An Exploratory Data Analysis on Nigeria's Conformance on Frequency Events to NERC's National Grid Code Requirement

Esan Ayodele Benjamin .C

20th, January 2020

This exploratory data analysis (EDA) is the second in the series for EDA's performed on the Energy situation in Nigeria. The dataset used for this analysis was scraped from the National Electricity Regulatory Commissions's (NERC) website.

In this analysis, frequency events on the National grid are shown across three years 2016 - 2018. According to NERC, a **normal-frequency** event should be between (50 +/- 0.25Hz), hence any frequency event greater than 50.25Hz is considered as an **over-frequency event** and frequency events less than 49.75Hz are considered **under-frequency event**.

Note:

- Under-frequency events occur when the available energy sent out by the electric utility is not sufficient to cater for the load requirements of consumers at any particular time.
- Over-frequeny events occur when the available energy sent out by the electric utility is more than the load requirements of consumers at any particular time.

There has been progressive improvements over the last three years in the number of under-frequency events. The analysis belows show these trends.

Step 1: Setting up preliminary workspaces.

```
setwd("C:/Users/Isioma/Desktop/New Data Science Project") # Sets the current wo
rking directory

rm(list = ls(all.names = TRUE)) #Removes all previously defined objects in work
space.

library(tidyverse) # Loads the tidyverse package including dplyr, ggplot2, tidy
r, etc.

library(lubridate) # Loads the lubridate package for dealing with date-time obj
ects

library(stringr) # Loads the stringr package for dealing with strings in R

library(scales) # Loads the scales package.

library(RColorBrewer) # Loads the color brewer pallete.

theme_set(theme_classic()) #Sets the general theme for plots with ggplot2.
```

Step 2: Importing and cleaning the data-set.

```
freq_data <- read_csv("Daily Frequency Fluctuations.csv")
freq_data <- as.data.frame(freq_data)
View(freq_data)

# Transforming the Date column into date format in R
freq_data$Date <- dmy(freq_data$Date)

# Renaming column names and extracting the year, month,day of week.
freq_data <- freq_data %>%
              rename("upper_freq" = "Upper Frequency (Hz)", "lower_freq" = "Low
er Frequency (Hz)") %>%
              mutate(Year = year(Date), Month = month(Date, label = TRUE), Day
 = wday(Date, label = TRUE))
```

Step 3: Feature engineering.

```
# Only selects from year 2016 - 2018.
freqdata <- freq_data %>%
              filter(Year %in% c(2016,2017,2018))

# Checks frequency values against grid code standard of 50Hz +/- 0.25
freqdata <- freqdata %>%
          select(-c(Date, lower_freq)) %>%
          mutate(
            status = case_when(
                        upper_freq > 50.25 ~ "Over frequency event",
                        upper_freq < 49.75 ~ "Under frequency event",
                        TRUE ~ "Normal frequency event"
            ),
            status = factor(status, levels = c("Normal frequency event", "Under
frequency event", "Over frequency event")),
            deviation = case_when(
                        upper_freq > 50.25 ~ (upper_freq - 50.25),
                        upper_freq < 49.75 ~ (49.75 - upper_freq),
                        TRUE ~ (upper_freq - 50)
            )
          )
```

Analysis I: Understanding the proportion of frequency events across the three years (2016 - 2018) considered.

```
# Frequency events across years 2016 - 2018
freqdata %>%
      count(status, sort = T)%>%
      mutate(proportion = n/sum(n))
```

```
## # A tibble: 3 x 3
##   status                     n proportion
##   <fct>                  <int>      <dbl>
## 1 Under frequency event    648    0.591
## 2 Normal frequency event   446    0.407
## 3 Over frequency event       2    0.00182
```

```
# Frequency events in the year 2016
freqdata %>%
      filter(Year == 2016)%>%
      count(status, sort = T)%>%
      mutate(proportion = n/sum(n))
```

```
## # A tibble: 2 x 3
##   status                     n proportion
##   <fct>                  <int>      <dbl>
## 1 Under frequency event    365    0.997
## 2 Normal frequency event     1    0.00273
```

```
# Frequency events in the year 2017
freqdata %>%
      filter(Year == 2017)%>%
      count(status, sort = T)%>%
      mutate(proportion = n/sum(n))
```

