

HW-3

Homework - 3 - STATS 506

Vasudha Rohatgi

Problem - 1

I have made extensive use of the pipe operator `%>%` throughout this exercise to chain a sequence of operations

```
library(haven)
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`

```
vix <- read_xpt("VIX_D.XPT")

## ncol(vix) --> 57

demo <- read_xpt("DEMO_D.XPT")
vision_data <- merge(vix, demo, by = "SEQN")
nrow(vision_data)
```

```
[1] 6980
```

```
## ncol(vision_data) --> 99
## summary.data.frame(vision_data)
```

```
vision_df<- data.frame(vision_data)
nrow(vision_df)
```

```
[1] 6980
```

```
ncol(vision_df)
```

```
[1] 99
```

```
str(vision_df)
```

```
'data.frame': 6980 obs. of 99 variables:
 $ SEQN : num 31129 31130 31131 31132 31133 ...
 $ VIQ110 : num 2 NA 2 2 2 2 NA 2 2 2 ...
 $ VIQ121 : num NA NA NA NA NA NA NA NA NA NA ...
 $ VIQ130 : num 2 NA 2 2 2 2 NA 2 2 2 ...
 $ VIQ141 : num NA NA NA NA NA NA NA NA NA NA ...
 $ VIQ150 : num NA NA NA 1 NA 1 NA NA NA NA ...
 $ VIQ160 : num NA NA NA 1 NA 1 NA NA NA NA ...
 $ VIQ170 : num NA NA NA 1 NA 1 NA NA NA NA ...
 $ VIXNC1 : num NA NA NA 5 NA 5 NA NA NA NA ...
 $ VIXNC2 : num NA NA NA 14 NA 15 NA NA NA NA ...
 $ VIQ180 : num 2 NA 2 2 2 2 NA 2 2 2 ...
 $ VIQ191 : num NA NA NA NA NA NA NA NA NA NA ...
 $ VIQ200 : num 2 NA 2 2 2 2 NA 2 2 2 ...
 $ VIQ211 : num NA NA NA NA NA NA NA NA NA NA ...
 $ VIQ220 : num 2 NA 1 1 1 1 NA 2 1 2 ...
 $ VIQ230 : num NA NA 1 1 1 1 NA NA 1 NA ...
 $ VIQ240 : num NA NA 1 1 2 1 NA NA 1 NA ...
 $ VIQ250 : num NA NA 1 1 NA 1 NA NA 1 NA ...
 $ VIXPRS : num NA NA -3 -5 NA -3.25 NA NA -4.25 NA ...
 $ VIXPRC : num NA NA 0.25 0.25 NA 3.25 NA NA 0 NA ...
 $ VIXPRA : num NA NA 51 160 NA 4 NA NA 0 NA ...
 $ VIXPLS : num NA NA -3.25 -5 NA -3.75 NA NA -6.25 NA ...
 $ VIXPLC : num NA NA 0.25 1.25 NA 3 NA NA 0.5 NA ...
 $ VIXPLA : num NA NA 148 178 NA 174 NA NA 81 NA ...
 $ VIXOCMT : num NA NA NA NA NA NA NA NA NA NA ...
 $ VIXORSM : num 1.75 NA -2.75 -4.5 -4.5 -3.5 NA -0.25 -4.75 -1 ...
 $ VIXORCM : num 0.25 NA 0.25 1.25 1 2.25 NA 0 0.75 0.75 ...
 $ VIXORAM : num 58 NA 29 4 62 180 NA 0 82 15 ...
 $ VIDORFM : num 9 NA 9 8.7 9 7.3 NA 9 8.7 9 ...
```

\$ VIXOLSM : num 1 NA -3.75 -4.75 -4.75 -4.5 NA -0.5 -6.75 -1 ...
 \$ VIXOLCM : num 0.75 NA 1 1 1.25 3.5 NA 0.75 1.25 0.5 ...
 \$ VIXOLAM : num 104 NA 114 174 117 174 NA 75 82 176 ...
 \$ VIDOLFM : num 8.3 NA 9 8.7 9 8.7 NA 9 9 9 ...
 \$ VIXKRM1 : num 8.17 NA 7.43 7.62 7.74 7.95 NA 7.93 8.11 7.99 ...
 \$ VIXKRD1 : num 41.2 NA 45.5 44.2 43.5 ...
 \$ VIXKRG1 : num 5 NA 12 156 179 99 NA 164 167 73 ...
 \$ VIXKRM2 : num 7.95 NA 7.21 7.55 7.53 7.62 NA 7.78 7.97 7.88 ...
 \$ VIXKRD2 : num 42.5 NA 46.8 44.8 44.8 ...
 \$ VIXKRG2 : num 95 NA 102 66 89 9 NA 74 77 163 ...
 \$ VIXKRMM : num 8.06 NA 7.32 7.59 7.64 7.79 NA 7.86 8.04 7.94 ...
 \$ VIXKRDM : num 41.8 NA 46 44.5 44.2 ...
 \$ VIXKRCD : num 1.25 NA 1.25 0.5 1.25 1.75 NA 1 0.75 0.5 ...
 \$ VIXKRCG : num 95 NA 102 66 89 9 NA 74 77 163 ...
 \$ VIXKLM1 : num 8.09 NA 7.44 7.65 7.74 8.03 NA 7.99 8.18 7.98 ...
 \$ VIXKLD1 : num 41.8 NA 45.2 44 43.5 ...
 \$ VIXKLG1 : num 19 NA 11 77 9 86 NA 48 7 104 ...
 \$ VIXKLM2 : num 7.93 NA 7.13 7.58 7.44 7.55 NA 7.83 7.89 7.91 ...
 \$ VIXKLD2 : num 42.5 NA 47.2 44.5 45.2 ...
 \$ VIXKLG2 : num 109 NA 101 167 99 176 NA 138 97 14 ...
 \$ VIXKLMM : num 8.01 NA 7.29 7.62 7.59 7.79 NA 7.91 8.04 7.95 ...
 \$ VIXKLDM : num 42.2 NA 46.2 44.2 44.5 ...
 \$ VIXKLCD : num 0.75 NA 2 0.5 1.75 2.75 NA 0.75 1.5 0.5 ...
 \$ VIXKLCG : num 109 NA 101 167 99 176 NA 138 97 14 ...
 \$ VIDRVA : num 20 NA 20 25 25 25 NA 40 20 25 ...
 \$ VIDLVA : num 20 NA 25 25 30 25 NA 20 20 30 ...
 \$ VIDROVA : num NA NA NA NA NA NA NA 25 NA NA ...
 \$ VIDLOVA : num NA NA NA NA 25 NA NA NA NA 25 ...
 \$ SDDSRVYR : num 4 4 4 4 4 4 4 4 4 4 ...
 \$ RIDSTATR : num 2 2 2 2 2 2 2 2 2 2 ...
 \$ RIDEXMON : num 2 2 2 2 2 2 1 2 2 2 ...
 \$ RIAGENDR : num 1 2 2 1 2 1 2 2 1 1 ...
 \$ RIDAGEYR : num 15 85 44 70 16 73 14 18 16 19 ...
 \$ RIDAGEMN : num 189 NA 535 842 193 882 178 217 195 229 ...
 \$ RIDAGEEX : num 190 NA 536 843 194 883 178 217 196 229 ...
 \$ RIDRETH1 : num 4 3 4 3 4 3 4 2 1 3 ...
 \$ DMQMILIT : num NA 2 2 1 NA 1 NA 2 NA 2 ...
 \$ DMDDBORN : num 1 1 1 1 1 1 1 1 1 1 ...
 \$ DMDCITZN : num 1 1 1 1 1 1 1 1 1 1 ...
 \$ DMDYRSUS : num NA NA NA NA NA NA NA NA NA NA ...
 \$ DMDEDUC3 : num 10 NA NA NA 9 NA 6 13 10 15 ...

```

$ DMDDEDUC2: num  NA 4 4 5 NA 3 NA NA NA NA ...
$ DMDSCHOL: num  1 NA NA NA 1 NA 1 3 2 2 ...
$ DMDMARTL: num  5 2 1 1 5 1 5 5 5 5 ...
$ DMDHHSIZ: num  6 1 4 2 3 2 7 3 4 3 ...
$ DMDFMSIZ: num  6 1 4 2 3 2 7 3 4 3 ...
$ INDHHINC: num 10 4 11 11 11 12 1 11 11 11 ...
$ INDFMINC: num 10 4 11 11 11 12 1 11 11 11 ...
$ INDFMPIR: num  2.71 1.99 4.65 5 5 NA 0.1 4.91 5 5 ...
$ RIDEXPRG: num  NA NA 2 NA 2 NA 2 2 NA NA ...
$ DMDHRGND: num  1 2 1 1 1 1 2 2 2 1 ...
$ DMDHRAGE: num 41 85 36 70 47 73 33 18 50 58 ...
$ DMDHRBRN: num  1 1 1 1 1 1 1 1 1 1 ...
$ DMDHREDU: num  4 4 5 5 2 3 1 3 5 3 ...
$ DMDHRMAR: num  1 2 1 1 1 1 NA 5 1 1 ...
$ DMDHSEDU: num  4 NA 4 5 4 3 NA NA 5 4 ...
$ SIALANG : num  1 1 1 1 1 1 1 1 1 1 ...
$ SIAPROXY: num  1 2 2 2 2 2 1 2 2 2 ...
$ SIAINTRP: num  2 2 2 2 2 2 2 2 2 2 ...
$ FIALANG : num  1 1 1 1 1 1 1 1 1 1 ...
$ FIAPROXY: num  2 2 2 2 2 2 2 2 2 2 ...
$ FIAINTRP: num  2 2 2 2 2 2 2 2 2 2 ...
$ MIALANG : num  1 NA 1 1 1 1 1 1 1 1 ...
$ MIAPROXY: num  2 NA 2 2 2 2 2 2 2 2 ...
$ MIAINTRP: num  2 NA 2 2 2 2 2 2 2 2 ...
$ AIALANG : num  1 NA 1 1 1 1 1 1 1 1 ...
$ WTINT2YR: num 5317 29961 26458 32962 5635 ...
$ WTMEC2YR: num 5587 34031 26771 35316 5921 ...
$ SDMVPSU : num  1 2 1 2 1 2 2 1 2 1 ...
$ SDMVSTRA: num  51 46 48 52 51 48 51 44 48 44 ...

```

```
## head(vision_df)
```

These are the actual column names, in which I have highlighted the ones I used for computation

| | | | | | |
|---------|--|----------|--|-----------|---|
| VIQ110 | SP has a severe eye infection? | VIXKRM1 | Right keratometry radius flat curve (mm) | RIDRETH1 | Race/Ethnicity |
| VIQ121 | Which eye(s) infected? | VIXKRD1 | Right keratometry power flat curve (D) | DMQMILIT | Veteran/Military Status |
| VIQ130 | Is SP wearing an eye patch? | VIXKRG1 | Right keratometry axis flat curve (deg) | DMDBORN | Country of Birth |
| VIQ141 | Which eye has eye patch? | VIXKRM2 | Right keratometry radius steep curve(mm) | DMDCITZN | Citizenship Status |
| VIQ150 | Glasses/contact lenses for near work? | VIXKRD2 | Right keratometry power steep curve (D) | DMDYRSUS | Length of time in US |
| VIQ160 | Glasses/contact lenses available? | VIXKRG2 | Right keratometry axis steep curve (deg) | DMDDEDUC3 | Education Level |
| VIQ170 | Glasses or contacts worn for near test? | VIXKRRM | Right keratometry radius, average (mm) | DMDDEDUC2 | Education Level |
| VIXNC1 | Line read on near card | VIXKRD | Right keratometry power, average (D) | DMDSCHOL | Now attending school? |
| VIXNC2 | Testing distance of near card | VIXKRCD | Right keratometry cylinder | DMDMARTL | Marital Status |
| VIQ180 | Eye surgery for near sightedness? | VIXKRCG | Right keratometry axis (deg) | DMDHHSIZ | Total number of people in the Household |
| VIQ191 | Which eye(s) surgery near sightedness? | VIXKLM1 | Left keratometry radius flat curve (mm) | DMDFMSIZ | Total number of people in the Family |
| VIQ200 | Eye surgery for cataracts? | VIXKLD1 | Left keratometry power flat curve (D) | INDHHINC | Annual Household Income |
| VIQ211 | Which eye(s) cataract surgery? | VIXKLG1 | Left keratometry axis flat curve (deg) | INDFMINC | Annual Family Income |
| VIQ220 | Glasses/contact lenses worn for distance | VIXKLM2 | Left keratometry radius steep curve (mm) | INDFMPIR | Family PIR |
| VIQ230 | Is distance correction available? | VIXKLD2 | Left keratometry power steep curve (D) | RIDEXPRG | Pregnancy Status at Exam |
| VIQ240 | Which type? Glasses or contacts? | VIXKLG2 | Left keratometry axis steep curve (deg) | DMDHRGND | HH Ref Person Gender |
| VIQ250 | Was prescription obtained? | VIXKLM | Left keratometry radius, average (mm) | DMDHRAGE | HH Ref Person Age |
| VIXPRS | Prescription, right sphere | VIXKLD | Left keratometry power, average (D) | DMDHRBRN | HH Ref Person Country of Birth |
| VIXPRC | Prescription, right cylinder | VIXKLCD | Left keratometry cylinder | DMDHREDU | HH Ref Person Education Level |
| VIXPRA | Prescription, right axis | VIXKLCG | Left keratometry axis (deg) | DMDHRMAR | HH Ref Person Marital Status |
| VIXPLS | Prescription, left sphere | VIDRVA | Right visual acuity, presenting | DMDHSEDU | HH Ref Person's Spouse Education Level |
| VIXPLC | Prescription, left cylinder | VIDLVA | Left visual acuity, presenting | SIALANG | Language of SP Interview |
| VIXPLA | Prescription, left axis | VIDROVA | Right visual acuity, w/ obj. refraction | SIAPROXY | Proxy used in SP Interview? |
| VIXOCMT | Visual Acuity Comments | VIDLOVA | Left visual acuity, w/ obj. refraction | SIAINTRP | Interpreter used in SP Interview? |
| VIXORSM | OR right sphere, median | SEQN | Respondent sequence number | FIALANG | Language of Family Interview |
| VIXORCM | OR right cylinder, median | SDDSRVYR | Data Release Number | FIAPROXY | Proxy used in Family Interview? |
| VIXORAM | OR right axis, median | RIDSTATR | Interview/Examination Status | FIAINTRP | Interpreter used in Family Interview? |
| VIDORFM | OR right confidence level reading | RIDEXMON | Six month time period | MIALANG | Language of MEC Interview |
| VIXOLSM | OR left sphere, median | RIAGENDR | Gender | MIAPROXY | Proxy used in MEC Interview? |
| VIXOLCM | OR left cylinder, median | RIDAGEYR | Age at Screening Adjudicated | MIAINTRP | Interpreter used in MEC Interview? |
| VIXOLAM | OR left axis, median | RIDAGEMN | Age in Months | AIALANG | Language of ACASI Interview |
| VIDOLFM | OR left confidence level reading | RIDAGEEX | Exam Age in Months | WTINT2YR | Full Sample 2 Year Interview Weight |
| | | | | WTMEC2YR | Full Sample 2 Year MEC Exam Weight |
| | | | | SDMVPSU | Masked Variance Pseudo |
| | | | | SDMVSTRA | Masked Variance Pseudo |

Create age brackets and display the data

