

# Projeto Análise de Dados

por Matheus Seiji, Ian de Almeida, Antônio Kotsubo

Análises sobre dados climáticos do CEPAGRI

## Introdução

O presente trabalho propõe-se a fazer uma análise sobre um grupo de dados captados pelo CEPAGRI, da Unicamp, sendo desenvolvido para a matéria de Tópicos inovadores em Computação, do Professor William Vicente, em 2021. Os dados foram disponibilizados pelo Professor Zanoni Dias e contam com temperatura, umidade, sensação térmica e vento registrados de dez em dez minutos, a partir do meio de 2014 até o presente.

Em um primeiro momento, delimitamos o período de tempo no qual faríamos as análises, do começo de 2015 até o final de 2020, para que tivéssemos todos os anos utilizados completos. Logo descartamos as extremidades da tabela de informações e, precedendo a análise própria e mais específica dos dados, excluimos também alguns valores inválidos. Nós olhamos para os valores da tabela como um todo, sumariamente, nos atentamos aos valores máximos e mínimos de cada coluna e observamos algumas incoerências, como o limite da sensação térmica a 99°C e alguns valores simplesmente não preenchidos, indicados como [Erro]. Portanto precisamos fazer uma filtragem na nossa tabela, a fim de eliminar discrepâncias gritantes.

```
##          data          sensa          temp          vento
## Min.      :2014-03-02   Min.      :-8.20   Min.      : 5.10   Min.      : 0.00
## 1st Qu.   :2016-01-21   1st Qu.   :16.40   1st Qu.   :18.60   1st Qu.   : 15.40
## Median    :2017-11-30   Median  :19.90   Median    :21.50   Median    : 23.40
## Mean      :2017-11-28   Mean     :19.76   Mean      :21.96   Mean      : 27.31
## 3rd Qu.   :2019-10-07   3rd Qu.  :23.90   3rd Qu.   :25.50   3rd Qu.   : 36.90
## Max.      :2021-08-20   Max.      :99.90   Max.      :38.10   Max.      :143.60
##          NA's          :30931   NA's       :30931   NA's       :30931
##          umid
## Min.      : 0.00
## 1st Qu.   : 56.40
## Median    : 72.80
## Mean      : 69.53
## 3rd Qu.   : 84.40
## Max.      :100.00
## NA's      :30931
```

No decorrer do projeto, construímos diversos gráficos e elaboramos determinadas justificativas para as correlações e discrepâncias nele encontradas. Para isso, decidimos tomar como parâmetro, para a escolha das nossas primeiras investigações, a correlação entre os dados de cada coluna, reproduzida abaixo.

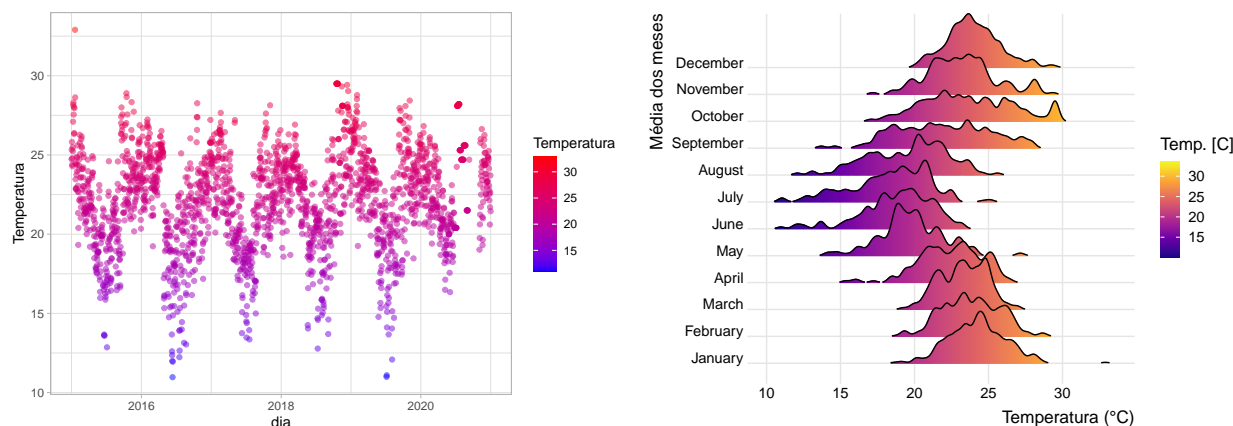
```
##          data          sensa          temp          vento          umid
## data      1.000000000    0.1522092    0.04973669   -0.002609083   -0.14009070
## sensa     0.152209226    1.0000000    0.93279024   -0.219528938   -0.49903408
## temp      0.049736686    0.9327902    1.000000000   -0.161875043   -0.59723515
## vento     -0.002609083   -0.2195289   -0.16187504    1.000000000    0.06168511
## umid      -0.140090701   -0.4990341   -0.59723515    0.061685106    1.000000000
```

