



# PROFITS OF THE INTERNET

DATA ANALYTICS IN ONLINE ADVERTISING

By Daniel Gale

2003

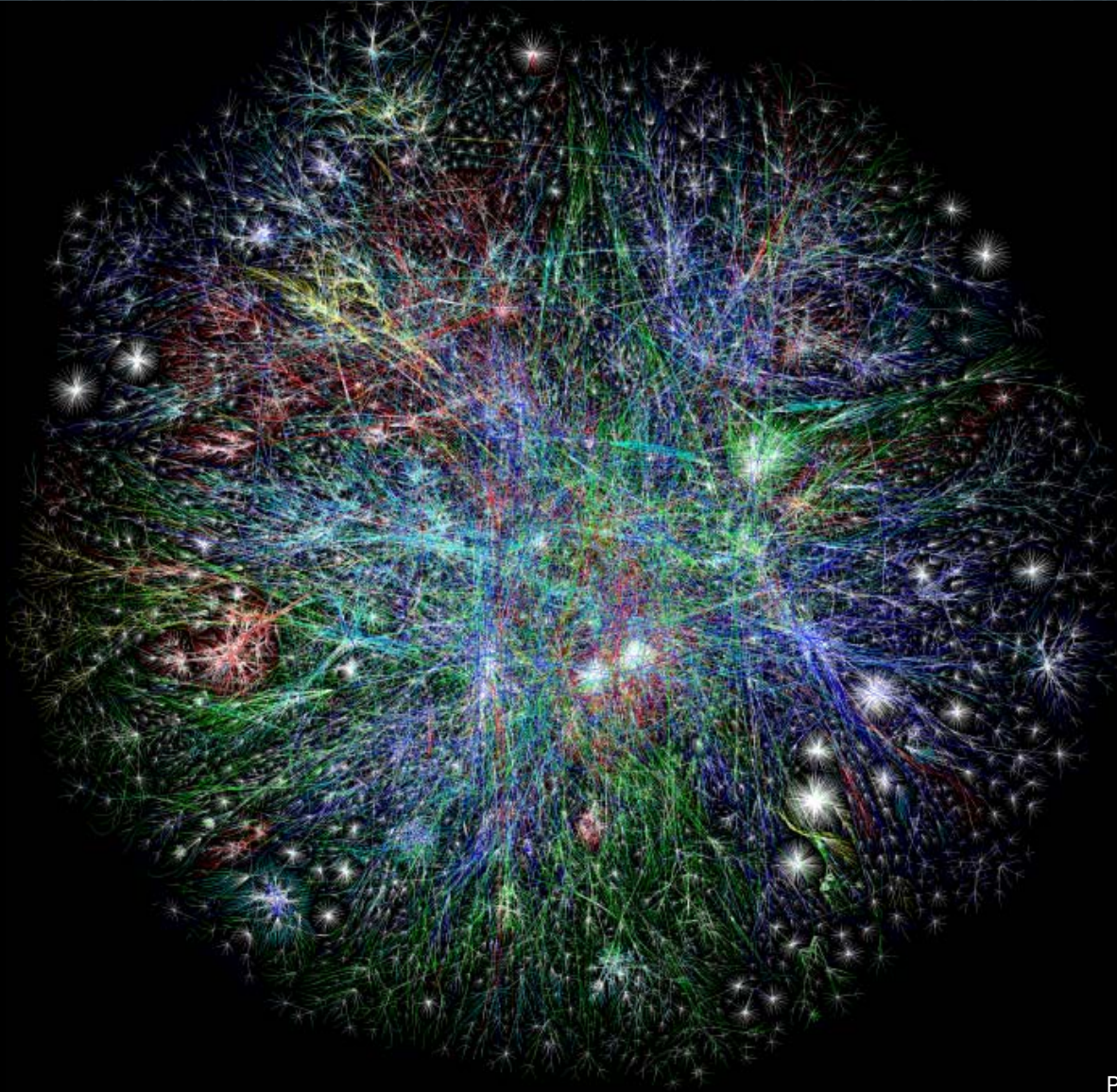


Photo by: Barrett Lyon / The Opte Project





**First Website 1991**

**2017**

**Last Restriction 1995**

**Google Founded 