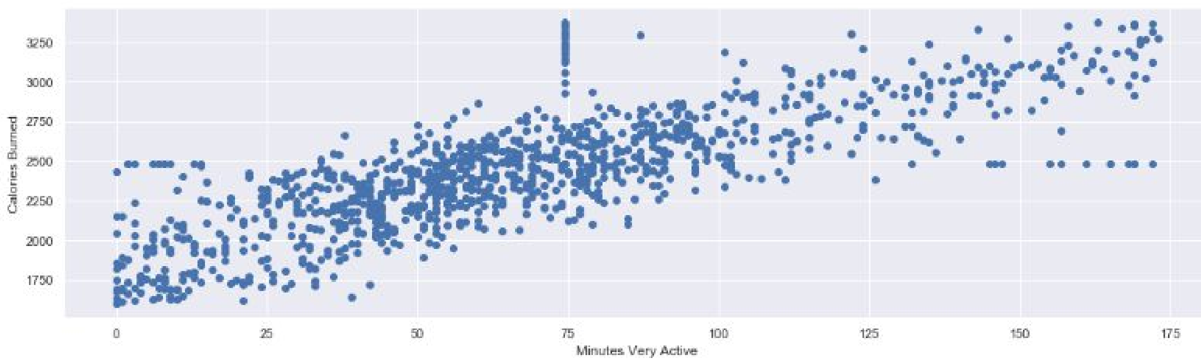
Capstone 1: Inferential Statistics

The data made available by Fitbit contains information regarding a user's activity and sleep data. For Tracy in particular, she also logs her food consumption onto Fitbit's app, so data about her food is also extractable. Most of the data used are her activity and sleep data, as they are most "complete" and provide information some insight to her health and lifestyle.

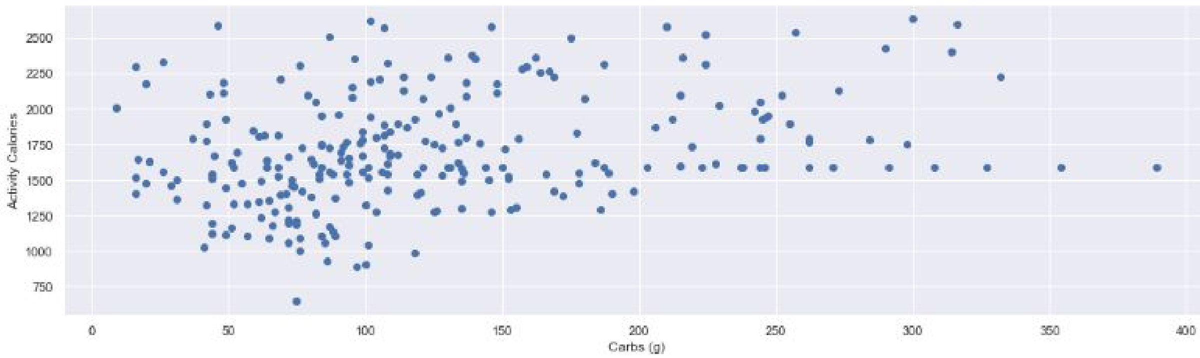



```
scipy_r, scipy_p = stats.pearsonr(activities['Minutes Very Active'], \
                                  activities['Calories Burned'])

print("Scipy's correlation coefficient:", scipy_r)
print("Scipy's p-value:", scipy_p)

Scipy's correlation coefficient: 0.7858467996449217
Scipy's p-value: 7.142083581912041e-233
```

Two variables that obviously have a strong correlation are "Minutes Very Active" and "Calories Burned", as exercising burns calories. Creating a scatter plot with "Minutes Very Active" as the independent variable and "Calories Burned" as the dependent variable, an upward trend is formed. This means that the longer Tracy is active, the more calories she burns. It is important to note that there is a separate variable called "Activity Calories" in the dataset, but was not chosen as it is more interesting to see if there are other factors that contribute to calorie expenditure. The calculated pearson correlation coefficient is 0.786, indicating a moderately strong linear relationship between active minutes and daily burned calories. Since the coefficient is not enough to represent an almost linear relationship, there are still other factors that contribute to the total amount of calories Tracy burns in a day.

