

SP-101
STATS &
PROBABILITY

LECTURE 1



WHY LEARN STATISTICS?

1

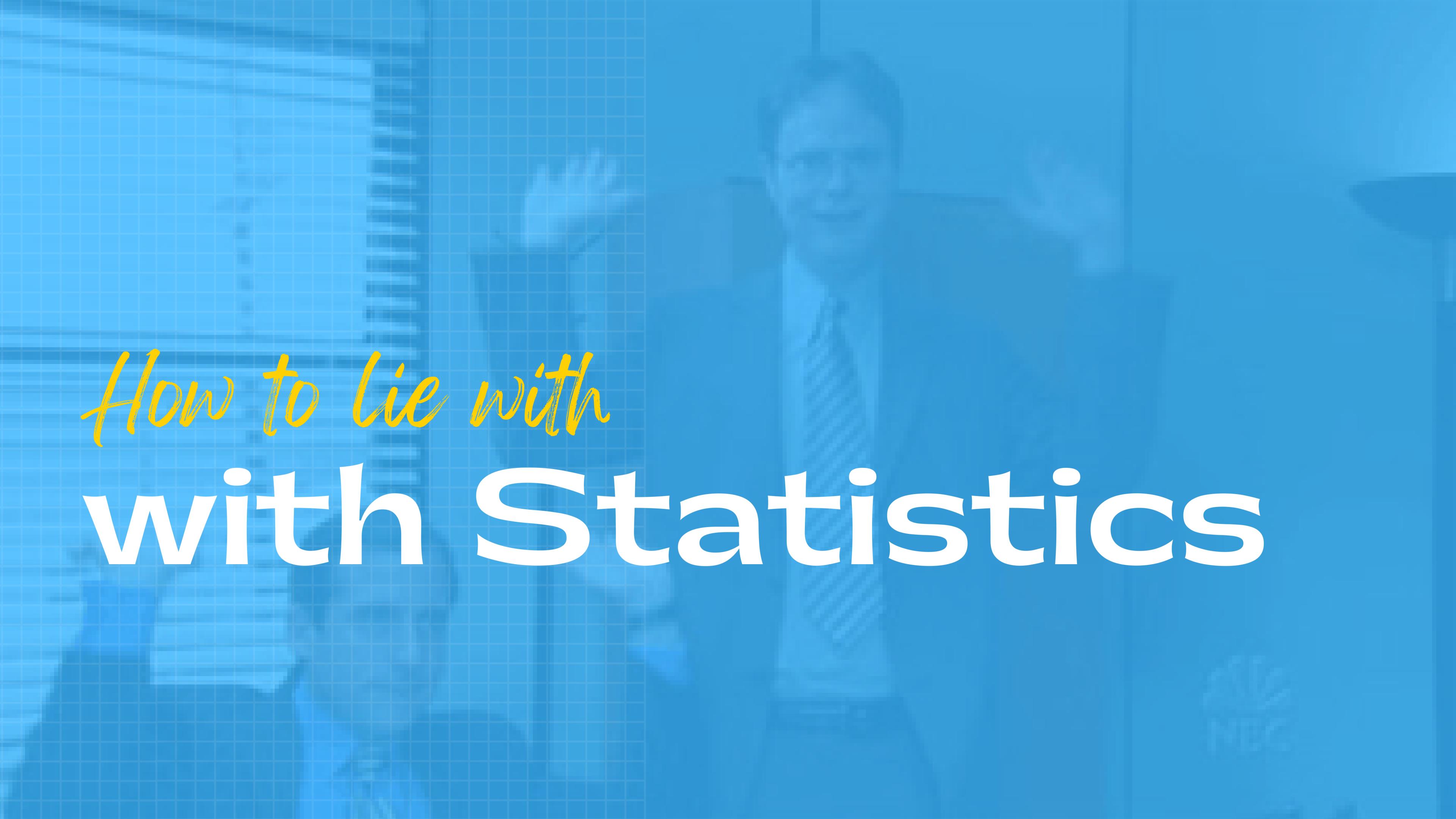
Without it, we can't plan our budgets, pay taxes or enjoy our games ...

2

Data is everywhere. Informed decision making and Critical Thinking ...

3

It will be difficult to lie, get easy promotion, or get elected without doing any work ...



How to lie with
with Statistics

There are
three kinds of
lies: lies,
damned lies,
and statistics

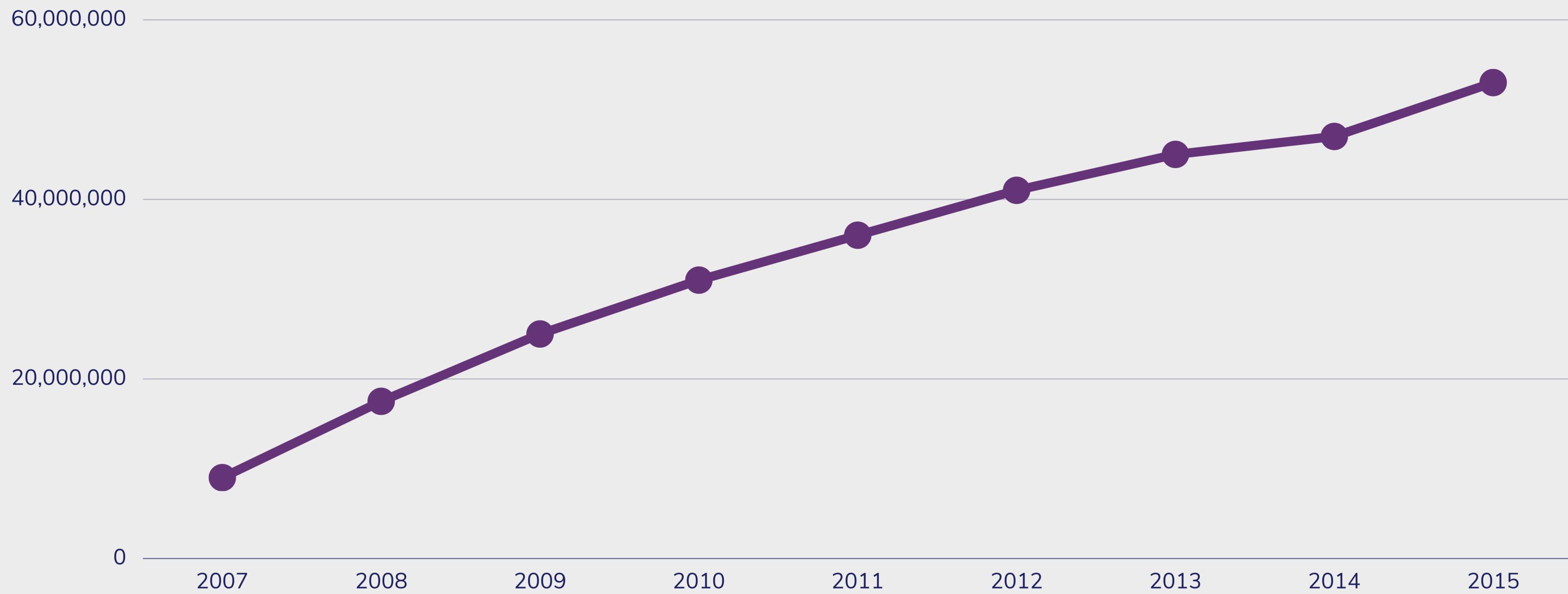
Mark Twain

Study reveals...

