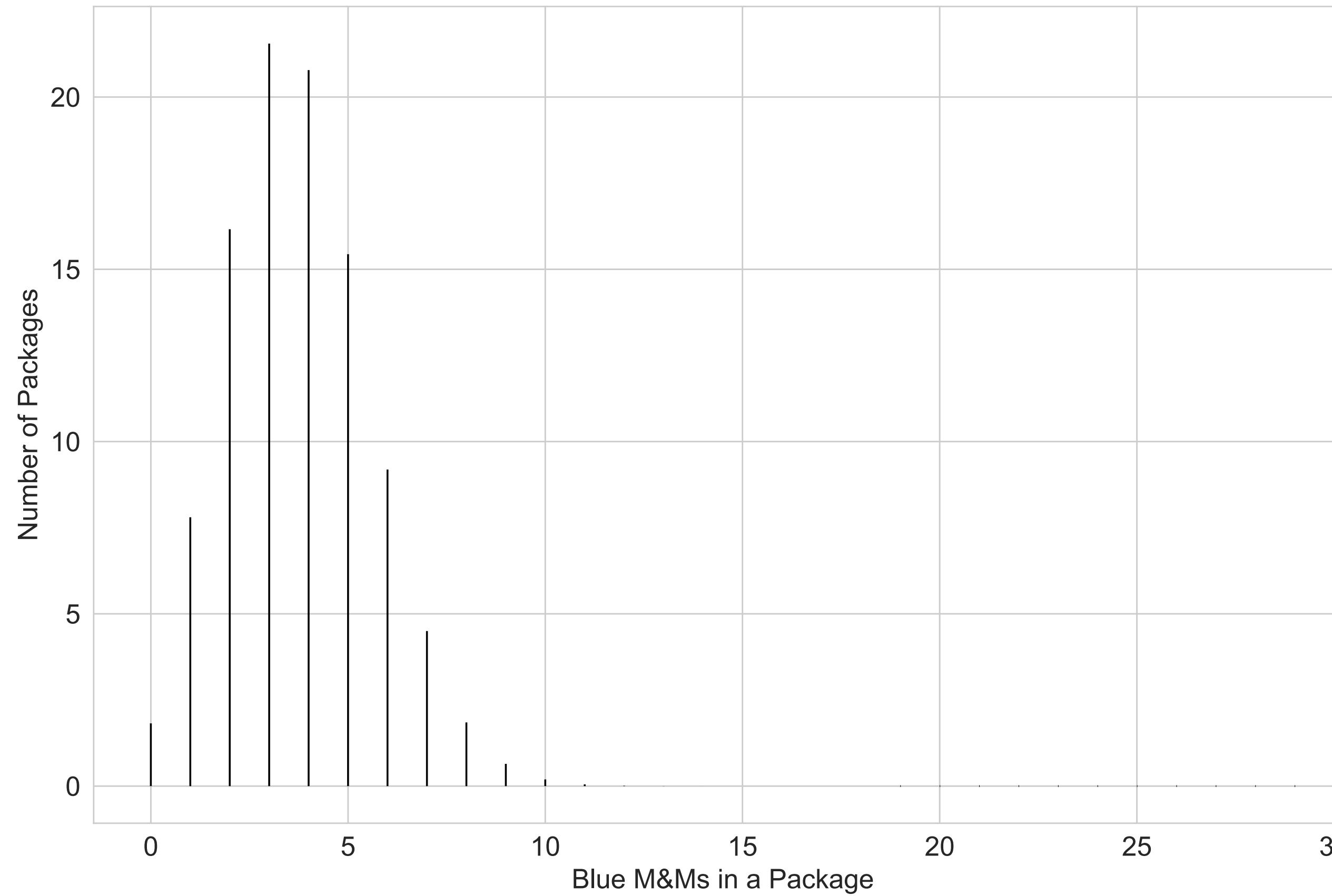
What do you think the fraction of blue M&Ms produced at the factory should be?

Join at
slido.com
#3643 573



First Assumption: The M&Ms are mixed together at the factory and then distributed into packages.

