2023Spring_UWE

LingLei

2023-01-17

Load in data

```
income = read_csv("/Users/linglei/housing_madison/median_house_income.csv")
price = read_csv("/Users/linglei/housing_madison/median_listing_price.csv")
```

EDA Process

```
head(income)
```

```
## # A tibble: 6 x 2
##   DATE       MHIWI55025A052NCEN
##   <date>     <chr>
## 1 1989-01-01 31986
## 2 1990-01-01 .
## 3 1991-01-01 .
## 4 1992-01-01 .
## 5 1993-01-01 38821
## 6 1994-01-01 .
```

```
head(price)
```

```
## # A tibble: 6 x 2
##   DATE       MEDLISPRI31540
##   <date>              <dbl>
## 1 2016-07-01         299546
## 2 2016-08-01         299900
## 3 2016-09-01         299900
## 4 2016-10-01         297250
## 5 2016-11-01         299900
## 6 2016-12-01         299000
```

First rename the columns.

```
income = income %>%
  rename(date = DATE,
         income = MHIWI55025A052NCEN) %>%
  mutate(date = date(date),
         year = year(date));

price = price %>%
  rename(date = DATE,
         house_price = MEDLISPRI31540) %>%
  mutate(date = date(date),
         year = year(date));
```

Double check the datasets.

```
head(income)
```

```
## # A tibble: 6 x 3
##   date       income  year
##   <date>     <chr>  <dbl>
## 1 1989-01-01 31986   1989
## 2 1990-01-01 .       1990
## 3 1991-01-01 .       1991
## 4 1992-01-01 .       1992
## 5 1993-01-01 38821   1993
## 6 1994-01-01 .       1994
```

```
head(price)
```

```
## # A tibble: 6 x 3
##   date       house_price  year
##   <date>           <dbl> <dbl>
## 1 2016-07-01      299546  2016
## 2 2016-08-01      299900  2016
## 3 2016-09-01      299900  2016
## 4 2016-10-01      297250  2016
## 5 2016-11-01      299900  2016
## 6 2016-12-01      299000  2016
```

Filterring all datasets from 2000 to 2022

```
income_2000 = income %>%
  filter(year >= 2016);

price_2000 = price %>%
  filter(year >= 2016);
```

Merge two tables

```
fulldf = full_join(income_2000, price_2000, by = "date") %>%
  select(-year.x) %>%
  mutate(year = year.y) %>%
  select(-year.y)
```

Implement NA values

```
price_16 =  70815
price_17 =  72385
price_18 =  71789
price_19 =  77828
price_20 =  74829
price_21 =  77653

fulldf = fulldf %>%
  mutate(income_new = case_when(
    year == 2016 ~ price_16,
    year == 2017 ~ price_17,
    year == 2018 ~ price_18,
    year == 2019 ~ price_19,
    year == 2020 ~ price_20,
    year == 2021 ~ price_21,
  )) %>%
  select(-income) %>%
  filter(!is.na(income_new)) %>%
  select(-year) %>%
  rename(income = income_new)
```

Calculate the price to income

```
fulldf = fulldf %>%
  mutate(price_to_income = house_price / income)
```

Data from Department of Numbers Madison Wisconsin

```
date = c("June 01 2012", "September 01 2012", "December 01 2012", "March 01 2013", "June 01 2013", "Sep
price_to_income = c(3.58, 3.58, 3.58, 3.32, 3.78, 3.89, 3.71, 3.57)
realtor_data = data.frame(date = mdy(date), price_to_income)
```

Full join the tables

```
df = full_join(realtor_data, fulldf, by = "date")
```

```r
df = df %>%
  mutate(
    year = year(date),
    price_ration = case_when(
    year <= 2014 ~ price_to_income.x,
    year >= 2016 ~ price_to_income.y
  )) %>%
  select(-price_to_income.x, -price_to_income.y) %>%
  rename(price_to_income = price_ration)
```

