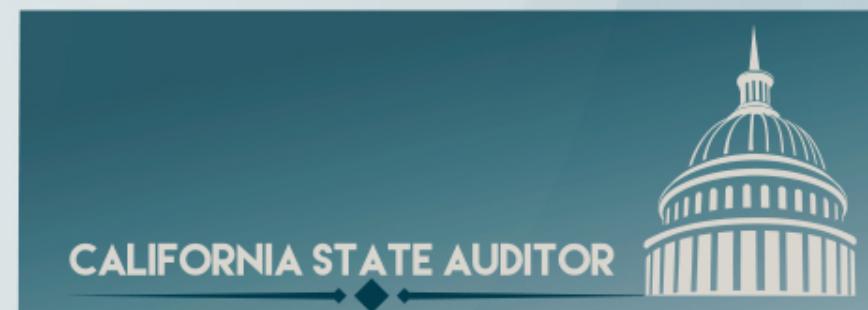


Data Analysis Tools and Innovative Data Analytics

Ben Ward, CISA, ACDA
California State Auditor

October 31, 2017



Clarifying Issues •
Formulating Solutions •
Positively Impacting Californians •



Agenda

- Identifying Data
- Evaluating Data Quality
- Using Tools for Analysis and Reporting
 - ACL
 - Tableau
 - ArcGIS

Background

- Conducts audits requested and approved by the Legislature or audits mandated by law
- Analyzes large sets of data to address audit objectives



Identifying Data

- Publicly available data
- Auditee data
- Surveys



Surveys

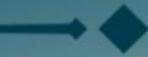
Audit of California's oversight of information security

- On-site reviews
- Survey



Surveys

- Surveyed entities for an audit of the federal Buy American requirement
 - Followed up with selection of 19 survey respondents
 - Only three respondents provided support for their answers



Evaluating Data Quality

- Reliability of computer-processed data
 - Logic Testing
 - Accuracy Testing
 - Completeness Testing

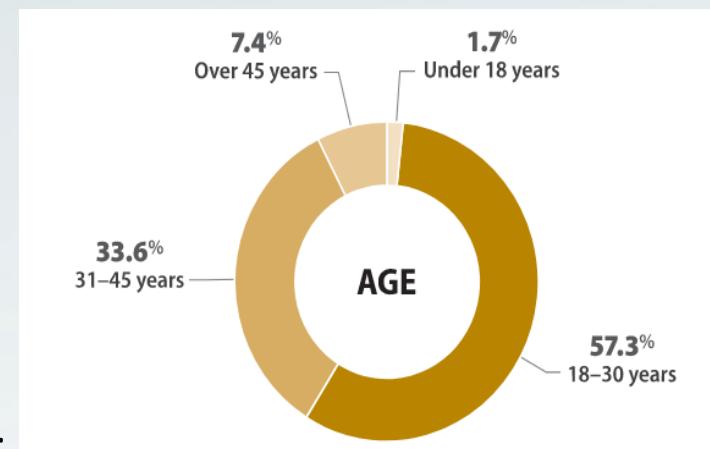


Logic Testing

- Audit of CalGang database
 - Shared criminal intelligence system
 - Tracks suspected gang members

Logic Testing

- Categorize gang members by age
 - 42 individuals in CalGang whose birthdates indicated that they were less than one year old
 - Found that 28 of these “babies” were in the system in part because—according to the data—they “admitted to being gang members”





Logic Testing

- Categorize gang members by location
 - City names were misspelled or abbreviated in inconsistent ways
 - More than 100 variations for Los Angeles, including “Los Angels” and “Los Angelesees”
 - Classified gang members based on county because the city data was unusable

→◆ Accuracy/Completeness

- Audit of the State's oversight of noncompetitive contracts
 - Data contained inaccurate dollar amounts, inaccurate procurement methods, and were missing contracts

