

Temporal and spatial variations of PM_{2.5} and PM₁₀ concentrations in Mongolia

Erdenebayar Munkhtsetseg^{1,2}, Atsushi Shimizu³

¹National University of Mongolia (NUM), Mongolia,

²Kanazawa University, Japan,

³National Institute for Environmental Studies (NIES), Japan,

Corresponding author: Atsushi Shimizu, shimizua@nies.go.jp

Abstract

PM2.5 and PM10 data for the 4 distinct sites of Mongolia from 2008 to 2020 is found

Plain Language Summary

PM2.5 and PM10 data for the 4 distinct sites of Mongolia from 2008 to 2020 is found ...

0.1 Data & Methods**1 01_datawork**

Munkhtsetseg

Library

1.1 Import the dataset and remove the duplicates

Import the dataset from the directory of: ~/Data Input/Preprocessing data/Preprocessing data.csv, assign the dataset as object of df:

Remove the duplicates with the function of distinct(), assign the dataset as df_01:

1.2 Produce a table with missing data

A tibble: 6 × 19

Rowwise:

	Year	Month	Day	Hour	correct_PM10	correct_PM2	Visibility	WD	WS	WS.u
	<int>	<int>	<int>	<int>	<dbl>	<dbl>	<int>	<int>	<dbl>	<dbl>
1	2009	1	1	0	0.087	0.073	3366	252	1.08	1.02
2	2009	1	1	1	0.2	0.175	1999	317	0.546	0.374
3	2009	1	1	2	0.309	0.266	6756	87	1.03	-1.03
4	2009	1	1	3	0.105	0.089	9559	121	1.03	-0.885
5	2009	1	1	4	0.063	0.052	17664	121	0.46	-0.393
6	2009	1	1	5	0.027	0.021	11348	92	2.05	-2.05

9 more variables: WS.v <dbl>, OPC <int>, Station.name <chr>, Date <chr>,
PM10 <dbl>, PM2 <dbl>, PM10_rel <dbl>, PM2_rel <dbl>, ratio <dbl>

1.2.0.1 For date options as year, month, etc:

A tibble: 35 × 9

Groups: Station.name [4]

	Station.name	Year	NA_date	NA_PM2	NA_PM10	NA_Vis	NA_WD	NA_WS	NA_OPC
	<chr>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>
1	Dalanzadgad	2009	8760	929	715	659	748	748	8760
2	Dalanzadgad	2010	8784	1086	921	756	787	787	8784
3	Dalanzadgad	2011	8760	3309	2652	1759	2394	2394	8760
4	Dalanzadgad	2012	5088	3016	1074	693	1412	1412	5088
5	Dalanzadgad	2013	6096	1809	1766	2479	1240	1240	6096
6	Dalanzadgad	2014	7800	921	843	6068	1482	1482	7800
7	Dalanzadgad	2015	8760	1587	1539	8115	2635	2635	8760
8	Dalanzadgad	2016	6288	1613	1654	5995	3306	3306	6288
9	Sainshand	2009	8688	424	376	423	587	587	8688
10	Sainshand	2010	8784	2577	2557	1113	1210	1210	8784

25 more rows

1.2.0.2 For station

A tibble: 4 × 8

	Station.name	NA_date	NA_PM2	NA_PM10	NA_Vis	NA_WD	NA_WS	NA_OPC
	<chr>	<int>	<int>	<int>	<int>	<int>	<int>	<int>
1	Dalanzadgad	60336	14270	11164	26524	14004	14004	60336
2	Sainshand	59040	11929	11727	9320	8527	8527	59040

```

56 3 UB          76656 8716 7879 3770 4053 4053 43415
57 4 Zamynuud   67392 10075 8880 3444 4960 4960 67392

```

58 1.2.1 By percentages

```

59 # A tibble: 4 × 6
60 # Groups:   Station.name [4]
61   Station.name missing_PM2 missing_PM10 missing_Vis missing_WS missing_WD
62   <chr>          <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
63 1 Dalanzadgad    19.2        25.7        44.5        24.3        24.3
64 2 Sainshand      19.7        20.0        15.7        14.6        14.6
65 3 UB             11.0        11.9         4.53         4.85         4.85
66 4 Zamynuud       12.7        14.4         5.49         7.44         7.44

```

67 1.3 Note that:

68 We use the data in the period of 2009-2018, which has been regarded as a monitor-
69 ing work stabilized since 2008 when is the beginning of the monitoring. According to
70 NIES, site maintenance was consistent up to 2018.

