

Chicago AQi EDA

Ally Racho

2/10/2022

```
library(jsonlite)
#chicago 2019 aqi
result <- fromJSON("./Data/ChicagoAQI2019.json")
chicagoAQI2019 <- as.data.frame(result)

## Warning in (function (... , row.names = NULL, check.rows = FALSE, check.names =
## TRUE, : row names were found from a short variable and have been discarded

#chicago 2020 aqi
result <- fromJSON("./Data/ChicagoAQI2020.json")
chicagoAQI2020 <- as.data.frame(result)

## Warning in (function (... , row.names = NULL, check.rows = FALSE, check.names =
## TRUE, : row names were found from a short variable and have been discarded

#chicago 2021 aqi
result <- fromJSON("./Data/ChicagoAQI2021.json")
chicagoAQI2021 <- as.data.frame(result)

## Warning in (function (... , row.names = NULL, check.rows = FALSE, check.names =
## TRUE, : row names were found from a short variable and have been discarded

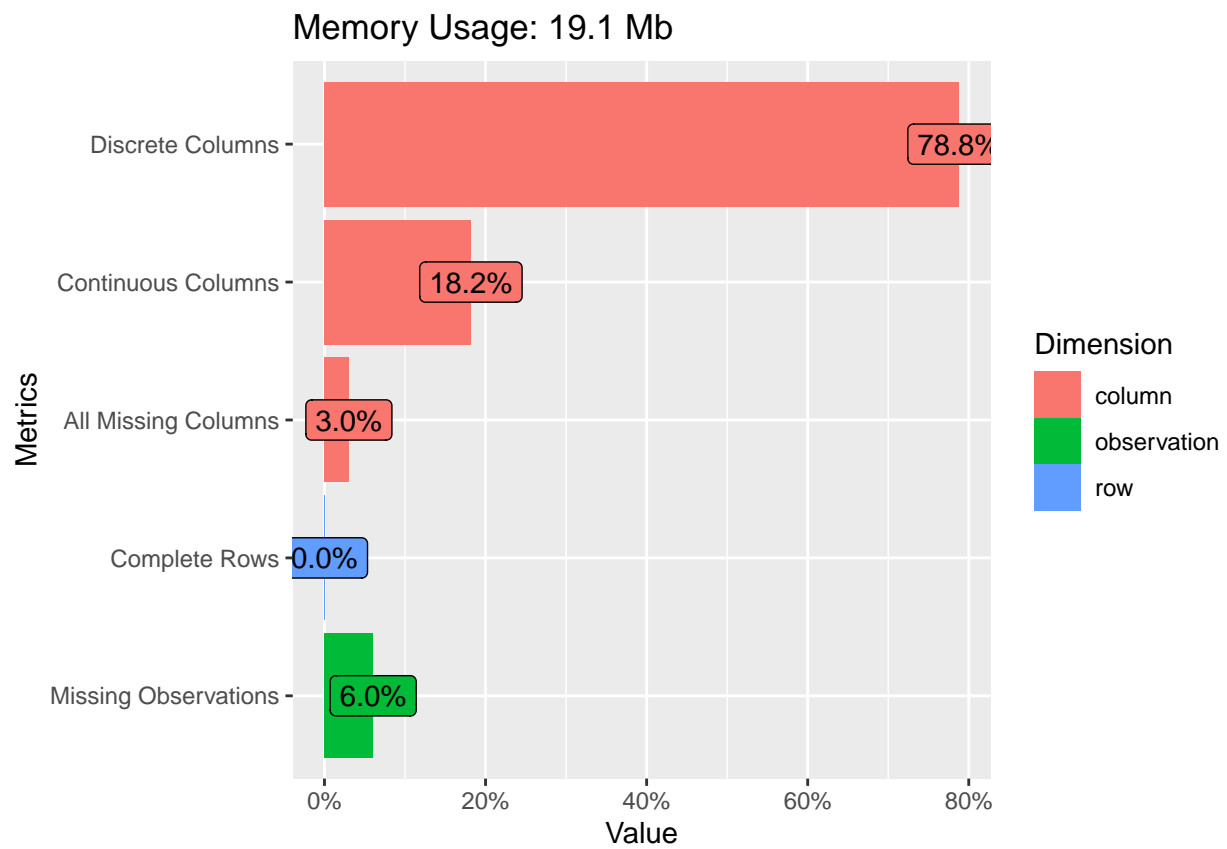
#combine all years to one dataset
df_list <- list(chicagoAQI2019, chicagoAQI2020, chicagoAQI2021)
ChicagoAQI <- Reduce(function(x,y)merge(x,y,all = TRUE), df_list, accumulate = FALSE)

library(DataExplorer)

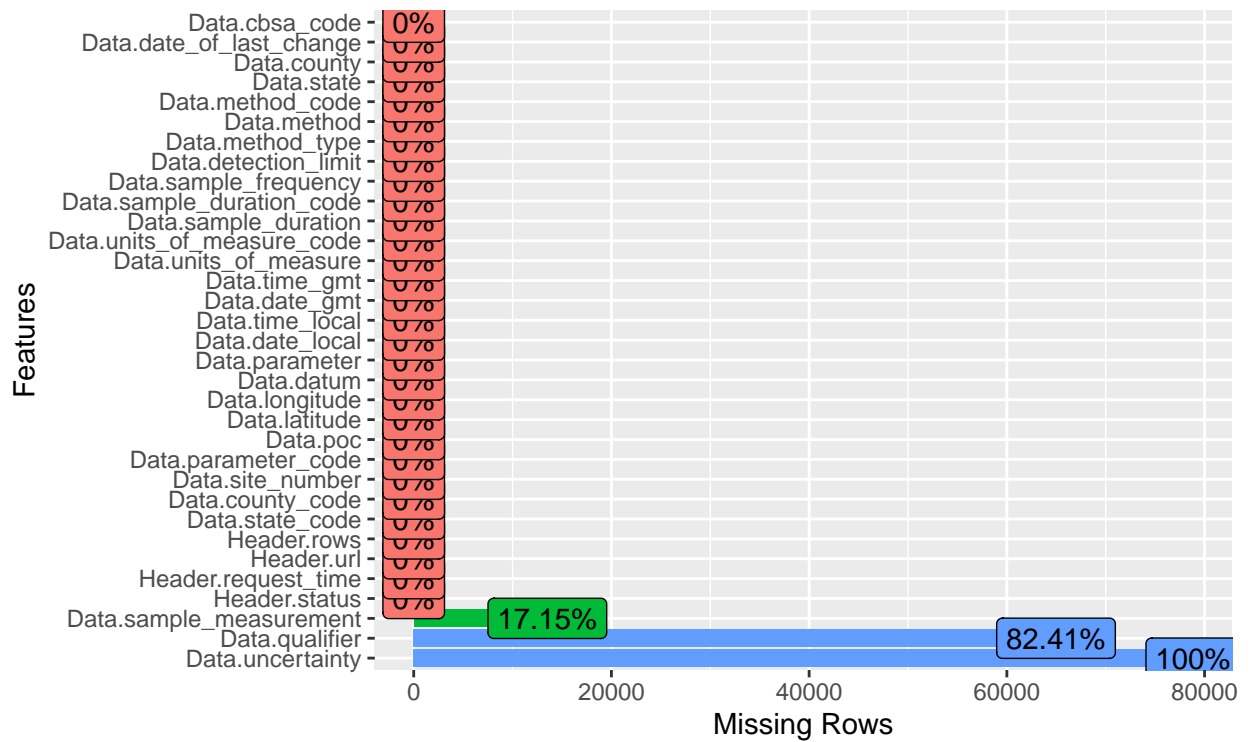
introduce(ChicagoAQI)

##      rows columns discrete_columns continuous_columns all_missing_columns
## 1 78838      33           26              6              1
##   total_missing_values complete_rows total_observations memory_usage
## 1           157330           0           2601654           20025096

plot_intro(ChicagoAQI)
```