```
## Filtering out NA valued rows for column VIQ220
```

```
vision_df <- vision_df %>%filter(!is.na(VIQ220))
```

```
## Filtering out NA values for some additional columns
```

```
vision_df <- vision_df %>%filter(!is.na(RIAGENDR))
```

```
vision_df <- vision_df %>%filter(!is.na(RIDAGEYR))
```

```
vision_df <- vision_df %>%filter(!is.na(INDFMPIR))
```

```
vision_df <- vision_df %>%filter(!is.na(RIDRETH1))
```

```
## VIQ220 isn't mapped as binary, so we'll convert
```

```
vision_df <- vision_df %>%
```

```
  mutate(VIQ220 = ifelse(VIQ220 == "1", 1, 0))
```

```
vision_df$VIQ220 <- as.numeric(vision_df$VIQ220)
```

```

vision_df <- vision_df %>% mutate(age_bracket = cut(RIDAGEYR, breaks = seq(0, 100, by
↵ = 10), right = FALSE, labels = paste0(seq(0, 90, by = 10), "-", seq(9, 99, by =
↵ 10))))

# Step 5: Calculate proportions of glasses/contact lenses wearers
proportion_table <- vision_df %>%
  group_by(age_bracket) %>%
  summarise(
    total_respondents = n(),
    wear_glasses_count = sum(VIQ220 == 1, na.rm = TRUE), # Assuming 'VIQ220' is
    ↵ binary (1 = yes)
    proportion_wear_glasses = wear_glasses_count / total_respondents
  )

# Display the proportion table
print(proportion_table)

```

```

# A tibble: 8 x 4
  age_bracket total_respondents wear_glasses_count proportion_wear_glasses
  <fct>          <int>          <int>          <dbl>
1 10-19          1988             647             0.325
2 20-29           894             299             0.334
3 30-39           733             267             0.364
4 40-49           750             276             0.368
5 50-59           583             322             0.552
6 60-69           601             375             0.624
7 70-79           415             280             0.675
8 80-89           285             189             0.663

```

Fit logistic regression models

```

# Model 1: Predicting glasses wear based on RIDAGEYR
model1 <- glm(VIQ220 ~ RIDAGEYR, data = vision_df, family = binomial)

# Model 2: Predicting glasses wear based on RIDAGEYR, race, and gender
model2 <- glm(VIQ220 ~ RIDAGEYR + RIDRETH1 + RIAGENDR, data = vision_df, family =
↵ binomial)

# Model 3: Predicting glasses wear based on RIDAGEYR, race, gender, and poverty
↵ income ratio
model3 <- glm(VIQ220 ~ RIDAGEYR + RIDRETH1 + RIAGENDR + INDFMPIR, data = vision_df,
↵ family = binomial)

```

```
summary(model1)
```

Call:

```
glm(formula = VIQ220 ~ RIDAGEYR, family = binomial, data = vision_df)
```

Coefficients:

| | Estimate | Std. Error | z value | Pr(> z) |
|-------------|----------|------------|---------|------------|
| (Intercept) | -1.22974 | 0.05464 | -22.51 | <2e-16 *** |
| RIDAGEYR | 0.02423 | 0.00124 | 19.55 | <2e-16 *** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 8521.3 on 6248 degrees of freedom
Residual deviance: 8117.7 on 6247 degrees of freedom
AIC: 8121.7

Number of Fisher Scoring iterations: 4

```
summary(model2)
```

Call:

```
glm(formula = VIQ220 ~ RIDAGEYR + RIDRETH1 + RIAGENDR, family = binomial,  
     data = vision_df)
```

Coefficients:

| | Estimate | Std. Error | z value | Pr(> z) |
|-------------|-----------|------------|---------|--------------|
| (Intercept) | -2.346217 | 0.122116 | -19.213 | < 2e-16 *** |
| RIDAGEYR | 0.024615 | 0.001258 | 19.569 | < 2e-16 *** |
| RIDRETH1 | 0.119747 | 0.023118 | 5.180 | 2.22e-07 *** |
| RIAGENDR | 0.498329 | 0.053727 | 9.275 | < 2e-16 *** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 8521.3 on 6248 degrees of freedom

Residual deviance: 8003.6 on 6245 degrees of freedom
AIC: 8011.6

Number of Fisher Scoring iterations: 4

```
summary(model3)
```

Call:

```
glm(formula = VIQ220 ~ RIDAGEYR + RIDRETH1 + RIAGENDR + INDFMPIR,  
     family = binomial, data = vision_df)
```

Coefficients:

| | Estimate | Std. Error | z value | Pr(> z) |
|-------------|-----------|------------|---------|--------------|
| (Intercept) | -2.634253 | 0.128466 | -20.505 | < 2e-16 *** |
| RIDAGEYR | 0.023775 | 0.001262 | 18.839 | < 2e-16 *** |
| RIDRETH1 | 0.093258 | 0.023559 | 3.958 | 7.54e-05 *** |
| RIAGENDR | 0.517449 | 0.054114 | 9.562 | < 2e-16 *** |
| INDFMPIR | 0.142353 | 0.017005 | 8.371 | < 2e-16 *** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 8521.3 on 6248 degrees of freedom
Residual deviance: 7933.0 on 6244 degrees of freedom
AIC: 7943

Number of Fisher Scoring iterations: 4

Now we can -

- **Convert coefficients to odds ratios** using the exponential function.
- **Create a summary table** for the odds ratios along with additional information about the models.

```
library(dplyr)
library(knitr)

# Create a function to extract odds ratios and prepare results
get_odds_ratios <- function(model) {
  odds_ratios <- exp(coef(model)) # Exponential of coefficients for odds ratios
```



```

odds_ratios_df <- data.frame(
  Term = names(odds_ratios),
  Odds_Ratio = odds_ratios
)
return(odds_ratios_df)
}

# Prepare a summary table for model results
model_results <- data.frame(
  Model = c("Age-Based", "Age,Race & Gender", "Age,Race,Gender & Poverty-Line"),
  Sample_Size = c(nrow(vision_df), nrow(vision_df), nrow(vision_df)),
  AIC = c(AIC(model1), AIC(model2), AIC(model3))
)

# Extract odds ratios for each model
odds_ratios_model1 <- get_odds_ratios(model1)
odds_ratios_model2 <- get_odds_ratios(model2)
odds_ratios_model3 <- get_odds_ratios(model3)

# Combine odds ratios into a single data frame with model names
odds_ratios_combined <- bind_rows(
  odds_ratios_model1 %>% mutate(Model = "Age-Based"),
  odds_ratios_model2 %>% mutate(Model = "Age,Race & Gender"),
  odds_ratios_model3 %>% mutate(Model = "Age,Race,Gender & Poverty-Line"))

# Combine model results with odds ratios for a final summary
final_summary <- model_results %>%
  left_join(odds_ratios_combined, by = c("Model" = "Model")) %>%
  select(Model, Sample_Size, AIC, Term, Odds_Ratio)

# Print the final summary in a readable format
kable(final_summary, caption = "Logistic Regression Models and Odds Ratios", digits =
  ↵ 3)

```

Table 1: Logistic Regression Models and Odds Ratios

| Model | Sample_Size | AIC | Term | Odds_Ratio |
|-------------------|-------------|----------|-------------|------------|
| Age-Based | 6249 | 8121.692 | (Intercept) | 0.292 |
| Age-Based | 6249 | 8121.692 | RIDAGEYR | 1.025 |
| Age,Race & Gender | 6249 | 8011.600 | (Intercept) | 0.096 |

| Model | Sample_Size | AIC | Term | Odds_Ratio |
|--------------------------------|-------------|----------|-------------|------------|
| Age,Race & Gender | 6249 | 8011.600 | RIDAGEYR | 1.025 |
| Age,Race & Gender | 6249 | 8011.600 | RIDRETH1 | 1.127 |
| Age,Race & Gender | 6249 | 8011.600 | RIAGENDR | 1.646 |
| Age,Race,Gender & Poverty-Line | 6249 | 7943.021 | (Intercept) | 0.072 |
| Age,Race,Gender & Poverty-Line | 6249 | 7943.021 | RIDAGEYR | 1.024 |
| Age,Race,Gender & Poverty-Line | 6249 | 7943.021 | RIDRETH1 | 1.098 |
| Age,Race,Gender & Poverty-Line | 6249 | 7943.021 | RIAGENDR | 1.678 |
| Age,Race,Gender & Poverty-Line | 6249 | 7943.021 | INDFMPIR | 1.153 |

```
# To ensure that RIAGENDR is a factor with two levels
vision_df$Gender <- as.factor(vision_df$RIAGENDR)
levels(vision_df$RIAGENDR) # Check the levels to make sure it's correct
```

NULL

```
# Logistic regression for RIAGENDR difference
gender_model <- glm(VIQ220 ~ RIAGENDR, data = vision_df, family = binomial)
gender_summary <- summary(gender_model)

# Display the logistic regression results
print(gender_summary)
```

Call:

```
glm(formula = VIQ220 ~ RIAGENDR, family = binomial, data = vision_df)
```

Coefficients:

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.95566    0.08286 -11.533  <2e-16 ***
RIAGENDR      0.42961    0.05155   8.334  <2e-16 ***
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 8521.3  on 6248  degrees of freedom
Residual deviance: 8451.4  on 6247  degrees of freedom
AIC: 8455.4
```

Number of Fisher Scoring iterations: 4

```
# Proportion test: summarize gender-based glasses usage
gender_proportions <- vision_df %>%
  group_by(RIAGENDR) %>%
  summarise(
    total_respondents = n(),
    wear_glasses_count = sum(VIQ220 == 1, na.rm = TRUE)
  )

# Proportion test between genders
prop_test_result <- prop.test(gender_proportions$wear_glasses_count,
  ↪ gender_proportions$total_respondents)

# Display the results of the proportion test
print(prop_test_result)
```

2-sample test for equality of proportions with continuity correction

```
data:  gender_proportions$wear_glasses_count out of gender_proportions$total_respondents
X-squared = 69.313, df = 1, p-value < 2.2e-16
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.12915275 -0.07978616
sample estimates:
   prop 1    prop 2 
0.3714379 0.4759074
```

Problem-2

Here, I will first establish connection to the SQL server so that I can run SQL queries

```
library(DBI)
library(RSQLite)

db_file <- "sakila_master.db"
conn <- dbConnect(SQLite(), dbname = db_file)
```

```
tables <- dbListTables(conn)
print(tables)
```

| | | |
|----------------------|--------------------------|------------------|
| [1] "actor" | "address" | "category" |
| [4] "city" | "country" | "customer" |
| [7] "customer_list" | "film" | "film_actor" |
| [10] "film_category" | "film_list" | "film_text" |
| [13] "inventory" | "language" | "payment" |
| [16] "rental" | "sales_by_film_category" | "sales_by_store" |
| [19] "staff" | "staff_list" | "store" |

```
SELECT release_year, COUNT(*) AS movie_count
FROM film
WHERE release_year = (SELECT MIN(release_year) FROM film)
GROUP BY release_year;
```

Table 2: 1 records

| release__year | movie__count |
|---------------|--------------|
| 2006 | 1000 |

Now that test run is done, we can run all the queries sequentially and through R:

I have used one executable cell to run all the queries in order to avoid SQL database connection errors

```
## few redundant calls but I added them because code was running into errors
```

```
library(RSQLite)
library(dplyr)

db_file <- "sakila_master.db"

con <- dbConnect(SQLite(), dbname = db_file)

if (is.null(con)) {
  stop("Failed to connect to the database.")
}
```

```
## Oldest Movie Data Query
```

```
query <- "
SELECT release_year, COUNT(*) AS movie_count
```

```

FROM film
WHERE release_year = (SELECT MIN(release_year) FROM film)
GROUP BY release_year;"

## Genre Data Query

genre_query <- "
SELECT category.name AS genre, COUNT(*) AS movie_count
FROM film
JOIN film_category ON film.film_id = film_category.film_id
JOIN category ON film_category.category_id = category.category_id
GROUP BY category.name;"

## Query to retrieve customer counts by country

country_customer_query <- "
  SELECT country.country AS country_name, COUNT(customer.customer_id) AS
  ↵ customer_count
  FROM country
  JOIN city ON country.country_id = city.country_id
  JOIN address ON city.city_id = address.city_id
  JOIN customer ON address.address_id = customer.address_id
  GROUP BY country.country;
"

country_customer_df <- dbGetQuery(con, country_customer_query)

# Filter countries with exactly 13 customers
countries_with_13_customers <- subset(country_customer_df, customer_count == 13)
print(countries_with_13_customers)