71 +Sainshand site, data 2009-2015 get used; + Dalanzad site: 2009-2016. + UB: 2009-
72 2018 + Zamyn uud: 2009-2018

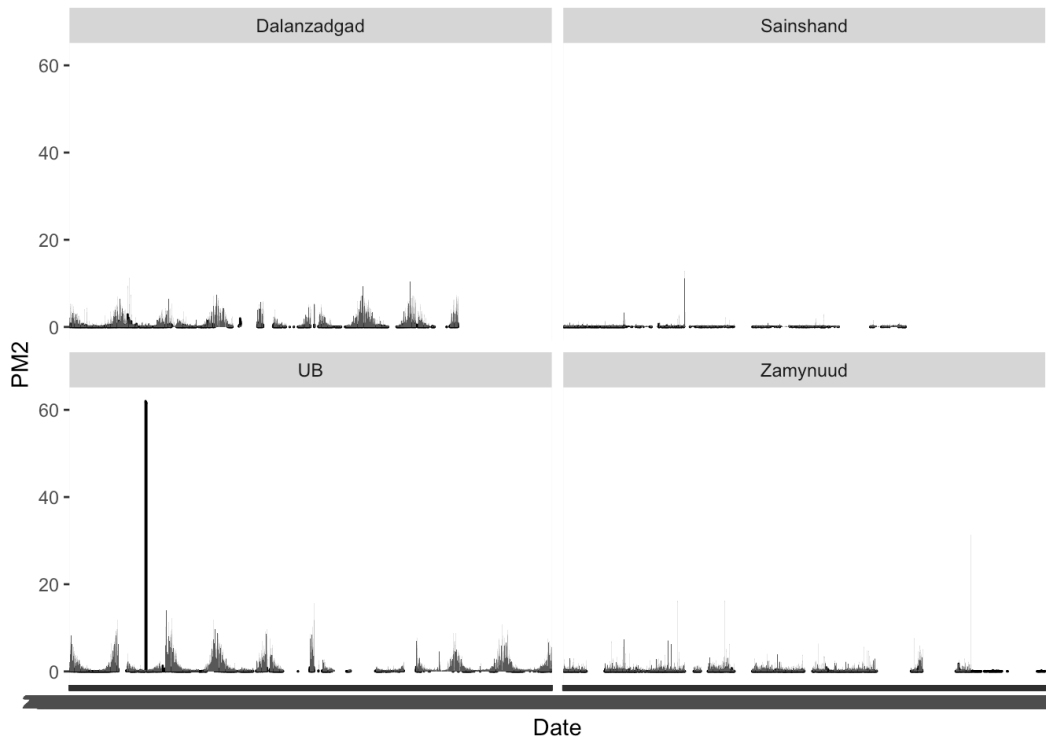
73 2 Remove the spikes, and produce an extended table

74 Remove the spikes in the datasets, and produce the table with NA, with removed
75 spikes; express it in a percentages. #| Comments*kedjdkjk ### Remove the spikes
76 Method 1. Mean value +- (3-5)SD - Find Monthly mean #| flow: 1st - Remove
77 spikes PMs >10 mgm/hour is unreasonable. #| - Exclude 0 values in PMs.

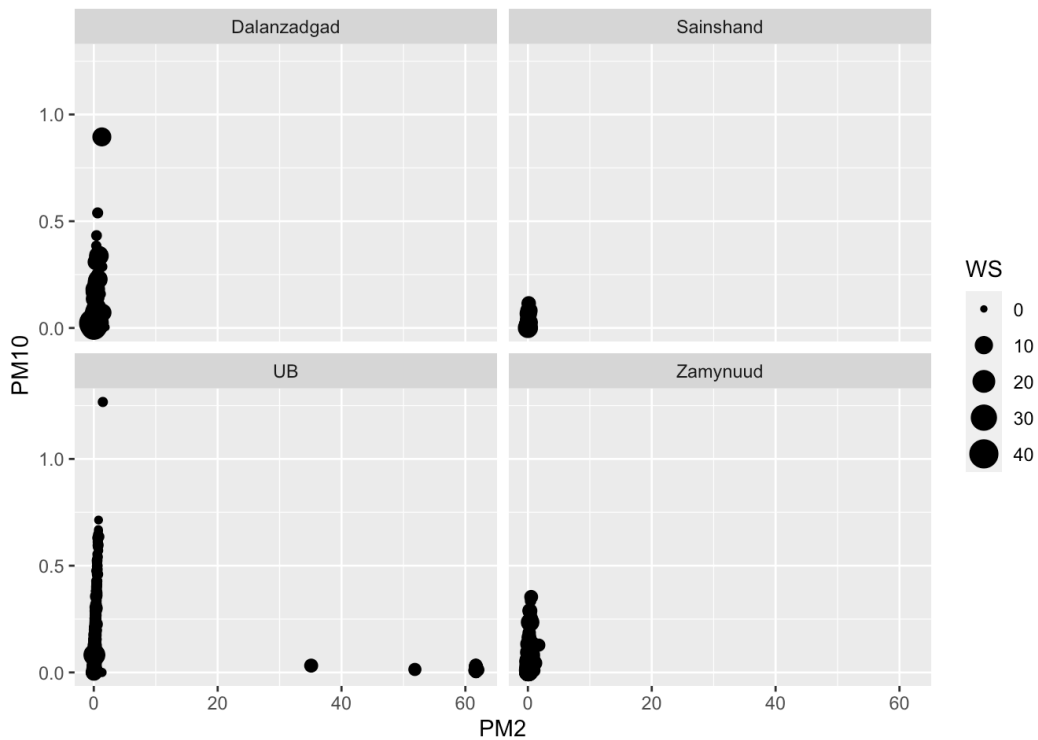
```

78 # A tibble: 6 × 19
79 # Rowwise:
80   Year Month Day Hour correct_PM10 correct_PM2 Visibility WD WS WS.u
81   <int> <int> <int> <int>      <dbl>      <dbl>      <int> <int> <dbl> <dbl>
82 1 2009     1     9    21        0.003        0.002    20000    NA    NA     NA
83 2 2009     1    16    19        0.006        0.001    20000    278   7.14   7.08
84 3 2009     1    16    20        0.005        0.001    20000    276   6.42   6.38
85 4 2009     1    16    22        0.008        0.002    20000    276   7.74   7.71
86 5 2009     1    16    23        0.007        0.002    20000    286   5.55   5.33
87 6 2009     2     5    19        0.006        0.001    20000    274   3.49   3.49
88 # 9 more variables: WS.v <dbl>, OPC <int>, Station.name <chr>, Date <chr>,
89 # PM10 <dbl>, PM2 <dbl>, PM10_rel <dbl>, PM2_rel <dbl>, ratio <dbl>

