# Data Wrangling
## Messy Hydrologic Data Transformed for Decision Making
## Kendri Salmans

# Importance of Water Data






GRACE-Based Shallow Groundwater Drought Indicator
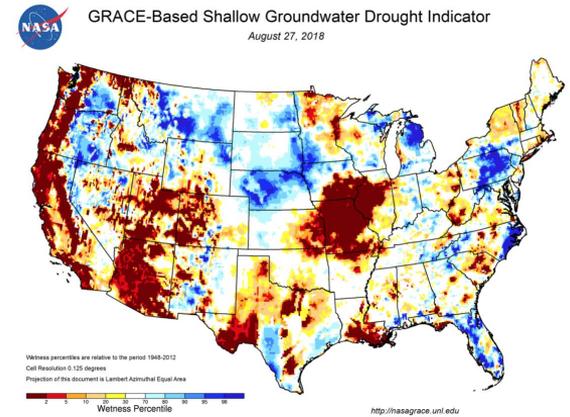August 27, 2018

## GROUNDWATER IS CRITICAL

67% irrigation supply
51% drinking supply

## RELIANCE GROWING

Growing population
Industry & Agricultural
Use

## RESOURCE CONFLICT

Ag. vs. Consumer vs. Business

WellInt

# Importance of Water Data

| RISK MANAGEMENT | SUSTAINABILITY | DISPUTE RESOLUTION |
|---|---|---|

**RISK MANAGEMENT**

- Protect harvest
- Plan operations
- Mitigate financial impact

**SUSTAINABILITY**

- Metrics planning
- Water footprint management
- ESG reporting

**DISPUTE RESOLUTION**

- Regulatory requirements
- Community dispute
- Lawsuit protection

WellInt

# Data Wrangling

| Collection Methods | Data Types | Data Formats | Data Providers |
|---|---|---|---|
| Manual Measurements | Precipitation | Handwritten | Internal Observations |
| Analog Meters | Groundwater Withdrawal | Local Storage | Public Database |
| Transducers | Surface Water Flow | Databases | Private/Restricted Access |
| Models | Groundwater Level | SCADA | |
| | Temperature | | |
| | Soil Moisture | | |

