Temporal and spatial variations of PM2 and PM10 concentrations in Mongolia

Erdenebayar Munkhtsetseg

Atsushi Shimizu

2024-01-26

Abstract

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   
# ℹ 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  
# ℹ 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  
3 UB 76656 8716 7879 3770 4053 4053 43415  
4 Zamynuud 67392 10075 8880 3444 4960 4960 67392

### 1.2.1 By percentages

# A tibble: 4 × 6  
# Groups: Station.name [4]  
 Station.name missing\_PM2 missing\_PM10 missing\_Vis missing\_WS missing\_WD  
 <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
1 Dalanzadgad 19.2 25.7 44.5 24.3 24.3   
2 Sainshand 19.7 20.0 15.7 14.6 14.6   
3 UB 11.0 11.9 4.53 4.85 4.85  
4 Zamynuud 12.7 14.4 5.49 7.44 7.44

## 1.3 Note that:

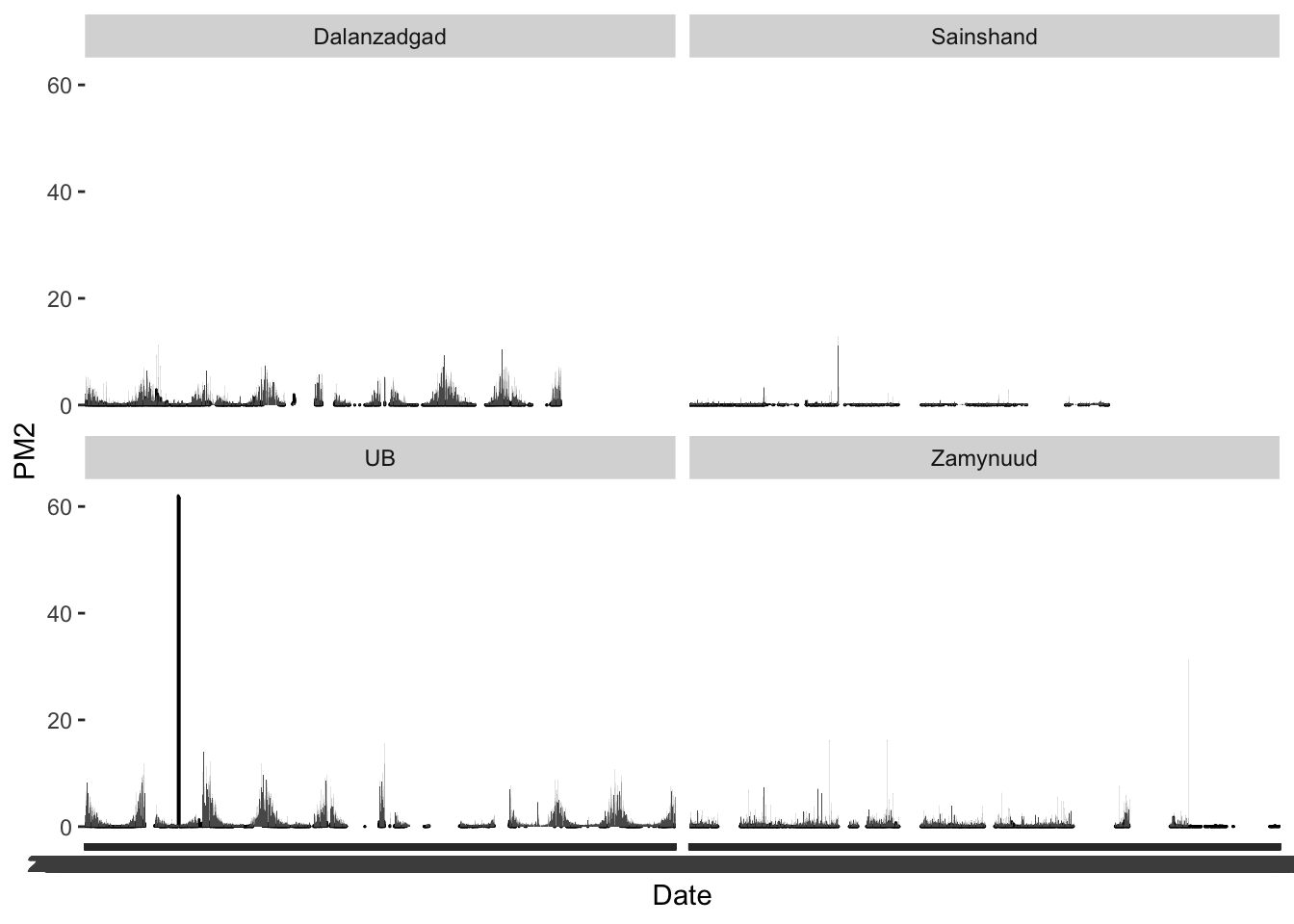
We use the data in the period of 2009-2018, which has been regarded as a monitoring work stabilized since 2008 when is the beginning of the monitoring. According to NIES, site maintenance was consistent up to 2018.