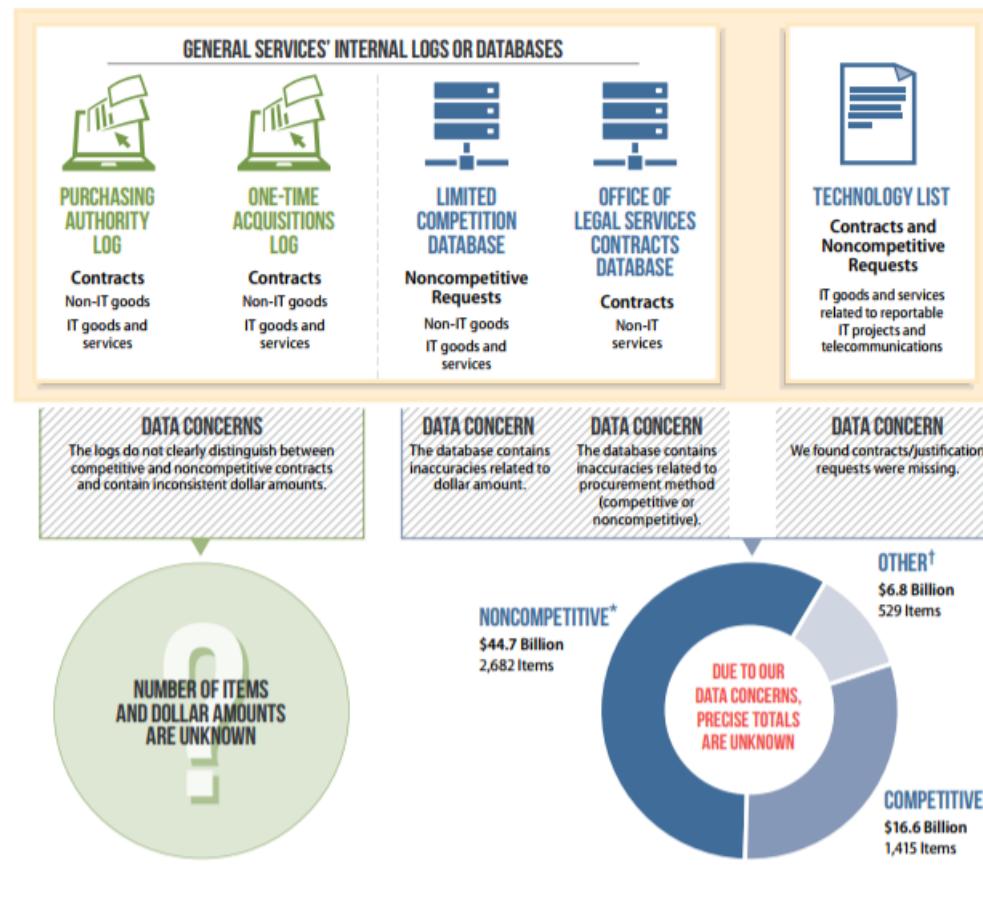
→◆ Accuracy/Completeness

- \$34 million contract identified as competitively bid
 - Original \$3 million contract was competitively bid
 - \$31 million in amendments was noncompetitive
- Twelve of 31 contracts and amendments were missing from data

Accuracy/Completeness

Figure 4

General Services' and Technology's Databases Are Inadequate for Accurately Identifying Contracts Over \$1 Million Fiscal Years 2011–12 Through 2015–16





ACL Analytics

- Main tool for data mining and descriptive analytics
- Designed for the audit community
 - Read-only access to original data
 - Log of all analysis performed



ACL Analytics

- Different ways to perform analysis
 - graphical user interface
 - scripting
- Dialogue boxes help customize analysis

ACL Analytics

WIAF 2017.acl - Analytics

File Edit Import Data Analyze Sampling Applications Tools Server Window Help

NAVIGATOR

- WIAF 2017.acl
- Logs
- Scripts
- Select_Sample_of_Orders
- Tables
 - Customer
 - Orders
 - Product

COMMAND LINE

Product

ProdID	ProdName	UnitPrice	Descript	ShipWt
1	Chai	18.00	Soft drinks, coffees, teas, beer, and ale	5.65
2	Chang	19.00	Soft drinks, coffees, teas, beer, and ale	6.15
3	Chef Anton's Cajun Seasoning	22.00	Sweet and savory sauces, relishes, spreads, and seasonings	12.80
4	Uncle Bob's Organic Dried Pears	30.00	Dried fruit and bean curd	3.85
5	Northwoods Cranberry Sauce	40.00	Sweet and savory sauces, relishes, spreads, and seasonings	2.15
6	Ikura	31.00	Seaweed and fish	18.90
7	Queso Cabrales	21.00	Cheeses	4.80
8	Queso Manchego La Pastora	38.00	Cheeses	5.20
9	Konbu	6.00	Seaweed and fish	8.40
10	Tofu	23.25	Dried fruit and bean curd	6.85
11	Genen Shouyu	15.50	Sweet and savory sauces, relishes, spreads, and seasonings	13.05
12	Pavlova	17.45	Desserts, candies, sweetbreads	3.70
13	Carnarvon Tigers	62.50	Seaweed and fish	6.00
14	Teatime Chocolate Biscuits	9.20	Desserts, candies, sweetbreads	8.35
15	Sir Rodney's Marmalade	81.00	Desserts, candies, sweetbreads	2.65



