

Regressão Linear múltipla - 3º Estágio - Exemplo Credit

- Autor
- Lucas Oliveira Belmiro - 118210068
 - Suelen Samara da Silva Felix - 121210333

1. Introdução

Este relatório tem como objetivo ajustar um modelo de regressão linear múltipla para investigar a influência de características específicas associadas ao comportamento dos consumidores na base de dados Credit. A regressão será realizada utilizando a base de dados Credit, que pode ser carregada a partir do pacote ISLR em R.

A base de dados Credit está disponível no pacote ISLR (Introduction to Statistical Learning with Applications in R), que é uma das referências principais em aprendizado estatístico e inclui diversas bases de dados para aplicação prática dos conceitos discutidos no livro.

Referência James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning with Applications in R. Springer. Disponível em: <https://www.statlearning.com/>

1.1. Iniciando nossa base de dados

Após instalar as bibliotecas necessárias, vamos listar todas as variáveis da nossa base de dados Credit.

```
install.packages("ISLR")
install.packages("corrplot")

library(corrplot)
library("ISLR")
library(dplyr)
library(knitr)
library(ggplot2)

data(Credit)
kable(Credit)
```

	ID	Income	Limit	Rating	Cards	Age	Education	Gender	Student	Married	Ethnicity	Balance
1	1	14.891	3606	283	2	34	11	Male	No	Yes	Caucasian	333
2	2	106.025	6645	483	3	82	15	Female	Yes	Yes	Asian	903
3	3	104.593	7075	514	4	71	11	Male	No	No	Asian	580
4	4	148.924	9504	681	3	36	11	Female	No	No	Asian	964
5	5	55.882	4897	357	2	68	16	Male	No	Yes	Caucasian	331
6	6	80.180	8047	569	4	77	10	Male	No	No	Caucasian	1151
7	7	20.996	3388	259	2	37	12	Female	No	No	African American	203
8	8	71.408	7114	512	2	87	9	Male	No	No	Asian	872
9	9	15.125	3300	266	5	66	13	Female	No	No	Caucasian	279
10	10	71.061	6819	491	3	41	19	Female	Yes	Yes	African American	1350
11	11	63.095	8117	589	4	30	14	Male	No	Yes	Caucasian	1407
12	12	15.045	1311	138	3	64	16	Male	No	No	Caucasian	0
13	13	80.616	5308	394	1	57	7	Female	No	Yes	Asian	204
14	14	43.682	6922	511	1	49	9	Male	No	Yes	Caucasian	1081
15	15	19.144	3291	269	2	75	13	Female	No	No	African American	148
16	16	20.089	2525	200	3	57	15	Female	No	Yes	African American	0
17	17	53.598	3714	286	3	73	17	Female	No	Yes	African American	0
18	18	36.496	4378	339	3	69	15	Female	No	Yes	Asian	368
19	19	49.570	6384	448	1	28	9	Female	No	Yes	Asian	891
20	20	42.079	6626	479	2	44	9	Male	No	No	Asian	1048
21	21	17.700	2860	235	4	63	16	Female	No	No	Asian	89
22	22	37.348	6378	458	1	72	17	Female	No	No	Caucasian	968
23	23	20.103	2631	213	3	61	10	Male	No	Yes	African American	0
24	24	64.027	5179	398	5	48	8	Male	No	Yes	African American	411

	25	10.742	1757	156	3	57	15 Female	No	No	Caucasian		0
	26	14.090	4323	326	5	25	16 Female	No	Yes	African American		671
	27	42.471	3625	289	6	44	12 Female	Yes	No	Caucasian		654
	28	32.793	4534	333	2	44	16 Male	No	No	African American		467
	29	186.634	13414	949	2	41	14 Female	No	Yes	African American		1809
	30	26.813	5611	411	4	55	16 Female	No	No	Caucasian		915
	31	34.142	5666	413	4	47	5 Female	No	Yes	Caucasian		863
	32	28.941	2733	210	5	43	16 Male	No	Yes	Asian		0
	33	134.181	7838	563	2	48	13 Female	No	No	Caucasian		526
	34	31.367	1829	162	4	30	10 Male	No	Yes	Caucasian		0
	35	20.150	2646	199	2	25	14 Female	No	Yes	Asian		0
	36	23.350	2558	220	3	49	12 Female	Yes	No	Caucasian		419
	37	62.413	6457	455	2	71	11 Female	No	Yes	Caucasian		762
	38	30.007	6481	462	2	69	9 Female	No	Yes	Caucasian		1093
	39	11.795	3899	300	4	25	10 Female	No	No	Caucasian		531
	40	13.647	3461	264	4	47	14 Male	No	Yes	Caucasian		344
	41	34.950	3327	253	3	54	14 Female	No	No	African American		50
	42	113.659	7659	538	2	66	15 Male	Yes	Yes	African American		1155
	43	44.158	4763	351	2	66	13 Female	No	Yes	Asian		385
	44	36.929	6257	445	1	24	14 Female	No	Yes	Asian		976
	45	31.861	6375	469	3	25	16 Female	No	Yes	Caucasian		1120
	46	77.380	7569	564	3	50	12 Female	No	Yes	Caucasian		997
	47	19.531	5043	376	2	64	16 Female	Yes	Yes	Asian		1241
	48	44.646	4431	320	2	49	15 Male	Yes	Yes	Caucasian		797
	49	44.522	2252	205	6	72	15 Male	No	Yes	Asian		0
	50	43.479	4569	354	4	49	13 Male	Yes	Yes	African American		902
	51	36.362	5183	376	3	49	15 Male	No	Yes	African American		654
	52	39.705	3969	301	2	27	20 Male	No	Yes	African American		211
	53	44.205	5441	394	1	32	12 Male	No	Yes	Caucasian		607
	54	16.304	5466	413	4	66	10 Male	No	Yes	Asian		957
	55	15.333	1499	138	2	47	9 Female	No	Yes	Asian		0
	56	32.916	1786	154	2	60	8 Female	No	Yes	Asian		0
	57	57.100	4742	372	7	79	18 Female	No	Yes	Asian		379
	58	76.273	4779	367	4	65	14 Female	No	Yes	Caucasian		133
	59	10.354	3480	281	2	70	17 Male	No	Yes	Caucasian		333
	60	51.872	5294	390	4	81	17 Female	No	No	Caucasian		531
	61	35.510	5198	364	2	35	20 Female	No	No	Asian		631
	62	21.238	3089	254	3	59	10 Female	No	No	Caucasian		108
	63	30.682	1671	160	2	77	7 Female	No	No	Caucasian		0
	64	14.132	2998	251	4	75	17 Male	No	No	Caucasian		133
	65	32.164	2937	223	2	79	15 Female	No	Yes	African American		0
	66	12.000	4160	320	4	28	14 Female	No	Yes	Caucasian		602
	67	113.829	9704	694	4	38	13 Female	No	Yes	Asian		1388
	68	11.187	5099	380	4	69	16 Female	No	No	African American		889
	69	27.847	5619	418	2	78	15 Female	No	Yes	Caucasian		822
	70	49.502	6819	505	4	55	14 Male	No	Yes	Caucasian		1084
	71	24.889	3954	318	4	75	12 Male	No	Yes	Caucasian		357
	72	58.781	7402	538	2	81	12 Female	No	Yes	Asian		1103
	73	22.939	4923	355	1	47	18 Female	No	Yes	Asian		663
	74	23.989	4523	338	4	31	15 Male	No	No	Caucasian		601
	75	16.103	5390	418	4	45	10 Female	No	Yes	Caucasian		945
	76	33.017	3180	224	2	28	16 Male	No	Yes	African American		29
	77	30.622	3293	251	1	68	16 Male	Yes	No	Caucasian		532
	78	20.936	3254	253	1	30	15 Female	No	No	Asian		145
	79	110.968	6662	468	3	45	11 Female	No	Yes	Caucasian		391
	80	15.354	2101	171	2	65	14 Male	No	No	Asian		0
	81	27.369	3449	288	3	40	9 Female	No	Yes	Caucasian		162
	82	53.480	4263	317	1	83	15 Male	No	No	Caucasian		99
	83	23.672	4433	344	3	63	11 Male	No	No	Caucasian		503
	84	19.225	1433	122	3	38	14 Female	No	No	Caucasian		0
	85	43.540	2906	232	4	69	11 Male	No	No	Caucasian		0
	86	152.298	12066	828	4	41	12 Female	No	Yes	Asian		1779
	87	55.367	6340	448	1	33	15 Male	No	Yes	Caucasian		815
	88	11.741	2271	182	4	59	12 Female	No	No	Asian		0
	89	15.560	4307	352	4	57	8 Male	No	Yes	African American		579
	90	59.530	7518	543	3	52	9 Female	No	No	African American		1176
	91	20.191	5767	431	4	42	16 Male	No	Yes	African American		1023
	92	48.498	6040	456	3	47	16 Male	No	Yes	Caucasian		812
	93	30.733	2832	249	4	51	13 Male	No	No	Caucasian		0
	94	16.479	5435	388	2	26	16 Male	No	No	African American		937
	95	38.009	3075	245	3	45	15 Female	No	No	African American		0
	96	14.084	855	120	5	46	17 Female	No	Yes	African American		0
	97	14.312	5382	367	1	59	17 Male	Yes	No	Asian		1380
	98	26.067	3388	266	4	74	17 Female	No	Yes	African American		155
	99	36.295	2963	241	2	68	14 Female	Yes	No	African American		375
	100	83.851	8494	607	5	47	18 Male	No	No	Caucasian		1311
	101	21.153	3736	256	1	41	11 Male	No	No	Caucasian		298
	102	17.976	2433	190	3	70	16 Female	Yes	No	Caucasian		431
	103	68.713	7582	531	2	56	16 Male	Yes	No	Caucasian		1587
	104	146.183	9540	682	6	66	15 Male	No	No	Caucasian		1050
	105	15.846	4768	365	4	53	12 Female	No	No	Caucasian		745
	106	12.031	3182	259	2	58	18 Female	No	Yes	Caucasian		210
	107	16.819	1337	115	2	74	15 Male	No	Yes	Asian		0

108	39.110	3189	263	3	72	12 Male	No	No	Asian		0
109	107.986	6033	449	4	64	14 Male	No	Yes	Caucasian		227
110	13.561	3261	279	5	37	19 Male	No	Yes	Asian		297
111	34.537	3271	250	3	57	17 Female	No	Yes	Asian		47
112	28.575	2959	231	2	60	11 Female	No	No	African American		0
113	46.007	6637	491	4	42	14 Male	No	Yes	Caucasian		1046
114	69.251	6386	474	4	30	12 Female	No	Yes	Asian		768
115	16.482	3326	268	4	41	15 Male	No	No	Caucasian		271
116	40.442	4828	369	5	81	8 Female	No	No	African American		510
117	35.177	2117	186	3	62	16 Female	No	No	Caucasian		0
118	91.362	9113	626	1	47	17 Male	No	Yes	Asian		1341
119	27.039	2161	173	3	40	17 Female	No	No	Caucasian		0
120	23.012	1410	137	3	81	16 Male	No	No	Caucasian		0
121	27.241	1402	128	2	67	15 Female	No	Yes	Asian		0
122	148.080	8157	599	2	83	13 Male	No	Yes	Caucasian		454
123	62.602	7056	481	1	84	11 Female	No	No	Caucasian		904
124	11.808	1300	117	3	77	14 Female	No	No	African American		0
125	29.564	2529	192	1	30	12 Female	No	Yes	Caucasian		0
126	27.578	2531	195	1	34	15 Female	No	Yes	Caucasian		0
127	26.427	5533	433	5	50	15 Female	Yes	Yes	Asian		1404
128	57.202	3411	259	3	72	11 Female	No	No	Caucasian		0
129	123.299	8376	610	2	89	17 Male	Yes	No	African American		1259
130	18.145	3461	279	3	56	15 Male	No	Yes	African American		255
131	23.793	3821	281	4	56	12 Female	Yes	Yes	African American		868
132	10.726	1568	162	5	46	19 Male	No	Yes	Asian		0
133	23.283	5443	407	4	49	13 Male	No	Yes	African American		912
134	21.455	5829	427	4	80	12 Female	No	Yes	African American		1018
135	34.664	5835	452	3	77	15 Female	No	Yes	African American		835
136	44.473	3500	257	3	81	16 Female	No	No	African American		8
137	54.663	4116	314	2	70	8 Female	No	No	African American		75
138	36.355	3613	278	4	35	9 Male	No	Yes	Asian		187
139	21.374	2073	175	2	74	11 Female	No	Yes	Caucasian		0
140	107.841	10384	728	3	87	7 Male	No	No	African American		1597
141	39.831	6045	459	3	32	12 Female	Yes	Yes	African American		1425
142	91.876	6754	483	2	33	10 Male	No	Yes	Caucasian		605
143	103.893	7416	549	3	84	17 Male	No	No	Asian		669
144	19.636	4896	387	3	64	10 Female	No	No	African American		710
145	17.392	2748	228	3	32	14 Male	No	Yes	Caucasian		68
146	19.529	4673	341	2	51	14 Male	No	No	Asian		642
147	17.055	5110	371	3	55	15 Female	No	Yes	Caucasian		805
148	23.857	1501	150	3	56	16 Male	No	Yes	Caucasian		0
149	15.184	2420	192	2	69	11 Female	No	Yes	Caucasian		0
150	13.444	886	121	5	44	10 Male	No	Yes	Asian		0
151	63.931	5728	435	3	28	14 Female	No	Yes	African American		581
152	35.864	4831	353	3	66	13 Female	No	Yes	Caucasian		534
153	41.419	2120	184	4	24	11 Female	Yes	No	Caucasian		156
154	92.112	4612	344	3	32	17 Male	No	No	Caucasian		0
155	55.056	3155	235	2	31	16 Male	No	Yes	African American		0
156	19.537	1362	143	4	34	9 Female	No	Yes	Asian		0
157	31.811	4284	338	5	75	13 Female	No	Yes	Caucasian		429
158	56.256	5521	406	2	72	16 Female	Yes	Yes	Caucasian		1020
159	42.357	5550	406	2	83	12 Female	No	Yes	Asian		653
160	53.319	3000	235	3	53	13 Male	No	No	Asian		0
161	12.238	4865	381	5	67	11 Female	No	No	Caucasian		836
162	31.353	1705	160	3	81	14 Male	No	Yes	Caucasian		0
163	63.809	7530	515	1	56	12 Male	No	Yes	Caucasian		1086
164	13.676	2330	203	5	80	16 Female	No	No	African American		0
165	76.782	5977	429	4	44	12 Male	No	Yes	Asian		548
166	25.383	4527	367	4	46	11 Male	No	Yes	Caucasian		570
167	35.691	2880	214	2	35	15 Male	No	No	African American		0
168	29.403	2327	178	1	37	14 Female	No	Yes	Caucasian		0
169	27.470	2820	219	1	32	11 Female	No	Yes	Asian		0
170	27.330	6179	459	4	36	12 Female	No	Yes	Caucasian		1099
171	34.772	2021	167	3	57	9 Male	No	No	Asian		0
172	36.934	4270	299	1	63	9 Female	No	Yes	Caucasian		283
173	76.348	4697	344	4	60	18 Male	No	No	Asian		108
174	14.887	4745	339	3	58	12 Male	No	Yes	African American		724
175	121.834	10673	750	3	54	16 Male	No	No	African American		1573
176	30.132	2168	206	3	52	17 Male	No	No	Caucasian		0
177	24.050	2607	221	4	32	18 Male	No	Yes	Caucasian		0
178	22.379	3965	292	2	34	14 Female	No	Yes	Asian		384
179	28.316	4391	316	2	29	10 Female	No	No	Caucasian		453
180	58.026	7499	560	5	67	11 Female	No	No	Caucasian		1237
181	10.635	3584	294	5	69	16 Male	No	Yes	Asian		423
182	46.102	5180	382	3	81	12 Male	No	Yes	African American		516
183	58.929	6420	459	2	66	9 Female	No	Yes	African American		789
184	80.861	4090	335	3	29	15 Female	No	Yes	Asian		0
185	158.889	11589	805	1	62	17 Female	No	Yes	Caucasian		1448
186	30.420	4442	316	1	30	14 Female	No	No	African American		450
187	36.472	3806	309	2	52	13 Male	No	No	African American		188
188	23.365	2179	167	2	75	15 Male	No	No	Asian		0
189	83.869	7667	554	2	83	11 Male	No	No	African American		930
190	58.351	4411	326	2	85	16 Female	No	Yes	Caucasian		126

