Modeling extreme values with a GEV mixture probability distributions

Application to a rain data in australia

Pascal Alain Dkengne Sielenou

2023-10-26

```
# library(xfun)
```

```
path <- ".."
```

```
xfun::in_dir(dir = path, expr = source("./src/generate_gev_sample.R"))
xfun::in_dir(dir = path, expr = source("./src/calculate_gev_inverse_cdf.R"))
xfun::in_dir(dir = path, expr = source("./src/estimate_gev_mixture_model_parameters.R"))
xfun::in_dir(dir = path, expr = source("./src/predict_gev_mixture_model_parameters.R"))
xfun::in_dir(dir = path, expr = source("./src/plot_gev_mixture_model_pdf.R"))
xfun::in_dir(dir = path, expr = source("./src/plot_gev_mixture_model_cdf.R"))
xfun::in_dir(dir = path, expr = source("./src/estimate_gev_mixture_model_quantile.R"))
xfun::in_dir(dir = path, expr = source("./src/estimate_gev_mixture_model_pdf.R"))
xfun::in_dir(dir = path, expr = source("./src/estimate_gev_mixture_model_cdf.R"))
xfun::in_dir(dir = path, expr = source("./src/estimate_gev_mixture_model_sample.R"))
```

```
library(readr)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v purrr     1.0.2
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(DataExplorer)
#library(tibble)
#library(explore)
```

```
weatherAUS <- xfun::in_dir(dir = path, expr = read_csv("./applications/weatherAUS.csv"))
```

```
## Rows: 235699 Columns: 24
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr   (6): Location, WindGustDir, WindDir9am, WindDir3pm, RainToday, RainTom...
## dbl  (17): MinTemp, MaxTemp, Rainfall, Evaporation, Sunshine, WindGustSpeed,...
## date  (1): Date
##
```

```
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# View(weatherAUS)
```

```
#str(weatherAUS)
```

```
names(weatherAUS)
```

```
