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.

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

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

source("../R/2_APN_edit.R") # Function used to edit APNs
```

2013

```r
cdfa_2013_raw = read_csv("../R_input/CSV/CDFA/excel_edited/CDFA_excelAPNs_2013.csv")

apn_params = list(id_cols = 2, # Column
                  apn_cols = 19:72,
                  col_names = c("company","APN_index","original_apn"),
                  year = 2013)

cdfa_2013_final_df = APN_edit_fun(cdfa_2013_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 210 rows [1,
## 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 20, 21, 22, 23, 24, 25, 26, ...].
```

2014

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

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

```r
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, ...].