

# Toxoplasmosis in Cats: Risk Factor Analysis

Jonathan Ndayisenga

#Descriptive Analysis #Summary of the Dataset

```
# Load Excel data
# Fixed: Read directly from the only sheet, which already contains everything
merged_data <- read_excel("merged_toxoplasmosis_data.xlsx", sheet = "Sheet1") %>%
  clean_names()

# Create toxoplasmosis status variable
merged_data <- merged_data %>%
  mutate(toxo_status = factor(ifelse(ig_m == "Positive" | ig_g == "Positive", "Yes", "No")))
```

```
summary(merged_data)
```

```
##   cat_number_1   owners_age   sex_3   occupation
##   Min.   : 1.00   Min.   : 9.00   Length:80   Length:80
##   1st Qu.:20.75   1st Qu.:26.00   Class :character   Class :character
##   Median :40.50   Median :30.00   Mode  :character   Mode  :character
##   Mean   :40.50   Mean   :34.94
##   3rd Qu.:60.25   3rd Qu.:45.00
##   Max.   :80.00   Max.   :70.00
##
##   education_level   pregnancy   hiv   housing_type
##   Length:80   Length:80   Length:80   Length:80
##   Class :character   Class :character   Class :character   Class :character
##   Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##   cat_number_9   name_of_cat   age_months   sex_12
##   Min.   : 1.00   Length:80   Min.   : 3.00   Length:80
##   1st Qu.:20.75   Class :character   1st Qu.: 12.50   Class :character
##   Median :40.50   Mode  :character   Median : 30.00   Mode  :character
##   Mean   :40.50   Mean   : 31.61
##   3rd Qu.:60.25   3rd Qu.: 48.00
##   Max.   :80.00   Max.   :192.00
##   NA's   :1
##   status   color   where_did_you_get_the_cat_from
##   Length:80   Length:80   Length:80
##   Class :character   Class :character   Class :character
##   Mode  :character   Mode  :character   Mode  :character
##
##
```

```
##
##
## other_cats_at_home if_yes_how_many type_of_food other_animals_at_home
## Length:80 Min. : 0.000 Length:80 Length:80
## Class :character 1st Qu.: 0.000 Class :character Class :character
## Mode :character Median : 2.000 Mode :character Mode :character
## Mean : 7.575
## 3rd Qu.:15.000
## Max. :30.000
##
## what_animal outdoor_access hours cat_disappear
## Length:80 Length:80 Min. : 0.000 Length:80
## Class :character Class :character 1st Qu.: 0.000 Class :character
## Mode :character Mode :character Median : 5.000 Mode :character
## Mean : 7.362
## 3rd Qu.:12.000
## Max. :24.000
##
## days interact_with_other_cats litterbox
## Min. : 0.0 Length:80 Length:80
## 1st Qu.: 0.0 Class :character Class :character
## Median : 0.0 Mode :character Mode :character
## Mean : 566.7
## 3rd Qu.: 0.0
## Max. :45323.0
##
## who_emptyes_it type_of_litter cat_number_29 ig_g
## Length:80 Length:80 Min. : 1.00 Length:80
## Class :character Class :character 1st Qu.:20.75 Class :character
## Mode :character Mode :character Median :40.50 Mode :character
## Mean :40.50
## 3rd Qu.:60.25
## Max. :80.00
##
## ig_m toxo_status
## Length:80 No :71
## Class :character Yes: 9
## Mode :character
##
##
##
```

```
total_missing <- sum(is.na(merged_data))
missing_by_column <- colSums(is.na(merged_data))

total_missing
```

```
## [1] 144
```

```
missing_by_column
```

