



# **City of Hope, Data Culture Mindset - Saving Lives and Solving Real-World Health Issues**

## **Facilitator**

Barry Chaiken, MD, Clinical Lead, Healthcare, Tableau

## **Lead Presenter**

James Lacey, Jr., Ph.D., Director, Division of Health Analytics, City of Hope



## **James Lacey, Jr., Ph.D.**

**Director of Division of Health Analytics, City of Hope**

- Leader in applying cloud computing, data commons and customer engagement to improve cancer research.
- Working at the intersection of epidemiology, data science, and population health informatics.
- Principal Investigator for the California Teachers Study exploring women's cancer.
- Studied causes of female reproductive cancers at National Cancer Institute.



## **Barry Chaiken, MD**

**Clinical Lead, Healthcare, Tableau**

- 25+ years of experience in medical research, epidemiology, clinical information technology, and analytics.
- Board-certified in general preventive medicine and public health and is a fellow, and former board member and chair of HIMSS.
- Experience in quality improvement studies, health IT clinical transformation projects, and clinical investigations.

# Observational Cohort Studies Are Essential in the Biomedical Research Enterprise



**Assemble a large group of participants**

**Collect diverse, real-world data at baseline and over years of follow-up**

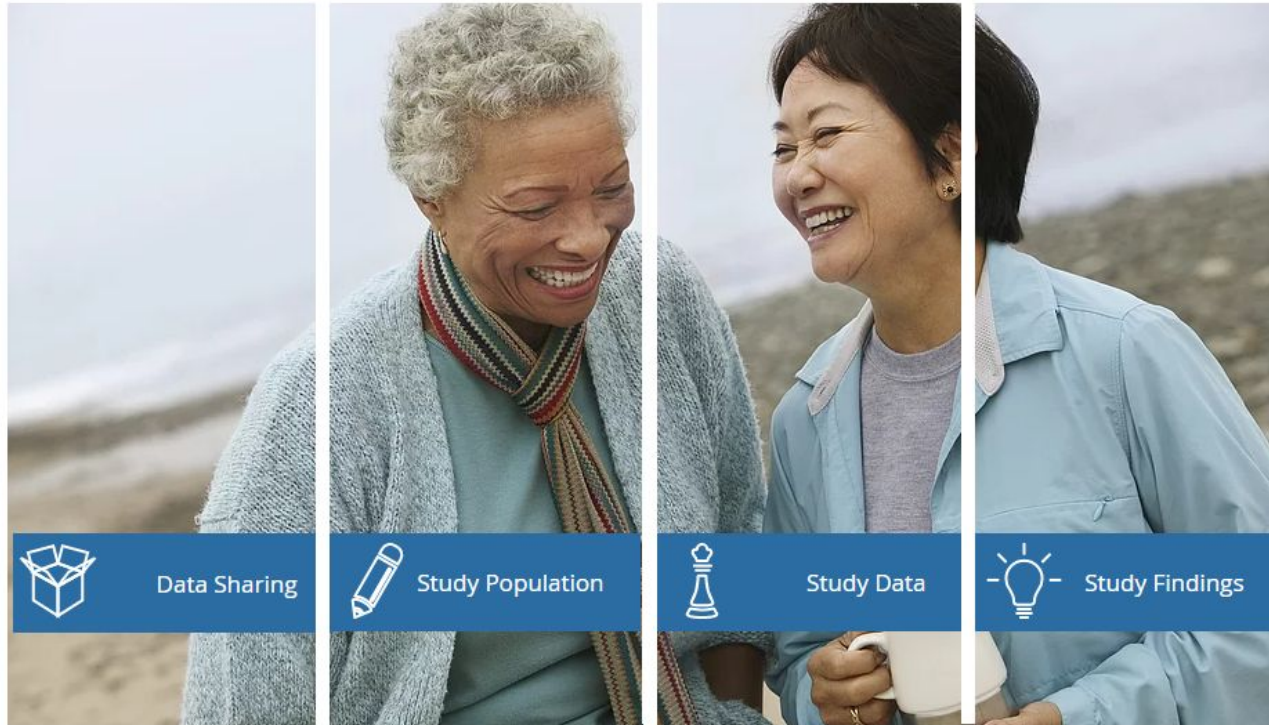
**Statistically analyze group differences to evaluate a broad range of exposures & outcomes**

# The California Teachers Study

[www.calteachersstudy.org](http://www.calteachersstudy.org)



[Home](#) | [About Us](#) | [For Researchers](#) | [For Participants](#) | [Blog](#)



With large, long-term studies like the California Teachers Study, we can learn a great deal about risk factors for cancer and what we can do as individuals and as Californians to control the disease.

[Learn More](#)

## Study Questionnaires

Participants have been asked to complete [five self-reported questionnaires](#), which collected data on:

- Anthropometric measurements
- Cancer screening and imaging
- Children's and family health history
- Diet and supplements
- Medical conditions and overall health
- Personal care
- Physical activity
- Prescription and over-the-counter meds
- Reproductive factors and menopausal status
- Residential history
- Sleep
- Smoking and second-hand smoke
- Stress and social support



UC San Diego



# California Teachers Study Endpoints

Baseline: 1995-1996

Today



**MORTALITY: Dates and Causes of Death**

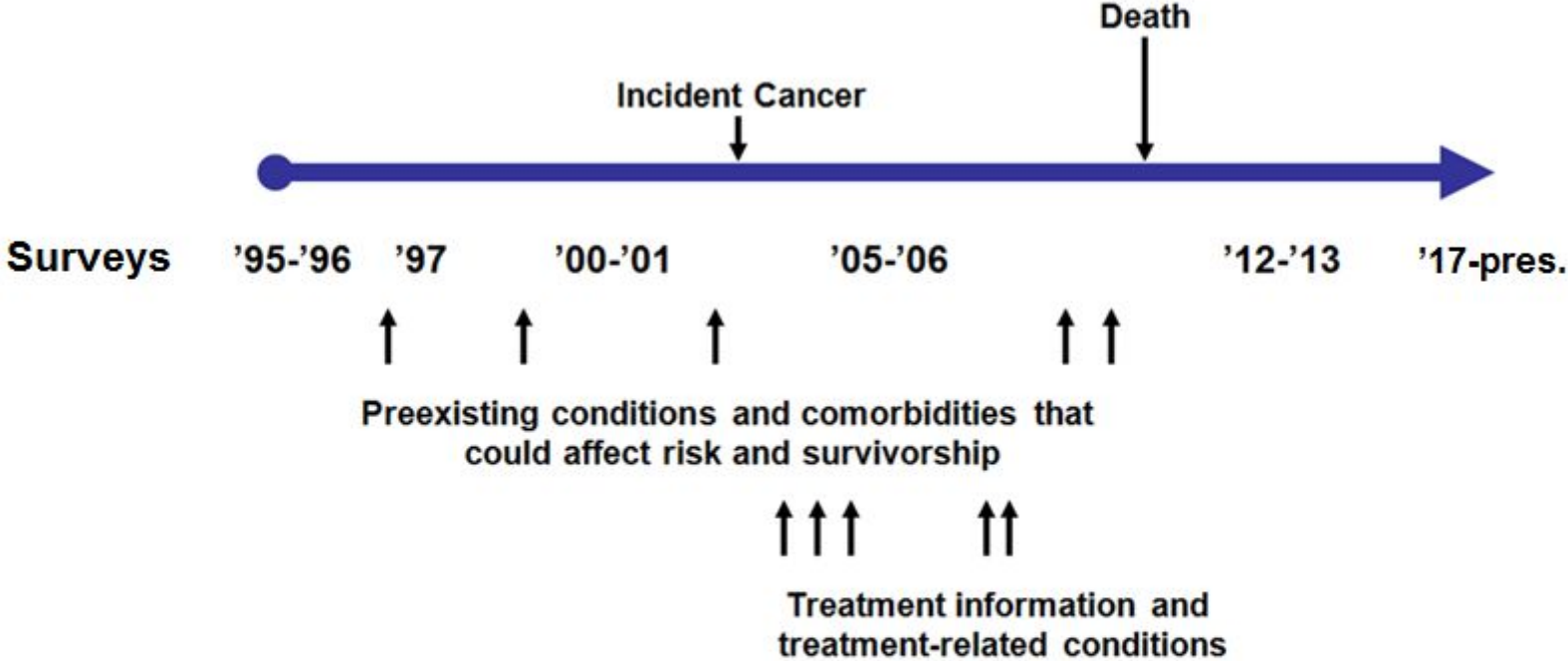
**CANCER: Incident Cancers and Details: Self-report, records, & registry**

**INPATIENT HOSPITALIZATION:**

*Inpatient hospitalizations (up to 25 ICD & DRG codes)*  
*Hospital-based procedures (up to 21 ICD codes)*  
*Other conditions present at diagnosis (up to 5 ICD codes)*

**Ambulatory Surgery & ER Visits**  
*Cause of injury (up to 5 ICD codes)*  
*Procedures (up to 5 CPT codes)*

# A Participant Journey



# Recently Launched CTS Data Analysis Projects



- Light at night and Non Hodgkin Lymphoma
- Light at night and hypertension & diabetes
- Air pollution and survival after breast cancer
- Sleep quality during follow-up and risk of multiple cancers
- Sugar-sweetened beverages and CVD, colorectal cancer, & death
- Dietary patterns and risk of premenopausal breast cancer
- Dietary patterns and risk of chronic kidney disease
- Rice consumption & risk of multiple cancers
- Census-tract educational levels and non-smoking lung cancer

