

# Trabalho prático de Introdução a R

Alunos: Riam Martinelli & Hézio Silva Dos Santos

## INTRODUÇÃO

O banco de dados ChEbi foi uma iniciativa muito importante para o mundo dos dados químicos e para a bioinformática, que por sua vez, nasceu inspirado na negligência da comunidade de biologia computacional, que em publicações quase mínimas vezes apresentavam dados químicos como algo digno de atenção.

Consideraram então, essa abordagem fundamentalmente falha, e que um recurso de acesso aberto e de boa qualidade possui um valor indiscutível, tanto para o contexto de vias metabólicas ou ligantes de proteínas.

A fim de abordar esta questão, em 2002, um projeto foi iniciado no Instituto Europeu de Bioinformática (EBI) para criar um dicionário definitivo, disponível gratuitamente, de Entidades Químicas de Interesse Biológico (ChEBI; pronuncia-se /'keb1/). A principal motivação era fornecer um produto de alta qualidade, vocabulário controlado minuciosamente anotado para promover o uso correto e consistente de terminologia bioquímica inequívoca em todos os bancos de dados de biologia molecular no EBI.

## DESCRIÇÃO DO BANCO DE DADOS

---

Entidades Químicas de Interesse Biológico (ChEBI), é um dicionário gratuito de entidades moleculares focado em 'pequenos' compostos químicos. As entidades moleculares em questão são produtos naturais ou produtos sintéticos usados para intervir nos processos dos organismos vivos. Além das entidades moleculares, o ChEBI contém grupos (partes de entidades moleculares) e classes de entidades.

ChEBI é projetado como um banco de dados relacional, que é implementado em um servidor de banco de dados Oracle. Vários aplicativos utilitários, implementados principalmente em scripts Java e Unix, fornece funcionalidade adicional ao banco de dados, como o carregamento de dados de fontes externas. Interfaces especializadas baseadas na Web fornecem acesso público aos dados e acesso restrito à ferramenta de anotação.

Para criar o ChEBI, dados de várias fontes foram incorporados e submetidos a procedimentos de fusão para eliminar a redundância. 4 dos principais dados em que são extraídas as informações: 1º IntEnz, 2º KEGG COMPOUND, 3º PDBeChem, 4º ChEMBL.

Nada mantido no banco de dados é proprietário ou derivado de uma fonte proprietária que limitaria sua distribuição/disponibilidade gratuita para qualquer pessoa. Cada item de dados no banco de dados é totalmente rastreável e explicitamente referenciado à fonte original. Embora o objetivo inicial do ChEBI fosse padronizar a terminologia bioquímica, a necessidade de armazenar e representar as estruturas químicas 2D foi reconhecida desde o início.

Dentro do site existem diversos banco de dados disponíveis para uso, o que utilizamos foi o arquivo [CHEBI complete\\_3star.sdf](#), contém todas as estruturas químicas e informações associadas. Observe que exclui qualquer informação ontológica, pois as classes ontológicas não podem ser representadas, pois não contêm uma estrutura.

## PRIMEIRO LANÇAMENTO & ÚLTIMA ATUALIZAÇÃO

---

Seu primeiro lançamento público foi em (21 de julho de 2004).

O ChEBI representava mais de 12.000 entidades moleculares, grupos e classes. O termo 'entidade molecular' refere-se a qualquer átomo, molécula, íon, par de íons, radical, íon radical, complexo, confórmero, etc.

Sua última atualização foi em (09 de 11 novembro de 2022).

Em sua versão mais recente, está introduzido ao ChEbi dezenas de milhares de entidades moleculares, grupos, classes e diversas outras informações relevantes.

## Descritores do RDKit

---

SMILES = É uma forma de representar estruturas químicas usando caracteres ASCII.

Molecular Formula = Indica os números de cada tipo de átomo em uma molécula.

ExactMolWt = O peso molecular exato da molécula.

MolLogP = Cálculo de LogP baseado em átomos.

(LogP = valor experimental do coeficiente de partição)<sup>o</sup>

NumRotatableBonds = Número de ligações que permitem a rotação livre em torno de si.

RingCount= A contagem exata de anéis da molécula.

TPSA= Área de superfície polar topológica, definida a partir da soma da superfície sobre todos os átomos ou moléculas polares.

Sobre os dois últimos descritores: Existe uma ligação de hidrogênio intramolecular que ocorre quando uma mesma molécula apresenta, simultaneamente, **um grupo doador e outro receptor de próton, em configuração espacial favorável à formação dessa interação.**

NumHAcceptors = Número de aceitadores de ligações de hidrogênio.

NumHDonors = Número de doadores de ligações de hidrogênio.

