Assignment_1

Juliette Verstaen

5/2/2019

Contents

A. Import and Tidy

```
###load packages
library(tidyverse)
```

```
## -- Attaching packages ----------------------------------------------------------------- tidyverse 1.2.1
```

```
## 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_conflicts()
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
###set working directory to Assignment_1 folder
getwd()
```

```
## [1] "/GitHub/ESM262/Assignment_1"
```

```
###load in raw data
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.
```

```
 ### tried col_types because Frew said to in class, but then I couldn't calculate anything later so I ha
#                         col_types = cols(.default = col_character())) %>%
  as_tibble()
```

```
## # A tibble: 0 x 0
```

```
###select only columns of interest
parcels <-
  transmute(parcels_raw,
    APN       = APN,
    Situs1    = Situs1,
    Situs2    = Situs2,
    Acreage   = Acreage,
    UseCode   = UseCode,
    NonTaxCode= NonTaxCode,
    AgPres    = AgPres,
    LandValue = LandValue,
    Net_Impr  = Net_Impr,
    Net_AV    = Net_AV,
    M_Address1 = M_Address1,
    M_Address2 = M_Address2)

###Convert all blanks in tibble to NAs
parcels[is.na(parcels)] <- "NA"

###write to CSV file
parcels <- write_delim(parcels,
  "parcels.csv",
  delim = "|",
  na = ""
```

```
)

parcels
```

```
## # A tibble: 128,566 x 12
##    APN   Situs1 Situs2 Acreage UseCode NonTaxCode AgPres LandValue Net_Impr
##    <chr> <chr>  <chr>    <dbl> <chr>   <chr>      <chr>      <dbl>    <dbl>
##  1 083-~ NA     NA       361.  5443    NA         72AP1~   3838662   768071
##  2 083-~ NA     NA       295.  5443    NA         72AP1~   1186685        0
##  3 083-~ NA     NA       153.  5413    NA         NA        518967        0
##  4 083-~ NA     NA        53.6 5443    NA         72AP1~    784694        0
##  5 083-~ NA     NA        60.9 5443    NA         70AP1~    784974        0
##  6 083-~ NA     NA        73   5413    NA         NA        233535        0
##  7 083-~ NA     NA       100   5443    NA         70AP1~    438298        0
##  8 083-~ NA     NA       275.  5443    NA         70AP1~    442216        0
##  9 083-~ NA     NA        16.6 8100    PU         NA             0        0
## 10 083-~ NA     NA       321.  5443    NA         70AP1~    117024        0
## # ... with 128,556 more rows, and 3 more variables: Net_AV <dbl>,
## #   M_Address1 <chr>, M_Address2 <chr>
```

B. Analyze

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

```
###load in use code data
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()
## )
```

```
                    #   col_types = cols(.default = col_character()))
use_code[is.na(use_code)] <- "NA"

###combine the parcels data with the use code data
parcels_codes <- left_join(parcels, use_code, by="UseCode")

###count the most 10 most frequenctly occuring land uses
top_10 <- parcels_codes %>%
  count(UseCode) %>%
  arrange(desc(n)) %>%
  head(10)

###add the code descriptions to the top 10 counted so you know what they are
B1_topfreq <-  left_join(top_10, use_code, by="UseCode") %>%
  select("UseCode", "n", "CodeDesc")
```

2. How many acres are in agricultural preserves?

```
###choose only parcels that are agricultural preserves and have acreage values
agpres <- parcels_codes %>%
  filter(AgPres != "NA") %>%
  filter(Acreage != "NA")

###add up the number of acres in an ag preserve
B2_agpres <- sum(agpres$Acreage)
B2_agpres
```

```
## [1] 549563.4
```

