EDA

Keshav

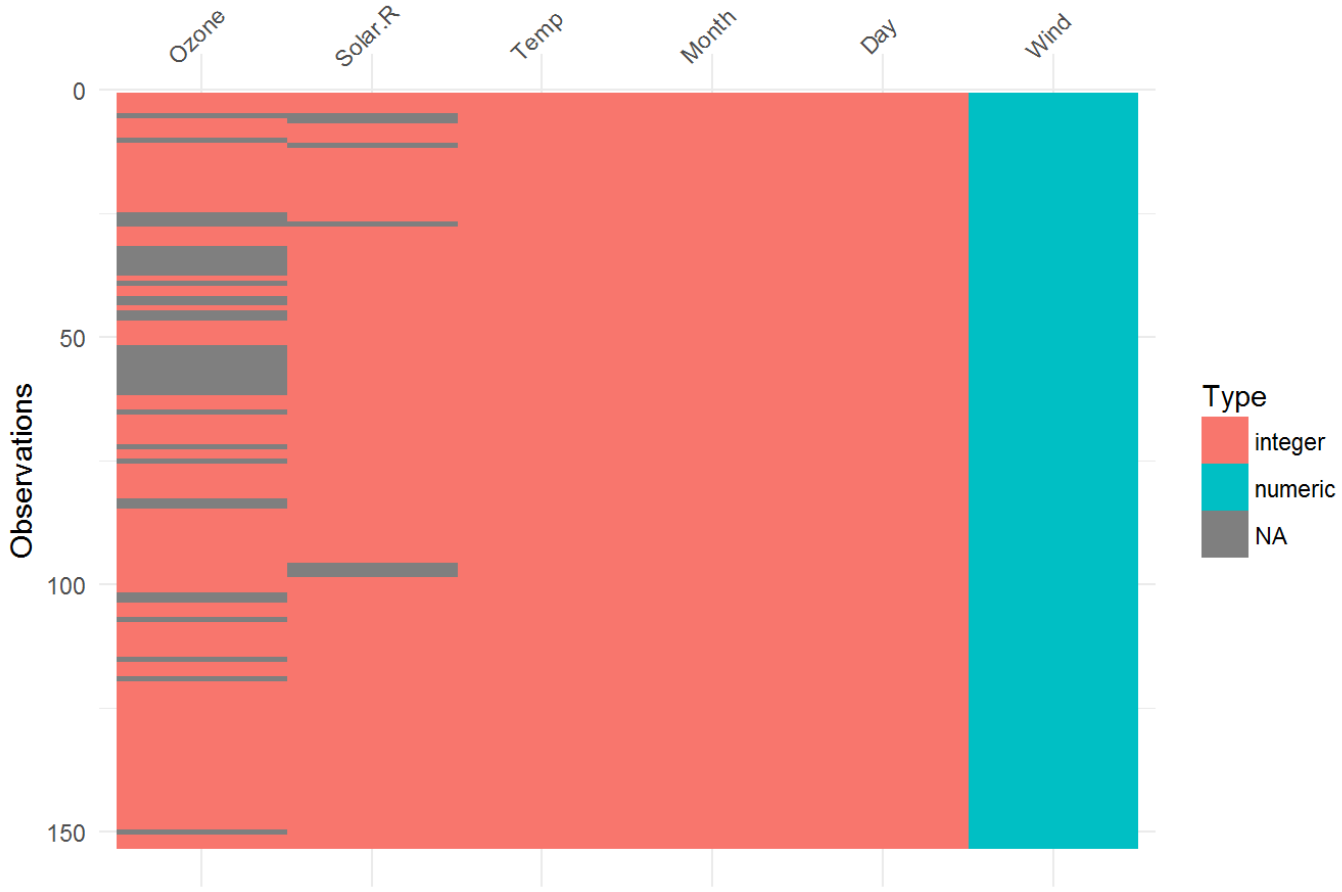
22 November 2017

Showing how EDA makes understanding of Missing value simple .The code is simple but powerful while understanding missing data.

