

# paper03\_imputations\_airstata

October 10, 2021

## 1 Imputations for Missing Data on Our Sensor.

Sensor data can be handled as sequential data: ie a timeseries.

The main issue handling timeseries is the missing observations or missing data. We have missing observations due to power outages, software updates or gathering system downtime.

	temperature	pressure	humidity	gasResistance	IAQ	iaqAccuracy	\
0	21.54	777.41	43.93	151328	37.5		1
1	21.56	777.41	43.89	152702	35.6		1
2	21.53	777.41	43.97	151328	37.5		1
3	21.51	777.41	44.03	151464	38.5		1
4	21.51	777.41	44.05	152425	36.9		1
...	...	...	...	...	...		...
6285098	25.84	782.96	56.64	928867	130.8		1
6285099	25.83	782.94	56.66	923130	131.5		1
6285100	25.83	782.94	56.63	925034	131.3		1
6285101	25.83	782.94	56.62	923130	131.9		1
6285102	25.83	782.96	56.63	925034	131.6		1

	datetime	year	month	day	hour	minute
0	2021-02-12 06:04:09.089621067	2021	2	12	6	4
1	2021-02-12 06:04:12.087778807	2021	2	12	6	4
2	2021-02-12 06:04:15.072475433	2021	2	12	6	4
3	2021-02-12 06:04:18.070170164	2021	2	12	6	4
4	2021-02-12 06:04:21.061994791	2021	2	12	6	4
...	...	...	...	...	...	...
6285098	2021-09-18 01:20:38.889113188	2021	9	18	1	20
6285099	2021-09-18 01:20:41.882042885	2021	9	18	1	20
6285100	2021-09-18 01:20:44.877856970	2021	9	18	1	20
6285101	2021-09-18 01:20:47.872255564	2021	9	18	1	20
6285102	2021-09-18 01:20:50.866486311	2021	9	18	1	20

[6285103 rows x 12 columns]

We can create a difference between the times of observations to find the missing observations, as we know the sensor creates a new record every 3 seconds.

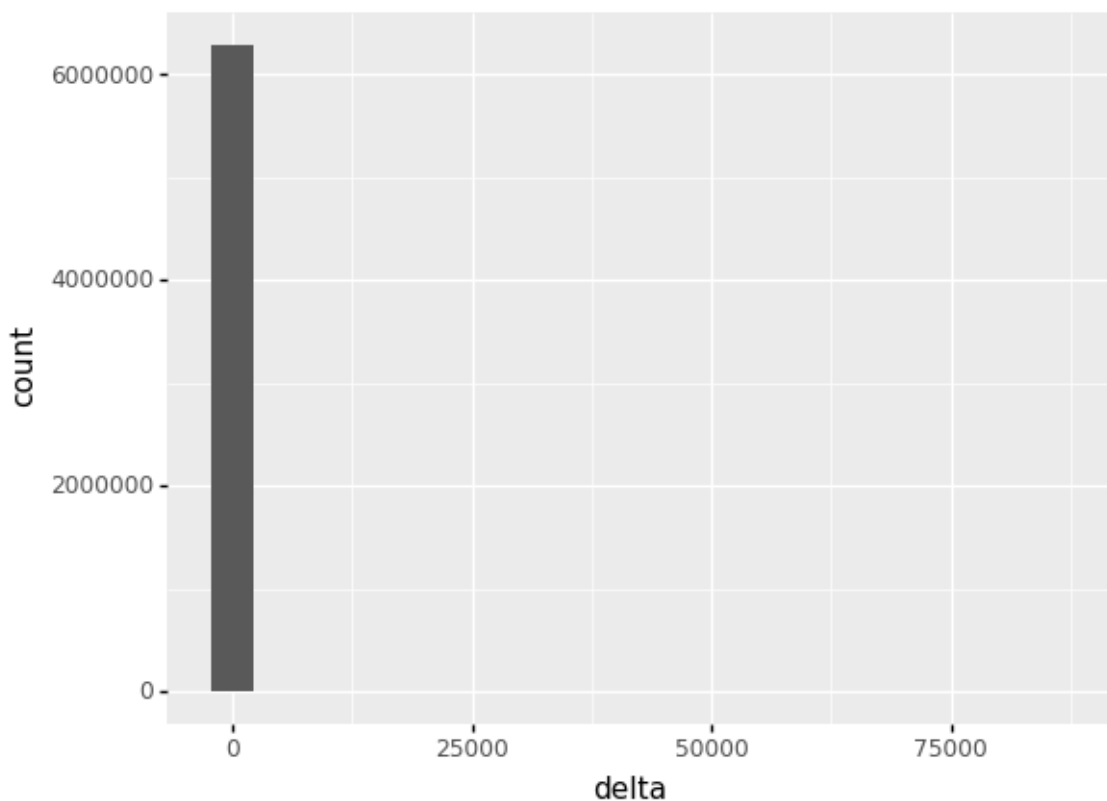
Then, if there's a missing observation, therefore the difference should be larger than those 3 seconds.

	temperature	pressure	humidity	gasResistance	IAQ	iaqAccuracy	\
0	21.51	777.41	44.04	152149	34.7		1
1	21.51	777.41	43.98	152841	33.6		1
2	21.54	777.41	43.73	153259	31.5		1
3	21.53	777.41	43.70	152841	31.5		1
4	21.52	777.41	43.70	153399	30.2		1
5	21.54	777.41	43.77	152702	30.9		1
6	21.55	777.40	43.76	152980	30.7		1
7	21.59	777.40	43.61	152841	30.8		1
8	21.59	777.41	43.56	152980	30.6		1
9	21.63	777.43	43.45	153679	28.8		1

	datetime	year	month	day	hour	minute	\
0	2021-02-12 06:05:35.846304417	2021	2	12	6	5	
1	2021-02-12 06:05:38.837326527	2021	2	12	6	5	
2	2021-02-12 06:05:47.812360048	2021	2	12	6	5	
3	2021-02-12 06:05:50.803695202	2021	2	12	6	5	
4	2021-02-12 06:05:53.795462847	2021	2	12	6	5	
5	2021-02-12 06:05:56.786891460	2021	2	12	6	5	
6	2021-02-12 06:05:59.778601646	2021	2	12	6	5	
7	2021-02-12 06:06:02.770255804	2021	2	12	6	6	
8	2021-02-12 06:06:05.761730671	2021	2	12	6	6	
9	2021-02-12 06:06:08.753019810	2021	2	12	6	6	

	datetime-1	delta	imputed
0	2021-02-12 06:05:29.856916904	5.0	False
1	2021-02-12 06:05:35.846304417	2.0	False
2	2021-02-12 06:05:38.837326527	8.0	False
3	2021-02-12 06:05:47.812360048	2.0	False
4	2021-02-12 06:05:50.803695202	2.0	False
5	2021-02-12 06:05:53.795462847	2.0	False
6	2021-02-12 06:05:56.786891460	2.0	False
7	2021-02-12 06:05:59.778601646	2.0	False
8	2021-02-12 06:06:02.770255804	2.0	False
9	2021-02-12 06:06:05.761730671	2.0	False

## 1.1 Missing Data



```
<ggplot: (8729120807801)>
```

Time difference between readings:

- Min: 2.0 seconds.
- Max: 85993.0 seconds.

The differences are mainly due to the following reasons:

