HomeWork Assignment 1

AUTHOR
Rohan Chitale (23206793)

Reading the dataset

```
# Load necessary libraries
library(tidyverse)
```

```
Warning: package 'tidyverse' was built under R version 4.3.2

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
✓ dplyr     1.1.3     ✓ readr     2.1.4
✓ forcats   1.0.0     ✓ stringr   1.5.0
✓ ggplot2   3.4.4     ✓ tibble    3.2.1
✓ lubridate 1.9.3     ✓ tidyr     1.3.0
✓ purrr     1.0.2
── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
errors
```

```
library(skimr)
```

```
Warning: package 'skimr' was built under R version 4.3.2
```

```
library(ggplot2)
```

```
data=read_csv('Sleep_Efficiency.csv')
```

```
Rows: 452 Columns: 15
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr   (2): Gender, Smoking status
dbl  (11): ID, Age, Sleep duration, Sleep efficiency, REM sleep percentage, ...
dttm  (2): Bedtime, Wakeup time

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(data)
```

```
# A tibble: 6 × 15
     ID   Age Gender Bedtime             `Wakeup time`       `Sleep duration`
  <dbl> <dbl> <chr>  <dttm>              <dttm>                         <dbl>
1     1    65 Female 2021-03-06 01:00:00 2021-03-06 07:00:00             6
2     2    69 Male   2021-12-05 02:00:00 2021-12-05 09:00:00             7
```

```
3     3    40 Female 2021-05-25 21:30:00 2021-05-25 05:30:00             8
4     4    40 Female 2021-11-03 02:30:00 2021-11-03 08:30:00             6
5     5    57 Male   2021-03-13 01:00:00 2021-03-13 09:00:00             8
6     6    36 Female 2021-07-01 21:00:00 2021-07-01 04:30:00             7.5
# ℹ 9 more variables: `Sleep efficiency` <dbl>, `REM sleep percentage` <dbl>,
#   `Deep sleep percentage` <dbl>, `Light sleep percentage` <dbl>,
#   Awakenings <dbl>, `Caffeine consumption` <dbl>,
#   `Alcohol consumption` <dbl>, `Smoking status` <chr>,
#   `Exercise frequency` <dbl>
```

```
summary(data)
```

