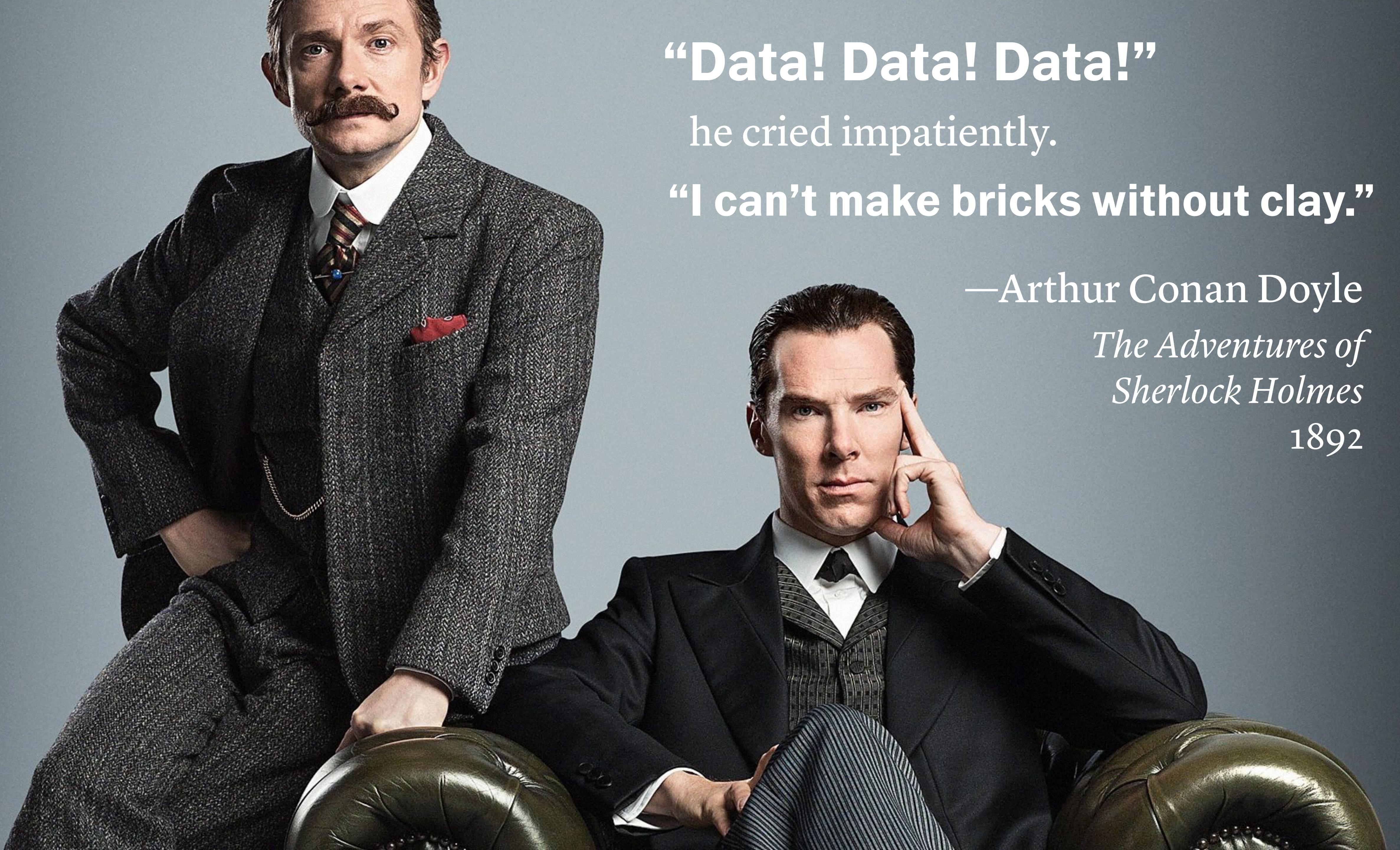


VSCO is an
ART & TECHNOLOGY company
empowering & inspiring people everywhere to
CREATE, DISCOVER & CONNECT

“Big data”

Data
Small data
Large data
Big data



“Data! Data! Data!”

he cried impatiently.