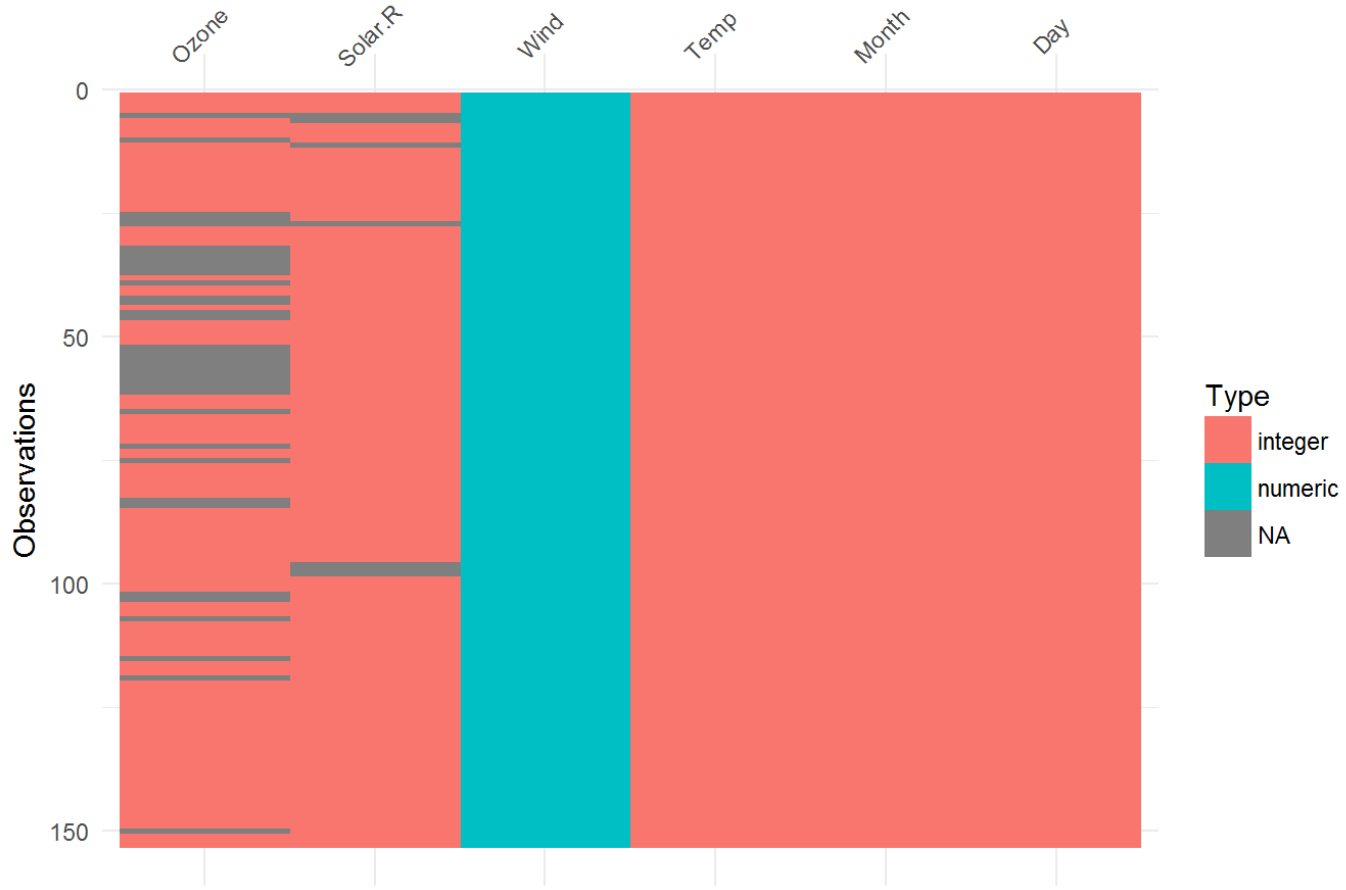
```
#install.packages("visdat")
library(visdat)
```

