

Universidade Estadual de Campinas  
Mineração de Dados Complexos - INF0612

## **Trabalho Final**

Grupo  
Nicole Nogueira Silva  
Rodolfo Dalla Costa

Campinas  
2021

# **Conteúdo**

<b>1</b>	<b>Introdução</b>	<b>2</b>
<b>2</b>	<b>Banco de dados</b>	<b>2</b>
2.1	Tratamento dos dados . . . . .	2
<b>3</b>	<b>Análise exploratória</b>	<b>3</b>
<b>4</b>	<b>Conclusão</b>	<b>8</b>

# 1 Introdução

O Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura (Cepagri) é um instituto de pesquisa que fornece informações meteorológicas da cidade de Campinas. Esse trabalho tem como objetivo extrair análises a partir dos dados provenientes do Cepagri e entender o comportamento do clima baseado nos indicadores climáticos medidos.

## 2 Banco de dados

O banco de dados utilizado para essa análise originou-se de consultas ao site do Cepagri ([www.cepagri.unicamp.br](http://www.cepagri.unicamp.br)) que a cada 10 minutos extraiu informações sobre a temperatura, velocidade do vento, umidade e sensação térmica. Na Tabela 2 pode-se observar os 5 primeiros registros da base. Originalmente a base contém 387.260 registros e 5 colunas.

Tabela 1: Primeiras linhas do banco de dados

Data	Temperatura	Vento	Umidade	Sensação Térmica
02/03/2014-19:08	23.7	59.3	77.1	22.6
02/03/2014-19:18	23.4	59.1	77.9	22.3
02/03/2014-19:28	23.2	56.7	78.9	22.1
02/03/2014-19:38	23.0	55.4	79.2	21.9
02/03/2014-19:48	22.8	52.6	79.7	21.7
02/03/2014-19:58	22.6	62.6	80.7	21.5

### 2.1 Tratamento dos dados

O tratamento dos dados foi realizado levando em consideração as orientações passadas em aula. Primeiramente foi necessário garantir que o comando de leitura do arquivo CSV fosse capaz de lidar com dados inesperados, como a palavra ERRO e a falta de dados nas colunas de algumas linhas. Para isso foi utilizado o argumento `fill` como `TRUE` da função `read.table()`.

Após isso, todos os dados que continham `NA` na coluna `Sensacao`, resultado da conversão das linhas com valores inadequados, foram removidos.

Logo em seguida, foi checado as classes de todas as colunas, constatou-se que a coluna `temperatura` estava com o tipo `char`, e deveria ser um tipo numérico, portanto foi realizada a conversão do tipo desta coluna para numérico para garantir a integridade dos dados no `dataframe`.

Depois disso, foi realizada a etapa de remoção de outliers. Primeiro foi avaliado por meio do `summary`, as estatísticas do `dataframe` conforme mostra a Tabela 2, com isso constatou-se que a coluna `sensacao` possuía valores fora do esperado como 99°. Assim, para garantir a integridade adequada do `dataframe`, foram removidos todas as linhas com `sensacao` acima de 60°.

Após essa etapa, foi então realizada a checagem de repetição de dados, para isso foi utilizado uma estratégia de checagem de repetição de linhas numa variação específica, por exemplo, para checar se existe repetição de dados num prazo de 10, nesse caso, a função percorre o `dataframe` observando se o elemento atual é igual aos 1440 elementos após ele. Percebeu-se que, em termos gerais, existe uma possível repetição de dados no equivalente a 10 dias.

Tabela 2: Estatísticas sumárias do banco de dados

data	temperatura	vento	umidade	sensacao	ano	mes
Min. :2015-01-01 00:00:00	Min. : 5.10	Min. : 0.00	Min. : 0.00	Min. :-8.00	Min. :2015	Min. : 1.000
1st Qu.:2016-05-29 12:22:30	1st Qu.:18.60	1st Qu.: 16.80	1st Qu.: 57.20	1st Qu.:16.70	1st Qu.:2016	1st Qu.: 3.000
Median :2017-10-27 15:15:00	Median :21.50	Median : 24.50	Median : 73.30	Median :20.10	Median :2017	Median : 6.000
Mean :2017-10-28 04:51:12	Mean :21.99	Mean : 28.43	Mean : 70.01	Mean :19.96	Mean :2017	Mean : 6.426
3rd Qu.:2019-03-20 18:37:30	3rd Qu.:25.50	3rd Qu.: 37.90	3rd Qu.: 84.80	3rd Qu.:24.10	3rd Qu.:2019	3rd Qu.:10.000
Max. :2020-12-31 23:50:00	Max. :37.40	Max. :143.60	Max. :100.00	Max. :34.30	Max. :2020	Max. :12.000
NA	NA	NA	NA	NA's :142	NA	NA

