R_Plots

Abu Bakar Siddiqur Rahman

10/6/2021

DataExploration

Scatter Plot with 2 Variables

this plot will help to understand by getting the average age for each program. The data visualization provide a quick view that which age group prefer which program.

```
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
dat <- read.csv("HFS.csv")
library(dplyr)
```

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

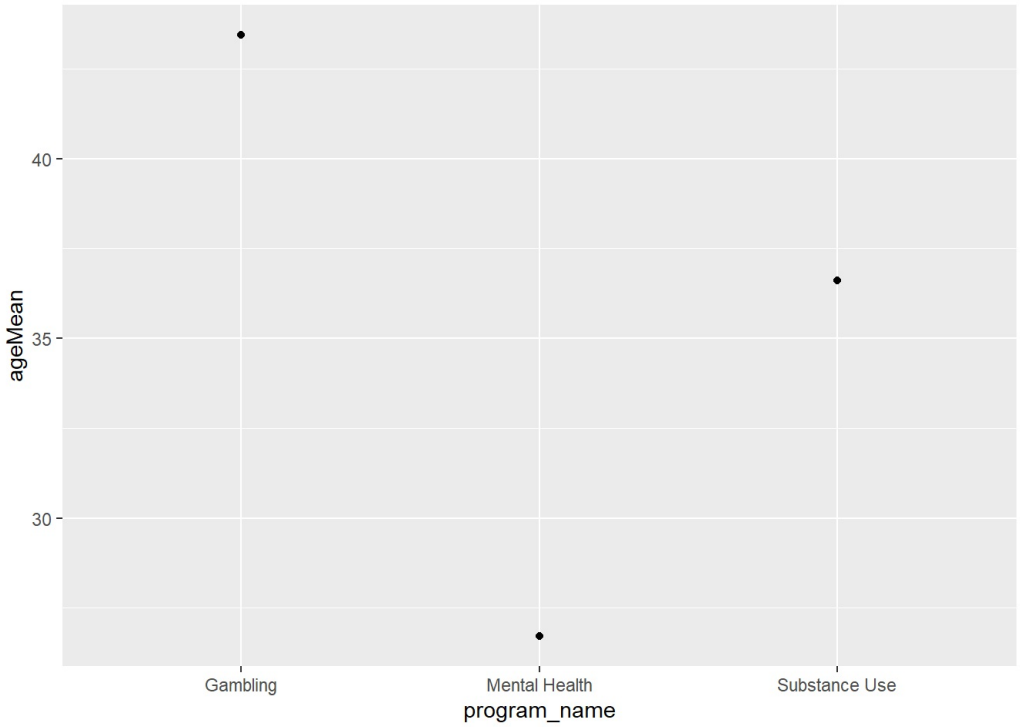
```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

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

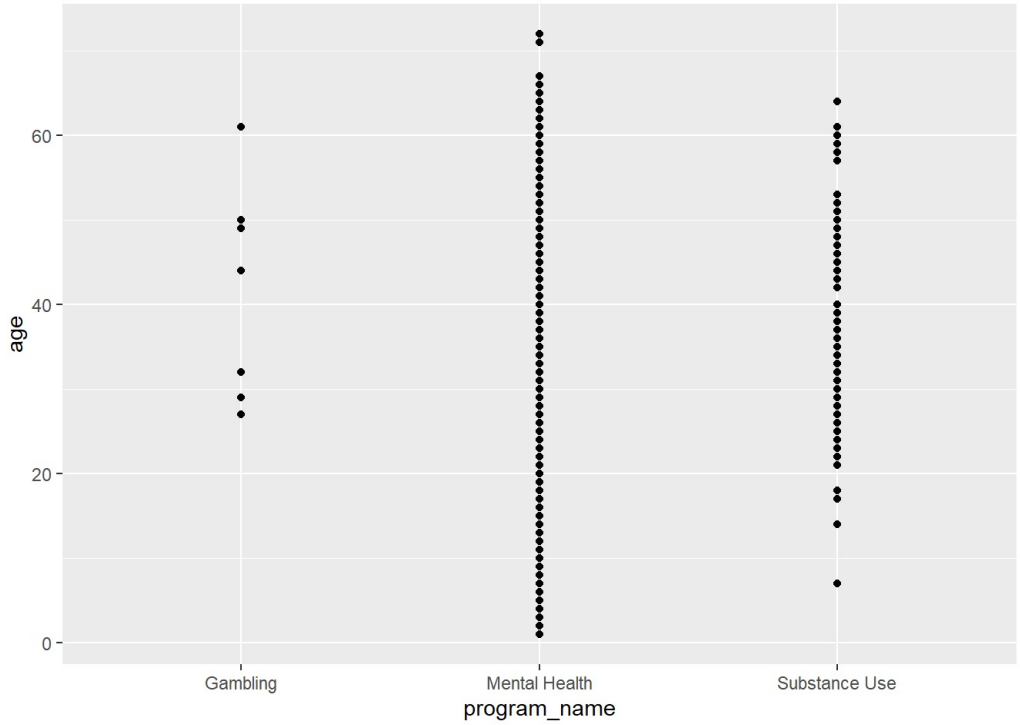
```
library(ggplot2)

df <- group_by(dat, program_name)
df2 <- summarise(df, ageMean = mean(age))

ggplot(data = df2, mapping = aes(x = program_name, y = ageMean)) +
  geom_point()
```