```
## Warning: package 'visdat' was built under R version 3.4.3
```

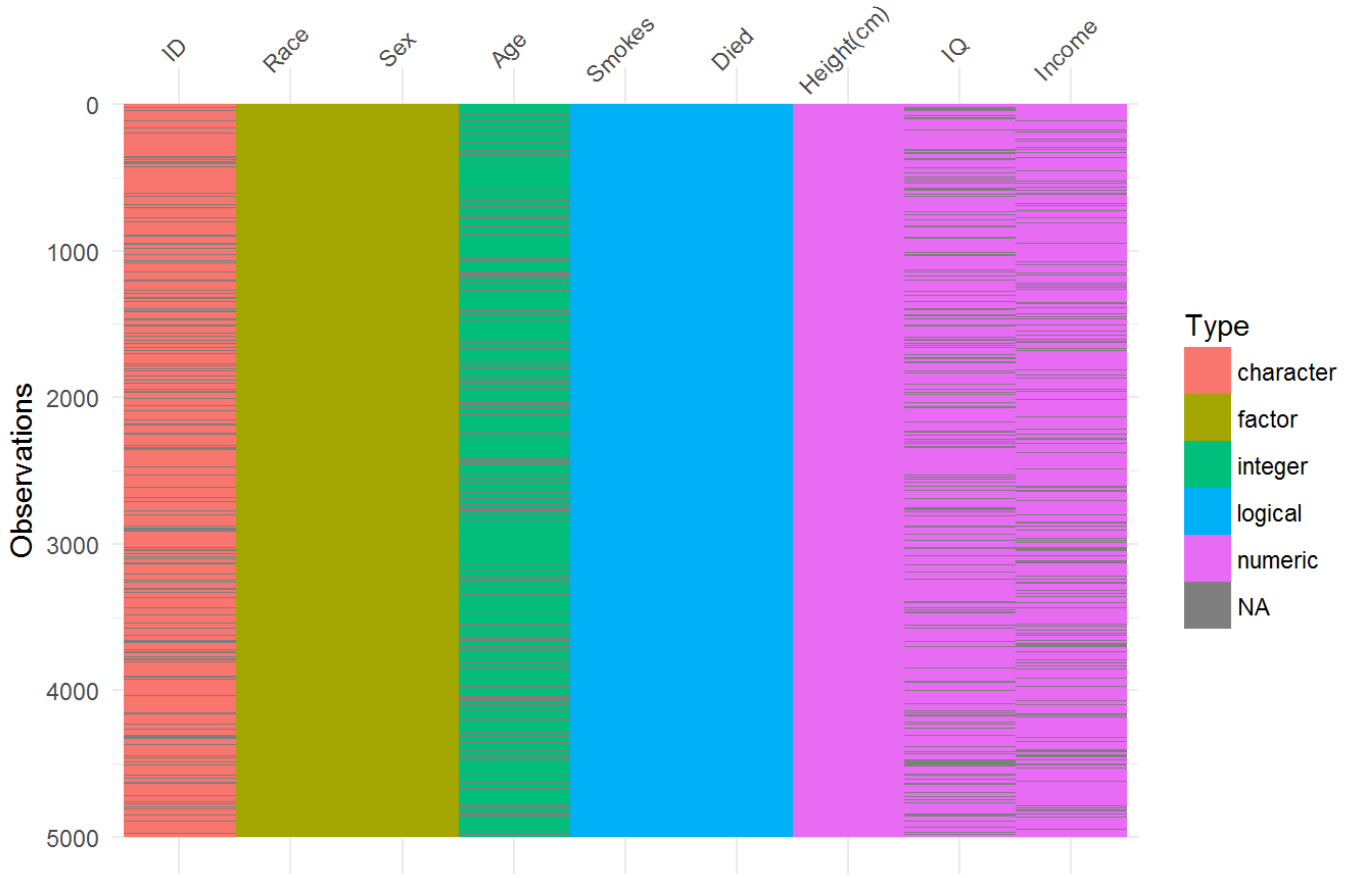
