

For my dataset, I first organized the data. There are some NA data, and I choose to delete them. I created the air quality column and use Chinese national standards(GB standards) to classify them. From GB 3095-2012, the 24-hour average value of PM2.5 lower than 75($\mu\text{g}/\text{m}^3$) means the air quality is healthy, the 24-hour average value of PM2.5 higher than 75($\mu\text{g}/\text{m}^3$) means the air quality is unhealthy [1]. I also created the season column because I need to use this attribute in my questions.

```
data("beijingpm2.5")
pm2.5_no_na <- na.omit(beijingpm2.5)
## delete NA
data("pm2.5_no_na")
head(pm2.5_no_na)
colnames(pm2.5_no_na) <-
c("No", "year", "month", "day", "hour", "pm2.5", "DEMP", "TEMP", "PRES", "cbwd", "lws", "ls", "lr")
## rename colnum
pm2.5_no_na$airquality <- as.character(pm2.5_no_na$pm2.5)
## create a new Column airquality
pm2.5_no_na <- pm2.5_no_na[-1,]
## delete first row
pm2.5_no_na$airquality[which(pm2.5_no_na$airquality >= 75)] <- "unhealthy"
pm2.5_no_na$airquality[which(pm2.5_no_na$airquality < 75)] <- "healthy"
## change airquality value to healthy and unhealthy
pm2.5_no_na$season <- as.character(pm2.5_no_na$month)
## create a new column season
pm2.5_no_na$season[which(pm2.5_no_na$season == 3)] <- "spring"
pm2.5_no_na$season[which(pm2.5_no_na$season == 4)] <- "spring"
pm2.5_no_na$season[which(pm2.5_no_na$season == 5)] <- "spring"
pm2.5_no_na$season[which(pm2.5_no_na$season == 6)] <- "summer"
pm2.5_no_na$season[which(pm2.5_no_na$season == 7)] <- "summer"
pm2.5_no_na$season[which(pm2.5_no_na$season == 8)] <- "summer"
pm2.5_no_na$season[which(pm2.5_no_na$season == 9)] <- "fall"
pm2.5_no_na$season[which(pm2.5_no_na$season == 10)] <- "fall"
pm2.5_no_na$season[which(pm2.5_no_na$season == 11)] <- "fall"
pm2.5_no_na$season[which(pm2.5_no_na$season == 12)] <- "winter"
pm2.5_no_na$season[which(pm2.5_no_na$season == 1)] <- "winter"
pm2.5_no_na$season[which(pm2.5_no_na$season == 2)] <- "winter"
## change season value
```

For the beginning, I build two figures to show the count and the percentage of the healthy air quality and unhealthy air quality for each hour as a summary. Because of the deleted NA data, I want to make sure that the total quantity of data for each year doesn't have too much difference. I think it seems fine for the quantity.

For the percentage of the healthy and unhealthy air quality, although the percentage increase of healthy air quality is small, it does increase year by year. This is due to the serious measures taken by Beijing to reduce coal consumption in early 2010[2].

From two figures, it seems that the air quality is not bad, but that is because I choose the Chinese national standards. In the research paper which this dataset comes from, they choose the US standards, and under that condition, the percentage of healthy air quality ($PM \leq 35$) is just around 20%[3].

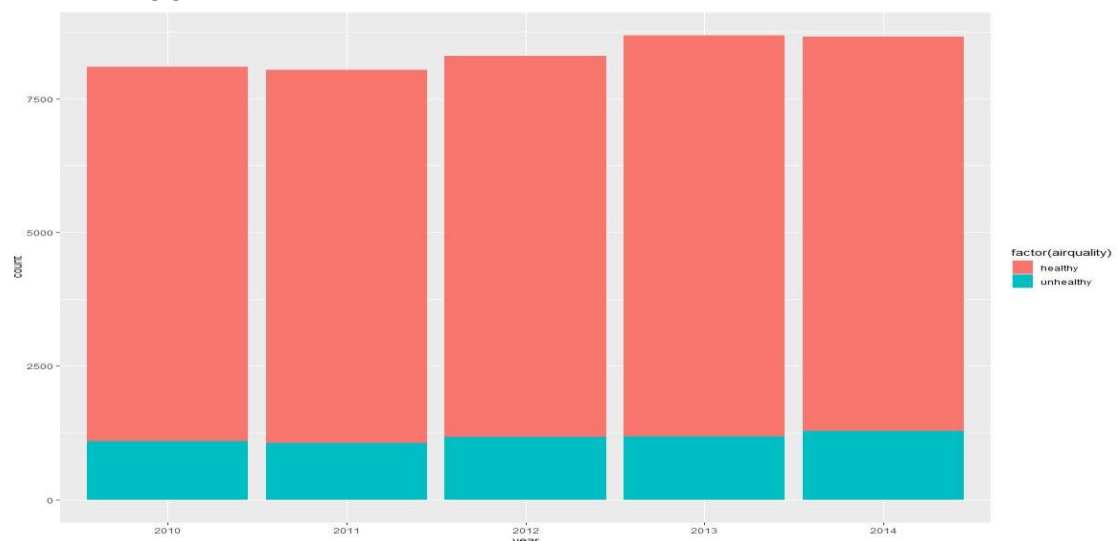


Fig. 1 The count of healthy and unhealthy air quality in each year(2010-2014)

