

**SIGCSE 2017**
Inspire, Innovate, Improve!



Data science for all!



**aspiringminds**
Employability Quantified

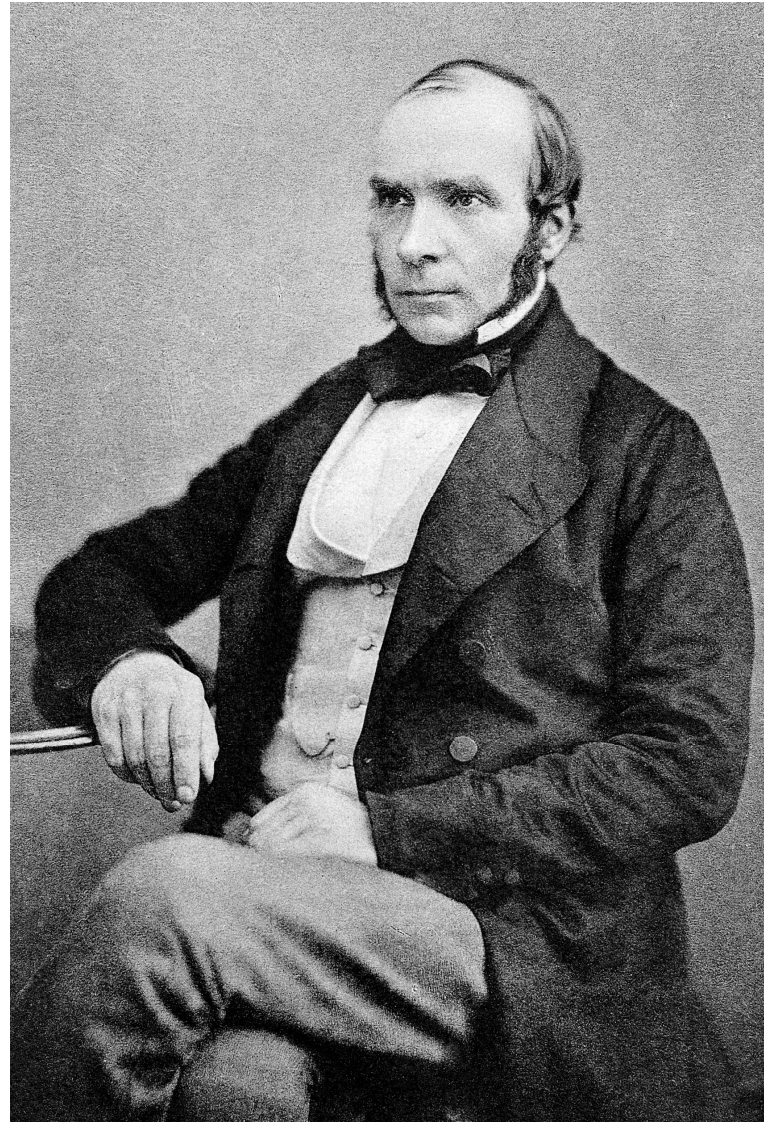
CHOLERA : AN ICKY PROBLEM

CHOLERA IS PRONOUNCED
KOL-cr-uh.

