Analytics Capstone

Kaniqua Outlaw

2024-01-05

Read in datasets

```
dailyActivity <- read_csv("data/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.
```

```
head(dailyActivity)
```

```
## # A tibble: 6 x 15
##           Id ActivityDate TotalSteps TotalDistance TrackerDistance
##        <dbl> <chr>             <dbl>         <dbl>           <dbl>
## 1 1503960366 4/12/2016         13162          8.5             8.5
## 2 1503960366 4/13/2016         10735          6.97            6.97
## 3 1503960366 4/14/2016         10460          6.74            6.74
## 4 1503960366 4/15/2016          9762          6.28            6.28
## 5 1503960366 4/16/2016         12669          8.16            8.16
## 6 1503960366 4/17/2016          9705          6.48            6.48
## # i 10 more variables: LoggedActivitiesDistance <dbl>,
## #   VeryActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
## #   LightActiveDistance <dbl>, SedentaryActiveDistance <dbl>,
## #   VeryActiveMinutes <dbl>, FairlyActiveMinutes <dbl>,
## #   LightlyActiveMinutes <dbl>, SedentaryMinutes <dbl>, Calories <dbl>
```

Examine the data and get summary statistics

Daily Activity*

```
#  Examine the number of unique participants and observations
n_distinct(dailyActivity$Id)
```

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

```
nrow(dailyActivity)
```

```
## [1] 940
```

```
#  Convert ActivityDate to date-time object
dailyActivity$ActivityDate <- mdy(dailyActivity$ActivityDate)
head(dailyActivity)
```

```
## # A tibble: 6 x 15
##           Id ActivityDate TotalSteps TotalDistance TrackerDistance
##        <dbl> <date>            <dbl>         <dbl>           <dbl>
## 1 1503960366 2016-04-12        13162          8.5             8.5
## 2 1503960366 2016-04-13        10735          6.97            6.97
## 3 1503960366 2016-04-14        10460          6.74            6.74
## 4 1503960366 2016-04-15         9762          6.28            6.28
## 5 1503960366 2016-04-16        12669          8.16            8.16
## 6 1503960366 2016-04-17         9705          6.48            6.48
## # i 10 more variables: LoggedActivitiesDistance <dbl>,
## #   VeryActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
## #   LightActiveDistance <dbl>, SedentaryActiveDistance <dbl>,
## #   VeryActiveMinutes <dbl>, FairlyActiveMinutes <dbl>,
## #   LightlyActiveMinutes <dbl>, SedentaryMinutes <dbl>, Calories <dbl>
```

```
# Check for missing values
sum(is.na(dailyActivity))
```

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

```
#  Run validation check for TotalDistance column.  The sum of Logged Activities, Very Active, Moderatel
dailyActivity %>%
  mutate(theoretical_total = LoggedActivitiesDistance + VeryActiveDistance + ModeratelyActiveDistance +
  filter(theoretical_total != TotalDistance) %>%
  select(TotalDistance, theoretical_total)
```

```
## # A tibble: 636 x 2
##    TotalDistance theoretical_total
##            <dbl>             <dbl>
##  1          8.5               8.49
##  2          6.97              6.97
##  3          6.74              6.75
##  4          6.28              6.23
##  5          8.16              8.16
##  6          6.48              6.48
##  7          8.59              8.60
##  8          9.88              9.88
##  9          6.68              6.68
## 10          6.34              6.34
## # i 626 more rows
```

Daily Steps

```
dailySteps <- read_csv("data/dailySteps_merged.csv")
```

```
## Rows: 940 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): ActivityDay
## dbl (2): Id, StepTotal
##
## 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.
```

```
head(dailySteps)
```

```
## # A tibble: 6 x 3
##           Id ActivityDay StepTotal
##        <dbl> <chr>           <dbl>
## 1 1503960366 4/12/2016       13162
## 2 1503960366 4/13/2016       10735
## 3 1503960366 4/14/2016       10460
## 4 1503960366 4/15/2016        9762
## 5 1503960366 4/16/2016       12669
## 6 1503960366 4/17/2016        9705
```

```
#  Examine the number of unique participants and observations
n_distinct(dailySteps$Id)
```

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

```
nrow(dailySteps)
```

```
## [1] 940
```

```
#  Convert ActivityDay from character to date-time object
dailySteps$ActivityDay <- mdy(dailySteps$ActivityDay)
head(dailySteps)
```

```
## # A tibble: 6 x 3
##           Id ActivityDay StepTotal
##        <dbl> <date>          <dbl>
## 1 1503960366 2016-04-12      13162
## 2 1503960366 2016-04-13      10735
## 3 1503960366 2016-04-14      10460
## 4 1503960366 2016-04-15       9762
## 5 1503960366 2016-04-16      12669
## 6 1503960366 2016-04-17       9705
```

```r
#  Check for missing values
sum(is.na(dailySteps))
```

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

Sleep Minutes

```r
Sleepminute <- read_csv("data/minuteSleep_merged.csv")
```

```
## Rows: 188521 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): date
## dbl (3): Id, value, logId
##
## 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.
```

```r
head(Sleepminute)
```

```
## # A tibble: 6 x 4
##           Id date                 value       logId
##        <dbl> <chr>                <dbl>       <dbl>
## 1 1503960366 4/12/2016 2:47:30 AM     3 11380564589
## 2 1503960366 4/12/2016 2:48:30 AM     2 11380564589
## 3 1503960366 4/12/2016 2:49:30 AM     1 11380564589
## 4 1503960366 4/12/2016 2:50:30 AM     1 11380564589
## 5 1503960366 4/12/2016 2:51:30 AM     1 11380564589
## 6 1503960366 4/12/2016 2:52:30 AM     1 11380564589
```

```r
#  Examine the number of unique participants and observations
n_distinct(Sleepminute$Id)
```

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

```r
nrow(Sleepminute)
```

```
## [1] 188521
```

```r
#  Convert date to date-time object
Sleepminute$date <- mdy_hms(Sleepminute$date)
head(Sleepminute)
```

```
## # A tibble: 6 x 4
##           Id date                value       logId
##        <dbl> <dttm>              <dbl>       <dbl>
## 1 1503960366 2016-04-12 02:47:30     3 11380564589
## 2 1503960366 2016-04-12 02:48:30     2 11380564589
## 3 1503960366 2016-04-12 02:49:30     1 11380564589
## 4 1503960366 2016-04-12 02:50:30     1 11380564589
## 5 1503960366 2016-04-12 02:51:30     1 11380564589
## 6 1503960366 2016-04-12 02:52:30     1 11380564589
```

```
#  Rename date column to sleep_date
Sleepminute <- Sleepminute %>%
rename(sleep_date = date)
head(Sleepminute)
```

```
## # A tibble: 6 x 4
##           Id sleep_date          value       logId
##        <dbl> <dttm>              <dbl>       <dbl>
## 1 1503960366 2016-04-12 02:47:30     3 11380564589
## 2 1503960366 2016-04-12 02:48:30     2 11380564589
## 3 1503960366 2016-04-12 02:49:30     1 11380564589
## 4 1503960366 2016-04-12 02:50:30     1 11380564589
## 5 1503960366 2016-04-12 02:51:30     1 11380564589
## 6 1503960366 2016-04-12 02:52:30     1 11380564589
```

```
#  Check for missing values
sum(is.na(Sleepminute))
```

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

Sleep Day

```
sleepDay <- read_csv("data/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.
```

```
head(sleepDay)
```

```
## # A tibble: 6 x 5
##           Id SleepDay        TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
##        <dbl> <chr>                       <dbl>              <dbl>          <dbl>
## 1 1503960366 4/12/2016 12:0~                 1                327            346
## 2 1503960366 4/13/2016 12:0~                 2                384            407
## 3 1503960366 4/15/2016 12:0~                 1                412            442
## 4 1503960366 4/16/2016 12:0~                 2                340            367
## 5 1503960366 4/17/2016 12:0~                 1                700            712
## 6 1503960366 4/19/2016 12:0~                 1                304            320
```

```
#  Examine the number of unique participants and observations
n_distinct(sleepDay$Id)
```

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

```
nrow(sleepDay)
```

```
## [1] 413
```

```
#  Convert SleepDay to date-time object
sleepDay$SleepDay <- mdy_hms(sleepDay$SleepDay)
head(sleepDay)
```

```
## # A tibble: 6 x 5
##       Id SleepDay            TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
##    <dbl> <dttm>                          <dbl>              <dbl>          <dbl>
## 1 1.50e9 2016-04-12 00:00:00                 1                327            346
## 2 1.50e9 2016-04-13 00:00:00                 2                384            407
## 3 1.50e9 2016-04-15 00:00:00                 1                412            442
## 4 1.50e9 2016-04-16 00:00:00                 2                340            367
## 5 1.50e9 2016-04-17 00:00:00                 1                700            712
## 6 1.50e9 2016-04-19 00:00:00                 1                304            320
```

```
#  Check for missing values
sum(is.na(sleepDay))
```

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

#Examine Daily Activity & Steps, Heart Rate, Hourly Steps, Intensities, & Calories, and Sleep. Explore trends in device usage.

###Summary statistics for DailyActivity

```
#  Examine summary statistics for total steps, total distance, sedentary
dailyActivity %>%
  select(TotalSteps, TotalDistance, VeryActiveDistance:Calories ) %>%
  summary()
```

```
##    TotalSteps    TotalDistance    VeryActiveDistance ModeratelyActiveDistance
##  Min.   :    0   Min.   : 0.000   Min.   : 0.000     Min.   :0.0000          
##  1st Qu.: 3790   1st Qu.: 2.620   1st Qu.: 0.000     1st Qu.:0.0000          
##  Median : 7406   Median : 5.245   Median : 0.210     Median :0.2400          
##  Mean   : 7638   Mean   : 5.490   Mean   : 1.503     Mean   :0.5675          
##  3rd Qu.:10727   3rd Qu.: 7.713   3rd Qu.: 2.053     3rd Qu.:0.8000          
##  Max.   :36019   Max.   :28.030   Max.   :21.920     Max.   :6.4800          
##  LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
##  Min.   : 0.000      Min.   :0.000000        Min.   :  0.00   
##  1st Qu.: 1.945      1st Qu.:0.000000        1st Qu.:  0.00   
##  Median : 3.365      Median :0.000000        Median :  4.00   
##  Mean   : 3.341      Mean   :0.001606        Mean   : 21.16   
##  3rd Qu.: 4.782      3rd Qu.:0.000000        3rd Qu.: 32.00   
##  Max.   :10.710      Max.   :0.110000       Max.   :210.00
##  FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes    Calories
##  Min.   :  0.00      Min.   :  0.0        Min.   :   0.0   Min.   :   0
##  1st Qu.:  0.00      1st Qu.:127.0        1st Qu.: 729.8   1st Qu.:1828
##  Median :  6.00      Median :199.0        Median :1057.5   Median :2134
##  Mean   : 13.56      Mean   :192.8        Mean   : 991.2   Mean   :2304
##  3rd Qu.: 19.00      3rd Qu.:264.0        3rd Qu.:1229.5   3rd Qu.:2793
##  Max.   :143.00      Max.   :518.0        Max.   :1440.0   Max.   :4900
```

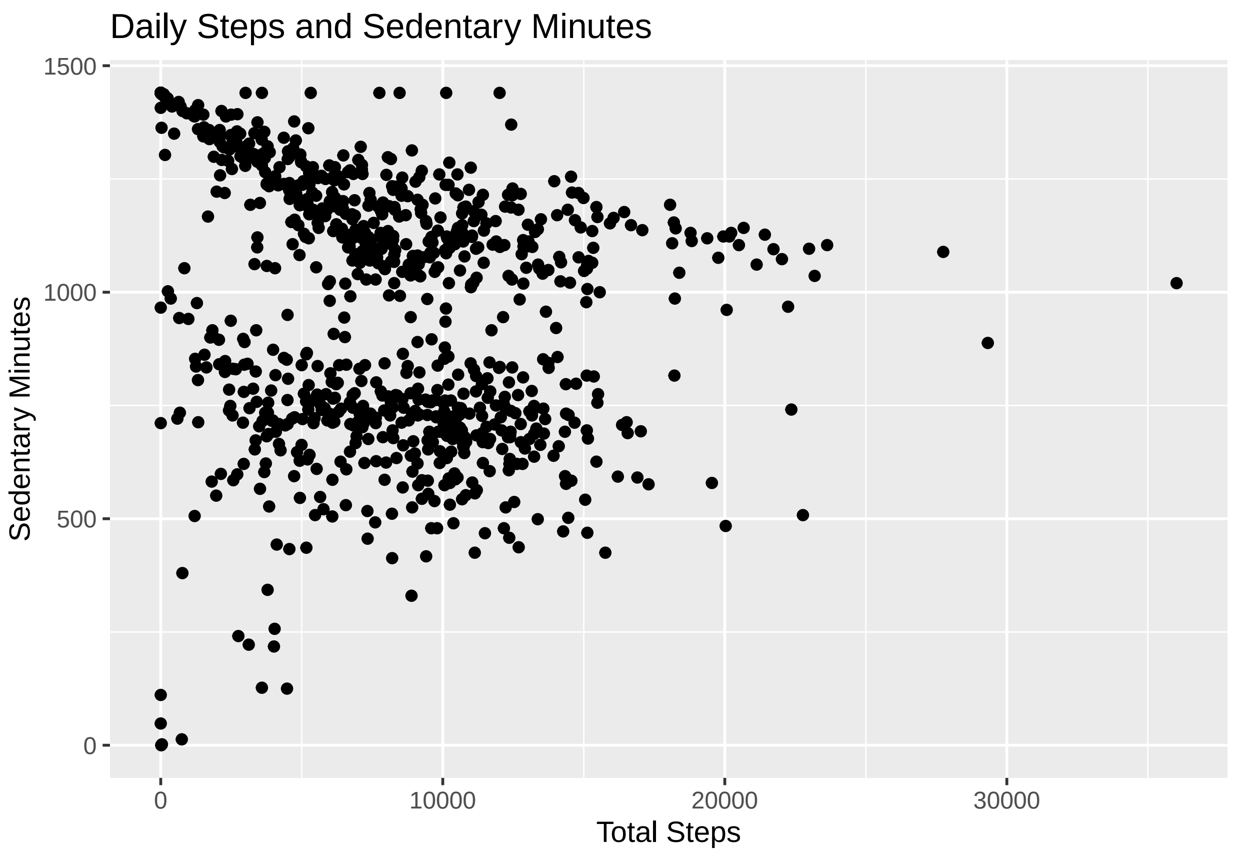
```r
#  Examine the average calories and median sedentary, lightly active, and fairly active minutes

dailyActivity %>%
  summarize(average_calories = mean(Calories),
            avg_sedentary_minutes = mean(SedentaryMinutes),
            avg_lightly_minutes = mean(LightlyActiveMinutes),
            avg_moderate_minutes = mean(FairlyActiveMinutes),
            avg_veryactive_minutes = mean(VeryActiveMinutes),
            avg_sedentary_dist = mean(SedentaryActiveDistance),
            avg_light_dist = mean(LightActiveDistance),
            avg_moderate_dist = mean(ModeratelyActiveDistance),
            avg_veryactive_dist = mean(VeryActiveDistance),
            avg_total_dist = mean(TotalDistance))
```

```
## # A tibble: 1 x 10
##   average_calories avg_sedentary_minutes avg_lightly_minutes
##              <dbl>                 <dbl>               <dbl>
## 1            2304.                  991.                193.
## # i 7 more variables: avg_moderate_minutes <dbl>, avg_veryactive_minutes <dbl>,
## #   avg_sedentary_dist <dbl>, avg_light_dist <dbl>, avg_moderate_dist <dbl>,
## #   avg_veryactive_dist <dbl>, avg_total_dist <dbl>
```

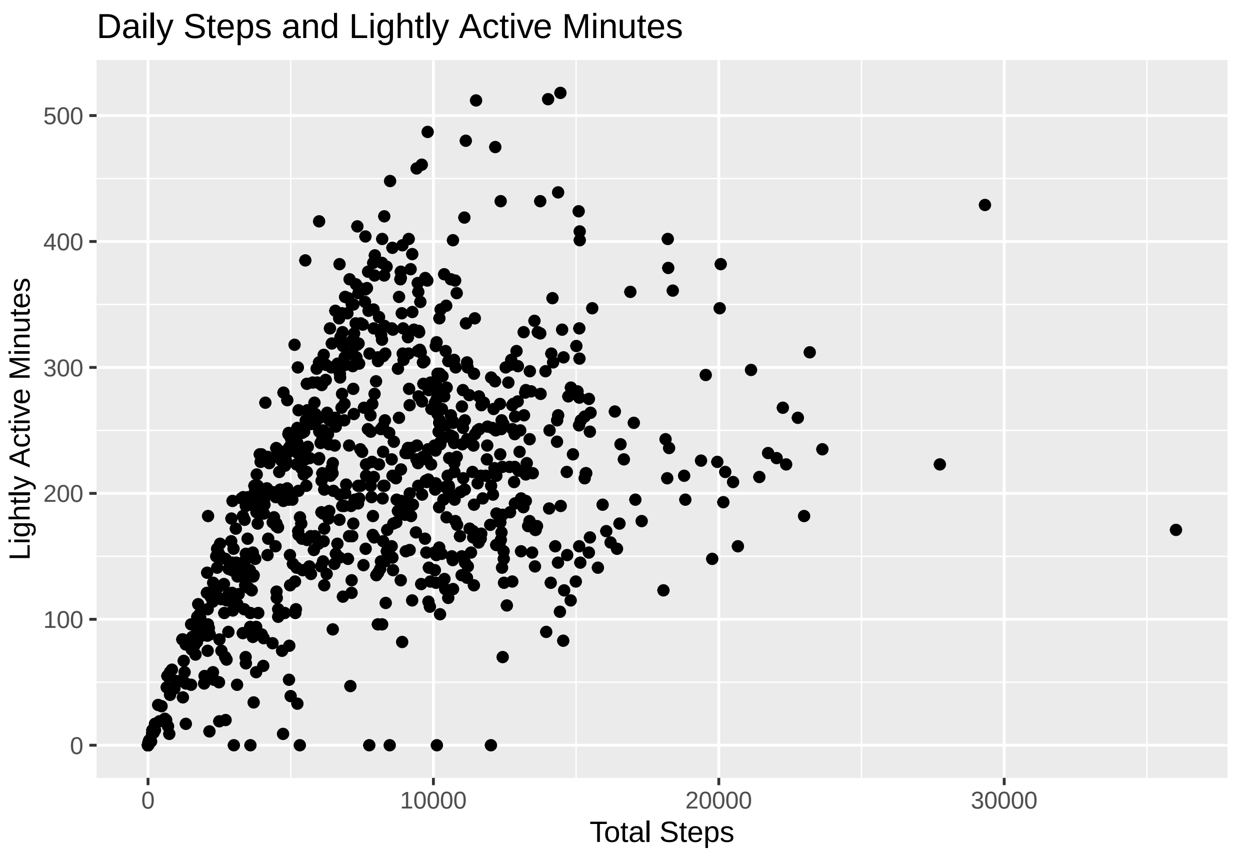
On average people spent more time engaging in light activities than any other activity type