Além disso, percebeu-se que havia dados além dos que interessavam para a análise, desse modo, foi removido da base qualquer registro de anos menores que 2015 ou maiores que 2020. Após a fase inicial de tratamento dos dados, o dataframe usado para o relatório pode ser conferido na Tabela 3, onde estão as 5 primeiras linhas da base já tratada.

Tabela 3: Primeiras linhas do banco de dados tratado

	Data	Temperatura	Vento	Umidade	Sensação Térmica	ano	mes
42083	2015-01-01 00:00:00	22.7	22.2	92.2	21.6	2015	1
42085	2015-01-01 00:20:00	22.6	23.1	92.5	21.5	2015	1
42086	2015-01-01 00:30:00	22.7	23.1	92.0	21.6	2015	1
42087	2015-01-01 00:40:00	22.8	22.7	91.3	21.7	2015	1
42088	2015-01-01 00:50:00	22.7	22.3	91.6	21.6	2015	1
42089	2015-01-01 01:00:00	22.9	21.3	90.5	21.7	2015	1

### 3 Análise exploratória

Como apresentado anteriormente, a base de dados contém informações sobre a temperatura ao longo do tempo para a cidade de Campinas. Podemos começar a análise entendendo o comportamento das temperaturas médias de cada mês ao longo dos anos de 2015 a 2020. A Figura 1 apresenta o boxplot dessa medida.

Pode-se observar que o mês que possui temperaturas mais constantes ao longo dos anos é o mês de Dezembro já que apresentou menor variação. Enquanto isso, os meses de Janeiro e Junho costumam ter alta amplitude na temperatura média.

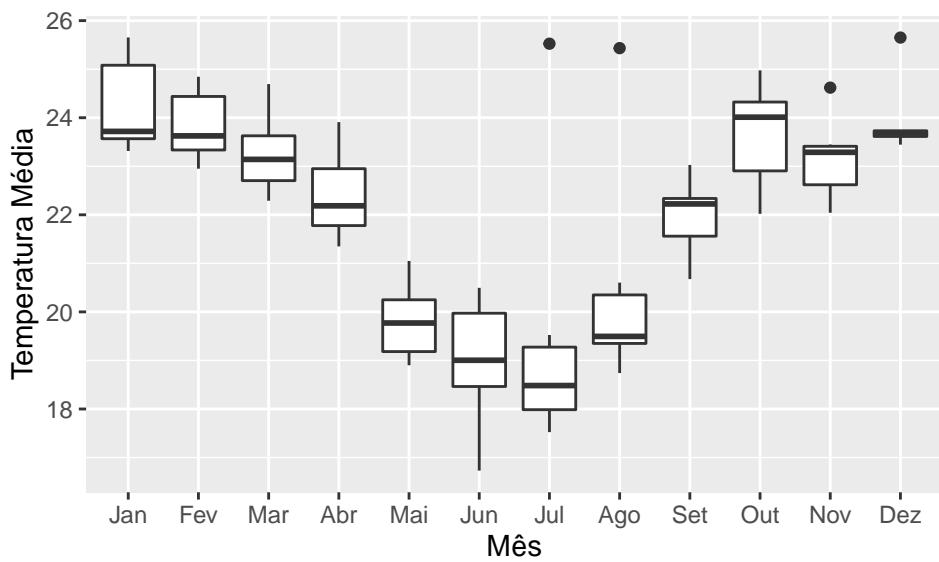


Figura 1: Boxplot da temperatura média de Campinas por mês.

Semelhante à Figura 1, podemos analisar o comportamento da Umidade do ar média em cada mês para todos os anos da base como mostra a Figura 2. Nota-se que os meses de Junho a Setembro, a cidade costuma ter menores médias de umidade relativa do ar, caracterizando o inverno seco comum ao bioma da região.

