Measurement & Analysis in the Real World

Tools for Cleaning Messy Data

Will Hayes – SEI

Robert Stoddard – SEI

Rhonda Brown – SEI

Software Solutions Conference 2015

November 16–18, 2015

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

DM-0003055

Agenda

- Introduction
- Matching Information Needs
- Getting to the Data
- Cumulative Flow Diagrams
- Tool Demonstration
- Predictive Modeling

Complementary but Different Focus

Government Program Office

- Assess forecasted risk
- Manage to outcomes
- Responsible for total cost of ownership (and current cost)
- Obliged to seek out and communicate user needs

Strive to avoid directing the contractor on HOW to work

Development Contractor

- Predict performance
- Control performance drivers
- Responsible for meeting current commitments
- Subject to re-direction based on user needs

Influence on WHAT to build may be constrained by contract

Different Audiences for Metrics and Status

Program office personnel who interact directly with contractors

- Generally need insight at a finer level of detail
- Must maintain visibility/continuity over time

Stakeholders in the program, beyond Program Management

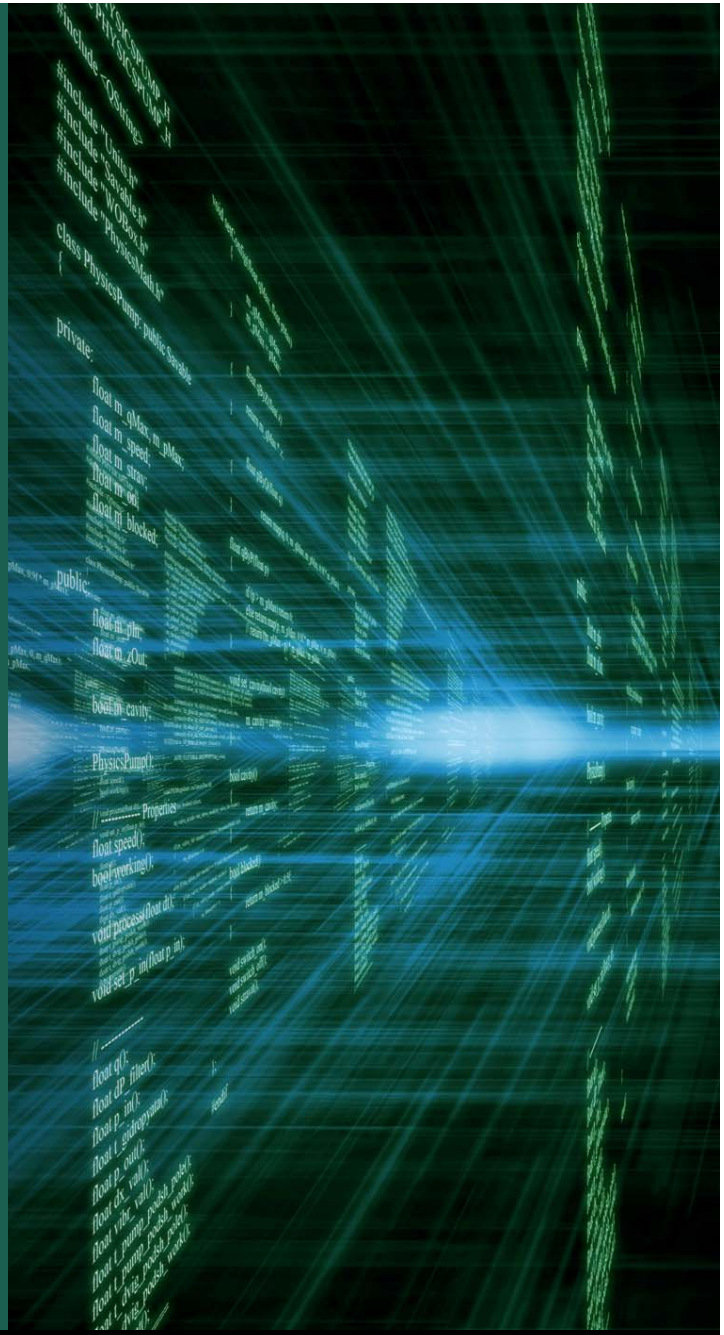
- May focus on specific topics to the exclusion of all else
- Participate in less frequent status discussions, perhaps

Senior leadership who oversee the program office

- Focus on performance of the program, not just this contract
- Frame of reference may be broader and more long term

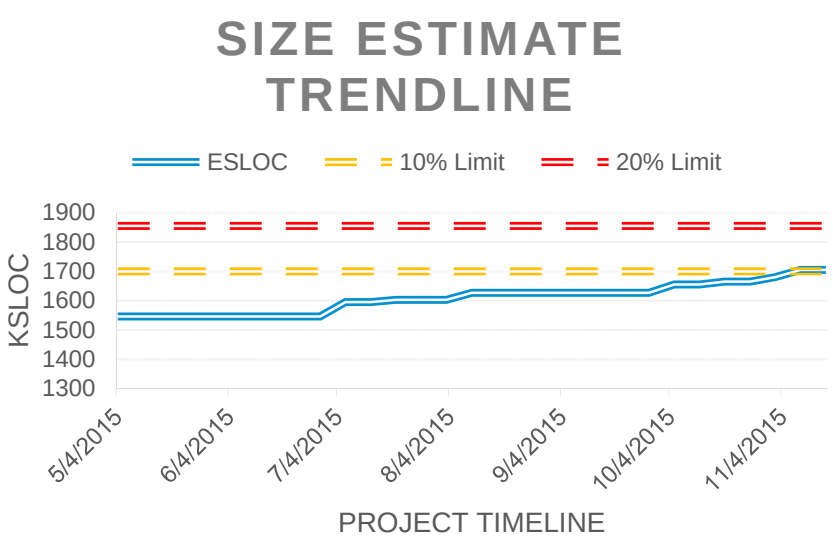
Matching Information with Needs

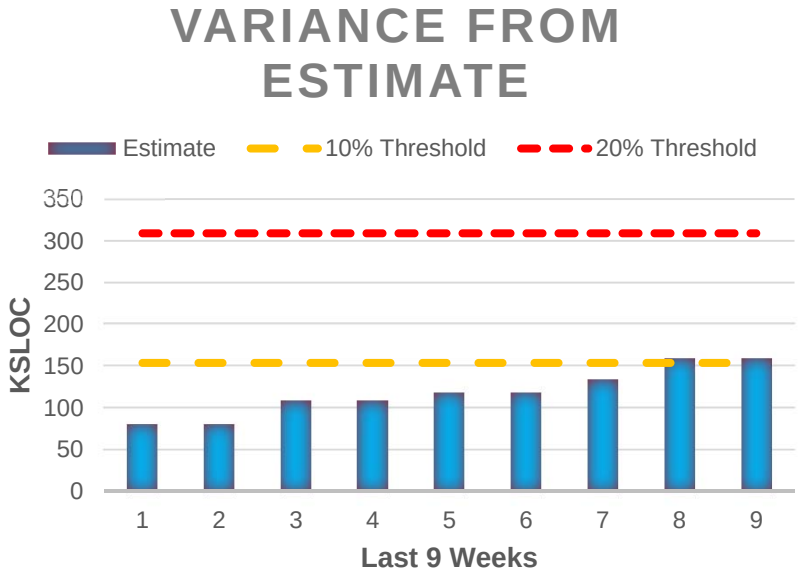
Re-Casting Metrics for the Target Audience



Time-Horizon and Specificity

The chart below shows the trend in estimated size, with thresholds for potential corrective action.