191	55.187	5352	385	4	50	17 Female	No	Yes	Caucasian		538
192	124.290	9560	701	3	52	17 Female	Yes	No	Asian		1687
193	28.508	3933	287	4	56	14 Male	No	Yes	Asian		336
194	130.209	10088	730	7	39	19 Female	No	Yes	Caucasian		1426
195	30.406	2120	181	2	79	14 Male	No	Yes	African American		0
196	23.883	5384	398	2	73	16 Female	No	Yes	African American		802
197	93.039	7398	517	1	67	12 Male	No	Yes	African American		749
198	50.699	3977	304	2	84	17 Female	No	No	African American		69
199	27.349	2000	169	4	51	16 Female	No	Yes	African American		0
200	10.403	4159	310	3	43	7 Male	No	Yes	Asian		571
201	23.949	5343	383	2	40	18 Male	No	Yes	African American		829
202	73.914	7333	529	6	67	15 Female	No	Yes	Caucasian		1048
203	21.038	1448	145	2	58	13 Female	No	Yes	Caucasian		0
204	68.206	6784	499	5	40	16 Female	Yes	No	African American		1411
205	57.337	5310	392	2	45	7 Female	No	No	Caucasian		456
206	10.793	3878	321	8	29	13 Male	No	No	Caucasian		638
207	23.450	2450	180	2	78	13 Male	No	No	Caucasian		0
208	10.842	4391	358	5	37	10 Female	Yes	Yes	Caucasian		1216
209	51.345	4327	320	3	46	15 Male	No	No	African American		230
210	151.947	9156	642	2	91	11 Female	No	Yes	African American		732
211	24.543	3206	243	2	62	12 Female	No	Yes	Caucasian		95
212	29.567	5309	397	3	25	15 Male	No	No	Caucasian		799
213	39.145	4351	323	2	66	13 Male	No	Yes	Caucasian		308
214	39.422	5245	383	2	44	19 Male	No	No	African American		637
215	34.909	5289	410	2	62	16 Female	No	Yes	Caucasian		681
216	41.025	4229	337	3	79	19 Female	No	Yes	Caucasian		246
217	15.476	2762	215	3	60	18 Male	No	No	Asian		52
218	12.456	5395	392	3	65	14 Male	No	Yes	Caucasian		955
219	10.627	1647	149	2	71	10 Female	Yes	Yes	Asian		195
220	38.954	5222	370	4	76	13 Female	No	No	Caucasian		653
221	44.847	5765	437	3	53	13 Female	Yes	No	Asian		1246
222	98.515	8760	633	5	78	11 Female	No	No	African American		1230
223	33.437	6207	451	4	44	9 Male	Yes	No	Caucasian		1549
224	27.512	4613	344	5	72	17 Male	No	Yes	Asian		573
225	121.709	7818	584	4	50	6 Male	No	Yes	Caucasian		701
226	15.079	5673	411	4	28	15 Female	No	Yes	Asian		1075
227	59.879	6906	527	6	78	15 Female	No	No	Caucasian		1032
228	66.989	5614	430	3	47	14 Female	No	Yes	Caucasian		482
229	69.165	4668	341	2	34	11 Female	No	No	African American		156
230	69.943	7555	547	3	76	9 Male	No	Yes	Asian		1058
231	33.214	5137	387	3	59	9 Male	No	No	African American		661
232	25.124	4776	378	4	29	12 Male	No	Yes	Caucasian		657
233	15.741	4788	360	1	39	14 Male	No	Yes	Asian		689
234	11.603	2278	187	3	71	11 Male	No	Yes	Caucasian		0
235	69.656	8244	579	3	41	14 Male	No	Yes	African American		1329
236	10.503	2923	232	3	25	18 Female	No	Yes	African American		191
237	42.529	4986	369	2	37	11 Male	No	Yes	Asian		489
238	60.579	5149	388	5	38	15 Male	No	Yes	Asian		443
239	26.532	2910	236	6	58	19 Female	No	Yes	Caucasian		52
240	27.952	3557	263	1	35	13 Female	No	Yes	Asian		163
241	29.705	3351	262	5	71	14 Female	No	Yes	Asian		148
242	15.602	906	103	2	36	11 Male	No	Yes	African American		0
243	20.918	1233	128	3	47	18 Female	Yes	Yes	Asian		16
244	58.165	6617	460	1	56	12 Female	No	Yes	Caucasian		856
245	22.561	1787	147	4	66	15 Female	No	No	Caucasian		0
246	34.509	2001	189	5	80	18 Female	No	Yes	African American		0
247	19.588	3211	265	4	59	14 Female	No	No	Asian		199
248	36.364	2220	188	3	50	19 Male	No	No	Caucasian		0
249	15.717	905	93	1	38	16 Male	Yes	Yes	Caucasian		0
250	22.574	1551	134	3	43	13 Female	Yes	Yes	Caucasian		98
251	10.363	2430	191	2	47	18 Female	No	Yes	Asian		0
252	28.474	3202	267	5	66	12 Male	No	Yes	Caucasian		132
253	72.945	8603	621	3	64	8 Female	No	No	Caucasian		1355
254	85.425	5182	402	6	60	12 Male	No	Yes	African American		218
255	36.508	6386	469	4	79	6 Female	No	Yes	Caucasian		1048
256	58.063	4221	304	3	50	8 Male	No	No	African American		118
257	25.936	1774	135	2	71	14 Female	No	No	Asian		0
258	15.629	2493	186	1	60	14 Male	No	Yes	Asian		0
259	41.400	2561	215	2	36	14 Male	No	Yes	Caucasian		0
260	33.657	6196	450	6	55	9 Female	No	No	Caucasian		1092
261	67.937	5184	383	4	63	12 Male	No	Yes	Asian		345
262	180.379	9310	665	3	67	8 Female	Yes	Yes	Asian		1050
263	10.588	4049	296	1	66	13 Female	No	Yes	Caucasian		465
264	29.725	3536	270	2	52	15 Female	No	No	African American		133
265	27.999	5107	380	1	55	10 Male	No	Yes	Caucasian		651
266	40.885	5013	379	3	46	13 Female	No	Yes	African American		549
267	88.830	4952	360	4	86	16 Female	No	Yes	Caucasian		15
268	29.638	5833	433	3	29	15 Female	No	Yes	Asian		942
269	25.988	1349	142	4	82	12 Male	No	No	Caucasian		0
270	39.055	5565	410	4	48	18 Female	No	Yes	Caucasian		772
271	15.866	3085	217	1	39	13 Male	No	No	Caucasian		136
272	44.978	4866	347	1	30	10 Female	No	No	Caucasian		436
273	30.413	3690	299	2	25	15 Female	Yes	No	Asian		728

274	16.751	4706	353	6	48	14 Male	Yes	No	Asian		1255
275	30.550	5869	439	5	81	9 Female	No	No	African American		967
276	163.329	8732	636	3	50	14 Male	No	Yes	Caucasian		529
277	23.106	3476	257	2	50	15 Female	No	No	Caucasian		209
278	41.532	5000	353	2	50	12 Male	No	Yes	Caucasian		531
279	128.040	6982	518	2	78	11 Female	No	Yes	Caucasian		250
280	54.319	3063	248	3	59	8 Female	Yes	No	Caucasian		269
281	53.401	5319	377	3	35	12 Female	No	No	African American		541
282	36.142	1852	183	3	33	13 Female	No	No	African American		0
283	63.534	8100	581	2	50	17 Female	No	Yes	Caucasian		1298
284	49.927	6396	485	3	75	17 Female	No	Yes	Caucasian		890
285	14.711	2047	167	2	67	6 Male	No	Yes	Caucasian		0
286	18.967	1626	156	2	41	11 Female	No	Yes	Asian		0
287	18.036	1552	142	2	48	15 Female	No	No	Caucasian		0
288	60.449	3098	272	4	69	8 Male	No	Yes	Caucasian		0
289	16.711	5274	387	3	42	16 Female	No	Yes	Asian		863
290	10.852	3907	296	2	30	9 Male	No	No	Caucasian		485
291	26.370	3235	268	5	78	11 Male	No	Yes	Asian		159
292	24.088	3665	287	4	56	13 Female	No	Yes	Caucasian		309
293	51.532	5096	380	2	31	15 Male	No	Yes	Caucasian		481
294	140.672	11200	817	7	46	9 Male	No	Yes	African American		1677
295	42.915	2532	205	4	42	13 Male	No	Yes	Asian		0
296	27.272	1389	149	5	67	10 Female	No	Yes	Caucasian		0
297	65.896	5140	370	1	49	17 Female	No	Yes	Caucasian		293
298	55.054	4381	321	3	74	17 Male	No	Yes	Asian		188
299	20.791	2672	204	1	70	18 Female	No	No	African American		0
300	24.919	5051	372	3	76	11 Female	No	Yes	African American		711
301	21.786	4632	355	1	50	17 Male	No	Yes	Caucasian		580
302	31.335	3526	289	3	38	7 Female	No	No	Caucasian		172
303	59.855	4964	365	1	46	13 Female	No	Yes	Caucasian		295
304	44.061	4970	352	1	79	11 Male	No	Yes	African American		414
305	82.706	7506	536	2	64	13 Female	No	Yes	Asian		905
306	24.460	1924	165	2	50	14 Female	No	Yes	Asian		0
307	45.120	3762	287	3	80	8 Male	No	Yes	Caucasian		70
308	75.406	3874	298	3	41	14 Female	No	Yes	Asian		0
309	14.956	4640	332	2	33	6 Male	No	No	Asian		681
310	75.257	7010	494	3	34	18 Female	No	Yes	Caucasian		885
311	33.694	4891	369	1	52	16 Male	Yes	No	African American		1036
312	23.375	5429	396	3	57	15 Female	No	Yes	Caucasian		844
313	27.825	5227	386	6	63	11 Male	No	Yes	Caucasian		823
314	92.386	7685	534	2	75	18 Female	No	Yes	Asian		843
315	115.520	9272	656	2	69	14 Male	No	No	African American		1140
316	14.479	3907	296	3	43	16 Male	No	Yes	Caucasian		463
317	52.179	7306	522	2	57	14 Male	No	No	Asian		1142
318	68.462	4712	340	2	71	16 Male	No	Yes	Caucasian		136
319	18.951	1485	129	3	82	13 Female	No	No	Caucasian		0
320	27.590	2586	229	5	54	16 Male	No	Yes	African American		0
321	16.279	1160	126	3	78	13 Male	Yes	Yes	African American		5
322	25.078	3096	236	2	27	15 Female	No	Yes	Caucasian		81
323	27.229	3484	282	6	51	11 Male	No	No	Caucasian		265
324	182.728	13913	982	4	98	17 Male	No	Yes	Caucasian		1999
325	31.029	2863	223	2	66	17 Male	Yes	Yes	Asian		415
326	17.765	5072	364	1	66	12 Female	No	Yes	Caucasian		732
327	125.480	10230	721	3	82	16 Male	No	Yes	Caucasian		1361
328	49.166	6662	508	3	68	14 Female	No	No	Asian		984
329	41.192	3673	297	3	54	16 Female	No	Yes	Caucasian		121
330	94.193	7576	527	2	44	16 Female	No	Yes	Caucasian		846
331	20.405	4543	329	2	72	17 Male	Yes	No	Asian		1054
332	12.581	3976	291	2	48	16 Male	No	Yes	Caucasian		474
333	62.328	5228	377	3	83	15 Male	No	No	Caucasian		380
334	21.011	3402	261	2	68	17 Male	No	Yes	African American		182
335	24.230	4756	351	2	64	15 Female	No	Yes	Caucasian		594
336	24.314	3409	270	2	23	7 Female	No	Yes	Caucasian		194
337	32.856	5884	438	4	68	13 Male	No	No	Caucasian		926
338	12.414	855	119	3	32	12 Male	No	Yes	African American		0
339	41.365	5303	377	1	45	14 Male	No	No	Caucasian		606
340	149.316	10278	707	1	80	16 Male	No	No	African American		1107
341	27.794	3807	301	4	35	8 Female	No	Yes	African American		320
342	13.234	3922	299	2	77	17 Female	No	Yes	Caucasian		426
343	14.595	2955	260	5	37	9 Male	No	Yes	African American		204
344	10.735	3746	280	2	44	17 Female	No	Yes	Caucasian		410
345	48.218	5199	401	7	39	10 Male	No	Yes	Asian		633
346	30.012	1511	137	2	33	17 Male	No	Yes	Caucasian		0
347	21.551	5380	420	5	51	18 Male	No	Yes	Asian		907
348	160.231	10748	754	2	69	17 Male	No	No	Caucasian		1192
349	13.433	1134	112	3	70	14 Male	No	Yes	Caucasian		0
350	48.577	5145	389	3	71	13 Female	No	Yes	Asian		503
351	30.002	1561	155	4	70	13 Female	No	Yes	Caucasian		0
352	61.620	5140	374	1	71	9 Male	No	Yes	Caucasian		302
353	104.483	7140	507	2	41	14 Male	No	Yes	African American		583
354	41.868	4716	342	2	47	18 Male	No	No	Caucasian		425
355	12.068	3873	292	1	44	18 Female	No	Yes	Asian		413
356	180.682	11966	832	2	58	8 Female	No	Yes	African American		1405