```



90

91 `#| flow: 2nd - ratio check. PM10 >= PM2.5`

92

93 `#| flow: 2nd - ratio check. remove data of ratio>1`94 `# A tibble: 6 × 19`95 `# Rowwise:`

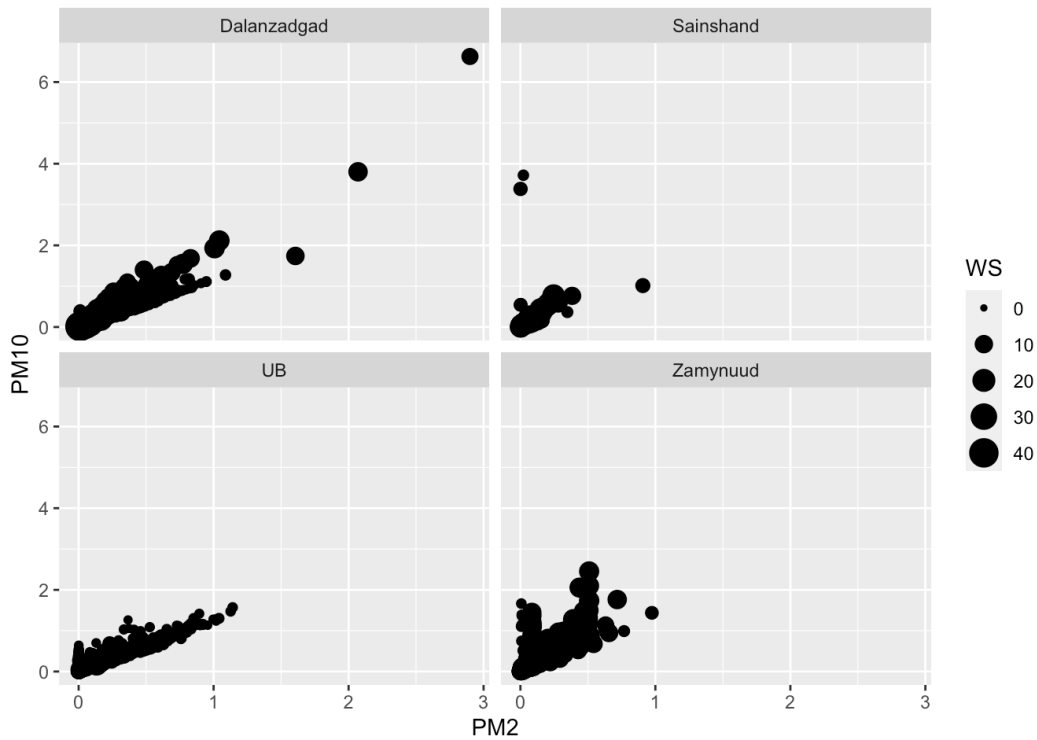
Year	Month	Day	Hour	correct_PM10	correct_PM2	Visibility	WD	WS	WS.u
<int>	<int>	<int>	<int>	<dbl>	<dbl>	<int>	<int>	<dbl>	<dbl>

97

```

98 1 2009 1 1 0 0.087 0.073 3366 252 1.08 1.02
99 2 2009 1 1 1 0.2 0.175 1999 317 0.546 0.374
100 3 2009 1 1 2 0.309 0.266 6756 87 1.03 -1.03
101 4 2009 1 1 3 0.105 0.089 9559 121 1.03 -0.885
102 5 2009 1 1 4 0.063 0.052 17664 121 0.46 -0.393
103 6 2009 1 1 5 0.027 0.021 11348 92 2.05 -2.05
104 # 9 more variables: WS.v <dbl>, OPC <int>, Station.name <chr>, Date <chr>,
105 # PM10 <dbl>, PM2 <dbl>, PM10_rel <dbl>, PM2_rel <dbl>, ratio <dbl>

```



```

106

```

```

107 #| flow: 2nd - Method 1. Mean value +- (3-5)SD - Monthly mean at stations #|
108 output: false

