Capstone-Data698

Dieudonne

11/23/2016

Introduction

The HPI is a broad measure of the movement of single-family house prices. The HPI is a weighted, repeat-sales index, meaning that it measures average price changes in repeat sales or refinancings on the same properties. This information is obtained by reviewing repeat mortgage transactions on single-family properties whose mortgages have been purchased or securitized by Fannie Mae or Freddie Mac since January 1975.

The HPI serves as a timely, accurate indicator of house price trends at various geographic levels. Because of the breadth of the sample, it provides more information than is available in other house price indexes. It also provides housing economists with an improved analytical tool that is useful for estimating changes in the rates of mortgage defaults, prepayments and housing affordability in specific geographic areas.

The HPI includes house price figures for the nine Census Bureau divisions, for the 50 states and the District of Columbia, and for Metropolitan Statistical Areas (MSAs) and Divisions.Divisons are below

Pacific:Hawaii, Alaska, Washington, Oregon, California

Mountain:Montana, Idaho, Wyoming, Nevada, Utah, Colorado, Arizona, New Mexico

West North Central:North Dakota, South Dakota, Minnesota, Nebraska, Iowa, Kansas, Missouri

West South Central:Oklahoma, Arkansas, Texas, Louisiana

East North Central:Michigan, Wisconsin, Illinois, Indiana, Ohio

East South Central:Kentucky, Tennessee, Mississippi, Alabama

New England:Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut

Middle Atlantic:New York, New Jersey, Pennsylvania

South Atlantic:Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida

In this paper I study the Home Price Index (HPI) for all states in the USA using R language ,I applied the analytical tools learned during my MSDA courseworks, I used shiny to visualize the 25 largest cities in the US.

```r
require(ggplot2) || install.packages("ggplot2")
```

```
## [1] TRUE
```

```r
require(reshape2) || install.packages("reshape2")
```

```
## [1] TRUE
```

```r
require(dplyr) || install.packages("dplyr")
```

```
## [1] TRUE
```

```r
require(GGally) || install.packages("GGally")
```

```
## [1] TRUE
```

```r
require(maps) || install.packages("maps")
```

```
## [1] TRUE
```

```r
require(scales) || install.packages("scales")
```

```
## [1] TRUE
```

```r
require(ggmap) || install.packages("ggmap")
```

```
## [1] TRUE
```

```r
require(lubridate) || install.packages("lubridate")
```

```
## [1] TRUE
```

The data for my study are provided by the Federal Housing Financial Agency and Bureau of Economy Analysis

US housing price index summary

http://www.fhfa.gov/DataTools/Downloads/Documents/HPI/HPI_PO_summary.xls

US housing price index by state

http://www.fhfa.gov/DataTools/Downloads/Documents/HPI/HPI_PO_state.txt

US housing price index for metropolitan areas

http://www.fhfa.gov/DataTools/Downloads/Documents/HPI/HPI_PO_metro.txt

Data for the largest cities in USA

http://travel.forumsee.com/a/m/s/p12-22806-0310459--metropolitan-gdp-billions.html www.city-data.com

Below are those cities #Chicago IL, Houston TX, Los Angeles CA, New York NY, DC DC, Atlanta GA, Boston MA, Dallas TX, Philadelphia PA, San Francisco CA, Detroit MI, Miami FL, Minneapolis MN, Phoenix AZ, Seattle WA, Baltimore MD, Denver CO, Portland OR, San Diego CA, San Jose CA, Charlotte NC, Kansas City KS,Pittsburgh PA, Saint Louis MO, Tampa FL

US personal income by Metropolitan Areas

http://www.bea.gov/regional/index.htm) http://www.bea.gov/iTable/iTable.cfm?reqid=70&step=1&isuri=1&acrdn=3#reqid=70&step=1&isuri=1&7028=30&7040=-1&7083=percentchange&7029=49&7022=49&7023=7&7024=non-industry&7025=5&7026=16740,28140,38300,41180,45300&7027=2012,2011,2010,2009,2008&7001=749&7031=5&7090=70&7033=-1"

On the website Bureau of Economic Analysis, choose Interactive Data tab and select Regional Data GDP and Personal Income on the right panel.Then choose Local Area Personal Income: Personal income, per capita personal income, and population (CA1-3) and then select Metropolitan Statistical Area. Next, choose 25 biggest cities that we are interested in and choose the unit of measurement: percent change from proceeding period. Choose Per Capita Personal Income (dollars) for statistic.

Data Preprocessing and Cleaning

HPI data for USA http://www.fhfa.gov/DataTools/Downloads/Documents/HPI/HPI_PO_summary.xls I Downloaded the summary xls file, into Excel to remove unnecessary information and save as csv file. In the dataset I only kept seasonally adjusted Purchase Only Index , HPI changes over previous 4 quarters, and HPI changes over previous quarter. The missing value in the dataset is filled with zeros.

```
us<-read.csv("dataset/USAHPI.csv")
colnames(us)[4:6]<-c("PurchaseOnlyIndex","change-Prev4Qt","change-PrevQtr")
```

HPI data for 51 different US states http://www.fhfa.gov/DataTools/Downloads/Documents/HPI/HPI_PO_state.txt

```
require(data.table)
STATES<-read.table("https://www.fhfa.gov/DataTools/Downloads/Documents/HPI/HPI_PO_state.txt",fill=TRUE,h
STATES<-select(STATES,state,yr,qtr,index_sa)
colnames(STATES)[c(1:4)]<-c("Region","Year","Quarter","PurchaseOnlyIndex")
print(paste("Unique Regions in the state Dataset:",length(unique(STATES$Region)),"areas"))
```

```
## [1] "Unique Regions in the state Dataset: 51 areas"
```

Then, we can import US state and city data for further plotting on US map. One thing to pay attention to is that there are no DC and USA entries in usmap data while in STATES data, region for DC appears. We have to add these two parts into US map data. This is because my preliminary study shows that there is one outlier whose HPI index soars during the years, which is identified exactly as DC.

US map and US cities data

```
#Import US state and cities Data and put DC and USA into an independent class
data("state")
data(us.cities)
class<-data.frame(state.abb,state.region,state.name,state.division)
colnames(class)<-c("Region","Class","Name","Division")
class<-rbind(class, data.frame(Region=c("USA","DC"), Class=c("USA","DC"), Name = c("USA","DC"),Division=
```

HPI for 100 metropolitan areas Select Q4 of recent five years HPI in 100 metropolitan cities from FHFA website and further process the data.Since we are only interested in the 25 biggest cities, we need to grep 25 cities out of 100. Since the name for each large city is quite messy, separating the city and state names and using grepl(income$country, metro2$name) does not work very well,so I selected the data of 25 large cities manually as long as there are not too much laborous work. The cleaned dataset is HPI-25metro-clean.csv.

```
library(pander)
metro<-read.table("https://www.fhfa.gov/DataTools/Downloads/Documents/HPI/HPI_PO_metro.txt",fill=TRUE,he
metro<-select(metro,metro_name,yr,qtr,index_sa)
colnames(metro)<-c("name","year","quarter","HPI")
metrosubset<-filter(metro,year %in% c(2006,2007,2008,2009,2010,2011,2012,2013,2014,2015,2016), quarter==
metro2<-dcast(metrosubset,name~year+quarter,value.var="HPI")
pander::pander(head(metro2))
```

Table 1: Table continues below

name	2006_4	2007_4	2008_4	2009_4	2010_4
Akron, OH	172.9	167.2	155.3	158.7	149.9
Albany-Schenectady-Troy, NY	182.8	183	179.6	176.4	177.9
Albuquerque, NM	236.4	234.9	221.4	215.1	201.7
Allentown-Bethlehem-Easton, PA-NJ	197.3	199	184	171.2	162

name	2006_4	2007_4	2008_4	2009_4	2010_4
Anaheim-Santa Ana-Irvine, CA (MSAD)	282.9	245.8	201.6	209.9	203.4
Atlanta-Sandy Springs-Roswell, GA	198.6	193.6	170.5	165.4	145.4

2011_4	2012_4	2013_4	2014_4	2015_4
141.6	154.8	157.9	160.6	170.8
171.7	173.4	177.1	179.7	181.3
189.3	195.2	201.5	209.4	211.9
152.8	156.6	161.2	163.2	164.5
197.8	212.4	244.2	257.7	276.7
143.8	158.7	180.2	192.4	209.5

Summary of House Price Index(Purchase Only Index)

A summary of House Price Index(Purchase Only Index) for the whole country shows that the maximum of HPI index is 225.2 and the median is 160.9. (where index for 1991-Quarter1 is set to 100)

```
pander::pander(summary(us$PurchaseOnlyIndex))
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
100	116	160.9	158.3	192.6	225.2

Analysis of HPI of 50 US states and DC

First, I used package lubridate to deal with Dates.Year and Quarter were switched into DATE. Each quarter is represented by the first day of the beginning of the month in the quarter, eg.2007-01-01 represents the first quarter, 2007-04-01 represents the second quarter, 2007-07-01 represents the third quarter, and 2007-10-01 represents the fourth quarter. Then I averaged over 4 quarters of each year and generated a parallel coordinate plot according to years as well as making a time series plot for all quarters.

