

탐색적 자료분석 및 시각화

등급에 따른 Skillcraft data와
winequality(red, white) data의 탐색적 자료분석

비즈니스 애널리틱스

김다영

2017.08.31

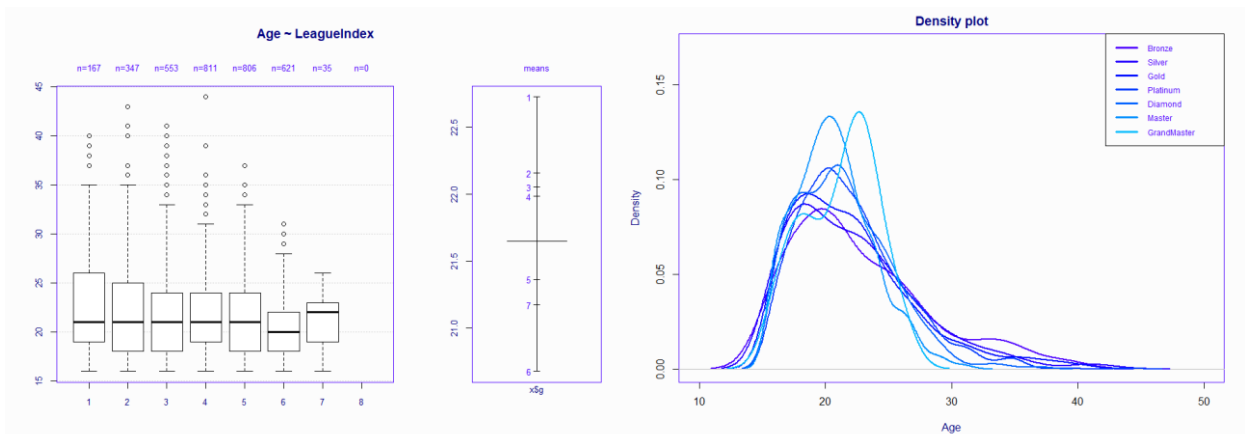
Skillcraft Data

식별자	"GameID"	
종속변수	"LeagueIndex"	'Bronze', 'Silver', 'Gold', 'Platinum', 'Diamond', 'Master', 'GrandMaster', 'Professional' 총 8개의 League로 범주화 해준다.
설명변수(18개)	"Age", "HoursPerWeek", "TotalHours", "APM", "SelectByHotkeys", "AssignToHotkeys", "UniqueHotkeys", "MinimapAttacks", "MinimapRightClicks", "NumberOfPACs", "GapBetweenPACs", "ActionLatency", "ActionsInPAC", "TotalMapExplored", "WorkersMade", "UniqueUnitsMade", "ComplexUnitsMade", "ComplexAbilitiesUsed"	"Age", "HoursPerWeek", "TotalHours"는 각각 55, 56, 57개의 결측값 '?'가 있어서 NA 처리 해준다.

A. 등급별로 각 변수에 대한 특징 파악

1. "Age"

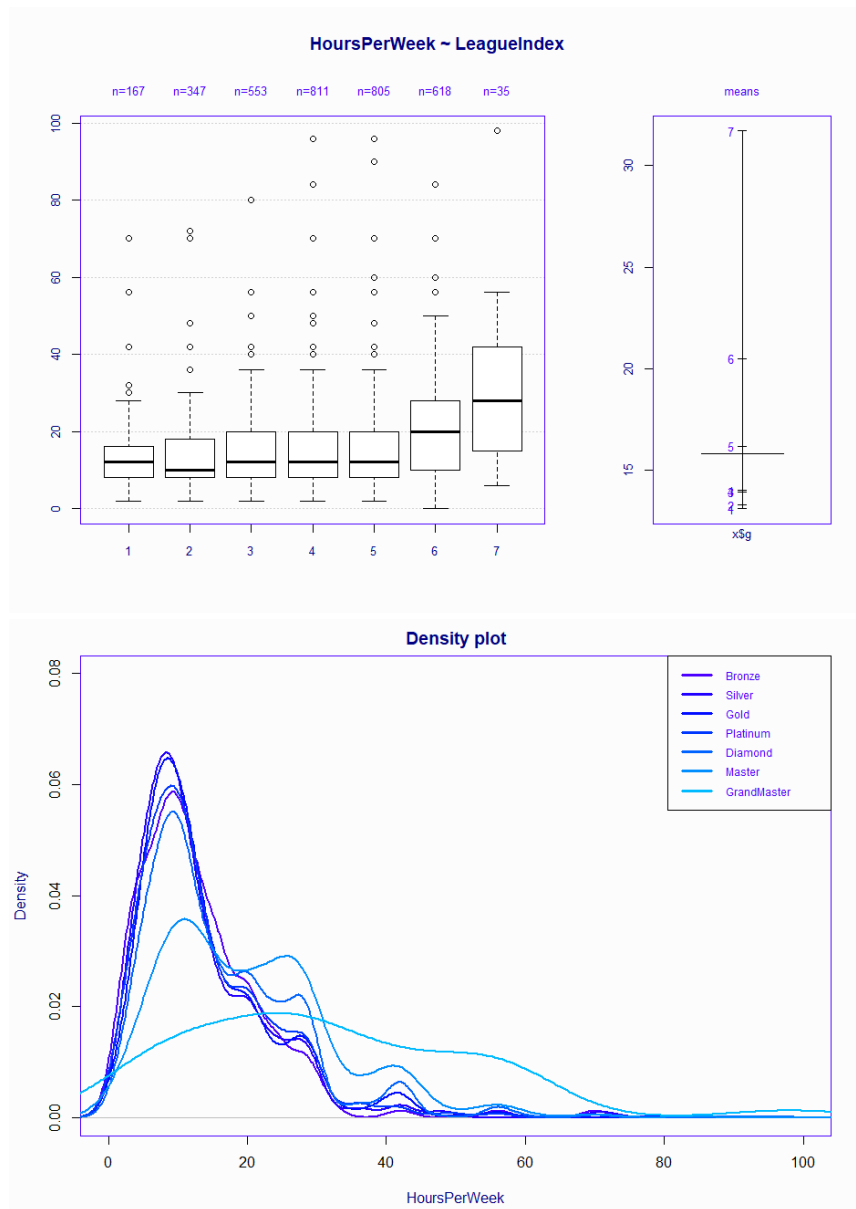
LeagueIndex	min	first_Qu	mean	median	third_Qu	max	sd	skewness	kurtosis
1	16	19	22.72455	21	26	40	5.52286	0.9813398	0.42198133
2	16	18	22.15562	21	25	43	5.091531	1.1309183	1.45830579
3	16	18	22.05063	21	24	41	4.901305	1.1869667	1.47957355
4	16	19	21.9815	21	24	44	4.141736	1.0016429	1.45699947
5	16	18	21.36228	21	24	37	3.662164	0.7382248	0.67978961
6	16	18	20.67794	20	22	31	3.030381	0.538481	0.02717121
7	16	19	21.17143	22	23	26	2.864444	-0.2329773	-1.0994761
8	Inf	NA	NaN	NA	NA	-Inf	NaN	NaN	NaN



등급별로 Age에 대한 boxplot을 보면 1~5등급('Bronze', 'Silver', 'Gold', 'Platinum', 'Diamond', 'Master')까지는 15~35세까지 넓게 분포했지만 6, 7등급('GrandMaster', 'Professional')은 15~22세까지 비교적 좁은 구간에 몰려있다. 가운데 그래프는 각 등급별로 평균을 표시한 그래프이다. 등급별로 Age에 대한 히스토그램을 기반으로 그린 density plot을 보면 각 등급에 따라 크게 다르게 분포되어있지 않다.

2. "HoursPerWeek"

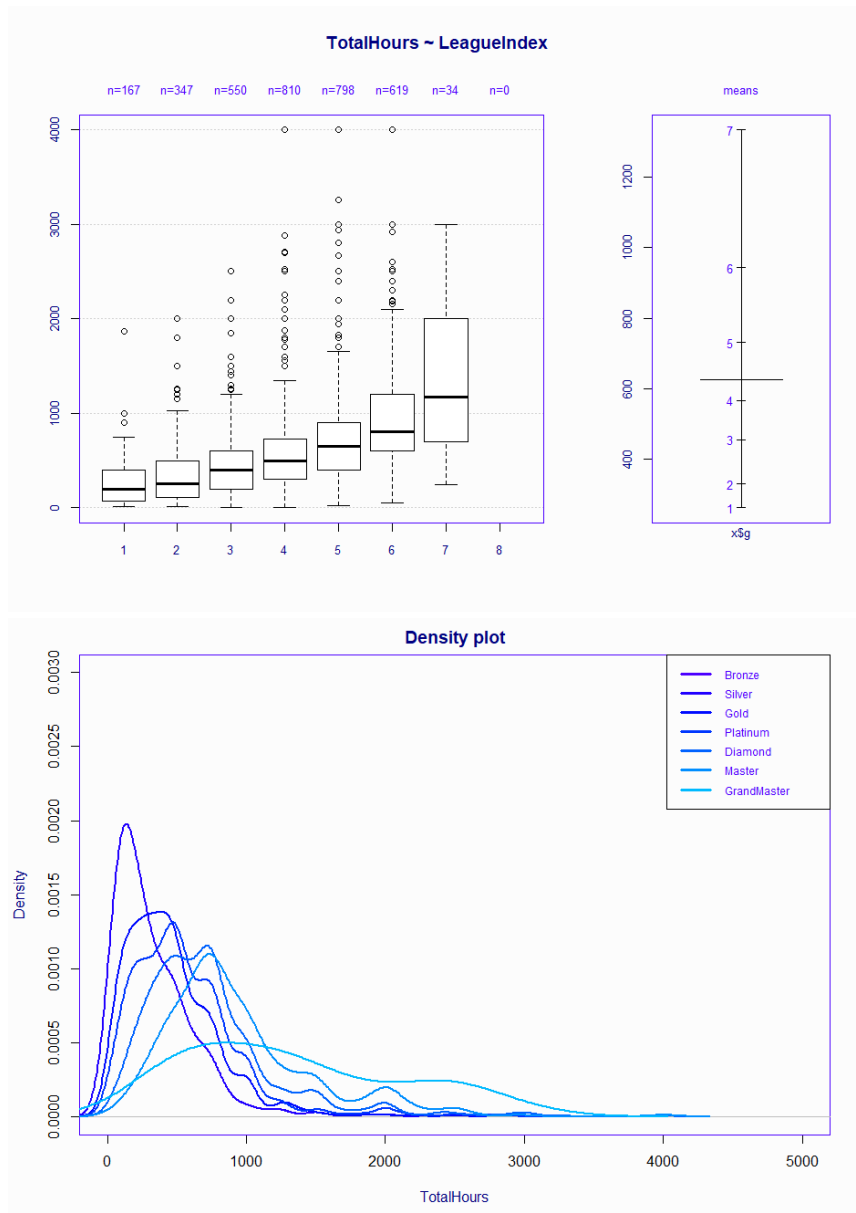
LeagueIndex	min	first_Qu	mean	median	third_Qu	max	sd	skewness	kurtosis
1	2	8	13.12575	12	16	70	9.405317	2.2229039	8.9966959
2	2	8	13.29683	10	18	72	9.59474	2.1033987	7.4834187
3	2	8	13.94937	12	20	80	9.744432	1.8761453	5.6107715
4	2	8	14.02219	12	20	96	10.205518	2.3073793	10.2939317
5	2	8	16.18385	12	20	96	11.370556	1.9236813	6.7366405
6	0	10	21.08857	20	28	168	15.12529	3.1608852	21.7684982
7	6	15	31.71429	28	42	98	20.626327	0.8968196	0.8811706
8	Inf	NA	NaN	NA	NA	-Inf	NaN	NaN	NaN



등급별로 HoursPerWeek에 대한 boxplot을 보면 1-7등급으로 갈수록 분포가 넓어지는 것을 볼 수 있다. 등급별로 HoursPerWeek에 대한 density plot을 보면 7등급이 눈에 띄게 다른 분포를 하고 있음을 확인할 수 있다. (등급별 boxplot의 차이를 확실하게 보기 위해 boxplot은 100이상의 값을 가진 이상치를 제외하고 그렸다.)

3. "TotalHours"

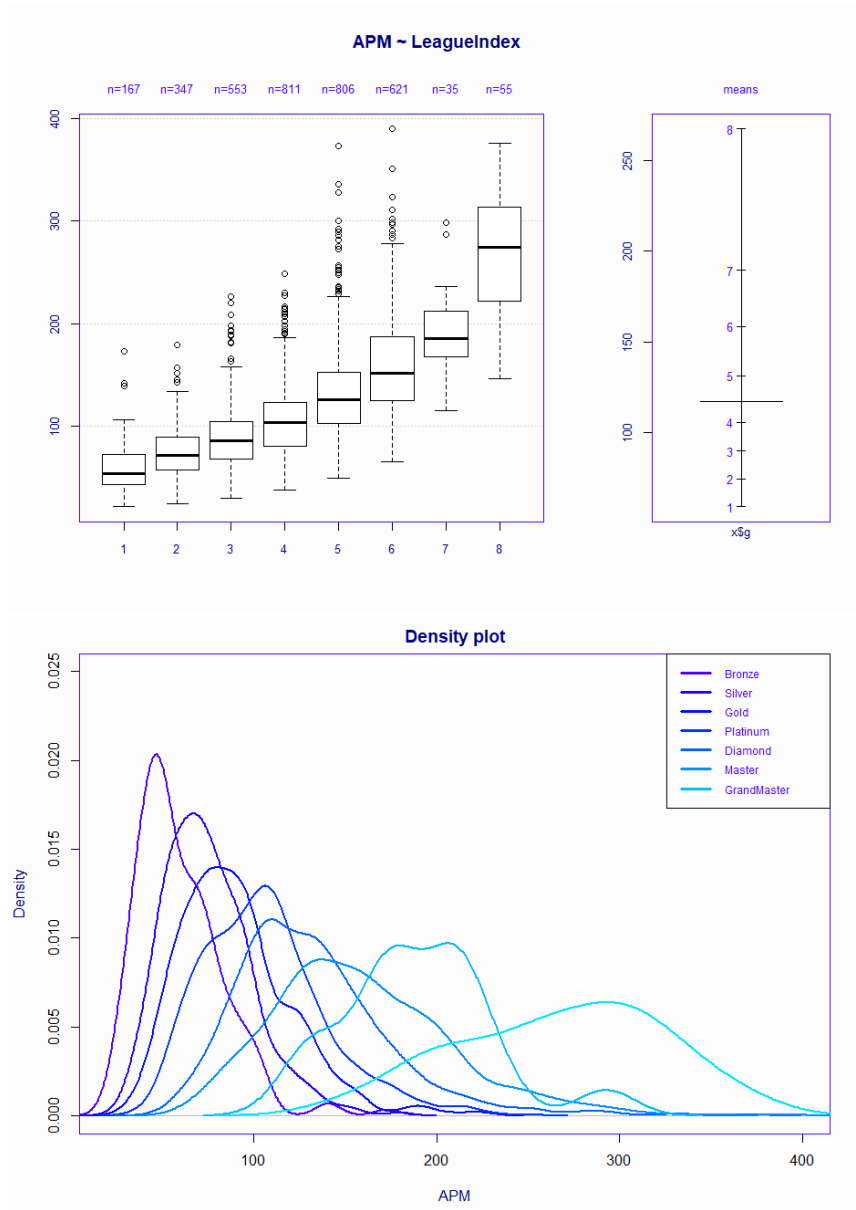
LeagueIndex	min	first_Qu	mean	median	third_Qu	max	sd	skewness	kurtosis
1	12	71	264.1916	200	400	1870	265.0602	2.035831	7.071268
2	10	110	331.4092	250	500	2000	289.4037	1.904884	5.702025
3	3	200	493.792	400	600	10260	624.2061	9.594175	127.815521
4	7	300	588.0062	500	730	18000	741.3068	16.380984	373.622688
5	25	400	781.6849	700	900	20000	923.1443	12.735118	241.076733
6	50	600	988.4058	800	1200	25000	1119.9386	16.089665	337.869467
7	240	715	1581.0286	1250	2000	10000	1652.0207	3.737328	16.362754
8	Inf	NA	NaN	NA	NA	-Inf	NaN	NaN	NaN



등급별로 TotalHours에 대한 boxplot과 density plot을 보면 1-7등급 모두 눈에 띄게 다른 분포를 하고 있음을 확인할 수 있다.

4. "APM"

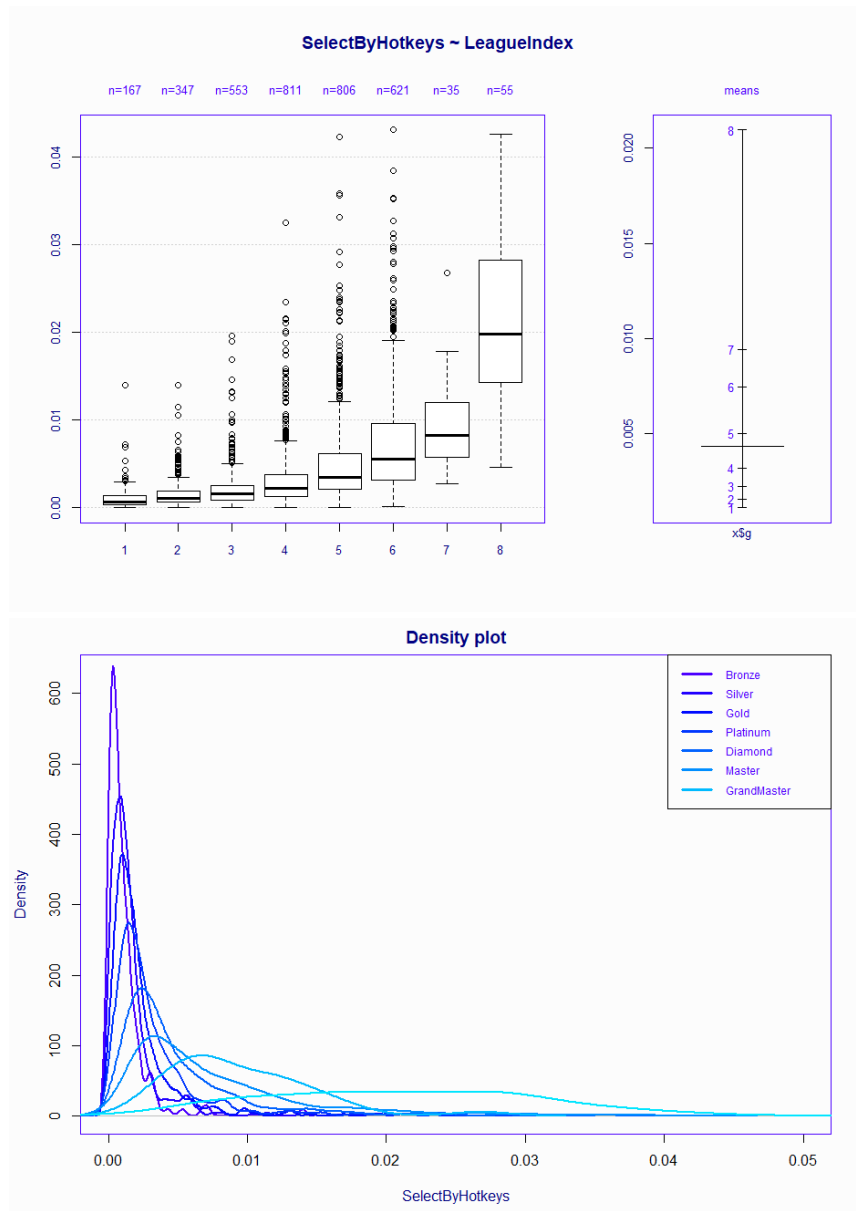
LeagueIndex	min	first_Qu	mean	median	third_Qu	max	sd	skewness	kurtosis
1	22.0596	43.3191	59.53928	54.045	72.5043	172.953	23.42944	1.3496191	3.2111521
2	24.6648	57.2142	74.78092	71.6754	89.2131	179.625	23.8554	0.8278485	1.1220153
3	29.8188	68.139	89.97126	85.959	104.853	226.6554	30.66116	1.0041821	1.7027597
4	38.028	80.7396	105.84717	103.8072	123.7281	249.021	33.81872	0.8046443	0.942452
5	49.7358	102.8381	131.52072	126.0075	152.5796	372.6426	41.69677	1.3007464	3.3950068
6	65.373	125.1288	158.68321	152.1852	187.1862	389.8314	48.3103	0.7931466	1.1327023
7	115.7556	168.1215	189.55569	185.3424	212.3751	298.7952	40.79249	0.5096221	0.3413219
8	146.3892	222.2286	267.34124	274.3404	313.2783	375.8664	56.18301	-0.2170089	-0.8200577



등급별로 APM 에 대한 boxplot과 density plot을 보면 1-7등급 모두 눈에 띄게 다른 분포를 하고 있음을 확인할 수 있다.

