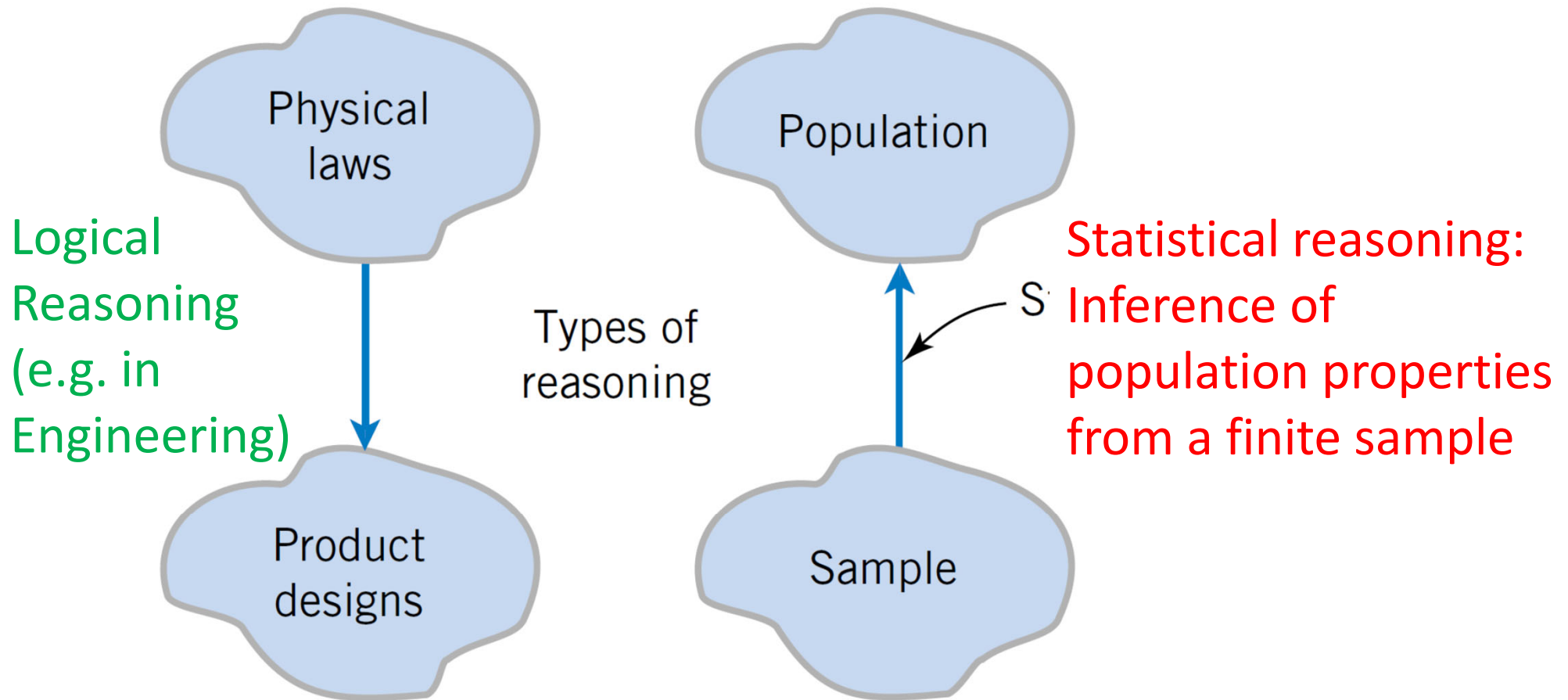


Descriptive statistics:  
Populations, Samples  
Histograms, Quartiles  
Sample mean and  
variance

# Two types of reasoning



# Numerical Summaries of Data

- Data are the **numerical observations** of a **phenomenon of interest**.
- The totality of all observations is a **population**.
  - **Population can be infinite** (e.g. abstract random variables)
  - **It can be very large** (e.g. 7 billion humans or all patients who have cancer of a given type)
- A (usually small) portion of the population collected for analysis is a random **sample**.
- We want to **use sample** to **infer facts about populations**
- The **inference** is not perfect but **gets better and better as sample size increases**.

