**Module 2 Report**

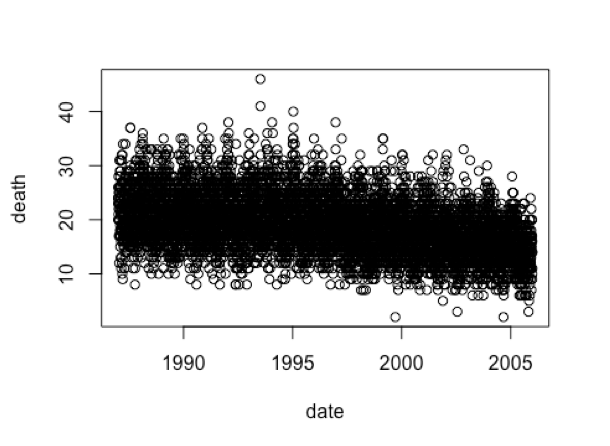
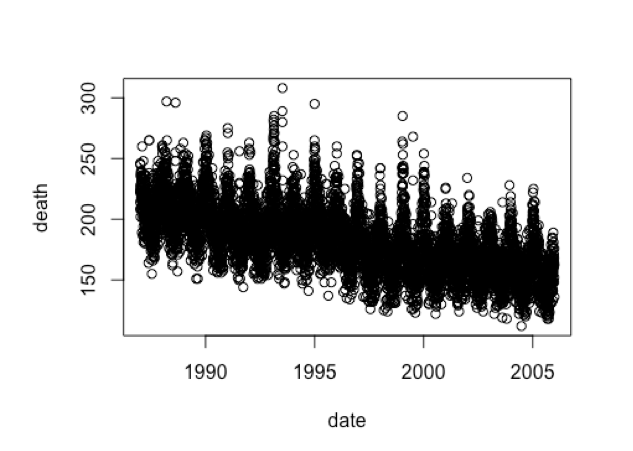
Module 2 emphasized what specific models we can select from to better predict relationships between mortality and air pollution in the form of particulate matter (PM) across four cities of interest. In our group, we chose Seattle, Los Angeles, New York, and Baltimore. Compared to Module 1, in which our data was centered on health outcomes for those who do and do not smoke, results this time around were more about accounting for a variety of variables and determining the “best” model.

We began with answering the first question of the three,

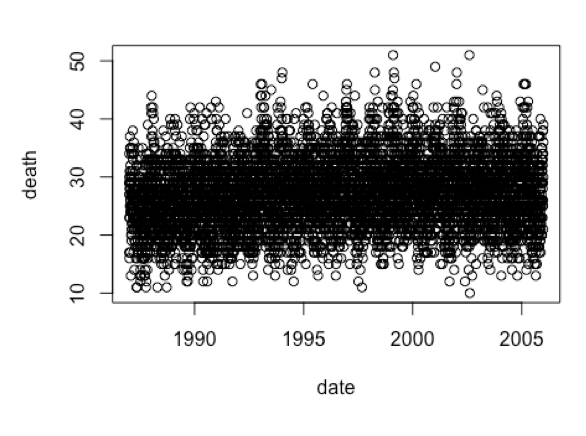
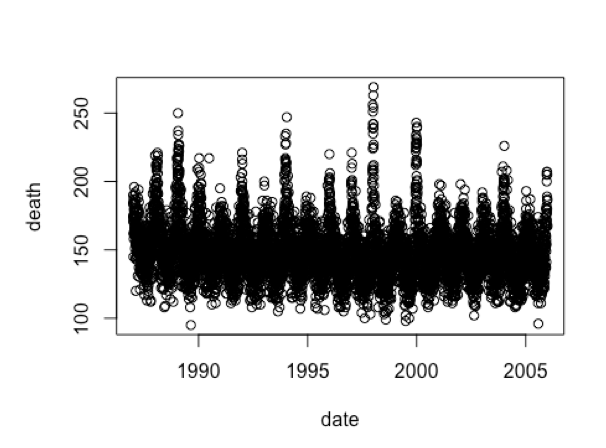
**Q1: How does the daily risk of death depend upon air pollution level in American cities?**

We began our general investigation with an overview of death across time (labeled ‘date’ on the x-axis) for our four cities.

Baltimore: New York:

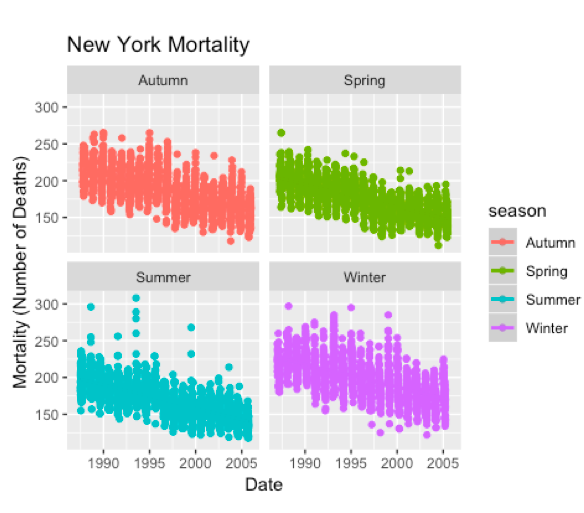
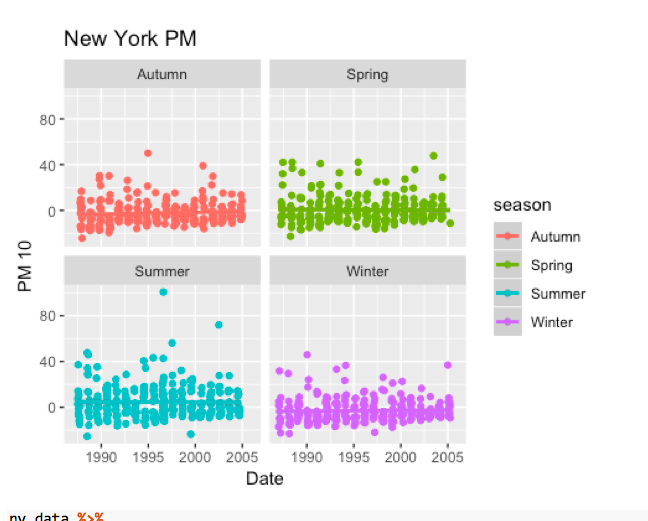
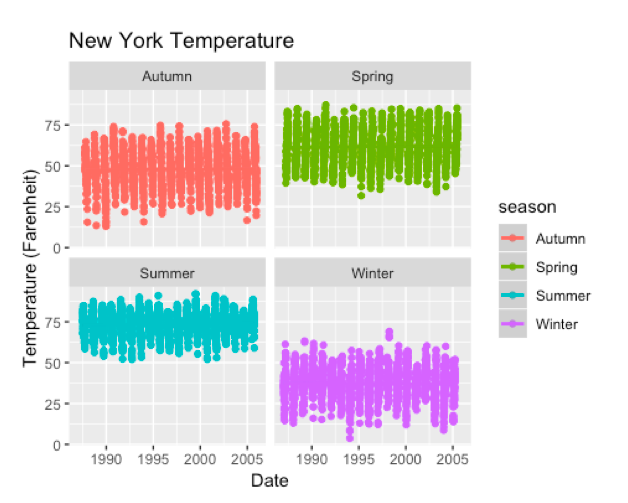


Seattle : LA:



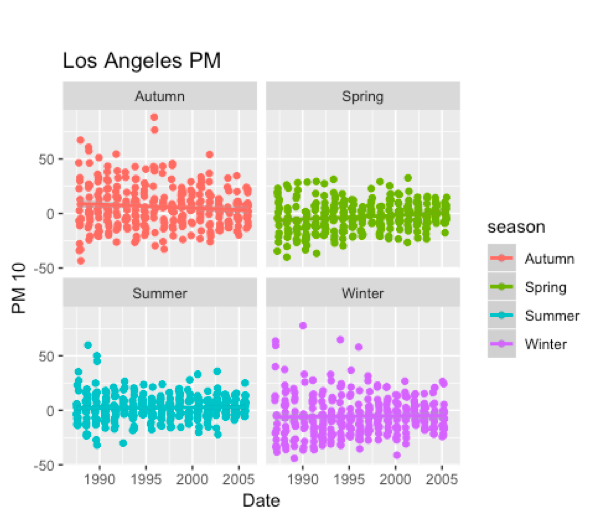
Now, equipped with an idea of the general mortality trends for context, we further analyzed the ways in which air pollution impacted death. This was done by analyzing PM10 levels and temperature alongside the mortality rates for the four cities.

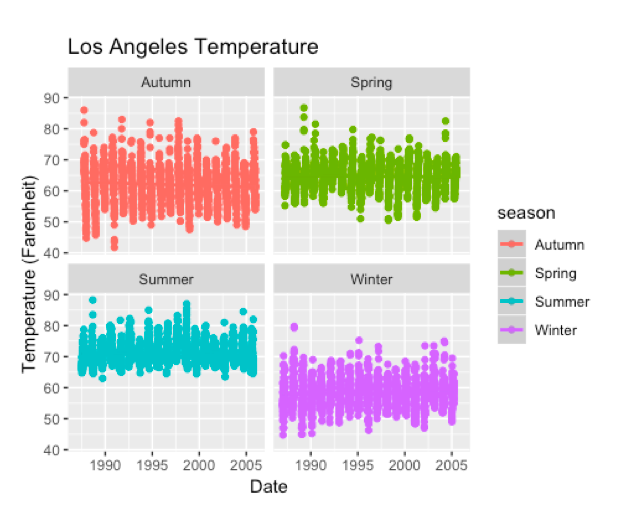
New York:

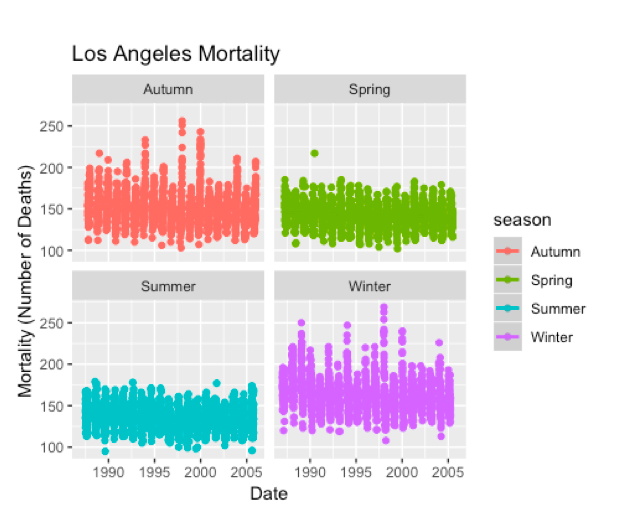


**(PM10)** In the top left corner, there is a chart that shows the PM10 for the city of New York across the 4 seasons (Autumn, Spring, Summer, Winter) from a time span of 20 years (1985 to 2005). Throughout the seasons, the PM10 level stays between a range of approximately -20 to 50 and are relatively consistent amongst the cities. Autumn seems to consistently have PM10 levels that most consistently hover around 0. On the other hand, Summer appears to be the season with the highest variation amongst PM10 levels throughout the decades. **(Mortality)** For mortality, the number of deaths per year typically ranges between 150 and 250 with some outliers across the seasons. It appears that spring has the most consistent mortality rate and while summer has a high variability when it comes to number of deaths. However, even though the number of deaths remains relatively consistent across the seasons, it is clear that the highest number of deaths occurs during the winter. **(Temperature)** Understandably so, it seems that the highest temperatures are recorded during the summer months with temperatures ranging from high 50s to high 90s and averaging around 75 degrees Fahrenheit. Contrarily, the lowest temperatures are found during the winter months with temperatures ranging from 0 to 50 and averaging around 35 degrees Fahrenheit. This is consistent with what we understand to be the temperature conditions in NYC during these times of the year. **(Inferences)** Given this information, one can deduce that the number of deaths goes down as the temperature increases.

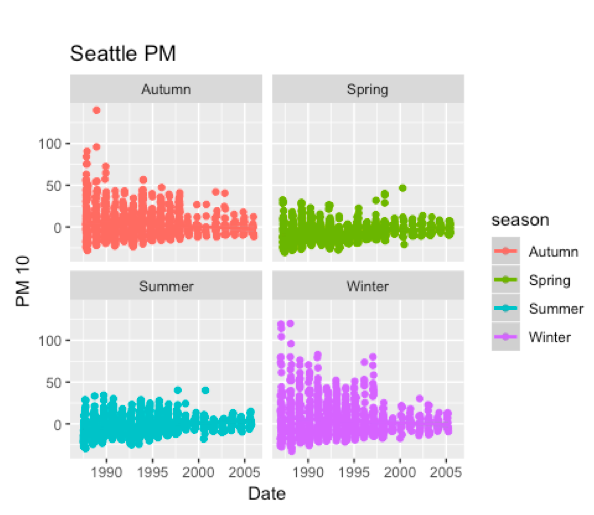
LA:

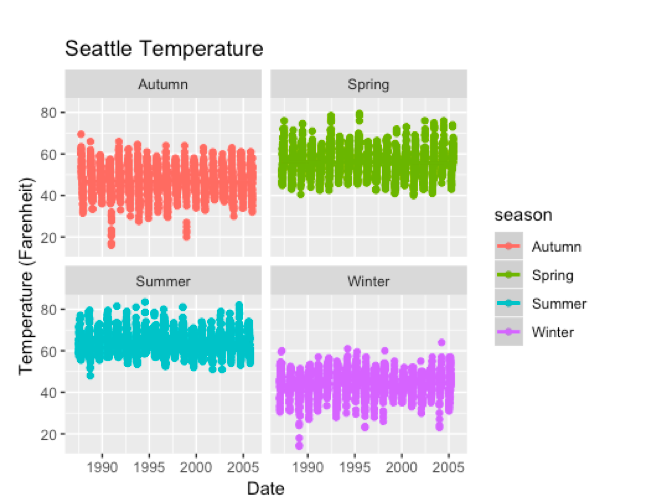


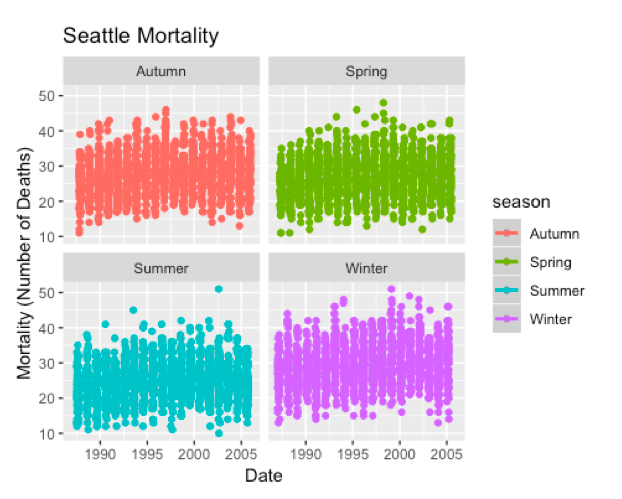


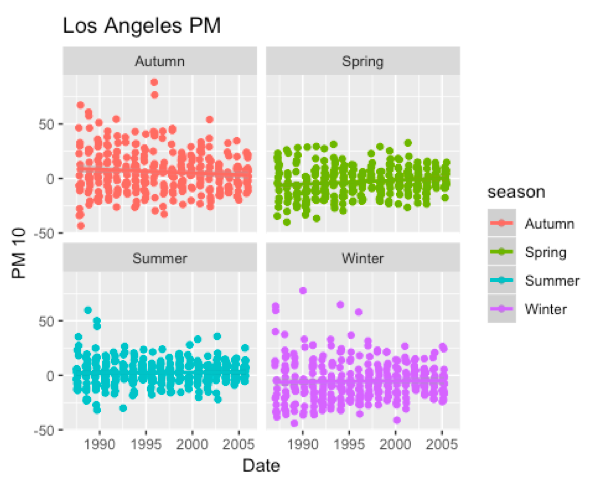


**(PM10)** For the most part, in the city of Los Angeles, PM10 levels typically vary between -25 and 25 across the seasons from 1985 to 2005. However, in Autumn and somewhat during winter, the PM10 values seem to deviate from this average more consistently (reaching values of 75 and above). Summer seems to have the most consistent values while spring has the values closest to 0 on average. **(Temperature)** Again, as in NYC, we see that summer has the highest temperatures (ranging from 65 to 90 degrees Fahrenheit) while winter has the lowest temperatures (40 to 75 degrees Fahrenheit). Summer has the most consistent temperatures across the seasons in contrast to Autumn’s high variability temperature recordings. **(Mortality)** Mortality appears to be particularly high in winter and autumn months and low in the summer. **(Inferences)** Again, as with NYC, the number of deaths seems to decrease with temperature increase. This time, we can see that PM10 levels that are closest to 0 (like during summer) coincides with lower death rates.

Seattle:

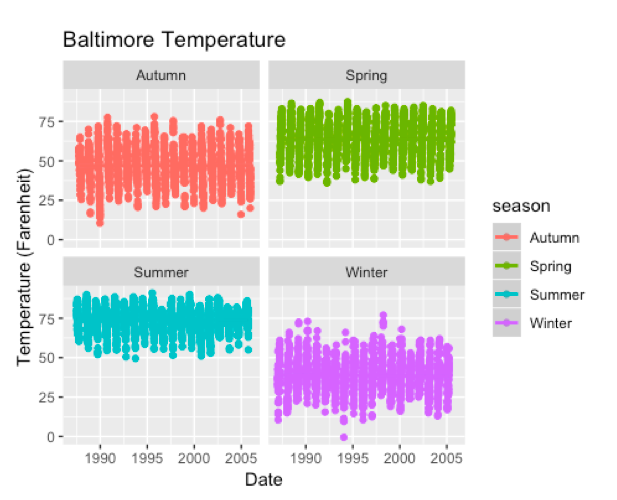
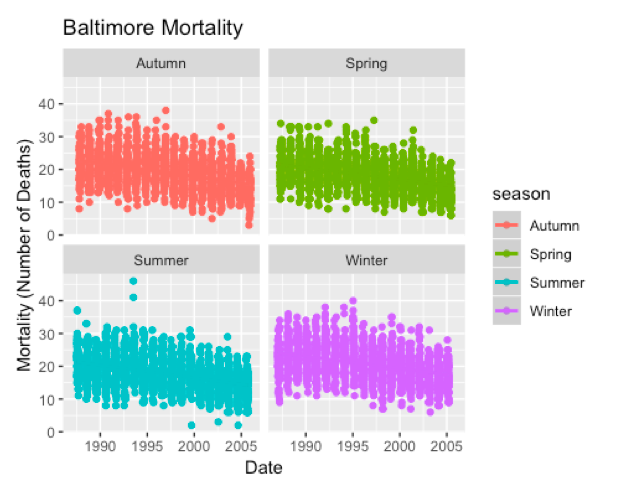
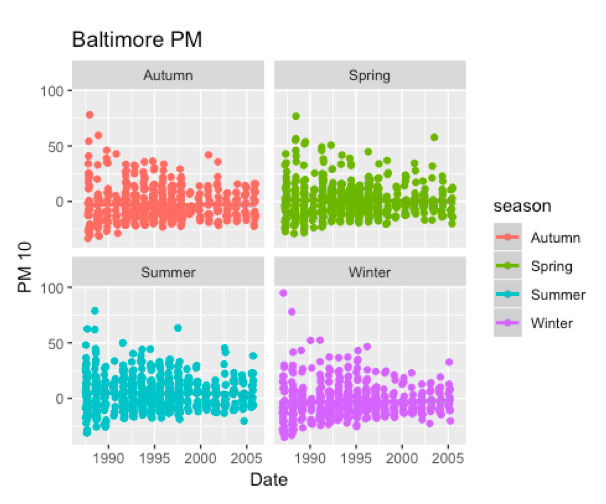






**(PM10)** Over the years, it appears that overall, PM10 levels decrease across the seasons. However, they are the most consistent and close to 0 during the summer months. Contrarily, they have the most variability during the winter time. **(Temperature)** Temperature trends in Seattle from 1985 to 2005 across the seasons mirror trends seen in NYC and LA and are consistent with what we understand to be true about temperature. **(Mortality)** Mortality across the seasons is very consistent amongst one another. However, the overall deaths are much lower than NYC and LA (possibly due to a lower population). If one were to determine a season with the lowest mortality, it could be argued that summer would be that season. The highest mortality rate would be more difficult to determine as the other 3 seasons have such comparable numbers of deaths. **(Inferences)** Again, the ~0 PM10 levels and high temperatures of summer appear to coincide with lower mortality rates.