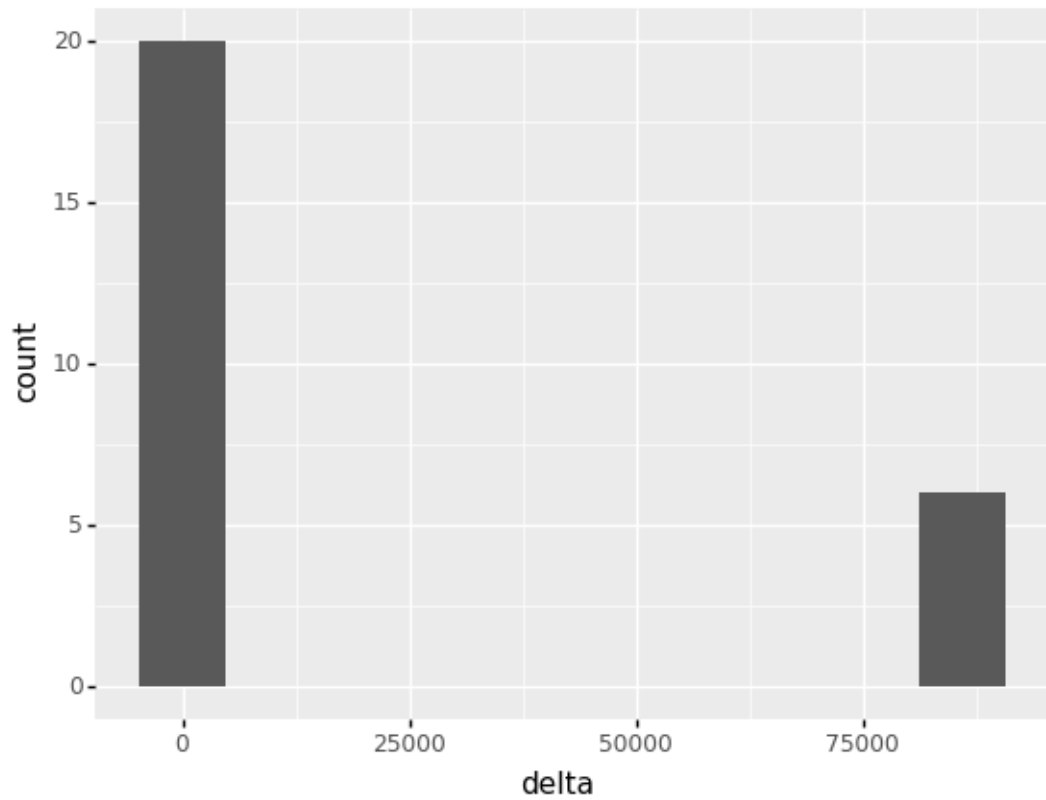
- It can be 2 seconds as the sensor and the systems that gathers and save to a permanent storage the records can be rounded, as our system is not a realtime system (a deterministic operating system).
- Differences larger than 3 seconds can be due to: system reboots, power outages, system updates. These events are seldom as you may appreciate:

	temperature	pressure	humidity	gasResistance	IAQ	iaqAccuracy \
0	21.51	777.41	44.04	152149	34.7	1
2	21.54	777.41	43.73	153259	31.5	1
11471	19.95	778.34	43.60	124814	236.8	1
11495	20.41	778.38	42.94	122095	243.2	1
194038	25.24	777.47	25.27	188409	28.7	1
711281	24.44	782.68	29.77	361856	57.8	3
711381	24.80	782.67	30.52	501364	25.0	0
711383	24.67	782.67	30.52	461738	25.0	0
1225137	25.72	781.48	25.65	499500	25.0	0
1816217	27.68	781.48	36.36	1022814	25.0	0
2613223	22.28	779.66	52.13	1187114	25.0	0
2613257	23.33	779.64	47.13	946517	25.0	0
2911711	28.23	780.46	32.44	850727	25.0	0
3921890	21.54	781.15	58.53	1136940	25.0	0
3921891	21.50	781.15	58.34	1114333	25.0	0
4655880	18.85	787.50	63.51	1167535	25.0	0
4655911	19.76	787.58	59.56	978894	25.0	0
5247822	25.03	784.42	55.43	1023592	25.0	0
5247823	25.00	784.44	55.23	1015094	25.0	0
5534130	22.10	779.38	64.59	1161485	25.0	0
5534160	23.73	779.30	57.95	993353	25.0	0
5671288	20.19	780.64	61.61	1236213	25.0	0
5671324	21.23	780.62	56.29	970418	25.0	0
5700342	20.09	781.58	63.31	943860	25.0	0
5700359	21.09	781.60	59.98	926947	25.0	0
5755798	23.11	780.95	57.93	1069157	25.0	0

	datetime	year	month	day	hour	minute \
0	2021-02-12 06:05:35.846304417	2021	2	12	6	5
2	2021-02-12 06:05:47.812360048	2021	2	12	6	5
11471	2021-02-12 15:38:09.454870701	2021	2	12	15	38
11495	2021-02-12 15:39:24.238069534	2021	2	12	15	39
194038	2021-02-18 23:30:02.871531487	2021	2	18	23	30
711281	2021-03-08 21:39:54.148881435	2021	3	8	21	39
711381	2021-03-08 21:45:12.908920765	2021	3	8	21	45
711383	2021-03-08 21:45:35.030911922	2021	3	8	21	45
1225137	2021-03-26 17:03:22.019438267	2021	3	26	17	3
1816217	2021-05-21 04:52:55.083148003	2021	5	21	4	52
2613223	2021-06-17 19:17:15.529208899	2021	6	17	19	17
2613257	2021-06-17 19:39:28.145390511	2021	6	17	19	39
2911711	2021-04-16 04:41:09.900250912	2021	4	16	4	41
3921890	2021-06-28 03:47:29.354257584	2021	6	28	3	47
3921891	2021-06-28 03:47:50.266227722	2021	6	28	3	47
4655880	2021-07-23 13:44:50.798935652	2021	7	23	13	44
4655911	2021-07-23 14:07:44.666577578	2021	7	23	14	7
5247822	2021-08-13 02:19:01.038081169	2021	8	13	2	19
5247823	2021-08-13 02:19:27.867798567	2021	8	13	2	19
5534130	2021-08-23 00:17:24.607170820	2021	8	23	0	17
5534160	2021-08-23 00:29:57.714995146	2021	8	23	0	29

5671288	2021-08-27	18:17:15.750379562	2021	8	27	18	17
5671324	2021-08-27	18:34:21.413355350	2021	8	27	18	34
5700342	2021-08-28	18:28:06.032265186	2021	8	28	18	28
5700359	2021-08-28	19:08:07.682873249	2021	8	28	19	8
5755798	2021-08-30	17:15:34.766223192	2021	8	30	17	15

		datetime-1	delta	imputed
0		2021-02-12 06:05:29.856916904	5.0	False
2		2021-02-12 06:05:38.837326527	8.0	False
11471		2021-02-12 15:37:33.558219671	35.0	False
11495		2021-02-12 15:39:18.254687548	5.0	False
194038		2021-02-18 23:23:30.376312494	392.0	False
711281		2021-03-08 21:39:39.176842928	14.0	False
711381		2021-03-08 21:44:50.610525370	22.0	False
711383		2021-03-08 21:45:15.906122684	19.0	False
1225137		2021-03-26 17:02:30.955729723	51.0	False
1816217		2021-04-16 04:40:10.764022350	764.0	False
2613223		2021-06-17 19:36:00.756372929	85274.0	False
2613257		2021-06-17 19:18:54.300177336	1233.0	False
2911711		2021-06-28 03:46:40.487612009	3269.0	False
3921890		2021-05-21 04:52:00.783312559	82528.0	False
3921891		2021-06-28 03:47:29.354257584	20.0	False
4655880		2021-07-23 14:04:10.164503574	85240.0	False
4655911		2021-07-23 13:46:20.478425026	1284.0	False
5247822		2021-08-13 02:18:20.963308096	40.0	False
5247823		2021-08-13 02:19:01.038081169	26.0	False
5534130		2021-08-23 00:24:10.624371529	85993.0	False
5534160		2021-08-23 00:18:51.275225639	666.0	False
5671288		2021-08-27 18:31:00.879312038	85574.0	False
5671324		2021-08-27 18:19:00.481747389	920.0	False
5700342		2021-08-28 18:41:50.590122700	85575.0	False
5700359		2021-08-28 18:28:53.921931505	2353.0	False
5755798		2021-08-30 17:14:20.699152231	74.0	False



```
<ggplot: (8729118032617)>
```

	temperature	pressure	humidity	gasResistance	IAQ	\
count	6285097.00	6285097.00	6285097.00	6285097.00	6285097.00	
mean	24.35	781.62	43.44	695073.49	157.43	
std	2.50	2.19	12.57	314473.23	71.43	
min	16.67	773.78	7.63	76404.00	0.00	
25%	22.52	780.21	32.68	503242.00	99.00	
50%	24.21	781.74	43.99	689762.00	172.90	
75%	26.12	783.16	54.35	863841.00	220.50	
max	31.29	788.19	71.37	2920738.00	500.00	

	iaqAccuracy	year	month	day	hour	minute	\
count	6285097.00	6285097.0	6285097.00	6285097.00	6285097.00	6285097.00	
mean	2.53	2021.0	5.51	15.63	11.51	29.50	
std	0.82	0.0	2.09	8.66	6.92	17.32	
min	0.00	2021.0	2.00	1.00	0.00	0.00	
25%	2.00	2021.0	4.00	8.00	6.00	14.00	
50%	3.00	2021.0	6.00	16.00	12.00	29.00	
75%	3.00	2021.0	7.00	23.00	18.00	44.00	
max	3.00	2021.0	9.00	31.00	23.00	59.00	

	delta
count	6285097.00
mean	2.08
std	83.11
min	2.00
25%	2.00
50%	2.00
75%	2.00
max	85993.00

	temperature	pressure	humidity	gasResistance	IAQ	iaqAccuracy	\
194038	25.24	777.47	25.27	188409	28.7	1	

	datetime	year	month	day	hour	minute	\
194038	2021-02-18 23:30:02.871531487	2021	2	18	23	30	

	datetime-1	delta	imputed
194038	2021-02-18 23:23:30.376312494	392.0	False

```
/home/jaa6766/.conda/envs/cuda/lib/python3.7/site-  
packages/ipykernel_launcher.py:1: UserWarning: Boolean Series key will be  
reindexed to match DataFrame index.
```

	temperature	pressure	humidity	gasResistance	IAQ	iaqAccuracy	\
194007	25.67	777.40	24.82	189056	35.0	3	
194008	25.66	777.43	24.83	187661	35.9	3	
194009	25.66	777.41	24.83	187661	36.4	3	
194010	25.64	777.41	24.82	188088	36.5	3	
194011	25.63	777.40	24.82	189925	34.8	3	
194012	25.64	777.41	24.81	189272	34.3	3	
194013	25.62	777.43	24.83	188409	34.8	3	
194014	25.61	777.38	24.90	187874	35.6	3	
194015	25.63	777.41	24.95	187342	36.6	3	
194016	25.64	777.41	24.97	187129	37.5	3	
194017	25.64	777.40	24.94	187448	37.8	3	

