

Data Abstraction and Review

Subject Level

- Edit Checks
- Monitoring, SDV
- Data Listings
- Other Misc.

Site/Country/Cohort level

- Custom Reports, Visualization
- Risk Based Monitoring, KRIs, KPIs
- **Central Statistical Monitoring (CSM)**

Anomaly Detection Process

Data Preparation

- **Data Integration**
- **Data Selection**
- **Feature Extraction**

Anomaly detection algorithms

- **Univariate Anomalies**
- **Complex patterns**

Analyze findings

- **Consolidate anomalous data patterns**
- **Visualize**
- **Dig into data to confirm findings**

Our Improvements



**Make data preparation
unsupervised**

Takes advantage of CDISC
standard



Deeper learning

Dissect data in many ways
Rule based anomalies



**Use simpler and sound
statistical approaches**

Non-parametric tests
Tukey fences



**Actionable findings along
with powerful visualization**

Consolidated findings in excel
Very detailed and consistent
visualizations

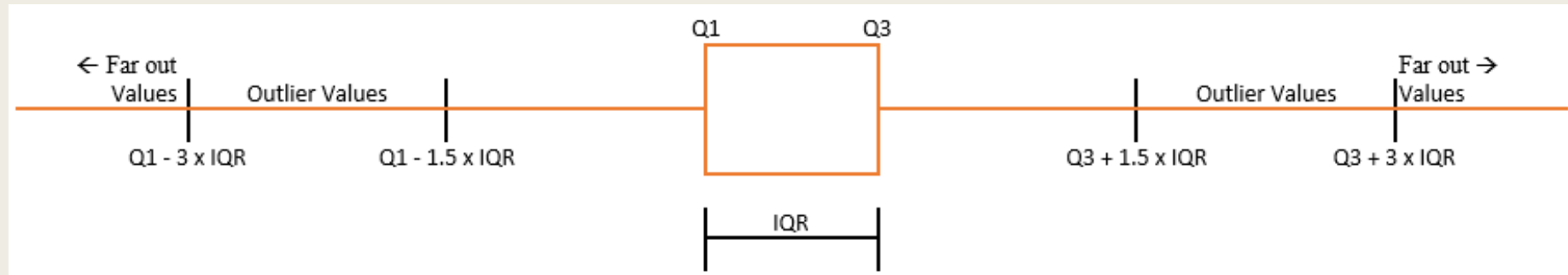
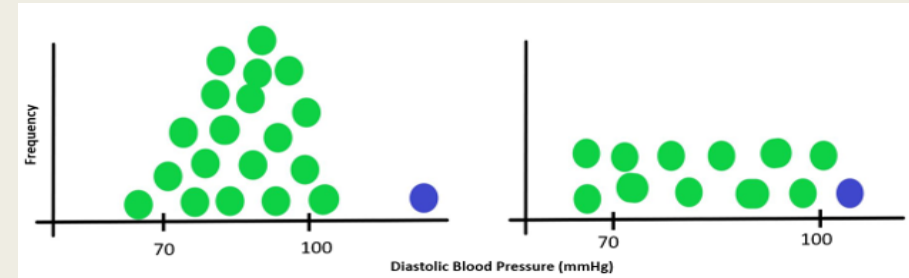
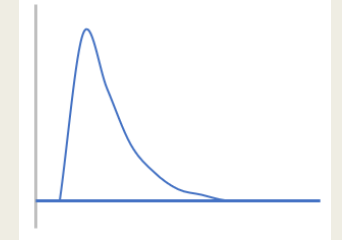
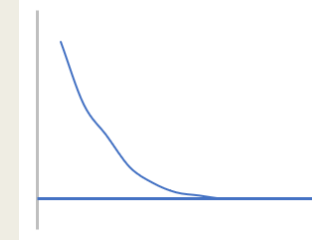
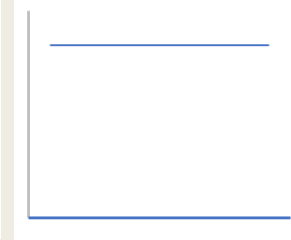


