Assignment_1

Juliette Verstaen

5/2/2019

Contents

Import and Tidy

```
library(tidyverse)
```

```
## -- Attaching packages ----------------------------------------------------------------------- tidyverse
```

```
## v ggplot2 3.1.0     v purrr   0.2.5
## v tibble  2.0.1     v dplyr   0.7.8
## v tidyr   0.8.2     v stringr 1.3.1
## v readr   1.3.1     v forcats 0.3.0
```

```
## -- Conflicts -------------------------------------------------------------------------- tidyverse_confli
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
parcels_raw <- read_csv("data/Santa_Barbara_County_parcels_2011.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   OBJECTID = col_double(),
##   Acreage = col_double(),
##   LandValue = col_double(),
##   StrImpr = col_double(),
##   TradeFix = col_double(),
##   LivImpr = col_double(),
##   PerPropDec = col_double(),
##   PersPropUn = col_double(),
##   MobileHome = col_double(),
##   Exemptions = col_double(),
##   HomeOwEx = col_double(),
##   NetSecVal = col_double(),
##   Net_Impr = col_double(),
##   Net_Pers = col_double(),
##   Net_UNX = col_double(),
##   Net_AV = col_double(),
##   MFrac = col_double(),
##   POBox = col_double(),
##   Country = col_logical(),
```

```
##   SNum = col_double()
##   # ... with 6 more columns
## )
```

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

```
## Warning: 566 parsing failures.
##  row     col               expected         actual                                        file
## 1388 MFrac   no trailing characters /2             'data/Santa_Barbara_County_parcels_2011.csv'
## 1511 Country 1/0/T/F/TRUE/FALSE     CHINA 200120   'data/Santa_Barbara_County_parcels_2011.csv'
## 1657 Country 1/0/T/F/TRUE/FALSE     09618-0039     'data/Santa_Barbara_County_parcels_2011.csv'
## 1972 Country 1/0/T/F/TRUE/FALSE     AUSTRALIA 4005 'data/Santa_Barbara_County_parcels_2011.csv'
## 2387 MFrac   no trailing characters /2             'data/Santa_Barbara_County_parcels_2011.csv'
## .... ....... ...................... ............. ........................................
## See problems(...) for more details.
```

```
as_tibble(parcels_raw)
```

```
## # A tibble: 128,566 x 56
##    OBJECTID APN   LAYER Situs1 Situs2 Acreage LandUse UseCode TRA
##       <dbl> <chr> <chr> <chr>  <chr>    <dbl> <chr>   <chr>   <chr>
##  1        1 083-~ Grou~ <NA>   <NA>     361.  PASTUR~ 5443    0720~
##  2        2 083-~ Grou~ <NA>   <NA>     295.  PASTUR~ 5443    0720~
##  3        3 083-~ Grou~ <NA>   <NA>     153.  PASTUR~ 5413    0720~
##  4        4 083-~ Grou~ <NA>   <NA>      53.6 PASTUR~ 5443    0720~
##  5        5 083-~ Grou~ <NA>   <NA>      60.9 PASTUR~ 5443    0720~
##  6        6 083-~ Grou~ <NA>   <NA>      73   PASTUR~ 5413    0720~
##  7        7 083-~ Grou~ <NA>   <NA>     100   PASTUR~ 5443    0720~
##  8        8 083-~ Grou~ <NA>   <NA>     275.  PASTUR~ 5443    0720~
##  9        9 083-~ Grou~ <NA>   <NA>      16.6 UTILIT~ 8100    0940~
## 10       10 083-~ Grou~ <NA>   <NA>     321.  PASTUR~ 5443    0720~
## # ... with 128,556 more rows, and 47 more variables: NonTaxCode <chr>,
## #   SBENo <chr>, AgPres <chr>, LandValue <dbl>, StrImpr <dbl>,
## #   TradeFix <dbl>, LivImpr <dbl>, PerPropDec <dbl>, PersPropUn <dbl>,
## #   MobileHome <dbl>, Exemptions <dbl>, ExempCode <chr>, HomeOwEx <dbl>,
## #   NetSecVal <dbl>, Net_Impr <dbl>, Net_Pers <dbl>, Net_UNX <dbl>,
## #   Net_AV <dbl>, MNumber <chr>, MFrac <dbl>, MDir <chr>, MStreet <chr>,
## #   MStrSuffix <chr>, MUnitType <chr>, MUnitNumb <chr>, POBox <dbl>,
## #   MCity <chr>, MZip <chr>, MZipExt <chr>, MState <chr>, Country <lgl>,
## #   SNum <dbl>, SFra <lgl>, SDir <chr>, SStreet <chr>, SStreetSuf <chr>,
## #   SUnitType <chr>, SUnitNumb <chr>, SCity <chr>, SZip <dbl>,
## #   SZipExt <dbl>, M_Address1 <chr>, M_Address2 <chr>, WEB_LINK <chr>,
## #   SHAPE_Leng <dbl>, Shape_area <dbl>, Shape_len <dbl>
```

```
parcels <- parcels_raw %>%
  select("APN", "Situs1", "Situs2", "Acreage", "UseCode", "NonTaxCode", "AgPres", "LandValue", "Net_Impr

parcels[is.na(parcels)] <- "NA"

#need to write to a csv file###
```