194018	25.64	777.43	24.90	189707	35.9	3
194019	25.63	777.41	24.91	189489	34.7	3
194020	25.63	777.41	24.90	187874	35.5	3
194021	25.63	777.41	24.89	188302	35.7	3
194022	25.61	777.40	24.91	188409	35.7	3
194023	25.59	777.40	24.94	189381	34.8	3
194024	25.60	777.41	25.01	188732	34.7	3
194025	25.61	777.40	25.05	186496	36.8	3
194026	25.64	777.43	25.01	186918	37.7	3
194027	25.67	777.40	25.01	187024	38.2	3
194028	25.68	777.41	24.95	187661	37.9	3
194029	25.68	777.41	24.93	188409	36.9	3
194030	25.69	777.40	24.90	187024	37.6	3
194031	25.70	777.43	24.89	186812	38.3	3
194032	25.70	777.43	24.91	188088	37.5	3
194033	25.71	777.43	24.92	187767	37.2	3
194034	25.69	777.43	24.95	186181	38.5	3
194035	25.69	777.43	24.89	187235	38.4	3
194036	25.69	777.43	24.85	189272	36.4	3
194037	25.71	777.41	24.77	187661	36.7	3
194038	25.24	777.47	25.27	188409	28.7	1

	datetime	year	month	day	hour	minute	\
194007	2021-02-18 23:22:00.555237055	2021	2	18	23	22	
194008	2021-02-18 23:22:03.552131891	2021	2	18	23	22	
194009	2021-02-18 23:22:06.543661118	2021	2	18	23	22	
194010	2021-02-18 23:22:09.544224024	2021	2	18	23	22	
194011	2021-02-18 23:22:12.538087845	2021	2	18	23	22	
194012	2021-02-18 23:22:15.531872988	2021	2	18	23	22	
194013	2021-02-18 23:22:18.525767088	2021	2	18	23	22	
194014	2021-02-18 23:22:21.519450902	2021	2	18	23	22	
194015	2021-02-18 23:22:24.513052940	2021	2	18	23	22	
194016	2021-02-18 23:22:27.506954193	2021	2	18	23	22	
194017	2021-02-18 23:22:30.500771999	2021	2	18	23	22	
194018	2021-02-18 23:22:33.488167286	2021	2	18	23	22	
194019	2021-02-18 23:22:36.488412857	2021	2	18	23	22	
194020	2021-02-18 23:22:39.481906414	2021	2	18	23	22	
194021	2021-02-18 23:22:42.475703716	2021	2	18	23	22	
194022	2021-02-18 23:22:45.463096142	2021	2	18	23	22	
194023	2021-02-18 23:22:48.457271814	2021	2	18	23	22	
194024	2021-02-18 23:22:51.451401234	2021	2	18	23	22	
194025	2021-02-18 23:22:54.445565462	2021	2	18	23	22	
194026	2021-02-18 23:22:57.442430258	2021	2	18	23	22	
194027	2021-02-18 23:23:00.439792395	2021	2	18	23	23	
194028	2021-02-18 23:23:03.433859587	2021	2	18	23	23	
194029	2021-02-18 23:23:06.427474499	2021	2	18	23	23	
194030	2021-02-18 23:23:09.421072960	2021	2	18	23	23	
194031	2021-02-18 23:23:12.414887190	2021	2	18	23	23	
194032	2021-02-18 23:23:15.408778191	2021	2	18	23	23	
194033	2021-02-18 23:23:18.402513504	2021	2	18	23	23	
194034	2021-02-18 23:23:21.390101194	2021	2	18	23	23	
194035	2021-02-18 23:23:24.390413523	2021	2	18	23	23	
194036	2021-02-18 23:23:27.384147644	2021	2	18	23	23	
194037	2021-02-18 23:23:30.376312494	2021	2	18	23	23	
194038	2021-02-18 23:30:02.871531487	2021	2	18	23	30	

	datetime-1	delta	imputed
194007	2021-02-18 23:21:57.560976982	2.0	False
194008	2021-02-18 23:22:00.555237055	2.0	False

194009	2021-02-18	23:22:03.552131891	2.0	False
194010	2021-02-18	23:22:06.543661118	3.0	False
194011	2021-02-18	23:22:09.544224024	2.0	False
194012	2021-02-18	23:22:12.538087845	2.0	False
194013	2021-02-18	23:22:15.531872988	2.0	False
194014	2021-02-18	23:22:18.525767088	2.0	False
194015	2021-02-18	23:22:21.519450902	2.0	False
194016	2021-02-18	23:22:24.513052940	2.0	False
194017	2021-02-18	23:22:27.506954193	2.0	False
194018	2021-02-18	23:22:30.500771999	2.0	False
194019	2021-02-18	23:22:33.488167286	3.0	False
194020	2021-02-18	23:22:36.488412857	2.0	False
194021	2021-02-18	23:22:39.481906414	2.0	False
194022	2021-02-18	23:22:42.475703716	2.0	False
194023	2021-02-18	23:22:45.463096142	2.0	False
194024	2021-02-18	23:22:48.457271814	2.0	False
194025	2021-02-18	23:22:51.451401234	2.0	False
194026	2021-02-18	23:22:54.445565462	2.0	False
194027	2021-02-18	23:22:57.442430258	2.0	False
194028	2021-02-18	23:23:00.439792395	2.0	False
194029	2021-02-18	23:23:03.433859587	2.0	False
194030	2021-02-18	23:23:06.427474499	2.0	False
194031	2021-02-18	23:23:09.421072960	2.0	False
194032	2021-02-18	23:23:12.414887190	2.0	False
194033	2021-02-18	23:23:15.408778191	2.0	False
194034	2021-02-18	23:23:18.402513504	2.0	False
194035	2021-02-18	23:23:21.390101194	3.0	False
194036	2021-02-18	23:23:24.390413523	2.0	False
194037	2021-02-18	23:23:27.384147644	2.0	False
194038	2021-02-18	23:23:30.376312494	392.0	False

	temperature	pressure	humidity	gasResistance	IAQ	iaqAccuracy	\
194024	25.60	777.41	25.01	188732	34.7		3
194025	25.61	777.40	25.05	186496	36.8		3
194026	25.64	777.43	25.01	186918	37.7		3
194027	25.67	777.40	25.01	187024	38.2		3
194028	25.68	777.41	24.95	187661	37.9		3
194029	25.68	777.41	24.93	188409	36.9		3
194030	25.69	777.40	24.90	187024	37.6		3
194031	25.70	777.43	24.89	186812	38.3		3
194032	25.70	777.43	24.91	188088	37.5		3
194033	25.71	777.43	24.92	187767	37.2		3
194034	25.69	777.43	24.95	186181	38.5		3
194035	25.69	777.43	24.89	187235	38.4		3
194036	25.69	777.43	24.85	189272	36.4		3
194037	25.71	777.41	24.77	187661	36.7		3
194038	25.24	777.47	25.27	188409	28.7		1

		datetime	year	month	day	hour	minute	\
194024	2021-02-18	23:22:51.451401234	2021	2	18	23	22	
194025	2021-02-18	23:22:54.445565462	2021	2	18	23	22	
194026	2021-02-18	23:22:57.442430258	2021	2	18	23	22	
194027	2021-02-18	23:23:00.439792395	2021	2	18	23	23	
194028	2021-02-18	23:23:03.433859587	2021	2	18	23	23	
194029	2021-02-18	23:23:06.427474499	2021	2	18	23	23	
194030	2021-02-18	23:23:09.421072960	2021	2	18	23	23	
194031	2021-02-18	23:23:12.414887190	2021	2	18	23	23	
194032	2021-02-18	23:23:15.408778191	2021	2	18	23	23	
194033	2021-02-18	23:23:18.402513504	2021	2	18	23	23	

