```
# load libraries

library("gmodels")
library("dplyr")

# Part 1: Does the term of the loan influence loan status? If so, how?
CrossTable(loans_excel$term, loans_excel$loan_status, chisq = TRUE, expected = TRUE,
sresid = TRUE, format = "SPSS")
# The p value of 0 is < 0.05 therefore The result is statistically significant.
# In Expected Values are more than 5
# The Std Residual is more than +- 2. Therefore we take the data with a grain of salt.
# The term of the loan influences loan status.

library("gmodels")
library("tidyr")

# Part 3: The news just ran a story that only 15% of homes are fully paid for in America, and that
another 10% have given up on paying it back, so the bank has "charged off" the loan.
# Does it seem likely that the data for this hands on came from the larger population of
America?

# Expected ratio of 15/10 fully paid for/charged off
# Data Wrangling:
loans_excel %>% group_by(loan_status) %>% summarise(count=n())
# Charged Off 3282, Current 502, Fully Paid 18173
observed = c(18173, 3282, 502)
expected = c(0.15, 0.10, 0.75)
chisq.test(x=observed, p=expected)
# p value is significant, therefore this dataset came from a larger population of America than the
news article.

# Part 2: How has the ability to own a home changed after 2009?

# specifying the path
path <- "/Users/music/Desktop/loans.csv"
# reading contents of csv file
loans <- read.csv(path)
# contents of the csv file
print (loans)

loans$DateR <- as.Date(paste(loans$Date), "%m/%d/%Y")

loans$DateR <- as.Date(paste(loans$Date), "%m/%d/%Y")
```

```
loans1 <- separate(loans, DateR, c("Year", "Month", "Day"), sep="-")

loans1$YearR <- NA
loans1$YearR[loans1$Year <= 2009] <- 0
loans1$YearR[loans1$Year > 2009] <- 1

loans1$RentvOwn <- NA
loans1$RentvOwn[loans1$home_ownership == "RENT"] <- 0
loans1$RentvOwn[loans1$home_ownership == "OWN"] <- 1

CrossTable(loans1$RentvOwn, loans1$YearR, fisher=TRUE, chisq = TRUE, mcnemar = TRUE,
expected = TRUE, sresid=TRUE, format="SPSS")

# The p value of 0 is < 0.05 therefore The result is statistically significant.
# In Expected Values are more than 5
# The Std Residual is less than +- 2.
# Housing market has changed for the worse with the std residual leaving little error.
```