Naval Center for Cost Analysis

Software Maintenance (SWMX)

Recommendations for Estimating and Data Collection

June 2014

Presenter: Shelley Dickson

Objective

Provide the Department of Defense with the ability to accurately estimate, budget, allocate, and justify the software maintenance resources required to meet evolving mission and service affordability requirements across the system life-cycle.

Source: Jones, Cheryl. *Estimating Software Maintenance Costs for U.S. Defense Systems.* Deputy Assistant Secretary of the Army for Cost and Economics. 1 May 2014.

Outline of Presentation

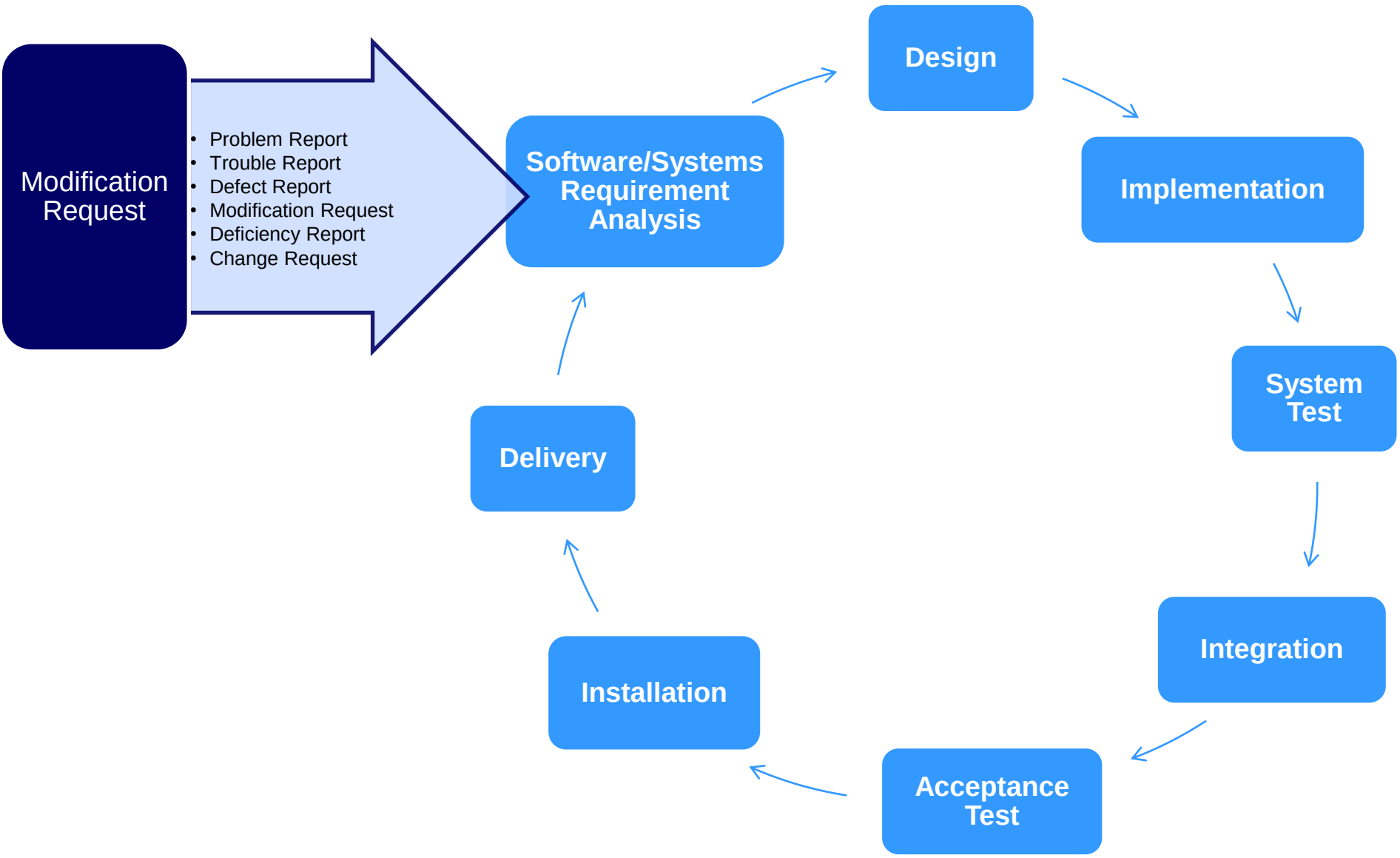
- Defining Software Maintenance
- Normalization
- Analysis
- Benchmarks
- Findings/Lessons Learned
- Demographics
- Impending Analysis

SWMX Definition

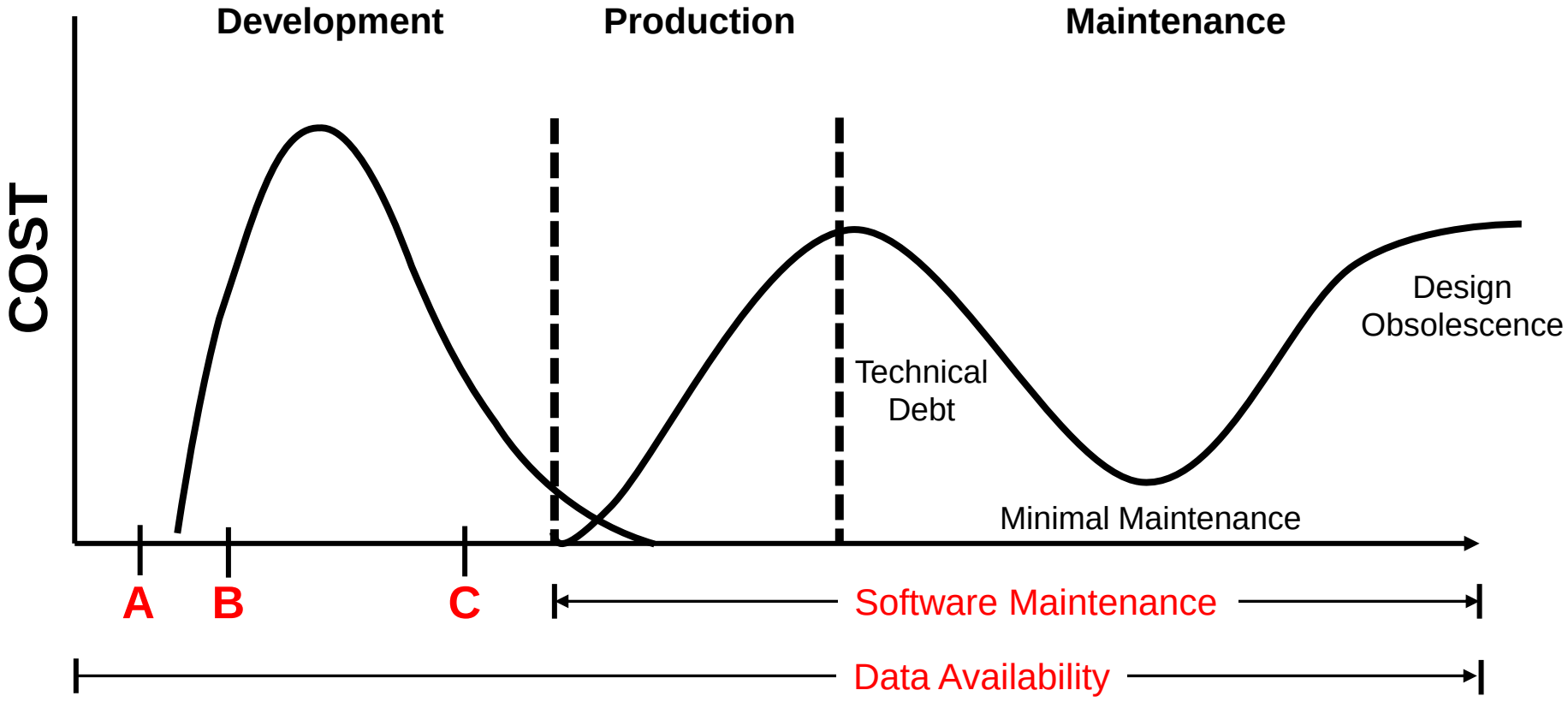
1) Correct defects and/or improve performance