194034	2021-02-18	23:23:21.390101194	2021	2	18	23	23
194035	2021-02-18	23:23:24.390413523	2021	2	18	23	23
194036	2021-02-18	23:23:27.384147644	2021	2	18	23	23
194037	2021-02-18	23:23:30.376312494	2021	2	18	23	23
194038	2021-02-18	23:30:02.871531487	2021	2	18	23	30

			datetime-1	delta	imputed
194024	2021-02-18	23:22:48.457271814		2.0	False
194025	2021-02-18	23:22:51.451401234		2.0	False
194026	2021-02-18	23:22:54.445565462		2.0	False
194027	2021-02-18	23:22:57.442430258		2.0	False
194028	2021-02-18	23:23:00.439792395		2.0	False
194029	2021-02-18	23:23:03.433859587		2.0	False
194030	2021-02-18	23:23:06.427474499		2.0	False
194031	2021-02-18	23:23:09.421072960		2.0	False
194032	2021-02-18	23:23:12.414887190		2.0	False
194033	2021-02-18	23:23:15.408778191		2.0	False
194034	2021-02-18	23:23:18.402513504		2.0	False
194035	2021-02-18	23:23:21.390101194		3.0	False
194036	2021-02-18	23:23:24.390413523		2.0	False
194037	2021-02-18	23:23:27.384147644		2.0	False
194038	2021-02-18	23:23:30.376312494		392.0	False

	temperature	pressure	humidity	gasResistance	IAQ	iaqAccuracy	\
194038	25.24	777.47	25.27	188409	28.7	1	
194039	25.23	777.47	25.21	187342	31.0	1	
194040	25.24	777.49	25.16	187342	32.7	1	
194041	25.26	777.49	25.11	188840	31.0	1	
194042	25.28	777.49	25.05	189056	29.5	1	
194043	25.31	777.47	25.03	188088	30.3	1	
194044	25.33	777.47	24.98	189707	27.8	1	
194045	25.36	777.47	24.94	188840	27.8	1	
194046	25.38	777.45	24.98	186391	32.3	1	
194047	25.39	777.47	25.00	187554	33.2	1	
194048	25.41	777.47	25.02	188195	32.5	1	
194049	25.43	777.47	24.97	188088	32.2	1	
194050	25.44	777.47	24.99	187235	33.5	1	
194051	25.45	777.45	25.01	188302	32.4	1	
194052	25.46	777.49	24.98	187767	32.5	1	

			datetime	year	month	day	hour	minute	\
194038	2021-02-18	23:30:02.871531487		2021	2	18	23	30	
194039	2021-02-18	23:30:05.860353708		2021	2	18	23	30	
194040	2021-02-18	23:30:08.853756189		2021	2	18	23	30	
194041	2021-02-18	23:30:11.847195148		2021	2	18	23	30	
194042	2021-02-18	23:30:14.840431452		2021	2	18	23	30	
194043	2021-02-18	23:30:17.833710194		2021	2	18	23	30	
194044	2021-02-18	23:30:20.826932430		2021	2	18	23	30	
194045	2021-02-18	23:30:23.820370197		2021	2	18	23	30	
194046	2021-02-18	23:30:26.814028502		2021	2	18	23	30	
194047	2021-02-18	23:30:29.807134628		2021	2	18	23	30	
194048	2021-02-18	23:30:32.800416470		2021	2	18	23	30	
194049	2021-02-18	23:30:35.794191837		2021	2	18	23	30	
194050	2021-02-18	23:30:38.786831379		2021	2	18	23	30	
194051	2021-02-18	23:30:41.780043602		2021	2	18	23	30	
194052	2021-02-18	23:30:44.773188829		2021	2	18	23	30	

			datetime-1	delta	imputed
194038	2021-02-18	23:23:30.376312494		392.0	False



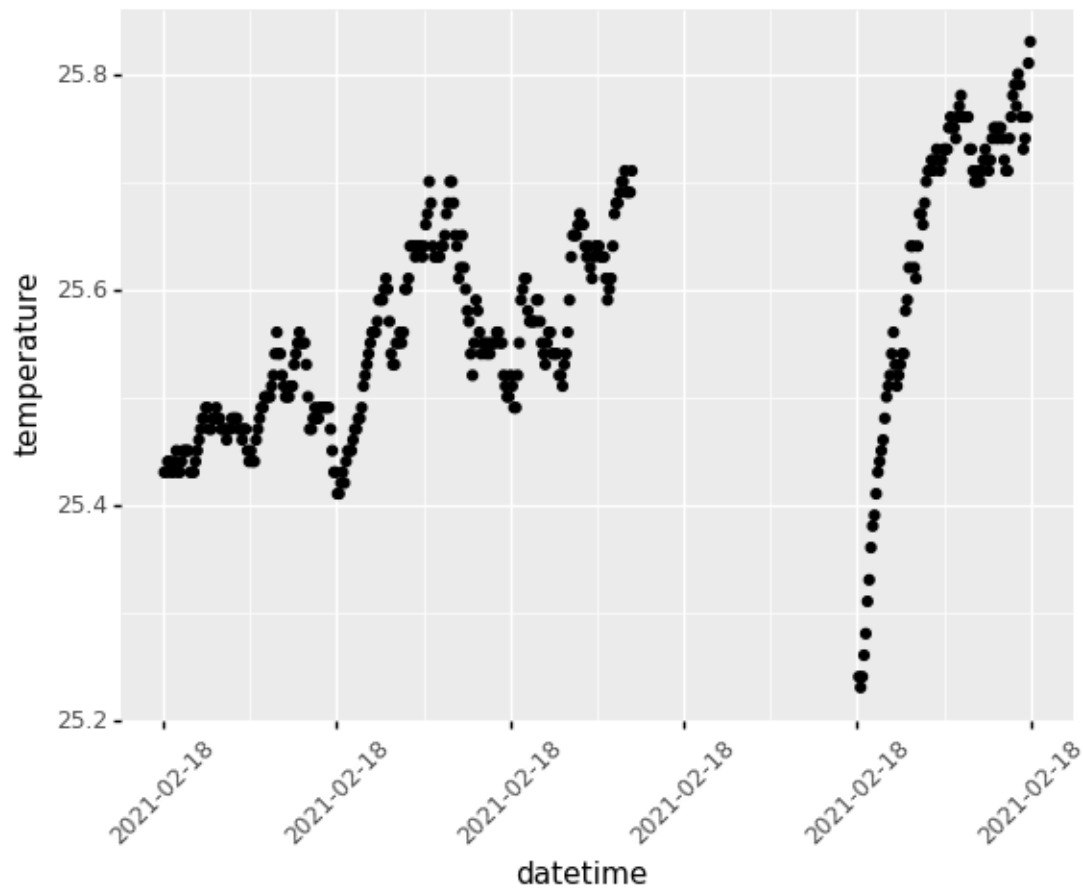
194039	2021-02-18	23:30:02.871531487	2.0	False
194040	2021-02-18	23:30:05.860353708	2.0	False
194041	2021-02-18	23:30:08.853756189	2.0	False
194042	2021-02-18	23:30:11.847195148	2.0	False
194043	2021-02-18	23:30:14.840431452	2.0	False
194044	2021-02-18	23:30:17.833710194	2.0	False
194045	2021-02-18	23:30:20.826932430	2.0	False
194046	2021-02-18	23:30:23.820370197	2.0	False
194047	2021-02-18	23:30:26.814028502	2.0	False
194048	2021-02-18	23:30:29.807134628	2.0	False
194049	2021-02-18	23:30:32.800416470	2.0	False
194050	2021-02-18	23:30:35.794191837	2.0	False
194051	2021-02-18	23:30:38.786831379	2.0	False
194052	2021-02-18	23:30:41.780043602	2.0	False

## 1.2 Missing Data Visualizations

```

/home/jaa6766/.conda/envs/cuda/lib/python3.7/site-
packages/ipykernel_launcher.py:3: UserWarning: Boolean Series key will be
reindexed to match DataFrame index.

