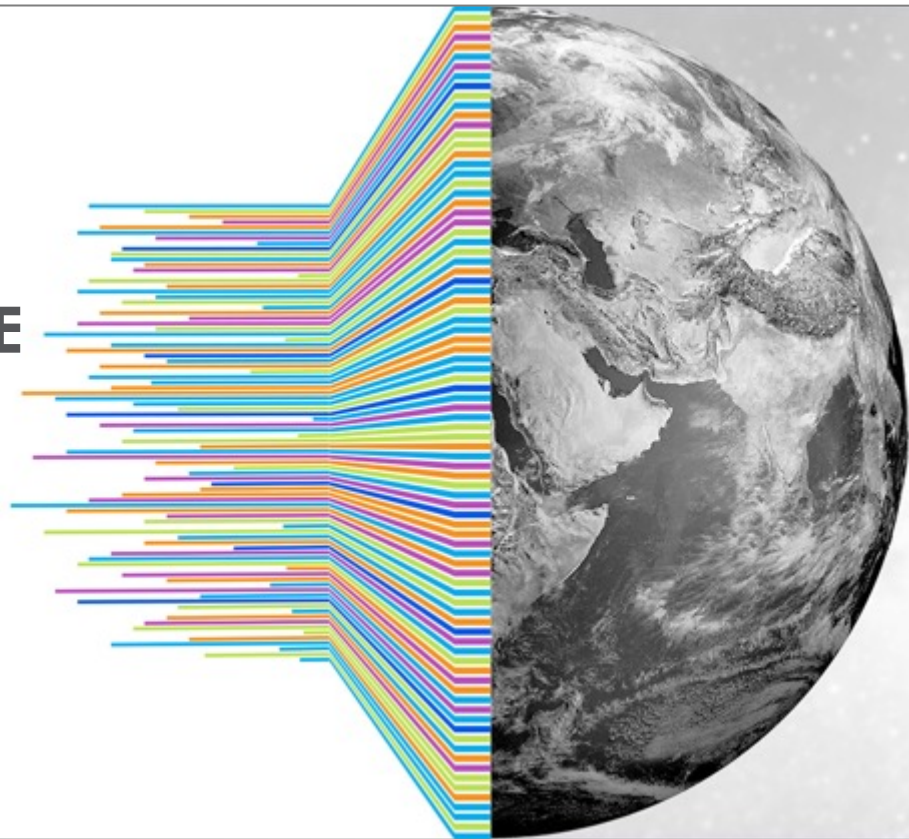


At Teradata, we believe...

Analytics and data  
**unleash the potential**  
of great companies

# 2016 ANALYTICS CHALLENGE Finalist Presentations – Part 1

Susan Baskin



# ANALYTICS CHALLENGE SESSIONS

- There are a total of 9 Analytics Challenge Finalists
  - 5 finalists presenting in Part ONE
  - 4 finalists presenting in Part TWO
- Each finalist will present for 5-7 minutes
- At conclusion of the session:
  - All teams will remain in the room for additional questions



---

**ANALYTICS CHALLENGE**

A1 - California State University Fullerton, CA, USA

A2 - Loyola University Chicago, IL, USA

A3 - National University of Singapore

A4 - Oklahoma State University, Stillwater, OK, USA

A5 – University of North Carolina, Charlotte, NC, USA

# VAERS Fears

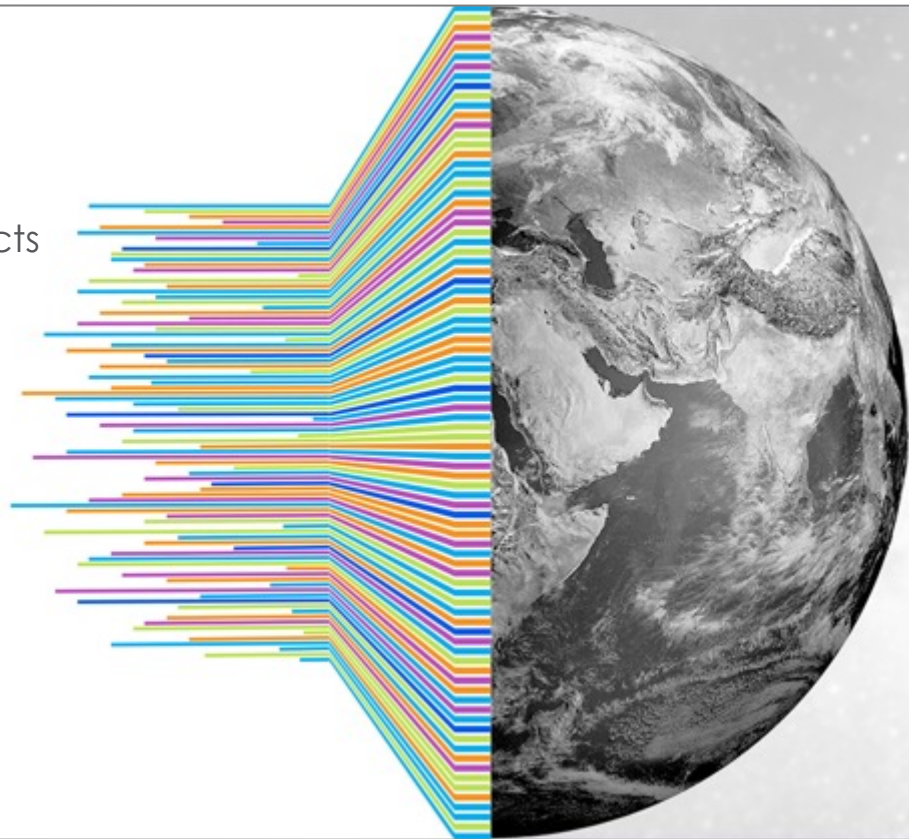
Understanding Vaccinations and Their Side Effects

Misti Vogt & Liam Watson

A1 –

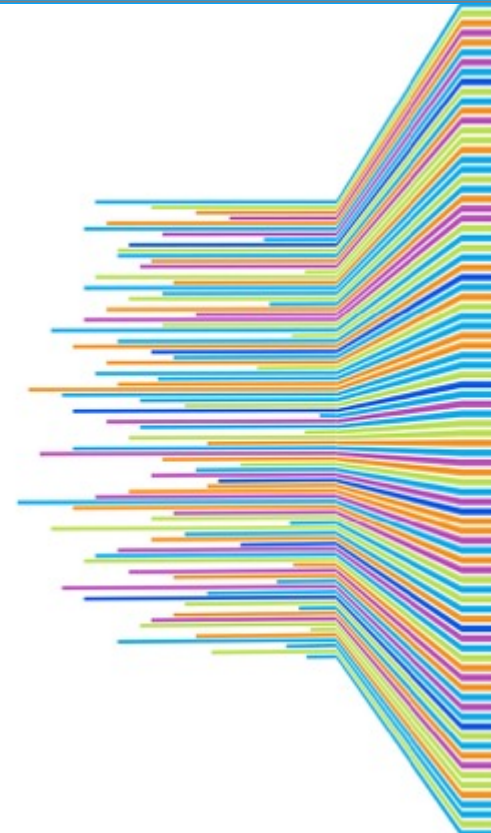


CALIFORNIA STATE UNIVERSITY  
**FULLERTON**

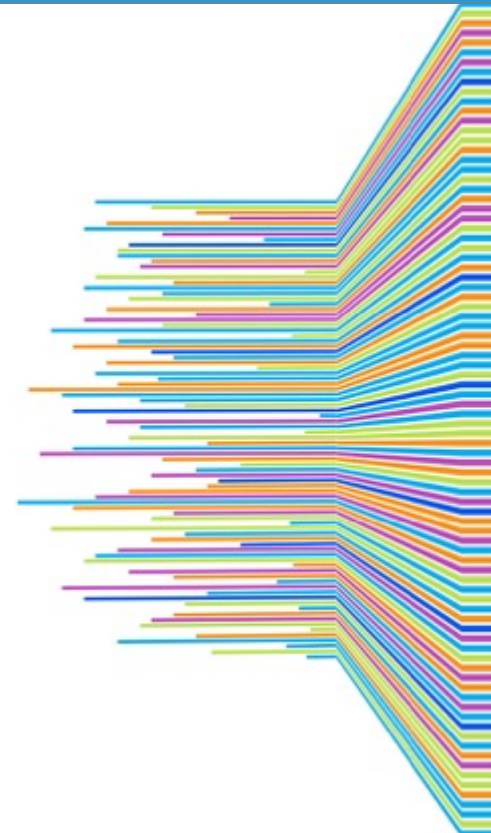


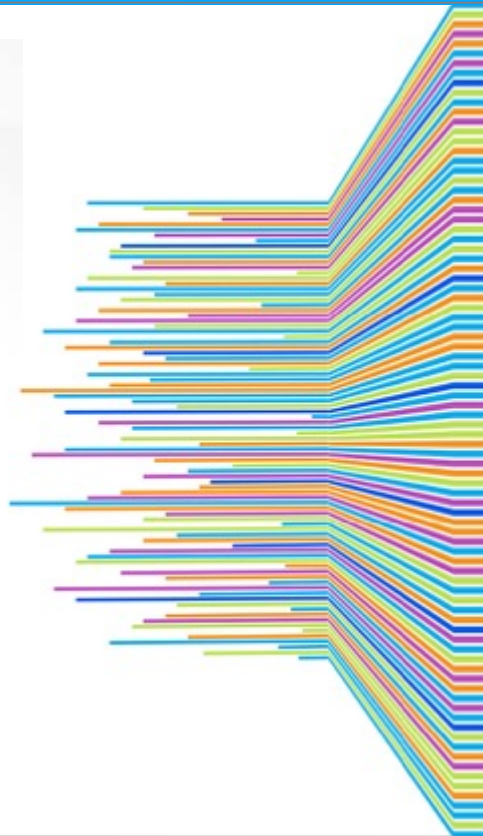
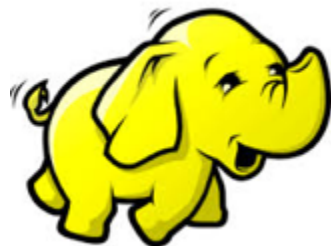
# Overview and Motivations

1. Vaccine related mortality aggregated by age, gender, and state
2. Compare most commonly reported side effects to those reported by the CDC/Drug Companies
3. 1998 Lancet study on vaccines and autism



- VAERS data (1990 to 2015) from 3 sources:
  - Biographical Data
  - Side Effects per patient
  - Vaccines (often multiple administered simultaneously)
- Data Related Issues:
  - Consistency
  - Multiple Vaccines (m:n)

