EAST BAY MUNICIPAL UTILITY DISTRICT

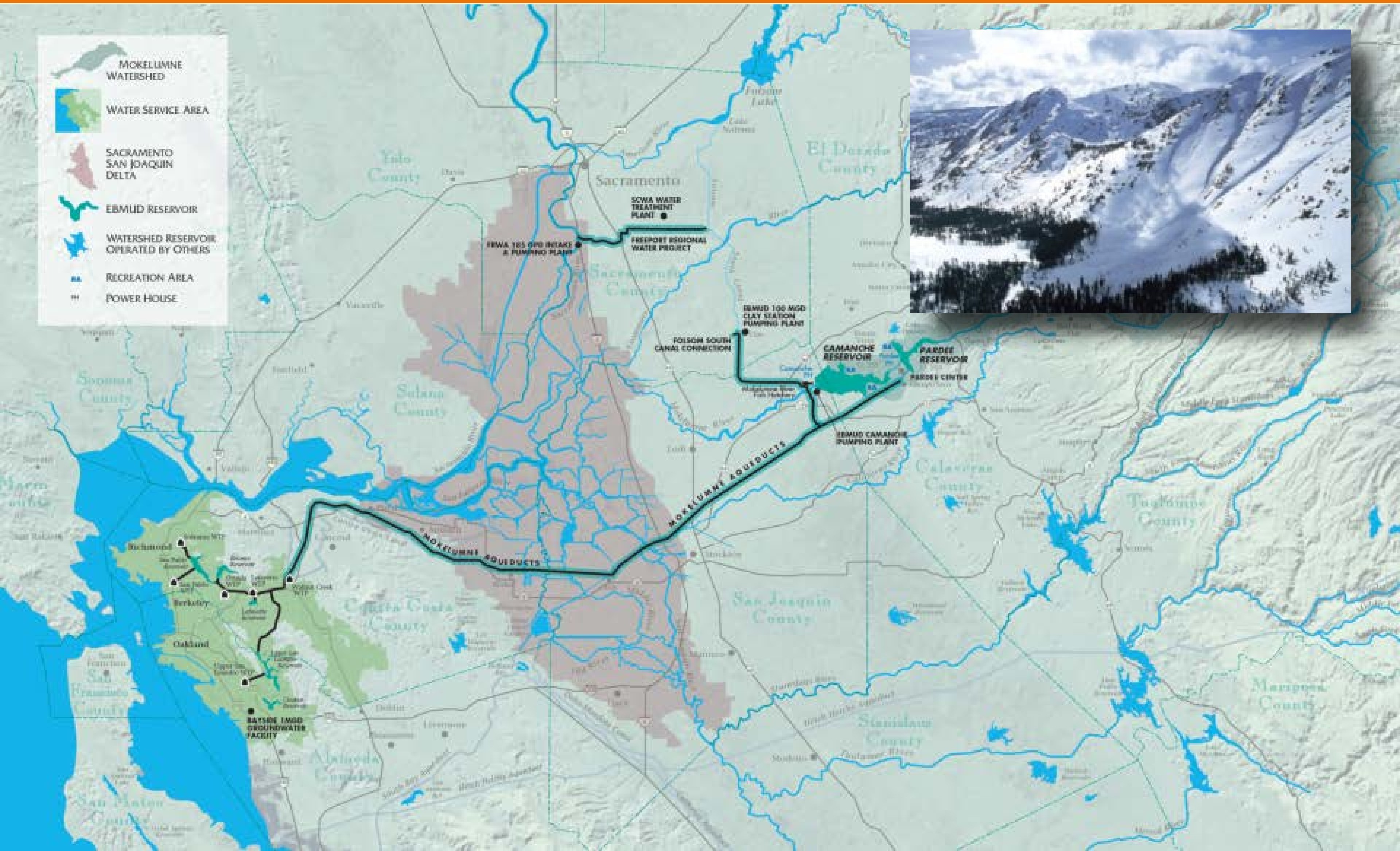
Machine Learning and Pipeline Replacement Prioritization