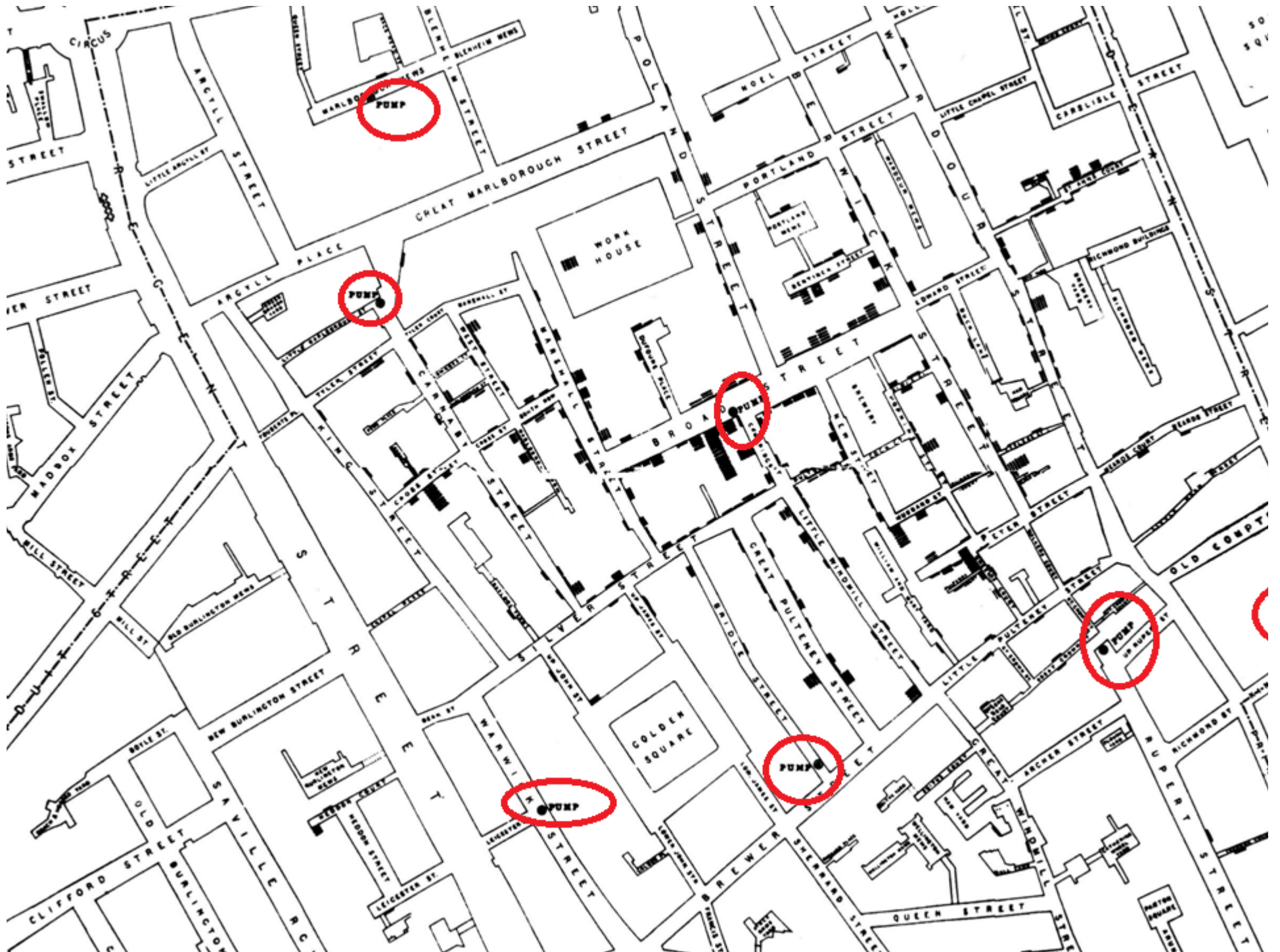
CONTAMINATED MEANS
IMPURE, HARMFUL,
AND UNUSABLE.

MIASMA MEANS A BAD
ATMOSPHERE; GERMS OR
BAD SMELLS OR
POLLUTANTS IN
THE AIR.

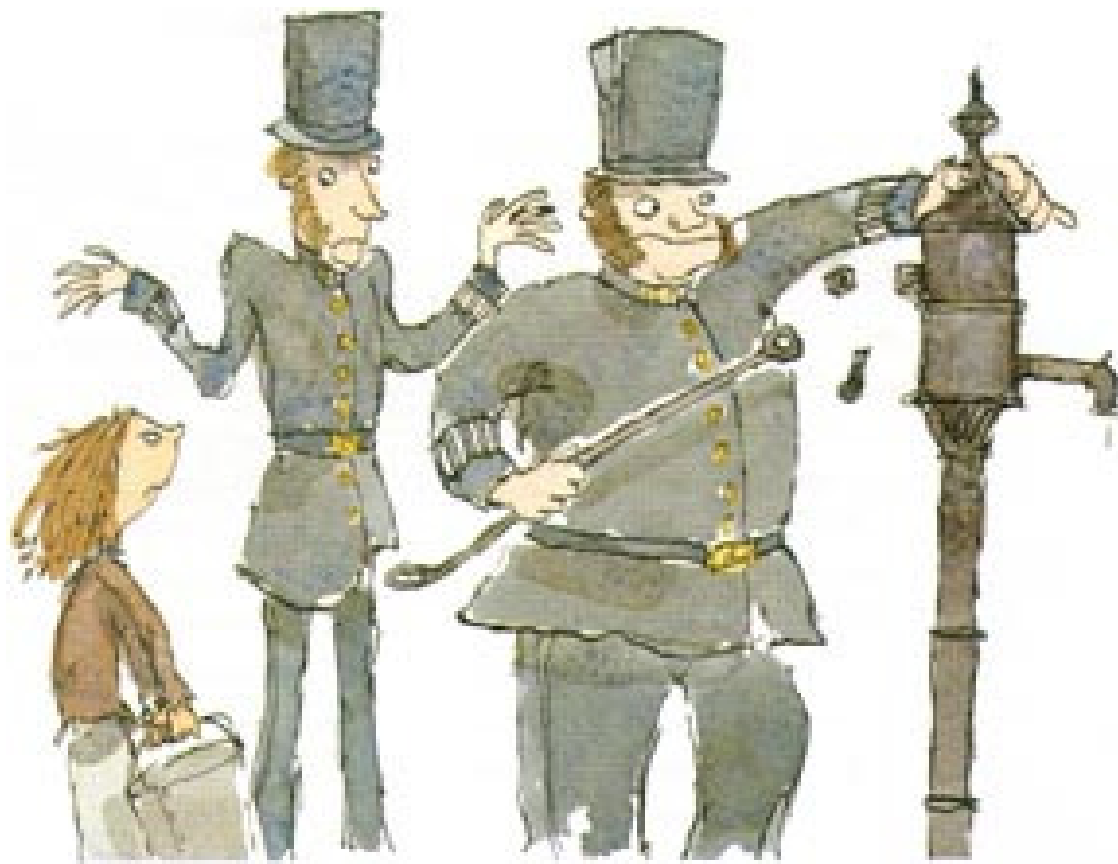
ICK.



