**Santander Customer Transaction Prediction Solution Using R**

***Report by AMAN ARYA***

**Index**

1. Load Libraries
2. Load data
3. Exploratory Data Analysis
4. Target
5. Features
6. Density plots of train data
7. Compare density plots train and test
8. Boxplots for train data
9. Compare boxplots train and test
10. Correlations
11. Modelling
12. Logistics Regression
13. Glm vs speedglm

Load Libraries

library(tidyverse)

library(data.table)

library(ggthemes)

library(speedglm)

library(MASS)

Load Data

train <- fread("C:/Users/DELL/Desktop/input/train.csv")

test <- fread("C:/Users/DELL/Desktop/input/test.csv")

Exploratory Data Analysis

glimpse(train)

Output

Observations: 200,000

Variables: 202

$ ID\_code *<chr>* "train\_0", "train\_1", "train\_2", "train\_3", "train\_4", "train\_5...

$ target *<int>* 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, ...

$ var\_0 *<dbl>* 8.9255, 11.5006, 8.6093, 11.0604, 9.8369, 11.4763, 11.8091, 13....

$ var\_1 *<dbl>* -6.7863, -4.1473, -2.7457, -2.1518, -1.4834, -2.3182, -0.0832, ...

$ var\_2 *<dbl>* 11.9081, 13.8588, 12.0805, 8.9522, 12.8746, 12.6080, 9.3494, 13...

$ var\_3 *<dbl>* 5.0930, 5.3890, 7.8928, 7.1957, 6.6375, 8.6264, 4.2916, 7.5985,...

$ var\_4 *<dbl>* 11.4607, 12.3622, 10.5825, 12.5846, 12.2772, 10.9621, 11.1355, ...

$ var\_5 *<dbl>* -9.2834, 7.0433, -9.0837, -1.8361, 2.4486, 3.5609, -8.0198, 0.8...

$ var\_6 *<dbl>* 5.1187, 5.6208, 6.9427, 5.8428, 5.9405, 4.5322, 6.1961, 5.6890,...

$ var\_7 *<dbl>* 18.6266, 16.5338, 14.6155, 14.9250, 19.2514, 15.2255, 12.0771, ...

$ var\_8 *<dbl>* -4.9200, 3.1468, -4.9193, -5.8609, 6.2654, 3.5855, -4.3781, 5.0...

$ var\_9 *<dbl>* 5.7470, 8.0851, 5.9525, 8.2450, 7.6784, 5.9790, 7.9232, 7.1971,...

$ var\_10 *<dbl>* 2.9252, -0.4032, -0.3249, 2.3061, -9.4458, 0.8010, -5.1288, 1.4...

$ var\_11 *<dbl>* 3.1821, 8.0585, -11.2648, 2.8102, -12.1419, -0.6192, -7.5271, -...

$ var\_12 *<dbl>* 14.0137, 14.0239, 14.1929, 13.8463, 13.8481, 13.6380, 14.1629, ...

$ var\_13 *<dbl>* 0.5745, 8.4135, 7.3124, 11.9704, 7.8895, 1.2589, 13.3058, 10.96...

$ var\_14 *<dbl>* 8.7989, 5.4345, 7.5244, 6.4569, 7.7894, 8.1939, 7.8412, 6.9190,...

$ var\_15 *<dbl>* 14.5691, 13.7003, 14.6472, 14.8372, 15.0553, 14.9894, 14.3363, ...

$ var\_16 *<dbl>* 5.7487, 13.8275, 7.6782, 10.7430, 8.4871, 12.0763, 7.5951, 9.53...

$ var\_17 *<dbl>* -7.2393, -15.5849, -1.7395, -0.4299, -3.0680, -1.4710, 11.0922,...

$ var\_18 *<dbl>* 4.2840, 7.8000, 4.7011, 15.9426, 6.5263, 6.7341, 21.1976, 5.154...

$ var\_19 *<dbl>* 30.7133, 28.5708, 20.4775, 13.7257, 11.3152, 14.8241, 6.2946, 1...

$ var\_20 *<dbl>* 10.5350, 3.4287, 17.7559, 20.3010, 21.4246, 19.7172, 15.8877, 1...

$ var\_21 *<dbl>* 16.2191, 2.7407, 18.1377, 12.5579, 18.9608, 11.9882, 24.2595, 1...

$ var\_22 *<dbl>* 2.5791, 8.5524, 1.2145, 6.8202, 10.1102, 1.0468, 8.1159, 1.2375...

$ var\_23 *<dbl>* 2.4716, 3.3716, 3.5137, 2.7229, 2.7142, 3.8663, 3.9769, 3.1711,...

$ var\_24 *<dbl>* 14.3831, 6.9779, 5.6777, 12.1354, 14.2080, 4.7252, 7.6851, 9.12...

$ var\_25 *<dbl>* 13.4325, 13.8910, 13.2177, 13.7367, 13.5433, 13.9427, 13.3600, ...

$ var\_26 *<dbl>* -5.1488, -11.7684, -7.9940, 0.8135, 3.1736, -1.2796, -0.5156, 3...

$ var\_27 *<dbl>* -0.4073, -2.5586, -2.9029, -0.9059, -3.3423, -4.3763, 0.0690, -...

$ var\_28 *<dbl>* 4.9306, 5.0464, 5.8463, 5.9070, 5.9015, 5.1494, 5.6452, 5.4501,...

$ var\_29 *<dbl>* 5.9965, 0.5481, 6.1439, 2.8407, 7.9352, 0.4124, 4.6140, 7.9894,...

$ var\_30 *<dbl>* -0.3085, -9.2987, -11.1025, -15.2398, -3.1582, -5.0732, -12.389...

$ var\_31 *<dbl>* 12.9041, 7.8755, 12.4858, 10.4407, 9.4668, 4.9010, 12.0880, 14....

$ var\_32 *<dbl>* -3.8766, 1.2859, -2.2871, -2.5731, -0.0083, 1.5459, -1.5290, -2...

$ var\_33 *<dbl>* 16.8911, 19.3710, 19.0422, 6.1796, 19.3239, 15.6423, 9.2376, 16...

$ var\_34 *<dbl>* 11.1920, 11.3702, 11.0449, 10.6093, 12.4057, 10.7209, 11.1510, ...

$ var\_35 *<dbl>* 10.5785, 0.7399, 4.1087, -5.9158, 0.6329, 15.1886, 6.6352, -0.4...

$ var\_36 *<dbl>* 0.6764, 2.7995, 4.6974, 8.1723, 2.7922, 1.8685, 4.8462, 2.8303,...

$ var\_37 *<dbl>* 7.8871, 5.8434, 6.9346, 2.8521, 5.8184, 7.7223, 7.0202, 7.5772,...

$ var\_38 *<dbl>* 4.6667, 10.8160, 10.8917, 9.1738, 19.3038, 5.5317, 19.9479, 9.0...

$ var\_39 *<dbl>* 3.8743, 3.6783, 0.9003, 0.6665, 1.4450, 2.2308, -6.2271, 1.2659...

$ var\_40 *<dbl>* -5.2387, -11.1147, -13.5174, -3.8294, -5.5963, 2.6553, 4.4616, ...

$ var\_41 *<dbl>* 7.3746, 1.8730, 2.2439, -1.0370, 14.0685, 13.9730, 9.0383, 14.1...

$ var\_42 *<dbl>* 11.5767, 9.8775, 11.5283, 11.7770, 11.9171, 11.5015, 10.6889, 1...

$ var\_43 *<dbl>* 12.0446, 11.7842, 12.0406, 11.2834, 11.5111, 11.6891, 10.9480, ...

$ var\_44 *<dbl>* 11.6418, 1.2444, 4.1006, 8.0485, 6.9087, 14.1062, 15.0431, 13.5...

$ var\_45 *<dbl>* -7.0170, -47.3797, -7.9078, -24.6840, -65.4863, -44.7257, 1.077...

$ var\_46 *<dbl>* 5.9226, 7.3718, 11.1405, 12.7404, 13.8657, 10.6470, 14.2049, 14...

$ var\_47 *<dbl>* -14.2136, 0.1948, -5.7864, -35.1659, 0.0444, -24.6935, -2.0894,...

$ var\_48 *<dbl>* 16.0283, 34.4014, 20.7477, 0.7613, -0.1346, 29.9912, 26.6997, 1...

$ var\_49 *<dbl>* 5.3253, 25.7037, 6.8874, 8.3838, 14.4268, 13.5894, 18.4393, 9.3...

$ var\_50 *<dbl>* 12.9194, 11.8343, 12.9143, 12.6832, 13.3273, 13.2355, 12.9840, ...

$ var\_51 *<dbl>* 29.0460, 13.2256, 19.5856, 9.5503, 10.4857, 5.5513, 23.2511, 11...

$ var\_52 *<dbl>* -0.6940, -4.1083, 0.7268, 1.7895, -1.4367, -7.7762, -3.4444, -8...

$ var\_53 *<dbl>* 5.1736, 6.6885, 6.4059, 5.2091, 5.7555, 5.2399, 5.2735, 6.6503,...

$ var\_54 *<dbl>* -0.7474, -8.0946, 9.3124, 8.0913, -8.5414, 7.0189, 3.7100, -14....

$ var\_55 *<dbl>* 14.8322, 18.5995, 6.2846, 12.3972, 14.1482, 15.2849, 12.4798, 9...

$ var\_56 *<dbl>* 11.2668, 19.3219, 15.6372, 14.4698, 16.9840, 11.8258, 18.4812, ...

$ var\_57 *<dbl>* 5.3822, 7.0118, 5.8200, 6.5850, 6.1812, 5.9587, 6.6104, 6.5482,...

$ var\_58 *<dbl>* 2.0183, 1.9210, 1.1000, 3.3164, 1.9548, 4.7676, 2.7292, 3.8325,...

$ var\_59 *<dbl>* 10.1166, 8.8682, 9.1854, 9.4638, 9.2048, 8.2253, 7.9378, 8.0006...

$ var\_60 *<dbl>* 16.1828, 8.0109, 12.5963, 15.7820, 8.6591, 11.1254, 15.3738, 13...

$ var\_61 *<dbl>* 4.9590, -7.2417, -10.3734, -25.0222, -27.7439, -0.7472, -12.668...

$ var\_62 *<dbl>* 2.0771, 1.7944, 0.8748, 3.4418, -0.4952, -1.6063, 0.6272, 0.158...

$ var\_63 *<dbl>* -0.2154, -1.3147, 5.8042, -4.3923, -1.7839, 6.0702, 2.7414, 4.2...

$ var\_64 *<dbl>* 8.6748, 8.1042, 3.7163, 8.6464, 5.2670, 6.0604, 6.7854, 4.4444,...

$ var\_65 *<dbl>* 9.5319, 1.5365, -1.1016, 6.3072, -4.3205, 5.4693, 0.4546, -8.47...

$ var\_66 *<dbl>* 5.8056, 5.4007, 7.3667, 5.6221, 6.9860, 4.9098, 5.8610, 5.2078,...

$ var\_67 *<dbl>* 22.4321, 7.9344, 9.8565, 23.6143, 1.6184, 14.1160, 13.5793, 8.7...

$ var\_68 *<dbl>* 5.0109, 5.0220, 5.0228, 5.0220, 5.0301, 5.0110, 5.0178, 5.0128,...

$ var\_69 *<dbl>* -4.7010, 2.2302, -5.7828, -3.9989, -3.2431, -5.6684, -5.7921, -...

$ var\_70 *<dbl>* 21.6374, 40.5632, 2.3612, 4.0462, 40.1236, 35.2664, 21.1640, 27...

$ var\_71 *<dbl>* 0.5663, 0.5134, 0.8520, 0.2500, 0.7737, 0.6780, 0.6386, 1.0085,...

$ var\_72 *<dbl>* 5.1999, 3.1701, 6.3577, 1.2516, -0.7264, 8.1653, 3.7051, 3.7054...

$ var\_73 *<dbl>* 8.8600, 20.1068, 12.1719, 24.4187, 4.5886, 9.9031, 19.3404, 16....

$ var\_74 *<dbl>* 43.1127, 7.7841, 19.7312, 4.5290, -4.5346, 10.5392, 29.6714, 20...

$ var\_75 *<dbl>* 18.3816, 7.0529, 19.4465, 15.4235, 23.3521, 14.4814, 13.7166, 1...

$ var\_76 *<dbl>* -2.3440, 3.2709, 4.5048, 11.6875, 1.0273, 12.7784, 12.5176, 0.7...

$ var\_77 *<dbl>* 23.4104, 23.4822, 23.2378, 23.6273, 19.1600, 17.4977, 16.9189, ...

$ var\_78 *<dbl>* 6.5199, 5.5075, 6.3191, 4.0806, 7.1734, 3.3793, 5.2232, 6.7311,...

$ var\_79 *<dbl>* 12.1983, 13.7814, 12.8046, 15.2733, 14.3937, 16.0082, 16.9441, ...

$ var\_80 *<dbl>* 13.6468, 2.5462, 7.4729, 0.7839, 2.9598, -5.4905, -2.7142, -3.8...

$ var\_81 *<dbl>* 13.8372, 18.1782, 15.7811, 10.5404, 13.3317, 15.3766, 12.3111, ...

$ var\_82 *<dbl>* 1.3675, 0.3683, 13.3529, 1.6212, -9.2587, 1.8135, -2.5448, -10....

$ var\_83 *<dbl>* 2.9423, -4.8210, 10.1852, -5.2896, -6.7075, -1.9153, 10.7269, 6...

$ var\_84 *<dbl>* -4.5213, -5.4850, 5.4604, 1.6027, 7.8984, -8.3058, -1.3374, -0....

$ var\_85 *<dbl>* 21.4669, 13.7867, 19.0773, 17.9762, 14.5265, 20.1690, 21.5486, ...

$ var\_86 *<dbl>* 9.3225, -13.5901, -4.4577, -2.3174, 7.0799, 2.6433, 12.6511, 6....

$ var\_87 *<dbl>* 16.4597, 11.0993, 9.5413, 15.6298, 20.1670, 19.9457, 6.3112, 6....

$ var\_88 *<dbl>* 7.9984, 7.9022, 11.9052, 4.5474, 8.0053, 9.4972, 5.9346, 6.3539...

$ var\_89 *<dbl>* -1.7069, 12.2301, 2.1447, 7.5509, 3.7954, -3.5982, 10.4663, 6.7...

$ var\_90 *<dbl>* -21.4494, 0.4768, -22.4038, -7.5866, -39.7997, -1.1717, -9.2271...

$ var\_91 *<dbl>* 6.7806, 6.8852, 7.0883, 7.0364, 7.0065, 6.9204, 7.0140, 6.8082,...

$ var\_92 *<dbl>* 11.0924, 8.0905, 14.1613, 14.4027, 9.3627, 8.8093, 11.1954, 9.1...

$ var\_93 *<dbl>* 9.9913, 10.9631, 10.5080, 10.7795, 10.4316, 10.5375, 10.4502, 1...

$ var\_94 *<dbl>* 14.8421, 11.7569, 14.2621, 7.2887, 14.0553, 10.2029, 11.1829, 1...

$ var\_95 *<dbl>* 0.1812, -1.2722, 0.2647, -1.0930, 0.0213, 1.1735, 0.6137, -0.66...

$ var\_96 *<dbl>* 8.9642, 24.7876, 20.4031, 11.3596, 14.7246, 14.0648, 23.7290, 1...

$ var\_97 *<dbl>* 16.2572, 26.6881, 17.0360, 18.1486, 35.2988, 17.9100, 32.8788, ...

$ var\_98 *<dbl>* 2.1743, 1.8944, 1.6981, 2.8344, 1.6844, 0.9217, 2.4715, 2.0848,...

$ var\_99 *<dbl>* -3.4132, 0.6939, -0.0269, 1.9480, 0.6715, 0.0886, 0.0483, 2.244...

$ var\_100 *<dbl>* 9.4763, -13.6950, -0.3939, -19.8592, -22.9264, -8.3794, -22.215...

$ var\_101 *<dbl>* 13.3102, 8.4068, 12.6317, 22.5316, 12.3562, 19.8236, 17.0682, 2...

$ var\_102 *<dbl>* 26.5376, 35.4734, 14.8863, 18.6129, 17.3410, 12.2780, 20.5977, ...

$ var\_103 *<dbl>* 1.4403, 1.7093, 1.3854, 1.3512, 1.6940, 1.6012, 1.8554, 1.4998,...

$ var\_104 *<dbl>* 14.7100, 15.1866, 15.0284, 9.3291, 7.1179, 15.4090, 12.1542, 14...

$ var\_105 *<dbl>* 6.0454, 2.6227, 3.9995, 4.2835, 5.1934, 4.2972, 4.3055, 4.1641,...

$ var\_106 *<dbl>* 9.5426, 7.3412, 5.3683, 10.3907, 8.8230, 10.5131, 8.3469, 8.513...

$ var\_107 *<dbl>* 17.1554, 32.0888, 8.6273, 7.0874, 10.6617, 16.4164, 23.3089, 16...

$ var\_108 *<dbl>* 14.1104, 13.9550, 14.1963, 14.3256, 14.0837, 14.1434, 14.4272, ...

$ var\_109 *<dbl>* 24.3627, 13.0858, 20.3882, 14.4135, 28.2749, 23.9731, 13.5718, ...

$ var\_110 *<dbl>* 2.0323, 6.6203, 3.2304, 4.2827, -0.1937, 7.0040, 11.6967, 2.518...

$ var\_111 *<dbl>* 6.7602, 7.1051, 5.7033, 6.9750, 5.9654, 7.1879, 6.6323, 7.1014,...

$ var\_112 *<dbl>* 3.9141, 5.3523, 4.5255, 1.6480, 1.0719, 3.9414, 2.7194, 3.3046,...

$ var\_113 *<dbl>* -0.4851, 8.5426, 2.1929, 11.6896, 7.9923, 9.9727, 4.6110, 5.749...

$ var\_114 *<dbl>* 2.5240, 3.6159, 3.1290, 2.5762, 2.9138, 3.6737, 2.3090, 2.8527,...

$ var\_115 *<dbl>* 1.5093, 4.1569, 2.9044, -2.5459, -3.6135, 0.8970, 0.3681, 4.195...

$ var\_116 *<dbl>* 2.5516, 3.0454, 1.1696, 5.3446, 1.4684, 2.2988, 2.9144, 4.9895,...

$ var\_117 *<dbl>* 15.5752, 7.8522, 28.7632, 38.1015, 25.6795, -18.1194, 19.4180, ...

$ var\_118 *<dbl>* -13.4221, -11.5100, -17.2738, 3.5732, 13.8224, 8.6835, 1.3141, ...

$ var\_119 *<dbl>* 7.2739, 7.5109, 2.1056, 5.0988, 4.7478, -2.8671, -0.9810, 0.079...

$ var\_120 *<dbl>* 16.0094, 31.5899, 21.1613, 30.5644, 41.1037, 11.2701, 23.3644, ...

$ var\_121 *<dbl>* 9.7268, 9.5018, 8.9573, 11.3025, 12.7140, 9.0465, 10.5135, 12.7...

$ var\_122 *<dbl>* 0.8897, 8.2736, 2.7768, 3.9618, 5.2964, -1.5511, 5.7981, 1.6484...

$ var\_123 *<dbl>* 0.7754, 10.1633, -2.1746, -8.2464, 9.7289, 9.9305, -5.0096, -1....

$ var\_124 *<dbl>* 4.2218, 0.1225, 3.6932, 2.7038, 3.9370, 4.3756, 0.5866, 3.0002,...

$ var\_125 *<dbl>* 12.0039, 12.5942, 12.4653, 12.3441, 12.1316, 12.3697, 11.8274, ...

$ var\_126 *<dbl>* 13.8571, 14.5697, 14.1978, 12.5431, 12.5815, 13.0031, 12.1880, ...

$ var\_127 *<dbl>* -0.7338, 2.4354, -2.5511, -1.3683, 7.0642, 5.0993, 1.7719, 0.76...

$ var\_128 *<dbl>* -1.9245, 0.8194, -0.9479, 3.5974, 5.6518, -5.8702, 2.7085, 1.74...

$ var\_129 *<dbl>* 15.4462, 16.5346, 17.1092, 13.9761, 10.9346, 15.0534, 21.7473, ...

$ var\_130 *<dbl>* 12.8287, 12.4205, 11.5419, 14.3003, 11.4266, 12.7212, 11.9410, ...

$ var\_131 *<dbl>* 0.3587, -0.1780, 0.0975, 1.0486, 0.9442, 0.5762, 0.5405, 0.4158...

$ var\_132 *<dbl>* 9.6508, 5.7582, 8.8186, 8.9500, 7.7532, 7.0965, 8.9913, 9.8819,...

$ var\_133 *<dbl>* 6.5674, 7.0513, 6.6231, 7.1954, 6.6173, 6.6143, 7.4506, 7.0754,...

$ var\_134 *<dbl>* 5.1726, 1.9568, 3.9358, -1.1984, -6.8304, -8.8256, -13.2871, -1...

$ var\_135 *<dbl>* 3.1345, -8.9921, -11.7218, 1.9586, 6.4730, -0.9844, -5.7890, -1...

$ var\_136 *<dbl>* 29.4547, 9.7797, 24.5437, 27.5609, 17.1728, 26.1603, 23.8387, 2...

$ var\_137 *<dbl>* 31.4045, 18.1577, 15.5827, 24.6065, 25.8128, 11.9963, 7.8704, 2...

$ var\_138 *<dbl>* 2.8279, -1.9721, 3.8212, -2.8233, 2.6791, -5.9018, 5.2588, -0.0...

$ var\_139 *<dbl>* 15.6599, 16.1622, 8.6674, 8.9821, 13.9547, 13.3065, 7.4428, 17....

$ var\_140 *<dbl>* 8.3307, 3.6937, 7.3834, 3.8873, 6.6289, 4.4419, 2.7612, 1.0003,...

$ var\_141 *<dbl>* -5.6011, 6.6803, -2.4438, 15.9638, -4.3965, -0.0496, 10.3221, -...

$ var\_142 *<dbl>* 19.0614, -0.3243, 10.2158, 10.0142, 11.7159, 18.7239, -1.7192, ...

$ var\_143 *<dbl>* 11.2663, 12.2806, 7.4844, 7.8388, 16.1080, 14.1586, 16.4603, 16...

$ var\_144 *<dbl>* 8.6989, 8.6086, 9.1104, 9.9718, 7.6874, 9.6607, 8.9949, 9.7450,...

$ var\_145 *<dbl>* 8.3694, 11.0738, 4.3649, 2.9253, 9.1570, 2.6134, 11.7397, 12.11...

$ var\_146 *<dbl>* 11.5659, 8.9231, 11.4934, 10.4994, 11.5670, 12.9650, 11.4148, 9...

$ var\_147 *<dbl>* -16.4727, 11.7700, 1.7624, 4.1622, -12.7047, 4.7280, 6.8520, -1...

$ var\_148 *<dbl>* 4.0288, 4.2578, 4.0714, 3.7613, 3.7574, 3.9799, 3.8912, 4.1435,...

$ var\_149 *<dbl>* 17.9244, -4.4223, -1.2681, 2.3701, 9.9110, 1.4462, -1.3446, 26....

$ var\_150 *<dbl>* 18.5177, 20.6294, 14.3330, 18.0984, 20.1461, 19.5171, 11.6169, ...