Baltimore:

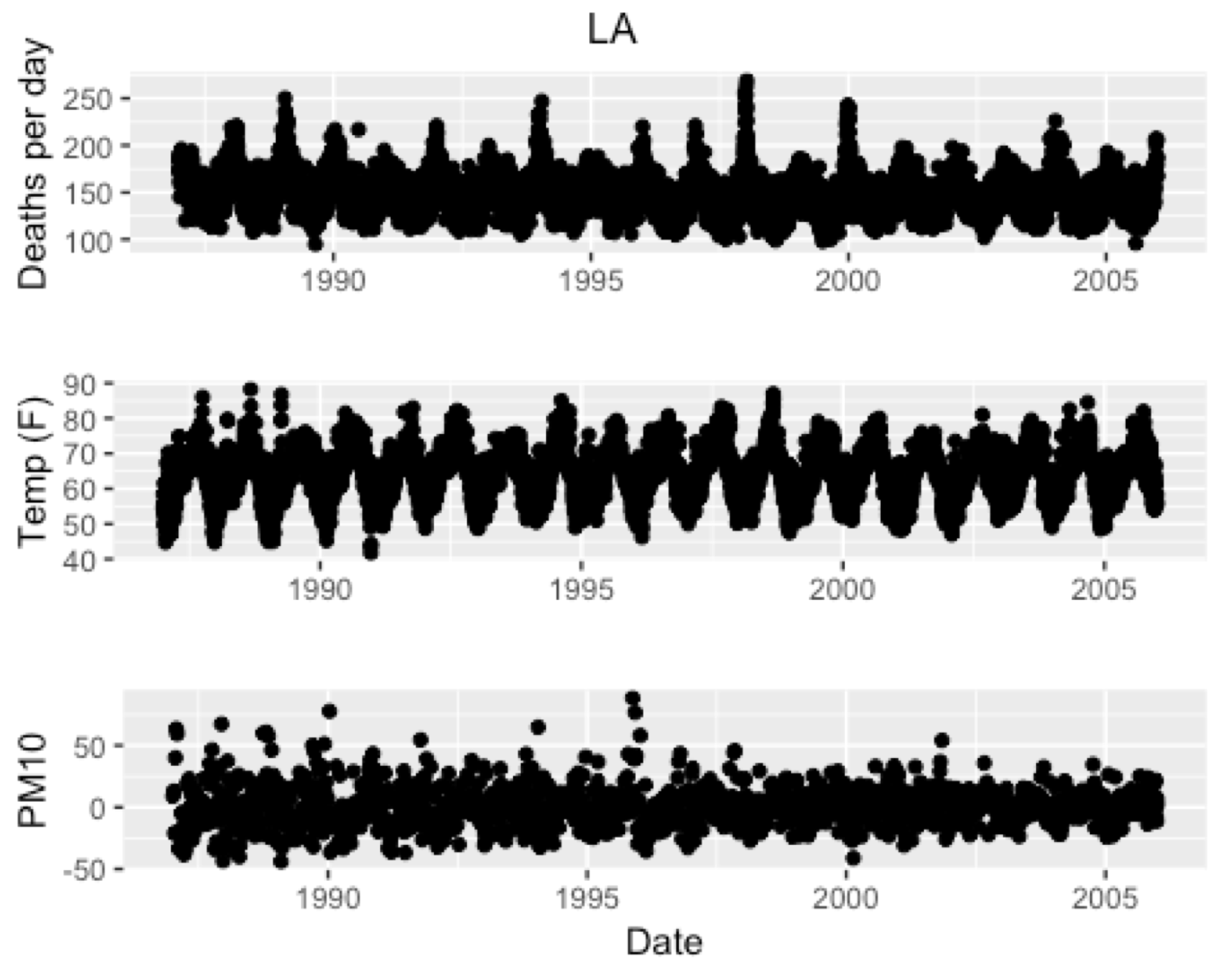
