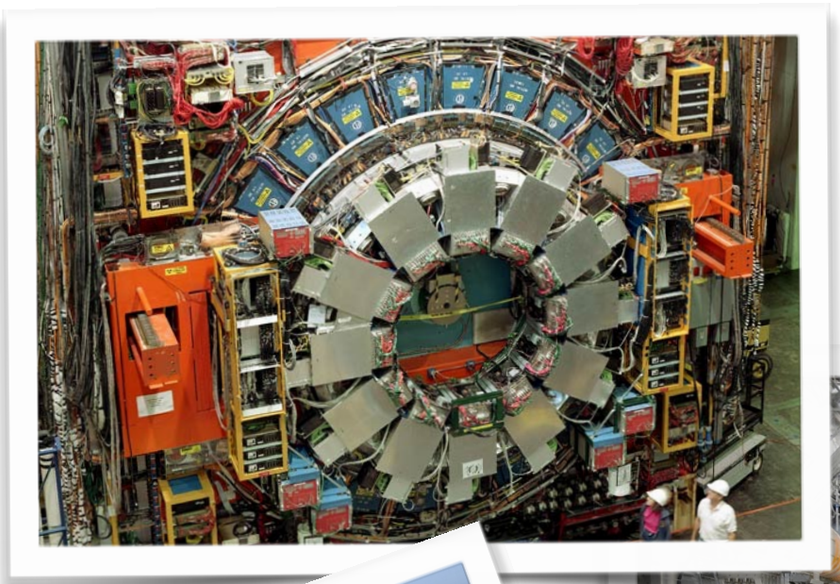
Kathy Copic

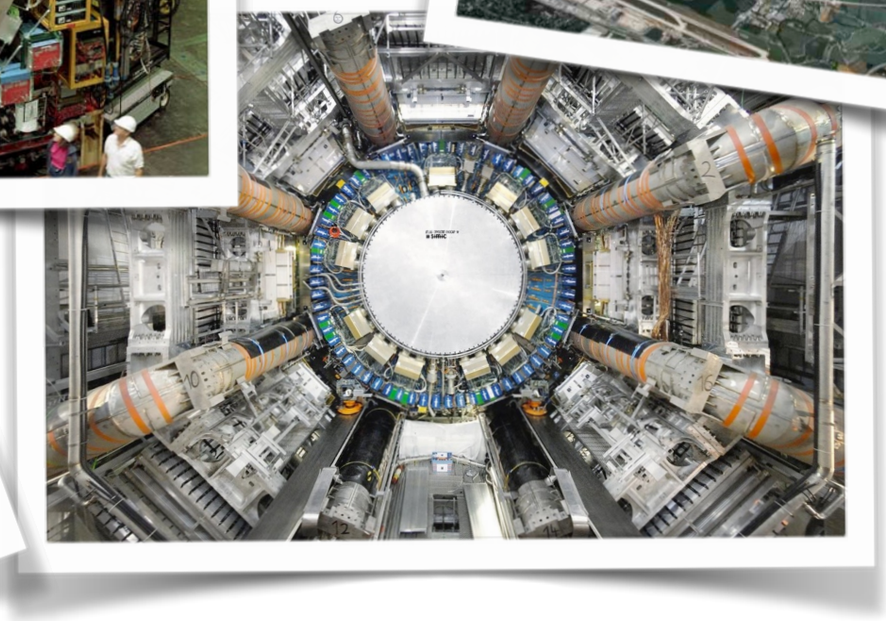
VP, Growth

The Insight Fellows Program

insightdatascience.com | insightdataengineering.com

insightdata.ai | insighthealthdata.com





INSIGHT

INSIGHT

Insight Fellows are **Data Scientists** and **Data Engineers** at:

Silicon Valley • New York • Boston • Seattle
+ many others…

Insight is an education startup that helps scientists and engineers start new careers.

Insight is an education startup that helps scientists and engineers start new careers.

Free 7-week fellowship programs

Insight is an education startup that helps scientists and engineers start new careers.

Emphasis on social learning

Insight is an education startup that helps scientists and engineers start new careers.

Learn from peers and industry mentors

Insight is an education startup that helps scientists and engineers start new careers.




Interview successfully for jobs!

Insight is an education startup that helps scientists and engineers start new careers.