John Snow









Oh, oh. There's a problem!



Let us solve this problem!



You!



Our aim



Is to find out what
factors can help
determine the cause
of the sickness

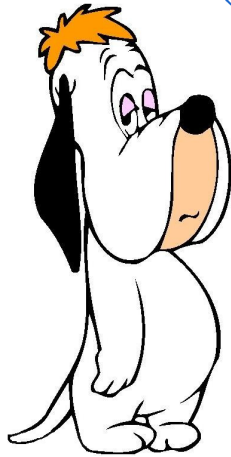


What could be causing this problem?



What could be causing this problem?

??



Same
school?

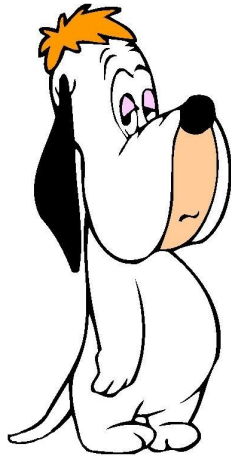
Same age
group?

Same
grocer?

Specific
houses?

Find patterns

Let's look at
details of those
who fell sick



Jim
5 yrs,
Baseball,
Cereals



Nancy
10 yrs,
Swimming,
Meat



Peter
11 yrs,
Swimming
Meat



Meena
15 yrs,
Hockey,
Veggies



Ali
10 yrs,
Swimming,
Sprouts



Dan
9 yrs,
Swimming,
Eggs



Beth
10 yrs,
Swimming,
Meat



Vince
6 years,
Swimming,
Fruits



Meg
11 years,
Basketball,
Eggs

**Finding
patterns**

Let's look at
games played by
those who fell sick



**Finding
patterns**

Jim
Baseball

Nancy
Swimming

Peter
Swimming

