R_analysis_Code.R

r1997351

2023-06-19

```
#install.packages("tidyverse")
#install.packages('janitor')

library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(here)
```

```
## here() starts at /cloud/project
```

```
library(skimr)
library(lubridate)
library(janitor)
```

```
##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
Daily_Activity=read.csv('dailyActivity_merged.csv')
Daily_Calories=read.csv('dailyCalories_merged.csv')
Daily_Intensities=read.csv('dailyIntensities_merged.csv')
Daily_Sleep=read.csv('sleepDay_merged.csv')
Daily_Sleep=read.csv('sleepDay_merged.csv')



Daily_Activity<-clean_names(Daily_Activity)
Daily_Activity<- rename_with(Daily_Activity, tolower)

Daily_Sleep<-clean_names(Daily_Sleep)
Daily_Sleep<- rename_with(Daily_Sleep, tolower)
```

```
Daily_Intensities<-clean_names(Daily_Intensities)
Daily_Intensities<- rename_with(Daily_Intensities, tolower)

## Let us now count how many unique Ids are there in each of these datasets :
n_distinct(Daily_Activity$id)
```

```
## [1] 33
```

```
n_distinct(Daily_Sleep$id)
```

```
## [1] 24
```

```
n_distinct(Daily_Intensities$id)
```

```
## [1] 33
```

```
## Drop all the duplicate bservations :
Daily_Activity <- Daily_Activity %>% distinct() %>% drop_na()
Daily_Intensities <- Daily_Intensities %>% distinct() %>% drop_na()
Daily_Sleep <- Daily_Sleep %>% distinct() %>% drop_na()

## check if still duplicate records exist
sum(duplicated(Daily_Sleep))
```

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

```
sum(duplicated(Daily_Activity))
```

