Building's price affected by pollution

Analysis on how pollution affects building's price.

Óscar Muñoz Garrigós
Stanislav Sitanskiy Igorevich
Carlos Frias Ruiz

Index

- Project purposes.
- Pollution data.
- Idealista API.
- Point One: Pollution Histogram
- Point Two: House Price Model
- Point Three: Pollution and Building's price
- Conclusions

Project purposes

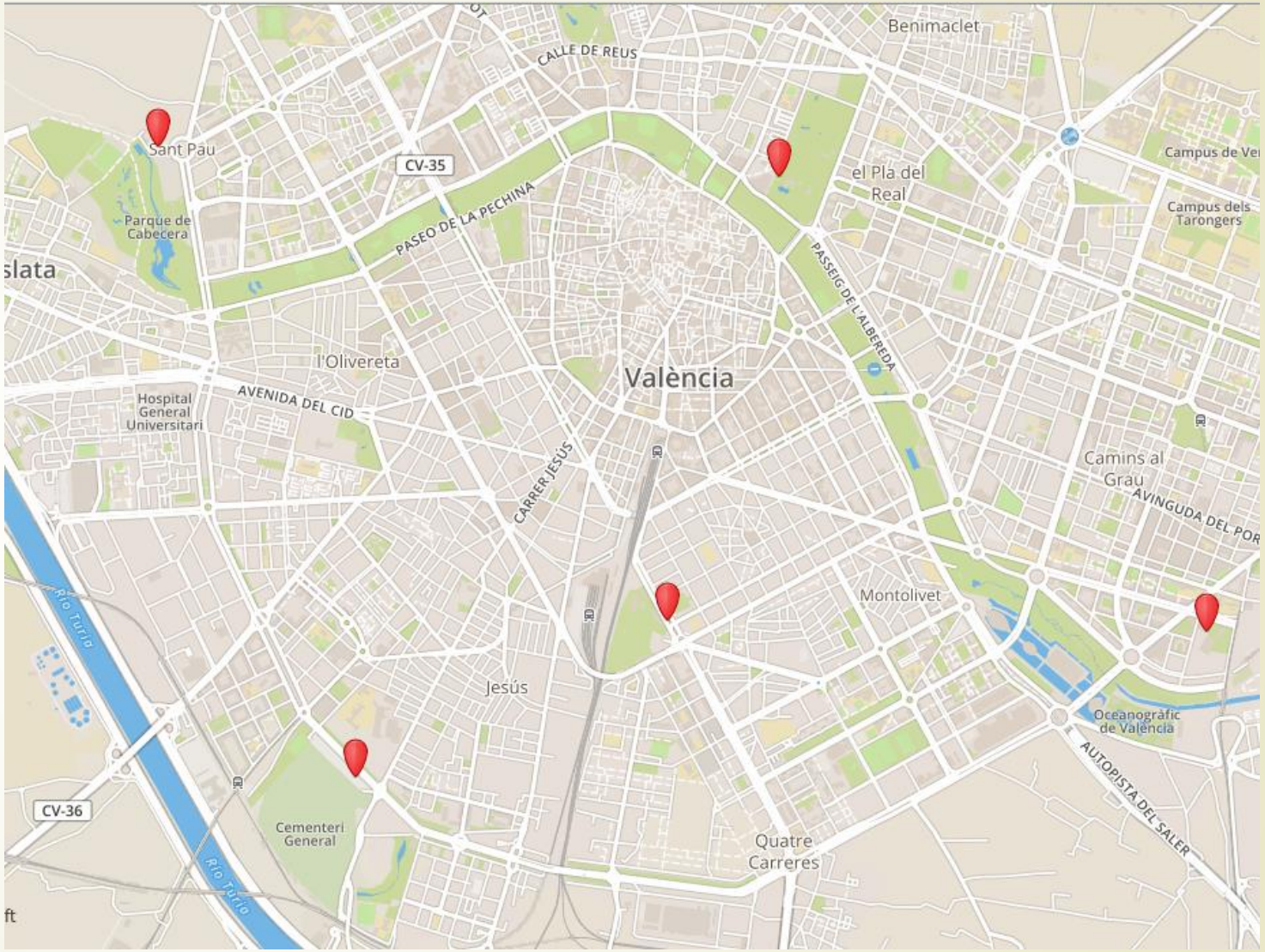
- Get a historical of contamination
- Get a model to explain building's price in Valencia
- Compare contamination with building's price
- Propose a new factor to set building's price

Pollution data

- Pollution data has been taken from Valencia's city hall public data.
- The measured data is from **01/01/2014** to **08/31/2017**.
- We have taken different formats of the.
  - 'csv' format for the main pollution data
  - 'kml' format for the environmental station's location.

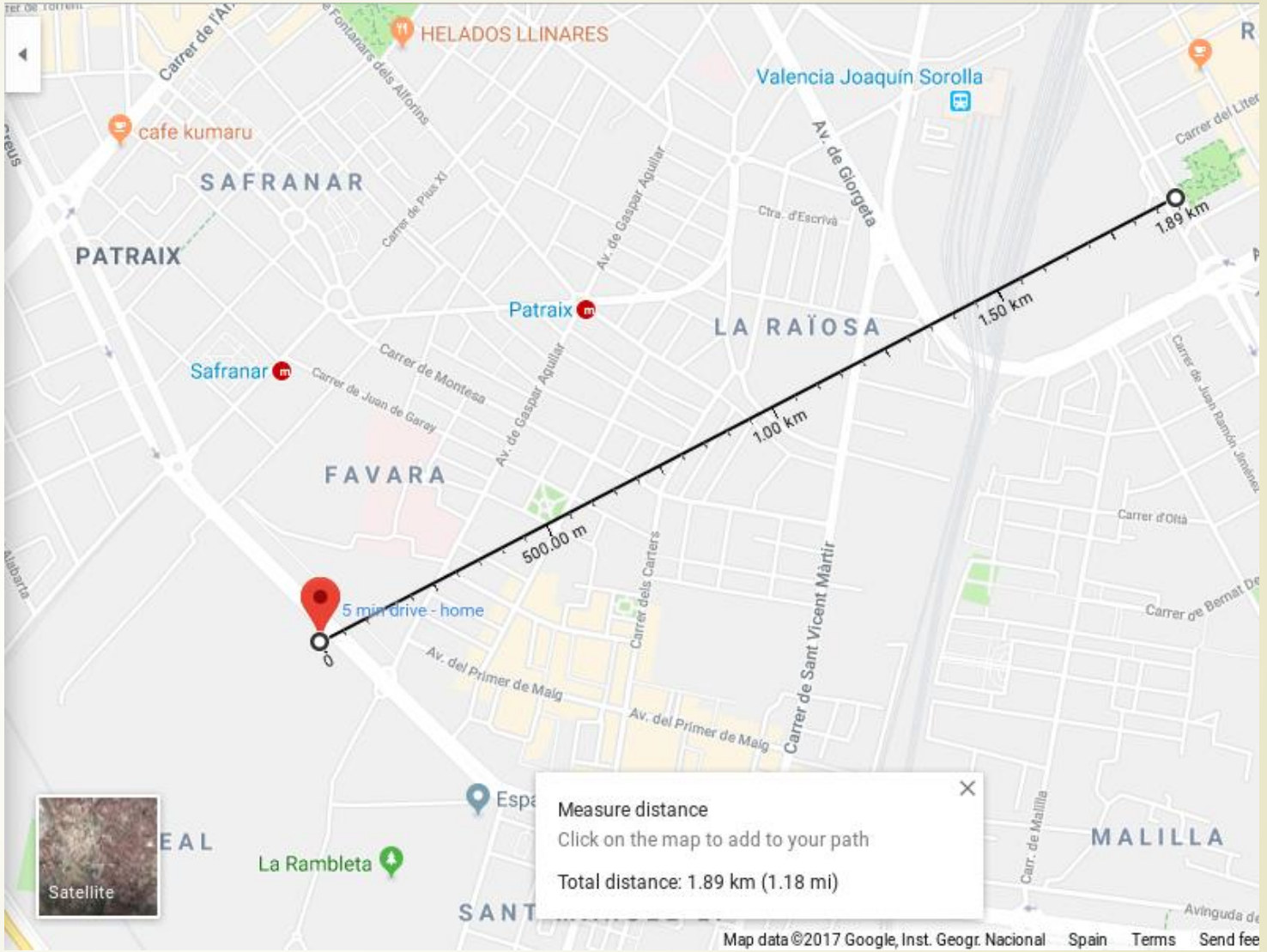
Pollution data

- Environmental stations location:



Pollution data

- Minimum distance between stations



Pollution data

Data in stations

Name	Data
Jardines de Viveros	Dióxido de azufre (SO2),Óxidos de nitrógeno totales (NOx),Monóxido de nitrógeno (NO),Dióxido de nitrógeno (NO2),Ozono,Benceno (Benc),Tolueno (Tolue),Xileno (Xilen),Nivel de presión sonora (SPL),Partículas PM2.5,Partículas PM10,Niquel (Ni),Arsénico (As),Plomo (Pb),Benzo(a)pireno (BaP),Cadmio (Cd)
Avda. de Francia, 60	Dióxido de azufre (SO2),Monóxido de carbono (CO),Ozono,Óxidos de nitrógeno totales (NOx),Monóxido de nitrógeno (NO),Dióxido de nitrógeno (NO2),Partículas PM10,Partículas PM2.5,Partículas PM1,Velocidad de viento (Veloc.)
Avda. Tres Cruces s/n (Parking Cementerio de Valencia)	Dióxido de azufre (SO2),Ozono,Óxidos de nitrógeno totales (NOx),Monóxido de nitrógeno (NO),Dióxido de nitrógeno (NO2),Partículas PM10,Niquel (Ni),Arsénico (As),Plomo (Pb),Benzo(a)pireno (BaP),Cadmio (Cd)
Campus de la U. Politécnica. Avda. Tarongers	Dióxido de azufre (SO2),Ozono,Óxidos de nitrógeno totales (NOx),Monóxido de nitrógeno (NO),Dióxido de nitrógeno (NO2),Partículas PM10,Partículas PM2.5,Partículas PM1
Avda. Pío Baroja - Avda. General Avilés	Dióxido de azufre (SO2),Monóxido de carbono (CO),Ozono,Óxidos de nitrógeno totales (NOx),Monóxido de nitrógeno (NO),Dióxido de nitrógeno (NO2),Partículas PM10,Partículas PM2.5,Partículas PM1
C/ Filipinas, s/n	Dióxido de azufre (SO2),Monóxido de carbono (CO),Ozono,Óxidos de nitrógeno totales (NOx),Monóxido de nitrógeno (NO),Dióxido de nitrógeno (NO2),Partículas PM10,Nivel de presión sonora (SPL),Velocidad de viento (Veloc.),Temperatura media (Temp.),Humedad relativa (H.Rel.),Radiación solar (R.Sol.),Presión barométrica (Pres.)

Pollution data

Common data

NAME	SO2	NOx	NO	NO2	O3	Benc	Tolue	Xilen	SPL	PM2.5	PM10	Ni
Jardines de viveros	x	x	x	x	x	x	x	x	x	x	x	x
Avda. de Francia	x	x	x	x	x						x	x
Avda. Tres Cruces	x	x	x	x	x						x	x
U. Politécnica	x	x	x	x	x					x	x	
Avda. Pío Baroja	x	x	x	x	x					x	x	
C/Filipinas	x	x	x	x	x				x		x	

NAME	As	Pb	BaP	Cd	CO	PM1	viento	Temp	H.Rel.	R.Sol.	Pres.
Jardines de viveros	x	x	x	x							
Avda. de Francia	x	x	x	x	x						
Avda. Tres Cruces	x	x	x	x							
U. Politécnica						x					
Avda. Pío Baroja					x						
C/Filipinas							x	x	x	x	x

Idealista API

- Idealista gave us access to it's api with a maximum of 100 requests per month.
- The requests have been optimized to obtain the desired data.
- A test has been made to find the spot with the minimum number of buildings
- The minimum was found at 359 by '*Avda. Pío Baroja'*
- The calls returns items by 50 each row
- So it's been decided to take 350 items from each point.

Idealista API

- To make the requests the following steps have been followed:
  - 1 - Obtain a token from idealista's auth method
  - 2 - Insert important parameters in request:
    - Central coordinates of request
    - Type of building requested: 'homes'
    - Max distance of the building from our central coordinates: 945 meters
    - Order items by: instance from center
    - Sort items: asc by distance from center
    - Max items per page (max allowed: 50)
    - Num of the page of the request

Idealista API

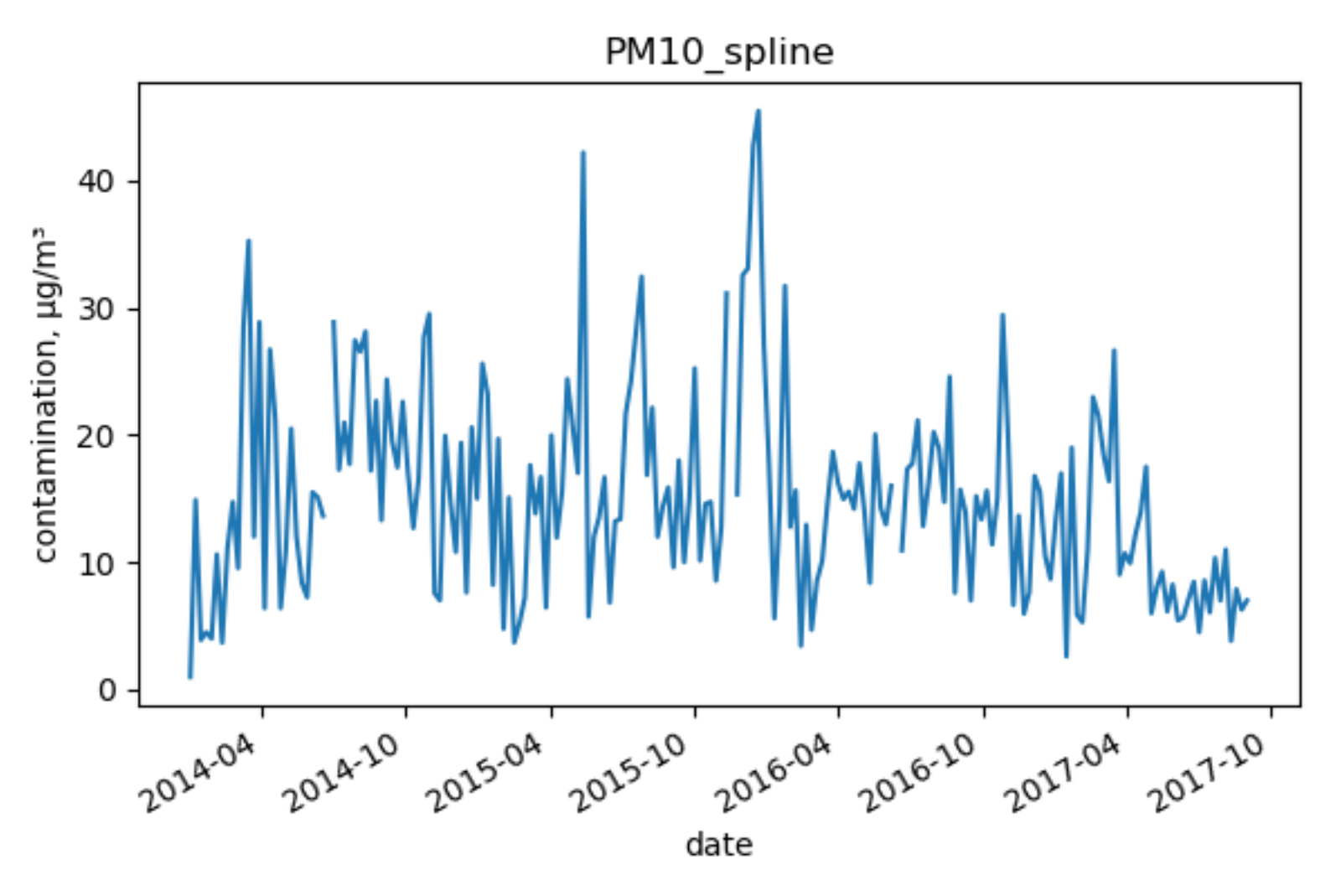
- To make the requests the following steps have been followed:
  - 3 - Clenaning data
    - Join all different responses depending on the central point
    - Anonimize data

Point One: Pollution Histogram

PM10

- anual – 50
- Zona 3 moderada(25%-50%)
- Partículas sólidas o líquidas dispersas en la atmósfera (diámetro < 10micras)
- Penetran en el aparato respiratorio hasta los alvéolos pulmonares.
- Fuentes principales
  - Combustión industrial y doméstica del carbón, procesos industriales, incendios, erosión eólica y erupciones volcánicas, construcciones, demoliciones, etc.
- Salud
  - Produce irritación en las vías respiratorias. Su acumulación en los pulmones origina enfermedades como la silicosis y la asbestosis. Agravan el asma y las enfermedades cardiovasculares.