```

```

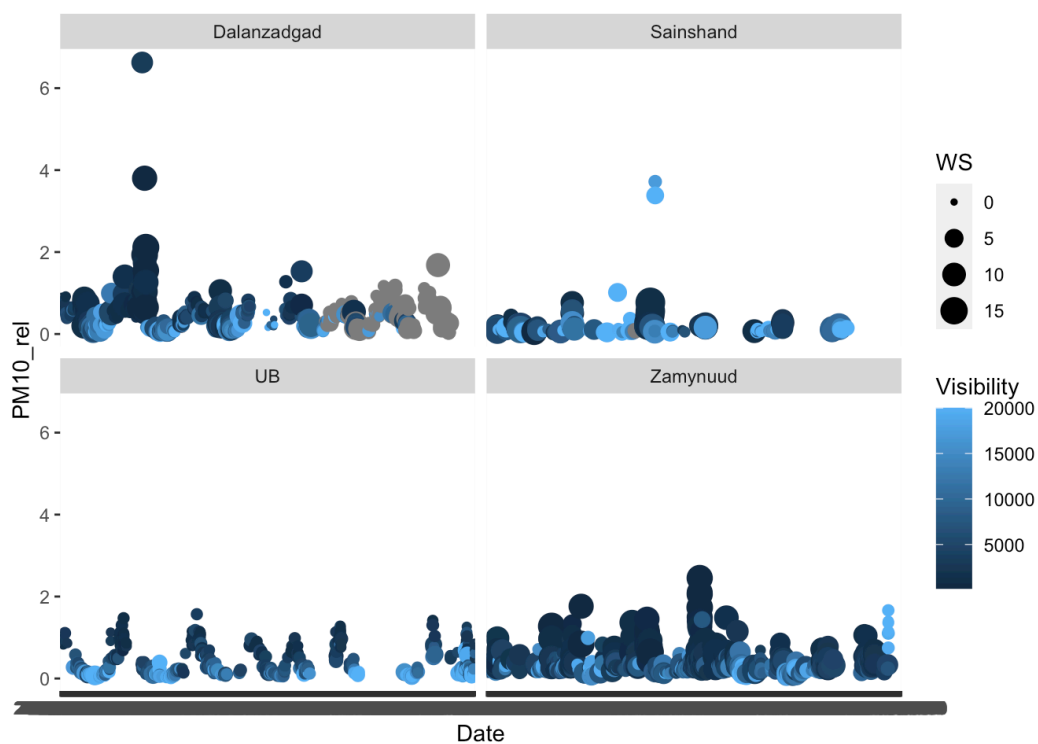
109 Option 1:

