

project

2024-09-23

```
mydata <- read.csv("brain_stroke.csv")  
#mydata
```

```
totalsize <- nrow(mydata)  
  
gender_counts <- table(mydata$gender)  
stroke_counts <- table(mydata$stroke)  
stroke_gender_counts <- table(mydata$gender[mydata$stroke == 1])  
nostroke_gender_counts <- table(mydata$gender[mydata$stroke == 0])  
  
gender_percentage <- prop.table(gender_counts) * 100  
stroke_gender_proportion <- prop.table(stroke_gender_counts) * 100  
nostroke_gender_proportion <- prop.table(nostroke_gender_counts) * 100  
  
cat("total size is:", totalsize, "\n")
```

```
## total size is: 4981
```

```
cat("gender counts is: ", gender_counts, "percentage is: ", gender_percentage, "\n")
```

```
## gender counts is: 2907 2074 percentage is: 58.36177 41.63823
```

```
stroke_counts
```

```
##  
##      0      1  
## 4733  248
```

```
stroke_gender_counts
```

```
##  
## Female    Male  
##    140    108
```

```
stroke_gender_proportion
```

```
##  
##   Female      Male  
## 56.45161 43.54839
```

```
nostroke_gender_counts
```

```
##  
## Female    Male  
##    2767    1966
```

```
nostroke_gender_proportion
```

```
##  
## Female      Male  
## 58.46186 41.53814
```

```
cat("Hypertension", "\n")
```

```
## Hypertension
```

```
hypertension_counts <- table(mydata$hypertension)  
hypertension_percentage <- prop.table(hypertension_counts) * 100  
hypertension_counts
```

```
##  
##    0    1  
## 4502  479
```

```
hypertension_percentage
```

```
##  
##          0          1  
## 90.383457  9.616543
```

```
hypertensionwithstroke <- table(mydata$hypertension[mydata$stroke == 1])  
hypertensionwithstroke
```

```
##  
##    0    1  
## 182   66
```

```
prop.table(hypertensionwithstroke) * 100
```

```
##  
##          0          1  
## 73.3871 26.6129
```

```
hypertensionwithoutstroke <- table(mydata$hypertension[mydata$stroke == 0])  
hypertensionwithoutstroke
```

```
##  
##    0    1  
## 4320  413
```

```
prop.table(hypertensionwithoutstroke) * 100
```

```
##  
##           0           1  
## 91.274033  8.725967
```

```
cat("\n")
```

```
cat("Heartdisease", "\n")
```

```
## Heartdisease
```

```
heart_counts <- table(mydata$heart_disease)  
heart_counts
```

```
##  
##      0      1  
## 4706  275
```

```
prop.table(heart_counts) * 100
```

```
##  
##           0           1  
## 94.47902  5.52098
```

```
heartwithstroke <- table(mydata$heart_disease[mydata$stroke == 1])  
heartwithstroke
```

```
##  
##      0      1  
## 201   47
```

```
prop.table(heartwithstroke) * 100
```

```
##  
##           0           1  
## 81.04839 18.95161
```

```
heartwithoutstroke <- table(mydata$heart_disease[mydata$stroke == 0])  
heartwithoutstroke
```

```
##  
##      0      1  
## 4505  228
```

```
prop.table(heartwithoutstroke) * 100
```

```
##  
##           0           1  
## 95.182759  4.817241
```

```
cat("\n")
```

```
cat("Marriage Status", "\n")
```

```
## Marriage Status
```

```
marriage <- table(mydata$ever_married)
marriage
```

```
##
##   No   Yes
## 1701 3280
```

```
prop.table(marriage) * 100
```

```
##
##      No      Yes
## 34.14977 65.85023
```

```
marriagewithstroke <- table(mydata$ever_married[mydata$stroke == 1])
marriagewithstroke
```

```
##
##   No   Yes
##   29  219
```

```
prop.table(marriagewithstroke) * 100
```

```
##
##      No      Yes
## 11.69355 88.30645
```

```
marriagewithoutstroke <- table(mydata$ever_married[mydata$stroke == 0])
marriagewithoutstroke
```

```
##
##   No   Yes
## 1672 3061
```

```
prop.table(marriagewithoutstroke) * 100
```

```
##
##      No      Yes
## 35.32643 64.67357
```

```
cat("Work Type", "\n")
```

```
## Work Type
```

```
work <- table(mydata$work_type)
work
```

```
##
##      children      Govt_job      Private Self-employed
##          673          644          2860          804
```

```
prop.table(work) * 100
```

```
##
##      children      Govt_job      Private Self-employed
##    13.51134    12.92913    57.41819    16.14134
```

```
workwithstroke <- table(mydata$work_type[mydata$stroke == 1])
workwithstroke
```

```
##
##      children      Govt_job      Private Self-employed
##           2          33          148          65
```

```
prop.table(workwithstroke) * 100
```

```
##
##      children      Govt_job      Private Self-employed
##    0.8064516    13.3064516    59.6774194    26.2096774
```