Point One: Pollution Histogram

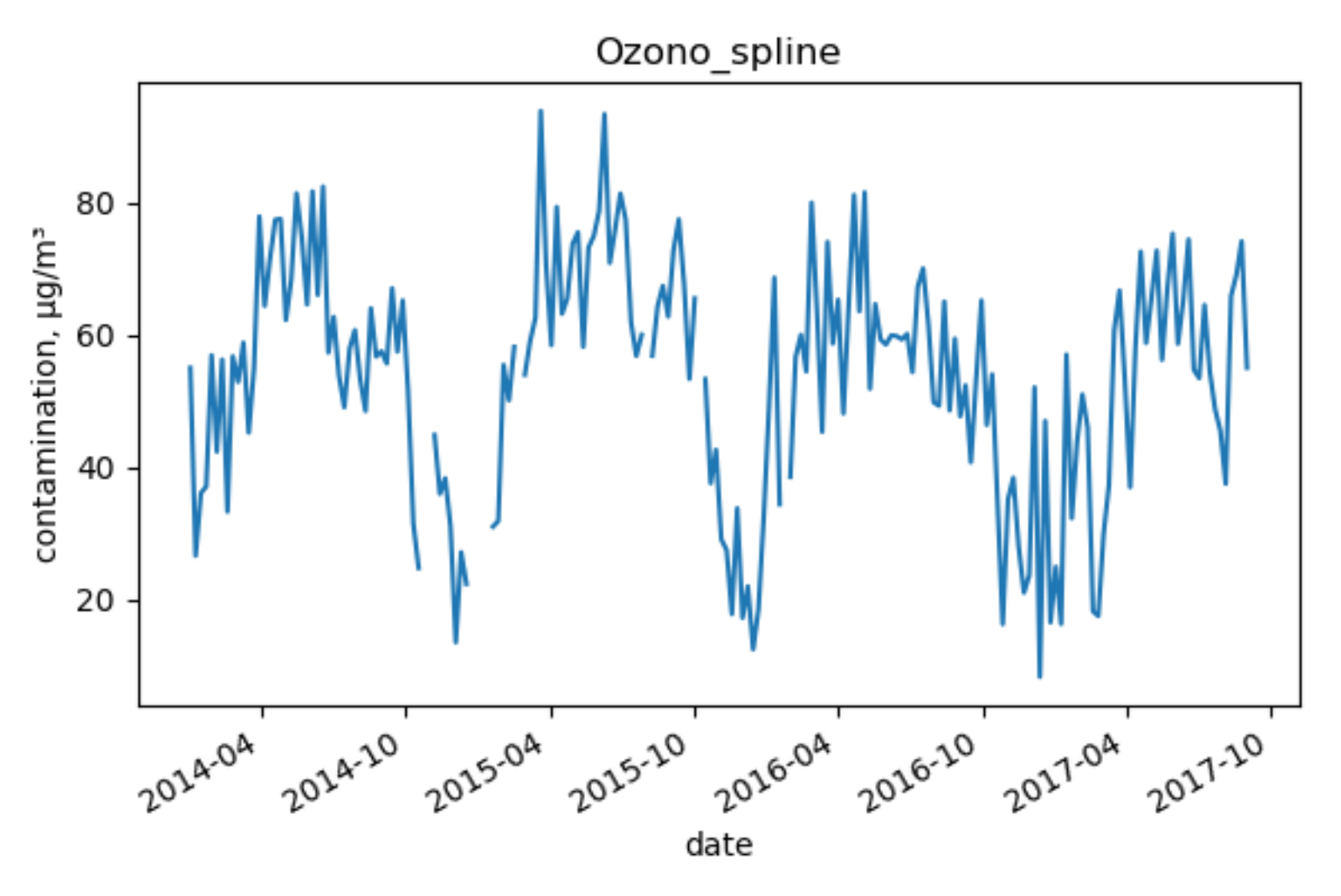


Point One: Pollution Histogram

Ozono

- 8horas-160
- Gas compuesto de tres átomos de oxígeno
- Es un oxidante y capta electrones de otras moléculas, iniciando con esto reacciones en cadenas y perturbando las estructuras vitales en las células
- Se forma en la baja atmósfera mediante reacciones químicas de compuestos orgánicos volátiles y el bióxido de nitrógeno en presencia de la luz
- Salud
  - Produce irritación de los ojos, nariz, y del tracto respiratorio. Agrava las enfermedades respiratorias y cardiovasculares. Los síntomas se manifiestan por dolor de pecho, tos, jadeo, congestión nasal y pulmonar, garganta irritada, náuseas, respiración acelerada.

Point One: Pollution Histogram



Point One: Pollution Histogram

Dióxido de nitrógeno (N02)

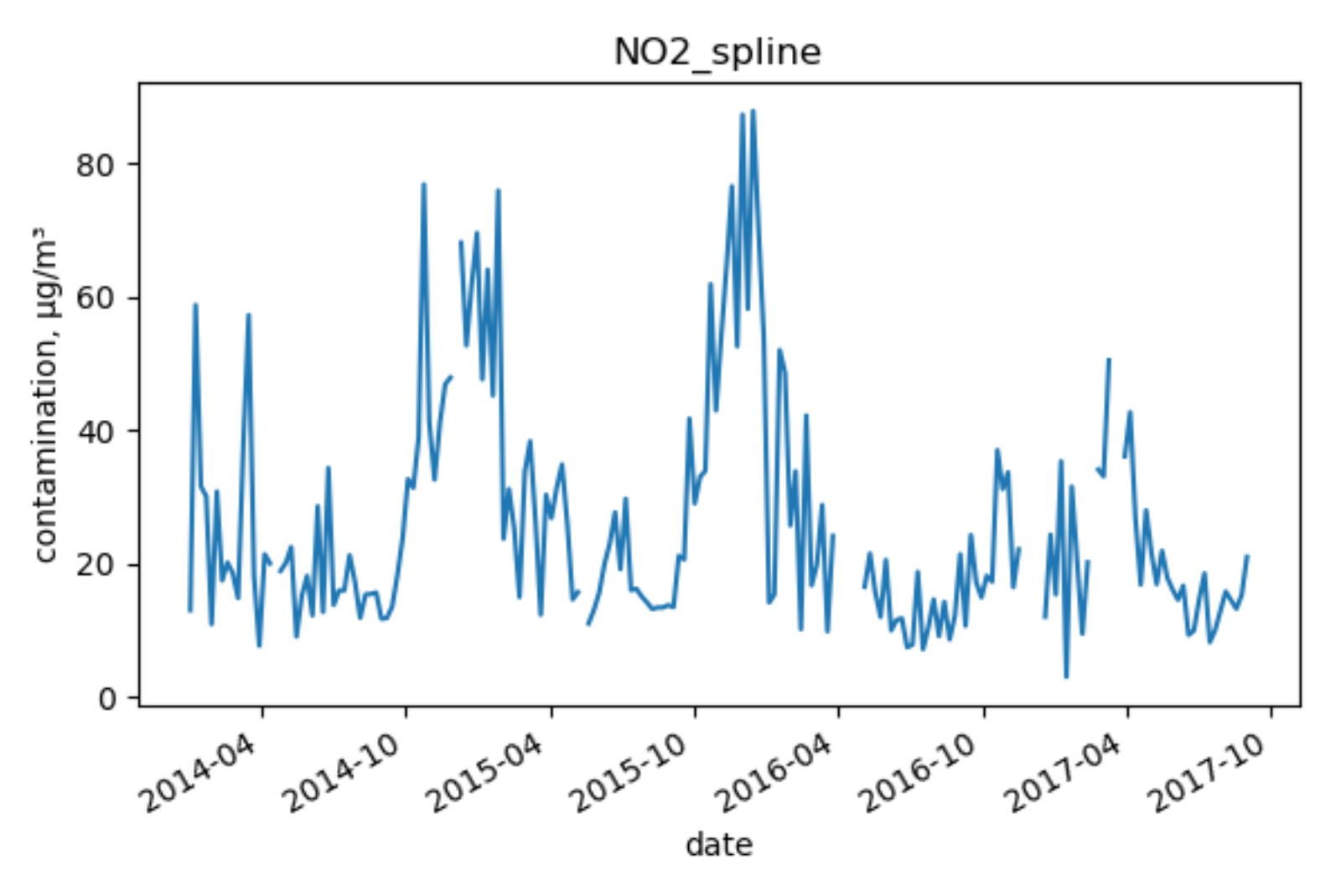
- anual – 100
- Zona 3 moderada (25%-50%)
- Zona 4 marginal (10%-25%)
- Gas amarillo parduzco picante que da al smog su característica color café.
- Se produce por reacción fotoquímica de óxido nitroso (NO) en el aire
- Es un oxidante, con capacidad de quitar electrones a otras moléculas

Point One: Pollution Histogram

Dióxido de nitrógeno (N02)

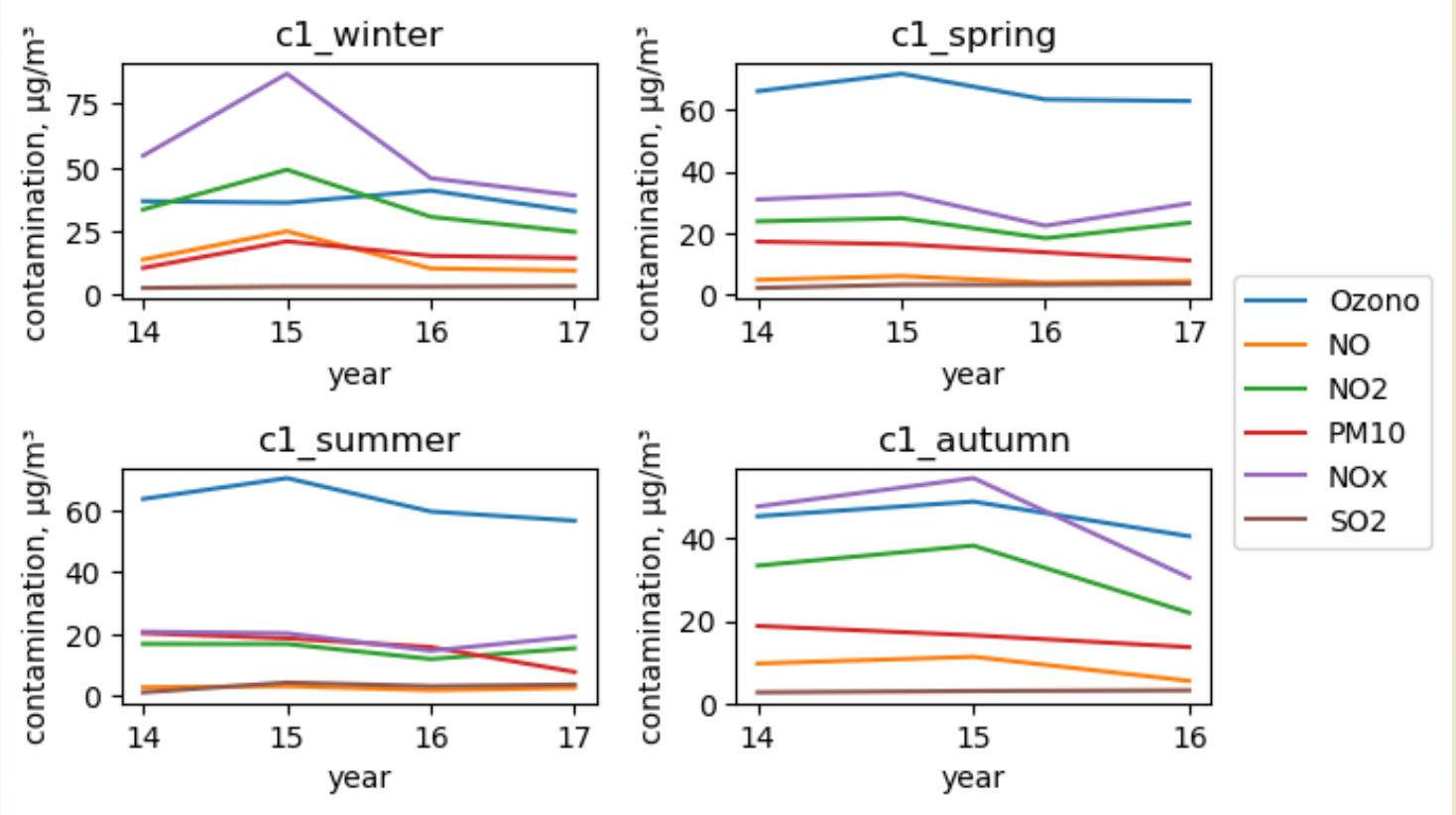
- Fuentes principales
  - Se produce en la combustión de altas temperaturas en industrias y vehículos, tormentas eléctricas y en las reacciones químicas atmosféricas.
- Salud
  - Produce irritación a los pulmones y daña las células que revisten los pulmones. No se presentan síntomas a menos que se trate de concentraciones muy altas; el mayor daño aparece de 5 a 72 horas después de la exposición causando edema pulmonar. Los niños que habitan en casas con calefacción presentan infecciones respiratorias (resfriados comunes). Algunos de los síntomas son: descarga nasal, dolor de cabeza, mareo y dificultad al respirar

