Coursera Fitabase Data Case Study

Joseph Lennan

2022-09-28

Methodology

The goal of this case study for the Google Analytics Certificate is to analyze the provided smart fitness data to provide insight into trends and potential opportunities with Bellabeat's current products in mind.

For this case study, the initial data to use for inspiration and a few graphs were provided to get people started. I ended up focusing my insights on the relationship for sleep and exercise, although there's probably a great deal more potential within the data.

Documentation

Documentation of Data https://www.fitabase.com/media/1930/fitabasedatadictionary102320.pdf Distance in Kilometers Calories is Calories burned

Data Fetching

```
#Get Dataframes

sleepDay = myfiles$sleepDay_merged.csv
dailyActivity = myfiles$dailyActivity_merged.csv
dailyIntensities = myfiles$dailyIntensities_merged.csv
dailySteps = myfiles$dailySteps_merged.csv
dailyCalories = myfiles$dailyCalories_merged.csv
hourlyCalories = myfiles$hourlyCalories_merged.csv
hourlyIntensities = myfiles$hourlyIntensities_merged.csv
weightLog = myfiles$weightLogInfo_merged.csv
```

#Data Skimming

Examining the skimr output to see if there's any interesting outliers, observations, or measures of central tendency. In addition, I also had to look for the documentation of the data to make more sense of some parts of the data, such as calories being burned or eaten as well as what unit distance was in.

```
sleepDay %>%
  select(TotalSleepRecords,
         TotalMinutesAsleep,
         TotalTimeInBed
  ) %>%
  skim_without_charts()
```

Table 1: Data summary

Name	Piped data
Number of rows	413
Number of columns	3
Column type frequency:	
numeric	3
Group variables	None

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
TotalSleepRecords	0	1	1.12	0.35	1	1	1	1	3
TotalMinutesAsleep	0	1	419.47	118.34	58	361	433	490	796
TotalTimeInBed	0	1	458.64	127.10	61	403	463	526	961

```
#Mean 419 Min Asleep, 39.2 Mean Extra Minutes in Bed
#Standard Deviations seem high, ~68% of people take ~86 minutes to go to bed?
#Might be counting time after waking and still lying in bed



dailyActivity %>%
  select(
    TotalDistance,
    TotalSteps,
    Calories,
    TrackerDistance,
    LoggedActivitiesDistance,
    SedentaryActiveDistance,
    SedentaryMinutes,
    LightActiveDistance,
    LightlyActiveMinutes,
    ModeratelyActiveDistance,
    FairlyActiveMinutes,
    VeryActiveDistance,
    VeryActiveMinutes
  ) %>%
  skim_without_charts()
```

Table 3: Data summary

Name	Piped data
Number of rows	940
Number of columns	13
Column type frequency:	
numeric	13

Table 3: Data summary

Group variables	None

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
TotalDistance	0	1	5.49	3.92	0	2.62	5.24	7.71	28.03
TotalSteps	0	1	7637.91	5087.15	0	3789.75	7405.50	10727.00	36019.00
Calories	0	1	2303.61	718.17	0	1828.50	2134.00	2793.25	4900.00
TrackerDistance	0	1	5.48	3.91	0	2.62	5.24	7.71	28.03
LoggedActivitiesDistance	0	1	0.11	0.62	0	0.00	0.00	0.00	4.94
SedentaryActiveDistance	0	1	0.00	0.01	0	0.00	0.00	0.00	0.11
SedentaryMinutes	0	1	991.21	301.27	0	729.75	1057.50	1229.50	1440.00
LightActiveDistance	0	1	3.34	2.04	0	1.95	3.36	4.78	10.71
LightlyActiveMinutes	0	1	192.81	109.17	0	127.00	199.00	264.00	518.00
ModeratelyActiveDistance	0	1	0.57	0.88	0	0.00	0.24	0.80	6.48
FairlyActiveMinutes	0	1	13.56	19.99	0	0.00	6.00	19.00	143.00
VeryActiveDistance	0	1	1.50	2.66	0	0.00	0.21	2.05	21.92
VeryActiveMinutes	0	1	21.16	32.84	0	0.00	4.00	32.00	210.00

```
#5.49 Units Mean Total Distance, 7638 Mean Steps, 2304 Mean Calories...

dailyIntensities %>%
  select(
    SedentaryMinutes,
    LightlyActiveMinutes,
    FairlyActiveMinutes,
    VeryActiveMinutes,
    SedentaryActiveDistance,
    LightActiveDistance,
    ModeratelyActiveDistance,
    VeryActiveDistance
  ) %>%
  skim_without_charts()
```

Table 5: Data summary

Name	Piped data
Number of rows	940
Number of columns	8
Column type frequency:	
numeric	8
Group variables	None

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
SedentaryMinutes	0	1	991.21	301.27	0	729.75	1057.50	1229.50	1440.00
LightlyActiveMinutes	0	1	192.81	109.17	0	127.00	199.00	264.00	518.00
FairlyActiveMinutes	0	1	13.56	19.99	0	0.00	6.00	19.00	143.00
VeryActiveMinutes	0	1	21.16	32.84	0	0.00	4.00	32.00	210.00
SedentaryActiveDistance	0	1	0.00	0.01	0	0.00	0.00	0.00	0.11
LightActiveDistance	0	1	3.34	2.04	0	1.95	3.36	4.78	10.71
ModeratelyActiveDistance	0	1	0.57	0.88	0	0.00	0.24	0.80	6.48
VeryActiveDistance	0	1	1.50	2.66	0	0.00	0.21	2.05	21.92

```r
#193 mean lightly active minutes, 32 fairly active mean minutes


dailySteps %>%
  select(
    StepTotal
  ) %>%
  skim_without_charts()
```

Table 7: Data summary

Name	Piped data
Number of rows	940
Number of columns	1
Column type frequency:	
numeric	1
Group variables	None

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
StepTotal	0	1	7637.91	5087.15	0	3789.75	7405.5	10727	36019

```r
dailyCalories %>%
  select(
    Calories
  ) %>%
  skim_without_charts()
```

Table 9: Data summary

Name	Piped data
Number of rows	940
Number of columns	1
Column type frequency:	

Table 9: Data summary

numeric	1
Group variables	None

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Calories	0	1	2303.61	718.17	0	1828.5	2134	2793.25	4900

#Data Editing

My aim was to join the data by IDs and Date in an attempt to match people's data across data frames where able. Since there was a bit of variance in terms of times and I wasn't quite sure if I could have easily joined them through posixct, I simply created new columns, separated date and time.

I had hoped to utilize other joined data frames however the sample size and inconsistent samples led to small, uninteresting, and likely misleading data frames.

```
#Separate Datetime to Date and Time

#Date Time - Posix*
sleepDay$Date_Time_Cleaned = strptime(sleepDay$SleepDay,"%m/%d/%Y %I:%M:%S %p",tz="UTC")
hourlyCalories$Date_Time_Cleaned = strptime(hourlyCalories$ActivityHour,"%m/%d/%Y %I:%M:%S %p",tz="UTC"
hourlyIntensities$Date_Time_Cleaned = strptime(hourlyIntensities$ActivityHour,"%m/%d/%Y %I:%M:%S %p",tz
weightLog$Date_Time_Cleaned = strptime(weightLog$Date,"%m/%d/%Y %I:%M:%S %p",tz="UTC")


#Date Only - Posix*
sleepDay$Date_Cleaned = strptime(sleepDay$SleepDay,"%m/%d/%Y")
hourlyCalories$Date_Cleaned = strptime(hourlyCalories$ActivityHour,"%m/%d/%Y")
hourlyIntensities$Date_Cleaned = strptime(hourlyIntensities$ActivityHour,"%m/%d/%Y")
weightLog$Date_Cleaned = strptime(weightLog$Date,"%m/%d/%Y")

dailyActivity$Date_Cleaned = strptime(dailyActivity$ActivityDate,"%m/%d/%Y")
dailyIntensities$Date_Cleaned = strptime(dailyIntensities$ActivityDay,"%m/%d/%Y")
dailySteps$Date_Cleaned = strptime(dailySteps$ActivityDay,"%m/%d/%Y")
dailyCalories$Date_Cleaned = strptime(dailyCalories$ActivityDay,"%m/%d/%Y")

#Time Only - Characters
sleepDay$Time_Cleaned = strftime(sleepDay$Date_Time_Cleaned,"%H:%M:%S")
hourlyCalories$Time_Cleaned = strftime(hourlyCalories$Date_Time_Cleaned,"%H:%M:%S")
hourlyIntensities$Time_Cleaned = strftime(hourlyIntensities$Date_Time_Cleaned,"%H:%M:%S")
weightLog$Time_Cleaned = strftime(weightLog$Date_Time_Cleaned,"%H:%M:%S")


#Joins/Merge

#Activity and Sleep
Activity_Day_Sleep_Day = left_join(dailyActivity,sleepDay,by=c("Id","Date_Cleaned"))
```

```
#Adding Columns for Analysis

#Time Awake in Bed
Activity_Day_Sleep_Day$SleepLatency = Activity_Day_Sleep_Day$TotalTimeInBed - Activity_Day_Sleep_Day$To

#Conversion Integer to Numeric - For Correlations
Activity_Day_Sleep_Day[c(3,4,5,6,7,8,9,10,11,12,13,14,15,19,20,23)] = sapply(Activity_Day_Sleep_Day[c(3
```

#Exploring Sleep & Exercise

Since the other data frames left a lot to be desired in ability to provide insight, especially when joining IDs across data tables, I decided to focus on the larger sample between exercise and sleep.

My first decision was to graph and check each ID for their standard deviation and mean, in an effort to check for and find out there seemed to be some notably higher variances and/or means by some individuals. The data set wasn't as large as I would've liked, but I chose to focus on the core parts of the data.

```
# Examining & Graphing Variables



#Sleep Variable
Activity_Day_Sleep_Day %>%
  select(TotalSleepRecords,
         TotalMinutesAsleep,
         TotalTimeInBed,
         SleepLatency
  ) %>%
  skim_without_charts()
```

Table 11: Data summary

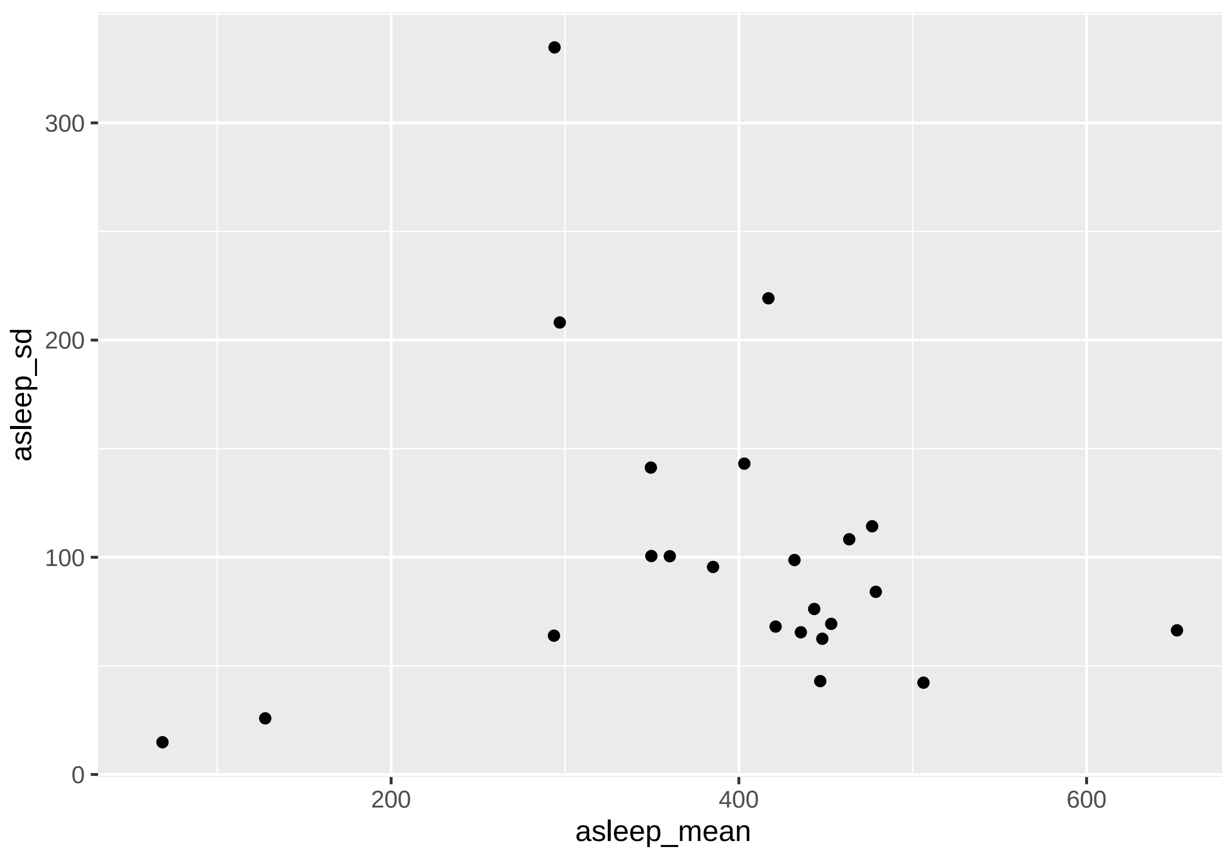
Name	Piped data
Number of rows	943
Number of columns	4
Column type frequency:	
numeric	4
Group variables	None