1998**

**97% of Telecom 2007**



12B

# WHAT IS PROGRAMMATIC ADVERTISING?





**MOBILE**



**VIDEO**



**DISPLAY**



**NATIVE**





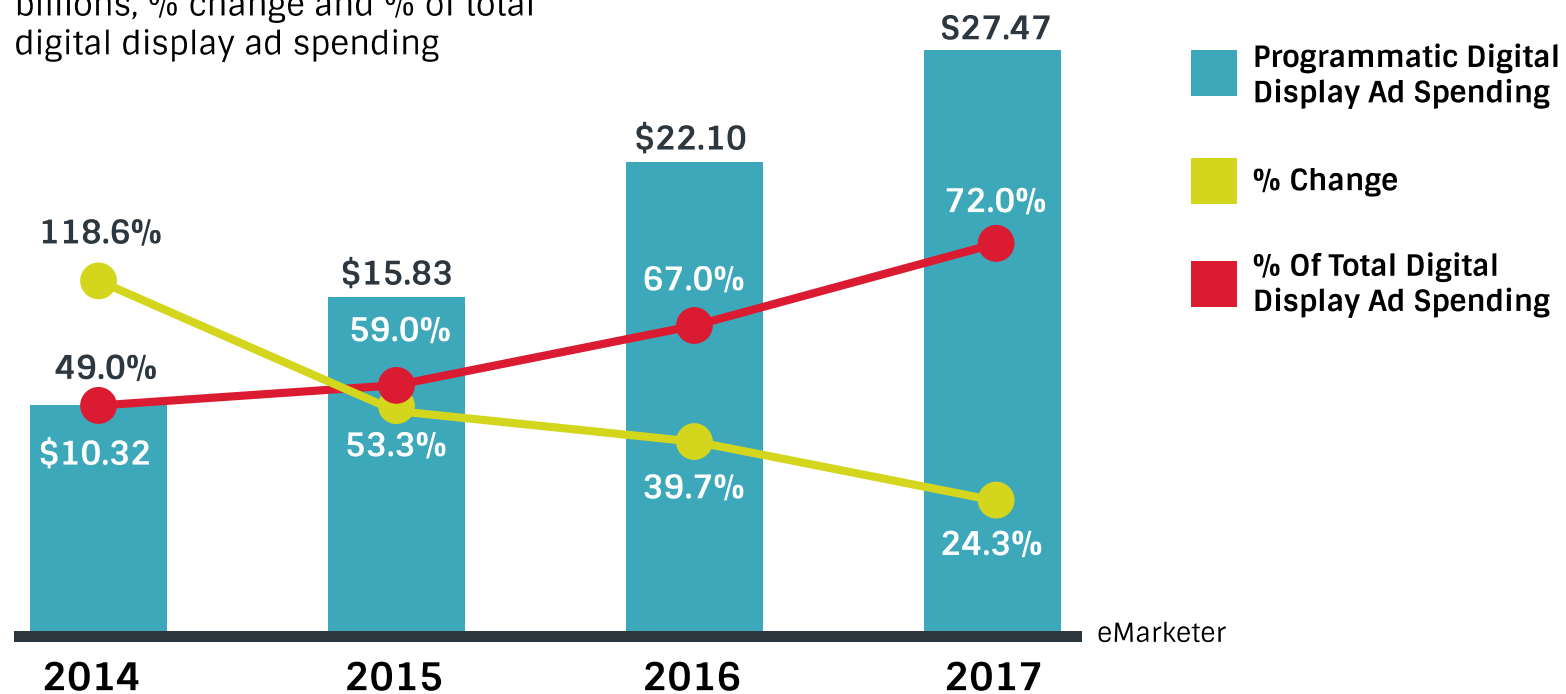




# PROGRAMMATIC TAKING OVER THE ADVERTISING WORLD

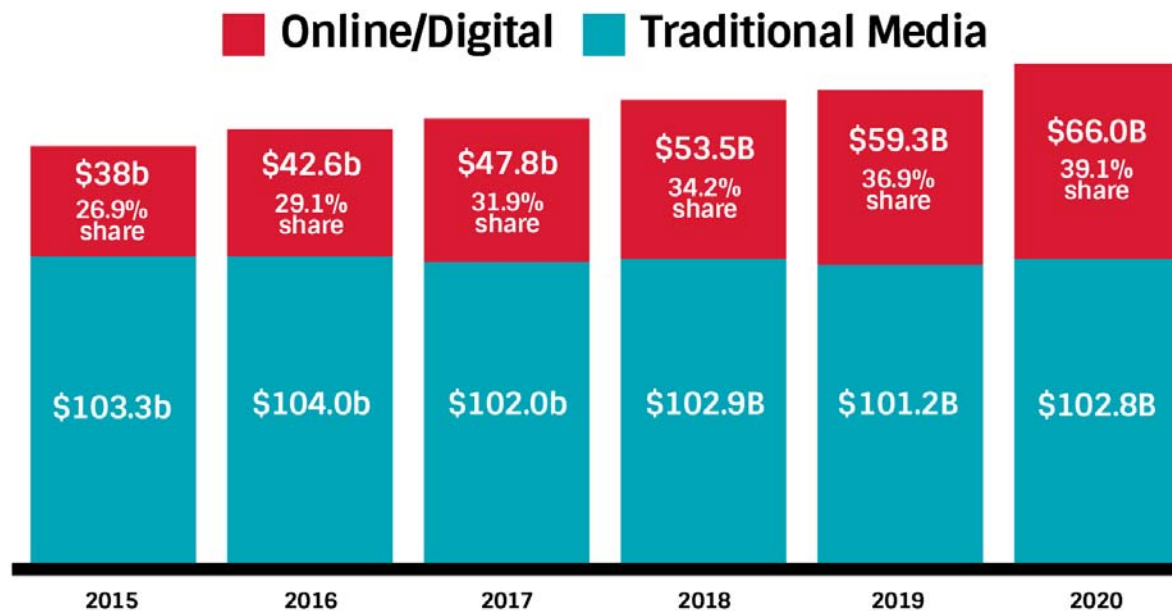
## US PROGRAMMATIC DIGITAL DISPLAY AD SPENDING, 2014-2017

billions, % change and % of total digital display ad spending



# LOCAL DIGITAL GROWTH TRENDS

Local Media Advertising Spend, Traditional vs. Digital Media: 2015 - 2020 (\$billions)