Point One: Pollution Histogram



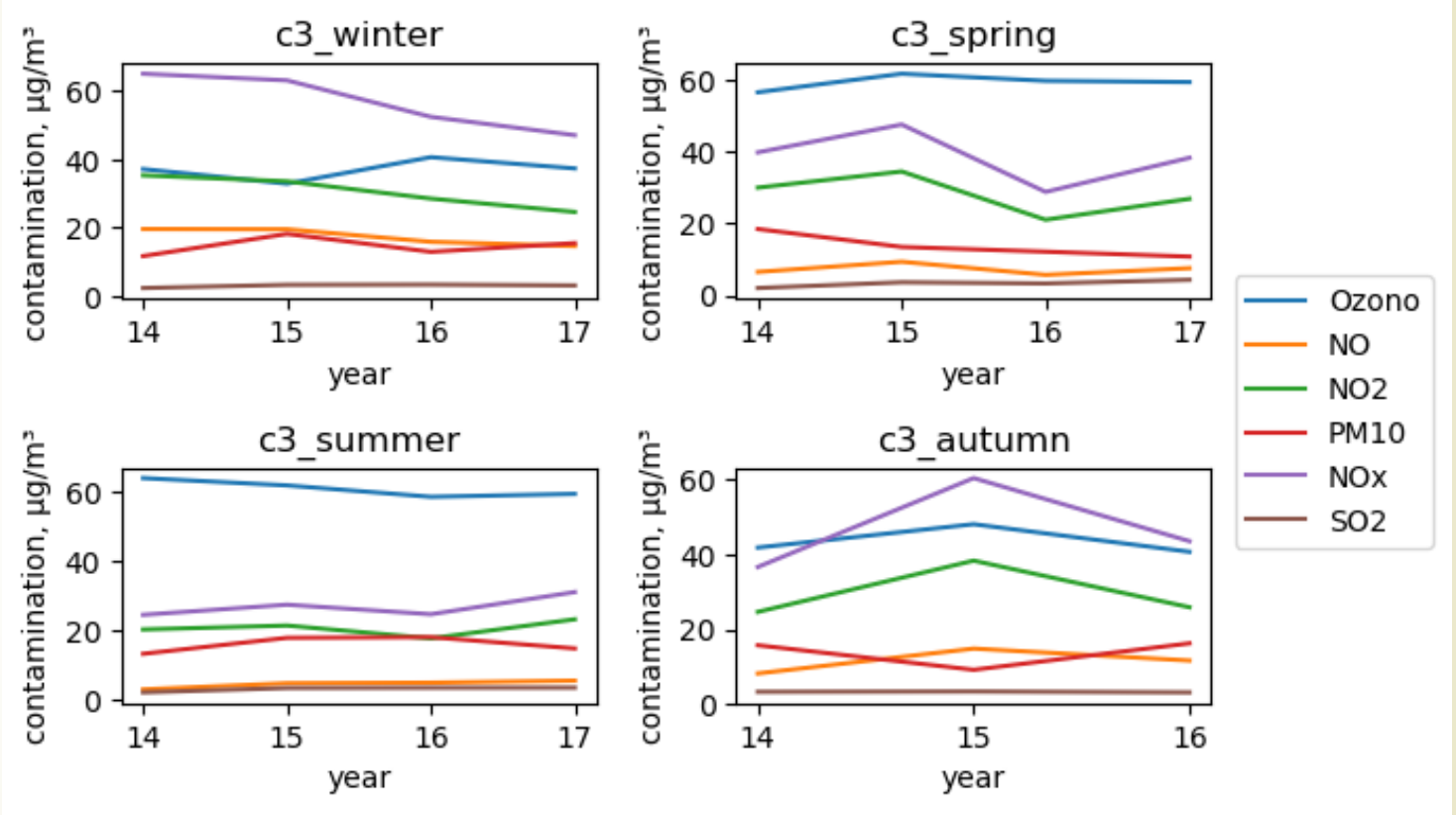
Point One: Pollution Histogram

- Evolucion de la zona 1
- Jardines de Viveros



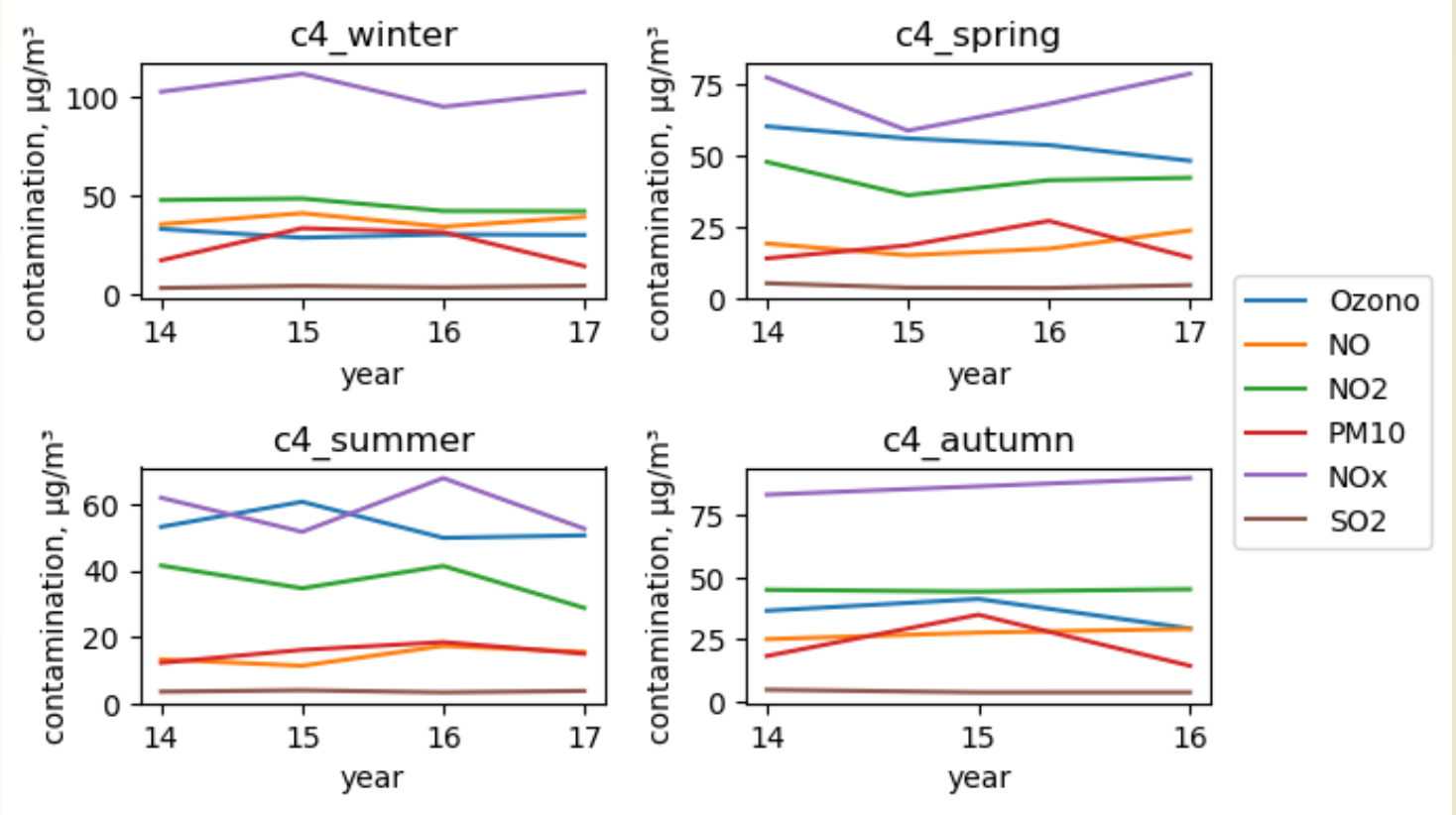
Point One: Pollution Histogram

- Evolucion de la zona 3
- Avda. de Francia, 60



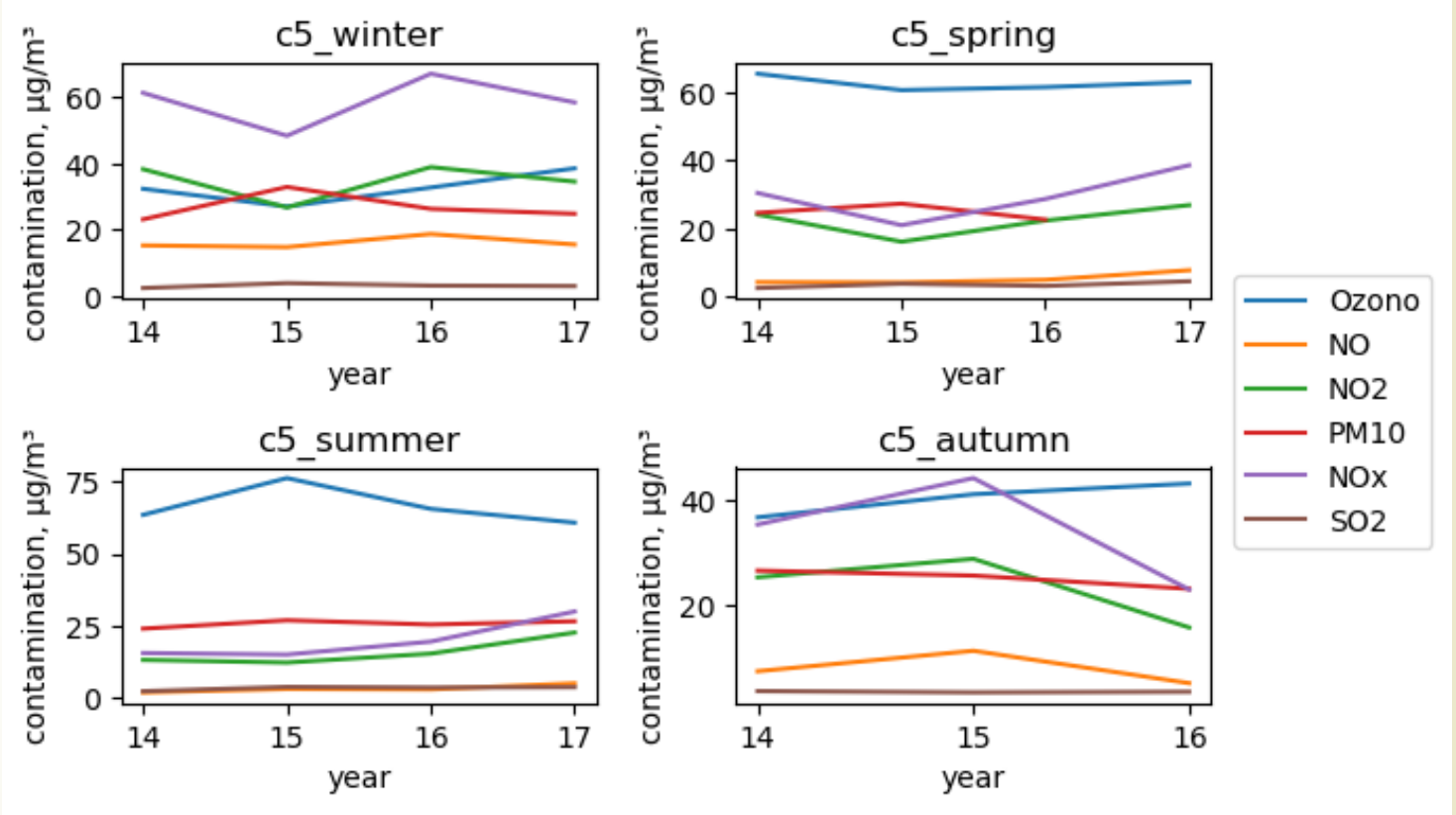
Point One: Pollution Histogram

- Evolucion de la zona 4
- Avda. Tres Cruces s/n (Parking Cementerio de Valencia)