Clifford Chan

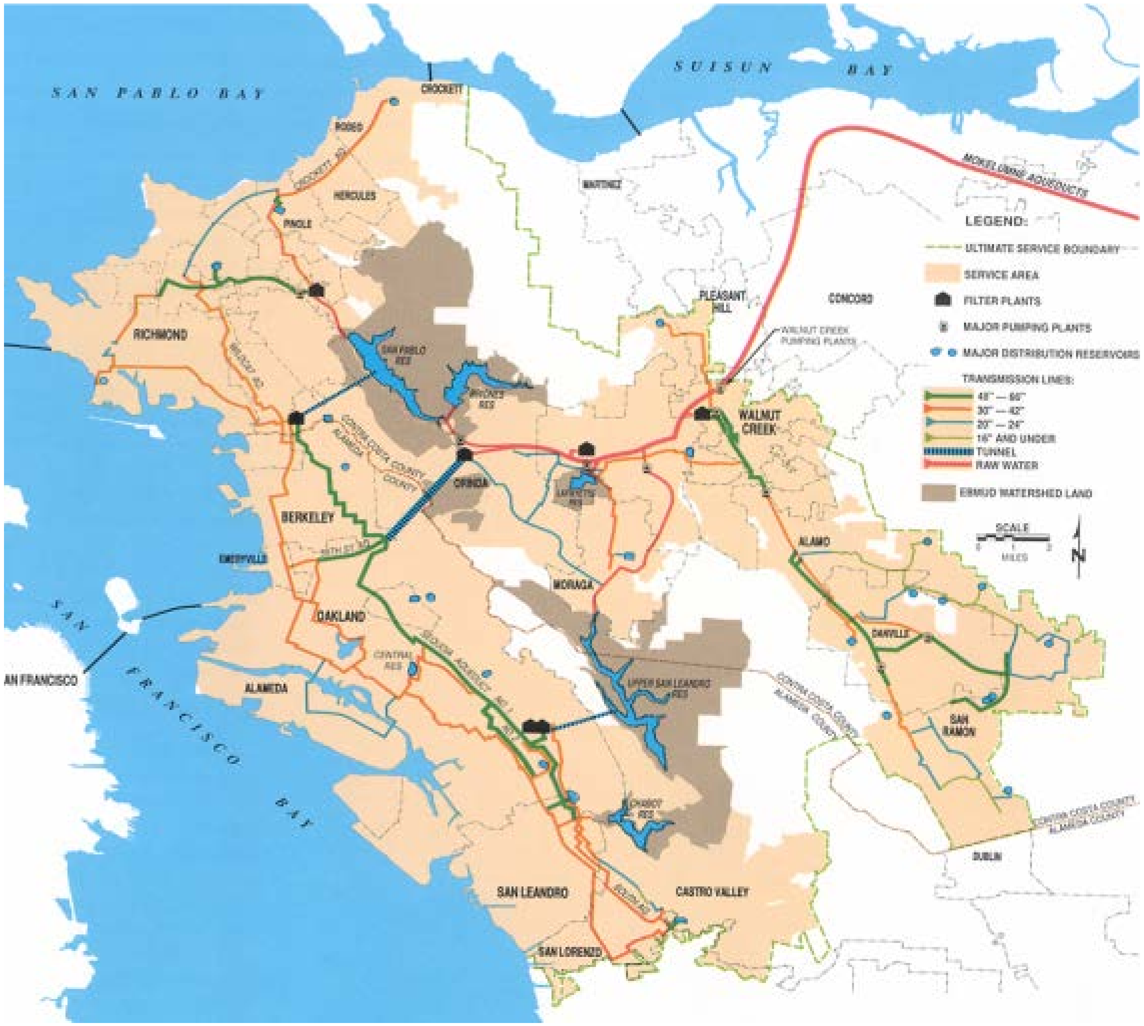
David Katzev

June 1, 2018

EBMUD Water System



EBMUD System & Service Area



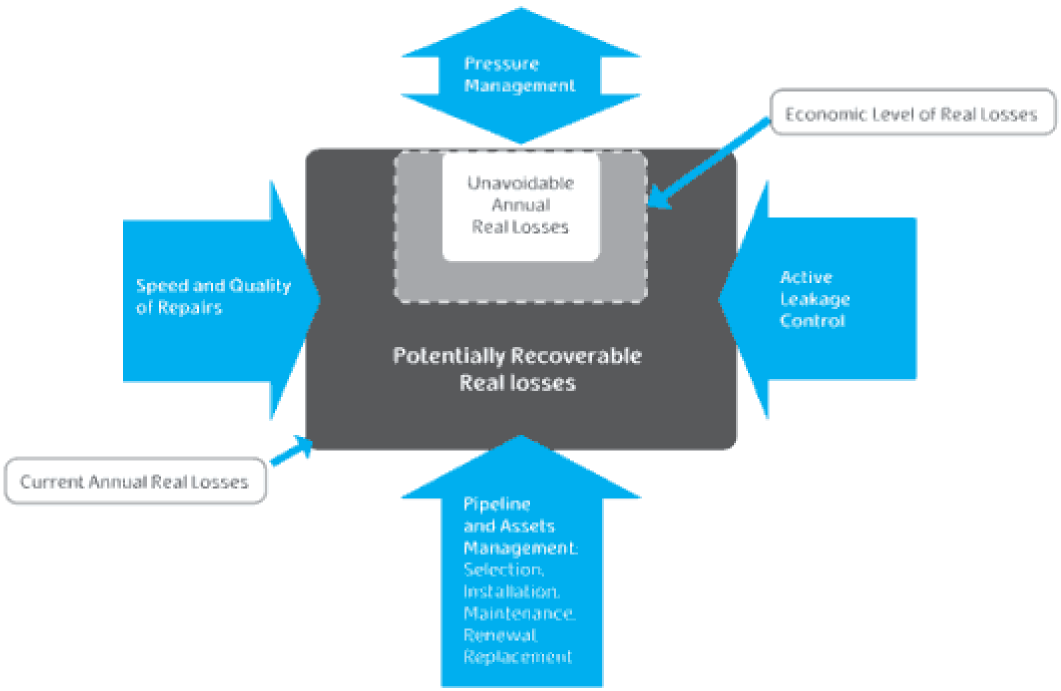
Customers
- 1,400,000 customers

Raw Water System
- 2 upcountry reservoirs
- 5 local reservoirs

Treatment System
- 6 water treatment plants

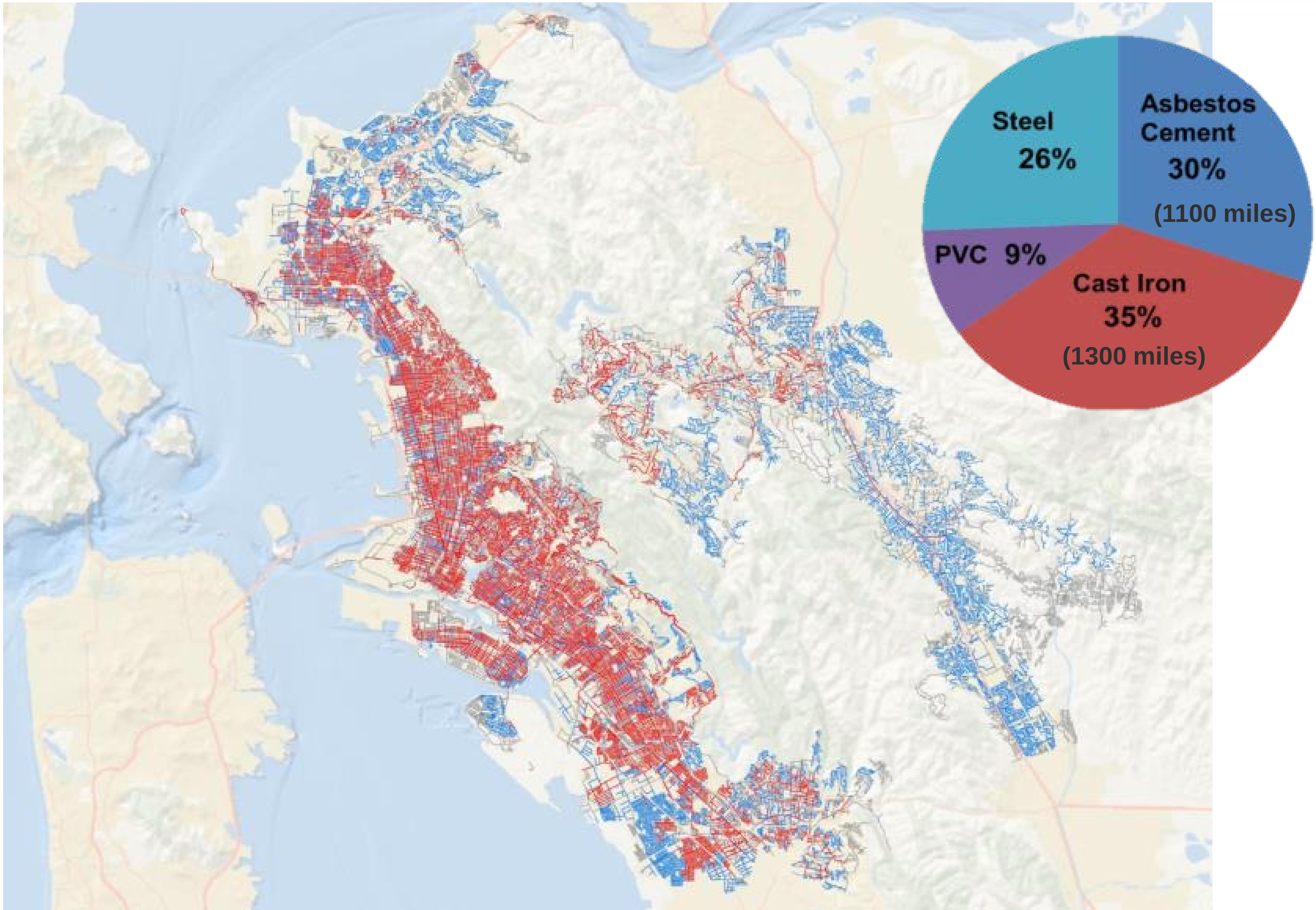
Distribution System
- 4,200 miles of pipeline
- 122 pressure zones
- 164 reservoirs
- 135 pumping plants
- 100 regulators/RCS