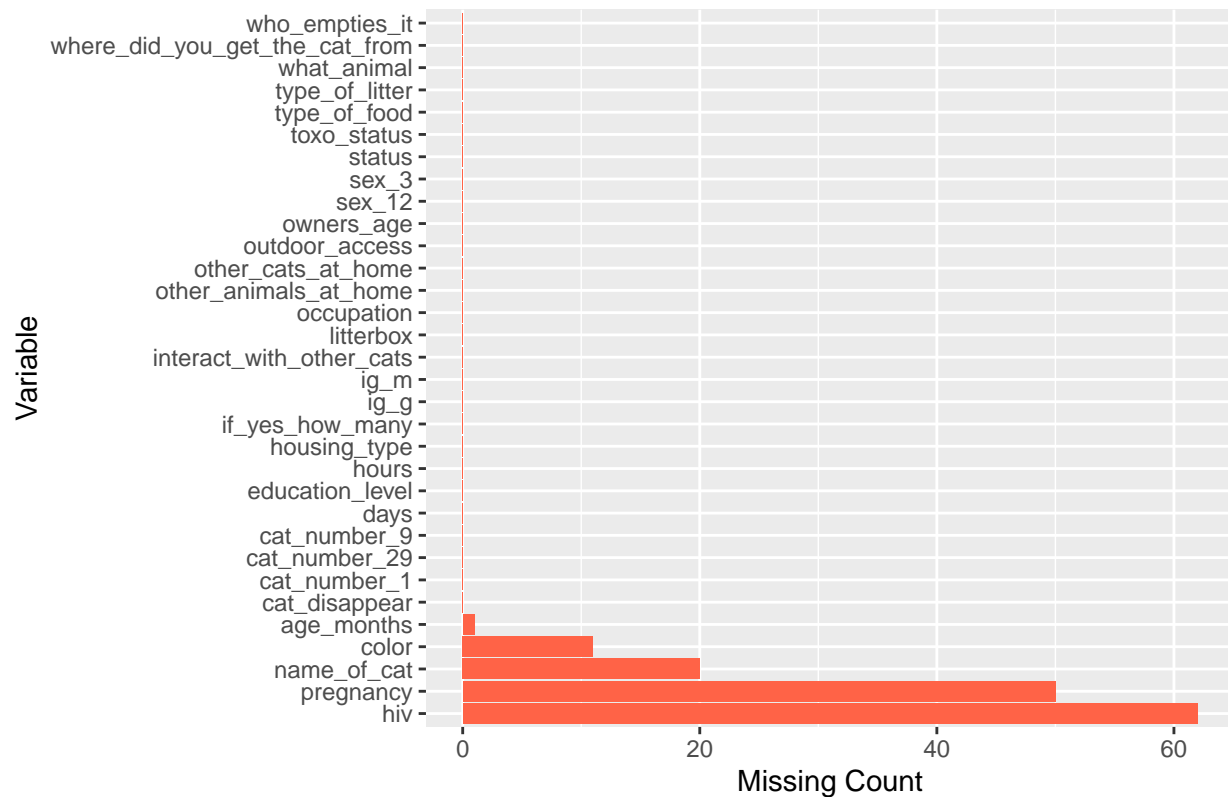
```
## cat_number_1 owners_age
```

```
##          0          0
##          sex_3      occupation
##          0          0
##          education_level pregnancy
##          0          50
##          hiv        housing_type
##          62         0
##          cat_number_9 name_of_cat
##          0          20
##          age_months   sex_12
##          1           0
##          status       color
##          0           11
## where_you_did_you_get_the_cat_from other_cats_at_home
##          0           0
##          if_yes_how_many type_of_food
##          0           0
##          other_animals_at_home what_animal
##          0           0
##          outdoor_access      hours
##          0           0
##          cat_disappear       days
##          0           0
##          interact_with_other_cats litterbox
##          0           0
##          who_emptyies_it      type_of_litter
##          0           0
##          cat_number_29       ig_g
##          0           0
##          ig_m               toxo_status
##          0           0
```

```
missing_df <- data.frame(
  variable = names(missing_by_column),
  missing_count = as.numeric(missing_by_column)
)

ggplot(missing_df, aes(x = reorder(variable, -missing_count), y = missing_count)) +
  geom_bar(stat = "identity", fill = "tomato") +
  coord_flip() +
  labs(title = "Missing Data by Column", x = "Variable", y = "Missing Count")
```

## Missing Data by Column



```
summary(merged_data$owners_age)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      9.00  26.00   30.00   34.94  45.00   70.00
```

```
summary(merged_data$age_months)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##      3.00  12.50   30.00   31.61  48.00  192.00      1
```

```
sd(merged_data$owners_age, na.rm = TRUE)
```

```
## [1] 14.82928
```

```
sd(merged_data$age_months, na.rm = TRUE)
```

```
## [1] 26.12073
```

```
table(merged_data$sex_3)
```

```
##
##  F  M
## 36 44
```

```
table(merged_data$housing_type)
```

```
##  
## Bungalow    Flat    Hotel    House    Hut  
##      61      9      1      3      6
```

```
table(merged_data$education_level)
```

```
##  
##      None    Primary Secondary  Tertiary  
##        4      21      8      47
```

```
prop.table(table(merged_data$sex_3))
```

```
##  
##      F      M  
## 0.45 0.55
```

```
prop.table(table(merged_data$housing_type))
```