Explore Daily Activity

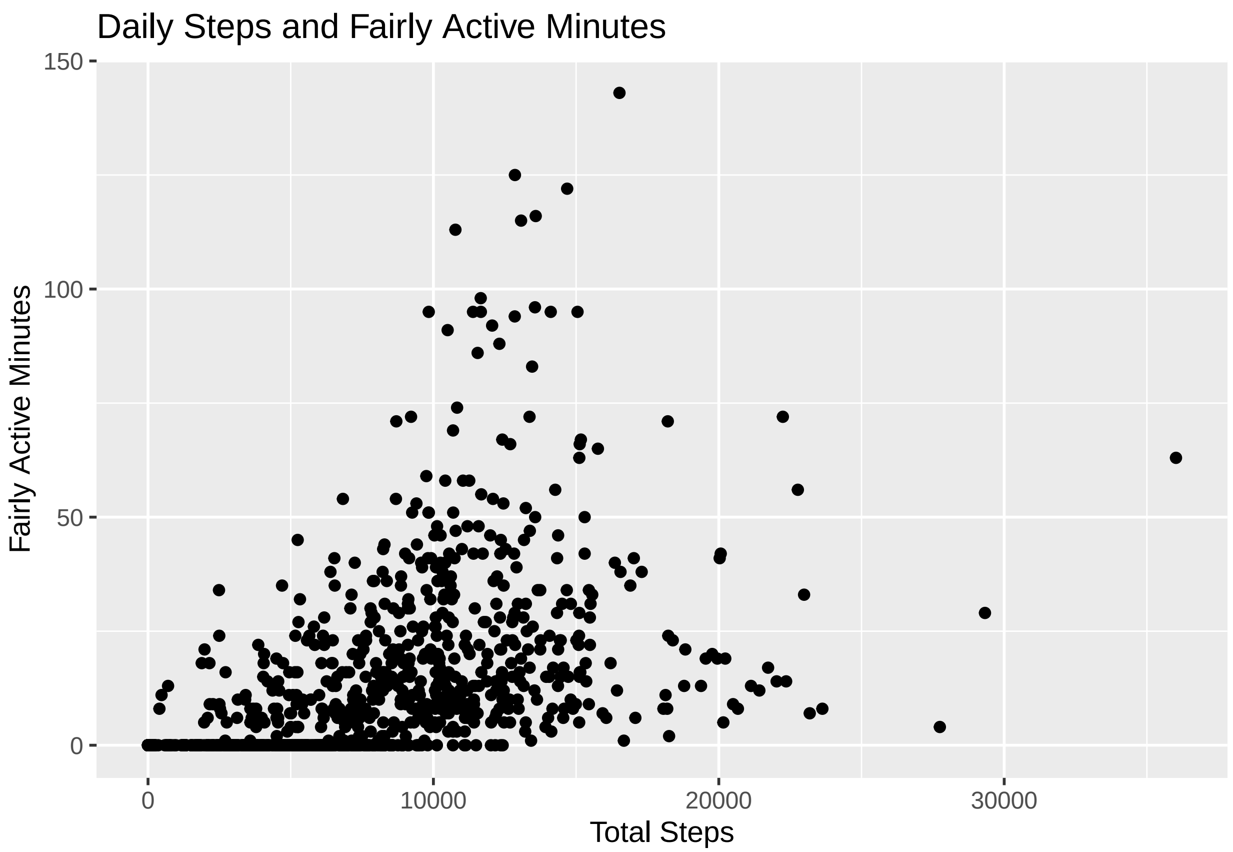
```r
ggplot(dailyActivity, aes(x = TotalSteps, y = SedentaryMinutes))+
  geom_point()+
  labs(x = "Total Steps",
       y = "Sedentary Minutes",
       title = "Daily Steps and Sedentary Minutes")
```



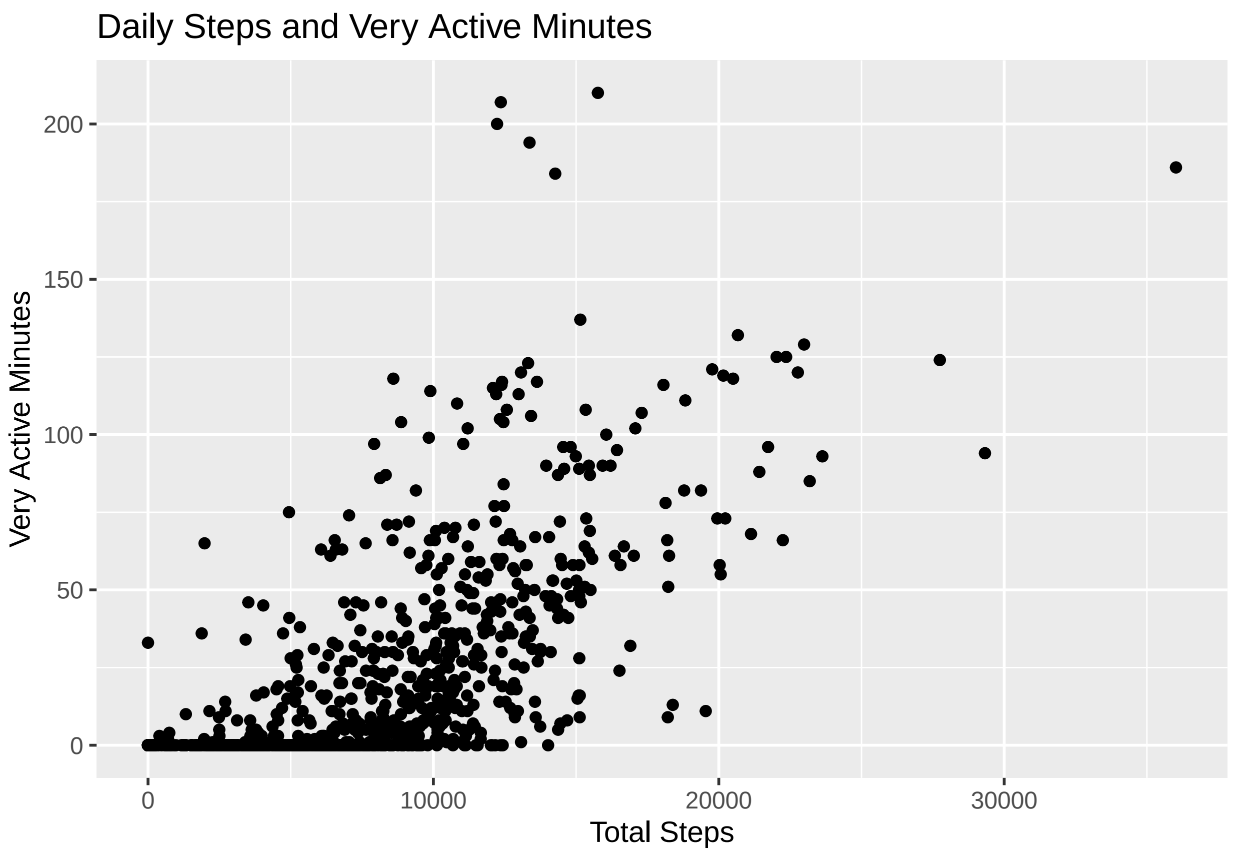
```r
ggplot(dailyActivity, aes(x = TotalSteps, y = LightlyActiveMinutes))+
  geom_point()+
  labs(x = "Total Steps",
       y = "Lightly Active Minutes",
       title = "Daily Steps and Lightly Active Minutes")
```



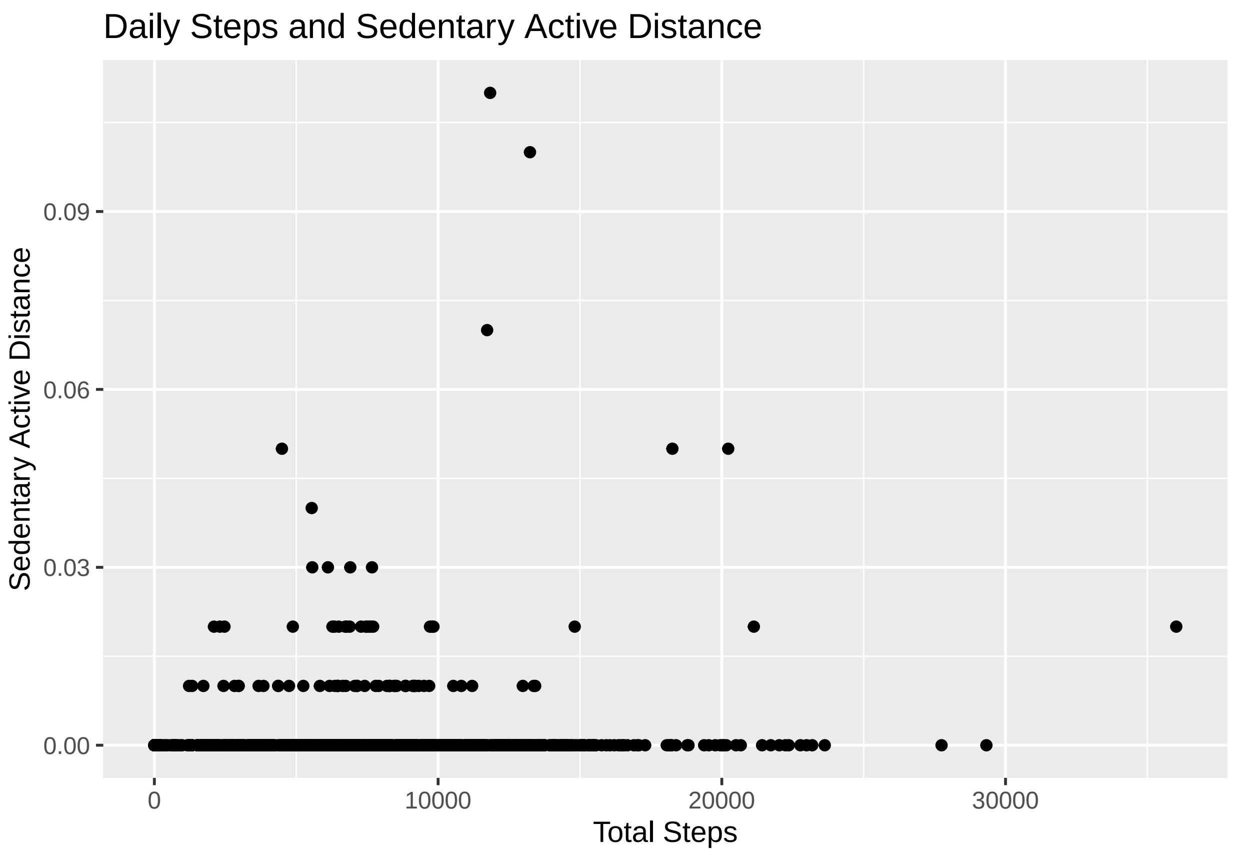
```r
ggplot(dailyActivity, aes(x = TotalSteps, y = FairlyActiveMinutes))+
  geom_point()+
  labs(x = "Total Steps",
       y = "Fairly Active Minutes",
       title = "Daily Steps and Fairly Active Minutes")
```



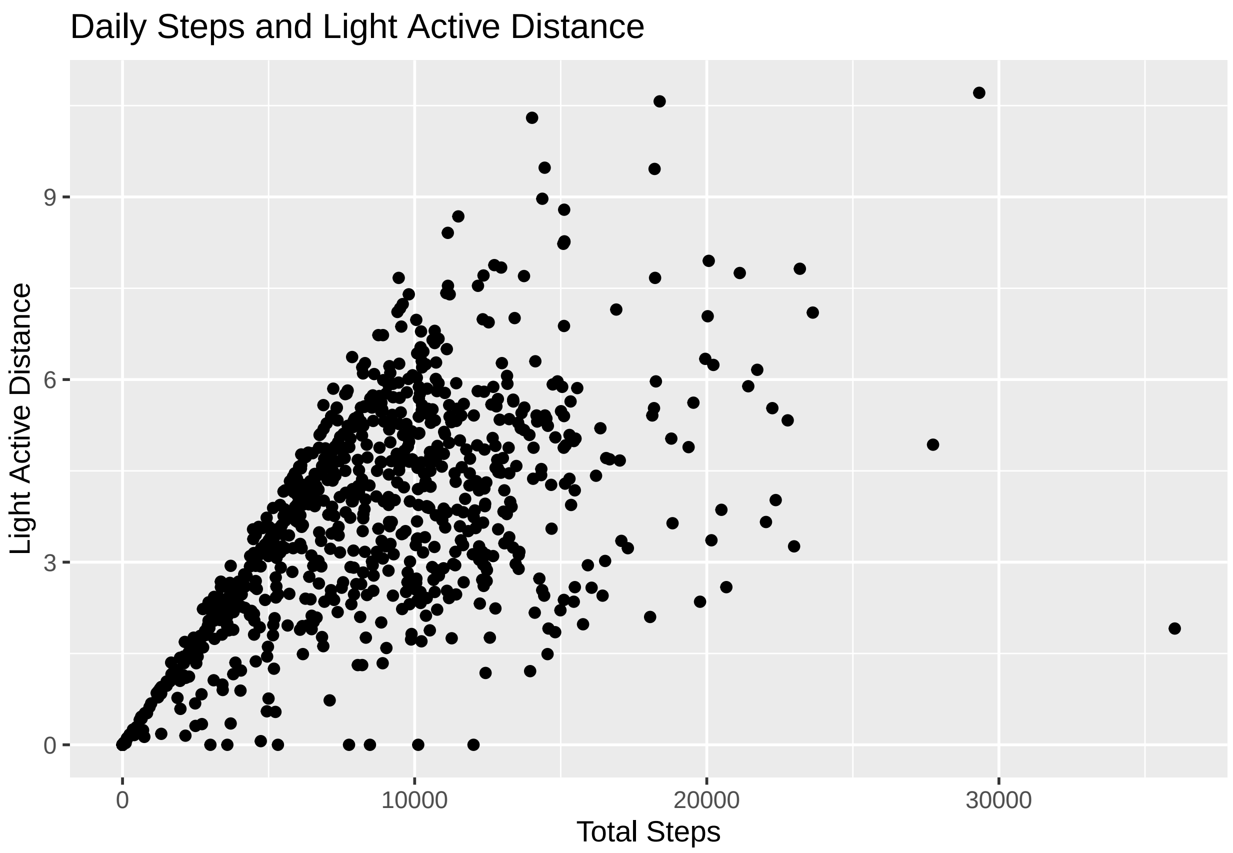
```
ggplot(dailyActivity, aes(x = TotalSteps, y = VeryActiveMinutes))+
  geom_point()+
  labs(x = "Total Steps",
       y = "Very Active Minutes",
       title = "Daily Steps and Very Active Minutes")
```



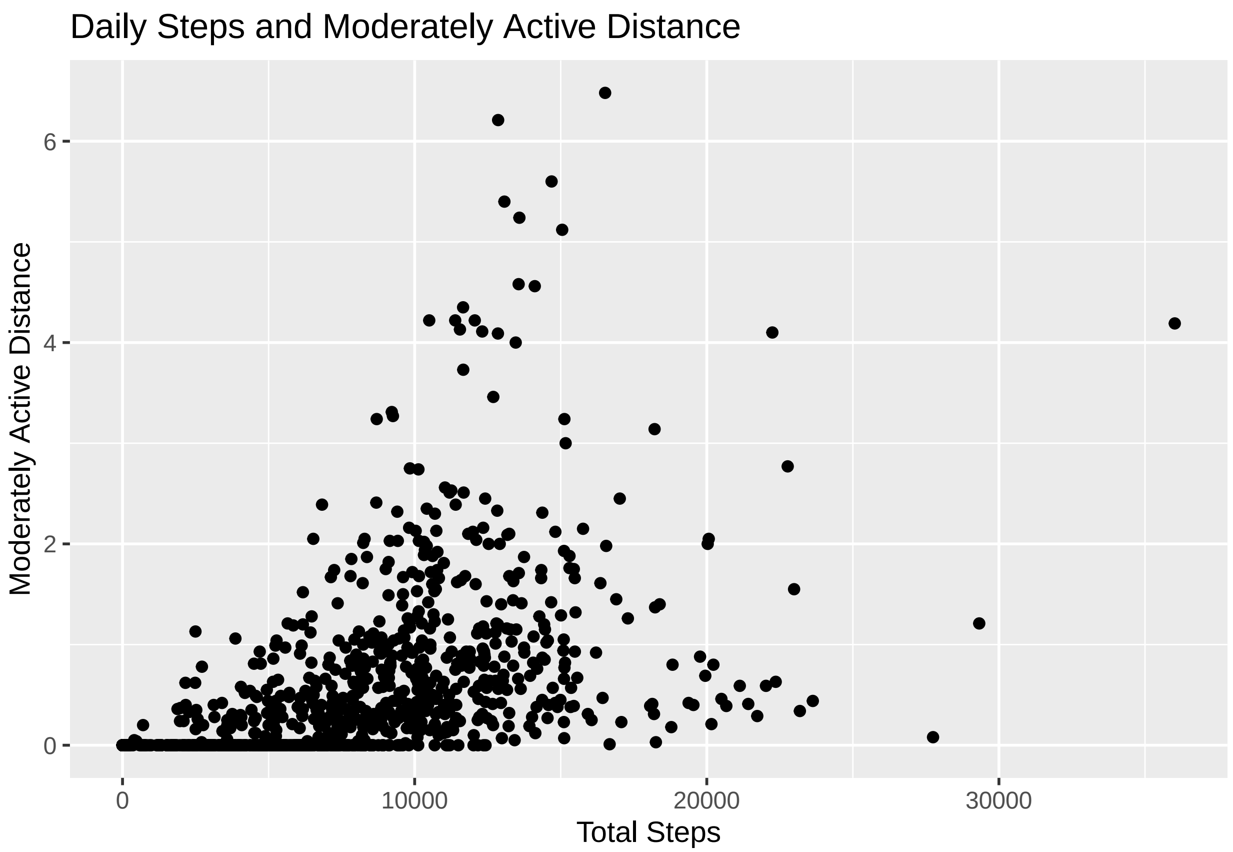
```
ggplot(dailyActivity, aes(x = TotalSteps, y = SedentaryActiveDistance))+
  geom_point()+
  labs(x = "Total Steps",
       y = "Sedentary Active Distance",
       title = "Daily Steps and Sedentary Active Distance")
```