WellInt

# Data Wrangling
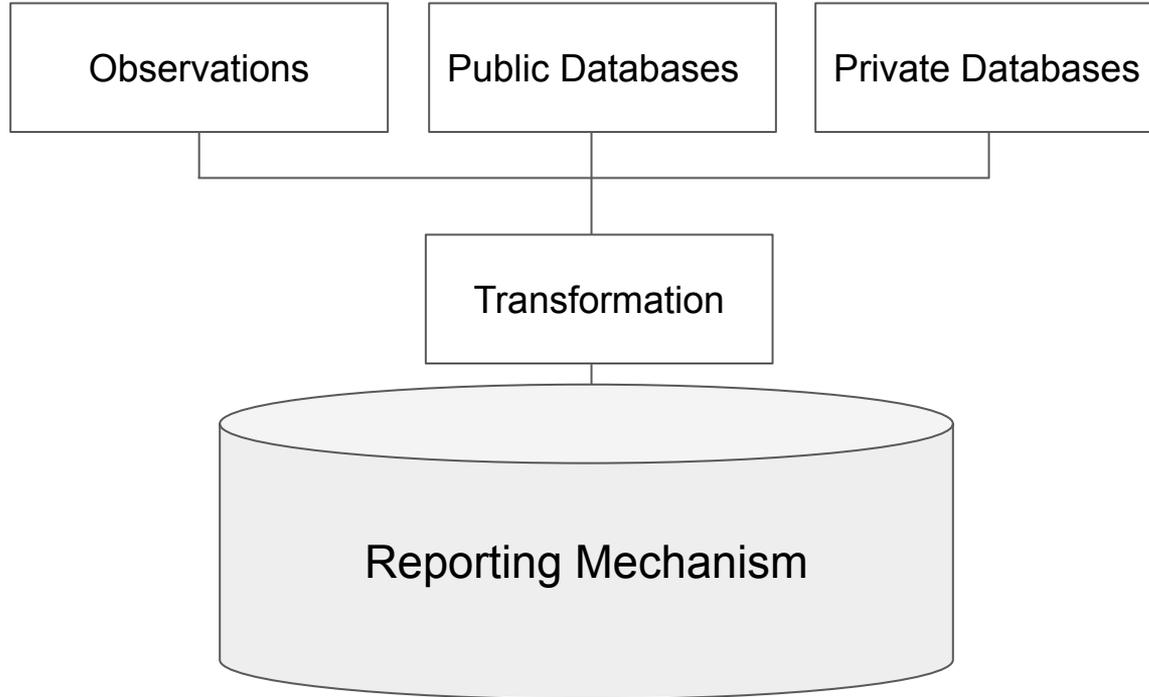
# Tidy Data

| Location | Value | Date | Data Type | Data Source | Unit Type |
|---|---|---|---|---|---|
| Well 1 | 50 | 8/30/23 | Groundwater Level | Pressure Transducer | PSI |
| Well 2 | 60 | 5/10/23 | Groundwater Level | Manual Measurement | Feet Below Ground Surface |
| Flume | 2 | 10/15/20 | Surface Flow | 4 Inch Cutthroat | Inches above ground surface |
| Rain Gauge | 10 | 6/17/18 | Precipitation | NOAA API | Inches |

Rules
1) Each Variable must have its own column
2) Each observation must have its own row
3) Each value must have is own cells

# Tidy vs Messy Data

| Location | Value | Date | Data Type | Data Source | Unit Type |
|---|---|---|---|---|---|
| Well 1 | 50 | 8/30/23 | Groundwater Level | Pressure Transducer | PSI |
| Well 2 | 60 | 5/10/23 | Groundwater Level | Manual Measurement | Feet Below Ground Surface |
| Flume | 2 | 10/15/20 | Surface Flow | 4 Inch Cutthroat | Inches above ground surface |
| Rain Gauge | 10 | 6/17/18 | Precipitation | NOAA API | Inches |

## Messy Data

| Monitoring Wells | | | | | | Surface Flow | | | Precipitation | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Well 1 | | | Well 2 | | | Flume | | | Rain gauge | | |
| Data Source | Value | Date | Data Source | Value | Date | Data Source | Value | Date | Data Source | Value | Date |
| Pressure Transducer | 50 | 8/30/23 | Manual Measurement | 60 | 5/10/23 | 4 Inch Cutthroat | 2 | 10/15/20 | NOAA | 10 | 6/17/18 |

