Class Activity 15

Your name here

April 27 2023

Group Activity 1

a. Write a for loop to iterate over the columns of the 'energy' dataset and print the names of all columns containing the string "House". Please use the function `colnames()` to extract the column names.

Answer:

```
for (col_name in colnames(energy)) {
  if (str_detect(col_name, "House")) {
    print(col_name)
  }
}
```

```
[1] "Allen_House"
[1] "Alumni_Guest_House/Johnson_House"
[1] "Benton_House"
[1] "Berg_House"
[1] "Bird_House"
[1] "Chaney_House"
[1] "Clader_House"
[1] "Dacie_Moses_House"
[1] "Douglas_House"
[1] "Farm_House"
[1] "Geffert_House"
[1] "Headley_House"
[1] "Henrickson_House"
[1] "Henry_House"
[1] "Hill_House"
[1] "Hilton_House"
[1] "Hoppin_House_(Alumni)"
[1] "Huntington_House"
[1] "Jewett_House"
[1] "Jones_House"
[1] "Nutting_House"
```

```
[1] "Page_House_West"
[1] "Parish_House_"
[1] "Parr_House"
[1] "Pollock_House"
[1] "Prentice_House"
[1] "Rayment_House"
[1] "Rice_House"
[1] "Rogers_House"
[1] "Ryberg_House"
[1] "Seccombe_House"
[1] "Sperry_House"
[1] "Stimson_House"
[1] "Strong_House"
[1] "Whittier_House"
[1] "Wilson_House"
```

b. Using a for loop, calculate and print the mean of the first 8 columns of the 'energy' dataset, excluding the 'Timestamp' and 'dayWeek' column.

```
for (i in c(2:5,7:8)) {
  col_mean <- mean(energy[[i]], na.rm = TRUE)
  cat("Mean of", colnames(energy)[i], ":", col_mean, "\n")
}
Mean of year : 2015.667
Mean of month : 6.513678
Mean of weekOfYear : 26.59575
Mean of dayOfMonth : 15.75704
Mean of timeHour : 11.5013
Mean of timeMinute : 22.49936
```

Group Activity 2

1. Make a data frame of quantiles for `energy` buildings in columns 9-90 (you will need `na.rm = TRUE`)

```
qdf <- energy %>% select(9:90) %>%
  map_dfc(quantile, probs = seq(.1,.9,.1), na.rm = TRUE)
qdf
# A tibble: 9 x 82
  100_Nevada_S~1 104_M~2 106_W~3 Allen~4 Alumn~5 Arbor~6 Art_S~7 Bento~8 Berg_~9
           <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
1         0.0972    1.04   0.601   0.756    17.0    0.13    0.23    1.59    1.06
2         0.120     1.11   0.632   0.781    18.1    0.23    0.28    1.60    1.25
3         0.183     1.18   0.673   0.941    18.4    0.25    0.33    1.70    1.28
4         0.461     1.18   0.681   0.983    20.3    0.28    0.4     1.79    1.31
5         0.710     1.42   0.692   1.00     21.0    0.32    0.47    1.79    1.34
6         0.795     1.42   0.865   1.01     21.8    0.38    0.57    2.10    1.49
7         0.915     1.54   1.10    1.07     21.9    0.44    0.73    2.21    1.56
8         1.11      1.56   1.20    1.07     22      0.52    0.88    2.27    1.57
```

```
9         1.24      1.67   1.27    1.25     22.5    0.71    1.09    2.33    1.58
# ... with 73 more variables: Bird_House <dbl>,
#   Boliou_Memorial_Art_Bldg. <dbl>, Burton_Hall <dbl>,
#   `Cassat_Hall_/_James_Hall` <dbl>,
#   `Center_for_Mathematics_&_Computing` <dbl>, Chaney_House <dbl>,
#   Clader_House <dbl>, College_Warehouse <dbl>, Cowling_Gym <dbl>,
#   Dacie_Moses_House <dbl>, Davis_Hall <dbl>, Douglas_House <dbl>,
#   Evans_Hall <dbl>, `Faculty_Club_/_Annex` <dbl>, Farm_House <dbl>, ...
```