```
workwithoutstroke <- table(mydata$work_type[mydata$stroke == 0])
workwithoutstroke
```

```
##
##      children      Govt_job      Private Self-employed
##          671          611          2712          739
```

```
prop.table(workwithoutstroke) * 100
```

```
##
##      children      Govt_job      Private Self-employed
##    14.17705    12.90936    57.29981    15.61378
```

```
cat("Residence Type", "\n")
```

```
## Residence Type
```

```
resident <- table(mydata$Residence_type)
resident
```

```
##
## Rural Urban
##  2449  2532
```

```
prop.table(resident) * 100
```

```
##  
##      Rural      Urban  
## 49.16683 50.83317
```

```
residentwithstroke <- table(mydata$Residence_type[mydata$stroke == 1])  
residentwithstroke
```

```
##  
## Rural Urban  
##   113   135
```

```
prop.table(residentwithstroke) * 100
```

```
##  
##      Rural      Urban  
## 45.56452 54.43548
```

```
residentwithoutstroke <- table(mydata$Residence_type[mydata$stroke == 0])  
residentwithoutstroke
```

```
##  
## Rural Urban  
## 2336 2397
```

```
prop.table(residentwithoutstroke) * 100
```

```
##  
##      Rural      Urban  
## 49.35559 50.64441
```

```
cat("Smoking Status", "\n")
```

```
## Smoking Status
```

```
smoking <- table(mydata$smoking_status)  
smoking
```

```
##  
## formerly smoked      never smoked      smokes      Unknown  
##           867           1838           776           1500
```

```
prop.table(smoking) * 100
```

```
##  
## formerly smoked      never smoked      smokes      Unknown  
##      17.40614      36.90022      15.57920      30.11443
```

```
smokingwithstroke <- table(mydata$smoking_status[mydata$stroke == 1])
smokingwithstroke
```

```
##
## formerly smoked    never smoked      smokes      Unknown
##              70              89              42              47
```

```
prop.table(smokingwithstroke) * 100
```

```
##
## formerly smoked    never smoked      smokes      Unknown
##      28.22581      35.88710      16.93548      18.95161
```

```
smokingwithoutstroke <- table(mydata$smoking_status[mydata$stroke == 0])
smokingwithoutstroke
```

```
##
## formerly smoked    never smoked      smokes      Unknown
##              797              1749              734              1453
```

```
prop.table(smokingwithoutstroke) * 100
```

```
##
## formerly smoked    never smoked      smokes      Unknown
##      16.83921      36.95331      15.50813      30.69935
```

```
sapply(mydata[, c("age", "avg_glucose_level", "bmi")], function(x) c(mean = mean(x, na.rm = TRUE),
                                                                    median = median(x, na.rm = TRUE),
                                                                    sd = sd(x, na.rm = TRUE),
                                                                    range = range(x, na.rm = TRUE)))
```

```
##          age avg_glucose_level      bmi
## mean   43.41986      105.94356 28.498173
## median 45.00000      91.85000 28.100000
## sd     22.66276      45.07537  6.790464
## range1  0.08000      55.12000 14.000000
## range2 82.00000      271.74000 48.900000
```

```
calc_stats <- function(data) {
  c(Mean = mean(data, na.rm = TRUE),
    Median = median(data, na.rm = TRUE),
    SD = sd(data, na.rm = TRUE),
    Range = range(data, na.rm = TRUE))
}
age_stats <- sapply(split(mydata$age, mydata$stroke), calc_stats)

cat("Statistics for Age:\n")
```

```
## Statistics for Age:
```

```
print(age_stats)
```

```
##           0           1
## Mean    42.14135 67.81984
## Median  43.00000 71.00000
## SD      22.34504 12.67056
## Range1   0.08000 1.32000
## Range2  82.00000 82.00000
```

```
glucose_stats <- sapply(split(mydata$avg_glucose_level, mydata$stroke), calc_stats)
cat("Statistics for glucose level:\n")
```

```
## Statistics for glucose level:
```

```
print(glucose_stats)
```

```
##           0           1
## Mean   104.56906 132.17556
## Median  91.45000 105.04000
## SD      43.60219 61.77108
## Range1  55.12000 56.11000
## Range2 267.76000 271.74000
```

```
bmi_stats <- sapply(split(mydata$bmi, mydata$stroke), calc_stats)
cat("Statistics for bmi:\n")
```

```
## Statistics for bmi:
```

```
print(bmi_stats)
```

```
##           0           1
## Mean    28.409698 30.186694
## Median  28.000000 29.450000
## SD       6.833637  5.658245
## Range1  14.000000 16.900000
## Range2  48.900000 48.900000
```