357	34.480	6090	442	3	36	14 Male	No	No	Caucasian		962
358	39.609	2539	188	1	40	14 Male	No	Yes	Asian		0
359	30.111	4336	339	1	81	18 Male	No	Yes	Caucasian		347
360	12.335	4471	344	3	79	12 Male	No	Yes	African American		611
361	53.566	5891	434	4	82	10 Female	No	No	Caucasian		712
362	53.217	4943	362	2	46	16 Female	No	Yes	Asian		382
363	26.162	5101	382	3	62	19 Female	No	No	African American		710
364	64.173	6127	433	1	80	10 Male	No	Yes	Caucasian		578
365	128.669	9824	685	3	67	16 Male	No	Yes	Asian		1243
366	113.772	6442	489	4	69	15 Male	Yes	Yes	Caucasian		790
367	61.069	7871	564	3	56	14 Male	No	Yes	Caucasian		1264
368	23.793	3615	263	2	70	14 Male	No	No	African American		216
369	89.000	5759	440	3	37	6 Female	No	No	Caucasian		345
370	71.682	8028	599	3	57	16 Male	No	Yes	Caucasian		1208
371	35.610	6135	466	4	40	12 Male	No	No	Caucasian		992
372	39.116	2150	173	4	75	15 Male	No	No	Caucasian		0
373	19.782	3782	293	2	46	16 Female	Yes	No	Caucasian		840
374	55.412	5354	383	2	37	16 Female	Yes	Yes	Caucasian		1003
375	29.400	4840	368	3	76	18 Female	No	Yes	Caucasian		588
376	20.974	5673	413	5	44	16 Female	No	Yes	Caucasian		1000
377	87.625	7167	515	2	46	10 Female	No	No	African American		767
378	28.144	1567	142	3	51	10 Male	No	Yes	Caucasian		0
379	19.349	4941	366	1	33	19 Male	No	Yes	Caucasian		717
380	53.308	2860	214	1	84	10 Male	No	Yes	Caucasian		0
381	115.123	7760	538	3	83	14 Female	No	No	African American		661
382	101.788	8029	574	2	84	11 Male	No	Yes	Caucasian		849
383	24.824	5495	409	1	33	9 Male	Yes	No	Caucasian		1352
384	14.292	3274	282	9	64	9 Male	No	Yes	Caucasian		382
385	20.088	1870	180	3	76	16 Male	No	No	African American		0
386	26.400	5640	398	3	58	15 Female	No	No	Asian		905
387	19.253	3683	287	4	57	10 Male	No	No	African American		371
388	16.529	1357	126	3	62	9 Male	No	No	Asian		0
389	37.878	6827	482	2	80	13 Female	No	No	Caucasian		1129
390	83.948	7100	503	2	44	18 Male	No	No	Caucasian		806
391	135.118	10578	747	3	81	15 Female	No	Yes	Asian		1393
392	73.327	6555	472	2	43	15 Female	No	No	Caucasian		721
393	25.974	2308	196	2	24	10 Male	No	No	Asian		0
394	17.316	1335	138	2	65	13 Male	No	No	African American		0
395	49.794	5758	410	4	40	8 Male	No	No	Caucasian		734
396	12.096	4100	307	3	32	13 Male	No	Yes	Caucasian		560
397	13.364	3838	296	5	65	17 Male	No	No	African American		480
398	57.872	4171	321	5	67	12 Female	No	Yes	Caucasian		138
399	37.728	2525	192	1	44	13 Male	No	Yes	Caucasian		0
400	18.701	5524	415	5	64	7 Female	No	No	Asian		966

2 . Descrevendo a base de dados

Nessa seção iremos analisar de maneira geral, os dados que foram fornecidas.

str(Credit)											
'data.frame': 400 obs. of 12 variables:											
\$ ID	: int	1 2 3	4 5	6 7 8	9 10 ...						
\$ Income	: num	14.9	106	104.6148.9	55.9 ...						
\$ Limit	: int	3606	6645	70759504	4897	8047	3388	7114	3300	6819 ...	
\$ Rating	: int	283	483	514	681	357	569	259	512	266 491 ...	
\$ Cards	: int	2 3 4	32422	53...							
\$ Age	: int	34	82	71	36	6877	37	87	66	41 ...	
\$ Education:	int	11	15	11	11	1610	12	9	13	19 ...	
\$ Gender	: Factor w/	2 levels " Male", "Female":				1212112122...					
▯ Student	: Factor w/ 2 levels "No", "Yes":	1	2	1	1	1	1	1	1	2 ...	
▯ Married	: Factor w/ 2 levels "No", "Yes":	2	2	1	1	2	1	1	1	2 ...	
▯ Ethnicity:	Factor w/ 3 levels "African American", ...:	3	2	2	2	3	3	1	2	3 1 ...	
▯ Balance	: int	333	903	580	964	331	1151	203	872	279 1350 ...	

Como vemos pelo código acima, é possível notar que temos 400 observações e 12 variáveis, dentre elas, temos:

ID: Um identificador numérico único para cada uma das observações realizadas.

Income: Variável numérica contínua que representa a renda em dólares das pessoas.

Limit: Variável numérica, representada por números inteiros, que refere-se ao limite máximo de crédito de uma pessoa.

Rating: É a variável quantitativa que indica a pontuação de crédito, semelhante ao score no padrão brasileiro. Ou seja, quanto maior o rating, melhor a capacidade de crédito e confiança do indivíduo em cumprir com obrigações financeiras.

Cards: Variável quantitativa que indica o número de cartões de crédito em posse de cada pessoa.

Age: Variável quantitativa que indica a idade das pessoas em anos.

Education: É uma variável quantitativa que indica o nível de escolaridade dos indivíduos em anos.

Gender: Variável qualitativa que indica o gênero das pessoas, categorizado como "Male" (masculino) ou "Female" (feminino).

Student: Variável qualitativa que indica se a pessoa é estudante ou não, com valores "Yes" (sim) ou "No" (não).

Married: Variável qualitativa que indica quando uma pessoa é casada, com valores "Yes" (casado) ou "No" (não casado).

Ethnicity: Variável qualitativa que indica a etnia das observações.

Balance: É uma variável quantitativa que indica o saldo devido das pessoas, em dólares.

3. Análise exploratória dos dados, outliers e correções na base de dados.

Iniciaremos limpando a base de dados, removendo a coluna ID que não será utilizada durante o restante deste trabalho.

```
# Filtrar o conjunto de dados excluindo a coluna "ID" df <- Credit %>%  
  
select(-ID) %>% transmute(renda  
= Income,  
                           limite = Limit, score =  
                           Rating, cartoes = Cards,  
                           idade = Age, educacao =  
                           Education, genero =  
                           Gender, estudante =  
                           Student, casado = Married,  
                           etnia = Ethnicity, saldo =  
                           Balance)
```

Para realizar uma análise exploratória de dados, iremos verificar também se existem valores ausentes no nosso dataframe e de outliers.

```
# Verifica possíveis valores ausentes  
colSums(is.na(Credit))
```

ID: 0 Income: 0 Limit: 0 Rating: 0 Cards: 0 Age: 0 Education: 0 Gender: 0 Student: 0 Married: 0 Ethnicity: 0 Balance: 0

```
install.packages("skimr", dependencies = TRUE)  
library(skimr)  
  
dados <- Credit %>%  
  select(-ID) %>%  
  transmute(renda = Income,  
            limite = Limit,  
            score = Rating,  
            saldo = Balance,  
            cartoes = Cards,  
            idade = Age,  
            educacao = Education,  
            genero = Gender,  
            estudante = Student,  
            casado = Married,  
            etnia = Ethnicity)  
  
skim(dados)
```

Data summary	
Name	dados
Number of rows	400
Number of columns	11

Column type frequency:








factor	4
numeric	7

Group variables None

Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
genero	0	1	FALSE	2	Fem: 207, Ma: 193
estudante	0	1	FALSE	2	No: 360, Yes: 40
casado	0	1	FALSE	2	No: 155, Yes: 245
etnia	0	1	FALSE	3	Cau: 199, Asi: 102, Afr: 99

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
renda	0	1	45.2	35.2	10.4	21.0	33.1	57.5	187	
limite	0	1	4736	2308	855	3088	4622	5873		
score	0	1	355	155	93	247	344	437	982	
saldo	0	1	520	460	0	68,8	460	863	1999	
cartoes	0	1	2,96	1,37	1	2	3	4	9	
idade	0	1	55,7	17,2	23	41,8	56	70	98	
educacao	0	1	13,4	3,13	5	11	14	16	20	

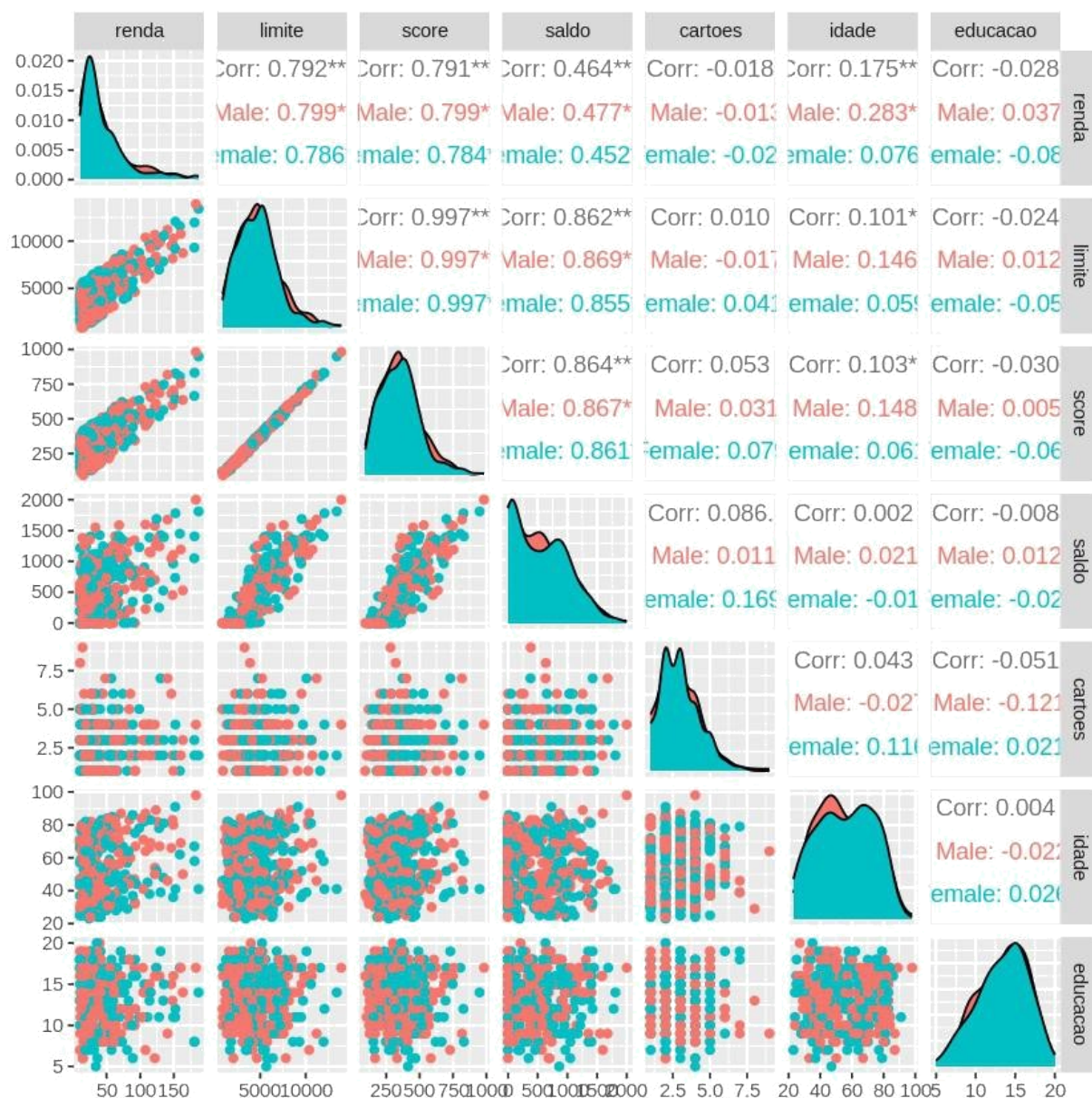
3.1 Análise de correlação

```
install.packages("GGally")
library(GGally)

### Multicolinearidade: r > 0.9 (ou 0.8)

graf1 <- ggpairs(dados, columns = 1:7, ggplot2::aes(colour=genero))

graf1
```

Através desses gráficos, podemos analisar a correlação com a variável resposta Balance (Saldo) e temos que:

- Renda:

Apresenta uma correlação significativa de 0,464*** com saldo. Isso indica uma relação positiva moderada, sugerindo que à medida que a renda aumenta, o saldo tende a aumentar também.

- Limite:

Tem uma correlação forte de 0,862*** com saldo, o que sugere que um limite maior está associado a um saldo mais elevado. Isso pode ser esperado, pois um maior limite de crédito pode proporcionar mais oportunidades de gastos e, consequentemente, um saldo mais elevado.

- Score:

A correlação entre "saldo" e "score" é de 0,864***. Isso indica que scores de crédito mais altos estão associados a saldos mais altos, reforçando a ideia de que consumidores com melhor crédito tendem a ter melhores condições financeiras.

- Cartões:

A correlação entre o número de cartões e o saldo é baixa, em torno de 0.086. Isso indica que o número de cartões de crédito de uma pessoa tem pouca influência direta sobre o saldo mantido.

- Idade:

A correlação entre "saldo" e "score" é de 0,864***. Isso indica que scores de crédito mais altos estão associados a saldos mais altos, reforçando a ideia de que consumidores com melhor crédito tendem a ter melhores condições financeiras.

- Educação:

A correlação entre "saldo" e "score" é de 0,864***. Isso indica que scores de crédito mais altos estão associados a saldos mais altos, reforçando a ideia de que consumidores com melhor crédito tendem a ter melhores condições financeira