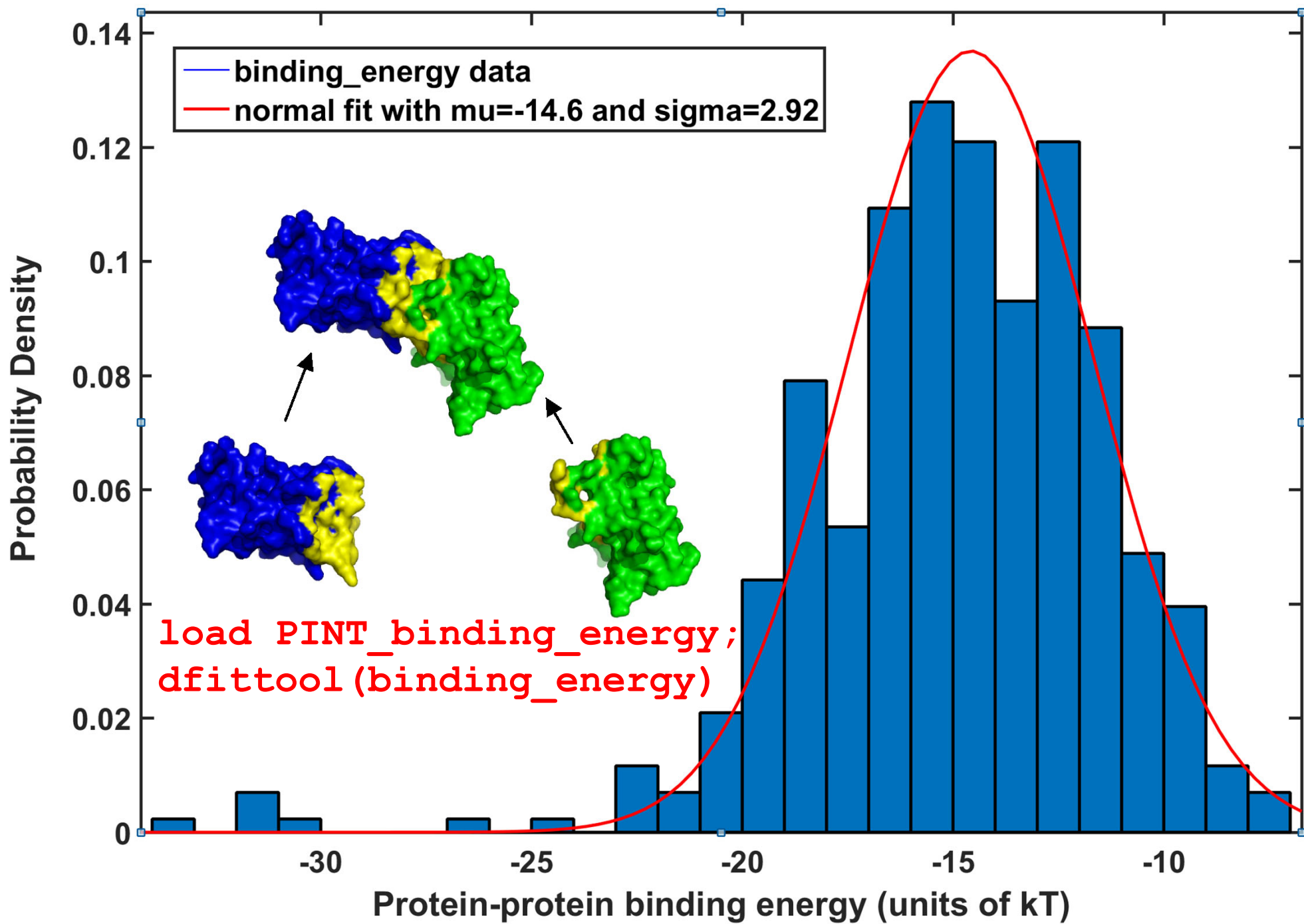
# Some Definitions

- The random variables  $X_1, X_2, \dots, X_n$  are a **random sample** of **size  $n$**  if:
  - a) The  $X_i$  are **independent** random variables.
  - b) Every  $X_i$  has **the same probability distribution**.
- Such  $X_1, X_2, \dots, X_n$  are also called **independent and identically distributed** (or **i. i. d.**) random variables

