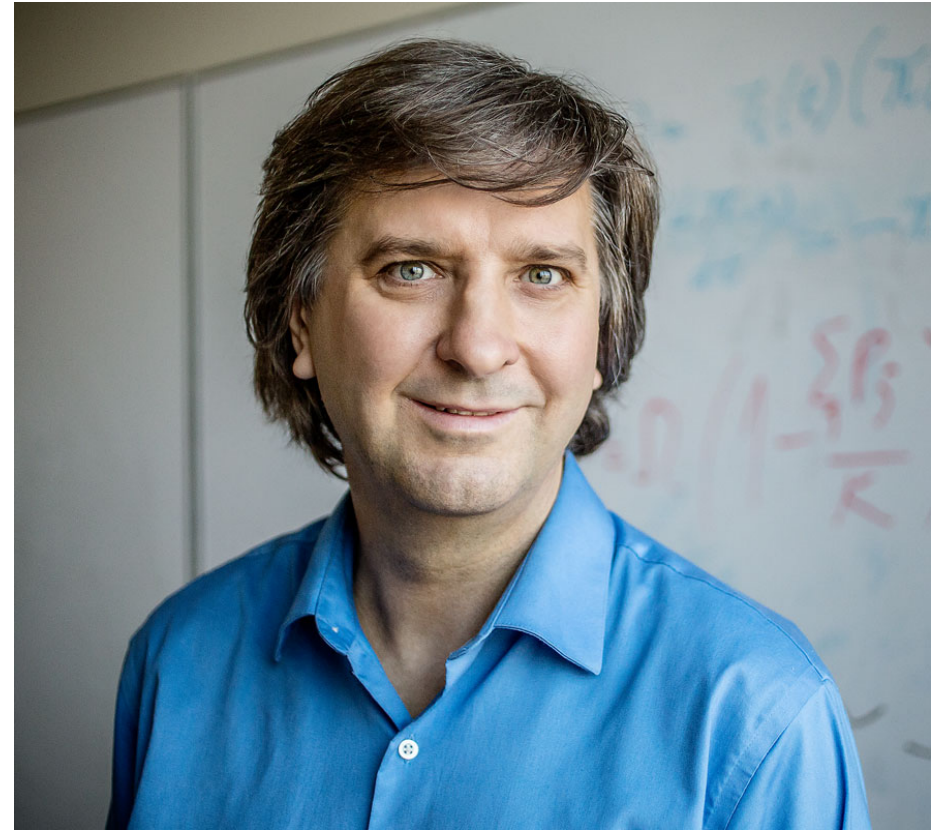


# BIOE 505: Computational Bioengineering

What this class is all about?

# Instructor

- Name: **Sergei Maslov**
- **Professor of Bioengineering, Physics, Carl R. Woese Institute for Genomic Biology, and National Center for Supercomputing Applications**
- Office: 3406 Carl Woese Institute for Genomic Biology and sometimes 3146C Everitt Laboratory (both by appointment)
- E-mail: [maslov@illinois.edu](mailto:maslov@illinois.edu)
- Phone: 217-265-5705



# Questions and Suggestions:

[maslov@Illinois.edu](mailto:maslov@Illinois.edu)

Start subject with [BIOE505]

# Grading

- Midterm exam 40%
- Final exam 60%
- Homework (ungraded) will be posted online. Solutions will be posted in a week.
- Homework will build on topics covered in lectures and will consist of problem sets related to topics covered in lectures
- Useful to prepare for exams

# Course Website

<https://courses.engr.illinois.edu/bioe505>

Grades will be on

<https://my.bioen.illinois.edu/gradebook>

← → ↻ 🏠 [courses.grainger.illinois.edu/bioe505/fa2019/index.html](https://courses.grainger.illinois.edu/bioe505/fa2019/index.html)

📱 Apps ★ ⌚ 📁 Recent 📁 Group 📁 Library@RC 📄 Ludwig • Find your... 🛠 WebEdit 🌐 BIOE

← → ↻ 🏠 [courses.grainger.illinois.edu/bioe505/fa2019/schedule.html](https://courses.grainger.illinois.edu/bioe505/fa2019/schedule.html) ☆ 📁 I 📅 15 🗣 2h

## BIOE 505 - Computational Bioengineering

📱 Apps ★ ⌚ 📁 Recent 📁 Group 📁 Library@RC 📄 Ludwig • Find your... 🛠 WebEdit 🌐 BIOE 505 | Prof. Ser... 🛠 BIOE 505 | Course E...

## Schedule

## BIOE 505 - Computational Bioengineering

### Instructor

Sergei Maslov: [maslov@illinois.edu](mailto:maslov@illinois.edu)  
Office: IGB 3406  
Office hours: by appointment

### Logistics

Tuesdays: 12:00AM - 1:50AM  
Thursdays: 12:00AM - 1:50PM

106B8 Engineering Hall

I WANT YOU TO BRING YOUR i-CLIP

I WANT YOU TO BRING YOUR OWN  
LAPTOPS SHOULD HAVE MATLAB  
STATISTICS AND MACHINE LEARN

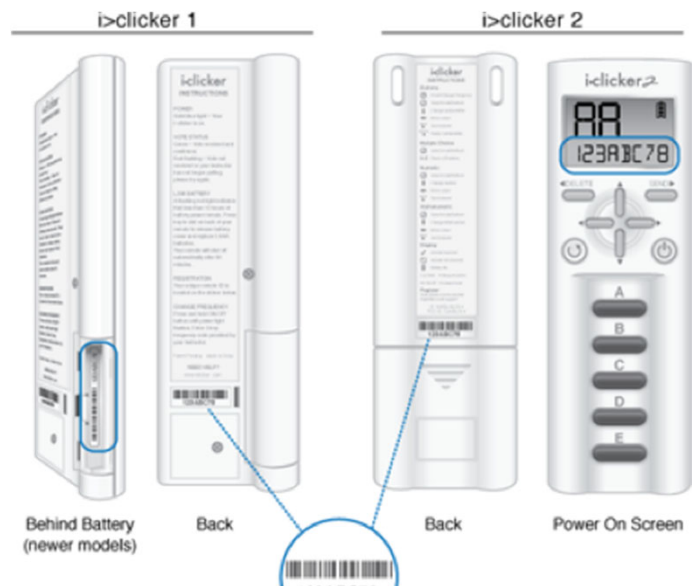
### Description

#	Date	Topics	Slides	Matlab	Homework	Exams
1	Aug 27					
2	Aug 29					

# Bring your iClickers to my lectures

- **Who knows what is an iClicker?**
- **Show of hands: who has an iClicker?**
- I would like you all to have an iClicker and bring it to every class. On **amazon.com** a new **iClicker** (1<sup>st</sup> generation is OK) costs around \$40. It is also sold at UIUC Bookstore. The used ones are cheaper.
- An alternative solution is using a mobile app:  
<https://www.iclicker.com/students/apps-and-remotes/apps>

- Your answers **WILL NOT** be used for grading.  
I need them to see if I lost some of you and what could I rephrase to better explain the material



# We will use Matlab in class

- Bring **your laptops to class**
- **Poll: who has Matlab?**
- Need to have **Matlab installed** and know the basic user interface (inline commands, plotting)
- We will use **Statistics and Machine Learning Toolbox and Bioinformatics Toolboxes**
- You can use CITRIX for UIUC students and connect to EWS Windows Lab Software
- **.m files and .mat** with Matlab commands and data **will be on the website** after the lecture

# Who has Matlab?

- A. Have on my own laptop
- B. Plan to use CITRIX
- C. I don't have matlab
- D. I don't know yet
- E. I will never use Matlab! Over my dead body

Get your i-clickers



Possible alternative to purchasing Matlab and toolboxes is to use campus resources.

Both Engineering Workstations (EWS) and ACES computers have Matlab. I don't think all of them offer the statistics and bioinformatics toolboxes (EWS should, ACES computers may not..).

See the following to access:

**Citrix for EWS, Matlab, and ACES computers -- links for all**

<https://it.engineering.illinois.edu/ews/lab-information/remote-connections/connecting-citrix>

<https://it.engineering.illinois.edu/services/instructional-services/remote-connections-citrix>

**Accessing Engineering Workstations (EWS)**

<https://it.engineering.illinois.edu/ews>

**Accessing ACES Academic Computing Workstations**

<http://acf.aces.illinois.edu/remote/>

<http://acf.aces.illinois.edu/remote/pc.html>

To access off campus use:

**CISCO Virtual Private Network -- For off-campus access to campus computer and network resources (software programs, files saved on the network, etc.)**

<https://techservices.illinois.edu/services/virtual-private-networking-vpn/download-and-set-up-the-vpn-client>

**CISCO VPN CLIENT**

<https://webstore.illinois.edu/shop/product.aspx?zpid=2600>

**CISCO AnyConnect VPN**

<https://webstore.illinois.edu/shop/product.aspx?zpid=1222>

# What will you learn in this course?

- Basics of probability and statistics
  - Basic concepts of probability, Bayes theorem
  - Discrete and continuous probability distributions
  - Multivariate statistics
  - Sampling distributions
  - Parameter estimation
  - Hypothesis testing
  - Regression
- How it is applied to biological data
  - Basics of genomics
  - Systems biology (gene expression, networks)

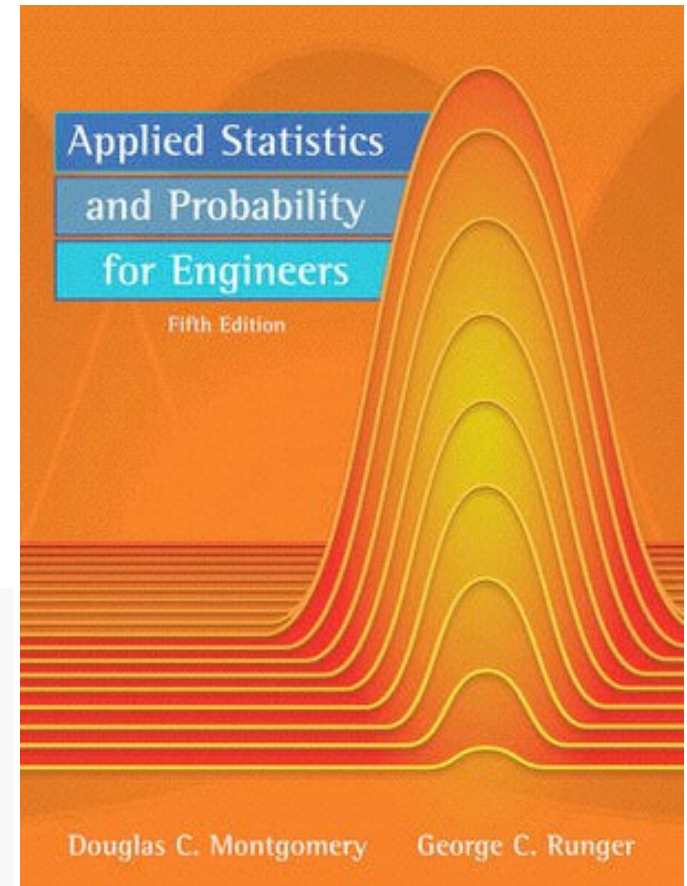
# The main Probability/Statistics Textbook