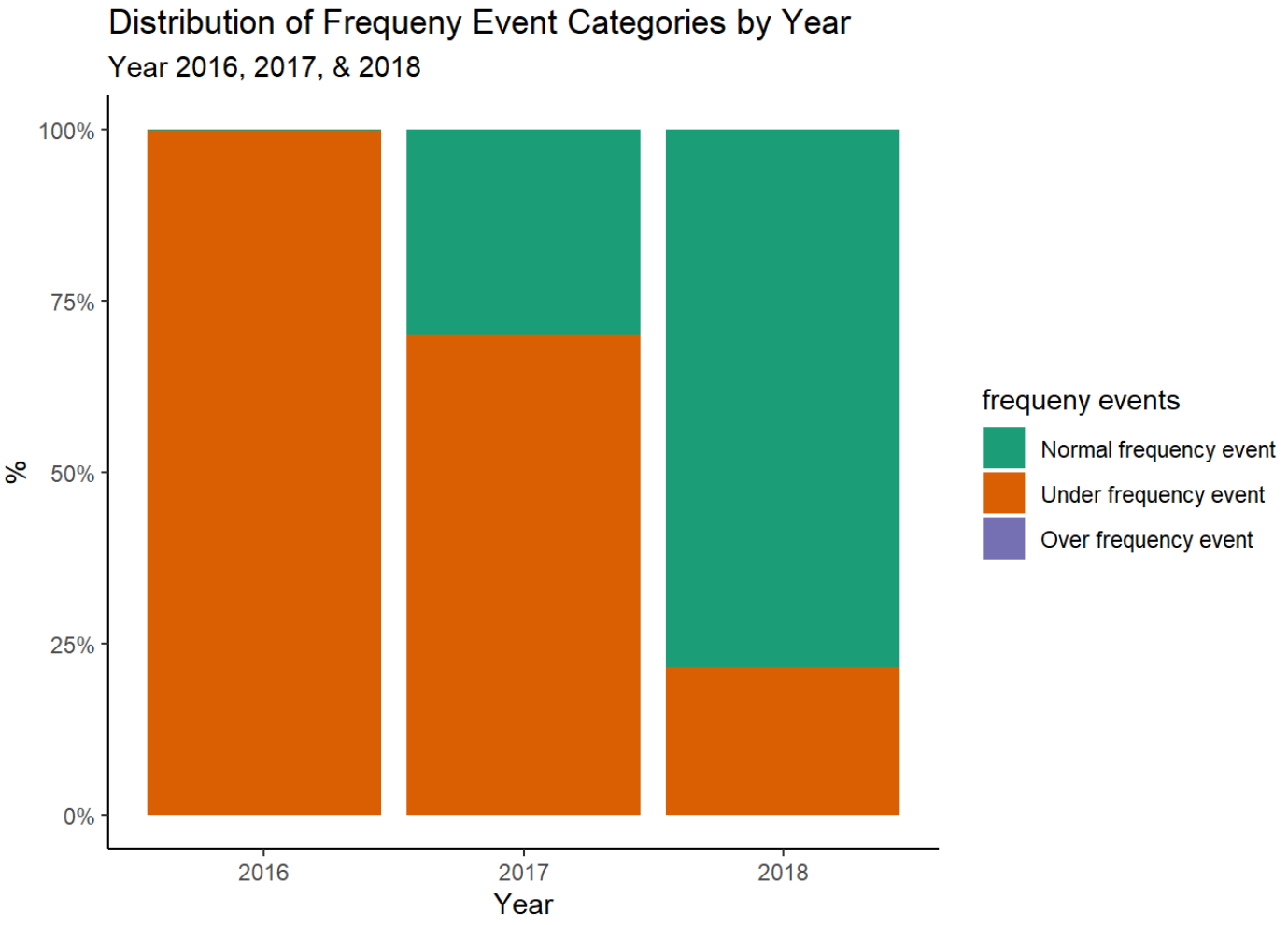
```
## # A tibble: 2 x 3
##   status                     n proportion
##   <fct>                  <int>      <dbl>
## 1 Under frequency event    225      0.616
## 2 Normal frequency event   140      0.384
```

```
# Frequency events in year 2018
freqdata %>%
      filter(Year == 2018)%>%
      count(status, sort = T)%>%
      mutate(proportion = n/sum(n))
```

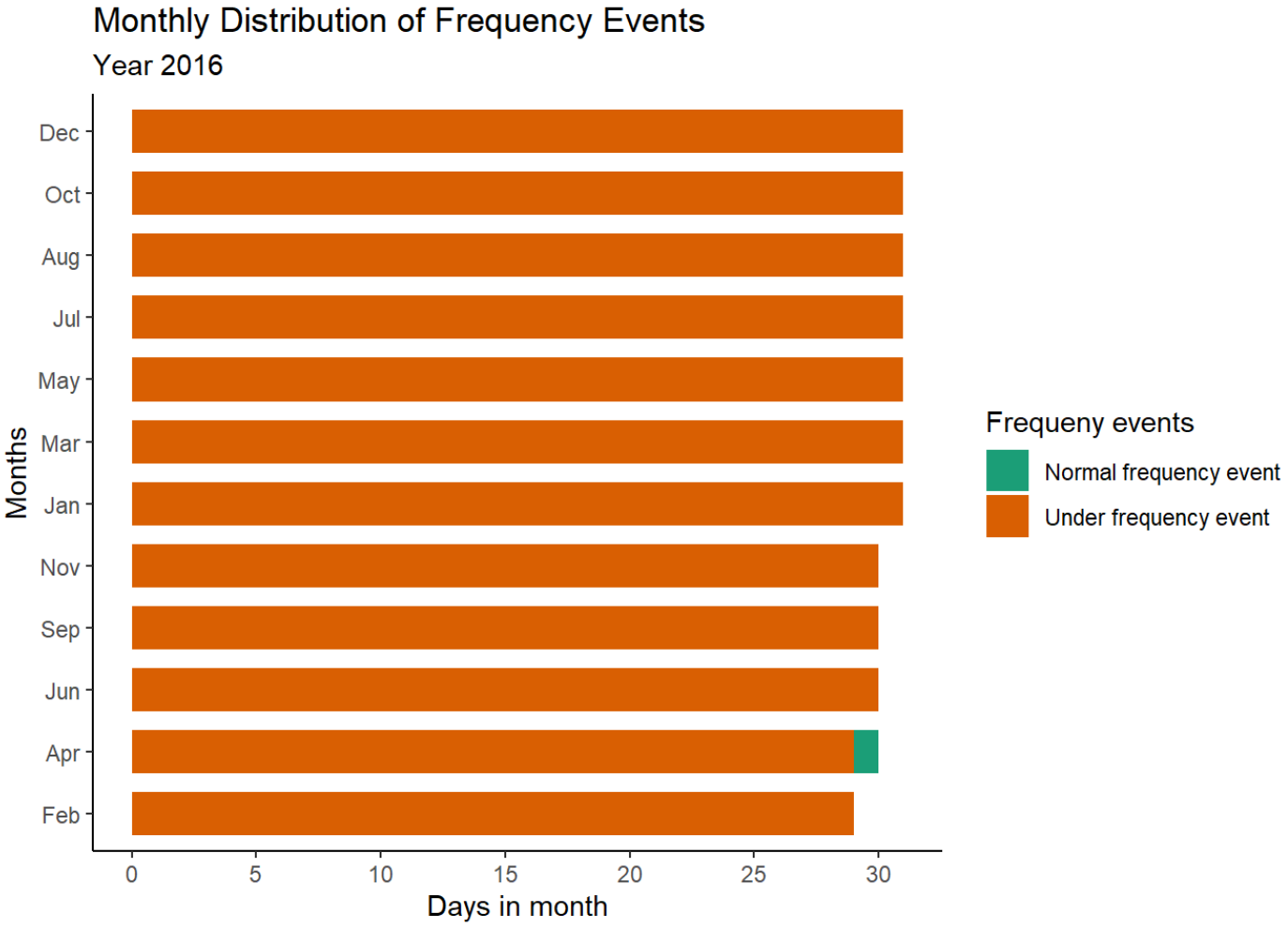
```
## # A tibble: 3 x 3
##   status                     n proportion
##   <fct>                  <int>      <dbl>
## 1 Normal frequency event   305    0.836
## 2 Under frequency event     58    0.159
## 3 Over frequency event       2    0.00548
```