Variable type: numeric

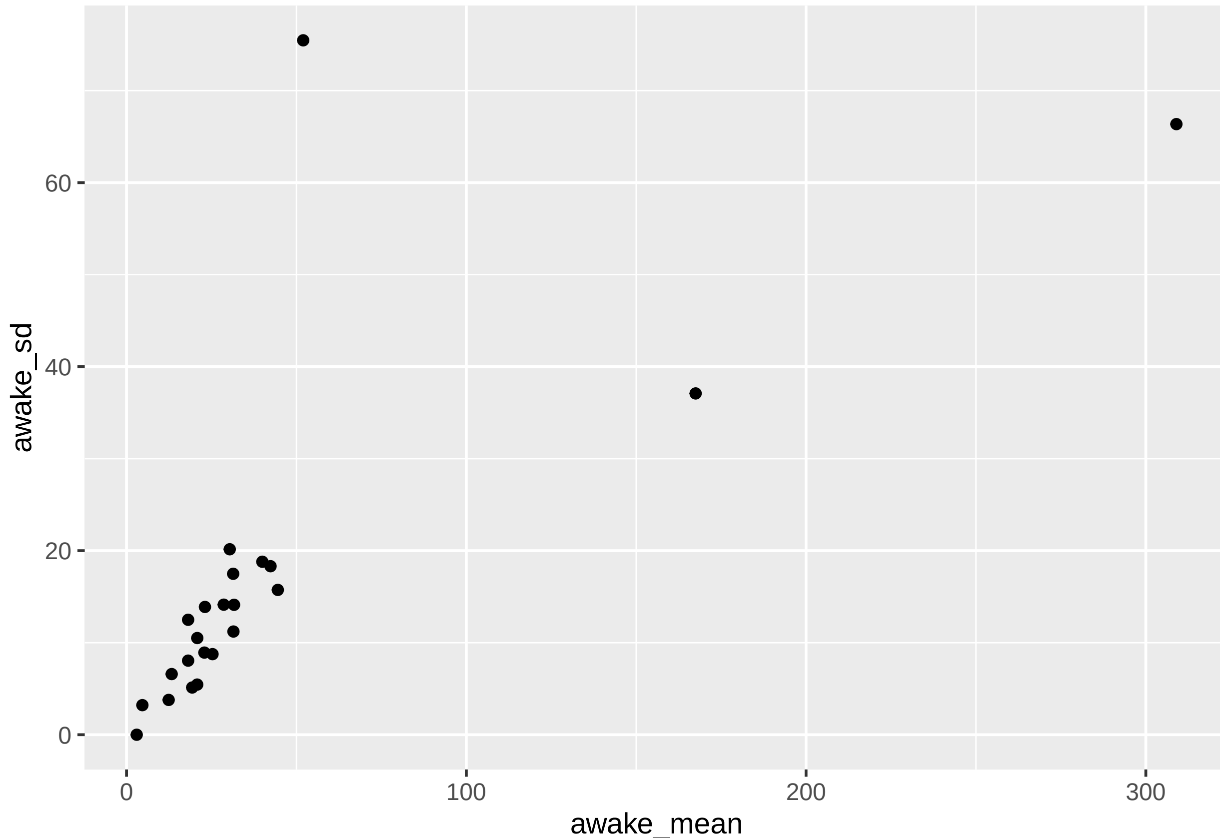
skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
TotalSleepRecords	530	0.44	1.12	0.35	1	1	1	1	3
TotalMinutesAsleep	530	0.44	419.47	118.34	58	361	433	490	796
TotalTimeInBed	530	0.44	458.64	127.10	61	403	463	526	961
SleepLatency	530	0.44	39.17	46.57	0	17	25	40	371

```
#Individual Means and Deviations - Some Notable Outliers
#Some IDs have no SD - Because only one observed sleep data.
#Notably, two individuals had substantially higher awake means,
```

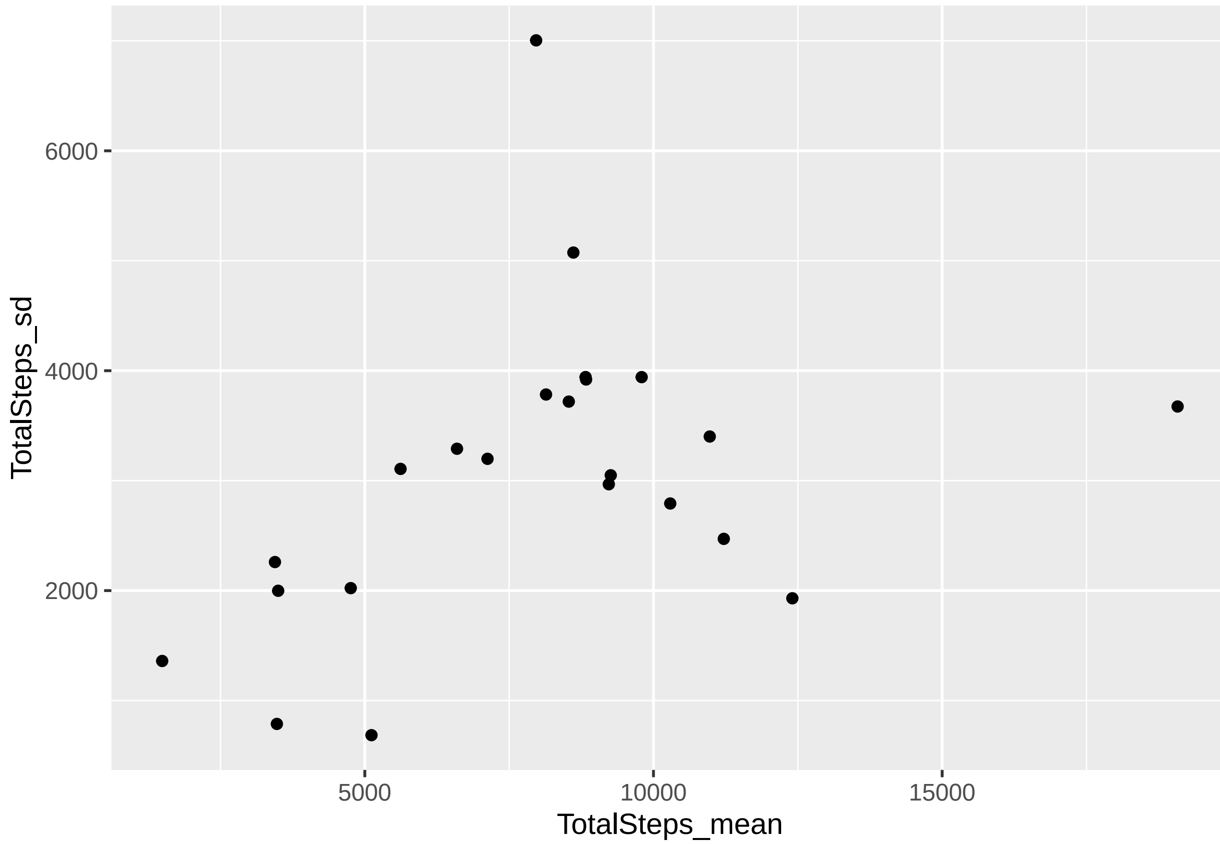
```r
#2 others had substantially higher awake standard deviations

Sleep_Awake_Mean_SD =
  Activity_Day_Sleep_Day %>%
  group_by(Id) %>%
  drop_na() %>%
  mutate(awake_mean=mean(SleepLatency,na.rm=TRUE),
         awake_sd = sd(SleepLatency,na.rm=TRUE),
         asleep_mean=mean(TotalMinutesAsleep,na.rm=TRUE),
         asleep_sd = sd(TotalMinutesAsleep,TRUE),
         .keep="used") %>%
  filter(!duplicated(Id))

#Visualization of Outliers around cluster of more normal means and deviations
ggplot(Sleep_Awake_Mean_SD,aes(x=asleep_mean,y=asleep_sd)) + geom_point()
```



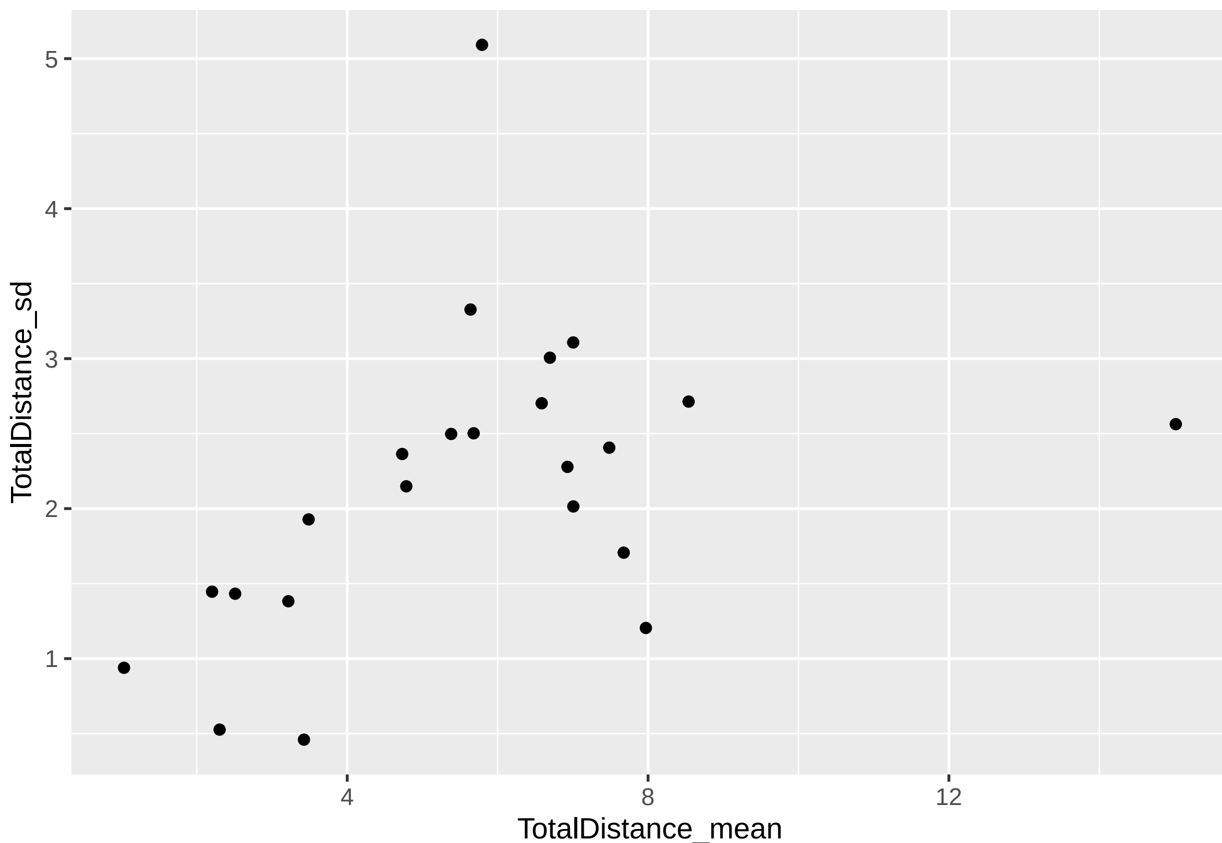

```r
ggplot(Sleep_Awake_Mean_SD,aes(x=awake_mean,y=awake_sd)) + geom_point()
```