```
#Get the average  Index for Each Year for all states (Average over 4 Quarters)
#Swith Year and Quarter Into Date
STATES$Date<-paste(STATES$Year,'-0',STATES$Quarter*3-2,'-01',sep="")
STATES$Date[seq(4,dim(STATES)[1],by=4)]<-paste(STATES$Year[seq(4,dim(STATES)[1],by=4)],'-',
                                              STATES$Quarter[seq(4,dim(STATES)[1],by=4)]*3-2,'-01',sep=
STATES$Date<-as.Date(STATES$Date)
STATES.merge<-merge(STATES,class,id="Region") #not used here, but below
# Time series by quarters
df.states<-dcast(select(STATES,Region,Date,PurchaseOnlyIndex),
       Region~Date,value.var="PurchaseOnlyIndex")
newdf.states<-merge(df.states,class,id="Region")
# Averaged by each year
df.states2<-dcast(select(STATES,Region,Year,Quarter,PurchaseOnlyIndex),
           Region~Year,mean,value.var="PurchaseOnlyIndex")
df.states2<-merge(df.states2,class,id="Region")
pander::pander(df.states2)
```

Table 4: Table continues below

Region	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
AK	100.9	103.5	107.3	111.2	116.2	121.1	124.2	128.5	132.3	135.7
AL	101.5	105.7	110.7	116	119.9	124.7	128.4	133	137.8	141.9
AR	101.1	104.3	109.9	116.9	121.6	125.3	128.1	131	135.6	139.6
AZ	100.4	102.5	106.3	113.1	119.6	125	129.5	135.9	144.2	152.5
CA	99.33	97.31	91.74	87.84	85.58	84.88	86.83	94.61	104.3	117.4
CO	101	109.1	121.6	136.5	144.9	152.8	160.2	170.5	187.2	208.5
CT	97.52	95.56	91.76	91.62	90.69	91.06	92.25	96.68	104.9	114.4
DC	98.94	99.43	94.58	95.5	91.31	94.27	92.69	102.2	113.1	130.3
DE	100	100.2	99.05	99.26	99.61	100.2	101.2	104.6	110.6	117.8
FL	100.2	101.6	104	107.1	109.7	112	114.7	119.5	126	135.4
GA	100.1	102.1	104.6	108.3	112.8	117.8	122.8	129.9	139.1	148.4
HI	98.78	101	101	99.22	95.8	90.62	82.72	83.43	83.75	90.28
IA	101.2	106.4	113.9	120.8	126.4	131.9	136	142.6	149.6	155.9
ID	102.2	110.3	120.6	129.9	135.4	137.9	140.3	143.5	147.9	151.9
IL	100.9	104.8	109.1	114.3	117.8	121.4	123.7	127.4	134	142.3
IN	100.3	104	108.9	114.1	119.3	124.3	127.7	132	136.8	141.7
KS	99.8	102.6	107.6	114.6	120.7	126	130.6	138.3	146.1	152.2
KY	100.7	104.8	109.9	116.2	121.1	126.1	131	135.9	143	149.2
LA	102.4	107.9	114.5	121.8	127.1	133.6	138.5	145.1	150.6	155.9
MA	98.05	96.83	96.02	97.46	99.15	102.9	107.7	117.2	130.7	149
MD	100.3	102	101.7	102.1	101.6	102.4	102.9	105.6	111.3	119
ME	100.8	100.7	97.47	97.83	98.23	101.6	103.3	109	117.5	128
MI	101.1	104.7	107.6	113.2	121.7	131.6	140.1	149.1	159.4	170.2
MN	99.62	102.9	107.7	112.7	116.8	122.5	127.3	135.2	148.6	165.6
MO	100.8	103.3	106.7	112.4	117.2	122.2	126.2	131.9	139.4	147.1
MS	99.8	103.8	107.3	113.6	118.4	122.7	126.6	131.7	137.4	141.4
MT	105	115.6	129.8	142.6	150.7	156.8	160.6	164	169.7	176.6
NC	100.3	103	106.2	111.9	117.1	122.7	127.9	133	138.9	144.2
ND	100.1	103.8	111	118	122.4	125.3	128.8	133.4	136.6	140.1
NE	101.3	108.2	115.8	122.2	128.5	134.8	141.6	149.1	155.8	160.7
NH	97.94	94.37	92.5	93.96	94.83	97.66	102.2	110.2	121.3	138.1
NJ	99.5	100.9	101.3	102.1	101.8	102.5	104.1	108.8	116.4	128.1
NM	101.4	107.9	116.6	129.2	136	138.4	139.5	141.4	144.7	145.7
NV	100.3	102.9	105.2	108.9	112.5	114.8	117.2	118.4	121.9	126.1
NY	99.45	100.8	100.3	99.2	98.58	99.18	100.4	105	113.3	124
OH	101	105.6	110.3	115.7	120.7	126	129.7	134.5	140.8	146.3
OK	100.7	103.2	108.2	113.5	116.6	120.6	124	129.5	136.1	141.8
OR	102.7	111.5	121.3	134.1	145.3	155.4	164.1	169.4	175.7	181.4
PA	100.1	102	103.2	104.8	104.7	105.8	107	109.4	113.7	119
RI	97.39	95.35	92.95	92.51	91.86	91.04	91.7	95.57	102.6	114.3
SC	101	104	106.6	110.4	114.1	119	123.4	129.2	136.5	143.2
SD	101.7	108.3	116.2	124.6	128.8	134.9	139.3	144.7	151.2	158.2
TN	100.5	103.3	107.3	113.6	120.1	126.1	130.8	135.9	141.6	145.9
TX	100	102.4	105.7	109.4	111.7	114.3	116.8	122.9	130.6	139.8
UT	101.5	109.7	125.2	145.7	159.3	171.9	178.2	184.4	189.1	193.2
VA	99.63	101	101.8	104	105.5	107.4	109.7	112.9	119.2	127.7
VT	98.78	100	100.7	101.2	99.84	102.8	101.8	106	111.8	121.7
WA	101.1	105.6	110.8	117.2	119.7	122	127.2	136.1	144.6	152
WI	101.9	108.5	117	125.7	131.2	136.2	140.6	146.3	154.3	163.6
WV	100.7	106	111.9	118.7	124	127.3	129.5	132.7	136	137.8
WY	103.1	108.9	117.2	130.4	139.8	145.3	148.9	153.5	158.1	165

Region	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000

Table 5: Table continues below

2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
144.3	153.6	164.7	180.9	200.8	216.6	223.9	222.2	218.2	219.8	224
145.8	150.5	156.9	164.5	176.6	192.1	200.3	195.7	191.4	180.5	172.1
144.5	150.2	158.2	168.9	180.2	190.4	194.2	188.4	185.4	178.5	176.9
161.9	171.6	185.9	213	277.2	318.1	308.3	251.8	204	182.5	166.2
133	152.2	178.6	221	269.9	281.4	256.8	190.5	165.6	164.7	153.9
227.4	236.4	242.3	251.6	264.3	273.7	275.8	265.2	264	261.4	254.3
125.6	139.6	154.6	171.8	189.3	196.4	196.5	185.9	176.1	169.5	164.4
150.5	179.8	207.6	255.3	311	327.8	338.6	322.6	313.2	323.1	331.2
127.5	139.7	154	175.3	199.4	217.9	218.7	206.3	197.8	187.9	175.7
149.3	166.5	188.1	222.1	278.7	310.4	295	229.4	189.5	178.7	167.8
156.5	163.4	169.6	175.8	186.2	196.2	199.7	186.2	174.3	160.8	149.5
98.67	108.8	125.8	157.1	194.8	212	212.7	202.8	185.3	176.4	167.3
161.5	167.3	173.7	181.2	188.4	195	198.6	196.9	195	193.9	191.2
157.4	162.2	170.8	186.6	214	247.7	261.3	247.7	228.3	200.9	182
152.1	162.2	173.7	185.8	198.6	208.9	210.7	198	184.6	177.4	167.1
145.3	148.8	153.1	158.1	163	166.9	168.4	162.6	159.2	157.5	156.4
158.9	164.9	171.3	178.6	185.5	193.5	198.8	196.1	194.6	190.8	186.1
154.2	159.4	166.5	174.4	181.8	188.7	191.9	190.1	188.2	186.9	183.1
161.5	168	177.1	187.9	202	224.4	234.5	230.9	228.1	227.2	222.3
170.7	193.1	214.2	235.8	251.1	246	236.9	223.3	218.2	215.5	210.4
130.9	148.5	169.7	201.2	241.5	263.8	266.2	235.8	218.2	210.1	201.4
142.3	159.1	176.7	195.9	215.4	219.6	220	213.3	208.1	202.5	198.8
178.9	186.1	192.4	198.4	201.4	195.9	183.2	161.6	152.9	145.3	141
184.2	201.5	218.7	234.6	247.9	252.2	247.8	227.9	217.3	209.4	197.8
155.6	163.4	172.6	182.6	193.1	201.7	203.9	195	190.4	184.3	176.6
145.4	149.4	154.3	160.6	171.3	186.4	194.3	188.7	180.4	175.1	173.1
186.9	198.6	215.5	236.5	264.3	294.1	312.6	310.3	301	290.1	283.2
149	153.3	158.3	165.6	177.3	191.6	200.9	198.4	193.4	183.8	175.3
144.8	152.5	162	173	186.8	198.4	207.5	212.5	215.6	221.1	229.4
165.2	171.1	178	185.4	191.7	196.9	198.7	192.7	193.3	191.1	189.7
157.5	177.2	197	216.3	234.1	234.6	229.1	212.3	204	196.8	189.5
143.2	163.6	186.3	213.3	243.5	257.9	255.1	238.5	224.8	218.9	207.4
150.3	157.3	167.1	180.7	203.8	230.7	242.1	236.2	223.3	214.1	202.8
134.6	145	163.1	211.2	256.8	271.1	252.7	190.1	141.1	129.4	114.9
136.6	152.5	170.8	189.8	207.3	216.4	217.6	213.2	206.5	204.8	199.5
151.6	157.1	162.8	168.3	173.1	174.8	172.3	162.8	157.7	154.4	148.4
147.3	152.7	158.5	164.8	173.5	183.3	192.2	192	193.1	192.8	188.9
190	200.3	214.7	237.5	277.1	319.3	335.5	314.9	286.5	265.7	245.9
126.5	136.5	148.8	164.7	182.5	195.6	200.5	195.5	189.5	187.3	182.8
130.6	155.2	182.3	212.2	234.4	236.5	225.2	205.2	193.9	184.9	176.3
148.4	153.6	158.2	167.1	179	192.5	200.4	196.7	191.1	181.3	172.3
164.9	171.1	179.5	189.5	201.5	210.9	218	221.5	221	218.3	219
149.6	154	160.3	168.4	180.2	193.7	201.9	196.6	189.9	183.9	179.4
147	151.9	155.6	160.5	168.5	179.2	188.6	189.4	188.3	188.1	185.9
196.8	200.1	205.2	217.1	241.6	282.2	316.4	301.4	269.5	253.8	238.4
139.3	152.1	168.6	192.3	223	243.5	246.1	225.2	215.9	210.9	204.5
133	144.7	156.4	178.2	199.9	212.2	215.9	211.8	208.5	203.1	202.9

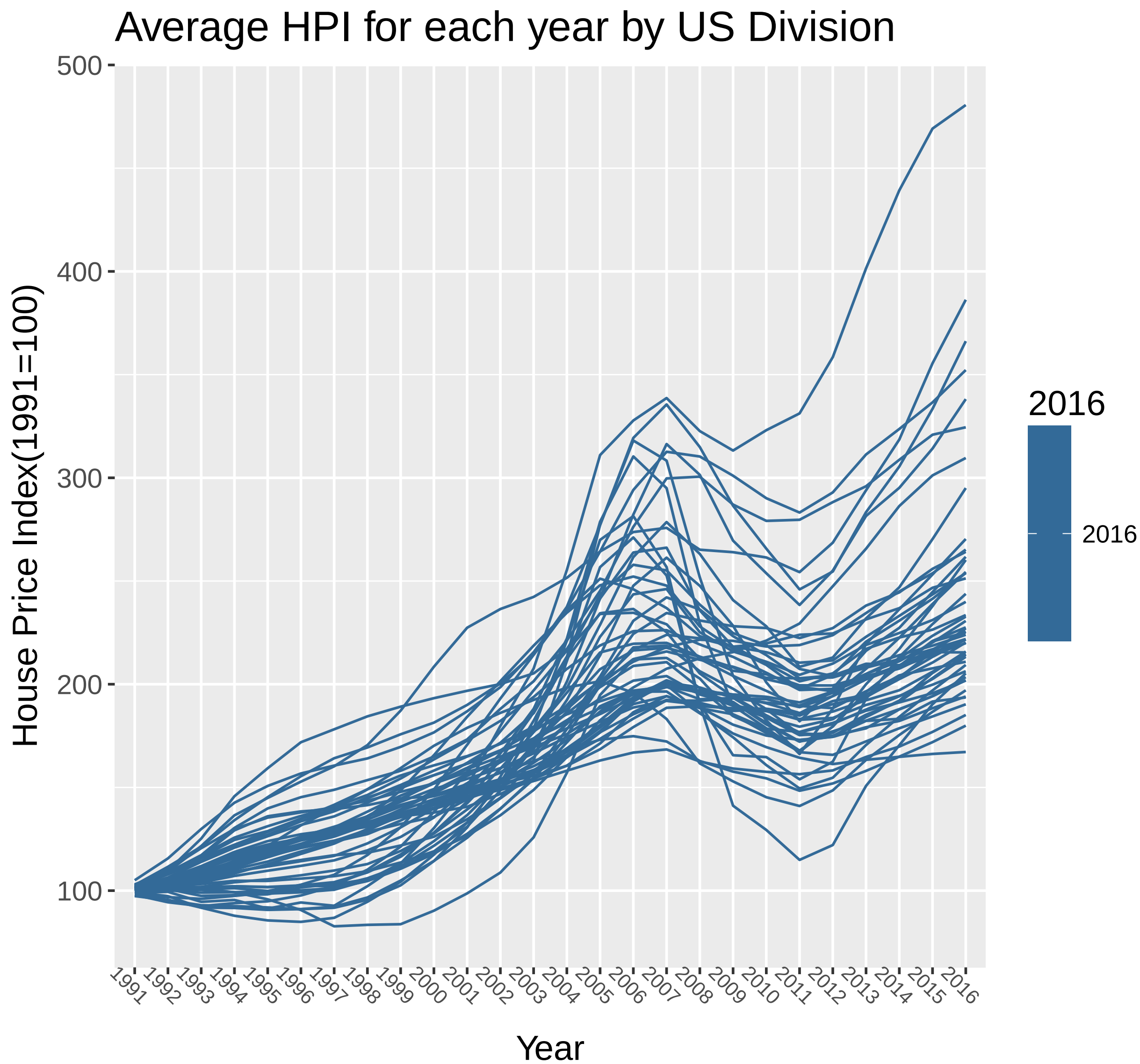
2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
159.5	167.7	178.4	198.2	228.8	261.6	278.6	263.1	240.7	228.1	208.4
172.6	182.2	193.8	207.4	218.8	225.7	226.2	219.3	213.2	205.6	197.2
140.6	146.9	153.9	164.8	176.7	185.4	192.4	191	187.1	187.9	184.8
173.2	188.1	201.4	221.1	245.2	276.2	299.7	300.6	287	279.1	279.7

2012	2013	2014	2015	2016	Class	Name	Division
224.6	231.4	236.8	246.7	251.3	West	Alaska	Pacific
177.2	182.4	187.9	194.2	201.9	South	Alabama	East South Central
181.3	187.2	191.2	195.7	201.8	South	Arkansas	West South Central
188.2	219.5	236.3	253.2	270.4	West	Arizona	Mountain
162.5	192.9	213.2	229.1	243.8	West	California	Pacific
268.6	294	318.5	355.5	386.3	West	Colorado	Mountain
161.3	163.4	164.8	166.2	167.2	Northeast	Connecticut	New England
358.5	401.5	439.3	469.2	480.6	DC	DC	DC
175	182.2	183.2	191.6	193.6	South	Delaware	South Atlantic
179.5	200	217.1	237.8	260.2	South	Florida	South Atlantic
154.7	170.6	184.2	197	209.4	South	Georgia	South Atlantic
179.5	196.2	208.8	221	227.5	West	Hawaii	Pacific
196.5	203.2	209.1	216.8	225.6	North Central	Iowa	West North Central
198	216.2	227.3	244.7	261.7	West	Idaho	Mountain
165.8	172.3	178.7	184.4	190.3	North Central	Illinois	East North Central
158.2	164.8	169.8	176.9	185.1	North Central	Indiana	East North Central
190.3	193.9	202.5	210.5	220.2	North Central	Kansas	West North Central
186.8	192.2	197.1	206	214.7	South	Kentucky	East South Central
227.2	238.1	244.5	255.9	264.1	South	Louisiana	West South Central
211.5	223	233.1	243.5	253.8	Northeast	Massachusetts	New England
205.3	217.1	222.6	226.4	233.5	South	Maryland	South Atlantic
196.6	203.5	207.6	215.2	221.8	Northeast	Maine	New England
148.6	163.3	175.4	186.6	197.1	North Central	Michigan	East North Central
204.4	220.6	230.8	241.4	253.9	North Central	Minnesota	West North Central
180.7	187.7	194.2	202.5	213.3	North Central	Missouri	West North Central
174.5	179.2	182.2	188.4	194.1	South	Mississippi	East South Central
292.9	311.3	323.7	336.5	352.2	West	Montana	Mountain
176.5	185.2	191.4	203	214.3	South	North Carolina	South Atlantic
247.5	265.7	286.3	301.2	309.6	North Central	North Dakota	West North Central
195.9	205	212.5	223.5	232.7	North Central	Nebraska	West North Central

2012	2013	2014	2015	2016	Class	Name	Division
188.4	196.6	203.8	213.3	221.7	Northeast	New Hampshire	New England
203.9	208.8	214.1	218.7	223.9	Northeast	New Jersey	Middle Atlantic
203.3	208	212.1	217	221.9	West	New Mexico	Mountain
122.1	150.8	170.4	190.3	205.3	West	Nevada	Mountain
199.4	203.9	208.1	214.2	219.9	Northeast	New York	Middle Atlantic
151.9	158	164.9	171.9	179.9	North Central	Ohio	East North Central
194.2	202.3	208.4	220.7	225.2	South	Oklahoma	West South Central
254.6	283.4	305.5	333.4	366.2	West	Oregon	Pacific
183.7	190	194.4	199.8	206.3	Northeast	Pennsylvania	Middle Atlantic
175.1	178.6	187.8	194.2	203.4	Northeast	Rhode Island	New England
174.7	184.1	192.1	205.5	215.8	South	South Carolina	South Atlantic
223.8	234.5	245	253.7	265.2	North Central	South Dakota	West North Central
183.1	193.7	203.2	215.7	227.1	South	Tennessee	East South Central
194.3	207.4	222.5	238.9	254.4	South	Texas	West South Central
255	281.5	295.2	314.2	338.1	West	Utah	Mountain
210	219	224.8	231.1	239.9	South	Virginia	South Atlantic
204.2	209.9	209.2	215.6	215.4	Northeast	Vermont	New England
212.9	231.9	246.9	270.3	295	West	Washington	Pacific
197.7	205	211.1	220.2	230.7	North Central	Wisconsin	East North Central
191.9	194.8	204.3	207.7	210.9	South	West Virginia	South Atlantic
288.2	295.9	308.7	320.9	324.5	West	Wyoming	Mountain

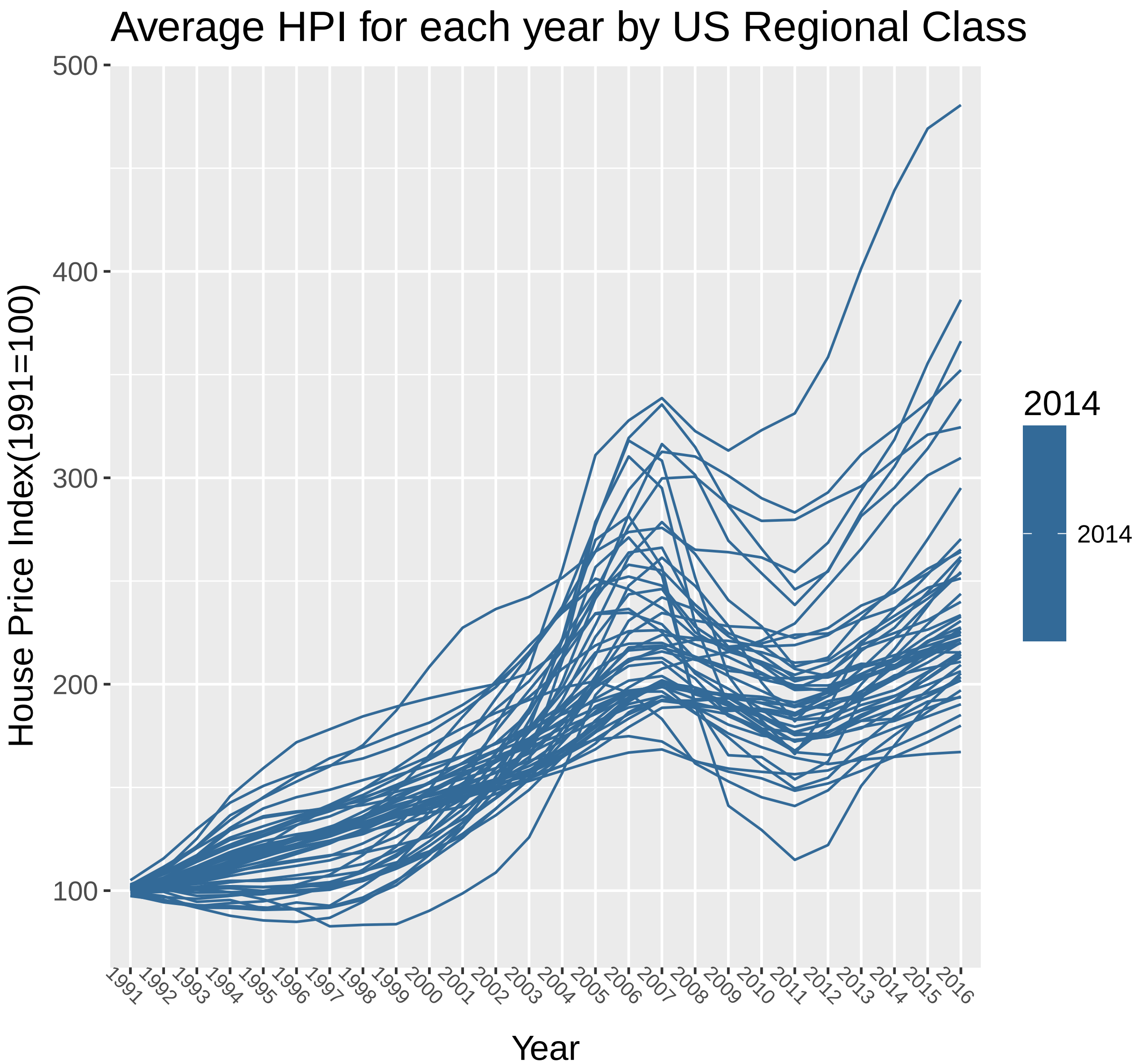
Next we can make a parallel Coordinate Plot of Purchase Only Index for each state in each year, grouped by US divisions and regional classes. From the result we can observe that HPI rises slowly from 1991 to 2000, and then increases exponentially since 2000 and reaches peak in 2008. After the year 2008, HPI drastically fell down and was fairly low in 2011-2012. However, we could also notice that from the year 2013, it seems that the house maket is thriving and perhaps it is turning over. This is a positive sign for the turn around of economy, although we need more data to determine whether the house market is recovering for sure. Moreover, the plot also shows that DC has not been affected by economy crisis significantly during 2007-2008, The HPI fell down a little bit and then went up immediately. HPI for DC is extremely high during these years compared to other regions in US based on the fact that the HPI is always higherest among all other areas since 2005.