Nela é possível perceber a intensa e evidente relação entre sensação térmica e temperatura. Ademais, encontramos algumas correlações menores, todavia notáveis, como as de umidade e vento com temperatura,

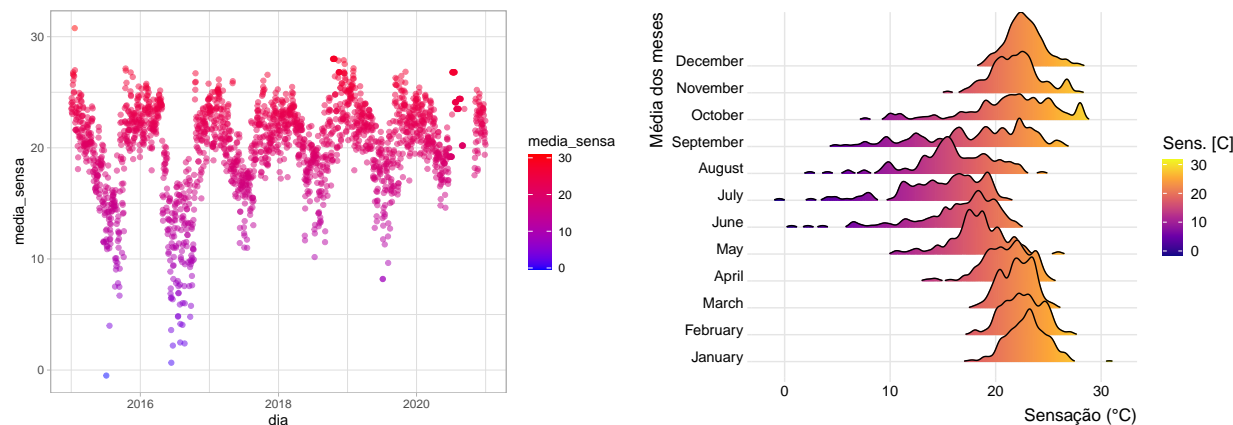
deixando de lado correlações quase insignificantes, como a entre umidade e vento.