2) Upgrade or modify to adapt and/or perfect the fielded software baseline to a changing/changed environment

Generic Software Maintenance Process

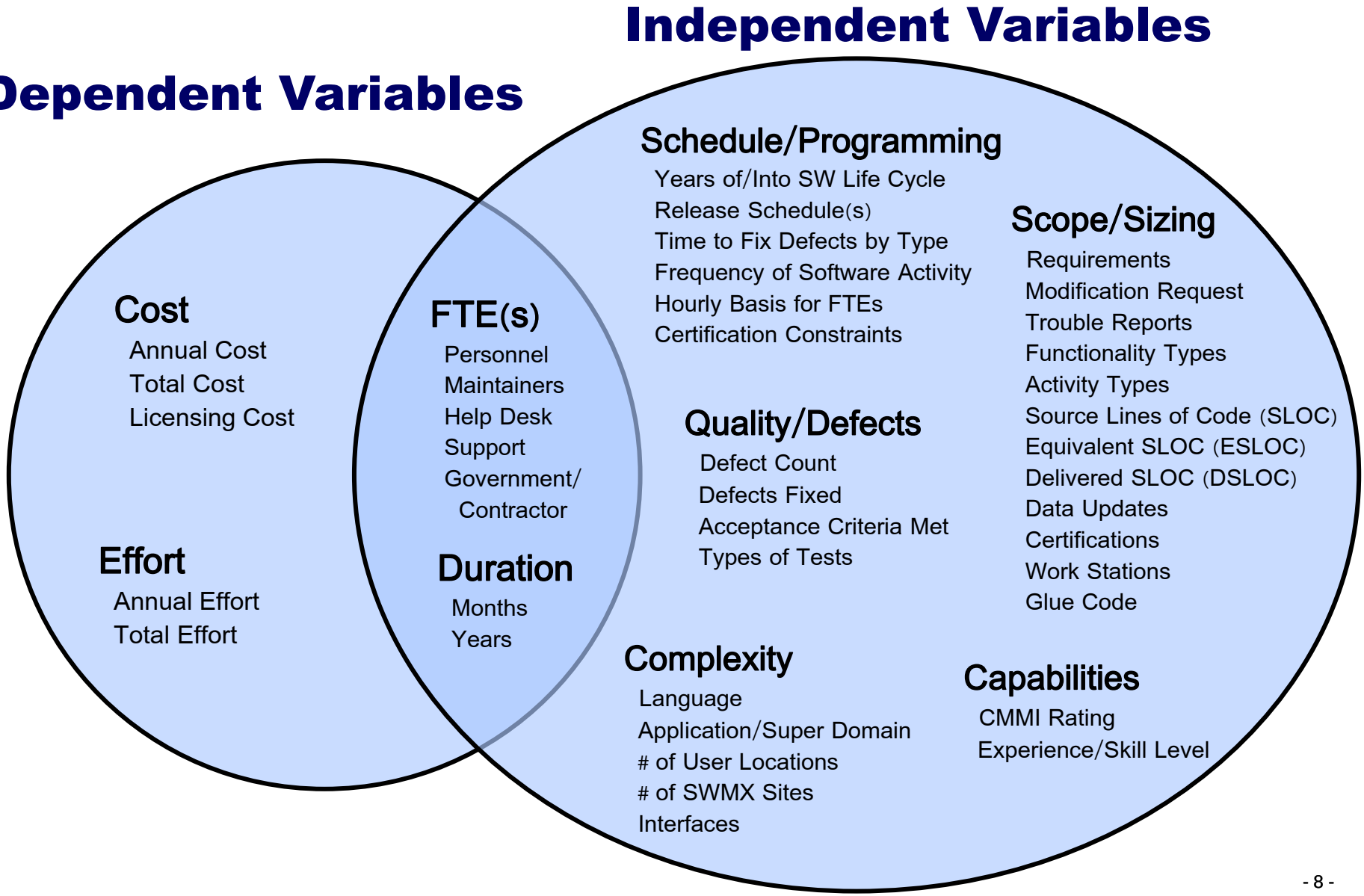


Notional Software Maintenance Life-Cycle Cost Model

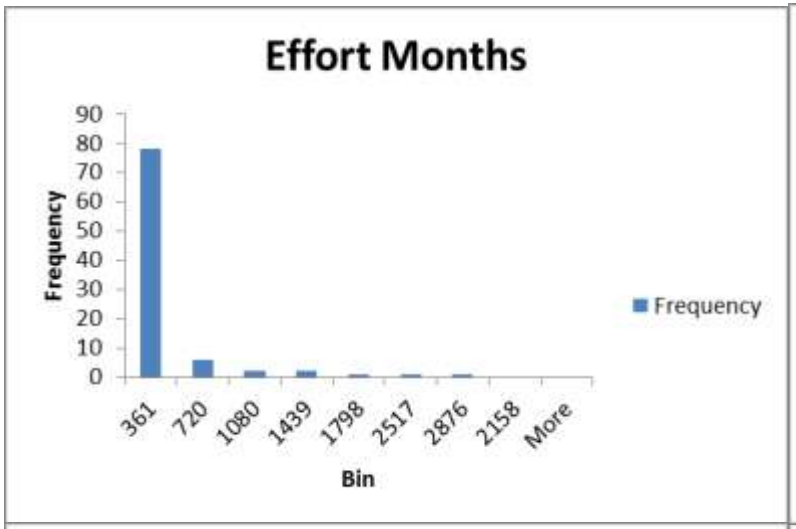


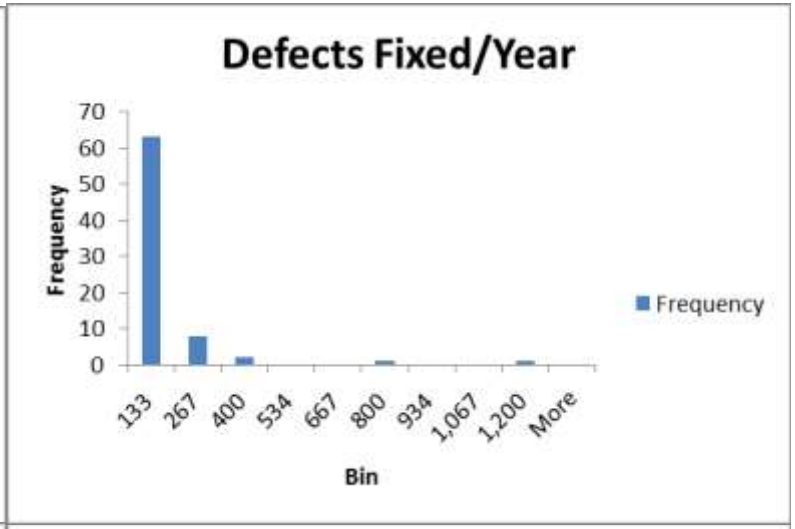
Source: Jones, Cheryl. *Estimating Software Maintenance Costs for U.S. Defense Systems.* Deputy Assistant Secretary of the Army for Cost and Economics. 1 May 2014.