Meena
Hockey

Ali
Swimming

Dan
Swimming

Beth
Swimming

Vince
Swimming

Meg
Basketball

Let's look at
games played by
those who fell sick



**Finding
patterns**

Baseball

Swimming

Swimming

Hockey

Swimming

Swimming

Swimming

Swimming

Basketball



Baseball

Basketball

Swimming

Swimming

Swimming

Swimming

Swimming

Swimming

Hockey

Finding patterns

Baseball

Swimming

Swimming

Hockey

Swimming

Swimming

Swimming

Swimming

Basketball



Baseball

Basketball

Swimming

Swimming

Swimming

Swimming

Swimming

Swimming

Hockey

Finding patterns

Baseball	1
Basketball	1
Swimming	6
Hockey	1



Baseball

Basketball

Swimming

Swimming

Swimming

Swimming

Swimming

Swimming

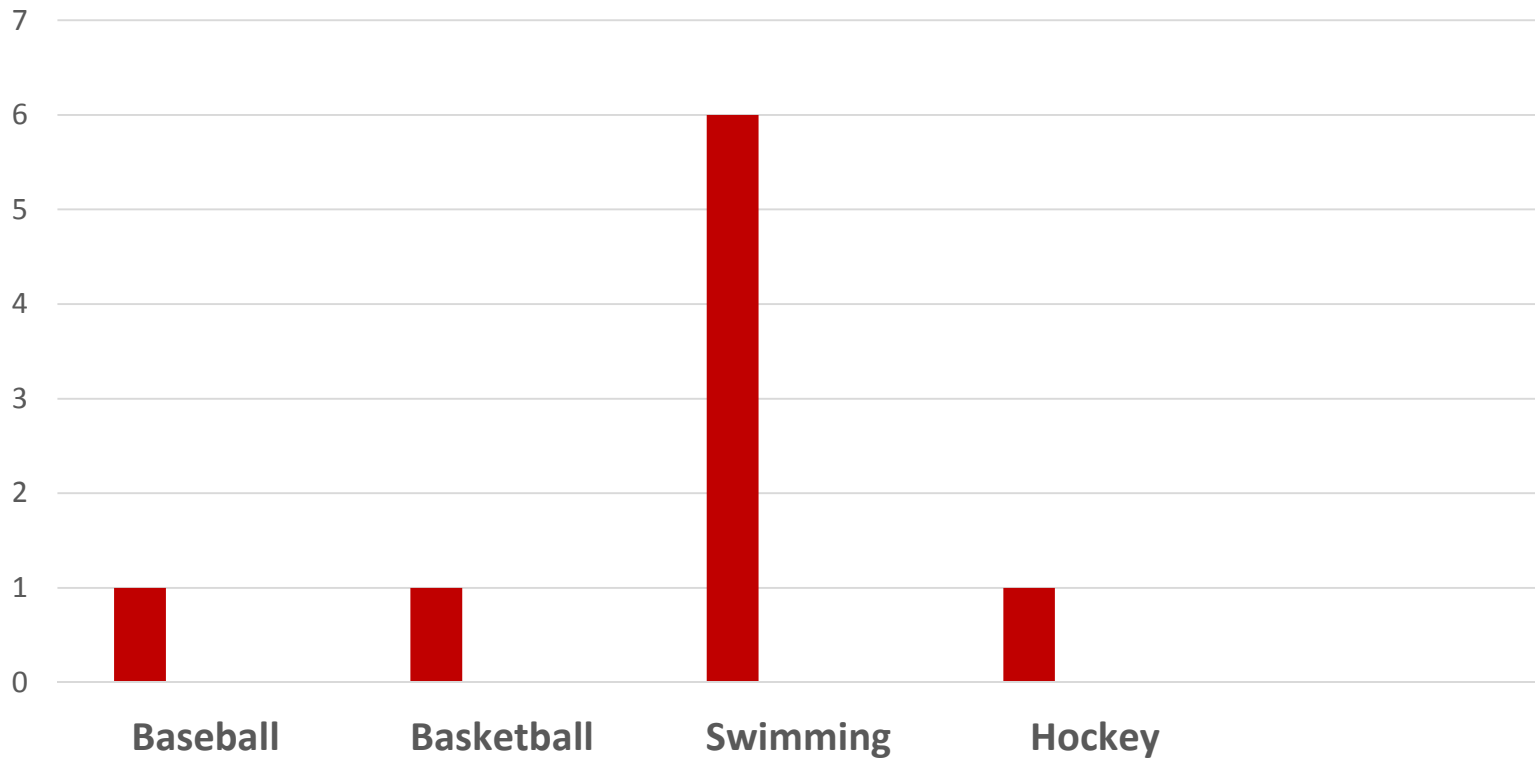
Hockey

Finding patterns

Baseball	1
Basketball	1
Swimming	6
Hockey	1

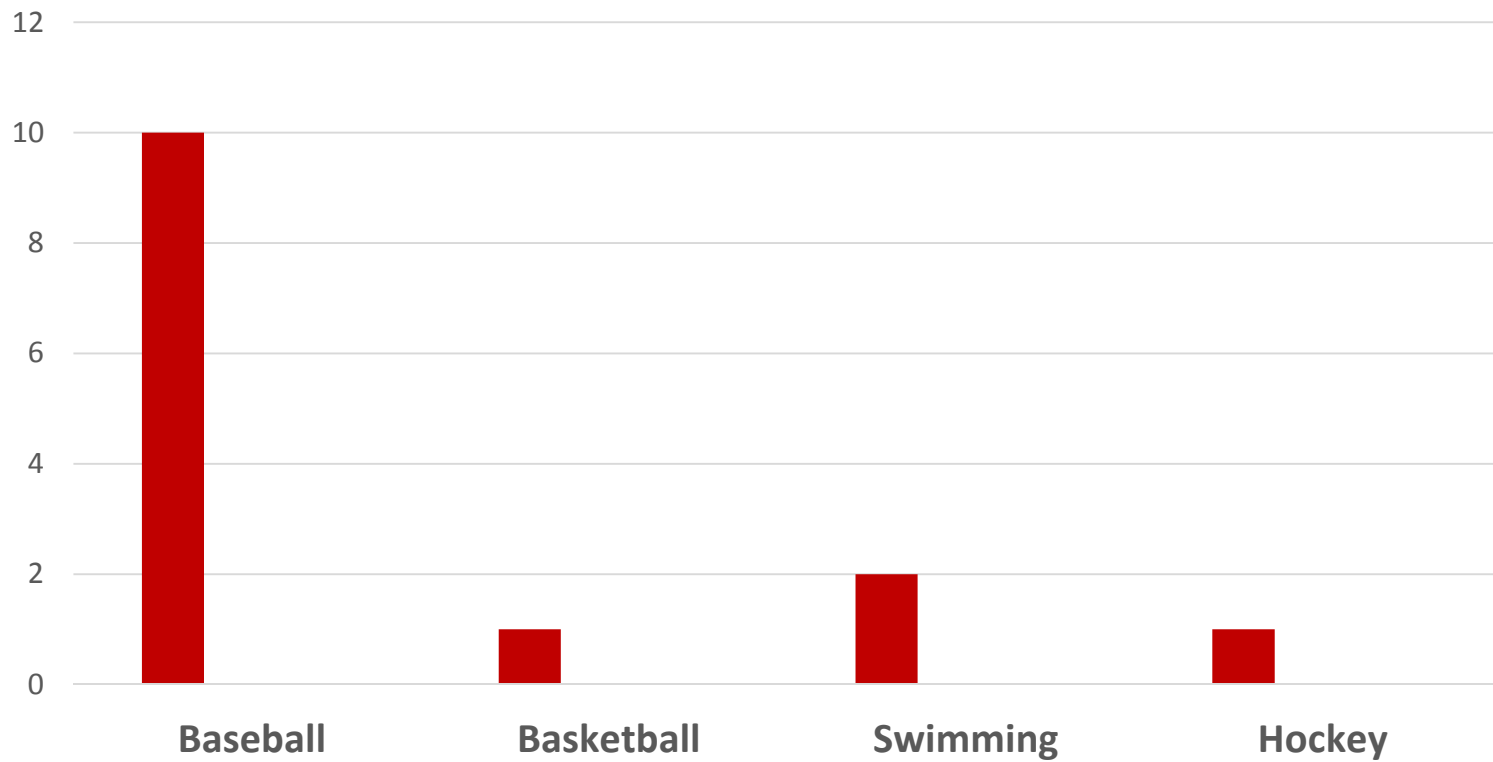
Baseball	1
Basketball	1
Swimming	6
Hockey	1