Second Assumption: All colours are equally distributed

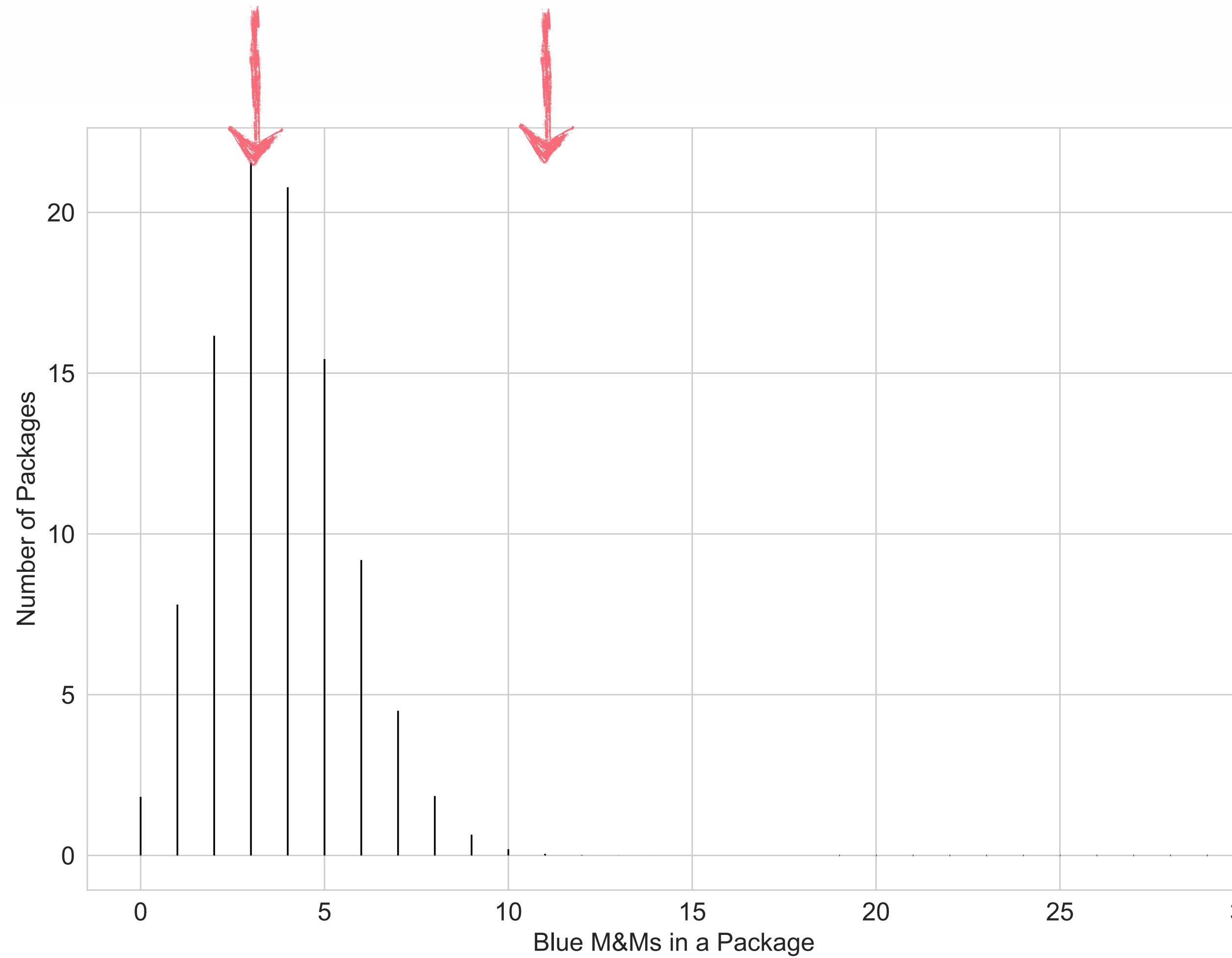


Binomial distribution

Blue Fraction: 0.125

First Assumption: The M&Ms are mixed together at the factory and then distributed into packages.

Second Assumption: All colours are equally distributed

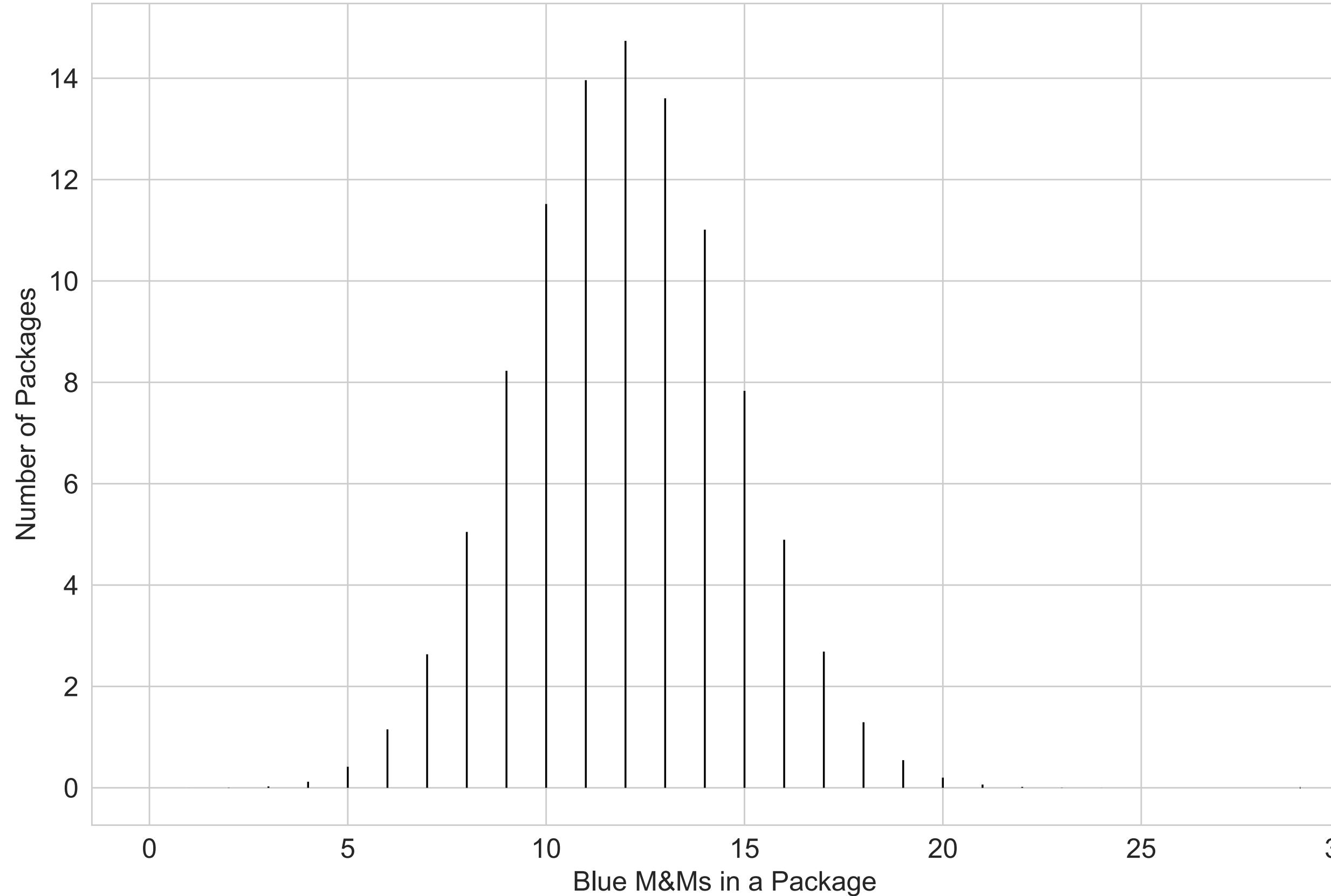


Binomial distribution

Blue Fraction: 0.125

First Assumption: The M&Ms are mixed together at the factory and then distributed into packages.