## Análises iniciais

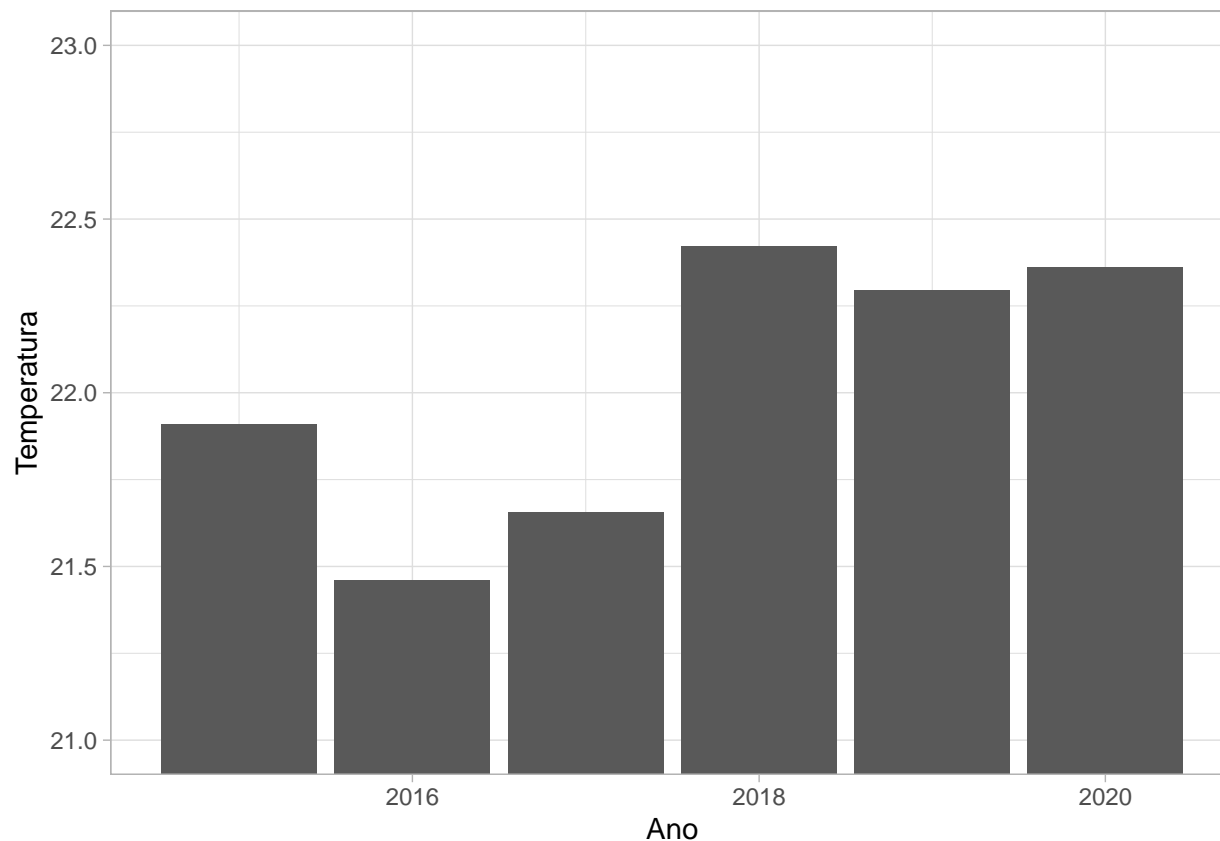
**Temperatura x Tempo e Temperatura x Sensação Térmica** Iniciando nossas análises, realizamos o simples, porém muito concreto, relacionamento entre a temperatura e o tempo (utilizamos as médias de temperatura a cada dia do período analisado, com o objetivo de reduzir a quantidade, estas sem apresentar variações representativas, dos dados analisados). Desta forma, esperava-se que o resultado representasse a variação normal de temperatura ao longo do ano e das estações. Isso de fato ocorreu, como mostrado no gráfico abaixo:



De maneira similar ao realizado privamente, montamos um gráfico que relaciona a sensação térmica média ao longo dos dias dos anos em questão. Obtivemos, então, resultados que, não só refletem a curva de temperatura observada previamente, como também acentua suas variações:



**Aumento gradual da temperatura anual** Uma vez que fizemos as duas análises já apresentadas, notamos um sutil aumento na temperatura média com o decorrer do tempo, portanto decidimos verificar essa observação produzindo um gráfico de barras com a média anual da temperatura dos últimos anos.