# Recently Launched CTS Data Analysis Projects



## During 25 years of CTS follow-up ...

- Approx. 180,000,000 survey data points
  - Over 32,000 cancers
  - Over 30,000 deaths
- Over 700,000 hospitalization records
  - Biospecimens
- Geospatial capabilities





NCI Cohort Consortium

<https://epi.grants.cancer.gov/Consortia/cohort.html>

Advancing Collaborative Cancer Epidemiology Research

59  
Epidemiology  
Cohorts

>7  
Million  
Participants

Global

North America\* (38 cohorts)  
Asia (7 cohorts)  
Europe (13 cohorts)  
Australia\* (1 cohort)

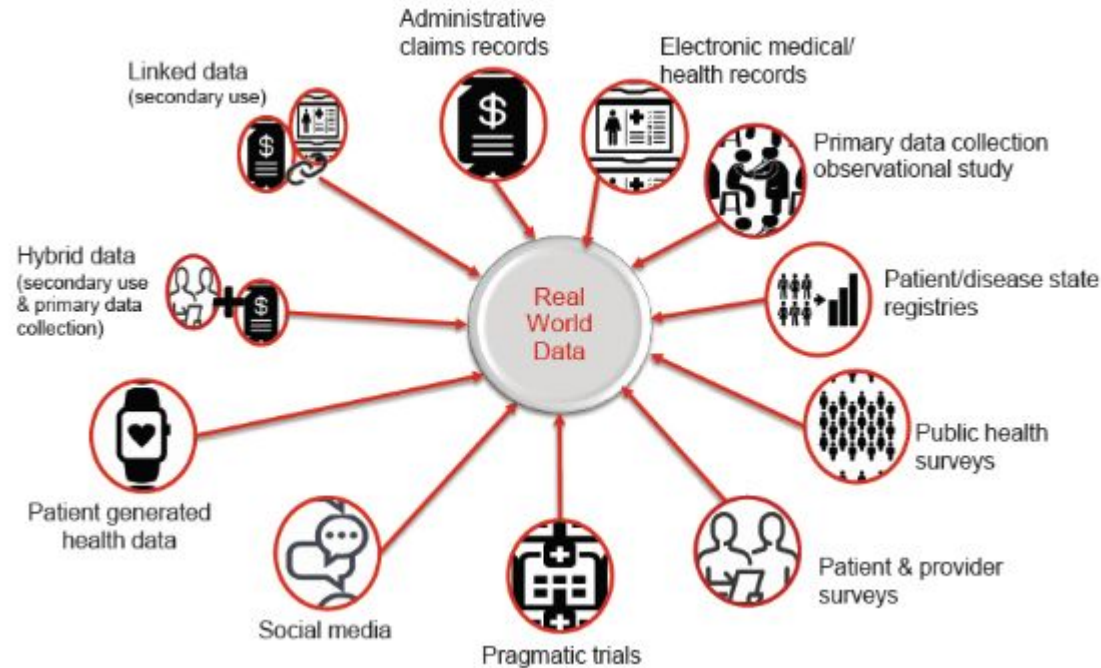
\* Two cohorts include study participants from both North America and Australia.

Biospecimens

- Plasma/Serum
- Saliva/Buccal Cells
- Urine
- Buffy Coat/Whole Blood
- Tissues
- Nails

have been collected on approximately 2 million individuals

# “Same thing only different ...”



## Real World Data

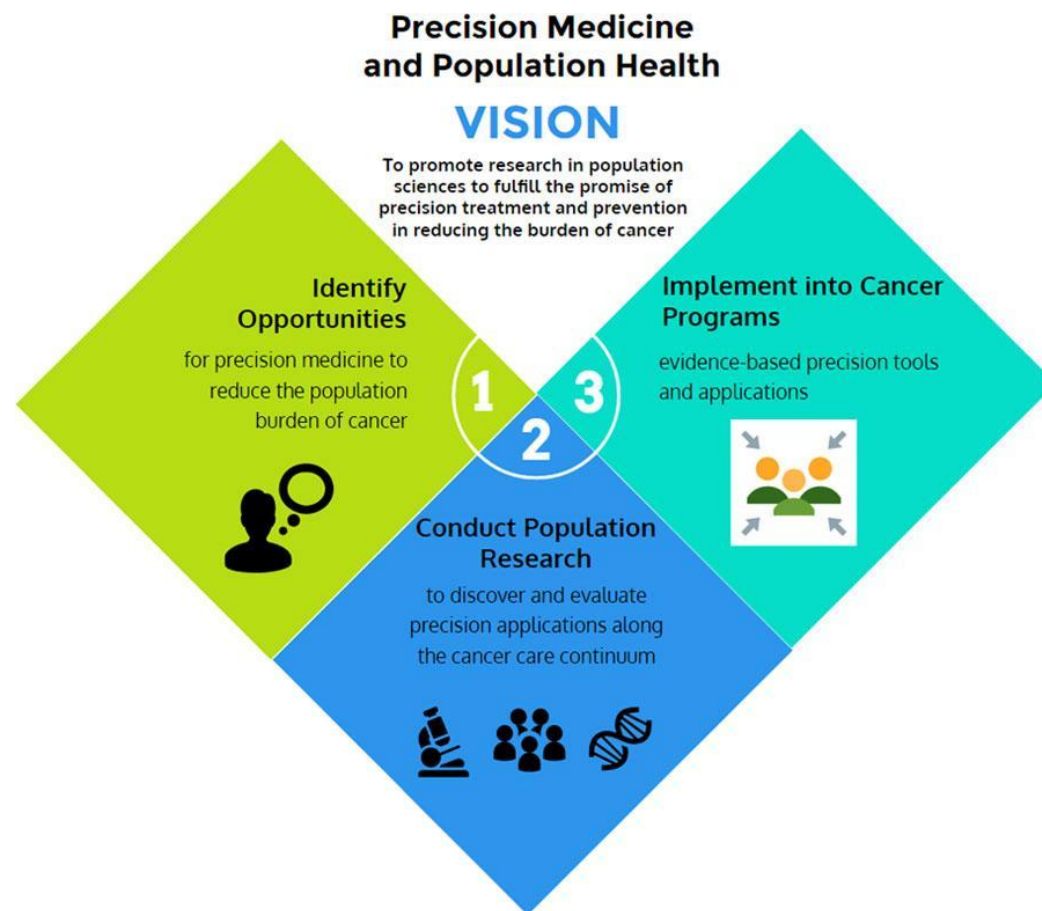
*National Academies Press 2019: Examining the Impact of Real-World Evidence on Medical Product Development: Proceedings of a Workshop Series*



## Social Determinants of Health

*DHHS/Office of Disease Prevention and Health Promotion: Healthypeople.gov/2020*

# Precision Medicine and Population Health Research



<https://cancercontrol.cancer.gov/research-emphasis/precision-medicine.html>

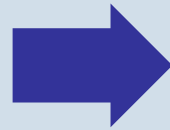
# How Does the CTS Do This?

*At every stage of this process we use*



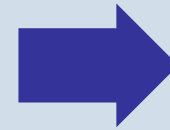
**U.S. Taxpayers pay us to build & maintain CTS data.**

- *We collect, combine, & standardize data.*



**Researchers propose ideas & use CTS data.**

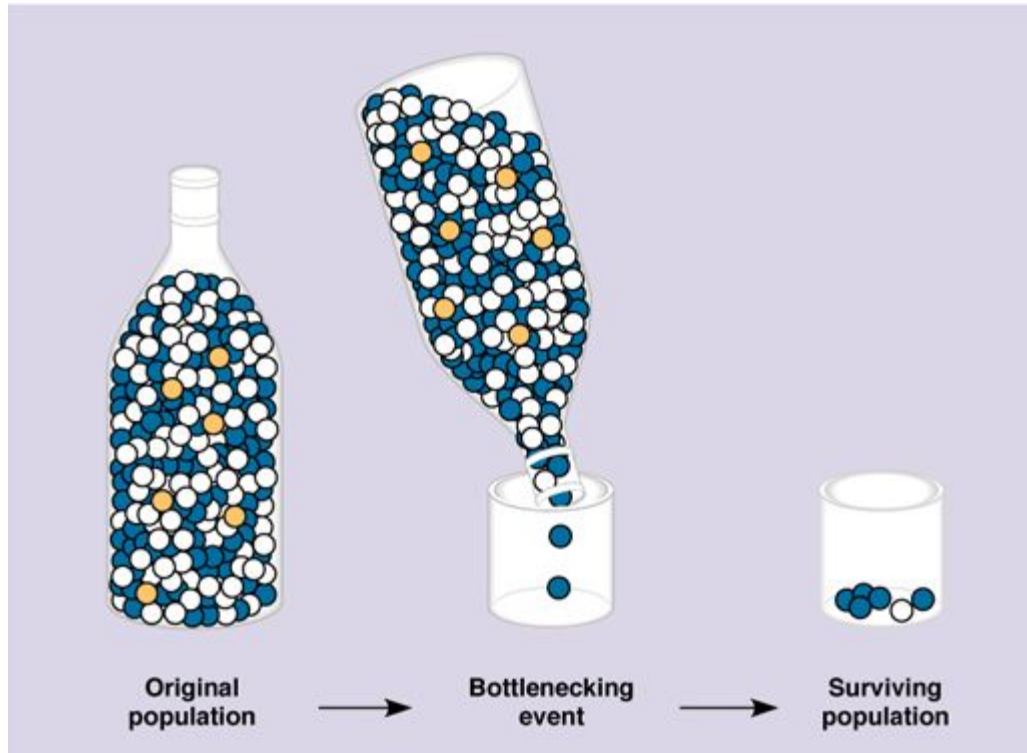
- *“I hypothesize that exposure to X affects risk of cancer Y...”*