##  [1] "Date"          "Location"      "MinTemp"       "MaxTemp"
##  [5] "Rainfall"      "Evaporation"   "Sunshine"      "WindGustDir"
##  [9] "WindGustSpeed" "WindDir9am"    "WindDir3pm"    "WindSpeed9am"
## [13] "WindSpeed3pm"  "Humidity9am"   "Humidity3pm"   "Pressure9am"
## [17] "Pressure3pm"   "Cloud9am"      "Cloud3pm"      "Temp9am"
## [21] "Temp3pm"       "RainToday"     "RISK_MM"       "RainTomorrow"
```

```
head(weatherAUS)
```

```
## # A tibble: 6 x 24
##   Date       Location MinTemp MaxTemp Rainfall Evaporation Sunshine WindGustDir
##   <date>     <chr>      <dbl>   <dbl>    <dbl>       <dbl>    <dbl> <chr>
## 1 2008-12-01 Albury      13.4    22.9      0.6          NA       NA W
## 2 2008-12-02 Albury       7.4    25.1      0            NA       NA WNW
## 3 2008-12-03 Albury      12.9    25.7      0            NA       NA WSW
## 4 2008-12-04 Albury       9.2    28        0            NA       NA NE
## 5 2008-12-05 Albury      17.5    32.3      1            NA       NA W
## 6 2008-12-06 Albury      14.6    29.7      0.2          NA       NA WNW
## # i 16 more variables: WindGustSpeed <dbl>, WindDir9am <chr>, WindDir3pm <chr>,
## #   WindSpeed9am <dbl>, WindSpeed3pm <dbl>, Humidity9am <dbl>,
## #   Humidity3pm <dbl>, Pressure9am <dbl>, Pressure3pm <dbl>, Cloud9am <dbl>,
## #   Cloud3pm <dbl>, Temp9am <dbl>, Temp3pm <dbl>, RainToday <chr>,
## #   RISK_MM <dbl>, RainTomorrow <chr>
```

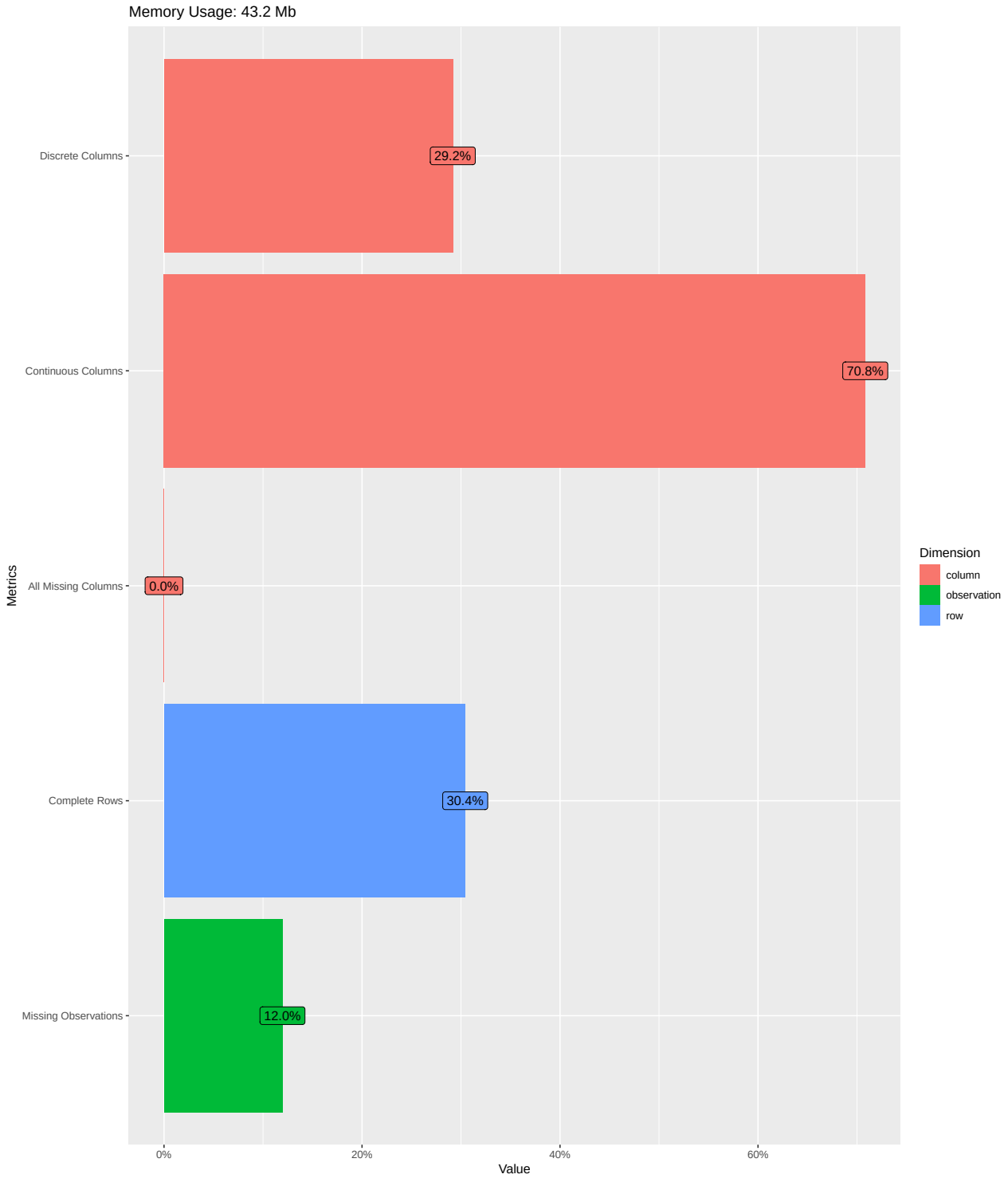
```
tail(weatherAUS)
```

```
## # A tibble: 6 x 24
##   Date       Location MinTemp MaxTemp Rainfall Evaporation Sunshine WindGustDir
##   <date>     <chr>      <dbl>   <dbl>    <dbl>       <dbl>    <dbl> <chr>
## 1 2023-09-24 Uluru       14.1    34          0          NA       