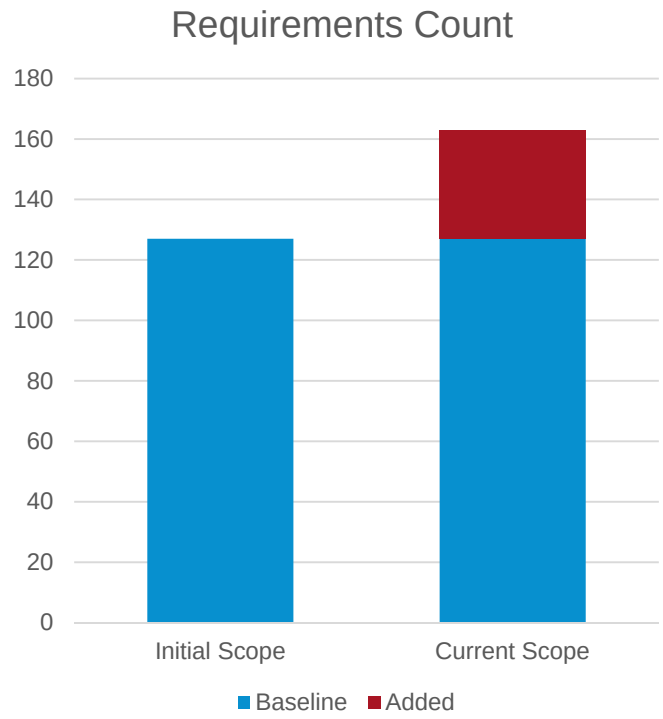


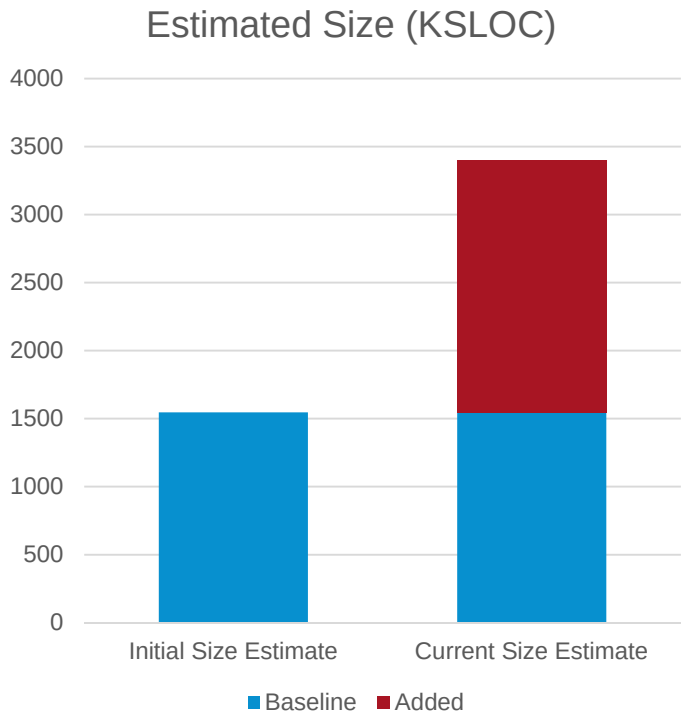


The simplified version above shows only 9 weeks, focusing only on variance from original estimate.

Choose time-horizon and specificity to meet audience needs

Converging Indicators

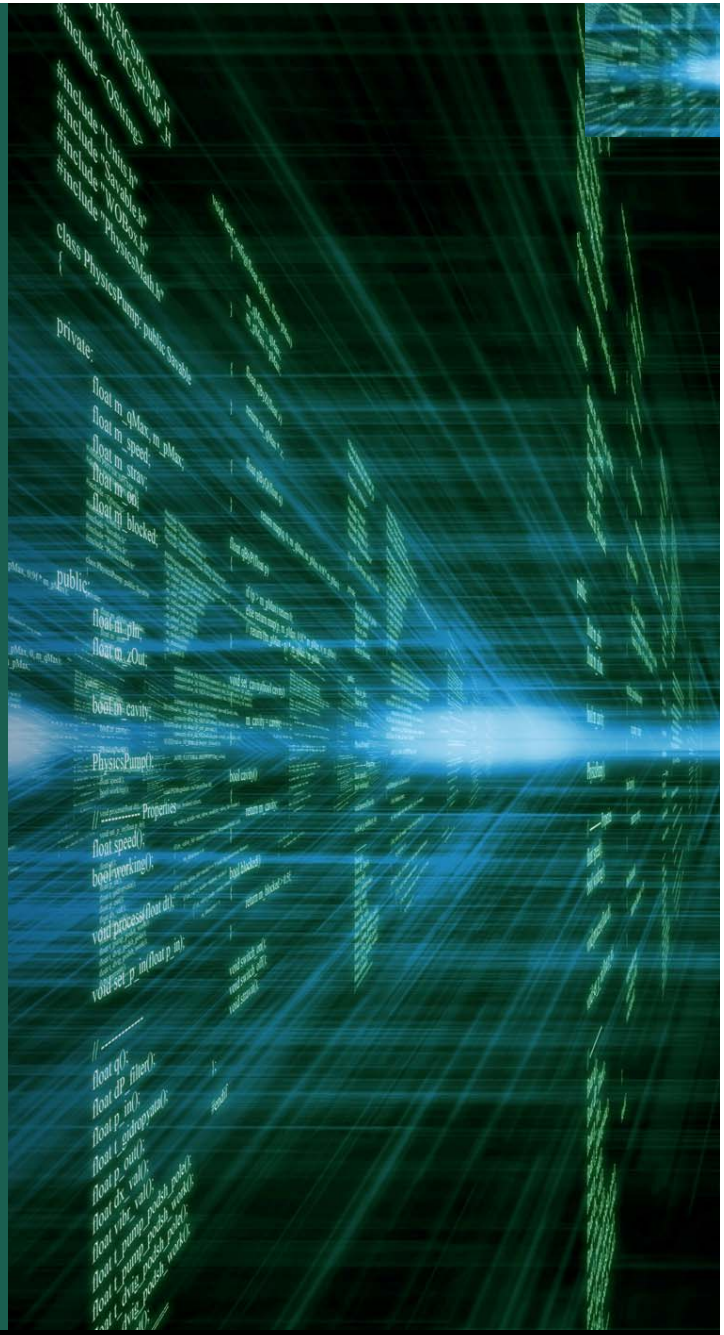




Some information is visible only when you combine data

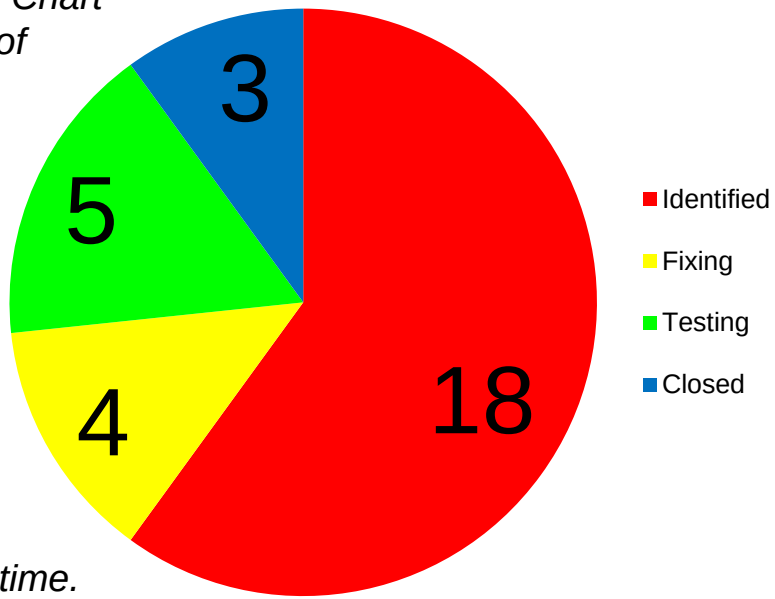
Useful Graphical Tool

Cumulative Flow Diagram



Constructing a Cumulative Flow Diagram$_1$

Here we have a Pie Chart showing the status of 30 defects across the four stages of the defect handling life-cycle.

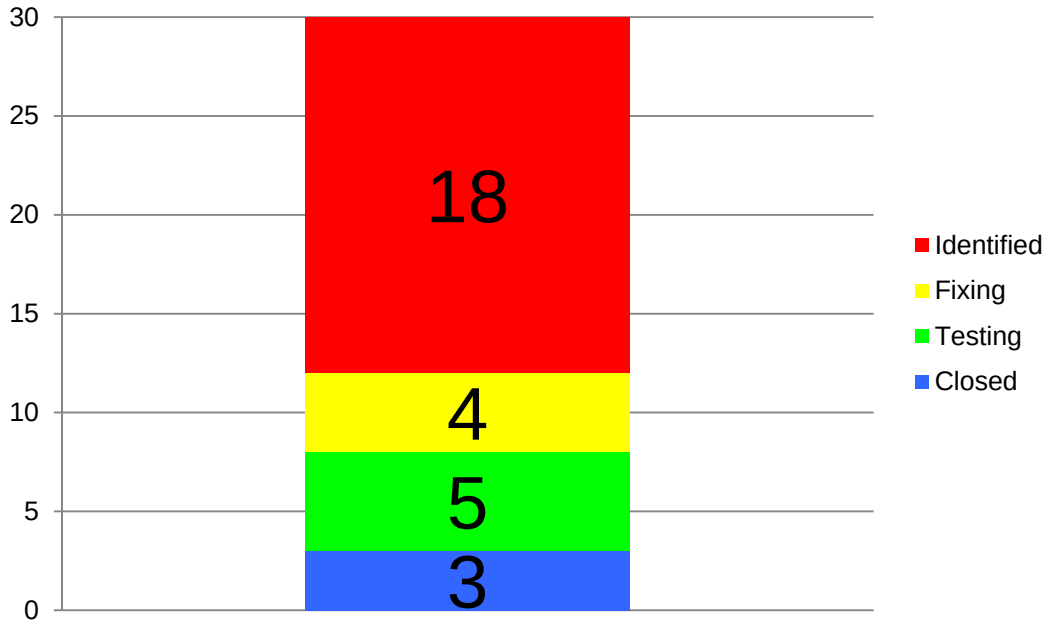


This is a snapshot for a single point in time.