```
# This plot will give the idea which program has participated by
# a variety of age groups.
dat <- read.csv("HFS.csv") # entire data set
ggplot(data = dat, mapping = aes(x = program_name, y = age)) +
  geom_point()
```

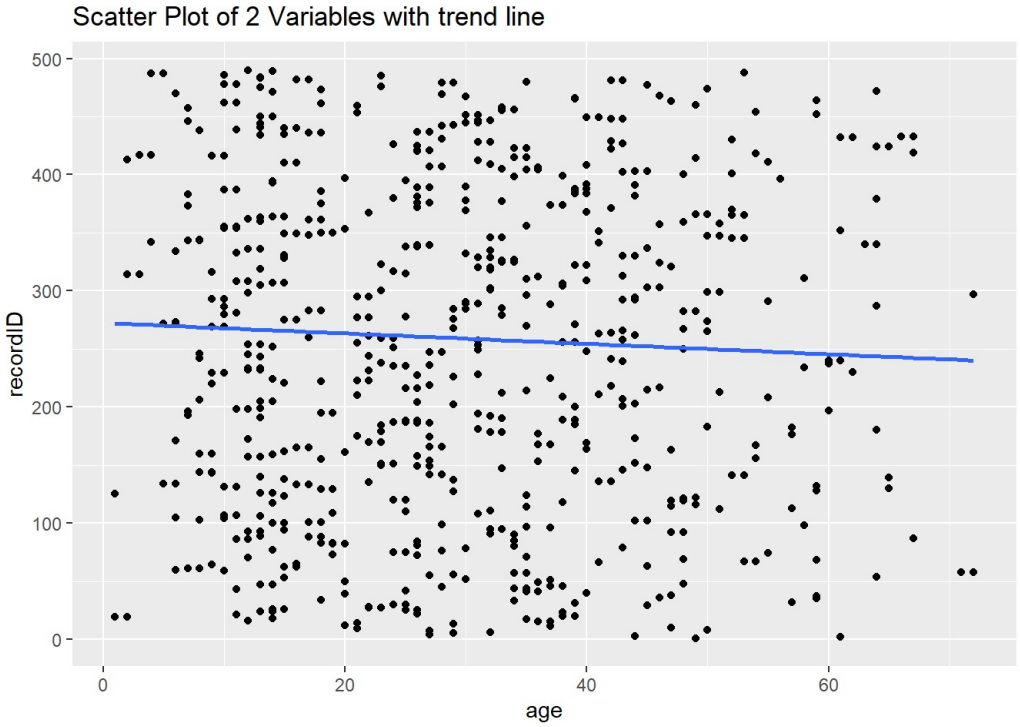


Scatter plot of 2 variables with trend line

This plot shows a linear relationship between age and recordID

```
ggplot(data = dat, mapping = aes(x = age, y = recordID)) +
  geom_point()+
  geom_smooth(method = 'lm', se= FALSE)+
  labs(title= "Scatter Plot of 2 Variables with trend line")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