```
#Sum of Each Year-by Division
ggparcoord(df.states2, columns = c(2:(length(df.states2)-3)), groupColumn = 27, scale = "globalminmax",
           missing = "exclude") + theme(legend.position = "right", text=element_text(size=14),legend.te
```



We can group the time series plot by different clustering criteria. Aftering dividing the US states roughly into 5 regions, we could easily find that on the whole west part has higer HPI index than other areas despite that DC stands out.

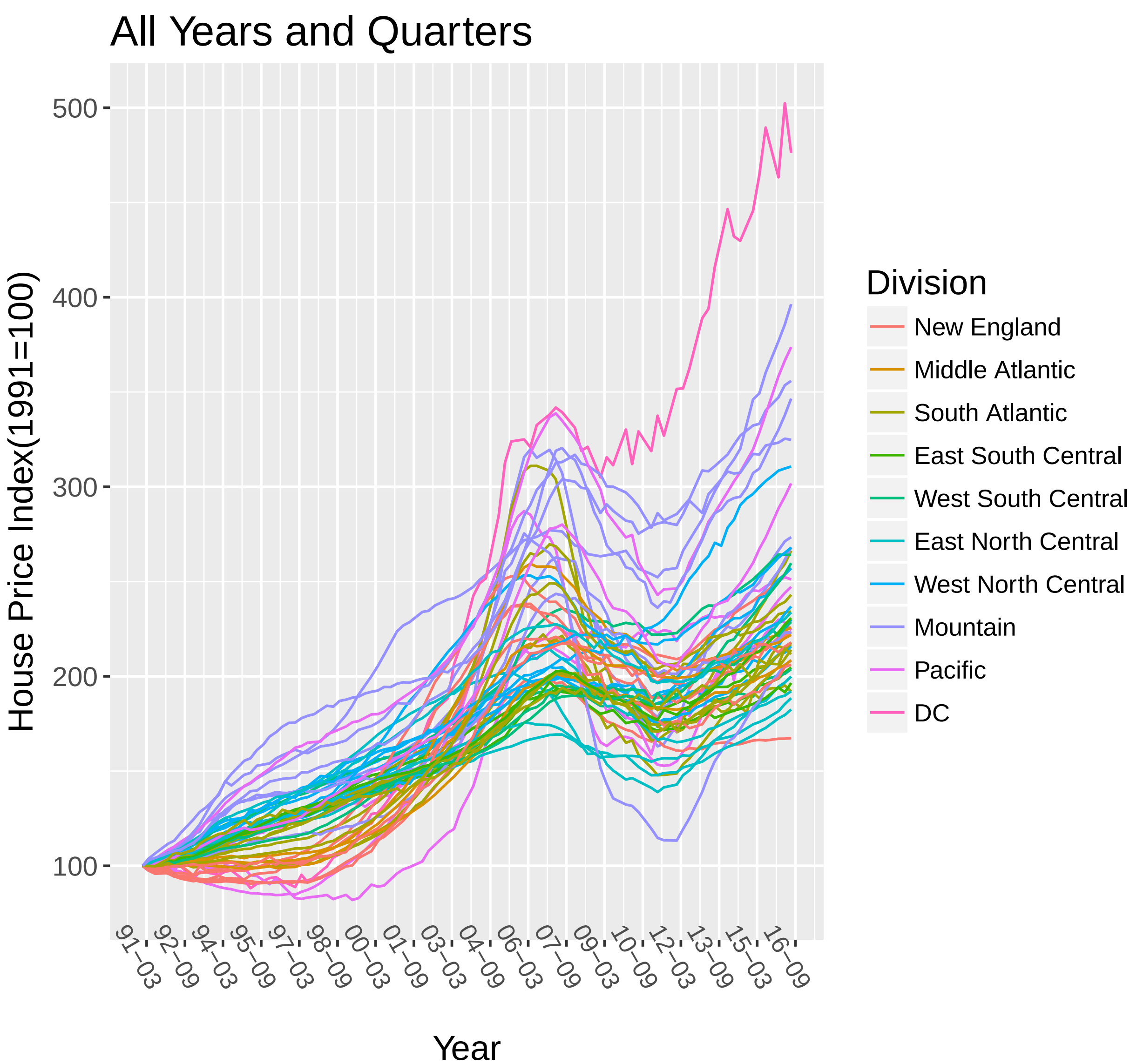
```
#Sum of Each Year-by 5 Region
ggparcoord(df.states2, columns = c(2:(length(df.states2)-3)), groupColumn = 25, scale = "globalminmax",
           missing = "exclude") + theme(legend.position = "right", text=element_text(size=14),legend.te
```



Finally, I also made a parallel coordinate plot for all four quarters in the period of 1991-2016 by division. DC still has the highest HPI and mountain and pacific areas have relatively high HPI during these years. Besides, different quarters does not make much difference within a year therefore mean value of HPI for each year is enough for analyzing the trend.

