

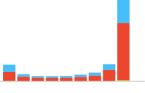






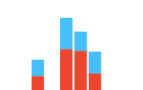














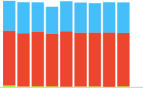
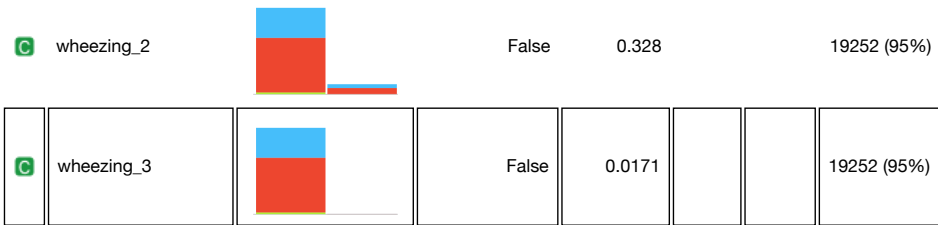


Name	Distribution	Center	Dispersion	Min.	Max.	Missing
<div><div>N</div><div>age</div></div>		34.5976	0.38752	1	102	9604 (47%)
<div><div>C</div><div>choking_1</div></div>		False	0.0171			19252 (95%)
<div><div>C</div><div>choking_2</div></div>		False	0.136			19252 (95%)
<div><div>C</div><div>choking_3</div></div>		False	0			19252 (95%)
<div><div>C</div><div>congestion_1</div></div>		False	0.0864			19252 (95%)
<div><div>C</div><div>congestion_2</div></div>		False	0.128			19252 (95%)
<div><div>C</div><div>congestion_3</div></div>		False	0.0242			19252 (95%)
<div><div>N</div><div>cough_detected</div></div>		0.641158	0.593992	0.00	0.9994	0 (0%)
<div><div>C</div><div>cough_type_1</div></div>		dry	0.921			19252 (95%)
<div><div>C</div><div>cough_type_2</div></div>		dry	0.76			19253 (95%)
<div><div>C</div><div>cough_type_3</div></div>		dry	1.04			19277 (96%)
<div><div>C</div><div>diagnosis_1</div></div>		COVID-19	1.26			19252 (95%)
<div><div>C</div><div>diagnosis_2</div></div>		COVID-19	1.51			19252 (95%)
<div><div>C</div><div>diagnosis_3</div></div>		upper_infection	1.19			19279 (96%)
<div><div>C</div><div>dyspnea_1</div></div>		False	0.0372			19252 (95%)
<div><div>C</div><div>dyspnea_2</div></div>		False	0.361			19252 (95%)
<div><div>C</div><div>dyspnea_3</div></div>		False	0.0171			19252 (95%)
<div><div>C</div><div>fever_muscle_pain</div></div>		False	0.391			8758 (43%)

