

FHFA HPI Time Series Analysis

Date: Sep 17, 2021

Wujian Xue

Content

1.Introduction.....	2
2.Data.....	3
3.Methodology.....	7
5.Reference	10
6.Appendix 1: R program.....	10
7.Appendix 2: GitHub, GitHub Desktop, PowerShell	10

“The land is the only thing in the world worth working for, worth fighting for, worth dying for, because it's the only thing that lasts.”

--Gerald O'Hara, Gone with The Wind

1.Introduction

For most people, the largest business in their life is to qualify mortgages from banks and buy houses. The Nobel price winner Robert Shiller states in his book “Irrational Exuberance” that people work for houses and the purpose of work is just to keep paying mortgages for 30 years. On average, the house price in the U.S. is around \$300,000. With 20% down payment and 3% interest rate, the loan amount is about \$240,000 and monthly payment is around \$1,300 plus other fees.

Besides these costs, buying houses maybe the simplest and the most successful investment for many people. On average, the annual growth rate for U.S housing price is 3-5%. Although the growth rate itself doesn't look exciting, considering the large principle, the profits from buying houses are still significant. For example, if a house worth \$300,000, the growth rate is 5%, then 10 years later, the target house price will be \$488,668. The \$188,668 gain from house is much easier and stabler than profit from other kind of investment such as stocks or cryptocurrencies.

Unlike the stock market, which has one price for one stock in a centralized trading platform such as New York Stock Exchange(NYSE), the housing markets are composed of lots of local markets. In the other words, 3-bed room single family houses have different prices in New York, Washington D.C, Atlanta or San Francisco. Therefore, measuring housing market is different from measuring the stock market.

However, just like the stock index, many institutions design, calculate and publish House (or Home) Price Index (HPI) to measure housing prices. The most well-known HPI is S&P/Case-Shiller indexes, which is proposed by Karl Case and Robert Shiller

and published by S&P Global Ratings. In addition, the U.S. government also creates HPI. For instance, Federal Housing Financing Agency(FHFA) publishes HPI using data from Fannie Mae and Freddie Mac since 1975.

In this project, I will analyze the HPI from FHFA, demonstrate HPI in different areas of US and forecast the future prices using time series model.

2.Data

The FHFA House Price Index is a method to measure the movement of single-family house prices. It is a weighted and repeat-sales HPI, which measures average price changes in repeat sales or refinancings on the same properties. This information is obtained by Fannie Mae or Freddie Mac since January 1975.

In this project, I will use FHFA House Price Indexes (HPIs) from Kaggle. It is located at <https://www.kaggle.com/tunguz/fhfa-house-price-indexes-hpis>. On FHFA website(<https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index.aspx>), there are more detailed HPI datasets.

Table 1 explains the columns in the dataset. Originally, the dataset has 11 variables and 114,372 observations. Among these variables, yr and period are integers. And index_nsa (non-seasonal adjusted) and index_sa (seasonal adjusted) are float. The rests are string. The examples of each variable are also listed in Table 1.

Table 1

	Variable	Type	Example
1	hpi_type	String	Traditional, non-metro, distress-free, developmental
2	hpi_flavor	String	purchase-only, all-transactions
3	frequency	String	Monthly, quarterly
4	level	String	MSA, State, USA or Census Division, Puerto Rico
5	place_name	String	Washington-Arlington-Alexandria, DC-VA-MD-WV (MSAD) New York-Jersey City-White Plains, NY-NJ (MSAD)
6	place_id	String	Zip code, state
7	yr	Integer	1975, 1976, ..., 2020
8	period	Integer	1,2,3,4,5.....,12
9	index_nsa	Float	100, 100.98, 101.22, ...
10	index_sa	Float	100, 101.06, 100.95,

Basically, the dataset contains non-seasonal adjusted and seasonal adjusted HPI time series data from 1975 to 2020. The HPI datasets can be grouped by several ways such as by state, zip code or Metropolitan Statistical Area (MSA). The original point for both non-seasonal adjusted and seasonal adjusted HPI are 100. From Figure 1 to Figure 4, HPIs in 4 MSAs are shown:

- New York-Jersey City-White Plains, NY-NJ
- Washington-Arlington-Alexandria, DC-VA-MD-WV
- Charlotte-Concord-Gastonia, NC-SC
- Francisco-San Mateo-Redwood City, CA

Figure 1

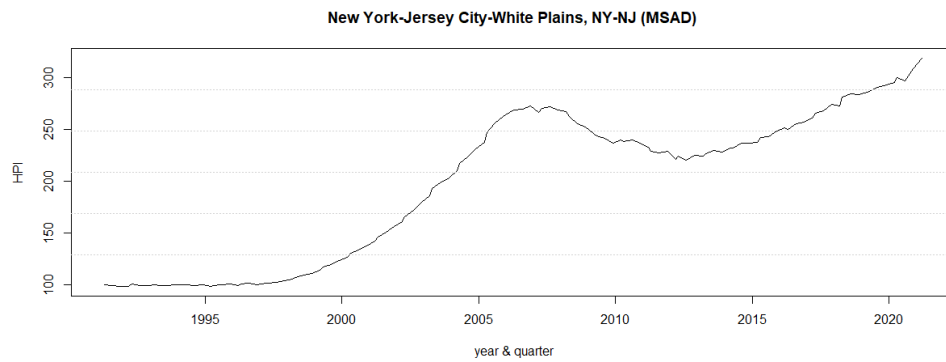


Figure 2

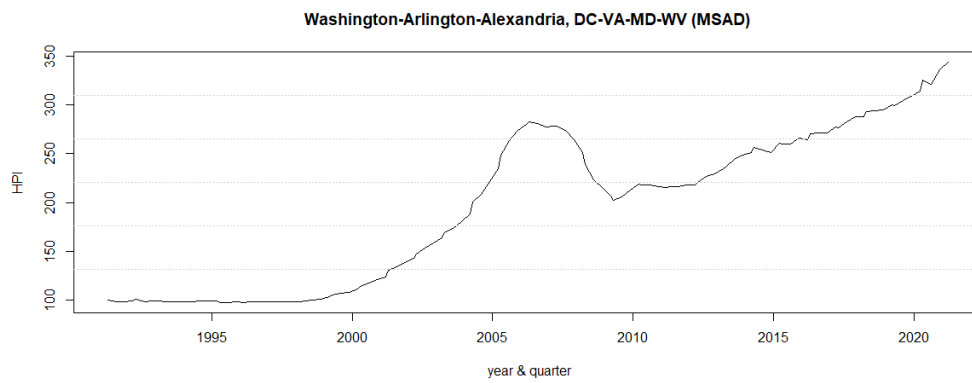


Figure 3

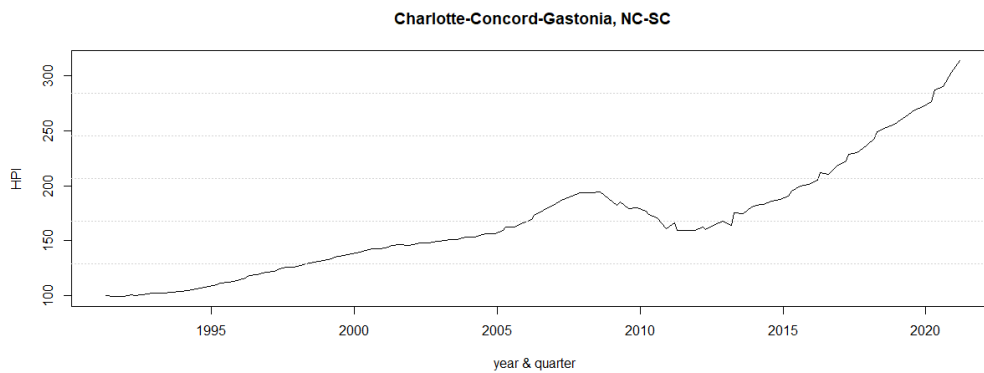
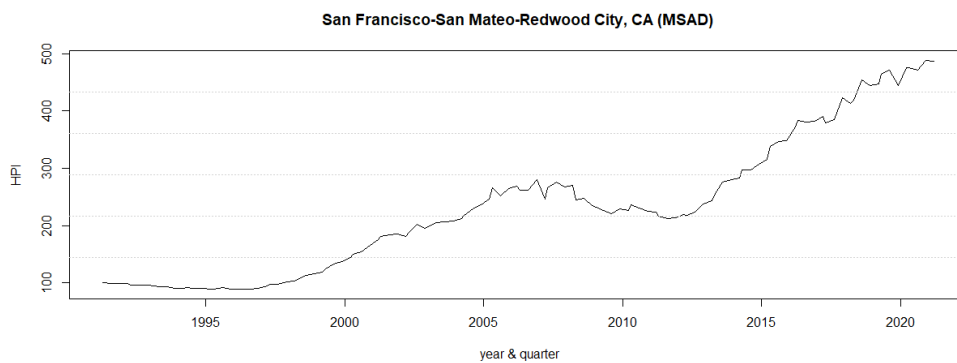