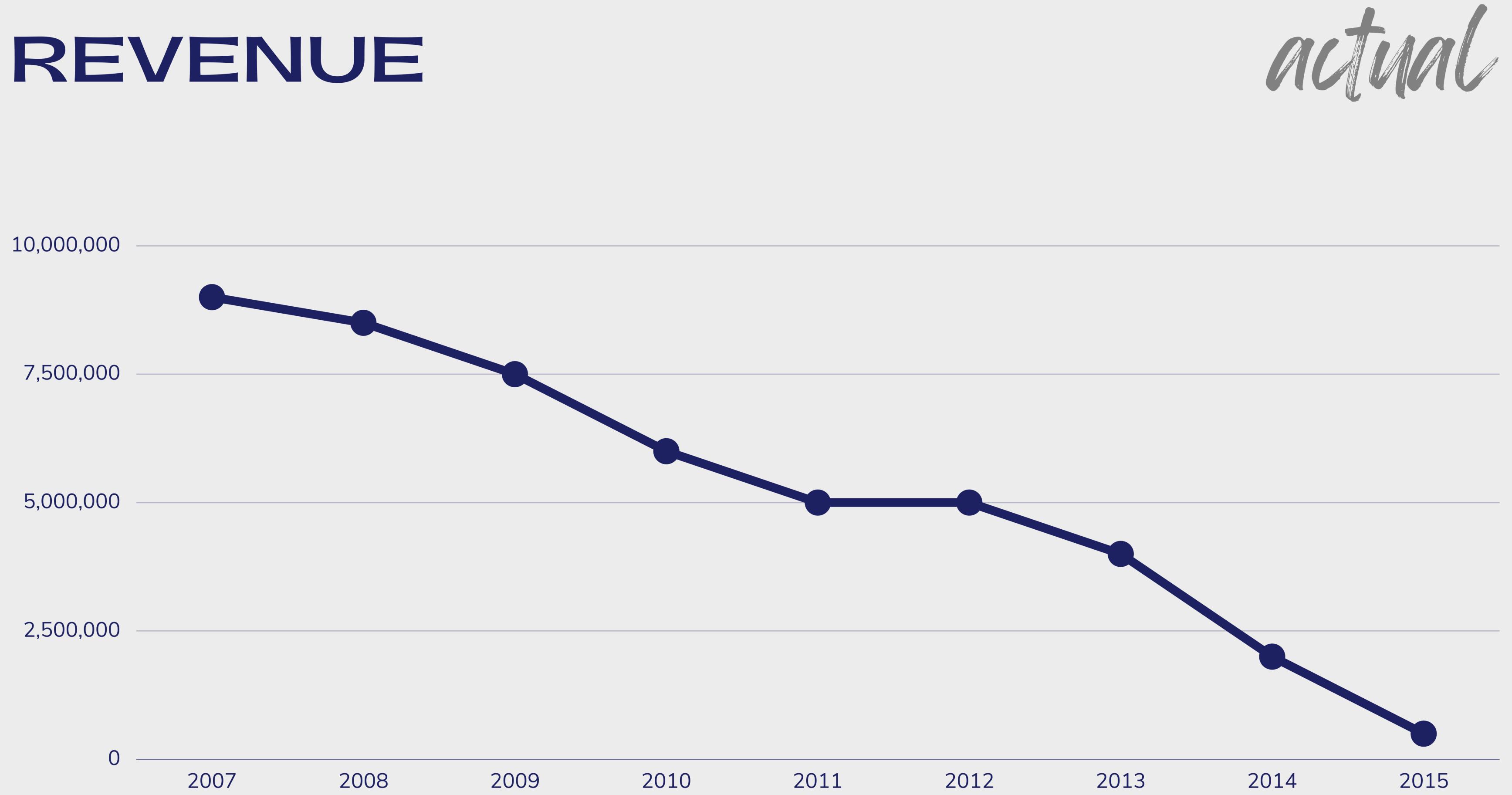
**100% of
new born
babies are
non-smokers**