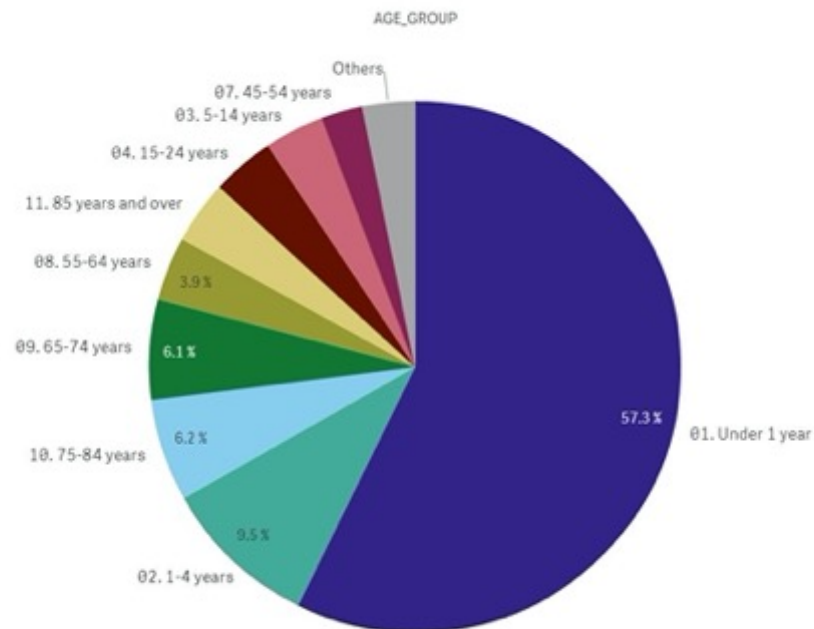






# Mortality by Age Group

## Demographics



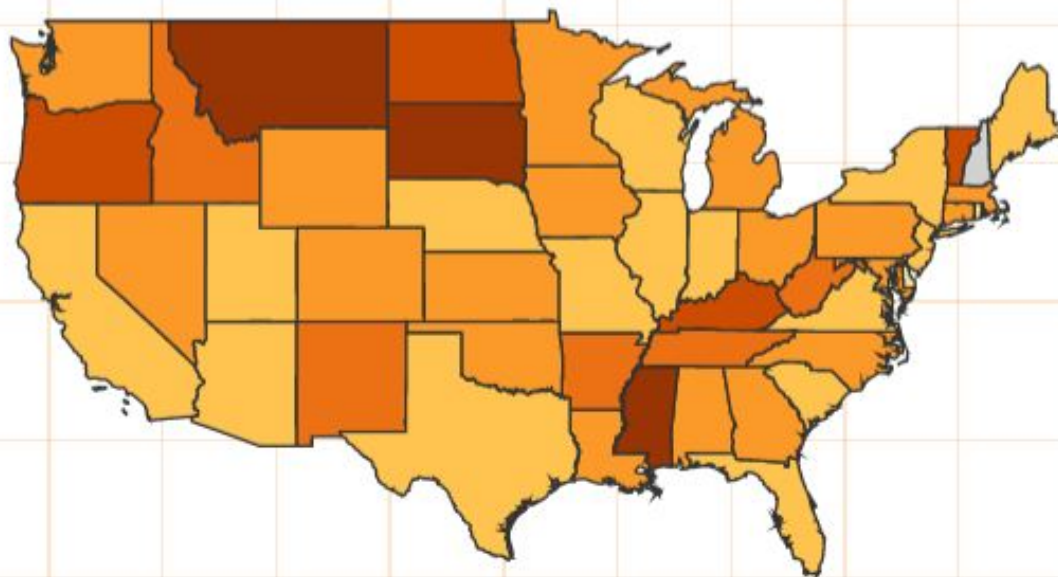
## ER Visit before Death



## Hospital Stay



# Mortality Rate Heat Map by State Population

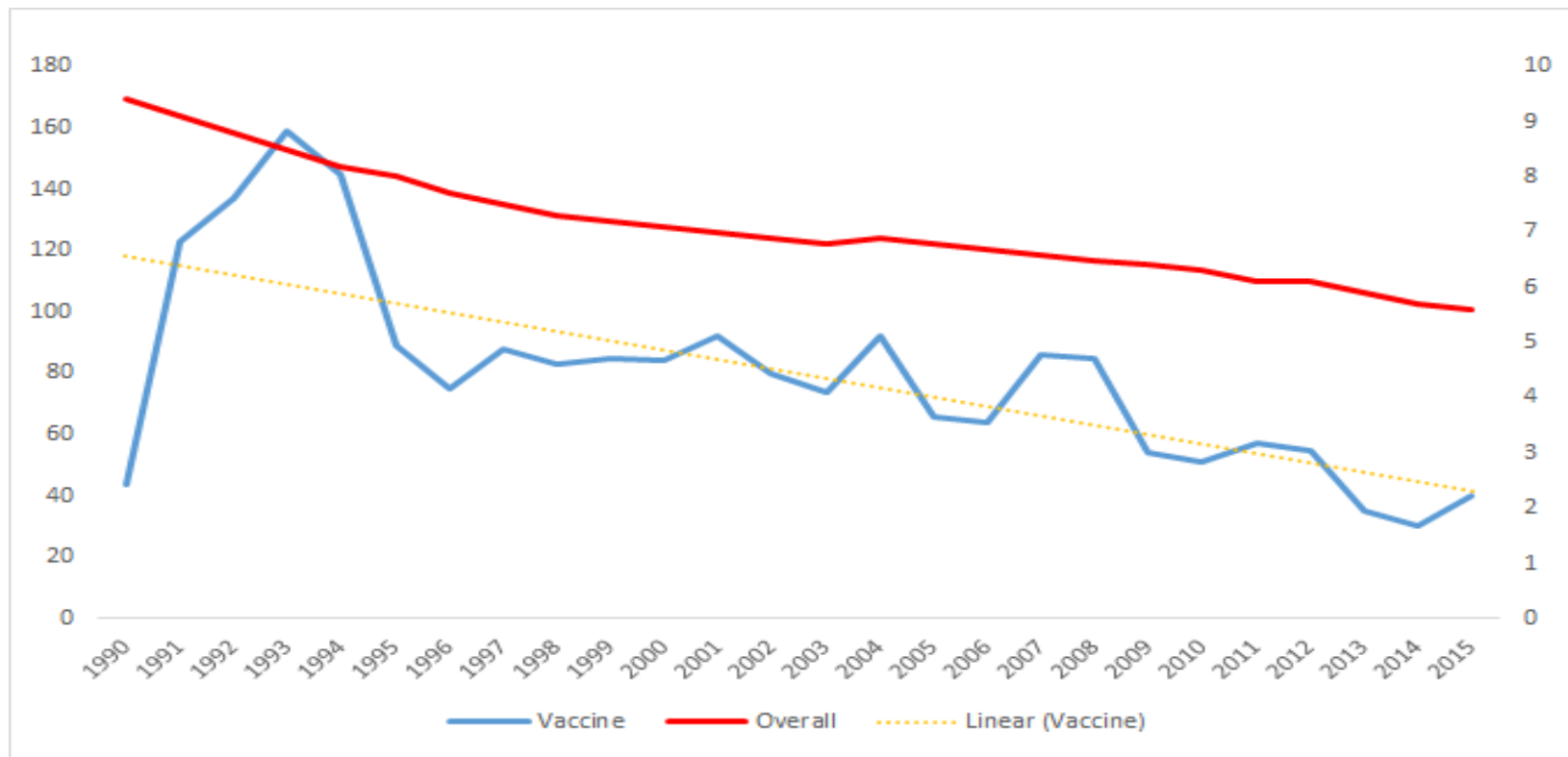


NH - 1 : 250,000

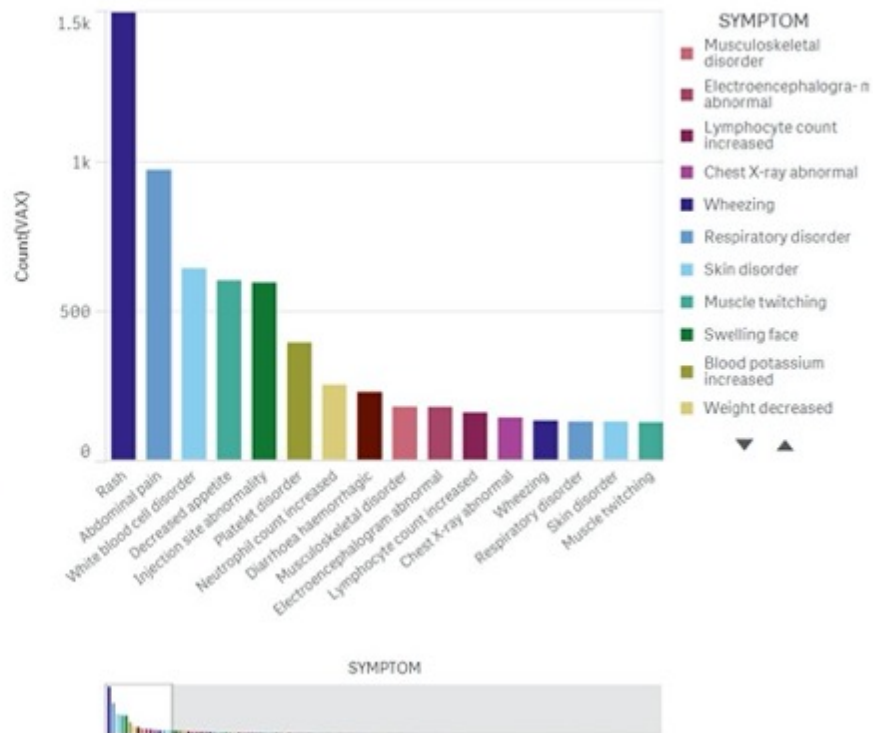
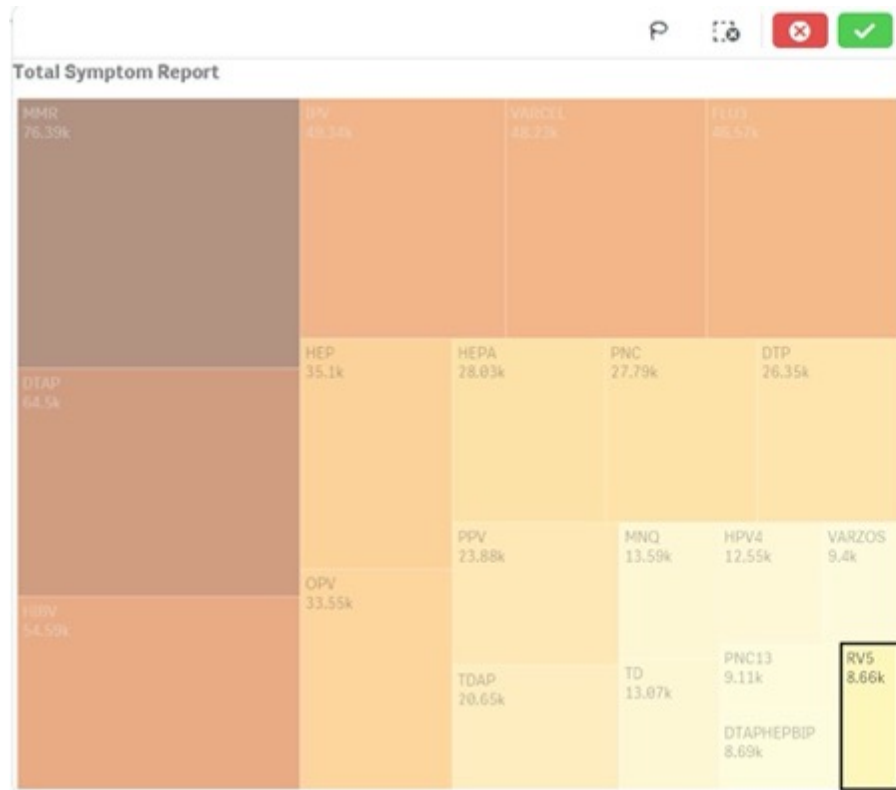
AL - 1 : 2M

National - 1 : 2.3M

# Infant Mortality Trend



# Rota Virus Symptoms



TERADATA  
PARTNERS  
CONFERENCE



# Vaccines, Autism and Death



Vaccine Related Infant Mortality – 1 : 213,000

Vaccine Related Overall Mortality – 1 : 3.5M

US Infant Mortality – 1 : 170

Shark Attacks – 1 : 3.7M



1741 total reported cases over 25 years with 13.1B  
vaccines administered

Risk of Autism from Vaccines – 1 : 7,500,000

US Autism Incidence – 1 : 68 (1% globally)

# Thank You

## Questions/Comments

Email: [mistivogt@csu.fullerton.edu](mailto:mistivogt@csu.fullerton.edu)  
[Liamwatson@csu.fullerton.edu](mailto:Liamwatson@csu.fullerton.edu)

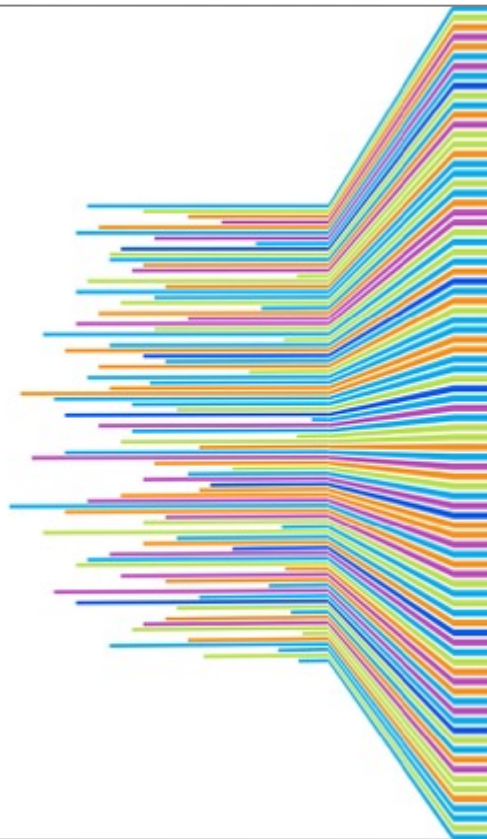
## Follow Me

Twitter @

## Rate This Session #

with the PARTNERS Mobile App

Remember To Share Your Virtual Passes



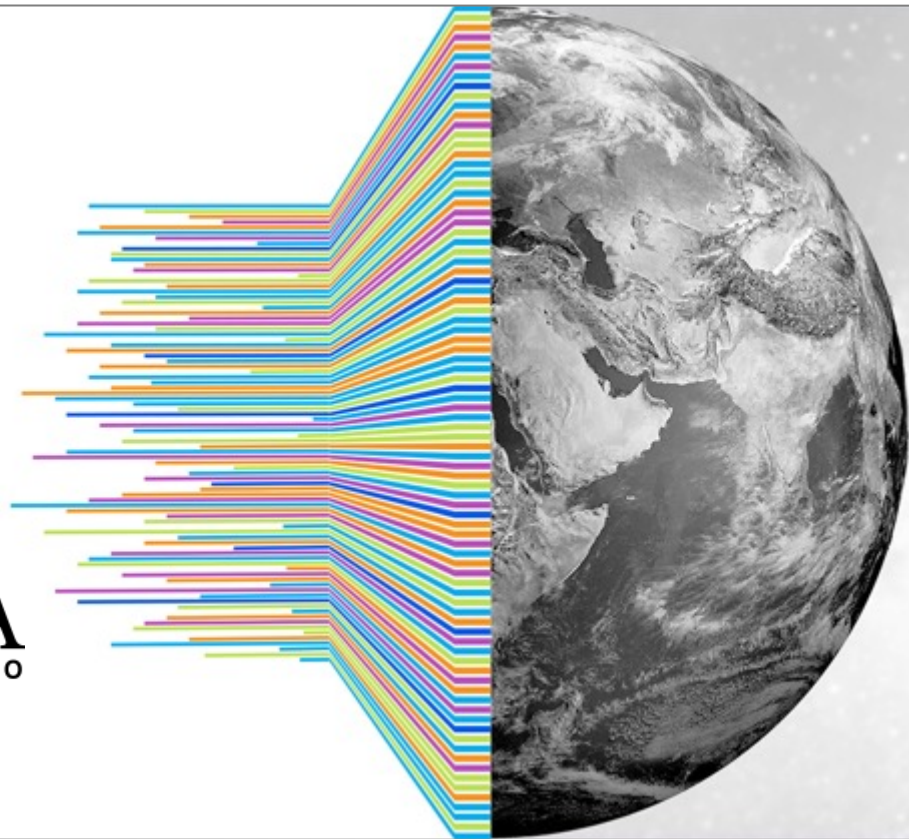
# Dan Needs a Kidney Transplant

Mit Patel  
Jonah Murray  
Spencer Christensen  
Dan Krapu  
Janki Patel



**LOYOLA**  
UNIVERSITY CHICAGO

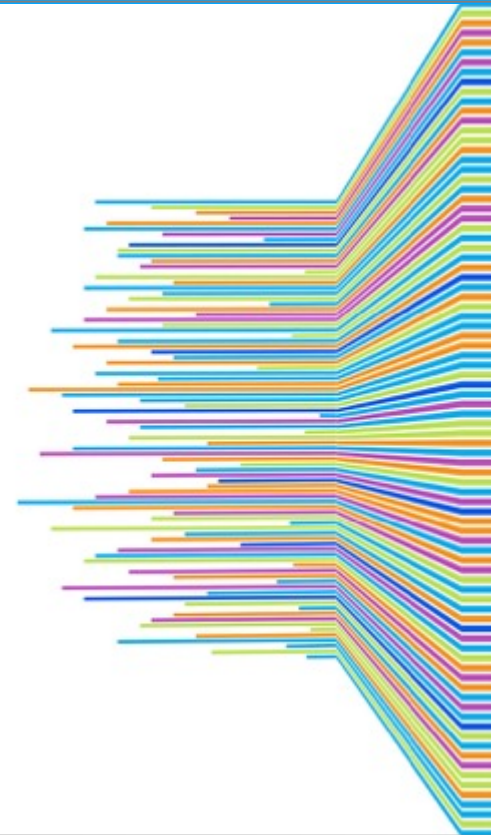
*A2 - Loyola University Chicago*





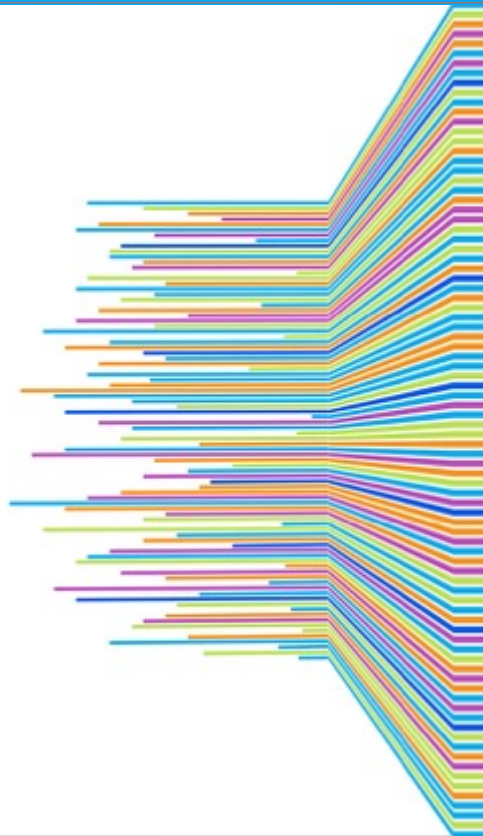
# My Friend Dan

- 21 year old male
- Early diagnosis of hypertension
- Loves Marlboros



# Project Objectives

- Find key contributors to transplant failure
- Maximize kidney and patient survival
- To help Dan live a longer, second life

