

Felipe dos Santos Barbosa | 04242029

Grigor Ferreira Mascarenhas | 04242067

## Atividade Análise de Empréstimo - Cálculo Computacional

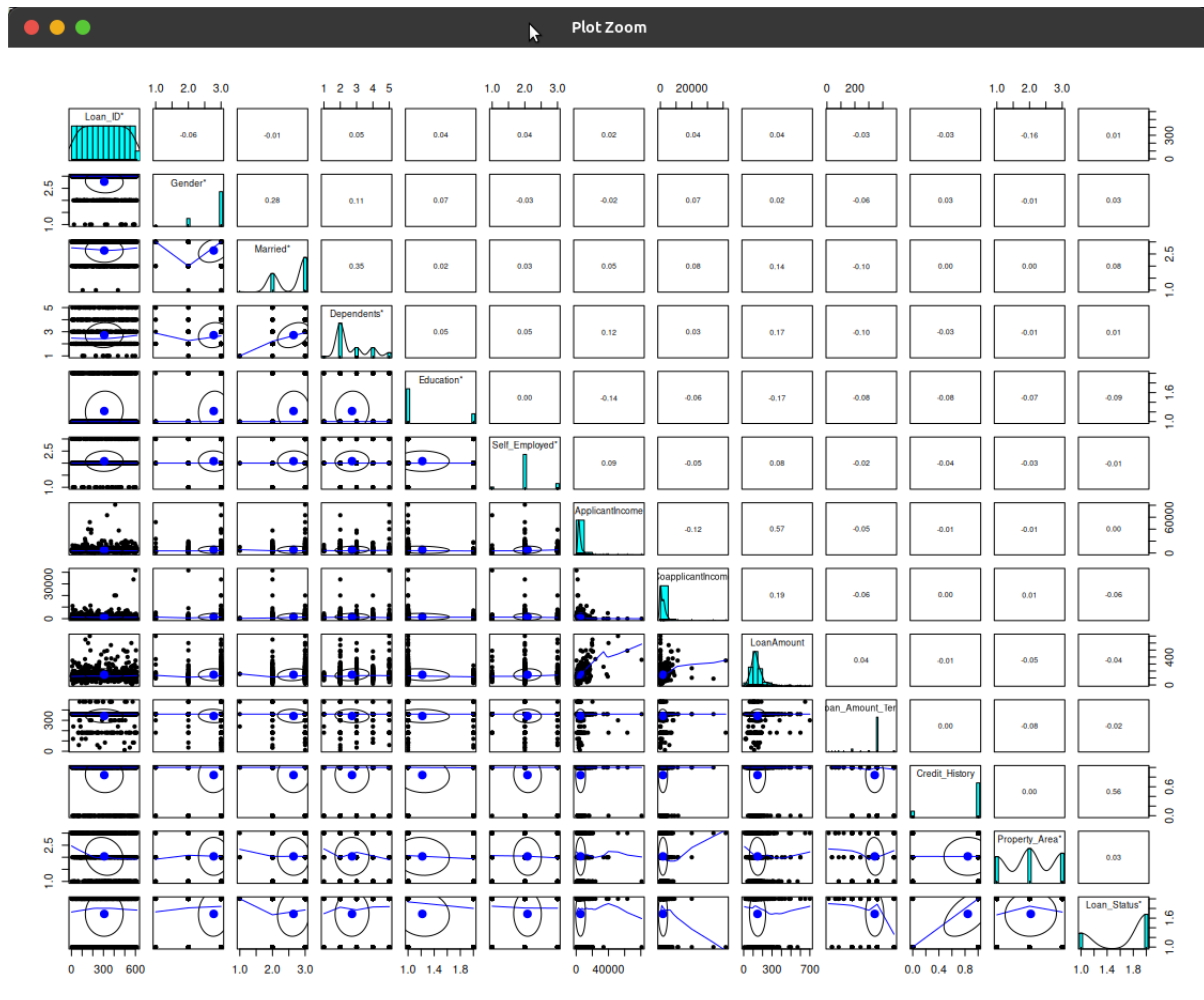
### Conclusão:

Utilizando o csv loan, conseguimos extrair possíveis insights que explicam o valor de empréstimo de variáveis independentes sobre variáveis dependentes.

Na aula anterior fizemos análises relacionadas ao “ApplicantIncome”, nesta atividade nosso foco foi sobre a variável “LoanAmount” que seria o valor de empréstimo.

