Report for R

- Task 2-a: Find out the gender with largest number of records

Answer: names(which.max(table(df_data$gender)))

Explanation: Selecting gender column from the given dataset.

table() function is used to create a categorical representation of data with the variable name and the frequency in the form of a table.

which.max() returns the position of the element with the maximal value in a vector.

names() function is used to get or set the name of an Object.

- Task 2-b: Find out the total number of Residence_type "Urban" who are Male

Answer: count(df_data %>% filter(gender == 'Male' & Residence_type == 'Urban'))

Explanation: First I applied the filter() function for the dataset to filter the data where gender is male and resident type is Urban. Then applied pipe to forward the result to the next function call and the count() function to print the final count.

- 2-c 1 question #Find the top 10 ages with highest av_glucose_level

Answer: a<- df_data %>% group_by(age) %>% summarise(Mean=mean(avg_glucose_level)) %>% arrange(desc(Mean))

head(a$age,10)

Explanation: Created a data frame from the dataset where it returns the data with age and highest average glucose level using summarise() function and then returning the first 10 rows from the resulting dataset.

- 2-d 2nd question top 10 ages with more number of strokes

Answer: a<- df_data %>% group_by(age) %>% summarise(Mean=mean(stroke)) %>% arrange(desc(Mean))

head(a$age,10)

Explanation: Created a data frame from the dataset where it returns the data with age and highest number of stokes using summarise() function and then returning the first 10 rows from the resulting dataset.

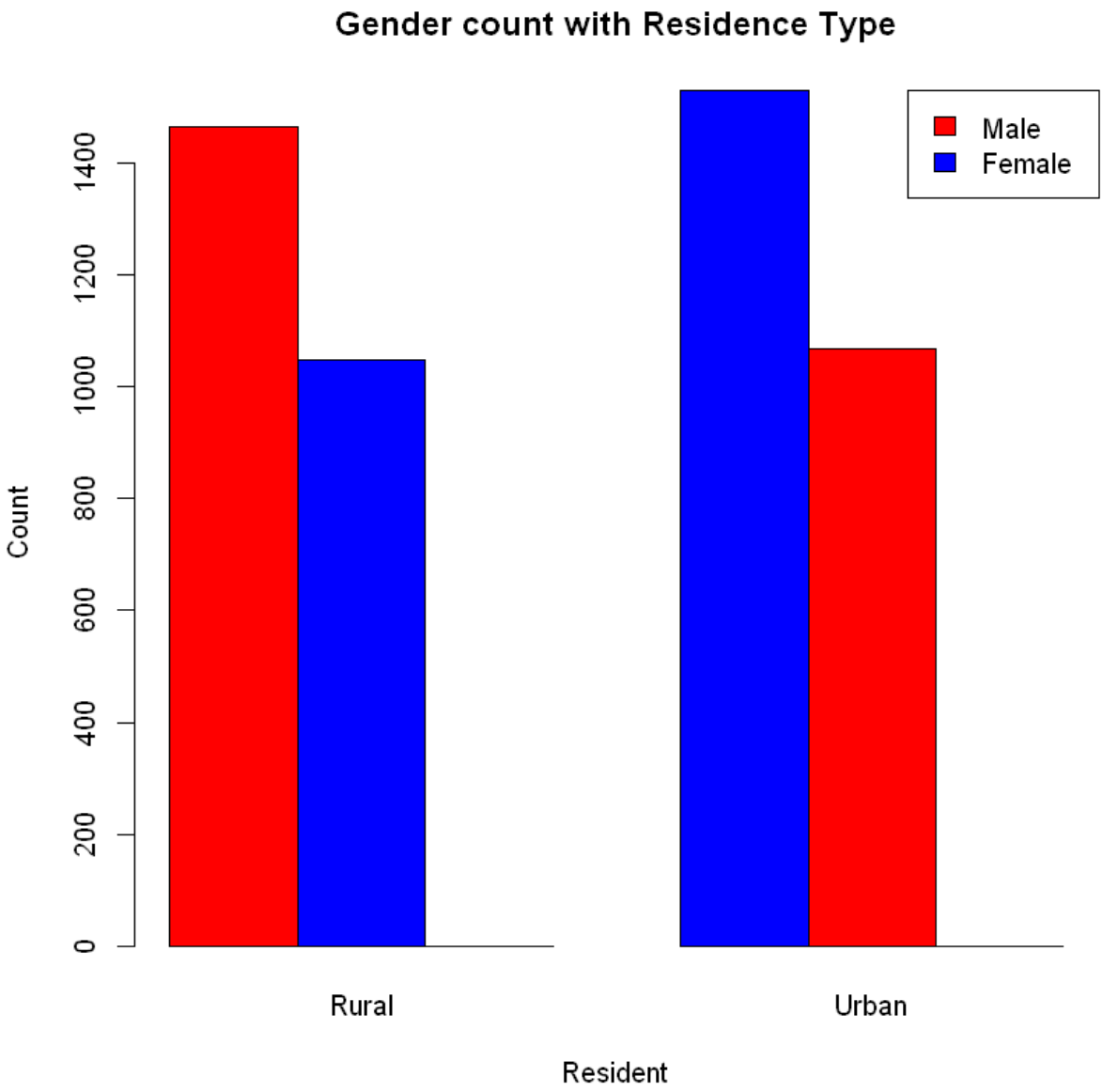
- Task 3-a: Create barplot showing gender with count with residence type