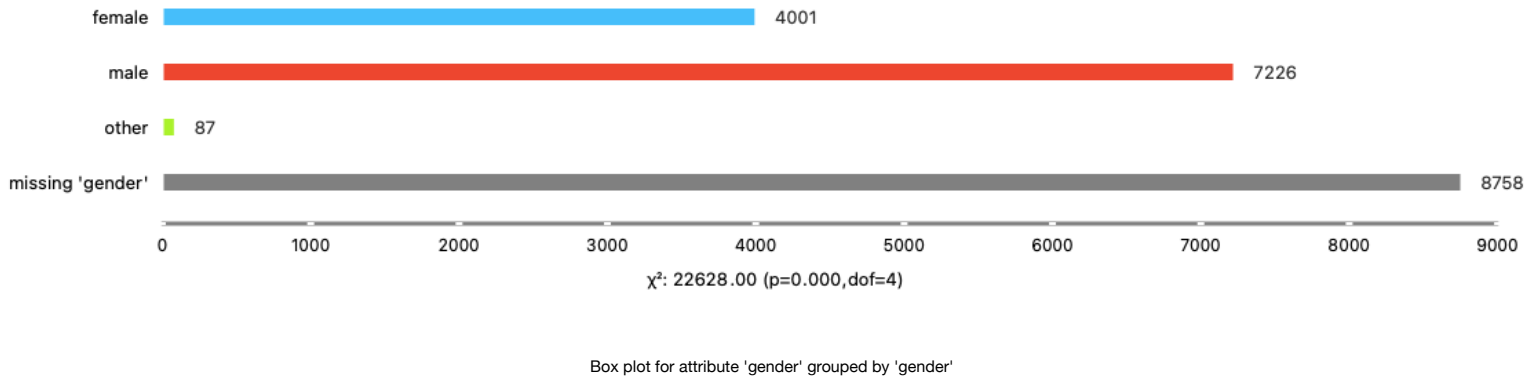
C	gender		male	0.691			8758 (43%)
N	latitude		30.1089	0.736108	-54.8	65.1	8326 (41%)
N	longitude		4.99349	10.924	-158.1	153.4	8326 (41%)
C	nothing_1		True	0.303			19252 (95%)
C	nothing_2		True	0.598			19252 (95%)
C	nothing_3		True	0.169			19252 (95%)
C	quality_1		ok	0.743			19252 (95%)
C	quality_2		good	1.04			19252 (95%)
C	quality_3		good	0.386			19252 (95%)
C	respiratory_condition		False	0.485			8758 (43%)
C	severity_1		mild	0.859			19252 (95%)
C	severity_2		mild	0.912			19252 (95%)
C	severity_3		mild	1.11			19276 (96%)
C	status		healthy	0.715			8758 (43%)
C	stridor_1		False	0			19252 (95%)
C	stridor_2		False	0.165			19252 (95%)
C	stridor_3		False	0.0094			19252 (95%)
N	Unnamed: 0		10035.50	0.58	0	20071	0 (0%)
C	wheezing_1		False	0.246			19252 (95%)



A base de dados COUGHVID possui mais de 20.000 áudios e 39 features ou variáveis. Parte dos áudios foram avaliados por três especialistas, o que resultou em classificações de condições de saúde, aparentemente diferentes entre eles, como podemos visualizar nos resultados da funcionalidade "Feature Statistics" para compor essa análise exploratória por gênero: azul (female), vermelho (male) e verde (other). Os "missing values" correspondem aos áudios não avaliados pelos especialistas e casos de não preenchimento.

Box Plot

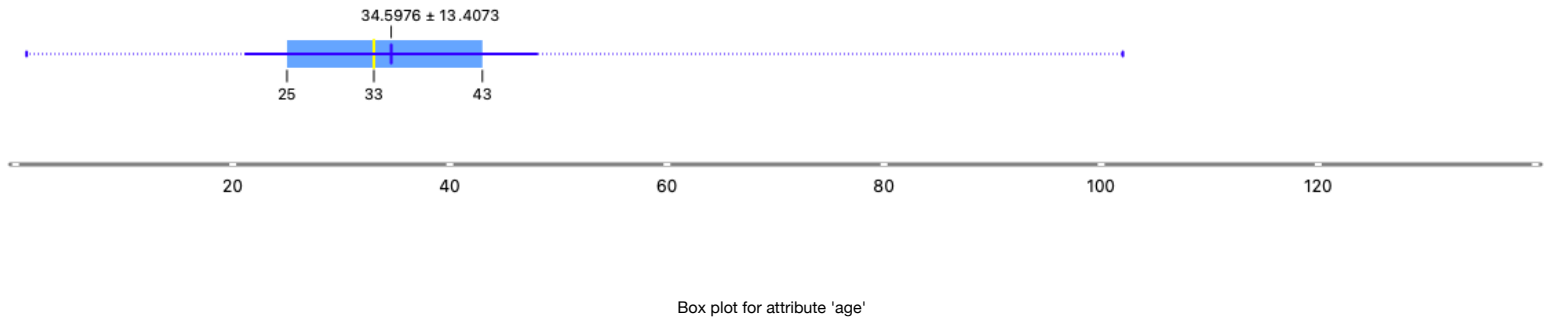
Sun Jun 20 21, 19:46:56



Com relação ao gênero, a maior frequência encontrada foi masculina, seguida de feminina e outros. Considerando as mais de 20.000 observações, em mais de 8700 não houve o preenchimento da variável gênero.