REVENUE

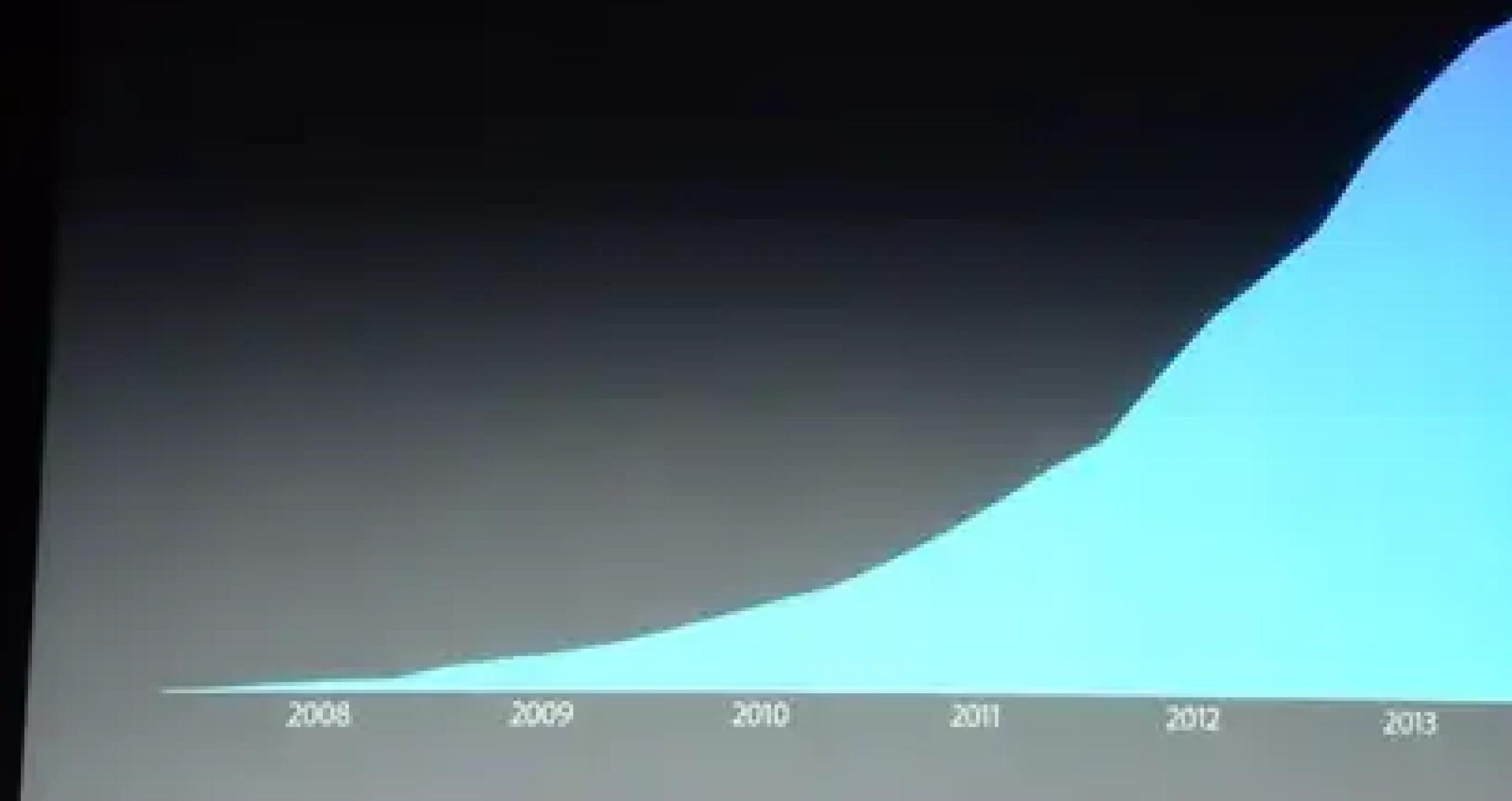
cumulative



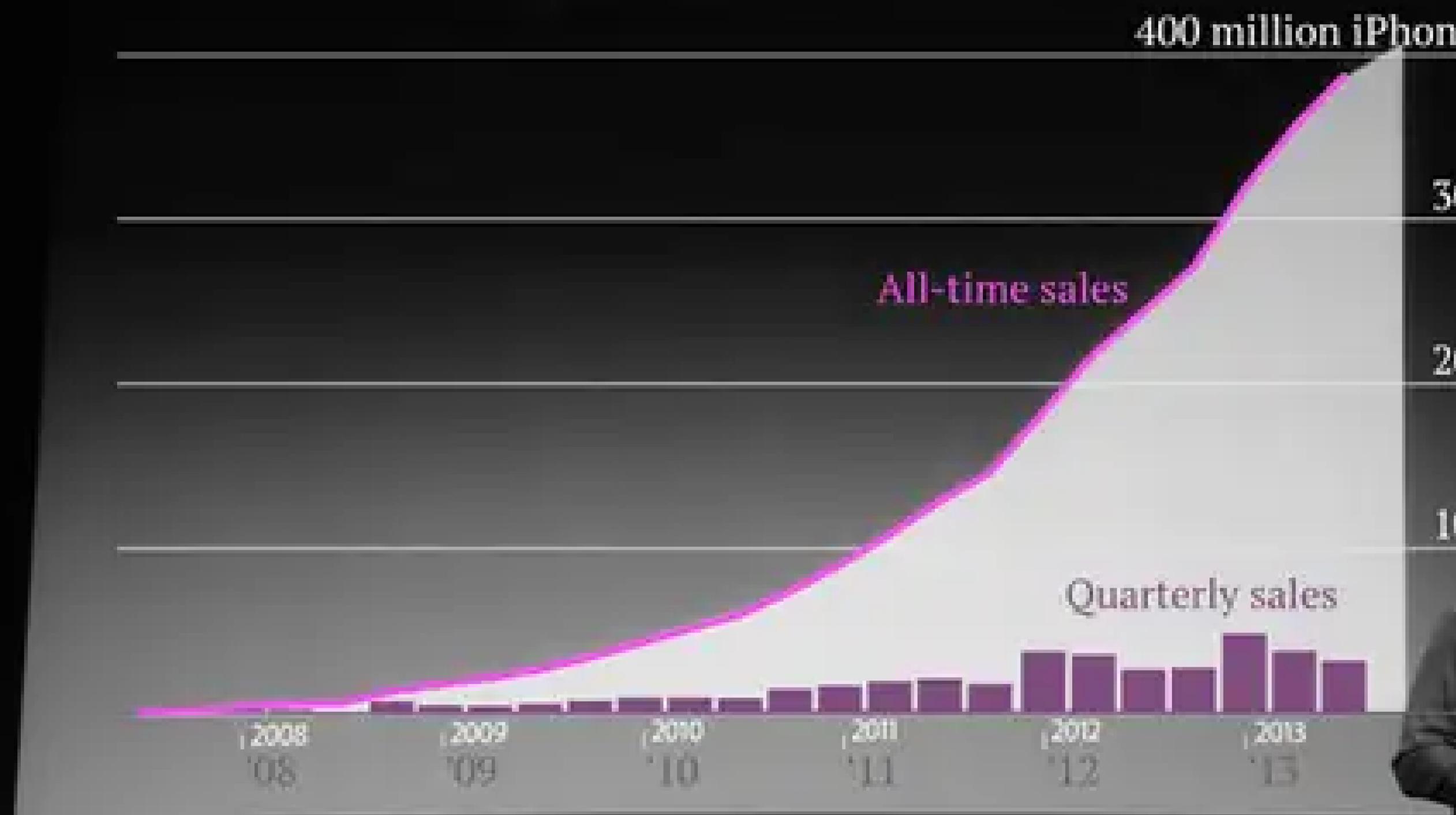
REVENUE



Cumulative iPhone sales



Cumulative iPhone sales



Quartz | qz.com

Data: Apple

Photo: The Verge

Cumulative iPhone sales

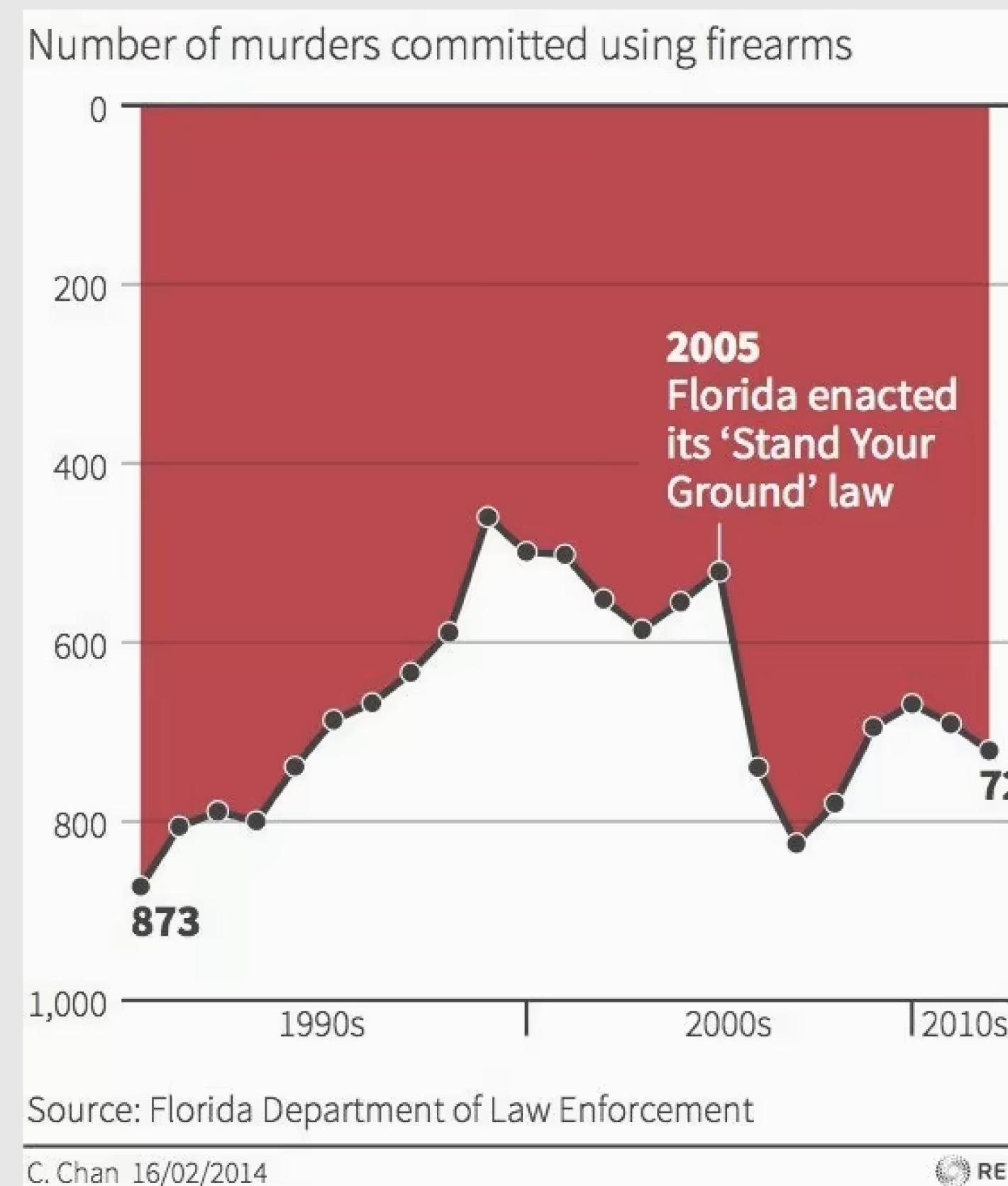


Quartz | qz.com

Data: Apple

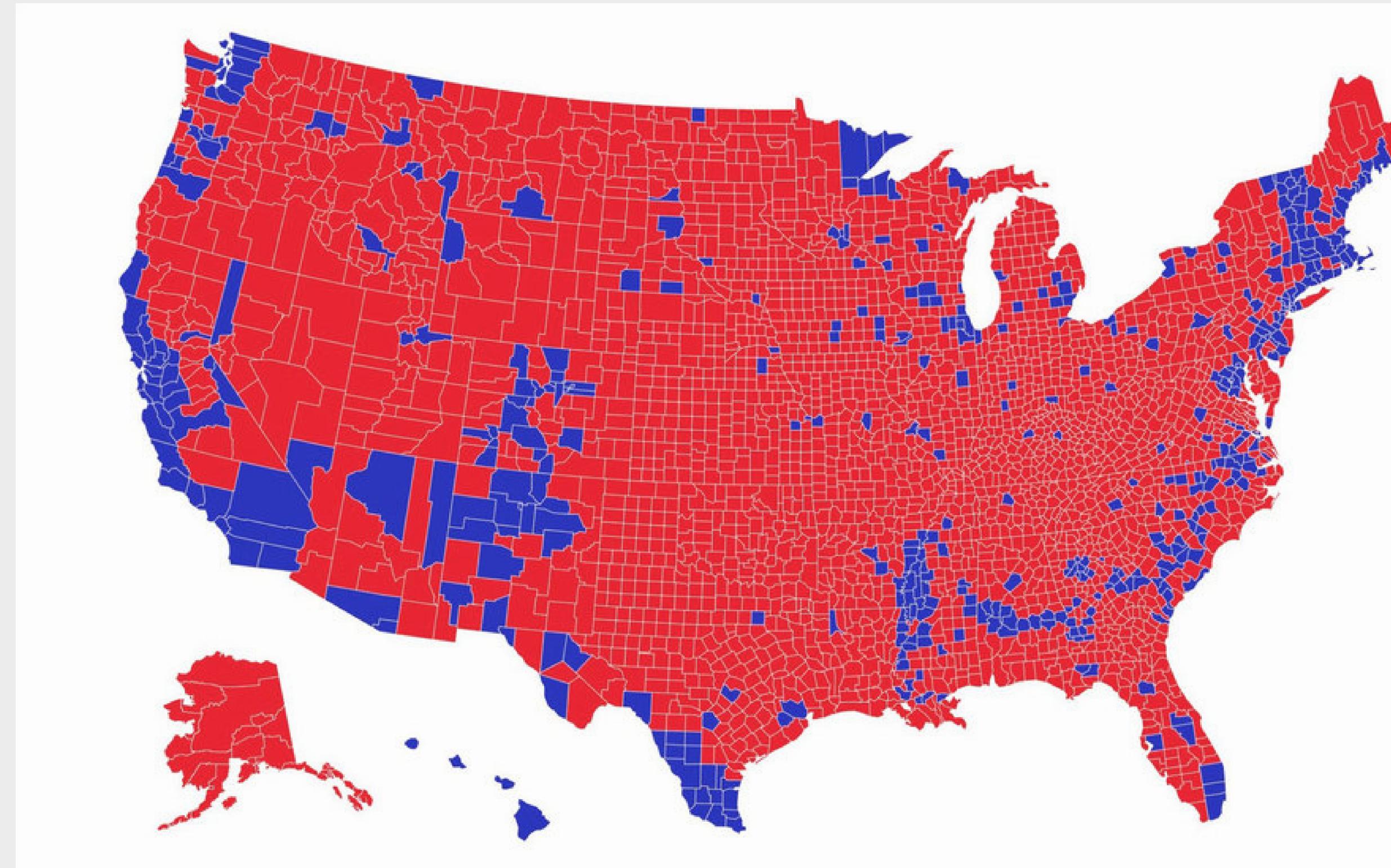