Resultados das análises do R

---

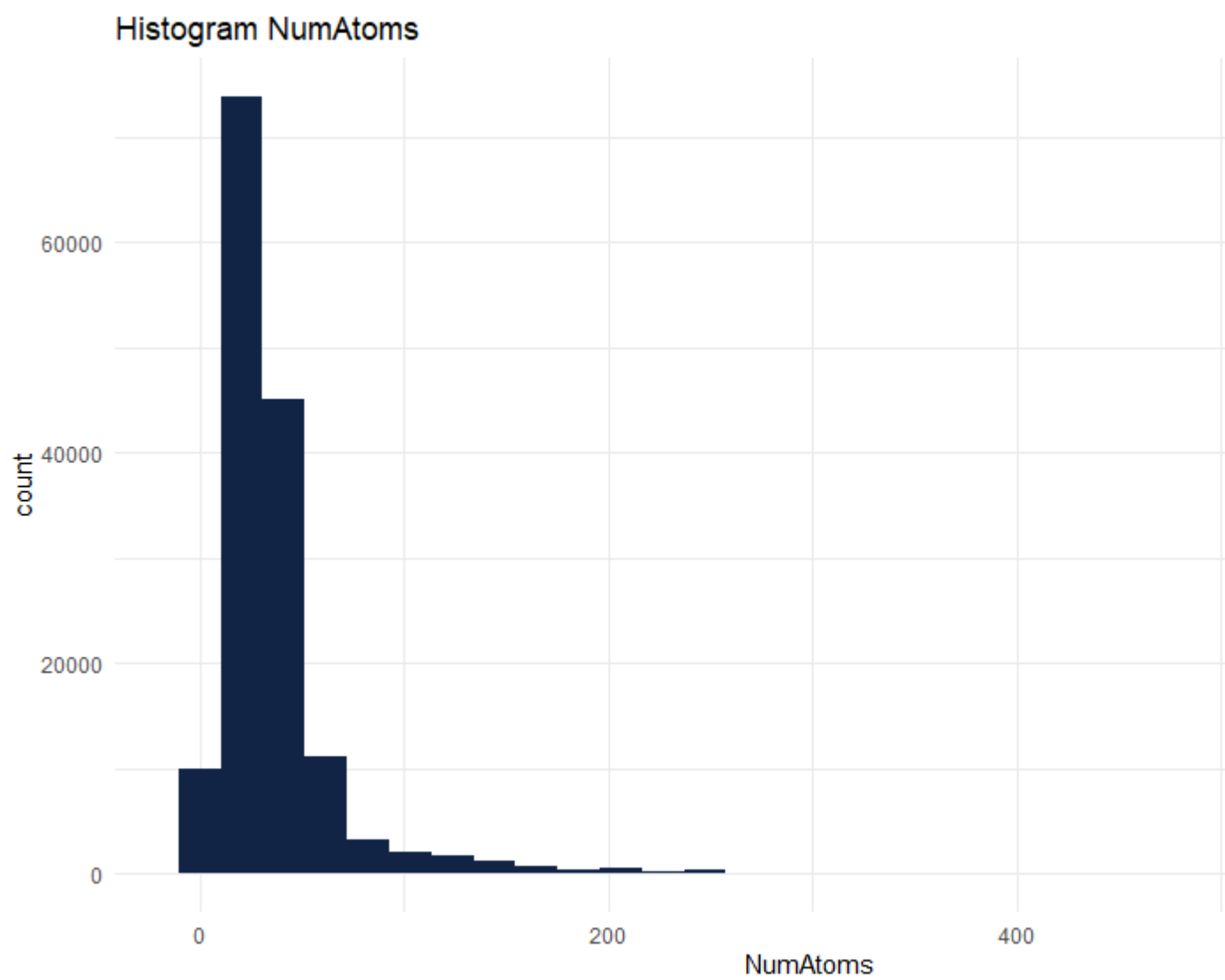
Valor máximo, média e mínimo de cada descritor.

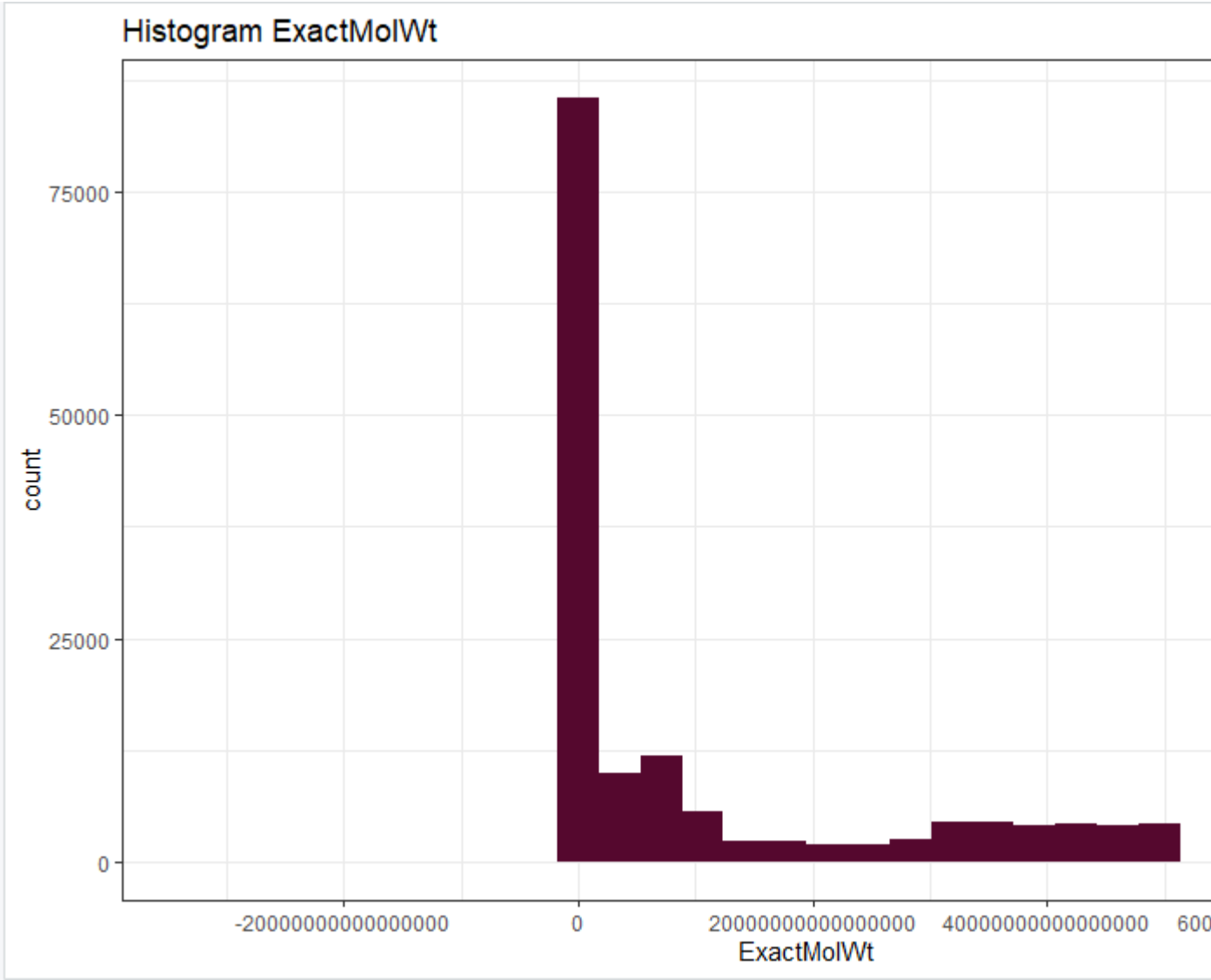
	▲ Descriptors ▼	Max ▼	Mean ▼	Min ▼
1	NumAtoms	5.980000e+02	3.539143e+01	0.000000e+00
2	ExactMolWt	6.974957e+16	9.802132e+15	-3.291479e+16
3	NumRotableBonds	2.600000e+02	1.048072e+01	0.000000e+00
4	RingCount	7.500000e+01	2.869599e+00	0.000000e+00
5	NumHAcceptors	2.510000e+02	9.036770e+00	0.000000e+00
6	NumHDonors	1.310000e+02	4.640772e+00	0.000000e+00

