

Fingerprinting Analysis

An initial view of the data show various ways to fingerprint the dicom pixel array based on included metadata. Missing values were filtered out to make figures.

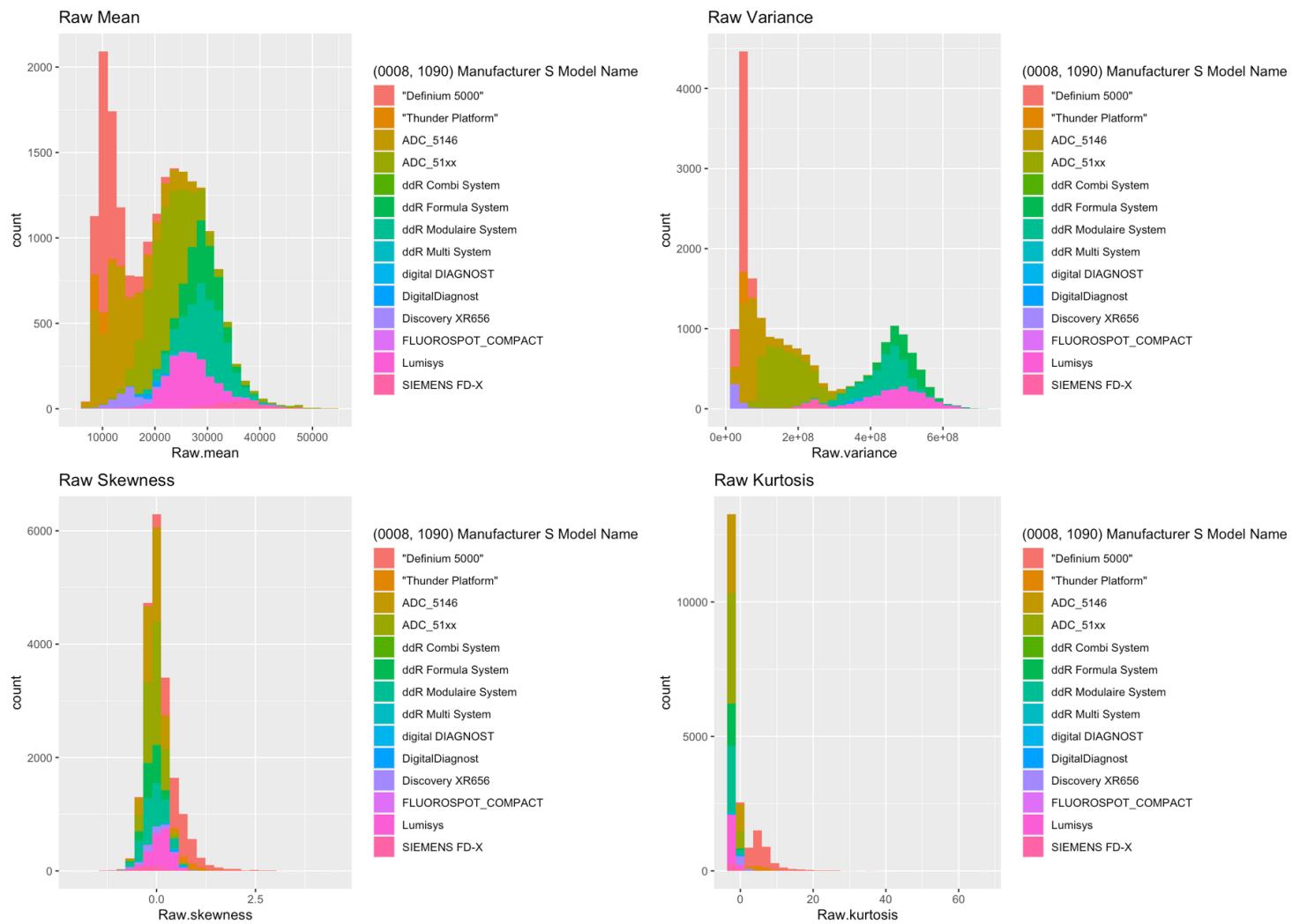
Manufacturer Model Name

Looking at the dicom attribute (0008, 1090) or Manufacturer model name. The variable is used as the color fill for scatter plots of pixel array mean and standard deviation. The plot is split according to successive LUT pixel data transformations. More information can be found here (<https://pydicom.github.io/pydicom/dev/reference/handlers.html#>). The greatest separability can be seen after applying the windowing operation.