```
#A plot of All Years #Use lubridate Date
ggplot(STATES.merge,aes(Date,PurchaseOnlyIndex,group=Region))+geom_line(aes(color=Division))+
   scale_x_date(breaks = date_breaks("18 month"), labels = date_format("%y-%m"))+
  theme(legend.position = "right", text=element_text(size=14),legend.text = element_text(size =10),axis
  ggtitle("All Years and Quarters")
```

```
## Warning: Removed 950 rows containing missing values (geom_path).
```



Analyzing HPI changes of 50 US states and DC

I then analyzed percentage change of HPI (Purchase Only Index) over previous quarter, previous year, past 5 years and since the first quarter of 1991. D.C and USA result is also included in the analysis. Period ended is 2016-Q3. In dataset US-HPI, since the region "USA" has leading and trailing whitespaces, we need to trim it off to combine US HPI data with US map data.

```
trim <- function (x) gsub("^\\s+|\\s+$", "", x) #returns string w/o leading or trailing whitespace
us$Region<-as.factor(trim(as.character(us$Region)))
combPOI<-rbind(select(us,Region,Year,Quarter,PurchaseOnlyIndex),
            select(STATES,Region,Year,Quarter,PurchaseOnlyIndex))

combPOI.expand<-merge(combPOI,class,id="Region")
comb.state<-dcast(combPOI.expand,Region~Year+Quarter,mean,value.var="PurchaseOnlyIndex")
#Use Mean of several states to indicate a Division
comb.division<-dcast(combPOI.expand,Division~Year+Quarter,mean,value.var="PurchaseOnlyIndex")
```

```r
#Define a Function to get HPI changes
Change_HPI<-function(data){
  mutdata<-mutate(data,
                  OneQtr=100*(data[,length(data)]-data[,length(data)-1])/data[,length(data)-1],
                OneYr=100*(data[,length(data)]-data[,length(data)-4*1])/data[,length(data)-4*1],
                FiveYr=100*(data[,length(data)]-data[,length(data)-4*5])/data[,length(data)-4*5],allYr=
  return(mutdata)
}

#HPI Change in Division
mutcomb.division<-Change_HPI(comb.division)
mutcomb.division<-mutcomb.division[,c(1,(length(comb.division)+1):length(mutcomb.division))]
```

We can then sort the HPI changes grouped by division over, for example, the past 1 year.

```r
#Sort the HPI changes over the past year: desc(OneYr)
mutcomb.sort<-arrange(mutcomb.division,desc(OneYr))
top<-mutcomb.sort$Division[1]
second<-mutcomb.sort$Division[2]
third<-mutcomb.sort$Division[3]
```

The sorted result shows that, Over the past one year from Q3 of 2015 to Q3 of 2016, the HPI change in Pacific area increases significantly. The second highest region is Mountain area. These two regions carry higher NPI increase than that of the whole country. States in Pacific regions are:Alaska,California,Hawaii,Oregon,Washington and states in mountain regions are: Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah and Wyoming. This makes sense since we know that housing prices in Califonia and Alaska are always expensive as well as in its surrounding areas such as Arizona and Nevada. This also indicates that over the past year, housing maket has come back to life originated from the west coast and pacific areas.

```r
pander::pander(head(mutcomb.sort,10))
```

Division	OneQtr	OneYr	FiveYr	allYr
Pacific	0.8744	6.54	39.52	179.6
Mountain	1.366	6.238	38.98	199.6
East North Central	1.532	5.694	22.84	99.35
East South Central	1.696	5.39	19.79	112.4
South Atlantic	0.9193	5.094	25.62	124.5
West North Central	1.267	5.063	24.95	148.8
West South Central	1.185	4.176	23.99	139.1
New England	0.7388	3.469	13.9	115.9
Middle Atlantic	0.977	3.129	11.33	118.8
DC	-5.21	-2.729	45.64	376.2

```r
pander::pander(print(class$Name[class$Division==as.character(top)], max.levels=0))
```

[1] Alaska California Hawaii Oregon Washington *2*, *5*, *11*, *37* and *47*

```r
pander::pander(print(class$Name[class$Division==as.character(second)], max.levels=0))
```

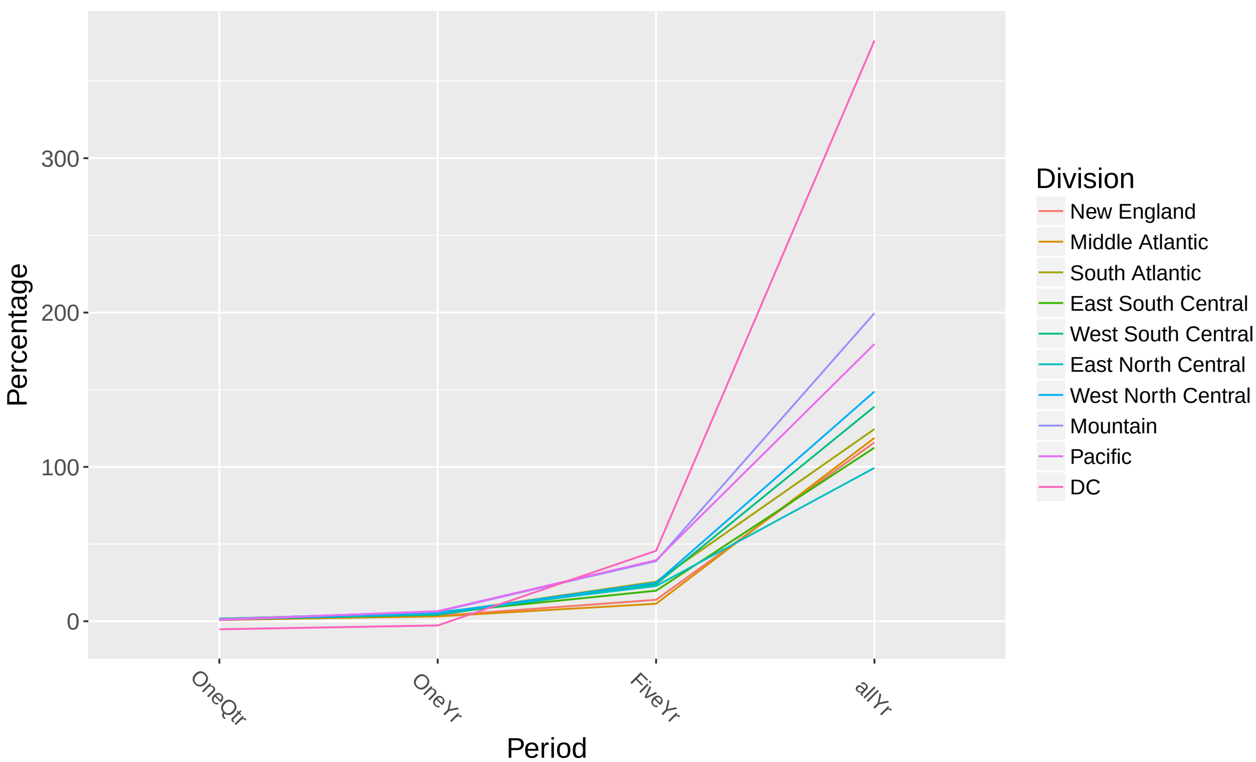
[1] Arizona Colorado Idaho Montana Nevada New Mexico [7] Utah Wyoming
3, *6*, *12*, *26*, *28*, *31*, *44* and *50*

```
pander::pander(print(class$Name[class$Division==as.character(third)], max.levels=0))
```

[1] Illinois Indiana Michigan Ohio Wisconsin *13*, *14*, *22*, *35* and *49*

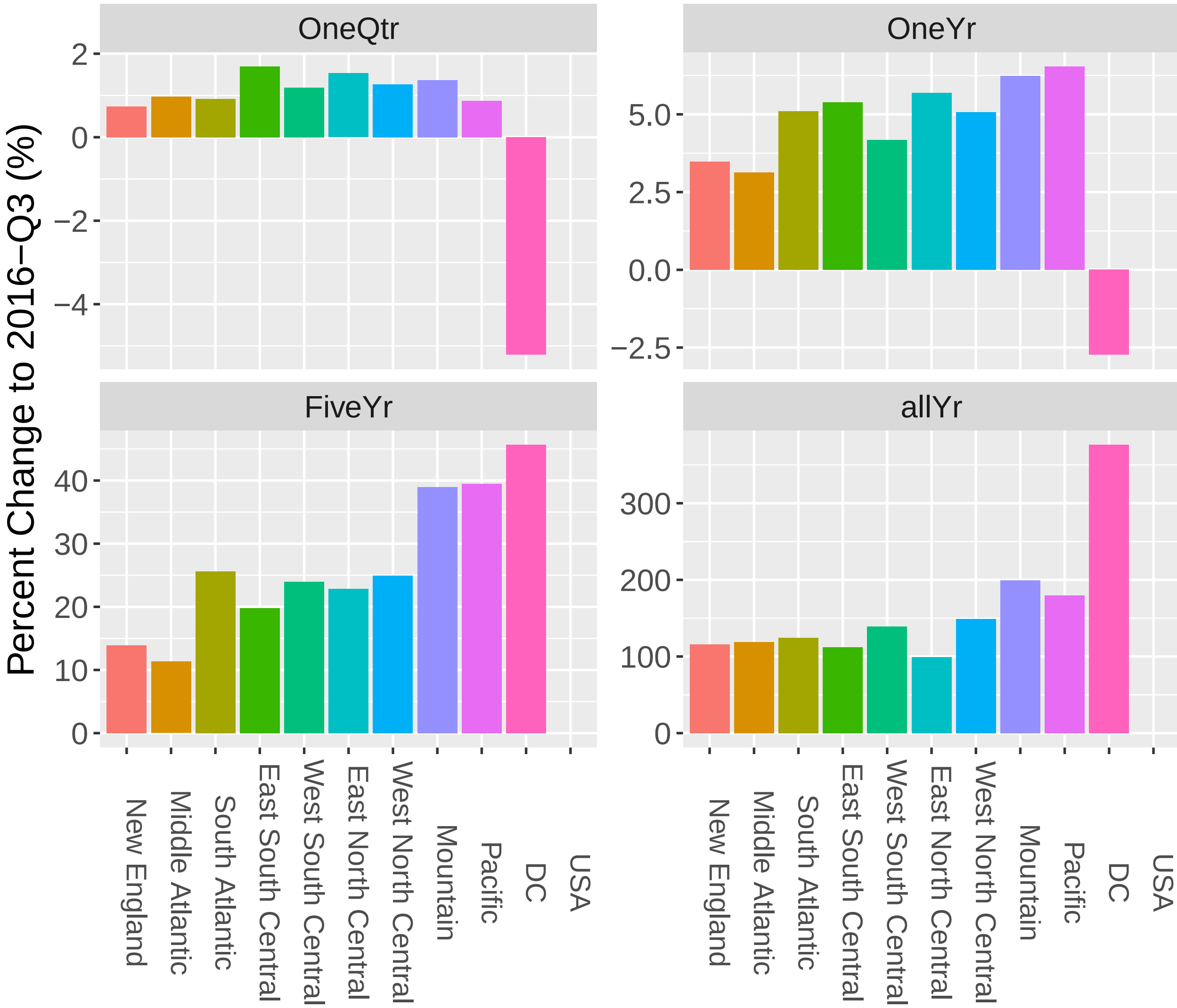
Take a look at HPI percentage changes for all divisions in the given time span: change over previous quarter, over previous year, over past five years and over all years since 1991-Q1. We found that over the past five years, HPI change is very small compared to changes over all years. This could be explained by the fact that the US has gone through an economy turndown since 2008. However, for the past 20 years, the Housing Price Index has increased by almost 100 percent for around half of the US states and HPI surge in DC has even increased by 300 percent.

