Philly AQI

Ally Racho

2/10/2022

```
library("jsonlite")

#########################
# Philadelphia AQI
#########################
#philly 2019 aqi
result <- fromJSON("./Data/PhillyAQI2019.json")
phillyAQI2019 <- 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
```

```
#philly 2021 aqi
result <- fromJSON("./Data/PhillyAQI2020.json")
phillyAQI2021 <- 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
```

```
#philly 2020 aqi
result <- fromJSON("./Data/PhilllyAQI2021.json")
phillyAQI2020 <- 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
```

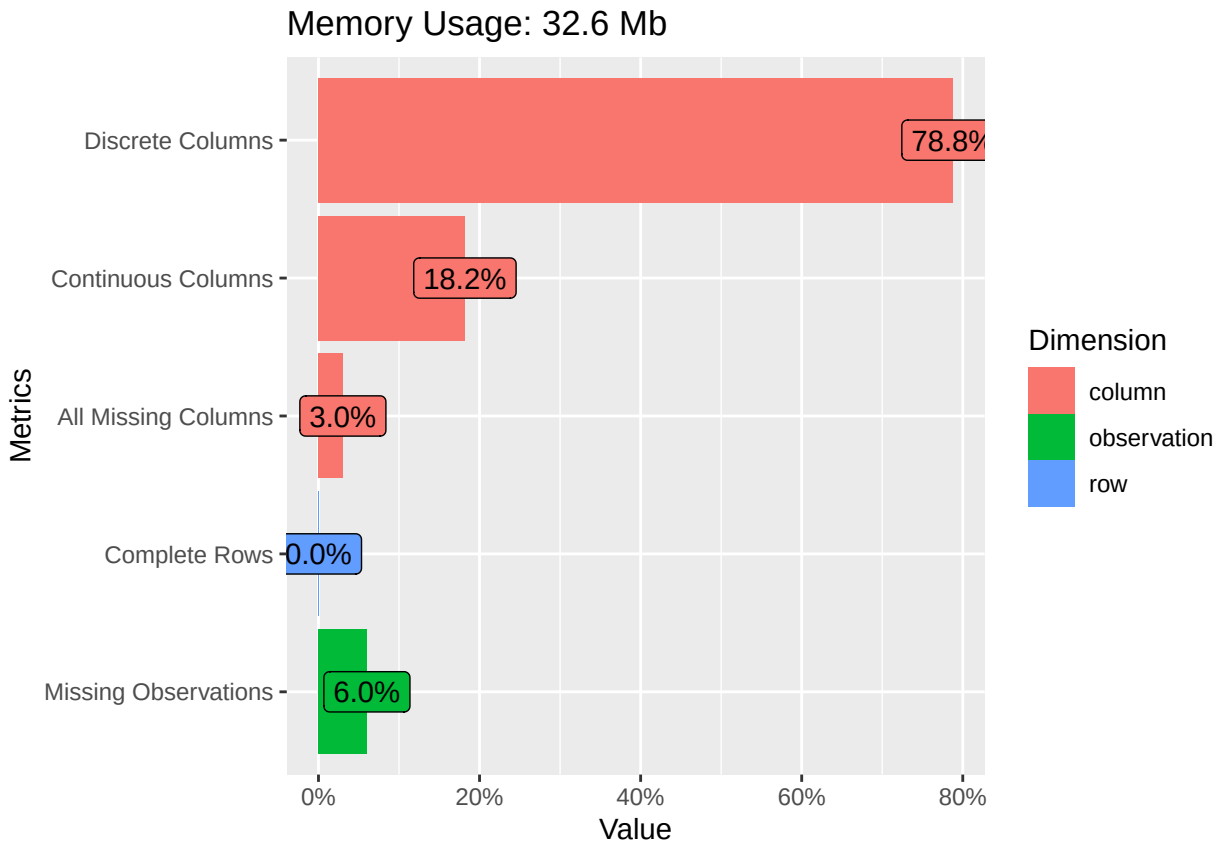
```
#combine all years to one dataset
df_list <- list(phillyAQI2019, phillyAQI2020, phillyAQI2021)
PhillyAQI <- Reduce(function(x,y)merge(x,y,all = TRUE), df_list, accumulate = FALSE)
```

```
library(DataExplorer)
```

