

A panoramic view of the Singapore skyline at dusk, featuring the Marina Bay Sands hotel, the Esplanade - Theatres on the Bay, and the ArtScience Museum. The water of Marina Bay is in the foreground, and the sky is a mix of blue and orange.

# Smart Data and Food Innovation

## Potential impact on Public Health Solutions

**Ralph Graichen, Director Food and Nutrition**  
**A\*STAR, Singapore**

# Big Data

**5 billion** mobile phones in use in 2010

**30 billion** pieces of content shared on Facebook every month

**40%** projected growth in global data generated per year

**\$300 billion** potential value to US health care

**€ 250 billion** potential value to Europe's public sector administration

**60%** potential increase in retailer's operating margins

**1.5 million** more data-savvy managers needed in the US

Creating an Innovation Economy

# Big Data to Smart Data

- Detection and exploitation of patterns
- Machine learning = computers develop algorithms and predictive models
- But humans filter still for:
  - statistical significance
  - anomaly detection
  - perform discovery analytics
- Data quality and completeness
- Misleading data
  
- Historically:  
80% data collection, 20% analyzing  
now reverse
- **90% of all data has been generated in the last 2 years**

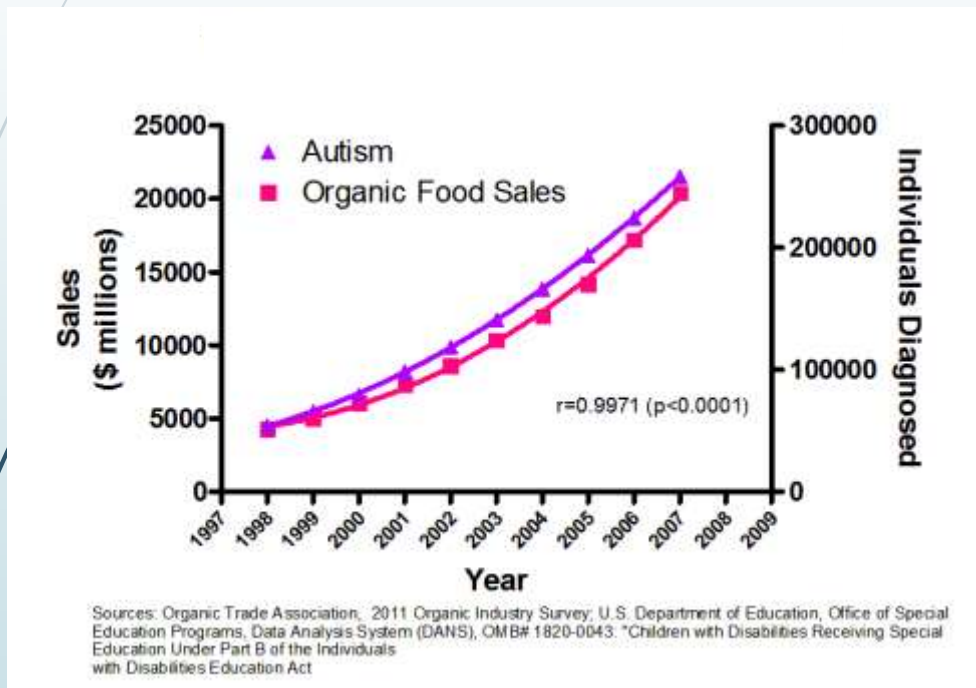


# Big Data to Smart Data



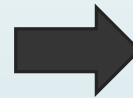
- ▶ **Statistical Correlation:** any statistical relationships involving dependence or how close two variables are having a linear relationship
- ▶ Correlation **does not** imply causation