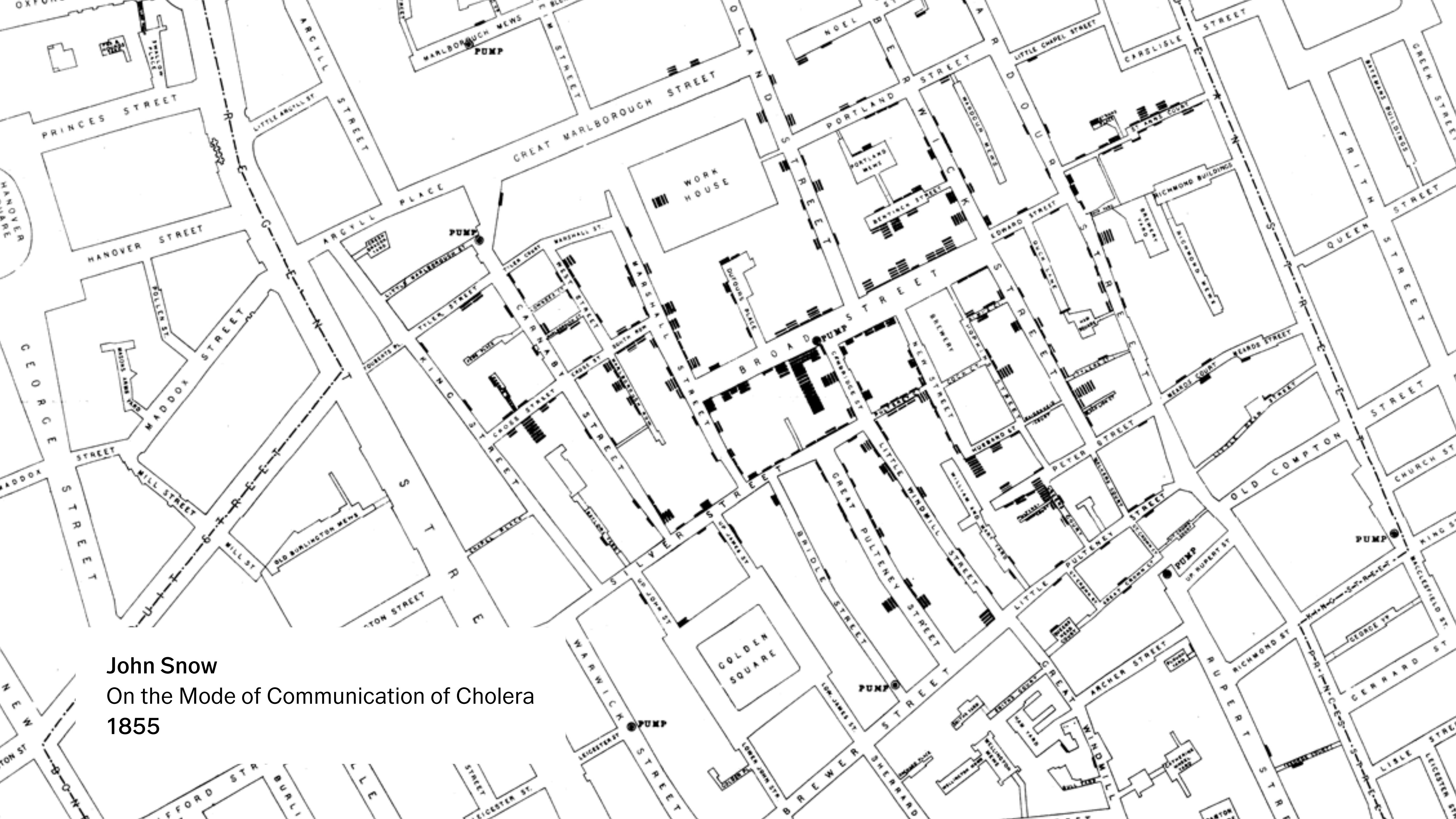
“I can’t make bricks without clay.”

