



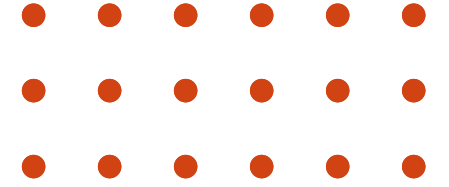
# Do More with Less: Low-Cost Tools for Public Health Analytics, Surveillance & AI

BY NACCHO



SHAPING TOMORROW'S PUBLIC HEALTH TODAY.

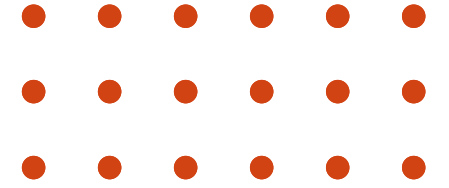




# Agenda

---

- Welcome
- Funding Acknowledgment
- About NACCHO
- Low-Cost Tools for Public Health Analytics, Surveillance & AI
- Q&A
- Evaluation
- Closing



**Recording:** This meeting is being recorded. A link to the recording will be posted on NACCHO's website.



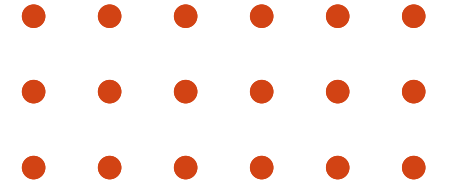
**Audio:** All participant should mute their microphones to minimize background noise. Please remain muted and raise your hands to ask questions.



**Questions:** We encourage your questions! Use chat feature to drop your questions or raise your hand to come off mute. Questions will be addressed during the designated Q&A portion or in the chat.



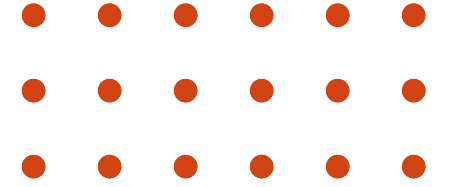
**Technical Issues:** If you experience any technical difficulties, try refreshing your browser or exiting and rejoining the session. You can also reach out in the chat for assistance.



# Funding Acknowledgement

---













This work is supported by funds made available in whole from the Centers for Disease Control and Prevention (CDC) of the U.S. Department of Health and Human Services (HHS), National Center for STLT Public Health Infrastructure and Workforce, through OE22-2203: Strengthening U.S. Public Health Infrastructure, Workforce, and Data Systems grant. The contents are those of the author(s) and do not necessarily represent the official views of, nor an endorsement, by CDC/HHS, or the U.S. Government.

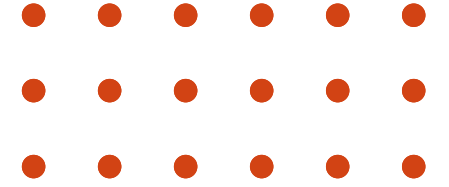


# About NACCHO

---

NACCHO is comprised of nearly **3,300 local health departments (LHDs)** across the United States. Our mission is to serve as a **leader, partner, catalyst, and voice** with local health departments.

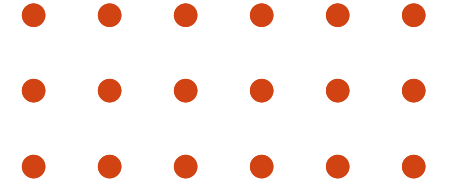
-   Advocacy
-   Partnerships
-   Funding
-   Training and Education
-   Networking
-   Resources, tools, and technical assistance



# Disclaimer

---

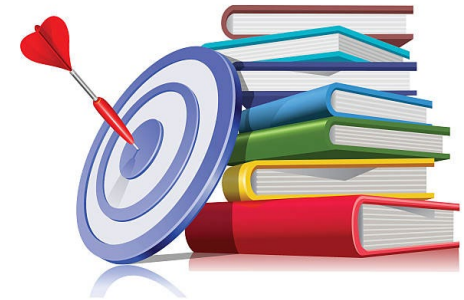
Reference to any specific company, products, processes, tools or services by trade name, trademark, manufacturer, or otherwise on this webinar does not necessarily constitute or imply its endorsement or recommendation by the National Association of County and City Health Officials (NACCHO) or NACCHO's funders.