4. Análises univariadas e bivariadas

```
install.packages("car")
library(GGally)
library(ggplot2)
library(dplyr)
library(car)

# Análise Univariada
# Estatísticas descritivas
summary(dados)

# Visualização das distribuições
ggplot(dados, aes(x = renda)) +
  geom_histogram(bins = 30, fill = "lightblue", color = "black") + ggtitle("Distribuição
  de Renda")

ggplot(dados, aes(x = limite)) +
  geom_histogram(bins = 30, fill = "lightgreen", color = "black") + ggtitle("Distribuição
  de Limite")

ggplot(dados, aes(x = score)) +
  geom_histogram(bins = 30, fill = "lightcoral", color = "black") + ggtitle("Distribuição
  de Score")

ggplot(dados, aes(x = cartoes)) +
  geom_bar(fill = "lightpink", color = "black") +
  ggtitle("Distribuição de Cartões")

# Análise Bivariada
# Matriz de correlação
cor_matrix <- cor(dados[, c("renda", "limite", "score", "cartoes", "saldo")])
print(cor_matrix)

# Gráficos de dispersão e correlação
ggpairs(dados, columns = c("renda", "limite", "score", "cartoes", "saldo"), aes(colour = casado))

# Gráficos de dispersão específicos
ggplot(dados,
  aes(x = renda, y = saldo)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", se = FALSE) +
  ggtitle("Dispersão entre Renda e Saldo")

ggplot(dados, aes(x = limite, y = saldo)) +
  geom_point(color = "green") +
  geom_smooth(method = "lm", se = FALSE) +
  ggtitle("Dispersão entre Limite e Saldo")

ggplot(dados, aes(x = score, y = saldo)) +
  geom_point(color = "red") +
  geom_smooth(method = "lm", se = FALSE) +
  ggtitle("Dispersão entre Score e Saldo")

ggplot(dados, aes(x = cartoes, y = saldo)) +
  geom_point(color = "purple") +
  geom_smooth(method = "lm", se = FALSE) +
  ggtitle("Dispersão entre Cartões e Saldo")

# Modelo de regressão linear
modelo <- lm(saldo ~ renda + limite + score + cartoes, data = dados)
# Exibir resultados de correlação de Pearson
cor_results_renda <- cor.test(dados$renda, dados$saldo)
print(cor_results_renda)

cor_results_limite <- cor.test(dados$limite, dados$saldo)
print(cor_results_limite)

cor_results_score <- cor.test(dados$score, dados$saldo)
print(cor_results_score)

cor_results_cartoes <- cor.test(dados$cartoes, dados$saldo)
print(cor_results_cartoes)

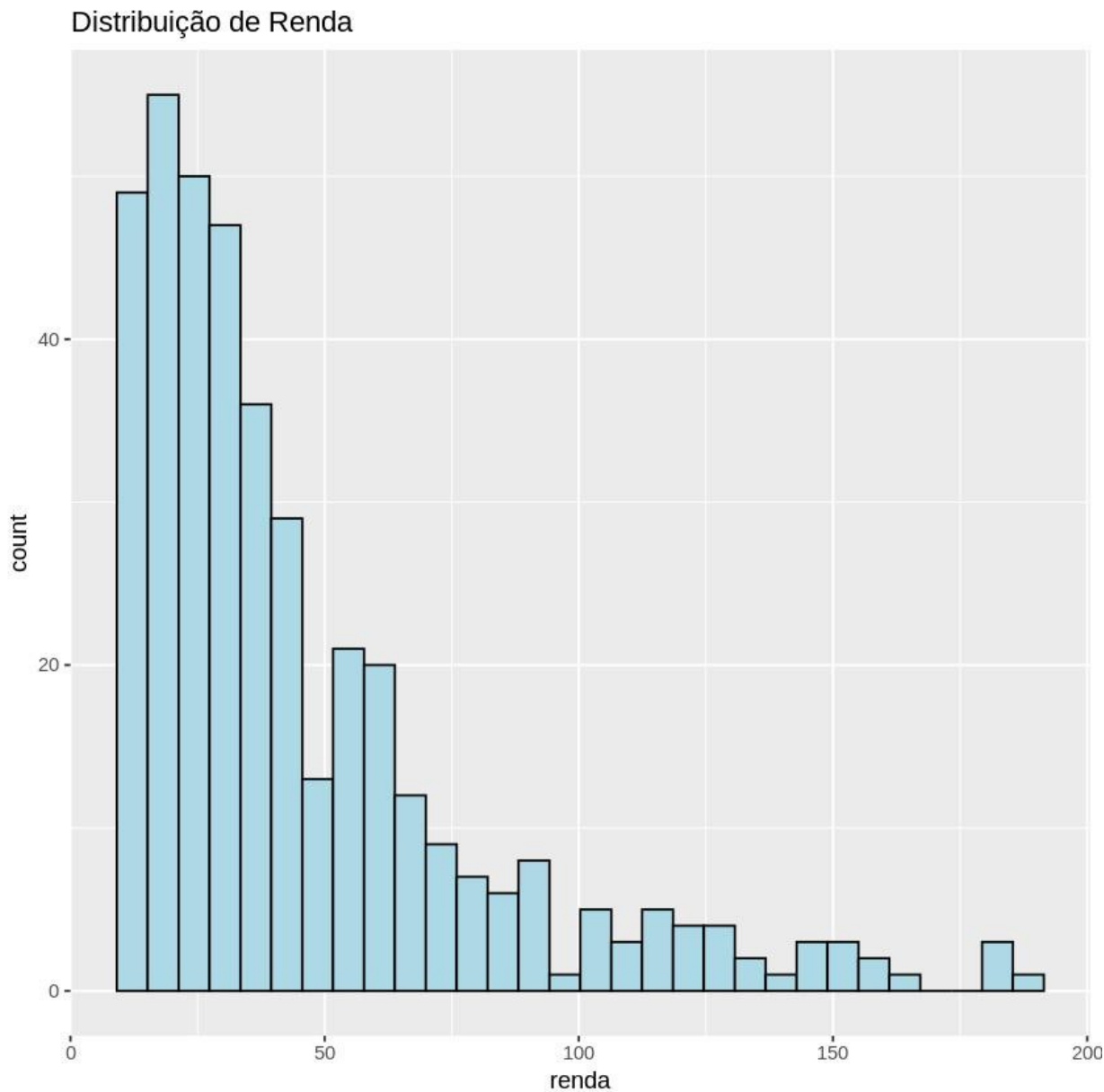
vif_results <- vif(modelo)
print(vif_results)

# Interpretação dos resultados de VIF
ifelse(any(vif_results > 10), "Atenção: Pode haver multicolinearidade!", "Não há problemas de multicolinearidade significativos")
```

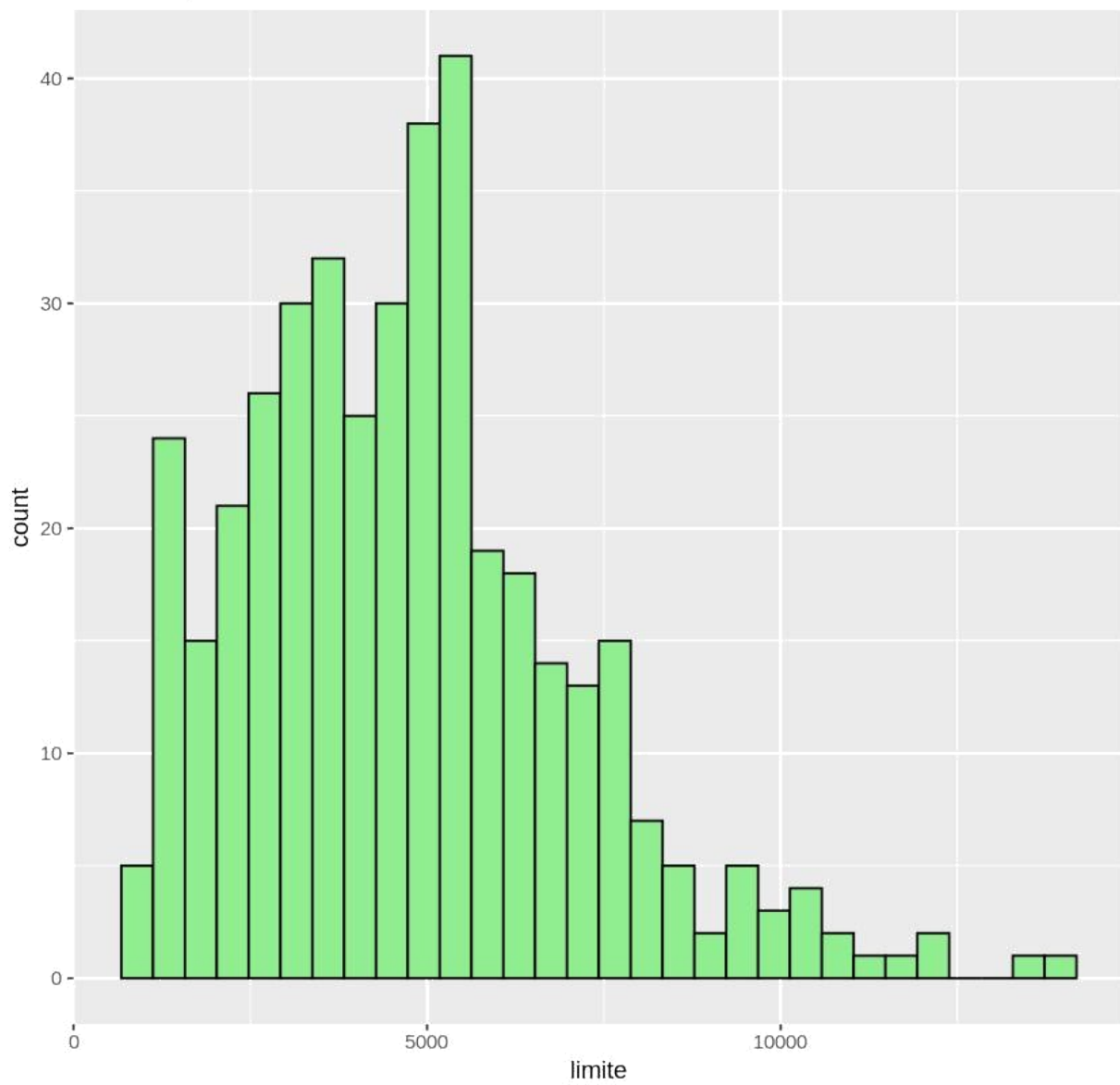
renda		limite		score		saldo	
Min.	: 10.35	Min.	: 855	Min.	: 93.0	Min.	: 0.00
1st Qu.:	21.01	1st Qu.:	3088	1st Qu.:	247.2	1st Qu.:	68.75
Median :	33.12	Median :	4622	Median :	344.0	Median :	459.50
Mean	: 45.22	Mean	: 4736	Mean	:354.9	Mean	: 520.01
3rd Qu.:	57.47	3rd Qu.:	5873	3rd Qu.:	437.2	3rd Qu.:	863.00
Max.	:186.63	Max.	:13913	Max.	:982.0	Max.	:1999.00

cartoes		idade		educacao		genero		estudante	
Min.	:1.000	Min.	:23.00	Min.	: 5.00		Male :	193	No :360
1st Qu.:	2.000	1st Qu.:	41.75	1st Qu.:	11.00		Female:	207	Yes: 40
Median :	3.000	Median :	56.00	Median :	14.00				
Mean	:2.958	Mean	:55.67	Mean	:13.45				
3rd Qu.:	4.000	3rd Qu.:	70.00	3rd Qu.:	16.00				
Max.	:9.000	Max.	:98.00	Max.	:20.00				

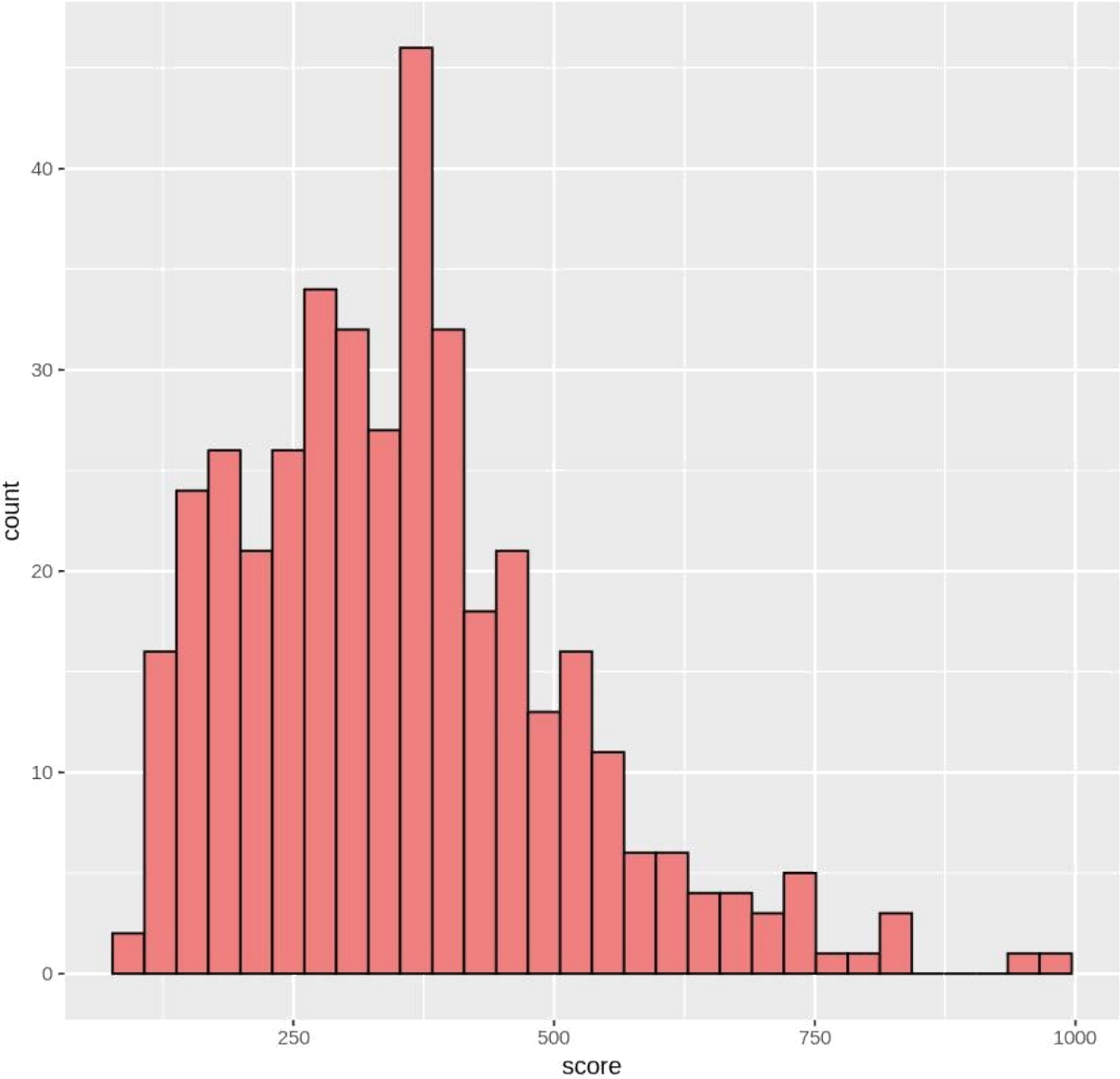
casado		etnia	
No :	155	African American:	99
Yes:	245	Asian	:102
		Caucasian	:199