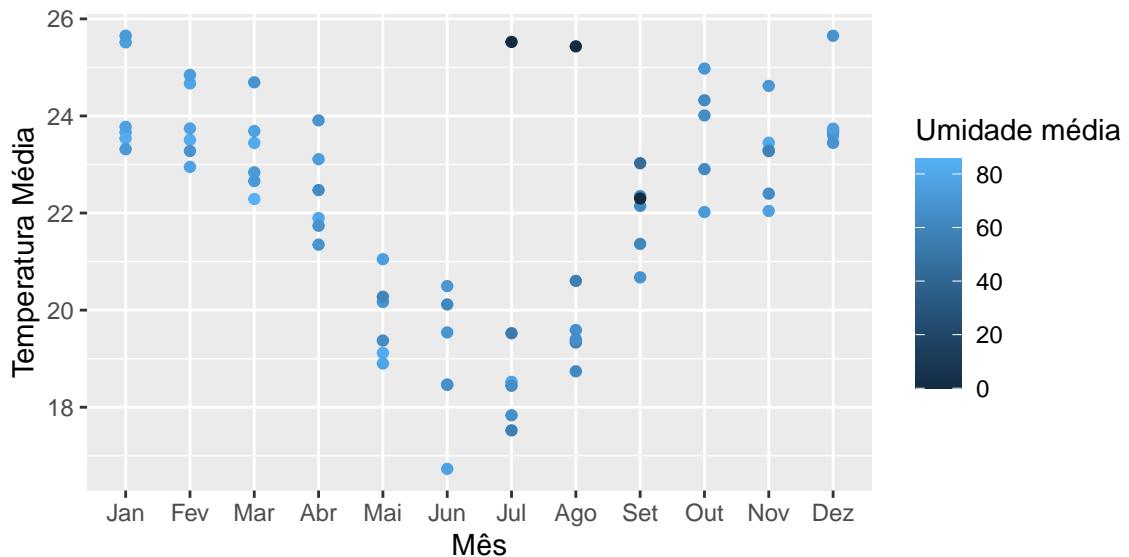


Figura 2: Temperatura e umidade média de Campinas por mês.

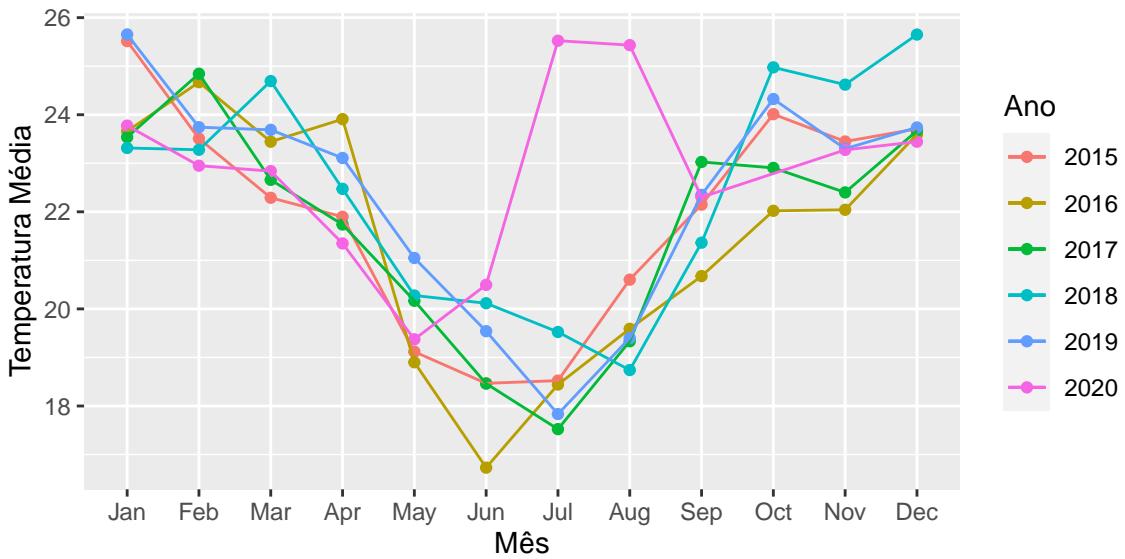


Figura 3: Média das temperaturas por ano

Para o gráfico da Figura 3 foi necessário elaborar um dataset filtrado que contém as médias de cada ano por mês, é possível ver o dataset na Tabela 4.