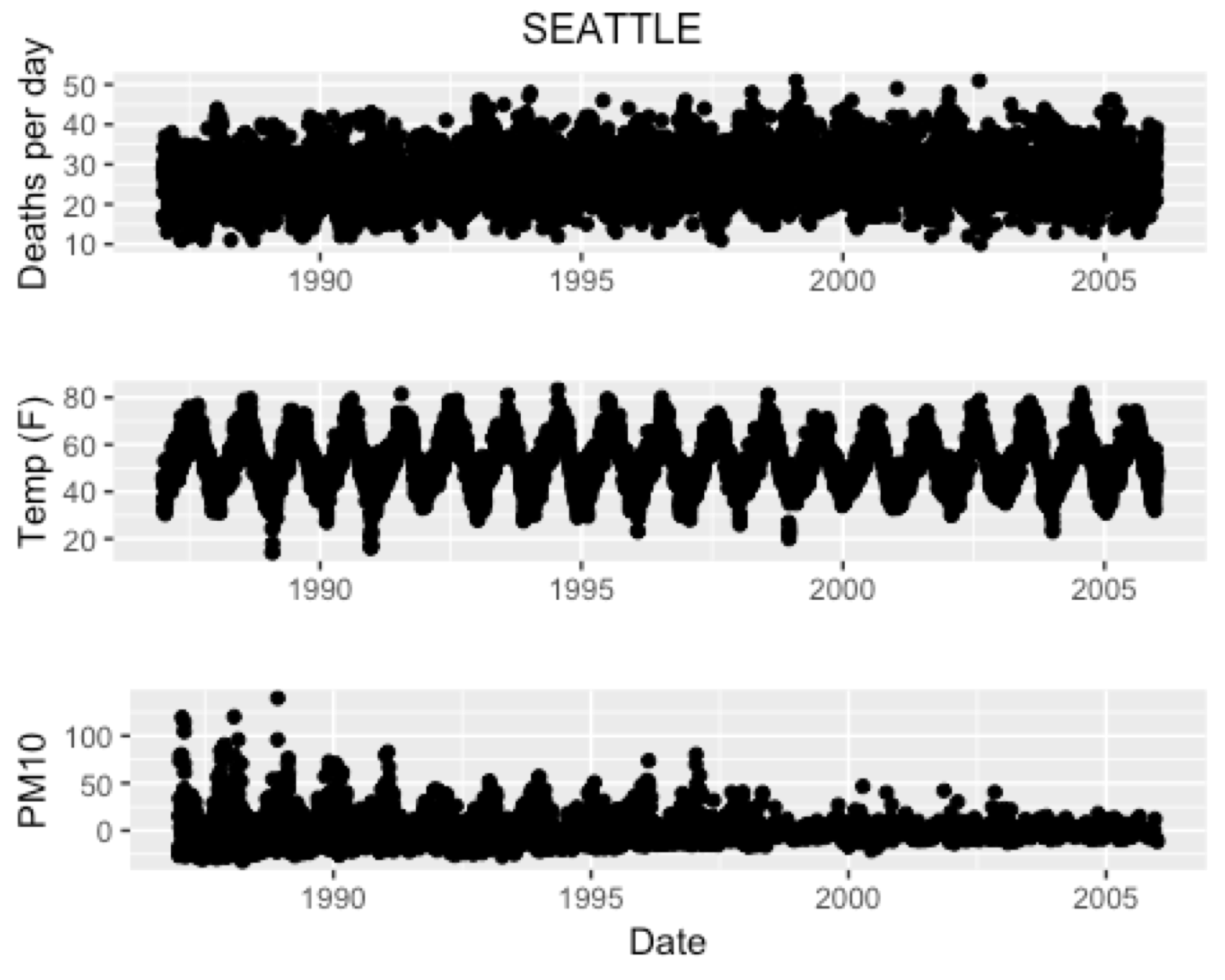
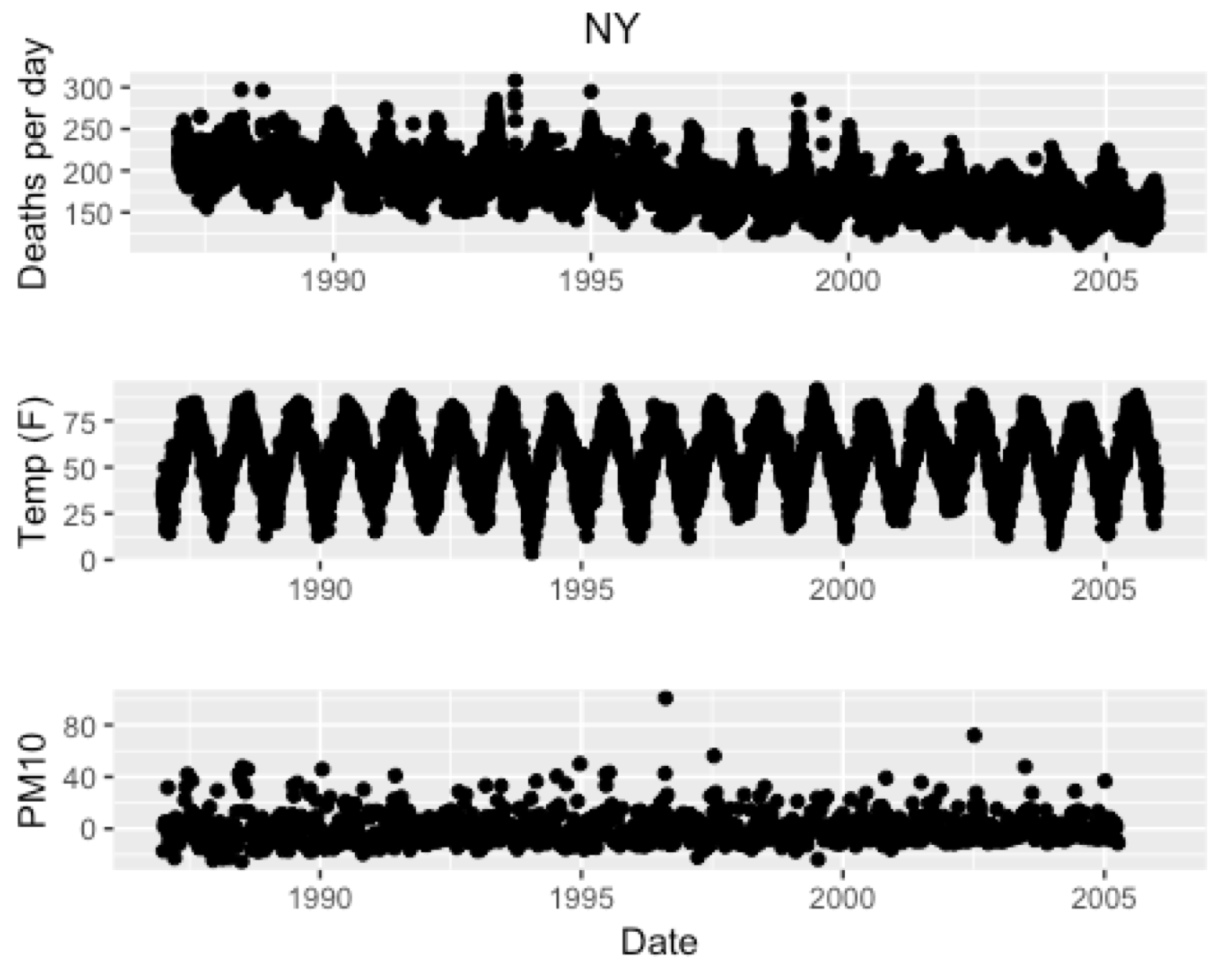