Analysis II: How are the frequency event categories distributed across the three years collectively and Individually?

```
# Stacked bar plot showing frequency events across 2016 - 2018.
freqdata %>%
      mutate(Year = as.factor(Year))%>%
      add_count(status)%>%
      ggplot(aes(Year, n, fill = status))+
      geom_bar(position = "fill", stat = "identity")+
      labs(y = "%", title = "Distribution of Frequeny Event Categories by Year"
, subtitle = "Year 2016, 2017, & 2018", fill = "frequeny events")+
      scale_y_continuous(labels = percent_format())+
      scale_fill_brewer(palette = "Dark2")+
      ggsave("Image1.png", width = 8, height = 6)
```

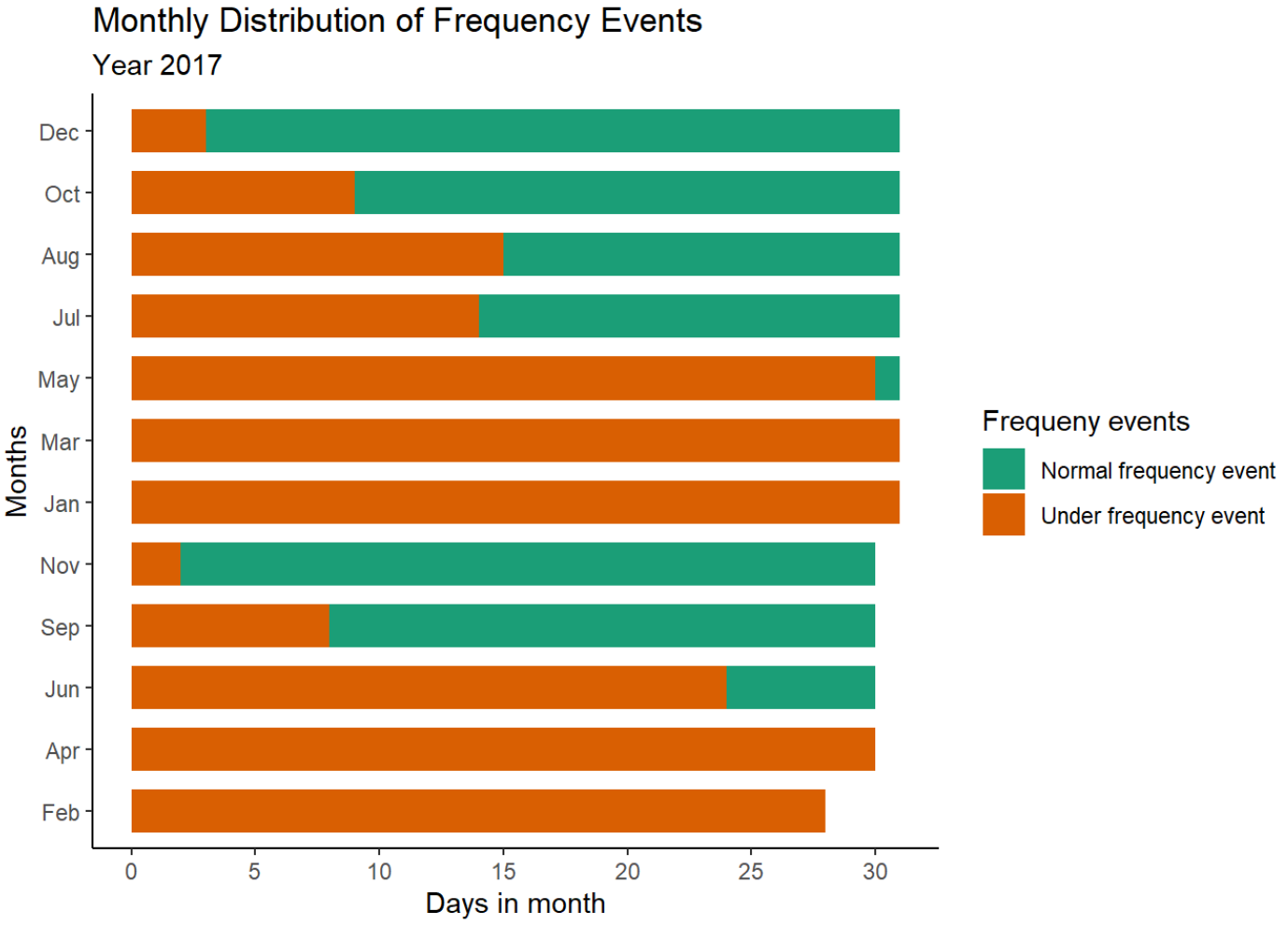


```
# Column plot showing frequency events for 2016.
freqdata %>%
      filter(Year == 2016) %>%
      mutate(Month = as.factor(Month)) %>%
      group_by(Month) %>%
      count(status) %>%
      ggplot(aes(fct_reorder(Month, n, sum), n, fill = status)) +
      geom_col(width = 0.7)+
      labs(x = "Months", y = "Days in month", title = "Monthly Distribution of
 Frequency Events", subtitle = "Year 2016", fill = "Frequeny events")+
      scale_y_continuous(breaks = c(0, 5, 10, 15, 20, 25, 30))+
      coord_flip()+
      scale_fill_brewer(palette = "Dark2")+
      ggsave("Image2.png", width = 8, height = 6)
```