Finding patterns



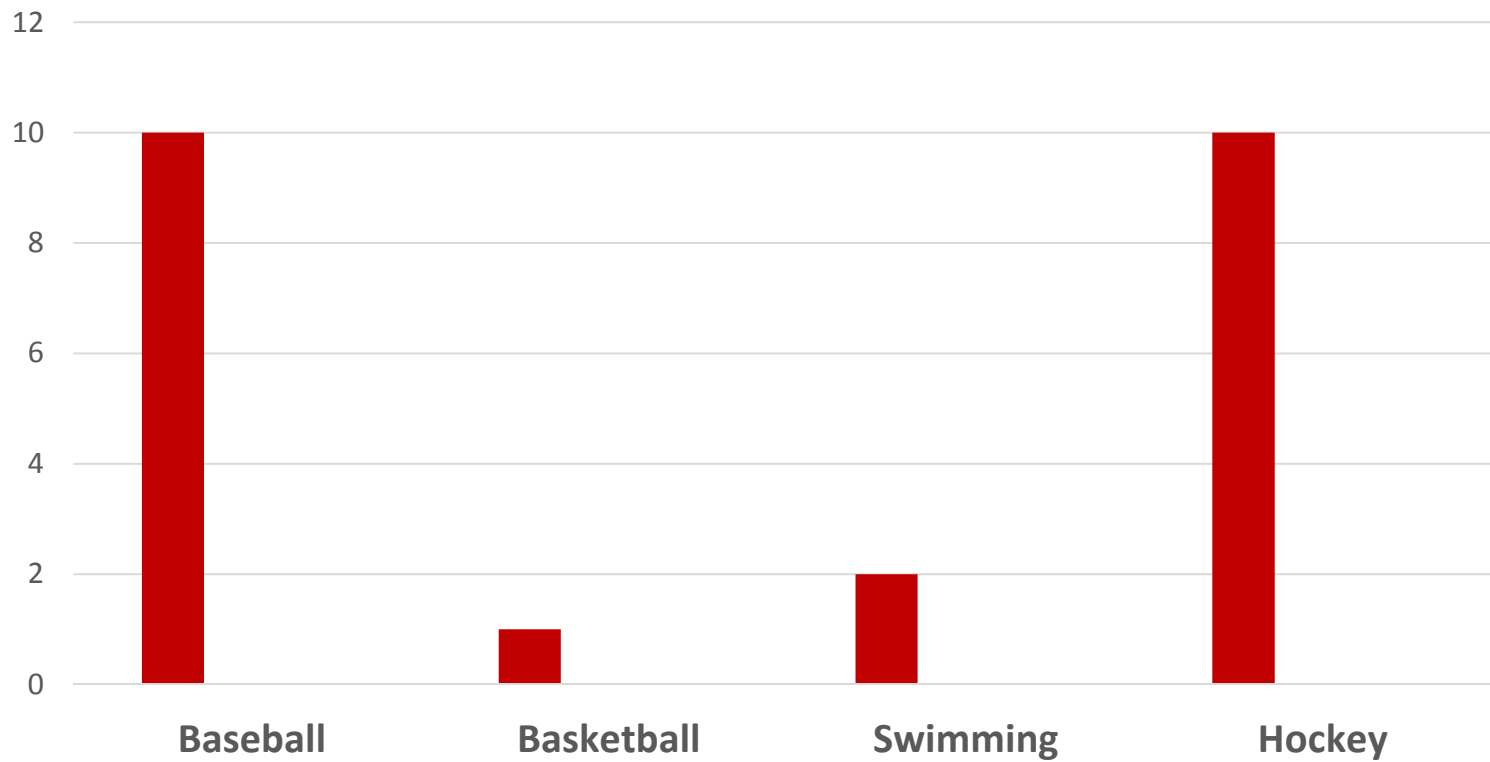
Finding patterns