De início, utilizamos a função “pairs.panels” da biblioteca “psych” para obter uma visão geral das correlações entre as variáveis do dataframe. Nos gráficos gerados, os histogramas são exibidos na diagonal principal, representando a distribuição de cada variável individualmente.

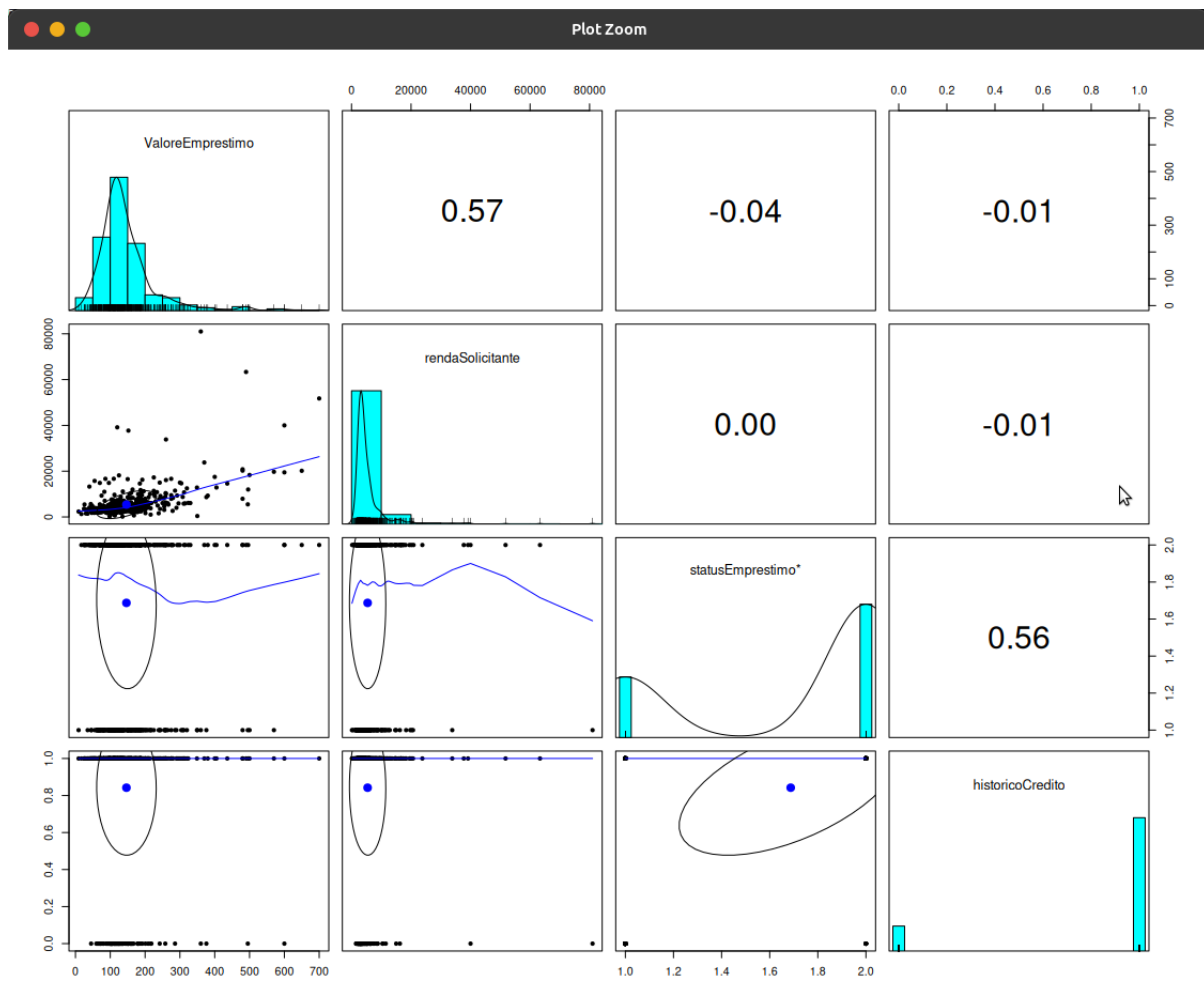
Ilustração do plot:



Como podemos ver, abaixo das diagonais estão os gráficos de dispersão de cada variável.