5. "SelectByHotkeys"

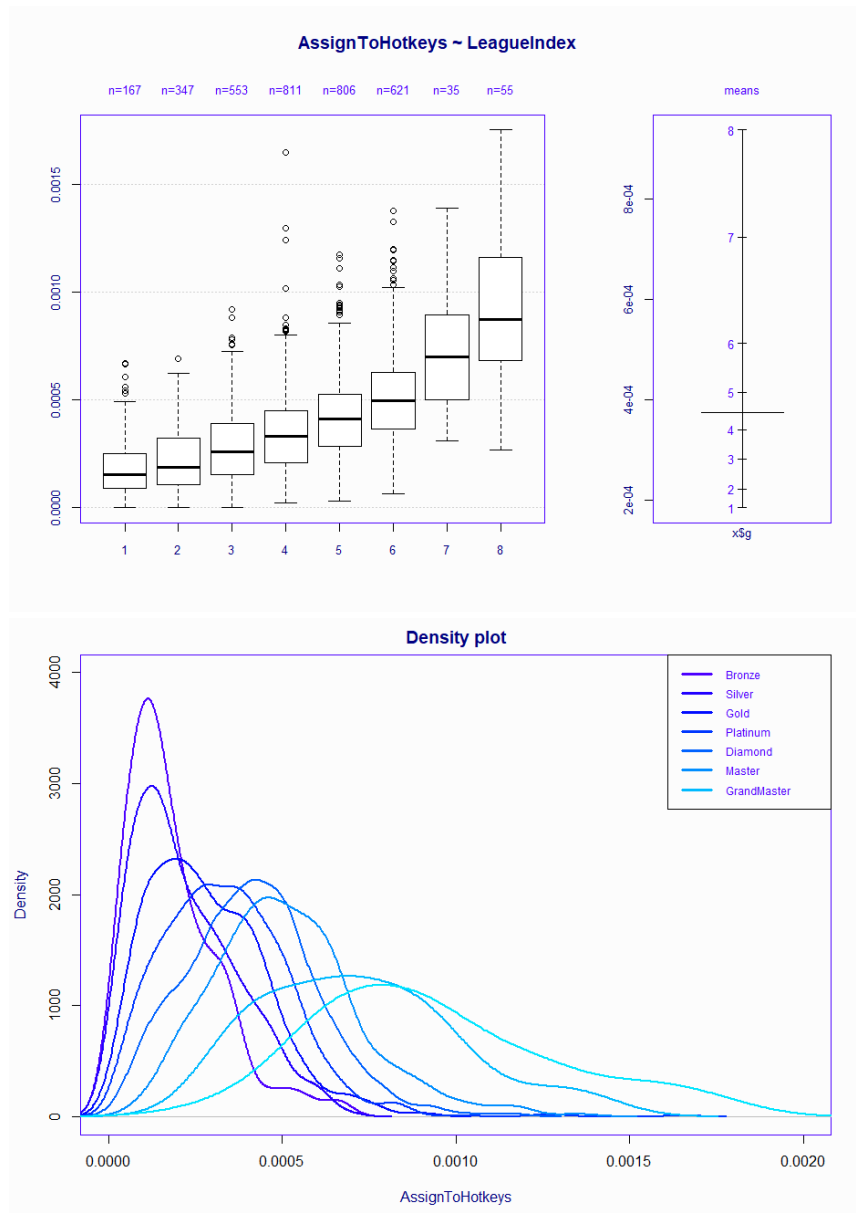
LeagueIndex	min	first_Qu	mean	median	third_Qu	max	sd	skewness	kurtosis
1	0.00E+00	0.00029	0.00108	0.00067	0.00137	0.01399	0.00150	4.83287	33.65233
2	0.00E+00	0.00058	0.00154	0.00109	0.00185	0.01399	0.00166	3.23031	15.23165
3	0.00E+00	0.00088	0.00219	0.00156	0.00254	0.01958	0.00237	3.38619	15.99585
4	0.00E+00	0.00126	0.00315	0.00217	0.00380	0.03245	0.00337	3.25644	15.08253
5	5.32E-05	0.00210	0.00498	0.00344	0.00611	0.04226	0.00492	2.89834	11.82924
6	1.22E-04	0.00309	0.00744	0.00547	0.00956	0.04309	0.00642	2.04959	5.26858
7	2.69E-03	0.00574	0.00942	0.00822	0.01195	0.02677	0.00499	1.22451	2.00748
8	4.58E-03	0.01431	0.02098	0.01982	0.02826	0.04258	0.00930	0.14248	-0.86628



등급별로 SelectByHotkeys 에 대한 boxplot과 density plot을 보면 1-7등급 모두 눈에 띄게 다른 분포를 하고 있음을 확인할 수 있다.

6. "AssignToHotkeys"

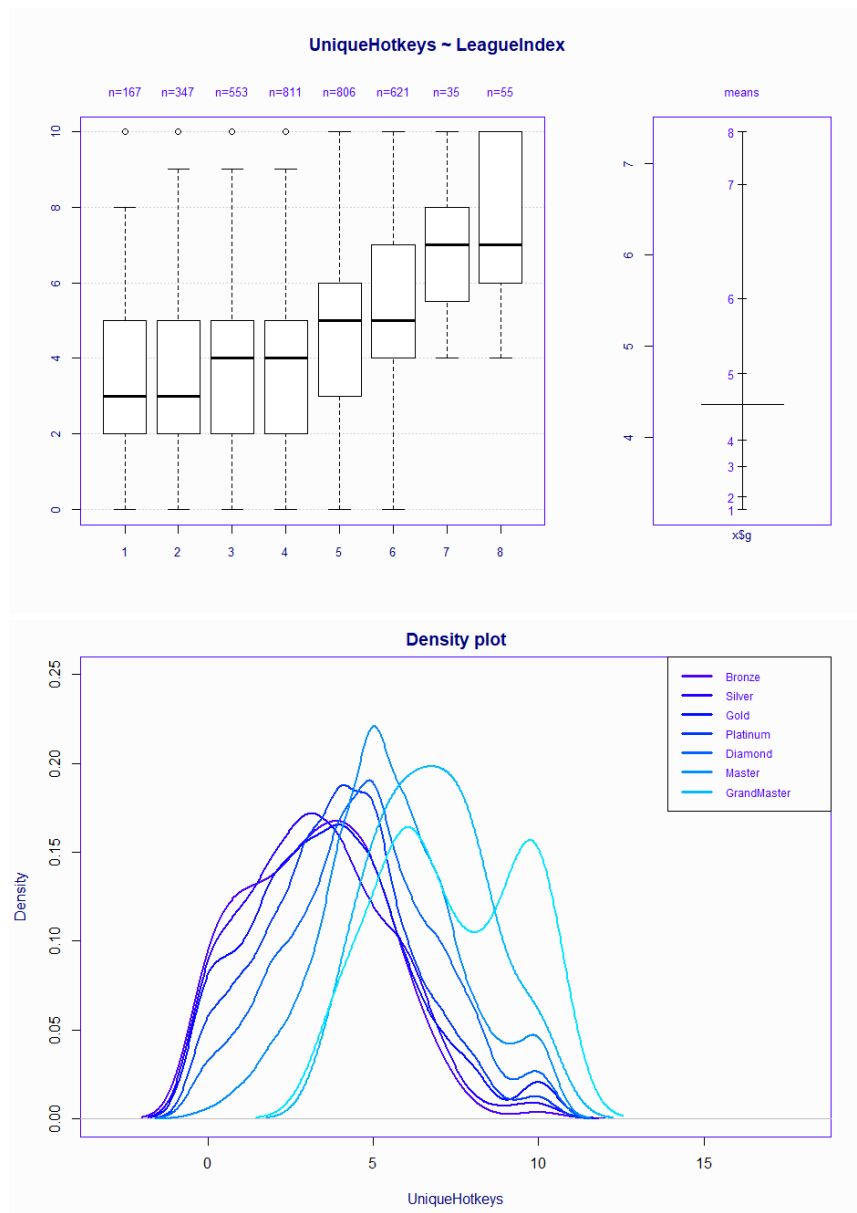
LeagueIndex	min	first_Qu	mean	median	third_Qu	max	sd	skewness	kurtosis
1	0.00E+00	9.06E-05	0.00019	0.00015	0.00025	0.00067	0.00013	1.26015	1.65062
2	0.00E+00	1.06E-04	0.00022	0.00019	0.00032	0.00069	0.00015	0.73959	-0.16972
3	0.00E+00	1.53E-04	0.00028	0.00026	0.00039	0.00092	0.00016	0.70654	0.37942
4	2.33E-05	2.05E-04	0.00034	0.00033	0.00045	0.00165	0.00018	1.09662	4.27009
5	3.11E-05	2.86E-04	0.00041	0.00041	0.00053	0.00117	0.00019	0.44874	0.50309
6	6.50E-05	3.66E-04	0.00051	0.00049	0.00063	0.00138	0.00021	0.71687	0.96153
7	3.10E-04	4.97E-04	0.00072	0.00070	0.00089	0.00139	0.00028	0.54092	-0.40038
8	2.66E-04	6.83E-04	0.00094	0.00087	0.00116	0.00175	0.00035	0.59083	-0.33890



등급별로 AssignToHotkeys 에 대한 boxplot과 density plot을 보면 1-7등급 모두 눈에 띄게 다른 분포를 하고 있음을 확인할 수 있다.

7. "UniqueHotkeys"

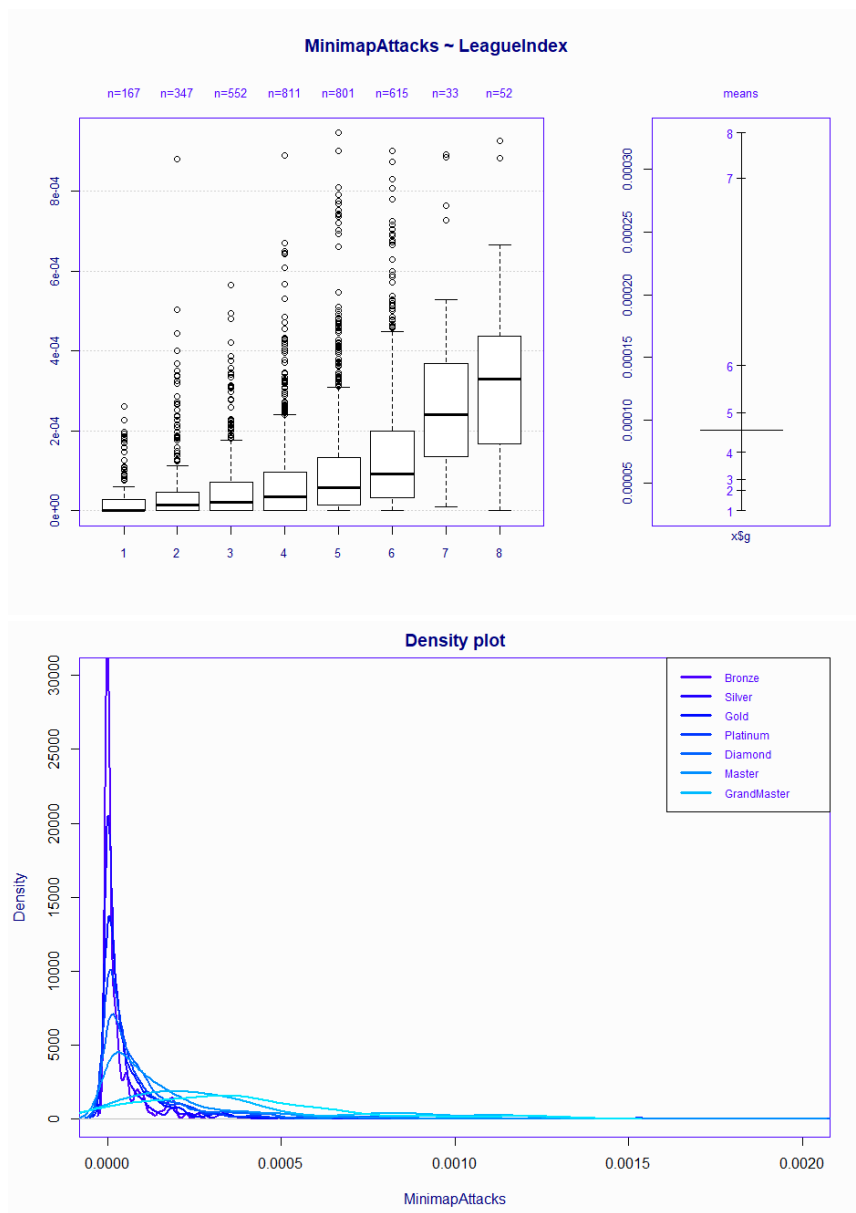
LeagueIndex	min	first_Qu	mean	median	third_Qu	max	sd	skewness	kurtosis
1	0	2	3.21557	3	5	10	2.05678	0.20659	-0.45736
2	0	2	3.35159	3	5	10	2.16989	0.41369	-0.12767
3	0	2	3.68716	4	5	10	2.34478	0.46770	0.00300
4	0	2	3.97164	4	5	10	2.17572	0.23097	-0.12945
5	0	3	4.69975	5	6	10	2.28072	0.11039	-0.27551
6	0	4	5.52174	5	7	10	2.05291	0.18723	-0.05012
7	4	5.5	6.77143	7	8	10	1.69923	0.20810	-0.84164
8	4	6	7.34546	7	10	10	2.11010	-0.03183	-1.41450



등급별로 UniqueHotkeys 에 대한 boxplot과 density plot을 보면 1-7등급의 분포는 비슷하지만 평균에 약간 차이가 있어 보인다.

8. "MinimapAttacks"

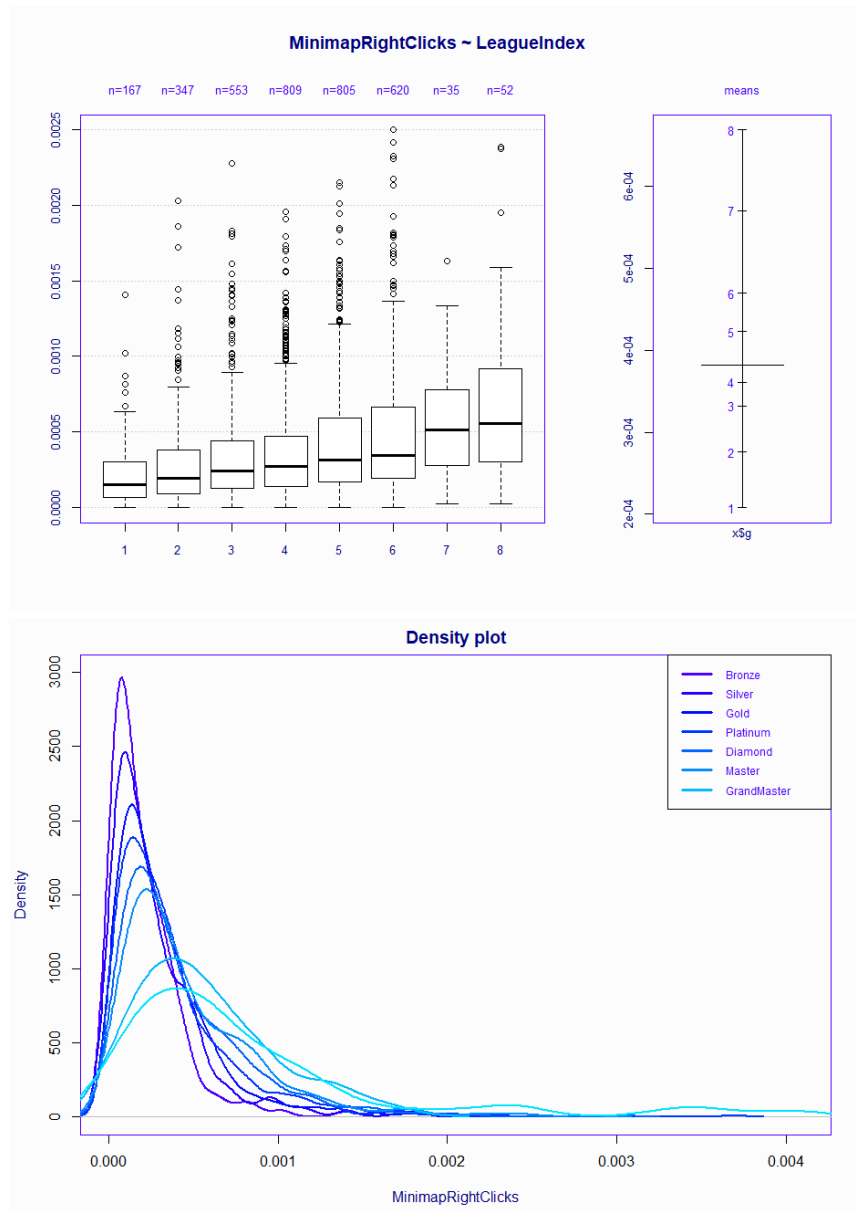
LeagueIndex	min	first_Qu	mean	median	third_Qu	max	sd	skewness	kurtosis
1	0.00E+00	0.00E+00	2.83E-05	0.00E+00	2.86E-05	0.00026	5.15E-05	2.32339	5.06518
2	0.00E+00	0.00E+00	4.46E-05	1.40E-05	4.52E-05	0.00088	8.91E-05	4.24176	26.22276
3	0.00E+00	0.00E+00	5.60E-05	2.13E-05	7.17E-05	0.00153	1.03E-04	6.55134	77.48515
4	0.00E+00	0.00E+00	7.48E-05	3.47E-05	9.60E-05	0.00089	1.09E-04	2.71248	9.95987
5	0.00E+00	1.36E-05	1.15E-04	5.73E-05	1.34E-04	0.00302	1.92E-04	6.09406	70.01977
6	0.00E+00	3.16E-05	1.56E-04	9.21E-05	2.03E-04	0.00197	2.03E-04	3.39449	18.59422
7	9.72E-06	1.38E-04	3.40E-04	2.41E-04	3.94E-04	0.00116	3.04E-04	1.22694	0.59770
8	0.00E+00	1.73E-04	3.76E-04	3.50E-04	4.98E-04	0.00124	2.95E-04	1.12703	1.17585



등급별로 MinimapAttacks 에 대한 boxplot과 density plot을 보면 1-7등급으로 갈수록 분포가 넓어지는 것을 볼 수 있다. (등급 별 boxplot의 차이를 확실하게 보기 위해 boxplot은 0.001이상의 값을 가진 이상치를 제외하고 그렸다.)