## Applied Statistics and Probability for Engineers, 5th Edition

*D. C. Montgomery and G. C. Runger*  
John Wiley & Sons, Inc. (2011)

You can also use other editions from  
4<sup>th</sup> (2007) to 6<sup>th</sup> (2014)

5<sup>th</sup> edition is available for free  
at our library



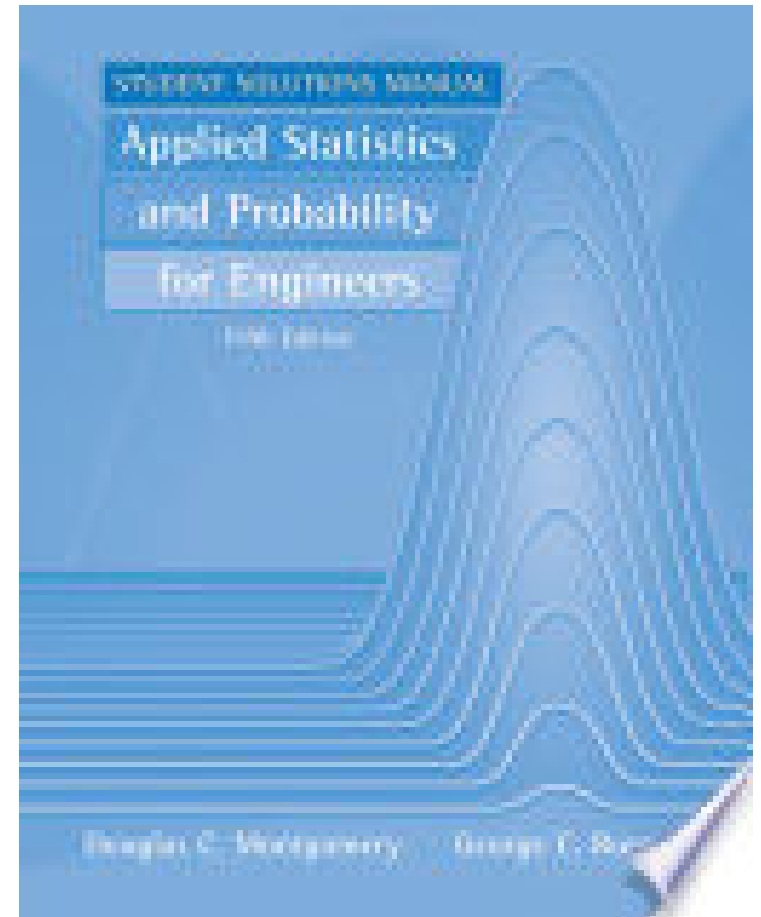
# Problems for our main Probability/Statistics Textbook

## Student Solutions Manual Applied Statistics and Probability for Engineers, 5th Edition

*D. C. Montgomery and G. C. Runger*  
John Wiley & Sons, Inc. (2010)

You can also use other editions from  
4<sup>th</sup> (2007) to 6<sup>th</sup> (2014)

5<sup>th</sup> edition is available  
for free at our library



# Probability/Statistics for Bioengineering with Matlab exercises

## Statistics for Bioengineering Sciences

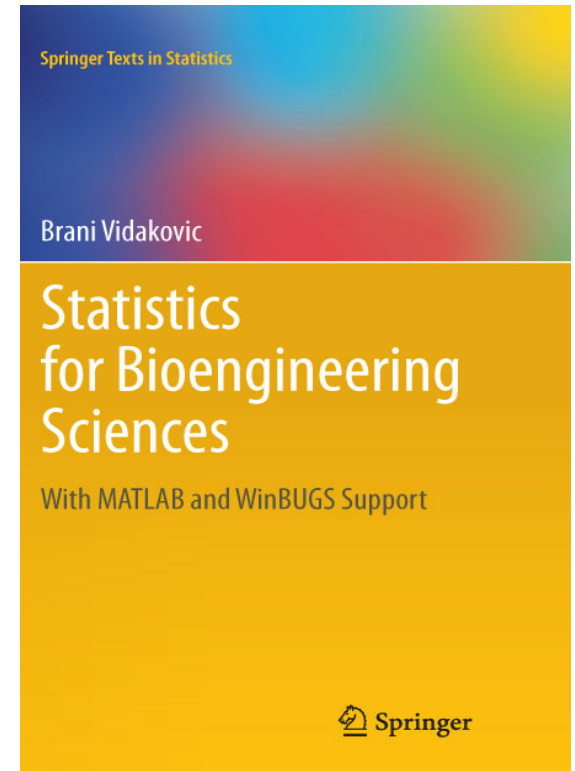
with MATLAB and WinBUGS Support

*Brani Vidakovic*

*Department of Biomedical Engineering, Georgia Tech*

*(2011) Springer, New York*

*It is constantly updated with the newest version at the link  
below.*



*Free as a PDF eBook at*

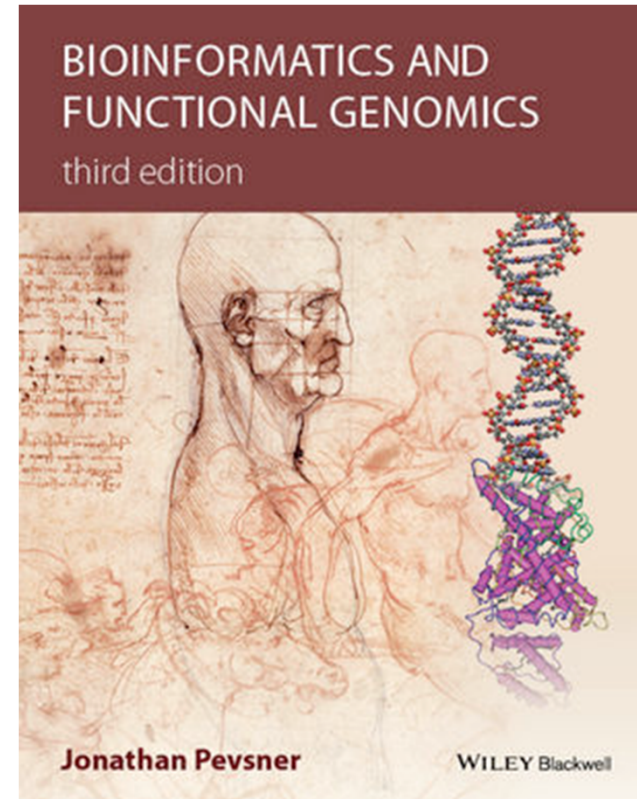
*<http://statbook.gatech.edu/statb4.pdf>*

*Matlab exercises and datasets are at*

*<http://springer.bme.gatech.edu>*

# Genomics/Systems Biology Textbook

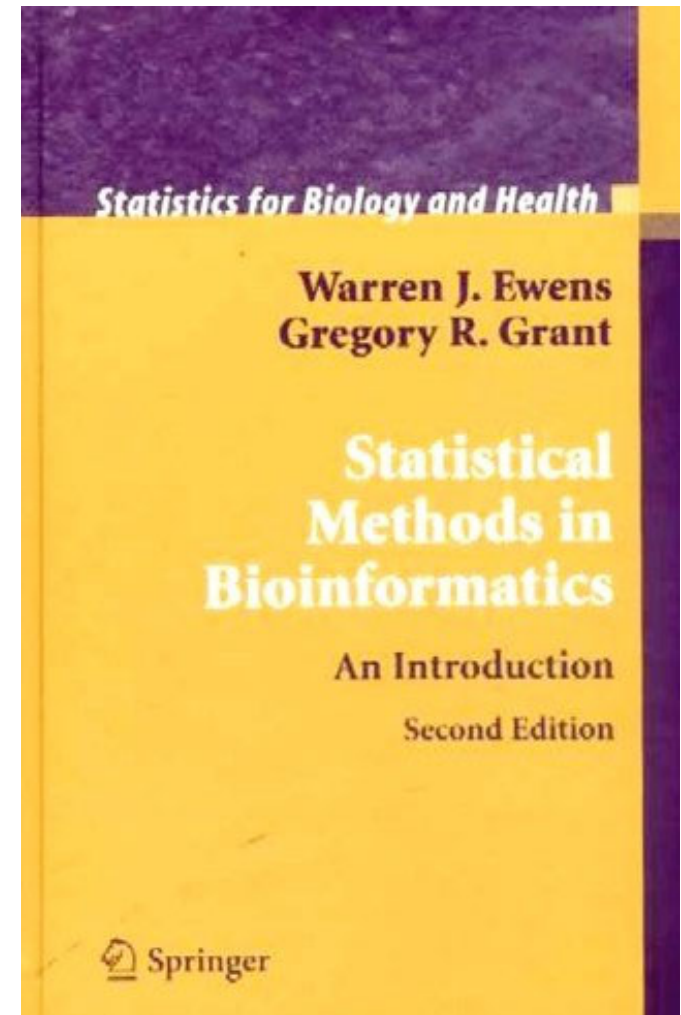
- *J Pevsner*  
***Bioinformatics and functional genomics***  
Wiley-Blackwell,  
2<sup>nd</sup> edition [2009] *exists in electronic form*  
3<sup>rd</sup> edition [2015] *has up-to-date*  
*information on NGS: RECOMMENDED*  
*(about \$60 on amazon)*
- *2<sup>nd</sup> edition is available for free*  
*in electronic form in our library*





# Another Bioinformatics/Statistics Textbook

- *Ewens, WJ and Grant, GR Statistical Methods in Bioinformatics: An Introduction, 2nd ed, Springer, 2005.*
- *2<sup>nd</sup> edition as PDF eBook*