$ var\_151 *<dbl>* 10.7800, 14.8743, 8.0088, 17.1765, 1.2995, 7.5903, 10.4602, 14....

$ var\_152 *<dbl>* 9.0056, 9.4317, 4.4015, 7.6508, 5.8493, 5.7223, 4.3166, 7.2698,...

$ var\_153 *<dbl>* 16.6964, 16.7242, 14.1479, 18.2452, 19.8234, 15.2331, 13.1753, ...

$ var\_154 *<dbl>* 10.4838, -0.5687, -5.1747, 17.0336, 4.7022, 15.4401, 9.9217, 13...

$ var\_155 *<dbl>* 1.6573, 0.1898, 0.5778, -10.9370, 10.6101, -2.0738, -2.5725, -1...

$ var\_156 *<dbl>* 12.1749, 12.2419, 14.5362, 12.0500, 13.0021, 12.4068, 13.7241, ...

$ var\_157 *<dbl>* -13.1324, -9.6953, -1.7624, -1.2155, -12.6068, -16.8151, -10.03...

$ var\_158 *<dbl>* 17.6054, 22.3949, 33.8820, 19.9750, 27.0846, 24.2054, 21.9814, ...

$ var\_159 *<dbl>* 11.5423, 10.6261, 11.6041, 12.3892, 8.0913, 4.9455, 11.1648, 19...

$ var\_160 *<dbl>* 15.4576, 29.4846, 13.2070, 31.8833, 33.5107, 16.5552, 39.9599, ...

$ var\_161 *<dbl>* 5.3133, 5.8683, 5.8442, 5.9684, 5.6953, 5.3739, 5.5552, 5.4098,...

$ var\_162 *<dbl>* 3.6159, 3.8208, 4.7086, 7.2084, 5.4663, 6.4487, 3.3459, 5.1402,...

$ var\_163 *<dbl>* 5.0384, 15.8348, 5.7141, 3.8899, 18.2201, 11.5631, 9.2661, 10.7...

$ var\_164 *<dbl>* 6.6760, -5.0121, -1.0410, -11.0882, 6.5769, 1.3847, 6.1213, -8....

$ var\_165 *<dbl>* 12.6644, 15.1345, 20.5092, 17.2502, 21.2607, 14.9638, 23.7558, ...

$ var\_166 *<dbl>* 2.7004, 3.2003, 3.2790, 2.5881, 3.2304, 2.8455, 3.0298, 2.6085,...

$ var\_167 *<dbl>* -0.6975, 9.3192, -5.5952, -2.7018, -1.7759, -9.0953, 5.9109, -1...

$ var\_168 *<dbl>* 9.5981, 3.8821, 7.3176, 0.5641, 3.1283, 3.8278, 8.1035, 8.7362,...

$ var\_169 *<dbl>* 5.4879, 5.7999, 5.7690, 5.3430, 5.5518, 5.9714, 6.1887, 5.2273,...

$ var\_170 *<dbl>* -4.7645, 5.5378, -7.0927, -7.1541, 1.4493, -6.1449, 0.2619, 8.9...

$ var\_171 *<dbl>* -8.4254, 5.0988, -3.9116, -6.1920, -2.6627, -2.0285, -1.1405, -...

$ var\_172 *<dbl>* 20.8773, 22.0330, 7.2569, 18.2366, 19.8056, 18.4106, 25.1675, 6...

$ var\_173 *<dbl>* 3.1531, 5.5134, -5.8234, 11.7134, 2.3705, 1.4457, 2.6965, 0.087...

$ var\_174 *<dbl>* 18.5618, 30.2645, 25.6820, 14.7483, 18.4685, 21.8853, 17.0152, ...

$ var\_175 *<dbl>* 7.7423, 10.4968, 10.9202, 8.1013, 16.3309, 9.2654, 12.7942, 13....

$ var\_176 *<dbl>* -10.1245, -7.2352, -0.3104, 11.8771, -3.3456, -6.5247, -3.0403,...

$ var\_177 *<dbl>* 13.7241, 16.5721, 8.8438, 13.9552, 13.5261, 10.7687, 8.1735, 17...

$ var\_178 *<dbl>* -3.5189, -7.3477, -9.7009, -10.4701, 1.7189, -7.6283, 4.5637, -...

$ var\_179 *<dbl>* 1.7202, 11.0752, 2.4013, 5.6961, 5.1743, 1.0208, 3.8973, 0.0592...

$ var\_180 *<dbl>* -8.4051, -5.5937, -4.2935, -3.7546, -7.6938, 7.1968, -8.1416, 5...

$ var\_181 *<dbl>* 9.0164, 9.4878, 9.3908, 8.4117, 9.7685, 11.1227, 10.0570, 10.54...

$ var\_182 *<dbl>* 3.0657, -14.9100, -13.2648, 1.8986, 4.8910, 2.2257, 15.7862, 6....

$ var\_183 *<dbl>* 14.3691, 9.4245, 3.1545, 7.2601, 12.2198, 6.4056, 3.3593, 6.972...

$ var\_184 *<dbl>* 25.8398, 22.5441, 23.0866, -0.4639, 11.8503, 21.0550, 11.9140, ...

$ var\_185 *<dbl>* 5.8764, -4.8622, -5.3000, -0.0498, -7.8931, -13.6509, -4.2870, ...

$ var\_186 *<dbl>* 11.8411, 7.6543, 5.3745, 7.9336, 6.4209, 4.7691, 7.5015, 8.4947...

$ var\_187 *<dbl>* -19.7159, -15.9319, -6.2660, -12.8279, 5.9270, -8.9114, -29.976...

$ var\_188 *<dbl>* 17.5743, 13.3175, 10.1934, 12.4124, 16.0201, 15.1007, 17.2867, ...

$ var\_189 *<dbl>* 0.5857, -0.3566, -0.8417, 1.8489, -0.2829, 2.4286, 1.8539, 1.97...

$ var\_190 *<dbl>* 4.4354, 7.6421, 2.9057, 4.4666, -1.4905, -6.3068, 8.7830, 13.17...

$ var\_191 *<dbl>* 3.9642, 7.7214, 9.7905, 4.7433, 9.5214, 6.6025, 6.4521, 6.5491,...

$ var\_192 *<dbl>* 3.1364, 2.5837, 1.6704, 0.7178, -0.1508, 5.2912, 3.5325, 3.9906...

$ var\_193 *<dbl>* 1.6910, 10.9516, 1.6858, 1.4214, 9.1942, 0.4403, 0.1777, 5.8061...

$ var\_194 *<dbl>* 18.5227, 15.4305, 21.6042, 23.0347, 13.2876, 14.9452, 18.3314, ...

$ var\_195 *<dbl>* -2.3978, 2.0339, 3.1417, -1.2706, -1.5121, 1.0314, 0.5845, -0.3...

$ var\_196 *<dbl>* 7.8784, 8.1267, -6.5213, -2.9275, 3.9267, -3.6241, 9.1104, 4.21...

$ var\_197 *<dbl>* 8.5635, 8.7889, 8.2675, 10.2922, 9.5031, 9.7670, 9.1143, 9.4237...

$ var\_198 *<dbl>* 12.7803, 18.3560, 14.7222, 17.9697, 17.9974, 12.5809, 10.8869, ...

$ var\_199 *<dbl>* -1.0914, 1.9518, 0.3965, -8.9996, -8.8104, -4.7602, -3.2097, 3....

Code

glimpse(test)

Output

Observations: 200,000

Variables: 201

$ ID\_code *<chr>* "test\_0", "test\_1", "test\_2", "test\_3", "test\_4", "test\_5", "te...

$ var\_0 *<dbl>* 11.0656, 8.5304, 5.4827, 8.5374, 11.7058, 5.9862, 8.4624, 17.30...

$ var\_1 *<dbl>* 7.7798, 1.2543, -10.3581, -1.3222, -0.1327, -2.2913, -6.1065, -...

$ var\_2 *<dbl>* 12.9536, 11.3047, 10.1407, 12.0220, 14.1295, 8.6058, 7.3603, 13...

$ var\_3 *<dbl>* 9.4292, 5.1858, 7.0479, 6.5749, 7.7506, 7.0685, 8.2627, 8.3998,...

$ var\_4 *<dbl>* 11.4327, 9.1974, 10.2628, 8.8458, 9.1035, 14.2465, 12.0104, 11....

$ var\_5 *<dbl>* -2.3805, -4.0117, 9.8052, 3.1744, -8.5848, -8.6761, -7.2073, 9....

$ var\_6 *<dbl>* 5.8493, 6.0196, 4.8950, 4.9397, 6.8595, 4.2467, 4.1670, 5.9596,...

$ var\_7 *<dbl>* 18.2675, 18.6316, 20.2537, 20.5660, 10.6048, 14.7632, 13.0809, ...

$ var\_8 *<dbl>* 2.1337, -4.4131, 1.5233, 3.3755, 2.9890, 1.8790, -4.3004, -4.80...

$ var\_9 *<dbl>* 8.8100, 5.9739, 8.3442, 7.4578, 7.1437, 7.2842, 6.3181, 7.4643,...

$ var\_10 *<dbl>* -2.0248, -1.3809, -4.7057, 0.0095, 5.1025, -4.9194, 3.3959, 4.0...

$ var\_11 *<dbl>* -4.3554, -0.3310, -3.0422, -5.0659, -3.2827, -9.1869, -2.0205, ...

$ var\_12 *<dbl>* 13.9696, 14.1129, 13.6751, 14.0526, 14.1013, 14.0581, 13.7682, ...

$ var\_13 *<dbl>* 0.3458, 2.5667, 3.8183, 13.5010, 8.9672, 11.4403, 12.0534, 9.22...

$ var\_14 *<dbl>* 7.5408, 5.4988, 10.8535, 8.7660, 4.7276, 4.2950, 7.2186, 10.320...

$ var\_15 *<dbl>* 14.5001, 14.1853, 14.2126, 14.7352, 14.5811, 13.8808, 14.6552, ...

$ var\_16 *<dbl>* 7.7028, 7.0196, 9.8837, 10.0383, 11.8615, 10.9588, 10.7060, 10....

$ var\_17 *<dbl>* -19.0919, 4.6564, 2.6541, -15.3508, 3.1480, -7.0124, -8.0046, -...

$ var\_18 *<dbl>* 15.5806, 29.1609, 21.2181, 2.1273, 18.0126, 15.1829, 16.5411, 2...

$ var\_19 *<dbl>* 16.1763, 0.0910, 20.8163, 21.4797, 13.8006, 18.2710, -1.9177, 1...

$ var\_20 *<dbl>* 3.7088, 12.1469, 12.4666, 14.5372, 1.6026, 9.7086, 23.4575, 26....

$ var\_21 *<dbl>* 18.8064, 3.1389, 12.3696, 12.5527, 16.3059, 22.9103, 20.4773, 1...

$ var\_22 *<dbl>* 1.5899, 5.2578, 4.7473, 2.9707, 6.7954, 6.5080, 3.9038, 5.8071,...

$ var\_23 *<dbl>* 3.0654, 2.4228, 2.7936, 4.2398, 3.6015, 2.3607, 2.4832, 4.0316,...

$ var\_24 *<dbl>* 6.4509, 16.2064, 5.2189, 13.7796, 13.6569, 1.7899, 18.0843, 19....

$ var\_25 *<dbl>* 14.1192, 13.5023, 13.5670, 14.1408, 13.8807, 13.9555, 13.2727, ...

$ var\_26 *<dbl>* -9.4902, -5.2341, -15.4246, 1.0061, 8.6228, -7.2154, -15.1986, ...

$ var\_27 *<dbl>* -2.1917, -3.6648, -0.1655, -1.3479, -2.2654, -2.4084, -0.4427, ...

$ var\_28 *<dbl>* 5.7107, 5.7080, 7.2633, 5.2570, 5.2255, 5.1616, 5.1813, 5.1571,...

$ var\_29 *<dbl>* 3.7864, 2.9965, 3.4310, 6.5911, 7.0165, 6.9291, 6.4509, 4.2024,...

$ var\_30 *<dbl>* -1.7981, -10.4720, -9.1508, 6.2161, -15.6961, 0.1141, 1.1644, -...

$ var\_31 *<dbl>* 9.2645, 11.4938, 9.7320, 9.5540, 10.6239, 9.7667, 8.1706, 11.02...

$ var\_32 *<dbl>* 2.0657, -0.9660, 3.1062, 2.3628, -4.7674, 1.2339, -0.3277, -4.0...

$ var\_33 *<dbl>* 12.7753, 15.3445, 22.3076, 10.2124, 17.5447, 17.7117, 18.6378, ...

$ var\_34 *<dbl>* 11.3334, 10.6361, 11.9593, 10.8047, 11.8668, 12.3433, 10.6855, ...

$ var\_35 *<dbl>* 8.1462, 0.8966, 9.9255, -2.5588, 3.0154, -0.1787, 11.9737, 12.2...

$ var\_36 *<dbl>* -0.0610, 6.7428, 4.0702, 6.0720, 4.2546, -1.2858, 5.8620, 2.611...

$ var\_37 *<dbl>* 3.5331, 2.3421, 4.9934, 3.2613, 6.7601, 5.5220, 1.6195, 3.7225,...

$ var\_38 *<dbl>* 9.7804, 12.8678, 8.0667, 16.5632, 5.9613, 13.0498, 16.8667, 11....

$ var\_39 *<dbl>* 8.7625, -1.5536, 0.8804, 8.8336, 0.3695, -5.3328, -6.1878, 7.48...

$ var\_40 *<dbl>* -15.6305, 10.0309, -19.0841, -4.8327, -14.4364, -10.7780, -11.4...

$ var\_41 *<dbl>* 18.8766, 3.1337, 5.2272, 0.9554, 5.1392, 0.0721, 12.0445, 12.96...

$ var\_42 *<dbl>* 11.2864, 10.5742, 9.5977, 12.3754, 11.6336, 10.3590, 10.6562, 9...

$ var\_43 *<dbl>* 11.8362, 11.7664, 12.1801, 11.4241, 12.0338, 11.6175, 11.4393, ...

$ var\_44 *<dbl>* 13.3680, 2.1782, 8.3565, 6.6917, 18.9670, 0.4534, 13.3974, 9.59...

$ var\_45 *<dbl>* -31.9891, -41.1924, 15.1170, -12.9761, 12.0144, -6.7184, 12.322...

$ var\_46 *<dbl>* 12.1776, 13.5322, 10.0921, 13.7343, 16.2096, 12.1589, 16.0258, ...

$ var\_47 *<dbl>* 8.7714, -17.3834, -20.8504, 5.0150, -2.1966, -13.8476, -13.4051...

$ var\_48 *<dbl>* 17.2011, 6.3806, 8.6758, 31.3923, 1.1174, 11.0962, 23.4127, 29....

$ var\_49 *<dbl>* 16.8508, 12.5589, 8.1292, 5.8555, 13.4532, 12.6084, 20.8895, 15...

$ var\_50 *<dbl>* 13.0534, 11.6887, 11.8932, 12.6082, 12.7925, 12.3914, 11.5413, ...

$ var\_51 *<dbl>* 14.4069, 25.3930, 10.6869, 1.4182, 4.3775, 1.2393, 17.2430, 18....

$ var\_52 *<dbl>* -4.8525, 1.5776, -0.6434, -4.1185, -0.1543, 1.5472, -3.2139, -0...

$ var\_53 *<dbl>* 7.3213, 6.8481, 5.6510, 6.2536, 5.6794, 5.2122, 5.9612, 6.4081,...

$ var\_54 *<dbl>* -0.5259, 8.7348, 9.3742, 1.4257, 0.8210, -6.7651, 3.8564, -4.09...

$ var\_55 *<dbl>* 16.6365, 16.4239, 25.8831, 13.5426, 19.1358, 22.7272, 9.5905, 9...

$ var\_56 *<dbl>* 19.3036, 21.7056, 19.8701, 15.4090, 12.6589, 16.7353, 17.7064, ...

$ var\_57 *<dbl>* 6.4129, 6.9345, 5.4834, 6.8761, 6.4394, 5.3347, 5.9169, 5.2576,...

$ var\_58 *<dbl>* -5.3948, 1.6678, -4.0304, 1.7476, 4.3425, 10.3657, 4.3098, -0.4...

$ var\_59 *<dbl>* 9.3269, 9.5249, 8.5160, 10.0413, 8.7003, 8.2551, 7.2078, 9.5163...

$ var\_60 *<dbl>* 11.9314, 5.3383, 8.9776, 15.2857, 12.0586, 11.4422, 16.1110, 9....

$ var\_61 *<dbl>* -3.5750, -18.7083, -5.6619, -4.1378, -10.4753, -8.7121, 6.1001,...

$ var\_62 *<dbl>* -0.7706, 1.3382, 2.8117, 0.7928, -0.0337, 3.8256, 5.7367, 1.349...

$ var\_63 *<dbl>* 0.8705, -1.7401, 2.5996, 2.5301, 5.6603, 1.1723, 1.7652, -1.224...

$ var\_64 *<dbl>* 6.9282, 5.8398, 9.0986, 8.1458, 6.2529, 6.9224, 4.4416, 7.8745,...

$ var\_65 *<dbl>* 2.8914, 3.1051, 7.1167, 2.5738, 1.5238, -1.5636, 1.8125, -1.439...

$ var\_66 *<dbl>* 5.9744, 4.4307, 4.9466, 5.9876, 4.5356, 4.7235, 7.2286, 3.6731,...

$ var\_67 *<dbl>* 17.4851, 16.0005, 13.8268, 13.0758, 20.1344, 23.2156, 9.5884, 1...

$ var\_68 *<dbl>* 5.0125, 5.0306, 5.0093, 5.0087, 5.0267, 5.0215, 5.0167, 5.0351,...

$ var\_69 *<dbl>* -1.4230, -7.3365, 4.7782, -9.7824, -1.8628, -5.2671, 1.5204, -4...

$ var\_70 *<dbl>* 33.3401, 12.2806, 19.2081, 8.9289, 39.8219, 10.7981, 37.4611, 2...

$ var\_71 *<dbl>* 0.8018, 0.6992, 0.4340, 0.4205, 1.0498, 0.8257, 0.4291, 0.4480,...

$ var\_72 *<dbl>* -4.7906, -0.7772, 0.8459, -2.5463, -0.9113, -5.5004, -6.8746, 4...

$ var\_73 *<dbl>* 30.2708, 21.5123, 34.8598, 2.9428, 38.5076, 24.4032, 13.9159, 3...

$ var\_74 *<dbl>* 26.8339, 6.7803, 20.7048, 10.7087, 2.2201, 8.4758, 9.8784, 29.9...

$ var\_75 *<dbl>* 21.7205, 18.1896, 16.4953, 12.2008, 9.5235, 23.8209, 20.3224, 2...

$ var\_76 *<dbl>* 7.3075, 6.9388, -9.7077, 12.5465, 8.1522, 4.2468, 4.5523, -13.5...

$ var\_77 *<dbl>* 14.0810, 22.1336, 19.6357, 19.4201, 14.9224, 19.5714, 20.5327, ...

$ var\_78 *<dbl>* 3.1192, 6.3755, 7.6587, 5.5060, 6.1573, 6.4908, 3.6862, 7.5424,...

$ var\_79 *<dbl>* 17.4265, 13.1525, 15.5744, 14.1586, 15.5221, 12.4225, 14.0201, ...

$ var\_80 *<dbl>* 9.4883, 1.9772, 16.1691, 17.5941, 11.8133, 6.4326, -0.8756, 3.8...

$ var\_81 *<dbl>* 16.9060, 14.0406, 14.3299, 15.4375, 16.7661, 16.0807, 12.5293, ...

$ var\_82 *<dbl>* 14.5117, 6.6904, 1.3360, -13.2668, -14.6524, 1.4890, 5.8005, -1...

$ var\_83 *<dbl>* 10.0276, 9.9732, -0.4412, 14.0885, -0.4469, 13.1412, 11.1811, -...

$ var\_84 *<dbl>* -0.9706, -11.5679, -0.2830, 4.0357, 0.0306, 3.8126, -2.8989, 3....

$ var\_85 *<dbl>* 20.4588, 20.4525, 14.9105, 22.3119, 22.5276, 19.7215, 26.0394, ...

$ var\_86 *<dbl>* 4.7945, 9.4951, -3.9016, 1.8571, 6.9774, -0.3531, 3.3605, 2.080...

$ var\_87 *<dbl>* 20.4160, 9.6343, 14.6881, 16.5210, 2.2563, 17.5688, 7.3378, 5.4...

$ var\_88 *<dbl>* 13.1633, 8.1252, 7.3220, 10.8149, 3.5779, 8.6988, 8.2451, 10.89...

$ var\_89 *<dbl>* 7.9307, 2.6059, -5.1443, 0.3256, 1.4268, 5.2293, 5.2095, 7.5485...

$ var\_90 *<dbl>* -7.6509, -17.4201, -34.3488, -21.4797, 9.0680, -13.2339, -18.74...

$ var\_91 *<dbl>* 7.0834, 7.1848, 7.0194, 6.9174, 7.0197, 6.9382, 6.8274, 7.1200,...

$ var\_92 *<dbl>* 15.2324, 15.3484, 12.4785, 9.9483, 19.7765, 12.9613, 13.9692, 1...

$ var\_93 *<dbl>* 10.1416, 10.6522, 9.6665, 10.3696, 10.0499, 10.6412, 10.2064, 1...

$ var\_94 *<dbl>* 5.9156, 5.9897, 13.2595, 11.0362, 11.4803, 14.2706, 5.3246, 9.2...

$ var\_95 *<dbl>* -0.5775, 0.3392, -0.5624, 0.1892, 0.2548, -0.4687, -1.0409, -0....

$ var\_96 *<dbl>* 5.7600, 10.3516, 5.6347, 19.4321, 16.7029, 9.9168, 24.2579, 12....

$ var\_97 *<dbl>* 30.3238, 29.8204, 9.5853, 40.3383, 45.5510, 2.3534, 5.3655, 31....

$ var\_98 *<dbl>* 2.1251, 1.9998, 1.4515, 1.4105, 1.5795, 0.2386, 2.2065, 1.7538,...

$ var\_99 *<dbl>* 1.8585, -1.4166, 1.7818, 2.6165, 0.1148, -0.6977, -1.7311, -0.7...

$ var\_100 *<dbl>* -9.2198, -1.7257, -3.5065, 1.7021, -14.3858, -19.8952, 6.1458, ...

$ var\_101 *<dbl>* 17.3089, 15.4712, 14.1663, 2.5363, 17.8630, 12.7305, 16.2640, 1...

$ var\_102 *<dbl>* 30.9548, 35.6020, 28.0256, 3.8763, 23.2274, 30.2647, 18.6954, 1...

$ var\_103 *<dbl>* 1.4918, 1.6570, 1.3935, 1.5173, 1.4375, 1.7676, 1.7570, 1.7108,...

$ var\_104 *<dbl>* 12.8721, 13.0783, 10.8257, 13.4083, 14.4838, 14.9591, 7.8975, 9...

$ var\_105 *<dbl>* 3.4902, 2.7752, 4.2954, 2.8965, 4.3806, 4.9202, 2.8033, 5.5384,...

$ var\_106 *<dbl>* 8.2856, 6.4986, 8.2125, 7.0919, 10.6976, 7.9543, 5.7950, 13.075...