—Arthur Conan Doyle

*The Adventures of
Sherlock Holmes*

1892

Data
Small data
Large data
Big data



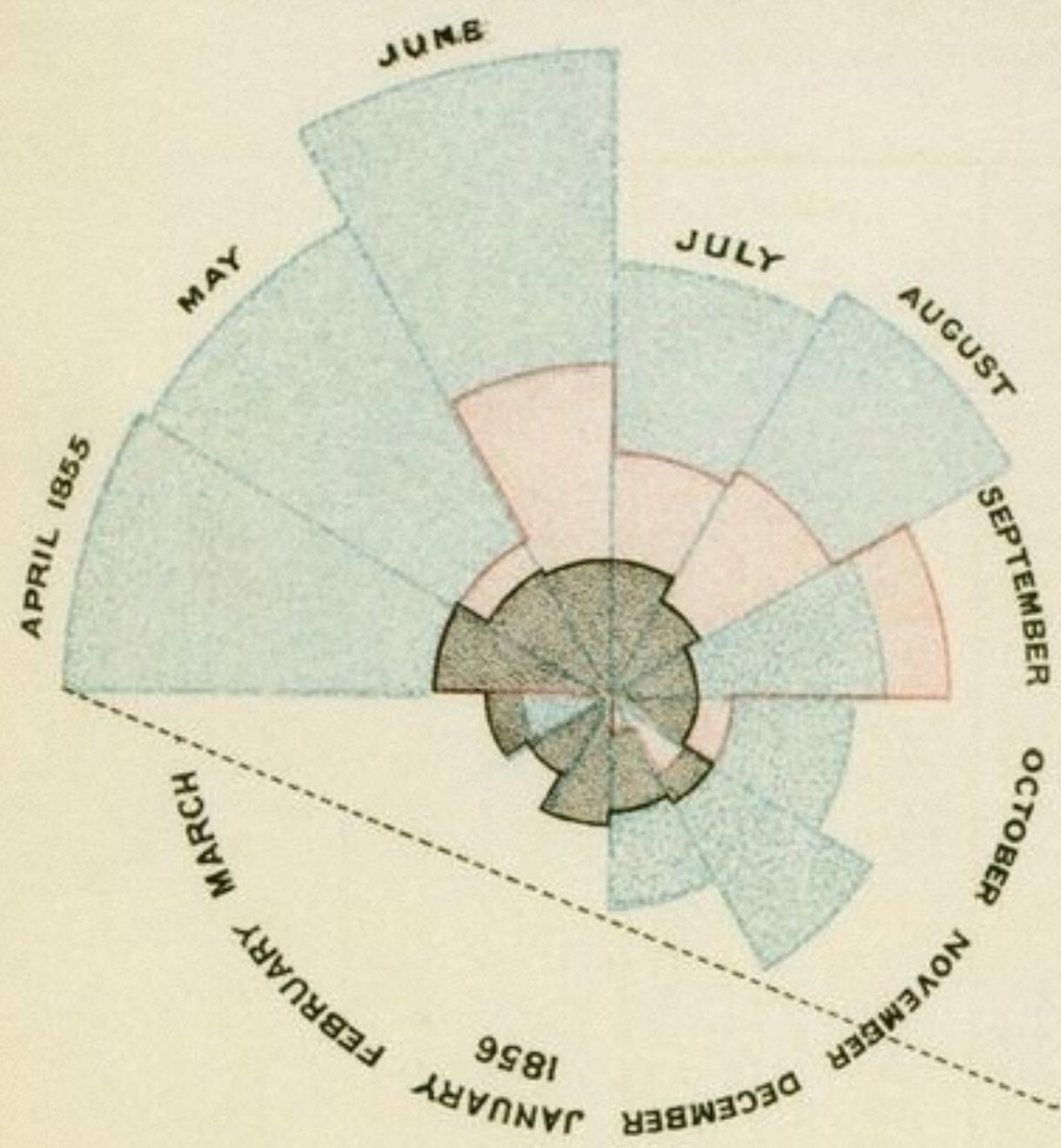
John Snow

On the Mode of Communication of Cholera

1855

DIAGRAM OF THE CAUSES OF MORTALITY
IN THE ARMY IN THE EAST.

APRIL 1855 to MARCH 1856.



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.

