Mexico\_city\_Toronto\_comparison

James Johnson

February 1, 2020

## Assignment 2

* Identify a highly polluted city in the world (or small geographic region)
* Justify your selection of being highly polluted based on available data and standards/resources -Describe condisitons for several air pollutants
* levesl, seasonalily, responsible sources, public awareness, current efforts the responsible government is taking to improve air quality.
* Compare levels in that highly polluted place with Toronto for PM2.5, NO2, and O3

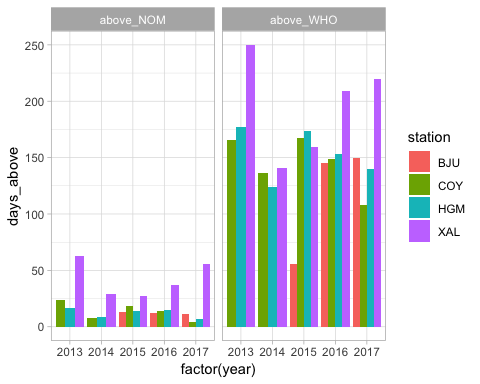
## Mexico City

### Topography

Mexico City is in an elevated basin surrounded on three sides by mountain ridges on to the east, south and west. There is a broad opening to the north and a narrow gap to the south-southwest. (Harvard) Two major volcanoes are located in the southeast. Within the basin, the mountains and frequent thermal inversions trap pollutants within the basin. The high altitude also makes combustion sources less efficient.

### Air quality

day\_Mc <- MC\_data %>%  
 mutate(year = year(date),  
 trimes = floor\_date(date, unit = "3 months"),  
 month = floor\_date(date,unit = "1 month"),  
 day = floor\_date(date, unit = "1 day")) %>%   
 filter(station == "XAL"|station == "COY"|station == "BJU"|station == "HGM") %>%  
 filter(PM25 >0) %>%   
 group\_by(station, year, trimes, day) %>%   
 summarize(mean = mean(PM25,na.rm = T),quint\_75 = quantile(PM25, 0.75,na.rm = T),quint\_98 = quantile(PM25, 0.98,na.rm = T)) %>%   
 ungroup()  
  
days\_above <- day\_Mc %>%  
 mutate(above\_25 = mean > 25,  
 above\_45 = mean > 45) %>%   
 group\_by(station, year) %>%   
 summarize(above\_WHO = sum(above\_25), above\_NOM = sum(above\_45)) %>%   
 ungroup() %>%   
 gather(standard,days\_above, above\_WHO:above\_NOM)  
  
ggplot(days\_above,aes(factor(year),days\_above, fill = station)) + geom\_bar(stat = "identity", position = "dodge") + facet\_wrap(~standard) + theme\_light()



Mc\_24\_hr <- day\_Mc %>%   
 group\_by(station, year) %>%  
 summarize(quant\_75 = quantile(mean, 0.75)) %>%   
 spread(station,quant\_75) %>%   
 gt()   
Mc\_24\_hr

<tr>  
 <th class="gt\_col\_heading gt\_columns\_bottom\_border gt\_right" rowspan="1" colspan="1">year</th>  
 <th class="gt\_col\_heading gt\_columns\_bottom\_border gt\_right" rowspan="1" colspan="1">BJU</th>  
 <th class="gt\_col\_heading gt\_columns\_bottom\_border gt\_right" rowspan="1" colspan="1">COY</th>  
 <th class="gt\_col\_heading gt\_columns\_bottom\_border gt\_right" rowspan="1" colspan="1">HGM</th>  
 <th class="gt\_col\_heading gt\_columns\_bottom\_border gt\_right" rowspan="1" colspan="1">XAL</th>  
</tr>

