

STAT 6021: Final Project EDA

Group 1

Libraries

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2     3.5.0      v tibble    3.2.1
## v lubridate  1.9.2      v tidyr     1.3.0
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggcorrplot)
```

Data

Looking at sleep efficiency, with hours of sleep as the response variable

```
sleep <- read.csv("Sleep_Efficiency.csv")
```

```
# Remove ID, bedtime, and wakeup time
sleep <- sleep[, -c(1, 4, 5)]
View(sleep)
```

```
colnames(sleep)
```

```
## [1] "Age" "Gender" "Sleep.duration"
## [4] "Sleep.efficiency" "REM.sleep.percentage" "Deep.sleep.percentage"
## [7] "Light.sleep.percentage" "Awakenings" "Caffeine.consumption"
## [10] "Alcohol.consumption" "Smoking.status" "Exercise.frequency"
```

Cleaning

```
# Turn percentages to proportions
sleep <- sleep %>%
  mutate(REM.sleep.percentage = REM.sleep.percentage / 100,
         Deep.sleep.percentage = Deep.sleep.percentage / 100,
         Light.sleep.percentage = Light.sleep.percentage / 100)
```

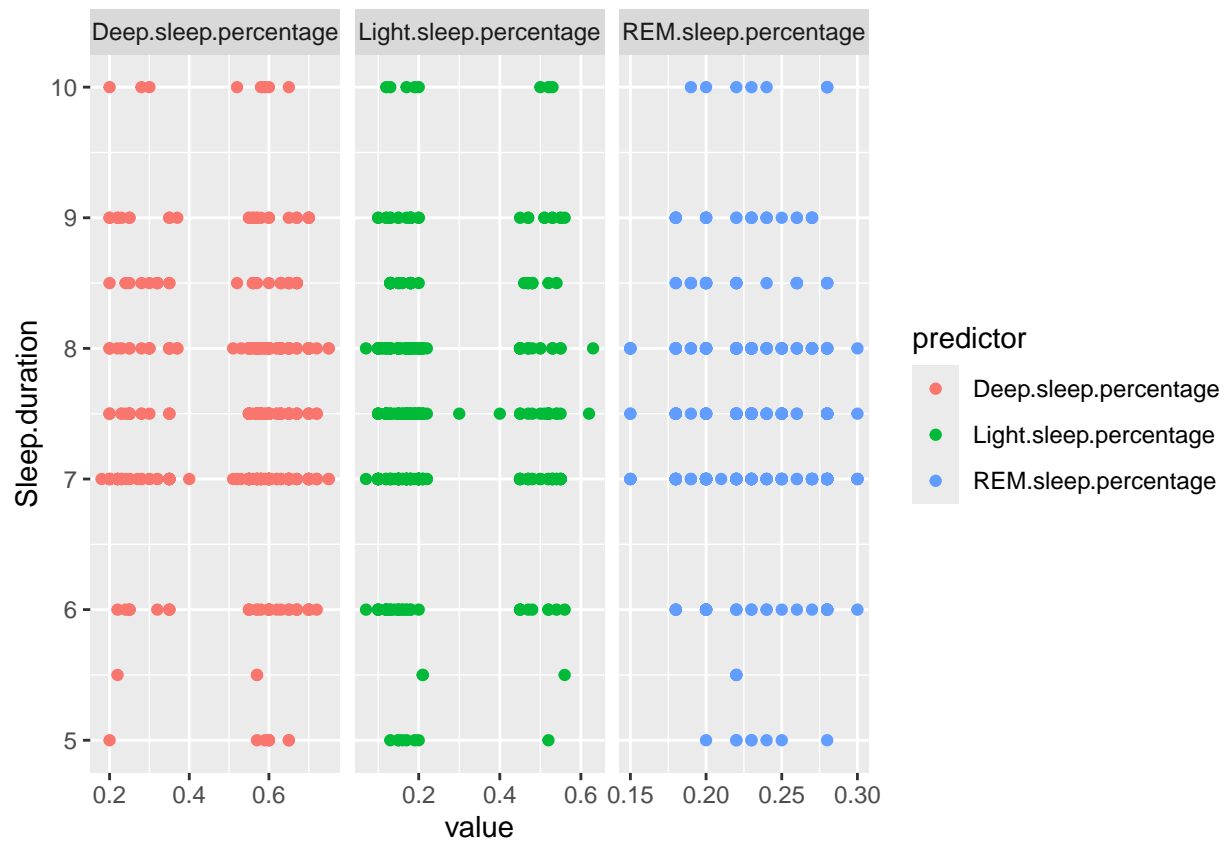
```
str(sleep)
```

