BOOTSTRAP*

HOW TO EVALUATE ACCURACY OF **THINGS**** ABOUT LOTS OF **STUFF*****
WHEN ONLY HAVING **NOT SO MUCH STUFF******

* A statistical algorithm extremely useful when only a small set of data is available. It gives a precise idea of the accuracy of the estimate

** Estimate of statistical parameters
*** A huge population
**** A small sample

START

We want to estimate the mean height of all the blueberry-loving people in Milan, but we only know the height of 10 of them. **How do we get there?**

STEP 01 SAMPLE

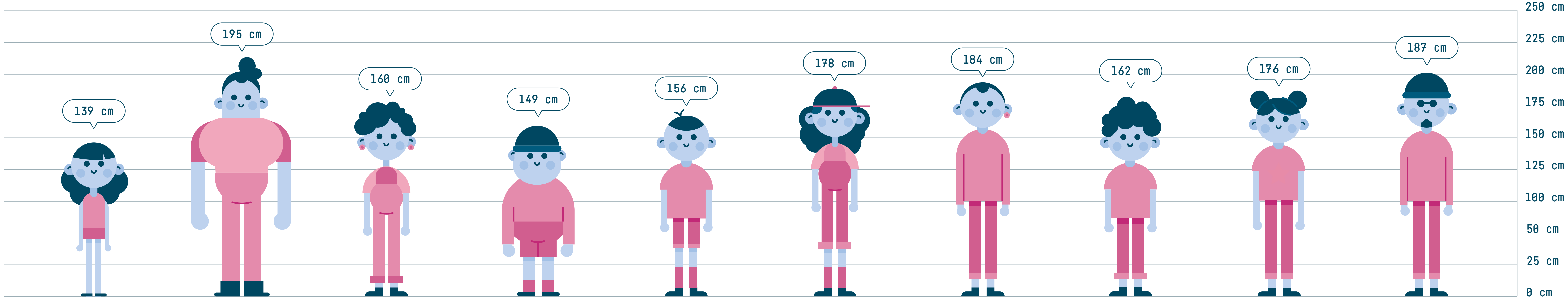
WE TAKE THE DATA FROM THE SAMPLE

As we all know, blueberry-loving people are very *very* shy, so we only managed to tackle 10 of them for our super-duper important research about the link between height and blueberryness. We'll call them our **sample**.

The bigger the obtainable sample, the better the estimate. In our case 10 people will do just fine.

STARTING SAMPLE

▼ **Law of Large Numbers**
As the sample size increases, its ability to give an accurate representation of the whole population increases too.



STEP 02 RESAMPLE

WE RESAMPLE OUR DATA AND CALCULATE THE MEAN *MANY MANY MANY* TIMES

We could just calculate the mean of the starting sample, but **we wouldn't have any information about the accuracy of the estimate**.

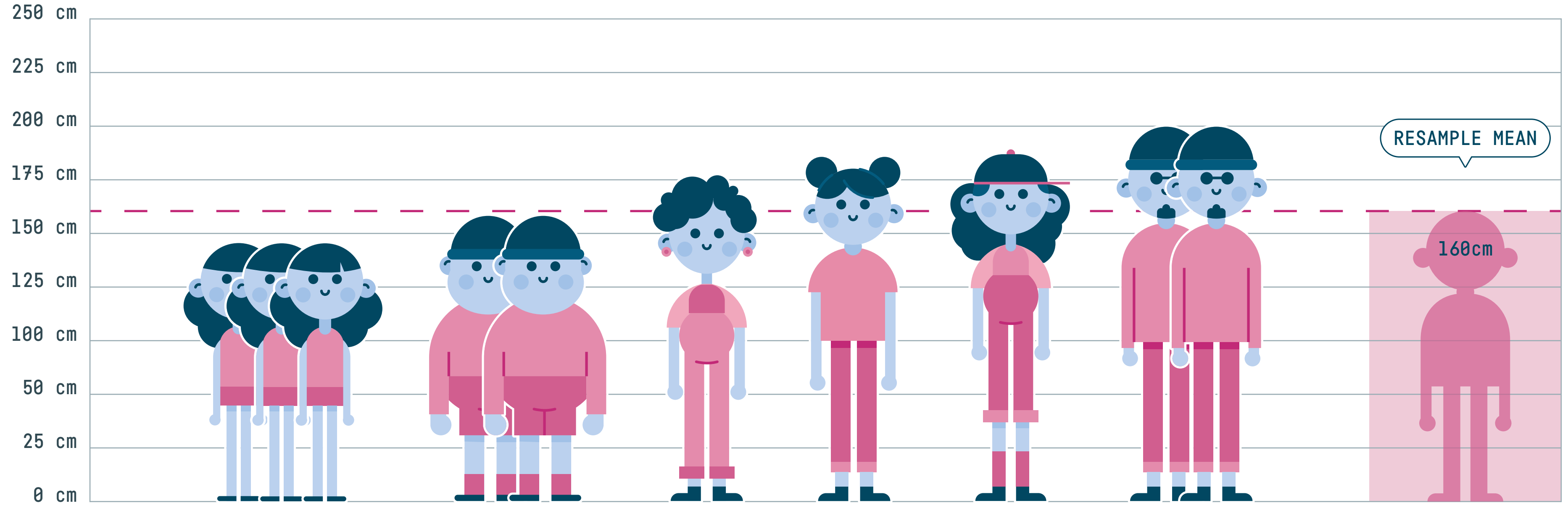
Instead, we **resample with replacement** many (*many!*) times: each resample is made of 10 height values, randomly picked from the original 10.

We calculate the **mean height value of each resample**, and we store them away for later. (in case we're hungry).

RESAMPLE #0001

◀ **Recommended number of resamples**
Ranges from 1K to 100K.

▼ **Resampling with replacement**
Replacement means that repetition is allowed, each value can be picked more than once or not picked at all.

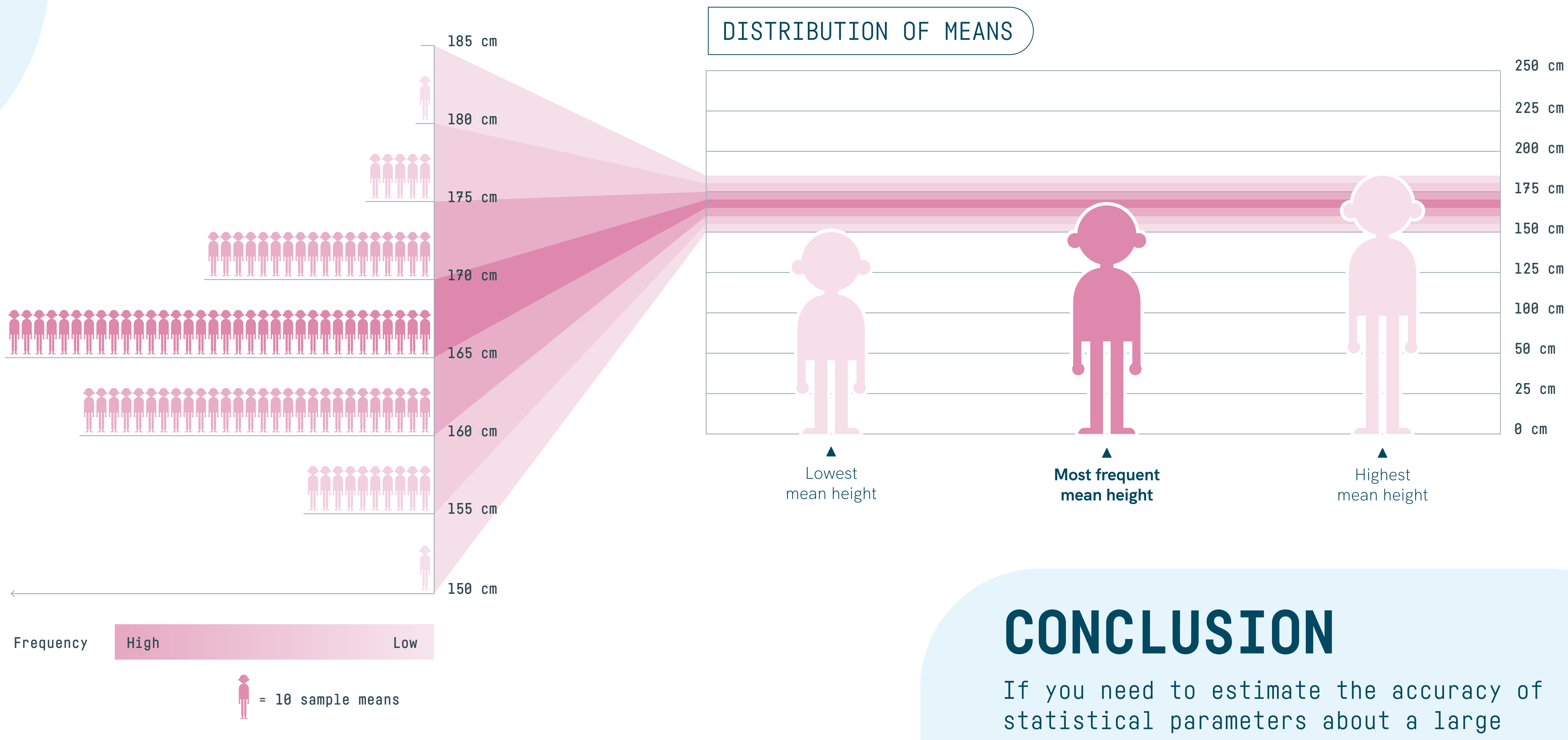


STEP 03 DISTRIBUTION

WE DRAW THE DISTRIBUTION OF THE MEAN HEIGHT VALUES

The distribution shows the **estimated mean height** and the **accuracy of the estimate**. The frequency is expected to be higher for values near the **real population mean**, while lower for further values.