```
ggparcoord(mutcomb.division, columns = c(2:length(mutcomb.division)), groupColumn = 1, scale = "globalmi
           missing = "exclude") + theme(legend.position = "right", legend.text = element_text(size=10))+
 theme(text = element_text(size=16),legend.text=element_text(size=12),
       axis.text.x = element_text(size=12,angle=-45,vjust=0.5))
```



We could also use bar plot to visualize the HPI inrease during different period of time. The upper left panel in the following result shows that for the past quarter, from 2016-Q2 to 2016-Q3, HPI decrease of DC is significant while HPI of mountain and pacific area has increased most. This is unusual since HPI of DC is always increasing and is always higher than other part of US states. The upper right panel is HPI changes over the past year for all divisions. Surprisingly, we found that the HPI increase is at least 2.5% during 2015-2016. This shows that housing market is starting to turning around in most areas in US. In Five-Year range in the lower left panel, most parts of US have low HPI percentage change (or increase) since the country has been just recovered from a crisis happened years ago and HPI was coming back slowly. From the figure I can conclude that the housing in mountain and pacific areas has been recovering most in recent years.

```
mutcomb.division.melt<-melt(mutcomb.division,id="Division")
ggplot(mutcomb.division.melt,aes(x=Division,y=value,group=variable))+geom_bar(stat="identity",aes(fill=
  theme(legend.position="None",text = element_text(size=16),
       axis.text.x = element_text(size=12,angle=-90, vjust=0),axis.title.x=element_blank())
```

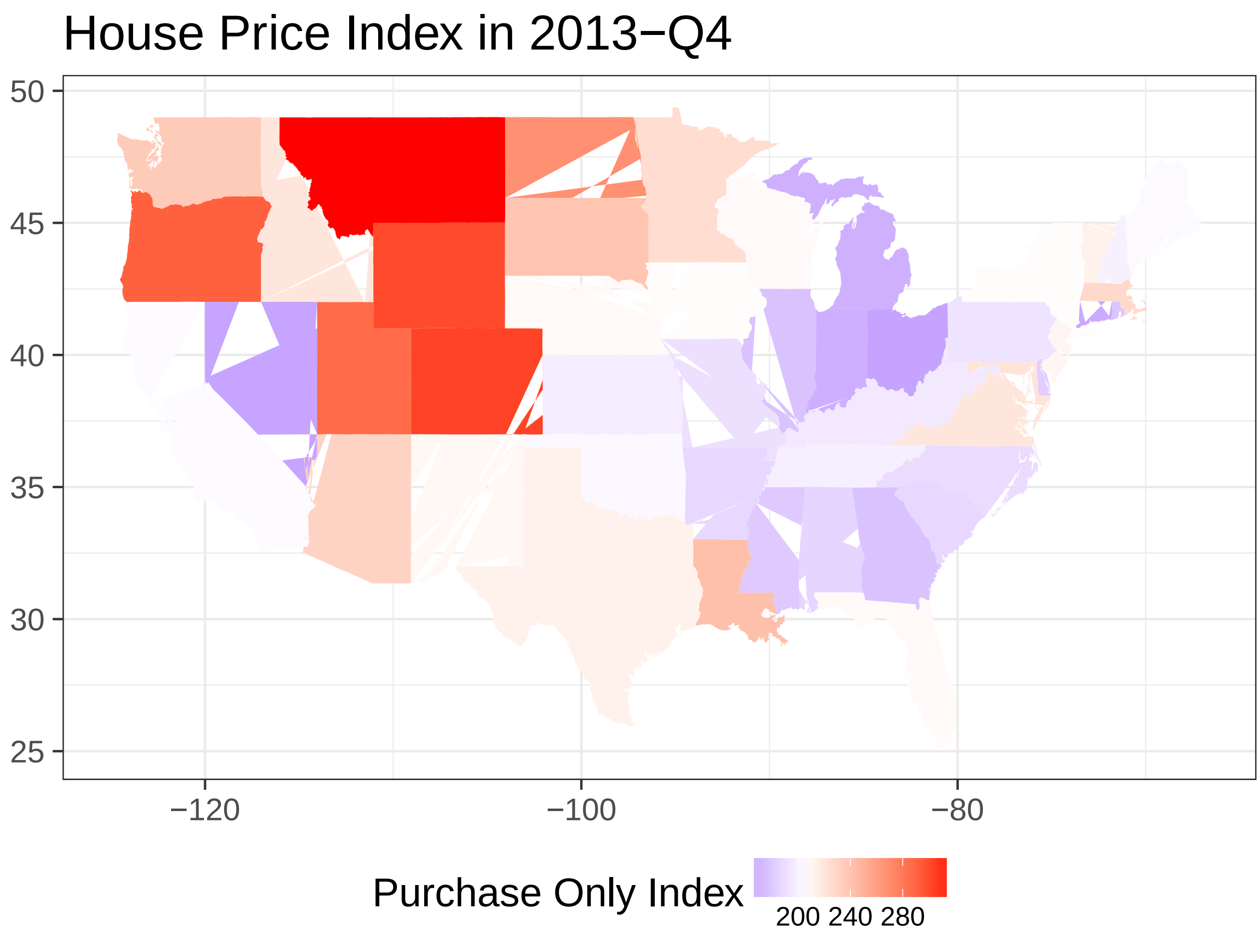


Visualizing HPI changes of 50 US states and DC

The HPI absolute value and HPI change for each state could be projected onto US map for better visualization. The following figure is a plot of HPI value for the most recent period 2016-Q3 for each US state on US Map. On the plot, we use HPI index of USA as the value of midpoint (white color on the map). Red represents high value and blue represents low value. In the most recent months, HPI of Montana, Wyoming, Colorado and Oregon have the highest HPI while Indiana, Michigan and Ohio have the lowest.

```
mapdf <- map_data("state")
combPOI.expand$Name<-as.factor(tolower(as.character(combPOI.expand$Name)))
mapdf$region<-as.factor(mapdf$region)
#Get Latest Year Subset dataset changes Only
subset<-select(filter(combPOI.expand,Year==2013 & Quarter==4),Name,PurchaseOnlyIndex)
colnames(subset)<-c("region","value")
midvalue=filter(subset,region=="usa")$value  #USE USA value as midvalue
mapdf2<-merge(subset,mapdf,by="region")
```

```
ggplot(mapdf2) +
  geom_polygon(aes(x=long, y=lat,group = group,order=order,fill=value))+
  scale_fill_gradient2("Purchase Only Index",low ="blue", mid = "#FFFFFF", high ="red",
                      midpoint = midvalue, space = "rgb", guide = "colourbar")+theme_bw(base_size=18)+t
```



Next, we also made plots of HPI changes for each state in US over a certain period of time and then project the percentage change on US map. The mapping code is written into a function and we can choose time span of interest, such as one quarter, one year, five years, etc.

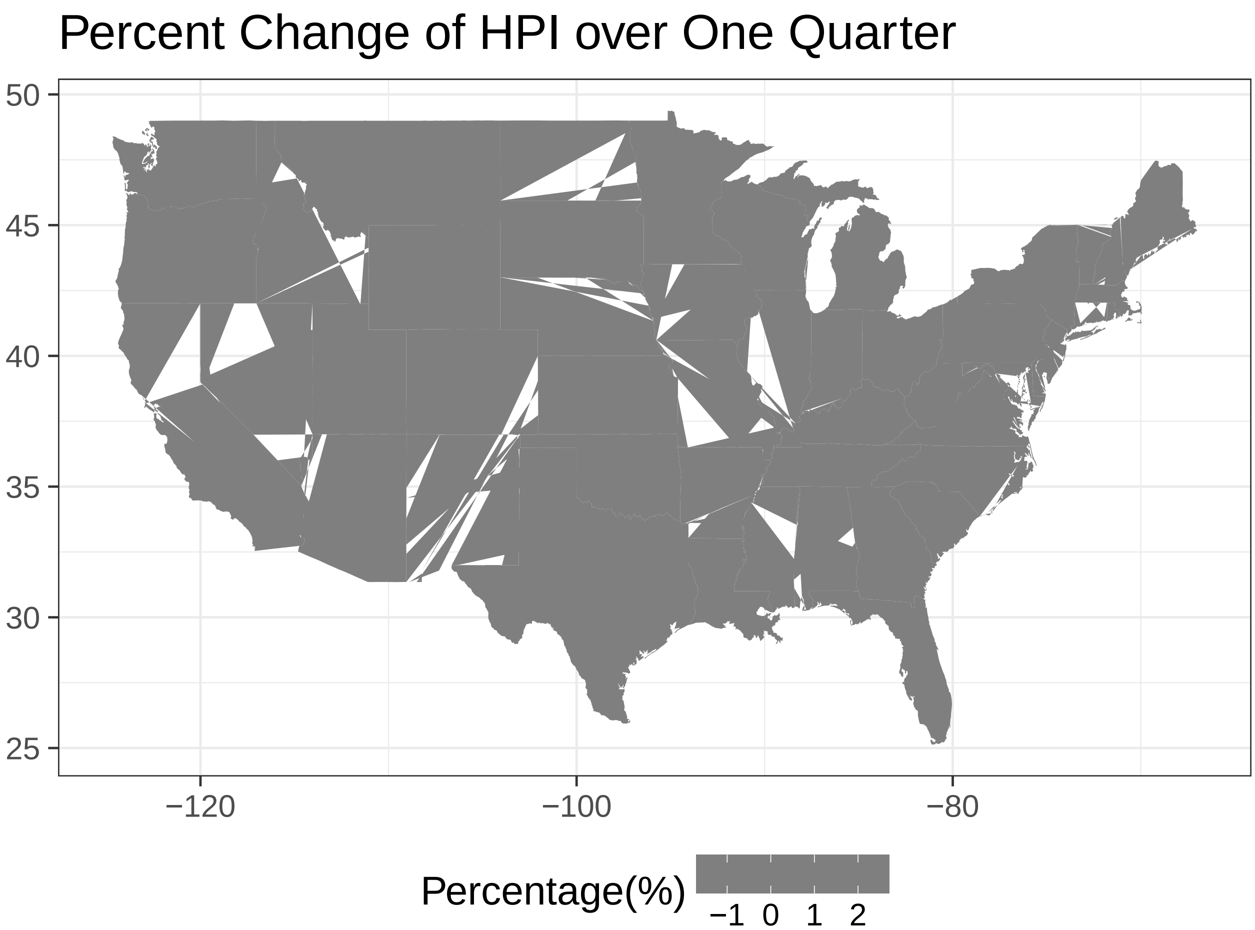
```
mutcomb.state<-Change_HPI(comb.state)
mutcomb.state<-mutcomb.state[,c(1,(length(comb.state)+1):length(mutcomb.state))]
mapdf <- map_data("state")
mutcomb.state.merge<-merge(mutcomb.state,class,by="Region")
mutcomb.state.merge$Name<-as.factor(tolower(as.character(mutcomb.state.merge$Name)))
mapdf$region<-as.factor(mapdf$region)
```

```
#Define a function to Plot HPI changes on US map during a certain period of time
plotChange_HPI<-function(subset,title){
  colnames(subset)<-c("region","value")
  midvalue=filter(subset,region=="usa")$value  #USE USA value as midvalue
  mapdf2<-merge(subset,mapdf,by="region")
  ggplot(mapdf2) +
  geom_polygon(aes(x=long, y=lat,group = group,order=order,fill=value))+
  scale_fill_gradient2("Percentage(%)",low ="blue", mid = "#FFFFFF", high ="red",
                      midpoint = midvalue, space = "rgb", guide = "colourbar")+theme_bw(base_size=18)+t
```

```
  ggtitle(title)
  }
```

The figure below is HPI change over one quarter during 2013-Q3 and 2013-Q4. Nevada, Arizona and Hawaii are the top three regions that have the largest HPI change in this period of time, with more than 3% of HPI increase.HPI change for overall USA is +1.20% and there are 18 states that has higher HPI change than the average HPI change for USA.

