

R -Project work

- Dataset: Sleep health and lifestyle-

Contents

- Introduction
- Coding with R
- Result

A large blue letter 'R' is centered within a circular frame. The frame has a thick orange border and a thin white inner border.

R

Introduction On Dataset : sleep health and lifestyle



- This project includes the analysis of sleep health and lifestyle dataset which is taken from Kaggle.
- predicting stress levels using r program.
- This dataset include comprises 400 rows and 13 columns.
- It includes details such as gender, age, occupation, sleep duration, quality of sleep, physical activity level, stress levels, BMI category, blood pressure, heart rate, daily steps, and sleep disorders.

OverView

- Sleep health analysis: Explore factors related to sleep duration
 - Lifestyle factors analysis: Investigate physical activity levels, stress levels, and BMI categories.
 - Cardiovascular health analysis: Examine blood pressure and resting heart rate measurements.
 - Sleep disorder analysis: Determine the presence of sleep disorders such as insomnia and sleep apnea.
- For adults minimum 8 hours of sleep is required.^T
- Stress Level (scale: 1-10): A subjective rating of the stress level experienced by the person, ranging from 1 to 10.
 - BMI Category: The BMI category of the person (e.g., Underweight, Normal, Overweight).
- A normal blood pressure reading for an adult is blood pressure that's below 120/80 mm Hg and above 90/60 mm Hg
- A None: No specific sleep disorder.
 - Insomnia: The individual experiences difficulty falling asleep or staying asleep
 - Sleep Apnea: The individual suffers from pauses in breathing during sleep, resulting in disrupted sleep patterns and potential health risks.

Coding

INSTALLING THE LIBRARY

```
install.packages("ggplot2")
```

```
install.packages("dplyr")
```

```
install.packages("ISLR")
```

```
install.packages("forcats")
```

```
install.packages("hrbrthemes")
```