```

| | country_name | customer_count |
|----|--------------|----------------|
| 6 | Argentina | 13 |
| 68 | Nigeria | 13 |

```

oldest_movie_data <- dbGetQuery(con, query)
genre_df <- dbGetQuery(con, genre_query)

print(oldest_movie_data)

```

```

release_year movie_count
1           2006           1000

```

```

least_common_genre <- genre_df[which.min(genre_df$movie_count), ]
print(least_common_genre)

```

```

genre movie_count
12 Music           51

```

```

# Disconnect from the database
dbDisconnect(con)

```

Problem - 3

```
install.packages("benford.analysis") install.packages("BenfordTests")
```

Using a similar approach as in previous problems, I will re-introduce libraries into the global environment and load the downloaded us-500 csv file into a dataframe

```

library(ggplot2)
library(benford.analysis)
library(dplyr)

data_500 <- read.csv("us-500.csv")
print(data_500[1:5,])

```

| | first_name | last_name | company_name | address | city |
|---|------------|-----------|------------------------|----------------------|-------------|
| 1 | James | Butt | Benton, John B Jr | 6649 N Blue Gum St | New Orleans |
| 2 | Josephine | Darakjy | Chanay, Jeffrey A Esq | 4 B Blue Ridge Blvd | Brighton |
| 3 | Art | Venere | Chemel, James L Cpa | 8 W Cerritos Ave #54 | Bridgeport |
| 4 | Lenna | Paprocki | Feltz Printing Service | 639 Main St | Anchorage |
| 5 | Donette | Foller | Printing Dimensions | 34 Center St | Hamilton |

| | county | state | zip | phone1 | phone2 |
|---|------------|-------|-------|--------------|--------------|
| 1 | Orleans | LA | 70116 | 504-621-8927 | 504-845-1427 |
| 2 | Livingston | MI | 48116 | 810-292-9388 | 810-374-9840 |
| 3 | Gloucester | NJ | 8014 | 856-636-8749 | 856-264-4130 |
| 4 | Anchorage | AK | 99501 | 907-385-4412 | 907-921-2010 |
| 5 | Butler | OH | 45011 | 513-570-1893 | 513-549-4561 |

| | email | web |
|---|-------------------------------|----------------------------------|
| 1 | jbutt@gmail.com | http://www.bentonjohnbjr.com |
| 2 | josephine_darakjy@darakjy.org | http://www.chanayjeffreyaesq.com |

```

3          art@venere.org      http://www.chemeljameslcpa.com
4      lpaprocki@hotmail.com http://www.feltzprintingservice.com
5      donette.foller@cox.net  http://www.printingdimensions.com

```

1. Proportion of email addresses with TLD “.com”

```

com_proportion <- data_500 %>%
  mutate(TLD = sub(".*@\.\.", "", email)) %>% # Extract TLD from email
  summarise(Proportion_Com = mean(TLD == "com", na.rm = TRUE))

print(paste("Proportion of emails with .com TLD:", com_proportion$Proportion_Com))

```

```
[1] "Proportion of emails with .com TLD: 0.732"
```

2. Proportion of email addresses with non-alphanumeric characters

```

non_alphanumeric_proportion <- data_500 %>%
  mutate(Non_Alphanumeric = grepl("[^a-zA-Z0-9@.]", email)) %>%
  summarise(Proportion_Non_Alpha = mean(Non_Alphanumeric, na.rm = TRUE))

print(paste("Proportion of emails with non-alphanumeric characters:",
  ↪ non_alphanumeric_proportion$Proportion_Non_Alpha))

```

```
[1] "Proportion of emails with non-alphanumeric characters: 0.248"
```

3. Top 5 most common area codes

We can compute this for column phone1 and phone2 separately

```

top_area_codes <- data_500 %>%
  mutate(Area_Code = substr(phone1, 1, 3)) %>%
  group_by(Area_Code) %>%
  summarise(Count = n(), .groups = "drop") %>%
  arrange(desc(Count)) %>%
  head(5)

print("Top 5 records with area codes:")

```

```
[1] "Top 5 records with area codes:"
```

```
print(top_area_codes)
```

```

# A tibble: 5 x 2
  Area_Code Count

```

| | <chr> | <int> |
|---|-------|-------|
| 1 | 973 | 18 |
| 2 | 212 | 14 |
| 3 | 215 | 14 |
| 4 | 410 | 14 |
| 5 | 201 | 12 |

```
top_area_codes <- data_500 %>%
  mutate(Area_Code = substr(phone2, 1, 3)) %>%
  count(Area_Code) %>%
  arrange(desc(n)) %>%
  head(5)
```

```
print("Top 5 most common area codes:")
```

```
[1] "Top 5 most common area codes:"
```

```
print(top_area_codes)
```

| | Area_Code | n |
|---|-----------|----|
| 1 | 973 | 18 |
| 2 | 212 | 14 |
| 3 | 215 | 14 |
| 4 | 410 | 14 |
| 5 | 201 | 12 |

4. Histogram of the log of apartment numbers

```
print(head(data_500$address, 500))
```

| | |
|----------------------------|-------------------------|
| [1] "6649 N Blue Gum St" | "4 B Blue Ridge Blvd" |
| [3] "8 W Cerritos Ave #54" | "639 Main St" |
| [5] "34 Center St" | "3 Mcauley Dr" |
| [7] "7 Eads St" | "7 W Jackson Blvd" |
| [9] "5 Boston Ave #88" | "228 Runamuck Pl #2808" |
| [11] "2371 Jerrold Ave" | "37275 St Rt 17m M" |
| [13] "25 E 75th St #69" | "98 Connecticut Ave Nw" |
| [15] "56 E Morehead St" | "73 State Road 434 E" |
| [17] "69734 E Carrillo St" | "322 New Horizon Blvd" |
| [19] "1 State Route 27" | "394 Manchester Blvd" |
| [21] "6 S 33rd St" | "6 Greenleaf Ave" |
| [23] "618 W Yakima Ave" | "74 S Westgate St" |
| [25] "3273 State St" | "1 Central Ave" |

| | |
|-----------------------------------|---------------------------|
| [27] "86 Nw 66th St #8673" | "2 Cedar Ave #84" |
| [29] "90991 Thorburn Ave" | "386 9th Ave N" |
| [31] "74874 Atlantic Ave" | "366 South Dr" |
| [33] "45 E Liberty St" | "4 Ralph Ct" |
| [35] "2742 Distribution Way" | "426 Wolf St" |
| [37] "128 Bransten Rd" | "17 Morena Blvd" |
| [39] "775 W 17th St" | "6980 Dorsett Rd" |
| [41] "2881 Lewis Rd" | "7219 Woodfield Rd" |
| [43] "1048 Main St" | "678 3rd Ave" |
| [45] "20 S Babcock St" | "2 Lighthouse Ave" |
| [47] "38938 Park Blvd" | "5 Tomahawk Dr" |
| [49] "762 S Main St" | "209 Decker Dr" |
| [51] "4486 W O St #1" | "39 S 7th St" |
| [53] "98839 Hawthorne Blvd #6101" | "71 San Mateo Ave" |
| [55] "76 Brooks St #9" | "4545 Courthouse Rd" |
| [57] "14288 Foster Ave #4121" | "4 Otis St" |
| [59] "65895 S 16th St" | "14302 Pennsylvania Ave" |
| [61] "201 Hawk Ct" | "53075 Sw 152nd Ter #615" |
| [63] "59 N Groesbeck Hwy" | "2664 Lewis Rd" |
| [65] "59 Shady Ln #53" | "3305 Nabell Ave #679" |
| [67] "18 Fountain St" | "7 W 32nd St" |
| [69] "2853 S Central Expy" | "74 W College St" |
| [71] "701 S Harrison Rd" | "1088 Pinehurst St" |
| [73] "30 W 80th St #1995" | "20932 Hedley St" |
| [75] "2737 Pistorio Rd #9230" | "74989 Brandon St" |
| [77] "6 Kains Ave" | "47565 W Grand Ave" |
| [79] "4284 Dorigo Ln" | "6794 Lake Dr E" |
| [81] "31 Douglas Blvd #950" | "44 W 4th St" |
| [83] "11279 Loytan St" | "69 Marquette Ave" |
| [85] "70 W Main St" | "461 Prospect Pl #316" |
| [87] "47154 Whipple Ave Nw" | "37 Alabama Ave" |
| [89] "3777 E Richmond St #900" | "3 Fort Worth Ave" |
| [91] "4800 Black Horse Pike" | "83649 W Belmont Ave" |
| [93] "840 15th Ave" | "1747 Calle Amanecer #2" |
| [95] "99385 Charity St #840" | "68556 Central Hwy" |
| [97] "55 Riverside Ave" | "7140 University Ave" |
| [99] "64 5th Ave #1153" | "3 Secor Rd" |
| [101] "4 Webbs Chapel Rd" | "524 Louisiana Ave Nw" |
| [103] "185 Blackstone Bldge" | "170 Wyoming Ave" |
| [105] "4 10th St W" | "7 W Pinhook Rd" |
| [107] "1 Commerce Way" | "64 Lakeview Ave" |

| | | |
|-------|------------------------------|----------------------------|
| [109] | "3 Aspen St" | "32860 Sierra Rd" |
| [111] | "555 Main St" | "2 Se 3rd Ave" |
| [113] | "2239 Shawnee Mission Pky" | "2726 Charcot Ave" |
| [115] | "5161 Dorsett Rd" | "55892 Jacksonville Rd" |
| [117] | "5 N Cleveland Massillon Rd" | "7 Benton Dr" |
| [119] | "9390 S Howell Ave" | "8 County Center Dr #647" |
| [121] | "4646 Kaahumanu St" | "2 Monroe St" |
| [123] | "52777 Leaders Heights Rd" | "72868 Blackington Ave" |
| [125] | "9 Norristown Rd" | "83 County Road 437 #8581" |
| [127] | "1 N Harlem Ave #9" | "90131 J St" |
| [129] | "8597 W National Ave" | "6 Gilson St" |
| [131] | "65 W Maple Ave" | "866 34th Ave" |
| [133] | "798 Lund Farm Way" | "9387 Charcot Ave" |
| [135] | "30553 Washington Rd" | "481 W Lemon St" |
| [137] | "4 Warehouse Point Rd #7" | "4940 Pulaski Park Dr" |
| [139] | "627 Walford Ave" | "137 Pioneer Way" |
| [141] | "61 13 Stoneridge #835" | "2409 Alabama Rd" |
| [143] | "8927 Vandever Ave" | "134 Lewis Rd" |
| [145] | "9 N College Ave #3" | "60480 Old Us Highway 51" |
| [147] | "4 Bloomfield Ave" | "429 Tiger Ln" |
| [149] | "54169 N Main St" | "92 Main St" |
| [151] | "72 Mannix Dr" | "12270 Caton Center Dr" |
| [153] | "749 W 18th St #45" | "8 Industry Ln" |
| [155] | "1 Huntwood Ave" | "55262 N French Rd" |
| [157] | "422 E 21st St" | "501 N 19th Ave" |
| [159] | "455 N Main Ave" | "1844 Southern Blvd" |
| [161] | "2023 Greg St" | "63381 Jenks Ave" |
| [163] | "6651 Municipal Rd" | "81 Norris Ave #525" |
| [165] | "6916 W Main St" | "9635 S Main St" |
| [167] | "17 Us Highway 111" | "992 Civic Center Dr" |
| [169] | "303 N Radcliffe St" | "73 Saint Ann St #86" |
| [171] | "44 58th St" | "9745 W Main St" |
| [173] | "84 Bloomfield Ave" | "287 Youngstown Warren Rd" |
| [175] | "6 Van Buren St" | "229 N Forty Driv" |
| [177] | "2887 Knowlton St #5435" | "523 Marquette Ave" |
| [179] | "3717 Hamann Industrial Pky" | "3 State Route 35 S" |
| [181] | "82 N Highway 67" | "9 Murfreesboro Rd" |
| [183] | "6 S Broadway St" | "6 Harry L Dr #6327" |
| [185] | "47939 Porter Ave" | "9 Wales Rd Ne #914" |
| [187] | "195 13n N" | "99 Tank Farm Rd" |
| [189] | "4671 Alemany Blvd" | "98 University Dr" |