Free Software

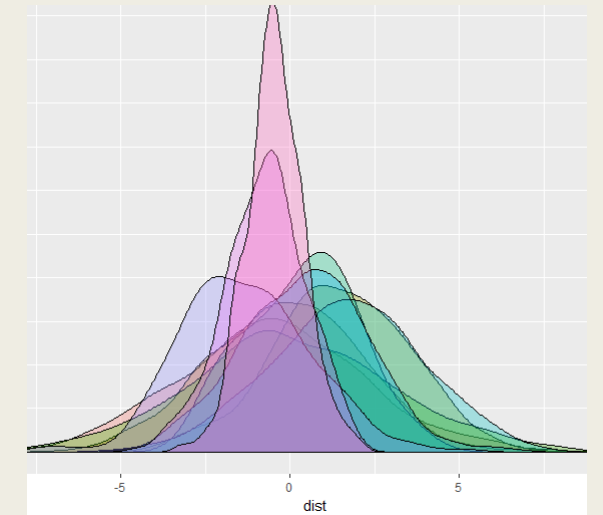
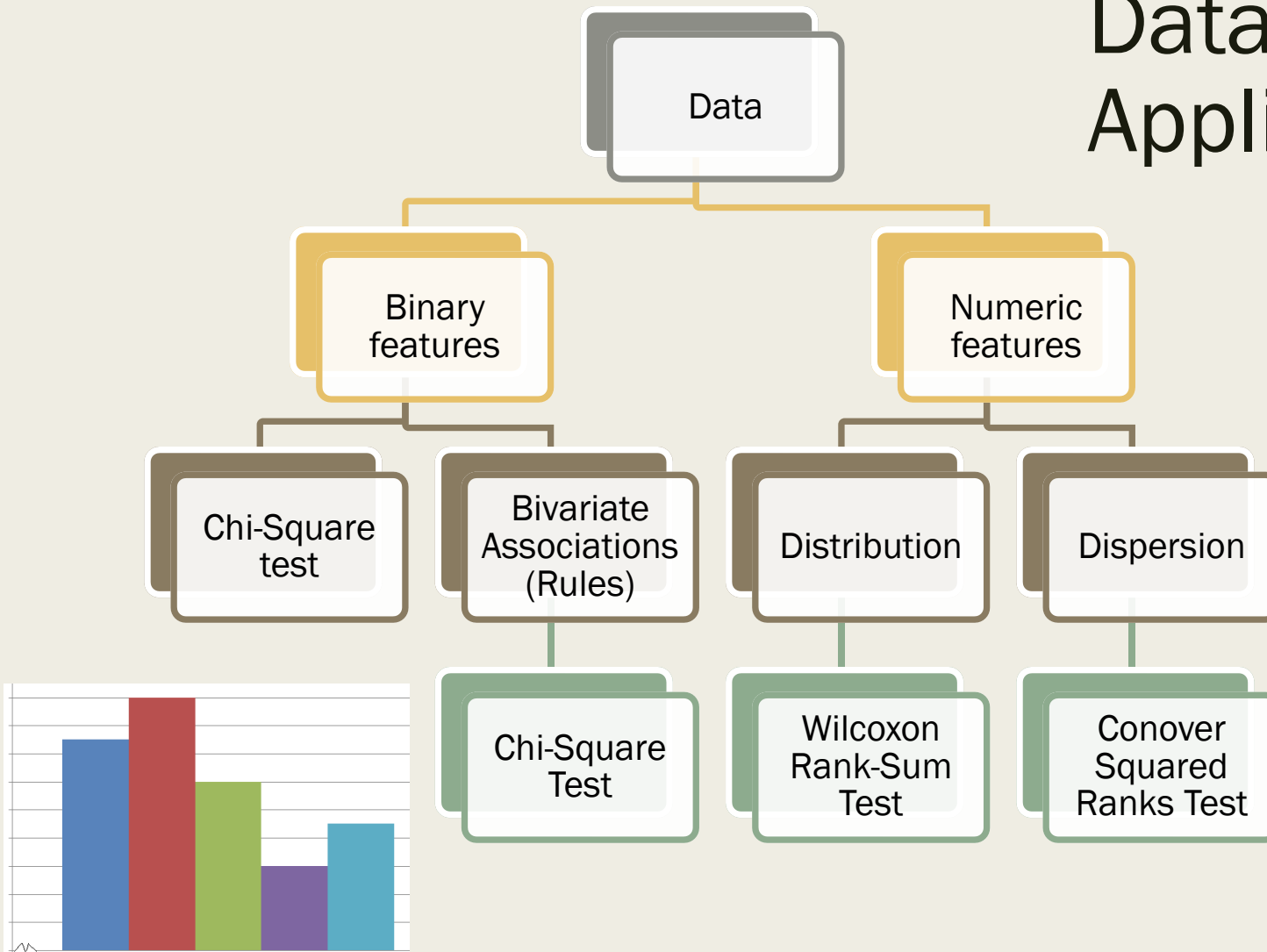
New study can be set-up in hours