# Big Data to Smart Data



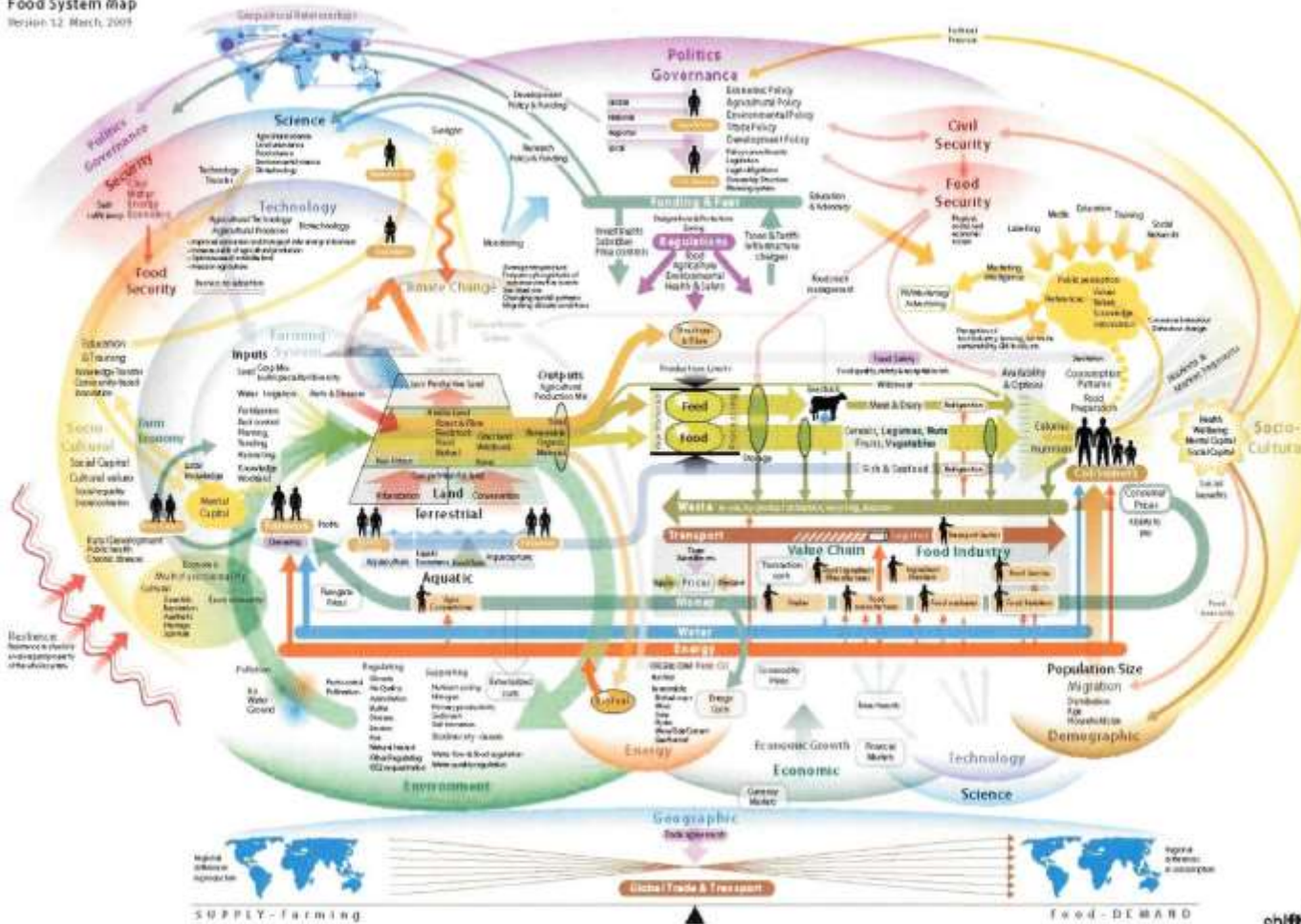
- **Statistical Correlation:** any statistical relationships involving dependence or how close two variables are having a linear relationship
- Correlation **does not** imply causation