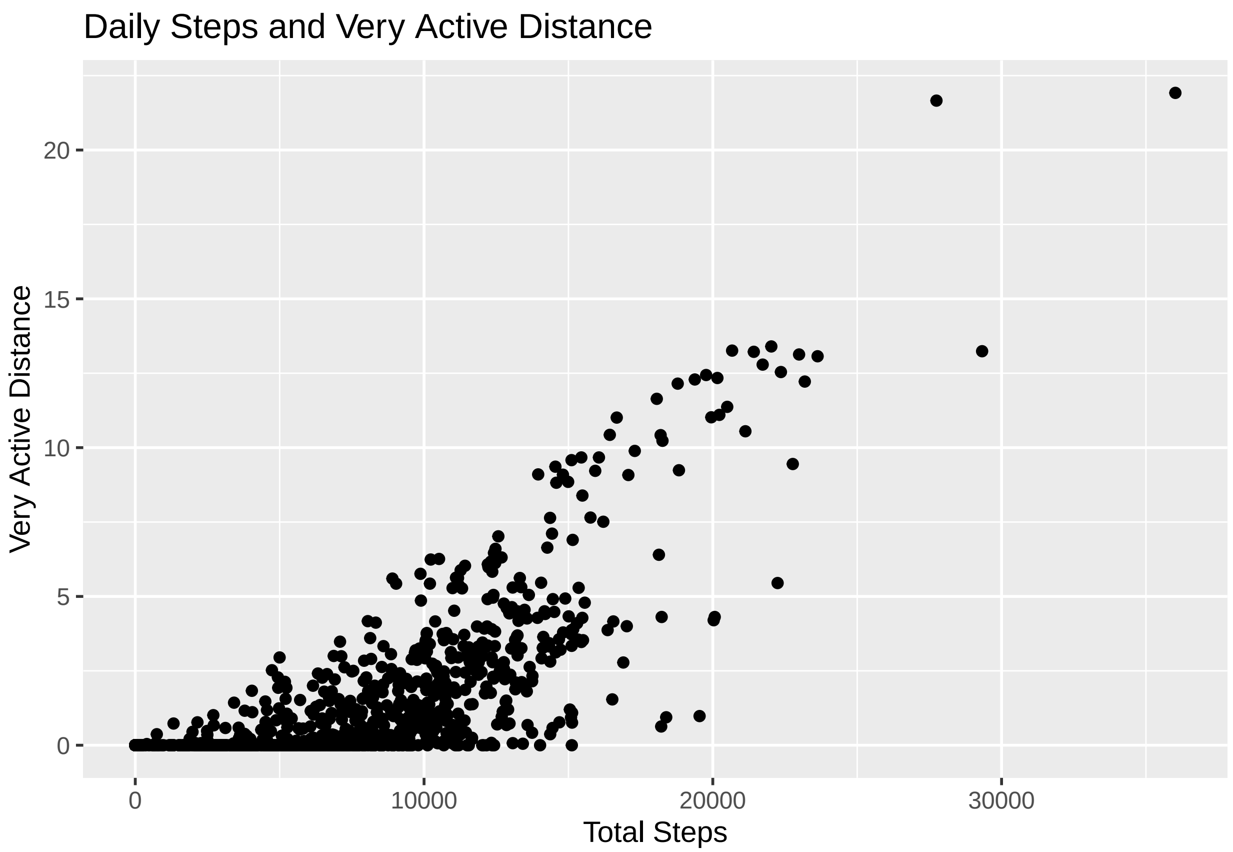
```
ggplot(dailyActivity, aes(x = TotalSteps, y = LightActiveDistance))+
  geom_point()+
  labs(x = "Total Steps",
       y = "Light Active Distance",
       title = "Daily Steps and Light Active Distance")
```



```
ggplot(dailyActivity, aes(x = TotalSteps, y = ModeratelyActiveDistance))+
  geom_point()+
  labs(x = "Total Steps",
       y = "Moderately Active Distance",
       title = "Daily Steps and Moderately Active Distance")
```



```r
ggplot(dailyActivity, aes(x = TotalSteps, y = VeryActiveDistance))+
  geom_point()+
  labs(x = "Total Steps",
       y = "Very Active Distance",
       title = "Daily Steps and Very Active Distance")
```



Summary statistics for Sleep Day

```
sleepDay %>%
  select(TotalMinutesAsleep, TotalTimeInBed) %>%
  summary()
```

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

```
sleepDay %>%
  summarise(avg_sleep_min = mean(TotalMinutesAsleep),
            avg_bed_time = mean(TotalTimeInBed))
```

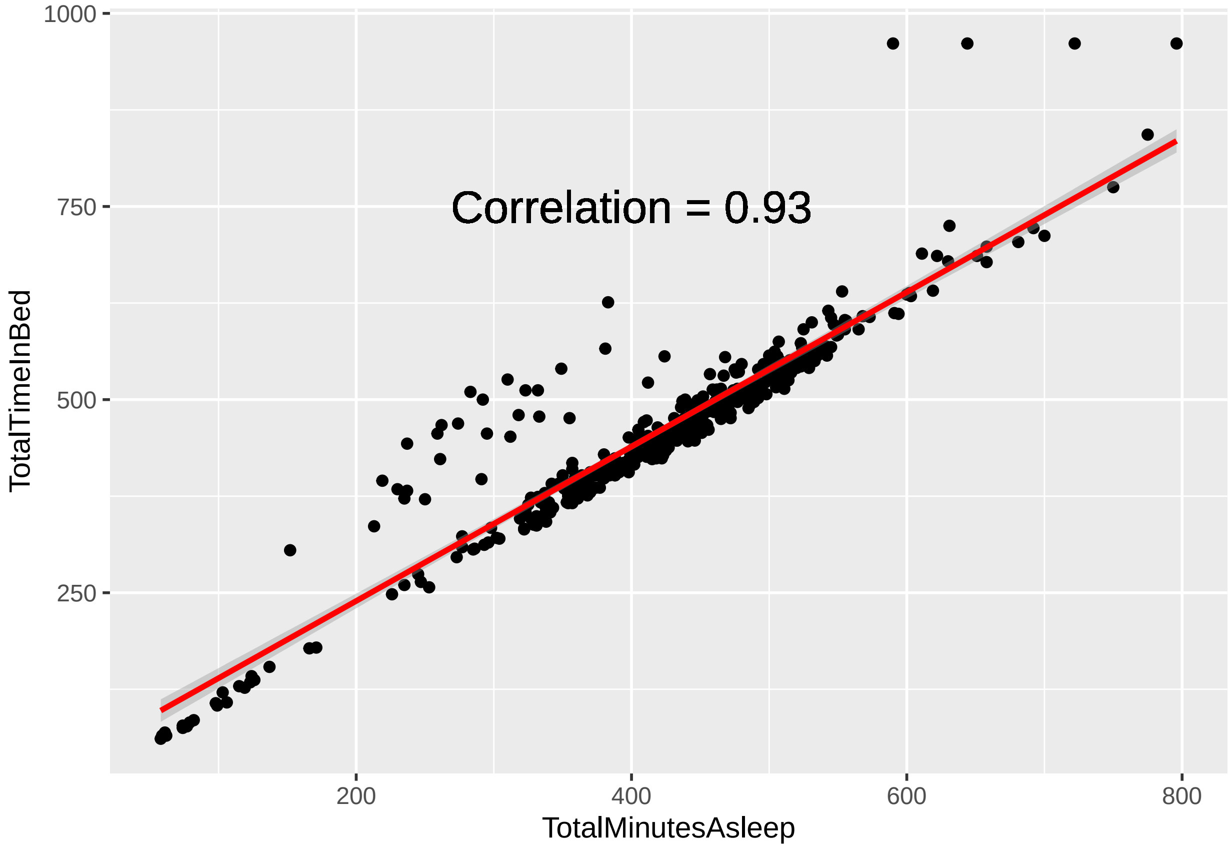
```
## # A tibble: 1 x 2
##   avg_sleep_min avg_bed_time
##           <dbl>        <dbl>
## 1          419.         459.
```

```
#  The average time asleep was 419 minutes, and total time in bed averaged 458 minutes.  The total time
```

Explore Sleep Activity

```
correlation <- cor(sleepDay$TotalMinutesAsleep, sleepDay$TotalTimeInBed, use = "complete.obs")
g <- ggplot(sleepDay, aes(x = TotalMinutesAsleep, y = TotalTimeInBed))
g+ geom_point() +
  geom_smooth(method = lm, col = "red") +
  geom_text(x = 400, y = 750, size = 6,
            label = paste0("Correlation = ", round(correlation, 2)))
```

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



There is a positive, linear correlation between time in bed and time spent asleep.

Join Sleep and Daily Activity data

```
combined_sleep_and_daily_activity <- merge(dailyActivity, sleepDay, by = "Id")
head(combined_sleep_and_daily_activity)
```

```
##           Id ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366   2016-05-07      11992          7.71            7.71
## 2 1503960366   2016-05-07      11992          7.71            7.71
## 3 1503960366   2016-05-07      11992          7.71            7.71
## 4 1503960366   2016-05-07      11992          7.71            7.71
## 5 1503960366   2016-05-07      11992          7.71            7.71
## 6 1503960366   2016-05-07      11992          7.71            7.71
##   LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                        0               2.46                     2.12
## 2                        0               2.46                     2.12
## 3                        0               2.46                     2.12
## 4                        0               2.46                     2.12
## 5                        0               2.46                     2.12
## 6                        0               2.46                     2.12
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1                3.13                       0                37
## 2                3.13                       0                37
## 3                3.13                       0                37
## 4                3.13                       0                37
## 5                3.13                       0                37
## 6                3.13                       0                37
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories   SleepDay
## 1                  46                  175              833     1821 2016-04-12
## 2                  46                  175              833     1821 2016-04-13
## 3                  46                  175              833     1821 2016-04-15
## 4                  46                  175              833     1821 2016-04-16
## 5                  46                  175              833     1821 2016-04-17
## 6                  46                  175              833     1821 2016-04-19
##   TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
## 1                 1                327            346
## 2                 2                384            407
## 3                 1                412            442
## 4                 2                340            367
## 5                 1                700            712
## 6                 1                304            320
```

```
#  Examine the number of unique participants and observations
n_distinct(combined_sleep_and_daily_activity$Id)
```

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

```
nrow(combined_sleep_and_daily_activity)
```

```
## [1] 12441
```

The original Daily Activity data set had 33 unique participants, while the combined daily activity and sleep dataset has only 24 unique particpants. I applied an outer join to the two datasets to keep the original participants in the dataset.

```
# Outer join daily activity and sleep day to get all the records in daily activity table

join_sleep_dailyactivity <- left_join(dailyActivity, sleepDay, by = c("Id"))
```

```
## Warning in left_join(dailyActivity, sleepDay, by = c("Id")): Detected an unexpected many-to-many rela
## i Row 1 of ‘x‘ matches multiple rows in ‘y‘.
## i Row 1 of ‘y‘ matches multiple rows in ‘x‘.
## i If a many-to-many relationship is expected, set ‘relationship =
##   "many-to-many"‘ to silence this warning.
```

```
head(join_sleep_dailyactivity)
```

```
## # A tibble: 6 x 19
##           Id ActivityDate TotalSteps TotalDistance TrackerDistance
##        <dbl> <date>            <dbl>         <dbl>           <dbl>
## 1 1503960366 2016-04-12        13162           8.5             8.5
## 2 1503960366 2016-04-12        13162           8.5             8.5
## 3 1503960366 2016-04-12        13162           8.5             8.5
## 4 1503960366 2016-04-12        13162           8.5             8.5
## 5 1503960366 2016-04-12        13162           8.5             8.5
## 6 1503960366 2016-04-12        13162           8.5             8.5
## # i 14 more variables: LoggedActivitiesDistance <dbl>,
## #   VeryActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
## #   LightActiveDistance <dbl>, SedentaryActiveDistance <dbl>,
## #   VeryActiveMinutes <dbl>, FairlyActiveMinutes <dbl>,
## #   LightlyActiveMinutes <dbl>, SedentaryMinutes <dbl>, Calories <dbl>,
## #   SleepDay <dttm>, TotalSleepRecords <dbl>, TotalMinutesAsleep <dbl>,
## #   TotalTimeInBed <dbl>
```

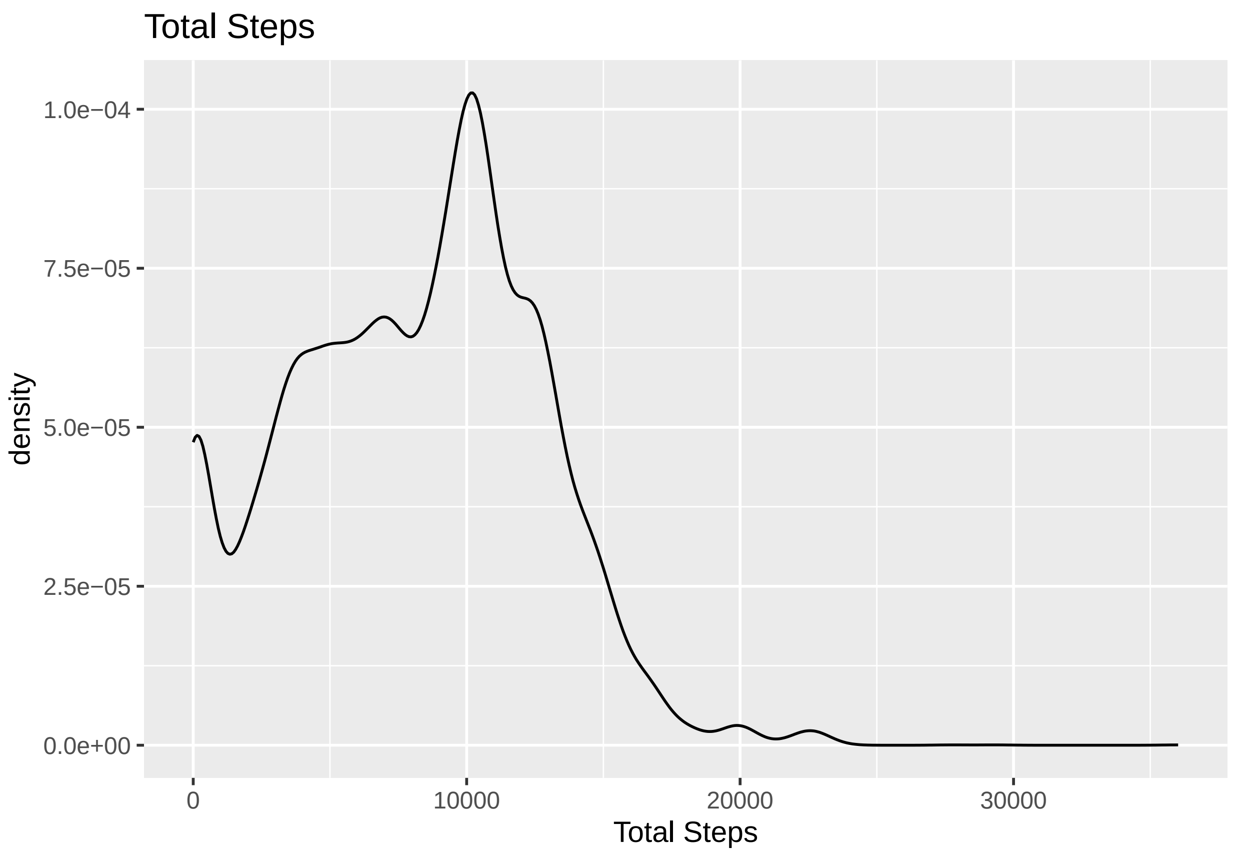
```
n_distinct(join_sleep_dailyactivity$Id)
```