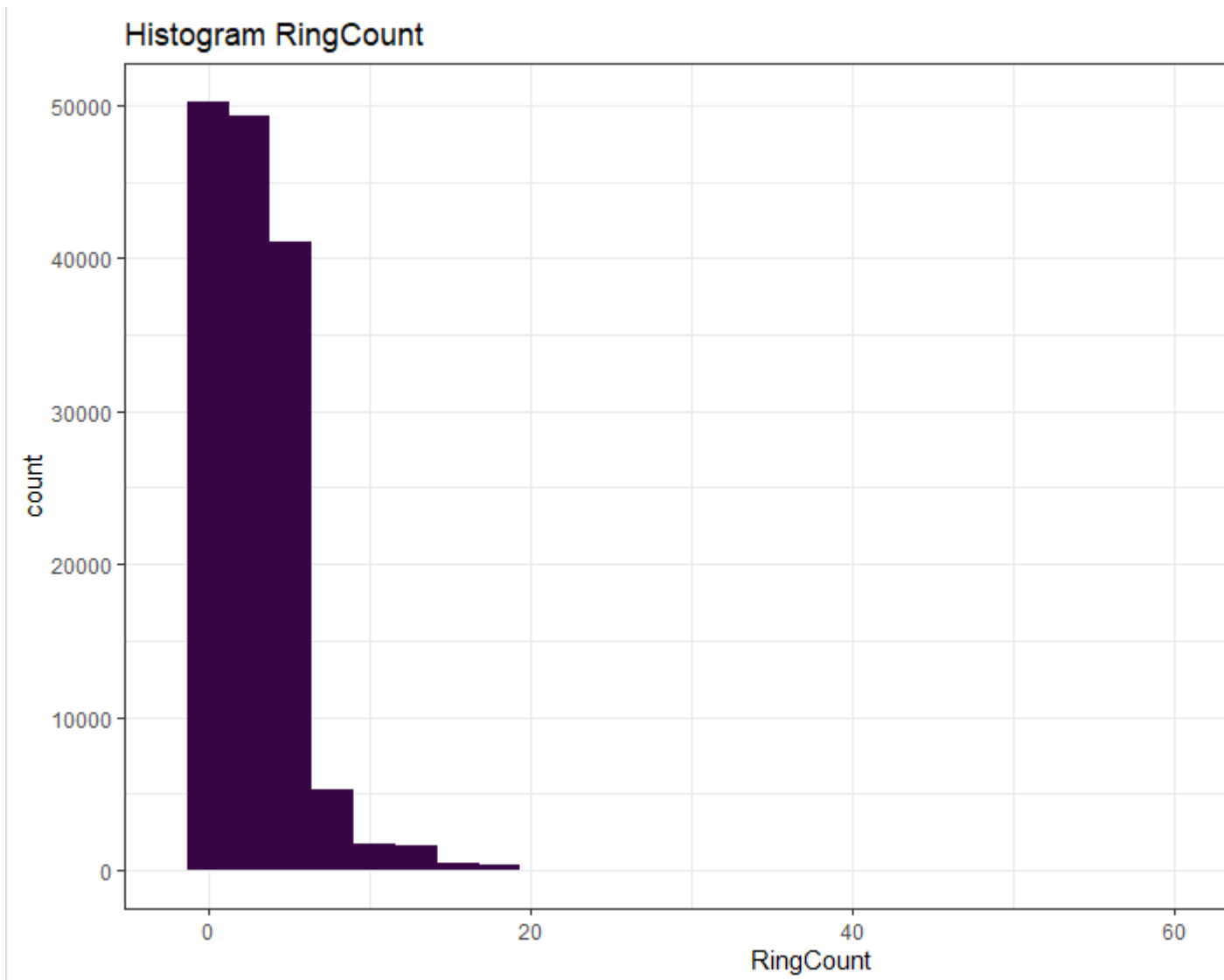
## Fim das páginas textuais

## Histogramas e Plots