2. Add a variable to identify the quantile

```
qdf <- energy %>% select(9:90) %>%
  map_dfc(quantile, probs = seq(.1,.9,.1), na.rm = TRUE) %>%
  mutate(stat = str_c("quantile_", seq(10,90,10)))
qdf
# A tibble: 9 x 83
  100_Nevada_S~1 104_M~2 106_W~3 Allen~4 Alumn~5 Arbor~6 Art_S~7 Bento~8 Berg_~9
           <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
1         0.0972    1.04   0.601   0.756    17.0    0.13    0.23    1.59    1.06
2         0.120     1.11   0.632   0.781    18.1    0.23    0.28    1.60    1.25
3         0.183     1.18   0.673   0.941    18.4    0.25    0.33    1.70    1.28
4         0.461     1.18   0.681   0.983    20.3    0.28    0.4     1.79    1.31
5         0.710     1.42   0.692   1.00     21.0    0.32    0.47    1.79    1.34
6         0.795     1.42   0.865   1.01     21.8    0.38    0.57    2.10    1.49
7         0.915     1.54   1.10    1.07     21.9    0.44    0.73    2.21    1.56
8         1.11      1.56   1.20    1.07     22      0.52    0.88    2.27    1.57
9         1.24      1.67   1.27    1.25     22.5    0.71    1.09    2.33    1.58
# ... with 74 more variables: Bird_House <dbl>,
#   Boliou_Memorial_Art_Bldg. <dbl>, Burton_Hall <dbl>,
#   `Cassat_Hall_/_James_Hall` <dbl>,
#   `Center_for_Mathematics_&_Computing` <dbl>, Chaney_House <dbl>,
#   Clader_House <dbl>, College_Warehouse <dbl>, Cowling_Gym <dbl>,
#   Dacie_Moses_House <dbl>, Davis_Hall <dbl>, Douglas_House <dbl>,
#   Evans_Hall <dbl>, `Faculty_Club_/_Annex` <dbl>, Farm_House <dbl>, ...
```