Source: Marketing Charts.com | Data Source: BIA/Kelsey

## LOCAL DIGITAL

2016

**\$42 Billion**

2020

**\$66 Billion**

# 250B

Data collection events daily



# 25B

Unique keywords-user pairs collected daily



# 550M

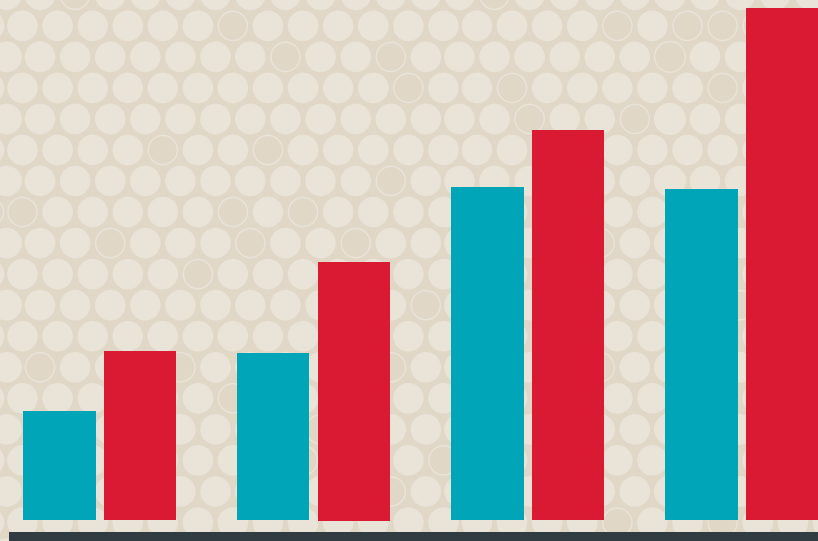
Unique GPS events daily



# 70k

Active advertising campaigns

# GETTING TO THE DATA



# SIMPLIFI DATA INFRASTRUCTURE: THE BIG THREE

## KAFKA

### Log storage and simple queue

- No local storage for most services

## VERTICA

### Valuable and Filtered Data

- Driving model building and internal analytics

### Client-Facing Ad-Hoc Analytics

- Utilizing Looker to visualize data

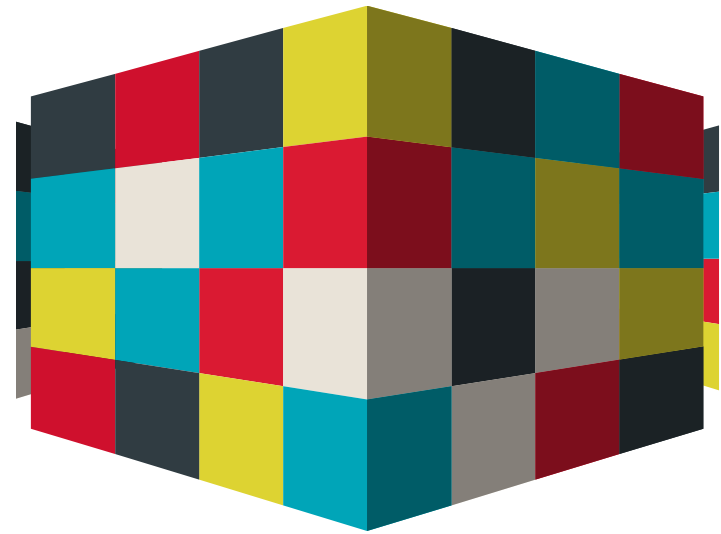
## HADOOP/SPARK

### Unknown Value of Data

- All data goes into Hadoop

### Offline Analytics

- Utilizing Spark to analyze and restructure data





# DATA SCIENCE IN PRODUCTION

## PERFORMANCE MODELS

Utilized at bid-time to predict events

- 50k+ active models

## KEYWORD RECOMMENDATIONS

### K-Means + PCA

- Clustering keywords to discover similar keywords

### Tau Stats

- Measure the overlap between keywords