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

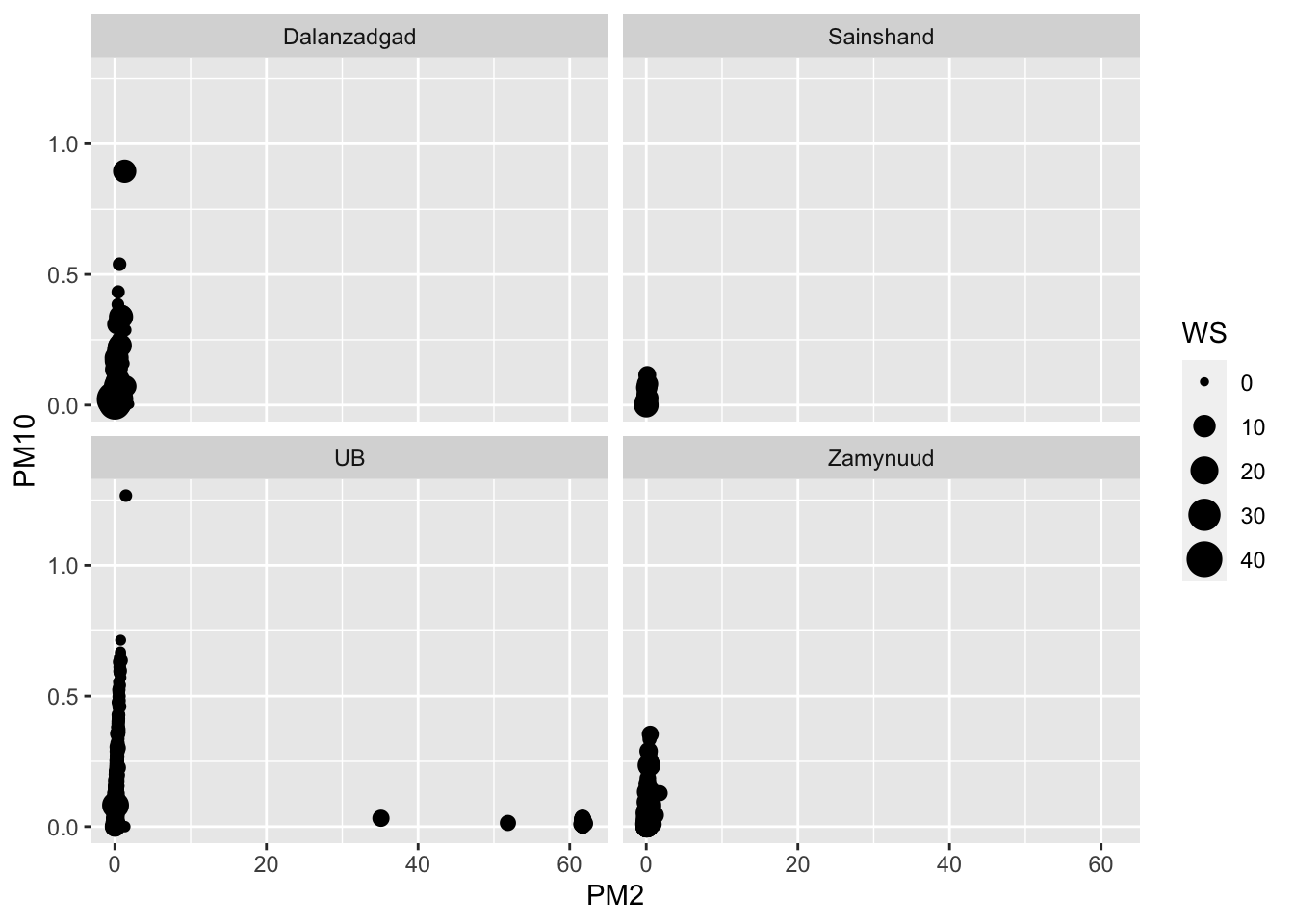
# 2. Remove the spikes, and produce an extended table