Constructing a Cumulative Flow Diagram$_2$

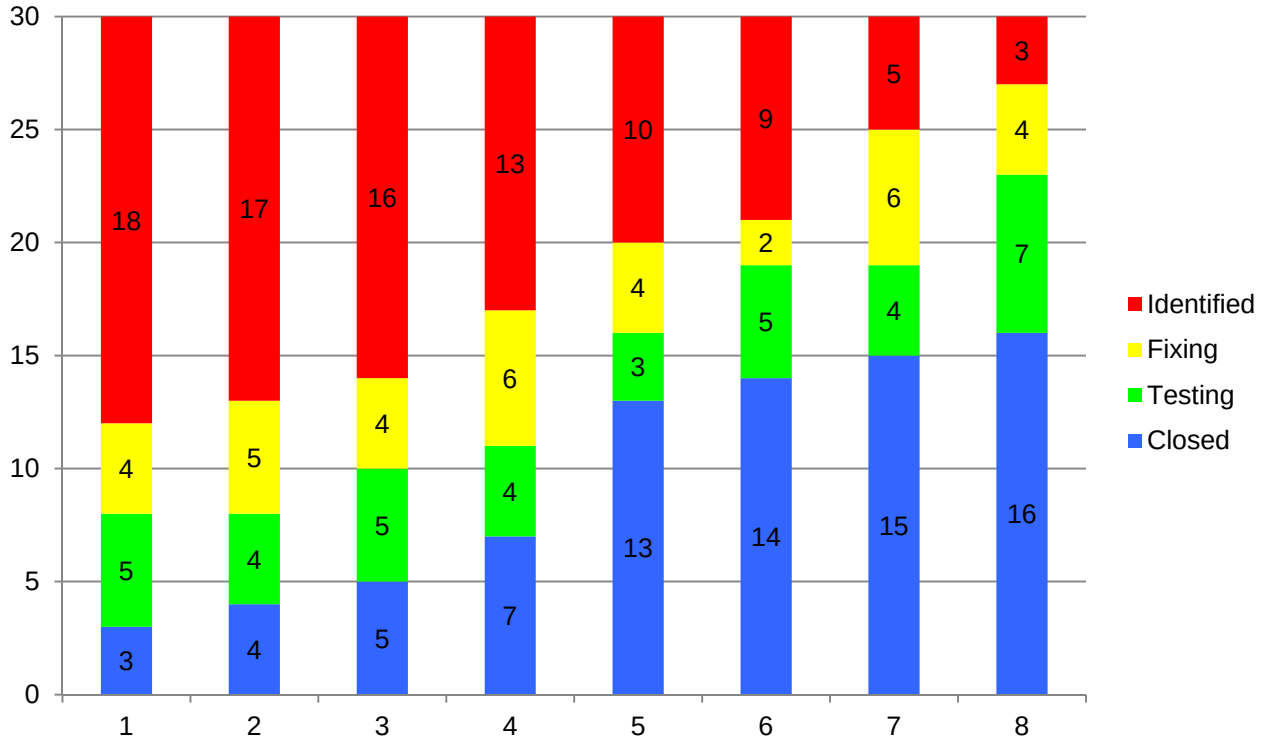
Same data, but presented in a stacked column chart

For a single point in time.



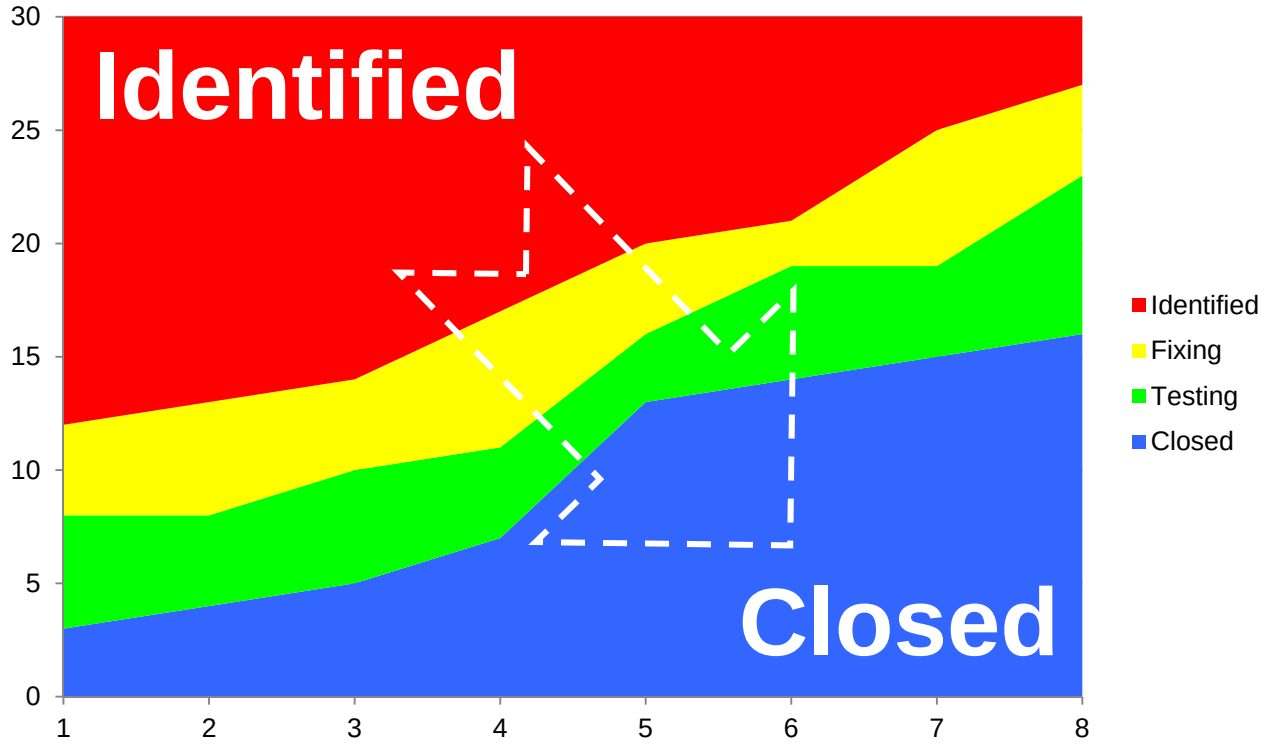
Constructing a Cumulative Flow Diagram$_3$

