

# Relatório parcial

Arthur Augusto Campos

Instituto Federal de Brasília - Estrutural

13 de junho de 2025

# Sumário

- 1 Ataque ao problema
- 2 Limpeza dos dados
  - Dados Iniciais
  - Dados da radiação
- 3 Primeiras análises
  - Resumo estadual
  - Visualizando NA's
  - Analisando boxplots por estado

# Estratégia

A estratégia de abordagem do problema foi dividida em quatro etapas:

- Limpeza dos dados
- Análises estatísticas básicas e de NA
- Reconstrução dos dados faltantes (se necessário)
  - por hora, penso que se essa etapa for necessário, pode ser interessante fazer um bootstrap das amostras horárias em outros anos com dados para extrair uma boa estatística para imputar nos dados faltantes
- Aplicação da transformada de Fourier para realizar a Análise Harmônica

# Relembrando os dados iniciais

- Pressão atmosférica
  - nível da estação horária; max na h ant; min na h ant. (mbar)
- Precipitação
  - Total horário. (mm)
- Radiação
  - global.  $\left(\frac{kJ}{m^2}\right)$
- Temperatura
  - do ar bulbo seco horaria; ponto de orvalho; max na h ant; min na h ant; orvalho max na h ant; orvalho min na h ant. ( $^{\circ}C$ )
- Umidade
  - relativa max na h ant, relativa min na h ant, relativa horária. (percentual)
- Vento
  - dir. horária (gr); rajada max; vel horária max.  $\left(\frac{m}{s}\right)$

# Radiação

```
> glimpse(dados_radiacao)
Rows: 1,077,096
Columns: 6
$ data_tempo      <dtm> 2001-01-01 00:00:00, 2001-01-01 01:00:00, 2001-01-01
  02:00:00, 2001-01-01 ...03:
$ data            <date> 2001-01-01, 2001-01-01, 2001-01-01, 2001-01-01,
  2001-01-01, 2001-01-01, ...2001-
$ hora_utc        <time> 00:00:00, 01:00:00, 02:00:00, 03:00:00, 04:00:00,
  05:00:00, 06:00:00, ...07:00:0
$ estado          <chr> "AM", "AM", "AM", "AM", "AM", "AM", "AM", "AM", "AM", "
  AM", "AM", "AM", "AM", ...
$ arquivo_origem  <chr> "INMET_N_AM_A101_MANAUS_01-01-2001_A_31-12-2001.CSV", "
  INMET_N_AM_A101_MANAUS_..."
$ radiacao_global_kj_m2 <chr> "-9999", "-9999", "-9999", "-9999", "-9999", "-9999", "
  -9999", "-9999", "...-9999
```

# Radiação limpo

```
> glimpse(dados_radiacao)
Rows: 1,077,096
Columns: 6
$ data_tempo      <dtm> 2001-01-01 00:00:00, 2001-01-01 01:00:00, 2001-01-01
  02:00:00, 2001-01-01 03:00:00, ...
$ data            <date> 2001-01-01, 2001-01-01, 2001-01-01, 2001-01-01, 2001-01-01,
  2001-01-01, 2001-01-01, ...
$ hora_utc        <time> 00:00:00, 01:00:00, 02:00:00, 03:00:00, 04:00:00, 05:00:00,
  06:00:00, 07:00:00, ...08:0
$ estado          <chr> "AM", "AM", "AM", "AM", "AM", "AM", "AM", "AM", "AM", "AM", "
  AM", "AM", "AM", "AM", ...
$ arquivo_origem  <chr> "INMET_N_AM_A101_MANAUS_01-01-2001_A_31-12-2001.CSV", "INMET_N
  _AM_A101_MANAUS_...01-01-2"
$ radiacao_kjm2   <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA
  , NA, NA, NA, NA, NA, ...N
```

# Considerações sobre o ruído

- Estamos analisando dados de estações de coleta automática em cinco estados diferentes ao longo de aproximadamente 25 anos

# Considerações sobre o ruído

- Estamos analisando dados de estações de coleta automática em cinco estados diferentes ao longo de aproximadamente 25 anos
- Acontecem falhas e a manutenção nem sempre é feita como o desejável



# Considerações sobre o ruído

- Estamos analisando dados de estações de coleta automática em cinco estados diferentes ao longo de aproximadamente 25 anos
- Acontecem falhas e a manutenção nem sempre é feita como o desejável
- Algumas informações a seguir são influenciadas pelo ruído, mas entendo que seja possível tratá-lo.