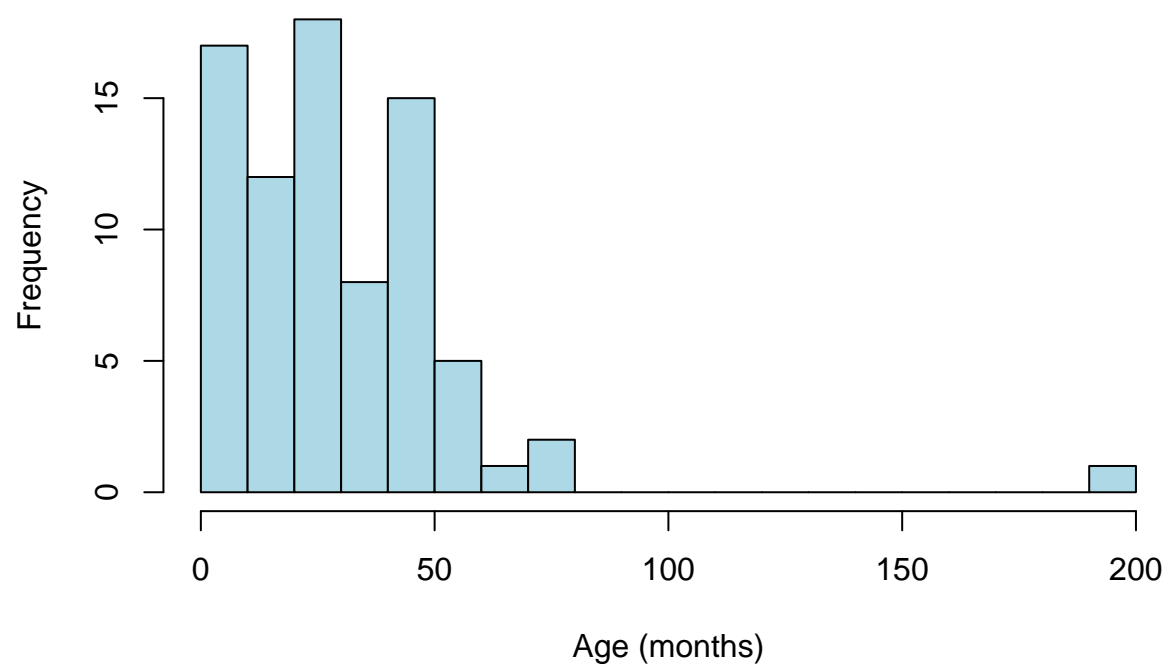
```
##  
## Bungalow    Flat    Hotel    House    Hut  
## 0.7625 0.1125 0.0125 0.0375 0.0750
```

```
prop.table(table(merged_data$education_level))
```

```
##  
##      None    Primary Secondary  Tertiary  
## 0.0500 0.2625 0.1000 0.5875
```

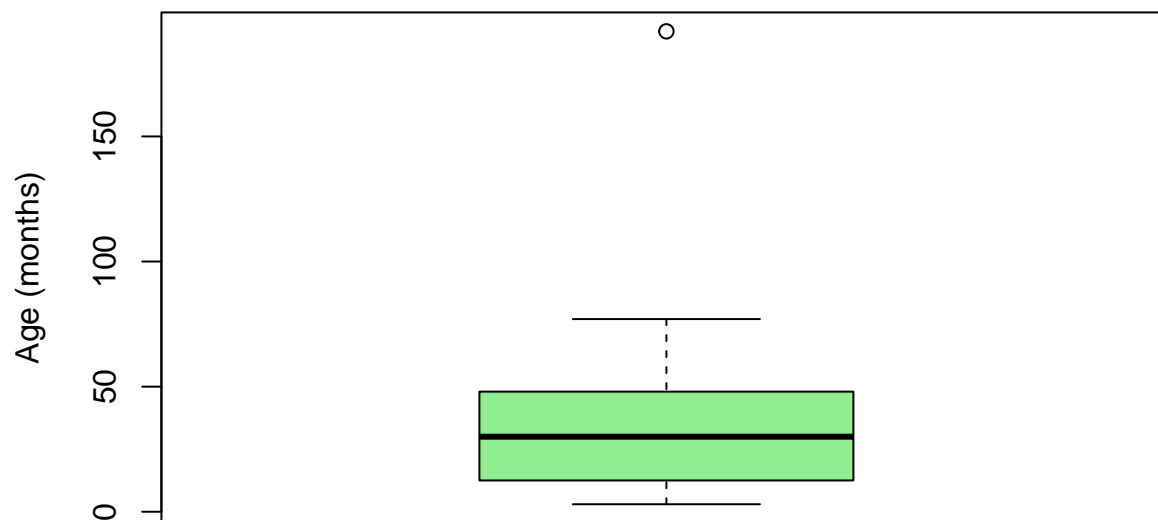
```
hist(merged_data$age_months, main = "Age Distribution of Cats", xlab = "Age (months)", col = "lightblue")
```