| | |
|-----------------------------------|------------------------------|
| [191] "50 E Wacker Dr" | "70 Euclid Ave #722" |
| [193] "326 E Main St #6496" | "406 Main St" |
| [195] "3 Elmwood Dr" | "9 Church St" |
| [197] "9939 N 14th St" | "5384 Southwyck Blvd" |
| [199] "97 Airport Loop Dr" | "37855 Nolan Rd" |
| [201] "4252 N Washington Ave #9" | "42754 S Ash Ave" |
| [203] "703 Beville Rd" | "5 Harrison Rd" |
| [205] "73 Southern Blvd" | "189 Village Park Rd" |
| [207] "6 Middlegate Rd #106" | "1128 Delaware St" |
| [209] "577 Parade St" | "70 Mechanic St" |
| [211] "4379 Highway 116" | "55 Hawthorne Blvd" |
| [213] "7116 Western Ave" | "2026 N Plankinton Ave #3" |
| [215] "99586 Main St" | "8739 Hudson St" |
| [217] "383 Gunderman Rd #197" | "4441 Point Term Mkt" |
| [219] "2972 Lafayette Ave" | "2140 Diamond Blvd" |
| [221] "93 Redmond Rd #492" | "3989 Portage Tr" |
| [223] "1 Midway Rd" | "77132 Coon Rapids Blvd Nw" |
| [225] "755 Harbor Way" | "87 Sierra Rd" |
| [227] "7667 S Hulen St #42" | "75684 S Withlapopka Dr #32" |
| [229] "5 Elmwood Park Blvd" | "23 Palo Alto Sq" |
| [231] "38062 E Main St" | "3958 S Dupont Hwy #7" |
| [233] "560 Civic Center Dr" | "3270 Dequindre Rd" |
| [235] "1 Garfield Ave #7" | "9122 Carpenter Ave" |
| [237] "48 Lenox St" | "5 Little River Tpke" |
| [239] "3 N Groesbeck Hwy" | "37 N Elm St #916" |
| [241] "433 Westminster Blvd #590" | "66697 Park Pl #3224" |
| [243] "96263 Greenwood Pl" | "8 Mcarthur Ln" |
| [245] "8 Fair Lawn Ave" | "9 N 14th St" |
| [247] "9 Vanowen St" | "18 Waterloo Geneva Rd" |
| [249] "506 S Hacienda Dr" | "3732 Sherman Ave" |
| [251] "25657 Live Oak St" | "4923 Carey Ave" |
| [253] "3196 S Rider Trl" | "3 Railway Ave #75" |
| [255] "87393 E Highland Rd" | "67 E Chestnut Hill Rd" |
| [257] "33 Lewis Rd #46" | "8100 Jacksonville Rd #7" |
| [259] "7 W Wabansia Ave #227" | "25 Minters Chapel Rd #9" |
| [261] "6882 Torresdale Ave" | "985 E 6th Ave" |
| [263] "7 West Ave #1" | "26659 N 13th St" |
| [265] "669 Packerland Dr #1438" | "759 Eldora St" |
| [267] "5 S Colorado Blvd #449" | "944 Gaither Dr" |
| [269] "66552 Malone Rd" | "77 Massillon Rd #822" |
| [271] "25346 New Rd" | "60 Fillmore Ave" |

| | | |
|-------|----------------------------|-----------------------------------|
| [273] | "57 Haven Ave #90" | "6538 E Pomona St #60" |
| [275] | "6535 Joyce St" | "78112 Morris Ave" |
| [277] | "96950 Hidden Ln" | "3718 S Main St" |
| [279] | "9677 Commerce Dr" | "5 Green Pond Rd #4" |
| [281] | "636 Commerce Dr #42" | "42744 Hamann Industrial Pky #82" |
| [283] | "1950 5th Ave" | "61304 N French Rd" |
| [285] | "87 Imperial Ct #79" | "94 W Dodge Rd" |
| [287] | "4 58th St #3519" | "5221 Bear Valley Rd" |
| [289] | "9648 S Main" | "7 S San Marcos Rd" |
| [291] | "812 S Haven St" | "3882 W Congress St #799" |
| [293] | "4 E Colonial Dr" | "45 2nd Ave #9759" |
| [295] | "57254 Brickell Ave #372" | "8977 Connecticut Ave Nw #3" |
| [297] | "9 Waydell St" | "43 Huey P Long Ave" |
| [299] | "7563 Cornwall Rd #4462" | "22 Bridle Ln" |
| [301] | "70099 E North Ave" | "3211 E Northeast Loop" |
| [303] | "26 Montgomery St" | "13252 Lighthouse Ave" |
| [305] | "206 Main St #2804" | "96541 W Central Blvd" |
| [307] | "34 Saint George Ave #2" | "47857 Coney Island Ave" |
| [309] | "8573 Lincoln Blvd" | "596 Santa Maria Ave #7913" |
| [311] | "3829 Ventura Blvd" | "13 S Hacienda Dr" |
| [313] | "40 9th Ave Sw #91" | "2845 Boulder Crescent St" |
| [315] | "33 State St" | "2 S 15th St" |
| [317] | "4 Kohler Memorial Dr" | "1 Rancho Del Mar Shopping C" |
| [319] | "3943 N Highland Ave" | "5 Williams St" |
| [321] | "60 Old Dover Rd" | "8 Sheridan Rd" |
| [323] | "85092 Southern Blvd" | "64 Newman Springs Rd E" |
| [325] | "48 Stratford Ave" | "80 Pittsford Victor Rd #9" |
| [327] | "87163 N Main Ave" | "393 Lafayette Ave" |
| [329] | "99 5th Ave #33" | "49 N Mays St" |
| [331] | "993 Washington Ave" | "88 15th Ave Ne" |
| [333] | "3381 E 40th Ave" | "201 Ridgewood Rd" |
| [335] | "39 Moccasin Dr" | "4 Carroll St" |
| [337] | "9581 E Arapahoe Rd" | "33 N Michigan Ave" |
| [339] | "2 S Biscayne Blvd" | "8 Us Highway 22" |
| [341] | "7422 Martin Ave #8" | "94 Chase Rd" |
| [343] | "8139 I Hwy 10 #92" | "5 Cabot Rd" |
| [345] | "3387 Ryan Dr" | "3125 Packer Ave #9851" |
| [347] | "347 Chestnut St" | "8116 Mount Vernon Ave" |
| [349] | "8772 Old County Rd #5410" | "868 State St #38" |
| [351] | "772 W River Dr" | "73 W Barstow Ave" |
| [353] | "61047 Mayfield Ave" | "2139 Santa Rosa Ave" |

| | |
|--------------------------------|----------------------------|
| [355] "598 43rd St" | "70295 Pioneer Ct" |
| [357] "92899 Kalakaua Ave" | "395 S 6th St #2" |
| [359] "9506 Edgemore Ave" | "72119 S Walker Ave #63" |
| [361] "369 Latham St #500" | "3158 Runamuck Pl" |
| [363] "9 Plainsboro Rd #598" | "8728 S Broad St" |
| [365] "2215 Prosperity Dr" | "1 S Pine St" |
| [367] "187 Market St" | "94290 S Buchanan St" |
| [369] "7061 N 2nd St" | "10759 Main St" |
| [371] "97 E 3rd St #9" | "82 Winsor St #54" |
| [373] "41 Steel Ct" | "49440 Dearborn St" |
| [375] "7 S Beverly Dr" | "919 Wall Blvd" |
| [377] "89 20th St E #779" | "721 Interstate 45 S" |
| [379] "3 Lawton St" | "38 Pleasant Hill Rd" |
| [381] "45 E Acacia Ct" | "63728 Poway Rd #1" |
| [383] "77 222 Dr" | "53 W Carey St" |
| [385] "617 Nw 36th Ave" | "539 Coldwater Canyon Ave" |
| [387] "735 Crawford Dr" | "910 Rahway Ave" |
| [389] "7 Tarrytown Rd" | "35433 Blake St #588" |
| [391] "29 Cherry St #7073" | "810 N La Brea Ave" |
| [393] "987 Main St" | "36 Enterprise St Se" |
| [395] "8429 Miller Rd" | "5 W 7th St" |
| [397] "2 Flynn Rd" | "2094 Ne 36th Ave" |
| [399] "649 Tulane Ave" | "2094 Montour Blvd" |
| [401] "393 Hammond Dr" | "8590 Lake Lizzie Dr" |
| [403] "87895 Concord Rd" | "46314 Route 130" |
| [405] "4 Cowesett Ave" | "95 Main Ave #2" |
| [407] "28 S 7th St #2824" | "79 S Howell Ave" |
| [409] "36 Lancaster Dr Se" | "2759 Livingston Ave" |
| [411] "17 Jersey Ave" | "2 W Grand Ave" |
| [413] "18 Coronado Ave #563" | "72 Beechwood Ter" |
| [415] "92 Broadway" | "39 Franklin Ave" |
| [417] "4 Iwaena St" | "32820 Corkwood Rd" |
| [419] "34 Raritan Center Pky" | "6201 S Nevada Ave" |
| [421] "78 Maryland Dr #146" | "76598 Rd I 95 #1" |
| [423] "1610 14th St Nw" | "86350 Roszel Rd" |
| [425] "1644 Clove Rd" | "9 W Central Ave" |
| [427] "27846 Lafayette Ave" | "10276 Brooks St" |
| [429] "1 Century Park E" | "9 State Highway 57 #22" |
| [431] "4 S Washington Ave" | "25 Se 176th Pl" |
| [433] "105 Richmond Valley Rd" | "22 Spruce St #595" |
| [435] "2 W Beverly Blvd" | "72 Southern Blvd" |

| | | |
|-------|---------------------------|---------------------------|
| [437] | "1 Washington St" | "90177 N 55th Ave" |
| [439] | "9 Tower Ave" | "278 Bayview Ave" |
| [441] | "80312 W 32nd St" | "82 Us Highway 46" |
| [443] | "4 Stovall St #72" | "19 Amboy Ave" |
| [445] | "63 Smith Ln #8343" | "11360 S Halsted St" |
| [447] | "26849 Jefferson Hwy" | "2500 Pringle Rd Se #508" |
| [449] | "65 Mountain View Dr" | "1 N San Saba" |
| [451] | "51120 State Route 18" | "1482 College Ave" |
| [453] | "4119 Metropolitan Dr" | "2167 Sierra Rd" |
| [455] | "6 Sunrise Ave" | "55713 Lake City Hwy" |
| [457] | "75698 N Fiesta Blvd" | "88 Sw 28th Ter" |
| [459] | "7 Flowers Rd #403" | "4 Nw 12th St #3849" |
| [461] | "2 A Kelley Dr" | "88827 Frankford Ave" |
| [463] | "2 W Scyene Rd #3" | "62260 Park Stre" |
| [465] | "3424 29th St Se" | "35 E Main St #43" |
| [467] | "7163 W Clark Rd" | "21575 S Apple Creek Rd" |
| [469] | "747 Leonis Blvd" | "13 Gunnison St" |
| [471] | "18 3rd Ave" | "62 W Austin St" |
| [473] | "177 S Rider Trl #52" | "2 W Mount Royal Ave" |
| [475] | "1953 Telegraph Rd" | "63517 Dupont St" |
| [477] | "5 E Truman Rd" | "251 Park Ave #979" |
| [479] | "43496 Commercial Dr #29" | "2184 Worth St" |
| [481] | "50126 N Plankinton Ave" | "38773 Gravois Ave" |
| [483] | "16452 Greenwich St" | "40 Cambridge Ave" |
| [485] | "20113 4th Ave E" | "6 Ridgewood Center Dr" |
| [487] | "469 Outwater Ln" | "62 Monroe St" |
| [489] | "3338 A Lockport Pl #6" | "9 Hwy" |
| [491] | "8284 Hart St" | "5 Washington St #1" |
| [493] | "8 S Haven St" | "9 Front St" |
| [495] | "1933 Packer Ave #2" | "67 Rv Cent" |
| [497] | "2 Sw Nyberg Rd" | "89992 E 15th St" |
| [499] | "61556 W 20th Ave" | "63 E Aurora Dr" |

Here I have tried different methods of pattern matching to determine which is yielding maximum apartment numbers

```
library(stringr)

addresses <- c(data_500$address)

pattern <-
  ↪ "^[\\d{1,5}]\\s+([NSEW]?\\s*)?([\\w\\s\\.]+)\\s+(St|Ave|Blvd|Rd|Ct|Ln|Dr|Pl|Ter|Pky|Cir|Way
```

```

pattern2<- "(Apt\\s*\\d+|Apartment\\s*\\d+|#\\d+|Apt\\s*[A-Za-z]\\d+|\\d+[A-Za-z])"

# Extracting apartment numbers

apartment_numbers <- sapply(addresses, function(address) {
  match <- regmatches(address, regex(pattern2, address))
  # Extract the second or third capturing group, prioritizing the one that exists
  if (length(match[[1]]) > 1) {
    return(ifelse(!is.na(match[[1]][3]), match[[1]][3], match[[1]][2]))
  }
  return(NA) # Return NA if no match
})

typeof(apartment_numbers)

```

```
[1] "character"
```

```

## determining apartment number using REGEX
data_apartments <- data_500 %>%
  mutate(Apartment_Number = str_extract(address,
    ↪ "^(\\d{1,5})\\s+([NSEW]?\\s*)?(\\w\\s\\.)+\\s+(St|Ave|Blvd|Rd|Ct|Ln|Dr|Pl|Ter|Pky|Cir|V
    ↪ %>%
  filter(!is.na(Apartment_Number))

nrow(data_apartments)

```

```
[1] 454
```

```

data_apartments <- data_500 %>%
  mutate(Apartment_Number = str_extract(address,
    ↪ "(Apt\\s*\\d+|Apartment\\s*\\d+|#\\d+|Apt\\s*[A-Za-z]\\d+|\\d+[A-Za-z])")) %>%
  mutate(Apartment_Number = str_extract(Apartment_Number, "\\d+")) %>%
  filter(!is.na(Apartment_Number))