# Learning Objectives

---

- Identify low-cost tools for public health analytics, surveillance, and AI
- Assess tools based on key features, best practices, and pros/cons
- Identify tools that align with your health department needs and capacity

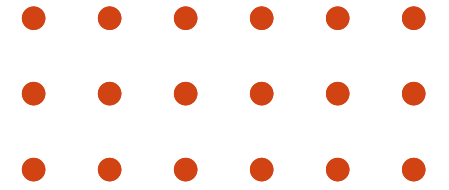


# Low-Cost Solutions

---

# Transforming Surveillance and Informatics for Local Health Departments **Post-Convening Report**

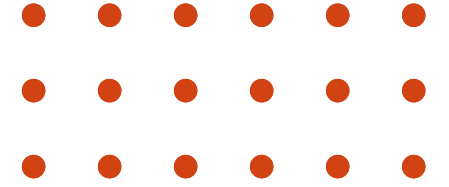
## Context



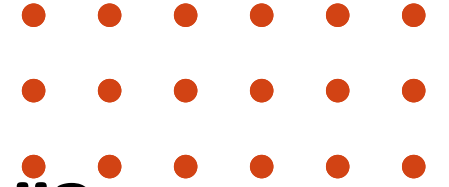
In 2025 NACCHO held a two-day workshop focused on increasing capacity of local public health departments through:

- Peer learning
- Presentations
- Group discussions





**What comes to your mind when you think of  
“low-cost solutions”?**



# What do we mean when we say, “low-cost tools”?

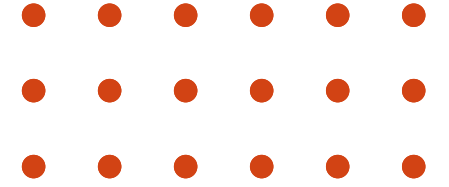
- Tools that are free
- Tools that come with existing licenses
- Tools with minimal implementation costs
- Tools that do not require dedicated IT team to maintain

## Why it matters:

- **Affordability**
- **Accessibility**
- **Simplicity**
- **Functional Alternative**

# Data Collection/ Electronic Data Capture

---



# REDCap – Research Electronic Data Capture

---

- **What it is:**

- REDCap is a secure, **web-based** platform for building and managing online surveys, patient registries, and clinical data collection instruments.

- **Why it's low cost:**

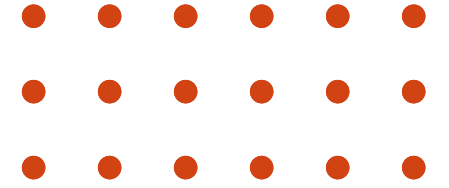
- REDcap software is free for non-profit (501 c3 organizations, including government or NGO institutions).

- **Used for:**

- Chronic disease patient registries
- Community health needs assessments
- Contact tracking & disease surveillance data capture
- Longitudinal cohort follow-up and tracking



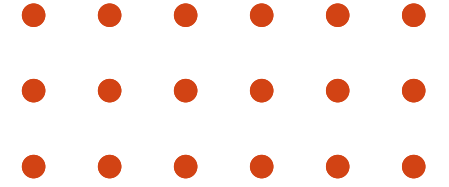
<http://dx.doi.org/10.1002/ce.10001>



# REDCap – Research Electronic Data Capture

---

- **Best Practices:**
  - Enable required fields to reduce missing data
  - Use REDCap's API to connect to visualization tools for automated reporting pipelines
  - Assign role-based access to limit PHI access to only the staff who need it
- **Pros**
  - Designed to be compliant with HIPAA, FISMA and 21 CFR Part 11
  - No coding required for most use cases (drag and drop from builder)
- **Cons**
  - Requires institutional server infrastructure to host (IT involvement needed)
  - Limited native visualization, requires exporting to another tool for reporting.

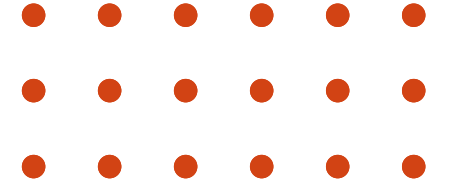


# KoBoToolbox

---

- **What it is:**
  - KoBoToolbox is an open-source platform for collecting, managing, and visualizing data.
- **Why it's low cost:**
  - Free tier for nonprofit and government agencies. Paid version also available for organizations with high volume data needs or tailor-made solutions.
  - Open-source version can be self-hosted.
- **Used for:**
  - Community health needs assessments and field surveys in rural or underserved areas
  - Rapid needs assessments during public health emergencies
  - Environmental health inspections and community screening events
  - Program monitoring and evaluation data collection

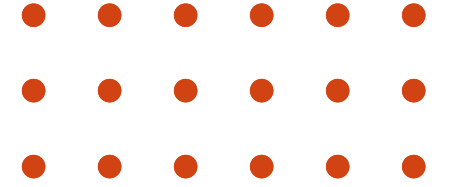




# KoBoToolbox

---

- **Best practices:**
  - Use GPS question type to auto-capture coordinates
  - Export data to CSV or SPSS for analysis
  - For PHI, use the self-hosted option (KoBoToolbox is fully open-source on GitHub)
- **Pros:**
  - Easy to use, able to design survey in minutes with minimal training
  - Supports GPS location capture, photo, audio, and video in surveys
  - Multi-language support for diverse community populations
- **Cons:**
  - Limited HIPAA compliance features, as such not ideal for clinical PHI collection
  - Free tier data stored on KoBo servers (not self-hosted) as such, need to consider data security.



