

Análise do Conjunto de Dados RCSB PDB

Explorando Estruturas e Propriedades Moleculares

RCSB PDB (Research Collaboratory for Structural Bioinformatics Protein Data Bank)

Gerenciamento e Atualização do Dados

Visualizações - Boxplots

Análise e Visualização de Dados

Visualizações - Scatterplots

Pré-processamento de Dados

Visualizações - Gráficos de Densidade

Visualizações - Histogramas

O Banco de Dados RCSB PDB

Armazenamento

A RCSB PDB é uma base de dados com informações tridimensionais de macromoléculas biológicas.

Gerenciamanento

Gerenciada por três instituições:
Universidade Rutgers,
Instituto de Tecnologia da Califórnia e Instituto de Pesquisa Scripps.

Financiamanento

A RCSB PDB é financiada pela National Science Foundation, pelo National Institutes of Health e pelo Departamento de Energia dos Estados Unidos.

Registros

A RCSB PDB possui mais de 200.000 registros, compostos por coordenadas atômicas que representam a posição dos átomos em estruturas moleculares tridimensionais.



Análise e Visualização de Dados

Bibliotecas para visualização

Utilizamos R para análise, com bibliotecas como ggplot2, dplyr e corrplot

Conjunto de dados

- Fórmula molecular
- Peso molecular exato
- LogP
- Número de aceptadores de hidrogênio
- Número de doadores de hidrogênio
- Número de ligações rotacionáveis
- Contagem de anéis
- Área de superfície polar topológica (TPSA)

Limpeza de dados

- Pré-processamos os dados para garantir consistência.