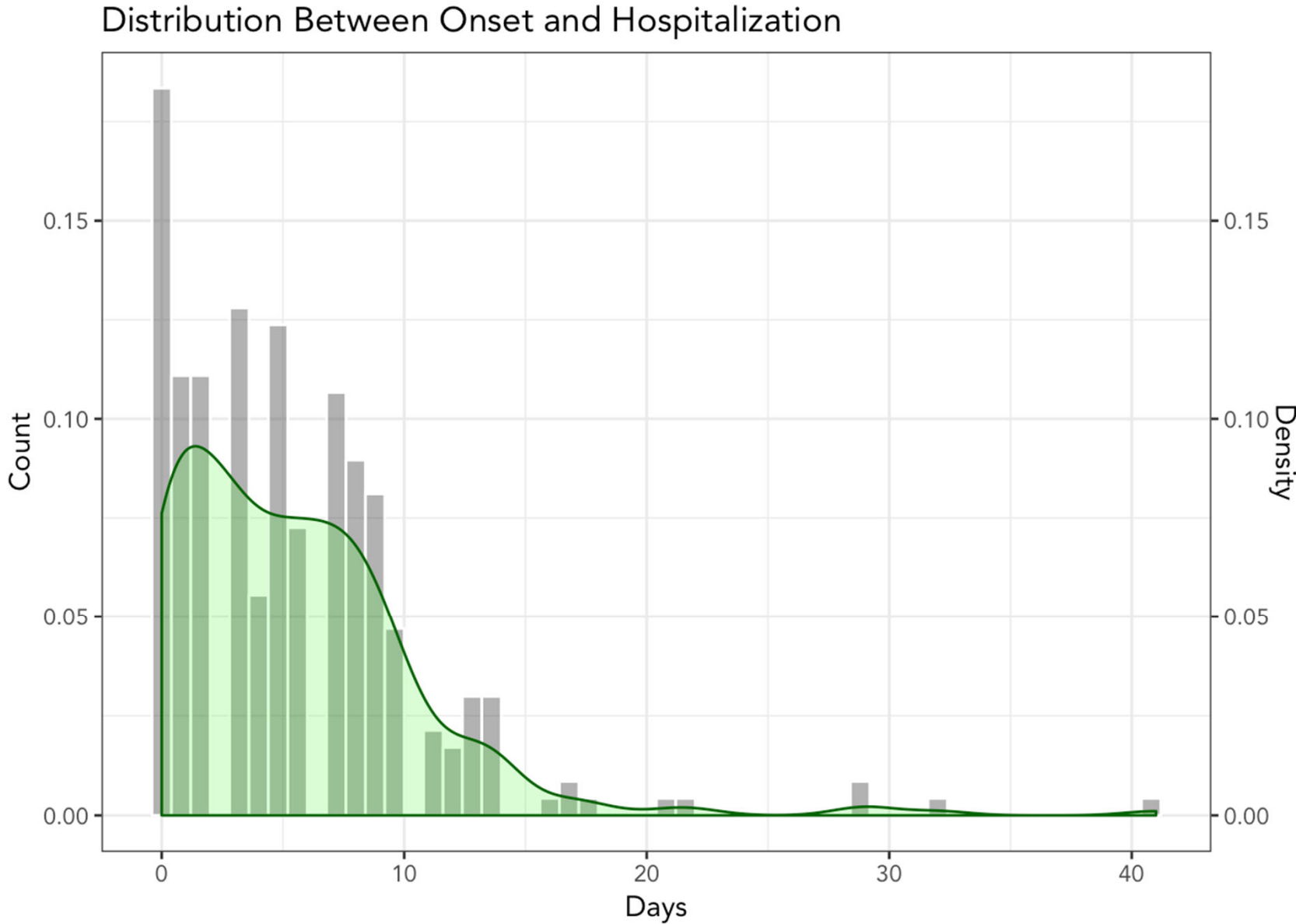
Ways to describe a sample:

**Histogram**

approximates PDF  
(or PMF)



