bellabeat.R

Brianna’s Laptop

2022-05-31

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──

## ✔ ggplot2 3.3.6 ✔ purrr 0.3.4  
## ✔ tibble 3.1.7 ✔ dplyr 1.0.9  
## ✔ tidyr 1.2.0 ✔ stringr 1.4.0  
## ✔ readr 2.1.2 ✔ forcats 0.5.1

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(lubridate)

##   
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':  
##   
## date, intersect, setdiff, union

library(skimr)  
library(janitor)

##   
## Attaching package: 'janitor'

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

library(ggplot2)  
library(dplyr)  
  
library(readr)  
daily\_activity <- read\_csv("Google Data Analytics/CAPSTONE BELLABEAT/dailyActivity\_merged.csv")

## Rows: 940 Columns: 15

## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): ActivityDate  
## dbl (14): Id, TotalSteps, TotalDistance, TrackerDistance, LoggedActivitiesDi...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

sleep\_day <- read\_csv("Google Data Analytics/CAPSTONE BELLABEAT/sleepDay\_merged.csv")

## Rows: 413 Columns: 5  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): SleepDay  
## dbl (4): Id, TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

daily\_calories <- read\_csv("Google Data Analytics/CAPSTONE BELLABEAT/dailyCalories\_merged.csv")

## Rows: 940 Columns: 3  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): ActivityDay  
## dbl (2): Id, Calories  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

heartrate\_seconds <- read\_csv("Google Data Analytics/CAPSTONE BELLABEAT/heartrate\_seconds\_merged.csv")

## Rows: 2483658 Columns: 3  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): Time  
## dbl (2): Id, Value  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

daily\_intensities <- read\_csv("Google Data Analytics/CAPSTONE BELLABEAT/dailyIntensities\_merged.csv")

## Rows: 940 Columns: 10  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): ActivityDay  
## dbl (9): Id, SedentaryMinutes, LightlyActiveMinutes, FairlyActiveMinutes, Ve...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

weight\_log <- read\_csv("Google Data Analytics/CAPSTONE BELLABEAT/weightLogInfo\_merged.csv")

## Rows: 67 Columns: 8  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): Date  
## dbl (6): Id, WeightKg, WeightPounds, Fat, BMI, LogId  
## lgl (1): IsManualReport  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

str(daily\_activity)

## spec\_tbl\_df [940 × 15] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ Id : num [1:940] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...  
## $ ActivityDate : chr [1:940] "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...  
## $ TotalSteps : num [1:940] 13162 10735 10460 9762 12669 ...  
## $ TotalDistance : num [1:940] 8.5 6.97 6.74 6.28 8.16 ...  
## $ TrackerDistance : num [1:940] 8.5 6.97 6.74 6.28 8.16 ...  
## $ LoggedActivitiesDistance: num [1:940] 0 0 0 0 0 0 0 0 0 0 ...  
## $ VeryActiveDistance : num [1:940] 1.88 1.57 2.44 2.14 2.71 ...  
## $ ModeratelyActiveDistance: num [1:940] 0.55 0.69 0.4 1.26 0.41 ...  
## $ LightActiveDistance : num [1:940] 6.06 4.71 3.91 2.83 5.04 ...  
## $ SedentaryActiveDistance : num [1:940] 0 0 0 0 0 0 0 0 0 0 ...  
## $ VeryActiveMinutes : num [1:940] 25 21 30 29 36 38 42 50 28 19 ...  
## $ FairlyActiveMinutes : num [1:940] 13 19 11 34 10 20 16 31 12 8 ...  
## $ LightlyActiveMinutes : num [1:940] 328 217 181 209 221 164 233 264 205 211 ...  
## $ SedentaryMinutes : num [1:940] 728 776 1218 726 773 ...  
## $ Calories : num [1:940] 1985 1797 1776 1745 1863 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. Id = col\_double(),  
## .. ActivityDate = col\_character(),  
## .. TotalSteps = col\_double(),  
## .. TotalDistance = col\_double(),  
## .. TrackerDistance = col\_double(),  
## .. LoggedActivitiesDistance = col\_double(),  
## .. VeryActiveDistance = col\_double(),  
## .. ModeratelyActiveDistance = col\_double(),  
## .. LightActiveDistance = col\_double(),  
## .. SedentaryActiveDistance = col\_double(),  
## .. VeryActiveMinutes = col\_double(),  
## .. FairlyActiveMinutes = col\_double(),  
## .. LightlyActiveMinutes = col\_double(),  
## .. SedentaryMinutes = col\_double(),  
## .. Calories = col\_double()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(sleep\_day)