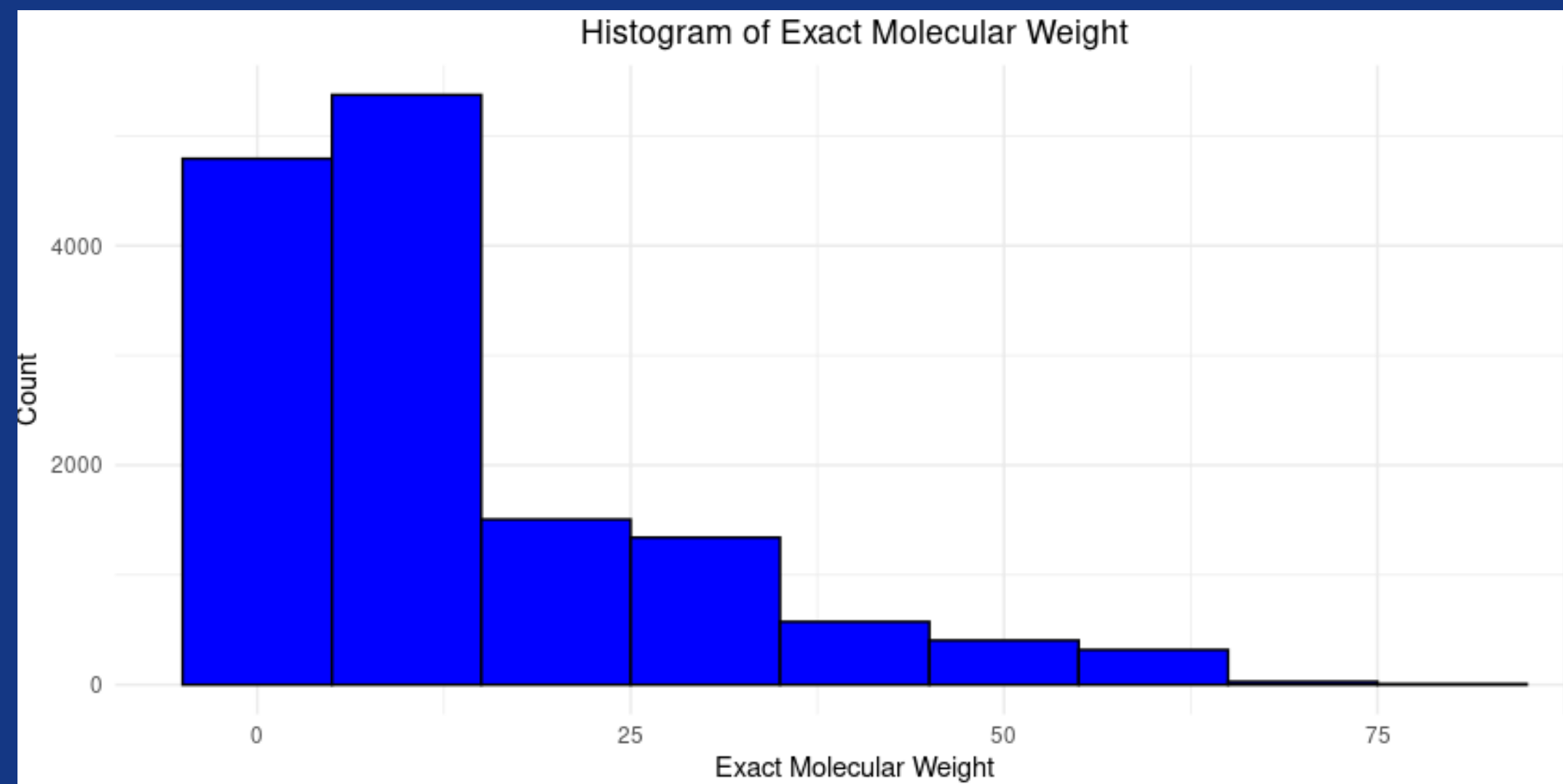
Tipo de visualização

Usamos histogramas para visualizar a distribuição de peso molecular exato e logP.

VISUALIZAÇÕES E INSIGHTS

1 Histograma de Peso

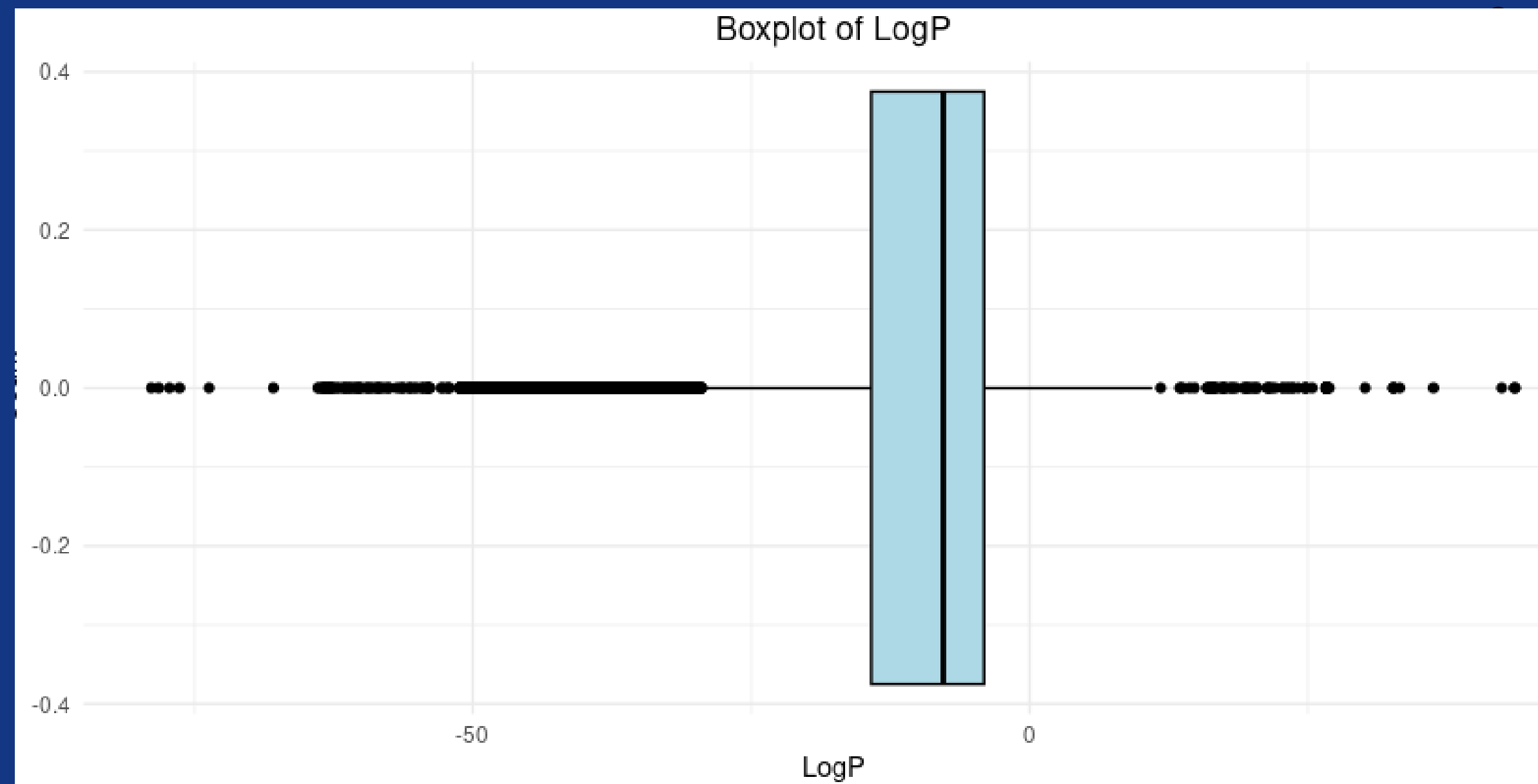
Molecular Exato: Maioria das moléculas têm peso entre 0 e 100 Daltons.