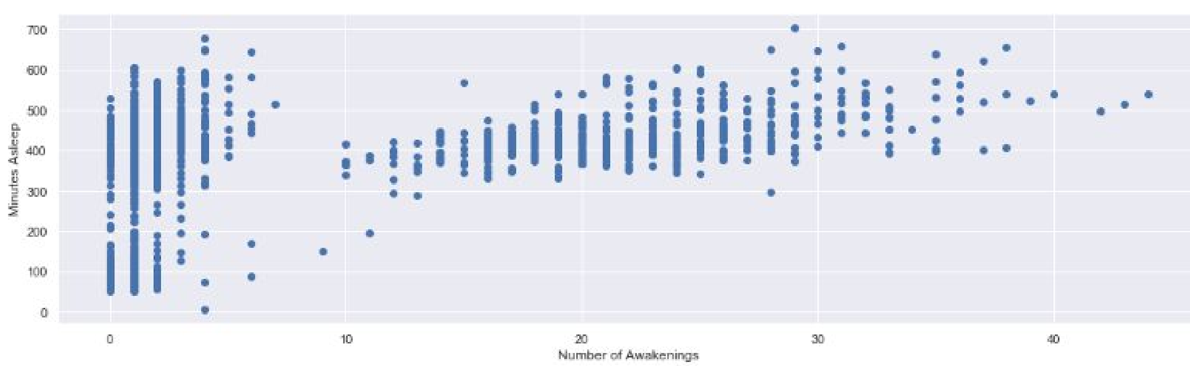



```
scipy_r, scipy_p = stats.pearsonr(df1['Activity Calories'], df1['Carbs (g)'])

print("Scipy's correlation coefficient:", scipy_r)
print("Scipy's p-value:", scipy_p)
```

```
Scipy's correlation coefficient: 0.301696397595359
Scipy's p-value: 9.62446731823555e-07
```

Other variables that were tested for correlation include “Activity Calories” with “Carb (g)” and “Number of Awakenings” with “Minutes Asleep”. With the former, the correlation was tested to determine if Tracy is an athlete who tends to load up on carbs on her activity-heavy days. The r-value is calculated to be 0.3, which is a somewhat weak correlation, but is still a positive one nonetheless. One thing to note is that the food data is incomplete (e.g., there are instances where dinner is not recorded). It is reasonable to assume that the r-value might actually be greater if all food intake is logged, but with the data we have available, it is still plausible to say there is a weak positive relationship between how many calories are burned from exercising and how much carbs Tracy eats.




```
scipy_r, scipy_p = stats.spearmanr(sleep['Number of Awakenings'], sleep['Minutes Asleep'])

print("Scipy's correlation coefficient:", scipy_r)
print("Scipy's p-value:", scipy_p)
```

```
Scipy's correlation coefficient: 0.38107089058809945
Scipy's p-value: 1.0564244602304605e-46
```

When looking at Tracy's sleep data, it was intriguing -- maybe even shocking -- to see nights where she wakes up more than 30 times. According to a study, the average number of awakenings per night for an adult is around 6. Tracy's number is quintuple the average, which is astounding. However, the r-value for "Number of Awakenings" and "Minutes Asleep" is 0.381. This indicates a weak, but nonetheless positive correlation. The more awakenings Tracy experiences, the longer her sleeps. It would be more worrying if the upward trend is absent because that would indicate a low quantity in sleep, which would be a health concern.

Only the pearson correlation test was performed, as the above three examples asks how strongly related the two variables are to each other. As an afterthought, it is possible to separate her data into winter and summer and see how her stats compare. For now, we are able to assume that Tracy is an active individual who has trouble staying asleep.