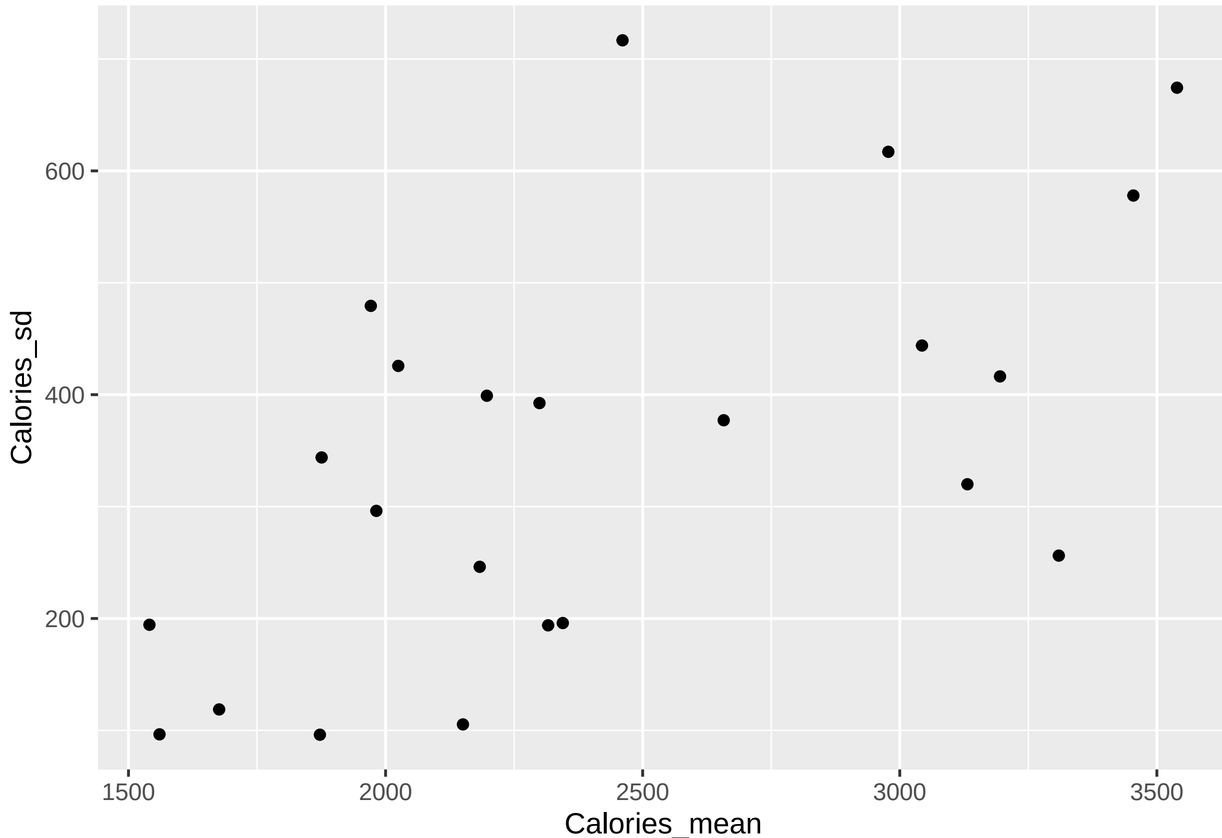

```
#Exercise/Activity Variable
Exercise_Sedentary_Mean_SD =
  Activity_Day_Sleep_Day %>%
  group_by(Id) %>%
  drop_na() %>%
  mutate(TotalSteps_mean=mean(TotalSteps,na.rm=TRUE),
         TotalSteps_sd = sd(TotalSteps,na.rm=TRUE),
         TotalDistance_mean=mean(TotalDistance,na.rm=TRUE),
         TotalDistance_sd = sd(TotalDistance,TRUE),
         Calories_mean=mean(Calories,na.rm=TRUE),
         Calories_sd = sd(Calories,na.rm=TRUE),
         SedentaryMinutes_mean=mean(SedentaryMinutes,na.rm=TRUE),
         SedentaryMinutes_sd = sd(SedentaryMinutes,TRUE),
         .keep="used") %>%
  filter(!duplicated(Id))

ggplot(Exercise_Sedentary_Mean_SD,aes(x=TotalSteps_mean,y=TotalSteps_sd)) + geom_point()
```



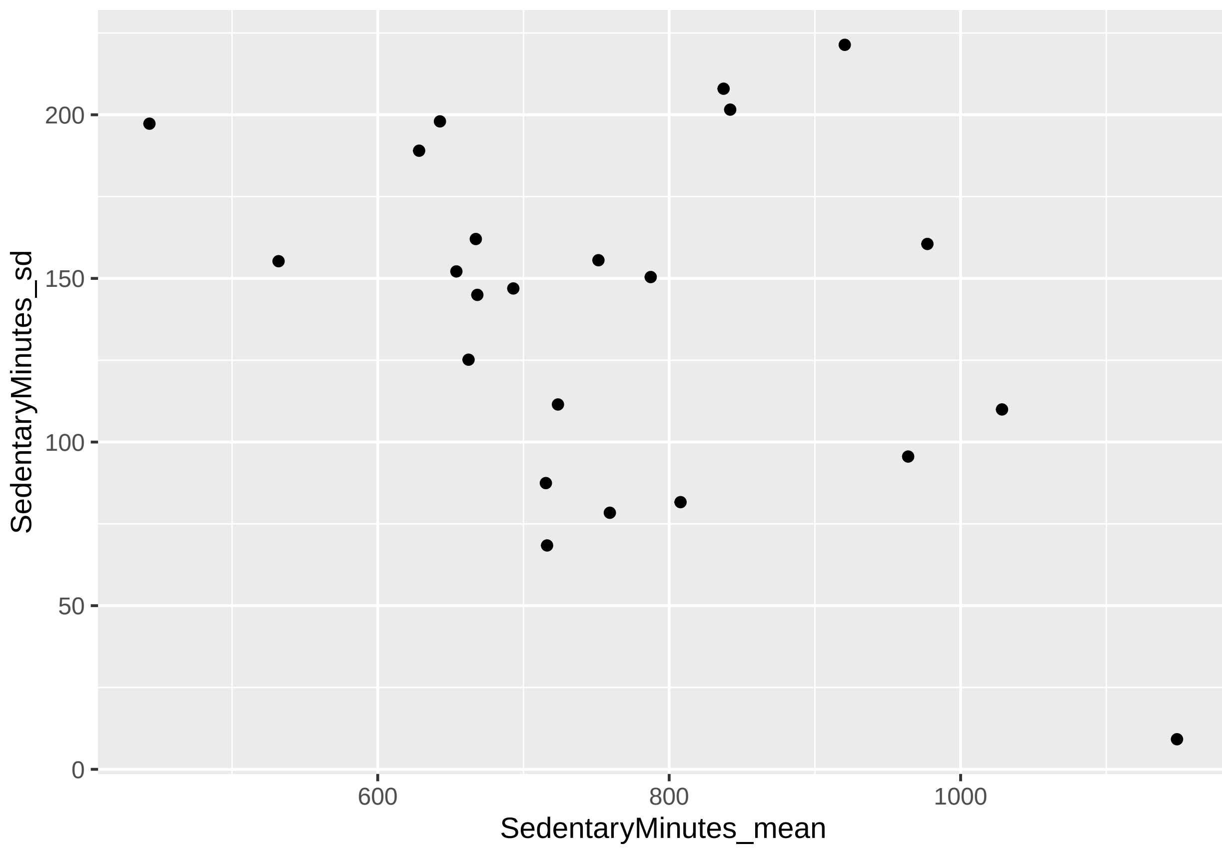

```
ggplot(Exercise_Sedentary_Mean_SD,aes(x=TotalDistance_mean,y=TotalDistance_sd)) + geom_point()
```



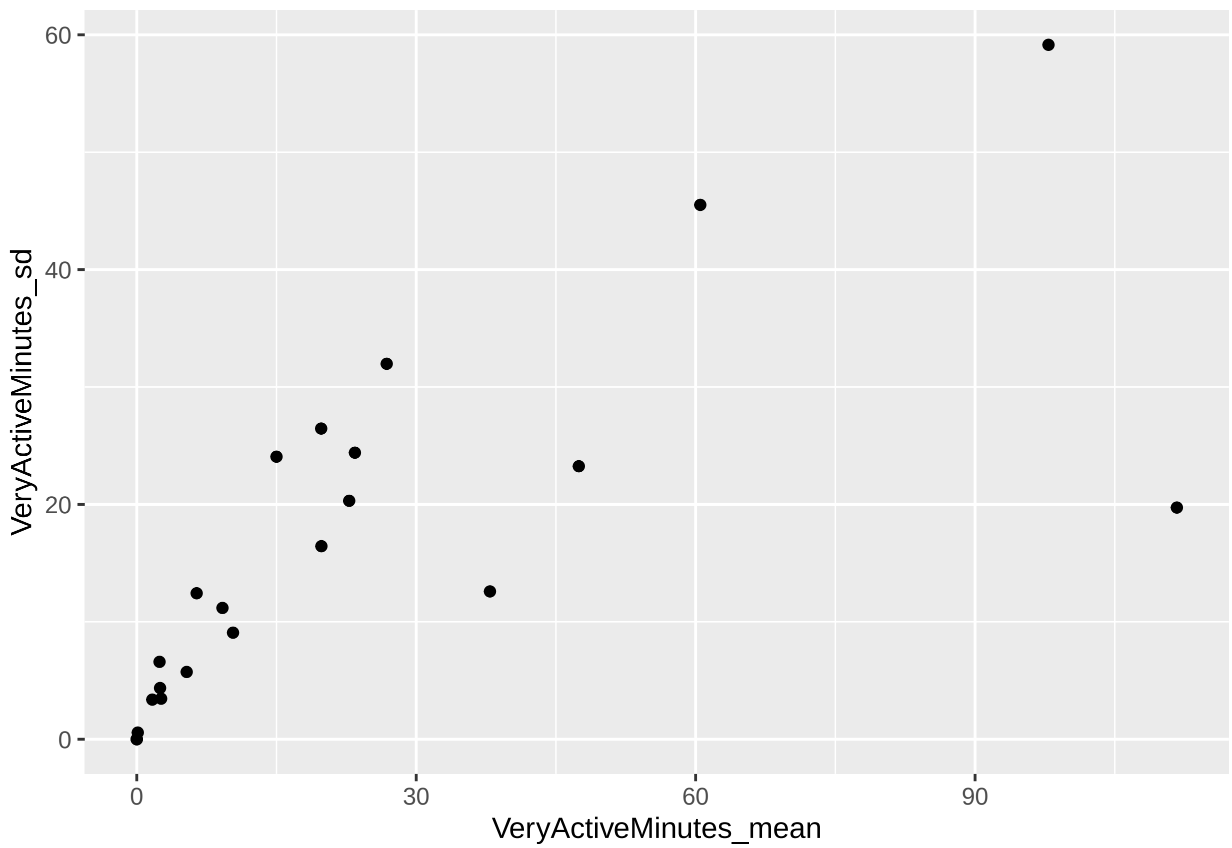

```
ggplot(Exercise_Sedentary_Mean_SD,aes(x=Calories_mean,y=Calories_sd)) + geom_point()
```



