

```
# Load required libraries
install.packages("randomForest")
install.packages("keras3")
install.packages("xgboost")
install.packages("ggplot2")
```

⇒ Installing package into '/usr/local/lib/R/site-library'  
(as 'lib' is unspecified)

Installing package into '/usr/local/lib/R/site-library'  
(as 'lib' is unspecified)

also installing the dependencies 'RcppTOML', 'here', 'png', 'config', 'tfautogra

Installing package into '/usr/local/lib/R/site-library'  
(as 'lib' is unspecified)

Installing package into '/usr/local/lib/R/site-library'  
(as 'lib' is unspecified)



```
library(keras3)
library(xgboost)
library(randomForest)
library(ggplot2)
```

⇒ randomForest 4.7-1.2

Type rfNews() to see new features/changes/bug fixes.

Attaching package: 'ggplot2'

The following object is masked from 'package:randomForest':

margin

```
# Load and summarize the dataset
cat("Loading datasets...\n")
dataset_files <- c("/content/station00.csv") # Add other files if needed
```

⇒ Loading datasets...

```
# Initialize empty list to store datasets
datasets <- list()

# Loop through each dataset file and load into a list
for (file in dataset_files) {
  dataset <- read.csv(file) # Load each dataset
  datasets[[file]] <- dataset # Store in the list
  cat("Loaded dataset: ", file, "\n")
}
```

Loaded dataset: /content/station00.csv

```
# Combine all datasets into one large dataset
combined_dataset <- do.call(rbind, datasets)
cat("Combined dataset dimensions: ", dim(combined_dataset), "\n")
```

Combined dataset dimensions: 28896 15

```
# Summary and structure of dataset
cat("Summary of combined dataset:\n")
print(summary(combined_dataset))
cat("Structure of combined dataset:\n")
print(str(combined_dataset))
```

Summary of combined dataset:

date_time	nwp_globalirrad	nwp_directirrad	nwp_temperature
Length:28896	Min. : 0.0	Min. : 0.0	Min. : -14.01
Class :character	1st Qu.: 0.0	1st Qu.: 0.0	1st Qu.: 2.29
Mode :character	Median : 0.0	Median : 0.0	Median : 10.51
	Mean :168.4	Mean :147.8	Mean : 11.06
	3rd Qu.:305.5	3rd Qu.:259.2	3rd Qu.: 19.58
	Max. :942.8	Max. :885.6	Max. : 41.09
nwp_humidity	nwp_windspeed	nwp_winddirection	nwp_pressure
Min. : 5.07	Min. : 0.050	Min. : 0.03	Min. : 987.8
1st Qu.: 23.33	1st Qu.: 2.070	1st Qu.: 89.50	1st Qu.:1007.6
Median : 35.05	Median : 3.140	Median :186.89	Median :1015.4
Mean : 40.83	Mean : 3.539	Mean :184.02	Mean :1014.8
3rd Qu.: 54.46	3rd Qu.: 4.510	3rd Qu.:280.99	3rd Qu.:1021.1
Max. :100.00	Max. :15.980	Max. :360.00	Max. :1040.8
lmd_totalirrad	lmd_diffuseirrad	lmd_temperature	lmd_pressure
Min. : 0.0	Min. : 0.00	Min. : -13.50	Min. : 988.5
1st Qu.: 0.0	1st Qu.: 0.00	1st Qu.: 2.80	1st Qu.:1006.8
Median : 0.0	Median : 0.00	Median : 11.40	Median :1014.6
Mean : 167.4	Mean : 96.43	Mean : 11.51	Mean :1014.4
3rd Qu.: 275.0	3rd Qu.:138.00	3rd Qu.: 20.50	3rd Qu.:1020.8
Max. :1122.0	Max. :927.00	Max. : 36.80	Max. :1043.8
lmd_winddirection	lmd_windspeed	power	
Min. : 0	Min. : 0.000	Min. :0.0000	
1st Qu.:114	1st Qu.: 0.400	1st Qu.:0.0000	
Median :177	Median : 1.100	Median :0.0000	
Mean :181	Mean : 1.471	Mean :0.8311	
3rd Qu.:265	3rd Qu.: 2.100	3rd Qu.:1.3957	

```

Max.      :359      Max.      :13.300      Max.      :5.5230
Structure of combined dataset:
'data.frame':  28896 obs. of  15 variables:
 $ date_time      : chr  "2018-08-15 16:00:00" "2018-08-15 16:15:00" "2018-08-
 $ nwp_globalirrad : num  0 0 0 0 0 0 0 0 0 0 ...
 $ nwp_directirrad : num  0 0 0 0 0 0 0 0 0 0 ...
 $ nwp_temperature : num  22.8 22.8 22.7 22.6 22.6 ...
 $ nwp_humidity    : num  96.8 96.9 97 97.1 97.2 ...
 $ nwp_windspeed   : num  4.28 4.3 4.28 4.28 4.33 4.39 4.44 4.55 4.74 5 ...
 $ nwp_winddirection: num  339 337 334 332 330 ...
 $ nwp_pressure    : num  1007 1007 1007 1007 1007 ...
 $ lmd_totalirrad  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ lmd_diffuseirrad : int  0 0 0 0 0 0 0 0 0 0 ...
 $ lmd_temperature : num  25.9 25.9 25.8 25.6 25.7 ...
 $ lmd_pressure    : num  1006 1006 1006 1006 1006 ...
 $ lmd_winddirection: int  353 330 1 309 335 343 5 342 331 359 ...
 $ lmd_windspeed   : num  1.1 0.9 1.9 0.4 1.6 1.1 1.4 1.5 2.5 2.8 ...
 $ power          : num  0 0 0 0 0 0 0 0 0 0 ...
NULL