```

```
print(data_apartments)
```

| | first_name | last_name | company_name |
|---|------------|-----------|---------------------------|
| 1 | Art | Venere | Chemel, James L Cpa |
| 2 | Sage | Wieser | Truhlar And Truhlar Attys |
| 3 | Kris | Marrier | King, Christopher A Esq |

| | | | |
|----|-----------|------------|--------------------------------|
| 4 | Abel | Maclead | Rangoni Of Florence |
| 5 | Kiley | Caldarera | Feiner Bros |
| 6 | Bette | Nicka | Sport En Art |
| 7 | Chanel | Caudy | Professional Image Inc |
| 8 | Ezekiel | Chui | Sider, Donald C Esq |
| 9 | Bernardo | Figeroa | Clark, Richard Cpa |
| 10 | Valentine | Gillian | Fbs Business Finance |
| 11 | Lavera | Perin | Abc Enterprises Inc |
| 12 | Brock | Bologna | Orinda News |
| 13 | Lorrie | Nestle | Ballard Spahr Andrews |
| 14 | Sabra | Uyetake | Lowy Limousine Service |
| 15 | Karl | Klonowski | Rossi, Michael M |
| 16 | Amber | Monarrez | Branford Wire & Mfg Co |
| 17 | Delmy | Ahle | Wye Technologies Inc |
| 18 | Jamal | Vanausdal | Hubbard, Bruce Esq |
| 19 | Maurine | Yglesias | Schultz, Thomas C Md |
| 20 | Tawna | Buvenus | H H H Enterprises Inc |
| 21 | Elly | Morocco | Killion Industries |
| 22 | Bobbie | Rhym | Smits, Patricia Garity |
| 23 | Tamar | Hoogland | A K Construction Co |
| 24 | Devorah | Chickering | Garrison Ind |
| 25 | Timothy | Mulqueen | Saronix Nymph Products |
| 26 | Myra | Munns | Anker Law Office |
| 27 | Stephen | Emigh | Sharp, J Daniel Esq |
| 28 | Danica | Bruschke | Stevens, Charles T |
| 29 | Wilda | Giguere | Mclaughlin, Luther W Cpa |
| 30 | Elvera | Benimadho | Tree Musketeers |
| 31 | Lisha | Centini | Industrial Paper Shredders Inc |
| 32 | Rosio | Cork | Green Goddess |
| 33 | Sue | Kownacki | Juno Chefs Incorporated |
| 34 | Oretha | Menter | Custom Engineering Inc |
| 35 | Rory | Papasergi | Bailey Cntl Co Div Babcock |
| 36 | Talia | Riopelle | Ford Brothers Wholesale Inc |
| 37 | Howard | Paulas | Asendorf, J Alan Esq |
| 38 | Marti | Maybury | Eldridge, Kristin K Esq |
| 39 | Melissa | Wiklund | Moapa Valley Federal Credit Un |
| 40 | Daren | Weirather | Panasystems |
| 41 | Lezlie | Craghead | Chang, Carolyn Esq |
| 42 | Yolando | Luczki | Dal Tile Corporation |
| 43 | Lizette | Stem | Edward S Katz |
| 44 | Latrice | Tolfree | United Van Lines Agent |

| | | | |
|----|-----------|-------------|--------------------------------|
| 45 | Clay | Hoa | Scat Enterprises |
| 46 | Jennifer | Fallick | Nagle, Daniel J Esq |
| 47 | Joesph | Degonia | A R Packaging |
| 48 | Claribel | Varriano | Meca |
| 49 | Hillary | Skulski | Replica I |
| 50 | Merilyn | Bayless | 20 20 Printing Inc |
| 51 | Lemuel | Latzke | Computer Repair Service |
| 52 | Melodie | Knipp | Fleetwood Building Block Inc |
| 53 | Felicidad | Poullion | Mccorkle, Tom S Esq |
| 54 | Barrett | Toyama | Case Foundation Co |
| 55 | Norah | Waymire | Carmichael, Jeffery L Esq |
| 56 | Rebecka | Gesick | Polykote Inc |
| 57 | Freeman | Gochal | Kellermann, William T Esq |
| 58 | Avery | Steier | Dill Dill Carr & Stonbraker Pc |
| 59 | Alaine | Bergesen | Hispanic Magazine |
| 60 | Earleen | Mai | Little Sheet Metal Co |
| 61 | Eladia | Saulter | Tyee Productions Inc |
| 62 | Salena | Karpel | Hammill Mfg Co |
| 63 | Alex | Loader | Sublett, Scott Esq |
| 64 | Lashon | Vizarro | Sentry Signs |
| 65 | Lauran | Burnard | Professionals Unlimited |
| 66 | Peggie | Sturiale | Henry County Middle School |
| 67 | Heike | Berganza | Cali Sportswear Cutting Dept |
| 68 | Deandrea | Hughey | Century 21 Krall Real Estate |
| 69 | Kimberlie | Duenas | Mid Contntl Rlty & Prop Mgmt |
| 70 | Martina | Staback | Ace Signs Inc |
| 71 | Skye | Fillingim | Rodeway Inn |
| 72 | Charlene | Hamilton | Oshins & Gibbons |
| 73 | Geoffrey | Acey | Price Business Services |
| 74 | Stevie | Westerbeck | Wise, Dennis W Md |
| 75 | Pamella | Fortino | Super 8 Motel |
| 76 | Johnna | Engelberg | Thrifty Oil Co |
| 77 | Jerry | Zurcher | J & F Lumber |
| 78 | Deonna | Kippley | Midas Muffler Shops |
| 79 | Raymon | Calvaresi | Seaboard Securities Inc |
| 80 | Garry | Keetch | Italian Express Franchise Corp |
| 81 | Matthew | Neither | American Council On Sci & Hlth |
| 82 | Theodora | Restrepo | Kleri, Patricia S Esq |
| 83 | Noah | Kalafatis | Twiggs Abrams Blanchard |
| 84 | Lavonda | Hengel | Bradley Nameplate Corp |
| 85 | Herminia | Nicolozakes | Sea Island Div Of Fstr Ind Inc |

| | | | |
|-----|-----------|-----------|--------------------------------|
| 86 | Filiberto | Tawil | Flash, Elena Salerno Esq |
| 87 | Sarah | Candlish | Alabama Educational Tv Comm |
| 88 | Lucy | Treston | Franz Inc |
| 89 | Judy | Aquas | Plantation Restaurant |
| 90 | Felix | Hirpara | American Speedy Printing Ctrs |
| 91 | Ligia | Reiber | Floral Expressions |
| 92 | Goldie | Schirpke | Reuter, Arthur C Jr |
| 93 | Amie | Perigo | General Foam Corporation |
| 94 | Cherry | Lietz | Sebring & Co |
| 95 | Whitley | Tomasulo | Freehold Fence Co |
| 96 | Adelina | Nabours | Courtyard By Marriott |
| 97 | Leota | Ragel | Mayar Silk Inc |
| 98 | Kirk | Herritt | Hasting, H Duane Esq |
| 99 | Leonora | Mauson | Insty Prints |
| 100 | Lashawnda | Stuer | Rodriguez, J Christopher Esq |
| 101 | Nobuko | Halsey | Goeman Wood Products Inc |
| 102 | Mariann | Bilden | H P G Industrys Inc |
| 103 | Samira | Heintzman | Mutual Fish Co |
| 104 | Margart | Meisel | Yeates, Arthur L Aia |
| 105 | Louvenia | Beech | John Ortiz Nts Therapy Center |
| 106 | Vincenza | Zepp | Kbor 1600 Am |
| 107 | Venita | Maillard | Wallace Church Assoc Inc |
| 108 | Kasandra | Semidey | Can Tron |
| 109 | Maile | Linahan | Thompson Steel Company Inc |
| 110 | Quentin | Birkner | Spoor Behrins Campbell & Young |
| 111 | Sheron | Louissant | Potter, Brenda J Cpa |
| 112 | Izetta | Funnell | Baird Kurtz & Dobson |
| 113 | Markus | Lukasik | M & M Store Fixtures Co Inc |
| 114 | Jennie | Drymon | Osborne, Michelle M Esq |
| 115 | Galen | Cantres | Del Charro Apartments |
| 116 | Lizbeth | Kohl | E T Balancing Co Inc |
| 117 | Glenn | Berray | Griswold, John E Esq |
| 118 | Annelle | Tagala | Vico Products Mfg Co |
| 119 | Levi | Munis | Farrell & Johnson Office Equip |
| 120 | Hubert | Walthall | Dee, Deanna |
| 121 | Arthur | Farrow | Young, Timothy L Esq |
| 122 | Staci | Schmaltz | Midwest Contracting & Mfg Inc |
| 123 | Catarina | Gleich | Terk, Robert E Esq |
| 124 | Virgie | Kiel | Cullen, Terrence P Esq |
| 125 | Jolene | Ostolaza | Central Die Casting Mfg Co Inc |
| 126 | Selma | Husser | Armon Communications |

| | | | |
|-----|----------|------------|-------------------------------|
| 127 | Luisa | Jurney | Forest Fire Laboratory |
| 128 | Dick | Wenzinger | Wheaton Plastic Products |
| 129 | Caprice | Suell | Egnor, W Dan Esq |
| 130 | Annmarie | Castros | Tipiak Inc |
| 131 | Cecil | Lapage | Hawkes, Douglas D |
| 132 | Josphine | Villanueva | Santa Cruz Community Internet |
| 133 | Britt | Galam | Wheatley Trucking Company |
| 134 | Nu | Mcnease | Amazonia Film Project |
| 135 | Daniela | Comnick | Water & Sewer Department |
| 136 | Cecilia | Colaizzo | Switchcraft Inc |
| 137 | Izetta | Dewar | Lisatoni, Jean Esq |
| 138 | Ruthann | Keener | Maiden Craft Inc |
| 139 | Joni | Breland | Carriage House Cllsn Rpr Inc |
| 140 | Jess | Chaffins | New York Public Library |
| 141 | Nickolas | Juvera | United Oil Co Inc |
| 142 | Rasheeda | Sayaphon | Kummerer, J Michael Esq |
| 143 | Alpha | Palaia | Stoffer, James M Jr |
| 144 | Lili | Paskin | Morgan Custom Homes |
| 145 | Catalina | Tillotson | Icn Pharmaceuticals Inc |
| 146 | Thaddeus | Ankeny | Atc Contracting |
| 147 | Lai | Harabedian | Buergi & Madden Scale |
| 148 | Flo | Bookamer | Simonton Howe & Schneider Pc |
| 149 | Jani | Biddy | Warehouse Office & Paper Prod |

| | address | city | county | state |
|----|----------------------------|---------------|----------------|-------|
| 1 | 8 W Cerritos Ave #54 | Bridgeport | Gloucester | NJ |
| 2 | 5 Boston Ave #88 | Sioux Falls | Minnehaha | SD |
| 3 | 228 Runamuck Pl #2808 | Baltimore | Baltimore City | MD |
| 4 | 37275 St Rt 17m M | Middle Island | Suffolk | NY |
| 5 | 25 E 75th St #69 | Los Angeles | Los Angeles | CA |
| 6 | 6 S 33rd St | Aston | Delaware | PA |
| 7 | 86 Nw 66th St #8673 | Shawnee | Johnson | KS |
| 8 | 2 Cedar Ave #84 | Easton | Talbot | MD |
| 9 | 386 9th Ave N | Conroe | Montgomery | TX |
| 10 | 775 W 17th St | San Antonio | Bexar | TX |
| 11 | 678 3rd Ave | Miami | Miami-Dade | FL |
| 12 | 4486 W O St #1 | New York | New York | NY |
| 13 | 39 S 7th St | Tullahoma | Coffee | TN |
| 14 | 98839 Hawthorne Blvd #6101 | Columbia | Richland | SC |
| 15 | 76 Brooks St #9 | Flemington | Hunterdon | NJ |
| 16 | 14288 Foster Ave #4121 | Jenkintown | Montgomery | PA |
| 17 | 65895 S 16th St | Providence | Providence | RI |

| | | | | |
|----|--------------------------|-----------------|---------------|----|
| 18 | 53075 Sw 152nd Ter #615 | Monroe Township | Middlesex | NJ |
| 19 | 59 Shady Ln #53 | Milwaukee | Milwaukee | WI |
| 20 | 3305 Nabell Ave #679 | New York | New York | NY |
| 21 | 7 W 32nd St | Erie | Erie | PA |
| 22 | 30 W 80th St #1995 | San Carlos | San Mateo | CA |
| 23 | 2737 Pistorio Rd #9230 | London | Madison | OH |
| 24 | 31 Douglas Blvd #950 | Clovis | Curry | NM |
| 25 | 44 W 4th St | Staten Island | Richmond | NY |
| 26 | 461 Prospect Pl #316 | Euless | Tarrant | TX |
| 27 | 3777 E Richmond St #900 | Akron | Summit | OH |
| 28 | 840 15th Ave | Waco | McLennan | TX |
| 29 | 1747 Calle Amanecer #2 | Anchorage | Anchorage | AK |
| 30 | 99385 Charity St #840 | San Jose | Santa Clara | CA |
| 31 | 64 5th Ave #1153 | Mc Lean | Fairfax | VA |
| 32 | 4 10th St W | High Point | Guilford | NC |
| 33 | 2 Se 3rd Ave | Mesquite | Dallas | TX |
| 34 | 8 County Center Dr #647 | Boston | Suffolk | MA |
| 35 | 83 County Road 437 #8581 | Clarks Summit | Lackawanna | PA |
| 36 | 1 N Harlem Ave #9 | Orange | Essex | NJ |
| 37 | 866 34th Ave | Denver | Denver | CO |
| 38 | 4 Warehouse Point Rd #7 | Chicago | Cook | IL |
| 39 | 61 13 Stoneridge #835 | Findlay | Hancock | OH |
| 40 | 9 N College Ave #3 | Milwaukee | Milwaukee | WI |
| 41 | 749 W 18th St #45 | Smithfield | Johnston | NC |
| 42 | 422 E 21st St | Syracuse | Onondaga | NY |
| 43 | 501 N 19th Ave | Cherry Hill | Camden | NJ |
| 44 | 81 Norris Ave #525 | Ronkonkoma | Suffolk | NY |
| 45 | 73 Saint Ann St #86 | Reno | Washoe | NV |
| 46 | 44 58th St | Wheeling | Cook | IL |
| 47 | 2887 Knowlton St #5435 | Berkeley | Alameda | CA |
| 48 | 6 Harry L Dr #6327 | Perrysburg | Wood | OH |
| 49 | 9 Wales Rd Ne #914 | Homosassa | Citrus | FL |
| 50 | 195 13n N | Santa Clara | Santa Clara | CA |
| 51 | 70 Euclid Ave #722 | Bohemia | Suffolk | NY |
| 52 | 326 E Main St #6496 | Thousand Oaks | Ventura | CA |
| 53 | 9939 N 14th St | Riverton | Burlington | NJ |
| 54 | 4252 N Washington Ave #9 | Kennedale | Tarrant | TX |
| 55 | 6 Middlegate Rd #106 | San Francisco | San Francisco | CA |
| 56 | 2026 N Plankinton Ave #3 | Austin | Travis | TX |
| 57 | 383 Gunderman Rd #197 | Coatesville | Chester | PA |
| 58 | 93 Redmond Rd #492 | Orlando | Orange | FL |

