

MINERAÇÃO DE DADOS

Exemplos de análise de dados



ANÁLISE DE DADOS

Entender as características dos dados

Censo escolar (2021)

Pesquisa Origem Destino 2017 (OD17)

CENSO ESCOLAR

- Pesquisa de 2021
- INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira)
- Dados abertos: [resultados do censo escolar 2021](#)

CENSO ESCOLAR

- **370 atributos:**
 - **Estrutura:** salas, laboratórios, esgoto, internet, ...
 - **Pessoas:** professores, alunos, enfermeiros, ...
 - **Local:** região, cidade, ...
 - **Tipo:** pública, privada, ...
 - **Matrícula:** fundamental, EaD, vagas, ...
- **221.140 objetos (estabelecimentos)**

CENSO ESCOLAR

- **Agrupamentos existentes:**
 - **Região, estado**
 - **Pública, privada**
 - **Fundamental, Médio**

CENSO ESCOLAR - DIST. DE FREQUÊNCIA

Total de matrículas por estabelecimento

Ensino infantil, fundamental, médio, técnico, ...

~20% de valores nulos ou zero

176.932 objetos válidos

CENSO ESCOLAR - DIST. DE FREQUÊNCIA

10 classes

Limite inferior: 1, limite superior: 6960

Amplitude: 696

CENSO ESCOLAR - DIST. DE FREQUÊNCIA

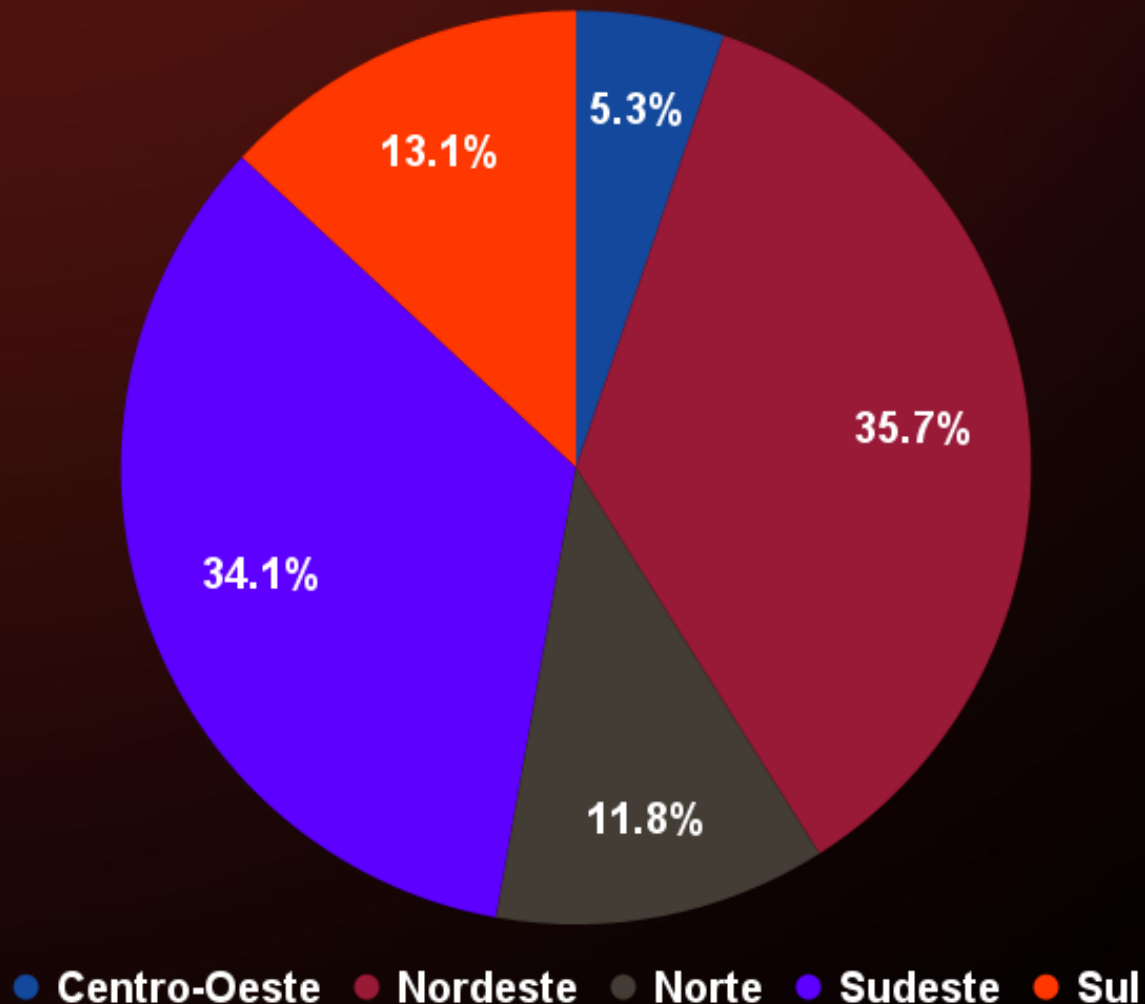
Classe	Limite inferior	Ponto médio	Limite superior	Freq. absoluta	Freq. relativa	Freq. acumulada
1	1	348	695	161474	91,26%	91,26%
2	696	1043,5	1391	13665	7,72%	98,98%
3	1392	1739,5	2087	1537	0,87%	99,85%
4	2088	2435,5	2783	183	0,1%	99,95%
5	2784	3131,5	3479	54	0,03%	99,98%
6	3480	3827,5	4175	7	0,004%	99,98%
7	4176	4523,5	4871	2	0,001%	99,98%
8	4872	5219,5	5567	6	0,003%	99,99%
9	5568	5915,5	6263	2	0,001%	99,99%
10	6264	6612	6960	2	0,001%	100%