## Age Distribution of Cats



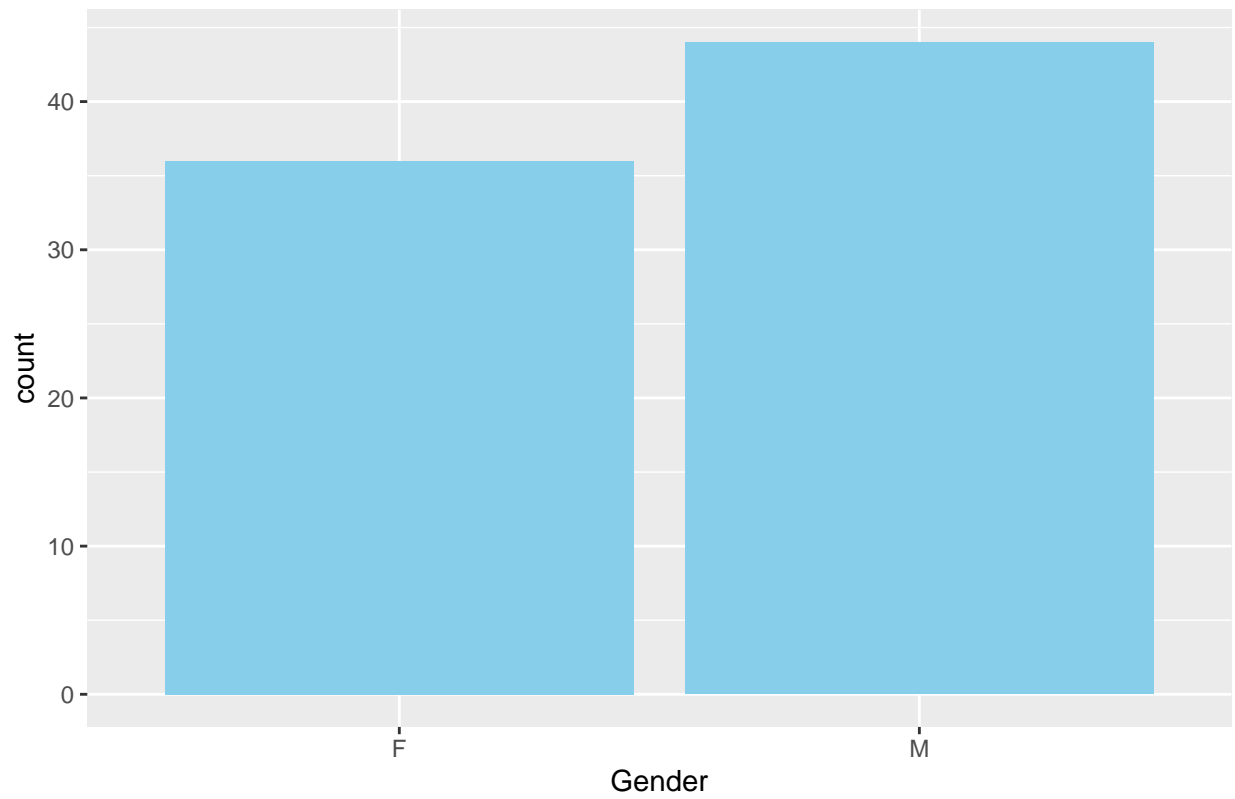
```
boxplot(merged_data$age_months, main = "Boxplot of Cat Ages", ylab = "Age (months)", col = "lightgreen")
```