ACL Analytics

- Audit of the Department of Motor Vehicles' disabled person parking placard program
 - Placards allow people to park in metered spaces without paying or time limitations
 - Significant incentive for misuse



ACL Analytics

- Obtained placard data for the entire State of California and identified areas of potential misuse
 - 8,000 individuals had two or more active placards

→◆ ACL Analytics

- Nearly 26,000 placard holders with active, permanent placards who were age 100 or older
- California's entire centenarian population: about 8,000 individuals as of 2014



ACL Analytics

- Compared DMV's active placard holders to the U.S. Social Security Administration's death information and found 35,000 matches
 - May not be precise because of limitations in the available data
 - Tested 10 names and found public information to corroborate that five individuals were deceased

ACL Analytics

Table 2

Replacements of Permanent Disabled Person Parking Placards DMV
Issued From July 2013 Through June 2016

NUMBER OF REPLACEMENTS PER INDIVIDUAL	NUMBER OF INDIVIDUALS WITH REPLACEMENTS	TOTAL NUMBER OF REPLACEMENTS
1	288,038	288,038
2	38,834	77,668
3	6,401	19,203
Subtotals of 3 or Fewer Replacements	333,273	384,909
4	1,501	6,004
5	437	2,185
6-10	332	2,274
11-15	24	295
16+	9	175
Subtotals of 4 or More Replacements	2,303	10,933
Totals	335,576	395,842

Source: California State Auditor's analysis of data obtained from DMV's Vehicle/Vessel Registration Master File.

Note: As of June 30, 2016, there were approximately 2.4 million active, permanent disabled person parking placards.



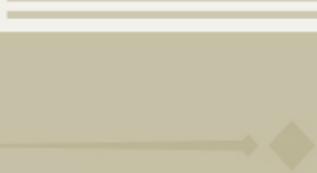
ACL Analytics

- Legislature passed and the Governor signed a bill requiring DMV to:
 - Conduct audits of disabled placard applications
 - Review additional death records
 - Limit number of replacement placards



Tableau

- Tableau Desktop and Tableau Reader
- Main tool for data visualization
- Gives non-technical auditors greater access to the data



Tableau

- Drag and drop interface
- Multiple graph types
- Dashboards with multiple interactive graphics that are linked together

Tableau

The screenshot shows the Tableau desktop application interface. The top menu bar includes File, Data, Worksheet, Dashboard, Story, Analysis, Map, Format, Server, Window, and Help. The ribbon bar contains various icons for navigation and data manipulation. The left sidebar contains the Data pane (Sales Commission, Sales Target, Sample - Superstore), Dimensions pane (Customer, Order, Location, Order Profitable?, Product, Ship Status, Measure Names), and Measures pane (Days to Ship Actual, Days to Ship Scheduled, Discount, Profit, Profit per Order, Profit Ratio, Quantity, Sales, Sales above Target?, Sales per Customer, Latitude (generated)). The center workspace is titled "Sheet 1" and features three "Drop field here" placeholder areas. To the right is a large library of visualization options, including various charts like bar charts, line graphs, and maps. A "Select or drag data" instruction and a note about using the Shift or Ctrl key to select multiple fields are also present.



Tableau

- Specialists use Tableau Desktop to create dashboards
- Audit teams use Tableau Reader to interact with dashboards