Photo: The Verge

GUN DEATHS IN FLORIDA

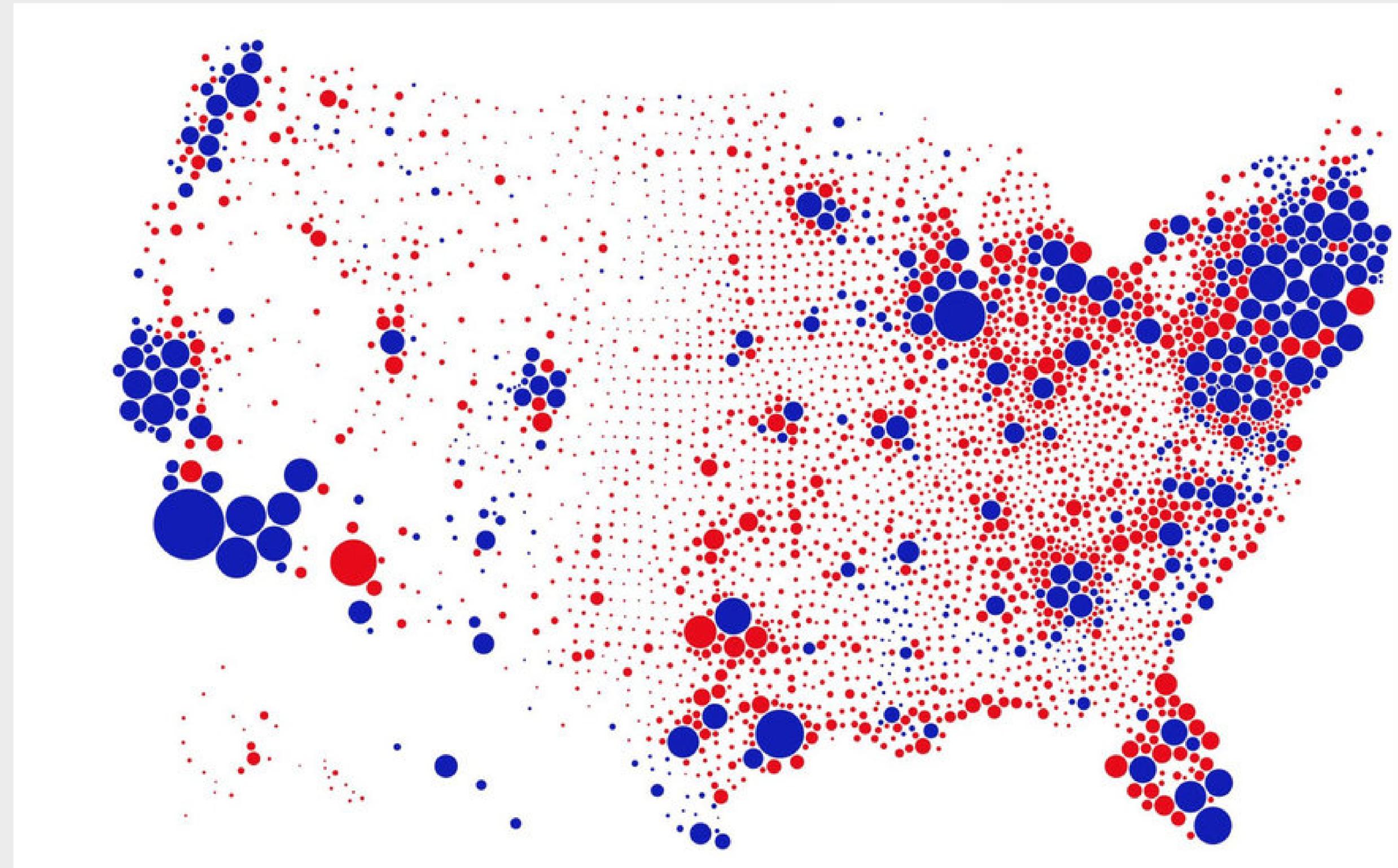


y-axis is upside-down, with zero at the top and the maximum value at the bottom.

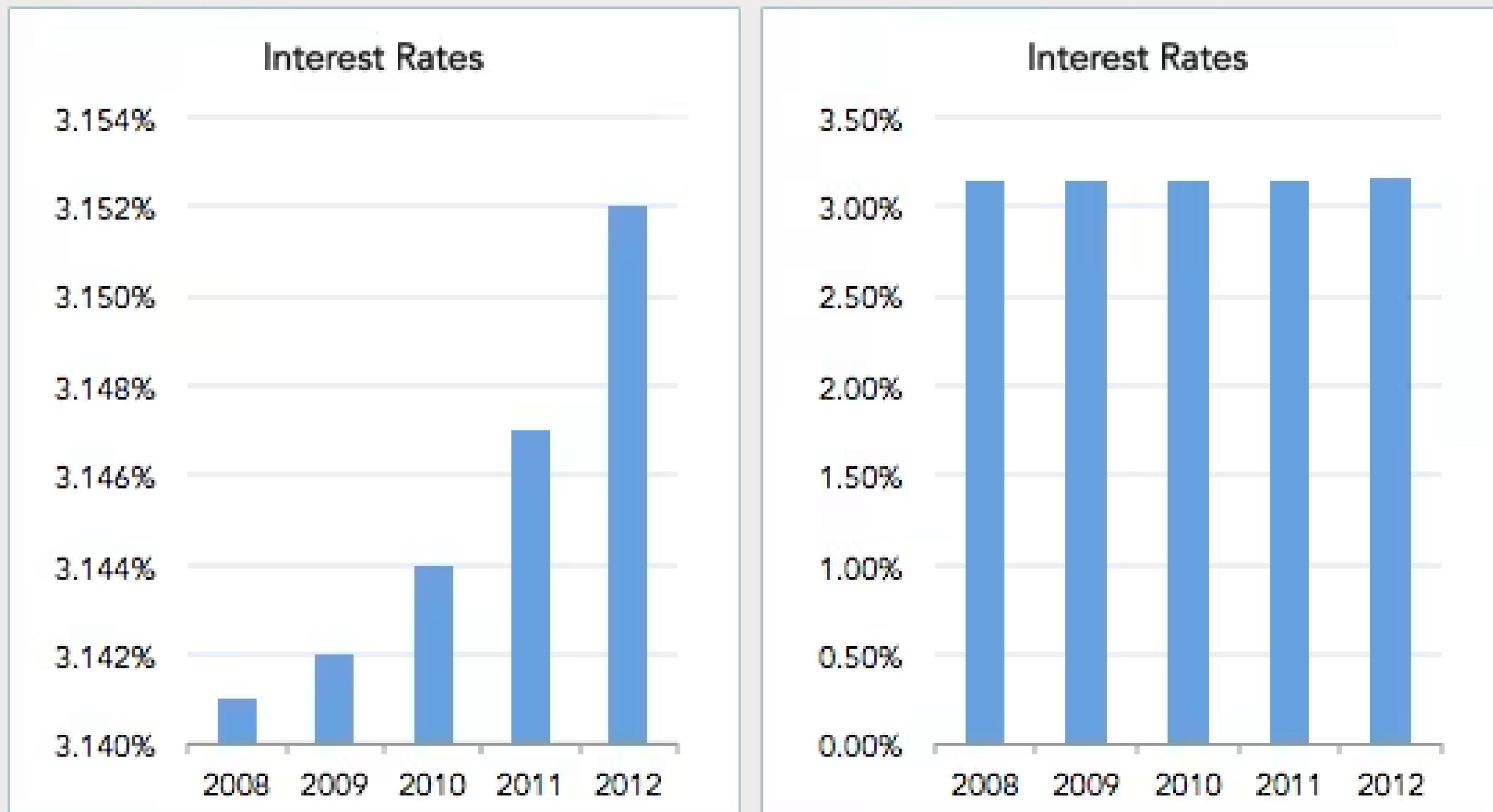
2016 PRESIDENTIAL ELECTION (RED = REPUBLICAN, BLUE = DEMOCRAT)



LAND DOESN'T VOTE. PEOPLE DO



SAME DATE. DIFFERENT Y-AXIS



Now you, too, can out-double-talk your accountant;
confuse your political opponent;

How to

prove that your product has secret built-in goodness!

LIE

with

Statistics

by Darrell Huff

PICTURES BY
Irving Geis



ASSIGNMENT
FIND CHARTS,
GRAPHS OR MAPS
*where you think data
analysts were paid extra
bonuses*

STATISTICS IS THE
STUDY OF COLLECTING,
ANALYZING, &
INTERPRETING DATA

GOAL.

