

ESM 262 Assignment 2

Rucha Thakar

May 11, 2017

PART 1

1. Reading in the data as-is

```
gaz_raw <- read.delim("C:/boxsync/rthakar/Courses/Spring2017/ESM262/EnvInformatics/Assignment2/CA_Features.txt",
  sep = "|",
  as.is = TRUE)
```

2. Selecting required columns and converting data frame to tibble

```
gaz <- gaz_raw %>% select (i..FEATURE_ID, FEATURE_NAME,
  FEATURE_CLASS, STATE_ALPHA, COUNTY_NAME,
  PRIM_LAT_DEC, PRIM_LONG_DEC, SOURCE_LAT_DEC,
  SOURCE_LONG_DEC, ELEV_IN_M, MAP_NAME,
  DATE_CREATED, DATE_EDITED)

colnames(gaz) <- c("featureID", "feature_name", "feature_class", "state_alpha",
  "county_name", "primary_latitude", "primary_longitude",
  "source_latitude", "source_longitude", "elevation",
  "map_name", "date_created", "date_edited")

gaz <- as.tibble(gaz)
```

3. Change class to appropriate types

Using https://geonames.usgs.gov/domestic/states_fileformat.htm as reference, data types of descriptors were changed.

```
gaz$primary_latitude <- as.numeric(gaz$primary_latitude)
gaz$primary_longitude <- as.numeric(gaz$primary_longitude)
gaz$source_latitude <- as.numeric(gaz$source_latitude)
gaz$source_longitude <- as.numeric(gaz$source_longitude)
gaz$elevation <- as.numeric(gaz$elevation)
gaz$date_created <- mdy (gaz$date_created)
gaz$date_edited <- mdy (gaz$date_edited)
```

4. Removing unknown data

According to https://geonames.usgs.gov/domestic/states_fileformat.htm, records showing “Unknown” and zeros for the latitude and longitude DMS and decimal fields, respectively, indicate that the coordinates of the feature are unknown. They are recorded in the database as zeros to satisfy the format requirements

of a numerical data type. They are not errors and do not reference the actual geographic coordinates at 0 latitude, 0 longitude.

```
gaz <- gaz %>% filter (primary_longitude != 0 | primary_latitude != 0)
```

In addition, drop NA values

```
gaz <- gaz %>%  
  drop_na(primary_latitude) %>%  
  drop_na(primary_longitude)
```

5. Selecting only features belonging to California

Removed all features that do not belong to the state of California, USA

```
California <- gaz %>%  
  filter(state_alpha == "CA")
```

6. Saving file to disk

Writing final file to the disk as final_file_part1.csv using “|” as separator

```
write.table(California, file="California_Data.csv", sep = "|")
```

PART 2

1. Most-frequently-occurring feature name in California

```
Frequent_Name <- California %>%  
  count (feature_name)%>% filter (n == max(n))
```

```
Fr_Name <- Frequent_Name$feature_name[1]  
Fr_Count <- Frequent_Name$n[1]
```

Church of Christ is the most frequently occurring feature name in California. It occurs 228 times.

2. least-frequently-occurring feature class in California

```
Rare_class <- California %>% count (feature_class)%>% filter (n == min(n))
```

Isthmus and Sea occur the least frequently in California. They occur 1 time and 1 time, respectively.

3. approximate center point of each county

Removing empty county names

```
California <- California %>% filter (county_name != "")
```

```
County_Group <- group_by(California, county_name)  
County_Mean <- summarize (County_Group, mean (primary_longitude), mean (primary_latitude))
```

```
head (County_Mean)
```

```
## # A tibble: 6 × 3
##   county_name `mean(primary_longitude)` `mean(primary_latitude)`
##   <chr>          <dbl>          <dbl>
## 1 Alameda        -122.1109          37.72641
## 2 Alpine         -119.8411          38.60157
## 3 Amador         -120.6859          38.43400
## 4 Butte          -121.5643          39.66573
## 5 Calaveras      -120.5478          38.17118
## 6 Colusa         -122.3094          39.19792
```

4. Fractions of the total number of features in each county that are natural vs. man-made

63 feature classes are categorized between natural and man-made according to the classification described in <https://geonames.usgs.gov/apex/f?p=gnispq:8:0::::>

```
manmade <- c("Airport", "Bridge", "Building",
             "Canal", "Cemetery", "Census", "Church",
             "Civil", "Crossing", "Dam",
             "Harbor", "Hospital", "Levee", "Locale",
             "Military", "Pillar", "Populated Place", "Post Pffice",
             "Reservoir", "School", "Tower", "Tunnel", "Well")

California$feature <- ifelse(California$feature_class %in% manmade, "Manmade", "Natural")

Display_Data <- data.frame(California$feature_class, California$feature)
head (Display_Data)
```

```
##   California.feature_class California.feature
## 1 Park                    Natural
## 2 School                  Manmade
## 3 Valley                  Natural
## 4 Valley                  Natural
## 5 Spring                  Natural
## 6 Tunnel                  Manmade
```

Calculate fractions of natural vs man-made feature classes

```
feature_count <- California %>% count (feature)

Manmade_Fraction <- feature_count$n[1]/sum(feature_count$n)
Natural_Fraction <- feature_count$n[2]/sum(feature_count$n)
```

Man-made feature classes are 0.5691267 in proportion to all feature classes in California. Natural feature classes 0.4308733 in proportion to all feature classes in California.