# ODK (Open Data Kit)

---

- **What it is:**

- Open Data Kit (ODK) is an open-source suite of tools for mobile data collection. ODK Central serves as the server, while ODK Collect is the Android app used in the field.

- **Why it's low cost:**

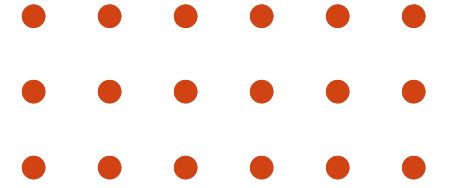
- ODK is free when self-hosted and the ODK Collect app is free on Android.
- Paid option available for those who need services for hosting, managing, and support.

- **Used for:**

- Field-based chronic disease screening events (blood pressure, BMI, glucose checks)
- Environmental health and food safety inspection data collection



<http://www.opendatakit.org>



# ODK (Open Data Kit)

---

- **Best Practice:**

- Use XLSForm format for survey design. It's portable and version-controllable
- Connect ODK Central to Power BI or Tableau via REST API for automated dashboards

- **Pros:**

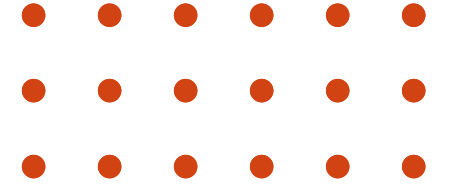
- Surveys built in Excel & highly customizable
- ODK Collect works on low-cost Android devices (smartphones or tablets)

- **Cons:**

- Requires more technical setup than other tools and has a higher learning curve (e.g., KoBoToolbox)

# Statistical Analysis

---



# R / RStudio

---

- **What it is:**

- An open-source programming language and integrated development environment used for statistical computing, data analysis, and visualization.

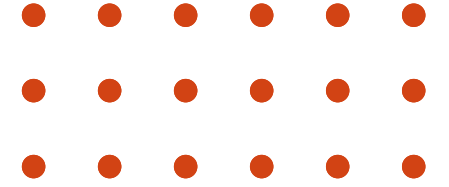
- **Why it's low cost:**

- Completely free and open-source, with no licensing fees and a large community providing free packages and support



- **Used for:**

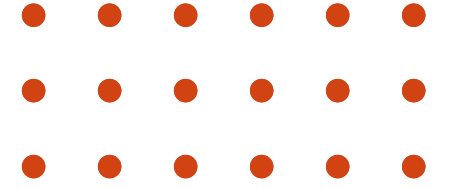
- Field-based chronic disease screening events (blood pressure, BMI, glucose checks), environmental health and food safety inspection data collection, vital statistics and birth/death registration in resource-limited settings, syndromic surveillance and outbreak investigation data collection



# R / RStudio

---

- **Best practices:**
  - Use reproducible workflows (R scripts, R Markdown, version control with Git)
  - Leverage vetted packages (e.g., tidyverse) and document assumptions
- **Pros:**
  - Free, open-source, and widely used in public health and academia
  - Strong ecosystem for advanced analytics, visualization, and reporting
- **Cons:**
  - Steeper learning curve for non-programmers
  - Memory-based processing can limit performance with very large datasets



# Python

---

- **What it is:**

- An open-source programming language widely used for data analysis, statistics, and machine learning, supported by powerful libraries like pandas, statsmodels, and scikit-learn.

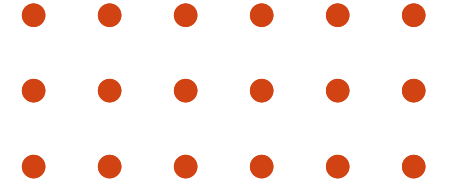
- **Why it's low cost:**

- Free and open-source with extensive libraries that eliminate the need for costly proprietary analytics software

- **Used for:**

- Data cleaning, transformation, and statistical analysis (pandas, statsmodels)
- Machine learning and predictive modeling (scikit-learn)

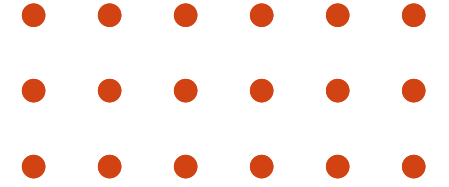




# Python

---

- **Best practices:**
  - Use reproducible workflows (Jupyter notebooks, version control with Git)
  - Structure code with clear documentation and validate models rigorously
- **Pros:**
  - Highly versatile across analytics, AI, and automation tasks
  - Large global community with strong documentation and support
- **Cons:**
  - Requires programming knowledge and setup time
  - Performance tuning may be needed for very large datasets



# PSPP

---

- **What it is:**

- An open-source statistical analysis software designed as a free alternative to IBM SPSS Statistics for basic and intermediate analytics.