```

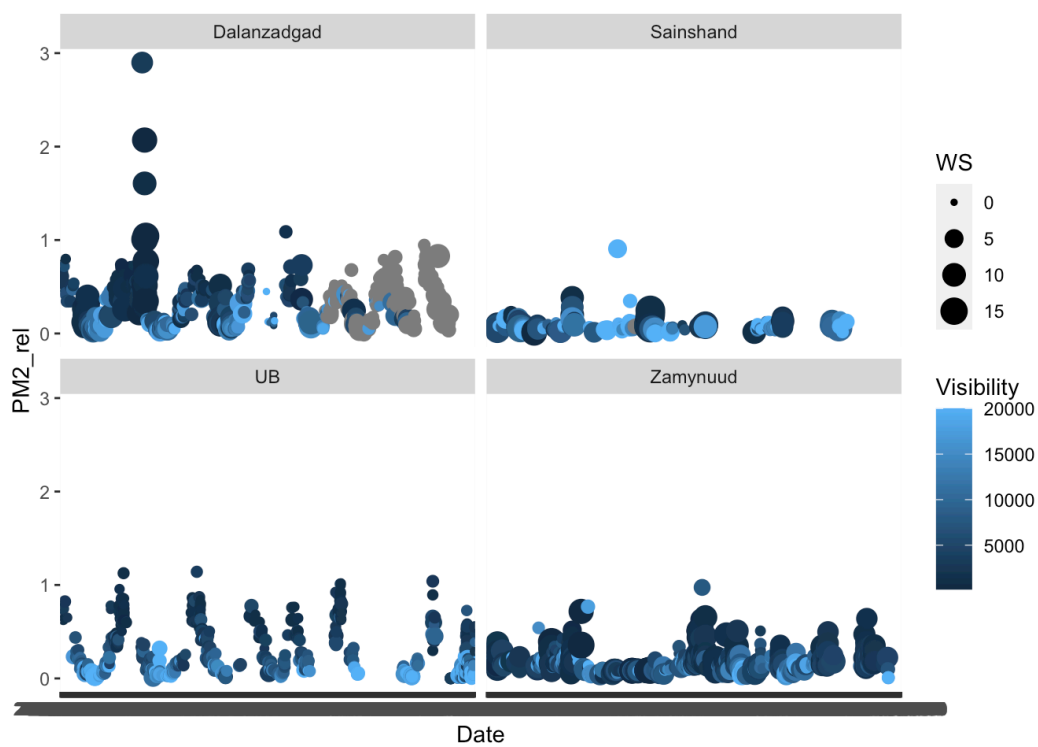
```

110 # A tibble: 2,637 × 19
111 # Groups:   Station.name, Month [48]
112   Year Month Day Hour correct_PM10 correct_PM2 Visibility WD WS
113   <int> <int> <int> <int> <dbl> <dbl> <int> <int> <dbl>
114 1 2009 1 5 16 0.594 0.509 1814 260 0.788
115 2 2009 1 6 0 0.641 0.547 744 248 1.23
116 3 2009 1 6 1 0.867 0.728 1093 277 0.738
117 4 2009 1 6 2 0.705 0.597 1723 0 1.62
118 5 2009 1 9 1 0.756 0.656 992 292 0.429
119 6 2009 1 9 2 0.709 0.612 2098 55 1
120 7 2009 1 15 12 0.818 0.718 5941 147 0.633
121 8 2009 1 16 1 0.742 0.556 2125 120 0.274
122 9 2009 1 30 1 0.915 0.797 1315 9 0.366
123 10 2009 1 30 2 0.601 0.507 2485 309 0.713
124 # 2,627 more rows
125 # 10 more variables: WS.u <dbl>, WS.v <dbl>, OPC <int>, Station.name <chr>,
126 # Date <chr>, PM10 <chr>, PM2 <chr>, PM10_rel <dbl>, PM2_rel <dbl>,
127 # ratio <dbl>

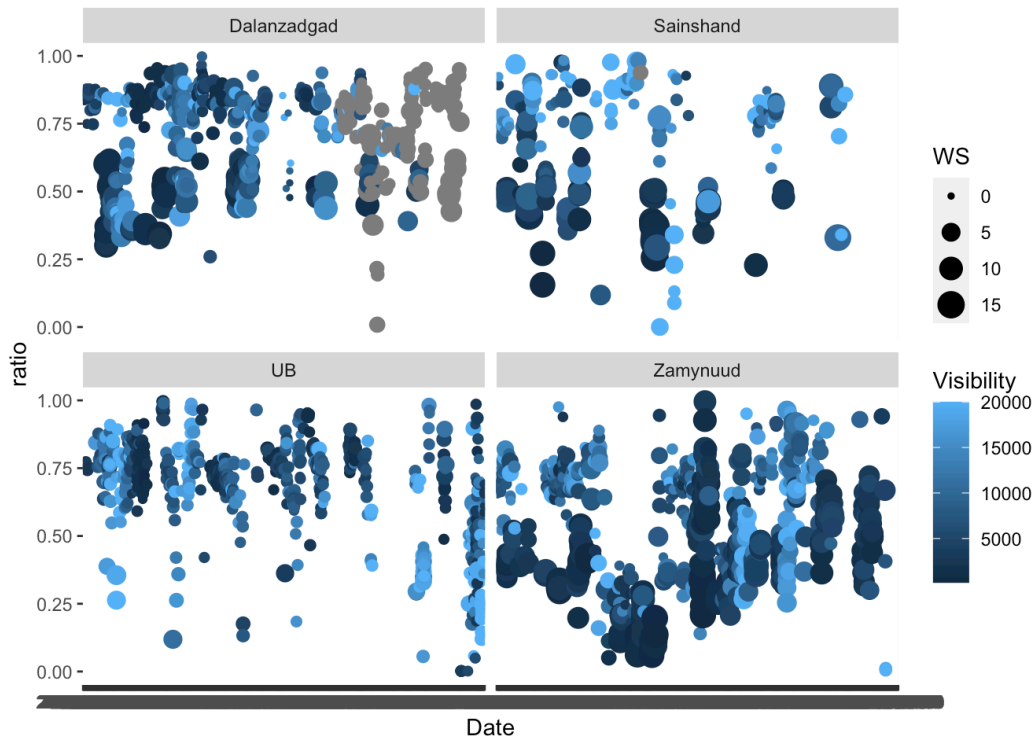
```



128



129



130

131 Option 2:

```

132 # A tibble: 118,482 × 19
133 # Groups:   Station.name, Month [48]
134   Year Month Day Hour correct_PM10 correct_PM2 Visibility WD WS
135   <int> <int> <int> <int>      <dbl>      <dbl>      <int> <int> <dbl>
136 1 2009     1     1     1     0.087     0.073     3366  252 1.08
137 2 2009     1     1     1     0.2      0.175     1999  317 0.546
138 3 2009     1     1     2     0.309     0.266     6756   87 1.03
139 4 2009     1     1     3     0.105     0.089     9559  121 1.03
140 5 2009     1     1     4     0.063     0.052    17664  121 0.46
141 6 2009     1     1     5     0.027     0.021    11348   92 2.05
142 7 2009     1     1     6     0.065     0.055    15397    0 1.6
143 8 2009     1     1     7     0.03      0.023    20000  300 1.74
144 9 2009     1     1     8     0.013     0.01     17384  305 1.29
145 10 2009     1     1     9     0.029     0.023     9322  324 2.96
146 # 118,472 more rows
147 # 10 more variables: WS.u <dbl>, WS.v <dbl>, OPC <int>, Station.name <chr>,
148 # Date <chr>, PM10 <dbl>, PM2 <dbl>, PM10_rel <dbl>, PM2_rel <dbl>,
149 # ratio <dbl>