## spec\_tbl\_df [413 × 5] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ Id : num [1:413] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...  
## $ SleepDay : chr [1:413] "4/12/2016 12:00:00 AM" "4/13/2016 12:00:00 AM" "4/15/2016 12:00:00 AM" "4/16/2016 12:00:00 AM" ...  
## $ TotalSleepRecords : num [1:413] 1 2 1 2 1 1 1 1 1 1 ...  
## $ TotalMinutesAsleep: num [1:413] 327 384 412 340 700 304 360 325 361 430 ...  
## $ TotalTimeInBed : num [1:413] 346 407 442 367 712 320 377 364 384 449 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. Id = col\_double(),  
## .. SleepDay = col\_character(),  
## .. TotalSleepRecords = col\_double(),  
## .. TotalMinutesAsleep = col\_double(),  
## .. TotalTimeInBed = col\_double()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(daily\_calories)

## spec\_tbl\_df [940 × 3] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ Id : num [1:940] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...  
## $ ActivityDay: chr [1:940] "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...  
## $ Calories : num [1:940] 1985 1797 1776 1745 1863 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. Id = col\_double(),  
## .. ActivityDay = col\_character(),  
## .. Calories = col\_double()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(heartrate\_seconds)

## spec\_tbl\_df [2,483,658 × 3] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ Id : num [1:2483658] 2.02e+09 2.02e+09 2.02e+09 2.02e+09 2.02e+09 ...  
## $ Time : chr [1:2483658] "4/12/2016 7:21:00 AM" "4/12/2016 7:21:05 AM" "4/12/2016 7:21:10 AM" "4/12/2016 7:21:20 AM" ...  
## $ Value: num [1:2483658] 97 102 105 103 101 95 91 93 94 93 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. Id = col\_double(),  
## .. Time = col\_character(),  
## .. Value = col\_double()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(daily\_intensities)

## spec\_tbl\_df [940 × 10] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ Id : num [1:940] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...  
## $ ActivityDay : chr [1:940] "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...  
## $ SedentaryMinutes : num [1:940] 728 776 1218 726 773 ...  
## $ LightlyActiveMinutes : num [1:940] 328 217 181 209 221 164 233 264 205 211 ...  
## $ FairlyActiveMinutes : num [1:940] 13 19 11 34 10 20 16 31 12 8 ...  
## $ VeryActiveMinutes : num [1:940] 25 21 30 29 36 38 42 50 28 19 ...  
## $ SedentaryActiveDistance : num [1:940] 0 0 0 0 0 0 0 0 0 0 ...  
## $ LightActiveDistance : num [1:940] 6.06 4.71 3.91 2.83 5.04 ...  
## $ ModeratelyActiveDistance: num [1:940] 0.55 0.69 0.4 1.26 0.41 ...  
## $ VeryActiveDistance : num [1:940] 1.88 1.57 2.44 2.14 2.71 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. Id = col\_double(),  
## .. ActivityDay = col\_character(),  
## .. SedentaryMinutes = col\_double(),  
## .. LightlyActiveMinutes = col\_double(),  
## .. FairlyActiveMinutes = col\_double(),  
## .. VeryActiveMinutes = col\_double(),  
## .. SedentaryActiveDistance = col\_double(),  
## .. LightActiveDistance = col\_double(),  
## .. ModeratelyActiveDistance = col\_double(),  
## .. VeryActiveDistance = col\_double()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(weight\_log)

## spec\_tbl\_df [67 × 8] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ Id : num [1:67] 1.50e+09 1.50e+09 1.93e+09 2.87e+09 2.87e+09 ...  
## $ Date : chr [1:67] "5/2/2016 11:59:59 PM" "5/3/2016 11:59:59 PM" "4/13/2016 1:08:52 AM" "4/21/2016 11:59:59 PM" ...  
## $ WeightKg : num [1:67] 52.6 52.6 133.5 56.7 57.3 ...  
## $ WeightPounds : num [1:67] 116 116 294 125 126 ...  
## $ Fat : num [1:67] 22 NA NA NA NA 25 NA NA NA NA ...  
## $ BMI : num [1:67] 22.6 22.6 47.5 21.5 21.7 ...  
## $ IsManualReport: logi [1:67] TRUE TRUE FALSE TRUE TRUE TRUE ...  
## $ LogId : num [1:67] 1.46e+12 1.46e+12 1.46e+12 1.46e+12 1.46e+12 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. Id = col\_double(),  
## .. Date = col\_character(),  
## .. WeightKg = col\_double(),  
## .. WeightPounds = col\_double(),  
## .. Fat = col\_double(),  
## .. BMI = col\_double(),  
## .. IsManualReport = col\_logical(),  
## .. LogId = col\_double()  
## .. )  
## - attr(\*, "problems")=<externalptr>

daily\_activity$ActivityDate <- as.Date(daily\_activity$ActivityDate, format = "%m/%d/%Y")  
sleep\_day$SleepDay <- as.Date(sleep\_day$SleepDay, format = "%m/%d/%Y")  
daily\_intensities$ActivityDay <- as.Date(daily\_intensities$ActivityDay, format = "%m/%d/%Y")  
daily\_calories$ActivityDay <- as.Date(daily\_calories$ActivityDay, format = "%m/%d/%Y")  
weight\_log$Date <- as.Date(weight\_log$Date, format = "%m/%d/%Y")  
  
  
class(daily\_activity$ActivityDate)