- **Why it's low cost:**

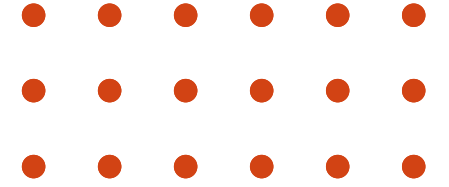
- Completely free and open-source, eliminating expensive proprietary licensing fees

- **Used for:**

- Descriptive statistics, hypothesis testing, and survey data analysis
- Data entry and analysis with a familiar SPSS-like interface



<https://www.gnu.org/software/pspp/pspplogo.png>



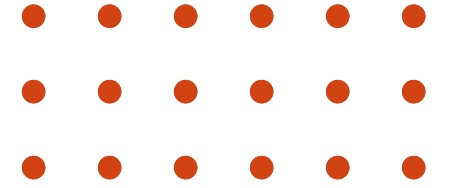
# PSPP

---

- **Best practices:**
  - Validate outputs against known methods when transitioning from SPSS
- **Pros:**
  - User-friendly for those already familiar with SPSS
  - No cost
- **Cons:**
  - Limited advanced analytics and modeling capabilities
  - Smaller user community and fewer updates compared to SPSS

# Geospatial Analysis

---



# QGIS

---

- **What it is:**



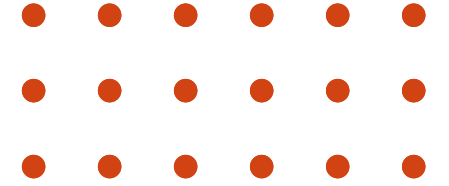
- QGIS is a free, open-source desktop GIS (Geographic Information System) application for creating, editing, visualizing, and analyzing geospatial data.

- **Why it's low cost:**

- Completely free and open-source. Runs on Windows, Mac, and Linux.

- **Used for:**

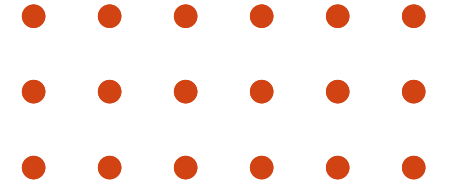
- Identifying food deserts, park access gaps, and social determinants of health
- Visualizing healthcare service locations relative to high-need populations



# QGIS

---

- **Best Practice**
  - Use the QuickMapServices plugin to add free basemaps
  - Export final maps to PDF or image for reports; use QGIS2Web plugin to create shareable web maps
- **Pros**
  - Has professional-grade capabilities at zero cost
  - Extensive plugin ecosystem with over 2000 plugins
  - Active Community with free tutorials
- **Cons**
  - Initial training investment required
  - Desktop only application
  - Performance can lag with very large datasets

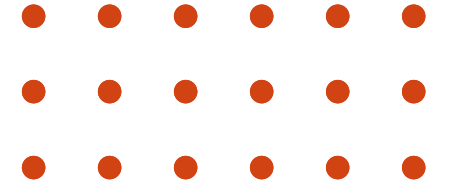


# GRASS GIS — Geographic Resources Analysis Support System

---

- **What it is:**
  - GRASS GIS is a powerful, open-source software designed for advanced geospatial analysis, raster processing, environmental modeling and visualization.
- **Why it's low cost:**
  - Completely free and open-source.
- **Used for:**
  - Disease cluster detection and hotspot modeling
  - Environmental health analysis: air quality, water access, land use exposure modeling





# GRASS GIS — Geographic Resources Analysis Support System

---

- **Best Practice:**

- Use GRASS GIS as a QGIS plugin rather than standalone. This reduces the learning curve significantly
- Use the OSGeo community forum to explore use cases or ask questions
- Pair with R's 'rgrass' package for statistical analysis of GRASS outputs in a familiar environment

- **Pros:**

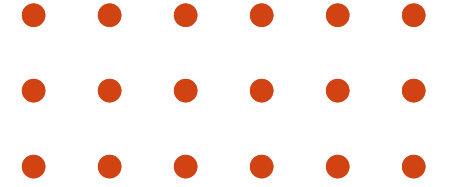
- Runs inside QGIS via the Processing Toolbox or GRASS plugin
- Handles very large datasets efficiently. Has better performance than QGIS for raster-heavy work

- **Cons:**

- Steep learning curve, as such, not recommended for casual GIS users or beginners
- Unnecessary for basic disease mapping tasks (QGIS is better for most county-level needs)

# Data Visualization

---

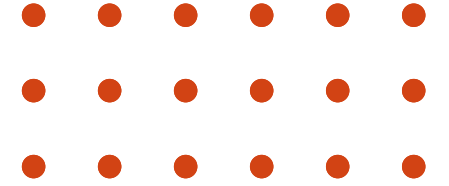


# Looker Studio

---

- **What it is:**
  - Looker Studio is a free, web-based business intelligence and data visualization platform by Google.
- **Why its low cost:**
  - Free for all users with a Google account.
- **Used for:**
  - Chronic disease surveillance dashboards
  - Public-facing health equity and community health indicator dashboards
  - Program performance reporting (vaccination rates, screenings)
  - Connecting and visualizing data from Google Sheets or CSV exports

