Haedam Cho

Professor Agnellini

DIDA 325

5 May 2024

Introduction

The rise of average global temperatures, especially in the 21st century, has led to increasingly dangerous air quality levels which can have serious, adverse health effects to vulnerable individuals. Thus, it is paramount to document the air quality of different communities so that action can be taken accordingly. Australia is an interesting case for this topic due to the fact that it is always ranked relatively high for air quality, but has recently been plagued with bushfires that emit toxic gasses into the atmosphere. So, in order to see the fruits of well implemented government policy as well as see how they dealt with obstacles such as the 2019-2020 bushfires, we take a deeper look into Australia's, specifically the Australian Capital Territory (ACT), air quality over the last decade.

Using R programming and basic data analysis, we can clearly see the effects of the '19 -'20 bushfires on Australia's air quality as well as the overall trend in AQI. We find that there is an extreme spike in AQI as well as levels of all major pollutants during the '19 - '20 bushfire season, but we also saw other interesting findings, such as trends in hourly and seasonal AQI as well as differences in air quality by location. Finally, using time series analyses and forecasting, we see that in the next few years, Australia is expected to continue to improve its air quality and reduce possible health risks through a reduction in PM2.5 levels, despite severe setbacks caused by the major bushfires.

Background Information

For a better understanding of the analysis, an explanation of air quality indicators and its implications will be helpful. A good measure of air quality is the U.S's Air Quality Index (AQI), which is used by the Environmental Protection Agency (EPA). The AQI is a uniform scale that runs from 0 to 500, where a higher AQI value corresponds to a greater level of air pollution, ergo a greater health concern (airnow.gov).

AQI Basics for Ozone and Particle Pollution

Daily AQI Color	Levels of Concern	Values of Index	Description of Air Quality
Green	Good	0 to 50	Air quality is satisfactory, and air pollution poses little or no risk.
Yellow	Moderate	51 to 100	Air quality is acceptable. However, there may be a risk for some people, particularly those who are unusually sensitive to air pollution.
Orange	Unhealthy for Sensitive Groups	101 to 150	Members of sensitive groups may experience health effects. The general public is less likely to be affected.
Red	Unhealthy	151 to 200	Some members of the general public may experience health effects; members of sensitive groups may experience more serious health effects.
Purple	Very Unhealthy	201 to 300	Health alert: The risk of health effects is increased for everyone.
Maroon	Hazardous	301 and higher	Health warning of emergency conditions: everyone is more likely to be affected.

Figure 1a

The five major pollutants that are regulated under the EPA by the Clean Air Act are ground-level ozone (O3) , particulate matter (PM), carbon monoxide (CO), sulfur dioxide (SO2), and nitrogen dioxide (NO2). PM in particular should be underscored due to the risks it imposes on peoples' health. There are two main types of PM that are studied, the first being PM10 (particulate matter that is around 10 micrometers) and the second being PM2.5 (2.5 micrometers or less). Although both cause harm to the human body, PM2.5 is proven to be far more hazardous than its 10 micrometer counterpart. While PM10 irritates the eyes, nose, and throat areas, finer particles such as PM2.5 “are more dangerous because they can get into the deep parts of your lungs - or even into your blood” (National Center for Environmental Health, 2023), potentially leading to serious health concerns. This is why in the forecasting portion of the data analysis, we will be focusing on PM2.5 levels. The rest of the major pollutants under the Clean Air Act all pose health hazards and also can be detrimental to the environment and affected ecosystems.

Australia is known to have some of the best air quality in the world, with three of Australia's major cities (Canberra, Sydney, and Melbourne) being in the top 15 best air quality in the world (IQAir, 2024). This is done through maintaining strong government policies such as the National Clean Air Agreement and the National Environment Protection (Ambient Air Quality) Measure, which are both enforced by the Australian Government by working with states and territories within the continent. Each major state, territory, and local government are responsible for monitoring the air quality in their respective jurisdictions.