… adding the next 7 times

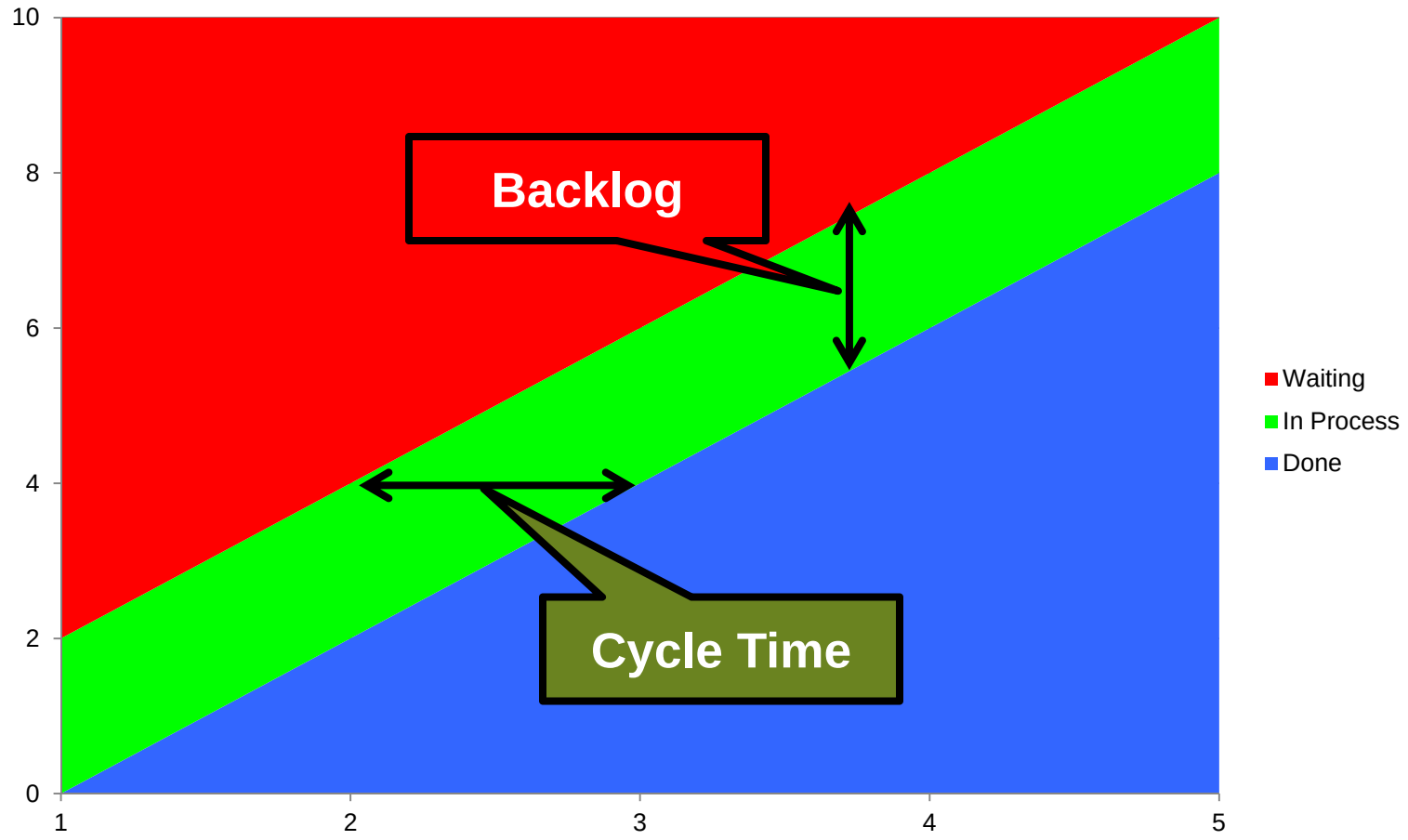


Constructing a Cumulative Flow Diagram$_4$

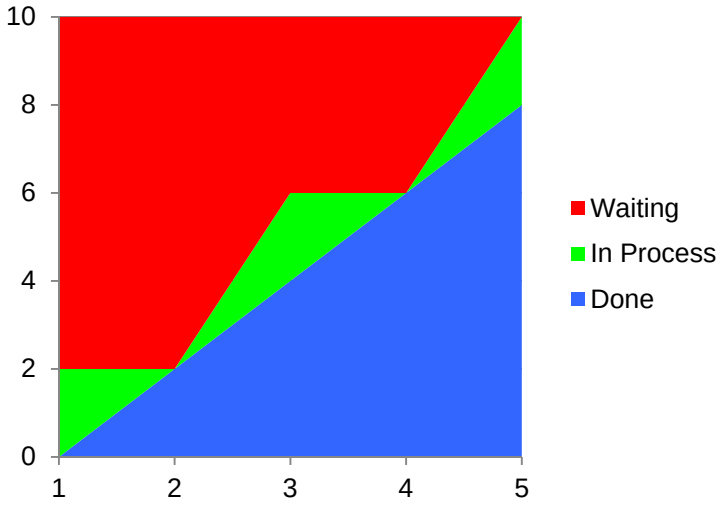
… now we are looking at the flow from "identified"… to "Closed"…
This view starts to show patterns a little easier…

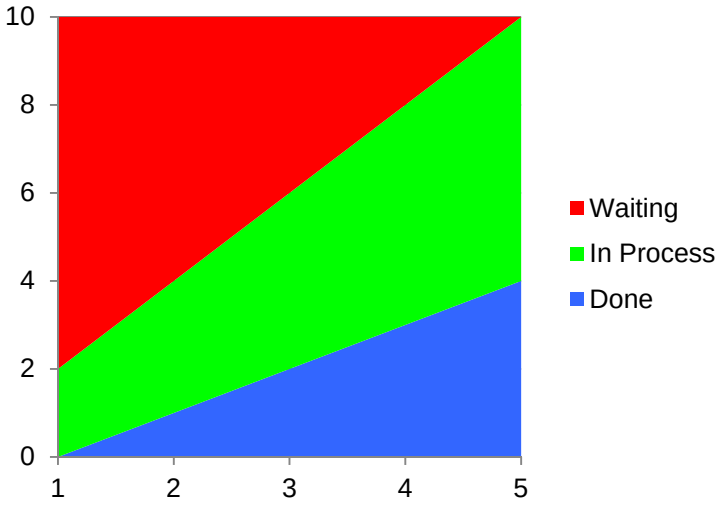


Tell-Tale Signals

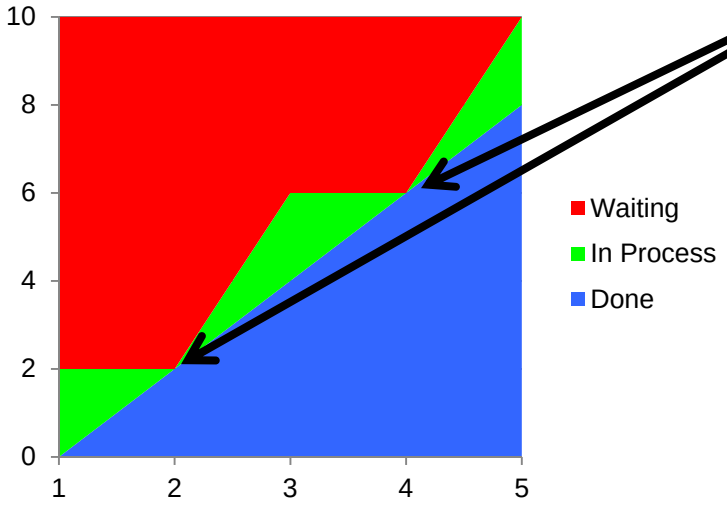


Exercise: What is Going on Here?





Exercise: What *MIGHT BE* Happening$_1$



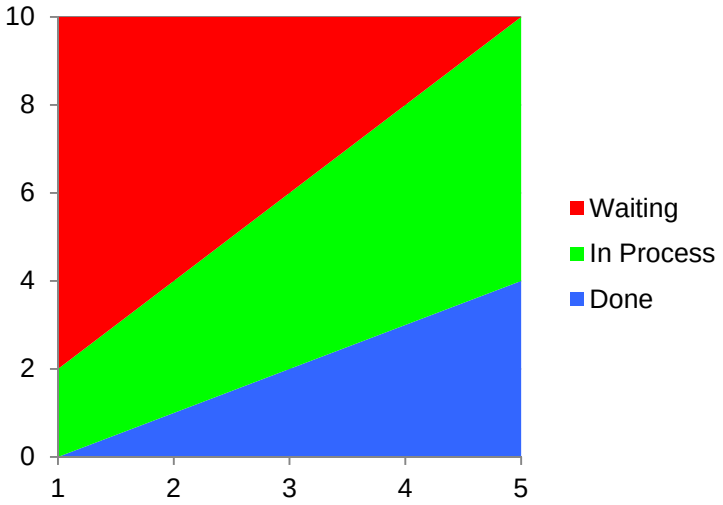
At time 2, and then again at time 4, the number of items "In Process" goes to zero.

- **Have we lost the resource(s) that were preparing the items in the "Waiting" state?**
- **Is this intentional, due to limited resource(s) who can work on items in the "In Process" state?**

Exercise: What *MIGHT BE* Happening$_2$

The number of items that are "In Process" is growing over time.