# PDF of time between COVID-19 symptoms onset and hospitalization in IL, April 2020



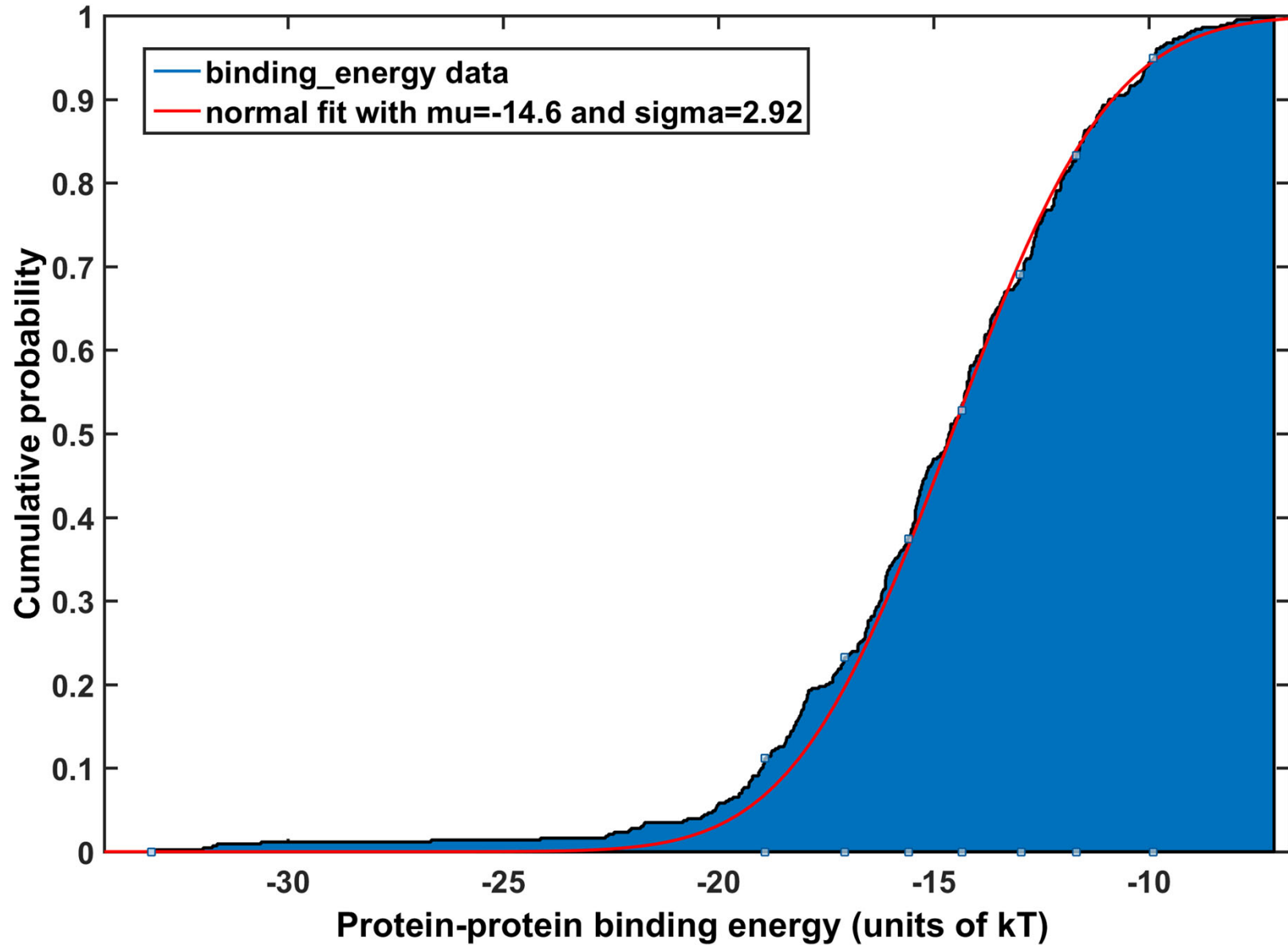
# Histograms with Unequal Bin Widths

- If the data is tightly clustered in some regions and scattered in others, it is visually helpful to use **narrow bin widths** in the **clustered region** and **wide bin widths** in the **scattered areas**.
- To approximate the PDF, the **rectangle area**, not the height, must be proportional to the **bin relative frequency**.

$$\text{Rectangle height} = \frac{\text{bin relative frequency}}{\text{bin width}}$$



# Cumulative Frequency Plot



# Median, Quartiles, Percentiles

- The **median**  $q_2$  divides the sample into two equal parts: 50% ( $n/2$ ) of sample points below  $q_2$  and 50% ( $n/2$ ) points above  $q_2$
- The **three quartiles** partition the data into four equally sized counts or segments.
  - 25% of the data is less than  $q_1$ .
  - 50% of the data is less than  $q_2$ , the median.
  - 75% of the data is less than  $q_3$ .
- There are **100 percentiles**.  $n$ -th percentile  $p_n$  is defined so that  $n\%$  of the data is less than  $p_n$

# Matlab exercise #1

- Find the **median** and **lower & upper quartiles** of a **n=100 sample** drawn from a **continuous uniform distribution** in **[0,1]**
- Do not use built-in Matlab functions for this exercise!
- Hint: use **[a,b]=sort(r1)**; to rank order your sample. The variable **a** returns r1 sorted in the increasing order.
- **How to find the median and both quartiles from a?**

# How to find the median & quartiles

- % Example: find median and lower quartile of
- % a sample with  $n=100$  drawn from uniform
- `r1=rand(100,1);`
- `[a,b]=sort(r1);`
- `med=(a(50)+a(51))./2`
- `sum(r1<med) % verify`
- `q1=(a(25)+a(26))./2`
- `sum(r1<q1) % verify`

Credit: XKCD  
comics

# WHY ARE THERE SLAVES IN THE BIBLE

WHY DO TWINS HAVE DIFFERENT FINGERPRINTS  
WHY ARE AMERICANS AFRAID OF DRAGONS  
WHY IS HTTPS CROSSED OUT IN RED  
WHY IS THERE A LINE THROUGH HTTPS  
WHY IS THERE A RED LINE THROUGH HTTPS ON FACEBOOK  
WHY IS HTTPS IMPORTANT