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 AUTOCOMLETE



WHY ARE THERE WEEKS  
WHY DO I FEEL DIZZY

WHY AREN'T ECONOMISTS RICH

WHY ARE THERE SWARMS OF GNATS  
WHY IS THERE PHLEGM  
WHY ARE THERE SO MANY CROWS IN ROCHESTER, MN

WHY DO AMERICANS CALL IT SOCCER

WHY IS PSYCHIC WEAK TO BUG

WHY ARE MY EARS RINGING

WHY DO CHILDREN GET CANCER

WHY ARE THERE SO MANY AVENGERS

WHY IS POSEIDON ANGRY WITH ODYSSEUS

WHY ARE THE AVENGERS FIGHTING THE X MEN

WHY IS THERE ICE IN SPACE

WHY IS WOLVERINE NOT IN THE AVENGERS

# WHY ARE THERE ANTS IN MY LAPTOP

WHY IS EARTH TILTED

WHY ARE THERE GHOSTS

WHY IS THERE AN OWL IN MY BACKYARD

WHY IS SPACE BLACK

WHY ARE THERE GHOSTS

WHY IS THERE AN OWL OUTSIDE MY WINDOW

WHY IS OUTER SPACE SO COLD

WHY ARE THERE GHOSTS

WHY IS THERE AN OWL ON THE DOLLAR BILL

WHY ARE THERE PYRAMIDS ON THE MOON

WHY ARE THERE GHOSTS

WHY DO OWLS ATTACK PEOPLE

WHY IS NASA SHUTTING DOWN

WHY ARE THERE GHOSTS

WHY ARE AK 47s SO EXPENSIVE

WHY ARE THERE MALE AND FEMALE BIKES

WHY ARE THERE GHOSTS

WHY ARE THERE HELICOPTERS CIRCLING MY HOUSE

WHY ARE THERE TINY SPIDERS IN MY HOUSE

WHY ARE THERE GHOSTS

WHY ARE THERE GODS

WHY DO SPIDERS COME INSIDE

WHY ARE THERE GHOSTS

WHY ARE THERE TWO SPOCKS

WHY ARE THERE HUGE SPIDERS IN MY HOUSE

WHY ARE THERE GHOSTS

WHY IS LIFE SO BORING

WHY ARE THERE LOTS OF SPIDERS IN MY HOUSE

WHY ARE THERE GHOSTS

WHY ARE CIGARETTES LEGAL

WHY ARE THERE SPIDERS IN MY ROOM

WHY ARE THERE GHOSTS

WHY ARE THERE DUCKS IN MY POOL

WHY ARE THERE SO MANY SPIDERS IN MY ROOM

WHY ARE THERE GHOSTS

WHY IS JESUS WHITE

WHY DO SPIDER BITES ITCH

WHY ARE THERE GHOSTS

WHY IS THERE LIQUID IN MY EAR

WHY IS DYING SO SCARY

WHY ARE THERE GHOSTS

WHY DO Q TIPS FEEL GOOD

WHY DO WHALES JUMP  
WHY ARE WITCHES GREEN  
WHY ARE THERE MIRRORS ABOVE BEDS

WHY AREN'T THERE

WHY DO I SAY UH

WHY DO IGUANAS DIE

WHY IS SEA SALT BETTER

DINOSAUR GHOSTS

WHY ARE THERE TREES IN THE MIDDLE OF FIELDS

WHY AREN'T THERE

WHY IS THERE NOT A POKEMON MMO

WHY AREN'T THERE

WHY IS THERE LAUGHING IN TV SHOWS

WHY AREN'T THERE

WHY ARE THERE DOORS ON THE FREEWAY

WHY AREN'T THERE

WHY ARE THERE SO MANY SVCHOST.EXE RUNNING

WHY AREN'T THERE

WHY AREN'T THERE ANY COUNTRIES IN ANTARCTICA

WHY AREN'T THERE

WHY ARE THERE SCARY SOUNDS IN MINECRAFT

WHY AREN'T THERE

WHY IS THERE KICKING IN MY STOMACH

WHY AREN'T THERE

WHY ARE THERE TWO SLASHES AFTER HTTP

WHY AREN'T THERE

WHY ARE THERE CELEBRITIES

WHY AREN'T THERE

WHY DO SNAKES EXIST

WHY AREN'T THERE

WHY DO OYSTERS HAVE PEARLS

WHY AREN'T THERE

WHY ARE DUCKS CALLED DUCKS

WHY AREN'T THERE

WHY DO THEY CALL IT THE CLAP

WHY AREN'T THERE

WHY ARE KYLE AND CARTMAN FRIENDS

WHY AREN'T THERE

WHY IS THERE AN ARROW ON AANG'S HEAD

WHY AREN'T THERE

WHY ARE TEXT MESSAGES BLUE

WHY AREN'T THERE

WHY ARE THERE MUSTACHES ON CLOTHES

WHY AREN'T THERE

WHY ARE THERE MUSTACHES ON CARS

WHY AREN'T THERE

WHY ARE THERE MUSTACHES EVERYWHERE

WHY AREN'T THERE

WHY ARE THERE SO MANY BIRDS IN OHIO

WHY AREN'T THERE

WHY IS THERE SO MUCH RAIN IN OHIO

WHY AREN'T THERE

WHY IS OHIO WEATHER SO WEIRD

WHY AREN'T THERE

WHY ARE THERE BRIDESMAIDS

WHY AREN'T THERE



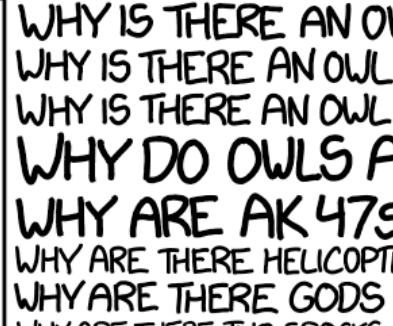
WHY ARE THERE BRIDESMAIDS  
WHY DO DYING PEOPLE REACH UP  
WHY AREN'T THERE VARICOSE ARTERIES  
WHY ARE OLD KUNGONS DIFFERENT

WHY IS PROGRAMMING SO HARD  
WHY IS THERE A 0 OHM RESISTOR  
WHY DO AMERICANS HATE SOCCER  
WHY DO RHYMES SOUND GOOD

WHY IS THERE HELL IF GOD FORGIVES  
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 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 THERE GHOSTS  
WHY ARE THERE GHOSTS  
WHY ARE THERE GHOSTS  
WHY ARE THERE GHOSTS



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 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  
WHY AREN'T MY ARMS GROWING  
WHY ARE THERE WEEKS  
WHY DO I FEEL DIZZY  
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 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 IS LIFE SO BORING  
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 AREN'T MY ARMS GROWING  
WHY AREN'T THERE GUNS IN HARRY POTTER  
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 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  
WHY AREN'T MY ARMS GROWING  
WHY ARE THERE WEEKS  
WHY DO I FEEL DIZZY  
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 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 IS LIFE SO BORING  
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 AREN'T MY ARMS GROWING  
WHY AREN'T THERE GUNS IN HARRY POTTER  
WHY AREN'T MY QUAIL LAYING EGGS  
WHY AREN'T MY QUAIL EGGS HATCHING  
WHY AREN'T THERE ANY FOREIGN MILITARY BASES IN AMERICA



This course is about **biological data**  
and **probability theory and statistics**  
concepts needed for its analysis

# What biological data will be discussed?

Will be covered in lectures or Matlab exercises:

- Genomic data: strings of letters ACGT
- Gene Expression data: messenger RNA copy numbers transcribed from genes
- Proteomic data: protein abundances
- Network data: pairs of interacting genes or proteins and protein-protein interaction strengths

Will not be covered:

- Imaging data such as e.g. fMRI brain scans, Brain connectome data, Ecosystem dynamics data

Why do you need  
probability and statistics  
to analyze  
modern biological data?

## Definition of **probability theory** by Encyclopedia Britannica

a branch of mathematics concerned  
with the analysis of **random  
phenomena**

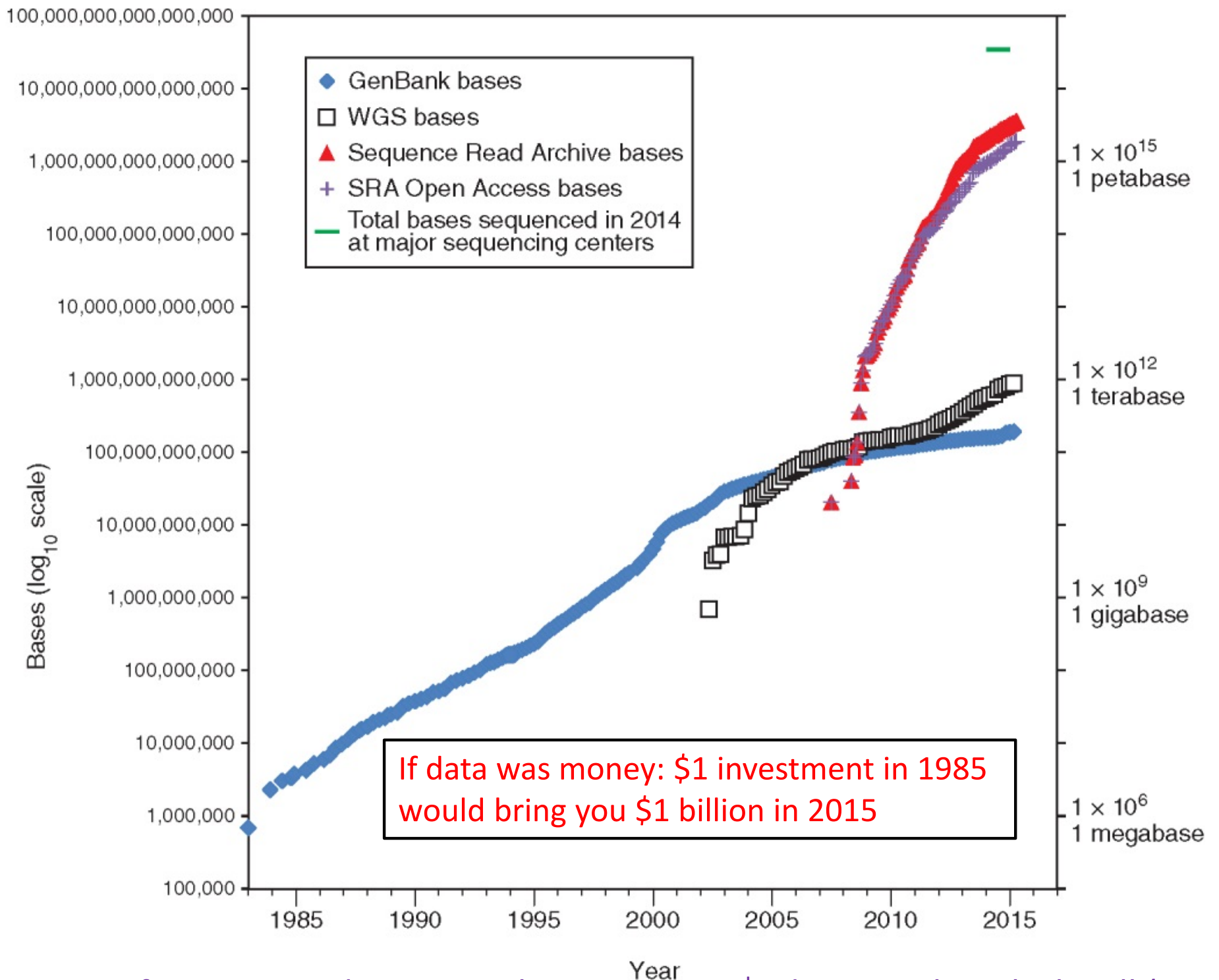
## Definition of ***statistics*** by Merriam-Webster