Distribuição de Limite

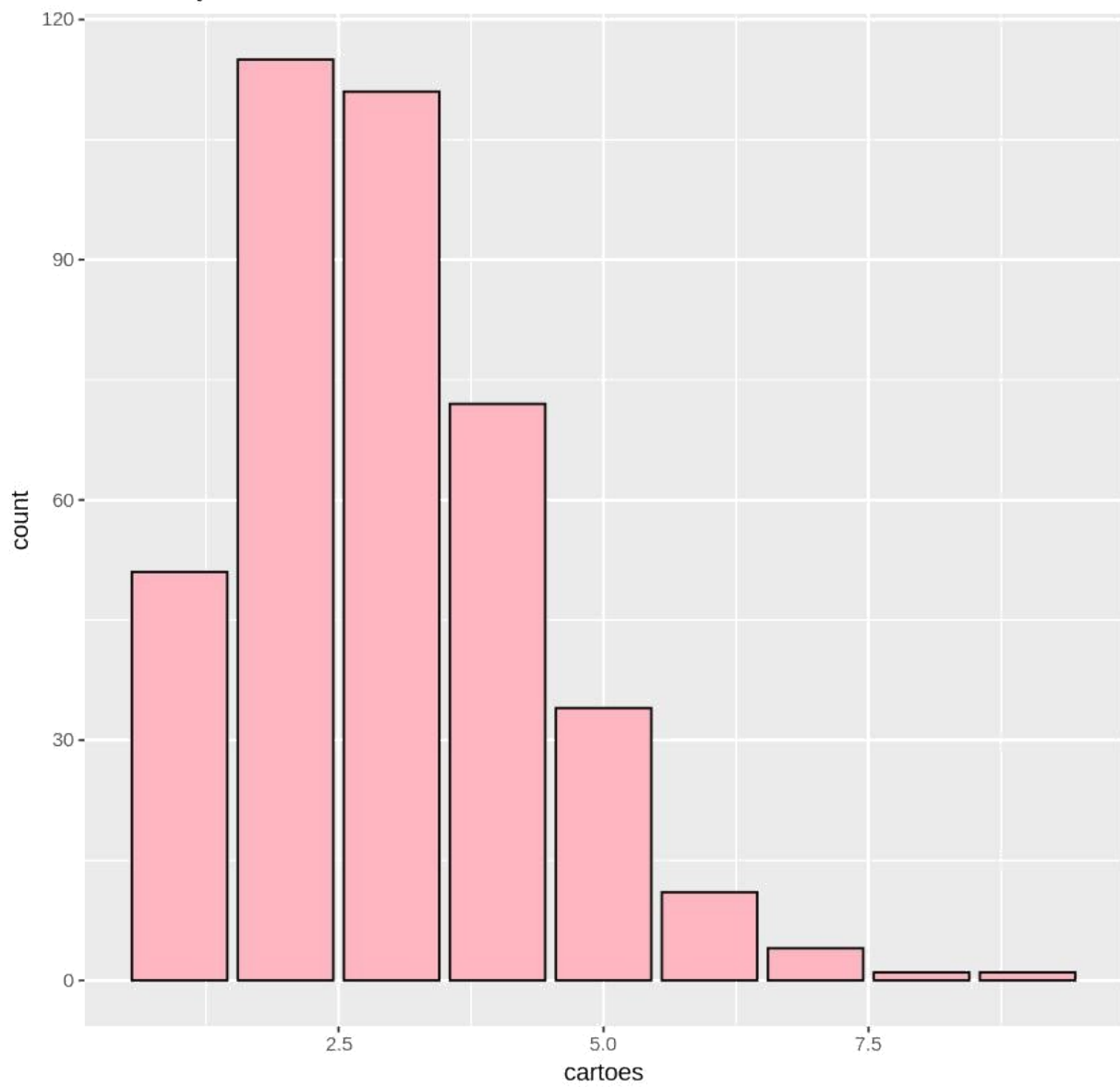


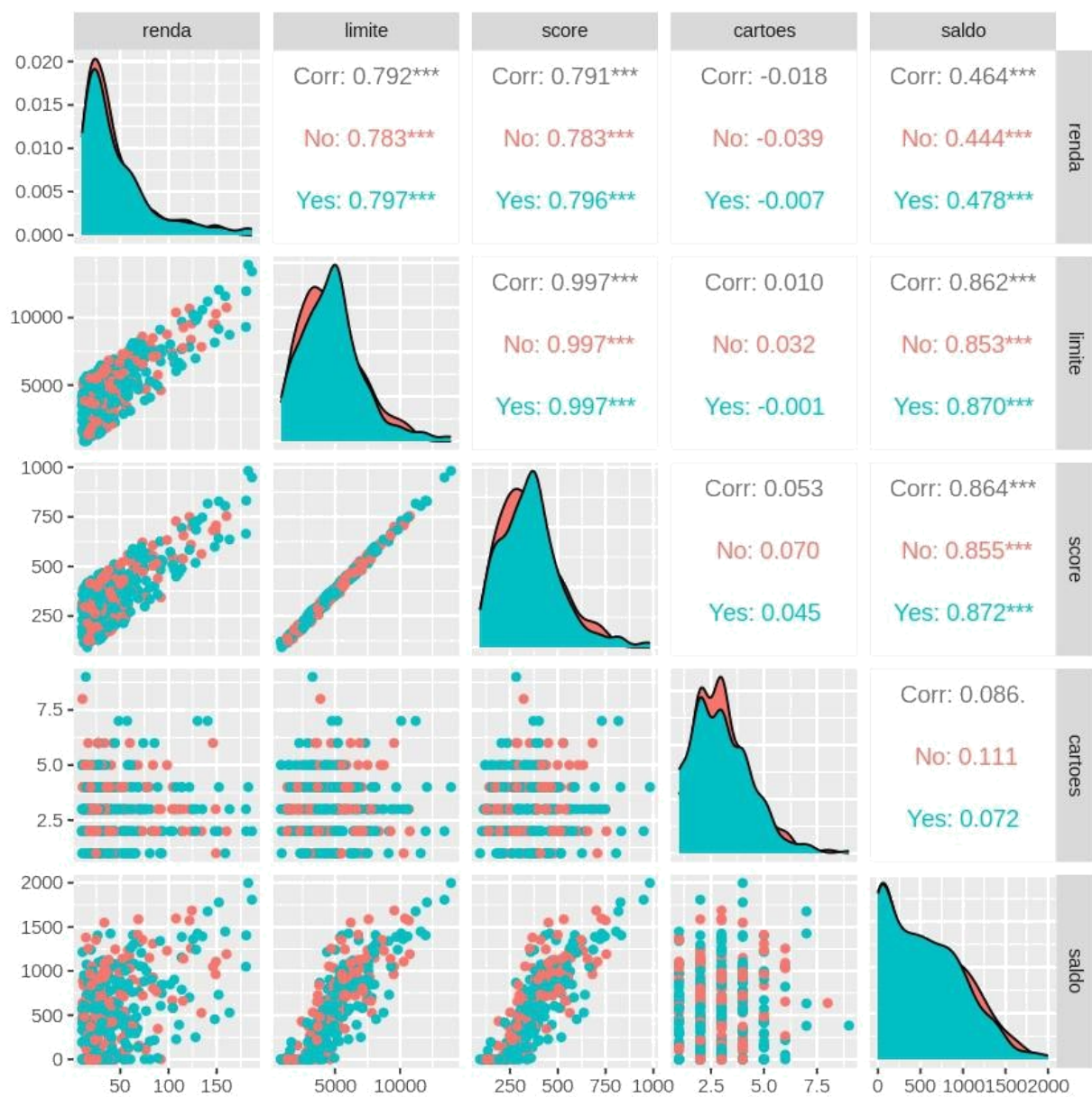
Distribuição de Score



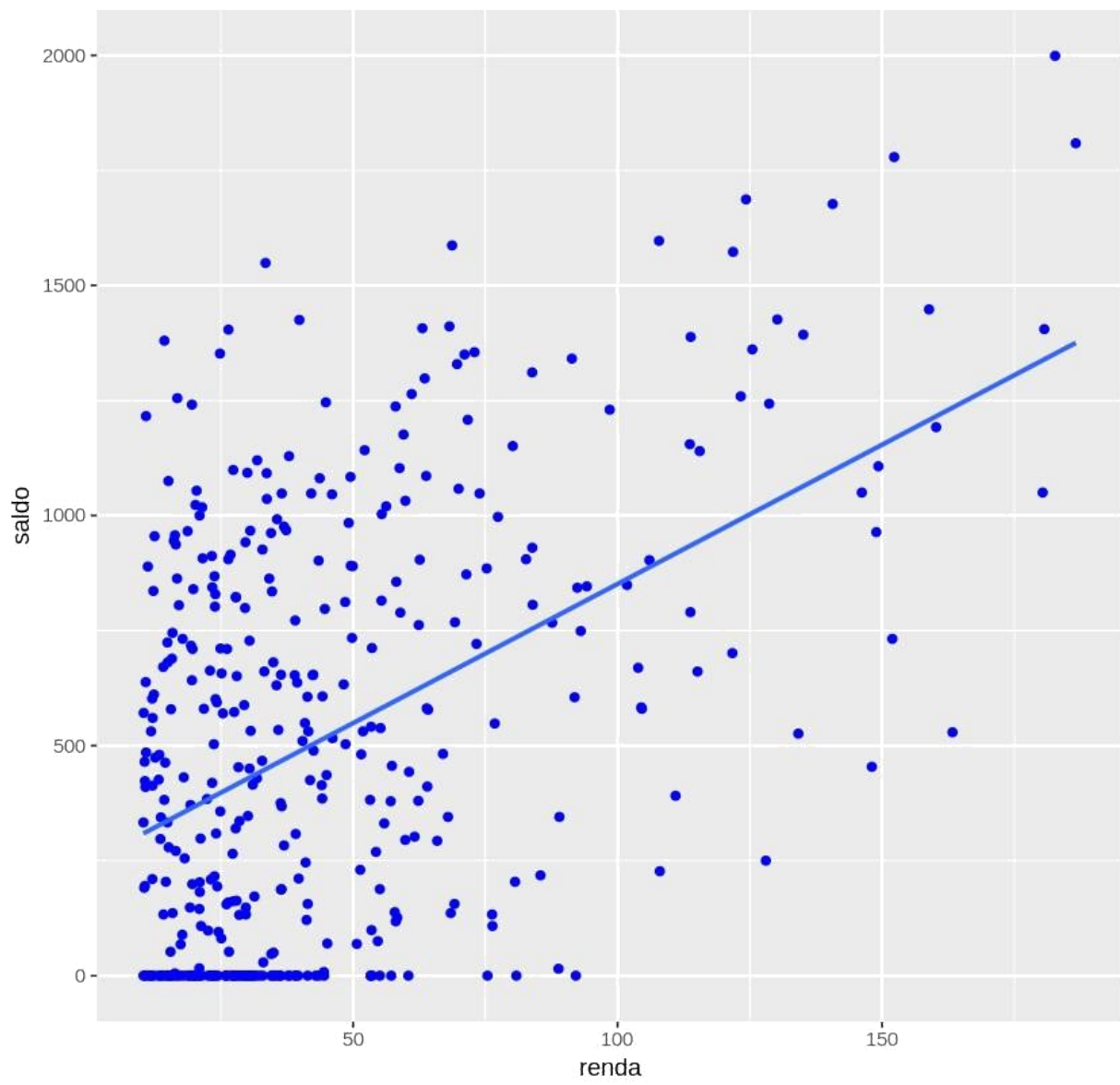
	renda	limite	score	cartoes	saldo
renda	1.00000000	0.79208834	0.79137763	-0.01827261	0.46365646
limite	0.79208834	1.00000000	0.99687974	0.01023133	0.86169727
score	0.79137763	0.99687974	1.00000000	0.05323903	0.86362516
cartoes	-0.01827261	0.01023133	0.05323903	1.00000000	0.08645635
saldo	0.46365646	0.86169727	0.86362516	0.08645635	1.00000000

