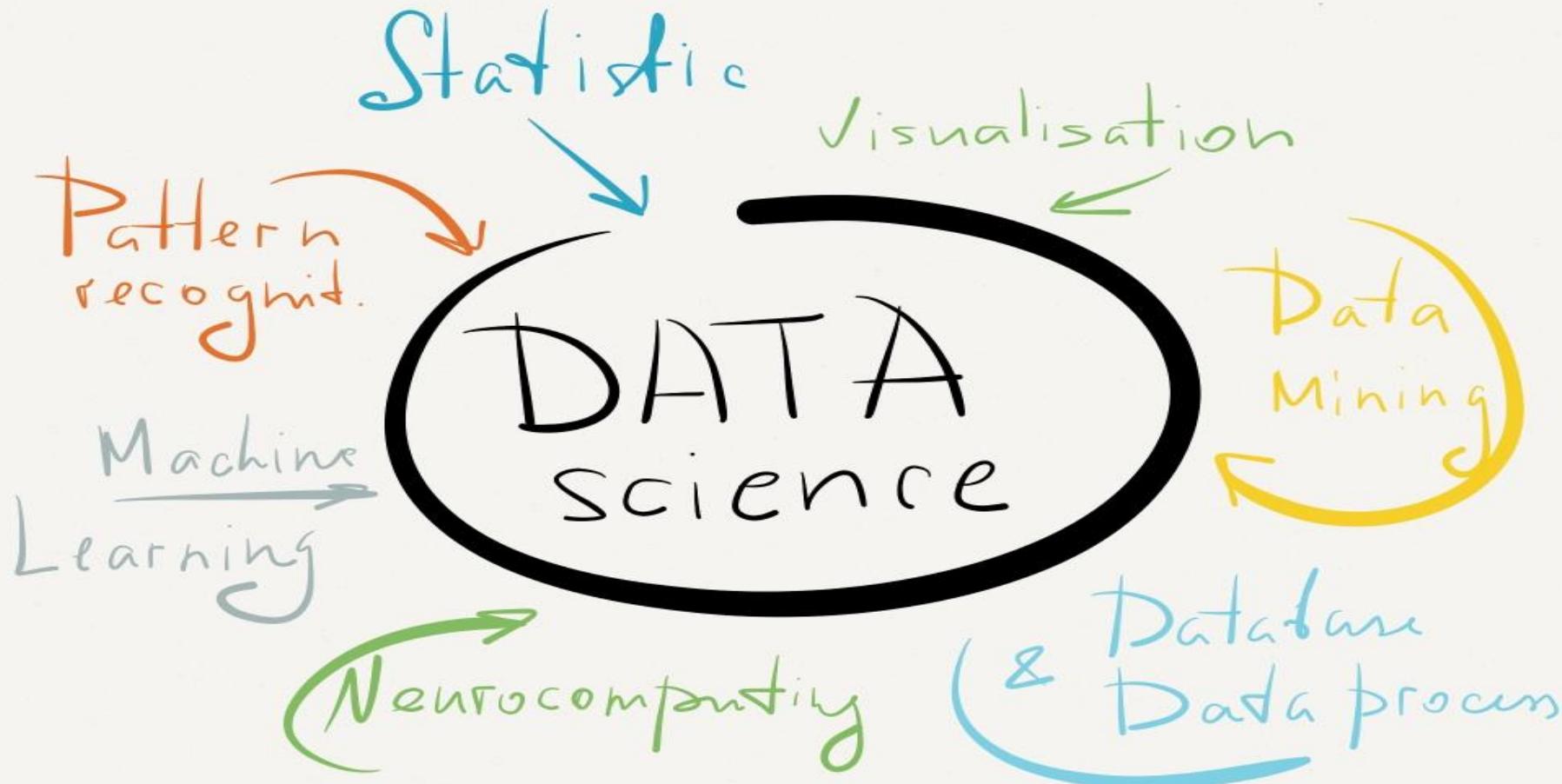
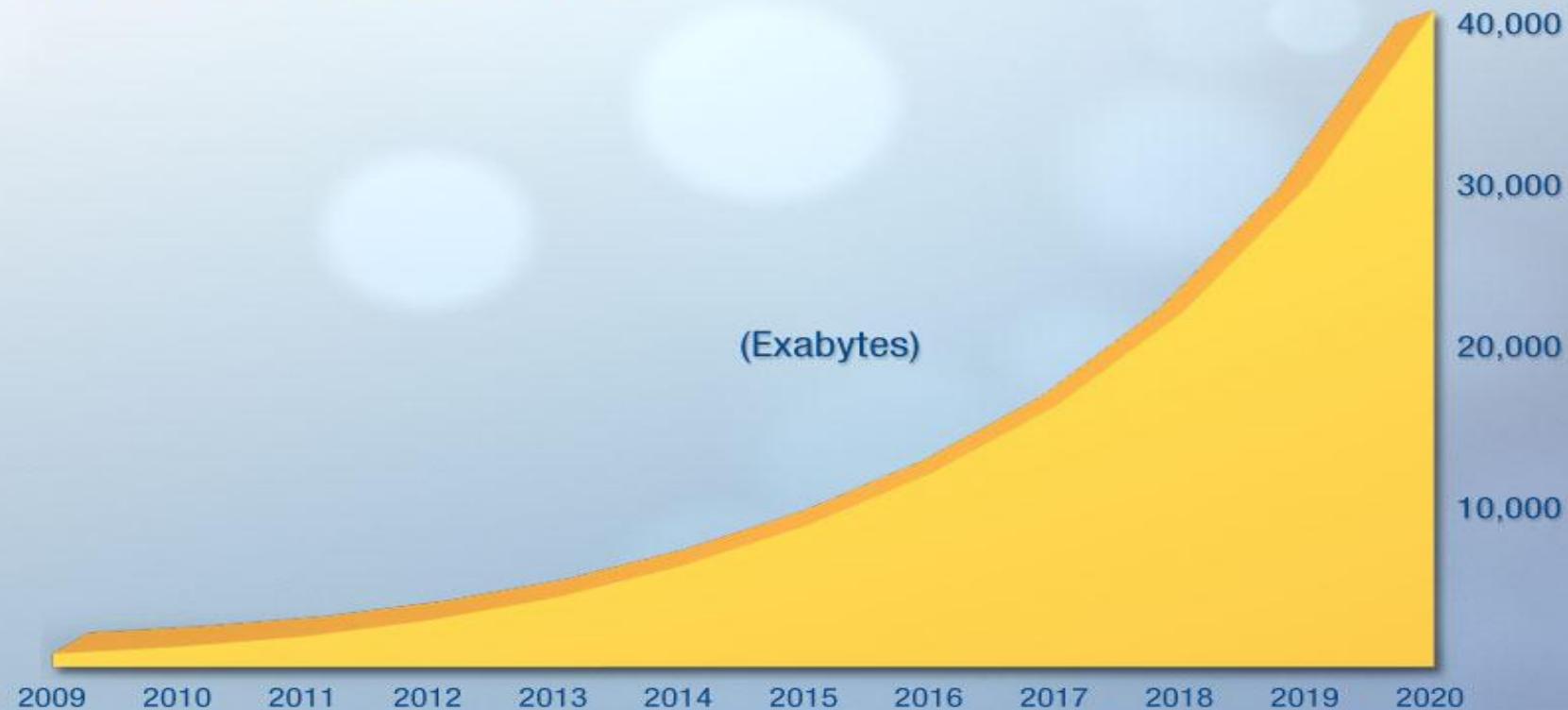


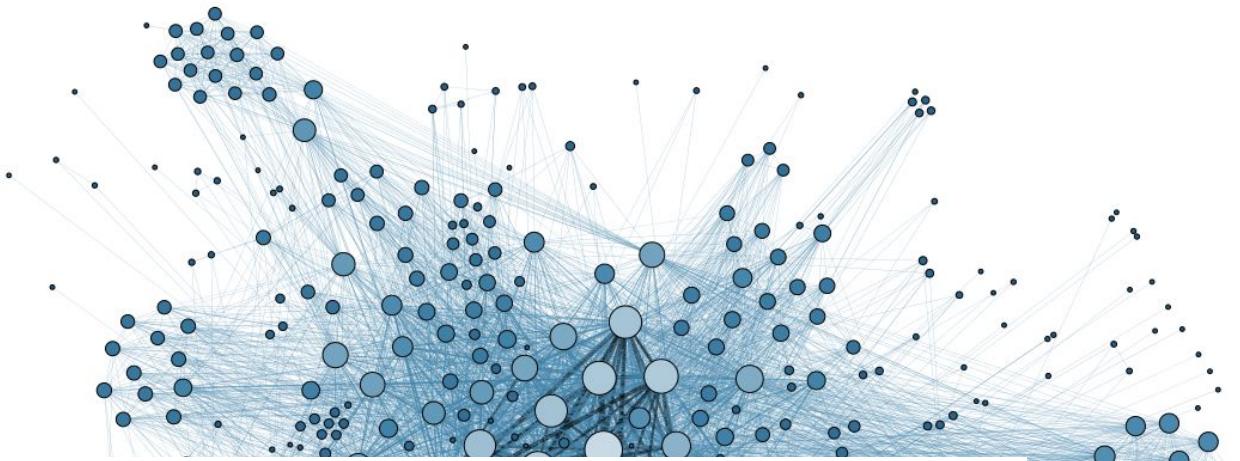


# What is Data Science?

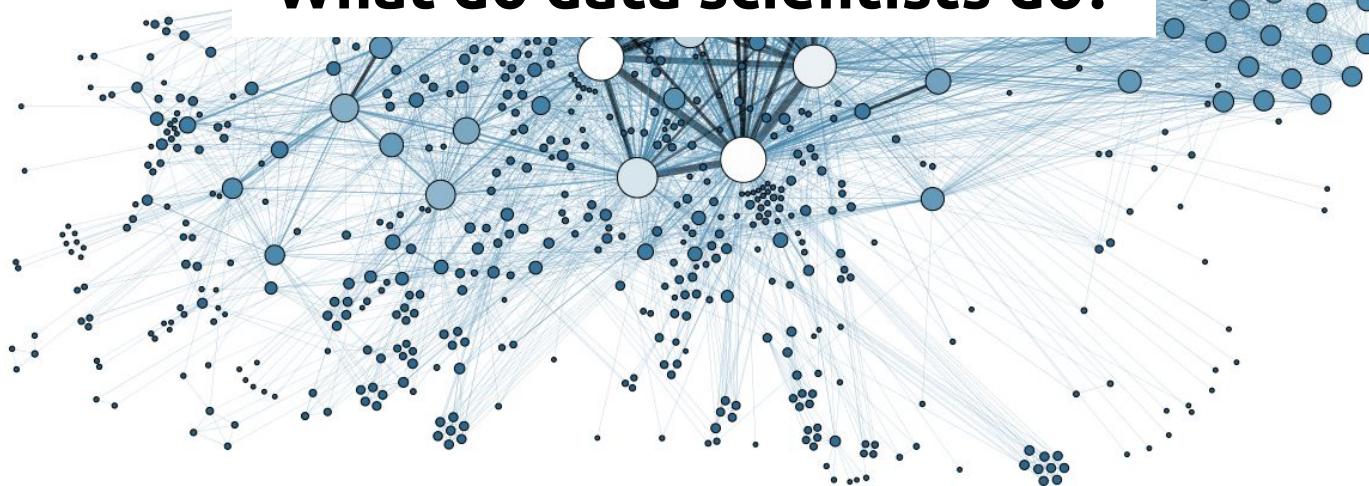


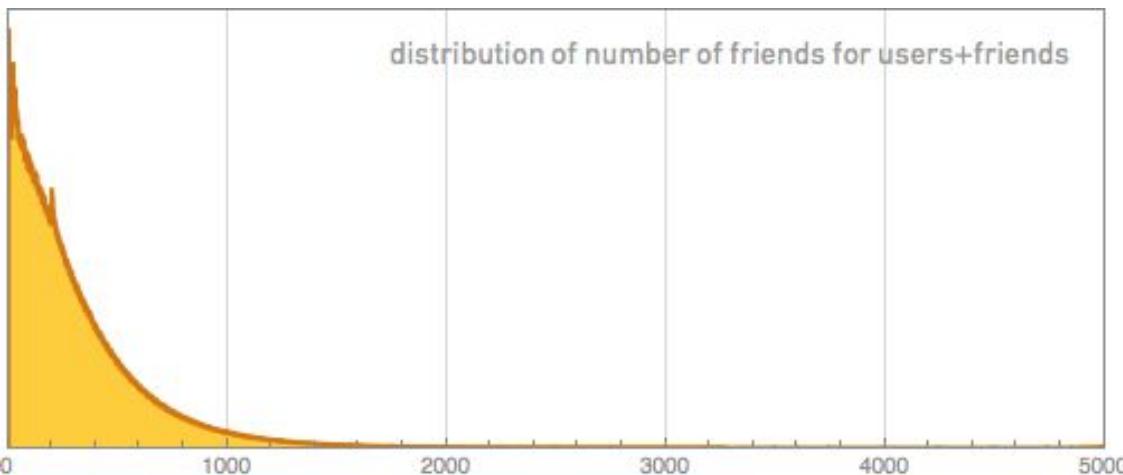
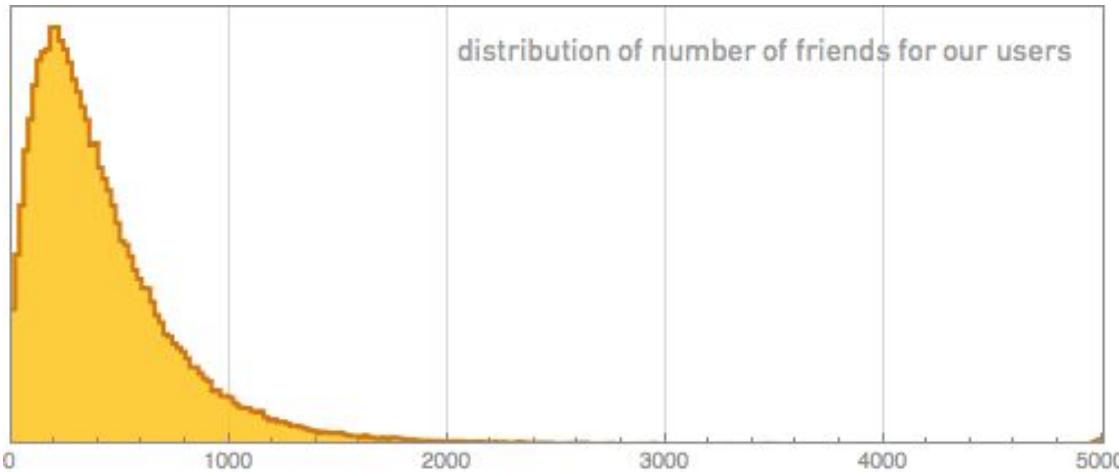
# The Digital Universe: 50-fold Growth from the Beginning of 2010 to the End of 2020