*1* : a branch of mathematics dealing with the  
collection, analysis, interpretation, and  
presentation of **masses of numerical data**

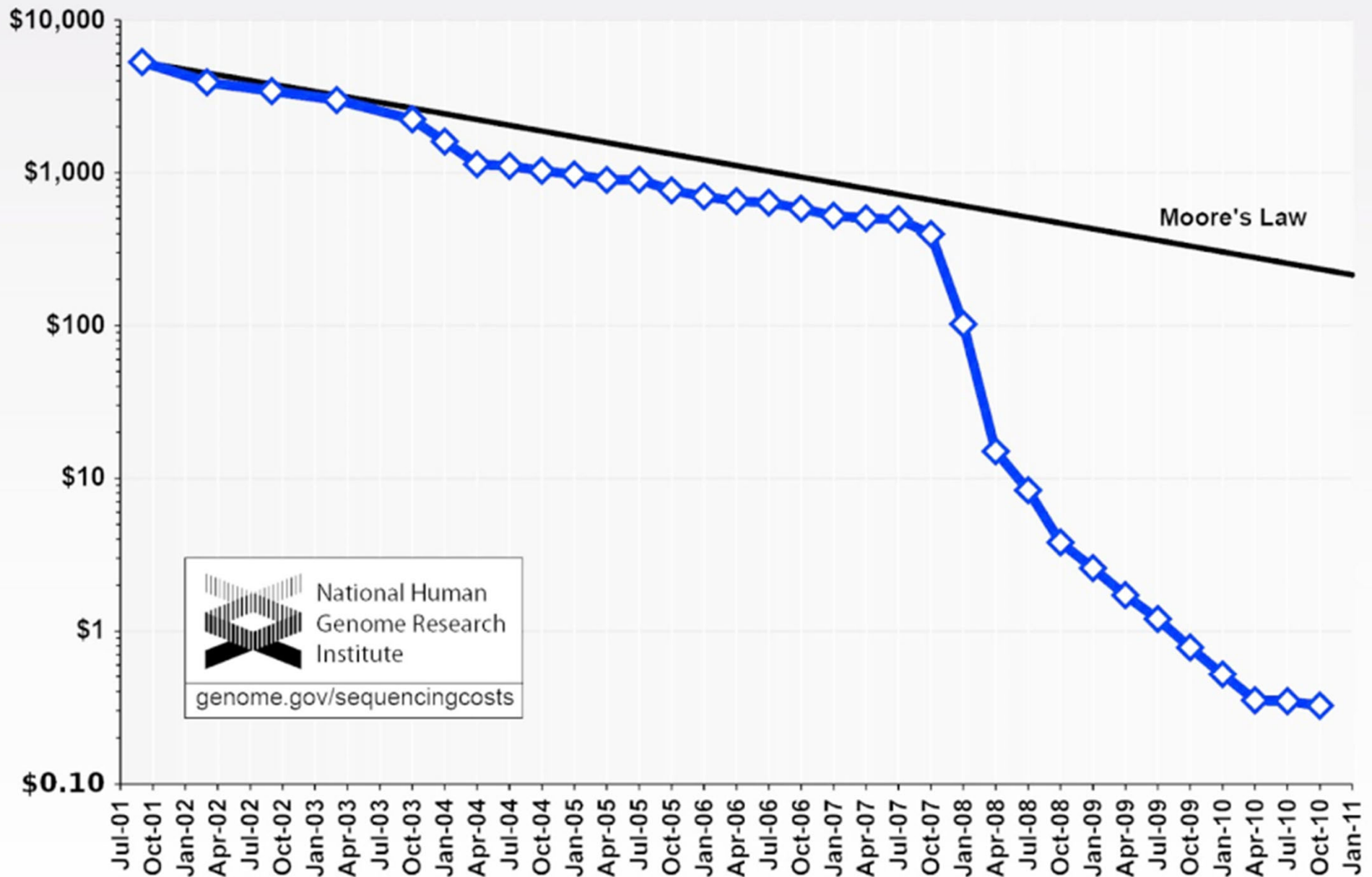
...

Why do you need  
probability and statistics  
to analyze  
modern biological data?

Reason 1:  
Biology now has Lots of Data



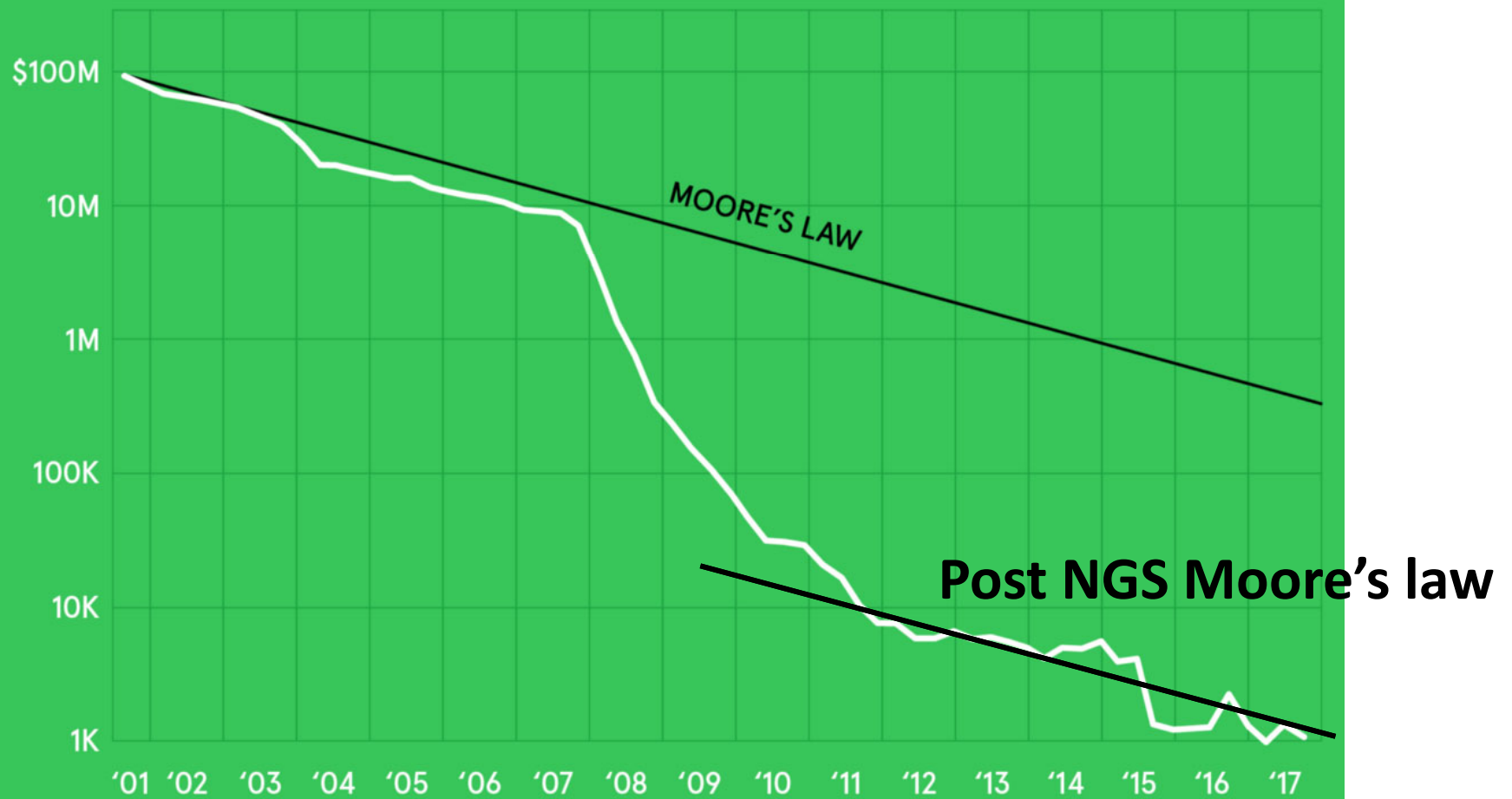
# Cost per Megabase of DNA Sequence



 National Human  
Genome Research  
Institute  
[genome.gov/sequencingcosts](http://genome.gov/sequencingcosts)

## Cost per Genome Sequenced

The cost of sequencing a human genome compared with the reductions that would be expected at the rate Moore's law predicts for computer chips. Over the past decade, next-generation sequencing and cloud computing drove the figure down. The average bumped higher in recent years because of brief slowdowns in production.



Source: NIH

NEO LIFE



# Who will have **bigger data** by 2025?

<u>Data Phase</u>	<u>Astronomy</u>	<u>Twitter</u>
Acquisition	25 zetta-bytes/year	0.5–15 billion tweets/year
Storage	1 EB/year	1–17 PB/year

Peta= $10^{15}$

Exa= $10^{18}$

Zetta= $10^{21}$

## YouTube

500–900 million hours/year

1–2 EB/year

## Genomics

1 zetta-bases/year

2–40 EB/year

Z. Stephens, S. Lee, F. Faghri, R. Campbell, C. Zhai, M. Efron,  
R. Iyer, M. Schatz, S. Sinha, and G. Robinson (2015) PLoS Biol 13: e1002195.