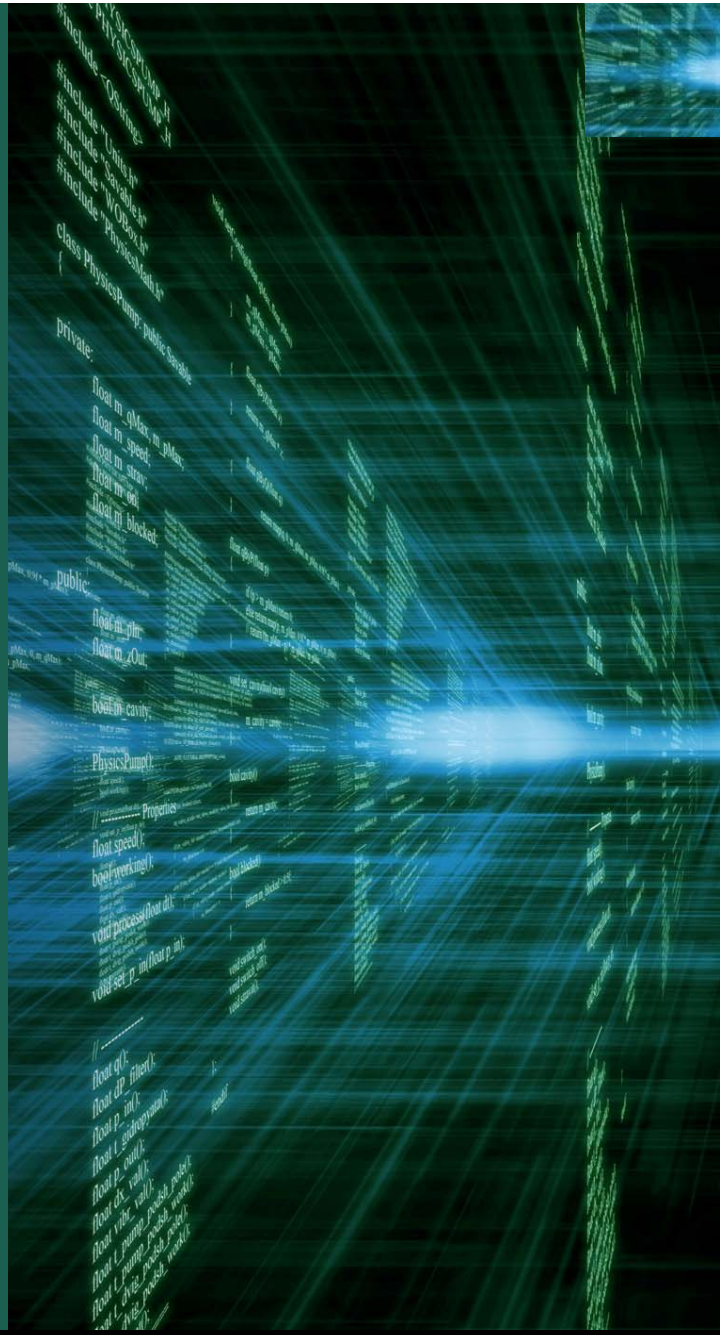
- **The rate at which things enter "In Process" is greater than the rate at which things leave "In Process."**
- **Are people moving onto new items without completing their work?**
- **Are new resources being added, who start new work at each time period?**
- **Are things moving into the "Done" state quickly enough?**



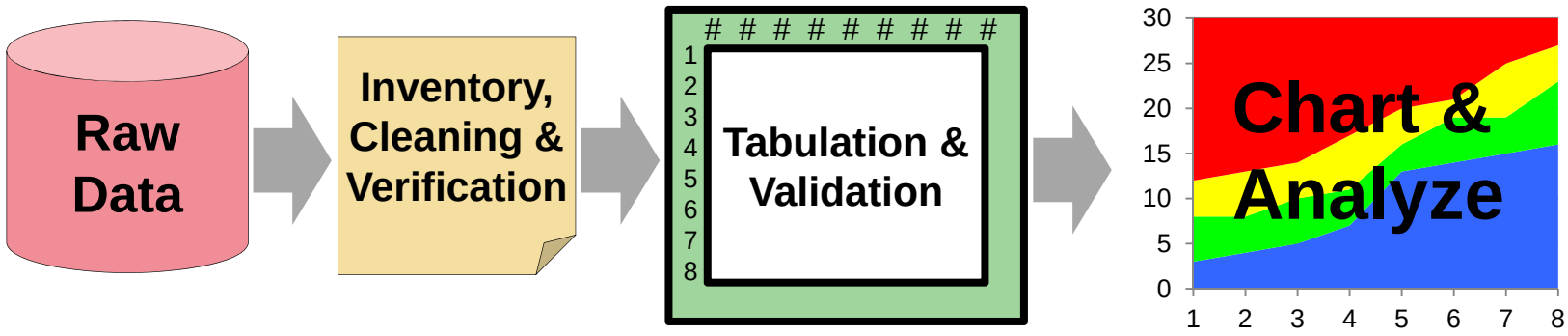
Getting to the Data

Mining a Configuration Management Database

or Application Lifecycle Management Tool



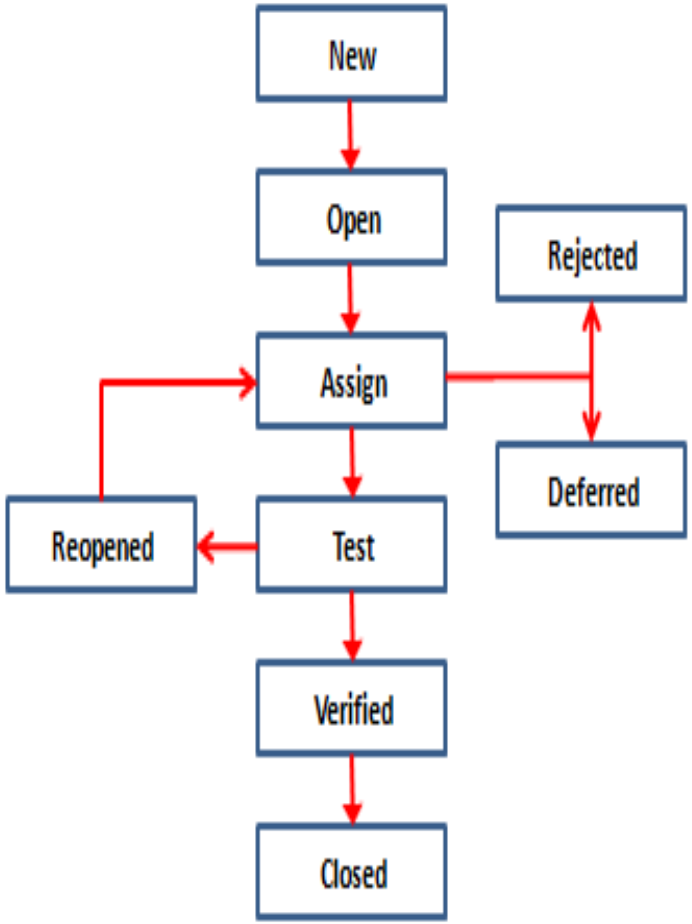
Activity Flow: Mining the Database



Weekly analysis activities comprised of these steps:

- Data pulled directly from Configuration Management system
- Inventory 'change records' to verify completeness and accuracy
- Tabulate by pre-defined time intervals and validate totals
- Chart data using Cumulative Flow Diagrams to analyze progress

Details: Process View



Process flow for a defect being worked

- Entry/exit criteria for each step
- Individual assigned to work each one
- Progress through the process tracked
- Database fields used to record
  - Current state in the process
  - History of progression through the states
  - Date/time stamp for each state change
  - … and lots of other information

Details: Raw Data

Main Data Table

Defect ID	Title	Description	Severity	…..
1000001	Dropped data…	Message traffic is overwritten when buffer size not specified in….	1	- - - -
1000002	Missing header…	File never read at initialization due to missing pointer in….	2	- - - -
1000003	Unpredictable close…	Process XYZ terminates while opening file…	1	- - - -
- - - -	- - - -	- - - -	- - - -	- - - -

Change Auditing Table

ID	Old State	New State	TimeStamp	… LOTS of other data …
1000001	New	Open	mm/dd/yy hh:mm:ss	- - - -
1000001	Open	Assign	mm/dd/yy hh:mm:ss	- - - -
1000001	Assign	Test	mm/dd/yy hh:mm:ss	- - - -
1000002	New	Open	mm/dd/yy hh:mm:ss	- - - -
- - - -	- - - -	- - - -	- - - -	- - - -

Details: Mining the Change Auditing Table

This database table provides:

- Date and time when each item entered a given state
- History of all such transitions since the record was created