DISTRIBUTION OF MEANS



CONCLUSION

If you need to estimate the accuracy of statistical parameters about a large population, but only have a small sample, the bootstrap algorithm is the way to go.

CAUTION! overconsumption of blueberries might result in your skin turning blue.

VISUAL EXPLANATIONS OF STATISTICAL METHODS

Bootstrap

AUTHORS

Daniele Dell'Orto
Martina Francella
Octavian Husoschi
Martina Melillo
Matteo Pini
Alessandro Quets
Shan Huang

FACULTY

Michele Mauri
Ángeles Briones
Gabriele Colombo
Simone Vantini
Salvatore Zingale

TEACHING ASSISTANTS

Elena Aversa
Andrea Benedetti
Tommaso Elli
Beatrice Gobbo
Anna Riboldi

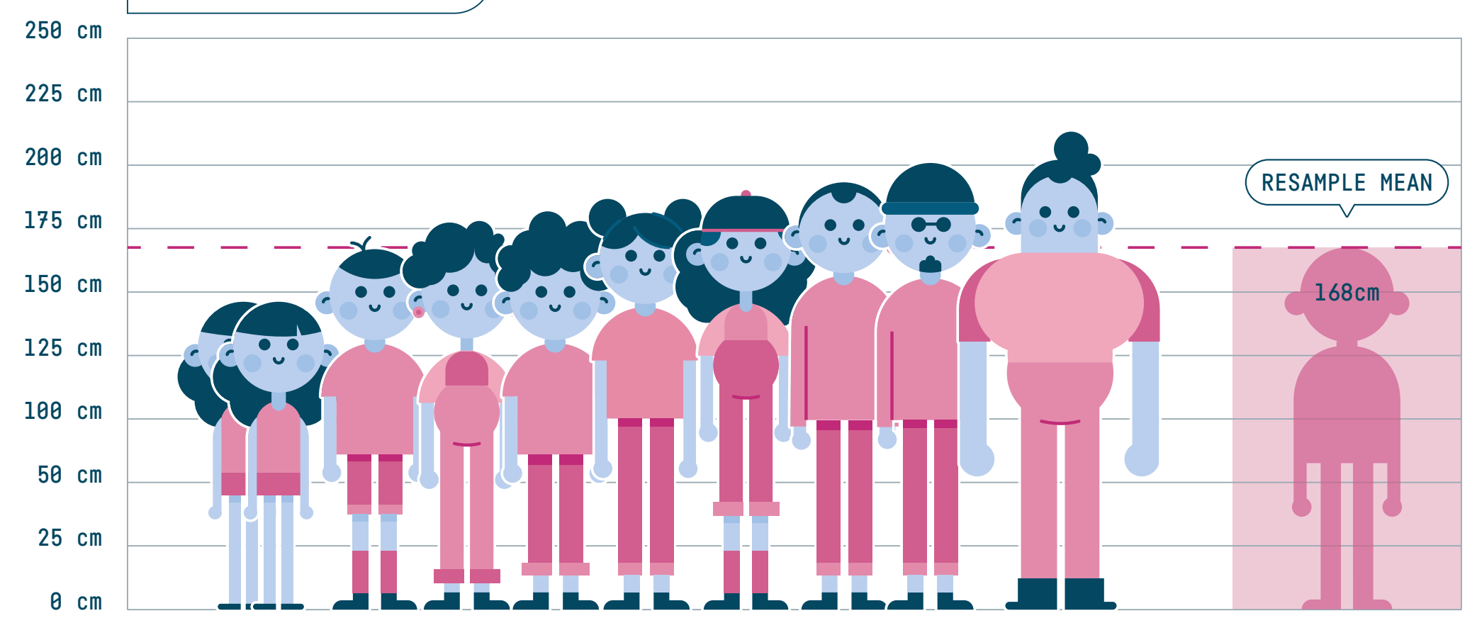
Final Synthesis Design Studio
Sect. C3

LM in Communication Design
A.A. 2021/2022

DEN-
SITY
GN+

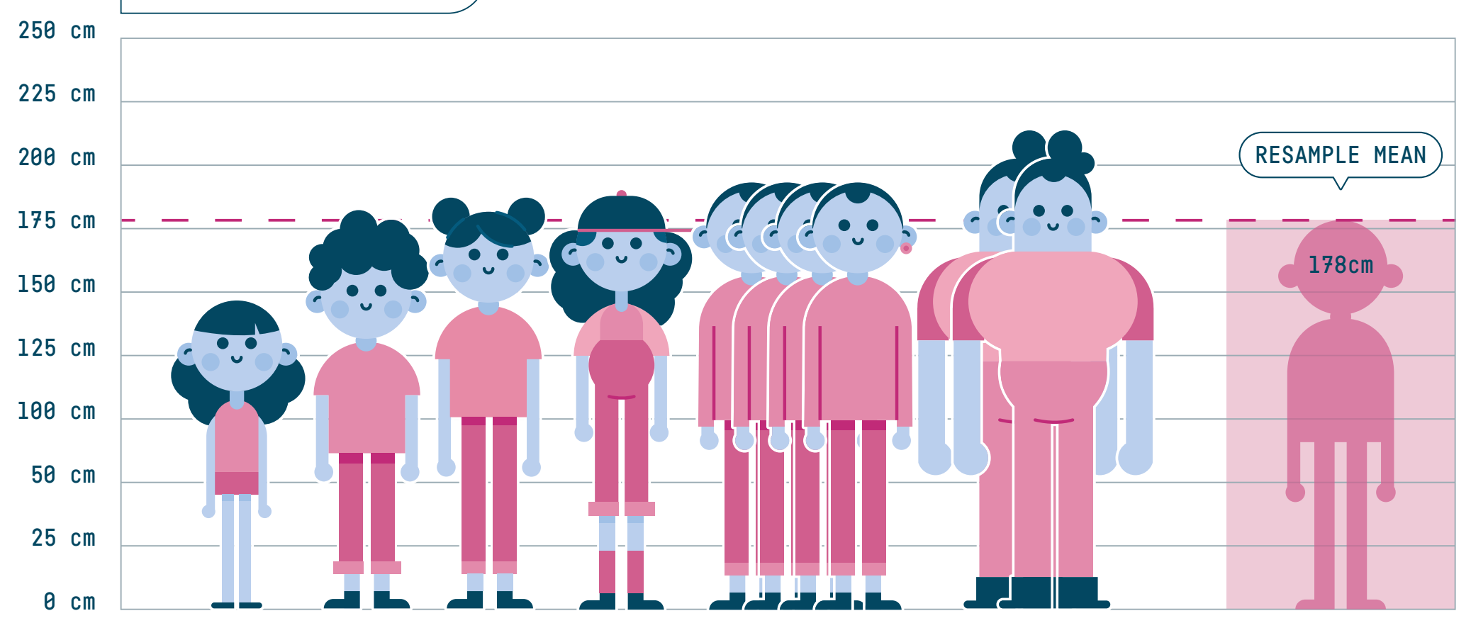
POLITECNICO
MILANO 1863
SCHOOL OF DESIGN

