City of Salem
Utility Data Health

Caleb Anderson | Phil Chapman | Anastasiya Grecu

Is your data A+?

• Previously state of data based on qualitative approach

• Need to quantify state of our data and identify deficiencies

• End-goal is great data

The Migration

- ArcInfo Coverages to SDE
- Hansen 7 to Infor Public Sector 8
- Data managed by many people over the last 20 years with limited documented workflows

The Situation

No surprise, there are issues with the data...

• Orphans between GIS and asset management

• Coverages now in Geometric network (flow direction, snapping, multi-part features, etc.)

• Spatial accuracy (average 10 ft off based on field testing)

• How are we supposed to know if our data is "good" or "bad"?!?

City of Salem   Utility Data Health

Is your data A+? The Solution The Metrics The Reports The Visualization The Dashboard The Future The Conclusion   A Story Map

The Solution

- Analyzing data health is not always fun or exciting

- Need to define and weight metrics (checks)

- Create tools (easy, standardized)

- Share results with organization

The Metrics

Sewer Asset Statistics

• 7 Feature Classes

• 107,420 features total

• 8 Check Types in Data Reviewer

• 2 Check Types in Python

• 5-10 Checks per Feature Class

• 46 Checks in Data Reviewer

• 80 Total Checks and growing

Groups of Checks

• Geometry and Polyline
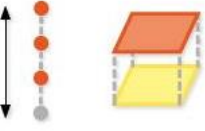
• Database Validation

• Table

• Custom (using Python)

City of Salem   Utility Data Health

Is your data A+?   The Solution   The Metrics   The Reports   The Visualization   The Dashboard   The Future   The Conclusion   A Story Map

The Metrics

field in the Reviewer table. The severity rating allows you to indicate how important the results from a check are in terms of your quality assurance/quality control processes. The lower the number, the greater the priority the check's results have.

Geometry Checks

· Duplicate Geometry

Returns geometries from features in Feature Class 1 that are co-located with geometries in Feature Class 2. Optionally, features must share identical non-spatial attributes with the option to ignore PLTS metadata.
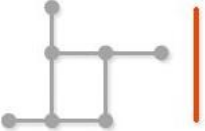
· Invalid Geometry

Returns features with invalid geometry. This includes null geometries, empty geometries, features with empty envelopes, and non-simple geometries.

· Orphan Check

Finds polyline features that are not connected to other polyline, polygon or point features.

▸ Polyline checks

▸ Spatial Parameter Evaluation checks

▸ Table checks

▸ Z-value checks

▸ Working with composite checks

1   Start ArcMap.

2   On the main menu, click **Customize > Toolbars > Data Reviewer**.

3   Click the **Select Data Check** drop-down arrow on the **Data Reviewer** toolbar, click the plus sign (+) next to **Duplicate Geometry Checks**, then click **Duplicate Geometry Check**.
The **Duplicate Geometry Check Properties** dialog box appears.

4   If necessary, type a unique name for the check in the **Check Title** text box.

The Metrics

provide a more speci~~fic description for the feature that has been~~ written to the Reviewer table and are copied to the Notes field in the Reviewer table. The severity rating allows you to indicate how important the results from a check are in terms of your quality assurance/quality control processes. The lower the number, the greater the priority the check's results have.

Domain Checks

Check Description

LANDUSE	DESCRIPTION
000	UNCLASSIFIED
AGR	AGRICULTURE
SDP	PLAN
IND	INDUSTRIAL

Returns rows or features that violate domain constraints.

1   Start ArcMap.

2   On the main menu, click **Customize > Toolbars > Data Reviewer.**

3   Click the **Select Data Check** drop-down arrow on the **Data Reviewer** toolbar, click the plus sign (+) next to **Database Validation Checks**, then click **Domain Check.**
    The **Domain Check Properties** dialog box appears.

Domain Check Properties dialog box

4   If necessary, type a unique name for the check in the **Check Title** text box.

▮ Note: The check title can be used to describe the conditions you are looking for with the check. This

The Metrics

Checking for unique IDs

Comparing table attributes

Finding features with a SQL query

▸ Z-value checks

▸ Working with composite checks

Unique ID Check

•Null Values •Unique Values

Check Description

FEATURE CLASS 1

7	25
8	18
9	18

FEATURE CLASS 2

1	36
2	21
3	18

Checks the values of a set of fields across a set of tables and feature classes within a given workspace for uniqueness within each field

3   Click the **Select Data Check** drop-down arrow on the **Data Reviewer** toolbar, click the plus sign (+) next to **Table Checks**, then click **Unique ID Check**.