```

```
# Checking for missing values
```

```
cat("Checking for missing values...\n")
```

```
missing_values <- sum(is.na(combined_dataset))
```

```
cat("Total missing values in dataset: ", missing_values, "\n")
```



```
Checking for missing values...
```

```
Total missing values in dataset: 0
```

```
# Visualize dataset histograms and boxplots
```

```
cat("Visualizing dataset distributions...\n")
```

```
for (i in 2:(ncol(combined_dataset)-1)) {
```

```
  column_name <- names(combined_dataset)[i]
```

```
  hist(combined_dataset[[i]], main = paste("Histogram of", column_name), xlab = colu
```

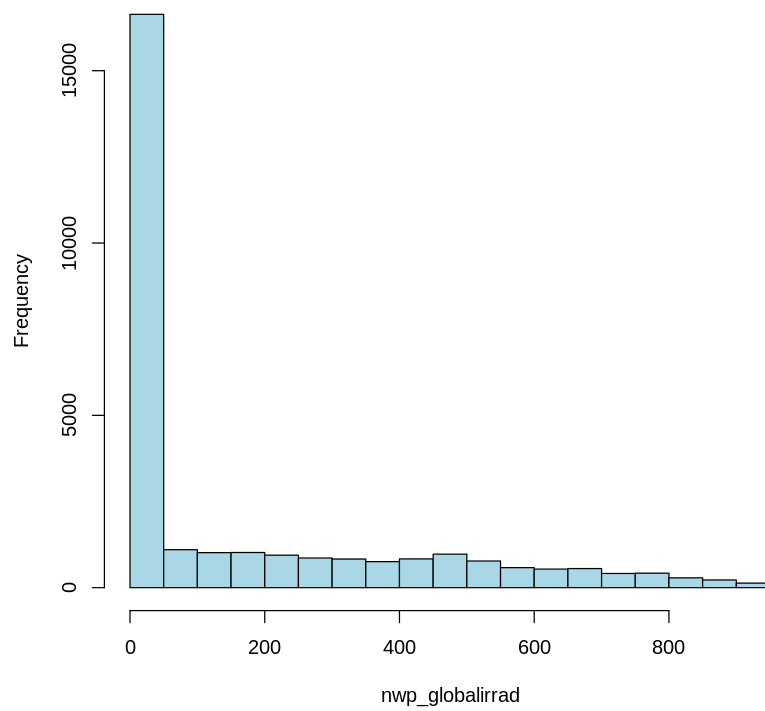
```
  boxplot(combined_dataset[[i]], main = paste("Boxplot of", column_name), col = "ora
```

```
}
```