IMPORTING THE LIBRARY

`library(ggplot2)`  used for Visualization

`library("dplyr")`  Grammer data manipulation

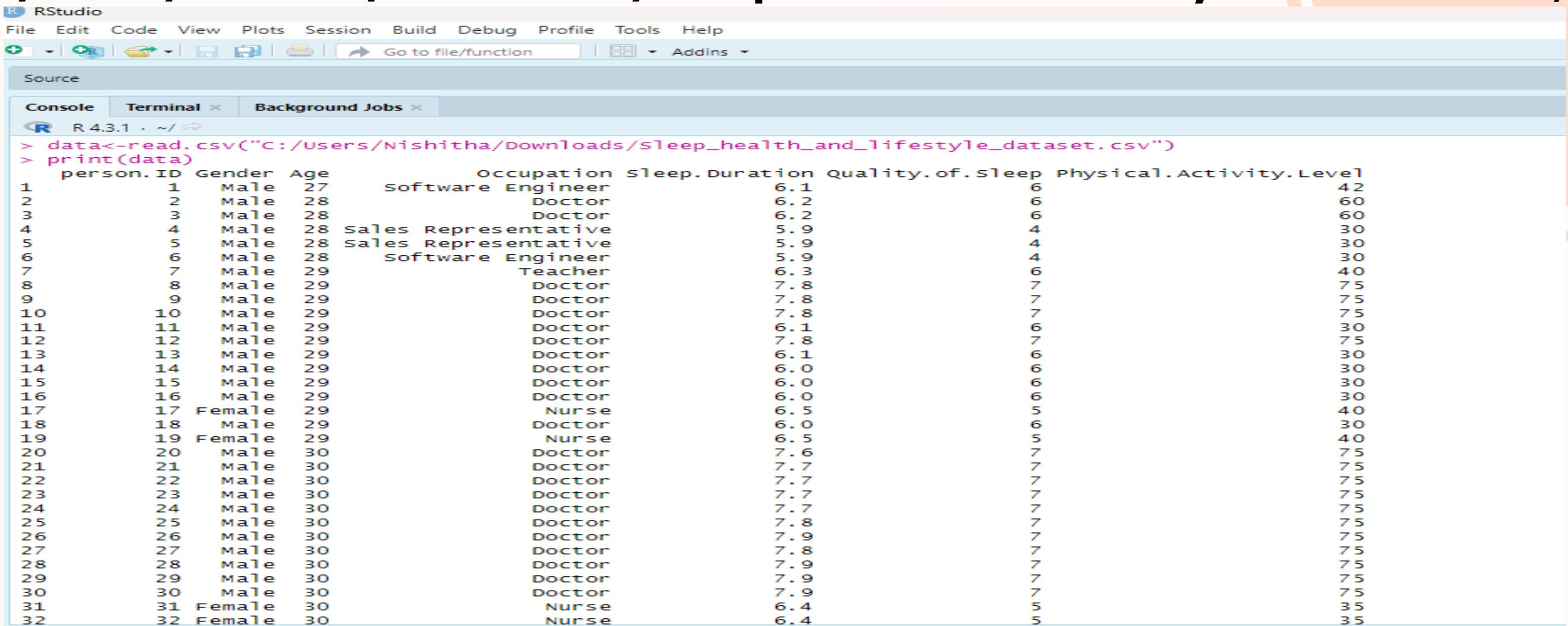
`library("ISLR")`  For dataset

`library(forcats)`  solve common problems with factors

`library(hrbrthemes)`  Additional Themes,Components for ggplot2

IMPORT THE DATASET

```
data<-  
read.csv("C:/Users/Nishitha/Downloads/Sleep_health_and_lifestyle_dataset.csv")  
print(data)
```



The screenshot shows the RStudio interface with the console tab selected. The console window displays the R code used to load the dataset and the resulting data frame. The data frame has 32 rows and 7 columns, with the first few rows showing gender, age, occupation, sleep duration, sleep quality, and physical activity level.

	person.ID	Gender	Age	Occupation	Sleep.Duration	Quality.of.Sleep	Physical.Activity.Level
1	1	Male	27	Software Engineer	6.1	6	42
2	2	Male	28	Doctor	6.2	6	60
3	3	Male	28	Doctor	6.2	6	60
4	4	Male	28	Sales Representative	5.9	4	30
5	5	Male	28	Sales Representative	5.9	4	30
6	6	Male	28	Software Engineer	5.9	4	30
7	7	Male	29	Teacher	6.3	6	40
8	8	Male	29	Doctor	7.8	7	75
9	9	Male	29	Doctor	7.8	7	75
10	10	Male	29	Doctor	7.8	7	75
11	11	Male	29	Doctor	6.1	6	30
12	12	Male	29	Doctor	7.8	7	75
13	13	Male	29	Doctor	6.1	6	30
14	14	Male	29	Doctor	6.0	6	30
15	15	Male	29	Doctor	6.0	6	30
16	16	Male	29	Doctor	6.0	6	30
17	17	Female	29	Nurse	6.5	5	40
18	18	Male	29	Doctor	6.0	6	30
19	19	Female	29	Nurse	6.5	5	40
20	20	Male	30	Doctor	7.6	7	75
21	21	Male	30	Doctor	7.7	7	75
22	22	Male	30	Doctor	7.7	7	75
23	23	Male	30	Doctor	7.7	7	75
24	24	Male	30	Doctor	7.7	7	75
25	25	Male	30	Doctor	7.8	7	75
26	26	Male	30	Doctor	7.9	7	75
27	27	Male	30	Doctor	7.8	7	75
28	28	Male	30	Doctor	7.9	7	75
29	29	Male	30	Doctor	7.9	7	75
30	30	Male	30	Doctor	7.9	7	75
31	31	Female	30	Nurse	6.4	5	35
32	32	Female	30	Nurse	6.4	5	35

HEAD

head(data) ←→ Here first six elements of data is displayed for the user

```
> head(data)
#> #> person.ID Gender Age Occupation sleep.Duration Quality.of.sleep Physical.Activity.Level
#> #> 1 1 Male 27 Software Engineer 6.1 6 42
#> #> 2 2 Male 28 Doctor 6.2 6 60
#> #> 3 3 Male 28 Doctor 6.2 6 60
#> #> 4 4 Male 28 Sales Representative 5.9 4 30
#> #> 5 5 Male 28 Sales Representative 5.9 4 30
#> #> 6 6 Male 28 Software Engineer 5.9 4 30
#> Stress.Level BMI.Category Blood.Pressure Heart.Rate Daily.Steps sleep.Disorder
#> #> 1 6 Overweight 126/83 77 4200 None
#> #> 2 8 Normal 125/80 75 10000 None
#> #> 3 8 Normal 125/80 75 10000 None
#> #> 4 8 Obese 140/90 85 3000 Sleep Apnea
#> #> 5 8 Obese 140/90 85 3000 Sleep Apnea
#> #> 6 8 Obese 140/90 85 3000 Insomnia
```