**Generate new discoveries & publish results.**

- *The value of these studies is their broad & deep data.*

# The Bottleneck Challenge Facing Cohorts Like the CTS

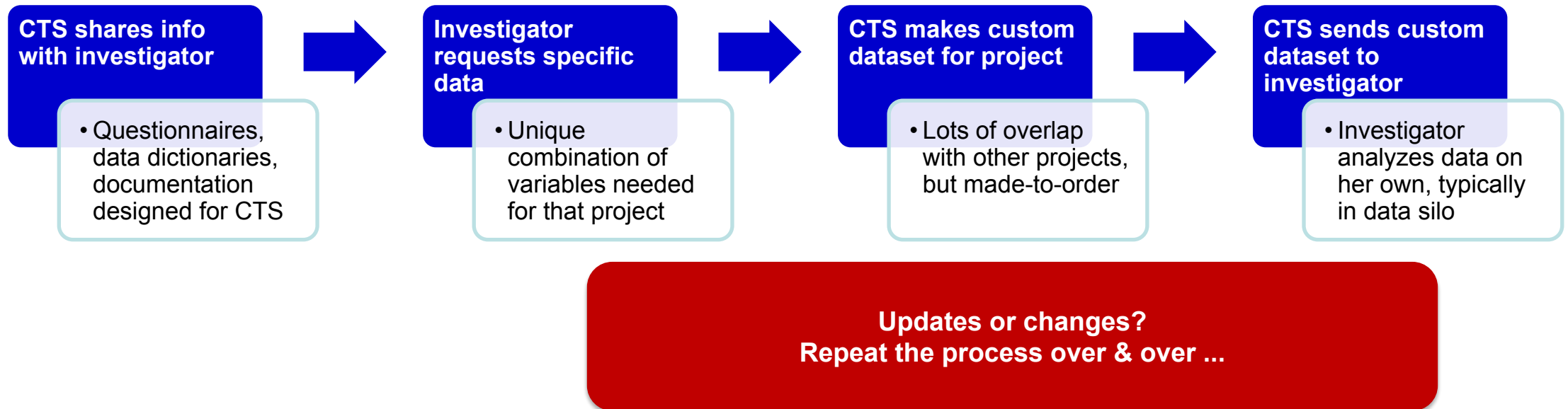


No one ever analyzes all cohort data at once.

Before: *Create bespoke data for every project.*

Now: *Provide scalable self-service.*

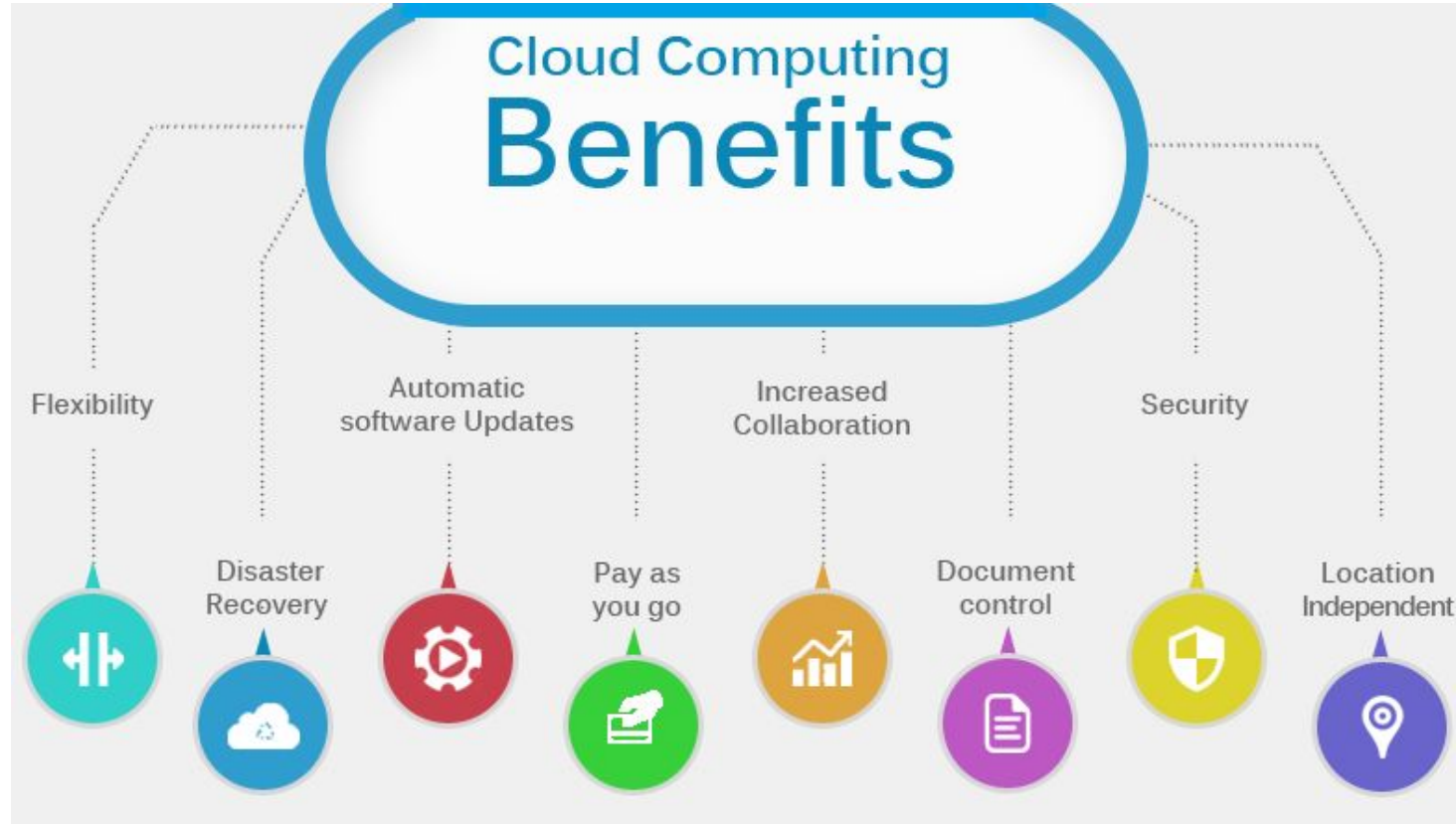
# Data Bottlenecks & Process Bottlenecks



# Voice of the Frustrated Researcher...



# We Needed to Bend our Data Curve

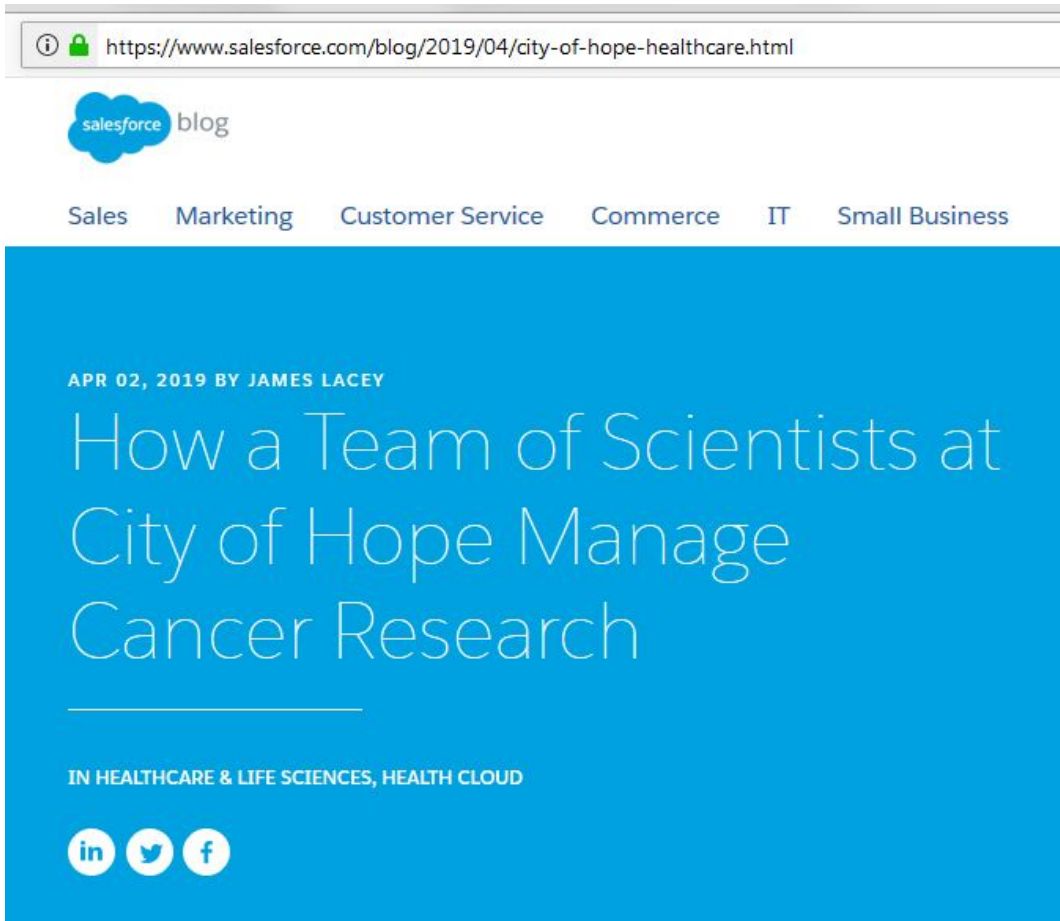


Spend less time *wrestling with & looking for* data.