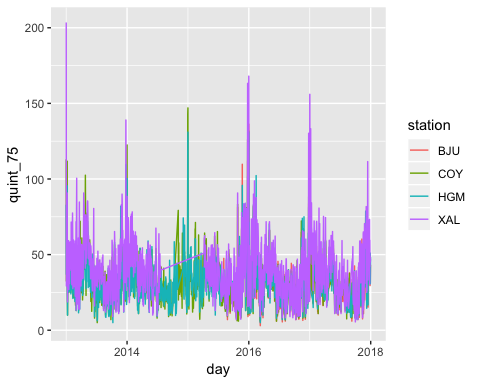
<tr>  
 <td class="gt\_row gt\_right">2013</td>  
 <td class="gt\_row gt\_right">NA</td>  
 <td class="gt\_row gt\_right">32.50000</td>  
 <td class="gt\_row gt\_right">32.70833</td>  
 <td class="gt\_row gt\_right">41.94792</td>  
</tr>  
<tr>  
 <td class="gt\_row gt\_right gt\_striped">2014</td>  
 <td class="gt\_row gt\_right gt\_striped">NA</td>  
 <td class="gt\_row gt\_right gt\_striped">29.47917</td>  
 <td class="gt\_row gt\_right gt\_striped">29.65341</td>  
 <td class="gt\_row gt\_right gt\_striped">40.70833</td>  
</tr>  
<tr>  
 <td class="gt\_row gt\_right">2015</td>  
 <td class="gt\_row gt\_right">30.40217</td>  
 <td class="gt\_row gt\_right">33.95833</td>  
 <td class="gt\_row gt\_right">32.19792</td>  
 <td class="gt\_row gt\_right">36.85417</td>  
</tr>  
<tr>  
 <td class="gt\_row gt\_right gt\_striped">2016</td>  
 <td class="gt\_row gt\_right gt\_striped">29.75000</td>  
 <td class="gt\_row gt\_right gt\_striped">30.54167</td>  
 <td class="gt\_row gt\_right gt\_striped">30.97917</td>  
 <td class="gt\_row gt\_right gt\_striped">36.06250</td>  
</tr>  
<tr>  
 <td class="gt\_row gt\_right">2017</td>  
 <td class="gt\_row gt\_right">30.44792</td>  
 <td class="gt\_row gt\_right">30.41667</td>  
 <td class="gt\_row gt\_right">30.97917</td>  
 <td class="gt\_row gt\_right">41.57292</td>  
</tr>

Mc\_annual <- day\_Mc %>%   
 group\_by(station, year,trimes) %>%  
 summarize(mean\_quint = quantile(mean,0.75, na.rm = T)) %>%   
 ungroup() %>%   
 group\_by(station,year) %>%   
 summarize(annual\_mean = mean(mean\_quint, na.rm = T)) %>%   
 spread(station, annual\_mean) %>%   
 gt()  
Mc\_annual

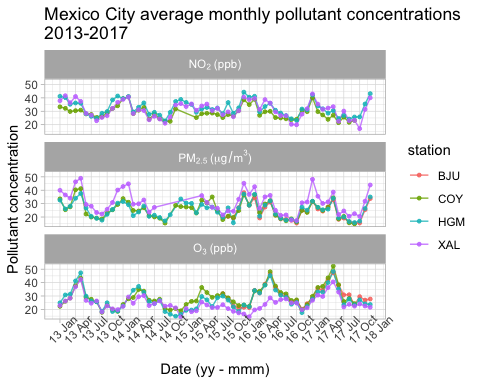
<tr>  
 <th class="gt\_col\_heading gt\_columns\_bottom\_border gt\_right" rowspan="1" colspan="1">year</th>  
 <th class="gt\_col\_heading gt\_columns\_bottom\_border gt\_right" rowspan="1" colspan="1">BJU</th>  
 <th class="gt\_col\_heading gt\_columns\_bottom\_border gt\_right" rowspan="1" colspan="1">COY</th>  
 <th class="gt\_col\_heading gt\_columns\_bottom\_border gt\_right" rowspan="1" colspan="1">HGM</th>  
 <th class="gt\_col\_heading gt\_columns\_bottom\_border gt\_right" rowspan="1" colspan="1">XAL</th>  
</tr>