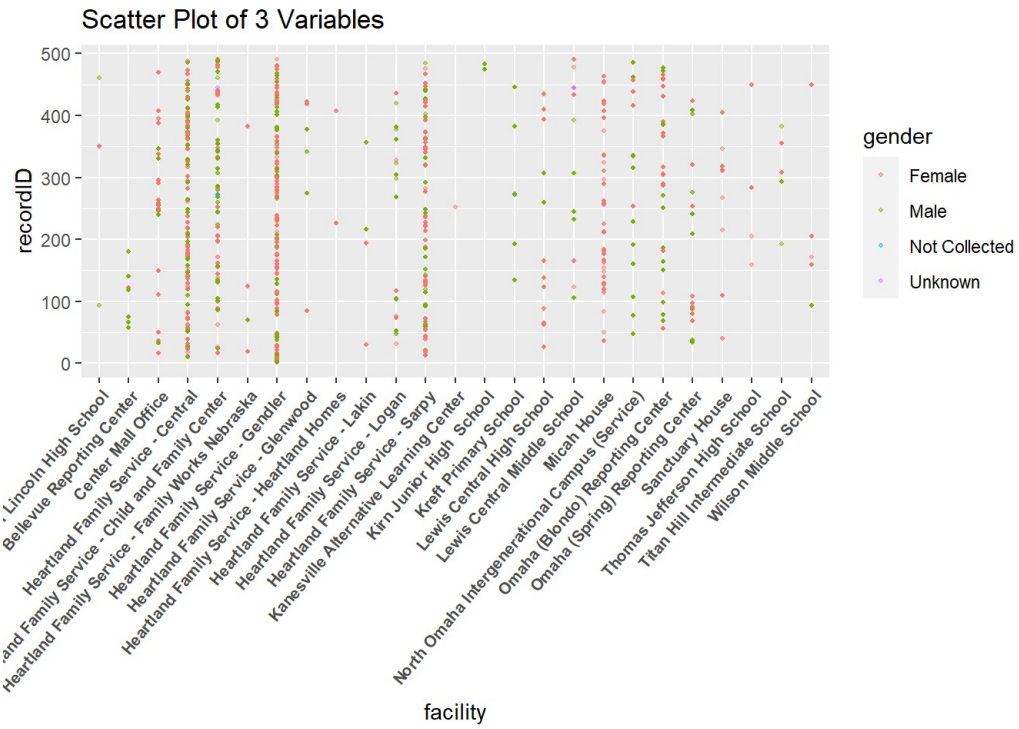


Scatter Plot with 3 Variables

This plot shows which facility has most of the recordID pariticipates based on the gender. The plot indicates which facility has more male or female no strange outliers in the graph

```
ggplot(data = dat, mapping = aes(facility, recordID, color = gender))+
  geom_point(size = 0.8, alpha = 0.5)+
  geom_smooth(method=lm, se=FALSE, col='red', size=2)+
  theme(axis.text.x = element_text(angle=50, size = 8, face = "bold",
                                   hjust=1, vjust = 1, lineheight = 5))+
  labs(title= "Scatter Plot of 3 Variables")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

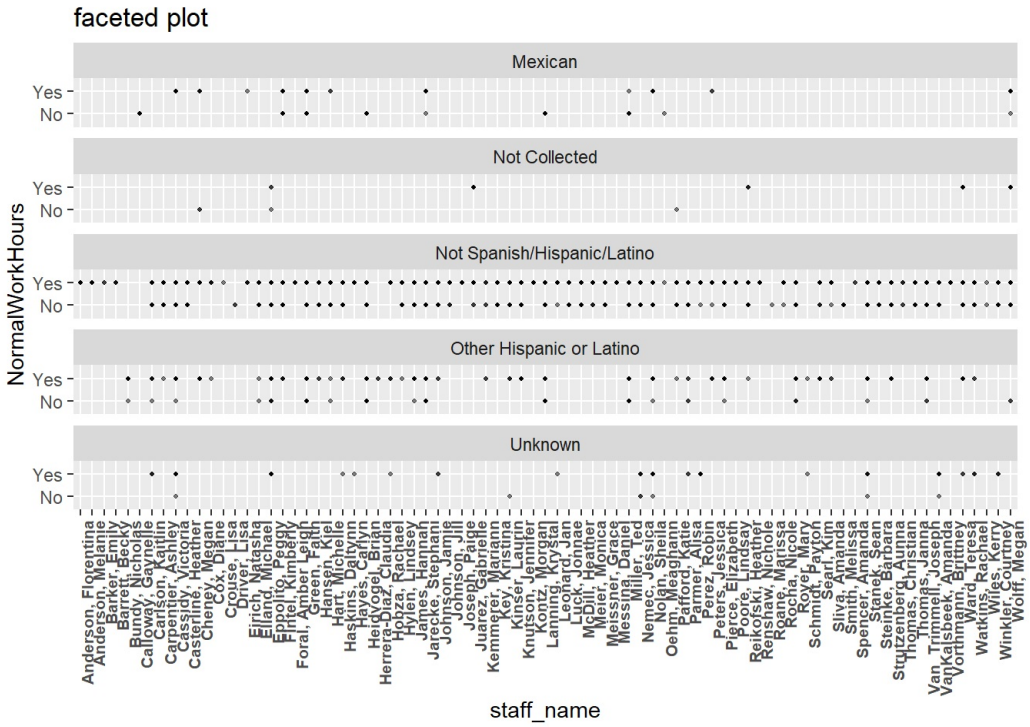


Facet plot

This plot shows which staff works mostly with all ethnicity (this will identify communication skills) and which staff works with his/her normal work schedule or not

```
ggplot(data = dat, mapping = aes(x = staff_name, y = NormalWorkHours
                                 )) +
  facet_wrap(~ ethnic_identity, nrow = 5)+
  theme(axis.text.x = element_text(angle=90, size = 8, face = "bold",
                                   hjust=1, vjust = 1, lineheight = 5))+
  geom_point(size = 0.8, alpha = 0.5)+
  geom_smooth(method = "lm", se = FALSE)+
  labs(title= "faceted plot")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Bar plot

This plot shows the average hours worked by each job_title this gives an idea for the organizer which job holders have more work in average

```
df5 <- dat %>% group_by(job_title) %>%
  summarize(AvgTotalDuration = mean(duration_num))

ggplot(data = df5, mapping = aes(x = job_title, y = AvgTotalDuration))+
  geom_col(aes(fill=job_title))+
  theme(axis.text.x = element_text(angle=90, size = 8, face = "bold",
                                   hjust=1, vjust = 1, lineheight = 5))+
labs(title= "bar plot")
```