?



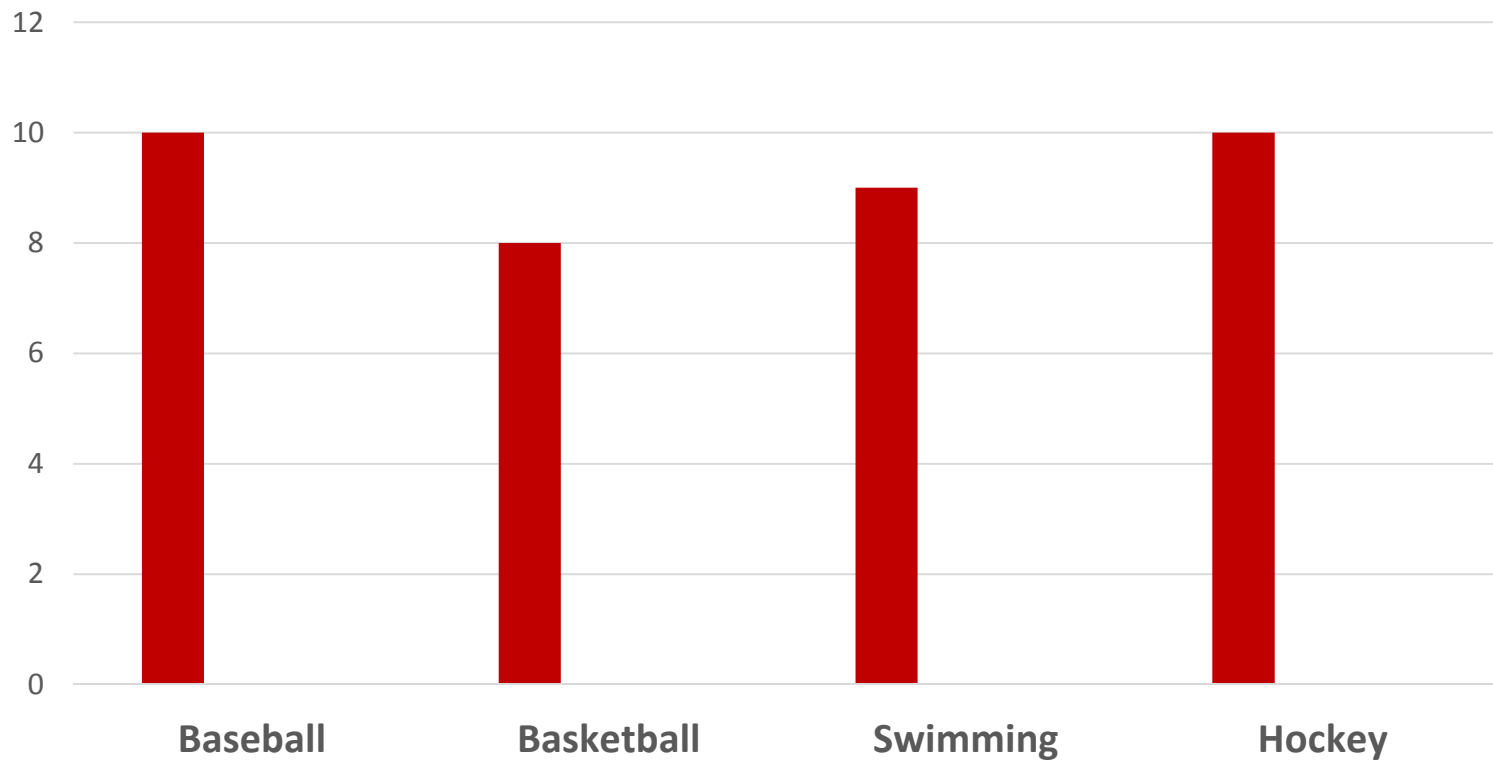
Finding patterns

?