RESAMPLE #N

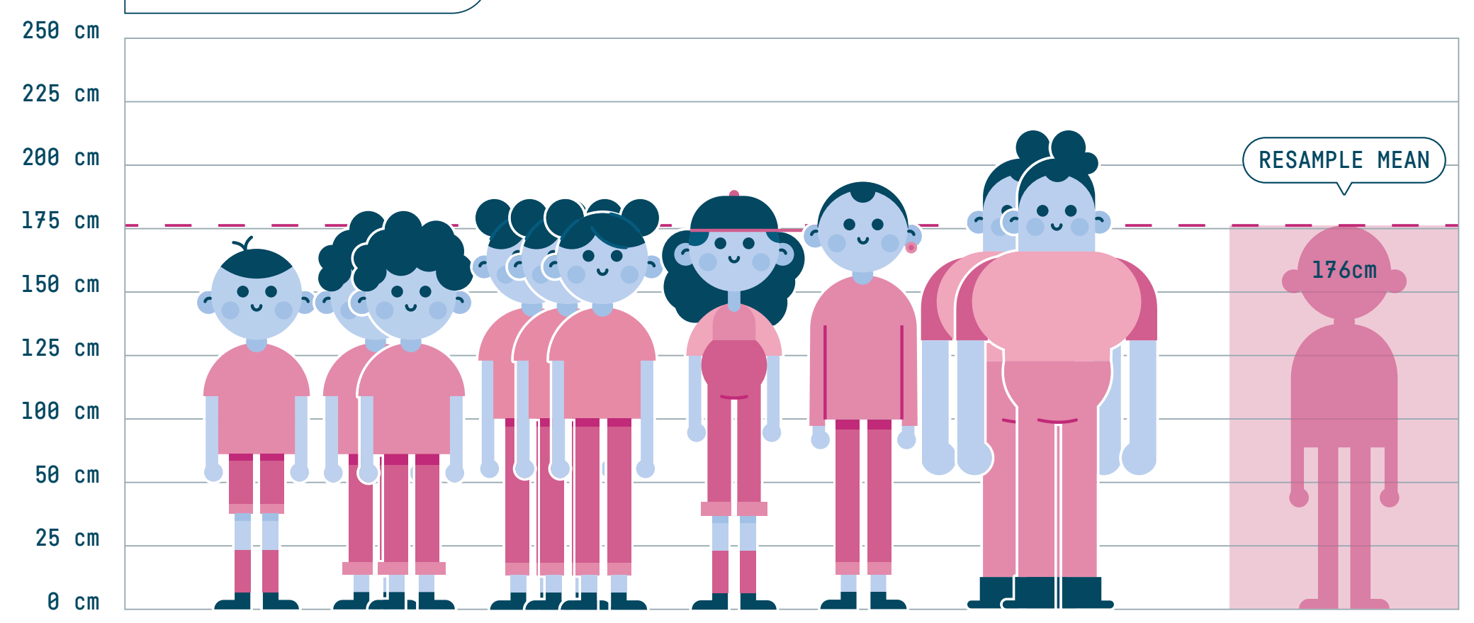


[...]

