

Taller 2 - Gestión de datos

Calidad de aire - Estación La Flora Santiago de Cali

Ronald Fernando Rodríguez Barbosa

Maestría en Ingeniería de Sistemas y Computación

Maestría en Analítica para la Inteligencia de Negocios

Pontificia Universidad Javeriana

18 de Mayo de 2019

1. Introducción

Contextualización

La estación La Flora ubicada en el norte de la ciudad de Santiago de Cali, es una estación automática que reporta información horaria al centro de control del DAGMA. Mide los niveles de Material Particulado Menor a 10 micrómetros (PM10), Dióxido de Azufre (SO2), Dióxido de Nitrógeno (NO2), Monóxido de Carbono (CO), Ozono (O3) y variables meteorológicas como velocidad del viento, dirección del viento, temperatura, humedad, radiación solar y precipitación.

La estación Escuela República de Argentina ubicada en el centro de la ciudad de Santiago de Cali, es una estación automática que reporta información horaria al centro de control del DAGMA. Mide los niveles de Material Particulado Menor a 10 micrómetros (PM10), Dióxido de Azufre (SO2), Dióxido de Nitrógeno (NO2), Monóxido de Carbono (CO) y variables meteorológicas como velocidad del viento, dirección del viento, temperatura, humedad, radiación solar y precipitación. El DAGMA mediante el grupo de calidad del aire en el transcurso de los años ha trabajado en robustecer el sistema de calidad del aire para garantizar la continuidad en los datos, sin embargo es normal que los sistemas de monitoreo de calidad del aire presenten discontinuidad en los datos (datos faltantes o espacios en blanco) debido a dos situaciones: la primera corresponde a las anomalías que se dan en las estaciones de monitoreo, tales como: Fallas en los equipos, falta de energía eléctrica en la zona, hurto de equipos o cableado, mantenimiento o cambio de equipos, etc y la segunda causa corresponde a la inclusión o exclusión de algunos contaminantes o variables meteorológicas, según criterio de los expertos y característica de la zona a monitorear.

Para conocer más sobre la temática, se invita a la comunidad interesada a consultar los boletines de calidad del aire, los cuales se encuentran cargados en el portal del DAGMA, sección grupo de Calidad del Aire: http://www.cali.gov.co/dagma/publicaciones/sistema_de_vigilancia_de_calidad_del_aire_de_cali_svcac_pub

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Estructura del documento

2. Carga y exploración

```
## 'data.frame':   84629 obs. of  12 variables:
## $ Fecha...Hora      : Factor w/ 75869 levels "01/01/2011 01:00:00 AM",...: 13130 13131 ...
## $ PM10...ug.m3.     : Factor w/ 1390 levels "0.1","0.2","0.3",...: 1390 1390 ...
## $ SO2...ug.m3.      : Factor w/ 3968 levels "0.03","0.05",...: 3968 3968 ...
```

```

## $ NO2...ug.m3.      : Factor w/ 5036 levels "0.02","0.13",...: 5036 5036 ...
## $ CO...ug.m3.       : Factor w/ 8027 levels "1000.10","1000.35",...: 8027 8027 ...
## $ O3...ug.m3.       : Factor w/ 6397 levels "0","0.0","0.02",...: 6397 6397 ...
## $ Vel.Viento...m.s.  : Factor w/ 63 levels "0","0.1","0.2",...: 4 3 ...
## $ Dir.Viento..Grados. : Factor w/ 3602 levels "0","0.1","0.2",...: 2063 2257 ...
## $ Temperatura..CÃ..  : Factor w/ 172 levels "16.2","16.3",...: 171 75 ...
## $ Humedad....       : Factor w/ 708 levels "100","100.3",...: 707 477 ...
## $ Radiacion.Solar..Watt.M2.: Factor w/ 7176 levels "0","0.1","0.2",...: 7175 7175 ...
## $ Lluvia..mm.       : Factor w/ 182 levels "#","0","0.1",...: 2 2 ...

##          Fecha...Hora      PM10...ug.m3.      SO2...ug.m3.
## 01/01/2011 01:00:00 AM:      2      ND      :17551      ND      :64788
## 01/01/2011 01:00:00 PM:      2      38      : 421      9.56      : 25
## 01/01/2011 02:00:00 AM:      2      48      : 374      11.09     : 24
## 01/01/2011 02:00:00 PM:      2      27      : 344      5.26      : 24
## 01/01/2011 03:00:00 AM:      2      32      : 342      6.77      : 24
## 01/01/2011 03:00:00 PM:      2      39      : 342      8.57      : 24
## (Other)          :84617      (Other):65255      (Other):19720