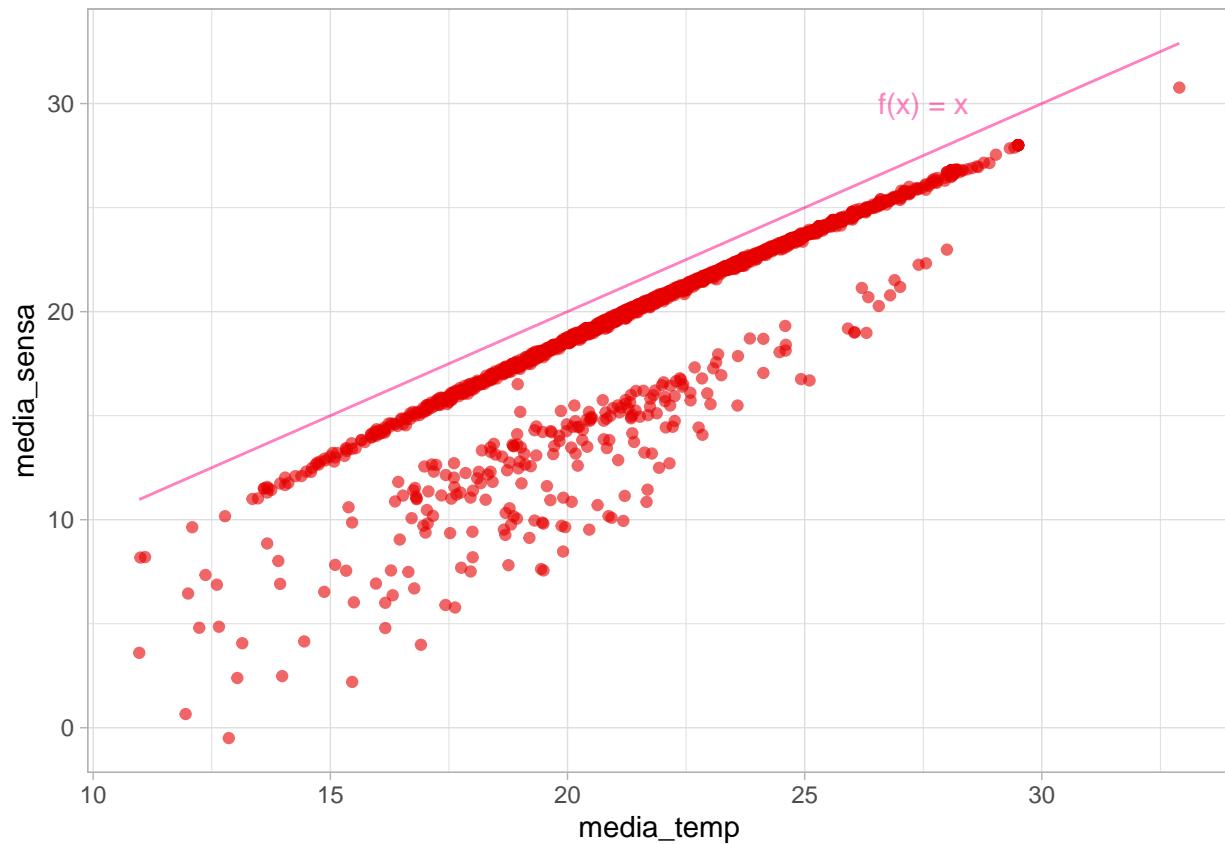
```
##      [,1]  [,2]    [,3]    [,4]    [,5]    [,6]    [,7]    [,8]
## [1,] ""    "January" "February" "March" "April" "May"    "June" "July"
## [2,] "2016" "23.66"   "24.67"   "23.44" "23.91" "18.91" "16.76" "18.47"
## [3,] "2017" "23.54"   "24.84"   "22.66" "21.74" "20.17" "18.46" "17.52"
## [4,] "2018" "23.32"   "23.28"   "24.69" "22.47" "20.28" "20.12" "19.52"
## [5,] "2019" "25.65"   "23.74"   "23.69" "23.11" "21.05" "19.54" "17.83"
## [6,] "2020" "23.78"   "22.95"   "22.84" "21.35" "19.37" "20.5"  "25.52"
##      [,9]    [,10]    [,11]    [,12]    [,13]
## [1,] "August" "September" "October" "November" "December"
## [2,] "19.63"  "20.71"    "22.04"  "22.04"    "23.59"
## [3,] "19.33"  "23.03"    "22.9"   "22.4"     "23.67"
## [4,] "18.74"  "21.37"    "24.98"  "24.62"    "25.65"
## [5,] "19.4"   "22.35"    "24.32"  "23.3"     "23.74"
## [6,] "25.43"  "22.3"     NA       "23.28"    "23.44"
```

```
##      Estação Temperatura anterior Temperatura atual Variação
## 1  2015 - inverno      23.75604      19.93557 -3.820473
## 2  2015 - verão       19.93557      23.84246  3.906886
## 3  2016 - inverno      23.84246      18.89208 -4.950377
## 4  2016 - verão       18.89208      23.67734  4.785257
## 5  2017 - inverno      23.67734      19.58639 -4.090948
## 6  2017 - verão       19.58639      24.23444  4.648047
## 7  2018 - inverno      24.23444      19.93826 -4.296178
## 8  2018 - verão       19.93826      24.20579  4.267533
## 9  2019 - inverno      24.20579      19.78088 -4.424911
## 10 2019 - verão       19.78088      23.25214  3.471263
## 11 2020 - inverno      23.25214      NA        NA
```

