ESM262-Assignment1-Vela

Alejandro Vela

5/2/2018

## Part 1: Import and Tidy

### Part 1-1: Load Libraries

library(plyr)  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────────────────────────────────────────────────── tidyverse 1.2.1 ──

## ✔ ggplot2 2.2.1 ✔ purrr 0.2.4  
## ✔ tibble 1.4.2 ✔ dplyr 0.7.4  
## ✔ tidyr 0.8.0 ✔ stringr 1.3.0  
## ✔ readr 1.1.1 ✔ forcats 0.3.0

## ── Conflicts ────────────────────────────────────────────────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::arrange() masks plyr::arrange()  
## ✖ purrr::compact() masks plyr::compact()  
## ✖ dplyr::count() masks plyr::count()  
## ✖ dplyr::failwith() masks plyr::failwith()  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::id() masks plyr::id()  
## ✖ dplyr::lag() masks stats::lag()  
## ✖ dplyr::mutate() masks plyr::mutate()  
## ✖ dplyr::rename() masks plyr::rename()  
## ✖ dplyr::summarise() masks plyr::summarise()  
## ✖ dplyr::summarize() masks plyr::summarize()

library(readr)  
library(tibble)  
library(dplyr)  
library(tidyr)

### Part 1-2: Import Raw Data

Raw\_CA\_Gazetteer\_Dataset <- read\_delim("~/Documents/Bren School Documents/3-Spring Quarter 2018/ESM262-Computing for Environmental Sciences/Assignment1/CA\_Features\_20180401.txt",   
 "|", escape\_double = FALSE, trim\_ws = TRUE)

## 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(Raw\_CA\_Gazetteer\_Dataset)

## Warning: running command ''/usr/bin/otool' -L '/Library/Frameworks/  
## R.framework/Resources/modules/R\_de.so'' had status 1

### Part 1-3: Copy Only Required Columns

CA\_Gazetteer\_NewColumns <- Raw\_CA\_Gazetteer\_Dataset  
  
names(CA\_Gazetteer\_NewColumns) <- c("ID", "Name", "Class", "State", "State\_ID", "County", "County\_ID", "Latitude\_DMS", "Longitude\_DMS", "Latitude\_DEC", "Longitude\_DEC", "Latitude\_DMS\_SRC", "Longitude\_DMS\_SRC", "Latitude\_DEC\_SRC", "Longitude\_DEC\_SRC", "Elevation\_MTS", "Elevation\_FT", "MapName", "DateCreated", "DateEdited")  
  
View(CA\_Gazetteer\_NewColumns)

## Warning: running command ''/usr/bin/otool' -L '/Library/Frameworks/  
## R.framework/Resources/modules/R\_de.so'' had status 1

CA\_Gazetter\_SelectedColumns <- select(CA\_Gazetteer\_NewColumns, ID, Name, Class, State, County, Latitude\_DEC, Longitude\_DEC, Latitude\_DEC\_SRC, Longitude\_DEC\_SRC, Elevation\_MTS, MapName, DateCreated, DateEdited)  
  
View(CA\_Gazetter\_SelectedColumns)

## Warning: running command ''/usr/bin/otool' -L '/Library/Frameworks/  
## R.framework/Resources/modules/R\_de.so'' had status 1

### Part 1-4: Convert the gaz columns to the appropiate type. Convert any placeholders for unknown data to NA

#The only columns needed to change types was the dates, so I proceeded to change both dates columns.  
  
str(CA\_Gazetter\_SelectedColumns)