Analyze

What are the 10 most-frequently-occuring land uses (in descending order)?

```
use_code <- read_delim("data/UseCodes.csv", delim="|", quote= "")
```

```
## Parsed with column specification:
## cols(
##   UseCode = col_character(),
##   CodeDesc = col_character(),
##   CdeRetireFlg = col_double(),
##   RecDateTime = col_character(),
##   RecUserId = col_double()
## )
```

```
use_code[is.na(use_code)] <- "NA"

parcels_codes <- left_join(parcels, use_code, by="UseCode")

top_10 <- parcels_codes %>%
  separate(CodeDesc, c("CodeName", "Desc1", "Desc2"), sep= ",") %>%
  mutate(n = 1) %>%
  group_by(CodeName) %>%
  summarise(frequency = sum(n)) %>%
  arrange(desc(frequency)) %>%
  head(10)
```

```
## Warning: Expected 3 pieces. Additional pieces discarded in 800 rows [401,
## 404, 460, 461, 467, 862, 863, 864, 1711, 1723, 1724, 1725, 1726, 1727,
## 1859, 1880, 1983, 2026, 2027, 2028, ...].
```

```
## Warning: Expected 3 pieces. Missing pieces filled with `NA` in 124371 rows
## [3, 6, 9, 13, 14, 27, 29, 32, 33, 40, 41, 42, 44, 45, 46, 47, 48, 49, 50,
## 51, ...].
```

```
top_10
```

```
## # A tibble: 10 x 2
##    CodeName                                frequency
##    <chr>                                       <dbl>
##  1 Single Family Residence                     73915
##  2 Residential Rentals                          5507
##  3 Sited inside M/H Park (Rented)               4540
##  4 Vacant Land                                  4382
##  5 Condo                                        3791
##  6 Rec. Area - 5-6 class                        3563
##  7 Proper improvement                           2535
##  8 Rec. Area - 7-up class                       2226
##  9 Ag preserve                                  2079
## 10 Double Wide New Const. (after 7/1/80)        1718
```

How many acres are in agricultural preserves?

```
agg_pre <- parcels_codes %>%
  separate(CodeDesc, c("CodeName", "Desc1", "Desc2"), sep= ",") %>%
  filter(CodeName == "Ag preserve")
```

```
## Warning: Expected 3 pieces. Additional pieces discarded in 800 rows [401,
## 404, 460, 461, 467, 862, 863, 864, 1711, 1723, 1724, 1725, 1726, 1727,
```

```
## 1859, 1880, 1983, 2026, 2027, 2028, ...].
```

```
## Warning: Expected 3 pieces. Missing pieces filled with `NA` in 124371 rows
## [3, 6, 9, 13, 14, 27, 29, 32, 33, 40, 41, 42, 44, 45, 46, 47, 48, 49, 50,
## 51, ...].
```

```
total_acre <- sum(agg_pre$Acreage)
total_acre
```

```
## [1] 465605.8
```

What is the mean net assessed value per acre of the entire county?

```
mean_county <- mean(parcels_codes$LandValue)
mean_county
```

```
## [1] 227627
```

What is the total net assessed value of all non-taxable parcels?