# The Global Food Value Chain



# The Global Food Value Chain

Food System Map  
Version 1.2 March, 2009



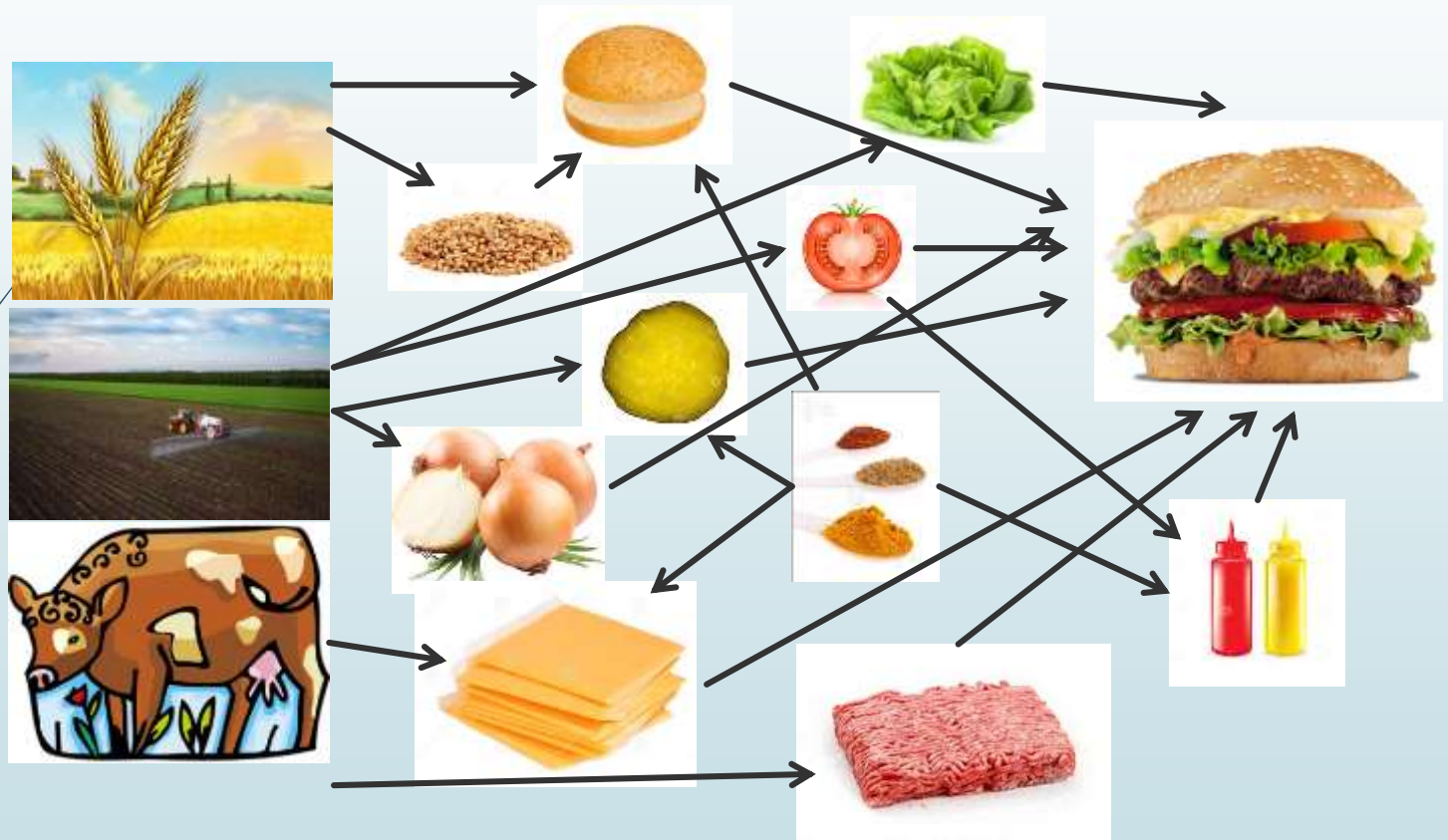
# Globalizing of Food: The Cheeseburger



Creating an Innovation Economy

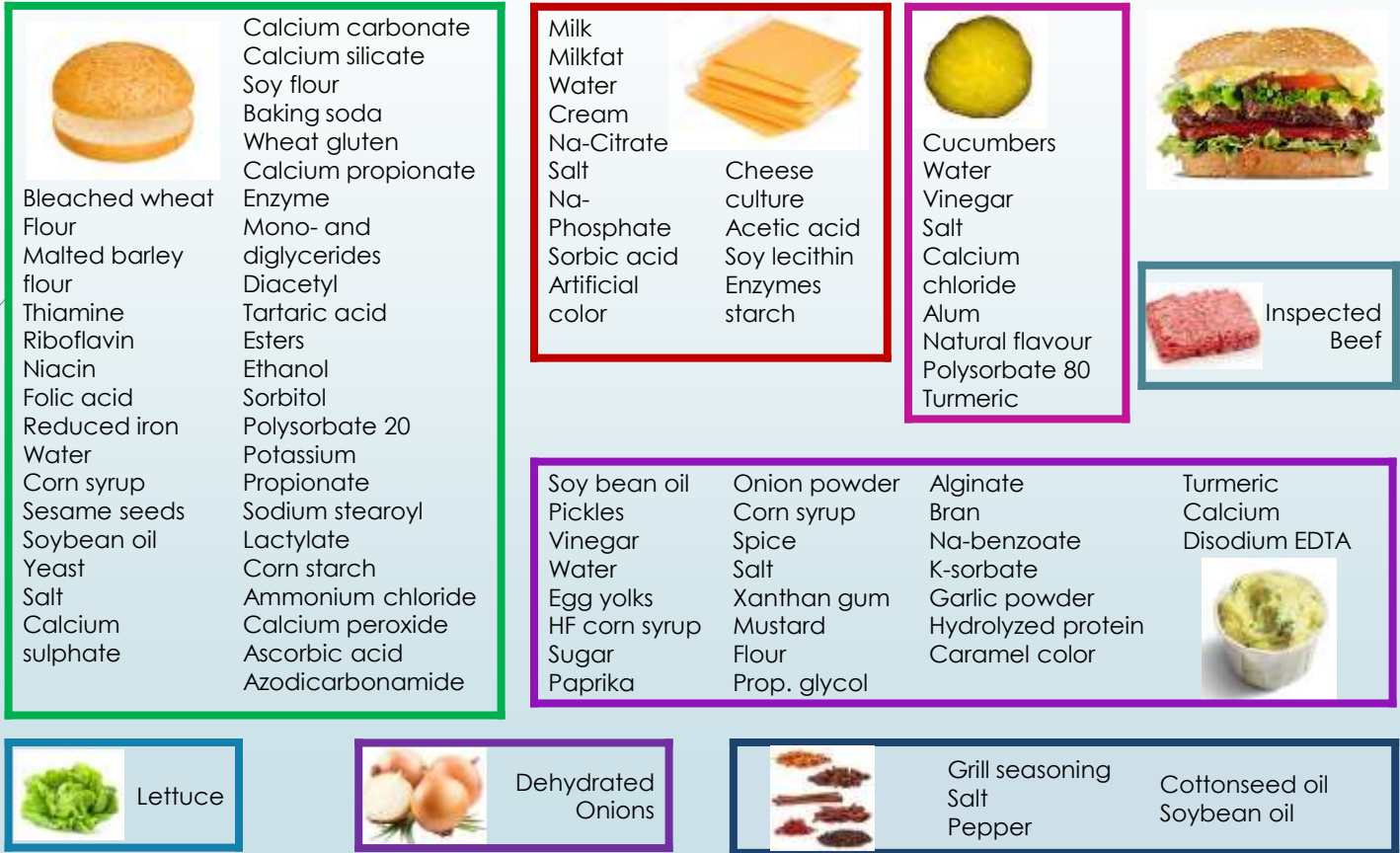


# Globalizing of Food: The Cheeseburger



# Globalizing the Cheeseburger

## The Components of a Cheeseburger



\* National Centre for Food Protection and Defense

# Globalizing the Cheeseburger

**Tartar Sauce**



**Garlic Powder**

**Tomatoes**

**Beef**

**Wheat Gluten**



Argentina	Japan	Brazil	Belgium	Australia	Australia
Australia	S. Korea	Canada	Canada	Canada	Belgium
Austria	Lebanon	China	Colombia	Chile	Canada
Belgium	Peru	Germany	Costa Rica	Costa Rica	China
Brazil	Poland	India	Dom. Rep	Japan	Chile
Canada	Portugal	Israel	Guatemala	Mexico	Czech Rep.
China	Serbia	Japan	Israel	Nicaragua	France
Chile	Philippines	S. Korea	Morocco	New Zealand	Germany
Colombia	Russia	Mexico	Mexico	Uruguay	Kazakhstan
Denmark	S. Africa		Netherlands		Lithuania
Dom. Rep	Singapore		New Zealand		Netherlands
France	Spain		Poland		Poland
Germany	Sweden		Spain		Russia
Greece	Turkey				Switzerland
Hong Kor	Taiwan				Thailand
Israel	U.K.				U.K.