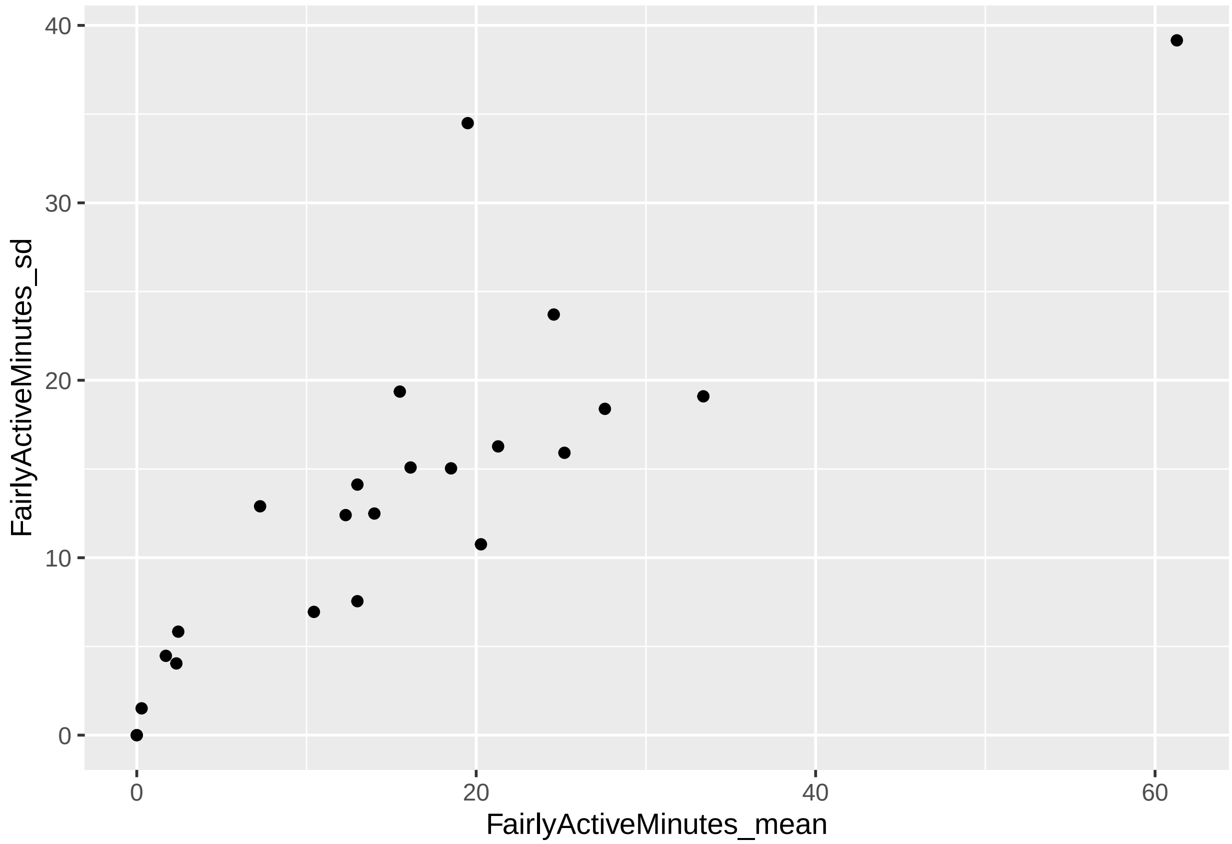

```
ggplot(Exercise_Sedentary_Mean_SD,aes(x=SedentaryMinutes_mean,y=SedentaryMinutes_sd)) + geom_point()
```



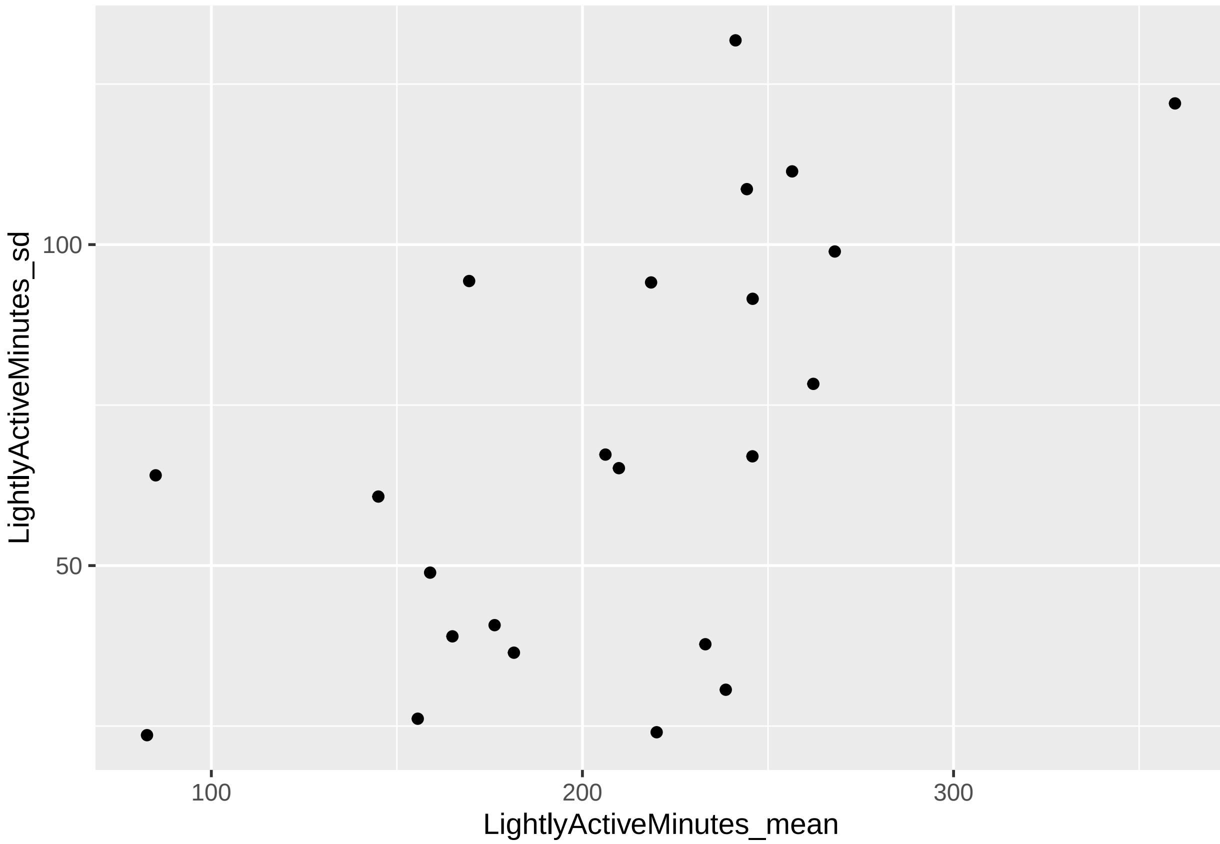

```
#Minutes - VeryActiveMinutes, FairlyActiveMinutes,LightlyActiveMinutes,SedentaryMinutes

Exercise_Minutes_Mean_SD =
  Activity_Day_Sleep_Day %>%
  group_by(Id) %>%
  drop_na() %>%
  mutate(VeryActiveMinutes_mean=mean(VeryActiveMinutes,na.rm=TRUE),
         VeryActiveMinutes_sd = sd(VeryActiveMinutes,na.rm=TRUE),
         FairlyActiveMinutes_mean=mean(FairlyActiveMinutes,na.rm=TRUE),
         FairlyActiveMinutes_sd = sd(FairlyActiveMinutes,TRUE),
         LightlyActiveMinutes_mean=mean(LightlyActiveMinutes,na.rm=TRUE),
         LightlyActiveMinutes_sd = sd(LightlyActiveMinutes,na.rm=TRUE),
         SedentaryMinutes_mean=mean(SedentaryMinutes,na.rm=TRUE),
         SedentaryMinutes_sd = sd(SedentaryMinutes,TRUE),
         .keep="used") %>%
  filter(!duplicated(Id))

ggplot(Exercise_Minutes_Mean_SD,aes(x=VeryActiveMinutes_mean,y=VeryActiveMinutes_sd)) + geom_point()
```



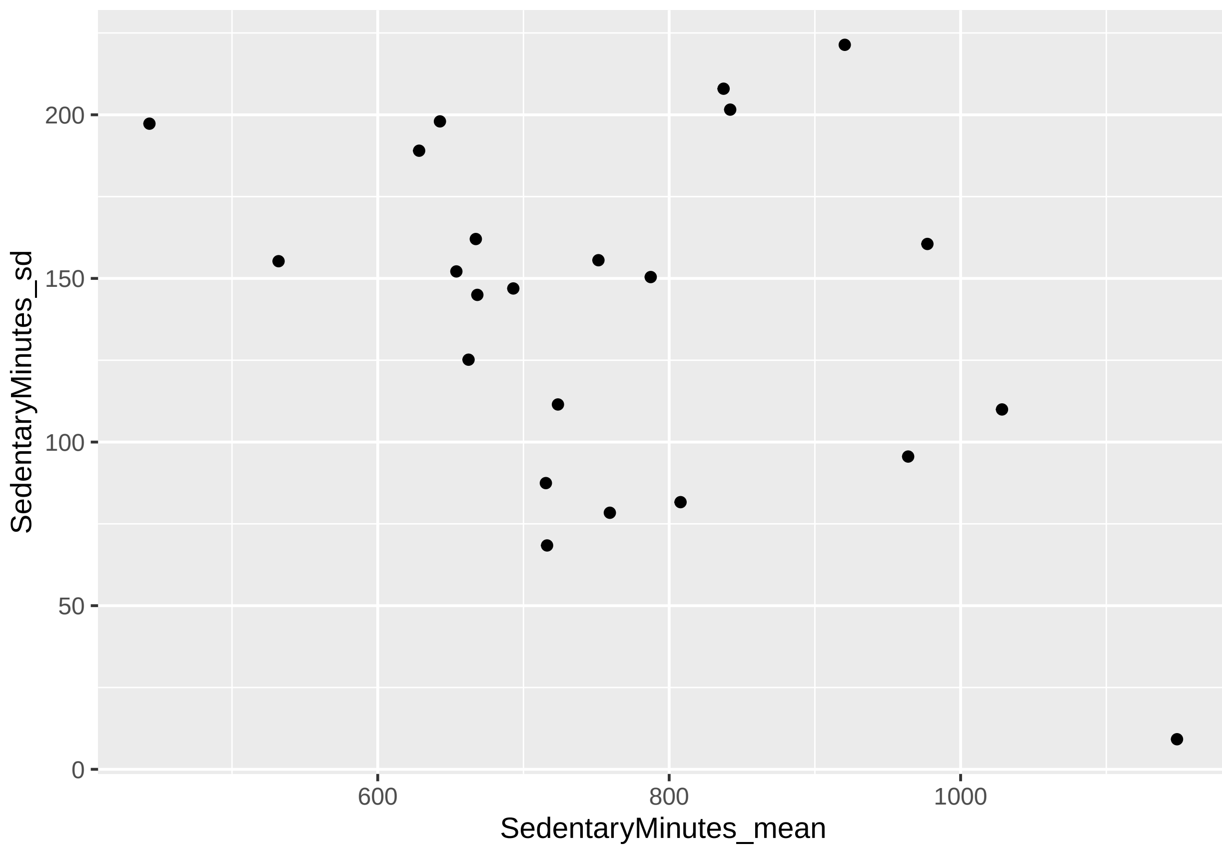

```
ggplot(Exercise_Minutes_Mean_SD,aes(x=FairlyActiveMinutes_mean,y=FairlyActiveMinutes_sd)) + geom_point()
```




```
ggplot(Exercise_Minutes_Mean_SD,aes(x=LightlyActiveMinutes_mean,y=LightlyActiveMinutes_sd)) + geom_poin
```




```
ggplot(Exercise_Minutes_Mean_SD,aes(x=SedentaryMinutes_mean,y=SedentaryMinutes_sd)) + geom_point()
```



Graphs and Correlation

After examining the selected data closer and getting a better read on it, I examined the data for potential correlations, pairing variables with one another. I started with an obvious correlation, Steps and Calories, which obviously had a high level since more exercise correlates with more calories burned. Likewise, Distance and calories had a similar result.

After that, I graphed sleep against sedentary minutes, which turned out to be a decently high negative correlation. People may either be feeling a lack of energy from sleep which translates to less motivation to less exercise and more sedentary time, or more sedentary time may lead be leading to less sleep. I also verified that sedentary time was not factored into time asleep.