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

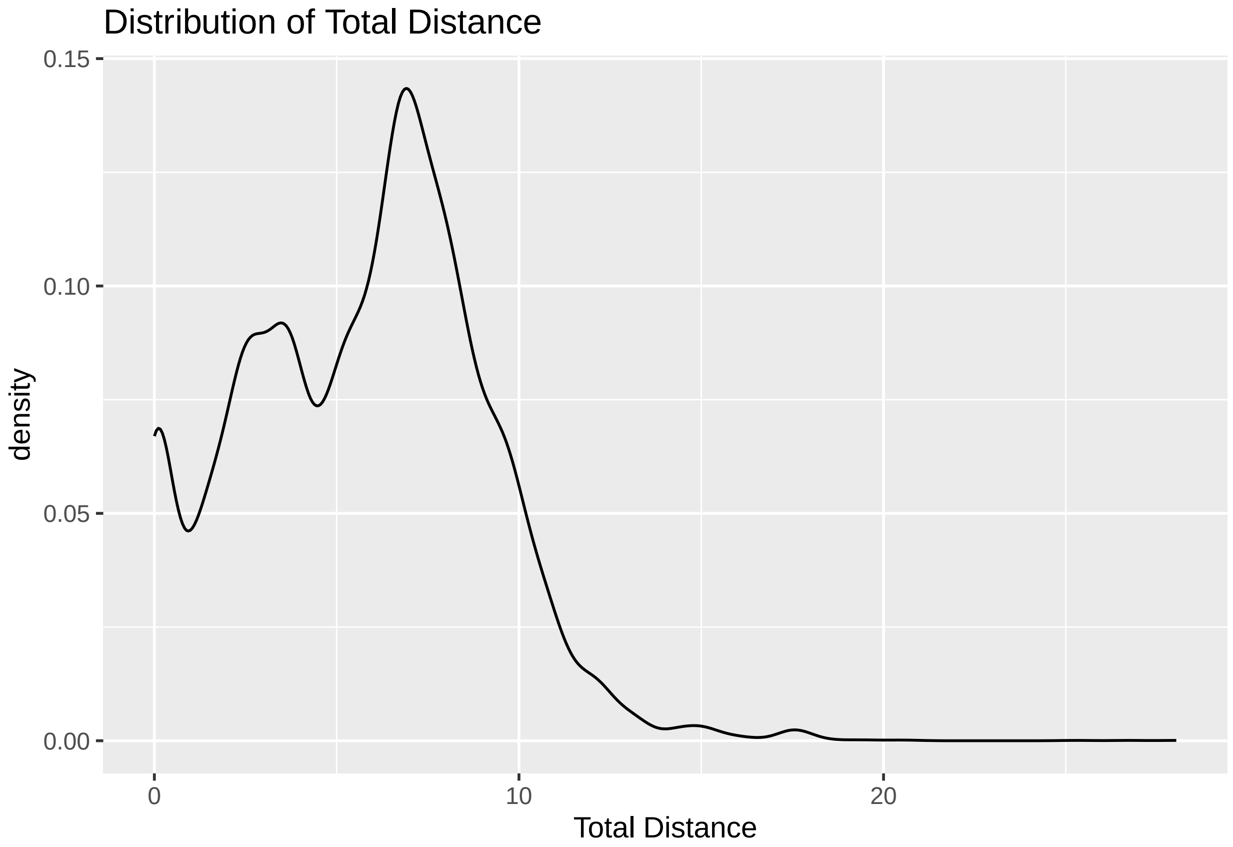
```
nrow(join_sleep_dailyactivity)
```

```
## [1] 12668
```

```
#  Distribution of Total Steps
ggplot(join_sleep_dailyactivity, aes(x = TotalSteps)) +
geom_density()+
  labs(x = "Total Steps",
       title = "Total Steps")
```



```
#  Distribution of total distance
ggplot(join_sleep_dailyactivity, aes(x = TotalDistance)) +
geom_density()+
  labs(x = "Total Distance",
       title = "Distribution of Total Distance")
```



Extract month from Activity Date

```
#  Extract month to visualize activity by month
join_sleep_dailyactivity <- join_sleep_dailyactivity %>%
  mutate(ActivityMonth = month(ActivityDate, label = TRUE)) %>%
  select(ActivityDate, ActivityMonth, everything())
```

Because the data looked at activity for only 2 months, I decided to drill down to weekdays to get a better idea of activities by day

Extract weekday to visualize activity by weekday

```
join_sleep_dailyactivity <- join_sleep_dailyactivity %>%
  mutate(ActivityDay = wday(ActivityDate, label = TRUE, abbr = FALSE)) %>%
  select(ActivityDate, ActivityDay, everything())
```

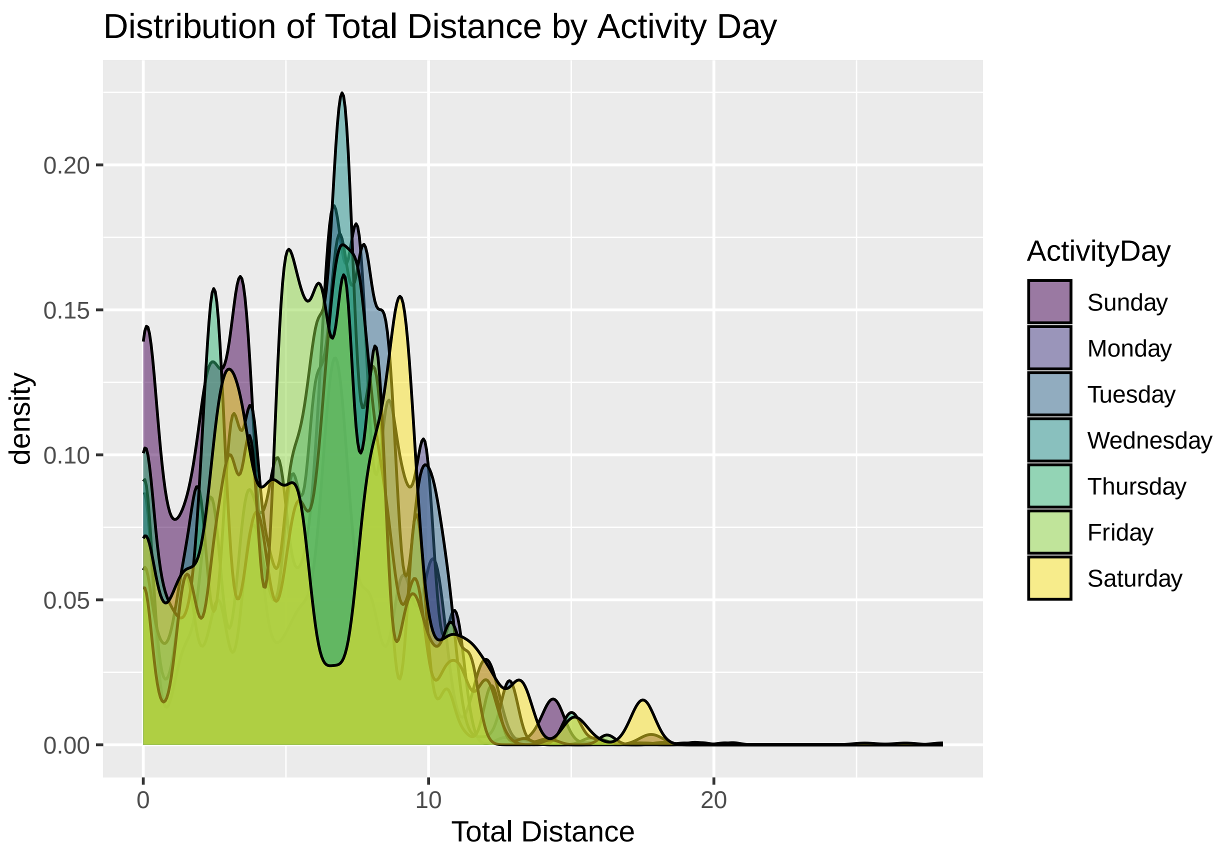
Summary Statistics (mean) for Steps, Distance, Activity Minutes

```
#  Summary statistics (mean) for sedentary, lightly, and moderately active users
join_sleep_dailyactivity %>%
  summarize(across(.fns = mean, .cols = c(TotalSteps, TotalDistance, SedentaryActiveDistance, Sedentary
```

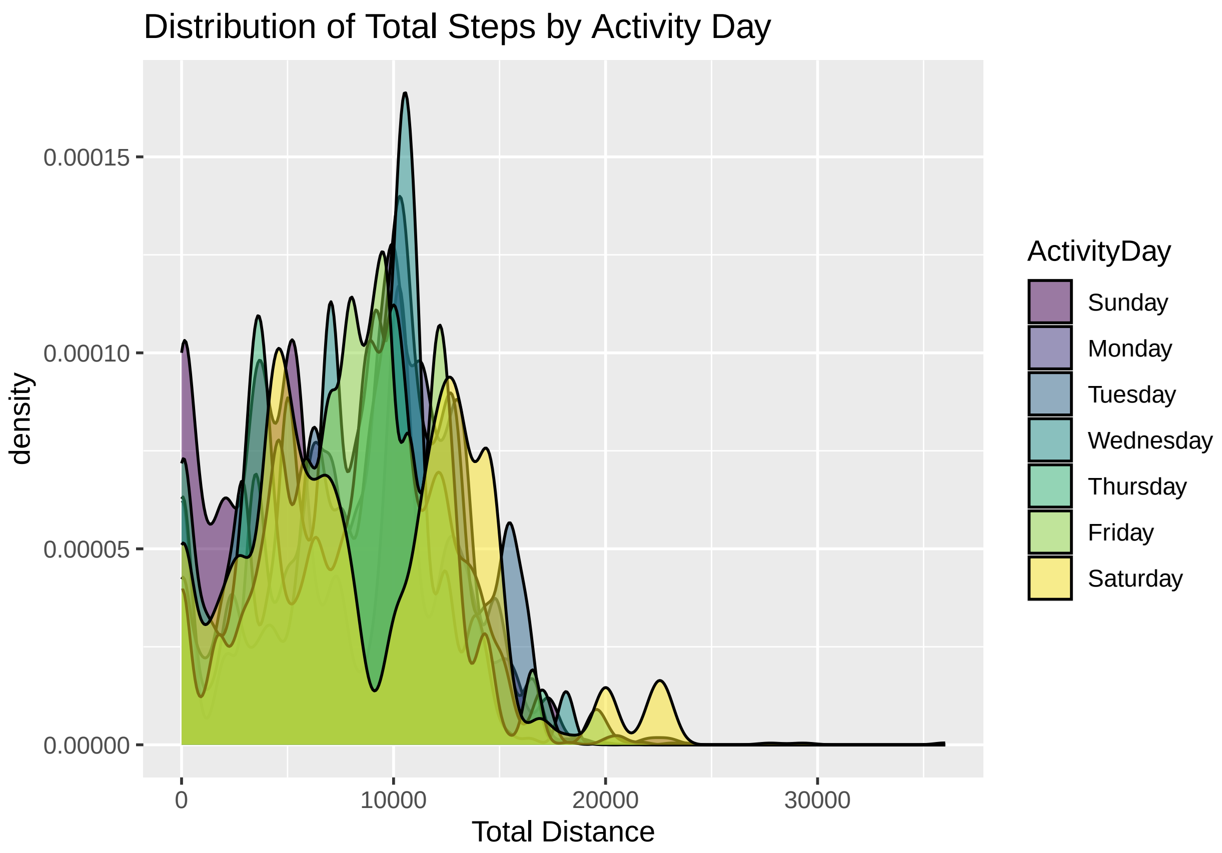
```
## Warning: There was 1 warning in `summarize()`.
## i In argument: `across(...)`.
## Caused by warning:
## ! The `...` argument of `across()` is deprecated as of dplyr 1.1.0.
## Supply arguments directly to `.fns` through an anonymous function instead.
##
##   # Previously
##   across(a:b, mean, na.rm = TRUE)
##
##   # Now
##   across(a:b, \(x) mean(x, na.rm = TRUE))
```

```
## # A tibble: 1 x 8
##   TotalSteps TotalDistance SedentaryActiveDistance SedentaryMinutes
##        <dbl>         <dbl>                   <dbl>            <dbl>
## 1      8124.          5.75                0.000740             806.
## # i 4 more variables: LightActiveDistance <dbl>, LightlyActiveMinutes <dbl>,
## #   ModeratelyActiveDistance <dbl>, FairlyActiveMinutes <dbl>
```

```
#  Distribution of Total Distance by Activity Day

ggplot(join_sleep_dailyactivity, aes(x = TotalDistance)) +
geom_density(adjust = 0.5, alpha = 0.5, aes(fill = ActivityDay))+
  labs(x = "Total Distance",
       title = "Distribution of Total Distance by Activity Day")
```

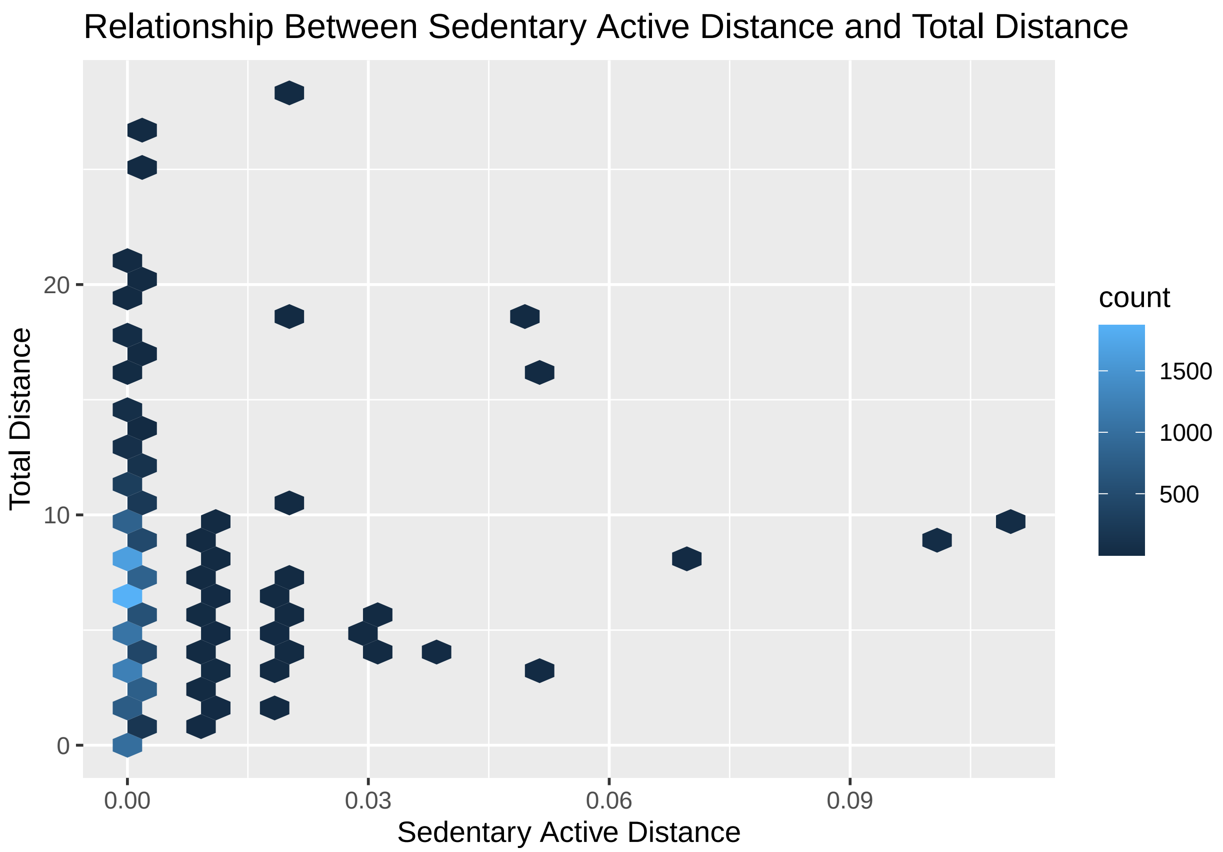


```
#  Distribution of Total Steps by Activity Day
ggplot(join_sleep_dailyactivity, aes(x = TotalSteps)) +
geom_density(adjust = 0.5, alpha = 0.5, aes(fill = ActivityDay))+
  labs(x = "Total Distance",
       title = "Distribution of Total Steps by Activity Day")
```



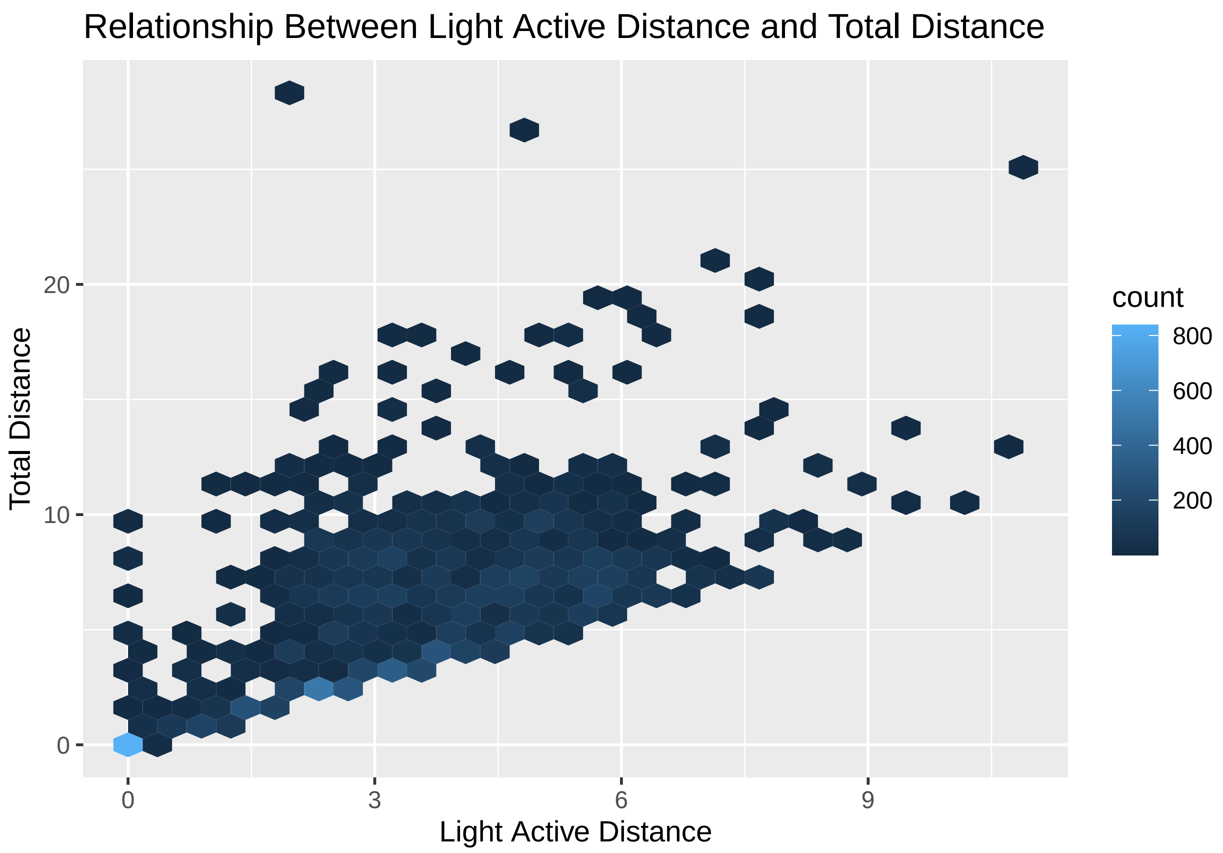
Plot Sedentary Active Distance and Total Distance

```
ggplot(join_sleep_dailyactivity, aes(x = SedentaryActiveDistance, y = TotalDistance))+
  geom_hex()+
  labs(x = "Sedentary Active Distance",
       y = "Total Distance",
       title = "Relationship Between Sedentary Active Distance and Total Distance")
```