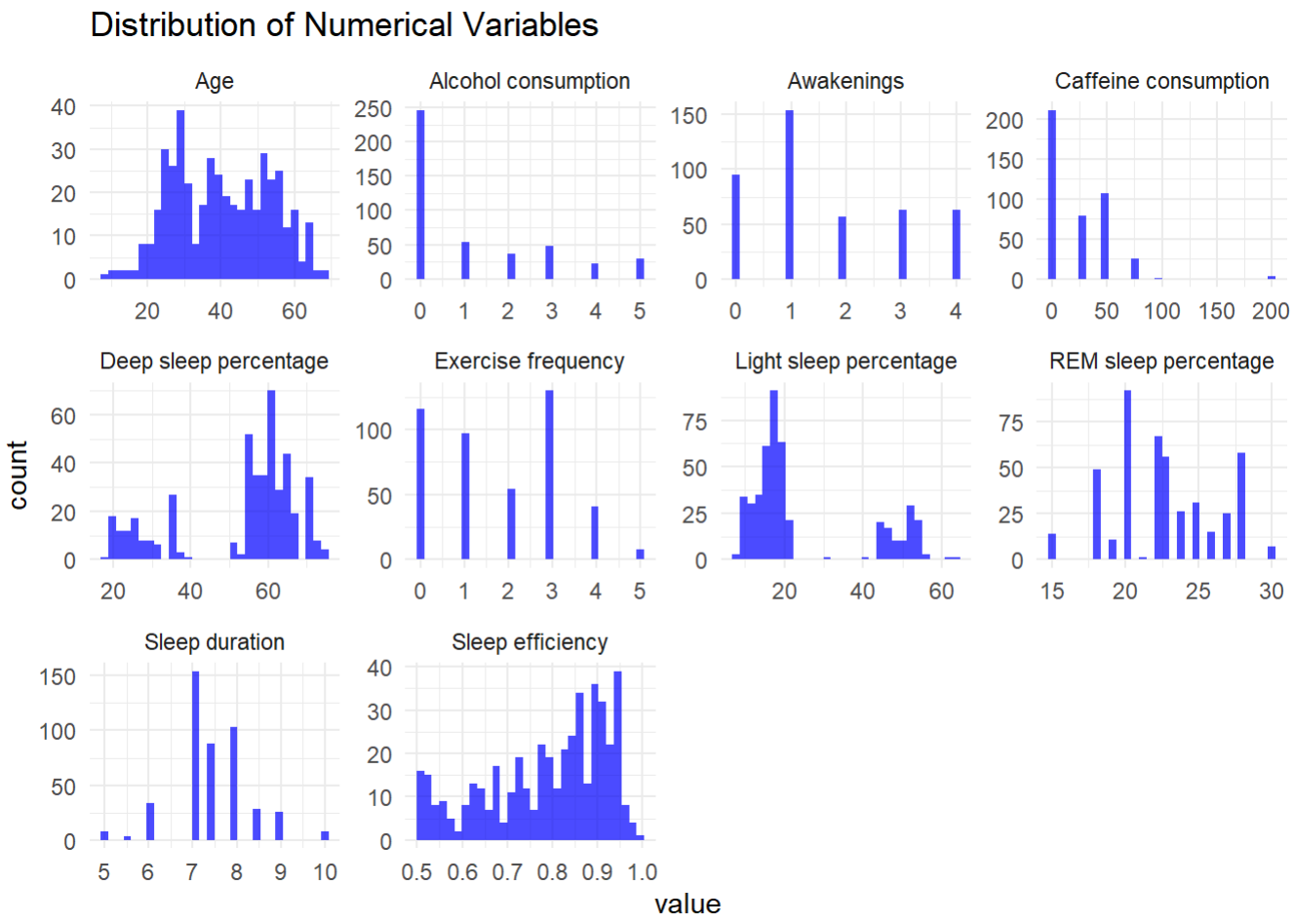
```
      ID              Age           Gender
 Min.   :  1.0   Min.   : 9.00   Length:452
 1st Qu.:113.8   1st Qu.:29.00   Class :character
 Median :226.5   Median :40.00   Mode  :character
 Mean   :226.5   Mean   :40.29
 3rd Qu.:339.2   3rd Qu.:52.00
 Max.   :452.0   Max.   :69.00

    Bedtime                         Wakeup time
 Min.   :2021-01-03 00:30:00.00   Min.   :2021-01-03 08:30:00.00
 1st Qu.:2021-04-14 01:07:30.00   1st Qu.:2021-04-14 07:52:30.00
 Median :2021-07-20 23:30:00.00   Median :2021-07-20 16:00:00.00
 Mean   :2021-07-13 00:03:39.02   Mean   :2021-07-12 20:19:22.82
 3rd Qu.:2021-10-11 05:22:30.00   3rd Qu.:2021-10-11 05:52:30.00
 Max.   :2021-12-31 21:00:00.00   Max.   :2021-12-31 06:30:00.00

 Sleep duration   Sleep efficiency REM sleep percentage Deep sleep percentage
 Min.   : 5.000   Min.   :0.5000   Min.   :15.00        Min.   :18.00
 1st Qu.: 7.000   1st Qu.:0.6975   1st Qu.:20.00        1st Qu.:48.25
 Median : 7.500   Median :0.8200   Median :22.00        Median :58.00
 Mean   : 7.466   Mean   :0.7889   Mean   :22.62        Mean   :52.82
 3rd Qu.: 8.000   3rd Qu.:0.9000   3rd Qu.:25.00        3rd Qu.:63.00
 Max.   :10.000   Max.   :0.9900   Max.   :30.00        Max.   :75.00

 Light sleep percentage   Awakenings    Caffeine consumption
 Min.   : 7.00          Min.   :0.000   Min.   :  0.00
 1st Qu.:15.00          1st Qu.:1.000   1st Qu.:  0.00
 Median :18.00          Median :1.000   Median : 25.00
 Mean   :24.56          Mean   :1.641   Mean   : 23.65
 3rd Qu.:32.50          3rd Qu.:3.000   3rd Qu.: 50.00
 Max.   :63.00          Max.   :4.000   Max.   :200.00
                        NA's   :20      NA's   :25
 Alcohol consumption Smoking status     Exercise frequency
 Min.   :0.000       Length:452         Min.   :0.000
 1st Qu.:0.000       Class :character   1st Qu.:0.000
 Median :0.000       Mode  :character   Median :2.000
 Mean   :1.174                          Mean   :1.791
 3rd Qu.:2.000                          3rd Qu.:3.000
 Max.   :5.000                          Max.   :5.000
 NA's   :14                             NA's   :6
```