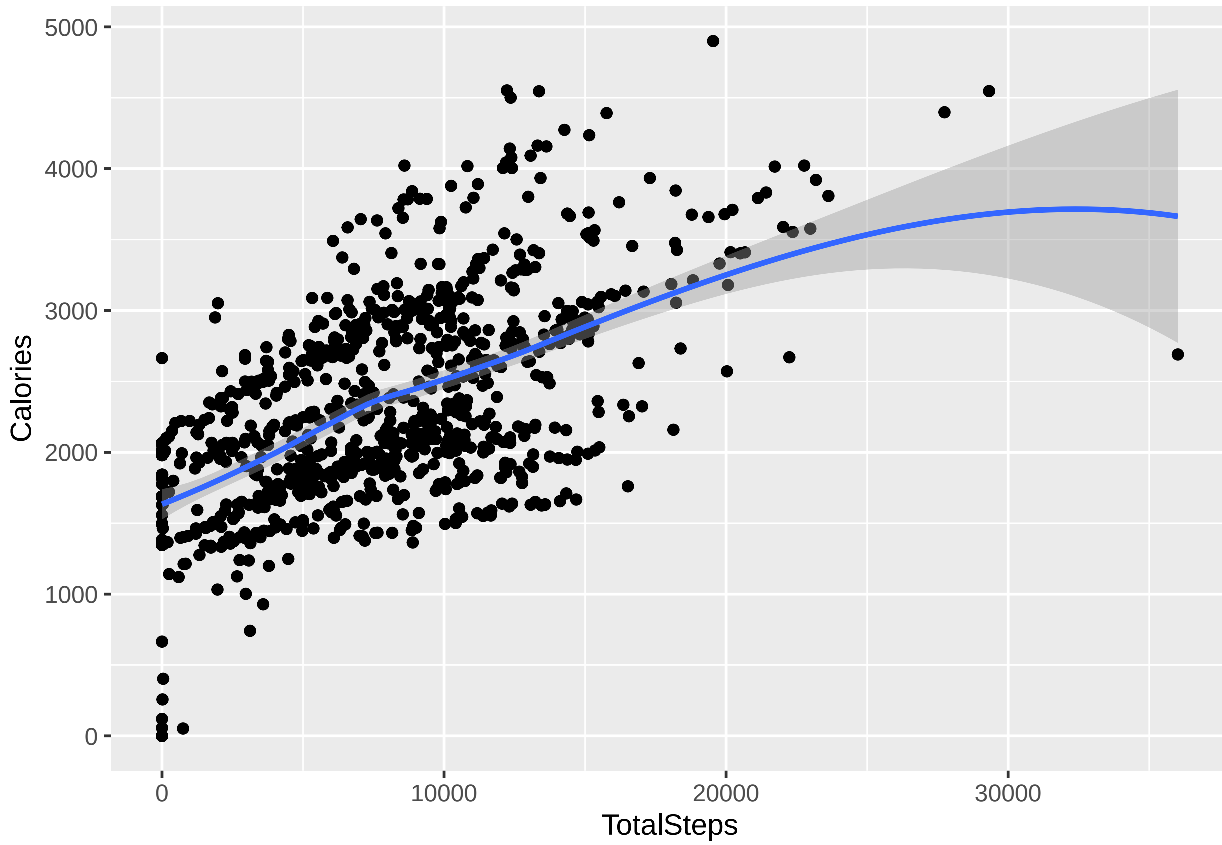
Graphing sleep latency against calories burned tells an interesting story, about how a fair amount of data points who only burned 1500 calories had notably high sleep latency. This might be due to an excess of energy, not burning enough calories throughout the day.

Next, there was a small negative correlation between total distance and time asleep and in bed. This may be due to more exercise causing better quality rest, and thus a lower amount of sleep. At worst, it means people are sacrificing time sleeping to exercise.

Last, there was graphs between sleep latency against distance and sedentary minutes, to try to better understand if either more exercise or less would predict sleep latency. However, both variables had a negative relationship with sleep latency, suggesting either a larger sample size or, more likely, the fact that sleep latency is likely to be predicted by a multitude of different variables, and might not be easily examined.

```
#Investigating Correlations alongside Graphs

#High Correlation between Steps and Calories Burned
ggplot(data=Activity_Day_Sleep_Day,aes(x=TotalSteps,y=Calories)) + geom_point() + geom_smooth()
```

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

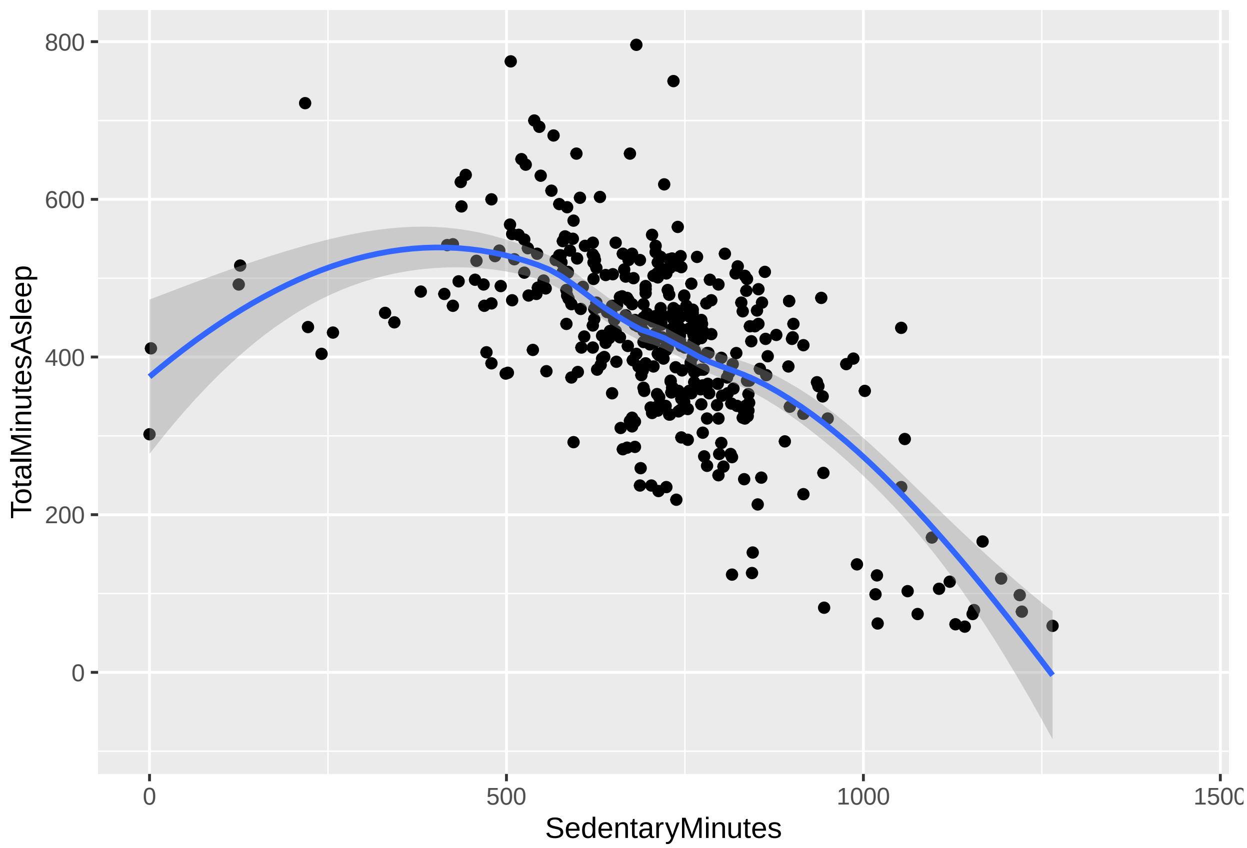



```
cor(Activity_Day_Sleep_Day$TotalSteps,Activity_Day_Sleep_Day$Calories,use="complete.obs")
```

```
## [1] 0.5929493
```

```
#High Correlation - Less Sleep - More Sedentary Minutes, but also some outliers with 0 Sedentary Minute
ggplot(data=Activity_Day_Sleep_Day, aes(x=SedentaryMinutes, y=TotalMinutesAsleep)) + geom_point() + geo
```

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

## Warning: Removed 530 rows containing non-finite values (stat_smooth).

## Warning: Removed 530 rows containing missing values (geom_point).
```

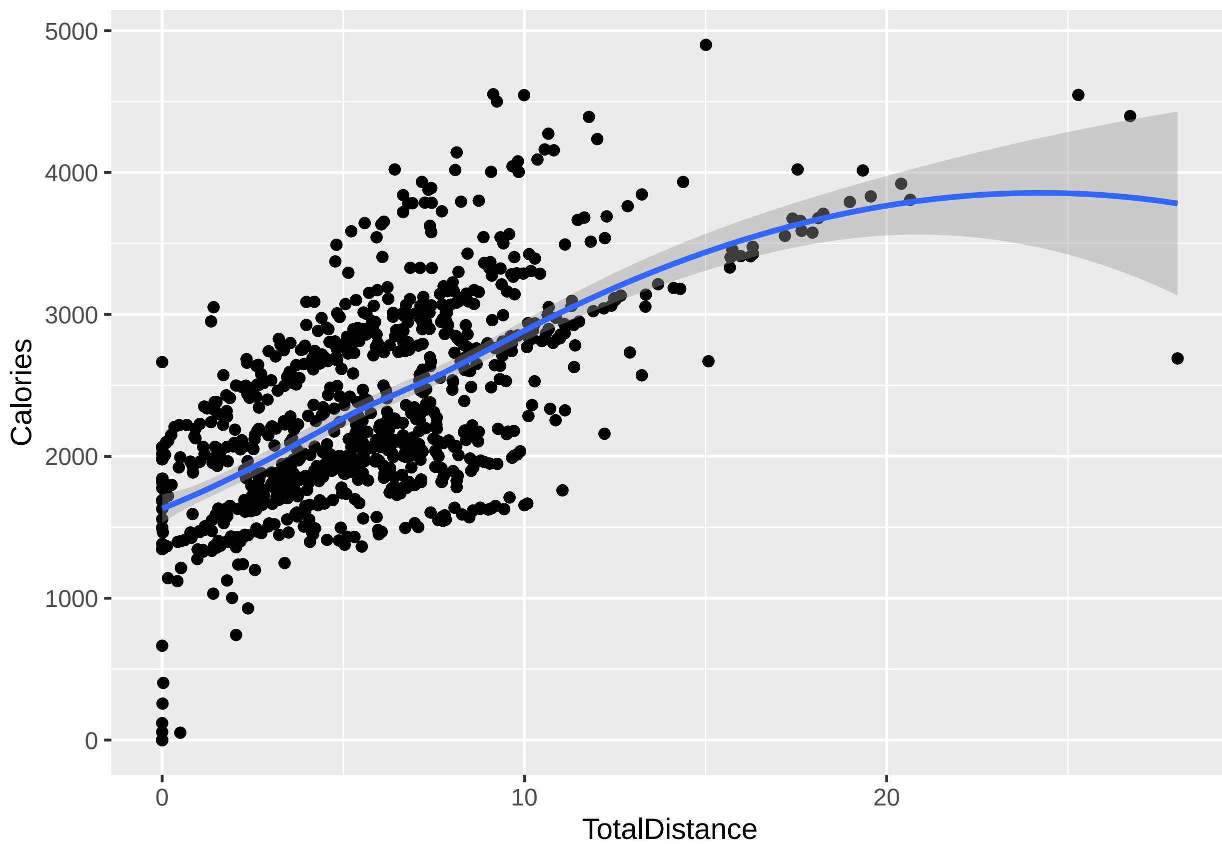



```
cor(Activity_Day_Sleep_Day$SedentaryMinutes,Activity_Day_Sleep_Day$TotalMinutesAsleep,use="complete.obs
```