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

add up cost of all peices and then divide by the area

```
###pick out acreage and net average value of the parcels
### remove any of the parcels that are less than or equal to 0, with the assumption that is incorrect/u
mean_county <- parcels_codes %>%
  select("Acreage", "Net_AV") %>%
  filter(Net_AV >=0)

###make vectors of sum of average value and sum of acreage
sum_av <- sum(mean_county$Net_AV)
sum_ac <- sum(mean_county$Acreage)

###math for average price/acre
B3_meannv <- sum_av/sum_ac
B3_meannv
```

```
## [1] 34206.61
```

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

```
###keep only parcels that are non-taxable
mean_nontax <- parcels_codes%>%
  filter(NonTaxCode != "NA")

###Find the total net assessed value
B4_totalnontax <- sum(mean_nontax$Net_AV)

B4_totalnontax
```

```
## [1] 1093026091
```

5. What are the 10 largest property holders, by acreage?

```
###Keep only properties that have full mailing addresses and keep top 10 acreage wise
B5_topacre <- parcels_codes%>%
  filter(M_Address1 != "NA") %>%
  filter(M_Address2 != "NA") %>%
```

```
  unite(address, c("M_Address1", "M_Address2"), sep = " ") %>%
  arrange(desc(Acreage)) %>%
  head(10)%>%
  select("address", "Net_AV", "Acreage")

B5_topacre
```

```
## # A tibble: 10 x 3
##    address                                              Net_AV Acreage
##    <chr>                                                 <dbl>   <dbl>
##  1 785 MARKET ST SAN FRANCISCO CA 94103                      0  25660
##  2 201 MISSION ST 4TH FLR SAN FRANCISCO CA 94105 1831    49000  16640
##  3 166 PARADISE RD SANTA BARBARA CA 93105              2760580  10517.
##  4 166 PARADISE RD SANTA BARBARA CA 93105               925010  10036
##  5 785 MARKET ST SAN FRANCISCO CA 94103                      0   9438
##  6 870 MARKET ST.SUITE 1100 SAN FRANCISCO CA 94102    2810022   6358
##  7 600 HARRISON ST #600 SAN FRANCISCO CA 94107 1372         0   6263.
##  8 2491 BULL CANYON RD SANTA MARIA CA 93454           4457154   6174
##  9 870 MARKET ST 1100 SAN FRANCISCO CA 94102           419520   6080
## 10 650 ALAMO PINTADO  203 SOLVANG CA 93463            4843311   5786.
```

```
###how did san fransisco make it into the SB county data??
```

6. What are the 10 largest property holders, by net assessed value?

```
###Keep only properties that have full mailing addresses and keep top 10 net assessed value wise
B6_topnv <- parcels_codes%>%
  filter(M_Address1 != "NA") %>%
  filter(M_Address2 != "NA") %>%
  unite(address, c("M_Address1", "M_Address2"), sep = " ") %>%
  arrange(desc(Net_AV)) %>%
  head(10) %>%
  select("address", "Net_AV", "Acreage")

B6_topnv
```

```
## # A tibble: 10 x 3
##    address                                          Net_AV Acreage
##    <chr>                                             <dbl>   <dbl>
##  1 737 GARDEN ST SANTA BARBARA CA 93101         242575885  826.
##  2 280 CHESTNUT WESTMONT IL 60559               130917962    5.9
##  3 735 ANACAPA ST SANTA BARBARA CA 93101        124353106   17.8
##  4 1260 CHANNEL DR SANTA BARBARA CA 93105       116311340   12.3
##  5 645 FIFTH AVE 8 NEW YORK NY 10022            108600000   30.0
##  6 500 STEVENS AVE 100 SOLANA BEACH CA 92075    107926369   20.8
##  7 PO BX 340 RAMSEY NJ 07446                     88323699   50.0
##  8 110 N CARPENTER ST CHICAGO IL 60607           85751872   40.2
##  9 633 E CABRILLO BLVD SANTA BARBARA CA 93103    82206252   21.8
## 10 1112 SANTA BARBARA ST SANTA BARBARA CA 93101  75963032    0.39
```