From a political standpoint, this close monitoring is likely what is keeping Australia’s air quality in check, and from taking a look at the Australian Capital Territory’s (ACT) air quality data, we are expecting to see the results of these efforts and potentially other interesting findings that will help individuals, or other government organizations make decisions regarding air quality. Regarding what other researchers have found about air quality trends, the Environmental Protection Agency found that the concentrations of air pollutants in the U.S. have dropped significantly since 1990, proving that policies that push for cleaner air practices do work, and with the data we will work with in this article, we hope to that see that same trend for Australia.

A major setback that Australia faced, however , and will most likely continue to face, are bushfires, specifically during the 2019-2020 season. It should be noted that wildfire is the general umbrella term that bushfire (wildfires that occur in bushlands) lies under. Wildfires pose such a risk to air quality because they emit wildfire smoke that includes nitrogen dioxide, ozone, and especially PM2.5 (World Health Organization, 2024) which all fall under the major pollutants of the Clean Air Act. Consequently, Australia is expected to have had a major spike in pollutants and a decrease in air quality (or increase in AQI) during 2019-2020. Unfortunately, Australia is also expected to continue to have severe bushfires due to climate change making wildfire seasons start earlier and end later. So, the implications of this article prove to be that much more crucial.

Data and Methods

The data that I chose, provided by the Australian Capital Territory (ACT) Government, shows information regarding the air quality in the ACT through 2014 to 2024. The dataset was created to support the National Environment Protection Measures to better communicate the ambient air quality to the ACT Community. The value of this data is immense, as it gives policy makers access to information that tells a story of progress made, or lack thereof, towards cleaner air in their jurisdiction. The main contributing agency to this dataset is the ACT health agency, which is only right due to how air quality directly affects human health. This dataset is frequently updated, as it was last modified just three days prior to the writing of this article. It should also be noted that air quality was measured hourly in several forms, such as the measurement of major air pollutants as well as the overall AQI.

The following variables NO2, O3_1hr, O3_4hr, O3_8hr, and CO are all measured in parts per million, which is a unit that measures concentrations of a substance. The different hours from the O3 variables represent how long the air quality monitors were rolling for. So logically, the 8 hour variable would be most accurate, therefore we will be omitting the 1 and 4 hour variables in this study (the EPA also uses the 8 hour rolling monitor data). The PM10 and PM2.5 variables are both measured in micrograms per cubic meter of air, making the numbers significantly bigger compared to how the other pollutants are measured. This is all accounted for when calculating the AQI, named AQI_Site in this dataset, which converts all of the concentrations of the pollutants into one uniform index on a scale from 0 to 500. Each pollutant also has an AQI of its own, which is calculated in its different way. Since we won't be looking at the individual AQI of each pollutant, these variables can be omitted as well (AQI_CO, AQI_NO2, etc.) The other variables of the dataset are self-explanatory, such as Name, GPS, DateTime, etc.

With this data, I want to answer the following questions: What are the general trends of air quality in the ACT area? Has Australia's efforts to achieve cleaner air quality been working despite the 2019-2020 wildfires? What should members of communities affected by wildfires do to avoid poor air quality? I also want to be able to test a hypothesis that supports my questions. Testing the significance of the wildfires' effect on air quality, however, has too obvious of an answer for me to test a hypothesis on it. So, the following hypotheses will be addressed instead.

<u>Null hypothesis:</u> AQI and pollutant levels following 2019-2020 has not decreased significantly compared to AQI and pollutant levels prior to 2019-2020.

<u>Alternative hypothesis:</u> AQI and pollutant levels following 2019-2020 has decreased significantly compared to AQI and pollutant levels prior to 2019-2020.

First, to answer the first question regarding general trends in air quality, I will be looking at a variety of data visualizations that also answer my other questions. In order to answer the question of what individuals should do in the midst of poor air quality, I will manipulate the data and visualize the hourly AQI to see what times have the highest AQI (worst air quality) as well as visualize AQI by location to potentially give members of society clues as to what kinds of areas to avoid and where to go instead when air quality is poor. Next, to answer if Australia's efforts have been successful, I will create a time series plot that models PM2.5 levels (arguably the most dangerous pollutant) over the last decade and forecast the next few years of PM2.5 levels. Finally, to address my hypothesis, I will conduct a Welch Two Sample t-test that will give us evidence to either reject, or fail to reject the null hypothesis. Throughout most of my visualizations, I will be comparing the air quality of various years to that of 2019-2020, to see the contrast of the effects of wildfires.