```
## [1] -0.599394
```

```
#High (Obvious) Correlation More Distance = More Calories Burned
ggplot(data=Activity_Day_Sleep_Day, aes(x=TotalDistance, y=Calories)) + geom_point() + geom_smooth()
```

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



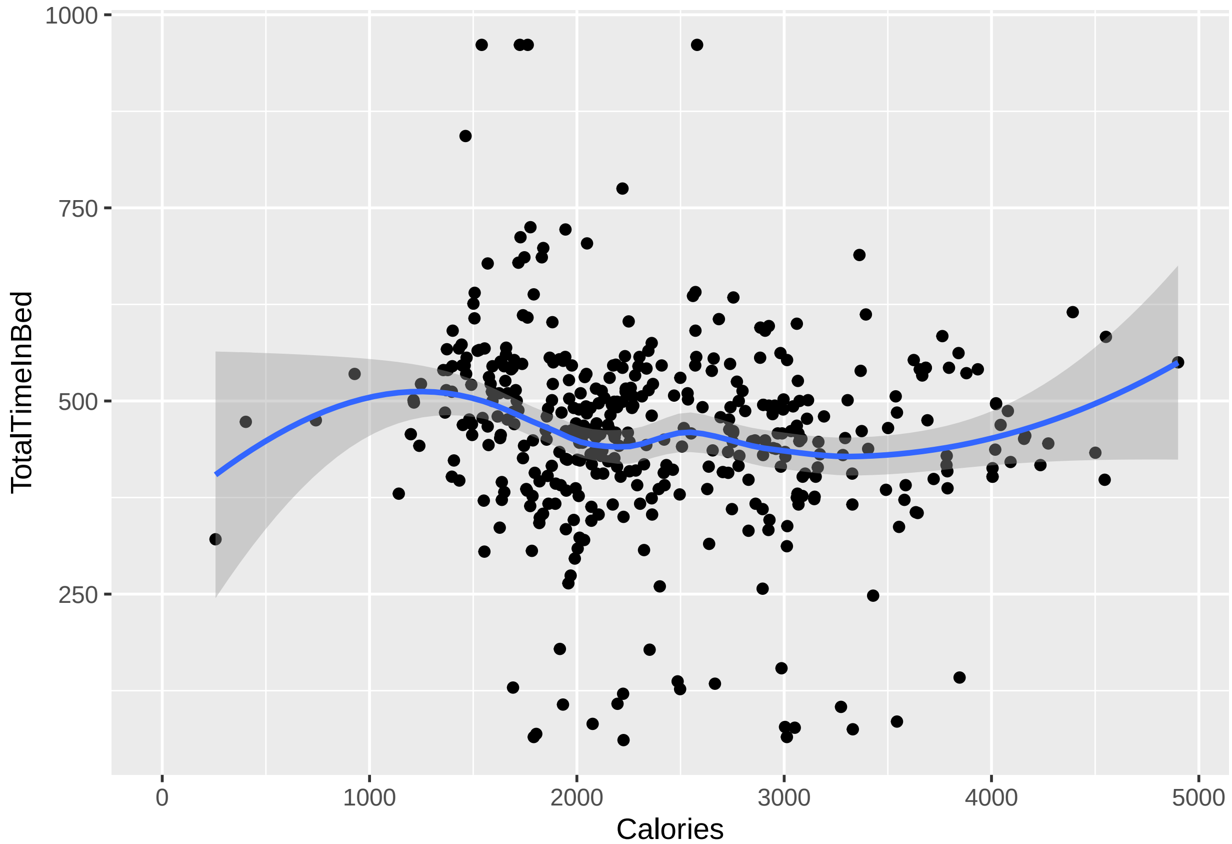

```
cor(Activity_Day_Sleep_Day$TotalDistance,Activity_Day_Sleep_Day$Calories,use="complete.obs")
```

```
## [1] 0.6466023
```

```
#Low Negative Correlation - Calories and Total Time in Bed
ggplot(data=Activity_Day_Sleep_Day,aes(x=Calories,y=TotalTimeInBed)) + geom_point() + geom_smooth()
```

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

```
## Warning: Removed 530 rows containing non-finite values (stat_smooth).
## Removed 530 rows containing missing values (geom_point).
```



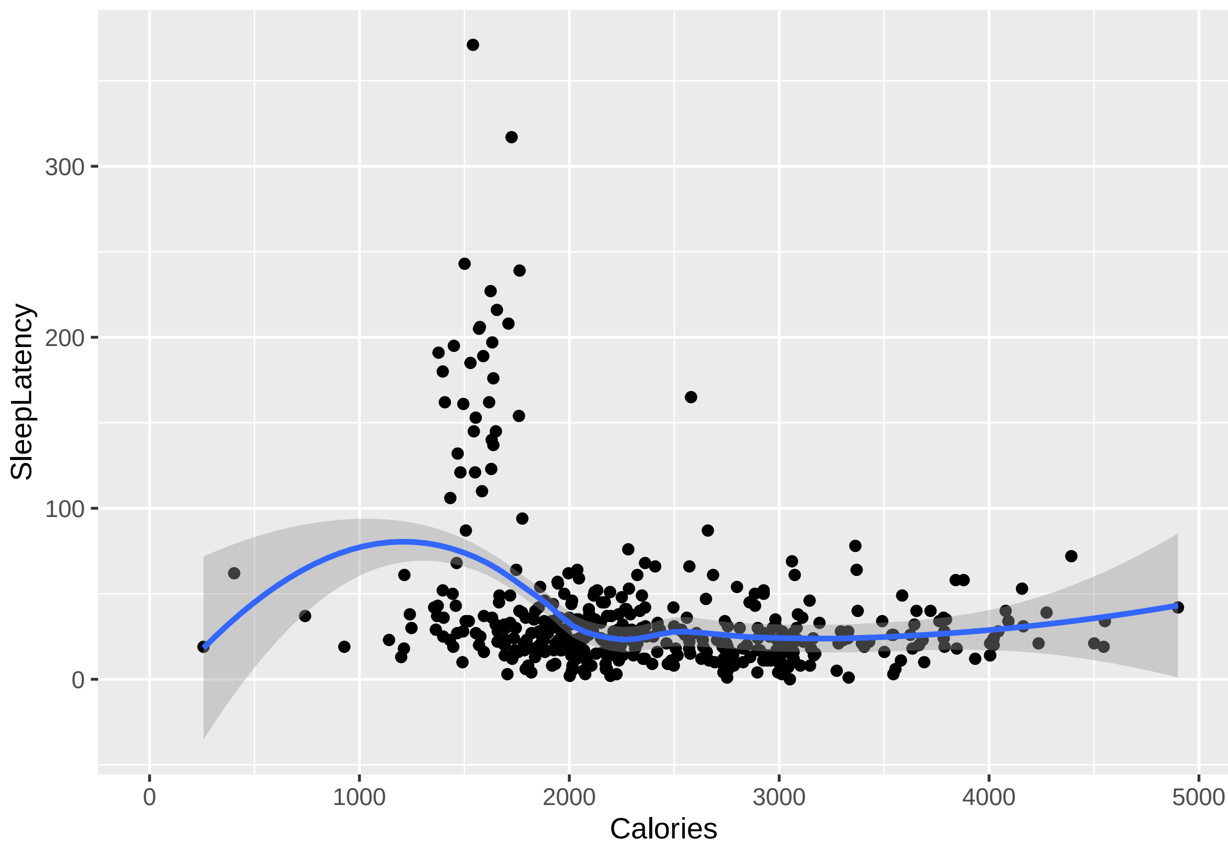

```
cor(Activity_Day_Sleep_Day$Calories,Activity_Day_Sleep_Day$TotalTimeInBed,use="complete.obs")
```

```
## [1] -0.1325071
```

```
#Low Negative Correlation - Sleep Latency might be caused by not enough calories burned for excess ener
ggplot(data=Activity_Day_Sleep_Day, aes(x=Calories, y=SleepLatency)) + geom_point() + geom_smooth()
```

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

```
## Warning: Removed 530 rows containing non-finite values (stat_smooth).
## Removed 530 rows containing missing values (geom_point).
```

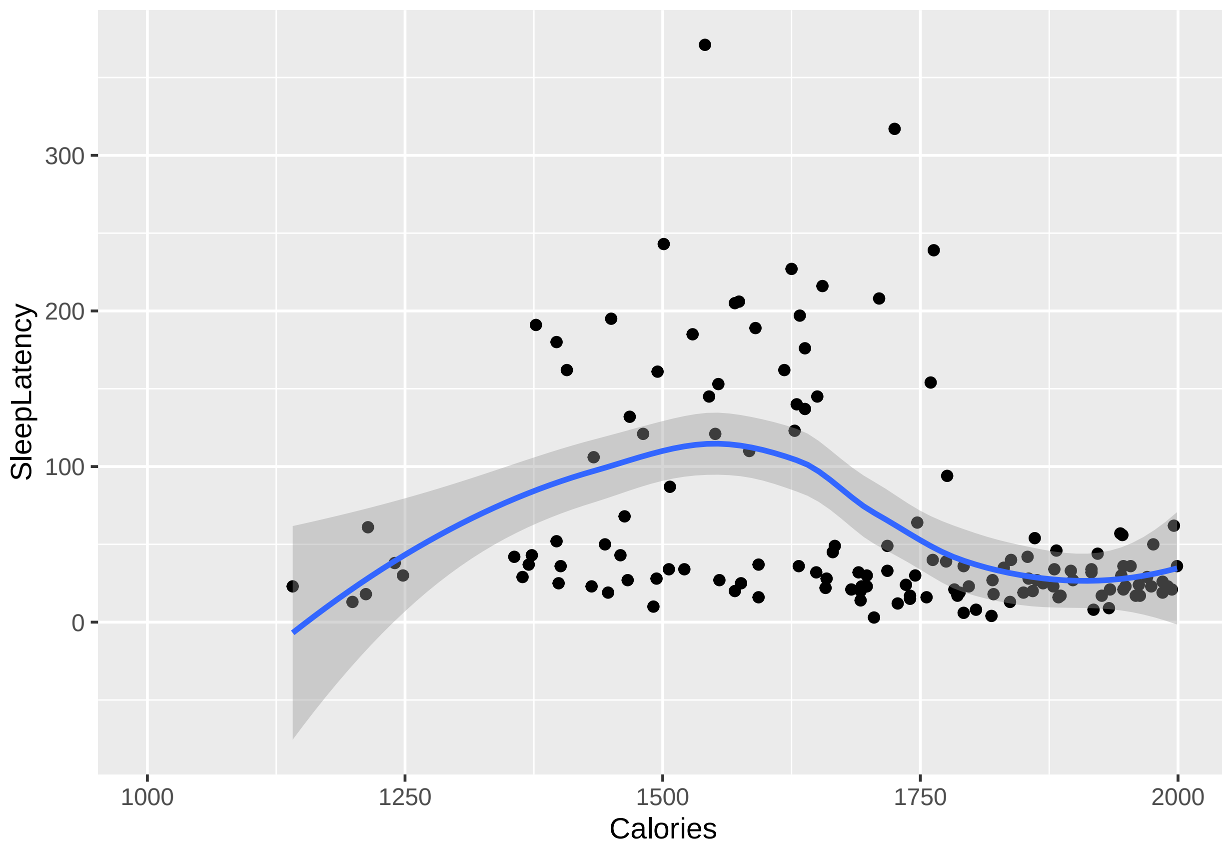



```
cor(Activity_Day_Sleep_Day$Calories,Activity_Day_Sleep_Day$SleepLatency,use="complete.obs")
```

```
## [1] -0.2891555
```

```
#What if we look at the range from 1000-2000 Calories?
Activity_Day_Sleep_Day %>%
  filter(between(Calories,1000,2000)) %>%
  ggplot(aes(x=Calories,y=SleepLatency)) + geom_point() + geom_smooth()
```

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

## Warning: Removed 218 rows containing non-finite values (stat_smooth).

## Warning: Removed 218 rows containing missing values (geom_point).
```




```
Activity_Day_Sleep_Day %>%
  filter(between(Calories,1000,2000)) %>%
  summarise(cor(Calories, SleepLatency, use = "complete.obs"))
```