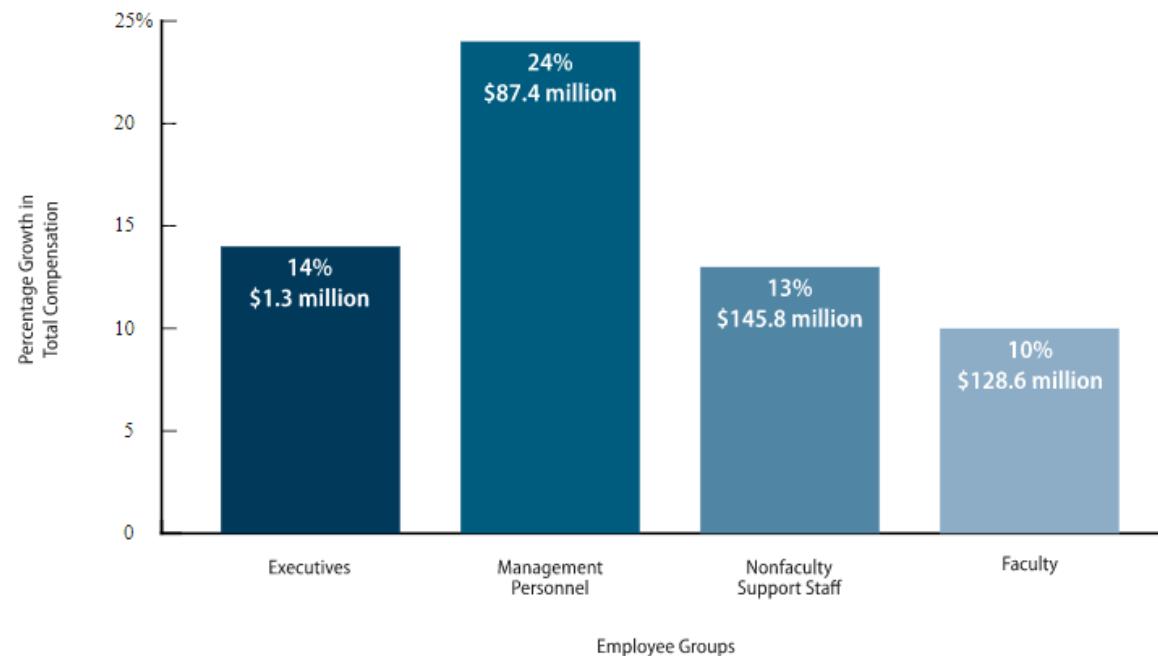
Tableau

- Audit of the California State University System's hiring and compensation
- Growth in number and compensation of management personnel significantly outpaced other employees

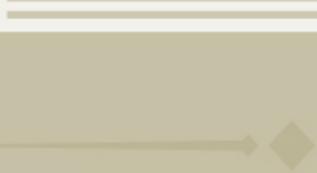
Tableau

Figure 6

Growth in *Total Compensation* for California State University's Management Personnel Positions Outpaced That of Other Employee Groups
Fiscal Years 2007–08 Through 2015–16

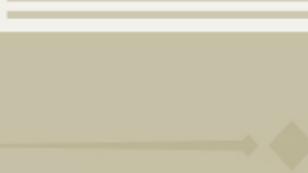


Source: California State Auditor's analysis of California State University payroll data as maintained in the State Controller's Office's Uniform State Payroll System.



Tableau

- Quickly compared trends at different colleges
- Answered public records act requests without additional programming