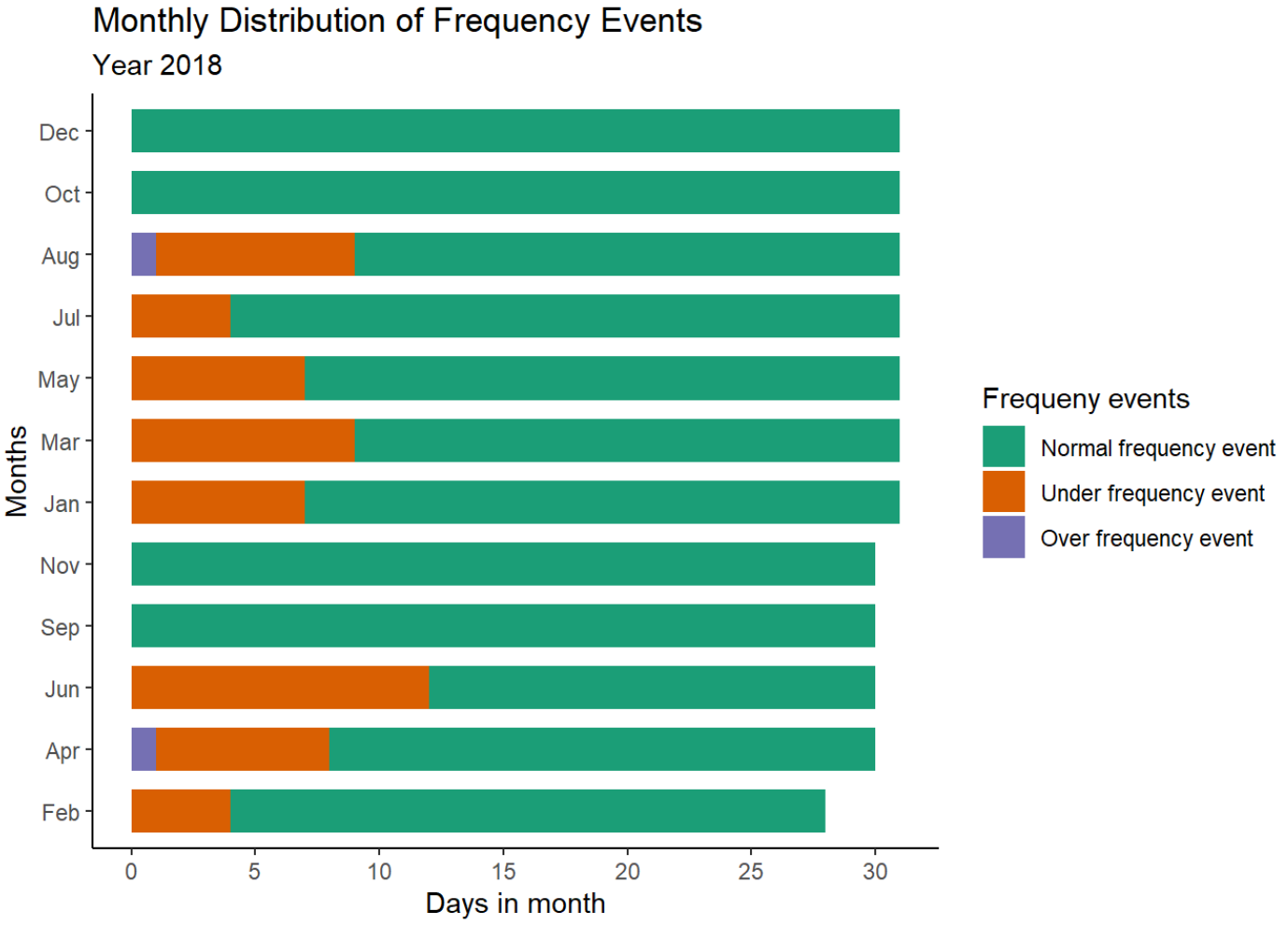
- All months in 2016, with the exception of maybe 1 or 2 days in the month of April had under-frequency events.

```
# Column plot showing frequency events for 2017.
freqdata %>%
      filter(Year == 2017) %>%
      mutate(Month = as.factor(Month)) %>%
      group_by(Month) %>%
      count(status) %>%
      ggplot(aes(fct_reorder(Month, n, sum), n, fill = status)) +
      geom_col(width = 0.7)+
      labs(x = "Months", y = "Days in month", title = "Monthly Distribution of
 Frequency Events", subtitle = "Year 2017", fill = "Frequeny events")+
      scale_y_continuous(breaks = c(0, 5, 10, 15, 20, 25, 30))+
      coord_flip()+
      scale_fill_brewer(palette = "Dark2")+
      ggsave("Image3.png", width = 8, height = 6)
```