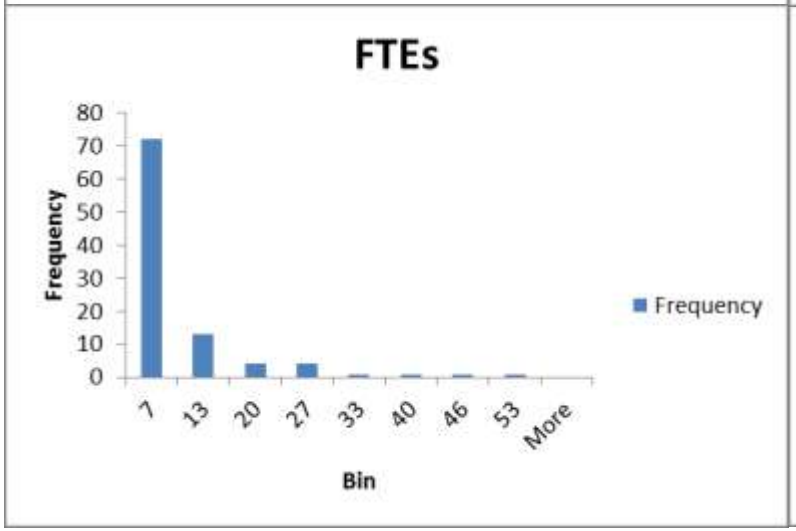
SWMX Variables

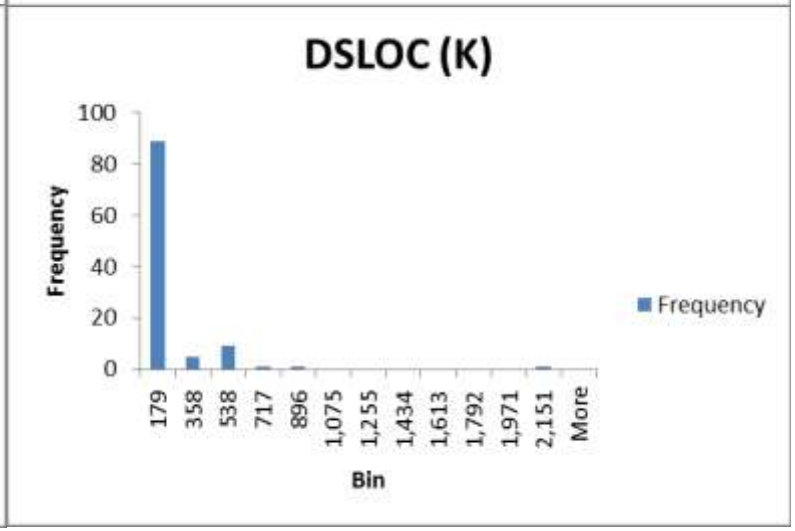


Data Demographics



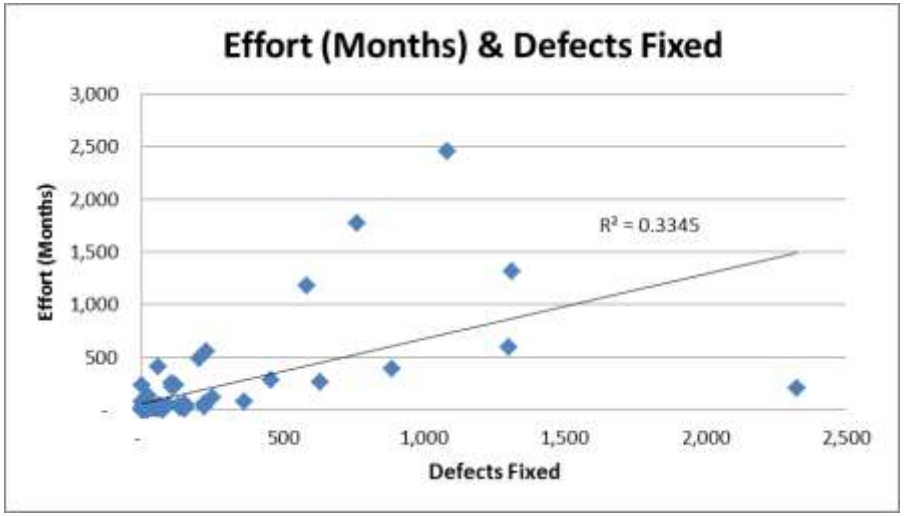


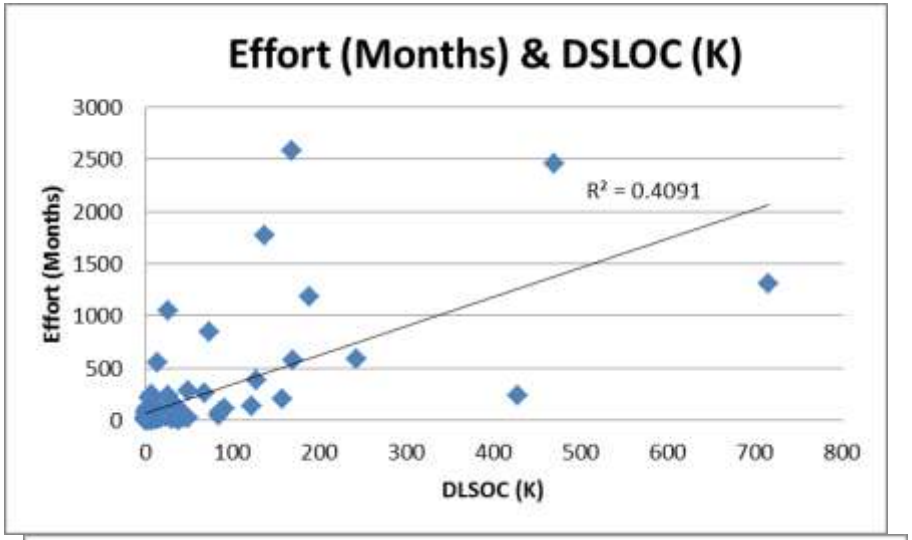


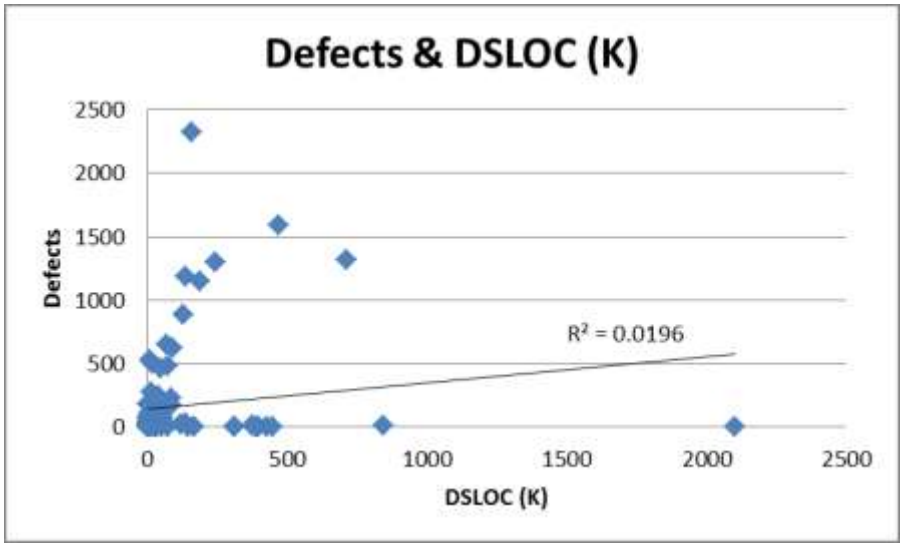


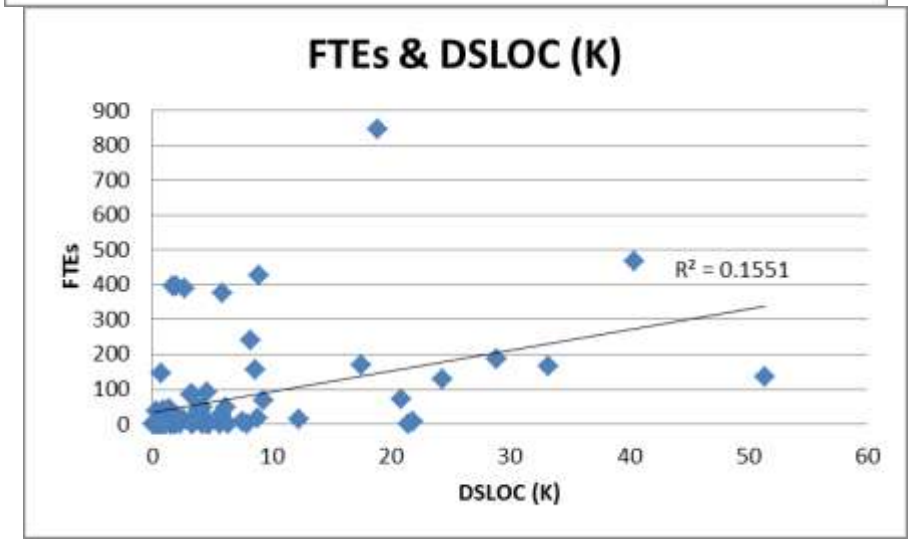
Data is highly skewed and is not normally distributed.

Exploring Data Subsets









Scatter plots show no clear trends.

Impact of Software Size

Dependent Variable	Independent Variable	Model	Type	Zero Intercept [Y/N]	n	T Stat (PNZ)	SE	R Sq	R Sq Adj	F Stat (PNZ)	DF	CoV	Range
Effort	DSLOC	Linear	Bivariate	No	83	1	59,656	41%	40%	1	81	182%	[1, 714617]
Effort	DSLOC	Linear	Bivariate	Yes	83	1	60,121	49%	48%	1	82	184%	[1, 714617]
Effort	DSLOC	Log Linear	Bivariate	No	83	1	68,163	23%	22%	1	81	208%	[0, 13.5]
Effort	DSLOC	Log Linear	Bivariate	Yes	83	1	72,360	26%	25%	1	82	221%	[0, 13.5]
Effort	ESLOC Normalized	Linear	Bivariate	No	41	1	63,562	26%	24%	1	39	200%	[15, 396598]
Effort	ESLOC Normalized	Linear	Bivariate	Yes	41	1	63,669	36%	34%	1	40	200%	[15, 396598]
Effort	ESLOC Normalized	Log Linear	Bivariate	No	41	1	63,385	26%	24%	1	39	199%	[2.7, 12.9]
Effort	ESLOC Normalized	Log Linear	Bivariate	Yes	41	1	68,608	26%	24%	1	40	216%	[2.7, 12.9]
Cost	DSLOC	Linear	Bivariate	No	24	1	9,291,482	5%	0%	1	22	166%	[580, 845000]
Cost	DSLOC	Linear	Bivariate	Yes	24	1	10,646,668	5%	1%	1	23	190%	[580, 845000]
Cost	DSLOC	Log Linear	Bivariate	No	24	0	9,509,164	0%	0%	0	22	170%	[6.4, 13.6]
Cost	DSLOC	Log Linear	Bivariate	Yes	24	1	9,383,448	26%	23%	1	23	168%	[6.4, 13.6]