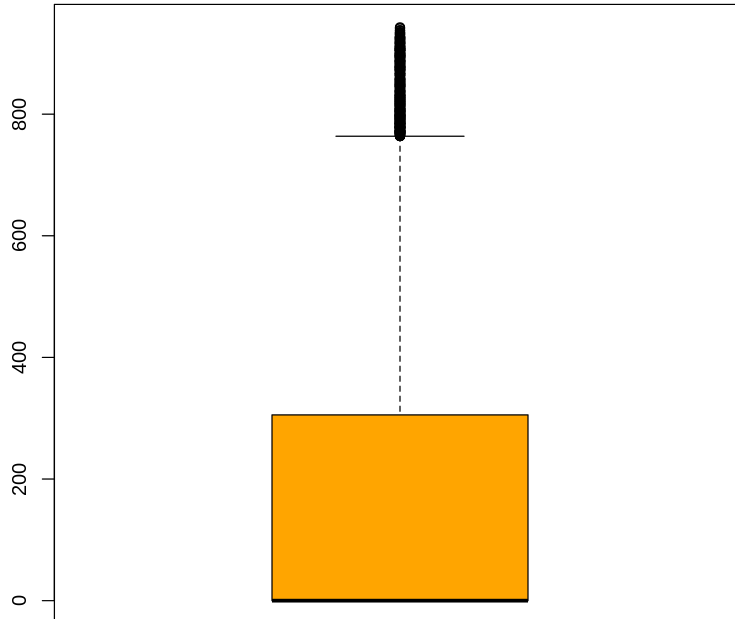


Visualizing dataset distributions...

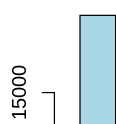
Histogram of nwp\_globalirrad

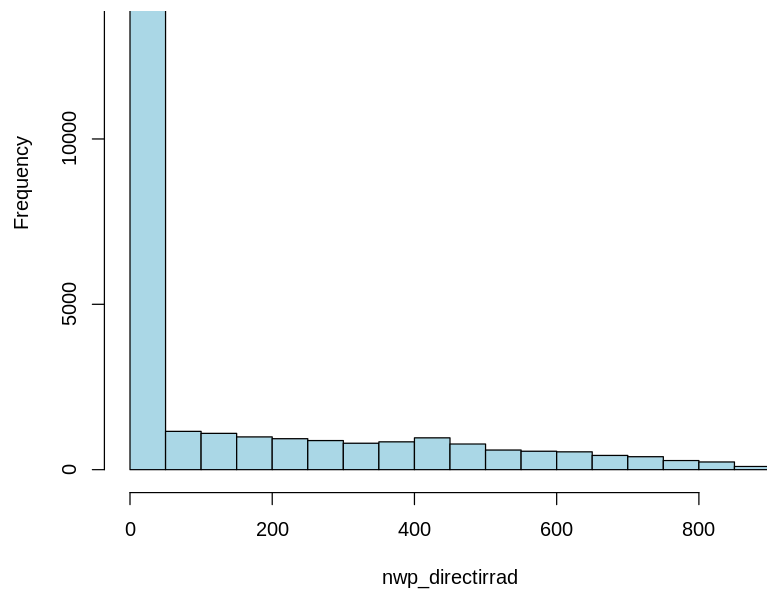
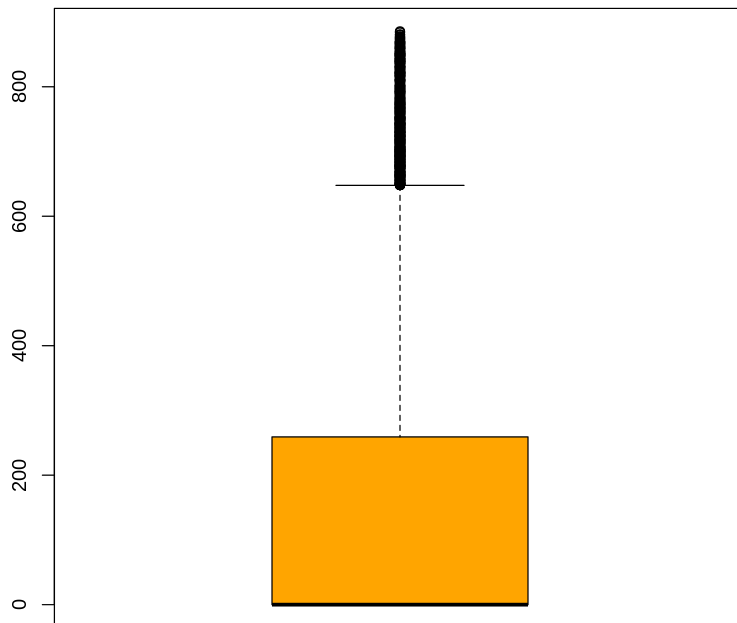
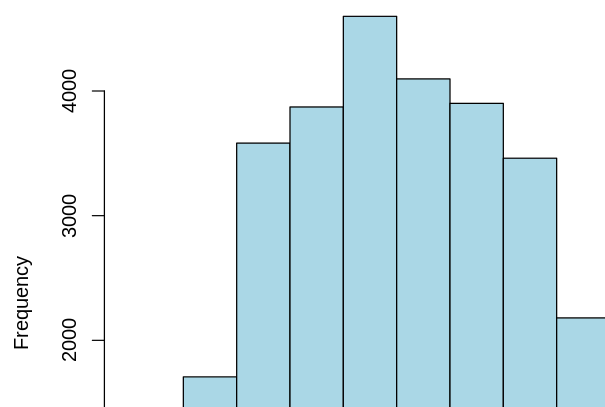