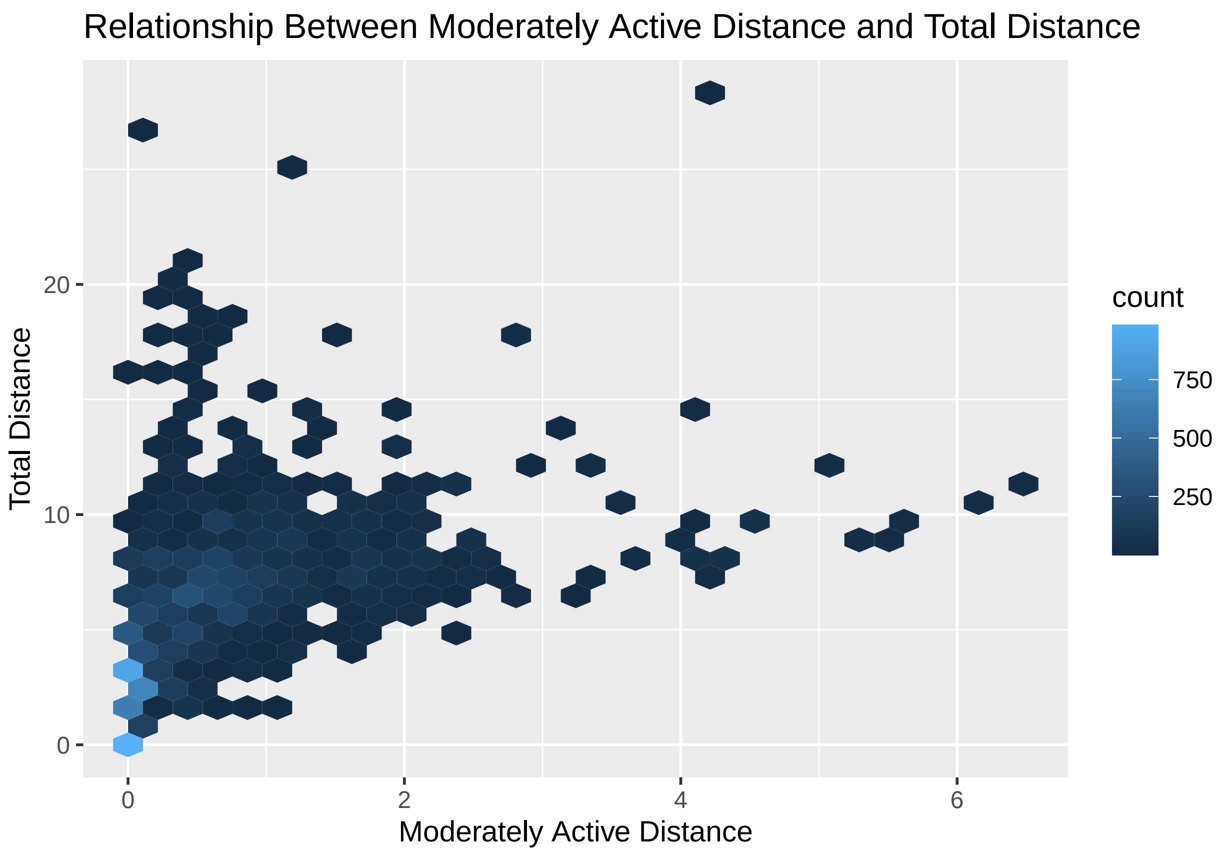
Plot Light Active Distance and Total Distance

```
ggplot(join_sleep_dailyactivity, aes(x = LightActiveDistance, y = TotalDistance))+
  geom_hex()+
  labs(x = "Light Active Distance",
       y = "Total Distance",
       title = "Relationship Between Light Active Distance and Total Distance")
```



Plot Moderately Active Distance and Total Distance

```
ggplot(join_sleep_dailyactivity, aes(x = ModeratelyActiveDistance, y = TotalDistance))+
  geom_hex()+
  labs(x = "Moderately Active Distance",
       y = "Total Distance",
       title = "Relationship Between Moderately Active Distance and Total Distance")
```



Summary Statistics: Distance, Activity Minutes, Calories, Sleep

```
join_sleep_dailyactivity %>%
  select(TotalSteps:Calories, TotalMinutesAsleep, TotalTimeInBed) %>%
  summary()
```

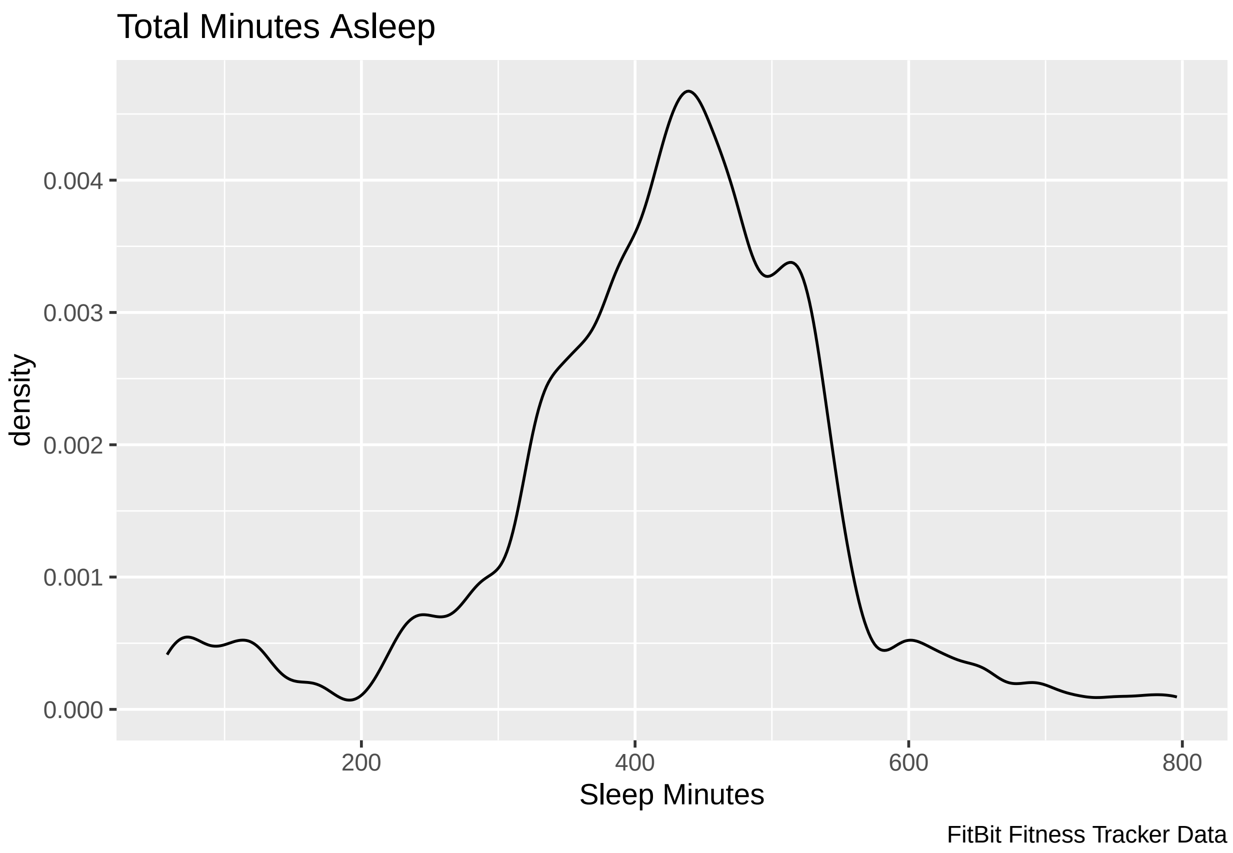
```
##    TotalSteps     TotalDistance    TrackerDistance  LoggedActivitiesDistance
##  Min.   :    0   Min.   : 0.000   Min.   : 0.000   Min.   :0.0000
##  1st Qu.: 4676   1st Qu.: 3.180   1st Qu.: 3.180   1st Qu.:0.0000
##  Median : 8582   Median : 6.120   Median : 6.120   Median :0.0000
##  Mean   : 8124   Mean   : 5.745   Mean   : 5.738   Mean   :0.1211
##  3rd Qu.:11207   3rd Qu.: 7.920   3rd Qu.: 7.890   3rd Qu.:0.0000
##  Max.   :36019   Max.   :28.030   Max.   :28.030   Max.   :4.9421
##
##  VeryActiveDistance ModeratelyActiveDistance LightActiveDistance
##  Min.   : 0.000     Min.   :0.0000           Min.   : 0.000
##  1st Qu.: 0.000     1st Qu.:0.0000           1st Qu.: 2.370
##  Median : 0.530     Median :0.4000           Median : 3.540
##  Mean   : 1.406     Mean   :0.7273           Mean   : 3.547
##  3rd Qu.: 2.310     3rd Qu.:1.0000           3rd Qu.: 4.850
##  Max.   :21.920     Max.   :6.4800           Max.   :10.710
##
##  SedentaryActiveDistance VeryActiveMinutes FairlyActiveMinutes
##  Min.   :0.0000000       Min.   :  0.00    Min.   :  0.00
##  1st Qu.:0.0000000       1st Qu.:  0.00    1st Qu.:  0.00
##  Median :0.0000000       Median :  8.00    Median : 10.00
##  Mean   :0.0007405       Mean   : 23.93    Mean   : 17.23
##  3rd Qu.:0.0000000       3rd Qu.: 36.00    3rd Qu.: 24.00
##  Max.   :0.1100000       Max.   :210.00    Max.   :143.00
##
##  LightlyActiveMinutes SedentaryMinutes    Calories    TotalMinutesAsleep
##  Min.   :  0.0        Min.   :   0.0   Min.   :   0   Min.   : 58.0
##  1st Qu.:145.0        1st Qu.: 660.0   1st Qu.:1783   1st Qu.:361.0
##  Median :201.0        Median : 738.0   Median :2162   Median :432.0
##  Mean   :200.2        Mean   : 805.9   Mean   :2329   Mean   :419.4
##  3rd Qu.:258.0        3rd Qu.: 878.0   3rd Qu.:2862   3rd Qu.:492.0
##  Max.   :518.0        Max.   :1440.0   Max.   :4900   Max.   :796.0
##                                                       NA's   :227
##  TotalTimeInBed
##  Min.   : 61.0
##  1st Qu.:402.0
##  Median :463.0
##  Mean   :458.4
##  3rd Qu.:526.0
##  Max.   :961.0
##  NA's   :227
```

```
#  Average sedentary minutes were greater than time spent being lightly active, fairly active, and very
```

Time trends by day

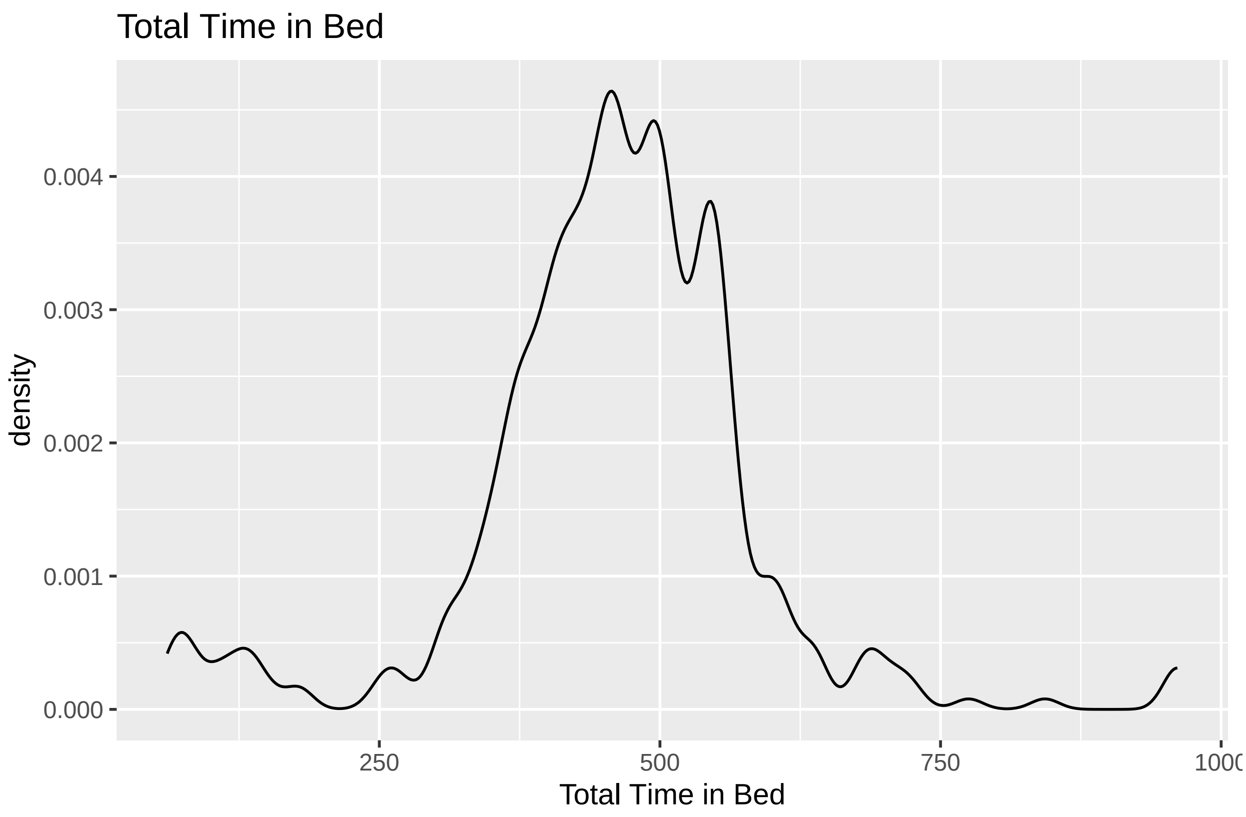
```
ggplot(join_sleep_dailyactivity, aes(x = TotalMinutesAsleep)) +
  geom_density() +
  labs(x = "Sleep Minutes",
       title = "Total Minutes Asleep",
       caption = "FitBit Fitness Tracker Data")
```

```
## Warning: Removed 227 rows containing non-finite values (`stat_density()`).
```



```r
ggplot(join_sleep_dailyactivity, aes(x = TotalTimeInBed))+
  geom_density()+
  labs(x = "Total Time in Bed",
       title = "Total Time in Bed",
       caption = "Source: FitBit Fitness Tracker Data")
```

```
## Warning: Removed 227 rows containing non-finite values (‘stat_density()‘).
```



Source: FitBit Fitness Tracker Data

Summarize Activity by User Id and Day of Week

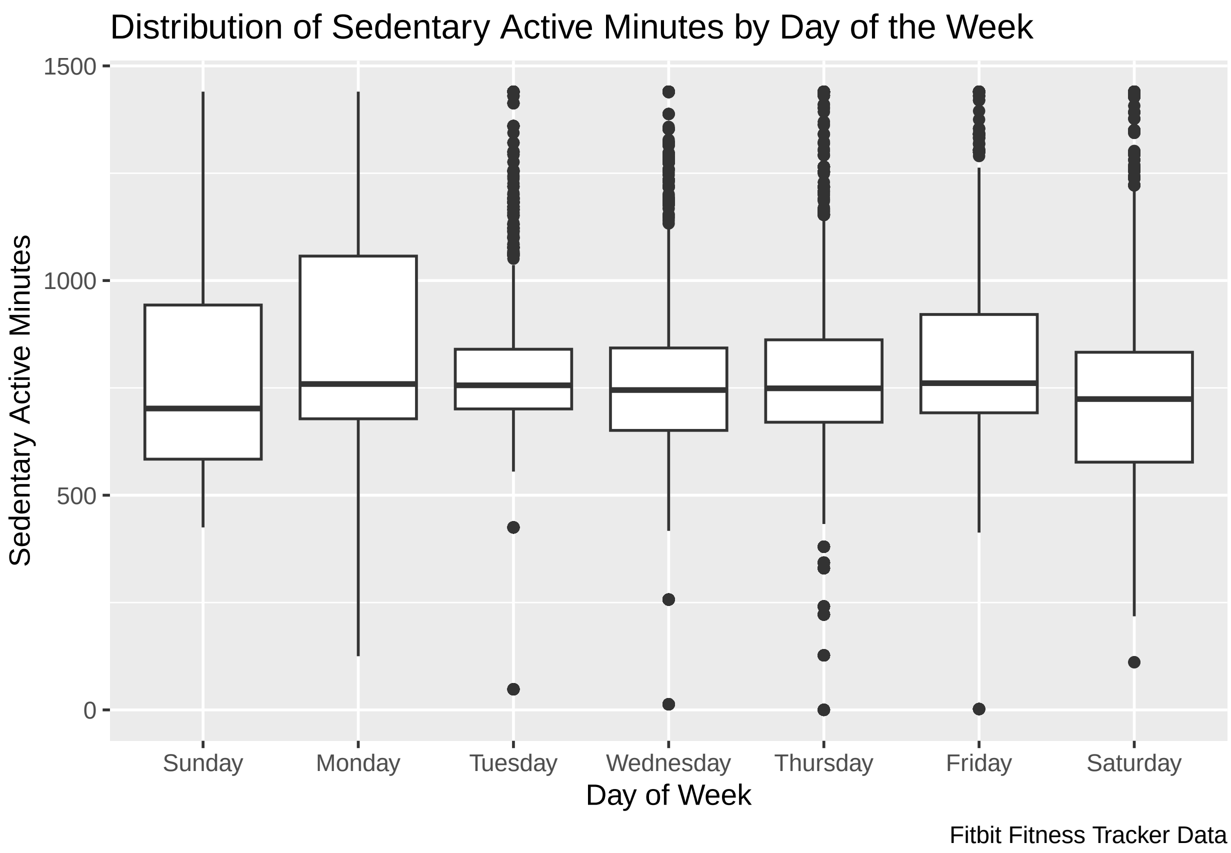
```
#  Look at day of the week to identify which days of the week users are more likely to be active.  Beca

join_sleep_dailyactivity %>%
  group_by(Id, ActivityDay) %>%
  summarize(across(.fns = mean, .cols = where(is.double)))
```

```
## ‘summarise()‘ has grouped output by ’Id’. You can override using the ‘.groups‘
## argument.
```