```
vis_dat(airquality)
```



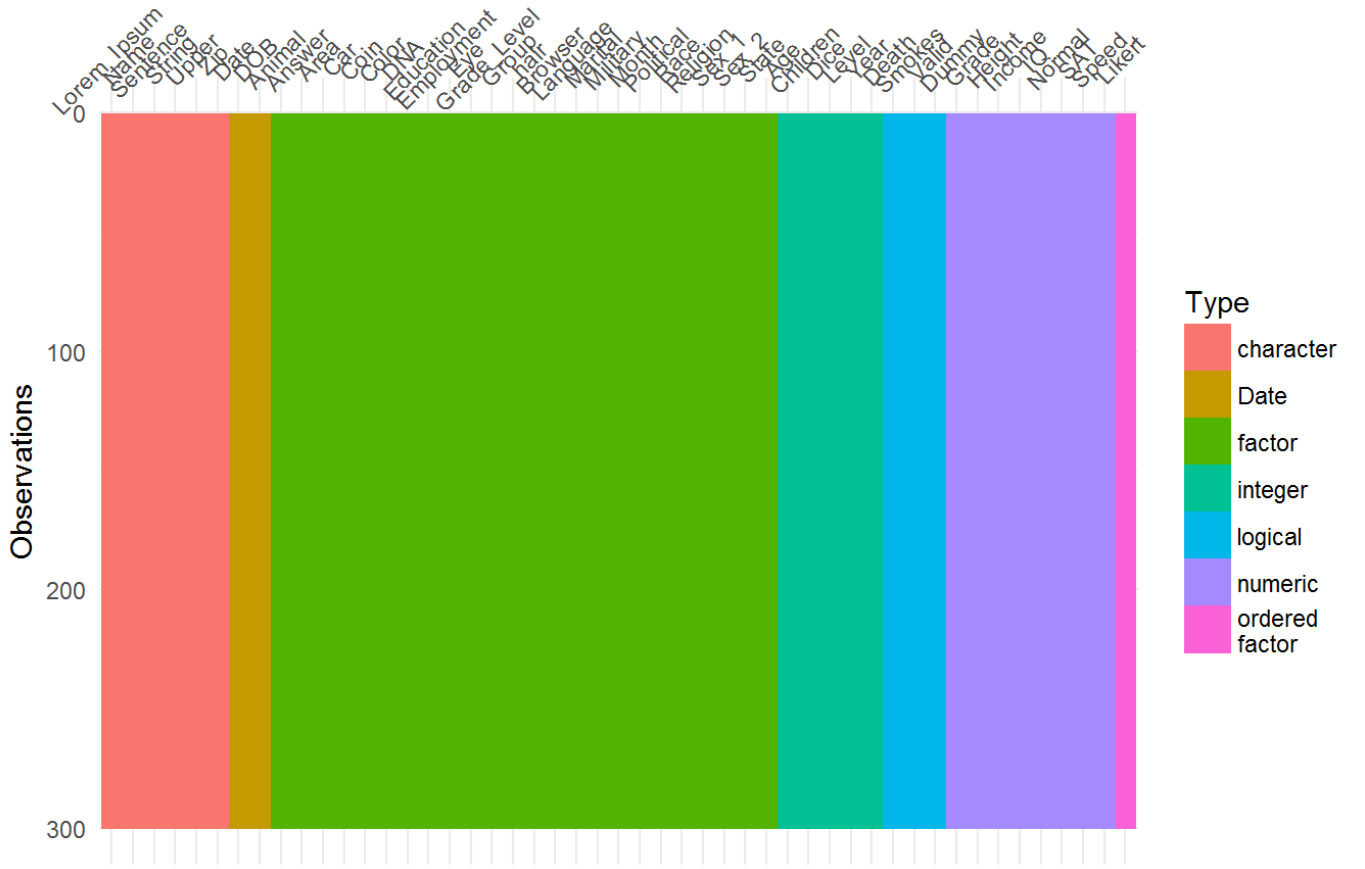