STR

Str(data) ←→ str is nothing but display the internal structure of data

In sleep dataset : 374 obs. of 13 variables

```
> str(data)
#> #> 'data.frame': 374 obs. of 13 variables:
#> #> $ person.ID : int 1 2 3 4 5 6 7 8 9 10 ...
#> #> $ Gender : chr "Male" "Male" "Male" "Male" ...
#> #> $ Age : int 27 28 28 28 28 28 29 29 29 29 ...
#> #> $ occupation : chr "Software Engineer" "Doctor" "Doctor" "Sales Representative" ...
#> #> $ sleep.Duration : num 6.1 6.2 6.2 5.9 5.9 5.9 6.3 7.8 7.8 7.8 ...
#> #> $ Quality.of.sleep : int 6 6 6 4 4 4 6 7 7 7 ...
#> #> $ Physical.Activity.Level: int 42 60 60 30 30 30 40 75 75 75 ...
#> #> $ Stress.Level : int 6 8 8 8 8 7 6 6 6 ...
#> #> $ BMI.Category : chr "Overweight" "Normal" "Normal" "Obese" ...
#> #> $ Blood.Pressure : chr "126/83" "125/80" "125/80" "140/90" ...
#> #> $ Heart.Rate : int 77 75 75 85 85 82 70 70 70 ...
#> #> $ Daily.Steps : int 4200 10000 10000 3000 3000 3000 3500 8000 8000 8000 ...
#> #> $ sleep.Disorder : chr "None" "None" "None" "Sleep Apnea" ...
```

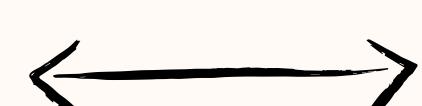
DISPLAYING THE ROW AND COLUMNS

```
dim(data)
```

```
[1] 374 13
```

There are 374 rows and 13 column

FINDING NULL VALUES



No null or missing values are found in this dataset

```
sapply(data,function(x) sum(is.na(x)))
```

```
> sapply(data,function(x) sum(is.na(x)))
```

person.ID	Gender	Age	Occupation	sleep.Duration
0	0	0	0	0
quality.of.sleep	Physical.Activity.Level	Stress.Level	BMI.Category	Blood.Pressure
0	0	0	0	0
Heart.Rate	Daily.steps	sleep.Disorder		
0	0	0		

SUMMARY OF DATASET

summary(data) ←

→ Summary of sleep dataset gives us a mean ,median ,mode
1 and 3 quartile

```
> summary(data)
```

person.ID	Gender	Age	Occupation	sleep.Duration	Quality.of.sleep
Min. : 1.00	Length:374	Min. :27.00	Length:374	Min. :5.800	Min. :4.000
1st Qu.: 94.25	class :character	1st Qu.:35.25	class :character	1st Qu.:6.400	1st Qu.:6.000
Median :187.50	Mode :character	Median :43.00	Mode :character	Median :7.200	Median :7.000
Mean :187.50		Mean :42.18		Mean :7.132	Mean :7.313
3rd Qu.:280.75		3rd Qu.:50.00		3rd Qu.:7.800	3rd Qu.:8.000
Max. :374.00		Max. :59.00		Max. :8.500	Max. :9.000
Physical.Activity.Level	Stress.Level	BMI.Category	Blood.Pressure	Heart.Rate	Daily.steps
Min. :30.00	Min. :3.000	Length:374	Length:374	Min. :65.00	Min. : 3000
1st Qu.:45.00	1st Qu.:4.000	class :character	class :character	1st Qu.:68.00	1st Qu.: 5600
Median :60.00	Median :5.000	Mode :character	Mode :character	Median :70.00	Median : 7000
Mean :59.17	Mean :5.385			Mean :70.17	Mean : 6817
3rd Qu.:75.00	3rd Qu.:7.000			3rd Qu.:72.00	3rd Qu.: 8000
Max. :90.00	Max. :8.000			Max. :86.00	Max. :10000
Sleep.Disorder					
Length:374					
Class :character					
Mode :character					