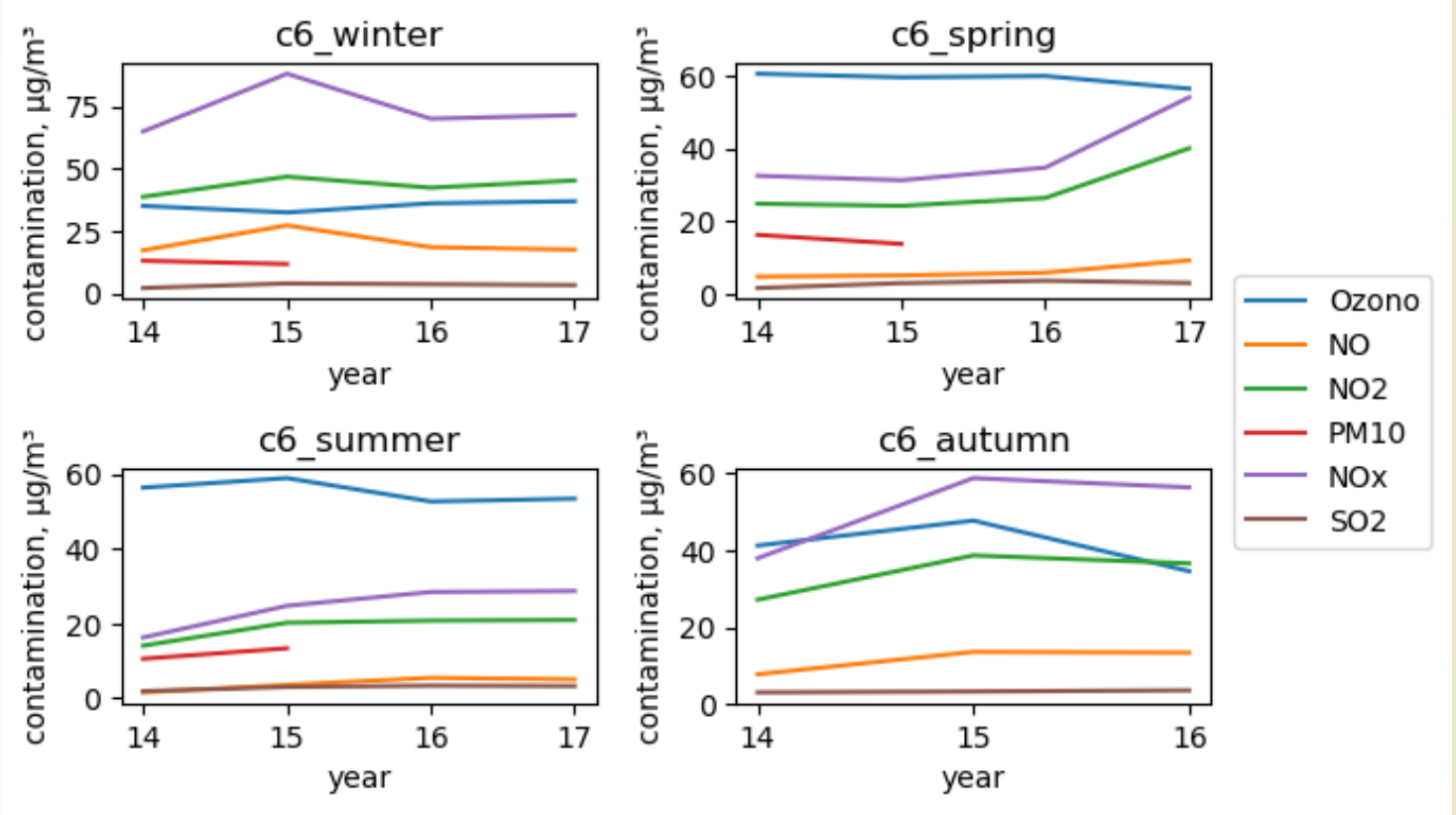
Point One: Pollution Histogram

- Evolucion de la zona 5
- Campus de la U. Politécnica. Avda. Tarongers



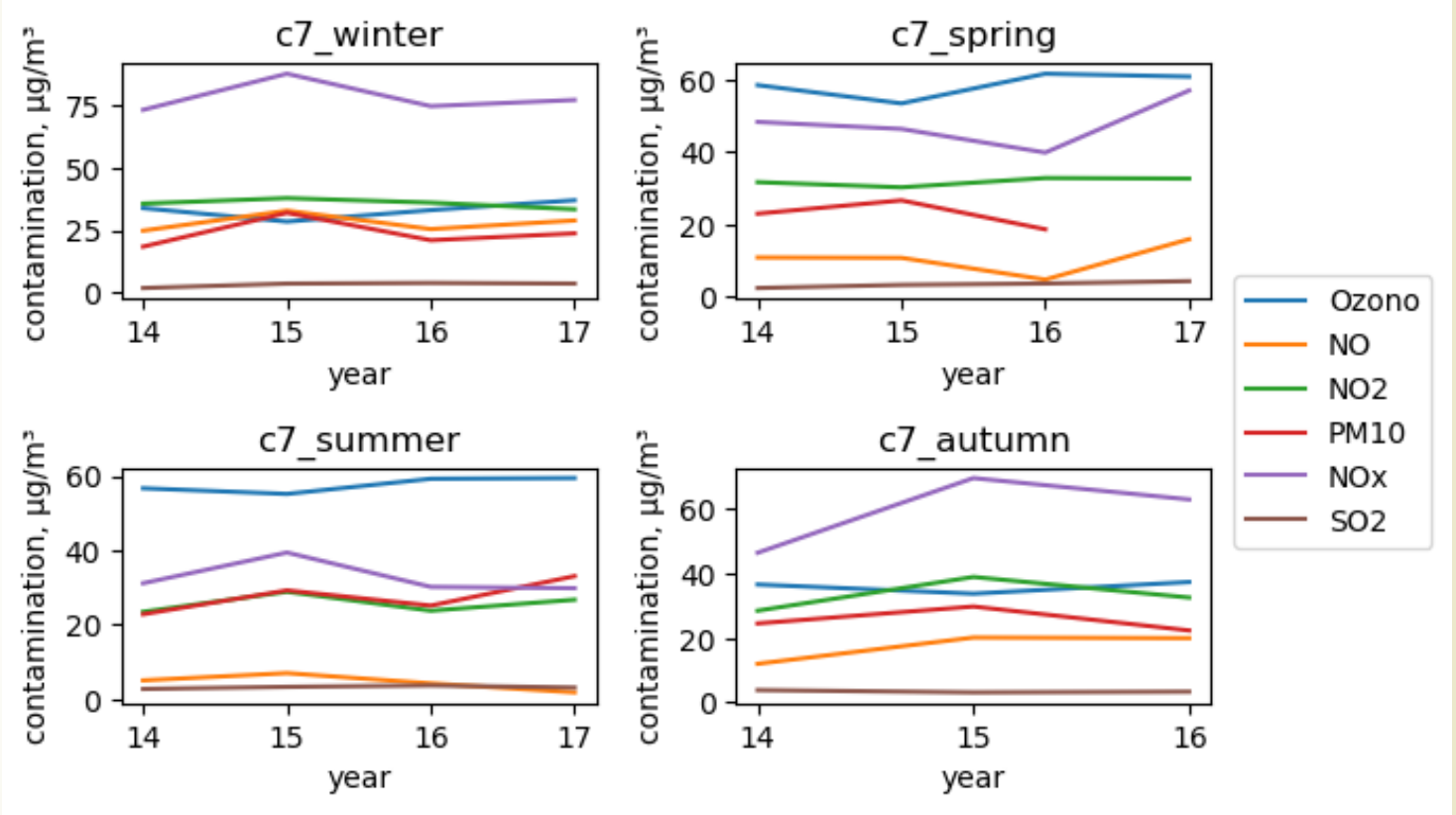
Point One: Pollution Histogram

- Evolucion de la zona 6
- Avda. Pío Baroja - Avda. General Avilés



Point One: Pollution Histogram

- Evolucion de la zona 7
- C/ Filipinas, s/n



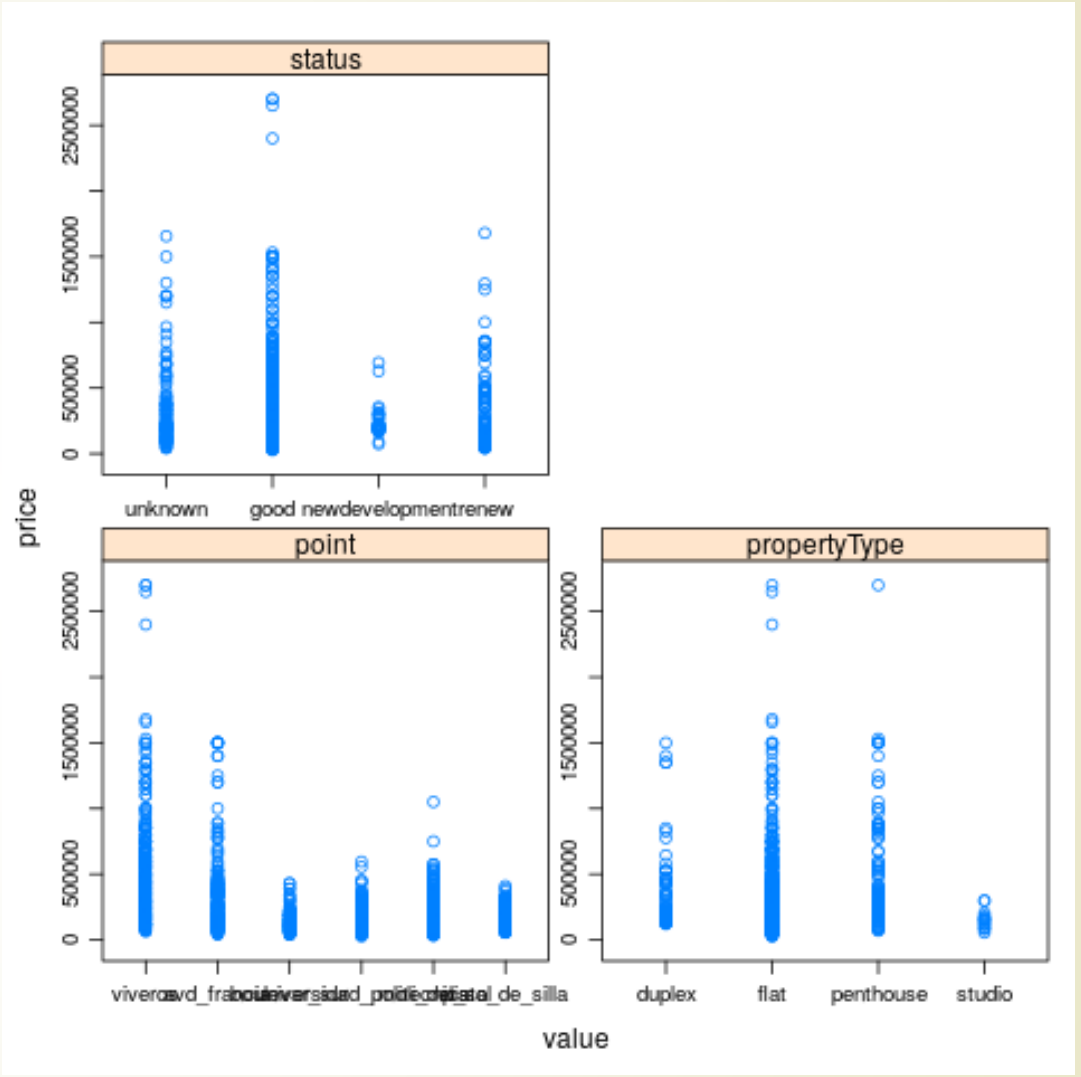
Point Two: House Price Model

- Data summary

```
> summary(idealista)
                     point       numPhotos         floor
 viveros                :341   Min.   : 0.00   Min.   : 1.000
 avd_francia            :343   1st Qu.:10.00   1st Qu.: 4.000
 boulevar_sur           :335   Median :16.00   Median :10.000
 universidad_politecnica:295   Mean   :17.89   Mean   : 9.073
 moli_del_sol           :330   3rd Qu.:24.00   3rd Qu.:12.000
 pista_de_silla         :341   Max.   :61.00   Max.   :20.000
     price              propertyType       size        exterior
 Min.   :  26600   chalet      :   0   Min.   : 32.0   Mode :logical
 1st Qu.: 121005   duplex      :  73   1st Qu.: 85.0   FALSE:311
 Median : 175000   flat        :1745   Median :108.0   TRUE :1674
 Mean   : 250460   penthouse   : 150   Mean   :119.3   NA's :0
 3rd Qu.: 285000   studio      :  17   3rd Qu.:134.0
 Max.   :2700000   countryHouse:   0   Max.   :655.0
     rooms         bathrooms        distance       hasVideo
 Min.   :0.000   Min.   :0.000   Min.   :  33.0   Mode :logical
 1st Qu.:2.000   1st Qu.:1.000   1st Qu.: 379.0   FALSE:1899
 Median :3.000   Median :2.000   Median : 545.0   TRUE :86
 Mean   :3.032   Mean   :1.778   Mean   : 541.5   NA's :0
 3rd Qu.:4.000   3rd Qu.:2.000   3rd Qu.: 665.0
 Max.   :8.000   Max.   :8.000   Max.   :1093.0
            status     newDevelopment   hasLift         priceByArea
 unknown       : 157   Mode :logical   Mode :logical   Min.   : 455
 good          :1572   FALSE:1894      FALSE:263       1st Qu.:1237
 newdevelopment:  55   TRUE :91        TRUE :1722      Median :1714
 renew         : 201   NA's :0         NA's :0         Mean   :1926
                                                       3rd Qu.:2411
                                                       Max.   :7614

  detailedType_typology
 chalet      :   0
 flat        :1985
 countryHouse:   0



> nrow(idealista)
[1] 1985
```

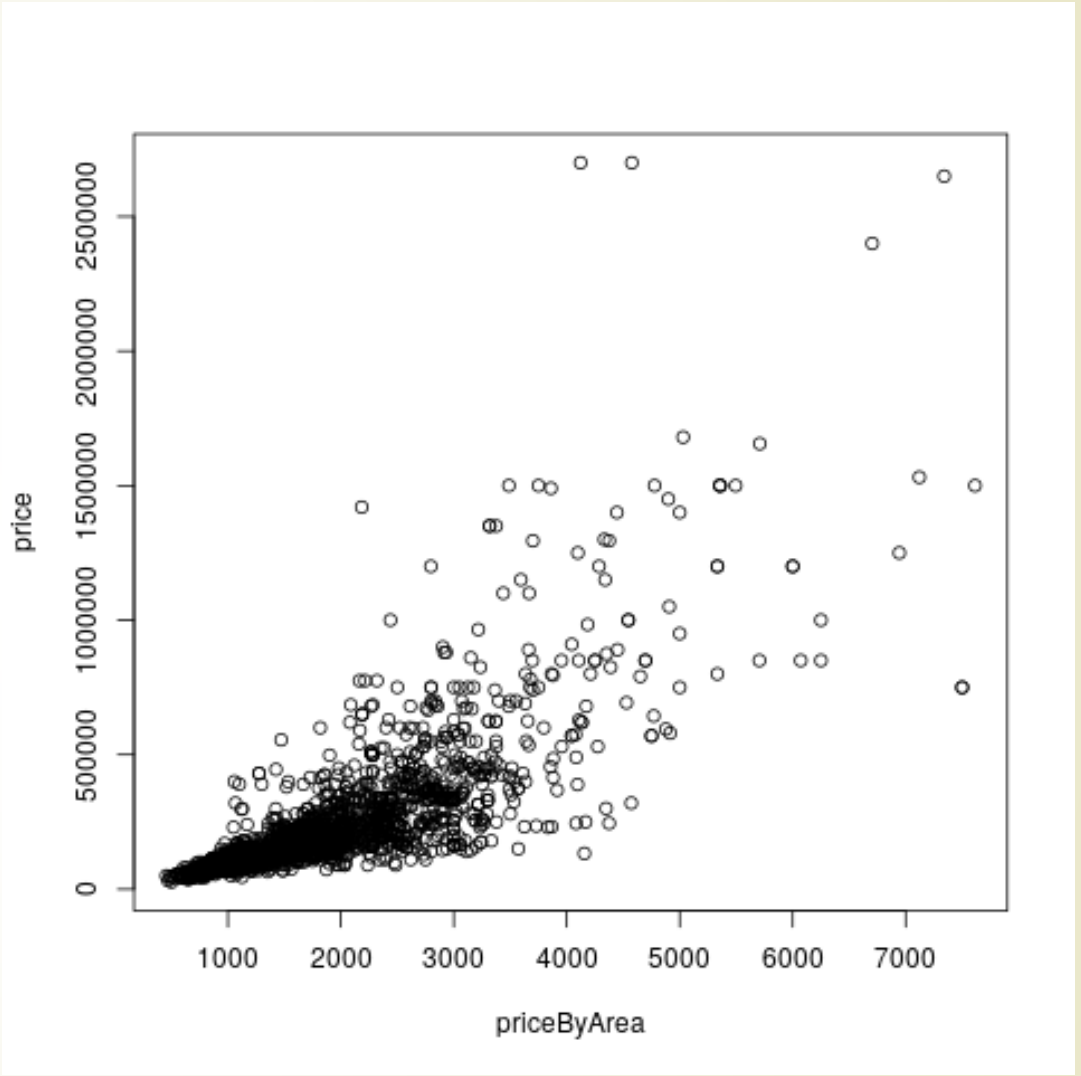
Point Two: House Price Model

- Some interesting plots which can give us a lot of information:



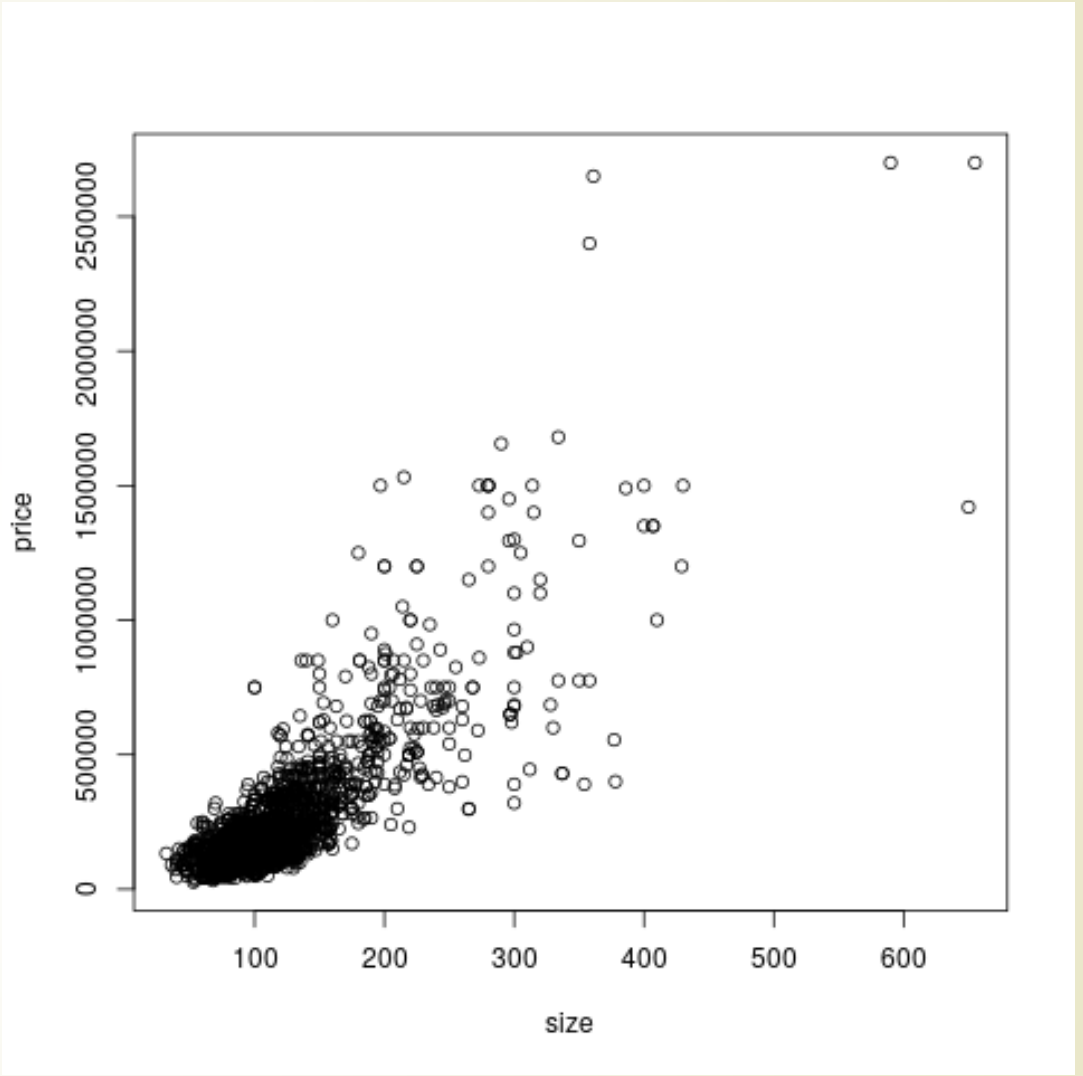
Point Two: House Price Model

- Price of he m² in the area



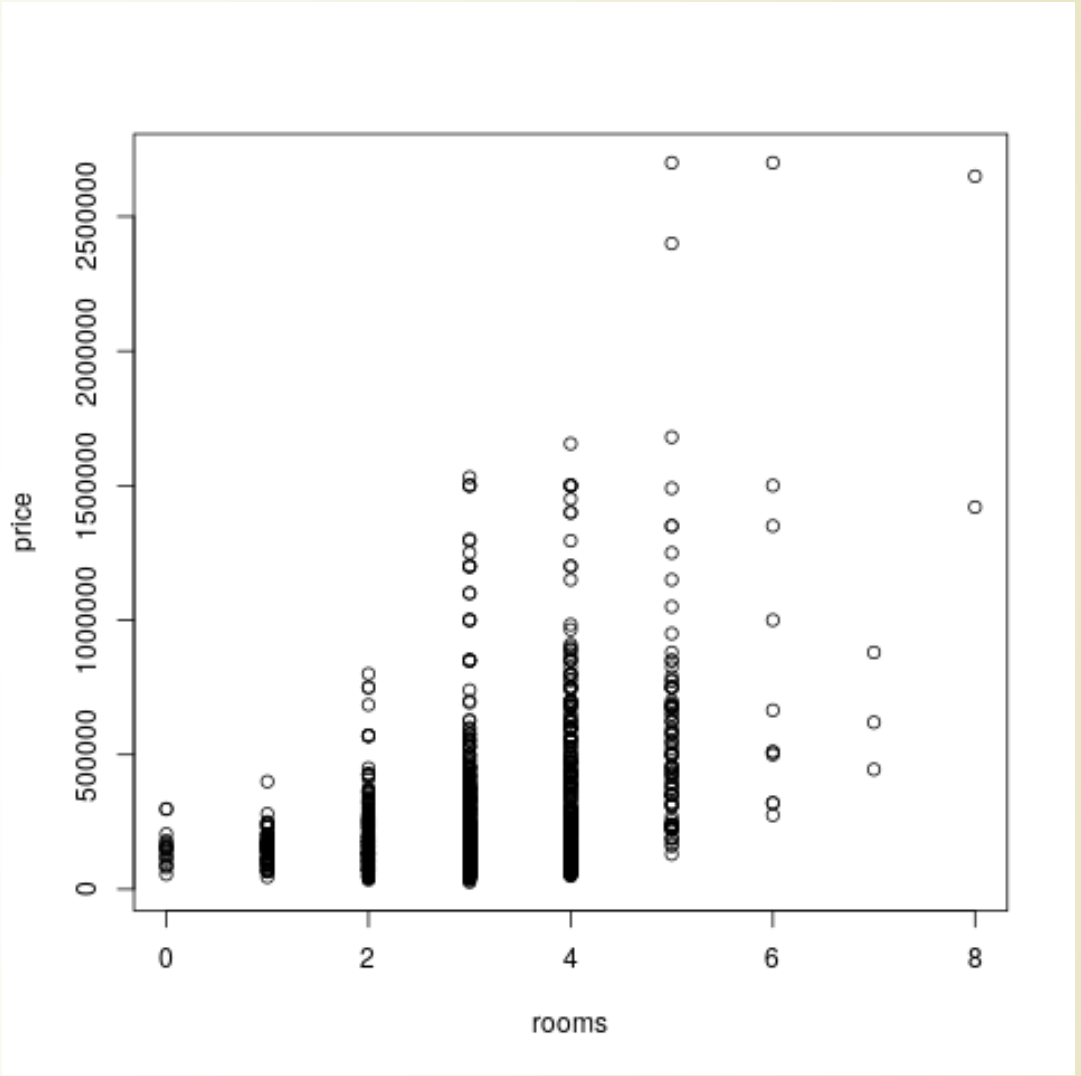
Point Two: House Price Model

- Sice of the building



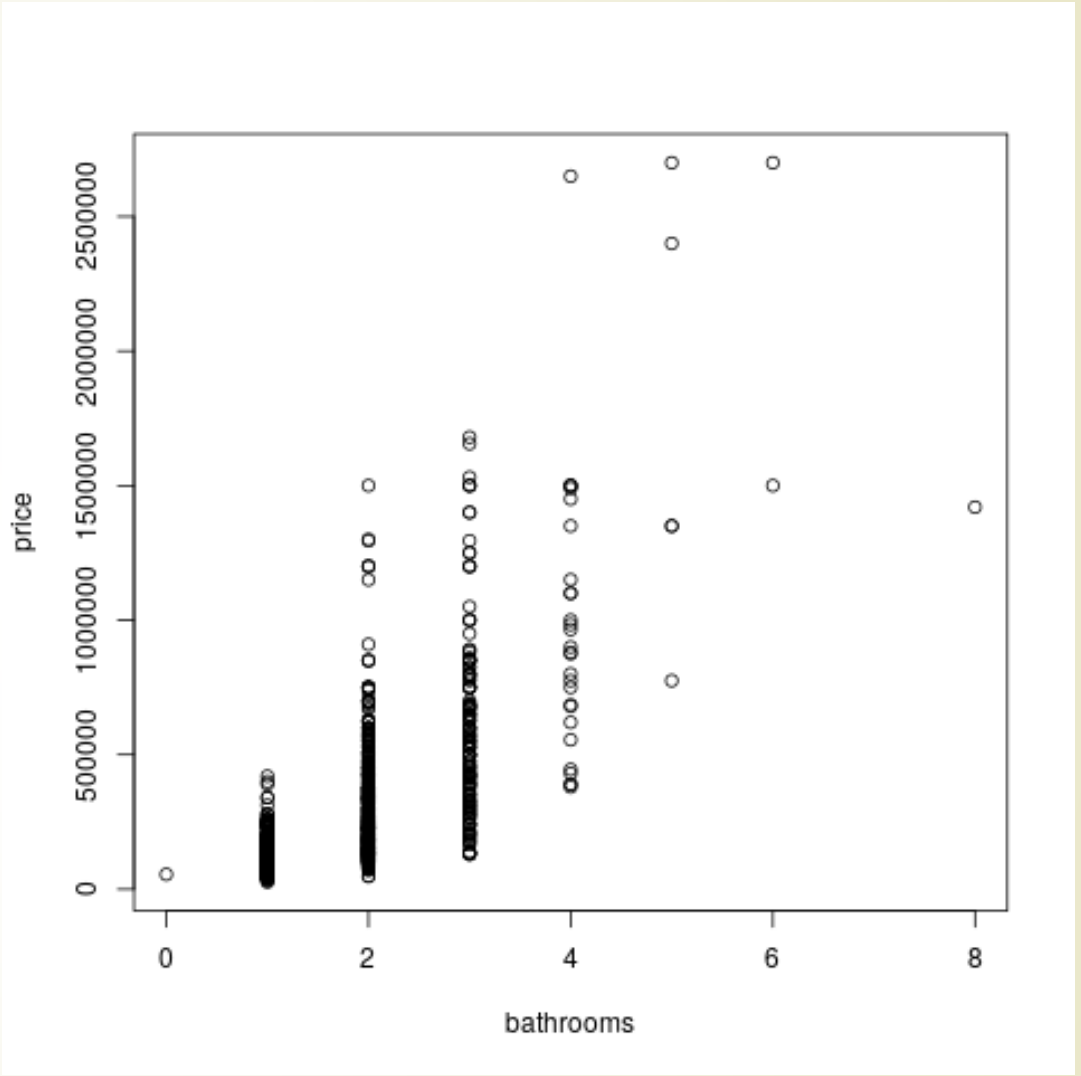
Point Two: House Price Model

- Number of rooms in the building



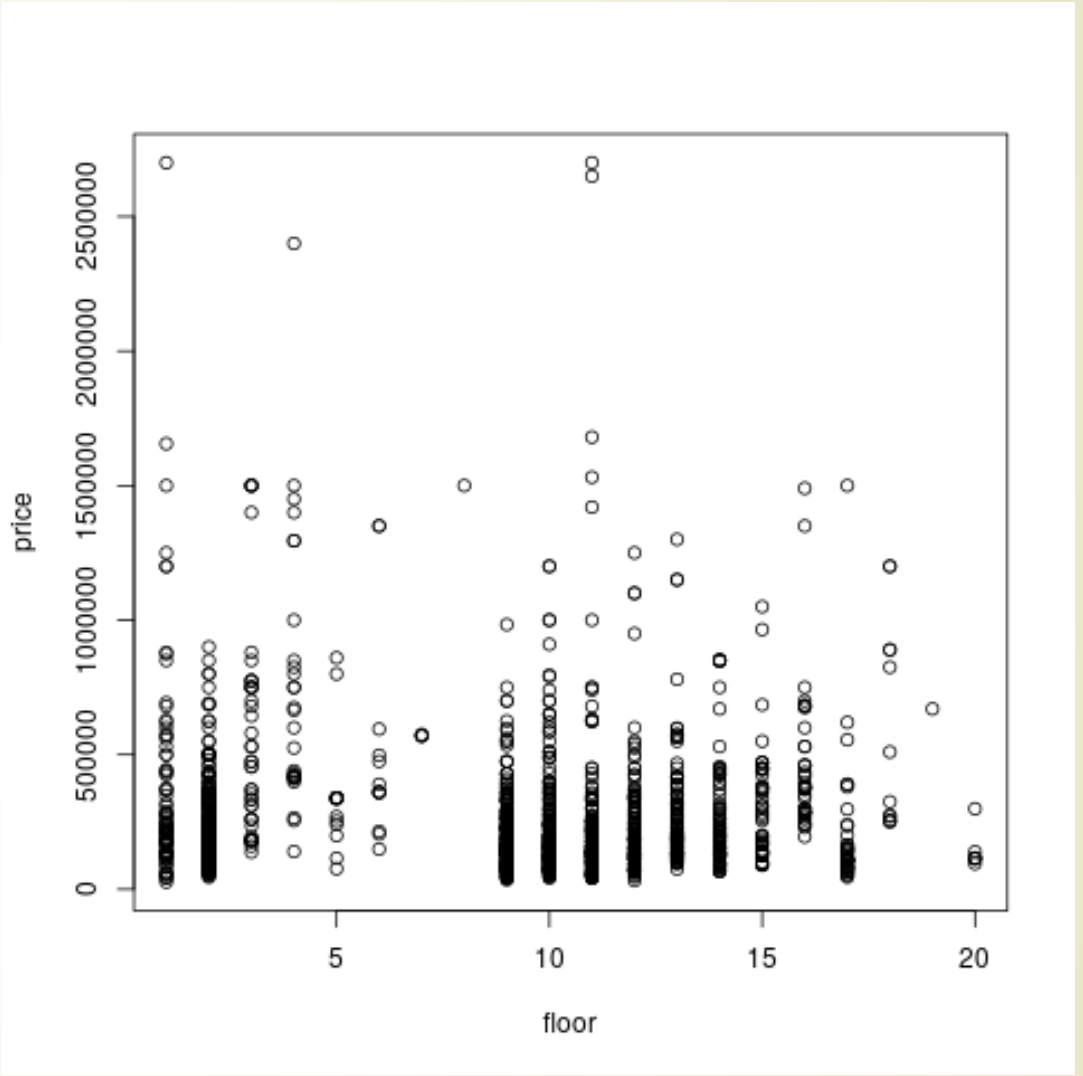
Point Two: House Price Model

- Number of bathrooms in the building



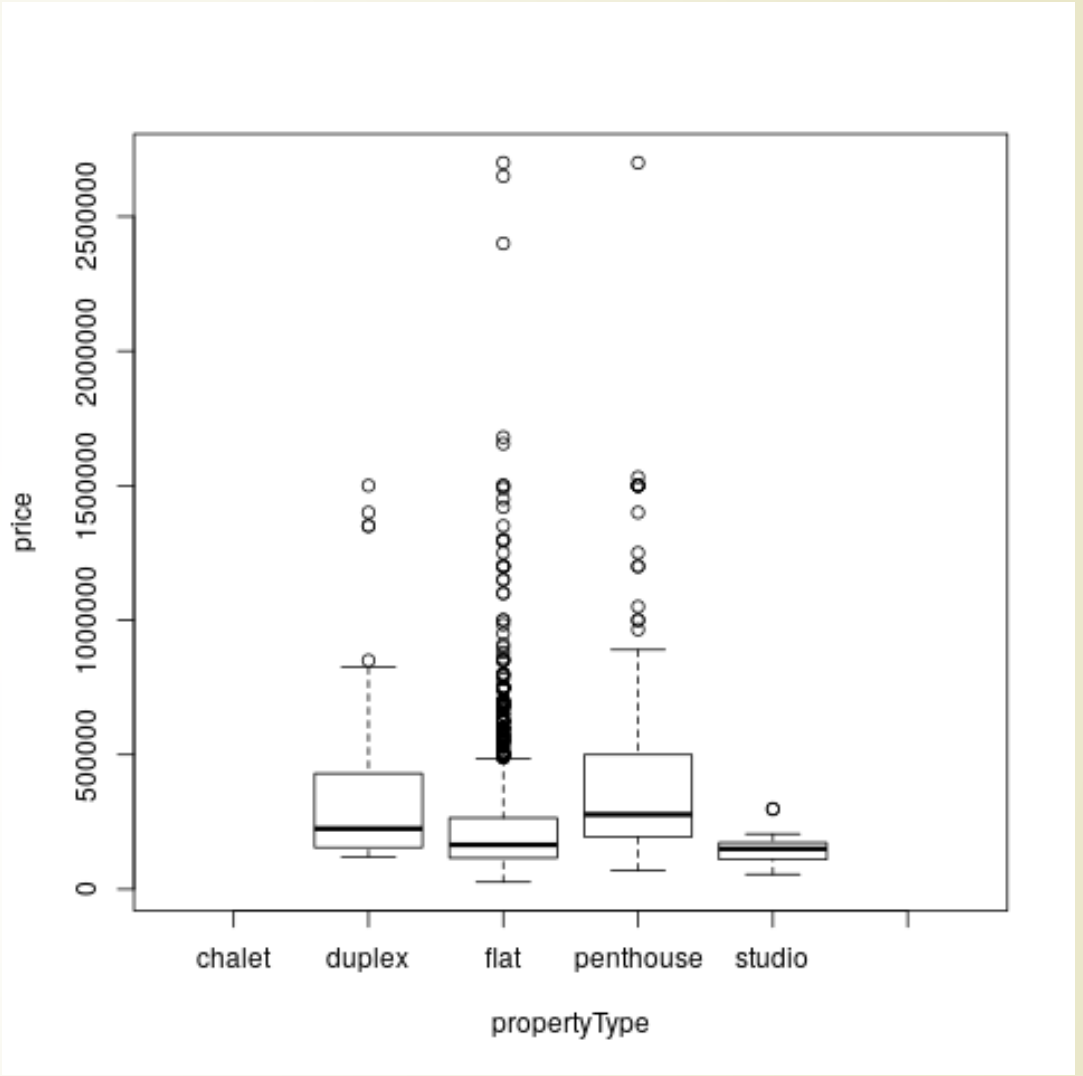
Point Two: House Price Model

- Building's floor



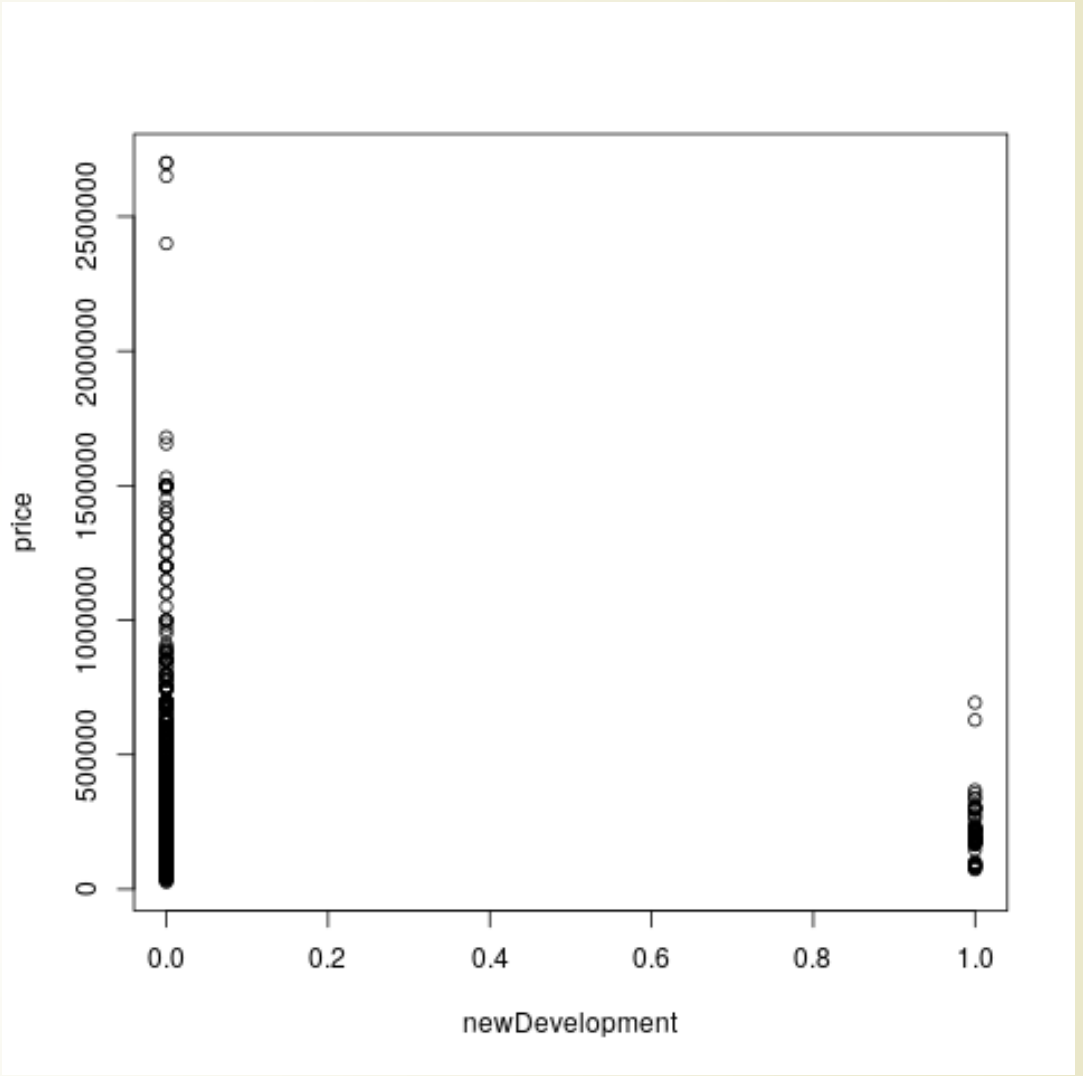
Point Two: House Price Model

- Building's type



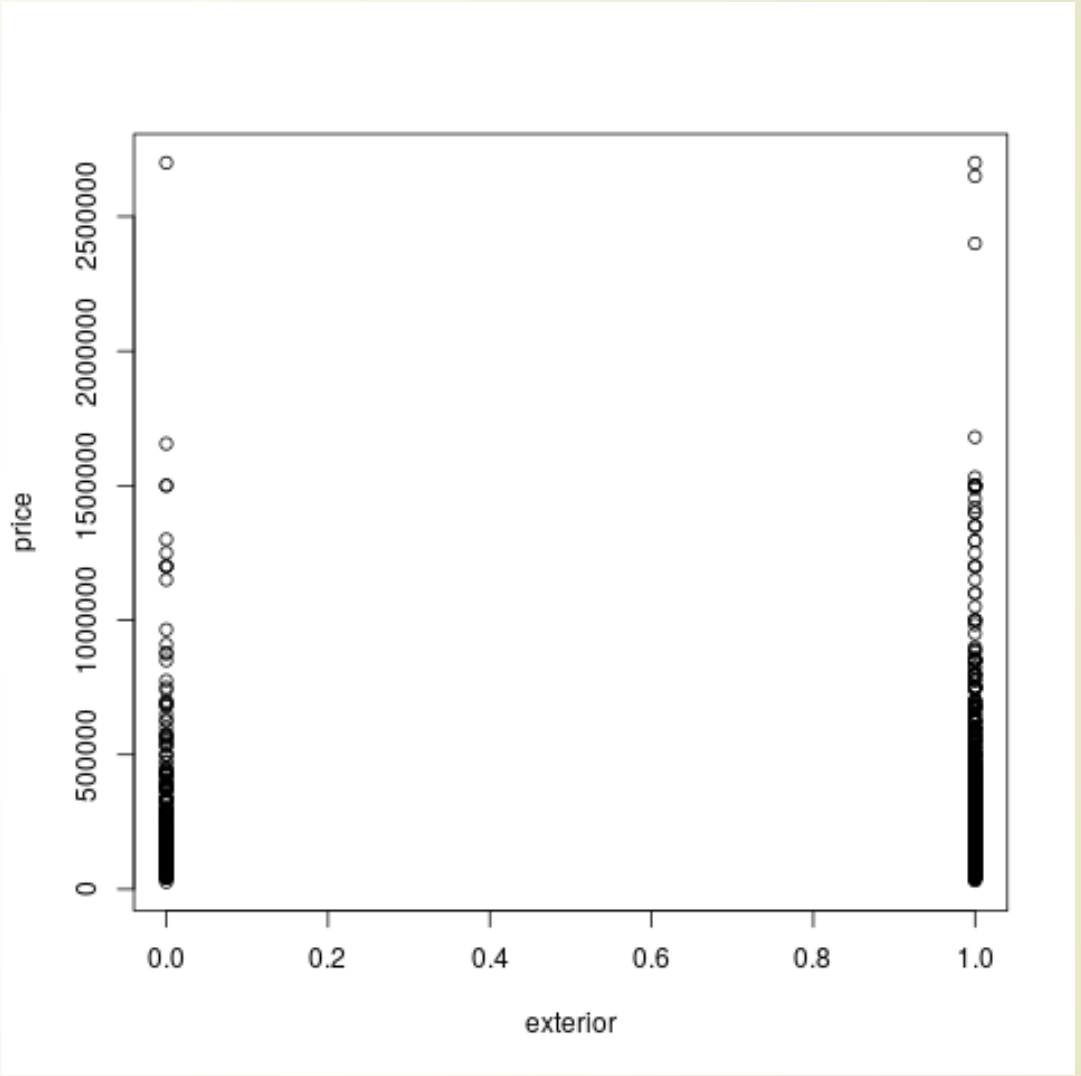
Point Two: House Price Model

- ¿Is a new development?