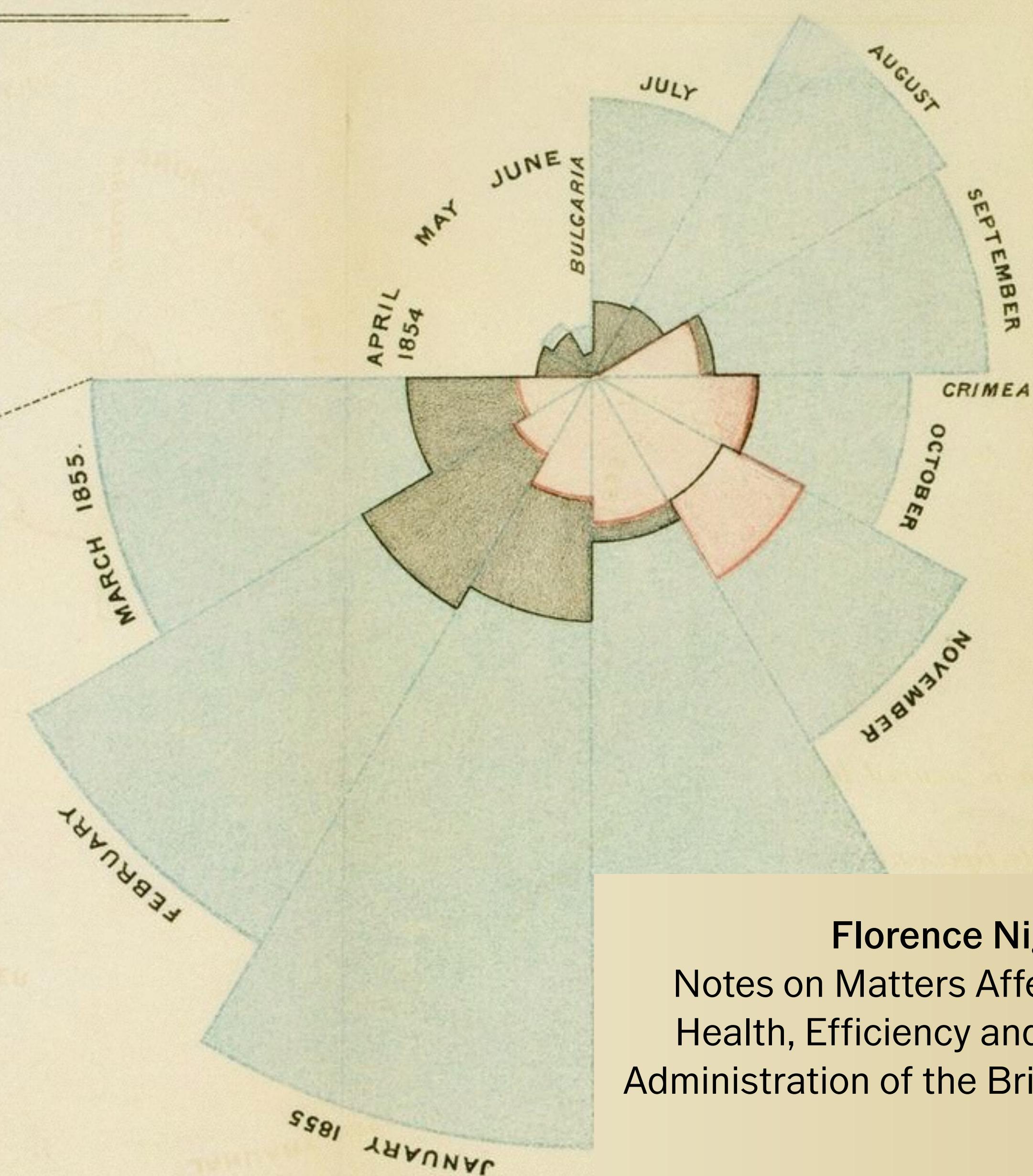
The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases; the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes.

The black line across the red triangle in Nov.^r 1854 marks the boundary of the deaths from all other causes during the month.

In October 1854, & April 1855, the black area coincides with the red;
in January & February 1855, the blue coincides with the black.

The entire areas may be compared by following the blue, the red & the black lines enclosing them.

APRIL 1854 TO MARCH 1855.