Using that information, we can derive:

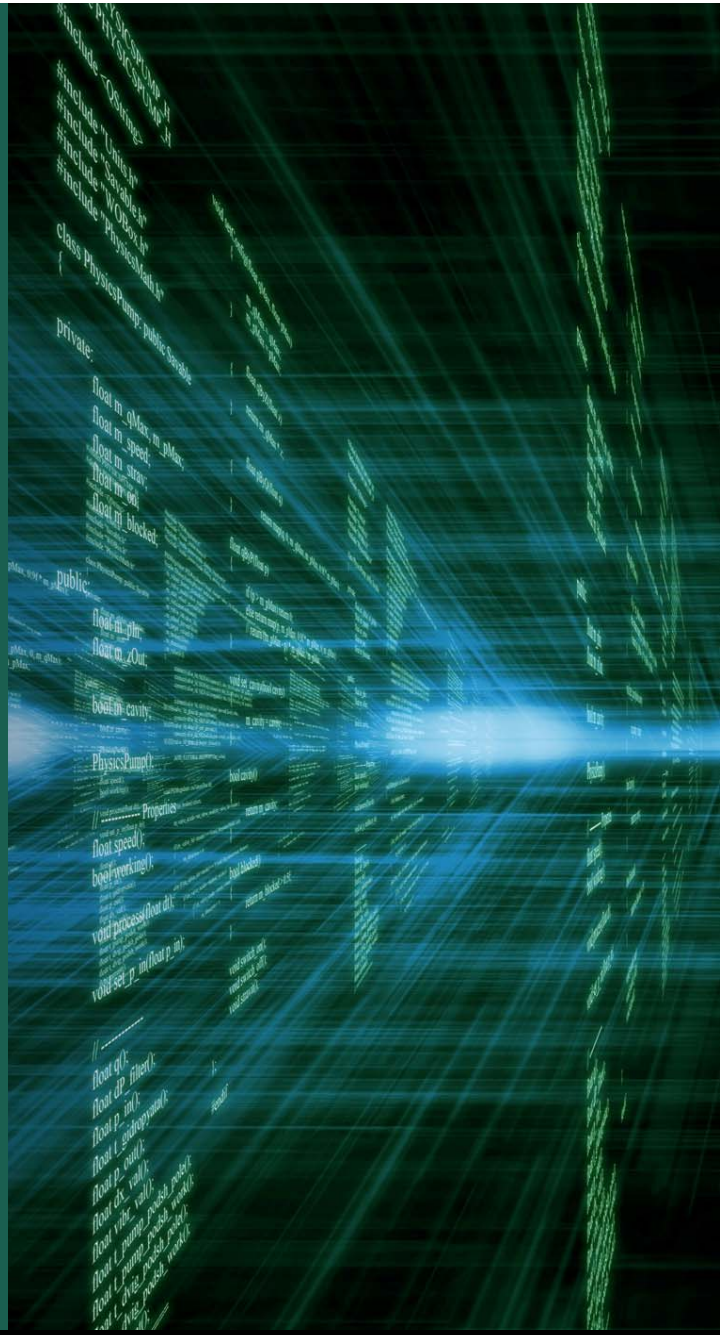
- How many records are in each state at a given time
- How long each item stayed in any particular state

This allows us to:

- Draw Cumulative Flow Diagrams to show flow
- Model the state-transition activity with a predictive model