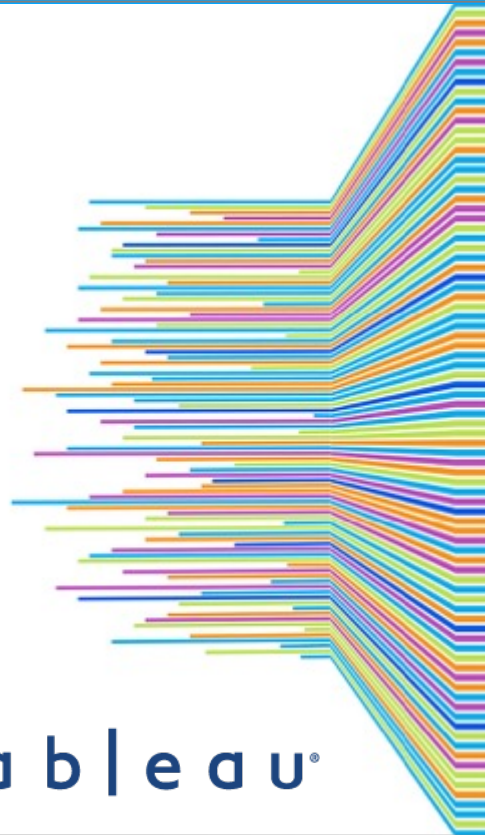


## Key Facts:

- 23 years of transplant data
- 188,000 + records of transplant data and follow-ups
- Current Waiting List for Kidneys ~ 100,000 people
- Donors ~ 8,000 this year

## Key Factors/Metrics:

- Length of Transplant Survival
- Length of Patient Survival
- Donor's History of Hypertension, Smoking, Alcohol Consumption
- Cocaine, IV Drugs, and Other Drugs History in Donor



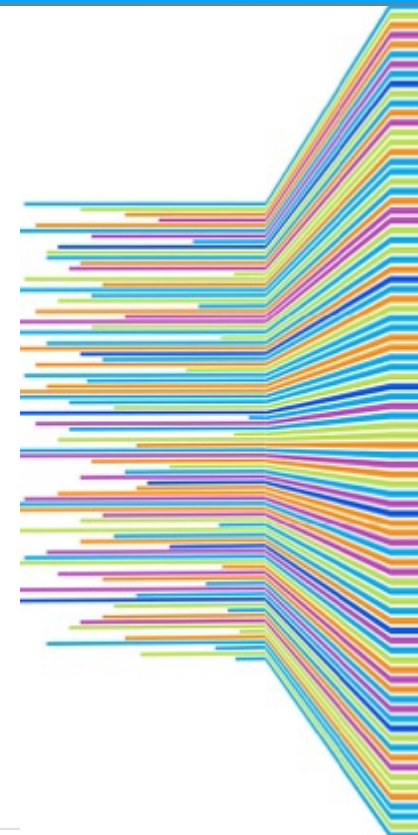
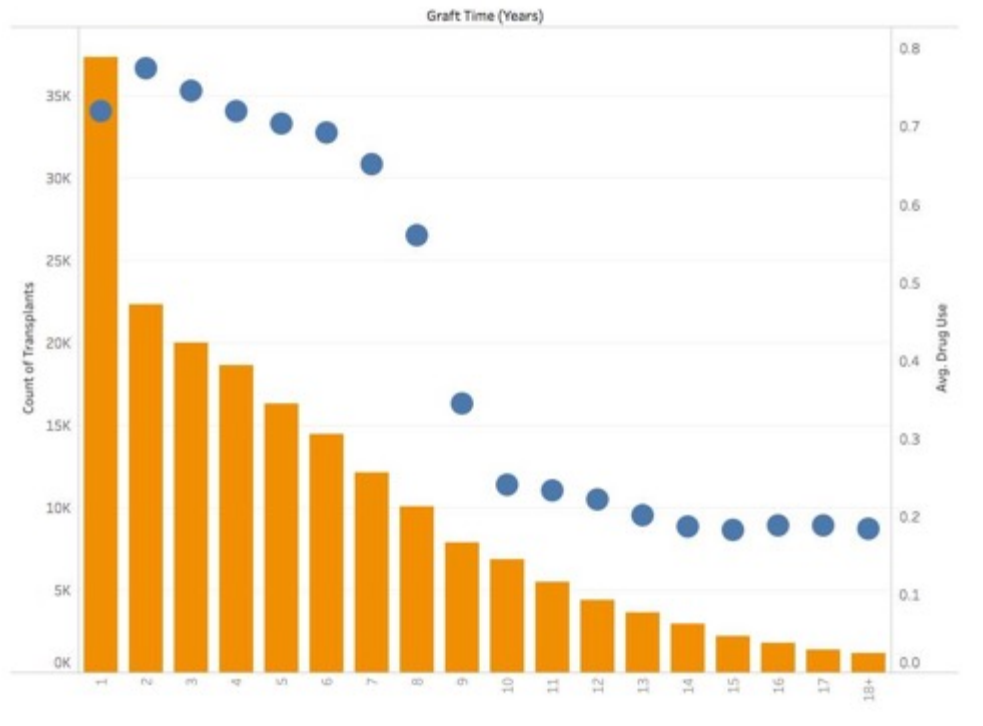
# Kidney Transplant Age vs Average Recipient Drug

Drug History recoded into:  
0 – no history of drugs  
1 – history of drugs

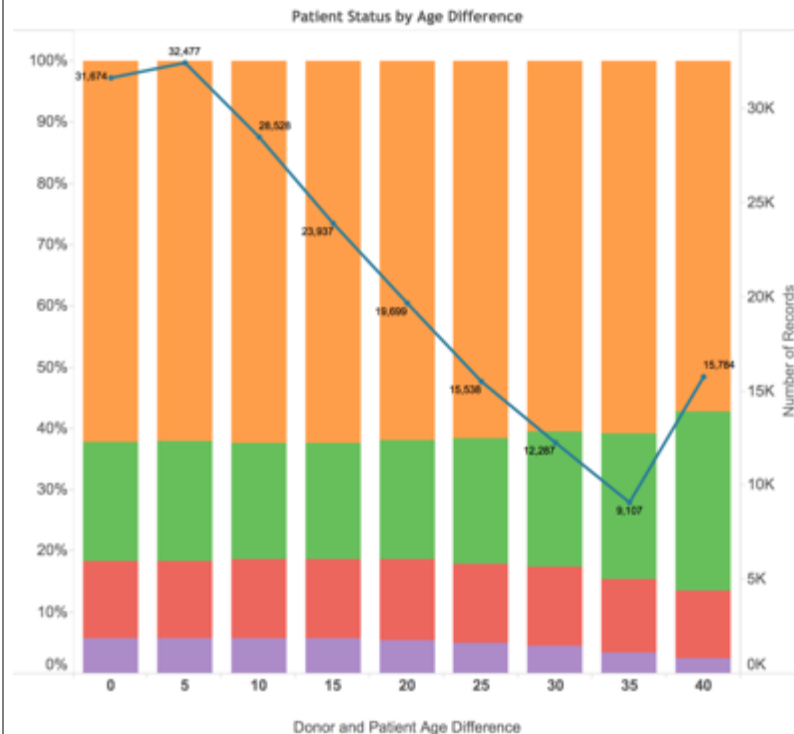
Average then taken of all  
recipient records

Drugs Include:  
Cocaine, IV Drugs, others

**Higher recipient drug usage  
correlates to lower  
kidney survival**



# Donor-Recipient Age Difference and Patient Survival

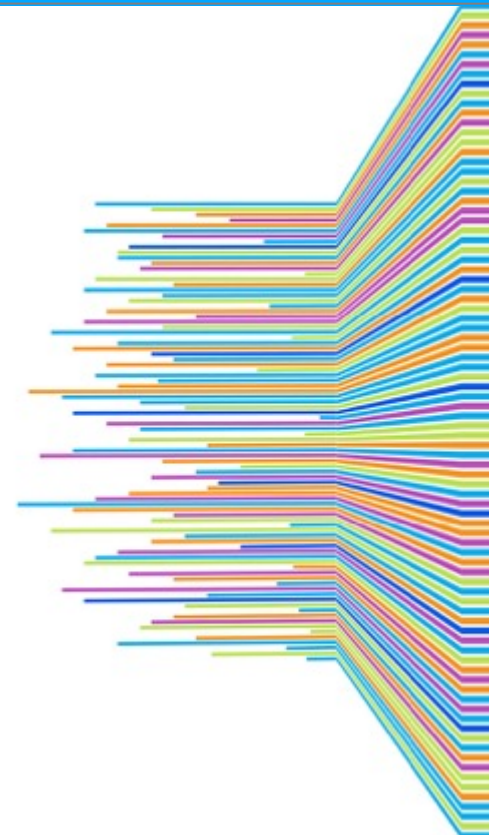


Blue Curve –  
Number of transplants done at  
each age difference

Colored Columns –  
Percentage of total transplants  
broken down by patient's  
status

**From 0 years – 25 years of age  
difference, similar outcomes  
achieved**

**From 25+ years of difference,  
Increased percentage face  
death**



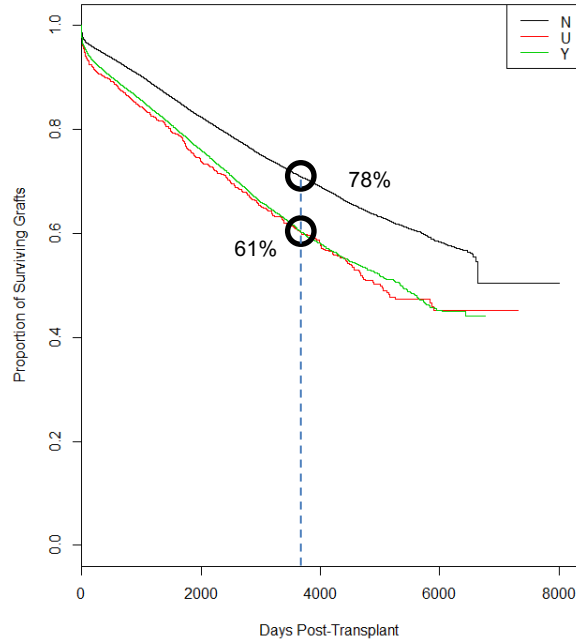
# Hypertension and Smoking

No hypertension in donor →  
higher survival rates

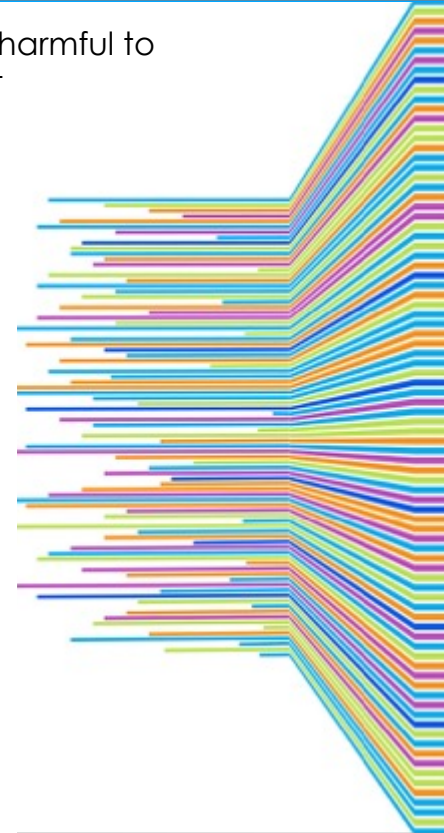
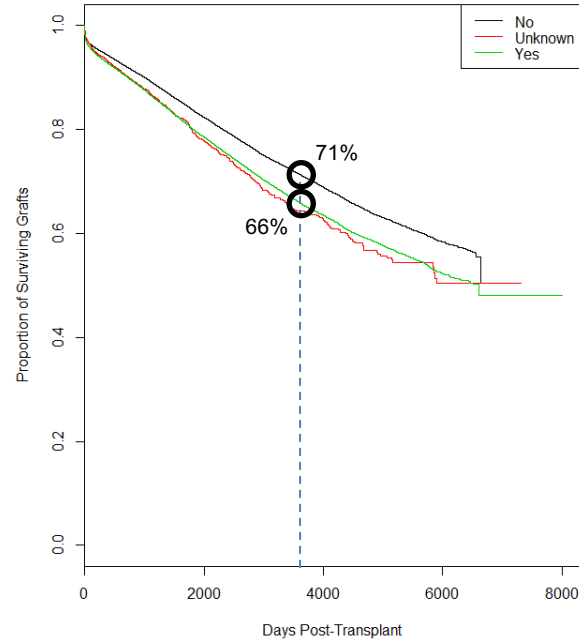
Donor doesn't smoke →  
higher survival rates