\* National Centre for Food Protection and Defense

Creating an Innovation Economy

# Big Data



+



**Accurate and Meaningful**

4 Key Aspects:

- ▶ Volume of data
- ▶ Speed of data generation
- ▶ Aggregation of distinctly different data types
- ▶ Validity and security of data

# Creating Value

## Transparency:

- Data Integration
- Reducing searching and processing time
- Identify root causes
- Distribution and warehousing

## Supporting Human Decision making:

- Customer segmentation
- Risk management
- Improved quality

## Innovation:

- New products and services
- Customer segmentation - Marketing and after-sales
- Product cycles



# Supply Chain Network Clustering



## Motivation

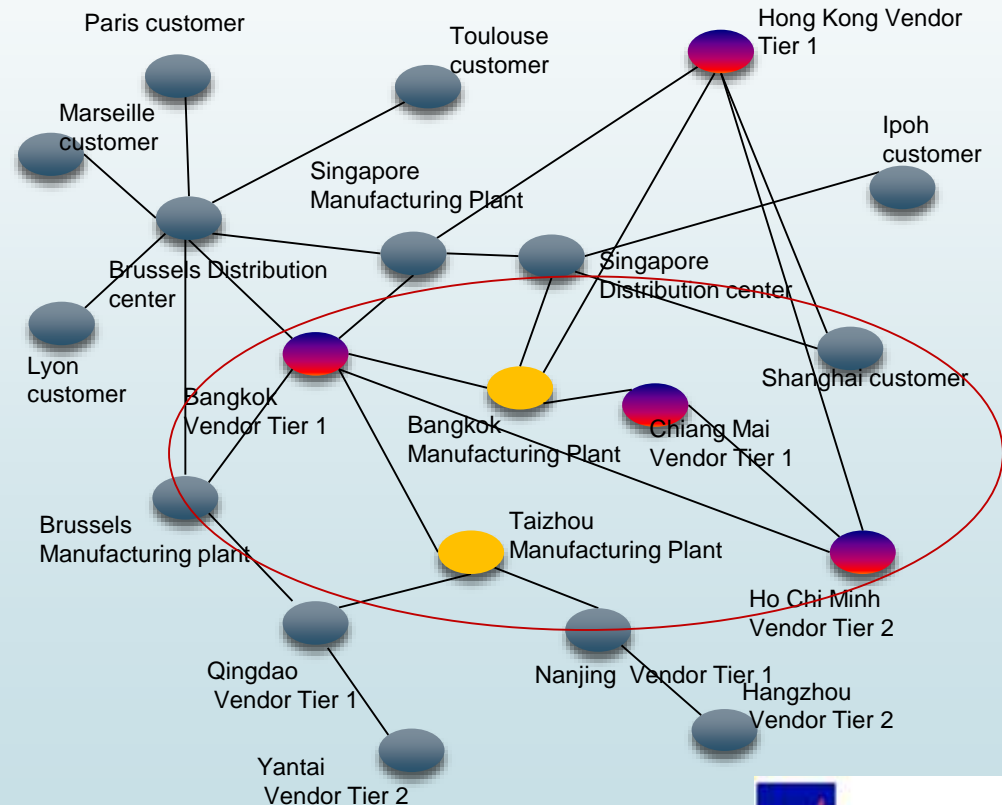
- Identify topological weak nodes considering internal and external factors and network properties, such as product flow and volume, and location based disruption (e.g. political and financial stability, natural disasters)
- Examine key groups

## Approach

- Machine learning approaches
- Clustering algorithms
- Network analytics

## Achievement & Impact

- Detection of at-risk clustering hidden
- Evaluation of robustness of supply chain networks



# A Glimpse into the Future

- ▶ Using “machine learning” on data sets **understanding** consumer **recommending** products **personalise** websites
- ▶ **Otto Group**, a German e-commerce merchant
- ▶ Predict customers purchases **a week before order**
- ▶ Analysing ~ 3bn past transactions  
200 variables
- ▶ 90% accuracy for a 30 day forecast
- ▶ AI purchasing 200,000 items a month  
no human intervention



# From Data to Insights



## Amazon

- **152 million customer** accounts
- Building recommender systems to improve customer relationship using customer click-stream data and historical purchase data
- Constantly monitor, track and secure **1.5 billion items** in retail over **200 fulfilment centres**
- Simple Storage Service (S3)
- receives **≥ 50 million** updates a week and every 30 minutes all data received are crunched and reported back
  
- **Google** and **Facebook** = data about consumers,  
**Amazon** manages customers directly