- The first six months in the the year 2017 had the most under-frequency events (January till June), but the second half of the year (July - December) saw some drastic reductions in under-frequency events, with normal frequency events peaking from July (representing about 55% percent of frequency events for the month) all the way till December(representing about 97% percent of frequency events for the month).

```
# Column plot showing frequency events for 2018.
freqdata %>%
      filter(Year == 2018) %>%
      mutate(Month = as.factor(Month)) %>%
      group_by(Month) %>%
      count(status) %>%
      ggplot(aes(fct_reorder(Month, n, sum), n, fill = status)) +
      geom_col(width = 0.7)+
      labs(x = "Months", y = "Days in month", title = "Monthly Distribution of
 Frequency Events", subtitle = "Year 2018", fill = "Frequeny events")+
      scale_y_continuous(breaks = c(0, 5, 10, 15, 20, 25, 30))+
      coord_flip()+
      scale_fill_brewer(palette = "Dark2")+
      ggsave("Image4.png", width = 8, height = 6)
```



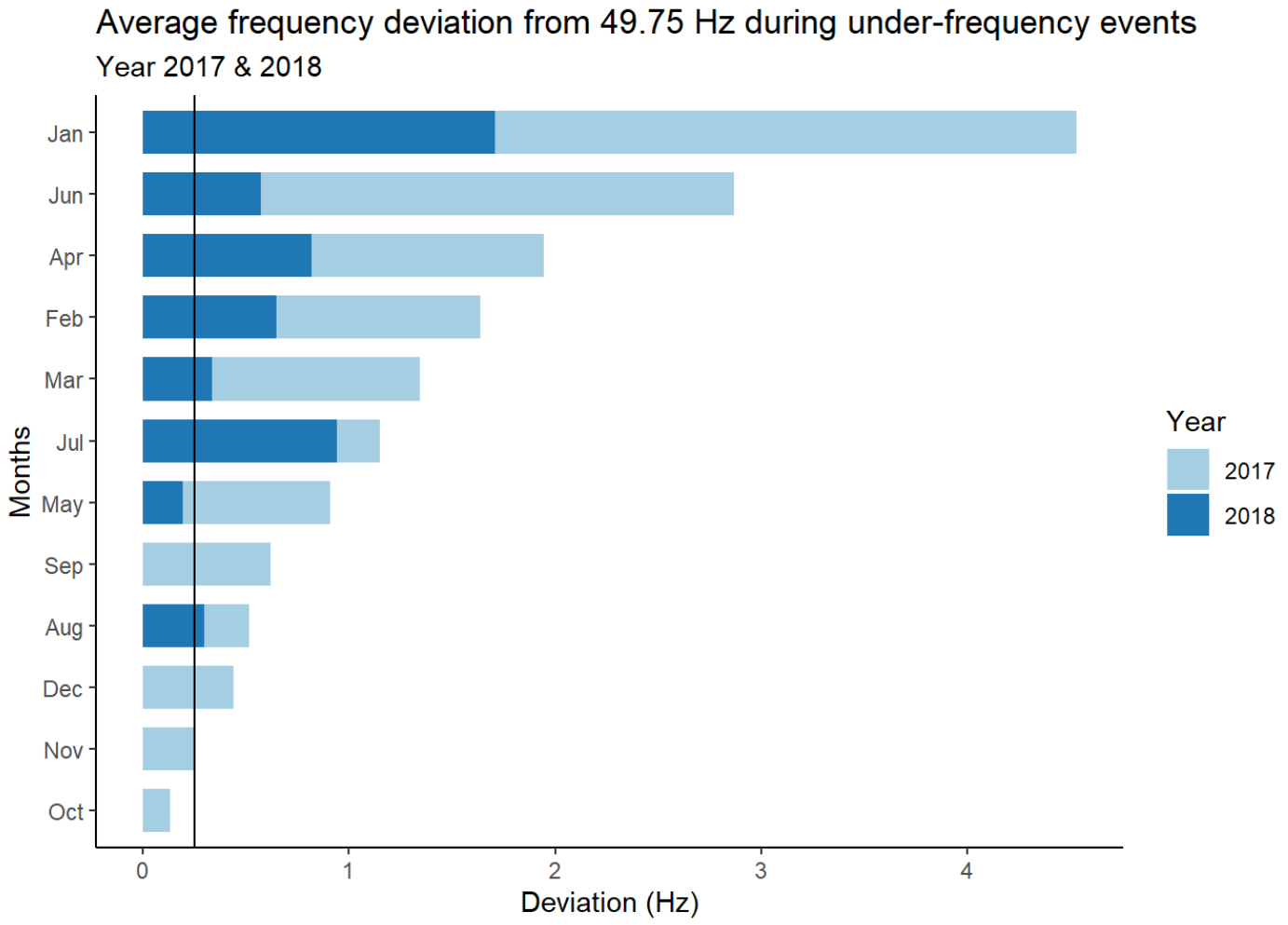
- The year 2018 saw dramatic improvements in frequency events, with most months having normal frequency events well above 50%. The first 8 months (January till August) recorded some under-frequency events (June experiencing the most, about 30% of frequency events). Still yet, two months (April and August) recorded some over frequency event though minimal.