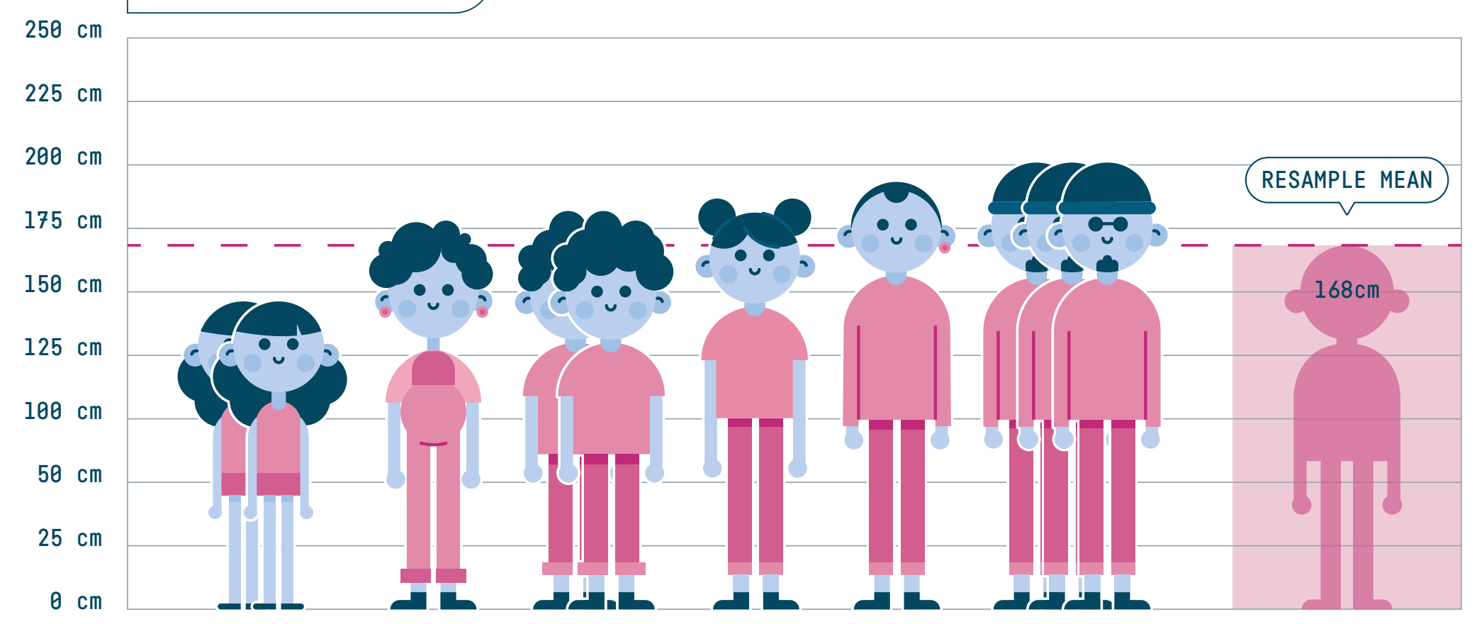
RESAMPLE #0019



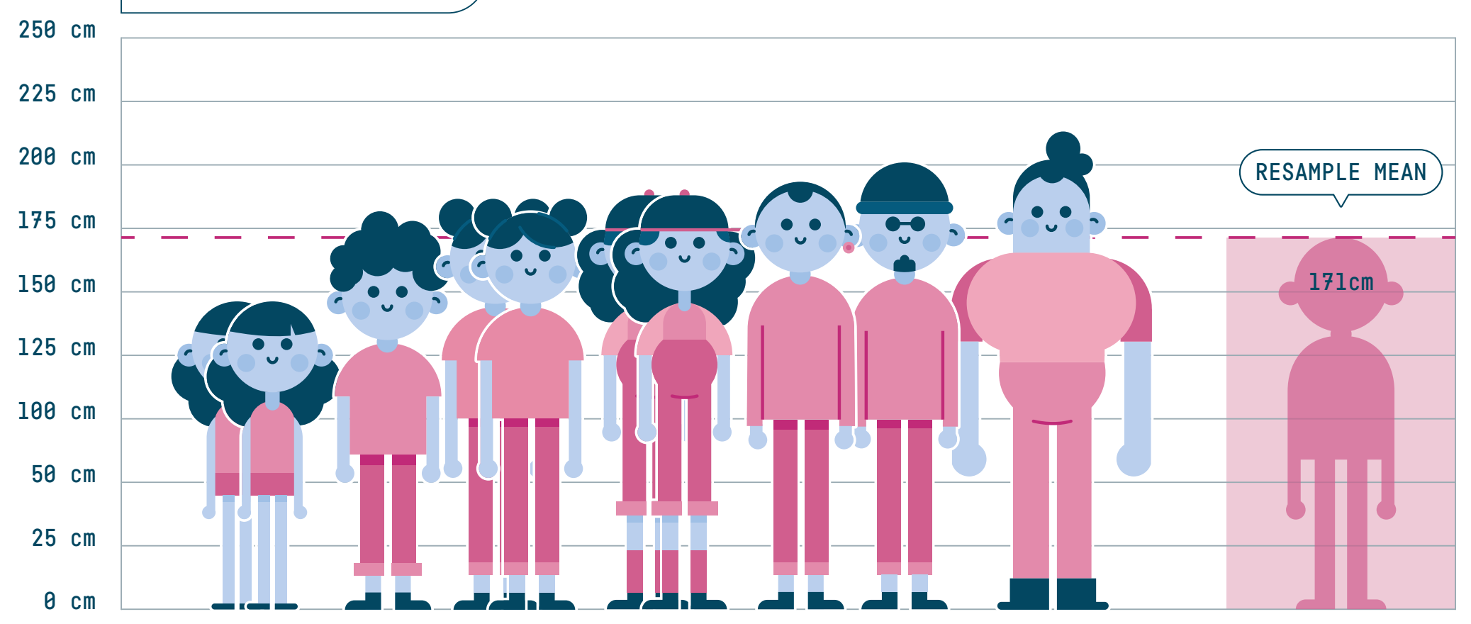
RESAMPLE #0018



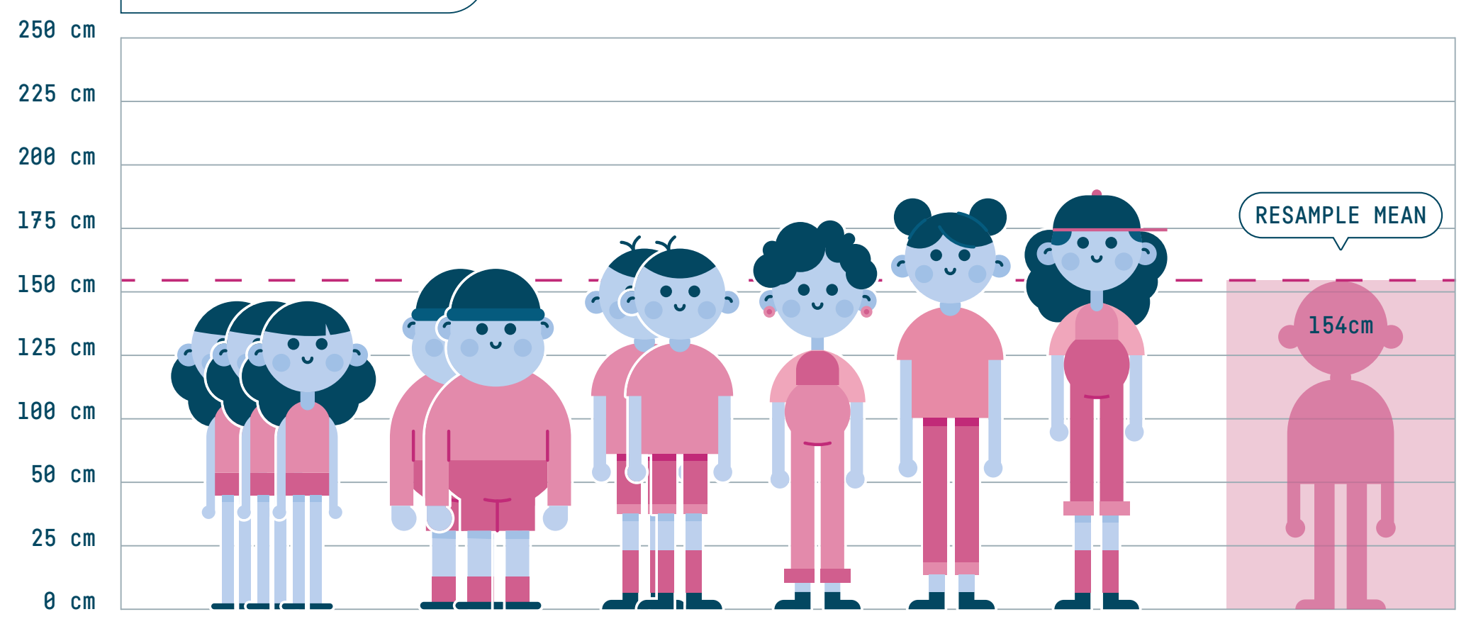
RESAMPLE #0017



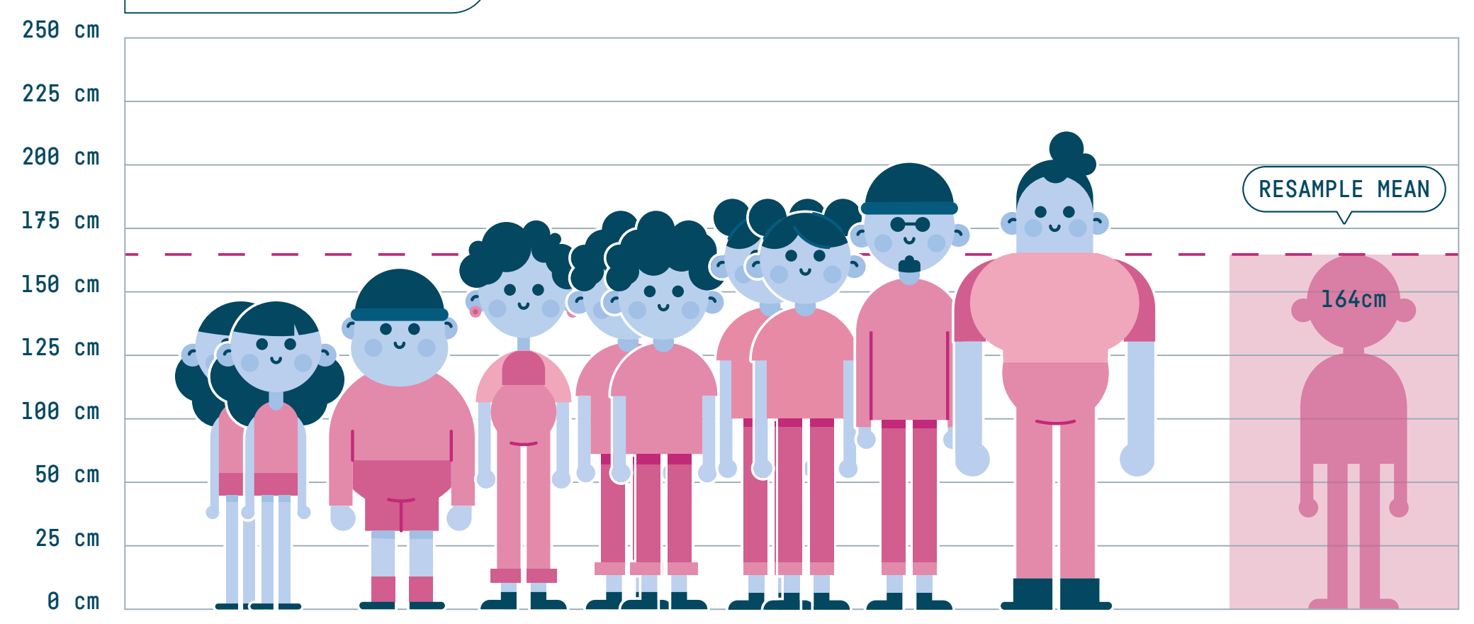
RESAMPLE #0016



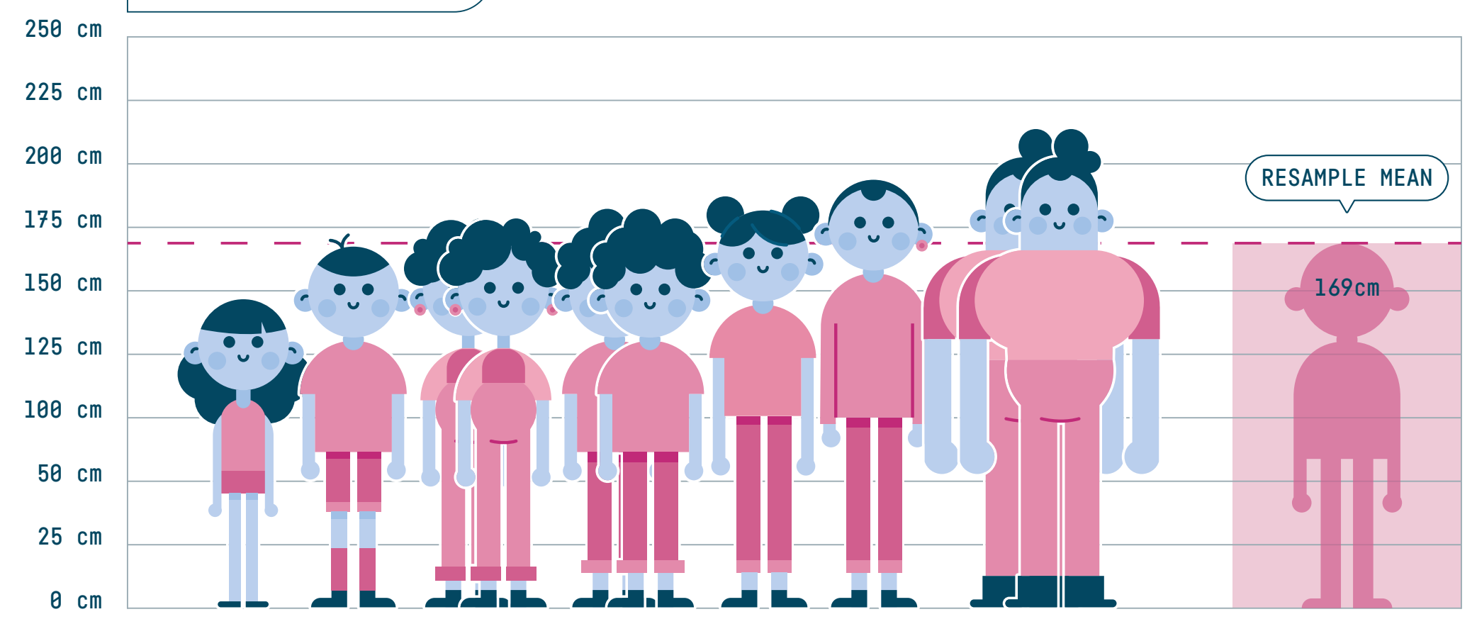