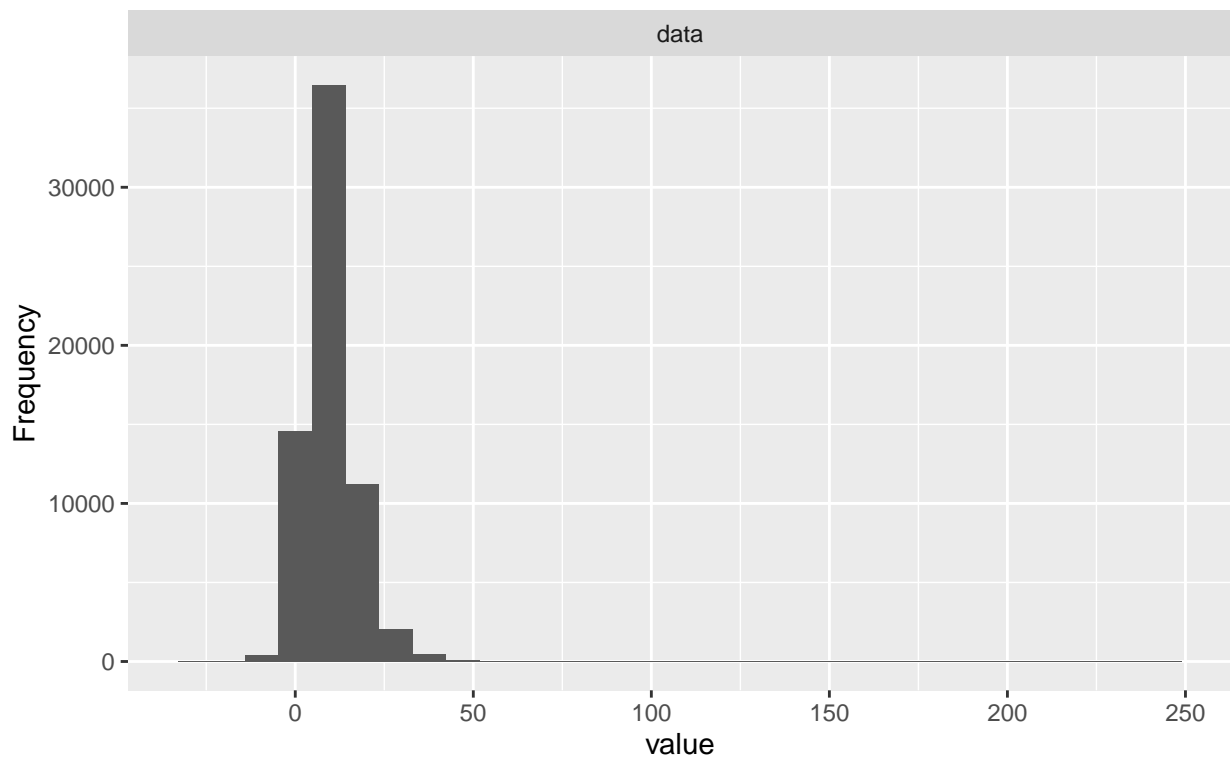
```
plot_missing(ChicagoAQI)
```



Band ■ Good ■ OK ■ Remove

```
plot_histogram(ChicagoAQI$Data.sample_measurement, title = 'Chiacao AQI Sample Measurements')
```

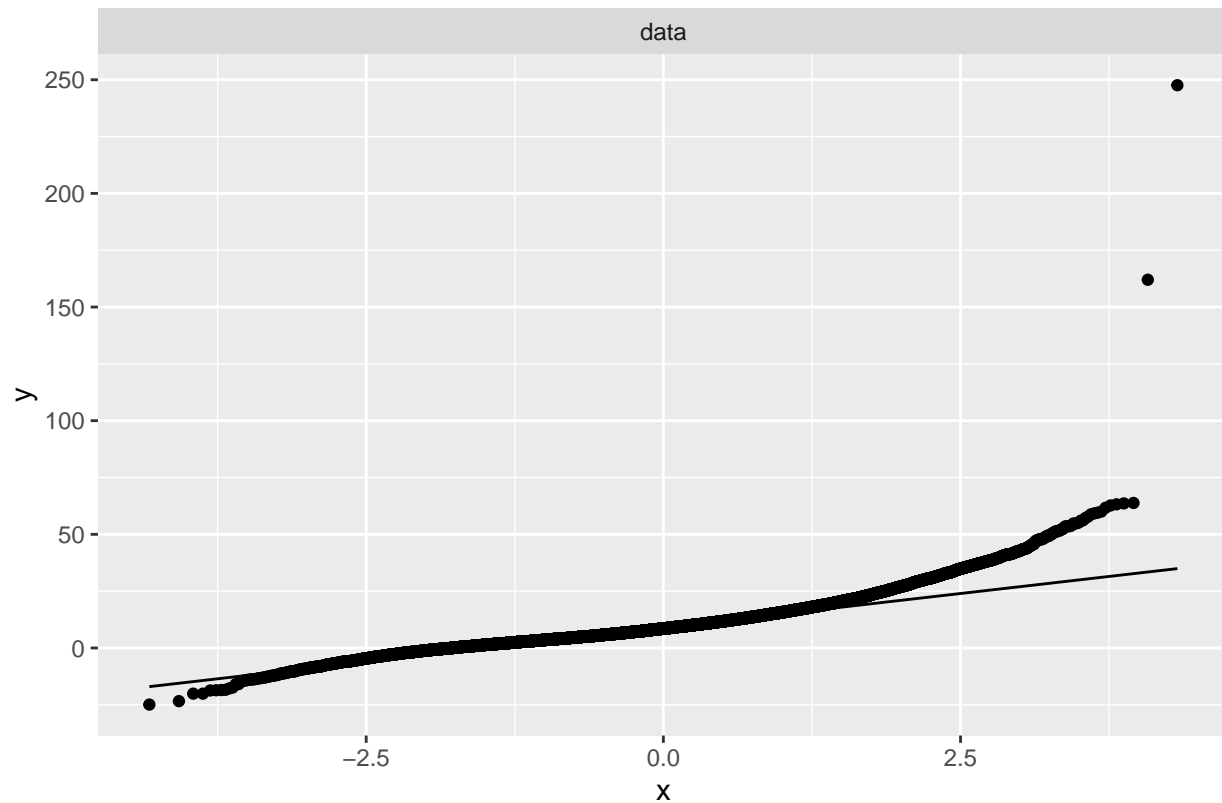
Chiacao AQI Sample Measurements



```
plot_qq(ChicagoAQI$Data.sample_measurement)
```

```
## Warning: Removed 13520 rows containing non-finite values (stat_qq).
```

```
## Warning: Removed 13520 rows containing non-finite values (stat_qq_line).
```



```
sample_meas <- ChicagoAQI[, c("Data.sample_measurement", "Data.date_local")]
```

```
plot_scatterplot(sample_meas, by = "Data.date_local", sampled_rows = 1000L)
```

```
## Warning: Removed 181 rows containing missing values (geom_point).
```

Data.date_local