Florence Nightingale

Notes on Matters Affecting the Health, Efficiency and Hospital Administration of the British Army

1858

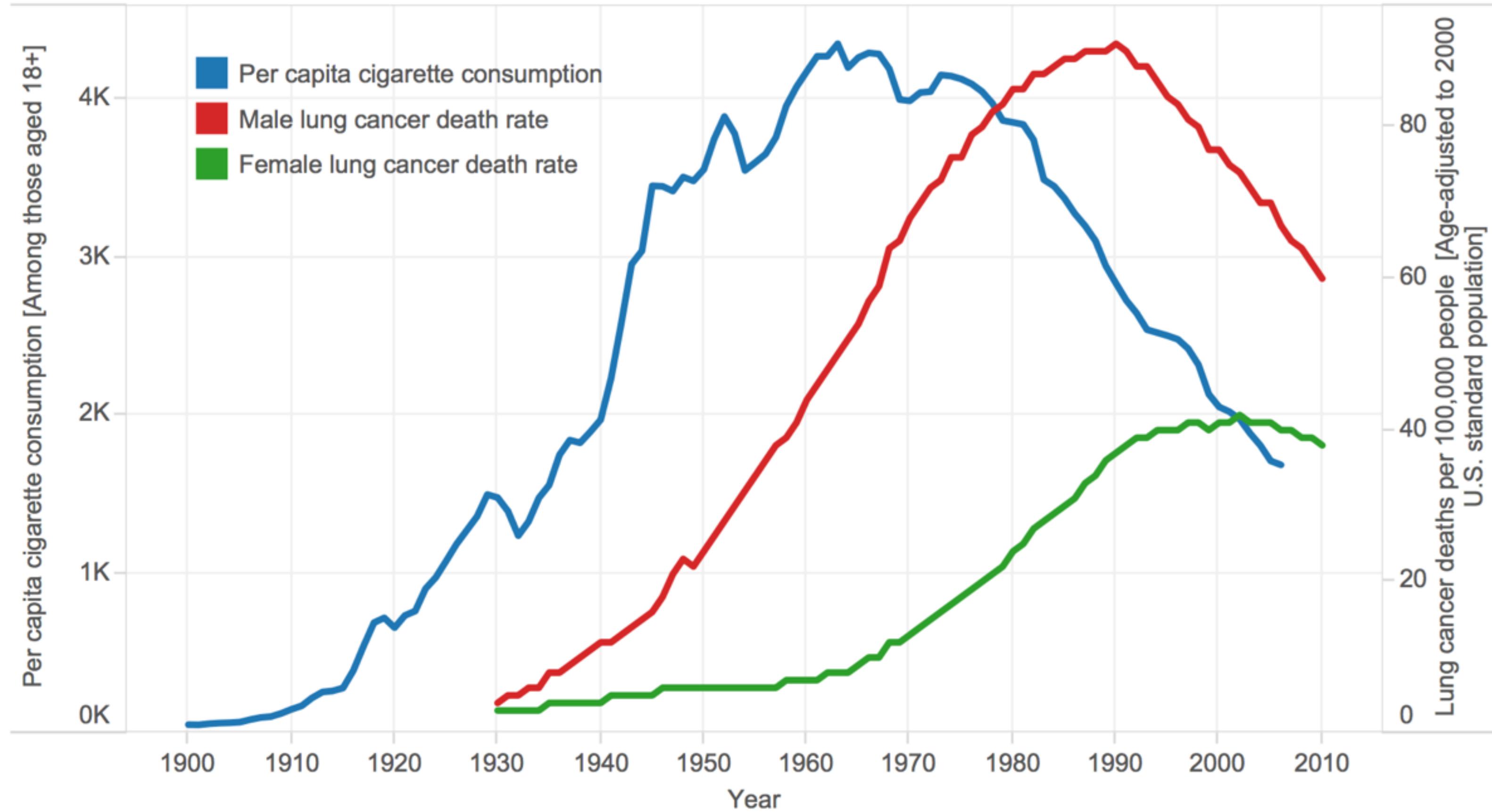
Data

Small data

Large data

Big data

Smoking causes lung cancer



Elizabeth Mendes
The Study That Helped Spur the U.S. Stop-Smoking Movement
2014





AP/Paul Sakura

Oakland Athletics' 20th consecutive win
2002

Data

Small data

Large data

Big data

insights
products

Data
Small data
Large data
Big data → insights

Sign Up

It's free and always will be.