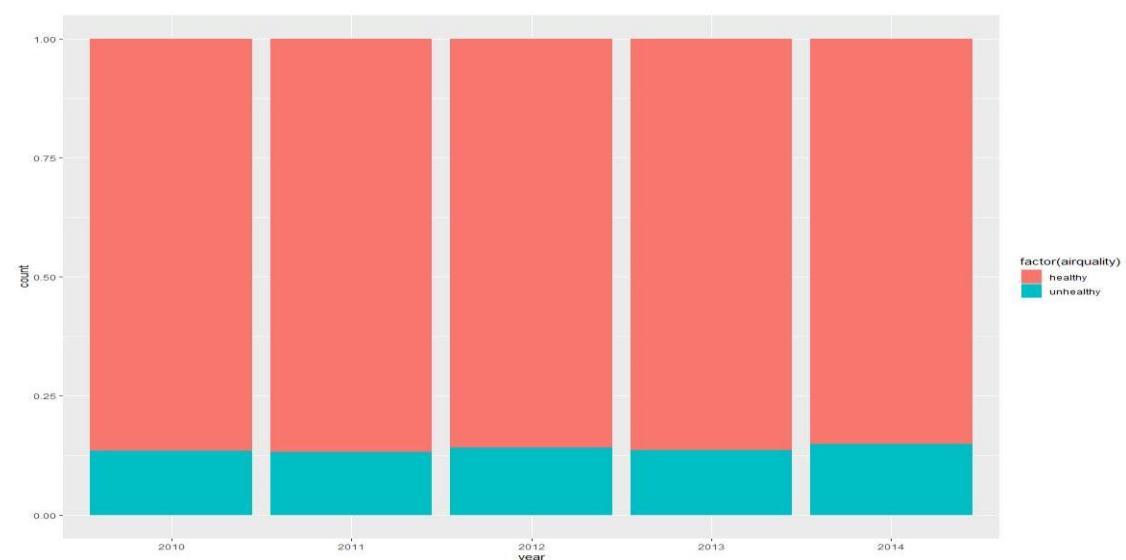


Fig. 2 The percentage of healthy and unhealthy air quality in each year(2010-2014)

For the second step, I build two figures to show the air quality in different month and different season. From fig.3, we can see that for the whole year, the percentage of healthy air quality shows a trend of first decline and then rise. Meanwhile, from fig. 4 we can see that summer has the lowest percentage of healthy air quality, and winter has the highest percentage of healthy air quality.

This conclusion is kind of weird, because it's the opposite of the science. We all know that winter has worse air quality than summer. From the research which the dataset from, I found the reason [3]. In the research paper, the percentage of PM2.5 between 35 and 150 is much higher in summer than in winter, but the percentage of PM2.5 higher than 150 is much lower in summer than in winter. I think this is because the frequent rainfall in summer cause the day of $PM_{2.5} \geq 150$ decrease, but the rain cannot reduce the particulate matter which has smaller diameter. Therefore, the percentage of healthy air quality is lower in summer than it in winter in my dataset.