```

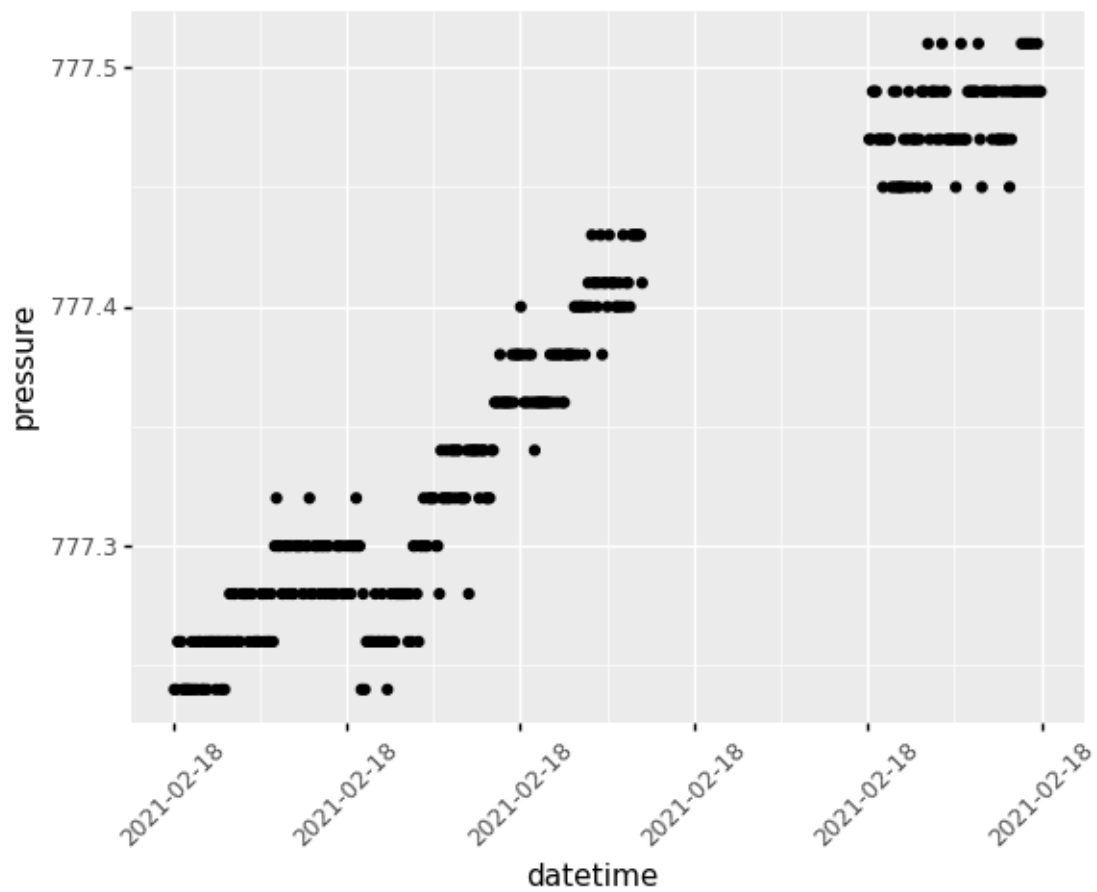


```
<ggplot: (8729117976041)>
```

```

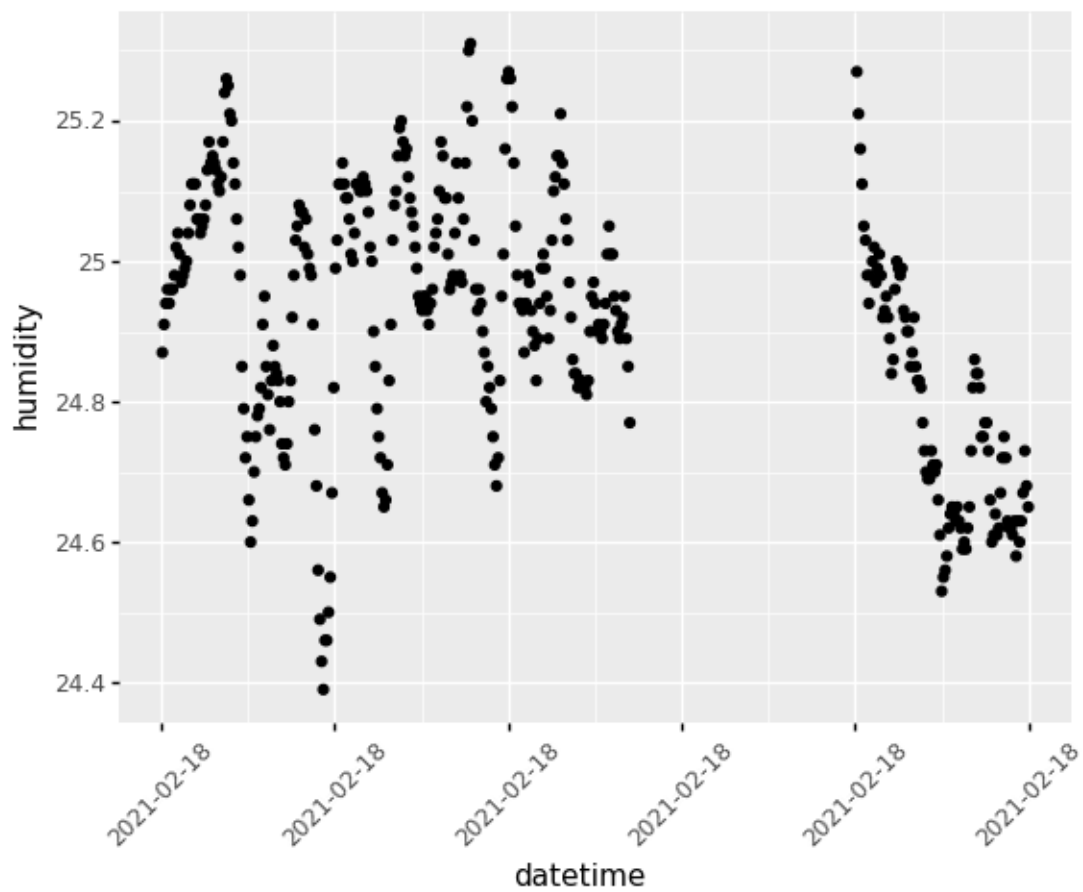
/home/jaa6766/.conda/envs/cuda/lib/python3.7/site-
packages/ipykernel_launcher.py:3: UserWarning: Boolean Series key will be
reindexed to match DataFrame index.

```



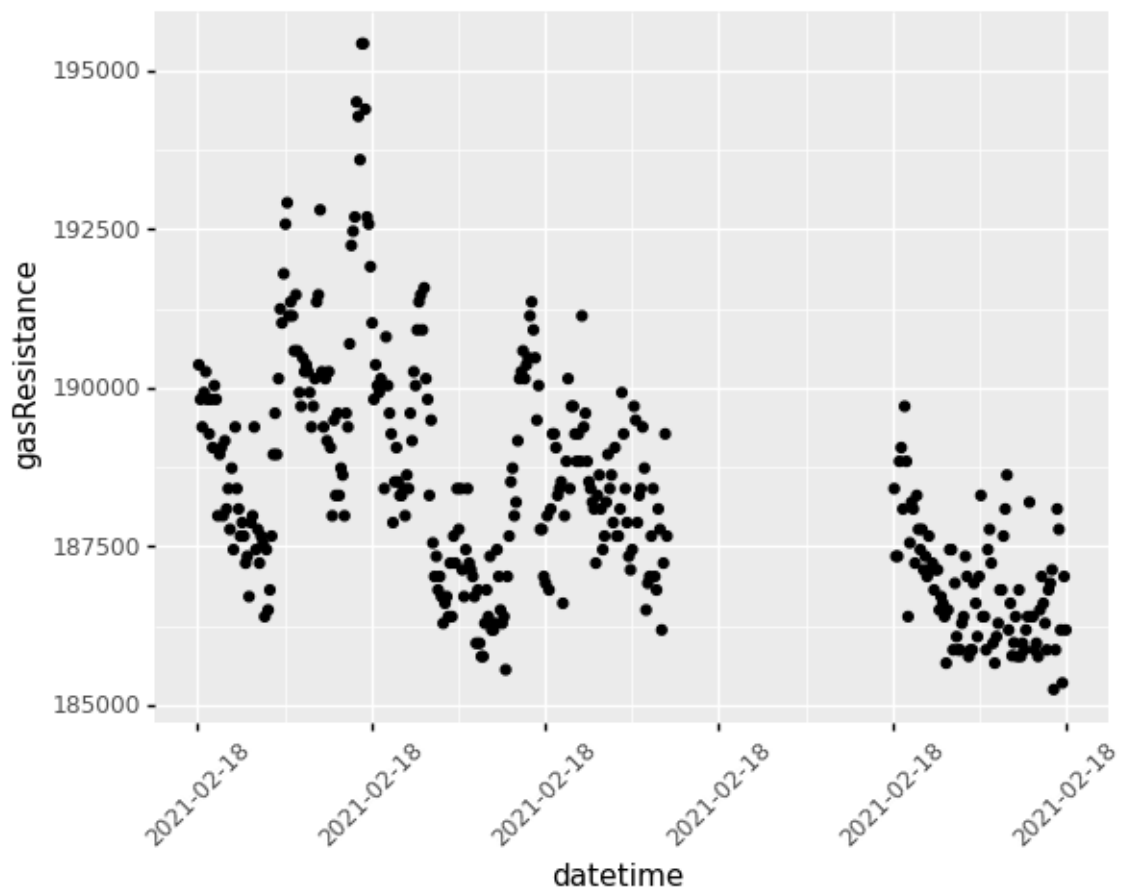
```
<ggplot: (8729117979797)>
```

```
/home/jaa6766/.conda/envs/cuda/lib/python3.7/site-  
packages/ipykernel_launcher.py:3: UserWarning: Boolean Series key will be  
reindexed to match DataFrame index.
```