~~Second Assumption: All colours are equally distributed~~

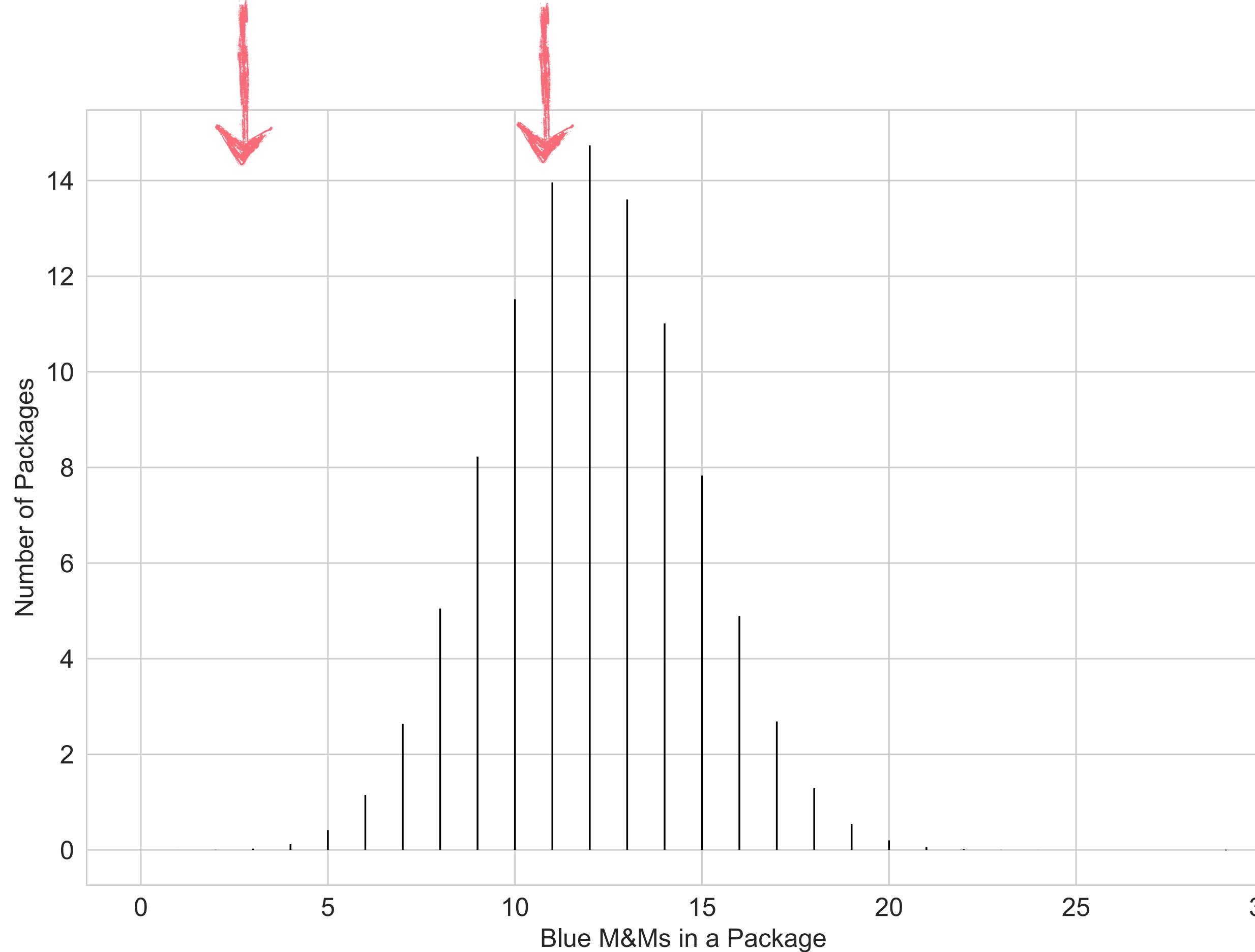


Binomial distribution

Blue Fraction: 0.4

First Assumption: The M&Ms are mixed together at the factory and then distributed into packages.

~~Second Assumption: All colours are equally distributed~~



Binomial distribution

Blue Fraction: 0.4



Question: Can we use this somehow
to estimate the fraction of blue
M&Ms?