Results

First, to get a better general understanding of the dataset, I created two tables that show relationships between air pollutant levels and categorical variables such as the year and location. Figure 2a shows a table that groups the dataset by the two locations Florey and Monash and shows their respective average air pollutant concentrations over the last ten years. We see that while Monash has slightly higher levels of NO2 and CO, Florey has significantly higher average concentrations of both PM10 and PM2.5. In turn, Florey has a higher average AQI score than Monash.

	location	avg_NO2	avg_CO	avg_O3	avg_PM10	avg_PM2.5	avg_AQI
1	Florey	0.002899981	0.1940720	0.01869859	19.45674	12.929156	64.53562
2	Monash	0.003810027	0.2209008	0.01848414	12.40786	8.957786	45.77216

Figure 2a

The difference in average AQI between the two locations is significant because when we refer back to the AQI scale (Figure 1a) we see the Monash has an average AQI score of 45.77, which is in the green zone, meaning that Monash has ‘good’ air quality. Florey, on the other hand, has an average AQI of 64.53 over the last decade, which is above the threshold of a score of 50, meaning that Florey is in the yellow zone of ‘moderate’ air quality, where the level of air pollutants may be a health risk to unusually sensitive individuals. This information can give individuals a general idea of what kind of areas have poorer air quality.

	year	avg_NO2	avg_CO	avg_O3	avg_PM10	avg_PM2.5	avg_AQI
1	2014	0.004821670	0.2042363	0.007504891	9.253574	4.493604	22.19714
2	2015	0.004812000	0.3089411	0.014026947	10.048147	8.170905	38.83411
3	2016	0.003987354	0.1969313	0.015961688	9.783548	7.271604	35.59422
4	2017	0.004549313	0.2107091	0.019341913	9.767170	7.747420	40.34030
5	2018	0.004242890	0.2190401	0.020335957	11.245747	6.901257	38.98629
6	2019	0.003976893	0.2173779	0.023492298	20.224074	14.238840	69.15716
7	2020	0.004152154	0.3614925	0.020705345	31.945013	25.285573	112.18080
8	2021	0.003053029	0.1309252	0.016924333	10.229088	6.531114	37.31817
9	2022	0.002799074	0.1302677	0.018578721	7.394138	4.920477	33.68195
10	2023	0.003493654	0.2226151	0.018666946	10.356005	6.423843	36.66265
11	2024	0.001709725	0.2093043	0.017961801	10.409607	6.043172	33.03229

Figure 2b

The second table I made (Figure 2b) shows the same measurements of air pollutant concentrations and AQI, except over each of the last ten years instead of location. We notice right away that the years 2019 and 2020 have significantly higher levels of all pollutants and a higher AQI score compared to every other year. This is likely attributable to the major wildfires of Australia. The air quality measurements of the other years also hold valuable information and could potentially tell interesting stories. To delve further into these stories, visualizations of the data are necessary.

Looking closer at the relationship between the two locations and air quality, we can compare the average AQI of Florey and Monash over the last two years and that of 2019-2020. Figures 3a and 3b shows through easily understandable bar charts. In Figure 3a, we can see that recently (2023-2024) there is little difference between the AQI of the two sites, both being in the green zone of the AQI scale (<50). Figure 3b, however, depicts a stark difference in the average AQI during the severe wildfires that plagued Australia. Monash has nearly over 100 AQI, which would put them in the orange zone of the scale. Florey on the other hand, has an AQI value of about 260, three times the AQI value of Monash, putting it in the purple zone where air quality is described as 'very unhealthy' and the situation is described as a 'health alert', according Figure 1a.