```

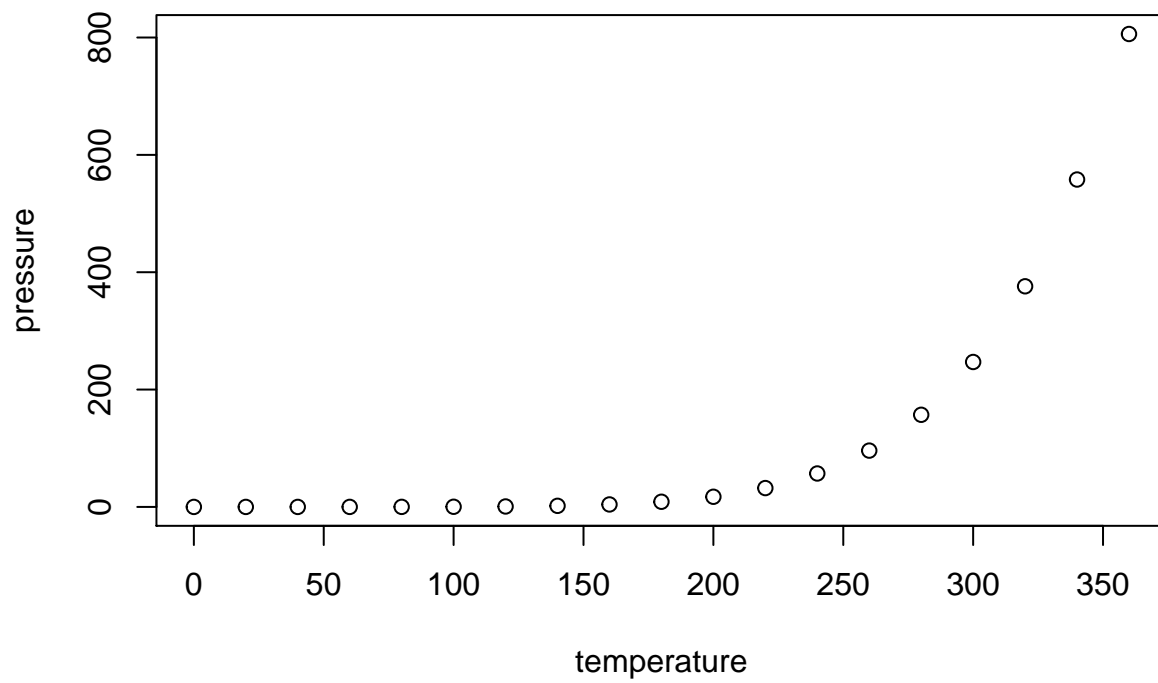
Información de la Entidad Área o dependencia Departamento Administrativo de Gestión del Medio Ambiental
- DAGMA Nombre de la Entidad Alcaldía Municipal de Santiago de Cali Departamento Valle del Cauca
Municipio Cali Orden Territorial Sector Ambiente y Desarrollo Sostenible

Información de Datos Idioma Español Cobertura Geográfica Municipal Frecuencia de Actualización Anual
Fecha Emisión (aaaa-mm-dd) 2016-12-13