Tabela 4: Dataset com as temperaturas médias por mês de cada ano

Mes	Avg (Temperatura Media)	Ano
1	25.51657	2015
2	23.50825	2015
3	22.28965	2015
4	21.89768	2015
5	19.11822	2015
6	18.46698	2015

Por meio da Tabela 4 é possível observar as variações das temperaturas médias no decorrer dos meses nos anos de 2015 a 2020. É possível notar que o ano de 2020 possui um padrão muito divergente com relação aos outros anos. Em partes, isso pode ser devido ao fato de haver menos dados de 2020 que dos outros anos como pode se ver na Tabela 5:

Tabela 5: Primeiras linhas do banco de dados

2020	2019	2018	2017	2016	2015
31334	52276	51971	50915	52130	51212

No entanto, apesar da falta de dados que pode ser observada, o ano de 2020 foi marcado por um número de queimadas acima do comum para o período de estiagem, que contemplam os meses de julho e agosto, essa informação pode ser conferida em diversos sites assim como este.

Desse modo, percebe-se que seria interessante cruzar os dados de queimadas na região de Campinas com outros

dados metereológicos para identificar o perfil de influência do aumento de queimadas no clima. Existe claramente outros fatores meteorológicos que podem influenciar nessa variação constatada, assim como pode se tratar de erro do equipamento, ou falhas nas leituras que demoraram a ter correção, porém, a análise dos dados de 2020 não indica uma anormalidade relacionada aos instrumentos utilizados para a coleta da informação, como é possível observar na Tabela 6.

Tabela 6: Estatísticas da temperatura por ano.

ano	Min	Mediana	Media	Max
2015	9.8	21.3	21.91088	37.4
2016	5.1	21.1	21.44157	36.0
2017	7.6	21.1	21.65546	35.5
2018	6.8	22.0	22.42030	34.6
2019	6.1	21.8	22.29458	36.4
2020	10.0	21.8	22.36251	35.8

A temperatura para todo o ano de 2020 apresenta uma média geral dentro do comum, porém, percebe-se que as menores temperaturas são mais altas do que nos outros anos, o que indica que, de fato, o ano de 2020 registrou maiores temperaturas com maior frequência e não se trata de um erro técnico.

Outra análise interessante de ser realizada é a análise da sensação termica, dada a velocidade do vento e a temperatura apresentada na Figura 4