VISUALIZAÇÕES E INSIGHTS

2 Boxplot de LogP: Mediana de LogP é -7.7, indicando tendência à hidrofiliicidade.

- Isso pode ter implicações para sua atividade biológica e comportamento em um sistema biológico, pois pode afetar suas propriedades de absorção, distribuição, metabolismo e excreção (ADME).

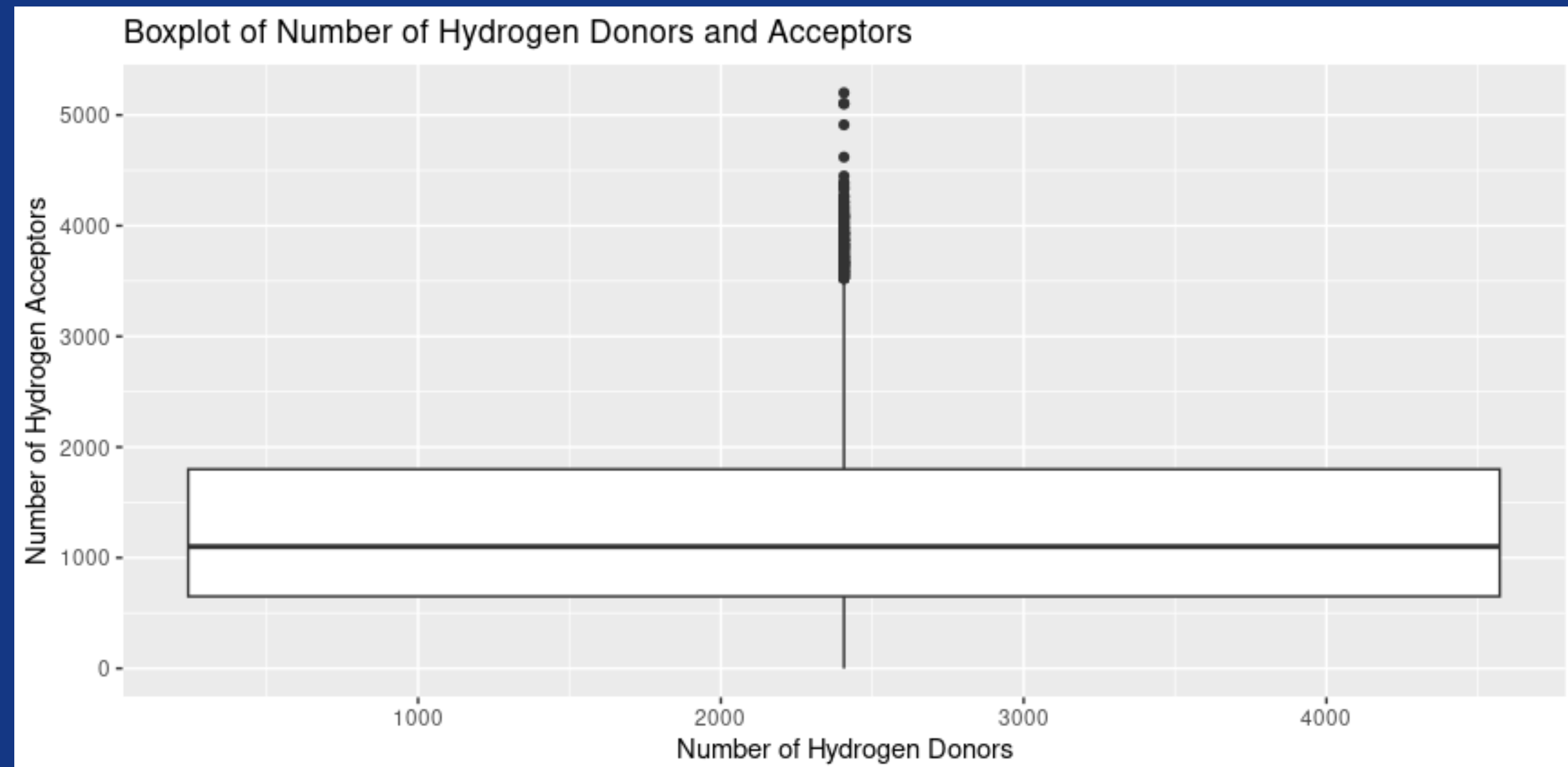


VISUALIZAÇÕES E INSIGHTS

Boxplot de Número de Aceitadores e Doadores de Hidrogênio: Maioria das moléculas têm número moderado de aceitadores e doadores de hidrogênio.

