DOING DATA SCIENCE IN SERVICE OF THE PUBLIC GOOD

SALLIE ANN KELLER
Division Director, Social and Decision Analytics
Distinguished Professor in Biocomplexity
Professor of Public Health Sciences

Datapalooza | November 1, 2019
Biocomplexity Institute

Team science exploring the behavior of massively interacting living systems to develop practical solutions to real-world problems.
Data science starts with a problem.
Without a problem, there’s no data science.
Data Science is a Team Sport

Thank you to my team!
Quiz

What came first: data or computers?
Answer

✅ **COMPUTER** *(circa 1613)* … A person who performs calculations or computations

**DATA** *(circa 1640)* ……….. Any fact assumed to be a matter of direct observation
DOING DATA SCIENCE
Where the journey begins
DOING DATA SCIENCE

Game Changers
Out of the Office into the Community
Fire/EMS Data Systems

Flow of Events:
- Incident Occurs
- Call Arrives @ EOC
- Unit(s) Dispatched
- Unit(s) Arrive
- Incident Closed
- Post Incident

Internal Data:
- CAD Call Summary
- Unit Movement Summary
- Incident Summary

External Data:
- Social Media
- Media Reports
- NFIRS, Med Report
From Big Data to All Data
How Data Are Born

**Designed Data**
- Microscope
- Document
- Satellite dish

**Administrative Data**
- Computer
- Medical records
- Ambulance

**Opportunity Data**
- Network
- Smartwatch
- GPS

**Procedural Data**
- Traffic light
- Law book
Early Data Insights

Minutes on Call

- Fire Units
- Medical Units
- Other
Game Changer: Data are surrogates for life in the city
911 Data Characterizes Rhythm of the City

Sunday, 2:00

High Call Volume

Low Call Volume
Emergency Management
Seconds Matter

Response Time by Distance

Fire Stations and Unit Dispatches
Data Science Helps Disentangle
Citizen Safety: Operation Firesafe

Ensure at-risk citizens have smoke detectors.
Predictive Probabilities of Smoke Detectors

Census Block Level Predictions

Housing Unit Level Predictions
Game Changer:
New language breaks down organization silos
The New Language of Data
Parks, Recreation, Human Services, and Schools
Science of All Data Comes Together

Proctor & Gamble

Federal Statistical Agencies

U.S. Army

Local Government
Data Science Framework

PROBLEM IDENTIFICATION: THEORIES AND WORKING HYPOTHESES

DATA DISCOVERY [INVENTORY | SCREENING | ACQUISITION]

DATA INGESTION AND GOVERNANCE

DATA WRANGLING
  PROFILING • PREPARATION • LINKAGE • EXPLORATION

FITNESS-FOR-USE

STATISTICAL MODELING AND ANALYSES

COMMUNICATION AND DISSEMINATION

ETHICS REVIEW
Game Changer:
Community-based research
vs.
Research based in the community
Community-based Research

- Bring data in service of the public good
- Deepen partnership between communities and public and land-grant universities
- Enable communities to become data-driven learning communities
Engage community leaders to identify issues. Formulate questions.

Communicate and disseminate. Share results with everyone.
Key Community-Based Desires

- **LOCATING** and describing a population
- **ESTIMATING** a statistic and a measure of its variability to evaluate its usefulness
- **FORECASTING** future needs
- **EVALUATING** a program, policy, or standard operating procedure
Life doesn’t happen in census tracts.
CommunityScapes
Identifies where to target programs and policies

Supervisor Districts
High School Districts
Census Tracts

Obesogenic vulnerability
Low High
Game Changer: Workforce Development
<table>
<thead>
<tr>
<th>POLICE DEPARTMENT, ARLINGTON COUNTY, VIRGINIA</th>
<th>USDA</th>
<th>US DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT</th>
<th>VIRGINIA DEPARTMENT OF EMERGENCY MANAGEMENT</th>
<th>SCHEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>COUNTY OF FAIRFAX, VIRGINIA</td>
<td>NCSES National Center for Science and Engineering Statistics</td>
<td>Data Science for the Public Good Projects</td>
<td>ROANOKE</td>
<td>United States Census Bureau</td>
</tr>
</tbody>
</table>
Game Changer: Mobilize National Movement
Democratization of Data
Sparking a National Effort
Data Science Research Organization
Where the journey is now
The Journey Continues

Discussion