Visualize of a dataset

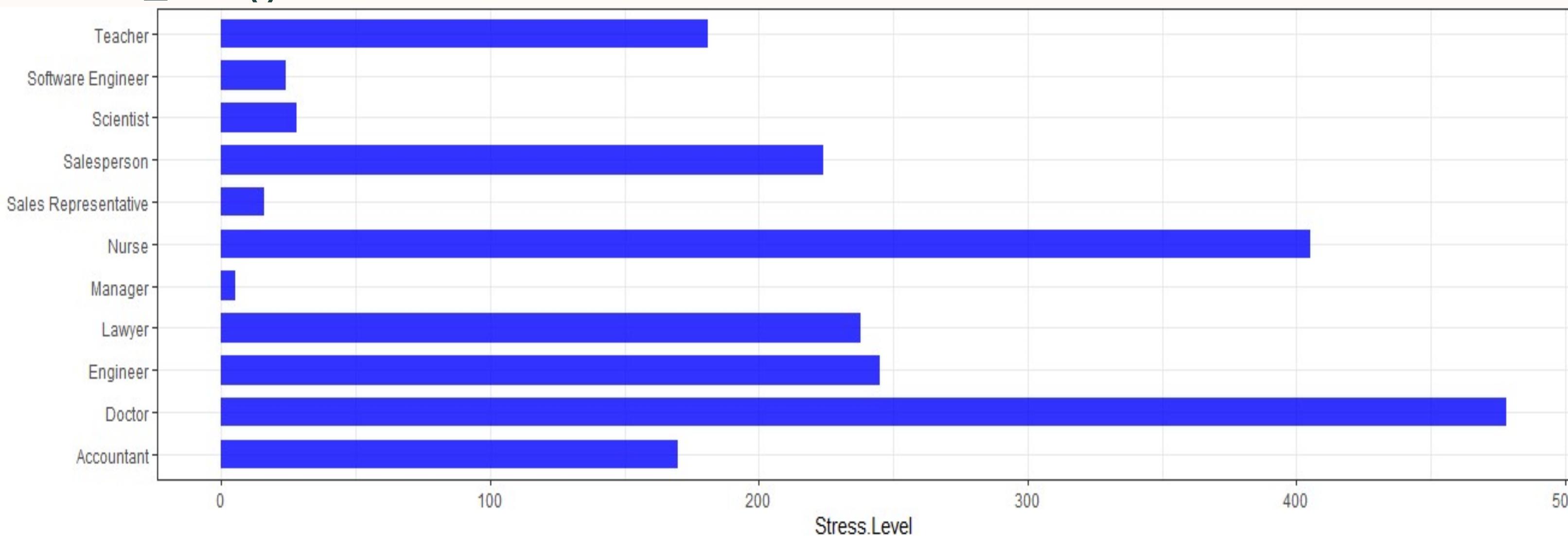
```
ggplot(data, aes(x=Stress.Level)) + ← here x is taken as stress level  
  geom_histogram(fill="pink",  
  color="blue",bins=6) +  
  ggtitle("Histogram of Stress.Level")
```