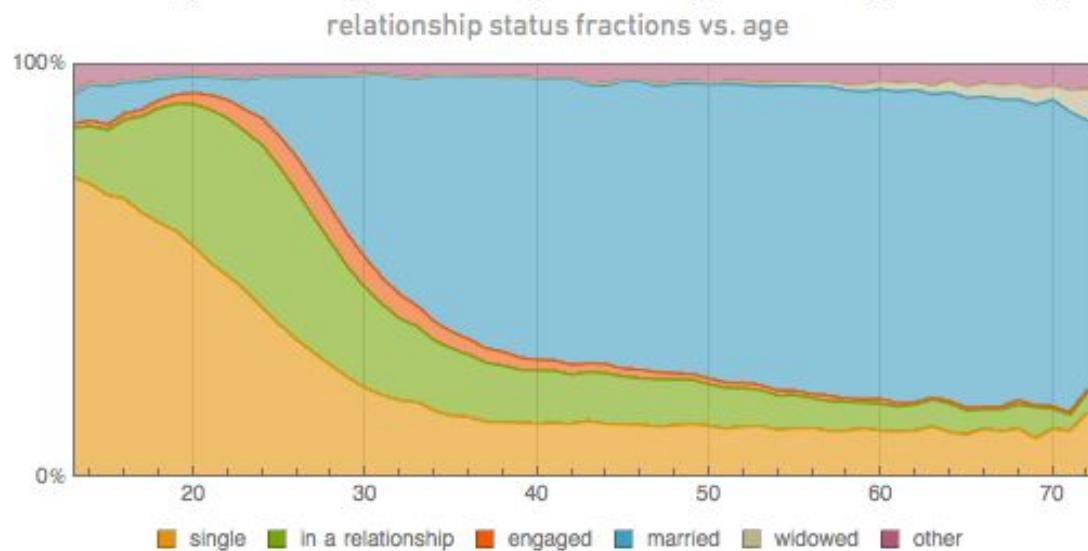
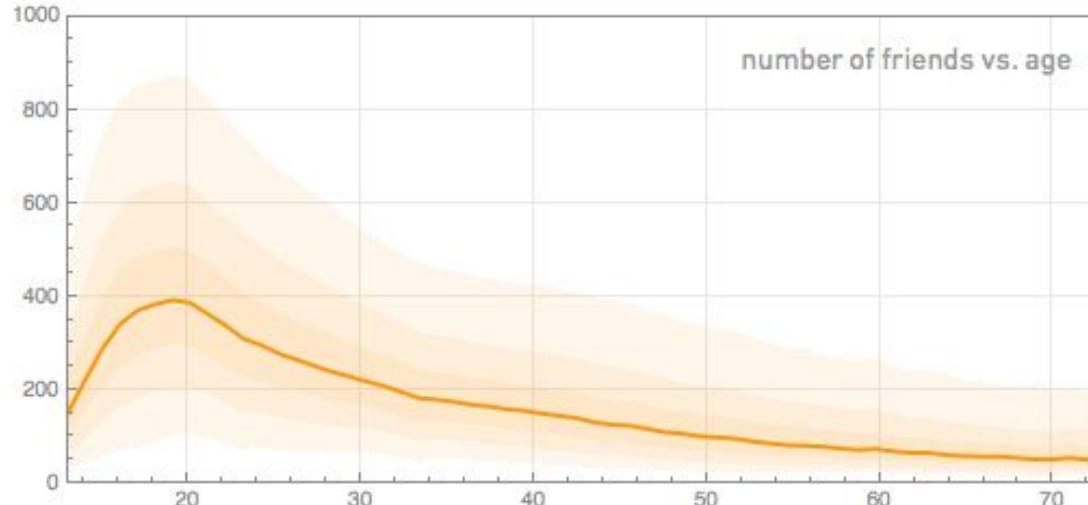