Addressing Real Water Loss

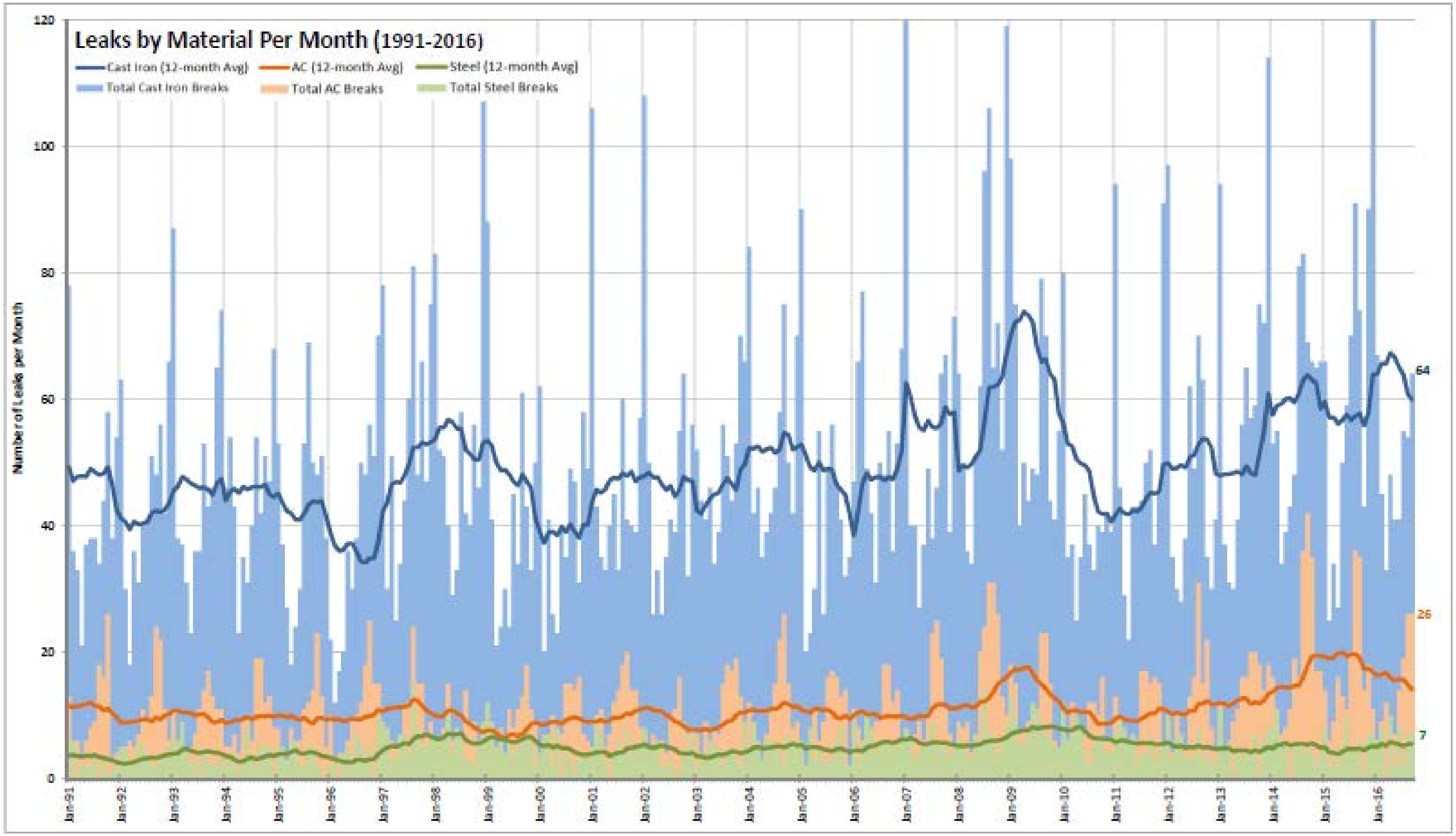


- Active leakage control
- Pressure management
- Speed and quality of repairs
- Infrastructure management