## Boxplot of Cat Ages



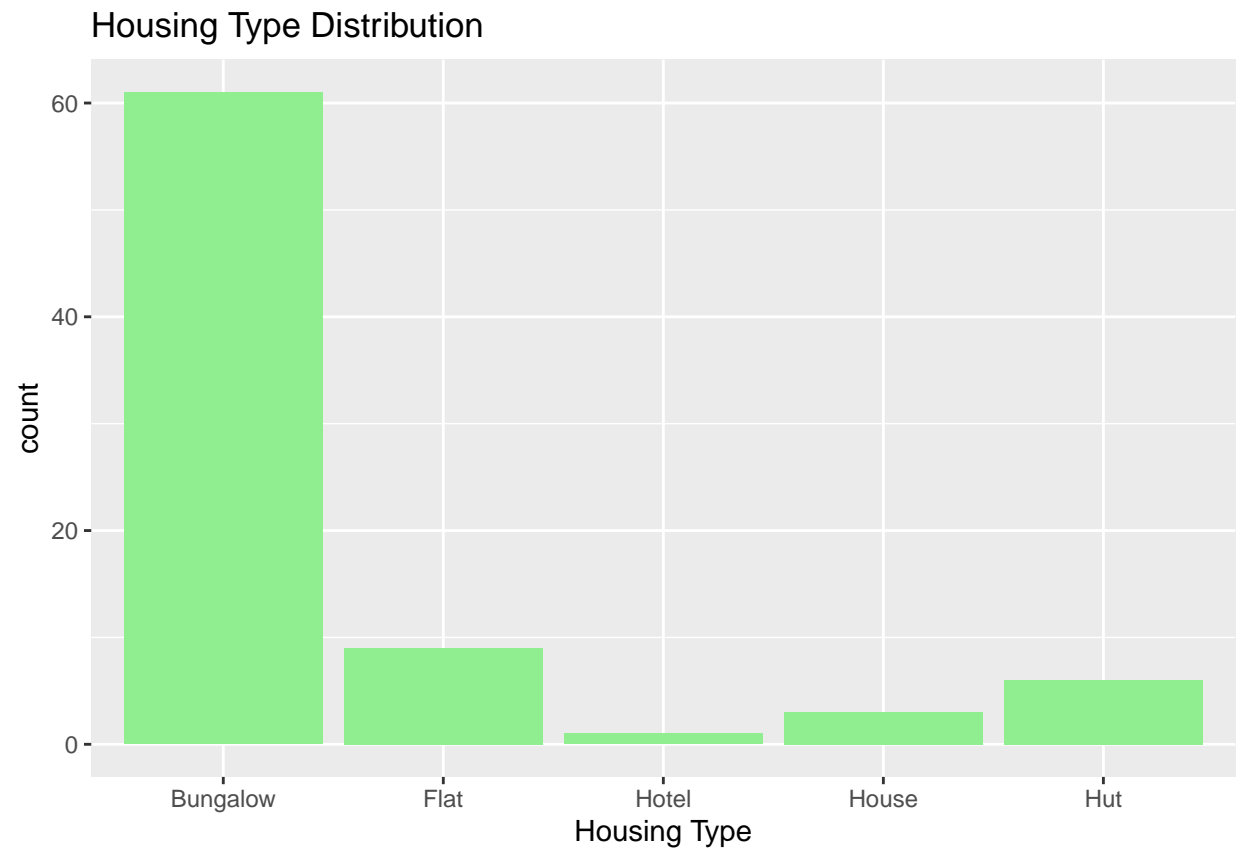
```
ggplot(merged_data, aes(x = sex_3)) +  
  geom_bar(fill = "skyblue") +  
  labs(title = "Cat Gender Distribution", x = "Gender")
```