Exercises 1, 2, 3

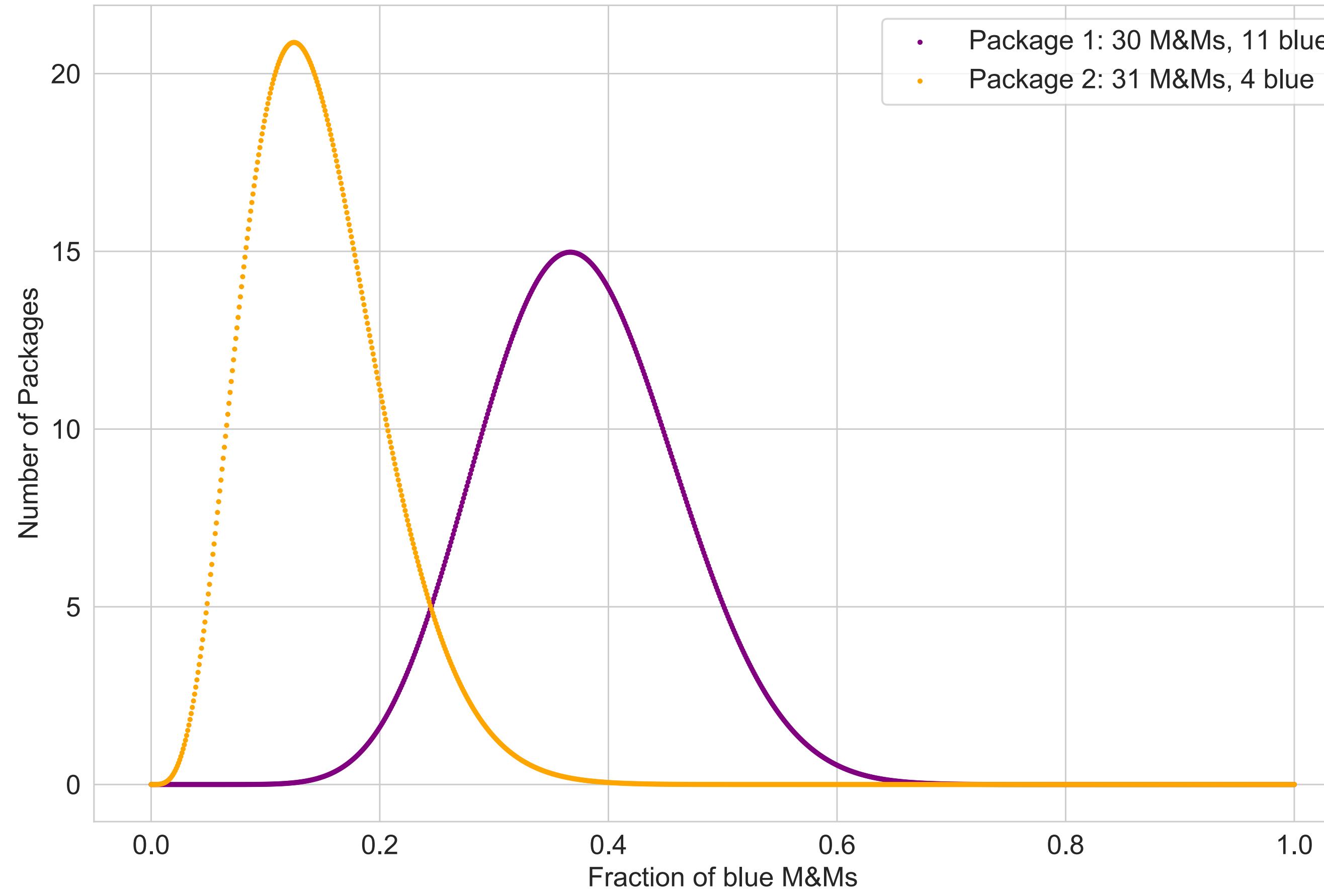


<https://bit.ly/ahw2022-bayestutorial>

Exercises 1, 2, 3

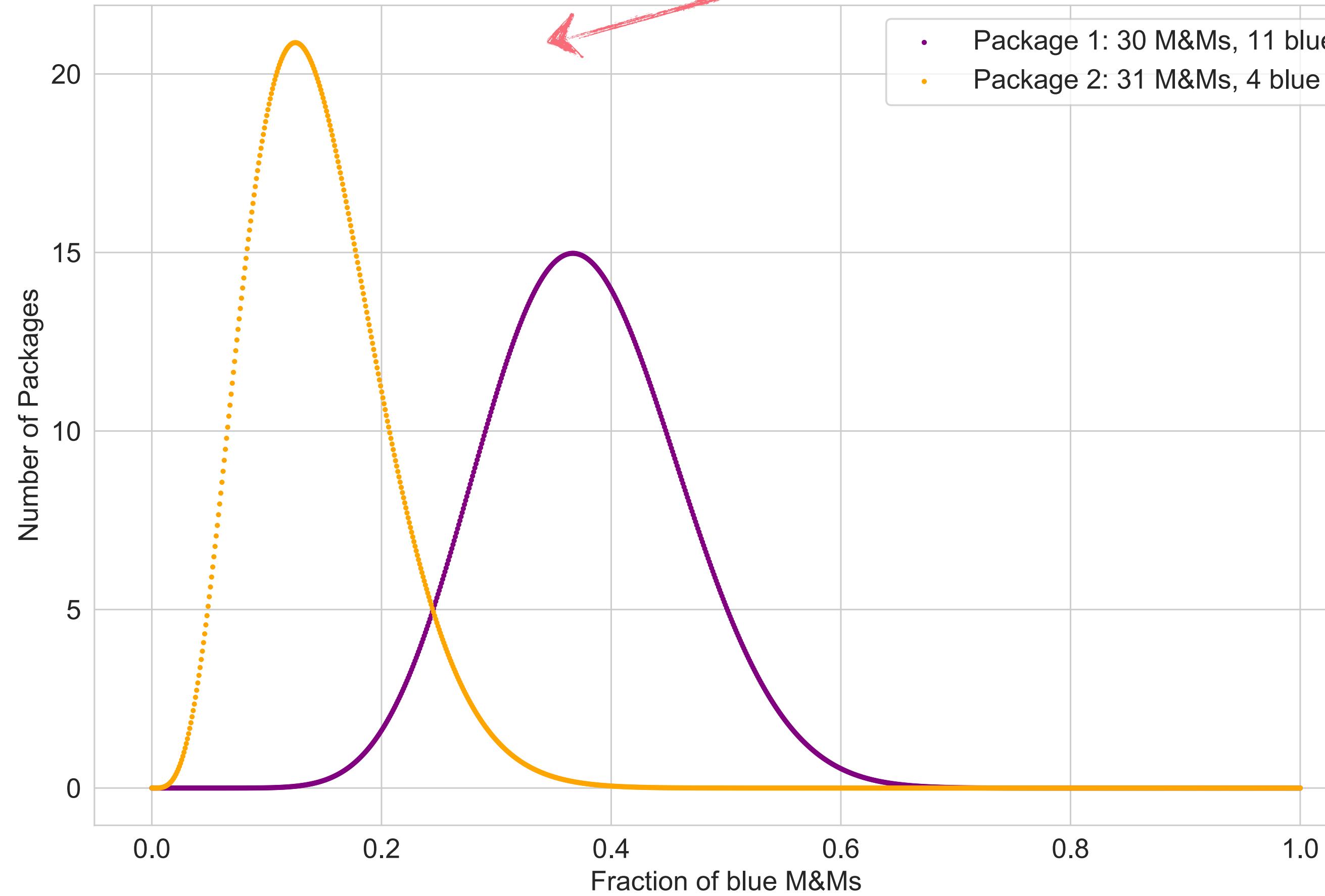


The Likelihood Function



The Likelihood Function

use optimization routines to
find the maximum



Problem: The Likelihood assumes we know
the true underlying fraction of M&Ms.

We're interested in $p(\text{工厂} | \text{糖果})$, not $p(\text{糖果} | \text{工厂})$

Conditional Probability Magic

$$p(\text{factory} \mid \text{candy}) = \frac{\text{likelihood} \quad \text{prior}}{\text{Evidence}}$$

posterior

 | 

 | 

 | 



Conditional Probability Magic

$$p(\text{factory} \mid \text{candy}) = \frac{\text{likelihood} \quad \text{prior}}{p(\text{candy})}$$

Evidence

$$p(\text{candy} \mid \text{factory}) \propto p(\text{candy} \mid \text{factory}) \ p(\text{factory})$$

Conditional Probability Magic

posterior

$$p(\text{factory} | \text{candy}) = \frac{\text{likelihood} \quad \text{prior}}{p(\text{candy})}$$

likelihood prior

$$p(\text{candy} | \text{factory}) p(\text{factory})$$

Evidence

$$p(\text{factory} | \text{candy}) \propto p(\text{candy} | \text{factory}) p(\text{factory})$$

Bayes theorem



Question

What prior information do we have about $p(\text{factory})$?

How many **different colours** of M&Ms are there?

What **percentage** of blue m&ms are made at the factory?

Do you think every bag of m&ms will have the **same percentage** of blue m&ms?

Sketch your prior



prior

$$P(\theta | y) \propto P(y | \theta) P(\theta)$$

prior

$$P(\theta | y) \propto P(y | \theta) P(\theta)$$

$$P(\theta) \propto \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

prior

$$P(\theta | y) \propto P(y | \theta) P(\theta)$$

beta distribution

$$P(\theta) \propto \theta^{\alpha-1} (1 - \theta)^{\beta-1}$$

prior

$$P(\theta | y) \propto P(y | \theta) P(\theta)$$

beta distribution

$$P(\theta) \propto \theta^{\alpha-1} (1 - \theta)^{\beta-1}$$

“conjugate
prior”

prior

$$P(\theta | y) \propto P(y | \theta) P(\theta)$$

beta distribution

$$P(\theta) \propto \theta^{\alpha-1} (1 - \theta)^{\beta-1}$$

“conjugate
prior”

$$\alpha = ?$$

prior

$$P(\theta | y) \propto P(y | \theta) P(\theta)$$

beta distribution

$$P(\theta) \propto \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

“conjugate
prior”

$$\alpha = ? \qquad \beta = ?$$

Write a function for the beta distribution and try out different values for alpha and beta

Some values to try:

$$\alpha = \beta$$

$$\alpha = \beta = 1$$

α, β very large

α, β very small

Write a function for the beta distribution and try out different values for alpha and beta

Some values to try:

$$\alpha = \beta$$

$$\alpha = \beta = 1$$

α, β very large

α, β very small

$$P(\theta \,|\, y) \propto P(y \,|\, \theta)P(\theta)$$

likelihood

$$p(y) \propto \theta^y(1 - \theta)^{n-y}$$

$$P(\theta | y) \propto P(y | \theta)P(\theta)$$

likelihood

$$p(y) \propto \theta^y(1 - \theta)^{n-y}$$

prior

$$P(\theta) \propto \theta^{\alpha-1}(1 - \theta)^{\beta-1}$$

$$P(\theta | y) \propto P(y | \theta)P(\theta)$$

likelihood

$$p(y) \propto \theta^y(1 - \theta)^{n-y}$$

prior

$$P(\theta) \propto \theta^{\alpha-1}(1 - \theta)^{\beta-1}$$

$$P(\theta | y) \propto P(y | \theta)P(\theta)$$

$$p(\theta) \propto \theta^{y+\alpha-1}(1 - \theta)^{n-y+\beta-1}$$

Gather some data!