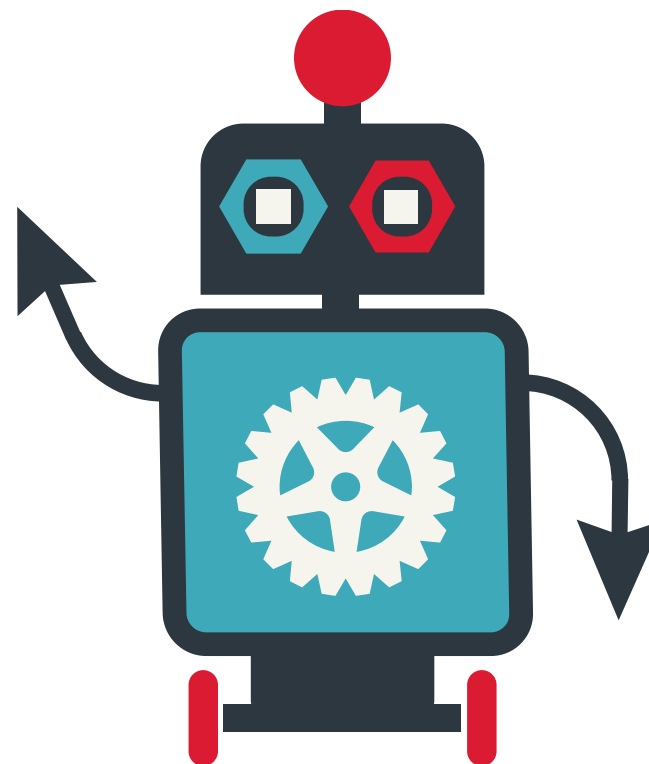
## CAMPAIGN ANALYTICS

### Audience Insights

- Demographic data using 3<sup>rd</sup> party segments

### Geo Conversion Lift

- Difference between campaign and natural population



# DATA SCIENCE IN DEVELOPMENT

## DATA VALUATION AND FILTERING

Prediction models for whether data is used

- Largest dataset so far

---

## LOCATION DATA CATEGORIZATION

Filtering bad data

- Stop targeting fraudulent lat/long data

Categorize remaining data

- Categorize GPS data for our various use cases

---

## AD HOC ESTIMATES

Campaign Estimates

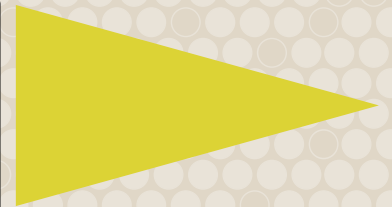
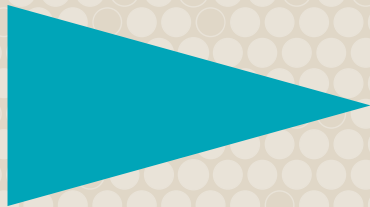
- Determine how many users we can target

Smarter Caching

- Better estimates allow for more intelligent caching



# LESSONS LEARNED



**ONE:**

Start simple.

ABC



**TWO:**

Do not be afraid to  
pay for software.



## THREE:

Understand the  
value of your data.



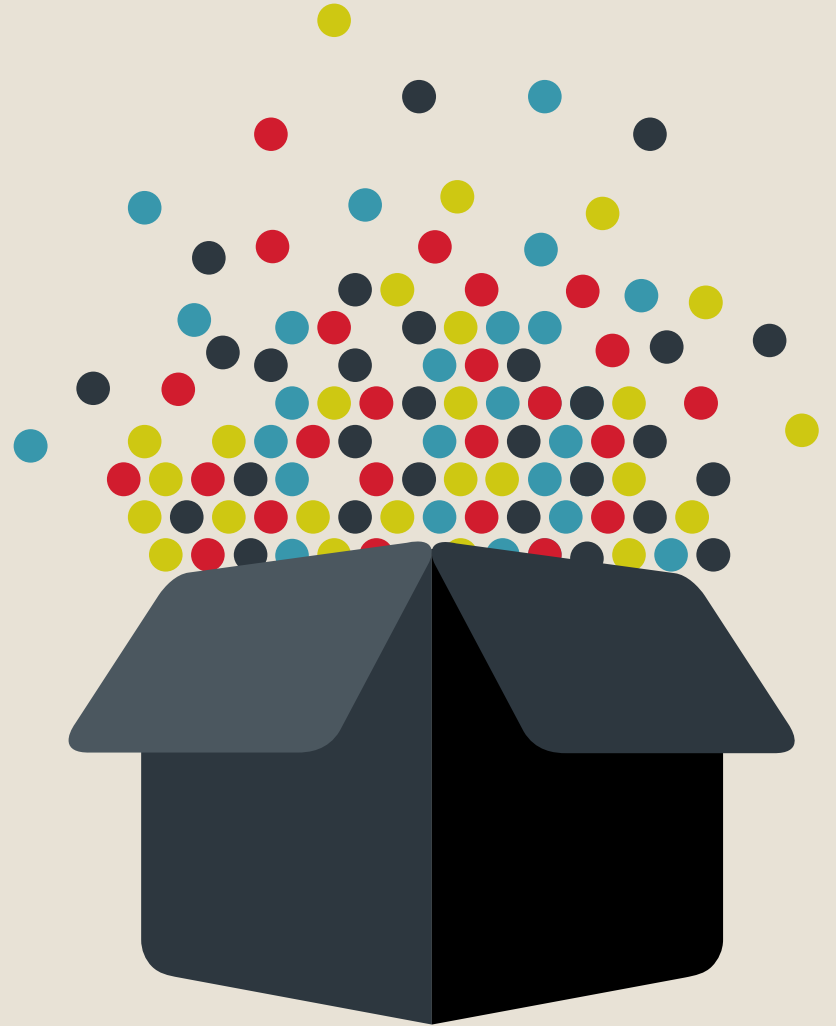
# FOUR:

Learn where data science fits in your data pipeline.



**FIVE:**

More analytics  
leads to more data.





Q&A



# THANK YOU

Simplifi