1

Identify + recruit potential data scientists

>> Since 2012, we've reviewed 20k applications and conducted 1000s of interviews.

Insight is an education startup that helps scientists and engineers start new careers.

1

Identify + recruit potential data scientists

2

Onboard + train them as a team

>> Each team member produces a functional data product prototype within their first 20 days.

Insight is an education startup that helps scientists and engineers start new careers.

1

Identify + recruit potential data scientists

2

Onboard + train them as a team

3

Release them back into the wild

>> Over 1,000 Insight alumni have joined teams at close to 300 companies in the last 5 years.

Insight is an education startup that helps scientists and engineers start new careers.

1

Identify + recruit potential data scientists

2

Onboard + train them as a team

3

Release them back into the wild

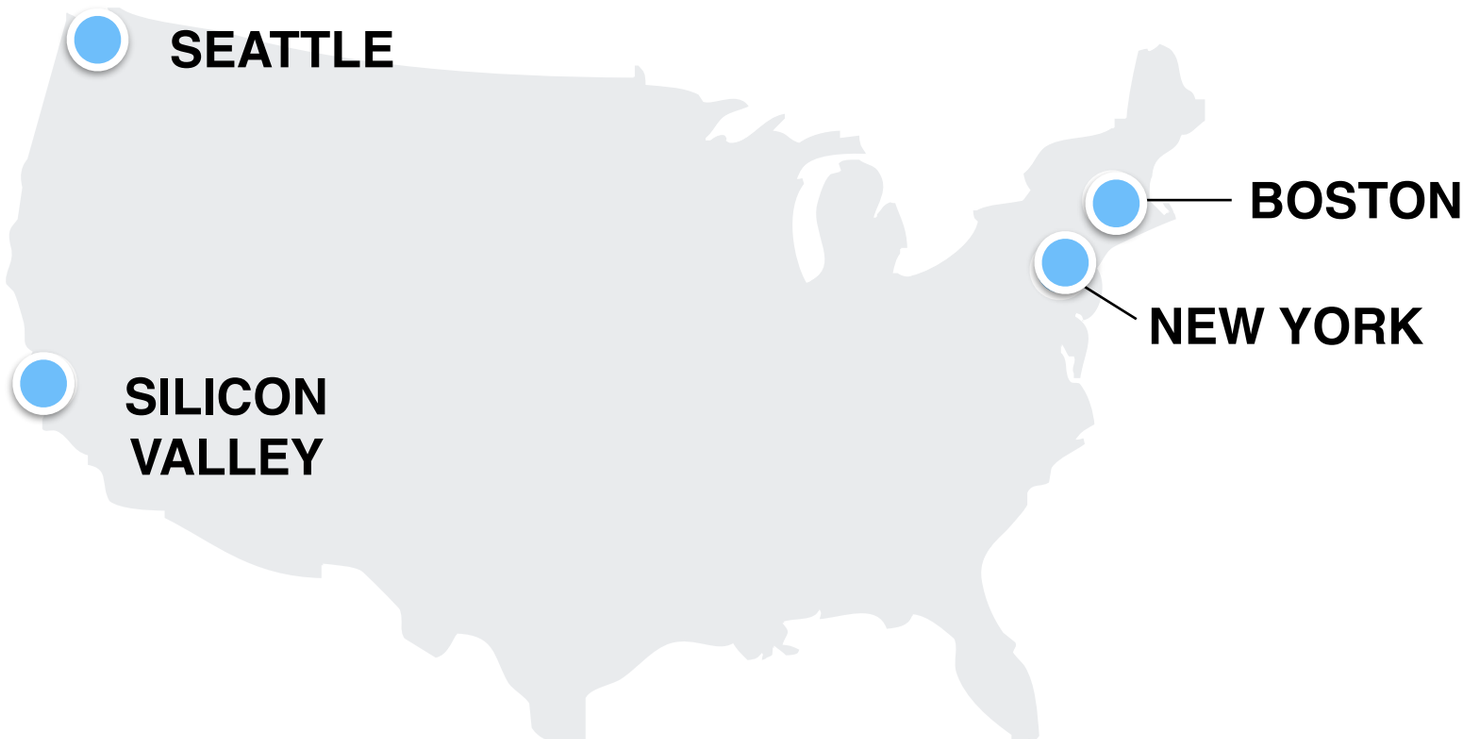
4

Repeat

>> 3x per year, in 4 cities, for 4 separate fields:
Data Science, Health Data Science, Data Engineering, & AI.