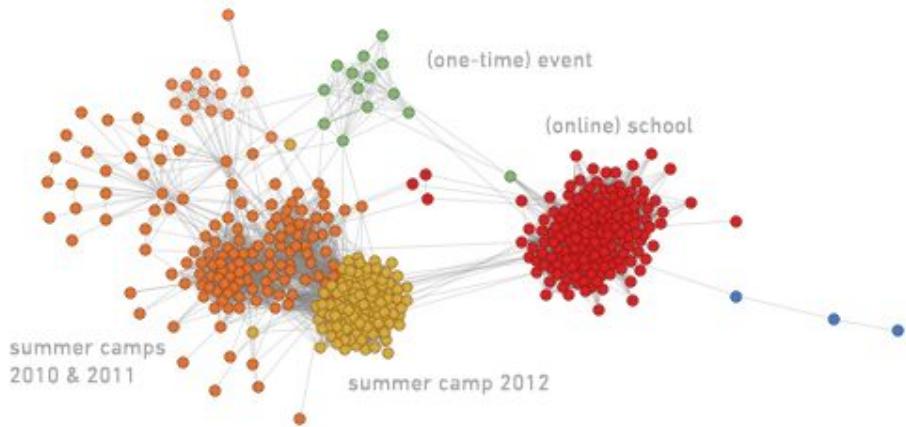


## What do data scientists do?







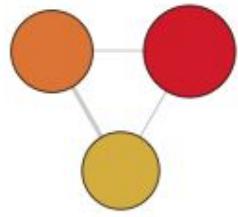


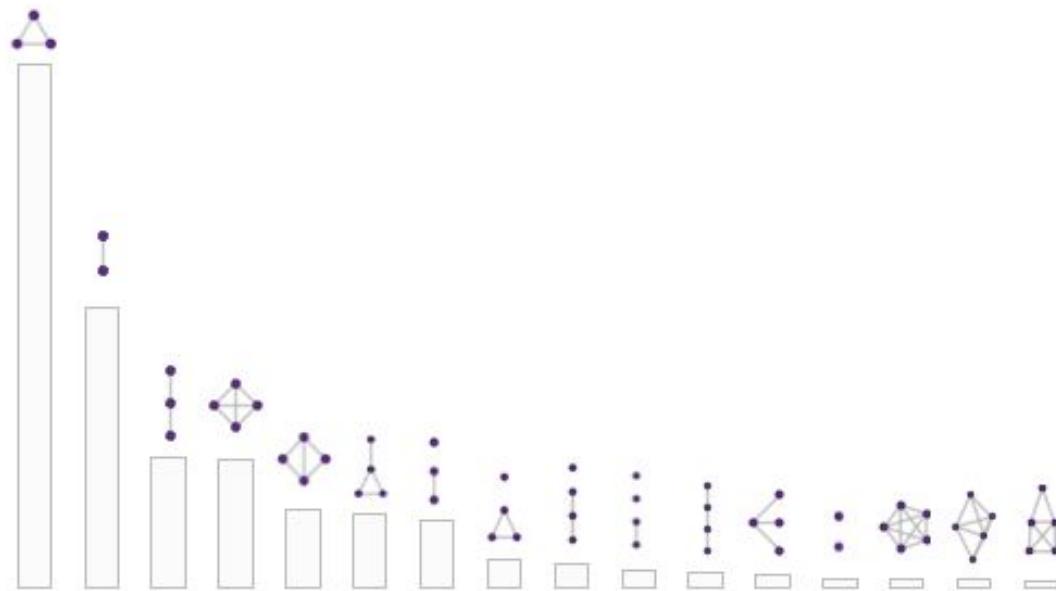
sports team

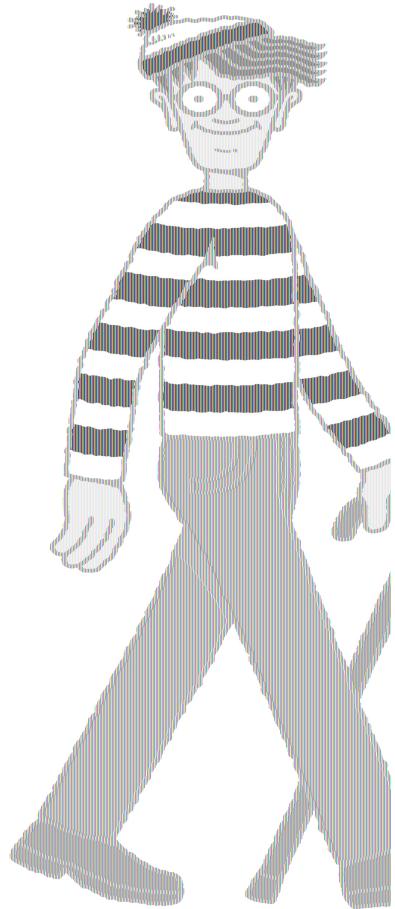


"random" people







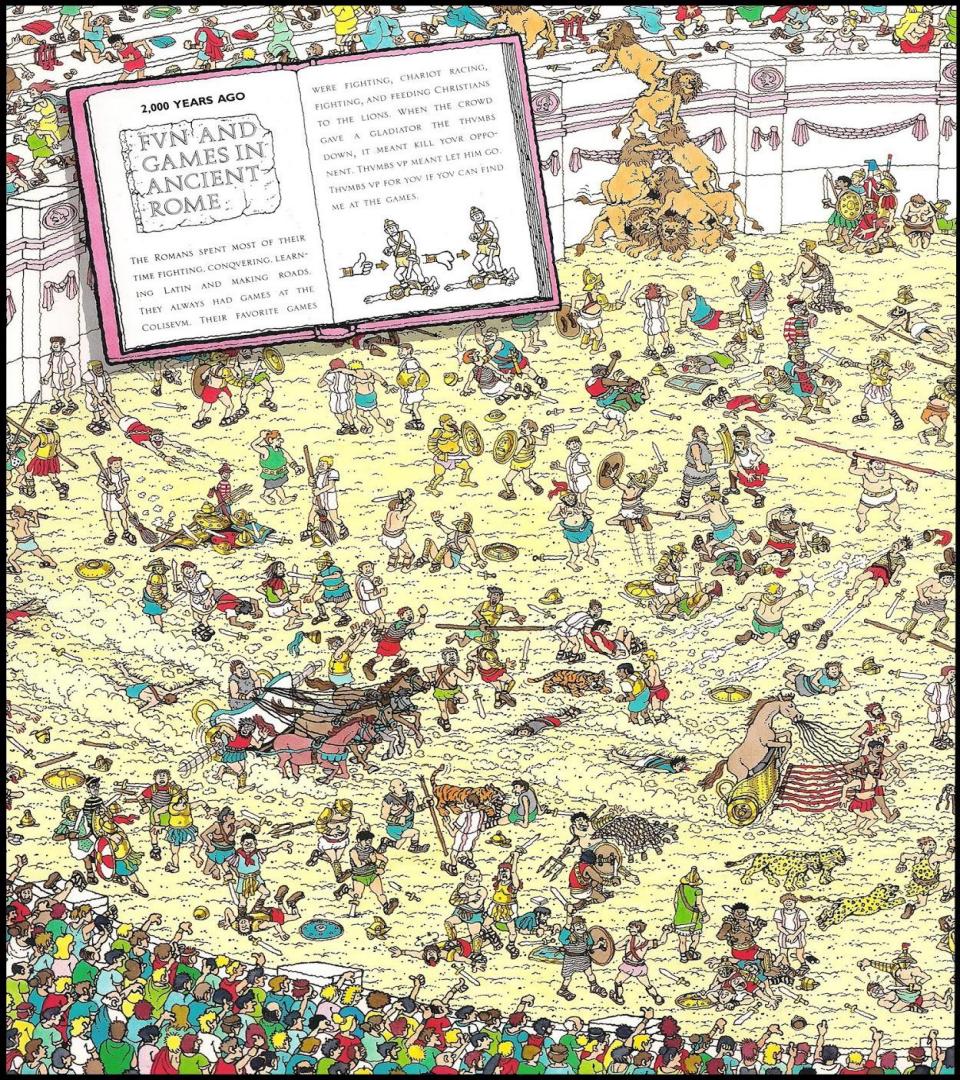


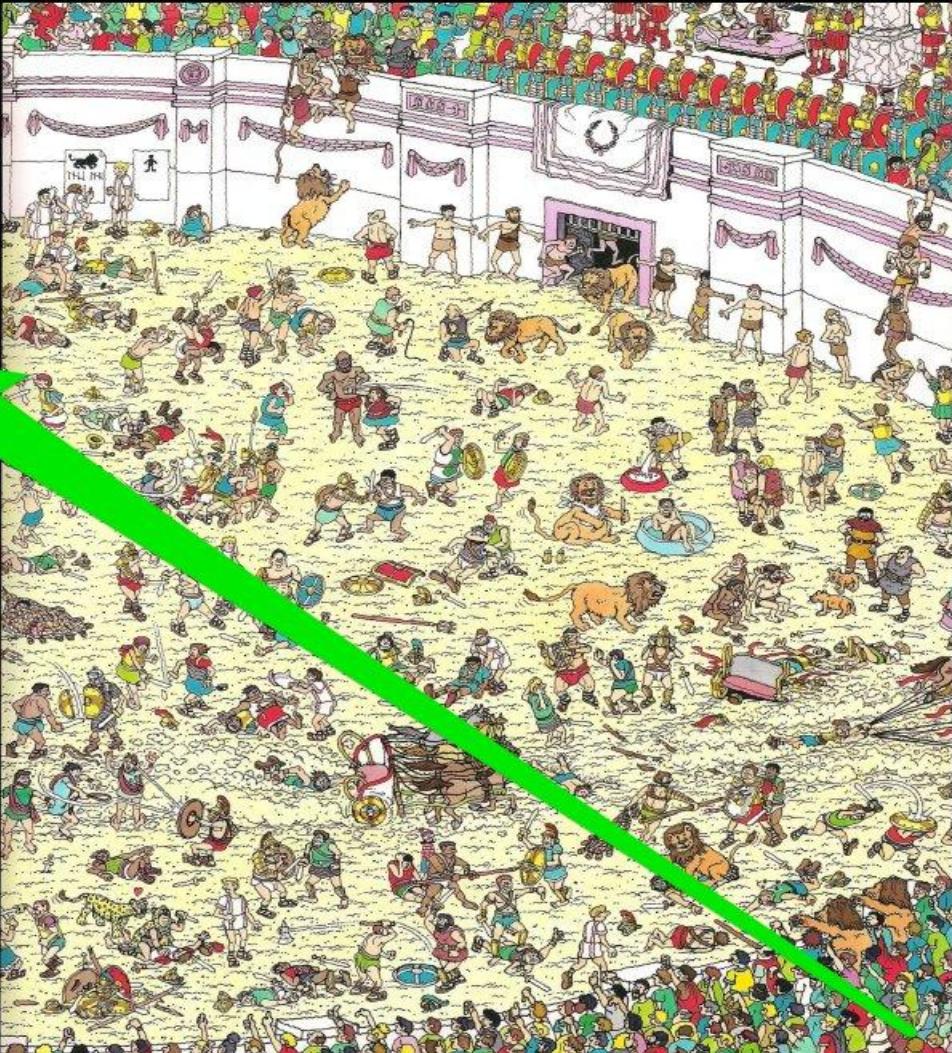
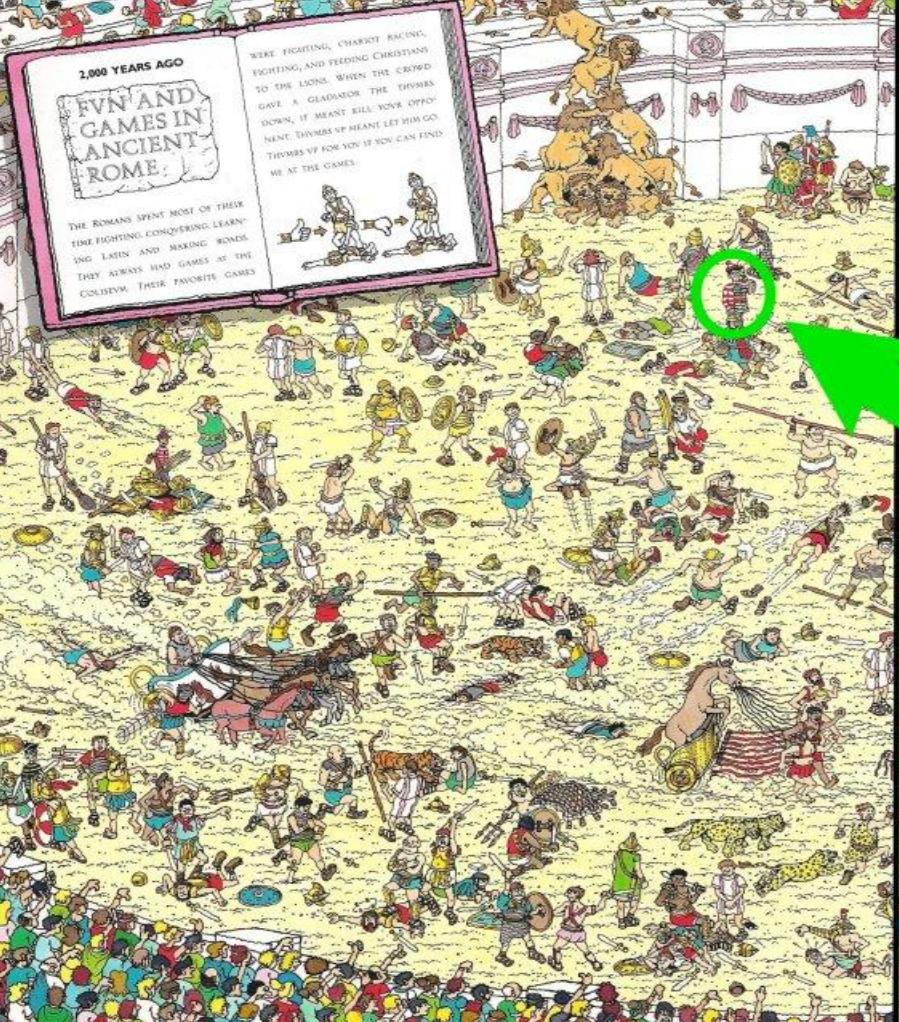
2,000 YEARS AGO

## FUN AND GAMES IN ANCIENT ROME

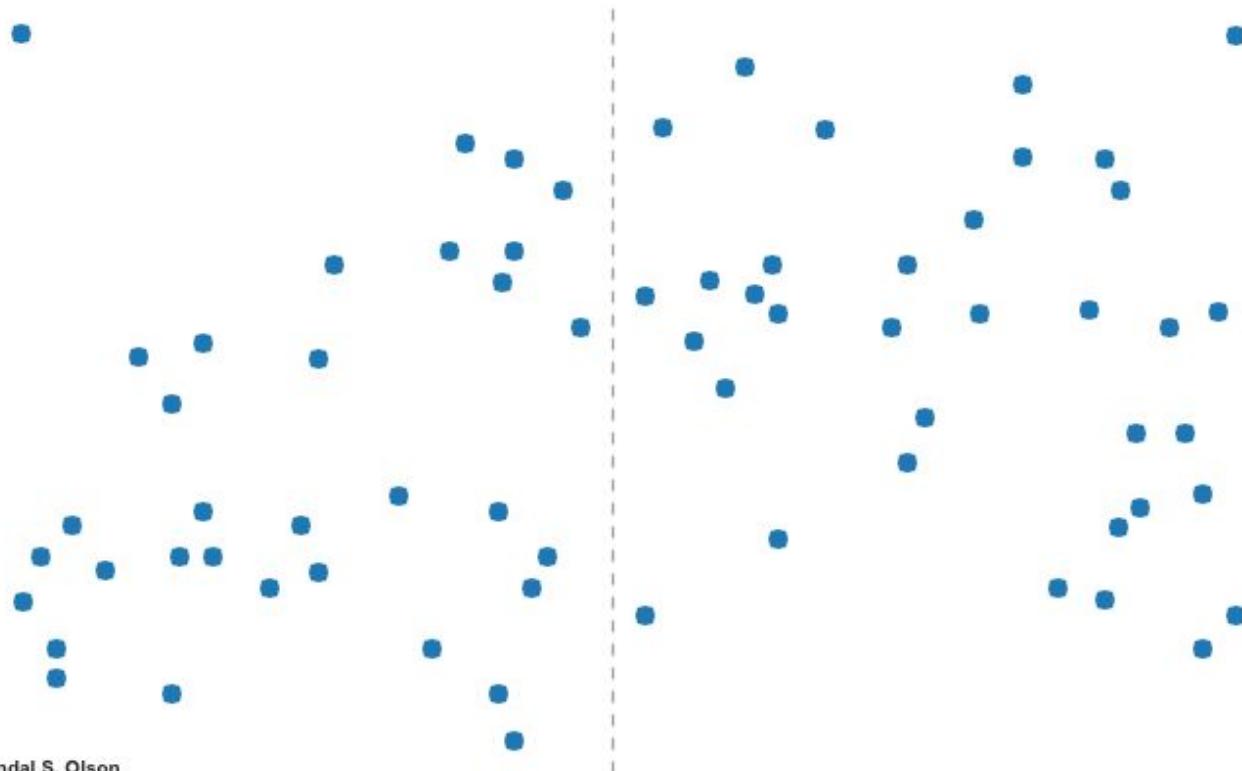
THE ROMANS SPENT MOST OF THEIR TIME FIGHTING, CONQUERING, LEARNING LATIN AND MAKING ROADS. THEY ALWAYS HAD GAMES AT THE COLOSSEUM. THEIR FAVORITE GAMES WERE FIGHTING, CHARIOT RACING, FIGHTING, AND FEEDING CHRISTIANS TO THE LIONS. WHEN THE CROWD GAVE A GLADIATOR THE THUMBS DOWN, IT MEANT KILL YOVR OPPONENT. THUMBS UP MEANT LET HIM GO.

ME AT THE GAMES.

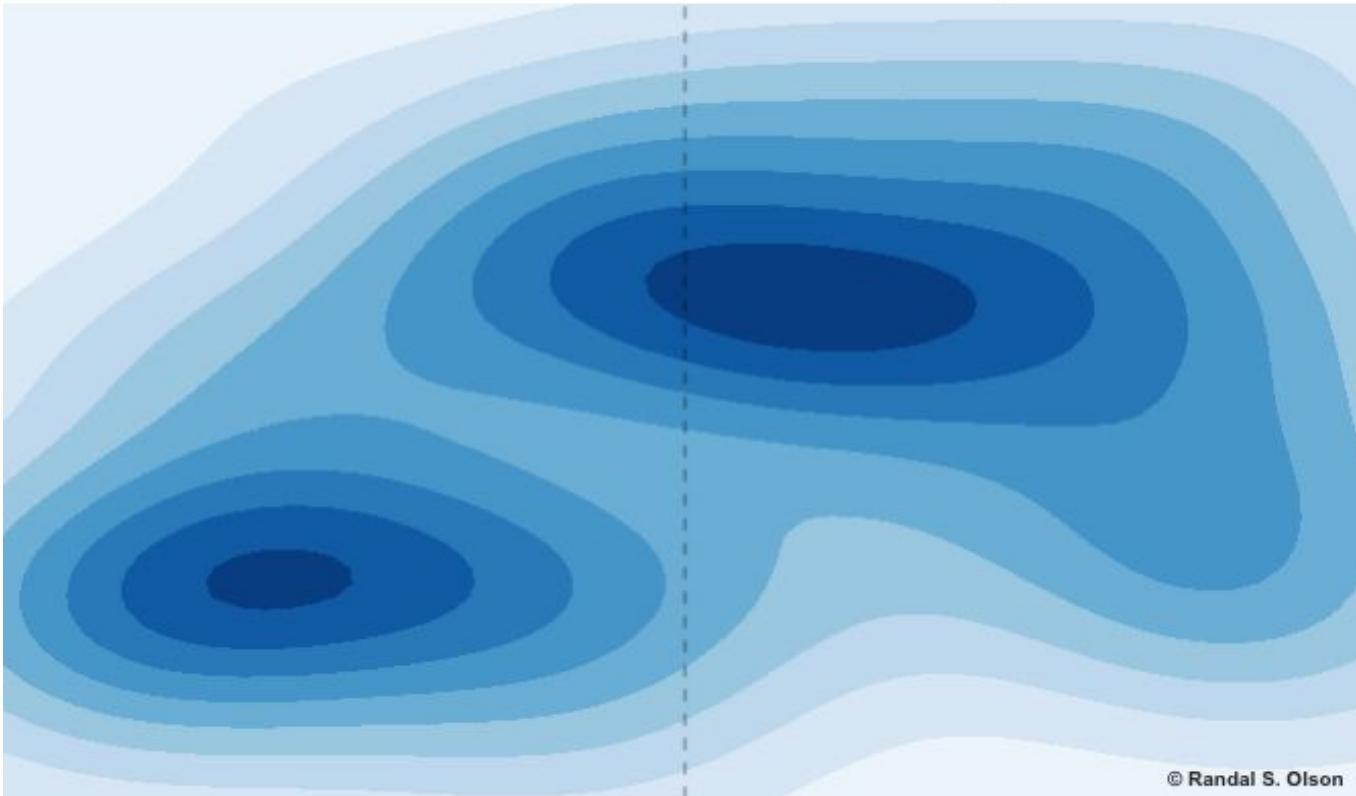




# Waldo locations in "Where's Waldo?"



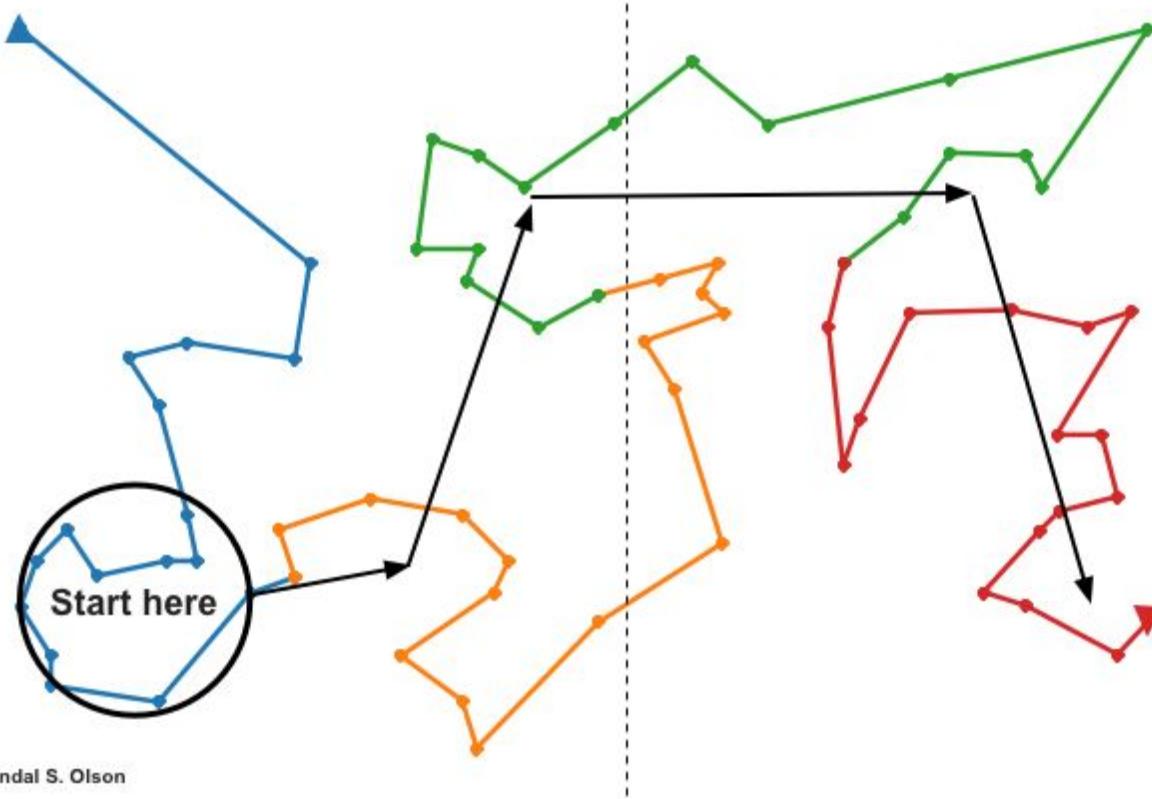
# "Where's Waldo?" kernel density estimate