In groups of 3, pool your M&Ms

record the counts for all colours



Use your function for the beta distribution to plot both the prior and the posterior in the same plot

- Is the posterior distribution what you expected?
- Compare the posterior distribution to the prior distribution: Is this the result you expected, given six different colours?
- How sensitive is the posterior to the prior?

Use your function for the beta distribution to plot both the prior and the posterior in the same plot

- Is the posterior distribution what you expected?
- Compare the posterior distribution to the prior distribution: Is this the result you expected, given six different colours?
- How sensitive is the posterior to the prior?

<https://bit.ly/ahw2022-bayestutorial>
Exercise 7+8

How would you expect the posterior to change with **more data**?

Let's **pool all the data** and find out!

<https://forms.gle/zHy8LSbJsB8Pm9XA9>

How would you expect the posterior to change with **more data**?

Let's **pool all the data** and find out!

<https://forms.gle/zHy8LSbJsB8Pm9XA9>

<https://bit.ly/ahw2022-bayestutorial>
Exercise 9

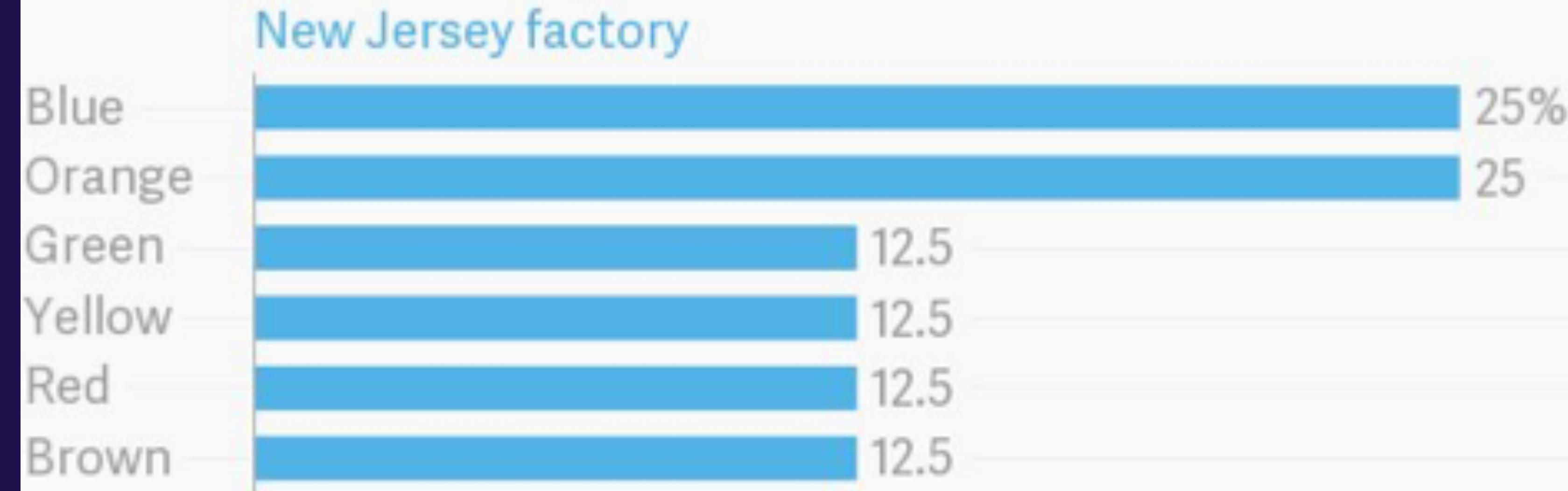
Priors are useful when data isn't very informative.

When the data is informative, the choice of prior doesn't matter much.

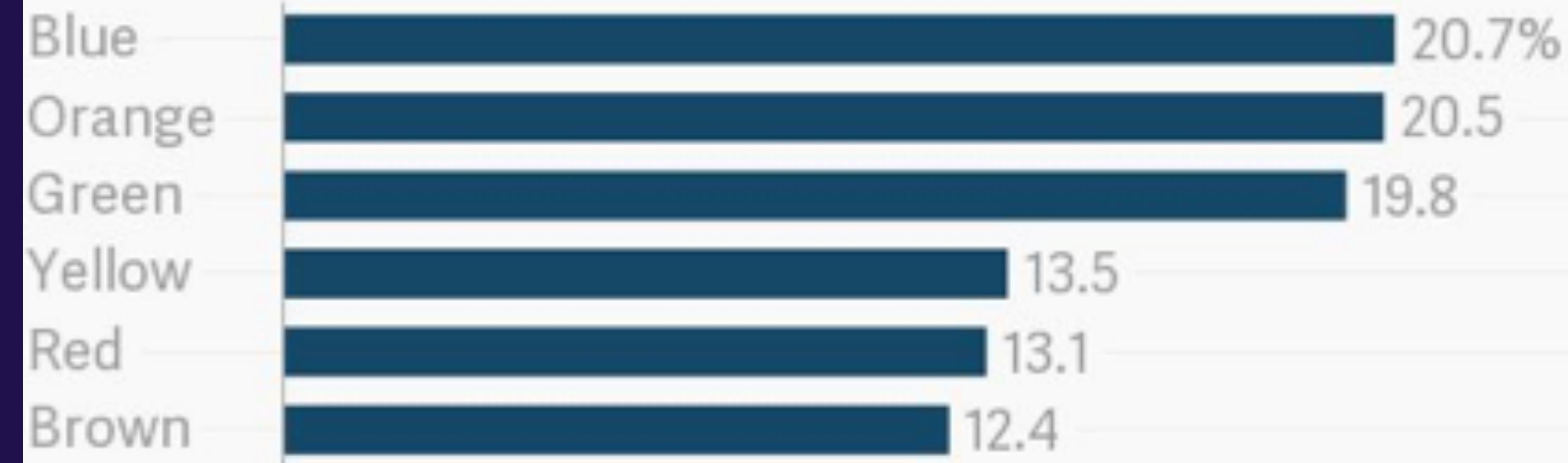
Surprise Twist!