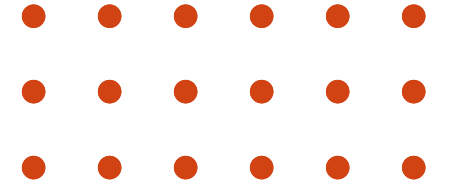




# Looker Studio

---

- **Best Practice:**
  - Use for aggregate, de-identified data only. Never connect to data sources with individual PHI
  - Link to Google Sheets as a live data source. When the Sheet is updated, the dashboard auto-refreshes
  - Use the 'Embed' feature to put dashboards directly on public facing website
- **Pros:**
  - Zero cost, zero installation, works entirely in a web browser
  - Drag-and-drop dashboard builder, limiting coding for standard visualizations
  - Connects natively to Google Sheets.
- **Cons:**
  - Requires Google account which may conflict with agency data governance policies
  - Not HIPAA-compliant in the free tier. Suitable for aggregate data only, not individual-level PHI

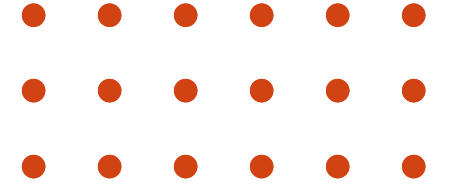


# Apache Superset

---

- **What it is:**
  - Apache Superset is a free, open-source data exploration and visualization platform.
- **Why it's low cost:**
  - Completely free and open-source. Must be self-hosted.
- **Used for:**
  - Chronic disease surveillance dashboards connected to secure databases
  - SQL-based exploration of vital records, or disease registry data
  - Internal team analytics portals for epidemiologists and program staff





# Apache Superset

---

- **Best Practice:**

- Connect directly to your PostgreSQL or MySQL data warehouse for real-time dashboards
- Write simple SQL queries by picking only needed columns.

- **Pros:**

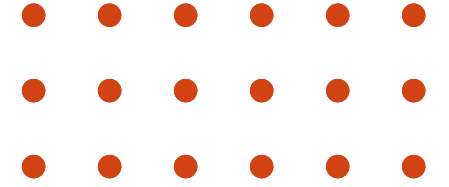
- A **no-code interface** for building charts quickly
- Can be self-hosted in a HIPAA-compliant environment
- Connects to virtually any SQL database: PostgreSQL, MySQL, BigQuery
- Enterprise-grade role-based access control

- **Cons:**

- Requires Docker or cloud deployment. More complex setup than other tools
- Needs IT support for installation and user management

# Artificial Intelligence (AI)

---



# Claude

---

- **What it is:**

- A generative AI assistant developed by Anthropic that supports natural language analysis, coding, and document review.

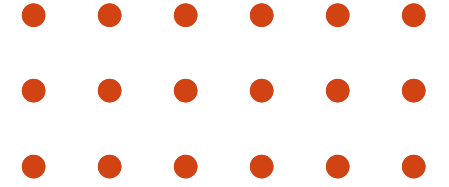
- **Why it's low cost:**

- Offers free or low-cost tiers compared to traditional analytics or consulting resources

- **Used for:**

- Summarizing reports, policies, and surveillance data
- Assisting with code generation, data cleaning logic, and documentation

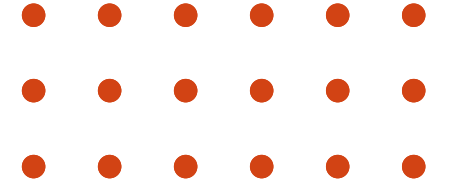




# Claude

---

- **Best practices:**
  - Do not input sensitive or identifiable public health data
  - Validate outputs and use human review for decision-making
- **Pros:**
  - Fast insights from unstructured data (text, reports, notes)
  - Low barrier to entry for non-technical users
- **Cons:**
  - Risk of inaccurate or fabricated outputs (hallucinations)
  - Data privacy and governance considerations



# Hugging Face

---

- **What it is:**

- An open-source platform and community for sharing and deploying machine learning models.