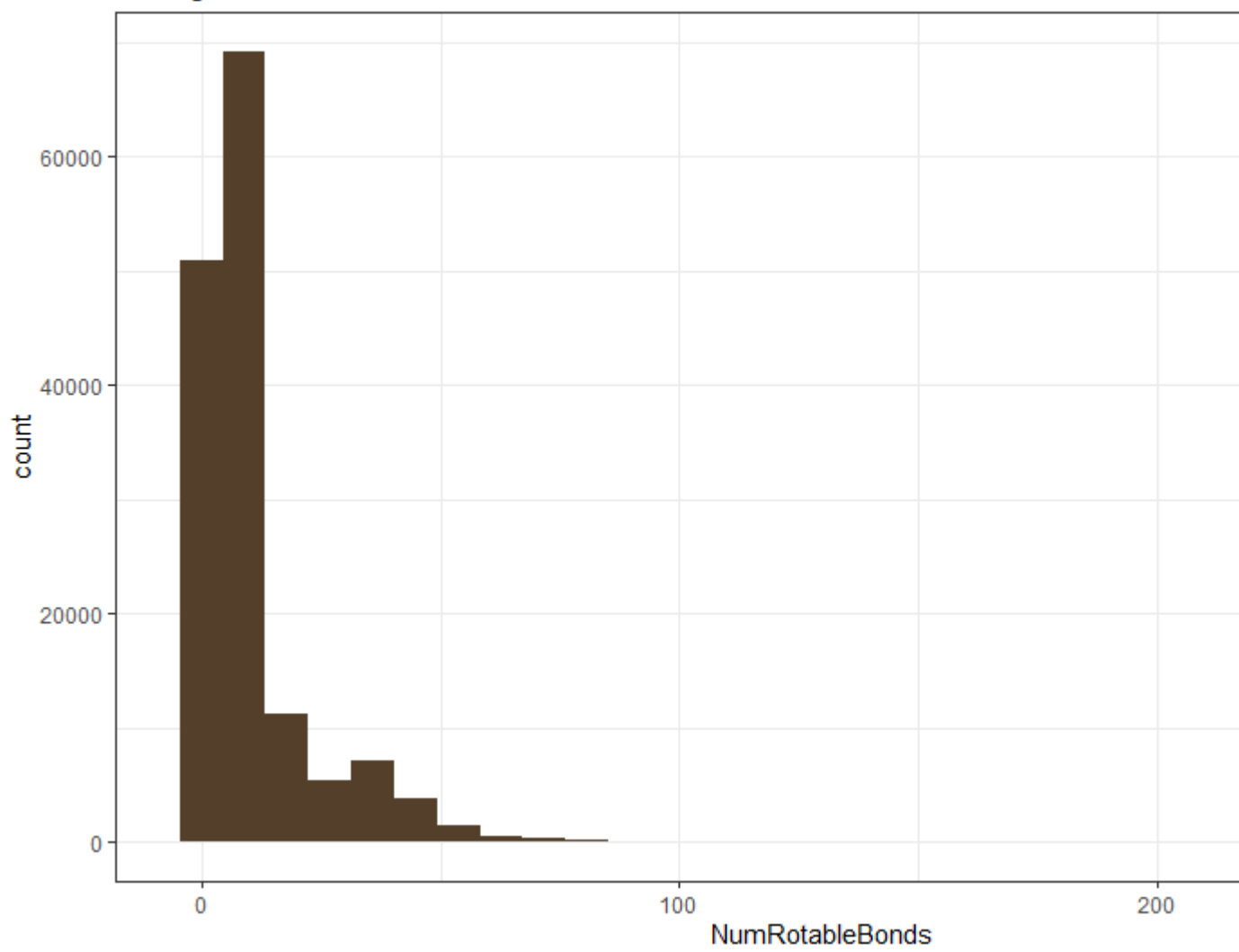
---

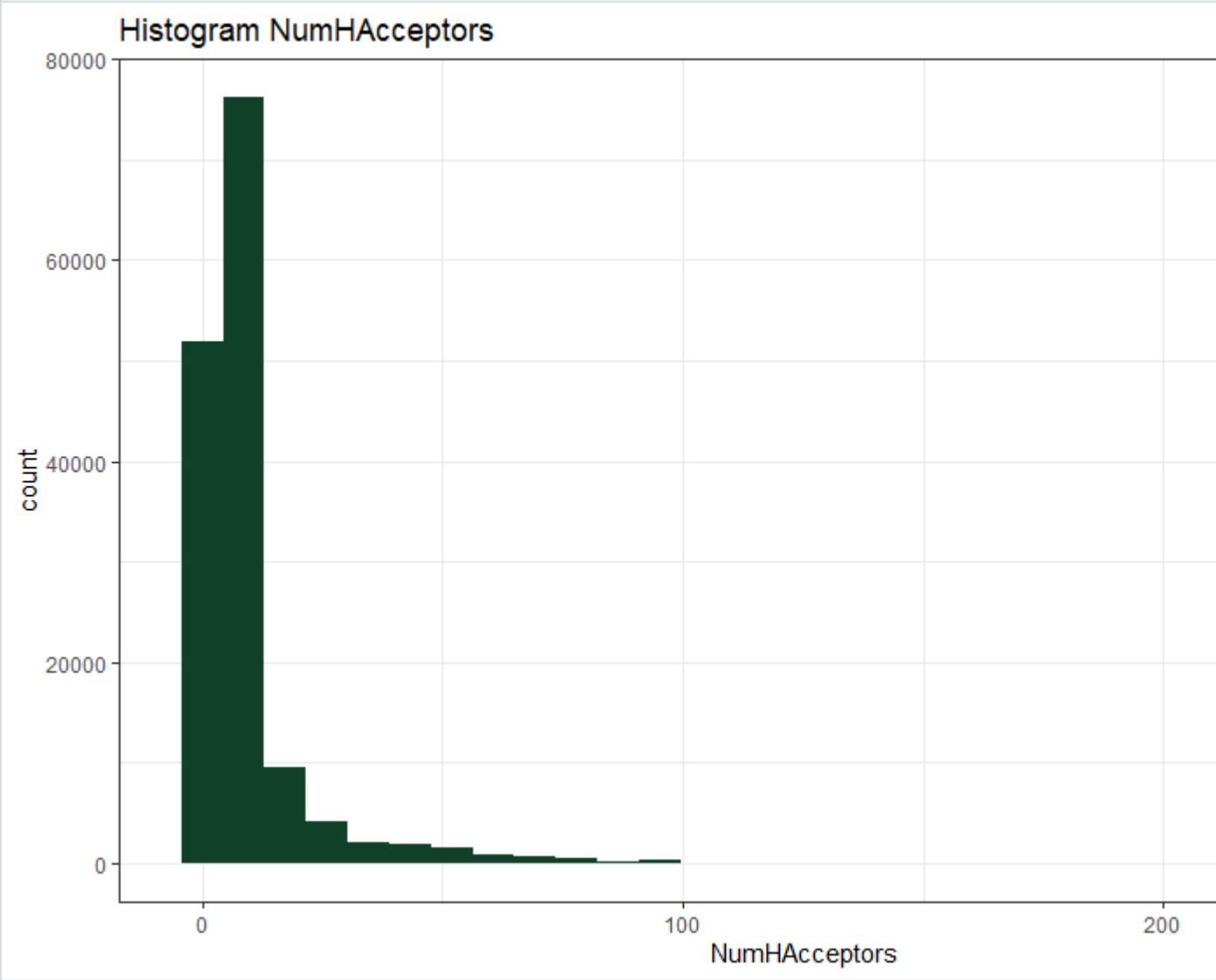




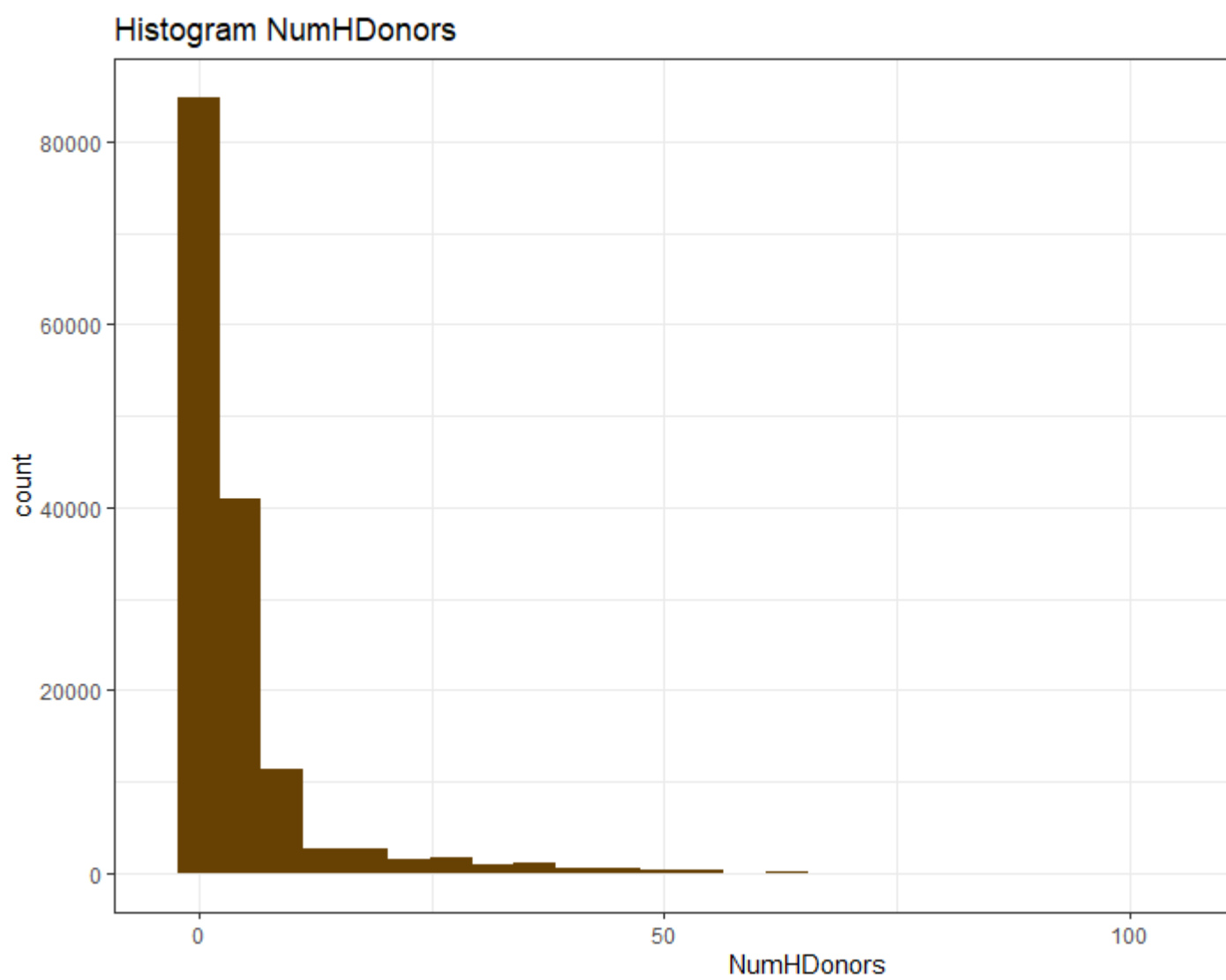