9. "MinimapRightClicks"

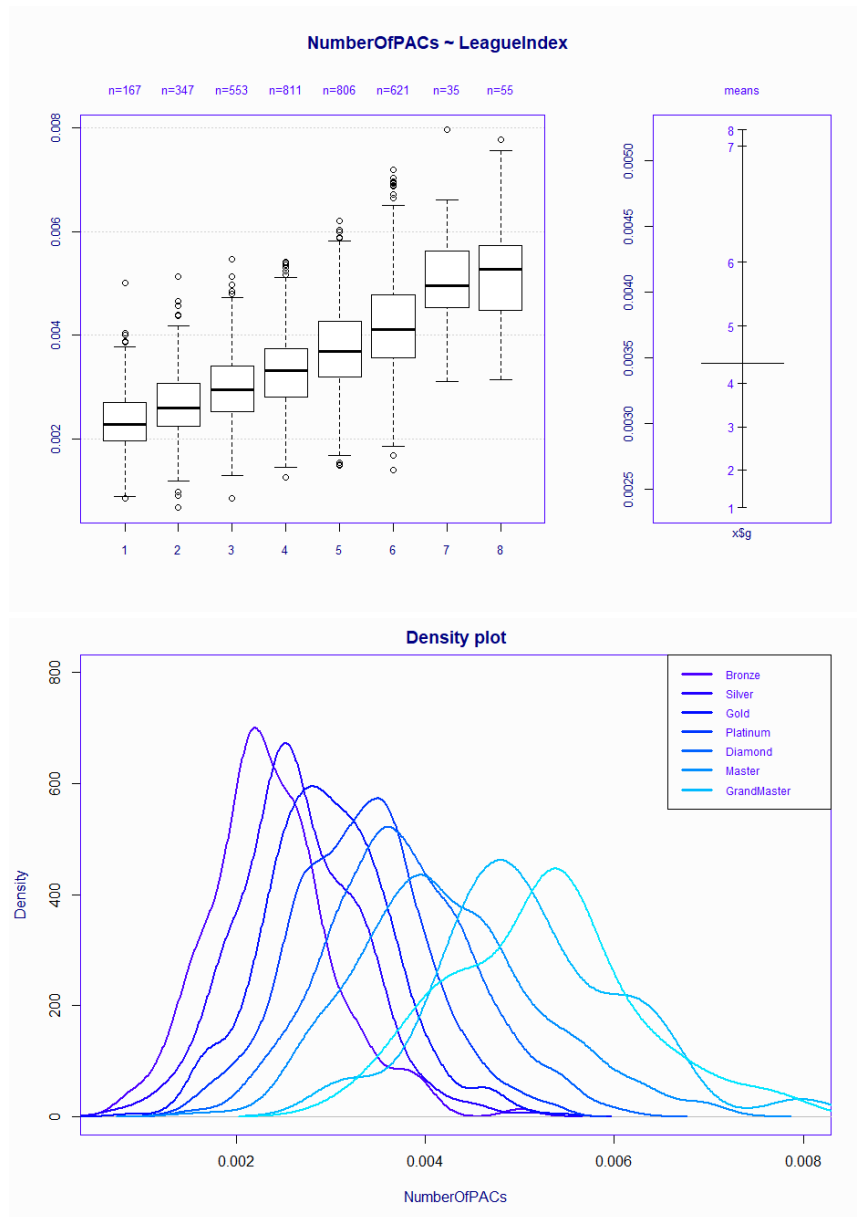
LeagueIndex	min	first_Qu	mean	median	third_Qu	max	sd	skewness	kurtosis
1	0.00E+00	6.50E-05	0.00021	0.00015	0.00030	0.00141	0.00020	2.26470	7.94697
2	0.00E+00	8.89E-05	0.00028	0.00020	0.00038	0.00203	0.00029	2.51830	9.15717
3	0.00E+00	1.25E-04	0.00033	0.00024	0.00044	0.00227	0.00031	2.21110	6.82061
4	0.00E+00	1.38E-04	0.00037	0.00027	0.00047	0.00369	0.00035	2.69615	13.89534
5	0.00E+00	1.68E-04	0.00043	0.00032	0.00059	0.00303	0.00037	1.85088	5.22397
6	0.00E+00	1.91E-04	0.00047	0.00035	0.00066	0.00282	0.00041	1.92555	5.20665
7	2.51E-05	2.78E-04	0.00057	0.00051	0.00078	0.00163	0.00039	0.85423	0.03962
8	2.33E-05	3.30E-04	0.00083	0.00057	0.00099	0.00404	0.00086	2.08358	4.15565



등급별로 MinimapRightClicks 에 대한 boxplot과 density plot을 보면 1-7등급으로 갈수록 분포가 넓어지는 것을 볼 수 있다. (등급별 boxplot의 차이를 확실하게 보기 위해 boxplot은 0.0025이상의 값을 가진 이상치를 제외하고 그렸다.)

10. "NumberOfPACs"

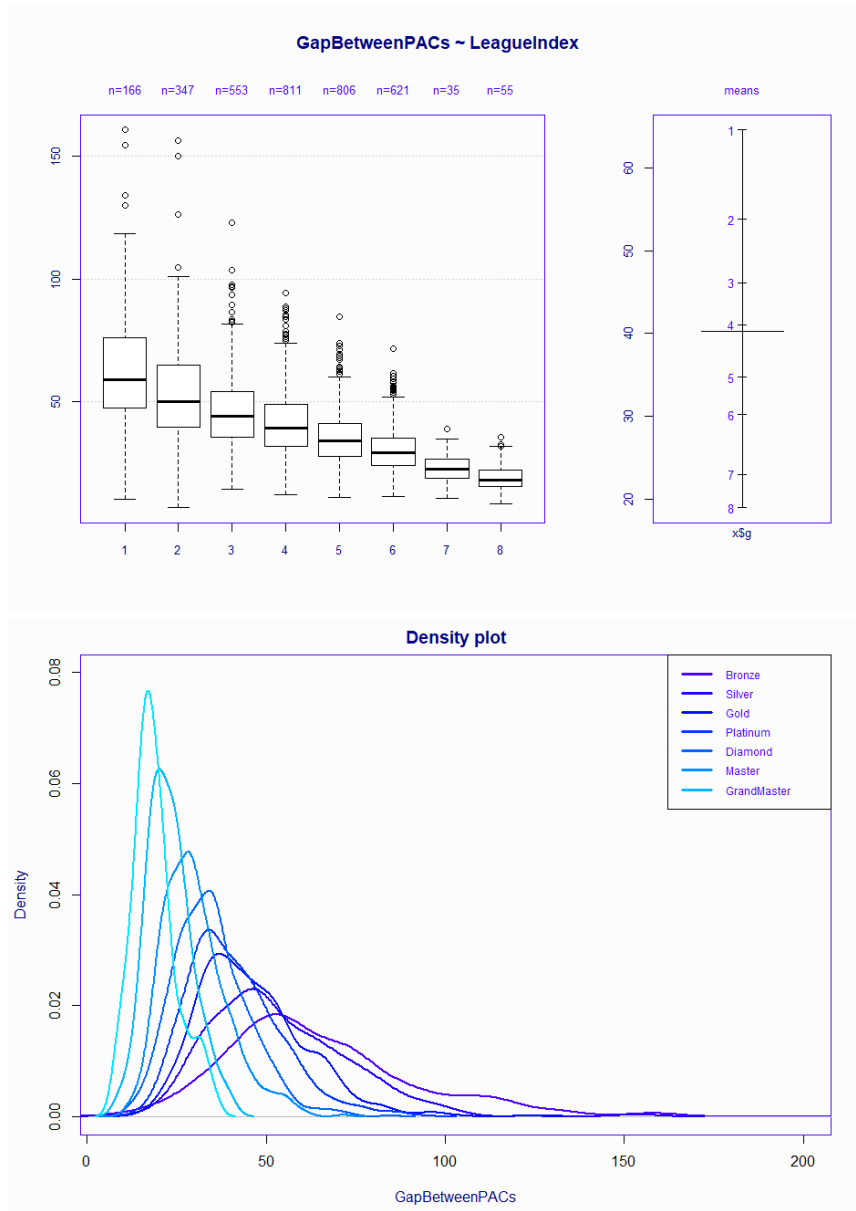
LeagueIndex	min	first_Qu	mean	median	third_Qu	max	sd	skewness	kurtosis
1	0.00085	0.00197	0.00236	0.00229	0.00271	0.00501	0.00066	0.59308	1.15063
2	0.00068	0.00225	0.00265	0.00260	0.00308	0.00514	0.00066	0.23427	0.49767
3	0.00086	0.00253	0.00298	0.00295	0.00341	0.00546	0.00068	0.26342	0.43988
4	0.00126	0.00281	0.00331	0.00332	0.00374	0.00542	0.00071	0.11591	0.08974
5	0.00149	0.00320	0.00374	0.00369	0.00426	0.00621	0.00079	0.18226	0.00069
6	0.00140	0.00358	0.00423	0.00411	0.00479	0.00719	0.00097	0.40009	0.06965
7	0.00311	0.00453	0.00511	0.00495	0.00564	0.00797	0.00100	0.46548	0.44023
8	0.00314	0.00449	0.00524	0.00528	0.00573	0.00778	0.00100	0.29589	-0.08674



등급별로 NumberOfPACs 에 대한 boxplot과 density plot을 보면 1-7등급으로 갈수록 분포가 약간씩 넓어지고 있고 평균도 조금씩 커지는 형태를 보인다..

11. "GapBetweenPACs"

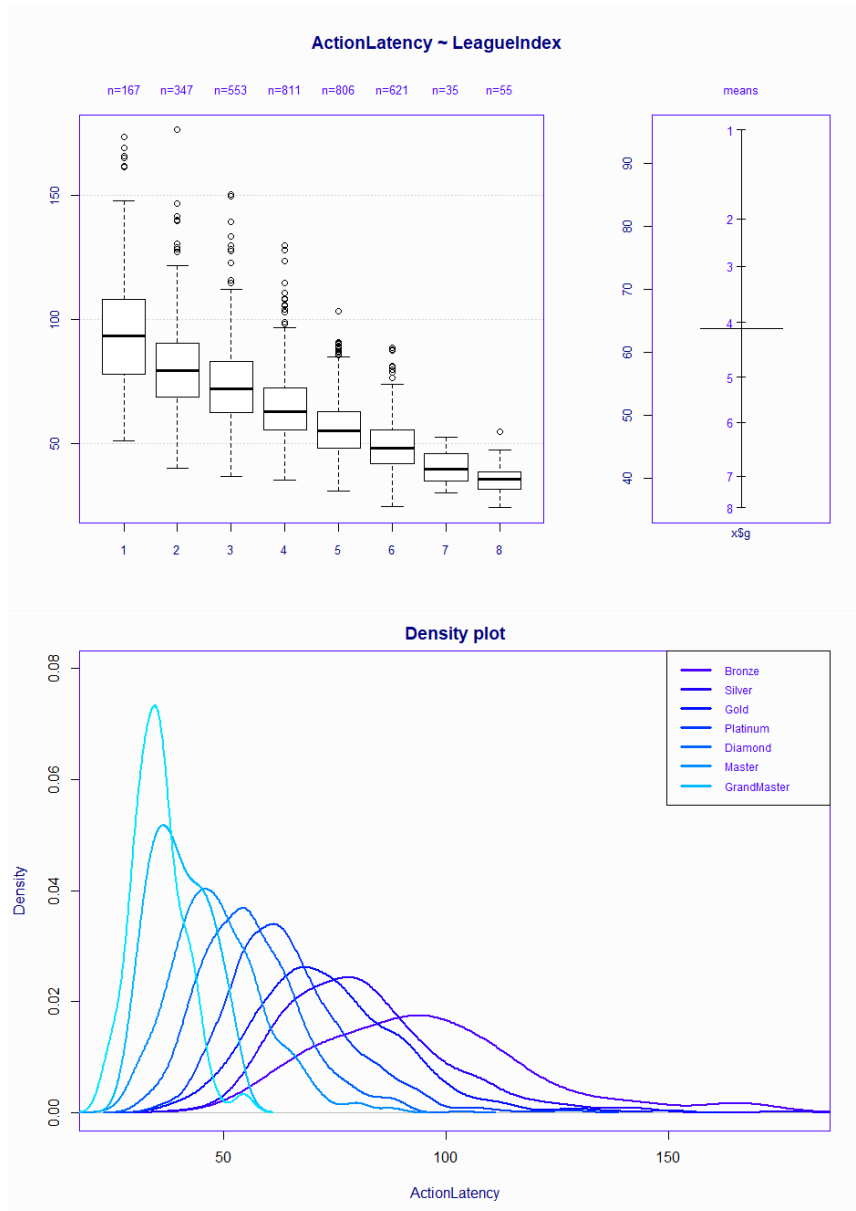
LeagueIndex	min	first_Qu	mean	median	third_Qu	max	sd	skewness	kurtosis
1	9.93550	47.33445	65.65499	58.94920	76.26410	237.14290	29.54617	1.76132	6.31802
2	6.66670	39.62770	53.78889	50.02900	64.83340	156.62340	19.81049	1.15247	3.03082
3	14.04880	35.23080	46.14388	43.80950	54.03640	122.80000	14.92129	0.93343	1.55556
4	12.04350	31.59415	41.04143	39.09090	48.61940	94.14810	13.01038	0.81537	0.86229
5	10.85710	27.39753	34.72647	33.80540	40.87980	84.60870	10.51014	0.69123	0.99432
6	11.32870	23.71220	30.20789	28.96730	35.08680	71.50000	9.16840	0.87826	1.01990
7	10.28570	18.61305	22.97816	22.48550	26.43180	38.78790	6.15757	0.45204	-0.11706
8	8.16420	15.10215	18.96977	17.99010	21.77410	35.41270	6.08393	0.70617	0.14917



등급별로 GapBetweenPACs 에 대한 boxplot을 보면 1-7등급으로 갈수록 분포가 좁아지고 평균은 낮아지고 있다. (등급별 boxplot의 차이를 확실하게 보기 위해 boxplot은 200이상의 값을 가진 이상치를 제외하고 그렸다.)

12. "ActionLatency"

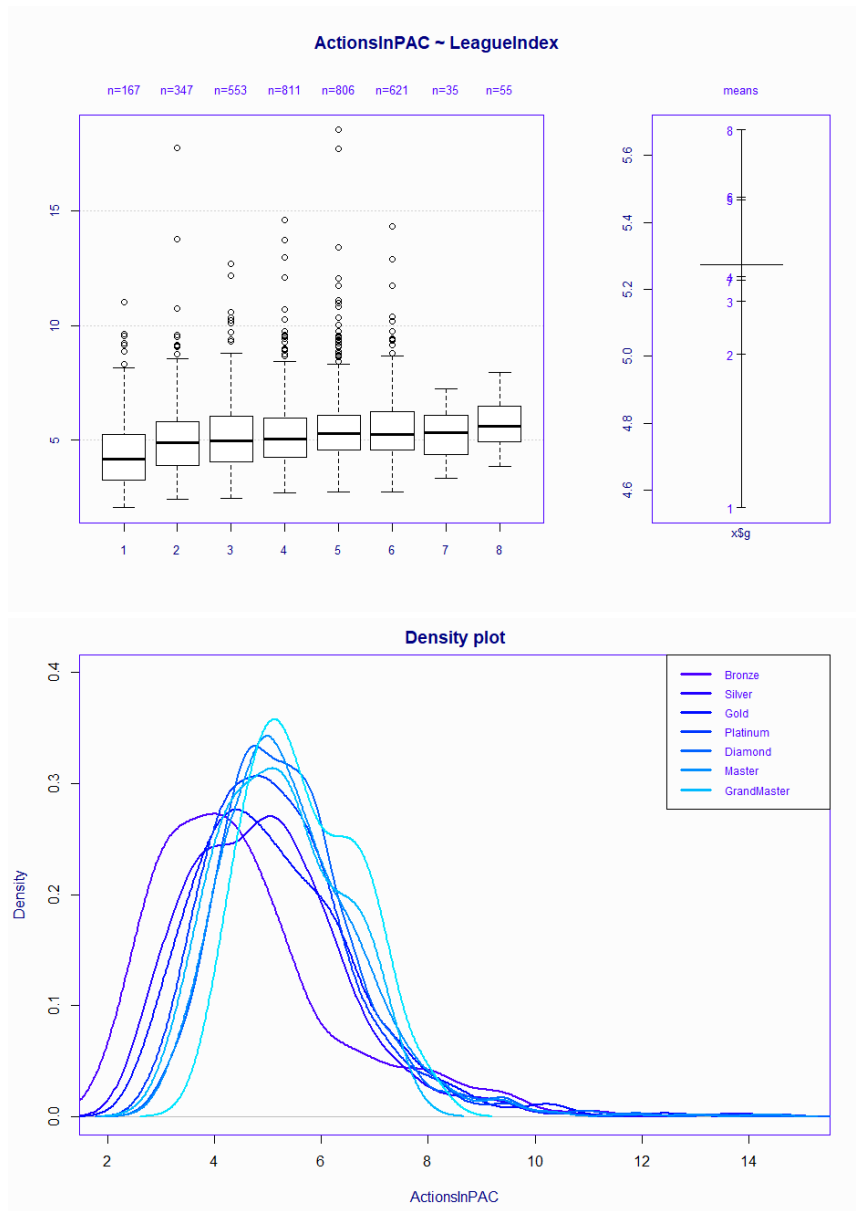
LeagueIndex	min	first_Qu	mean	median	third_Qu	max	sd	skewness	kurtosis
1	51.10650	77.88310	95.40498	93.32710	107.82420	173.55560	24.26315	0.85383	0.92780
2	40.01090	68.48400	81.27005	79.23750	90.17595	176.37210	18.12375	1.12694	2.56426
3	36.58480	62.46150	73.69730	72.09520	83.07340	150.29510	16.57936	0.88003	1.93555
4	35.13640	55.52360	64.79376	62.77000	72.23600	129.84620	13.38149	0.92019	1.71800
5	30.75930	48.10230	56.08501	54.98955	62.71603	103.37600	11.21285	0.58199	0.50466
6	24.63260	41.83080	48.94784	47.96450	55.32620	88.32260	10.42236	0.63395	0.69226
7	29.98710	34.90720	40.34290	39.73130	45.72630	52.31110	6.49089	0.20759	-1.15380
8	24.09360	31.28295	35.39015	35.40940	38.57910	54.55880	5.79208	0.60864	0.83535



등급별로 ActionLatency 에 대한 boxplot을 보면 1-7등급으로 갈수록 분포가 좁아지고 평균은 낮아지고 있다.

