Assignment\_2.Informatics.

Wagner Quiros Pereira

5/9/2017

## ESM ASSIGNMENT 2.

## Part1. Data Ingest.

**1. Reading the gazetteer data as-is (all columns; no type conversion) into a gaz\_raw tibble and keeping only specific columns**

suppressPackageStartupMessages({  
 library(tidyverse)  
 library(dplyr)  
 library(magrittr)  
 library(knitr)  
})  
  
gaz\_raw<- read\_delim("CA\_Features\_20170401.zip", delim = "|")

## Parsed with column specification:  
## cols(  
## .default = col\_character(),  
## `﻿FEATURE\_ID` = col\_integer(),  
## PRIM\_LAT\_DEC = col\_double(),  
## PRIM\_LONG\_DEC = col\_double(),  
## SOURCE\_LAT\_DEC = col\_double(),  
## SOURCE\_LONG\_DEC = col\_double(),  
## ELEV\_IN\_M = col\_integer(),  
## ELEV\_IN\_FT = col\_integer()  
## )

## See spec(...) for full column specifications.

View(gaz\_raw)  
gaz <- gaz\_raw %>%   
 select (-COUNTY\_NUMERIC, STATE\_NUMERIC, PRIMARY\_LAT\_DMS, PRIM\_LONG\_DMS, SOURCE\_LAT\_DMS, SOURCE\_LONG\_DMS, ELEV\_IN\_FT) %>%   
 mutate(DATE\_CREATED = as.Date(DATE\_CREATED, format= "%m/%d/%Y"),   
 DATE\_EDITED = as.Date(DATE\_EDITED, format= "%m/%d/%Y" ))  
as\_tibble(gaz\_raw)

## # A tibble: 122,051 × 20  
## `﻿FEATURE\_ID` FEATURE\_NAME FEATURE\_CLASS  
## <int> <chr> <chr>  
## 1 2928 Cibola Bridge Bridge  
## 2 6185 Imperial National Wildlife Refuge Park  
## 3 8164 Mohave Canyon Valley  
## 4 8174 Mohave Valley Valley  
## 5 9144 Palo Verde Dam Dam  
## 6 9146 Palo Verde Intake Locale  
## 7 9227 Parker Valley Valley  
## 8 12628 Topock Gorge Valley  
## 9 14114 Yuma Main Canal Canal  
## 10 22751 Cibola National Wildlife Refuge Park  
## # ... with 122,041 more rows, and 17 more variables: STATE\_ALPHA <chr>,  
## # STATE\_NUMERIC <chr>, COUNTY\_NAME <chr>, COUNTY\_NUMERIC <chr>,  
## # PRIMARY\_LAT\_DMS <chr>, PRIM\_LONG\_DMS <chr>, PRIM\_LAT\_DEC <dbl>,  
## # PRIM\_LONG\_DEC <dbl>, SOURCE\_LAT\_DMS <chr>, SOURCE\_LONG\_DMS <chr>,  
## # SOURCE\_LAT\_DEC <dbl>, SOURCE\_LONG\_DEC <dbl>, ELEV\_IN\_M <int>,  
## # ELEV\_IN\_FT <int>, MAP\_NAME <chr>, DATE\_CREATED <chr>,  
## # DATE\_EDITED <chr>

**2. Converting the gaz columns to the appropriate type, and any placeholders for unknown data to NA**

gaz$PRIM\_LAT\_DEC <- ifelse(gaz$PRIM\_LAT\_DEC == 0, NA, gaz$PRIM\_LAT\_DEC)  
gaz$PRIM\_LONG\_DEC <- ifelse(gaz$PRIM\_LONG\_DEC == 0, NA, gaz$PRIM\_LONG\_DEC)  
gaz$MAP\_NAME <- ifelse(gaz$MAP\_NAME =='Unknown', NA, gaz$MAP\_NAME)  
View(gaz)

**3. Selecting main observations**

gaz <- gaz %>%  
 filter(!(PRIM\_LAT\_DEC == 'NA' | PRIM\_LONG\_DEC == 'NA')) %>%   
 filter(STATE\_ALPHA == 'CA')  
View(gaz)

**4. Saving clean up doc with delim |**

write\_delim(gaz, "/Users/wagnerquiros/Desktop/UCSB/Courses/Spring 2017/Informatics/newprojects/gaz.csv", delim = "|")  
View(gaz)

## Part2. Analyze.

**1.What is the most-frequently-occuring feature name ?**

Common\_feature\_name <- count(gaz, FEATURE\_NAME)  
Common\_feature\_name[which.max(Common\_feature\_name$n), 1]