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

```
## LEt us now analyze the summary statistics :
summary(Daily_Activity)
```

```
##        id            activity_date       total_steps    total_distance
##  Min.   :1.504e+09   Length:940         Min.   :    0   Min.   : 0.000
##  1st Qu.:2.320e+09   Class :character   1st Qu.: 3790   1st Qu.: 2.620
##  Median :4.445e+09   Mode  :character   Median : 7406   Median : 5.245
##  Mean   :4.855e+09                      Mean   : 7638   Mean   : 5.490
##  3rd Qu.:6.962e+09                      3rd Qu.:10727   3rd Qu.: 7.713
##  Max.   :8.878e+09                      Max.   :36019   Max.   :28.030
##  tracker_distance logged_activities_distance very_active_distance
##  Min.   : 0.000   Min.   :0.0000             Min.   : 0.000
##  1st Qu.: 2.620   1st Qu.:0.0000             1st Qu.: 0.000
##  Median : 5.245   Median :0.0000             Median : 0.210
##  Mean   : 5.475   Mean   :0.1082             Mean   : 1.503
##  3rd Qu.: 7.710   3rd Qu.:0.0000             3rd Qu.: 2.053
##  Max.   :28.030   Max.   :4.9421             Max.   :21.920
##  moderately_active_distance light_active_distance sedentary_active_distance
##  Min.   :0.0000             Min.   : 0.000        Min.   :0.000000
##  1st Qu.:0.0000             1st Qu.: 1.945        1st Qu.:0.000000
##  Median :0.2400             Median : 3.365        Median :0.000000
##  Mean   :0.5675             Mean   : 3.341        Mean   :0.001606
##  3rd Qu.:0.8000             3rd Qu.: 4.782        3rd Qu.:0.000000
##  Max.   :6.4800             Max.   :10.710        Max.   :0.110000
##  very_active_minutes fairly_active_minutes lightly_active_minutes
##  Min.   :  0.00      Min.   :  0.00        Min.   :  0.0
##  1st Qu.:  0.00      1st Qu.:  0.00        1st Qu.:127.0
```

```
##  Median :  4.00      Median :  6.00        Median :199.0
##  Mean   : 21.16      Mean   : 13.56        Mean   :192.8
##  3rd Qu.: 32.00      3rd Qu.: 19.00        3rd Qu.:264.0
##  Max.   :210.00      Max.   :143.00        Max.   :518.0
##  sedentary_minutes    calories
##  Min.   :   0.0    Min.   :   0
##  1st Qu.: 729.8    1st Qu.:1828
##  Median :1057.5    Median :2134
##  Mean   : 991.2    Mean   :2304
##  3rd Qu.:1229.5    3rd Qu.:2793
##  Max.   :1440.0    Max.   :4900
```

```
summary(Daily_Sleep)
```

```
##        id             sleep_day         total_sleep_records
##  Min.   :1.504e+09   Length:410         Min.   :1.00
##  1st Qu.:3.977e+09   Class :character   1st Qu.:1.00
##  Median :4.703e+09   Mode  :character   Median :1.00
##  Mean   :4.995e+09                      Mean   :1.12
##  3rd Qu.:6.962e+09                      3rd Qu.:1.00
##  Max.   :8.792e+09                      Max.   :3.00
##  total_minutes_asleep total_time_in_bed
##  Min.   : 58.0        Min.   : 61.0
##  1st Qu.:361.0        1st Qu.:403.8
##  Median :432.5        Median :463.0
##  Mean   :419.2        Mean   :458.5
##  3rd Qu.:490.0        3rd Qu.:526.0
##  Max.   :796.0        Max.   :961.0
```

```
##Select few specific variables and perform summary stats analysis on them:
Daily_Activity %>%
  select(total_steps, total_distance, sedentary_minutes) %>% summary()
```

```
##   total_steps    total_distance   sedentary_minutes
##  Min.   :    0   Min.   : 0.000   Min.   :   0.0
##  1st Qu.: 3790   1st Qu.: 2.620   1st Qu.: 729.8
##  Median : 7406   Median : 5.245   Median :1057.5
##  Mean   : 7638   Mean   : 5.490   Mean   : 991.2
##  3rd Qu.:10727   3rd Qu.: 7.713   3rd Qu.:1229.5
##  Max.   :36019   Max.   :28.030   Max.   :1440.0
```

```
Daily_Sleep %>%
  select(total_sleep_records, total_minutes_asleep, total_time_in_bed) %>% summary()
```

```
##  total_sleep_records total_minutes_asleep total_time_in_bed
##  Min.   :1.00        Min.   : 58.0        Min.   : 61.0
##  1st Qu.:1.00        1st Qu.:361.0        1st Qu.:403.8
##  Median :1.00        Median :432.5        Median :463.0
##  Mean   :1.12        Mean   :419.2        Mean   :458.5
##  3rd Qu.:1.00        3rd Qu.:490.0        3rd Qu.:526.0
##  Max.   :3.00        Max.   :796.0        Max.   :961.0
```

