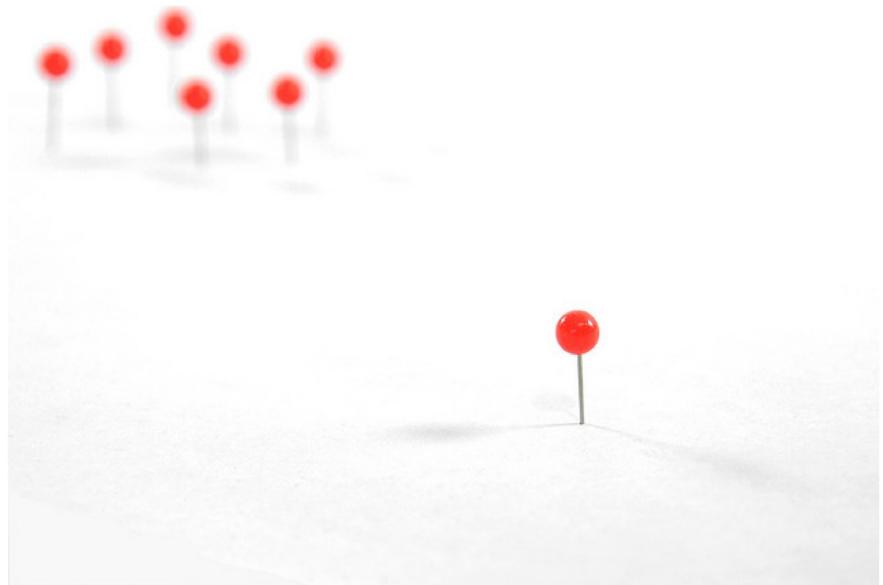


# Tutorials on Data Management

## Lesson 5: Data Quality Control and Assurance



CC image by Shane Melaugh on Flickr

# Lesson Topics

- Definitions
  - Quality assurance and Quality control
  - Data contamination
  - Types of errors
- QA/QC best practices
  - Before data collection
  - During data collection/entry
  - After data collection/entry