# Anomaly Detection

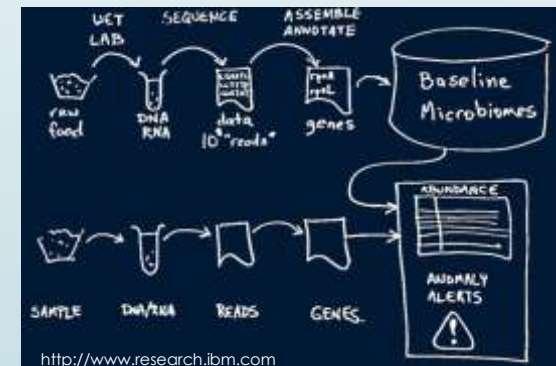


- Primary Data versus Smart Data
- Nestle = **100 M** Analytical Data per year
- Lack of mining and mapping against metadata (e.g. geographical information)
  
- Securing the supply chain
- Impact on food quality/safety
- Stakeholder access – sharing/access
  
- Companies/Process Engineers need to look at big picture **not** departmentally

# Consortium for Sequencing The Food Supply Chain



- Lead by IBM Research and Mars, Inc
- DNA and RNA sequencing, profiling microbiome in the food supply chain
- Data: Genomic and microbiome of ingredient samples combined with **contextual data** (weather, shipping)
- Creating baseline of safe ingredient microbiome
- Comparison with new data:
  - Detection of new genes
  - Gene variants
  - Undetected anomalies
- Ultimately to cover plants, livestock, bacteria, any other organism **from origin to consumer**



# Anomaly Detection



- ▶ Whole Genome Sequencing
- ▶ General impact on food safety
  - ❑ FDA: identifying causative pathogens causing foodborne illness
  - ❑ Industry: quality control
  - ❑ Consumer: traceability
- ▶ Data upload to GenomeTrakr
- ▶ Pairing foodborne pathogen's with geographic location

Creating an Innovation Economy

# Food and Crime



## Infamous **Top Ten**

- I. Olive Oil
- II. Honey
- III. Fish
- IV. Scallops
- V. Balsamic Vinegar
- VI. Saffron
- VII. Vanilla
- VIII. Coffee
- IX. Cinnamon
- X. Black Pepper

- Economic Cost: **US\$30 to \$40 billion** every year
- Drivers are **economic gains** but can cause serious **public health risks**
- 1981: Toxic Oil Syndrome
- 2007: Melamine in animal feed
- 2008: Chinese Milk scandal

# BlockChain and The Supply Chain

## Blockchain:

- Universal transaction processing tool
- Provides proof of ownership at any given moment
- Distributed consensus mechanism
- No single ownership/control
  
- Information has to be verified
- Scale up issues/energy efficiency
- Institutional resistance

Walmart – Origin of Mangoes

- Currently = **6d, 18hrs, 26min**
- Blockchain = **2.2 sec**

IBM partners with Nestle, Unilever and other food giants to trace food contamination with blockchain

- IBM has announced a blockchain collaboration with food giants including Nestle, Unilever and Walmart.
- The corporation said blockchain would enable food businesses to trace the source of contaminated produce in mere seconds.
- Blockchain maintains a digital ledger of transactions rather than a physical one.

Ryan Browne | @Ryan\_Browne\_ |  
Published 5:00 AM ET Tue, 22 Aug 2017 | Updated 2:13 AM ET Wed, 4 Oct 2017

CNBC



Creating an Innovation Economy

# From Data to Insights

- ▶ The world's food supply depends on about **150 plant** species.



- ▶ Just **12** provide **three-quarters** of the world's food.



- ▶ There are **over 300,000 plant** species.
- ▶ Providing over **18 billion plant proteins**,
- ▶ **108 million lipids**,
- ▶ and **4 million polysaccharides**.