INSIGHT

LOCATIONS

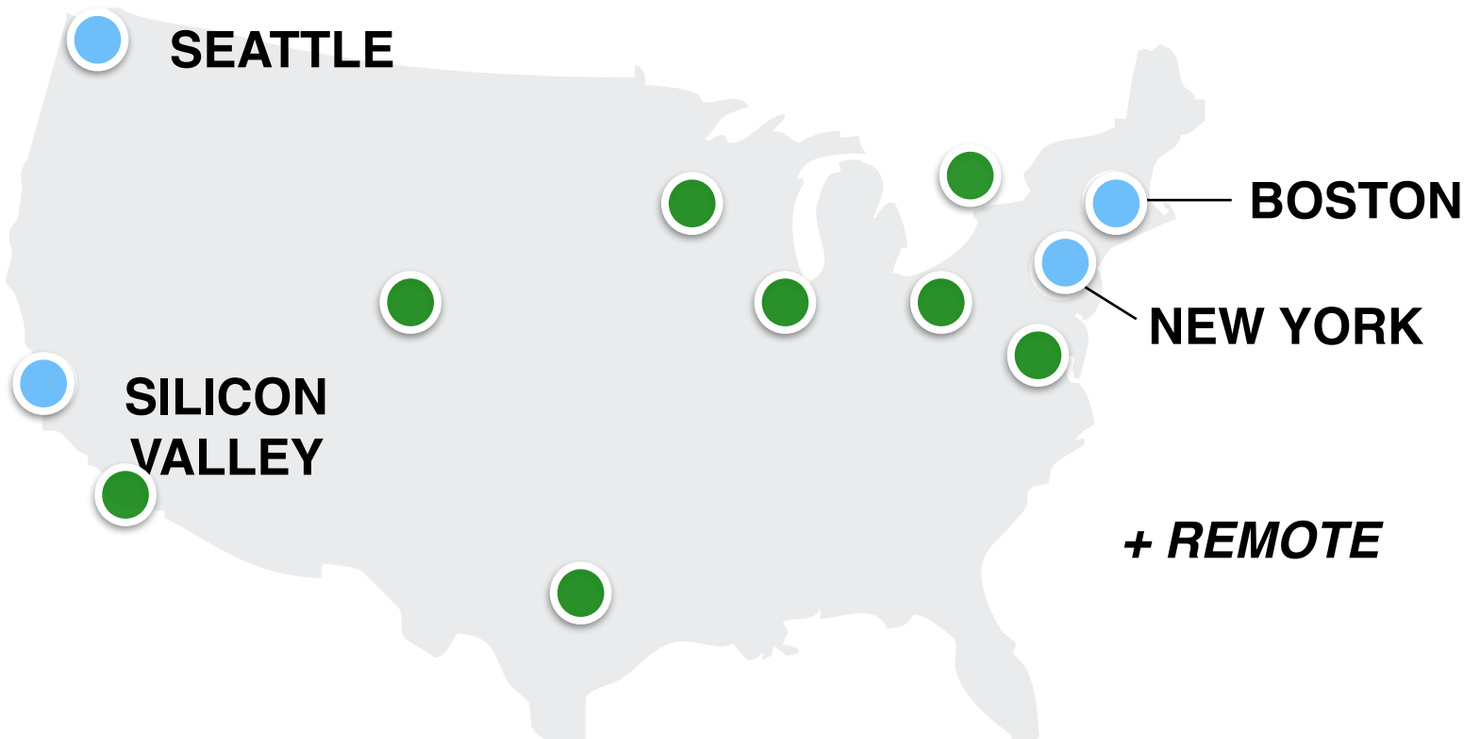


PROGRAMS

- DATA SCIENCE
- DATA ENGINEERING
- HEALTH DATA SCIENCE
- ARTIFICIAL INTELLIGENCE

INSIGHT

LOCATIONS



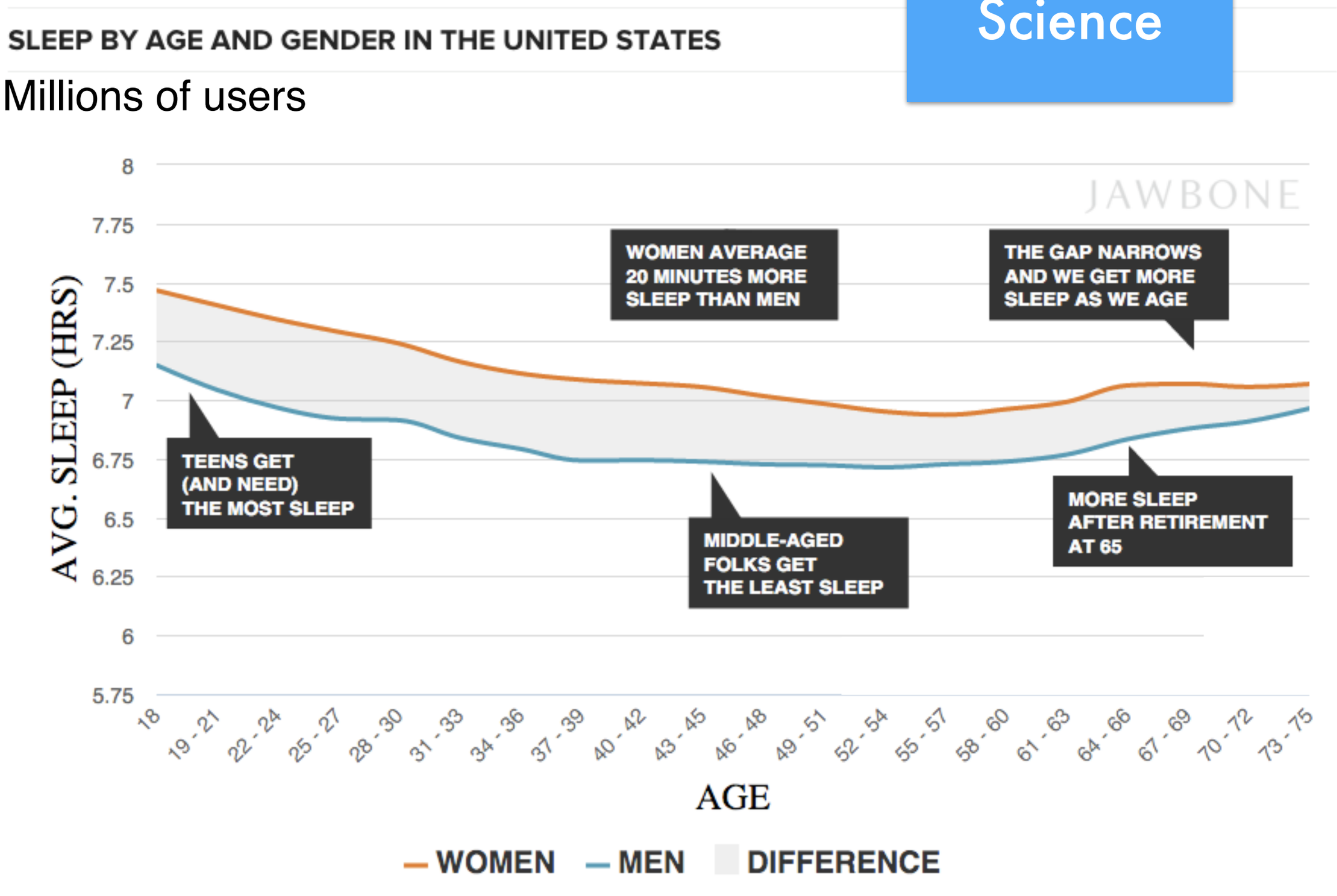
PROGRAMS

- DATA SCIENCE
- DATA ENGINEERING
- HEALTH DATA SCIENCE
- ARTIFICIAL INTELLIGENCE

Data Careers cover a wide space

Analytics & Decision Science	Data Products	AI & Deep Learning	Data Engineering

Analytics & Decision Science



[Featured in Nature]

[Last week's NYT]