13. "ActionsInPAC"

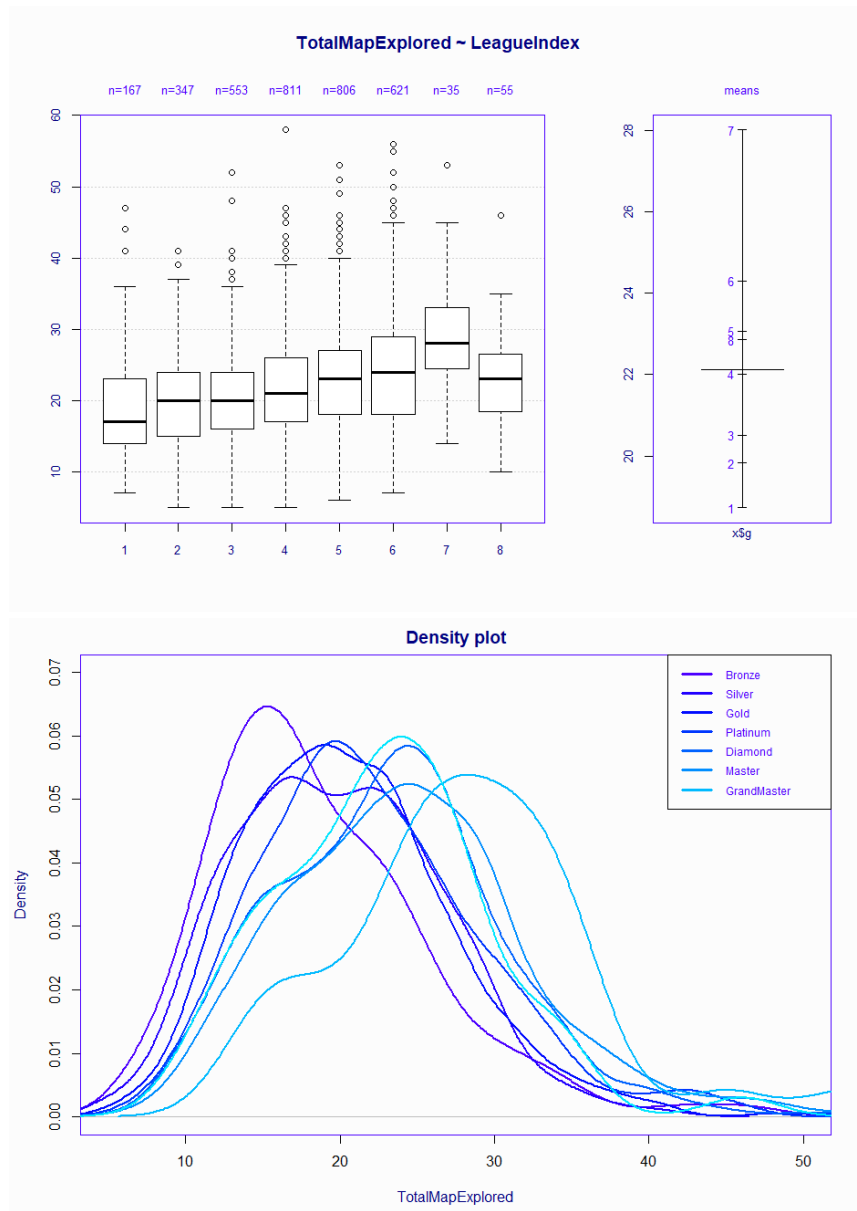
LeagueIndex	min	first_Qu	mean	median	third_Qu	max	sd	skewness	kurtosis
1	2.03890	3.25275	4.54829	4.15880	5.24615	11.01590	1.71767	1.20209	1.32563
2	2.39550	3.85855	5.00595	4.87900	5.77965	17.76190	1.65360	2.15278	11.54732
3	2.45400	4.04350	5.16328	4.93570	6.03900	12.68000	1.53827	1.05244	1.97083
4	2.67220	4.25295	5.23697	5.04910	5.95785	14.60710	1.41424	1.51238	5.47290
5	2.73240	4.53810	5.46803	5.26355	6.07805	18.55810	1.48900	2.47131	14.48470
6	2.71080	4.56930	5.47556	5.24890	6.23810	14.34330	1.36570	1.39563	4.67261
7	3.32210	4.34495	5.22609	5.30620	6.07440	7.23640	1.07831	0.17396	-1.07908
8	3.83710	4.92950	5.67673	5.58720	6.45750	7.96150	1.01712	0.31685	-0.86068



등급별로 ActionsInPAC 에 대한 boxplot과 density plot을 보면 1-7등급 모두 비슷한 분포를 가지고 있다. boxplot을 보면 6, 7 등급은 다른 등급에 비해 이상치가 적다.

14. "TotalMapExplored"

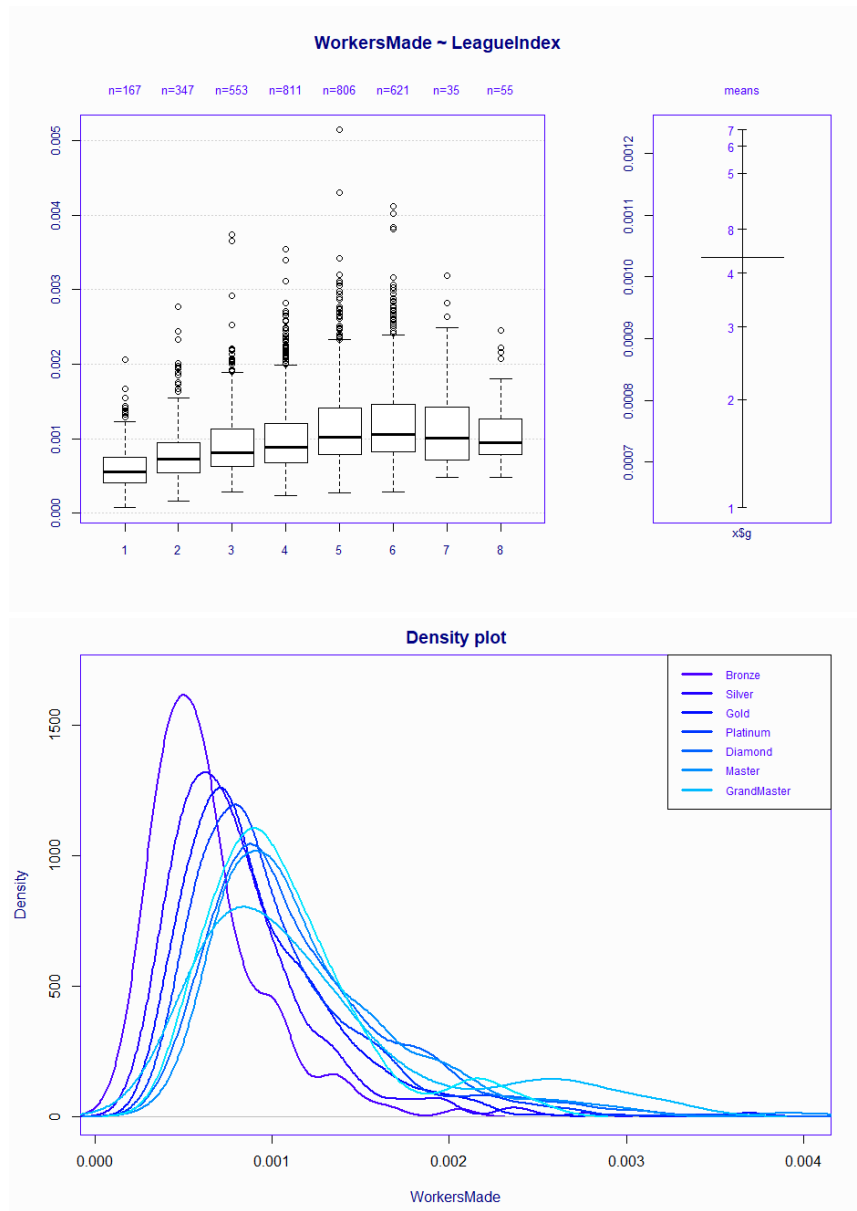
LeagueIndex	min	first_Qu	mean	median	third_Qu	max	sd	skewness	kurtosis
1	7	14	18.73054	17	23	47	7.07655	1.09364	1.72919
2	5	15	19.82421	20	24	41	6.51372	0.27425	-0.27563
3	5	16	20.50633	20	24	52	6.71336	0.68701	1.03178
4	5	17	22.01973	21	26	58	7.37059	0.75000	0.99494
5	6	18	23.07568	23	27	53	7.25356	0.38583	0.40072
6	7	18	24.30757	24	29	56	7.85393	0.64492	0.88869
7	14	24.5	28.02857	28	33	53	8.11493	0.63748	1.05176
8	10	18.5	22.85455	23	26.5	46	6.91862	0.52826	0.74774



등급별로 TotalMapExplored 에 대한 boxplot과 density plot을 보면 1-7등급 모두 비슷한 분포를 가지고 있다.

15. "WorkersMade"

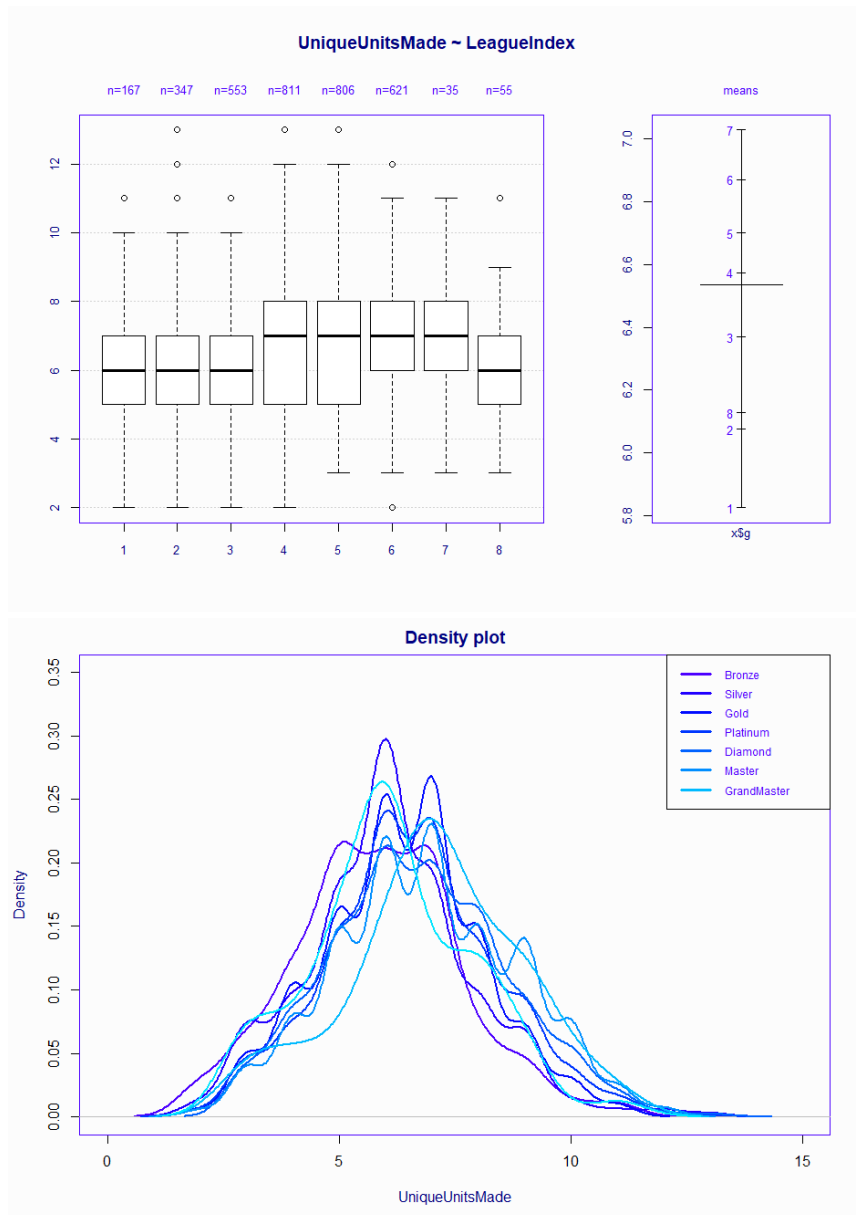
LeagueIndex	min	first_Qu	mean	median	third_Qu	max	sd	skewness	kurtosis
1	7.70E-05	0.00041	0.00063	0.00055	0.00076	0.00206	0.00032	1.36727	2.57611
2	1.65E-04	0.00054	0.00080	0.00072	0.00095	0.00277	0.00039	1.58917	3.78089
3	2.84E-04	0.00063	0.00092	0.00081	0.00113	0.00374	0.00044	1.81569	6.21561
4	2.42E-04	0.00068	0.00101	0.00088	0.00121	0.00355	0.00048	1.54472	3.20292
5	2.79E-04	0.00079	0.00117	0.00102	0.00141	0.00515	0.00055	1.70389	5.32218
6	2.82E-04	0.00082	0.00121	0.00106	0.00146	0.00412	0.00056	1.63109	3.80341
7	4.84E-04	0.00071	0.00124	0.00101	0.00143	0.00319	0.00070	1.25994	0.64799
8	4.78E-04	0.00079	0.00108	0.00095	0.00127	0.00246	0.00044	1.23040	1.28530



등급별로 WorkersMade 에 대한 boxplot과 density plot을 보면 1-7등급으로 갈수록 분포가 조금씩 넓어지고 평균이 커지는 것을 확인할 수 있다.

16. "UniqueUnitsMade"

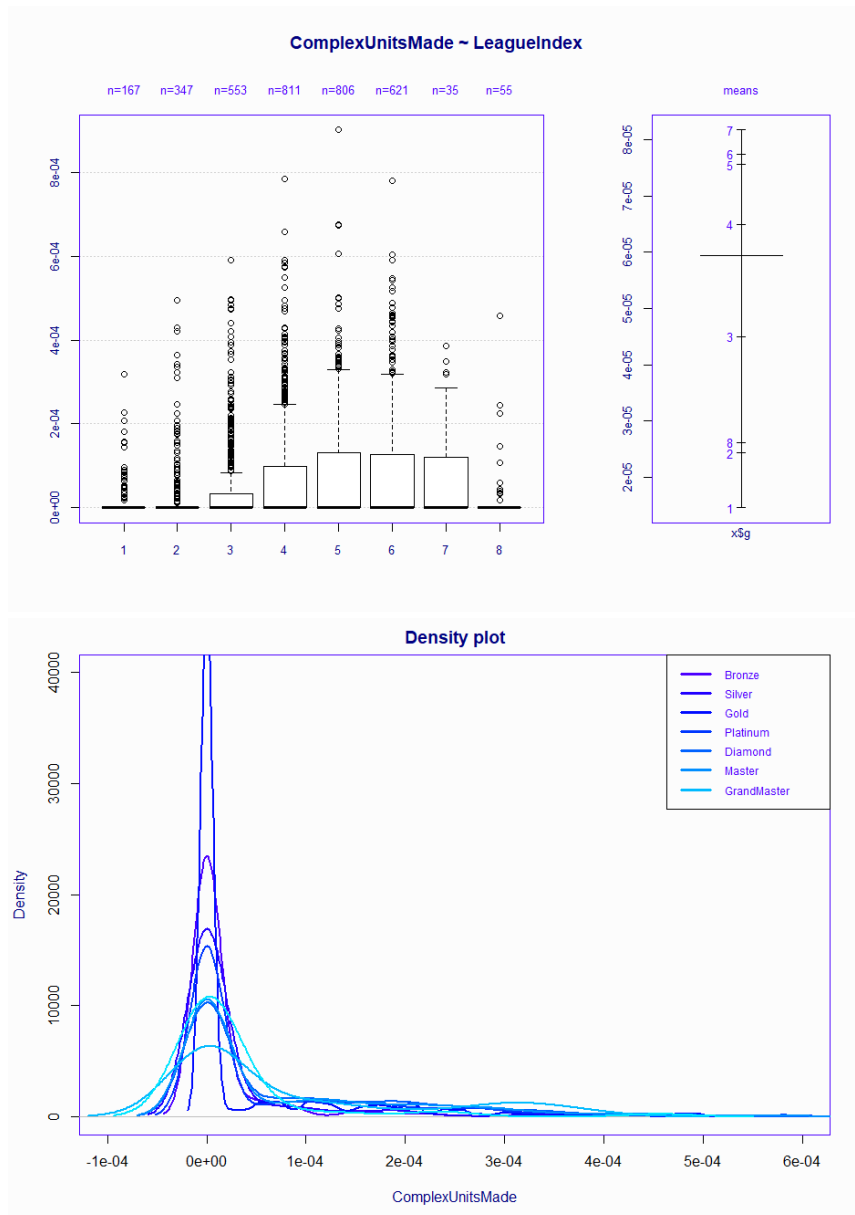
LeagueIndex	min	first_Qu	mean	median	third_Qu	max	sd	skewness	kurtosis
1	2	5	5.82635	6	7	11	1.78002	0.19098	0.09008
2	2	5	6.07493	6	7	13	1.78307	0.36936	0.71746
3	2	5	6.36890	6	7	11	1.70610	0.04154	-0.13670
4	2	5	6.57337	7	8	13	1.81910	0.18018	0.07083
5	3	5	6.69975	7	8	13	1.88730	0.21678	-0.22028
6	2	6	6.86957	7	8	12	1.95037	0.08926	-0.40761
7	3	6	7.02857	7	8	11	1.90179	-0.21409	-0.33471
8	3	5	6.12727	6	7	11	1.76441	0.20754	-0.16206



등급별로 UniqueUnitsMade 에 대한 boxplot과 density plot을 보면 1-7등급 모두 비슷한 분포를 가지고 있다.