Different factories make different distributions of m&ms!

M&Ms color distribution, c. 2017



Tennessee factory

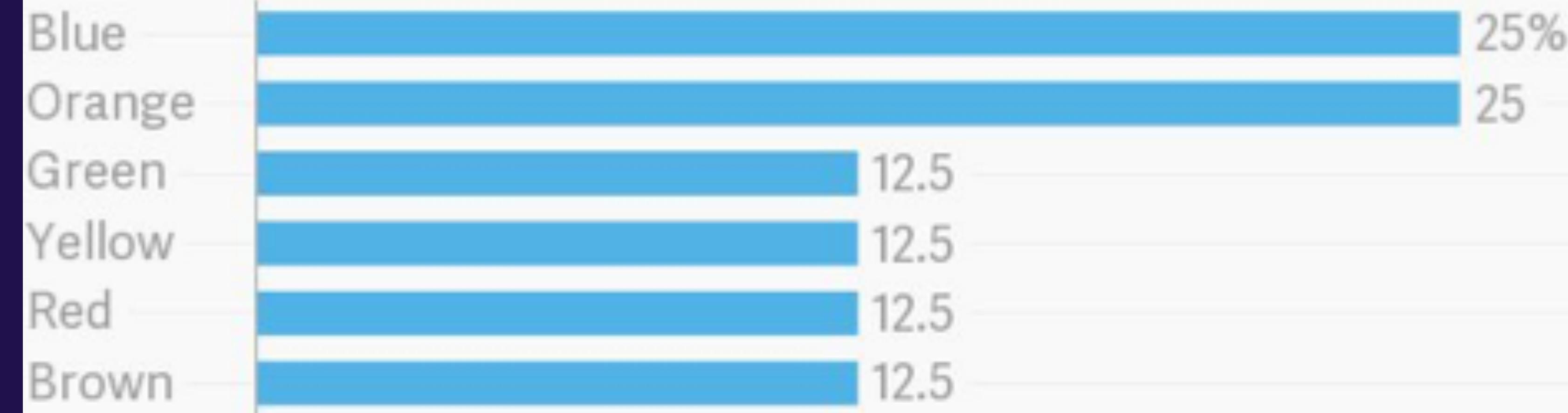


Surprise Twist!

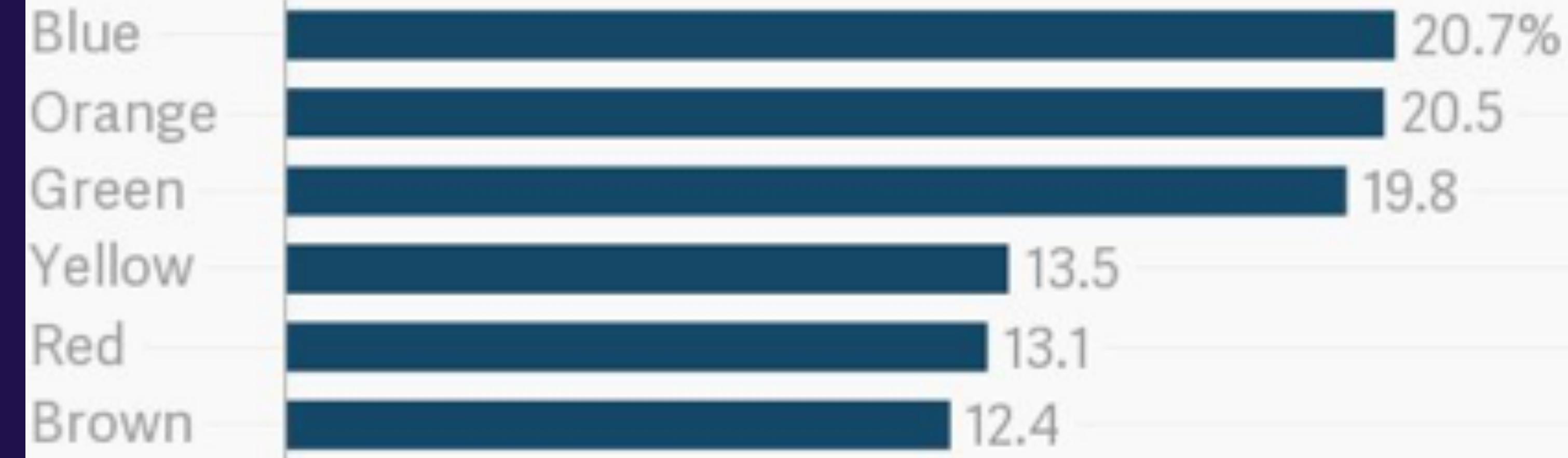
Which factory
did your m&ms
come from?