Base pairs	Unit	Abbreviation	Example
1	1 base pair	1 bp	A, C, G, T = 2 bits = 0.25 bytes
1000	1 kilobase pair	1 kb	
1,000,000	1 megabase pair	1 Mb	
10 <sup>9</sup>	1 gigabase pair	1 Gb	
10 <sup>12</sup>	1 terabase pair	1 Tb	
10 <sup>15</sup>	1 petabase pair	1 Pb	

Size	Abbreviation	No. bytes	Examples
Bytes	–	1	1 byte is typically 8 bits, used to encode a single character of text
Kilobytes	1 kb	10 <sup>3</sup>	Size of a text file with up to 1000 characters
Megabytes	1 MB	10 <sup>6</sup>	Size of a text file with 1 million characters <i>← Bacterial genome = "War &amp; Peace"</i>
Gigabytes	1 GB	10 <sup>9</sup>	600 GB: size of GenBank (uncompressed flat files) ftp://ftp.ncbi.nih.gov/genbank/gbrel.txt (WebLink 2.84) <i>← Human Genome</i>
Terabytes	1 TB	10 <sup>12</sup>	385 TB: <u>United States Library of Congress web archive</u> ( <a href="http://www.loc.gov/webarchiving/faq.html">http://www.loc.gov/webarchiving/faq.html</a> ) (WebLink 2.85) 464 TB: Data generated by the <u>1000 Genomes Project</u> ( <a href="http://www.1000genomes.org/faq/how-much-disk-space-used-1000-genomes-project">http://www.1000genomes.org/faq/how-much-disk-space-used-1000-genomes-project</a> ) (WebLink 2.86)
Petabytes	1 PB	10 <sup>15</sup>	1 PB: size of dataset available from <u>The Cancer Genome Atlas (TCGA)</u> 5 PB: size of <u>SRA data available for download from NCBI</u> 15 PB: amount of data produced <u>each year at the physics facility CERN (near Geneva)</u> ( <a href="http://home.web.cern.ch/about/computing">http://home.web.cern.ch/about/computing</a> ) (WebLink 2.87)
Exabytes	1 EB	10 <sup>18</sup>	<u>2.5 exabytes of data are produced worldwide (Lampitt, 2014)</u>

# What makes genomic data so big?

- There are **~9 millions species** each with its own genome
- **Each of us humans** (7.5 billions and counting) has **unique DNA**: we want to compare them all to each other
- Each cell has **just 1 genome (DNA)** but **multitude of transcriptomes (RNA levels)** and **proteomes (protein levels)**
- **Cancer cells acquire mutations** in their genomes: need to track **multiple lineages in a tumor vs time** to understand cancer
- **DNA** was proposed as a **long-term storage medium** of information

Farfetched? Storage standards evolve fast but DNA standard remained unchanged for 4 billion years

Note: Nature article started the comparison with a hard drive and flash memory skipping the floppy disk





## How DNA could store all the world's data





Modern archiving technology may hold an answer to that problem

Andy Extance

31 August 2016

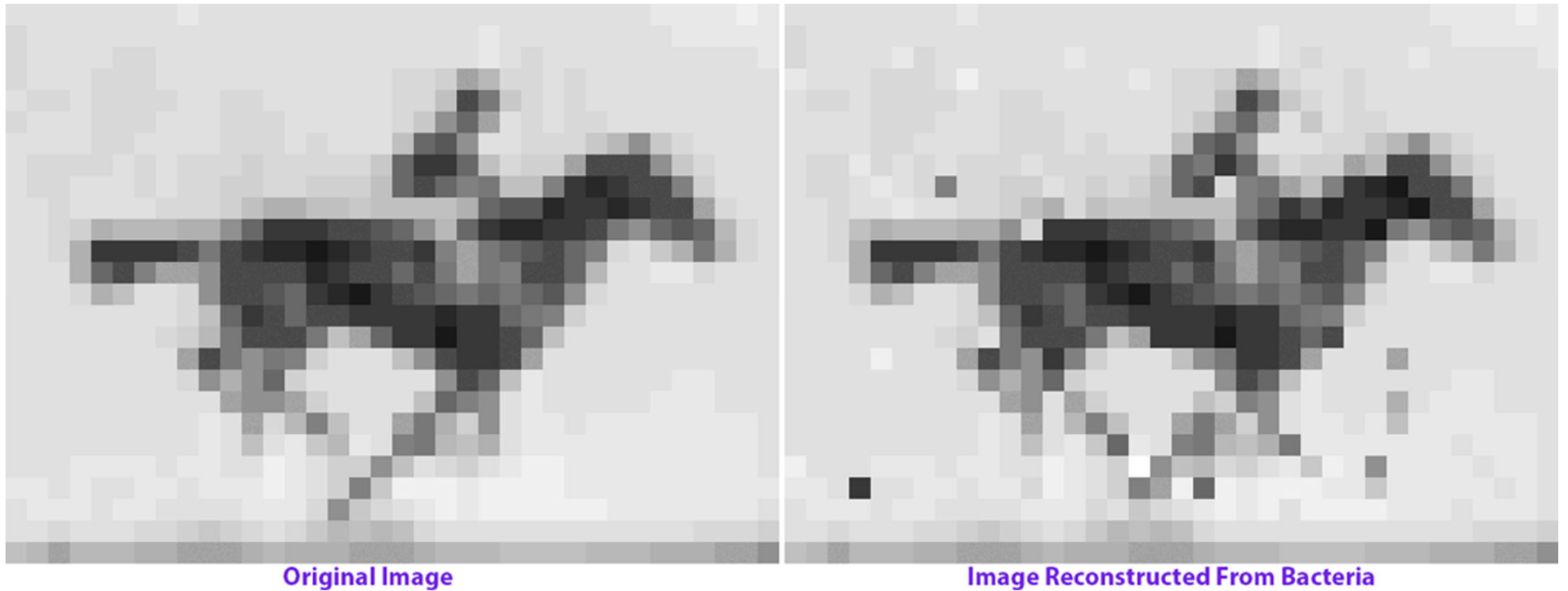
### STORAGE LIMITS

Estimates based on bacterial genetics suggest that digital DNA could one day rival or exceed today's storage technology.

	 Hard disk	 Flash memory	 Bacterial DNA	WEIGHT OF DNA NEEDED TO STORE WORLD'S DATA 
Read-write speed (µs per bit)	~3,000–5,000	~100	<100	
Data retention (years)	>10	>10	>100	
Power usage (watts per gigabyte)	~0.04	~0.01–0.04	<10 <sup>-10</sup>	
Data density (bits per cm <sup>3</sup> )	~10 <sup>13</sup>	~10 <sup>16</sup>	~10 <sup>19</sup>	

- Prof Olgica Milenkovic from Electrical and Computer Engineering UIUC is a local expert on this topic
- Profs. George Church and Sri Kosuri (Harvard Medical School) explains a potential use of DNA as storage medium in 2012
- <https://www.youtube.com/watch?v=IJAdqAVjQqY>

# Fast-forward from 2012 to 2017



Shipman SL, Nivala J, Macklis JD, Church GM.  
CRISPR–Cas encoding of a digital movie into the genomes  
of a population of living bacteria. *Nature*. 2017;547: 345–349. doi:10.1038/nature23017

Why do you need  
probability and statistics  
to analyze  
modern biological data?

Reason 2:  
Life is random and messy



# Show video “Cell organelles”

- Made at the Walter and Eliza Hall Institute of Medical Research at Victoria, Australia
- Animated by award-winning artist Dr. Drew Berry
- Go to <https://www.wehi.edu.au/wehi-tv> for other videos

Life is messy, random, and noisy

Yet it is beautifully complex  
and has many parts  
(see statistics)

# Why life is so random?

- Biomolecules are very small (nano- to micro-meters) → Brownian noise
- # molecules/cell is often small → Large cell-to-cell variations
- Genomic data comes from biological evolution
  - the Mother of all random processes
- Genomic data involves (random) samples
  - We have genomes of some (not all) organisms
  - We have tissue samples of some (not all) cancer patients

Why life is so complex?

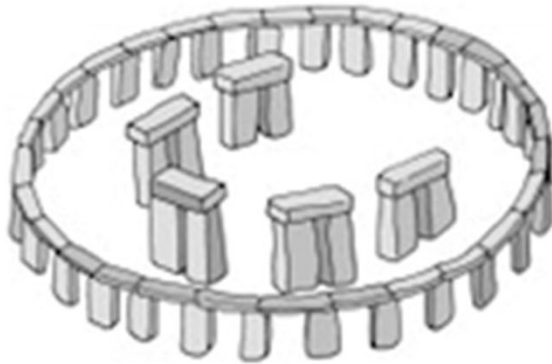
Primer on complex system

# Complex systems have many interacting parts

- All **parts** are **different** from each other
  - 10s thousands ( $10^4$ ) types of **proteins** in an organism
  - 100 thousands ( $10^5$ ) **organizations (AS)** in the Internet
  - 1 billion ( $10^9$ ) people on **Facebook**
  - 10 billion ( $10^{10}$ ) **web pages** in the WWW
  - 100 billion ( $10^{11}$ ) **neurons** in a human brain
  - **NOT  $10^{23}$  electrons or quarks studied by physics: they are all the same and boring!**
- Yet they **share** the same **basic design**
  - All proteins are strings of the **same 20 amino acids**
  - All WWW pages use **HTML**, JavaScript, etc.
  - All neurons generate and receive **electric spikes**

# Example: a complex system with many parts

# HËNJ



80x



30x



30x



10x



5x



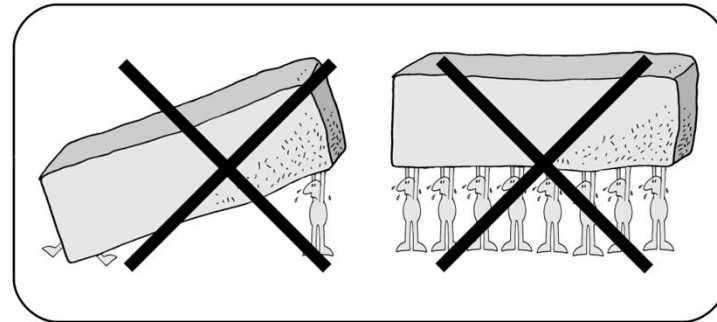
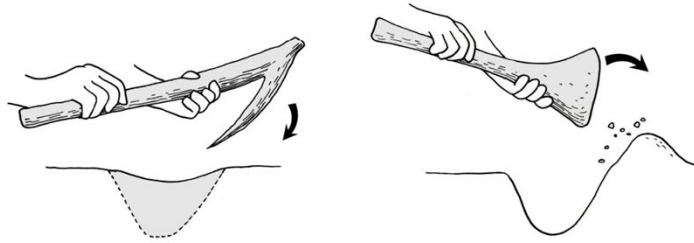
1x