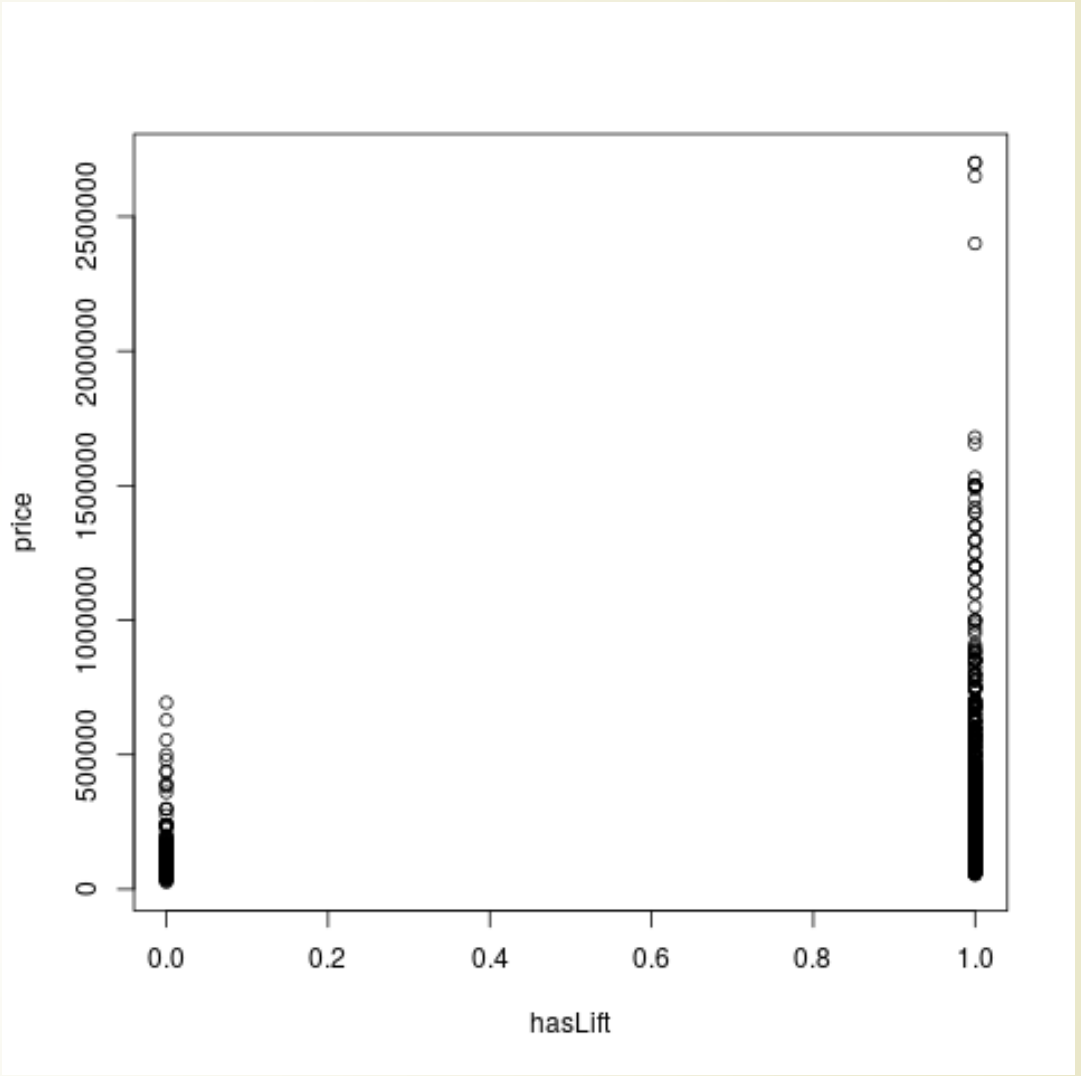
Point Two: House Price Model

- ¿Does the building has exterior?



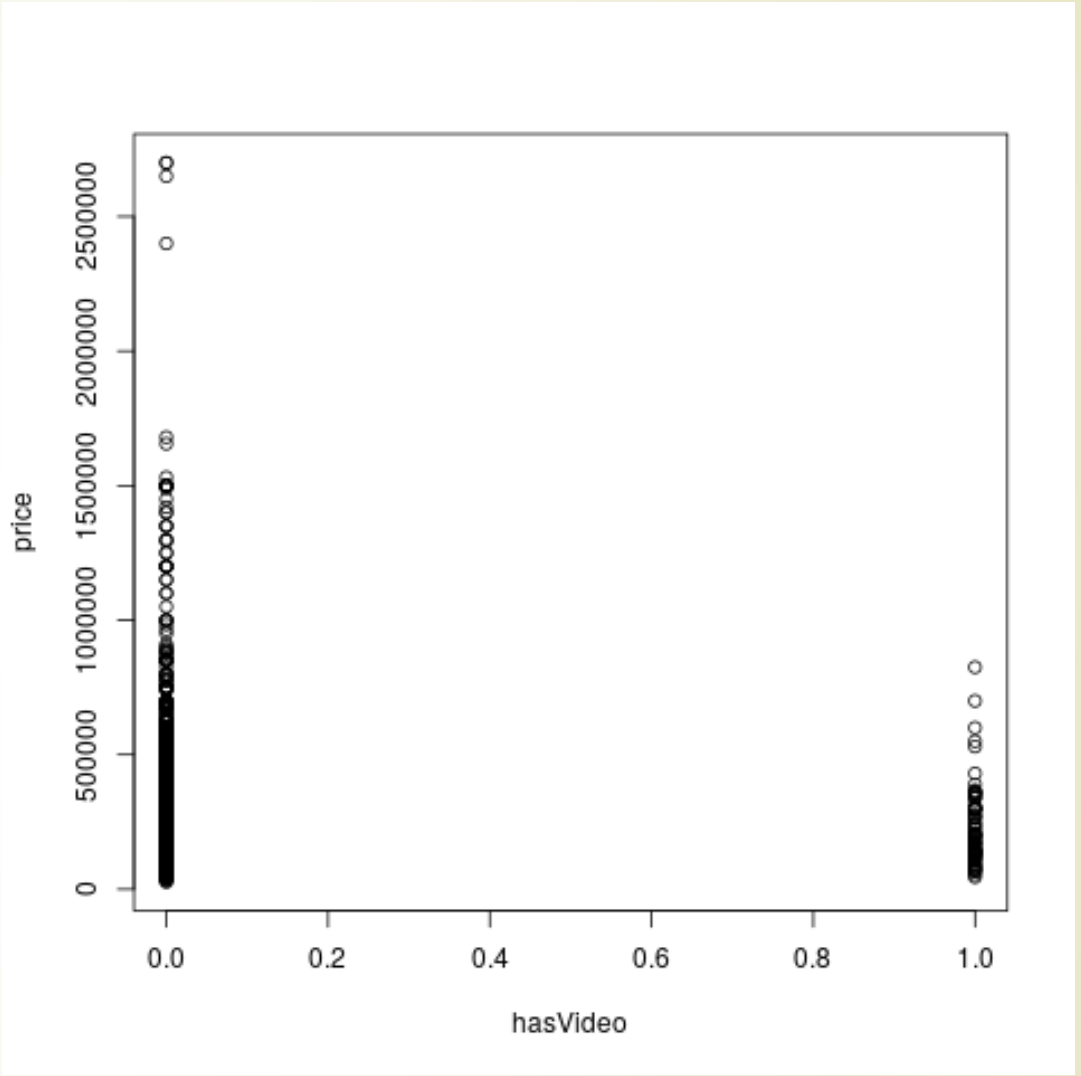
Point Two: House Price Model

- ¿Does the building has lift?



Point Two: House Price Model

- ¿Does the building's annuncement has video?



Point Two: House Price Model

Correlation matrix

	numPhotos	floor	price	size	rooms	bathrooms	distance	priceByArea
numPhotos	1	0,02286639	0,20012760	0,21713154	0,15805619	0,24358395	0,02651322	0,15477167
floor	0,02286639	1	-0,06298738	-0,06784338	-0,11193510	-0,07852130	0,03399334	-0,03247983
price	0,20012760	-0,06298738	1	0,81993945	0,38421684	0,66427945	-0,15318137	0,77242896
size	0,21713154	-0,06784338	0,81993945	1	0,59869868	0,75540254	-0,11988583	0,38544488
rooms	0,15805619	-0,11193510	0,38421684	0,59869868	1	0,48350355	-0,10270013	0,02360684
bathrooms	0,24358395	-0,07852130	0,66427945	0,75540254	0,48350355	1	-0,06518356	0,42624090
distance	0,02651322	0,03399334	-0,15318137	-0,11988583	-0,10270013	-0,06518356	1	-0,17625529
priceByArea	0,15477167	-0,03247983	0,77242896	0,38544488	0,02360684	0,42624090	-0,17625529	1

Point Two: House Price Model

House price model

Coefficients

	Estimate	Std. Error	t value	Pr(>mod(t))
(Intercept)	-3,43E+08	1,08E+07	-31.831	< 2e-16 ***
priceByArea	1,45E+05	2,36E+03	61.170	< 2e-16 ***
size	2,90E+06	5,69E+04	50.872	< 2e-16 ***
rooms	3,48E+06	2,59E+06	1.344	0.1791
bathrooms	-2,44E+07	4,34E+06	-5.612	2.38e-08 ***
floor	-4,35E+05	4,18E+05	-1.042	0.2977
numPhotos	-2,70E+05	1,99E+05	-1.358	0.1748
distance	2,07E+04	9,08E+03	2.280	0.0227 *

Point Two: House Price Model

House price model

Residuals

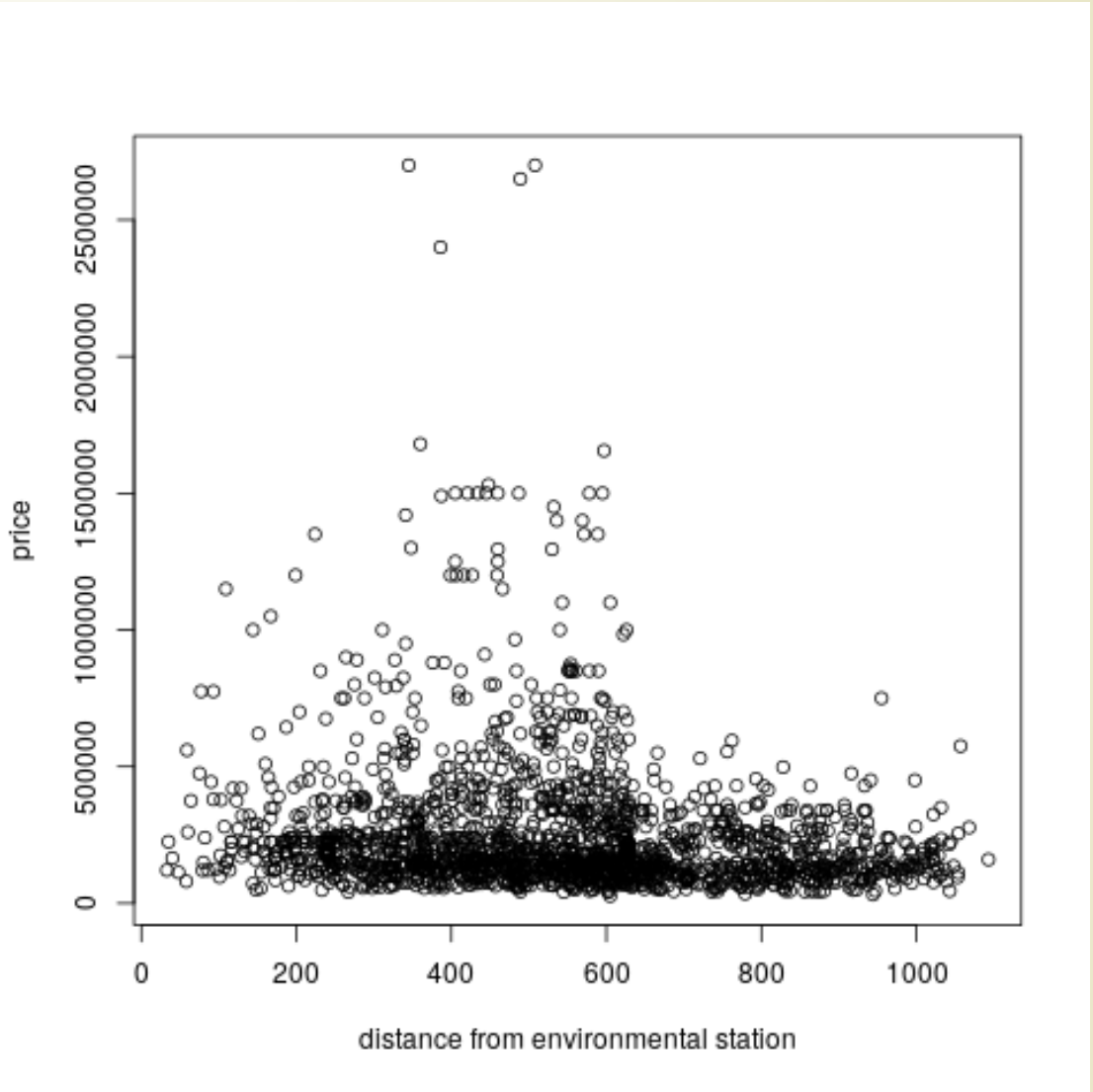
Min	1Q	Median	3Q	Max
-490213	-34060	-4722	27031	961095

- Residual standard error: 73830 on 1481 degrees of freedom
- Multiple R-squared: 0.9152
- Adjusted R-squared: 0.9148
- F-statistic: 2283 on 7 and 1481 DF
- p-value: < 2.2e-16

RMSE	MAE
71173.29 €	43755.22 €

Point Three: Pollution and Building's price

- Building distance from environmental stations



Point Three: Pollution and Building's price

	numPhotos	floor	price	size	rooms	bathrooms	distance	priceByArea	Ozono	NO	NO2	PM10	NOx
numPhotos	1,00000	0,02287	0,20013	0,21713	0,15806	0,24358	0,02651	0,15477	0,02990	-0,06847	-0,04437	-0,15718	-0,05964
floor	0,02287	1,00000	-0,06299	-0,06784	-0,11194	-0,07852	0,03399	-0,03248	-0,06403	0,06629	0,06215	0,05432	0,06501
price	0,20013	-0,06299	1,00000	0,81994	0,38422	0,66428	-0,15318	0,77243	0,30959	-0,33133	-0,33844	-0,30618	-0,33818
size	0,21713	-0,06784	0,81994	1,00000	0,59870	0,75540	-0,11989	0,38544	0,18312	-0,20133	-0,20499	-0,20772	-0,20523
rooms	0,15806	-0,11194	0,38422	0,59870	1,00000	0,48350	-0,10270	0,02361	0,03539	-0,02549	-0,03234	-0,01853	-0,02821
bathrooms	0,24358	-0,07852	0,66428	0,75540	0,48350	1,00000	-0,06518	0,42624	0,11056	-0,14843	-0,13334	-0,22660	-0,14418
distance	0,02651	0,03399	-0,15318	-0,11989	-0,10270	-0,06518	1,00000	-0,17626	-0,01981	-0,10732	-0,01485	-0,06602	-0,07118
priceByArea	0,15477	-0,03248	0,77243	0,38544	0,02361	0,42624	-0,17626	1,00000	0,37019	-0,39677	-0,39760	-0,37185	-0,40164
Ozono	0,02990	-0,06403	0,30959	0,18312	0,03539	0,11056	-0,01981	0,37019	1,00000	-0,95828	-0,98592	-0,24478	-0,97783
NO	-0,06847	0,06629	-0,33133	-0,20133	-0,02549	-0,14843	-0,10732	-0,39677	-0,95828	1,00000	0,96209	0,46244	0,99325
NO2	-0,04437	0,06215	-0,33844	-0,20499	-0,03234	-0,13334	-0,01485	-0,39760	-0,98592	0,96209	1,00000	0,33190	0,98717
PM10	-0,15718	0,05432	-0,30618	-0,20772	-0,01853	-0,22660	-0,06602	-0,37185	-0,24478	0,46244	0,33190	1,00000	0,41572
NOx	-0,05964	0,06501	-0,33818	-0,20523	-0,02821	-0,14418	-0,07118	-0,40164	-0,97783	0,99325	0,98717	0,41572	1,00000

Call: lm(formula = price ~ priceByArea + size + rooms + bathrooms + floor + numPhotos + distance + Ozono + NO + NO2 + PM10 + NOx, data = train)

Point Three: Pollution and Building's price

Coefficients

	Estimate	Std. Error	t value	Pr(>mod(t))
(Intercept)	-6,59E+08	1,36E+09	-0.484	0.62862
priceByArea	1,49E+05	2,50E+03	59.648	< 2e-16 ***
size	2,76E+06	5,66E+04	48.676	< 2e-16 ***
rooms	7,41E+06	2,49E+06	2.979	0.00294 **
bathrooms	-2,43E+07	4,22E+06	-5.767	9.8e-09 ***
floor	1,46E+04	4,05E+05	0.036	0.97131
numPhotos	-1,26E+05	1,92E+05	-0.654	0.51320
distance	1,76E+04	1,35E+04	1.312	0.18986
Ozono	4,40E+06	2,30E+07	0.191	0.84817
NO	1,22E+08	1,70E+08	0.715	0.47501
NO2	7,51E+07	1,09E+08	0.690	0.49015
PM10	2,19E+06	1,15E+06	1.913	0.05600 .
NOx	-7,63E+07	1,05E+08	-0.724	0.46924