Spend more time *generating insights from* data.



# An Eye-Opening Win-Win



The screenshot shows a web browser window with the URL <https://www.salesforce.com/blog/2019/04/city-of-hope-healthcare.html>. The page header includes the Salesforce logo and the word "blog". Below the header are navigation links for Sales, Marketing, Customer Service, Commerce, IT, and Small Business. The main content area has a blue background with the text "APR 02, 2019 BY JAMES LACEY" and the title "How a Team of Scientists at City of Hope Manage Cancer Research". Below the title is a horizontal line and the text "IN HEALTHCARE & LIFE SCIENCES, HEALTH CLOUD". At the bottom left of the blue area are three circular social media icons for LinkedIn, Twitter, and Facebook.

[salesforce.com/blog/2019/04/city-of-hope-healthcare.html](https://www.salesforce.com/blog/2019/04/city-of-hope-healthcare.html)

I wasn't aware of a research solution that would suit our needs. However, a member of the team had recently attended a dinner party in the Bay Area. She told the group about CTS and mentioned how my team at City of Hope was looking to modernize our data collection process. She explained our challenge and one person suggested we look into Salesforce.

Why not? Salesforce helps companies successfully manage and connect with customers in other industries. It should be able to help us manage and connect with CTS participants.

With Salesforce, we've consolidated processes that previously took weeks, even months, into a consistent process that can take minutes.

Now, instead of wrangling data and wrestling with paper forms, we can focus more on our cancer research and our interactions with our research volunteers. We can dive deeper into the data to uncover new findings and share and inspire others in the field of epidemiology and beyond.

"Why not?" For my team at City of Hope, asking and answering that question made all the difference.

# TECHNOLOGY PLATFORM

Best of Class Enterprise & Open Source Products

**Storage:** NetApp, Dell  
(File, Block, Database)

**Compute:** Commodity Hardware  
(Dell, IBM, etc.)

**Virtualization:** VMWare

**OS Support:** Windows and Linux



**SHERLOCK**  
SOLVED.

**Shared Services:** Backups (CommVault),  
Authentication (RSA, Active Directory),  
Configuration Mgmt. (Windows GP, Chef),  
Monitoring (Nagios), Managed File Transfer  
(GlobalScape), Secure Access (HW and SW VPNs),  
Anti Virus (Sophos), Certificate Authority,  
Account Auditing (Netwrix), Patch Management  
(SolarWinds)

**Security:** Firewall Segmentation (Juniper, Cisco,  
Vyatta, Sophos), IDS/IPS (Bro, Snort), Auditing  
and Log Management (Splunk), Vulnerability  
Scanning (Nessus), Web Application Firewall (F5)

# Academic Research & a Data Culture



- **Biomedical research can often ...**
  - **discourage trust**
  - **favor people over ideas**
  - **make new talent wait too long**
  - **discourage sharing, and**
  - **promote status quo rather than data**
- **Our use-cases drive our transition to a data culture**
  - **Commitment to common challenges we need to solve**
  - **Talented people who needed new tools**
  - **Deploying & sharing changes the mindset & builds trust**

# Top Priorities: Enable Better CTS Research

## Original CTS Objectives

- Better storage: *all data in one place*
- More consistency: *single source of data*
- Fresher data: *automated updates*
- More access: *everyone sees same data*
- More secure: *reduce data moving*
  
- Faster starts: *repeatable processes*
- Easier collaboration: *shared workspaces*
- More sharing: *actually share same data*

“Great.

How will you help people  
consume & use your data?”

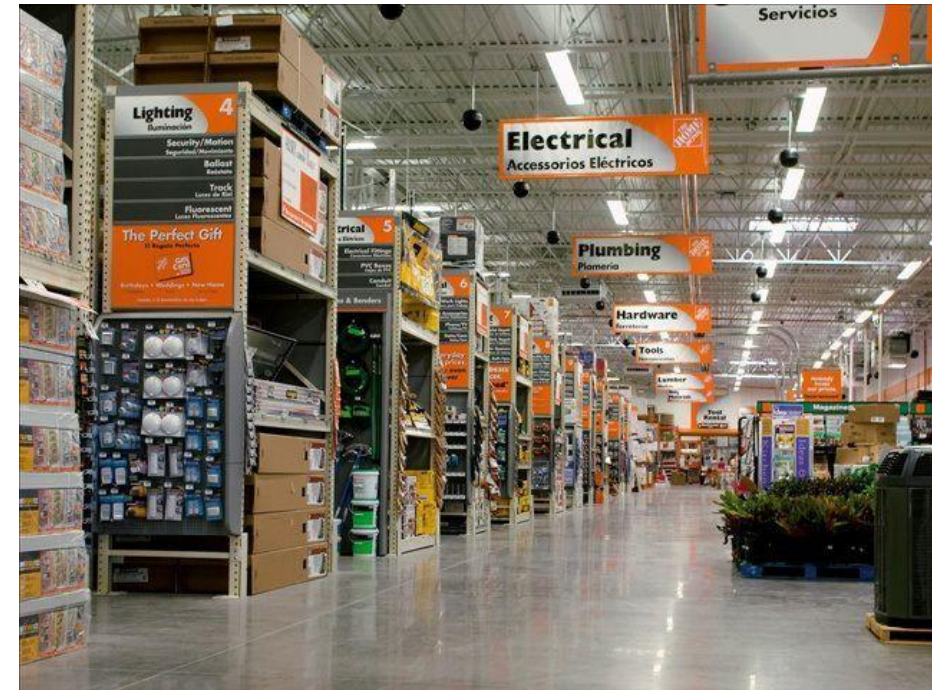
# Our Approach: Think Like a Hardware Store...

*Most trips to the hardware store fall into one of two categories:*

*I think I know what I want, but I need help. Find smart staff person.*



*I have a job to do. I know what I need. Get in, get out.*



# Our Approach: Think Like a Hardware Store...

Current challenge for research-enabling observational cohort studies:

*Provide knowledgeable customer-service-associates when needed.*

*Organize our data in ways that enable self-service.*



# Standardized, Efficient, and Flexible Support for all Types of Observational Study Activities

Explore

QC

Analyze

Share

Before:

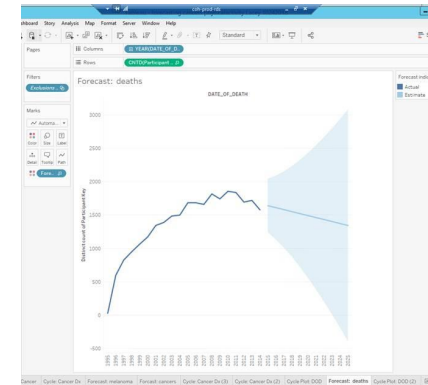
Static PDFs

## CALIFORNIA TEACHERS STUDY USEFUL NUMBERS

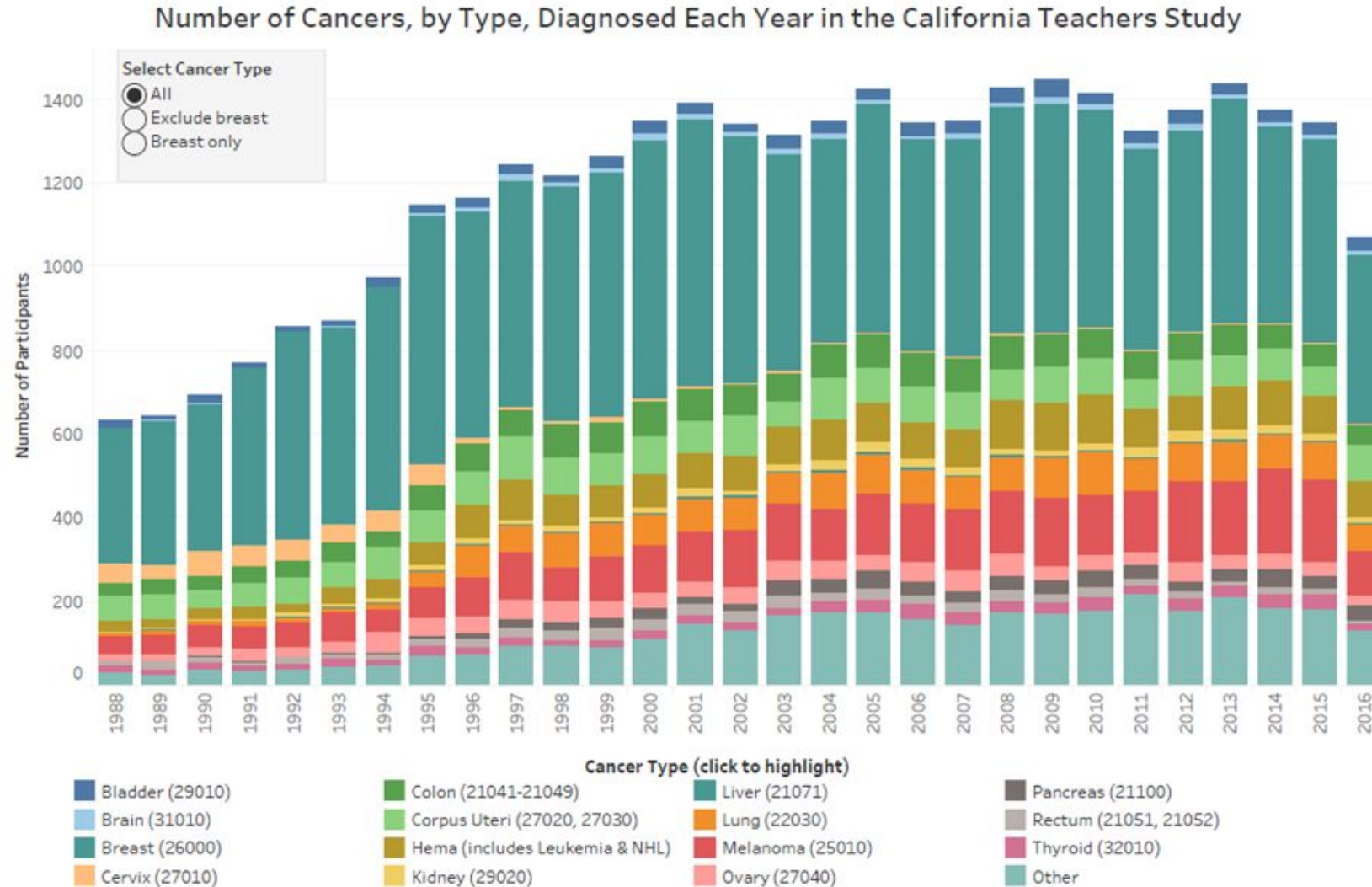
Top 50 Principal Hospital Procedures, OSHPD 1995-2012