Finding patterns

?



What pattern did we find so far?

Games played by those
who fell sick



Baseball	1
Basketball	1
Swimming	6
Hockey	1

**Those who swim regularly seemed to
have fallen sick the most**

Recap - Our aim



Is to find out what factors can help determine the cause of the sickness



Assumption



Let us assume that age is the only factor we found which “explains” the sickness



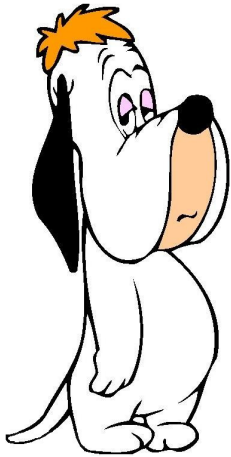
What then?



How can you use this information?

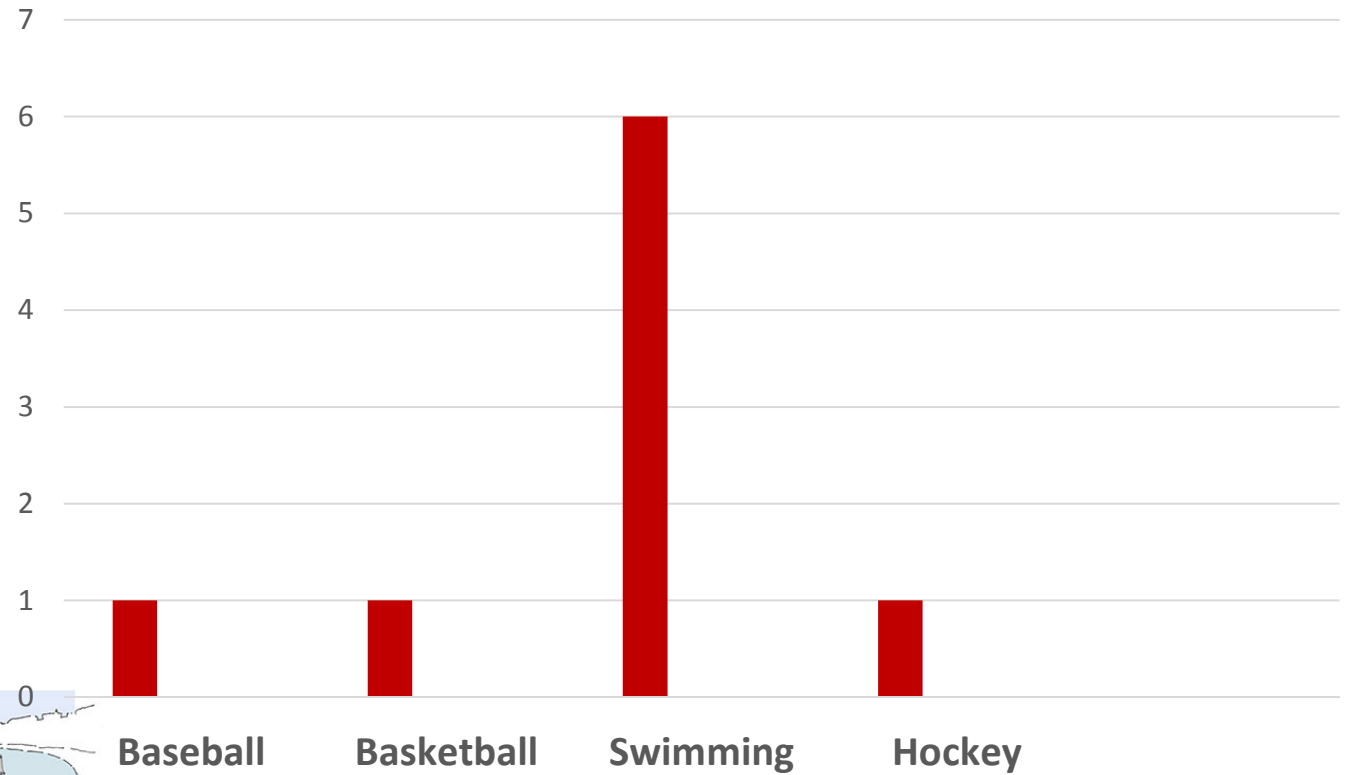


Will she be sick soon?