Pipeline Inventory



Pipeline Break History

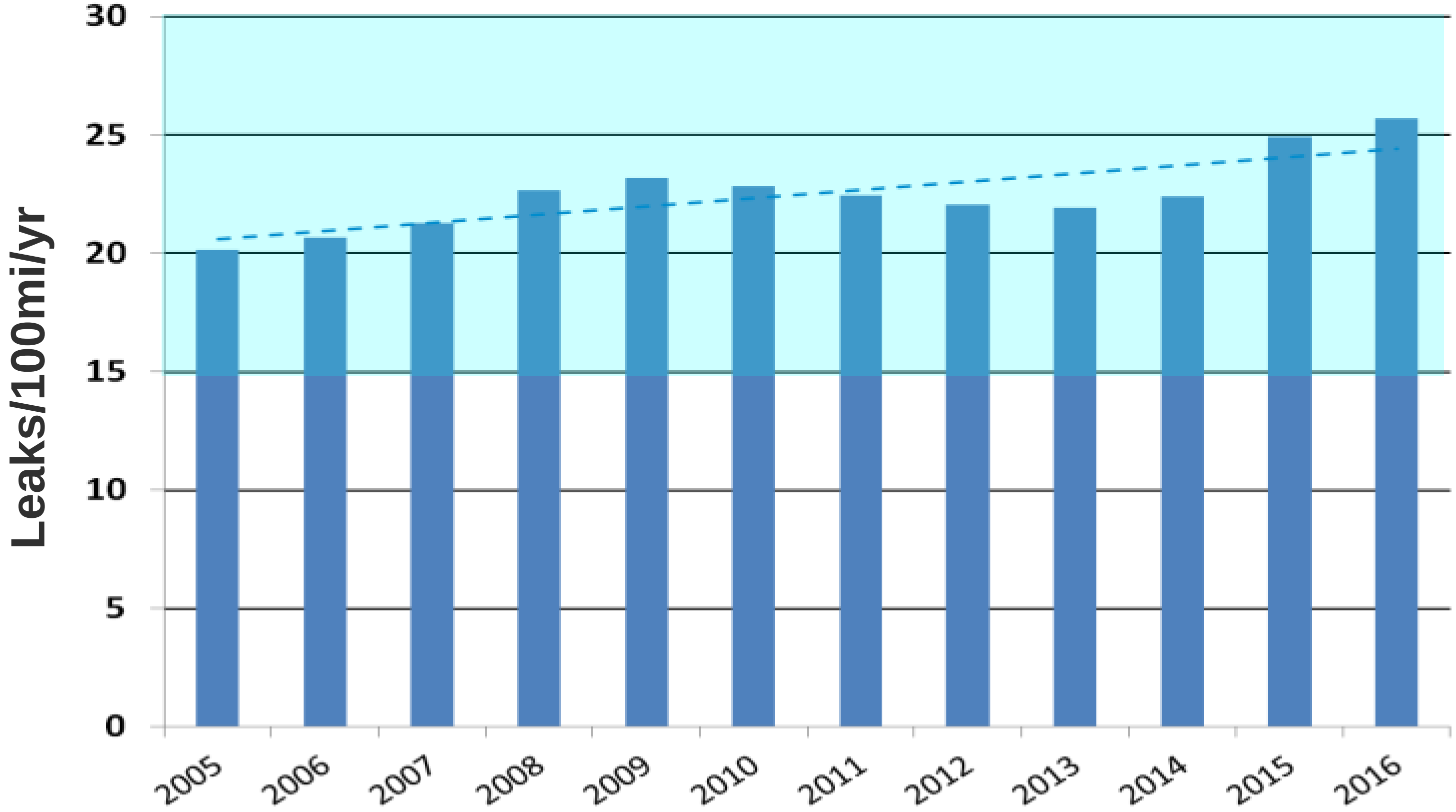


Infrastructure Renewal Program -- Ramping Up Replacements

EBMUD

Industry Benchmarks for a well maintained system

= < 15-30 leaks/100 miles/year

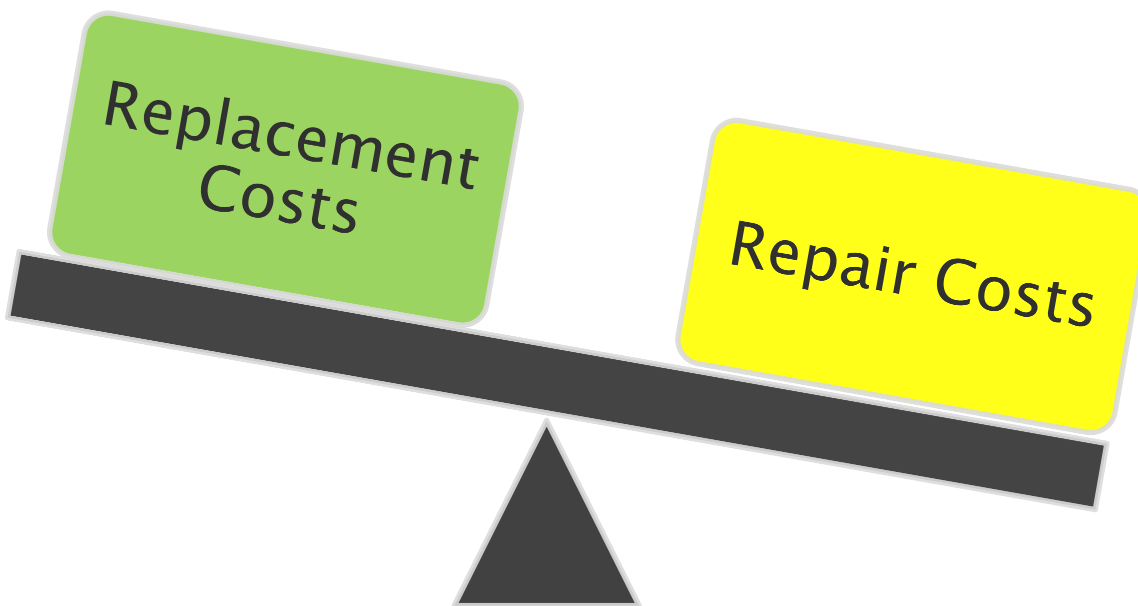


How do we make sure we are replacing the right pipe?

Prior Pipeline Replacement Model

$$\text{Cost Benefit Ratio} = \frac{\text{Repair Costs}}{\text{Replacement Costs}}$$

> 1.0 , more beneficial to replace pipe



New Pipe Replacement Risk Model

EBMUD

LOF **COF**

RISK = (Likelihood of Failure) x (Consequence of Failure)

The probability a pipeline will leak.