# Considerações sobre o ruído

- Estamos analisando dados de estações de coleta automática em cinco estados diferentes ao longo de aproximadamente 25 anos
- Acontecem falhas e a manutenção nem sempre é feita como o desejável
- Algumas informações a seguir são influenciadas pelo ruído, mas entendo que seja possível tratá-lo.
- Vamos a primeira análise: Após verificação dos índices de radiação máxima, mínima e a mediana, foi possível constatar:

# AM - Manaus

Hora	Rad. Máx	Rad. Mín	Rad. Mediana
0	490	0	0
1	443	0	0
2	480	0	0
3	531	0	0
4	527	0	0
5	538	0	0
6	505	0	0
7	538	0	0
8	537	0	0
9	518	0	0
10	514	0	1.8
11	744	0	151
12	2909	0	715
13	4286	0	1404
14	4637	0	1994
15	3922	0	2311
16	6485	0	2435
17	6455	1	2430
18	4098	1	2171
19	5059	1	1747
20	4820	0	1082
21	4389	0	488
22	1164	0	112
23	560	0	0.6

# Análise breve Manaus

- A priori, é visível o ruído com observações de radiação máxima diferentes de zero em períodos de madrugada
- Contudo, é possível perceber, com a mediana, o padrão do nascer e do por do sol entre 6h e 18h. (Lembrando que as horas estão no padrão: Tempo Universal Coordenado - UTC)
- O maior registro de máxima acontece ao meio dia, o que também parece ser um bom sinal.

# BA - Salvador

Hora	Rad. Máx	Rad. Mín	Rad. Mediana
0	4205	0	2
1	4480	0	2.8
2	4401	0	2.7
3	4392	0	1.85
4	3984	0	2
5	4214	0	2
6	4405	0	2
7	4227	0	2
8	4340	0	1.2
9	4305	0	21.1
10	4114	0	211
11	3641	0.1	821
12	4106	0	1541
13	3945	3.4	2075
14	4013	1.2	2541
15	4202	1.1	2837
16	4082	4.2	2835
17	4446	1.4	2555
18	4316	0.2	2039
19	3921	0.8	1338
20	4365	0	465
21	4295	0	29
22	4266	0	1.5
23	4000	0	2.95

# Análise breve Salvador

- Aqui temos um quadro um pouco mais preocupante, com análises de máxima relativamente próximas ao longo de todas as horas do dia, indicando claramente um ruído.
- Além disso, podemos perceber medianas que, embora próximas de zero, não se anulam na madrugada. Isso nos leva a perceber que existe uma boa quantidade de dados equivocados para esses horários nesse estado.
- Uma boa notícia é que a mediana revela um comportamento de aumento da radiação durante os horários de sol.

# DF - Brasília

Hora	Rad. Máx	Rad. Mín	Rad. Mediana
0	134	0	0.5
1	2.1	0	0.1
2	3.3	0	0.3
3	1.6	0	0.1
4	1.9	0	0.2
5	6	0	0.2
6	2.2	0	0.2
7	1.9	0	0.15
8	3.5	0	0.2
9	655	0	8.5
10	4472	0.1	88.7
11	13556	13	588
12	24322	4	1350
13	33272	49.6	2060
14	42546	111	2575
15	43439	36	2827
16	43969	46.6	2833
17	42015	32	2602
18	35207	6	2160
19	26745	9.2	1540
20	16279	8.6	850
21	5014	0	239
22	2585	0	22
23	5.1	0	0.55

# Análise breve Brasília

- Novamente, temos medianas diferentes de zero (mas relativamente próximas) ao longo da madrugada
- Conseguimos identificar um padrão do período solar pelas medianas
- As radiações máximas a partir de 8h começam a ter valores impraticáveis. Radiações de mais de 40 mil  $\text{kJm}^2$ , ao passo que outros estados tem radiações sempre menores que 10 mil.