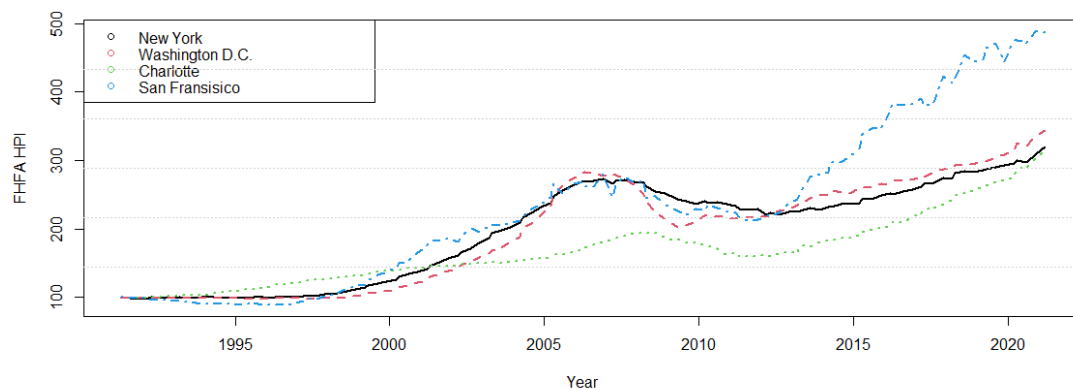
Figure 4



As you can see from these figures, the HPI increased in all the 4 MSAs from 1990 to 2020. And obviously, these 30 years can be divided into 3 periods: 1990-2000, 2001-2007 and 2008-2020. Before 2000, the HPIs are relatively stable in the 4 MSAs, although Charlotte has a little bit increasing trend. In the second period, the HPIs in all MSAs raised dramatically with 2-3 times increment. However, a new cycle started from 2008. The HPI dropped for several years and then create new high.

The HPIs for the 3 MSAs in the east coast have similar pattern. Washington D.C metropolitan area has the highest seasonal adjusted HPI at 344.05, followed by New York at 319.40 and Charlotte at 313.82. However, when it comes to the west coast, San Francisco has an HPI at 486.75. This is significantly higher than those 3 MSAs in the east coast, as shown in Figure 5.