3

- Estrutura e estabilidade de moléculas biológicas
- Interação entre moléculas



VISUALIZAÇÕES E INSIGHTS

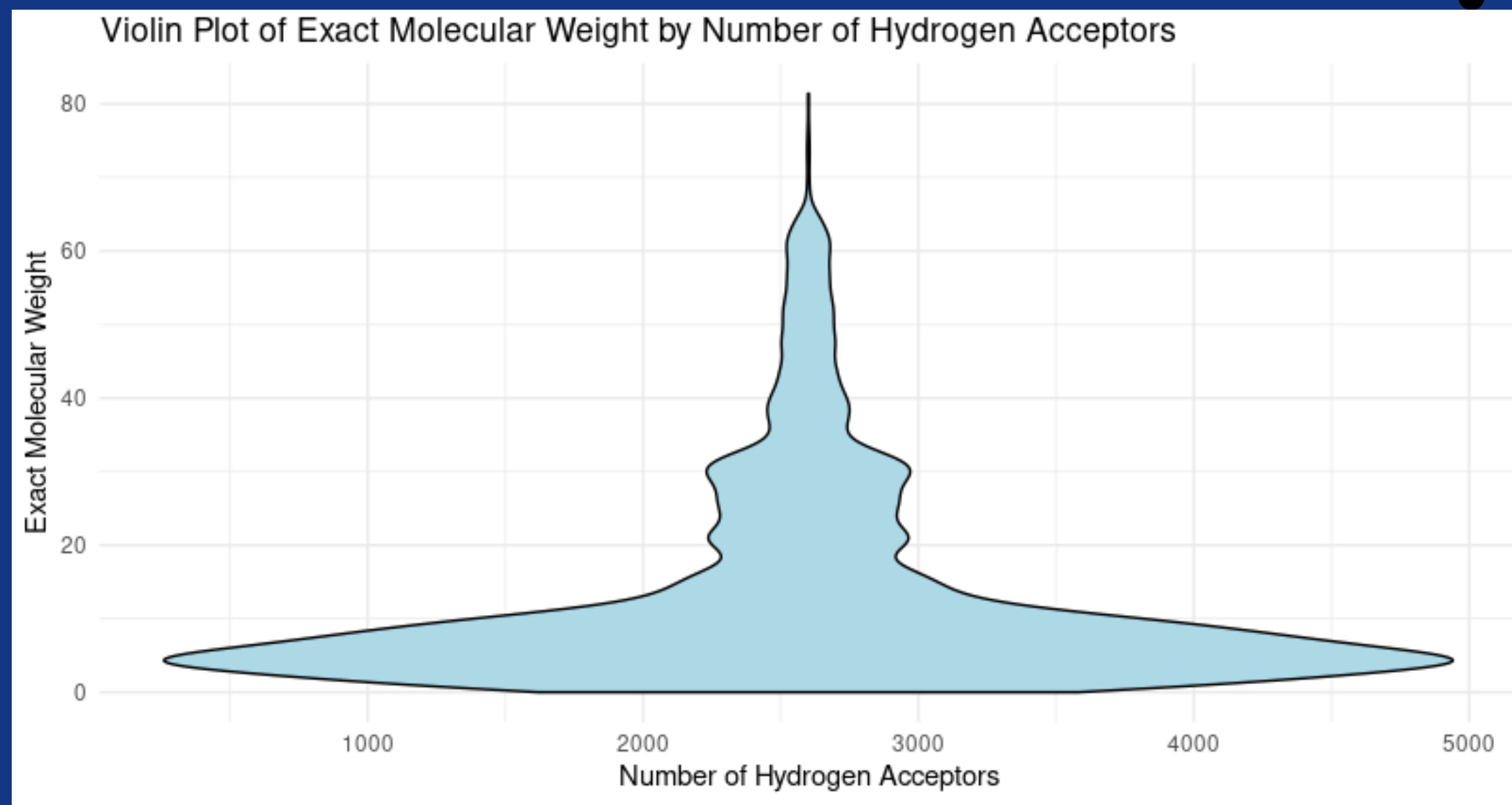
Violin Plot de Peso Molecular