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

# 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 9 21 0.003 0.002 20000 NA NA NA   
2 2009 1 16 19 0.006 0.001 20000 278 7.14 7.08  
3 2009 1 16 20 0.005 0.001 20000 276 6.42 6.38  
4 2009 1 16 22 0.008 0.002 20000 276 7.74 7.71  
5 2009 1 16 23 0.007 0.002 20000 286 5.55 5.33  
6 2009 2 5 19 0.006 0.001 20000 274 3.49 3.49  
# ℹ 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>

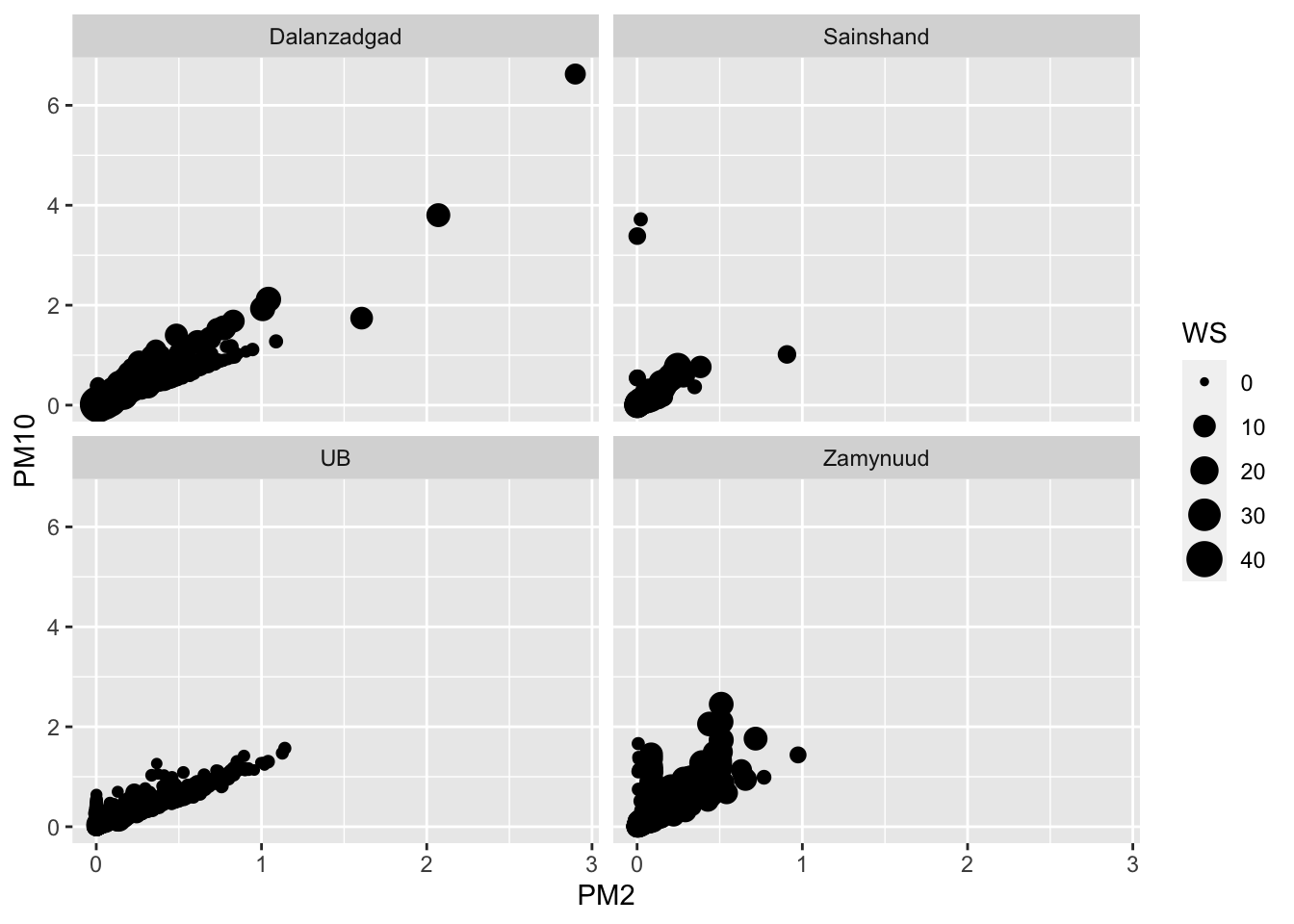


#| flow: 2nd - ratio check. PM10 >= PM2.5