- **Why it's low cost:**

- Many pre-trained models and tools are freely available, reducing development time and cost

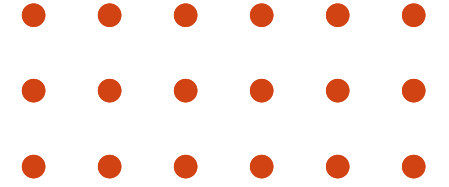


**Hugging Face**

<https://www.github.com/HuggingFace/huggingface>

- **Used for:**

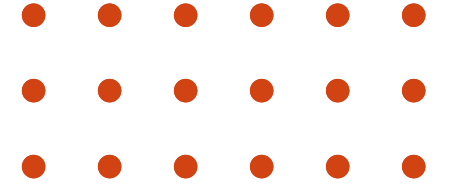
- Natural language processing (e.g., classification, entity recognition)
- Rapid prototyping of AI models for public health use cases



# Hugging Face

---

- **Best practices:**
  - Evaluate model bias and performance on your specific population/data
  - Use secure environments when working with sensitive datasets
- **Pros:**
  - Large library of ready-to-use models and datasets
  - Strong open-source community and documentation
- **Cons:**
  - Requires technical expertise for customization and deployment
  - Some advanced features or hosting options may incur costs



# H2O.ai

---

- **What it is:**

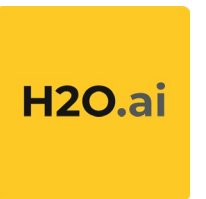
- An open-source AI and machine learning platform from H2O.ai that enables scalable predictive modeling and automation.

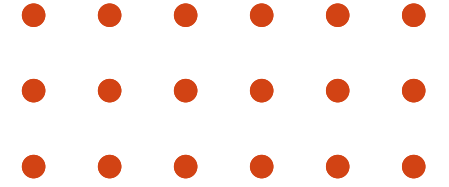
- **Why it's low cost:**

- Core platform is free and open-source, with optional paid enterprise features

- **Used for:**

- Predictive analytics (e.g., outbreak forecasting, risk modeling)
- Automated machine learning (AutoML) for rapid model development





# H2O.ai

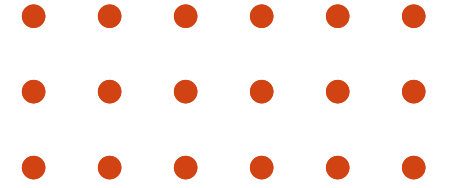
---

- **Best practices:**
  - Start with AutoML for baseline models, then refine with domain expertise
  - Monitor model performance and retrain regularly with updated data
- **Pros:**
  - Scalable for large datasets and complex models
  - Supports automation, reducing need for deep ML expertise
- **Cons:**
  - Setup and integration can require technical skills
  - Advanced enterprise features may require paid licensing

EXAMPLE

# A Health Department Outbreak Investigation Response

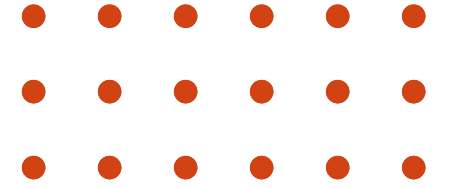
---



# Outbreak Detection and Data Collection (REDCap)

---

- Health department receives reports of gastrointestinal illness linked to a community event
- Initial case definition established; rapid increase in suspected cases triggers investigation
- Need for standardized, rapid data collection across multiple interviewers
- Use REDCap to:
  - Design and deploy a structured case investigation questionnaire
  - Include exposure history (food, location, timing), symptoms, and demographics
  - Enable mobile/offline data entry for field teams
- Result: Centralized, real-time dataset to support timely decision-making

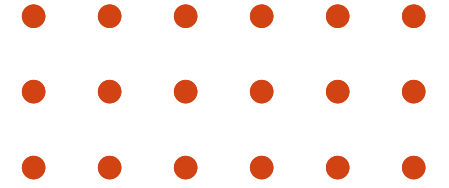


# Descriptive Analysis (REDCap Reports)

---

- Generate basic summaries directly within REDCap
- Use built-in reporting tools to:
  - Calculate frequencies (symptoms, exposures, demographics)
  - Identify most commonly reported food items or locations
  - Construct an epidemic curve and stratify cases by time , place, and person
- Quickly export line lists for situational awareness
- Result: Early signals of likely exposure source and populations at risk

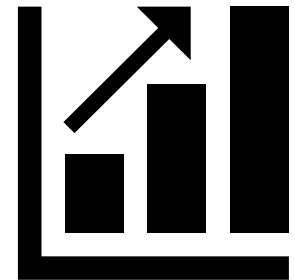


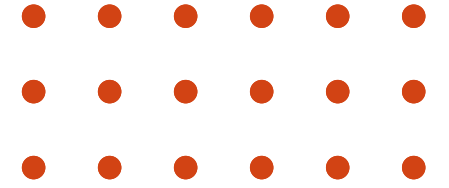


# Statistical Analysis and Hypothesis Testing (R)

---

- Preliminary findings suggest a common exposure, prompting further analysis and investigative actions (coordination with environmental health for potential inspections, collection of samples, situational reports)
- Export cleaned dataset from REDCap into R
- Conduct deeper analysis to test hypotheses:
  - Compare exposure vs. non-exposure groups
  - Calculate odds ratios or relative risks
  - Perform chi-square or Fisher's exact tests
- Assess statistical significance of suspected sources (e.g., specific food item)
- Result: Evidence-based confirmation of outbreak source to guide interventions