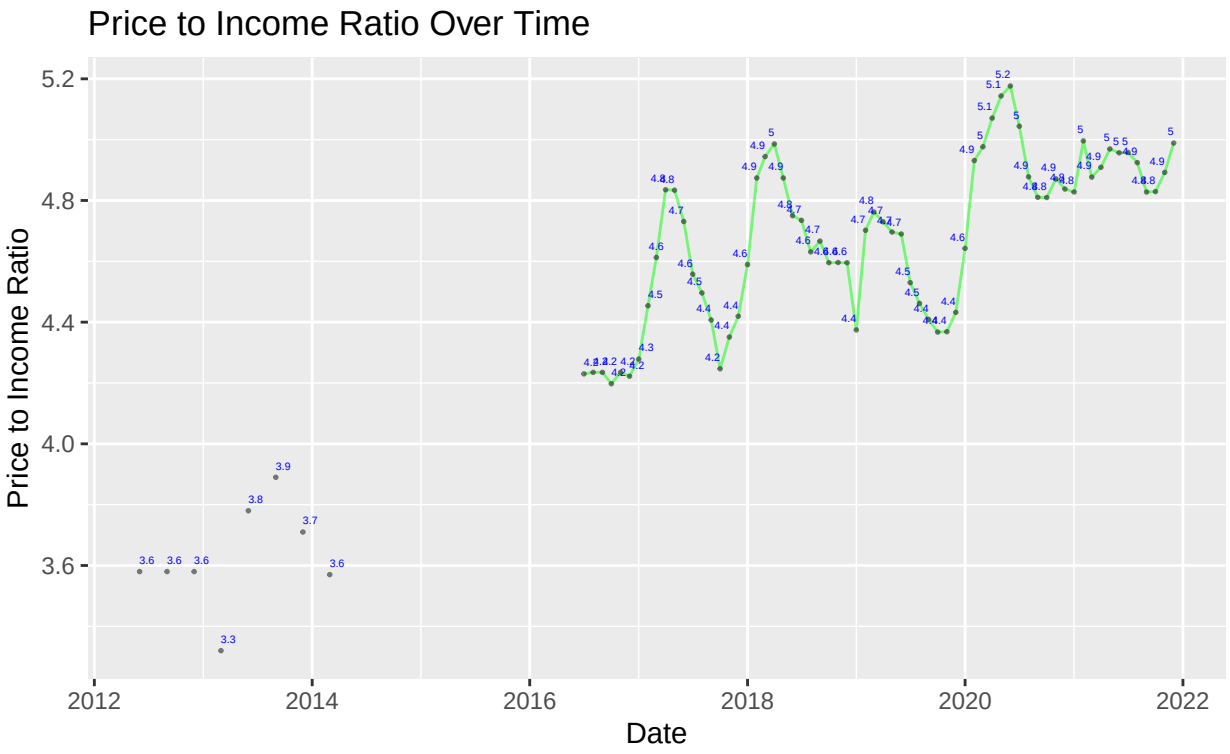
```r
seq_date = seq(from = dmy("01 06 2012"), to = dmy("01 12 2021"), by = "month")
c = c(1:115)
data = data.frame(date = seq_date, c)
df = full_join(data, df, by = "date")
```

Graph of the Price_to_income index

```r
df %>%
  ggplot(aes(x = date, y = price_to_income)) +
  geom_line(alpha = 0.5, color = "green") +
  geom_point(size = 0.3, alpha = 0.5) +
  geom_text(aes(label = round(price_to_income, 1)),
            vjust = -1, hjust = "inward",
            show.legend = FALSE, size = 1.4, color = "blue") +
  xlab("Date") +
  ylab("Price to Income Ratio") +
  ggtitle("Price to Income Ratio Over Time")
```

```
## Warning: Removed 41 rows containing missing values (`geom_point()`).
```

```
## Warning: Removed 41 rows containing missing values (`geom_text()`).
```



Price to rent ratio

```
date = c("01 Jan 2023", "01 December 2022", "01 Nov 2022", "01 October 2022", "01 September 2022", "01 A
rent = c(1429, 1665, 2124, 1423, 1361, 1194, 1005, 1067, 1015, 1051, 1039, 1014, 1006, 1032, 1070, 1138
zumper_data = data.frame(date = dmy(date), rent)
```

```
price_to_rent = full_join(zumper_data, price, by = "date") %>%
  drop_na()%>%
  mutate(annul_rent = rent * 12,
         price_to_rent = house_price / annul_rent)
```

```
seq_date = seq(from = dmy("01 06 2016"), to = dmy("01 11 2022"), by = "month")
c = c(1:78)
data = data.frame(date = seq_date, c)
price_to_rent = full_join(data, price_to_rent, by = "date")
```

```
price_to_rent %>%
  ggplot(aes(x = date, y = price_to_rent)) +
  geom_line(alpha = 0.5, color = "green") +
  geom_point(size = 0.3, alpha = 0.5) +
  geom_text(aes(label = round(price_to_rent, 1)),
            vjust = -1, hjust = "inward",
            show.legend = FALSE, size = 1.4, color = "blue") +
  xlab("Date") +
  ylab("Price to Rent Ratio") +
  ggtitle("Price to Rent Ratio Over Time")
```

```
## Warning: Removed 1 row containing missing values (`geom_line()`).
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
```

```
## Warning: Removed 1 rows containing missing values (`geom_text()`).
```