```
subset<-select(mutcomb.state.merge,Name,OneQtr)
title<-"Percent Change of HPI over One Quarter"
plotChange_HPI(subset,title)
```

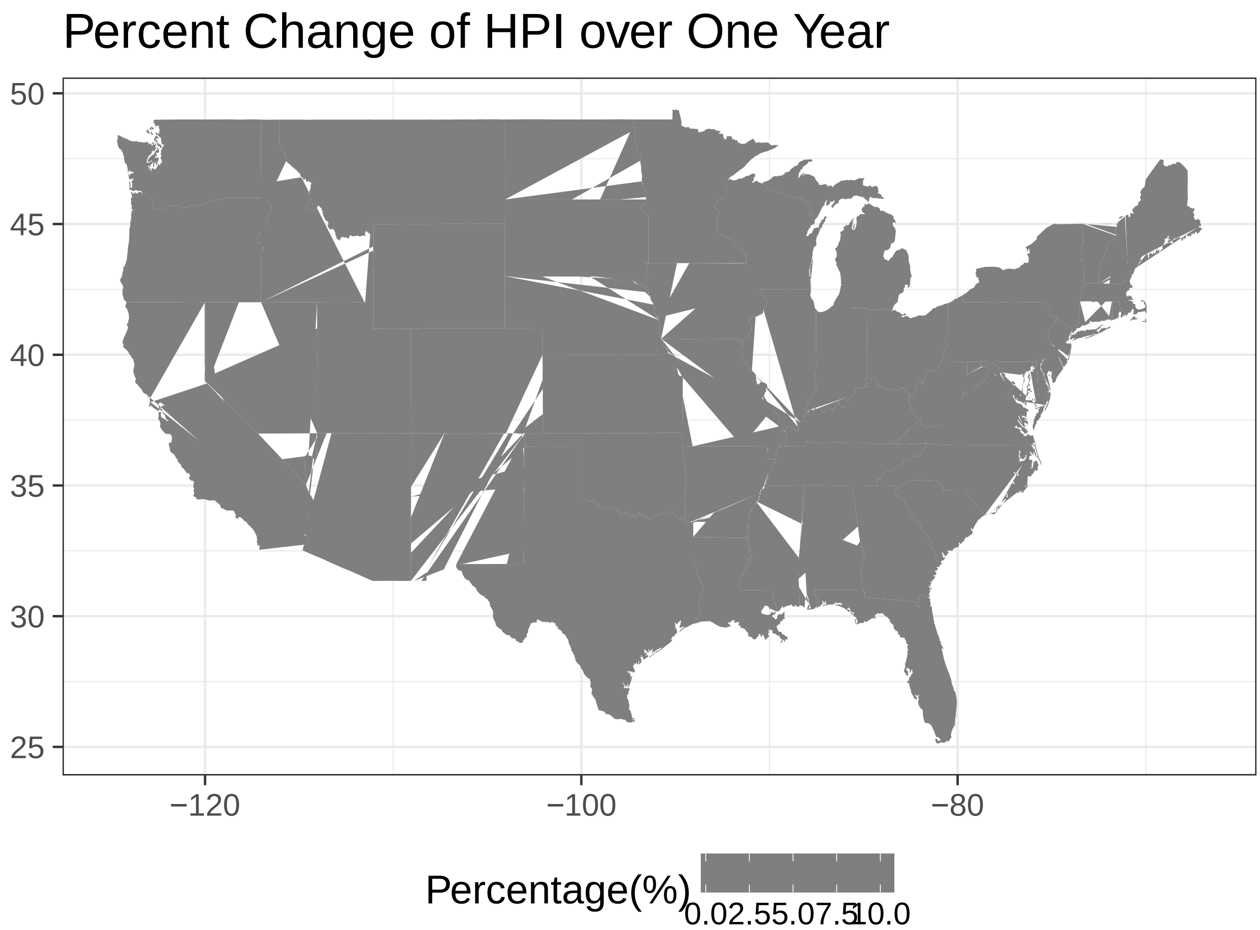



```
mutcomb.sort<-arrange(mutcomb.state.merge,desc(OneQtr))
pander::pander(head(mutcomb.sort))
```

Region	OneQtr	OneYr	FiveYr	allYr	Class	Name	Division
CO	2.79	10.03	54.76	296.4	West	colorado	Mountain
FL	2.707	10.73	57.57	166.3	South	florida	South Atlantic
UT	2.593	9.529	44.64	246.5	West	utah	Mountain
MS	2.558	4.157	14.36	96.46	South	mississippi	East South Central
ID	2.524	8.899	45.89	168.1	West	idaho	Mountain
WA	2.378	10.44	46.09	201.8	West	washington	Pacific

As for one year duration for the whole year 2013, HPI change of Nevada, Califonia and Arizona rank the top three among all other states. The HPI increase is 24.3%, 19.5% and 15.2% respectively.

```
subset<-select(mutcomb.state.merge,Name,OneYr)
title<-"Percent Change of HPI over One Year"
plotChange_HPI(subset,title)
```



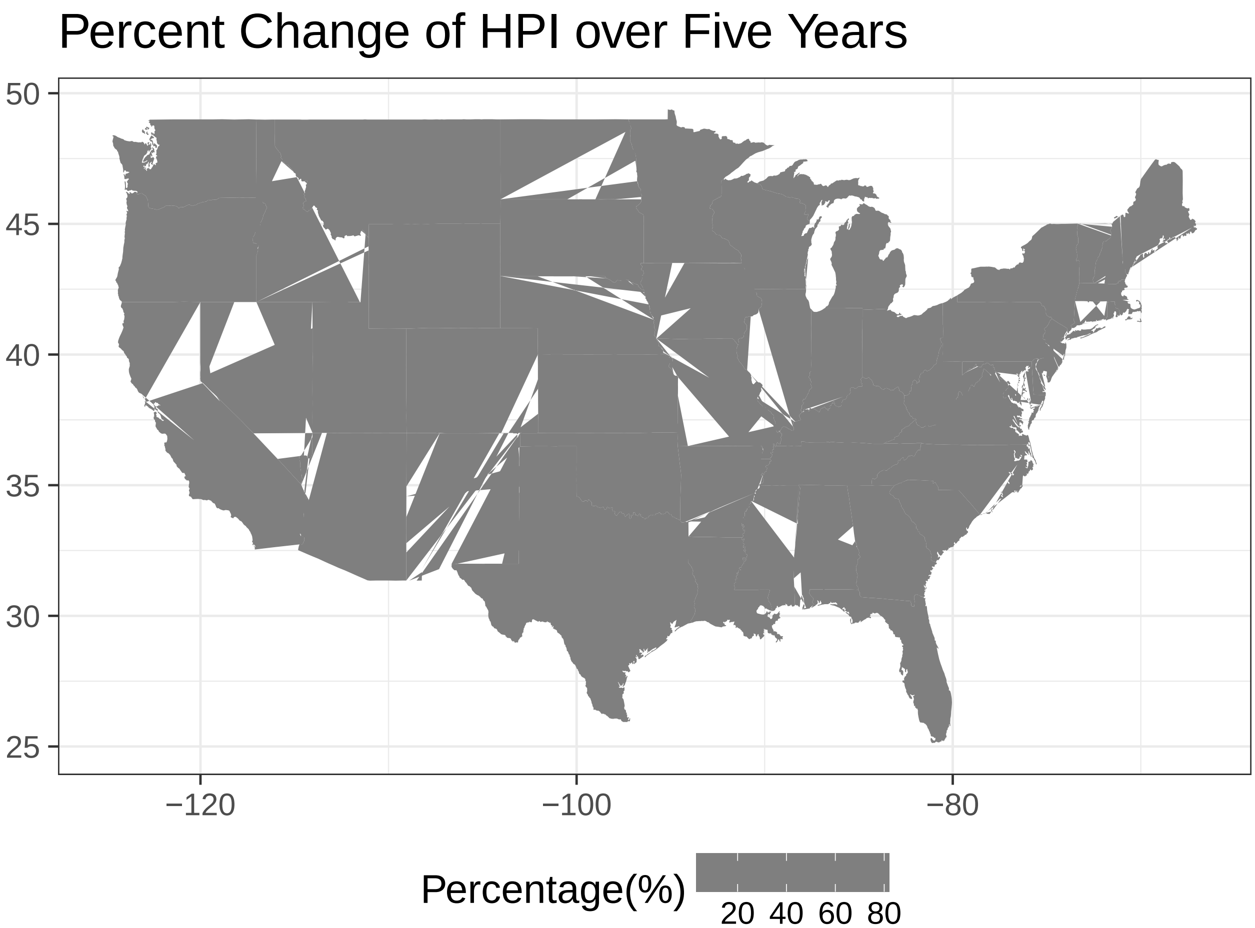

```
mutcomb.sort<-arrange(mutcomb.state.merge,desc(OneYr))
pander::pander(head(mutcomb.sort))
```

Region	OneQtr	OneYr	FiveYr	allYr	Class	Name	Division
FL	2.707	10.73	57.57	166.3	South	florida	South Atlantic
OR	1.951	10.45	51.99	273.7	West	oregon	Pacific
WA	2.378	10.44	46.09	201.8	West	washington	Pacific
CO	2.79	10.03	54.76	296.4	West	colorado	Mountain
UT	2.593	9.529	44.64	246.5	West	utah	Mountain
ID	2.524	8.899	45.89	168.1	West	idaho	Mountain

For HPI change over a long span 5 years since 2008,Washington D.C, North Dakota and Califonia rank top three, with HPI increase of 27.8%, 27.7% and 19.4%. The amount is surprisingly almost the same as one year increase of HPI in 2013 for West coast.

