ANA 515 Practicum

Sakshi Kubba

2023-07-03

##Description of the data

```
The data is related to mental health data
```

##Install the following packages install.packages("dplyr") install.packages("tidyverse") install.packages("knitr") install.packages("bslib") install.packages("utils")

##Reading the data into R.

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.2.3
```

```
## get the dataset from excel
mentalspreadsheet1 <- read_excel("C:/users/hkubb/Downloads/survey.xlsx", sheet = "Sheet1",
                                 col_types = c("date", "numeric", "text",
                                               "text", "text", "text", "text", "text",
                                               "text", "text", "text", "text", "text",
                                               "text", "text", "text", "text", "text",
                                               "text", "text", "text", "text", "text",
                                               "text", "text", "text", "text"))

mentalspreadsheet2 <- read_excel("C:/users/hkubb/Downloads/survey.xlsx", sheet = "Sheet 2",
                                 col_types = c("date", "numeric", "text",
                                               "text", "text", "text", "text", "text",
                                               "text", "text", "text", "text", "text",
                                               "text", "text", "text", "text", "text",
                                               "text", "text", "text", "text", "text",
                                               "text", "text", "text", "text"))
```

```
## Warning: Coercing numeric to date in A5 / R5C1
```

```
#Merging the datasheets
mentalhealth <- merge(mentalspreadsheet1, mentalspreadsheet2, all.x= TRUE, all.y = TRUE)
```

##Clean the data

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 4.2.3
```

```
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 4.2.3
```

```
library(bslib)
```

```
## Warning: package 'bslib' was built under R version 4.2.3
```

```
##
## Attaching package: 'bslib'
```

```
## The following object is masked from 'package:utils':
##
##     page
```

```
library(utils)
mentalhealthclean<-filter(mentalhealth, Country =="United States", Age>=1|Age<=100)
mentalhealthclean <- rename (mentalhealthclean, c(Sex = Gender))
mentalhealthcleanfinal <- mentalhealthclean %>%
  select(c(Age, Sex, state, self_employed, family_history, treatment, no_employees, remote_work,
tech_company, benefits, care_options, wellness_program, seek_help, mental_health_consequence, ph
ys_health_consequence, coworkers, supervisor, mental_health_interview, phys_health_interview, me
ntal_vs_physical, obs_consequence))
```