Figure 5



Besides this visualization, I also explore the data by check missing data, computing descriptive statistics, and other methods.

```
> anyNA(hpi)
[1] TRUE
```

```
> summary(hpi)
hpi_type      hpi_flavor      frequency      level      place_name
Length:114731 Length:114731 Length:114731 Length:114731 Length:114731
Class :character Class :character Class :character Class :character Class :character
Mode  :character Mode  :character Mode  :character Mode  :character Mode  :character

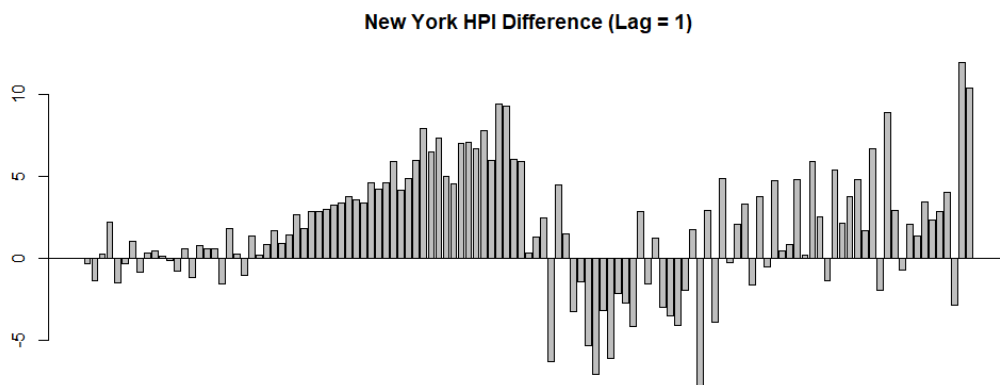
place_id      yr      period      index_nsa      index_sa
Length:114731 Min.   :1975 Min.   : 1.00 Min.   : 15.08 Min.   : 70.72
Class :character 1st Qu.:1995 1st Qu.: 2.00 1st Qu.:106.72 1st Qu.:126.17
Mode  :character Median:2003 Median: 3.00 Median:153.25 Median:175.04
Mean  :2003 Mean  : 2.63 Mean  :164.18 Mean  :182.03
3rd Qu.:2012 3rd Qu.: 4.00 3rd Qu.:200.03 3rd Qu.:219.03
Max.   :2021 Max.   :12.00 Max.   :951.26 Max.   :599.26
NA's   :76937
```

3.Methodology

In this section, I use HPI in New York as a time series example and apply ARMA model to identify and describe its structure and forecast future HPI values.

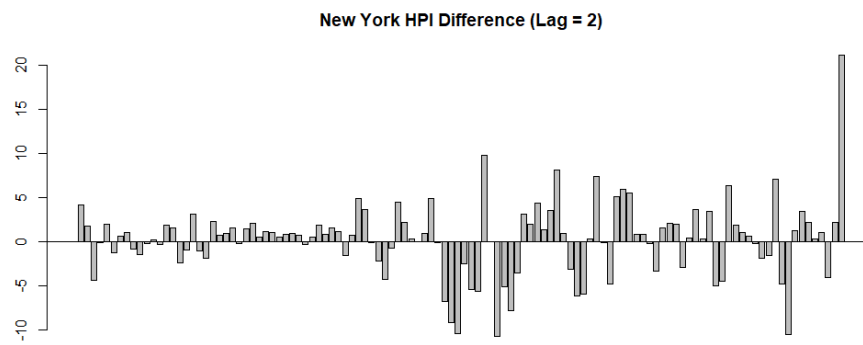
To apply an ARMA model, the dataset needs to be stationary time series. Using `diff()` function in R, the New York HPI time series is differenced once as shown in Figure 6. But it doesn't appear to be a stationary process.

Figure 6



By changing lag to 2 and difference to 2 in the `diff()` function, we can see this issue can be solved. As shown in Figure 7, the differenced time series has a constant mean closed to zero and a fairly constant variance over time. Therefore, a stationary time series has been obtained.

Figure 7



Then, the ACF and PACF for the differenced series are plotted in Figure 8 and Figure 9. As we can see from these figures, AR(1) and MA(1) can be considered.