Statistical Models

- Unknown Distribution of data
- Non-parametric
 - *Wilcoxon Rank Sum Test*
 - *Conover Squared Ranks Test*
 - *Chi-Square Test*
- Supplemented by Tukey Fences



Data features and Application of models



Rule based Anomalies

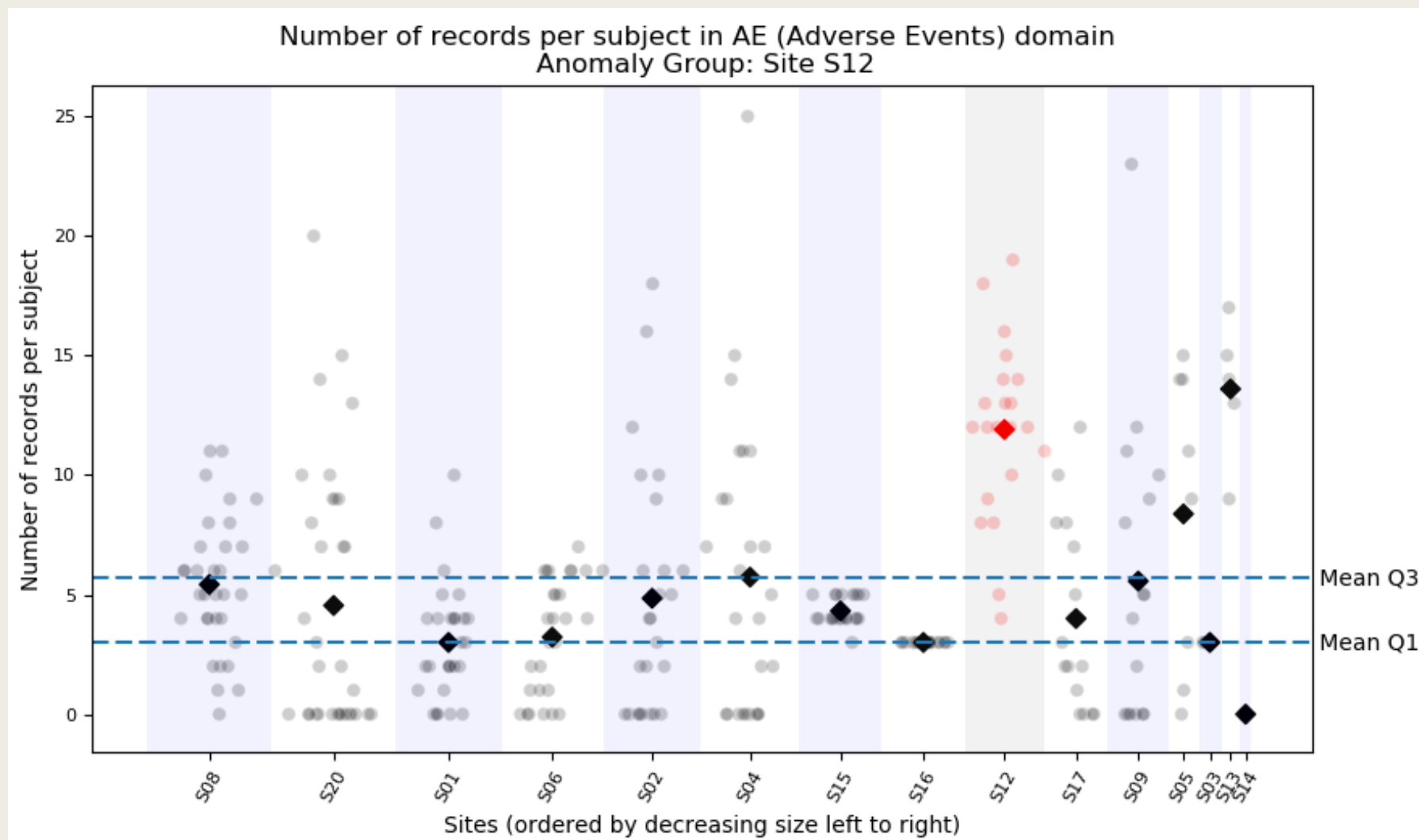
Overall AE Data			
		Severe	
		Yes	No
Serious	Yes	30	10
	No	25	75