$ var\_107 *<dbl>* 11.9794, 4.6835, 26.2595, 21.6304, 18.4023, 26.5086, 24.5927, 2...

$ var\_108 *<dbl>* 14.0176, 13.7963, 14.0232, 14.2000, 14.2212, 14.2583, 14.1619, ...

$ var\_109 *<dbl>* 15.0763, 17.7261, 19.4604, 23.0368, 16.0638, 13.2841, 14.7296, ...

$ var\_110 *<dbl>* 3.7662, 1.7375, 8.6896, 10.3445, 6.3933, 3.3909, 5.8805, 5.5510...

$ var\_111 *<dbl>* 6.0426, 5.5689, 8.1036, 6.0369, 6.8699, 4.7015, 6.4036, 7.7358,...

$ var\_112 *<dbl>* 4.4243, 3.6609, 1.2057, 5.0227, 2.7253, 5.3538, 3.4604, 5.2163,...

$ var\_113 *<dbl>* 14.1799, 8.9725, 8.9156, 12.6600, 12.6458, 4.4087, 4.6195, 13.7...

$ var\_114 *<dbl>* 2.0921, 4.1159, 0.9777, 2.1278, 3.2376, 2.0986, 3.5566, 1.5873,...

$ var\_115 *<dbl>* 1.5493, 1.0693, 2.3797, 4.0592, 3.4218, 0.2104, 6.6851, 6.0762,...

$ var\_116 *<dbl>* 3.2206, 2.0234, 3.1638, 1.9084, -0.5658, 0.8690, -0.0033, 4.904...

$ var\_117 *<dbl>* 0.0172, 8.2760, 37.8664, 11.6095, -5.6840, 20.6923, 20.4088, 20...

$ var\_118 *<dbl>* -6.6602, -6.8610, -3.3864, 7.5397, 4.7753, -1.4024, -0.8950, -1...

$ var\_119 *<dbl>* 8.4785, 0.2780, -2.4090, 8.1972, 10.3320, 2.7168, -0.1137, 7.39...

$ var\_120 *<dbl>* 42.0248, 17.0488, 29.7978, 20.0844, 39.7127, 21.1759, 26.2992, ...

$ var\_121 *<dbl>* 11.4164, 11.6704, 12.2056, 10.4440, 11.2319, 10.7737, 9.9379, 8...

$ var\_122 *<dbl>* 0.4564, 3.1215, 4.7688, 8.4676, -1.2978, 10.1884, 2.2976, 0.985...

$ var\_123 *<dbl>* 9.4006, 8.5093, 7.9344, 5.0350, 12.4827, 4.9242, 6.6694, 11.857...

$ var\_124 *<dbl>* 0.9685, 5.6367, 2.2102, 4.3103, 6.5034, 2.0627, -0.8026, 3.4626...

$ var\_125 *<dbl>* 12.4929, 12.0099, 12.6482, 12.0067, 12.7157, 11.8315, 12.1630, ...

$ var\_126 *<dbl>* 14.1240, 14.2372, 14.3377, 13.7149, 13.3054, 14.0724, 13.1157, ...

$ var\_127 *<dbl>* 4.0388, -6.1600, 2.3268, 1.6143, -1.9678, 0.7827, 1.8116, -0.83...

$ var\_128 *<dbl>* -4.4442, -5.6690, 2.3930, -1.2328, -1.2363, -6.7349, -0.5829, -...

$ var\_129 *<dbl>* 16.6684, 8.9094, 13.7005, 22.7248, 11.5686, 18.2366, 11.1905, 1...

$ var\_130 *<dbl>* 12.5380, 11.0605, 12.7047, 12.6609, 12.6428, 13.4610, 11.2068, ...

$ var\_131 *<dbl>* 0.9205, 0.4583, 0.7507, 0.8039, 0.4792, 1.1437, 0.8252, 1.0311,...

$ var\_132 *<dbl>* 10.5998, 9.7974, 7.7726, 4.7666, 7.1984, 8.8194, 8.3042, 8.3112...

$ var\_133 *<dbl>* 7.5147, 7.0891, 6.5950, 6.7888, 7.1434, 6.4073, 6.6573, 7.4290,...

$ var\_134 *<dbl>* -4.1748, 2.6849, 0.2990, 5.8537, -0.2056, -2.4266, -7.7321, -3....

$ var\_135 *<dbl>* -0.4824, 8.4970, 12.9154, -4.5434, -16.3908, -1.2834, -6.2635, ...

$ var\_136 *<dbl>* 10.5267, 15.7774, 29.9162, 19.0111, 27.1589, 16.7356, 7.4701, 1...

$ var\_137 *<dbl>* 17.7547, 4.8775, 6.8031, 12.6907, 23.5997, 15.0301, 0.4679, 1.1...

$ var\_138 *<dbl>* -6.5226, 3.6129, 10.5031, -2.9322, -4.6175, 2.0551, 2.1011, -3....

$ var\_139 *<dbl>* -2.5502, 6.7530, -6.0452, 12.7898, 11.7989, -0.7009, 3.5594, 5....

$ var\_140 *<dbl>* -5.1547, 11.1003, -4.5298, 12.0466, 12.5683, -5.9772, 5.1848, 9...

$ var\_141 *<dbl>* -2.1246, 15.3593, 1.3903, 13.1646, -3.6145, 5.3172, 9.4451, 6.9...

$ var\_142 *<dbl>* 19.8319, 2.2105, 5.0469, 7.7063, 22.1069, 18.2809, 4.8391, 15.8...

$ var\_143 *<dbl>* 13.0752, 8.2280, 12.9740, 11.6549, 9.5539, 14.0878, 13.5847, 11...

$ var\_144 *<dbl>* 9.2275, 9.0717, 9.3878, 9.8274, 9.2721, 8.8363, 9.1868, 7.9618,...

$ var\_145 *<dbl>* 3.0213, -5.0947, -0.1113, 1.8061, -1.6214, 10.1988, 0.0095, 2.5...

$ var\_146 *<dbl>* 11.6793, 8.7644, 11.6749, 8.6963, 12.9327, 11.0207, 5.8965, 14....

$ var\_147 *<dbl>* -11.6827, -2.2873, 16.8588, 1.8057, 6.8080, 7.5835, 4.7403, -0....

$ var\_148 *<dbl>* 4.1017, 4.1240, 4.2600, 3.8265, 4.2135, 4.0888, 3.9294, 4.0720,...

$ var\_149 *<dbl>* 5.2954, -13.3006, 14.6476, -16.3027, 22.1044, 10.2551, 19.9754,...

$ var\_150 *<dbl>* 18.7741, 18.7454, 14.4431, 13.7106, 20.0502, 14.5978, 17.7241, ...

$ var\_151 *<dbl>* 9.8892, 9.3783, 14.1649, 9.7908, 6.9953, 12.0263, 6.0450, 11.69...

$ var\_152 *<dbl>* 7.5219, 1.5284, 9.4875, 5.8497, 9.3823, 13.4995, 8.5249, 4.3652...

$ var\_153 *<dbl>* 14.9745, 16.0407, 16.5769, 15.4378, 20.5534, 17.4899, 17.4007, ...

$ var\_154 *<dbl>* 18.9880, 7.7732, 7.2638, 5.0372, 3.4368, 3.2240, 10.0876, 8.227...

$ var\_155 *<dbl>* 1.0842, 1.4316, -2.2008, -8.7673, -15.2208, 4.3380, -1.9114, -2...

$ var\_156 *<dbl>* 11.9125, 14.8679, 12.5953, 13.6035, 13.0974, 11.7515, 13.3160, ...

$ var\_157 *<dbl>* -4.5103, 3.3619, 7.4487, -3.5002, -14.0888, -11.6321, -7.9423, ...

$ var\_158 *<dbl>* 16.1361, 11.5799, 23.1407, 13.9785, 11.7586, 8.8682, 20.4256, 2...

$ var\_159 *<dbl>* 11.0067, 14.2058, 10.4597, 14.6118, 14.5259, 1.4625, 17.8738, 8...

$ var\_160 *<dbl>* 5.9232, 30.9641, 39.3654, 19.7251, 22.8700, 10.4064, 16.6887, 3...

$ var\_161 *<dbl>* 5.4113, 5.6723, 5.5228, 5.3882, 5.6688, 5.6731, 5.5245, 5.5475,...

$ var\_162 *<dbl>* 3.8302, 3.6873, 3.3159, 3.6775, 6.1159, 5.2368, 4.1781, 5.0035,...

$ var\_163 *<dbl>* 5.7380, 13.0429, 4.3324, 7.4753, 13.2433, 5.2973, 16.6762, 13.2...

$ var\_164 *<dbl>* -8.6105, -10.6572, -0.5382, -11.0780, -11.9785, -13.2809, -6.21...

$ var\_165 *<dbl>* 22.9530, 15.5134, 13.3009, 24.8712, 26.2040, 26.6861, 23.8006, ...

$ var\_166 *<dbl>* 2.5531, 3.2185, 3.1243, 2.6415, 3.2348, 3.0711, 2.8187, 3.1809,...

$ var\_167 *<dbl>* -0.2836, 9.0535, -4.1731, 2.2673, -5.5775, -5.8410, -16.7911, -...

$ var\_168 *<dbl>* 4.3416, 7.0535, 1.2330, 7.2788, 5.7036, 3.2268, 4.1165, 0.7334,...

$ var\_169 *<dbl>* 5.1855, 5.3924, 6.1513, 5.6406, 6.1717, 5.9344, 5.7013, 5.7269,...

$ var\_170 *<dbl>* 4.2603, -0.7720, -0.0391, 7.2048, -1.6039, 0.6292, 8.8525, 2.58...

$ var\_171 *<dbl>* 1.6779, -8.1783, 1.4950, 3.4504, -2.4866, -4.0885, 0.2582, -1.2...

$ var\_172 *<dbl>* 29.0849, 29.9227, 16.8874, 2.4130, 17.2728, 15.8299, 17.7310, 2...

$ var\_173 *<dbl>* 8.4685, -5.6274, -2.9787, 11.1674, 2.3640, 1.9671, -10.3297, 6....

$ var\_174 *<dbl>* 18.1317, 10.5018, 27.4035, 14.5499, 14.0037, 27.9116, 20.0649, ...

$ var\_175 *<dbl>* 12.2818, 9.6083, 15.8819, 10.6151, 12.9165, 12.1392, 11.6405, 1...

$ var\_176 *<dbl>* -0.6912, -0.4935, -10.9660, -5.7922, -12.0311, -5.8638, -4.2624...

$ var\_177 *<dbl>* 10.2226, 8.1696, 15.6415, 13.9407, 10.1161, 7.9274, 12.0176, 13...

$ var\_178 *<dbl>* -5.5579, -4.3605, -9.4056, 7.1078, -8.7562, 3.2828, 10.9073, -1...

$ var\_179 *<dbl>* 2.2926, 5.2110, 4.4611, 1.1019, 6.0889, 1.6738, 2.2162, 2.0752,...

$ var\_180 *<dbl>* -4.5358, 0.4087, -3.0835, 9.4590, -1.3620, -6.6870, -4.0787, -0...

$ var\_181 *<dbl>* 10.3903, 12.0030, 8.5549, 9.8243, 10.3559, 10.8213, 8.1047, 9.4...

$ var\_182 *<dbl>* -15.4937, -10.3812, -2.8517, 5.9917, -7.4915, -9.0377, 0.3254, ...

$ var\_183 *<dbl>* 3.9697, 5.8496, 13.4770, 5.1634, 9.4588, 13.9614, 13.1701, 9.34...

$ var\_184 *<dbl>* 31.3521, 25.1958, 24.4721, 8.1154, 3.9829, 5.7416, 16.1184, 20....

$ var\_185 *<dbl>* -1.1651, -8.8468, -3.4824, 3.6638, 5.8580, 3.6823, -3.1572, 0.9...

$ var\_186 *<dbl>* 9.2874, 11.8263, 4.9178, 3.3102, 8.3635, 5.1080, 10.4842, 9.660...

$ var\_187 *<dbl>* -23.5705, -8.7112, -2.0720, -19.7819, -24.8254, -7.5139, -10.69...

$ var\_188 *<dbl>* 13.2643, 15.9072, 11.5390, 13.4499, 11.4928, 15.2017, 12.8029, ...

$ var\_189 *<dbl>* 1.6591, 0.9812, 1.1821, 1.3104, 1.6321, 2.6198, 2.1781, -1.0241...

$ var\_190 *<dbl>* -2.1556, 10.6165, -0.7484, 9.5702, 4.2259, -2.1115, 12.3609, 4....

$ var\_191 *<dbl>* 11.8495, 8.8349, 10.9935, 9.0766, 9.1723, 7.1178, 6.8661, 4.421...

$ var\_192 *<dbl>* -1.4300, 0.9403, 1.9803, 1.6580, 1.2835, -0.4249, 4.0971, 0.930...

$ var\_193 *<dbl>* 2.4508, 10.1282, 2.1800, 3.5813, 3.3778, 8.8781, 8.8484, 1.4994...

$ var\_194 *<dbl>* 13.7112, 15.5765, 12.9813, 15.1874, 19.5542, 14.9438, 17.5010, ...

$ var\_195 *<dbl>* 2.4669, 0.4773, 2.1281, 3.1656, -0.2860, -2.2151, 0.0295, -1.79...

$ var\_196 *<dbl>* 4.3654, -1.4852, -7.1086, 3.9567, -5.1612, -6.0233, 7.7443, 6.5...

$ var\_197 *<dbl>* 10.7200, 9.8714, 7.0618, 9.2295, 7.2882, 9.8117, 9.1509, 10.485...

$ var\_198 *<dbl>* 15.4722, 19.1293, 19.8956, 13.0168, 13.9260, 17.1127, 18.4736, ...

$ var\_199 *<dbl>* -8.7197, -20.9760, -23.1794, -4.2108, -9.1846, 10.8240, 5.1499,...

* In train set, there are 200,000 observations and 200 features(var\_0~var\_199) with target variable.
* In test set, there are 200,000 observations and 200 features without target variable.

Code

train$ID\_code %>% unique %>% length

[1] 200000

test$ID\_code %>% unique %>% length

[1] 200000

* ID\_code variable has unique values for each dataset.

**Code**

colSums(is.na(train))

ID\_code target var\_0 var\_1 var\_2 var\_3 var\_4 var\_5 var\_6 var\_7

0 0 0 0 0 0 0 0 0 0

var\_8 var\_9 var\_10 var\_11 var\_12 var\_13 var\_14 var\_15 var\_16 var\_17

0 0 0 0 0 0 0 0 0 0

var\_18 var\_19 var\_20 var\_21 var\_22 var\_23 var\_24 var\_25 var\_26 var\_27

0 0 0 0 0 0 0 0 0 0

var\_28 var\_29 var\_30 var\_31 var\_32 var\_33 var\_34 var\_35 var\_36 var\_37

0 0 0 0 0 0 0 0 0 0

var\_38 var\_39 var\_40 var\_41 var\_42 var\_43 var\_44 var\_45 var\_46 var\_47

0 0 0 0 0 0 0 0 0 0

var\_48 var\_49 var\_50 var\_51 var\_52 var\_53 var\_54 var\_55 var\_56 var\_57

0 0 0 0 0 0 0 0 0 0

var\_58 var\_59 var\_60 var\_61 var\_62 var\_63 var\_64 var\_65 var\_66 var\_67

0 0 0 0 0 0 0 0 0 0

var\_68 var\_69 var\_70 var\_71 var\_72 var\_73 var\_74 var\_75 var\_76 var\_77

0 0 0 0 0 0 0 0 0 0

var\_78 var\_79 var\_80 var\_81 var\_82 var\_83 var\_84 var\_85 var\_86 var\_87

0 0 0 0 0 0 0 0 0 0

var\_88 var\_89 var\_90 var\_91 var\_92 var\_93 var\_94 var\_95 var\_96 var\_97

0 0 0 0 0 0 0 0 0 0

var\_98 var\_99 var\_100 var\_101 var\_102 var\_103 var\_104 var\_105 var\_106 var\_107

0 0 0 0 0 0 0 0 0 0

var\_108 var\_109 var\_110 var\_111 var\_112 var\_113 var\_114 var\_115 var\_116 var\_117

0 0 0 0 0 0 0 0 0 0

var\_118 var\_119 var\_120 var\_121 var\_122 var\_123 var\_124 var\_125 var\_126 var\_127

0 0 0 0 0 0 0 0 0 0

var\_128 var\_129 var\_130 var\_131 var\_132 var\_133 var\_134 var\_135 var\_136 var\_137

0 0 0 0 0 0 0 0 0 0

var\_138 var\_139 var\_140 var\_141 var\_142 var\_143 var\_144 var\_145 var\_146 var\_147

0 0 0 0 0 0 0 0 0 0

var\_148 var\_149 var\_150 var\_151 var\_152 var\_153 var\_154 var\_155 var\_156 var\_157

0 0 0 0 0 0 0 0 0 0

var\_158 var\_159 var\_160 var\_161 var\_162 var\_163 var\_164 var\_165 var\_166 var\_167

0 0 0 0 0 0 0 0 0 0

var\_168 var\_169 var\_170 var\_171 var\_172 var\_173 var\_174 var\_175 var\_176 var\_177

0 0 0 0 0 0 0 0 0 0

var\_178 var\_179 var\_180 var\_181 var\_182 var\_183 var\_184 var\_185 var\_186 var\_187

0 0 0 0 0 0 0 0 0 0

var\_188 var\_189 var\_190 var\_191 var\_192 var\_193 var\_194 var\_195 var\_196 var\_197

0 0 0 0 0 0 0 0 0 0

var\_198 var\_199

0 0

**Code**

colSums(is.na(test))

ID\_code var\_0 var\_1 var\_2 var\_3 var\_4 var\_5 var\_6 var\_7 var\_8

0 0 0 0 0 0 0 0 0 0

var\_9 var\_10 var\_11 var\_12 var\_13 var\_14 var\_15 var\_16 var\_17 var\_18

0 0 0 0 0 0 0 0 0 0

var\_19 var\_20 var\_21 var\_22 var\_23 var\_24 var\_25 var\_26 var\_27 var\_28

0 0 0 0 0 0 0 0 0 0

var\_29 var\_30 var\_31 var\_32 var\_33 var\_34 var\_35 var\_36 var\_37 var\_38

0 0 0 0 0 0 0 0 0 0

var\_39 var\_40 var\_41 var\_42 var\_43 var\_44 var\_45 var\_46 var\_47 var\_48

0 0 0 0 0 0 0 0 0 0

var\_49 var\_50 var\_51 var\_52 var\_53 var\_54 var\_55 var\_56 var\_57 var\_58

0 0 0 0 0 0 0 0 0 0

var\_59 var\_60 var\_61 var\_62 var\_63 var\_64 var\_65 var\_66 var\_67 var\_68

0 0 0 0 0 0 0 0 0 0

var\_69 var\_70 var\_71 var\_72 var\_73 var\_74 var\_75 var\_76 var\_77 var\_78

0 0 0 0 0 0 0 0 0 0

var\_79 var\_80 var\_81 var\_82 var\_83 var\_84 var\_85 var\_86 var\_87 var\_88

0 0 0 0 0 0 0 0 0 0

var\_89 var\_90 var\_91 var\_92 var\_93 var\_94 var\_95 var\_96 var\_97 var\_98

0 0 0 0 0 0 0 0 0 0

var\_99 var\_100 var\_101 var\_102 var\_103 var\_104 var\_105 var\_106 var\_107 var\_108

0 0 0 0 0 0 0 0 0 0

var\_109 var\_110 var\_111 var\_112 var\_113 var\_114 var\_115 var\_116 var\_117 var\_118

0 0 0 0 0 0 0 0 0 0

var\_119 var\_120 var\_121 var\_122 var\_123 var\_124 var\_125 var\_126 var\_127 var\_128

0 0 0 0 0 0 0 0 0 0

var\_129 var\_130 var\_131 var\_132 var\_133 var\_134 var\_135 var\_136 var\_137 var\_138

0 0 0 0 0 0 0 0 0 0

var\_139 var\_140 var\_141 var\_142 var\_143 var\_144 var\_145 var\_146 var\_147 var\_148

0 0 0 0 0 0 0 0 0 0

var\_149 var\_150 var\_151 var\_152 var\_153 var\_154 var\_155 var\_156 var\_157 var\_158

0 0 0 0 0 0 0 0 0 0

var\_159 var\_160 var\_161 var\_162 var\_163 var\_164 var\_165 var\_166 var\_167 var\_168

0 0 0 0 0 0 0 0 0 0

var\_169 var\_170 var\_171 var\_172 var\_173 var\_174 var\_175 var\_176 var\_177 var\_178

0 0 0 0 0 0 0 0 0 0

var\_179 var\_180 var\_181 var\_182 var\_183 var\_184 var\_185 var\_186 var\_187 var\_188

0 0 0 0 0 0 0 0 0 0

var\_189 var\_190 var\_191 var\_192 var\_193 var\_194 var\_195 var\_196 var\_197 var\_198

0 0 0 0 0 0 0 0 0 0

var\_199

0

Target

table(train$target)

0 1

179902 20098

target\_df <- data.frame(table(train$target))

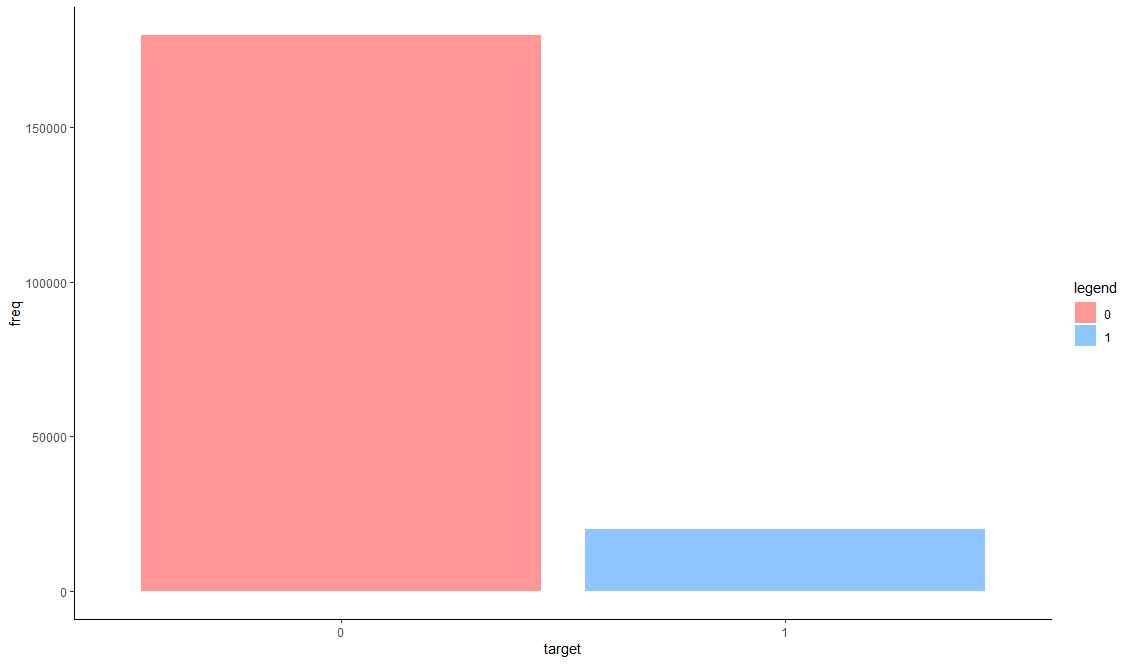
colnames(target\_df) <- c("target", "freq")

ggplot(data=target\_df, aes(x=target, y=freq, fill=target)) +

geom\_bar(position = 'dodge', stat='identity', alpha=0.5) +

scale\_fill\_manual("legend", values = c("1" = "dodgerblue", "0"="firebrick1")) +

theme\_classic()