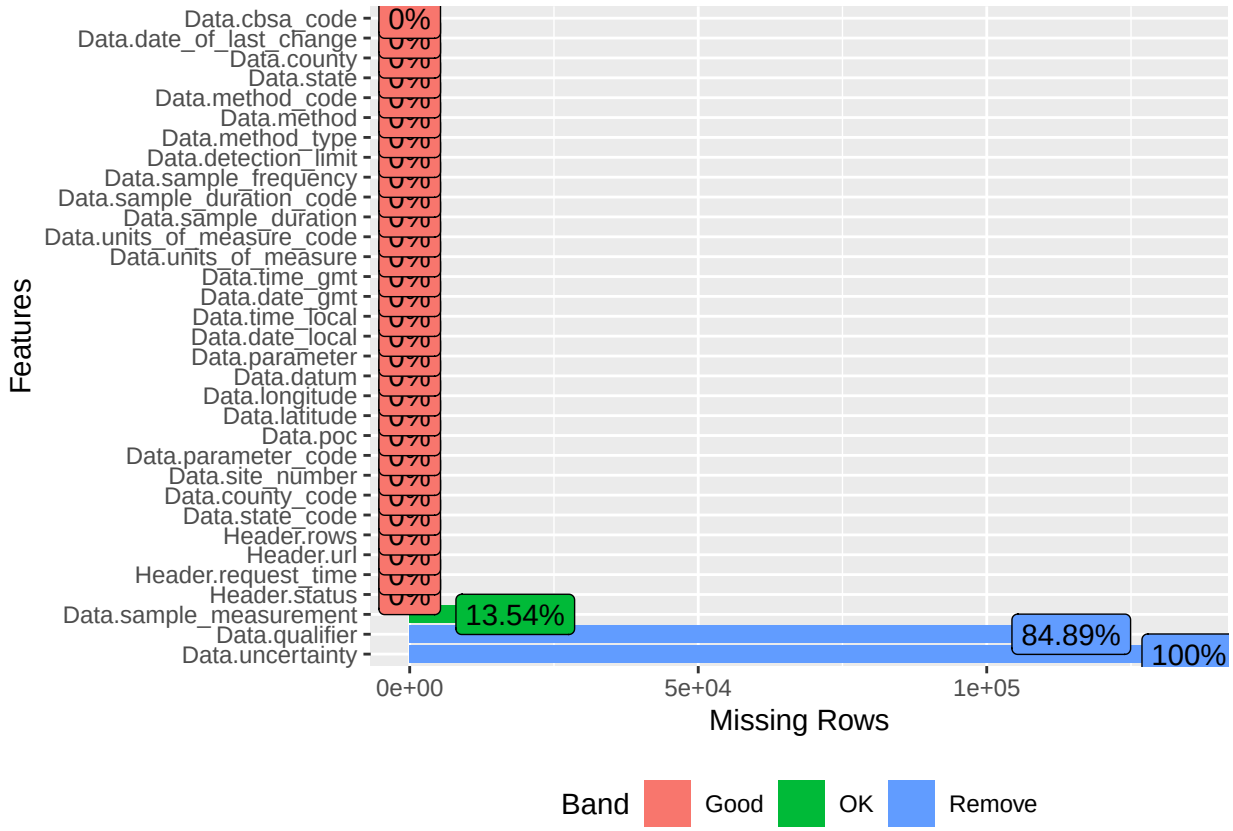
```
introduce(PhillyAQI)
```

```
##     rows columns discrete_columns continuous_columns all_missing_columns
## 1 135064      33               26                   6                   1
##   total_missing_values complete_rows total_observations memory_usage
## 1               268008             0            4457112     34182080
```