Exato por Número de

4

Aceitadores de Hidrogênio:

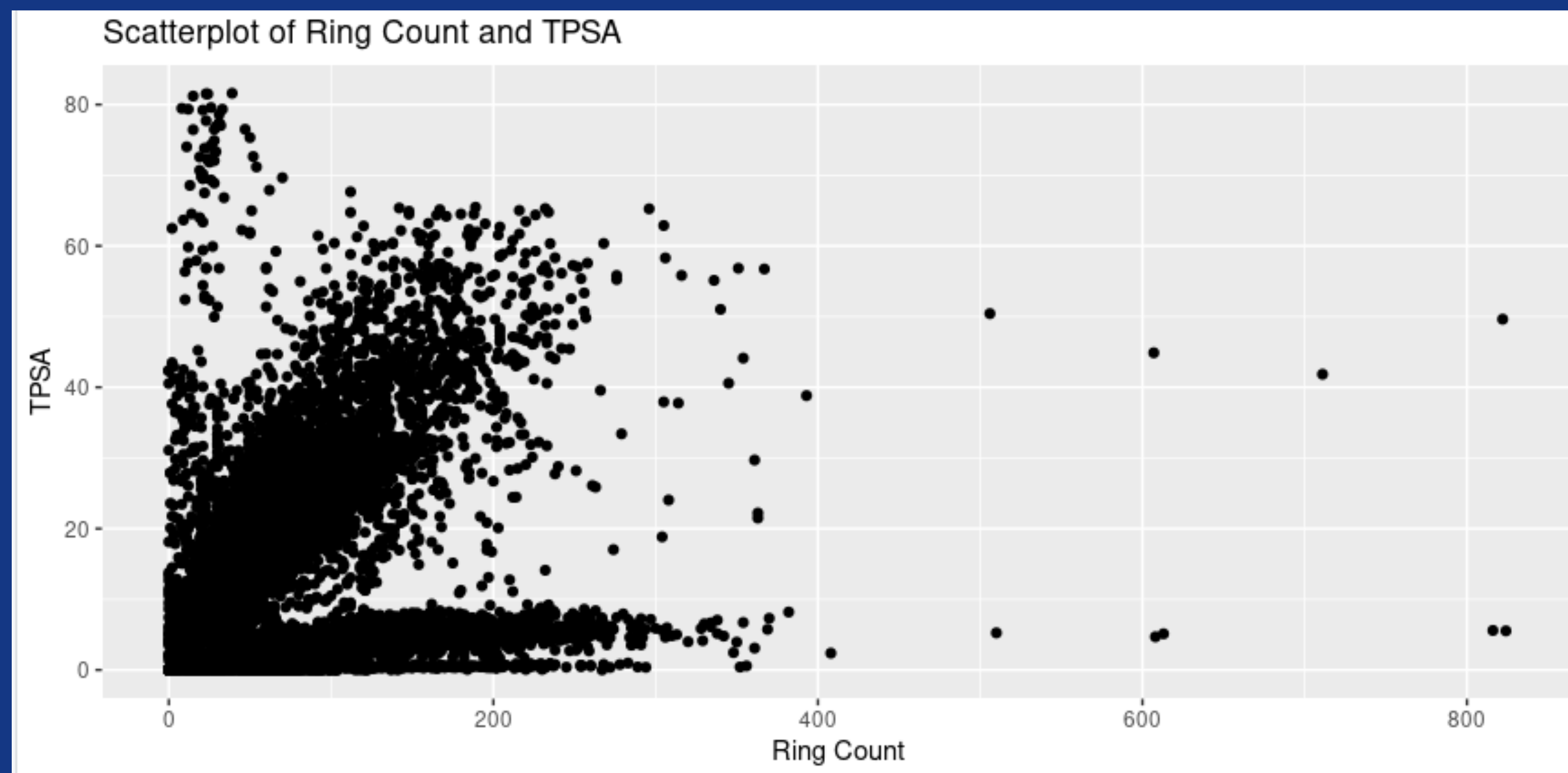
A largura do "violino" indica a densidade de moléculas com um peso molecular específico.