- As seen from figures 2, 3, and 4, over-frequency events represents a very minuscle portion of frequency events between 2016 and 2018. Hence, in our further analysis, we specifically look at just under-frequency events.

Analysis III: During under-frequency events alone, how does the average deviation from 49.75Hz vary across months in 2017 and 2018?

- As seen from previous plots, it's apparent that in 2016, under-frequency events were common occurence virtually all year round. Thus, we look at under-frequency events in 2017 and 2018 and try to understand their distribution across months based on their deviations from the standard NERC frequency grid code requirement.

```
freqdata %>%
      filter(status == "Under frequency event" & Year != 2016) %>%
      mutate(Year = as.factor(Year), Month = as.factor(Month)) %>%
      group_by(Year, Month) %>%
      summarize(average_deviation = mean(deviation)) %>%
      ungroup()%>%
      ggplot(aes(fct_reorder(Month, average_deviation, sum), average_deviation,
fill = Year))+
      geom_col(width = 0.7)+
      geom_hline(yintercept = 0.25)+
      coord_flip()+
      labs(y = "Deviation (Hz)", x = "Months", title = "Average frequency devia
tion from 49.75 Hz during under-frequency events", subtitle = "Year 2017 & 201
8")+
      scale_fill_brewer(palette = "Paired")+
      ggsave("Image5.png", width = 8, height = 6)
```



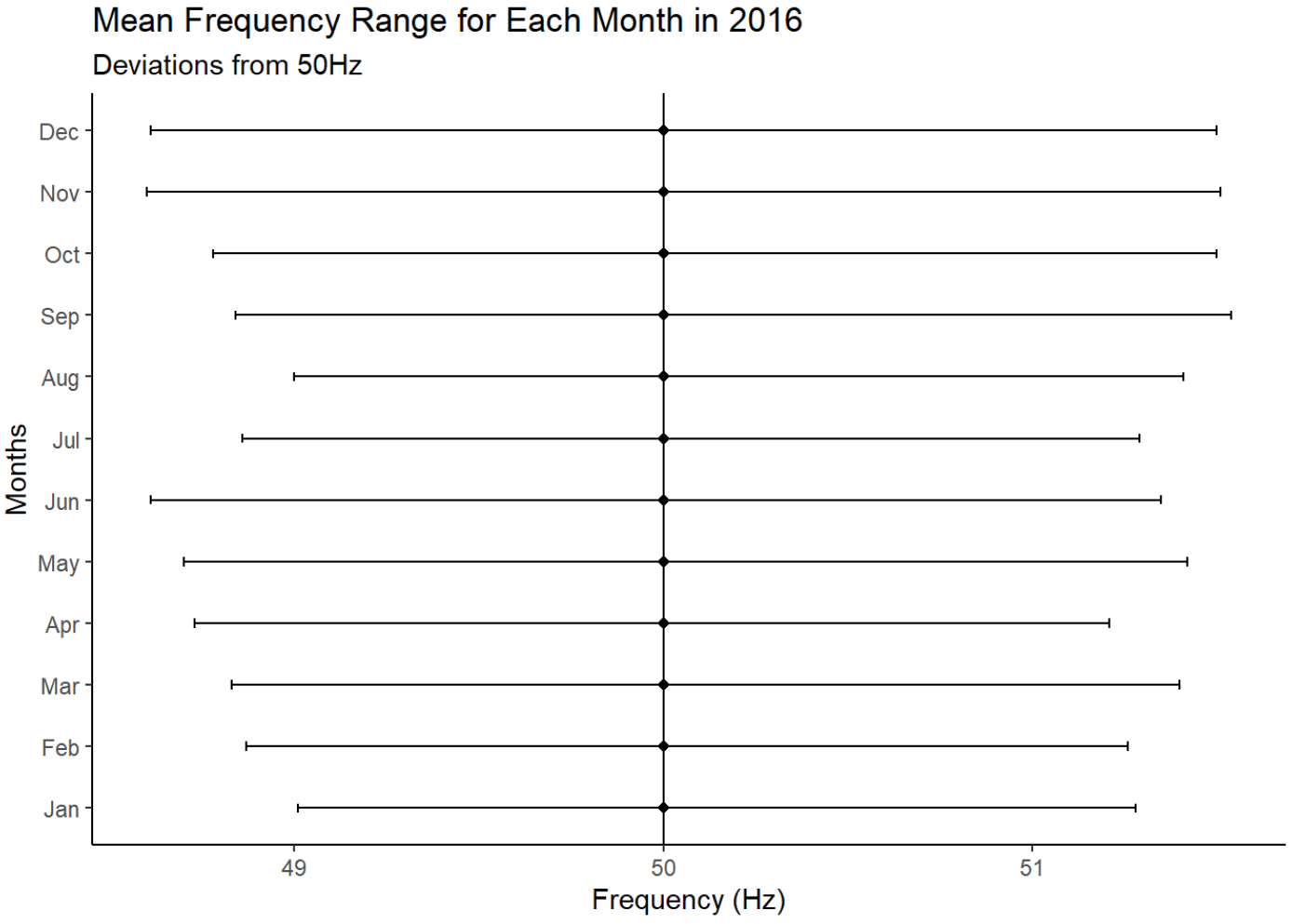
Note: The deviation feature of under-frequency events represent shifts or deviance from the grid code i.e. 49.75Hz @ the lower frequency rung.

- The vertical line in figure 5 represents the threshold deviation of 0.25 as set by NERC. It is seen that in 2017, only October and November experienced normal frequency events. Meanwhile, in 2018, only the month of May recorded normal frequency events within the grid frequency margin.