Boxplots

```
par(mfrow = c(1, 2))

boxplot(age ~ stroke, data = mydata,
        xlab = "Stroke Status",
        ylab = "Age",
        names = c("No Stroke", "Stroke"),
        main = "Age vs Stroke Status")

boxplot(avg_glucose_level ~ stroke, data = mydata,
```

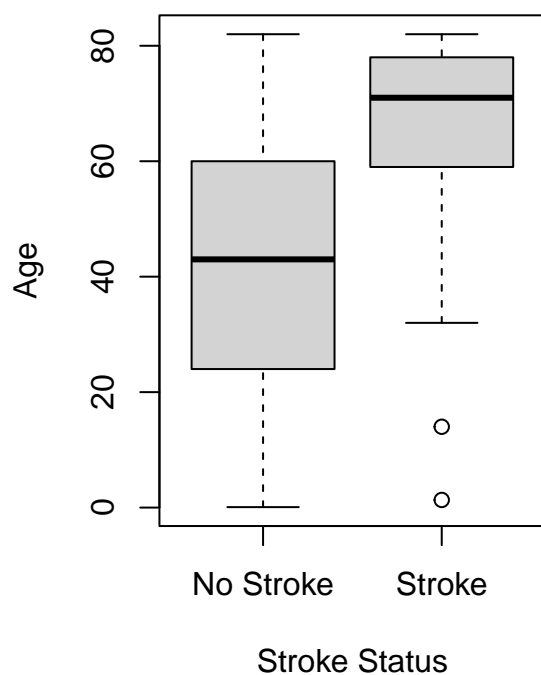


```

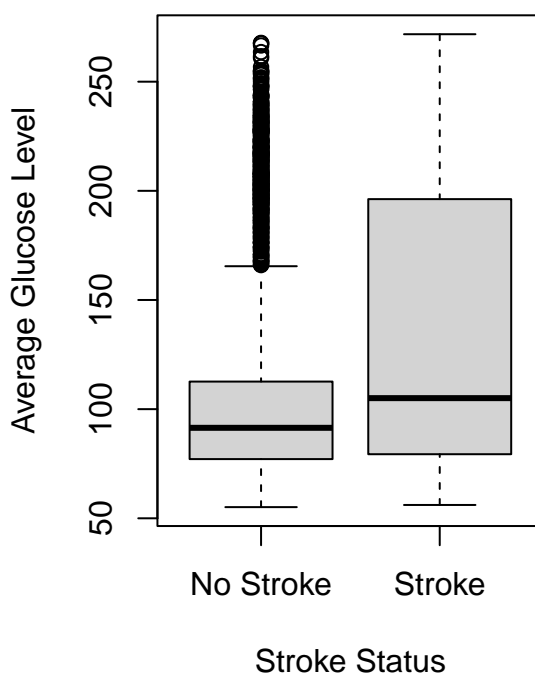
xlab = "Stroke Status",
ylab = "Average Glucose Level",
names = c("No Stroke", "Stroke"),
main = "Avg Glucose Level vs Stroke Status")