Cat Gender Distribution

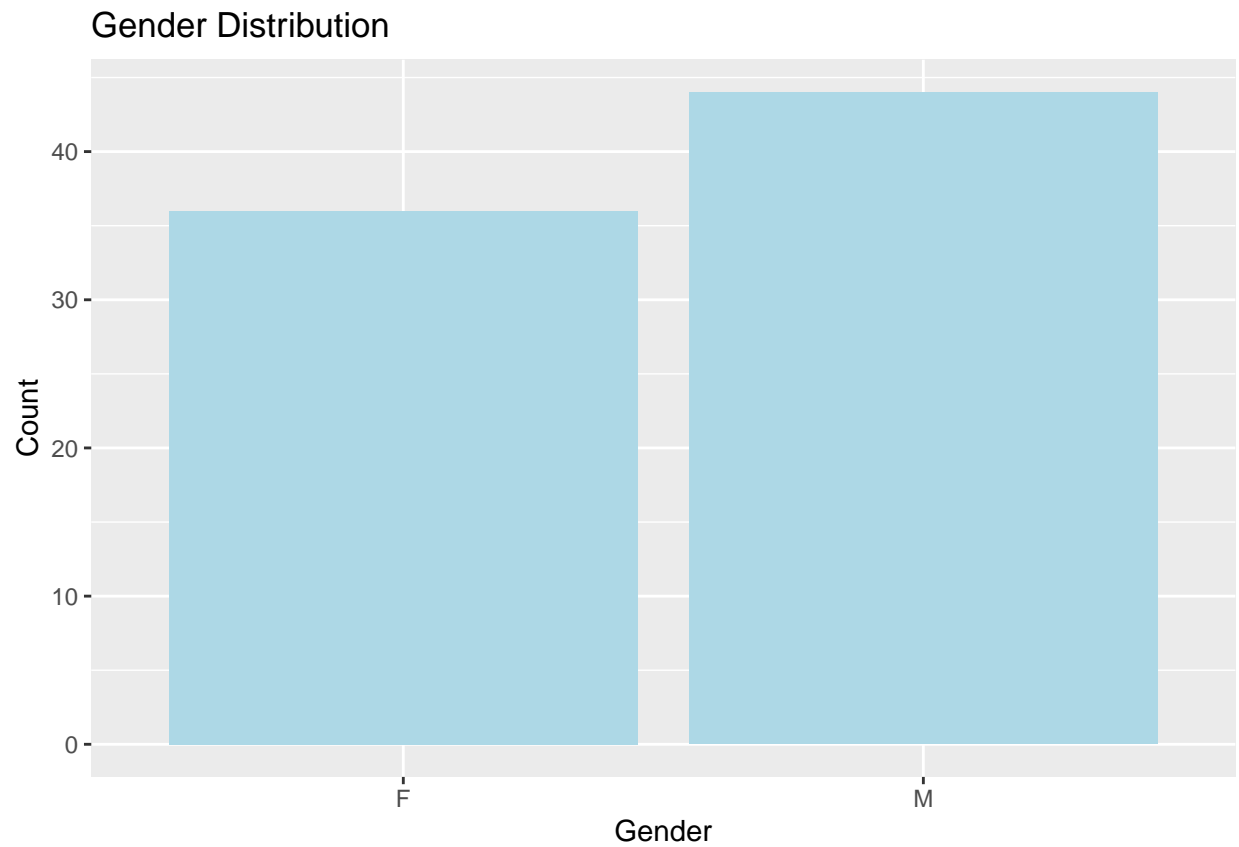


```
ggplot(merged_data, aes(x = housing_type)) +  
  geom_bar(fill = "lightgreen") +  
  labs(title = "Housing Type Distribution", x = "Housing Type")
```