M&Ms color distribution, c. 2017

New Jersey factory



Tennessee factory



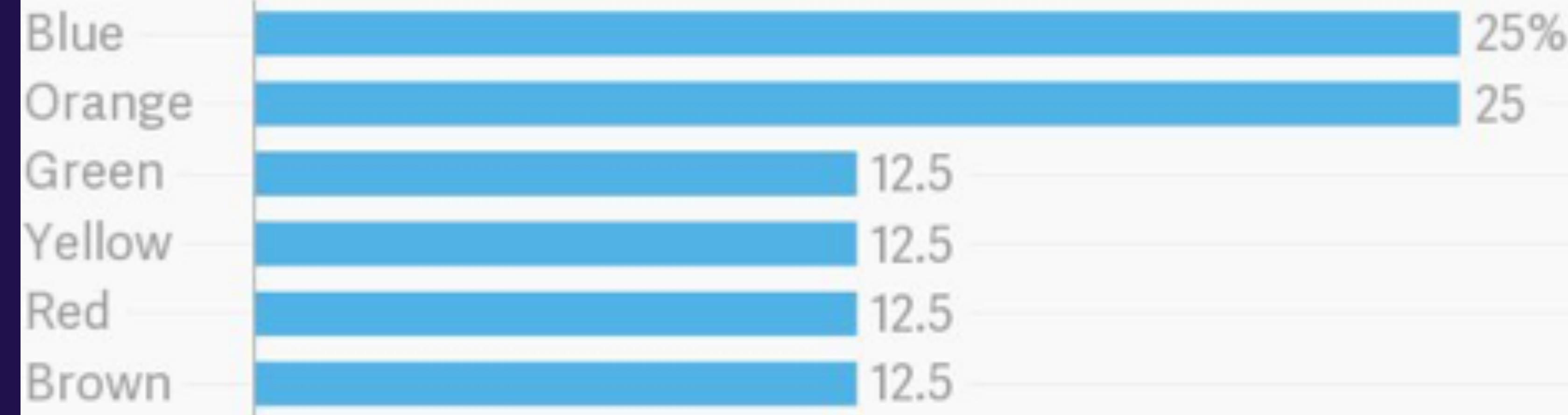
**Surprise
Twist!**

**New Jersey =
HKL**

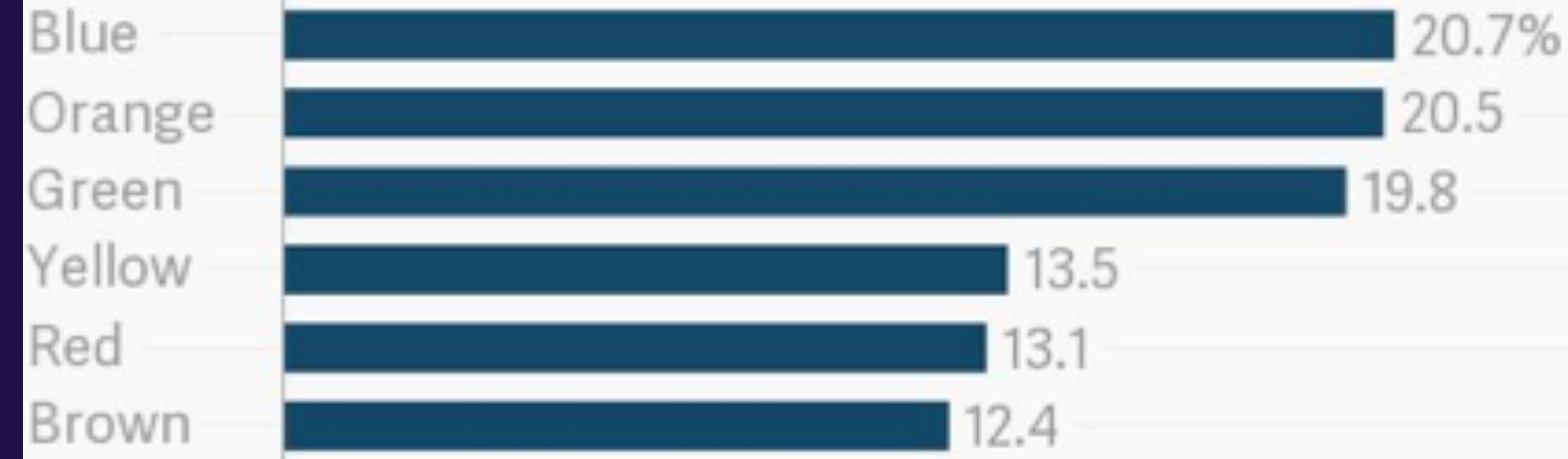
**Tennessee =
CLV**

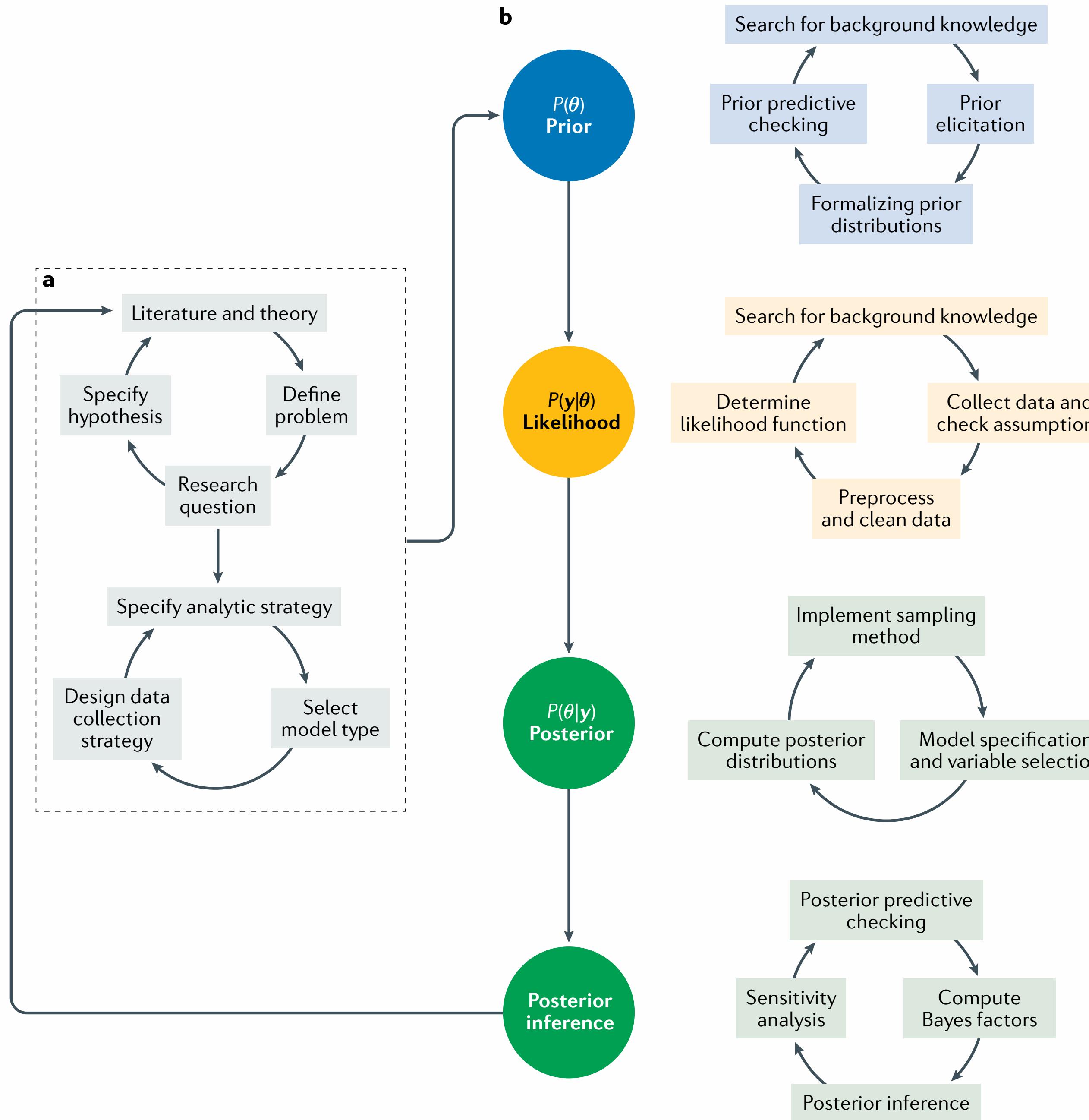
M&Ms color distribution, c. 2017

New Jersey factory



Tennessee factory



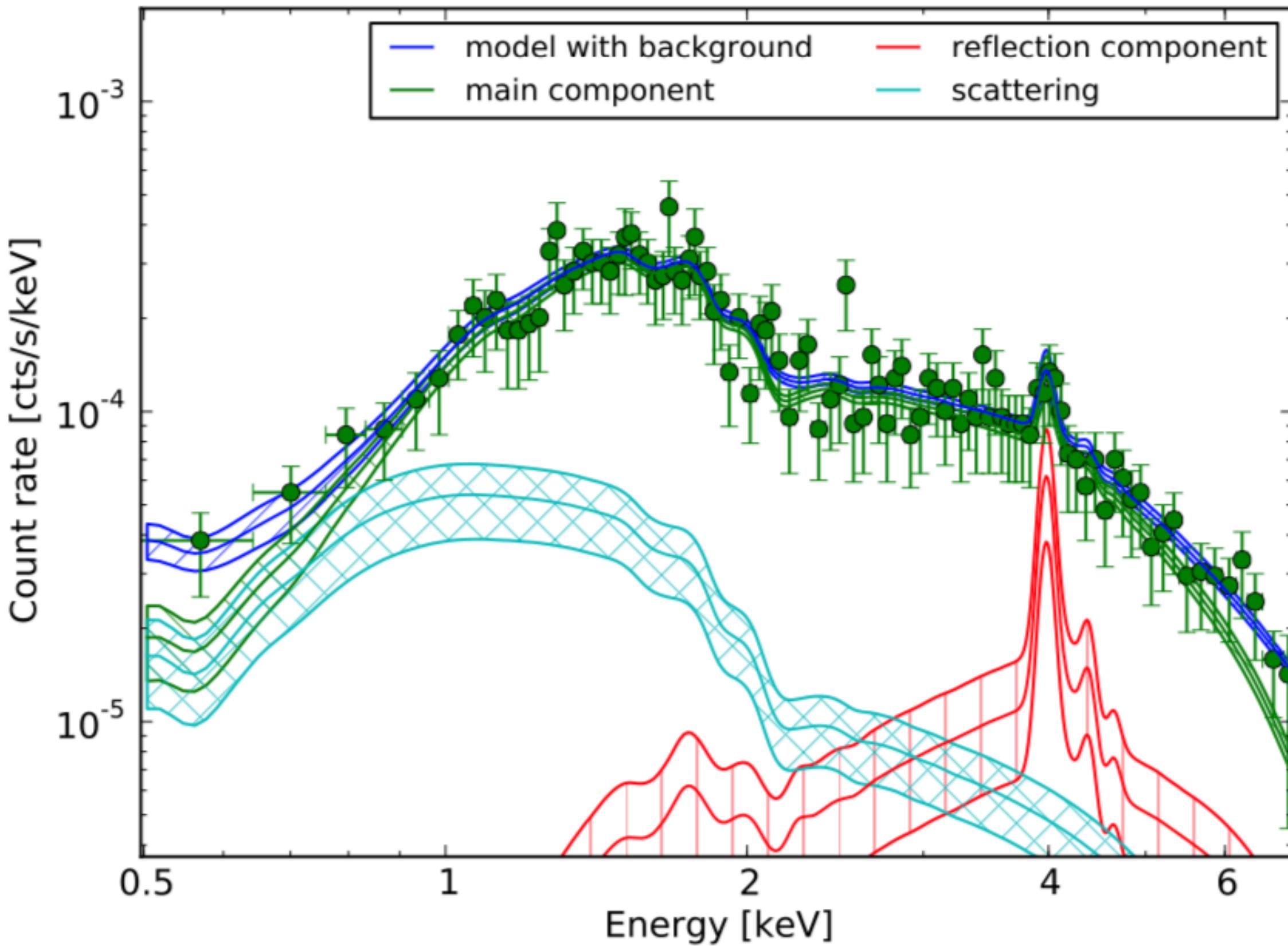


Have other people calculated the distribution of colours? What can we reason about $p(\text{factory} | \text{color})$?

Assumption that all colours are mixed together at the factory and then distributed into packages: **binomial likelihood**

What's the probability of a certain fraction of blue smarties given our observations, $p(\text{factory} | \text{smarties})$?

What about astronomy?



Data:

Model:

Likelihood: Poisson

X-ray spectrum
absorption +
scattering + reflection

What if we don't have conjugate priors?

Markov Chain Monte Carlo*

- * pick a parameter vector θ_1
- * Compute the posterior $p(\theta_1 | D)$
- * Pick another parameter vector θ_2 according to $p(\theta_2 | \theta_1)$
- * Compute the posterior $p(\theta_2 | D)$
- * Compute $p_{\text{accept}} = p(\theta_2 | D) / p(\theta_1 | D)$
- * Draw a random, uniform number r
- * Keep θ_2 if $p_{\text{accept}} > r$, otherwise keep θ_2 with probability r

* Metropolis-Hastings Algorithm

Ethical considerations in statistics

data and algorithms are social constructs

data and algorithms are social constructs

... are created by humans

data and algorithms are social constructs

... are created by humans

... encode biases

data and algorithms are social constructs

... are created by humans

... encode biases

**... can be (mis-)used to
serve an agenda**

Based on the M&Ms exercise, can you think of ways our statistical procedure could be mis-used or be made to be misleading?