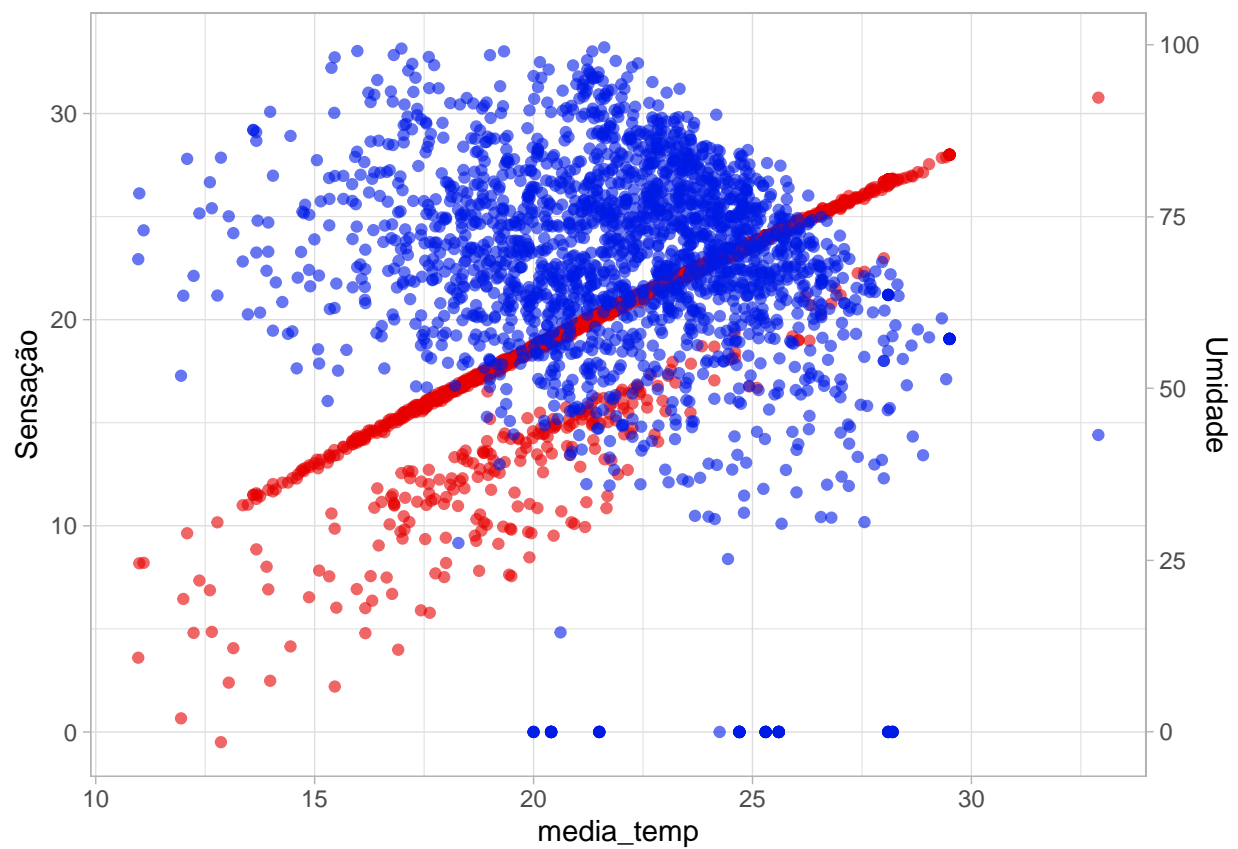
Através disso, foi possível verificar um crescimento considerável na temperatura média anual para o período objeto da pesquisa. Dissidentemente, 2015 foi sutilmente mais quente do que 2016, entretanto, ao ver o gráfico mais amplamente, fica claro que enfrentamos um fenômeno climático que pode estar associado com o aquecimento global.