```

```

150 # A tibble: 3,267 × 19
151   Year Month Day Hour correct_PM10 correct_PM2 Visibility WD WS
152   <int> <int> <int> <int>      <dbl>      <dbl>      <int> <int> <dbl>
153 1 2009     1     1     2     0.309     0.266     6756   87 1.03
154 2 2009     1     3    15     0.351     0.292     3444  119 0.856
155 3 2009     1     5     1     0.313     0.271     3392  275 2.56
156 4 2009     1     5    13     0.482     0.419     1383  260 1.7
157 5 2009     1     5    14     0.482     0.415     1072  266 1.84
158 6 2009     1     5    15     0.549     0.466     1099  261 0.83
159 7 2009     1     5    16     0.594     0.509     1814  260 0.788
160 8 2009     1     6     0     0.641     0.547      744  248 1.23

```

```

161 9 2009 1 6 1 0.867 0.728 1093 277 0.738
162 10 2009 1 6 2 0.705 0.597 1723 0 1.62
163 # 3,257 more rows
164 # 10 more variables: WS.u <dbl>, WS.v <dbl>, OPC <int>, Station.name <chr>,
165 # Date <chr>, PM10 <chr>, PM2 <chr>, PM10_rel <dbl>, PM2_rel <dbl>,
166 # ratio <dbl>
167 # A tibble: 0 × 19
168 # 19 variables: Year <int>, Month <int>, Day <int>, Hour <int>,
169 # correct_PM10 <dbl>, correct_PM2 <dbl>, Visibility <int>, WD <int>,
170 # WS <dbl>, WS.u <dbl>, WS.v <dbl>, OPC <int>, Station.name <chr>,
171 # Date <chr>, PM10 <dbl>, PM2 <dbl>, PM10_rel <dbl>, PM2_rel <dbl>,
172 # ratio <dbl>

```

173 **2.1 Save dataset in folder: 01_data_raw**

174 **3 Tidy data**

175 **3.1 Fill the missing data**

176 Method 1. Fill the gap Method 2. Relationship equation Method 3. Look-up table

177 **3.2 Save dataset in folder: 02_data_tidy**

178 Source: [01_datawork](#)

179 **3.3 Introduction**

180 Source: [Article Notebook](#)

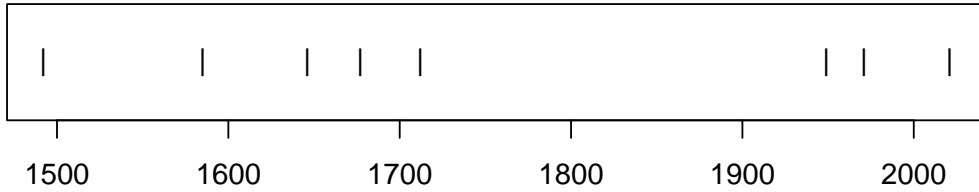


Figure 1: Timeline of recent earthquakes on La Palma

181 Source: [Article Notebook](#)

182 Source: [Article Notebook](#)

183 Based on data up to and including 1971, eruptions on La Palma happen every 79.8
184 years on average.

185 Studies of the magma systems feeding the volcano, such as Marrero et al. (2019),
186 have proposed that there are two main magma reservoirs feeding the Cumbre Vieja
187 volcano; one in the mantle (30-40km depth) which charges and in turn feeds a shal-
188 lower crustal reservoir (10-20km depth).

189 Eight eruptions have been recorded since the late 1400s (Figure 1).

190 Data and methods are discussed in Section 0.1.

191 Let x denote the number of eruptions in a year. Then, x can be modeled by a Pois-
192 son distribution

$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad (1)$$

193 where λ is the rate of eruptions per year. Using Equation 1, the probability of an
 194 eruption in the next t years can be calculated.

Table 1: Recent historic eruptions on La Palma

Name	Year
Current	2021
Teneguía	1971
Nambroque	1949
El Charco	1712
Volcán San Antonio	1677
Volcán San Martin	1646
Tajuya near El Paso	1585
Montaña Quemada	1492