Histogram NumRotableBonds

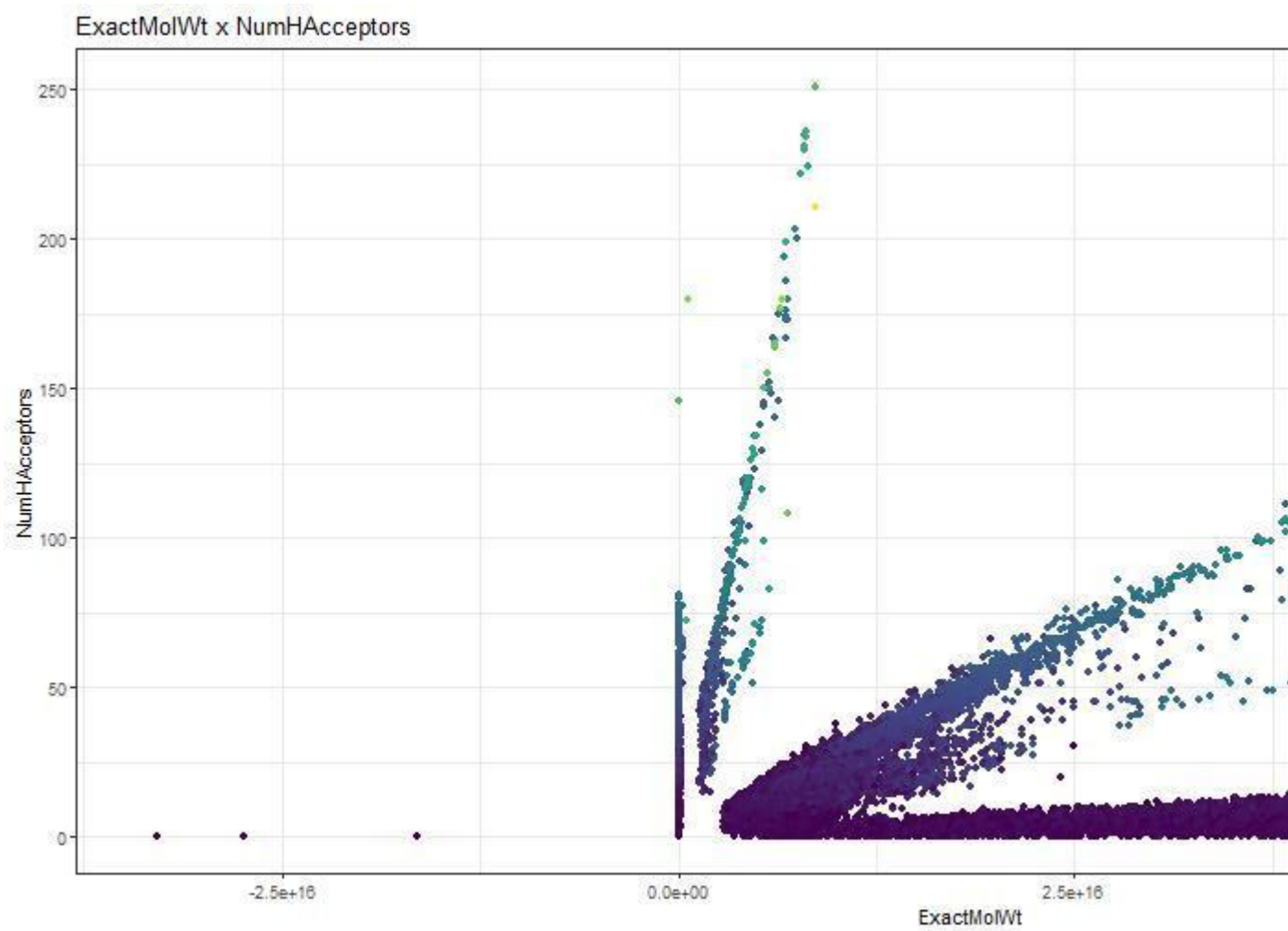




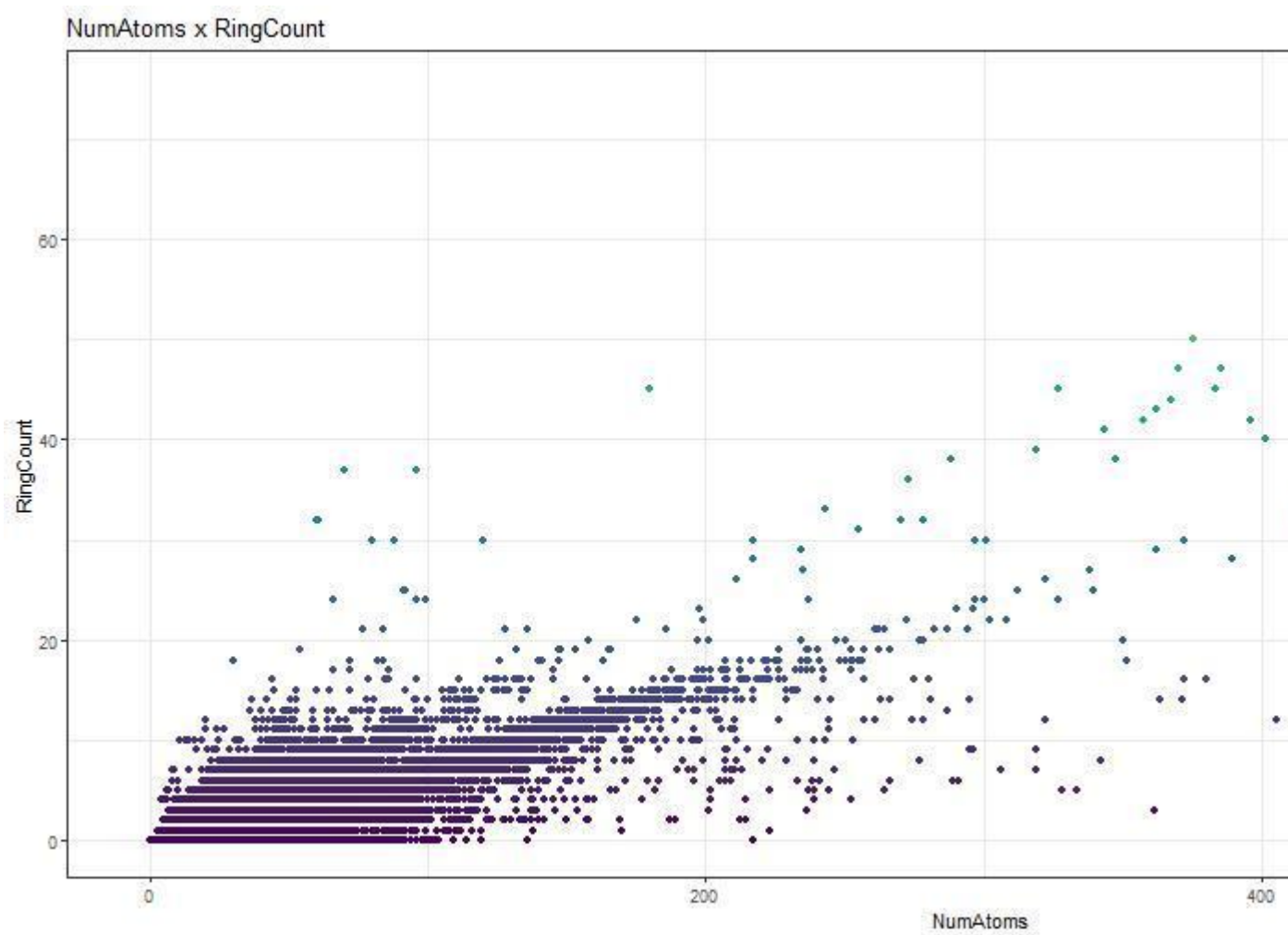