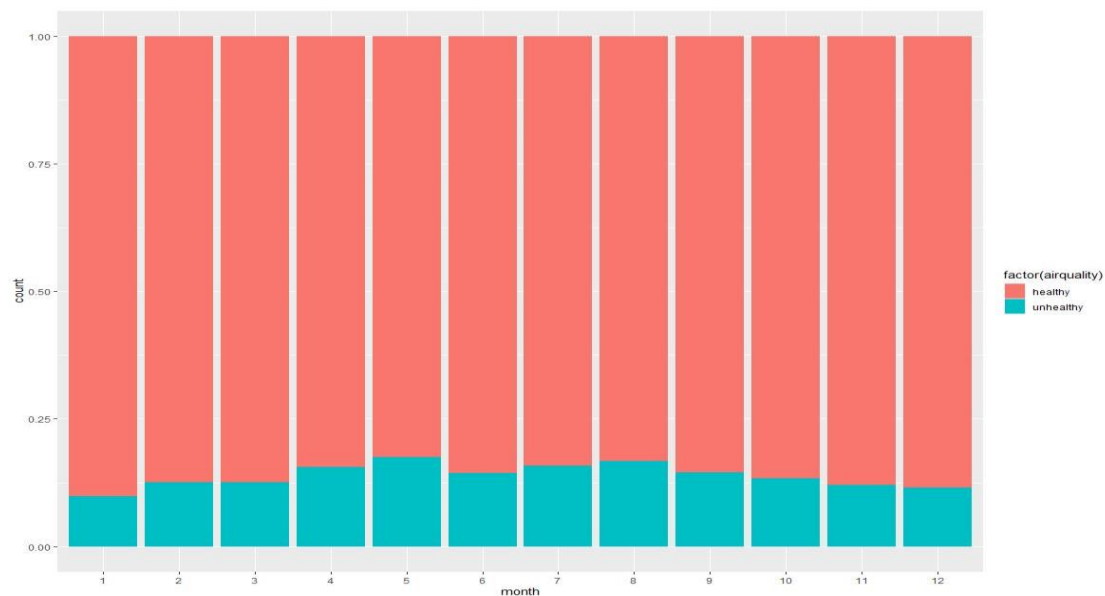


Fig. 3 The percentage of healthy and unhealthy air quality in 12 months

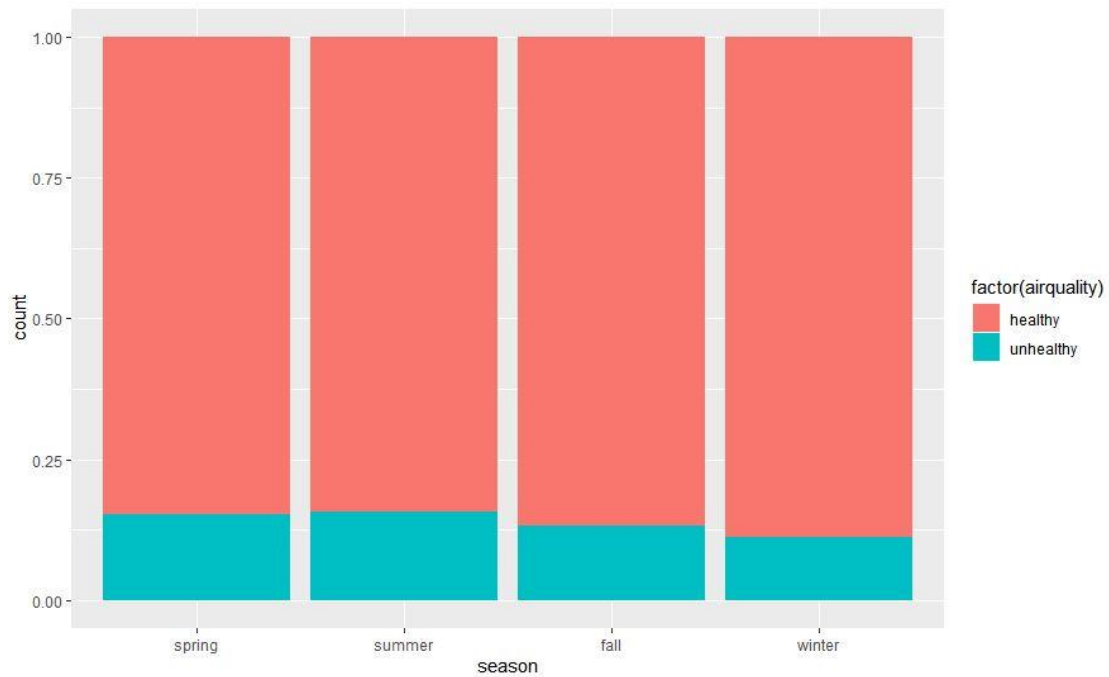


Fig. 4 The percentage of healthy and unhealthy air quality in four seasons (spring, summer, fall, winter)

Reference:

- [1] Ministry of Ecology and Environment of the People's Republic of China. 2012. Ambient air quality standards. Retrieved from <https://www.mee.gov.cn/>
- [2] China Power Team. Is Air Quality in China a Social Problem China Power. (February 2016). Retrieved August 26, 2020 from <https://chinapower.csis.org/air-quality/>
- [3] He H, Tie X, Zhang Q, Liu X, Gao Q, Li X, Gao Y. 2015. Analysis of the causes of heavy aerosol pollution in Beijing, China: a case study with the WRF-Chem model. Particuology 20, 32–40. DOI: <https://doi.org/10.1016/j.partic.2014.06.004>