195 Table 1 summarises the eruptions recorded since the colonization of the islands by
 196 Europeans in the late 1400s.

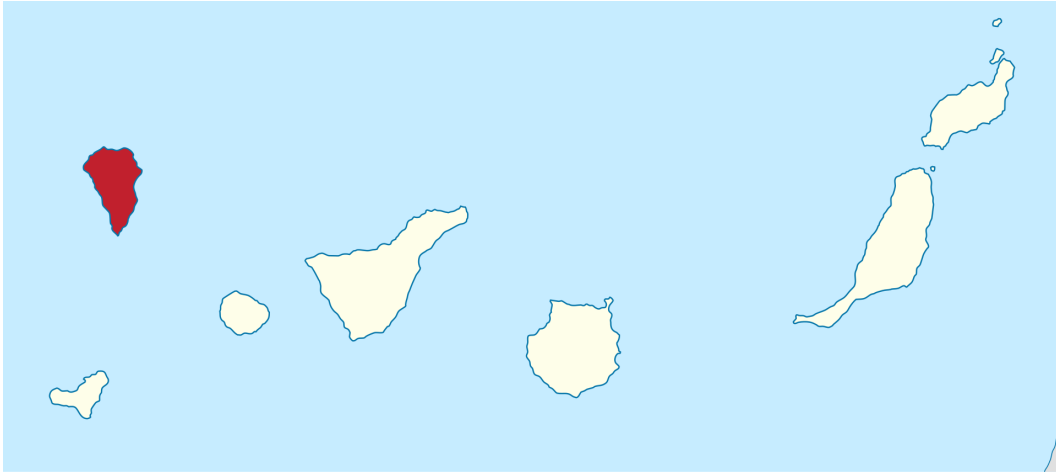


Figure 2: Map of La Palma

197 La Palma is one of the west most islands in the Volcanic Archipelago of the Canary
 198 Islands (Figure 2).

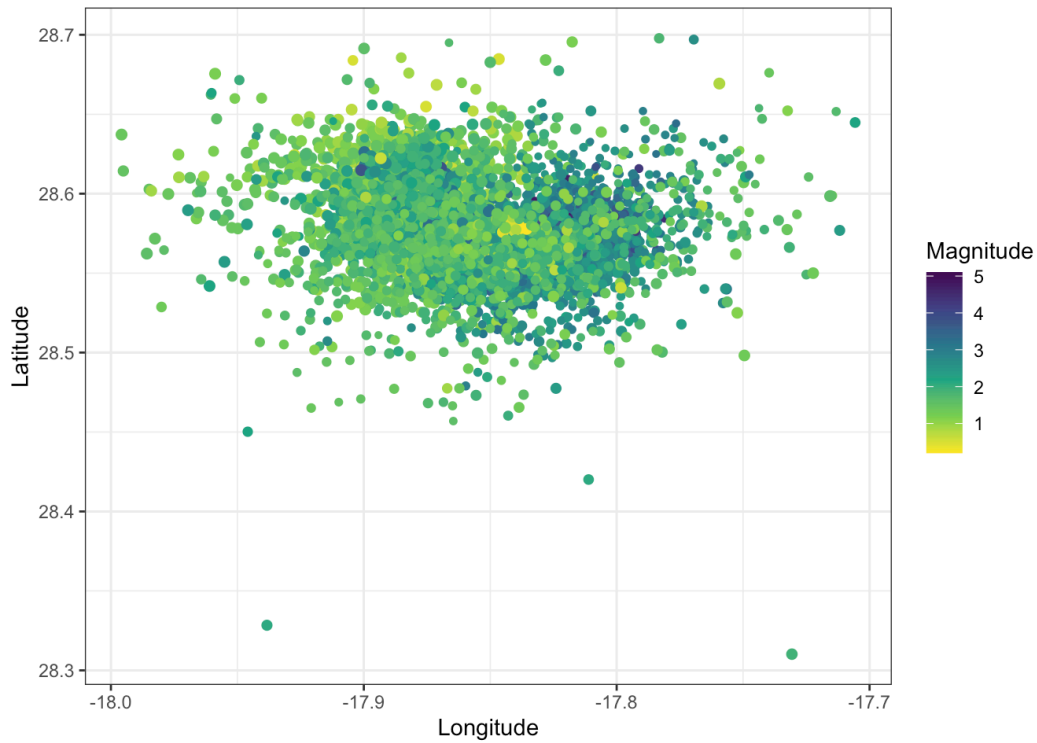


Figure 3: Locations of earthquakes on La Palma since 2017

199 Source: [Explore Earthquakes](#)

200 kk

201 4 Explore Earthquakes

202 Munkhtsetseg

203 Library

204 5 Import the dataset and remove the duplicates

205 Import the dataset from the directory of: ~/Data Input/Preprocessing data/Preprocessing
206 data.csv, assign the dataset as object of df:

207 Remove the duplicates with the function of distinct(), assign the dataset as df_01:

208 5.1 Produce a table with missing data

209 For date options as year, month, etc:

210 # A tibble: 52 × 9

211 # Groups: Station.name [4]

	Station.name	Year	NA_date	NA_PM2	NA_PM10	NA_Vis	NA_WD	NA_WS	NA_OPC
	<chr>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>
213	1 Dalanzadgad	2008	4630	1543	1672	1463	1566	1566	4630
214	2 Dalanzadgad	2009	8760	715	929	659	748	748	8760
215	3 Dalanzadgad	2010	8784	921	1086	756	787	787	8784
216	4 Dalanzadgad	2011	8760	2652	3309	1759	2394	2394	8760
217	5 Dalanzadgad	2012	5088	1074	3016	693	1412	1412	5088
218	6 Dalanzadgad	2013	6096	1766	1809	2479	1240	1240	6096
219	7 Dalanzadgad	2014	7800	843	921	6068	1482	1482	7800

```

221     8 Dalanzadgad    2015    8760    1539    1587    8115    2635    2635    8760
222     9 Dalanzadgad    2016    6288    1654    1613    5995    3306    3306    6288
223    10 Dalanzadgad    2017    3264     36     45    3264    3264    3264    3264
224    # 42 more rows