Principal Procedure	Count	Percent
8154, TOTAL KNEE REPLACEMENT	9,321	4.4
8151, TOTAL HIP REPLACEMENT	5,480	2.6
7359, MANUAL ASSIST DELIV NEC	3,950	1.9
684, TOTAL ABD HYSTERECTOMY	3,739	1.8
9904, PACKED CELL TRANSFUSION	3,738	1.8
741, LOW CERVICAL CSECTION	3,642	1.7
9339, PHYSICAL THERAPY NEC	3,405	1.6
8872, DX ULTRASOUNDHEART	2,545	1.2
5123, LAPAROSCOPIC CHOLECYSTEC	2,517	1.2
6859, OTHER VAG HYSTERECTOMY	2,109	1.0
3722, LEFT HEART CARDIAC CATH	2,083	1.0
7569, REPAIR OB LACERATION NEC	2,079	1.0
8703, CAT SCAN OF HEAD	1,992	0.9
4516, EGD WITH CLOSED BIOPSY	1,948	0.9
7935, OPEN REDUCINT FIX FEMUR	1,937	0.9
3893, VENOUS CATH NEC	1,916	0.9
8152, PARTIAL HIP REPLACEMENT	1,674	0.8
736, EPISIOTOMY	1,621	0.8
8543, UNILAT EXTEN SIMP MASTEC	1,338	0.6
4513, SM BOWEL ENDOSCOPY NEC	1,325	0.6
0309, SPINAL CANAL EXPLOR NEC	1,285	0.6
7936, OP REDINT FIX TIB/FIBUL	1,208	0.6
3491, THORACENTESIS	1,164	0.6

All-in on data visualization



# Accelerate Research by Facilitating Data Exploration



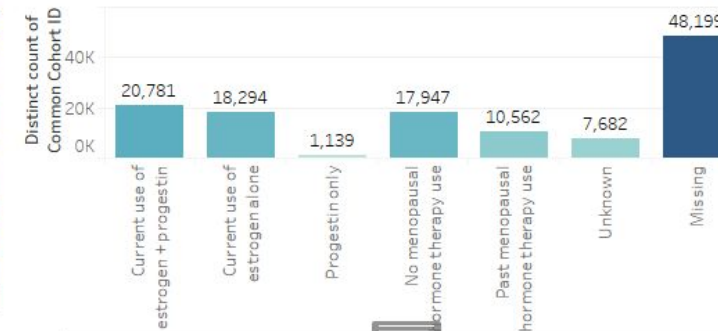


# Accelerate Research by Facilitating Data Exploration

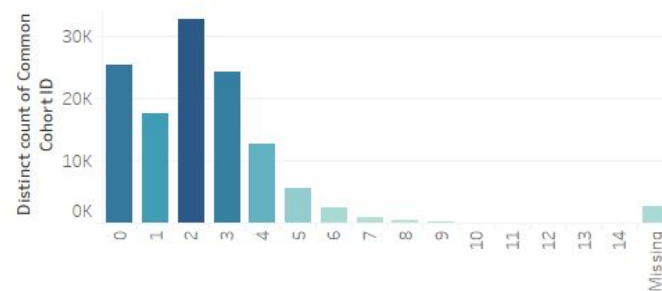
Menopausal status at baseline

Premenopausal	48,199
Peri-menopausal	2,519
Postmenopausal due to medication/chemo/radiation	1,489
Postmenopausal due to other reason	4,002
Postmenopausal unknown type of menopause	1,764
Postmenopausal: bilateral oophorectomy	10,425
Postmenopausal: hysterectomy at age 56+	156
Postmenopausal: hysterectomy at age <56 but aged 56+ at baseline	8,063
Postmenopausal: natural menopause	37,240
Unknown menopausal status	10,747
<b>Grand Total</b>	<b>124,604</b>