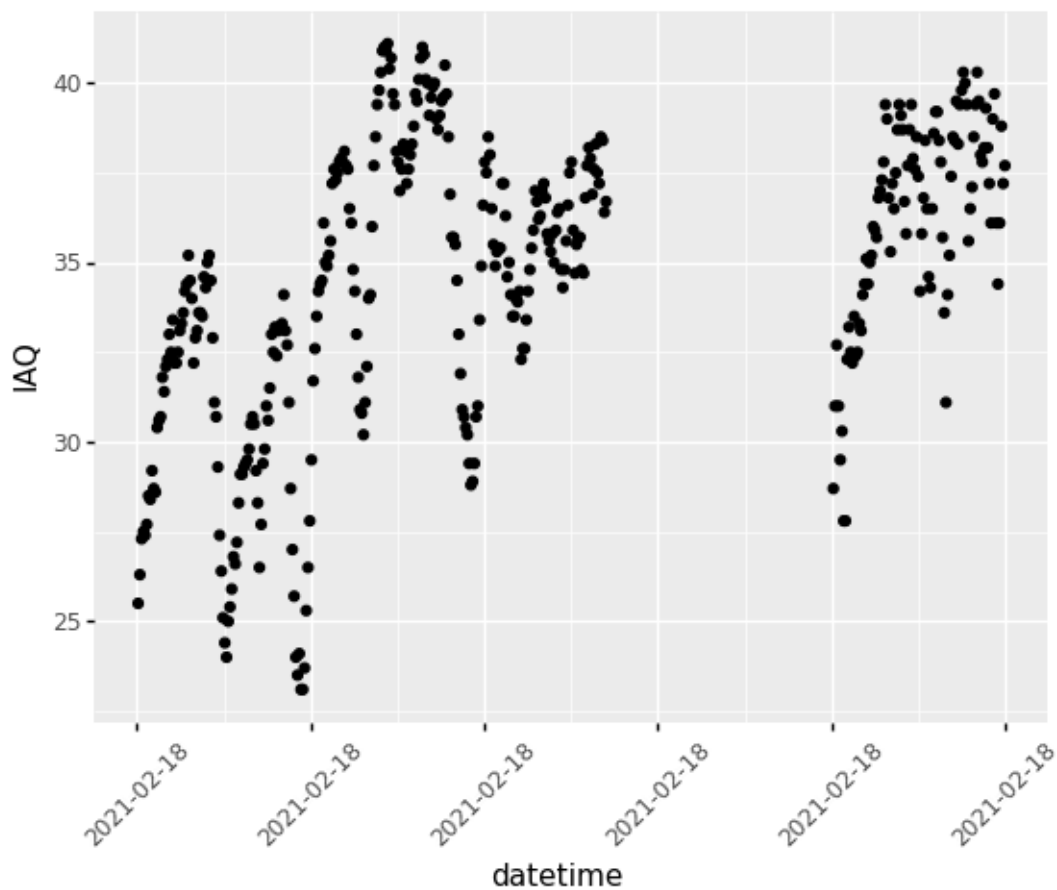
```
<ggplot: (8729117912305)>
```

```
/home/jaa6766/.conda/envs/cuda/lib/python3.7/site-  
packages/ipykernel_launcher.py:3: UserWarning: Boolean Series key will be  
reindexed to match DataFrame index.
```



```
<ggplot: (8729117975997)>
```

```
/home/jaa6766/.conda/envs/cuda/lib/python3.7/site-  
packages/ipykernel_launcher.py:3: UserWarning: Boolean Series key will be  
reindexed to match DataFrame index.
```



```
<ggplot: (8729118601317)>
```

One option could be to discard previous data and only having 6,090,958 records out of (6,285,097).

Another options is to impute the missing data.

### 1.3 Imputations

```
/home/jaa6766/.conda/envs/cuda/lib/python3.7/site-
packages/ipykernel_launcher.py:29: RuntimeWarning: divide by zero encountered in
double_scalars
```

```
/home/jaa6766/.conda/envs/cuda/lib/python3.7/site-
packages/ipykernel_launcher.py:29: RuntimeWarning: invalid value encountered in
double_scalars
```

Table with missing data:

	temperature	pressure	humidity	gasResistance	IAQ	iaqAccuracy	\
2	21.54	777.41	43.73	153259.0	31.5	1	
11506	20.41	778.38	42.94	122095.0	243.2	1	
2826589	27.68	781.48	36.36	1022814.0	25.0	0	
3623629	23.33	779.64	47.13	946517.0	25.0	0	

	datetime	year	month	day	hour	minute	\
2	2021-02-12 06:05:47.812360048	2021.0	2.0	12.0	6.0	5.0	
11506	2021-02-12 15:39:24.238069534	2021.0	2.0	12.0	15.0	39.0	
2826589	2021-05-21 04:52:55.083148003	2021.0	5.0	21.0	4.0	52.0	
3623629	2021-06-17 19:39:28.145390511	2021.0	6.0	17.0	19.0	39.0	

datetime-1 delta imputed

2	2021-02-12 06:05:38.837326527	8.0	False
11506	2021-02-12 15:39:18.254687548	5.0	False
2826589	2021-05-21 04:52:00.783312559	54.0	False
3623629	2021-06-17 19:36:00.756372929	207.0	False

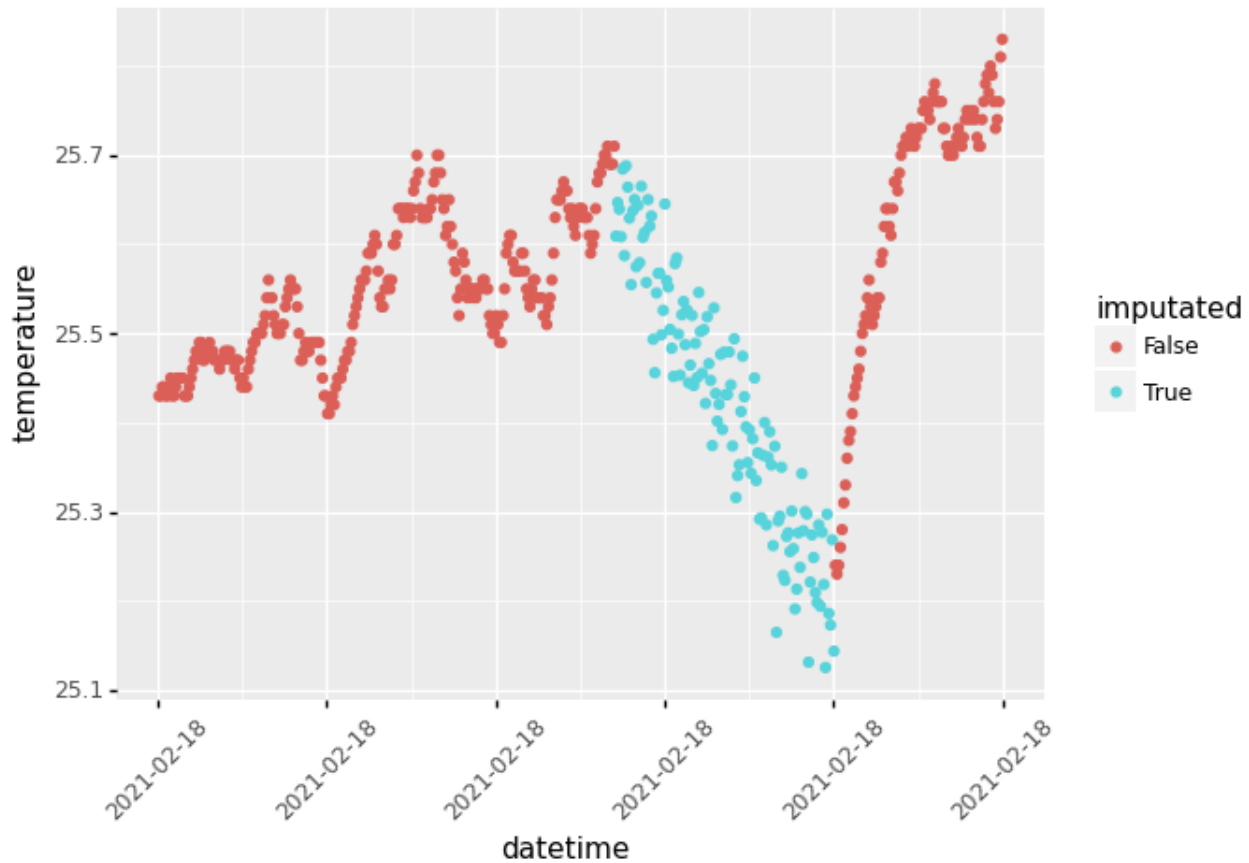
It is noteworthy that these values are very small (lesser than 10 seconds) that will be mitigated with the resampling of the data we will see in next sections.