```
## 'data.frame': 452 obs. of 12 variables:
## $ Age : int 65 69 40 40 57 36 27 53 41 11 ...
## $ Gender : chr "Female" "Male" "Female" "Female" ...
## $ Sleep.duration : num 6 7 8 6 8 7.5 6 10 6 9 ...
## $ Sleep.efficiency : num 0.88 0.66 0.89 0.51 0.76 0.9 0.54 0.9 0.79 0.55 ...
## $ REM.sleep.percentage : num 0.18 0.19 0.2 0.23 0.27 0.23 0.28 0.28 0.28 0.18 ...
## $ Deep.sleep.percentage : num 0.7 0.28 0.7 0.25 0.55 0.6 0.25 0.52 0.55 0.37 ...
## $ Light.sleep.percentage: num 0.12 0.53 0.1 0.52 0.18 0.17 0.47 0.2 0.17 0.45 ...
## $ Awakenings : num 0 3 1 3 3 0 2 0 3 4 ...
## $ Caffeine.consumption : num 0 0 0 50 0 NA 50 50 50 0 ...
## $ Alcohol.consumption : num 0 3 0 5 3 0 0 0 0 0 ...
## $ Smoking.status : chr "Yes" "Yes" "No" "Yes" ...
## $ Exercise.frequency : num 3 3 3 1 3 1 1 3 1 0 ...
```

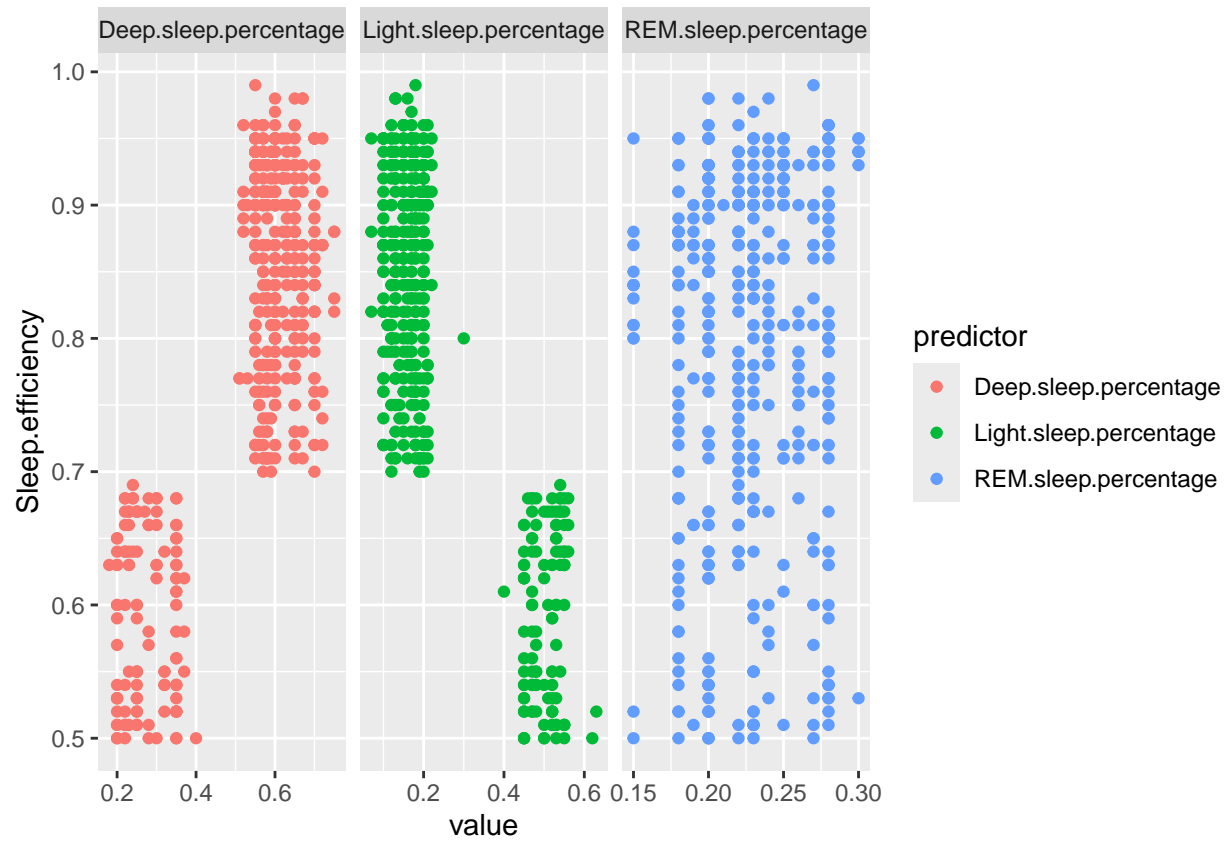
Scatter