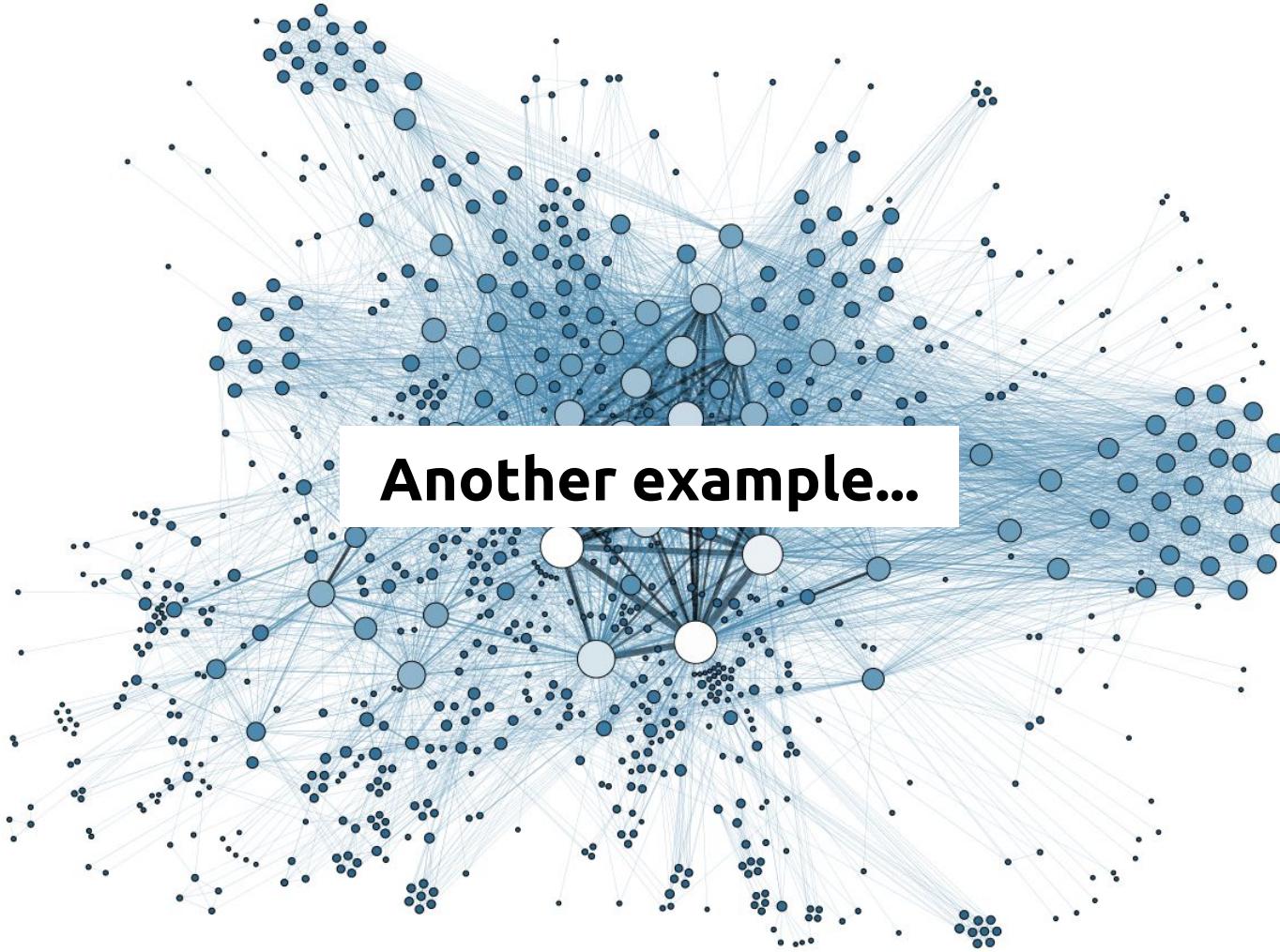


© Randal S. Olson

<http://gfycat.com/DeadlyJampackedFishingcat>

# "Where's Waldo?" optimal search path





Another example...