Connect with friends and the world around you on Facebook.



See photos and updates from friends in News Feed.



Share what's new in your life on your Timeline.



Find more of what you're looking for with Facebook Search.

 First name Last name Mobile number or email Re-enter mobile number or email New password

Birthday

 Month Day Year[Why do I need to provide my birthday?](#) Female Male

By clicking Sign Up, you agree to our [Terms](#) and that you have read our [Data Policy](#), including our [Cookie Use](#).

[Sign Up](#)

[Create a Page for a celebrity, band or business.](#)

[English \(US\)](#) [Español](#) [Français \(France\)](#) [中文\(简体\)](#) [العربية](#) [Português \(Brasil\)](#) [Italiano](#) [한국어](#) [Deutsch](#) [हिन्दी](#) [日本語](#)

[Sign Up](#)[Log In](#)[Messenger](#)[Facebook Lite](#)[Mobile](#)[Find Friends](#)[Badges](#)[People](#)[Pages](#)[Places](#)[Games](#)[Locations](#)[Celebrities](#)[Groups](#)[About](#)[Create Ad](#)[Create Page](#)[Developers](#)[Careers](#)[Privacy](#)[Cookies](#)[Ad Choices](#) ▾[Terms](#)[Help](#)

Data

Small data

Large data

Big data → products

fire

[Top](#) | [Live](#) | [News](#) | [Photos](#) | [Videos](#) | [More options ▾](#)

Related searches

[fire san jose](#)[california fire](#)[fire singapore](#)[#bluecutfire](#)[fire la](#)Who to follow · [Refresh](#) · [View all](#)

Adobe Lightroom @Light... [Follow](#)

Lennox Bishop @LennoxBis... [Follow](#)

500px @500px [Follow](#)

[Find friends](#)

Top news

[View all](#)

Southern California Fire Flared Up at 5 Acres. A Day Later, It Covered 30,000.

The Blue Cut Fire “hit with an intensity that we haven’t seen before,” the San Bernardino County fire chief said.

The New York Times

Aug 16, 2016

Photos

[View all](#)

Related articles

[View all](#)

'It hit with an intensity we haven't seen before,' San Bernardino County fire...

Update on 'Live updates: Devastating Blue Cut fire in Cajon Pass consumes homes at...

Los Angeles Times

Aug 17, 2016

'Distracting' American Flags Ordered Removed from Town's Fire Trucks

A decision by the board of fire commissioners serving Poughkeepsie, New York sparked...

FoxNewsInsider

Aug 17, 2016

Big-data product: Amazon recommendations



(I just wanted a kitchen scale — now Amazon thinks I sell pot)



Beadaholique 100 Count



Dope Turtle - FOURTH



Zip Grinders - Large Spice



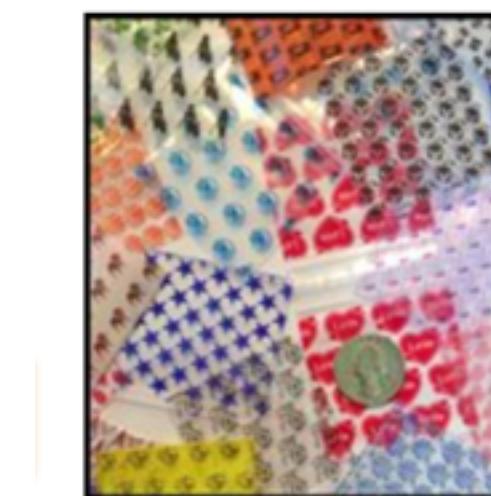
Raw Rolling Tray (Full)



Chromium Crusher 2.5 Inch



3 Boxes of Raw PRE-



2020 Apple Mini Ziplock



American Weigh Scales



Smart Weigh Calibration



1000 1.25" x 1.25" Zip



KingTop Herb Spice



1515 Apple Mini Ziplock



Space Case Grinder / Sifter
Large Titanium - 4 pc.
 136
\$79.00 ✓Prime



ZIG ZAG WRAPS 30
PACKS-15 DIFFERENT
FLAVORS- 2 PACKS
EACH



American Weigh Scales
GEMINI-20 Portable
MilliGram Scale, 20 by
0.001 G



Formax420 Zinc Alloy Herb
Grinder Leaf Designed on
Top Part 50 mm 4 Piece
with Free Pollen Presser...