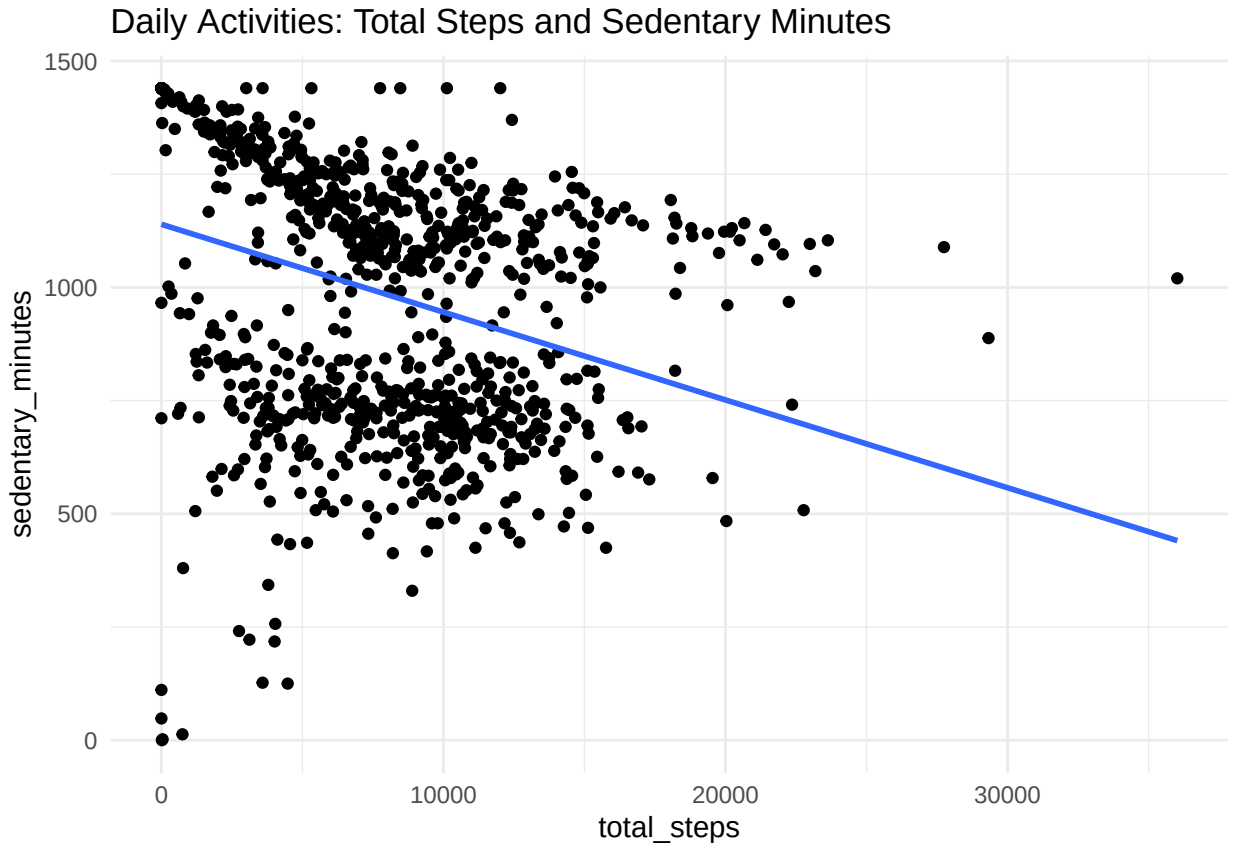
```
Daily_Intensities %>%
  select(lightly_active_minutes, fairly_active_minutes, very_active_minutes) %>% summary()
```

```
##  lightly_active_minutes fairly_active_minutes very_active_minutes
##  Min.   :  0.0          Min.   :  0.00        Min.   :  0.00
```

```
##  1st Qu.:127.0          1st Qu.:  0.00        1st Qu.:  0.00
##  Median :199.0          Median :  6.00        Median :  4.00
##  Mean   :192.8          Mean   : 13.56        Mean   : 21.16
##  3rd Qu.:264.0          3rd Qu.: 19.00        3rd Qu.: 32.00
##  Max.   :518.0          Max.   :143.00        Max.   :210.00
```