Hypertension more harmful to  
transplant

20 Year Graft Survival  
by Donor Hypertension History

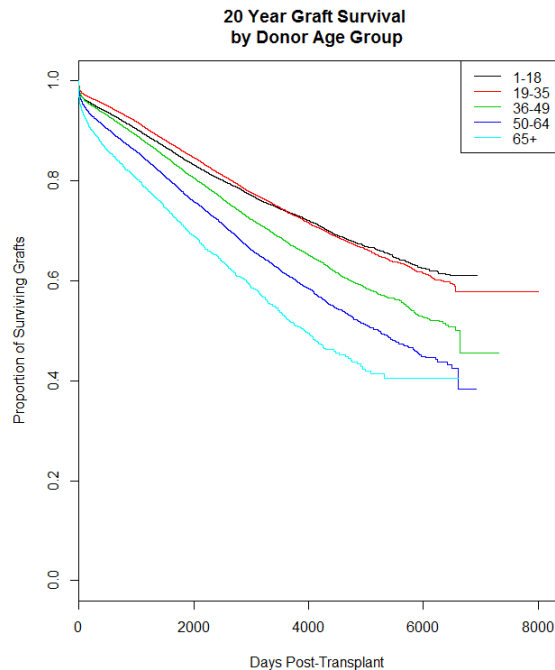


20 Year Survival Analysis in Grafts  
by Donor Smoking History

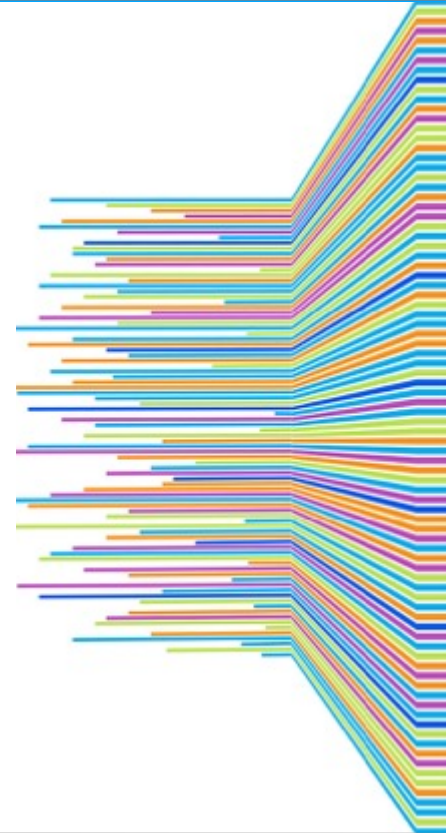
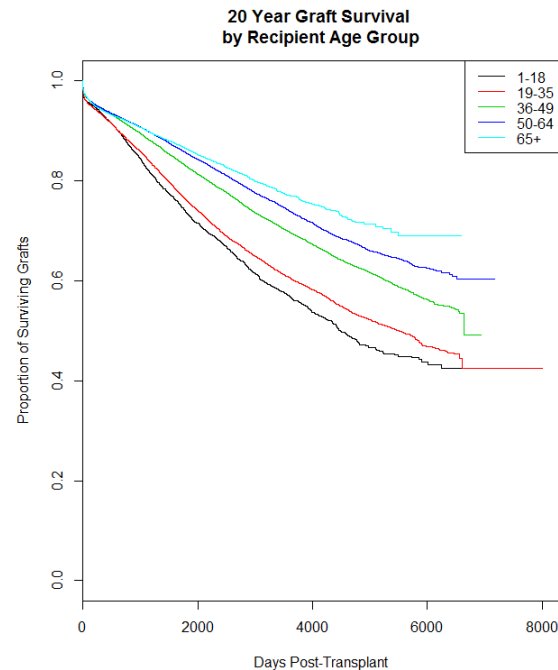


# Age of Donor and Recipient

Younger kidneys in older patients  
expected for longer kidney survival



Health of kidney determinant  
of long-term kidney success





- ~20% of population's kidney failed within 1 year
  - Reduce drug consumption
  - Consider quitting
- Up to 25 years of age difference between Donor and Recipient—similar outcomes
  - Could open up new donor-recipient matching possibilities.
- Reduce risk factors for hypertension
  - 78% vs. 61%
- Consider smoking cessation
  - 71% vs. 66%
- Educate the public early on disease prevention





# Thank You

## Questions/Comments

Email: [mpatel26@luc.edu](mailto:mpatel26@luc.edu)

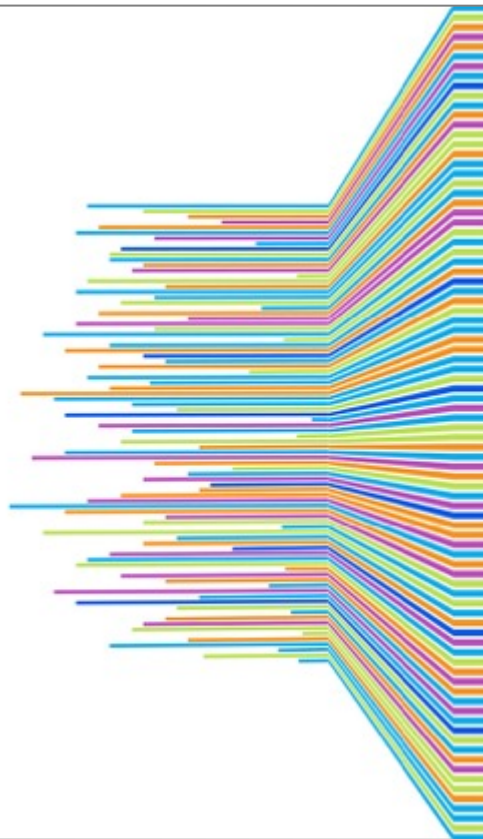
## Follow Me

Twitter @ Instagram @ letaptim23

## Rate This Session #

with the PARTNERS Mobile App

Remember To Share Your Virtual Passes

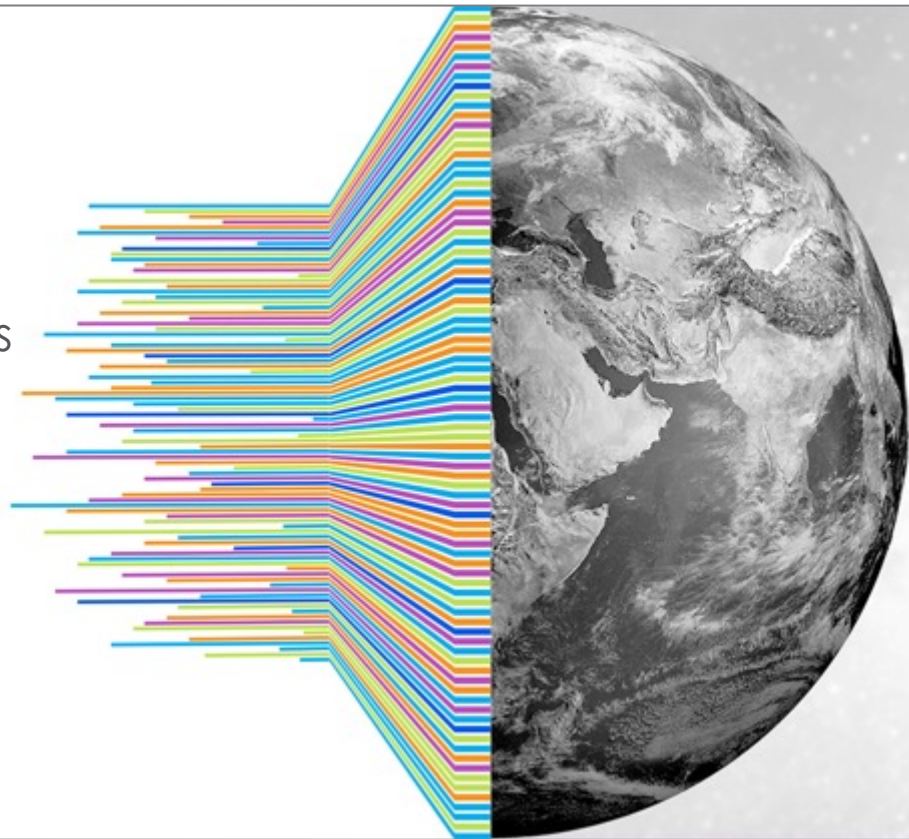


# Terror-Eyes

An Immediate Response System to  
Proactively Deal with Terrorist Threats

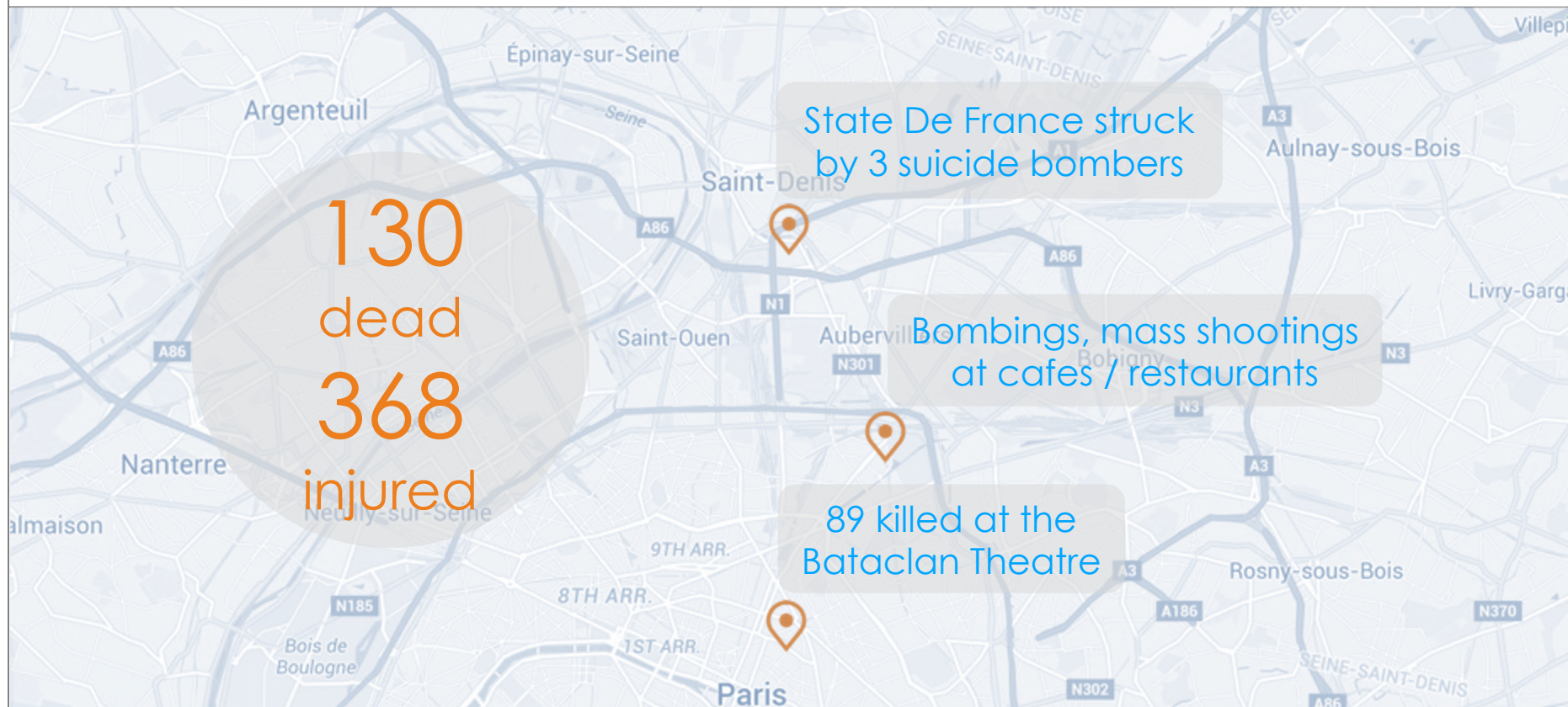
Andre Tan, Eddy Pang, M Thirukkumaran

A3 – National University of Singapore



- Global terrorism on rise: Fivefold increase in terror-related deaths since 2000
- Almost 18,000 people were killed in **terrorist attacks** in 2013, a 61% **increase** from 2012
- The threat of terrorism is real, we need to fight against it **proactively, not reactively**.

# November 2015 Paris Attacks



# What could we have done?

## Timeline (CET)

21:20

Bomb 1, State De France

21:25

Shooting, Rues Bichat & Alibert

21:30

Bomb 2, State De France

21:40

Shooting and hostages, Bataclan

21:53

Bomb 3, State De France

Is there a pattern to these attacks?

What if...

The first attack could be used to prevent subsequent ones?

---

We already have information documenting terrorist attacks.

What we need, is **fast** and **actionable** information.