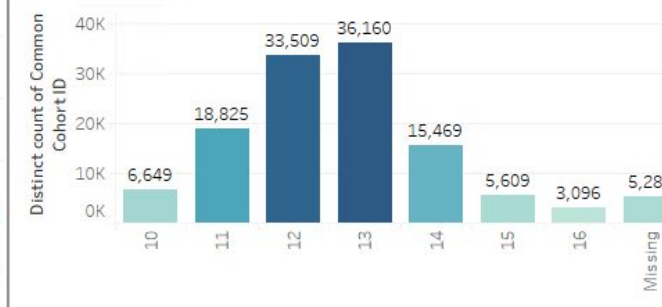
Hormone therapy use



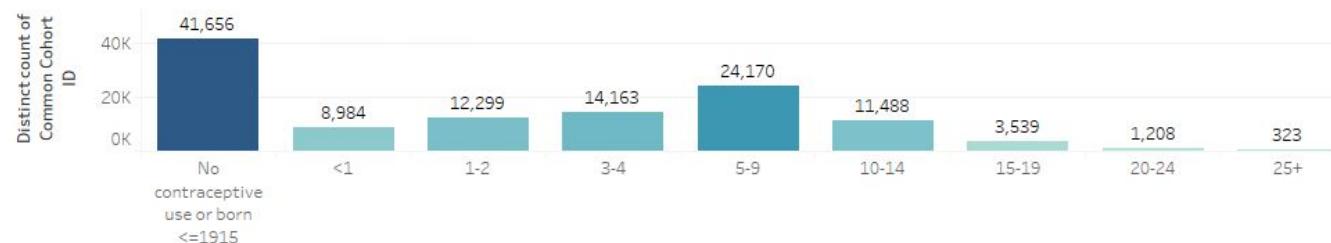
Number of pregnancies at baseline



Age at menarche



Years taking oral contraceptives



# Standardized, Efficient, and Flexible Support for all Types of Cohort Data Activities

Explore

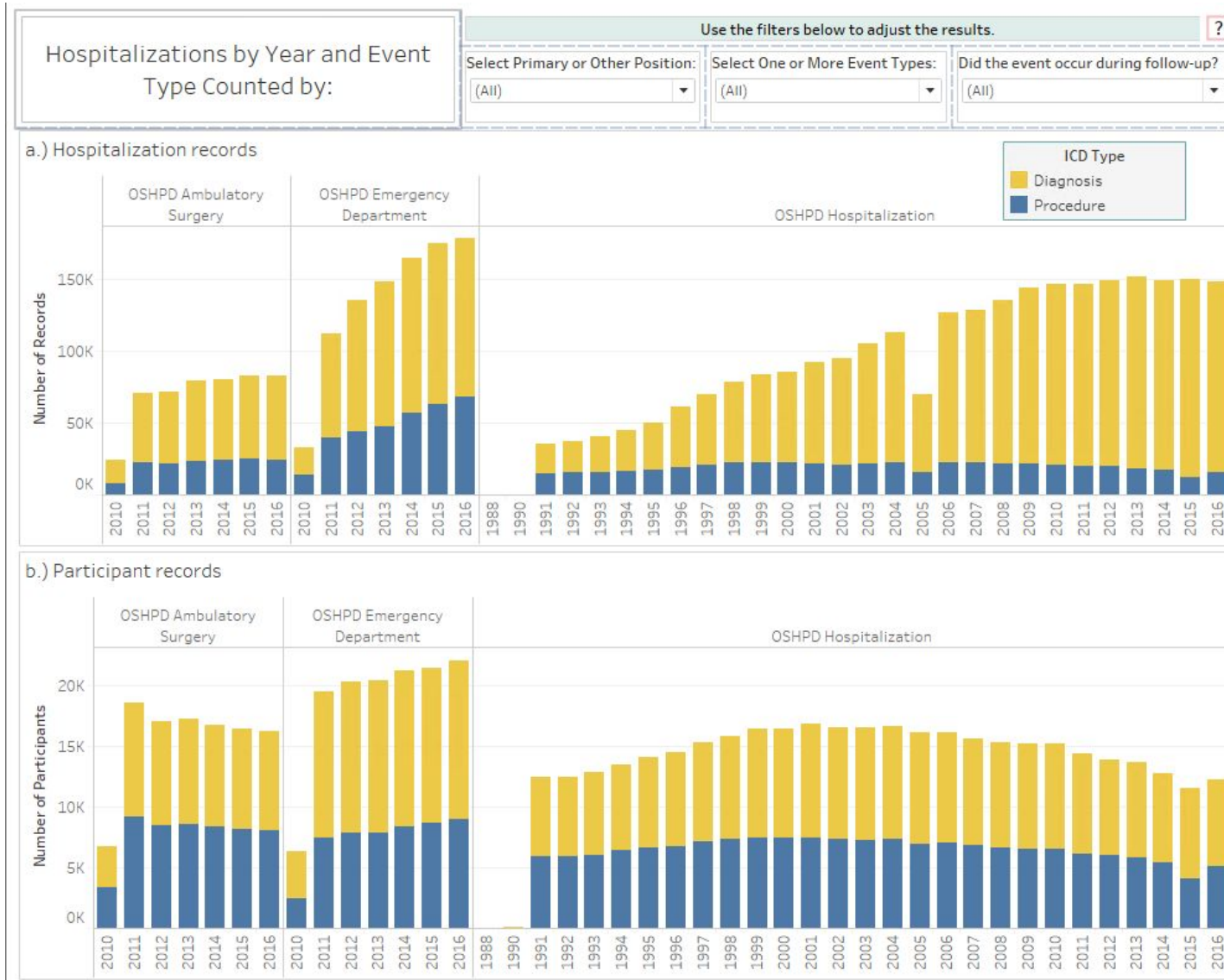
QC

Analyze

Share

**Visually inspect all new data sources for quality, coherence, and patterns**

# Quality Control via Data Inspection



# Tell Data Stories to Facilitate Data Analysis

## Eligibility Criteria

The original CTS population that enrolled at baseline & customary exclusions for analyses	Informed consent: for all CTS research, or just breast cancer research?	Did participants live in California or outside California at baseline?	<b>Baseline population for analyses of all cancer and hospitalization endpoints</b>	Exclude participants who were cancer survivors at baseline?	Analysis censoring date
---	---	--	---	---	-------------------------

124,604

Analyses that rely on the California Cancer Registry for cancer endpoints or the California Office of Statewide Health Planning and Development for hospitalization endpoints can use this population of 124,604 as the number of CTS participants who were eligible for follow-up at baseline.

# Simplify the Complex

View Hospitalization Data by ICD Code

Hover here for definitions

**1. Start Here!**  
 Select your ICD code(s) of interest by checking the boxes below. To search, use the magnifying glass at the top right.

Select your ICD code(s) of interest:

- 25030
- 25031
- 25032
- 25033
- 25040
- 25041
- 25042
- 25043
- 25050
- 25051
- 25052
- 25053
- 25060
- 25061
- 25062
- 25063
- 25070
- 25071
- 25072
- 25073
- 25075
- 25080
- 25081
- 25082
- 25083
- 25085
- 25090
- 25091
- 25092
- 25093

**2. View the CCS code, description and frequency in the CTS population of your chosen ICD code(s):**

### ICD Codes and Descriptions

Ccs Single Cat Nme	ICD CDE	ICD Code Description	
Diabetes mellitus without complication	25000	Diabetes Mellitus Without Mention Of Complication, Type Ii Or Unspecified Type, Not Stated As Uncontrolled	10,397
	25001	Diabetes Mellitus Without Mention Of Complication, Type I [Juvenile Type], Not Stated As Uncontrolled	841
Diabetes mellitus with complications	25002	Diabetes Mellitus Without Mention Of Complication, Type Ii Or Unspecified Type, Uncontrolled	795
	25070	Diabetes With Peripheral Circulatory Disorders, Type Ii Or Unspecified Type, Not Stated As Uncontrolled	474

**3. View the Event Type, ICD Type, Position and ICD Version for your chosen ICD code(s):**

### Event Types and ICD Code Information

EventType_Description	Code Position Description	Icd Version Nme	
OSHPD Ambulatory Surgery	Primary Diagnosis	ICD9	4
	Other Diagnosis	ICD9	2,338
	Primary Procedure	HCPCS	63
	Other Procedure	HCPCS	13
OSHPD Emergency Department	Primary Diagnosis	ICD9	107
	Other Diagnosis	ICD9	3,492
OSHPD Hospitalization	Primary Diagnosis	ICD9	381
	Other Diagnosis	ICD9	8,597
Grand Total			11,029

**Click-lists and menus for over 200,000 ICD codes and 700,000 hospital visits**

# Lower Barriers to Entry: Geospatial Data

Geospatial workflow standardizes process of linking external, publicly available data to:  
 the correct CTS participants,  
 at the correct point in time, &  
 at the correct residence location.

https://data.ca.gov/dataset/poverty-rate-defined-us-census-california-regions/resource/6561849c-  
 CA CALIFORNIA OPEN DATA PORTAL

Datasets Topics Groups About State Portals Civic Engagement Documentation Log in

Home / Datasets / Poverty Rate (defined by U.S. Census) by California Regions  
 / Poverty Rate (defined by U.S. Census) by California Regions, 2000-2010 (CSV)

View published Back to dataset Download

Poverty Rate (defined by U.S. Census) by California Regions, 2000-2010 (CSV)

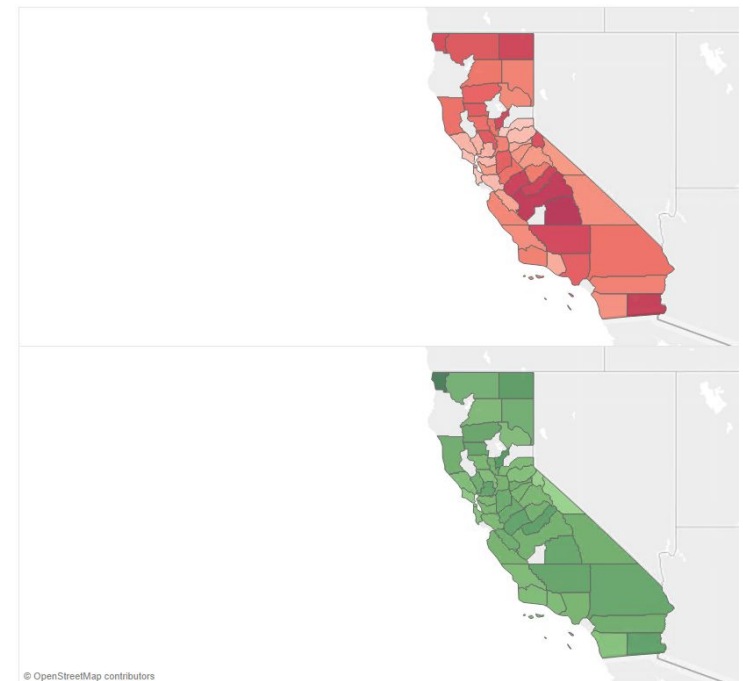
poverty-rate-defined-by-u.s.-census-by-california-regions-2000-2010.csv

Data Preview: Note that by default the preview only displays up to 100 records. Use the pager to flip through more records or adjust the start and end fields to display the number of records you wish to see.

Grid Graph Map 136242 records 1 - 100 Search data ... Go Filters Fields

ind_id	ind_d...	repor...	race_...	race_...	geotype	geoty...	geona...	count...	count...	regio...	regio...	Povert
754	Overa...	2006-...	1	AIAN	CA	6	Califo...			Conce...	30379	
754	Overa...	2006-...	2	Asian	CA	6	Califo...			Conce...	491802	
754	Overa...	2006-...	3	Africa...	CA	6	Califo...			Conce...	441447	
754	Overa...	2006-...	4	Latino	CA	6	Califo...			Conce...	26777...	
754	Overa...	2006-...	5	NHOPI	CA	6	Califo...			Conce...	16438	