<http://www.residentmar.io/2016/02/09/average-chain-distance.html>

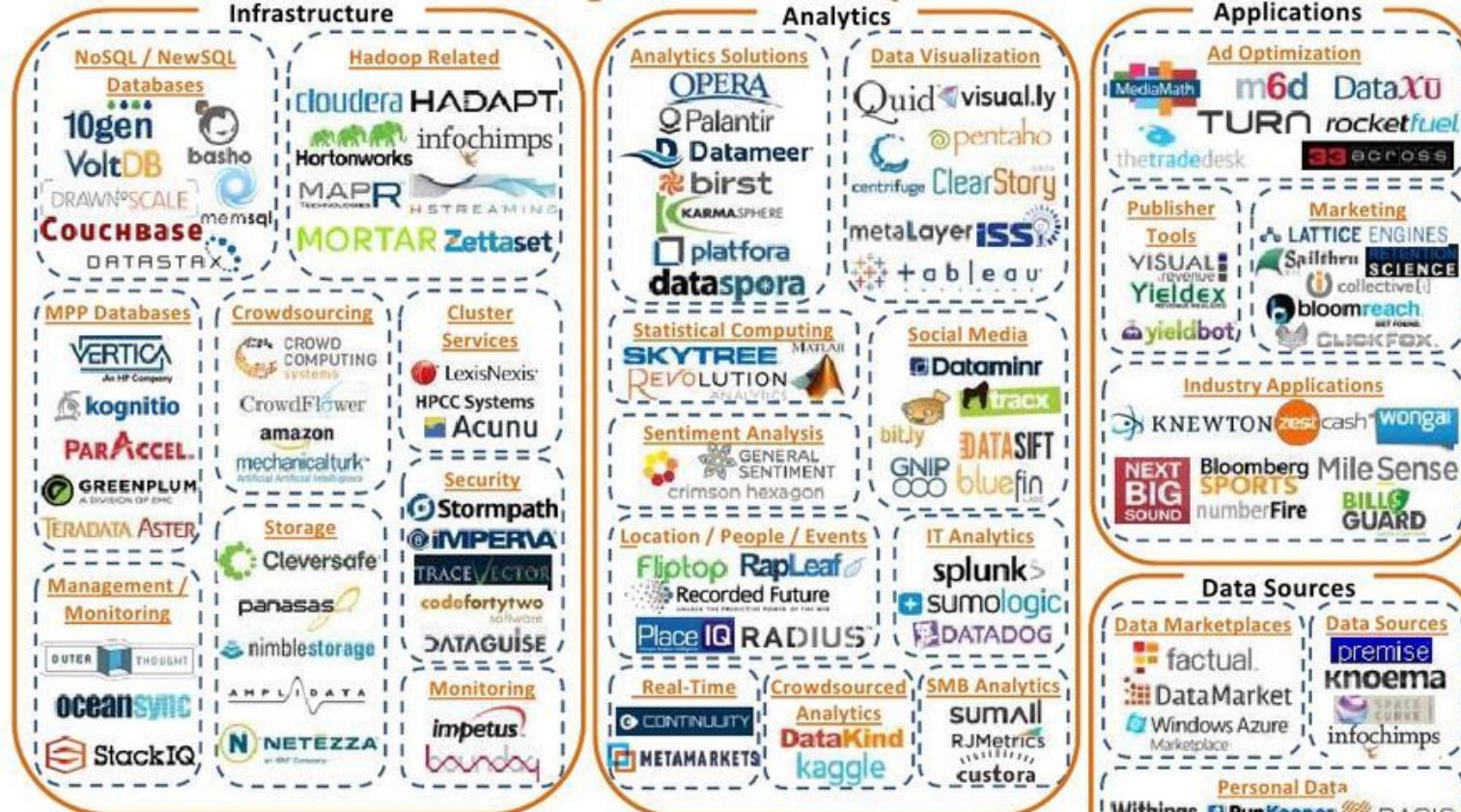


**How is data science done?**

julia

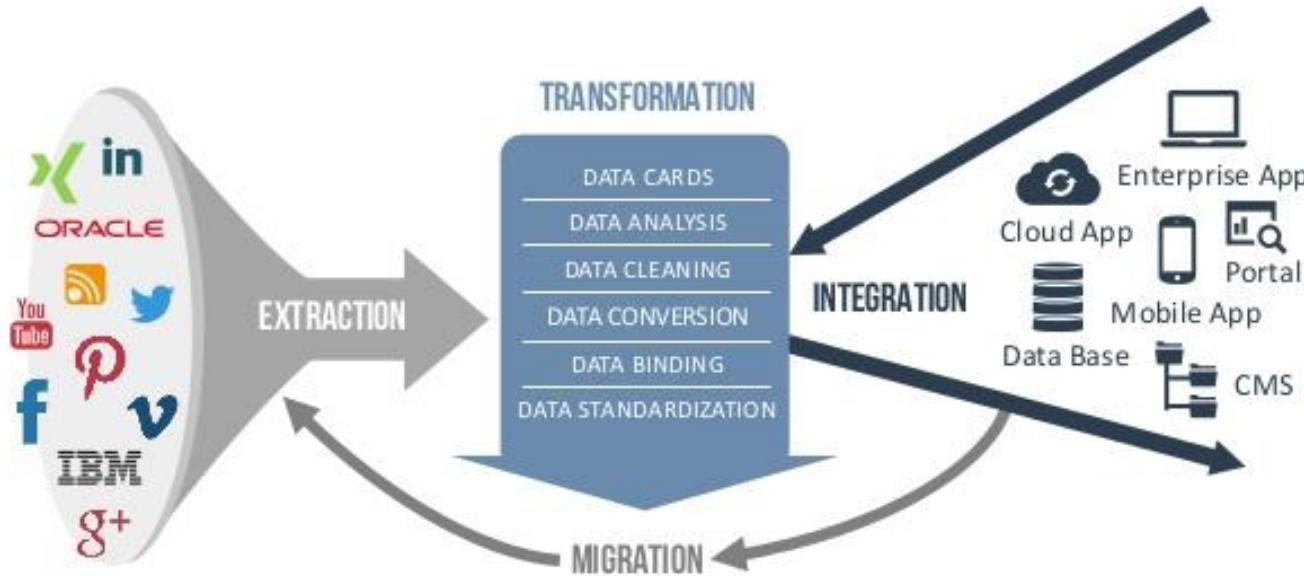
R





# EXTRACT – TRANSFORM – LOAD

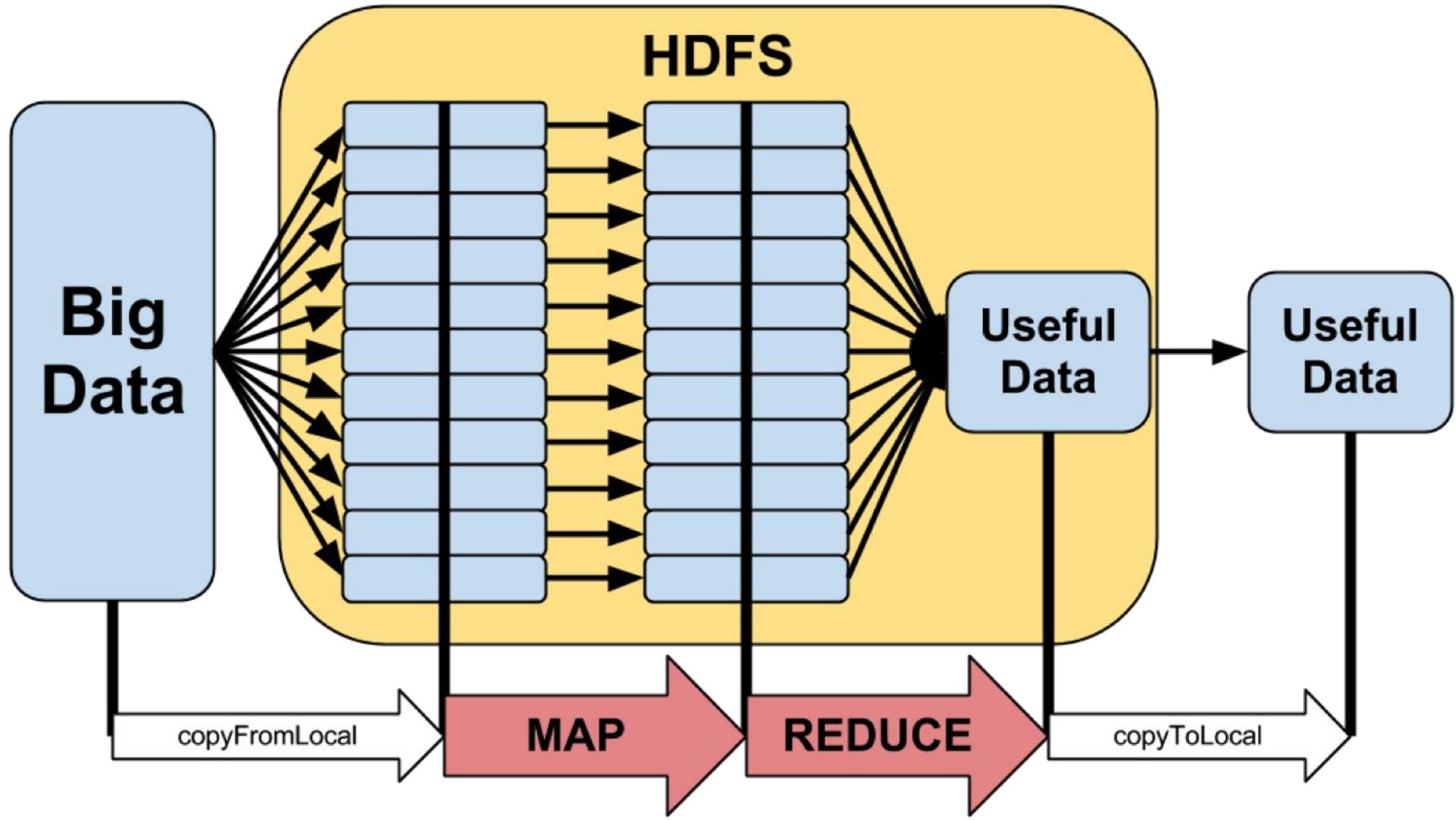
The ETL Process

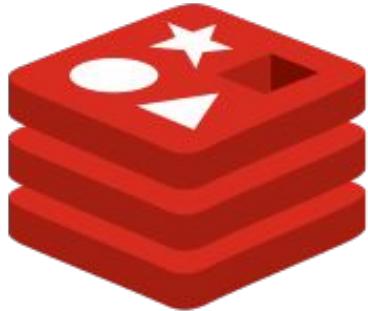


```
def fetch_businesses(name, area='New York', manual_override=0):
    area = area.lower().replace(' ', '-')
    name = name.lower().replace(' ', '-')
    i = 2
    # Run the first one through by hand.
    try:
        responses = [client.get_business("{0}-{1}".format(name, area))]
    # This can happen, and did, in the Dunkin' Donuts case.
    except BusinessUnavailable:
        responses = []
        pass
    # The rest are handled by a loop.
    while True:
        bus_id = "{0}-{1}-{2}".format(name, area, i)
        try:
            response = client.get_business(bus_id)
        except BusinessUnavailable:
            # We manually check trouble spots.
            # But see the TODO.
            if requests.get('http://www.yelp.com/biz/' + bus_id).status_code != requests.codes.ok:
                break
            else:
                # Increment the counter but don't include the troubled ID.
                i += 1
                continue
        responses += [response]
        i += 1
    print("Ended `fetch_businesses()` on:", "http://www.yelp.com/biz/" + bus_id)
    return responses
```

```
def frame(responses):
    """
    Given a list of yelp.obj.business_response.BusinessResponse objects like the one returns by fetch_businesses(),
    builds a coordinate-logging DataFrame out of them.
    """
    latitudes = [response.business.location.coordinate.latitude for response in responses]
    longitudes = [response.business.location.coordinate.longitude for response in responses]
    df = DataFrame({'latitude': latitudes, 'longitude': longitudes})
    df.index.name=responses[0].business.name
    return df

def map_coordinates(df):
    """
    Returns a folium map of all of the coordinates stored in a coordinate DataFrame, like the one returned by frame().
    """
    ret = folium.Map(location=[40.753889, -73.983611], zoom_start=11)
    for row in df.iterrows():
        ret.simple_marker([row[1]['latitude'], row[1]['longitude']])
    return ret
```



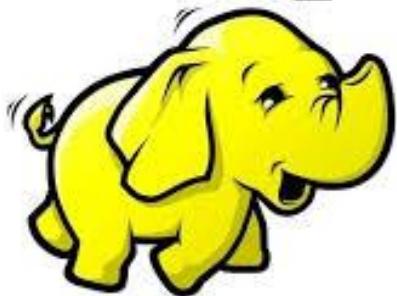


redis

MySQL®

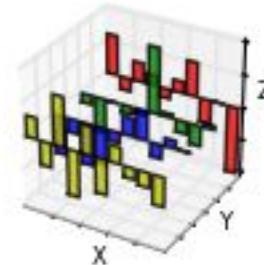
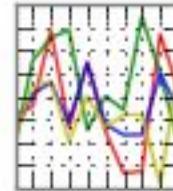
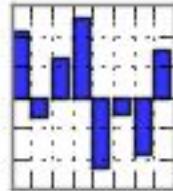
The MongoDB logo features a single green leaf icon followed by the word "mongoDB" in a lowercase, sans-serif font. The letters "mongo" are in a dark gray color, while "DB" is in a lighter gray shade.

*hadoop*

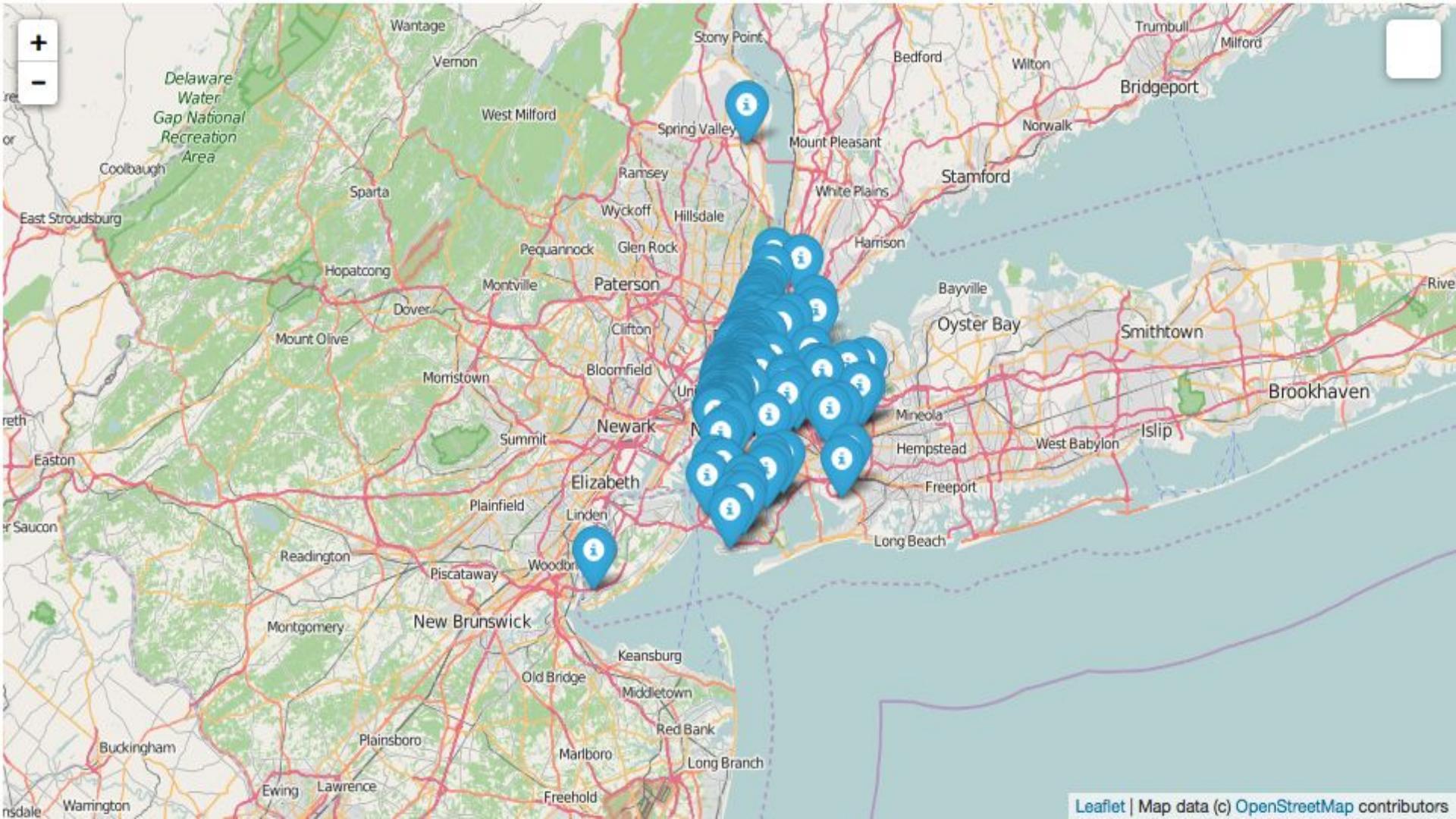


# pandas

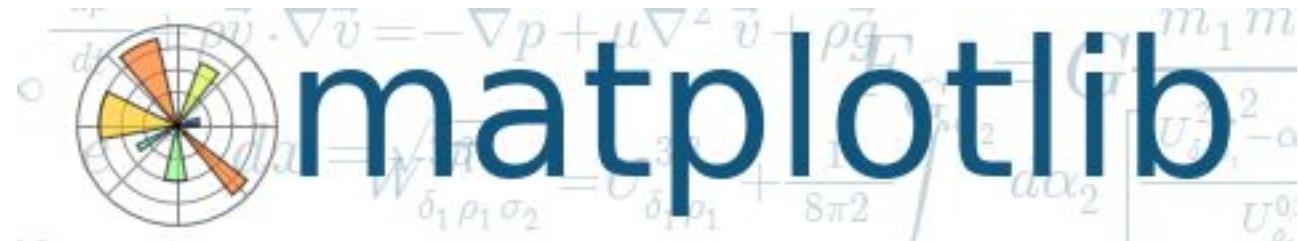
$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$

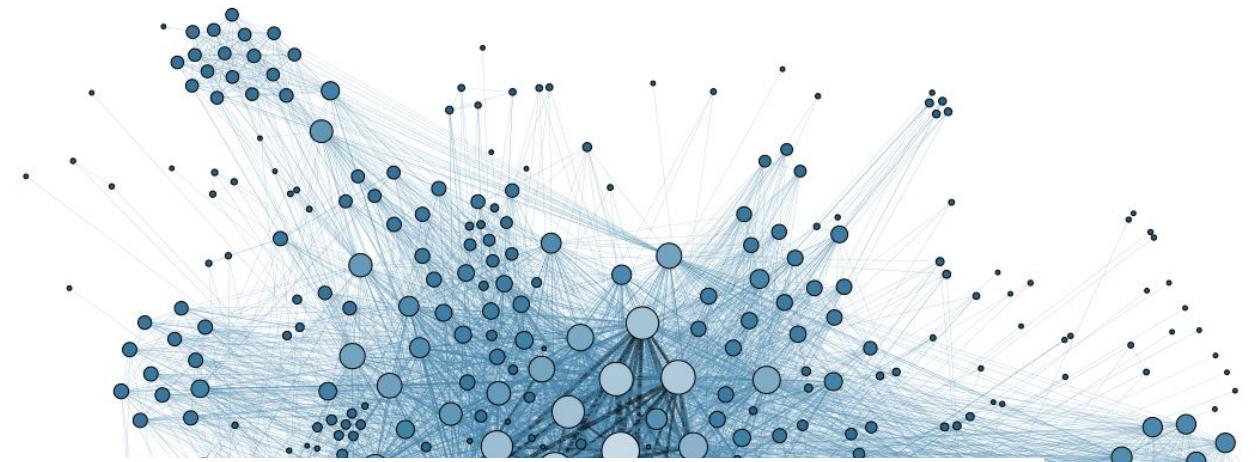


|     | account | campaign     | date                      | successes | trials | rate     |
|-----|---------|--------------|---------------------------|-----------|--------|----------|
| 455 | 1       | Campaign #76 | 2012-08-14 11:56:20 -0400 | 2         | 2      | 1.000000 |
| 449 | 1       | Campaign #78 | 2012-08-14 12:06:20 -0400 | 2         | 2      | 1.000000 |
| 438 | 1       | Campaign #87 | 2012-08-14 18:06:30 -0400 | 27        | 118    | 0.228814 |
| 431 | 1       | Campaign #95 | 2012-08-15 00:07:42 -0400 | 22        | 118    | 0.186441 |
| 422 | 1       | Campaign #99 | 2012-08-15 01:27:48 -0400 | 25        | 120    | 0.208333 |