The **Unique ID Check Properties** dialog box appears.
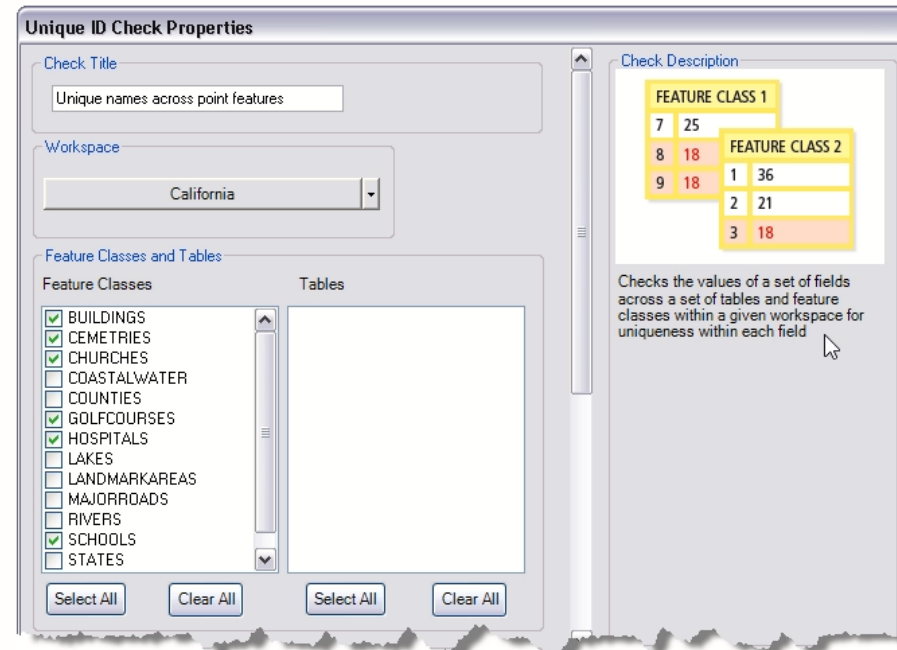
Unique ID Check Properties

Check Title

Unique names across point features

Workspace

California

Feature Classes and Tables

Feature Classes

☑ BUILDINGS
☑ CEMETRIES
☑ CHURCHES
☐ COASTALWATER
☐ COUNTIES
☑ GOLFCOURSES
☑ HOSPITALS
☐ LAKES
☐ LANDMARKAREAS
☐ MAJORROADS
☐ RIVERS
☑ SCHOOLS
☐ STATES

Tables

Select All   Clear All   Select All   Clear All

Check Description

FEATURE CLASS 1

7	25
8	18
9	18

FEATURE CLASS 2

1	36
2	21
3	18

Checks the values of a set of fields across a set of tables and feature classes within a given workspace for uniqueness within each field

4   If necessary, type a unique name for the check in the **Check Title** text box.

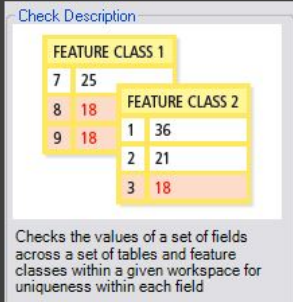▌ Note: The check title can be used to describe the conditions you are looking for with the check. This is useful when you have multiple instances of the same check to validate the same feature classes or tables but with different validation parameters.

City of Salem   Utility Data Health

Is your data A+? The Solution The Metrics The Reports The Visualization The Dashboard The Future The Conclusion   A Story Map