```
# We can sort the missing values in decreasing form.
vis_dat(airquality,
        sort_type = FALSE)
```



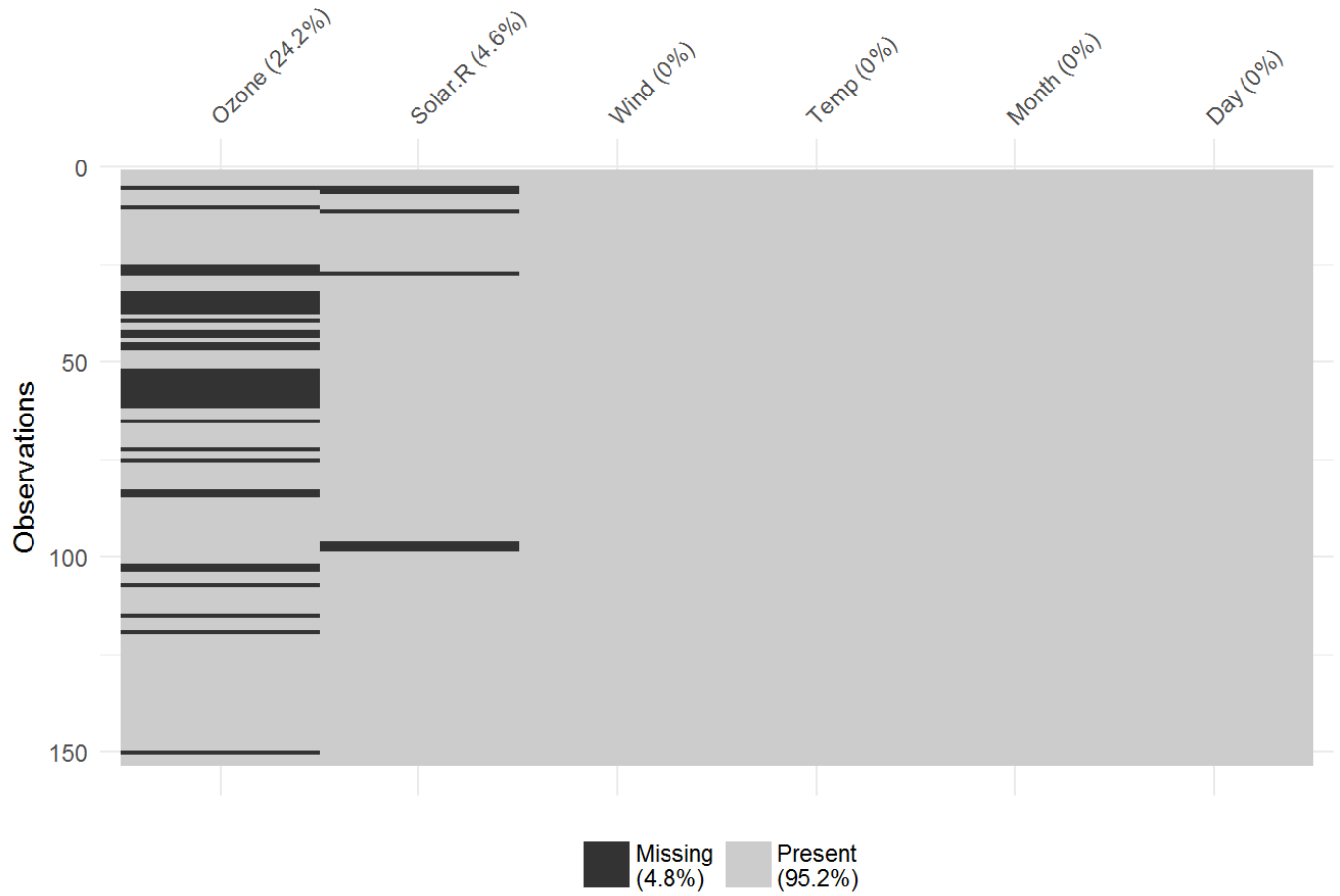