NA S
## 2 2023-09-25 Uluru       15.5    33.8        0          NA       NA ENE
## 3 2023-09-26 Uluru       15      34.1        0          NA       NA SSE
## 4 2023-09-27 Uluru       18.8    33          0          NA       NA SE
## 5 2023-09-28 Uluru       19.6    33.6        0          NA       NA E
## 6 2023-09-29 Uluru       17.2    34.2        0          NA       NA ENE
## # i 16 more variables: WindGustSpeed <dbl>, WindDir9am <chr>, WindDir3pm <chr>,
## #   WindSpeed9am <dbl>, WindSpeed3pm <dbl>, Humidity9am <dbl>,
## #   Humidity3pm <dbl>, Pressure9am <dbl>, Pressure3pm <dbl>, Cloud9am <dbl>,
## #   Cloud3pm <dbl>, Temp9am <dbl>, Temp3pm <dbl>, RainToday <chr>,
## #   RISK_MM <dbl>, RainTomorrow <chr>
```

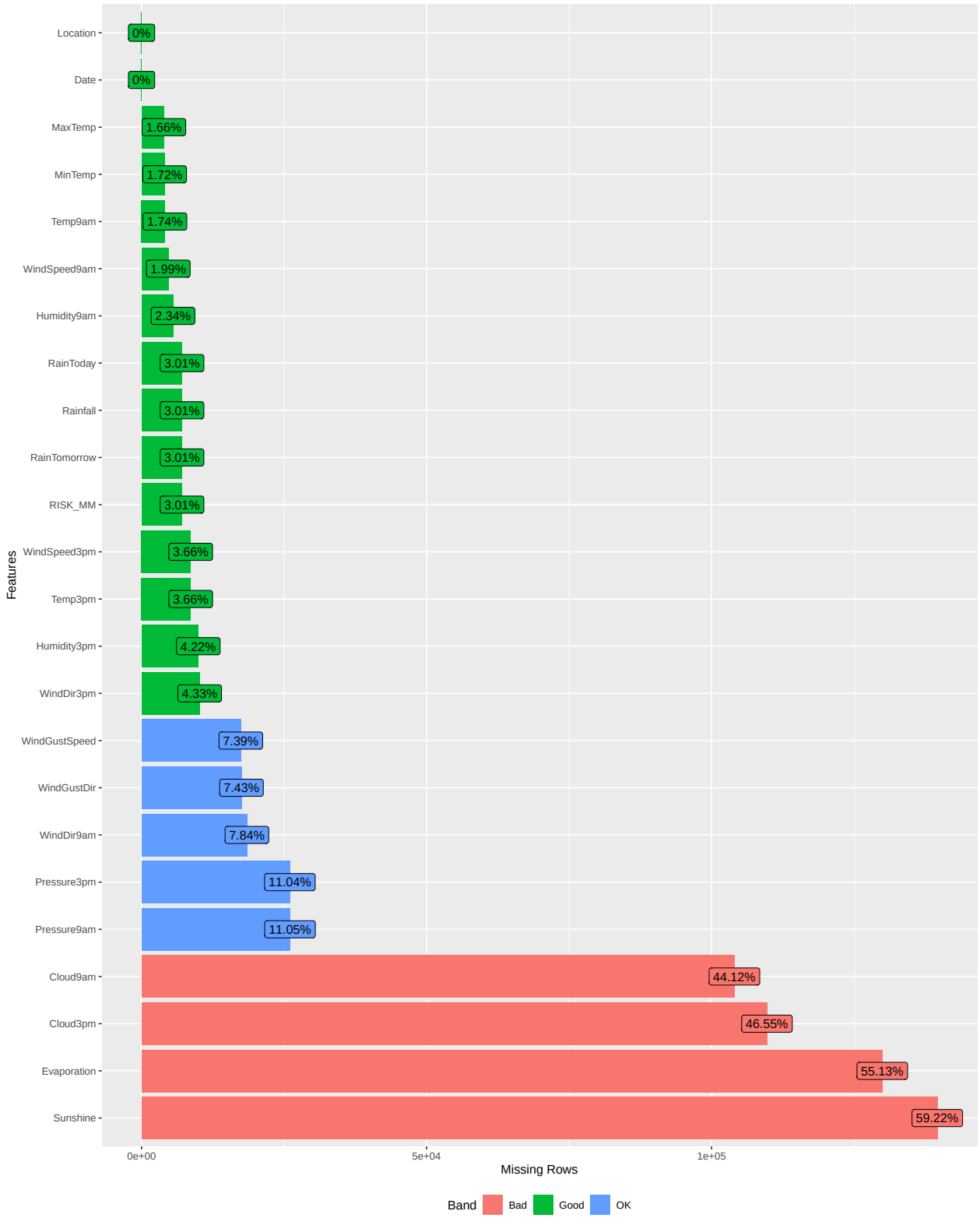
```
# plot_str(weatherAUS)

# introduce(weatherAUS)
```

```
plot_intro(weatherAUS)
```




```