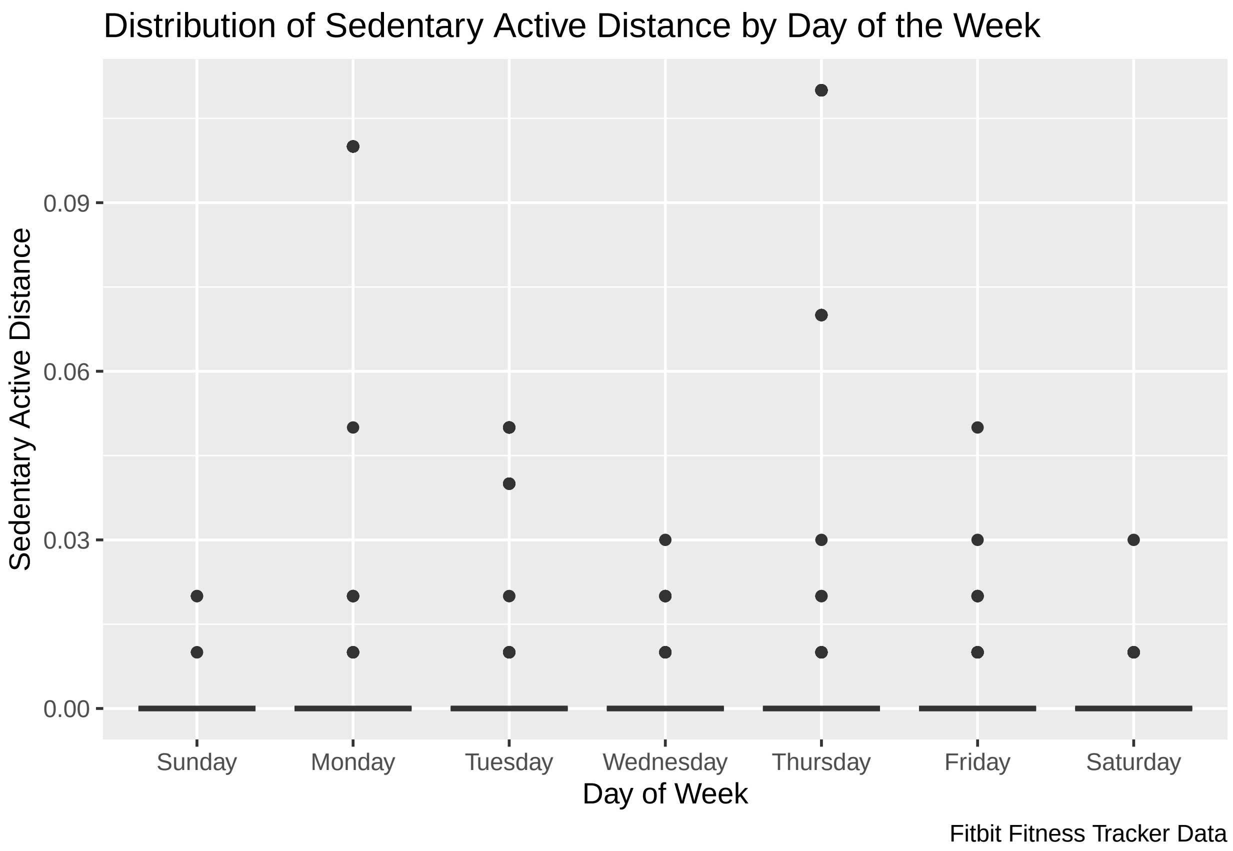
```
## # A tibble: 228 x 20
## # Groups:   Id [33]
##            Id ActivityDay ActivityDate TotalSteps TotalDistance TrackerDistance
##         <dbl> <ord>       <date>            <dbl>         <dbl>           <dbl>
##  1 1503960366 Sunday      2016-04-27       10102.          6.57            6.57
##  2 1503960366 Monday      2016-04-28       13781.          8.96            8.96
##  3 1503960366 Tuesday     2016-04-26       13947.          8.92            8.92
##  4 1503960366 Wednesday   2016-04-27       12657.          8.23            8.23
##  5 1503960366 Thursday    2016-04-28        9501.          6.10            6.10
##  6 1503960366 Friday      2016-04-25       11466.          7.40            7.40
##  7 1503960366 Saturday    2016-04-26       13426.          8.54            8.54
##  8 1624580081 Sunday      2016-04-27       12924.          9.57            9.57
##  9 1624580081 Monday      2016-04-28        6480          4.42            4.42
## 10 1624580081 Tuesday     2016-04-26        3795.         2.47            2.47
## # i 218 more rows
## # i 14 more variables: LoggedActivitiesDistance <dbl>,
## #   VeryActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
## #   LightActiveDistance <dbl>, SedentaryActiveDistance <dbl>,
## #   VeryActiveMinutes <dbl>, FairlyActiveMinutes <dbl>,
## #   LightlyActiveMinutes <dbl>, SedentaryMinutes <dbl>, Calories <dbl>,
## #   SleepDay <dttm>, TotalSleepRecords <dbl>, TotalMinutesAsleep <dbl>, ...
```

Activity Distribution by Day of Week

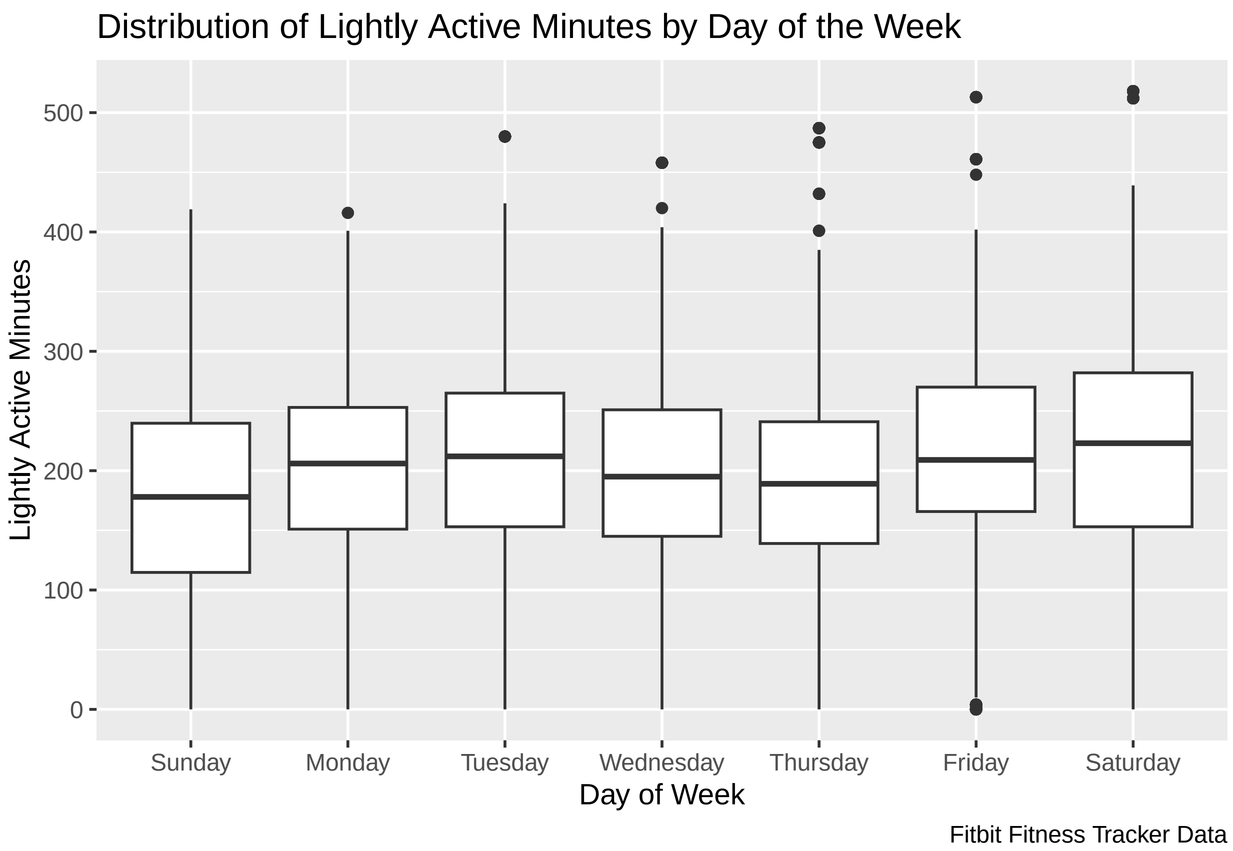
```
ggplot(join_sleep_dailyactivity, aes(x = ActivityDay, y = SedentaryMinutes))+
  geom_boxplot()+
  labs(x = "Day of Week",
       y = "Sedentary Active Minutes",
       title = "Distribution of Sedentary Active Minutes by Day of the Week",
       caption = "Fitbit Fitness Tracker Data")
```



```
ggplot(join_sleep_dailyactivity, aes(x = ActivityDay, y = SedentaryActiveDistance))+
  geom_boxplot()+
  labs(x = "Day of Week",
       y = "Sedentary Active Distance",
       title = "Distribution of Sedentary Active Distance by Day of the Week",
       caption = "Fitbit Fitness Tracker Data")
```



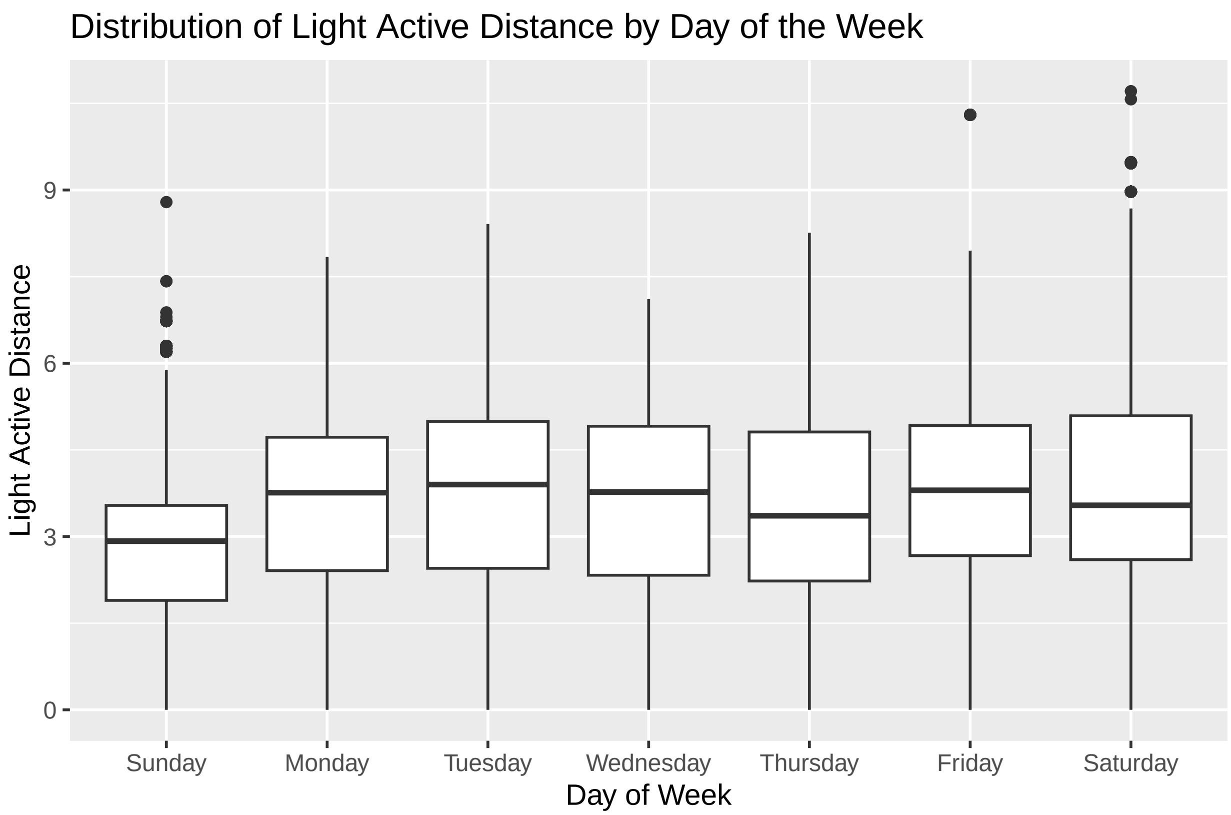
```
ggplot(join_sleep_dailyactivity, aes(x = ActivityDay, y = LightlyActiveMinutes))+
  geom_boxplot()+
  labs(x = "Day of Week",
       y = "Lightly Active Minutes",
       title = "Distribution of Lightly Active Minutes by Day of the Week",
       caption = "Fitbit Fitness Tracker Data")
```



Distribution of Lightly Active Minutes by Day of the Week

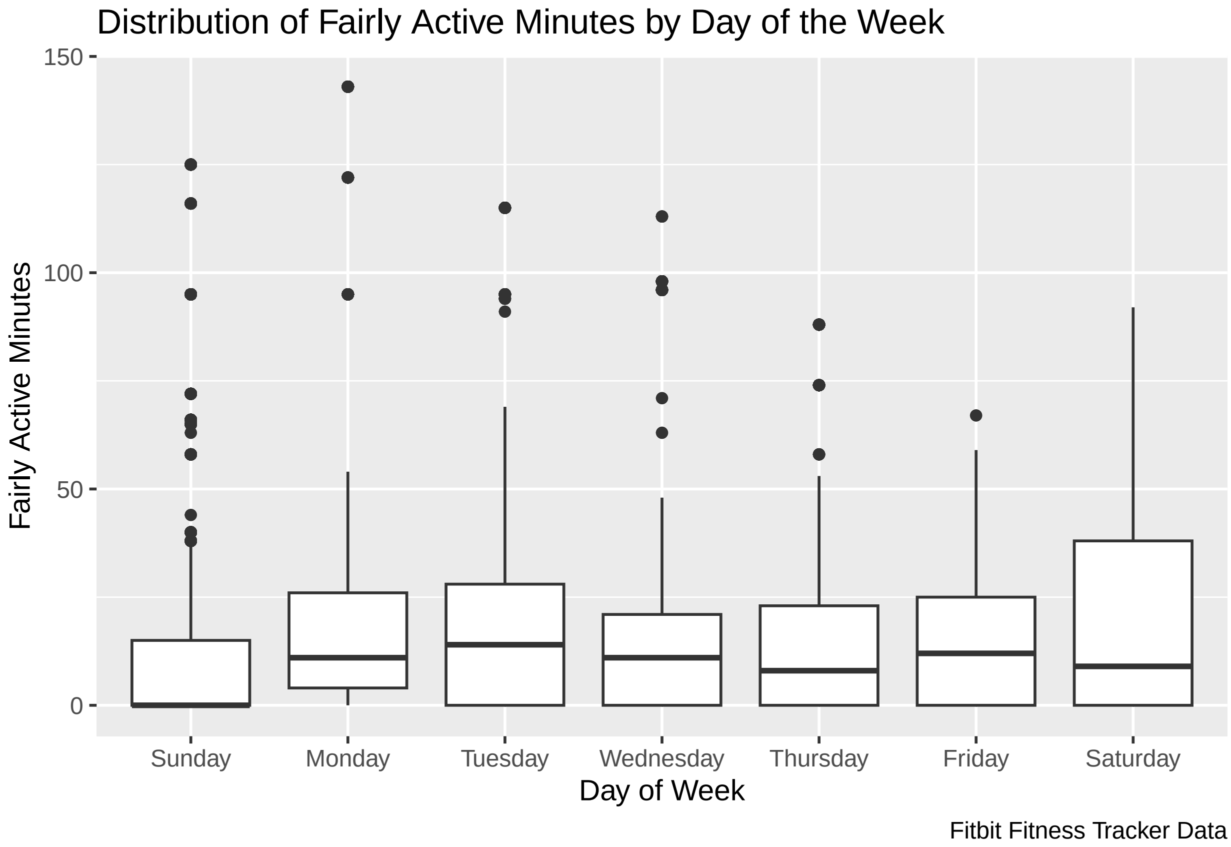
Fitbit Fitness Tracker Data

```r
ggplot(join_sleep_dailyactivity, aes(x = ActivityDay, y = LightActiveDistance))+
  geom_boxplot()+
  labs(x = "Day of Week",
       y = "Light Active Distance",
       title = "Distribution of Light Active Distance by Day of the Week",
       caption = "Fitbit Fitness Tracker Data")
```