3. Reshape the data frame to make variables `stat` (describing the quantile), `building` and `quant` (quantile value)

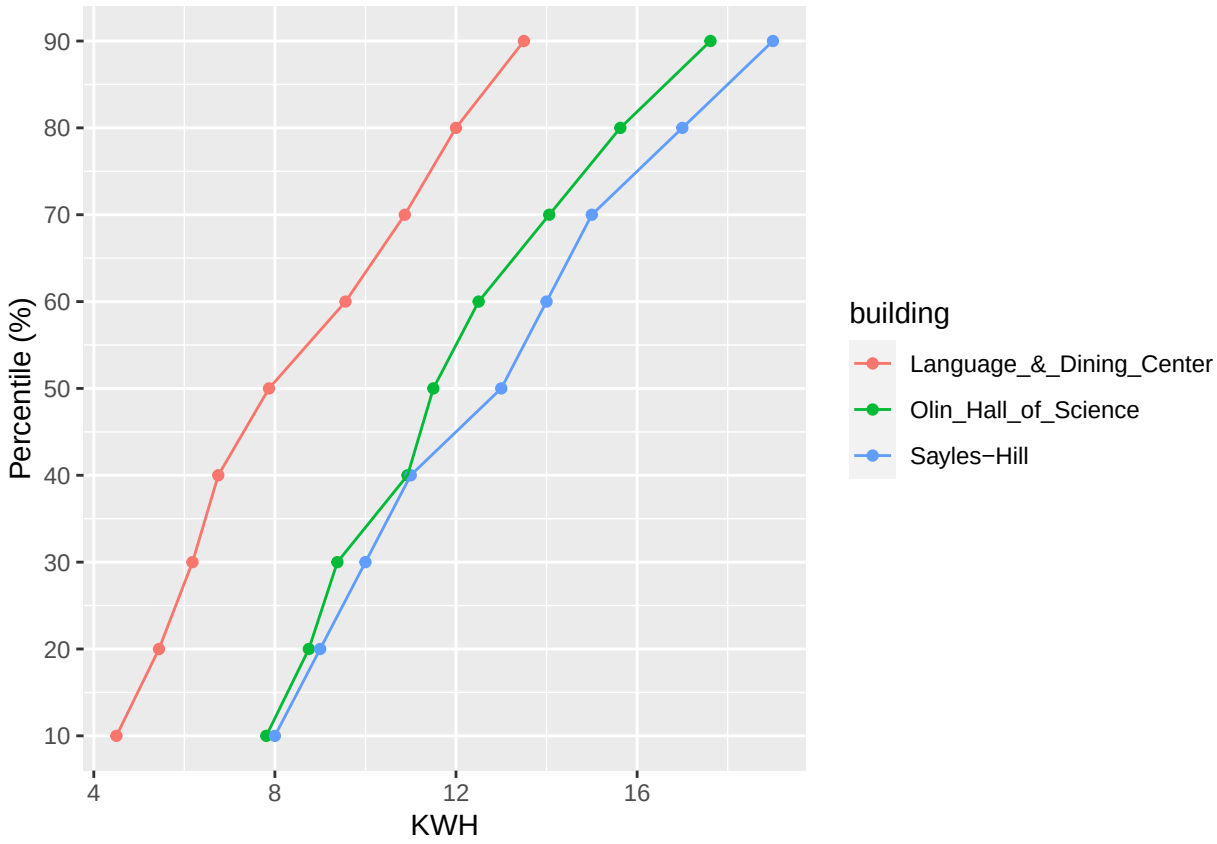
```
qdf <- energy %>% select(9:90) %>%
  map_dfc(quantile, probs = seq(.1,.9,.1), na.rm = TRUE) %>%
  mutate(stat = str_c("quantile_", seq(10,90,10))) %>%
  gather(key = building, value = q, 1:82)
qdf
# A tibble: 738 x 3
   stat        building              q
   <chr>       <chr>             <dbl>
 1 quantile_10 100_Nevada_Street 0.0972
 2 quantile_20 100_Nevada_Street 0.120
 3 quantile_30 100_Nevada_Street 0.183
 4 quantile_40 100_Nevada_Street 0.461
 5 quantile_50 100_Nevada_Street 0.710
 6 quantile_60 100_Nevada_Street 0.795
 7 quantile_70 100_Nevada_Street 0.915
```

```
 8 quantile_80 100_Nevada_Street 1.11
 9 quantile_90 100_Nevada_Street 1.24
10 quantile_10 104_Maple_St.     1.04
# ... with 728 more rows
```

```
qdf1 <- energy %>% select(9:90) %>%
  map_dfc(quantile, probs = seq(.1,.9,.1), na.rm = TRUE) %>%
  mutate(stat = str_c("quantile_", seq(10,90,10))) %>%
  pivot_longer(names_to = "building", values_to = "quantiles", 1:82)
qdf1
# A tibble: 738 x 3
   stat        building                          quantiles
   <chr>       <chr>                                 <dbl>
 1 quantile_10 100_Nevada_Street                    0.0972
 2 quantile_10 104_Maple_St.                        1.04
 3 quantile_10 106_Winona_St.                       0.601
 4 quantile_10 Allen_House                          0.756
 5 quantile_10 Alumni_Guest_House/Johnson_House    17.0
 6 quantile_10 Arboretum_Office                     0.13
 7 quantile_10 Art_Studios                          0.23
 8 quantile_10 Benton_House                         1.59
 9 quantile_10 Berg_House                           1.06
10 quantile_10 Bird_House                           1.42
# ... with 728 more rows
```

4. Plot the KWH value for each quantile on the x-axis for the buildings `Sayles-Hill`, `Language_&_Dining_Center`, `Olin_Hall_of_Science`

```
qdf %>%
  filter(building %in% c("Sayles-Hill" ,"Language_&_Dining_Center", "Olin_Hall_of_Science"))  %>%
  ggplot(aes(x=q, y=parse_number(stat), color=building)) +
  geom_point() +
  geom_line(aes(group=building)) +
  labs(y="Percentile (%)",x="KWH") +
  scale_y_continuous(breaks=seq(10,90,by=10))
```


Percentile (%)
90
80
70
60
50
40
30
20
10
4
8
12
16
KWH
building
Language_&_Dining_Center
Olin_Hall_of_Science
Sayles−Hill