Chromium Crusher 1.6 Inch
4 Piece Tobacco Spice
Herb Grinder - Black
 558



Beadaholique 500 Count
Resealable Zipper Poly
Bags, 2 by 3-Inch, 50mm
by 100mm, Clear

Big-data product: everything at Google

The image displays a map of San Francisco and its surrounding areas, specifically highlighting KFC locations. The map shows several KFC outlets, each marked with a red icon of a fork and knife. Labels for these locations include "KFC Fast-food fried chicken chain" and "Boston Market Counter-serve chain for chicken & more". The map also features major roads like I-80, I-580, and I-880, and neighborhoods like WEST OAKLAND, CHINATOWN, and TEMESCAL.

Below the map is a horizontal row of colored buttons used for filtering search results:

- pink
- blue
- white
- red
- black
- brown
- black and white
- wild
- caged
- happy
- silky
- drunk
- sultan
- ninja

To the right of the map is a sidebar titled "See results about" which provides information on two topics:

Chicken meat (Poultry)
Chicken is the most common type of poultry in the world. In developed countries, chickens are usually ...

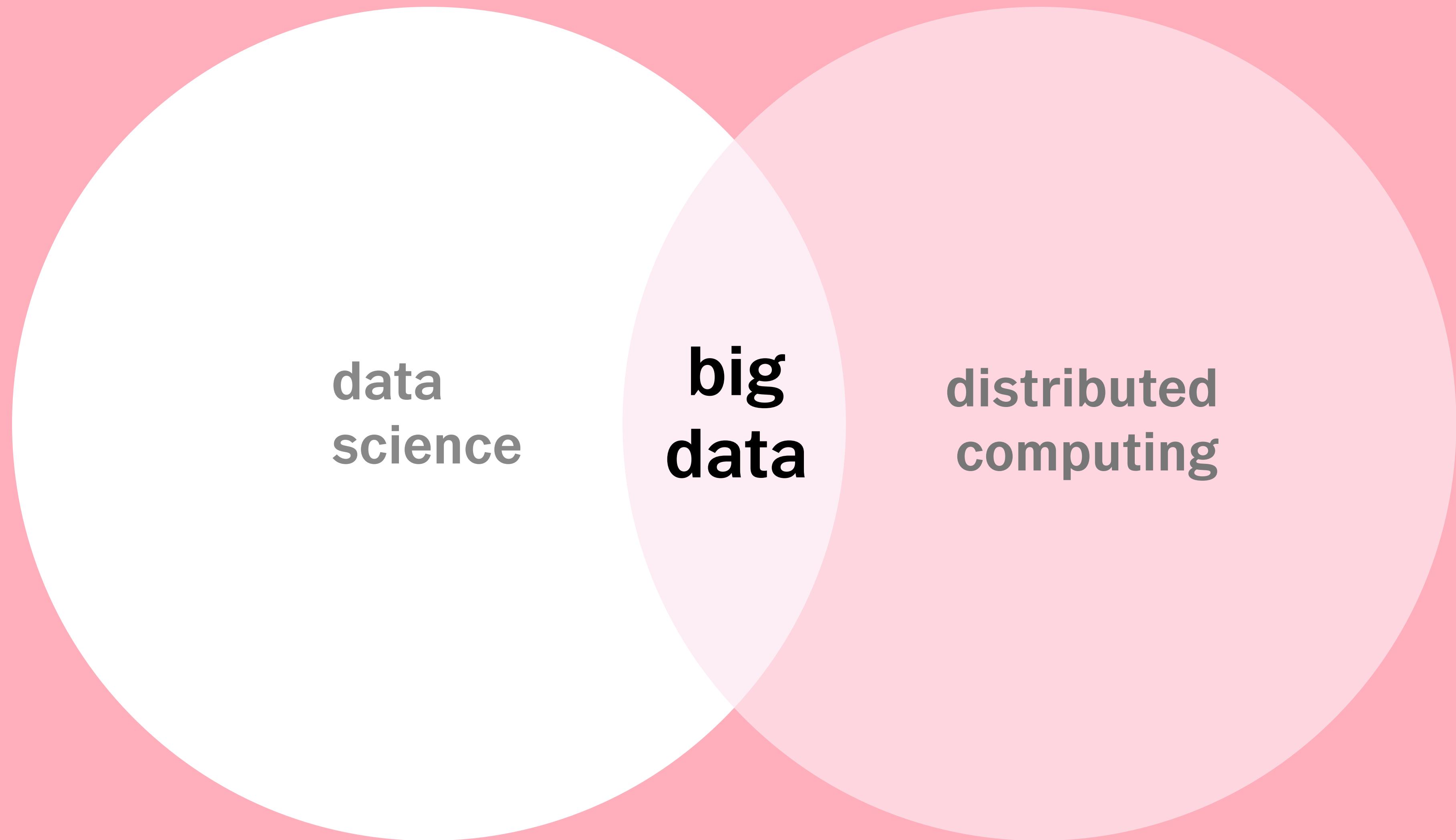


Chicken (Bird)
Breeds: Plymouth Rock chicken, Orpington chicken, Silkie
Scientific name: Gallus gallus domesticus



Big data

→ what is big data?



Parts of a system

databases/distributed filesystem
distributed computing
modeling/machine learning
visualization
data integration
services/model deployment

databases/distributed filesystem

HDFS

Red Hat Gluster Storage

Ceph

Amazon S3

databases/distributed filesystem

relational

MySQL

PostgreSQL

Oracle

document

MongoDB

RethinkDB

Espresso

Apollo

columnar

Cassandra

BigTable

key-value

Redis

Riak

graph

Titan

Neo4j

databases/distributed filesystem
distributed computing
modeling/machine learning
visualization
data integration
services/model deployment

distributed computing

general-purpose

Hadoop MapReduce

Spark

Flink

Storm

graph analytics

Spark GraphX

Elastic

NetworkX

Titan

approximate queries

Spark SQL