CC image by cobalt123 on Flickr

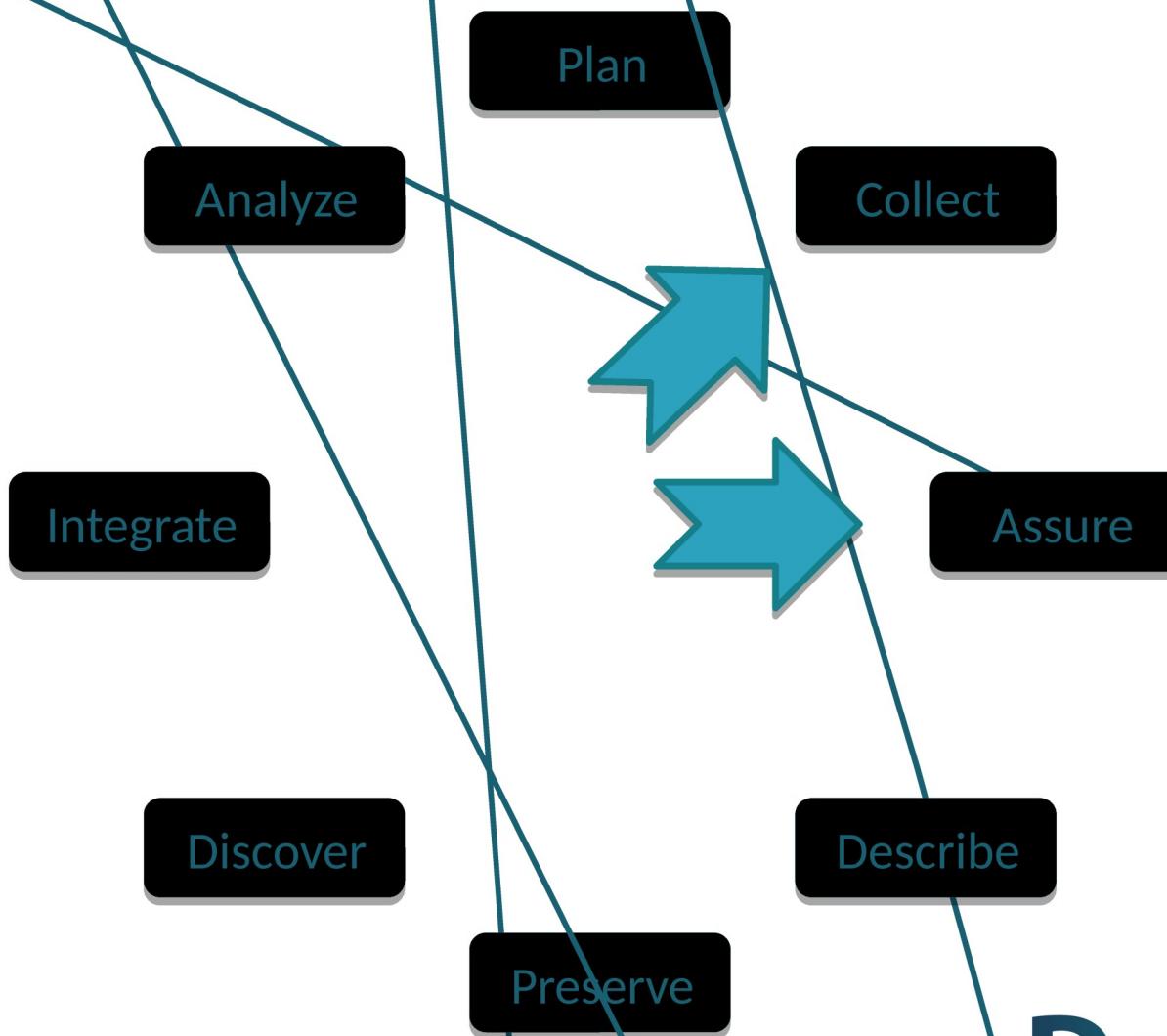
# Learning Objectives

- After completing this lesson, the participant will be able to:
  - Define data quality control and data quality assurance
  - Perform quality control and assurance on their data at all stages of the research cycle



CC image by 0xFCAF on Flickr

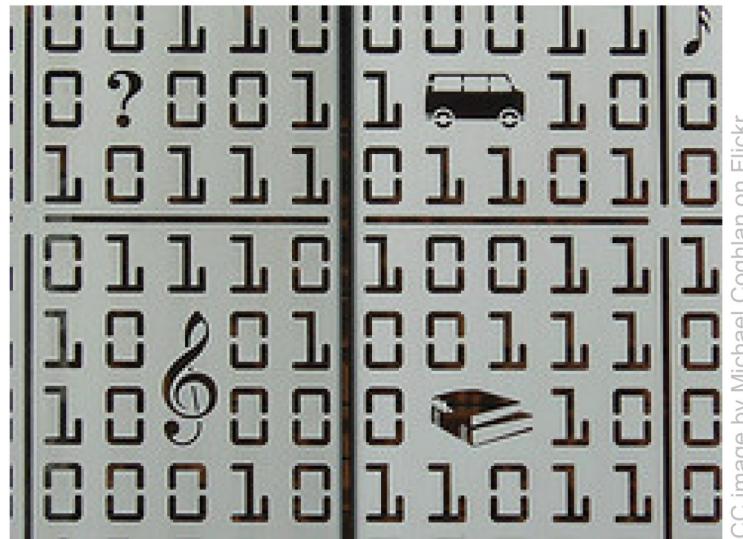
# The Data Life Cycle



# Definitions

## Data Contamination

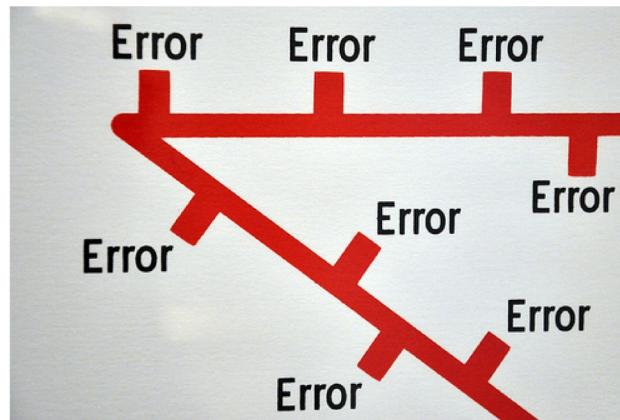
- Process or phenomenon, other than the one of interest, that affects the variable value
- Erroneous values