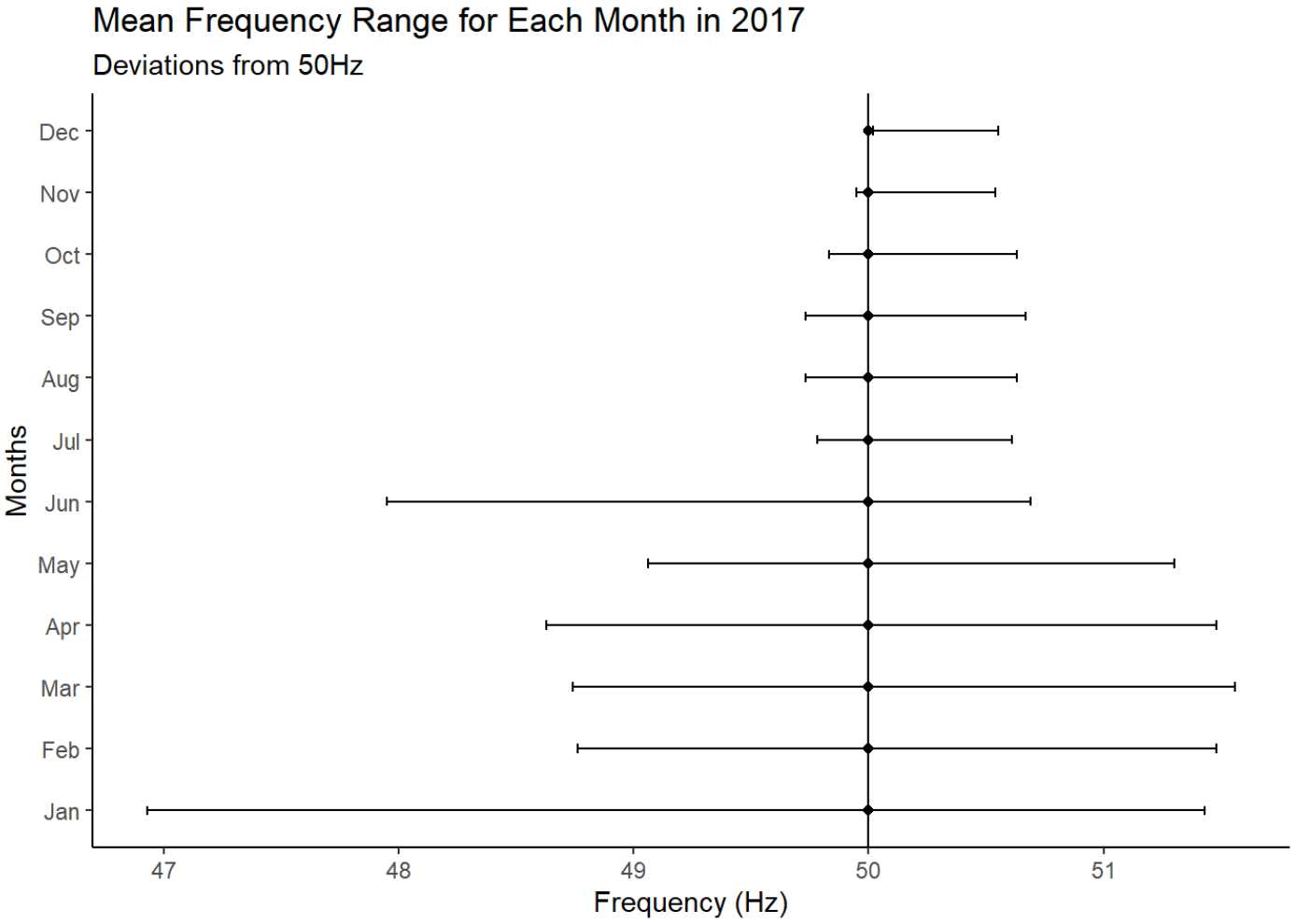
Analysis IV: How does the daily frequency records by NERC between 2016 and 2018 vary about the ideal frequency (50Hz) adopted in Nigeria?

Considering the daily lower and upper frequency records as obtained from NERC between 2016 and 2018, figures 6 to 8, show an error_graph depicting the monthly spread of the grid's frequency from the NERC's requirements of (50 +/- 0.25 Hz).