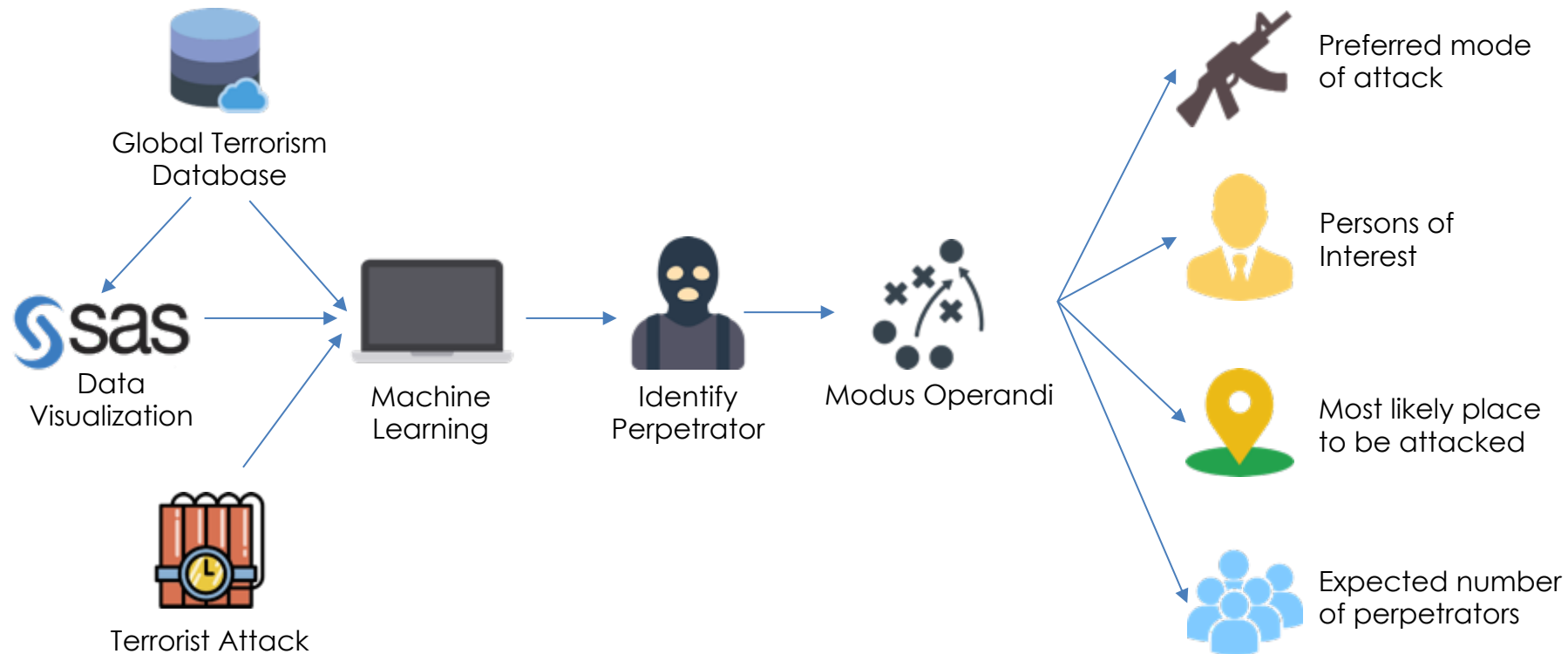
Terror-Eyes

An **Immediate Response**  
System to Proactively Deal  
with Terrorist Threats

# The Problem: Information Asymmetry

- Who is responsible
  - Where they will attack next
  - How they plan to attack
  - What other information will help prevent unnecessary loss of life
- 
- Make sense of large amounts of historical data to determine how best to allocate manpower and resources

# Methodology





# The Terror-Eyes Ecosystem



- Holistic overview
- Develop a mental model of the perpetrator
- Visually identify areas that are at higher risk
- Improved allocation of manpower and resources



- Twitter bot to crowdsource anti-terrorism efforts by empowering the public to help identify suspicious activity
- Collects and classifies images with pre-trained Neural Nets to provide authorities with an on-the-ground perspective

# Implementation

## Data Exploration



- Finding patterns in the data
- Develop an intuition on what features might be important



## Web Application

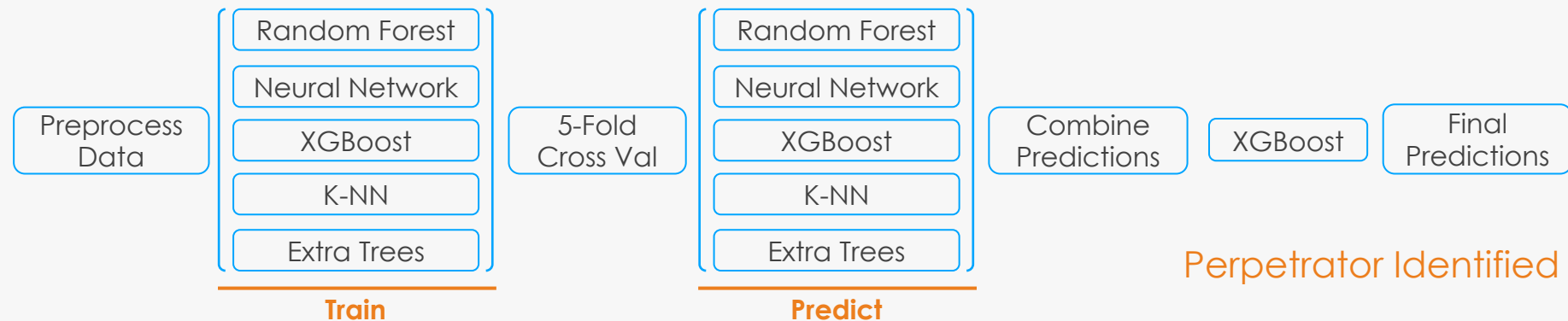


Flask



- Dashboard interface
- Help authorities understand quickly what has to be done

## Predictive Modeling



# Terror-Eyes in Kabul, Afghanistan

**The New York Times**

"Heavily armed insurgents wearing suicide vests struck Tuesday, demonstrating the Taliban's ability to infiltrate even the most heavily fortified districts of the capital."

**BBC  
WORLD**

"Tuesday's attacks appear to be a complex operation. At 13:30 local time, insurgents fired rockets on targets in Kabul's upmarket embassy district."

## Predicted

## Actual



Most likely perpetrator predicted to be:  
**Haqqani Network**

Responsibility claimed by Taliban.



Probability of multiple attacks:  
**65%**

Initial attack was followed up by 2 subsequent attacks.



Most likely place to be attacked:  
**Military**

Subsequent attacks at Police Stations and a jail run by intelligence services.



Preferred choice of weapon:  
**Explosive / Bombs / Dynamite**

Suicide bombings and rockets.

# Terror-Eyes in Kabul, Afghanistan

The New York Times

BBC  
WORLD

"KABUL, Afghanistan — The American University of Afghanistan in Kabul came under attack by bomb and gunfire on Wednesday night, in a siege that lasted for hours as pockets of people trapped on campus tried to escape."

"Thirteen people, including seven students, died in an attack by gunmen on the American University in the Afghan capital Kabul, police say."

## Predicted

## Actual



Most likely perpetrator predicted to be:  
Taliban

Responsibility claimed by Taliban **2 days later**.



Probability of multiple attacks:  
24.53%

No multiple attacks



Most likely place to be attacked:  
Private Citizens & Property

No multiple attacks



Preferred choice of weapon:  
Explosive / Bombs / Dynamite

No multiple attacks

# Thank You

## Questions/Comments

Email: [andre@u.nus.edu](mailto:andre@u.nus.edu), [ttt@u.nus.edu](mailto:ttt@u.nus.edu), [eddytang93@gmail.com](mailto:eddytang93@gmail.com)

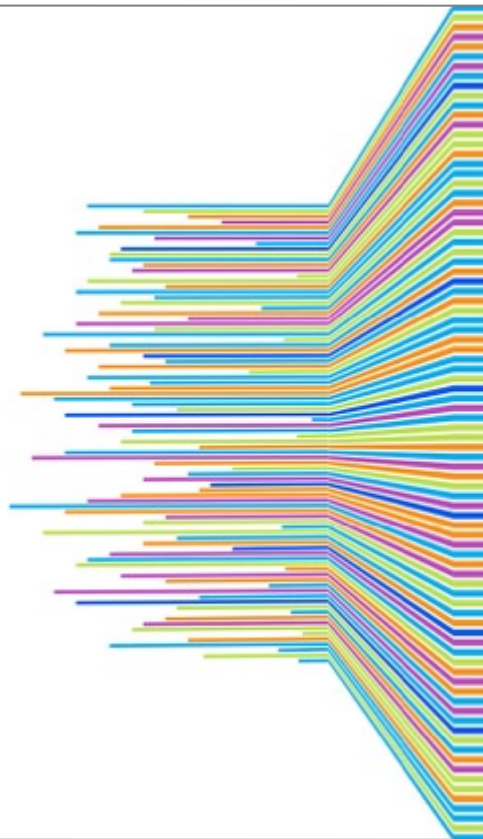
## Follow Me

Twitter @ [terrorbgone](#)

## Rate This Session #

with the PARTNERS Mobile App

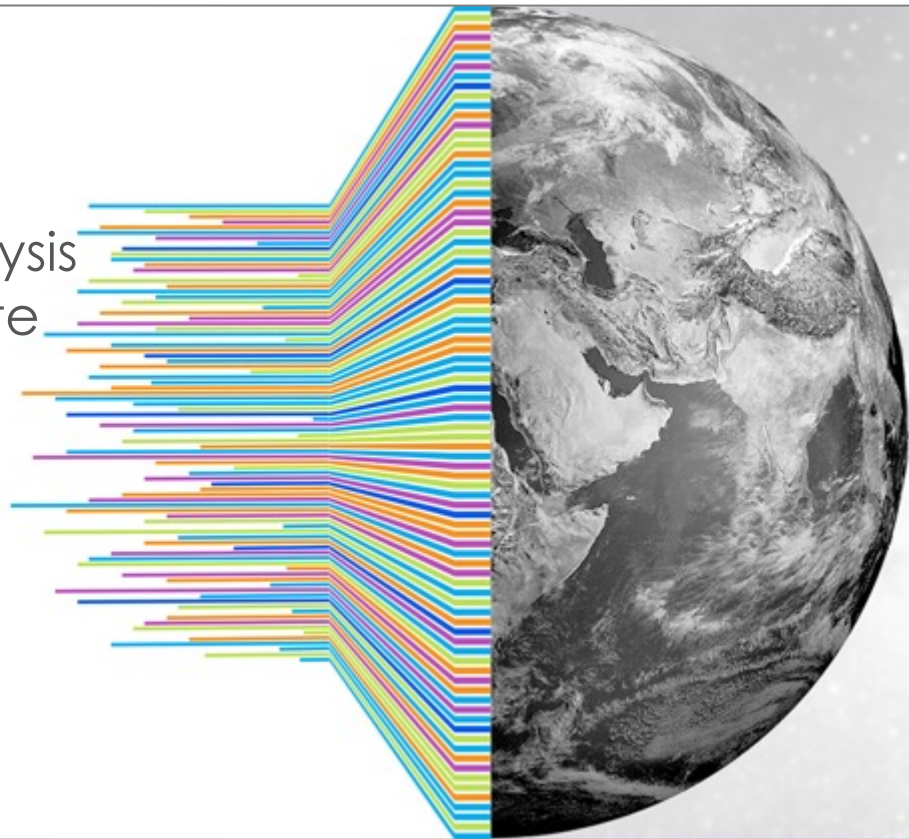
Remember To Share Your Virtual Passes



# HEALTH ANALYTICS ON TOBACCO USE DISORDERS

Aster Npath Analysis, Network Analysis  
and Prediction of Readmission rate

**Bhargav Molaka**  
**Samsheel Kathuri**



# Project Objective and Scope

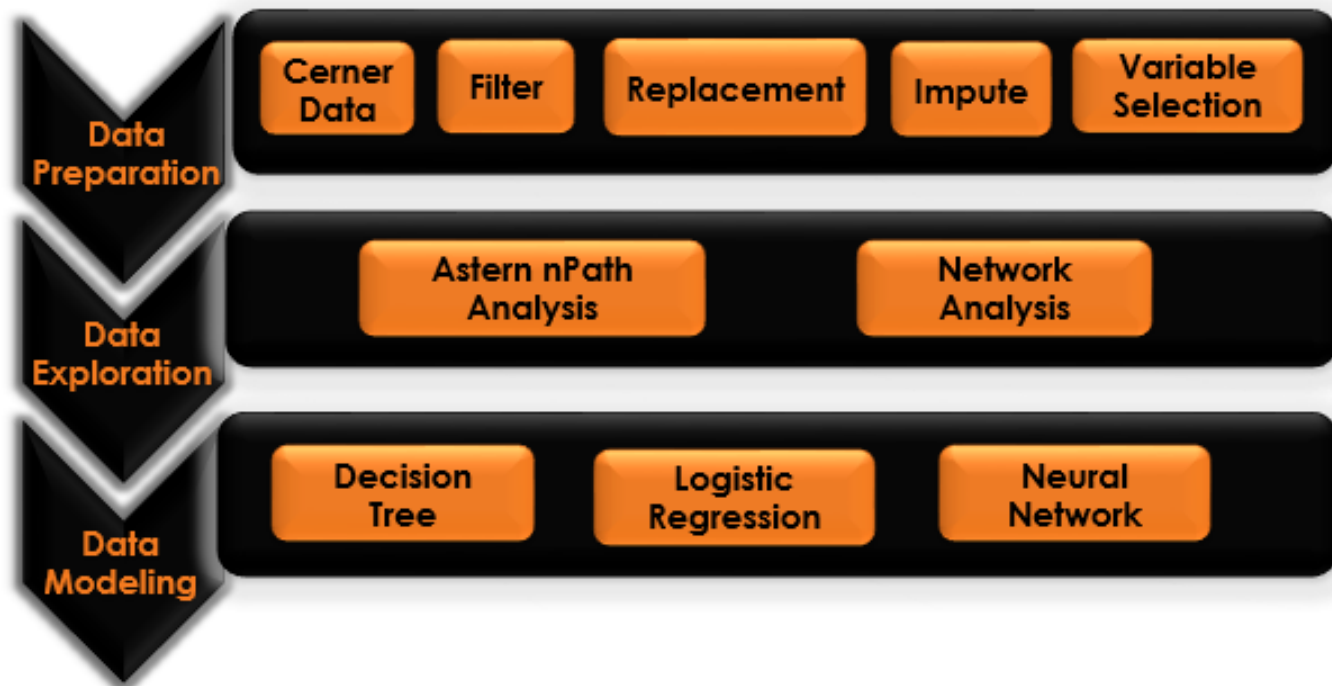
## ➤ Objective

- To help the hospitals to reduce the readmission penalty

## ➤ Scope

- To explore patient's path using Teradata Aster Npath
- To analyze the comorbidity effect among the patients
- To model and predict patient's readmission rate