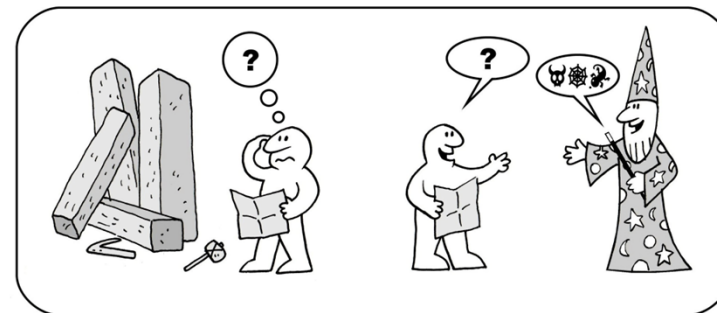
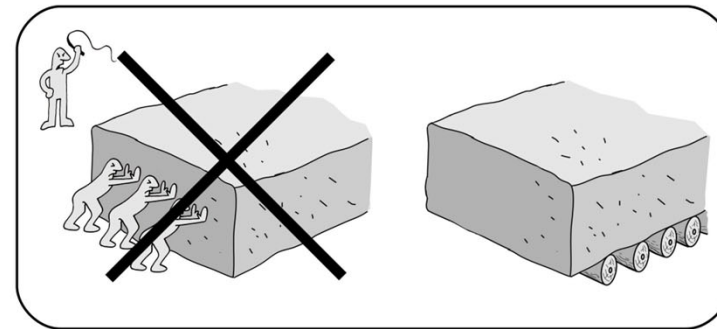
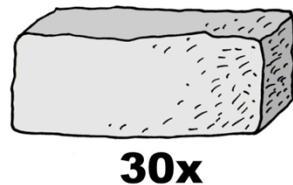
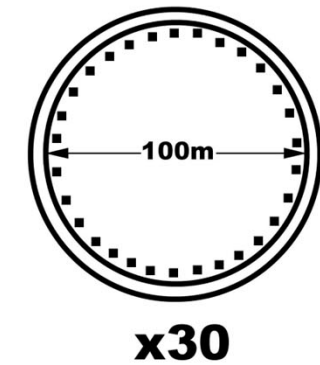
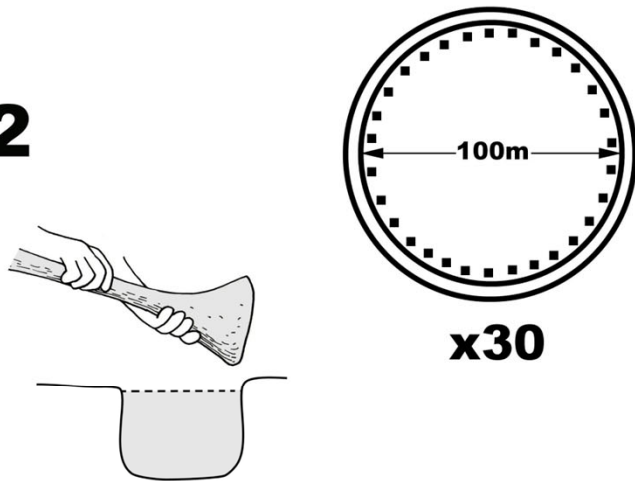
3x