The resulting magnitude of consequence if the pipe does leak.

Pipe Replacement Risk Model

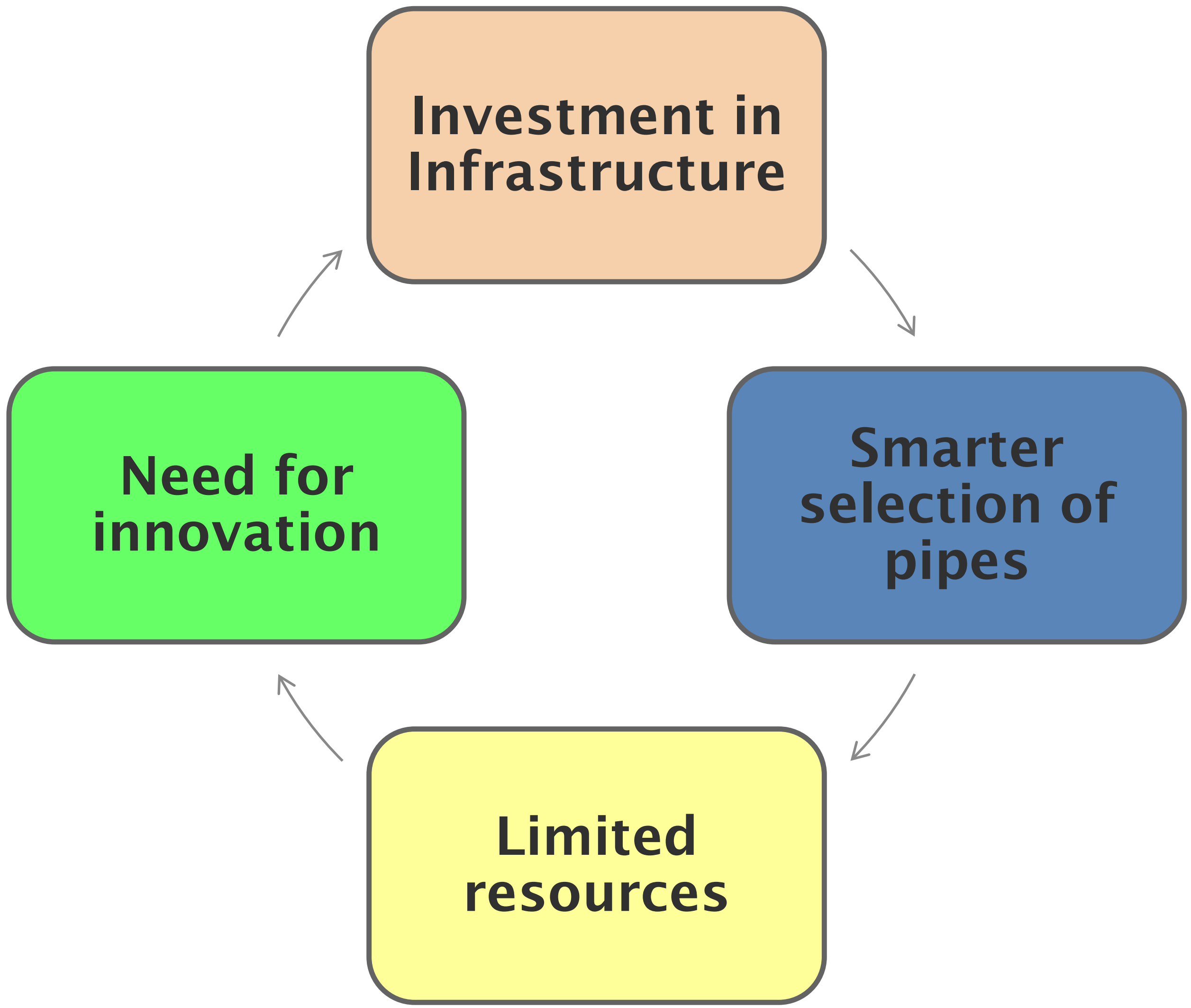
- Pipe Leak History
- Pipe Age

- Creek crossing
- Diameter
- Consumption
- Access
- Slope
- Backbone/Critical
- Highway crossing
- Railroad crossing

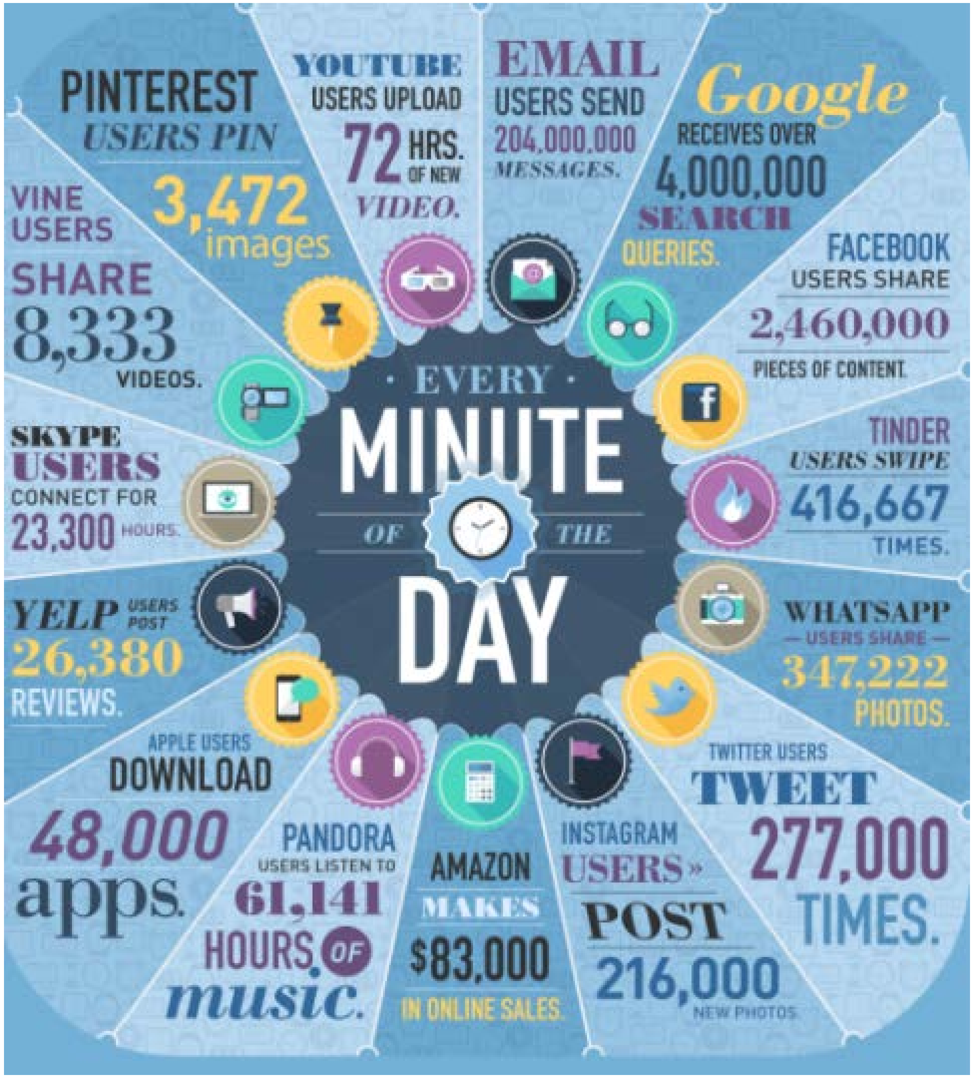
		Likelihood of Failure				
		Very Low	Low	Medium	High	Very High
Consequence of Failure	Very Low					
	Low					
	Medium					
	High					
	Very High					