**Features**

### Density plots for train data

feature\_groups <- 3:22

col\_names <- colnames(train)[c(2,feature\_groups)]

temp <- gather(train[,col\_names, with=F], key="features", value="value", -target)

temp$target <- factor(temp$target)

temp$features <- factor(temp$features, levels=col\_names[-1], labels=col\_names[-1])

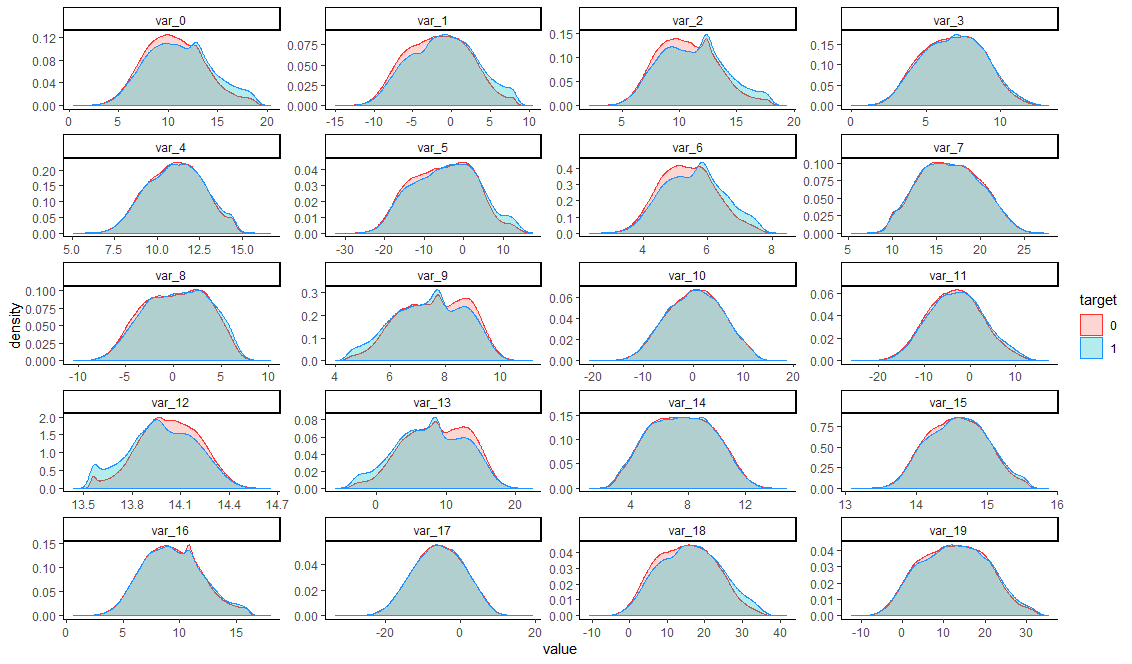
ggplot(data=temp, aes(x=value)) +

geom\_density(aes(fill=target, color=target), alpha=0.3) +

scale\_color\_manual(values = c("1" = "dodgerblue", "0"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

col\_names <- colnames(train)[c(2,feature\_groups+20)]

temp <- gather(train[,col\_names, with=F], key="features", value="value", -target)

temp$target <- factor(temp$target)

temp$features <- factor(temp$features, levels=col\_names[-1], labels=col\_names[-1])

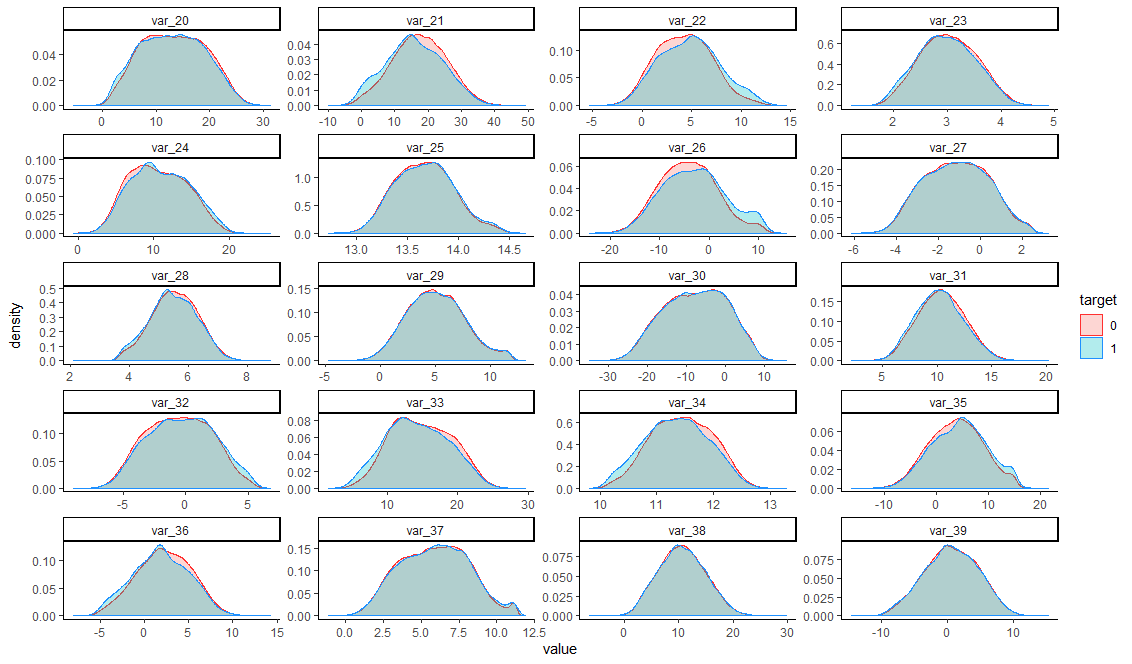
ggplot(data=temp, aes(x=value)) +

geom\_density(aes(fill=target, color=target), alpha=0.3) +

scale\_color\_manual(values = c("1" = "dodgerblue", "0"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")



col\_names <- colnames(train)[c(2,feature\_groups+40)]

temp <- gather(train[,col\_names, with=F], key="features", value="value", -target)

temp$target <- factor(temp$target)

temp$features <- factor(temp$features, levels=col\_names[-1], labels=col\_names[-1])

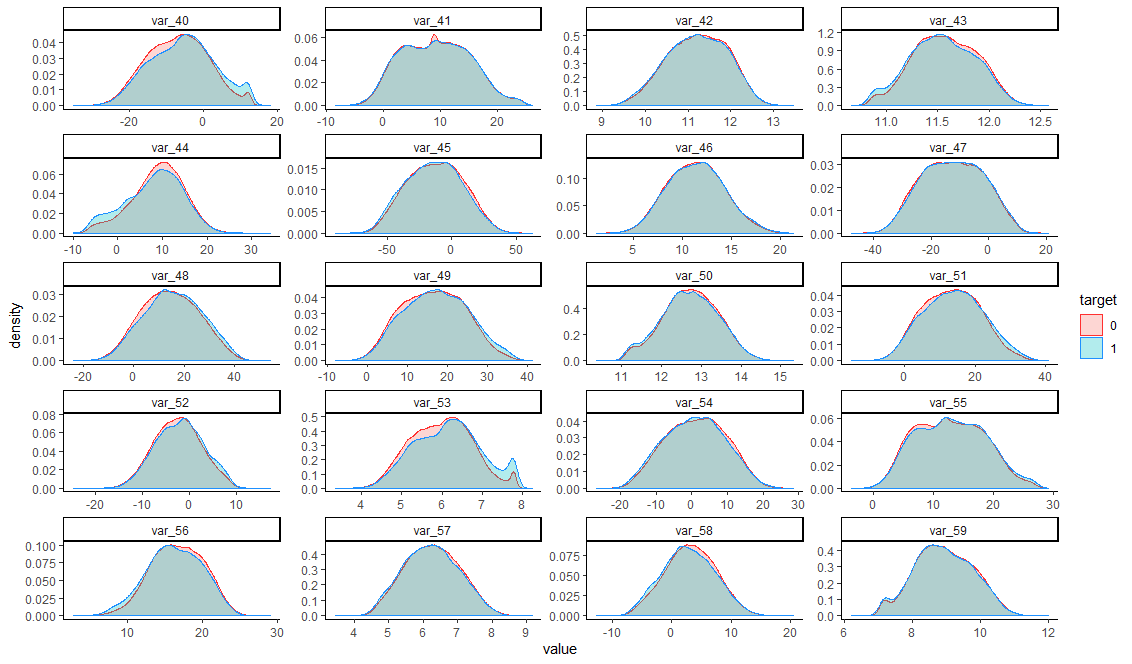
ggplot(data=temp, aes(x=value)) +

geom\_density(aes(fill=target, color=target), alpha=0.3) +

scale\_color\_manual(values = c("1" = "dodgerblue", "0"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")



col\_names <- colnames(train)[c(2,feature\_groups+60)]

temp <- gather(train[,col\_names, with=F], key="features", value="value", -target)

temp$target <- factor(temp$target)

temp$features <- factor(temp$features, levels=col\_names[-1], labels=col\_names[-1])

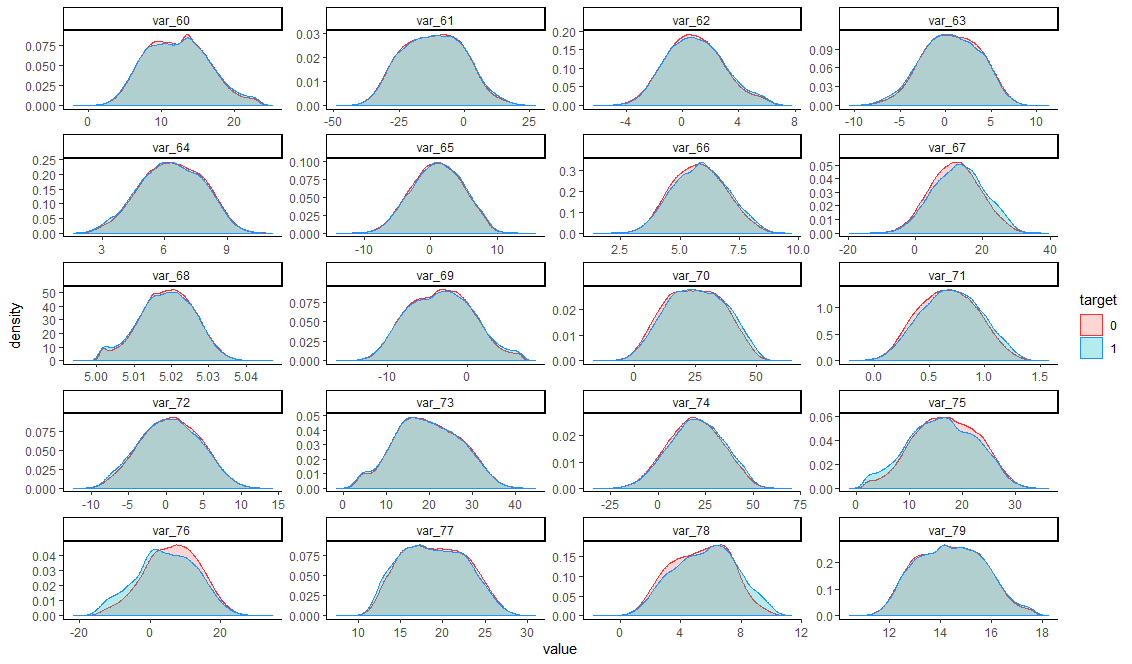
ggplot(data=temp, aes(x=value)) +

geom\_density(aes(fill=target, color=target), alpha=0.3) +

scale\_color\_manual(values = c("1" = "dodgerblue", "0"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")



col\_names <- colnames(train)[c(2,feature\_groups+80)]

temp <- gather(train[,col\_names, with=F], key="features", value="value", -target)

temp$target <- factor(temp$target)

temp$features <- factor(temp$features, levels=col\_names[-1], labels=col\_names[-1])

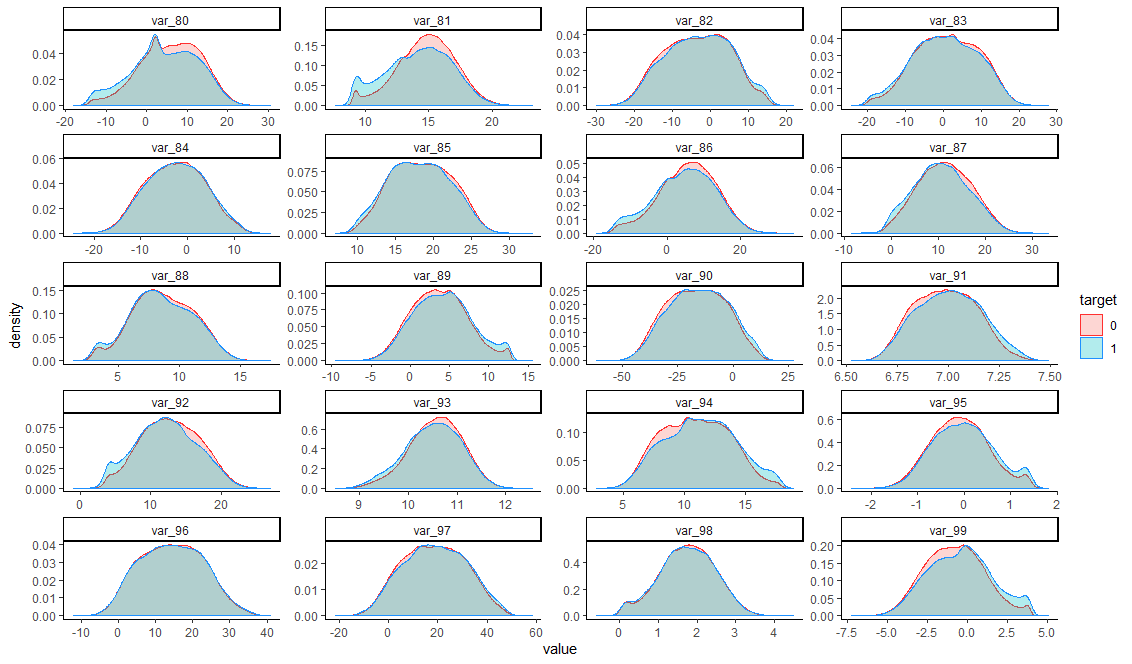
ggplot(data=temp, aes(x=value)) +

geom\_density(aes(fill=target, color=target), alpha=0.3) +

scale\_color\_manual(values = c("1" = "dodgerblue", "0"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")



col\_names <- colnames(train)[c(2,feature\_groups+100)]

temp <- gather(train[,col\_names, with=F], key="features", value="value", -target)

temp$target <- factor(temp$target)

temp$features <- factor(temp$features, levels=col\_names[-1], labels=col\_names[-1])

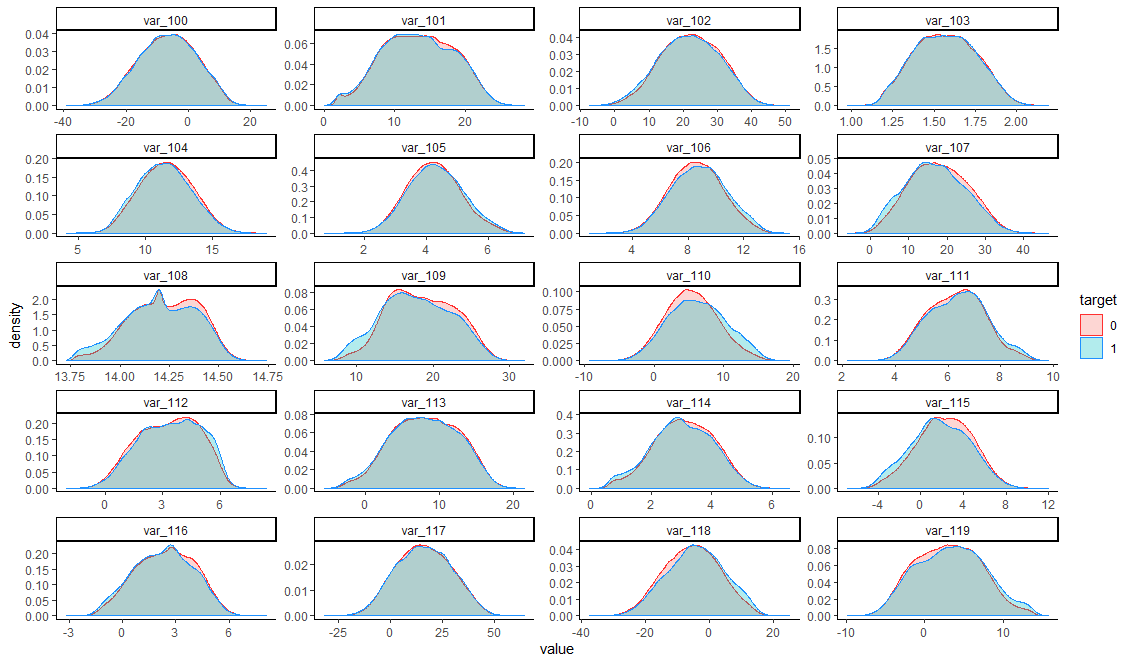
ggplot(data=temp, aes(x=value)) +

geom\_density(aes(fill=target, color=target), alpha=0.3) +

scale\_color\_manual(values = c("1" = "dodgerblue", "0"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")



col\_names <- colnames(train)[c(2,feature\_groups+120)]

temp <- gather(train[,col\_names, with=F], key="features", value="value", -target)

temp$target <- factor(temp$target)

temp$features <- factor(temp$features, levels=col\_names[-1], labels=col\_names[-1])

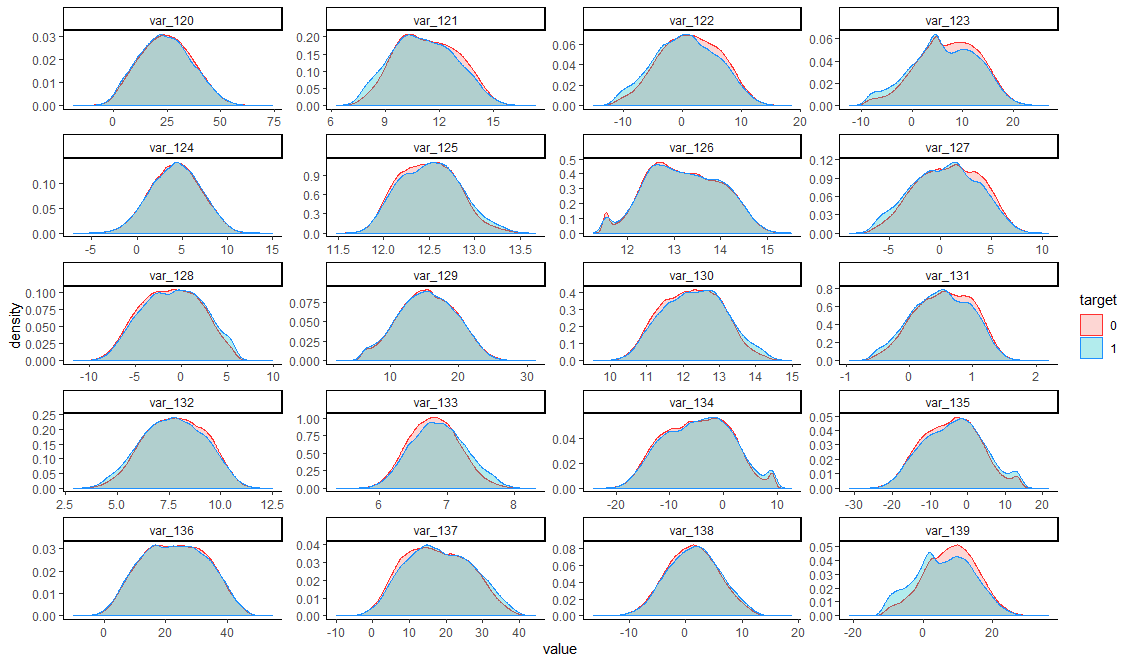
ggplot(data=temp, aes(x=value)) +

geom\_density(aes(fill=target, color=target), alpha=0.3) +

scale\_color\_manual(values = c("1" = "dodgerblue", "0"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")



col\_names <- colnames(train)[c(2,feature\_groups+140)]

temp <- gather(train[,col\_names, with=F], key="features", value="value", -target)

temp$target <- factor(temp$target)

temp$features <- factor(temp$features, levels=col\_names[-1], labels=col\_names[-1])

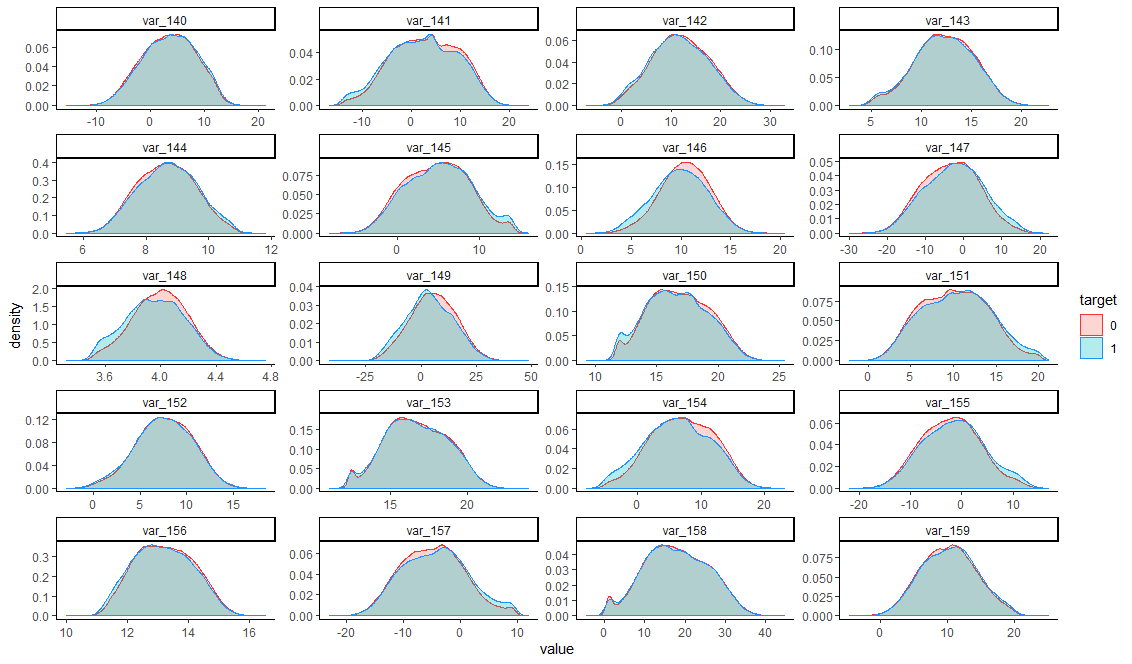
ggplot(data=temp, aes(x=value)) +

geom\_density(aes(fill=target, color=target), alpha=0.3) +

scale\_color\_manual(values = c("1" = "dodgerblue", "0"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")



col\_names <- colnames(train)[c(2,feature\_groups+160)]

temp <- gather(train[,col\_names, with=F], key="features", value="value", -target)

temp$target <- factor(temp$target)

temp$features <- factor(temp$features, levels=col\_names[-1], labels=col\_names[-1])