The limited data for the Rescale transformation is due to there being only two cases with specified (0028,1052) Rescale Intercept and (0028,1053) Rescale Slope values and a non empty manufacturer model name. Mean and Variance give better linear separability than other statistics of the pixel array. Below

are histograms of both various distribution statistics for the pixel arrays. There is considerably more overlap of different manufacturer models in regards to array skewness and kurtosis than mean and variance .



A table of the image counts by model name shows an unequal distribution of X-rays by manufacturer models. The values for mean raw pixel value and VOI raw pixel value indicate the presence of a noop .

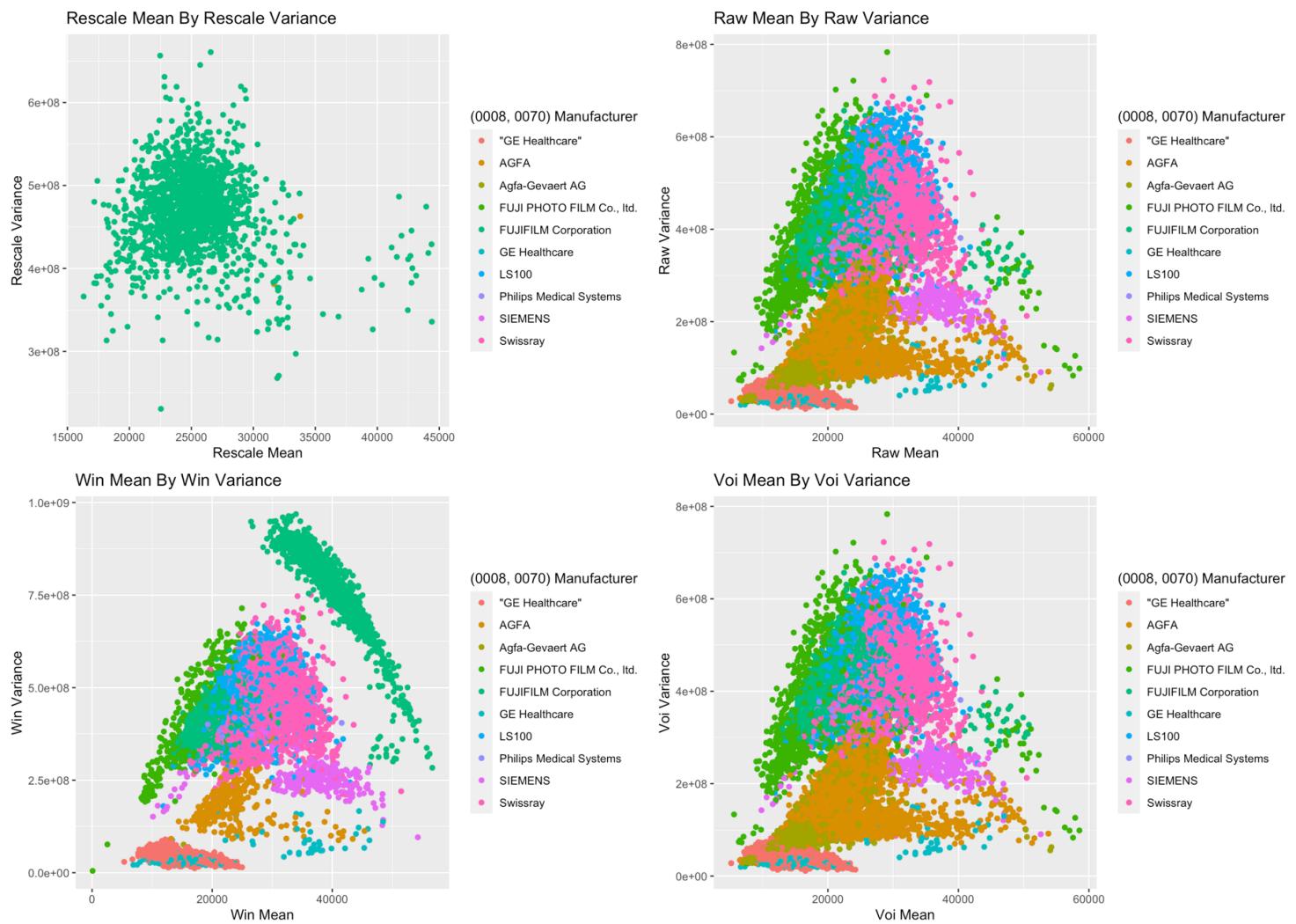
Manufacturer Model Name

X.0008..1090..Manufacturer.s.Model.Name	N	Raw.mean	Rescale.mean	Win.mean	VOI.mean
ADC_51xx	4731	22864.89	32722.94	NA	22864.89
ADC_5146	3938	14194.47	NaN	NA	14194.47
"Definium 5000"	3470	11611.16	NaN	NA	11611.16
ddR Modulaire System	2693	29539.21	NaN	29705.76	29539.21
Lumisys	1966	26387.33	NaN	26387.22	26387.33
ddR Formula System	1635	28947.30	NaN	29769.43	28947.30

"Thunder Platform"	416	10176.43	NaN	NA	10176.43
Discovery XR656	406	16375.63	NaN	16885.51	16375.63
SIEMENS FD-X	324	34717.77	NaN	35499.17	34717.77
DigitalDiagnost	93	19919.66	NaN	20540.19	19919.66
ddR Multi System	29	26192.57	NaN	26192.57	26192.57
ddR Combi System	4	23794.39	NaN	24544.64	23794.39
digital DIAGNOST	2	19886.14	NaN	20517.15	19886.14
FLUOROSPOT_COMPACT	1	52626.12	NaN	54251.81	52626.12

Manufacturer

Looking at (0008, 0070) shows similar results. Interestingly, only two manufacturers (FUJIFILM Corporation and AGFA) have specified rescale intercept/slope values. There appears to be better separability in comparison to (0008, 1090) Manufacturer Model Name especially when viewing the windowed frame (values for (0028,1050) Window Center and (0028,1051) Window Width specified).



There is overlap with manufacturer model name and a similarly uneven distribution of counts.