```
subset<-select(mutcomb.state.merge,Name,FiveYr)
title<-"Percent Change of HPI over Five Years"
```

```
plotChange_HPI(subset,title)
```

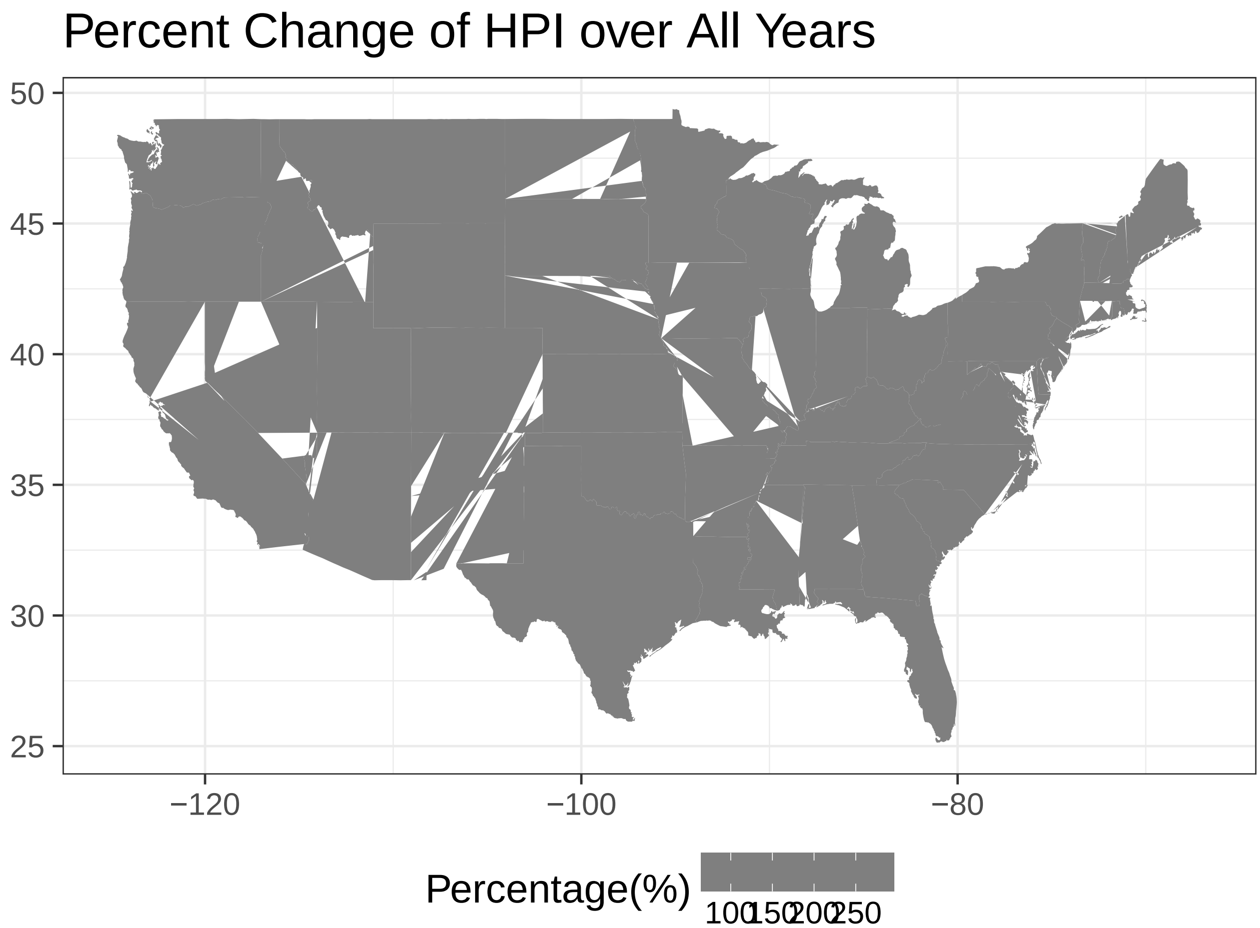



```
mutcomb.sort<-arrange(mutcomb.state.merge,desc(FiveYr))
pander::pander(head(mutcomb.sort))
```

Region	OneQtr	OneYr	FiveYr	allYr	Class	Name	Division
NV	1.073	7.823	83.51	108.3	West	nevada	Mountain
AZ	0.8296	6.662	66.26	173.4	West	arizona	Mountain
CA	1.324	7.213	61.83	147.2	West	california	Pacific
FL	2.707	10.73	57.57	166.3	South	florida	South Atlantic
CO	2.79	10.03	54.76	296.4	West	colorado	Mountain
OR	1.951	10.45	51.99	273.7	West	oregon	Pacific

During 20 years since 1991, HPI change increase most in Washington D.C area, followed by Montana and Colorado.

```
subset<-select(mutcomb.state.merge,Name,allYr)  #OneQtr
title<-"Percent Change of HPI over All Years"
plotChange_HPI(subset,title)
```




```
mutcomb.sort<-arrange(mutcomb.state.merge,desc(allYr))
head(mutcomb.sort)
```

```
##   Region     OneQtr      OneYr   FiveYr  allYr Class     Name Division
## 1     DC -5.2095153 -2.7290926 45.64306 376.18    DC       dc       DC
## 2     CO  2.7903841 10.0324792 54.75949 296.37  West colorado Mountain
## 3     OR  1.9505142 10.4471437 51.98666 273.72  West   oregon  Pacific
## 4     MT  0.6589926  4.5934111 26.08049 255.90  West  montana Mountain
## 5     UT  2.5934040  9.5293783 44.63876 246.54  West     utah Mountain
## 6     WY -0.1813376  0.9292063 15.42863 224.77  West  wyoming Mountain
```

Housing Price Index and Per Capita Personal Income for 25 Metropolitan Cities

Next, relationship of HPI change and Per Capita personal income is analyzed. The first thing to do is to download Per Capita Personal Income for 25 metropolitan cities on Bereau of Economic Analysis.(http://www.bea.gov) The exact steps to download the data of interest are described in section2(4). Income change is preferred and we are only interested in the past five years and 25 biggest cities as well. One thing to note is that on this website, the latest income change is from 2011-2012. Since defining of regions or areas is quite messy on different websites, I matched the city and county.etc in the downloaded dataset with information of 25 biggest cities. Then we need to preprocess the us.cities data by adding D.C entry. Here I used Arlignton,VA as the lat and lon of D.C since Arlington is quite close to D.C.

```
#Data1-US Cities:
data(us.cities)
us.cities$country.etc <- factor(us.cities$country.etc)
n <- nchar(us.cities$name)
us.cities$name <- substr(us.cities$name, 1, n-3) #Extract City Names
addDC<-data.frame(us.cities[us.cities$name=="Arlington" & us.cities$country.etc=="VA",])
addDC[,1:2]<-c("DC","DC")#ABOVE ALL: Use arlington,VA as DC value
us.cities<-rbind(us.cities,addDC)
pander::pander(tail(us.cities,2))
```

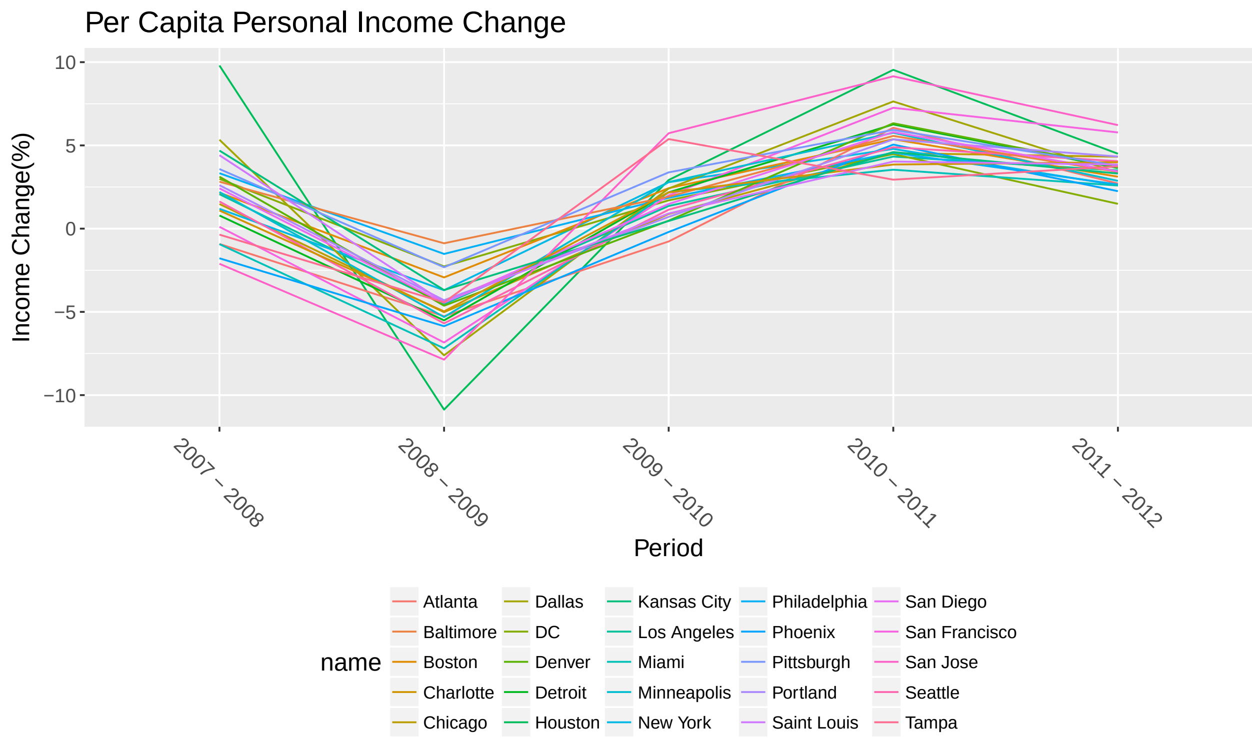
	name	country.etc	pop	lat	long	capital
1005	Yuma	AZ	85720	32.68	-114.6	0
3610	DC	DC	184603	38.88	-77.1	0

```
#Data2-Income.Metro:
income<-read.csv("dataset/metroIncome.csv")
income<-income[,-1]
colnames(income)[3:7]<-c("income1","income2","income3","income4","income5")
pander::pander(head(income))
```

name	country.etc	income1	income2	income3	income4	income5
Chicago	IL	1.48	-5.02	0.88	4.54	4.32
Houston	TX	9.8	-10.87	2.92	9.54	4.5
Los Angeles	CA	2.07	-4.53	1.36	4.33	3.5
New York	NY	1.19	-3.7	2.79	4.79	2.6
DC	DC	2.99	-2.27	1.7	4.48	1.49
Atlanta	GA	-0.92	-5.26	-0.77	6.06	2.71

The following figure is a parallel coordinate plot of Personal Income Change during the past years.Since year 2007, the personal income change first decreases to negative and then increases. During the period of 2008-2009, income change rate reaches lowest-below zero for all 25 biggest cities. However during these years, personal income change rate speeds up and exceeds the change rate before economy turndown.

```
ggparcoord(income, columns = c(3:7), groupColumn = 1, scale = "globalminmax",
           missing = "exclude") + theme(legend.position = "bottom", text=element_text(size=14),legend.t
  scale_x_discrete("Period",labels=c("2007 - 2008", "2008 - 2009", "2009 - 2010","2010 - 2011","2011 - 
```