CPU times: user 1min, sys: 1min 24s, total: 2min 24s

Wall time: 2min 24s

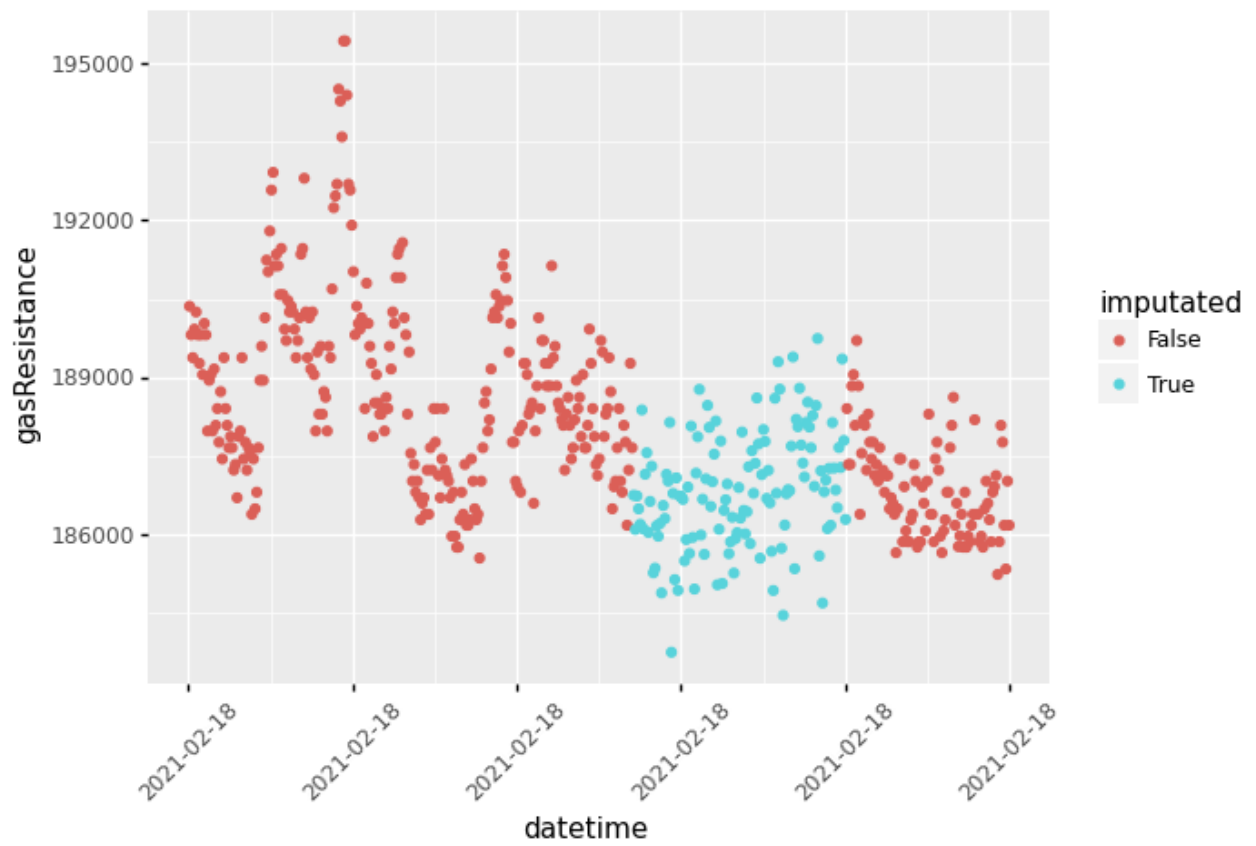
## 1.4 Imputed Values Visualizations

```
/home/jaa6766/.conda/envs/cuda/lib/python3.7/site-
packages/ipykernel_launcher.py:4: UserWarning: Boolean Series key will be
reindexed to match DataFrame index.
```



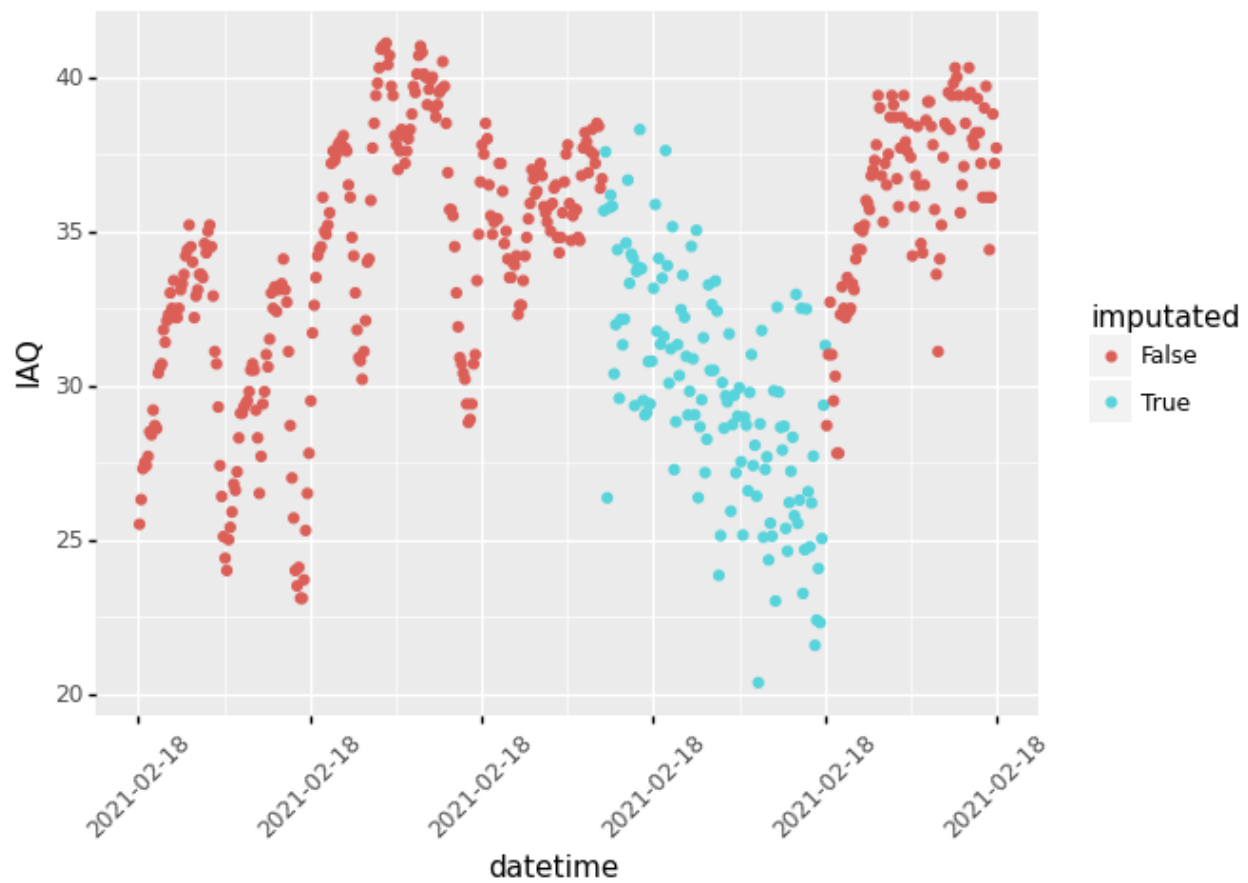
```
<ggplot: (8729118013281)>
```

```
/home/jaa6766/.conda/envs/cuda/lib/python3.7/site-
packages/ipykernel_launcher.py:4: UserWarning: Boolean Series key will be
reindexed to match DataFrame index.
```



```
<ggplot: (8729116297793)>
```

