

Rmarkdown example

Nicholas Tierney

Contents

Introduction	1
Method	2
Results	3
Conclusion	4
References	4

```
library(tidyverse)
```

```
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr

## Conflicts with tidy packages -----

## filter(): dplyr, stats
## lag():    dplyr, stats
```

Introduction

This is a simple analysis of the New York Air Quality Measurements using the R statistical programming language [1]. As stated in the helpfile `?airquality`, this dataset contains:

Daily air quality measurements in New York, May to September 1973.

And the dataset is sourced from:

... the New York State Department of Conservation (ozone data) and the National Weather Service (meteorological data).

It contains the following variables

- Ozone: Mean ozone in parts per billion from 1300 to 1500 hours at Roosevelt Island.
- Solar.R: Solar radiation in Langleys in the frequency band 4000–7700 Angstroms from 0800 to 1200 hours at Central Park.
- Wind: Average wind speed in miles per hour at 0700 and 1000 hours at LaGuardia Airport.
- Temp: Maximum daily temperature in degrees Fahrenheit at La Guardia Airport.
- Month: Month (1–12)
- Day: Day of month (1–31)

We are going to explore the relationship between solar radiation and other selected variables, solar radiation, wind, and temperature.

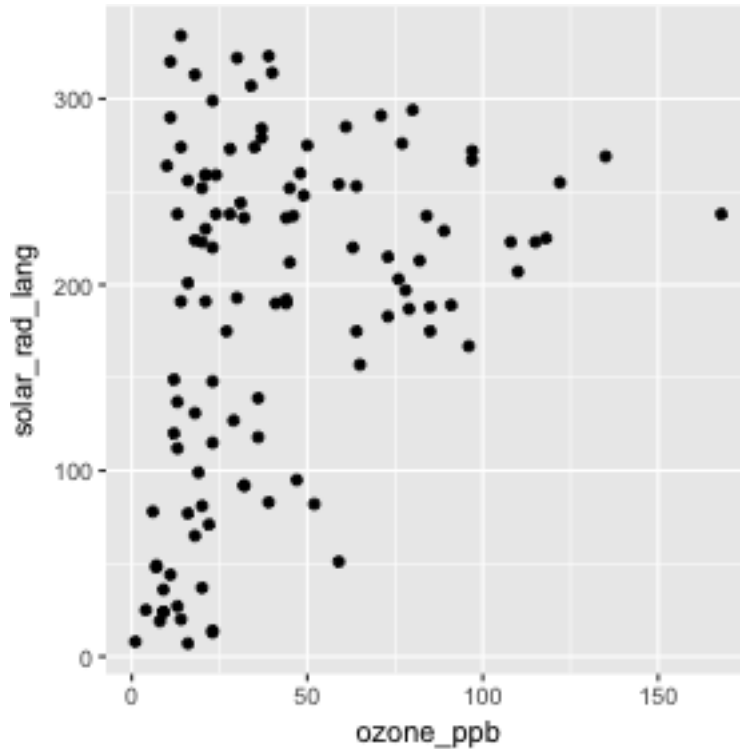


Figure 1: This is a figure of Ozone and Solar Radiation. Good question!

Method

First, we tidy the names of the dataset, to provide information about the units of measurement for Ozone, Solar Radiation, Wind, and Temperature. We do this by renaming the variables and adding a suffix at the end to describe the units. To do this we use the `rename` function from the `dplyr` package[2].

```
tidy_aq <- rename(.data = airquality,
  ozone_ppb = Ozone,
  solar_rad_lang = Solar.R,
  wind_mph = Wind,
  temp_fah = Temp,
  month = Month,
  day = Day)
```

We can see that there is an interesting relationship between ozone and solar radiation in figure 1 below, plotted using `ggplot2` [3]

```
ggplot(tidy_aq,
  aes(x = ozone_ppb,
    y = solar_rad_lang)) +
  geom_point()
```