```
#Load cleaned metro data and combine it with income data as well as the US cities coordinate data.
metro<-read.csv("dataset/HPI-25metro-clean.csv",header=TRUE)
colnames(metro)[3:8]<-c("seven","eight","nine","ten","eleven","twelve")
metrohpi<-mutate(metro,"hpi1"=100*(eight-seven)/seven,"hpi2"=100*(nine-eight)/eight,
                      "hpi3"=100*(ten-nine)/nine,"hpi4"=100*(eleven-ten)/ten,"hpi5"=100*(twelve-eleven)/
metrohpi<-metrohpi[,-c(3:8)]
#Combine HPI and Income
income.hpi<-cbind(income, metrohpi[,-c(1:2)])
income.hpi.map<-merge(income.hpi,us.cities,id=c("name","country.etc"))
pander::pander(head(income.hpi.map))
```

Table 13: Table continues below

name	country.etc	income1	income2	income3	income4	income5
Atlanta	GA	-0.92	-5.26	-0.77	6.06	2.71
Baltimore	MD	2.82	-0.88	1.92	5.57	3.41
Boston	MA	2.16	-2.93	2.43	5.37	3.12
Charlotte	NC	1.09	-4.97	2.18	3.85	3.99
Chicago	IL	1.48	-5.02	0.88	4.54	4.32
Dallas	TX	5.34	-7.61	2.41	7.64	3.38

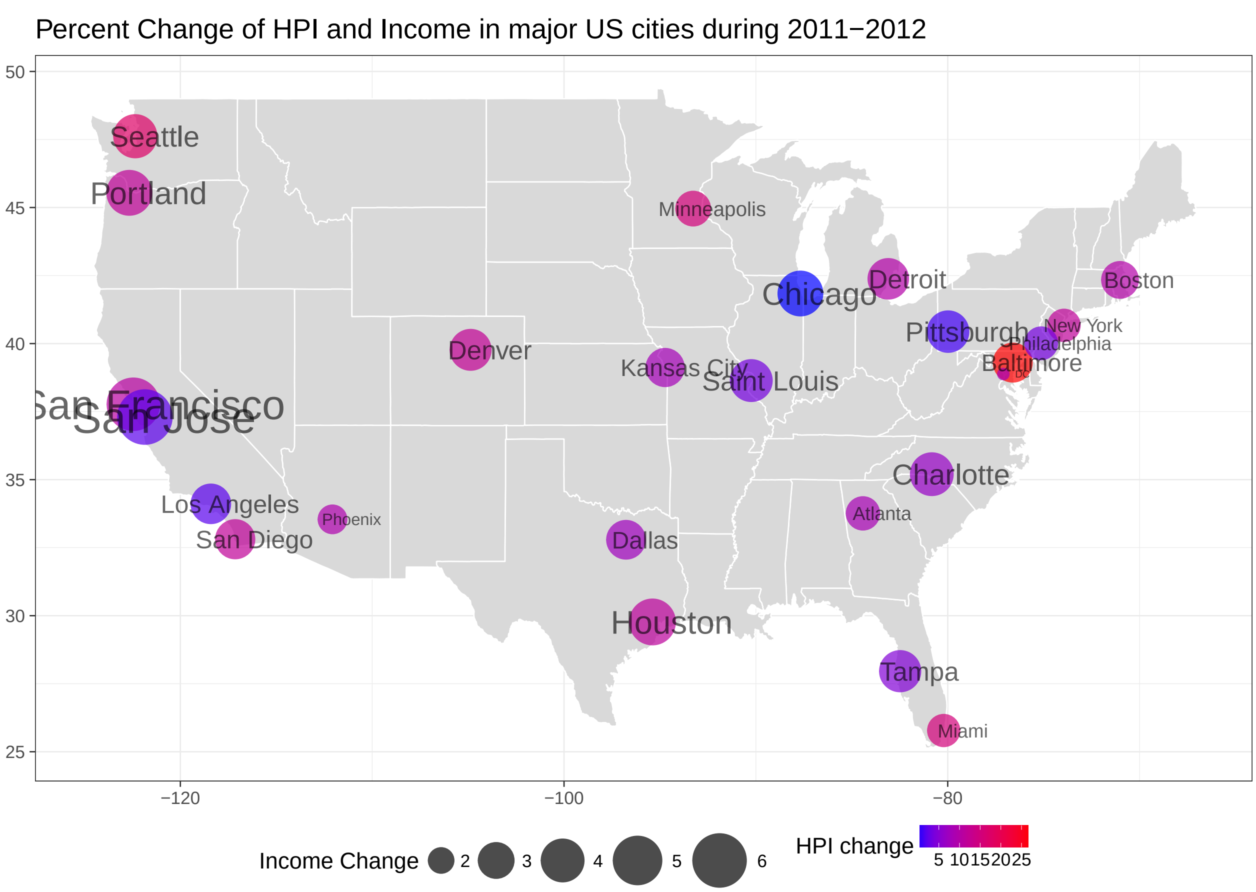
hpi1	hpi2	hpi3	hpi4	hpi5	pop	lat	long	capital
-0.08107	0.5911	-3.866	0.6892	8.535	424096	33.76	-84.42	2
-27.53	-12.84	-14.53	2.088	26.64	602658	39.3	-76.61	0
-14.07	-1.393	-1.905	-1.291	9.924	567759	42.34	-71.02	2
-6.694	0.6064	-7.346	-3.451	6.556	607111	35.2	-80.83	0
-6.738	-4.473	-1.088	-4.94	0.2882	2830144	41.84	-87.68	0
-3.375	0.9581	-3.641	-6.889	7.553	1216543	32.79	-96.77	0

After loading the dataset, processed and combined them, we could map HPI change and Per Capita Income change on US map for a time period such as 2011-2012.The following figure shows percent change of HPI and income in year 2011-2012 shows the relationship of HPI and income changes in these large cities.Blue represents low HPI change and red represents high HPI change. The change in income over one year is represented by size of points. The bigger the points, the higher the increase of Per Capita Personal Income. We can tell that Personal income increases most in west coast such as San Francisco and San Jose with relatively low HPI increase.The HPI increases most in Baltimore,MD which is near DC area with an increase of 26.6%.

```
mapdf <- map_data("state")
map<-ggplot(mapdf)+geom_polygon(aes(x=long, y=lat,group = group,order=order),fill=I("grey85"),
                                color=I("white"),size=0.5)+theme_bw()

labeldata<- data.frame(x2=income.hpi.map$long, y2=income.hpi.map$lat,y3=income.hpi.map$income5,
                       texthere=income.hpi.map$name)

map+geom_point(data=income.hpi.map,mapping=aes(x=long, y=lat,color=hpi5,size=income5),alpha=I(0.7))+
scale_colour_gradient("HPI change",low=("blue"),high=("red"))+
  scale_size_continuous("Income Change",range=c(4,20))+
 theme_bw(base_size=18)+
  theme(legend.position="bottom",axis.title.x=element_blank(),axis.title.y=element_blank())+
annotate("text",x=labeldata$x2+1,y=labeldata$y2,label=as.character(labeldata$texthere),
         size=2*labeldata$y3,
         alpha=I(0.6))+ggtitle("Percent Change of HPI and Income in major US cities during 2011-2012")
```

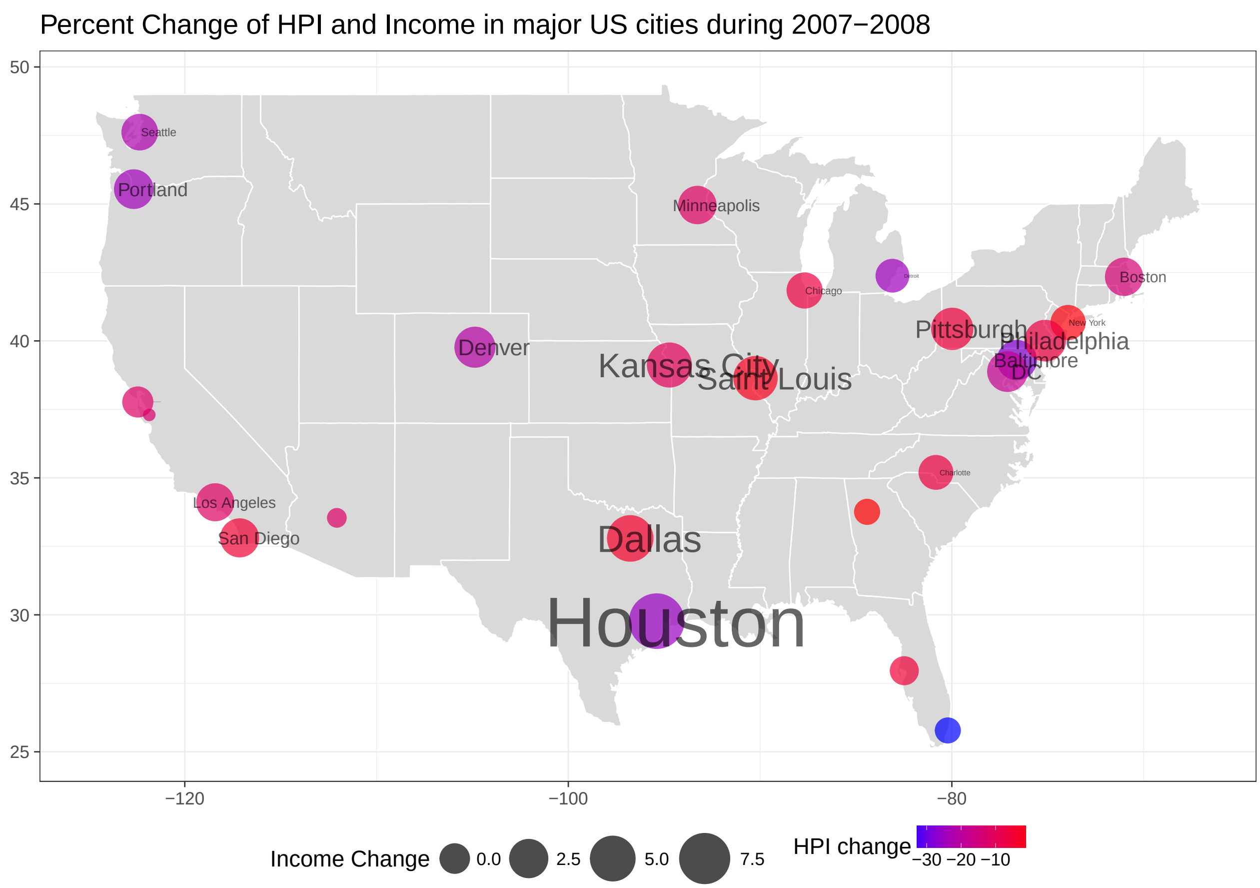


A glimpse of HPI and income change during year 2007-2008 is shown below. Compared to last two years' boom, HPI decreased drastically in year 2007-2008 and personal income increase was quite low in most areas. Houston has the highest HPI decrease rate of -25.3% although its income increase is 9.80%, highest among other areas. Most areas have undergone an HPI decrease which marks the fell down of housing market despite that some popular areas such as San Diego and New York were not significantly affected by the economy turn down.

```
mapdf <- map_data("state")
map<-ggplot(mapdf)+geom_polygon(aes(x=long, y=lat,group = group,order=order),fill=I("grey85"),
                                  color=I("white"),size=0.5)+theme_bw()

labeldata<- data.frame(x2=income.hpi.map$long, y2=income.hpi.map$lat,y3=income.hpi.map$income1,
                          texthere=income.hpi.map$name)

map+geom_point(data=income.hpi.map,mapping=aes(x=long, y=lat,color=hpi1,size=income1),alpha=I(0.7))+
scale_colour_gradient("HPI change",low=("blue"),high=("red"))+
  scale_size_continuous("Income Change",range=c(4,20))+
 theme_bw(base_size=18)+
  theme(legend.position="bottom",axis.title.x=element_blank(),axis.title.y=element_blank())+
annotate("text",x=labeldata$x2+1,y=labeldata$y2,label=as.character(labeldata$texthere),
         size=2*labeldata$y3,
         alpha=I(0.6))+ggtitle("Percent Change of HPI and Income in major US cities during 2007-2008")
```



This analysis on HPI and income change was also turned into a vivid Shiny App that can show the relationship of HPI change and income change in major US cities during a certain period of time of interest. The following two screenshots were taken from my Shiny App. Before running the app, we need to install packages including shiny, maps, ggmap, grid, and dplyr. In this easy App, there are three tabs. The first tab Plot shows HPI and personal income changes in a given period of time.On the left panel, user can select the period of time they are interested in. The visualization can also be adjusted by hiding/showing legend and labels as well as changing size and transparancy of points and label texts. The second tab Summary displays a summary of dataset for HPI and income during a selected period. We can know about min(max) of HPI and Income change. The third tab is a table view of the dataset, I used renderDataTable which is quite fancy to show results including sorting, change number of observations, search by keyword, etc.

5.Discussion and future work

More research to further investigate the statistical properties of repeat transactions house price indexes is needed. Many of these issues are discussed in the paper by Stephens et al. (1996). Topics for future research include: geographic and temporal aggregation, revision volatility, the use of appraisal values, sample selection, and comparisons with alternative methods.

This research project helps me to get a deeper knowledge and substantial practice in data technology skills. The skills applied in the project which I have learned in the MSDA program. After analyzing the datasets, I could conclude that during the past two years US economy has been turning around with a step-by-step rise in housing market. In this project, I have analyzed the HPI and HPI changes during different period of time as well as changes in per capita personal income at the meanwhile. I have decided to choose personal income rather than Gross Domestic(State) Product (mentioned in the proposal) in that personal income is more specific and better to reflect the purchase power of people in the big cities. In terms of HPI and HPI changes, Washington D.C stands out with highest HPI value as well as high HPI changes except for the 4th quarter in 2013. One year's HPI increase is quite much in most regions in US with Mountain and Pacific areas rank top. Compared to five years' period, HPI increase is fairly low below 10% in most areas except that D.C experienced an increase of 26%. This shows that US housing market has fell down since 2007-2008 and now it has just come back to life. In the coming 2016-2017 period, the housing price may be expected to keep rising and perhaps it is just the time to buy.