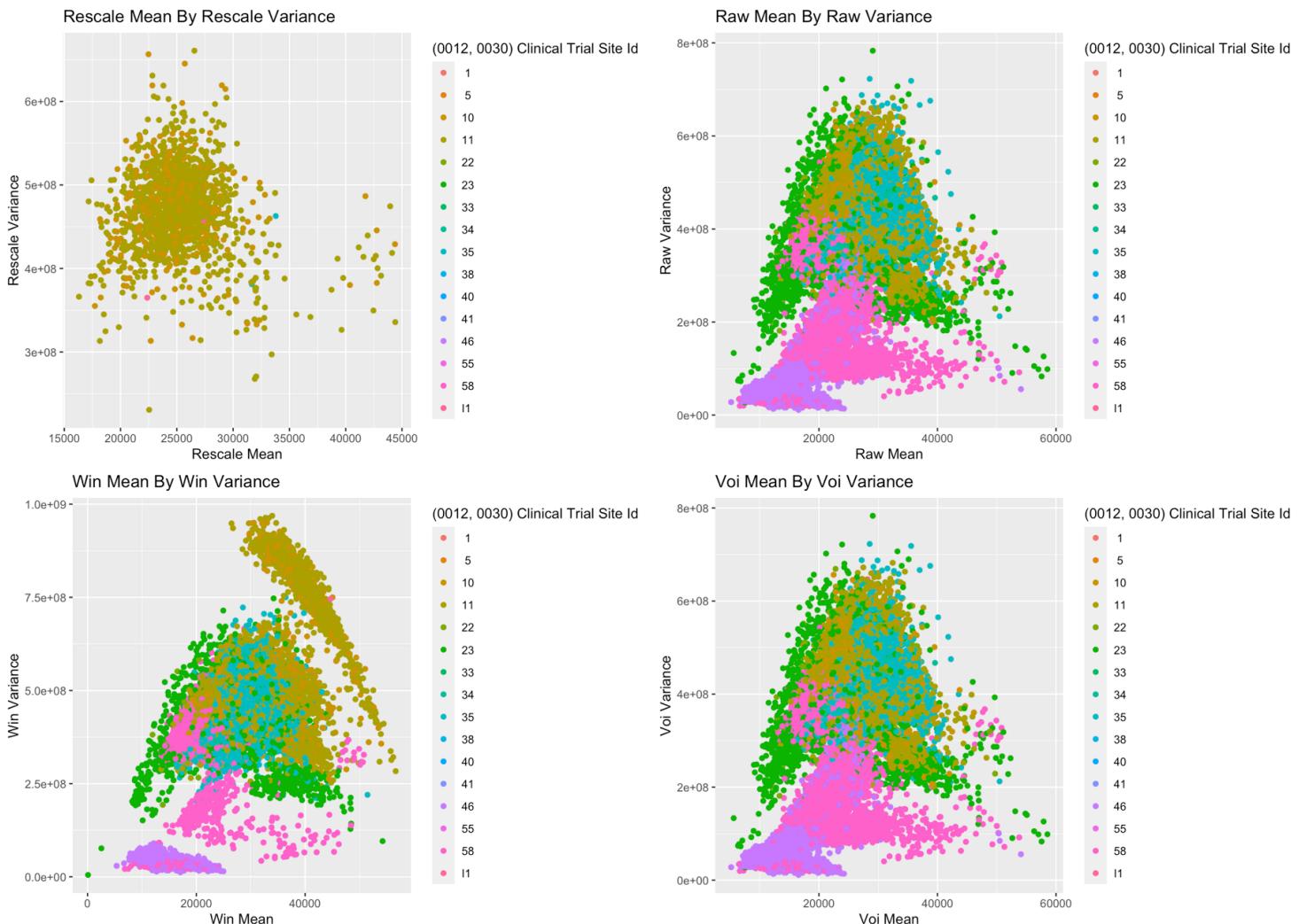
Manufacturer

X.0008..0070..Manufacturer	N	Raw.mean	Rescale.mean	Win.mean	VOI.mean
AGFA	4731	22864.89	32722.94	NA	22864.89
Swissray	4361	29289.77	NaN	29701.53	29289.77
Agfa-Gevaert AG	3938	14194.47	NaN	NA	14194.47
"GE Healthcare"	3886	11457.57	NaN	NA	11457.57
FUJIFILM Corporation	3703	22233.54	25250.90	NA	22233.54
LS100	2137	26354.07	NaN	26353.96	26354.07
FUJI PHOTO FILM Co., Ltd.	1202	18694.26	NaN	NA	18694.26
GE Healthcare	406	16375.63	NaN	16885.51	16375.63

SIEMENS	325	34772.88	NaN	35556.87	34772.88
Philips Medical Systems	95	19918.95	NaN	20539.71	19918.95

Clinical Trial Site ID

Looking at other possible partitions of the pixel arrays, the trial site ID appears to have similar linear separability as the X-ray manufacturer. A contingency table of both variables gives evidence of the same effect as seen in the low probability of independence (**Chi-squared**: $8.6426008^{[4]}$). This intuitively makes sense as clinical sites would most likely use a single manufacturer.