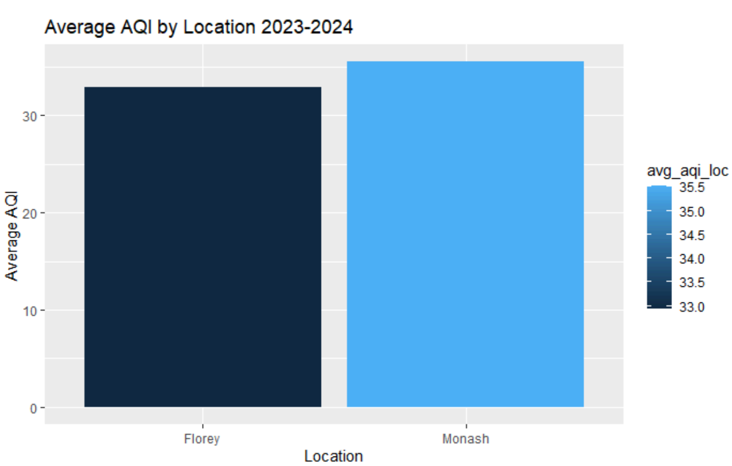

Figure 3a

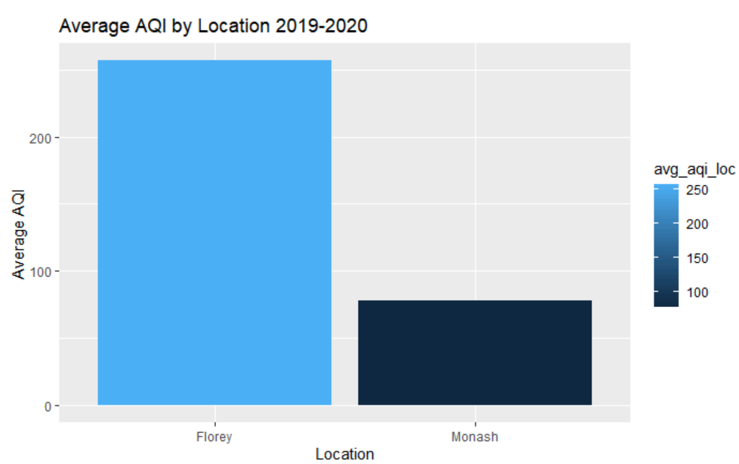

Figure 3b

This information is valuable because it describes what locations are most affected by wildfire (at least in terms of air quality) so next time Australia deals with severe wildfire seasons, the government will know what locations to prioritize for air quality remediation. Additionally, members of these communities will know what areas to avoid (Florey and places similar or near) and go to instead during severe bushfires.

To further answer the questions of what general trends of air quality are in the ACT area and what members of the community should do to avoid poor air quality, I took a look at the hourly trends of AQI in the ACT area. Figures 4a and 4b show average AQI levels throughout a 24-hour span in the same respective time periods of Figures 3a and 3b. Both Figures 4a and 4b show a similar trend where AQI slightly peaks early in the morning at around 12:00am to 2:00am, dips later into the morning, and peaks again at around 2:00pm to 7:00pm. It is clear that air quality is worst (high AQI) during the late afternoon and early evening, therefore outdoor air should be avoided at these times, especially during wildfire season. We see in Figure 4b that this trend in hourly AQI is seen with less strength due to the wildfires that make air quality that much worse all day.