BAR GRAPH

```
data %>%
```

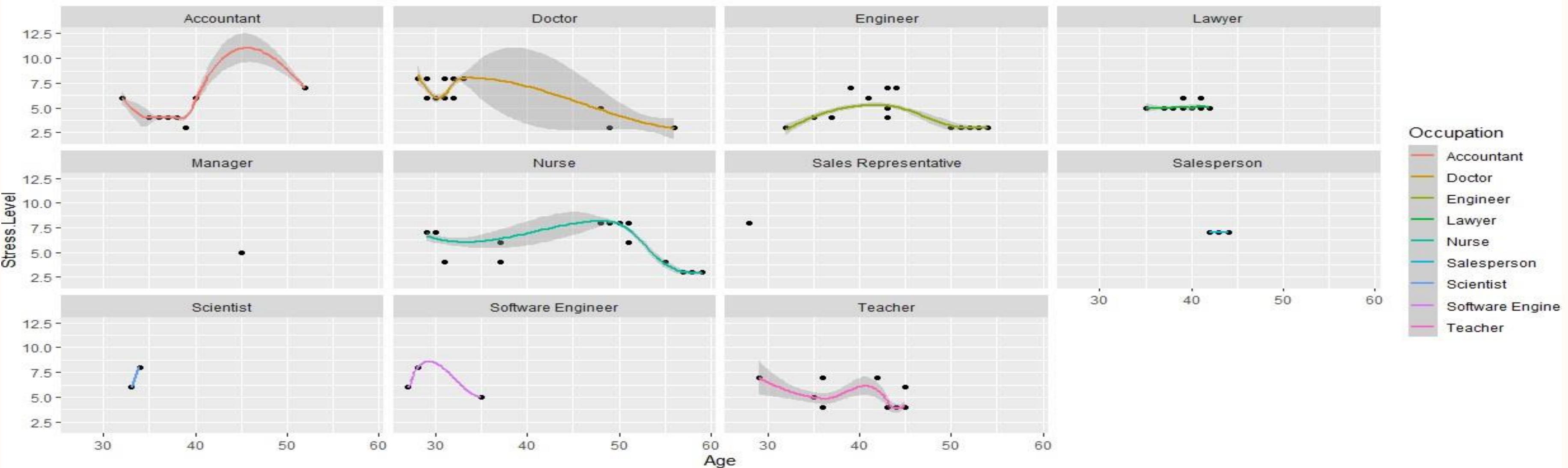
```
ggplot( aes(x=Occupation, y=Stress.Level)) +  
  geom_bar(stat="identity", fill="blue", alpha=.8, width=.7) +  
  coord_flip() +  
  xlab("") +  
  theme_bw()
```



Here in this graph we can find that stress level is more in doctor and Nurse and least in manager

MAPING

```
ggplot(data = data, mapping = aes(x = Age, y = Stress.Level)) +  
  geom_point() +  
  geom_smooth(aes(color = Occupation)) +  
  facet_wrap(~ Occupation)
```