```

Age vs Stroke Status



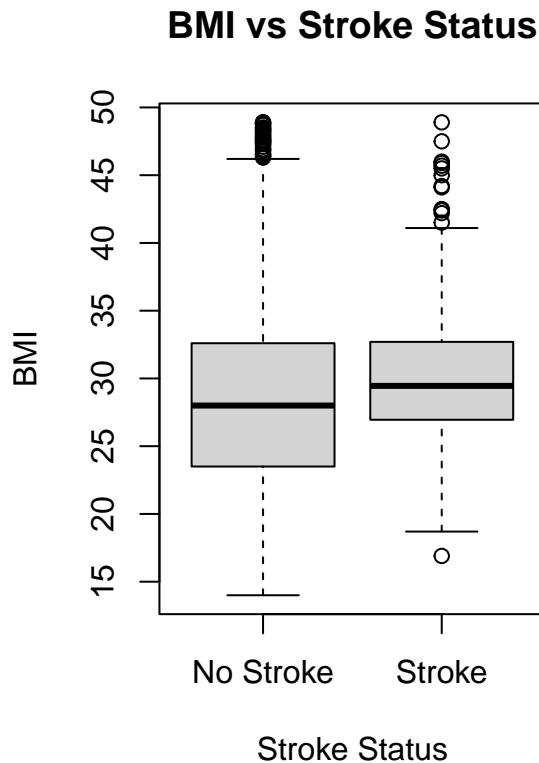
Avg Glucose Level vs Stroke Stat



```

boxplot(bmi ~ stroke, data = mydata,
xlab = "Stroke Status",
ylab = "BMI",
names = c("No Stroke", "Stroke"),
main = "BMI vs Stroke Status")

```



Stacked Barplots

```

categorical_vars <- c("gender", "hypertension", "heart_disease", "ever_married",
                      "work_type", "Residence_type", "smoking_status")

par(mfrow = c(1,2), mar = c(5, 3, 8, 8))
for (var in categorical_vars) {
  table_data <- table(mydata[[var]], mydata$stroke)
  prop_data <- prop.table(table_data, margin = 2)

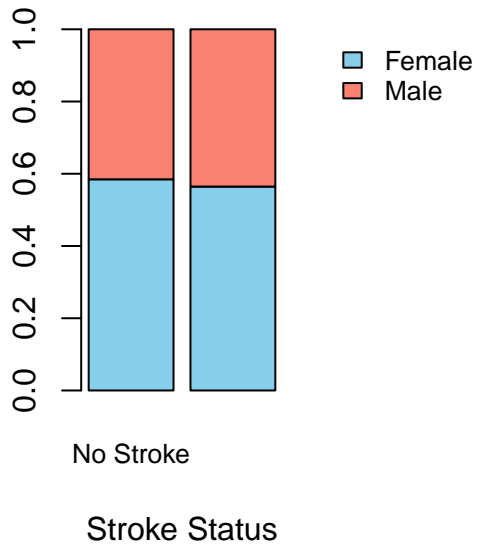
  barplot(prop_data,
    beside = FALSE,
    col = c("skyblue", "salmon", "lightgreen", "lightpink", "orange"),
    xlab = "Stroke Status",
    ylab = "Proportion",
    main = paste("Stroke Status for", var),
    names.arg = c("No Stroke", "Stroke"),
    cex.names = 0.8,
    cex.main = 0.9)

  legend("topright", inset = c(-1, 0),
    legend = rownames(prop_data),
    fill = c("skyblue", "salmon", "lightgreen", "lightpink", "orange"),
    xpd = TRUE, bty = "n", cex = 0.8)
}

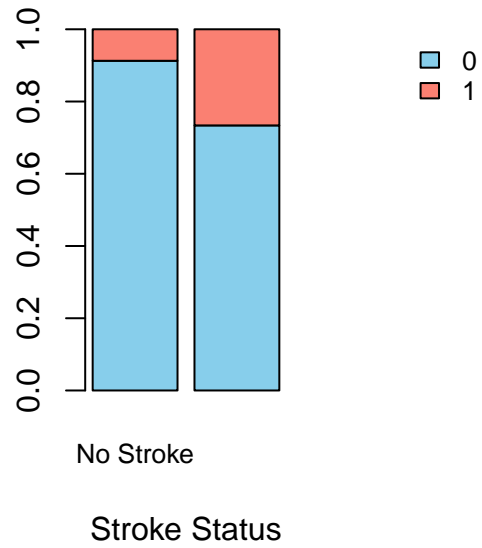
```

