CDFA APN edit

The purpose of this Rmarkdown document is to edit CDFA APNs so that they are in a format that can be matched to APNs in the Kern County Parcel shapefile attribute table.

Notes:

- APNs are self-reported by farmers to the CDFA, and so exist in a wide variety of formats in the CDFA database
- This code uses a function stored in the "2_APN_edit.R" file to edited APNs. The function takes the original APN value and puts them into a 123-456-78 format to match the format of those in the Kern County Parcel shapefile attribute table
- APN edit methods were evaluated on the number of matches made between CDFA APNs and Kern County Parcel data, and the accuracy of those matches when the Permittee and Company columns, respectively, were compared.

Output:

- CSV files of CDFA original and edited APNs in R_input and R_output files

---

```
library(tidyverse)
library(reshape2)
library(rgdal)

# Set working directory to ensure relative file pathways are functional
setwd("~/Desktop/Organics_Final/Working/R_files/Rmarkdown")

# Bring in the function stored in the '1_APN_edit.R' script. If editing that function,
# check the box that says 'source on save' to upload that latest version of the function
# in this Rmarkdown document
source("../R/1_APN_edit.R")
```

2013

```
# Read in csv that contains APNs for Kern County in 2013. Prior to being read in to this
# Rmarkdown document, APNs may have been edited in excel in one or more of the following
# ways: (1) if multiple  APNs in one cell, seperated into individual cells across multiple
# columns, (2) dashes added to those numbers that didn't have them. If dashes were added I
# always started counting from the left and added a dash after the 3rd number and after the
# 6th number. Original versions of the APNs were kept for comparison
cdfa_2013_raw = read_csv("../R_input/CSV/CDFA/excel_edited/CDFA_excelAPNs_2013.csv")

# Define parameters for APN_edit function in a list of elements. These also could be defined
# directly in the function arguments below.
apn_params = list(id_cols = 2, # ID Column used to go from wide to long format
                  apn_cols = 19:72, # These are the APN columns what will be merged into one
                  col_names = c("company","APN_index","original_apn"), # Column names of the new long t
                  year = 2013) # Year being evaluated

# Use the APN_edit_fun() with the list of parameters for the 2013 APNs.
cdfa_2013_final_df = APN_edit_fun(cdfa_2013_raw, # 2013 CDFA Dataframe
                                  id_cols = apn_params$id_cols,
                                  apn_cols = apn_params$apn_cols,
```

```
                                    col_names = apn_params$col_names,
                                    year = apn_params$year)
```

```
## Warning: Expected 5 pieces. Missing pieces filled with `NA` in 210 rows [1,
## 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 20, 21, 22, 23, 24, 25, 26, ...].
```

```
# The warning message refers to empty cells that results when APNs are separated by '-' into
# multiple columns. Those that have fewer than the maximum number of dahes recieve NAs in
# those columns. The warning is not problematic and does not effect those APN values.
```

All subsequent years follow the same order of operations as those explained in the code chunk above for 2013.

2014

```
cdfa_2014_raw = read_csv("../R_input/CSV/CDFA/excel_edited/CDFA_excelAPNs_2014.csv")
```

```
## Warning: Missing column names filled in: 'X20' [20]
```

```
apn_params = list(id_cols = 2,
                  apn_cols = 22:75,
                  col_names = c("company","APN_index","original_apn"),
                  year = 2014)

cdfa_2014_final_df = APN_edit_fun(cdfa_2014_raw,
                                  id_cols = apn_params$id_cols,
                                  apn_cols = apn_params$apn_cols,
                                  col_names = apn_params$col_names,
                                  year = apn_params$year)
```

```
## Warning: Expected 5 pieces. Additional pieces discarded in 1 rows [12].
```

```
## Warning: Expected 5 pieces. Missing pieces filled with `NA` in 221 rows [1,
## 2, 3, 4, 5, 6, 7, 11, 13, 14, 15, 17, 18, 19, 20, 21, 22, 24, 25, 26, ...].
```

2015

```
cdfa_2015_raw = read_csv("../R_input/CSV/CDFA/excel_edited/CDFA_excelAPNs_2015.csv")
```

```
## Warning: Missing column names filled in: 'X20' [20]
```

```
## Warning: Duplicated column names deduplicated: 'SiteAPN' =>
## 'SiteAPN_1' [21]
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   SiteAcreage = col_double(),
##   SiteZip = col_double(),
##   X20 = col_logical()
## )
```

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

```
apn_params = list(id_cols = 2,
                  apn_cols = 22:75,
                  col_names = c("company","APN_index","original_apn"),
                  year = 2015)

cdfa_2015_final = APN_edit_fun(cdfa_2015_raw,
```

```
                                  id_cols = apn_params$id_cols,
                                  apn_cols = apn_params$apn_cols,
                                  col_names = apn_params$col_names,
                                  year = apn_params$year)
```

```
## Warning: Expected 5 pieces. Missing pieces filled with `NA` in 250 rows
## [10, 11, 12, 13, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 30,
## 31, 32, ...].
```

2016

```
cdfa_2016_raw = read_csv("../R_input/CSV/CDFA/excel_edited/CDFA_excelAPNs_2016.csv")
```

```
## Warning: Missing column names filled in: 'X20' [20]
```

```
apn_params = list(id_cols = 2,
                  apn_cols = 22:75,
                  col_names = c("company","APN_index","original_apn"),
                  year = 2016)

cdfa_2016_final_df = APN_edit_fun(cdfa_2016_raw,
                                  id_cols = apn_params$id_cols,
                                  apn_cols = apn_params$apn_cols,
                                  col_names = apn_params$col_names,
                                  year = apn_params$year)
```

```
## Warning: Expected 5 pieces. Additional pieces discarded in 2 rows [77, 81].
```

```
## Warning: Expected 5 pieces. Missing pieces filled with `NA` in 348 rows
## [11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 23, 25, 26, 27, 28, 29, 30,
## 31, 32, ...].
```

2017

```
cdfa_2017_raw = read_csv("../R_input/CSV/CDFA/excel_edited/CDFA_excelAPNs_2017.csv")
```

```
## Warning: Duplicated column names deduplicated: 'Site APN' => 'Site
## APN_1' [20]
```

```
apn_params = list(id_cols = 2,
                  apn_cols = 21:74,
                  col_names = c("company","APN_index","original_apn"),
                  year = 2017)

cdfa_2017_final_df = APN_edit_fun(cdfa_2017_raw,
                                  id_cols = apn_params$id_cols,
                                  apn_cols = apn_params$apn_cols,
                                  col_names = apn_params$col_names,
                                  year = apn_params$year)
```

```
## Warning: Expected 5 pieces. Additional pieces discarded in 1 rows [323].
```

```
## Warning: Expected 5 pieces. Missing pieces filled with `NA` in 341 rows
## [11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 23, 25, 26, 27, 28, 29, 30,
## 31, 32, ...].
```