# Tools and Techniques Used



TERADATA. ASTER

Teradata Studio

TERADATA AppCenter

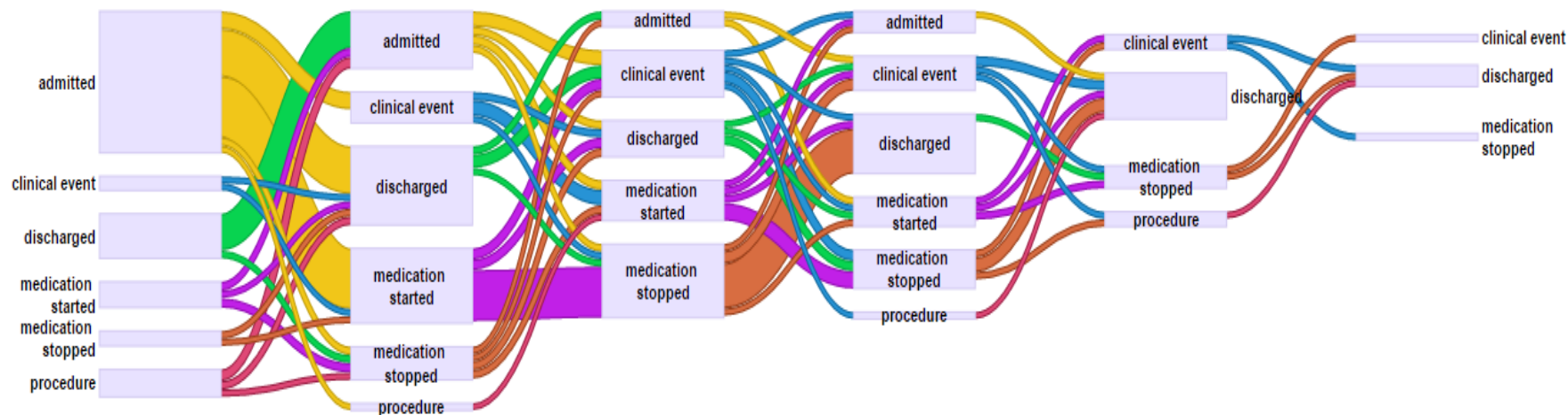




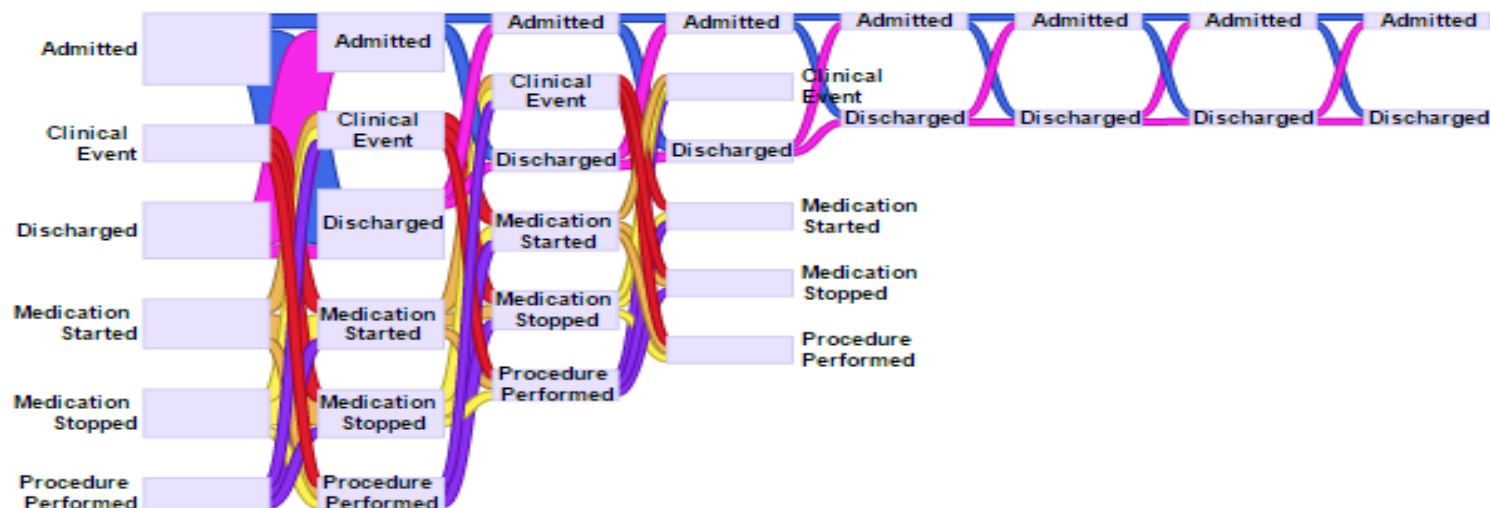
# Data Preparation



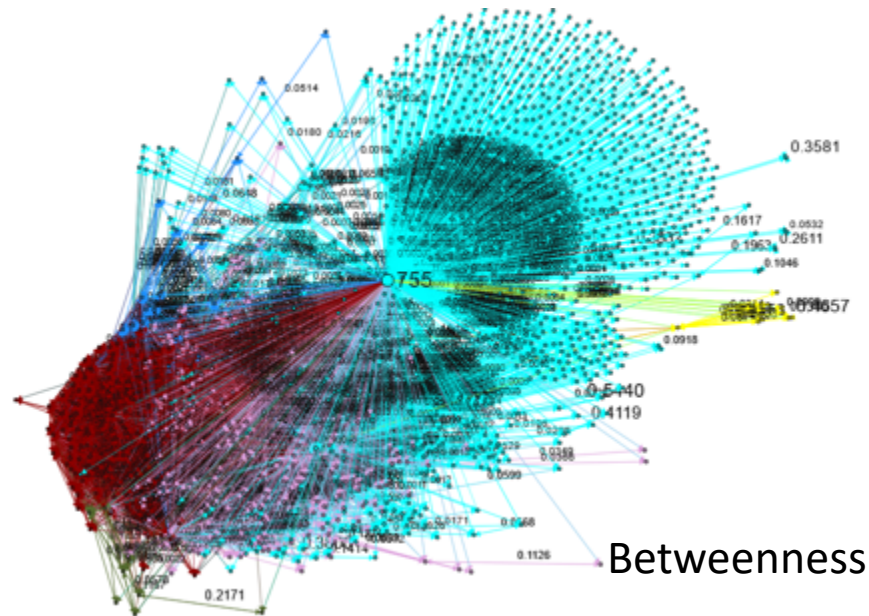
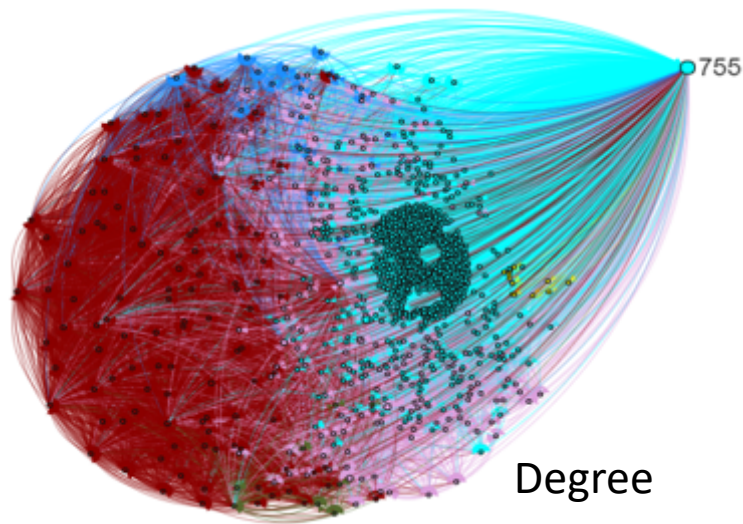
## ➤ Aster Npath Analysis: Rural Region



## ➤ Aster Npath Analysis: Urban Region

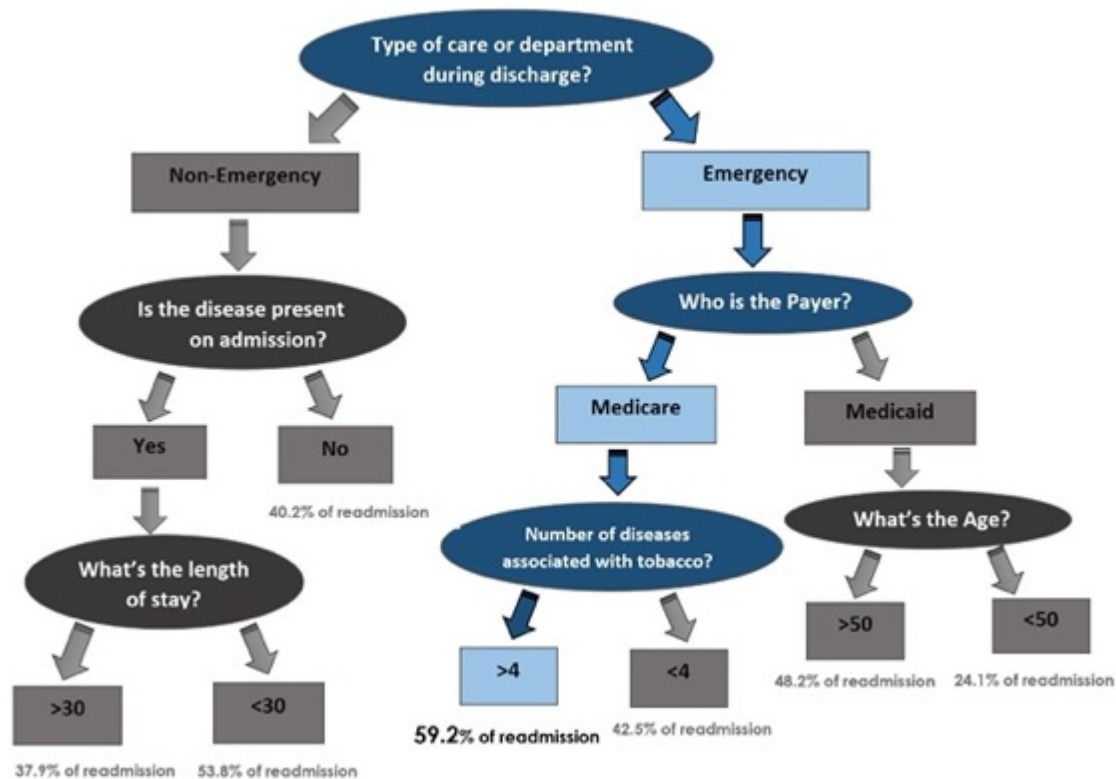


## ➤ Network Analysis



- Diagnosis Id “755” represents the tobacco use disorders
- Tobacco Disorder has high Degree and high Betweenness centralities
- Degree centrality is one of the important variables in prediction

# Decision Tree Model

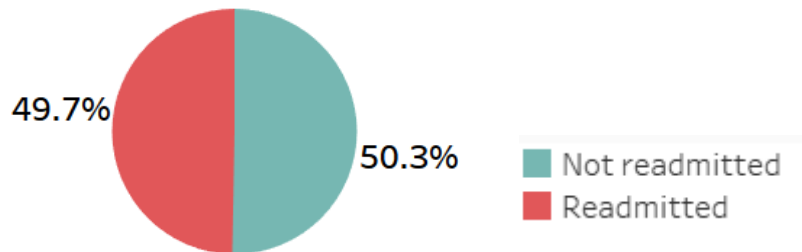


## Variable Importance

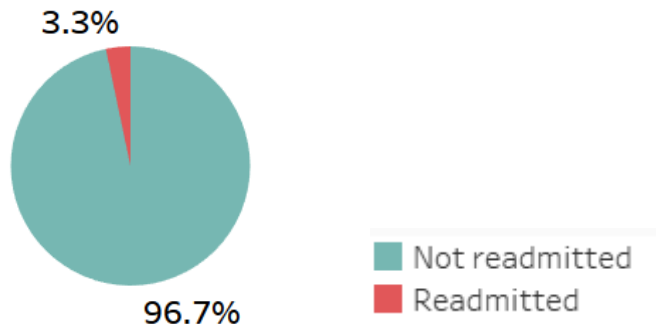
Variable Name	Variable Importance
Discharge Care Setting	1.0000
Present on Admit	0.7400
Payer Code	0.4155
Length of Stay	0.3564
Degree	0.2089
Age	0.1600

# Modeling Results

## ➤ Prediction of Readmission Rate



## ➤ Prediction of 30 day Readmission Rate



## ➤ Accuracy of the Models

Models	Validation Misclassification Rate	Training Misclassification Rate
Decision Tree	35.0 %	34.7 %
Neural Network	35.3 %	35.0 %
Regression	35.8 %	35.5 %
Entropy Tree	36.5 %	36.1 %

- Used stratified balanced sampling for model building
- The results shown are with respect to the Validation Data

- Through Aster NPath, we observed that the patients were getting readmitted after their discharge
- Through Network analysis, the comorbidity effect on tobacco use disorders were observed
- Discharge Care setting and Network Measures like Degree, Eigen centrality were important variables
- Almost 50% of the patients are readmitted
- Around 3% of the patients are readmitted within 30 days

# Thank You

## Questions/Comments

Email: [bhargav.Molaka@okstate.edu](mailto:bhargav.Molaka@okstate.edu) , [samsheel.kathuri@okstate.edu](mailto:samsheel.kathuri@okstate.edu)