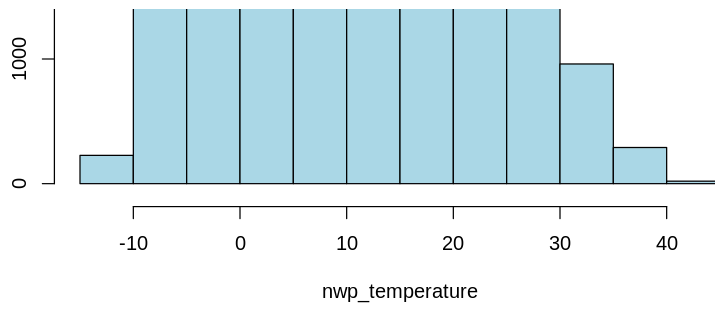
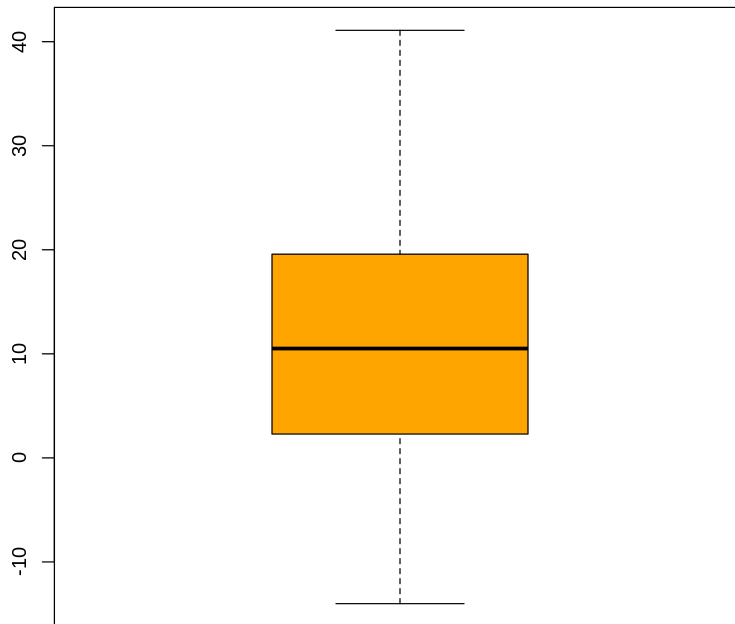
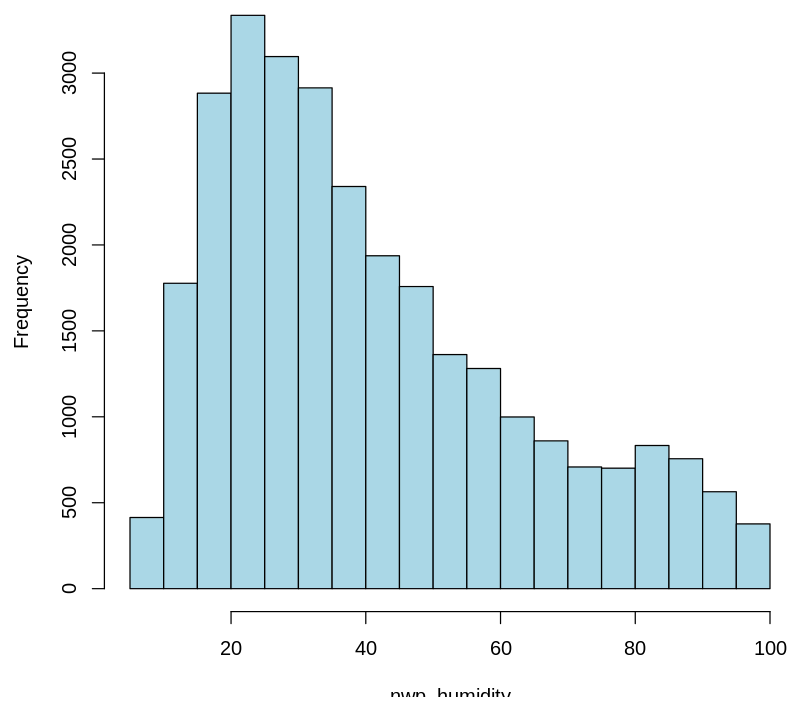
Boxplot of nwp\_globalirrad