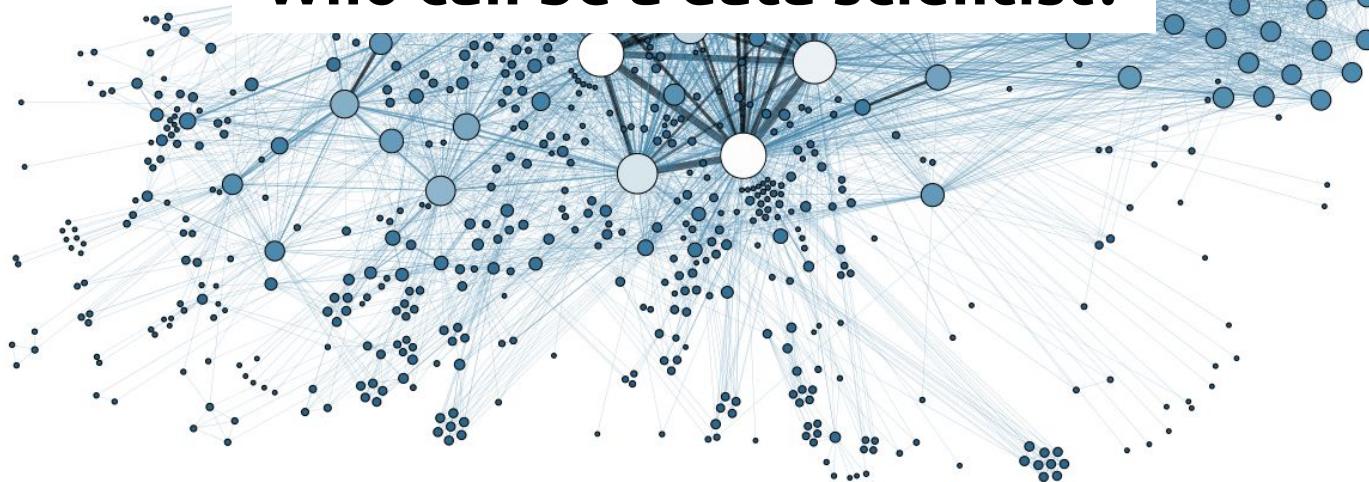






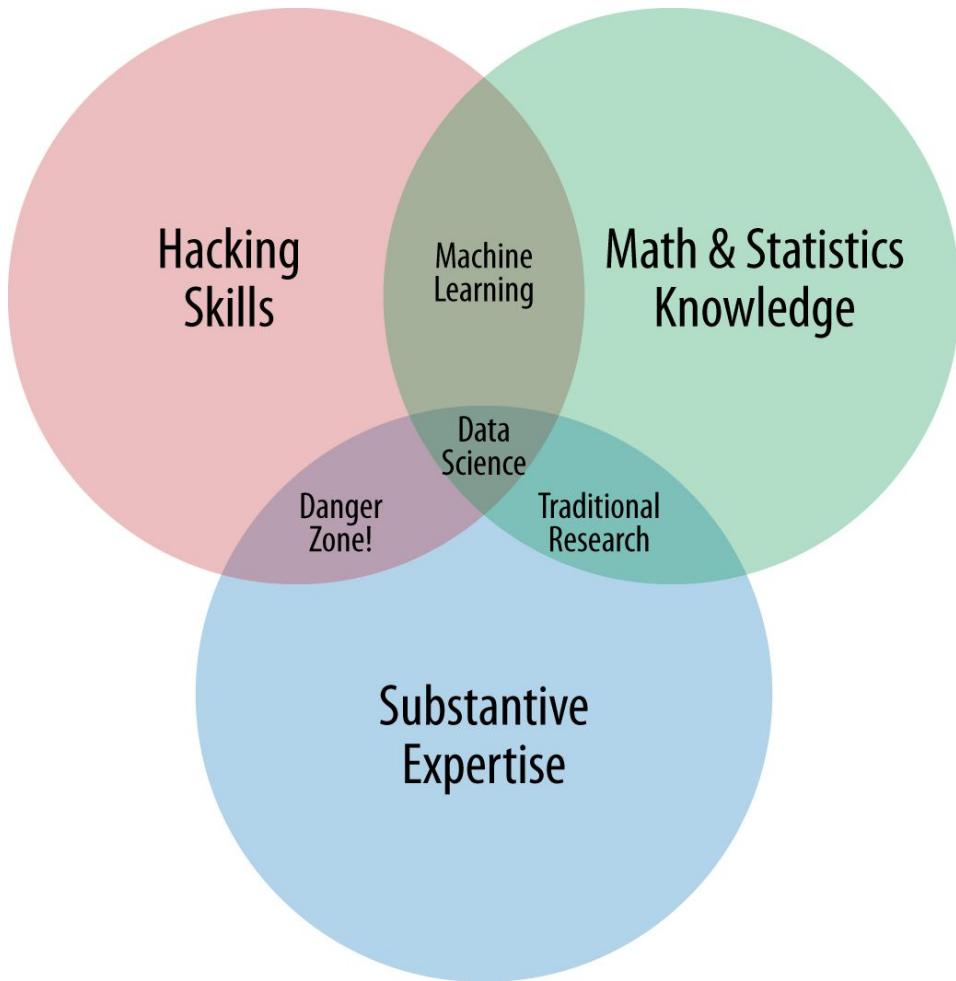


**Who can be a data scientist?**



You.





# **More than one way to bake a cake...**

## **MTH**

- \*Calc I/Calc II/Calc III
- MTH 4100 (Linear Algebra)
- MTH 4120 (Probability)
- MTH 4125 (Stochastic Processes)
- MTH 4130 (Mathematics of Statistics)
- MTH 4135 (Methods of Analysis)
- MTH 4140 (Graph Theory)
- MTH 4320 (Fundamental Algorithms)

## **CIS**

- \*CIS 3120 (Programming for Analytics)
- CIS 3400 (Database Mgmt. Systems)
- CIS 3710 (Business Analytics)
- CIS 3920 (Data Mining for Analytics)
- CIS 4400 (Data Warehousing)
- CIS 4170 (Data Visualization)

## **STA**

- \*STA 2100/2000 (Business Stats I)
- STA 3154 (Business Stats II)
- STA 3155 (Regression Models)
- Sneaking into graduate stats lectures.

Learning  
by Doing!

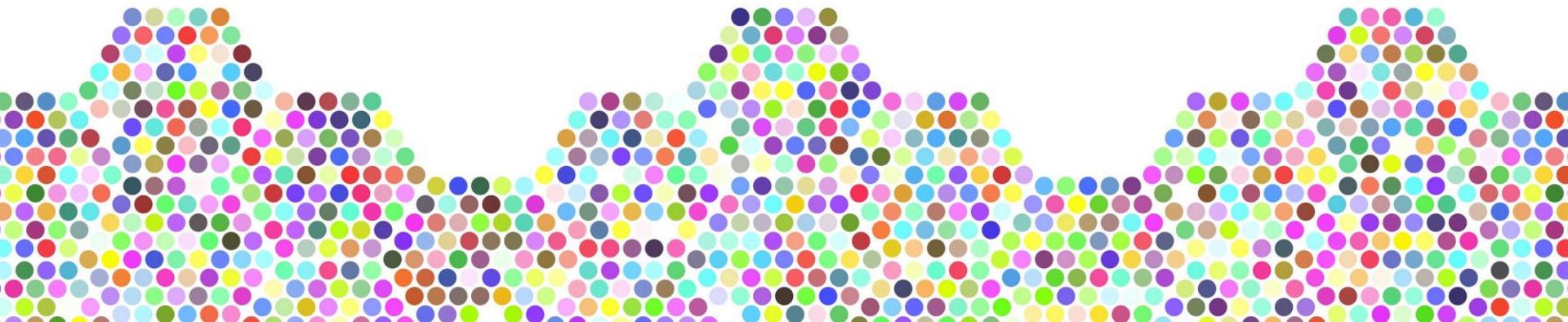


What's a Kaggle?

# The Home of Data Science

COMPETITIONS • DATASETS • CUSTOMER SOLUTIONS • JOBS BOARD

[Get started »](#)





Knowledge • 3,575 teams

## Titanic: Machine Learning from Disaster

Fri 28 Sep 2012

Sat 31 Dec 2016 (10 months to go)

### Dashboard

Home



Data



Make a submission



### Information



Description

Evaluation

Rules

Prizes

Frequently Asked Question...

Further Reading / Watching

Getting Started With Excel

Getting Started With Python

Getting Started With Python...

Getting Started With Rand...

New: Getting Started with R

Submission Instructions

### Forum



### Scripts



New Script

New Notebook

### Leaderboard



### Visualization



### 3,519 Scripts

Kaggle\_First\_pythonNotebook  
10 Votes / 15 hours ago / PythonDecision Tree Visualization &  
Submission  
10 Votes / 17 hours ago / Python[Competition Details](#) » [Get the Data](#) » [Make a submission](#)

## Predict survival on the Titanic using Excel, Python, R & Random Forests

See best practice code and explore visualizations of the Titanic dataset on Kaggle Scripts. Submit directly to the competition, no data download or local environment needed!

The sinking of the RMS Titanic is one of the most infamous shipwrecks in history. On April 15, 1912, during her maiden voyage, the Titanic sank after colliding with an iceberg, killing 1502 out of 2224 passengers and crew. This sensational tragedy shocked the international community and led to better safety regulations for ships.

One of the reasons that the shipwreck led to such loss of life was that there were not enough lifeboats for the passengers and crew. Although there was some element of luck involved in surviving the sinking, some groups of people were more likely to survive than others, such as women, children, and the upper-class.

In this challenge, we ask you to complete the analysis of what sorts of people were likely to survive. In particular, we ask you to apply the tools of machine learning to predict which passengers survived the tragedy.



# Titanic Dataset

| pclass | survived | name                             | sex    | age   | sibsp | parch | ticket   | fare     | cabin   | embarked | boat | body | home.dest                       |
|--------|----------|----------------------------------|--------|-------|-------|-------|----------|----------|---------|----------|------|------|---------------------------------|
| 1      | 1        | Allen, Miss. Elisabeth Walton    | female | 29    | 0     | 0     | 24160    | 211.3375 | B5      | S        | 2    |      | St Louis, MO                    |
| 1      | 1        | Allison, Master. Hudson Trevor   | male   | 0.917 | 1     | 2     | 113781   | 151.5500 | C22 C26 | S        | 11   |      | Montreal, PQ / Chesterville, ON |
| 1      | 0        | Allison, Miss. Helen Loraine     | female | 2     | 1     | 2     | 113781   | 151.5500 | C22 C26 | S        |      |      | Montreal, PQ / Chesterville, ON |
| 1      | 0        | Allison, Mr. Hudson Joshua Crei  | male   | 30    | 1     | 2     | 113781   | 151.5500 | C22 C26 | S        |      | 135  | Montreal, PQ / Chesterville, ON |
| 1      | 0        | Allison, Mrs. Hudson J C (Bessie | female | 25    | 1     | 2     | 113781   | 151.5500 | C22 C26 | S        |      |      | Montreal, PQ / Chesterville, ON |
| 1      | 1        | Anderson, Mr. Harry              | male   | 48    | 0     | 0     | 19952    | 26.5500  | E12     | S        | 3    |      | New York, NY                    |
| 1      | 1        | Andrews, Miss. Kornelia Theodo   | female | 63    | 1     | 0     | 13502    | 77.9583  | D7      | S        | 10   |      | Hudson, NY                      |
| 1      | 0        | Andrews, Mr. Thomas Jr           | male   | 39    | 0     | 0     | 112050   | 0.0000   | A36     | S        |      |      | Belfast, NI                     |
| 1      | 1        | Appleton, Mrs. Edward Dale (Cha  | female | 53    | 2     | 0     | 11769    | 51.4792  | C101    | S        | D    |      | Bayside, Queens, NY             |
| 1      | 0        | Artagaveytia, Mr. Ramon          | male   | 71    | 0     | 0     | PC 17609 | 49.5042  |         | C        |      | 22   | Montevideo, Uruguay             |
| 1      | 0        | Astor, Col. John Jacob           | male   | 47    | 1     | 0     | PC 17757 | 227.5250 | C62 C64 | C        |      | 124  | New York, NY                    |
|        |          |                                  |        |       |       |       |          | 7.5250   | C62 C64 | C        | 4    |      | New York, NY                    |
|        |          |                                  |        |       |       |       |          | 9.3000   | B35     | C        | 9    |      | Paris, France                   |
|        |          |                                  |        |       |       |       |          | 8.8500   |         | S        | 6    |      |                                 |

Identify data types of different attributes...

|   |   |                                 |        |    |   |   |         |
|---|---|---------------------------------|--------|----|---|---|---------|
| 1 | 1 | Barkworth, Mr. Algernon Henry   | male   | 80 | 0 | 0 | 27042   |
| 1 | 0 | Baumann, Mr. John D             | male   |    | 0 | 0 | PC 1731 |
| 1 | 0 | Baxter, Mr. Quigg Edmond        | male   | 24 | 0 | 1 | PC 1755 |
| 1 | 1 | Baxter, Mrs. James (Helene DeL  | female | 50 | 0 | 1 | PC 1755 |
| 1 | 1 | Bazzani, Miss. Albina           | female | 32 | 0 | 0 | 11813   |
| 1 | 0 | Beattie, Mr. Thomson            | male   | 36 | 0 | 0 | 13050   |
| 1 | 1 | Beckwith, Mr. Richard Leonard   | male   | 37 | 1 | 1 | 11751   |
| 1 | 1 | Beckwith, Mrs. Richard Leonard  | female | 47 | 1 | 1 | 11751   |
| 1 | 1 | Behr, Mr. Karl Howell           | male   | 26 | 0 | 0 | 111369  |
| 1 | 1 | Bidois, Miss. Rosalie           | female | 42 | 0 | 0 | PC 1775 |
| 1 | 1 | Bird, Miss. Ellen               | female | 29 | 0 | 0 | PC 1748 |
| 1 | 0 | Birnbaum, Mr. Jakob             | male   | 25 | 0 | 0 | 13905   |
| 1 | 1 | Bishop, Mr. Dickinson H         | male   | 25 | 1 | 0 | 11967   |
| 1 | 1 | Bishop, Mrs. Dickinson H (Heler | female | 19 |   |   | 11967   |

## VARIABLE DESCRIPTIONS

|           |  |
|-----------|--|
| Pclass    | Passenger Class (1 = 1st; 2 = 2nd; 3 = 3rd)                          |
| survival  | Survival (0 = No; 1 = Yes)   |
| name      | Name   |
| sex       | Sex  |
| age       | Age  |
| sibsp     | Number of Siblings/Spouses Aboard                                    |
| parch     | Number of Parents/Children Aboard                                    |
| ticket    | Ticket Number  |
| fare      | Passenger Fare (British pound)                                       |
| cabin     | Cabin  |
| embarked  | Port of Embarkation (C = Cherbourg; Q = Queenstown; S = Southampton) |
| boat      | Lifeboat   |
| body      | Body Identification Number   |
| home.dest | Home/Destination   |

**Titanic: Machine Learning from Disaster**

Fri 28 Sep 2012

Sat 31 Dec 2016 (10 months to go)

Home

Forum

Scripts

Leaderboard

New Notebook

New Script

## Scripts

Hottest

All Languages

All Output Types

My Scripts

- |      |  |        |  |             |
|------|--|--------|--|-------------|
| ▲ 10 | Kaggle_First_IpythonNotebook<br>last run 16 hours ago by <a href="#">Malai</a>                     | Python |  |             |
| ▲ 49 | Decision Tree Visualization & Submission<br>last run 2 months ago by <a href="#">Arda Yildirim</a> | R      |  | 2 comments  |
| ▲ 40 | A Journey through Titanic<br>last run 4 months ago by <a href="#">Omar El Gabry</a>                | Python |  | 17 comments |
| ▲ 3  | Random Forest Survivors<br>last run 1 week ago by <a href="#">Kushal Agrawal</a> (+29 / -7 / ~14)  | Python |  | 7 comments  |
| ▲ 4  | Titanic ML<br>last run 1 month ago by <a href="#">TomGlaser</a>                                    | Python |  |             |
| ▲ 14 | Large families not good for Survival<br>last run 3 months ago by <a href="#">Jason</a>             | R      |  |             |
| ▲ 2  | Popular Man Names<br>last run 4 days ago by <a href="#">Saba</a>                                   | Python |  |             |
| ▲ 2  | Popular Lady Name<br>last run 4 days ago by <a href="#">Saba</a>                                   | Python |  |             |



Knowledge • 3,575 teams

## Titanic: Machine Learning from Disaster

Fri 28 Sep 2012

Sat 31 Dec 2016 (10 months to go)

Dashboard

Competition Forum

Start Watching

Search

1 2 3 4 5 6 7 8 9 10 ... 23 24 &gt;

| Votes | Topics  | Replies | Views | Last Post                    |
|-------|---|---------|-------|------------------------------|
| 24    | <a href="#">Revised (and added) Tutorials</a><br>by Ramzi R, 22 months ago                        | 13      | 15303 | hehe<br>5 months ago         |
| 25    | <a href="#">Rolling Leaderboards</a><br>by William Cukierski, 2 years ago                         | 10      | 7762  | John Uckele<br>20 months ago |
| 0     | <a href="#">How to improve test accuracy score with Random fore...</a><br>by ajayram, 10 days ago | 2       | 184   | Masaal Dosai<br>15 hours ago |
| 0     | <a href="#"> A Journey through Titanic</a><br>by Omar El Gabry, 3 months ago                      | 16      | 0     | dalu610<br>2 days ago        |
| 0     | <a href="#">Creating the survival_table</a><br>by Ian McCarthy, 3 years ago                       | 7       | 2902  | mashiro<br>3 days ago        |
| 3     | <a href="#"> Random Forest Survivors</a><br>by Kushal Agrawal, 16 days ago                        | 6       | 0     | Kushal Agrawal<br>8 days ago |



Completed • \$10,000

## Allstate Claim Prediction Challenge

Wed 13 Jul 2011 – Wed 12 Oct 2011 (4 years ago)

### Dashboard

Home

Data

### Information

Description

Evaluation

Rules

Prizes

### Forum

### Leaderboard

Public

Private

### Private Leaderboard

1. Matt C
2. Owen
3. PlanetThanet
4. Old Dogs With New Tricks
5. dmvse/hgmi
6. Winter is Coming
7. Optimus
8. Kausing Much Damage
9. Glendog
10. Stefan

### Forum (33 topics)

A key part of insurance is charging each customer the appropriate price for the risk they represent.