CENSO ESCOLAR - DIST. DE FREQUÊNCIA

Classe	Limite inferior	Ponto médio	Limite superior	Freq. absoluta	Freq. relativa	Freq. acumulada
1	1	348	695	161474	91,26%	91,26%
2	696	1043,5	1391	13665	7,72%	98,98%
3	1392	1739,5	2087	1537	0,87%	99,85%
4	2088	2435,5	2783	183	0,1%	99,95%
5	2784	3131,5	3479	54	0,03%	99,98%
6	3480	3827,5	4175	7	0,004%	99,98%
7	4176	4523,5	4871	2	0,001%	99,98%
8	4872	5219,5	5567	6	0,003%	99,99%
9	5568	5915,5	6263	2	0,001%	99,99%
10	6264	6612	6960	2	0,001%	100%

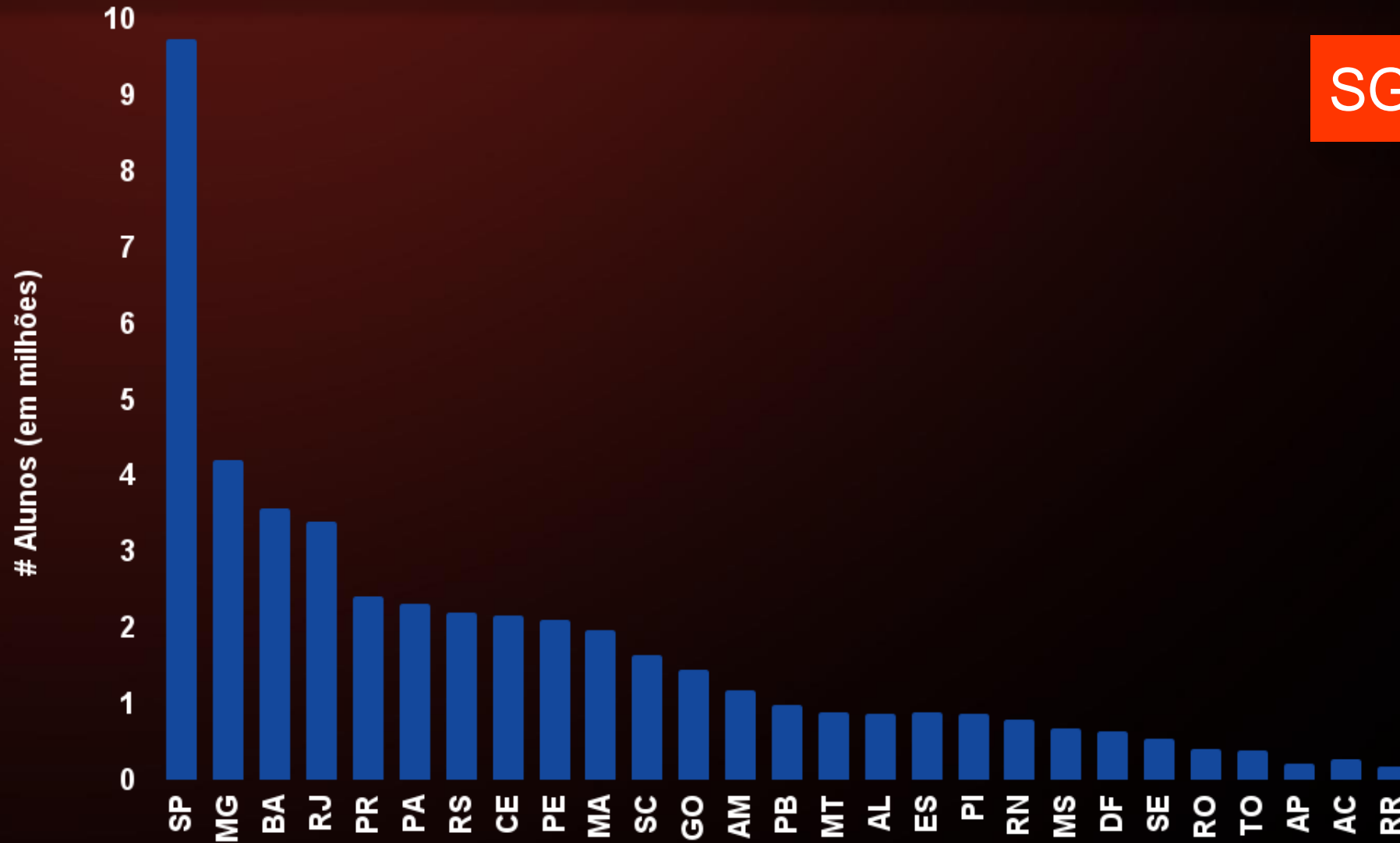