Distribuição de Cartões

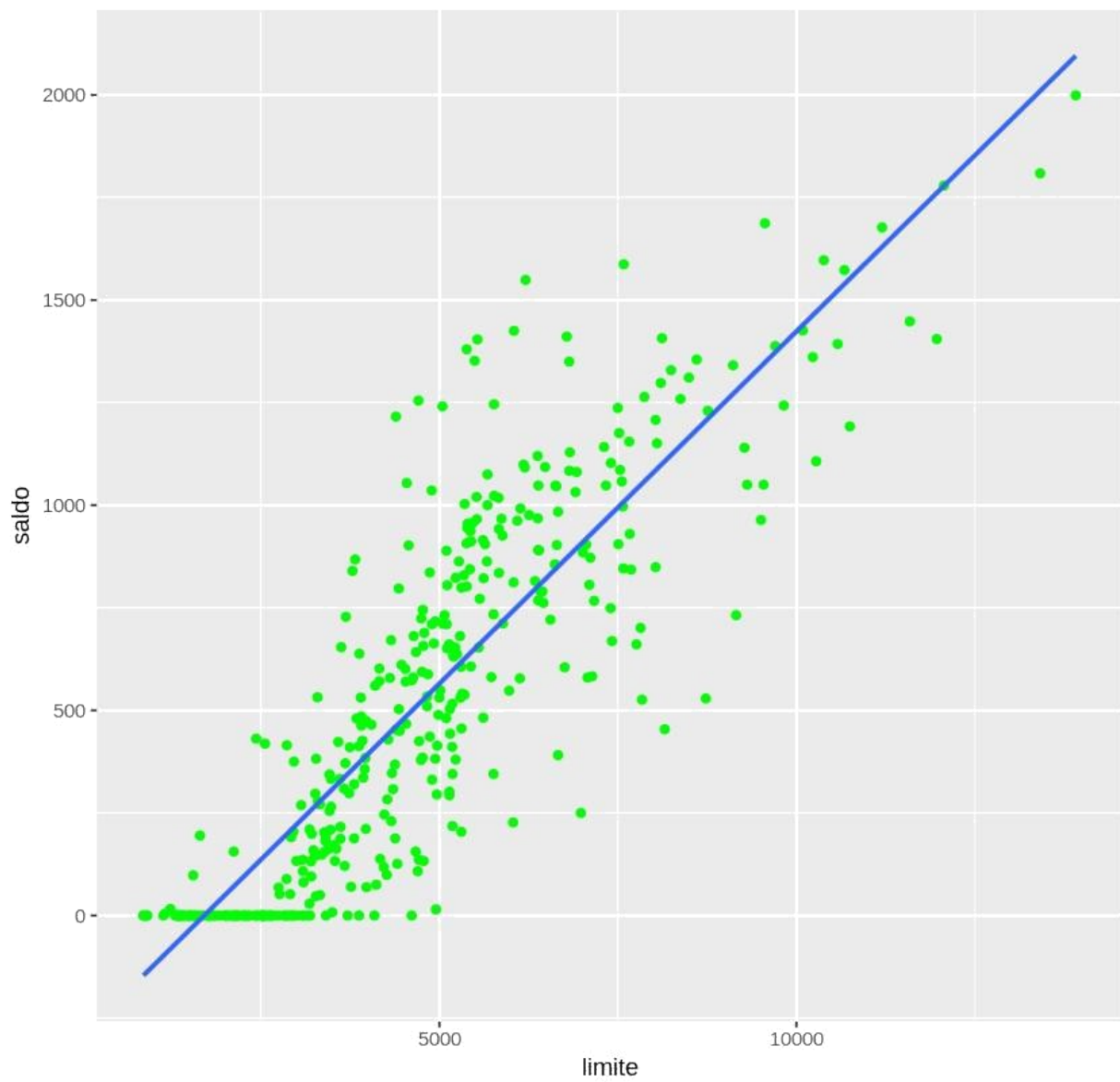




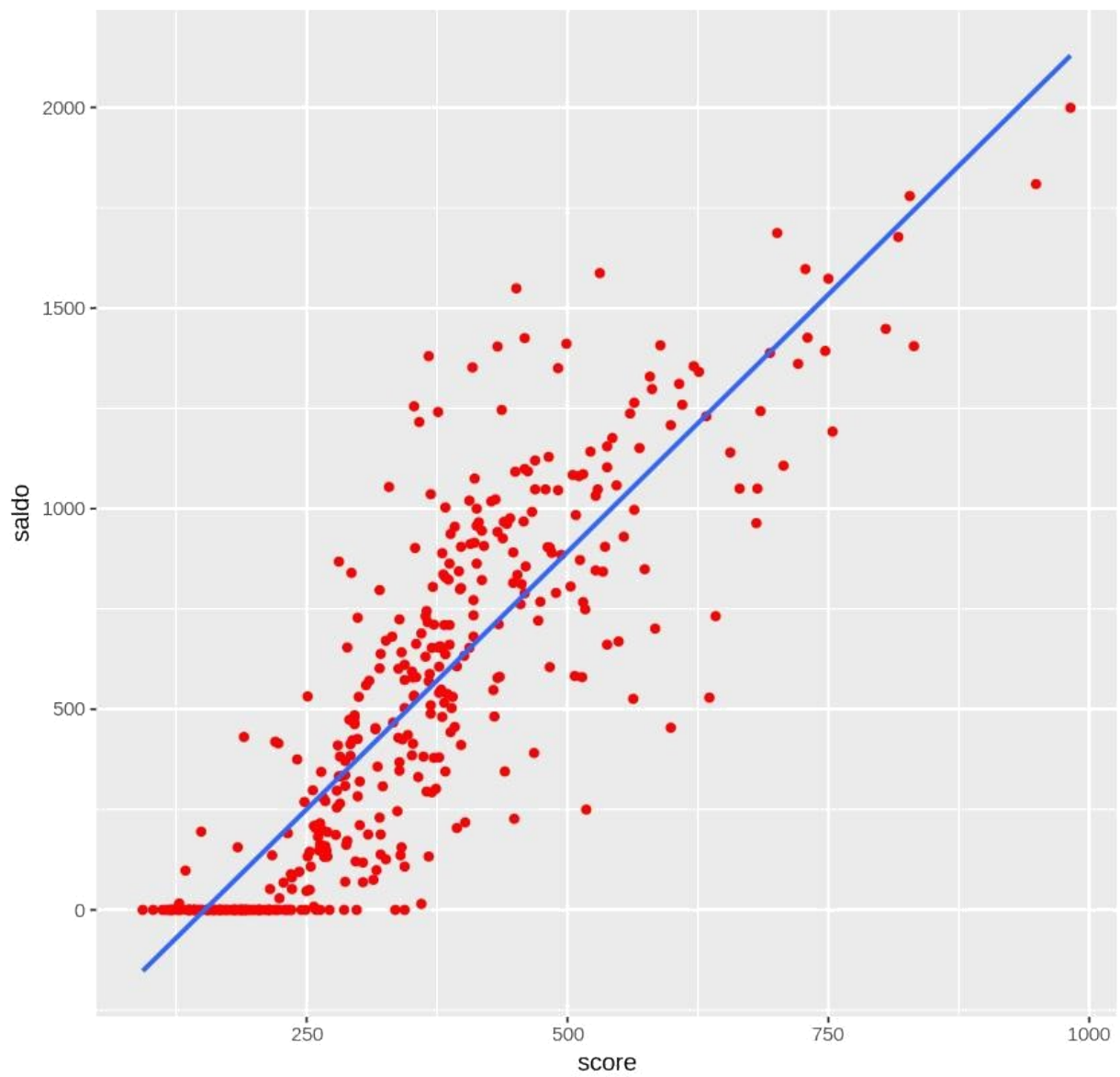
Dispersão entre Renda e Saldo



Dispersão entre Limite e Saldo



Dispersão entre Score e Saldo



Pearson's product-moment correlation

```
data:      dados$renda and dados$saldo
t = 10.44, df = 398, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.3830176 0.5372821
sample estimates:
      cor
0.4636565
```

Pearson's product-moment correlation

```
data:      dados$limite and dados$saldo
t = 33.879, df = 398, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.8341212 0.8849765
sample estimates:
      cor
0.8616973
```

Pearson's product-moment correlation

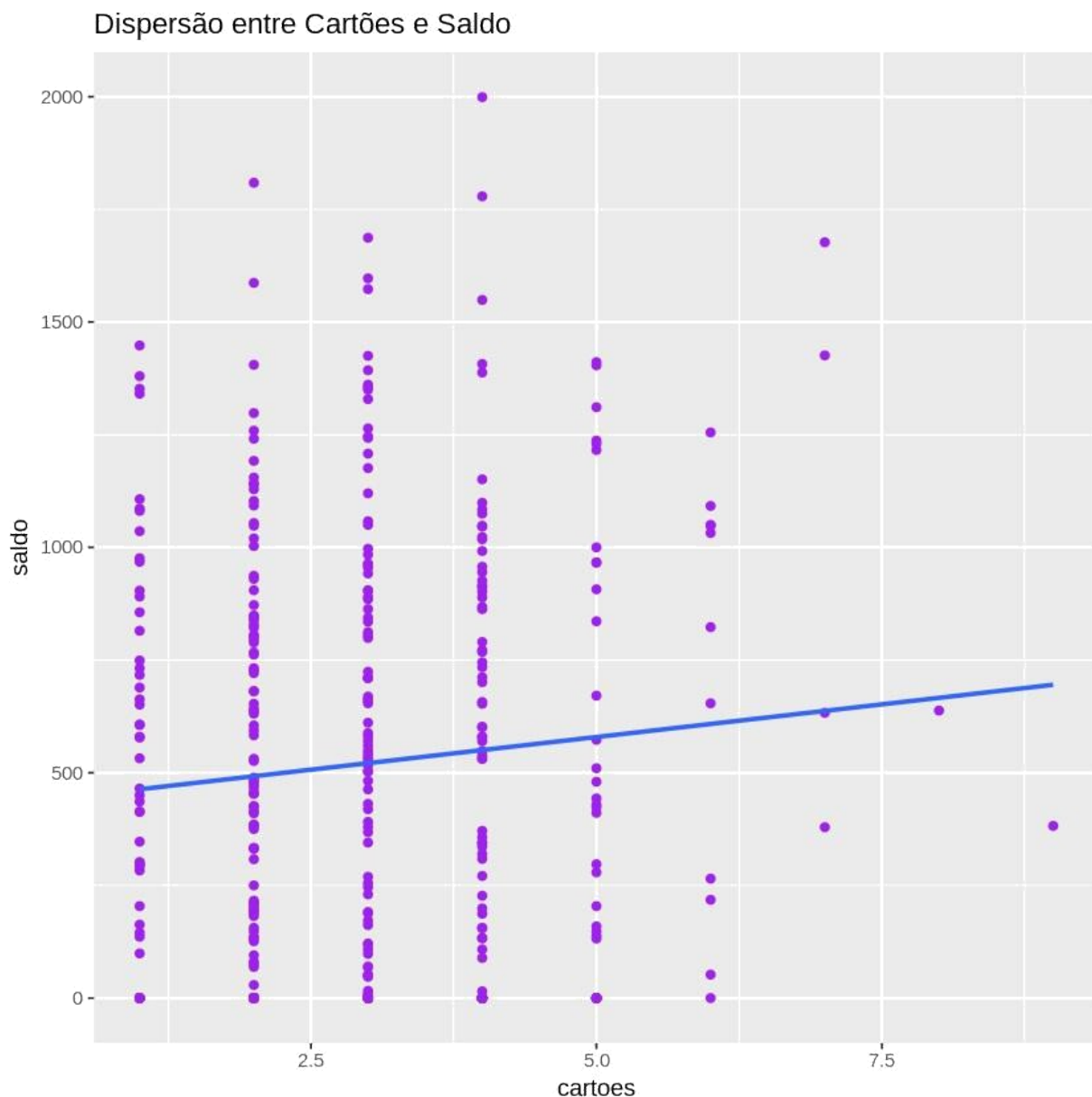
```
data:      dados$score and dados$saldo
t = 34.176, df = 398, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.8363997 0.8865997
sample estimates:
      cor
0.8636252
```

Pearson's product-moment correlation

```
data:      dados$cartoes and dados$saldo
t = 1.7313, df = 398, p-value = 0.08418
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
-0.01169451  0.18295710
sample estimates:
      cor
0.08645635
```

renda	limite	score	2.702816	cartoes
227.829464	229.587751			1.431425

'Atenção: Pode haver multicolinearidade!'



4.1. Análise Univariada

As distribuições das variáveis foram visualizadas através de histogramas e gráficos de barras. A análise sugere que:

- Renda: Apresenta uma distribuição assimétrica à direita, com a maioria dos valores concentrados em torno de faixas menores, indicando que a maioria dos clientes possui rendas menores.
- Limite: Também apresenta uma distribuição assimétrica, com muitos clientes tendo limites baixos e poucos com limites elevados.
- Score: Mostra uma distribuição que pode indicar uma concentração de clientes com scores relativamente altos, mas com alguns outliers.
- Cartões: A distribuição é representada como um gráfico de barras, indicando a contagem de cartões por cliente, sugerindo a presença de um número considerável de clientes que possuem vários cartões.

4.2. Análise Bivariada

A matriz de correlação revelou o seguinte:

- Renda e Saldo: A correlação é positiva, mas não muito alta, sugerindo que aumentos na renda estão associados a aumentos no saldo, mas outros fatores também influenciam essa relação.
- Limite e Saldo: Apresenta uma correlação forte e positiva, indicando que limites maiores tendem a estar associados a saldos maiores.
- Score e Saldo: A correlação é moderadamente alta, sugerindo que scores melhores estão associados a saldos mais elevados.
- Cartões e Saldo: A correlação é baixa, indicando que o número de cartões não tem uma relação significativa com o saldo.

Os gráficos de dispersão (exibindo a relação entre cartões e saldo) mostram uma tendência geral de aumento, mas com muitos outliers e dispersão, o que reforça a baixa correlação observada.

4.3. VIF e Multicolinearidade

Os resultados do VIF (Variance Inflation Factor) mostram que:

- Renda, Limite, e Score têm valores de VIF bem acima de 10, indicando potencial multicolinearidade entre essas variáveis. Isso pode ser um problema para a interpretação dos coeficientes do modelo de regressão, pois a presença de multicolinearidade dificulta a distinção do impacto de cada variável explicativa sobre a variável resposta.
- Cartões também apresenta um VIF elevado, sugerindo que pode haver uma relação complexa entre o número de cartões e outras variáveis.

A mensagem de alerta sobre a multicolinearidade indica que as conclusões tiradas a partir do modelo de regressão linear devem ser interpretadas com cautela. É recomendável considerar a exclusão de variáveis altamente correlacionadas ou a aplicação de técnicas de regularização para resolver esses problemas.

4.4. Teste de Correlação de Pearson

São realizados testes de correlação de Pearson entre cada variável explicativa e a variável resposta (saldo). Isso fornece insights sobre a força e a direção da relação linear entre as variáveis.

Dessa forma, o coeficiente de correlação de Pearson pode ser classificado das seguintes formas:

- Correlação menor que zero: Se a correlação for menor que zero, significa que é negativo, isto é, que as variáveis são inversamente relacionadas. Quando o valor de alguma variável é alto, o valor da outra variável é baixo.
- Correlação maior que zero: Quando a correlação é maior que zero, significa que a correlação é positiva, isto é, que as variáveis estão diretamente correlacionadas. Quando o valor de uma variável é alto, o valor da outra variável também é alto, o mesmo acontece quando eles são baixos.
- Correlação igual a zero: Quando a correlação é igual a zero, significa que não é possível determinar qualquer senso de covariação.
- Correlação igual a +1 ou -1: Quando a correlação é igual a +1 ou -1, significa uma correlação positiva perfeita ou uma correlação negativa perfeita, respectivamente.

Adicionalmente, ainda podemos classificar o coeficiente de correlação de três maneiras, de acordo com o intervalo no qual o coeficiente está contido:

- Correlação entre 0 e 0,3: Quando o módulo da correlação está entre 0 e 0,3, isso sugere uma relação linear fraca entre as variáveis. Embora exista uma associação, ela é bastante limitada.
- Correlação entre 0,3 e 0,7: Quando o módulo da correlação está entre 0,3 e 0,7, indica uma relação linear moderada entre as variáveis. Nesse caso, as variáveis tendem a se mover juntas em certa medida, mas não de forma extremamente forte. Correlação
- entre 0,7 e 1: Quando o módulo da correlação é maior ou igual a 0,7, indica uma forte relação linear entre as variáveis. Isso significa que as variáveis estão altamente correlacionadas e tendem a se mover juntas de forma consistente. Essa forte associação pode ser útil para prever ou explicar o comportamento das variáveis dependentes com base nos valores das variáveis independentes.

4.5. Conclusão

A análise univariada e bivariada forneceu insights valiosos sobre as relações entre variáveis explicativas e a variável resposta Balance (saldo). A presença de multicolinearidade sugere que uma investigação adicional ou ajustes ao modelo são necessários para obter interpretações precisas.

5. Ajustando o modelo inicial cosiderando todas as variáveis explicativas

```
install.packages("lmtest", repos = 'http://cran.us.r-project.org')
# Ajustar o modelo inicial
modelo_inicial <- lm(saldo ~ renda + limite + score + cartoes + idade + educacao + genero + casado + etnia, data

# Resumo do modelo
summary(modelo_inicial)

# Análise dos resíduos
par(mfrow=c(2, 2)) # Configuração do layout do gráfico
plot(modelo_inicial, which = 1, col = "lightgreen") # Histograma dos resíduos plot(modelo_inicial, which = 2, col =
"lightgreen") # Gráfico QQ plot dos resíduos plot(modelo_inicial, which = 3, col = "lightgreen") # Resíduos vs. Valores
Ajustados plot(modelo_inicial, which = 5, col = "lightgreen") # Resíduos vs. Variáveis Explicativas

# Gráfico de Dispersão dos Resíduos residuos <-
resid(modelo_inicial) valores_ajustados <-
fitted(modelo_inicial)
plot(valores_ajustados, residuos, main = "Gráfico de Dispersão dos Resíduos",
```

```

xlab = "Valores Ajustados", ylab = "Resíduos", col = "lightgreen")
abline(h = 0, col = "red", lty = 2) # Adiciona linha horizontal em y=0

# Teste de normalidade dos resíduos
shapiro_test <- shapiro.test(residuals(modelo_inicial))
cat("Teste de Shapiro-Wilk para normalidade dos resíduos:\n")
print(shapiro_test)

# Teste de homocedasticidade
library(lmtest)
bptest_result <- bptest(modelo_inicial)
cat("Teste de Breusch-Pagan para homocedasticidade:\n")
print(bptest_result)

```

Call:

```
lm(formula = saldo ~ renda + limite + score + cartoes + idade + educacao + genero +
    casado + etnia, data = dados)
```

Residuals:

Min	1Q	Median	3Q	Max
-214.80	-110.36	-46.20	51.76	530.73

Coefficients: Estimate Std. Error t value Pr(>|t|)

(Intercept)	-478.49892	58.38617	-8.195	3.65e-15	***
renda	-7.55003	0.38194	-19.767	< 2e-16	***
limite	0.11114	0.05325	2.087	0.03754	*
score	2.28874	0.79777	2.869	0.00434	**
cartoes	10.40326	7.06938	1.472	0.14194	
idade	-0.92453	0.47940	-1.929	0.05452	.
educacao	2.22610	2.59927	0.856	0.39229	
generoFemale	3.52732	16.15496	0.218	0.82728	
casadoYes	-34.43100	16.83144	-2.046	0.04146	*
etniaAsian	32.11041	23.02264	1.395	0.16389	
etniaCaucasian	6.21498	19.92610	0.312	0.75528	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 161.2 on 389 degrees of freedom

Multiple R-squared: 0.8801, Adjusted R-squared: 0.877

F-statistic: 285.5 on 10 and 389 DF, p-value: < 2.2e-16

Teste de Shapiro-Wilk para normalidade dos resíduos:

Shapiro-Wilk normality test

data: residuals(modelo_inicial)