*To make sense of numerical
information and draw conclusions
based on that information.*

RESULT

Final (D/N), Wankhede, April 02, 2011, ICC Cricket World Cup



Sri Lanka

274/6



India

(48.2/50 ov, T:275) 277/4

India won by 6 wickets (with 10 balls remaining)

PLAYER OF THE MATCH

MS Dhoni, IND



91* (79)

PLAYER OF THE SERIES

Yuvraj Singh, IND



362 runs • 15 wkts

BEST PERFORMANCES - BATTERS



DPM Jayawardene SL

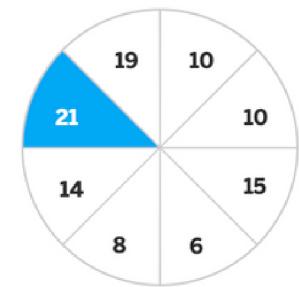
103 runs (88)

13 fours 0 six

CONTROL
83%

PRODUCTIVE SHOT
Leg Glance

23 runs 0 four 0 six



G Gambhir IND

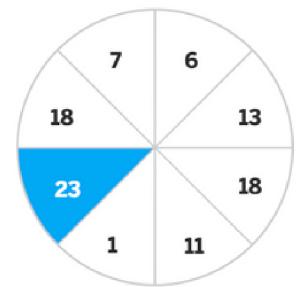
97 runs (122)