Site 101 AE Data	
Severe	
Yes	No
5	8
2	5

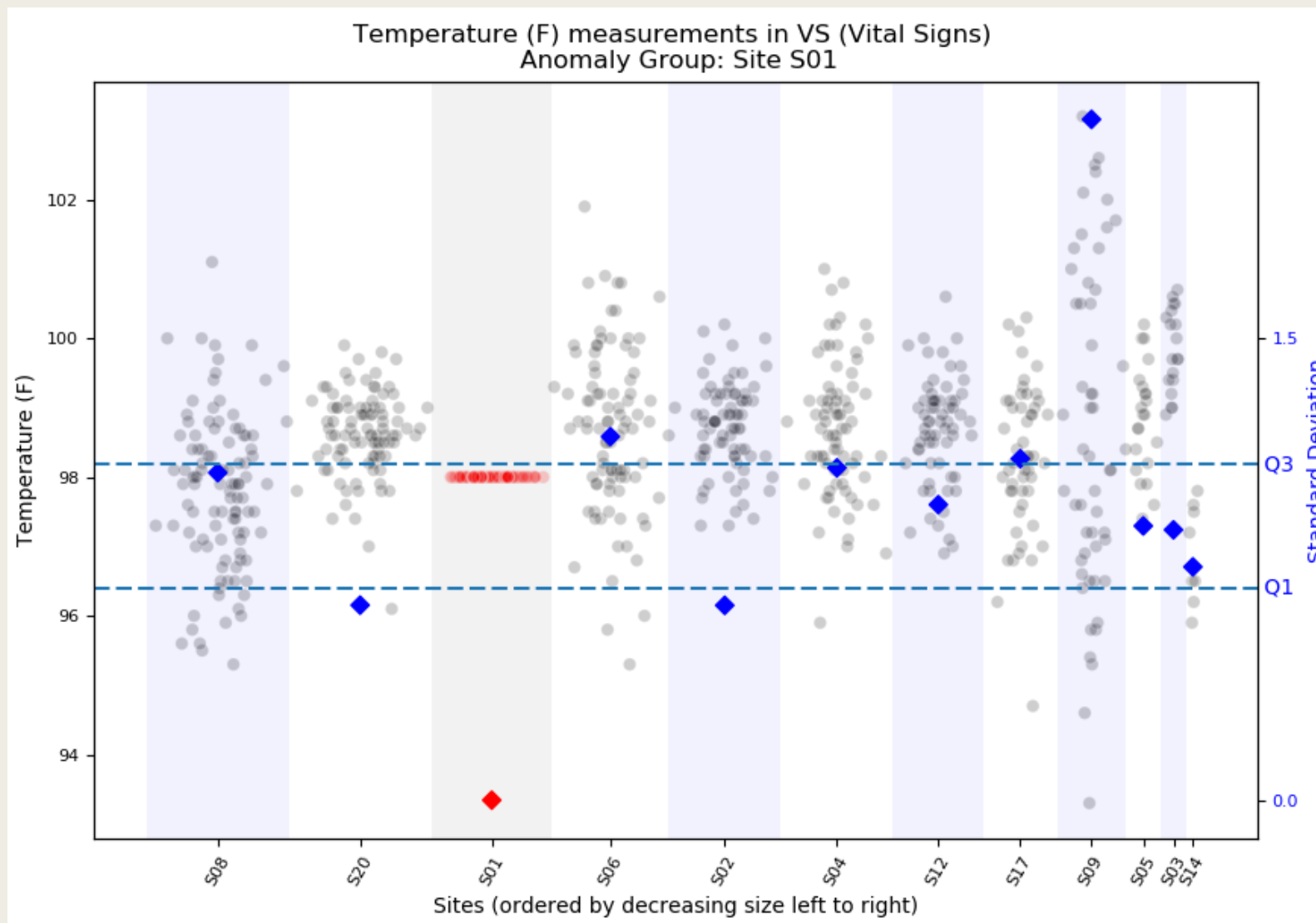
Presentation of Results

Group	Data				Finding Type	Records		Non-Parametric Statistics		Tukey Fences		
	Dataset	Data Split	Variables	Data Selection		Numbers	%	Model	p-value	Statistic	Q1, Q3	Distance from IQR
SITEID = '101'	CM	Category for Medication (CMCAT) = 'CONCOM-ITANT MEDS'	Start Date (CMSTDTC) w.r.t to Treatment Start Date (Relative)		Numeric Distribution	20 out of 523	3.8	Wilcoxon rank-sum	0.00005	Mean: -6	2, 5	2.0
SITEID = '103'	VS	Vital Signs Test Name (VSTEST) = 'Temperature'	Numeric Result/Finding in Standard Units (VSSTRESN)		Numeric Dispersion	210 out of 2212	9.5	Conover squared ranks	<.00001	SD: 1.13	.11, 0.27	5.4
SITEID = '105'	AE	NA	Serious Event (AESER)	AESER= 'Mild'	Univariate Binary Distribution	5 out of 34 vs. 352 out of 383	8.9	2x2 Chi-Square	<.00001	Ratio: .15	.75, 1	2.4
SITEID = '107'	CM	Category for Medication (CMCAT) = 'Rescue Medication'	Reported Name (CMTRT) & Dose per Administration (CMDOSE)	CMTRT = 'Drug A' & CMDOSE > 5	Bivariate Binary Distribution	15 out of 100 vs. 5 out of 353	22.1	2x2 Chi-Square	<.00001	Ratio: .15	0, .03	4.0