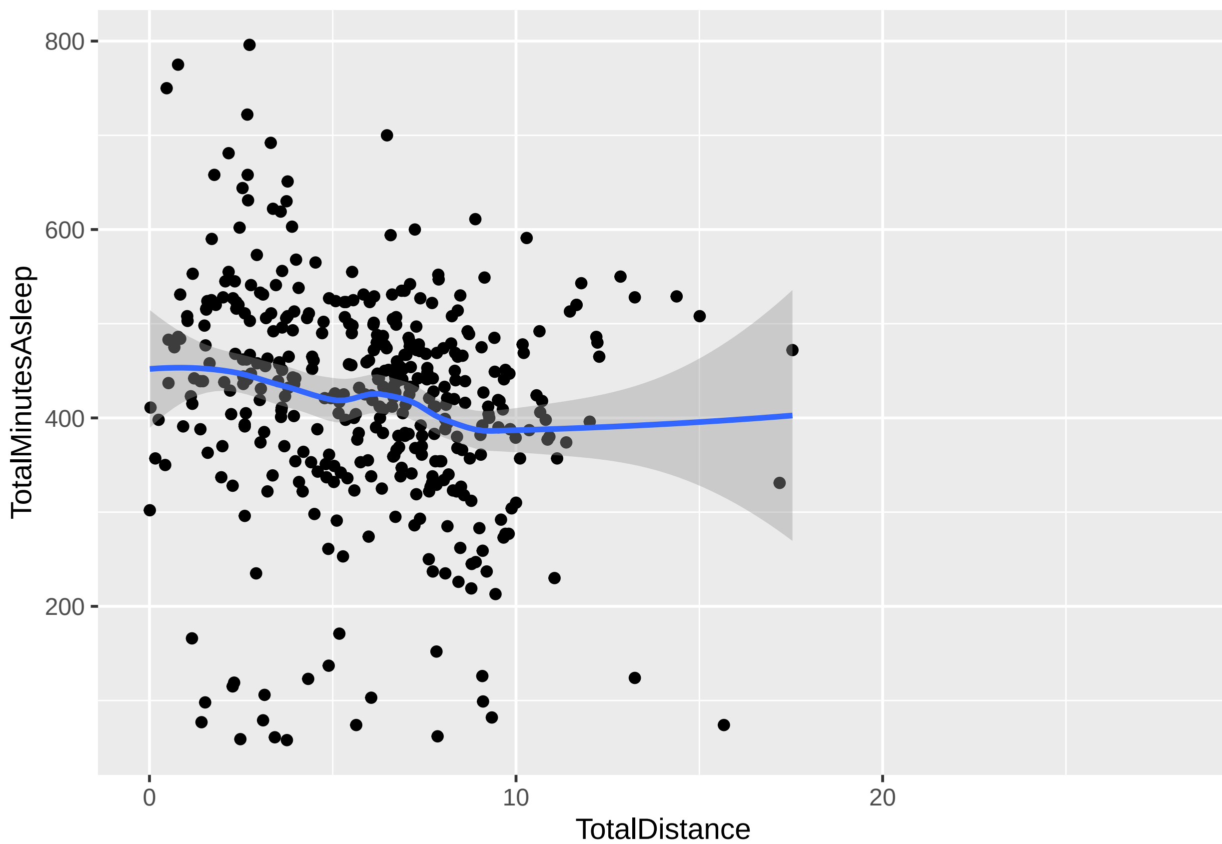
```
##   cor(Calories, SleepLatency, use = "complete.obs")
## 1                                         -0.277088
```

```
#More Exercise and Less Sedentary Minutes = Less Minutes Asleep (Possibly more rested or better quality
ggplot(data=Activity_Day_Sleep_Day, aes(x=TotalDistance, y=TotalMinutesAsleep)) + geom_point() + geom_sm
```

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

```
## Warning: Removed 530 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 530 rows containing missing values (geom_point).
```



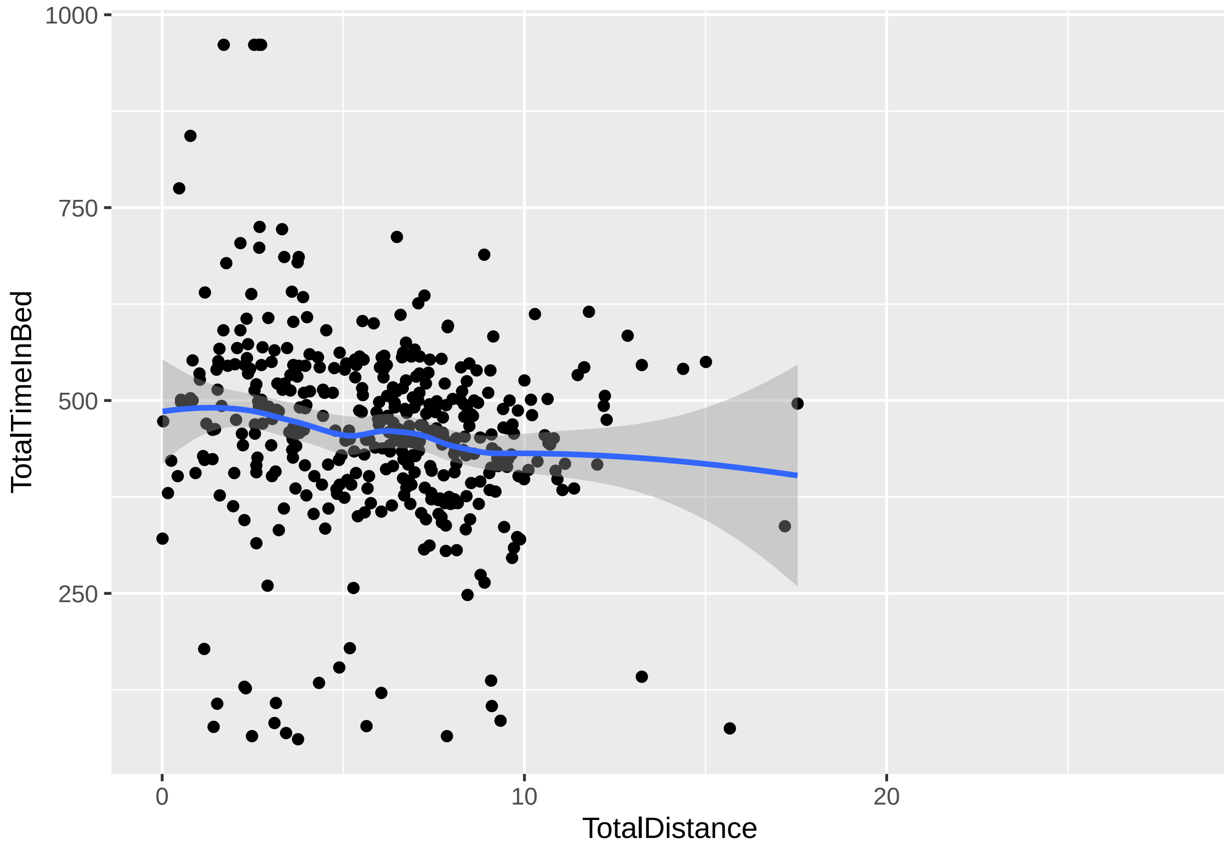

```
cor(Activity_Day_Sleep_Day$TotalDistance,Activity_Day_Sleep_Day$TotalMinutesAsleep,use="complete.obs")
```

```
## [1] -0.1721427
```

```
ggplot(data=Activity_Day_Sleep_Day, aes(x=TotalDistance, y=TotalTimeInBed)) + geom_point() + geom_smoot
```

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

```
## Warning: Removed 530 rows containing non-finite values (stat_smooth).
## Removed 530 rows containing missing values (geom_point).
```



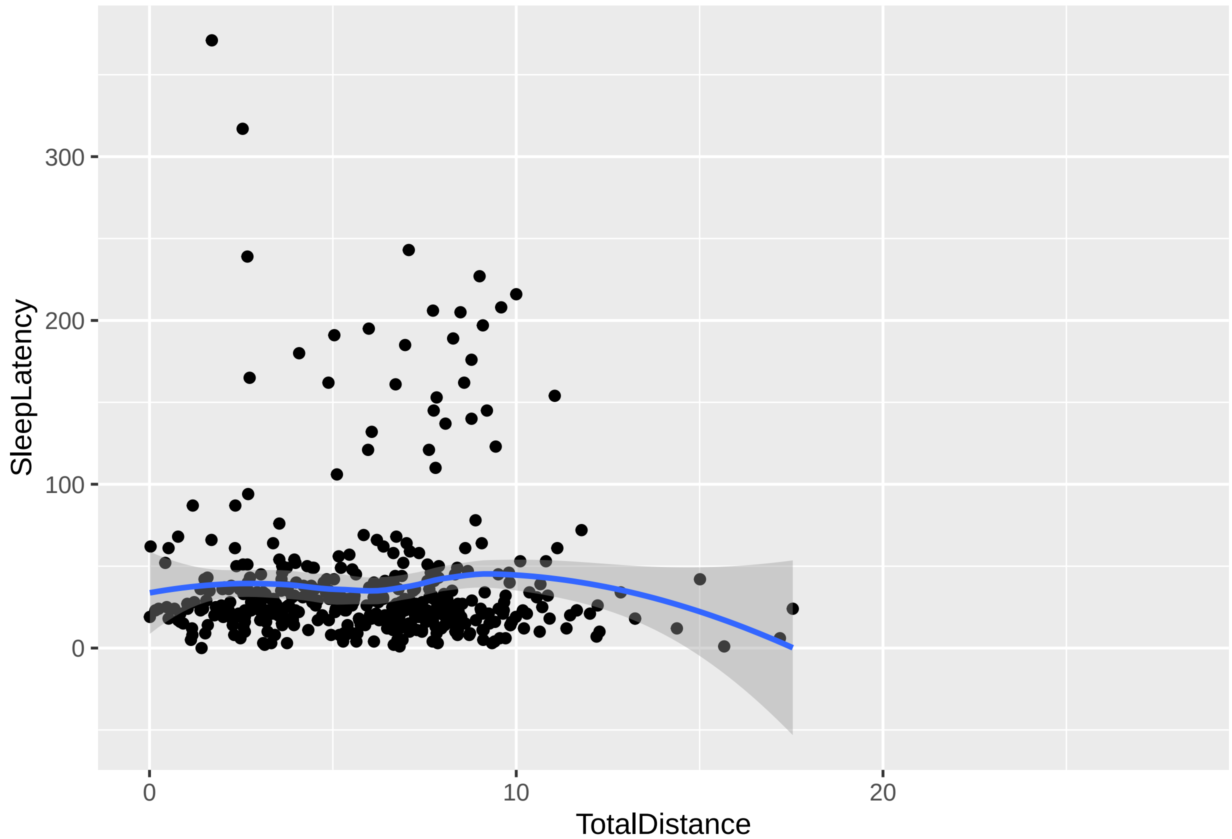

```
cor(Activity_Day_Sleep_Day$TotalDistance,Activity_Day_Sleep_Day$TotalTimeInBed,use="complete.obs")
```

```
## [1] -0.1580949
```

```
#Does More Exercise Lower Sleep Latency? Not such a clear relationship, possibly too many external vari
ggplot(data=Activity_Day_Sleep_Day, aes(x=TotalDistance, y=SleepLatency)) + geom_point() + geom_smooth(
```

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

```
## Warning: Removed 530 rows containing non-finite values (stat_smooth).
## Removed 530 rows containing missing values (geom_point).
```