## [1] "Date"

class(sleep\_day$SleepDay)

## [1] "Date"

class(daily\_intensities$ActivityDay)

## [1] "Date"

class(daily\_calories$ActivityDay)

## [1] "Date"

class(weight\_log$Date)

## [1] "Date"

sum(is.na(daily\_activity))

## [1] 0

sum(is.na(sleep\_day))

## [1] 0

sum(is.na(daily\_calories))

## [1] 0

sum(is.na(heartrate\_seconds))

## [1] 0

sum(is.na(daily\_intensities))

## [1] 0

sum(is.na(weight\_log))

## [1] 65

n\_distinct(daily\_activity$Id)

## [1] 33

n\_distinct(sleep\_day$Id)

## [1] 24

n\_distinct(daily\_intensities$Id)

## [1] 33

n\_distinct(daily\_calories$Id)

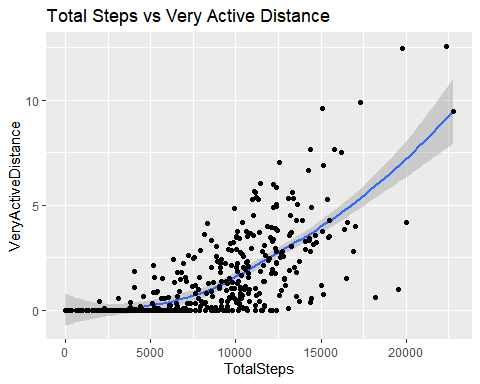
## [1] 33

n\_distinct(weight\_log$Id)

## [1] 8

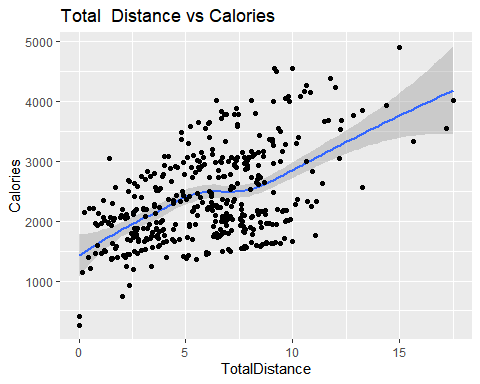
sleep\_day <- sleep\_day %>% rename(Date = SleepDay)  
daily\_activity <- daily\_activity %>% rename(Date = ActivityDate)  
daily\_intensities <- daily\_intensities %>% rename(Date = ActivityDay)  
  
active\_sleep <- merge(sleep\_day, daily\_activity, by=c("Id", "Date"))  
intensity\_weight <- merge(daily\_intensities, weight\_log, by=c("Id", "Date"))  
  
ggplot(data = active\_sleep)+   
 geom\_smooth(mapping = aes(x=TotalSteps, y=VeryActiveDistance))+ geom\_point(mapping = aes(x=TotalSteps, y=VeryActiveDistance))+labs(title = "Total Steps vs Very Active Distance")

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



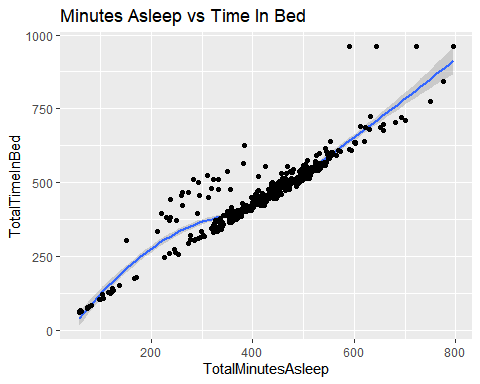
ggplot(data = active\_sleep)+ geom\_smooth(mapping = aes(x=TotalDistance, y=Calories))+geom\_point(mapping = aes(x=TotalDistance, y=Calories))+labs(title = "Total Distance vs Calories")

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



ggplot(data = active\_sleep)+ geom\_smooth(mapping = aes(x=TotalMinutesAsleep, y=TotalTimeInBed))+geom\_point(mapping = aes(x=TotalMinutesAsleep, y=TotalTimeInBed))+labs(title = "Minutes Asleep vs Time In Bed")