FTEs	DSLOC	Linear	Bivariate	No	82	1	8.6	16%	14%	1	80	141%	[1, 845000]
FTEs	DSLOC	Linear	Bivariate	Yes	82	1	9.4	29%	28%	1	81	154%	[1, 845000]
FTEs	DSLOC	Log Linear	Bivariate	No	82	1	8.5	17%	16%	1	80	140%	[0, 13.6]
FTEs	DSLOC	Log Linear	Bivariate	Yes	82	1	9	40%	39%	1	81	142%	[0, 13.6]
FTEs	ESLOC Normalized	Linear	Bivariate	No	46	1	6.3	29%	28%	1	44	107%	[15, 396598]
FTEs	ESLOC Normalized	Linear	Bivariate	Yes	46	1	6.9	47%	46%	1	45	117%	[15, 396598]
FTEs	ESLOC Normalized	Log Linear	Bivariate	No	46	1	6.4	27%	26%	1	44	108%	[2.7, 12.9]
FTEs	ESLOC Normalized	Log Linear	Bivariate	Yes	46	1	6.7	50%	49%	1	45	114%	[2.7, 12.9]

While the models reflect large variability, they are statistically significant.

Impact of Defects Fixed

Dependent Variable	Independent Variable	Model	Type	Zero Intercept [Y/N]	n	T Stat (PNZ)	SE	R Sq	R Sq Adj	F Stat (PNZ)	DF	CoV	Range
Effort	Defects Fixed	Linear	Bivariate	No	62	1	57,679	33%	32%	1	60	187%	[1, 2324]
Effort	Defects Fixed	Linear	Bivariate	Yes	62	1	57,912	43%	42%	1	61	188%	[1, 2324]
Effort	Defects Fixed	Log Linear	Bivariate	No	62	1	59,200	30%	28%	1	60	192%	[0, 7.8]
Effort	Defects Fixed	Log Linear	Bivariate	Yes	62	1	62,648	33%	32%	1	61	203%	[0, 7.8]
Cost	Defects Fixed	Linear	Bivariate	No	49	1	3,329,810	12%	10%	1	47	160%	[1, 631]
Cost	Defects Fixed	Linear	Bivariate	Yes	49	1	3,699,430	18%	16%	1	48	178%	[1, 631]
Cost	Defects Fixed	Log Linear	Bivariate	No	49	1	3,186,478	19%	17%	1	47	153%	[0, 6.4]
Cost	Defects Fixed	Log Linear	Bivariate	Yes	49	1	3,167,516	40%	39%	1	48	152%	[0, 6.4]

While the models reflect large variability, they are statistically significant.