CC image by Michael Coghlan on Flickr

# Definitions: Types of Errors

- **Errors of Commission**
  - Incorrect or inaccurate data entered
  - Examples: malfunctioning instrument, mistyped data
- **Errors of Omission**
  - Data or metadata not recorded
  - Examples: inadequate documentation, human error, anomalies in the field



CC image by Nick J Webb on Flickr

# Defining QA/QC

- Strategies for preventing errors from entering a dataset
- Activities to ensure quality of data before collection
- Activities that involve monitoring and maintaining the quality of data during the study

# QA/QC Before Collection

- Define & enforce standards
  - Formats
  - Codes
  - Measurement units
  - Metadata
- Assign responsibility for data quality
  - Be sure assigned person is educated in QA/QC

# QA/QC During Data Entry

- Double entry
  - Data keyed in by two independent people
  - Check for agreement with computer verification
- Record a reading of the data and transcribe from the recording
- Use text-to-speech program to read data back



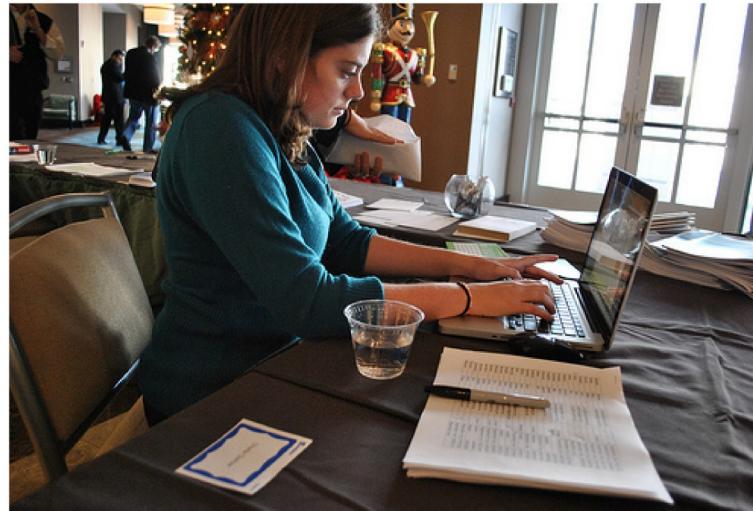
CC image by weskriesel on Flickr

# QA/QC During Data Entry

- Design data storage well
  - Minimize number of times items that must be entered repeatedly
  - Use consistent terminology
  - Atomize data: one cell per piece of information
- Document changes to data
  - Avoids duplicate error checking
  - Allows undo if necessary

# QA/QC After Data Entry

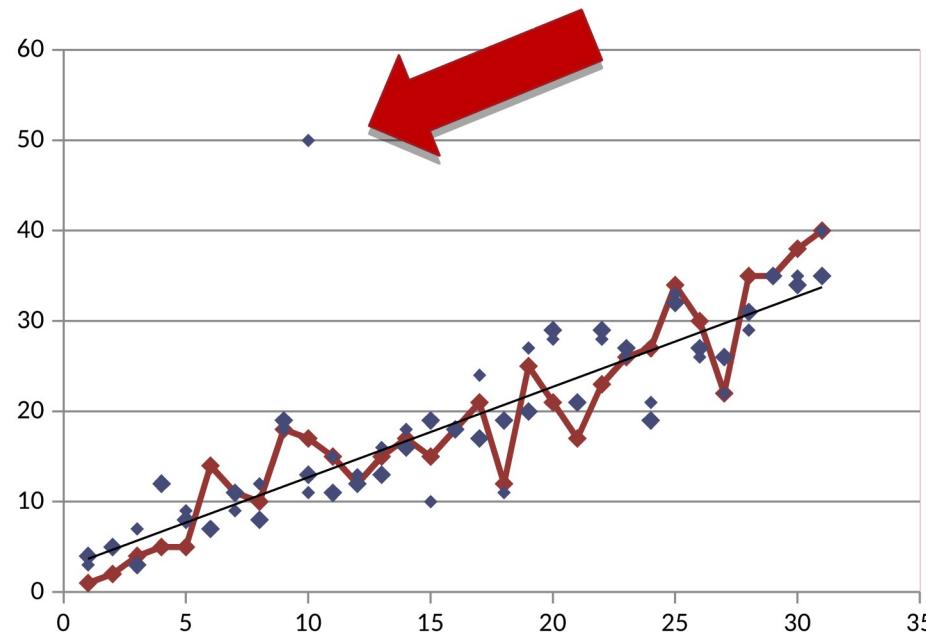
- Make sure data line up in proper columns
- No missing, impossible, or anomalous values
- Perform statistical summaries