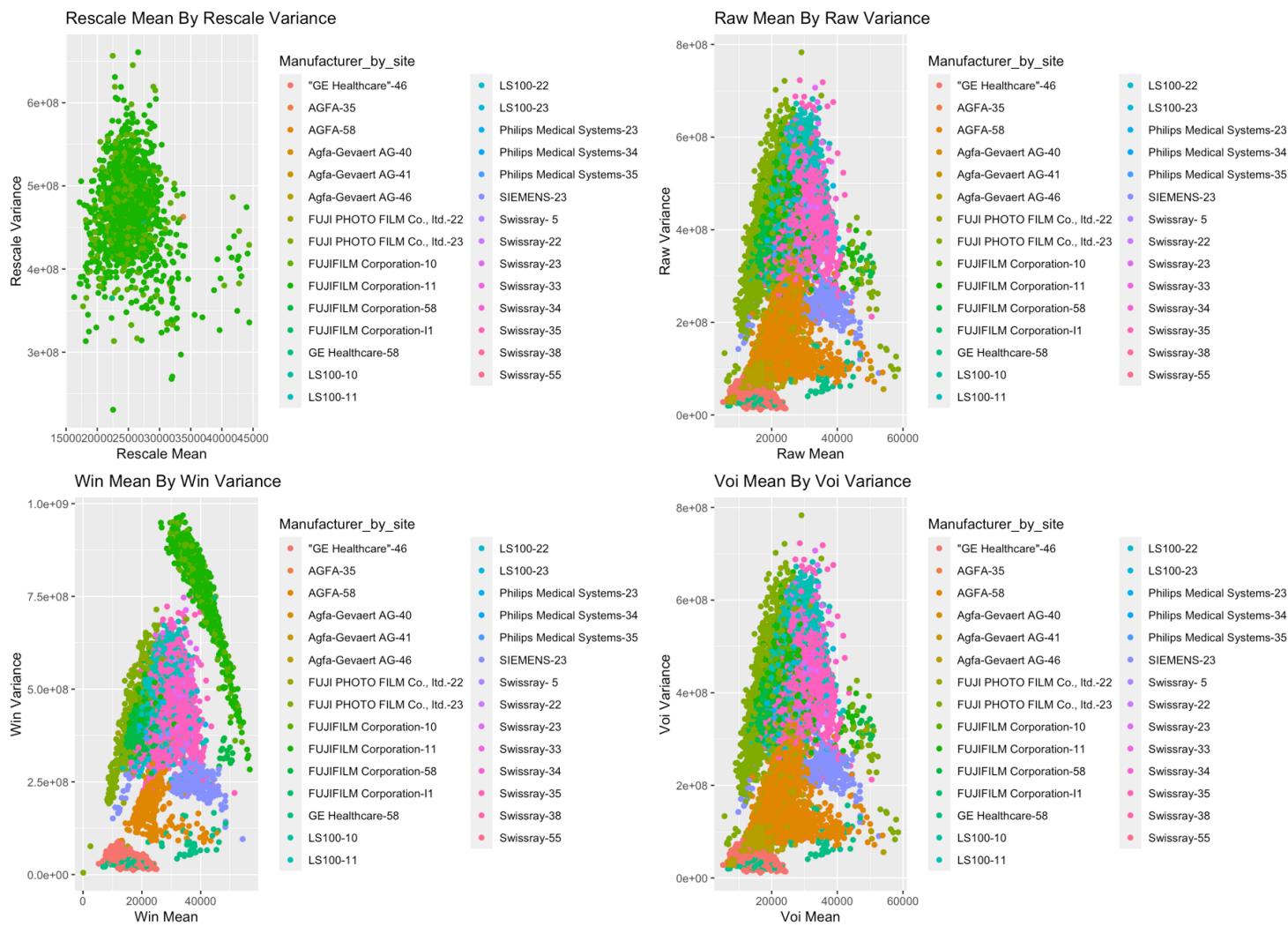
Clinical Trial Site ID

X.0012..0030..Clinical.Trial.Site.ID	N	Raw.mean	Rescale.mean	Win.mean	VOI.mean
46	7819	12829.41	NaN	NA	12829.41
58	6574	21447.82	NaN	NA	21447.82