##Summary Staistics

```
        kable(str(mentalhealthcleanfinal))
```

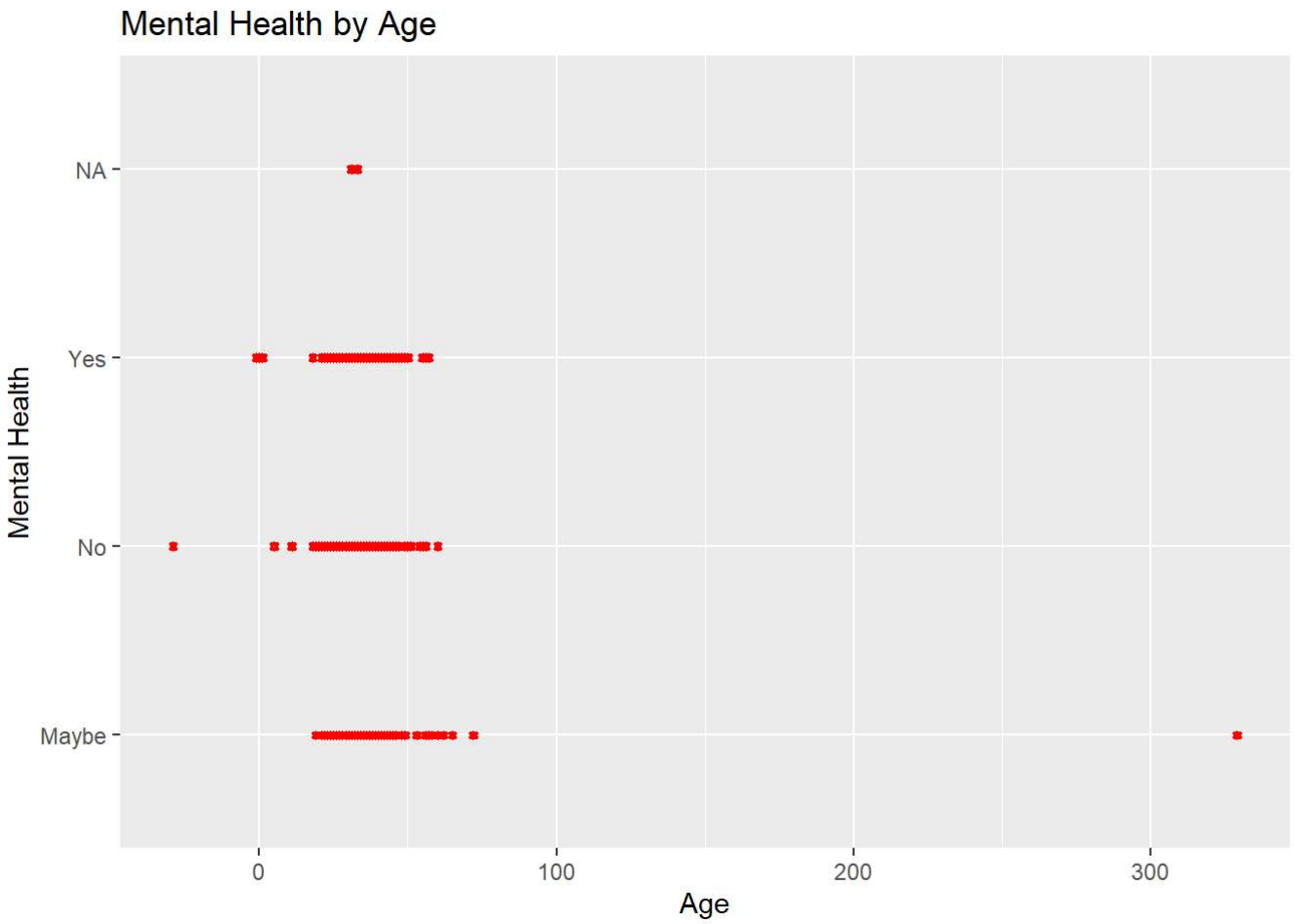
```
## 'data.frame':    745 obs. of  21 variables:
##  $ Age                      : num  37 44 37 44 31 33 35 31 42 36 ...
##  $ Sex                      : chr  "male" "Female" "Female" "M" ...
##  $ state                    : chr  "MA" "MA" "IL" "IN" ...
##  $ self_employed            : chr  "No" "No" "NA" "NA" ...
##  $ family_history           : chr  "No" "No" "No" "No" ...
##  $ treatment                : chr  "No" "No" "Yes" "No" ...
##  $ no_employees             : chr  "500-1000" "100-500" "44372" "More than 1000" ...
##  $ remote_work              : chr  "Yes" "No" "No" "No" ...
##  $ tech_company             : chr  "Yes" "Yes" "Yes" "No" ...
##  $ benefits                 : chr  "Yes" "Yes" "Yes" "Don't know" ...
##  $ care_options             : chr  "No" "Not sure" "Not sure" "No" ...
##  $ wellness_program         : chr  "No" "Don't know" "No" "Don't know" ...
##  $ seek_help                : chr  "No" "Don't know" "Yes" "Don't know" ...
##  $ mental_health_consequence: chr  "Yes" "Maybe" "No" "Maybe" ...
##  $ phys_health_consequence  : chr  "No" "Maybe" "No" "No" ...
##  $ coworkers                : chr  "Some of them" "No" "Some of them" "No" ...
##  $ supervisor               : chr  "No" "Some of them" "Yes" "No" ...
##  $ mental_health_interview  : chr  "Maybe" "No" "No" "No" ...
##  $ phys_health_interview    : chr  "Yes" "No" "Maybe" "No" ...
##  $ mental_vs_physical       : chr  "No" "Don't know" "Yes" "Don't know" ...
##  $ obs_consequence          : chr  "No" "No" "No" "No" ...
```