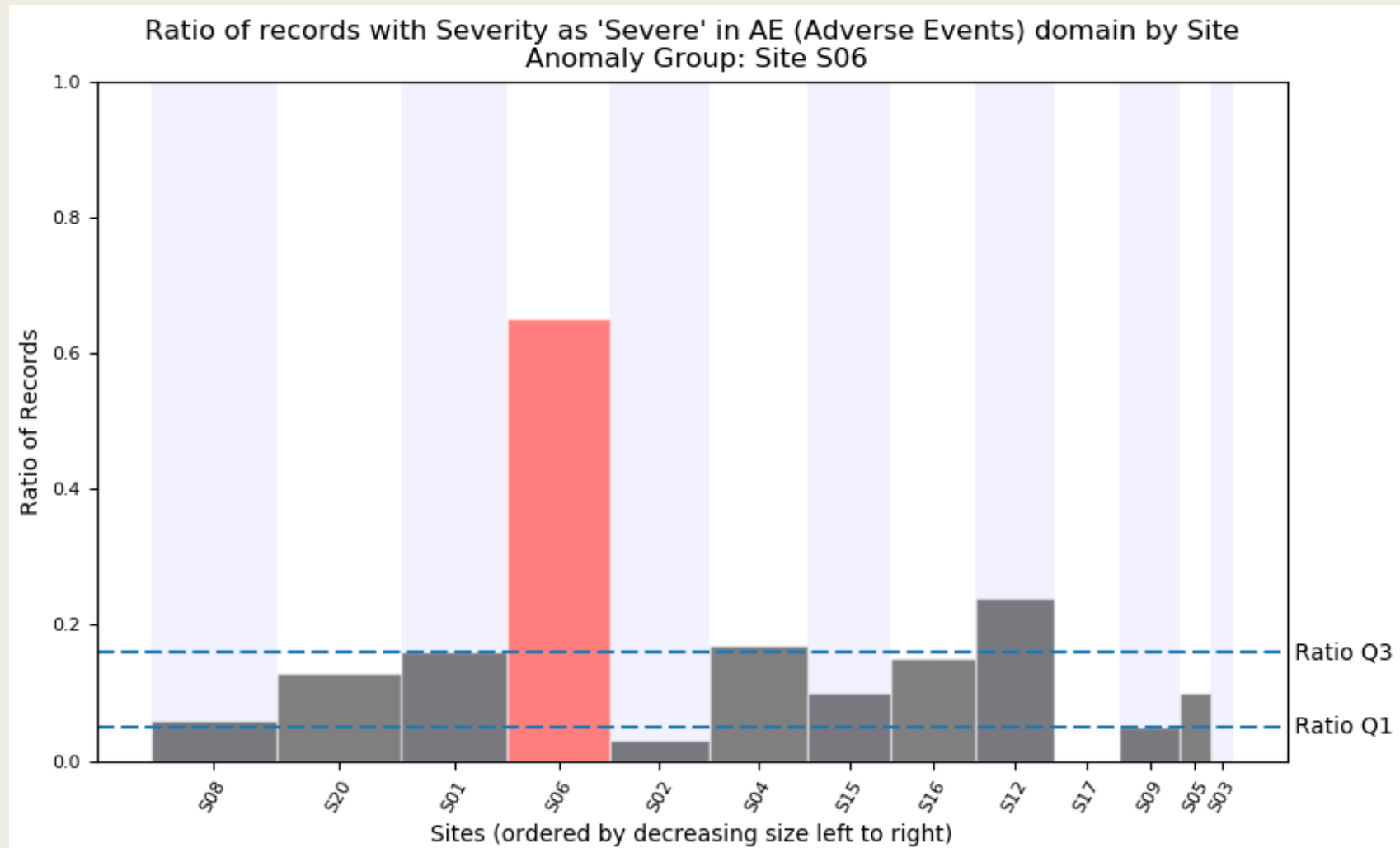
Visualization: Distribution of numeric data



