Data Visualization: Sleep Project

Jeffrey Li

November 25, 2015

Introduction

During the month of May in 2015, I kept a log of how many hours I slept and the activities I performed throughout each day. The goal was to determine whether or not certain activities during my day affected how many hours I required to sleep each night.

Methodology

The following methodology was used for this project:

- Keeping a log of my sleeping hours and activities in an Excel spreadsheet,
- Cleaning the dataset for missing values, and
- Creating a data visualization in R to identify any trends.

Tracking & Cleaning

Shown below is a sample of the log used in this project. Two key variables used are “Interrupted” and “Exercise”. The variable “Interrupted” is a Boolean value of whether or not I woke up in the middle of the night. The variable “Exercise” is also a Boolean value of whether or not I exercised that day. Feel Score, Mood Score, and Productivity score were not used in the project. Missing values from forgetting to log into the diary was also omitted. After compiling one month’s worth of data, the spreadsheet was converted into a csv file to be analyzed in R.

	A	B	C	D	E	F	G	H	I
1	Date	Sleep time	Wake up time	Sleep duration	Interrupted	Feel score	Mood score	Productivity score	Exercise
2	5/25/2015	0:30	11:00	10.5	TRUE	70%	75%	65%	FALSE
3	5/26/2015	23:00	5:30	6.5	FALSE	75%	75%	75%	FALSE
4	5/27/2015	23:00	10:00	11	TRUE	75%	75%	80%	FALSE
5	5/28/2015	23:00	10:00	11	TRUE	70%	80%	85%	TRUE

Figure 1. Excel spreadsheet of sleep log

Data Visualization Findings

At a glance, it seems that I slept the most on the days I exercised. This could be attributed to the extra rest needed for the body to recuperate. Secondly, it appears that I needed extra rest to compensate for the days that I was interrupted during the night.

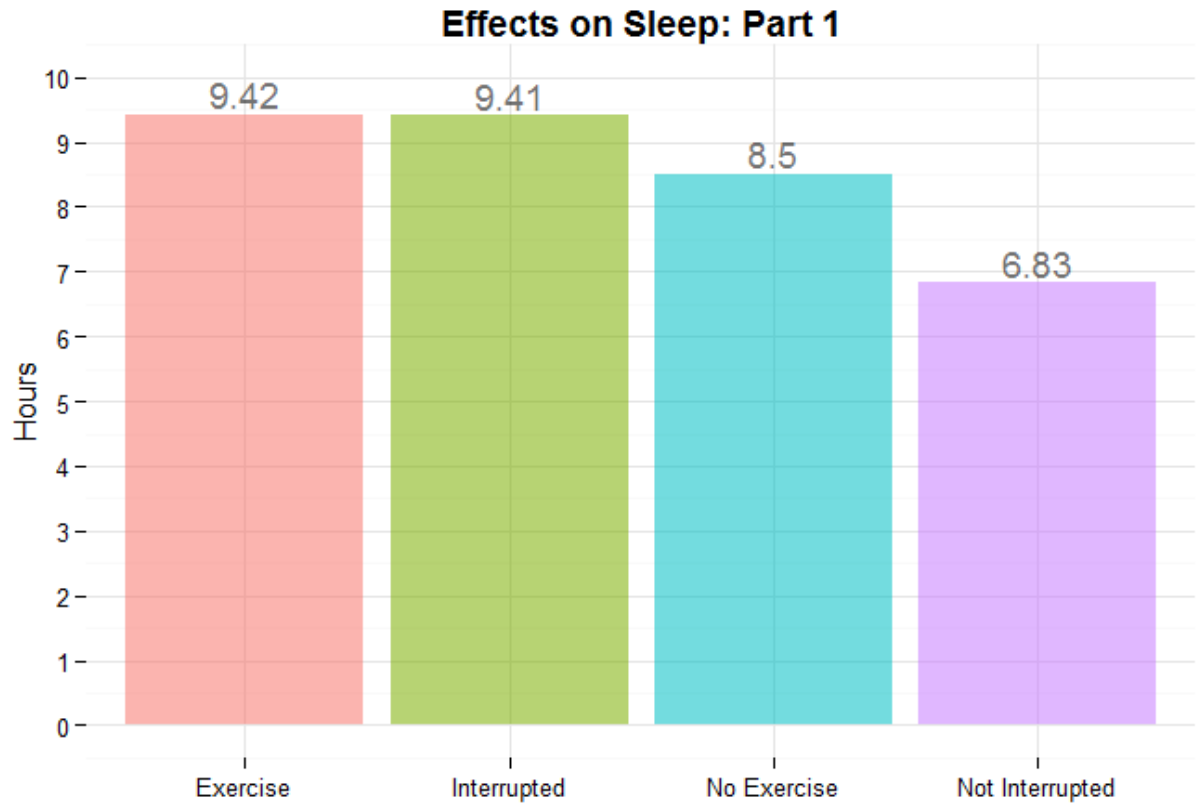

Figure 2. Segmented sleep log

Based on this visualization, I can infer that exercising and being interrupted may cause me to sleep more. As well, not expending energy to exercise and having uninterrupted sleep lowers my sleep needs.