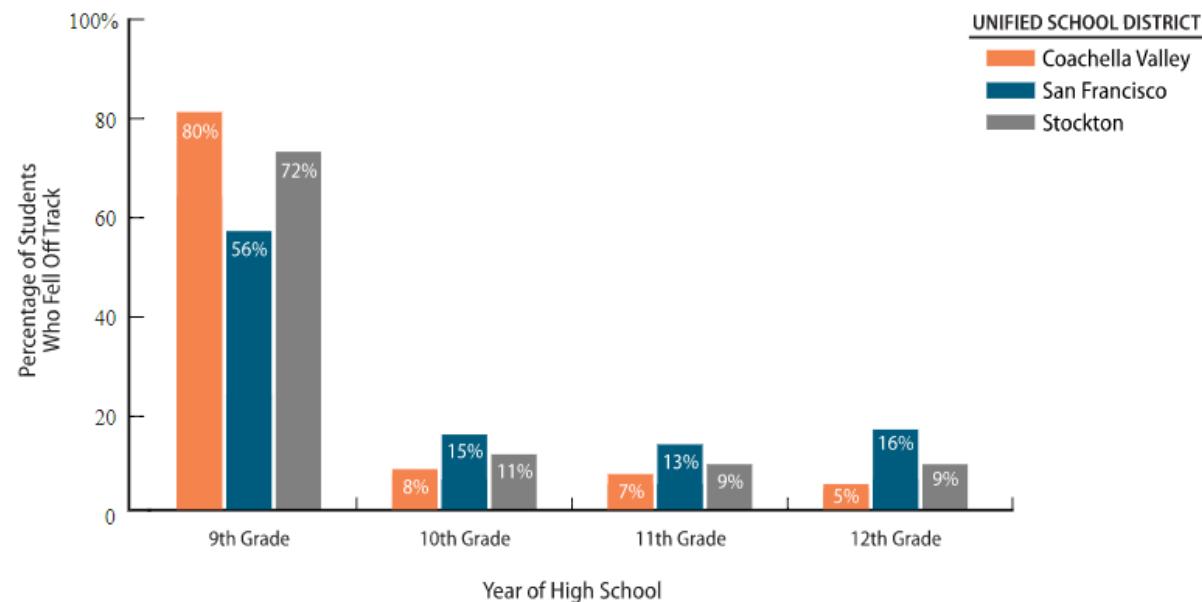
Tableau

- Audit of access to and completion of college preparatory coursework
 - Students must complete certain college preparatory classes to attend UC and CSU
 - Coursework completion rates heavily influenced by students' performance in ninth grade

Tableau

Figure 4

**Most Students Who Fell Off Track Did So in Grade Nine
Graduation Years 2013 Through 2015**



Source: California State Auditor's analysis of student data provided by Coachella Valley, San Francisco, and Stockton Unified School Districts.

Notes: We calculated the number of students who fell off track to meet college preparatory requirements during each year of school for students in our cohorts who graduated high school in years 2013 through 2015.

We excluded students who left the district.

Tableau

- Interactive dashboard allowing public to filter by demographic information
- <http://www.auditor.ca.gov/reports/2016-114/supplemental.html>

Tableau

Status and GPA Subject Area Enrollment

Student A-G Status and Grade Point Average

Use the drop-down menus below to filter the data. Additionally, you can gain further information about a specific group of students by clicking on a column or percentage. For example, if you click on the column heading "Off Track" on the "High School Year 1" table, the system filters the data for students off track after their first year of high school.

School District	Gender	Race/Ethnicity	English Learner Status	Unduplicated Pupil Status*
(All)	(All)	(All)	(All)	(All)

High School Year 1

Graduation Year	Left District	Off Track	On Track
2013	9%	46%	46%
2014	6%	46%	48%
2015	10%	42%	47%

High School Year 2

Graduation Year	Left District	Off Track	On Track
2013	15%	46%	40%
2014	14%	43%	43%
2015	18%	41%	42%

High School Year 3

Graduation Year	Left District	Off Track	On Track
2013	22%	42%	36%
2014	20%	41%	39%
2015	24%	37%	38%

High School Year 4

Graduation Year	Left District	Off Track	On Track
2013	27%	42%	32%
2014	24%	41%	35%
2015	28%	38%	34%

A-G Completion Rate (Excludes Students Who Left the District)

Graduation Year	A-G Completion Rate
2013	43%
2014	47%
2015	47%

Student Grade Point Average

Graduation Year	1	2	3	4
2013	2.43	2.47	2.67	2.88
2014	2.52	2.57	2.71	2.87
2015	2.51	2.58	2.73	2.89

Range of Grade Point Averages
2.43 → 2.89

Source: California State Auditor's analysis of student data provided by Coachella Valley, San Francisco, and Stockton Unified School Districts.

Notes: "Left District" indicates that a student either did not complete the current school year or left in a previous school year.

A blank cell for A-G status indicates that we omitted information because it would identify 50 or fewer students, or 100 percent of students. A blank cell for grade point average indicates that we omitted information because it would identify fewer than 10 students. These omissions are in accordance with the aggregate data reporting guidance issued by the California Department of Education and the United States Department of Education.