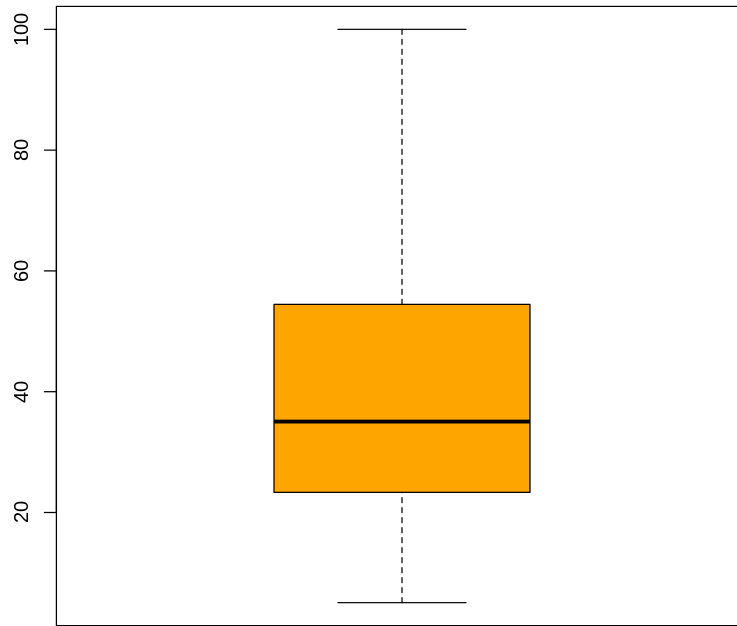
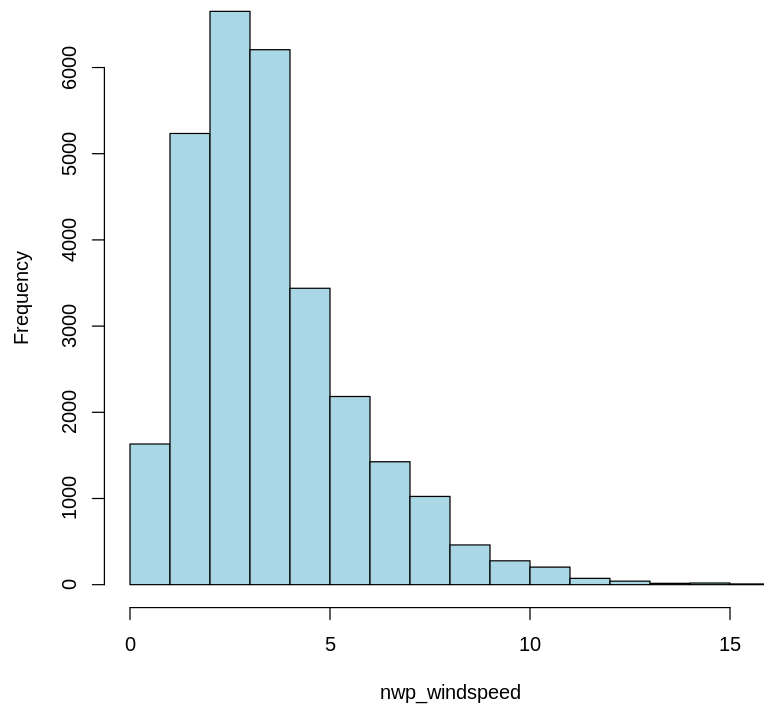
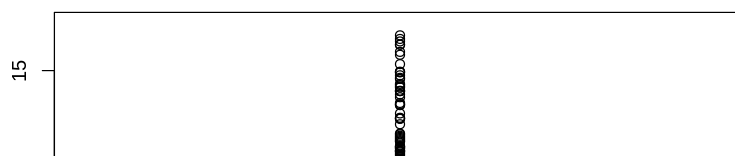
Histogram of nwp\_directirrad