CENSO ESCOLAR – ESCOLAS POR REGIÃO

NO_REGIAO



CENSO ESCOLAR - ALUNOS POR ESTADO

SG_UF

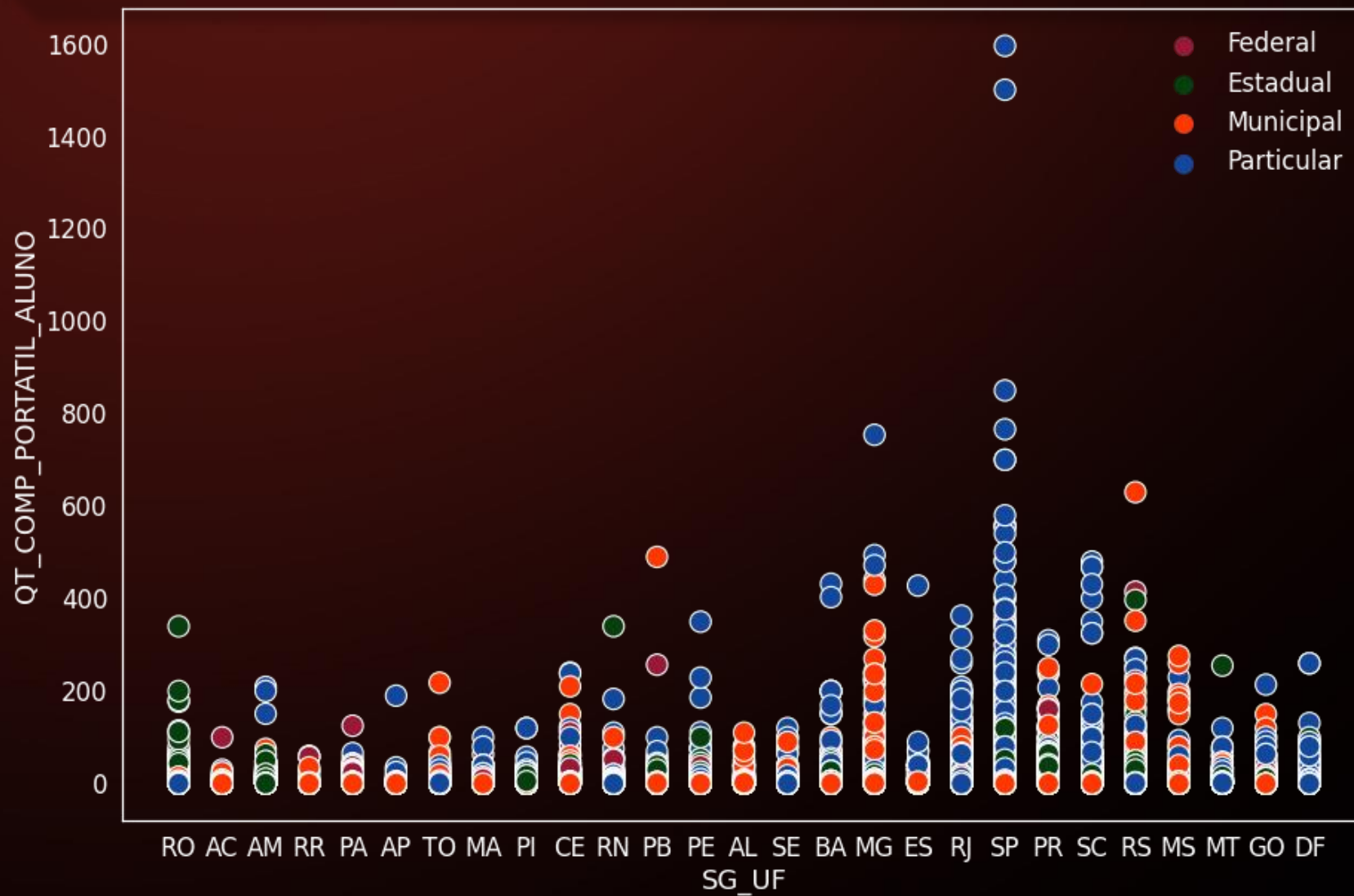


CENSO ESCOLAR – NOTEBOOKS

Quantidade de notebooks para uso dos alunos por tipo de estabelecimento e por estado

"Trivariada": UF, tipo de estabelecimento, quantidade de notebooks

CENSO ESCOLAR – NOTEBOOKS



PESQUISA ORIGEM DESTINO 2017 (OD17)

- Mobilidade urbana da região metropolitana de SP
- Metrô (Companhia do Metropolitano de São Paulo)
- Dados abertos: [resultados da pesquisa OD 17](#)



