

```
install.packages('tidyverse')
install.packages('reshape2')
install.packages('scales')
install.packages('skimr')
install.packages('janitor')
install.packages('lubridate')
install.packages('dplyr')
install.packages('readr')
library(tidyverse)
library(reshape2)
library(scales)
library(skimr)
library(janitor)
library(lubridate)
library(dplyr)
library(readr)
daily_activity <- read.csv("dailyActivity_merged.csv")
daily_calories <- read.csv("dailyCalories_merged.csv")
daily_intensities <- read.csv("dailyIntensities_merged.csv")
daily_steps <- read.csv("dailySteps_merged.csv")
daily_weight <- read.csv("weightLogInfo_merged.csv")
daily_sleep <- read.csv("sleepDay_merged.csv")
head(daily_activity)
head(daily_intensities)
head(daily_calories)
head(daily_steps)
head(daily_weight)
head(daily_sleep)
daily_activity$ActivityDate <- as.Date(daily_activity$ActivityDate, "%m/%d/%Y")
daily_sleep$ActivityDate <- parse_date_time(daily_sleep$SleepDay,
                                             orders = "mdy HMS")
daily_sleep$ActivityDate <- as.Date(daily_sleep$ActivityDate,
                                     "%m/%d/%y %h:%m:%s")
merge_1 <- merge(daily_activity, daily_calories, by = c("Id", "Calories"))
merge_2 <- merge(daily_intensities, daily_intensities, by =
c("Id", "ActivityDay", "SedentaryMinutes",
  "LightlyActiveMinutes", "FairlyActiveMinutes", "VeryActiveMinutes",
  "SedentaryActiveDistance", "LightActiveDistance", "ModeratelyActiveDistance",
  "VeryActiveDistance"))

merge_daily <- merge(merge_1, merge_2, by = c("Id", "ActivityDay", "SedentaryMinutes",
  "LightlyActiveMinutes", "FairlyActiveMinutes", "VeryActiveMinutes",
  "SedentaryActiveDistance", "LightActiveDistance", "ModeratelyActiveDistance",
  "VeryActiveDistance")) %>%
  select(-ActivityDay) %>% rename(Date = ActivityDate)

daily_data <- merge(merge_daily, daily_sleep, by = "Id", all=TRUE) %>% drop_na() %>%
select(-SleepDay, -TrackerDistance)

options(repr.plot.width=30)
summary(daily_data)

sleepType_by_userType <- daily_data %>%
  group_by(Id) %>%
  summarise(
    user_type = factor(case_when(
      SedentaryMinutes > mean(SedentaryMinutes) & LightlyActiveMinutes <
mean(LightlyActiveMinutes) & FairlyActiveMinutes < mean(FairlyActiveMinutes) &
VeryActiveMinutes < mean(VeryActiveMinutes) ~ "Sedentary",
      SedentaryMinutes < mean(SedentaryMinutes) & LightlyActiveMinutes >
mean(LightlyActiveMinutes) & FairlyActiveMinutes < mean(FairlyActiveMinutes) &
VeryActiveMinutes < mean(VeryActiveMinutes) ~ "Lightly Active",
      SedentaryMinutes < mean(SedentaryMinutes) & LightlyActiveMinutes <
```

```

mean(LightlyActiveMinutes) & FairlyActiveMinutes > mean(FairlyActiveMinutes) &
VeryActiveMinutes < mean(VeryActiveMinutes) ~ "Fairly Active",
  SedentaryMinutes < mean(SedentaryMinutes) & LightlyActiveMinutes <
mean(LightlyActiveMinutes) & FairlyActiveMinutes < mean(FairlyActiveMinutes) &
VeryActiveMinutes > mean(VeryActiveMinutes) ~ "Very Active",
),levels=c("Sedentary", "Lightly Active", "Fairly Active", "Very Active")),
sleep_type = factor(case_when(
  mean(TotalMinutesAsleep) < 360 ~ "Bad Sleep",
  mean(TotalMinutesAsleep) > 360 & mean(TotalMinutesAsleep) <= 480 ~ "Normal
Sleep",
  mean(TotalMinutesAsleep) > 480 ~ "Over Sleep",
),levels=c("Bad Sleep", "Normal Sleep", "Over Sleep")), total_sleep =
sum(TotalMinutesAsleep) ,.groups="drop"
) %>%
drop_na() %>%
group_by(user_type) %>%
summarise(bad_sleepers = sum(sleep_type == "Bad Sleep"), normal_sleepers =
sum(sleep_type == "Normal Sleep"),over_sleepers = sum(sleep_type == "Over
Sleep"),total=n(),.groups="drop") %>%
group_by(user_type) %>%
summarise(
  bad_sleepers = bad_sleepers / total,
  normal_sleepers = normal_sleepers / total,
  over_sleepers = over_sleepers / total,
  .groups="drop"
)

```

#in R, I can use ggplot2 to create a bar chart of that data:

```
sleepType_by_userType_melted<- melt(sleepType_by_userType, id.vars = "user_type")
```

```

ggplot(sleepType_by_userType_melted, aes(user_type, value, fill = variable)) +
  geom_bar(position = "dodge", stat = "identity", color= "black") +
  scale_y_continuous(labels = scales::percent) +
  labs(x=NULL, fill="Sleep type") +
  theme(legend.position="right",text = element_text(size = 25),plot.title =
element_text(hjust = 0.5))

```

```

ggplot(data = daily_calories, aes (x = ActivityDay, y = Calories, colour =
(factor(Id)), group = 33)) +
  theme(axis.text.x=element_blank()) +
  geom_point() + geom_smooth()

```