# Parts interact → they need to be assembled to work

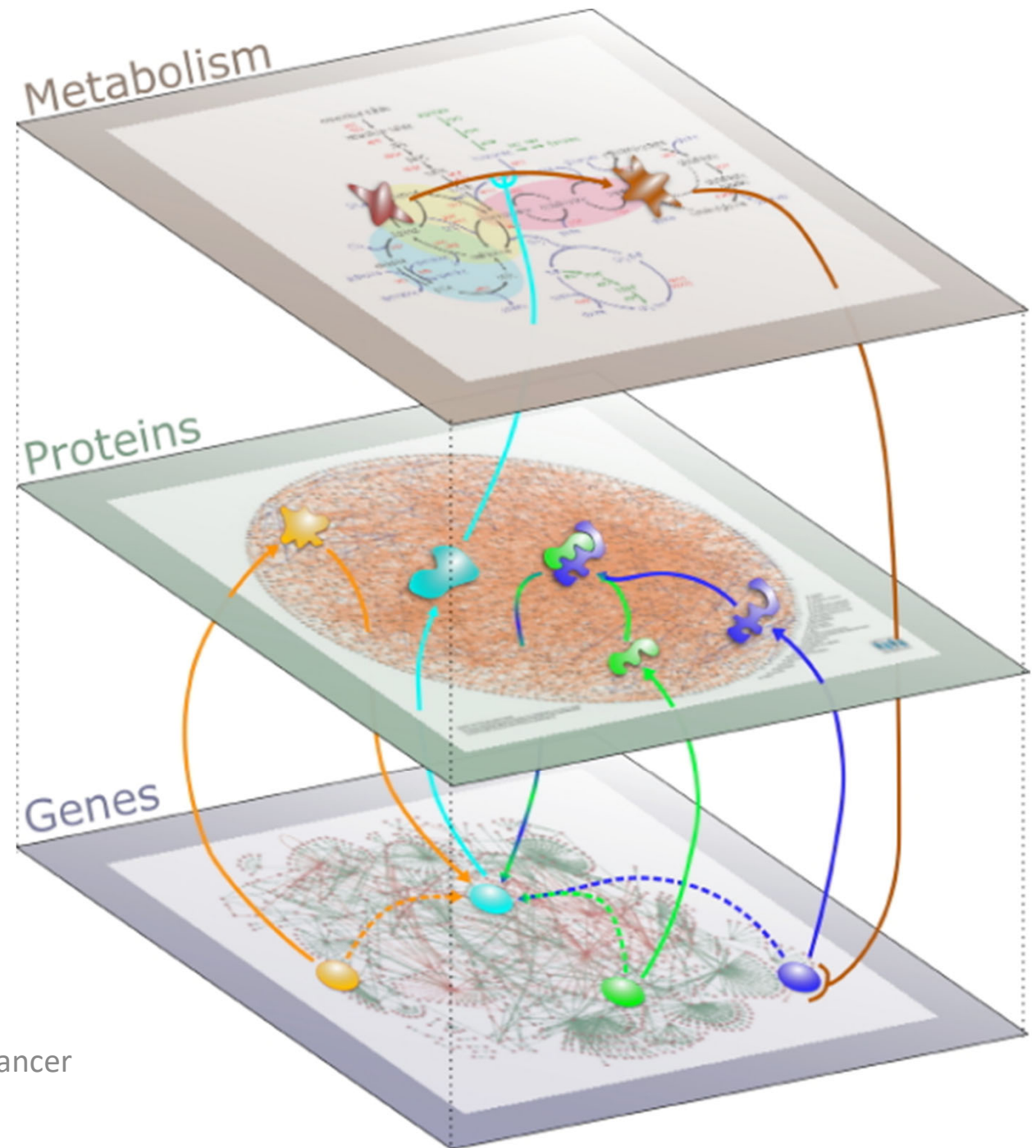
1



2



# Intra-cellular Networks operate on multiple levels



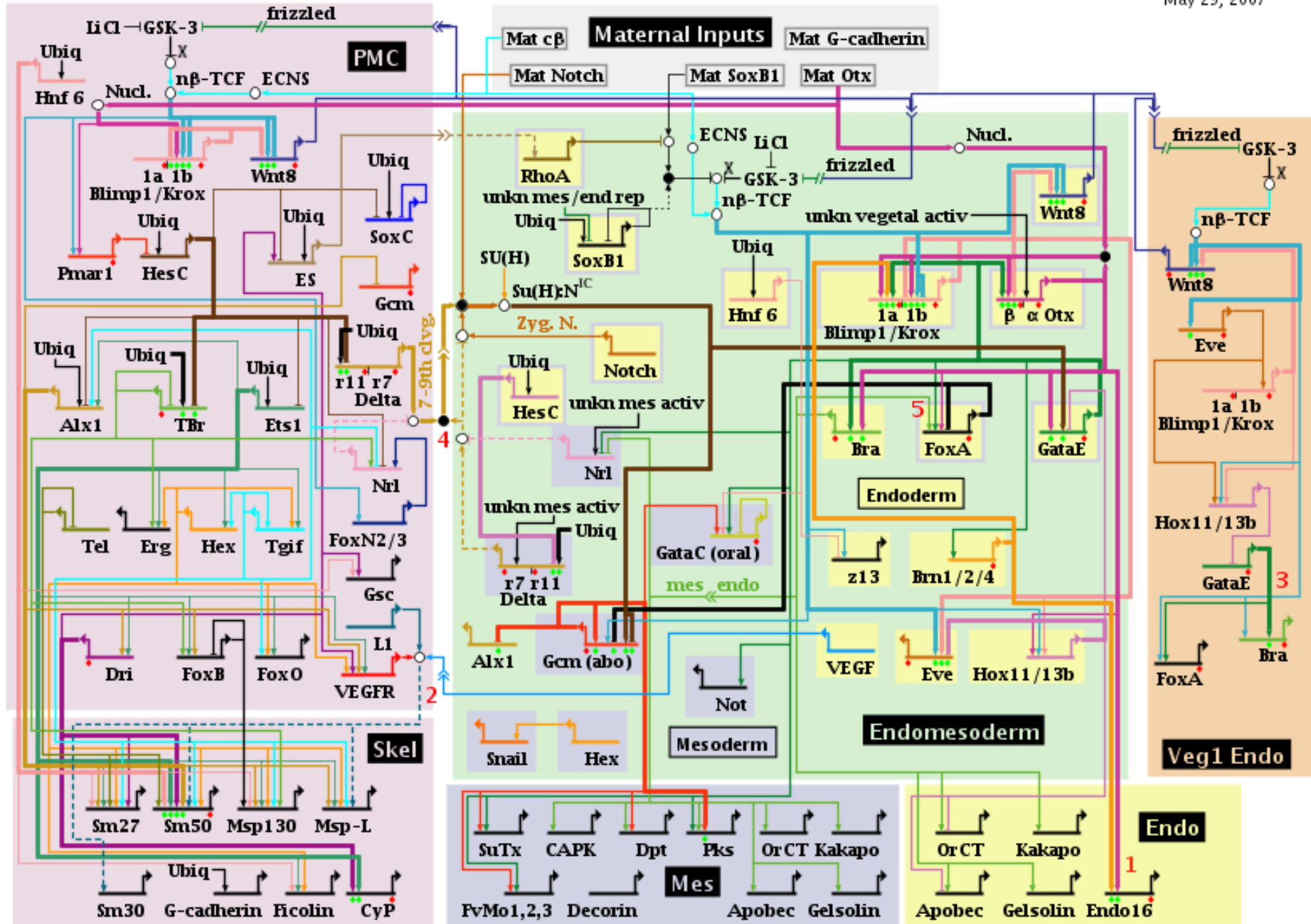
Slides by Amitabh Sharma, PhD

Northeastern University & Dana Farber Cancer  
Institute



# Sea urchin embryonic development (from endomesoderm up to 30 hours) by Davidson's lab

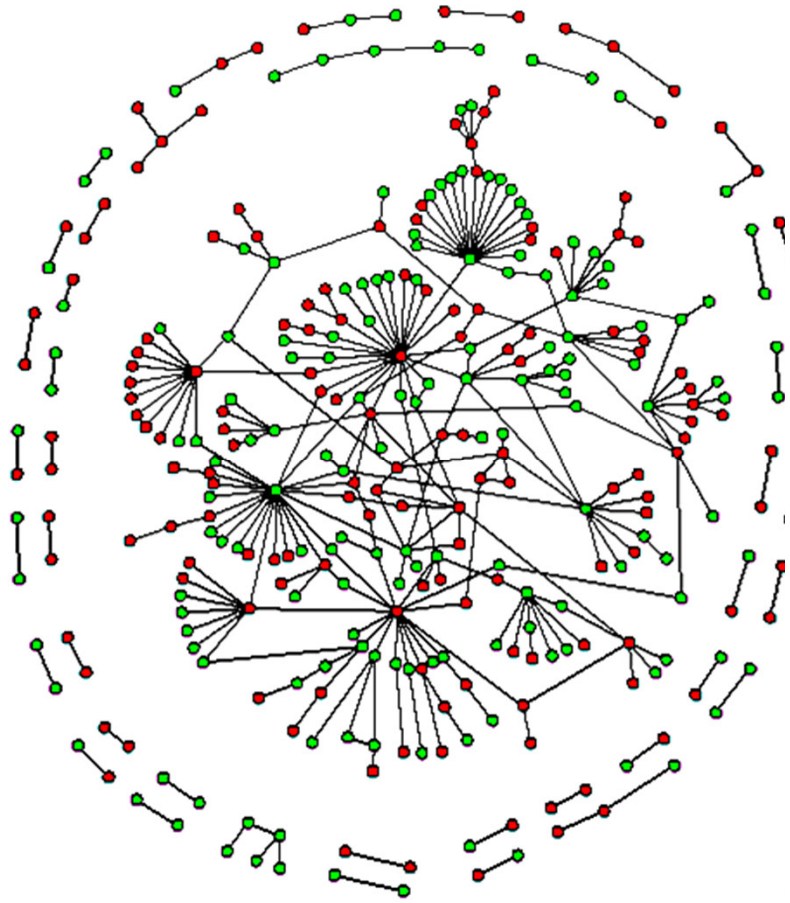
May 29, 2007



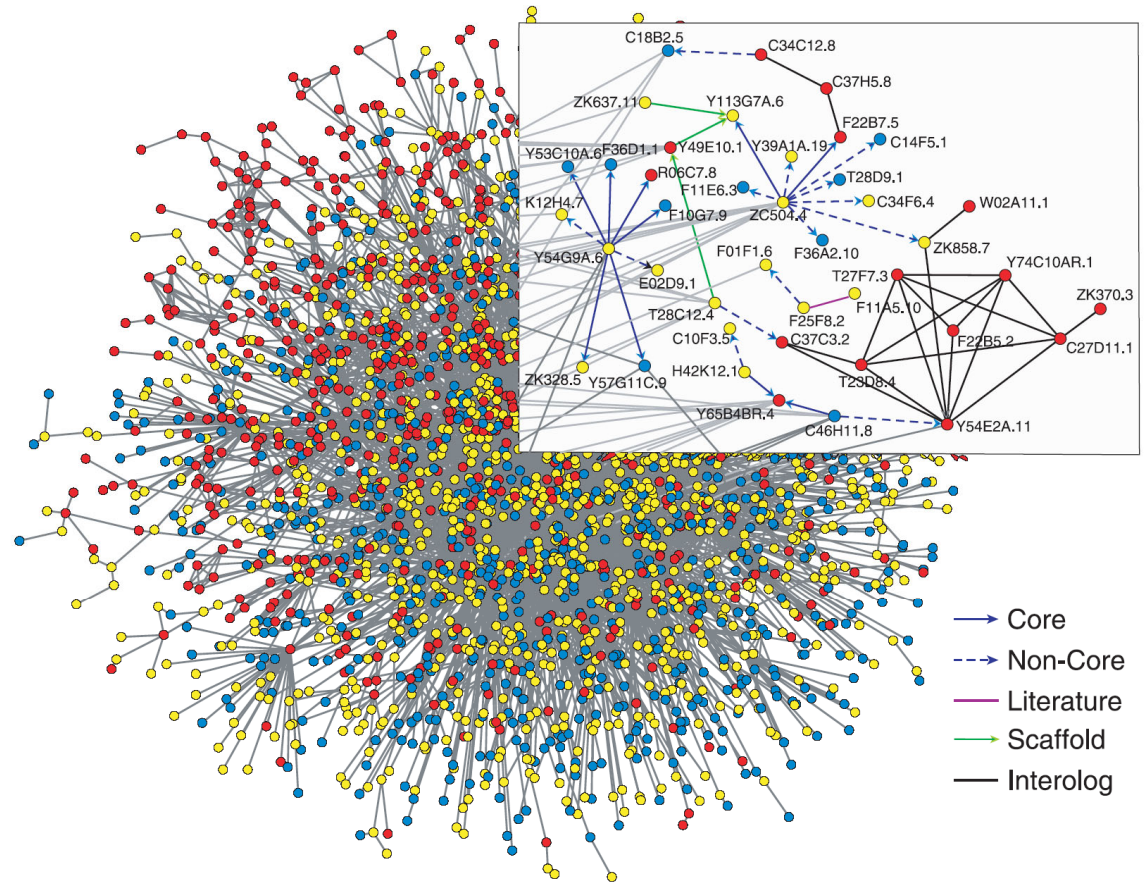
Ubiq=ubiquitous; Mat = maternal; activ = activator; rep = repressor;  
 unkn = unknown; Nucl. = nuclearization; X = β-catenin source;  
 nβ-TCF = nuclearized b-β-catenin-Tcf1; ES = early signal;  
 ECNS = early cytoplasmic nuclearization system; Zyg. N. = zygotic Notch

Copyright © 2001-2007 Hamid Bolouri and Eric Davidson

# Protein-Protein binding IntAct Database (Dec 2015) Interactions: 577,297 Proteins: 89,716



Baker's yeast *S. cerevisiae* (only nuclear proteins shown)  
From S. Maslov, K. Sneppen, Science 2002