# QUESTIONS FOUND IN GOOGLE AUTOCOMPLETE



WHY ARE THERE WEEKS  
WHY DO I FEEL DIZZY  
WHY ARE THERE ZIPPERS  
ON ALL ZIPPERS

WHY DO WHALES JUMP  
WHY ARE WITCHES GREEN  
WHY ARE THERE MIRRORS ABOVE BEDS  
WHY DO I SAY UH  
WHY IS SEA SALT BETTER  
WHY ARE THERE TREES IN THE MIDDLE OF FIELDS  
WHY IS THERE NOT A POKEMON MMO  
WHY IS THERE LAUGHING IN TV SHOWS  
WHY ARE THERE DOORS ON THE FREEWAY  
WHY ARE THERE SO MANY SVCHOST.EXE RUNNING  
WHY AREN'T THERE ANY COUNTRIES IN ANTARCTICA  
WHY ARE THERE SCARY SOUNDS IN MINECRAFT  
WHY IS THERE KICKING IN MY STOMACH  
WHY ARE THERE TWO SLASHES AFTER HTTP  
WHY ARE THERE CELEBRITIES  
WHY DO SNAKES EXIST  
WHY DO OYSTERS HAVE PEARLS  
WHY ARE DUCKS CALLED DUCKS  
WHY DO THEY CALL IT THE CLAP  
WHY ARE KYLE AND CARTMAN FRIENDS  
WHY IS THERE AN ARROW ON AANG'S HEAD  
WHY ARE TEXT MESSAGES BLUE  
WHY ARE THERE MUSTACHES ON CLOTHES  
WHY ARE THERE MUSTACHES ON CARS  
WHY ARE THERE MUSTACHES EVERYWHERE  
WHY ARE THERE SO MANY BIRDS IN OHIO  
WHY IS THERE SO MUCH RAIN IN OHIO  
WHY IS OHIO WEATHER SO WEIRD

WHY AREN'T ECONOMISTS RICH  
WHY DO AMERICANS CALL IT SOCCER  
WHY ARE MY EARS RINGING  
WHY ARE THERE SO MANY AVENGERS  
WHY ARE THE AVENGERS FIGHTING THE X MEN  
WHY IS WOLVERINE NOT IN THE AVENGERS

WHY ARE THERE SWARMS OF GNATS  
WHY IS THERE PHLEGM  
WHY ARE THERE SO MANY CROWS IN ROCHESTER, MN  
WHY IS PSYCHIC WEAK TO BUG  
WHY DO CHILDREN GET CANCER  
WHY IS POSEIDON ANGRY WITH ODYSSEUS  
WHY IS THERE ICE IN SPACE

# WHY ARE THERE ANTS IN MY LAPTOP

WHY ARE THERE BRIDESMAIDS  
WHY DO DYING PEOPLE REACH UP  
WHY AREN'T THERE VARICOSE ARTERIES  
WHY ARE OLD KUNGONS DIFFERENT

WHY IS EARTH TILTED  
WHY IS SPACE BLACK  
WHY IS OUTER SPACE SO COLD  
WHY ARE THERE PYRAMIDS ON THE MOON  
WHY IS NASA SHUTTING DOWN



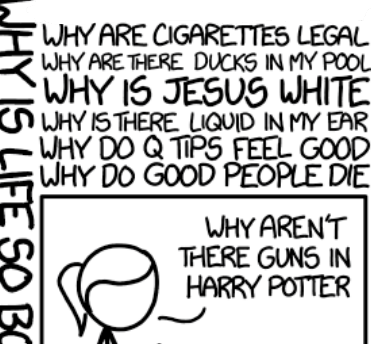
WHY IS THERE AN OWL IN MY BACKYARD  
WHY IS THERE AN OWL OUTSIDE MY WINDOW  
WHY IS THERE AN OWL ON THE DOLLAR BILL  
WHY DO OWLS ATTACK PEOPLE  
WHY ARE AK 47s SO EXPENSIVE  
WHY ARE THERE HELICOPTERS CIRCLING MY HOUSE  
WHY ARE THERE GODS  
WHY ARE THERE TWO SPOCKS

WHY ARE DOGS AFRAID OF FIREWORKS  
WHY IS THERE NO KING IN ENGLAND



WHY ARE THERE TINY SPIDERS IN MY HOUSE  
WHY DO SPIDERS COME INSIDE  
WHY ARE THERE HUGE SPIDERS IN MY HOUSE  
WHY ARE THERE LOTS OF SPIDERS IN MY HOUSE  
WHY ARE THERE SPIDERS IN MY ROOM  
WHY ARE THERE SO MANY SPIDERS IN MY ROOM  
WHY DO SPIDER BITES ITCH  
WHY IS DYING SO SCARY

WHY IS MT VESUVIUS THERE  
WHY DO THEY SAY T MINUS  
WHY ARE THERE OBELISKS  
WHY ARE WRESTLERS ALWAYS WET  
WHY ARE OCEANS BECOMING MORE ACIDIC  
WHY IS ARWEN DYING  
WHY AREN'T MY QUAIL LAYING EGGS  
WHY AREN'T MY QUAIL EGGS HATCHING  
WHY AREN'T THERE ANY FOREIGN MILITARY BASES IN AMERICA