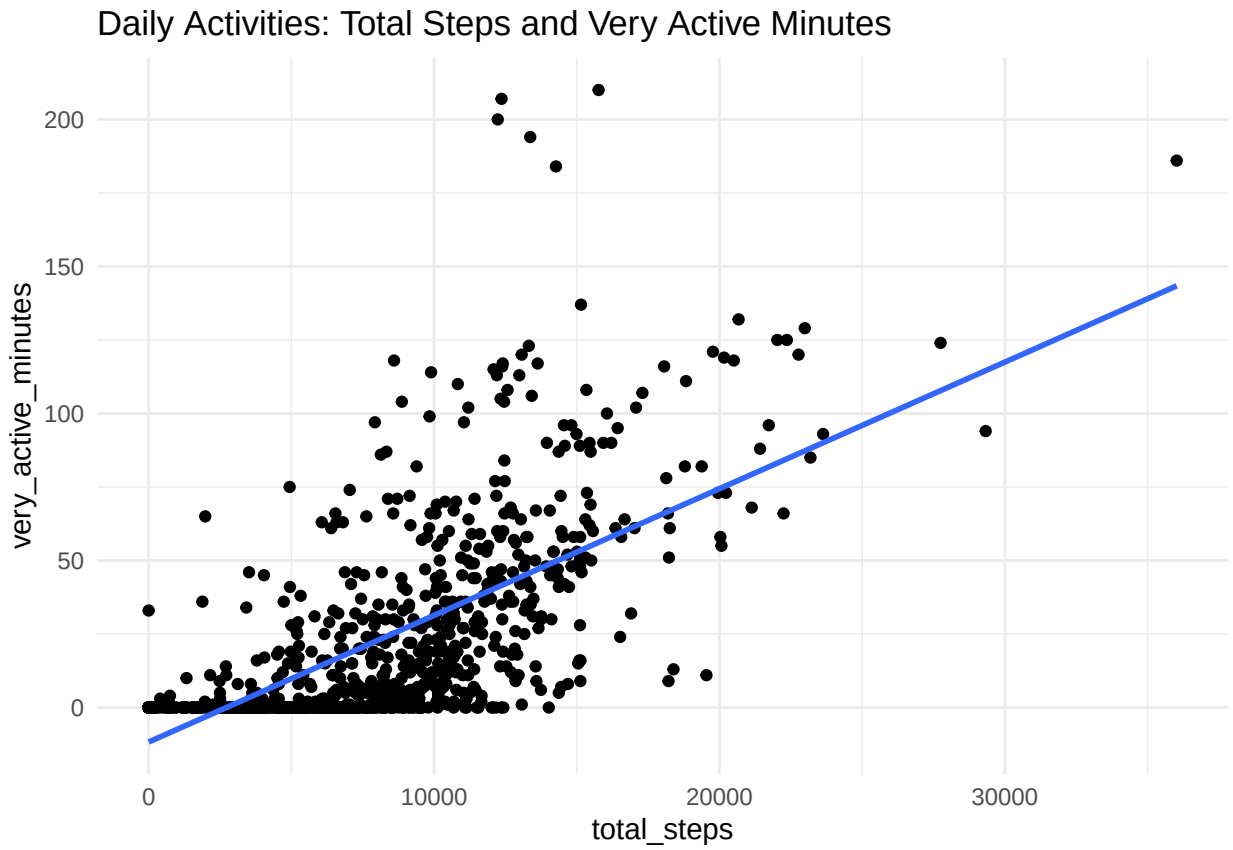
```
## analysing total steps vs sedentary minutes from the daily activity dataset :
ggplot(data=Daily_Activity, aes(x=total_steps, y=sedentary_minutes)) +
  geom_point() + geom_smooth(method = "lm", se = FALSE) +