The Methodology
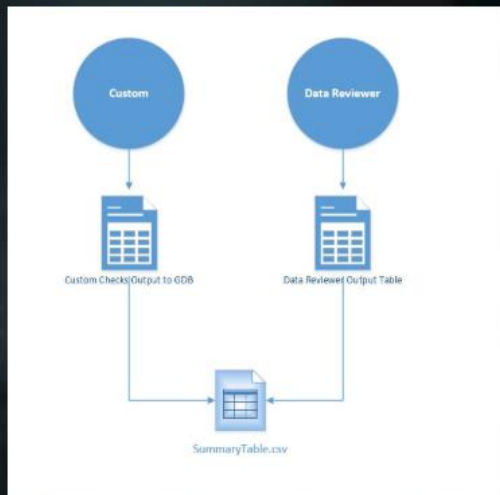
Score Factors

• Number of Features

• Check Criticality

The Reports

Output: SummaryTable.csv

• Derived from Data Reviewer output table and custom check output using Python.
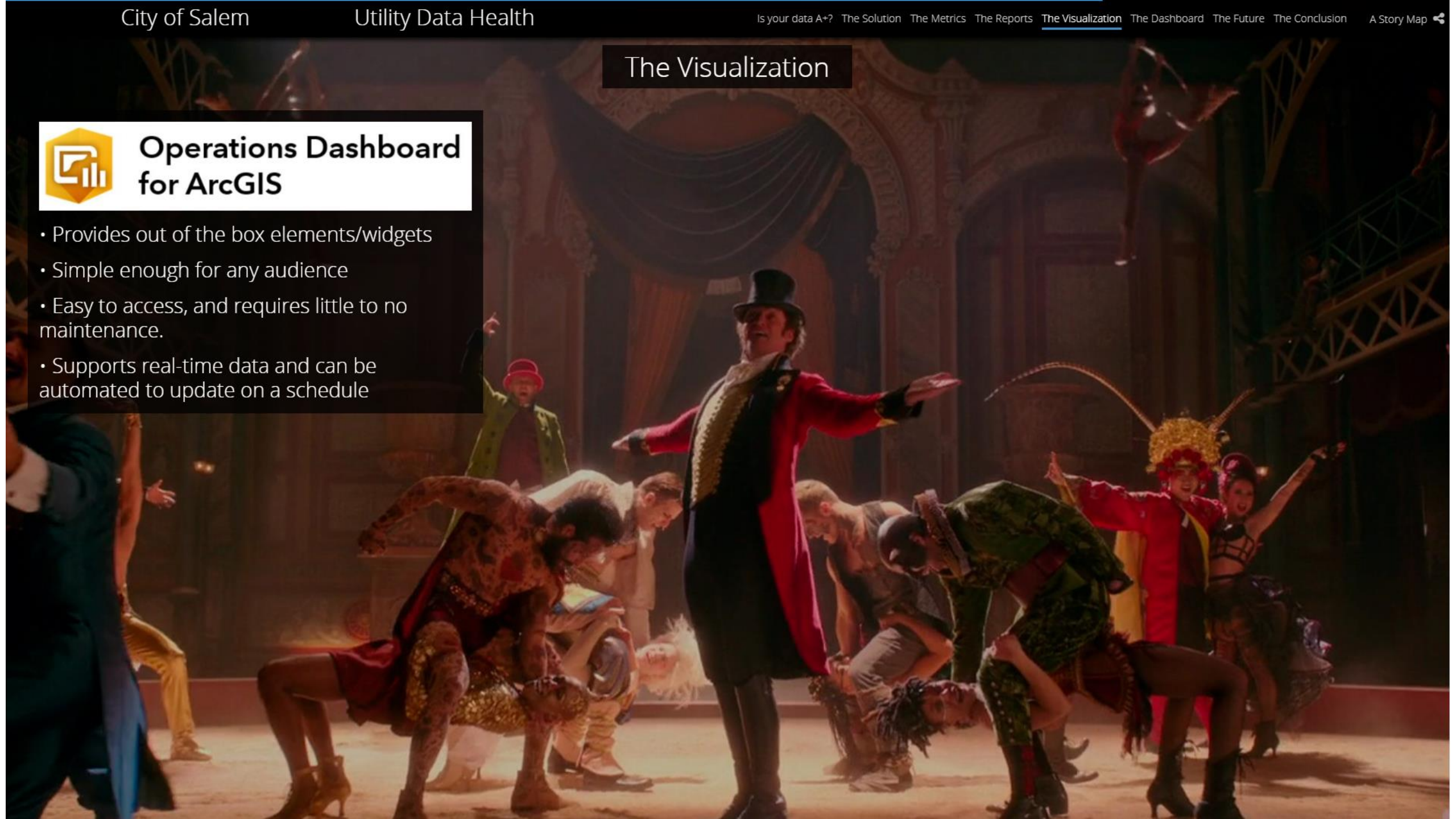

• Table is then hosted in ArcGIS Online where we have the ability to overwrite and feed our operations dashboard.

The Visualization

Operations Dashboard for ArcGIS

- Provides out of the box elements/widgets

- Simple enough for any audience

- Easy to access, and requires little to no maintenance.

- Supports real-time data and can be automated to update on a schedule

City of Salem   Utility Data Health

Is your data A+?   The Solution   The Metrics   The Reports   The Visualization   **The Dashboard**   The Future   The Conclusion   A Story Map

1 2

Sewer Utilities Totals ☰

Vaults
🟧V **76**
Active, City Maintained

Catch Basins
⬚ **9**
Active, City Maintained

Manholes
⊕ **15,944**
Active, City Maintained

City Maint. | Not City Maint.