After segmenting my sleep log into four categories, I began to wonder if these would be related at all. I hypothesized that I would need less sleep if I exercised and was not interrupted compared to when I exercised and was interrupted in the night.

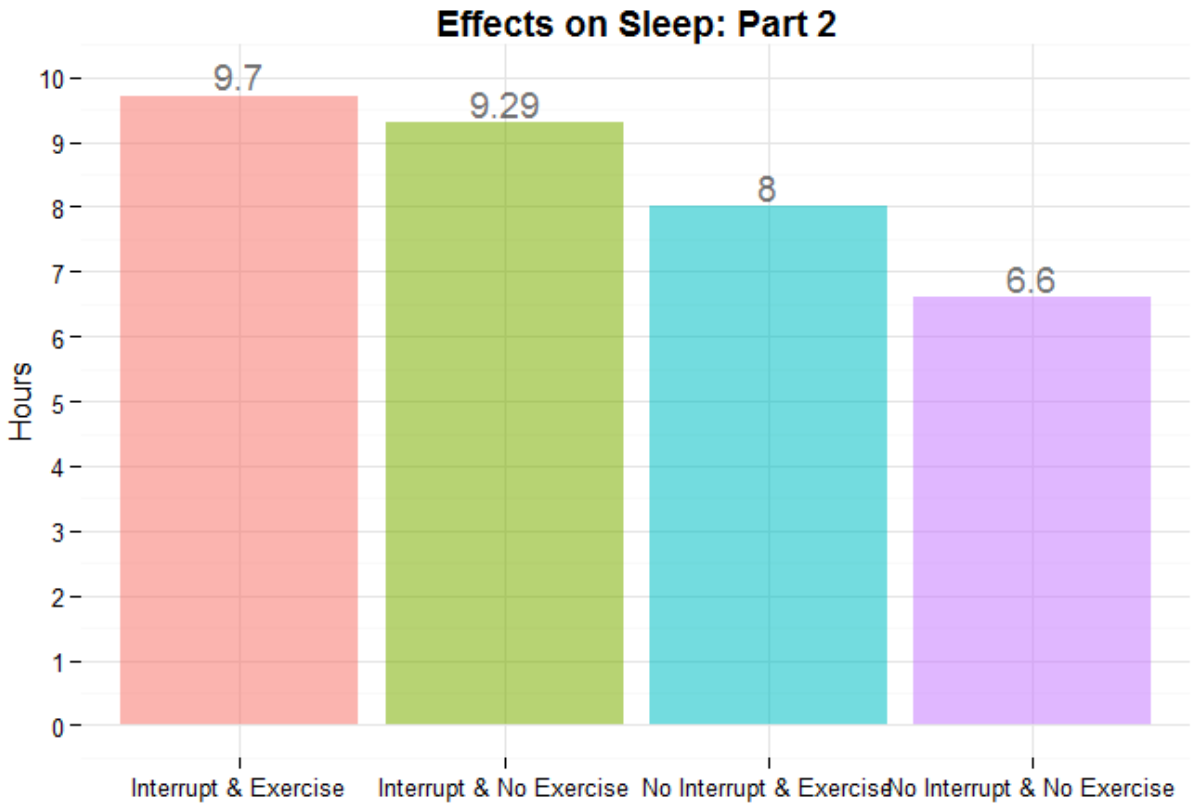

Figure 3. Segmented sleep log modified

By adjusting the dataset filters using the dplyr package, I was able to take an average of both "Interrupted" and "Exercise" on my sleep. Looking at the bar chart, my hypothesis was correct. Normally I need more sleep after exercising and being interrupted. If I were to exercise and be interrupted during the night, I would need more sleep than not exercising and being interrupted.

The most interesting find is not discovering that I needed more sleep after exercising and being interrupted in the same day, however, it was that I needed only 7 hours of sleep if I did not exercise and was not interrupted during the night.

Conclusion

Through doing this personal project, I was able to determine on average how many hours of sleep I required depending on whether I exercised and was interrupted. A downfall of this project is my dataset only contains 30 observations and even fewer accounting for missing observations. A future project could be to conduct the same project, but include more observations for a more conclusive finding.