<tr>  
 <td class="gt\_row gt\_right">2013</td>  
 <td class="gt\_row gt\_right">NA</td>  
 <td class="gt\_row gt\_right">33.34511</td>  
 <td class="gt\_row gt\_right">32.05529</td>  
 <td class="gt\_row gt\_right">40.09209</td>  
</tr>  
<tr>  
 <td class="gt\_row gt\_right gt\_striped">2014</td>  
 <td class="gt\_row gt\_right gt\_striped">NA</td>  
 <td class="gt\_row gt\_right gt\_striped">29.28733</td>  
 <td class="gt\_row gt\_right gt\_striped">29.53727</td>  
 <td class="gt\_row gt\_right gt\_striped">38.15458</td>  
</tr>  
<tr>  
 <td class="gt\_row gt\_right">2015</td>  
 <td class="gt\_row gt\_right">29.37749</td>  
 <td class="gt\_row gt\_right">32.57552</td>  
 <td class="gt\_row gt\_right">33.90208</td>  
 <td class="gt\_row gt\_right">37.26389</td>  
</tr>  
<tr>  
 <td class="gt\_row gt\_right gt\_striped">2016</td>  
 <td class="gt\_row gt\_right gt\_striped">29.04427</td>  
 <td class="gt\_row gt\_right gt\_striped">29.94531</td>  
 <td class="gt\_row gt\_right gt\_striped">30.45194</td>  
 <td class="gt\_row gt\_right gt\_striped">36.26036</td>  
</tr>  
<tr>  
 <td class="gt\_row gt\_right">2017</td>  
 <td class="gt\_row gt\_right">29.88281</td>  
 <td class="gt\_row gt\_right">29.74558</td>  
 <td class="gt\_row gt\_right">30.37344</td>  
 <td class="gt\_row gt\_right">39.73130</td>  
</tr>

ggplot(day\_Mc, aes(day,quint\_75,col = station)) + geom\_line()



monthly <- MC\_data %>%  
 mutate(year = year(date),  
 month = floor\_date(date,unit = "1 month"))%>%   
 filter(station == "XAL"|station == "COY"|station == "BJU"|station == "HGM") %>%  
 gather(pollutant,value, c(NO2:O3)) %>%   
 filter(value > 0) %>%   
 group\_by(year,month,pollutant,station) %>%   
 summarize(mean = mean(value)) %>%   
 ungroup %>%   
 mutate(month2 = as.Date(paste0("2013-", month,"-01"),"%Y-%m-%d")) %>%   
 mutate(pollutant = factor(pollutant, levels = c("NO2","PM25","O3"), labels = MC.poll.labs ))  
  
  
ggplot(monthly, aes(month,mean, col = station)) +  
 geom\_path(alpha = 0.8) +  
 geom\_point(size = 1, alpha = 0.8) +  
 facet\_wrap(~pollutant,nrow = 3, labeller = label\_parsed) +  
 scale\_x\_date(date\_breaks = "3 months",date\_labels = "%y %b") +  
 theme\_light() +  
 theme(axis.text.x = element\_text(angle = 45))+  
 labs(title = "Mexico City average monthly pollutant concentrations\n2013-2017", y = "Pollutant concentration", x = "Date (yy - mmm)")



MC\_data %>%   
 filter(!is.na(PM25)) %>%  
 filter(PM25>0) %>%   
 count(station)

## # A tibble: 19 x 2  
## station n  
## <chr> <int>  
## 1 AJM 23928  
## 2 BJU 20343  
## 3 CAM 34904  
## 4 CCA 27519  
## 5 COY 37697  
## 6 HGM 37610  
## 7 INN 14660  
## 8 MER 37787  
## 9 MGH 24927  
## 10 MPA 13928  
## 11 NEZ 39035  
## 12 PED 22301  
## 13 SAG 36563  
## 14 SFE 28693  
## 15 SJA 21355  
## 16 TLA 36723  
## 17 UAX 40041  
## 18 UIZ 29279  
## 19 XAL 34615