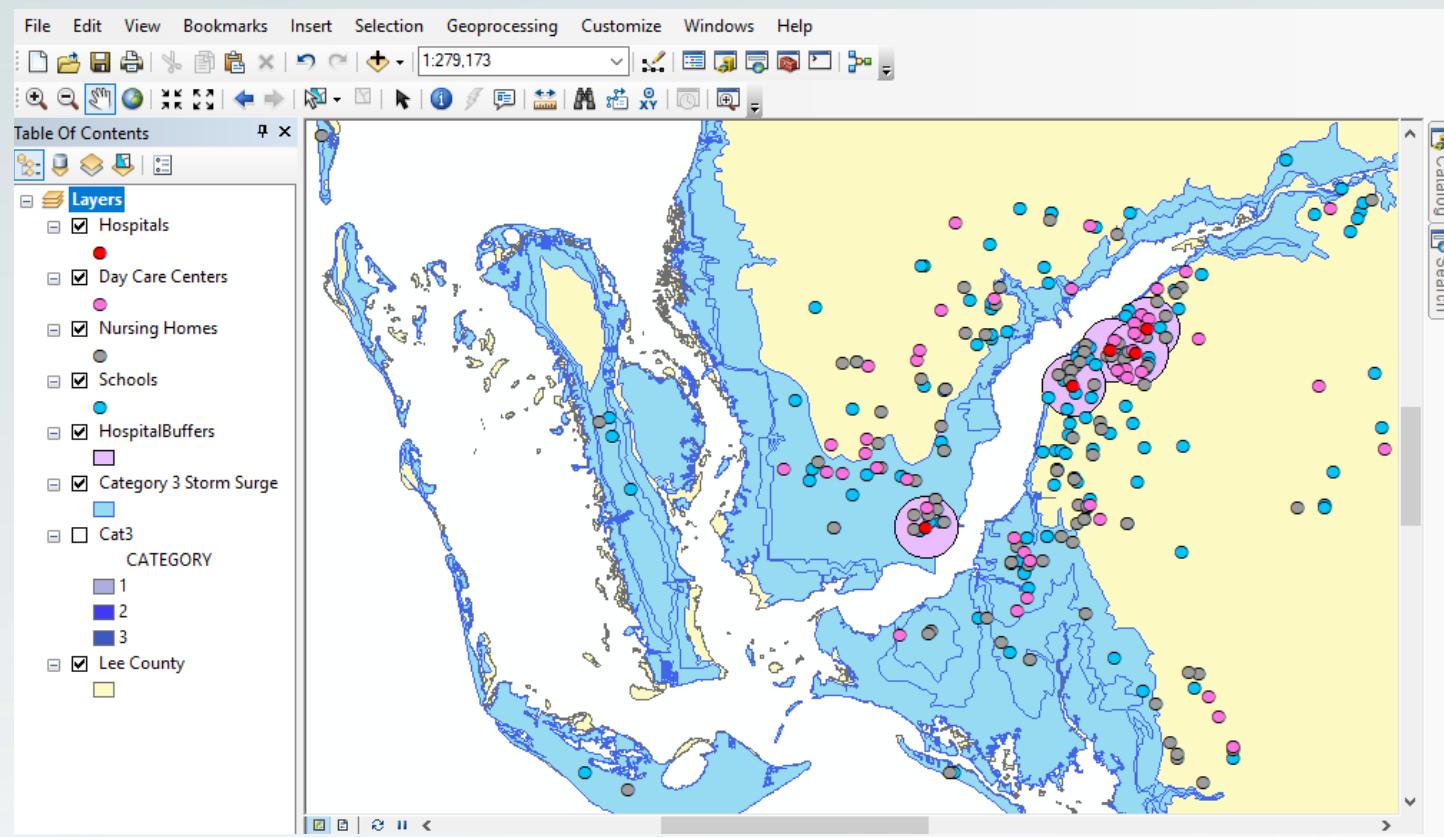
*Unduplicated students consist of English learners, students who are classified as eligible for free or reduced price meals, or students in foster care.