Impact of Size & Defects Fixed

Dependent Variable	Independent		Model	Zero Intercept [Y/N]	n	Variable 1	Variable 2	SE	R Sq	R Sq Adj	F Stat (PNZ)	DF	CoV	Range	Multi-Collinearity
	Variable (1)	Variable (2)				T Stat (PNZ)	T Stat (PNZ)								
Effort	Defects Fixed	DSLOC	Linear	No	59	0.5	1.0	46,268	59%	58%	1	56	155%	[1, 2324]	No
Effort	Defects Fixed	DSLOC	Linear	Yes	59	0.6	1.0	45,970	65%	64%	1	57	154%	[1, 2324]	No
Effort	Defects Fixed	DSLOC	Log Linear	No	59	0.9	0.7	60,019	31%	29%	1	56	201%	[0, 7.8]	Yes
Effort	Defects Fixed	DSLOC	Log Linear	Yes	59	1.0	1.0	61,979	37%	35%	1	57	208%	[0, 7.8]	No
Effort	Defects Fixed	DSLOC	Log Linear - Ridge Regression	No	59	1.0	1.0	60,323	31%	28%	1	56	202%	[0, 7.8]	No
Effort	ESLOC Normalized	Defects Fixed	Linear	No	19	0.9	0.2	23,758	45%	39%	1.0	16	118%	[884, 232877]	Yes
Effort	ESLOC Normalized	Defects Fixed	Linear	Yes	19	1.0	0.2	23,787	60%	56%	1.0	17	118%	[884, 232877]	Yes
Effort	ESLOC Normalized	Defects Fixed	Log Linear	No	19	0.8	0.9	23,968	44%	37%	1.0	16	119%	[6.8, 12.4]	Yes
Effort	ESLOC Normalized	Defects Fixed	Log Linear	Yes	19	0.5	1.0	24,918	56%	51%	1.0	17	124%	[6.8, 12.4]	Yes
Effort	ESLOC Normalized	Defects Fixed	Log Linear - Ridge Regression	No	19	0.9	1.0	24,212	43%	36%	1.0	16	120%	[6.8, 12.4]	No
Cost	Defects Fixed	DSLOC	Linear	No	12	0.7	0.1	6,128,663	14%	-5%	0.5	9	198%	[1, 631]	No
Cost	Defects Fixed	DSLOC	Linear	Yes	12	0.9	0.5	6,066,578	28%	13%	0.8	10	196%	[1, 631]	No
Cost	Defects Fixed	DSLOC	Log Linear	No	12	1.0	0.4	5,153,344	39%	26%	0.9	9	167%	[0, 6.45]	No
Cost	Defects Fixed	DSLOC	Log Linear	Yes	12	1.0	0.4	4,938,366	52%	42%	1.0	10	160%	[0, 6.45]	No
FTEs	Defects Fixed	DSLOC	Linear	No	56	1.0	1.0	8	39%	37%	1.0	53	134%	[1, 2324]	No
FTEs	Defects Fixed	DSLOC	Linear	Yes	56	1.0	1.0	8	53%	51%	1.0	54	136%	[0, 2324]	No
FTEs	Defects Fixed	DSLOC	Log Linear	No	56	1.0	0.9	8	30%	28%	1.0	53	144%	[0, 7.75]	Yes
FTEs	Defects Fixed	DSLOC	Log Linear	Yes	56	1.0	0.4	9	42%	40%	1.0	54	151%	[0, 7.75]	Yes

While the models reflect large variability, they are statistically significant.

SWMX Grouping

OPERATING ENVIRONMENT	
PLATFORM	**OPERATING ENVIRONMENT**
GROUND SITE	Manned Ground Site (MGS)
GROUND SURFACE	Manned Ground Vehicles (MGV)
	Unmanned Ground Vehicles (UGV)
MARITIME	Manned Maritime Vessel (MMV)
	Unmanned Maritime Vessel (UMV)
AIRCRAFT	Manned Aerial Vehicle (MAV)
	Unmanned Aerial Vehicle (UAV)
	Unmanned Ordinance Vehicle (UOV)
SPACECRAFT	Manned Space Vehicle (MSV)
	Unmanned Space Vehicle (USV)

SUPER DOMAIN	
MISSION CRITICAL	Embedded (MCEmb)
	Non-Embedded (MCNEmb)
MISSION SUPPORT	Embedded and Non-Embedded (MS)

Operating Environments

	ACRONYM	OPERATING ENVIRONMENT	EXAMPLES
GROUND SITE	MGS	Manned Ground Site	Command Post, Ground Operations Center, Ground Terminal, Testing Centers
GROUND SURFACE	MGV	Manned Ground Vehicles	Tanks
	UGV	Unmanned Ground Vehicles	Robots
MARITIME	MMV	Manned Maritime Vessel	Aircraft Carriers, Destroyers, Supply Ships, Submarines
	UMV	Unmanned Maritime Vessel	Mine Hunting Systems
AIRCRAFT	MAV	Manned Aerial Vehicle	Fixed-wing Aircraft, Helicopters
	UAV	Unmanned Aerial Vehicle	Remotely Piloted Vehicles
	UOV	Unmanned Ordinance Vehicle	Air-to-Air Missiles, Air-to-Ground Missiles, Smart Bombs, Strategic Missiles, Container Launch Unit
SPACECRAFT	MSV	Manned Space Vehicle	Space Shuttle, Space Passenger Vehicle, Manned Space Stations
	USV	Unmanned Space Vehicle	Orbiting satellites (for weather, communications, etc.), Exploratory Space Vehicles

Super Domains

ACRONYM	SUPER DOMAIN	DESCRIPTION	APPLICATION DOMAIN
MCEmb	Mission Critical, Embedded	Tightly coupled interfaces	Sensor Control & Signal Processing
		Real-time response required	Vehicle Control
		Very high reliability required (life critical)	Vehicle Payload
		Often severe memory and throughput constraints	Other Real Time Embedded
		Often executed on special-purpose hardware	
MCNEmb	Mission Critical, Non-Embedded	Multiple interfaces with other systems	Mission Processing
		Constrained response time required	Systems Software
		High reliability but not life critical	Automation and Process Control
		Generally executed on COTS	Simulation and Modeling
MS	Mission Support	Relatively less complex	Test
		Self-contained or few interfaces	Training
		Less stringent reliability required	Data Processing

SLOC Classifications

DSLOC = (New) + (Base) + (Converted) + (Generated) + (Modified) + (Rehost) + (Reuse)

ESLOC = 1.00 (New) + 0.03 (Base) + 0.20 (Converted) + 0.24 (Generated) + 0.03 (Deleted) + 0.80 (Modified) + 0.10 (Rehost) + 0.01 (Reuse)

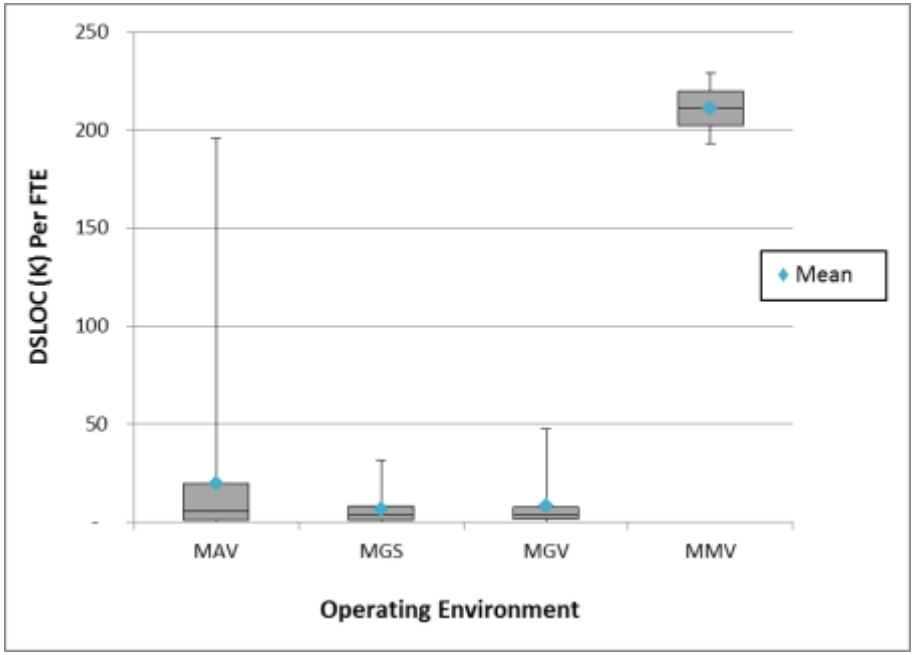
Term	Definition
New	SLOC developed from scratch
Deleted	Deleted from the previous version or release
Coder Generated	List the number of new human-generated SLOC added to the new version or release
Auto Generated	Auto-generated code produced using specialized tools at a pace far exceeding manual development
Reused (Carryover)	List the number of SLOC from the previous version that were carried over as is. These lines are not changed in any way
Modified (Carryover)	SLOC from previous releases that were changed and included in the new version or release.
Base	SLOC count from the initial starting point
Converted	Code translated to another language
Rehost	Moving SLOC from one operating systems/platform to another