## RJ - Rio de Janeiro

Hora	Rad. Máx	Rad. Mín	Rad. Mediana
0	490	0	0
1	813	0	0
2	932	0	0
3	920	0	0
4	791	0	0
5	659	0	0
6	335	0	0
7	722	0	0
8	1538	0	0
9	2327	0	20.7
10	3076	0	133
11	3470	7	592
12	3718	40.6	1173
13	3703	24.2	1727
14	3999	51.5	2203
15	4346	61.3	2406
16	4091	45	2378
17	3790	10	2141
18	6300	32	1668
19	3049	3	1042
20	2004	2.3	392
21	1019	0	62.5
22	261	0	8.9
23	267	0	0

# Análise breve Rio de Janeiro

- Conseguimos perceber ruídos devido a máximas diferentes de zero na madrugada
- Contudo, a mediana nos revela um bom padrão solar com os índices de radiação aumentando por volta de 12h.

# RS - Porto Alegre

Hora	Rad. Máx	Rad. Mín	Rad. Mediana
0	3.3	0	0.2
1	3.6	0	0.1
2	5.5	0	0.2
3	4.9	0	0.1
4	2.8	0	0
5	3.6	0	0.1
6	2.7	0	0.1
7	4.5	0	0.1
8	74	0	0.3
9	679	0	23.8
10	1518	0	113
11	2336	0	305
12	2978	2	696
13	3507	1.2	1328
14	3862	9.7	1872
15	4000	3.3	2136
16	4000	6.6	2208
17	3998	11	2098
18	3741	3.7	1769
19	3113	5.3	1284
20	2537	0	647
21	1670	0	209
22	662	0	72.9
23	108	0	10

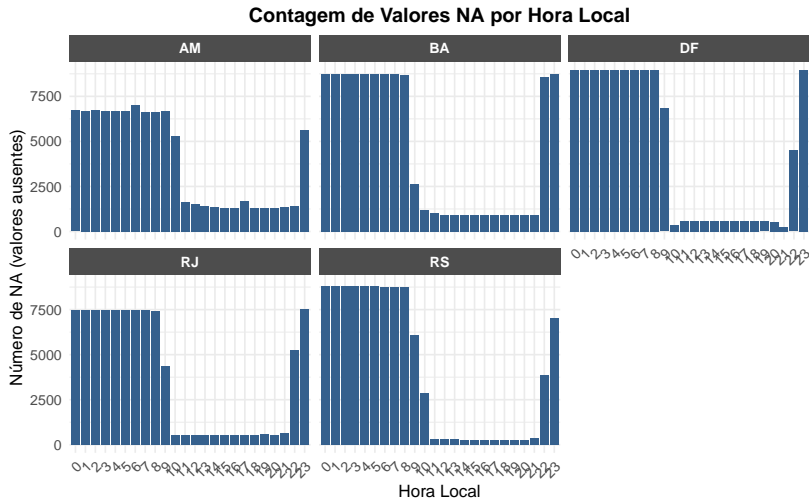
# Análise breve Porto Alegre

- Medianas na madrugada diferentes de zero levantam preocupações, assim como as mínimas muito baixas ao longo do dia
- Período solar bem evidenciado pelas medianas e temperaturas máximas

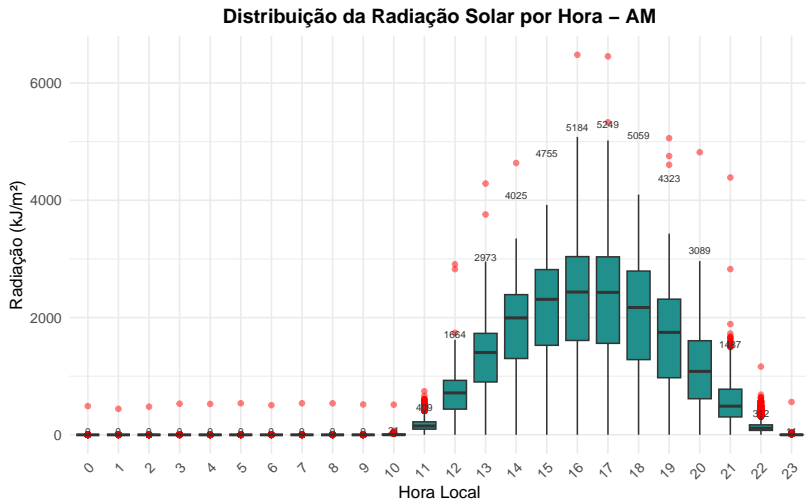
# Apanhado geral

- Dados de máxima e mínima no momento não dizem nada além de: existem ruídos no banco de dados
- As medianas já fornecem alguma informação que pode nos servir de alívio, indicando que boa parte dos dados seguem um padrão esperado.
- A questão das medianas não nulas de madrugada pode ser consequência do alto número de NA's nesses horários. Isso é, quando temos muitas entradas iguais a zero, a mediana será zero. Mas, com poucos dados, a mediana fica mais sensível.