ArcGIS

- Analysis based on location data
- Requires specialized training

ArcGIS



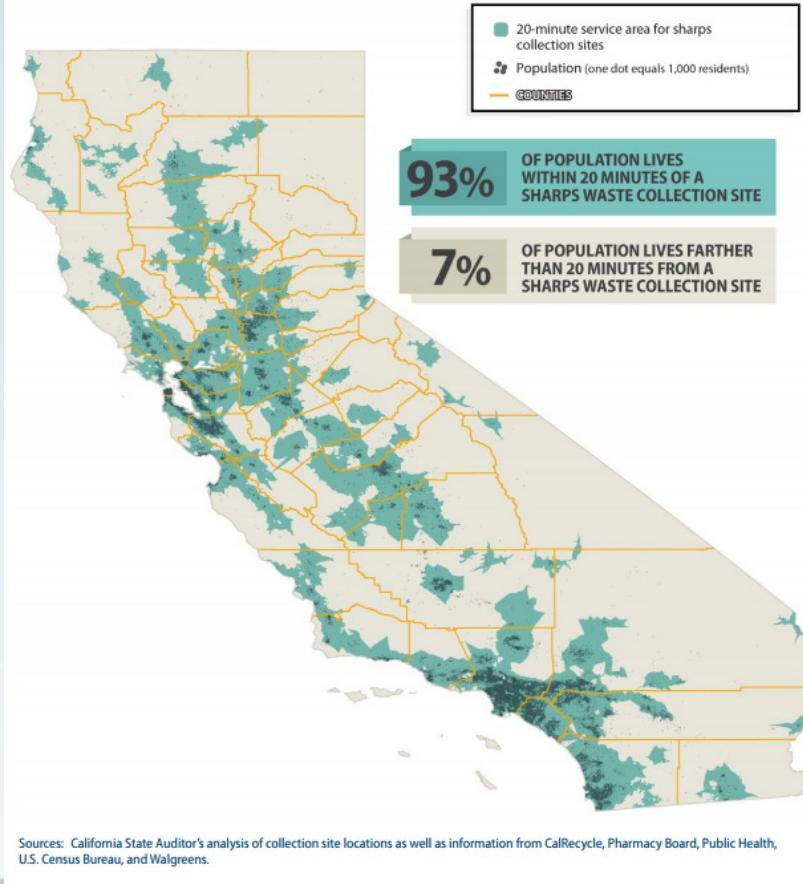


ArcGIS

- Audit of home-generated sharps and pharmaceutical waste
 - Lack of lead agency to oversee waste disposal
 - Lack of adequate information about collection sites

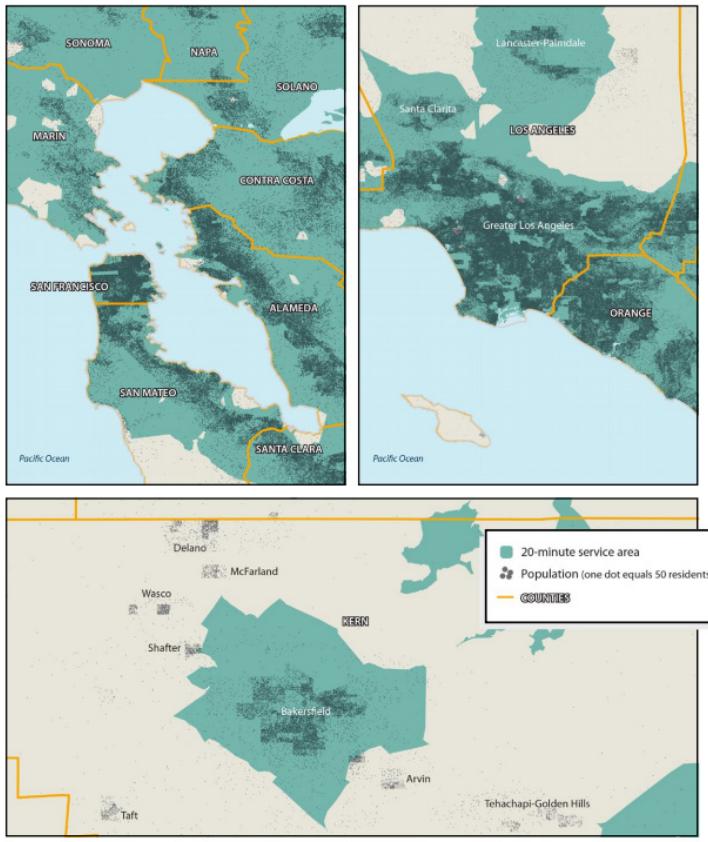
ArcGIS

Figure 5
Most Californians Live Within a 20-Minute Drive of Collection Sites for Home-Generated Sharps Waste



ArcGIS

Figure 6
Access to Home-Generated Sharps Waste Collection Sites Varies Among Different Regions



Sources: California State Auditor's analysis of collection site locations as well as information from CalRecycle, Pharmacy Board, Public Health, U.S. Census Bureau, and Walgreens.

Report Links

- CalGang (www.auditor.ca.gov/reports/2015-130/index.html)
- DGS and Technology (www.auditor.ca.gov/reports/2016-124/index.html)
- DMV (www.auditor.ca.gov/reports/2016-121/index.html)
- CSU (www.auditor.ca.gov/reports/2016-122/index.html)
- High School Coursework (www.auditor.ca.gov/reports/2016-114/index.html)
- Pharmaceutical Waste
(<http://www.auditor.ca.gov/reports/2016-127/index.html>)



Questions?