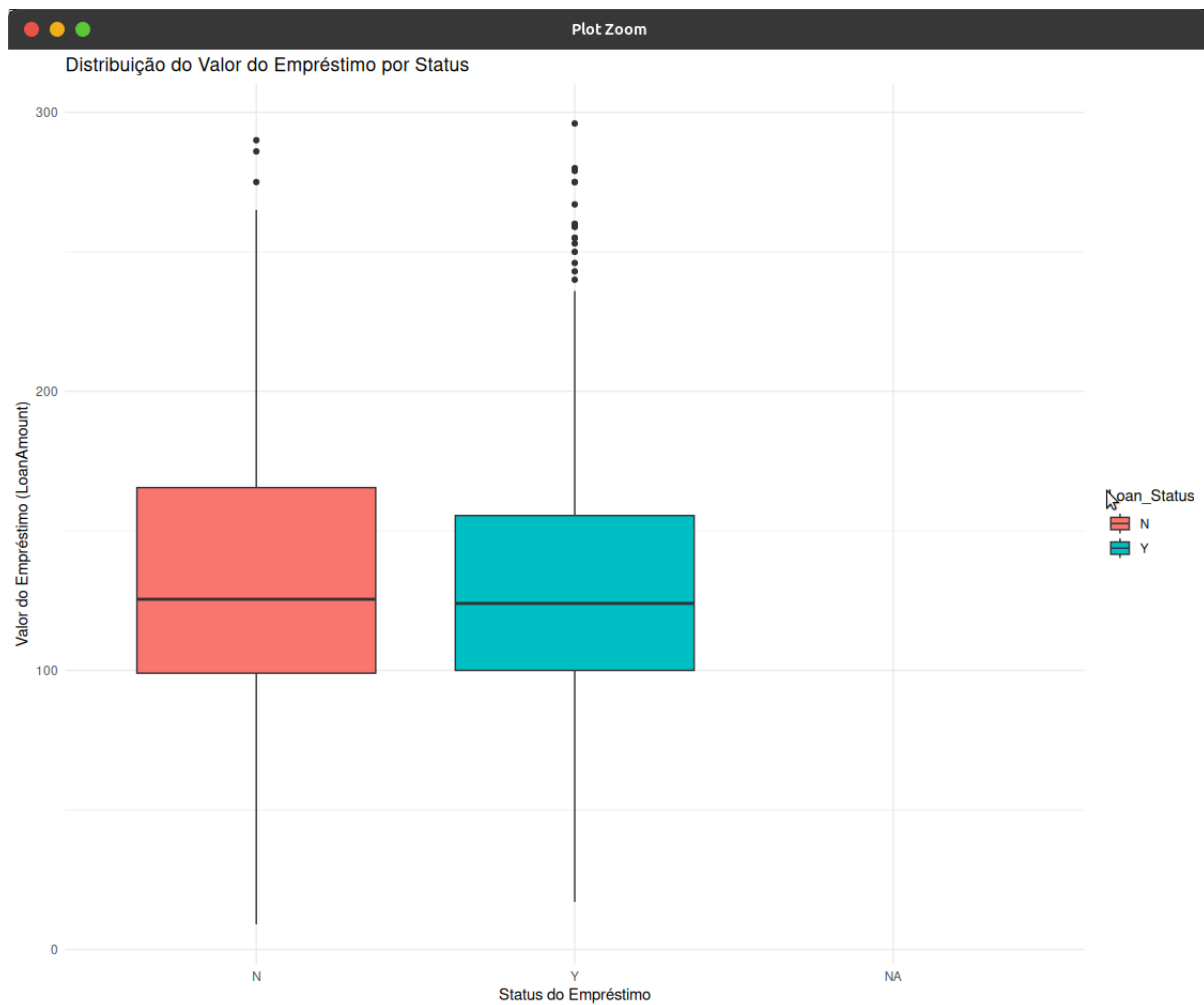
Em seguida, fizemos um novo dataframe com as colunas que iremos usar nessas análises (LoanAmount, ApplicantIncome, Loan\_Status e Credit\_History).

Veja a seguir como ficou o pairs.panels desse novo dataframe:



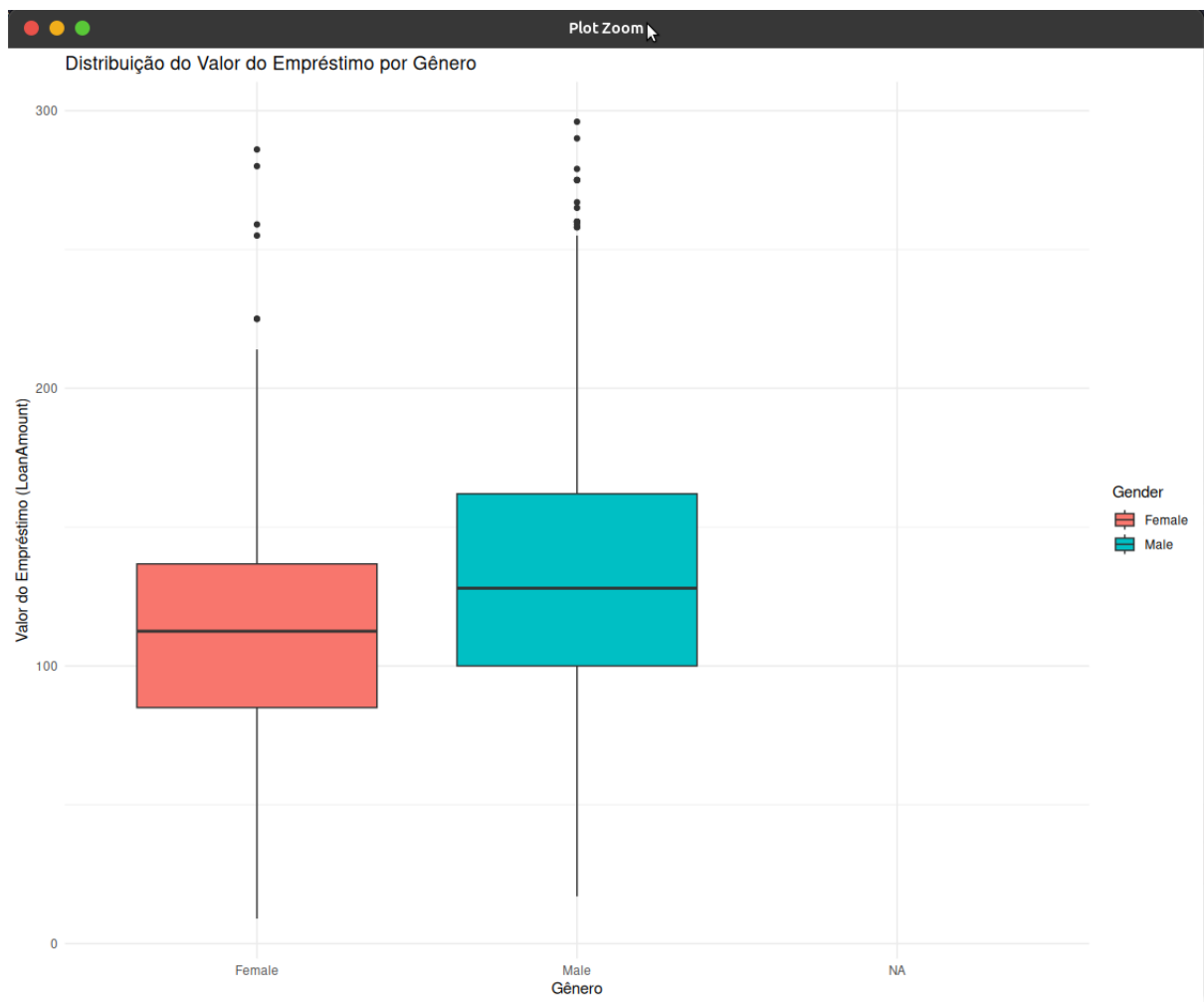
O próximo passo é filtrar a exibição do status de empréstimo (y = possui empréstimo, n = não possui empréstimo) com o valor do empréstimo (LoanAmount)

Veja a seguir como ficou a exibição desse plot:



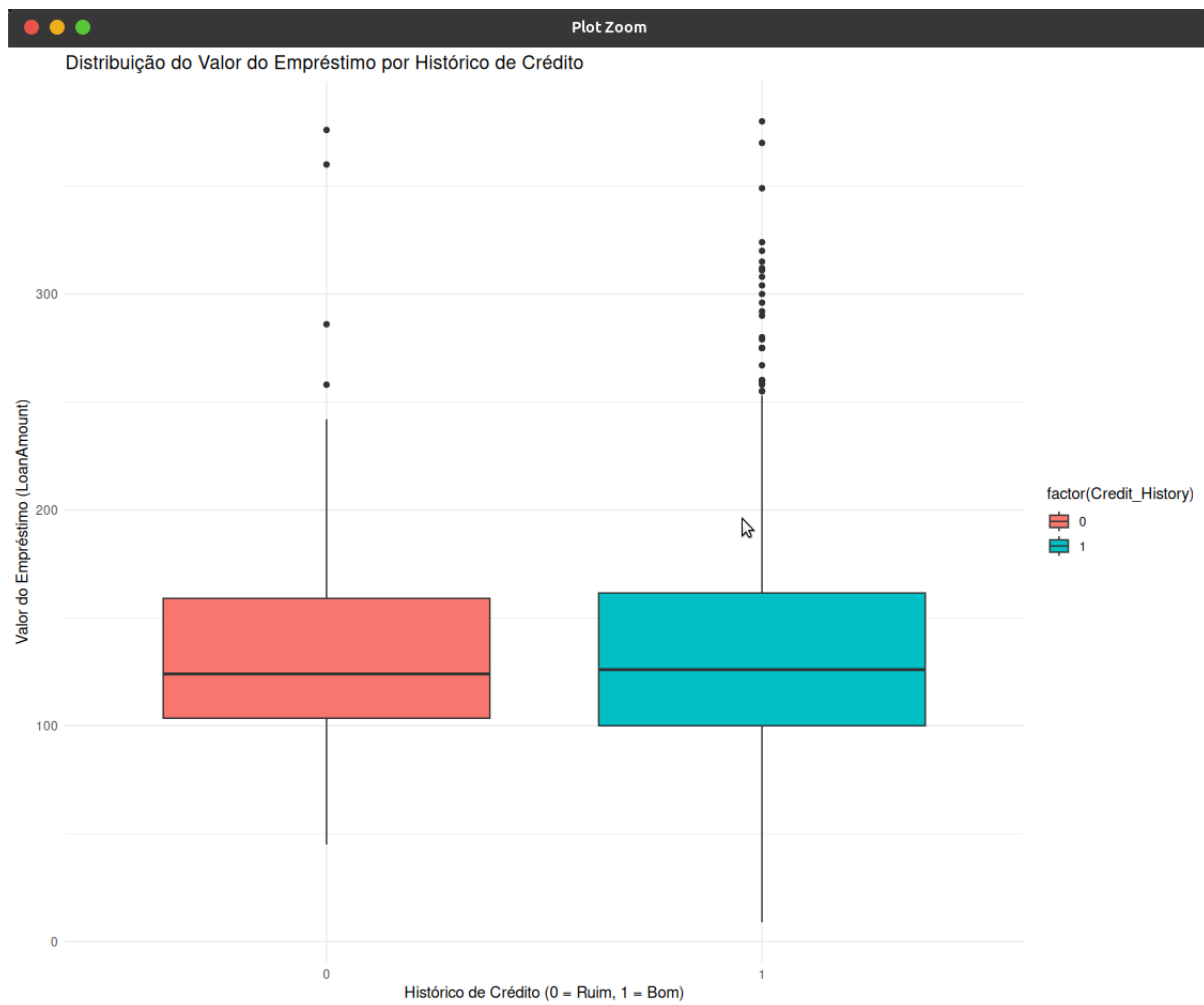
A análise revelou que existe uma distribuição diferenciada dos valores de empréstimo entre aprovados e rejeitados. Empréstimos aprovados tendem a ter valores ligeiramente menores, sugerindo que instituições financeiras são mais conservadoras com valores maiores.

Prosseguindo com a filtragem, vamos ver como fica a distribuição dos valores de empréstimos por gênero (male, female)



A análise por gênero mostrou diferenças na distribuição dos valores solicitados, indicando possíveis padrões comportamentais distintos entre homens e mulheres na solicitação de empréstimos.

Além disso, também foi feita uma análise sobre bons pagantes e devedores, a nível de obter alguma resposta relacionada com a situação do cliente.



O histórico de crédito demonstrou ser um fator crucial, com clara diferenciação entre solicitantes com bom e mau histórico. A regressão logística confirmou a significância estatística desta variável.

Outro insight que foi retirado do modelo (Credit\_History ~ ApplicantIncome) foi o seguinte:

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	1.708e+00	1.501e-01	11.373	<2e-16 ***
ApplicantIncome	-5.940e-06	1.703e-05	-0.349	0.727

Contrariamente ao esperado, a renda do solicitante não tem relação estatisticamente significativa com o histórico de crédito

Por fim, teve outra análise com o modelo (Credit\_History ~ Dependents) a nível de descobrir se a variável dependente influencia com o histórico de crédito, o resultado foi o seguinte:

Coefficients:				
	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	0.6931	0.5477	1.266	0.2057
Dependents0	1.1023	0.5709	1.931	0.0535 .
Dependents1	0.9985	0.6202	1.610	0.1074
Dependents2	1.0622	0.6195	1.715	0.0864 .
Dependents3+	0.5199	0.6465	0.804	0.4213

Com isso, chegamos à conclusão (de acordo com o modelo) que pessoas que não possuem dependentes tendem a ter melhores históricos de crédito.