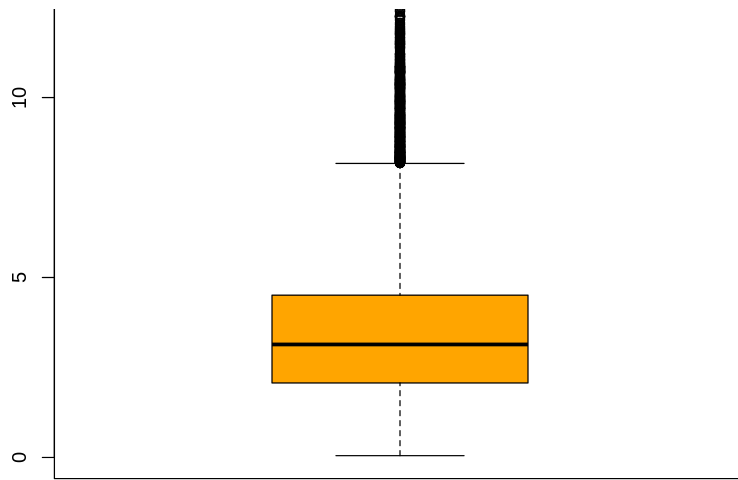
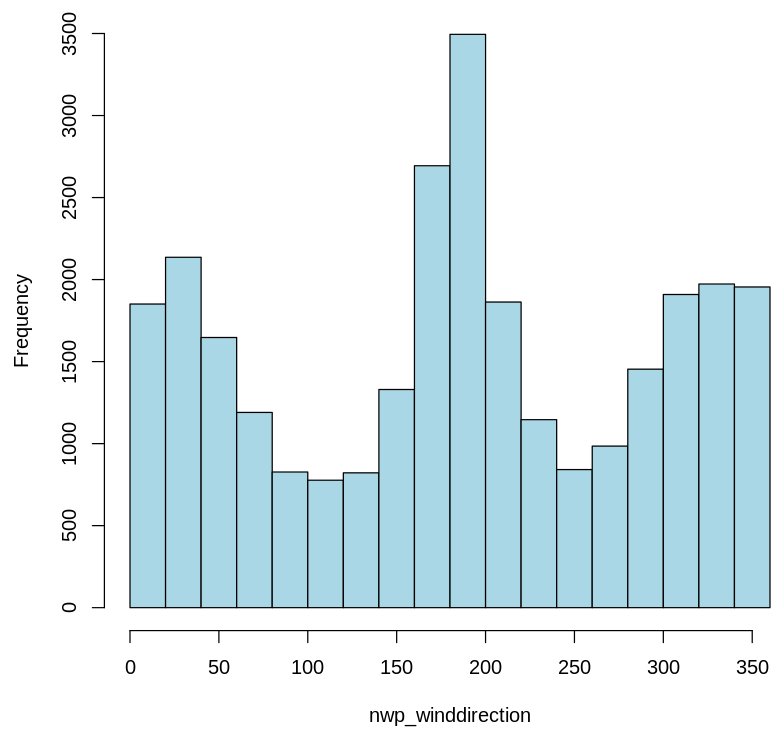
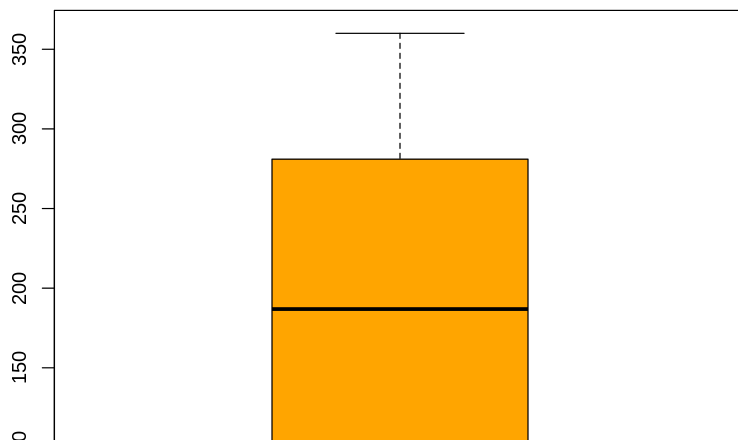