Answer:

```
df_data_barplot <- table(df_data$gender, df_data$Residence_type)
barplot(df_data_barplot,
      main="Gender count with Residence Type",
      xlab="Resident",
      ylab="Count",
      col = c("red","blue"),
      beside=TRUE
)
legend("topright",
    c("Male","Female"),
    fill=c("red","blue"))
```

Explanation:

Created a table that represents the gender and Resident type from the data set and then used barplot() function plotted the bar graph with the data from the resulting table with Resident type on X-axis and Count on the y-axis. The bars indicate the count of males and females with their resident types.



- Task 3-b: Display pie chart for the smoking status data

Answer:

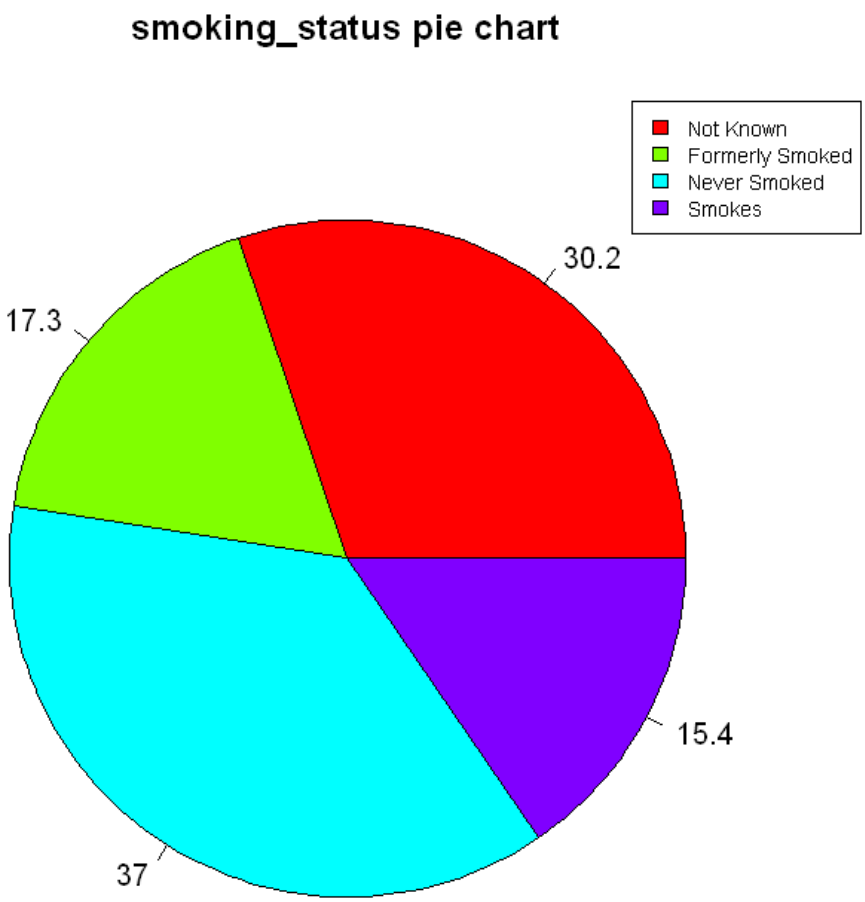
```
c<- df_data %>% count(df_data$smoking_status)

b_count <- round(c$n*100/nrow(df_data),1)

pie(c$n, labels=b_count,main= "smoking_status pie chart",col= rainbow(length(c$n)))

legend("topright", c("Not Known","Formerly Smoked","Never Smoked","Smokes"),cex = 0.7,fill =
rainbow(length(c$n)))
```

Explanation: Count the smoking status data from the dataset using the pie() function.



#Task4 finding an interesting pattern

atleast two visualization with explanation

Answer:

a <- table(df_data$gender,df_data$avg_glucose_level)

barplot(a, main = "Average Glucose Level of Gender", xlab = "Avg Glucose Level", ylab = "Gender", col = c("orange","green"))

legend("topright",c("Male","Female"),fill = c("orange","green"))

Explanation: Created a table that represents the gender and avg_glucose_level from the data set and then used barplot() function plotted the bar graph with the data from the resulting table with Resident type on X-axis and Count on the y-axis.