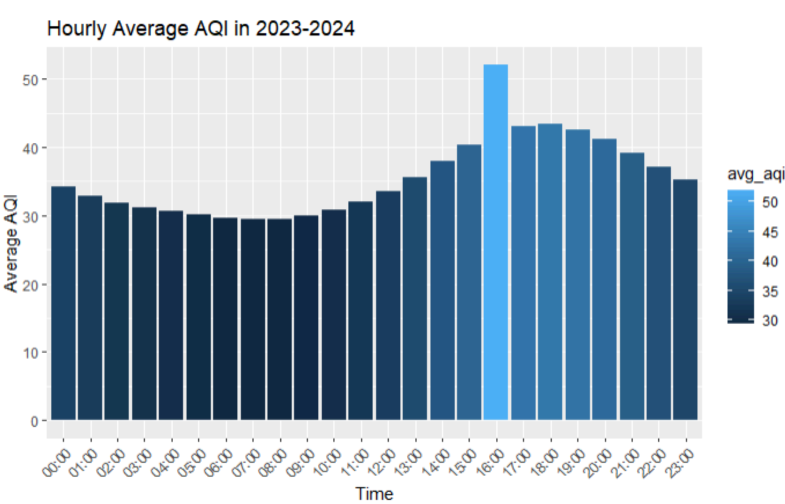

Figure 4a

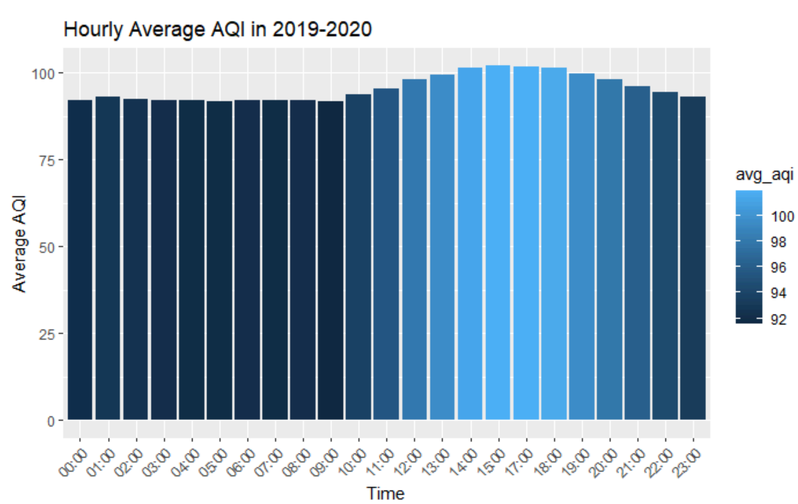

Figure 4b

Next, we can see if our data has a seasonal pattern to it so we can make conclusions on what time of year has the worst air quality on average. Figure 5a shows a season plot with month on the x-axis, PM2.5 levels on the y-axis, and lines that plot the relationship between the two variables for each year of the dataset. The reason why I am taking a look at PM2.5 specifically is because it is arguably the most dangerous of the five major pollutants, as mentioned before. Disregarding the spike in PM2.5 during the major bushfire season, we can see that mainly the first half of the year has the highest levels of PM2.5, therefore the worst air quality. Thus, sensitive individuals such as those with health conditions or old/young individuals should avoid excessive exposure to outdoor air during those months. Government organizations should also take this into consideration when implementing policies or new technologies that seek to reduce air pollutant concentrations in their jurisdiction.