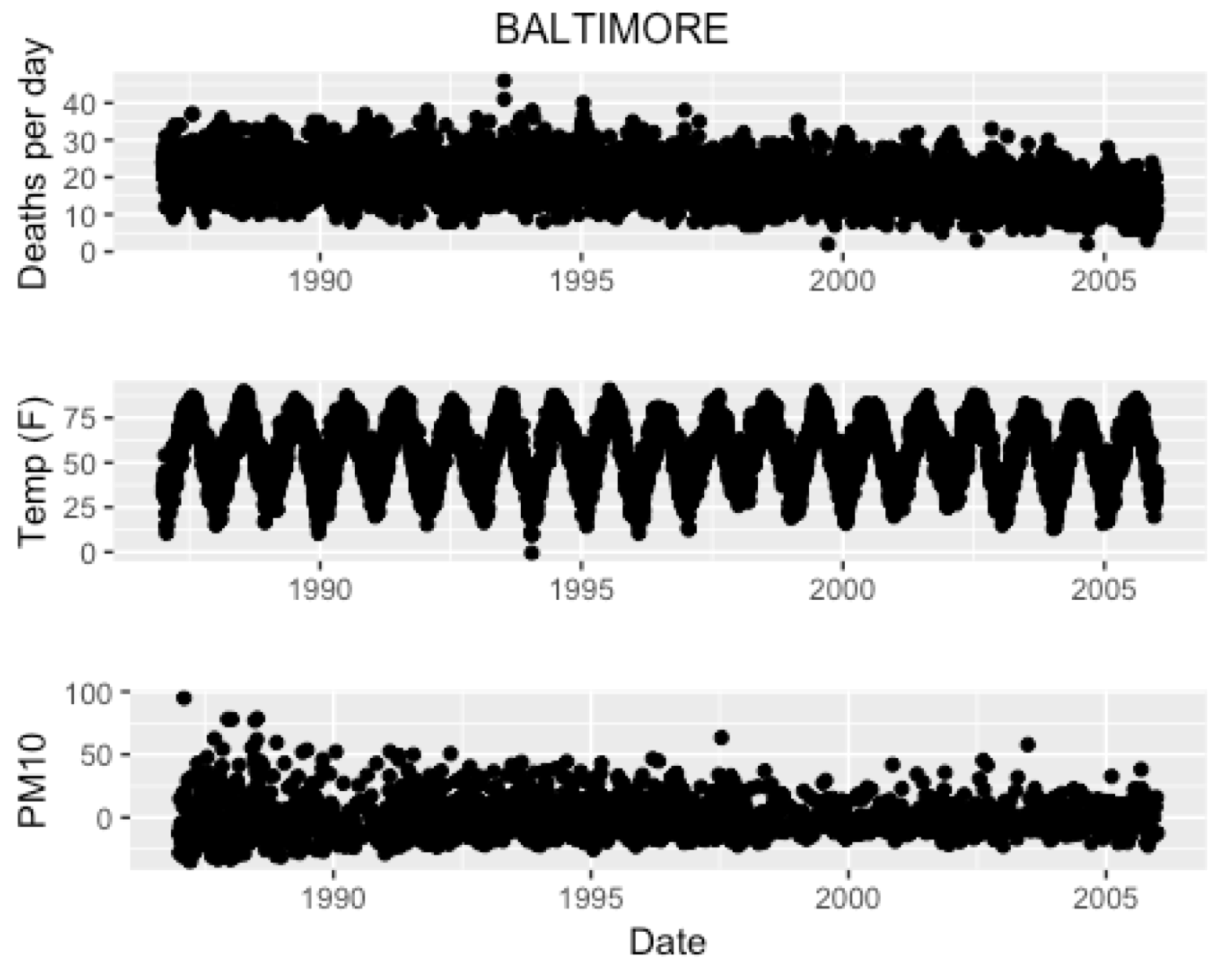
Similarly, Baltimore experiences an increase in morality in the summer – one can note two points near the year 1993 surpassing 40 **(Mortality)**. Regarding temperature, we are not surprised to find a uniform increase during the summer season in the city **(Temperature)**. What is interesting to note is the general pattern across Baltimore’s PM levels – in the 20 year span, we see a spike in PM levels at the start (year 1985) that declines steadily till the final year (2005) **(PM)**. Along with the other cities, these set of graphics help support a relationship does in fact exist between these variables and inspires a deeper investigation when answering Question 2.2.

