Lab_2

Meng-Kang Kao

7/20/2022

```r
tuesdata <- tt_load('2022-07-05')
```

```
## --- Compiling #TidyTuesday Information for 2022-07-05 ----

## --- There are 3 files available ---

## --- Starting Download ---

##
##  Downloading file 1 of 3: 'rent.csv'
##  Downloading file 2 of 3: 'sf_permits.csv'
##  Downloading file 3 of 3: 'new_construction.csv'

## --- Download complete ---
```

```r
rent <- tuesdata$rent
```

```r
#summary(rent)
#200796 total rows
rent %>%
  group_by(year) %>%
  summarize(
    count = n()
  ) %>%
  arrange(desc(count))
```

```
## # A tibble: 19 x 2
##     year count
##    <dbl> <int>
##  1  2004 41847
##  2  2012 35206
##  3  2006 22533
##  4  2011 21230
##  5  2016 19400
##  6  2015 13801
##  7  2005 11285
##  8  2007  7804
##  9  2014  6231
## 10  2013  5495
```

```
## 11   2008   4580
## 12   2018   2741
## 13   2017   2406
## 14   2003   2141
## 15   2009   1667
## 16   2010   1380
## 17   2001    735
## 18   2002    291
## 19   2000     23
```

```r
#year 2012 has 35206 rows

rent_2012 <- rent %>%
  filter(
    year == 2012,
  )

sum(!is.na(rent_2012$price)) #35206 rows has price information
```

```
## [1] 35206
```

```r
sum(!is.na(rent_2012$beds)) #34023 rows has number of bedrooms information
```

```
## [1] 34023
```

```r
sum(!is.na(rent_2012$baths)) #6498 rows has number of bathrooms information
```

```
## [1] 6498
```

```r
sum(!is.na(rent_2012$sqft)) #20144 rows has number of sqft information
```

```
## [1] 20144
```

```r
sum(!is.na(rent_2012$county)) #35011 rows has number of county information
```

```
## [1] 35011
```

```r
rent_2012_clean <- rent %>%
  filter(
    year == 2012,
    !is.na(price),
    !is.na(beds),
    !is.na(baths),
    !is.na(sqft),
    !is.na(county)
  )

count(rent_2012_clean) #3764 rows of clean data after removing na rows for any of the fields used in li
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1  3764
```

```
mod <- lm(price ~ county + beds + baths + sqft, data = rent_2012_clean)
coeftest(mod, vcov=vcovHC(mod))
```

```
##
## t test of coefficients:
##
##                      Estimate Std. Error  t value  Pr(>|t|)
## (Intercept)          95.43652   78.82734   1.2107  0.226085
## countycontra costa -379.80065   38.07493  -9.9751 < 2.2e-16 ***
## countymarin         663.62547   81.17985   8.1748 4.014e-16 ***
## countynapa         -145.50206   74.75204  -1.9465  0.051674 .
## countysan francisco 1427.85277  69.83938  20.4448 < 2.2e-16 ***
## countysan mateo     829.52742   53.36767  15.5436 < 2.2e-16 ***
## countysanta clara   373.46336   35.54262  10.5075 < 2.2e-16 ***
## countysanta cruz     76.23349  167.44133   0.4553  0.648931
## countysolano       -583.66926   44.50017 -13.1161 < 2.2e-16 ***
## countysonoma       -331.88688   63.31445  -5.2419 1.677e-07 ***
## beds                 55.34450   74.15522   0.7463  0.455513
## baths               458.32614  105.63227   4.3389 1.470e-05 ***
## sqft                  0.66692    0.25197   2.6468  0.008159 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```