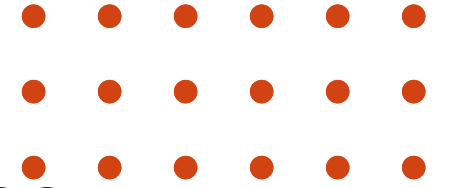




# Statistical Analysis and Cluster Detection (QGIS)

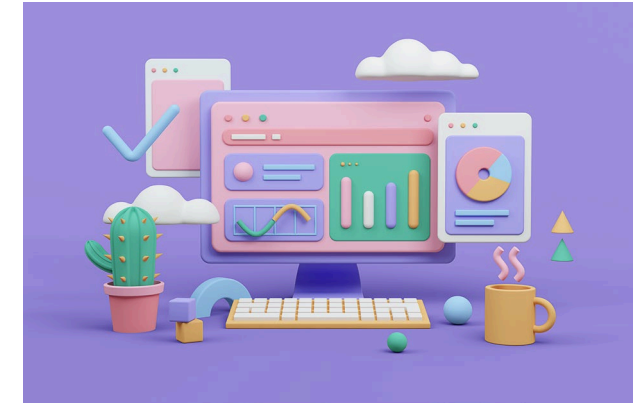
- Import case data with geographic identifiers into QGIS
- Map cases by residence or exposure location
- Conduct spatial analysis to:
  - Identify geographic clustering or hotspots
  - Overlay environmental or facility data (e.g., restaurant locations)
- Visualize spread patterns and potential point-source origin
- Result: Geographic insights to target inspections and control measures



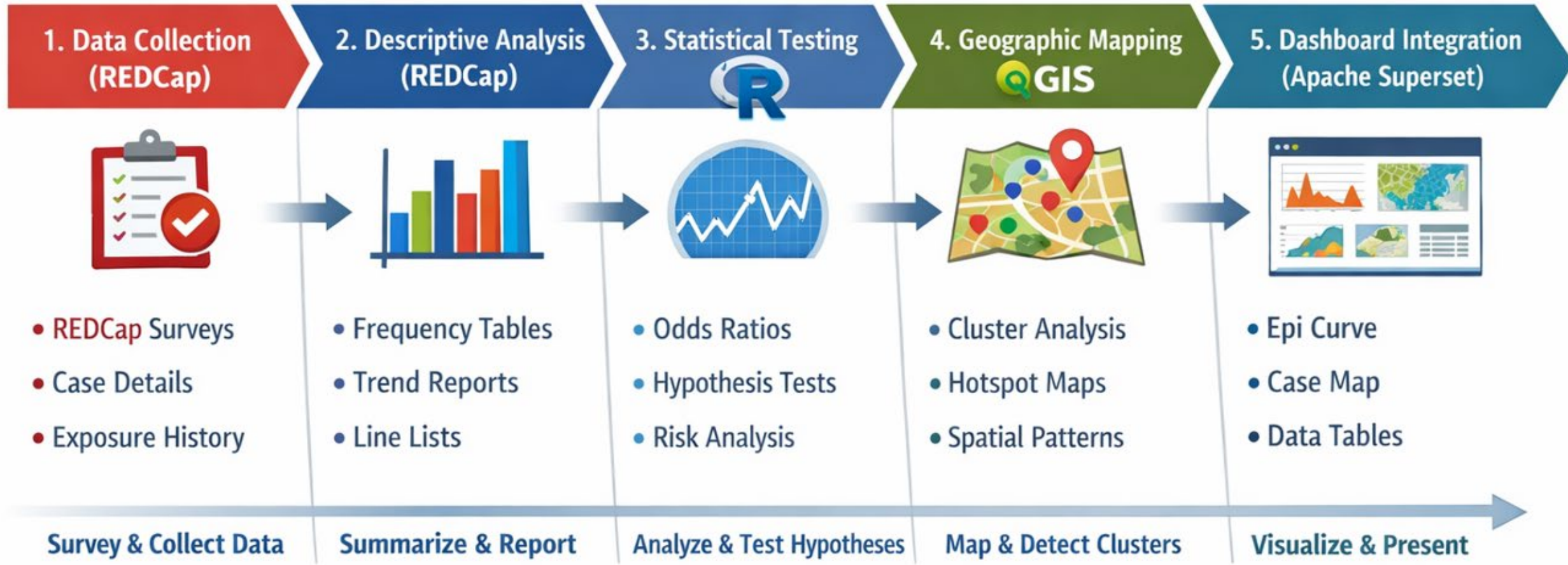


# Integrated Dashboard & Communication (Apache Superset)

- Combine datasets and outputs into Apache Superset
- Build an interactive dashboard displaying:
  - Epidemiologic curve (cases over time)
  - Map of case distribution
  - Table of key exposures and demographics
- Enable real-time updates for leadership and partners
- Support transparent communication and rapid decision-making
- Result: One centralized view for monitoring the outbreak and coordinating response