Data Products

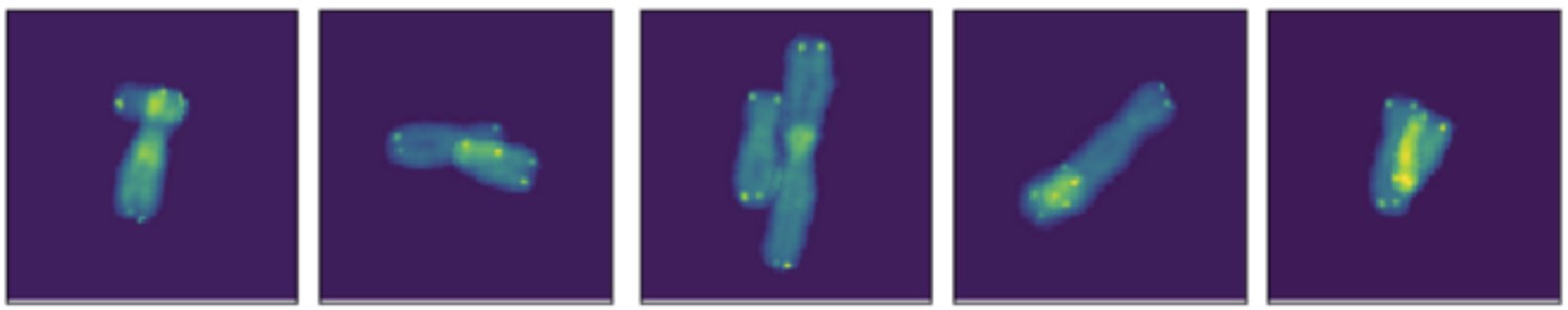
Products
where Data is at the
Core

AI & Deep Learning

When the usual tools just won't do

[Lily Hu's Project]

Input



Ground Truth

Prediction

[Data Ecosystem Map]

Data Engineering

Volume
Velocity
Variety

INSIGHT

Find out more about the Insight Data Engineering Fellows Program and Data Labs

Mouse over each box for more details.

Click on each technology for resources to get started.

Last updated Oct 26th, 2015
You can also find the previous version here.

Ingestion
1. Kafka
2. Logstash
3. RabbitMQ
4. Fluentd
5. AWS Kinesis

File Format
1. Avro
2. ProtoBuf
3. Thrift
4. Parquet
5. ORC Files

File System
1. HDFS
2. AWS S3
3. Azure
4. Tachyon
5. Ceph

Batch Processing
1. Spark
2. Hadoop MapReduce
3. AWS EMR
4. Flink
5. Tez

Stream Processing
1. Storm
2. Spark Streaming
3. AWS Lambda
4. Samza
5. Flink

High-Level MR
1. Pig
2. Cascading
3. Hadoop Streaming
4. Cascalog

Batch ML
1. H2O
2. Mahout
3. Spark MLlib
4. FlinkML

Batch Graph
1. GraphLab
2. Giraph
3. Spark GraphX
4. Hama

Batch SQL
1. Hive
2. Presto
3. Drill
4. Impala

General management tools for data pipelines

Cluster Management
1. Docker
2. Zookeeper
3. YARN
4. Mesos
5. Hue

Scheduling/Monitoring
1. Luigi
2. Airflow
3. Nagios
4. Graphite
5. Azkaban

Data Store
Transactions
Analytics
Uptime Critical
Search
Graph
Geospatial
Time Series
Cache

Web Framework
1. Ruby on Rails
2. Node.js
3. Django
4. AngularJS
5. Flask

Data Visualization
1. D3
2. Tableau
3. Leaflet
4. Highcharts
5. Kibana

Transactions
1. MySQL
2. Oracle
3. PostgreSQL

Analytics
1. AWS Redshift
2. Vertica
3. HBase

Uptime Critical
1. Cassandra
2. Riak
3. AWS DynamoDB

Search
1. Elasticsearch
2. Solr
3. MongoDB

Graph
1. Neo4j
2. OrientDB
3. ArangoDB

Geospatial
1. CouchDB
2. PostGIS
3. Elasticsearch

Time Series
1. InfluxDB
2. Cassandra
3. Druid

Cache
1. Redis
2. Memcached
3. Hazelcast

4 Reasons People want to Hire Physicists....
and they are NOT the Reasons you'd Expect!!!
(Well, OK maybe 1 of them is.)

Michael Kirby

Jan 4 at 10:45am • Chicago, Illinois •

Today's combinatorics problem: The cafeteria has 7 types of cookies. Kirby likes 3 of the 7 types of cookies. Everyday, the cafeteria puts out 4 types of cookies for the day. How often does Kirby get to enjoy all of the cookies that he likes? Never. It has never happened in 5 years.