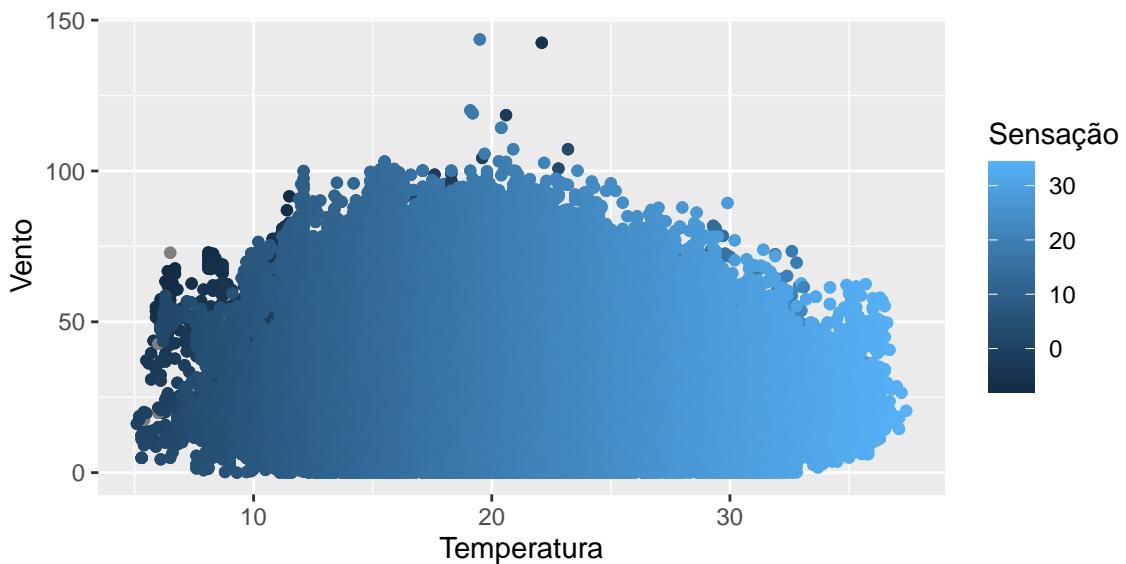


Figura 4: Temperatura conforme velocidade do vento e sensação térmica.

Como pode ser observado no gráfico da Figura 4 nota-se quanto mais fraco os ventos e maior a temperatura, maior é a sensação térmica, de certo modo, pode-se dizer que o vento influencia sim na sensação térmica, no entanto, para uma análise mais precisa e adequada, seria interessante ter o dado da temperatura do vento, claramente ventos muito fortes costumam ter temperaturas mais baixa, no entanto, uma massa de ar quente pode influenciar a temperatura a subir, o que poderia causar uma sensação térmica mais alta do que o esperado para ventos mais fortes.

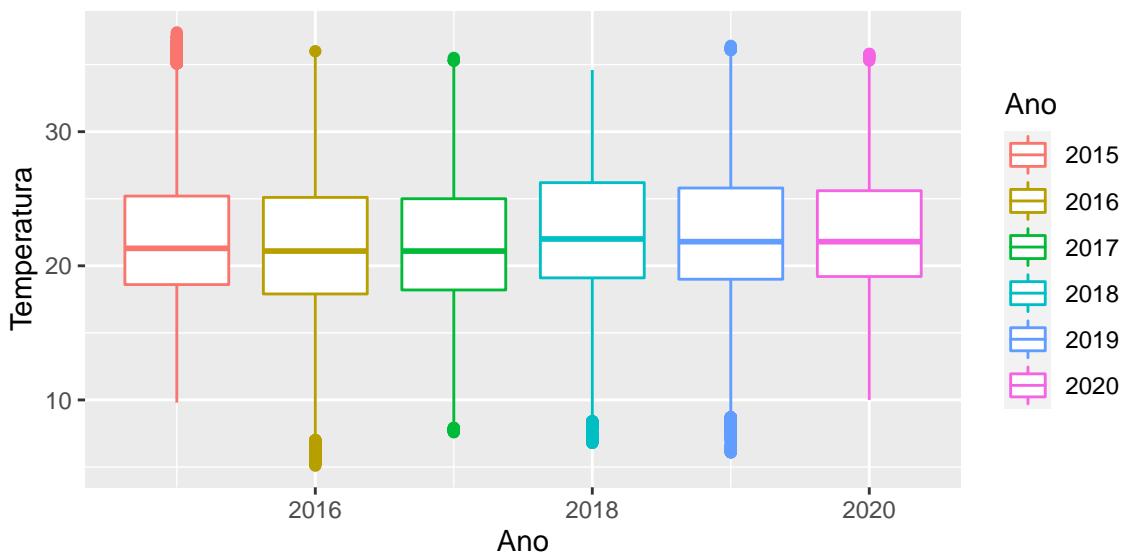


Figura 5: Variação da temperatura conforme o ano.

A observação da Figura 5 traz uma perspectiva muito interessante pois é possível perceber que de 2018 a 2020, houve um aumento na media de temperaturas por ano, nota-se que a variavao se manteve muito parecida em todos os anos, mas a media subiu notavelmente. Mesmo com mais outliers pequenos para o ano de 2019 e 2018, ainda sim a media se manteve mais alta.

## **4 Conclusão**

Por fim, após as análises da base, reforça-se que parece haver um aquecimento geral no clima da região monitorada. Logo, é fundamental obter dados qualitativos a respeito da região em específico para identificar se o aquecimento pode ser local ou devido a adversidades climáticas naturais. No entanto, a discussão sobre o aquecimento do clima de modo geral, realizada por especialistas, independente de região, não é um fato inédito e isolado e vem sendo destaque nos últimos anos dada a sua importância.