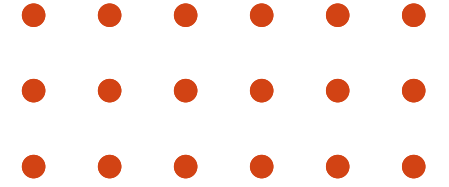
# Outbreak Investigation & Analysis Pipeline



*REDCap captured it. R confirmed it. QGIS mapped it. Superset communicated it.*

# Key Takeaways

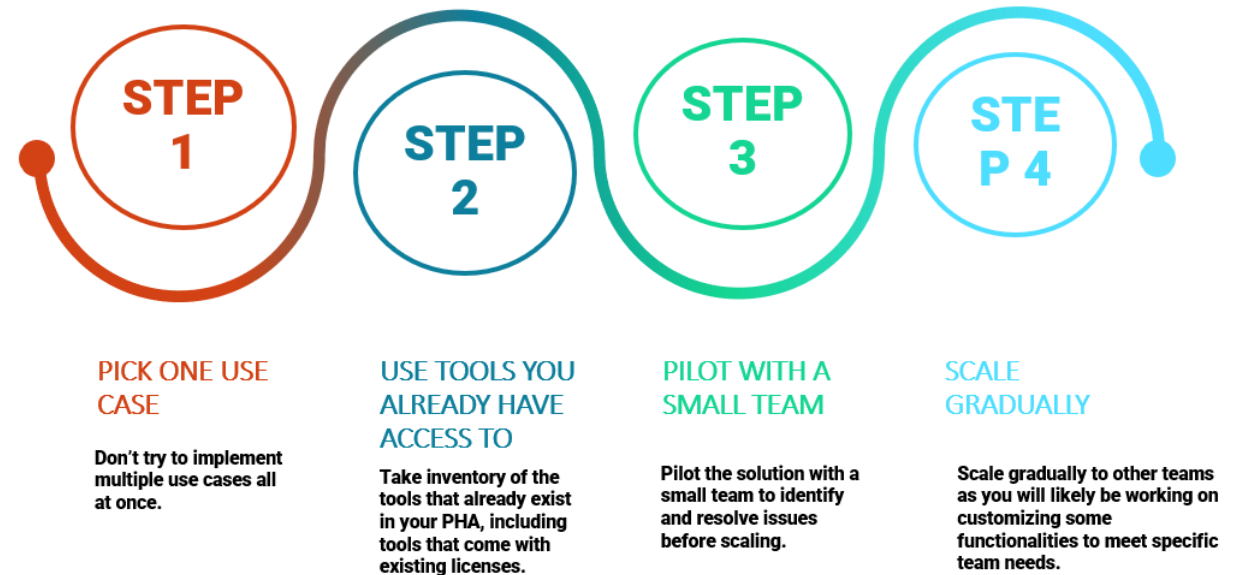
---



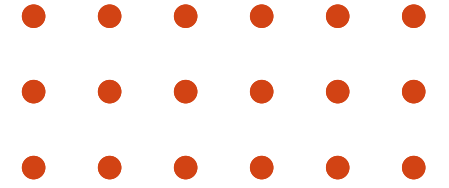
# Key takeaways

---

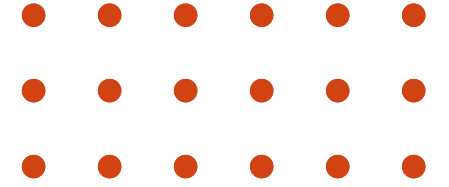
- FREE ≠ INFERIOR
- MATCH TOOL TO TASK
- HIPAA FIRST
- START WITH WHAT YOU HAVE



# References



- REDCap
    - <https://project-redcap.org/>
    - <https://project-redcap.org/software/>
    - <https://projectredcap.org/wp-content/resources/REDCapTechnicalOverview.pdf>
  - KoboToolbox
    - <https://www.kobotoolbox.org/>
    - [https://support.kobotoolbox.org/is\\_my\\_data\\_safe.html](https://support.kobotoolbox.org/is_my_data_safe.html)
  - ODK
    - <https://getodk.org/>
    - <https://docs.getodk.org/central-install/>
    - <https://docs.getodk.org/xlsform/>
  - QGIS
    - <https://qgis.org/>
    - <https://plugins.qgis.org/>
    - <https://pmc.ncbi.nlm.nih.gov/articles/PMC12416080/pdf/main.pdf>
  - Looker Studio
    - <https://cloud.google.com/looker-studio>
    - <https://docs.cloud.google.com/data-studio/data-governance-in-an-overview>
    - <https://cloud.google.com/looker/studio-hipaa-implementation-guide>
  - Apache Superset
    - <https://superset.apache.org/>
    - <https://superset.apache.org/user-docs/>
-



# Thank you!



*For questions, please reach out to [gtamrat@naccho.org](mailto:gtamrat@naccho.org) and [aayala@naccho.org](mailto:aayala@naccho.org)!*