17. "ComplexUnitsMade"

LeagueIndex	min	first_Qu	mean	median	third_Qu	max	sd	skewness	kurtosis
1	0	0	1.47E-05	0	0.00E+00	0.00032	4.46E-05	4.00320	18.11990
2	0	0	2.44E-05	0	0.00E+00	0.00049	7.07E-05	3.75178	15.55968
3	0	0	4.50E-05	0	3.36E-05	0.00059	9.61E-05	2.58826	7.09891
4	0	0	6.50E-05	0	9.86E-05	0.00079	1.18E-04	2.21341	5.46023
5	0	0	7.56E-05	0	1.32E-04	0.00090	1.19E-04	1.98239	5.24096
6	0	0	7.74E-05	0	1.27E-04	0.00078	1.26E-04	1.85515	3.49597
7	0	0	8.18E-05	0	1.21E-04	0.00039	1.25E-04	1.20223	-0.11914
8	0	0	2.62E-05	0	0.00E+00	0.00046	7.76E-05	3.88534	16.35932

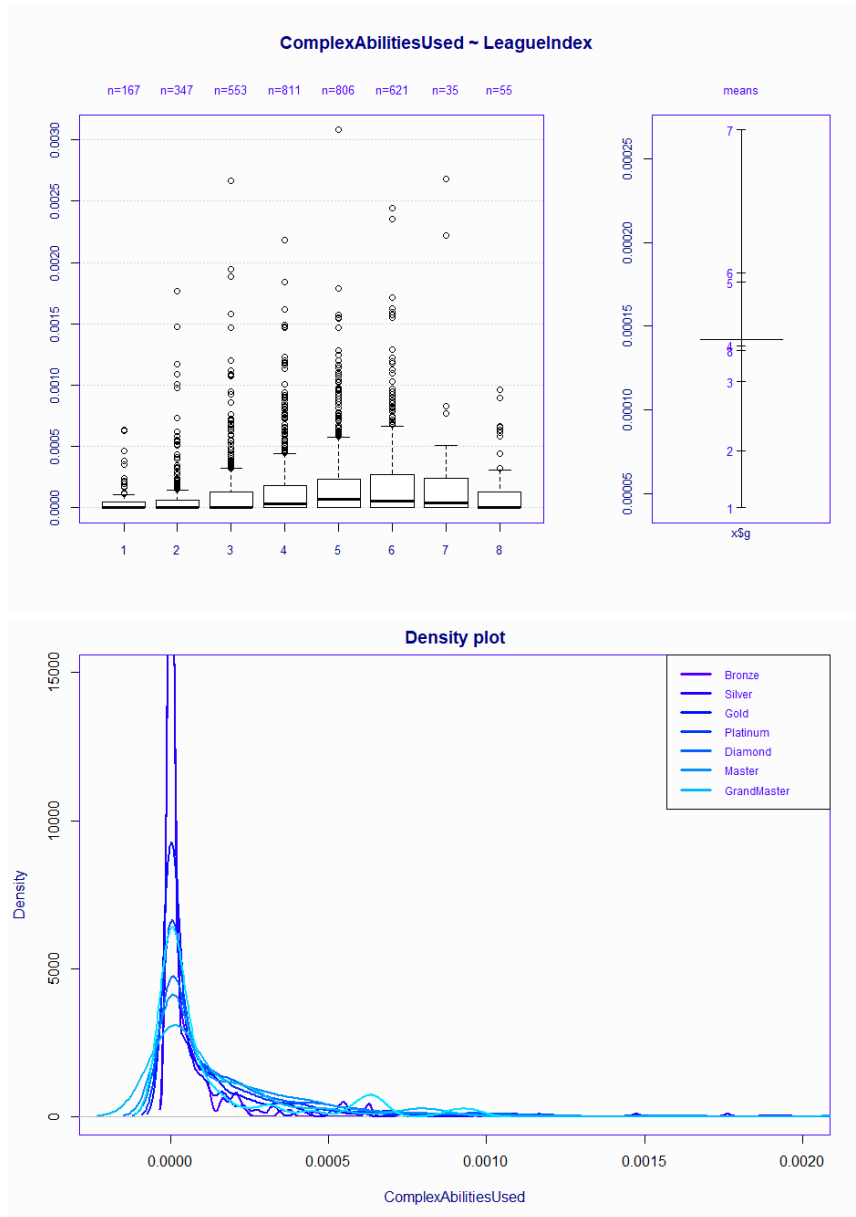


ComplexUnitsMade은 모든 등급에서 반 이상 0의 값을 가진다.

등급별로 ComplexUnitsMade 에 대한 boxplot과 density plot을 보면 1-6등급으로 갈수록 분포가 조금씩 넓어지지만 평균은 비슷한 것을 확인할 수 있다.

18. "ComplexAbilitiesUsed"

LeagueIndex	min	first_Qu	mean	median	third_Qu	max	sd	skewness	kurtosis
1	0	0	4.17E-05	0.00E+00	4.5081E-05	0.00062987	9.59E-05	3.892596	17.820972
2	0	0	7.56E-05	0.00E+00	0.00006	0.00176	2.00E-04	4.69046	27.15037
3	0	0	1.17E-04	0.00E+00	0.00013	0.00266	2.57E-04	4.62372	29.83607
4	0	0	1.38E-04	2.86E-05	0.00018	0.00219	2.45E-04	3.35515	15.46170
5	0	0	1.76E-04	6.60E-05	0.00023	0.00308	2.82E-04	3.28157	18.58957
6	0	0	1.82E-04	5.26E-05	0.00027	0.00244	2.93E-04	3.14286	14.62203
7	0	0	2.67E-04	4.00E-05	0.00024	0.00268	5.88E-04	3.02587	8.71156
8	0	0	1.35E-04	0.00E+00	0.00013	0.00096	2.46E-04	1.91441	2.56400

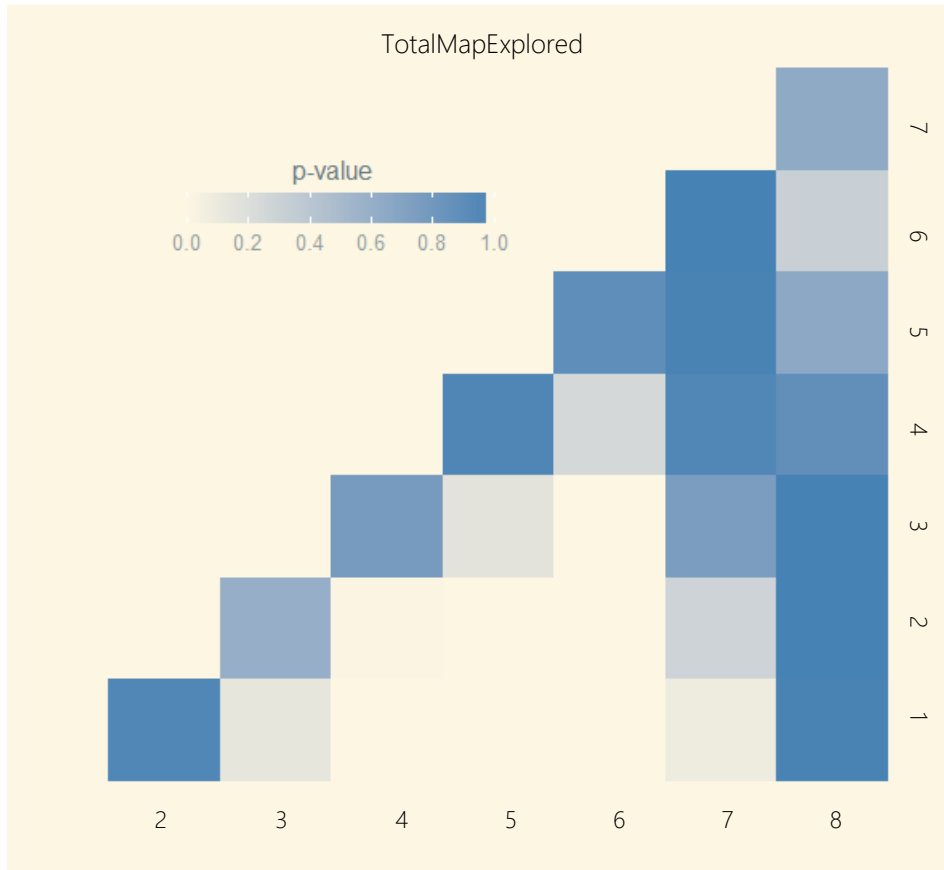


ComplexAbilitiesUsed 은 모든 등급에서 25%이상 0의 값을 가진다.

등급별로 ComplexAbilitiesUsed 에 대한 boxplot과 density plot을 보면 1-5등급으로 갈수록 분포가 조금씩 넓어지지만 평균은 비슷한 것을 확인할 수 있다.

B. 각 변수가 등급에 따라 차이가 있는지에 대한 검정

Levene등분산 검정을 해서 등분산 가정을 만족하는 변수들의 anova 검정을 하면 총 1개의 변수("TotalMapExplored")가 등급에 따라 유의미한 차이가 있다고 나온다. anova 사후검정을 하고, 그 결과로 각 등급 간의 변수의 평균에 차이가 있는지 검정에 대한 p-value를 매트릭스에 저장해서 plot을 그리면 아래와 같다.



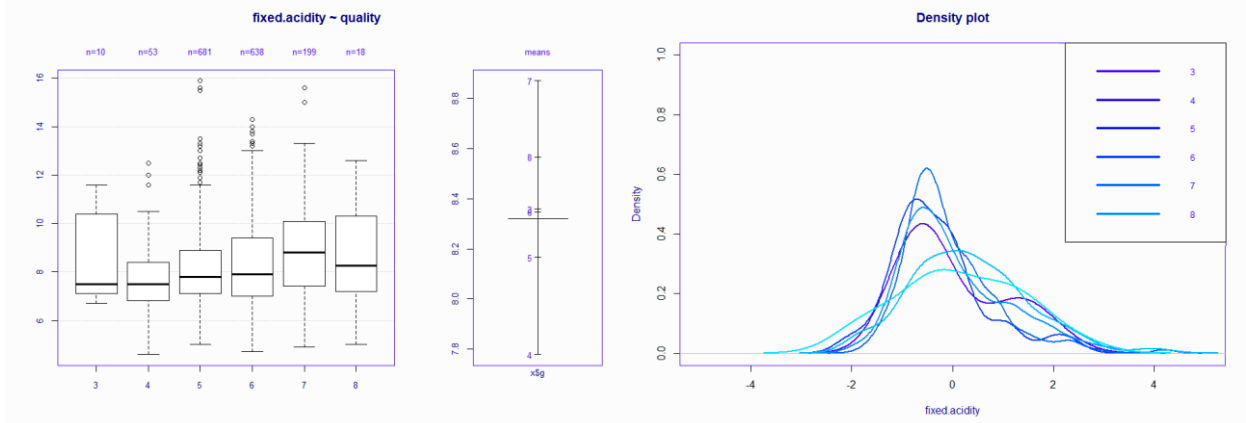
Wine Quality - RED

종속변수	"quality"
설명변수(11개)	"fixed.acidity", "volatile.acidity", "citric.acid", "residual.sugar", "chlorides", "free.sulfur.dioxide", "total.sulfur.dioxide", "density", "pH", "sulphates", "alcohol"

A. 등급별로 각 변수에 대한 특징 파악

1. "fixed.acidity"

quality	min	Q1	mean	median	Q3	max	sd	skewness	kurtosis
3	6.7	7.15	8.36	7.5	9.875	11.6	1.77088	0.68616	-1.33932
4	4.6	6.8	7.77925	7.5	8.4	12.5	1.62662	0.86894	0.82540
5	5	7.1	8.16725	7.8	8.9	15.9	1.56399	1.37686	3.07279
6	4.7	7	8.34718	7.9	9.4	14.3	1.79785	0.83019	0.20189
7	4.9	7.4	8.87236	8.8	10.1	15.6	1.99248	0.50668	0.40459
8	5	7.25	8.56667	8.25	10.225	12.6	2.11966	0.03662	-1.03776

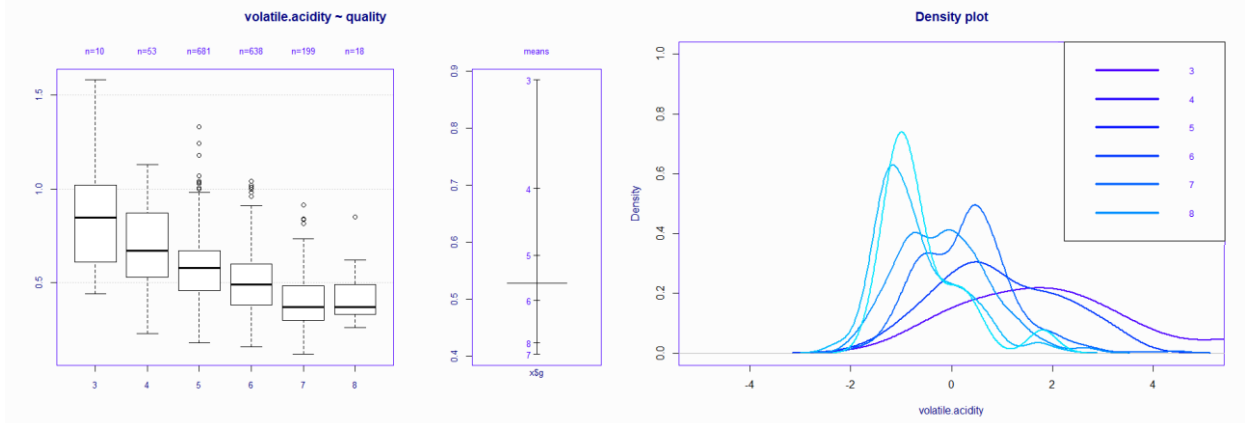


Wine Quality – RED 데이터의 모든 boxplot과 density plot은 변수의 값을 scale하고 그렸다.

등급별로 fixed.acidity 에 대한 boxplot과 density plot을 보면 3-8등급 모두 비슷한 분포를 가지고 있다.

2. "volatile.acidity"

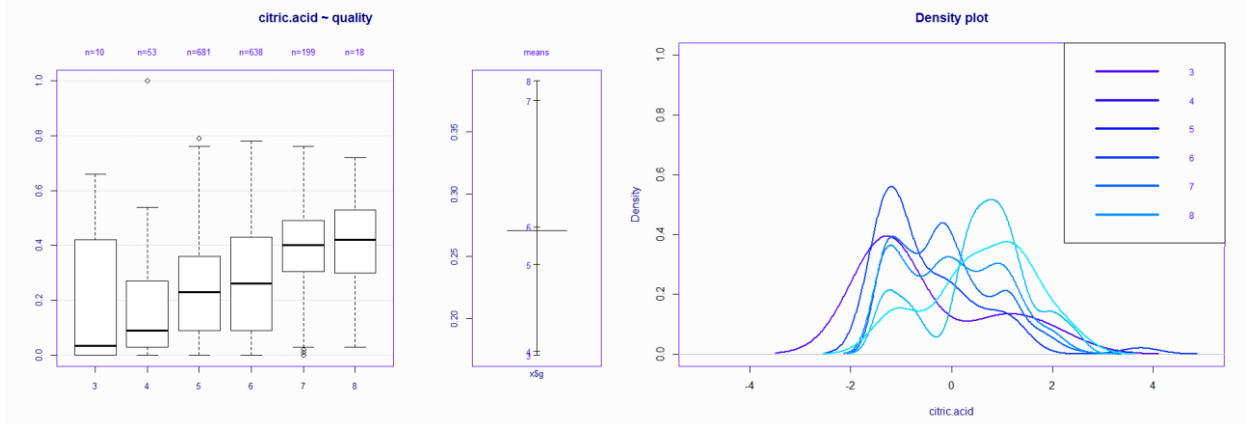
quality	min	Q1	mean	median	Q3	max	sd	skewness	kurtosis
3	0.44	0.6475	0.88450	0.845	1.01	1.58	0.33126	0.62702	-0.54055
4	0.23	0.53	0.69396	0.67	0.87	1.13	0.22011	0.14883	-0.81530
5	0.18	0.46	0.57704	0.58	0.67	1.33	0.16480	0.59110	1.36340
6	0.16	0.38	0.49748	0.49	0.6	1.04	0.16096	0.43275	0.08844
7	0.12	0.3	0.40392	0.37	0.485	0.915	0.14522	0.94289	0.76023
8	0.26	0.335	0.42333	0.37	0.4725	0.85	0.14491	1.44551	1.62440



와인 등급별로 volatile.acidity 에 대한 boxplot과 density plot을 보면 3-8등급으로 갈수록 분포가 조금씩 좁아지고 평균이 작아지는 것을 확인할 수 있다.

3. "citric.acid"

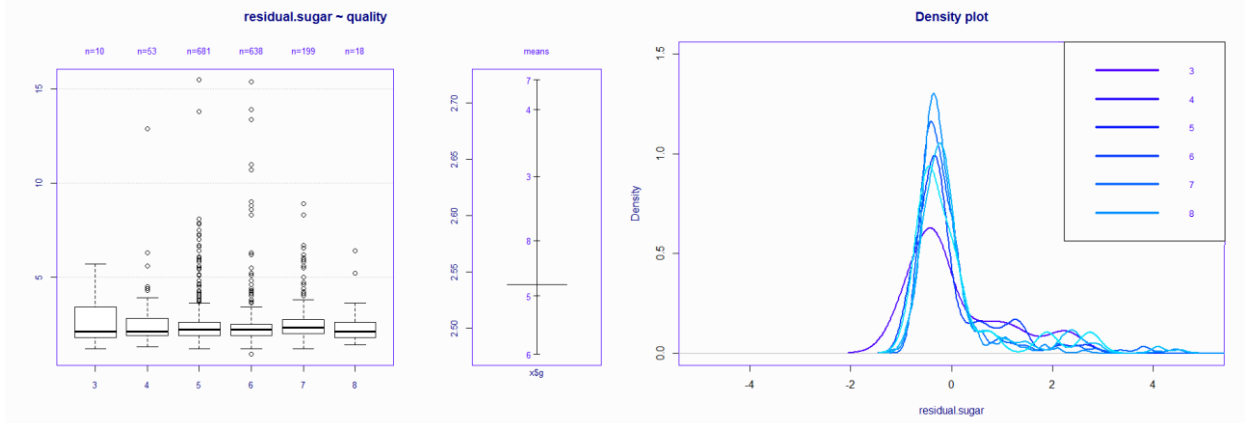
quality	min	Q1	mean	median	Q3	max	sd	skewness	kurtosis
3	0	0.005	0.171	0.035	0.3275	0.66	0.25066	0.88510	-1.08982
4	0	0.03	0.17415	0.09	0.27	1	0.20103	1.61983	3.34279
5	0	0.09	0.24369	0.23	0.36	0.79	0.18000	0.52227	-0.51632
6	0	0.09	0.27382	0.26	0.43	0.78	0.19511	0.22133	-0.98801
7	0	0.305	0.37518	0.4	0.49	0.76	0.19443	-0.37552	-0.45955
8	0.03	0.3025	0.39111	0.42	0.53	0.72	0.19953	-0.32605	-0.91807



와인 등급별로 citric.acid 에 대한 boxplot을 보면 3-8등급으로 갈수록 평균이 높아지고 있다. density plot을 보면 3-8등급 모두 비슷한 분포를 가지고 있다.