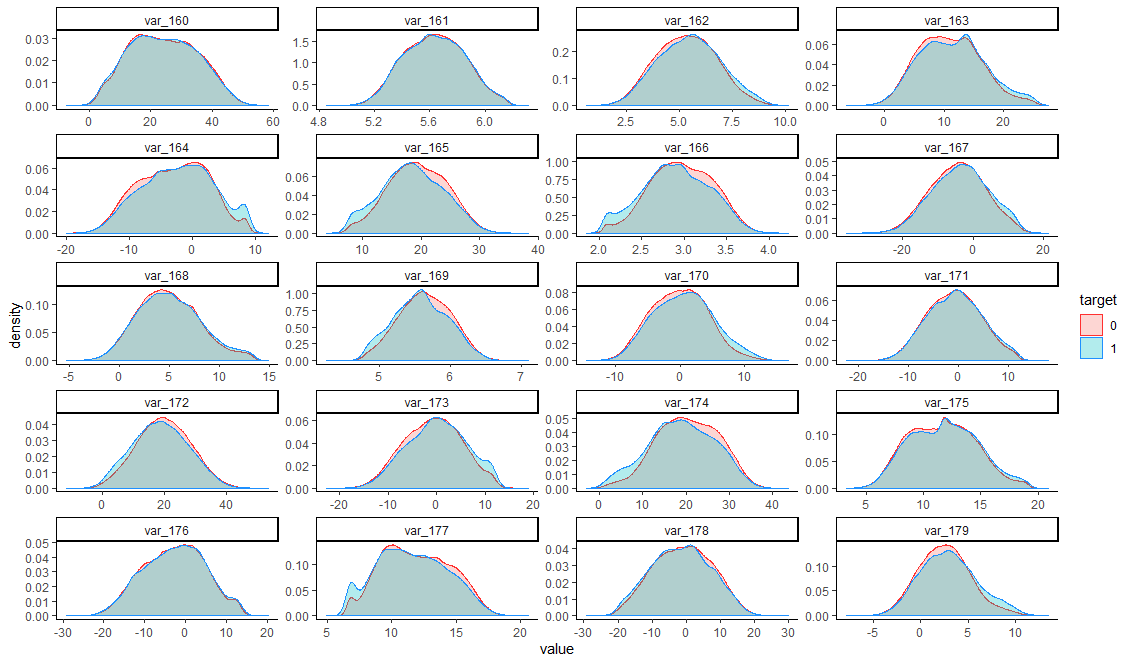
ggplot(data=temp, aes(x=value)) +

geom\_density(aes(fill=target, color=target), alpha=0.3) +

scale\_color\_manual(values = c("1" = "dodgerblue", "0"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")



col\_names <- colnames(train)[c(2,feature\_groups+180)]

temp <- gather(train[,col\_names, with=F], key="features", value="value", -target)

temp$target <- factor(temp$target)

temp$features <- factor(temp$features, levels=col\_names[-1], labels=col\_names[-1])

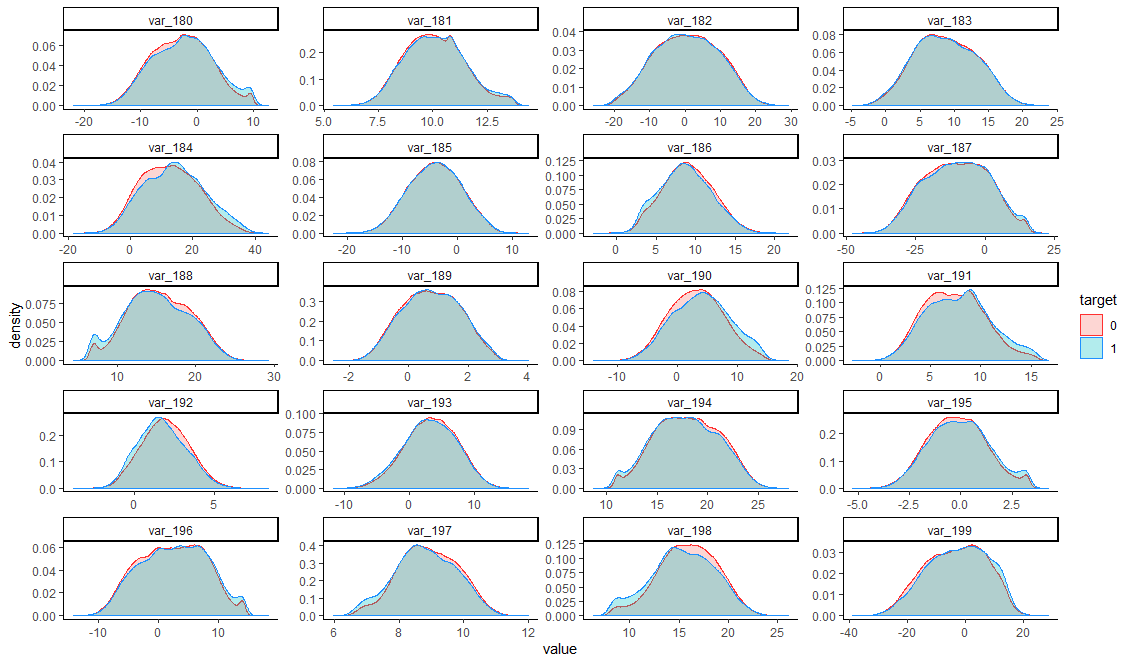
ggplot(data=temp, aes(x=value)) +

geom\_density(aes(fill=target, color=target), alpha=0.3) +

scale\_color\_manual(values = c("1" = "dodgerblue", "0"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")



**Compare density plots train and test**

train$key <- "train"

test$key <- "test"

full <- rbind(train[,-(1:2)], test[,-1])

full$key <- factor(full$key)

feature\_groups <- 1:20

col\_names <- colnames(full)[feature\_groups]

temp <- gather(full[,c("key",col\_names), with=F], key="features", value="value", -key)

temp$features <- factor(temp$features, levels=col\_names, labels=col\_names)

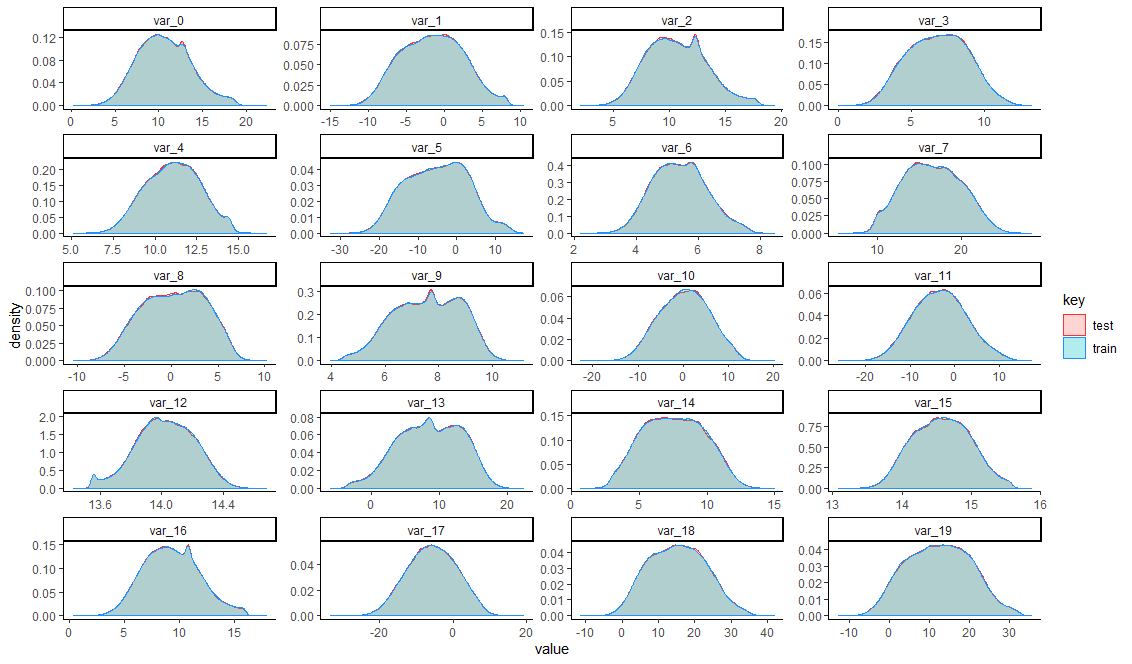
ggplot(data=temp, aes(x=value)) +

geom\_density(aes(fill=key, color=key), alpha=0.3) +

scale\_color\_manual(values = c("train" = "dodgerblue", "test"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

feature\_groups <- 1:20

col\_names <- colnames(full)[feature\_groups+20]

temp <- gather(full[,c("key",col\_names), with=F], key="features", value="value", -key)

temp$features <- factor(temp$features, levels=col\_names, labels=col\_names)

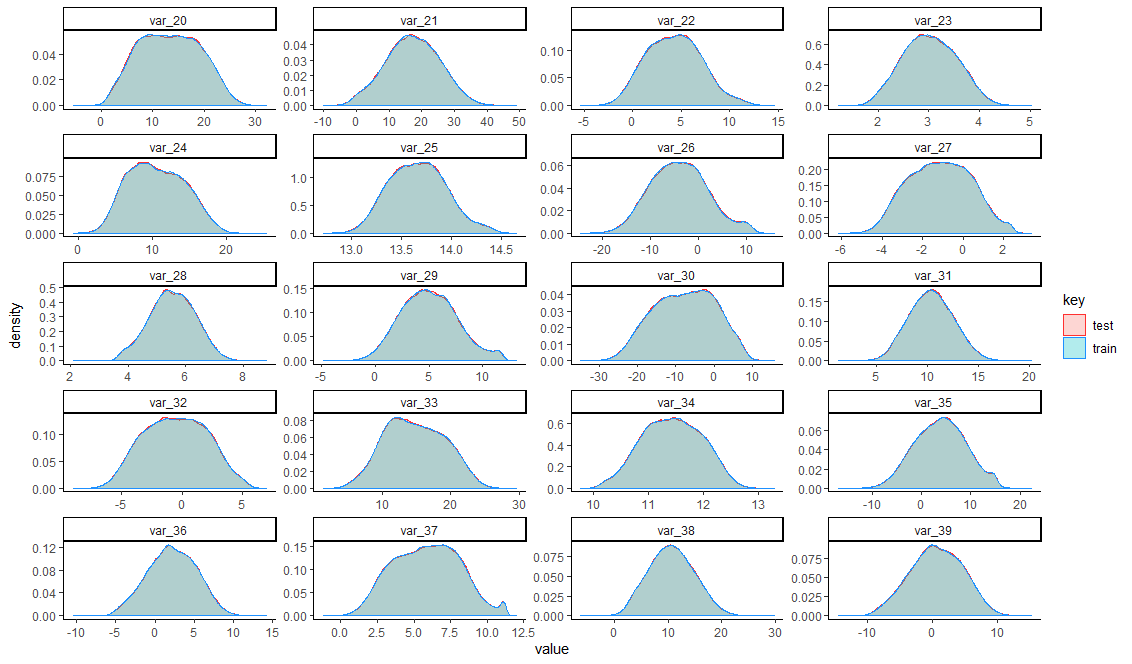
ggplot(data=temp, aes(x=value)) +

geom\_density(aes(fill=key, color=key), alpha=0.3) +

scale\_color\_manual(values = c("train" = "dodgerblue", "test"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

feature\_groups <- 1:20

col\_names <- colnames(full)[feature\_groups+40]

temp <- gather(full[,c("key",col\_names), with=F], key="features", value="value", -key)

temp$features <- factor(temp$features, levels=col\_names, labels=col\_names)

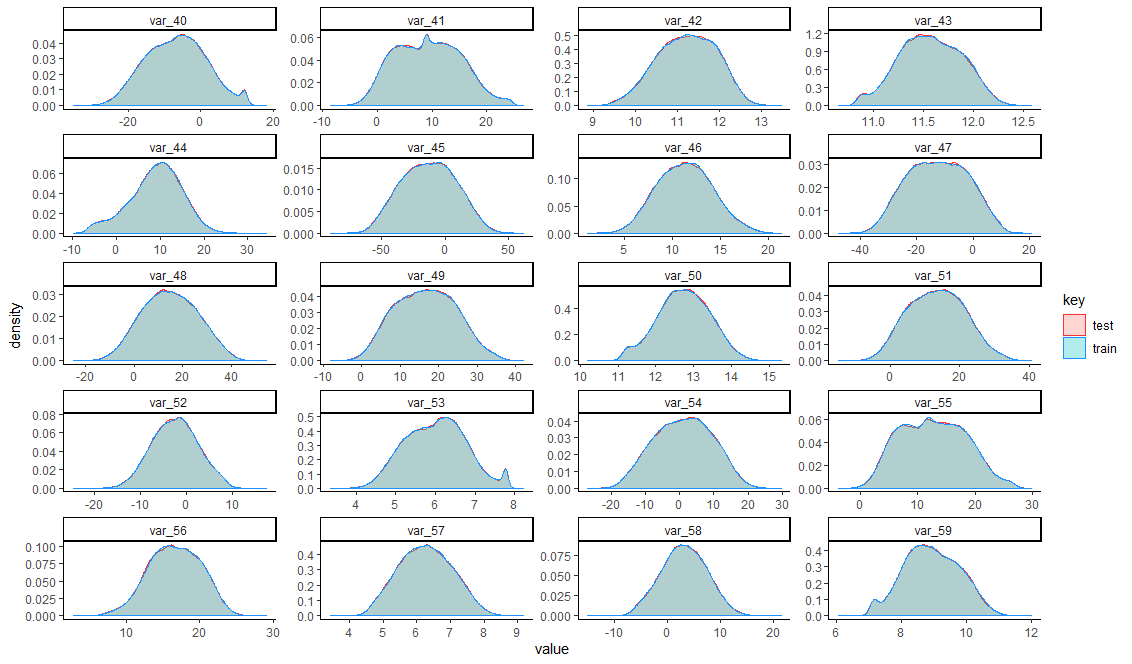
ggplot(data=temp, aes(x=value)) +

geom\_density(aes(fill=key, color=key), alpha=0.3) +

scale\_color\_manual(values = c("train" = "dodgerblue", "test"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

feature\_groups <- 1:20

col\_names <- colnames(full)[feature\_groups+60]

temp <- gather(full[,c("key",col\_names), with=F], key="features", value="value", -key)

temp$features <- factor(temp$features, levels=col\_names, labels=col\_names)

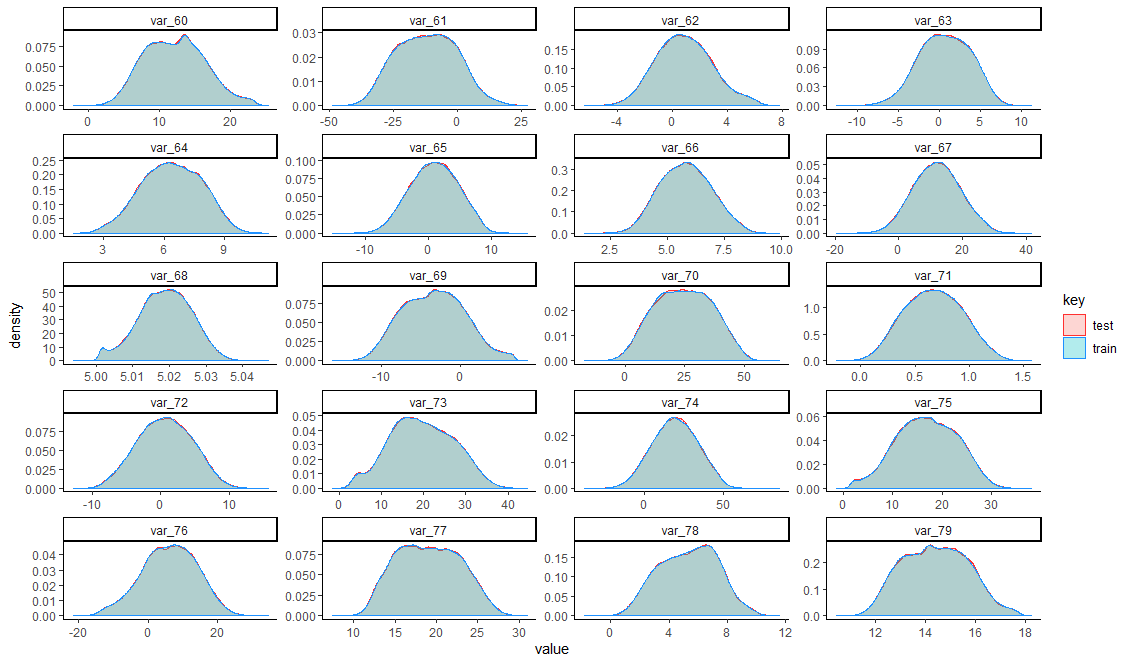
ggplot(data=temp, aes(x=value)) +

geom\_density(aes(fill=key, color=key), alpha=0.3) +

scale\_color\_manual(values = c("train" = "dodgerblue", "test"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

feature\_groups <- 1:20

col\_names <- colnames(full)[feature\_groups+80]

temp <- gather(full[,c("key",col\_names), with=F], key="features", value="value", -key)

temp$features <- factor(temp$features, levels=col\_names, labels=col\_names)

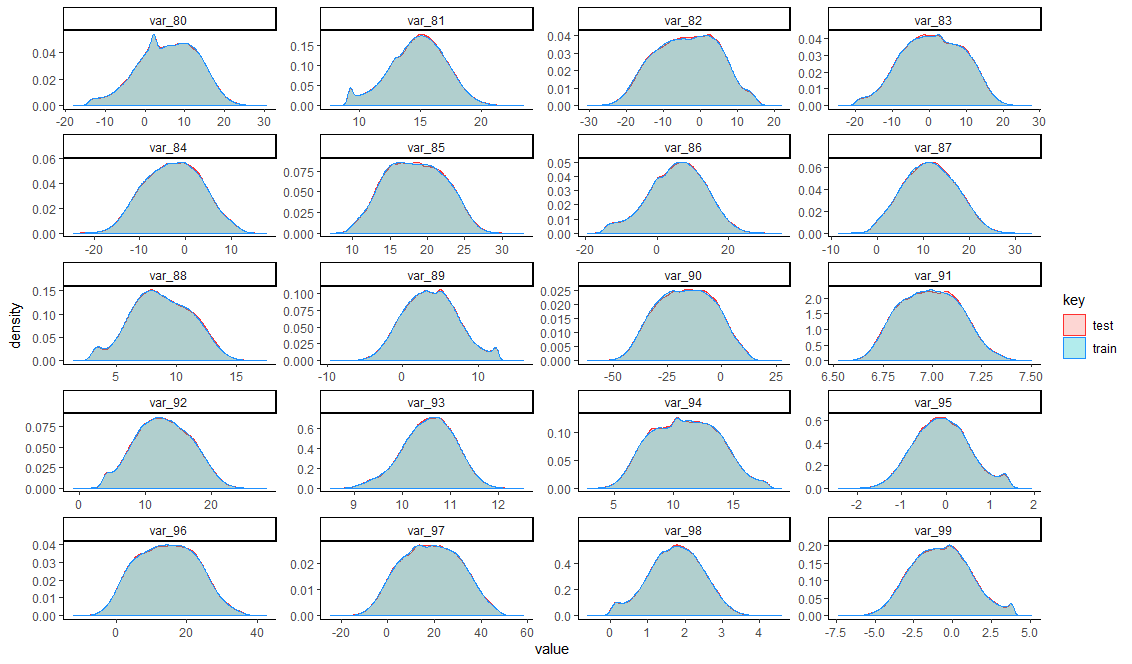
ggplot(data=temp, aes(x=value)) +

geom\_density(aes(fill=key, color=key), alpha=0.3) +

scale\_color\_manual(values = c("train" = "dodgerblue", "test"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

feature\_groups <- 1:20

col\_names <- colnames(full)[feature\_groups+100]

temp <- gather(full[,c("key",col\_names), with=F], key="features", value="value", -key)

temp$features <- factor(temp$features, levels=col\_names, labels=col\_names)

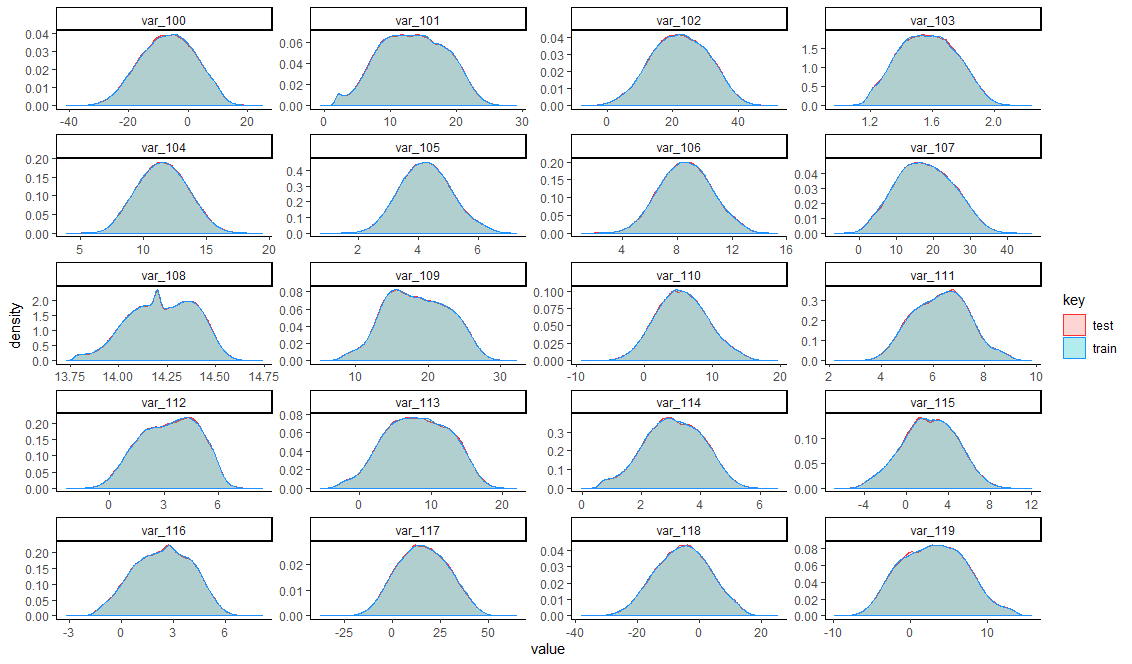
ggplot(data=temp, aes(x=value)) +

geom\_density(aes(fill=key, color=key), alpha=0.3) +

scale\_color\_manual(values = c("train" = "dodgerblue", "test"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

feature\_groups <- 1:20

col\_names <- colnames(full)[feature\_groups+120]

temp <- gather(full[,c("key",col\_names), with=F], key="features", value="value", -key)

temp$features <- factor(temp$features, levels=col\_names, labels=col\_names)