```
long<-gather(sleep, key="predictor", value = "value",
             REM.sleep.percentage, Deep.sleep.percentage, Light.sleep.percentage)

ggplot(long, aes(x=value, y=Sleep.duration, color=predictor))+geom_point()+
  facet_wrap(~predictor,scales = "free_x")
```



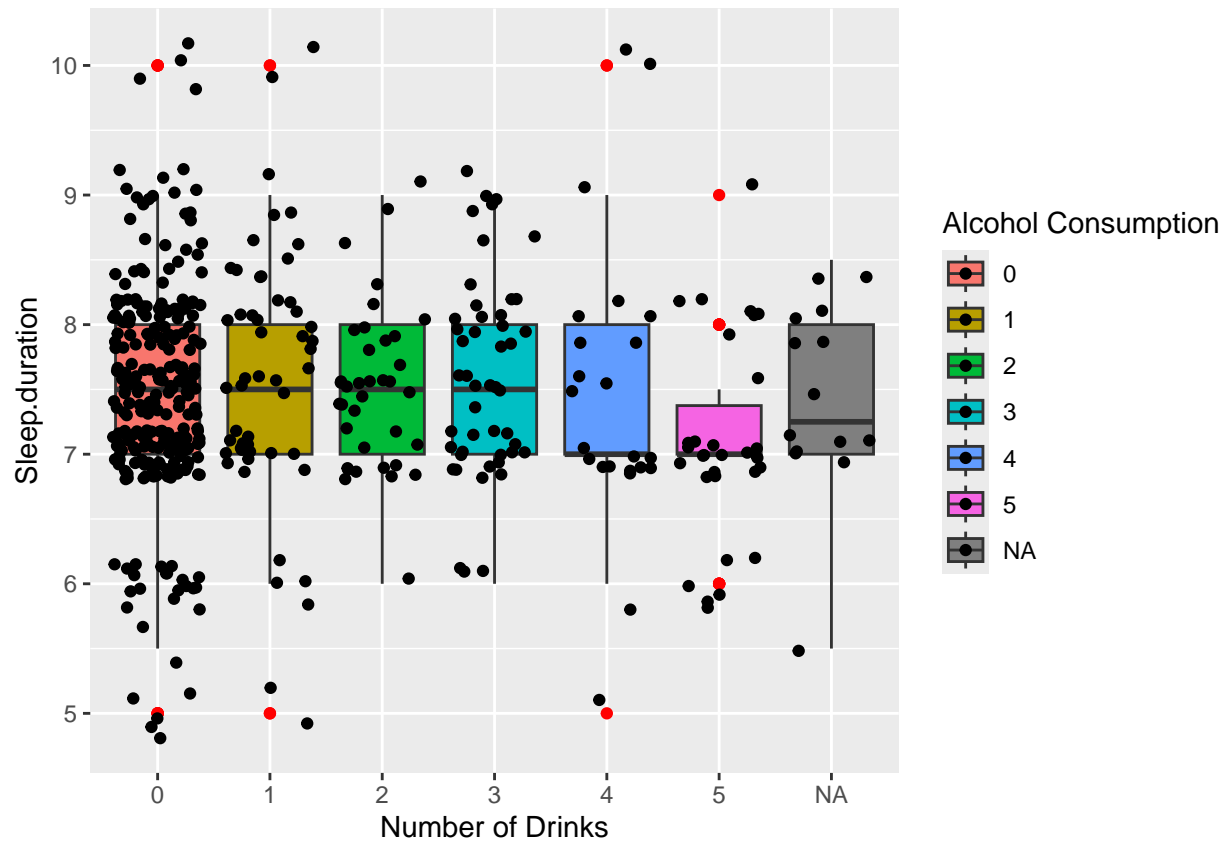
```
ggplot(long, aes(x=value, y=Sleep.duration, color=predictor))+geom_point()+
  facet_wrap(~predictor,scales = "free_x")
```



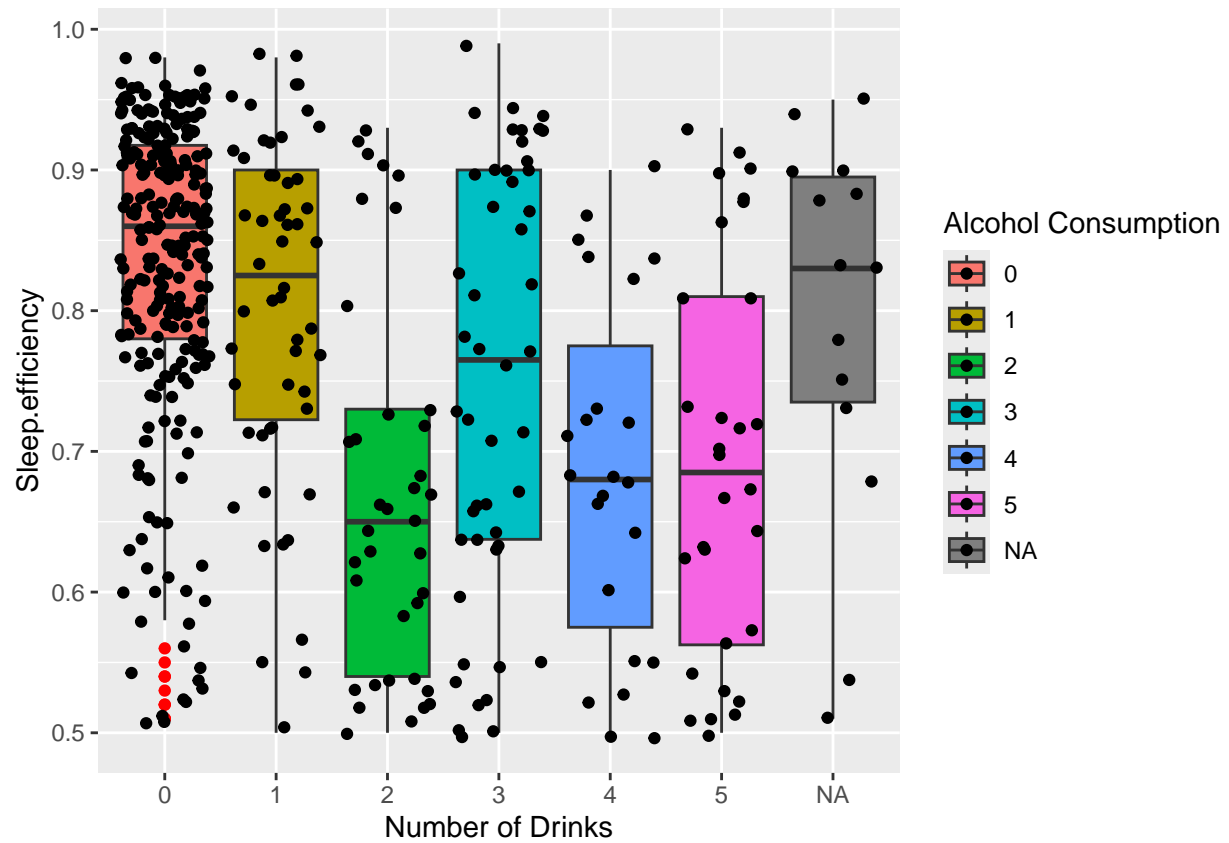
Boxplots

Alcohol Consumption