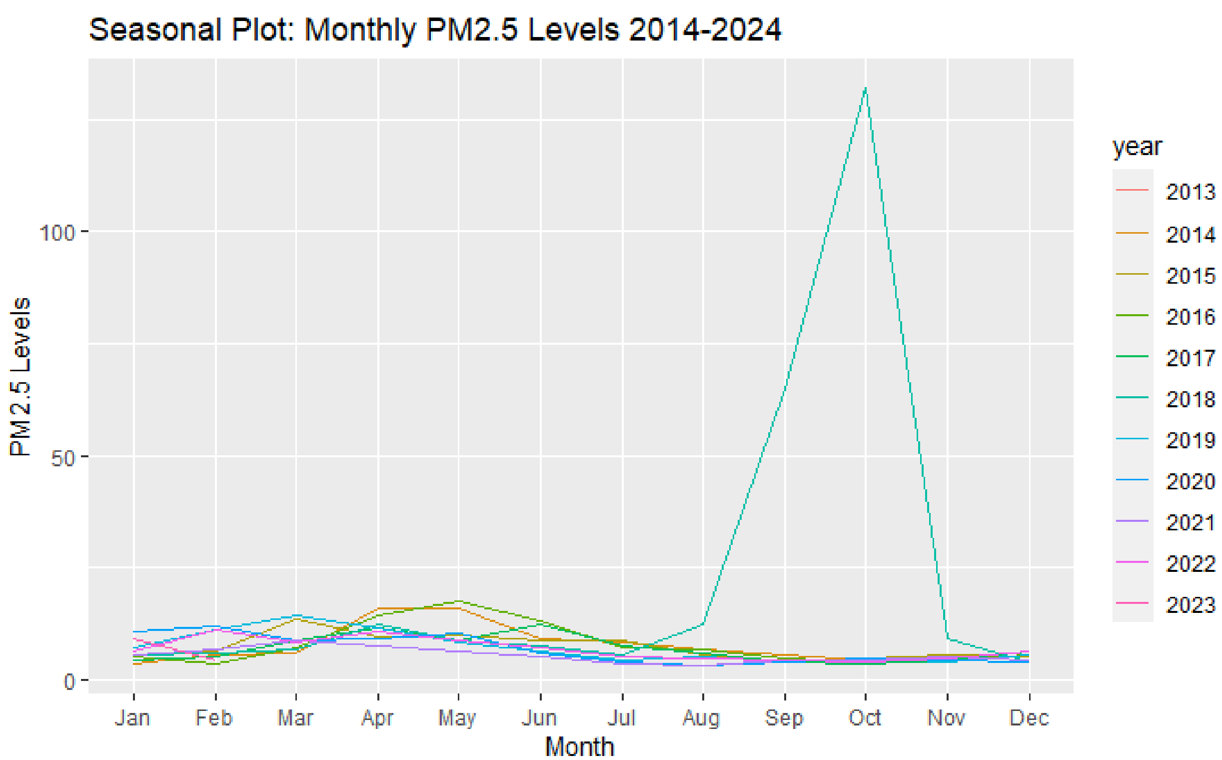

Figure 5a

In order to answer my last question of whether Australia's efforts to improve air quality has succeeded or not, I created a time series model using PM2.5 as the dependent variable and years as the independent. Figure 6a shows this by depicting average PM2.5 concentrations over a span of ten years. Despite the extreme peak in PM2.5 levels around 2019, due to bushfires running, we can see that the smaller peaks that occur between 2020-2024 are smaller than that of 2014-2018, meaning that PM2.5 levels saw a decrease over the years.