Print Scatter Plot

```
## Print Scatterplot
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

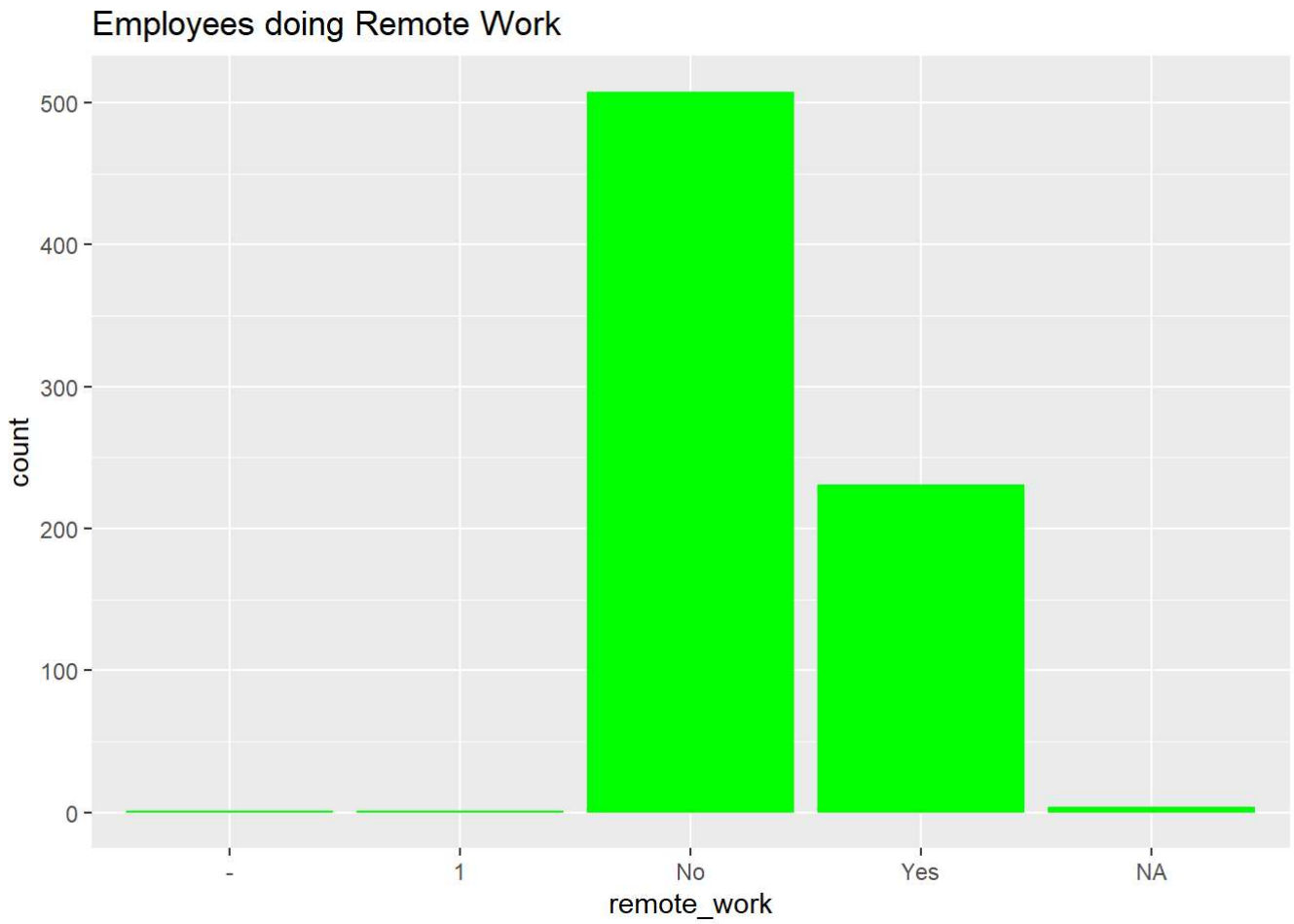
```
ggplot(data = mentalhealthcleanfinal, aes(x = Age, y = mental_health_consequence))+ geom_point(
    color="red",shape=11,size=0.2,stroke = 1)+
  labs(title = "Mental Health by Age", x = "Age", y="Mental Health")
```



Print Chart

```
mentalhealthcleanfinal %>% select(no_employees, remote_work)%>%group_by(remote_work, no_employee
s) %>% summarise(c=n()) %>% ggplot(aes(x=remote_work,y=c,fill=no_employees))+geom_bar(fill='gree
n',stat='identity')+labs(title="Employees doing Remote Work",y="count")
```

```
## `summarise()` has grouped output by 'remote_work'. You can override using the
## `.groups` argument.
```



saving clean data

```
write.csv(mentalhealthcleanfinal,"mentalhalthcleanfinal.csv")
```