```
# Columns Data type and their missing values are also represnted.
vis_dat(typical_data)
```



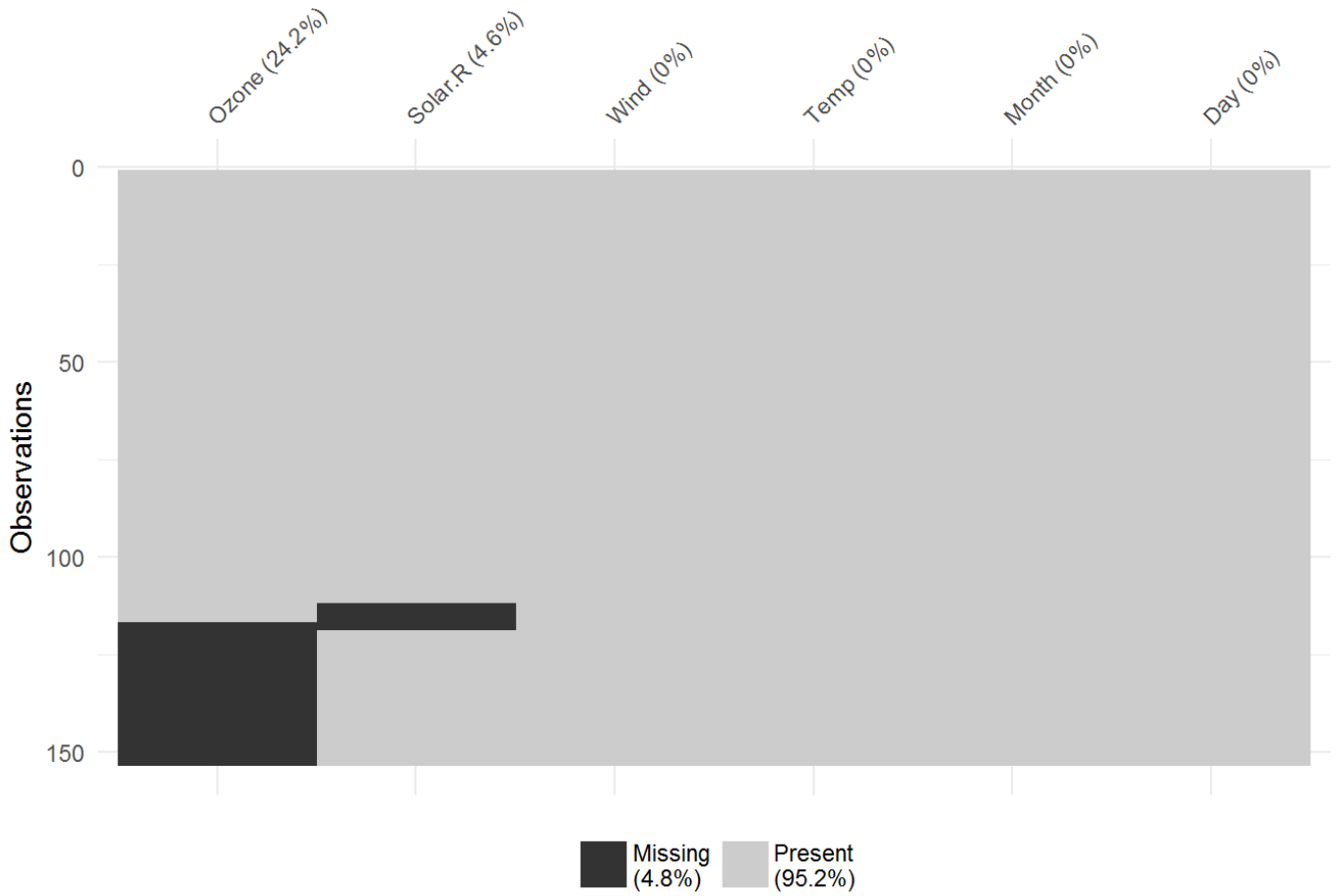

```