9 fours 0 six

CONTROL
79%

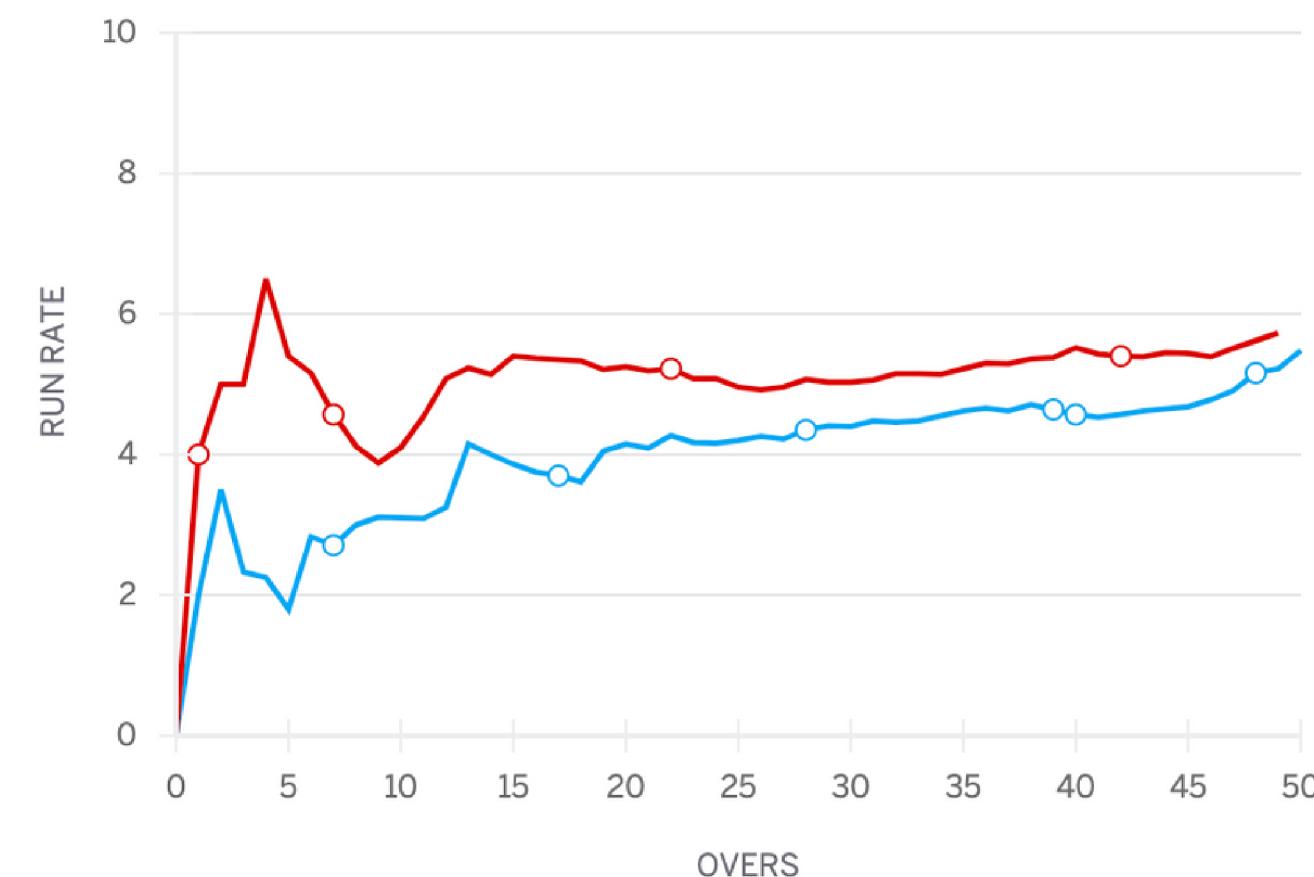
PRODUCTIVE SHOT
Leg Glance

32 runs 3 fours 0 six



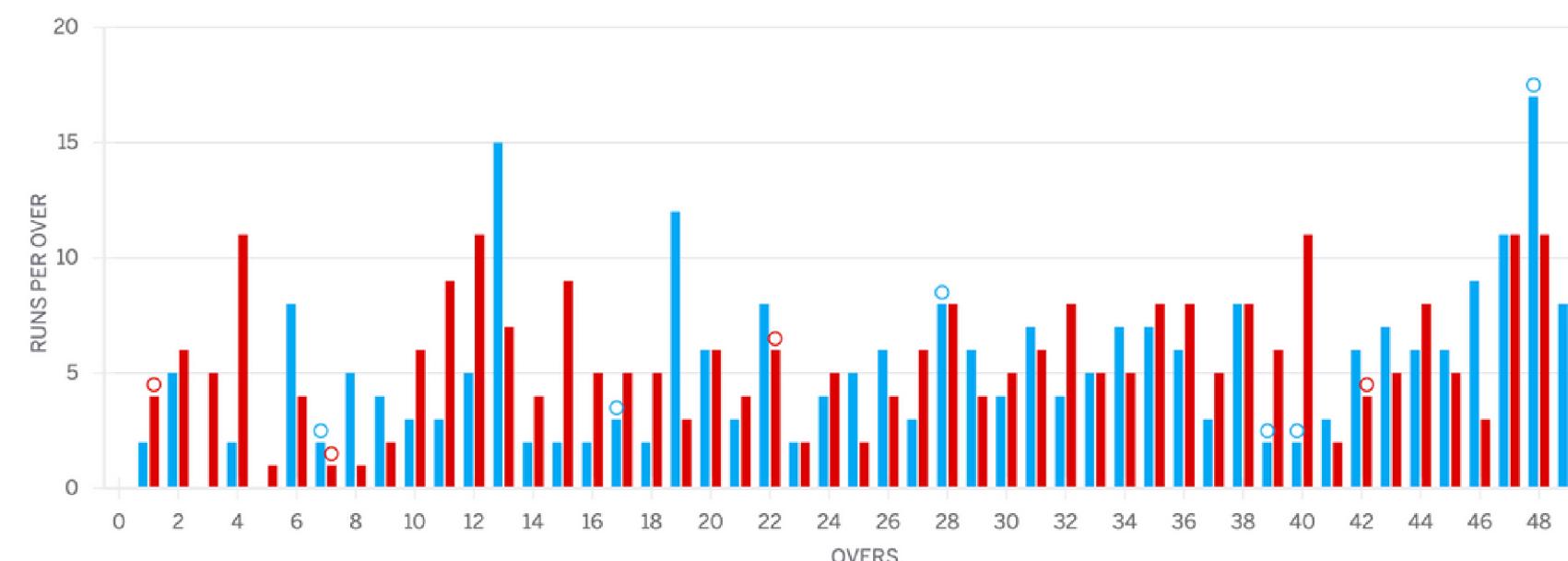
RUN RATE GRAPH

● Sri Lanka ● India



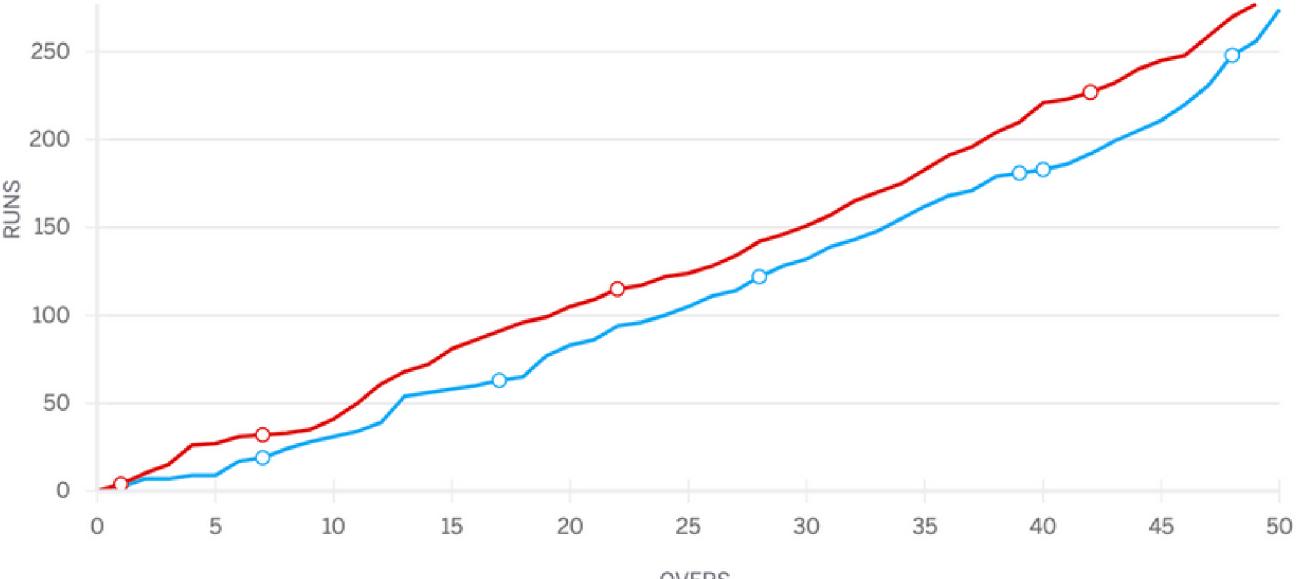
MANHATTAN

● Sri Lanka ● India



WORM

● Sri Lanka ● India



Fundamentals of Statistics

*The basic vocabulary
of Statistics*

Population

Sample

Variable

Parameter

Distribution

POPULATION

Population refers to the entire group of individuals, objects, or events that we are interested in studying.

This group can be as large or as small as we want, depending on the research question we are trying to answer.

Example

For example, if we want to study the health of all people living in a particular city, the population would be all individuals living in that city. This would include people of all ages, genders, races, and ethnicities.



SAMPLE

Smaller group of individuals, objects, or events that we select from the larger population to study or analyze.

The goal of taking a sample is to draw conclusions about the entire population using a smaller, more manageable dataset.

Example

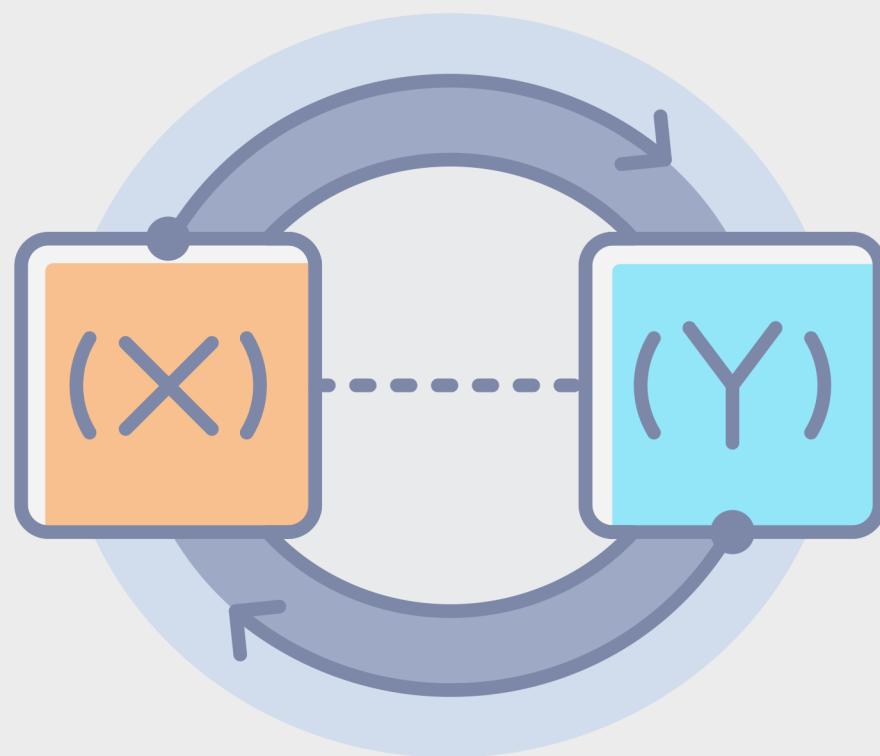
to study the average age of employees at a company, we could take a sample of employees and analyze their ages.



VARIABLES

A variable is an attribute that we can measure or observe in order to learn more about a particular group of people, objects, or events.

Variables can be either numerical or categorical,



Example

if we want to study the relationship between studying and grades, we can use two variables: the amount of time someone studies (measured in hours per week) and their grades (measured as an A, B, C, or D).

By measuring these two variables, we can identify patterns and relationships between studying and grades.