Worm *C. elegans*  
From S. Lee et al, Science 2004



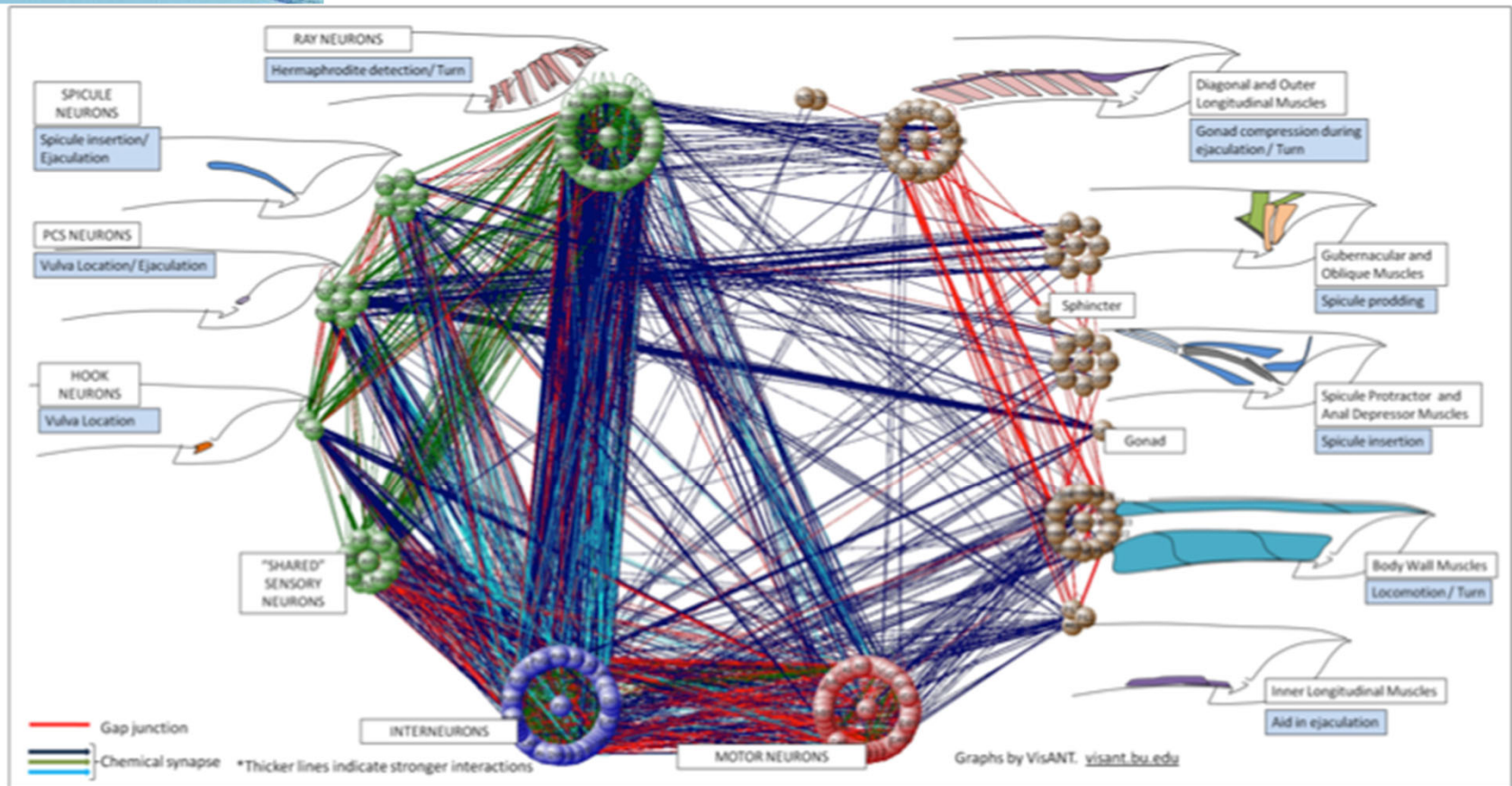


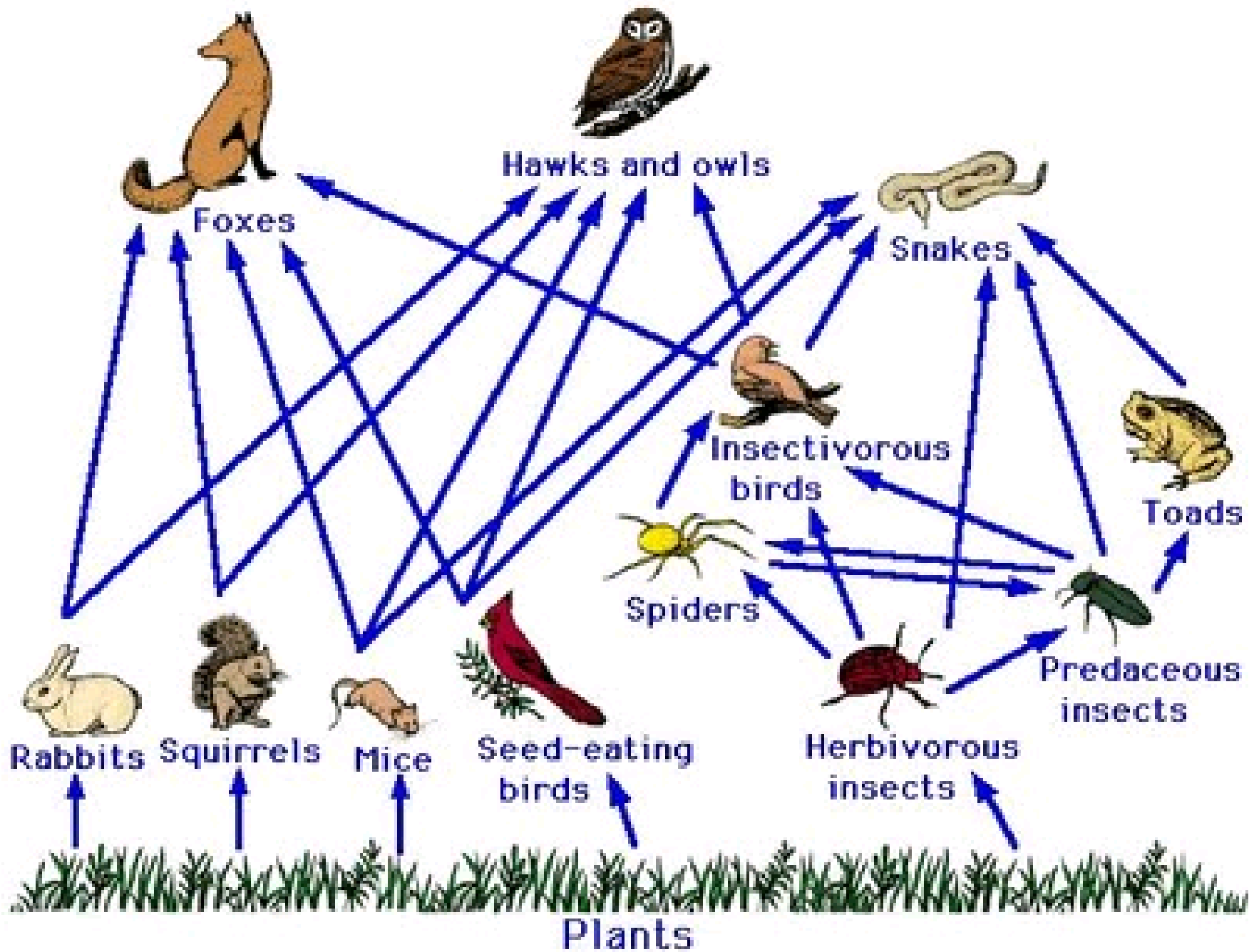


# Brain and nerves of a worm



- Worm (*C. elegans*) has 302 neurons
- Our brain has 100 billion ( $10^{11}$ ) neurons



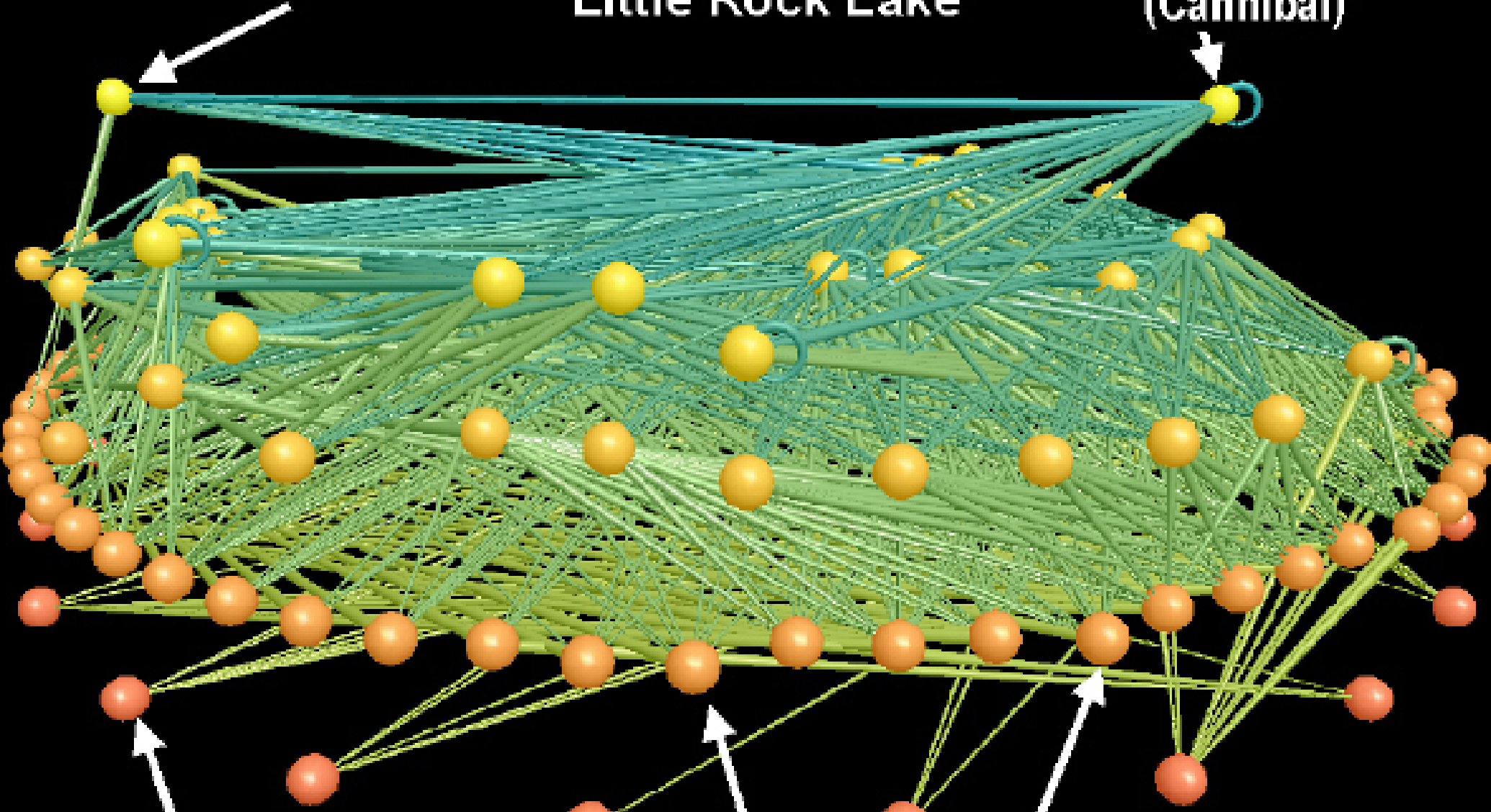


by Neo Martinez and Richard Williams

# Food Web of Little Rock Lake

## Smallmouth Bass (Cannibal)

### Leech



**1st Trophic Level**  
**Mostly Phytoplankton**

**2nd Trophic Level**  
**Many Zooplankton**



# QUESTIONS

FOUND IN GOOGLE AUTOCOMPLETE

WHY ARE THERE SLAVES IN THE BIBLE  
 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  
 WHY ARE THERE WEEKS  
 WHY DO I FEEL DIZZY  
 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 DO TESTICLES MOVE  
 WHY ARE THERE PSYCHICS  
 WHY ARE HATS SO EXPENSIVE  
 WHY IS THERE CAFFEINE IN MY SHAMPOO  
 WHY DO YOUR BOOBS HURT  
 WHY DO TWINS HAVE DIFFERENT FINGERPRINTS  
 WHY ARE AMERICANS AFRAID OF DRAGONS

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 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 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 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 ULTRASOUNDS IMPORTANT  
 WHY ARE ULTRASOUND MACHINES EXPENSIVE  
 WHY IS STEALING WRONG

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 ARE THERE MALE AND FEMALE BIKES  
 WHY ARE THERE BRIDESMAIDS  
 WHY DO DYING PEOPLE REACH UP  
 WHY AREN'T THERE VARIKOSE ARTERIES  
 WHY ARE OLD KLINGONS DIFFERENT



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

Credit: XKCD  
 comics

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 DOGS AFRAID OF FIREWORKS  
 WHY IS THERE NO KING IN ENGLAND