Point Three: Pollution and Building's price

Residuals

Min	1Q	Median	3Q	Max
-430597	-31335	-5248	26248	952945

- Residual standard error: 71350 on 1476 degrees of freedom
- Multiple R-squared: 0,9194
- Adjusted R-squared: 0, 9188
- F-statistic: 1404 on 12 and 1476 DF
- p-value: < 2.2e-16

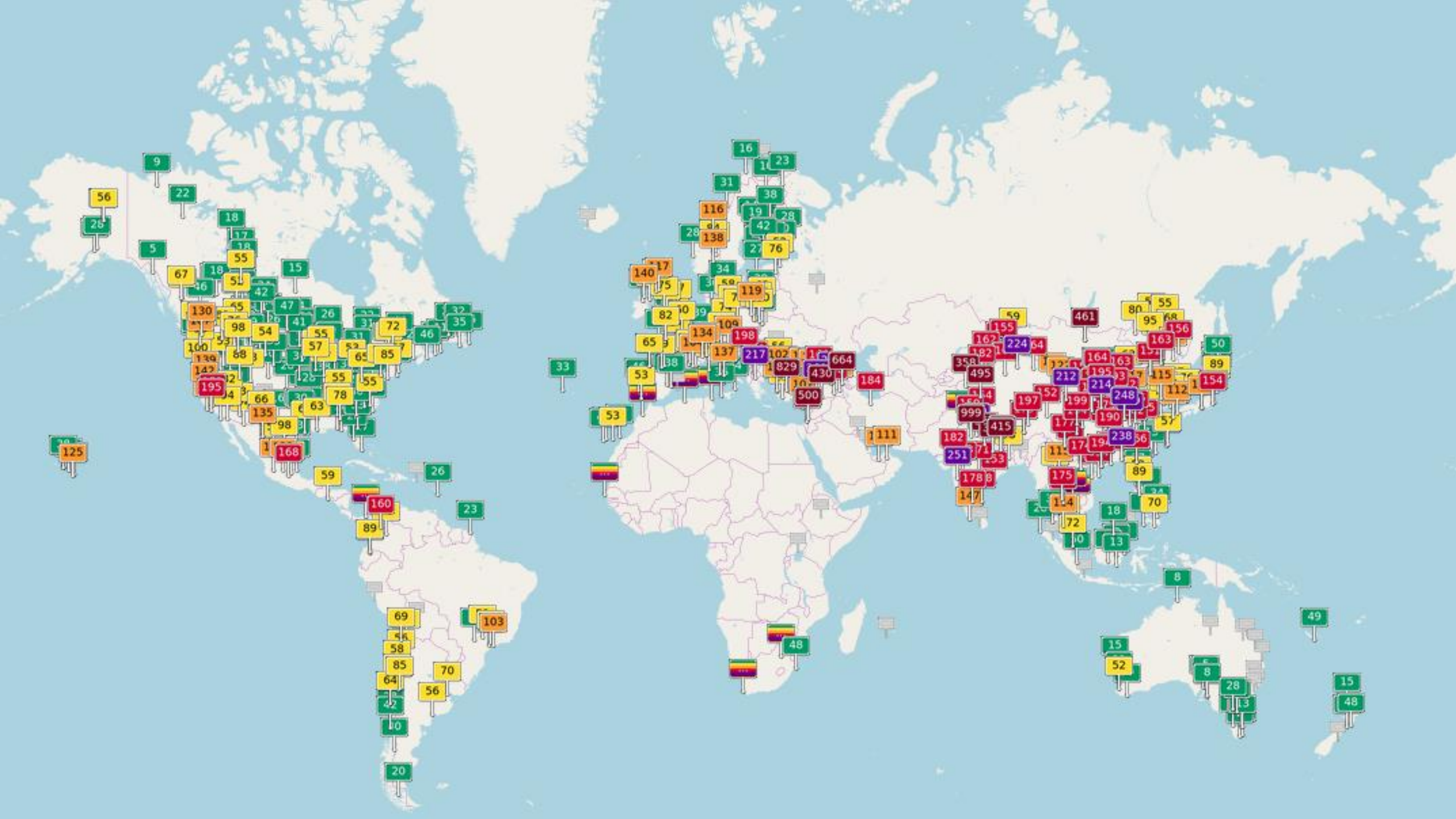
RMSE	MAE
65065.34858925	40183.70166710

Conclusions

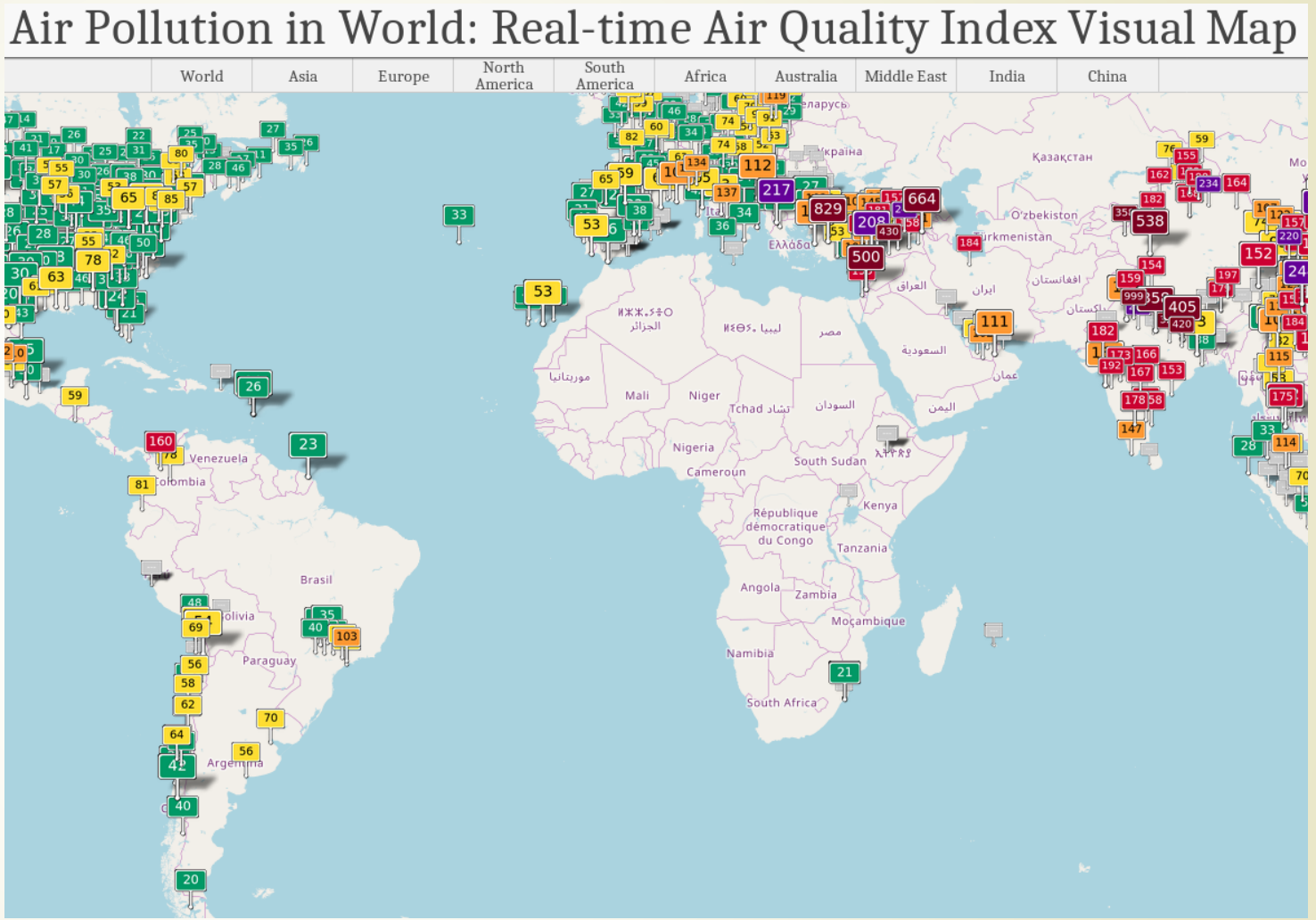
- As we could see, the corelation matrix doesn't give us a statistically clear correlation between contaminants and building's price.
- In the other hand we could see the regression model has been improved because the RMSE and the MAE have decreased, but not in a signifficant way.
- Air pollution seems not to be actually an imporant characteristic which influences in the price of buildings, but now we have the technology and the information to make this possible.
- The question is, would the prices of valencia's buildings decrease or increase? In this report we made a local comparison in air pollution, but when getting some global info, this could result in a increase of the price of valencia's buildings.

Global pollution

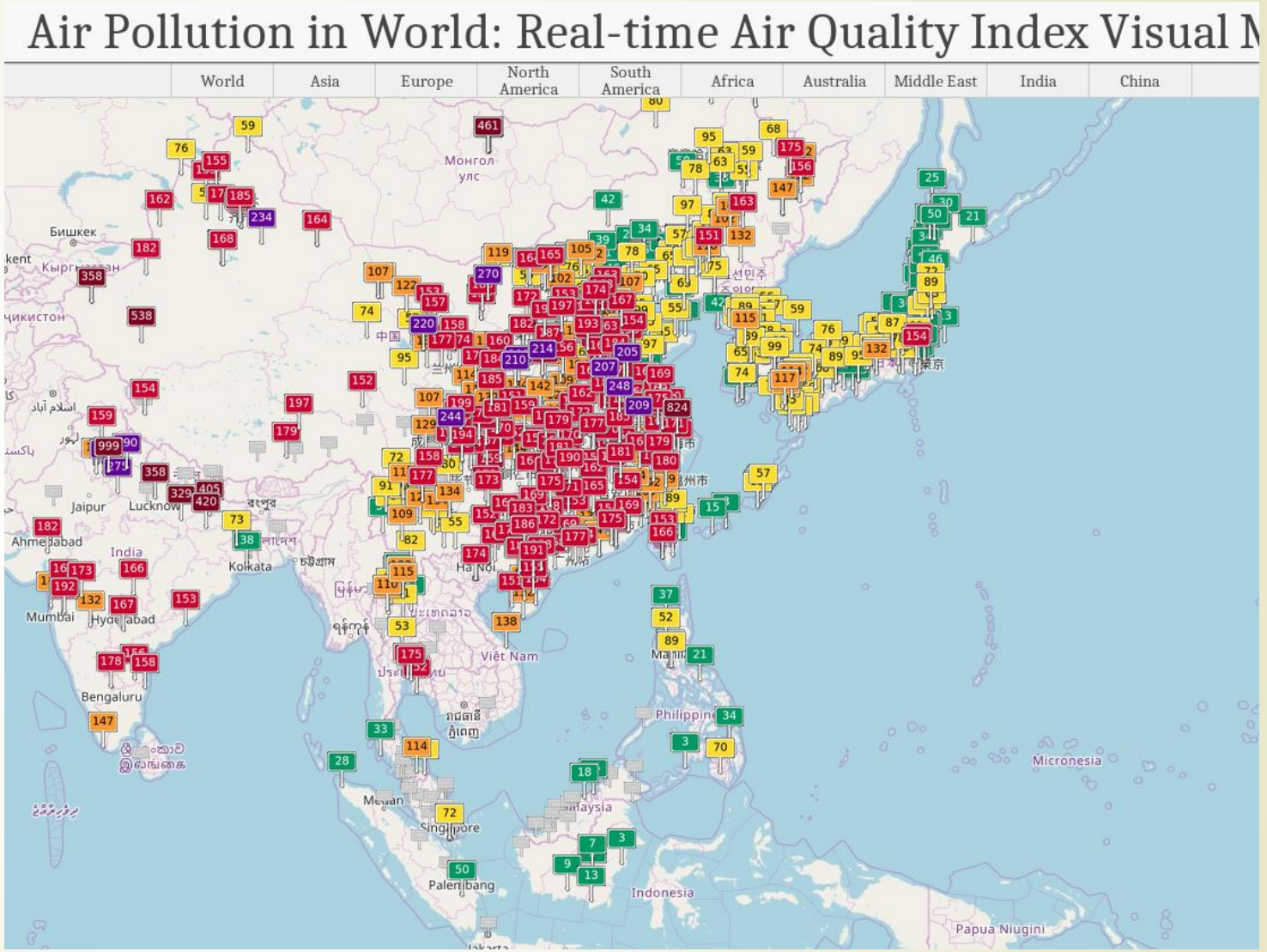
How much polluted are other cities arround the world



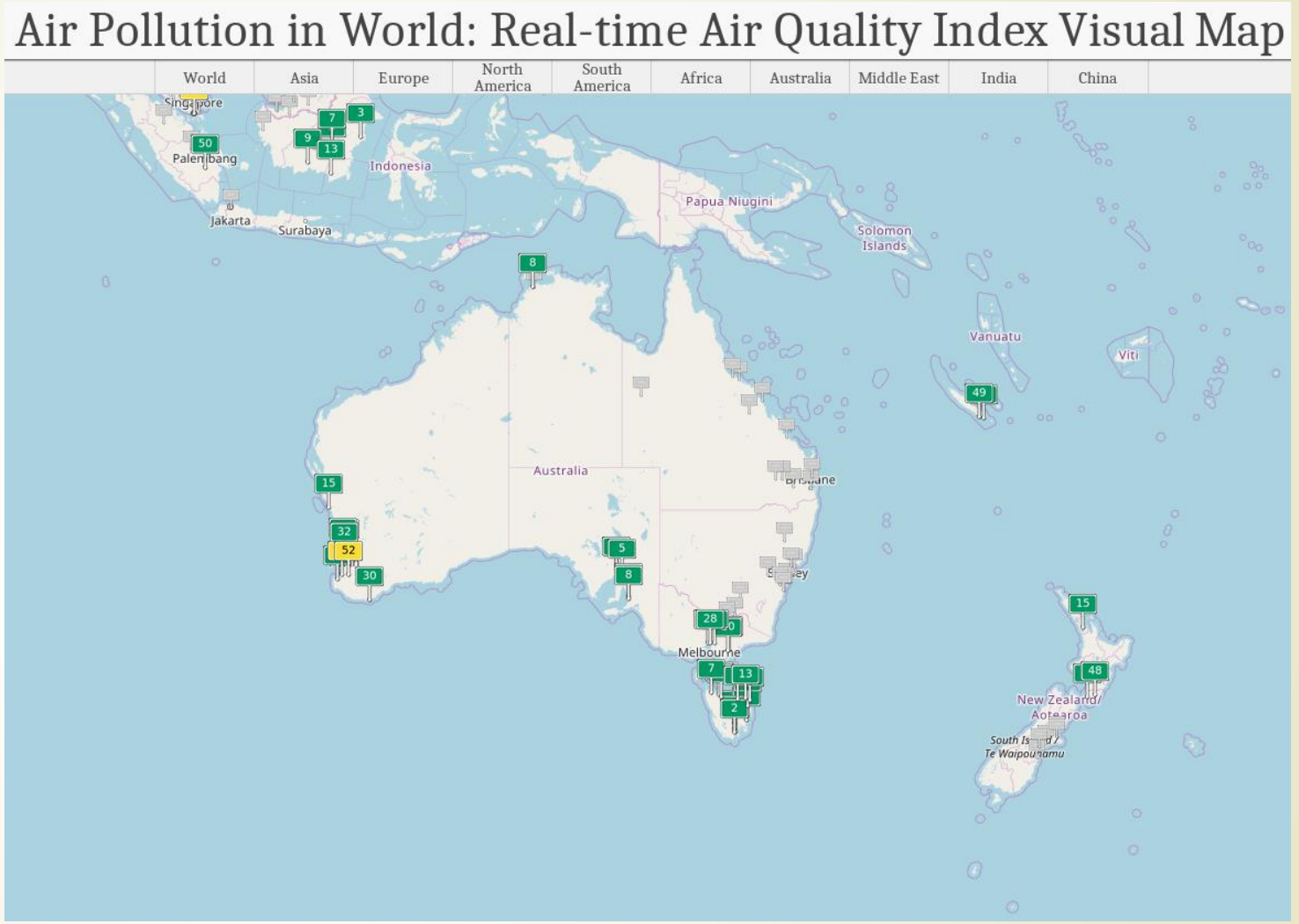
Africa


Air Pollution in World: Real-time Air Quality Index Visual Map
World
Asia
Europe
North America
South America
Africa
Australia
Middle East
India
China
Venezuela
Brasil
Paraguay
Argentina
Mali
Niger
Nigeria
Cameroun
South Sudan
Kenya
Tanzania
Angola
Zambia
Moçambique
Namibia
South Africa
Tchad
Казақстан
O'zbekiston
Ελλάδα
العراق
ایران
السعودية
اليمن
مصر
ليبيا
الجزائر
موريتانيا
السودان
عمان
افغانستان
پاکستان