How we choose projects using the Risk Model

New Decision-Making Tools



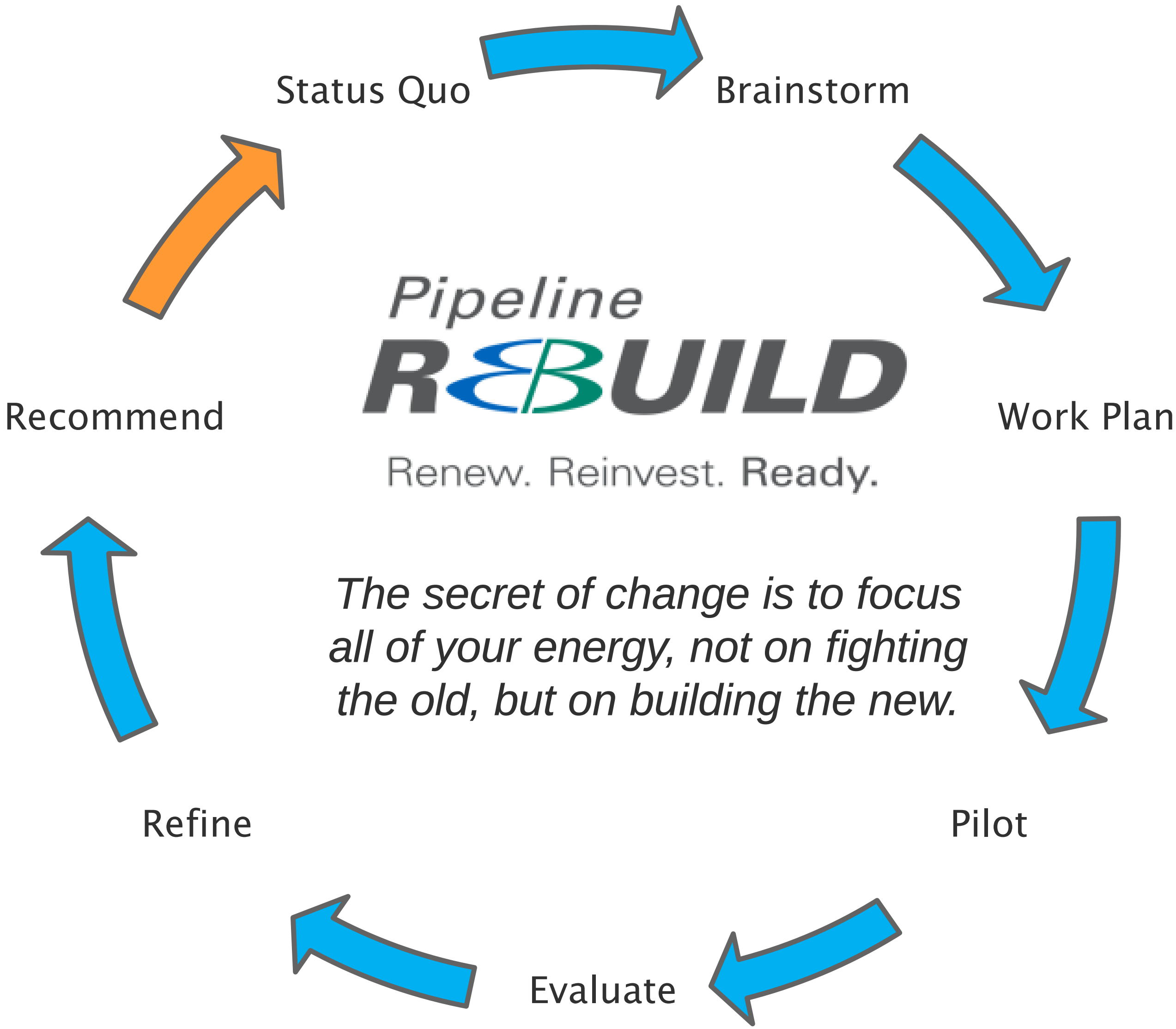
Data Never Sleeps



Operational Data

- SCADA
  - Enterprise database: ~150 million readings per year
  - Data historian: ~7 billion readings in the database
- What's Next
  - Pipeline data
  - Maintenance data
  - AMI data
  - Pressure and leak detection data

Pipeline Rebuild



The secret of change is to focus all of your energy, not on fighting the old, but on building the new.

Machine Learning
How It Works?