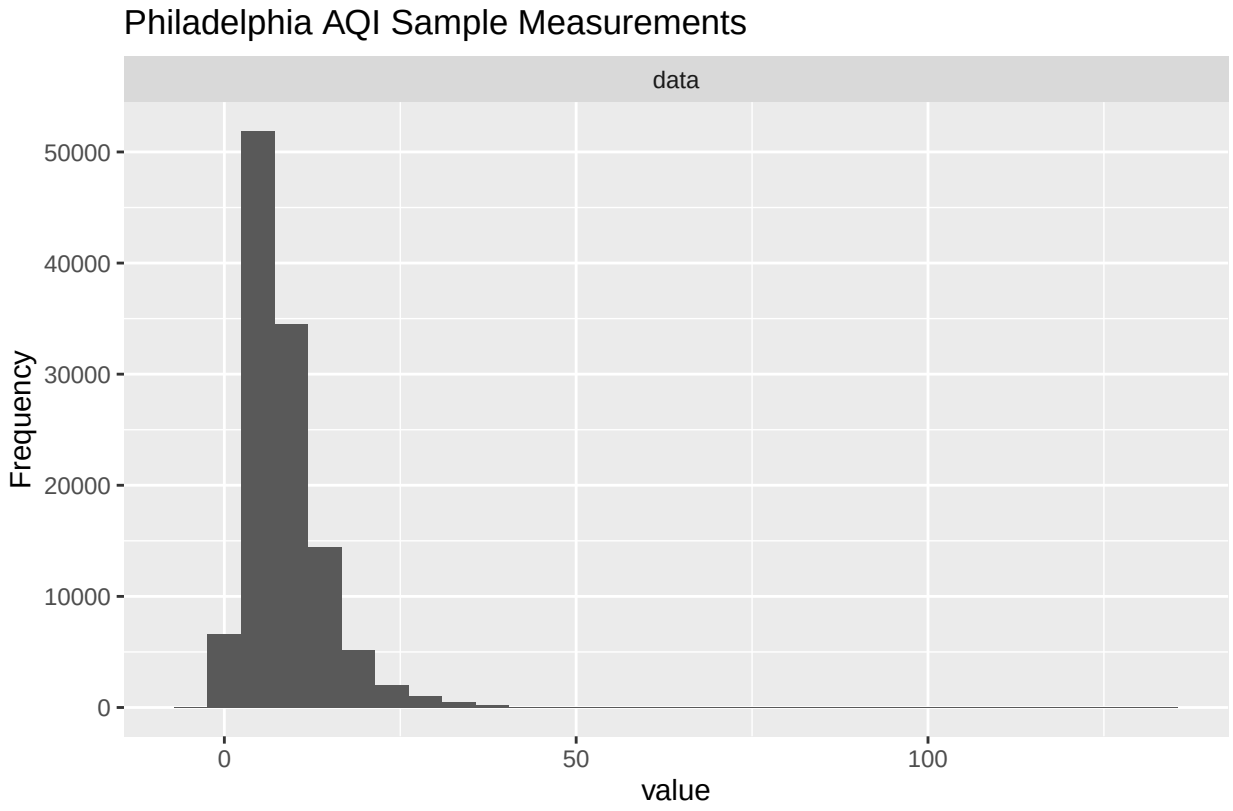
```
plot_intro(PhillyAQI)
```




```
plot_missing(PhillyAQI)
```



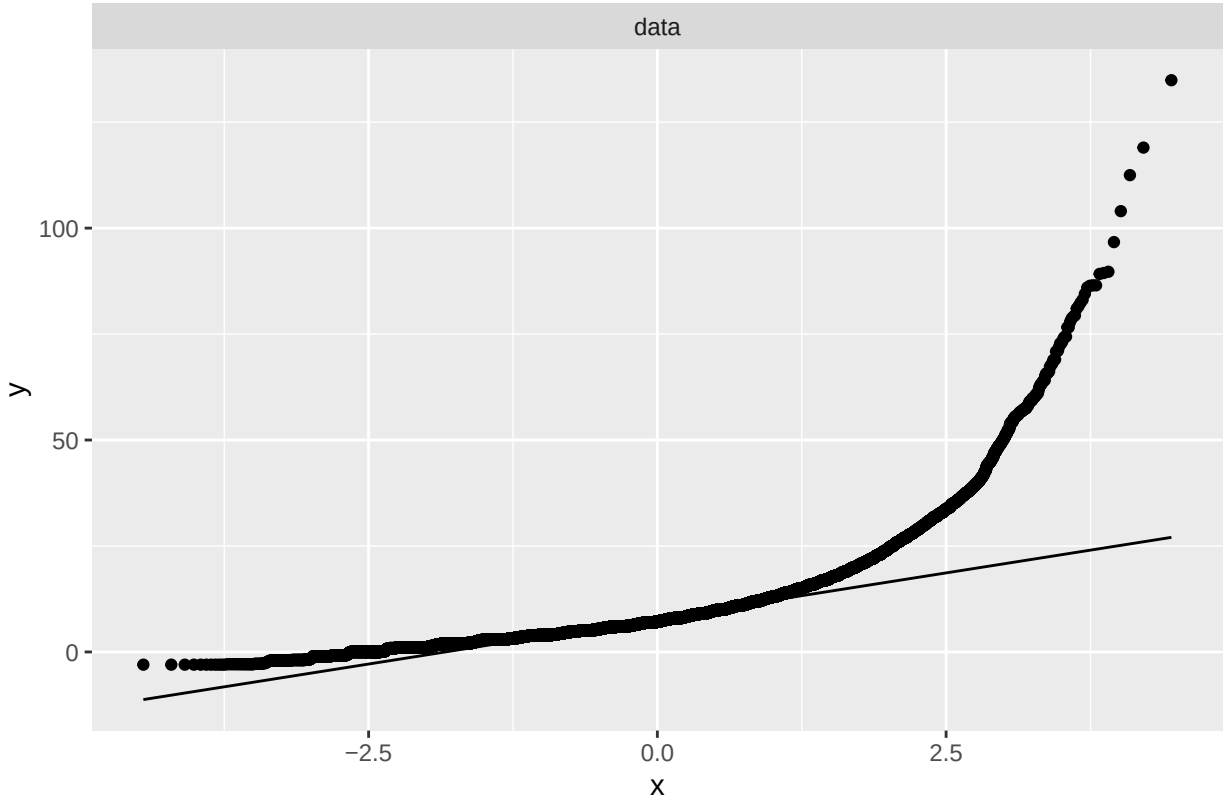

```
plot_histogram(PhillyAQI$Data.sample_measurement, title = 'Philadelphia AQI Sample Measurements')
```