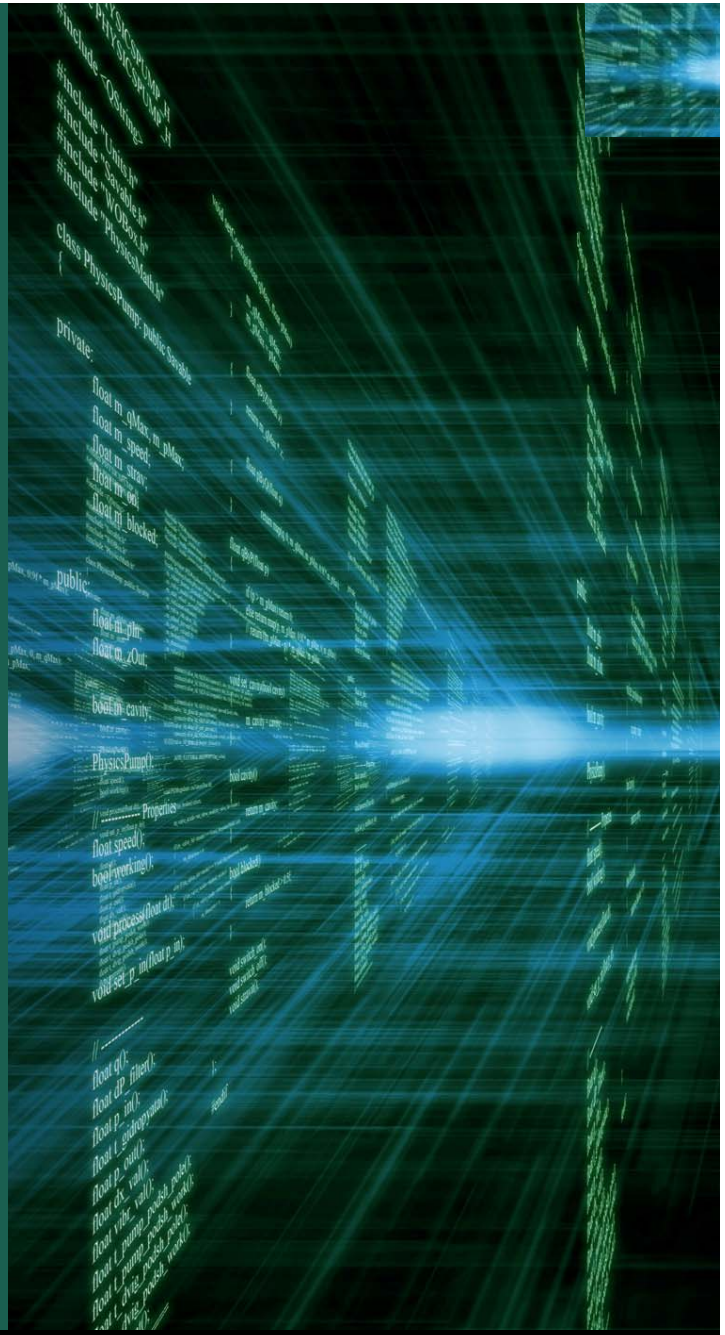
Leveraging *Excel* and *Access* with VB

Tool Demonstration



Useful Statistical Tool

Predictive Modeling



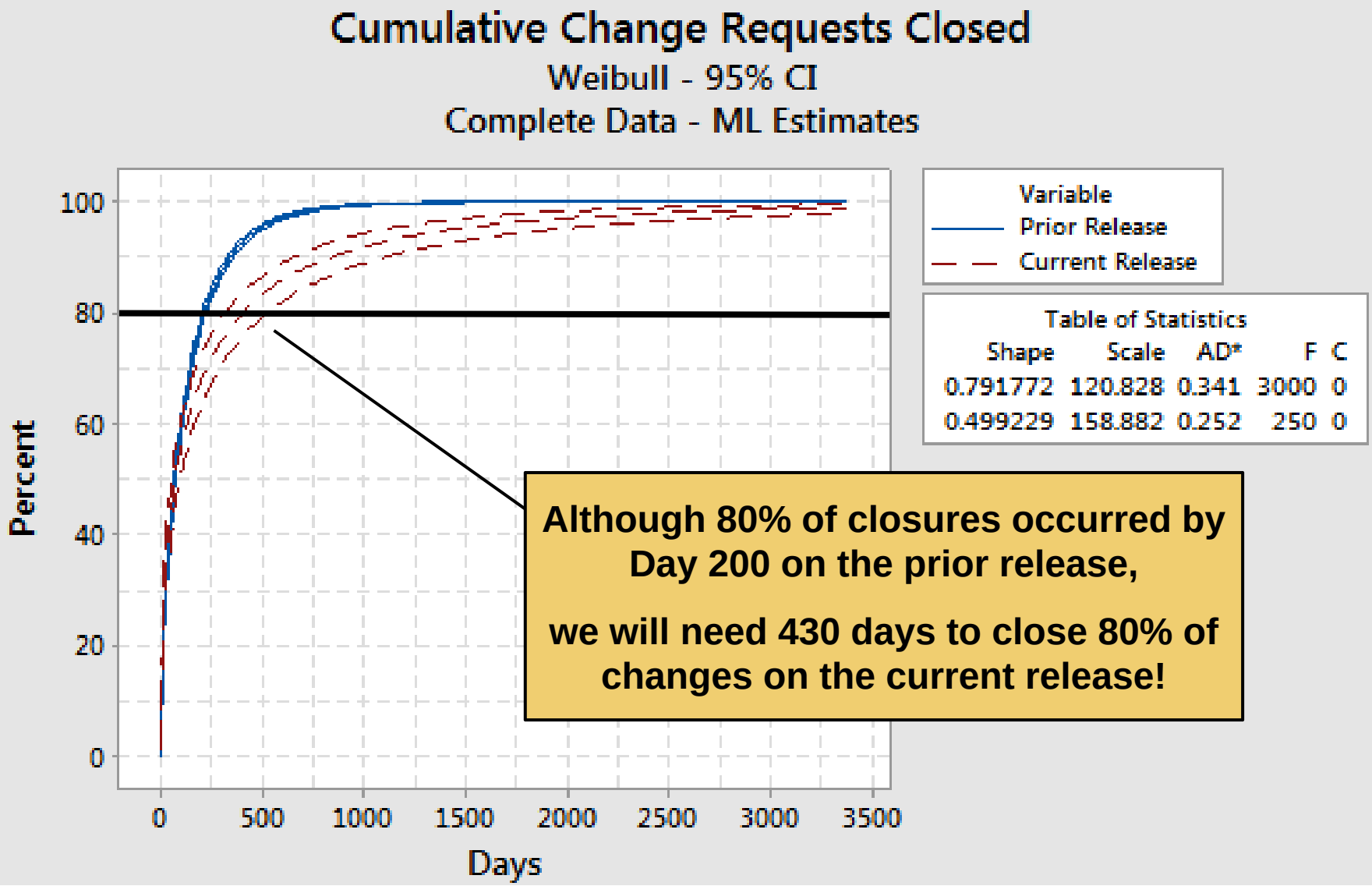
Building Models

Data derived from time stamps

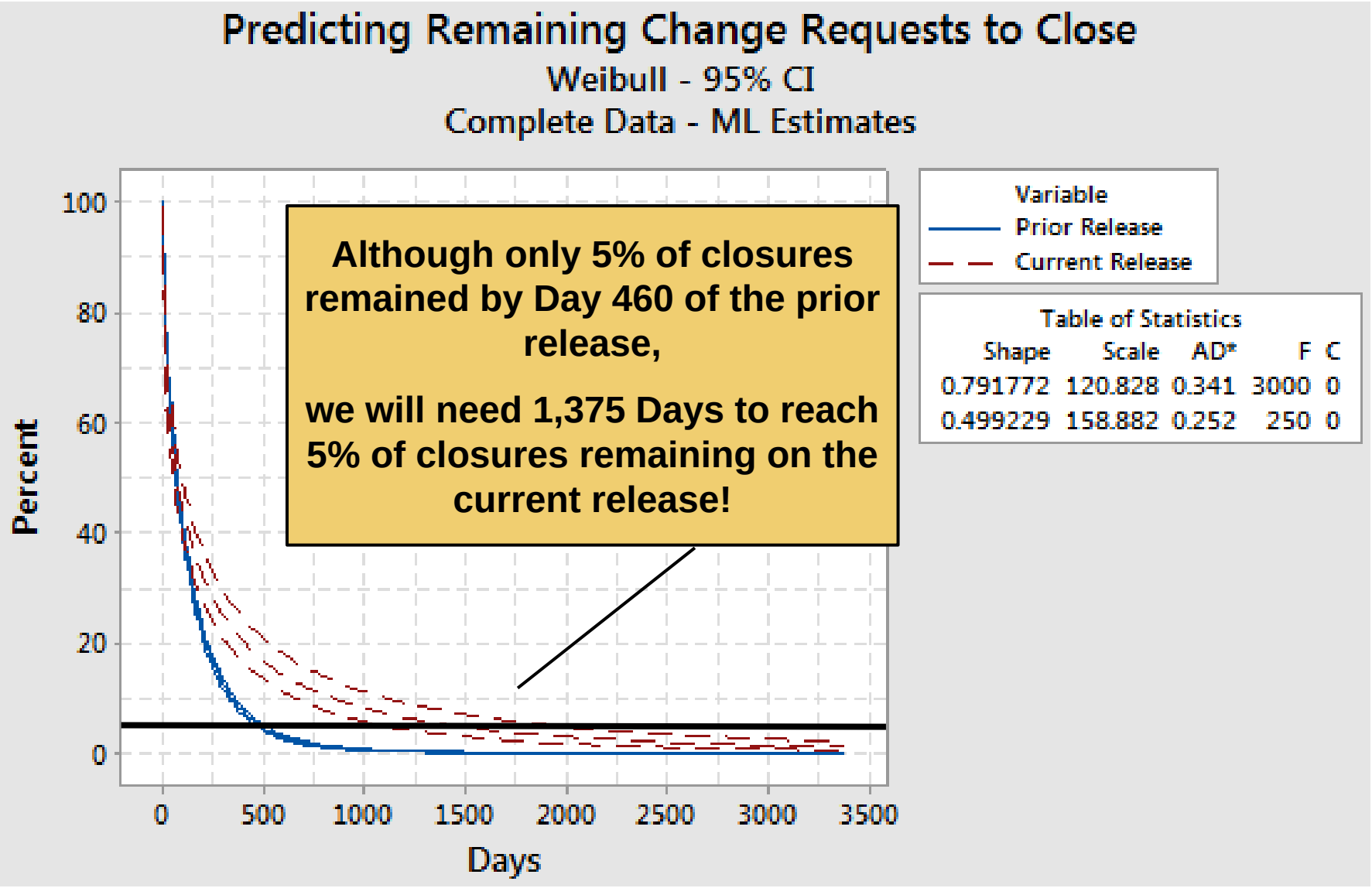
- Duration associated with each state in the sequence
- Information about range of time seen in the past
- Benchmarks for durations can aid in planning