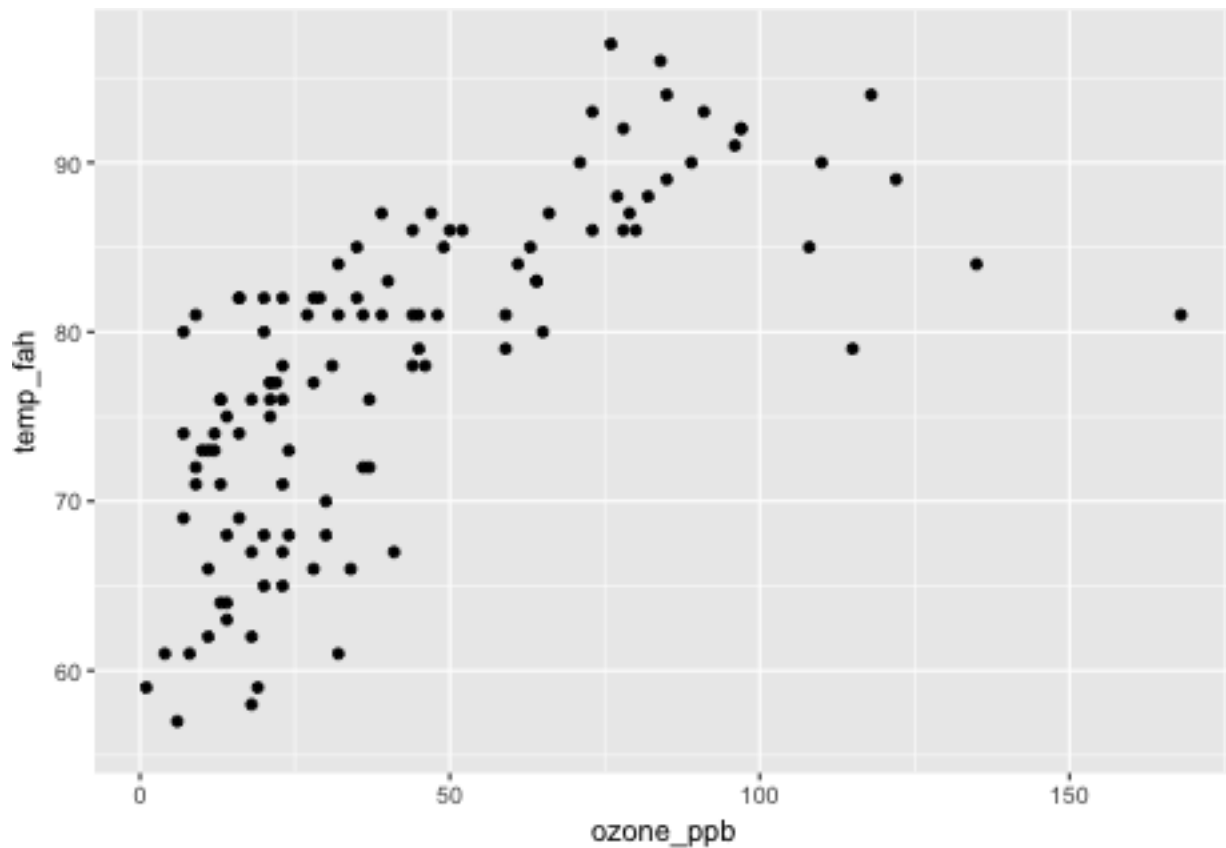
Warning: Removed 42 rows containing missing values (geom_point).

We can also see that there is an interesting relationship between Ozone and temperature.

```
ggplot(tidy_aq,
  aes(x = ozone_ppb,
    y = temp_fah)) +
```

```
geom_point()
```

```
## Warning: Removed 37 rows containing missing values (geom_point).
```



To explore the relationships between Ozone and all of the variables in the dataset, we can fit a basic linear model, with Ozone as the outcome, and all other variables as the predictors. We can express this as:

$$Ozone \sim \beta_0 + \beta_1 Solar + \beta_2 Wind + \beta_3 Temp + \gamma + \epsilon$$

And we can fit this model using the code below.

```
lm_aq <- lm(ozone_ppb ~ solar_rad_lang + wind_mph + temp_fah,  
            data = tidy_aq)
```

Results

The key results are given below, using the `tidy` function from the `broom` package [4] to clean up the data.

```
library(broom)
```

```
tidy_lm_aq <- tidy(lm_aq)
```

```
tidy_lm_aq
```

```
##           term      estimate std.error statistic    p.value  
## 1 (Intercept) -64.34207893 23.05472435  -2.790841 6.226638e-03
```

```
## 2 solar_rad_lang 0.05982059 0.02318647 2.579979 1.123664e-02
## 3      wind_mph  -3.3359131 0.65440710 -5.094063 1.515934e-06
## 4      temp_fah  1.65209291 0.25352979 6.516366 2.423506e-09
```

We can present this result in a nice table using the `kable` function from the `knitr` package [5]

```
knitr::kable(tidy_lm_aq,
             digits = 3,
             caption = "Table of results from the linear model")
```

Table 1: Table of results from the linear model

term	estimate	std.error	statistic	p.value
(Intercept)	-64.342	23.055	-2.791	0.006
solar_rad_lang	0.060	0.023	2.580	0.011
wind_mph	-3.334	0.654	-5.094	0.000
temp_fah	1.652	0.254	6.516	0.000

We can also refer to individual results of the model inside the text. For example, we can say that the estimated coefficient of Wind miles per hour is -3.334, and the P value of this is 0.

Conclusion

We have explored the relationship of Ozone with other variables in the airquality dataset

References

```
knitr::kable(tidy_lm_aq,
             digits = 3,
             caption = "Table of results from the linear model")
```

- 1 R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria.: R Foundation for Statistical Computing 2017. <https://www.R-project.org/>
- 2 Wickham H, Francois R, Henry L *et al.* *Dplyr: A grammar of data manipulation*. 2017. <https://CRAN.R-project.org/package=dplyr>
- 3 Wickham H, Chang W. *Ggplot2: Create elegant data visualisations using the grammar of graphics*. 2017.
- 4 Robinson D. *Broom: Convert statistical analysis objects into tidy data frames*. 2017. <https://CRAN.R-project.org/package=broom>
- 5 Xie Y. *Knitr: A general-purpose package for dynamic report generation in r*. 2017. <https://CRAN.R-project.org/package=knitr>