Figure 8

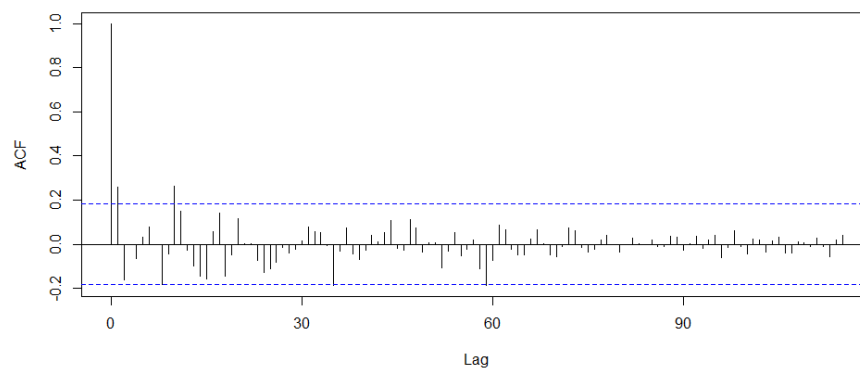
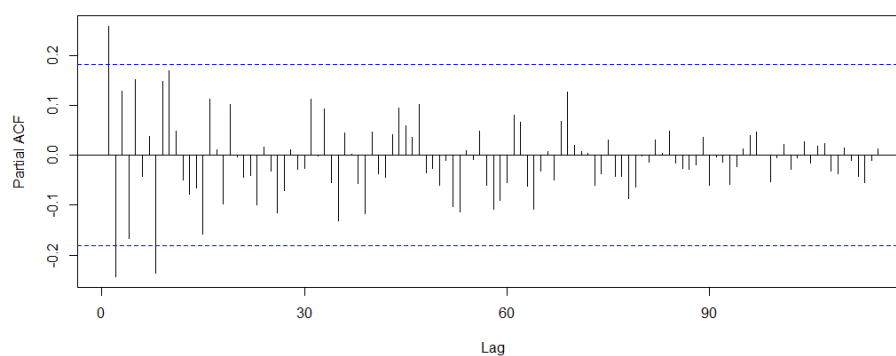


Figure 9



Using `arima()` function in R to fit this model, parameters can be estimated as follow.


```

> arima_ny_fhfa_hpi
Call: arima(x = fhfa_hpi[, 2], order = c(1, 2, 1))
Coefficients:
          ar1          ma1
      -0.4352    -0.437
s.e. 0.1112      0.114
sigma^2 estimated as 7.69: log likelihood = -288.18, aic = 582.36

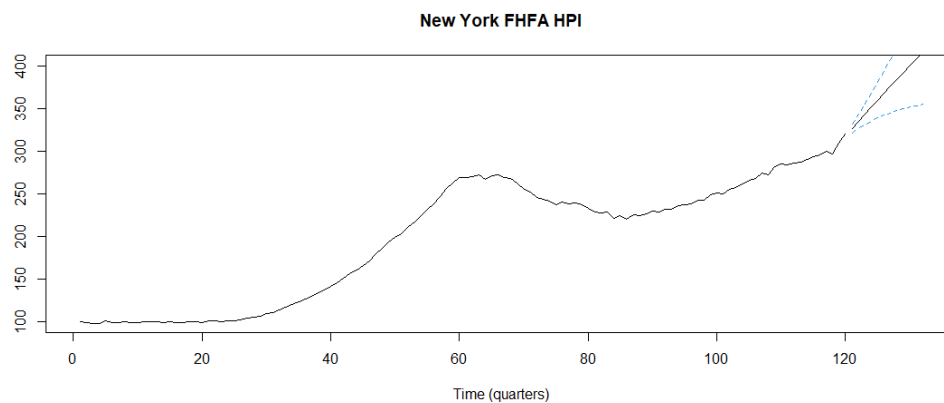
```

The next step is to forecast the future New York FHFA HPI. This can be obtained using the R predict() function. The predicted values together with lower and upper bounds at 95% confidence intervals are shown in Table 2 and Figure 10. The real New York HPI values in 2021 Q1 and Q2 announced by FHFA are also shown in Table 2.

Table 2

	Low Bound	Predict Values	FHFA Values	Upper Bound
2021Q1	321.12	326.56	329.27	332.00
2021Q2	326.93	335.12	341.43	343.31
2021Q3	330.98	343.07	-	355.16
2021Q4	335.11	351.29	-	367.46
2022Q1	338.64	359.39	-	380.14
2022Q2	341.89	367.54	-	393.19
2022Q3	344.77	375.67	-	406.57
2022Q4	347.34	383.81	-	420.28
2023Q1	349.62	391.94	-	434.27
2023Q2	351.61	400.08	-	448.56
2023Q3	353.32	408.22	-	463.11
2023Q4	354.78	416.35	-	477.93

Figure 10



Obviously, the upward trending is similar to the current housing market. The real New York HPI values in 2021 Q1 and Q2 are 329.27 and 341.43 respectively. They are closed to but different from the predicted values, although both of them are in the 95% confidence interval.

One reason results in these differences is that the data on Kaggle is dated since these data are seasonal adjusted. To solve this issue, I can use the non-seasonal adjusted data instead of seasonal adjusted HPI. Another method to solve the issue is to change the data source to FHFA, as it updates the data frequently.