4. "residual.sugar"

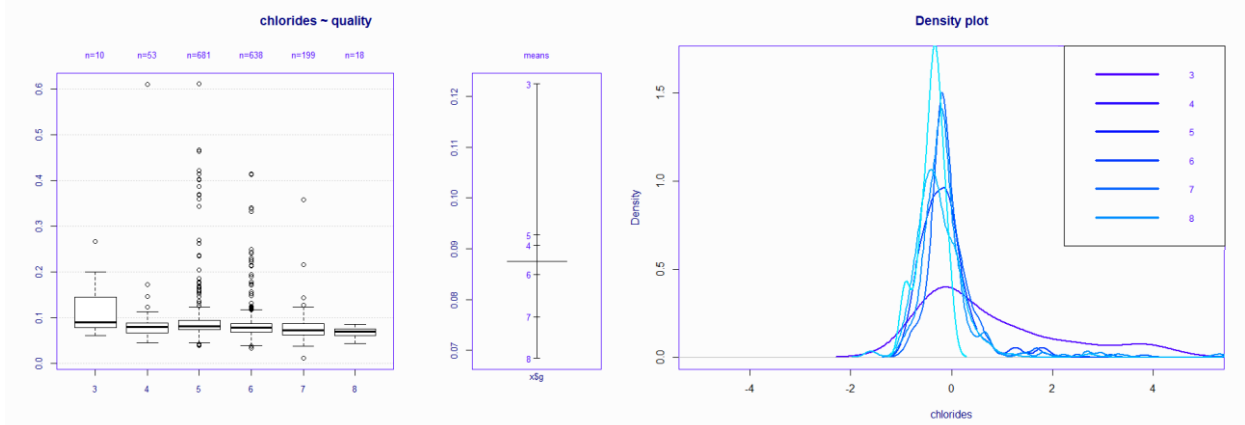
quality	min	Q1	mean	median	Q3	max	sd	skewness	kurtosis
3	1.2	1.875	2.635	2.1	3.1	5.7	1.40160	1.01374	-0.35526
4	1.3	1.9	2.69434	2.1	2.8	12.9	1.78944	3.73560	17.52507
5	1.2	1.9	2.52885	2.2	2.6	15.5	1.35975	4.38217	27.99537
6	0.9	1.9	2.47719	2.2	2.5	15.4	1.44158	5.46529	37.29144
7	1.2	2	2.72060	2.3	2.75	8.9	1.37151	2.18601	4.74273
8	1.4	1.8	2.57778	2.1	2.6	6.4	1.29504	1.79013	2.25881



와인 등급별로 residual.sugar 에 대한 boxplot과 density plot을 보면 3-8등급 모두 비슷한 분포와 평균을 가지고 있다. 하지만 5, 6, 7등급은 다른 등급에 비해 많은 이상치를 가지고 있다.

5. "chlorides"

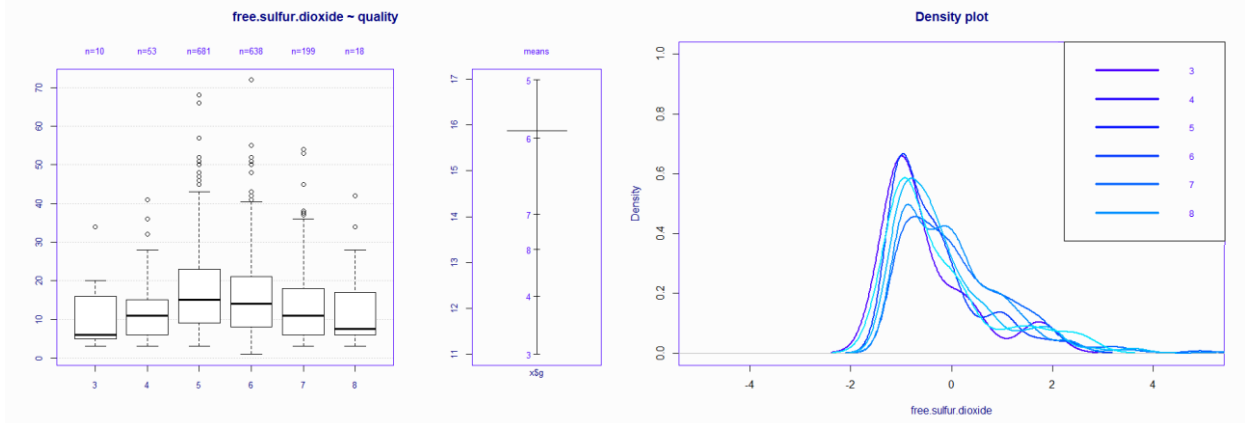
quality	min	Q1	mean	median	Q3	max	sd	skewness	kurtosis
3	0.061	0.079	0.1225	0.0905	0.143	0.267	0.06624	1.00540	-0.39558
4	0.045	0.067	0.09068	0.08	0.089	0.61	0.07619	5.98097	37.76511
5	0.039	0.074	0.09274	0.081	0.094	0.611	0.05371	5.31734	33.14192
6	0.034	0.06825	0.08496	0.078	0.088	0.415	0.03956	4.83818	30.68164
7	0.012	0.062	0.07659	0.073	0.087	0.358	0.02946	4.91050	42.10520
8	0.044	0.062	0.06844	0.0705	0.0755	0.086	0.01168	-0.65575	-0.42425



와인 등급별로 chlorides 에 대한 boxplot과 density plot을 보면 4-8등급 모두 비슷한 분포를 가지고 있다. 3등급만 다른 등급에 비해 넓은 분포를 가진다.

6. "free.sulfur.dioxide"

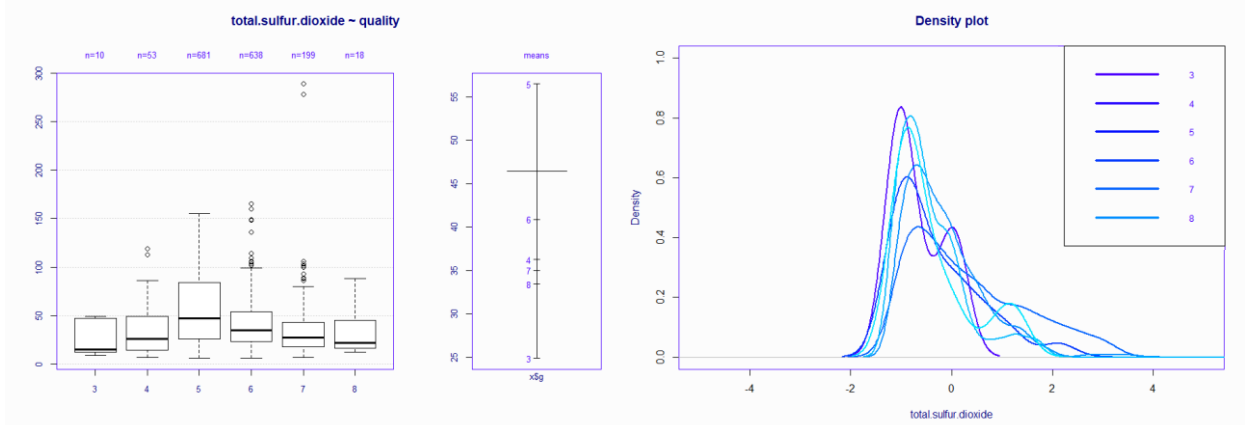
quality	min	Q1	mean	median	Q3	max	sd	skewness	kurtosis
3	3	5	11	6	14.5	34	9.76388	1.24728	0.25977
4	3	6	12.26415	11	15	41	9.02593	1.29504	1.06421
5	3	9	16.98385	15	23	68	10.95545	1.19577	1.84038
6	1	8	15.71160	14	21	72	9.94091	1.21765	2.24802
7	3	6	14.04523	11	18	54	10.17526	1.46938	2.01291
8	3	6	13.27778	7.5	16.5	42	11.15561	1.25284	0.41507



와인 등급별로 free.sulfur.dioxide 에 대한 boxplot을 보면 5-7등급이 조금 더 넓게 분포되어 있어 보이는데, density plot을 보면 3-8등급 모두 비슷한 분포를 가지고 있다.

7. "total.sulfur.dioxide"

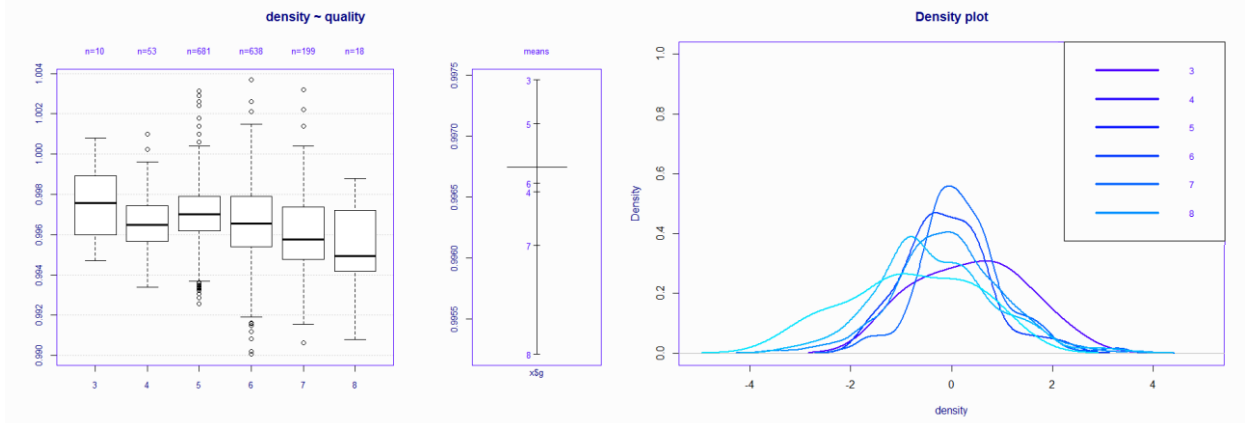
quality	min	Q1	mean	median	Q3	max	sd	skewness	kurtosis
3	9	12.5	24.9	15	42.5	49	16.82888	0.52537	-1.72289
4	7	14	36.24528	26	49	119	27.58337	1.13903	0.65027
5	6	26	56.51395	47	84	155	36.99312	0.75401	-0.40689
6	6	23	40.86991	35	54	165	25.03825	1.34132	2.57276
7	7	17.5	35.02010	27	43	289	33.19121	4.52456	29.43799
8	12	16	33.44444	21.5	43	88	25.43324	1.12455	-0.22225



와인 등급별로 total.sulfur.dioxide 에 대한 boxplot을 보면 5등급이 조금 더 넓게 분포되어 있어 보인다. density plot을 보면 5등급을 제외한 3-8등급 모두 비슷한 분포를 가지고 있다.

8. "density"

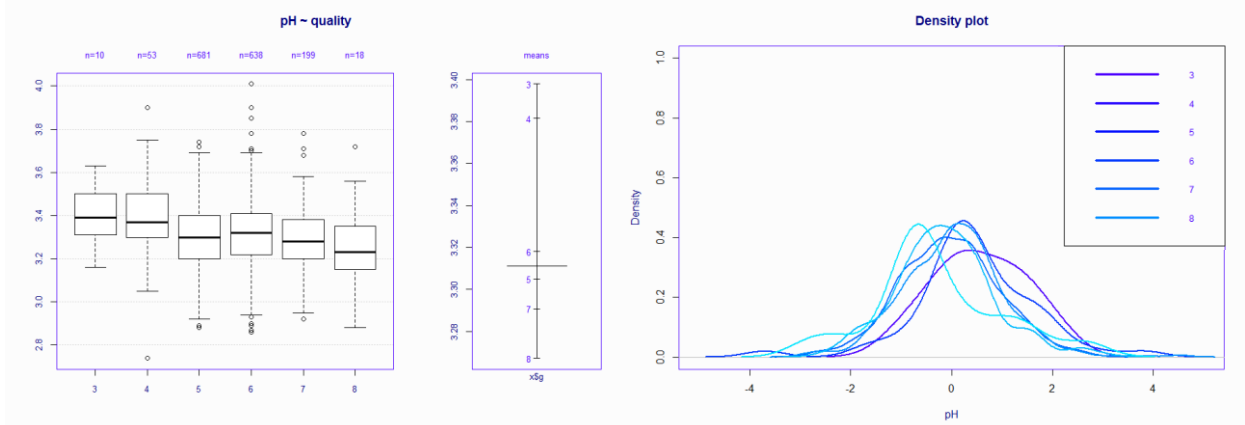
quality	min	Q1	mean	median	Q3	max	sd	skewness	kurtosis
3	0.99471	0.99615	0.99746	0.99757	0.99877	1.00080	0.00200	0.04759	-1.38571
4	0.99340	0.99565	0.99654	0.99650	0.99745	1.00100	0.00158	0.41374	0.25937
5	0.99256	0.99620	0.99710	0.99700	0.99790	1.00315	0.00159	0.47446	1.35329
6	0.99007	0.99540	0.99662	0.99656	0.99789	1.00369	0.00200	0.04706	0.68982
7	0.99064	0.99477	0.99610	0.99577	0.99736	1.00320	0.00218	0.33801	0.34659
8	0.99080	0.99418	0.99521	0.99494	0.99720	0.99880	0.00238	-0.20066	-1.15531



와인 등급별로 density 에 대한 boxplot과 density plot을 보면 3-8등급 모두 약간 다른 분포를 가지고 있지만 특정한 패턴이 보이지는 않는다.

9. "pH"

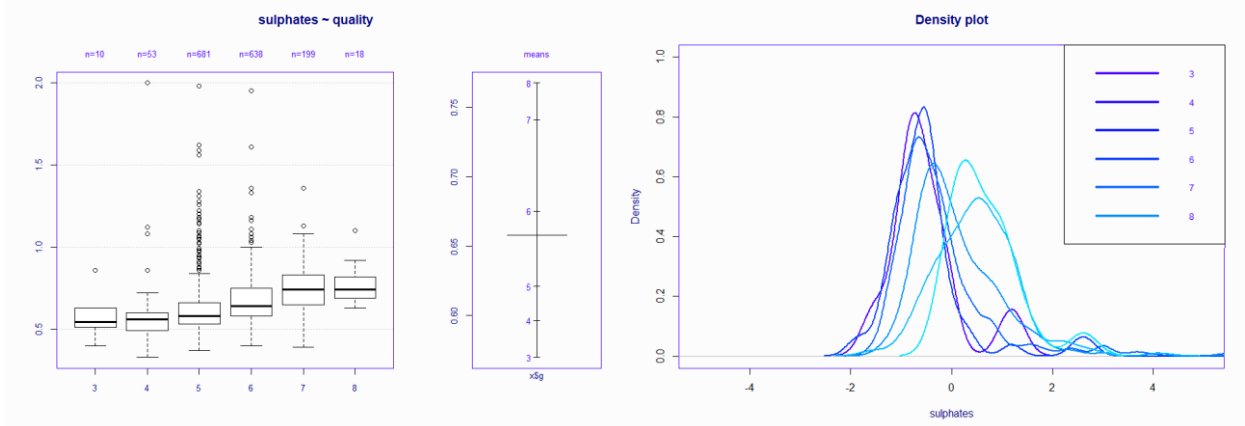
quality	min	Q1	mean	median	Q3	max	sd	skewness	kurtosis
3	3.16	3.3125	3.398	3.39	3.495	3.63	0.14405	-0.00914	-1.28854
4	2.74	3.3	3.38151	3.37	3.5	3.9	0.18144	-0.29306	2.35444
5	2.88	3.2	3.30495	3.3	3.4	3.74	0.15062	0.05305	-0.01996
6	2.86	3.22	3.31807	3.32	3.41	4.01	0.15400	0.30138	1.44202
7	2.92	3.2	3.29075	3.28	3.38	3.78	0.15010	0.37717	0.58982
8	2.88	3.1625	3.26722	3.23	3.35	3.72	0.20064	0.35404	-0.13123



와인 등급별로 pH 에 대한 boxplot과 density plot을 보면 3-8등급 모두 비슷한 분포를 가지고 있다.

10. "sulphates"

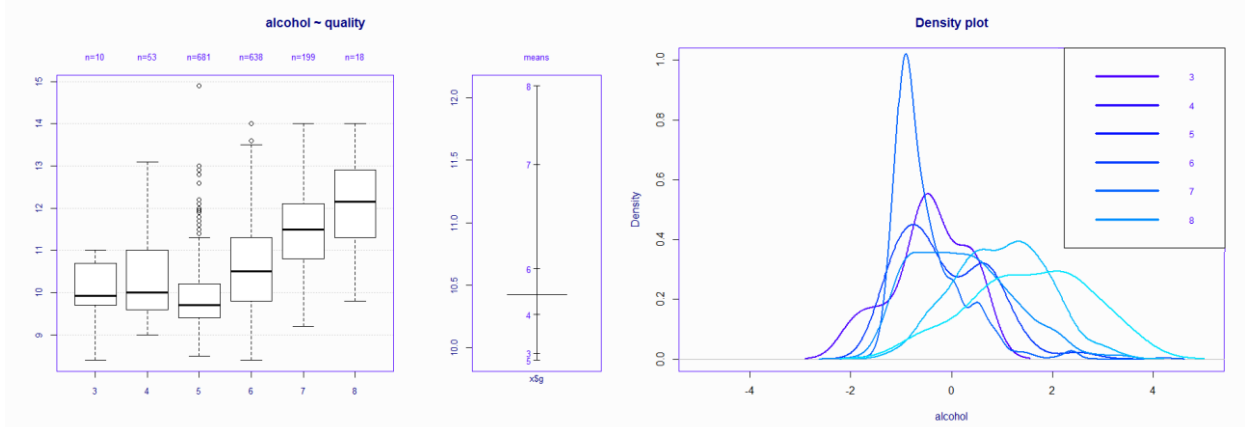
quality	min	Q1	mean	median	Q3	max	sd	skewness	kurtosis
3	0.4	0.5125	0.57	0.545	0.615	0.86	0.12202	1.04693	0.60660
4	0.33	0.49	0.59642	0.56	0.6	2	0.23939	4.09809	20.14243
5	0.37	0.53	0.62097	0.58	0.66	1.98	0.17106	2.80671	12.03771
6	0.4	0.58	0.67533	0.64	0.75	1.95	0.15865	2.70491	15.13806
7	0.39	0.65	0.74126	0.74	0.83	1.36	0.13564	0.59990	1.93595
8	0.63	0.69	0.76778	0.74	0.82	1.1	0.11538	1.22220	1.29577



와인 등급별로 sulphates에 대한 boxplot과 density plot을 보면 3-6등급 모두 비슷한 분포를 가지고 있다. 그리고 7-8등급이 비슷한 분포를 가지고 있다.