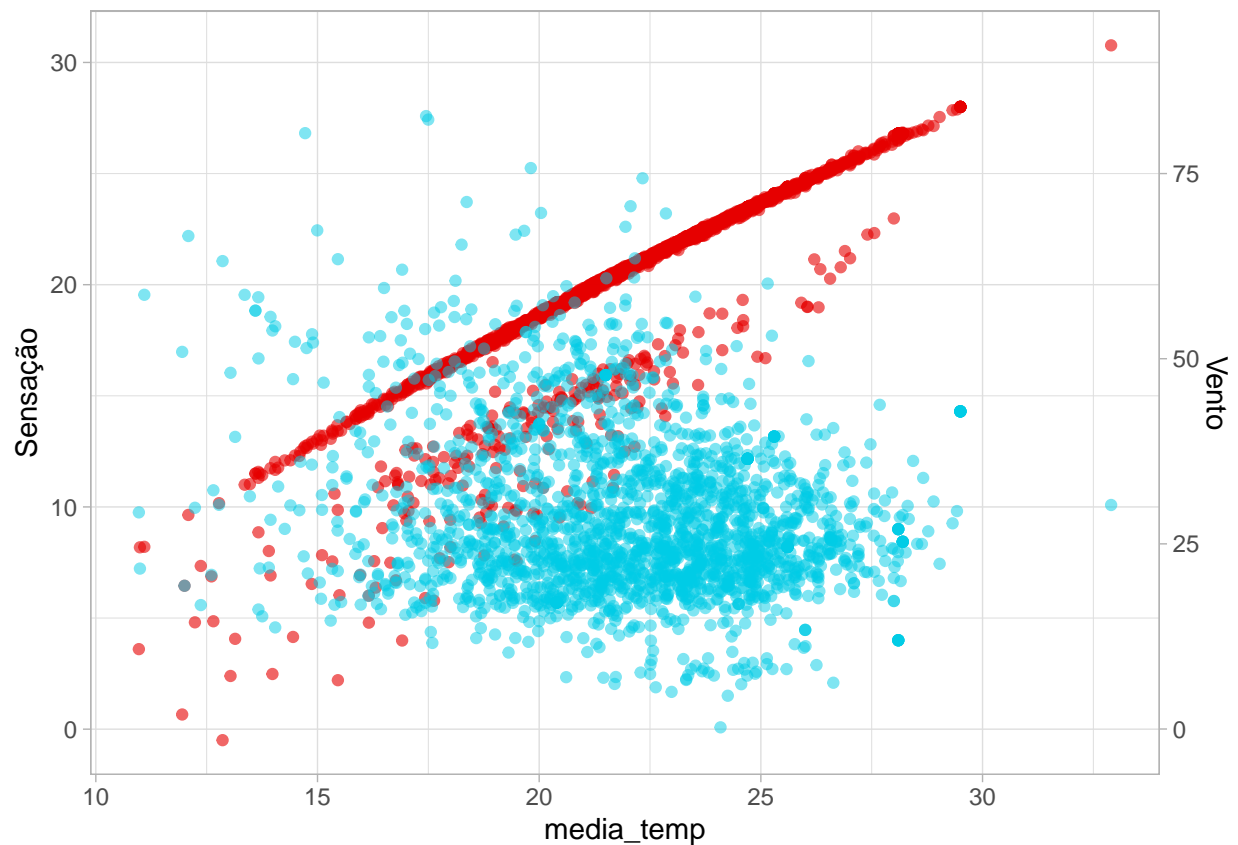
### Influência de umidade e vento na relação Temperatura e Sensação Térmica

**O gráfico Temperatura x Sensação Térmica** Através do gráfico da relação, pode-se observar uma forte correlação entre a sensação térmica e a temperatura, algo já esperado, o que forma uma grande concentração de medições próximas à reta  $f(x) = x$ , porém sempre abaixo dela. No entanto, existem alguns pontos dissidentes do curso principal, o que implica em uma influência de algum fator externo à temperatura nos índices da sensação térmica, que ocasionalmente puxa-a bem abaixo do desvio esperado.