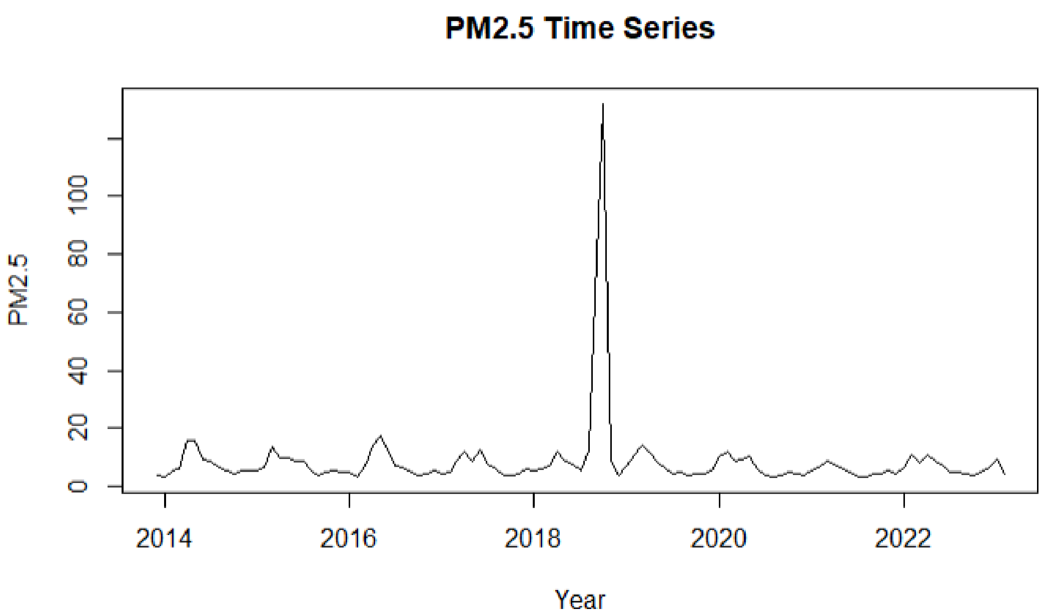

Figure 6a

Nevertheless, what's important is that Australia continues to enforce and create policies that prioritize improving air quality despite the negative effects of climate change. To predict this over the next few years, we can conduct a forecast using the time series model from Figure 6a. Figure 6b shows this forecasting model where the blue line (seasonal naïve) represents the future trajectory of PM2.5 levels based on the data and trends in the model. We can see that a steady decrease in PM2.5 concentrations continue, meaning that based on our data, Australia will continue to be successful in decreasing PM2.5 levels and improving air quality through the next few years.