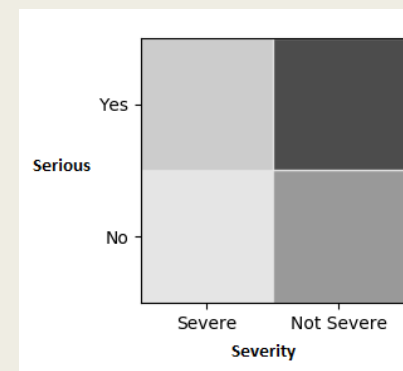
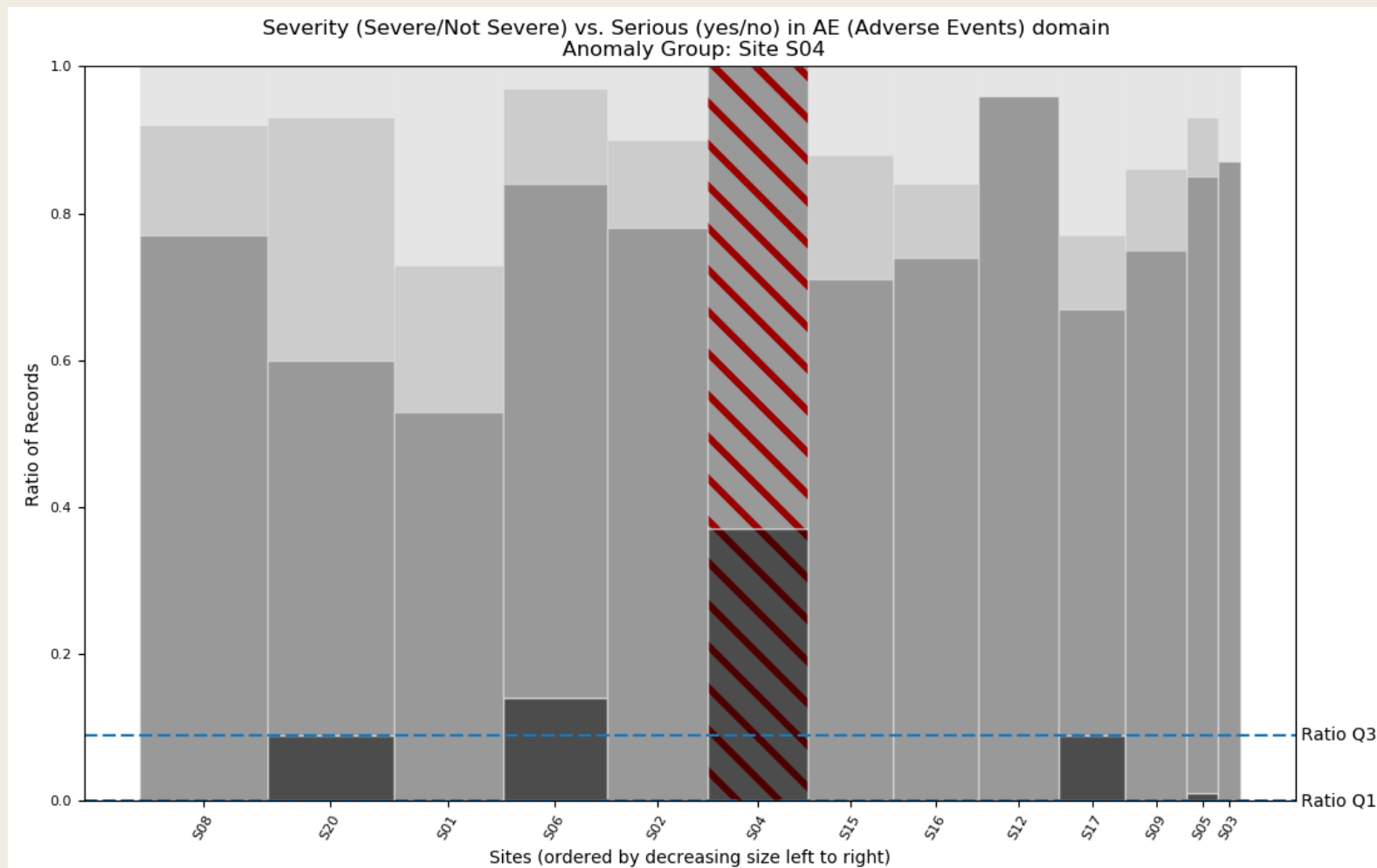
Visualization: Dispersion in numeric data



Visualization: Binary Distribution



Visualization: Rule based Bivariate anomalies



Summary

- CSM process can be **daunting** and **costly** since:
 - *No common standard, tedious & study-specific*
 - *May require lot of cross functional support*

- It can be made lot **simpler** and **economical** by:
 - *Automating the data preparation*
 - *Use of simpler and intuitive statistical techniques*
 - *Clear & actionable findings*
 - *Powerful and consistent visualizations*
 - *Free downloadable software*



THANK YOU
&
QUESTIONS