| | | | | |
|----|---------------------------------|-----------------|-------------|----|
| 59 | 7667 S Hulen St #42 | Yonkers | Westchester | NY |
| 60 | 75684 S Withlapopka Dr #32 | Dallas | Dallas | TX |
| 61 | 3958 S Dupont Hwy #7 | Ramsey | Bergen | NJ |
| 62 | 1 Garfield Ave #7 | Canton | Stark | OH |
| 63 | 37 N Elm St #916 | Tacoma | Pierce | WA |
| 64 | 433 Westminster Blvd #590 | Roseville | Placer | CA |
| 65 | 66697 Park Pl #3224 | Riverton | Fremont | WY |
| 66 | 9 N 14th St | El Cajon | San Diego | CA |
| 67 | 3 Railway Ave #75 | Little Falls | Passaic | NJ |
| 68 | 33 Lewis Rd #46 | Burlington | Alamance | NC |
| 69 | 8100 Jacksonville Rd #7 | Hays | Ellis | KS |
| 70 | 7 W Wabansia Ave #227 | Orlando | Orange | FL |
| 71 | 25 Minters Chapel Rd #9 | Minneapolis | Hennepin | MN |
| 72 | 985 E 6th Ave | Santa Rosa | Sonoma | CA |
| 73 | 7 West Ave #1 | Palatine | Cook | IL |
| 74 | 26659 N 13th St | Costa Mesa | Orange | CA |
| 75 | 669 Packerland Dr #1438 | Denver | Denver | CO |
| 76 | 5 S Colorado Blvd #449 | Bothell | Snohomish | WA |
| 77 | 77 Massillon Rd #822 | Satellite Beach | Brevard | FL |
| 78 | 57 Haven Ave #90 | Southfield | Oakland | MI |
| 79 | 6538 E Pomona St #60 | Indianapolis | Marion | IN |
| 80 | 5 Green Pond Rd #4 | Southampton | Bucks | PA |
| 81 | 636 Commerce Dr #42 | Shakopee | Scott | MN |
| 82 | 42744 Hamann Industrial Pky #82 | Miami | Miami-Dade | FL |
| 83 | 1950 5th Ave | Milwaukee | Milwaukee | WI |
| 84 | 87 Imperial Ct #79 | Fargo | Cass | ND |
| 85 | 4 58th St #3519 | Scottsdale | Maricopa | AZ |
| 86 | 3882 W Congress St #799 | Los Angeles | Los Angeles | CA |
| 87 | 45 2nd Ave #9759 | Atlanta | Fulton | GA |
| 88 | 57254 Brickell Ave #372 | Worcester | Worcester | MA |
| 89 | 8977 Connecticut Ave Nw #3 | Niles | Berrien | MI |
| 90 | 7563 Cornwall Rd #4462 | Denver | Lancaster | PA |
| 91 | 206 Main St #2804 | Lansing | Ingham | MI |
| 92 | 34 Saint George Ave #2 | Bangor | Penobscot | ME |
| 93 | 596 Santa Maria Ave #7913 | Mesquite | Dallas | TX |
| 94 | 40 9th Ave Sw #91 | Waterford | Oakland | MI |
| 95 | 2 S 15th St | Fort Worth | Tarrant | TX |
| 96 | 80 Pittsford Victor Rd #9 | Cleveland | Cuyahoga | OH |
| 97 | 99 5th Ave #33 | Trion | Chattooga | GA |
| 98 | 88 15th Ave Ne | Vestal | Broome | NY |
| 99 | 3381 E 40th Ave | Passaic | Passaic | NJ |

| | | | | |
|-----|--------------------------|-------------------|-------------------|----|
| 100 | 7422 Martin Ave #8 | Toledo | Lucas | OH |
| 101 | 8139 I Hwy 10 #92 | New Bedford | Bristol | MA |
| 102 | 3125 Packer Ave #9851 | Austin | Travis | TX |
| 103 | 8772 Old County Rd #5410 | Kent | King | WA |
| 104 | 868 State St #38 | Cincinnati | Hamilton | OH |
| 105 | 598 43rd St | Beverly Hills | Los Angeles | CA |
| 106 | 395 S 6th St #2 | El Cajon | San Diego | CA |
| 107 | 72119 S Walker Ave #63 | Anaheim | Orange | CA |
| 108 | 369 Latham St #500 | Saint Louis | Saint Louis City | MO |
| 109 | 9 Plainsboro Rd #598 | Greensboro | Guilford | NC |
| 110 | 7061 N 2nd St | Burnsville | Dakota | MN |
| 111 | 97 E 3rd St #9 | Long Island City | Queens | NY |
| 112 | 82 Winsor St #54 | Atlanta | Dekalb | GA |
| 113 | 89 20th St E #779 | Sterling Heights | Macomb | MI |
| 114 | 63728 Poway Rd #1 | Scranton | Lackawanna | PA |
| 115 | 617 Nw 36th Ave | Brook Park | Cuyahoga | OH |
| 116 | 35433 Blake St #588 | Gardena | Los Angeles | CA |
| 117 | 29 Cherry St #7073 | Des Moines | Polk | IA |
| 118 | 5 W 7th St | Parkville | Baltimore | MD |
| 119 | 2094 Ne 36th Ave | Worcester | Worcester | MA |
| 120 | 95 Main Ave #2 | Barberton | Summit | OH |
| 121 | 28 S 7th St #2824 | Englewood | Bergen | NJ |
| 122 | 18 Coronado Ave #563 | Pasadena | Los Angeles | CA |
| 123 | 78 Maryland Dr #146 | Denville | Morris | NJ |
| 124 | 76598 Rd I 95 #1 | Denver | Denver | CO |
| 125 | 1610 14th St Nw | Newport News | Newport News City | VA |
| 126 | 9 State Highway 57 #22 | Jersey City | Hudson | NJ |
| 127 | 25 Se 176th Pl | Cambridge | Middlesex | MA |
| 128 | 22 Spruce St #595 | Gardena | Los Angeles | CA |
| 129 | 90177 N 55th Ave | Nashville | Davidson | TN |
| 130 | 80312 W 32nd St | Conroe | Montgomery | TX |
| 131 | 4 Stovall St #72 | Union City | Hudson | NJ |
| 132 | 63 Smith Ln #8343 | Moss | Clay | TN |
| 133 | 2500 Pringle Rd Se #508 | Hatfield | Montgomery | PA |
| 134 | 88 Sw 28th Ter | Harrison | Hudson | NJ |
| 135 | 7 Flowers Rd #403 | Trenton | Mercer | NJ |
| 136 | 4 Nw 12th St #3849 | Madison | Dane | WI |
| 137 | 2 W Scyene Rd #3 | Baltimore | Baltimore City | MD |
| 138 | 3424 29th St Se | Kerrville | Kerr | TX |
| 139 | 35 E Main St #43 | Elk Grove Village | Cook | IL |
| 140 | 18 3rd Ave | New York | New York | NY |

| | | | | |
|-----|---------------------------------|---------------|-------------------------------|----|
| 141 | 177 S Rider Trl #52 | Crystal River | Citrus | FL |
| 142 | 251 Park Ave #979 | Saratoga | Santa Clara | CA |
| 143 | 43496 Commercial Dr #29 | Cherry Hill | Camden | NJ |
| 144 | 20113 4th Ave E | Kearny | Hudson | NJ |
| 145 | 3338 A Lockport Pl #6 | Margate City | Atlantic | NJ |
| 146 | 5 Washington St #1 | Roseville | Placer | CA |
| 147 | 1933 Packer Ave #2 | Novato | Marin | CA |
| 148 | 89992 E 15th St | Alliance | Box Butte | NE |
| 149 | 61556 W 20th Ave | Seattle | King | WA |
| | zip phone1 phone2 | | email | |
| 1 | 8014 856-636-8749 856-264-4130 | | art@venere.org | |
| 2 | 57105 605-414-2147 605-794-4895 | | sage_wieser@cox.net | |
| 3 | 21224 410-655-8723 410-804-4694 | | kris@gmail.com | |
| 4 | 11953 631-335-3414 631-677-3675 | | amaclead@gmail.com | |
| 5 | 90034 310-498-5651 310-254-3084 | | kiley.caldarera@aol.com | |
| 6 | 19014 610-545-3615 610-492-4643 | | bette_nicka@cox.net | |
| 7 | 66218 913-388-2079 913-899-1103 | | chanel.caudy@caudy.org | |
| 8 | 21601 410-669-1642 410-235-8738 | | ezeziel@chui.com | |
| 9 | 77301 936-336-3951 936-597-3614 | | bfigeroa@aol.com | |
| 10 | 78204 210-812-9597 210-300-6244 | | valentine_gillian@gmail.com | |
| 11 | 33196 305-606-7291 305-995-2078 | | lperin@perin.org | |
| 12 | 10003 212-402-9216 212-617-5063 | | bbologna@yahoo.com | |
| 13 | 37388 931-875-6644 931-303-6041 | | lnestle@hotmail.com | |
| 14 | 29201 803-925-5213 803-681-3678 | | sabra@uyetake.org | |
| 15 | 8822 908-877-6135 908-470-4661 | | karl_klonowski@yahoo.com | |
| 16 | 19046 215-934-8655 215-329-6386 | | amber_monarrez@monarrez.org | |
| 17 | 2909 401-458-2547 401-559-8961 | | delmy.ahle@hotmail.com | |
| 18 | 8831 732-234-1546 732-904-2931 | | jamal@vanausdal.org | |
| 19 | 53214 414-748-1374 414-573-7719 | | maurine_yglesias@yglesias.com | |
| 20 | 10009 212-674-9610 212-462-9157 | | tawna@gmail.com | |
| 21 | 16502 814-393-5571 814-420-3553 | | elly_morocco@gmail.com | |
| 22 | 94070 650-528-5783 650-811-9032 | | brhym@rhym.com | |
| 23 | 43140 740-343-8575 740-526-5410 | | tamar@hotmail.com | |
| 24 | 88101 505-975-8559 505-950-1763 | | devorah@hotmail.com | |
| 25 | 10309 718-332-6527 718-654-7063 | | timothy_mulqueen@mulqueen.org | |
| 26 | 76040 817-914-7518 817-451-3518 | | mmunns@cox.net | |
| 27 | 44302 330-537-5358 330-700-2312 | | stephen_emigh@hotmail.com | |
| 28 | 76708 254-782-8569 254-205-1422 | | danica_bruschke@gmail.com | |
| 29 | 99501 907-870-5536 907-914-9482 | | wilda@cox.net | |
| 30 | 95110 408-703-8505 408-440-8447 | | elvera.benimadho@cox.net | |
| 31 | 22102 703-235-3937 703-475-7568 | | lisha@centini.org | |

| | | | | |
|----|-------|--------------|--------------|-------------------------------|
| 32 | 27263 | 336-243-5659 | 336-497-4407 | rosio.cork@gmail.com |
| 33 | 75149 | 972-666-3413 | 972-742-4000 | sue@aol.com |
| 34 | 2210 | 617-418-5043 | 617-697-6024 | oretha_menter@yahoo.com |
| 35 | 18411 | 570-867-7489 | 570-469-8401 | rpapasergi@cox.net |
| 36 | 7050 | 973-245-2133 | 973-818-9788 | talia_riopelle@aol.com |
| 37 | 80231 | 303-623-4241 | 303-692-3118 | hpaulas@gmail.com |
| 38 | 60638 | 773-775-4522 | 773-539-1058 | marti.maybury@yahoo.com |
| 39 | 45840 | 419-939-3613 | 419-254-4591 | melissa@cox.net |
| 40 | 53216 | 414-959-2540 | 414-838-3151 | dweirather@aol.com |
| 41 | 27577 | 919-533-3762 | 919-885-2453 | lezzlie.craghead@craghead.org |
| 42 | 13214 | 315-304-4759 | 315-640-6357 | yolando@cox.net |
| 43 | 8002 | 856-487-5412 | 856-702-3676 | lizette.stem@aol.com |
| 44 | 11779 | 631-957-7624 | 631-998-2102 | latrice.tolfree@hotmail.com |
| 45 | 89502 | 775-501-8109 | 775-848-9135 | choa@hoa.org |
| 46 | 60090 | 847-979-9545 | 847-800-3054 | jfallick@yahoo.com |
| 47 | 94710 | 510-677-9785 | 510-942-5916 | joesph_degonia@degonia.org |
| 48 | 43551 | 419-544-4900 | 419-573-2033 | claribel_varriano@cox.net |
| 49 | 34448 | 352-242-2570 | 352-990-5946 | hillary.skulski@aol.com |
| 50 | 95054 | 408-758-5015 | 408-346-2180 | merilyn_bayless@cox.net |
| 51 | 11716 | 631-748-6479 | 631-291-4976 | lemuel.latzke@gmail.com |
| 52 | 91362 | 805-690-1682 | 805-810-8964 | mknipp@gmail.com |
| 53 | 8077 | 856-305-9731 | 856-828-6021 | fpoullion@poullion.com |
| 54 | 76060 | 817-765-5781 | 817-577-6151 | barrett.toyama@toyama.org |
| 55 | 94107 | 415-306-7897 | 415-874-2984 | norah.waymire@gmail.com |
| 56 | 78754 | 512-213-8574 | 512-693-8345 | rgesick@gesick.org |
| 57 | 19320 | 610-476-3501 | 610-752-2683 | freeman_gochal@aol.com |
| 58 | 32803 | 407-808-9439 | 407-945-8566 | avery@cox.net |
| 59 | 10701 | 914-300-9193 | 914-654-1426 | alaine_bergesen@cox.net |
| 60 | 75227 | 214-289-1973 | 214-785-6750 | earleen_mai@cox.net |
| 61 | 7446 | 201-474-4924 | 201-365-8698 | eladia@saulter.com |
| 62 | 44707 | 330-791-8557 | 330-618-2579 | skarpel@cox.net |
| 63 | 98409 | 253-660-7821 | 253-875-9222 | alex@loader.com |
| 64 | 95661 | 916-741-7884 | 916-289-4526 | lashon@aol.com |
| 65 | 82501 | 307-342-7795 | 307-453-7589 | lburnard@burnard.com |
| 66 | 92020 | 619-608-1763 | 619-695-8086 | peggie@cox.net |
| 67 | 7424 | 973-936-5095 | 973-822-8827 | heike@gmail.com |
| 68 | 27215 | 336-822-7652 | 336-467-3095 | deandrea@yahoo.com |
| 69 | 67601 | 785-629-8542 | 785-616-1685 | kimberlie_duenas@yahoo.com |
| 70 | 32822 | 407-471-6908 | 407-429-2145 | martina_staback@staback.com |
| 71 | 55401 | 612-508-2655 | 612-664-6304 | skye_fillingim@yahoo.com |
| 72 | 95407 | 707-300-1771 | 707-821-8037 | charlene.hamilton@hotmail.com |