**Boxplot of nwp\_directirrad****Histogram of nwp\_temperature**

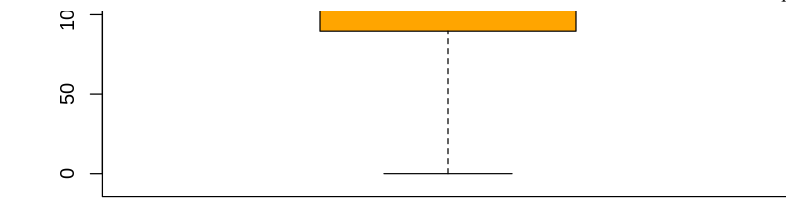
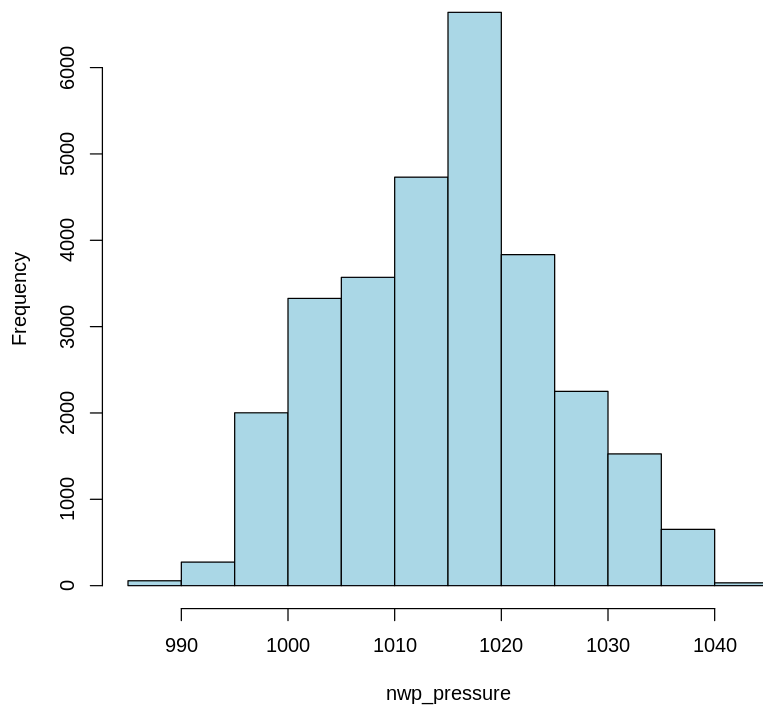
**Boxplot of nwp\_temperature****Histogram of nwp\_humidity**

nwp\_humidity

**Boxplot of nwp\_humidity****Histogram of nwp\_windspeed****Boxplot of nwp\_windspeed**

**Histogram of nwp\_winddirection****Boxplot of nwp\_winddirection**



**Histogram of nwp\_pressure****Boxplot of nwp\_pressure**