**Question 2.2:** Is the estimate of the pollution effect sensitive to assumptions about seasonal or weather effects?

As referenced before, Module 2 was far more detailed in regards to accounting to effect changes by seasonal/weather effects compared to our investigation of health outcomes across smoking and non-smoking populations.

We begin with an overall assessment of our cities in time displays across the variables of interest: death per day, temperature, and PM.

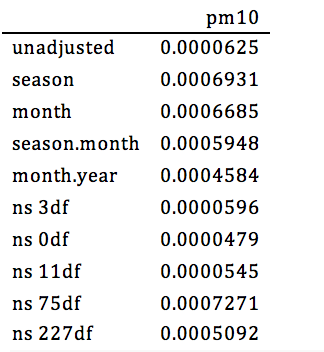
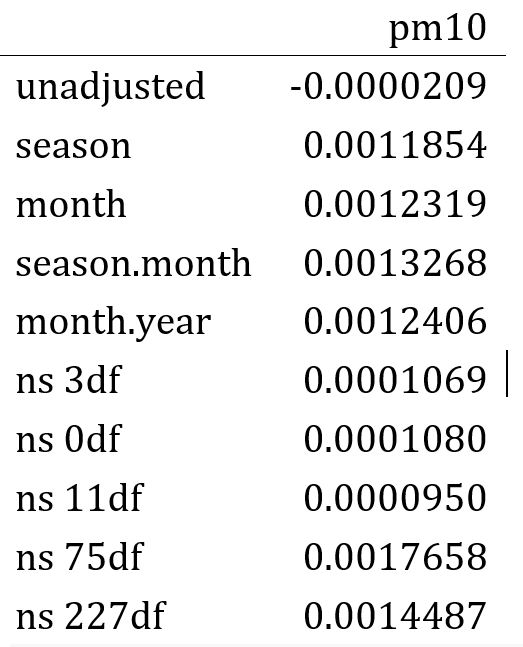
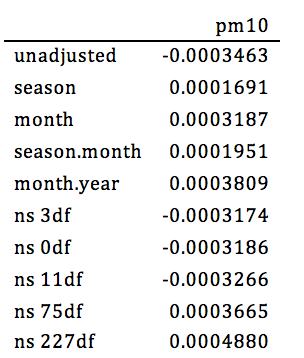




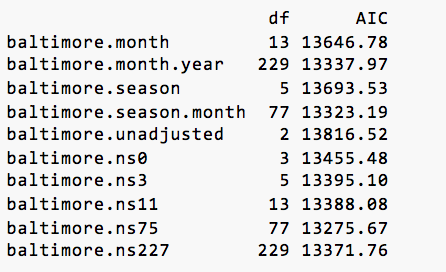
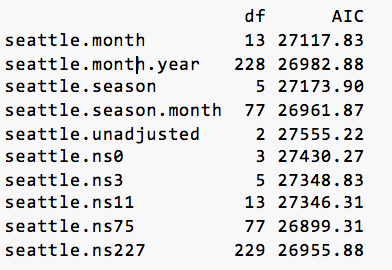
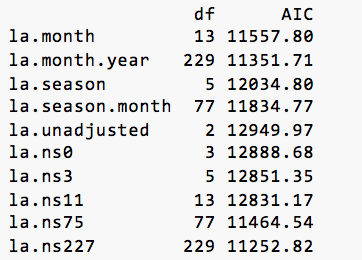
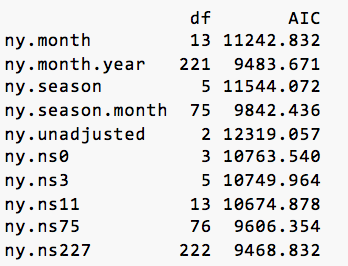
This time around, ten models were determined from the previous section of the module to have potential influence over our research objective. These were an unadjusted model measuring PM10 values, one impacted by seasonal variation, another impacted by month variation, one impacted by season and month simultaneously, year and month simultaneously, and five varying degrees of freedom (0, 3, 11, 75, and 225).

Coefficients were determined across the 10 models of interest per city:

New York: Seattle: LA: Baltimore:



Question 2.3: How do you pool PM effect (log relative rate) estimates from multiple cities taking account of both natural geographic variability in the true effects and statistical errors that might differ among cities?

In this section of the project, determining the exact model beyond just a mere assumption based on intercept values was determined by AIC coefficients across the four cities. These were our tabular displays per respective city: 

The two models of interest with alternating lowest AIC values were that of **.ns75** and **ns.227**, the former having the lowest AIC value for New York and LA while the latter had the lowest AIC values for Baltimore and Seattle. We decided to go with .ns75 because we found a greater deal of variation in this model than found in .ns227.

When writing a pooling calculation of our own using weighted averages, our updated coeffcients were as follows across New York, LA, Baltimore, and Seattle: 0.00178, 0.00037, 0.00073, and 0.00004. In total, they give us a pooled estimate of 0.00292.

In summary, we’ve found a model for linear regression predictions of daily mortality based on pollution levels and extenuating factors, such as seasonal and weather changes. We’ve chosen Graphical displays demonstrate trends across variables of interest and in the final section, we determined a manner to measure quality of statistical value between several potential models.