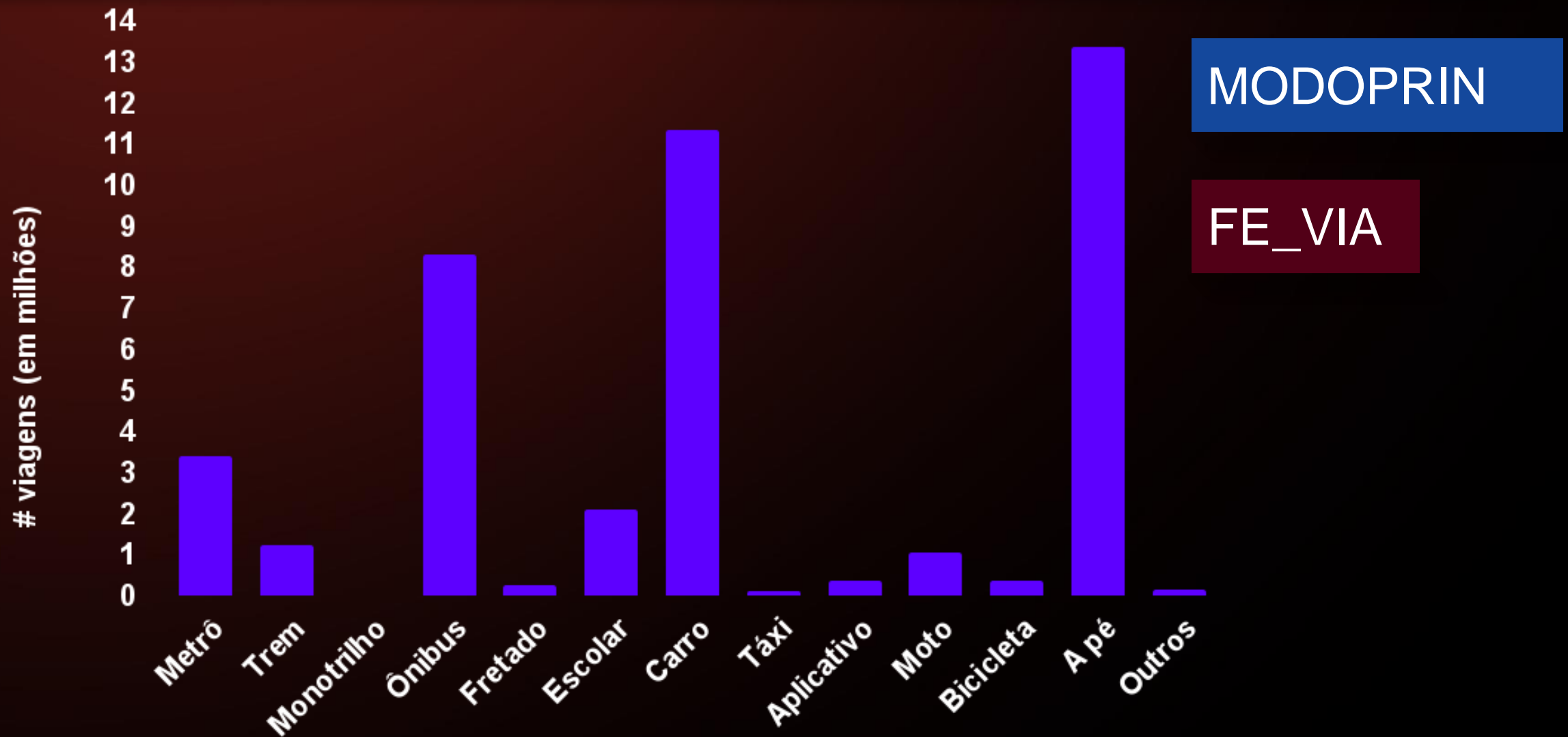
OD17

- **128 atributos:**
 - **Viagens:** origem, destino, motivo, horário, ...
 - **Modo:** a pé, carro, ônibus, ...
 - **Socioeconômicos:**
 - escolaridade, renda, bens, idade, ...
- **183.092 objetos (entrevistas)**
 - **42 milhões de viagens diárias**