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

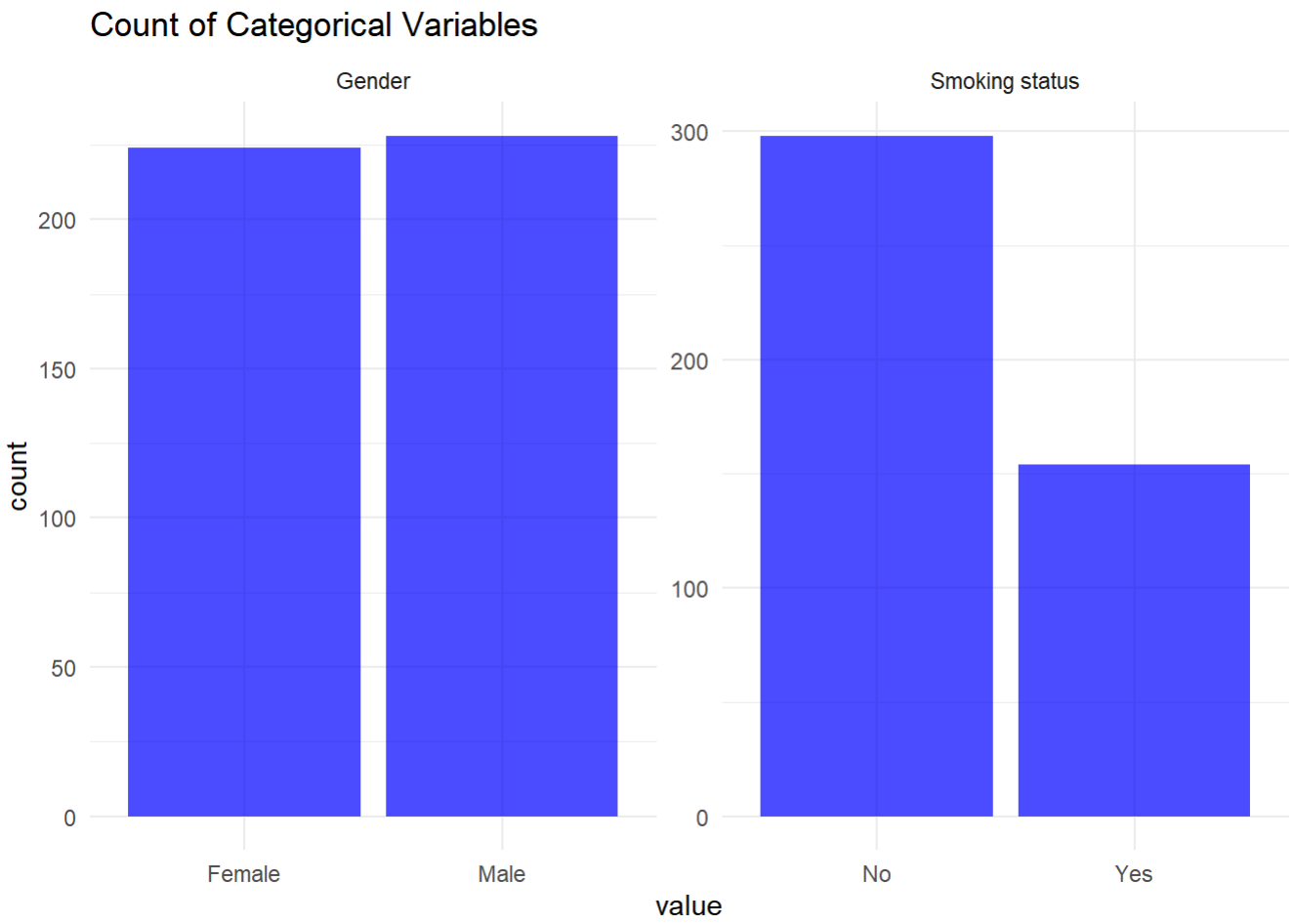
```
[1] 65
```

```
# Distributions of all variables
data %>% select(Age,`Sleep duration`,`Sleep efficiency`,`REM sleep percentage`, `Deep sleep pe
  labs(title='Distribution of Numerical Variables')
```

```
Warning: Removed 65 rows containing non-finite values (`stat_bin()`).
```




```
# Count of Categorical varaibles
data%>%select(Gender,`Smoking status`)%>% gather(key="variable",value="value")%>%
ggplot(aes(x=value))+geom_bar(fill='blue',alpha=0.7)+facet_wrap(~variable,scales ='free')+them
```


Count of Categorical Variables
Gender
Smoking status
count
0
50
100
150
200
0
100
200
300
Female
Male
No
Yes
value