Algebird

Hive

Presto

deep learning

TensorFlow

Torch

Theano

Caffe

non-deep learning

Spark MLLib

NLTK

databases/distributed filesystem
distributed computing
modeling/machine learning
visualization
data integration
services/model deployment

modeling/machine learning

mostly
statistics
applied math

modeling/machine learning

 **Pinboard**
@Pinboard

Machine learning is like a deep-fat fryer. First time you try it you think “Amazing, I bet this will work on anything!” And it kind of does

RETWEETS LIKES

255 **367**



9:16 PM - 6 Jul 2016

  255  367 

modeling/machine learning

regression

linear
logistic
stepwise
ordinary-least-squares

decision tree

chi-squared AID
conditional DTs
decision stump

clustering

k -means
 k -medians
hierarchical

dimensionality reduction

principal component analysis
principal component regression
multidimensional scaling
linear discriminant analysis
mixture discriminant analysis

deep learning/ neural networks

convolutional neural networks
recurrent neural networks
LSTMs
stacked auto-encoders
deep belief networks

databases/distributed filesystem
distributed computing
modeling/machine learning
visualization
data integration
services/model deployment

visualization

“The simple graph has brought more information to the data analyst’s mind than any other device.”

— John Tukey

ggplot2 (R)

D3 (JavaScript)

Tableau

matplotlib (Python)

Zeppelin

IPython/Jupyter

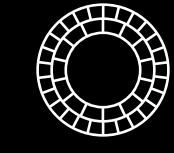
databases/distributed filesystem
distributed computing
modeling/machine learning
visualization
data integration
services/model deployment

databases/distributed filesystem
distributed computing
modeling/machine learning
visualization
data integration
services/model deployment

Don't forget

The point is to solve a problem

- Not to jam together as much cool technology as possible
- Often simpler is better



17 AUG 2016
MELINDA LU

→ MELINDA@VSCO.CO
→ @SKEWLEFT

Please visit

→ github.com/melindalu/wwcode-august2016

for the next part of this evening.