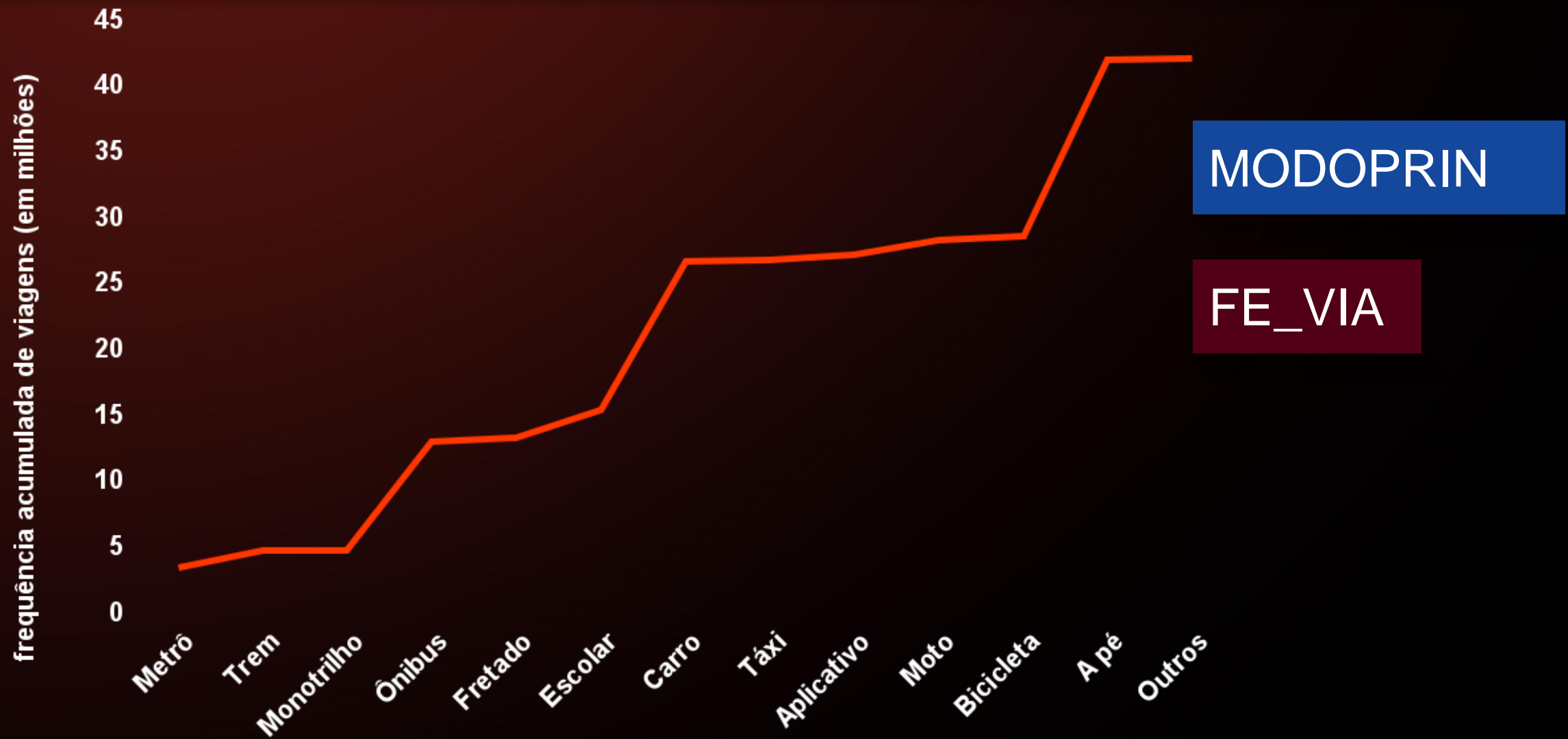
OD17

- **Agrupamentos existentes:**
 - **Modo de transporte**
 - **Motivo da viagem**
 - **Faixa de renda**
 - **Regiões administrativas**
 - **Cidade, distrito, ...**