RR to rest of CA and BMI at baseline



# Standardized, Efficient, and Flexible Support for all Types of Cohort Data Activities

Explore

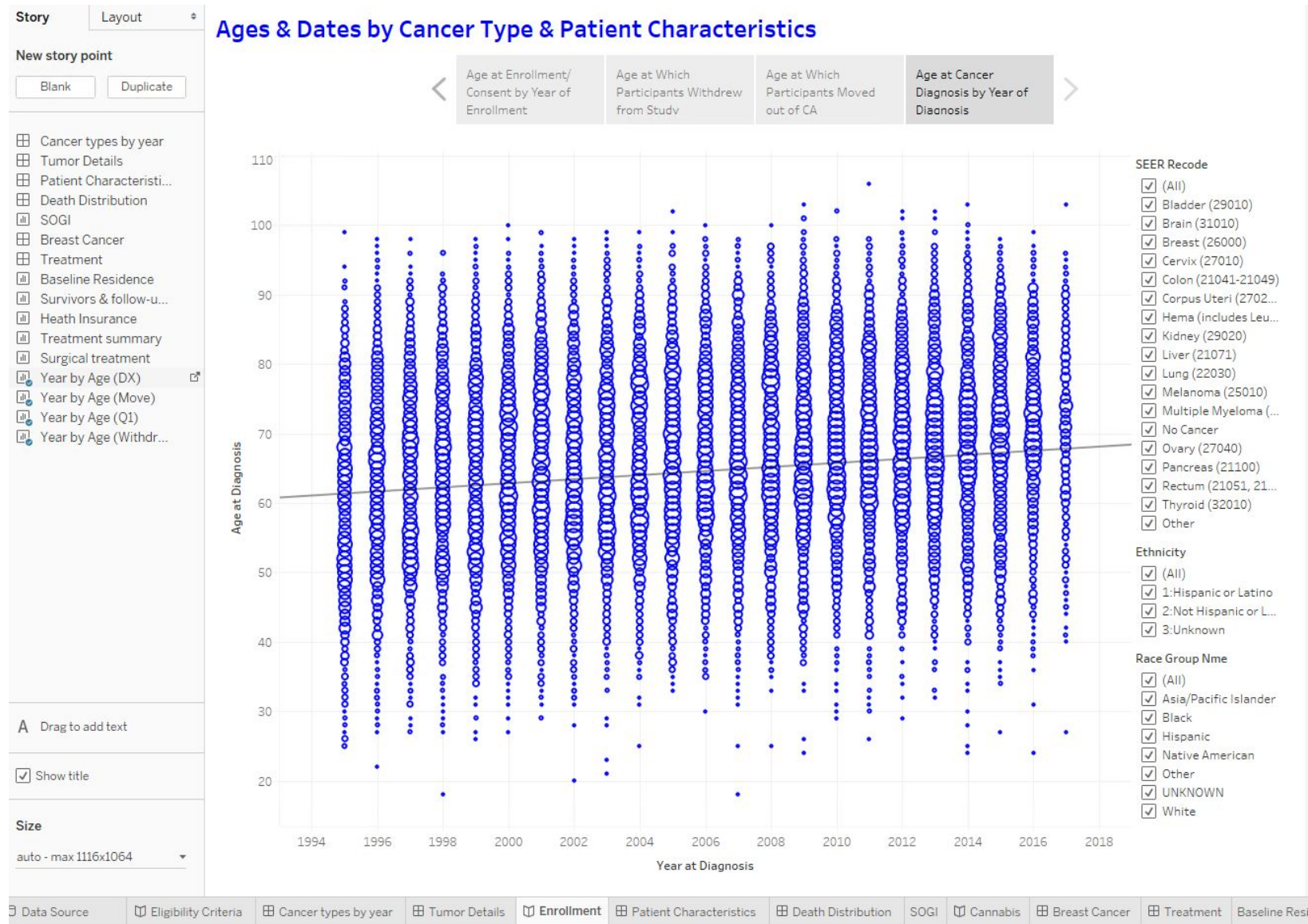
QC

Analyze

Share

**Find ways to provide fit-for-purpose data through scalable and reusable workflows and tools**

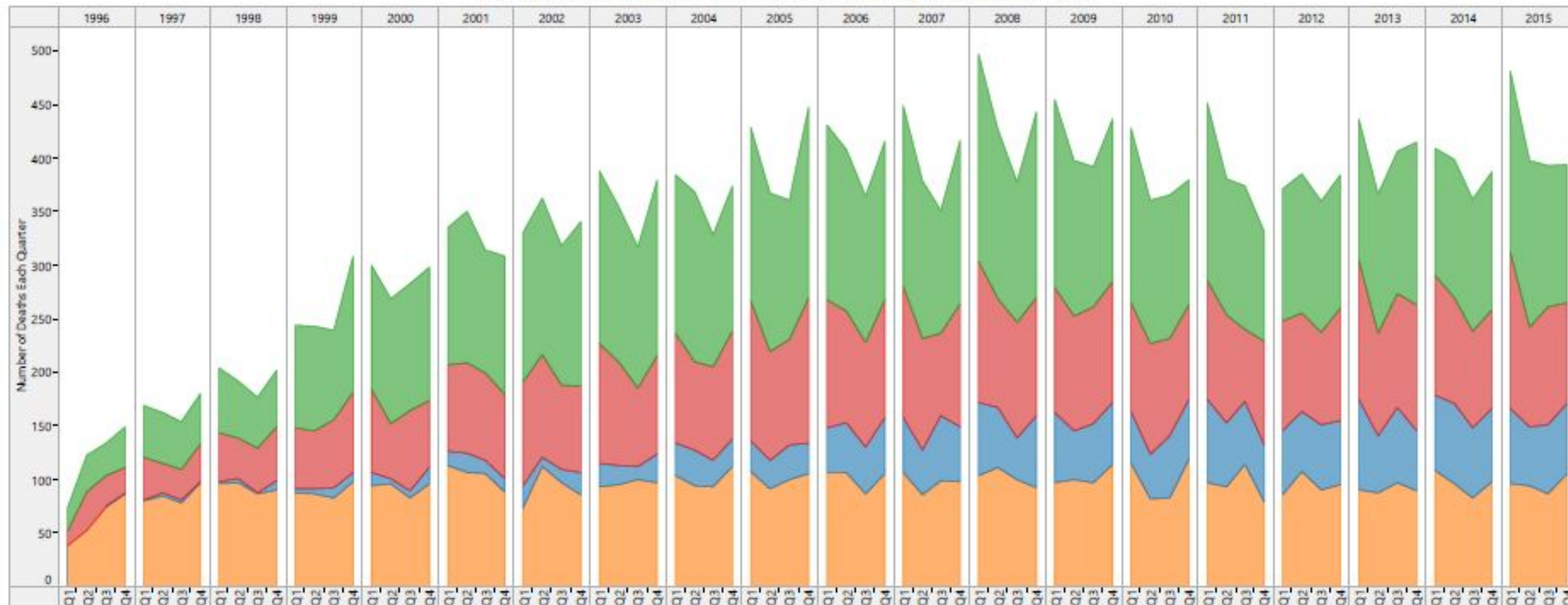
# Disease-Specific Registries vs. Scalable Platform





# Identify New Priorities & Opportunities

*Do More Women Die in the First Few Months of the Year? If So, Why?*



Deaths due to **Heart Disease** and **All Other Causes** are regularly higher in Q1 & Q4 than in Q2 & Q3.

Deaths due to **Dementia** and **Cancer** are more constant from quarter to quarter.

Temporal trends in deaths due to **Heart Disease** and **All Other Causes** represent a "Healthy Participant Bias."

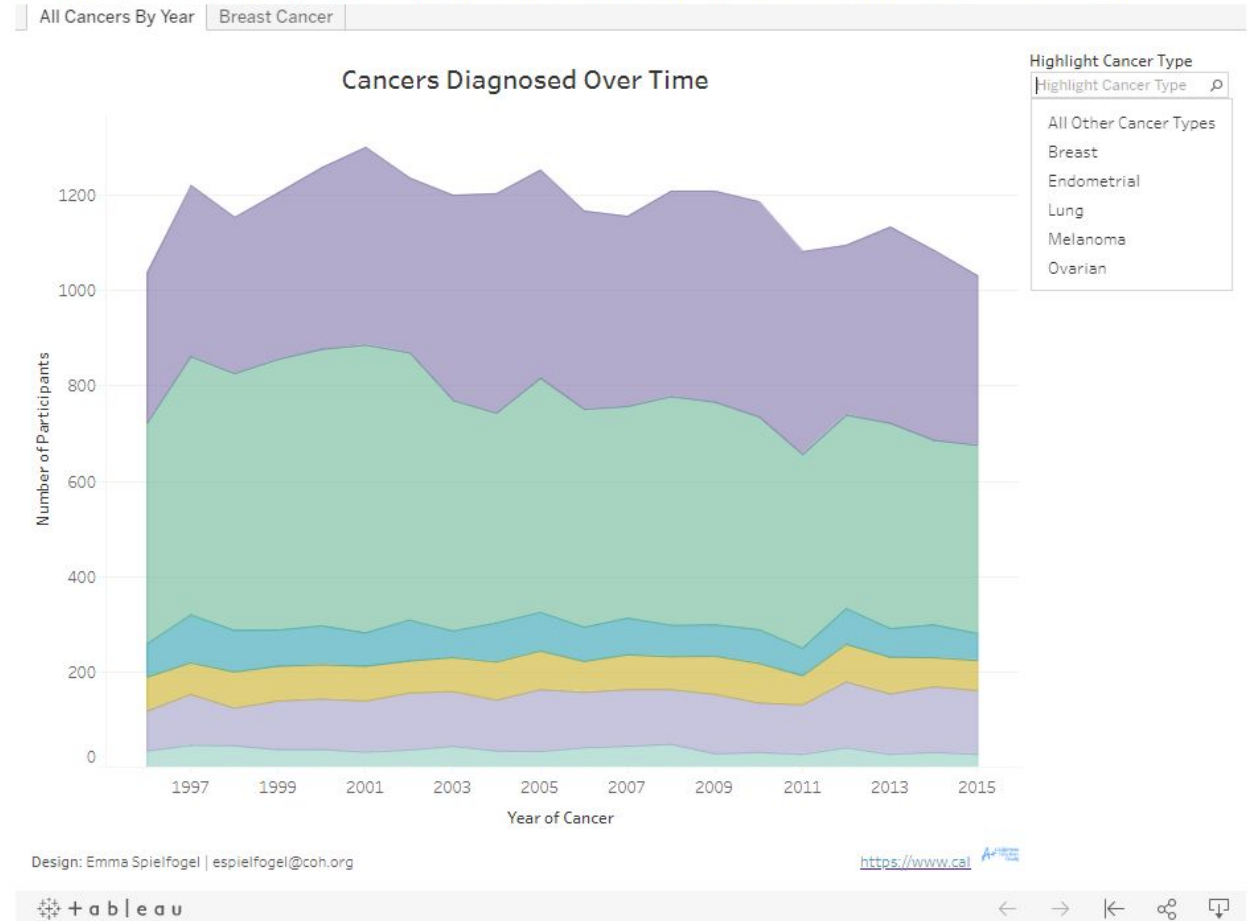
Temporal trends in deaths due to **Dementia** represent an aging population.

