HW2: Data Ingest

Ignacia Rivera

May 9, 2017

```r
library(tidyverse)
library(knitr)
```

Importing and tidying data

```r
#Importing

gaz_raw <- read_delim("C:/Users/Ignacia Rivera/Dropbox/MESM/Quarter VI/ComputingForEnvScience/Rivera_ESP

#Selecting variables of interest and formating date

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" ))

#Assigning NA to missed values
#Is this is the fastest way of replece them ?

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)

#Selecting observations of interest

gaz <- gaz %>%
  filter(!(PRIM_LAT_DEC == 'NA' | PRIM_LONG_DEC == 'NA')) %>%
  filter(STATE_ALPHA == 'CA')

#Saving the clean tibble

write_delim(gaz, "C:/Users/Ignacia Rivera/Dropbox/MESM/Quarter VI/ComputingForEnvScience/Rivera_ESM262/
```

Analyzing data

```r
#What is the most-frequently-occuring feature name?
```

```
count_feature_name <- count(gaz, FEATURE_NAME)
Ans1 <- count_feature_name[which.max(count_feature_name$n), 1]

#What is the least-frequently-occuring feature class?

count_feature_class <- count(gaz, FEATURE_CLASS)
Ans2 <- count_feature_class[which.min(count_feature_class$n), 1]

#What is the approximate center point of each county?

county_center <- 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_NAME, lat_center, long_center)

#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",

Feature_category <- tibble(FEATURE_CLASS= unique(gaz$FEATURE_CLASS)) %>%
  mutate(category = ifelse(FEATURE_CLASS %in% natural == "TRUE", 1, 0)) #1 = natural, 0= man-made

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_NAME, fraction_natural, fraction_manmade)
```

What is the most-frequently-occuring feature name?

The most-frequently- occuring feature in California is Church of Christ.

What is the least-frequently-occuring feature class?

The least-frequently-occuring feature in California is Isthmus.

What is the approximate center point of each county?

Table 1: Longitude and latitude of the center of each county in California.

COUNTY_NAME	lat_center	long_center
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

COUNTY_NAME	lat_center	long_center
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

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

Table 2: Fraction of natural and man-made features in each county in California.

COUNTY_NAME	fraction_natural	fraction_manmade
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

COUNTY_NAME	fraction_natural	fraction_manmade
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