11. "alcohol"

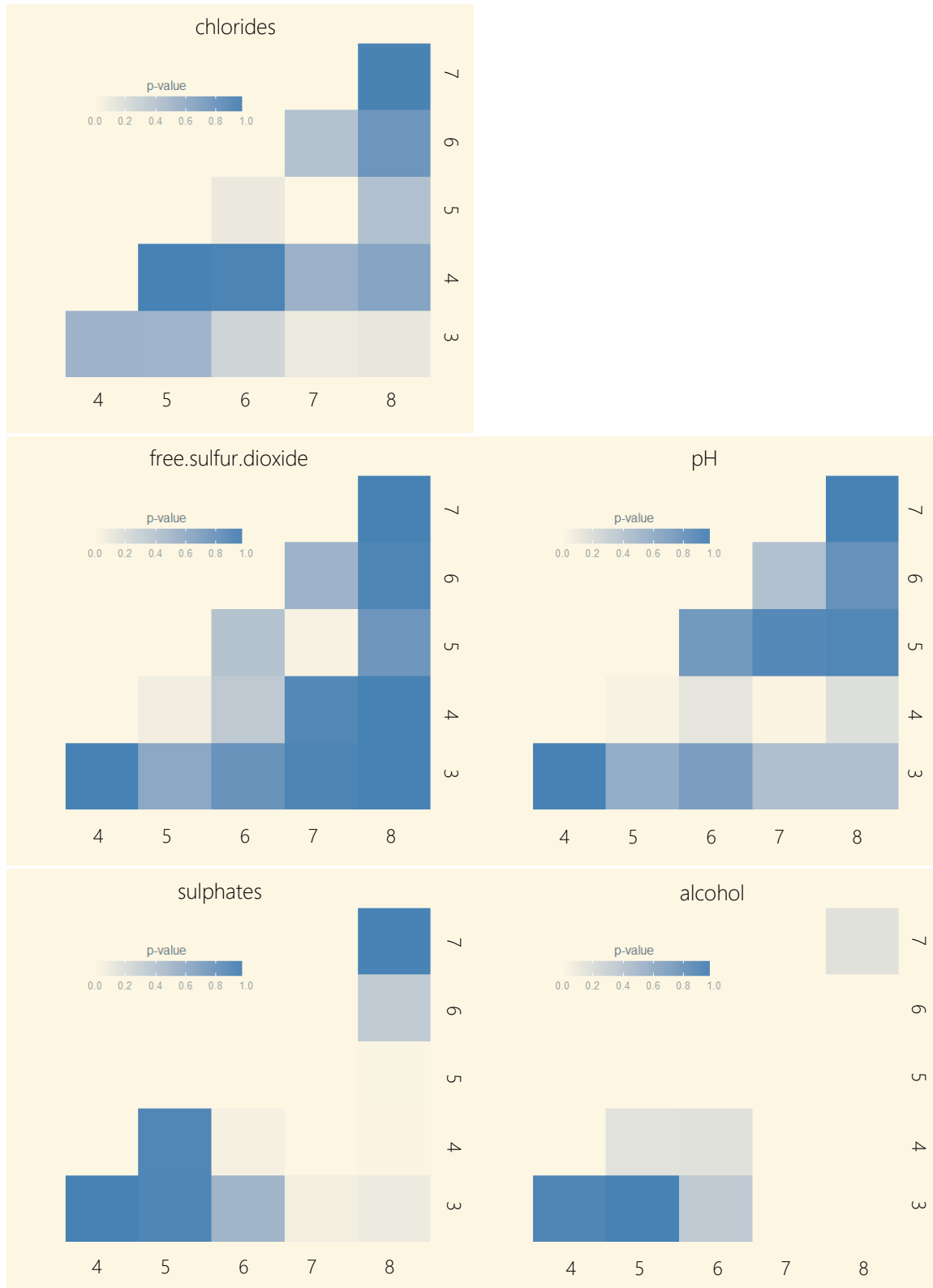
quality	min	Q1	mean	median	Q3	max	sd	skewness	kurtosis
3	8.4	9.725	9.955	9.925	10.575	11	0.81801	-0.40891	-0.99326
4	9	9.6	10.26509	10	11	13.1	0.93478	0.61327	-0.23206
5	8.5	9.4	9.89971	9.7	10.2	14.9	0.73652	1.83042	5.25442
6	8.4	9.8	10.62952	10.5	11.3	14	1.04964	0.54443	-0.15799
7	9.2	10.8	11.46591	11.5	12.1	14	0.96193	0.01033	-0.46770
8	9.8	11.325	12.09444	12.15	12.875	14	1.22401	-0.19693	-0.97768



와인 등급별로 alcohol 에 대한 boxplot과 density plot을 보면 3-8등급 모두 약간 다른 분포를 가지고 있지만 특정한 패턴이 보이지는 않는다.

B. 각 변수가 등급에 따라 차이가 있는가에 대한 검정

Levene등분산 검정을 해서 등분산 가정을 만족하는 변수들의 anova 검정을 하면 총 5개의 변수("chlorides", "free.sulfur.dioxide", "pH", "sulphates", "alcohol")들이 등급에 따라 유의미한 차이가 있다고 나온다. anova 사후검정을 하고, 그 결과로 각 등급 간의 변수의 평균에 차이가 있는지 검정에 대한 p-value를 매트릭스에 저장해서 plot을 그리면 아래와 같다.



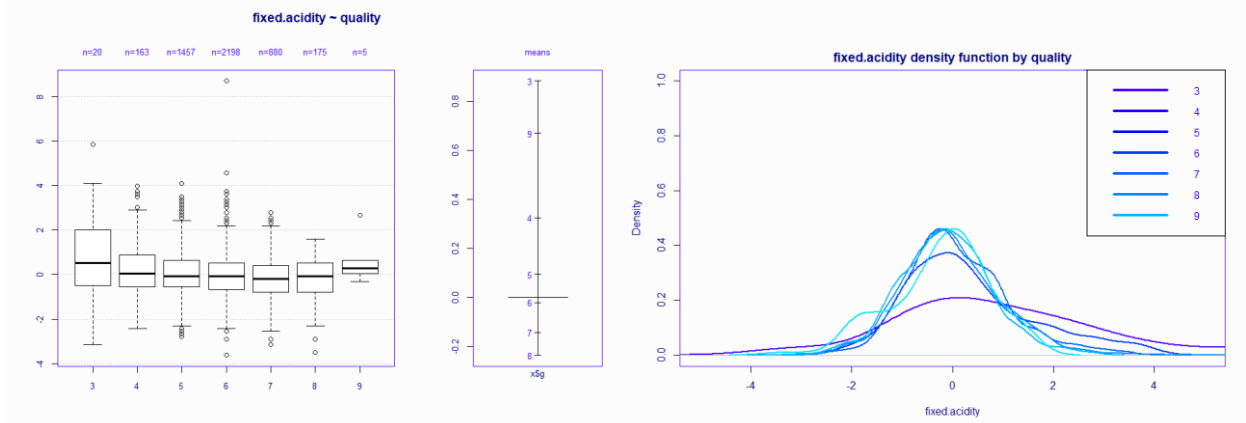
Wine Quality - WHITE

종속변수	"quality"
설명변수(11개)	"fixed.acidity", "volatile.acidity", "citric.acid", "residual.sugar", "chlorides", "free.sulfur.dioxide", "total.sulfur.dioxide", "density", "pH", "sulphates", "alcohol"

A. 등급별로 각 변수에 대한 특징 파악

1. "fixed.acidity"

quality	min	Q1	mean	median	Q3	max	sd	skewness	kurtosis
3	4.2	6.575	7.6	7.3	8.525	11.8	1.724743	0.473741	0.058556
4	4.8	6.4	7.129448	6.9	7.6	10.2	1.08149	0.762176	0.219055
5	4.5	6.4	6.933974	6.8	7.4	10.3	0.838433	0.595548	0.805067
6	3.8	6.3	6.837671	6.8	7.3	14.2	0.837442	0.748263	3.259482
7	4.2	6.2	6.734716	6.7	7.2	9.2	0.75567	0.140023	0.415816
8	3.9	6.2	6.657143	6.8	7.3	8.2	0.819337	-0.50186	0.066542
9	6.6	6.9	7.42	7.1	7.4	9.1	0.983362	0.844849	-1.18163



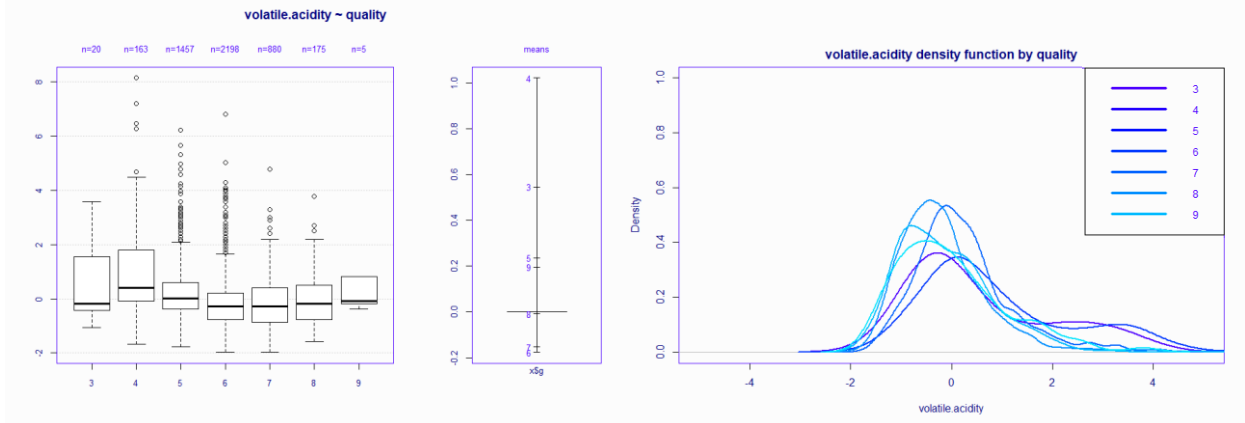
Wine Quality – WHITE 데이터의 모든 boxplot과 density plot은 변수의 값을 scale하고 그렸다.

3등급과 9등급의 데이터는 관측치의 수가 적다. 다른 등급과 분포를 비교하는데 참고해야한다.

와인 등급별로 fixed.acidity 에 대한 boxplot과 density plot을 보면 4-9등급 모두 비슷한 분포를 가지고 있다. 3등급만 조금 더 넓은 분포를 가진다.

2. "volatile.acidity"

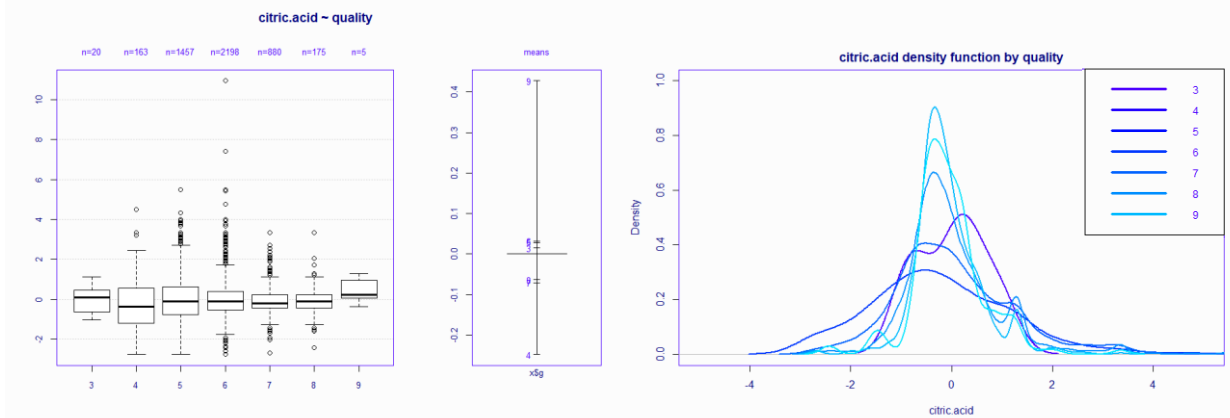
quality	min	Q1	mean	median	Q3	max	sd	skewness	kurtosis
3	0.17	0.2375	0.33325	0.26	0.4125	0.64	0.140827	0.88102	-0.684
4	0.11	0.27	0.381227	0.32	0.46	1.1	0.173463	1.37504	2.152822
5	0.1	0.24	0.302011	0.28	0.34	0.905	0.100066	1.426031	3.751129
6	0.08	0.2	0.260564	0.25	0.3	0.965	0.088142	1.531597	5.030737
7	0.08	0.19	0.262767	0.25	0.32	0.76	0.091106	0.808666	0.983864
8	0.12	0.2	0.2774	0.26	0.33	0.66	0.108029	0.974598	0.86169
9	0.24	0.26	0.298	0.27	0.36	0.36	0.057619	0.214034	-2.20945



와인 등급별로 volatile.acidity 에 대한 boxplot과 density plot을 보면 3-9등급 모두 비슷한 분포를 가지고 있다.

3. "citric.acid"

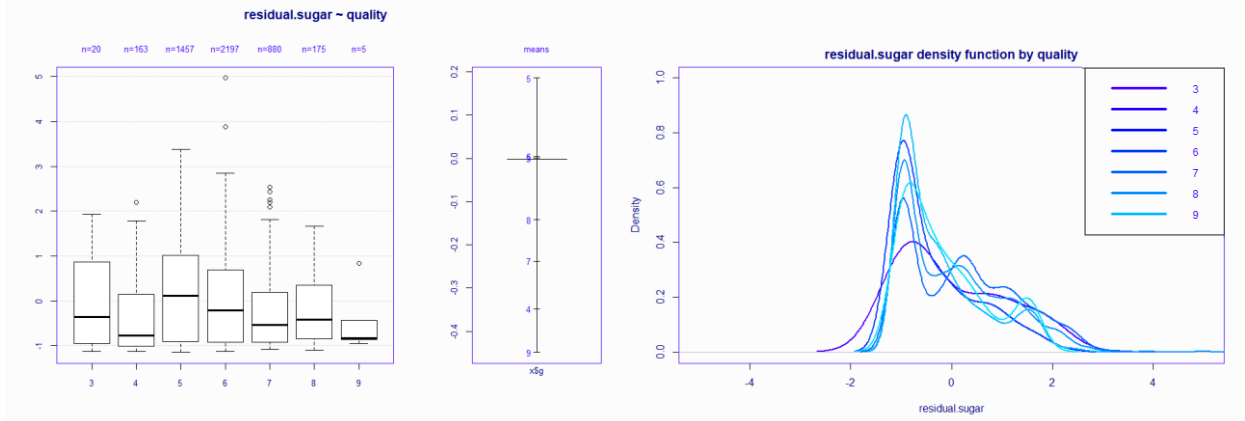
quality	min	Q1	mean	median	Q3	max	sd	skewness	kurtosis
3	0.21	0.2575	0.336	0.345	0.385	0.47	0.08146	-0.02828	-1.31016
4	0	0.19	0.304233	0.29	0.4	0.88	0.163857	0.463521	0.375582
5	0	0.24	0.337653	0.32	0.41	1	0.140814	0.701459	0.991756
6	0	0.27	0.338025	0.32	0.38	1.66	0.119325	1.995303	11.68706
7	0.01	0.28	0.325625	0.31	0.36	0.74	0.079183	0.750046	2.837847
8	0.04	0.28	0.326514	0.32	0.36	0.74	0.085439	0.536132	3.773468
9	0.29	0.34	0.386	0.36	0.45	0.49	0.082037	0.140321	-2.01251



와인 등급별로 citric.acid 에 대한 boxplot과 density plot을 보면 4-9등급으로 갈수록 분포가 조금씩 좁아진다.

4. "residual.sugar"

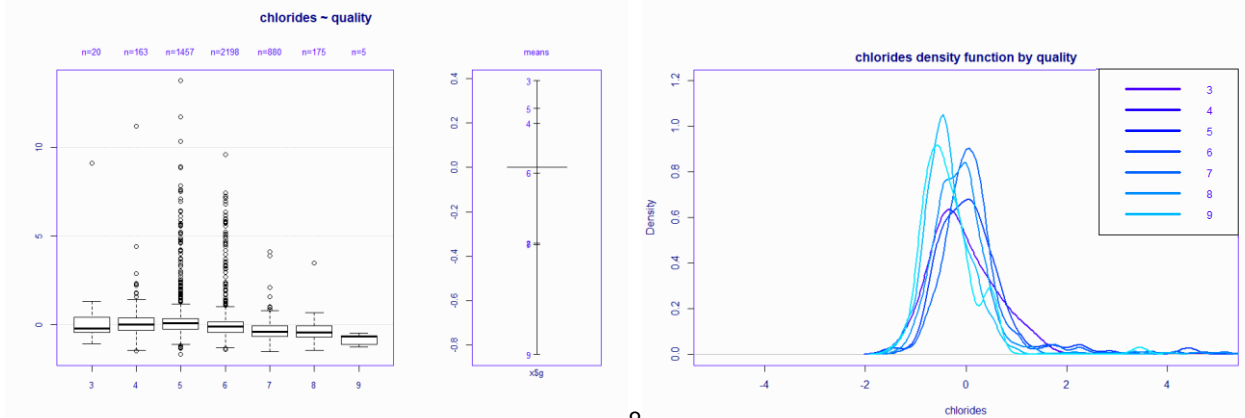
quality	min	Q1	mean	median	Q3	max	sd	skewness	kurtosis
3	0.7	1.5875	6.3925	4.6	10.7	16.2	5.317332	0.567256	-1.23329
4	0.7	1.3	4.628221	2.5	7.1	17.55	4.16361	1.090376	0.114669
5	0.6	1.8	7.334969	7	11.5	23.5	5.327619	0.486383	-0.79201
6	0.7	1.7	6.441606	5.3	9.9	65.8	5.166625	1.425423	7.626049
7	0.9	1.7	5.186477	3.65	7.325	19.25	4.297989	1.129676	0.236029
8	0.8	2.1	5.671429	4.3	8.2	14.8	4.263391	0.830933	-0.54598
9	1.6	2	4.12	2.2	4.2	10.6	3.759255	0.901593	-1.16005



와인 등급별로 residual.sugar 에 대한 boxplot과 density plot을 보면 3-9등급 모두 비슷한 분포를 가지고 있다.

5. "chlorides"

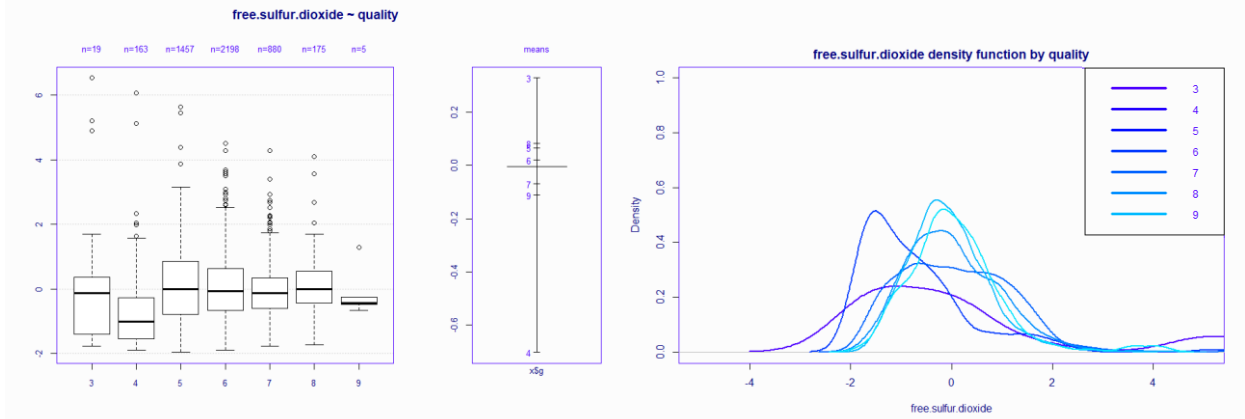
quality	min	Q1	mean	median	Q3	max	sd	skewness	kurtosis
3	0.022	0.03625	0.0543	0.041	0.054	0.244	0.046468	3.357266	10.91499
4	0.013	0.038	0.050098	0.046	0.054	0.29	0.025888	5.539899	44.68539
5	0.009	0.04	0.051546	0.047	0.053	0.346	0.026496	4.746673	31.04742
6	0.015	0.036	0.045217	0.043	0.049	0.255	0.020453	4.605542	28.75659
7	0.012	0.031	0.038191	0.037	0.044	0.135	0.010697	1.988853	13.61093
8	0.014	0.03	0.038314	0.036	0.044	0.121	0.013164	2.910764	15.55594
9	0.018	0.021	0.0274	0.031	0.032	0.035	0.007436	-0.24792	-2.12121



와인 등급별로 chlorides 에 대한 boxplot과 density plot을 보면 3-9등급 모두 비슷한 분포를 가지고 있다.