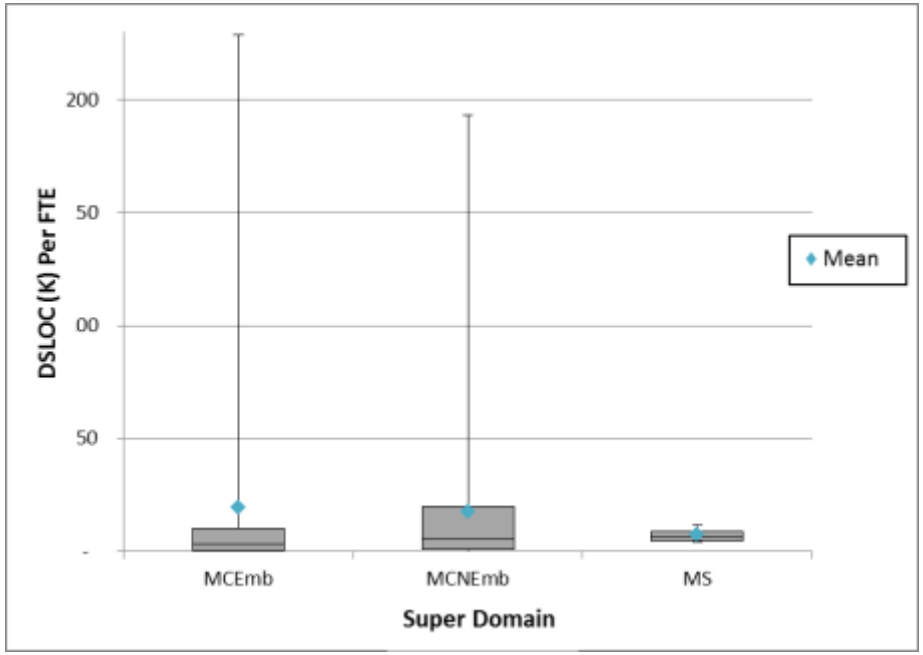
DSLOC (K)/FTE Benchmarks

DSLOC = (New) + (Base) + (Converted) + (Generated) + (Modified) + (Rehost) + (Reuse)

Operating Enviroment				
DSLOC (K) / FTE				
	MAV	MGS	MGV	MMV
Count	46	12	11	2
Q1	1.1	0.9	1.9	202.1
Median	5.3	3.7	3.8	211.0
Q3	19.6	8.1	7.9	220.0

Super Domain			
DSLOC (K) / FTE			
	MCEmb	MCNEmb	MS
Count	57	21	4
Q1	0.3	0.8	4.7
Median	3.1	5.3	6.4
Q3	10.2	19.7	8.9





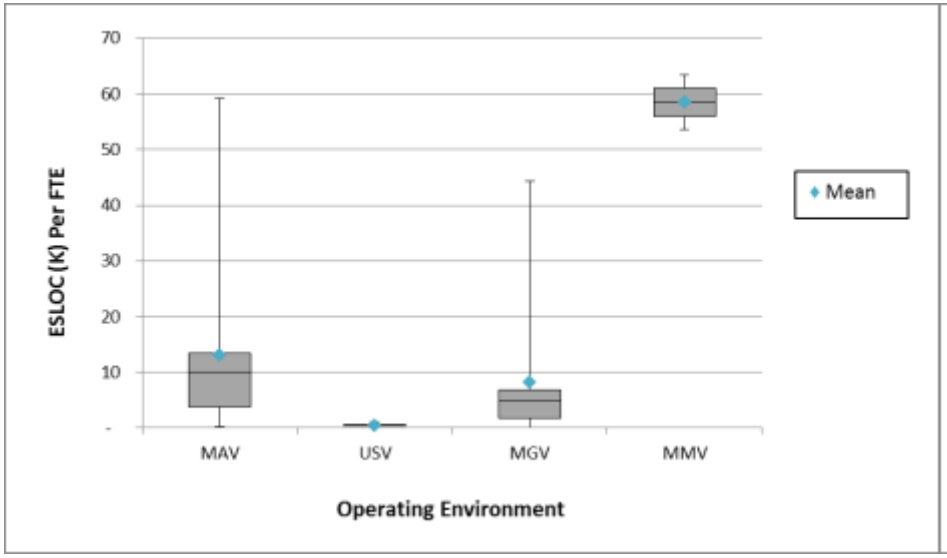
Grouping data helped decrease variability for certain subcategories. However, more data is still needed.

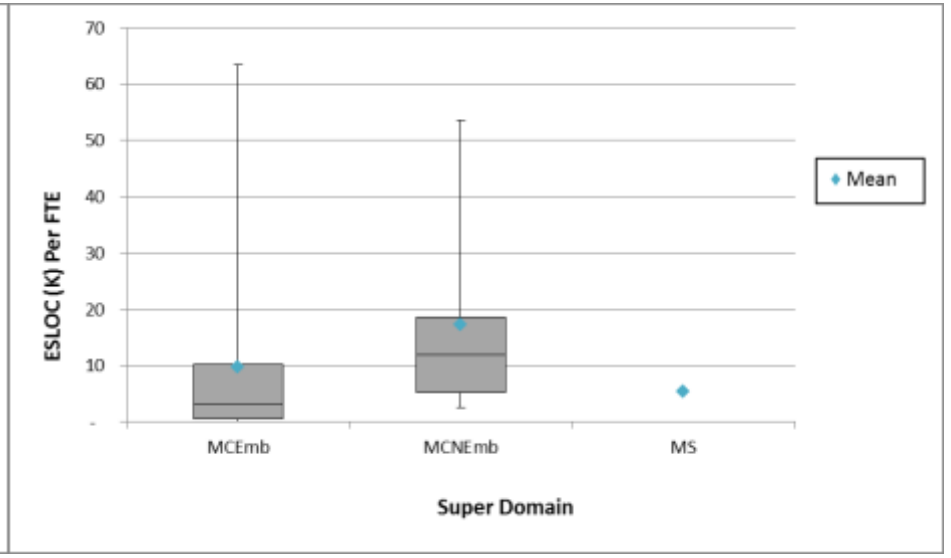
ESLOC (K)/FTE Benchmarks

ESLOC = 1.00 (New) + 0.03 (Base) + 0.20 (Converted) + 0.24 (Generated) + 0.03 (Deleted) + 0.80 (Modified) + 0.10 (Rehost) + 0.01 (Reuse)

Operating Enviroment				
ESLOC (K) / FTE				
	MAV	USV	MGV	MMV
Count	22	12	10	2
Q1	3.8	0.3	1.7	56.0
Median	10.0	0.4	4.8	58.5
Q3	13	1	7	61