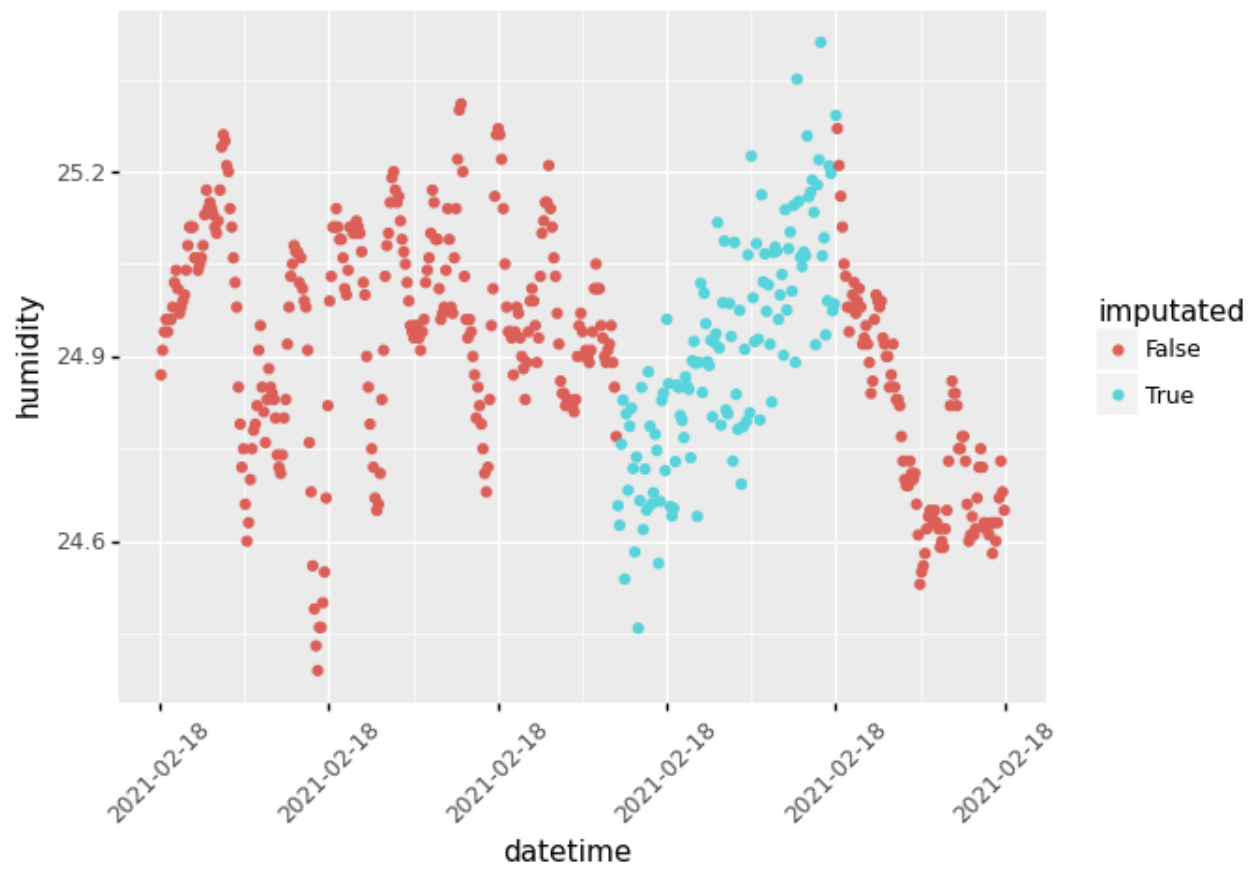
```
/home/jaa6766/.conda/envs/cuda/lib/python3.7/site-  
packages/ipykernel_launcher.py:4: UserWarning: Boolean Series key will be  
reindexed to match DataFrame index.
```



```
<ggplot: (8729117893681)>
```

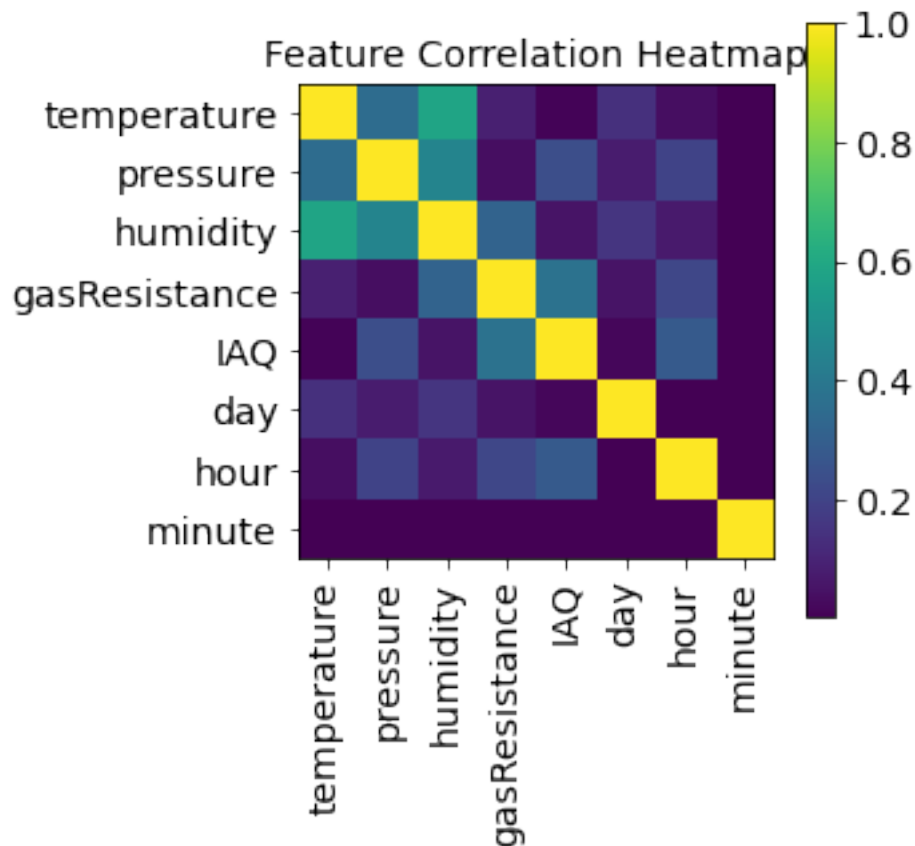
```
/home/jaa6766/.conda/envs/cuda/lib/python3.7/site-packages/ipykernel_launcher.py:4: UserWarning: Boolean Series key will be  
reindexed to match DataFrame index.
```





<ggplot: (8729116288241)>

### 1.4.1 Relationships Between Variables



	temperature	pressure	humidity	gasResistance	IAQ	day	\
temperature	1.00	0.35	0.58	0.09	0.01	0.14	
pressure	0.35	1.00	0.45	0.04	0.24	0.08	
humidity	0.58	0.45	1.00	0.32	0.05	0.15	
gasResistance	0.09	0.04	0.32	1.00	0.38	0.05	
IAQ	0.01	0.24	0.05	0.38	1.00	0.02	
day	0.14	0.08	0.15	0.05	0.02	1.00	
hour	0.04	0.20	0.07	0.21	0.28	0.00	
minute	0.00	0.00	0.00	0.00	0.00	0.00	

	hour	minute
temperature	0.04	0.0
pressure	0.20	0.0
humidity	0.07	0.0
gasResistance	0.21	0.0
IAQ	0.28	0.0
day	0.00	0.0
hour	1.00	0.0
minute	0.00	1.0

## 2 References

- [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/basics.html#essential-basic-functionality](https://pandas.pydata.org/pandas-docs/stable/user_guide/basics.html#essential-basic-functionality)
- [https://pandas.pydata.org/docs/user\\_guide/timeseries.html](https://pandas.pydata.org/docs/user_guide/timeseries.html)
- [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/timedeltas.html#attributes](https://pandas.pydata.org/pandas-docs/stable/user_guide/timedeltas.html#attributes)
- <https://www.datasciencemadesimple.com/re-arrange-or-re-order-the-column-of-dataframe-in-pandas-python-2/>
- <https://medium.com/@NatalieOlivo/use-pandas-to-lag-your-timeseries-data-in-order-to-examine-causal-relat>

- <https://www.geeksforgeeks.org/ways-to-filter-pandas-dataframe-by-column-values/>
- <https://stackoverflow.com/a/55557758/7323086>