OD 17 - VIAGENS POR MODO DE TRANSPORTE



OD 17 - VIAGENS POR MODO DE TRANSPORTE



OD 17 - DISTRIBUIÇÃO DA DURAÇÃO

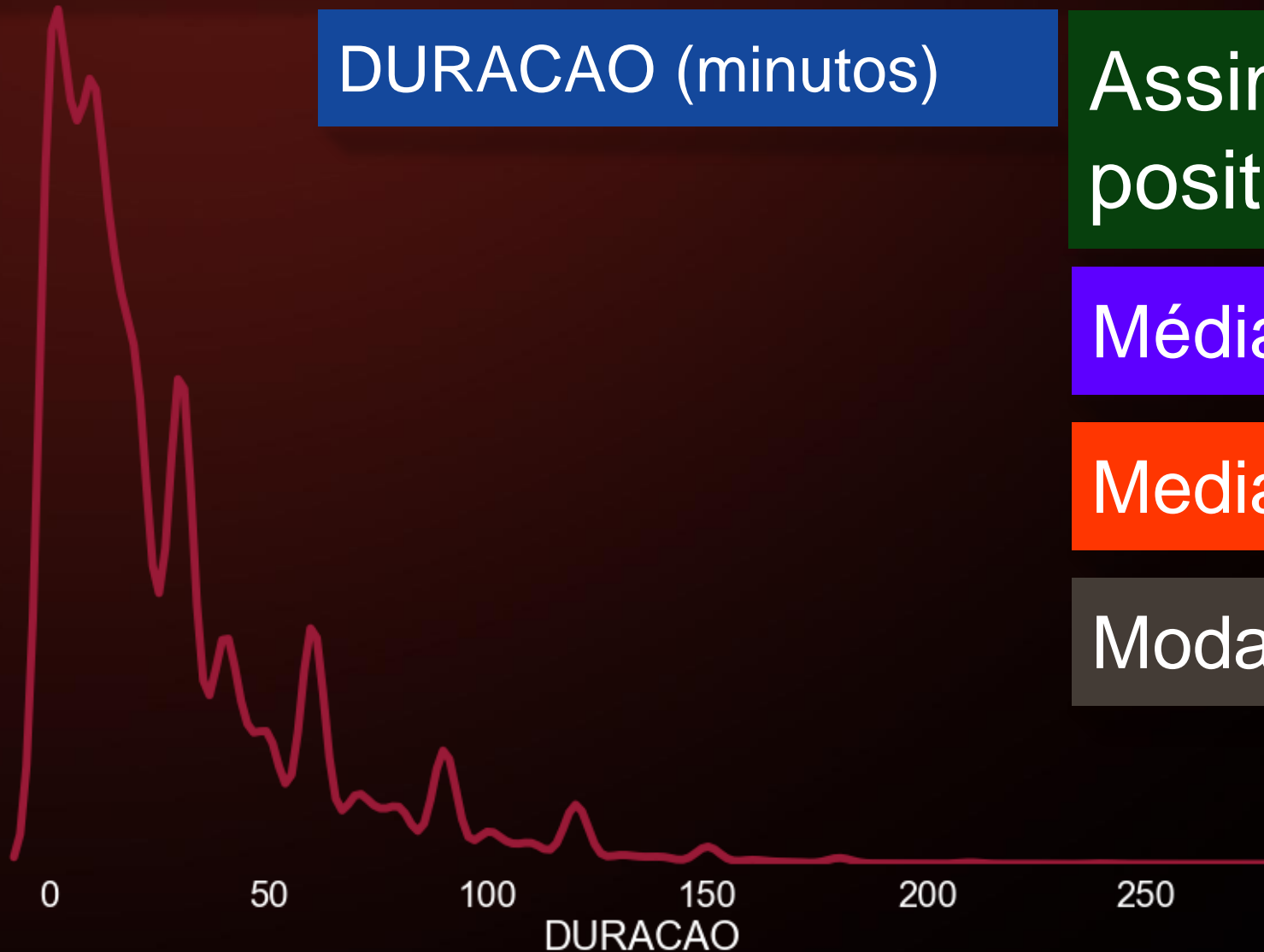
DURACAO (minutos)