Lift Stations
🟧LS **29**
Active, City Maintained

Lift Stations | Wet Wells | Pumps

Complexes
✉ **1**
Active, City Maintained

Complexes | Facilities

Miles of Mains
— **347.96**
Active, City Maintained

Cleanouts/Lampholes
🟠 **1,914**
Active, City Maintained

City Maint. | Not City Maint.

Valves
⧖ **146**
Active, City Maintained

Valves | Valve Box
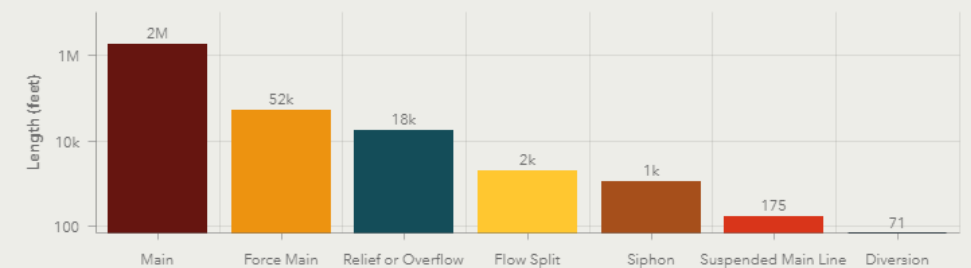
◁ Mains ▷

Mains Installed Per Year (50 Year)

No. of Mains

1k
500

1970   1980   1990   2000   2010

🟠 Main   🔵 Relief or Overflow   🟢 Force Main   🔴 Siphon   🟢 Flow Split
🟣 Suspended Main Line   🔵 Diversion
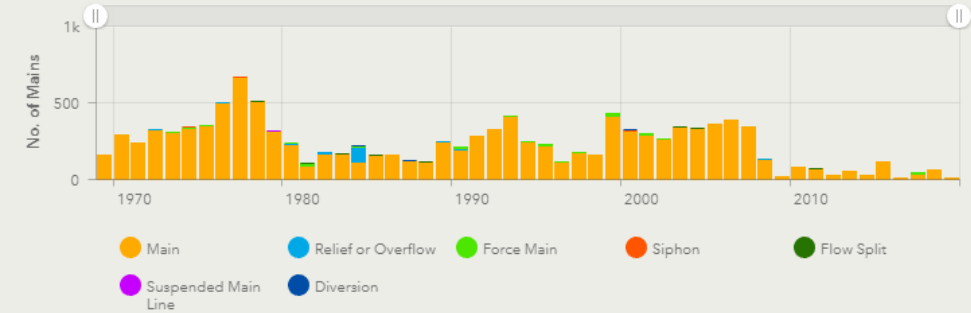
Mains Installed Per Year | Structures Installed Per Year

Pipe Types (Active, City Maintained)

🔵 Unknown 4.52%
🔴 Asbestos Cement 1.35%
🟣 Null 1.47%
🟠 Concrete Pipe 41.75%
🔴 Ductile Iron Pipe 1.34%
⚫ Polyvinyl Chloride 40.21%
🟢 Reinforced Concrete Pipe 7.24%

Length of Acive Mains by Type

Length (feet)

1M
10k
100

2M
52k   18k
2k   1k   175   71

Main   Force Main   Relief or Overflow   Flow Split   Siphon   Suspended Main Line   Diversion

Last update: a few seconds ago

City of Salem   Utility Data Health

Is your data A+?   The Solution   The Metrics   The Reports   The Visualization   The Dashboard   The Future   The Conclusion   A Story Map

1 2

Data Health Summary I Sewer Assets

☰

Data Health Score

A

Total Errors

5,914

107,401
Total Sewer Assets

Overall Data Health (%)

97.95

40 60
20   80
0   100
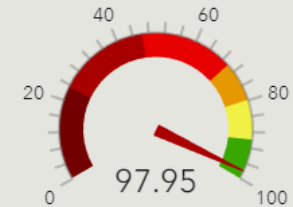
Error Counts

◁ 1 of 50 ▷

Data Health Check Statistics

Feature PW.ssNode
Check DR - Null values, key
Error Count 1,139.00
Feature Total 2,127.00
Percent Total 53.55
Weighted 28.68
Count Percent
Final Percent 71.32
Weight 3.00