}

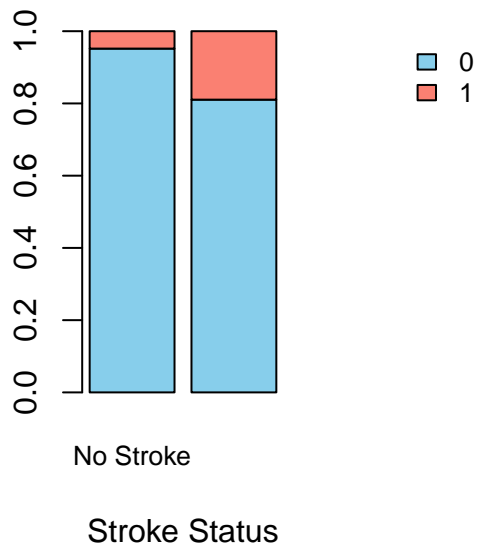
Stroke Status for gender



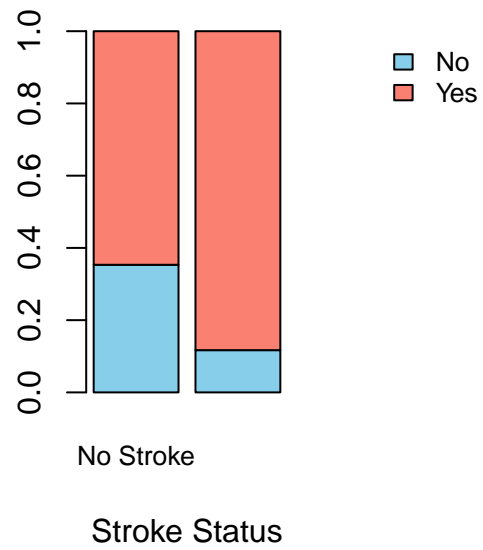
Stroke Status for hypertension



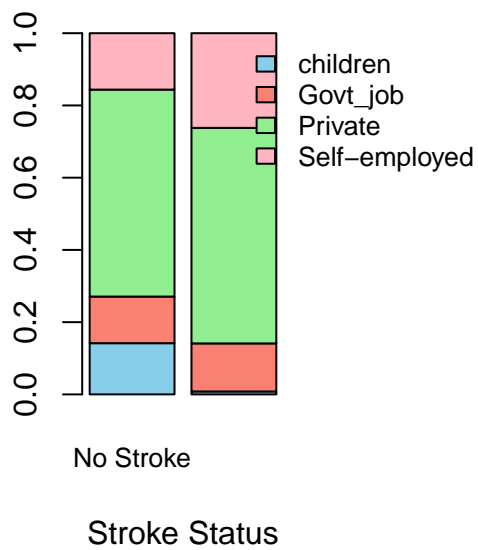
Stroke Status for heart_disease



Stroke Status for ever_married



Stroke Status for work_type



Stroke Status for Residence_type



stroke Status for smoking_status