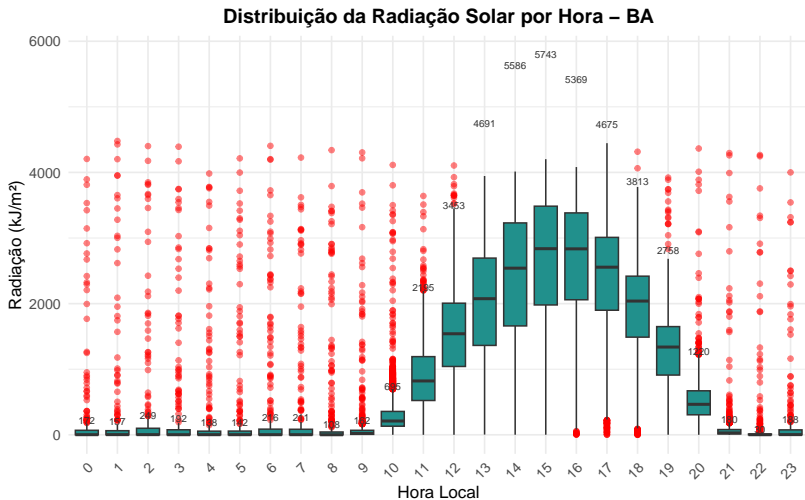
# Distribuição de NA por Estado e Hora



# Boxplot AM



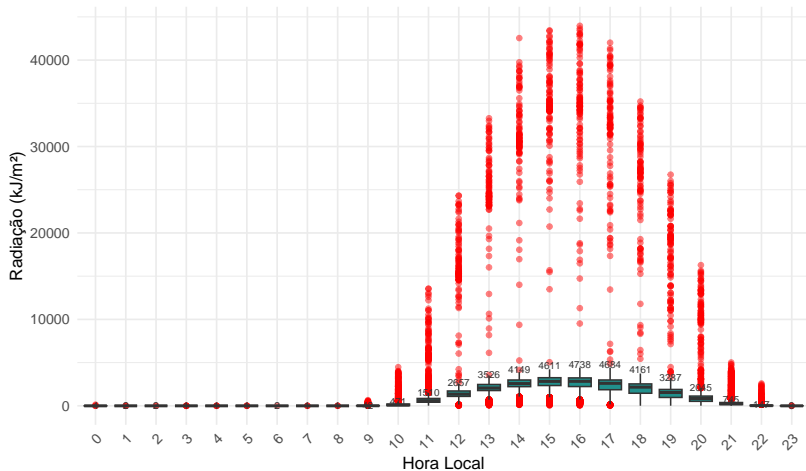
# Boxplot BA





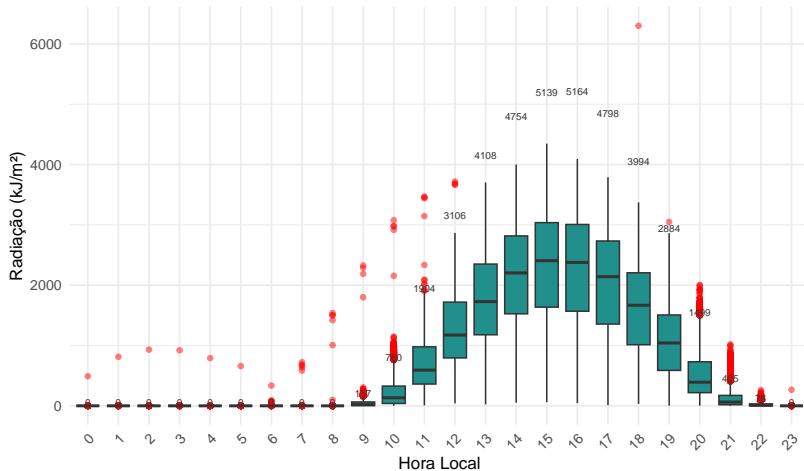
# Boxplot DF

Distribuição da Radiação Solar por Hora – DF

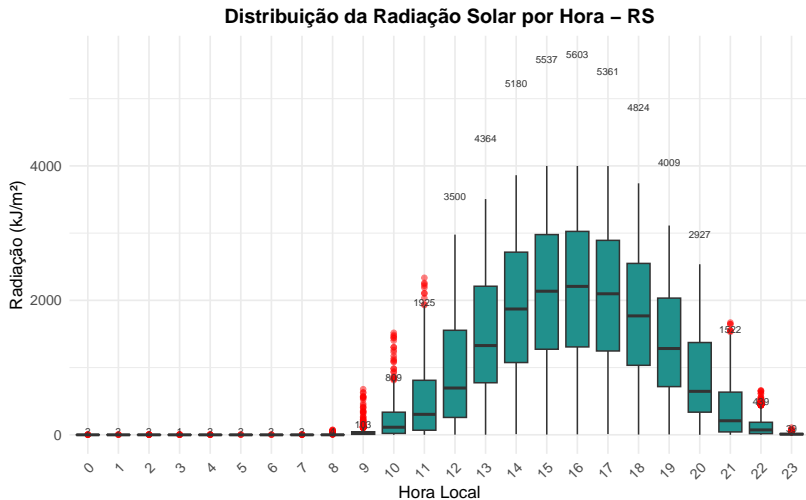


# Boxplot RJ

## Distribuição da Radiação Solar por Hora – RJ



# Boxplot RS



# Considerações

- Os boxplots estão aqui apenas como apoio visual. Ainda é preciso tratar os NA's e os outliers
- O cálculo dos limites superiores é meramente teórico ( $Q_3 + 1.5IQR = LS$ ). Inclusive, tratando-se da radiação é preferível repensar o fator 1.5 para o limite superior, porque a radiação pode ser altamente assimétrica.

# AM - Manaus

Hora	Obs. Totais	NA's	% NA	Acima Limite	% NA + Outliers
0	9003	6715	74.58625	145	76.19682
1	9003	6704	74.46407	151	76.14129
2	9003	6720	74.64179	156	76.37454
3	9003	6692	74.33078	144	75.93025
4	9003	6668	74.06420	126	75.46373
5	9003	6668	74.06420	122	75.41930
6	9003	7018	77.95179	116	79.24025
7	9003	6658	73.95313	140	75.50816
8	9003	6651	73.87537	167	75.73031
9	9003	6665	74.03088	148	75.67478
10	9003	5317	59.05809	75	59.89115
11	9003	1621	18.00511	217	20.41542
12	9003	1531	17.00544	3	17.03876