CC image by chesapeakeclimate on Flickr

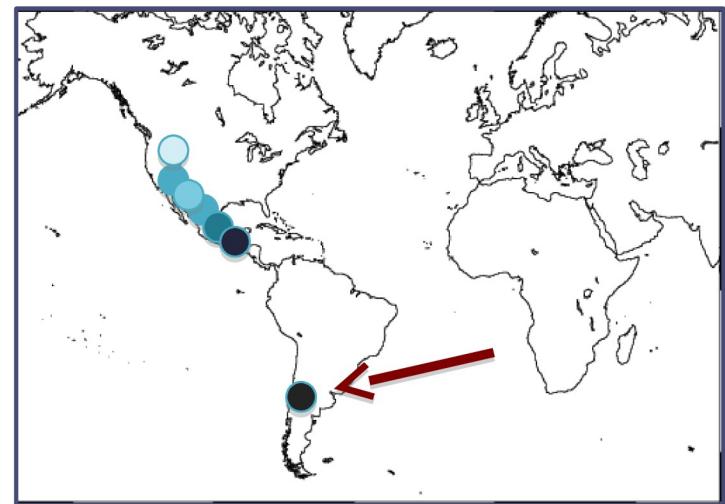
# QA/QC After Data Entry

- Look for outliers
  - Outliers are extreme values for a variable given the statistical model being used
  - The goal is not to eliminate outliers but to identify potential data contamination



# QA/QC After Data Entry

- Methods to look for outliers
  - Graphical
    - Normal probability plots
    - Regression
    - Scatter plots
  - Maps
  - Subtract values from mean



# Summary

- Data contamination is data that results from a factor not examined by the study that results in altered data values
- Data error types: commission or omission
- Quality assurance and quality control are strategies for
  - preventing errors from entering a dataset
  - ensuring data quality for entered data
  - monitoring, and maintaining data quality throughout the project
- Identify and enforce quality assurance and quality control measures throughout the Data Life Cycle

# References

1. D. Edwards, in *Ecological Data: Design, Management and Processing*, WK Michener and JW Brunt, Eds. (Blackwell, New York, 2000), pp. 70-91. Available at [www.ecoinformatics.org/pubs](http://www.ecoinformatics.org/pubs)
2. R. B. Cook, R. J. Olson, P. Kanciruk, L. A. Hook, Best practices for preparing ecological data sets to share and archive. *Bull. Ecol. Soc. Amer.* **82**, 138-141 (2001).
3. A. D. Chapman, “Principles of Data Quality: Report for the Global Biodiversity Information Facility” (Global Biodiversity Information Facility, Copenhagen, 2004). Available at <http://www.gbif.org/communications/resources/print-and-online-resources/download-publications/booklets/>

The full slide deck may be downloaded from:  
<http://www.dataone.org/education-modules>

Suggested citation:

DataONE Education Module: Data Quality Control and Assurance. DataONE. Retrieved Nov12, 2012. From [http://www.dataone.org/sites/all/documents/L05\\_DataQualityControlAssurance.pptx](http://www.dataone.org/sites/all/documents/L05_DataQualityControlAssurance.pptx)

Copyright license information:



No rights reserved; you may enhance and reuse for your own purposes. We do ask that you provide appropriate citation and attribution to DataONE.