plot_missing(weatherAUS)
```



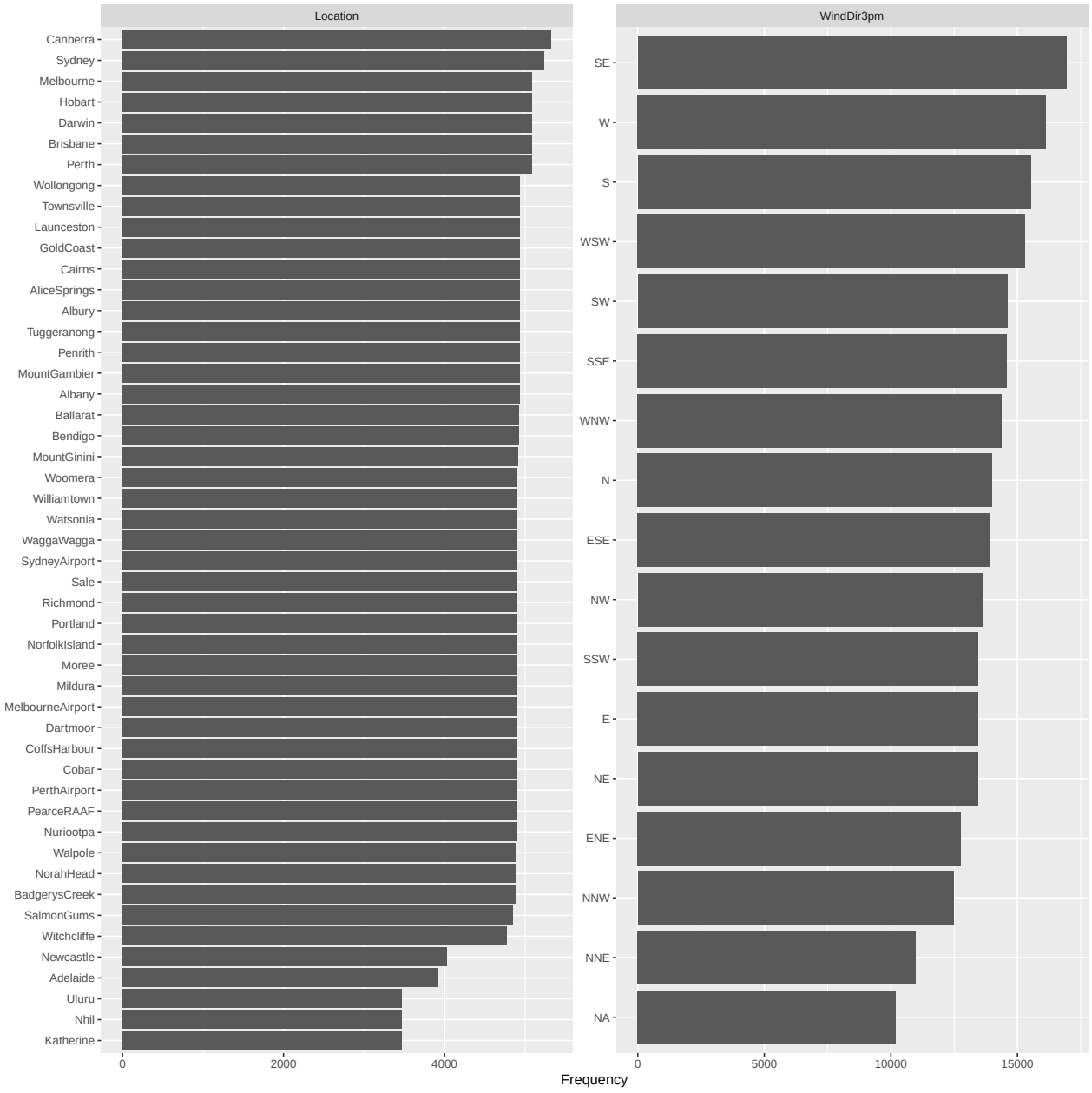

```
final_data <- drop_columns(weatherAUS, c("Date", "RISK_MM", "Cloud9am", "Cloud3pm", "Evaporation", "Suns
                                         "WindGustSpeed", "WindDir9am", "Pressure9am", "Pressure3pm", "

# names(final_data)
```

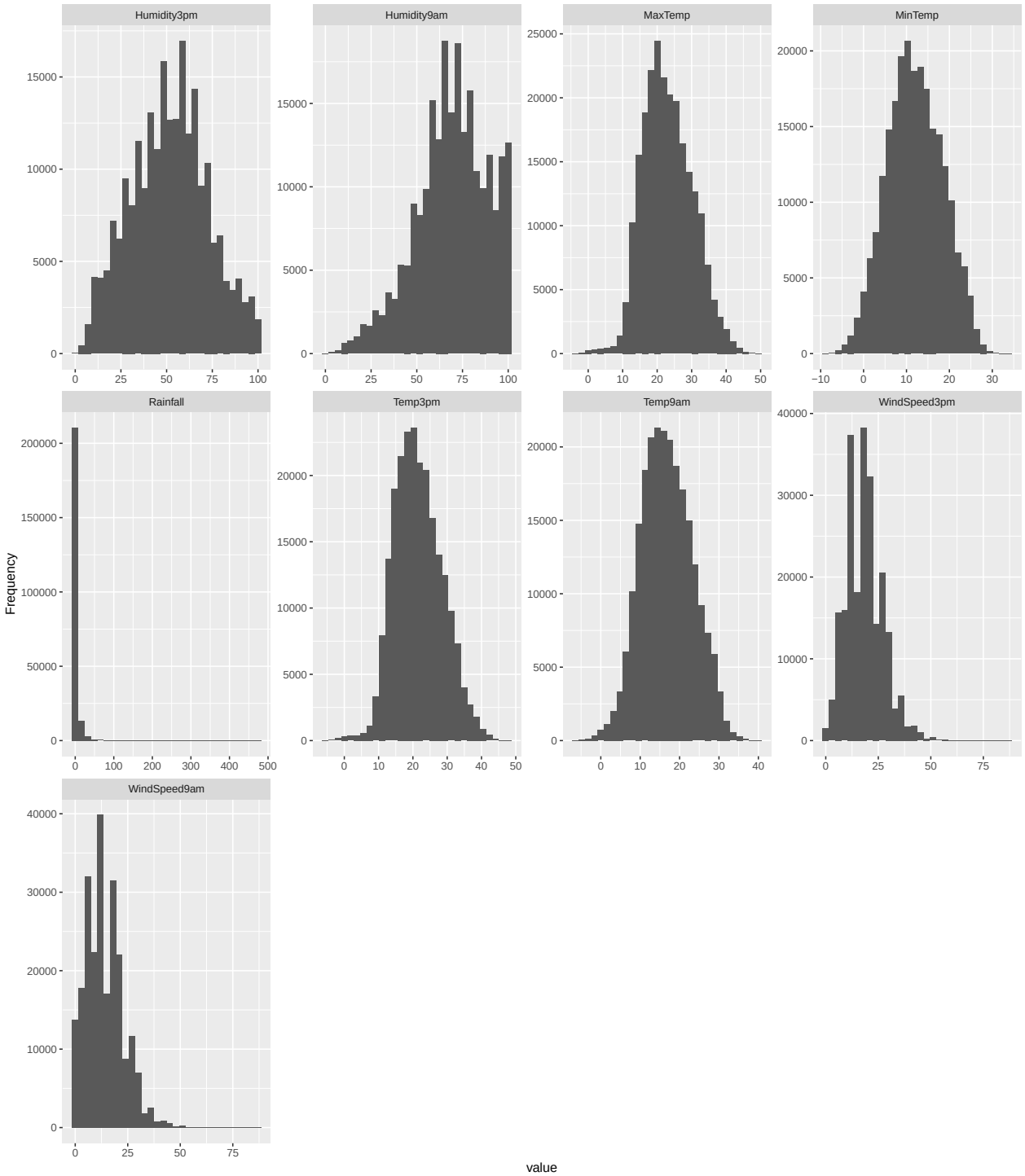
```
# View(final_data)

# profile_missing(final_data)
```

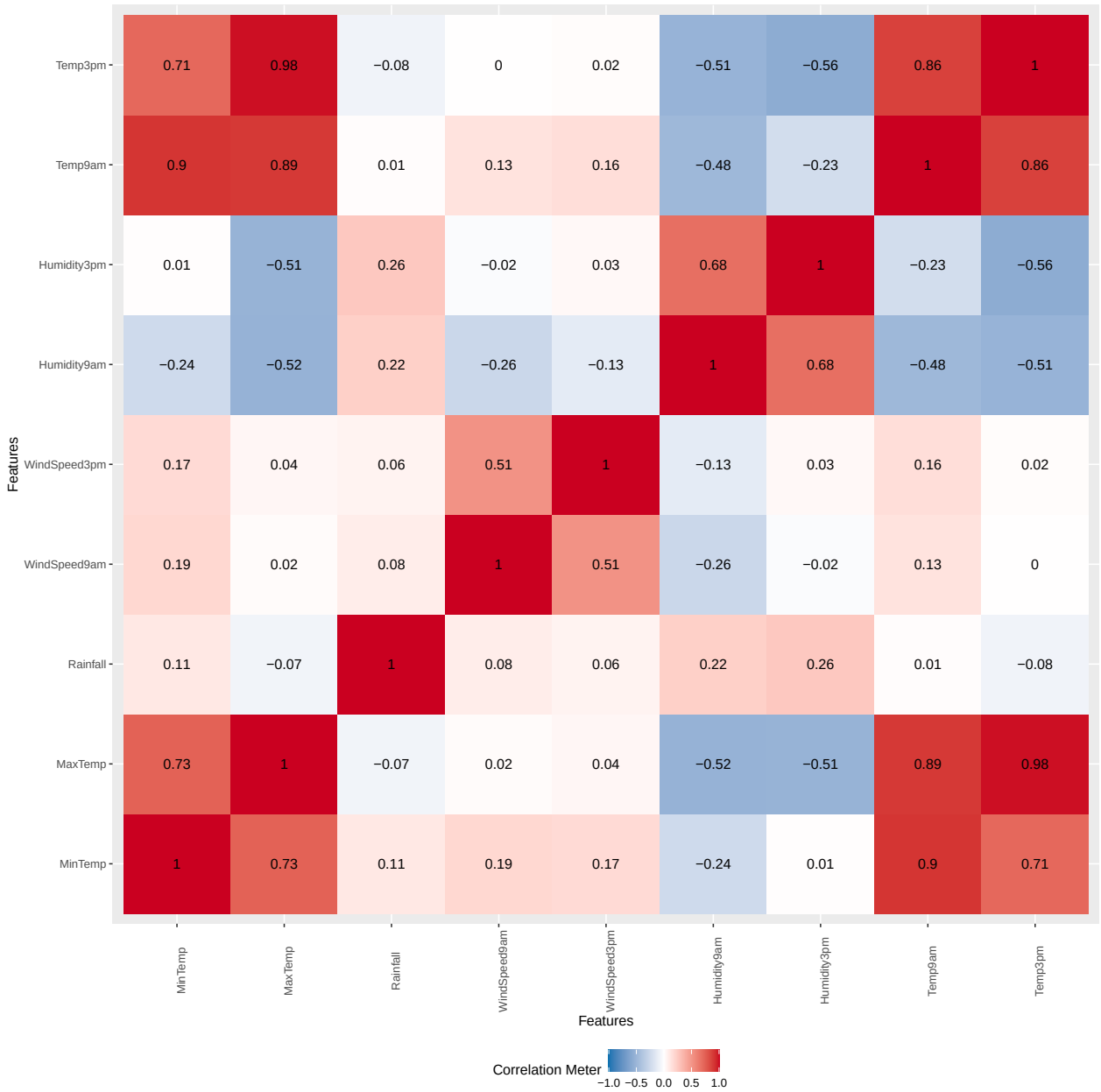
```
plot_bar(final_data)
```




```
plot_histogram(final_data)
```

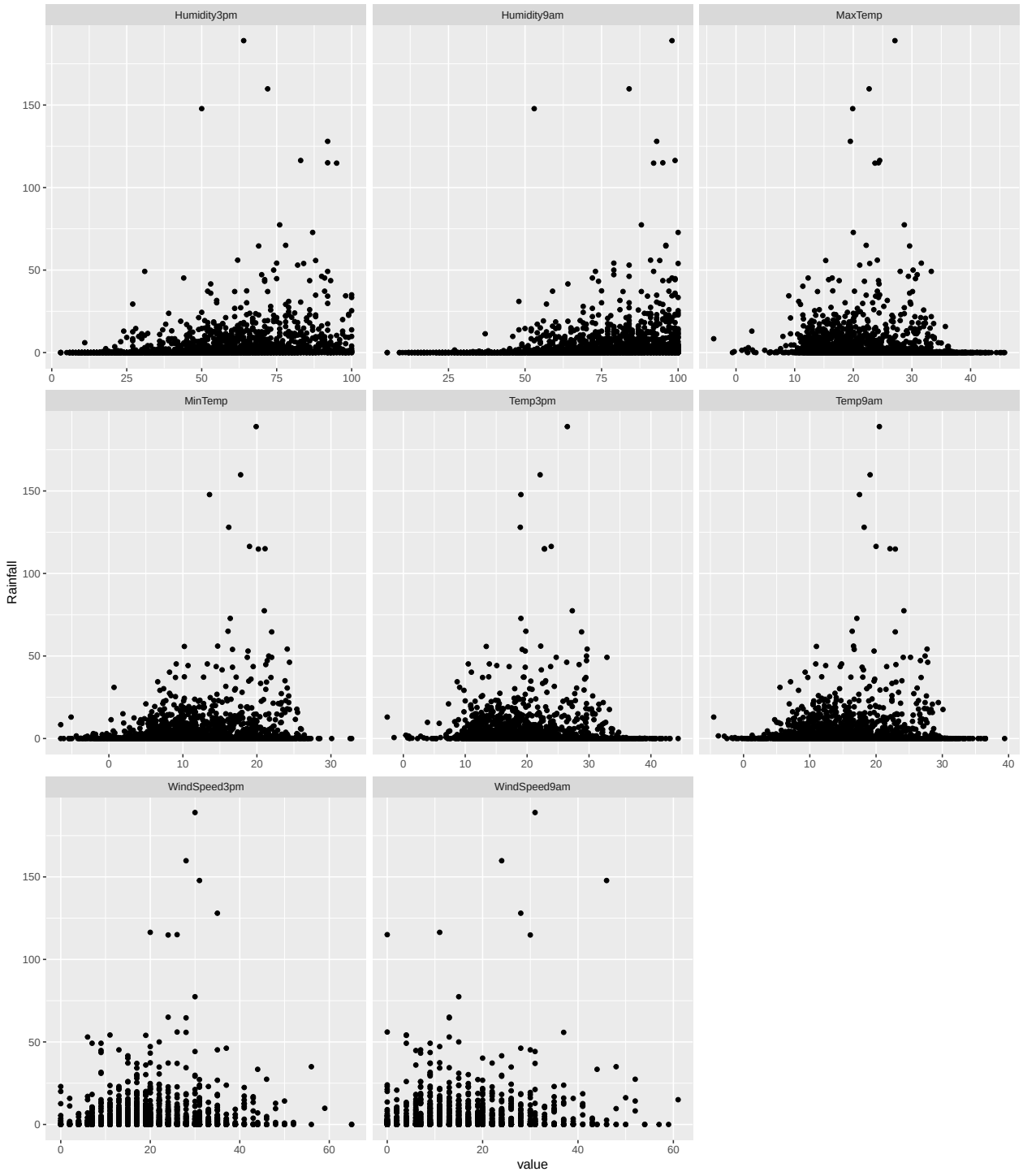



```
plot_correlation(na.omit(final_data), type = "c")
```




```
plot_scatterplot(final_data[, c("MinTemp", "MaxTemp", "Rainfall",  "WindSpeed9am",  "WindSpeed3pm",  "Hu
                                "Humidity3pm",  "Temp9am", "Temp3pm")],
                 by = "Rainfall",
                 sampled_rows = 3000L)