| | | | | |
|-----|-------|--------------|--------------|--------------------------------|
| 73 | 60067 | 847-222-1734 | 847-556-2909 | geoffrey@gmail.com |
| 74 | 92626 | 949-867-4077 | 949-903-3898 | stevie.westerbeck@yahoo.com |
| 75 | 80212 | 303-404-2210 | 303-794-1341 | pamella@fortino.com |
| 76 | 98021 | 425-986-7573 | 425-700-3751 | jengelberg@engelberg.org |
| 77 | 32937 | 321-518-5938 | 321-597-2159 | jzurcher@zurcher.org |
| 78 | 48075 | 248-913-4677 | 248-793-4966 | deonna_kippley@hotmail.com |
| 79 | 46222 | 317-825-4724 | 317-342-1532 | raymon.calvaresi@gmail.com |
| 80 | 18966 | 215-979-8776 | 215-846-9046 | garry_keetch@hotmail.com |
| 81 | 55379 | 952-651-7597 | 952-906-4597 | mneither@yahoo.com |
| 82 | 33136 | 305-936-8226 | 305-573-1085 | theodora.restrepo@restrepo.com |
| 83 | 53209 | 414-263-5287 | 414-660-9766 | noah.kalafatis@aol.com |
| 84 | 58102 | 701-898-2154 | 701-421-7080 | lavonda@cox.net |
| 85 | 85254 | 602-954-5141 | 602-304-6433 | herminia@nicolozakes.org |
| 86 | 90016 | 323-765-2528 | 323-842-8226 | ftawil@hotmail.com |
| 87 | 30328 | 770-732-1194 | 770-531-2842 | sarah.candlish@gmail.com |
| 88 | 1602 | 508-769-5250 | 508-502-5634 | lucy@cox.net |
| 89 | 49120 | 269-756-7222 | 269-431-9464 | jaquas@aquas.com |
| 90 | 17517 | 717-491-5643 | 717-583-1497 | felix_hirpara@cox.net |
| 91 | 48933 | 517-906-1108 | 517-747-7664 | lreiber@cox.net |
| 92 | 4401 | 207-295-7569 | 207-748-3722 | goldie.schirpke@yahoo.com |
| 93 | 75150 | 972-419-7946 | 972-898-1033 | amie.perigo@yahoo.com |
| 94 | 48329 | 248-980-6904 | 248-697-7722 | cherry@lietz.com |
| 95 | 76107 | 817-526-4408 | 817-819-7799 | whitley.tomasulo@aol.com |
| 96 | 44103 | 216-230-4892 | 216-937-5320 | adelina_nabours@gmail.com |
| 97 | 30753 | 706-221-4243 | 706-616-5131 | leota.ragel@gmail.com |
| 98 | 13850 | 607-407-3716 | 607-350-7690 | kirk.herritt@aol.com |
| 99 | 7055 | 973-412-2995 | 973-355-2120 | leonora@yahoo.com |
| 100 | 43607 | 419-588-8719 | 419-399-1744 | lstuer@cox.net |
| 101 | 2745 | 508-855-9887 | 508-897-7916 | nobuko.halsey@yahoo.com |
| 102 | 78753 | 512-223-4791 | 512-742-1149 | mariann.bilden@aol.com |
| 103 | 98032 | 206-311-4137 | 206-923-6042 | sheintzman@hotmail.com |
| 104 | 45251 | 513-617-2362 | 513-747-9603 | margart_meisel@yahoo.com |
| 105 | 90210 | 310-820-2117 | 310-652-2379 | louvenia.beech@beech.com |
| 106 | 92020 | 619-603-5125 | 619-935-6661 | vzepp@gmail.com |
| 107 | 92801 | 714-523-6653 | 714-663-9740 | venita_maillard@gmail.com |
| 108 | 63102 | 314-732-9131 | 314-697-3652 | kasandra_semidey@semidey.com |
| 109 | 27409 | 336-670-2640 | 336-364-6037 | mlinahan@yahoo.com |
| 110 | 55337 | 952-702-7993 | 952-314-5871 | qbirkner@aol.com |
| 111 | 11101 | 718-976-8610 | 718-613-9994 | sheron@aol.com |
| 112 | 30340 | 770-844-3447 | 770-584-4119 | izetta.funnell@hotmail.com |
| 113 | 48310 | 586-970-7380 | 586-247-1614 | markus@yahoo.com |

| | | | | |
|-----|-------|--------------|--------------|------------------------------------|
| 114 | 18509 | 570-218-4831 | 570-868-8688 | jennie@cox.net |
| 115 | 44142 | 216-600-6111 | 216-871-6876 | galen@yahoo.com |
| 116 | 90248 | 310-699-1222 | 310-955-5788 | lizbeth@yahoo.com |
| 117 | 50315 | 515-370-7348 | 515-372-1738 | gberray@gmail.com |
| 118 | 21234 | 410-757-1035 | 410-234-2267 | annelle@yahoo.com |
| 119 | 1603 | 508-456-4907 | 508-658-7802 | levi.munis@gmail.com |
| 120 | 44203 | 330-903-1345 | 330-566-8898 | hubert@walthall.org |
| 121 | 7631 | 201-238-5688 | 201-772-4377 | arthur.farrow@yahoo.com |
| 122 | 91106 | 626-866-2339 | 626-293-7678 | staci_schmaltz@aol.com |
| 123 | 7834 | 973-210-3994 | 973-491-8723 | catarina_gleich@hotmail.com |
| 124 | 80216 | 303-776-7548 | 303-845-5408 | vkiel@hotmail.com |
| 125 | 23608 | 757-682-7116 | 757-940-1741 | jolene@yahoo.com |
| 126 | 7306 | 201-991-8369 | 201-772-7699 | selma.husser@cox.net |
| 127 | 2138 | 617-365-2134 | 617-544-2541 | ljurney@hotmail.com |
| 128 | 90248 | 310-510-9713 | 310-936-2258 | dick@yahoo.com |
| 129 | 37211 | 615-246-1824 | 615-726-4537 | caprice@aol.com |
| 130 | 77301 | 936-751-7961 | 936-937-2334 | annmarie_castros@gmail.com |
| 131 | 7087 | 201-693-3967 | 201-856-2720 | clapage@lapage.com |
| 132 | 38575 | 931-553-9774 | 931-486-6946 | josphine_villanueva@villanueva.com |
| 133 | 19440 | 215-888-3304 | 215-351-8523 | britt@galam.org |
| 134 | 7029 | 973-751-9003 | 973-903-4175 | nu@gmail.com |
| 135 | 8611 | 609-200-8577 | 609-398-2805 | dcomnick@cox.net |
| 136 | 53717 | 608-382-4541 | 608-302-3387 | cecilia_colaizzo@colaizzo.com |
| 137 | 21217 | 410-473-1708 | 410-522-7621 | idewar@dewar.com |
| 138 | 78028 | 830-258-2769 | 830-919-5991 | ruthann@hotmail.com |
| 139 | 60007 | 847-519-5906 | 847-740-5304 | joni_breland@cox.net |
| 140 | 10016 | 212-510-4633 | 212-428-9538 | jess.chaffins@chaffins.org |
| 141 | 34429 | 352-598-8301 | 352-947-6152 | nickolas_juvera@cox.net |
| 142 | 95070 | 408-805-4309 | 408-997-7490 | rasheeda@aol.com |
| 143 | 8003 | 856-312-2629 | 856-513-7024 | alpha@yahoo.com |
| 144 | 7032 | 201-431-2989 | 201-478-8540 | lili.paskin@cox.net |
| 145 | 8402 | 609-373-3332 | 609-826-4990 | catalina@hotmail.com |
| 146 | 95678 | 916-920-3571 | 916-459-2433 | tankeny@ankeny.org |
| 147 | 94945 | 415-423-3294 | 415-926-6089 | lai@gmail.com |
| 148 | 69301 | 308-726-2182 | 308-250-6987 | flo.bookamer@cox.net |
| 149 | 98104 | 206-711-6498 | 206-395-6284 | jbiddy@yahoo.com |

web Apartment_Number

| | | |
|---|---|------|
| 1 | http://www.chemeljameslcpa.com | 54 |
| 2 | http://www.truhlarandtruhlarattys.com | 88 |
| 3 | http://www.kingchristopheraesq.com | 2808 |
| 4 | http://www.rangoniofflorence.com | 17 |

| | | |
|----|---|------|
| 5 | http://www.feinerbros.com | 75 |
| 6 | http://www.sportenart.com | 33 |
| 7 | http://www.professionalimageinc.com | 66 |
| 8 | http://www.siderdonaldcesq.com | 84 |
| 9 | http://www.clarkrichardcpa.com | 9 |
| 10 | http://www.fbsbusinessfinance.com | 17 |
| 11 | http://www.abcenterprisesinc.com | 3 |
| 12 | http://www.orindanews.com | 1 |
| 13 | http://www.ballardspahrandrews.com | 7 |
| 14 | http://www.lowylimousineservice.com | 6101 |
| 15 | http://www.rossimichaelm.com | 9 |
| 16 | http://www.branfordwiremfgco.com | 4121 |
| 17 | http://www.wyetechnologiesinc.com | 16 |
| 18 | http://www.hubbardbruceesq.com | 152 |
| 19 | http://www.schultzthomascmd.com | 53 |
| 20 | http://www.hhhenterprisesinc.com | 679 |
| 21 | http://www.killionindustries.com | 32 |
| 22 | http://www.smitspatriciagarity.com | 80 |
| 23 | http://www.akconstructionco.com | 9230 |
| 24 | http://www.garrisonind.com | 950 |
| 25 | http://www.saronixnymphproducts.com | 4 |
| 26 | http://www.ankerlawoffice.com | 316 |
| 27 | http://www.sharpjdanielesq.com | 900 |
| 28 | http://www.stevenscharlest.com | 15 |
| 29 | http://www.mclaughlinlutherwcpa.com | 2 |
| 30 | http://www.treemusketeers.com | 840 |
| 31 | http://www.industrialpapershreddersinc.com | 5 |
| 32 | http://www.greengoddess.com | 10 |
| 33 | http://www.junochefsincorporated.com | 3 |
| 34 | http://www.customengineeringinc.com | 647 |
| 35 | http://www.baileycntlcodivbabcock.com | 8581 |
| 36 | http://www.fordbrotherswholesaleinc.com | 9 |
| 37 | http://www.asendorffjalanesq.com | 34 |
| 38 | http://www.eldridgekristinkesq.com | 7 |
| 39 | http://www.moapavalleyfederalcreditun.com | 835 |
| 40 | http://www.panasystems.com | 3 |
| 41 | http://www.changcarolynesq.com | 18 |
| 42 | http://www.daltilecorporation.com | 21 |
| 43 | http://www.edwardskatz.com | 19 |
| 44 | http://www.unitedvanlinesagent.com | 525 |
| 45 | http://www.scaterprises.com | 86 |

| | | |
|----|---|------|
| 46 | http://www.nagledanieljesq.com | 58 |
| 47 | http://www.arpackaging.com | 5435 |
| 48 | http://www.meca.com | 6327 |
| 49 | http://www.replicai.com | 914 |
| 50 | http://www.printinginc.com | 13 |
| 51 | http://www.computerrepairservice.com | 722 |
| 52 | http://www.fleetwoodbuildingblockinc.com | 6496 |
| 53 | http://www.mccorkletomsesq.com | 14 |
| 54 | http://www.casefoundationco.com | 9 |
| 55 | http://www.carmichaeljefferylesq.com | 106 |
| 56 | http://www.polykoteinc.com | 3 |
| 57 | http://www.kellermannwilliamtesq.com | 197 |
| 58 | http://www.dilldillcarrstonbrakerpc.com | 492 |
| 59 | http://www.hispanicmagazine.com | 42 |
| 60 | http://www.littlesheetmetalco.com | 32 |
| 61 | http://www.tyee productionsinc.com | 7 |
| 62 | http://www.hammillmfgco.com | 7 |
| 63 | http://www.sublettscottesq.com | 916 |
| 64 | http://www.sentrysigns.com | 590 |
| 65 | http://www.professionalsunlimited.com | 3224 |
| 66 | http://www.henrycountymiddleschool.com | 14 |
| 67 | http://www.calisportswearcuttingdept.com | 75 |
| 68 | http://www.centurykrallrealestate.com | 46 |
| 69 | http://www.midcontntlrltypropmgmt.com | 7 |
| 70 | http://www.acesignsinc.com | 227 |
| 71 | http://www.rodewayinn.com | 9 |
| 72 | http://www.oshinsgibbons.com | 6 |
| 73 | http://www.pricebusinessservices.com | 1 |
| 74 | http://www.wisedenniswmd.com | 13 |
| 75 | http://www.supermotel.com | 1438 |
| 76 | http://www.thriftyoilco.com | 449 |
| 77 | http://www.jflumber.com | 822 |
| 78 | http://www.midasmufflershops.com | 90 |
| 79 | http://www.seaboardsecuritiesinc.com | 60 |
| 80 | http://www.italianexpressfranchisecorp.com | 4 |
| 81 | http://www.americancouncilonscihlth.com | 42 |
| 82 | http://www.kleripatriciasesq.com | 82 |
| 83 | http://www.twiggsabramsblanchard.com | 5 |
| 84 | http://www.bradleynameplatecorp.com | 79 |
| 85 | http://www.seaislanddivoffstrindinc.com | 58 |
| 86 | http://www.flashelenasalernoeshq.com | 799 |