Risk varies widely from customer to customer, and a deep understanding of different risk factors helps predict the likelihood and cost of insurance claims. The goal of this competition is to better predict Bodily Injury Liability Insurance claim payments based on the characteristics of the insured customer's vehicle.

Many factors contribute to the frequency and severity of car accidents including how, where and under what conditions people drive, as well as what they are driving.

Bodily Injury Liability Insurance covers other people's bodily injury or death for which the insured is responsible. **The goal of this competition is to predict Bodily Injury Liability Insurance claim payments based on the characteristics of the insured's vehicle.**

**Started:** 3:45 am, Wednesday 13 July 2011 UTC

**Ended:** 11:59 pm, Wednesday 12 October 2011 UTC (91 total days)

**Points:** this competition awarded standard ranking points

**Tiers:** this competition counted towards tiers

XPRIZE IS AN INNOVATION ENGINE. SEE WHO WE ARE &gt;

ABOUT  
XPRIZE

PRIZES

GRAND  
CHALLENGES

NEWS

BENEFACTORS

COMMUNITY

DONATE

HOME / NEWS &amp; DISCUSSION / HOW TO SAVE BILLIONS WITH A \$10K COMPETITION... IN 3 MONTHS

# HOW TO SAVE BILLIONS WITH A \$10K COMPETITION... IN 3 MONTHS

By Peter H. Diamandis M.D. on February 14, 2013

*This blog will show you how a \$10,000 Kaggle competition drove HUGE value for Allstate -- the U.S.'s largest publicly held insurer. I'll then share with you how YOU can use Kaggle to solve your challenges.*

As a reminder, Kaggle is a machine-learning, data-competition platform that has more than 45,000 machine-learning data specialists in its worldwide network who compete to analyze YOUR data and come up with insights and breakthrough algorithms. These 45,000 competing data scientists work during the day at some of the top Silicon Valley and global tech companies while, at night, driven by the challenge, they compete at Kaggle to demonstrate they are the best. As my friend Jeremy Howard, president and chief scientist at Kaggle told me, "It turns out that these people are applying themselves every day to real-world problems, and desire to apply what they've learned to other problems. That's why they come to Kaggle, and on Kaggle they find dozens and dozens of real-world challenges that people are desperate to solve."

During my interview with Jeremy, I asked him to tell me about Kaggle's most impressive success story. "With little doubt, it's the competition we did with Allstate." So here's the deal: Allstate, founded in 1931, is one of the world's largest insurance companies with \$32 billion in revenue and 70,000 employees. You can bet that this company, which lives and dies based on the quality of its data and algorithms, employs some of the most gifted actuarial specialists in the world. What I mean by this is that Allstate wants to be able to "know," based on your age, marital status, the kind of car you drive and where you live, exactly what the probability is that you will have an accident. This is how they set your rates and how the company makes its money. It's really that simple.

Allstate's formula for making a prediction based on the data is called a vehicle-risk assessment algorithm.

The competition that they chose to run with Allstate had to come up with a way to predict all the

JOIN THE REVOLUTION

GET EMAIL UPDATES ABOUT XPRIZE, TEAMS AND UPCOMING EVENTS.



**LOOP™**

POWERING XPRIZE

JOIN XPRIZE WITH A MONTHLY MEMBERSHIP.

**JOIN TODAY**

SUBMIT YOUR OWN

Intro to Data Science W... Database Courses Notes | H... Is a sundial accurate | ... History of timekeeping ... Atomic clock - Wikipedia File:Clock accuracy.jpg ... Sundial - Wikipedia, the... Allstate Purchase Predi... How to Save Billions wi...

www.xprize.org/news/blog/how-save-billions-10k-competition-3-months

XPRIZE IS AN INNOVATION ENGINE. SEE WHO WE ARE >

ABOUT XPRIZE PRIZES GRAND CHALLENGES NEWS BENEFACTORS COMMUNITY DONATE

GET UPDATES Email Address GO

LinkedIn YouTube Facebook Twitter



Kaggle to demonstrate they are the best. As my friend Jeremy Howard, president and chief scientist at Kaggle told me, "It turns out that these people are applying themselves every day to real-world problems, and desire to apply what they've learned to other problems. That's why they come to Kaggle, and on Kaggle they find dozens and dozens of real-world challenges that people are desperate to solve."

During my interview with Jeremy, I asked him to tell me about Kaggle's most impressive success story. "With little doubt, it's the competition we did with Allstate." So here's the deal: Allstate, founded in 1931, is one of the world's largest insurance companies with \$32 billion in revenue and 70,000 employees. You can bet that this company, which lives and dies based on the quality of its data and algorithms, employs some of the most gifted actuarial specialists in the world. What I mean by this is that Allstate wants to be able to "know," based on your age, marital status, the kind of car you drive and where you live, exactly what the probability is that you will have an accident. This is how they set your rates and how the company makes its money. It's really that simple.

Allstate's formula for making a prediction based on the data is called a vehicle-risk assessment algorithm. The competition that they chose to run with Allstate asked teams to improve on the company's internally developed algorithm.

The prize offered was only \$10,000 but, remarkably, somewhere around 600 data scientists, composing 300 teams, took part in the competition. "The \$10,000 had little to do with it. That just makes it a bit of fun," continued Jeremy. "The reason the data scientists chose to compete was all about demonstrating their creativity. They wanted to reach the top of the leaderboard against other brilliant people in the field."

The results of this competition were nothing less than stunning. In the end, the outside experts (i.e., tapping into the global cognitive surplus) blew away Allstate's internal experts. The winners of the Kaggle competition demonstrated a 340 percent improvement in predictive accuracy over Allstate's best internal algorithm!

I can't imagine how much such an improvement is worth, but I have little doubt that this \$10,000 purse will eventually drive hundreds of millions if not billions of dollars of additional profits. Jeremy, who spent 10 years in the insurance business himself, was shocked at the result. "I can tell you that Allstate's actuarial department is amongst the best in the world, and unless you are familiar with the insurance marketplace it's hard to understand how huge of an achievement this Kaggle competition represents."



SUBMIT YOUR OWN  
**PRIZE IDEA**

Email Address

 Receive XPRIZE newsletters and announcements.

**SUBMIT**



Completed • \$10,000 • 364 teams

## Right Whale Recognition

Thu 27 Aug 2015 – Thu 7 Jan 2016 (50 days ago)

### Dashboard

[Home](#)[Data](#)[Make a submission](#)[Information](#)[Description](#)[Evaluation](#)[Rules](#)[Prizes](#)[Creating a "Face Detector"](#)[Timeline](#)[Forum](#)[Leaderboard](#)[Public](#)[Private](#)

### Private Leaderboard

## 1. deepsense.io

## 2. felixlaumon

## 3. SKE

## 4. threedB

## 5. AbdulWahab

## 6. Tsakalis Kostas

## 7. bawdyb

## 8. Left Whales

## 9. Anil Thomas

## 10. Doug Koch

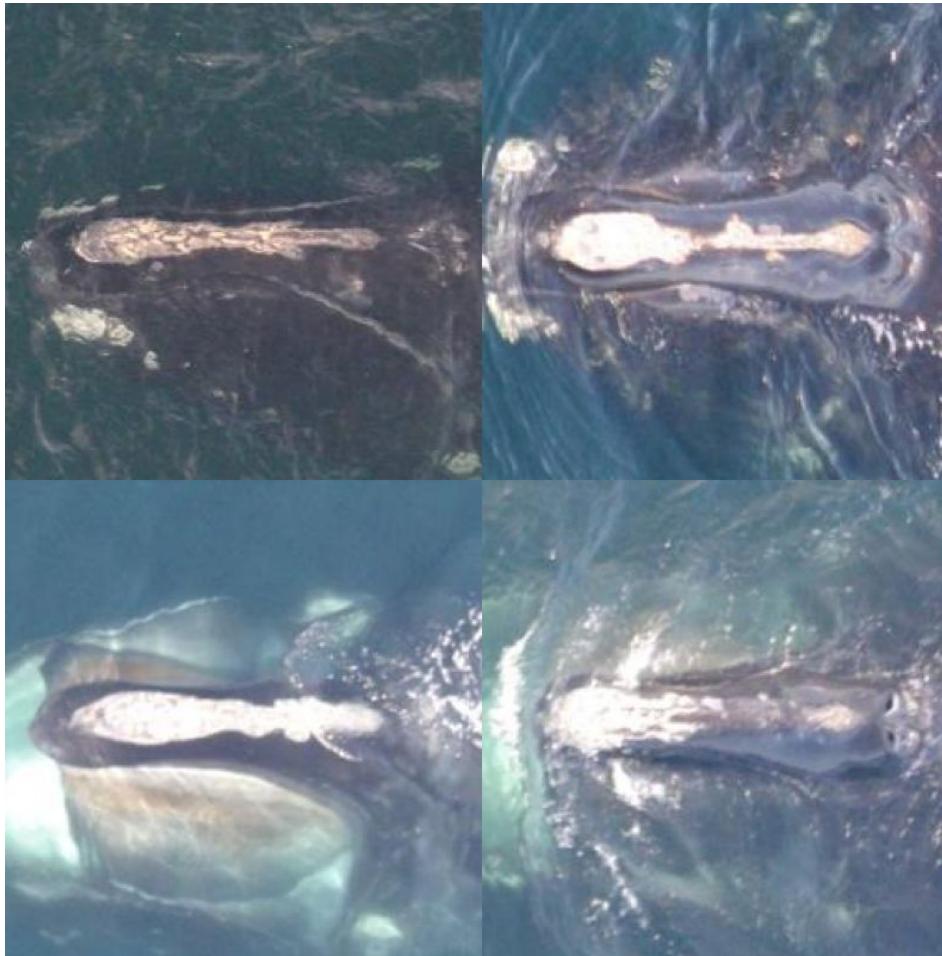
[Competition Details](#) » [Get the Data](#) » [Make a submission](#)

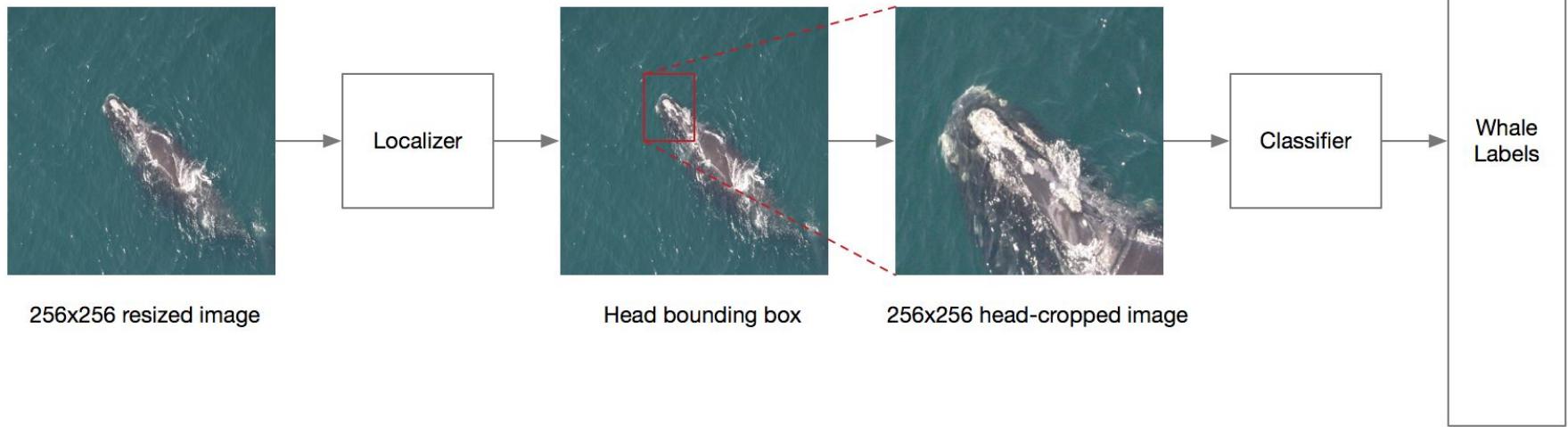
## Identify endangered right whales in aerial photographs

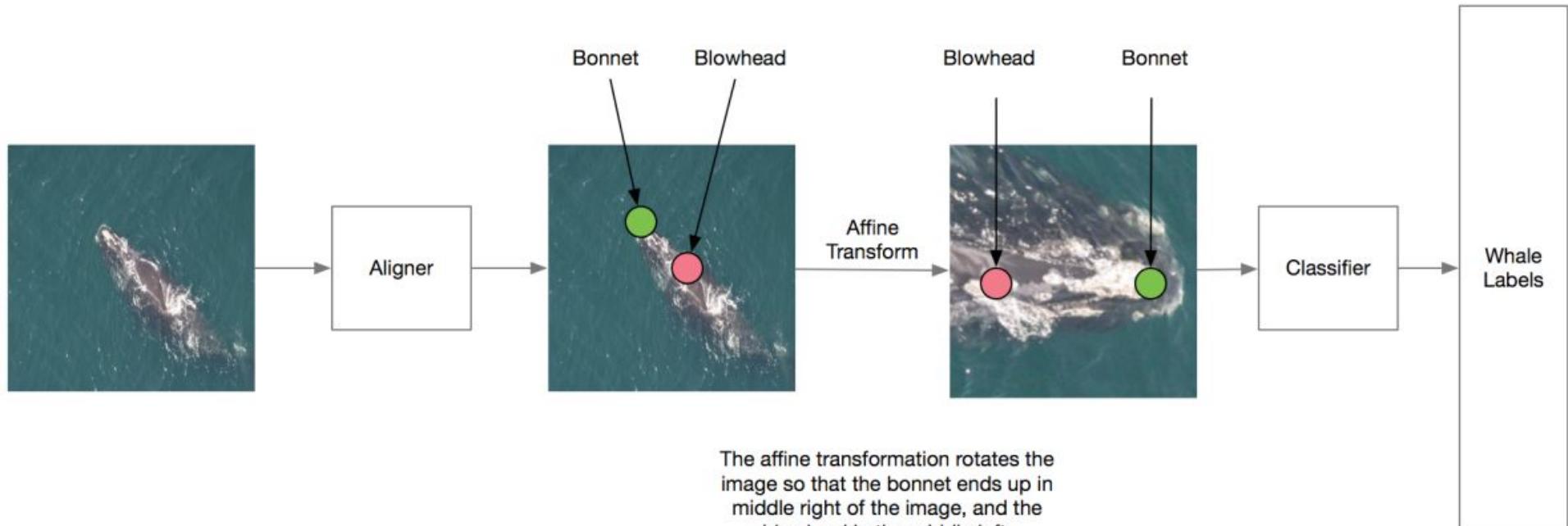
With fewer than 500 North Atlantic right whales left in the world's oceans, knowing the health and status of each whale is integral to the efforts of researchers working to protect the species from extinction.