WHY ARE CIGARETTES LEGAL  
WHY ARE THERE DUCKS IN MY POOL  
WHY IS JESUS WHITE  
WHY IS THERE LIQUID IN MY EAR  
WHY DO Q TIPS FEEL GOOD  
WHY DO GOOD PEOPLE DIE  
WHY ARE ULTRASOUNDS IMPORTANT  
WHY ARE ULTRASOUND MACHINES EXPENSIVE  
WHY IS STEALING WRONG

WHY IS PROGRAMMING SO HARD  
WHY IS THERE A 0 OHM RESISTOR  
WHY DO AMERICANS HATE SOCCER  
WHY DO RHYMES SOUND GOOD  
WHY DO TREES DIE  
WHY IS THERE NO SOUND ON CNN  
WHY AREN'T POKEMON REAL  
WHY AREN'T BULLETS SHARP  
WHY DO DREAMS SEEM SO REAL

WHY IS THERE NO GPS IN LAPTOPS  
WHY DO KNEES CLICK  
WHY AREN'T THERE E GRADES  
WHY IS ISOLATION BAD  
WHY DO BOYS LIKE ME  
WHY DON'T BOYS LIKE ME  
WHY IS THERE ALWAYS A JAVA UPDATE  
WHY ARE THERE RED DOTS ON MY THIGHS  
WHY IS LYING GOOD



WHY IS SEX SO IMPORTANT



# Box-and-Whisker Plot

- A box plot is a graphical display showing **S**pread, **O**utliers, **C**enter, and **S**hape (**SOCS**).
- It displays the **5-number summary**: *min*,  $q_1$ , *median*,  $q_3$ , and *max*.

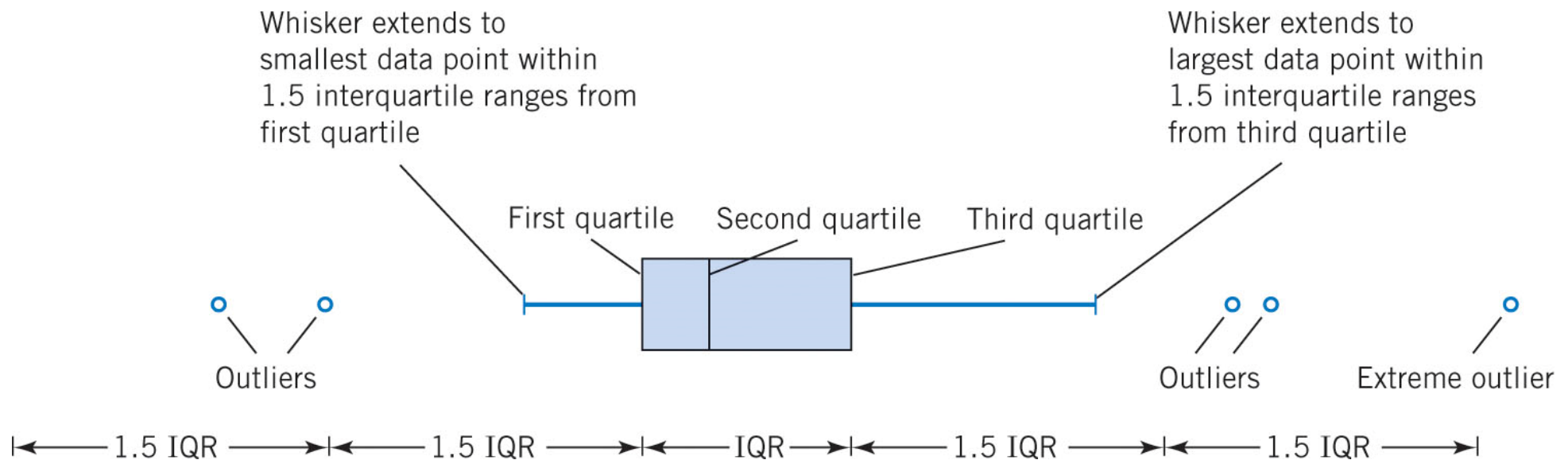
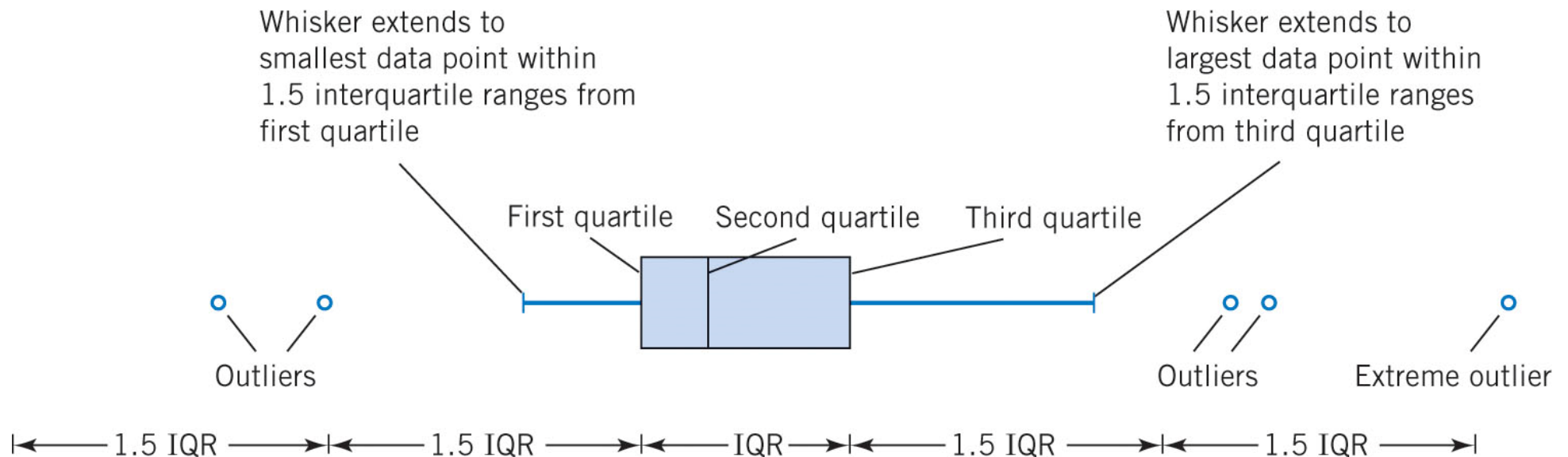