Doctors and nurses -from early 20s to late 50s there are stress level are high

Teacher-late 20s to mid 40s there are stress high

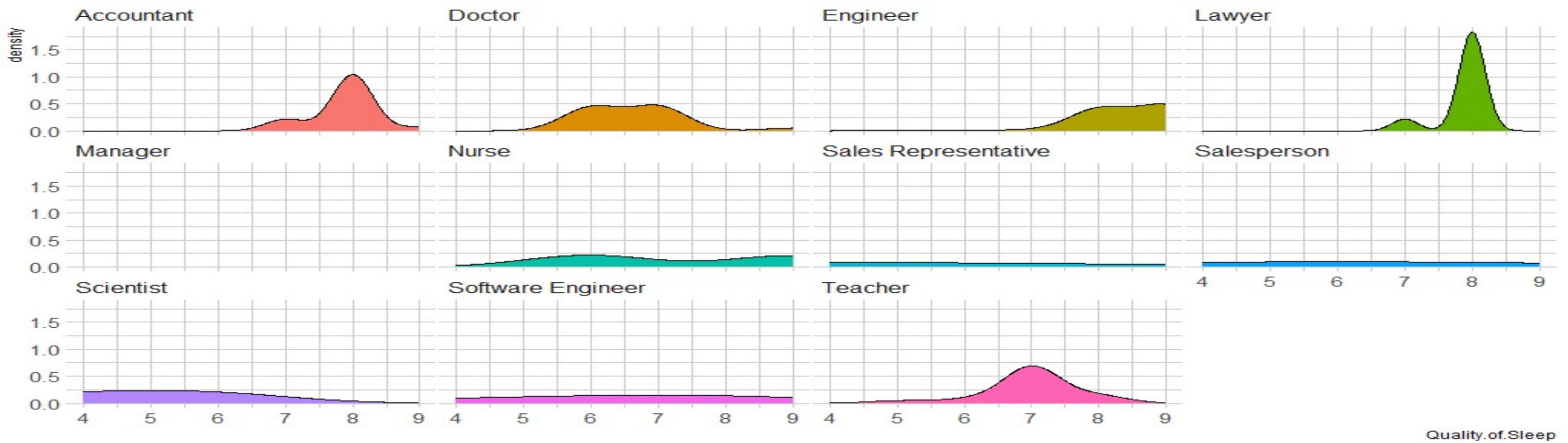
lawyer-have a stagnant stress level

software engineer and manager- have less stress level

```

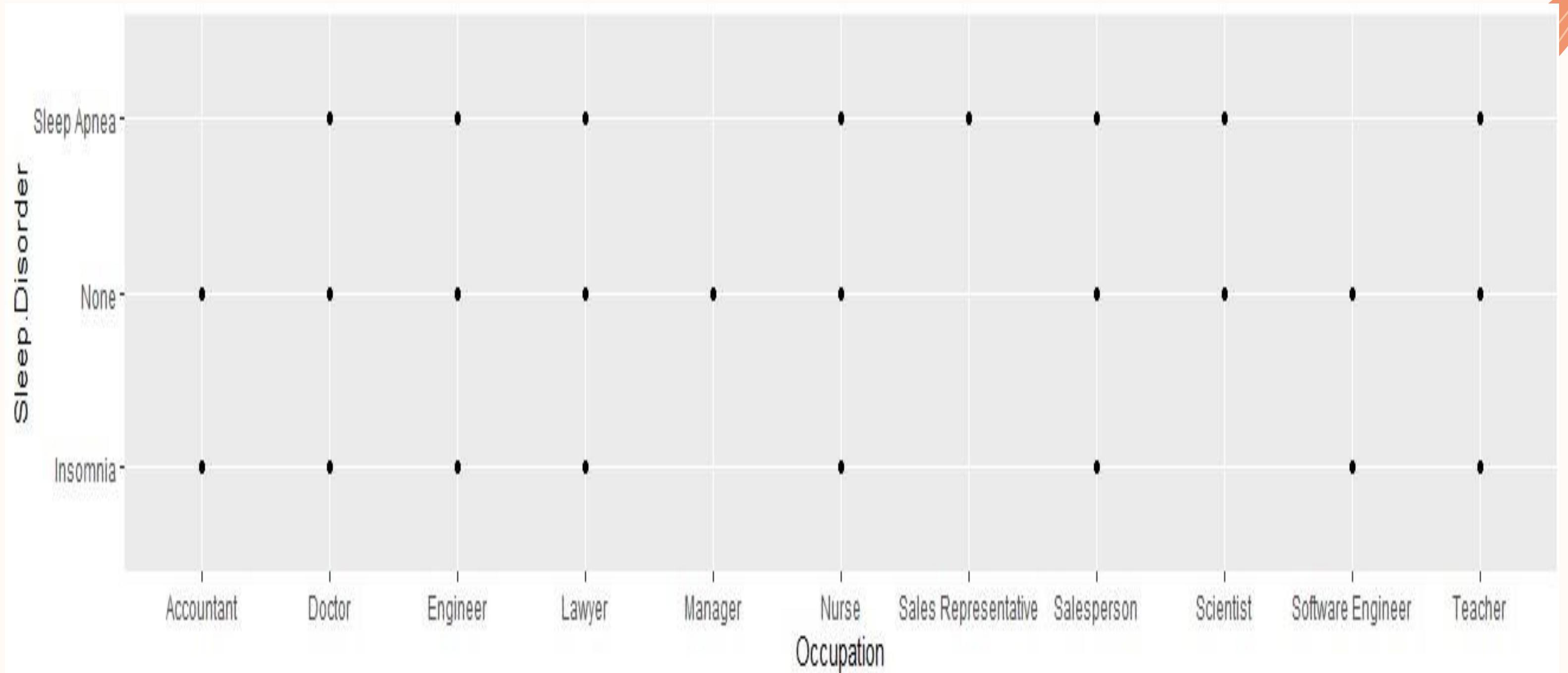
plot(data, aes(x=Sleep.Duration, group=Occupation,, fill=Occupation)) +
  geom_density(adjust=1.5) +
  theme_ipsum() +
  facet_wrap(~Occupation,) +
  theme(
    legend.position="none",
    panel.spacing = unit(0.1, "lines"),
  )

```



Here doctors and nurses have least sleep duration sales and accountant have good amount sleep duration

```
ggplot(data, aes(x=Occupation, y=Sleep.Disorder)) +  
  geom_point()
```





Thank
you