RESAMPLE #0015



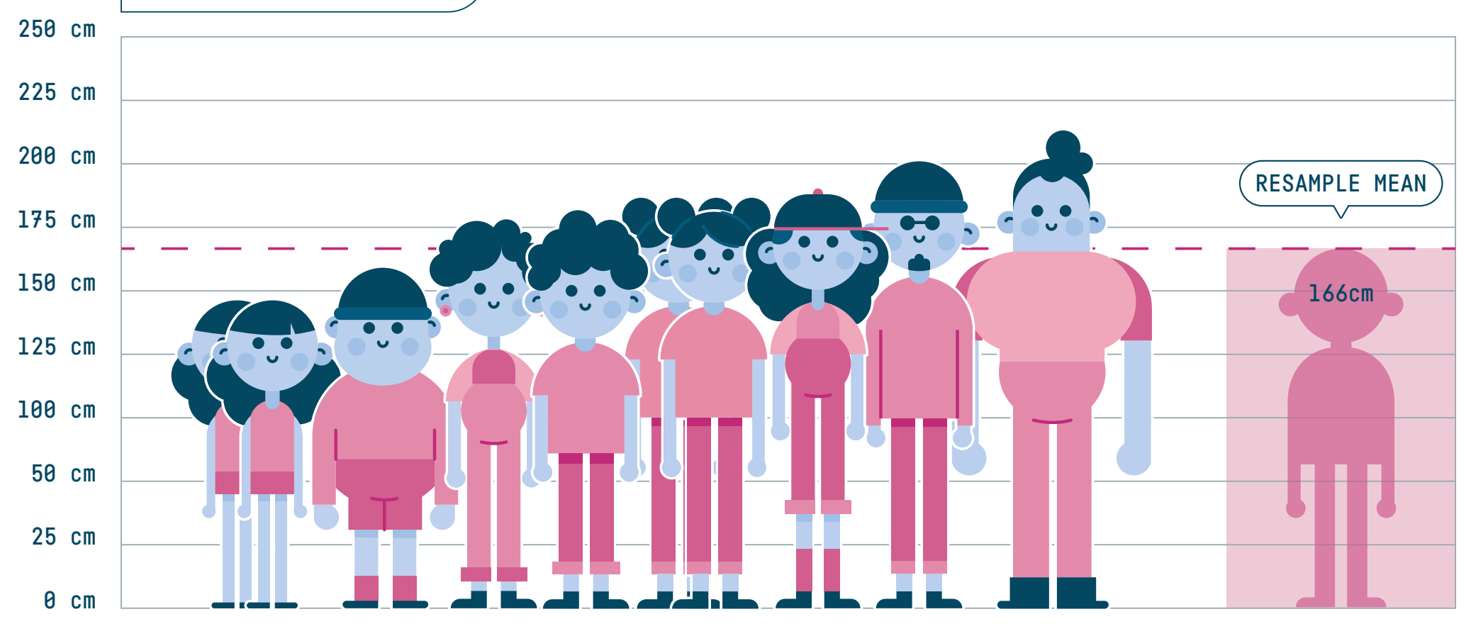
RESAMPLE #0014



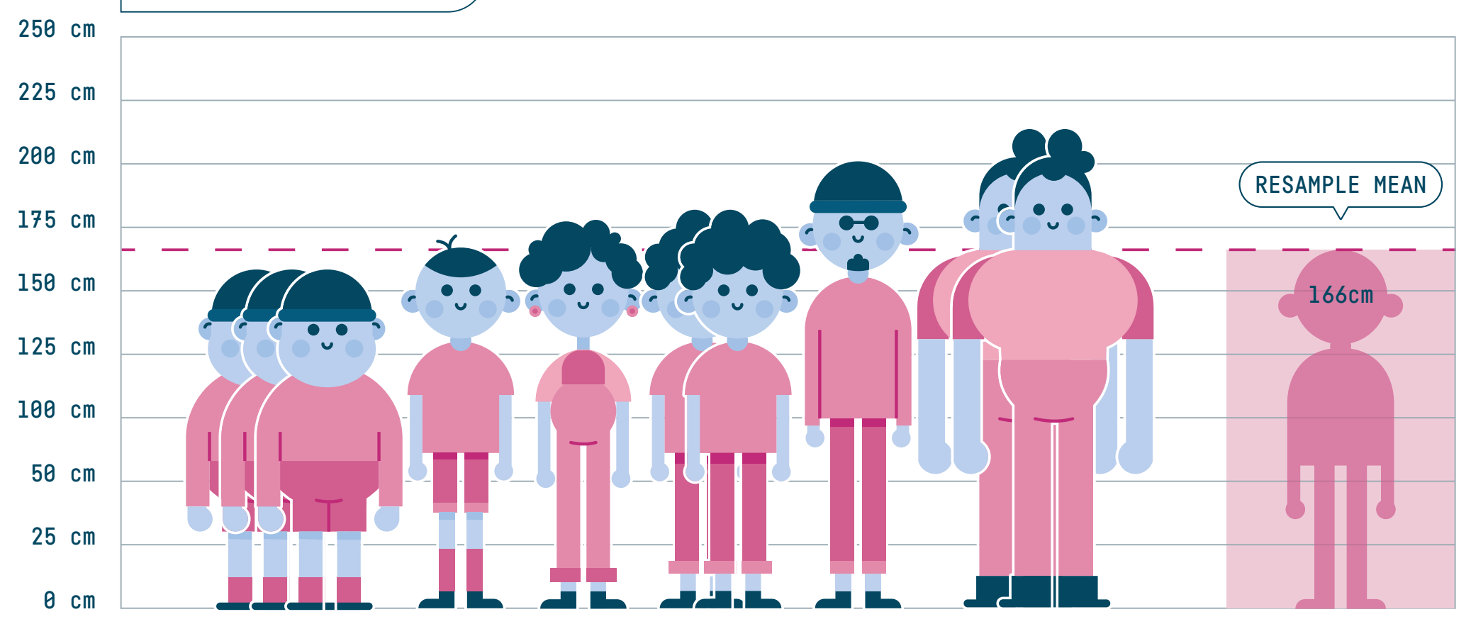
RESAMPLE #0013



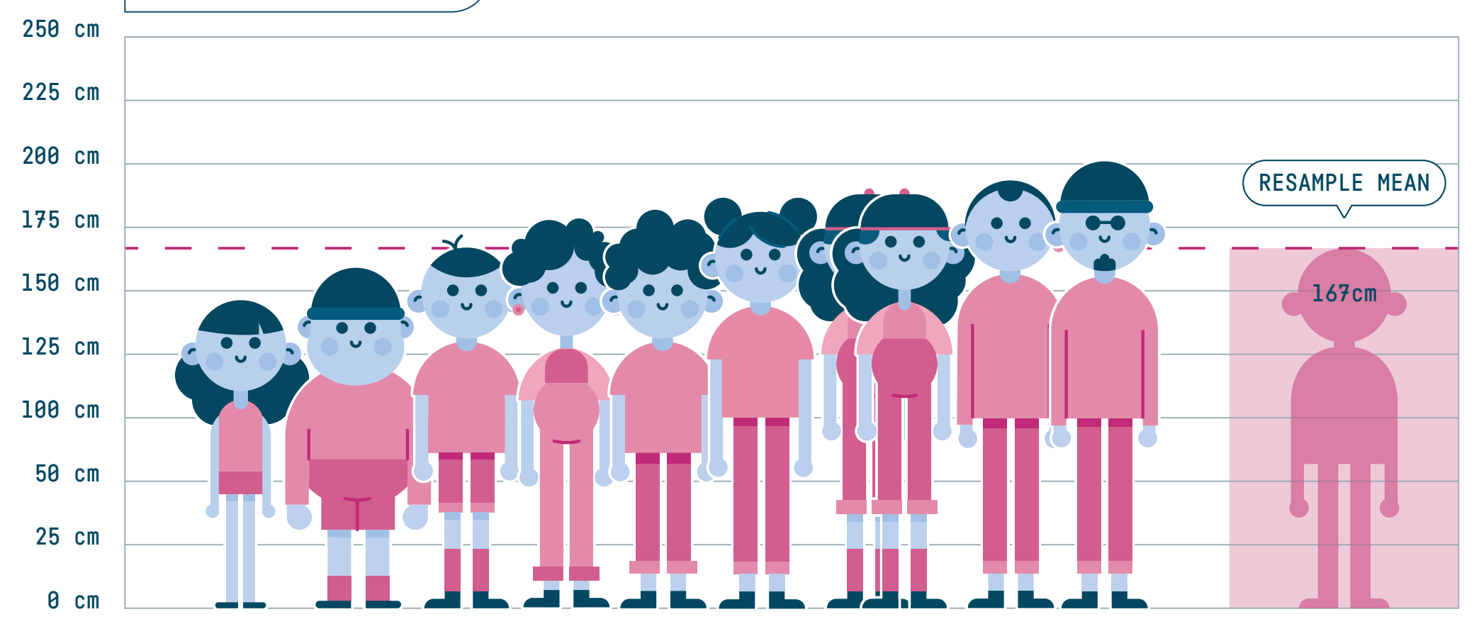
RESAMPLE #0012



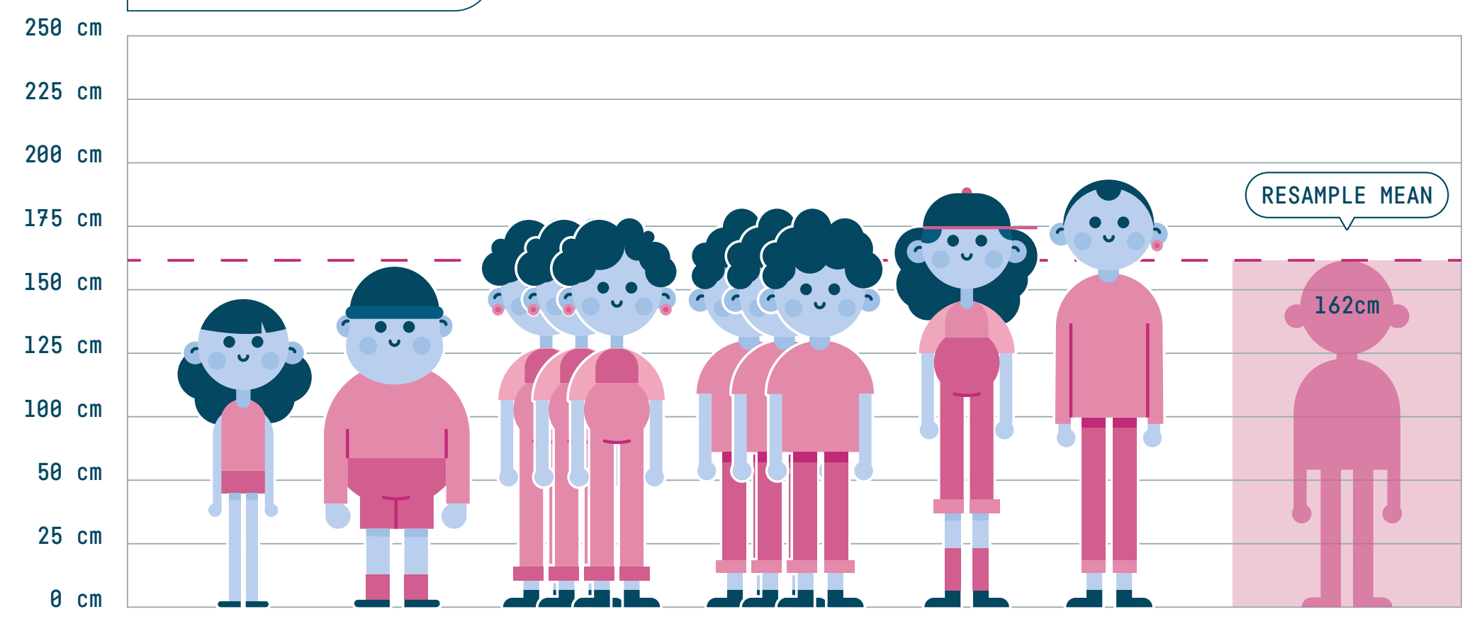
RESAMPLE #0011



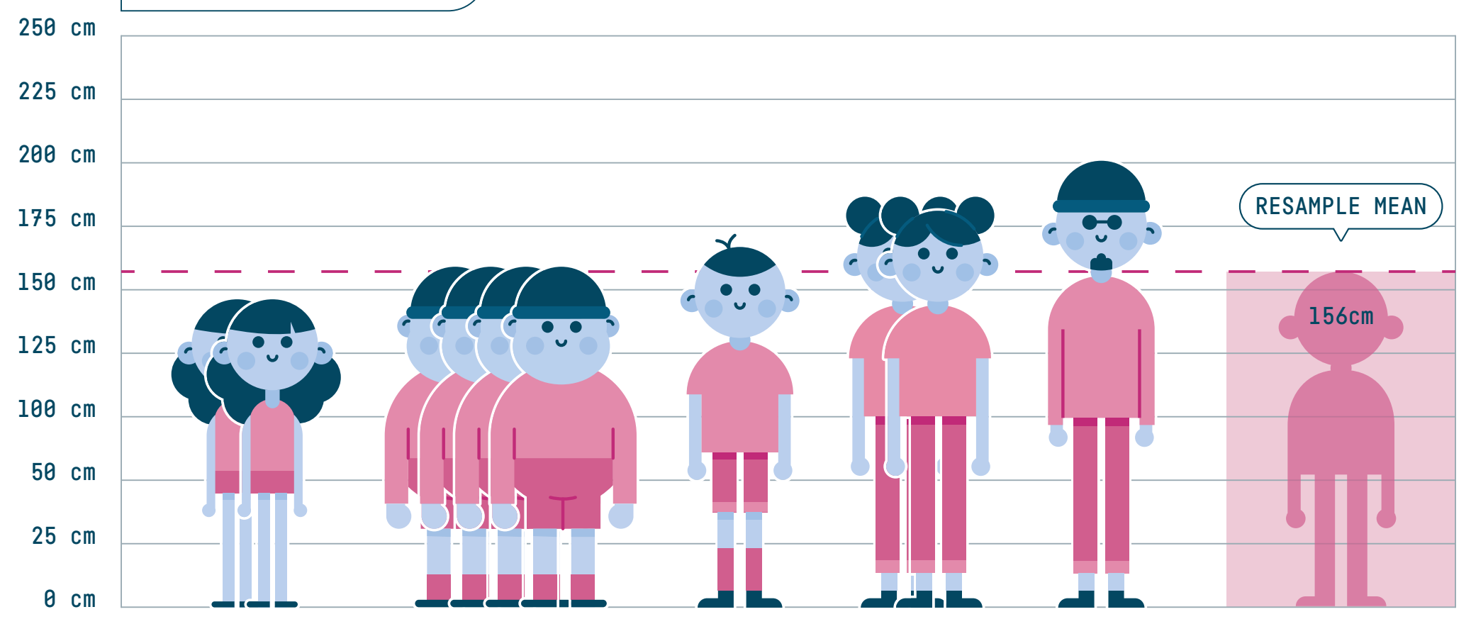