41 likes 23 comments

Like Comment Share

Michael Kirby
Jan 4 at 10:45am • Chicago, Illinois •

Today's combinatorics problem: The cafeteria has 7 types of cookies. Kirby likes 3 of the 7 types of cookies. Everyday, the cafeteria puts out 4 types of cookies for the day. How often does Kirby get to enjoy all of the cookies that he likes? Never. It has never happened in 5 years.

41 likes 23 comments

Like Comment Share

Jon Wilson

There are 35 possible combinations (7 choose 4). Of those, 4 combinations contain all three types you enjoy. Thus, on any given day, there is about an 11.4% chance that you will get all three types you enjoy.

There are about 260 (5*52) weekdays per year. In five years, there are about 1300 weekdays. So, from the binomial distribution, we can calculate that the odds of having 0 successes in 1300 trials, with a success probability of 11.4% is....

3 x 10^-69

Jan 4 at 11:32 AM • Like • 9 • Reply

Michael Kirby
Jan 4 at 10:45am • Chicago, Illinois •

Today's combinatorics problem: The cafeteria has 7 types of cookies. Kirby likes 3 of the 7 types of cookies. Everyday, the cafeteria puts out 4 types of cookies for the day. How often does Kirby get to enjoy all of the cookies that he likes? Never. It has never happened in 5 years.

41 likes 23 comments

Like Comment Share

Jon Wilson
There are 35 possible combinations (7 choose 4). Of those, 4 combinations contain all three types you enjoy. Thus, on any given day, there is about an 11.4% chance that you will get all three types you enjoy.

There are about 260 (5*52) weekdays per year. In five years, there are about 1300 weekdays. So, from the binomial distribution, we can calculate that the odds of having 0 successes in 1300 trials, with a success probability of 11.4% is....

3 x 10^-69

Jan 4 at 11:32 AM • Like • 9 • Reply

Michael Kirby
Jan 4 at 10:45am • Chicago, Illinois •

Today's combinatorics problem: The cafeteria has 7 types of cookies. Kirby likes 3 of the 7 types of cookies. Everyday, the cafeteria puts out 4 types of cookies for the day. How often does Kirby get to enjoy all of the cookies that he likes? Never. It has never happened in 5 years.

41 likes 23 comments

Like Comment Share

Jon Wilson
There are 35 possible combinations (7 choose 4). Of those, 4 combinations contain all three types you enjoy. Thus, on any given day, there is about an 11.4% chance that you will get all three types you enjoy.

There are about 260 (5*52) weekdays per year. In five years, there are about 1300 weekdays. So, from the binomial distribution, we can calculate that the odds of having 0 successes in 1300 trials, with a success probability of 11.4% is....

3 x 10^-69

Jan 4 at 11:32 AM • Like • 9 • Reply

Andy Hocker
I knew someone would crunch the numbers

Jan 4 at 12:53 PM • Like • 1 • Reply

Michael Kirby
Jan 4 at 10:45am • Chicago, Illinois •

Today's combinatorics problem: The cafeteria has 7 types of cookies. Kirby likes 3 of the 7 types of cookies. Everyday, the cafeteria puts out 4 types of cookies for the day. How often does Kirby get to enjoy all of the cookies that he likes? Never. It has never happened in 5 years.

41 likes 23 comments

Like Comment Share

Jon Wilson
There are 35 possible combinations (7 choose 4). Of those, 4 combinations contain all three types you enjoy. Thus, on any given day, there is about an 11.4% chance that you will get all three types you enjoy.

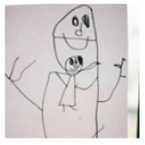
There are about 260 (5*52) weekdays per year. In five years, there are about 1300 weekdays. So, from the binomial distribution, we can calculate that the odds of having 0 successes in 1300 trials, with a success probability of 11.4% is....