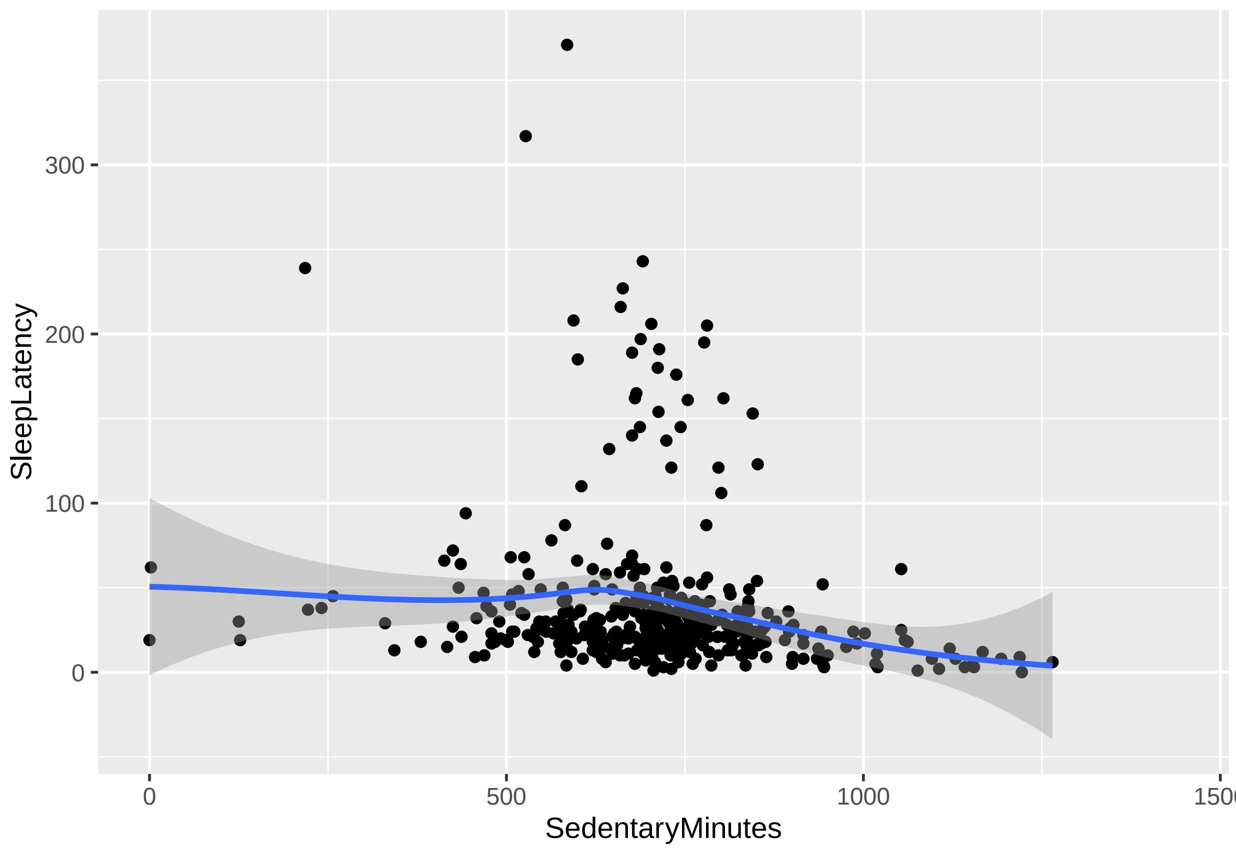

```
cor(Activity_Day_Sleep_Day$Calories,Activity_Day_Sleep_Day$TotalTimeInBed,use="complete.obs")
```

```
## [1] -0.1325071
```

```
ggplot(data=Activity_Day_Sleep_Day, aes(x=SedentaryMinutes, y=SleepLatency)) + geom_point() + geom_smoo
```

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

```
## Warning: Removed 530 rows containing non-finite values (stat_smooth).
## Removed 530 rows containing missing values (geom_point).
```




```
cor(Activity_Day_Sleep_Day$Calories,Activity_Day_Sleep_Day$TotalTimeInBed,use="complete.obs")
```

```
## [1] -0.1325071
```

Low Correlations

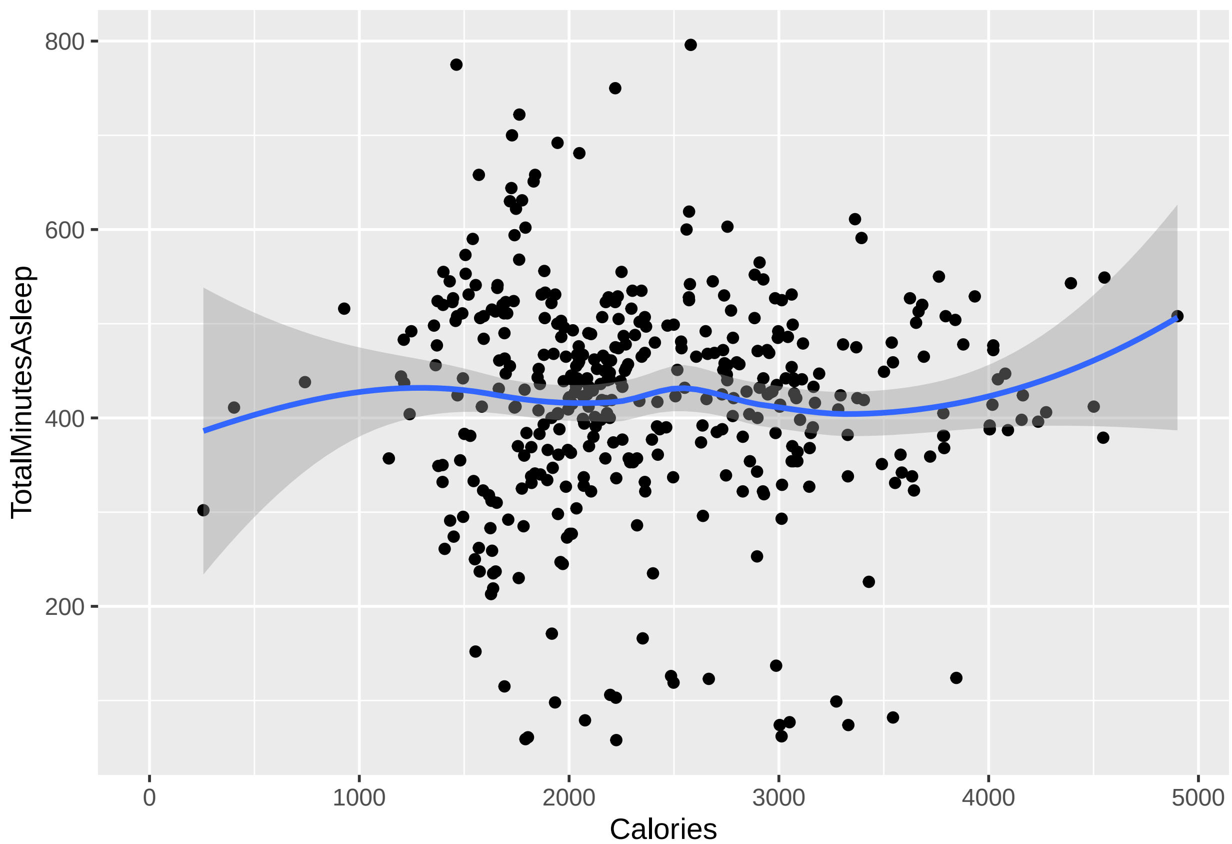
These are a few of the other graphs that I created that held little to no correlation, and as such I was unable to draw any conclusions from. However, they may inspire other graphs or further data examination.

```
#Low Correlation Data

#Calories Burned vs Sleep -
#Maybe having too much energy from not burning enough for some people?
#No clear relationship between Minutes Asleep and Calories burned here
ggplot(data=Activity_Day_Sleep_Day, aes(x=Calories, y=TotalMinutesAsleep)) + geom_point() + geom_smooth
```

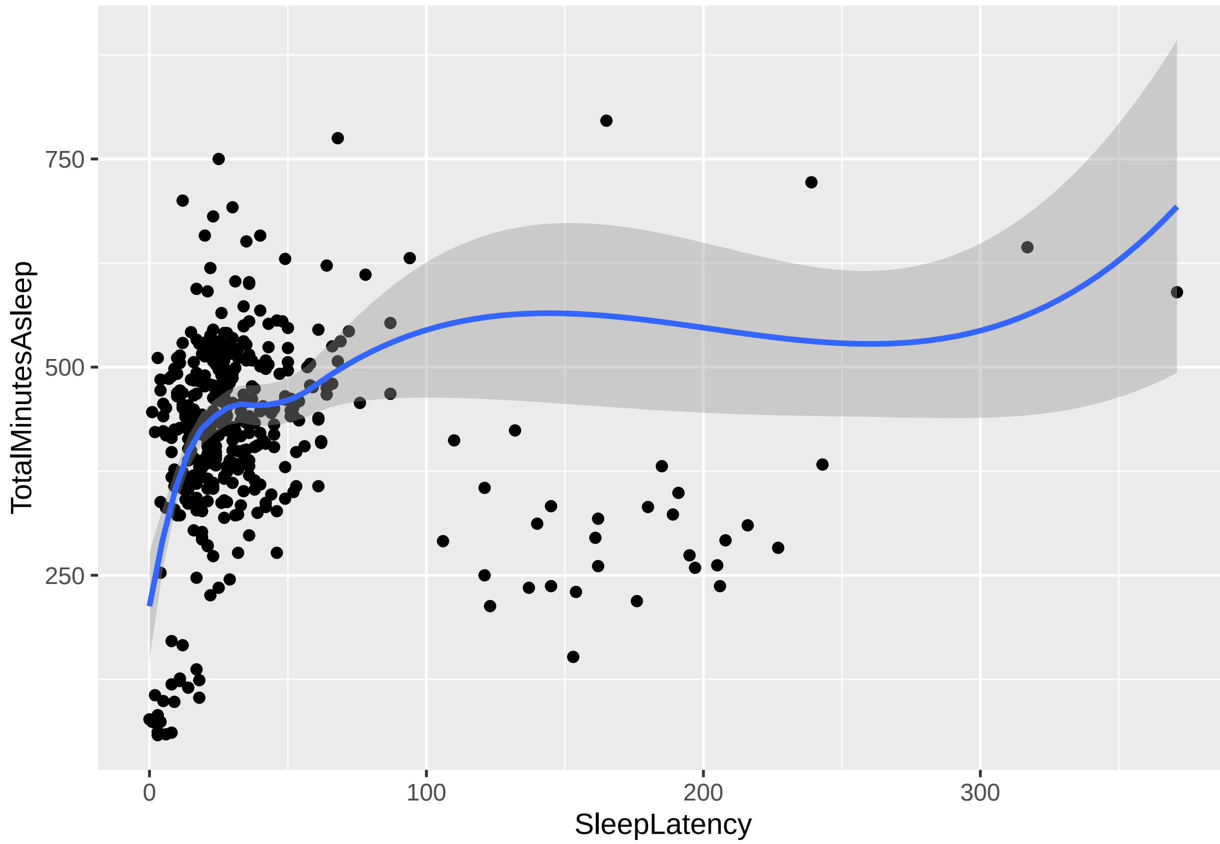
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```
## Warning: Removed 530 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 530 rows containing missing values (geom_point).
```




```
cor(Activity_Day_Sleep_Day$Calories,Activity_Day_Sleep_Day$TotalMinutesAsleep,use="complete.obs")

## [1] -0.02852571

#Normal Minutes Asleep (~5-8 hrs), people that sleep less more likely to spend more time in bed (trying
ggplot(data=Activity_Day_Sleep_Day, aes(x=SleepLatency, y=TotalMinutesAsleep)) + geom_point() + geom_sm

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

## Warning: Removed 530 rows containing non-finite values (stat_smooth).
## Removed 530 rows containing missing values (geom_point).
```




```
cor(Activity_Day_Sleep_Day$SleepLatency,Activity_Day_Sleep_Day$TotalMinutesAsleep,use="complete.obs")
```

```
## [1] -0.001761677
```

Data Takeaways - Bulletpoints

More Distance and Steps Correlated to more Calories Burned Less Sleep Correlated with More Sedentary Minutes that day Less sleep, lower energy for exercise? More Exercise and Less Sedentary Minutes = Less Minutes Asleep More exercise, higher sleep quality and thus less minutes asleep? Low Negative Correlation between Calories and Sleep Latency Could this be due to less calories burned keeping people awake? Lack of Food or Exercise?

Product Suggestions

My product suggestion or improvement for Bellabeat would be directed at aiming to improve sleep and exercise, working to balance and help schedule Bellabeat users both exercise consistently and sleep consistently for better motivation and higher quality sleep.

By further analyzing what general data says about the quality and quantity of sleep in correlation with various factors of exercise, such as the time exercise takes place as well as the length and intensity of it, we can draw some ideas on what general suggestions to make to people.

Depending on the future analysis, we might break down users by their fitness and make suggestions such as getting at least 30 minutes of intense exercise per day and encourage them to consistently sleep for healthy amounts to help them reinforce their exercise habits with the extra willpower and motivation that can come from it.

I'm not an expert when it comes to mobile apps and so I think it would be best to explore how mobile apps encourage users' habits utilizing rewards and other methods to see what would work best for Bellabeat and their users. One small way I can think of to encourage users is that once they have been active enough, showing off the data we've tracked about their progress and success, such as streaks or improvements in exercise and sleep.

Figuring out how to implement these app improvements would be highly beneficial to Bellabeat and their users by helping reinforce and benefit exercise and sleep, and thereby the use of the app as a means to do so.