RESAMPLE #0010



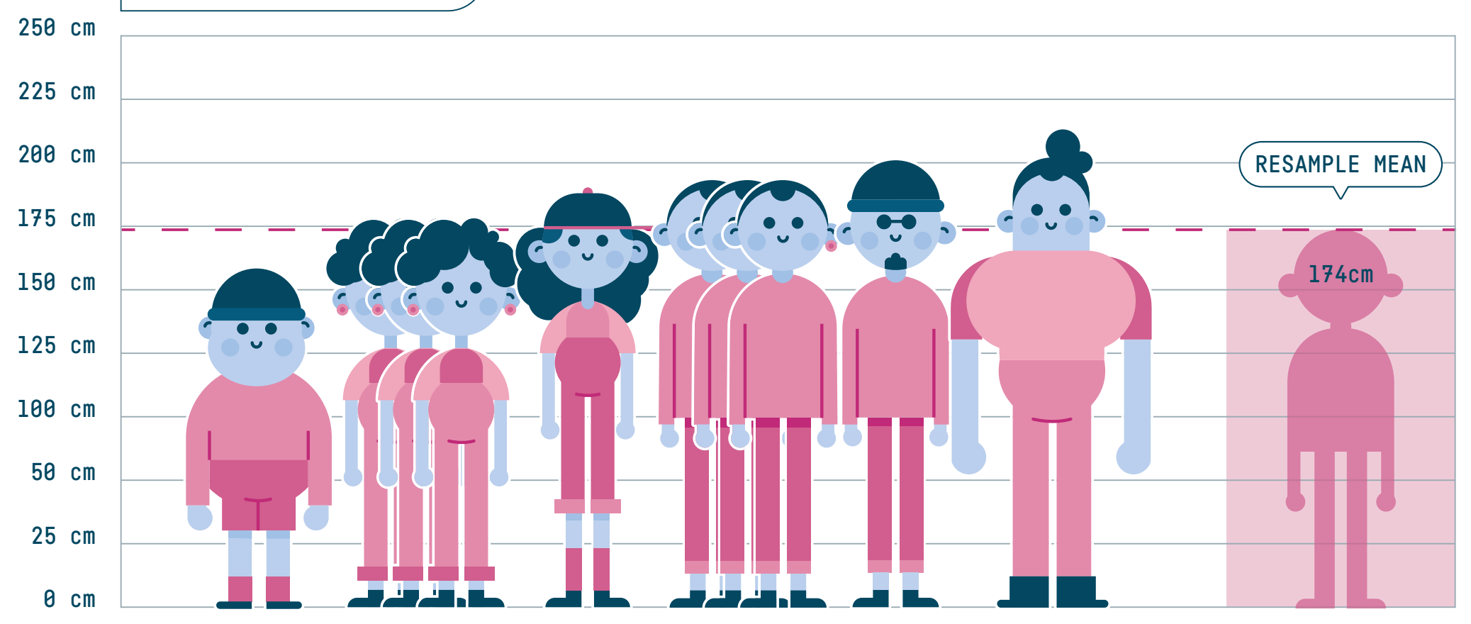
RESAMPLE #0009



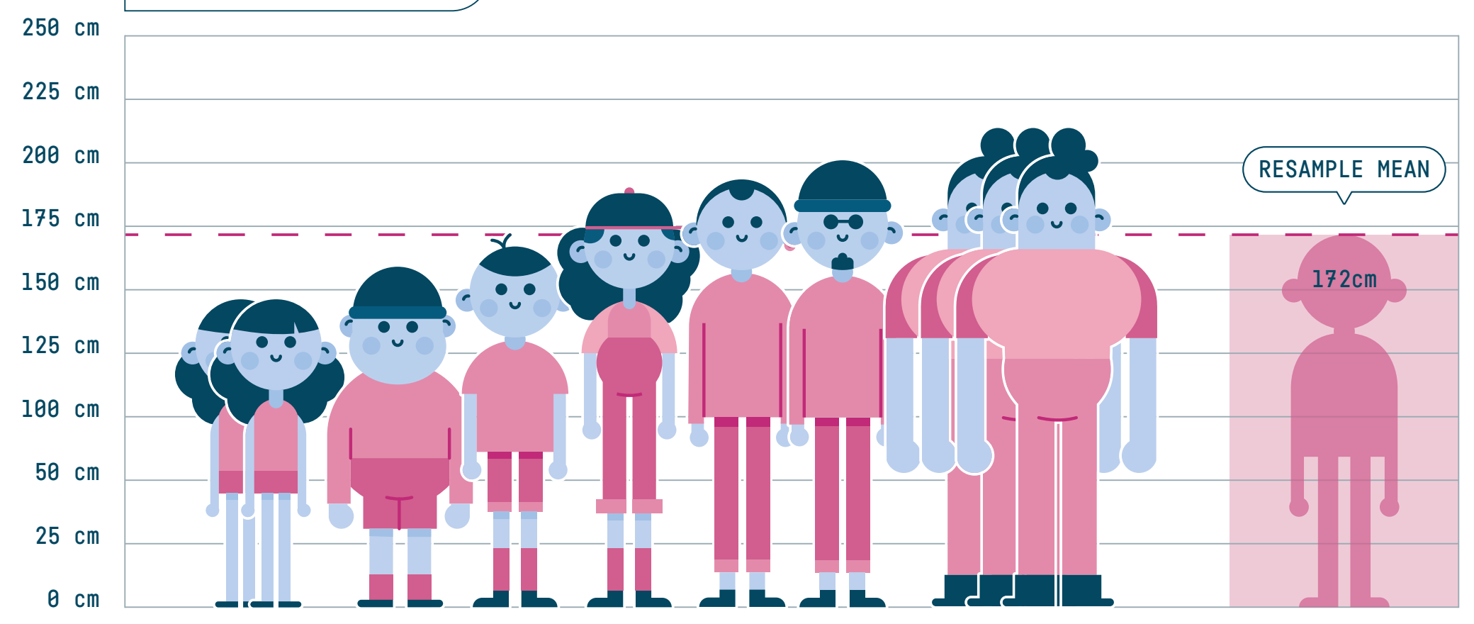
RESAMPLE #0008



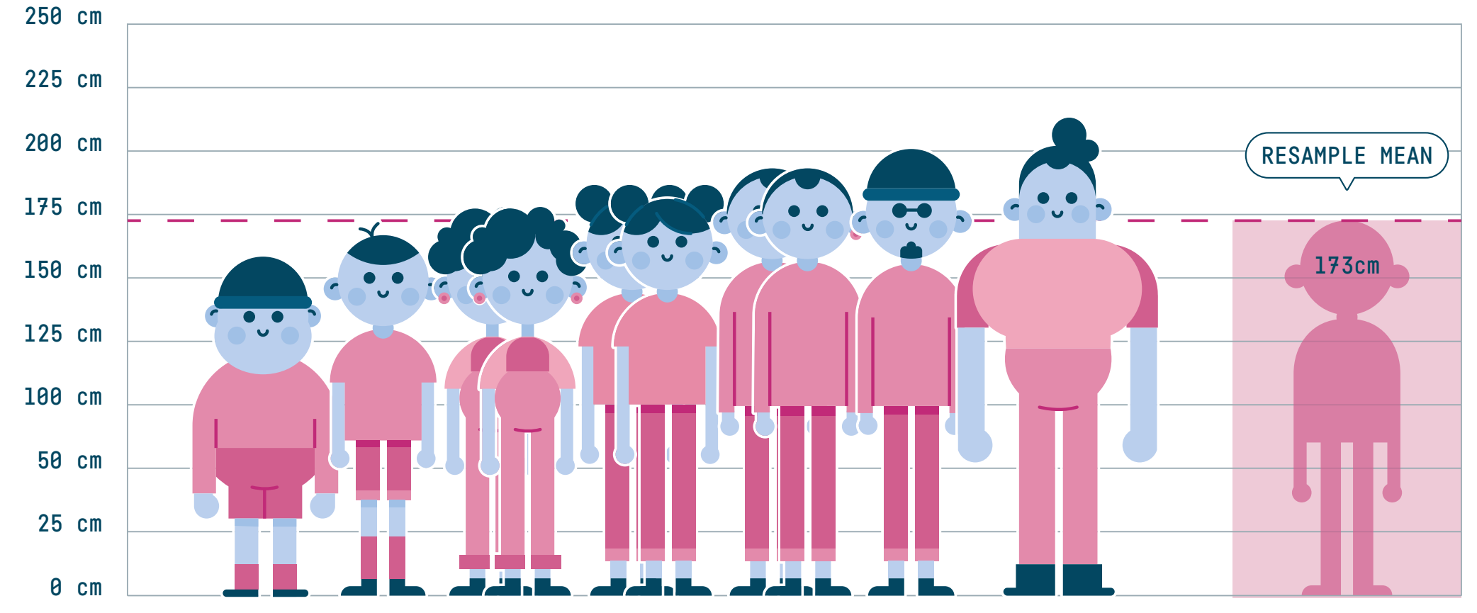
RESAMPLE #0007



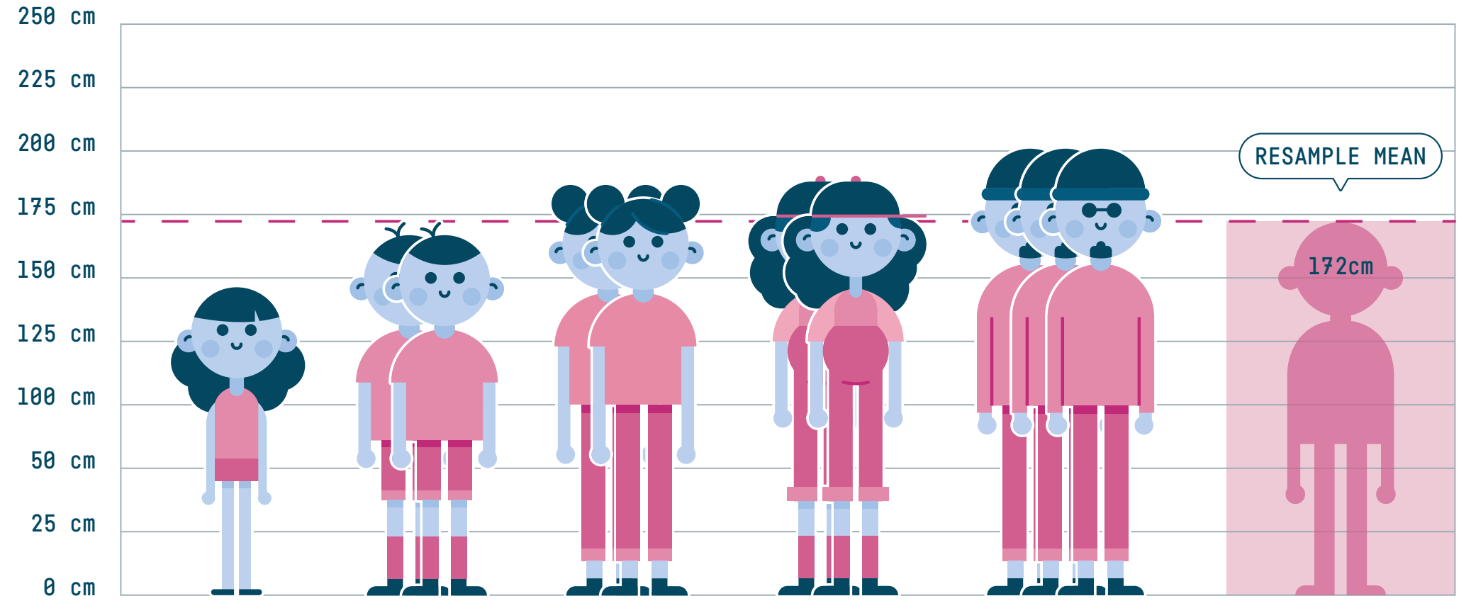
RESAMPLE #0006



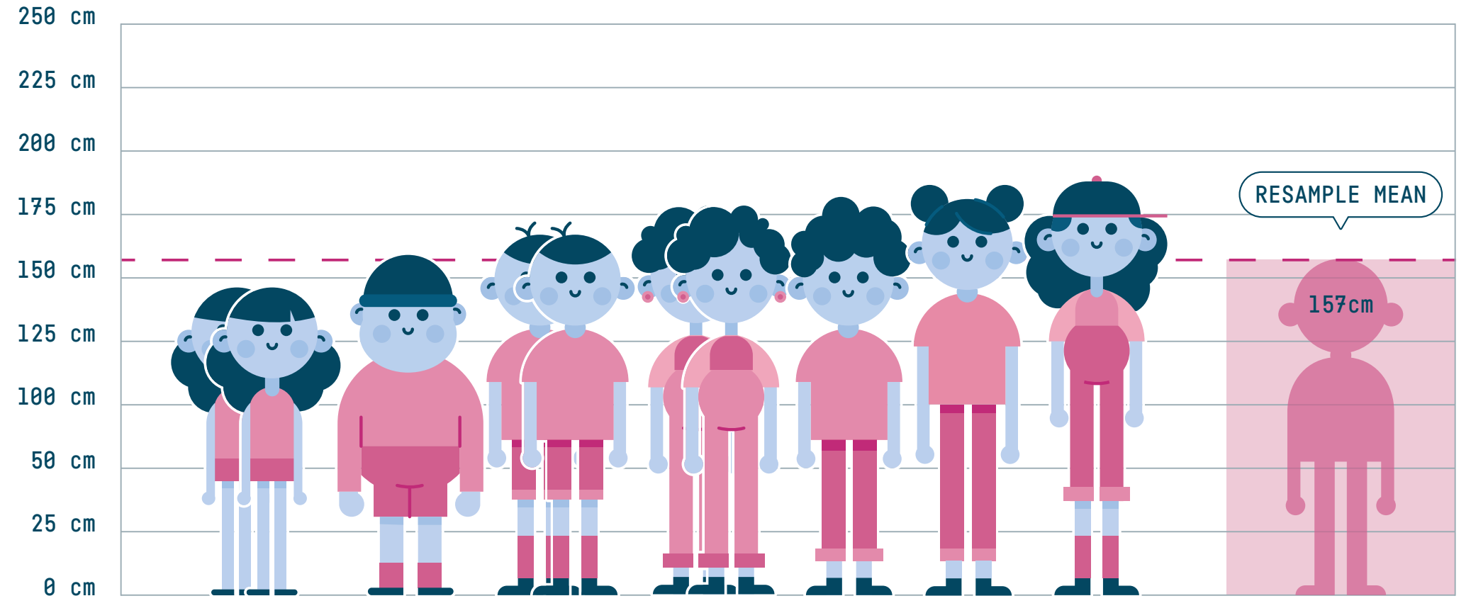