R Source

#Dependencies

```{r, echo=FALSE, warning=FALSE, message=FALSE}
require(plyr)
require(dplyr)
require(ggplot2)
require(gridExtra)
require(lubridate)
require(GGally)
require(stringr)
require(scales)
```

#Load csv file

```{r, echo=FALSE}
sleep.data <-
read.csv("C:/Users/Jeffrey/Desktop/Jeffrey/R/Portfolio/Personal
Project.csv")
```

#Clean dataset of missing values

#Find mean of sleep duration for days of no exercise and exercise

```{r, echo=FALSE}
sleep.vs.exercise <- sleep.data %>%
  select(Sleep.duration, Exercise) %>%
  filter(Sleep.duration != "", Sleep.duration != "NA", Exercise !=
"") %>%
  group_by(Exercise) %>%
  mutate(Hours = round(mean(Sleep.duration), digits = 2)) %>%
  distinct(Exercise) %>%
  select(- Sleep.duration) %>%
  rename(Status = Exercise)
```

```
sleep.vs.exercise[1, 1] <- "No Exercise"
sleep.vs.exercise[2, 1] <- "Exercise"
```

#Clean dataset of missing values
#Find mean of sleep duration for days of no interruptions and interruptions

```{r, echo=FALSE}
sleep.vs.interruption <- sleep.data %>%
  select(Sleep.duration, Interrupted) %>%
  filter(Sleep.duration != "", Sleep.duration != "NA", Interrupted != "") %>%
  group_by(Interrupted) %>%
  mutate(Hours = round(mean(Sleep.duration), digits = 2)) %>%
  distinct(Interrupted) %>%
  select(-Sleep.duration) %>%
  rename(Status = Interrupted)

sleep.vs.interruption[1, 1] <- "Interrupted"
sleep.vs.interruption[2, 1] <- "Not Interrupted"

#Row bind both objects for a data frame
sleep.condition <- rbind(sleep.vs.exercise, sleep.vs.interruption)

#Plotting bar chart
sleep.condition.barplot <- ggplot(data = sleep.condition, aes(x = Status, y = Hours, fill = Status, alpha = 0.3)) +
    geom_bar(stat = "identity") +
    theme_minimal() +
    theme(plot.title = element_text(lineheight = 1, face = "bold"), legend.position = "none") +
    geom_text(aes(label = Hours), position = position_dodge(width = 0.9), vjust = -0.25) +
```

```
    ggtitle("Effects on Sleep: Part 1") +
    xlab("") +
    ylab("Hours") +
    scale_y_continuous(breaks = seq(0, 10, 1)) +
    expand_limits(y = 10)
```

```{r, echo=FALSE, warning=FALSE, message=FALSE, error=FALSE}
sleep.condition.barplot
```

#Filtering variables to see relations

```{r, echo=FALSE}
sleep.data.2 <- sleep.data %>%
  select(Sleep.duration, Interrupted, Exercise) %>%
  filter(Interrupted == "FALSE", Exercise == "FALSE",
Sleep.duration != "NA") %>%
  summarize(sleep.duration = mean(Sleep.duration))

sleep.data.3 <- sleep.data %>%
  select(Sleep.duration, Interrupted, Exercise) %>%
  filter(Interrupted == "TRUE", Exercise == "FALSE", Sleep.duration !=
"NA") %>%
  summarize(sleep.duration = mean(Sleep.duration))

sleep.data.4 <- sleep.data %>%
  select(Sleep.duration, Interrupted, Exercise) %>%
  filter(Interrupted == "TRUE", Exercise == "TRUE", Sleep.duration !=
"NA") %>%
  summarize(sleep.duration = mean(Sleep.duration))

sleep.data.5 <- sleep.data %>%
  select(Sleep.duration, Interrupted, Exercise) %>%
```

```
  filter(Interrupted == "FALSE", Exercise == "TRUE", Sleep.duration != "NA") %>%
  summarize(sleep.duration = mean(Sleep.duration))

#Creating new matrix
sleep.condition.2 <- matrix(nrow = 4, ncol = 2)
list.1 <- c("No Interrupt & No Exercise", "Interrupt & No Exercise", "Interrupt & Exercise", "No Interrupt & Exercise")

#Put average hours into list
data.list <- c(sleep.data.2, sleep.data.3, sleep.data.4, sleep.data.5)
```

```{r, echo=FALSE}
#Used loop to put list values into matrix
n <- c(0, 0, 0, 0)

for(i in 1:4){
  n[i] <- as.numeric(data.list[i])
  sleep.condition.2[i, 2] <- n[i]
}

for(n in 1:4){
  sleep.condition.2[n, 1] <- list.1[n]
}
```

```{r, echo=FALSE, warning=FALSE, message=FALSE, error=FALSE}
#Convert matrix into data frame for ggplot
sleep.condition.3 <- as.data.frame(sleep.condition.2, stringsAsFactors = FALSE) %>%
  rename(Status = V1, Hours = V2) %>%
  mutate(Hours = round(as.numeric(Hours), digits =2))
```

```
sleep.condition.3.barplot <- ggplot(data = sleep.condition.3, aes(x =
Status, y = Hours, fill = Status, alpha = 0.3)) +
    geom_bar(stat = "identity") +
    theme_minimal() +
    theme(plot.title = element_text(lineheight = 1, face = "bold"),
legend.position = "none") +
    geom_text(aes(label = Hours), position = position_dodge(width =
0.9), vjust = -0.25) +
    ggtitle("Effects on Sleep: Part 2") +
    xlab("") +
    ylab("Hours") +
    scale_y_continuous(breaks = seq(0, 10, 1)) +
    expand_limits(y = 10)

sleep.condition.3.barplot
```