11		4479	27369.10	25315.52	NA	27369.10
35		3627	29275.97	32722.94	29665.94	29275.97
23		2213	24151.61	NaN	NA	24151.61
10		1247	27023.53	25072.64	NA	27023.53
34		281	28177.25	NaN	28449.63	28177.25
22		92	23584.67	NaN	NA	23584.67
I1		4	26669.22	24333.36	40342.84	26669.22
40		3	14359.63	NaN	NA	14359.63
1		1	35742.97	NaN	40303.06	35742.97
5		1	32161.76	NaN	32161.76	32161.76
33		1	20499.57	NaN	20499.57	20499.57
38		1	22458.14	NaN	22458.14	22458.14
41		1	11450.15	NaN	NA	11450.15
55		1	26960.97	NaN	26960.97	26960.97

Manufacturer By ClinicalTrial Site ID

Due to the degree of overlap between clinical trial site and the image capture manufacturer, the combination of the two variables does not add a significant increase in the ability to fingerprint.



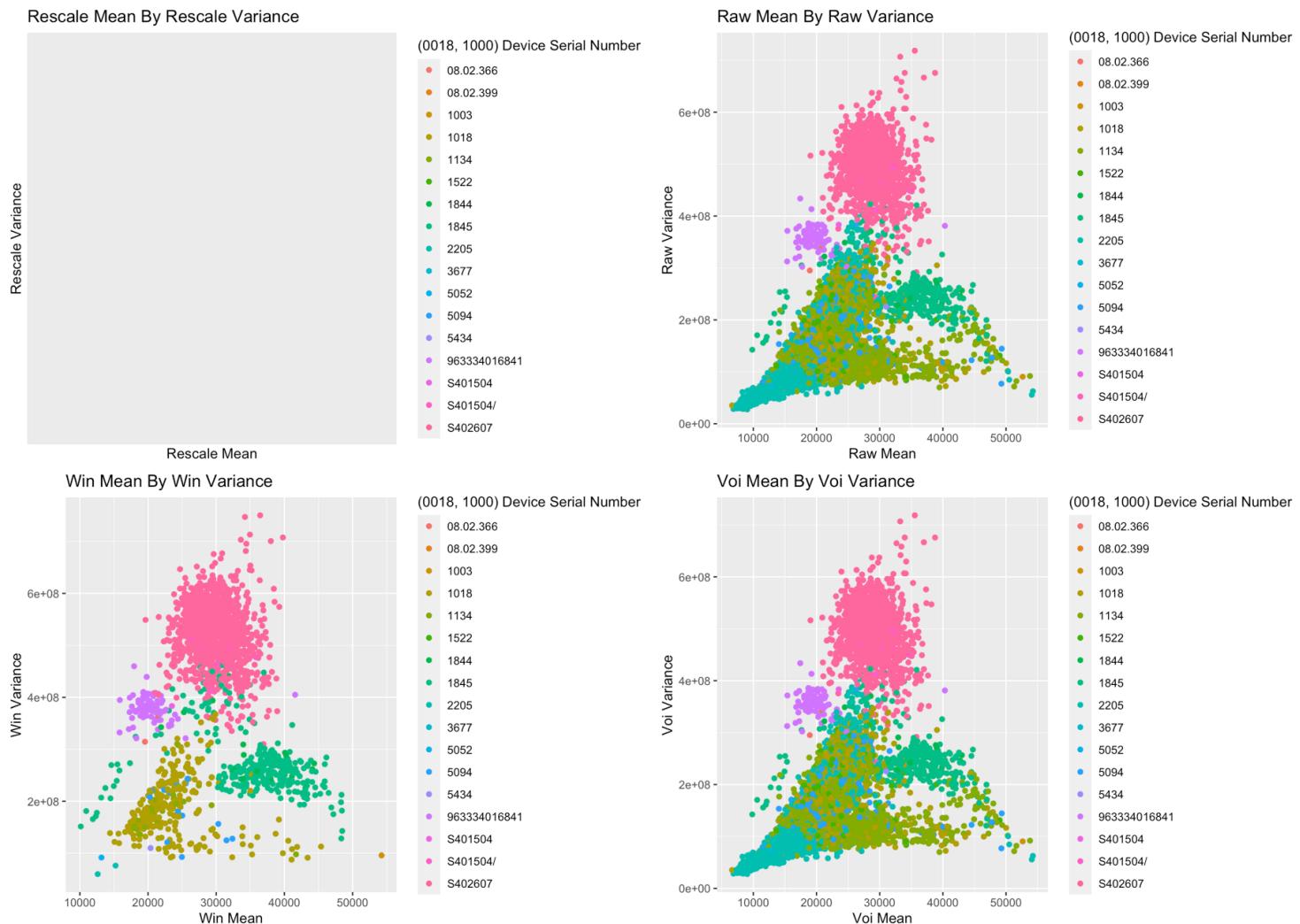
Manufacturer Clinical Trial Site ID

manufacturer_by_site	N	Raw.mean	Rescale.mean	Win.mean	VOI.mean
AGFA-58	4673	22618.96	NaN	NA	22618.96
Agfa-Gevaert AG-46	3933	14184.86	NaN	NA	14184.86
"GE Healthcare"-46	3886	11457.57	NaN	NA	11457.57
Swissray-35	3559	29445.57	NaN	29829.18	29445.57
FUJIFILM Corporation-11	1637	24340.14	25315.52	NA	24340.14
LS100-11	1593	26423.39	NaN	26423.33	26423.39
FUJIFILM Corporation-58	1495	19164.60	NaN	19753.74	19164.60
FUJI PHOTO FILM Co., Ltd.-23	1149	18681.73	NaN	NA	18681.73