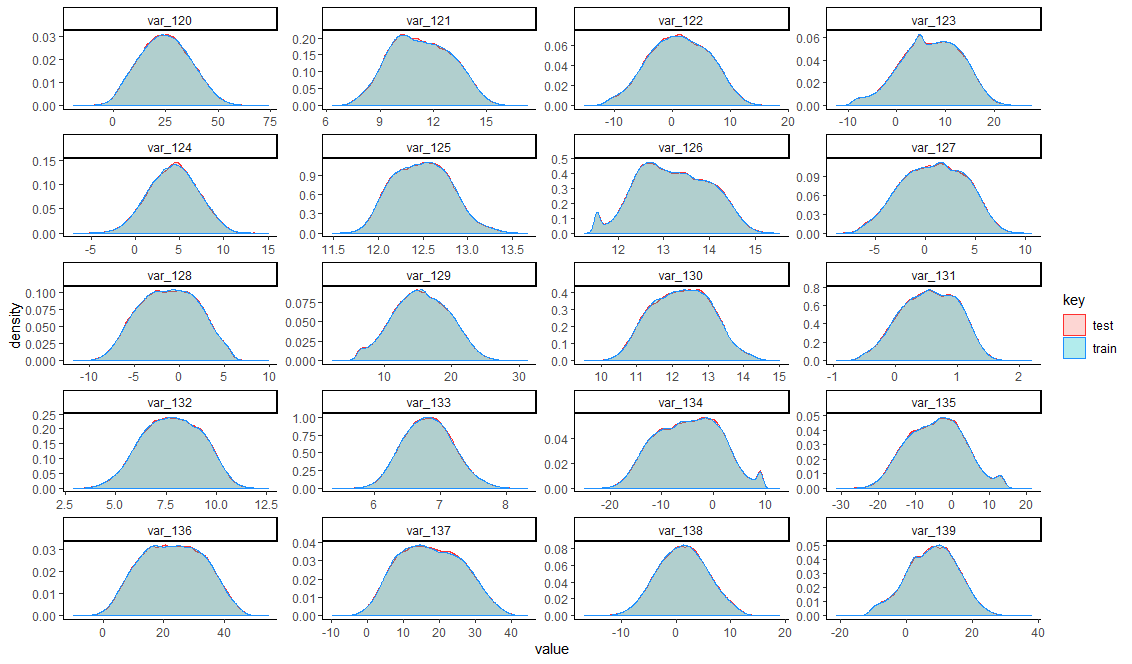
ggplot(data=temp, aes(x=value)) +

geom\_density(aes(fill=key, color=key), alpha=0.3) +

scale\_color\_manual(values = c("train" = "dodgerblue", "test"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

feature\_groups <- 1:20

col\_names <- colnames(full)[feature\_groups+140]

temp <- gather(full[,c("key",col\_names), with=F], key="features", value="value", -key)

temp$features <- factor(temp$features, levels=col\_names, labels=col\_names)

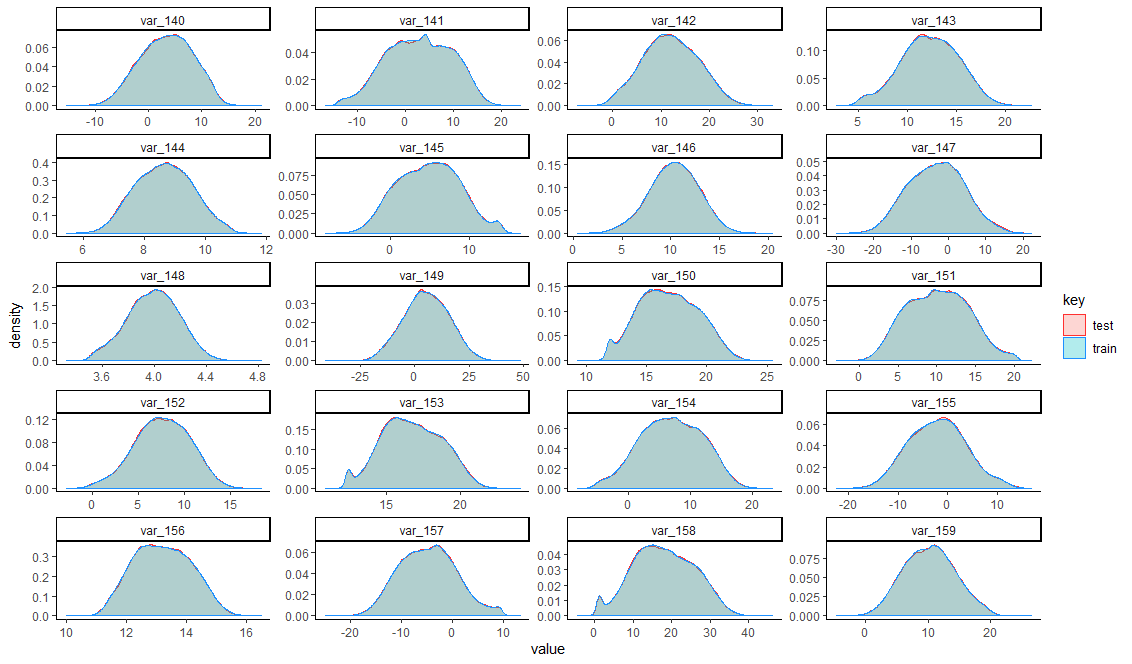
ggplot(data=temp, aes(x=value)) +

geom\_density(aes(fill=key, color=key), alpha=0.3) +

scale\_color\_manual(values = c("train" = "dodgerblue", "test"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

feature\_groups <- 1:20

col\_names <- colnames(full)[feature\_groups+160]

temp <- gather(full[,c("key",col\_names), with=F], key="features", value="value", -key)

temp$features <- factor(temp$features, levels=col\_names, labels=col\_names)

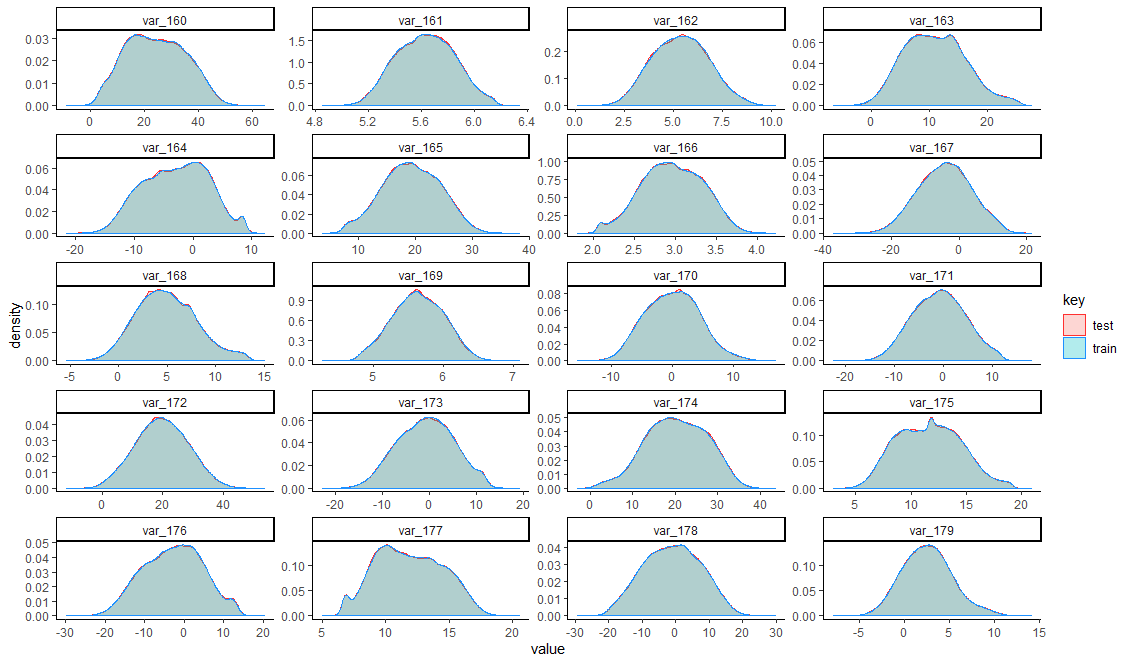
ggplot(data=temp, aes(x=value)) +

geom\_density(aes(fill=key, color=key), alpha=0.3) +

scale\_color\_manual(values = c("train" = "dodgerblue", "test"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

feature\_groups <- 1:20

col\_names <- colnames(full)[feature\_groups+180]

temp <- gather(full[,c("key",col\_names), with=F], key="features", value="value", -key)

temp$features <- factor(temp$features, levels=col\_names, labels=col\_names)

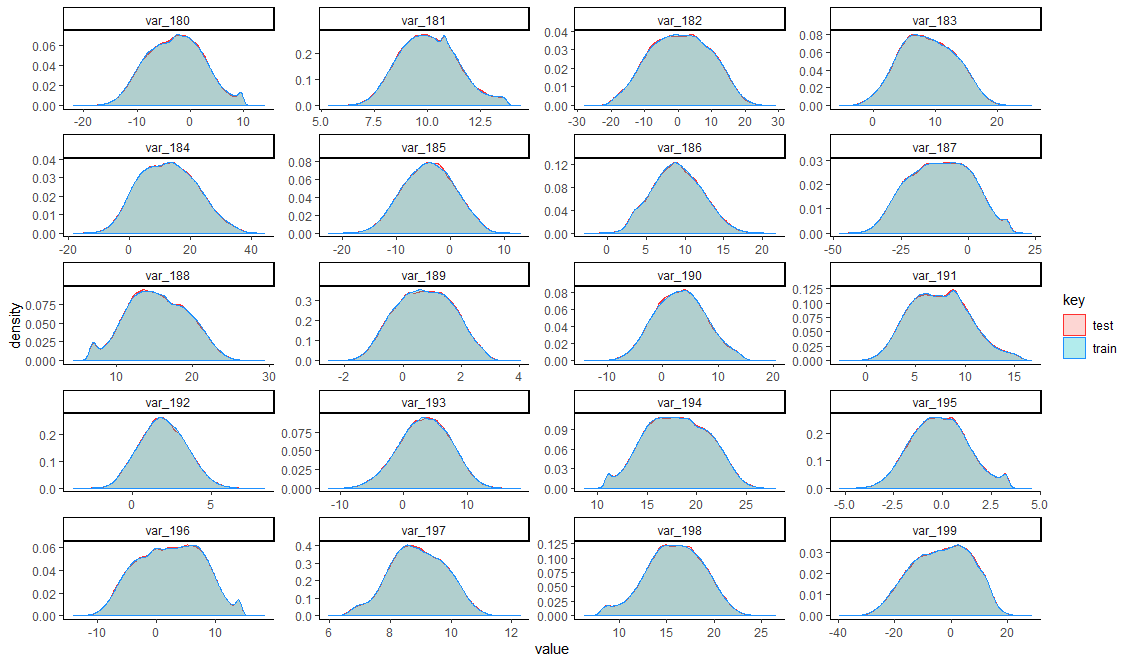
ggplot(data=temp, aes(x=value)) +

geom\_density(aes(fill=key, color=key), alpha=0.3) +

scale\_color\_manual(values = c("train" = "dodgerblue", "test"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

**Boxplots for train data**

feature\_groups <- 3:22

col\_names <- colnames(train)[c(2,feature\_groups)]

temp <- gather(train[,col\_names, with=F], key="features", value="value", -target)

temp$target <- factor(temp$target)

temp$features <- factor(temp$features, levels=col\_names[-1], labels=col\_names[-1])

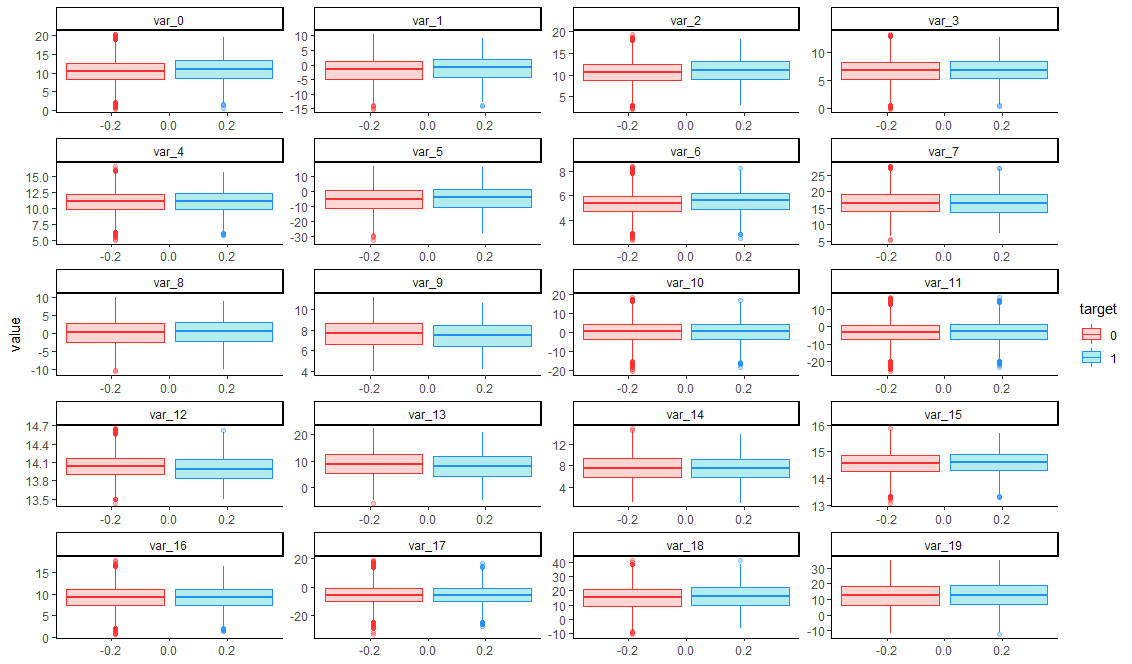
ggplot(data=temp, aes(y=value)) +

geom\_boxplot(aes(fill=target, color=target), alpha=0.3) +

scale\_color\_manual(values = c("1" = "dodgerblue", "0"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

col\_names <- colnames(train)[c(2,feature\_groups+20)]

temp <- gather(train[,col\_names, with=F], key="features", value="value", -target)

temp$target <- factor(temp$target)

temp$features <- factor(temp$features, levels=col\_names[-1], labels=col\_names[-1])

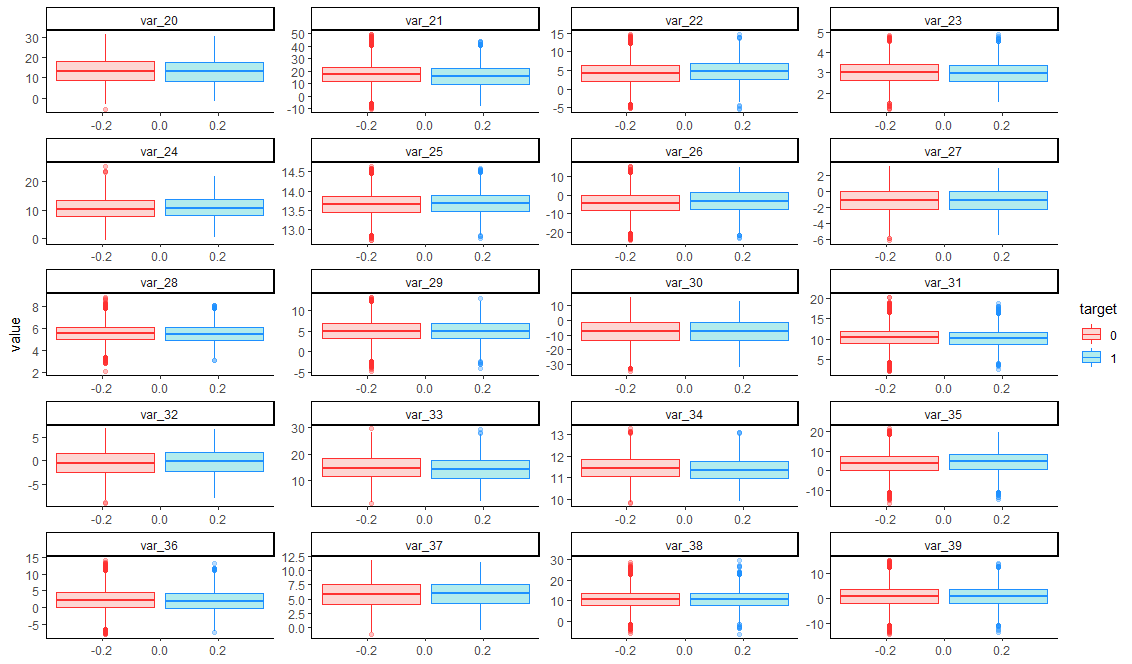
ggplot(data=temp, aes(y=value)) +

geom\_boxplot(aes(fill=target, color=target), alpha=0.3) +

scale\_color\_manual(values = c("1" = "dodgerblue", "0"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

col\_names <- colnames(train)[c(2,feature\_groups+40)]

temp <- gather(train[,col\_names, with=F], key="features", value="value", -target)

temp$target <- factor(temp$target)

temp$features <- factor(temp$features, levels=col\_names[-1], labels=col\_names[-1])

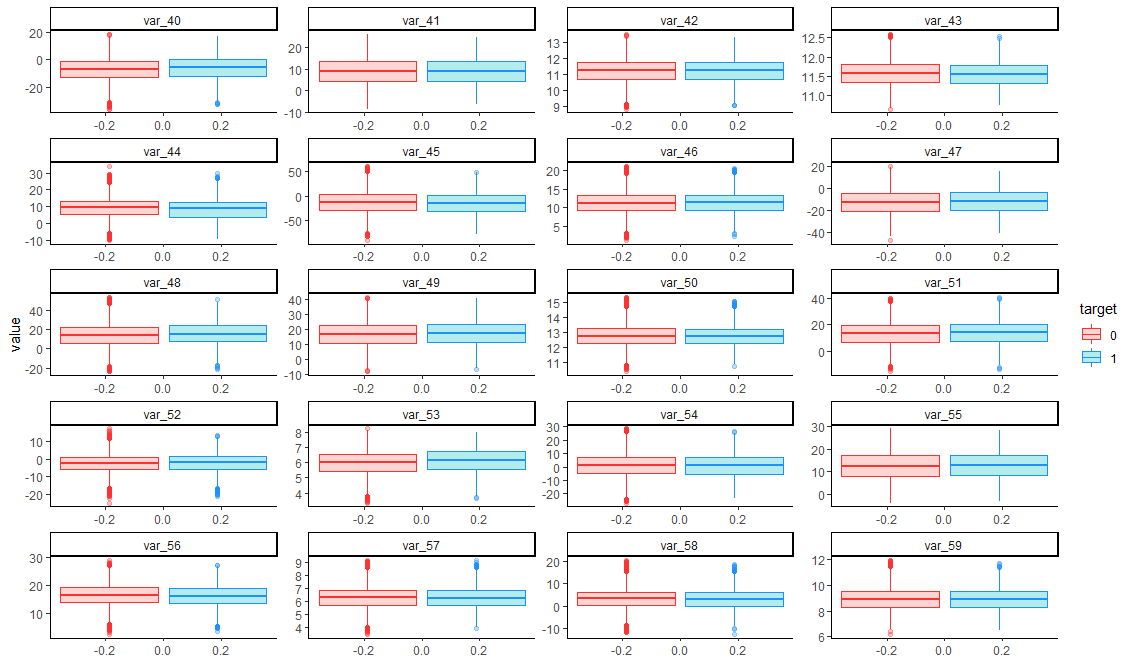
ggplot(data=temp, aes(y=value)) +

geom\_boxplot(aes(fill=target, color=target), alpha=0.3) +

scale\_color\_manual(values = c("1" = "dodgerblue", "0"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

col\_names <- colnames(train)[c(2,feature\_groups+60)]

temp <- gather(train[,col\_names, with=F], key="features", value="value", -target)

temp$target <- factor(temp$target)

temp$features <- factor(temp$features, levels=col\_names[-1], labels=col\_names[-1])

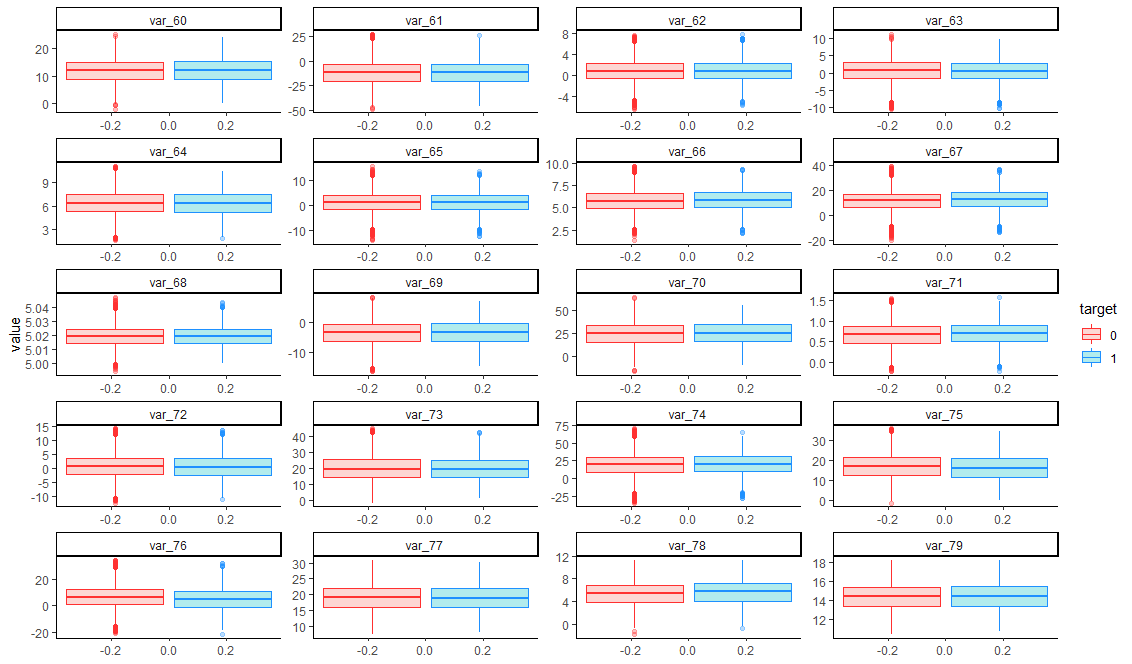
ggplot(data=temp, aes(y=value)) +

geom\_boxplot(aes(fill=target, color=target), alpha=0.3) +

scale\_color\_manual(values = c("1" = "dodgerblue", "0"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

col\_names <- colnames(train)[c(2,feature\_groups+80)]

temp <- gather(train[,col\_names, with=F], key="features", value="value", -target)

temp$target <- factor(temp$target)

temp$features <- factor(temp$features, levels=col\_names[-1], labels=col\_names[-1])