## Classes 'tbl\_df', 'tbl' and 'data.frame': 122308 obs. of 13 variables:  
## $ ID : int 2928 6185 8164 8174 9144 9146 9227 12628 14114 22751 ...  
## $ Name : chr "Cibola Bridge" "Imperial National Wildlife Refuge" "Mohave Canyon" "Mohave Valley" ...  
## $ Class : chr "Bridge" "Park" "Valley" "Valley" ...  
## $ State : chr "AZ" "AZ" "AZ" "AZ" ...  
## $ County : chr "La Paz" "La Paz" "Mohave" "Mohave" ...  
## $ Latitude\_DEC : num 33.4 33.1 34.6 35.1 33.7 ...  
## $ Longitude\_DEC : num -115 -115 -114 -115 -115 ...  
## $ Latitude\_DEC\_SRC : num NA NA 34.7 34.7 NA ...  
## $ Longitude\_DEC\_SRC: num NA NA -114 -114 NA ...  
## $ Elevation\_MTS : int 71 87 138 159 84 86 81 138 43 67 ...  
## $ MapName : chr "Palo Verde" "Picacho SW" "Castle Rock" "Davis Dam SE" ...  
## $ DateCreated : chr "02/08/1980" "02/08/1980" "02/08/1980" "02/08/1980" ...  
## $ DateEdited : chr NA "05/20/2017" "07/12/2016" NA ...  
## - attr(\*, "spec")=List of 2  
## ..$ cols :List of 20  
## .. ..$ FEATURE\_ID : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_integer" "collector"  
## .. ..$ FEATURE\_NAME : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## .. ..$ FEATURE\_CLASS : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## .. ..$ STATE\_ALPHA : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## .. ..$ STATE\_NUMERIC : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## .. ..$ COUNTY\_NAME : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## .. ..$ COUNTY\_NUMERIC : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## .. ..$ PRIMARY\_LAT\_DMS: list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## .. ..$ PRIM\_LONG\_DMS : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## .. ..$ PRIM\_LAT\_DEC : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_double" "collector"  
## .. ..$ PRIM\_LONG\_DEC : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_double" "collector"  
## .. ..$ SOURCE\_LAT\_DMS : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## .. ..$ SOURCE\_LONG\_DMS: list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## .. ..$ SOURCE\_LAT\_DEC : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_double" "collector"  
## .. ..$ SOURCE\_LONG\_DEC: list()  
## .. .. ..- attr(\*, "class")= chr "collector\_double" "collector"  
## .. ..$ ELEV\_IN\_M : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_integer" "collector"  
## .. ..$ ELEV\_IN\_FT : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_integer" "collector"  
## .. ..$ MAP\_NAME : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## .. ..$ DATE\_CREATED : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## .. ..$ DATE\_EDITED : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## ..$ default: list()  
## .. ..- attr(\*, "class")= chr "collector\_guess" "collector"  
## ..- attr(\*, "class")= chr "col\_spec"

CA\_Gazetter\_SelectedColumns$DateCreated <- as.Date(CA\_Gazetter\_SelectedColumns$DateCreated, format = "%m/%d/%Y")  
CA\_Gazetter\_SelectedColumns$DateEdited <- as.Date(CA\_Gazetter\_SelectedColumns$DateEdited, format = "%m/%d/%Y")  
  
str(CA\_Gazetter\_SelectedColumns)

## Classes 'tbl\_df', 'tbl' and 'data.frame': 122308 obs. of 13 variables:  
## $ ID : int 2928 6185 8164 8174 9144 9146 9227 12628 14114 22751 ...  
## $ Name : chr "Cibola Bridge" "Imperial National Wildlife Refuge" "Mohave Canyon" "Mohave Valley" ...  
## $ Class : chr "Bridge" "Park" "Valley" "Valley" ...  
## $ State : chr "AZ" "AZ" "AZ" "AZ" ...  
## $ County : chr "La Paz" "La Paz" "Mohave" "Mohave" ...  
## $ Latitude\_DEC : num 33.4 33.1 34.6 35.1 33.7 ...  
## $ Longitude\_DEC : num -115 -115 -114 -115 -115 ...  
## $ Latitude\_DEC\_SRC : num NA NA 34.7 34.7 NA ...  
## $ Longitude\_DEC\_SRC: num NA NA -114 -114 NA ...  
## $ Elevation\_MTS : int 71 87 138 159 84 86 81 138 43 67 ...  
## $ MapName : chr "Palo Verde" "Picacho SW" "Castle Rock" "Davis Dam SE" ...  
## $ DateCreated : Date, format: "1980-02-08" "1980-02-08" ...  
## $ DateEdited : Date, format: NA "2017-05-20" ...  
## - attr(\*, "spec")=List of 2  
## ..$ cols :List of 20  
## .. ..$ FEATURE\_ID : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_integer" "collector"  
## .. ..$ FEATURE\_NAME : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## .. ..$ FEATURE\_CLASS : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## .. ..$ STATE\_ALPHA : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## .. ..$ STATE\_NUMERIC : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## .. ..$ COUNTY\_NAME : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## .. ..$ COUNTY\_NUMERIC : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## .. ..$ PRIMARY\_LAT\_DMS: list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## .. ..$ PRIM\_LONG\_DMS : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## .. ..$ PRIM\_LAT\_DEC : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_double" "collector"  
## .. ..$ PRIM\_LONG\_DEC : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_double" "collector"  
## .. ..$ SOURCE\_LAT\_DMS : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## .. ..$ SOURCE\_LONG\_DMS: list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## .. ..$ SOURCE\_LAT\_DEC : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_double" "collector"  
## .. ..$ SOURCE\_LONG\_DEC: list()  
## .. .. ..- attr(\*, "class")= chr "collector\_double" "collector"  
## .. ..$ ELEV\_IN\_M : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_integer" "collector"  
## .. ..$ ELEV\_IN\_FT : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_integer" "collector"  
## .. ..$ MAP\_NAME : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## .. ..$ DATE\_CREATED : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## .. ..$ DATE\_EDITED : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## ..$ default: list()  
## .. ..- attr(\*, "class")= chr "collector\_guess" "collector"  
## ..- attr(\*, "class")= chr "col\_spec"