FUJIFILM Corporation-10	568	24228.72	25072.64	NA	24228.72

GE Healthcare-58	406	16375.63	NaN	16885.51	16375.63
Swissray-23	381	29220.61	NaN	30113.83	29220.61
SIEMENS-23	325	34772.88	NaN	35556.87	34772.88
LS100-10	294	26109.96	NaN	26109.92	26109.96
Swissray-34	279	28191.33	NaN	28446.32	28191.33
LS100-23	233	26007.39	NaN	26006.95	26007.39
Philips Medical Systems-35	64	19579.88	NaN	20190.15	19579.88
FUJI PHOTO FILM Co., Ltd.-22	53	18965.89	NaN	NA	18965.89
Philips Medical Systems-23	30	20794.62	NaN	21442.50	20794.62
Swissray-22	20	30258.58	NaN	31205.97	30258.58
LS100-22	17	28831.68	NaN	28830.87	28831.68
Agfa-Gevaert AG-40	3	14359.63	NaN	NA	14359.63
FUJIFILM Corporation-I1	3	24333.36	24333.36	41041.19	24333.36
AGFA-35	2	30379.23	32722.94	30546.24	30379.23
Agfa-Gevaert AG-41	1	11450.15	NaN	NA	11450.15
Philips Medical Systems-34	1	15349.05	NaN	15827.43	15349.05
Swissray- 5	1	32161.76	NaN	32161.76	32161.76
Swissray-33	1	20499.57	NaN	20499.57	20499.57
Swissray-38	1	22458.14	NaN	22458.14	22458.14
Swissray-55	1	26960.97	NaN	26960.97	26960.97

Device Serial Number

There are no cases with a specified `Rescale` values and an associated device serial number. The pixel data seems to indicate some degree of separability, although not necessarily linear. Again this is most apparent when a `Window` value is included.