Filas 84,6K Columnas 12 Cada fila es un Registro horario

Fecha & Hora
PM10 (ug/m3)
SO2 (ug/m3) NO2 (ug/m3) CO (ug/m3)
O3 (ug/m3)
Vel Viento (m/s)
Dir Viento (Grados) Temperatura (C°)
Humedad (%) Radiacion Solar (Watt/M2)
Lluvia (mm)

Tabla resumen



Métricas de tendencia central

Métricas de dispersión

Visualización de métricas de dispersión

Análisis de las visualizaciones

Análisis de Correlación

Recuerde incluir análisis de todo lo visualizado y no solo el gráfico

3. Limpieza de datos

Reconocimiento y tratamiento de atributos con valores únicos o distintos

Reconocimiento y tratamiento de atributos con valores faltantes

Reconocimiento y tratamiento de atributos con valores atípicos

Reconocimiento y tratamiento de registros atípicos

Reconocimiento y tratamiento de atributos redundantes

4. Creación de la vista minable

Generación de variables derivadas tipo 1 y 2

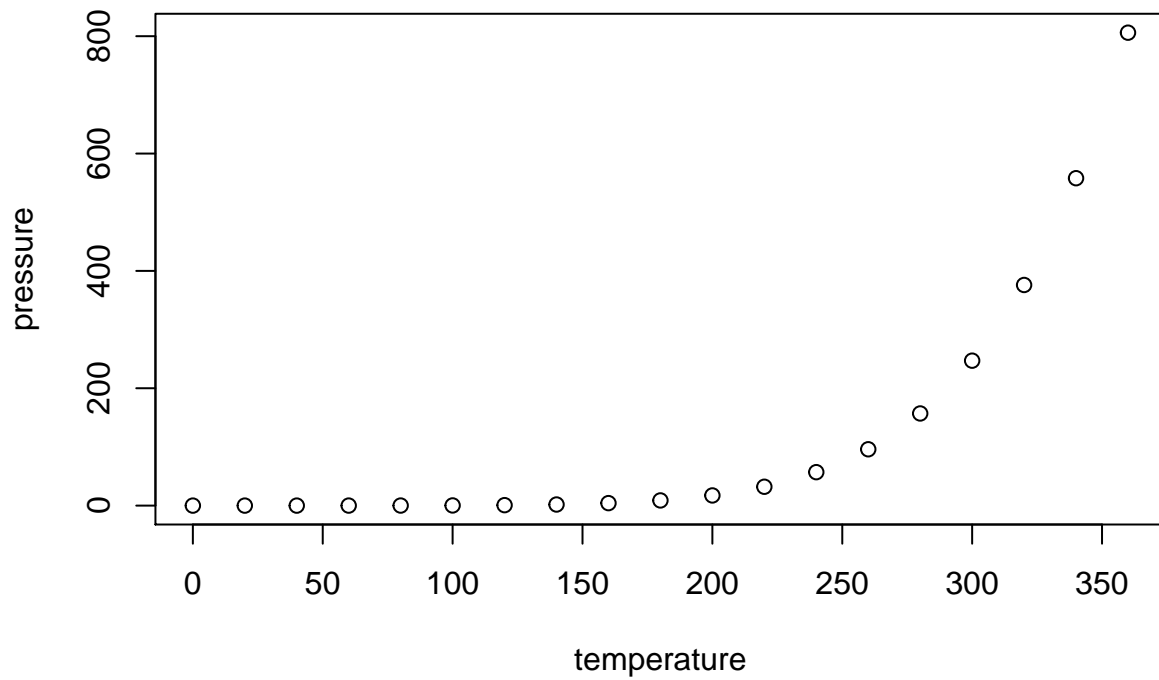
Normalización de al menos un atributo

Discretización de al menos un atributo

Numerización 1 a n de al menos un atributo

5. Conclusiones e Infografía

You can also embed plots, for example:



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