View(CA\_Gazetter\_SelectedColumns)

## Warning: running command ''/usr/bin/otool' -L '/Library/Frameworks/  
## R.framework/Resources/modules/R\_de.so'' had status 1

### Part 1-5: Extract California and Delete Unknown Primary Latitude or Longitude

CA\_Gazetteer\_CAOnly <- CA\_Gazetter\_SelectedColumns %>%  
 filter(State == "CA") %>%  
 filter(Latitude\_DEC != "") %>%  
 filter(Longitude\_DEC != "")

## Warning: package 'bindrcpp' was built under R version 3.4.4

View(CA\_Gazetteer\_CAOnly)

## Warning: running command ''/usr/bin/otool' -L '/Library/Frameworks/  
## R.framework/Resources/modules/R\_de.so'' had status 1

### Part 1-6: Write the Gaz Tibble to a CVS file (Using “|” as a delimiter)

write.table(CA\_Gazetteer\_CAOnly, file = "Gazetteer\_CAOnly.txt", sep = "|",  
 row.names = TRUE, col.names = NA)

## Part 2: Analyze

### Part 2-1: What is the most frequently occuring feature name?

# ```{r Most\_FeatureName}

Names\_Frequency <- count(CA\_Gazetteer\_CAOnly, ‘Name’)

Name\_Frequency\_Descending <- Names\_Frequency[with(Names\_Frequency, order(-freq)),]

View(Name\_Frequency\_Descending)

head(Name\_Frequency\_Descending)

# The most common Feature Name is ‘Church of Christ’

### Part 2-2: What is the least frequently occuring feature class?  
  
#```{r Least\_FeatureClass}  
  
Class\_Frequency <- count(CA\_Gazetteer\_CAOnly, 'Class')  
  
Class\_Frequency\_Ascending <- Class\_Frequency[with(Class\_Frequency, order(freq)),]  
  
View(Class\_Frequency\_Ascending)  
  
head(Class\_Frequency\_Ascending)  
  
#The least common Feature Class is 'Isthmus' and 'Sea'

### Part 2-4: What are the fractions of the total number of features in each county that are natural? man-made?

# ```{r Fraction\_Class}

CA\_Gazetter\_Simplified <- select(CA\_Gazetteer\_CAOnly, ID, Class, County)

View(CA\_Gazetter\_Simplified)

CA\_Gazetteer\_ByCounty <- spread(CA\_Gazetter\_Simplified, key = County, value = Class)

View(CA\_Gazetteer\_ByCounty)

# Alameda

Alameda\_County <- select(CA\_Gazetteer\_ByCounty, ID, Alameda)

View(Alameda\_County)

AlamedaCounty <- Alameda\_County[complete.cases(Alameda\_County),]

