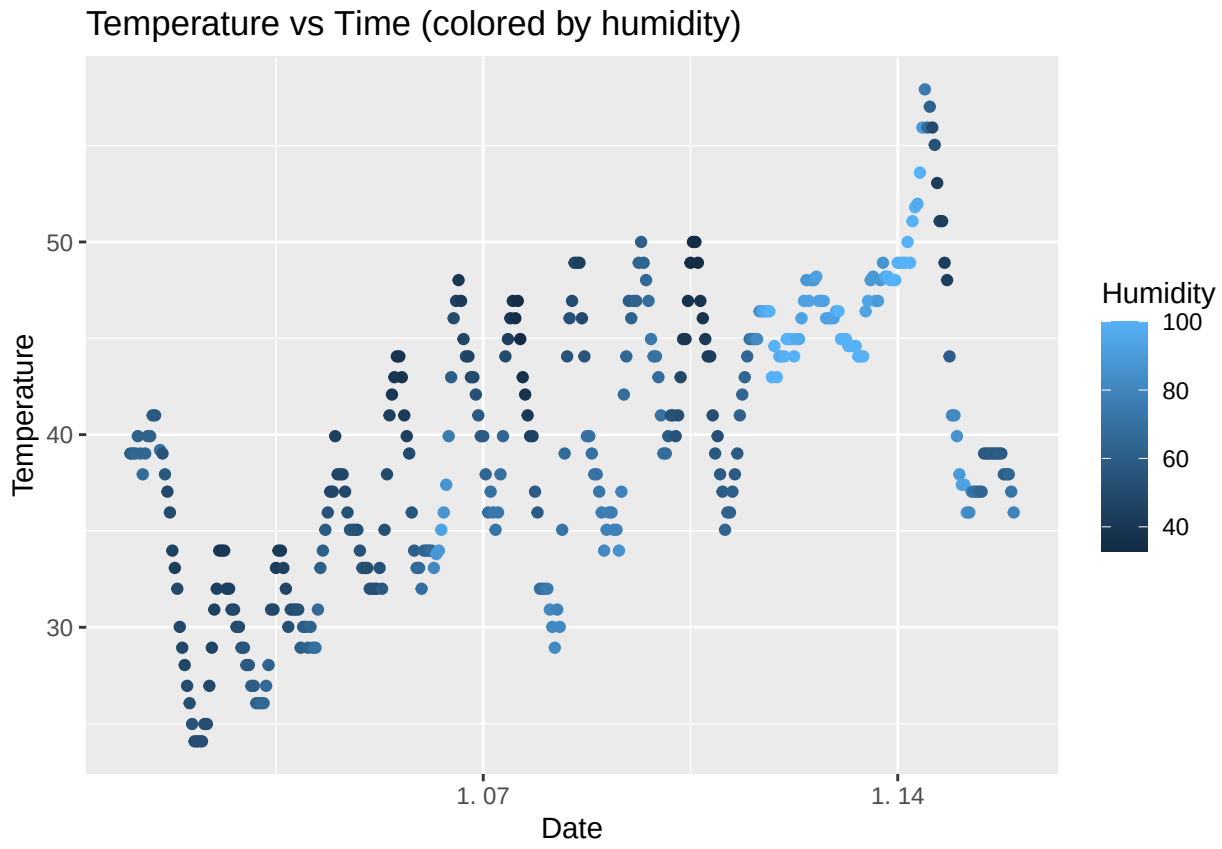
p8105_hw1_yg3096

Problem 1

The data has 358 rows and 15 columns.
Variables names are: origin, year, month, day, hour, temp, dewp, humid, wind_dir, wind_speed, wind_gust, precip, pressure, visib, time_hour.
And the average temperature is 39.58.

```
# Plot temperature vs time, using color of points to show humidity
p <- ggplot(early_january_weather, aes(x = time_hour, y = temp,
                                       color = humid)) +
     geom_point() + labs(x = "Date", y = "Temperature", color = "Humidity",
                         title = "Temperature vs Time (colored by humidity)")
print(p)
```




```
# Save the plot
ggsave(filename = "plot.png", plot = p)
```

```
## Saving 6.5 x 4.5 in image
```

The plot shows that the temperature has a overall trend of increasing. Also, the change of temperature in a single day has a periodic pattern.

Problem 2

```r
set.seed(123)

# A random sample of size 10 from a standard normal distribution
norm_sample <- rnorm(10)

# A logical vector indicating whether elements of the sample are greater than 0
logical_vec <- norm_sample > 0

# A character vector of length 10
char_vec <- c("a","b","c","d","e","f","g","h","i","j")

# A factor vector of length 10 with 3 different factor levels
factor_vec <- factor(sample(c("low","medium","high"), 10, replace = TRUE),
                     levels = c("low","medium","high"))

df <- data.frame(norm_sample = norm_sample, logical_vec = logical_vec,
                 char_vec = char_vec, factor_vec = factor_vec)

mean_norm <- mean(df %>% pull(norm_sample)) # working, returns 0.07462564
mean_logical <- mean(df %>% pull(logical_vec)) # working, returns 0.5
# Not working: mean_char and mean_factor

# Logical to numeric
as.numeric(df$logical_vec) # TRUE -> 1, FALSE -> 0
```

```
##  [1] 0 0 1 1 1 1 1 0 0 0
```

```r
# Factor to numeric
as.numeric(df$factor_vec) # returns (1,2,3), not the specific level
```

```
##  [1] 1 1 1 1 3 2 3 2 1 2
```

```r
# character to numeric (only if the character can be transfred into numbers)
# as.numeric(df$char_vec) works if char_vec contains "1", "2"...
```