#| flow: 2nd - ratio check. remove data of ratio>1

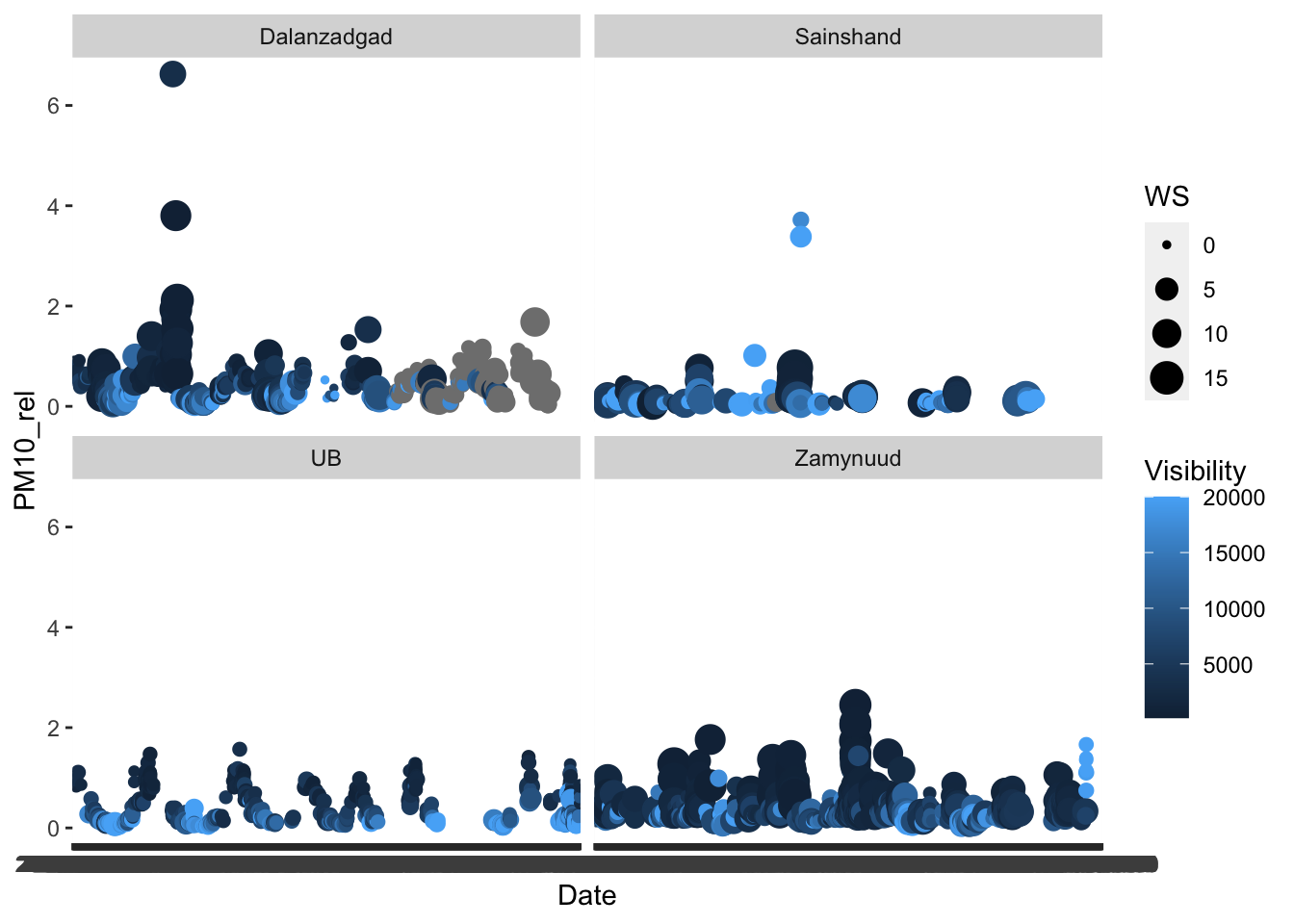
# A tibble: 6 × 19  
# Rowwise:   
 Year Month Day Hour correct\_PM10 correct\_PM2 Visibility WD WS WS.u  
 <int> <int> <int> <int> <dbl> <dbl> <int> <int> <dbl> <dbl>  
1 2009 1 1 0 0.087 0.073 3366 252 1.08 1.02   
2 2009 1 1 1 0.2 0.175 1999 317 0.546 0.374  
3 2009 1 1 2 0.309 0.266 6756 87 1.03 -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   
# ℹ 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>