```

```
## Warning: Removed 1079 rows containing missing values (`geom_point()`).
```




```
final_data_clean <- na.omit(final_data)

summary(final_data_clean)
```

```
##    Location           MinTemp           MaxTemp           Rainfall
##  Length:215553      Min.   :-8.70000   Min.   :-3.70000   Min.   :  0.000000
##  Class :character   1st Qu.: 7.40000   1st Qu.:17.90000   1st Qu.:  0.000000
##  Mode  :character   Median :11.80000   Median :22.60000   Median :  0.000000
```

```
##                     Mean   :12.02255   Mean   :23.20927   Mean   :  2.296852
##                     3rd Qu.:16.70000   3rd Qu.:28.20000   3rd Qu.:  0.600000
##                     Max.   :33.90000   Max.   :48.90000   Max.   :474.000000
##   WindDir3pm         WindSpeed9am       WindSpeed3pm       Humidity9am
##  Length:215553      Min.   : 0.00000   Min.   : 2.00000   Min.   :  0.00000
##  Class :character   1st Qu.: 7.00000   1st Qu.:13.00000   1st Qu.: 57.00000
##  Mode  :character   Median :13.00000   Median :19.00000   Median : 70.00000
##                     Mean   :14.09672   Mean   :18.69989   Mean   : 68.78909
##                     3rd Qu.:19.00000   3rd Qu.:24.00000   3rd Qu.: 83.00000
##                     Max.   :87.00000   Max.   :87.00000   Max.   :100.00000
##   Humidity3pm          Temp9am           Temp3pm
##  Min.   :  0.00000   Min.   :-6.20000   Min.   :-5.10000
##  1st Qu.: 36.00000   1st Qu.:12.10000   1st Qu.:16.60000
##  Median : 52.00000   Median :16.60000   Median :21.10000
##  Mean   : 51.23395   Mean   :16.86624   Mean   :21.69488
##  3rd Qu.: 66.00000   3rd Qu.:21.40000   3rd Qu.:26.40000
##  Max.   :100.00000   Max.   :40.20000   Max.   :48.20000
```

```
# View(final_data_clean)
```