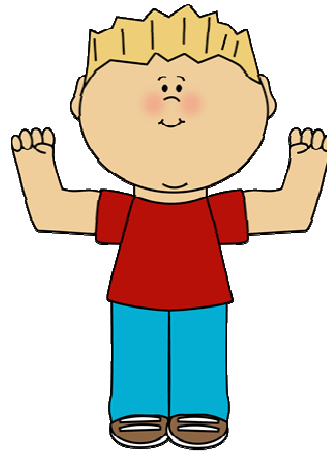


**Rachel
Swims**

Remember



What about him?



Tarun
Basketball

What is data?



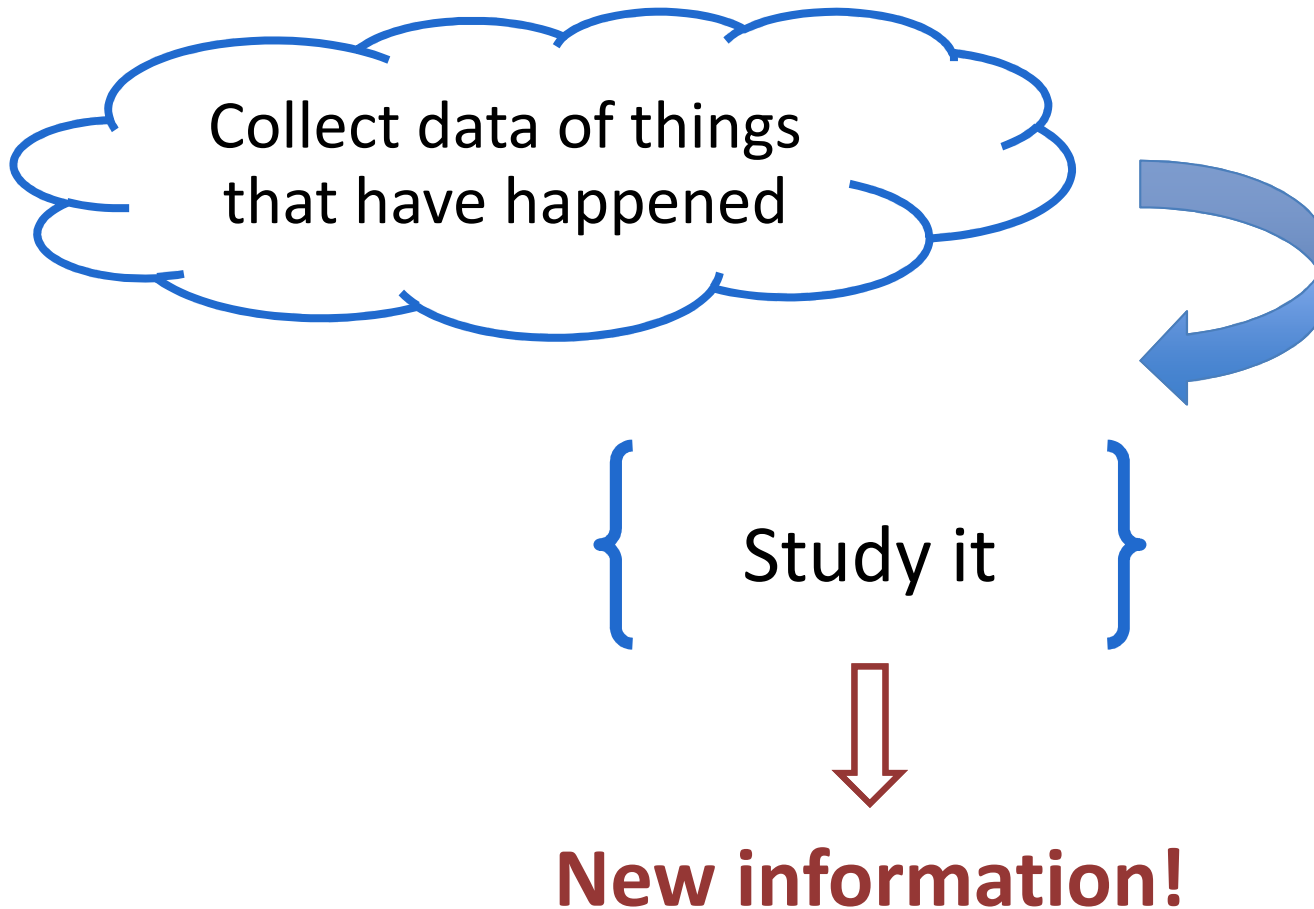
Units of
information

What is data?

Facts!

Anything you
can measure

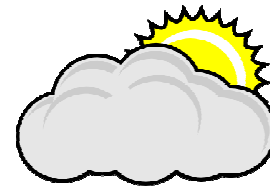
What is data science?



Agriculture



- Weather information
- Soil information
- Geographical information



Ways to improve crop growth

Shopping



- What you have bought
- When you have bought



Help you find more things you like

**What would you like
to predict?**