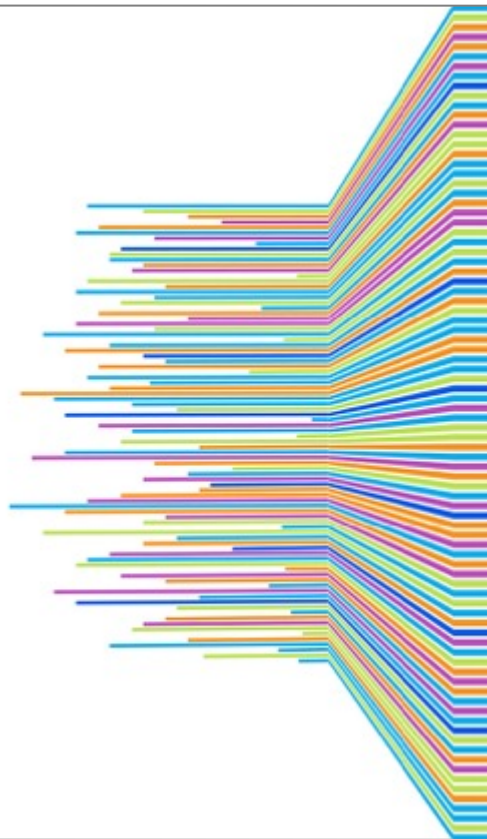
## Follow Me

Twitter @ MolakaBhargav, samsheel

## Rate This Session #

with the PARTNERS Mobile App

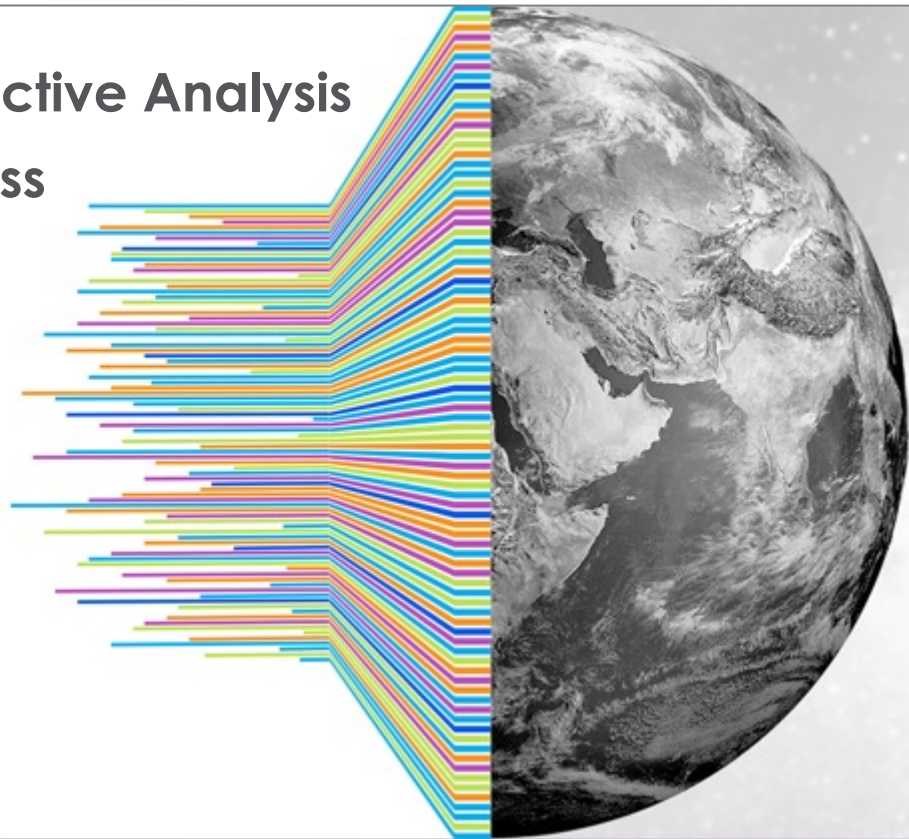
Remember To Share Your Virtual Passes



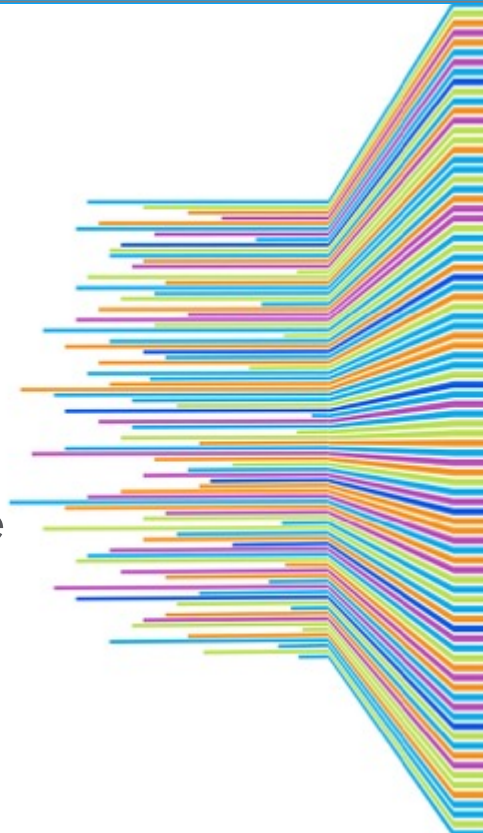


## A Big Data Approach to the Predictive Analysis on Fortune 500 company's success

Divya Ravi  
Xiazhi Fang  
Diana Kinney  
Malden Ivanova  
Aravindharaj Rajendran  
Kerri Easterbrook  
Mahalakshmi Vishnampettai Raghuraman



- Objective & Approach
- Data Sourcing
- Data Cleansing & Data Mining
- Relationships between variables
- Analysis of company with most patents filed
- Analysis of company with highest Total Revenue
- Impact of Presidential Parties



# Objective & Approach

To gain insight into the relationships between Fortune 500 company performance, analytics, innovation, social responsibility, patent development, and current political climate in order to make valuable predictions.

## Objective

- Impacts of social responsibility, patent development and political climate on Fortune 500 company performance

## Data Sourcing

- Innovation – Patents, Economy – GDP, Politics – Election Data

## Data Staging

- Structured - Excel / MS Access
- Semi-structured – File System

## Data Preparation

- Tokenization & Stemming, Stop Words Removal – Patent Abstract & Titles, Missing Data – Omission/Imputation

## Data Mining

- Patents – MapReduce, Word Cloud, Topic Modeling – Mallet

## Measures

- Testing - Recall
- Descriptive Measures - R Studio

## Visualization

- Patterns, Variable Relationship Analysis - Tableau

## Insights / Recommendations

- Variables, Relations, Relevant Data Consolidation, Hypothesis Prioritization

# Data Sourcing

TERADATA  
PARTNERS  
CONFERENCE



## Data Collection

### Structured Data:

- Federal Archives
- CompuStat
- FEC.gov
- US Census Website (NAICS)

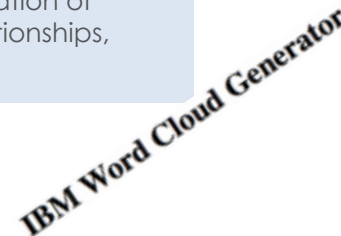
### Semi-Structured Data (Patents):

- US Patent Trademark Office
- Data.epo.org



**Compustat®**

A Standard & Poor's Capital IQ Business



### Technology stack

Java – MapReduce

Eclipse IDE - Development and Testing

SAS - Analysis/Company name sounds like

MS Access - Data preparation and mapping

IBM Word Cloud - Visualization using word cloud

Mallet - Topic modeling

Tableau - Visualization of variable relationships, trends



# Data Cleansing & Data Mining

**Deployment/Execution**  
DSBA Hadoop Server

## INPUT

- XML Patent Files

## PROCESSING

- MapReduce

## OUTPUT

- Org name & patent count
- Patent Title
- Patent Abstract

**Development**  
Eclipse IDE on Cloudera Hadoop distribution

Output:  
Company  
Name &  
Patent Count

Pull out  
Fortune 500  
Companies'  
data

Structured  
Data  
Imported  
into MS  
Access DB

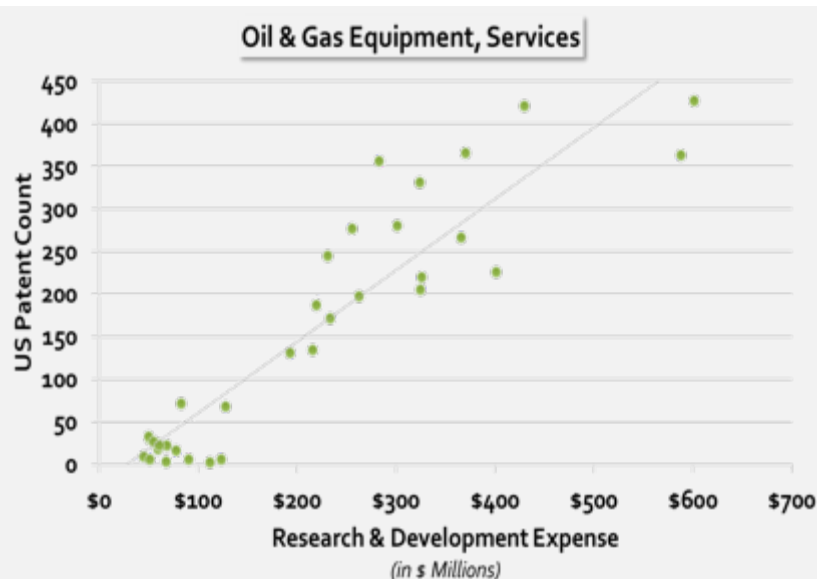
Join with  
other data  
sets

# Data Descriptive Statistics

- 14,897 Total Records
- 40+ Variables Collected & Examined
- 485 Fortune 500 Companies (97%)
- Timeframe analyzed:
  - Corporate Financial and Presidential Election Data: 1950 – 2015
  - Patent Application Filings: 2005 - 2014

Avg. Revenue	Avg. Net Income	Avg. R&D Expense	Avg. Patent Filings
\$11,574 million	\$658 million	\$436 million	170 (med. 23)

# Relationships between variables

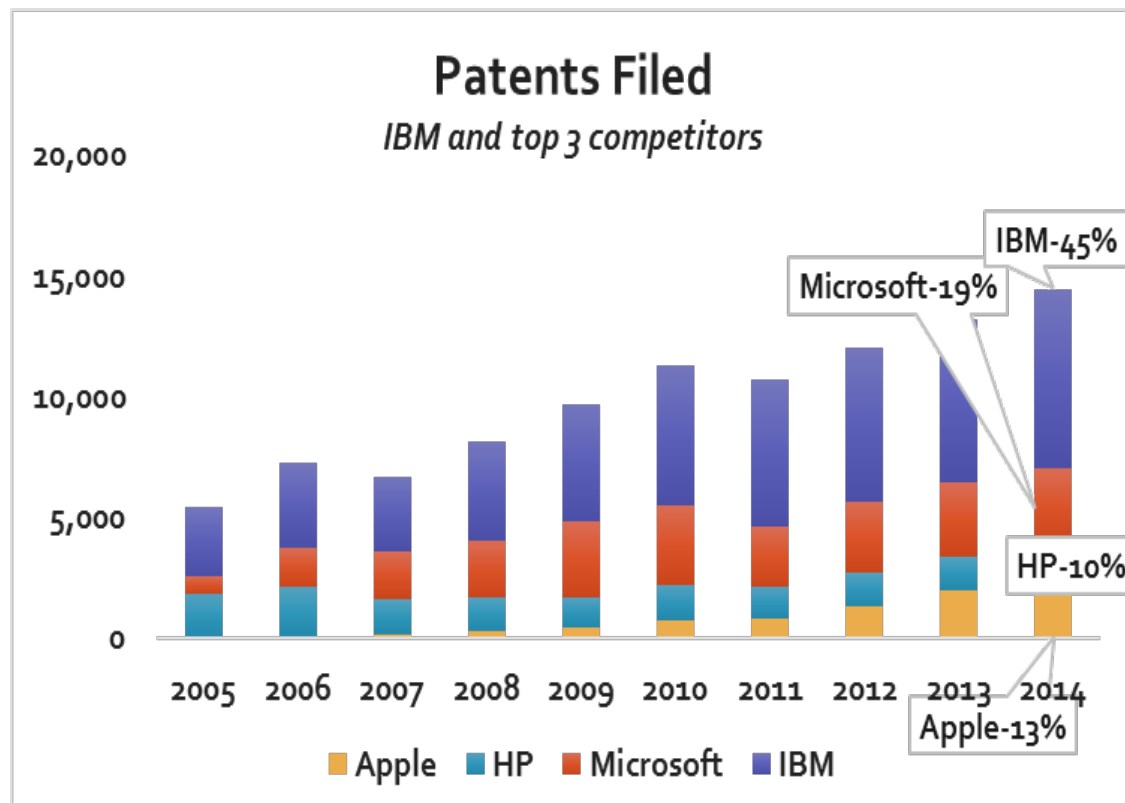


**IV: R&D Expense, DV: US Patent Count**  
It costs \$1.2 million R&D to produce 1 patent.

**IV: US Patent Count, DV: Revenue**  
For every 1 patent, will see \$43 million increase in revenue.

## Cluster Analysis

Cluster Analysis (with k = 4)	Cluster 1	Cluster 2	Cluster 3	Cluster 4
	Most Environment Concerns	No CSR	Somewhat CSR Committed	Most CSR Committed
Community Concern	0	0	0	0
Community Strength	0	0	1	1
Corporate Governance Concern	0	0	0	1
Corporate Governance Strength	0	0	0	0
Diversity Concern	0	0	0	0
Diversity Strength	1	0	2	4
Environment Concern	3	0	0	1
Environment Strength	1	0	1	1
Product Concern	1	0	1	1
Product Strength	0	0	0	0
Average Revenue (in Millions)	\$ 29,191	\$ 9,882	\$ 21,194	\$ 37,926
Average Patent Count	96	92	137	302