## # A tibble: 1 × 1  
## FEATURE\_NAME  
## <chr>  
## 1 Church of Christ

\*\*The most frequestly ocurring feature name is the church of Christ\*

**2.What is the most-frequently-occuring feature class ?**

Common\_feature\_class <- count(gaz, FEATURE\_CLASS)  
Common\_feature\_class[which.min(Common\_feature\_class$n), 1]

## # A tibble: 1 × 1  
## FEATURE\_CLASS  
## <chr>  
## 1 Isthmus

\*\*The most frequestly ocurring feature class is Isthmus\*

**3.What is the approximate center point of each county CPC ?**

c\_county\_point <- gaz %>%  
 group\_by(COUNTY\_NAME) %>%  
 summarise(latmin = min(PRIM\_LAT\_DEC, na.rm= TRUE),  
 latmax = max(PRIM\_LAT\_DEC, na.rm= TRUE),  
 longmin = min(PRIM\_LONG\_DEC, na.rm= TRUE),  
 longmax = max(PRIM\_LONG\_DEC, na.rm= TRUE)) %>%   
 mutate(lat\_center = (latmin + latmax) /2)%>%   
 mutate(long\_center = (longmin + longmax)/2) %>%   
 select(County = COUNTY\_NAME, Latitud = lat\_center,   
 Longitude= long\_center)  
View(c\_county\_point)  
Table1 <- table(c\_county\_point)  
View(Table1)

**4.What are the fractions of the total number of features in each county that are natural? man-made?**

natural <- c("Valley", "Spring" ,"Stream", "Gut", "Canal", "Cape", "Lake", "Summit", "Plain", "Woods", "Gap", "Ridge","Bay", "Beach", "Channel", "Slope", "Island", "Basin", "Cliff", "Swamp","Area" , "Bend", "Range", "Falls", "Forest", "Lava", "Rapids", "Crater", "Glacier", "Oilfield","Sea", "Arroyo", "Isthmus")  
  
Feature\_category <- tibble(FEATURE\_CLASS= unique(gaz$FEATURE\_CLASS)) %>%   
 mutate(category = ifelse(FEATURE\_CLASS %in% natural == "TRUE", 1, 0))   
  
#0=manmade, 1= natural  
   
gaz\_complete <- inner\_join(gaz, Feature\_category, by = "FEATURE\_CLASS")  
  
Features\_type <- gaz\_complete%>%   
 group\_by(COUNTY\_NAME) %>%   
 summarise(total = n(), natural = sum(category)) %>%   
 mutate(fraction\_natural = natural/total) %>%   
 mutate(fraction\_manmade = 1 - fraction\_natural) %>%   
 select(County= COUNTY\_NAME, Natural= fraction\_natural, Man\_Made= fraction\_manmade)

kable(c\_county\_point)

|  |  |  |
| --- | --- | --- |
| County | Latitud | Longitude |
| Alameda | 37.68525 | -121.9243 |
| Alpine | 37.61799 | -118.2290 |
| Amador | 38.35542 | -121.0613 |
| Butte | 39.72335 | -121.5716 |
| Calaveras | 36.46287 | -119.8929 |
| Colusa | 39.16739 | -122.2780 |
| Contra Costa | 37.90659 | -121.9944 |
| Del Norte | 41.69998 | -123.9550 |
| El Dorado | 37.97298 | -121.4447 |
| Fresno | 36.74745 | -119.6338 |
| Glenn | 39.62933 | -122.4071 |
| Humboldt | 40.65793 | -122.0243 |
| Imperial | 33.05796 | -115.2855 |
| Inyo | 36.60175 | -117.2923 |
| Kern | 35.34304 | -119.4605 |
| Kings | 36.13049 | -119.8870 |
| Lake | 39.13503 | -122.7503 |
| Lassen | 40.46185 | -120.8094 |
| Los Angeles | 35.08041 | -118.9973 |
| Madera | 36.04061 | -119.7934 |
| Marin | 36.83411 | -121.9622 |
| Mariposa | 37.24062 | -119.3346 |
| Mendocino | 39.38642 | -123.4288 |
| Merced | 37.18383 | -120.6907 |
| Modoc | 41.58628 | -120.7315 |
| Mono | 38.06252 | -118.9393 |
| Monterey | 36.33260 | -121.1135 |
| Napa | 38.49838 | -122.3625 |
| Nevada | 39.26712 | -120.6413 |
| Orange | 33.66613 | -117.7801 |
| Placer | 39.03074 | -120.7767 |
| Plumas | 37.34911 | -119.4346 |
| Riverside | 36.14524 | -118.4051 |
| Sacramento | 39.09157 | -121.6143 |
| San Benito | 36.59107 | -121.1142 |
| San Bernardino | 35.87137 | -117.8211 |
| San Diego | 35.16475 | -119.2360 |
| San Francisco | 36.00691 | -120.8974 |
| San Joaquin | 37.89854 | -121.2530 |
| San Luis Obispo | 35.35580 | -120.4077 |
| San Mateo | 37.39077 | -122.3197 |
| Santa Barbara | 33.96514 | -119.5775 |
| Santa Clara | 38.88102 | -121.8937 |
| Santa Cruz | 35.63555 | -120.4298 |
| Shasta | 40.78738 | -121.6522 |
| Sierra | 37.37716 | -120.5399 |
| Siskiyou | 41.50110 | -122.5810 |
| Solano | 36.13968 | -120.5618 |
| Sonoma | 38.46991 | -122.5055 |
| Stanislaus | 37.61389 | -120.9406 |
| Sutter | 39.03162 | -121.6965 |
| Tehama | 40.05988 | -122.1986 |
| Trinity | 40.66529 | -123.0399 |
| Tulare | 36.29215 | -118.7800 |
| Tuolumne | 39.00205 | -121.5801 |
| Ventura | 34.04778 | -119.1644 |
| Yolo | 38.63463 | -121.9447 |
| Yuba | 39.27781 | -121.3127 |
| NA | 37.96936 | -122.4480 |