# Engage Our Participants to Improve Retention



Use our interactive graphs to learn more about the CTS!

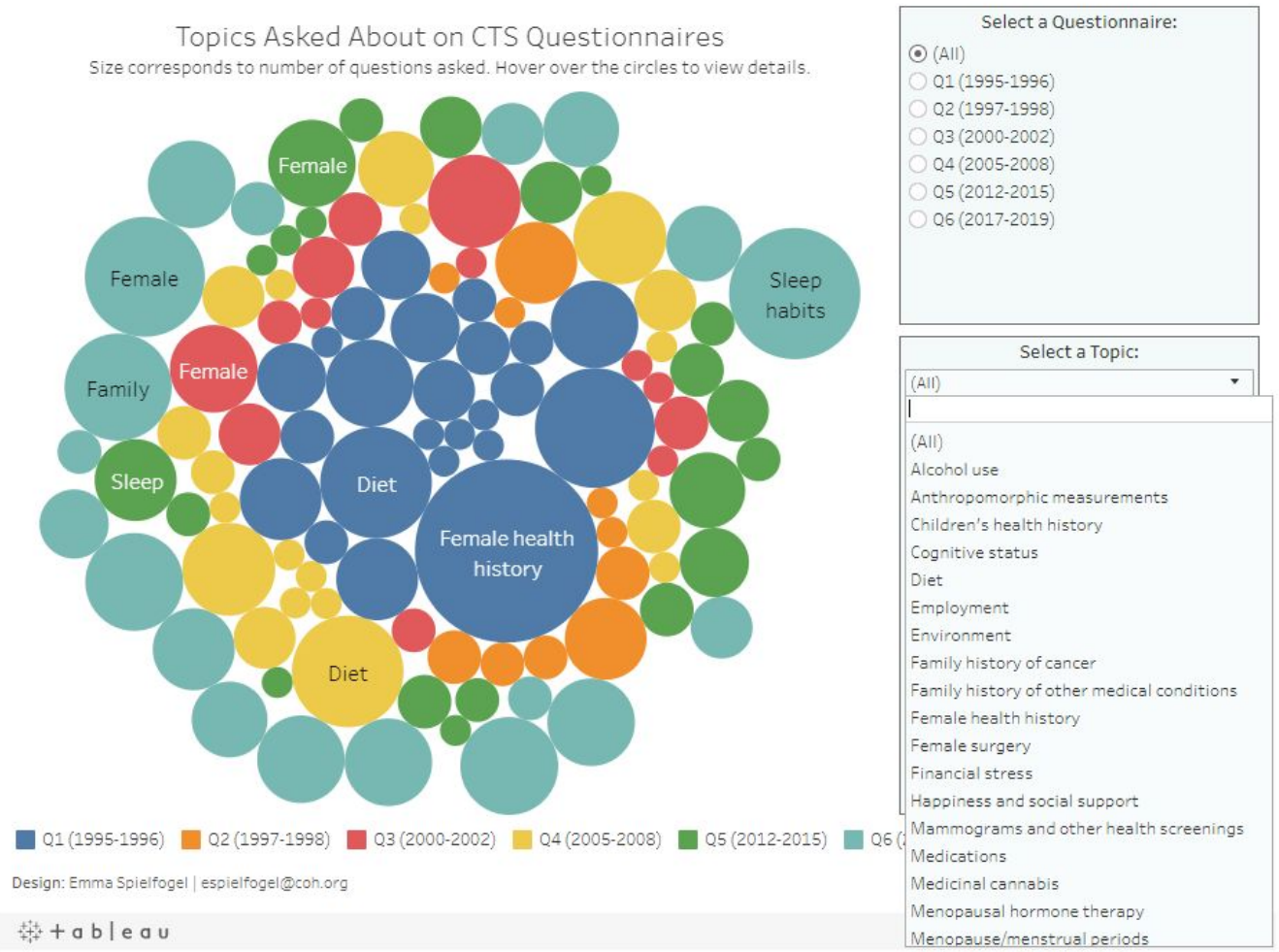
Click on the tab titled "Breast Cancer" for more information on the number of breast cancer cases in the California Teachers Study.



# Engage Our Partners to Facilitate Research

## Study Questionnaires

Participants have been asked to complete six [self-reported questionnaires](#), which collected data on various health conditions and risk factors.



# The Journey Hasn't Quite Been the Reward...



- Modernized our data & eliminated silos
- Introduced visualization for exploration, QC, & discovery
- Standardized data-analysis processes
- Democratized all our data in user-friendly ways

... and saw moderate uptake among the team

# Outdated Infrastructure has Enabled Bad Processes

Harvard  
Business  
Review

IT

## Why Process Is U.S. Health Care's Biggest Problem

by John S. Toussaint and Kathryn Correia

MARCH 19, 2018

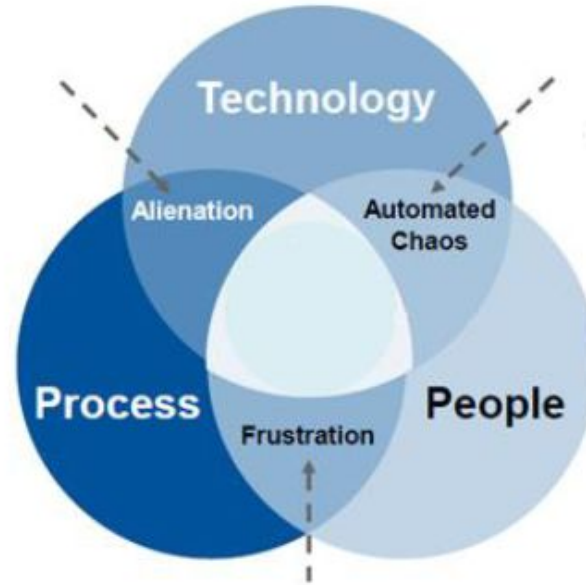
*It only takes 10 minutes of direct observation of a nurse in a hospital to understand care-delivery processes are not standardized and are dependent on individuals, not systems. This lack of reproducibility leads to errors. Since every caregiver does it his or her own way, it's difficult to improve anything. Stable systems that are reproducible are required to deliver consistently high quality. Industrial companies figured this out 50 years ago.*

*What that nurse and doctor do and how they do it are still what will determine successful outcomes of care.*

*Every nurse and doctor does not get to do it his or her own way. Standards are established about how the work is performed, and those standards are followed by all until a better way is determined collectively by the team.*

# A Counterclockwise Journey, From the Bottom Up

**Process and Technology Without People**  
Alienation and Turnover  
Underutilized Systems



**People and Technology Without Process**  
Automated Chaos and Confusion  
Poor Service Delivered

**People and Process Without Technology**  
Frustration and Inefficiency  
High Cost of Operation

# An Example of a Learning Healthcare System

**“... science, informatics, incentives, and culture are aligned for continuous improvement and innovation, ... with best practices seamlessly embedded ...”**



<https://www.calteachersstudy.org/team>

# “Same thing only different” is spot-on

*Healthcare organizations rarely have the exact same data or patients.  
Healthcare organizations shouldn't have the same data solutions.*

TECHNOLOGY FEATURE · 13 JANUARY 2020

## Eleven tips for working with large data sets

Big data are difficult to handle. These tips and tricks can smooth the way.  
Here are 11 tips for making the most of your large data sets.

nature

**Cherish your data**

**Visualize the information**

**Show your workflow**

**Use version control**

**Record metadata**

**Don't download the data**

**Start early**

**Make computing time count**

**Get help**

**Automate, automate, automate**

**Capture your environment**



# California Teachers Study Team

<https://www.calteachersstudy.org/team>



Cheryl Anderson



Jennifer Benbow



Nadia Chung



Sandeep Chandra

## Principal Investigators



Jessica Clague  
DeHart



Christine Duffy



Ann Hamilton



Susan Hurley



Jim Lacey



Elena Martinez



Hannah Lui Park



Lani Park



Rich Pinder



Kristen Savage



Emma Spielfogel



Sophia Wang

# Thank you!

---

## Additional resources:

- **California Teachers Study Data:** [www.calteachersstudy.org/cts-data](http://www.calteachersstudy.org/cts-data)
- **Tableau for Healthcare:** [tableau.com/solutions/healthcare-analytics](http://tableau.com/solutions/healthcare-analytics)
- **Visit Tableau at HIMSS:** [tableau.com/community/events/himss20-conference-2020-03-09](http://tableau.com/community/events/himss20-conference-2020-03-09)
- **On-Demand Recording on** [tableau.com](http://tableau.com)