Super Domain			
ESLOC (K) / FTE			
	MCEmb	MCNEmb	MS
Count	39	6	1
Q1	0.6	5.3	5.4
Median	3.2	12.0	5.4
Q3	10	19	5



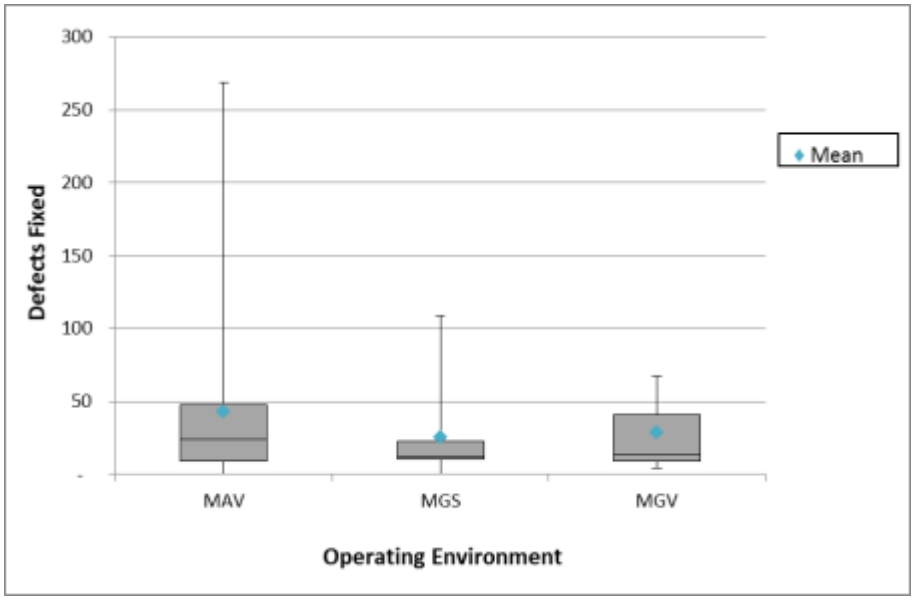


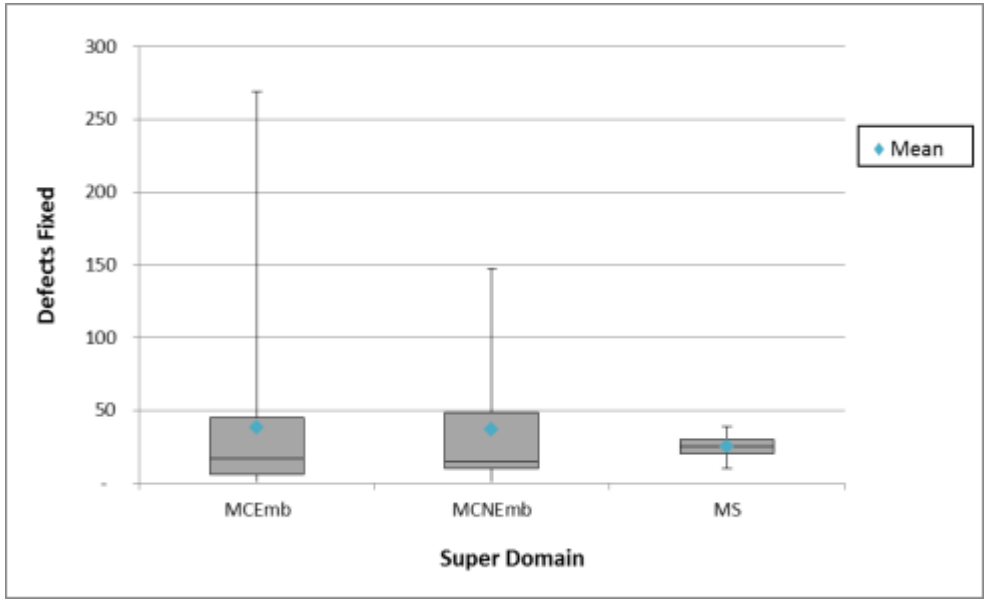
Grouping data helped decrease variability for certain subcategories. However, more data is still needed.

Defects Fixed/FTE Benchmarks

Operating Enviroment			
Defects Fixed / FTE			
	MAV	MGS	MGV
Count	43	12	3
Q1	8.7	10.2	8.7
Median	23.9	12.5	13.5
Q3	47.9	22.7	40.5

Super Domain			
Defects Fixed / FTE			
	MCEmb	MCNEmb	MS
Count	57	21	4
Q1	5.7	10.2	20.4
Median	16.7	14.6	25.3
Q3	44.9	48.3	29.8





Grouping data helped decrease variability for certain subcategories. However, more data is still needed.

CER for an Operating Environment Using Size

I. Model Form and Equation Table

Model Form:	Unweighted Linear model
Number of Observations Used:	12
Equation in Unit Space:	FTEs = Coefficient * ESLOC(K)

II. Fit Measures (in Fit Space)

Coefficient Statistics Summary

Variable	Coefficient	Std Dev of Coef	Beta Value	T-Statistic (Coef/SD)	P-Value	Prob Not Zero
Intercept						
ESLOC(K)	****	0.2747	0.9416	9.2734	0.0000	1.0000

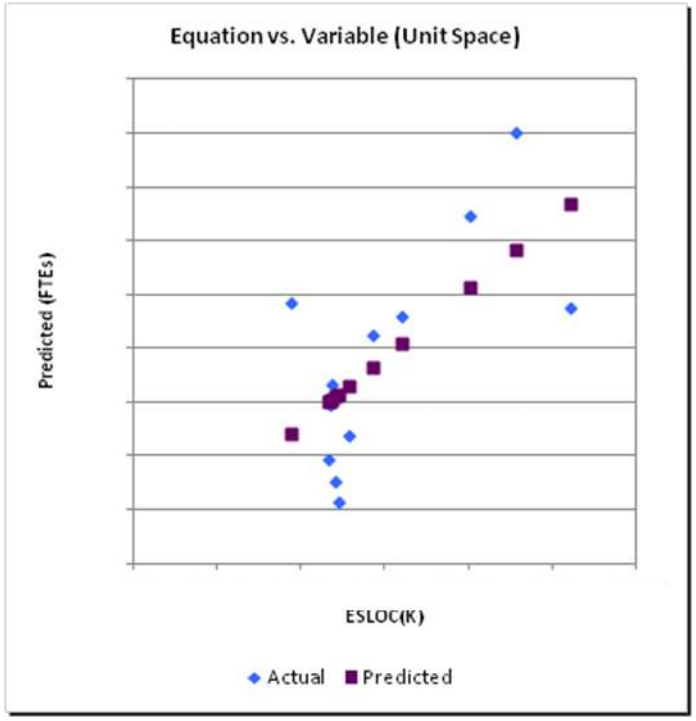
Goodness-of-Fit Statistics

Std Error (SE)	R-Squared	R-Squared (Adj)	Pearson's Corr Coef
1.5153	88.66%	87.63%	0.9416

Analysis of Variance

Due To	DF	Sum of Sqr (SS)	Mean SQ = SS/DF	F-Stat	P-Value	Prob Not Zero
Regression	1	197.4700	197.4700	85.9954	0.0000	1.0000
Residual (Error)	11	25.2591	2.2963			
Total	12	222.7291				

CV (Coefficient of Variation based on Standard Error (SE/Avg Act)) 51%



Lessons Learned

- Varied definitions for software maintenance processes impacted data variability
  - Data reported from different agencies were not consistent
  - More specific data collection request form
- Continued data and metadata documentation and collection to improve SWMX cost estimating

Next Steps

- Continue data and information gathering
- Partner with software engineers and their management
- Add to cost analysts' knowledge to create usable, useful factors, EERs, and CERs over time

Contributors:

NCCA

Corinne Wallshein

Shelley Dickson

Alex Thiel

Bruce Parker

ARDEC

Cheryl Jones

Technomics, Inc.