```

225 For station

```

226    # A tibble: 4 × 8
227      Station.name NA_date NA_PM2 NA_PM10 NA_Vis NA_WD NA_WS NA_OPC
228      <chr>         <int> <int> <int> <int> <int> <int> <int>
229    1 Dalanzadgad    69454 13081 16327 32475 20058 20058 69454
230    2 Sainshand     101230 27588 36117 28986 13768 13768 101230
231    3 UB            95662  7895  8785  3775  4121  4121  62421
232    4 Zamynnuud     99742 32281 33597 22525  5373  5373  99742

```

233 By percentages

```

234    # A tibble: 4 × 2
235    # Groups:   Station.name [4]
236      Station.name    sdq
237      <chr>         <dbl>
238    1 Dalanzadgad    10.7
239    2 Sainshand      25.9
240    3 UB             17.9
241    4 Zamynnuud      39.6

```

242 Note that the `echo = FALSE` parameter was added to the code chunk to prevent
 243 printing of the R code that generated the plot.

244 **6 Remove the spikes, and produce an extended table**

245 Remove the spikes in the datasets, and produce the table with NA, with removed
 246 spikes; express it in a percentages.

247 **6.0.1 Remove the spikes Method 1. Mean value $\pm (3-5)SD$**

248 Method 2. Seasonal variations, and trend-mean

249 **6.1 Save dataset in folder: 01_data_raw**

250 **7 Tidy data**

251 **7.1 Fill the missing data**

252 Method 1. Fill the gap Method 2. Relationship equation Method 3. Look-up table

253 **7.2 Save dataset in folder: 02_data_tidy**

254 Read a clean version of data:

255 Create spatial plot:

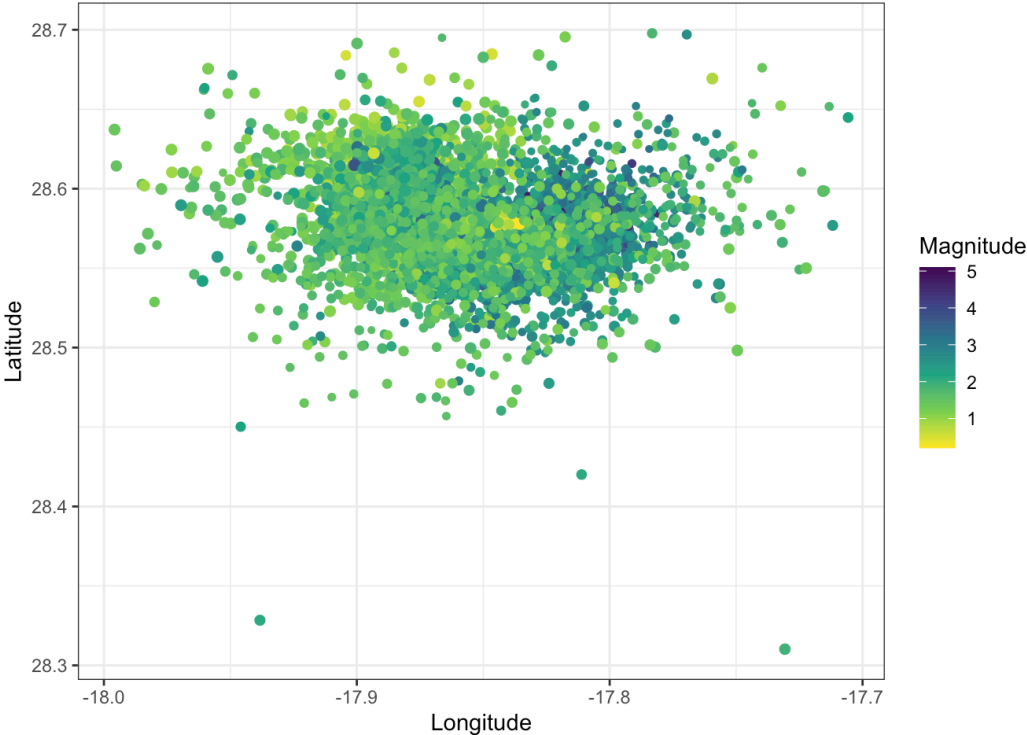


Figure 4: Locations of earthquakes on La Palma since 2017

Source: [Explore Earthquakes](#)

Figure 4 shows the location of recent Earthquakes on La Palma.

7.3 Results

7.4 Discussion

7.5 Conclusions

References

Marrero, J., García, A., Berrocoso, M., Llinares, Á., Rodríguez-Losada, A., & Ortiz, R. (2019). Strategies for the development of volcanic hazard maps in monogenetic volcanic fields: The example of La Palma (Canary Islands). *Journal of Applied Volcanology*, 8. <https://doi.org/10.1186/s13617-019-0085-5>