RESAMPLE #0005



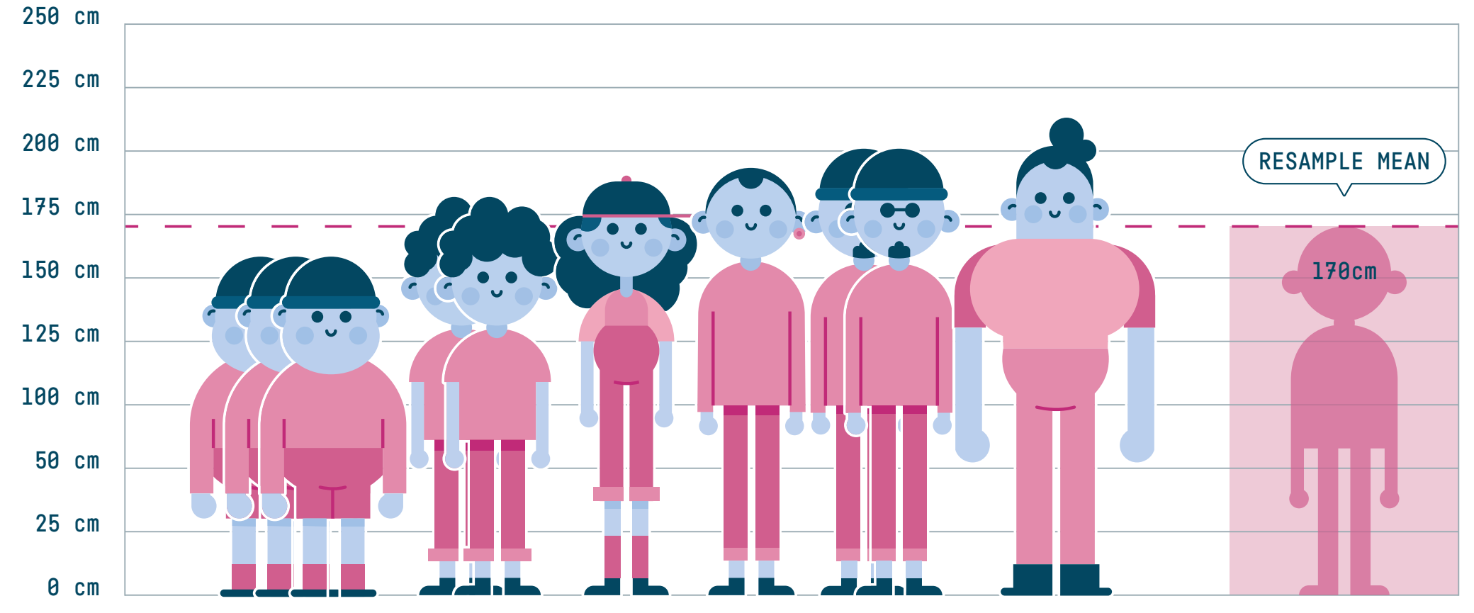
RESAMPLE #0004

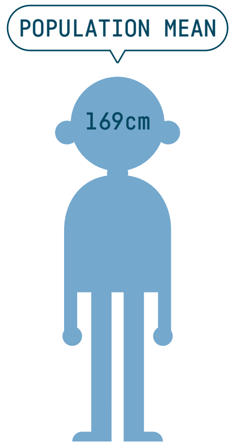


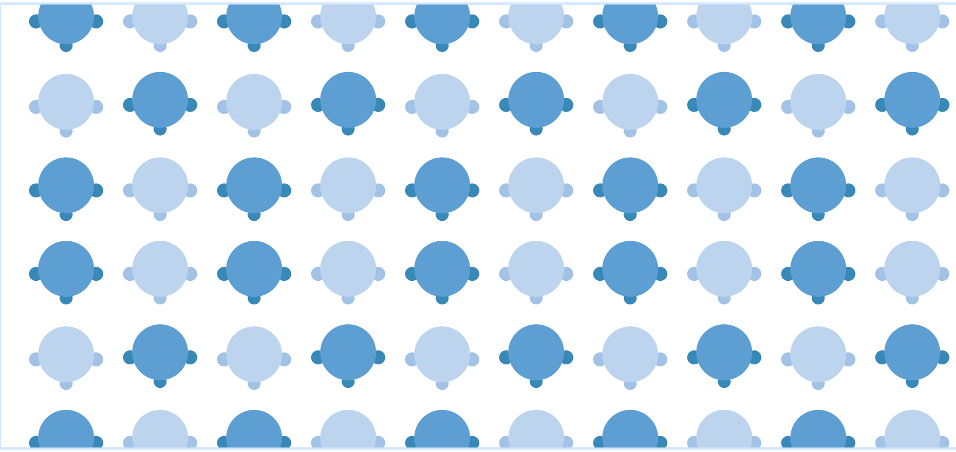
RESAMPLE #0003



RESAMPLE #0002







HOW ACCURATE IS IT THO?

In an ideal world, in which we have all the height values of the entire blueberry-loving population, we could just calculate the real mean height of the population.

We can compare the real mean height with the estimated distribution to verify if the algorithm did its job. (if it did, give it a cookie or something).

PULL