13	9003	1432	15.90581	2	15.92802
14	9003	1347	14.96168	1	14.97279
15	9003	1316	14.61735	0	14.61735
16	9003	1301	14.45074	1	14.46185
17	9003	1667	18.51605	2	18.53826
18	9003	1302	14.46185	0	14.46185
19	9003	1295	14.38409	3	14.41742
20	9003	1301	14.45074	1	14.46185
21	9003	1341	14.89504	20	15.11718
22	9003	1399	15.53926	357	19.50461
23	9003	5649	62.74575	228	65.27824

# BA - Salvador

Hora	Obs. Totais	NA's	% NA	Acima Limite	% NA + Outliers
0	8999	8755	97.288588	46	97.79976
1	8999	8755	97.288588	41	97.74419
2	8999	8757	97.310812	39	97.74419
3	8999	8741	97.133015	45	97.63307
4	8999	8738	97.099678	48	97.63307
5	8999	8746	97.188576	48	97.72197
6	8999	8722	96.921880	53	97.51083
7	8999	8726	96.966330	51	97.53306
8	8999	8666	96.299589	69	97.06634
9	8999	2652	29.469941	144	31.07012
10	8999	1190	13.223691	417	17.85754
11	8999	1009	11.212357	29	11.53462
12	8999	907	10.078898	12	10.21225
13	8999	915	10.167796	0	10.16780
14	8999	920	10.223358	0	10.22336
15	8999	926	10.290032	0	10.29003
16	8999	915	10.167796	0	10.16780
17	8999	924	10.267807	0	10.26781
18	8999	929	10.323369	2	10.34559
19	8999	931	10.345594	17	10.53450
20	8999	892	9.912213	52	10.49005
21	8999	939	10.434493	259	13.31259
22	8999	8561	95.132792	85	96.07734
23	8999	8763	97.377486	42	97.84420