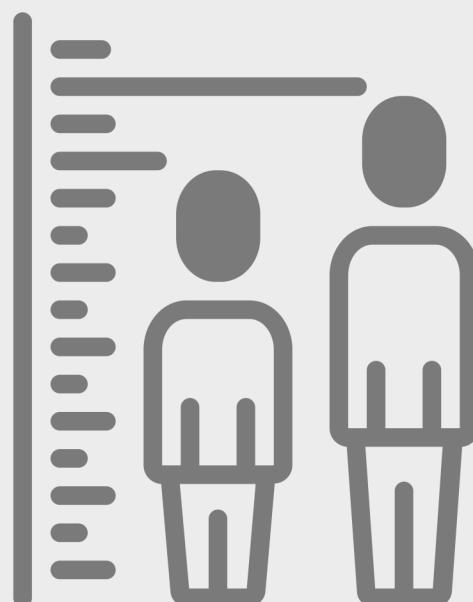
PARAMETERS

A parameter in statistics is a number that tells us something about a whole group of things or people.

A parameter is a single number that describes the whole group. A variable, on the other hand, is something that can vary or change from person to person or thing to thing in the entire group.

Example

Average height: Let's say we want to know the average height of all the kids in a school. The average height would be the parameter that tells us something about the whole group of kids.



DISTRIBUTION

A distribution in statistics is a way of showing how the values of a variable are spread out or distributed among a group.

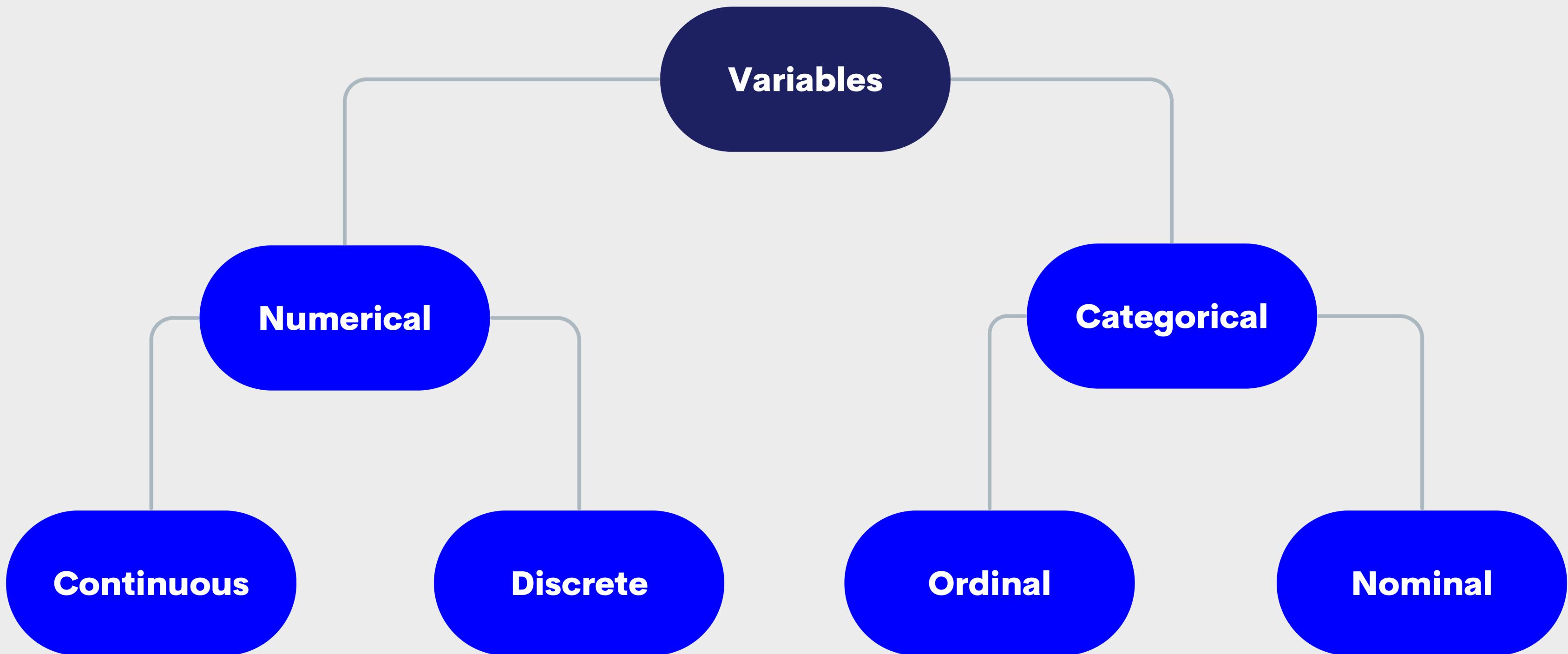
We can show a distribution visually using graphs like a histogram, which shows the frequency of each value in the data. This can help us see patterns in the data and how values are spread out.



Example

Grade distribution: If we were looking at the grades of all students in a class, the distribution would show how many students received each grade (e.g. A, B, C, D, or F).

Income distribution: If we were looking at the incomes of all people in a city, the distribution would show how many people earn each amount of money (e.g. \$0-\$25,000, \$25,001-\$50,000, \$50,001-\$75,000, etc.).





NUMERICAL

Numeric variables are a type of variable that represent numerical values.

These values can be manipulated using mathematical operations.

For example, let's consider two numeric variables, weight and height. We can add, subtract, multiply, and divide these values to perform various calculations.

CONTINUOUS

A continuous variable is a type of variable that can take on any value within a certain range.

This means that there are an infinite number of possible values that the variable can have.

Example

Time can be any value between a certain minimum and maximum value. So, it can be 3:00 PM, or 3:01 PM, or 3:01:30 PM.

There are an infinite number of possible values for time because it can be any value within a certain range.

DISCRETE

A discrete variable is a type of numeric variable that takes on specific, individual values.

In other words, there are only certain possible values that the variable can have.

Example

- The number of siblings someone has: This can only take on specific integer values, such as 0, 1, 2, 3, etc.
- The number of pets someone owns: This can only take on specific integer values, such as 0, 1, 2, 3, etc.
- The number of children in a family: This can only take on specific integer values, such as 0, 1, 2, 3, etc.



CATEGORICAL

Categorical variables are a type of variable in statistics that represent characteristics or attributes that can be placed into categories.

These categories are often named and have no numerical order.

For example, a categorical variable could be gender, where the categories are male and female. Another example is color, where the categories could be red, blue, green, and so on.

NOMINAL

A nominal variable is a type of categorical variable that describes a characteristic or attribute without any order or ranking

In other words, there is no specific order or hierarchy to the values of the variable.

Example

- Favourite color: There is no inherent order or ranking to different colors.
- Animal species: Different types of animals do not have an inherent order or ranking.
- Gender: There is no inherent order or ranking to the different genders.

ORDINAL

An ordinal variable is a type of categorical variable that describes a characteristic or attribute with a natural order or ranking.

In other words, there is a specific order or hierarchy to the values of the variable.

Example

- Education level: Different levels of education (high school diploma, associate's degree, bachelor's degree, etc.) have a natural order or ranking.
- Income brackets: Different levels of income (low, middle, high) have a natural order or ranking.
- Grades: Different letter grades (A, B, C, D, F) have a natural order or ranking.



Your Turn

Questions