| | | |
|-----|---|------|
| 87 | http://www.alabamaeducationaltvcomm.com | 2 |
| 88 | http://www.franzinc.com | 372 |
| 89 | http://www.plantationrestaurant.com | 3 |
| 90 | http://www.americanspeedyprintingctrs.com | 4462 |
| 91 | http://www.floralexpressions.com | 2804 |
| 92 | http://www.reuterarthurcjr.com | 2 |
| 93 | http://www.generalfoamcorporation.com | 7913 |
| 94 | http://www.sebringco.com | 9 |
| 95 | http://www.freeholdfenceco.com | 15 |
| 96 | http://www.courtyardbymarriott.com | 9 |
| 97 | http://www.mayarsilkinc.com | 5 |
| 98 | http://www.hastingshduaneesq.com | 15 |
| 99 | http://www.instyprints.com | 40 |
| 100 | http://www.rodriquezjchristophersq.com | 8 |
| 101 | http://www.goemanwoodproductsinc.com | 92 |
| 102 | http://www.hpgindustriesinc.com | 9851 |
| 103 | http://www.mutualfishco.com | 5410 |
| 104 | http://www.yeatesarthurlaia.com | 38 |
| 105 | http://www.johnortizntstherapycenter.com | 43 |
| 106 | http://www.kboram.com | 6 |
| 107 | http://www.wallacechurchassocinc.com | 63 |
| 108 | http://www.cantron.com | 500 |
| 109 | http://www.thompsonsteelcompanyinc.com | 598 |
| 110 | http://www.spoorbehrinscampbellyoung.com | 2 |
| 111 | http://www.potterbrendajcpa.com | 3 |
| 112 | http://www.bairdkurtzdobson.com | 54 |
| 113 | http://www.mmstorefixturescoinc.com | 20 |
| 114 | http://www.osbornemichellemesq.com | 1 |
| 115 | http://www.delcharroapartments.com | 36 |
| 116 | http://www.etbalancingcoinc.com | 588 |
| 117 | http://www.griswoldjohneesq.com | 7073 |
| 118 | http://www.vicoproductsmfgco.com | 7 |
| 119 | http://www.farrelljohnsonofficeequip.com | 36 |
| 120 | http://www.deedeanna.com | 2 |
| 121 | http://www.youngtimothylesq.com | 7 |
| 122 | http://www.midwestcontractingmfginc.com | 563 |
| 123 | http://www.terkroberteesq.com | 146 |
| 124 | http://www.cullenterrencepesq.com | 1 |
| 125 | http://www.centraldiecastingmfgcoinc.com | 14 |
| 126 | http://www.armoncommunications.com | 22 |
| 127 | http://www.forestfirelaboratory.com | 176 |

| | | |
|-----|---|------|
| 128 | http://www.wheatonplasticproducts.com | 595 |
| 129 | http://www.egnorwdanesq.com | 55 |
| 130 | http://www.tipiakinco.com | 32 |
| 131 | http://www.hawkesdouglasd.com | 72 |
| 132 | http://www.santacruzcommunityinternet.com | 8343 |
| 133 | http://www.wheatleytruckingcompany.com | 508 |
| 134 | http://www.amazoniafilmproject.com | 28 |
| 135 | http://www.watersewerdepartment.com | 403 |
| 136 | http://www.switchcraftinc.com | 12 |
| 137 | http://www.lisatonijeanesq.com | 3 |
| 138 | http://www.maidencraftinc.com | 29 |
| 139 | http://www.carriagehousecllsnrprinc.com | 43 |
| 140 | http://www.newyorkpubliclibrary.com | 3 |
| 141 | http://www.unitedoilcoinc.com | 52 |
| 142 | http://www.kummererjmichaellesq.com | 979 |
| 143 | http://www.stofferjamesmjr.com | 29 |
| 144 | http://www.morgancustomhomes.com | 4 |
| 145 | http://www.icnpharmaceuticalsinc.com | 6 |
| 146 | http://www.atccontracting.com | 1 |
| 147 | http://www.buergimaddenscale.com | 2 |
| 148 | http://www.simontonhoweschneiderpc.com | 15 |
| 149 | http://www.warehouseofficepaperprod.com | 20 |

Finally, we are able to extract 149 apartment numbers out of 500 addresses, which is sufficient for the purpose of analysis

NOTE: For the below sections I took help of the GPT tool to conceptualize, write and understand the code.

```
library(dplyr)
library(ggplot2)

if (nrow(data_apartments) > 0) {
  # Convert Apartment_Number to numeric, removing NAs if any
  data_apartments <- data_apartments %>%
    mutate(Apartment_Number = as.numeric(Apartment_Number)) %>%
    filter(!is.na(Apartment_Number), Apartment_Number > 0) # Filter out NA and
    ↪ non-positive numbers

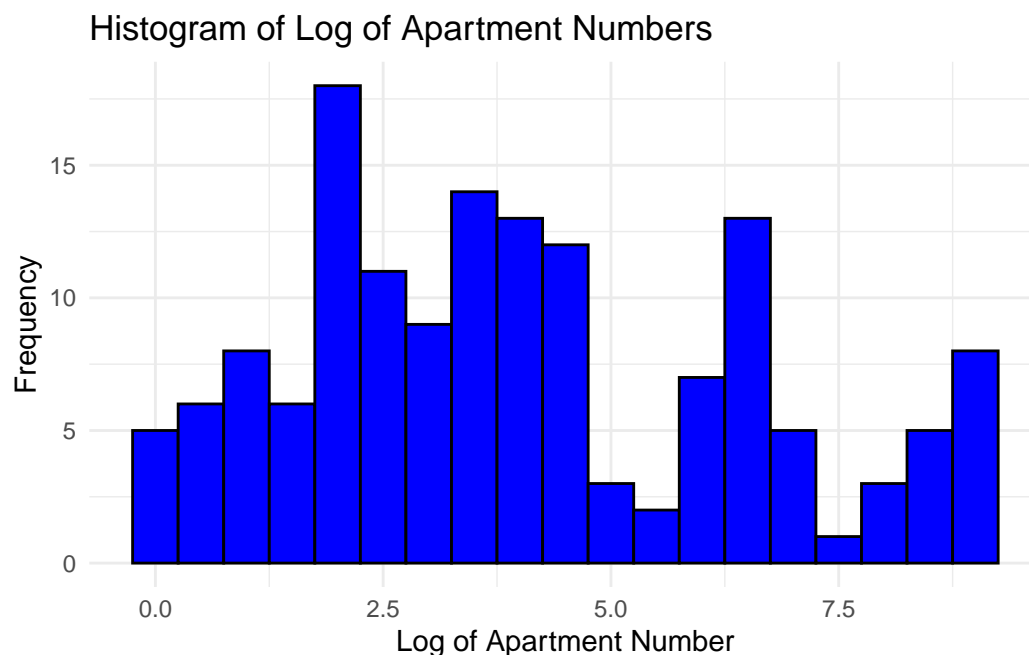
  # Calculate the log of apartment numbers
  data_apartments <- data_apartments %>%
    mutate(Log_Apartment_Number = log(Apartment_Number)) %>%
```

```

    filter(!is.infinite(Log_Apartment_Number), !is.na(Log_Apartment_Number))

# Produce a histogram of the log of the apartment numbers
ggplot(data_apartments, aes(x = Log_Apartment_Number)) +
  geom_histogram(binwidth = 0.5, fill = "blue", color = "black") +
  labs(title = "Histogram of Log of Apartment Numbers",
       x = "Log of Apartment Number",
       y = "Frequency") +
  theme_minimal() # Optional: a cleaner theme for the plot
} else {
  print("No valid apartment numbers extracted.")
}

```



5. Examine apartment numbers for Benford's Law

```

if (nrow(data_apartments) > 0) {
  # Calculate the log of apartment numbers
  data_apartments <- data_apartments %>%
    mutate(Log_Apartment_Number = log(Apartment_Number)) %>%
    filter(!is.infinite(Log_Apartment_Number), !is.na(Log_Apartment_Number))

  ggplot(data_apartments, aes(x = Log_Apartment_Number)) +
    geom_histogram(binwidth = 0.5, fill = "blue", color = "black") +

```

```

labs(title = "Histogram of Log of Apartment Numbers",
      x = "Log of Apartment Number",
      y = "Frequency") +
theme_minimal()

# Extract the leading digit of each apartment number
leading_digits <- data_apartments %>%
  mutate(Leading_Digit = as.numeric(substr(Apartment_Number, 1, 1))) %>%
  group_by(Leading_Digit) %>%
  summarise(Count = n()) %>%
  ungroup()

# Calculate the expected frequencies according to Benford's Law
expected_frequencies <- data.frame(
  Leading_Digit = 1:9,
  Expected_Probability = log10((1 + 1/(1:9))),
  Expected_Count = sum(leading_digits$Count) * log10((1 + 1/(1:9)))
)

# Combine observed and expected frequencies
benford_data <- leading_digits %>%
  right_join(expected_frequencies, by = "Leading_Digit") %>%
  mutate(Observed_Count = ifelse(is.na(Count), 0, Count),
         Leading_Digit = as.factor(Leading_Digit))

# Plot the observed and expected frequencies
ggplot(benford_data, aes(x = Leading_Digit)) +
  geom_bar(aes(y = Observed_Count), stat = "identity", fill = "blue", alpha =
    ↪ 0.6, position = "dodge") +
  geom_bar(aes(y = Expected_Count), stat = "identity", fill = "red", alpha = 0.4,
    ↪ position = "dodge") +
  labs(title = "Benford's Law Comparison",
       x = "Leading Digit",
       y = "Count") +
  scale_y_continuous(labels = scales::comma) +
  theme_minimal() +
  geom_text(aes(y = Observed_Count, label = Observed_Count), vjust = -0.5, size =
    ↪ 3) +
  geom_text(aes(y = Expected_Count, label = round(Expected_Count)), vjust = -0.5,
    ↪ size = 3, color = "red")

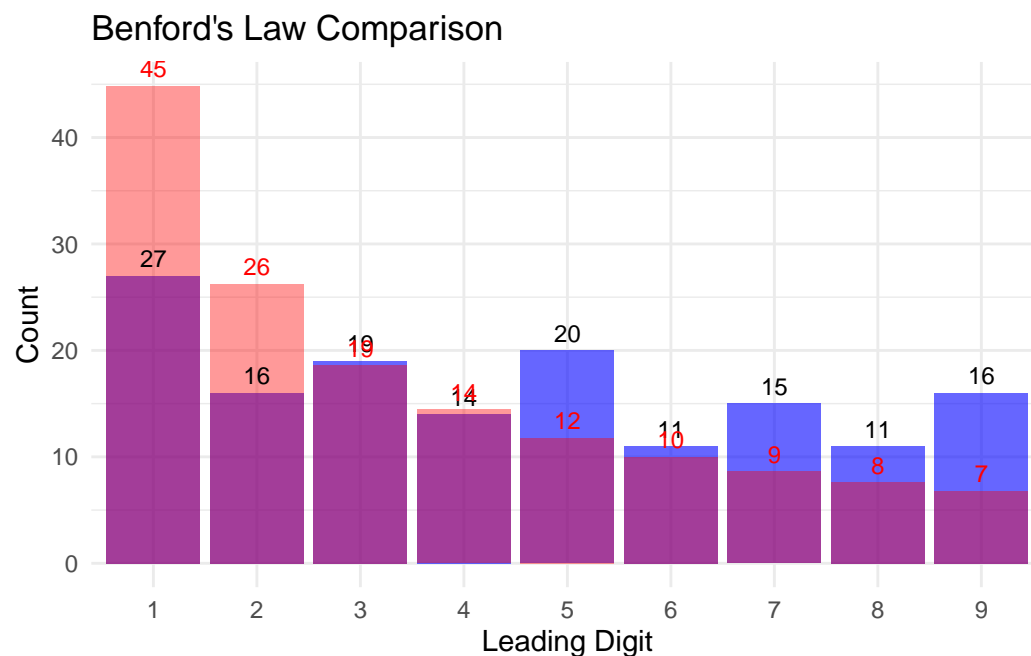
```



```

} else {
  print("No valid apartment numbers extracted.")
}

```



```

benford_result <- chisq.test(benford_data$Observed_Count, p =
  ↪ benford_data$Expected_Probability, rescale.p = TRUE)

# Conditional check for p-value
if (!is.null(benford_result$p.value)) {
  if (benford_result$p.value < 0.05) {
    print("Apartment numbers do not follow Benford's Law; they may not
      ↪ represent real data.")
  } else {
    print("Apartment numbers appear to follow Benford's Law; they may
      ↪ represent real data.")
  }
} else {
  print("Benford's Law analysis did not return a valid p-value.")
}

```

```
[1] "Apartment numbers do not follow Benford's Law; they may not represent real data."
```

The code implies that the given numbers might not represent real world data.