Opinion: Is science really facing a reproducibility crisis, and do we need it to?

Daniele Fanelli

PNAS March 13, 2018 115 (11) 2628-2631; first published March 12, 2018
<https://doi.org/10.1073/pnas.1708272114>

Edited by David B. Allison, Indiana University Bloomington, Bloomington, IN, and accepted by Editorial Board Member Susan T. Fiske November 3, 2017 (received for review June 30, 2017)

Article

Figures & SI

Info & Metrics

PDF

Abstract

Efforts to improve the reproducibility and integrity of science are typically justified by a narrative of crisis, according to which most published results are unreliable due to growing problems with research and publication practices. This article provides an overview of recent evidence suggesting that this narrative is mistaken, and argues that a narrative of epochal changes and empowerment of scientists would be more accurate, inspiring, and compelling.



nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

Archive > Volume 519 > Issue 7541 > Research Highlights: Social Selection > Article

NATURE | RESEARCH HIGHLIGHTS: SOCIAL SELECTION

Psychology journal bans *P* values

Test for reliability of results ‘too easy to pass’, say editors.

Chris Woolston

26 February 2015 | Clarified: 09 March 2015

PDF

Rights & Permissions

A controversial statistical test has finally met its end, at least in one journal. Earlier this month, the editors of *Basic and Applied Social Psychology* (BASP) announced that the journal would no longer publish papers containing *P* values because the statistics were too often used to support lower-quality research.¹

Most mistakes made in statistical procedures are either **mis-applications** of methods or **mis-interpretation** of the results

Questions?