kable(Features\_type)

|  |  |  |
| --- | --- | --- |
| County | Natural | Man\_Made |
| Alameda | 0.0823032 | 0.9176968 |
| Alpine | 0.5418895 | 0.4581105 |
| Amador | 0.2852405 | 0.7147595 |
| Butte | 0.3648449 | 0.6351551 |
| Calaveras | 0.2957746 | 0.7042254 |
| Colusa | 0.4952199 | 0.5047801 |
| Contra Costa | 0.1646374 | 0.8353626 |
| Del Norte | 0.5154799 | 0.4845201 |
| El Dorado | 0.3872449 | 0.6127551 |
| Fresno | 0.4125594 | 0.5874406 |
| Glenn | 0.4557439 | 0.5442561 |
| Humboldt | 0.5501607 | 0.4498393 |
| Imperial | 0.6144144 | 0.3855856 |
| Inyo | 0.5236092 | 0.4763908 |
| Kern | 0.3113703 | 0.6886297 |
| Kings | 0.3360544 | 0.6639456 |
| Lake | 0.5402884 | 0.4597116 |
| Lassen | 0.5313577 | 0.4686423 |
| Los Angeles | 0.1084098 | 0.8915902 |
| Madera | 0.3800000 | 0.6200000 |
| Marin | 0.2585079 | 0.7414921 |
| Mariposa | 0.4965920 | 0.5034080 |
| Mendocino | 0.5610329 | 0.4389671 |
| Merced | 0.3476357 | 0.6523643 |
| Modoc | 0.5182972 | 0.4817028 |
| Mono | 0.5180636 | 0.4819364 |
| Monterey | 0.3540573 | 0.6459427 |
| Napa | 0.3318486 | 0.6681514 |
| Nevada | 0.3674797 | 0.6325203 |
| Orange | 0.0847971 | 0.9152029 |
| Placer | 0.3444508 | 0.6555492 |
| Plumas | 0.5429757 | 0.4570243 |
| Riverside | 0.2173482 | 0.7826518 |
| Sacramento | 0.0446970 | 0.9553030 |
| San Benito | 0.4872180 | 0.5127820 |
| San Bernardino | 0.2640118 | 0.7359882 |
| San Diego | 0.1724342 | 0.8275658 |
| San Francisco | 0.0692156 | 0.9307844 |
| San Joaquin | 0.1422652 | 0.8577348 |
| San Luis Obispo | 0.3723012 | 0.6276988 |
| San Mateo | 0.1711230 | 0.8288770 |
| Santa Barbara | 0.3333333 | 0.6666667 |
| Santa Clara | 0.1416877 | 0.8583123 |
| Santa Cruz | 0.2543253 | 0.7456747 |
| Shasta | 0.5345603 | 0.4654397 |
| Sierra | 0.4207650 | 0.5792350 |
| Siskiyou | 0.6145475 | 0.3854525 |
| Solano | 0.2327869 | 0.7672131 |
| Sonoma | 0.2923833 | 0.7076167 |
| Stanislaus | 0.2580927 | 0.7419073 |
| Sutter | 0.2390671 | 0.7609329 |
| Tehama | 0.5362319 | 0.4637681 |
| Trinity | 0.6309897 | 0.3690103 |
| Tulare | 0.4009885 | 0.5990115 |
| Tuolumne | 0.4601449 | 0.5398551 |
| Ventura | 0.2307359 | 0.7692641 |
| Yolo | 0.2865248 | 0.7134752 |
| Yuba | 0.3316750 | 0.6683250 |
| NA | 0.5000000 | 0.5000000 |