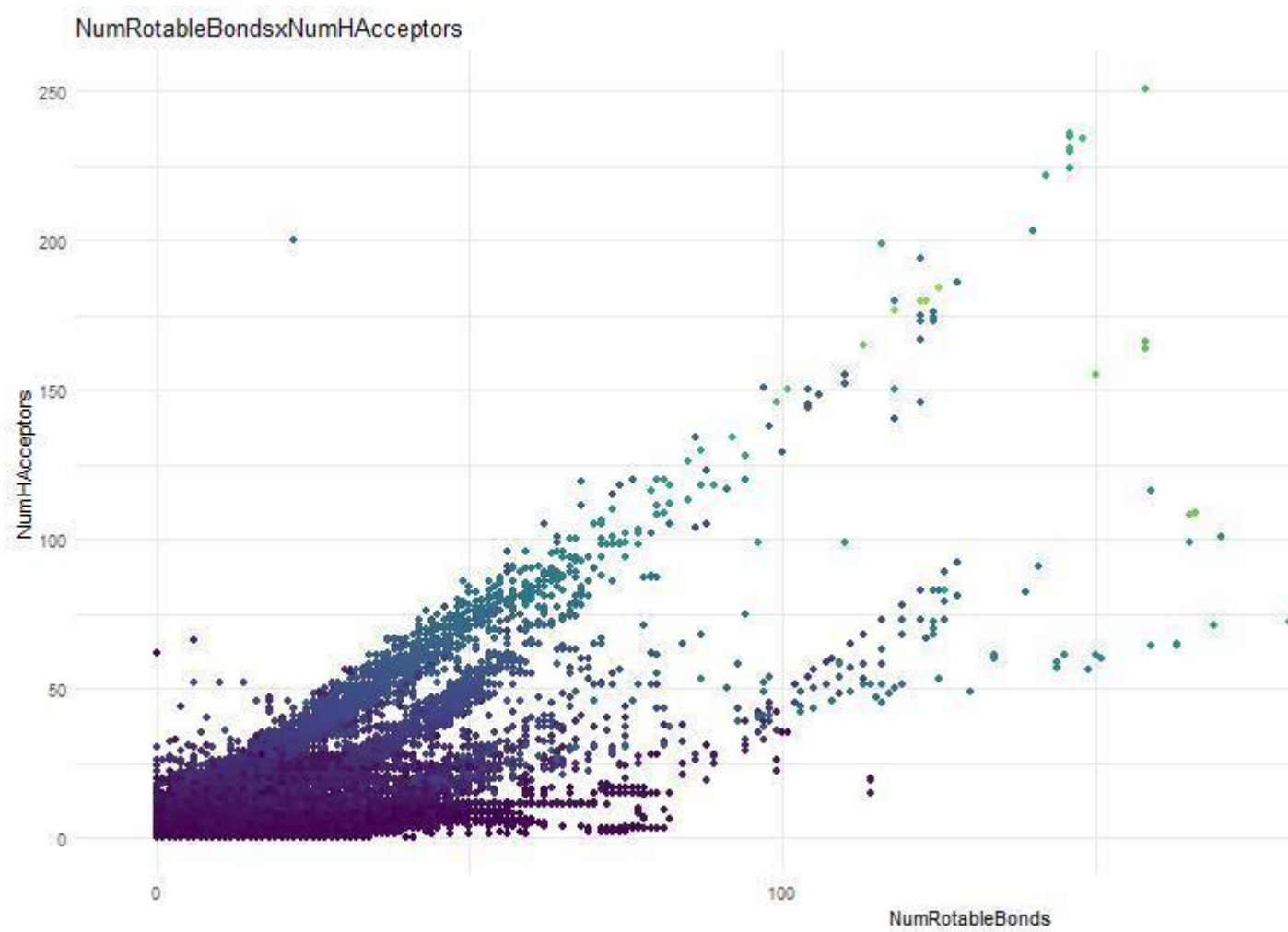
Peso molecular x Números de aceptadores