Box Plot

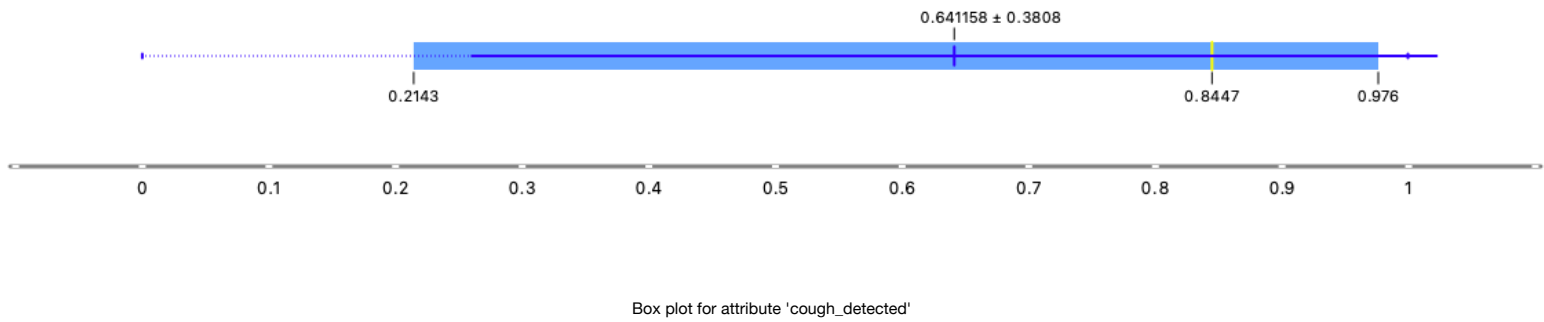
Sun Jun 20 21, 19:51:33



Foram observadas idade mínima de 1 ano e máxima de 102, com média de 34.6 anos de idade. A maior frequência do gênero masculino se manteve em todas as faixas etárias.

Box Plot

Sun Jun 20 21, 19:53:14

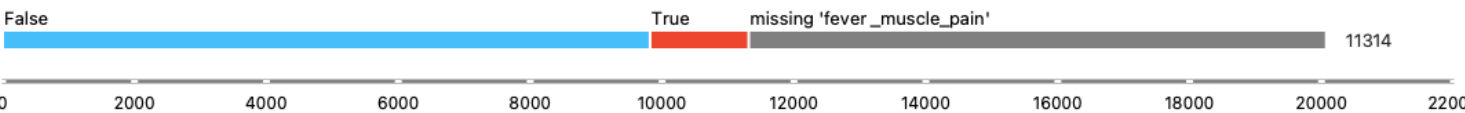


A variável "cough_detected" possui valor mínimo 0.00 e máximo 0.9994, e quanto mais próximo de 1 maior a chance do áudio ser de tosse, ou seja, áudio que permite a interpretação da tosse.



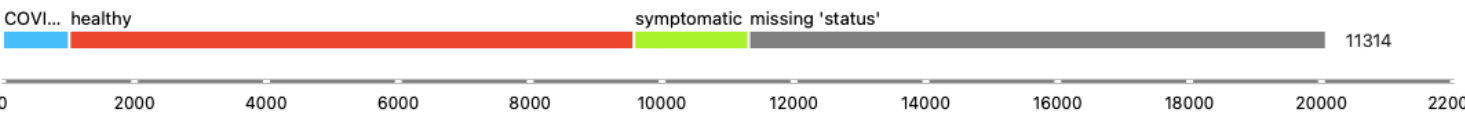
Box plot for attribute 'respiratory_condition'

Com relação à variável ter ou não uma condição respiratória, a maior frequência pode ser observada na opção "falsa".