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



daily\_activity %>%   
 select(TotalSteps, TotalDistance, SedentaryMinutes,LightlyActiveMinutes,FairlyActiveMinutes,VeryActiveMinutes,Calories) %>%   
 summary()

## TotalSteps TotalDistance SedentaryMinutes LightlyActiveMinutes  
## Min. : 0 Min. : 0.000 Min. : 0.0 Min. : 0.0   
## 1st Qu.: 3790 1st Qu.: 2.620 1st Qu.: 729.8 1st Qu.:127.0   
## Median : 7406 Median : 5.245 Median :1057.5 Median :199.0   
## Mean : 7638 Mean : 5.490 Mean : 991.2 Mean :192.8   
## 3rd Qu.:10727 3rd Qu.: 7.713 3rd Qu.:1229.5 3rd Qu.:264.0   
## Max. :36019 Max. :28.030 Max. :1440.0 Max. :518.0   
## FairlyActiveMinutes VeryActiveMinutes Calories   
## Min. : 0.00 Min. : 0.00 Min. : 0   
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.:1828   
## Median : 6.00 Median : 4.00 Median :2134   
## Mean : 13.56 Mean : 21.16 Mean :2304   
## 3rd Qu.: 19.00 3rd Qu.: 32.00 3rd Qu.:2793   
## Max. :143.00 Max. :210.00 Max. :4900

heartrate\_seconds %>%   
 select(Value) %>%   
 summary()

## Value   
## Min. : 36.00   
## 1st Qu.: 63.00   
## Median : 73.00   
## Mean : 77.33   
## 3rd Qu.: 88.00   
## Max. :203.00

sleep\_day %>%   
 select(TotalSleepRecords,TotalMinutesAsleep,TotalTimeInBed) %>%   
 summary()

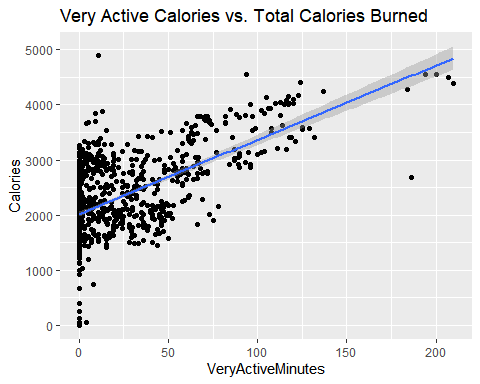
## TotalSleepRecords TotalMinutesAsleep TotalTimeInBed   
## Min. :1.000 Min. : 58.0 Min. : 61.0   
## 1st Qu.:1.000 1st Qu.:361.0 1st Qu.:403.0   
## Median :1.000 Median :433.0 Median :463.0   
## Mean :1.119 Mean :419.5 Mean :458.6   
## 3rd Qu.:1.000 3rd Qu.:490.0 3rd Qu.:526.0   
## Max. :3.000 Max. :796.0 Max. :961.0

weight\_log %>%   
 select(WeightPounds,BMI) %>%   
 summary()

## WeightPounds BMI   
## Min. :116.0 Min. :21.45   
## 1st Qu.:135.4 1st Qu.:23.96   
## Median :137.8 Median :24.39   
## Mean :158.8 Mean :25.19   
## 3rd Qu.:187.5 3rd Qu.:25.56   
## Max. :294.3 Max. :47.54

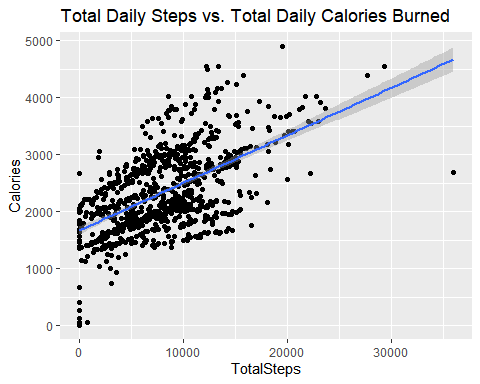
ggplot(data=daily\_activity,aes(x=VeryActiveMinutes,y=Calories))+  
 geom\_point()+stat\_smooth(method=lm)+labs(title="Very Active Calories vs. Total Calories Burned")

## `geom\_smooth()` using formula 'y ~ x'



ggplot(data=daily\_activity, aes(x=TotalSteps, y=Calories))+  
 geom\_point()+stat\_smooth(method=lm)+labs(title="Total Daily Steps vs. Total Daily Calories Burned")

## `geom\_smooth()` using formula 'y ~ x'



library(rmarkdown)