AlamedaCountyAlameda==“Airport”] <- “ManMade” AlamedaCountyAlameda==“Arch”] <- “Natural” AlamedaCountyAlameda==“Area”] <- “Natural” AlamedaCountyAlameda==“Arroyo”] <- “Natural” AlamedaCountyAlameda==“Bar”] <- “Natural” AlamedaCountyAlameda==“Basin”] <- “Natural” AlamedaCountyAlameda==“Bay”] <- “Natural” AlamedaCountyAlameda==“Beach”] <- “Natural” AlamedaCountyAlameda==“Bench”] <- “Natural” AlamedaCountyAlameda==“Bend”] <- “Natural” AlamedaCountyAlameda==“Bridge”] <- “ManMade” AlamedaCountyAlameda==“Building”] <- “ManMade” AlamedaCountyAlameda==“Canal”] <- “ManMade” AlamedaCountyAlameda==“Cape”] <- “ManMade” AlamedaCountyAlameda==“Cave”] <- “Natural” AlamedaCountyAlameda==“Cementery”] <- “ManMade” AlamedaCountyAlameda==“Census”] <- “ManMade” AlamedaCountyAlameda==“Channel”] <- “Natural” AlamedaCountyAlameda==“Church”] <- “ManMade” AlamedaCountyAlameda==“Civil”] <- “ManMade” AlamedaCountyAlameda==“Cliff”] <- “Natural” AlamedaCountyAlameda==“Crater”] <- “Natural” AlamedaCountyAlameda==“Crossing”] <- “ManMade” AlamedaCountyAlameda==“Dam”] <- “ManMade” AlamedaCountyAlameda==“Falls”] <- “Natural” AlamedaCountyAlameda==“Flat”] <- “Natural” AlamedaCountyAlameda==“Forest”] <- “Natural” AlamedaCountyAlameda==“Gap”] <- “ManMade” AlamedaCountyAlameda==“Glacier”] <- “Natural” AlamedaCountyAlameda==“Gut”] <- “Natural” AlamedaCountyAlameda==“Harbor”] <- “ManMade” AlamedaCountyAlameda==“Hospital”] <- “ManMade” AlamedaCountyAlameda==“Island”] <- “Natural” AlamedaCountyAlameda==“Isthmus”] <- “Natural” AlamedaCountyAlameda==“Lake”] <- “Natural” AlamedaCountyAlameda==“Lava”] <- “Natural” AlamedaCountyAlameda==“Levee”] <- “Natural” AlamedaCountyAlameda==“Locale”] <- “ManMade” AlamedaCountyAlameda==“Military”] <- “ManMade” AlamedaCountyAlameda==“Mine”] <- “ManMade” AlamedaCountyAlameda==“Oilfield”] <- “ManMade” AlamedaCountyAlameda==“Park”] <- “ManMade” AlamedaCountyAlameda==“Pillar”] <- “ManMade” AlamedaCountyAlameda==“Plain”] <- “ManMade” AlamedaCountyAlameda==“Populated Place”] <- “ManMade” AlamedaCountyAlameda==“Post Office”] <- “ManMade” AlamedaCountyAlameda==“Range”] <- “Natural” AlamedaCountyAlameda==“Rapids”] <- “Natural” AlamedaCountyAlameda==“Reserve”] <- “Natural” AlamedaCountyAlameda==“Reservoirs”] <- “ManMade” AlamedaCountyAlameda==“Ridge”] <- “ManMade” AlamedaCountyAlameda==“School”] <- “ManMade” AlamedaCountyAlameda==“Sea”] <- “Natural” AlamedaCountyAlameda==“Slope”] <- “Natural” AlamedaCountyAlameda==“Spring”] <- “Natural” AlamedaCountyAlameda==“Stream”] <- “Natural” AlamedaCountyAlameda==“Summit”] <- “Natural” AlamedaCountyAlameda==“Swamp”] <- “Natural” AlamedaCountyAlameda==“Tower”] <- “ManMade” AlamedaCountyAlameda==“Trail”] <- “ManMade” AlamedaCountyAlameda==“Tunnel”] <- “ManMade” AlamedaCountyAlameda==“Unknown”] <- “ManMade” AlamedaCountyAlameda==“Valley”] <- “Natural” AlamedaCountyAlameda==“Well”] <- “ManMade” AlamedaCountyAlameda==“Woods”] <- “Natural” AlamedaCountyAlameda==“Reservoir”] <- “ManMade” AlamedaCountyAlameda==“Cementery”] <- “ManMade” AlamedaCountyAlameda==“Cemetery”] <- “ManMade”

View(AlamedaCounty)

Alameda\_Frequency <- count(AlamedaCounty, ‘Status’)

View(Alameda\_Frequency)

Alameda\_Frequencyfreq/3075\*100

View(Alameda\_Frequency) #In this table it can be seen the percentaje of ManMande and Natural

# Professor, this is the only solution for this problem I could come up with, if I would do the same for every county I would take several hours. Please I would ask you to see that I came up with the solution.

```