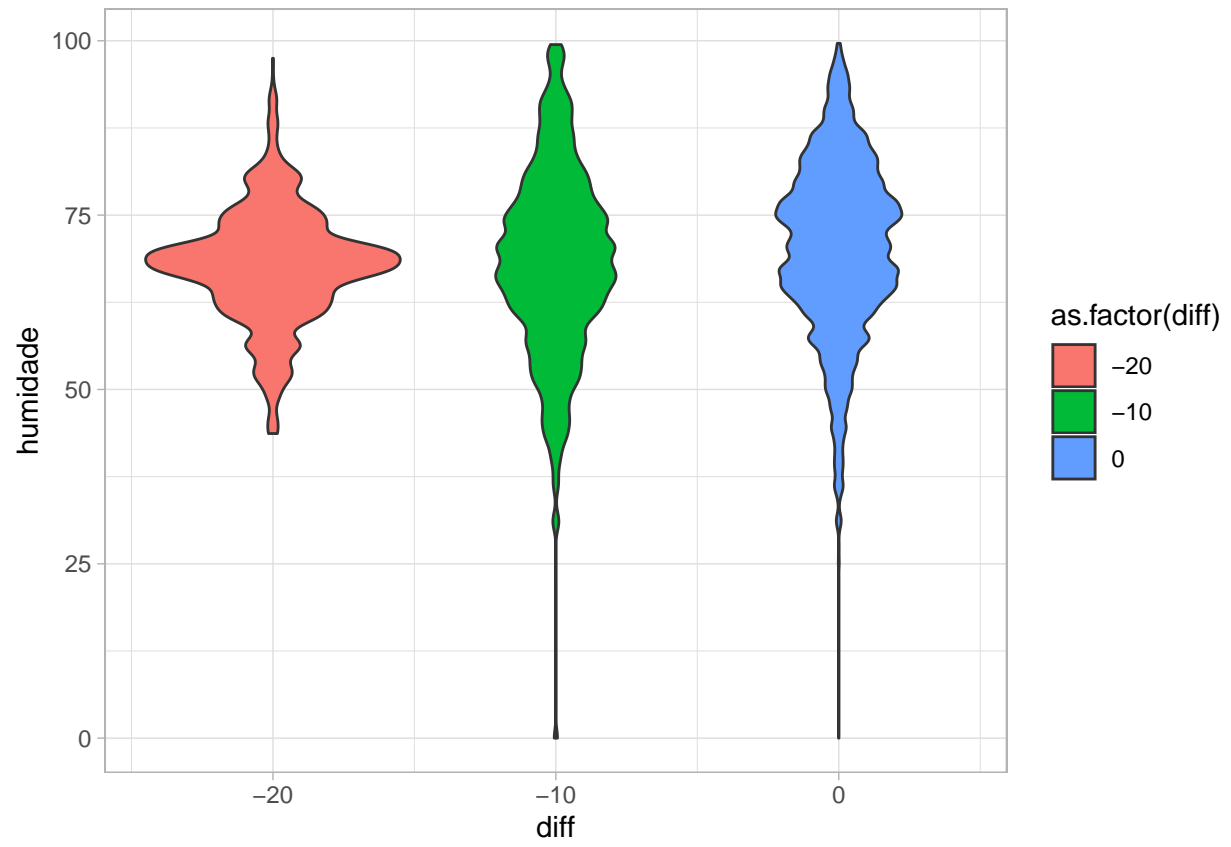


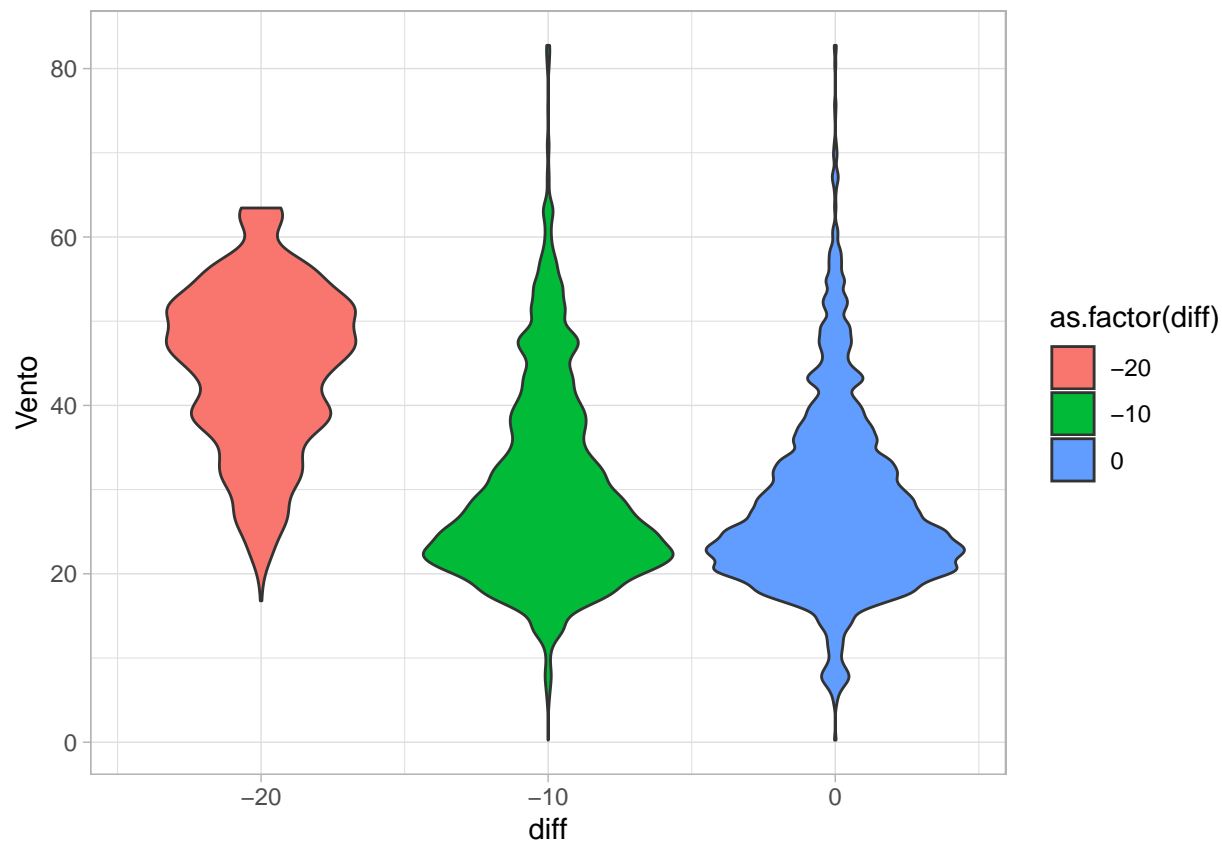
**A influência do Vento e da Umidade** Para buscar explicação aos pontos que se desviam da relação direta entre temperatura e sensação térmica, decidimos gerar outros quatro gráficos que trouxessem à tona dois fatores que poderiam ter relação com a dissidência encontrada: vento e umidade. Ambos tinham uma correlação mediana com a sensação térmica, mas talvez suficiente para, em algum grau, provocar sua variação.





Feitos os primeiros gráficos, obtivemos resultados positivos. Foi possível observar primeiramente que a região de sensação mais próxima a linha padrão, na qual há uma dissidência mais densa, a umidade é consistentemente mais elevada, indicando que ela pode influir mais intensamente nos resultados (umidades mais elevadas também elevam a temperatura). Algo semelhante ocorre com o vento, cujos valores consistentes encontram-se na mesma região de dissidentes, porém exercendo papel contrário à umidade, baixando a temperatura.