3 x 10^-69

Jan 4 at 11:32 AM • Like • 9 • Reply

Andy Hocker
I knew someone would crunch the numbers

Jan 4 at 12:53 PM • Like • 1 • Reply



Michael Mulhearn
Clearly our assumption of cookie variety universality is incorrect. For instance, suppose the varieties you like are all made by the same person, and the most varieties that one person can make is two varieties. Or perhaps you like the most expensive varieties, and the most the manager will allow them to put out is two expensive varieties.

Jan 4 at 1:00 PM • Like • 6 • Reply

Michael Kirby
Jan 4 at 10:45am • Chicago, Illinois •

Today's combinatorics problem: The cafeteria has 7 types of cookies. Kirby likes 3 of the 7 types of cookies. Everyday, the cafeteria puts out 4 types of cookies for the day. How often does Kirby get to enjoy all of the cookies that he likes? Never. It has never happened in 5 years.

41 likes 23 comments

Like Comment Share

Jon Wilson
There are 35 possible combinations (7 choose 4). Of those, 4 combinations contain all three types you enjoy. Thus, on any given day, there is about an 11.4% chance that you will get all three types you enjoy.

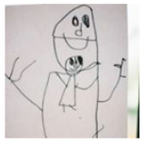
There are about 260 (5*52) weekdays per year. In five years, there are about 1300 weekdays. So, from the binomial distribution, we can calculate that the odds of having 0 successes in 1300 trials, with a success probability of 11.4% is....

3 x 10^-69

Jan 4 at 11:32 AM • Like • 9 • Reply

Andy Hocker
I knew someone would crunch the numbers

Jan 4 at 12:53 PM • Like • 1 • Reply



Michael Mulhearn
Clearly our assumption of cookie variety universality is incorrect. For instance, suppose the varieties you like are all made by the same person, and the most varieties that one person can make is two varieties. Or perhaps you like the most expensive varieties, and the most the manager will allow them to put out is two expensive varieties.

Jan 4 at 1:00 PM • Like • 6 • Reply

Robert M Roser
Talk to debbie. In cafeteria and make a request. Management problem not math

Jan 4 at 1:29 PM • Like • 5 • Reply

Someone great to work with has:

Someone great to work with has:

Curiosity

Someone great to work with has:

Curiosity

Critical thinking

Someone great to work with has:

Curiosity

Critical thinking

Clear communication

Someone great to work with has:

Curiosity

Critical thinking

Clear communication

Fast pace - Iterates quickly

Someone great to work with has:

Curiosity

Critical thinking

Clear communication

Fast pace - Iterates quickly

Technical skill - drawing conclusions from data

Someone great to work with has:

Curiosity

Critical thinking

Clear communication

Fast pace - Iterates quickly

Technical skill - drawing conclusions from data

Ability to translate a question into a program of research

Someone great to work with has:

Curiosity

Critical thinking

Clear communication

Fast pace - Iterates quickly

Technical skill - drawing conclusions from data

Ability to translate a question into a program of research

Domain experience - needed to work quickly

Someone great to work with has:

Curiosity

Critical thinking

Clear communication

Fast pace - Iterates quickly

Technical skill - drawing conclusions from data

Ability to translate a question into a program of research

Domain experience - needed to work quickly

Someone great to work with has:

Curiosity

Critical thinking

Clear communication

Develop through working on hard problems
- not homework -
higher stakes = better

Fast pace - Iterates quickly

Technical skill - drawing conclusions from data

Ability to translate a question into a program of research

Domain experience - needed to work quickly

HEP Alumni impact many fields

Tech: Facebook, Microsoft, LinkedIn, Intuit, Salesforce

Media/Entertainment: Netflix, Twitch, Spotify

Energy: Stem, Agentis

Health: Trace Genomics, 23andMe, Zenysis, FitBit

Consulting: Gartner, SVDS, Premise

Security: Vectra, Proofpoint, Amphor

Lending/Payments: LendUp, Prosper, LearnVest, Square

Build Connections between Science + Tech

Insight Blog + Notification List

Data Science Weekly

O’Reilly Data / AI / Solid

Fast Forward Labs Newsletter

Schemaless

Software Engineering Daily

Masters of Scale

Thank you!

kathy@insightdatascience.com
Kathy Copic: LinkedIn
@KathyScientist