6. "free.sulfur.dioxide"

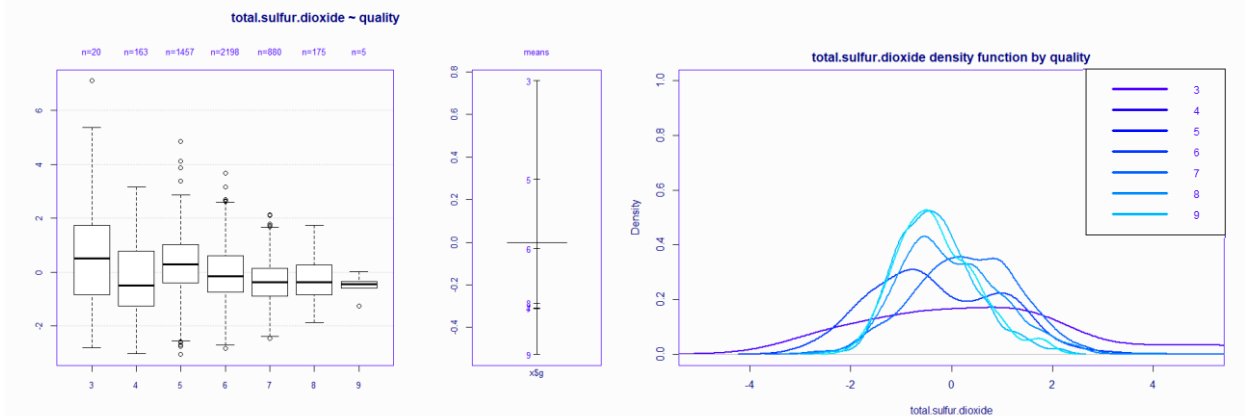
quality	min	Q1	mean	median	Q3	max	sd	skewness	kurtosis
3	5	13.25	53.325	33.5	47.5	289	69.42078	2.059403	3.967251
4	3	9	23.3589	18	30.5	138.5	20.39135	2.319598	8.201079
5	2	22	36.43205	35	50	131	18.14599	0.476534	0.386136
6	3	24	35.65059	34	46	112	15.73568	0.679683	0.778059
7	5	25	34.12557	33	41	108	13.24474	0.803375	1.71148
8	6	28	36.72	35	44.5	105	16.20367	1.439538	3.983473
9	24	27	33.4	28	31	57	13.42758	0.981435	-1.02773



와인 등급별로 free.sulfur.dioxide 에 대한 boxplot과 density plot을 보면 3-8등급 모두 약간 다른 분포를 가지고 있지만 특정한 패턴이 보이지는 않는다.

7. "total.sulfur.dioxide"

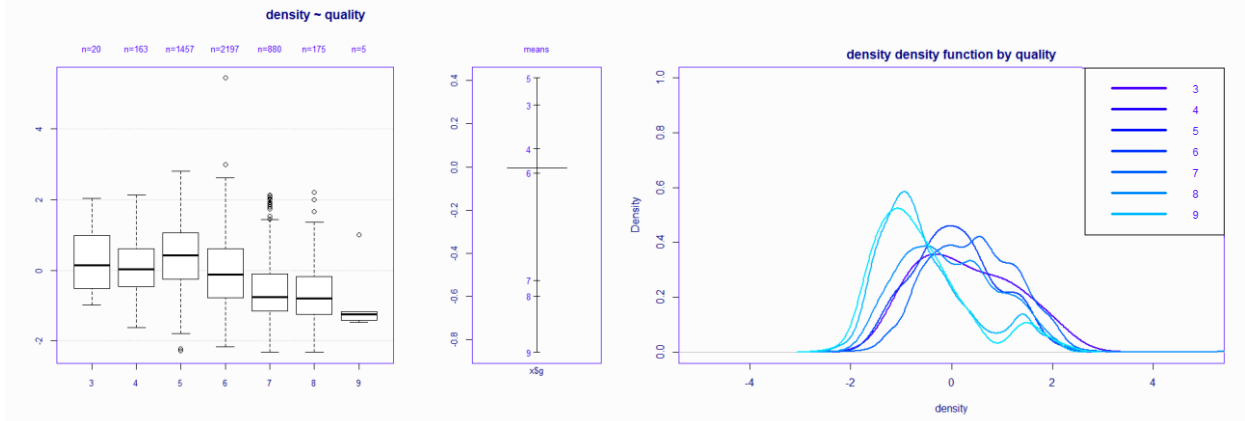
quality	min	Q1	mean	median	Q3	max	sd	skewness	kurtosis
3	19	105.75	170.6	159.5	210	440	107.7583	0.810716	0.095231
4	10	85	125.2791	117	171.5	272	52.75377	0.206418	-0.69121
5	9	121	150.9046	151	182	344	44.08619	-0.0317	0.073808
6	18	107.25	137.0473	132	164	294	41.28622	0.355914	-0.14955
7	34	101	125.1148	122	144.25	229	32.74298	0.500365	0.27305
8	59	102.5	126.1657	122	150	212.5	33.00633	0.536932	-0.0024
9	85	113	116	119	124	139	19.82423	-0.43928	-1.43622



와인 등급별로 total.sulfur.dioxide 에 대한 boxplot과 density plot을 보면 3-8등급 모두 약간 다른 분포를 가지고 있지만 특정한 패턴이 보이지는 않는다.

8. "density"

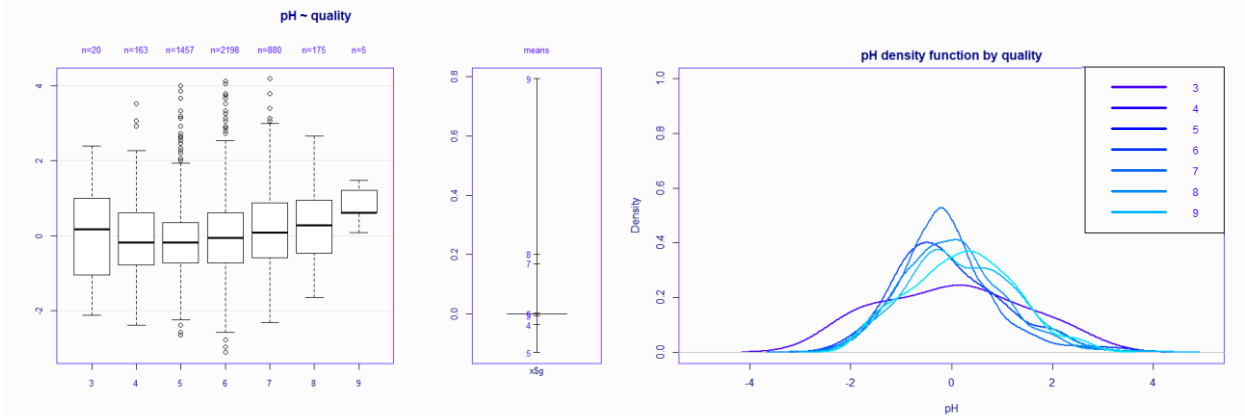
quality	min	Q1	mean	median	Q3	max	sd	skewness	kurtosis
3	0.9911	0.99252	0.994884	0.994425	0.996923	1.0001	0.002831	0.305667	-1.24112
4	0.9892	0.99264	0.994277	0.9941	0.99584	1.0004	0.002462	0.181955	-0.62478
5	0.98722	0.9933	0.995263	0.9953	0.9972	1.00241	0.002545	0.061411	-0.57947
6	0.98758	0.9917	0.993961	0.99366	0.99587	1.03898	0.003024	1.850747	21.8922
7	0.98711	0.99058	0.992452	0.99176	0.9937	1.0004	0.002768	0.977472	0.332863
8	0.98713	0.990305	0.992236	0.99164	0.993485	1.0006	0.002788	1.116485	0.799742
9	0.98965	0.9898	0.99146	0.9903	0.99055	0.997	0.003118	1.036894	-0.96365



와인 등급별로 density 에 대한 boxplot과 density plot을 보면 3-9등급 모두 비슷한 분포를 가지고 있다.

9. "pH"

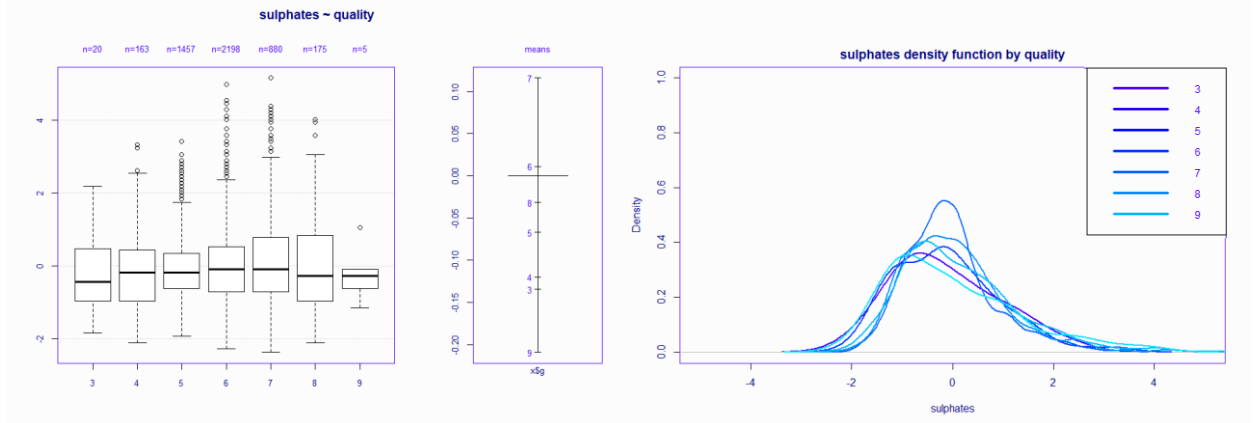
quality	min	Q1	mean	median	Q3	max	sd	skewness	kurtosis
3	2.87	3.035	3.1875	3.215	3.325	3.55	0.209834	0.08124	-1.18876
4	2.83	3.07	3.182883	3.16	3.28	3.72	0.163691	0.6527	0.300166
5	2.79	3.08	3.168833	3.16	3.24	3.79	0.140895	0.651182	1.426471
6	2.72	3.08	3.188599	3.18	3.28	3.81	0.150996	0.422826	0.581041
7	2.84	3.1	3.213898	3.2	3.32	3.82	0.158402	0.269148	-0.13354
8	2.94	3.12	3.218686	3.23	3.33	3.59	0.152289	0.088281	-0.51468
9	3.2	3.28	3.308	3.28	3.37	3.41	0.082885	-0.00143	-1.89697



와인 등급별로 pH 에 대한 boxplot과 density plot을 보면 4-9등급 모두 비슷한 분포를 가지고 있다. 3등급만 약간 넓은 분포를 가진다.

10. "sulphates"

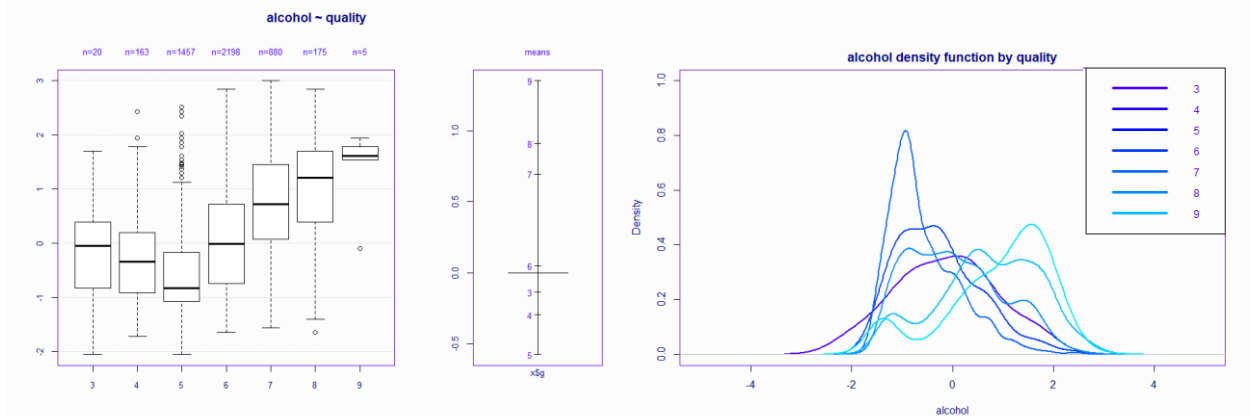
quality	min	Q1	mean	median	Q3	max	sd	skewness	kurtosis
3	0.28	0.38	0.4745	0.44	0.5425	0.74	0.119845	0.462408	-0.72534
4	0.25	0.38	0.476135	0.47	0.54	0.87	0.117973	0.688353	0.440207
5	0.27	0.42	0.482203	0.47	0.53	0.88	0.098205	0.844652	1.092119
6	0.23	0.41	0.491106	0.48	0.55	1.06	0.113357	0.979109	1.691733
7	0.22	0.41	0.503102	0.48	0.58	1.08	0.130195	0.950781	1.197369
8	0.25	0.38	0.486229	0.46	0.585	0.95	0.147073	0.956123	0.544029
9	0.36	0.42	0.466	0.46	0.48	0.61	0.092628	0.427847	-1.47657



와인 등급별로 sulphates 에 대한 boxplot과 density plot을 보면 3-9등급 모두 비슷한 분포를 가지고 있다.

11. "alcohol"

quality	min	Q1	mean	median	Q3	max	sd	skewness	kurtosis
3	8	9.55	10.345	10.45	11	12.6	1.224089	0.015478	-0.82687
4	8.4	9.4	10.15245	10.1	10.75	13.5	1.003217	0.697469	0.153388
5	8	9.2	9.80884	9.5	10.3	13.6	0.847065	1.078928	1.074779
6	8.5	9.6	10.57537	10.5	11.4	14	1.147776	0.397675	-0.71507
7	8.6	10.6	11.36794	11.4	12.3	14.2	1.246536	-0.30466	-0.56222
8	8.5	11	11.636	12	12.6	14	1.280138	-0.89139	0.011718
9	10.4	12.4	12.18	12.5	12.7	12.9	1.01341	-0.97667	-1.02917



와인 등급별로 alcohol 에 대한 boxplot과 density plot을 보면 3-8등급 모두 약간 다른 분포를 가지고 있지만 특정한 패턴이 보이지는 않는다.

사용한 R code

데이터 불러오기

```
Skillcraft <- read.table('SkillCraft1_Dataset.csv', header = T, sep = ',', stringsAsFactors = F)
Skillcraftmm <- read.table('SkillCraft1_Dataset.csv', header = T, sep = ',', stringsAsFactors = F)
which(Skillcraftmm$Age == '?')
which(Skillcraftmm$HoursPerWeek == '?')
which(Skillcraftmm$TotalHours == '?')
```

?를 NA처리하기

```
Skillcraft$Age[Skillcraft$Age == '?'] = NA
Skillcraft$HoursPerWeek[Skillcraft$HoursPerWeek == '?'] = NA
Skillcraft$TotalHours[Skillcraft$TotalHours == '?'] = NA
```

범주형 변수 범주화

```
Skillcraft$LeagueIndex = as.factor(Skillcraft$LeagueIndex)
```

문자형 변수 숫자형 변수로 변환

```
Skillcraft$Age = as.integer(Skillcraft$Age)
Skillcraft$HoursPerWeek = as.integer(Skillcraft$HoursPerWeek)
Skillcraft$TotalHours = as.integer(Skillcraft$TotalHours)
```

```
League_name <- c('Bronze', 'Silver', 'Gold', 'Platinum', 'Diamond', 'Master', 'GrandMaster', 'Professional')
```

기초통계 코드

```
Skillcraft %>%
  group_by(LeagueIndex) %>%
  summarize(min = min(Age, na.rm = TRUE), # 최솟값
             first_Qu = quantile(Age, probs=0.25, na.rm = TRUE), # first quantile
             mean = mean(Age, na.rm = TRUE), # 평균
             median = median(Age, na.rm = TRUE), # 중앙값
             third_Qu = quantile(Age, probs=0.75, na.rm = TRUE), # third quantile
             max = max(Age, na.rm = TRUE), #최댓값
             sd = sd(Age, na.rm = TRUE), # 표준편차
             skewness = skewness(Age, na.rm = TRUE), # 왜도(정규분포->0)
             kurtosis = kurtosis(Age, na.rm = TRUE)) %>% # 첨도(정규분포->0)
  View()
```

density plot 코드

```
par(mfrow = c(1,1), bg = 'grey99', col = topo.colors(20, alpha = 1)[1], col.main = 'navyblue', col.axis = 'navyblue', col.lab = 'navyblue')
plot(density(Skillcraft[Skillcraft$LeagueIndex == 1, 'HoursPerWeek']), xlim = c(0, 100), ylim = c(0,0.08),
     lwd = 2, xlab="HoursPerWeek", main = "Density plot", col.axis = 'black', col = topo.colors(20, alpha = 1)[1])
for (i in 2:7){
  lines(density(Skillcraft[Skillcraft$LeagueIndex == i, 'HoursPerWeek'], na.rm = TRUE), lwd = 2, col = topo.colors(20, alpha = 1)[i])
}
legend("topright", League_name[1:7],
      col = topo.colors(20, alpha = 1),
      lwd = 3, cex = 0.75)
```

Boxplot

```
Desc(Age~LeagueIndex,Skillcraft,plot=TRUE)
```

```
# 등분산 검정
```

```
levene <- leveneTest(ComplexAbilitiesUsed~LeagueIndex, Data)
if(levene$`Pr(>F)`[1] <= 0.05){print('이분산')} else{print('등분산')}
```

```
# 등분산 일때 anova검정
```

```
anova = oneway.test(TotalMapExplored~LeagueIndex, data=Data, na.action=na.omit, var.equal= TRUE)
if(anova$p.value <= 0.05){print('차이있다')} else{print('차이없다')}
```

```
# anova 사후검정 : ScheffeTest
```

```
aov_test <- aov(UniqueUnitsMade~LeagueIndex, data=Data)
```

```
comparison <- ScheffeTest(aov_test, # ANOVA model
```

```
                        "LeagueIndex", # vector treatment applied to each experimental unit
```

```
                        conf.level=NA)
```

```
{
```

```
p_value <- comparison$LeagueIndex
```

```
# 매트릭스 그래프 그리는 코드
```

```
theme_set(theme_solarized())
```

```
ggplot(p_value.data, aes(X1, X2, fill = value)) +
```

```
  geom_tile() +
```

```
  scale_fill_gradient(name=expression("p-value"), low = "#fdf6e3", high = "steelblue",
                      breaks=seq(0, 1, by = 0.2), limits = c(0, 1)) +
```

```
  scale_x_discrete(expand = c(0, 0.5)) +
```

```
  scale_y_discrete(expand = c(0, 0.5)) +
```

```
  labs(x = "", y = "") +
```

```
  guides(fill = guide_colorbar(barwidth = 7, barheight = 1, title.position = "top",
                               title.hjust = 0.5)) +
```

```
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1),
```

```
        panel.grid.major = element_blank(),
```

```
        panel.border = element_blank(),
```

```
        panel.background = element_blank(),
```

```
        axis.ticks = element_blank(),
```

```
        legend.justification = c(1, 0),
```

```
        legend.position = c(0.5, 0.7),
```

```
        legend.direction = "horizontal") +
```

```
  guides(fill = guide_colorbar(barwidth = 10, barheight = 1, title.position = "top",
                               title.hjust = 0.5))
```