```
ggplot(sleep, aes(x=as.factor(Alcohol.consumption), y=Sleep.duration, fill=as.factor(Alcohol.consumption)))
  labs(x="Number of Drinks", fill="Alcohol Consumption")
```

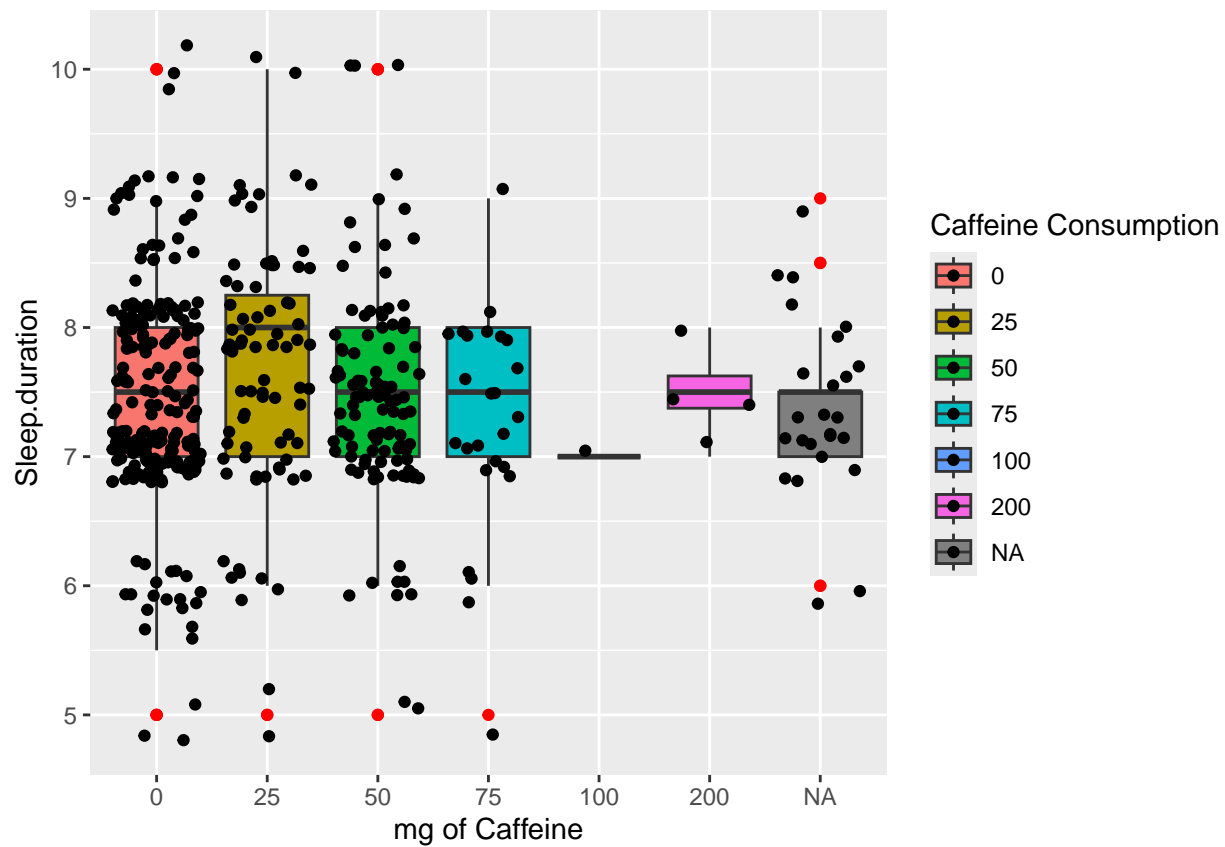