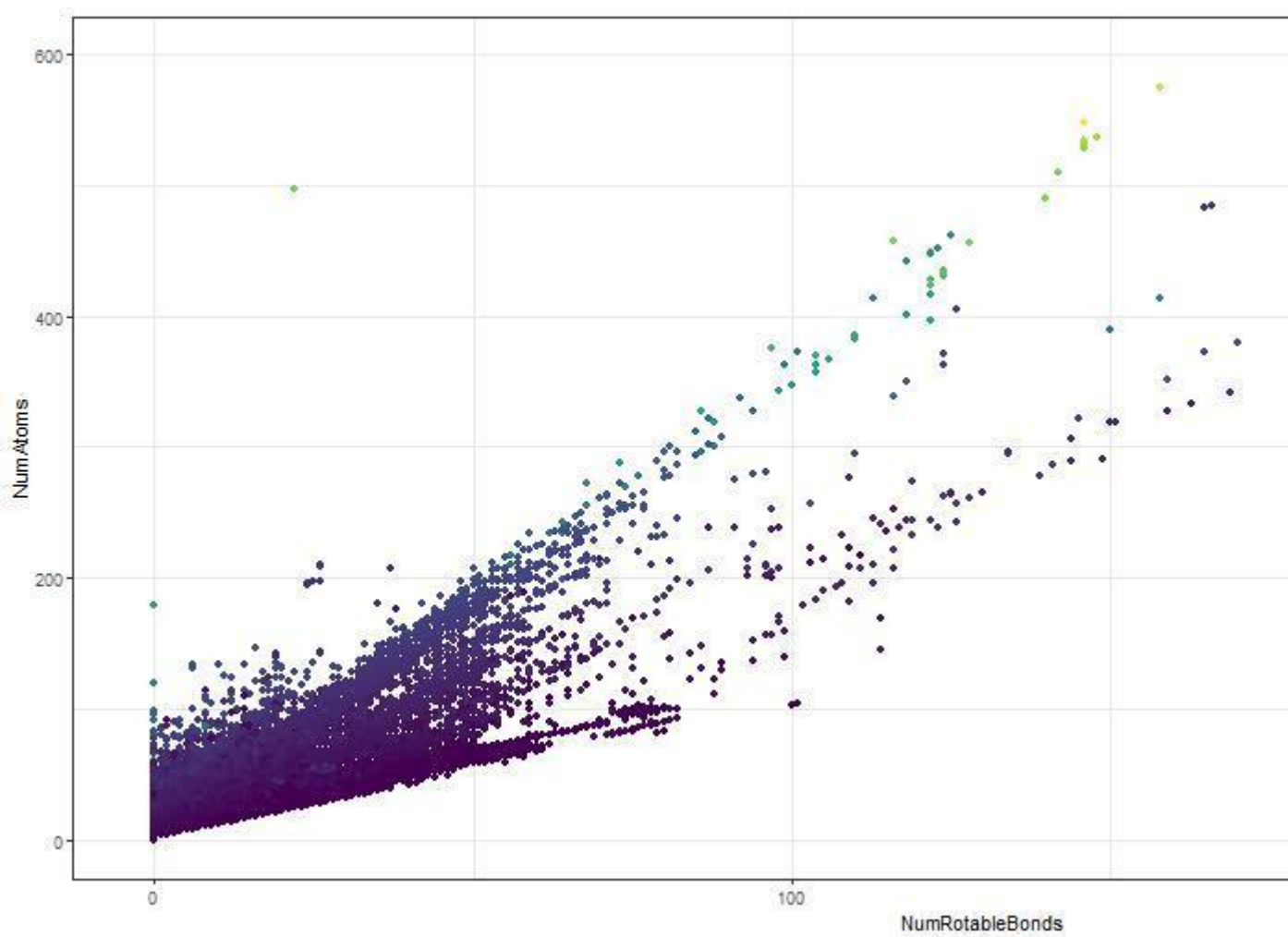
Número de átomos x Quantidade de anéis



Número de ligações x Número de aceptadores



Número de ligações x Número de átomos



Número de aceptadores x Número de donadores

NumHAcceptors x NumHDonors