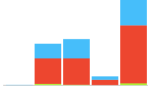
Box plot for attribute 'fever_muscle_pain'

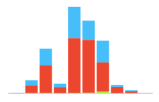

A alternativa "true" para febre ou dor muscular foi menos frequente quando comparada com "false".

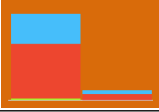
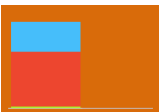


Box plot for attribute 'status'

Considerando o status, a maioria da população estudada estava saudável, e sintomáticos e com covid ficaram em segundo e terceiro lugar respectivamente.

Name	Distribution	Center	Dispersion	Min.	Max.	Missing
<div>N</div> age		34.5976	0.38752	1	102	9604 (47%)
<div>C</div> choking_1		False	0.0171			19252 (95%)
<div>C</div> choking_2		False	0.136			19252 (95%)
<div>C</div> choking_3		False	0			19252 (95%)
<div>C</div> congestion_1		False	0.0864			19252 (95%)
<div>C</div> congestion_2		False	0.128			19252 (95%)
<div>C</div> congestion_3		False	0.0242			19252 (95%)
<div>N</div> cough_detected		0.641158	0.593992	0.00	0.9994	0 (0%)
<div>C</div> cough_type_1		dry	0.921			19252 (95%)
<div>C</div> cough_type_2		dry	0.76			19253 (95%)
<div>C</div> cough_type_3		dry	1.04			19277 (96%)
<div>C</div> diagnosis_1		COVID-19	1.26			19252 (95%)
<div>C</div> diagnosis_2		COVID-19	1.51			19252 (95%)
<div>C</div> diagnosis_3		upper_infection	1.19			19279 (96%)
<div>C</div> dyspnea_1		False	0.0372			19252 (95%)
<div>C</div> dyspnea_2		False	0.361			19252 (95%)
<div>C</div> dyspnea_3		False	0.0171			19252 (95%)
<div>C</div> fever_muscle_pain		False	0.391			8758 (43%)
gender		male	0.691			8758 (43%)

<div>C</div>									
<div>N</div>	latitude		30.1089	0.736108	-54.8	65.1	8326 (41%)		
<div>N</div>	longitude		4.99349	10.924	-158.1	153.4	8326 (41%)		
<div>C</div>	nothing_1		True	0.303			19252 (95%)		
<div>C</div>	nothing_2		True	0.598			19252 (95%)		
<div>C</div>	nothing_3		True	0.169			19252 (95%)		
<div>C</div>	quality_1		ok	0.743			19252 (95%)		
<div>C</div>	quality_2		good	1.04			19252 (95%)		
<div>C</div>	quality_3		good	0.386			19252 (95%)		
<div>C</div>	respiratory_condition		False	0.485			8758 (43%)		
<div>C</div>	severity_1		mild	0.859			19252 (95%)		
<div>C</div>	severity_2		mild	0.912			19252 (95%)		
<div>C</div>	severity_3		mild	1.11			19276 (96%)		
<div>C</div>	status		healthy	0.715			8758 (43%)		
<div>C</div>	stridor_1		False	0			19252 (95%)		
<div>C</div>	stridor_2		False	0.165			19252 (95%)		
<div>C</div>	stridor_3		False	0.0094			19252 (95%)		
<div>N</div>	Unnamed: 0		10035.50	0.58	0	20071	0 (0%)		
<div>C</div>	wheezing_1		False	0.246			19252 (95%)		

C	wheezing_2		False	0.328			19252 (95%)
C	wheezing_3		False	0.0171			19252 (95%)

Considerando a avaliação dos áudios da tosse pelos três especialistas, podemos inferir que as variáveis "congestion"(congestão), "dyspnea" (dispneia ou falta de ar) e "wheezing" (respiração ofegante) se distribuem de maneira mais parecida. As demais variáveis pareceram ter classificações bastante diferentes entre os especialistas: "choking"(asfixia), "cough_type"(tipo de tosse), "diagnosis"(diagnóstico), "nothing", "quality", "severity"(severidade = desconhecida, pseudo tosse, média e severa).