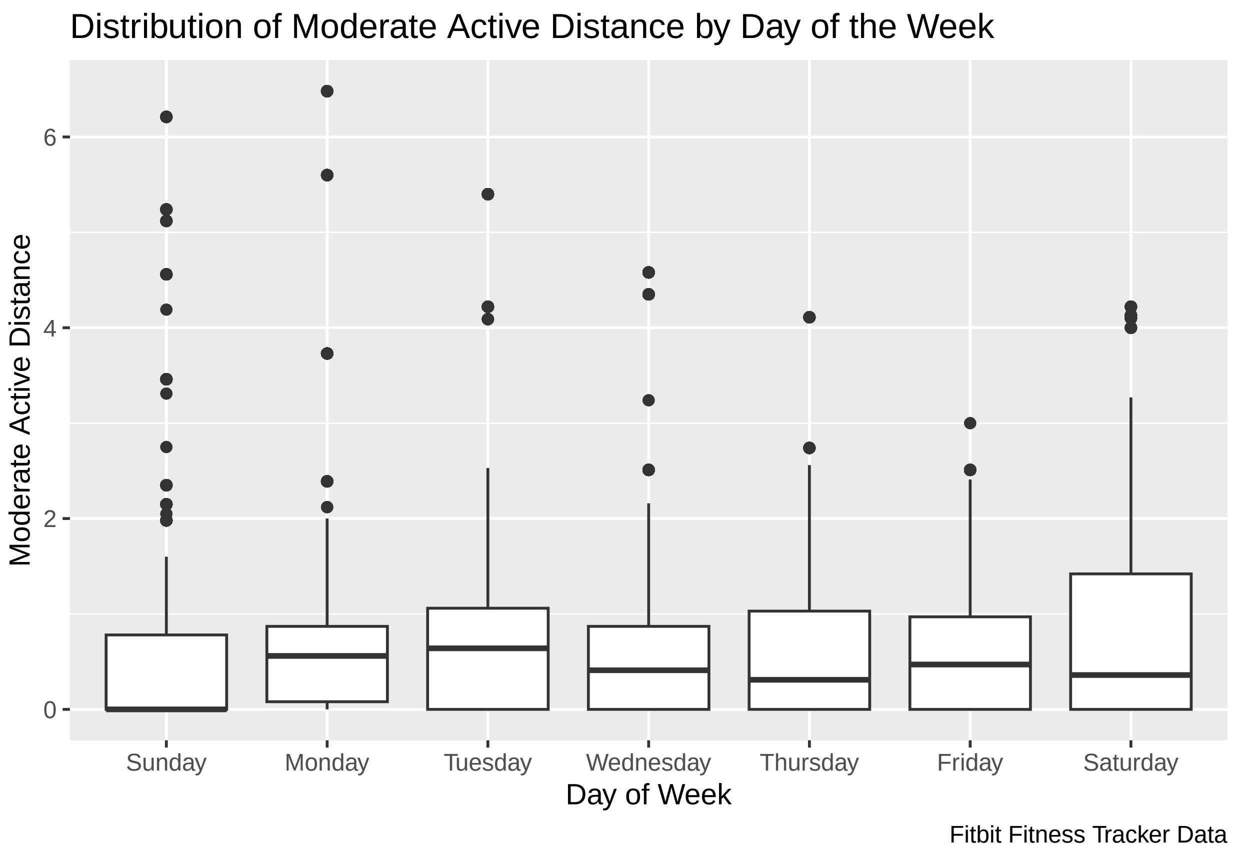


Fitbit Fitness Tracker Data

```
ggplot(join_sleep_dailyactivity, aes(x = ActivityDay, y = FairlyActiveMinutes))+
  geom_boxplot()+
  labs(x = "Day of Week",
       y = "Fairly Active Minutes",
       title = "Distribution of Fairly Active Minutes by Day of the Week",
       caption = "Fitbit Fitness Tracker Data")
```

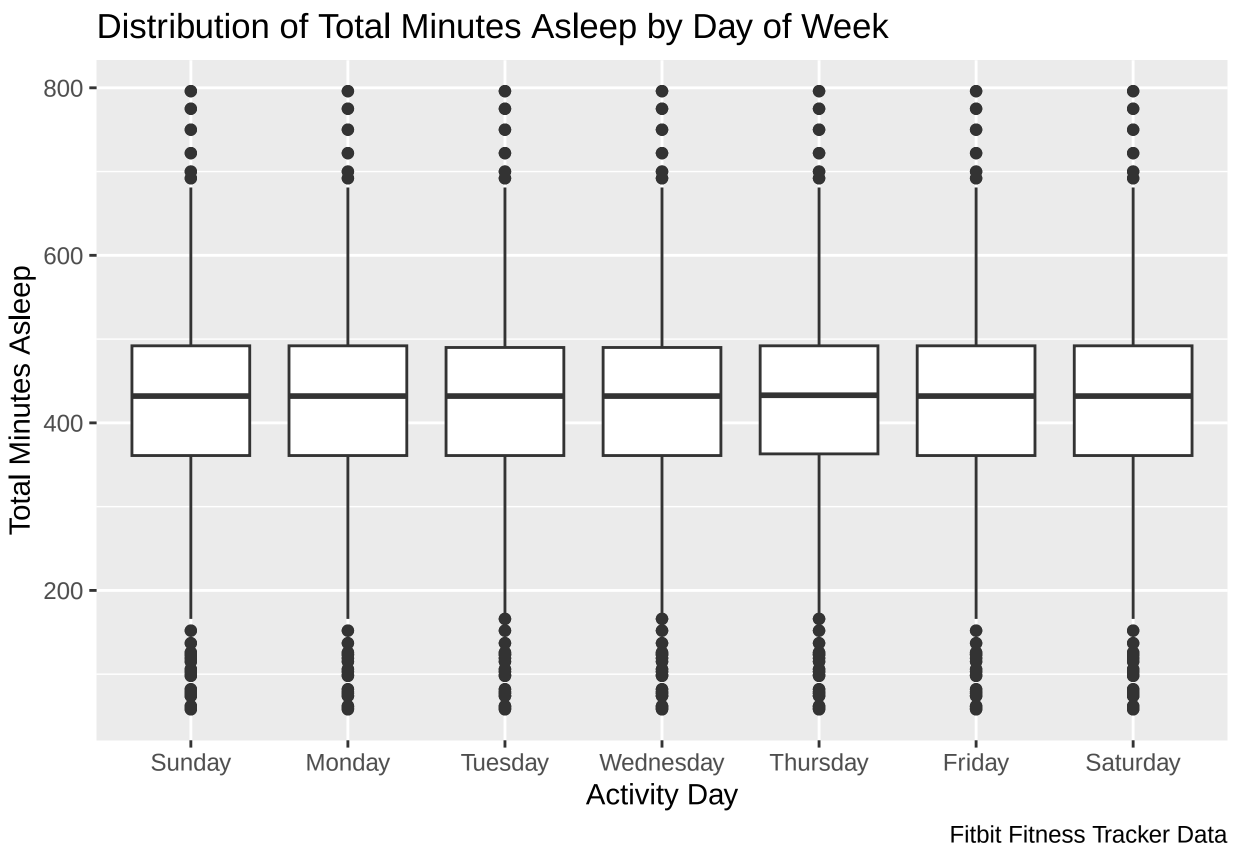


```r
#  Established Fairly Active Minutes as the respective equivalent for Moderately Active Distance

ggplot(join_sleep_dailyactivity, aes(x = ActivityDay, y = ModeratelyActiveDistance))+
  geom_boxplot()+
  labs(x = "Day of Week",
       y = "Moderate Active Distance",
       title = "Distribution of Moderate Active Distance by Day of the Week",
       caption = "Fitbit Fitness Tracker Data")
```



```
ggplot(join_sleep_dailyactivity, aes(x = ActivityDay, y = TotalMinutesAsleep))+
  geom_boxplot()+
  labs(x = "Activity Day",
       y = "Total Minutes Asleep",
       title = "Distribution of Total Minutes Asleep by Day of Week",
       caption = "Fitbit Fitness Tracker Data")
```

```
## Warning: Removed 227 rows containing non-finite values (‘stat_boxplot()‘).
```

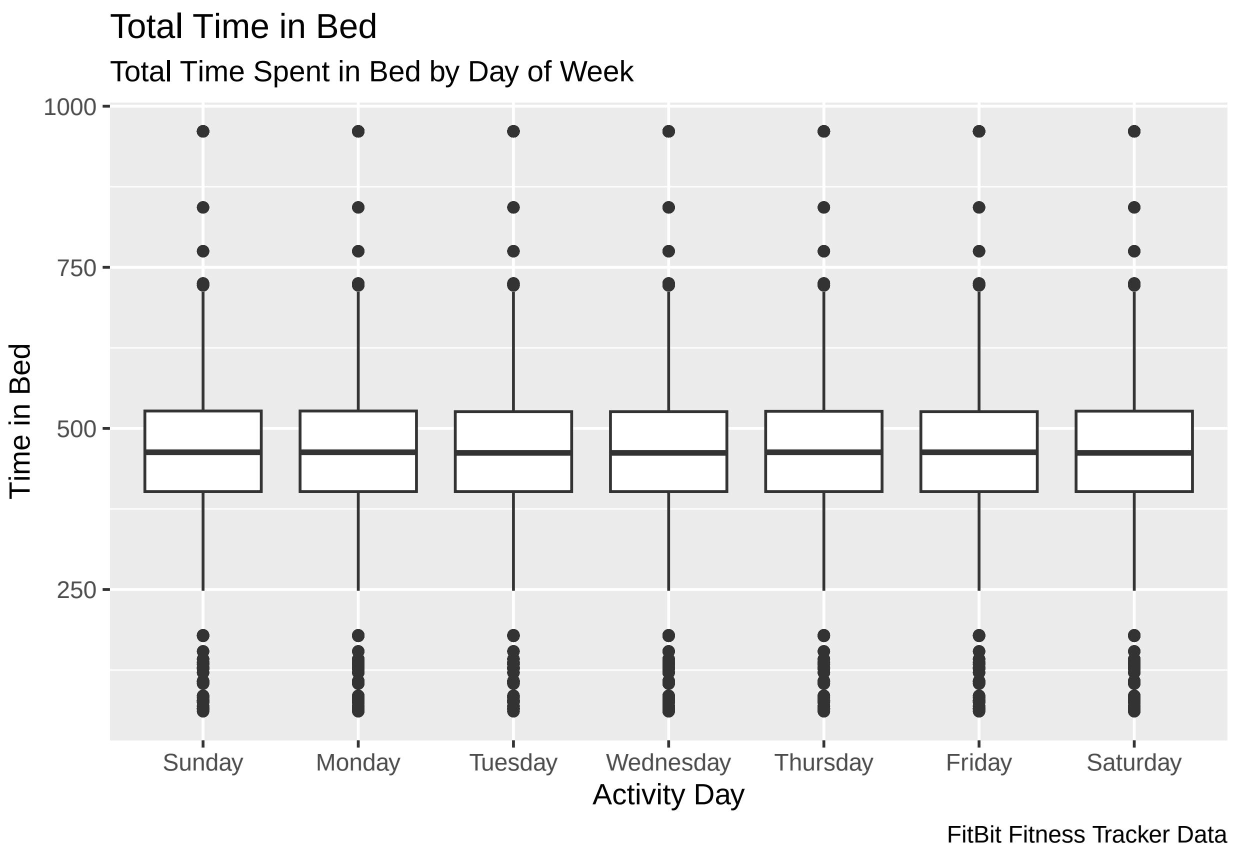


Distribution of Total Minutes Asleep by Day of Week

Fitbit Fitness Tracker Data

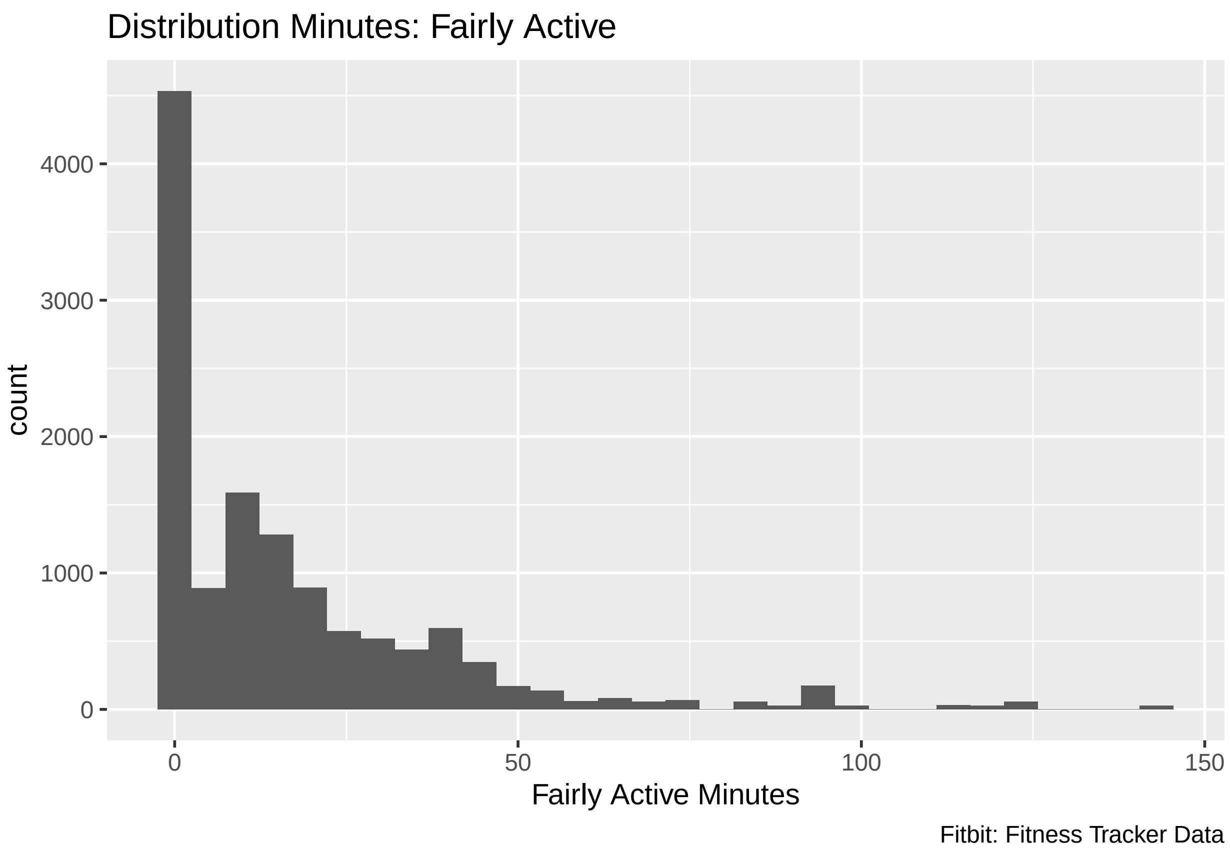
```
ggplot(join_sleep_dailyactivity, aes(x = ActivityDay, y = TotalTimeInBed))+
  geom_boxplot()+
  labs(x = "Activity Day",
       y = "Time in Bed",
       title = "Total Time in Bed",
       subtitle = "Total Time Spent in Bed by Day of Week",
       caption = "FitBit Fitness Tracker Data")
```

```
## Warning: Removed 227 rows containing non-finite values (‘stat_boxplot()‘).
```



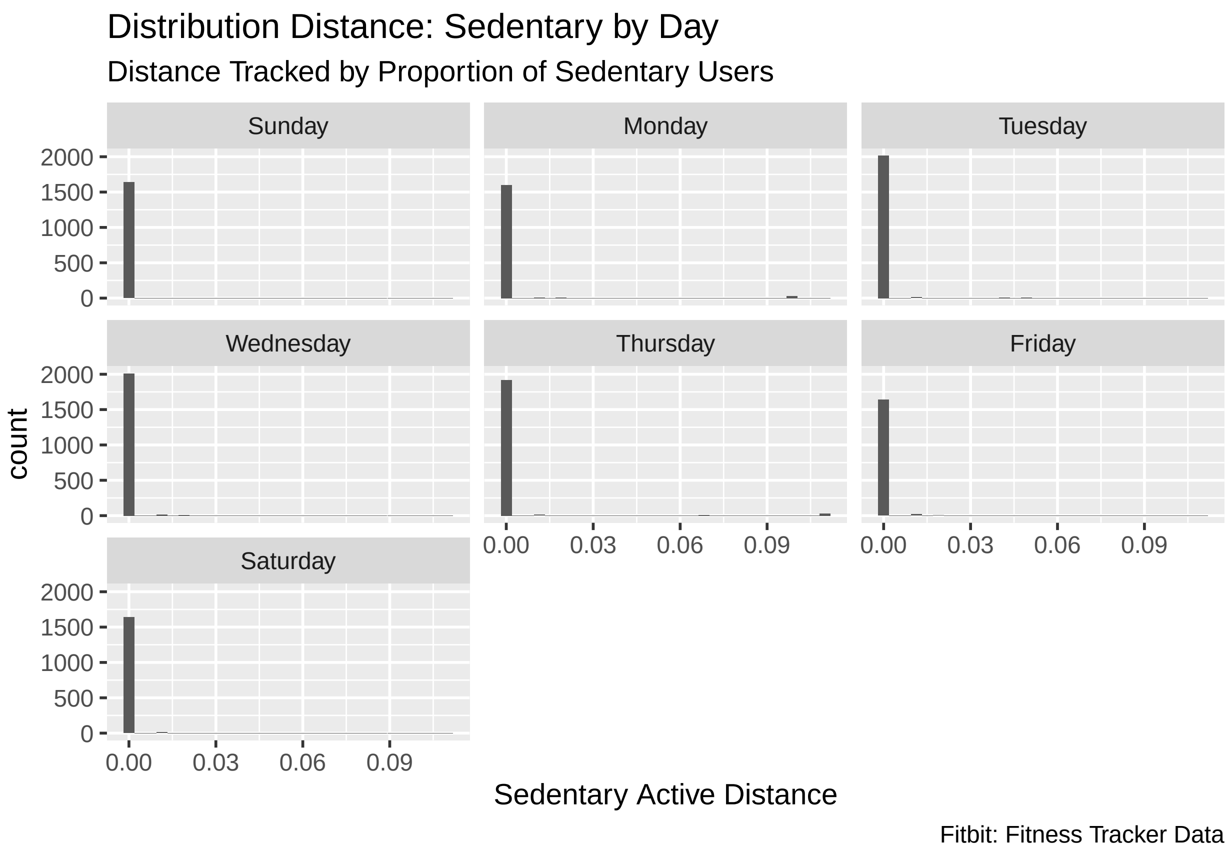
```
ggplot(join_sleep_dailyactivity, aes(x = FairlyActiveMinutes))+
  geom_histogram()+
  labs(x = "Fairly Active Minutes",
       title = "Distribution Minutes: Fairly Active ",
       caption = "Fitbit: Fitness Tracker Data")
```

```
## ‘stat_bin()‘ using ‘bins = 30‘. Pick better value with ‘binwidth‘.
```