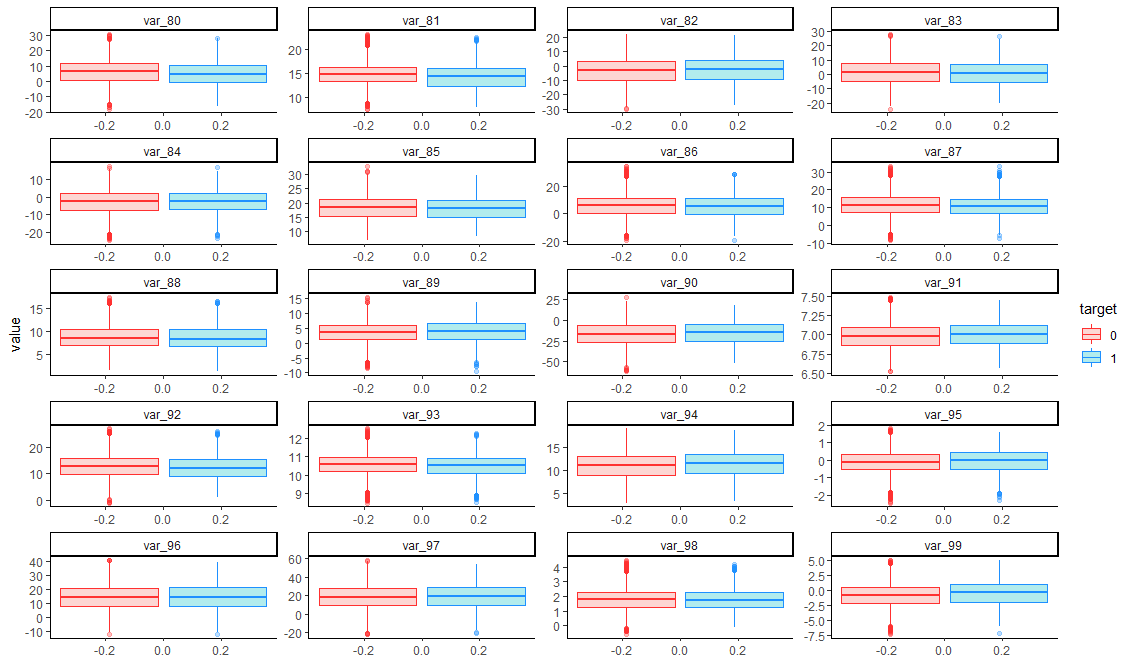
ggplot(data=temp, aes(y=value)) +

geom\_boxplot(aes(fill=target, color=target), alpha=0.3) +

scale\_color\_manual(values = c("1" = "dodgerblue", "0"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

col\_names <- colnames(train)[c(2,feature\_groups+100)]

temp <- gather(train[,col\_names, with=F], key="features", value="value", -target)

temp$target <- factor(temp$target)

temp$features <- factor(temp$features, levels=col\_names[-1], labels=col\_names[-1])

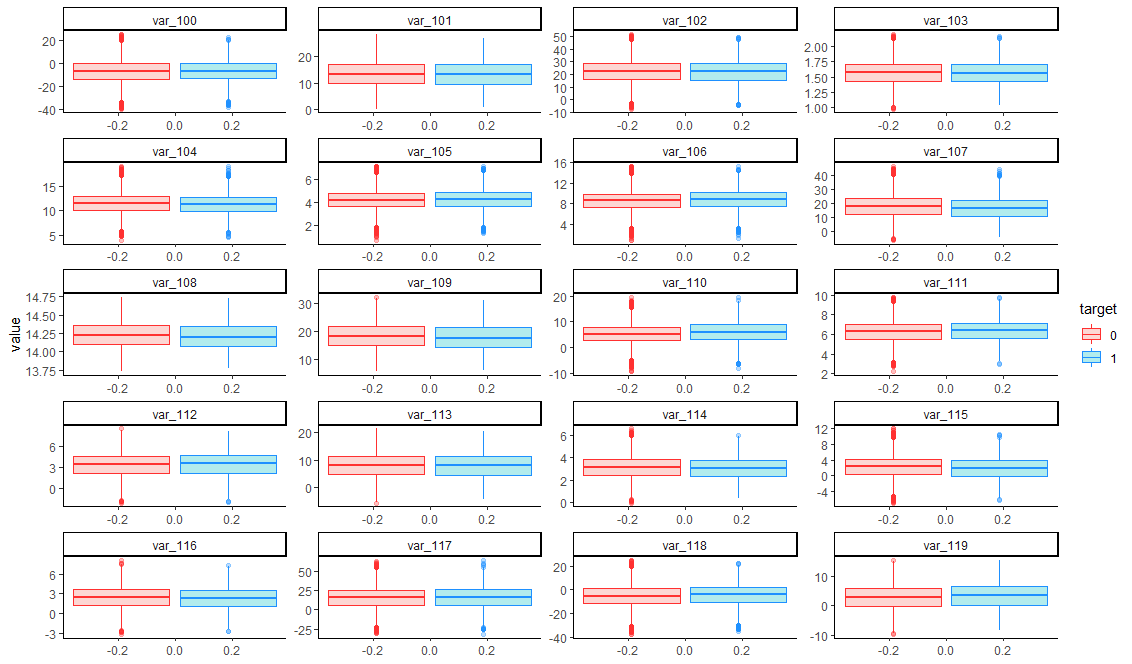
ggplot(data=temp, aes(y=value)) +

geom\_boxplot(aes(fill=target, color=target), alpha=0.3) +

scale\_color\_manual(values = c("1" = "dodgerblue", "0"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

col\_names <- colnames(train)[c(2,feature\_groups+120)]

temp <- gather(train[,col\_names, with=F], key="features", value="value", -target)

temp$target <- factor(temp$target)

temp$features <- factor(temp$features, levels=col\_names[-1], labels=col\_names[-1])

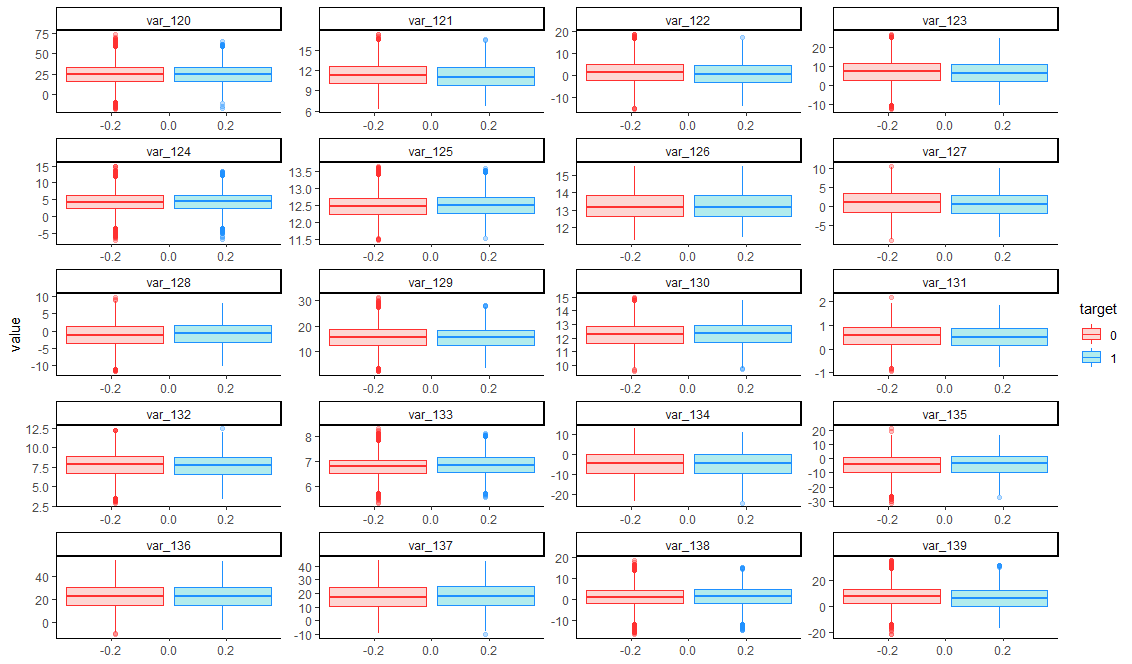
ggplot(data=temp, aes(y=value)) +

geom\_boxplot(aes(fill=target, color=target), alpha=0.3) +

scale\_color\_manual(values = c("1" = "dodgerblue", "0"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

col\_names <- colnames(train)[c(2,feature\_groups+140)]

temp <- gather(train[,col\_names, with=F], key="features", value="value", -target)

temp$target <- factor(temp$target)

temp$features <- factor(temp$features, levels=col\_names[-1], labels=col\_names[-1])

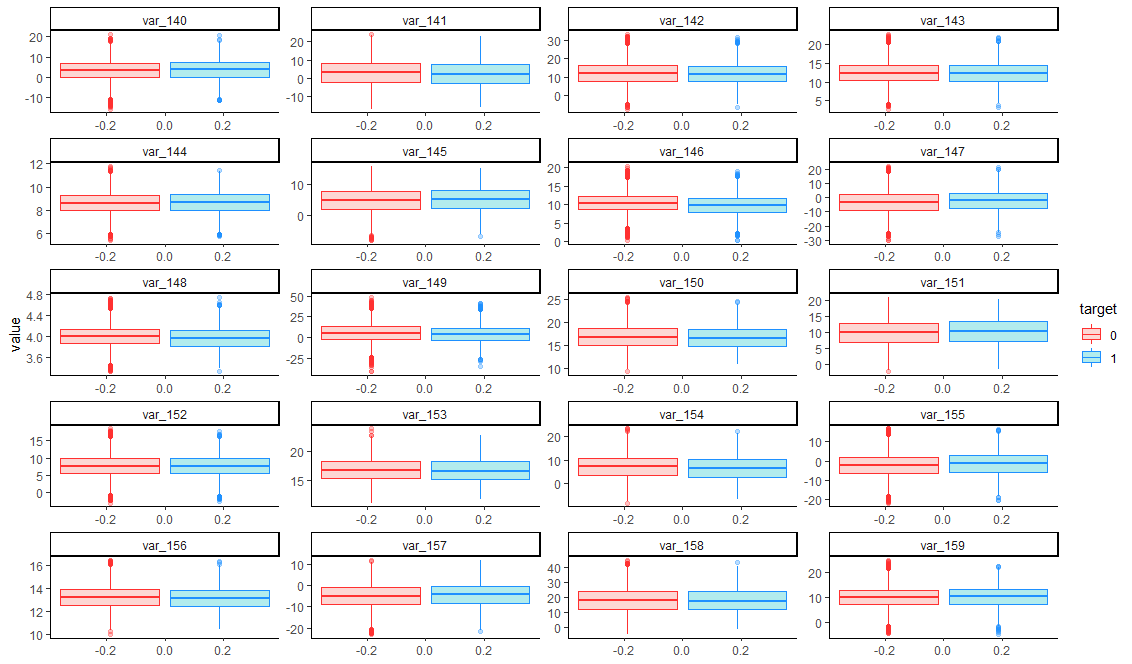
ggplot(data=temp, aes(y=value)) +

geom\_boxplot(aes(fill=target, color=target), alpha=0.3) +

scale\_color\_manual(values = c("1" = "dodgerblue", "0"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

col\_names <- colnames(train)[c(2,feature\_groups+160)]

temp <- gather(train[,col\_names, with=F], key="features", value="value", -target)

temp$target <- factor(temp$target)

temp$features <- factor(temp$features, levels=col\_names[-1], labels=col\_names[-1])

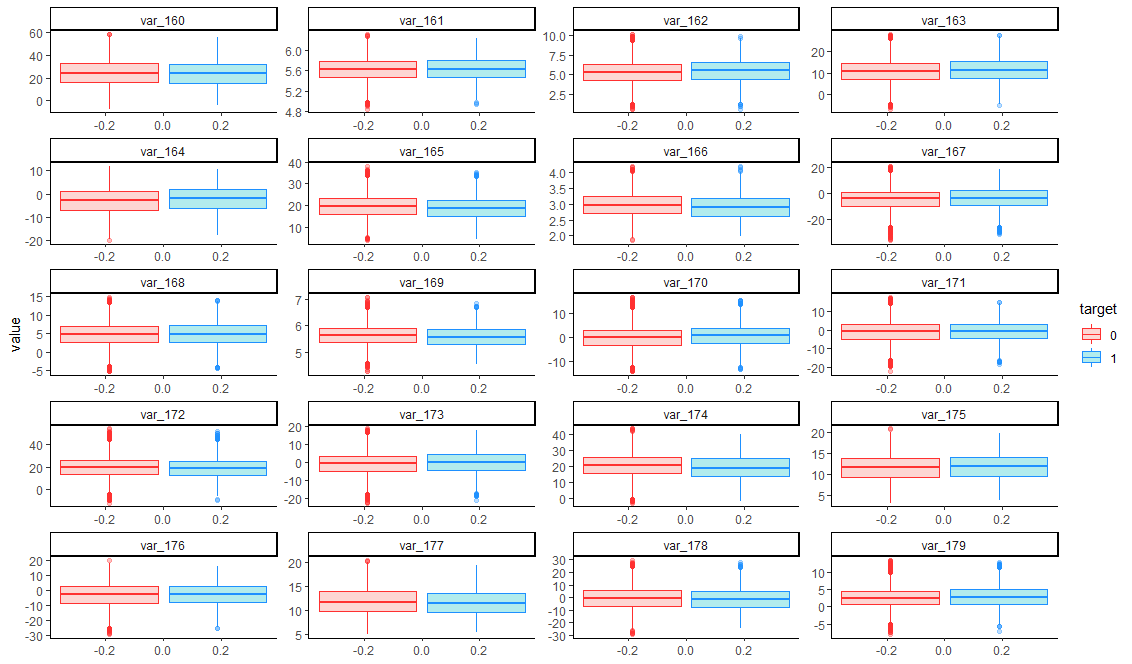
ggplot(data=temp, aes(y=value)) +

geom\_boxplot(aes(fill=target, color=target), alpha=0.3) +

scale\_color\_manual(values = c("1" = "dodgerblue", "0"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

col\_names <- colnames(train)[c(2,feature\_groups+180)]

temp <- gather(train[,col\_names, with=F], key="features", value="value", -target)

temp$target <- factor(temp$target)

temp$features <- factor(temp$features, levels=col\_names[-1], labels=col\_names[-1])

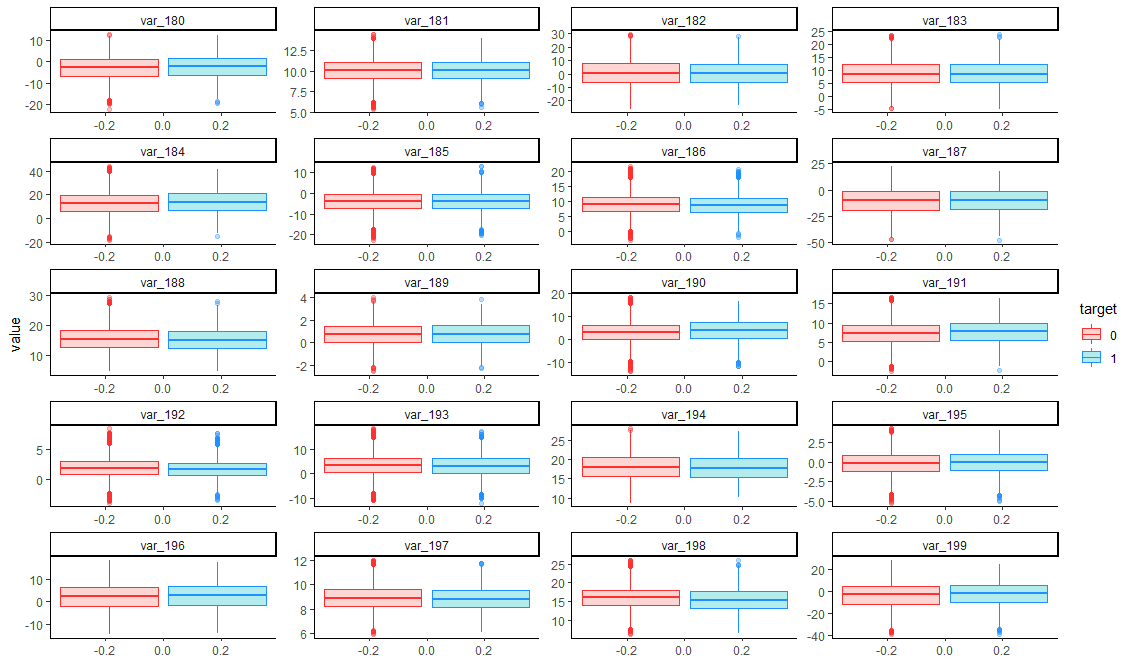
ggplot(data=temp, aes(y=value)) +

geom\_boxplot(aes(fill=target, color=target), alpha=0.3) +

scale\_color\_manual(values = c("1" = "dodgerblue", "0"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

**Compare boxplots train and test**

feature\_groups <- 1:20

col\_names <- colnames(full)[feature\_groups]

temp <- gather(full[,c("key",col\_names), with=F], key="features", value="value", -key)

temp$features <- factor(temp$features, levels=col\_names, labels=col\_names)

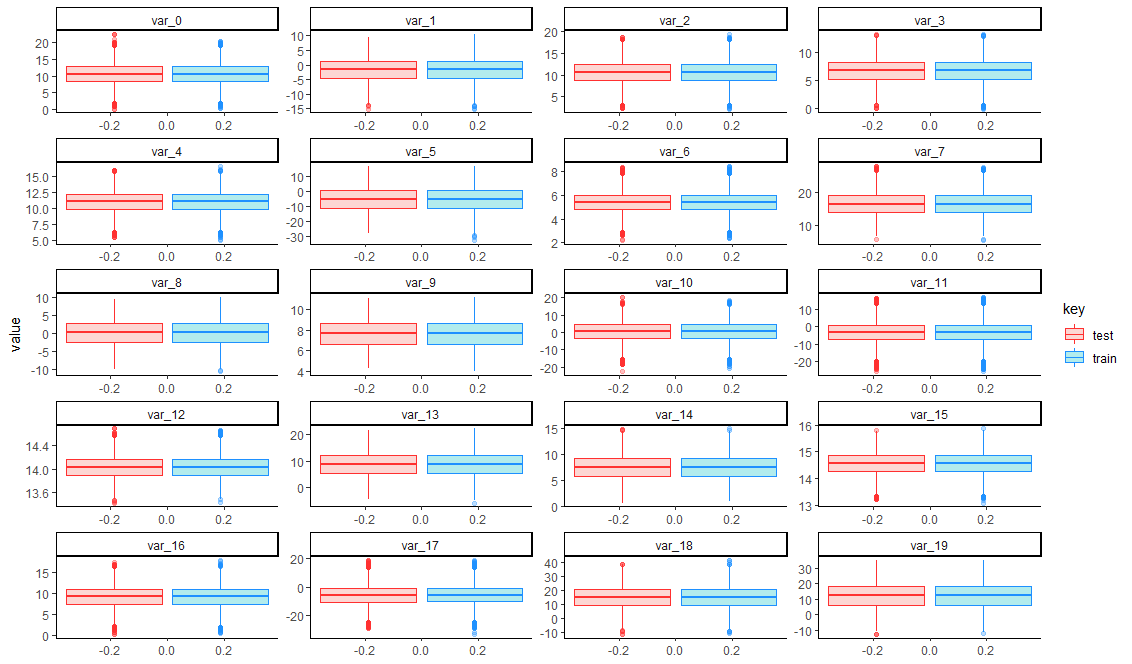
ggplot(data=temp, aes(y=value)) +

geom\_boxplot(aes(fill=key, color=key), alpha=0.3) +

scale\_color\_manual(values = c("train" = "dodgerblue", "test"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

feature\_groups <- 1:20

col\_names <- colnames(full)[feature\_groups+20]

temp <- gather(full[,c("key",col\_names), with=F], key="features", value="value", -key)

temp$features <- factor(temp$features, levels=col\_names, labels=col\_names)

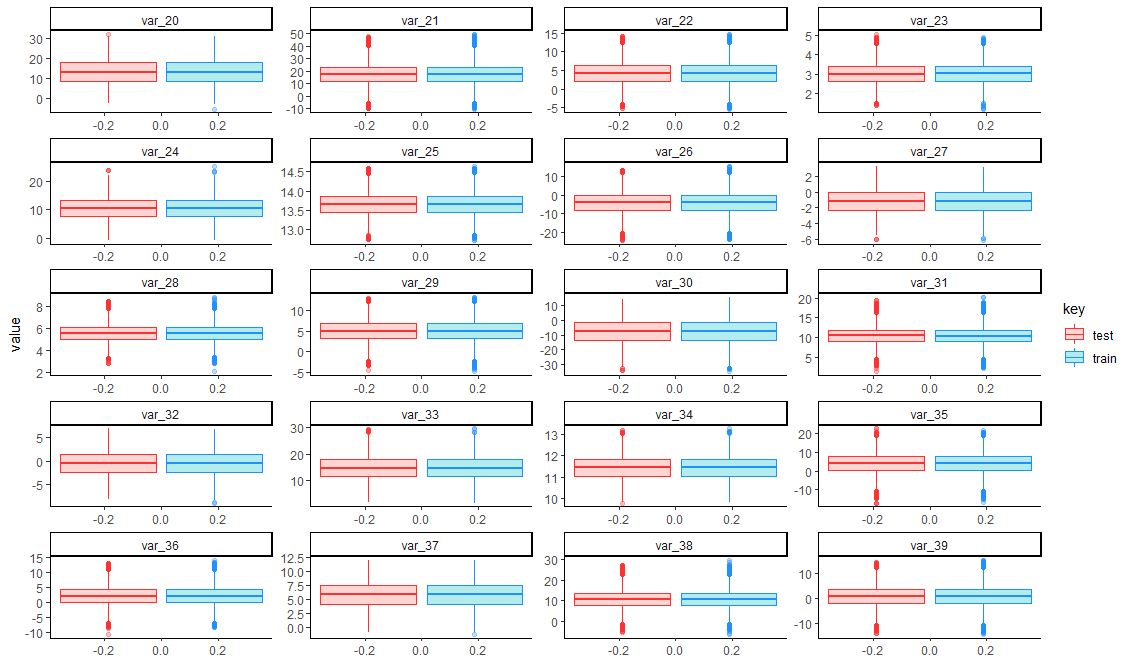
ggplot(data=temp, aes(y=value)) +

geom\_boxplot(aes(fill=key, color=key), alpha=0.3) +

scale\_color\_manual(values = c("train" = "dodgerblue", "test"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

feature\_groups <- 1:20

col\_names <- colnames(full)[feature\_groups+40]

temp <- gather(full[,c("key",col\_names), with=F], key="features", value="value", -key)

temp$features <- factor(temp$features, levels=col\_names, labels=col\_names)