Peter Braxton

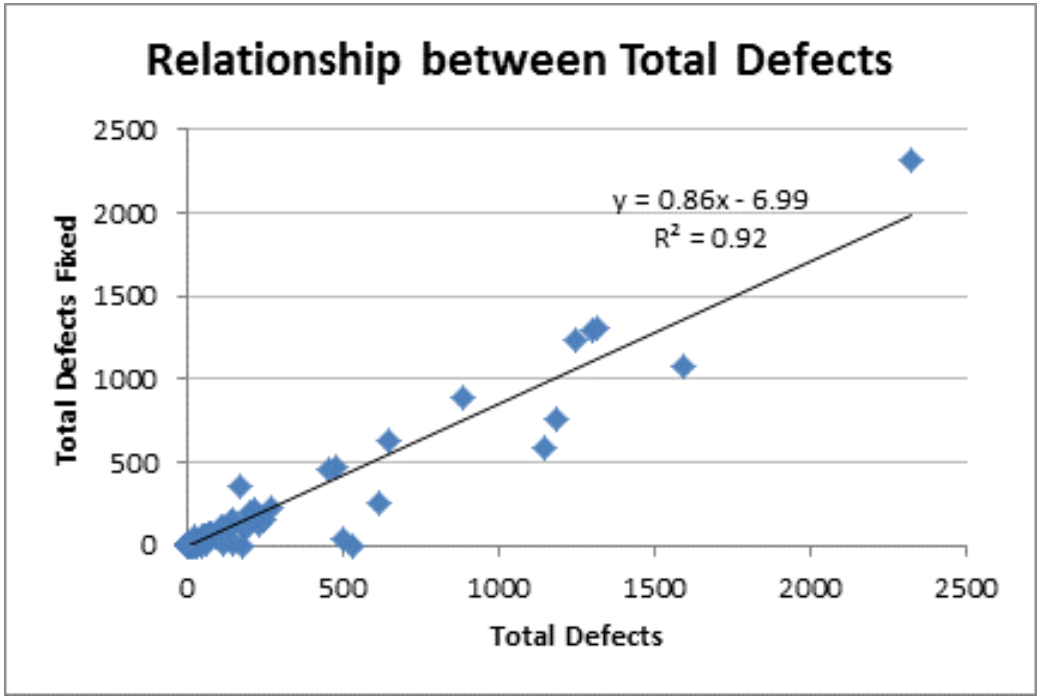
Thomas Harless

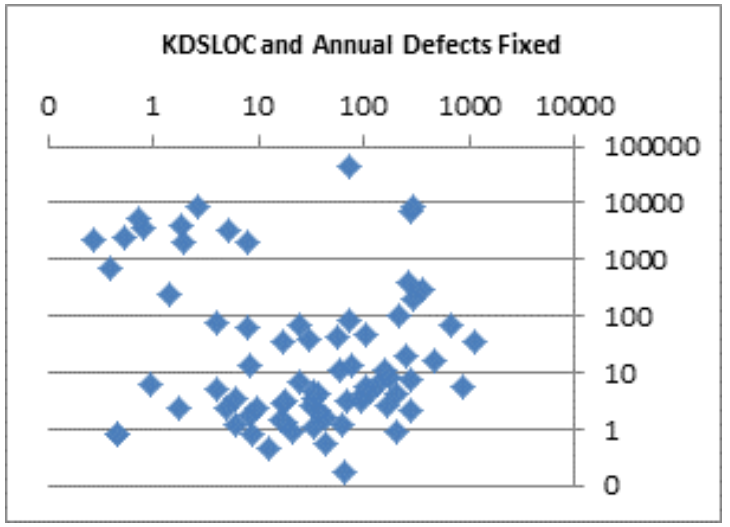
Vanessa Welker

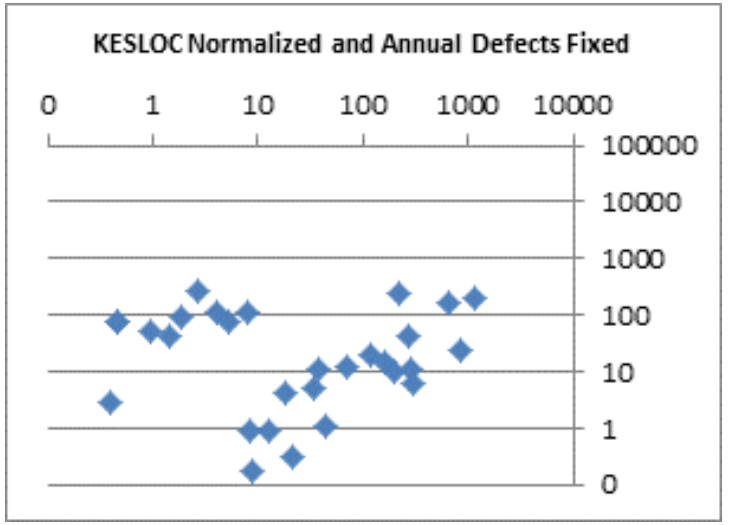
If you would like to contribute towards this effort or have any further questions, please contact Shelley Dickson at shelley.dickson@navy.mil or 703-604-3548.

Backup Slides

Data Relationship Checks







Due to this relationship <u>total defects fixed</u> or <u>total defects</u> may be used in regression.

Application Domains

ACRONYM	APPLICATION DOMAIN	DESCRIPTION	EXAMPLES
SCP	Sensor Control and Signal Processing	Software requiring timing-dependent device coding to enhance, transform, filter, convert, or compress data signals	Signal Processing, Sonar Signals, Radar Signals
VC	Vehicle Control	Hardware and software to control vehicle primary and secondary mechanical devices and surfaces	Bus, Platform, Executive, Operational Flight Program (OFP)
VP	Vehicle Payload	Hardware and software to control and monitor vehicle payloads and to provide communication to other vehicle subsystems and payloads	Payload, Weapons Delivery
RTE-Other	Real Time Embedded – Other	Real time data processing software embedded on platform/device designed to operate with tight resource constraints	Communication, Navigation, Electronic Warfare, Sensor Data Processing, Controls & Displays
MP	Mission Processing	Onboard master data processing unit(s) responsible for coordinating and directing major mission systems	Situational Awareness, Mission Management, Launch & Recovery, Environmental Control, Bombing Computer, Display Processors, Flight Control Computers, Electronic Tactical Data System
SYS	Software Systems	Software layers between the computing hardware and applications	Command and Control, Information Assurance, Infrastructure, Middleware, Maintenance and Diagnostics, Telecommunications
PC	Automation and Process Control	Software for automated systems	Process Control
S&M	Simulation and Modeling	Software to evaluate numerous scenarios by simulating events and situations with live personnel	Simulation, Modeling
TRN	Training	Applications used for educational and training including the required hardware configurations and software applications	Training for various situations (e.g., Mission Planning)
Test	Test	Applications used for testing purposes including their required hardware and software configurations	Automated Test Equipment (ATE) and Test Package Sets (TPS)
DP	Data Processing	Software to automate a common business function	Payroll, Financial Transactions, Personnel Management, Order Entry, Inventory Management, Logistics, Database