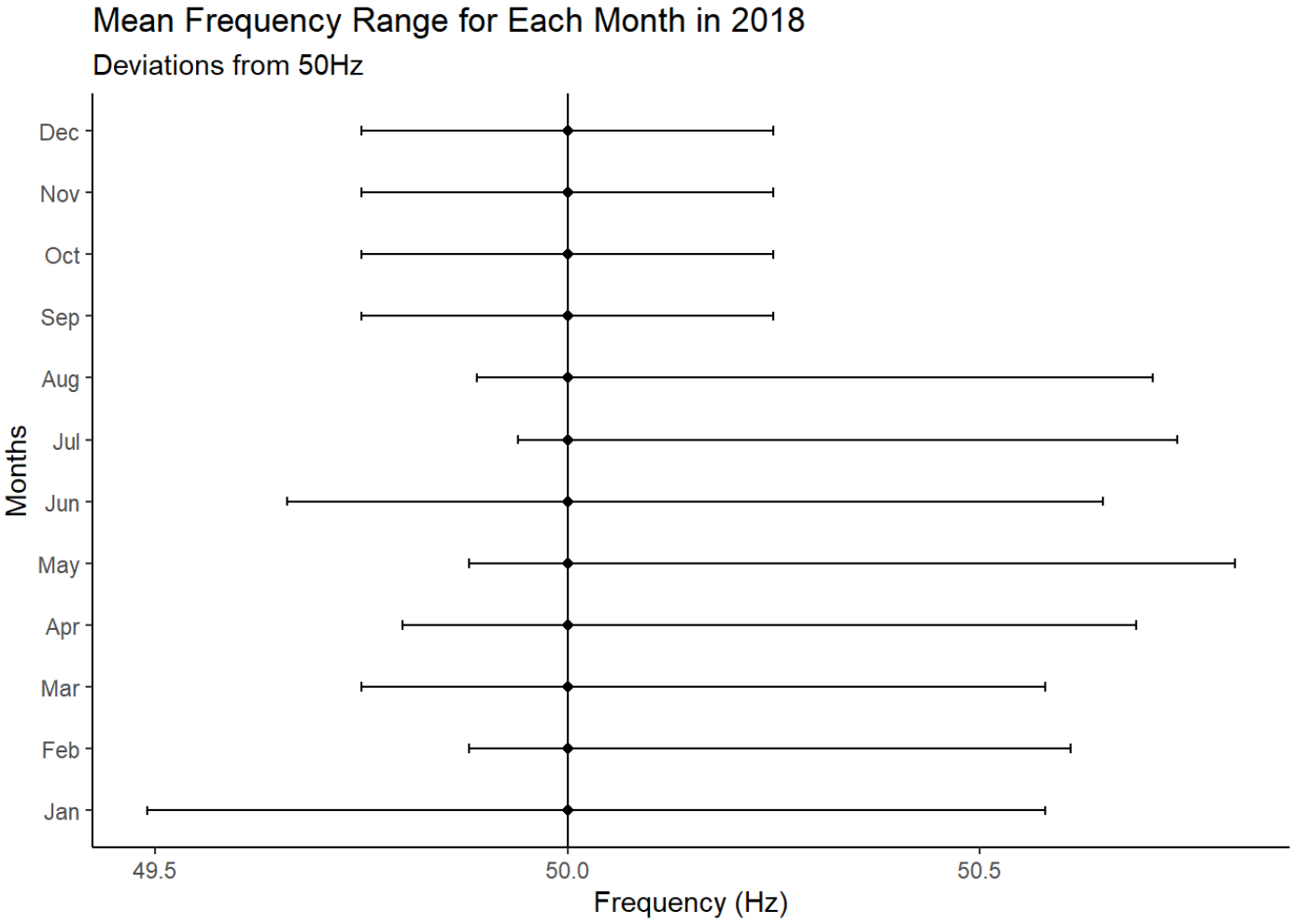
```
freq_data %>%
   filter(Year == 2016) %>%
   group_by(Month) %>%
   summarize(upperFreq = round(mean(upper_freq),2), lowerFreq = round(mean(lowe
r_freq),2), normalFreq = 50)%>%
   ggplot(aes(x = Month, y = normalFreq))+
   geom_point()+
   geom_errorbar(aes(ymin = lowerFreq, ymax = upperFreq), width = 0.15)+
   geom_hline(yintercept = 50)+
   coord_flip()+
   labs(y = "Frequency (Hz)", x = "Months", title = "Mean Frequency Range for E
ach Month in 2016", subtitle = "Deviations from 50Hz") +
   theme(legend.position = "none")+
   ggsave("Image6.png", width = 8, height = 6)
```



```
freq_data %>%
   filter(Year == 2017) %>%
   group_by(Month) %>%
   summarize(upperFreq = round(mean(upper_freq),2), lowerFreq = round(mean(lowe
r_freq),2), normalFreq = 50)%>%
   ggplot(aes(x = Month, y = normalFreq))+
   geom_point()+
   geom_errorbar(aes(ymin = lowerFreq, ymax = upperFreq), width = 0.15)+
   geom_hline(yintercept = 50)+
   coord_flip()+
   labs(y = "Frequency (Hz)", x = "Months", title = "Mean Frequency Range for E
ach Month in 2017", subtitle = "Deviations from 50Hz") +
   theme(legend.position = "none")+
      ggsave("Image7.png", width = 8, height = 6)
```



```
freq_data %>%
   filter(Year == 2018) %>%
   group_by(Month) %>%
   summarize(upperFreq = round(mean(upper_freq),2), lowerFreq = round(mean(lowe
r_freq),2), normalFreq = 50)%>%
   ggplot(aes(x = Month, y = normalFreq))+
   geom_point()+
   geom_errorbar(aes(ymin = lowerFreq, ymax = upperFreq), width = 0.15)+
   geom_hline(yintercept = 50)+
   coord_flip()+
   labs(y = "Frequency (Hz)", x = "Months", title = "Mean Frequency Range for E
ach Month in 2018", subtitle = "Deviations from 50Hz") +
   theme(legend.position = "none")+
      ggsave("Image8.png", width = 8, height = 6)
```



Conclusion.

Looking at the frequency events for the three years considered in this EDA, the trend shows a gradual decline in under-frequency events between 2016 - 2018 and a gradual increase in normal frequency events.

- Between 2016 and 2017, under-frequency events dropped by approximately **38%**, while normal-frequency events increased **way over 100%**.
- Between 2017 and 2018, under-frequency events dropped by approximately **74%**, while normal frequency events increased by approximately **118%**.
- In summary, while Nigeria still has a long way in upgrading her Generation and Transmission capacity to cater for her teeming population, she has made considerable progress in curtailing the number of under-frequency events experienced between 2016 and 2018. As a result of this, the **quality** of electricity supply has improved in recent years even-though the **quantity** is still grossly inadequate.