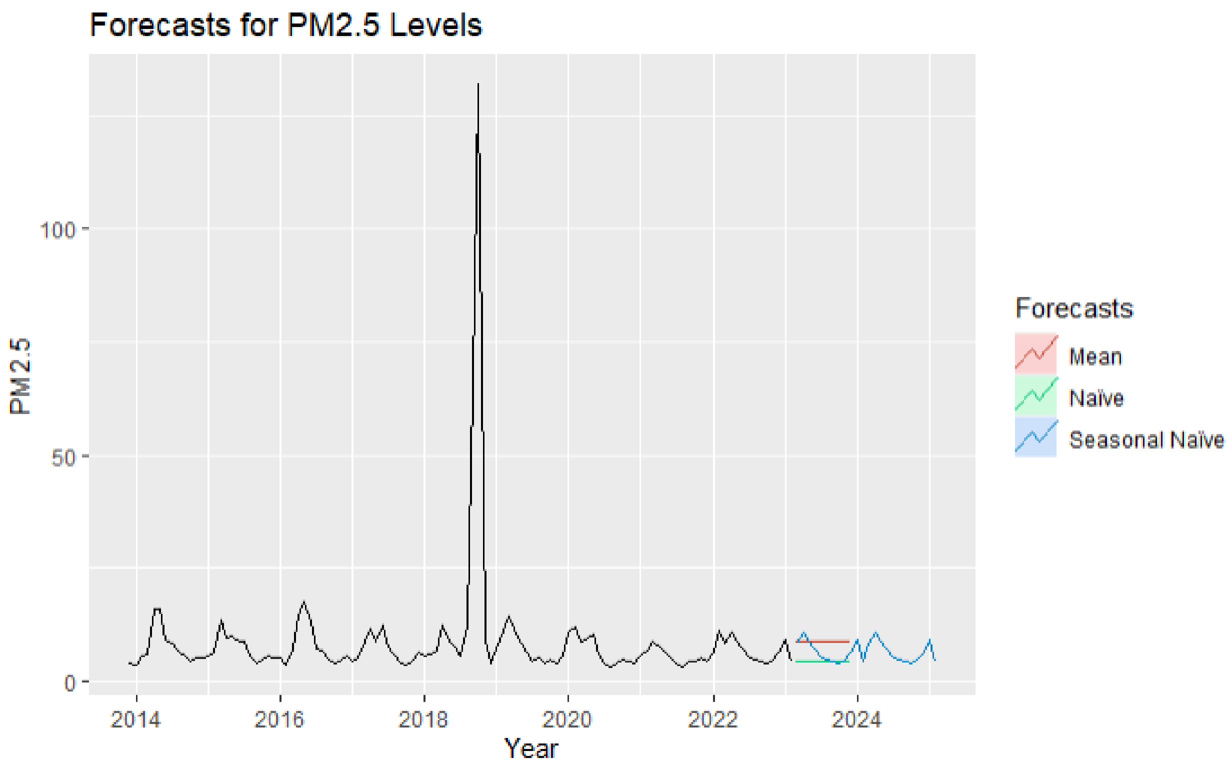

Figure 6b

Finally, I will be testing my hypothesis to see if there is really a significant difference in PM2.5 levels between the ranges 2014-2018 and 2020-2024. The results of this test will truly determine where Australia's air quality has been improving over the last decade. The reason why I am leaving out 2019 and 2020 is because the extreme data caused by the wildfires is not only irrelevant to the purposes of this test, but it will also throw off the final results. To test this, I used the Welch Two Sample t-test, where the two samples used are the PM2.5 concentrations from the two ranges of years. Figure 7a below shows the results of this test.

```
        Welch Two Sample t-test

data:  aqm_2014_2018$AQI_site and aqm_other_yrs2$AQI_site
t = -13.213, df = 24524, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.850979 -2.856060
sample estimates:
mean of x mean of y
 36.56230  39.91582
```

Figure 7a

The results reveal that the level of significance (p-value) must be less than 2.2e-16 in order to show a significant difference between the two samples. Fortunately, we see that our results show a p-value of 1.004761e-39, which is far less than the level of significance, meaning that our results are statistically significant. Therefore, Australia has been improving its air quality the past decade by a significant amount.

Conclusions

Through various data visualizations and analyses, we learned that Australia, in general, has good air quality, which can be measured by the AQI scale, where Australia (for the most part excluding the bushfire season) has AQI levels of less than 50. Through visualizations and machine learning models, we can also forecast future levels of PM2.5 based on the data from previous years, where we learn that PM2.5 concentrations will continue to see a decrease, meaning Australia's current efforts toward cleaner air quality has been successful. We also looked at a two sample t-test which showed that there was a statistically significant difference between the first half of the last decade and the second half (most recent). The importance of this will most likely be found in future research when seeing what strategies are most effective at improving air quality and Australia's tactics are considered. Then, Australia will be deemed a credible source in these other research projects and could be used as a standard. We also found that many of the visualizations were helpful to guide individuals how to take action in the midst of poor air quality and perhaps during bushfire season where PM2.5 levels are at its highest. With this information, members of the ACT community can act accordingly and the government could provide them with additional advice and suggestions to avoid potential health risks.

In regards to what I was unable to include in my analysis, there were many missing data values that all happened to be for the "Civic" location, a third location that was included in the dataset. So, unfortunately all of the data for the Civic location ended up being omitted. This, however, was for the sake of consistency in the data used in the analyses. If I did have all of the data on the Civic location, I would have been able to have more locations to compare results to. Additionally, more data would have given us more information on different areas within the ACT. I also would have liked if there were variables that potentially affected air pollutant concentration levels, such as the number of wildfires that occurred per year, so that I could conduct a linear regression on what has the most significant effect on AQI. Overall, the results of this study have major implications in the suggested behavior of ACT residents, potential changes in government policy, and current as well as future state of Australia's air quality that can set the stage for the rest of the world's progress toward cleaner air.

References

Government, A. (2024, May 6). *Air Quality Monitoring Data: Open data portal*. Air Quality Monitoring Data | Open Data Portal. https://www.data.act.gov.au/Environment/Air-Quality-Monitoring-Data/94a5-zqnn/about_data

Environmental Protection Agency. (n.d.). EPA. https://gispub.epa.gov/air/trendsreport/2022/#air_trends

World's cleanest cities. IQAir. (n.d.). https://www.iqair.com/us/world-air-quality-ranking/cleanest-cities

Air Quality. DCCEEW. (n.d.). https://www.dcceew.gov.au/environment/protection/air-quality

World Health Organization. (n.d.). *Wildfires*. World Health Organization. https://www.who.int/health-topics/wildfires#tab=tab_1

Centers for Disease Control and Prevention. (2023, February 16). *Particle pollution*. Centers for Disease Control and Prevention. https://www.cdc.gov/air/particulate_matter.html

AirNow.gov, U.S. EPA. (n.d.). *Aqi Basics*. AQI Basics | AirNow.gov. https://www.airnow.gov/aqi/aqi-basics/