1

Wrangle and import water main and geographic data

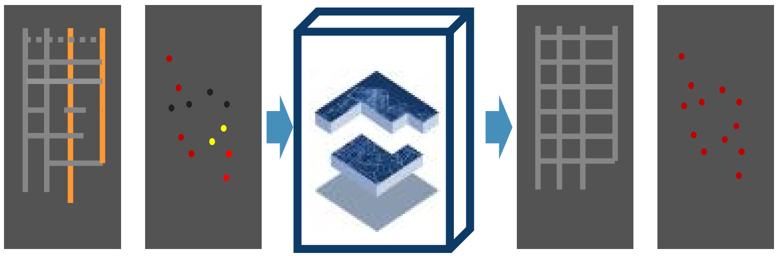
2

Layer geodata and run machine learning algorithms

3

Visualize vulnerabilities and apply LOF results

Step 1: Wrangle and Import Data



Utility Pipe Asset Data | Utility Break History | **Data Wrangling** | Cleaned Pipe Asset Data | Cleaned Break History

+ Cleaning and normalizing the data
+ Correcting wrong/outlier data points and filling in missing values
+ Geocoding breaks with pipe segments

<u>Example :</u>

<u>Received Pipe Data as ESRI Shapefile</u>

23876 Pipe Segments

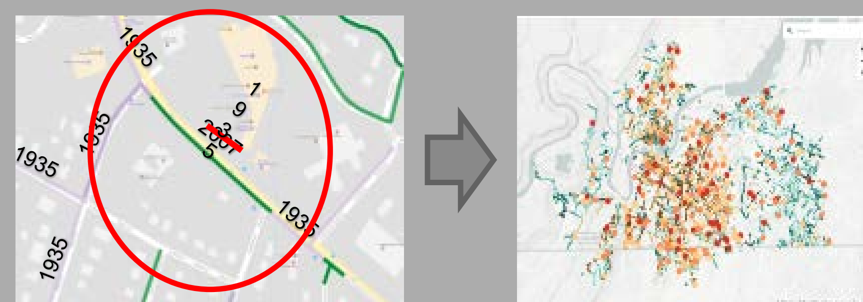
~15% Total Missing/Error

<u>Received Break data as two Excel sheets + GIS file</u>

Clean

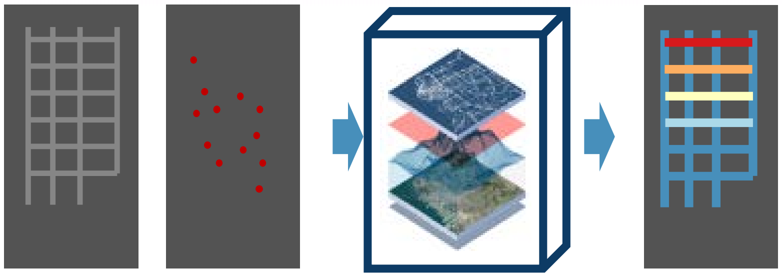
Correct

Geocode and assign to pipes



Step 2: Machine Learning Analysis

EBMUD



Cleaned Pipe Asset Data | Cleaned Break History | **Machine Learning Process** | Pipe Asset with LOF Scores

Utility Data

- Pipe Parameters
  - Length, Material, Diameter, Install Year, Pressure etc.
- Break History
  - Break info assigned to pipe segment

Variables derived from Utility Data

- Age-Derived Variables
- Pipe and Leak Density

Geo Variables

- Environment
  - 40+ Soil Properties from USGS
- Location
  - Slope, Elevation
  - Proximity to Transportation Features (roads, rail etc.)
  - Proximity to Water (salt water, river etc.)
- Population and Buildings
  - Population, Zoning, Buildings etc.

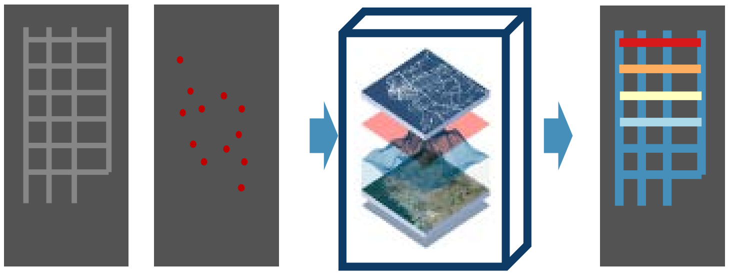
Variables derived from Geo Data

- Min/Max/Mean Distance
- Density of Soil Type Changes

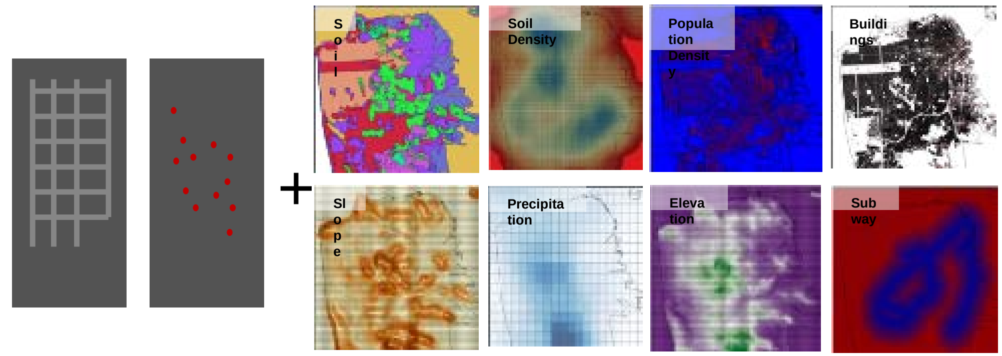
Utility variables | **Derived Utility variables** | **Geo variables** | **Derived Geo variables**