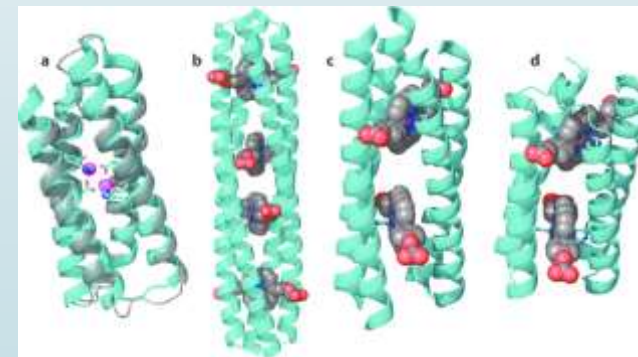
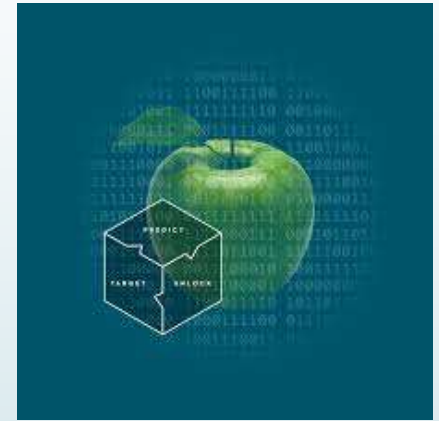
# From Data to Insights

- ▶ Data Mining, computational analysis to explore proteins, for **functions** and **nutrient density**
- ▶ Screening for molecular properties like
  - *protein yield,*
  - *thermal stability,*
  - *physical stability*
  - *physiological function*



# From Data to Insights

- **Nuritas:** combining artificial intelligence and genomics to discover Bioactive Peptides with health benefits
- **Amai Proteins:** computer-aided engineering for partial or full de novo designs of proteins of interest  
e.g. improving properties of taste-enhancing proteins





# From Insights to Products

- ▶ **Just / Hampton Creek:** raised more than **\$239 million** in funding data analytics to function and characteristics
- ▶ **Impossible Foods:** plant based burger investigating the molecular basis of food flavors and textures



# Moving Forward

## Big Data:

- Hard to maintain or incompatible databases
- High transaction costs
- Data Security

## BlockChain/Bitcoin

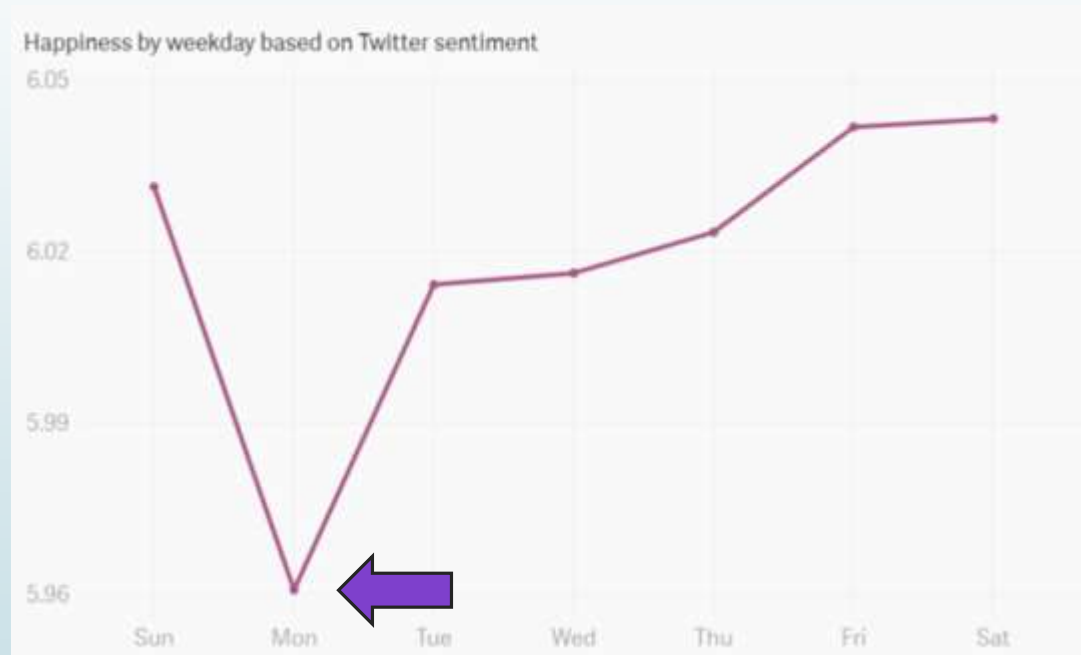
- 304,000 transactions a day
- 9,000,000 trillion hashes /second
- 20.03 terawatt hours per year (2017)



## Internet of Things:

- $\geq 25$  billion connected devices in 2020
- Security and trust
- Scaling transaction processing needs

# Big Data and Happiness





# Thank You

Creating an Innovation Economy

Creating an Innovation Economy