W = 0.87287, p-value < 2.2e-16

Teste de Breusch-Pagan para homocedasticidade:

studentized Breusch-Pagan test

data: modelo_inicial

BP = 12.636, df = 10, p-value = 0.2447

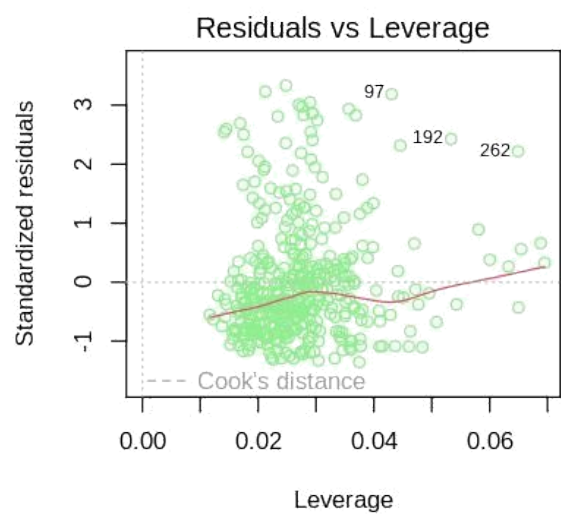
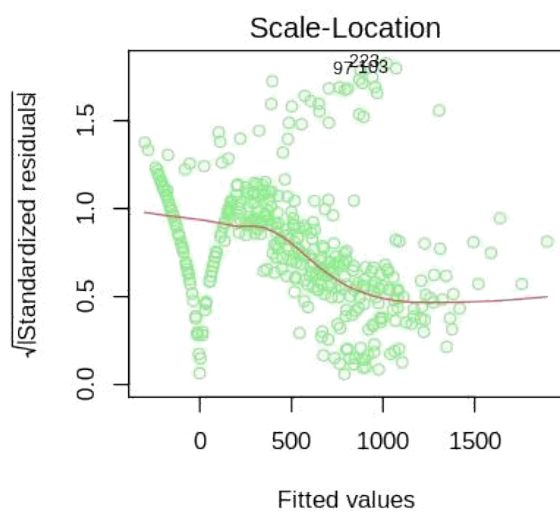
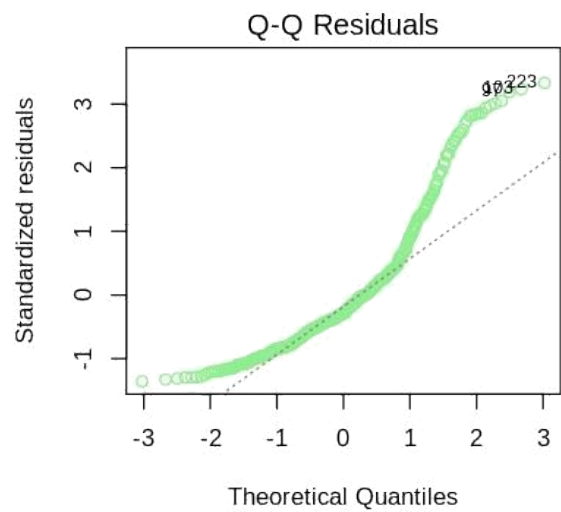
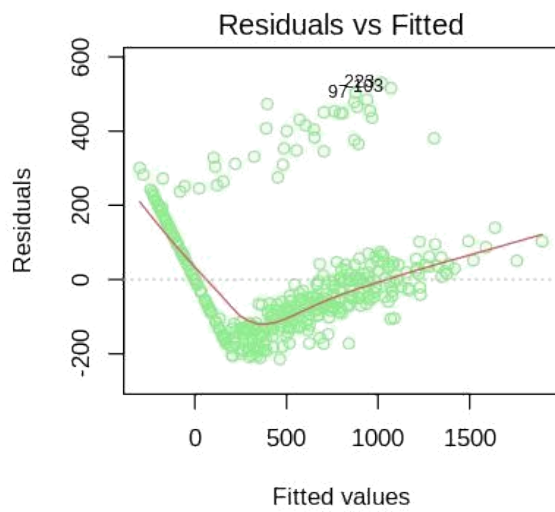
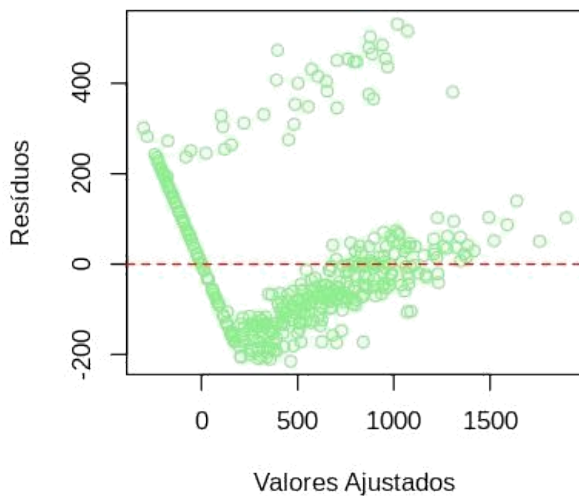


Gráfico de Dispersão dos Resíduos



Gráficos de Diagnóstico

Os gráficos de diagnóstico fornecem informações importantes sobre a qualidade do ajuste do modelo de regressão linear. Vamos analisar cada um deles:

1. Residuals vs Fitted:

Padrão: Os pontos não se distribuem aleatoriamente em torno da linha horizontal zero, indicando a presença de heterocedasticidade (variância dos resíduos não constante). Interpretação: A variância dos resíduos aumenta conforme os valores ajustados aumentam, sugerindo que o modelo pode não estar capturando toda a variabilidade dos dados nos valores mais altos.

2. Q-Q Plot: Padrão: Os pontos se desviam significativamente da linha diagonal, especialmente nas caudas. Interpretação: Os resíduos não seguem uma distribuição normal, o que viola uma das premissas da regressão linear.

3. Scale-Location:

Padrão: Similar ao gráfico "Residuals vs Fitted", mostra heterocedasticidade.

4. Residuals vs Leverage:

Padrão: Há alguns pontos com alta alavancagem, mas nenhum parece influenciar significativamente os resultados.

5. Gráfico de dispersão dos resíduos indica problemas como heterocedasticidade, falta de linearidade e presença de outliers. Demonstrando que o modelo estatístico avaliado não é adequado.

6. Escolhendo o melhor modelo

```
# Ajustando vários modelos
modelo_1 <- lm(saldo ~ renda + limite, data = dados)
modelo_2 <- lm(saldo ~ renda + limite + score, data = dados)
modelo_3 <- lm(saldo ~ renda + limite + score + cartoes + idade, data = dados)

# Resumos dos modelos
summary(modelo_1)
summary(modelo_2)
summary(modelo_3)

# Comparando AIC e BIC
aic_values <- c(AIC(modelo_1), AIC(modelo_2), AIC(modelo_3))
bic_values <- c(BIC(modelo_1), BIC(modelo_2), BIC(modelo_3))

cat("AIC dos modelos:\n", aic_values, "\n")
cat("BIC dos modelos:\n", bic_values, "\n")
```


Call:

lm(formula = saldo ~ renda + limite, data = dados)

Residuals:

Min	1Q	Median	3Q	Max
-232.79	-115.45	-48.20	53.36	549.77

Coefficients: Estimate Std. Error t value Pr(>|t|)

(Intercept)	-385.17926		19.46480	-19.79	<2e-16	***
renda		-7.66332	0.38507	-19.90	<2e-16	***
limite		0.26432	0.00588	44.95	<2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 165.5 on 397 degrees of freedom

Multiple R-squared: 0.8711, Adjusted R-squared: 0.8705

F-statistic: 1342 on 2 and 397 DF, p-value: < 2.2e-16

Call:

lm(formula = saldo ~ renda + limite + score, data = dados)

Residuals:

Min	1Q	Median	3Q	Max
-260.93	-113.14	-36.27	49.35	554.23

Coefficients: Estimate Std. Error t value Pr(>|t|)

(Intercept)	-489.72748		32.09892	-15.257	< 2e-16	***
renda		-7.71931	0.37806	-20.418	< 2e-16	***
limite		0.08467	0.04471	1.894	0.059 .	
score		2.69858	0.66594	4.052	6.11e-05	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 162.4 on 396 degrees of freedom

Multiple R-squared: 0.8762, Adjusted R-squared: 0.8753

F-statistic: 934.6 on 3 and 396 DF, p-value: < 2.2e-16

Call:

lm(formula = saldo ~ renda + limite + score + cartoes + idade, data = dados)

Residuals:

Min	1Q	Median	3Q	Max
-231.37	-113.46	-39.55	41.66	544.35

Coefficients: Estimate Std. Error t value Pr(>|t|)

(Intercept)	-449.36101		40.57409	-11.075	<2e-16	***
renda		-7.56211	0.38214	-19.789	<2e-16	***
limite		0.12855	0.05289	2.430	0.0155	*
score		2.02240	0.79208	2.553	0.0110	*
cartoes		11.55272	7.06285	1.636	0.1027	
idade		-0.88832	0.47781	-1.859	0.0638	.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 161.6 on 394 degrees of freedom

Multiple R-squared: 0.8781, Adjusted R-squared: 0.8765

F-statistic: 567.4 on 5 and 394 DF, p-value: < 2.2e-16

AIC dos modelos:

5227.203 5212.95 5211.013

BIC dos modelos:

5243.168 5232.908 5238.954

AIC e BIC dos Modelos

Modelo 1 (saldo ~ renda + limite):

AIC: 5227.203 BIC: 5243.168

Modelo 2 (saldo ~ renda + limite + score):

AIC: 5212.95 BIC: 5232.908

Modelo 3 (saldo ~ renda + limite + score + cartoes + idade):

AIC: 5211.013 BIC: 5238.954

6.1. Análise dos Resultados

- AIC (Akaike Information Criterion):

O Modelo 3 apresenta o menor valor de AIC (5211.013), indicando que é o modelo que melhor se ajusta aos dados entre os três modelos analisados, penalizando a complexidade. O Modelo 2 também tem um AIC relativamente baixo (5212.95), enquanto o Modelo 1 tem um AIC significativamente mais alto (5227.203), sugerindo que ele é o menos adequado.

- BIC (Bayesian Information Criterion):

O Modelo 3 novamente apresenta o menor valor de BIC (5238.954). Embora o Modelo 2 tenha um BIC ligeiramente superior (5232.908), ele ainda é melhor que o Modelo 1 (5243.168). O BIC é mais rigoroso em penalizar a inclusão de variáveis, e como o Modelo 3 ainda é o melhor aqui, isso sugere que ele não só se ajusta bem aos dados, mas também mantém uma complexidade razoável.

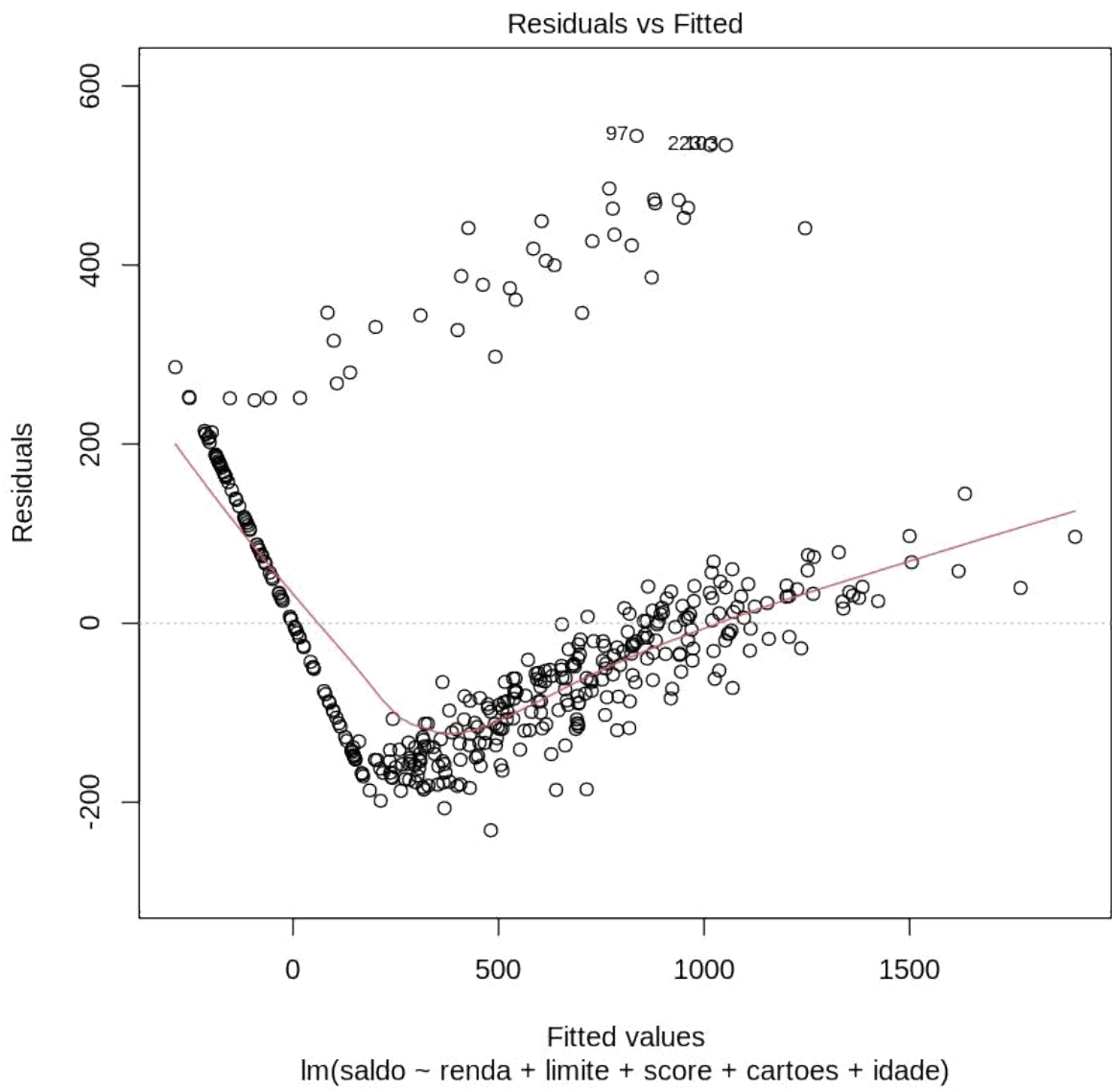
Interpretação:

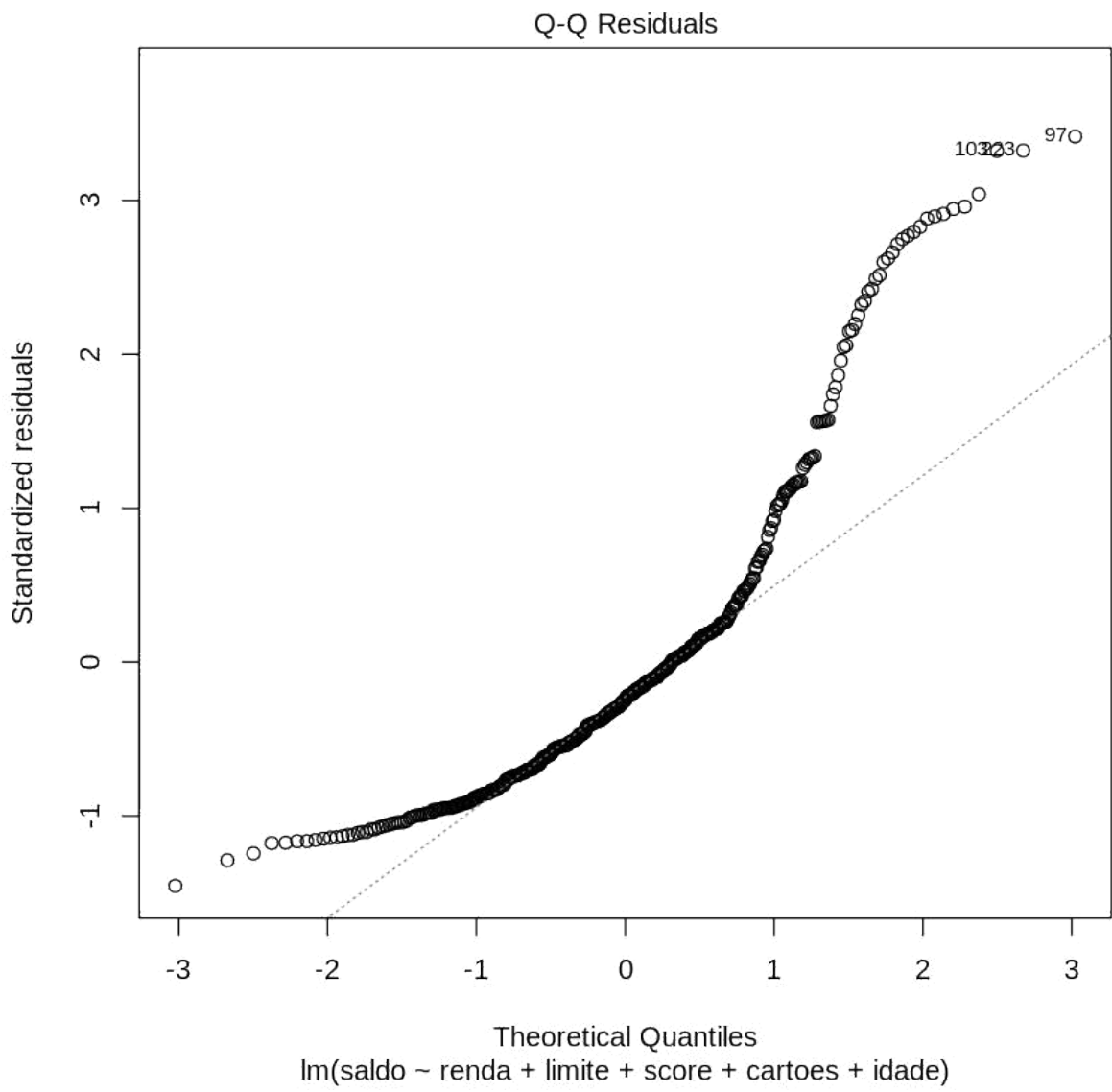
Como vimos, o modelo 3 mostrou-se mais adequado estatisticamente, pois oferece o melhor equilíbrio entre ajuste e complexidade.

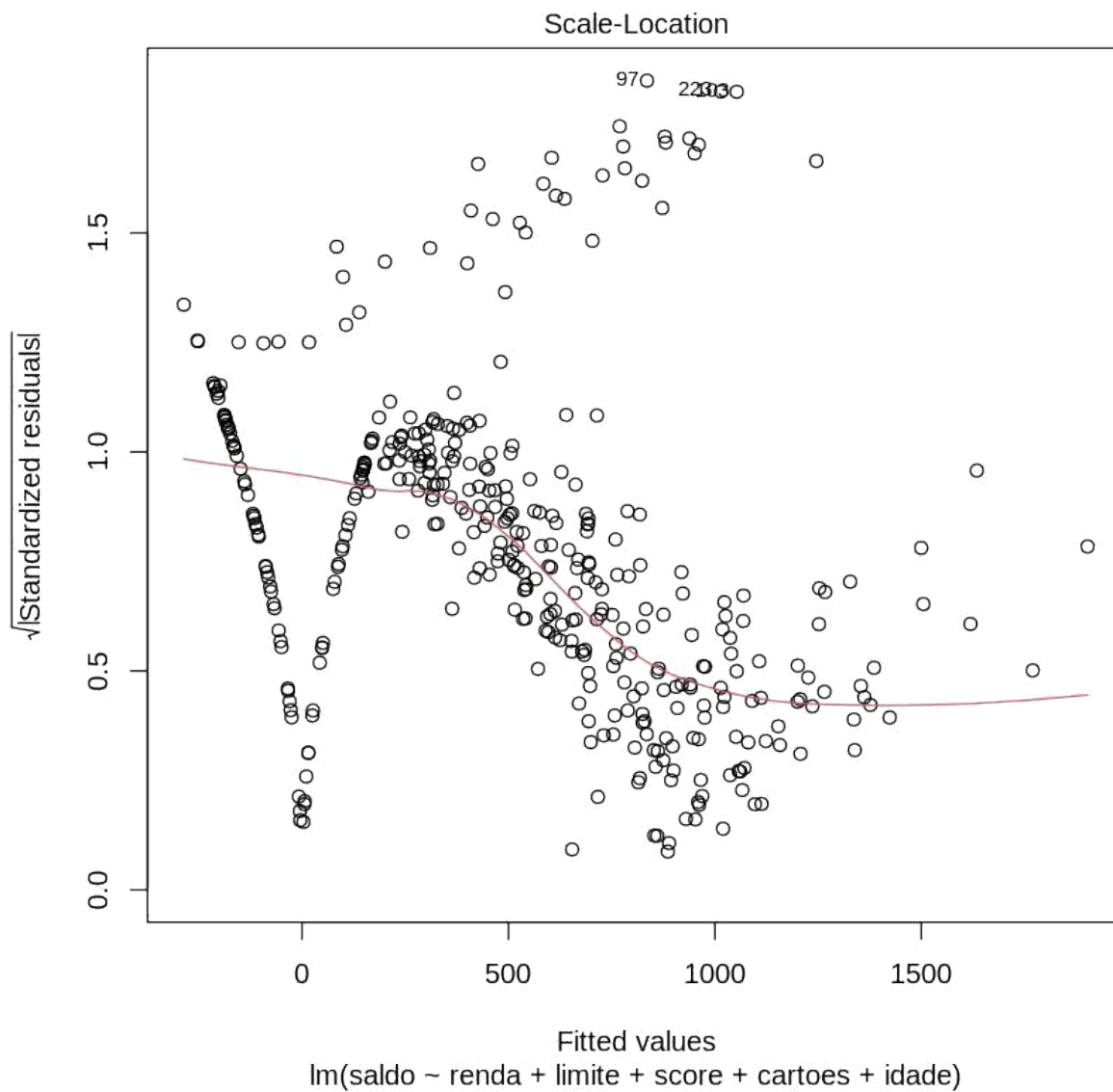
7. Pressupostos do MRLM

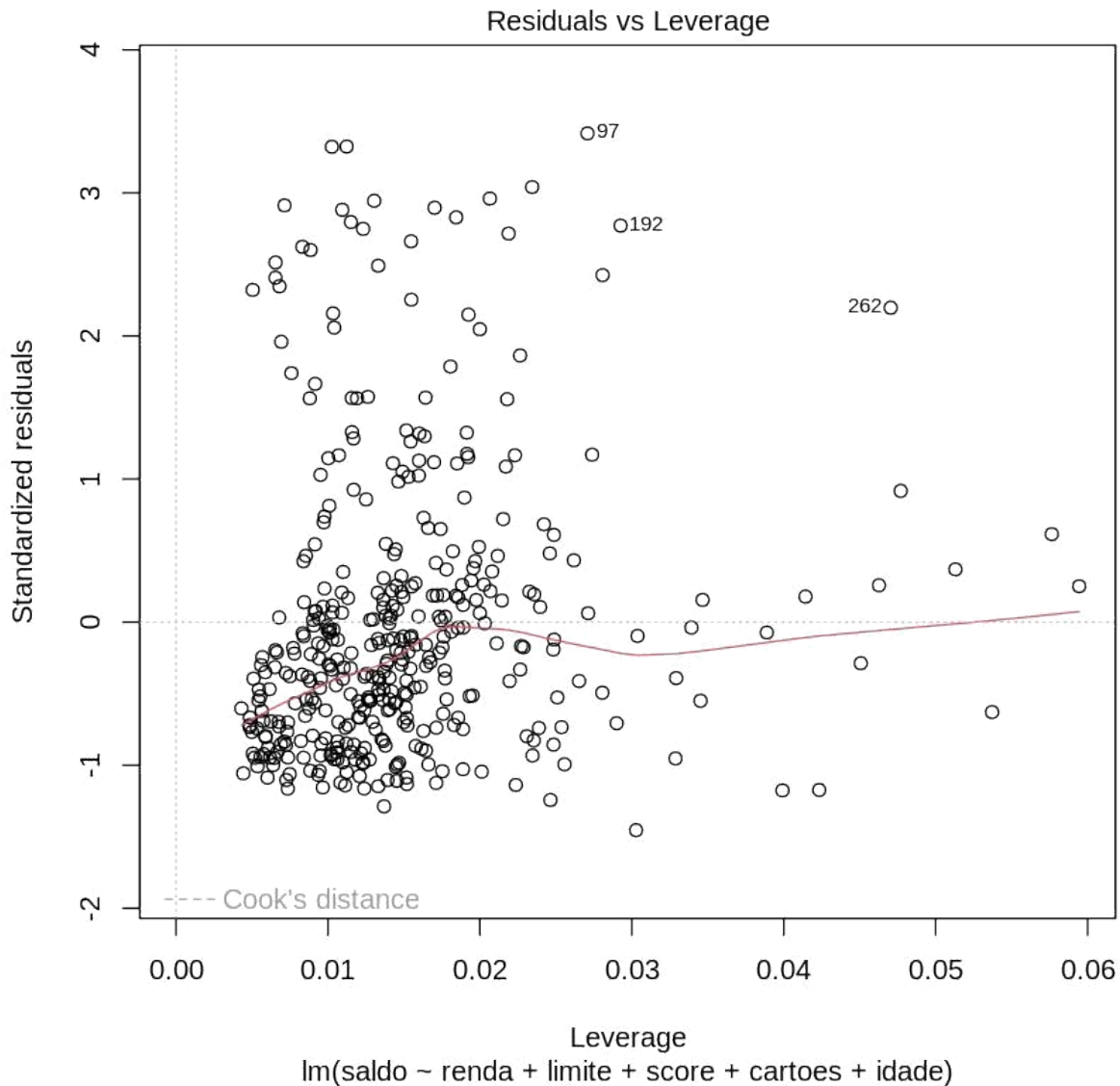
```
install.packages("report")
plot(modelo_3)

library(report)
report(modelo_3)
```









Tradução da função report(modelo_3):

Ajustamos um modelo linear (estimado usando OLS) para prever equilíbrio com renda, limite, score, cartoes e idade (fórmula: saldo ~ renda + limite + score + cartoes + idade). O modelo explica uma proporção estatisticamente significativa e substancial da variância ($R^2 = 0,88$, $F(5, 394) = 567,43$, $p < 0,001$, $R^2 \text{ adj.} = 0,88$). O intercepto do modelo, correspondente a renda = 0, limite = 0, score = 0, cartoes = 0 e idade = 0, está em -449,36 (IC 95% [-529,13, -369,59], $t(394) = -11,08$, $p < 0,001$).

Dentro deste modelo:

- O efeito de renda é estatisticamente significativo e negativo (beta = -7,56, IC 95% [-8,31, -6,81], $t(394) = -19,79$, $p < 0,001$; beta padrão = -0,58, IC 95% [-0,64, -0,52])
- O efeito de limite é estatisticamente significativo e positivo (beta = 0,13, IC 95% [0,02, 0,23], $t(394) = 2,43$, $p = 0,016$; beta padrão = 0,65, IC 95% [0,12, 1,17])
- O efeito de pontuação (score) é estatisticamente significativo e positivo (beta = 2,02, IC 95% [0,47, 3,58], $t(394) = 2,55$, $p = 0,011$; beta padrão = 0,68, IC 95% [0,16, 1,20])
- O efeito dos cartoes é estatisticamente não significativo e positivo (beta = 11,55, IC 95% [-2,33, 25,44], $t(394) = 1,64$, $p = 0,103$; beta padrão = 0,03, IC 95% [-6,96e-03, 0,08])
- O efeito da idade é estatisticamente não significativo e negativo (beta = -0,89, IC 95% [-1,83, 0,05], $t(394) = -1,86$, $p = 0,064$; beta padrão = -0,03, IC 95% [-0,07, 1,92e-03])

Os parâmetros padronizados foram obtidos ajustando o modelo em uma versão padronizada do conjunto de dados. Intervalos de confiança (ICs) de 95% e valores de p foram calculados usando uma aproximação de distribuição t de Wald.

8. Previsões

Para realizar previsões sobre valores para a variável resposta, iniciamos com o uso de valores para as variáveis explicativas dentro dos respectivos intervalos observados. Daí a importância de um breve resumo sobre os dados observados:

```
summary(modelo_3)
```

Call:

```
lm(formula = saldo ~ renda + limite + score + cartoes + idade, data = dados)
```

Residuals:

Min	1Q	Median	3Q	Max
-231.37	-113.46	-39.55	41.66	544.35

Coefficients: Estimate Std. Error t value Pr(>|t|)

(Intercept)	-449.36101		40.57409	-11.075	<2e-16	***
renda		-7.56211	0.38214	-19.789	<2e-16	***
limite		0.12855	0.05289	2.430	0.0155	*
score		2.02240	0.79208	2.553	0.0110	*
cartoes		11.55272	7.06285	1.636	0.1027	
idade		-0.88832	0.47781	-1.859	0.0638	.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 161.6 on 394 degrees of freedom

Multiple R-squared: 0.8781, Adjusted R-squared: 0.8765

F-statistic: 567.4 on 5 and 394 DF, p-value: < 2.2e-16

Vamos calcular a estimativa pontual do saldo considerando a média das variáveis explicativas:

```
# Criando um novo data frame com as médias das variáveis explicativas novas.preditoras <-  
data.frame(  
  renda = c(14.9, 104.6), limite =  
  c(3606, 7075), score = c(283,  
  514), cartoes = c(1, 4), idade =  
  c(18, 82)  
)  
  
# Previsão pontual  
predict <- predict(modelo_3, novas.preditoras, interval = "confidence")  
print(predict)
```

	fit	lwr	upr
1	469.4303	420.0009	518.8597
2	682.0419	642.2053	721.8785

- Estimativas Pontuais (fit):

- A estimativa pontual do saldo para a primeira observação é 469.43. Isso significa que, com base nas variáveis explicativas utilizadas, espera-se que o saldo médio seja de aproximadamente 469.43.
- Para a segunda observação, a estimativa pontual é 682.04. Este valor reflete a previsão média do saldo dado o conjunto de variáveis explicativas especificadas para essa observação.

- Intervalos de Confiança (lwr e upr):

- O intervalo de confiança para a primeira previsão varia entre 420.00 e 518.86. Isso indica que estamos 95% confiantes de que a média do saldo real (considerando o mesmo conjunto de variáveis) estará entre esses dois valores. A amplitude relativamente estreita sugere uma boa precisão na estimativa. *Para a segunda previsão, o intervalo de confiança se estende de 642.21 a 721.88. Novamente, isso indica um alto nível de confiança de que a média do saldo real estará dentro desse intervalo.

Além disso, tem-se que todas as variáveis explicam de forma estatisticamente significativa e com um bom ajuste indicado pelo Coeficiente de determinação ajustado (Adjusted R-squared: 0.8765)