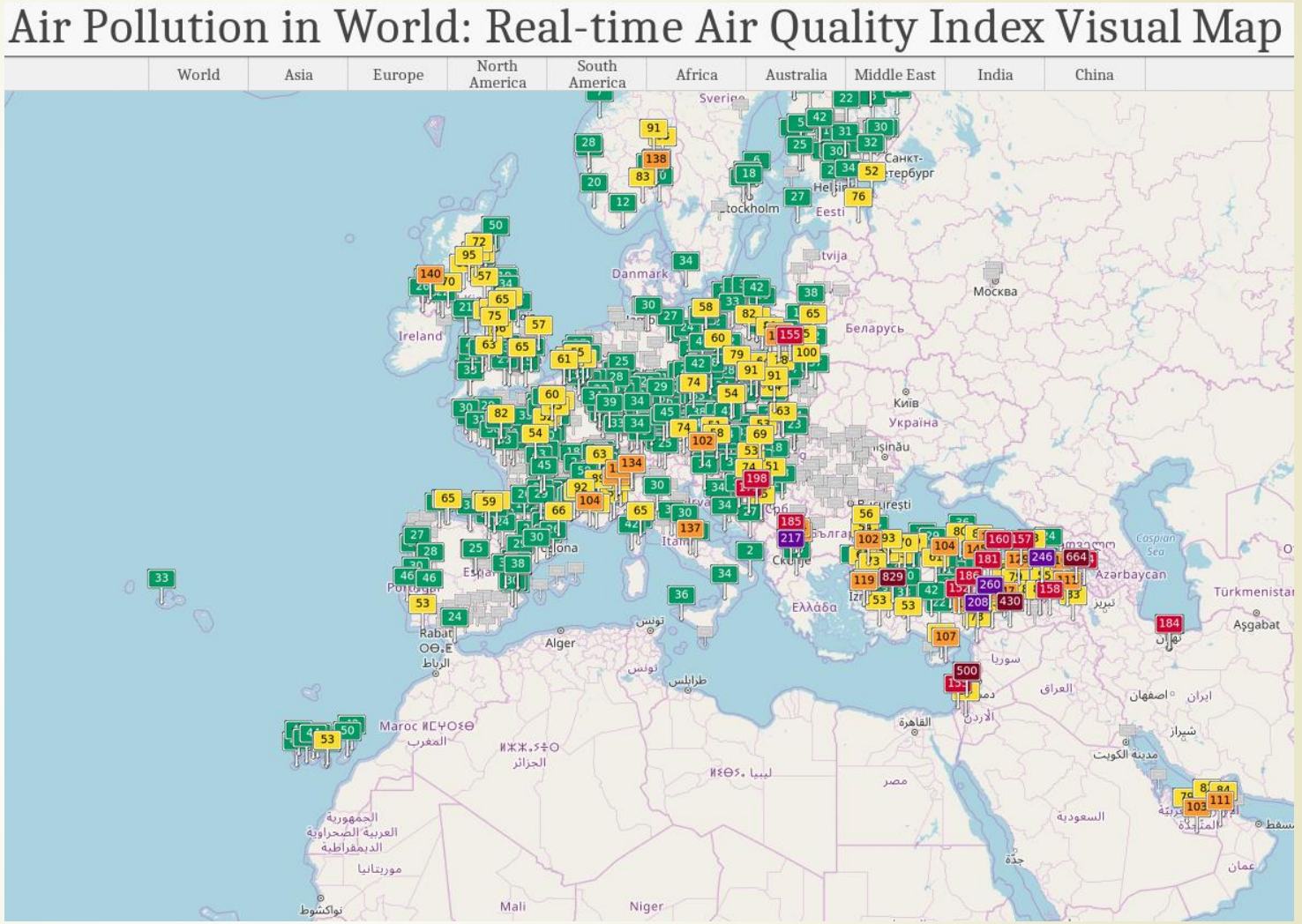
Asia


Air Pollution in World: Real-time Air Quality Index Visual M
World
Asia
Europe
North America
South America
Africa
Australia
Middle East
India
China
Монгол улс
Бишкек
Jaipur
Lucknow
Ahmedabad
India
Kolkata
Mumbai
Bengaluru
Viêt Nam
Philippines
Micronesia
Malaysia
Singapore
Palembang
Indonesia
Papua Niugini

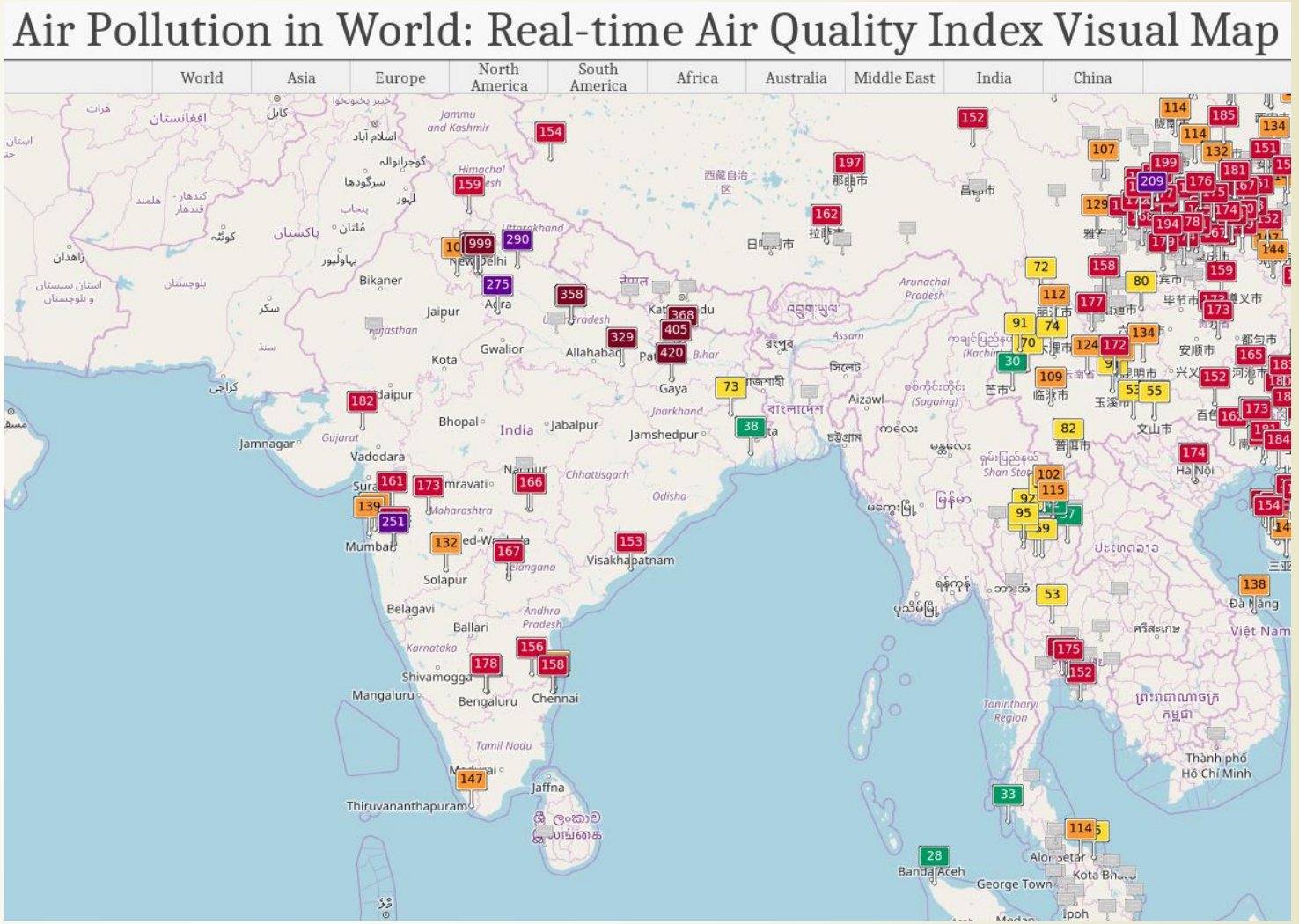
Australia


Air Pollution in World: Real-time Air Quality Index Visual Map
World
Asia
Europe
North America
South America
Africa
Australia
Middle East
India
China
Singapore
Palembang
Jakarta
Surabaya
Indonesia
Papua Niugini
Solomon Islands
Vanuatu
Viti
Australia
Melbourne
New Zealand/ Aotearoa
50
7
3
9
13
8
49
15
32
52
30
5
8
28
7
13
2
15
48

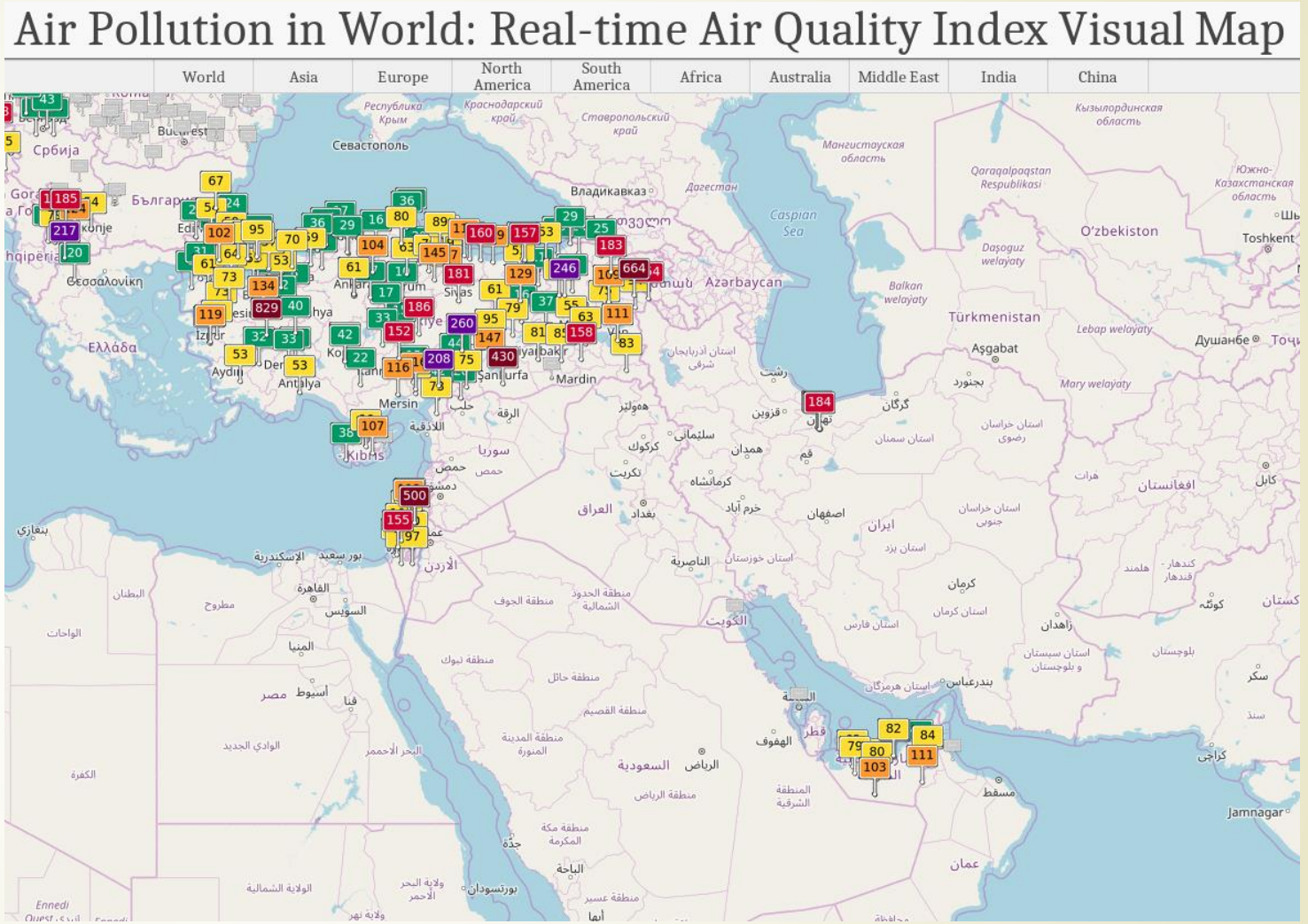
Europe


Air Pollution in World: Real-time Air Quality Index Visual Map
World
Asia
Europe
North America
South America
Africa
Australia
Middle East
India
China
Ireland
Danmark
Eesti
Беларусь
Україна
Москва
Санкт-Петербург
Stockholm
Rabat
Alger
تونس
طرابلس
Maroc
المغرب
الجزائر
ليبيا
مصر
القاهرة
الأردن
سوريا
العراق
ایران
اصفهان
شیراز
السعودية
جدة
عمان
Azarbaycan
Türkmenistan
Aşgabat
Ελλάδα
الجمهورية العربية الصحراوية الديمقراطية
موريتانيا
نواكشوط
Mali
Niger

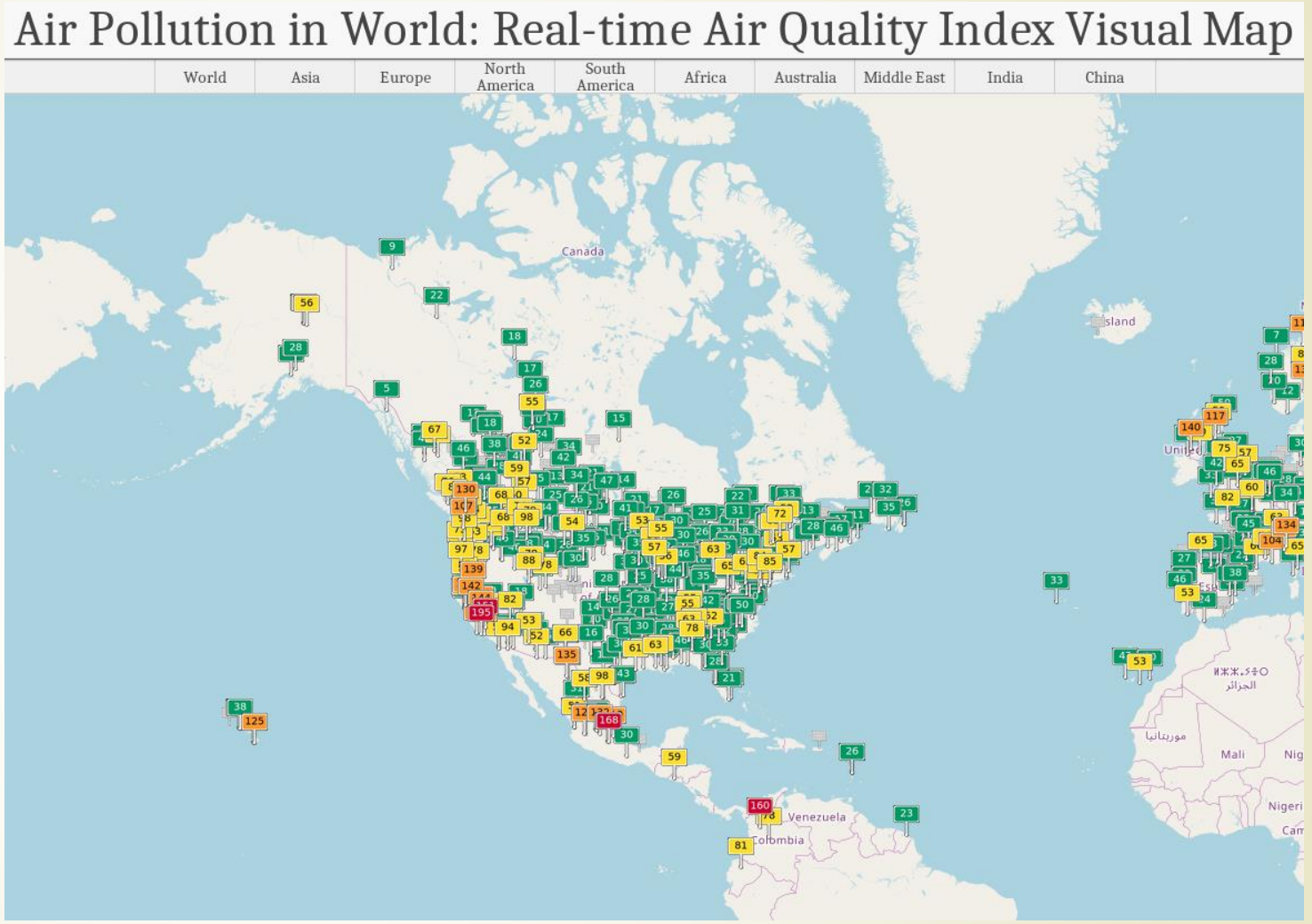
India


Air Pollution in World: Real-time Air Quality Index Visual Map
World
Asia
Europe
North America
South America
Africa
Australia
Middle East
India
China
Jammu and Kashmir
Himachal Pradesh
Uttarakhand
New Delhi
Bikaner
Jaipur
Agra
Rajasthan
Kota
Gwalior
Allahabad
Gaya
Jharkhand
Bihar
Assam
Arunachal Pradesh
Aizawl
Jamnagar
Gujarat
Vadodara
Bhopal
India
Jabalpur
Jamshedpur
Chhattisgarh
Odisha
Maharashtra
Mumbai
Solapur
Visakhapatnam
Telangana
Andhra Pradesh
Belagavi
Ballari
Karnataka
Shivamogga
Mangaluru
Bengaluru
Chennai
Tamil Nadu
Jaffna
Thiruvananthapuram
Banda Aceh
George Town
Ipoh
Hà Nội
Việt Nam
Thành phố Hồ Chí Minh
Tanintharyi Region

Midle East


Air Pollution in World: Real-time Air Quality Index Visual Map
World
Asia
Europe
North America
South America
Africa
Australia
Middle East
India
China

North America



South America

Thanks for your attention!

Óscar Muñoz Garrigós

Stanislav Sitanskiy Igorevich

Carlos Frias Ruiz