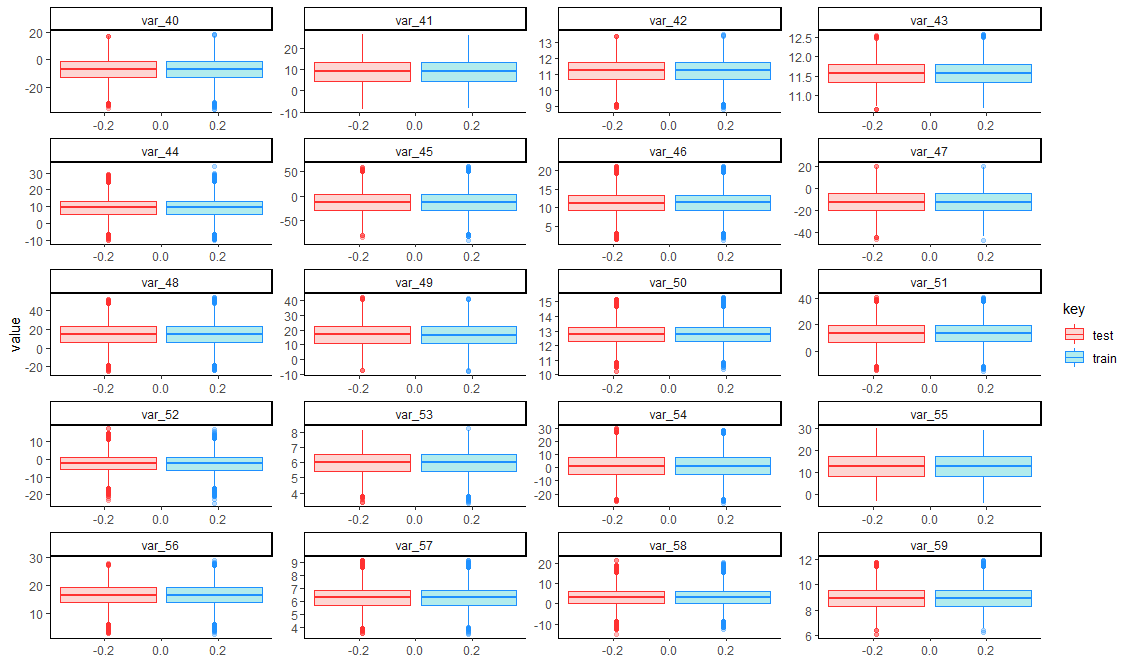
ggplot(data=temp, aes(y=value)) +

geom\_boxplot(aes(fill=key, color=key), alpha=0.3) +

scale\_color\_manual(values = c("train" = "dodgerblue", "test"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

feature\_groups <- 1:20

col\_names <- colnames(full)[feature\_groups+60]

temp <- gather(full[,c("key",col\_names), with=F], key="features", value="value", -key)

temp$features <- factor(temp$features, levels=col\_names, labels=col\_names)

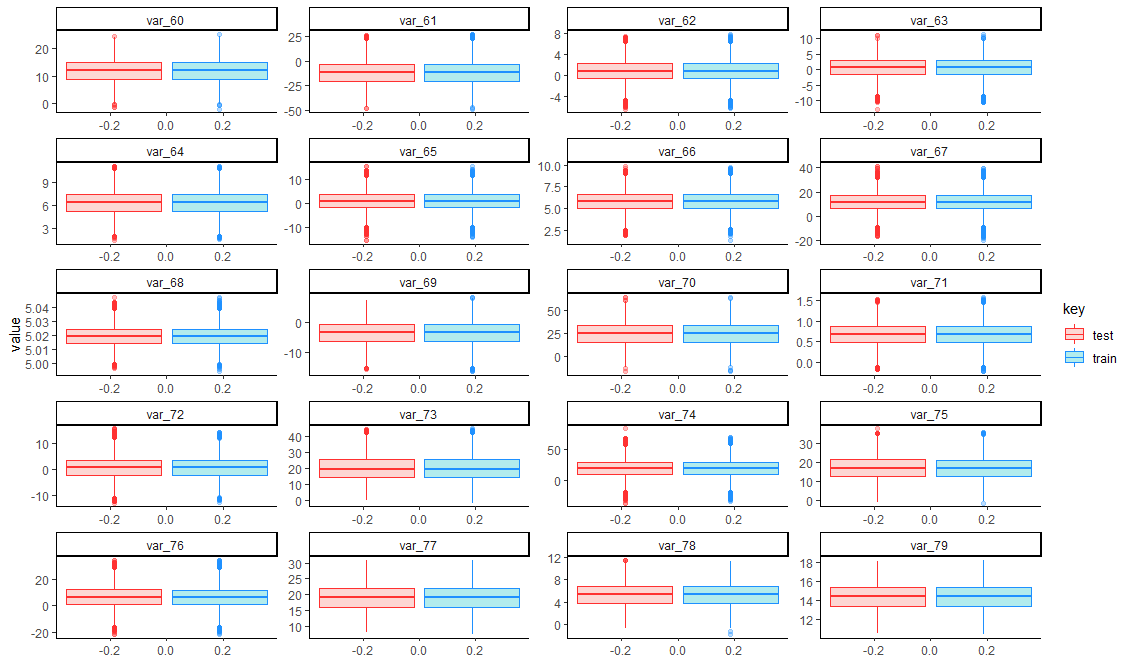
ggplot(data=temp, aes(y=value)) +

geom\_boxplot(aes(fill=key, color=key), alpha=0.3) +

scale\_color\_manual(values = c("train" = "dodgerblue", "test"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

feature\_groups <- 1:20

col\_names <- colnames(full)[feature\_groups+80]

temp <- gather(full[,c("key",col\_names), with=F], key="features", value="value", -key)

temp$features <- factor(temp$features, levels=col\_names, labels=col\_names)

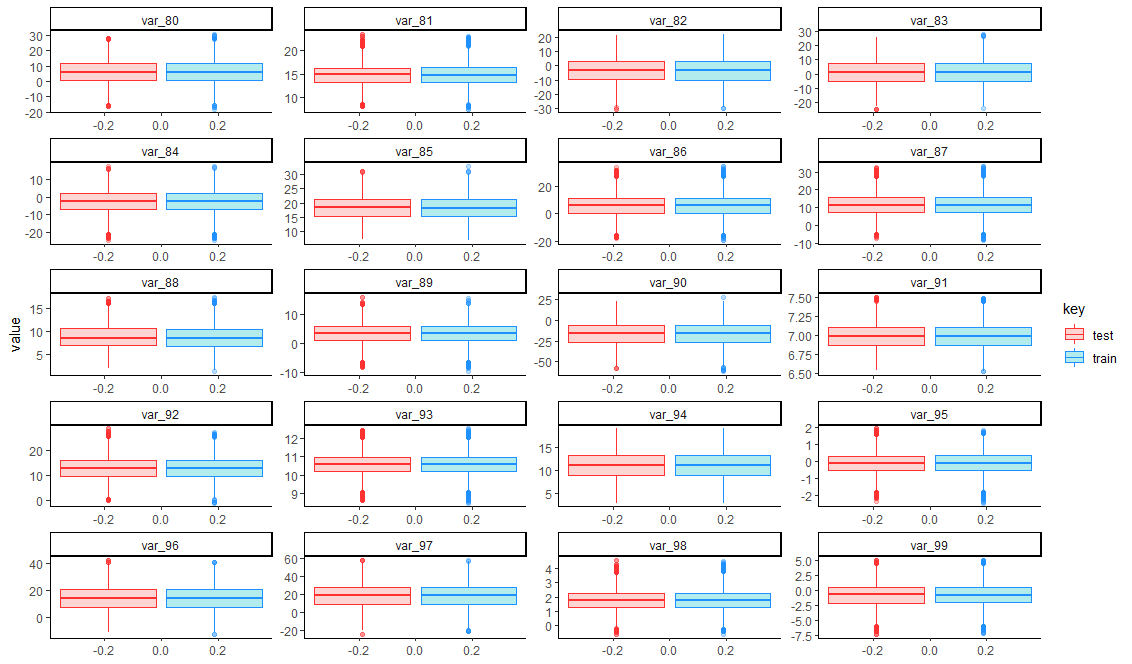
ggplot(data=temp, aes(y=value)) +

geom\_boxplot(aes(fill=key, color=key), alpha=0.3) +

scale\_color\_manual(values = c("train" = "dodgerblue", "test"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

feature\_groups <- 1:20

col\_names <- colnames(full)[feature\_groups+100]

temp <- gather(full[,c("key",col\_names), with=F], key="features", value="value", -key)

temp$features <- factor(temp$features, levels=col\_names, labels=col\_names)

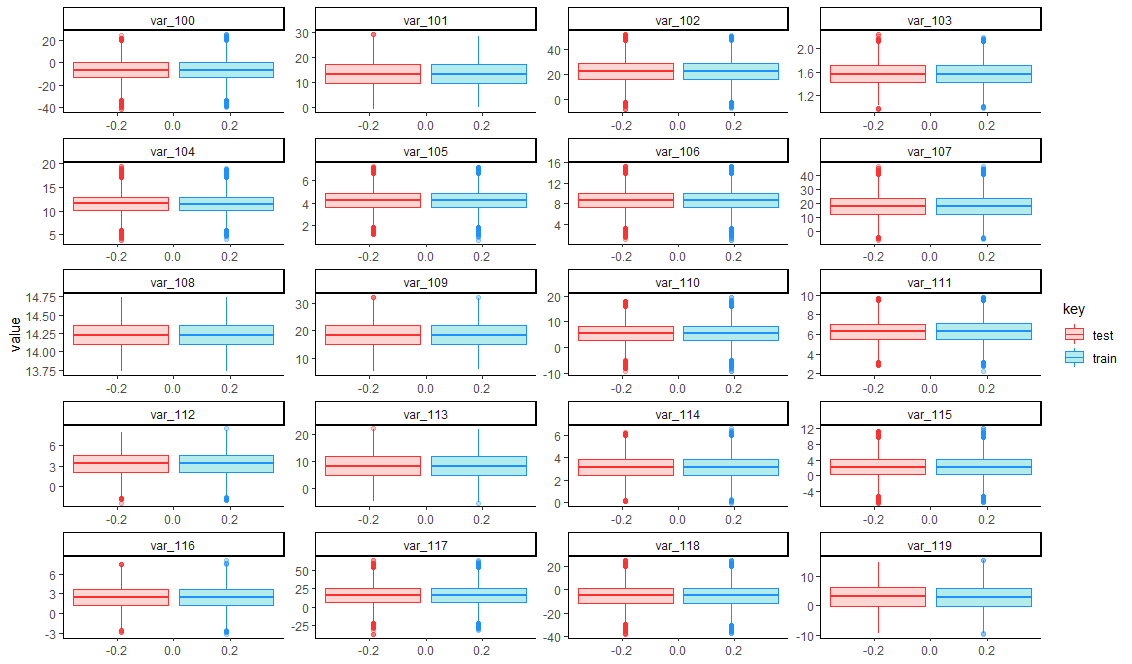
ggplot(data=temp, aes(y=value)) +

geom\_boxplot(aes(fill=key, color=key), alpha=0.3) +

scale\_color\_manual(values = c("train" = "dodgerblue", "test"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

feature\_groups <- 1:20

col\_names <- colnames(full)[feature\_groups+120]

temp <- gather(full[,c("key",col\_names), with=F], key="features", value="value", -key)

temp$features <- factor(temp$features, levels=col\_names, labels=col\_names)

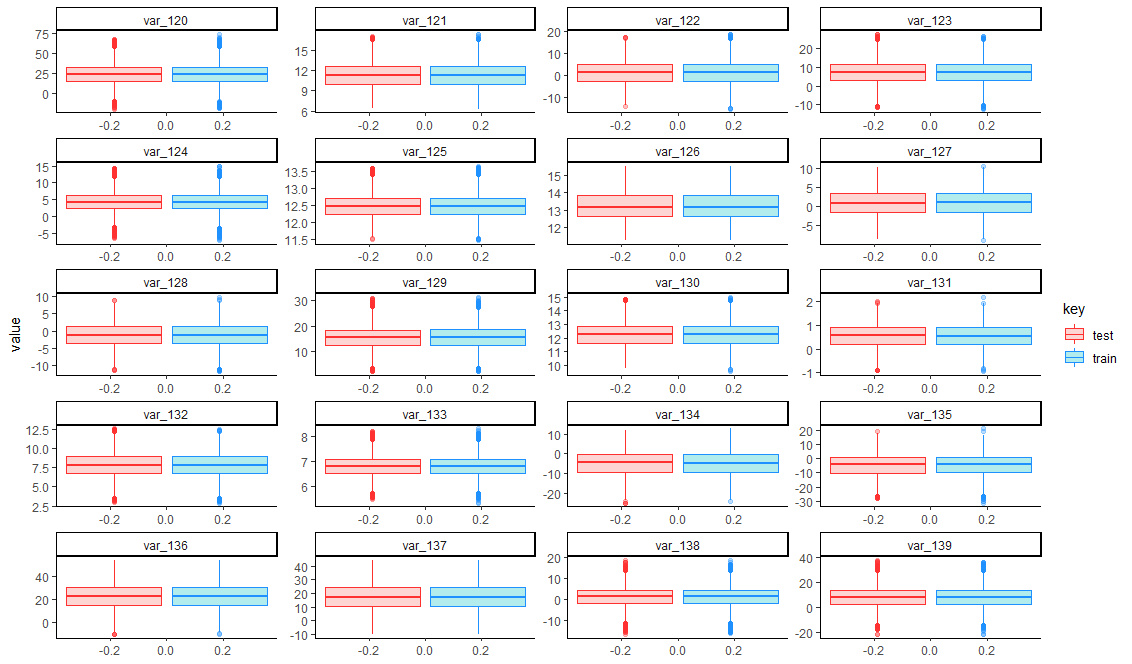
ggplot(data=temp, aes(y=value)) +

geom\_boxplot(aes(fill=key, color=key), alpha=0.3) +

scale\_color\_manual(values = c("train" = "dodgerblue", "test"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

feature\_groups <- 1:20

col\_names <- colnames(full)[feature\_groups+140]

temp <- gather(full[,c("key",col\_names), with=F], key="features", value="value", -key)

temp$features <- factor(temp$features, levels=col\_names, labels=col\_names)

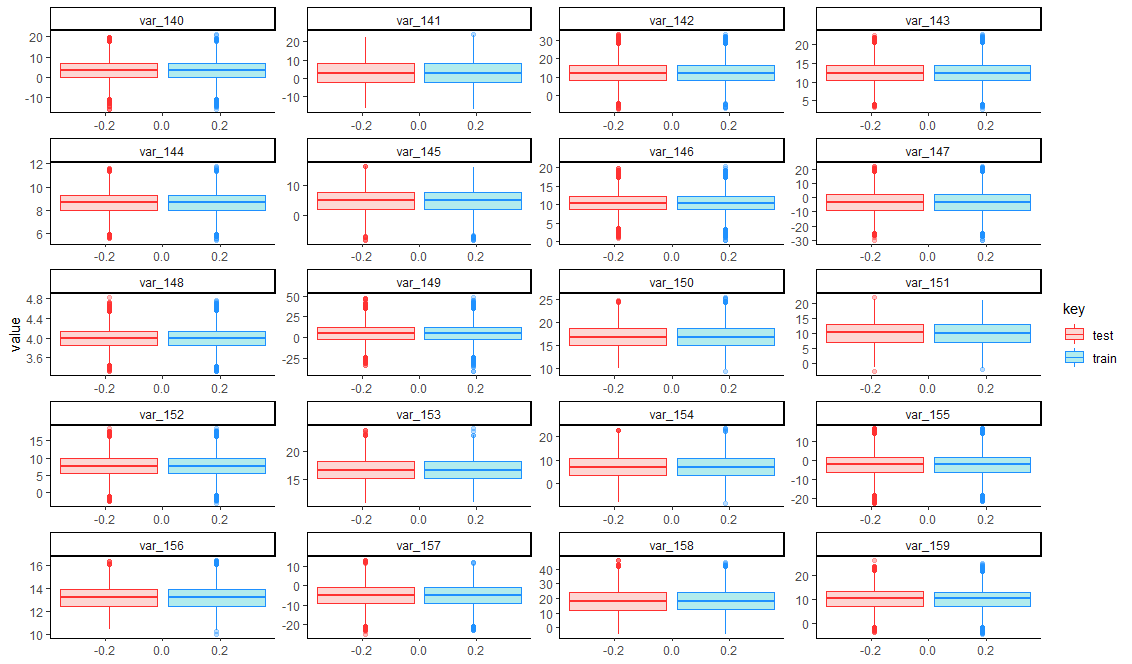
ggplot(data=temp, aes(y=value)) +

geom\_boxplot(aes(fill=key, color=key), alpha=0.3) +

scale\_color\_manual(values = c("train" = "dodgerblue", "test"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

feature\_groups <- 1:20

col\_names <- colnames(full)[feature\_groups+160]

temp <- gather(full[,c("key",col\_names), with=F], key="features", value="value", -key)

temp$features <- factor(temp$features, levels=col\_names, labels=col\_names)

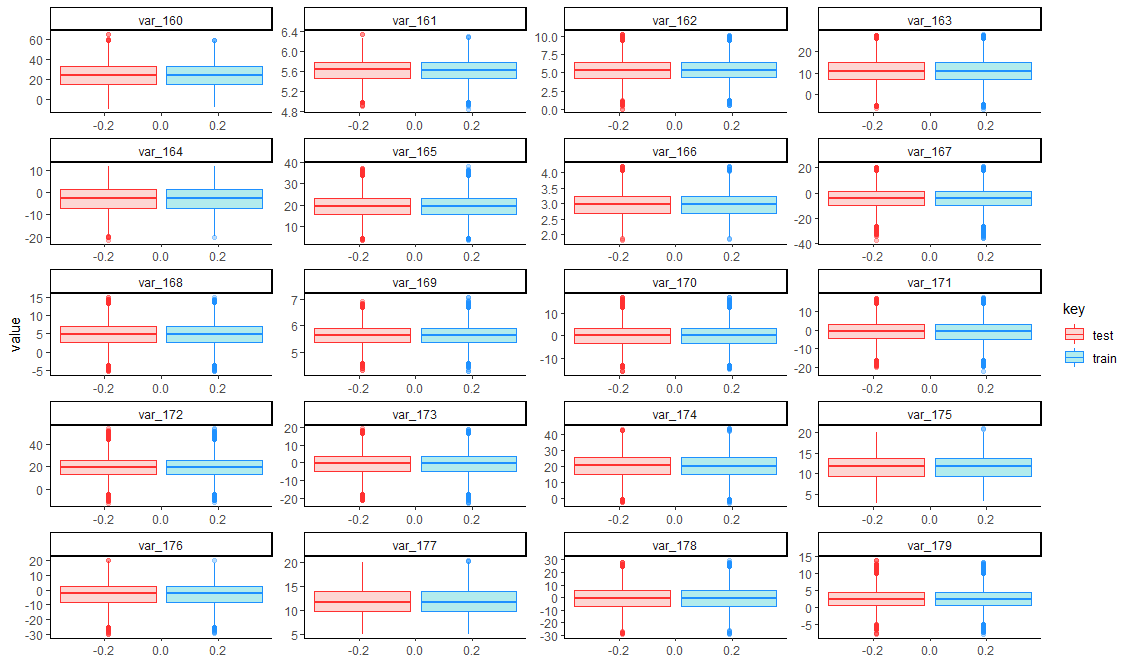
ggplot(data=temp, aes(y=value)) +

geom\_boxplot(aes(fill=key, color=key), alpha=0.3) +

scale\_color\_manual(values = c("train" = "dodgerblue", "test"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

feature\_groups <- 1:20

col\_names <- colnames(full)[feature\_groups+180]

temp <- gather(full[,c("key",col\_names), with=F], key="features", value="value", -key)

temp$features <- factor(temp$features, levels=col\_names, labels=col\_names)

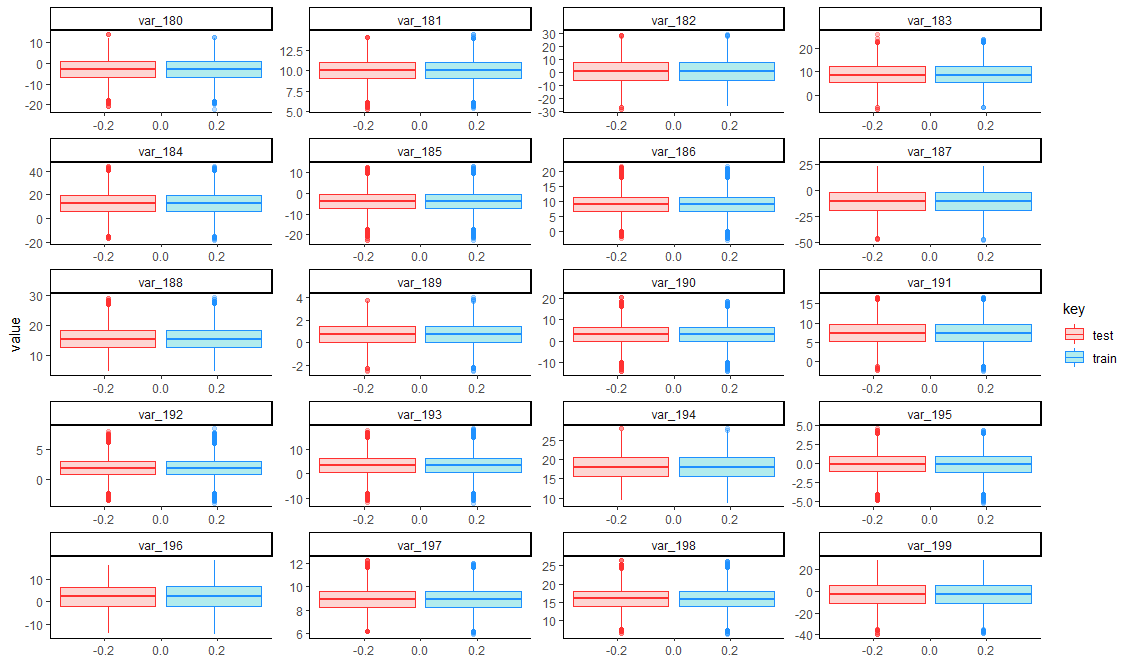
ggplot(data=temp, aes(y=value)) +

geom\_boxplot(aes(fill=key, color=key), alpha=0.3) +

scale\_color\_manual(values = c("train" = "dodgerblue", "test"="firebrick1")) +

theme\_classic() +

facet\_wrap(~ features, ncol = 4, scales = "free")

****

train <- train %>% dplyr::select(-key)

test <- test %>% dplyr::select(-key)

**Correlations**

cormat <- cor(train[,-c(1,2)])

summary(cormat[upper.tri(cormat)])

Min. 1st Qu. Median Mean 3rd Qu. Max.

-9.844e-03 -1.640e-03 1.808e-05 2.781e-05 1.723e-03 9.714e-03

* Correlation between features nearly zero

**Modeling**

X\_train <- scale(train[,-(1:2)]) %>% data.frame

X\_test <- scale(test[,-1]) %>% data.frame

target <- train$target

**Logistic Regression**

**Glm vs speedglm**

start.time <- Sys.time()

fit.logit <- glm(target~., data=X\_train, family=binomial)

end.time <- Sys.time()

end.time - start.time

Time difference of 1.808374 mins

start.time <- Sys.time()

fit.logit <- speedglm(target~., data=X\_train, family=binomial())

end.time <- Sys.time()

end.time - start.time

Time difference of 1.096498 mins

* Speedglm function is almost equal to glm function

pred.logit <- predict(fit.logit, newdata=X\_test, type="response")

submission <- read.csv("C:/Users/DELL/Desktop/input/sample\_submission.csv")

submission$target <- pred.logit

write.csv(submission, file="submission\_logit.csv", row.names=F)