Worksheet 03 - Intro to R programming - NCBS MSc WL (Answers)

Akshay Surendra, Anand M Osuri

25 September 2020

Problem set (graded)

Download the file *Valparai_weather_data_2015_summary.csv* to your working directory and import it as a tibble into your RStudio environment (remember to use `read_csv`). This file contains daily data from a weather station in rainforests of the Western Ghats. Columns include `Date`, `Month`, `Rain` (daily rainfall), `Temp_mean`, `Temp_min`, `Temp_max` (daily mean, minimum and maximum temperature, respectively), and `RH_mean` (mean relative humidity).

1. Plot a histogram of mean temperature (column `Temp_mean`), specify that the histogram must contain *exactly 10 bins*, rename the x-axis label as *"Mean daily temperature (deg C)"*, and save it to an object named `Plot_1`. Examine the plot and as a comment in your R code, make a rough estimate of the most frequent value of mean daily temperature in this dataset.

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.2     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
dat <- read_csv('Valparai_weather_data_2015_summary.csv')
```

```
## Parsed with column specification:
## cols(
##   Date = col_date(format = ""),
##   Month = col_double(),
##   Rain = col_double(),
##   Temp_min = col_double(),
##   Temp_mean = col_double(),
##   Temp_max = col_double(),
##   RH_mean = col_double()
## )
```

```
#Plot 1
Plot_1 <-
  ggplot(data=dat) +
  geom_histogram(mapping = aes(x = Temp_mean),
                 bins = 10) +
  labs(x = "Mean daily temperature (deg C)")
# 20-21 deg C
```

2. Plot a line graph showing the accumulation of rainfall by month from January to December. This is called a plot of **cumulative rainfall**. The graph would have the sequence of months on the X-axis. The Y-axis value for each month would be the total amount of rainfall that has occurred from the beginning of January until the end of that month. Rename the Y-axis label as *"Cumulative rainfall (mm)"*. Save this graph as object `Plot_2`. As a comment, estimate roughly how much rain the site received over the calendar year

Additional question: Are you happy with the X-axis of `Plot_2` graph, or would you improve it in some way? [hint: to perform this task you may need to transform the underlying data. First, use the `summarize()` function of `dplyr` to calculate total rainfall per month. Next, use the function `cumsum()` (cumulative sum) to calculate cumulative rainfall]

```
# Plot 2
Cum_table <- dat %>% group_by(Month) %>%
  summarise(Tot_Rain = sum(Rain,na.rm = T)) %>%
  mutate(Cum_rain = cumsum(Tot_Rain))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
Plot_2 <- ggplot(data = Cum_table) +
  geom_line(mapping = aes(x = Month,y = Cum_rain)) +
  labs(y="Cumulative rainfall (mm)") +
  scale_x_continuous(breaks = 1:12)
```

3. Make a scatterplot with minimum temperature (`Temp_min`) on the X-axis and mean relative humidity on the Y-axis. Rename the X- and Y-axes as *"Minimum temperature (deg C)"* and *"Relative humidity (%)"*, respectively. Colour code the graph to depict each month as a different colour. Save this graph as object `Plot_3`. As a comment, what relationship do you observe between the two variables?

```
# Plot 3
Plot_3 <- ggplot(data = dat)+
  geom_point(mapping = aes(x = Temp_min,y = RH_mean,color = as_factor(Month))) +
  labs(x = "Minimum temperature (deg C)", y = "Relative humidity (%)")
```

4. Redraw the above graph using facets, with separate facets for each month. Rename the X- and Y-axes as above. Call it `Plot_4`.

```
#Plot 4
Plot_4 <- ggplot(data = dat) +
  geom_point(mapping = aes(x = Temp_min,y = RH_mean)) +
  facet_wrap(~ Month) +
  labs(x = "Minimum temperature (deg C)",y = "Relative humidity (%)")
```

5. Draw a bar graph of mean monthly minimum and mean monthly maximum temperature. The graph should contain 12 pairs of bars corresponding to the 12 months. Each pair of bars should represent mean minimum and mean maximum temperatures for a given month, respectively. Save this graph to the object `Plot_6`. [hint: use `summarize()` to calculate monthly means for minimum and maximum temperature. Then use `pivot_longer()` to bring the data into the correct format for the graph]

```
# Plot 5
dat_long <-
  dat %>%
  select(Month,Temp_min,Temp_max) %>%
  pivot_longer(cols = c(Temp_min,Temp_max),
               names_to = "Type",
               values_to = "Temperature")

Plot_5 <-
  ggplot(data = dat_long) +
  geom_bar(mapping = aes(x = as_factor(Month),y = Temperature,fill = Type),
           stat = 'identity', position = 'dodge')
```