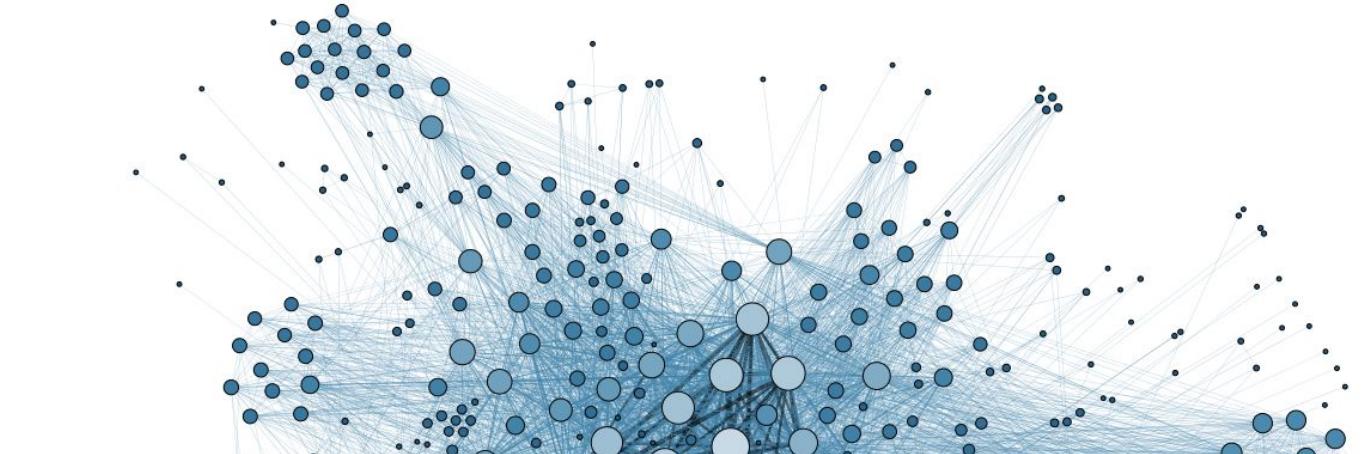
Currently, only a handful of very experienced researchers can identify individual whales on sight while out on the water. For the majority of researchers, identifying individual whales takes time, making it difficult to effectively target whales for biological samples, acoustic recordings, and necessary health assessments.



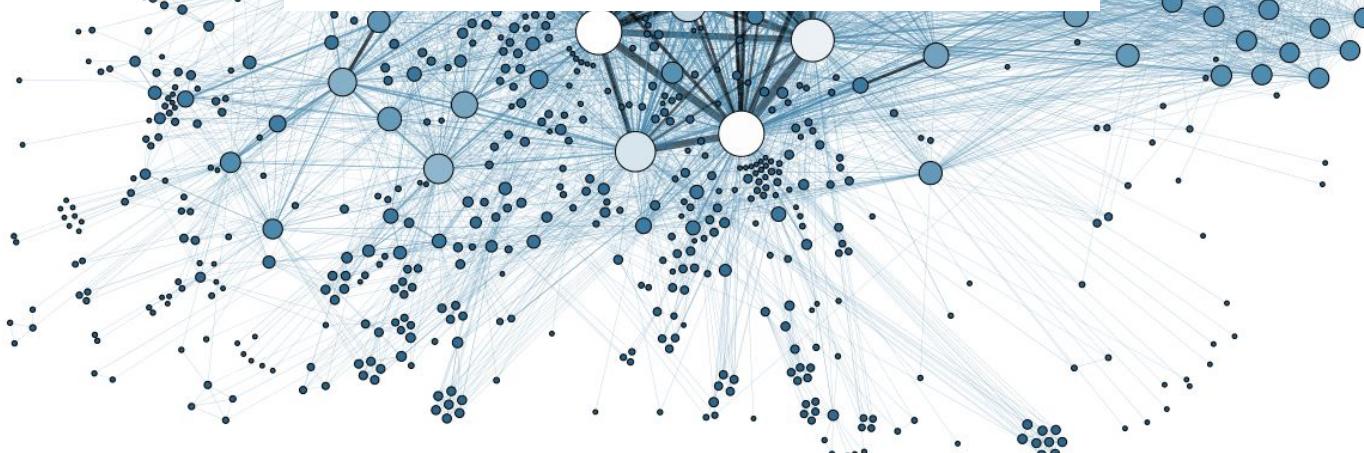








**What are you proposing?**



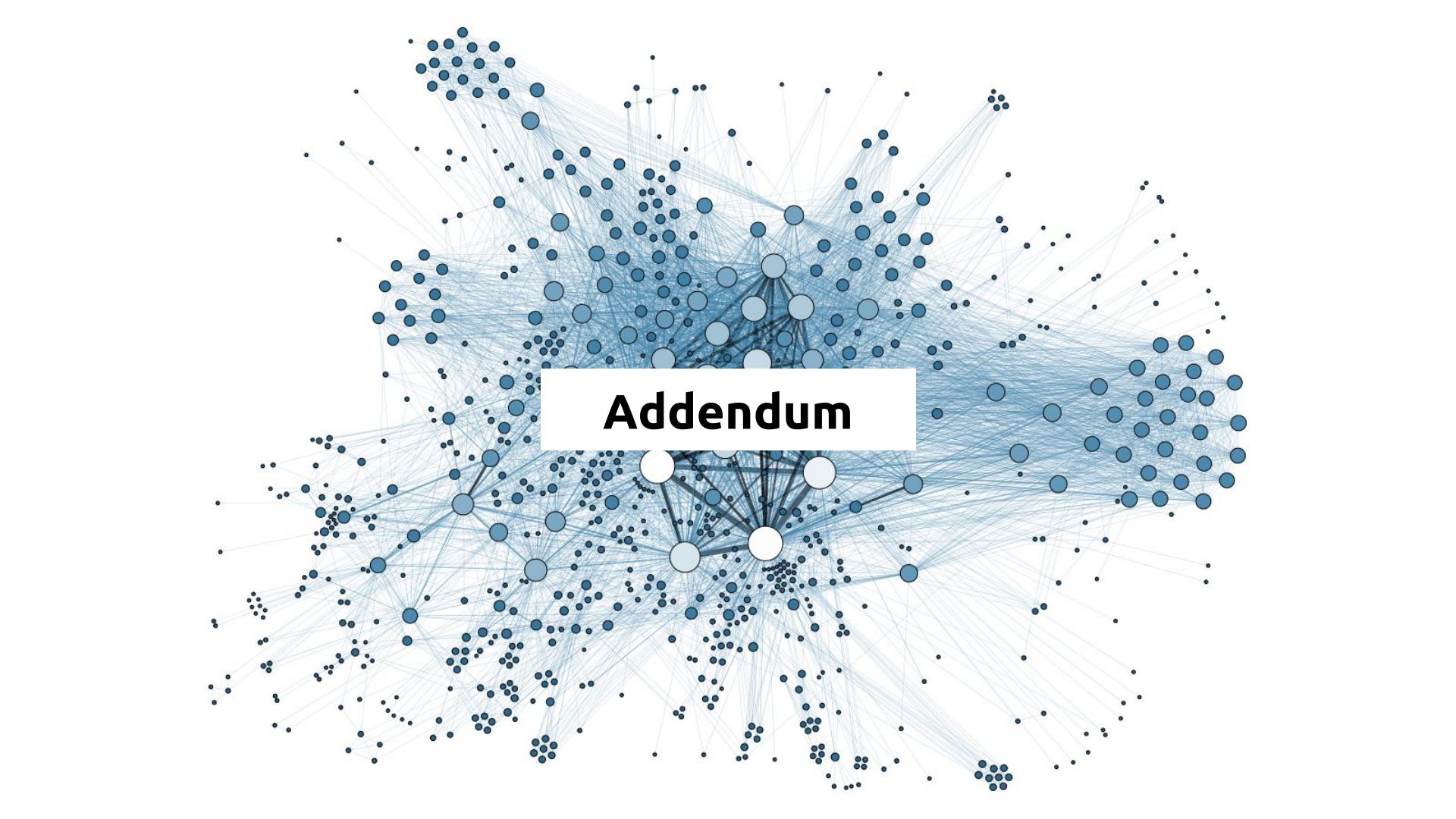
- ヴ (ツ) / -

- Professorial involvement?
  - Boot camps?
  - Outside meetups?
  - Travel budget?
- In-school competitions?
  - Hackathon?

All are possible...everything depends on the level of student (e.g. YOUR) interest.



**Thank you.**



# Addendum



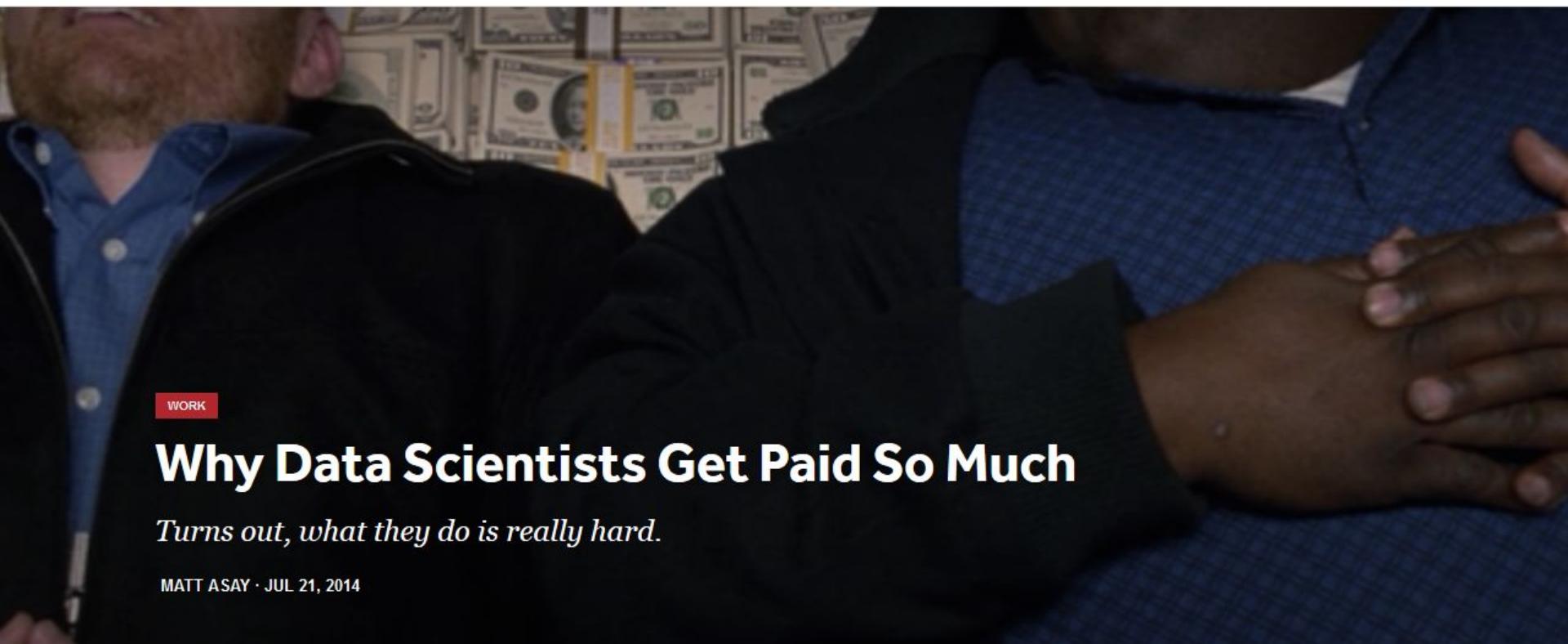
**Job outlook?**





NEWSLETTER  
Sign Up To Get  
ReadWrite In Your  
Inbox

CODE BUILD STRUCTURE CONNECT LEAD FUND OPERATE GROW



WORK

## Why Data Scientists Get Paid So Much

*Turns out, what they do is really hard.*

MATT ASAY · JUL 21, 2014



A complex network graph composed of numerous small black dots representing individual nodes, connected by a dense web of thin grey lines representing relationships or edges. The graph is highly interconnected, with several large, distinct clusters of nodes. A prominent feature is a central cluster of nodes, some of which are highlighted with larger, semi-transparent blue circles, indicating they are part of a specific group or community. This visual metaphor represents the interconnectedness and complexity of social networks or data relationships.

**Who are you anyway?**



O-Ring Damage Sustained by Launch Temperature (°F)