  labs(x ="total_steps", y = "sedentary_minutes", title = "Daily Activities: Total Steps and Sedentary M
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

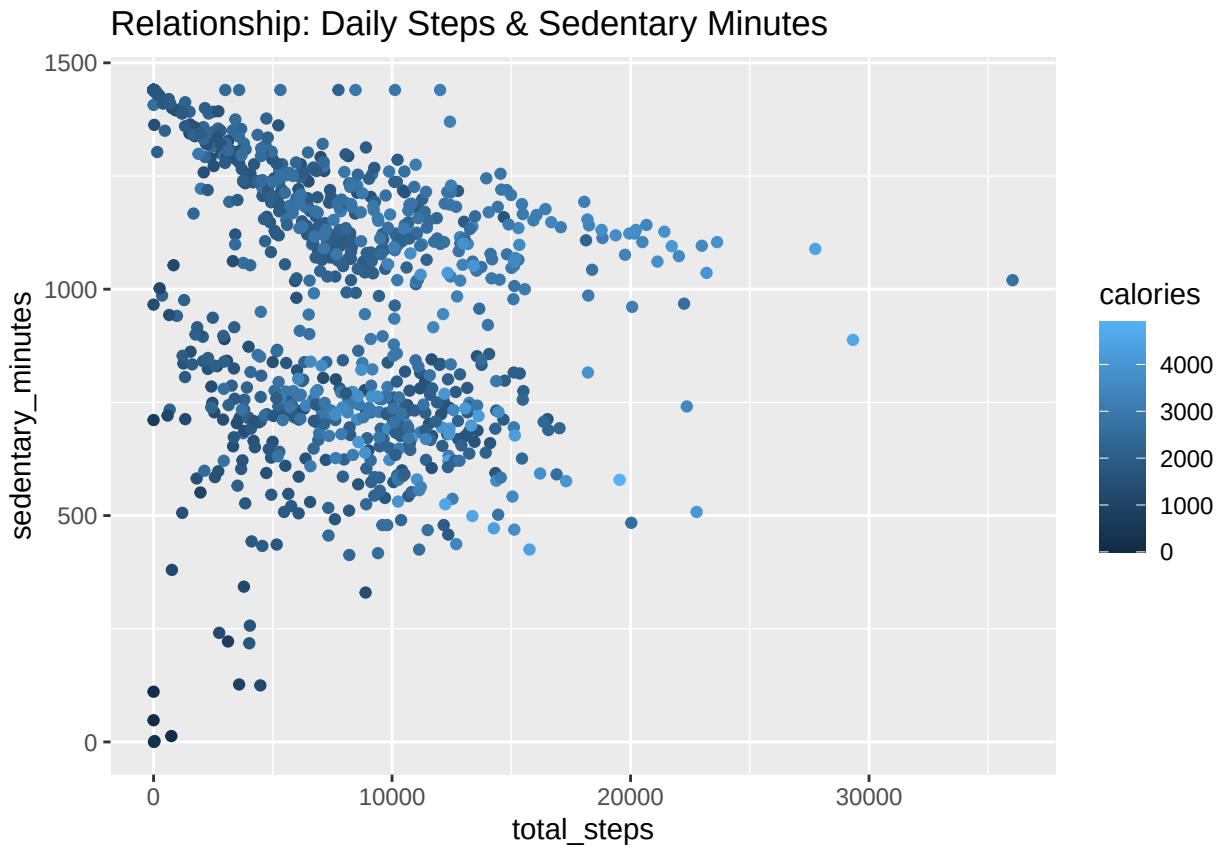