Assimetria
positiva: ~1,91

Média: ~ 27

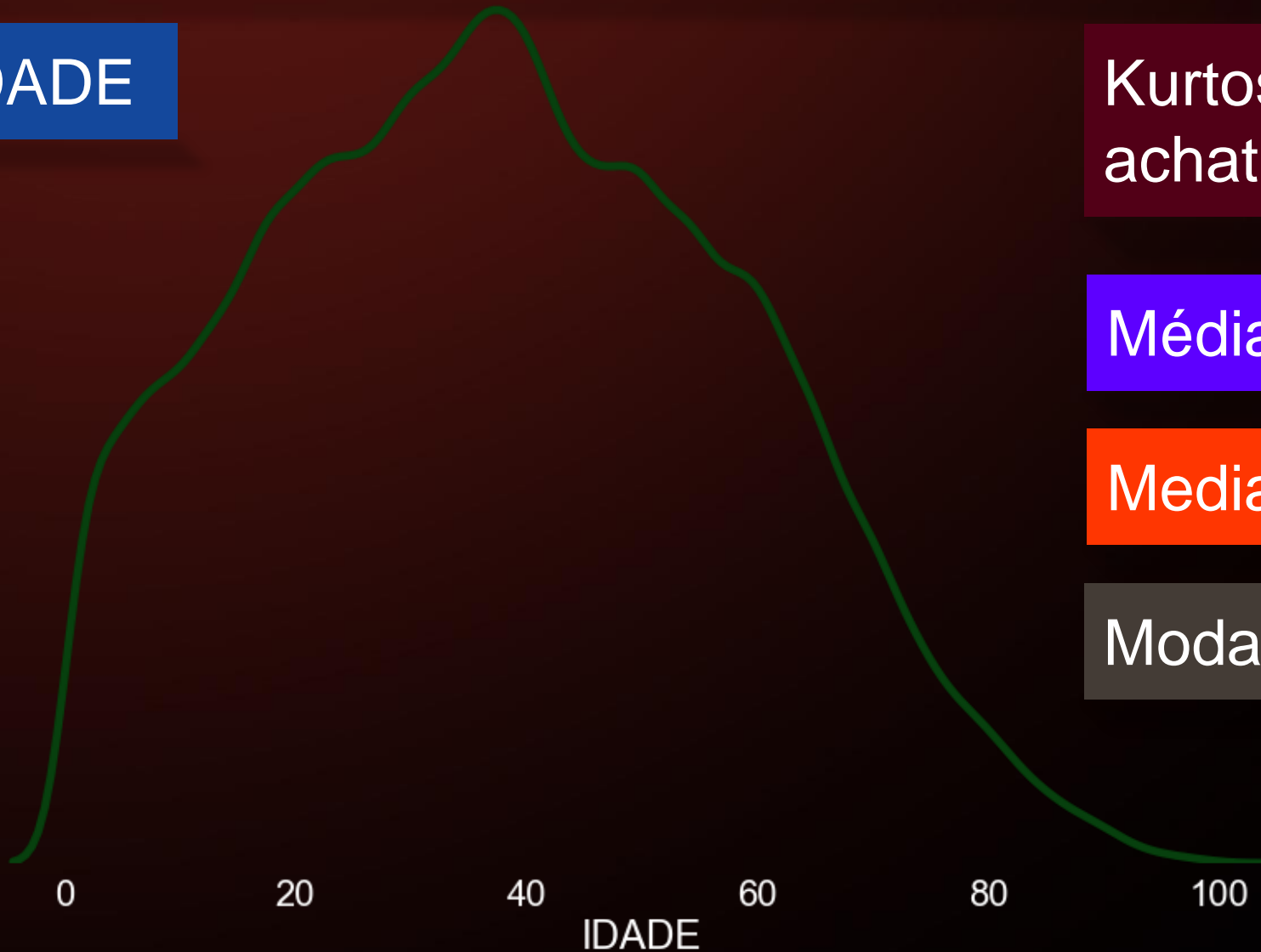
Mediana: 15

Moda: ~ 0



OD 17 - DISTRIBUIÇÃO DA IDADE

IDADE



Kurtose: ~ -0.69
achatada

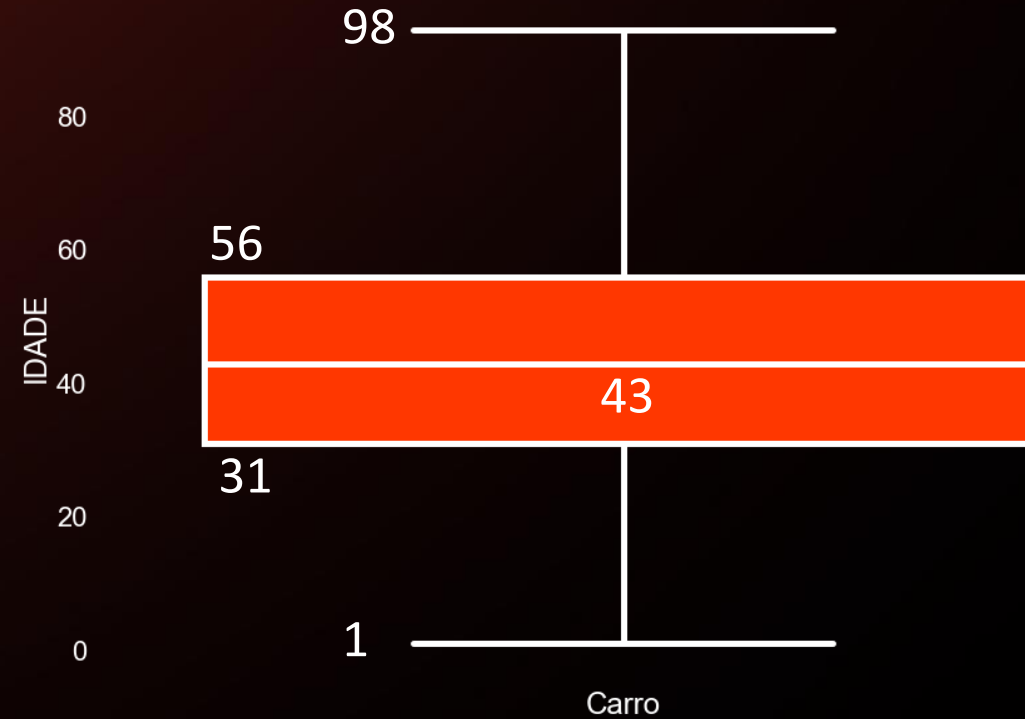
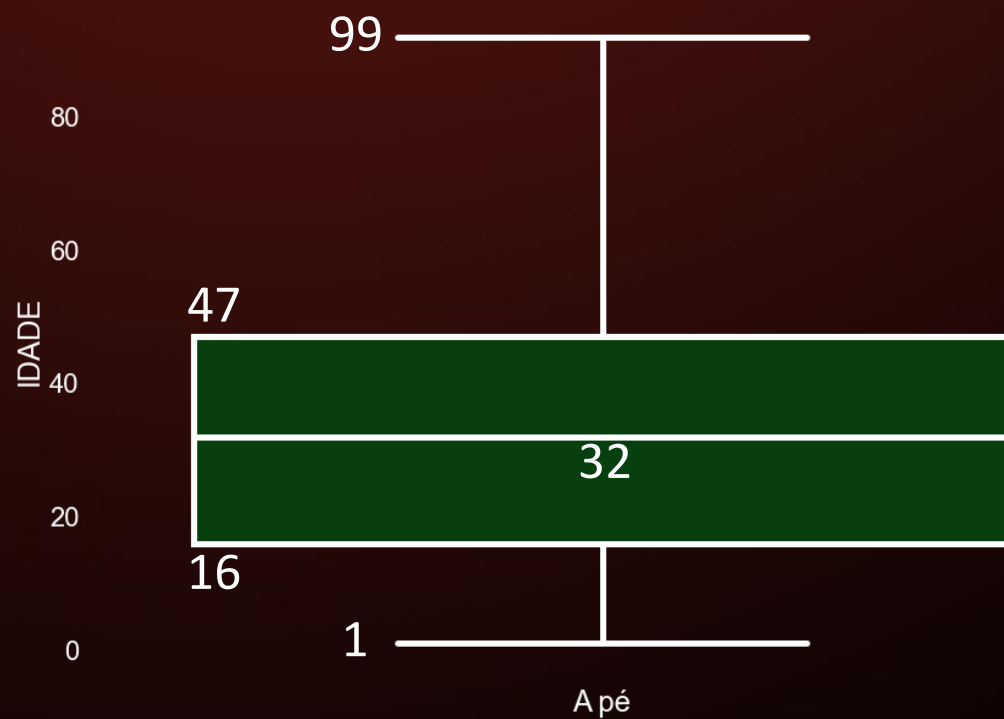
Média: ~ 38