Mains
— 3,245

Manholes
425

Valves
⧖ 63

Treatment Plants
T 0

Services
··· 790

Pumps
⬡ 0

Lift Stations
LS 2

Nodes
• 1,389

GIS and IPS Orphans
939

Geometry Checks
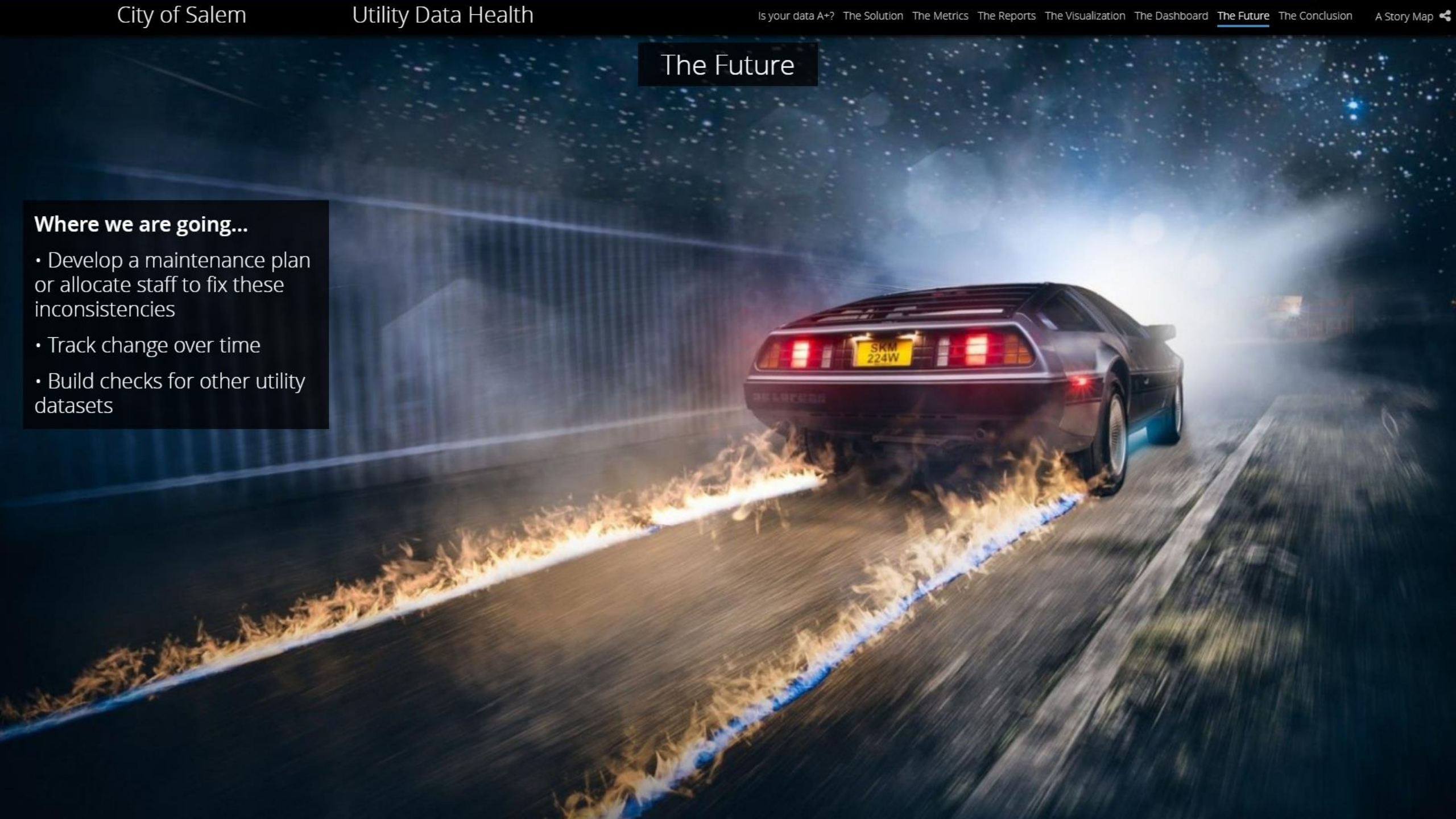73

Null Value Checks
2,751

Unique ID Check
1,142

The Future

Where we are going...

· Develop a maintenance plan or allocate staff to fix these inconsistencies

· Track change over time

· Build checks for other utility datasets

The Conclusion

• State of data was based on an approximation/gut feeling but is now quantifiable

• Deficiencies have been identified and can be fixed

• Tools developed for ease of analysis

• Results are shared with organization

• Still more to do (other data sets, spatial analysis, track change over time, etc.)

Image and Movie Credits

The Cowboys (1972)

Annie (1982)

Princess Bride (1987)

The Matrix (1999)

The Hitchhiker's Guide to the Galaxy (2005)

Minority Report (2002)

The Greatest Showman (2017)

Back to the Future (1985)

The Good, the Bad and the Ugly (1966)

City of Salem Capitol | https://traveloregon.com/