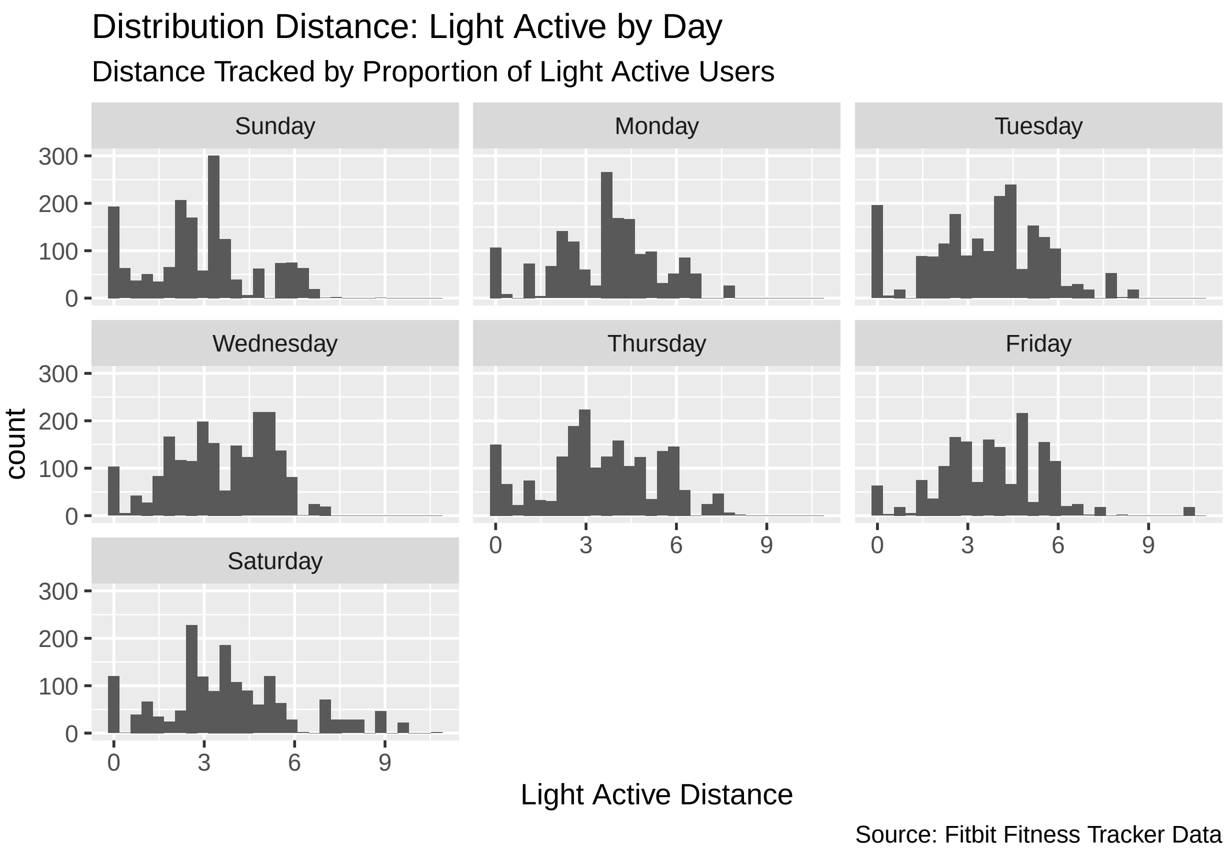
```r
ggplot(join_sleep_dailyactivity, aes(x = SedentaryActiveDistance))+
  geom_histogram()+
  facet_wrap(~ActivityDay)+
  labs(x = "Sedentary Active Distance",
       title = "Distribution Distance: Sedentary by Day",
       subtitle = "Distance Tracked by Proportion of Sedentary Users",
       caption = "Fitbit: Fitness Tracker Data")
```

```
## ‘stat_bin()‘ using ‘bins = 30‘. Pick better value with ‘binwidth‘.
```



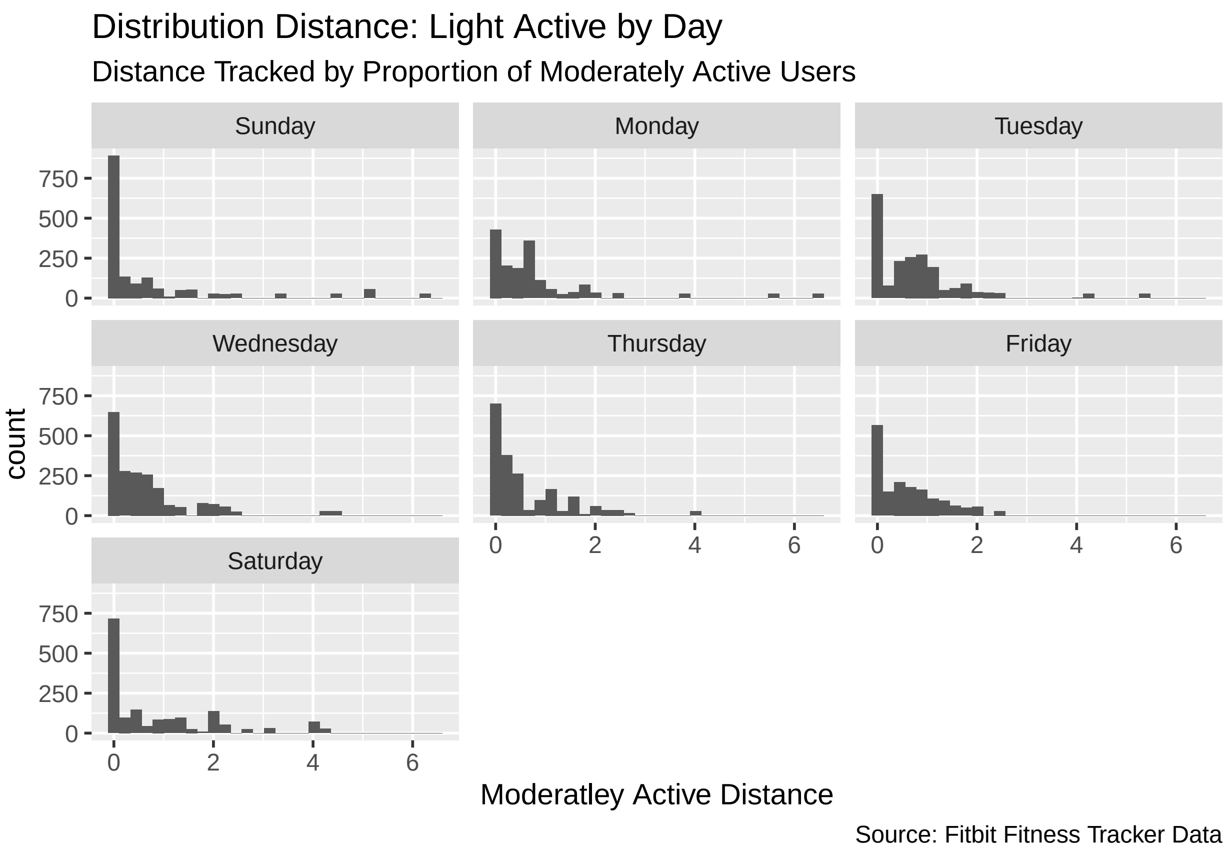
```
ggplot(join_sleep_dailyactivity, aes(x = LightActiveDistance))+
  geom_histogram() +
  facet_wrap(~ActivityDay) +
  labs(x = "Light Active Distance",
       title = "Distribution Distance: Light Active by Day",
       subtitle = "Distance Tracked by Proportion of Light Active Users",
       caption = "Source: Fitbit Fitness Tracker Data")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```r
ggplot(join_sleep_dailyactivity, aes(x = ModeratelyActiveDistance))+
  geom_histogram() +
  facet_wrap(~ActivityDay) +
  labs(x = "Moderatley Active Distance",
       title = "Distribution Distance: Light Active by Day",
       subtitle = "Distance Tracked by Proportion of Moderately Active Users",
       caption = "Source: Fitbit Fitness Tracker Data")
```

```
## ‘stat_bin()‘ using ‘bins = 30‘. Pick better value with ‘binwidth‘.
```



Distribution Distance: Light Active by Day

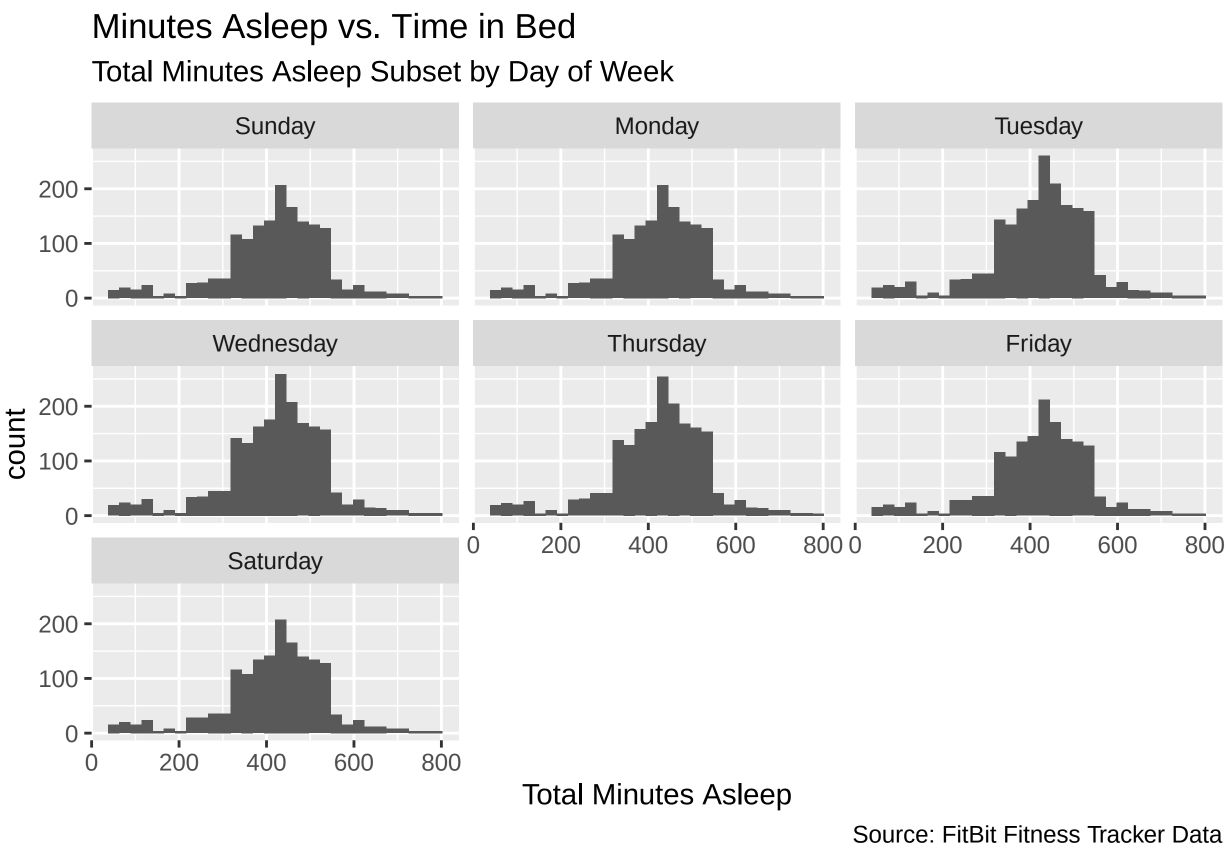
Distance Tracked by Proportion of Moderately Active Users

Source: Fitbit Fitness Tracker Data

```r
ggplot(join_sleep_dailyactivity, aes(x = TotalMinutesAsleep))+
  geom_histogram()+
  facet_wrap(~ActivityDay)+
  labs(x = "Total Minutes Asleep",
       title = "Minutes Asleep vs. Time in Bed",
       subtitle = "Total Minutes Asleep Subset by Day of Week",
       caption = "Source: FitBit Fitness Tracker Data")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 227 rows containing non-finite values (`stat_bin()`).
```


Minutes Asleep vs. Time in Bed

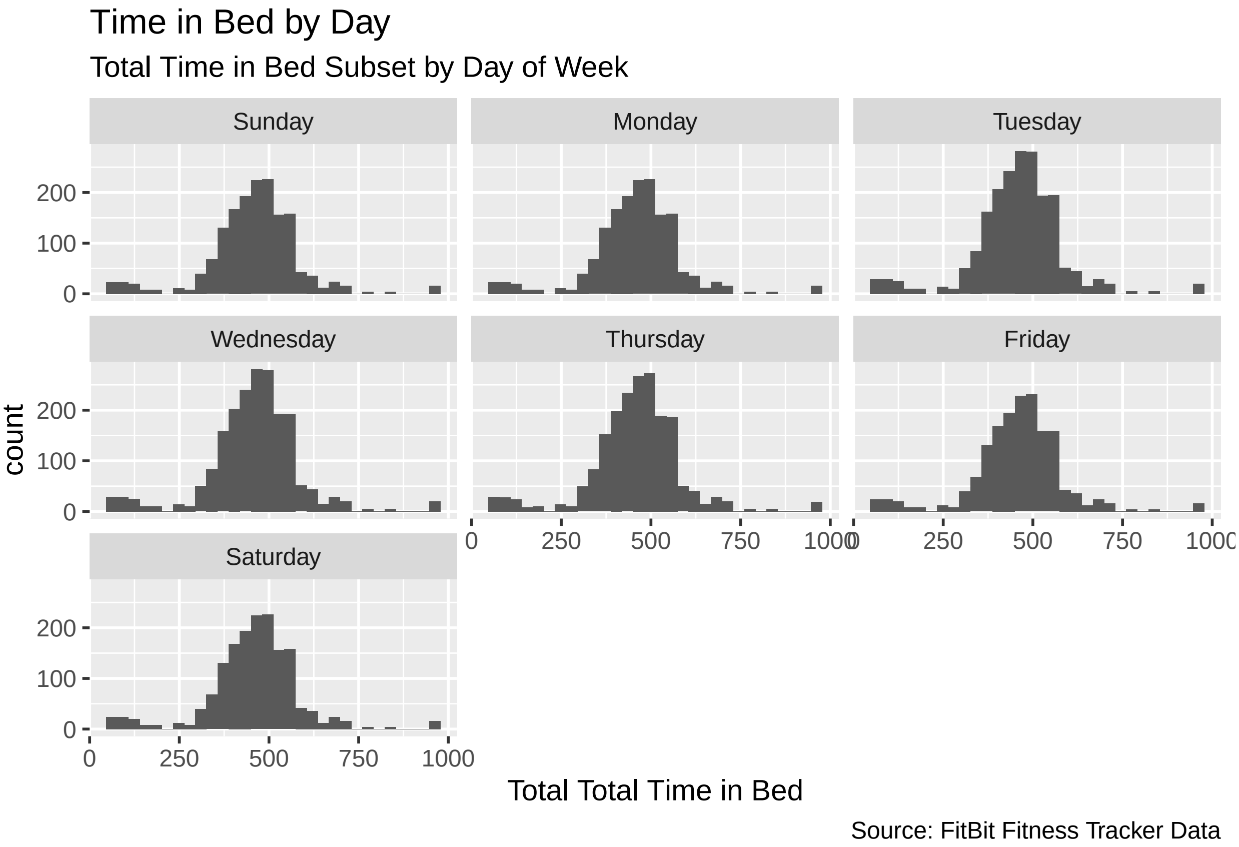
Total Minutes Asleep Subset by Day of Week

Source: FitBit Fitness Tracker Data

```
ggplot(join_sleep_dailyactivity, aes(x = TotalTimeInBed))+
  geom_histogram()+
  facet_wrap(~ActivityDay)+
  labs(x = "Total Total Time in Bed",
       title = "Time in Bed by Day",
       subtitle = "Total Time in Bed Subset by Day of Week",
       caption = "Source: FitBit Fitness Tracker Data")
```

```
## ‘stat_bin()‘ using ‘bins = 30‘. Pick better value with ‘binwidth‘.
```

```
## Warning: Removed 227 rows containing non-finite values (‘stat_bin()‘).
```



Time in Bed by Day

Total Time in Bed Subset by Day of Week

Source: FitBit Fitness Tracker Data

Summary ##Users engaged in more light active activity compared to sedentary and moderately active activity in terms of distance. Bellabeat could use this as an opportunity to market toward sedentary users and position the messaging that the app could assist in making small, but meaningful steps to increase activity.