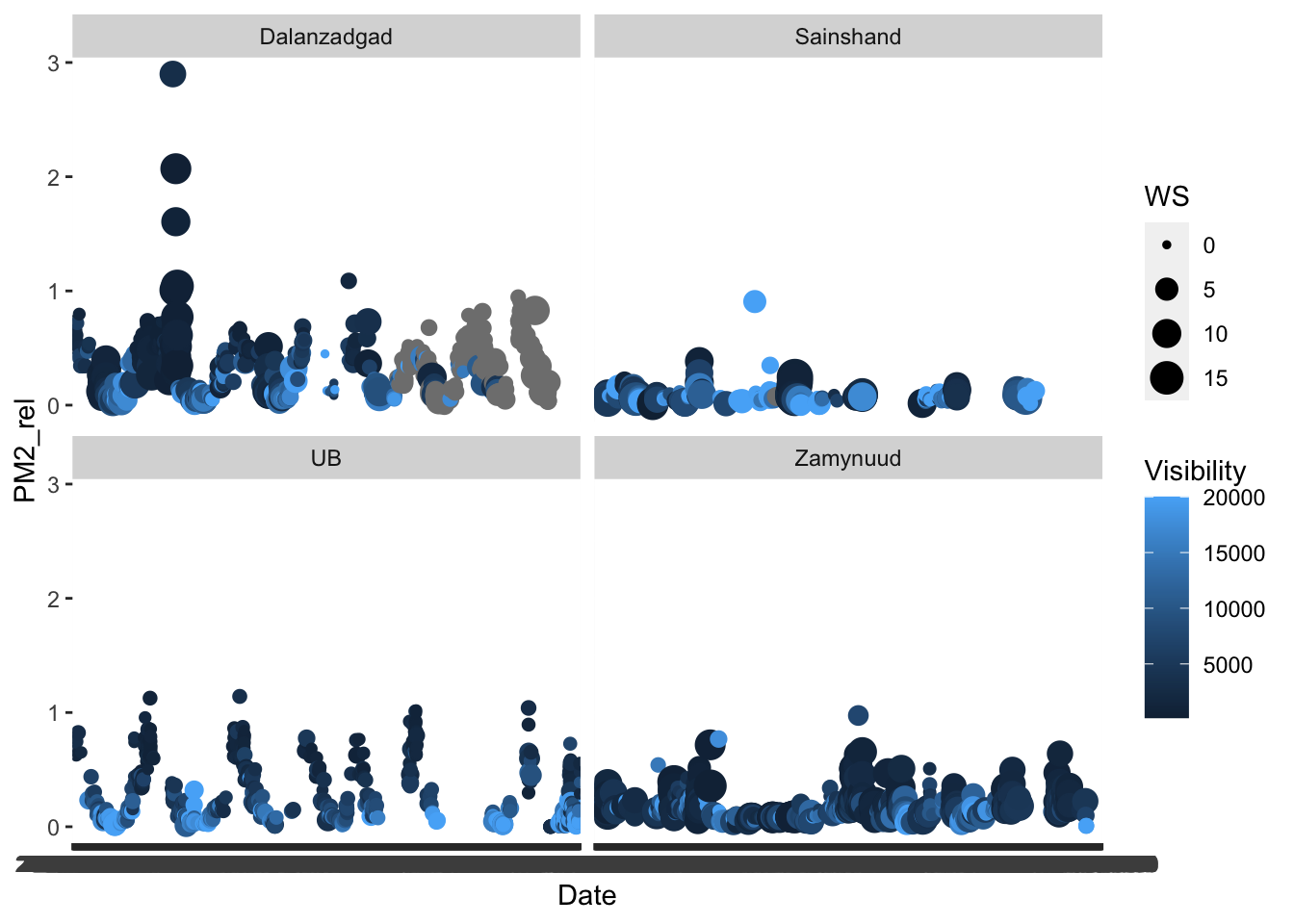


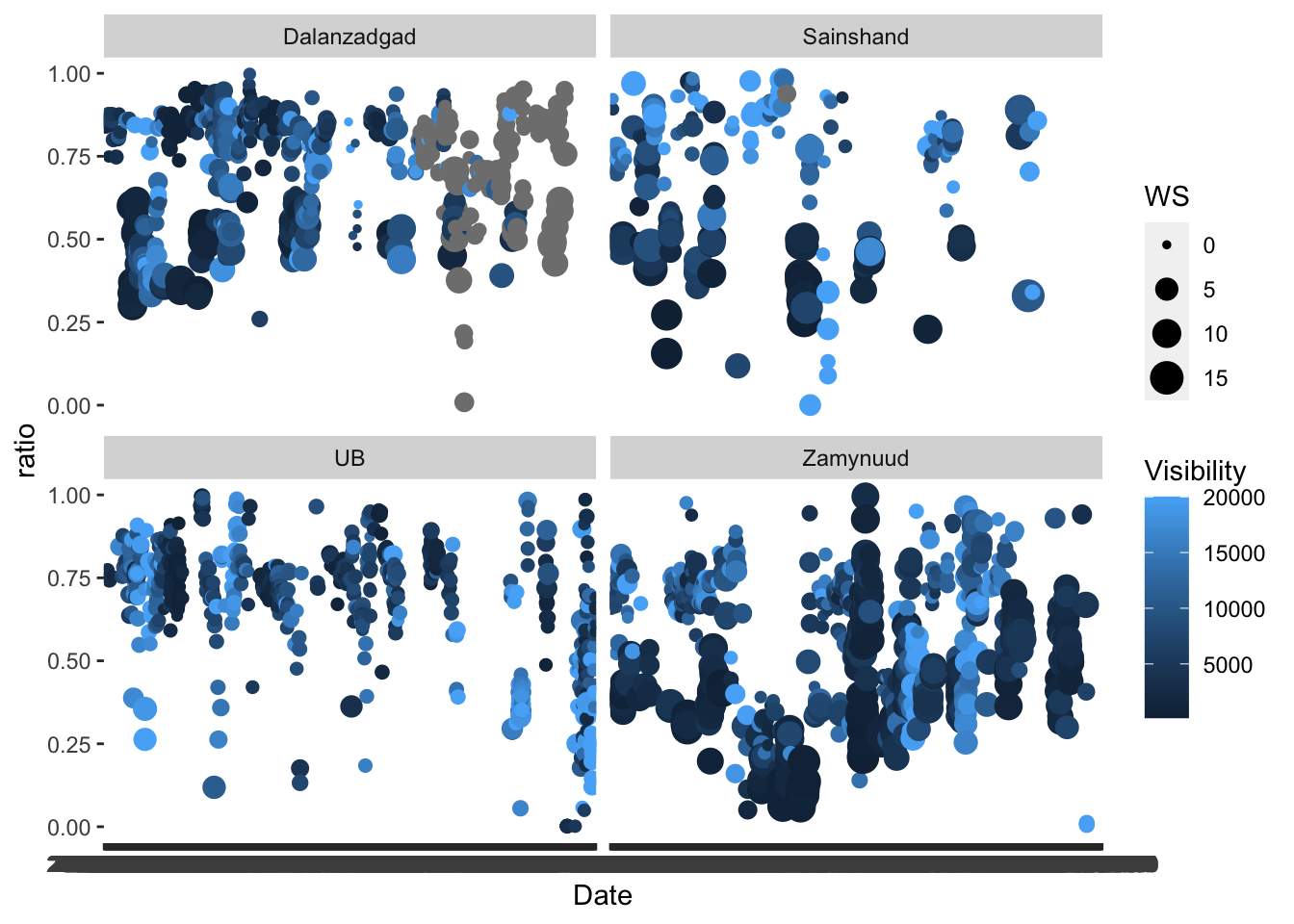
#| flow: 2nd - Method 1. Mean value +- (3-5)SD - Monthly mean at stations #| output: false

Option 1:

# A tibble: 2,637 × 19  
# Groups: Station.name, Month [48]  
 Year Month Day Hour correct\_PM10 correct\_PM2 Visibility WD WS  
 <int> <int> <int> <int> <dbl> <dbl> <int> <int> <dbl>  
 1 2009 1 5 16 0.594 0.509 1814 260 0.788  
 2 2009 1 6 0 0.641 0.547 744 248 1.23   
 3 2009 1 6 1 0.867 0.728 1093 277 0.738  
 4 2009 1 6 2 0.705 0.597 1723 0 1.62   
 5 2009 1 9 1 0.756 0.656 992 292 0.429  
 6 2009 1 9 2 0.709 0.612 2098 55 1   
 7 2009 1 15 12 0.818 0.718 5941 147 0.633  
 8 2009 1 16 1 0.742 0.556 2125 120 0.274  
 9 2009 1 30 1 0.915 0.797 1315 9 0.366  
10 2009 1 30 2 0.601 0.507 2485 309 0.713  
# ℹ 2,627 more rows  
# ℹ 10 more variables: WS.u <dbl>, WS.v <dbl>, OPC <int>, Station.name <chr>,  
# Date <chr>, PM10 <chr>, PM2 <chr>, PM10\_rel <dbl>, PM2\_rel <dbl>,  
# ratio <dbl>







Option 2:

# A tibble: 118,482 × 19  
# Groups: Station.name, Month [48]  
 Year Month Day Hour correct\_PM10 correct\_PM2 Visibility WD WS  
 <int> <int> <int> <int> <dbl> <dbl> <int> <int> <dbl>  
 1 2009 1 1 0 0.087 0.073 3366 252 1.08   
 2 2009 1 1 1 0.2 0.175 1999 317 0.546  
 3 2009 1 1 2 0.309 0.266 6756 87 1.03   
 4 2009 1 1 3 0.105 0.089 9559 121 1.03   
 5 2009 1 1 4 0.063 0.052 17664 121 0.46   
 6 2009 1 1 5 0.027 0.021 11348 92 2.05   
 7 2009 1 1 6 0.065 0.055 15397 0 1.6   
 8 2009 1 1 7 0.03 0.023 20000 300 1.74   
 9 2009 1 1 8 0.013 0.01 17384 305 1.29   
10 2009 1 1 9 0.029 0.023 9322 324 2.96   
# ℹ 118,472 more rows  
# ℹ 10 more variables: WS.u <dbl>, WS.v <dbl>, OPC <int>, Station.name <chr>,  
# Date <chr>, PM10 <dbl>, PM2 <dbl>, PM10\_rel <dbl>, PM2\_rel <dbl>,  
# ratio <dbl>

# A tibble: 3,267 × 19  
 Year Month Day Hour correct\_PM10 correct\_PM2 Visibility WD WS  
 <int> <int> <int> <int> <dbl> <dbl> <int> <int> <dbl>  
 1 2009 1 1 2 0.309 0.266 6756 87 1.03   
 2 2009 1 3 15 0.351 0.292 3444 119 0.856  
 3 2009 1 5 1 0.313 0.271 3392 275 2.56   
 4 2009 1 5 13 0.482 0.419 1383 260 1.7   
 5 2009 1 5 14 0.482 0.415 1072 266 1.84   
 6 2009 1 5 15 0.549 0.466 1099 261 0.83   
 7 2009 1 5 16 0.594 0.509 1814 260 0.788  
 8 2009 1 6 0 0.641 0.547 744 248 1.23   
 9 2009 1 6 1 0.867 0.728 1093 277 0.738  
10 2009 1 6 2 0.705 0.597 1723 0 1.62   
# ℹ 3,257 more rows  
# ℹ 10 more variables: WS.u <dbl>, WS.v <dbl>, OPC <int>, Station.name <chr>,  
# Date <chr>, PM10 <chr>, PM2 <chr>, PM10\_rel <dbl>, PM2\_rel <dbl>,  
# ratio <dbl>

# A tibble: 0 × 19  
# ℹ 19 variables: Year <int>, Month <int>, Day <int>, Hour <int>,  
# correct\_PM10 <dbl>, correct\_PM2 <dbl>, Visibility <int>, WD <int>,  
# WS <dbl>, WS.u <dbl>, WS.v <dbl>, OPC <int>, Station.name <chr>,  
# Date <chr>, PM10 <dbl>, PM2 <dbl>, PM10\_rel <dbl>, PM2\_rel <dbl>,  
# ratio <dbl>

## 2.1 Save dataset in folder: 01\_data\_raw

# 3. Tidy data

## 3.1 Fill the missing data

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

## 3.2 Save dataset in folder: 02\_data\_tidy

Source: [01\_datawork](https://EmouAcademy.github.io/my-awesome-manuscripts/notebooks/01_datawork-preview.html#c335df36-7197-45eb-a1d9-0a90e5099ce8)