- **Continue business analytics and research**

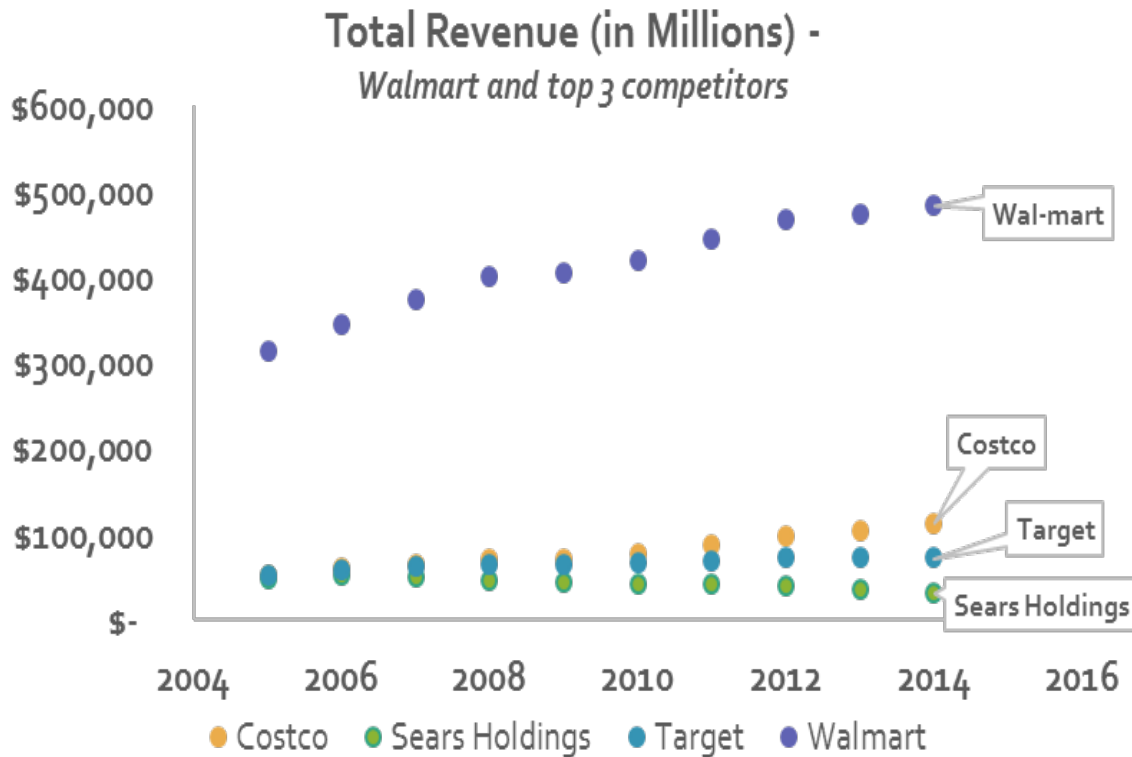
- Rising revenue within this sector- Every \$1 spent in R&D accounts for \$ 23 in revenue

- **Attract smaller and mid-sized clients, while keeping larger customers satisfied**

- **Invest more in developing Global Services and its people**

- **Retain existing experienced Domestic workforce**





- **Develop Analytics Further to Identify Improvements**
- **Attract and Retain In-store/Online Customers**
- **Ecommerce Startup Acquisition strategy**
- **Improve Mobile Analytics**



# Impact of Presidential Party on Fortune 500 Companies

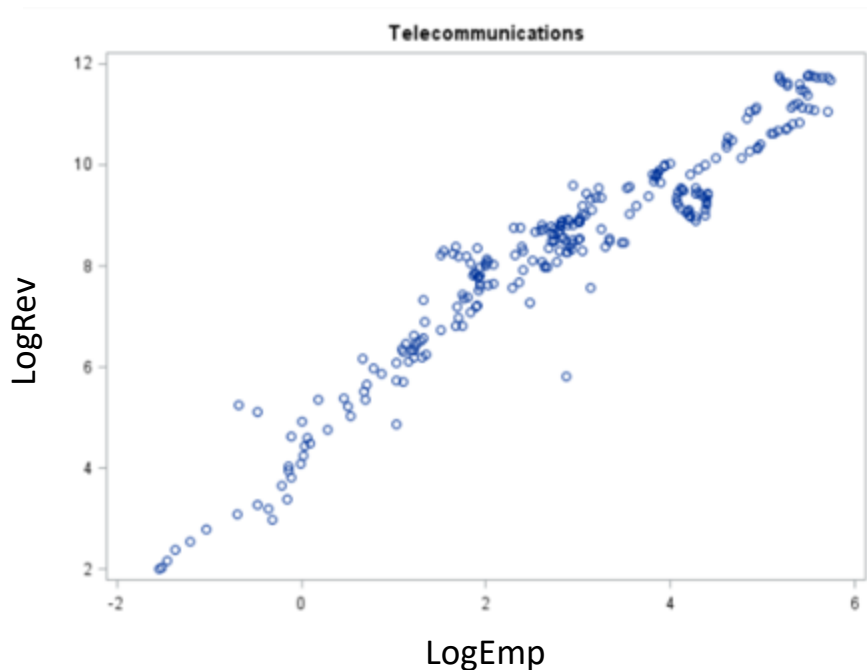
## **Republican:**

Aerospace and Defense  
Chemicals  
Commercial Banks  
Electronics, Electrical Equip.  
Energy  
General Merchandisers

...

## **Democrat:**

Hotels, Casinos, Resorts



$$\text{LogRev} = b + a_1 * \text{LogTime} + a_2 * \text{logEmp} + a_3 * \text{RIndicator}$$

# Thank You

## Questions/Comments

Email: [dravi@uncc.edu](mailto:dravi@uncc.edu)

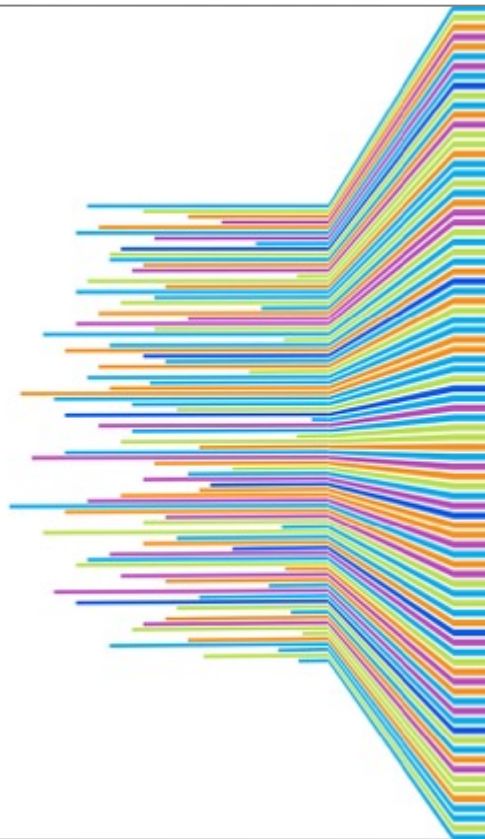
## Follow Me

Twitter @

## Rate This Session #

with the PARTNERS Mobile App

Remember To Share Your Virtual Passes





**ANALYTICS** CHALLENGE

# SESSION WRAP UP

# Student Poster Presentations

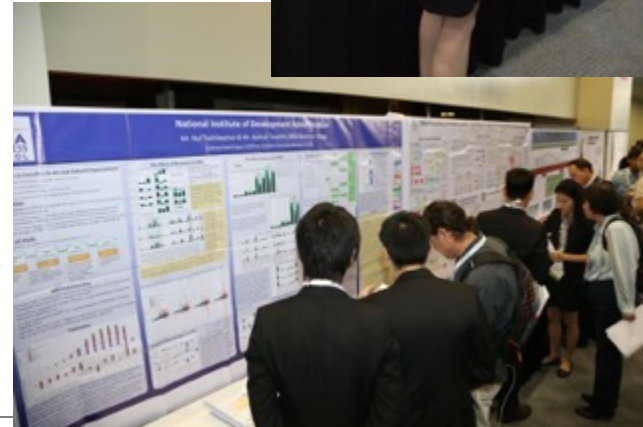
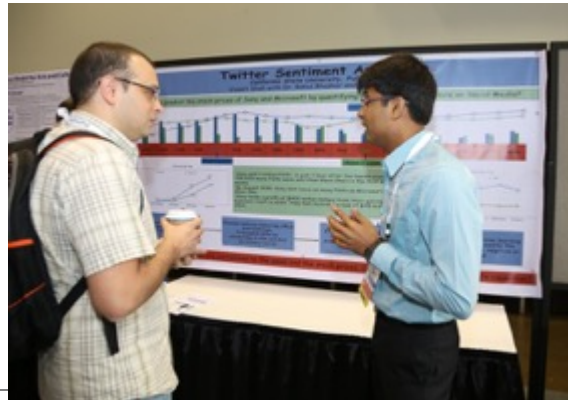
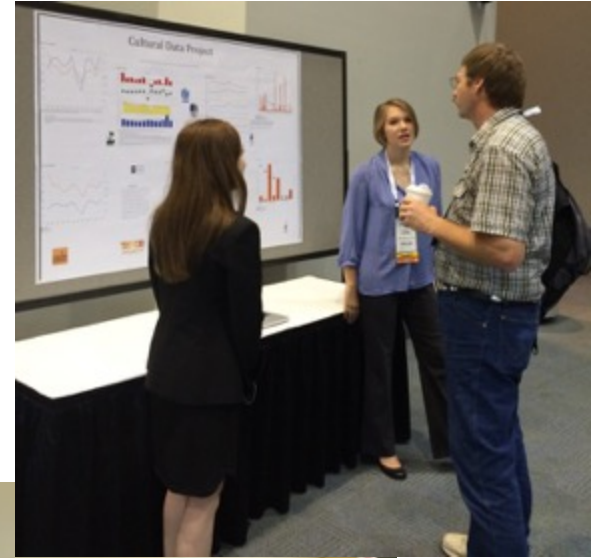
**Monday, September 12**

7:30-8:45 AM and 11:00 AM – 1:30 PM

**C HALL - LOBBY**

**Meet ALL 2016 Finalists from Both Challenges!**

**ALL TEAMS presenting prior to and after the opening session!**



# Student Competition Awards

## TUN ANALYTIC CHALLENGE

People's Choice - Best Presentation

Best Use of Analytics and Visualization

Overall Winner

Selected by

Attendees

Attendees

*TUN Selection  
Committee*

## TUN DATA CHALLENGE

People's Choice - Best Presentation

Most Value to Hire Heroes USA

Overall Winner

Selected by

Attendees



*TUN Selection  
Committee*





# CAST YOUR VOTES

## TUN ANALYTICS CHALLENGE

People's Choice - Best Presentation

Attendees vote

Best Use of Analytics and Visualization

Attendees vote

## TUN DATA CHALLENGE

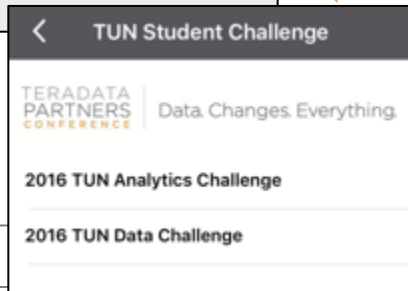
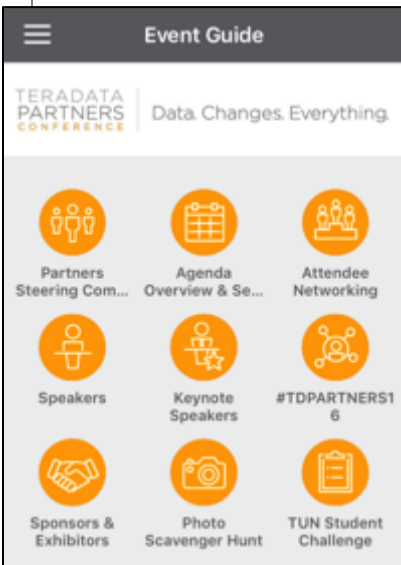
People's Choice - Best Presentation

Attendees vote

**CAST YOUR VOTES**

*on the PARTNERS Mobile App!*

**VOTING CLOSSES AT 2PM MONDAY**



2016 TUN Analytics Challenge	
2016 TUN Analytics Challenge	
A1 - Cal State Fullerton (Vogt)	
A2 - Loyola University Chicago (Patel)	
A3 - National University of Singapore (Tan)	
A4 - Oklahoma State Univ. (Molaka)	
A5 - UNC Charlotte (Ravi)	
A6 - UNC Charlotte (Naga)	
A7 - UNC Charlotte (Nirnanjan)	
A8 - UNLV(Girard)	
A9 - Univ. of Cincinnati (Samudrala)	
A10 - Univ. of Oklahoma(Myint)	
Voting: ( 2 Awards)	
Best Use of Analytics and Visualization	
People's Choice – Best Presentation	

2016 TUN Data Challenge	
Description	
2016 TUN Data Challenge	
D1 - Carnegie Mellon Univ. - Australia (Sanghvi)	
D2 - Loyola University Chicago (Vollan)	
D3 - Missouri Univ. of Science & Technology(Sen)	
D4 - NIDA – Thailand (Prateepvattanavit)	
D5 - UNCC (Sawant)	
D6 - UNCC (Withers)	
D7 - Waterloo Univ. (Lo)	
Voting: (1 Award)	
People's Choice – Best Presentation	

# TUN Student Celebration Event

TERADATA  
PARTNERS  
CONFERENCE

**Monday 6:30-9:30pm**

**Sheraton Hotel Atlanta – Capital North**

***All attendees are invited!***

*Wear your college colors & join us  
for a casual night of fun and excitement as we announce the*

**2016 Analytics Challenge  
&  
2016 Data Challenge  
WINNERS!**



# Thank You

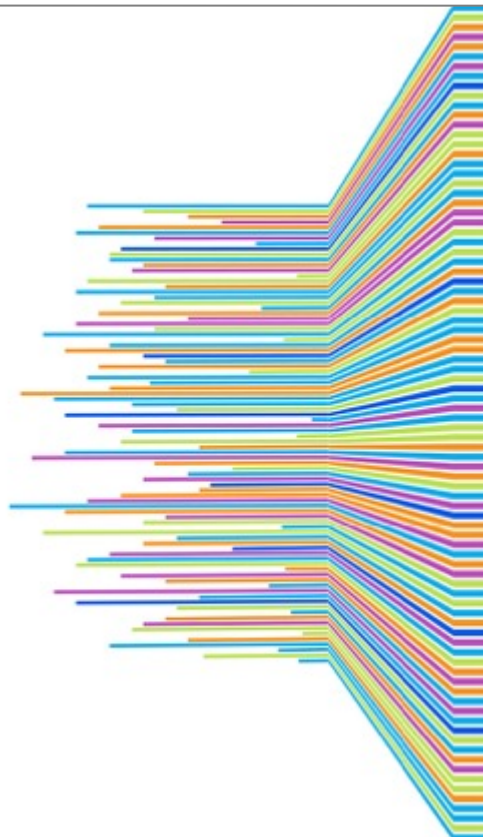
## Questions/Comments

Email: [Yenny.Yang@teradata.com](mailto:Yenny.Yang@teradata.com)

## Rate This Session # **550**

with the PARTNERS Mobile App

Remember To Share Your Virtual Passes



We empower companies to achieve  
**high-impact business outcomes**  
through analytics at scale  
on an agile data foundation