- A maior densidade sugere que moléculas com mais aceptores de hidrogênio tendem a ter um peso molecular maior.
- A maioria das moléculas apresenta um número moderado de aceptores de hidrogênio.

VISUALIZAÇÕES E INSIGHTS

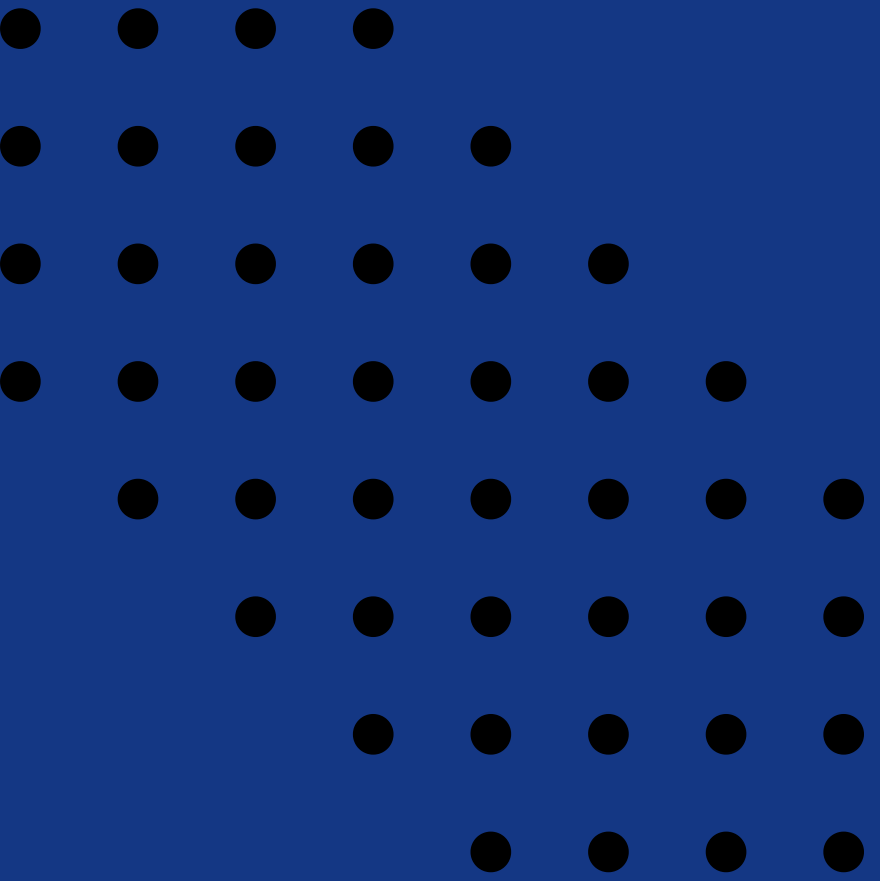
5 Gráfico de dispersão de Contagem de Anéis e TPSA: Correlação positiva entre contagem de anéis e TPSA, sugerindo conexão entre complexidade molecular e área de superfície polar.



CONSIDERAÇÕES FINAIS



- A solubilidade, influenciada pelo peso molecular e propriedades de LogP, é crucial para a eficácia de fármacos.
- A capacidade de formar ligações de hidrogênio, indicada pelo número de doadores e aceptadores, pode afetar a afinidade do fármaco pelo alvo.



Obrigado

