8 Bins of Building Types
Star Clean Data
N= 5573 & 21 Variables

The FREQ Procedure

bname				
bname	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Education	42	0.75	42	0.75
Facility	123	2.21	165	2.96
General Medical	18	0.32	183	3.28
Hotel	136	2.44	319	5.72
Multifamily	4652	83.47	4971	89.20
Office	545	9.78	5516	98.98
Parking	2	0.04	5518	99.01
Retail	55	0.99	5573	100.00

Input vs. Target
Correlations (Star)
N= 5573 & 21 Variables

The CORR Procedure

13 With Variables:	bin1 bin3 bin6 bin7 bin8 bin9 bin11 bin12 occupy floor_area water_use ws_ngas ws_elec
1 Variables:	site_eui

Simple Statistics							
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
bin1	5573	0.00754	0.08649	42.00000	0	1.00000	D1 Education
bin3	5573	0.02207	0.14693	123.00000	0	1.00000	D3 Facility
bin6	5573	0.00323	0.05675	18.00000	0	1.00000	D6 Medical
bin7	5573	0.02440	0.15431	136.00000	0	1.00000	D7 Hotel
bin8	5573	0.83474	0.37145	4652	0	1.00000	D8 Multifamily
bin9	5573	0.09779	0.29706	545.00000	0	1.00000	D9 Office
bin11	5573	0.0003589	0.01894	2.00000	0	1.00000	D11 Parking
bin12	5573	0.00987	0.09886	55.00000	0	1.00000	D12 Retail
occupy	5573	99.30648	4.18404	553435	0	100.00000	occupy
floor_area	5573	135498	171338	755129971	8000	2376683	floor_area
water_use	5573	14475	113582	80666885	0	3414616	water_use
ws_ngas	5573	2.17795	125.81696	12138	0	9393	ws_ngas
ws_elec	5573	1403817	9666257	7823474906	0	452023582	ws_elec
site_eui	5573	281.28868	11790	1567622	0	869265	site_eui

Pearson Correlation Coefficients, N = 5573	
Prob > |r| under H0: Rho=0	
	site_eui
bin1 D1 Education	-0.00148 0.9118
bin3 D3 Facility	-0.00295 0.8255
bin6 D6 Medical	-0.00010 0.9941
bin7 D7 Hotel	0.00459 0.7317
bin8 D8 Multifamily	0.00423 0.7525
bin9	-0.00530

D9 Office	0.6921
bin11	-0.00034
D11 Parking	0.9797
bin12	-0.00130
D12 Retail	0.9230
occupy	-0.03060
occupy	0.0223
floor_area	-0.00702
floor_area	0.6005
water_use	-0.00139
water_use	0.9174
ws_ngas	0.98747
ws_ngas	<.0001
ws_elec	0.02847
ws_elec	0.0335

Categorical vs. Categorical
Correlations (Star)
N= 5573 & 21 Variables

The CORR Procedure

8 With Variables:	bin1 bin3 bin6 bin7 bin8 bin9 bin11 bin12
8 Variables:	bin1 bin3 bin6 bin7 bin8 bin9 bin11 bin12

Simple Statistics							
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
bin1	5573	0.00754	0.08649	42.00000	0	1.00000	D1 Education
bin3	5573	0.02207	0.14693	123.00000	0	1.00000	D3 Facility
bin6	5573	0.00323	0.05675	18.00000	0	1.00000	D6 Medical
bin7	5573	0.02440	0.15431	136.00000	0	1.00000	D7 Hotel
bin8	5573	0.83474	0.37145	4652	0	1.00000	D8 Multifamily
bin9	5573	0.09779	0.29706	545.00000	0	1.00000	D9 Office
bin11	5573	0.0003589	0.01894	2.00000	0	1.00000	D11 Parking
bin12	5573	0.00987	0.09886	55.00000	0	1.00000	D12 Retail

Pearson Correlation Coefficients, N = 5573 Prob > |r| under H0: Rho=0								
	bin1	bin3	bin6	bin7	bin8	bin9	bin11	bin12
bin1 D1 Education	1.00000	-0.01309 0.3285	-0.00496 0.7112	-0.01378 0.3036	-0.19585 <.0001	-0.02869 0.0322	-0.00165 0.9019	-0.00870 0.5161
bin3 D3 Facility	-0.01309 0.3285	1.00000	-0.00855 0.5233	-0.02376 0.0761	-0.33763 <.0001	-0.04946 0.0002	-0.00285 0.8318	-0.01500 0.2629
bin6 D6 Medical	-0.00496 0.7112	-0.00855 0.5233	1.00000	-0.00900 0.5016	-0.12793 <.0001	-0.01874 0.1618	-0.00108 0.9358	-0.00568 0.6714
bin7 D7 Hotel	-0.01378 0.3036	-0.02376 0.0761	-0.00900 0.5016	1.00000	-0.35545 <.0001	-0.05207 0.0001	-0.00300 0.8230	-0.01579 0.2386
bin8 D8 Multifamily	-0.19585 <.0001	-0.33763 <.0001	-0.12793 <.0001	-0.35545 <.0001	1.00000	-0.73993 <.0001	-0.04258 0.0015	-0.22438 <.0001
bin9 D9 Office	-0.02869 0.0322	-0.04946 0.0002	-0.01874 0.1618	-0.05207 0.0001	-0.73993 <.0001	1.00000	-0.00624 0.6415	-0.03287 0.0141
bin11 D11 Parking	-0.00165 0.9019	-0.00285 0.8318	-0.00108 0.9358	-0.00300 0.8230	-0.04258 0.0015	-0.00624 0.6415	1.00000	-0.00189 0.8877
bin12 D12 Retail	-0.00870 0.5161	-0.01500 0.2629	-0.00568 0.6714	-0.01579 0.2386	-0.22438 <.0001	-0.03287 0.0141	-0.00189 0.8877	1.00000