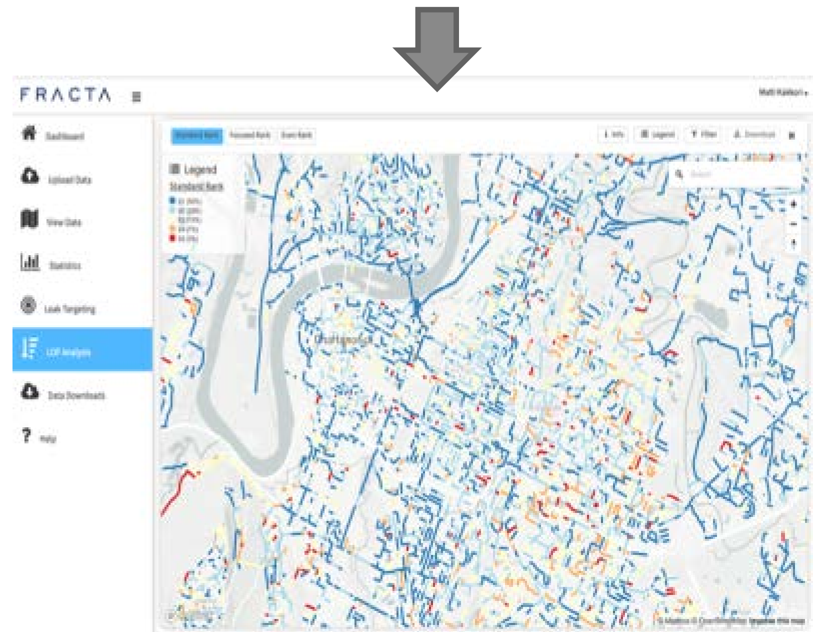
1000+ variables

Step 2: Layer Geodata and Run Machine Learn



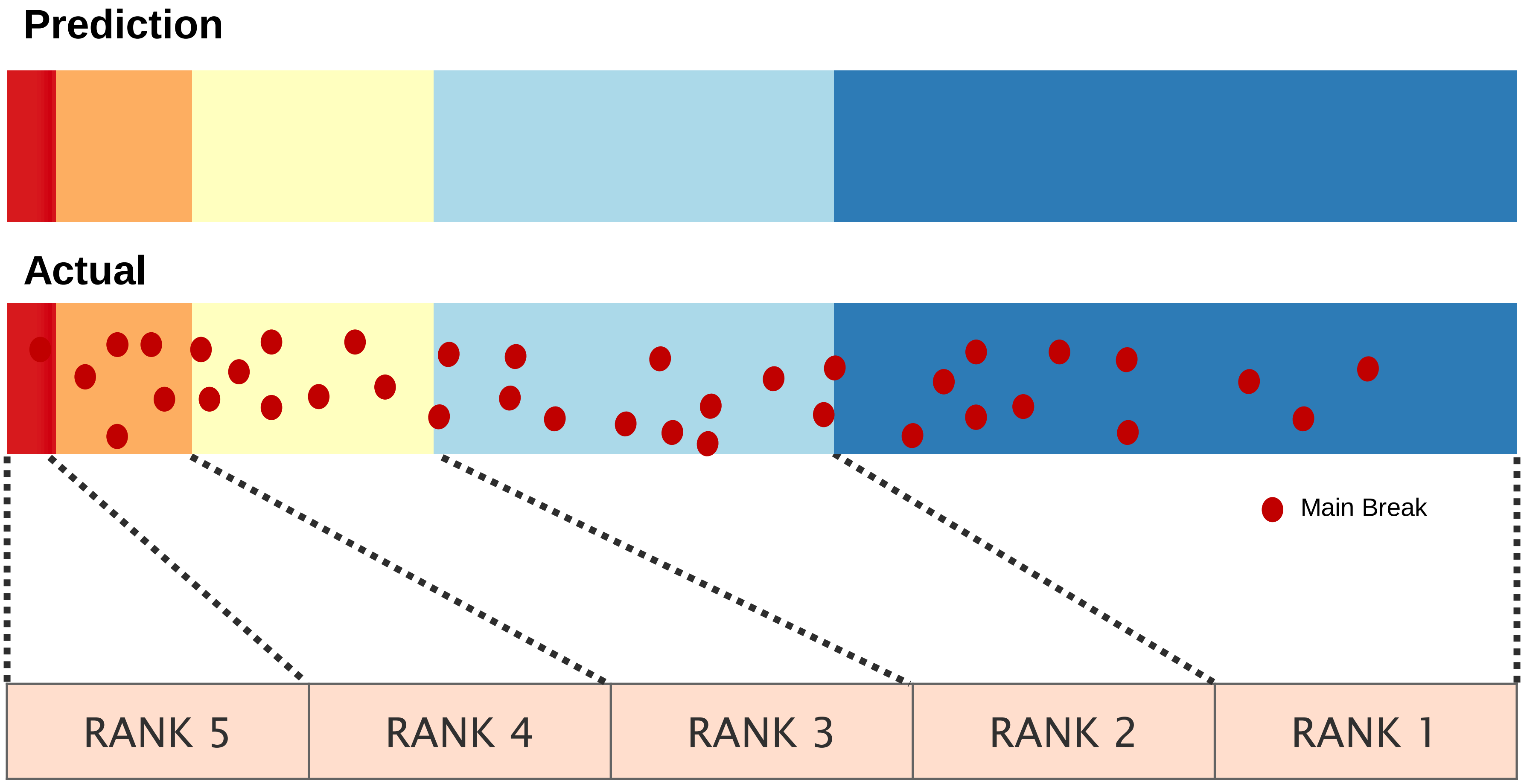
Cleaned Pipe Asset Data | Cleaned Break History | **Machine Learning Process** | Pipe Asset with LOF Scores





- Process in a repeatable and scalable manner
- Layer additional variables
- Look for correlations
- Calculate 5-year LOF probability scores

How to Measure LOF Accuracy



Measuring Accuracy

Fracta LOF Focused Ranking	Rank F5	Rank F4	Rank F3	Rank F2	Rank F1
	41.6 miles (1%)	166.3 miles (4%)	415.8 miles (10%)	831.8 miles (20%)	2702.9 miles (65%)
Pipe Segments	979	2692	7151	17462	76245
5-yr forecasted breaks	397	901	1160	1251	1183
5-yr forecasted break rate (x/100mi/yr)	191.0	108.3	55.8	30.1	8.8

Used historical data, 1990 → 2011,
to predict next 5 years LOF

Comparing to Actual Events
Break Data (2012-2016)

EBMUD

Fracta LOF Focused Ranking	Rank F5	Rank F4	Rank F3	Rank F2	Rank F1
	41.6 miles **(1%)**	166.3 miles **(4%)**	415.8 miles **(10%)**	831.8 miles **(20%)**	2702.9 miles **(65%)**
Pipe Segments	979	2692	7151	17462	76245
5-yr forecasted breaks	397	901	1160	1251	1183
5-yr forecasted break rate (x/100mi/yr)	191.0	108.3	55.8	30.1	8.8
Actual Breaks 2012-2016	436	991	1177	1038	838
Actual Break Rate 2012-2016	209.5	119.2	56.6	25.0	6.2

Broad Correlation of Projected & Actual break rates

Demo & Questions