#By using vis_dat() we can look at many columns
vis_dat(typical_data_large)
```



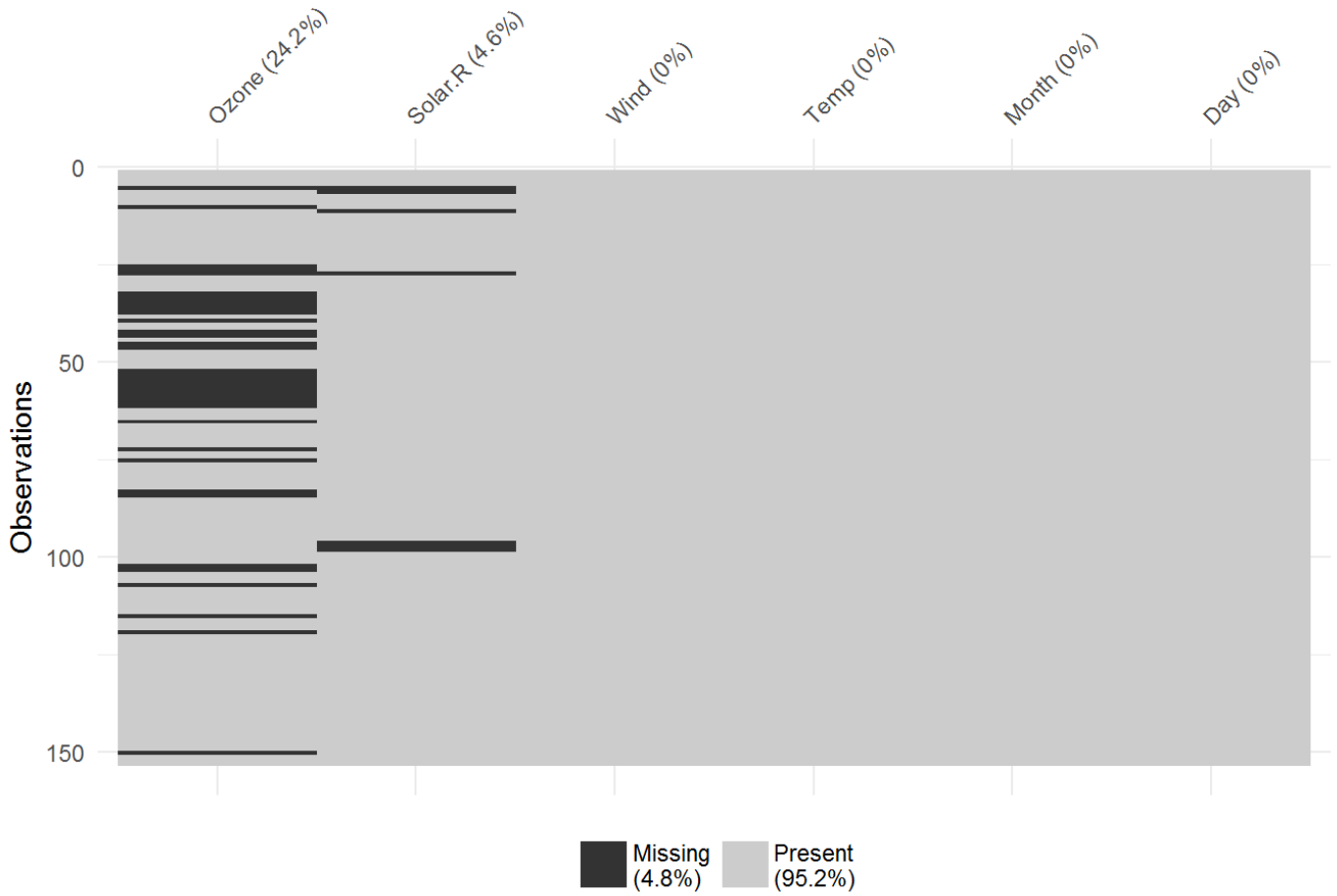

```
#We can explore the missing data further using vis_miss().