Figure 6-13 Description of a box plot.

# Matlab exercise #2:

- Generate a sample with  $n = 1000$  following **standard normal distribution**
- Calculate **median, first, and third quartiles**
- Calculate **IQR** and find ranges shown below
- Find and count **left and right outliers**
- **Do not use built-in Matlab functions for this!**
- Make box and whisker plot: use **boxplot**



How many right outliers one expects in a sample of  $n=1000$  following normal distribution?

- **% find the third quartile of a standard distribution**
- **norminv(0.75) %ans = 0.6745**
- **% Calculate IQR - Inter Quartale Range**
- **IQR=2.\*norminv(0.75) % 1.3490**
- **% Calculate  $0.5*IQR+1.5*IQR$  - the right whisker position**
- **whisker=0.5.\*IQR+1.5\*IQR %ans = 2.6980**
- **% Find the probability to be above the right whisker**
- **1-normcdf(whisker) %ans = 0.00349**
- **% Find number of right outliers in a sample of 1000 points**
- **1000.\*(1-normcdf(whisker)) %ans = 3.49**



Credit: XKCD  
comics

# WHY ARE THERE SLAVES IN THE BIBLE

WHY DO TWINS HAVE DIFFERENT FINGERPRINTS  
WHY ARE AMERICANS AFRAID OF DRAGONS

WHY IS HTTPS CROSSED OUT IN RED  
WHY IS THERE A LINE THROUGH HTTPS  
WHY IS THERE A RED LINE THROUGH HTTPS ON FACEBOOK  
WHY IS HTTPS IMPORTANT

# QUESTIONS FOUND IN GOOGLE AUTOCOMPLETE



WHY ARE THERE WEEKS  
WHY DO I FEEL DIZZY

WHY DO WHALES JUMP  
WHY ARE WITCHES GREEN  
WHY ARE THERE MIRRORS ABOVE BEDS  
WHY DO I SAY UH  
WHY IS SEA SALT BETTER  
WHY ARE THERE TREES IN THE MIDDLE OF FIELDS  
WHY IS THERE NOT A POKEMON MMO  
WHY IS THERE LAUGHING IN TV SHOWS  
WHY ARE THERE DOORS ON THE FREEWAY  
WHY ARE THERE SO MANY SVCHOST.EXE RUNNING  
WHY AREN'T THERE ANY COUNTRIES IN ANTARCTICA  
WHY ARE THERE SCARY SOUNDS IN MINECRAFT  
WHY IS THERE KICKING IN MY STOMACH  
WHY ARE THERE TWO SLASHES AFTER HTTP  
WHY ARE THERE CELEBRITIES  
WHY DO SNAKES EXIST  
WHY DO OYSTERS HAVE PEARLS  
WHY ARE DUCKS CALLED DUCKS  
WHY DO THEY CALL IT THE CLAP  
WHY ARE KYLE AND CARTMAN FRIENDS  
WHY IS THERE AN ARROW ON AANG'S HEAD  
WHY ARE TEXT MESSAGES BLUE  
WHY ARE THERE MUSTACHES ON CLOTHES  
WHY ARE THERE MUSTACHES ON CARS  
WHY ARE THERE MUSTACHES EVERYWHERE  
WHY ARE THERE SO MANY BIRDS IN OHIO  
WHY IS THERE SO MUCH RAIN IN OHIO  
WHY IS OHIO WEATHER SO WEIRD