5.Reference

- Aderibigbe, T. & Chi, H. (2018). Investigation of Florida Housing Prices using Predictive Time Series Model, Proceedings of the Practice and Experience on Advanced Research Computing, July 2018, Article No. 92, Pages 1–4
- Smola, A. & Vishwanathan, S.V.N. (2008). Introduction to Machine Learning, Cambridge Press.
- Shiller, R., (2015). Irrational exuberance, 3rd edition, Princeton University Press.
- Chacon, S. & Straub, B. (2014). Pro Git, <https://git-scm.com/book/en/v2>

6.Appendix 1: R program

R program for this project is located at
https://github.com/wx2123/FHFA_HPI/blob/master/FHFA_HPI.R

7.Appendix 2: GitHub, GitHub Desktop, PowerShell

To complete this project, I used Git tools such as GitHub, GitHub Desktop and PowerShell.

- My GitHub: <https://github.com/wx2123>

Search or jump to... Pull requests Issues Marketplace Explore

Overview Repositories Projects Packages

Panda Bear
wx2123
Data, Modelling and Real World!
Edit profile
1 follower • 7 following • 13

Pinned

- FHFA_HPI**
"The land is the only thing in the world worth working for, worth fighting for, worth dying for, because it's the only thing that lasts." -- Gone with the Wind
R
- Chinese_Stock_Data**
The value investment methods from superinvestors of Graham-and-Doddsville.
R
- Financial_Time_Series**
Financial time series and asset allocation for US stocks and ETFs.
Python 1 2
- Top-Investment-Books**
Wall Street classical investment books
1
- Data**
I put some data here
- R**
Statistical computing and graphics!
HTML

- Page for this project: https://github.com/wx2123/FHFA_HPI

Search or jump to... Pull requests Issues Marketplace Explore

wx2123 / FHFA_HPI Public Unwatch 1 Star 0 Fork 0

<> Code Issues Pull requests Actions Projects Wiki Security Insights Settings

master 1 branch 0 tags Go to file Add file Code

wx2123 delete junk ee88d89 30 minutes ago 46 commits

FHFA 2021Q2_HPI.pdf	Create FHFA 2021Q2_HPI.pdf	last month
FHFA HP time series analysis.docx	new commit	38 minutes ago
FHFA_HPI.R	new commit	1 hour ago
HPI_PO_metro.xls	Create HPI_PO_metro.xls	last month
HPI_master.csv	update at Aug 16, 2021	2 months ago
README.md	Update README.md	last month
gone_with_the_wind.pdf	Create gone_with_the_wind.pdf	last month
predict.xlsx	new commit	1 hour ago

About
"The land is the only thing in the world worth working for, worth fighting for, worth dying for, because it's the only thing that lasts." -- Gone with the Wind
Readme

Releases
No releases published
[Create a new release](#)

Packages
No packages published
[Publish your first package](#)

GitHub APP

10:41 65%

Panda Bear
wx2123
Data, Modelling and Real World!
New York, NY
1 follower • 7 following