Estes dois novos gráficos de “violino” apresentam a dispersão das intensidades desses fatores (humidade e vento) pela diferença observada entre sensação térmica e temperatura. Desta forma, temos outra maneira de enxergar a correlação observada entre as quatro variáveis, com grandes variações sendo comuns a menos faixas de humidade, e com ventos fortes influenciando, por sua vez, na redução da sensação térmica.

Em suma, podemos inferir que a influência de vento e temperatura na sensação térmica, mesmo que obviamente inferior a de temperatura, não pode ser descartada, pois, em sua própria intensidade, esta proporciona desvios muitas vezes consideráveis do esperado dada apenas a temperatura. Logo estes fatores são necessários para averiguar-se mais detalhadamente a sensação térmica de um determinado momento.

```
## # A tibble: 71 x 4
## # Groups:   temperatura [71]
##   temperatura vento umidade sensacao
##   <dbl> <dbl>   <dbl>   <dbl>
## 1     25.5  26.2    74.6    24.2
## 2     23.5  24.4    82.7    22.3
## 3     22.3  23.8    85.8    21.1
## 4     21.9  32.0    79.5    20.6
## 5     19.1  26.7    82.5    17.7
## 6     18.5  31.0    77.4    16.9
## 7     18.5  24.7    78.7    12.7
## 8     20.6  26.5    55.4    14.2
## 9     22.1  31.1    61.4    15.0
## 10    24.0  36.2    64.7    20.9
## # ... with 61 more rows
```