A variety of modeling techniques can be applied

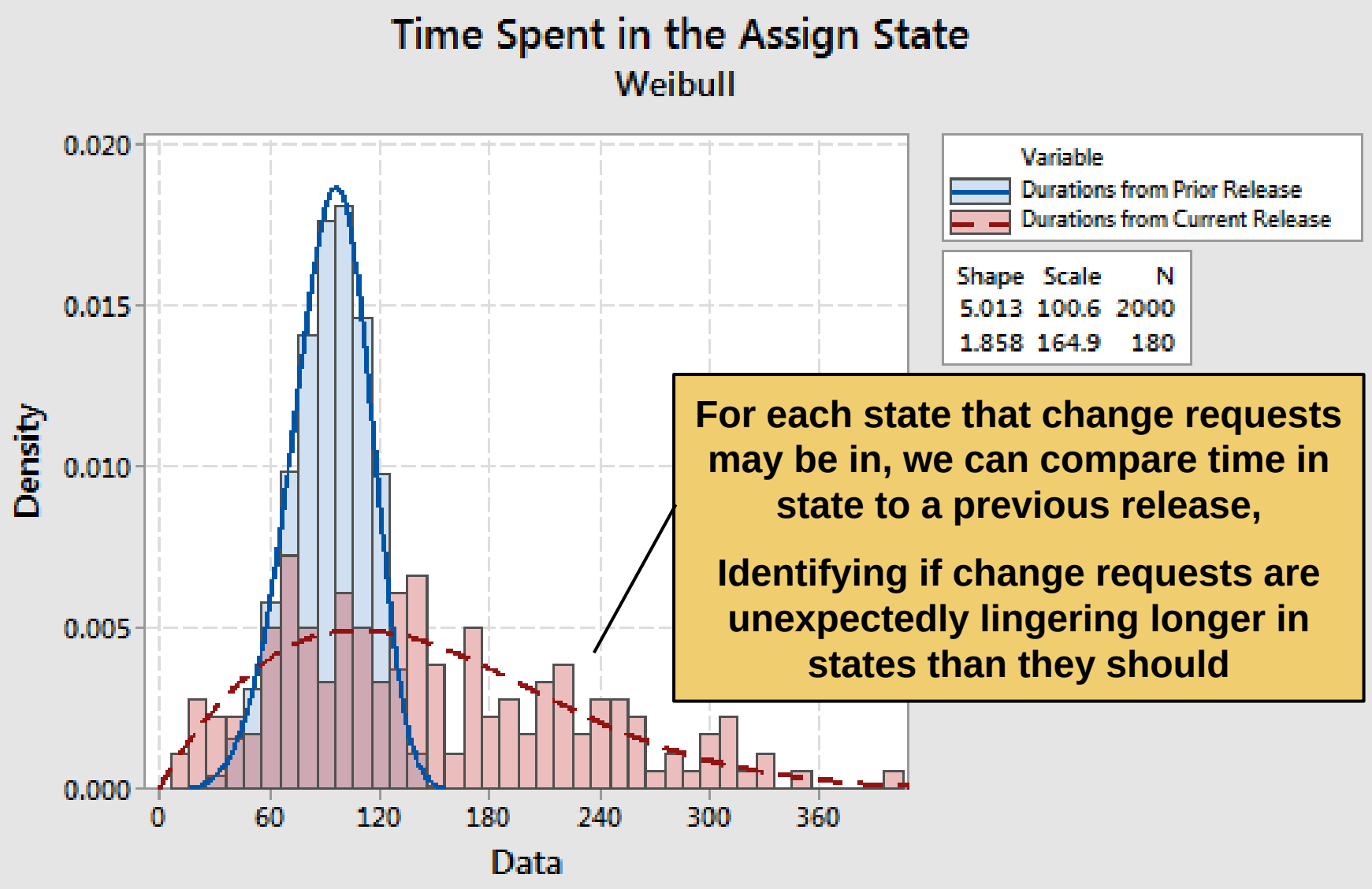
Predicting Change Request Closure



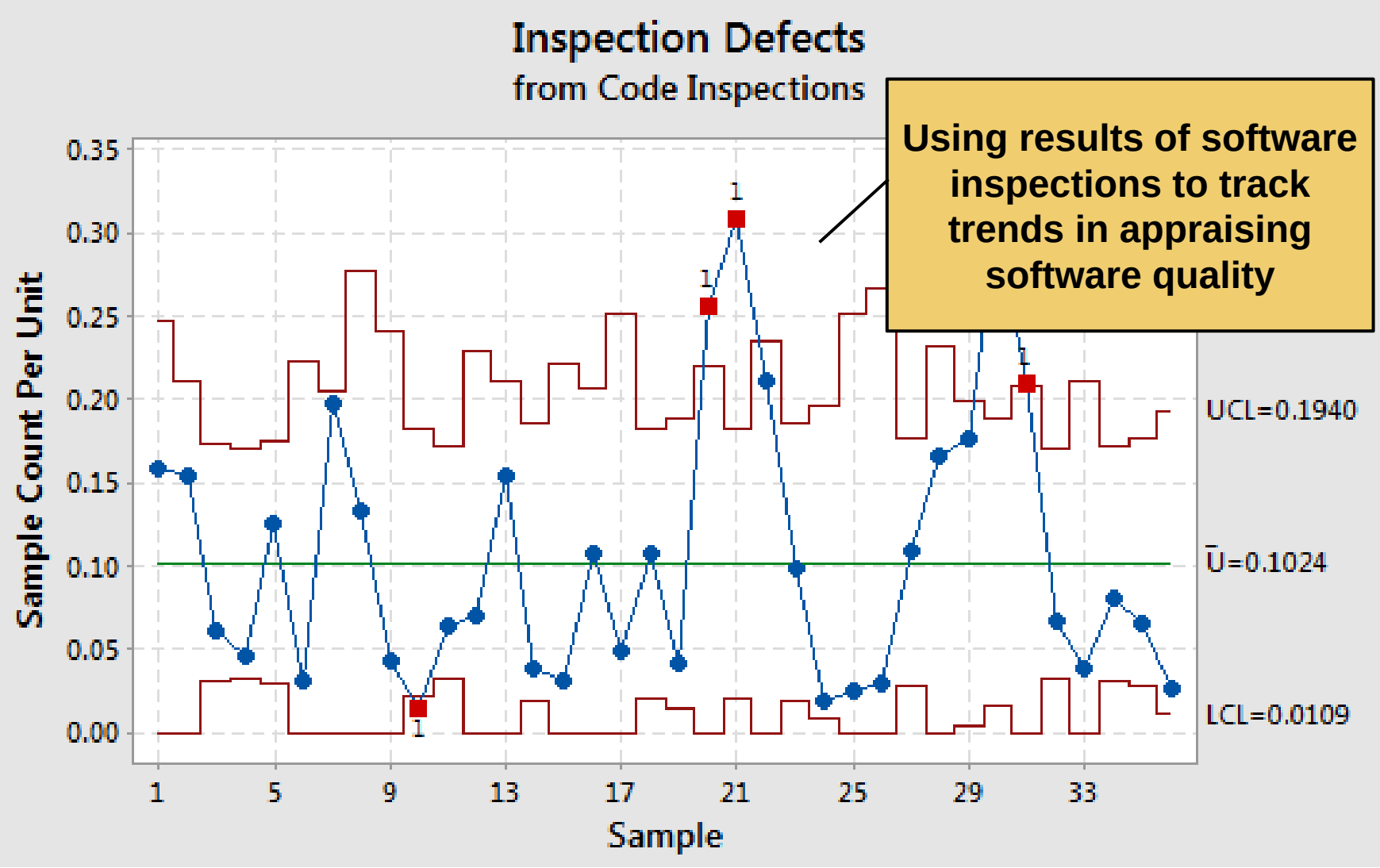
Predicting Remaining Changes to Close



Time in State Compared to Past Release



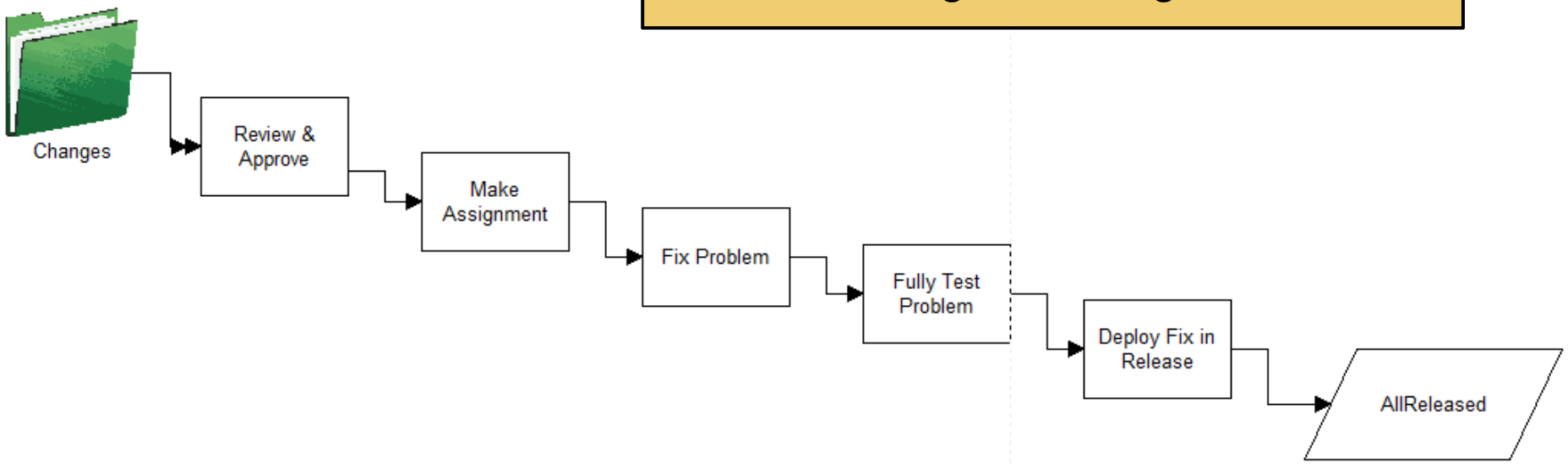
Tracking Software Quality Trends



Modeling Flow of Software Change Requests

Using Discrete Event Simulation, we can create simulations of the flow of software change requests, and

Conduct what-if analysis of various strategies to work off the change requests including staff assignments



Contact Information

Presenters

Will Hayes

Client Technical Solutions Division

Telephone: +1 412.268.6398

Email: wh@sei.cmu.edu

Rhonda Brown

Software Engineering
Measurement and Analysis

Telephone: + 1 412.268.3963

Email: rbrown@sei.cmu.edu

Robert Stoddard

Software Engineering
Measurement and Analysis

Telephone: + 1 412.268.1121

Email: rws@sei.cmu.edu