## 3.3 Introduction

Source: [Article Notebook](https://EmouAcademy.github.io/my-awesome-manuscripts/index.qmd.html)

|  |
| --- |
| Figure 1: Timeline of recent earthquakes on La Palma |

Source: [Article Notebook](https://EmouAcademy.github.io/my-awesome-manuscripts/index.qmd.html)

Source: [Article Notebook](https://EmouAcademy.github.io/my-awesome-manuscripts/index.qmd.html)

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

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

Eight eruptions have been recorded since the late 1400s ([Figure 1](#fig-timeline)).

Data and methods are discussed in [Section 0.1](#sec-data-methods).

Let denote the number of eruptions in a year. Then, can be modeled by a Poisson distribution

where is the rate of eruptions per year. Using [Equation 1](#eq-poisson), the probability of an eruption in the next 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 | |

[Table 1](#tbl-history) summarises the eruptions recorded since the colonization of the islands by Europeans in the late 1400s.

|  |
| --- |
| Figure 2: Map of La Palma |

La Palma is one of the west most islands in the Volcanic Archipelago of the Canary Islands ([Figure 2](#fig-map)).

|  |
| --- |
| Figure 3: Locations of earthquakes on La Palma since 2017 |

Source: [Explore Earthquakes](https://EmouAcademy.github.io/my-awesome-manuscripts/notebooks/explore-earthquakes-preview.html#cell-fig-spatial-plot)

kk

# 4. Explore Earthquakes

Munkhtsetseg

Library

# 5. 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:

## 5.1 Produce a table with missing data

For date options as year, month, etc:

# A tibble: 52 × 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 2008 4630 1543 1672 1463 1566 1566 4630  
 2 Dalanzadgad 2009 8760 715 929 659 748 748 8760  
 3 Dalanzadgad 2010 8784 921 1086 756 787 787 8784  
 4 Dalanzadgad 2011 8760 2652 3309 1759 2394 2394 8760  
 5 Dalanzadgad 2012 5088 1074 3016 693 1412 1412 5088  
 6 Dalanzadgad 2013 6096 1766 1809 2479 1240 1240 6096  
 7 Dalanzadgad 2014 7800 843 921 6068 1482 1482 7800  
 8 Dalanzadgad 2015 8760 1539 1587 8115 2635 2635 8760  
 9 Dalanzadgad 2016 6288 1654 1613 5995 3306 3306 6288  
10 Dalanzadgad 2017 3264 36 45 3264 3264 3264 3264  
# ℹ 42 more rows

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 69454 13081 16327 32475 20058 20058 69454  
2 Sainshand 101230 27588 36117 28986 13768 13768 101230  
3 UB 95662 7895 8785 3775 4121 4121 62421  
4 Zamynuud 99742 32281 33597 22525 5373 5373 99742

By percentages

# A tibble: 4 × 2  
# Groups: Station.name [4]  
 Station.name sdq  
 <chr> <dbl>  
1 Dalanzadgad 10.7  
2 Sainshand 25.9  
3 UB 17.9  
4 Zamynuud 39.6

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

# 6. Remove the spikes, and produce an extended table

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

### 6.0.1 Remove the spikes Method 1. Mean value +- (3-5)SD

Method 2. Seasonal variations, and trend-mean

## 6.1 Save dataset in folder: 01\_data\_raw

# 7. Tidy data

## 7.1 Fill the missing data

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

## 7.2 Save dataset in folder: 02\_data\_tidy

Read a clean version of data:

Create spatial plot:

|  |
| --- |
| Figure 4: Locations of earthquakes on La Palma since 2017 |

Source: [Explore Earthquakes](https://EmouAcademy.github.io/my-awesome-manuscripts/notebooks/01-earthquakes-preview.html#041ca1bf-04ed-4a8f-986d-dcdf65a9ec4d)

[Figure 4](#fig-spatial-plot) shows the location of recent Earthquakes on La Palma.

## 7.3 Results

## 7.4 Discussion

## 7.5 Conclusions

## References

Marrero, José, Alicia García, Manuel Berrocoso, Ángeles Llinares, Antonio Rodríguez-Losada, and R. Ortiz. 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 (July). <https://doi.org/10.1186/s13617-019-0085-5>.