# DF - Brasília

Hora	Obs. Totais	NA's	% NA	Acima Limite	% NA + Outliers
0	9005	8951	99.400333	4	99.444753
1	9005	8966	99.566907	3	99.600222
2	9005	8966	99.566907	2	99.589117
3	9005	8970	99.611327	6	99.677957
4	9005	8972	99.633537	4	99.677957
5	9005	8978	99.700167	4	99.744586
6	9005	8978	99.700167	1	99.711271
7	9005	8965	99.555802	3	99.589117
8	9005	8963	99.533592	3	99.566907
9	9005	6826	75.802332	115	77.079400
10	9005	339	3.764575	573	10.127707
11	9005	562	6.240977	186	8.306496
12	9005	589	6.540811	102	7.673515
13	9005	594	6.596335	100	7.706830
14	9005	586	6.507496	101	7.629095
15	9005	579	6.429761	101	7.551360
16	9005	576	6.396446	102	7.529151
17	9005	575	6.385342	104	7.540255
18	9005	569	6.318712	104	7.473626
19	9005	555	6.163243	106	7.340367
20	9005	520	5.774570	144	7.373681
21	9005	257	2.853970	804	11.782343
22	9005	4496	49.927818	319	53.470294
23	9005	8955	99.444753	3	99.478068

# RJ - Rio de Janeiro

Hora	Obs. Totais	NA's	% NA	Acima Limite	% NA + Outliers
0	9005	7489	83.164908	196	85.341477
1	9005	7488	83.153803	221	85.607996
2	9005	7467	82.920600	233	85.508051
3	9005	7488	83.153803	210	85.485841
4	9005	7482	83.087174	193	85.230428
5	9005	7480	83.064964	208	85.374792
6	9005	7467	82.920600	213	85.285952
7	9005	7472	82.976124	208	85.285952
8	9005	7400	82.176569	335	85.896724
9	9005	4337	48.162132	205	50.438645
10	9005	513	5.696835	413	10.283176
11	9005	513	5.696835	12	5.830094
12	9005	502	5.574681	4	5.619101
13	9005	504	5.596891	0	5.596891
14	9005	509	5.652415	0	5.652415
15	9005	514	5.707940	0	5.707940
16	9005	515	5.719045	0	5.719045
17	9005	534	5.930039	0	5.930039
18	9005	550	6.107718	1	6.118823
19	9005	559	6.207662	1	6.218767
20	9005	554	6.152138	254	8.972793
21	9005	623	6.918379	831	16.146585
22	9005	5244	58.234314	264	61.166019
23	9005	7505	83.342587	220	85.785675



# RS - Porto Alegre

Hora	Obs. Totais	NA's	% NA	Acima Limite	% NA + Outliers
0	8867	8800	99.244389	6	99.312056
1	8867	8798	99.221834	4	99.266945
2	8867	8784	99.063945	8	99.154167
3	8867	8796	99.199278	7	99.278223
4	8867	8773	98.939890	6	99.007556
5	8867	8779	99.007556	3	99.041389
6	8867	8757	98.759445	3	98.793278
7	8867	8763	98.827112	6	98.894778
8	8867	8717	98.308334	28	98.624112
9	8867	6062	68.365851	70	69.155295
10	8867	2883	32.513815	25	32.795760
11	8867	318	3.586331	10	3.699109
12	8867	294	3.315665	0	3.315665
13	8867	283	3.191609	0	3.191609
14	8867	260	2.932221	0	2.932221
15	8867	265	2.988609	0	2.988609
16	8867	276	3.112665	0	3.112665
17	8867	272	3.067554	0	3.067554
18	8867	265	2.988609	0	2.988609
19	8867	261	2.943498	0	2.943498
20	8867	267	3.011165	0	3.011165
21	8867	347	3.913387	7	3.992331
22	8867	3834	43.238976	103	44.400586
23	8867	7043	79.429345	36	79.835345

- A quantidade de NA's supera dez por cento em momentos de sol apenas em Manaus e Salvador.
- É possível considerar a possibilidade de trabalhar com esses dados e fazer um tratamento para melhorar sua qualidade, caso o interesse não seja perder os dados que apresentam erro.
- Para todo efeito, este é apenas um resultado parcial das análises realizadas até o momento

# Estudos futuros

- Para compreender melhor a questão da radiação, realizei uma coleção de artigos e documentos que possam fornecer bons insights. Destaco:
  - Atlas Brasileiro de Energia solar
  - Nasa power
  - guide to instruments and methods of observation (WMO - No 8)
  - Modelo de Estação Meteorológica Automática de Superfície.