vis_miss(airquality)
```



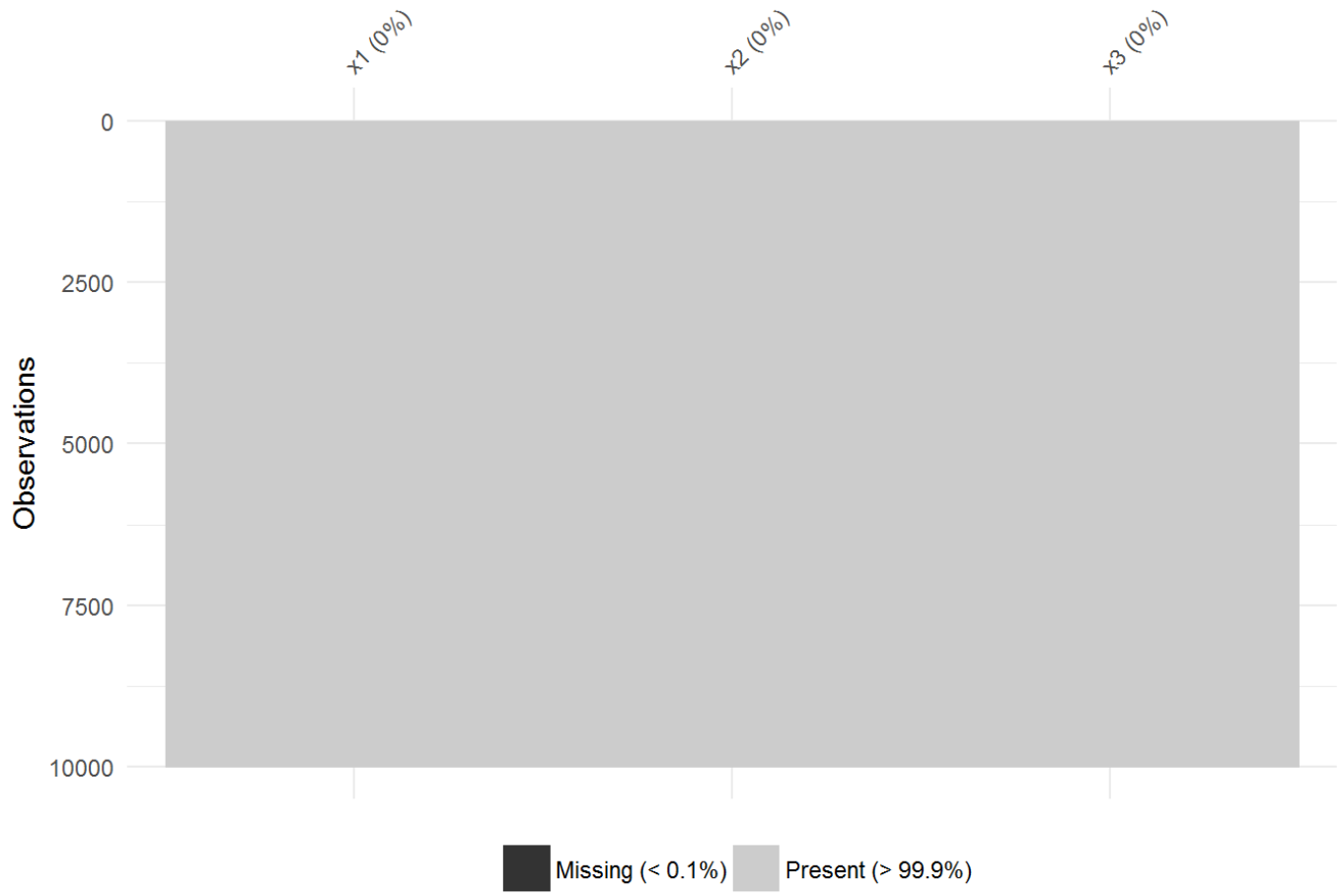

```
# It will cluster missing Data for simple understanding.
vis_miss(airquality,
         cluster = TRUE)
```




```
#The columns can also just be arranged by columns with most missingness, by setting
sort_miss = TRUE.

vis_miss(airquality,
         sort_miss = TRUE)
```




```
#vis_miss indicates when there is a very small amount of missing data at <0.1% miss
ingness.

test_miss_df <- data.frame(x1 = 1:10000,
                           x2 = rep("A", 10000),
                           x3 = c(rep(1L, 9999), NA))

vis_miss(test_miss_df)
```




```
#vis_miss will also indicate when there is no missing data at all.
vis_miss(mtcars)
```


mpg (0%)
cyl (0%)
disp (0%)
hp (0%)
drat (0%)
wt (0%)
qsec (0%)
vs (0%)
am (0%)
gear (0%)
carb (0%)
0
10
20
30
Observations
Present (100%)