```
ggplot(sleep, aes(x=as.factor(Alcohol.consumption), y=Sleep.duration, fill=as.factor(Alcohol.consumption)))
labs(x="Number of Drinks", fill="Alcohol Consumption")
```

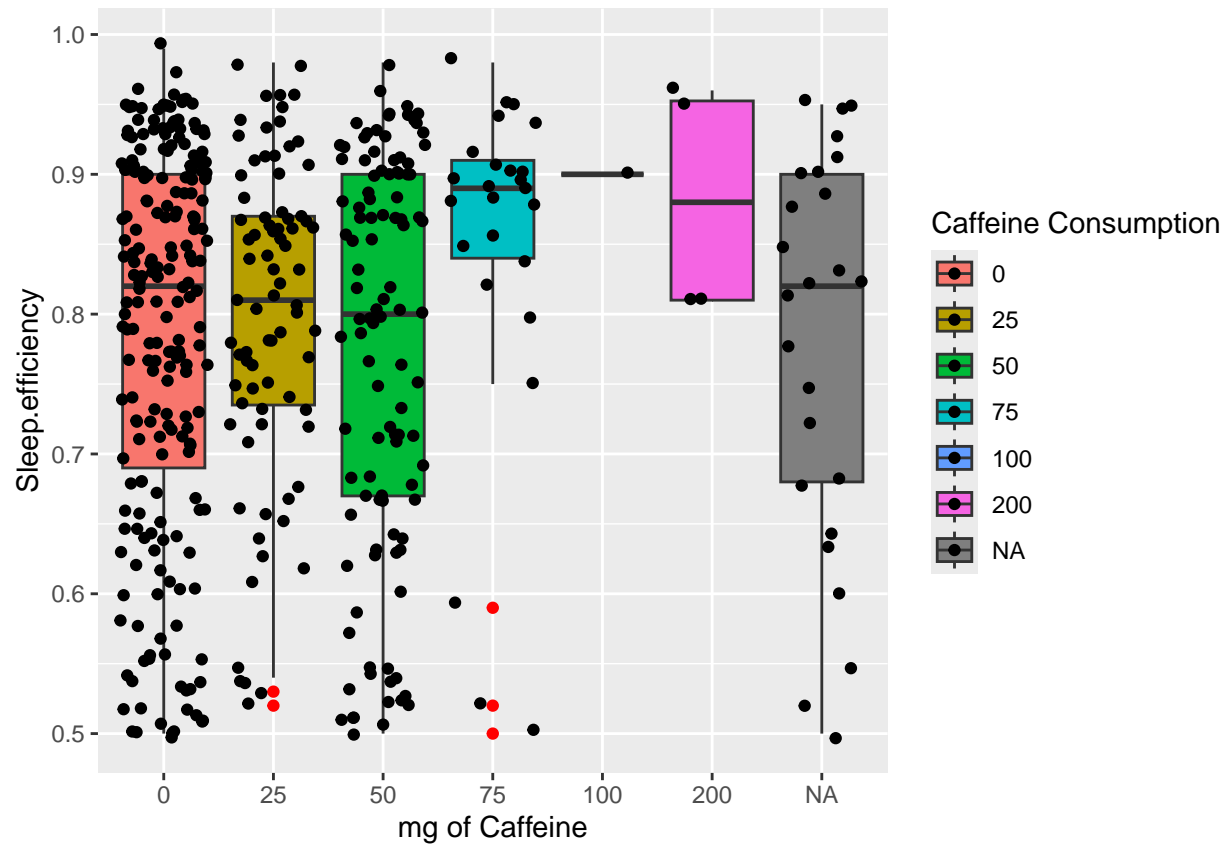


Caffeine Consumption

```
ggplot(sleep, aes(x=as.factor(Caffeine.consumption), y=Sleep.duration, fill=as.factor(Caffeine.consumption)))
labs(x="mg of Caffeine", fill="Caffeine Consumption")
```



```
ggplot(sleep, aes(x=as.factor(Caffeine.consumption), y=Sleep.duration, fill=as.factor(Caffeine.consumption)))
  labs(x="mg of Caffeine", fill="Caffeine Consumption")
```



Correlation Matrix

```
sleep.num <- sleep[-c(2, 11)]
cor_mat<-round(cor(sleep.num),2)
ggcorrplot(cor_mat,lab=TRUE, type="lower", method="circle")
```