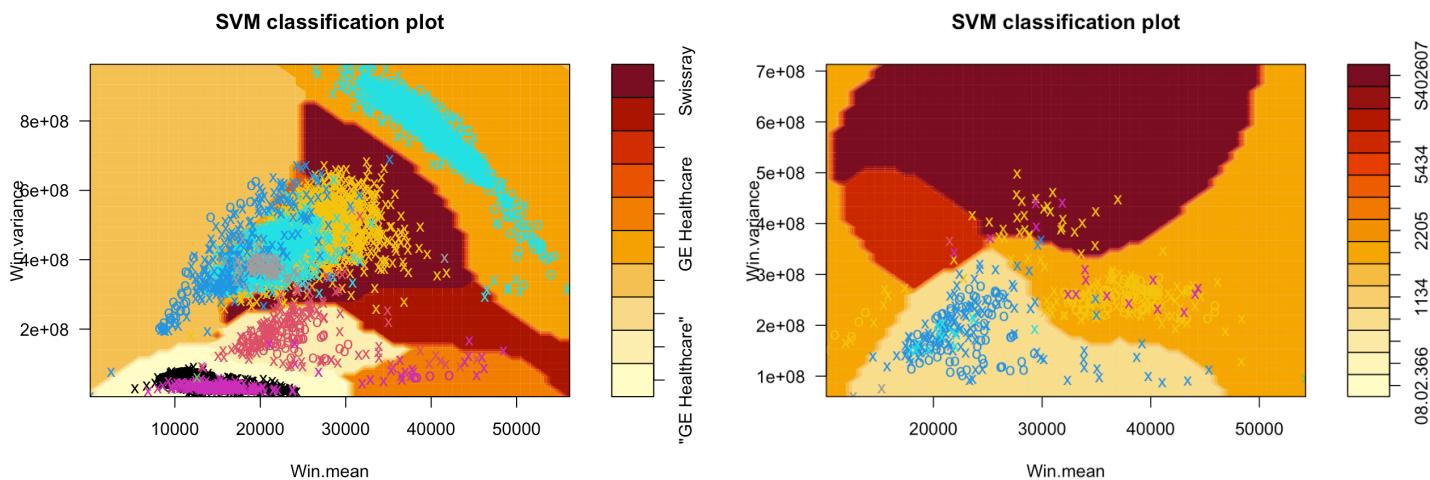
Device Serial Number

X.0018..1000..Device.Serial.Number	N	Raw.mean	Rescale.mean	Win.mean	VOI.mean
2205	3901	14195.06	NaN	NA	14195.06
1134	2582	23133.85	NaN	NA	23133.85
S402607	1385	28968.56	NaN	29860.54	28968.56
1018	1334	22730.45	NaN	NA	22730.45
5094	476	21924.11	NaN	NA	21924.11
1522	322	22657.66	NaN	NA	22657.66
1845	304	34731.47	NaN	35492.94	34731.47
963334016841	93	19919.66	NaN	20540.19	19919.66
3677	37	14132.52	NaN	NA	14132.52

1844	20	34509.58	NaN	35593.88	34509.58
S401504/	19	30648.66	NaN	30901.07	30648.66
5434	14	22469.94	NaN	NA	22469.94
08.02.366	1	18941.77	NaN	19553.36	18941.77
08.02.399	1	20830.50	NaN	21480.94	20830.50
1003	1	52626.12	NaN	54251.81	52626.12
5052	1	12771.17	NaN	13168.88	12771.17
S401504	1	23508.07	NaN	23508.07	23508.07

SVM Analysis

To calculate the performance of separability, an SVM model was ran using the windowed pixel values to predict either manufacturer or device ID class. The data is subset by valid windowing values but the analysis holds for raw pixels. Calculating the accuracy from a test/train split gives some indication on linear separability (or separability using another kernel) as a proxy for fingerprint potential.



Model Accuracies

Manufacturer Accuracy: 0.7665034 **Device Accuracy:** 0.9156051