Mediana: 37

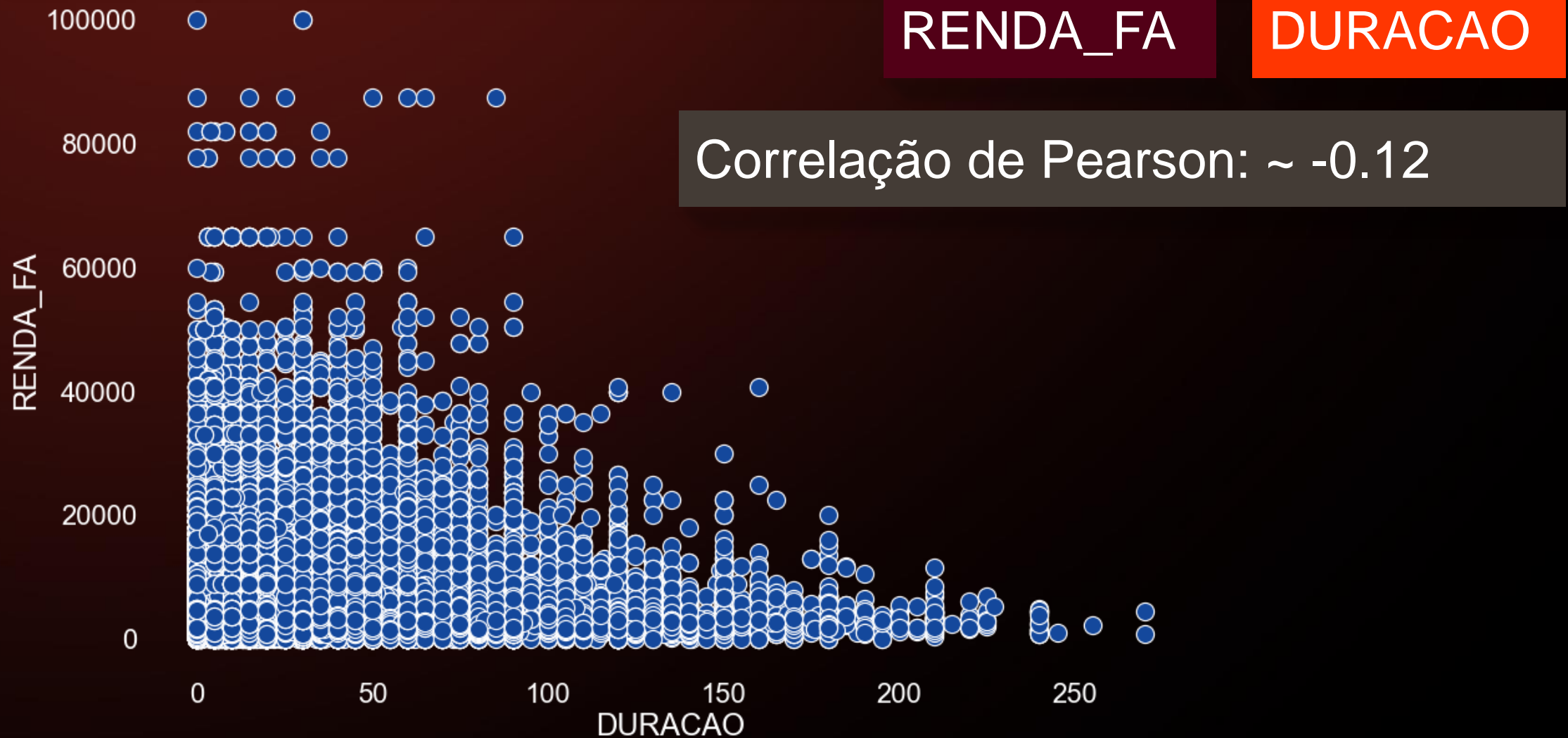
Moda: 40

OD 17 - POSIÇÃO RELATIVA DA IDADE

MODOPRIN: A pé X carro



OD 17 - CORRELAÇÃO: RENDA X DURAÇÃO



REFERÊNCIAS

Introdução à Mineração de Dados: Conceitos Básicos, Algoritmos e Aplicações: Cap. 3: Análise descritiva de dados.
Leandro Nunes de Castro e Daniel Gomes Ferrari. Editora Saraiva, 2016.