Multiple Columns per Variable        Multiple Observations per Row

# Sharing Data

| Observations | Public Databases | Private Databases |
|---|---|---|

## Data Processor

| Models | Reports | Databases |
|---|---|---|
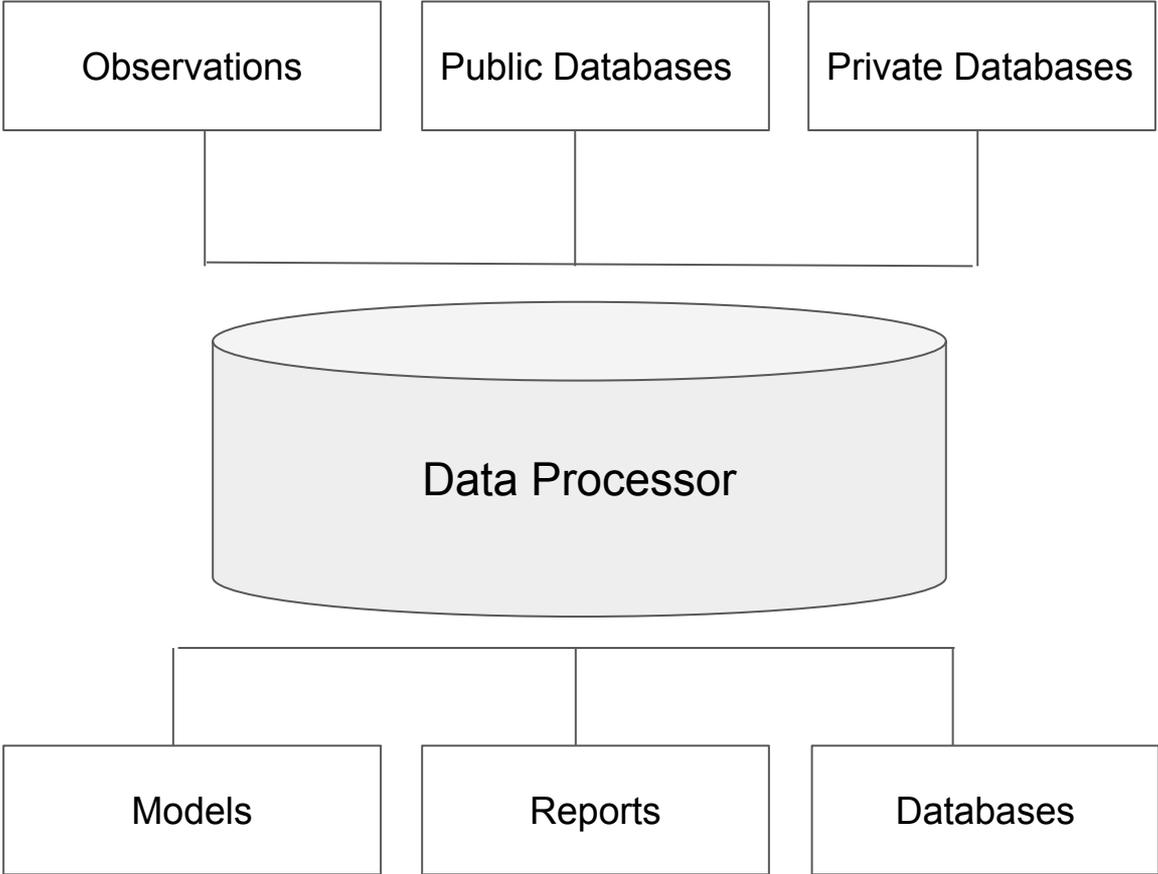
# Visualizing Dense Datasets
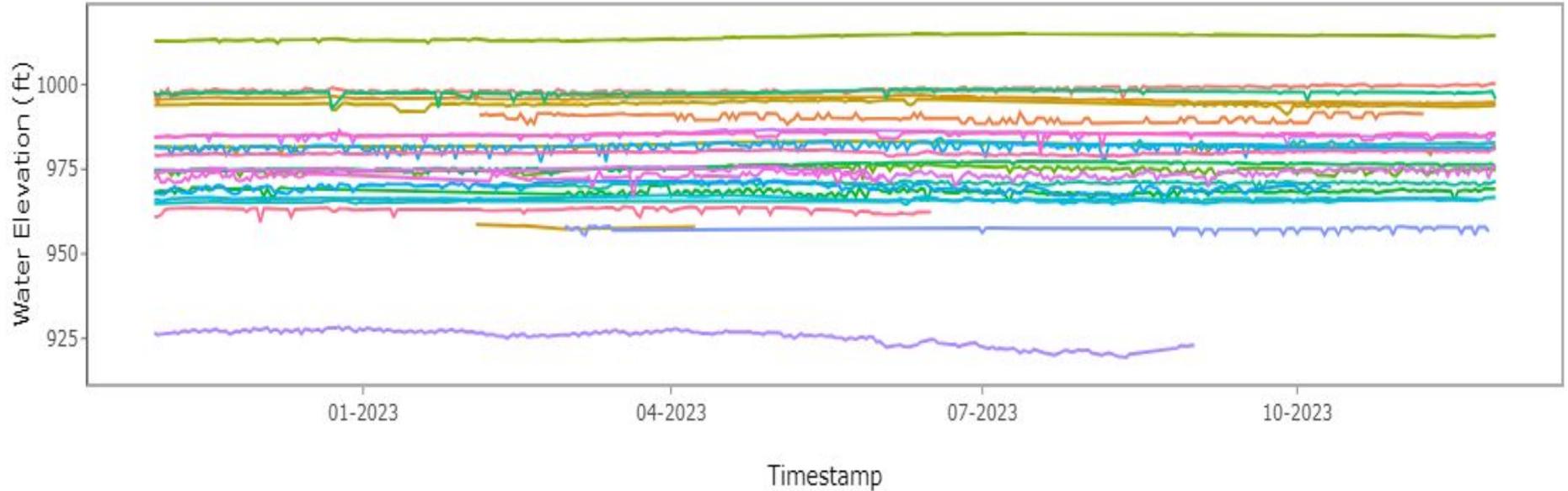
# Minimized Water Risk

**Public Agencies**

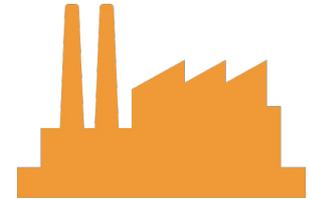Inform permitting and growth decisions

**Agriculture & Dairy**

Minimize operating risk and maximize output

**Water Utilities**

Protect operating assets, ensure production

**Industry**

Simplify compliance reporting and inform planning

# Questions?