Categorical vs. Feature
Correlations (Star)
N= 5573 & 21 Variables

The CORR Procedure

6 With Variables:	star occupy floor_area water_use ws_ngas ws_elec
8 Variables:	bin1 bin3 bin6 bin7 bin8 bin9 bin11 bin12

Simple Statistics							
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
star	5573	59.69926	28.96266	332704	1.00000	100.00000	star
occupy	5573	99.30648	4.18404	553435	0	100.00000	occupy
floor_area	5573	135498	171338	755129971	8000	2376683	floor_area
water_use	5573	14475	113582	80666885	0	3414616	water_use
ws_ngas	5573	2.17795	125.81696	12138	0	9393	ws_ngas
ws_elec	5573	1403817	9666257	7823474906	0	452023582	ws_elec
bin1	5573	0.00754	0.08649	42.00000	0	1.00000	D1 Education
bin3	5573	0.02207	0.14693	123.00000	0	1.00000	D3 Facility
bin6	5573	0.00323	0.05675	18.00000	0	1.00000	D6 Medical
bin7	5573	0.02440	0.15431	136.00000	0	1.00000	D7 Hotel
bin8	5573	0.83474	0.37145	4652	0	1.00000	D8 Multifamily
bin9	5573	0.09779	0.29706	545.00000	0	1.00000	D9 Office
bin11	5573	0.0003589	0.01894	2.00000	0	1.00000	D11 Parking
bin12	5573	0.00987	0.09886	55.00000	0	1.00000	D12 Retail

Pearson Correlation Coefficients, N = 5573 Prob > |r| under H0: Rho=0								
	bin1	bin3	bin6	bin7	bin8	bin9	bin11	bin12
star star	-0.03162 0.0182	-0.00413 0.7577	-0.00694 0.6043	-0.11100 <.0001	-0.04472 0.0008	0.12332 <.0001	0.00936 0.4850	0.00674 0.6149
occupy occupy	0.00205 0.8786	-0.00283 0.8327	0.00944 0.4812	-0.10304 <.0001	0.12602 <.0001	-0.10570 <.0001	0.00314 0.8147	0.00136 0.9189
floor_area floor_area	-0.02096 0.1177	-0.02801 0.0365	0.06600 <.0001	0.08549 <.0001	-0.22695 <.0001	0.23783 <.0001	-0.00418 0.7553	0.02749 0.0402
water_use water_use	-0.00954 0.4764	-0.01247 0.3520	0.20411 <.0001	0.02927 0.0289	-0.03492 0.0091	-0.02095 0.1180	-0.00216 0.8717	0.05859 <.0001
ws_ngas ws_ngas	-0.00120 0.9285	-0.00223 0.8681	-0.00040 0.9762	-0.00220 0.8694	0.00660 0.6221	-0.00511 0.7029	-0.00029 0.9827	-0.00138 0.9181
ws_elec ws_elec	-0.00422 0.7526	-0.01141 0.3942	0.05059 0.0002	0.06753 <.0001	-0.11992 <.0001	0.10508 <.0001	0.00048 0.9716	0.02094 0.1181

Feature vs. Feature
Correlations (Star)
N= 5573 & 21 Variables

The CORR Procedure

6 With Variables:	star occupy floor_area water_use ws_ngas ws_elec
5 Variables:	occupy floor_area water_use ws_ngas ws_elec

Simple Statistics							
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
star	5573	59.69926	28.96266	332704	1.00000	100.00000	star
occupy	5573	99.30648	4.18404	553435	0	100.00000	occupy
floor_area	5573	135498	171338	755129971	8000	2376683	floor_area
water_use	5573	14475	113582	80666885	0	3414616	water_use
ws_ngas	5573	2.17795	125.81696	12138	0	9393	ws_ngas
ws_elec	5573	1403817	9666257	7823474906	0	452023582	ws_elec

Pearson Correlation Coefficients, N = 5573 Prob > |r| under H0: Rho=0					
	occupy	floor_area	water_use	ws_ngas	ws_elec
star star	-0.00816 0.5427	0.00903 0.5005	-0.04268 0.0014	-0.02823 0.0351	-0.06823 <.0001
occupy occupy	1.00000	-0.07526 <.0001	-0.06998 <.0001	-0.02962 0.0270	-0.02403 0.0729
floor_area floor_area	-0.07526 <.0001	1.00000	0.10382 <.0001	-0.00645 0.6301	0.25404 <.0001
water_use water_use	-0.06998 <.0001	0.10382 <.0001	1.00000	-0.00127 0.9242	0.03458 0.0098
ws_ngas ws_ngas	-0.02962 0.0270	-0.00645 0.6301	-0.00127 0.9242	1.00000	-0.00121 0.9281
ws_elec ws_elec	-0.02403 0.0729	0.25404 <.0001	0.03458 0.0098	-0.00121 0.9281	1.00000