```
## analysing total steps vs Very active minutes from the daily activity dataset :
ggplot(data=Daily_Activity, aes(x=total_steps, y=very_active_minutes)) +
  geom_point() + geom_smooth(method = "lm", se = FALSE) +
  labs(x ="total_steps", y = "very_active_minutes", title = "Daily Activities: Total Steps and Very Acti
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
ggplot(data=Daily_Activity, aes(x=total_steps, y=sedentary_minutes, color = calories)) +
geom_point() +
labs(title = "Relationship: Daily Steps & Sedentary Minutes")
```



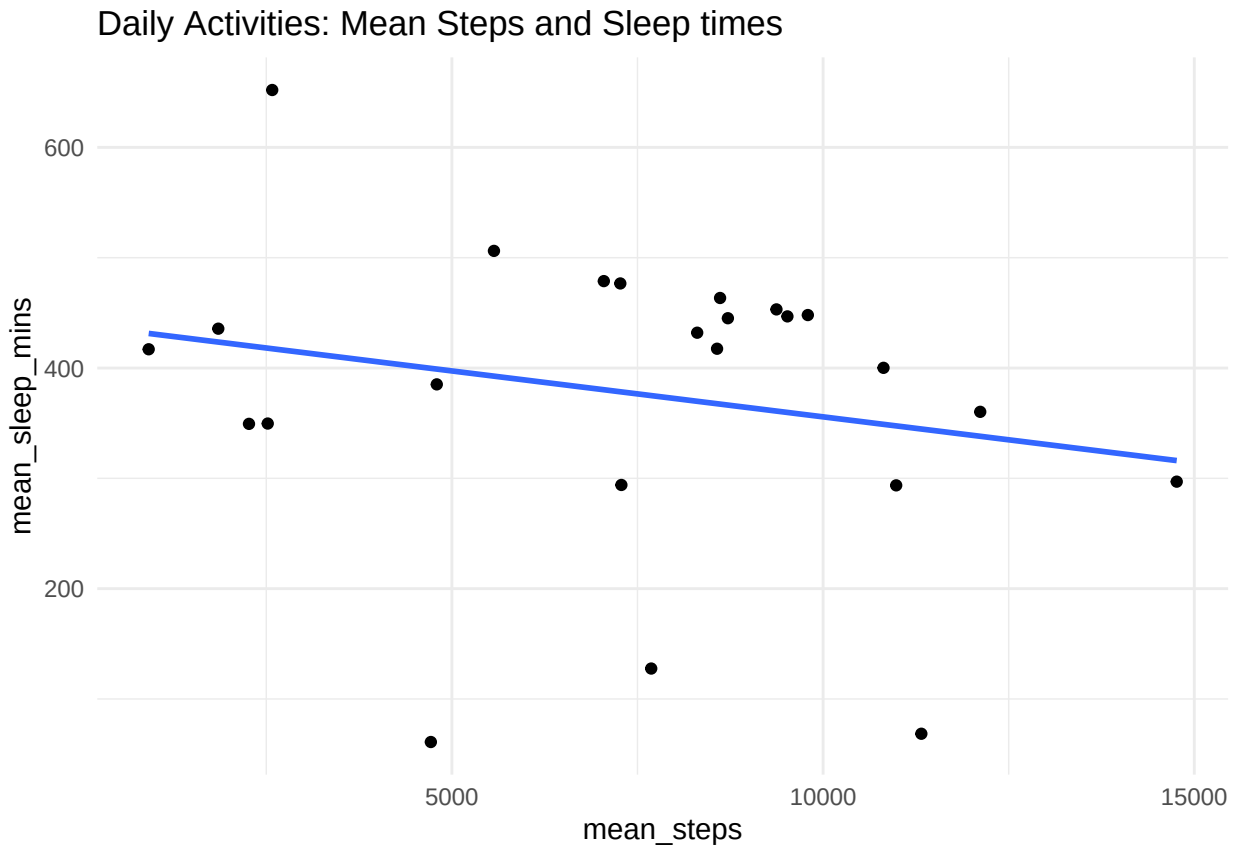

```
Activity_unique <- Daily_Activity %>%
  group_by(id) %>%
  summarise(mean_steps=mean(total_steps),
            mean_distance=mean(total_distance),
            mean_sed_mins=mean(sedentary_minutes),
            mean_cal=mean(calories)) %>%
  arrange(desc(mean_cal))

Sleep_unique <- Daily_Sleep %>%
  group_by(id) %>%
  summarise(mean_sleep_mins=mean(total_minutes_asleep))%>%
  arrange(desc(mean_sleep_mins))


## Merging activity and intesnity information :
ActivityandSleep <- merge(Activity_unique, Sleep_unique, by="id", how='inner')

## analysing mean teps vs sleep times from the merged dataset:
ggplot(data=ActivityandSleep, aes(x=mean_steps, y=mean_sleep_mins)) +
  geom_point() + geom_smooth(method = "lm", se = FALSE) +
  labs(x ="mean_steps", y = "mean_sleep_mins", title = "Daily Activities: Mean Steps and Sleep times")+
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
## find out the correlation :
cor(ActivityandSleep$mean_steps,ActivityandSleep$mean_sleep_mins)
```

```
## [1] -0.219686
```