Nina Spezzaferro

[ninaspezz@gmail.com](mailto:ninaspezz@gmail.com)

March 7, 2018

**Capstone Project: Data Story**

Introduction to the Problem

Are there relationships between my own activity, sleep patterns, and sleep quality?

Can I use my own activity and sleep data to improve my activity and sleep habits?

Data Set

**What important fields and information does the data set have?**

I would argue that all remaining fields and information in my cleaned data set are important. My data set includes 148 observations of 16 variables:

1. Date
2. Calories burned
3. Steps
4. Distance
5. Minutes sedentary
6. Minutes lightly active
7. Minutes fairly active
8. Minutes very active
9. Activity calories
10. Sleep start time
11. Sleep end time
12. Minutes asleep
13. Minutes awake
14. Number of awakenings
15. Time in bed
16. Day of week

**What are its limitations i.e. what are some questions that you cannot answer with this data set?**

This data set does not account for other factors that affect sleep such as travel, stress, and nutrition.

**What kind of cleaning and wrangling did you need to do?**

I needed to combine smaller batches of data into one cohesive set. In addition, I removed columns with blank values, filled in missing data for one night of unrecorded sleep, and removed smaller bouts of sleep considered to be naps.

**Any preliminary exploration you’ve performed and your initial findings.**

I began my exploration by calculating the mean minutes of sleep for each day of the week. I observed that Saturday had the highest mean minutes of sleep (517) and Tuesday had the lowest mean minutes of sleep (470).

With a difference of 47 minutes between the highest and lowest mean minutes of sleep by day of week, I may want to strive for more even minutes of sleep throughout the week. In addition, I may want to look at the variance of the minutes asleep for the time period studies. The minimum minutes of sleep observed in this time period was 294 minutes, while the maximum minutes of sleep observed was 656 minutes. When did the minimum and maximum events occur and what other events, outside of activity, may have caused fluctuations?

Next I calculated the mean steps for each day of the week. I observed that Sunday had the highest mean steps (16,608) and Monday had the lowest mean steps (9,160). This seemed odd to me as I typically complete long training runs on Saturdays. I then realized I have an outlier in the data set because I ran a marathon on a Sunday during the time period studied. This one day of 60,681 steps is enough to skew the rest of the data.

With that in mind, I then calculated the median steps for each day of the week. I observed that Wednesday had the highest median steps (14,978) and Monday had the lowest median steps (9,621). This seemed more in line with my step habits during the time period studied.

**Based on these findings, what approach are you going to take? How has your approach changed from what you initially proposed, if applicable?**

1. Explore activity data:

a. What is the mean number of steps per day during this time period?

b. What is the minimum and maximum steps per day during this time period?

c. Which day of the week has the highest mean steps?

d. Which day of the week has the highest mean active steps?

e. Plot mean steps by day of week

f. What does this mean?

g. What do these results make me want to do? Will behavioral changes should this inspire?

2. Explore sleep data:

a. What is the mean minutes asleep during this time period?

b. What is the mean minutes restless?

c. Histogram of sleep end time

d. Histogram of times awake

e. What does this mean?

f. What do these results make me want to do? Will behavioral changes should this inspire?

3. Explore activity data in relation to sleep data:

a. What relationships, if any, exist between minutes asleep and steps?

b. What relationships, if any, exist between minutes awake and steps?

c. …Minutes asleep and minutes active?

d. …Minutes asleep and minutes sedentary?