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

```
## Warning: Removed 18294 rows containing non-finite values (stat_qq).
```

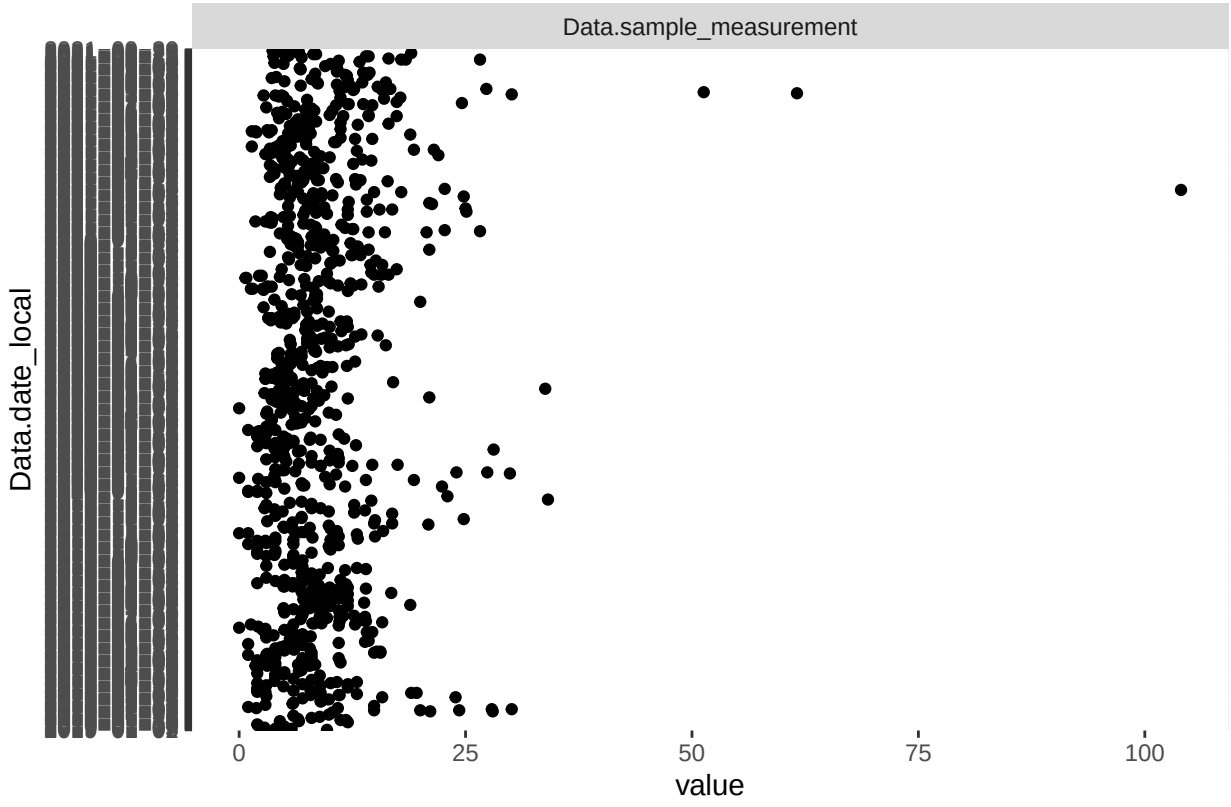
```
## Warning: Removed 18294 rows containing non-finite values (stat_qq_line).
```




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

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

```
## Warning: Removed 140 rows containing missing values (geom_point).
```


Data.sample_measurement
Data.date_local
0
25
50
75
100
value