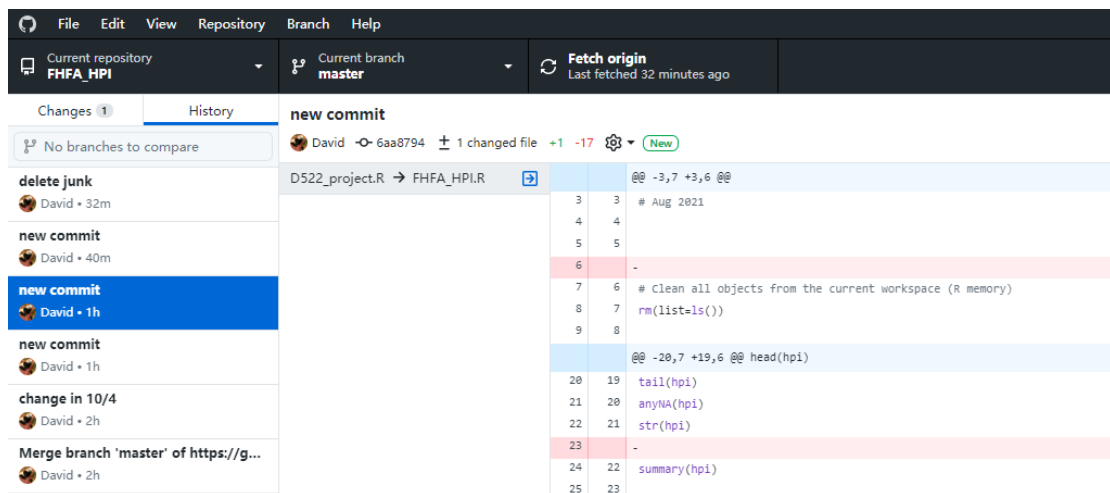
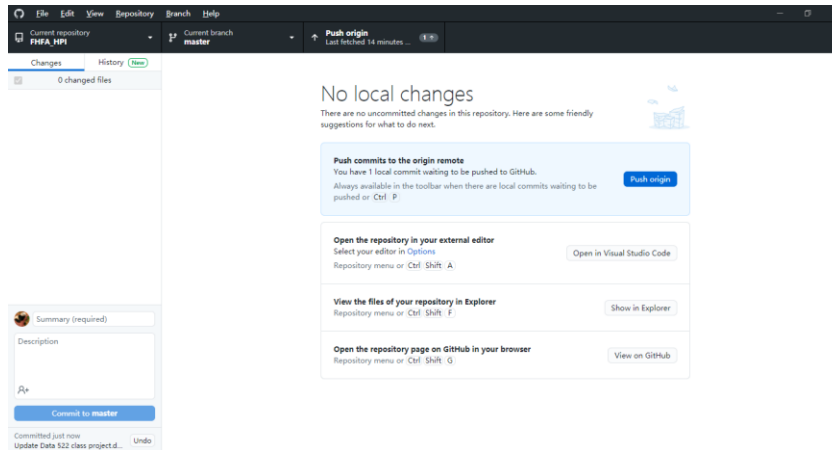
Pinned

- wx2123 FHFA_HPI "The land is the only thing in th..." 0 R
- wx2 Chinese The va 0

Repositories 10
Organizations 0
Starred 3

Home Notifications Explore Profile

● GitHub Desktop



● Git functions in PowerShell

Windows PowerShell

```
PS D:\0_0 Careers\2019\1910 UoNA\2107 Data 522> cd .\class_project
PS D:\0_0 Careers\2019\1910 UoNA\2107 Data 522\class_project> ls

    目录: D:\0_0 Careers\2019\1910 UoNA\2107 Data 522\class_project

Mode                LastWriteTime         Length Name
----                -
-a-----         2021/7/13         17:30      189841 Big Data Project Abstract.docx
-a-----         2021/9/2          22:47         2771 D522_project.R
-a-----         2021/9/2          22:48         856273 Data 522 class project.docx
-a-----         2021/8/28         23:45         335799 Data 522 mid point project presentation update.docx
-a-----         2021/9/1          16:43         2092462 FHFA 2021Q2_HPI.pdf
-a-----         2021/9/1          17:17         2925889 gone_with_the_wind.pdf
-a-----         2021/7/13         21:38        10088405 HPI_master.csv
-a-----         2021/8/16         17:10          51 mycode.ipynb
-a-----         2021/8/16         17:10          30 mycode2.ipynb
-a-----         2021/8/16         17:10         761 mycode4.ipynb
-a-----         2021/8/16         17:10        1309 mycode5.ipynb
-a-----         2021/9/2          22:26          100 README.md

PS D:\0_0 Careers\2019\1910 UoNA\2107 Data 522\class_project> git status
On branch master
Your branch is up to date with 'origin/master'.

nothing to commit, working tree clean
PS D:\0_0 Careers\2019\1910 UoNA\2107 Data 522\class_project> git diff
PS D:\0_0 Careers\2019\1910 UoNA\2107 Data 522\class_project> git add gone_with_the_wind.pdf
PS D:\0_0 Careers\2019\1910 UoNA\2107 Data 522\class_project> git commit
On branch master
Your branch is up to date with 'origin/master'.
```