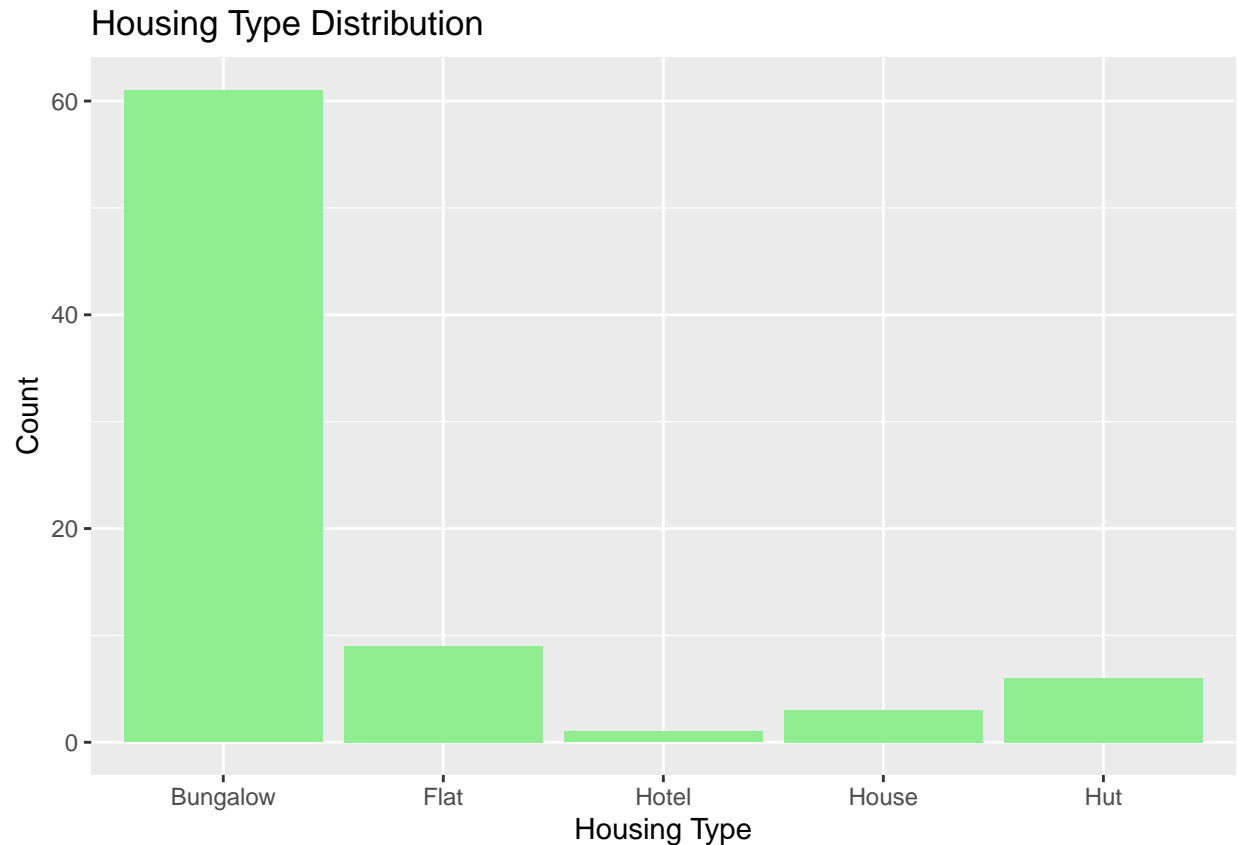




```
ggplot(merged_data, aes(x = sex_3)) +  
  geom_bar(fill = "lightblue") +  
  labs(title = "Gender Distribution", x = "Gender", y = "Count")
```



```
ggplot(merged_data, aes(x = `housing_type`)) +  
  geom_bar(fill = "lightgreen") +  
  labs(title = "Housing Type Distribution", x = "Housing Type", y = "Count")
```



```
model <- glm(toxo_status ~ age_months + sex_3 + housing_type + education_level,
             data = merged_data, family = binomial)
```

```
summary(model)
```

```
##
## Call:
## glm(formula = toxo_status ~ age_months + sex_3 + housing_type +
##      education_level, family = binomial, data = merged_data)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -8.239e-01  1.925e+00  -0.428   0.669
## age_months     9.033e-03  1.158e-02   0.780   0.435
## sex_3M        -2.067e-01  8.440e-01  -0.245   0.807
## housing_typeFlat -1.739e+01  3.787e+03  -0.005   0.996
## housing_typeHotel -1.708e+01  1.075e+04  -0.002   0.999
## housing_typeHouse -1.731e+01  6.204e+03  -0.003   0.998
## housing_typeHut  -6.807e-01  1.542e+00  -0.441   0.659
## education_levelPrimary -5.966e-01  1.651e+00  -0.361   0.718
## education_levelSecondary -1.892e+01  3.796e+03  -0.005   0.996
## education_levelTertiary -1.504e+00  1.786e+00  -0.842   0.400
##
## (Dispersion parameter for binomial family taken to be 1)
##
```





```

## incorrect

##
## Pearson's Chi-squared test
##
## data:  table(merged_data$housing_type, merged_data$toxoplasmosis_status)
## X-squared = 2.0367, df = 4, p-value = 0.729

chisq.test(table(merged_data$sex_3, merged_data$toxoplasmosis_status))

## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect

##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(merged_data$sex_3, merged_data$toxoplasmosis_status)
## X-squared = 3.5824e-30, df = 1, p-value = 1

fisher.test(table(merged_data$education_level, merged_data$toxoplasmosis_status))

##
## Fisher's Exact Test for Count Data
##
## data:  table(merged_data$education_level, merged_data$toxoplasmosis_status)
## p-value = 0.2829
## alternative hypothesis: two.sided

write_xlsx(merged_data, "final_merged_toxoplasmosis_data.xlsx")

```