WHY DO IGUANAS DIE

WHY AREN'T ECONOMISTS RICH  
WHY DO AMERICANS CALL IT SOCCER  
WHY ARE MY EARS RINGING  
WHY ARE THERE SO MANY AVENGERS  
WHY ARE THE AVENGERS FIGHTING THE X MEN  
WHY IS WOLVERINE NOT IN THE AVENGERS

WHY ARE THERE ANTS IN MY LAPTOP  
WHY IS EARTH TILTED  
WHY IS SPACE BLACK  
WHY IS OUTER SPACE SO COLD  
WHY ARE THERE PYRAMIDS ON THE MOON  
WHY IS NASA SHUTTING DOWN



WHY IS THERE AN OWL IN MY BACKYARD  
WHY IS THERE AN OWL OUTSIDE MY WINDOW  
WHY IS THERE AN OWL ON THE DOLLAR BILL  
WHY DO OWLS ATTACK PEOPLE  
WHY ARE AK 47s SO EXPENSIVE  
WHY ARE THERE HELICOPTERS CIRCLING MY HOUSE  
WHY ARE THERE GODS  
WHY ARE THERE TWO SPOCKS

WHY ARE CIGARETTES LEGAL  
WHY ARE THERE DUCKS IN MY POOL  
WHY IS JESUS WHITE  
WHY IS THERE LIQUID IN MY EAR  
WHY DO Q TIPS FEEL GOOD  
WHY DO GOOD PEOPLE DIE



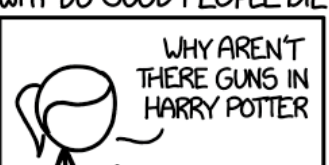
WHY ARE THERE BRIDESMAIDS  
WHY DO DYING PEOPLE REACH UP  
WHY AREN'T THERE VARICOSE ARTERIES  
WHY ARE OLD KUNGONS DIFFERENT  
WHY ARE THERE TINY SPIDERS IN MY HOUSE  
WHY DO SPIDERS COME INSIDE  
WHY ARE THERE HUGE SPIDERS IN MY HOUSE  
WHY ARE THERE LOTS OF SPIDERS IN MY HOUSE  
WHY ARE THERE SPIDERS IN MY ROOM  
WHY ARE THERE SO MANY SPIDERS IN MY ROOM  
WHY DO SPIDER BITES ITCH  
WHY IS DYING SO SCARY

WHY ARE THERE FEMALE MR NIMES



WHY IS MT VESUVIUS THERE  
WHY DO THEY SAY T MINUS  
WHY ARE THERE OBELISKS  
WHY ARE WRESTLERS ALWAYS WET  
WHY ARE OCEANS BECOMING MORE ACIDIC  
WHY IS ARWEN DYING  
WHY AREN'T MY QUAIL LAYING EGGS  
WHY AREN'T MY QUAIL EGGS HATCHING  
WHY AREN'T THERE ANY FOREIGN MILITARY BASES IN AMERICA

WHY IS LIFE SO BORING



WHY ARE DOGS AFRAID OF FIREWORKS  
WHY IS THERE NO KING IN ENGLAND

WHY IS PROGRAMMING SO HARD  
WHY IS THERE A 0 OHM RESISTOR  
WHY DO AMERICANS HATE SOCCER  
WHY DO RHYMES SOUND GOOD  
WHY DO TREES DIE  
WHY IS THERE NO SOUND ON CNN  
WHY AREN'T POKEMON REAL  
WHY AREN'T BULLETS SHARP  
WHY DO DREAMS SEEM SO REAL

WHY IS THERE NO GPS IN LAPTOPS  
WHY DO KNEES CLICK  
WHY AREN'T THERE E GRADES  
WHY IS ISOLATION BAD  
WHY DO BOYS LIKE ME  
WHY DON'T BOYS LIKE ME  
WHY IS THERE ALWAYS A JAVA UPDATE  
WHY ARE THERE RED DOTS ON MY THIGHS  
WHY IS LYING GOOD

WHY IS GPS FREE

WHY ARE ULTRASOUNDS IMPORTANT  
WHY ARE ULTRASOUND MACHINES EXPENSIVE  
WHY IS STEALING WRONG